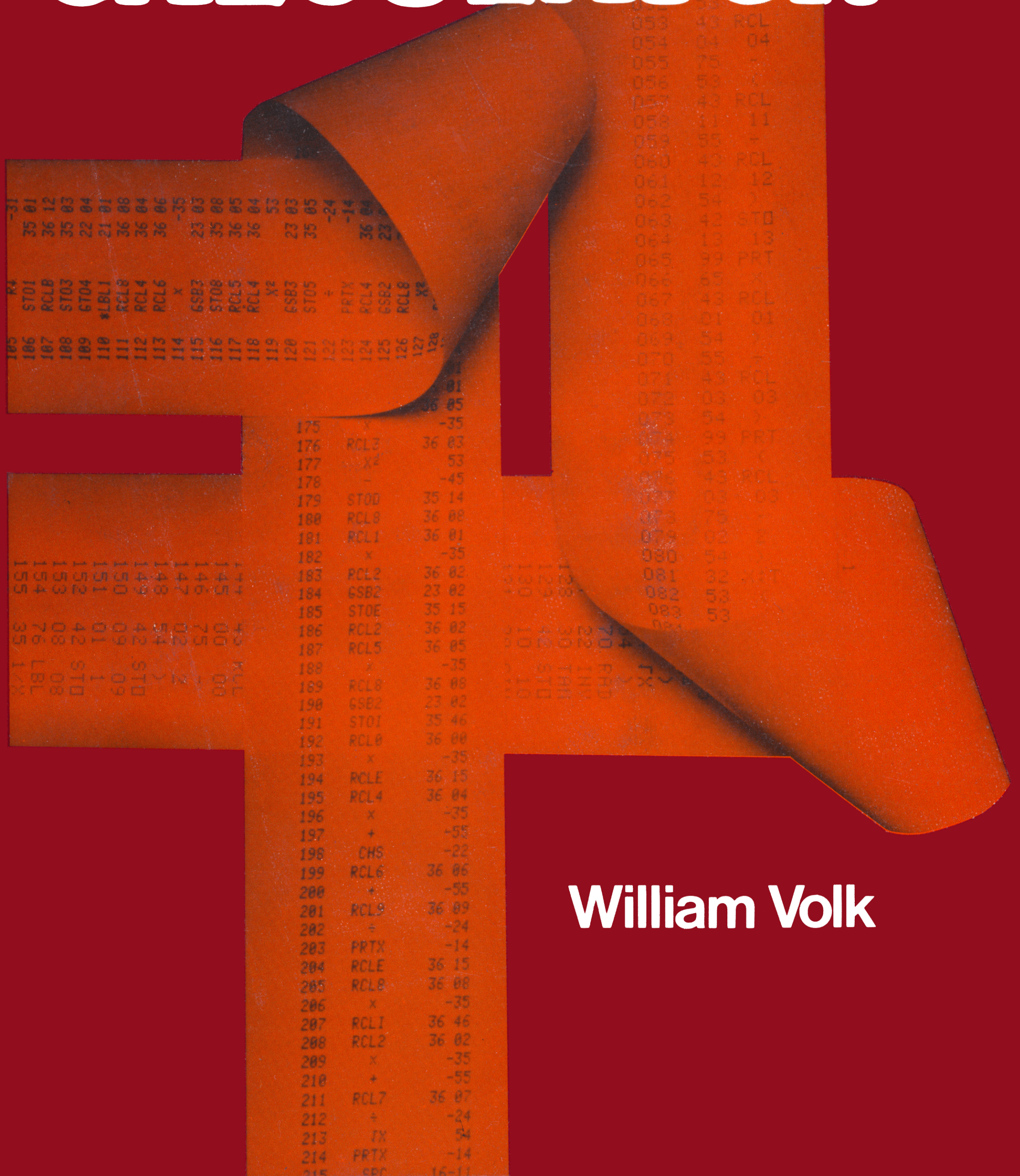
ENGINEERING STATISTICS with a PROGRAMMABLE CALCULATOR



William Volk

ISBN 0-07-067552-X

ENGINEERING STATISTICS with a PROGRAMMABLE CALCULATOR

By William Volk

There *are* other sources of calculator programs that carry out some of the statistical analyses detailed here. But none of the usual "statistical packages" was designed specifically for engineering problems, and none of them offers you—as this one does—

- programs that not only carry out statistical analyses but also calculate associated probabilities
- explanations of the particular functions employed
- programs in both RPN and AOS languages

In other words, this is no mere "cookbook." It is, rather, an all-round guide that enables you to calculate all statistical parameters and corresponding probabilities without reference to math tables. It provides enough theory and explanation to enable you to modify programs and solutions without referring to any other source. And because each program calculates the probability associated with the calculation, there is no need to refer to probability tables.

Armed with this unique guide you don't have to be a math expert to have your own statistical programs and to be in control of the size of the risks and the magnitude of the errors in your conclusions.

Each chapter deals with a different type of statistical procedure and provides: an explanation of the underlying probability function . . . a discussion of its application to engineering data . . . an outline of the types of problems to which it applies . . . and examples from typical engineering problems for each program.

In addition, programs are presented in detail with a line by line explanation as well as logic flow diagrams. And each is illustrated by one or more examples in both RPN and AOS languages—examples that you can use to check the accuracy of the programs in your calculator.

(continued on back flap)

William Volk

ENGINEERING STATISTICS WITH A PROGRAMMABLE CALCULATOR

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá Hamburg Johannesburg London Madrid Mexico Montreal New Delhi Panama Paris São Paulo Singapore Sydney Tokyo Toronto

Library of Congress Cataloging in Publication Data

Volk, William.
Engineering statistics with a programmable calculator.
Includes bibliographical references and index.
1. Engineering—Statistical methods. 2. Engineering
—Data processing. I. Title.
TA340.V66 519.5'02462 80–28153

2 3 4 5 6 7 8 9 0 KPKP 8 9 8 7 6 5 4 3 2

ISBN 0–07–067552–X

The editors for this book were Robert L. Davidson and James T. Halston, the designer was Elliot Epstein, and the production supervisor was Paul A. Malchow. It was set in Times Roman by The Kingsport Press.

Printed and bound by The Kingsport Press

CONTENTS

1
INTRODUCTION

1.1 OBJECT

The object of this book is to present those methods of statistics that will be helpful to the engineer in the analysis of experimental data. The statistical methods are presented from an applications point of view without detailed theoretical development. Sufficient theory is included so that the application may be understood.

The application of statistics usually involves reference to tables of statistical functions. In this book emphasis is placed on the solution of statistical problems by means of small, programmable calculators. Programs for the statistical functions and for the statistical analysis of data are given. Although reference to statistical tables is made, programs for the generation of all the tables required are given and once these programs are available on magnetic cards no reference to tables is required.

Programs are presented in detail so they may be run by any engineer familiar with the calculator. All programs and calculator references are given for both the Hewlett-Packard (HP) and the Texas Instruments (TI) calculators. Whenever it is practical, the same symbols and program steps are used in both types of programs so that the translation from one to the other is simplified. Adequate explanations and flow diagrams of the programs are given so that the programs can be used in larger computer systems if the engineer desires to do so.

All examples in the text are given for both types of calculators when there is any difference in data input or program performance. All examples show the actual running time of the calculator in the form (HP/TI), with the time in seconds. These times do not include the time of data input. The calculator times are given not only to show the short time required for most of the calculations, but also to indicate when the calculator time may be longer than usual, e.g., when a number of iterations is required for a solution. The longer times are indicated so that the user does not terminate the program under the impression that it has gone into an endless loop.

All programs terminate with a clearing of all program flags so that repeated calculations can be made without error from residual values. However, if programs are interrupted before termination, the calculator may not be clear for further calculation. And if the programs are run immediately following other calculator operations, there may be interference from prior flag settings. To ensure correct operation, all flags should be cleared either before the first operation of the program or for operation after an interruption. KEY RST in the Texas Instruments calculator or the CLEAR FLAG 0, 1, 2 operation in the Hewlett-Packard calculator will provide for correct operation.

It is suggested that all programs be tested with examples from the text the first time they are run before they are used with experimental data. Such testing will not only ensure that the program has been copied correctly, but will also give the operator an idea of the calculator running time.

References to program steps are given in upper-case type: KEY A, STO B (store in register B), etc.

1.2 STATISTICAL METHODOLOGY

The statistical methods discussed in this book are those in which data are observed and compared with some probability distribution so that statements with mathematically defined confidence may be made about the results. For example, an engineer makes several determinations of a process yield. There is a small amount of variation among the results and it would be useful to state with 95% certainty, or with 99% certainty, where the true yield lies (Sec. 4.3).

There are two types of quantitative data: measurements and counts. The measurements usually form a continuous variable with the continuity limited only by the precision of the measurements. The counted data are in discrete intervals. Different probability functions are used for these different types of data (Secs. 3.2, 3.5, and 3.9).

Two aspects of experimental results are subject to statistical analysis: the mean, some central tendency of the data; and the variation, the difference among results. Methods are provided for testing hypotheses about the mean: Is the mean such and such (Sec. 4.2)? Is there a difference between means from two sources (Sec. 4.2)? Where does the true mean lie (Sec. 4.3)?

Methods are also provided for dealing with questions about the variation of data. Does one set of measurements vary more than another set (Sec. 6.4)? Is the variation more than could be attributable to experimental error (Sec. 6.8)? Is the variation associated with changes in temperature, pressure, flow rate, etc. (Sec. 6.9)?

Another statistical relation discussed is that between two variables. How is one variable related to another (Sec. 7.2)? What is the best curve to draw through some data (Sec. 7.4)?

In all cases hand-calculator programs are presented both to process the data for making the necessary statistical calculations and to provide the probability associated with the calculated results. Programs for all the statistical functions are given in the text so that no reference to statistical tables is necessary.

1.3 CALCULATOR PROGRAMS

The calculator programs are written for both the Hewlett-Packard HP-97 and for the Texas Instruments TI-59, and they are in a form suitable to be copied directly into the calculator or stored on magnetic cards. The actual program listings and line-by-line descriptions are given in the Appendix. A logic diagram for each program is given as it is presented in the text so that the programs may be translated into any other language suitable for use in another type of calculator.

The programs as presented for the HP-97 will run equally well in the Hewlett-Packard models HP-67 and HP-41C. The programs given for the TI-59 will run equally well in the TI-58.

This book is not intended to give instruction in calculator programming, but some suggestions for modification or expansion are included with some of the programs. Some of the programs might be written in a more simple format, but in the presentation of a series of programs to do several similar tasks, it was considered better to keep the format uniform with respect to data entry, program operation, and output of results. Each program is complete and operable as presented, and several of the programs have subroutines for statistical function calculations that can be used either separately or in conjunction with other programs. These subroutines are sufficiently identified so that they can be used as separate programs.

Several relatively short programs may be stored with others on the same magnetic card. If this is done, it may be necessary to relabel the programs and any subroutines present to distinguish clearly between the programs stored on the same magnetic cards.

All the examples in the text are given for both the Hewlett-Packard programs and for the Texas Instruments programs. All of the programs are written for calculators with printing facility. The programs will run equally well in calculators without printing facility, but some modification is required to have the program stop or pause long enough for the results to be written. The changes are pointed out in the Appendix where the programs are discussed in detail.

Some of the calculator programs presented may be similar to those available from other sources. The solutions to statistical problems are presented in many texts and the programming of these solutions is fairly straightforward. However, all of the programs in this book were written specifically for this text.

2 STATISTICAL PARAMETERS

2.1 MEASURES OF CENTRAL TENDENCY

It is the general practice to use one number to describe a process or technique. The octane number is 92; the carbon content is 15 ppm; the gas mileage is 16.2 mi/gal. When the single number is a result of some experimental data it is also important to give some indication of how much variation there is in the data. The American Society for Testing and Materials (ASTM) recommends[1*] presenting as a minimum the mean, the standard deviation, and the number of observations. The mean is a measure of the central tendency. The standard deviation is a measure of the variation.

The single value most often used to report some data or process is the mean, the *arithmetic* mean. This number is intended to show the value around which all of the data cluster. There are other measures of the central tendency of the data and a number of these are listed in Table 2.1.

This book deals principally with the arithmetic mean, referred to in most cases simply as "the mean." It is not always the most suitable measure, however, and examples illustrating where one of the other measures is preferable are given in the sections that follow.

Table 2.1 Measures of Central Tendency

Arithmetic mean
Geometric mean
Harmonic mean
Weighted mean
Median
Mode
Midrange

* All references are listed after the Appendix.

2.1.1 The Arithmetic Mean

The arithmetic mean is the sum of all the data (Σx_i) divided by the number of observations (n):

$$\bar{x} = \frac{\Sigma x_i}{n} \tag{2.1}$$

The arithmetic mean is the most commonly reported measure of central tendency. It is relatively easy to calculate and its calculation includes all of the data. Most programmable calculators have a single key that both sums and counts the data and a second key that carries out the division to give the mean. In fact, the key that sums and counts the data will sum two sets of data entered at the same time so that the arithmetic mean of two sets of data may be obtained by just entering the data with the proper key.

One very important characteristic of the arithmetic mean is that its distribution tends toward a normal distribution. The larger the value of n the closer is the approximation of the distribution of the means toward the normal. The significance of this fact is that means of samples taken from any type of distribution (with minor restraints) will approach the normal distribution, and the statistical functions, which are discussed in the following chapters, can be used with confidence even when the distribution of the population from which the samples are taken is not known.

2.1.2 The Geometric Mean

The geometric mean is the nth root of the product of n observations:

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} \tag{2.2}$$

The logarithm of the geometric mean tends to be normally distributed inasmuch as the log of $\bar{x}_G$ is the arithmetic mean of the logarithms of the variable values:

$$\log(\bar{x}_G) = \Sigma \log(x_i)/n \tag{2.3}$$

and the transformation to log functions is sometimes employed when there is a wide range of the variable values.

With growth data, if the mean calculation is made to find one value that best represents all of the data, the geometric mean rather than the arithmetic mean should be used. This application occurs more frequently with business data than with engineering data. A simple illustration is with interest rates. If interest rates vary at regular intervals, the geometric mean, rather than the arithmetic mean, will give the correct average rate. See Example 2.4.1.

2.1.3 The Harmonic Mean

The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the data:

$$\bar{x}_{\mathrm{H}} = \frac{1}{\Sigma(1/x_i)/n} \tag{2.4}$$

The harmonic mean is used with rate data under certain conditions. If the average gasoline mileage is to be determined and in three test runs of 150 mi each a car gets 15, 17, and 19 mi/gal, what is the average mileage? If the test runs were made with fixed amounts of gasoline—say 10 gal in each run—and the same data were obtained, what would be the average mileage? In the first case the harmonic mean is used and in the second the arithmetic mean.

The proper selection is most easily explained in terms of the dimensions of the measurements. If the data are obtained at constant values of the variable that is the numerator of the rate measurements, the harmonic mean is used. If the data are at constant values of the denominator variable, the arithmetic mean is used.

With a process tested to determine yield in pounds per hour and data taken at regular time intervals, the arithmetic mean gives the average yield. If the data are taken in fixed quantities of product, the harmonic mean gives the correct average yield.

The reciprocals of harmonic means will tend toward normality since they are the arithmetic means of the reciprocals of the data, and the means from any distribution tend to be normally distributed.

2.1.4 Weighted Means

In most statistical analyses each measurement gets equal weight. There are occasions, however, when some data are considered more reliable than others and are given more weight in calculating means and other statistical parameters.

On other occasions, where there are large amounts of data, the measurements are grouped into classes and the statistical parameters are calculated by using the class midpoint and a weight corresponding to the number of measurements in the class. The presentation of large amounts of data is simplified and made easier to comprehend when the data are grouped into classes and shown as a frequency distribution with the frequency plotted as ordinate and the class values as abscissa. An example of a frequency distribution is given in the next chapter.

The ASTM manual on quality control suggests using between 13 and 20 classes. When there are very large amounts of data, as with income tax returns and other government statistics, or even with small-parts manufacture or catalyst testing, the grouping of the results is the only practical way to present the data.

If f_i is the weight or frequency of measurement x_i, the weighted arithmetic mean $\bar{x}_{\mathrm{w}}$, geometric mean $\bar{x}_{\mathrm{G_w}}$, and harmonic mean $\bar{x}_{\mathrm{H_w}}$, corresponding to the three means discussed before, are:

$$\bar{x}_{\mathrm{w}} = \frac{\Sigma f_i x_i}{\Sigma f_i} \tag{2.5}$$

$$\bar{x}_{\mathrm{G_w}} = \sqrt[\Sigma f_i]{x_1^{f_1} \cdot x_2^{f_2} \cdot \cdot \cdot x_n^{f_n}} \tag{2.6}$$

$$\bar{x}_{\mathrm{H_w}} = \frac{1}{\Sigma(1/f_i x_i)/\Sigma f_i} \tag{2.7}$$

2.1.5 Median and Mode

The median and mode are not discussed in this text beyond what is written here. The median and mode cannot be obtained readily with a hand calculator. They are used, however, in some statistical evaluations and it is useful to be familiar with what they represent.

The *median* is the middle value when the data are arranged in numerical sequence. Approximately 50 percent of the observations are greater than the median and 50 percent are less. With an odd number of observations, the median is a unique value. With an even number, the median is the arithmetic mean of the two middle values.

The median has particular value in life testing. Its practical advantage is that it is not necessary to wait for all of the data before obtaining its value. In flex-life testing of electric cable, when half the test pieces have failed the median life is obtained. In radioactive decay, the half-life is the median time of decay of the starting atoms.

The *mode* is the most frequent value. If the statistical parameter that corresponds to most of the population is desired, the mode is needed rather than a mean. In the social sciences it is often more desirable to know the mode—the income of the largest group or the test score achieved by the largest number of contestants—rather than the mean income or the mean score.

If a frequency distribution of data is symmetrical, the arithmetic mean, the median, and the mode will coincide. If the data are not symmetrical, the median will be between the mean and the mode. If the median and mode are less than the mean, the distribution is skewed to the left. If the median and mode are larger than the mean, the distribution is skewed to the right.

2.1.6 Midrange

The midrange is the arithmetic mean of the extreme values. The range is often used in quality control, and the midrange is easily calculated. It is seldom used with engineering data and is merely mentioned to complete the list. The range and the midrange are calculated from only two values in a sample and hence are less representative of the whole lot than the other parameters.

2.1.7 Log Mean

The log mean is not a statistical parameter, but is a mean used by engineers for calculating the mean driving force for a transfer problem where the rate of transfer is a function of the driving force and the driving force varies over the length of the system. In heat transfer, if the rate of transfer is proportional to a temperature difference, and the temperature difference varies over the length of the reactor in proportion to the heat transferred, the log mean temperature difference gives the mean driving force. The log mean is the difference between the differences at the ends of the system divided by the natural log of their ratio:

$$\text{log mean } \Delta T = \frac{\Delta T_2 - \Delta T_1}{\ln (\Delta T_2/\Delta T_1)} \tag{2.8}$$

In the statistical parameter program discussed later in this chapter the log mean is calculated from the maximum and minimum values of the data input. It is simply available in the program if the log mean is desired.

2.2 MEASURES OF DISPERSION

In addition to a measurement of the central tendency of the data, it is important to have a measure of the dispersion or variation. It is obvious that two sets of data could have the same mean but one set could have much more scatter, indicating less process control. Or the scatter might be caused by variation in operating factors.

As with means, there are several measures of dispersion. The most common are listed in Table 2.2. The one that is most useful statistically is the standard deviation.

The standard deviation is the square root of the variance, and the variance will be discussed first. But a few words need to be said about the sample and the population.

A *sample* is a portion of information taken from a much larger population. The population has certain defining parameters that are usually not known: the mean, the range, the variance, etc. Measurements are usually made on the sample with the aim of estimating the population parameters. The measurements on the sample are called *statistics.* Sometimes the population is a theoretical one and the parameters are known by definition: a population of random numbers, for example; or the distribution of winning numbers in a game of chance. In situations of this latter type, the samples are tested to determine if they conform with the defined population. Most of this text deals with problems of the first kind.

Probability theory permits the experimenter to make statements about the population parameters from the sample statistics with calculable chance of error. This book does not discuss probability theory, but the risks assumed in evaluating sample measurements are discussed when the statistical tests are applied.

Table 2.2 Measures of Dispersion

Variance
Standard deviation
Coefficient of variation
Range
Mean deviation
Weighted measures
Percentiles

2.2.1 Variance, Population

The variance is defined as the arithmetic mean of the squares of the deviations of all of the values from the arithmetic mean of the population. The usual symbol for the variance is sigma square, σ^2:

$$\sigma^2(x) = \frac{\Sigma(x_i - m)^2}{n} \tag{2.9}$$

The population mean is designated m to differentiate it from a sample mean which is designated $\bar{x}$. The value of m is usually not known and is only estimated by the sample value $\bar{x}$.

The variance is always positive, and, as is shown in Chap. 6, subject to rigorous mathematical operation.

2.2.2 Standard Deviation, Population

The standard deviation is the positive square root of the variance. The mean and the standard deviation are the only parameters required to define the distribution of a normal population. The normal distribution is discussed in the next chapter.

2.2.3 Variance, Sample

The sample variance is calculated usually as an estimate of the population variance from which the sample was taken. The population variance, as defined in Eq. (2.9), is calculated from the sum of squares of deviation from the population mean. The sum of squares of deviation of a set of numbers from their mean is less than the sum of squares of deviation from any other value. Therefore the sum of squares of deviation of sample values from the sample mean will be less than (or at most equal to) the sum of squares of deviation from the population mean. If Eq. (2.9) is used to calculate an estimate of the population variance from the sample values, the sum of squares of deviation would be calculated from the sample mean and the resulting answer would be smaller than if the sum of squares were calculated from the population mean. To correct for this bias in the calculation from the sample data, the sum of squares is divided by $n - 1$ instead of by n. The variance estimate, calculated from the sample, designated s^2 to differentiate it from the true variance σ^2, is calculated as follows:

$$s^2(x) = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} \tag{2.10}$$

The calculation of the sum of squares of deviation from the sample mean may be performed by calculating the mean, then squaring the difference between the mean and each value, and obtaining the sum. There is an easier method, and it is the one used in most calculator programs. The following equality is the basis of the calculation:

$$\begin{aligned} \Sigma(x_i - \bar{x})^2 &= \Sigma x^2 - (\Sigma x)^2/n \\ &= \Sigma x^2 - \bar{x} \cdot \Sigma x \end{aligned} \tag{2.11}$$

Most programmable calculators have a single key, KEY $\Sigma+$, which calculates Σx^2, Σx, and n with a single entry of the data.

2.2.4 Standard Deviation, Sample

The standard deviation estimated from the sample is the positive square root of the estimated variance from Eq. (2.10). If the data are entered into a programmable calculator with KEY $\Sigma+$, the calculator usually has keys labeled $\bar{x}$ and s. These keys will give the arithmetic mean and the estimated standard deviation. If the data entered are not a sample but are the total population, the value given by KEY s in the calculator will have been calculated by Eq. (2.10) and must be corrected by the factor $\sqrt{(n-1)/n}$ to get the correct σ value.

The standard deviation estimate calculated from sample data is a very important statistic. It, together with the mean and the sample size, are specified by the ASTM as the minimum information that should be presented with all data. The standard deviation has the same units as the sample measurements. It is used to set confidence limits to the mean and to compare means from two sets of data, it is used to set quality control limits in production processes, and it has other uses that will be mentioned further on in the text.

2.2.5 Coefficient of Variation

The coefficient of variation is the ratio of the standard deviation to the mean, usually expressed as a percentage:

$$\mathrm{CV} = \frac{100\, s(x)}{\bar{x}} \tag{2.12}$$

The coefficient of variation provides a measure for comparing the dispersion of measurements having different means and different units. The coefficient of variation is dimensionless. It is not used in any of the material in this book.

2.2.6 The Range

The range is the difference between the smallest and the largest values in a set of data. It involves only two observations and is simple to determine. It is used for comparing large numbers of small samples. It is used to a large extent in quality control, and constants are available for estimating population standard deviation and variance from the mean range of a number of samples.

2.2.7 Mean Deviation

The mean deviation is the mean absolute deviation of the sample measurements from the sample mean:

$$\mathrm{MD} = \frac{\Sigma|x_i - \bar{x}|}{n} \tag{2.13}$$

The mean deviation is less sensitive to extreme values than is the standard deviation inasmuch as the deviations are not squared. It is sometimes used in reporting data because its value is always smaller than that of the standard deviation. It has limited use in statistical calculations and is not used in any of the calculations in this book.

2.2.8 Weighted Measures

As discussed in Sec. 2.1.4, with large quantities of data the observations are often grouped into classes and the calculations are made by using the class means and the class frequencies rather than all of the individual values. Equations for calculating the arithmetic mean and the variance from grouped data are given below, where f_i is the frequency or weight of the x_i measurement:

$$\bar{x}_w = \frac{\Sigma f_i x_i}{\Sigma f_i} \tag{2.5}$$

$$\sigma_w{}^2(x) = \frac{\Sigma f_i(x_i - \bar{x}_w)^2}{\Sigma f_i} = \frac{\Sigma f_i x_i{}^2 - \bar{x}_w \cdot \Sigma f_i x_i}{\Sigma f_i} \tag{2.14}$$

Inasmuch as data are seldom grouped into classes for calculating purposes unless there are at least 100 or more observations, the $n - 1$ correction for the estimated variance calculation usually does not make a significant difference in the result.

The calculator KEY $\Sigma+$ cannot be used to calculate the means and standard deviations of weighted or grouped data.

2.2.9 Percentiles

Percentiles are used principally with social science data. They define where the specified percentages of the data fall: 10 percent of the observations fall below the 10th percentile; 90 percent of the data will lie between the 5th and the 95th percentiles. The 25th percentile is sometimes called the first quartile, the 50th percentile the second quartile, and so forth. This nomenclature is seldom used with engineering data.

2.3 STATISTICAL PARAMETER PROGRAM

The statistical parameter program calculates all of the values in Table 2.3 for either individual data or for weighted data.

When all the data have been entered and the calculator program has been run, all the results in Table 2.3 are available in the machine. The program as presented will print all the values in the order listed. However, for practical usage, only the values desired need be printed. The modification of the program to get specific values is explained in the next section.

The harmonic mean, using $1/x$, and the geometric mean, using $\log(x)$, cannot be obtained for zero values. The geometric mean cannot be obtained if one of

Table 2.3 Values Obtainable with Statistical Parameter Program

Arithmetic mean
Midrange
Geometric mean
Harmonic mean
Log mean
Estimated standard deviation
Data standard deviation
Estimated variance
Data variance
Range
Minimum value
Maximum value
Σx
Σx^2
(Number of groups for weighted data)
Total number of data points

the values is negative. The program provides for this eventuality. If either zero or negative values are entered, the program makes all the other calculations and gives zero values for the means that cannot be calculated.

2.3.1 Discussion of Statistical Parameter Program

Figure 2.1 is a flow diagram for the statistical parameter program. A listing and detailed description are given in the Appendix. Appendix A has the Hewlett-Packard listing and description, and Appendix B has the Texas Instruments listing and description. In the discussions that follow, any differences between the programs with respect to data input or other operation are pointed out in the text. When no difference is mentioned, there is no difference.

For equally weighted data each value is entered with KEY A: x_1, KEY A, x_2, KEY A, . . . , x_n, KEY A. When all the data are entered, the program is run with KEY C.

For weighted data, the weighting factor and the data value are entered with KEY B:

With the Hewlett-Packard calculator:

f_i, ENTER ↑, x_i, KEY B, f_2, ENTER ↑, x_2, KEY B, . . . , f_n, ENTER ↑, x_n, KEY B.

When all the data are entered, KEY C carries out the calculation.

With the Texas Instruments calculator:

f_1, KEY $x \rightleftharpoons t$, x_1, KEY B, f_2, KEY $x \rightleftharpoons t$, x_2, KEY B, . . . , f_n, KEY $x \rightleftharpoons t$, x_n, KEY B.

KEY C starts the calculation when all the data are in.

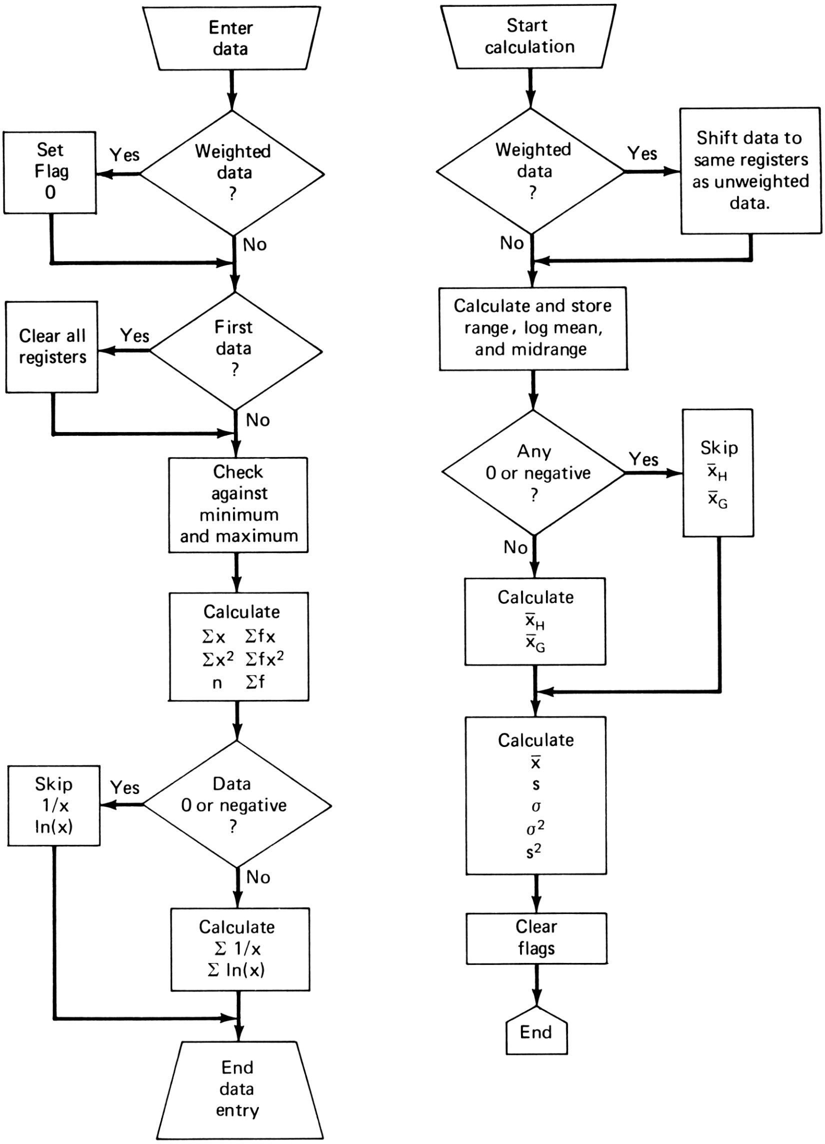


Figure 2.1 Flow diagram of statistical parameter program.

When weighted data are used, each x value must be preceded by a weighting factor even if it is unity. If no weighting value is entered, the program will use the last value that was in the calculator with the Hewlett-Packard program, or the last value in the t register with the Texas Instruments program.

At the end of the calculation, the final values are in the storage registers shown in Table 2.4. The programs presented call for printing all of these values in the

Table 2.4 Location of Results after Running Statistical Parameter Program

Register HP	Register TI	Result
0	00	$\bar{x}$, Arithmetic mean
1	01	Midrange
2	02	$\bar{x}_G$, Geometric mean
3	03	$\bar{x}_H$, Harmonic mean
4	04	Log mean of two extreme values
5	05	s, Estimated standard deviation
6	06	σ, Data standard deviation
7	07	s^2, Estimated variance
8	08	σ^2, Data variance
9	09	Range
A	10	Minimum value
B	11	Maximum value
C	12	Σx (or Σfx)
D	13	Σx^2 (or Σfx^2)
E	14	Number of groups for weighted data
I	15	Number of data values, n (or Σf)

order listed. If only specific results are wanted, the programs can be readily modified. In the Hewlett-Packard program the PREG (Print Register) instruction at line 155 can be replaced with statements calling the particular result desired and printing only that. In the Texas Instruments program, the print routine starts at line 306. Instructions calling particular results and printing these can be placed at this location in the program. See Example 2.4.1.

If zero values are entered, the harmonic mean, the geometric mean, and the log mean are not calculated, and the values in Registers 2, 3, and 4 will be zero. If negative values are entered, the geometric mean and the log mean are not calculated, and the values in Registers 2 and 4 will be zero. If the data are not weighted, the value in Register E (HP) or Register 14 (TI) will be zero.

With the first data entry the program clears all the registers so that any previous calculations will not interfere with the results. The first entry is stored in two registers, one for the minimum input and one for the maximum. Each subsequent data input is compared with the values in these registers; if it is larger than the maximum, in Register B, or 11; or if it is smaller than the previous minimum, in Register A, or 10, a substitution is made. The minimum and maximum values are used to calculate the log mean and the range.

If unweighted data are entered, KEY $\Sigma+$ is used to accumulate Σx and Σx^2; $\Sigma 1/x$ and Σ log (x) are accumulated separately. If there are negative or zero values, the last two calculations are bypassed.

With weighted data, the summations are made separately. The use of KEY B for weighted data sets Flag 0 in the calculator to direct the program to the proper summation routines.

When all the data are entered, with unweighted data the built-in functions of KEY $\bar{x}$ and KEY s are used to obtain the mean and the estimated standard deviation. The data variance is obtained directly in the Texas Instruments program with a built-in function, and from the estimated standard deviation value in the Hewlett-Packard program.

The geometric mean and harmonic mean are calculated from Eqs. (2.3) and (2.4), taking the antilog of the result of Eq. (2.3) for the geometric mean. With weighted data, Eqs. (2.5), (2.6), (2.7), and (2.14) are used.

The program gives more information than is usually desired, but the total calculator time is less than 30 seconds (s), so there is not much to be gained by having a smaller program.

2.4 EXAMPLES OF THE STATISTICAL PARAMETER PROGRAM

2.4.1 If a money market certificate pays 7.193 percent interest the first 6 months (mo), 11.237 percent the second and 10.734 percent the third, what is the average interest over the 18 mo? For an answer the geometric mean is required, and this value is stored in Register 2. Since none of the other values is of interest, the program will be modified to give only the geometric mean:

HP	TI
Delete line 155 Replace with: RCL 2 (Recall Register 2) PRTX (Print x)	Starting at line 306 replace the program lines with the following: 306 RCL (Recall) 307 02 (Register 02) 308 PRT (Print) 309 GTO (Go to) 310 – (Label –)

With these changes in the program, the problem is solved as follows:

1.0793, KEY A, 1.11237, KEY A, 1.10734, KEY A, KEY C. Answer: 1.09706. (Time:6/10)

The average interest rate over the three periods is 9.706 percent. The arithmetic mean of the three rates is 9.721, which is an incorrect answer to the problem.

2.4.2 In the body of the text the question was raised about the average gasoline mileage for a car that was driven 150 mi in each of three runs and obtained 15, 17, and 19 mi/gal in the three runs. The average mileage will be the harmonic mean. Change the program to recall storage register 3 and proceed as follows:

15, KEY A, 17, KEY A, 19, KEY A, KEY C. Answer: 16.84 mi/gal (Time:5/10)

The arithmetic mean of the three rates is obviously 17, which is too high an answer. The geometric mean and the harmonic mean are always less than the arithmetic mean.

Table 2.5 Weights of Coating of 100 Sheets of Galvanized Iron, ounces per square foot

1.467	1.603	1.577	1.563	1.437	1.337	1.543
1.623	1.603	1.577	1.393	1.350	1.637	1.473
1.520	1.383	1.323	1.647	1.530	1.753	1.603
1.767	1.730	1.620	1.620	1.383	1.567	1.570
1.550	1.700	1.473	1.530	1.457	1.633	1.467
1.533	1.600	1.420	1.470	1.443	1.373	1.490
1.377	1.603	1.450	1.337	1.473	1.617	1.763
1.373	1.477	1.337	1.580	1.433	1.563	1.457
1.637	1.513	1.440	1.493	1.637	1.550	1.477
1.460	1.533	1.557	1.563	1.500	1.573	1.503
1.627	1.593	1.480	1.543	1.607	1.660	1.577
1.537	1.503	1.477	1.567	1.423	1.750	1.537
1.533	1.600	1.550	1.670	1.573	1.550	1.323
1.483	1.497	1.420	1.647	1.647	1.600	1.717
1.513	1.690					

Table 2.6 Grouping Data from Table 2.5

Cell range	Cell midpoint	Number in cell
1.3000–1.3399	1.3200	5
1.3400–1.3799	1.3600	4
1.3800–1.4199	1.4000	3
1.4200–1.4599	1.4400	10
1.4600–1.4999	1.4800	15
1.5000–1.5399	1.5200	13
1.5400–1.5799	1.5600	18
1.5800–1.6199	1.6000	11
1.6200–1.6599	1.6400	11
1.6600–1.6999	1.6800	3
1.7000–1.7399	1.7200	3
1.7400–1.7799	1.7600	4

2.4.3 Table 2.5 shows some data on laboratory measurements of galvanized iron coating weights.[1] The tabulated data are difficult to evaluate and they will be grouped into classes for easier presentation. Table 2.6 shows the same data grouped into 12 classes. The smallest value in Table 2.5 is 1.323 and the largest is 1.767, giving a range of 0.444. Each of the 12 classes of the grouped data has a width of 0.04 giving a total range of 0.480. The first class starts at 1.300, just below the smallest value, and the last class ends at 1.780, above the largest value.

The data from Table 2.6 are run with the statistics parameter program as follows:

5, ENTER ↑, 1.32, KEY B, 4, ENTER ↑, 1.36, KEY B, . . . , 4, ENTER ↑, 1.76, KEY B

for the Hewlett-Packard program, or

5, KEY $x \rightleftharpoons t$, 1.32, KEY B, 4, KEY $x \rightleftharpoons t$, 1.36, KEY B, . . . , 4, KEY $x \rightleftharpoons t$, 1.76, KEY B

for the Texas Instruments program.

With all the data entered, KEY C carries out the calculation. The output is shown in Table 2.7. Just for comparison, the data of Table 2.5 were run without grouping, using KEY A. These results are also given in Table 2.7.

The calculator time for both calculations was 22 to 32 s, but of course the time to enter the data for the individual value calculation was greater than for the grouped data. There is very little difference between the results for the two different types of calculation, and the grouped data are easier to comprehend visually. The statistical parameter program provides for either calculation. The engineer can make the choice.

Table 2.7 Comparison of Individual and Grouped Data Calculations with Statistics Parameter Program for Data in Tables 2.5 and 2.6

	Individual	Grouped
Arithmetic mean	1.5351	1.5356
Midrange	1.5450	1.5400
Geometric mean	1.5316	1.5320
Harmonic mean	1.5281	1.5284
Log mean	1.5343	1.5295
Estimated standard deviation	0.1038	0.1051
Data standard deviation	0.1033	0.1046
Estimated variance	0.0108	0.0111
Data variance	0.0107	0.0109
Range	0.4440	0.4400
Minimum value	1.3230	1.3200
Maximum value	1.7670	1.7600
Σx (Σfx)	153.508	153.5600
Σx^2 (Σfx^2)	236.7134	236.9008
Number of cells	0	12
n (Σf)	100	100

3 PROBABILITY DISTRIBUTIONS

3.1 FREQUENCY DISTRIBUTIONS

A popular way of presenting data for visual appreciation is by means of a frequency diagram. In a diagram of this nature the value of the observations is scaled along the abscissa and the number of observations is shown as the ordinate. The measurement of variable shown as the abscissa is often indicated in fixed intervals, and all the observations falling within an interval are grouped together. The number of observations for each interval or class is indicated by a line or rectangle with length proportional to the number. An example of a frequency diagram for some income tax return data is shown in Fig. 3.1. The ordinate is the number of returns, in units of 10,000, and the abscissa is the amount of interest claimed as an itemized deduction for a particular income class.

If the frequency scale is changed from the actual number of observations to the relative frequency—dividing each value by the total number of observations—

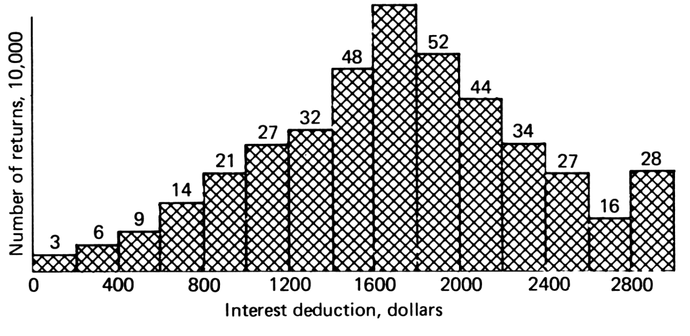


Figure 3.1 Interest deductions on income tax returns.

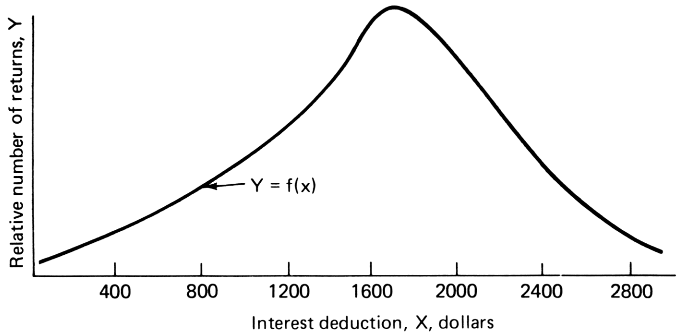


Figure 3.2 Probability distribution for income tax data.

the shape is not changed, but the diagram would then be called a probability distribution diagram. The ordinate scale would be the probability of an observation occurring in the class indicated by the abscissa.

If it were possible to write a mathematical function that would give the relative frequency of returns in terms of the amount of deductions as in Eq. (3.1),

$$\text{Relative frequency of returns} = f\,(\text{Amount of interest claimed}) \qquad (3.1)$$

this would be a probability distribution function, more commonly called a probability distribution. Such a function would be represented by a continuous curve that smoothed out the rectangular frequency distribution figure.

Figure 3.2 shows the probability distribution function for the income tax data of Fig. 3.1.

The fraction of the area under the curve up to any value of the abscissa represents the probability of a return having up to that amount of interest claimed. The fraction of the area beyond that point represents the probability of having that much or more interest claimed. The fraction of the area under the total curve is 1.0, equal to the probability of the data falling somewhere in the range depicted. Any continuous positive function enclosing a finite area may be regarded as a probability function if the fraction of the area under the curve of the function up to some value of the abscissa is taken as the probability of that value or less.

In general, a sample or samples are observed and their values are compared with some mathematical probability distribution. If they fall in an area of low probability, then questions are raised as to whether they came from the assumed distribution. If the distribution of Fig. 3.2 is taken to represent the tax returns for the \$20,000 to \$25,000 income group and a return from someone in that group has an interest deduction of \$400 or less, which lies in an area covering about 2 percent of the total area, the question can be raised as to whether this return is from the general population. Of course the Internal Revenue Service would be more interested in questioning returns with low probability from the other end of the curve.

3.2 NORMAL DISTRIBUTION

The normal distribution is the most frequently used probability function for the statistical analysis of data. It represents the mathematical function obtained when a large number of independent variables are contributing to the variation of a single result. The normal function is defined by the equation:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma} \tag{3.2}$$

which produces the familiar bell-shaped curve shown in Fig. 3.3.

In Eq. (3.2) m is the mean (sometimes called the average) and is equal to the sum of all the values divided by the number of values:

$$m = \frac{\Sigma x}{n} \tag{3.3}$$

where n is the number of observations.

Again, σ is the standard deviation, the square root of the mean of the squares of deviation of the individual values from the mean:

$$\sigma = \sqrt{\frac{\Sigma(x-m)^2}{n}} \tag{3.4}$$

If a random variable x is normally distributed with mean m and a standard deviation σ, the probability of observing a value equal to or greater than X_2 is the integral of Eq. (3.2) from X_2 to ∞.

$$\Pr(x \geqq X_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{X_2}^{\infty} e^{-(x-m)^2/2\sigma}\, dx \tag{3.5}$$

This probability is represented by the single cross-hatched area in Fig. 3.3.

The probability of a value equal to or less than X_1 is equal to the integral from $-\infty$ to X_1, represented by the double-crossed-hatched area of Fig. 3.3.

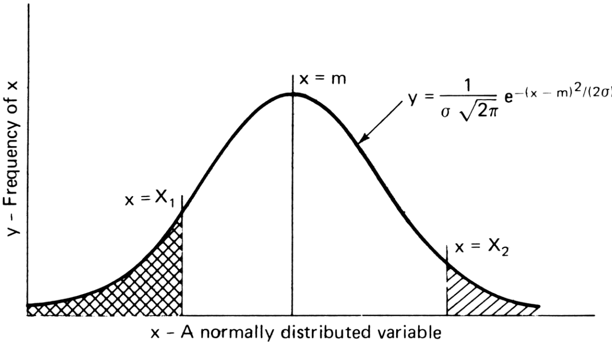


Figure 3.3 Normal distribution.

The probability of a value between X_1 and X_2 is the integral from X_1 to X_2 of Eq. (3.2):

$$\Pr(X_1 \leqq x \leqq X_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{X_1}^{X_2} e^{-(x-m)^2/2\sigma}\, dx \tag{3.6}$$

This probability is represented by the clear area under the curve of Fig. 3.3.

The probability values associated with the normal distribution are usually obtained from tables of the function. One set of values can be used for all values of m and σ if the variable x is transformed by the relation:

$$z = \frac{x - m}{\sigma} \tag{3.7}$$

If x has a normal distribution with mean m and a standard deviation σ, z will have a normal distribution with mean zero and a standard deviation equal to 1.0. Thus any normally distributed variable may be transformed into a normal distribution with mean zero and standard deviation 1.0 by subtracting the actual mean and dividing by the actual standard deviation. The transformed variable is called the standardized normal distribution; z is the standard deviate, and it is the distribution of z that is given in normal distribution tables.

In the example cited just above, the area beyond X_2 in Fig. 3.3 would correspond to the area beyond $(X_2 - m)/\sigma$ in the standardized form, and the probability would be found in a normal table at a value of z equal to $(X_2 - m)/\sigma$.

If the value below which 95 percent of a normally distributed result were expected to occur was wanted, the z value corresponding to 0.05 (that is, 5 percent of the normal area lies beyond this point) would be located in the normal table. The corresponding x value from Eq. (3.7) would then be calculated:

$$x_{0.95} = m + \sigma \cdot z_{0.05} \tag{3.8}$$

If 5 percent of the area lies above $z_{0.05}$, then 95 percent lies below. The normal distribution is symmetrical about the mean m, and only positive values are given in tables. The value below which 5 percent of the area lies is the negative value of the tabulated $z_{0.05}$. If it was desired to find the value of the normally distributed variable below which 5 percent of the results would be found, the same value of $z_{0.05}$ would be obtained from the normal table and the variable value would be $m - \sigma \cdot z_{0.05}$.

Ninety percent of the observations would be in the range of $m \pm \sigma \cdot z_{0.05}$. The generally accepted nomenclature is that for a normally distributed variable, $1 - \alpha$ fraction of the observations is in the range $m \pm (z_{\alpha/2})\sigma$.

For example, if the income tax data from Fig. 3.2 were normal with a mean of \$1762 and a standard deviation of \$633, and one desired to know the probability of a deduction of \$3200 or more, the z value would be calculated from Eq. (3.7) and the probability associated with that z determined from a table of the normal function:

$$z = (3200 - 1762)/633 = 2.27 \tag{3.9}$$

The probability value is 0.01.

The calculator program for the normal distribution solves problems of this type. The program is entered with the mean and the standard deviation and some value of the variable for which the probability is desired. The program will give either the probability of a value equal to or greater than the input value, or the probability of a value equal or less, depending on the key used. The program will also give the probability between two values if two values are entered.

The program can also be used in the inverse way. If along with the mean and the standard deviation, a probability is entered, the program will calculate the maximum or minimum value corresponding to that probability. With the income tax data, the program would be entered with 1762 (the mean), 633 (the standard deviation), and some probability, say 0.05. The calculator will find either the value which was the lower boundary of the top 5 percent, or the upper boundary of the lowest 5 percent, depending on the key used.

3.3 NORMAL PROBABILITY PROGRAM

A flow diagram for the normal program is given in Fig. 3.4, and a listing and detailed description are given in the Appendix. Table 3.1 shows the operation of the program.

The mean is entered, the standard deviation is entered, and then either one or two values of the variable, or a probability value (decimal fraction). With the Hewlett-Packard program the data are entered with the ENTER ↑ key, and then, depending on the problem to be solved, one of the program keys is activated. With the Texas Instruments program all the data are entered with the particular program key for the problem. The calculation starts with the last data entry. KEY A gives the probability between two variable values. KEY C gives the probability of an equal or greater value. KEY E gives the probability of an equal or lesser value.

If a probability value was entered, KEY c (in the Hewlett-Packard, or C' in the Texas Instruments) gives the variable value which is equaled or exceeded with the input probability. KEY e (KEY E') gives the variable value that is the upper boundary of the lower portion of the distribution that occurs with the input probability.

When the program has been run, the input data (in the order of entry) are in Registers A, B, C, and D in the Hewlett-Packard calculator and in Registers 1, 2, 3, and 4 in the Texas Instruments calculator for verification.

For the calculation of the probability between two values of a variable, the order of input of the two variable values is not important. The program gives the absolute value of the probability although it calculates the difference between the probability of the second value subtracted from the probability of the first.

The formulas used for calculating the normal probability function values are taken from the *Handbook of Mathematical Functions.*[2] The results from the program are good to four significant figures, which is the accuracy in most of the published normal tables. Greater accuracy can be obtained with the same program if some

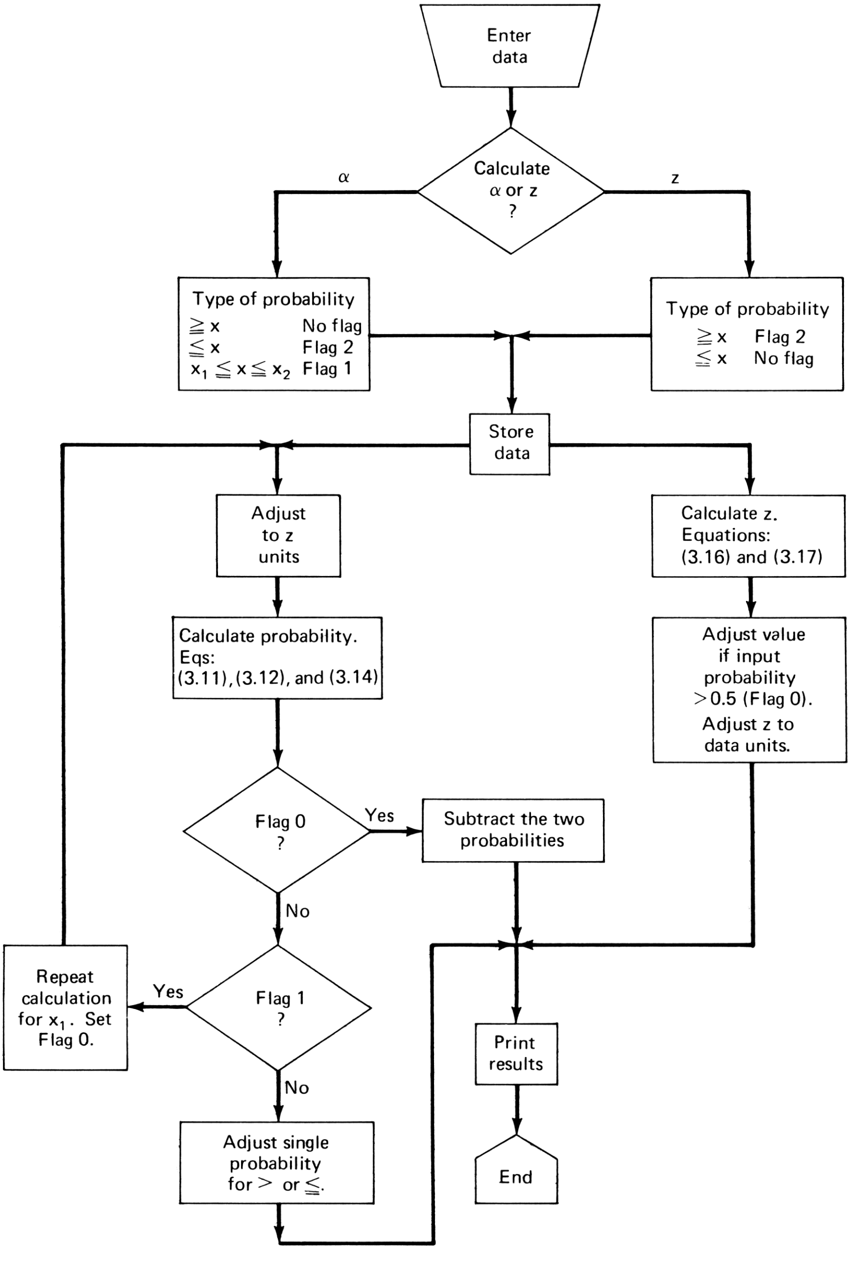


Figure 3.4 Flow diagram of normal probability program.

Table 3.1 Normal Probability Program Operation

		HP	TI
For: $\Pr(x_1 \leqq x \leqq x_2)$	Mean	ENTER ↑	KEY A
	Standard deviation	ENTER ↑	KEY A
	x_1	ENTER ↑	KEY A
	x_2	KEY A	KEY A
For: $\Pr(\geqq x)$	Mean	ENTER ↑	KEY C
	Standard deviation	ENTER ↑	KEY C
	x	KEY C	KEY C
For: $\Pr(\leqq x)$	Mean	ENTER ↑	KEY E
	Standard deviation	ENTER ↑	KEY E
	x	KEY E	KEY E
For: x that $P = \Pr(\geqq x)$	Mean	ENTER ↑	KEY C'
	Standard deviation	ENTER ↑	KEY C'
	P	KEY c	KEY C'
For: x that $P = \Pr(\leqq x)$	Mean	ENTER ↑	KEY E'
	Standard deviation	ENTER ↑	KEY E'
	P	Key e	KEY E'

of the constants are changed. The necessary changes are mentioned in the next section. Calculator time for obtaining probability values is 5 to 15 s.

With an input value of mean = 0 and standard deviation = 1, the program will give the values tabulated for the standardized normal function. The program may therefore be used to interpolate for values that are not given in the commonly available normal tables.

3.3.1 Normal Program Discussion

The normal probability program is actually two separate programs (except for a common "print" statement) in both the Texas Instruments and the Hewlett-Packard programs, and a common routine to store the input data in the Hewlett-Packard program. The first part of the program calculates the normal probability with input of the mean, standard deviation, and variable value. The second part calculates the variable value with input of mean, standard deviation, and probability. Either part could be used as a complete program and attached to some other program that made a calculation and produced the necessary input values.

The first part, which calculates the normal probability, uses the following equations from the cited reference:

$$\Pr(\geqq z) = W(z)(a_1 u + a_2 u^2 + a_3 u^3) \tag{3.10}$$

$$\mathrm{W}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{3.11}$$

$$u = 1/(1 + a_0 \cdot z) \tag{3.12}$$

Table 3.2 Constants for Normal Probability Calculation

Constants	Program values	Reference values
a_0	0.3327	0.33267
a_1	0.4362	0.4361836
a_2	−0.1202	−0.1201676
a_3	0.9373	0.9372980

$$z = \frac{|x - m|}{\sigma} \quad \text{the absolute value of the standardized input variable} \tag{3.13}$$

where a_0 to a_3 are numerical constants

The values of the numerical constants are given in the cited reference to seven significant figures. Only four significant figures are used in the normal program. If greater accuracy is desired, the additional significant figures could be added to the program. The two sets of values are given in Table 3.2.

Equation (3.10) is Eq. (26.2.16) of the reference. For easier programming, the equation form is modified to the following:

$$\Pr(\geqq z) = \mathrm{W}(z)(u(a_1 + u(a_2 + \mathrm{u} \cdot \mathrm{a}_3))) \tag{3.14}$$

The probability $\leqq x$ is equal to one minus the probability $\geqq x$. The program uses Eq. (3.14) for both types of probability calculation and simply subtracts the result from one if the $\leqq$ probability is called for.

The second part of the program, to calculate a variable value for a given probability, uses Eq. (26.2.23) from the reference. This equation is:

$$z_P = w - \frac{c_0 + c_1 \cdot w + c_2 \cdot w^2}{1 + c_3 \cdot w + c_4 \cdot w^2 + c_5 \cdot w^3} \tag{3.15}$$

$$w = \ln \sqrt{\frac{1}{P^2}} \tag{3.16}$$

The values of the c constants are:

c_0	2.515517
c_1	0.802853
c_2	0.010328
c_3	1.432788
c_4	0.189269
c_5	0.001308

P is the input probability, as a decimal fraction, and is the area under the standardized normal curve beyond the value z_P.

Table 3.3 Alternative Solution to the Normal Program

	Pr $\leqq x$	Pr $\geqq x$
Pr $\leqq$ 0.5	$m - z\sigma$	$m + z\sigma$
Pr $\geqq$ 0.5	$m + z\sigma$	$m - z\sigma$

For calculation purposes Eq. (3.15) is rearranged in the program to the following:

$$z_P = w - \frac{c_0 + w(c_1 + c_2 \cdot w)}{1 + w(c_3 + w(c_4 + c_5 \cdot w))} \tag{3.17}$$

After z_P, the standardized value, has been calculated, it is transformed into the dimension of the input variable by the reverse of Eq. (3.13):

$$x_P = m + z_p \cdot \sigma \tag{3.18}$$

The equations above give the correct answer if the input probability is less than 0.5 and is the probability for values $\geqq x$. If the input probability is greater than 0.5, or the probability value sought is for a probability $\leqq x$, a correction is required. Table 3.3 shows the possibilities and the change in the solution.

The program keeps track of the input value of P and the probability sought and makes the proper correction to Eq. (3.18) to obtain the correct result.

3.4 EXAMPLES OF THE NORMAL PROGRAM

3.4.1 A normal population has a mean of 395.1 and a standard deviation of 9.98. What fraction of the population is equal to or greater than 412.5?

With the program in the calculator:

	HP	TI	
395.1	ENTER ↑	KEY C	
9.98	ENTER ↑	KEY C	
412.5	KEY C	KEY C	Output: 0.0406 (Time:5/7)

3.4.2 With the income tax data, what value is exceeded by only 25 percent of the returns:

	HP	TI		
1762	ENTER ↑	KEY C'	The mean	
633	ENTER ↑	KEY C'	Standard deviation	
0.25	KEY c	KEY C'	*P* value	Output: 2188 (Time:9/13)

3.4.3 If a variable is normally distributed about a mean of 0.007 with a standard deviation of 0.85, what is the probability of a value equal to or less than -0.03?

	HP	TI			
0.007	ENTER ↑	KEY E			
0.85	ENTER ↑	KEY E			
−0.03	KEY E	KEY E	Output:	0.4826	(Time:6/8)

3.4.4 With the same variable as in Example 3.4.3, below what value does 5 percent of the data lie?

	HP	TI			
0.007	ENTER ↑	KEY E'			
0.85	ENTER ↑	KEY E'			
0.05	KEY e	KEY E'	Output:	−1.39	(Time:7/9)

3.4.5 For the standardized normal function, what fraction of the area under the normal curve lies below 1.0?

	HP	TI				
0.0	ENTER ↑	KEY E	The mean			
1.0	ENTER ↑	KEY E	Standard deviation			
1.0	KEY E	KEY E	Variable	Output:	0.8414	(Time:6/9)

What fraction lies above 2.0?

	HP	TI			
0.0	ENTER ↑	KEY C			
1.0	ENTER ↑	KEY C			
2.0	KEY C	KEY C	Output:	0.0228	(Time:6/8)

3.4.6 A normal population has a mean of 0.745 and a standard deviation of 0.008. What fraction of the population lies between 0.735 and 0.760?

	HP	TI			
0.745	ENTER ↑	KEY A			
0.008	ENTER ↑	KEY A			
0.735	ENTER ↑	KEY A			
0.760	KEY A	KEY A	Output:	0.8640	(Time:12/16)

Note: If the larger value of the variable was entered before the smaller value, the answer is the same.

3.5 BINOMIAL PROBABILITY DISTRIBUTION

Quantitative data are of two kinds: measured values and counted values. The normal distribution of the previous sections applies to measured data. The normal distribution is a continuous function and is used where the interval between measurements is limited only by the accuracy of the measurements.

Counted data, on the other hand, have only discrete values. With the continuous normal probability, the probability equal to or greater than some value (or the probability between two values) is equal to the area under the normal curve corresponding to the problem. The probability for any specific value is infinitesimal. It is represented by the area under one point on the curve. With a discrete probability there is a probability associated with each discrete number. The probability of some value or a greater one is the sum of the probabilities for each possible discrete value equal to or greater than the one in question. When the binomial probability distribution is shown as a series of rectangles or horizontal lines, as illustrated later in Fig. 3.5, the probability associated with a particular event is the fraction of the area under the curve from half the distance to the previous value to half the distance to the next value. The probability of three events is the fraction of the area under the curve from 2.5 to 3.5. The probability of three or more events is the fraction of the area from 2.5 to the end of the curve. The probability of three or fewer events, is the fraction of the area from the start to 3.5.

Two probabilities for discrete events are discussed in this book: the binomial and the Poisson. The binomial probability is that which is obtained with a fixed sample size when the probability of an event is constant for each item in the sample. The Poisson probability applies when the expected number of events is fixed but the sample size is undefined. For example, the expected number of accidents over a long weekend is 547, what is the probability of 600 or more? The binomial probability is discussed first.

If a coin is tossed, the probability of "heads" is 0.5. The distribution of the number of heads when 10 coins are tossed a number of times will be a binomial distribution. If two dice are tossed, the probability of their exposed faces adding to seven is 1/6. The distribution of the number of "sevens" with 10 tosses of a pair of dice will be a binomial distribution. The probability is constant for each event, and the sample size is fixed. The distributions for these two examples are shown in Fig. 3.5, and the probabilities are given in Table 3.4.

The probabilities for the binomial distribution are calculated by the formula of Eq. (3.19):

$$\Pr(r) = \frac{n!}{r!(n-r)!}(p^r)(1-p)^{n-r} \tag{3.19}$$

where $n =$ the sample size, the total number of trials
$r =$ the number of "events" occurring in n trials
$p =$ the constant probability for the event

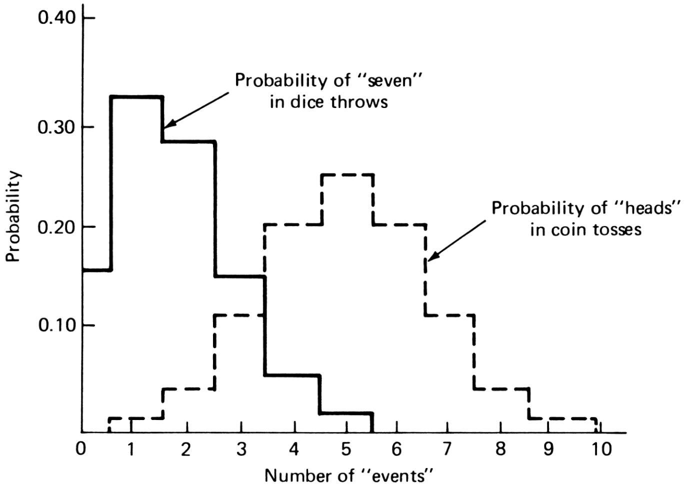


Figure 3.5 Binomial probability distributions.

In the example of tossing 10 coins, the probability of observing exactly four heads is:

$$\Pr(4) = \frac{10!}{4!\ \ 6!}(0.50)^4\ (0.50)^6 = 0.2051$$

The probability of observing 3 sevens in the dice example, where the probability of a seven is 1/6, is calculated as follows:

$$\Pr(3) = \frac{10!}{3!\ \ 7!}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = 0.1550$$

Table 3.4 Calculated Binomial Probabilities

Number of heads or sevens	10 coin tosses	10 dice throws
0	0.0010	0.1615
1	0.0098	0.3230
2	0.0439	0.2907
3	0.1172	0.1550
4	0.2051	0.0543
5	0.2461	0.0130
6	0.2051	0.0022
7	0.1172	$\left.\right\}$ 0.0003
8	0.0439	
9	0.0098	
10	0.0010	
	1.0001	1.0000

Usually of more interest than the probability of some specific number of observations is the probability of a specific quantity or a larger one; or a specific quantity and a smaller one. For example, if a shipment of small parts is supposed to have only 1 percent defectives, what is the probability of finding 2 or more defective ones in a sample of 50? What is the probability of finding 4 or more defective ones in a sample of 100? Or, if an organization claims that 80 percent of its students usually pass the Professional Engineers' examinations, what is the probability that out of a class of 90 only 65 or fewer students will pass?

To solve problems of this type it is necessary to obtain the sum of the binomial probabilities over the range of interest. With a sample of size n and a constant probability p, the probability of r or fewer events is:

$$\Pr(\leqq r) = \sum_{i=0}^{r} \left(\frac{n!}{i!\,(n-i)!}\right) p^i (1-p)^{n-i} \tag{3.20}$$

The probability of r or less than r equals unity minus the probability of more than r. The probability of r, given n and p, written $\Pr(=r|n,p)$, is equal to the probability of $(n-r)$ given n and $1-p$: $\Pr(=(n-r)|n, 1-p)$. In the calculator program discussed in the next section, the probability of Eq (3.20) is calculated, starting with the probability of zero events and summing the probabilities to the probability of r. Use is made of the following equalities to obtain other probabilities when they are desired.

$$\Pr(\leqq r|n,p) = \Pr(\geqq(n-r)|n, 1-p) = 1 - \Pr(>r|n,p) \tag{3.21}$$

$$\Pr(\geqq r|n,p) = \Pr(\leqq(n-r)|n, 1-p) = 1 - \Pr(<r|n,p) \tag{3.22}$$

The binomial program, in addition to calculating the probability associated with an input of n, p, and r, can also calculate the number of events expected with an input of n, p, and a probability value. The second type of solution is for problems exemplified by the two that follow: If a shipment of material is supposed to be not more than 0.1 percent defective, what maximum number of defectives would be expected in samples of 1000 95 percent of the time? If a candidate assumed that 51 percent of the electorate was on his side, what is the minimum number in his favor that he would expect in a poll of 100 voters 90 percent of the time? Answers to questions of this type are usually taken from curves of the binomial distribution and require interpolation to fit the particular problem. The binomial program gives exact answers.

3.6 BINOMIAL PROBABILITY PROGRAM

The binomial program carries out two general types of calculations. It calculates the probability for a particular binomial distribution, and, with a given probability, it calculates the binomial value. For both solutions the program follows the same formula, Eq. (3.20), so the two programs cannot be readily separated as is the case with the normal program.

Table 3.5 shows the calculations that can be made with the binomial program:

Table 3.5 Binomial Probability Program Operation

HP	TI	Gives
n, ENTER ↑, p, ENTER ↑, r, KEY A	n, KEY A, p, KEY A, r, KEY A	$\Pr(r)$
n, ENTER ↑, p, ENTER ↑, r, KEY B	n, KEY B, p, KEY B, r, KEY B	$\Pr(>r)$
n, ENTER ↑, p, ENTER ↑, r, KEY C	n, KEY C, p, KEY C, r, KEY C	$\Pr(\geqq r)$
n, ENTER ↑, p, ENTER ↑, r, KEY D	n, KEY D, p, KEY D, r, KEY D	$\Pr(<r)$
n, ENTER ↑, p, ENTER ↑, r, KEY E	n, KEY E, p, KEY E, r, KEY E	$\Pr(\leqq r)$
n, ENTER ↑, p, ENTER ↑, P, KEY c	n, KEY C', p, KEY C', P, KEY C'	r such that $\Pr(\geqq r) = P$
n, ENTER ↑, p, ENTER ↑, P, KEY e	n, KEY E', p, KEY E', P, KEY E'	r such that $\Pr(\leqq r) = P$

Note: n = sample size; p = constant probability; r = number of events; P = desired probability.

The actual calculation in all cases is for the probability equal to or less than r. All the other probabilities are obtained from that calculation. If $\Pr(r)$ is desired, KEY A, the difference between the last two calculations in the program: $\Pr(\leqq r)$ and $\Pr(\leqq r - 1)$ is used. If $\Pr(<r)$ is desired, the summation before the last term is taken. If $\Pr(>r)$ is wanted, $1 - \Pr(\leqq r)$ is used. If $\Pr(\geqq r)$ is wanted, $1 - \Pr(<r)$ is used. At the end of the program all of these values are stored in the calculator but only the one called for is printed.

The summation calculation for the probability equal to or less than r starts with the probability of zero:

$$\Pr(0) = \frac{n!}{0!\ n!}\, p^0(1-p)^n = (1-p)^n \tag{3.23}$$

Each term in the binomial summation is related to the previous term by the simple relation:

$$\Pr(i) = \Pr(i-1)\left(\frac{p}{1-p}\right)\left(\frac{n-i+1}{i}\right) \tag{3.24}$$

The use of Eq. (3.24) for calculating the individual terms after the first one not only simplifies the programming but also avoids the possibility of exceeding the factorial limits of the calculator if Eq. (3.19) were used for each term.

The calculation of the first term of the cumulative probability is Eq. (3.23). If p is close to one and n is large, $(1 - p)^n$ may be less than the lower limits of the calculator and give a zero answer. If this occurs, all subsequent terms will be zero since they are obtained by multiplication of the first term. The program has a check of the initial calculation, and if it is zero, the program substitutes one of the equivalent probabilities from Eq. (3.21) or (3.22).

The program also checks for impossible requests: $r > n$, or a probability input greater than one, for example. In these cases the program stops with an "error message." In the Hewlett-Packard, this is an actual display of the word "Error." In the Texas Instruments calculator, it is a flashing light. The input values are stored in Registers A, B, and C in the Hewlett-Packard and in Registers 01, 02, and 03 for verification after the program has been run.

The output for the first part of the program is a single value, the probability as a decimal fraction.

In the second type of operation, with an input of probability that is to be matched, the program checks the calculated probability after each calculation

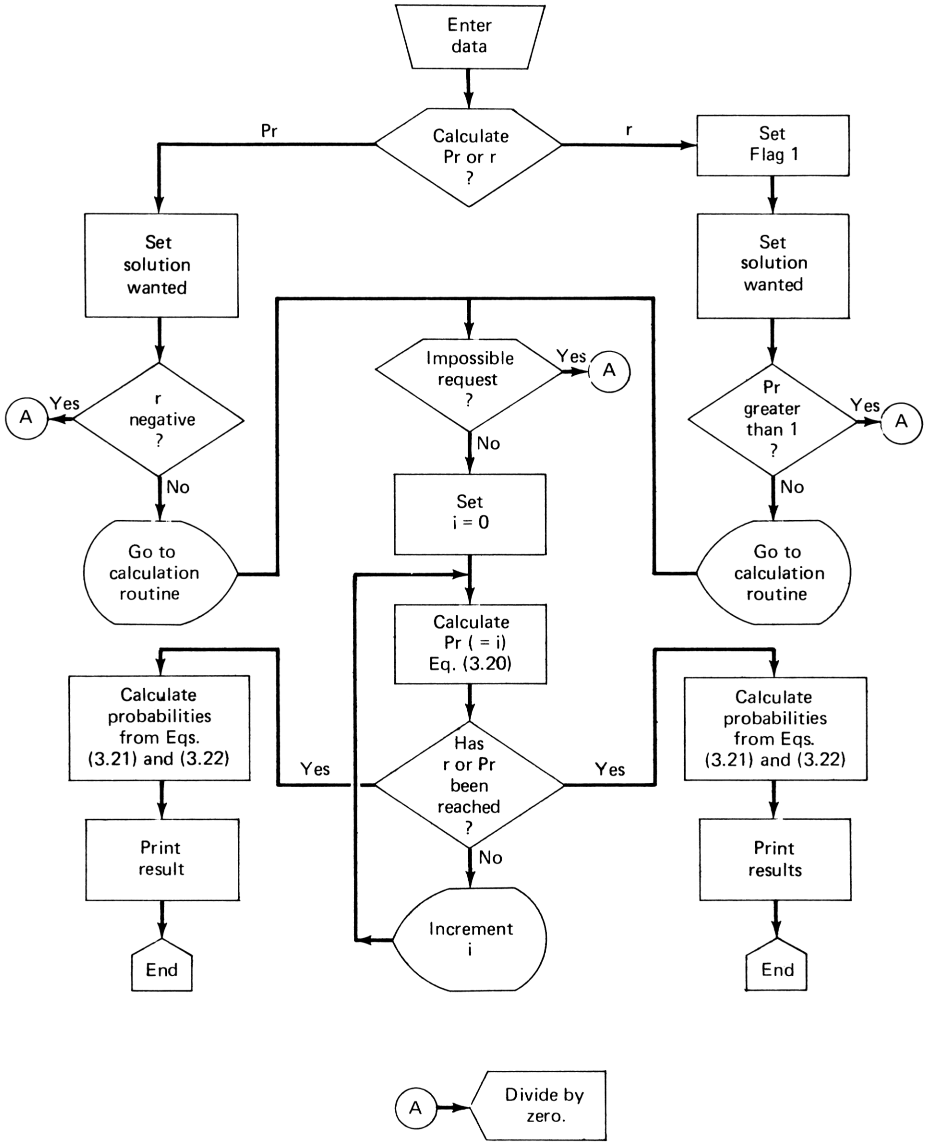


Figure 3.6 Flow diagram for binomial probability program.

Table 3.6 Register Values at Conclusion of Binomial Program

	Register		
KEYS A, B, C, D, E	HP	TI	KEYS c (C′), e (E′)
n	A	01	n
p	B	02	p
r	C	03	P
Pr(r)	1	04	Pr(r)
Pr($>r$)	2	05	
Pr($\geqq r$)	3	06	
Pr($<r$)	4	07	Pr($<r$)
Pr($\leqq r$)	5	08	Pr($\leqq r$)
r	I	09	r

against the input value. Since it is unlikely that the calculated probability will exactly equal the input value to the number of significant digits used by the calculator, the calculation continues until the input probability is exceeded. The program then prints the two calculated probabilities that bracket the input value along with the r values for each.

When the program ends, it clears all flags that were used to identify the different types of calculations so that the calculator may be used immediately for another binomial calculation. A flow diagram of the program is given in Fig. 3.6, and a listing and detailed description are given in the Appendix.

At the end of either type of calculation, the program prints the answer asked for. The values of the other probabilities are stored in the calculator and may be recovered if desired. Table 3.6 shows the location of the various values at the end of the program.

The binomial probability calculation is a summation of $r + 1$ terms. Each term takes slightly more than 1 s of calculator time; so if r is large, the calculator time may be longer than expected. It is possible to make approximations to the binomial probability, and the normal approximation is discussed in Sec. 3.8. However, the solution with the binomial program is exact, and the calculator time is not unreasonable as indicated by some of the following examples. The approximate calculation is available if several minutes of calculator time seem exorbitant.

3.7 EXAMPLES OF THE BINOMIAL PROBABILITY PROGRAM

3.7.1 In 100 tosses of a coin, what is the probability of observing 60 or more heads? $n = 100$, $p = 0.5$, $r = 60$; wanted: Pr($\geqq 60$).

	HP	TI	
100	ENTER ↑	KEY C	
0.5	ENTER ↑	KEY C	
60	KEY C	KEY C	Output: 0.0284 (Time:106/195)

3.7.2 A manufacturing process averages 0.2 percent defectives. What is the probability of finding less than 2 defectives in a sample of 1000? $n = 1000$, $p = 0.002$, $r = 2$; wanted: Pr(<2).

	HP	TI		
1000	ENTER ↑	KEY D		
0.002	ENTER ↑	KEY D		
2	KEY D	KEY D	Output:	0.4057 (Time:7/11)

3.7.3 If you toss 10 coins, what is the probability of exactly 5 heads? $n = 10$, $p = 0.5$, $r = 5$; wanted: Pr(5).

	HP	TI		
10	ENTER ↑	KEY A		
0.5	ENTER ↑	KEY A		
5	KEY A	KEY A	Output:	0.2461 (Time:11/20)

3.7.4 A person rolls 30 dice and you wager 3 to 1 that he or she will not roll a certain number, say "sixes," more than X times. What is the minimum value of X so that you will come out about even?

If you give 3:1 odds, you cannot afford to lose more than 25 percent of the time; that is, Pr $\leqq 0.25$. The probability of any specific number for one roll of a die is 1/6. $n = 30$, $p = 1/6$, $P \leqq 0.25$; wanted: the value of r such that Pr($\geqq r$) $\leqq 0.25$.

	HP	TI				
30	ENTER ↑	KEY C'				
1	ENTER ↑	÷				
6	÷	=				
		KEY C'				
0.25	KEY c	KEY C'	Output:	0.2235	(Pr $\geqq r$)	
				7	(r)	
				0.3836	(Pr $\geqq r$)	
				6	(r)	(Time:17/34)

The probability of 7 or more is 0.2235
The probability of 6 or more is 0.3836

You will be safe with a bet at 3:1 odds if you wager that your opponent will not roll a specific number more than six times in 30 rolls.

3.7.5 If a shipment of small parts is supposed to have only 1 percent defectives, what is the probability of finding 2 or more defectives in a sample of 50? $n = 50$, $p = 0.01$, $r = 2$; wanted: Pr($\geqq$2).

	HP	TI	
50	ENTER ↑	KEY C	
0.01	ENTER ↑	KEY C	
2	KEY C	KEY C	Output: 0.0894 (Time:6/12)

3.7.6 If the previous problem had been phrased differently, asking for the probability of finding no more than 48 nondefects, the procedure would be: $n = 50$, $p = 0.99$, $r = 48$; wanted: Pr(≦48).

	HP	TI	
50	ENTER ↑	KEY E	
0.99	ENTER ↑	KEY E	
48	KEY E	KEY E	Output: 0.0894 (Time:8/17)

The answer is the same as in Example 3.7.5. The actual calculation was made for $p = 0.01$ and $r = 2$, and the probability ≧ 2 was calculated. The reason is that the program made a substitution based on Eqs. (3.18) and (3.19) inasmuch as Pr(0) based on the original figures, $(0.01)^{50}$, is less than the smallest number the calculator can handle. This substitution is pointed out in the description of the program in the Appendix.

3.7.7 If an organization claims that 80 percent of its students usually pass the Professional Engineers' examinations, what is the probability that out of a class of 90 only 65 or less will pass? $n = 90$, $p = 0.80$, $r = 65$; wanted: Pr(≦65).

	HP	TI	
90	ENTER ↑	KEY E	
0.80	ENTER ↑	KEY E	
65	KEY E	KEY E	Output: 0.0474 (Time:115/240)

3.7.8 If a shipment of material is supposed to be not more than 0.1 percent defective, what maximum number of defects in samples of 1000 would be expected 95 percent of the time? $n = 1000$, $p = 0.001$, $P = 0.95$; wanted: r such that Pr(≦r) = 0.95.

	HP	TI			
1000	ENTER ↑	KEY E'			
0.001	ENTER ↑	KEY E'			
0.95	KEY e	KEY E'	Output:	0.9811 (Pr ≦ r)	
				3 (r)	
				0.9198 (Pr ≦ r)	
				2 (r)	(Time:7/19)

The probability of three or fewer defects is 0.9811, and the probability of two or fewer defects is 0.9198. Two defects could be expected 92 percent of the time and three defects 98 percent of the time.

3.8 NORMAL APPROXIMATION TO THE BINOMIAL

As mentioned earlier, because of the cumulative nature of the binomial probability calculation, the calculator time may be more than expected. One way to cut down on calculator time is to use the normal program for an approximation to the binomial probability calculation. How close the approximation will be depends on the values of n, p, and r. Some examples are given in the following paragraphs.

The larger the value of n, the closer p is to 0.5, and the closer r is to $n \cdot p$, the better will be the approximation. No exact rule can be given as to when the normal approximation will be satisfactory. Examination of the following examples and some experience with specific data will indicate when to substitute the approximate calculation for the exact binomial value in order to save calculator time.

The mean of the binomial distribution is $n \cdot p$, and the standard deviation is $\sqrt{n \cdot p \cdot (1-p)}$. These values are exact and may be substituted in the normal program for the corresponding values of m and σ. The value of r needs to be adjusted for the normal estimate of the binomial because of the difference in the two types of probability distribution functions.

The binomial function deals with discrete values of the variable. The probability of a variable being equal to or greater than the value r is one minus the probability of the variable being equal to or less than the value $r - 1$. $\Pr|\mathrm{B}(\geqq r) = 1 - \Pr|\mathrm{B}(\leqq r - 1)$, where Pr|B indicates binomial probability. The normal distribution function is a continuous function and there is no gap in the normal probability between successive values. The probability of a variable being equal to or greater than the value r, in the normal distribution, is one minus the probability of the variable being equal to or less than the value r. $\Pr|\mathrm{N}(\geqq r) = 1 - \Pr|\mathrm{N}(\leqq r)$, where Pr|N indicates normal probability. Therefore, when the normal probability is used to estimate the binomial, the variable values used are adjusted for one-half the difference between successive numbers. For the normal estimate of the binomial probability for a value equal to or greater than 6, we use a value of 5.5. For the probability of a value greater than 6, estimated by the normal probability, we would calculate the probability of a value greater than 6.5.

Table 3.7, a table of comparison of the probabilities calculated by both programs, illustrates both the method and the degree of agreement.

The agreement between the two calculations can be seen from Table 3.7. When r is large, 60 and 65 in Examples 3.7.1 and 3.7.3, the agreement is reasonably good. When r is small, in Example 3.7.2, the agreement is poor. Whether the time saved, perhaps 100s, is worth the loss in accuracy will depend on the problem and the engineer running it. If the normal program is to be used as an estimate of the binomial probability calculation, it would be simple to add a short routine to calculate the mean and standard deviation directly from the input binomial parameters so no preliminary calculations would be required.

Table 3.7 Comparison of Normal and Binomial Probability Program Calculations

Data	Binomial			Normal			
		HP	TI			HP	TI
EXAMPLE 3.7.1							
$n=100$	100	ENTER ↑	KEY C	Mean $=(100)(0.5)=50$	50	ENTER ↑	KEY C
$p=0.5$	0.5	ENTER ↑	KEY C	$\sigma=\sqrt{(100)(0.5)(0.5)}=5$	5	ENTER ↑	KEY C
$r=60$	60	KEY C	KEY C	$x=r-0.5=59.5$	59.5	KEY C	KEY C
$\Pr(\geqq r)=?$				$\Pr(\geqq x)=?$			
	Answer: 0.0284				Answer: 0.0287		
	Time: 106/195				Time: 6/8		
EXAMPLE 3.7.2							
$n=1000$	1000	ENTER ↑	KEY D	Mean $=(1000)(0.002)=2$	2	ENTER ↑	KEY E
$p=0.002$	0.002	ENTER ↑	KEY D	$\sigma=\sqrt{(1000)(0.002)(0.998)}=1.4128$	1.4128	ENTER ↑	KEY E
$r=2$	2	KEY D	KEY D	$x=r-0.5=1.5$	1.5	KEY E	KEY E
$\Pr(<r)=?$				$\Pr(\leqq x)=?$			
	Answer: 0.4057				Answer: 0.3617		
	Time: 7/11				Time: 6/8		

EXAMPLE 3.7.4

$n=30$	30	ENTER ↑	KEY C'	Mean $=(30)(1/6)=5$	5	ENTER ↑	KEY C'
$p=1/6$	1	ENTER ↑	÷	$\sigma=\sqrt{(30)(1/6)(5/6)}=2.0412$	2.0412	ENTER ↑	KEY C'
$P=0.25$	6	÷	=	$P\|N=P\|B=0.25$	0.25	KEY c	KEY C'
$P=$ Pr(≧?)			KEY C'	$P=$ Pr(≧?)			
	0.25	KEY c	KEY C'				
	Answer:	7 for Pr = 0.2235			Answer:	6.38	
		6 for Pr = 0.3836			Time:	7/9	
	Time:	17/34					

EXAMPLE 3.7.7

$n=90$	90	ENTER ↑	KEY E	Mean $=(90)(0.8)=72$	72	ENTER ↑	KEY E
$p=0.8$	0.8	ENTER ↑	KEY E	$\sigma=\sqrt{(90)(0.8)(0.2)}=3.7947$	3.7947	ENTER ↑	KEY E
$r=65$	65	KEY E	KEY E	$x=r+0.5=65.5$	65.5	KEY E	KEY E
Pr(≦r) = ?				Pr(≦x) = ?			
	Answer:	0.0474			Answer:	0.0434	
	Time:	115/240			Time:	6/8	

3.9 POISSON PROBABILITY DISTRIBUTION

The binomial distribution describes the frequency of events from a population of a definite size with a constant probability. The Poisson distribution describes the distribution of events from a large but indefinite population, where the probability is small but a definite expected number can be calculated; for example, the number of accidents occurring on a section of an interstate highway during half-hour periods from 8 A.M. to 6 P.M., or the number of Supreme Court Justices appointed in a calendar year.

The Poisson distribution involves only one parameter, the "expected value." In the 1978 baseball season of 1936 games, there were 2956 home runs, or 1.527 home runs per game. If this value is taken as the expected number of home runs per game, the Poisson distribution function can be used to calculate the probability that in the next game you see there will be no home runs, or more than three, or any other number. The expected value, usually designated m in the Poisson distribution, corresponds to the mean of the normal distribution and to $n \cdot p$ of the binomial.

The Poisson probability distribution is based on the following equation:

$$\Pr(r) = e^{-m}\frac{m^r}{r!} \tag{3.25}$$

The use of Eq. (3.25) might best be described by an example. In the 188 years from 1790 to 1978, 100 Justices have been appointed to the U.S. Supreme Court.[3] This is an average of 100/188 or 0.5319 Justices per year. With this value and Eq. (3.25) it is possible to calculate the probability of no Justices being appointed during a year, of one Justice, and so forth.

$$\Pr(0) = e^{-0.5319}\,(0.5319)^0/0! = 0.5875$$
$$\Pr(1) = e^{-0.5319}\,(0.5319)^1/1! = 0.3125$$
$$\vdots$$

The product of the calculated probabilities and the total number of years gives the predicted number of years that that number of Justices would be appointed. The probability of zero Justices, 0.5875, times the total number of years, 188, equals a predicted 110.4 as the number of years (out of 188) that no Justices would be appointed. Table 3.8 shows the predicted number and the actual number.

Although the data from the appointment of Supreme Court Justices is an extremely good fit, it is not unusual. The Poisson distribution has been fitted to the number of men killed by the kick of a horse, to the occurrence of rainstorms, and to the number of lost articles turned in at a busy office building. With data of an accidental nature with low probability of occurrence in a large population, the Poisson distribution often gives a reliable prediction.

One use of the Poisson probability is to size facilities so as to prepare for the expected number of events, but not to oversize them to handle the events that

Table 3.8 Number of U.S. Supreme Court Justices Appointed per Year

No. appointed a year	Poisson probability	Number of years for each number: Predicted	Actual
0	0.5875	110.4	110
1	0.3125	58.8	60
2	0.0831	15.6	15
3	0.0147	2.8	2
4	0.0020	0.4	1

have a low probability of occurrence. In designing a truck loading facility, the design should accommodate the usual number of trucks, but it need not be so large as to handle the number that would occur only once in 100 times.

As with the binomial probability, the Poisson calculation of most interest is the probability of the number of events equal to or greater than some specification; for example, the probability of 10 or more trucks in a half-hour period. Of equal interest is the probability of some number or one that is less; for example, what is the probability of a blood count lower than some specified value?

The calculation of a value greater than or less than probability involves the summation of the individual probabilities. For the probability of r or fewer events, the calculation sums the values of Eq. (3.25) from zero to r:

$$\Pr(\leqq r) = \sum_{i=0}^{r} e^{-m} \frac{m^i}{i!} \tag{3.26}$$

The cumulative Poisson probability is easily computed in a calculator from Eq. (3.26), since the probability of zero events is e^{-m}, and the probability of each successive term is the probability of the previous term multiplied by m/i:

$$\Pr(i+1) = \Pr(i) \frac{m}{i+1} \tag{3.27}$$

Most calculators have a limitation in the numbers they can handle, about 9×10^{99}. This limitation would prevent a Poisson probability calculation for m greater than 227. A way around this limitation is included in the Poisson probability program discussed in the next section. However, with expected values greater than about 100, a normal approximation to the Poisson may be close enough. A comparison between the Poisson probabilities and the normal approximation is given in Sec. 3.12.

Individual Poisson probabilities calculated from Eq. (3.25) require a factorial evaluation. Factorials of numbers larger than 69 exceed most calculators' capacity, so that if the multiplying factor of Eq. (3.27) was not used, a very strict limitation would be put on calculating the Poisson probabilities.

3.10 POISSON PROBABILITY PROGRAM

The Poisson calculation requires input of only the expected value m, and then either a variable value r, or a probability value P. With a variable value the program will calculate the probability; and with a probability input, the program will calculate the variable value. As with the binomial program, the program will calculate probabilities equal to, less than, or greater than the input variable. Table 3.9 shows the operation of the Poisson program.

All calculations start with the probability of zero events: $e^{-m}(m^0/0!) = e^{-m}$. Each successive term is calculated by Eq. (3.27). The calculation continues until the value of r is obtained, or until the cumulative probability first exceeds P. The last two cumulative probabilities are stored in the calculator after each calculation, and all the probabilities in Table 3.9 are calculated from these two. Thus

$$\Pr(\leqq r) = \Pr(r) + \Pr(<r) \tag{3.28}$$

$$\Pr(\geqq r) = 1 - \Pr(<r) \tag{3.29}$$

$$\Pr(>r) = \Pr(\geqq r) - \Pr(r) \tag{3.30}$$

All of the probabilities are available in the calculator at the end of the calculation, but only the one asked for is printed.

If there is probability input to find the variable value to match the probability, there is output of two values, the calculated probability that just exceeds the input value together with the corresponding r value, and the probability just before with its variable value, so that the input probability is bracketed.

The initial calculation for the probability summation is e^{-m}. If m is greater than 227, it will exceed the capacity of the calculator. Before the start of the calculation the program checks the value of m and, if it is greater than 227, it is factored in smaller increments which are easily handled. The following relation is used:

$$e^{-m} = e^{-x} \cdot e^{-y} \cdot e^{-z}$$

where $x + y + z = m$.

Table 3.9 Poisson Probability Program Operation

HP	TI	Gives
m, ENTER ↑, r, KEY A	m, KEY A, r, KEY A	$\Pr(r)$
m, ENTER ↑, r, KEY B	m, KEY B, r, KEY B	$\Pr(>r)$
m, ENTER ↑, r, KEY C	m, KEY C, r, KEY C	$\Pr(\geqq r)$
m, ENTER ↑, r, KEY D	m, KEY D, r, KEY D	$\Pr(<r)$
m, ENTER ↑, r, KEY E	m, KEY E, r, KEY E	$\Pr(\leqq r)$
m, ENTER ↑, P, KEY c	m, KEY C', P, KEY C'	r such that $\Pr(\geqq r) = P$
m, ENTER ↑, P, KEY e	m, KEY E', P, KEY E'	r such that $\Pr(\leqq r) = P$

Note: m = exected value; r = number of events; P = desired probability.

Table 3.10 Register Values at Conclusion of Poisson Program

	Register		
KEYS A, B, C, D, E	HP	TI	KEYS c, (C′), e, (E′)
m	A	01	m
r	B	02	P
Pr(r)	1	03	Pr(r)
Pr($>r$)	2	04	
Pr($\geqq r$)	3	05	
Pr($<r$)	4	06	Pr($<r$)
Pr($\leqq r$)	5	07	Pr($\leqq r$)
r	I	08	r

With the operation just described there is no limit to the Poisson probability calculation, aside from calculator time. There is no other readily available source for getting Poisson probabilities for expected values greater than 100. The calculator time is about 1.5 s for each calculation in the summation, so that with r values of 50 or greater, the calculator time may be more than is usually required, but there is no other source for an exact Poisson probability.

At the end of the calculation, the input data and the unprinted probability values are available in the calculator. The location of these values is shown in Table 3.10.

Figure 3.7 gives the flow of the Poisson program. Details and a listing are given in the Appendix.

3.11 EXAMPLES OF THE POISSON PROBABILITY PROGRAM

3.11.1 If the average red-cell count in a certain volume of blood is 20, what is the probability of a normal person having a count below 16? $m = 20$; wanted: Pr($\leqq$15).

	HP	TI	
20	ENTER ↑	KEY E	
15	KEY E	KEY E	Output: 0.1565 (Time:23/33)

3.11.2 From the data of Table 3.8, the expected number of Justices appointed per year is 0.5319. What is the probability of three Justices being appointed next year? $m = 0.5319$; wanted: Pr(3).

	HP	TI	
0.5319	ENTER ↑	KEY A	
3	KEY A	KEY A	Output: 0.0147 (Time:6/10)

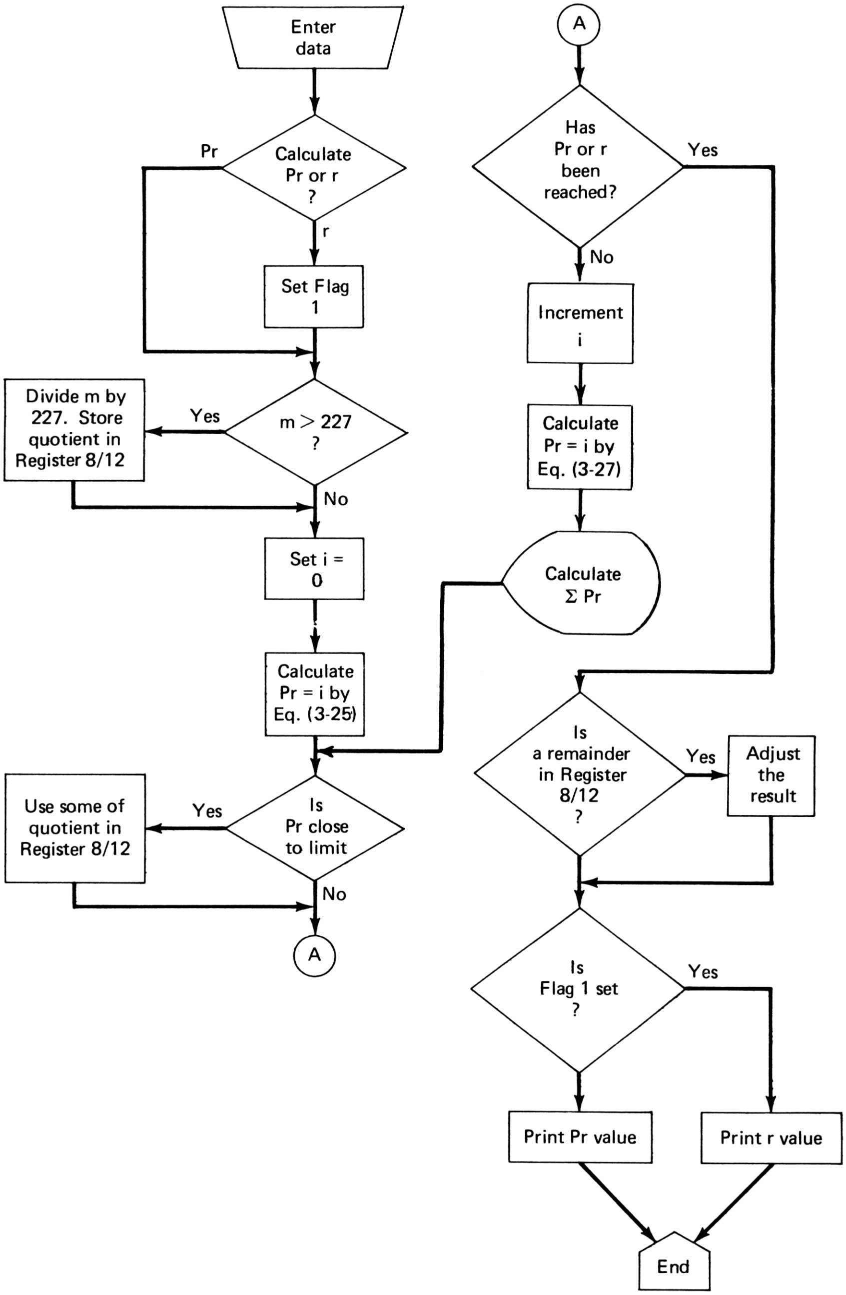


Figure 3.7 Flow diagram for Poisson probability program.

3.11.3 A switchboard usually gets 10 calls every minute between 9:30 and 11:30 A.M. It will be overtaxed with 20 calls. What is the probability that in the next minute it will be overtaxed? $m = 10$; wanted: Pr(≧20).

	HP	TI	
10	ENTER ↑	KEY C	
20	KEY C	KEY C	Output: 0.0035 (Time:31/48)

3.11.4 With the data of Example 3.11.3, what is the probability the switchboard will be overtaxed at least once in the next hour?

The probability it will be overtaxed at least once is one minus the probability it will not be overtaxed at all. The probability it will not be overtaxed in the next minute (from Example 3.11.3) is $1 - 0.0035 = 0.9965$. The probability it will not be overtaxed in the next 60 minutes (min) is $(0.9965)^{60} = 0.8103$. The probability it will be overtaxed at least once in the next hour is $1 - 0.8103 = 0.1897$.

3.11.5 If the usual number of accidents under certain conditions is 14, what number should be prepared for to keep the probability of being overtaxed at 0.05 or less? $m = 14$; wanted: r such that Pr($\geq r$) ≦ 0.05.

	HP	TI		
14	ENTER ↑	KEY C'		
0.05	KEY c	KEY C'	Output:	0.0479 (Pr ≧ r)
				21 (r)
				0.0765 (Pr ≧ r)
				20 (r) (Time:32/54)

The probability of 20 or more accidents is 0.0765.
The probability of 21 or more accidents is 0.0479.

If preparation is made for 21 accidents, when the expected number is 14, there is only a 0.0479 probability of being overtaxed.

3.12 NORMAL APPROXIMATION TO THE POISSON

The mean of the Poisson distribution is m, the expected value, and the standard deviation is $\sqrt{m}$. These values may be used with the normal program to estimate probabilities corresponding to the Poisson calculation. When the normal probability is used to approximate the Poisson, it is necessary to correct for the continuous nature of the normal compared to the discrete values for which the Poisson is applicable. The correction is the same as that discussed for the binomial program in Sec. 3.8. A correction of half a unit is made to the input value for the normal approximation. Table 3.11 shows the values used for the various cases:

Table 3.11 Adjustments to Input Values for the Normal Approximation to the Poisson Probability

Poisson calculation	Normal approximation
Pr(5)	Pr(4.5 $\leqq r \leqq$ 5.5)
Pr(>5)	Pr(>5.5)
Pr(≧5)	Pr(≧4.5)
Pr(<5)	Pr(<4.5)
Pr(≦5)	Pr(≦5.5)

The closeness of the normal approximation to the Poisson calculation will depend on the values of m and r. The larger the value of m and the closer r is to m, the closer will be the normal approximation. The following comparisons show the amount of agreement. The same data that are used for the binomial and normal comparison are used for the Poisson and normal so that the similarity of the three distributions might be observed. The comparison is shown in Table 3.12.

A very simple adjustment to the normal program permits the mean and standard deviation to be calculated directly with input of m. Again, however, the Poisson calculation, although it is somewhat longer, gives the exact probabilities without any exact calculations.

Table 3.12 Comparison of Normal and Poisson Probability Program Calculations

Data	Poisson calculation HP	Poisson calculation TI	Normal calculation	Normal calculation HP	Normal calculation TI
$m = 50$	50 ENTER ↑	KEY C	Mean = 50	50 ENTER ↑	KEY C
$r = 60$	60 KEY C	KEY C	$\sigma = \sqrt{50} = 7.071$	7.071 ENTER ↑	KEY C
Pr(≧r) = ?				59.5 KEY C	KEY C
	Answer: 0.0923			Answer: 0.0895	
	Time: 90/132			Time: 6/8	
$m = 2$	2 ENTER ↑	KEY D	Mean = 2	2 ENTER ↑	KEY E
$r = 2$	2 KEY D	KEY D	$\sigma = \sqrt{2} = 1.414$	1.414 ENTER ↑	KEY E
Pr(<r) = ?				1.5 KEY E	KEY E
	Answer: 0.4060			Answer: 0.3618	
	Time: 6/9			Time: 6/8	
$m = 0.5$	0.5 ENTER ↑	KEY C	Mean = 0.5	0.5 ENTER ↑	KEY C
$r = 2$	2 KEY C	KEY C	$\sigma = \sqrt{0.5} = 0.7071$	0.7071 ENTER ↑	KEY C
Pr(≧r) = ?				1.5 KEY C	KEY C
	Answer: 0.0902			Answer: 0.0786	
	Time: 6/9			Time: 6/8	
$m = 72$	72 ENTER ↑	KEY E	Mean = 72	72 ENTER ↑	KEY E
$r = 65$	65 KEY E	KEY E	$\sigma = \sqrt{72} = 8.485$	8.485 ENTER ↑	KEY E
Pr(≦5) = ?				65.5 KEY E	KEY E
	Answer: 0.2242			Answer: 0.2218	
	Time: 96/148			Time: 6/8	
$m = 5$	5 ENTER ↑	KEY C'	Mean = 5	5 ENTER ↑	KEY C'
$P = 0.25$	0.25 KEY c	KEY C'	$\sigma = \sqrt{5} = 2.236$	2.236 ENTER ↑	KEY C'
?:P = Pr(≧?)				0.25 KEY c	KEY C'
	Answer: 7 for Pr = 0.2378				
	6 for Pr = 0.3840			Answer: 6.508	
	Time: 15/22			Time: 7/9	

4

THE *t* TEST

4.1 INTRODUCTION

In the discussion of the normal distribution it was pointed out that if the mean and the standard deviation of the distribution were known, the distribution was defined and probability statements could be made about the value of observations drawn from the population. The t distribution is used in a similar manner where the mean of samples and the standard deviation estimated from the samples are used to make probability statements about the value of observations in the population from which the samples were drawn. The estimate of the standard deviation is calculated as described in Chap. 2:

$$s(x) = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \tag{4.1}$$

This chapter gives two t programs that provide for making the following calculations based on the t distribution and data from samples.

1. Comparison of a mean calculated from input data with a target value.
2. Comparison of two means each calculated from input data.
3. Calculation of the probability value associated with an input value of t and the amount of data from which t was calculated. (This calculation is equivalent to finding the value in a t table.)
4. Finding the confidence interval for a mean from input data and the desired confidence probability.
5. Finding the number of samples required when a specific confidence interval and probability are set.
6. Finding the t value associated with an input probability level and sample size. (This calculation is equivalent to using the t tables in the reverse way from item 3, above.)

4.2 THE *t* FUNCTION

With a variable x that is normally distributed with a mean m and a standard deviation σ, the probability of observing a value equal to or greater than some value x_1 is equal to the area under the standardized normal curve beyond the value of z_1, where z_1 was defined in Chap. 3 as:

$$z_1 = \frac{x_1 - m}{\sigma} \tag{3.7}$$

The t distribution is used in a similar manner. The probability of the mean of a sample being equal to or greater than $\bar{x}$ when the population mean is m, is the area under the t curve beyond the value indicated by the following equation:

$$t_\nu = \frac{\bar{x} - m}{s(\bar{x})} \tag{4.2}$$

In Eq. (4.2) there are two new terms, ν and $s(\bar{x})$. The term ν represents the degrees of freedom, the number of items in the sample that could be independently varied while the mean from which the deviations are measured to calculate the variance estimate remains fixed. In this example, ν is equal to $n - 1$. The term $s(\bar{x})$ is the standard deviation estimate of the mean, and is related to the standard deviation estimate by the relation: $s(\bar{x}) = s(x)/\sqrt{n}$

The probability distribution of t varies with the number of degrees of freedom associated with its calculation. The t values approach the normal function values as the degrees of freedom increase. For degrees of freedom greater than about 120, the difference is usually taken to be insignificant, and t table values are usually not given for degrees of freedom beyond 120. In the programs given later, there is no limit to the degrees of freedom that may be used.

The most common use of the t function is to establish some hypothesis, called a "null hypothesis," for which t may be calculated. If the resultant t is one of low probability when the hypothesis is true, the hypothesis is rejected. The steps in this procedure are listed below:

1. A null hypothesis is made regarding some parameter of a population from which a sample is taken. (Example: H_0: Population mean equals 100.)

2. A probability value is selected for an acceptable risk of a false rejection of the hypothesis. This would be rejecting the hypothesis, saying the mean was not 100, when it really was. The probability is usually designated α. (Example: $\alpha = 0.05$; a 5 percent chance of a false rejection.)

3. The value of t is calculated from a suitable equation, with statistics calculated from the sample. [Example: Use Eq. (4.2), where $\bar{x}$ is the sample mean and $s(\bar{x})$ is the sample estimate of the standard deviation of the mean, $t = (\bar{x} - 100)/s(\bar{x})$.]

4. The calculated t from step 3 is compared with values in the t tables at the degrees of freedom of the calculation, $n - 1$.

5. If the calculated t is larger than the tabulated t for the degrees of freedom involved at the probability α set at step 2, the hypothesis is rejected.

In the t programs that follow, the actual probability associated with the calculated t is given as output. If this probability is less than the α value set for the test, the hypothesis is rejected.

4.2.1 Two-Sided t Test

The t distribution is symmetrical about a zero mean, and only the positive values of t are tabulated. The probabilities associated with these values are for the sum of the areas under the t probability curve beyond the positive value and below the negative value. In step 5 of the procedure given above, the hypothesis is rejected if t is larger because $\bar{x}$ is greater than m, or if $-t$ is larger because $\bar{x}$ is smaller than m. There are occasions when only one side of the test is of interest. A discussion of the one-sided test is given in the next section. Numerical examples are given at the end of the chapter. The following is typical of the two-sided test.

With a continuing manufacturing process the hypothesis is made that the mean of some quantity is equal to m:

$$H_0\text{: mean} = m$$

The manufacturer is willing to take a 1 percent risk of falsely rejecting this hypothesis, that is, saying the mean is not m when it really is:

$$\alpha = 0.01$$

Samples are taken from the process, and measurements are made of n items. The sample measurements, along with the value of m are entered into the calculator. The program calculates t from the relation $t = (\bar{x} - m)/(s(x)/\sqrt{n})$, although t from Eq. (4.3) is actually the value of interest:

$$t = \frac{|\bar{x} - m|}{s(x)/\sqrt{n}} \tag{4.3}$$

where $s(x)$ is the estimated standard deviation, calculated from the data as explained in Chap. 2, and n is the number of data points, counted by the calculator.

As output, the program gives the mean $\bar{x}$, the t value, the degrees of freedom, $n - 1$, and the probability associated with t and $n - 1$. If the probability is less than α the hypothesis is rejected.

Note that the hypothesis would be rejected if $\bar{x}$ was significantly larger or significantly smaller than m and t from the calculation was positive or negative, although only the absolute value of t is of interest.

4.2.2 One-Sided *t* Tests

If purity of product concerned the manufacturer in the previous example, impurity values lower than standard would be of no interest, but too-high impurity values would be of concern. The hypothesis in this case would be:

$$H_0\text{: mean} \leqq m$$

and after a risk of a false rejection was set, t would be calculated in the same manner by the calculator, but only the value from Eq. (4.4) would be of interest:

$$t = \frac{\bar{x} - m}{s(x)/\sqrt{n}} \tag{4.4}$$

The hypothesis would be rejected if the calculated α value was more than twice the critical value set before the test, *and* if t was positive. The hypothesis would be accepted for all negative values of t, indicating that the mean was less than m. Table 4.1 at the end of this section summarizes the different t tests.

The same reasoning would apply if the result was of interest only if it showed the sample mean to be too low. A manufacturer who was interested in yield might not care if the yield exceeded specifications, but would be interested in finding if it was too low. In this case the hypothesis and the t equation would be:

$$H_0\text{: mean} \geqq m \qquad t = \frac{m - \bar{x}}{s(x)/\sqrt{n}} \tag{4.5}$$

In this case the hypothesis would be rejected if the probability value calculated for t was more than twice the critical value *and* t was negative. The value of t is calculated by the calculator program from Eq. (4.4), so although the proper t calculation is that of Eq. (4.5), it would have to be indicated by a negative value of t.

4.2.3 Comparison of Two Means

Another method of using the t function is to compare the means of measurements from two populations. The hypothesis in this case is that the two population means are equal:

$$H_0\text{: mean}_x = \text{mean}_y$$

As before, a probability level is set for a false rejection of the hypothesis, and a t value is calculated from the following equation:

$$t = \frac{|\bar{x} - \bar{y}|}{\bar{s}(x)\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}} \tag{4.6}$$

where $\bar{x}$ and $\bar{y}$ are the means calculated from samples from the two different populations, $\bar{s}(x)$ is a mean estimate of the standard deviation calculated from the estimates of the variances of the two samples and from weighting the variance estimates by the degrees of freedom of each. The equation for $\bar{s}(x)$ is:

$$\bar{s}(x) = \sqrt{\frac{(n_x - 1)\, s^2(x) + (n_y - 1)\, s^2(y)}{(n_x - 1) + (n_y - 1)}} \tag{4.7}$$

The evaluation of the calculated t is similar to that in the previous test. If the calculated t exceeds the tabulated t for the probability selected and $(n_x + n_y - 2)$ degrees of freedom, the hypothesis is rejected.

The probability associated with the t calculation from Eq. (4.6) allows for rejection of the hypothesis when either mean is significantly larger than the other. This is the two-sided test. As in the previous discussion, it is possible to test whether one mean is greater than the other. If a vendor's material is claimed to be better in some way than the standard material, it might be reasonable for a customer to test whether or not this was the case and not be interested in determining if the new material was poorer than the standard. To test whether mean_x was greater than mean_y, the hypothesis that it was not larger would be set; H_o: $\text{mean}_x \leqq \text{mean}_y$, and an α value would be selected. Samples from the two materials would be taken and the means determined; t would be calculated from the following equation:

$$t = \frac{\bar{x} - \bar{y}}{\bar{s}(x)\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}} \tag{4.8}$$

The calculated t would be compared with $t_{2\alpha}$ and only the positive value of t would be considered. If $\bar{y}$ was greater than $\bar{x}$, giving a negative t from Eq. (4.8), the hypothesis would be accepted that mean_x was not greater than mean_y. Or, if the positive value of t had a probability greater than 2α, the hypothesis would be accepted. If the positive value of t had a probability equal to or less than 2α, the hypothesis would be rejected.

Table 4.1 gives a summary of the one- and two-sided hypothesis tests that have been discussed and shows their calculation with program t-1.

If the variance estimates are significantly different, standard deviation estimates should not be pooled by Eq. (4.7). This restriction is often overlooked in t tests because data from similar sources usually have variances that are not significantly different. Chapter 6, dealing with variance testing, provides a method of testing the homogeneity of variances. If the standard deviations are not poolable by the variance test, the validity of the t test is not affected, but the interpretation is slightly different. If the t value is not significant, the reason may not be that there is not a real difference between the means but that the variation within the measurements is so different that the difference between the means cannot be detected. The Variance-1 program for testing means gives the standard deviation estimates for the two sets of data. Familiarity with the variance test procedure will indicate when the pooling of the standard deviations needs to be questioned.

Table 4.1 Summary of *t* Tests for Means

Test:	Population mean $\neq m$	Population mean $> m$	Population mean $< m$
Hypothesis, H_0	Population mean $= m$	Population mean $\leqq m$	Population mean $\geqq m$
	Set α, probability of a false rejection of H_0		
	Run program t-1,: Data, KEY A; m, KEY C		
	Program calculates: $t = \dfrac{x - m}{(s(x)/\sqrt{n})}$		
	$P = \Pr(\geqq \lvert t \rvert)$		
Accept H_0 if:	$P > \alpha$	t is negative or t is positive and $P > 2\alpha$	t is positive or t is negative and $P > 2\alpha$
Reject H_0 if:	$P \leqq \alpha$	t is positive and $P \leqq 2\alpha$	t is negative and $P \leqq 2\alpha$
Test:	$\text{Mean}_1 \neq \text{Mean}_2$	$\text{Mean}_1 > \text{Mean}_2$	$\text{Mean}_1 < \text{Mean}_2$
Hypothesis, H_0	$\text{Mean}_1 = \text{Mean}_2$	$\text{Mean}_1 \leqq \text{Mean}_2$	$\text{Mean}_1 \geqq \text{Mean}_2$
	Set α, probability of a false rejection of H_0		
	Run t program 1,: Data_1, KEY A; Data_2, KEY B; KEY C		
	Program calculates: $= (x_1 - x_2)/\left(\bar{s}(x)\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$		
	$P = \Pr(\geqq \lvert t \rvert)$		
Accept H_0 if:	$P > \alpha$	t is negative or t is positive and $P > 2\alpha$	t is positive or t is negative and $P > 2\alpha$
Reject H_0 if:	$P \leqq \alpha$	t is positive and $P \leqq 2\alpha$	t is negative and $P \leqq 2\alpha$

If the ratio of the standard deviation estimates is much greater than 2.0, depending on the degrees of freedom involved, the pooling might be questioned.

4.3 CONFIDENCE INTERVALS

Another use of the t function is to set confidence limits to a measured mean. Equation (4.3) may be rearranged to the following form:

$$m = \bar{x} \pm \frac{t_\alpha \cdot s(x)}{\sqrt{n}} \tag{4.9}$$

If the hypothesis involved would be rejected with an α chance of error, then Eq. (4.9) shows the $1 - \alpha$ confidence interval within which the mean m may be said to lie.

The usual procedure is to take samples from a population of material and calculate the mean and estimated standard deviation from the sample. An α value is selected, and the tabulated t for that value and the degrees of freedom $(n - 1)$, associated with the calculation, is found from tables. The $1 - \alpha$ confidence as to the location of the true mean is calculated from Eq. (4.9). The values of the lower and upper limits are obtained:

$$x_l = \bar{x} - \frac{t_\alpha \cdot s(x)}{\sqrt{n}} \tag{4.10}$$

$$x_u = \bar{x} + \frac{t_\alpha \cdot s(x)}{\sqrt{n}} \tag{4.11}$$

All of these uses of the t function are included in the t programs that are described in Sec. 4.5. The programs take the experimental data and calculate the means and estimated standard deviations, the t values from the proper equations, and the probability level associated with the calculated t. With the proper input, the programs calculate the confidence interval, and also calculate the sample size required to give a specific confidence interval. The calculation of the sample size is discussed in the next section.

4.4 CALCULATION OF SAMPLE SIZE

A question that often arises in experimental work is how many samples to take. The answer depends on what confidence is to be placed on the result. Equation (4.9) illustrated the calculation of confidence limits. If the confidence C is set such that the mean is known to be in the range $\bar{x} \pm C$ with a certain confidence, then the sample size can be estimated by rearranging Eq. (4.9) as follows:

$$C = \frac{t_\alpha \cdot s(x)}{\sqrt{n}} \tag{4.12}$$

$$n = \left(\frac{t_\alpha \cdot s(x)}{C}\right)^2 \tag{4.13}$$

The solution of Eq. (4.13) requires a knowledge of $s(x)$ and a trial-and-error approach inasmuch as t_α depends on the value of n. The calculator program presented next contains a routine for calculating the sample size if C, α, and $s(x)$ are fixed.

The t test problems discussed thus far are of the type illustrated in Fig. 4.1. A sample is observed and the hypothesis that it came from a population with mean m is tested. If the sample mean $\bar{x}$ is greater than some value x_u, or less than some value x_l, t will exceed the critical value and the hypothesis will be rejected. For example, in Fig. 4.1, if the value x_1 actually came from the distribution illustrated, the hypothesis that it did would be rejected because it falls outside the limit x_u set by the t test. We have a false rejection. It is called an error of Type I.

There is the possibility of a second type of error, illustrated by Fig. 4.2. If the value x_2 came from the population on the right of the figure with a mean of m_1, the hypothesis that it came from the population with mean m would be accepted because the value x_2 is between the acceptable limits of x_l and x_u. This is an error of Type II, a false acceptance of the hypothesis, for which the probability is designated β. The magnitude of the probability of this type of error is a function of the difference between m and m_1, the sample size, and $s(x)$.

If the probability of a Type I error is set at α, the critical value of x_u can be calculated by a rearrangement of Eq. (4.3) to give:

$$x_u = \frac{t_\alpha \cdot s(x)}{\sqrt{n}} + m \tag{4.14}$$

The probability β of a Type II error will be the probability associated with t_β, where:

$$t_\beta = \frac{m_1 - x_u}{s(x)/\sqrt{n}} \tag{4.15}$$

If it is desirable to fix the size of the Type II error (that is, if one wishes to take only a β chance of saying the mean is m when it is really m_1), it is necessary to adjust the sample size so that Eq. (4.16) holds.

$$\sqrt{n} = \frac{s(x) \cdot (t_\alpha + t_\beta)}{m_1 - m} \tag{4.16}$$

To be precise, it is not possible to solve Eq. (4.16) for n because $s(x)$, the estimated standard deviation, is not known until the sample has been obtained. However, in practice it is not unusual to have a reasonable estimate of the standard

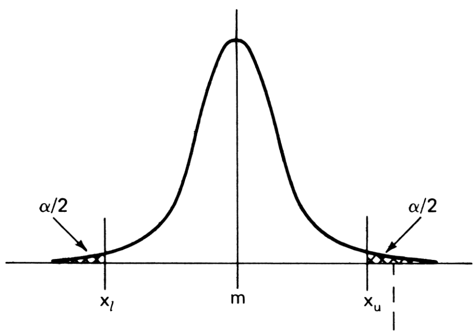


Figure 4.1 Error of Type I.

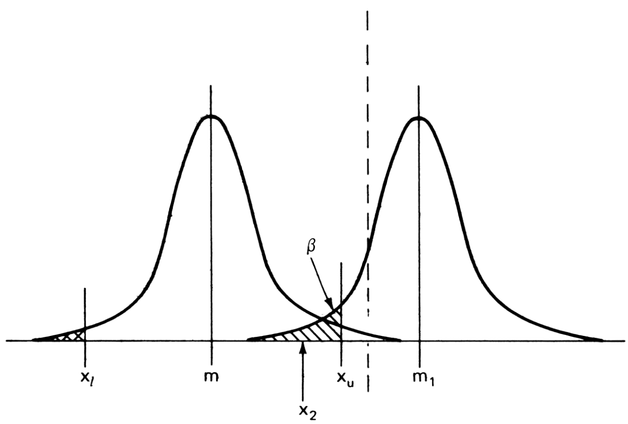


Figure 4.2 Error of Type II.

deviation from previous experience. The required sample size increases with increase in the estimated standard deviation; therefore if $s(x)$ is overestimated, n will be larger than necessary and the probability of Type II error will be decreased.

If a process has a target value of 45.0 and the probability of a Type I error is set at 0.05 (i.e., there will be only a 5 percent chance of rejecting material that truly has a mean of 45.0), and if it is desirable not to say the mean is 45.0 when it is as large as 46.0 or larger, and a probability of this type of error (Type II) is acceptable at 0.10, the sample size required can be calculated if $s(x)$ is known.

The t programs given in the next section solve problems of this type with input of probability values for α and β, the difference between m and m_1, and an estimate of $s(x)$.

4.5 THE *t* PROGRAMS

There are two t programs. One is for problems in which t and the probability of t are calculated. The other is for problems in which a probability is given, t is calculated, and then a variable value or sample size is determined. The two programs could be combined, but for simplicity of presentation they are given separately. They handle the following types of problems:

1. Calculating t values for hypotheses associated with Eq. (4.3) or (4.6). These are problems of comparing one mean with a target value or two means with each other. The program processes the original experimental data and calculates not only the t value but the associated probability.
2. Calculating the probability on input of a t value and degrees of freedom.
3. Calculating the confidence interval of the mean for a given probability level with input of experimental data.
4. Determining the sample size required to fix the probabilities of Type I and Type II errors when the standard deviation is available.
5. Calculating the t value corresponding to a given probability and degrees of freedom.

4.5.1 *t*-1 Program

The first t program, t-1, handles the first two types of problems, and the second t program, t-2, handles the other three. As previously stated, the two programs could be combined if desired.

For testing hypotheses about one mean, Eqs. (4.3) to (4.5), the program t-1 takes the original data: x_1, KEY A, x_2, KEY A, . . . , x_n, KEY A. When all the data are entered, the target mean is entered with KEY C: m, KEY C. This operation starts the calculation. The output is the mean $\bar{x}$, the estimated standard deviation $s(x)$, the t value from Eq. (4.4), the degrees of freedom ν, and the probability value.

If m is greater than $\bar{x}$, the t value will be negative. The probability calculated is the two-sided probability. If one-sided test results are desired, the calculation is the same but the interpretation of the results must be modified as explained in the text.

For testing hypotheses about two means $\bar{x}$ and $\bar{y}$, Eq. (4.6) and (4.8), two sets of data are input to the calculator. The number of observations in each set do not have to be the same. The data are entered: x_1, KEY A, x_2, KEY A, . . . , x_{n_x}, KEY A. When all the first variable data are entered, the second variable is entered with KEY B: y_1, KEY B, y_2, KEY B, . . . , y_{n_y}, KEY B. When all the data are entered, KEY C again starts the calculation. The output is $\bar{x}$, $s(x)$, $\bar{y}$, $s(y)$, t from Eq. (4.8), the degrees of freedom ν, and the probability value.

Again, the probability value is for the two-sided test, and if $\bar{y}$ is larger than $\bar{x}$, t will be negative. If a one-sided test is desired, the results need to be interpreted as indicated in Table 4.1.

4.5.2 *t*-3 Program

The t-1 program may also be used to find the probability associated with an input of t and ν degrees of freedom. This calculation is done with KEY D. With the Hewlett-Packard program, the data are entered: t, ENTER ↑, ν, KEY D. With the Texas Instruments program, the data are entered: t, KEY $x \rightleftharpoons t$, ν, KEY D. The output is the probability value. This portion is run automatically if the data are entered with KEYs A, B, and C. However, the probability portion may be run independently if desired.

The t-3 portion of the program is not listed separately but is included in the t-1 program. It may be used separately, and it may be included with other programs making a t calculation where the probability is wanted. The inputs of t and ν have to be in the proper sequence to obtain the correct answer. There are references to the probability section of the t program later in the text, and it is referred to as t-3, but is actually the latter portion of the t-1 program.

The operation of program t-1 is summarized in Table 4.2. A flow diagram of the program is in Fig. 4.3, and the listing and detailed description are given in the Appendix.

4.5.3 *t*-2 Program

To determine the confidence range for a calculated mean with t-2, the data are entered: x_1, KEY a, x_2, KEY a, . . . , x_n, KEY a, in the Hewlett-Packard program, or with KEY A' in the Texas Instruments program. When all the data have been entered, the desired confidence probability, as a decimal fraction, is entered—in the Hewlett-Packard program with KEY b and in the Texas Instruments program

Table 4.2 t-1 Program Operation

Hypothesis	$\bar{x} = m$	$m_x = m_y$
Data input	x_1, KEY A x_2, KEY A ⋮ x_n, KEY A m, KEY C	x_1, KEY A x_2, KEY A ⋮ x_{n_x}, KEY A y_1, KEY B y_2, KEY B ⋮ y_{n_y}, KEY B KEY C
Output	$\bar{x}$, mean $s(x)$ standard deviation t for $\dfrac{\bar{x} - m}{s(\bar{x})}$ ν degrees of freedom α probability	$\bar{x}$, mean of first set of data $s(x)$ standard deviation $\bar{y}$, mean of second set of data $s(y)$ standard deviation t for $\dfrac{\bar{x} - \bar{y}}{\bar{s}(x)\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}}$ ν degrees of freedom α probability

	t-3 PROGRAM		HP	TI
Data input	t		ENTER ↑	KEY $x \leftrightharpoons t$
	ν	degrees of freedom	KEY D	KEY D
Output	α	two-sided t probability		

with KEY B'. The program output is the mean, the lower level of the confidence range, and the upper level of the range. The t value, the estimated standard deviation, and the input probability value are available in storage registers if they are desired.

To use the t-2 program to obtain the sample size, the input is the estimated standard deviation, $s(x)$, the acceptable risks of Type I and Type II errors, α and β, and the difference between the mean value to be detected at the probability level set, Δ:

Input	HP	TI
$s(x)$	ENTER ↑	KEY C'
α	ENTER ↑	KEY C'
β	ENTER ↑	KEY C'
Δ	KEY c	KEY C'

The output is the sample size n required to satisfy the conditions set.

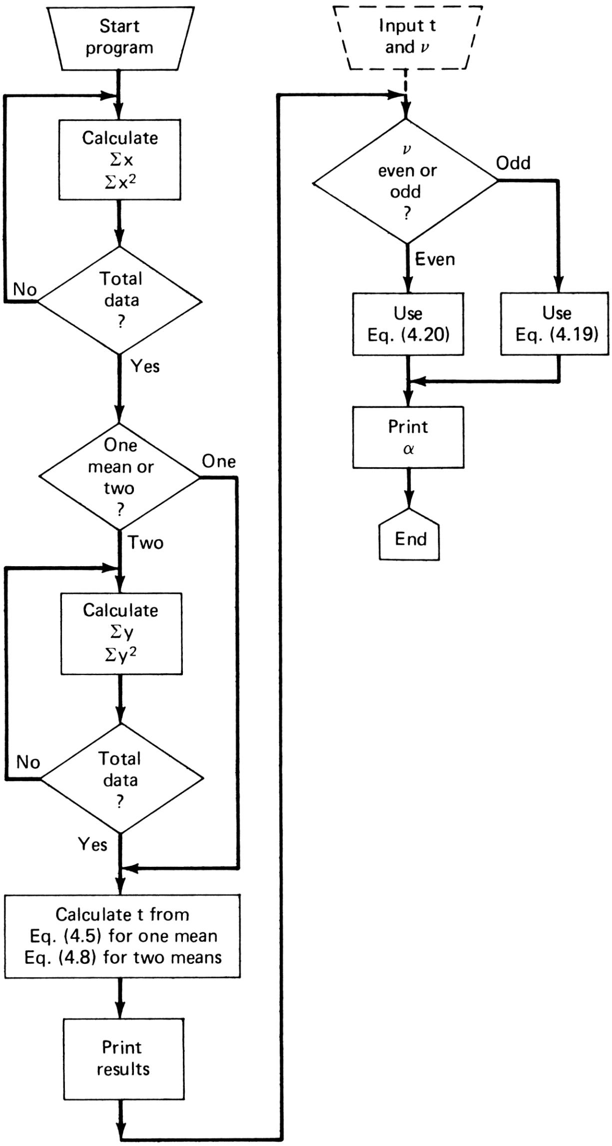


Figure 4.3 Flow diagram for t-1 program: calculation of t and probability.

4.5.4 *t*-4 Program

The two calculations mentioned just above include the calculation of a t value for a particular degrees of freedom and probability. This calculation may be made directly in the t-2 program with KEY d (KEY D') if the input is the probability, expressed as a decimal fraction, and ν, the degrees of freedom. With the Hewlett-

Table 4.3 t-2 Program Operation

	Confidence range		Sample size		t value	
	HP	TI	HP	TI	HP	TI
INPUT						
x_1,	KEY a	KEY A'	$s(x)$, ENTER ↑	KEY C'	α, ENTER ↑	KEY $x \leftrightharpoons t$
x_2,	KEY a	KEY A'	α, ENTER ↑	KEY C'	ν, KEY d	KEY D'
⋮			β, ENTER ↑	KEY C'		
			Δ, KEY c	KEY C'		
x_n,	KEY a	KEY A'				
P,	KEY b	KEY B'				
OUTPUT						
	$\bar{x}$ (mean)		n (sample size)		t	
	x_l (lower limit)					
	x_u (upper limit)					

Packard program: α, ENTER ↑; ν, KEY d. With the Texas Instruments program: α, KEY $x \leftrightharpoons t$; ν, KEY D'. The output is the t value. This calculation is equivalent to looking up t in a t table. The t corresponds to the probability as a two-sided value.

The t value is calculated if the program is started with KEYs a, b, or c (A', B', or C'). However, the program started with KEY d (D') can be used separately and is referred to in the text as the t-4 program although it is actually presented as a portion of the t-2 program.

Table 4.3 summarizes the operation of the t-2 program. A flow diagram is shown in Fig. 4.4, and the listing and detailed description are given in the Appendix.

At the termination of the programs, the following values are in storage areas in the calculator.

Storage area		Program		
HP	TI	Confidence range	Sample size	t value
A	12	t	Δ	α
B	13	$\bar{x}$	2β	—
C	14	$s(\bar{x})$	α	—
D	10	P	$s(x)$	—
I	09	ν	ν	ν

4.5.5 Discussion of the t Programs

The calculation of the mean, the standard deviation, and the number of data values required to determine t from either Eq. (4.5) or Eq. (4.8) is done by means

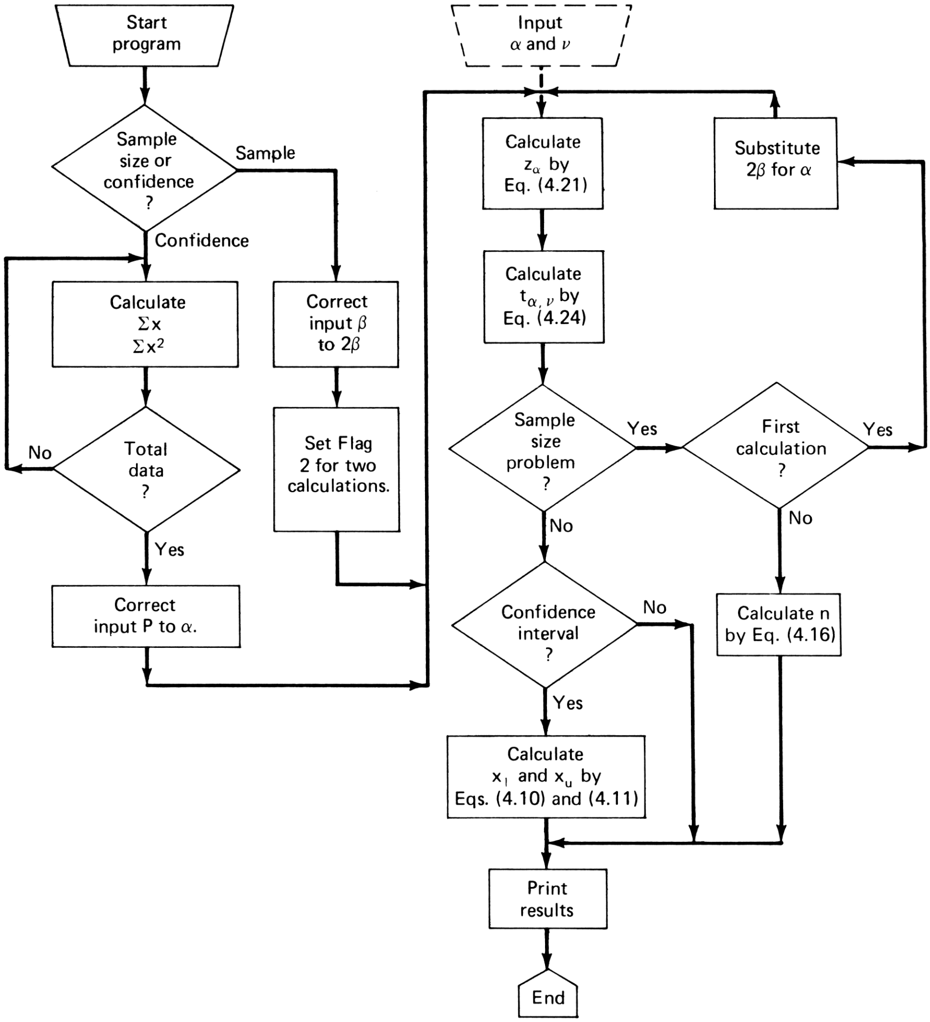


Figure 4.4 Flow diagram for *t*-2 program: given the probability, calculate *t*.

of the built-in function KEY Σ+. The pooled standard deviation estimate $\bar{s}(x)$ for Eq. (4.8) is calculated from Eq. (4.7), using the built-in routine for calculating $s(x)$, squaring this value, and multiplying by one less than the number of data entries, also accumulated with KEY Σ+. The only new technique in the programs consists of calculating the probability value for t and calculating t from the probability value.

The probability value for t is calculated using equations from Ref. 2. The equations used, numbers 26.7.3 and 26.7.4 of the reference are:

ν is odd:

$$A \left| \begin{array}{ll} = \dfrac{2\theta}{\pi} & \nu = 1 \\ = \left(\dfrac{2}{\pi}\right)\left(\theta + \sin\theta(\cos\theta + \dfrac{2}{3}\cos^3\theta + \cdots \quad + \dfrac{2 \cdot 4 \cdots (\nu-3)}{3 \cdot 5 \cdots (\nu-2)}\cos^{\nu-2}\theta\right) & \end{array} \right. \tag{4.17}$$

ν is even:

$$A = \sin\,(1 + \frac{1}{2}\cos^2\theta + \frac{1 \cdot 3}{2 \cdot 4}\cos^4\theta + \cdots \quad + \frac{1 \cdot 3 \cdot 5 \cdots (\nu-3)}{2 \cdot 4 \cdot 6 \cdots (\nu-2)}\cos^{\nu-2}\theta) \tag{4.18}$$

where $\theta = \arctan \dfrac{|t|}{\sqrt{\nu}}$

ν = degrees of freedom

A = area under the t distribution from $-t$ to $+t$

$\alpha = 1 - A$ (the probability used in the present text)

For the calculator program these equations are rearranged to simplify the programming. The arrangement in the t-1 program is as follows:

ν is odd:

$$A = \left(\frac{2}{\pi}\right)\left(\theta + \sin\theta\cos\theta\left(1 + \frac{2}{3}\cos^2\theta\left(1 + \frac{4}{5}\cos^2\theta\left(1 + \cdots\left(1 + \frac{\nu-3}{\nu-2}\cos^2\theta\right)\cdots\right)\right)\right)\right) \tag{4.19}$$

ν is even:

$$A = \sin\theta\left(1 + \frac{1}{2}\cos^2\theta\left(1 + \frac{3}{4}\cos^2\left(1 + \frac{5}{6}\cos^2\theta\left(1 + \cdots\left(1 + \frac{\nu-3}{\nu-2}\cos^2\theta\right)\cdots\right)\right)\right)\right) \tag{4.20}$$

The calculation of the t value from an input probability, in the t-2 program, involves calculating the normal distribution value for the same probability and then calculating t from the normal deviate z. Again, the equations used are from Ref. 2. The equation number for the normal deviate is 26.2.22 in the reference and that for the t value is 26.7.5. These equations are given below.

For the normal deviate z:

$$z = u - \frac{a_0 + a_1 u}{1 + a_2 u + a_3 u^2} \tag{4.21}$$

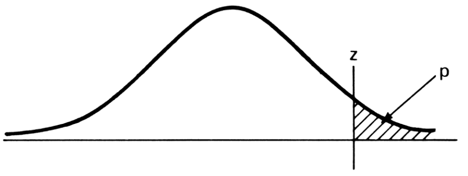


Figure 4.5 Representation of p in calculation of Eq. (4.21).

where $u = \sqrt{\ln(1/p^2)}$
$a_0 = 2.30753$
$a_1 = 0.27061$
$a_2 = 0.99229$
$a_3 = 0.04481$
$p =$ probability equal to $\alpha/2$, where α is the two-sided probability of this text (see Fig. 4.5)

The relation of t to the z value is given by the following equation with the degrees of freedom ν as the additional parameter:

$$t = z + \frac{z^3 + z}{4\nu} + \frac{5z^5 + 16z^3 + 3z}{96\nu^2} + \frac{3z^7 + 19z^5 + 17z^3 - 15z}{384\nu^3} + g \qquad (4.22)$$

The term g at the end of Eq. (4.22) is added to improve the accuracy of the calculation. With degrees of freedom greater than two the loss of accuracy with the g term omitted is in the second decimal place of the calculated t value or smaller. The g term is not included in the t-2 program, but the following equation for g could readily be added if the additional accuracy was desired.

$$g = \frac{79z^9 + 776z^7 + 1482z^5 - 1920z^3 - 945z}{92160\nu^4} \qquad (4.23)$$

The t-2 program uses a rearrangement of Eq. (4.22) to simplify the programming. The equation in t-2 is the following:

$$t = (384\nu^3 z + 96\nu^2(z^3 + z) + 4\nu(z(z^2(5z^2 + 16) + 3)) + z(z^2(z^2(3z^2 + 19) + 17) - 15))/(384\nu^3) \qquad (4.24)$$

4.6 EXAMPLES OF THE t PROGRAMS

4.6.1 A process is assumed to produce a gasoline with an octane number of 87.5. Samples are taken every hour over an 11-h period and the hypothesis that the mean is 87.5 is to be rejected with a 5 percent chance of an error based on the 11 samples.

Hypothesis—H_0: $m = 87.5$
$\alpha = 0.05$

Data: 86.6, 87.1, 86.4, 86.8, 87.2, 87.3, 86.1, 88.4, 86.4, 87.2, 88.6

With the t-1 program in the calculator, the data are entered: 86.6, KEY A, 87.1, KEY A, . . . , 88.6, KEY A. With all the data entered, the target mean is entered: 87.5, KEY C.

The calculator output is:

87.1	Mean
0.7950	Estimated standard deviation
−1.669	t value [Negative value indicates $m > \bar{x}$ (Eq. 4.2)]
10	Degrees of freedom ν
0.126	Probability associated with t. Since this value is larger than 0.05, hypothesis is not rejected. (Time:17/22)

4.6.2 Two processes are to be compared to determine whether one is better than the other. In this case, a lower value is better; i.e., time of filtration. The hypothesis, H_o: $m_1 = m_2$, is that the means of the two processes are the same. An α value of 0.05 is set for a false rejection of the hypothesis. The data are as follows:

Process 1	Process 2
8	9
10	10
12	10
13	4
13	7
9	9
14	

Using the t-1 program the data are entered, first one set and then the other: 8, KEY A, 10, KEY A, 12, KEY A, . . . 14, KEY A (for the first set of data); 9, KEY B, 10, KEY B, 10, KEY B, . . . 9, KEY B (for the second set of data); KEY C (runs the program).

The calculator output is:

11.28	Mean of first set of data
2.289	Estimated standard deviation of the first set
8.17	Mean of second set
2.317	Estimated standard deviation of the second set
2.436	t value (Eq. 4.6)
11	Degrees of freedom ν
0.033	Probability associated with t. Since this is less than the preestablished 0.05 value, the hypothesis is rejected. (Time:24/30)

4.6.3 Find the probability value associated with $t = 2.47$ and 14 degrees of freedom. With the t-1 program:

Calculator	Input	Output
HP	2.47, ENTER ↑, 14, KEY D.	0.027
TI	2.47, KEY $x \rightleftharpoons t$, 14, KEY D.	0.027

This is equivalent to looking the value up in a t table for which an interpolation would have to be made. (Time:17/25)

4.6.4 If the question to be answered is whether the mean of a set of data is larger (or smaller) than some target value and not just equal to the value, the procedure is the same as in Example 4.6.1, but the interpretation of the result is different. Table 4.1 illustrates the different tests for one and two sets of data.

The current filtration rate is 350 lb/(h)(ft^2), and a new process is to be tested to determine if it has a greater rate. Some data are taken with the new process and the hypothesis that the mean is equal to or less than 350 is tested at an α value of 0.05. The hypothesis will be rejected only if the calculated t is positive with a probability $= 0.10$.

Data: 409, 411, 388, 394, 369, 354 lb/(h)(ft^2)

$$H_0\colon\ m = 350 \qquad \alpha = 0.05$$

With the t-1 program, the data are entered: 409, KEY A, 411, KEY A, . . . , 354, KEY A, 350, KEY C.

The output is:

388	Mean
22.5	Estimated standard deviation
4.10	t
5	Degrees of freedom
0.01	Probability level (Time:12/17)

The hypothesis that the new process has a mean equal to or less than 350 is rejected; the t value is positive with a probability value of 0.01.

4.6.5 If two means are to be compared, not to see if they are equal but to determine whether a particular one is larger (or smaller) than the other, the procedure is similar to that in Example 4.6.2, but the interpretation of the result is different. Table 4.1 shows the different analyses for tests of two means.

A process usually gives a product with a mean impurity of 0.006. A new method is proposed that is claimed to yield a product with a lower impurity. The two are compared, with the results shown below, and a test is made as to whether the new method yields a product of lower impurity than the old. The hypothesis to be tested is that $\text{Mean}_{\text{new}} \geqq \text{Mean}_{\text{old}}$, and the α level of a false rejection (i.e., a false acceptance of the new method as being better than the old) is set at 0.05.

New method impurity data	Old method impurity data
0.0060	0.0063
0.0051	0.0050
0.0064	0.0062
0.0053	0.0060
0.0051	0.0070
0.0063	0.0070
0.0035	0.0033

$$H_0\colon\ m_{\text{new}} \geqq m_{\text{old}}$$
$$\alpha\colon\ 0.05$$

Calculator run using t-1:

0.0060, KEY A, 0.0051, KEY A, . . . , 0.0035, KEY A
0.0063, KEY B, 0.0050, KEY B, . . . , 0.0033, KEY B
KEY C

The output is:

0.0054	Mean, new	
0.0010	Estimated standard deviation	
0.0058	Mean, old	
0.0013	Estimated standard deviation	
−0.713	t	
12	Degrees of freedom	
0.489	Probability level	(Time:25/31)

The hypothesis is not rejected. The difference between the new mean and the old is not significant enough to reject the hypothesis that the methods are either equal or the new really yields a poorer product than the old.

4.6.6 The data of Example 4.6.1 will be used to place a 95 percent confidence range on the mean.

With the t-2 program in the calculator, the data are entered: 86.6 KEY a, 87.1, KEY a, . . . , 88.6, KEY a with the Hewlett-Packard program or using KEY A' in the Texas Instruments calculator. When all the data are entered, the probability level is entered: 0.95, KEY b (or KEY B' in the TI).

The calculator output is:

87.10	Mean	
86.57	Lower range	
87.63	Upper range	(Time:14/13)

4.6.7 Find the t value corresponding to an α of 0.025 and 16 degrees of freedom. With the t-2 program:

	HP	TI			
0.025	ENTER ↑	KEY $x \rightleftharpoons t$			
16	KEY d	KEY D'	Output:	2.474	(Time:10/11)

4.6.8 How many samples would be required to establish a mean was equal to 0.0050, with an α level of 0.05, and have only an 0.01 chance that it was really as high as 0.0058? Assume the standard deviation is 0.0012. Using the t-2 program:

Data	HP	TI	
0.0012	ENTER ↑	KEY C'	Estimated standard deviation
0.05	ENTER ↑	KEY C'	α value
0.01	ENTER ↑	KEY C'	β value
0.0008	KEY c	KEY C'	Minimum difference to be detected

Output: 45 (Number of samples required) (Time:60/75)

Note: this is not the solution to a problem to determine the number of samples in each of two groups to establish a difference between them of 0.008. The calculator program does not address that problem. However, the number of samples in each group is usually slightly less than twice the number given for the solution to the problem of one mean.

4.6.9 How many samples are required to establish a 95 percent confidence range of ±0.6 if the standard deviation is 0.8? The 95 percent confidence range corresponds to an α value of 0.05. Since there is no minimum value to be detected, a β value of 0.50 is used. With the t-2 program:

Data	HP	TI	
0.8	ENTER	KEY C'	Estimated standard deviation
0.05	ENTER	KEY C'	α value
0.50	ENTER	KEY C'	β value
0.6	KEY c	KEY C'	Confidence range

Output: 10 (Number of samples) (Time:60/75)

5

CHI-SQUARE (χ^2) TEST

5.1 INTRODUCTION

The χ^2 distribution is a mathematically derived function that has an important role in the theory of statistics. Its principal uses are in "goodness-of-fit" tests and in making probability evaluations of variances. This chapter deals exclusively with the use of the goodness-of-fit test. Some reference to the application of χ^2 to variances is made in the next chapter.

The goodness-of-fit test applies to counted data or frequency of observations. The observed frequencies, or the number of events of a particular type, are compared with the expected number and the deviations measured in terms of χ^2. χ^2 is defined as the sum of the squares of the deviations of the observed frequencies from the expected frequencies divided by the expected values.

$$\chi^2 = \sum \frac{(O - E)^2}{E} \tag{5.1}$$

The χ^2 test is made in relation to a null hypothesis and the mathematical χ^2 distribution in a similar manner to the t test discussed in the preceding chapter. The null hypothesis is that there is no difference between the observed and expected frequencies. If the χ^2 calculated from Eq. (5.1) is larger than a χ^2 value associated with a low probability, the hypothesis can be rejected with that probability of error.

The mathematical χ^2 distribution is dependent on the number of degrees of freedom, and this is the number of independent deviations available in the calculation of Eq. (5.1).

The usual procedure is to calculate χ^2 from some data and to compare the result with tabulated χ^2 values at the number of degrees of freedom involved. If the calculated χ^2 is larger than the value tabulated for 0.05 or 0.01 probability, the hypothesis of agreement of observed with expected values is rejected with the associated probability of a false rejection.

The procedure with the calculator programs is simply to enter the data in one of the χ^2 programs. The calculator produces the number of degrees of freedom, the χ^2 value, and the probability of equal or larger χ^2. Several different χ^2 tests are given in the following sections and programs for their use are presented.

5.2 RATIO TEST

Data that are expected to conform to some ratio may be tested by the χ^2 function. A person breeding long-haired animals expects to obtain a ratio of 3/1 long-haired to short-haired young in a litter. Of 85 animals the count is 59 long-haired and 26 short-haired young. Does this result conform to the breeder's expectation? It is claimed that 8 out of 10 people prefer the product of a particular manufacturer. Out of a sample of 125, 95 prefer this product and 30 do not. Does this sample conform with the claim?

If the observed ratio is (a/b), and the expected ratio is (p/q), the χ^2 Eq. (5.1) may be written:

$$\chi^2 = \frac{(a-(p/(p+q))(a+b))^2}{(p/(p+q))(a+b)} + \frac{(b-(q/(p+q))(a+b))^2}{(q/(p+q))(a+b)} \tag{5.2}$$

Equation (5.2) may be simplified to:

$$\chi^2 = \frac{(a-(p/q)b)^2}{(p/q)(a+b)} \tag{5.3}$$

From the long-haired animal data above, Eq. (5.3) gives:

$$\chi^2 = \frac{(59-(3/1)(26)^2}{(3/1)(85)} = \frac{(19)^2}{255} = 1.416$$

However, there is an adjustment to be made in this particular solution. There is only one independent difference involved in this χ^2 calculation. For a fixed sample size, either a or b may be independently determined. The other term is then established by difference from the total. Hence, there is only one degree of freedom for this χ^2 calculation. When there is only one degree of freedom, the χ^2 calculated from Eq. (5.1) is biased high, and it is customary to make a correction for each of the differences involved of 0.5 units. For one degree of freedom, Eq. (5.1) for χ^2 becomes:

$$\chi^2 = \sum \frac{(|O-E|-0.5)^2}{E} \tag{5.4}$$

All of the programs presented later in this chapter include the correction when there is one degree of freedom.

The correction to the ratio test changes Eq. (5.3) to the following:

$$\chi^2 = \frac{(|a-(p/q)b|-((p/q)+1)/2)^2}{(p/q)(a+b)} \tag{5.5}$$

With this correction the adjusted χ^2 calculation for the long-haired animal data is:

$$\chi^2 = \frac{(|59 - (3/1)(26)| - ((3/1) + 1)/2)^2}{(3/1)(85)} = \frac{(19 - 2)^2}{255} = 1.133$$

At one degree of freedom a χ^2 value of 1.133 corresponds to a probability level of 0.29; that is, there is a 29 percent chance of error if the hypothesis of a 3/1 ratio is rejected on the basis of these data.

With the χ^2-1 program, the observed ratio values are entered, the expected ratio is entered, and the calculator gives the number of degrees of freedom, the χ^2 value (adjusted for one degree of freedom), and the probability value.

5.3 2 × 2 TABLE

The 2 × 2 table χ^2 test is similar to the ratio test except that the expected ratio is determined from the data. This may best be illustrated by an example.

Table 5.1 shows the number of successful and unsuccessful pilot unit runs made by the day and night shifts over the period of a month. Are the data consistent?

In this case the expected ratio of successful to unsuccessful runs would be taken from the data: 48/14. The comparisons with the day and night shifts would be made on the total runs for each shift. The expected number of successful runs for the day shift would be 32(48/62). Here, again, there is only one degree of freedom. Since the totals of the rows and columns are fixed, only one value in the body of the data may be independently changed. All the other values would be fixed once that one value was determined. The equation for χ^2 would be Eq. (5.4), and the calculation would be as follows:

$$\begin{aligned}\chi^2 &= \frac{(|26 - (32)(48/62)| - 0.5)^2}{(32)(48/62)} + \frac{(|6 - (32)(14/62)| - 0.5)^2}{(32)(14/62)} \\ &\quad + \frac{(|22 - (30)(48/62)| - 0.5)^2}{(30)(48/62)} + \frac{(|8 - (30)(14/62)| - 0.5)^2}{(30)(14/62)} \\ &= 0.195\end{aligned}$$

The probability associated with a χ^2 of 0.195 and one degree of freedom is 0.659. The hypothesis of consistency would be accepted. If it were rejected, there would be a 66 percent chance of error based on these data.

Table 5.1 Successful and Unsuccessful Pilot Unit Runs

Runs	Day shift	Night shift	Total
Successful	26	22	48
Unsuccessful	6	8	14
Total	32	30	62

Where the expected values are contingent on the tabulated data, the χ^2 test is called a "contingency test." Contingency tables of various sizes, including the 2×2, are discussed in more detail in Sec. 5.6. One of the χ^2 programs is capable of calculating χ^2 for $r \times c$ contingency tables up to $6 \times c$ in the Hewlett-Packard system or $16 \times c$ in the Texas Instruments system, with no limit to the value of c.

5.4 EQUAL EXPECTATION: $r \times 1$ TABLE

A special case exists when the expected value is the mean of the observations. Four different types of computer terminals are tested over a 6-mo period, and with essentially the same usage they have different numbers of breakdowns: 44, 36, 32, 46. Is there a significant difference among them? The χ^2 test would be against the mean number of breakdowns: 39.5. Using Eq. (5.1), the calculated χ^2 is 3.316, with a probability value of 0.345, that is, a 34 percent chance of error if the assumption of equality is rejected.

When the assumption of equal expectation is made, Eq. (5.1) takes the form:

$$\chi^2 = \sum \frac{(x - \bar{x})^2}{\bar{x}} \tag{5.6}$$

The χ^2 calculation from Eq. (5.6) is very readily done in a calculator inasmuch as the numerator is equal to $\Sigma x^2 - (\Sigma x)^2/n$, and all of these values are obtained with the calculator KEY $\Sigma+$.

For χ^2 calculated for equal expectation, the number of degrees of freedom is $n - 1$. Of the values, $n - 1$ may be independently set and the last one is fixed by the total.

χ^2-2 program calculates the mean, the degrees of freedom, the χ^2 value and the probability for problems of equal expectation. The input is the observed data.

5.5 UNEQUAL EXPECTATION TABLES

It is often the case that the mean expectation calculated from the observed results is not the simple arithmetic mean but is a weighted mean, weighted proportionally to the amount of use of the different items. If the computer terminals mentioned in the previous section were used different amounts of time during the 6-mo test period, say 400, 600, 650, and 550 h, it would not be expected that the number of breakdowns would be equal. The total operating time is 2200 h, the sum of the four values just given. The total number of breakdowns was 158. The first unit ran 400 h; therefore, the expected number of breakdowns for the first unit would be $(400)(158/2200) = 28.72$.

If x_i is the observed frequency, and f_i is the weighting factor for the ith unit, χ^2 for unequal expectation would be calculated:

$$\chi^2 = \sum \frac{(x_i - f_i(\Sigma x/\Sigma f))^2}{f_i(\Sigma x/\Sigma f)} \tag{5.7}$$

For the computer terminal data, χ^2 would be calculated from Eq. (5.7) as follows:

$$\chi^2 = \frac{(44 - 400(158/2200))^2}{400(158/2200)} + \frac{(36 - 600(158/2200))^2}{600(158/2200)} + \cdots$$
$$= 14.97$$

With three degrees of freedom this χ^2 has a probability of 0.002, allowing for discarding the hypothesis of equal performance with an 0.002 chance of error.

The χ^2-3 program makes the calculation for unequal expectation with the entry of the weighting factor and the data for up to 11 items of comparison in the Hewlett-Packard program, and for up to 24 items in the Texas Instruments program. In cases of unequal expectation it is often of interest to see which items contribute most to the calculation of χ^2. The calculator program gives the χ^2 value for each observation as well as the total χ^2.

5.6 CONTINGENCY TABLES: $r \times c$ TABLES

When data may be classified according to two sets of categories, arranged in rows and columns, and the expected number in each combination of classifications is contingent on the row and column total, the data for χ^2 analysis are said to be in a "contingency table." For example: the results of a poll of voters classified as Republicans, Democrats, and Independents and counted as For, Against, and Undecided on an issue. The results would form a 3×3 contingency table.

The expected number in the category designated by row (i) and column (j) is the total of row (i) times the total of column (j) divided by the total number of observations. χ^2 can be calculated by Eq. (5.1) and the number of degrees of freedom will be the number of rows less 1 times the number of columns less 1. Table 5.2 illustrates a 5×4 table of some data on compressor failures.

The expected number of valve failures for compressor no. 3 would be (57)(49)/191 = 14.62. The χ^2 contribution for this item would be $(14.62 - 12)^2/14.62 = 0.4695$.

To use Eq. (5.1), each expected value would be calculated and the difference from the observed would be squared and divided by the expected value. The sum of all these results would be the χ^2 for the data and the number of degrees of freedom would be $(5 - 1)(4 - 1) = 12$.

Table 5.2 Comparison of Failures of Four Compressors

	Compressor no.				
Types of failures	**1**	**2**	**3**	**4**	**Totals**
Ring	8	10	11	8	37
Valve	14	17	12	14	57
Electric	5	7	13	6	31
Lubrication	13	12	10	11	46
Other	4	7	3	6	20
Totals	44	53	49	45	191

An alternative method of calculation, and the one that is used in the calculator program, is that of Eq. (5.8), which is equivalent to Eq. (5.1).

$$\chi^2 = T\left(\sum \frac{x_{ij}^2}{C_j R_i} - 1\right) \tag{5.8}$$

where T = total of all the data
C_j = total of the jth column
R_i = total of the ith row
x_{ij} = the observation in the ith row and the jth column

The use of Eq. (5.1) requires storing most of the data until the row and column totals can be obtained. With Eq. (5.8), if the data are calculated by columns, $\Sigma(x_{ij}^2/C_j)$ for column j may be calculated as soon as all of column j data are available, and the amount of calculator storage required is reduced. Equation (5.1) permits the calculation of the contribution to χ^2 of each item in the table of data, and this information is not available with Eq. (5.8).

The χ^2-4 program uses Eq. (5.8) so that it can be used with larger amounts of data with the limited storage capacity of the calculators. The limitation of the program as presented is six rows of data with the Hewlett-Packard program and 16 rows with the Texas Instruments program. There is no restriction on the number of columns. Row and column designation is arbitrary and may be interchanged.

Reference to Eq. (5.1) will show that if the expected value E is zero, or very small, the calculated χ^2 will be large. It is customary not to use expected values less than 5. With a large number of degrees of freedom, this rule is relaxed, and expected values of 2 or 3 are permitted. Zero, of course, leads to an impossible answer. In a contingency table, if a row or column of data consists of very small numbers, these data may be eliminated or combined with another row before calculating χ^2.

For example, if the "Electric" failures in Table 5.2 were all either 1 or 0, these values could be combined with "Other" to give a 4 × 4 table.

5.7 GOODNESS-OF-FIT TEST

χ^2 can be used to determine whether a distribution of data fit some mathematical model. It could be used to check the data in Table 3.8 for agreement with the Poisson distribution. The test may be used to check whether data are normally distributed. In games of chance, where the probability of an event is supposed to be constant, the distribution of events can be checked against the binomial distribution.

χ^2 for the goodness-of-fit test is calculated in the same manner as the other tests, using Eq. (5.1): the sum of the squares of deviation between the expected values and the observed values divided by the expected values. The number of degrees of freedom requires a little explanation.

For a goodness-of-fit test, the data are assembled into a number of groups,

Table 5.3 Number of Degrees of Freedom in Goodness-of-Fit χ^2 Tests

Conditions	Degrees of freedom*
Normal Distribution	
Mean and standard deviation from the data	$n-3$
Mean and standard deviation from outside source	$n-1$
Binomial Distribution	
Constant probability from the data	$n-2$
Constant probability from outside source	$n-1$
Poisson Distribution	
Expected value from the data	$n-2$
Expected value from outside source	$n-1$

*n is the number of categories of data tested by χ^2.

and the expected number in each group is calculated from the distribution being tested. The number of degrees of freedom will be the number of groups less the number of restraints placed on the calculation. If the data are compared to a normal distribution, and the mean and standard deviation are calculated from the data, three restraints are placed on the calculation of the expected values. Three degrees of freedom are lost: one for the total, one for the mean, and one for the standard deviation—all calculated from the data. If the data are compared with a normal distribution of preassigned mean and standard deviation, only one degree of freedom is lost.

Similar reasoning applies to comparing data with other distributions. If a parameter of the expected distribution calculation is based on the data, it places a restraint on the calculation. If the parameter is assumed, no restraint is placed on the calculation. One degree of freedom is always used in having the total of the expected values match the total of the data. Table 5.3 shows the number of degrees of freedom for most situations arising in comparisons of data with the three distributions of Chap. 3.

The calculation of a χ^2 goodness-of-fit test is probably best explained by an example. The data in Table 5.4 are from the distribution of income tax returns shown in Fig. 3.1, but the number of observations is given in single units rather than in units of 10,000 to simplify the arithmetic. χ^2 is directly proportional to the units of the counted data so if the units are changed, as in this example, the χ^2 calculated for the adjusted data would have to be readjusted by the same factor to apply to the original data.

5.7.1 Example of Goodness-of-Fit Calculation

The data are shown in Table 5.4. Question: are these data normally distributed?

With the statistical parameter program, the mean and standard deviation for weighted data are calculated:

Table 5.4 Data for Goodness-of-Fit Calculation

Range of deductions	Number of returns	Midrange
0–199	3	99.5
200–399	6	299.5
300–599	9	499.5
600–799	14	699.5
800–999	21	899.5
1000–1199	27	1099.5
1200–1399	32	1299.5
1400–1599	48	1499.5
1600–1799	63	1699.5
1800–1999	52	1899.5
2000–2199	44	2099.5
2200–2399	34	2299.5
2400–2599	27	2499.5
2600–2799	16	2699.5
2800–2999	28	2899.5

HP	TI
3, ENTER ↑, 99.5, KEY B	3, KEY $x \rightleftharpoons t$, 99.5, KEY B
6, ENTER ↑, 299.5, KEY B	6, KEY $x \rightleftharpoons t$, 299.5, KEY B
⋮	⋮
28, ENTER ↑, 2899.5, KEY B	28, KEY $x \rightleftharpoons t$, 2899.5, KEY B
KEY C	KEY C

Answer: Mean 1761, standard deviation, 633 with 15 groups and 424 observations. (Time:23/32)

If the midranges had been rounded to even hundreds the results would have been 1762 and 633; not significantly different.

To determine whether the distribution is normal, the expected number in each range is calculated with the normal program. The mean and the standard deviation are entered, and the lower and upper values of each group are entered. KEY A gives the probability of a value falling in that range. The probability is then multiplied by the total number of observations, 424, to get the expected number for that group if the total distribution was normal. In selecting the boundaries of the ranges, in order to provide for the continuity of the normal function, the upper end of one group must be the lower end of the next group, so the first group would range from 0 to 199.5, and the second group from 199.5 to 399.5.

One other adjustment needs to be made. The normal probability is figured from $-\infty$ to $+\infty$. To ensure that the probabilities will add to 1.0, the lower boundary of the first group should be four or five standard deviations below the mean, and the upper boundary of the last group should be four or five standard deviations above the mean. For the example, the first lower boundary is taken as −1000, and the last upper boundary is taken as +5000.

The normal program can be made to carry out the entire calculation with a slight adjustment. The mean and standard deviation are the same for all of the calculations, and the program stores these values in Registers A and B in the Hewlett-Packard program and in Registers 01 and 02 in the Texas Instruments program. The upper boundary of each group calculation is the lower boundary for the next group, and the expected number in the group is the calculated probability times a constant: the number of observations. These factors can be temporarily included into the program, and the program made to print the probability and expected number for each group with the entry of the upper boundary. The changes are as follows:

	HP	TI
1761	STORE A	STORE 01
633	STORE B	STORE 02
424	STORE 5	STORE 08
−1000	STORE 4	STORE 05
	Between lines 207 and 208 add:	Between lines 401 and 402 add:
	RCL 5	×
	×	RCL 08
	PRINT	=
	RCL D	RCL 04
	STO 4	STO 05

These changes multiply the probability by the number of observations and then replace the lower boundary with the upper boundary for the next calculation.

HP	TI
Between lines 3 and 4 add:	Between lines 4 and 5 add:
RCL A	GO TO CE
GO TO B	

These changes bypass the input data storing routines and start the probability calculation directly.

With these changes, the probabilities and expected values are calculated after the entry of each upper boundary value with KEY A. With 5000 used for the last upper boundary, the results are as shown in Table 5.5.

The purpose of all of this was to compare the observed values with values calculated for a normal distribution applied to the same data. And as a secondary benefit, to see various uses of the programs and the manner of modifying them. For the χ^2 calculation the comparison is made between the last column of Table

Table 5.5 Actual and Expected Number of Observations for Data of Table 5.4, Based on Normal Distribution

		Expectation from normal	
Range	Observed number	Probability	Number
(−1000) 0–199	3	0.007	2.9
200–399	6	0.009	3.8
400–599	9	0.018	7.4
600–799	14	0.031	13.2
800–999	21	0.050	21.2
1000–1199	27	0.073	30.9
1200–1399	32	0.096	40.9
1400–1599	48	0.115	48.9
1600–1799	63	0.125	53.0
1800–1999	52	0.123	52.0
2000–2199	44	0.109	46.2
2200–2399	34	0.088	37.2
2400–2599	27	0.064	27.1
2600–2799	16	0.042	17.9
2800–2999 (5000)	28	0.050	21.4
	424	1.000	424.0

5.5, the expected values, and the second column, the observed values. A very simple program will calculate χ^2 with an entry of the observed and expected values, following Eq. (5.1):

HP	TI	
ENTER ↑	KEY $x \rightleftharpoons t$	Observed value
KEY A	KEY A	Expected value
LABEL A	LABEL A	
STO 4	STO 04	
−	(	
x^2	$x \rightleftharpoons t$	
RCL 4	−	
÷	RCL 04	
STO +1	)	
1	x^2	
STO +2	÷	
R/S	RCL 04	
	=	
	SUM 01	
	1	
	SUM 02	
	R/S	

With this program the χ^2 for the data of Table 5.5 is 8.62 (stored in Register 1, or 01), with 15 data entries (stored in Register 2 or 02), giving 12 degrees of freedom for the calculation.

The probability for the χ^2 can be obtained from the χ^2-5 program. The short program listed above could be added to χ^2-5, and the probability would be obtained directly after all the data were entered.

While this description of the goodness-of-fit test may seem quite involved, requiring the use of three different calculator programs, the total time, including the modifications to the programs and the data input was less than 20 min from the original data in Table 5.4 to the calculation of χ^2 and the probability value, which is 0.735.

5.8 THE χ^2 CALCULATOR PROGRAMS

The χ^2 programs are presented as five separate programs. They are summarized in Table 5.6.

Although the programs are shown as five separate items, they may be combined (merged) in any combination and still be operable without any changes. A discussion of the individual programs follows.

5.8.1 Ratio Test: χ^2-1 Program

The χ^2-1 program calculates χ^2 from a comparison of an observed ratio, a/b, to an expected ratio, p/q. The input is: a, ENTER ↑, b, ENTER ↑, p, ENTER ↑, q, KEY A with the Hewlett-Packard program or a, KEY A, b, KEY A, p, KEY A, q, KEY A with the Texas Instruments program. The output is the number of degrees of freedom, which is one and the χ^2 value; and, if the χ^2-5 program is merged with χ^2-1, the probability associated with χ^2 is also calculated.

The χ^2-1 program follows Eq. (5.5) and includes the correction for a χ^2 with one degree of freedom. It is essential that the observed values, a and b, be entered before the expected values, p and q. The answer will be incorrect if the order is reversed. It is also important to emphasize that the χ^2 test is only applicable to ratios of numbers of observations. It is not applicable to ratios obtained from measured data: ratio of iron to nickel in the steel, for example.

After the program has been run, the first value entered will be in Register A, the second in B and the ratio p/q in C in the Hewlett-Packard calculator; the corresponding values will be in Registers 01, 02, and 03 in the Texas Instruments calculator. This permits a check on the input data if there is question about the result of the calculation. If the observed ratio is entered as b/a and the expected ratio as q/p, the answer will be the same as it is when the data are entered in the prescribed order.

5.8.2 Equal Expectancy Test: χ^2-2 Program

The χ^2-2 program calculates χ^2 following Eq. (5.6) but uses an equivalent form that is more adaptable to a calculator solution. The following equations show the equivalent formulations:

Table 5.6 χ^2 **Programs**

1 **Ratio test**	2 **Equal expectancy** $r \times 1$ **table**	3 **Unequal expectancy**	4 **Contingency tables** $r \times c$ **tables**	5 χ^2 **probability**
H_0: $a/b = p/q$	x_1 x_2 $\vdots$ x_r H_0: $x_i = x$	$f_1 \quad x_1$ $f_2 \quad x_2$ $\vdots \quad \vdots$ $f_r \quad x_r$ H_0: $\dfrac{x_i}{f_i} = \dfrac{\Sigma x_i}{\Sigma f_i}$ ($r \leqq 11$) in HP ($r \leqq 24$) in TI	$x_{11}\ x_{12}\ \ldots\ x_{1j}\ \ldots\ x_{1c}$ $x_{21}\ x_{22}\ \ldots\ x_{2j}\ \ldots\ x_{2c}$ $\vdots$ $x_{i1}\ x_{i2}\ \ldots\ x_{ij}\ \ldots\ x_{ic}$ $\vdots$ $x_{r1}\ x_{r2}\ \ldots\ x_{rj}\ \ldots\ x_{rc}$ H_0: Data are consistent ($r \leqq 6$) in HP ($r \leqq 16$) in TI	Pr($\geqq \chi^2$)

DATA ENTRY

1 HP	1 TI	2	3 HP	3 TI	4 HP	4 TI	5 HP	5 TI
a, ENTER ↑ b, ENTER ↑ p, ENTER ↑ q, KEY A	KEY A KEY A KEY A KEY A	x_1, KEY B x_2, KEY B $\vdots$ x_r, KEY B KEY C	f_1, ENTER ↑ x_1, KEY D f_2, ENTER ↑ x_2, KEY D $\vdots$ f_r, ENTER ↑ x_r, KEY D KEY E	$x \rightleftharpoons t$ KEY D $x \rightleftharpoons t$ KEY D $\vdots$ $x \rightleftharpoons t$ KEY D KEY E	r, KEY a x_{11}, KEY a x_{21}, KEY a $\vdots$ x_{r1}, KEY a (all data by columns) x_{rc}, KEY a KEY b	KEY A' KEY A' KEY A' $\vdots$ KEY A' KEY A' KEY B'	ν, ENTER ↑ χ^2, KEY e	$x \rightleftharpoons t$ KEY E'

CALCULATOR OUTPUT

1	2	3	4	5
Degrees of freedom χ^2	$\bar{x}$ (the mean) Degrees of freedom χ^2	$\chi_1^2, \chi_2^2, \ldots \chi_r^2$ Degrees of freedom χ^2 (total)	χ^2 for each row Degrees of freedom χ^2 (total)	α where $\alpha = \Pr(\geqq \chi^2)$

$$\chi^2 = \sum \frac{(x-\bar{x})^2}{\bar{x}} = \frac{\Sigma(x-\bar{x})^2}{\bar{x}} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{\Sigma x/n}$$

$$= \frac{n(\Sigma x^2)}{\Sigma x} - \Sigma x \tag{5.9}$$

The final expression of Eq. (5.9) can be readily calculated with the built-in function of $\Sigma+$ which gives Σx^2, Σx, and n directly. χ^2-2 uses this method.

The individual observations are entered with KEY B, and when all the data are entered, KEY C calculates χ^2. The output is the mean, the degrees of freedom, and the χ^2 value. Program χ^2-5 can be added at the end of χ^2-2 and the probability would also be calculated.

If an incorrect value is accidently entered with KEY B, it can be deleted by repeating the same entry with KEY $\Sigma-$. The data could be entered with KEY $\Sigma+$, but KEY B is used because the program has a routine that clears the registers with the first data entry to ensure that no error is introduced from values remaining in the calculator.

5.8.3 Unequal Expectancy Test: χ^2-3 Program

The χ^3-3 program is for tests of data where both the number of observations and the size of the sample to which the number applies are involved. The number of accidents in groups of different sizes; the number of failures over different time intervals; the number of winners in different groups of players in games of chance are examples of unequal expectation.

The data are entered: first group size and then the number of observations in the group.

	HP	TI
f_1	ENTER ↑	KEY $x \rightleftharpoons t$
x_1	KEY D	KEY D
f_2	ENTER ↑	KEY $x \rightleftharpoons t$
x_2	KEY D	KEY D
⋮		
f_r	ENTER ↑	KEY $x \rightleftharpoons t$
x_r	KEY D	KEY D
When all the data are entered:	KEY E	KEY E

With the Hewlett-Packard calculator the limit to r, the number of groups, is 11. With the Texas Instruments instrument, it is 24 with the normal partition of registers and program memory, and the number could be increased to 40 without any difficulty.

The output is χ^2 for each pair of values, the total number of degrees of freedom and the total χ^2. If the total χ^2 is large enough to reject the hypothesis of homogeneity, the individual χ^2 values will be of interest to determine which values are out of line.

If the probability value associated with the total χ^2 is desired, the χ^2-5 program may be run with an input of the total number of degrees of freedom and the total χ^2 value, or the χ^2-5 program can be merged at the end of the χ^2-3 program and the probability will be part of the output.

The program as written is limited to 11 pairs of groups and observations in the Hewlett-Packard program, and to 24 in the Texas Instruments program. The individual values are stored until all the data are entered and the expected ratio of $\Sigma x/\Sigma f$ is calculated. Storing the input data requires two storage registers for each pair so that with the other requirements of the program, 25 storage areas are used in the Hewlett-Packard and 51 in the Texas Instruments. The Texas Instruments system may be modified to handle more than 24 sets of data, but the Hewlett-Packard 67/97 system cannot be changed.

The calculation is made using Eq. (5.7), and as each individual χ^2 value is calculated it is printed, and the sum is accumulated. The number of pairs of input data was counted as the data were entered, and when the number of individual χ^2 calculations is the same as the data input the calculation is terminated and the total χ^2 printed. The total number of degrees of freedom is one less than the number of pairs of input data.

Figure 5.1 shows the flow diagrams for the χ^2-1, χ^2-2, and χ^2-3 programs. The listing and detailed description of these programs are given in the Appendix. The three programs are written so that they may be combined on one magnetic card without any ambiguity of storage areas or operation. The probability program, χ^2-5, could be stored on the same card and run with any or all of the three. To follow a program with the probability calculation would require only the insertion of a GO TO e or GO TO E' instruction at the end of each one. The probability program, χ^2-5, is identified by the label e in the Hewlett-Packard program and E' in the Texas Instruments program.

5.8.4 Contingency Tables: χ^2-4 Program

A contingency table is one in which the number of observations may be identified by two categories or indices. An example was shown in Table 5.2. The homogeneity is tested by calculating the expected value for each observation in the table from the row and column totals. The χ^2-4 program uses Eq. (5.8) which is a modification of the general χ^2 equation (5.1).

In the χ^2-4 program the number of rows is entered first. The data are then entered by columns. The calculator stores the number of rows and matches the data input by columns to the number of rows. When the number of data entries equals the number of rows, the program "knows" that one column has been entered and applies the next data entry to the start of the next column. The number of rows may not be more than 6 in the Hewlett-Packard program, or more than 16 in the Texas Instruments program, and the number of columns is unlimited. If

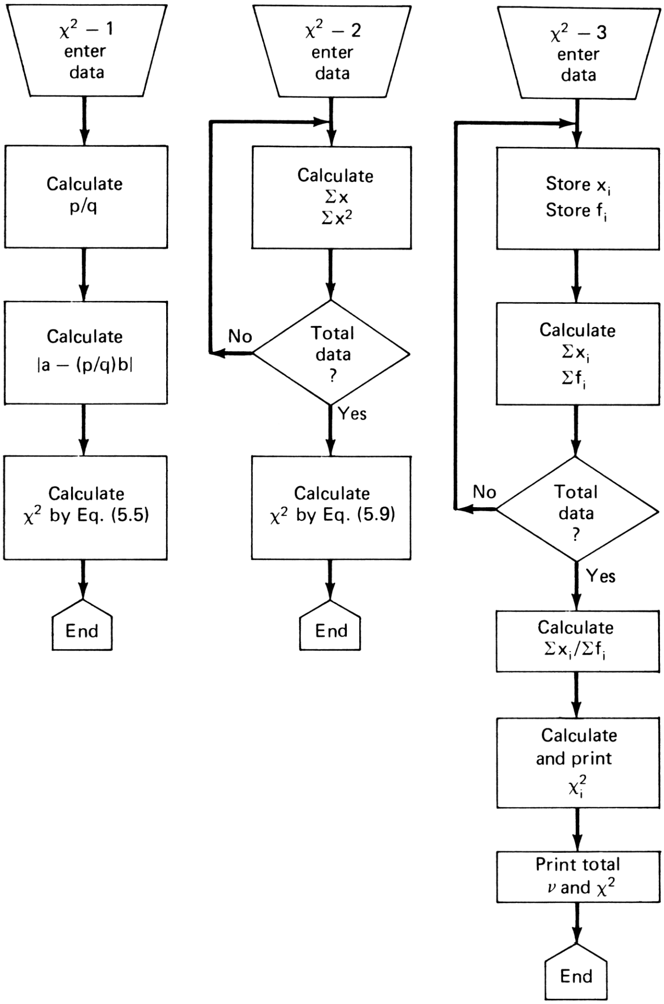


Figure 5.1 Flow diagrams for programs χ^2-1, χ^2-2, and χ^2-3.

the number of columns is 6 (16) or less and there are more than 6 (16) rows, the row and column designation may be interchanged.

After the number of rows is entered: r, KEY a, the data are entered by columns with the same KEY a: x_{11}, KEY a, x_{21}, KEY a, . . . , x_{r1}, KEY a, x_{12}, KEY a, x_{22}, KEY a, . . . , x_{rc}, KEY a in the Hewlett-Packard program, or KEY A' in the Texas Instruments program. At the completion of one column data input, the program carries out some calculations with the data for that column and there is a slightly longer time (a few seconds) before the calculator is ready for the start of the next column input than between data values within a column.

When all the data are entered, KEY b (or KEY B') starts the χ^2 calculation. The output is the χ^2 value for each row, the total number of degrees of freedom, and the total χ^2. If it had been anticipated that the χ^2 values for the column factor would be of more interest than that of the row factor, and there were no more than 6 (16) columns, the rows and columns could have been interchanged. The sum of the row χ^2 values will equal the total χ^2, except in a 2×2 table as discussed below.

If after the program has been run and the χ^2 is significantly large to reject the hypothesis of homogeneity, it is sometimes possible to find the observations which are most out of line by noting which rows make the largest contribution to the total χ^2 value. If there are less than 7 (17) columns, it is possible to pinpoint the observations that are most inconsistent by entering the data a second time with the rows and columns interchanged. The intersection of the rows and columns with the largest χ^2 values will show the observations in the table with the largest contribution to the total χ^2.

In the unique case of a 2×2 contingency table, the number of degrees of freedom is one, and a correction is made to the total χ^2 as discussed in Sec. 5.3. The correction does not apply to the χ^2 for each row, thus in the case of a 2×2 table the sum of the row χ^2's will not equal the total χ^2.

Again, the χ^2-5 program can be merged with the χ^2-4 program to give the probability associated with the total χ^2 value.

Figure 5.2 shows the flow of the χ^2-4 program, and a listing and detailed description will be found in the Appendix.

5.8.5 χ^2 Probability Program: χ^2-5

The χ^2-5 program calculates the probability of an equal or larger χ^2 from the number of degrees of freedom and the χ^2 value. The probability calculated is the probability of obtaining a χ^2 as large as or larger than that observed if the hypothesis of the observed values equaling the expected values was true and χ^2 was calculated with the number of degrees of freedom involved in the probability calculation. The calculation is based on Eq. 26.4.6 from Ref. 2. The equation is as follows:

$$\Pr(\geqq \chi^2) = 1 - \frac{\left(\frac{\chi^2}{2}\right)^{\nu/2} e^{-\chi^2/2}}{\Gamma\left(\frac{\nu+2}{2}\right)} \left\{1 + \sum_{n=1}^{\infty} \frac{\chi^{2r}}{(\nu+2)(\nu+4)(\;.\;.\;.\;)(\nu+2r)}\right\} \quad (5.10)$$

Where ν is the number of degrees of freedom and Γ is the gamma function:

$$\Gamma(n+1) = n!$$
$$\Gamma(1/2) = \sqrt{\pi}$$
$$\Gamma(3/2) = (1/2)\,\Gamma(1/2)$$
$$\Gamma(5/2) = (3/2)(1/2)\,\Gamma(1/2)$$

The program carries out the calculation until successive summation terms are less than 0.0001. This limitation gives the χ^2 probability to four significant figures.

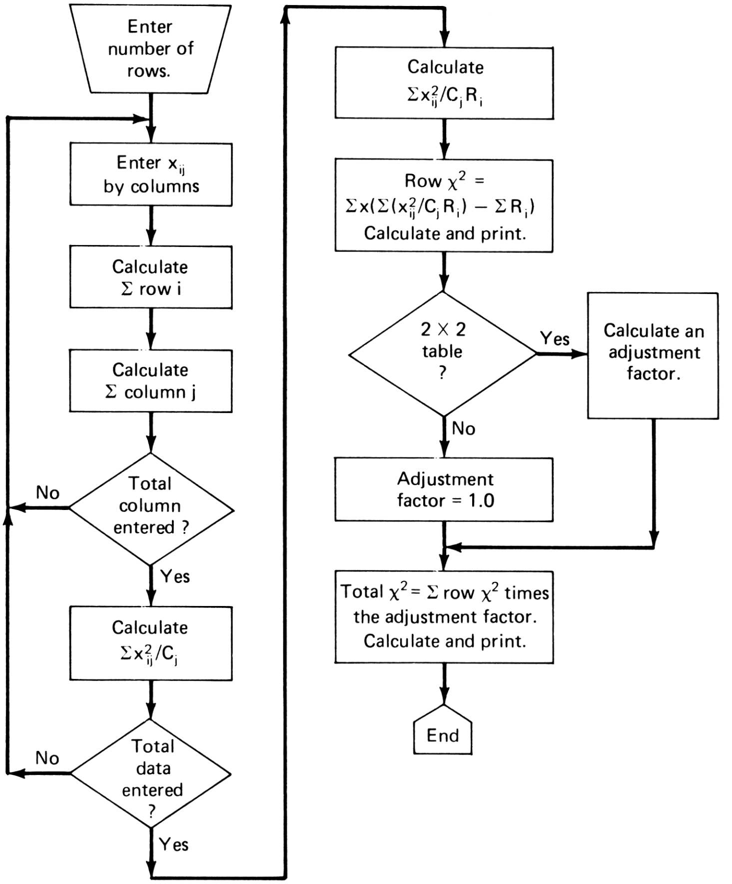


Figure 5.2 Flow diagram for program χ^2-4.

If greater accuracy is desired, it is a simple matter to change the calculation so that a smaller value is required to terminate the summation.

The program is run with: ν, ENTER ↑, χ^2, KEY e or KEY E'. The output is the probability. At the end of the calculation the values of ν and χ^2 are stored in Registers 1 and 2 for verification. If χ^2-5 is merged with any of the other χ^2 programs, it may be run directly by having the first line of χ^2-5 replace the "end" statement of the program with which it is merged.

The time required for the calculation of the χ^2 probability depends principally on the magnitude of χ^2. With large values of χ^2 the calculator time could be several minutes. However, with large values of χ^2 the probabilities are usually very small, unless the degrees of freedom are very large. Most tabulations of χ^2 do not give values for greater than 30 degrees of freedom or for χ^2 values greater than about 60. One reason for not including the χ^2-5 program with each of the χ^2 calculation programs is that if χ^2 is very large, the probability is obviously small and there is no necessity of taking the calculator time to calculate it.

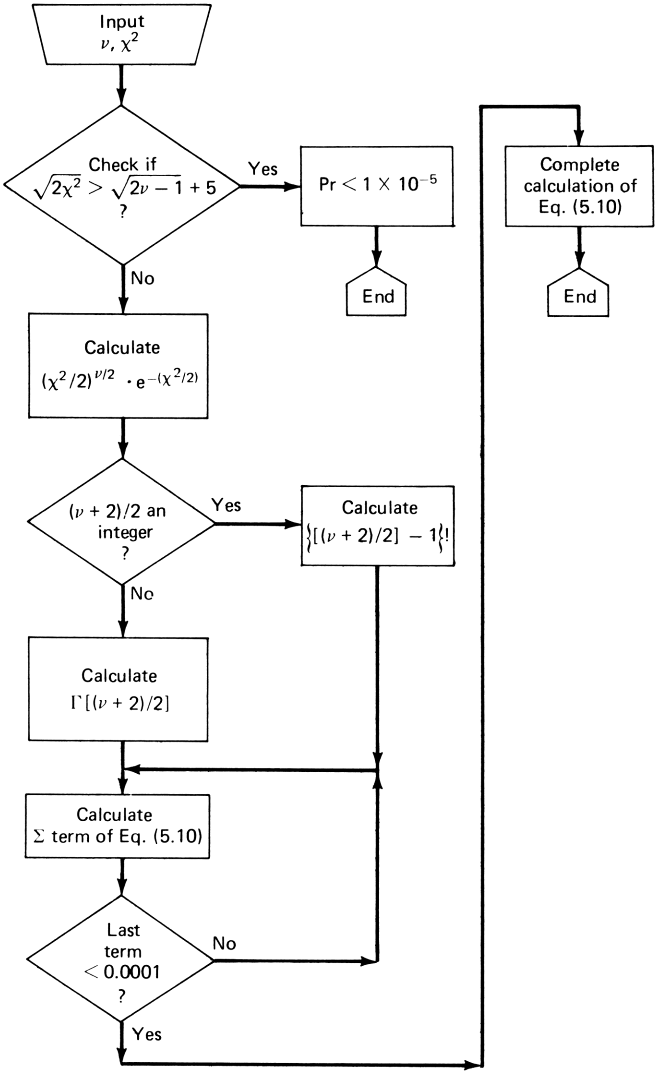


Figure 5.3 Flow diagram for program χ^2-5.

The value of χ^2 is related to the normal distribution, and as the number of degrees of freedom increases the value of $\sqrt{2\chi^2}$ approaches a normal distribution with a mean of $\sqrt{2\nu - 1}$ and a standard deviation of 1, where ν is the number of degrees of freedom. The χ^2-5 program contains a test of the input values, and if the χ^2 value is such that $\sqrt{2\chi^2}$ exceeds the mean, $\sqrt{2\nu - 1}$, by more than five standard deviations, for which the normal probability is less than 1×10^{-6}, the program output is a probability value of 1×10^{-5}. This test is inserted in the program to prevent the calculator from getting into a time-consuming calculation for which the probability calculated would be very small.

A flow diagram of χ^2-5 is given in Fig. 5.3 and the listing and detailed description are given in the Appendix.

5.9 EXAMPLES OF THE χ^2 PROGRAMS

All of the examples that follow are worked on the assumption that the χ^2-5 program is part of the program illustrated; that is, it has been merged at the end of the program. If it were not part of the program, the probability in each could be obtained by putting χ^2-5 in the calculator at the end of the χ^2 calculation and activating KEY e (KEY E'). An alternative would be to carry out the examples for the different χ^2 programs, accumulate the number of degrees of freedom and χ^2 answers from each problem, and then to run the χ^2-5 program separately with an input of the number of degrees of freedom and χ^2 for the probability of each example.

5.9.1 If a school usually passes two-thirds of a class in a particular subject and a new instructor fails 14 out of a class of 30, is the result in reasonable conformity with the ⅔ experience?

The observed failure/pass ratio is 14/16. The expected ratio is 1/2. H_0:14/16 agrees with 1/2 expectation. Using the χ^2-1 program

	HP	TI			
14,	ENTER ↑	KEY A			
16,	ENTER ↑	KEY A			
1,	ENTER ↑	KEY A			
2,	KEY A	KEY A	Output:	1.00	Degrees of freedom
				1.838	χ^2
				0.175	Probability

If the hypothesis of conformity is rejected, there is a 17.5 percent chance of error. (Time: 15/25)

Note: The ratio test could be made as 14/16 compared to 1/2, or 14/30 to 1/3, or 16/30 compared to 2/3. The answer would be different in each case because of the 0.5 correction for one degree of freedom. However, as the reader may confirm, none of the ratio tests is significant. The largest χ^2 value will be obtained if the ratio of the two parts is used rather than the ratio of one part to the total.

5.9.2 A candidate claims that 55 percent of the voters favor him, and a poll of 700 voters shows 358 in his favor and 342 opposed. Is his claim acceptable?

Observed ratio: 358/342. Expected ratio: 55/45. With program χ^2-1:

	HP	TI			
358	ENTER ↑	KEY A			
342	ENTER ↑	KEY A			
55	ENTER ↑	KEY A			
45	KEY A	KEY A	Output:	1.00	Number of degrees of freedom
				4.053	χ^2
				0.0441	Probability

There is less than a 5 percent chance of error if the hypothesis is rejected. (Time: 20/31)

5.9.3 In seven samples of treated refuse the following bacteria counts were found: 78, 76, 72, 78, 81, 74, 95. Are these results internally consistent? Hypothesis: There is no significant difference between the measurements and their mean. With the χ^2-2 program:

78, KEY B, 76, KEY B, 72, KEY B, 78, KEY B, 81, KEY B, 74, KEY B, 95, KEY B.

KEY C	Output:	79.1	Mean
		6.0	Number of degrees of freedom
		4.36	χ^2
		0.628	Probability

Accept the hypothesis. There is a 63 percent probability of error if it is rejected. (Time: 17/26)

5.9.4 Suppose the bacteria count had been 10 times as much in each case as in Example 5.9.3. Would the data still be consistent?

780, KEY B, 760, KEY B, 720, KEY B, 780, KEY B, 810, KEY B, 740, KEY B, 950, KEY B.

KEY C	Output:	791.4	Mean
		6.0	Number of degrees of freedom
		43.6	χ^2
		1×10^{-5}	Probability

(Time: 4/5)

In this case the hypothesis would be rejected with insignificant chance of error. Note that with the increase in the number of observations by a factor of 10, the χ^2 value increased by the same amount.

5.9.5 The following data show the number of employees and the number of lost-time accidents over a 12-mo period for four different categories of employees.

Department	No. of employees	No. of accidents
Welding shop	45	5
General maintenance	120	22
Research laboratory	83	6
Pilot plant	72	10

Is there significant difference in the accidents/worker-year for the different departments? Use the χ^2-3 program for unequal expectation of events.

	HP	TI
45	ENTER ↑	KEY $x \rightleftharpoons t$
5	KEY D	KEY D
120	ENTER ↑	KEY $x \rightleftharpoons t$
22	KEY D	KEY D
83	ENTER ↑	KEY $x \rightleftharpoons t$
6	KEY D	KEY D
72	ENTER ↑	KEY $x \rightleftharpoons t$
10	KEY D	KEY D
	KEY E	KEY E

The output is

0.181	χ^2 for welding shop
2.140	χ^2 for general maintenance
2.381	χ^2 for research
0.011	χ^2 for pilot plant
3.	Number of degrees of freedom
4.714	Total χ^2
0.194	Probability (Time: 35/35)

There is no significant difference. There would be a 19 percent chance of error if a significant difference was reported.

The possibility exists in an example of this type that while the overall χ^2 is not significant for the total number of degrees of freedom, one of the individual χ^2 values may exceed 3.841, the critical value for one degree of freedom at the 0.05 probability level. Strictly speaking, each individual χ^2 does not have one degree of freedom since the total number of degrees of freedom is one less than the total number of sets of observations. However, the individual χ^2 values give an indication of the conformity of each set of observations with the group. The χ^2-5 program can be used to obtain the probability value associated with the individual χ^2 based on one degree of freedom. The engineer will then have to make a judgment based on the results.

5.9.6 In the text an example was cited of a test on four different computer terminals. The data are tabulated below:

Terminal	Hours of testing	No. of failures
1	400	44
2	600	36
3	650	32
4	550	46

The χ^2-3 program is used to test whether the performance of the terminals is consistent.

HP	TI
400, ENTER ↑, 44, KEY D	400, KEY $x \rightleftharpoons t$, 44, KEY D
600, ENTER ↑, 36, KEY D	600, KEY $x \rightleftharpoons t$, 36, KEY D
650, ENTER ↑, 32, KEY D	650, KEY $x \rightleftharpoons t$, 32, KEY D
550, ENTER ↑, 46, KEY D	550, KEY $x \rightleftharpoons t$, 46, KEY D
KEY E	KEY E

The output is:

8.120	χ^2 for Terminal 1	
1.167	χ^2 for Terminal 2	
4.618	χ^2 for Terminal 3	
1.070	χ^2 for Terminal 4	
3	Degrees of freedom	
14.974	Total χ^2	
0.002	Probability value	(Time: 45/60)

There is significant difference, fewer than two chances in a thousand of being in error in saying so. The individual χ^2 values indicate that Terminal 1 had more than average failures and Terminal 3 had fewer.

5.9.7 As an example of a contingency table test, the data in Table 5.2 are used. The data list five different sources of failure for four compressors—a 5×4 table. With the χ^2-4 program, the data are tested for internal consistency.

	HP	TI
Number of rows:	5, KEY a	5, KEY A'
Data by columns:	8, KEY a, 14, KEY a 5, KEY a, 13, KEY a . . . , 11, KEY a, 6, KEY a	8, KEY A', 14, KEY A' 5, KEY A', 13, KEY A' . . . , 11, KEY A', 6, KEY A'
When all the data are entered:	KEY b	KEY B'

The output is:

0.338	χ^2 for row 1
0.641	χ^2 for row 2
4.376	χ^2 for row 3
1.696	χ^2 for row 4

12.	Number of degrees of freedom	
7.919	Total χ^2	
0.792	Probability	(Time: 50/82)

Note: The probability is calculated with the χ^2-5 program. In the Texas Instruments calculator, the χ^2-5 program is appended to the χ^2-4 program after the last statement in that program and replaces the RST instruction with GO TO E'. This change directs the operation to the probability calculation instead of ending the program with the output of the total χ^2 value. In order to make this change, the partition between storage area and program area must be changed since the combination of the two programs is too large for the normal partition of 479/59. (Texas Instruments users will understand this change.) With the Hewlett-Packard program, it is not possible to include the χ^2-5 program with program χ^2-4 since the two exceed the capacity of the HP-67/97 calculator and there is no provision for exchanging storage area for program area in these Hewlett-Packard calculators. However, the calculation of the total χ^2 value ends the program with the number of degrees of freedom and the χ^2 in the proper sequence for use with program χ^2-5. At the end of the calculation with the χ^2-4 program in the Hewlett-Packard system, the χ^2-5 program is entered into the calculator and the probability is calculated by simply activating KEY e.

To return to the example, the total χ^2 value of 7.919 with a probability of 0.792, is not significant. There would be a 79 percent chance of error if it were said that the data are not consistent. If there is any doubt about the third-row value of 4.376, the probability for this χ^2 value, at three degrees of freedom (four columns provide three degrees of freedom for each row χ^2 calculation), may be checked with the χ^2-5 program. The probability is 0.225—not significant.

If one of the row χ^2 values was significant, and it was not obvious which observation(s) were out of line, the data could be reentered using the columns as rows, and the rows as columns, to pinpoint the value making the most contribution to the χ^2. The next example illustrates this calculation.

5.9.8 The following data show some qualitative test scores of a number of applicants with different levels of education:

Applicant education level	Scoring			
	Low	Medium	High	Very high
College	18	28	75	115
High school	17	28	29	42
Grade school	11	11	11	19

The χ^2-4 program is used to determine whether there is any significant difference in the results for the different education levels. The data are entered as follows: 3 (number of rows), KEY a, 18, KEY a, 17, KEY a, . . . , 42, KEY a, 19, KEY a. With all the data entered, KEY b starts the calculation. With the Texas Instruments program the procedure is the same with KEY A' and KEY B'.

The output is:

8.454	χ^2 for row 1	
7.010	χ^2 for row 2	
6.578	χ^2 for row 3	
6.0	Number of degrees of freedom	
22.043	χ^2 total	
0.0012	Probability	(Time: 54/103)

A test with program χ^2-5 of the individual row χ^2 values, using three degrees of freedom, show only the first-row value significant—probability 0.038. The probabilities for the other two row values are 0.071 and 0.087.

To find which values in the first row make the most contribution to the total χ^2, the data are entered again with the rows and columns interchanged: 4 (number of columns) KEY a, 18, KEY a, 28, KEY a, 75, KEY a, . . . , 11, KEY a, 11, KEY a, 19, KEY a. Again, with all the data entered, KEY b starts the calculation. With the TI program use KEYs A' and B'.

The output is:

8.375	χ^2 for column 1	
7.816	χ^2 for column 2	
2.377	χ^2 for column 3	
3.476	χ^2 for column 4	
6.0	Number of degrees of freedom	
22.043	χ^2 total	
0.0012	Probability	(Time: 50/103)

The χ^2 values for the first two columns are significant: probabilities of 0.015 and 0.020. These values were obtained from the χ^2-5 program, using two degrees of freedom for the column χ^2 values. With the first row and the first two columns making the principal contribution to the total χ^2, the number of college-level applicants having low and medium scores are the values that are out of line.

5.9.9 In the two previous examples the sum of the individual χ^2 values for either the rows or the columns equaled the total χ^2. In a 2×2 table this will not be the case inasmuch as the total χ^2 is corrected for the calculation with one degree of freedom. The following data show a comparison of the number of successful and unsuccessful runs made on a modified and unmodified pilot plant.

	Runs	
Plant	Successful	Unsuccessful
Modified	115	24
Unmodified	98	36

Do the results substantiate the claim that the modified unit is better?

Using program χ^2-4, the data are entered: 2 (number of rows), KEY a, 115, KEY a, 98, KEY a, 24, KEY a, 36, KEY a. KEY b. (KEYs A' and B' with the Texas Instruments.)

The output is:

1.800	χ^2 for first row	
1.867	χ^2 for second row	
1	Number of degrees of freedom	
3.128	χ^2 total	
0.077	Probability	(Time: 30/50)

The total χ^2 is just not significant. A few more data are required to permit the claim that there is a difference between the units with less than a 5 percent chance of error. Note that if it wasn't for the correction for one degree of freedom, the sum of the two row χ^2's would be significant.

5.10 TOO GOOD A FIT

All of the χ^2 examples discussed have dealt with testing whether the calculated χ^2 was larger than would be expected if the observed values matched the expected values. If the calculated χ^2 was equal to or larger than a value that would occur with 0.05 probability or less, the χ^2 was said to be significant, and the hypothesis that the observed values equaled the expected values was rejected.

There is another possibility. The χ^2 value may be smaller than would be expected by chance. If the observations match the expectations exactly, the χ^2 value is zero. If a very small χ^2 results from the calculation, the possibility exists that the data are not random values from a population of results but were selected to fit the hypothesis. A χ^2 value corresponding to a probability of 0.95 (or 0.99) represents a result that would not be expected to occur more than 5 percent (or 1 percent) of the time if the observed values were taken from a population having the expected values as a mean.

If experimental data show too good a fit and result in a χ^2 value with a probability greater than 0.95, or certainly if the probability is greater than 0.99, the data should be examined for a forced conformity to the expected values.

6 VARIANCE AND THE ANALYSIS OF VARIANCE

6.1 INTRODUCTION

The variance was defined in Chap. 2, Sec. 2.2.1. It is the mean-squared deviation from the mean:

$$\sigma^2(x) = \frac{\Sigma(x_i - m)^2}{n} \tag{2.9}$$

An estimate of the variance calculated from a sample is the sum of the squares of deviations of the sample measurements from the sample mean divided by one less than the number of samples:

$$s^2(x) = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \tag{2.10}$$

The difference between these two equations is that the population variance is calculated from differences measured from the true population mean m, and the sample estimate is calculated from differences from the sample mean $\bar{x}$, which is only an estimate of the population mean. The division by $n - 1$ instead of by n is to compensate for the bias of calculating the differences from an estimate of the mean.

Although the standard deviation, the square root of the variance, is usually reported when one is reporting data, it is the variance that is a more basic mathematical function.

If a variable X is a function of a number of normally distributed, independent variables x_i, the variance of X is directly related to the variances of the variables x_i. If X is equal to the sum or difference of a number of variables x_i:

$$X = x_1 \pm x_2 \pm x_3 \pm \cdots \tag{6.1}$$

the variance of X is equal to the sum of the variances of x_i:

$$\sigma^2(X) = \sigma^2(x_1) + \sigma^2(x_2) + \sigma^2(x_3) + \cdots \tag{6.2}$$

This statement is not true for the standard deviation.

The general relationship for the variance of a variable X is that if X is a function of a number of normally distributed independent variables:

$$X = f(x_1, x_2, x_3, \ldots) \tag{6.3}$$

X will be normally distributed with a variance equal to the sum of the variances of the individual variables multiplied by the square of the partial derivative of the function with respect to each variable:

$$\sigma^2(X) = \sigma^2(x_1)\left(\frac{\partial f}{\partial x_1}\right)^2 + \sigma^2(x_2)\left(\frac{\partial f}{\partial x_2}\right)^2 + \cdots \tag{6.4}$$

The analysis of variance is a method of separating the total variance of some variable into the sum of the contributing variances in order to find which are important.

Another attribute of variances is that the variances of two variables may be rigorously compared in a manner similar to that in which the t test was used to compare two means. Confidence limits may be placed on variances, and groups of variances may be tested for homogeneity. The function for testing variances, similar to the t function for means, is called the F function.

6.2 THE F TEST FOR VARIANCES

The statistical procedure outlined for the t test in Chap. 4 was to set a hypothesis about a mean, to calculate the t value based on the hypothesis, and if the calculated t value was one of low probability, the hypothesis was rejected. A similar procedure is followed with variances using the F distribution. A hypothesis about two variances is established and an F statistic is calculated, usually from a ratio of the two variance estimates. If the calculated value of F is one of low probability, the hypothesis is rejected.

The t probability distribution included as one parameter the number of degrees of freedom associated with its calculation. The F distribution calculation involves two degrees-of-freedom terms, one for each of the variances being compared. The tabulated values of F are commonly available for only two probability levels, 0.05 and 0.01, because of the necessity of having values for combinations of two different degrees of freedom. In the calculator programs discussed later there is no restriction on the probability levels.

The rationale behind the F ratio test of two variances is that the larger, which is usually the numerator, may contain a variance factor that is not included in the smaller variance, the denominator. If the variances were in fact equal, their ratio would be 1.0. If the samples from which the variance estimates are calculated are from the same population, then the variance estimates would be estimates of the same value. They would not be expected to be exactly the same, but their ratio would not be much different from 1. The amount of difference would depend on the size of the samples which is reflected in the number of degrees of freedom of the variance estimates. The F statistic is a mathematical function based on the variation of variances of different sample sizes taken from normal populations. Calculator program Variance-9 calculates the probability for a given F value if the two different degrees of freedom are specified.

6.3 COMPARISON OF A VARIANCE WITH A STANDARD VALUE

Two general types of F tests are considered: first, the comparison of the variance estimate with some standard or target value; second, the comparison of two variance estimates with each other.

In the first test a sample is obtained and the variance of the population, estimated from the sample, is hypothesized to equal some predetermined value:

$$H_0 : \sigma^2(x) = V \tag{6.5}$$

The variance is estimated from the sample data, using one of the forms in Eq. (6.6):

$$\sigma^2(x) \approx s^2(x) = \frac{\Sigma(x_1 - \bar{x})^2}{n-1} = \frac{\Sigma x_i^2 - \bar{x}\Sigma x_i}{n-1} \tag{6.6}$$

Some critical value, 0.05 or 0.01, is set for the probability, which is the minimum acceptable probability for a false rejection of H_0. F is calculated as:

$$F = \frac{s^2(x)}{V} \tag{6.7}$$

If the calculated value of F exceeds F at the critical probability level, calculated with $n - 1$ and ∞ degrees of freedom, the hypothesis may be rejected with the critical probability as the chance of error. With the calculator program, if the calculated probability is less than the selected critical probability, the hypothesis is rejected.

The ∞ degrees of freedom are associated with the standard or target variance V since that value is assumed to be known. Variance V represents the variance of a hypothetical population to which the population from which the sample was taken is being compared.

6.3.1 Variance-1 Program

The usual procedure, after making the hypothesis to be tested, is to observe the sample data, calculate F from Eq. (6.7), and then look at an F table for the degrees of freedom corresponding to the variance estimate calculation.

The Variance-1 program does all of the calculations and gives the probability value. It takes the original observations and the target variance as input data:

x_1, KEY A, x_2, KEY A, . . . , x_n, KEY A, V, KEY B.

The program output is $s^2(x)$, the degrees of freedom ν, F, and α, the probability associated with F at ν and ∞ degrees of freedom.

The calculations are made using the built-in functions of KEY $\Sigma+$. If during the data entry an incorrect value is entered, it may be deleted by reentering the same value with KEY Σ−. The deletions may be made before or after using KEY B for the F calculation. If the deletions are made after having made an F calculation, ν must be reentered with KEY B to recalculate F.

The Variance-1 program will also calculate the F value and the probability if the input is the estimated variance from some source, the target variance, and the number of degrees of freedom. The F value is merely the quotient of the observed variance estimate and the target value, and a calculator is not needed. But, together with the number of degrees of freedom, the program will also calculate the probability value.

For this type of operation the input with the Hewlett-Packard program is: ν, degrees of freedom, ENTER ↑, $s^2(x)$, the estimated variance, ENTER ↑, and V, the target variance, KEY C; and the output is the F value and the probability.

With the Texas Instruments program, the input is: ν, KEY C, $s^2(x)$, KEY C, and V, KEY C. The output is F and the probability value.

At the end of this chapter, there is a discussion of a calculator program Variance-9, which calculates the probability for an F value if two variance estimates are available, each with its degrees of freedom. That program is not suitable if one of the variances is known with infinite degrees of freedom. The Variance-1 program uses an alternate calculation which is applicable if one variance has infinite degrees of freedom.

There is a relation between the F and the χ^2 distributions that is beyond the scope of this text, but the F value at infinite degrees of freedom for one variance estimate is equal to χ^2 divided by the degrees of freedom of the other variance estimate at the same probability level:

$$F_{\nu,\infty,\alpha} = \frac{\chi^2_{\nu,\alpha}}{\nu} \tag{6.8}$$

The χ^2-5 program given in Chap. 5 calculates the probability for χ^2 at a given number of degrees of freedom. That same calculation is used in the Variance-1 program. After the degrees of freedom and F are calculated in the Variance-1 program, they are multiplied to obtain χ^2 from Eq. (6.8). The probability for that value of χ^2 is calculated from the χ^2-5 program, which is made part of the Variance-1 program, and the probability associated with F at ν, and ∞ degrees of freedom is obtained.

The outputs from Variance-1 are stored in Registers A, B, C, and D in the Hewlett-Packard program or Registers 10, 11, 12, and 13 in the Texas Instrument program, hence they are available to more significant figures if desired.

A flow diagram of the Variance-1 program is given in Fig. 6.1 as well as a flow diagram of the Variance-2 program, described in the next two sections. A listing and a detailed description of the Variance-1 program are given in the Appendix.

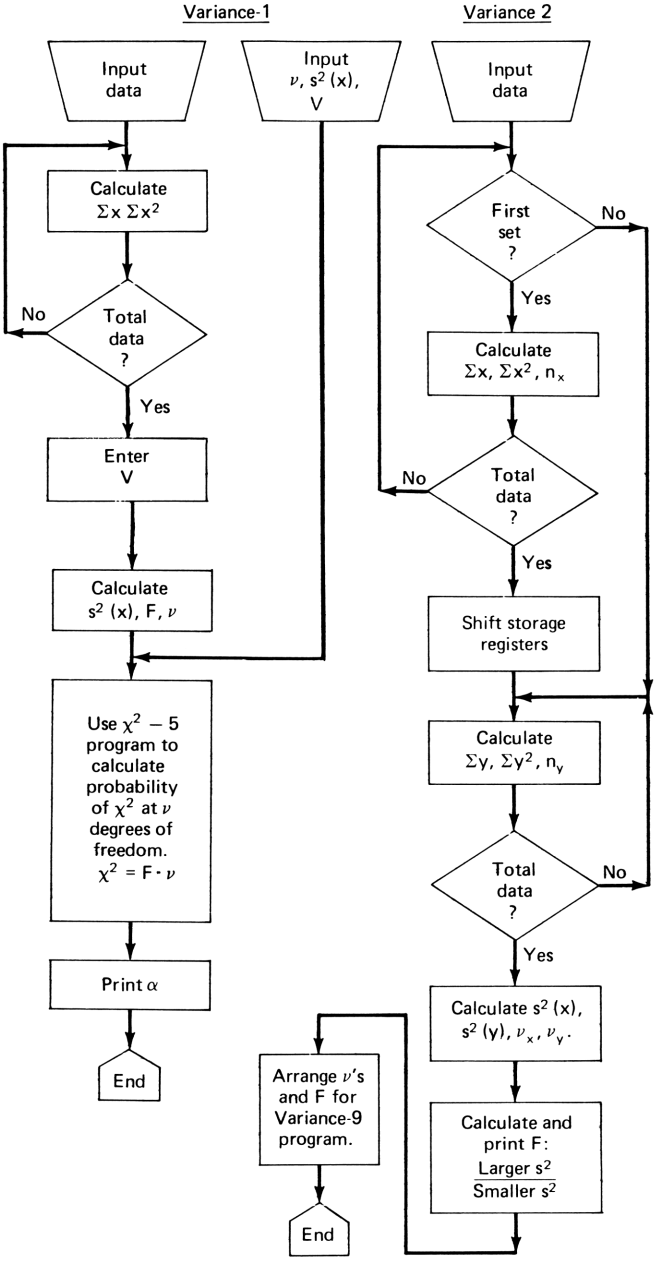


Figure 6.1 Flow diagrams for the Variance-1 and Variance-2 programs.

6.4 COMPARISON OF TWO VARIANCES

The most common F test is that between two variance estimates. The usual hypothesis is that the two variances are equal:

$$H_0: \sigma_1^2 = \sigma_2^2 \tag{6.9}$$

The two variance estimates are calculated from two sets of data, following Eq. (6.6) for each, and the ratio of the two estimates is F:

$$F = \frac{s^2(x_1)}{s^2(x_2)} \tag{6.10}$$

where the larger variance is in the numerator, giving an F value greater than unity.

Again, the usual procedure, after calculating the two variance estimates and the F value, is to compare the calculated F with tabulated values. If F exceeds a tabulated value at the critical probability level, the hypothesis is rejected. Usually only two critical probability values are available in tables, 0.05 and 0.01. With the calculator program, a probability value is obtained so the exact chance of a false rejection of the hypothesis is available.

6.4.1 Variance-2 Program

The Variance-2 program takes as input the original data for two sets of observations. It calculates the variance estimate of each, and the F value for the ratio of the larger to the smaller. The probability associated with the F value and the two degrees of freedom is calculated with the Variance-9 program. Variance-9 could be merged with Variance-2 if sufficient calculator facility is available. The calculations from Variance-2 are in the correct order for running Variance-9 immediately after the completion of the F calculation.

One set of data is entered with KEY A: x_1, KEY A, x_2, KEY A, . . . , x_{n_x}, KEY A. The second set of data is entered with KEY B: y_1, KEY B, y_2, KEY B, ... , y_{n_x}, KEY B. The number of observations in the two sets of data do not have to be the same. After the data are entered, KEY C starts the F calculation.

In the Hewlett-Packard program the variance estimates of both sets of data are calculated using the built-in function of KEY $\Sigma+$. If an error is made in entering either set of data, a correction can be made by reentering the same value with KEY $\Sigma-$. At the end of the input of the first set of data, the program shifts the storage registers to receive the second set of data. Any corrections to the first set of data must be made before KEY B is used, and any corrections to the second set must be made before KEY C is used.

In the Texas Instruments program, only the first variance estimate is calculated directly with the KEY $\Sigma+$. The sums for the second set of data are accumulated by instructions in the program and are later shifted to the storage registers so the built-in function of the variance estimate may be used. Only errors in the first set of data may be corrected by using KEY $\Sigma-$.

Both programs select the larger variance for the numerator in making the F calculation and put the numbers of degrees of freedom in the proper order for the probability calculation: the numerator degrees of freedom first, the denominator degrees of freedom second, and the F value third.

The output from the programs is the variance estimate and the degrees of freedom of the two sets of data in order of input and the F value. If the means of the two sets of data are desired, these values can be obtained by adding an $\bar{x}$ and PRINT instruction: in the Hewlett-Packard program between lines 20 and 21 and between lines 34 and 35; between lines 25 and 26 and between lines 53 and 54 in the Texas Instruments program.

To obtain the probability value in the Hewlett-Packard calculators, the Variance-9 program can be put in the calculator after the completion of Variance-2, and KEY A will get the probability value. The combination of Variance-2 and Variance-9 is too large for the HP-67/97 calculators.

In the Texas Instruments system, the Variance-9 program can be appended to Variance-2 to obtain the probability value directly after the calculation of F. The two programs together are larger than the normal partition of storage and program memory. Ten storage areas would have to be shifted to program space to accommodate the combined programs. If the programs are combined, line 91 in Variance-2 should be changed to GO TO LNX; lines 0 to 21 should be deleted from Variance-9; and the first line of the modified Variance-9 should replace line 109, the end line of Variance-2. If the calculation is not made very often, it may be simpler to run the two programs separately in both calculators.

6.5 COMPARISON OF MORE THAN TWO VARIANCES

The F test may be used to compare two variance estimates, but it is not applicable to testing the homogeneity of more than two. For more than two the Bartlett χ^2 test is used.[4] χ^2 is calculated from the formula of Eq. (6.11) for the comparison of k variance estimates and has $k - 1$ degrees of freedom. The χ^2 calculated from Eq. (6.11) is evaluated in the same way as the other χ^2's discussed in Chap. 5. If the calculated χ^2 has a value with low probability, the hypothesis of homogeneity of the variances is rejected.

The Bartlett χ^2 formula is:

$$\chi^2 = \frac{(\ln \bar{s}^2) \sum_1^k (\nu_i) - \sum_1^k (\nu_i \ln s_i^2)}{1 + \dfrac{1}{3(k-1)} \left(\sum_1^k \dfrac{1}{\nu_i} - \dfrac{1}{\sum_1^k \nu_i} \right)} \tag{6.11}$$

where $\bar{s}^2$ is the pooled variance estimate following Eq. (4.7) for the pooled estimate of two variances; $\bar{s}^2$ is the weighted mean of the variance estimates with each estimate weighted by its degrees of freedom:

$$\bar{s}^2 = \frac{\Sigma(s_i^2)(\nu_i)}{\Sigma(\nu_i)} \tag{6.12}$$

ν_i is the degrees of freedom for the ith set of data ($n_i - 1$), and k is the number of estimates.

Equation (6.11) may appear somewhat formidable, but it is simply the natural logarithm of the pooled estimate of variances multiplied by the total degrees of freedom minus the sum of the ln of each estimate multiplied by its degrees of freedom, and the difference divided by a correction factor depending on the degrees of freedom and the number of sets of data.

The usual procedure is to calculate the different variance estimates and to tabulate and sum the terms necessary to calculate χ^2. The Variance-3 program carries out all the calculations, including calculating the variance estimates, and gives the probability value.

6.5.1 Variance-3 Program

The Variance-3 program may be used in two ways. It can take as input the individual observations. With the input of the individual observations the program calculates the variance estimates for each set, calculates χ^2 from Eq. (6.11), and then calculates the probability associated with the result.

Or, the program can take input of variance estimates and degrees of freedom. With this input, the program calculates the χ^2 from Eq. (6.11) and the probability value.

For the first type of operation the input is the individual data values with KEY A: x_1, KEY A, x_2, KEY A, . . . , x_n, KEY A. When a complete set of data is entered, KEY B initiates the calculation of the variance estimate for that set and stores the result.

The next set of data is entered in the same manner as the first, following KEY B.

After all the data are entered, and KEY B has been used after the last set of data, KEY C starts the χ^2 calculation.

The output is the variance estimate and degrees of freedom for each set of data, and then the total degrees of freedom for the χ^2 calculation, the χ^2 value and the probability.

For the second type of operation, the input is the degrees of freedom and the variance estimate for each set of data with KEY D:, ν_1, ENTER ↑, $s^2(x_1)$, KEY D, ν_2, ENTER ↑, $s^2(x_2)$, KEY D, . . . , ν_k, ENTER ↑, $s^2(x_k)$, KEY D. When all the pairs of degrees of freedom and variance estimates have been entered, KEY C starts the χ^2 calculation. The TI program uses KEY $x \rightleftharpoons t$ in place of ENTER ↑.

The output is the degrees of freedom for the χ^2 calculation, ($k - 1$), the χ^2 value, and the probability.

When individual data values are entered with KEY A, the program calculates the variance estimate using the built-in function of KEY $\Sigma+$. When all the data from one set have been entered and KEY B is actuated, the program calculates the variance estimate and clears the registers for the next set of data. If an error has been made in entering the data, it can be corrected with KEY $\Sigma-$ by reentering the same data, but before KEY B is used.

With either type of usage, the program uses the χ^2-5 routine to calculate the probability value.

A flow diagram of Variance-3 is given in Fig. 6.2, and the listing and detailed description are given in the Appendix.

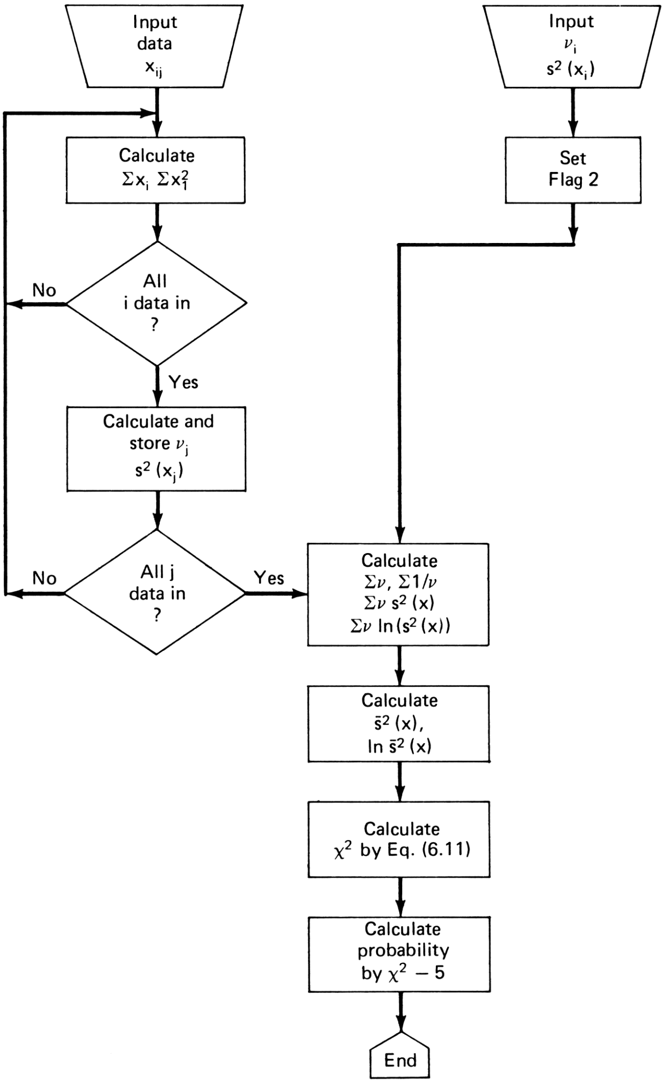


Figure 6.2 Flow diagram for the Variance-3 program.

6.6 EXAMPLES OF THE PROGRAMS: VARIANCE-1, VARIANCE-2, AND VARIANCE-3

6.6.1 A mass spectrometer is known from long experience to have a variance for a particular component of 0.24. A new analyst runs a series of determinations, and for 25 analyses obtains an estimated variance of 0.326. Has the new analyst contributed another variance factor to that of the instrument, or is the value 0.326 not significantly different from the standard value of 0.24?

With program Variance-1, the data are entered:

	HP	TI	
24	ENTER ↑,	KEY C	Degrees of freedom
0.326,	ENTER ↑,	KEY C	Variance estimate
0.24,	KEY C	KEY C	Target variance

Output:

1.358	*F* value	
0.113	Probability	(Time:37/60)

If the hypothesis that there is no difference between the variance obtained by the new analyst and the usual value is rejected, there is an 11 percent chance of being in error.

6.6.2 A raw material received in drums usually has a variance within a shipment of 7.0. A new shipment is received, and 10 drums are sampled. Analysis of the samples gives the following:

Drum	Analysis
1	62.8
2	60.0
3	58.7
4	57.1
5	55.1
6	63.4
7	62.5
8	59.2
9	54.8
10	58.3

Does the variation in the shipment conform to the usual value?

With program Variance-1, the data are entered: 62.8, KEY A, 60, KEY A, . . . , 54.8, KEY A, 58.3, KEY A, 7, (the target variance) KEY B.

The calculator output is:

9.308	Estimated variance	
9.0	Degrees of freedom	
1.330	F value	
0.215	Probability	(Time:29/39)

The shipment is accepted as meeting the variance specification. If it were rejected, there would be a 21 percent chance of error. If a better agreement were desired, more samples would have to be taken.

6.6.3 A special maternal diet is proposed to reduce the variation in the weight of newborn animals. One animal is fed the normal diet and another is fed the special diet. The weight at birth of the young in the litter from each of the animals is as follows:

Regular diet: 3.2, 2.8, 3.2, 3.5, 2.3, 2.4, 2.0, 1.6 oz
Special diet: 3.3, 3.6, 2.6, 3.1, 3.2, 3.3, 2.9, 3.4, 3.2, 3.2 oz

Is one litter more consistent than the other?

Using program Variance-2, the data are entered: 3.2, KEY A, 2.8, KEY A, . . . , 2.0, KEY A, 1.6, KEY A, completing the entry of the first set of data; 3.3, KEY B, 3.6, KEY B, . . . , 3.2, KEY B, 3.2, KEY B. With both sets of data entered, KEY C makes the F calculation.

The output is:

0.4364	Variance of first litter	
7.0	Degrees of freedom	
0.0751	Variance of second litter	
9.0	Degrees of freedom	
5.810	F value	
0.009	Probability from Variance-9	(Total time, including entering Variance-9: 40/75) (Time for merged programs in TI: 52 s)

The F value is significant. There is less than 1 percent chance of being in error if the hypothesis of equal variances is rejected.

6.6.4 Table 6.1 shows the replicate results of several analysts. The analyses were run on a control material, and the question asked is whether there is significant difference in the variation of the results obtained by the different analysts. Program Variance-3 provides an answer to questions of this type.

The data are entered in program Variance-3, each set with KEY A followed by KEY B. When all the data are in, KEY C makes the calculation. 32.0, KEY A, 32.8, KEY A, . . . , 31.3, KEY A, KEY B. 33.5, KEY A, 32.8, KEY A,

Table 6.1 Analytical Determinations from Five Analysts

1	2	3	4	5
32.0	33.5	33.3	33.2	32.6
32.8	32.8	33.6	33.3	32.6
33.3	33.2	32.6	33.2	32.9
33.2	33.5	31.3	32.9	32.0
34.4	32.3	33.3	33.3	32.0
33.6	32.4	32.9	32.5	32.1
31.9	32.0	33.4	32.6	
33.3	31.6	33.2	32.8	
32.8		33.2		
31.3				

. . . , 31.6, KEY B, . . . 32.6, KEY A, 32.6, KEY A, . . . 32.1, KEY A, KEY B, KEY C.

The output is the variance estimate and degrees of freedom of each set of data:

Variance estimate	Degrees of freedom
0.8360	9.0
0.4970	7.0
0.4794	8.0
0.1021	7.0
0.1467	5.0

And finally the degrees of freedom for the χ^2 calculation, the χ^2 value, and the probability:

4.0	Degrees of freedom	
8.887	χ^2	
0.064	Probability	(Time:27/32)

If a 5 percent risk is the maximum that is allowed, the hypothesis of equal variances cannot be rejected. The χ^2 probability is 0.064. The last two variance estimates appear to be lower than the others, but the test indicates they are not statistically lower. More data would have to be taken to support the rejection of the hypothesis.

6.6.5 The other calculation available with Variance-3 is the direct comparison of a number of variances that have already been calculated. If the variance estimates and the degrees of freedom for each are available, Variance-3 can be used to check their homogeneity.

The data in Table 6.2 show the variance estimate calculated from 90 tests each on the crack growth in vulcanized rubber from five different experiments. Are the variances homogeneous?

Table 6.2 Variance Estimates from Tests on Crack Growth

Experiment	Variance estimate
1	0.0268
2	0.0387
3	0.0482
4	0.0859
5	0.0552

The data are entered in Variance-3 with KEY D, and the calculation is run with KEY C. The number of degrees of freedom in each case is 89, one less than the number of tests. With the Hewlett-Packard program, the routine is:

89, ENTER ↑, 0.0268, KEY D, 89, ENTER ↑, .0387, KEY D, . . . , 89, ENTER ↑, 0.0552, KEY D. KEY C.

With the Texas Instruments program the data are entered with KEY $x \rightleftharpoons t$ and KEY D.

The output is:

4.0	Degrees of freedom for χ^2	
32.91	χ^2	
1.00×10^{-5}	Probability	(Time:5/7)

The χ^2 value is highly significant, and the hypothesis of equal variances is readily rejected. It is interesting to note that if the measurements were made on 25 samples with the same results, the resulting χ^2 would have a probability of 0.067, which is just not significant. The point is made merely to emphasize the increase in sensitivity of the test with increase in sample size.

6.7 ANALYSIS OF VARIANCE

The analysis of variance is a technique whereby the total variation of data may be separated into components contributed by different factors. If a number of similar experiments are made in three different reactors and there is variation in the results, how much of the variation is contributed by the reactors and how much is inherent in the experiment? With the analysis-of-variance technique it is possible to design an experimental program such that the maximum amount of information on the variation of the results may be obtained from a minimum number of experiments. There are several quite voluminous texts dealing with design and analysis of the data.[5,6]

The number of experimental designs is essentially unlimited and depends on the ingenuity of the experimenter, the problem to be solved, and also the facilities available. The purpose of these few sections is to present some basic designs and calculator programs for the analysis-of-variance solution to these designs. Some indication is given for the expansion of the programs to solve more involved designs.

Programs are given for five different types of analysis of variance. Table 6.11 at the end of this chapter summarizes the programs.

6.8 ONE FACTOR WITH REPLICATION

The basis for the analysis-of-variance calculation is described in this section for a one-factor experiment with replication of the experiments. More involved designs follow a similar analysis but the details are not presented in this text. The interested student will find them in texts dealing specifically with the design of experiments or with analysis of variance. (See the previously mentioned references or Ref. 7.)

Table 6.3 shows an arrangement of data in which r reactors make a number of replicate runs, each result indicated by x_{ij}—the ith run in the jth reactor. There are three sources for an estimate of the variance of the results. All the data could be pooled and a variance could be calculated following Eq. (2.10):

$$s^2(x) = \frac{\Sigma(x_{ij} - \bar{x})^2}{n - 1} \tag{6.13}$$

The replicate data from each reactor could be calculated separately and a pooled estimate of variance could be calculated, following Eq. (4.7):

$$\bar{s}^2(x) = \sum_{j=1}^{r} \sum_{i=1}^{k_j} \frac{(x_{ij} - \bar{x}_j)^2}{k_j - 1} \tag{6.14}$$

Or, the variance could be estimated from the variation of the means of the results in each reactor from the overall mean, and the subsequent calculation of the estimated variance from the estimated variance of the mean:

$$s^2(x) = n \cdot s^2(\bar{x}) \tag{6.15}$$

Table 6.3 Data for One-Factor Analysis of Variance with Replicates

Reactor			
R_1	R_2 . . .	R_j . . .	R_r
x_{11}	z_{12}	x_{1j}	x_{1r}
x_{21}	x_{22}	x_{2j}	x_{2r}
.	.	.	.
.	.	.	.
x_{i1}	x_{i2}	x_{ij}	x_{ir}
.	.	.	.
.	$x_{k_2 2}$	.	.
.			
$x_{k_1 1}$		.	$x_{k_r r}$
		.	
		$x_{k_j j}$	

which follows from the discussion in Sec. 4.2 where the relation $s(\bar{x}) = s(x)/\sqrt{n}$ was presented. The estimated variance calculated from the variance of the means has an equivalent form:

$$s^2(x) = \frac{\Sigma k_j(x_j - \bar{x})^2}{r - 1} \tag{6.16}$$

The pooled estimate of variance calculated from each set of replicates, Eq. (6.14), is the minimum variance estimate. It is an estimate of the error of the measurements, of how well each reactor repeats its results. The variance estimated from the means of the results from each reactor, Eq. (6.16), includes the error variance and any additional variance contributed by differences among the reactors. The analysis-of-variance test, the F test of $s^2(x)$ from Eq. (6.16) divided by the error variance estimate from Eq. (6.14) establishes whether there is a reactor effect on the variation of the data.

The total sum of squares of deviation, the numerator of Eq. (6.13) is equal to the sum of the numerators (the sums of squares) of the other two equations, (6.14) and (6.16). This fact can be seen by writing the sums of squares of deviations in their equivalent form shown in Eq. (2.11). The equivalence shown in Eqs. (6.17), (6.18), and (6.19) below might be more readily seen if it is recalled that $\bar{x}_j = (\Sigma x_{ij})/k_j$.

$$\Sigma(x_{ij} - \bar{x})^2 = \Sigma x_{ij}^2 - \frac{(\Sigma x_{ij})^2}{n} \tag{6.17}$$

$$\sum_j^r \sum_i^{k_j} (x_{ij} - \bar{x}_j)^2 = \Sigma x_{ij}^2 - \sum_j^r \frac{\left(\sum_i^{k_j} x_{ij}\right)^2}{k_j} \tag{6.18}$$

$$\sum_j^r k_j(x_j - \bar{x})^2 = \sum_j^r \frac{\left(\sum_i^{k_j} x_{ij}\right)^2}{k_j} - \frac{(\Sigma x_{ij})^2}{n} \tag{6.19}$$

The general arrangement is illustrated in Table 6.4. The total sum of squares is obtained from the total data. The group sum of squares is obtained from each group total. The error sum of squares is obtained by difference. The F test establishes whether the group variance estimate is significantly larger than the error variance estimate. If it is, the actual variance among groups $\sigma^2(G)$ may be estimated by subtracting the error estimate from the group estimate and dividing by k.

6.8.1 Variance-4 Program: One-Factor Analysis of Variance

The Variance-4 program carries out all the calculations indicated in Table 6.4 with input of the individual data values. The program follows Eq. (6.17) for the total sum of squares and Eq. (6.19) for the factor sum of squares. The error sum of squares is obtained by difference. The flow sequence of Variance-4 is shown in Fig. 6.3.

Table 6.4 Analysis-of-Variance Table for r Groups with k Observations in Each Group

Source	Sum of squares	Degrees of freedom, DF	Variance estimate	Variance estimated
Total	$\Sigma x^2 - \frac{(\Sigma x)^2}{n}$	$n-1$		
Among groups	$\Sigma \frac{(\text{Group total})^2}{\text{No. in group}} - \frac{(\Sigma x)^2}{n}$	$r-1$	$\frac{\text{Group SS}}{\text{Group DF}}$	$\sigma_0^2 + k\sigma^2(G)$
Within groups (error)	Difference	Difference	$\frac{\text{Difference SS}}{\text{Difference DF}}$	σ_0^2

NOTE: F test: Group variance estimate/Error variance estimate at $(r-1)$ and $n(k-1)$ degrees of freedom.

$n = r \cdot k$

$\sigma^2(G)$ = variance among groups

σ_0^2 = error variance

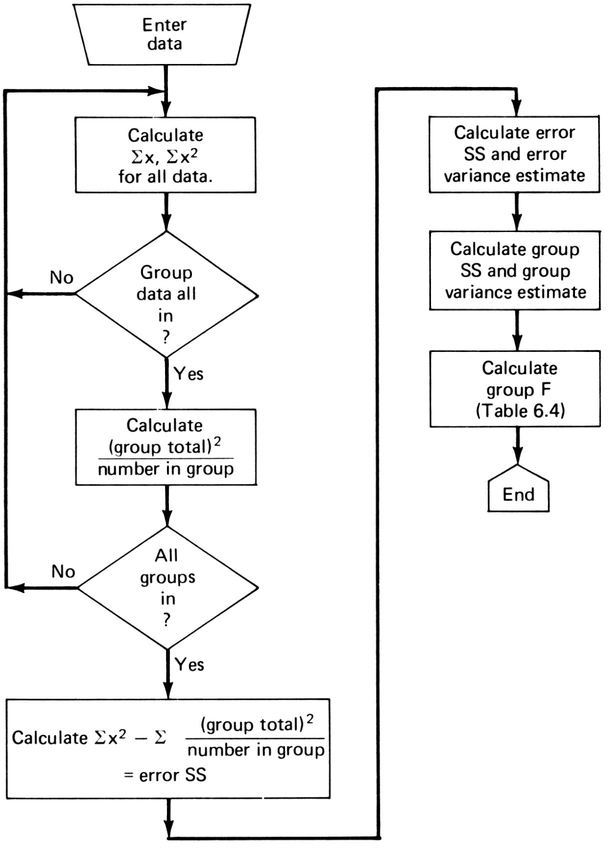


Figure 6.3 Flow diagram for the Variance-4 program: one factor with replication.

The individual data values are entered with KEY A. When one set of data is completed, KEY B is used. The (group total)^2/(number in the group) is calculated and the sum is accumulated. The mean for each group is calculated and printed.

When all the data are entered and the last group has been summed, KEY C is used and the analysis of variance is calculated. The output is each group mean, the group degrees of freedom, the group variance estimate, the error degrees of freedom, and the error variance estimate and the F value equal to the group variance estimate divided by the error variance estimate.

If the probability associated with the F value is desired, program Variance-9 may be run directly after program Variance-4 without any additional data input.

The listing and detailed description of Variance-4 are given in the Appendix.

6.9 TWO-FACTOR ANALYSIS OF VARIANCE, WITHOUT REPLICATION

A two-factor experiment might be illustrated as in Table 6.5 where data are obtained from five reactors, each run once by four different operators. In this case the total variation in the results could include a factor contributed by the different reactors and a factor contributed by the different operators as well as the inherent error variance of the experiment.

The analysis of variance follows a pattern similar to that for the one-factor experiments. The error variance estimate is calculated from the difference between the total sum of squares and the sums of squares for rows and columns. The calculation is illustrated in Table 6.6.

Each of the variance estimates in Table 6.6 may be tested by the F ratio to determine whether either or both of the factors makes a significant contribution to the total variance. The test is the same as for the single-factor analysis. If the F value calculated from the variance estimate ratio is significant, that is, has a probability value less than some preestablished level, usually 0.05 or 0.01, the variance of that factor may be estimated by the formulas in the last column of Table 6.6.

6.9.1 Variance-5 Program: Two-Factor Analysis of Variance

The program Variance-5 carries out the calculations indicated in Table 6.6. The program requires input of the number of rows and then the individual data values by columns. The Hewlett-Packard program as presented can handle up to 13

Table 6.5 Data for Two-Factor Analysis of Variance

	Reactor				
Operator	**R_1**	**R_2**	**R_3**	**R_4**	**R_5**
O_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
O_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
O_3	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
O_4	x_{41}	x_{42}	x_{43}	x_{44}	x_{45}

Table 6.6 Analysis of Variance Table for Two-Factor Experiment

Source	Sum of squares	Degrees of freedom, DF	Variance estimate	Variance estimated
Total	$\Sigma x^2 - \frac{(\Sigma x)^2}{n}$	$n-1$		
Row factor	$\Sigma \frac{(\text{Row total})^2}{\text{No. in row}} - \frac{(\Sigma x)^2}{n}$	$r-1$	$\frac{\text{Row SS}}{\text{Row DF}}$	$\sigma_0^2 + c\sigma^2(R)$
Column factor	$\Sigma \frac{(\text{Column total})^2}{\text{No. in column}} - \frac{(\Sigma x)^2}{n}$	$c-1$	$\frac{\text{Column SS}}{\text{Column DF}}$	$\sigma_0^2 + r\sigma^2(C)$
Error	Difference	Difference	$\frac{\text{Difference SS}}{\text{Difference DF}}$	σ_0^2

NOTE: F test: Group variance estimate/Error variance estimate at group degrees of freedom and error degrees of freedom.

$n = r \cdot c$
r = number of rows
c = number of columns

rows. The Texas Instruments program can handle up to 24 rows without any change in the partitioning between storage and program memory. Additional rows can be handled if the partitioning is changed. Two storage areas are required for each row. There is no limit to the number of columns, and the designation of rows and columns is arbitrary.

The data are entered by columns, also with KEY A. With the data of Table 6.5, the entries would be:

4, KEY A, x_{11}, KEY A, x_{21}, KEY A, . . . , x_{35}, KEY A, x_{45}, KEY A

When all the data are entered, KEY B starts the analysis of variance calculation.

The program accumulates the row totals by assigning a different storage area for each row after the information as to the number of rows has been entered. The total data are counted, and the individual values are summed and squared using the built-in function of KEY $\Sigma+$. Incorrect entries cannot be corrected with KEY $\Sigma-$, however, because the row accumulations would not be corrected.

Figure 6.4 gives a flow diagram of Variance-5.

The program output is the degrees of freedom, the variance estimate, and the F value; first for the row factor and then for the column factor; and then the degrees of freedom and the error variance estimate. The individual row and column means are readily available from the calculator if they are desired. The location of these values is pointed out in the detailed description of the program in the Appendix.

The calculation of the probabilities associated with the F value is not included as part of the program. Variance-9, which calculates the probability, could readily be merged with Variance-5, but it is a simple matter to put Variance-9 in the calculator after Variance-5 has been run, and only the probabilities of interest are calculated.

The listing of Variance-5 and a detailed description are given in the Appendix.

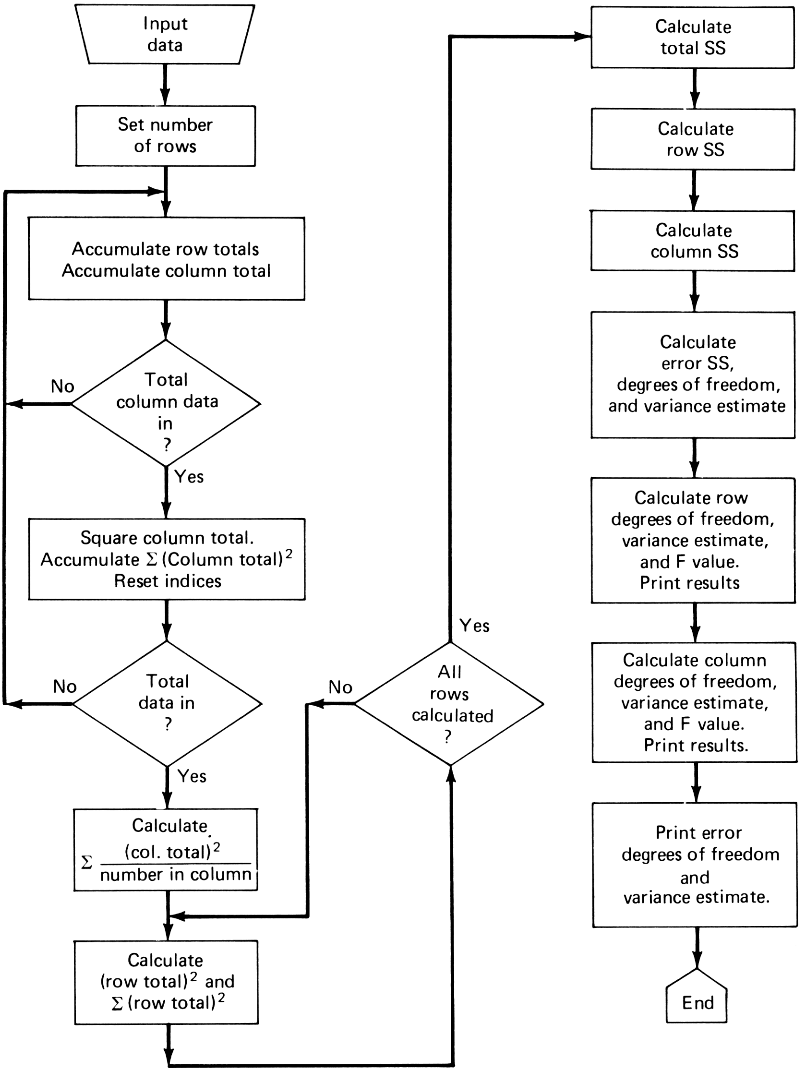


Figure 6.4 Flow diagram for the Variance-5 program: two-factor experiment without replication.

6.10 TWO-FACTOR ANALYSIS OF VARIANCE, WITH REPLICATION

If each operator listed in Table 6.5 had made two or more runs in each reactor, another factor would be added to the analysis of variance. The error variance would be calculated from the agreement within the replicates. The additional factor would be an "interaction" between the operator and the reactor effects. An interaction would occur if the differences among the operators was different with different reactors. If this difference was significant, it would contribute to the total variance and could be found by the analysis of variance.

Table 6.7 Analysis of Variance Table for Two Factors with Replication

Source	Sum of squares	Degrees of freedom, DF	Variance estimate	Variance estimated
Total	$\Sigma x^2 - \frac{(\Sigma x)^2}{n}$ (1)	$n-1$		
Row factor	$\Sigma \frac{(\text{Row total})^2}{ck} - \frac{(\Sigma x)^2}{n}$ (2)	$r-1$	$\frac{\text{Row SS}}{\text{Row DF}}$	$\sigma_0^2 + k\sigma^2(I) + kc\sigma^2(R)$
Column factor	$\Sigma \frac{(\text{Column total})^2}{rk} - \frac{(\Sigma x)^2}{n}$ (3)	$c-1$	$\frac{\text{Column SS}}{\text{Column DF}}$	$\sigma_0^2 + k\sigma^2(I) + kr\sigma^2(C)$
Interaction	Difference: (4) − (2) − (3)	$(c-1)(r-1)$	$\frac{\text{Interaction SS}}{\text{Interaction DF}}$	$\sigma_0^2 + k\sigma^2(I)$
Subtotal	$\Sigma \frac{(\text{Replication total})^2}{k} - \frac{(\Sigma x)^2}{n}$ (4)			
Error	Difference: (1) − (4)	$rc(k-1)$	$\frac{\text{Error SS}}{\text{Error DF}}$	σ_0^2

NOTE: r = number of rows, c = number of columns, k = replicates, $n = r \cdot c \cdot k$.

Table 6.7 illustrates the calculation of an interaction effect in a two-factor experiment with replication.

The F test for significance is first made with the interaction estimate against the error variance estimate. If the interaction effect is significant, then it is used to test the row and column effects. If the interaction is not significant, the row and column effects may be tested against the error variance, or the sums of squares of the error and interaction terms may be added, and their degrees of freedom added, and a new error variance estimate obtained by dividing the total of the two sums of squares by the total of the two degrees of freedom.

The ratio of variance estimates for the F test can be seen from the last column of Table 6.7.

6.10.1 Variance-6 Program: Two-Factor Analysis of Variance with Replication

The calculations of Table 6.7 are carried out by the Variance-6 program with input of the number of rows, the number of replicates, and then the experimental data by columns. To use the program as presented, the number of replicates must be the same in all cases. It would be possible to modify the program to process data with different numbers of replicates, but it would be necessary either to add the number of each set of replicates to the input before the replicates were entered or to indicate the end of the replicate input by a specific KEY. Program Variance-6 applies only to data where the number of replicates is constant.

Figure 6.5 shows a flow diagram for the program. A listing and a line-by-line description are given in the Appendix.

The number of rows and the number of replicates are entered first: Number of rows, r, ENTER ↑, number of replicates, k, KEY A. Then the data are entered by columns, with the replicates entered in sequence, all with KEY A. When all the data are entered, KEY B makes the analysis of variance calculation. The Texas Instruments program uses KEY $x \rightleftharpoons t$ where the Hewlett-Packard program uses ENTER ↑.

The program output follows Table 6.7: row degrees of freedom, row variance estimate, row F value; column degrees of freedom, column variance estimate, column F value. The two F values are calculated by comparison of the row and column variance estimates with the interaction variance estimate.

After the column F value, the output is the interaction degrees of freedom, the interaction variance estimate, and the interaction F value, calculated with reference to the error variance estimate.

Finally the error degrees of freedom and the error variance estimate are printed.

The program stores the initial input, the number of rows r, and the number of replicates k. With each group of k entries, the program sums and squares the total and accumulates the sum of the squares of the replicates. It is therefore essential that the data be entered with the replicates in sequence. The program goes on to sum each group of $r \cdot k$ entries for the column total. At this point in the program, the column mean could be calculated if desired. The program also stores each k set of entries in a different one of r registers to obtain the row

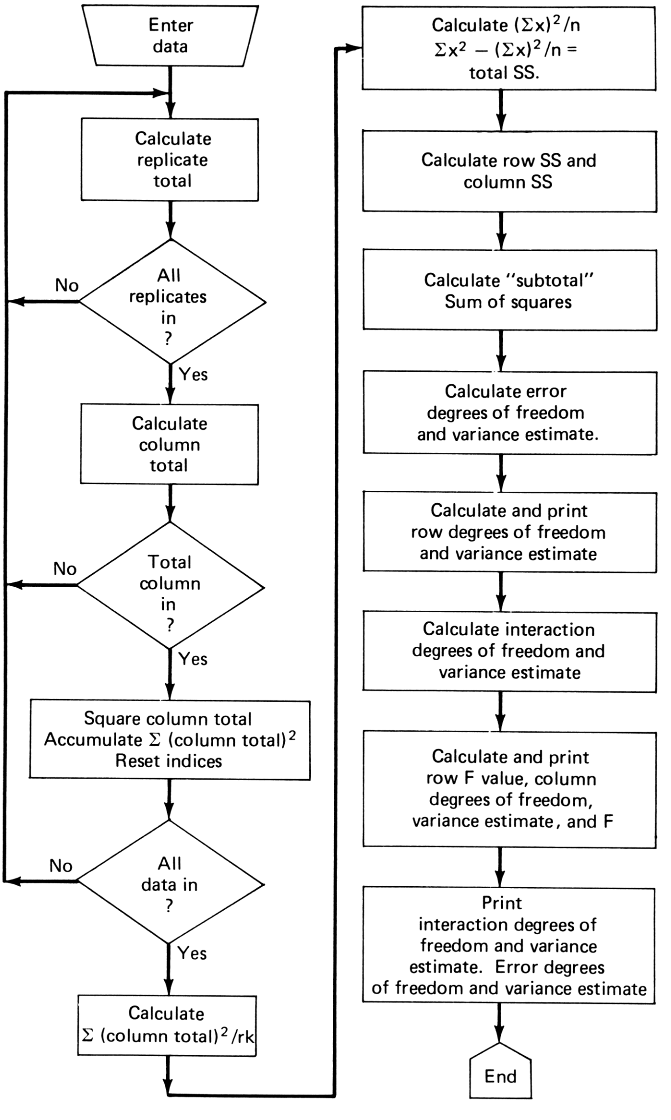


Figure 6.5 Flow diagram for the Variance-6 program: two factors with replication.

totals when all the data have been entered. With the row, column, and replicate totals, the calculation of the analysis follows the equations of Table 6.7.

The Hewlett-Packard program can handle up to 12 rows of data, and the Texas Instruments program can handle up to 23 rows with the normal partition of memory between storage and program. There is no limit to the number of columns, and the row and column designation is arbitrary. Variance-5, without replication, can handle one more row of data than Variance-6. The storage area of one row has been set aside for the replicate totals.

After the analysis of variance results have been printed, if the interaction F test is not significant, the interaction variance estimate and the error variance estimate may be pooled by adding the two sums of squares and dividing by the sum of the degrees of freedom. The error sum of squares and the interaction sum of squares can be obtained by multiplying the respective variance estimates by the corresponding degrees of freedom.

The probability associated with any of the F values may be obtained with the Variance-9 program.

6.11 LATIN SQUARES

A Latin square experimental design is one in which there are an equal number of row and column factor levels, and superimposed on the grid of row and column levels is a third factor with the same number of levels arranged so that each level of each factor is used in one and only one of the experiments. Table 6.8 illustrates a Latin square with five operators running experiments in five different reactors and using five different catalysts in one of the experiments.

The arrangements of the third factor in the grid is usually indicated by letters of the alphabet, hence the name Latin square. It is possible to add a fourth factor, sometimes indicated by letters of the Greek alphabet, and this design is called a Greco-Latin square. The example in Table 6.8 is for a three-factor, five-level Latin square. The program discussed later in this section is for calculating the analysis of variance for a three-factor Latin square experiment. It could readily be expanded to handle more than three factors, as will be obvious from the description.

Inspection of Table 6.8 will show that each reactor is run once by each operator with one of the catalysts. The result is 25 experiments from which the effect of three factors: reactors, operators, and catalysts may be obtained. The 25 experiments provide 24 degrees of freedom for calculation. Each of the factors, at five levels, uses four degrees of freedom, leaving 12 degrees of freedom for error variance esti-

Table 6.8 Latin Square Design of 5 × 5 Experiment

	Reactors				
Operators	**1**	**2**	**3**	**4**	**5**
1	A	B	C	D	E
2	B	E	A	C	D
3	C	D	E	B	A
4	D	A	B	E	C
5	E	C	D	A	B

Five catalysts designated A, B, C, D, and E.

Table 6.9 Analysis of Variance Table for a Latin Square Experiment Size: $k \times k$

Source	Sum of squares	Degrees of freedom, DF	Variance estimate	Estimated variance
Total	$\Sigma x^2 - \frac{(\Sigma x)^2}{n}$	$n-1$		
Row factor	$\Sigma \frac{(\text{Row total})^2}{k} - \frac{(\Sigma x)^2}{n}$	$k-1$	$\frac{\text{Row SS}}{\text{Row DF}}$	$\sigma_0^2 + k\sigma^2(R)$
Column factor	$\Sigma \frac{(\text{Column total})^2}{k} - \frac{(\Sigma x)^2}{n}$	$k-1$	$\frac{\text{Column SS}}{\text{Column DF}}$	$\sigma_0^2 + k\sigma^2(C)$
Letter factor	$\Sigma \frac{(\text{Letter total})^2}{k} - \frac{(\Sigma x)^2}{n}$	$k-1$	$\frac{\text{Letter SS}}{\text{Letter DF}}$	$\sigma_0^2 + k\sigma^2(L)$
Error	Difference	Difference	$\frac{\text{Error SS}}{\text{Error DF}}$	σ_0^2

mation. If an additional factor was imposed on the design it would use four more degrees of freedom, still leaving eight for the error estimate.

Texts dealing with experimental design give tables of Latin squares for various sizes of experiments. The references previously cited have a number of these designs.

The analysis of variance calculation for a Latin square experiment follows the pattern indicated in Tables 6.4, 6.6, and 6.7. The sum of squares for each factor is calculated from the sum of the squares of the factor totals, and the error sum of squares is calculated by the difference between the total and that of each factor. The calculation is shown in Table 6.9.

The F test for significance of each factor variance is made against the error variance as indicated by the last column of Table 6.9. If any of the factors makes a significant contribution to the variance, the variance for that term may be estimated by subtracting the error variance and dividing by the number of levels of measurement k.

The program for the calculator solution of the Latin square analysis of variance requires the data to be entered twice for a three-factor experiment, or three times for a four-factor experiment. It would be possible to modify the Latin square program to make the calculation with one input of the data. However, for the statistical evaluation of the results to be strictly applicable, the selection of the factor levels applied to the different symbols should be randomized. A computer program that needed only one input of data would require the same Latin square arrangement for each test.

6.11.1 Variance-7 Program, Latin Square Analysis of Variance

The calculation of Variance-7 follows Table 6.9. The row and column sums of squares are calculated with the first input of data. The third factor (letters) sum of squares is calculated with a second input of the data. The program contains a check, so if the total of the second input of data is different from the first, an error message appears to alert the operator. The calculation can be continued or a new start may be made at the discretion of the operator.

To make the calculation, the size of the Latin square is entered with KEY A. Then the data by columns, also with KEY A, just as with Variance-5.

After all the data have been entered, they are entered again with KEY B. The second time the data are entered by letter symbol: all the A's, then all the B's, etc., not necessarily in alphabetical order, but all the data represented by one letter must be entered in sequence. The order of entry of the values associated with each letter is of no importance, the calculation is made with the totals for each letter symbol.

When all the data have been entered a second time, the program automatically completes the calculation. The sum of the second entry of data is compared with the first. If they are different, the program prints the two sums and stops. The operator has a choice at this point of either continuing the calculation, if the difference is small, or starting over.

When the calculation is completed, the output is the degrees of freedom, the variance estimate, and the F value for each factor; first rows, then columns, then letters. Finally, the calculator gives the error degrees of freedom and the error variance estimate. If a probability value is desired for all or any of the F values, it can be obtained with the Variance-9 program. If by coincidence the error variance is zero so that the F values cannot be calculated, the program omits the F calculation but prints all the other results.

A flow diagram of Variance-7 is shown in Fig. 6.6. The Hewlett-Packard program can handle Latin squares up to size 13 by 13. The Texas Instruments program can handle them up to 24 by 24, with the normal partition, and larger squares if the partition is changed. If an additional factor beyond the letter factor is to be included, for a Greco-Latin square, or multiple-balanced-block experiment, the program routine starting with KEY B would have to be duplicated, and one storage area would have to be assigned to the additional factor sums. The additional factors would require additional input of the data.

6.12 2^n FACTORIAL DESIGN

The 2^n experimental design is a special case of an experiment to test the effect of n factors, each at two levels. The complete design requires testing all of the factors at each level of all of the others, making a total of 2^n experiments. When there are a large number of factors involved, the testing at two levels of each provides a method of eliminating unimportant factors with a minimum of experimentation.

There is a large literature on 2^n factorial design, and a number of ways of dealing with fractions of complete 2^n factorials are given that permit the determination of the main effects with a minimum loss of information. The analysis of variance of a 2^n factorial is handled in a different manner from that described in the preceding sections dealing with experiments done at more than two levels of the factors. The handling of data from a 2^n experiment also involves a different

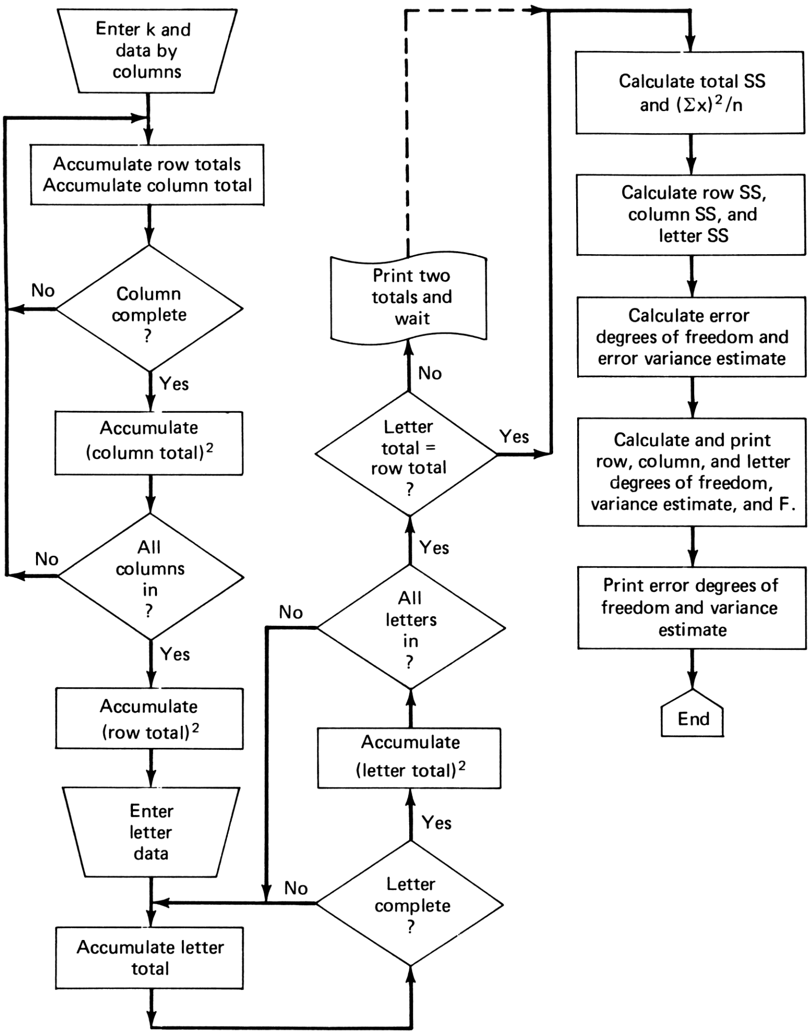


Figure 6.6 Flow diagram for the Variance-7 program: Latin squares.

nomenclature from that used for the analysis of variance of the other experimental designs presented.

This text provides a calculator program for the solution of a 2^n experiment and gives sufficient discussion of the nomenclature so that the program may be applied to data from such a design. No attempt is made to go into the many ramifications of the design, but it is suggested that engineers involved in experimental design become familiar with the powerful capability of the 2^n design. The program presented, Variance-8, will handle data from four, eight, or sixteen experiments; and the procedure can be expanded to handle larger designs. With a knowledge of the nomenclature, it will be seen that the same program will handle fractions of full 2^n factorials with equal facility.

6.12.1 2^n Factorial Nomenclature

Uppercase (capital) letters are used to designate the factors being tested: temperature, pressure, catalyst concentration, etc. Uppercase letters are also used to designate the particular effects being determined. A is the effect of the A factor, i.e., the difference between doing the experiment at the upper level and doing it at the lower level of A. AB designates the interaction of the A and B factors, i.e., the difference in the results between experiments with both A and B at their higher level or both at the lower level and experiments with one at its high level and the other at its low level.

Lowercase letters, $a, b, c, ab, \ldots$, are used to designate both the results obtained and the experiment to be run. For example, ab in an experiment involving factors A, B, C, and D designates the experiment with A and B at their upper level and factors C and D at their lower level. It also designates the result obtained under those conditions; c designates the experiment done with the C factor at its upper level and all the others at the lower level. The symbol (1) designates the experiment done at the low level of all factors.

A complete factorial involving four factors at two levels each, a 2^4 factorial, would have 16 experiments: (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, $abcd$.

The A effect, designated A, is the difference between all the experiments run at the higher level of A and all those at the lower level. In a 2^3 factorial this would be:

$$A = (a + ab + ac + abc) - ((1) + b + c + bc) \tag{6.20}$$

For the AC interaction, the AC effect would be the difference between the experiments run with A and C at the same levels and the experiments with A and C at different levels:

$$AC = ((1) + b + ac + abc) - (a + c + ab + bc) \tag{6.21}$$

Each of the effects is calculated from a difference involving all of the experiments.

The third-order interaction ABC would be calculated from the difference between experiments run at the high level of two factors and the low level of the third, and the experiments run at the low level of the same two factors and the high level of the third plus the difference between the experiments run at the high level of all and the low level of all.

$$ABC = (a + b + c + abc) - ((1) + ab + ac + bc) \tag{6.22}$$

6.12.2 2^n Factorial Calculation

The calculation of all of the effects can be obtained by a simple process of addition and subtraction if the experimental results are arranged in the proper order. Table 6.10 illustrates the calculation. The results are listed in the order: (1), a, b, ab, c, ac, bc, abc, . . . (If there is a D factor, the next results in order would be d times each of the previous terms: d, ad, bd, abd, . . . $abcd$. If there were an E factor, the next terms would be all the previous ones multiplied by e.) The tabulated

Table 6.10 Calculation of a 2^3 Factorial Experiment

Response	Calculation 1	Calculation 2	Calculation 3
(1)	$(1)+a$	$(1)+a+b+ab$	$(1)+a+b+ab+c+ac+bc+abc$
a	$b+ab$	$c+ac+bc+abc$	$a-(1)+ab-b+ac-c+abc-bc$
b	$c+ac$	$a-(1)+ab-b$	$b+ab-(1)-a+bc+abc-c-ac$
ab	$bc+abc$	$ac-c+abc-bc$	$ab-b-a+(1)+abc-bc-ac+c$
c	$a-(1)$	$b+ab-(1)-a$	$c+ac+bc+abc-(1)-a-b-ab$
ac	$ab-b$	$bc+abc-c-ac$	$ac-c+abc-bc-a+(1)-ab+b$
bc	$ac-c$	$ab-b-a+(1)$	$bc+abc-c-ac-b-ab+(1)+a$
abc	$abc-bc$	$abc-bc-ac+c$	$abc-bc-ac+c-ab+b+a-(1)$

results are then added in pairs and subtracted, the first from the second, in pairs; and the operation repeated n times for a 2^n experiment. The last column will be the effects indicated by the letters designated in the order of the data.

In Table 6.10, the last column, calculation 3, gives the total effects in the following order: Total of all the data, the A effect, the B effect, the AB interaction, the C effect, the AC interaction, the BC interaction, and the ABC triple interaction. These correspond to the order in which the responses are listed in the first column. If the interaction effects are not large, or in a high-order experiment where the interpretation of a three- or four-factor interaction might be difficult to evaluate, the interactions are taken as a basis for the error variance calculation.

When there are only two pieces of data the sum of squares of deviation from the mean is equal to half the square of the difference between the values:

$$\Sigma(x_i - \bar{x})^2 = \frac{(x_1 - x_2)^2}{2} \qquad (6.23)$$

when $\bar{x} = \dfrac{x_1 + x_2}{2}$

With the 2^n factorial calculation illustrated in Table 6.10, the last column is the difference between observations at two levels. The square of this difference divided by 2^n is the sum of squares and the variance estimate since each calculation has only one degree of freedom. For example, the second line in the last column of Table 6.10 is the calculation for the difference between four experiments run at the upper level of A and four run at the lower level. This difference squared divided by 8 is the A sum of squares and is the A variance estimate. The significance of the A variance estimate would be tested by the F test against an error variance estimate or one of the interaction variance estimates following the example of Table 6.7.

The actual A effect, the difference between operating at the upper and lower levels of A, is the value in the last column divided by 2^{n-1}. The difference is the sum of 2^{n-1} experiments made at the upper and lower level and 2^{n-1} made at the lower level. The mean difference is the total difference divided by 2^{n-1}.

The 2^n calculator program, Variance-8, calculates both the variance estimate and the mean difference for all of the factors.

6.12.3 2^n Factorial Confounding

One valuable feature of the 2^n factorial design is that the main effect of more than n factors may be obtained with 2^n experiments with suitable arrangement of the design. The method is generally known as "confounding," in that the main effects are confounded with high-order interactions which are assumed to be not significant. There is a large literature on 2^n factorial experiments. In addition to the references already cited, there is the National Bureau of Standards Handbook 91 on Experimental Statistics.[8]

As a simple illustration, refer to Table 6.10. The last row in column 4 shows the calculation for the third-order interaction, the *ABC* effect. The *abc, c, b,* and *a* results are added and the *bc, ac, ab,* and (1) results are subtracted to find the *ABC* interaction effect. If a fourth factor, *D,* were to be tested with the same experiment, the experiment would be run at the upper level of the *D* factor in all the experiments having a positive sign in the calculation of the *ABC* interaction: *abc, c, b,* and *a,* giving experiments $abc(d)$, $c(d)$, $b(d)$, and $a(d)$. *D* would be run at its lower level for the other four experiments, *bc, ac, ab,* and (1), all of which have a negative sign in the calculation of the *ABC* interaction. The experiment would still consist of only 2^3 runs, but 4 factors could be tested for their main effects. The *D* factor would be confounded with the *ABC* interaction, and each of the other factors, *A, B,* and *C* would be confounded with other third-order interactions.

The calculation of the effects and sums of squares is carried out in the same manner as indicated in Table 6.10 and provided for in the Variance-8 program. The interpretation of the results would depend on the confounding involved. The 2^n literature provides numerous tables, charts, and formulas for determining the experiments to be run to obtain specific confounding. To use program Variance-8 for the calculation of a confounded experiment it is only necessary to have the data in the order shown in the first column of Table 6.10 for 4, 8, or 16 experiments.

6.12.4 Variance-8 Program: 2^n Factorial

The flow diagram for Variance-8 is given in Fig. 6.7. The calculation follows Table 6.10. The input to the program is the value of 2^n, the number of experiments involved, 4, 8, or 16, with KEY A. This value establishes the number of times the sums and differences will be calculated and also sets the number of data items that will be entered.

The data are entered, also with KEY A, in the order indicated in Table 6.10, starting with the value for experiment (1). When all the data have been entered, the calculation will start without further input inasmuch as the amount of data has been put into the calculator with the first entry of 2^n.

The output is the variance estimate for each term starting with *A*. There is

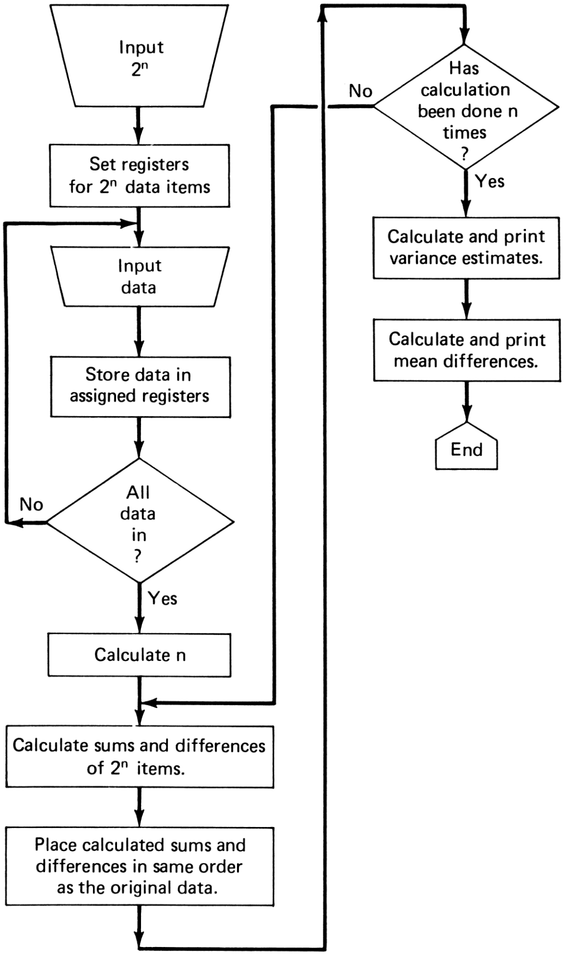


Figure 6.7 Flow diagram of the Variance-8 program: 2^n analysis of variance.

no variance estimate associated with the first entry (1). The variance estimate is calculated by squaring the differences calculated in column 3 of Table 6.10 and dividing by 2^n. Each variance estimate will have one degree of freedom.

After the variance estimate output, the program gives the mean difference for each factor, again starting with *A* and ending with *ABC* of Table 6.10. If the program is calculating a 2^4 experiment, it will continue on beyond *ABC* to give the variance estimate and mean difference for all terms. The mean difference is calculated from the total difference divided by 2^{n-1}. This value is the mean effect of the difference between operating at the two levels of the factor involved.

The variance estimates may be tested for significance by the *F* test. Each variance estimate has one degree of freedom, which will apply to the numerator of the *F* test ratio. The degrees of freedom for the denominator will depend on the number of interaction terms that are pooled to give an error variance estimate. There will be one degree of freedom for each.

The F values at 1 and ν degrees of freedom are the same as t^2 where t is calculated for ν degrees of freedom and the same probability level as F. The significance of any factor effect may be determined either by the F test or by using the t relation of Eq. (4.6) for the difference between two means. The mean difference, calculated from the third column in Table 6.10 divided by 2^{n-1}, is the difference between the mean of the experiments at the high level of the factor and the mean of the experiments at the low level. If this difference is significant, it will have a significant t value when divided by $\bar{s}(x)\sqrt{1/n_1 + 1/n_2}$. But $\sqrt{1/n_1 + 1/n_2}$ is equal to $\sqrt{1/2^{n-2}}$ for the 2^n factorial. Therefore a t test, equivalent to the F test, may be made by calculating t from the mean difference by the following equation:

$$t = (\text{mean difference}) \sqrt{\frac{2^{n-2}}{\text{Error variance estimate}}} \tag{6.24}$$

The engineer may choose whichever test is easier to run. The t program, t-5, is about half the length of the F program, Variance-9, and might more readily be merged with an analysis-of-variance program. If the probability is to be calculated separately after the analysis of variance program is run, Variance-9 is easier inasmuch as no change need be made in the calculations. If the t-5 program is used, the square root of the error variance times $1/\sqrt{2^{n-2}}$ needs to be calculated. However, this calculation can be readily added to the program.

Details and a listing of the 2^n Variance-8 program are given in the Appendix. A summary of all the analysis-of-variance programs is given in Table 6.11.

6.13 EXAMPLES OF ANALYSIS-OF-VARIANCE CALCULATOR PROGRAMS

6.13.1 The data in Table 6.12 show the weight gain for five litters of animal young, each being given a different diet. Is there significant difference among the different diets?

The Variance-4 program is used for testing the variance among groups with one factor being measured. If the F test is significant, it will indicate that there is more difference among the means of the different diets than there is within the litters.

With Variance-4 in the calculator the data are entered: 2.77, KEY A, 2.58, KEY A, . . . , 2.94, KEY A, 2.28, KEY A. With all the data from column one, KEY B. 3.01, KEY A, 2.22, KEY A, . . . , 2.71, KEY A, KEY B, . . . , 2.80, KEY A, 2.46, KEY A, KEY B. With all the data entered, after KEY B following the end of the last set, KEY C.

The output is: 2.590, 2.605, 2.524, 2.512, 3.073—the means for each diet. This output follows immediately after KEY B for each group. 4 (the degrees of freedom for the variance estimate), 0.361 (the variance estimate for the difference among

Table 6.11 Summary of Analysis-of-Variance Programs

Program				
Variance-4: 1 factor with replication	**Variance-5: 2 factors, no replication**	**Variance-6: 2 factors with replication**	**Variance-7: Latin square**	**Variance-8: 2^n factorial**
x_i (by columns), KEY A	No. of rows, KEY A	No. of rows, ENTER ↑, (KEY A in TI)	No. of rows, KEY A	2^n, KEY A
At end of column, KEY B	x_i (by columns), KEY A	No. of replicates, KEY A	x_i (by columns), KEY A	x_i, KEY A
At end of all data, KEY C	At end of all data, KEY B	x_i (replicates, by columns), KEY A	x_i (second input), KEY B	
		At end of all data, KEY B		
OUTPUT				
Factor means Factor DF Factor variance estimate Factor F value Error DF Error variance estimate	Row DF Row variance estimate Row F value Column DF Column variance estimate Column F value Error DF Error variance estimate	Row DF Row variance estimate Row F value Column DF Column variance estimate Column F value $R \times C$ DF $R \times C$ variance estimate $R \times C$ F value Error DF Error variance estimate	Row DF Row variance estimate Row F value Column DF Column variance estimate Column F value Letter DF Letter variance estimate Letter F value Error DF Error variance estimate	Variance estimates in order of input Mean difference in order of input

LIMITS				
HP: None	13 rows	12 rows	13×13 square	2^4
TI: None	24 rows	23 rows	24×24 square	2^4

NOTE: DF = degrees of freedom, R = row factor, C =column factor.

Table 6.12 Weight Gain for Animal Young

Diet				
1	2	3	4	5
2.77	3.01	2.74	2.74	3.28
2.58	2.22	2.87	2.50	3.52
2.38	2.61	2.46	2.56	2.88
2.94	2.36	2.31	2.48	3.09
2.28	2.72	2.24	2.32	3.48
	2.71		2.47	2.80
				2.46

diets), 24 (the degrees of freedom for the error variance), 0.0825 (the error variance estimate), and 4.385 (the F value). (Time:8/4)

If the probability value is desired, program Variance-9 may be run:

> In the Hewlett-Packard calculators: 4, ENTER ↑, 24, ENTER ↑, 4.385, KEY A.
> In the Texas Instruments calculators: 4, KEY A, 24, KEY A, 4.385, KEY A.

The output in both cases is 0.0084, the probability associated with an F value of 4.385 at 4 and 24 degrees of freedom. There is significant difference among the diets. There would be less than one chance in a hundred of being in error if the hypothesis of equal variances was rejected.

6.13.2 The data in Table 6.13 are from a series of 25 runs made at five different temperatures and for five different durations to measure a reaction. Is there a significant effect of either or both temperature and time on the conversion, measured in milligrams per gram?

The Variance-5 program is used for a two-factor experiment without replications. The data are entered: 5 (the number of rows), KEY A, and then the data by

Table 6.13 Measurement of Reaction, Milligrams per Gram

Temperature, °F	Time, minutes 30	60	90	120	150
100	16	40	50	20	15
125	30	25	62	67	30
150	50	50	83	85	45
175	80	80	95	98	70
200	90	92	98	100	88

columns: 16, KEY A, 30, KEY A, 50, KEY A, . . . , 45, KEY A, 70, KEY A, 88, KEY A. When all the data have been entered, KEY B.

The output is:

4.0	Row, temperature, degrees of freedom
3775.	Row variance estimate
37.6	Row F value
4.0	Column, duration, degrees of freedom
799.	Column variance estimate
7.97	Column F value
16.0	Error degrees of freedom
100.2	Error variance estimate

(Time:14/12)

Both the F values are so large that there is no need to find their probability values. (If Variance-9 were run, the values would be 6×18^{-8} for rows and 0.001 for columns.) Both factors are highly significant.

6.13.3 The data in Table 6.14 give the percent moisture absorbed by water-repellent cottons tested by four different laundries under four different test conditions. Each test was run in duplicate. Is there a difference among the laundries, and are there differences among the tests?

Variance-6 program is used for a two-factor experiment with replication. The data are entered: 4, the number of rows, ENTER ↑, in the Hewlett-Packard (or KEY A, in the Texas Instruments), 2, the number of replicates, KEY A. And then the data by columns: 7.20, KEY A, 9.06, KEY A, 11.7, KEY A, . . . , 5.27, KEY A, 2.74, KEY A, 2.31, KEY A. When all the data have been entered, KEY B.

Table 6.14 Percent Moisture Absorbed by Water-Repellent Cottons

	Laundry			
Test	**A**	**B**	**C**	**D**
1	7.20	2.40	2.19	1.22
	9.06	2.14	2.69	2.43
2	11.70	7.76	4.92	2.62
	11.79	7.76	1.86	3.90
3	15.12	6.13	5.34	5.50
	14.38	6.89	4.88	5.27
4	8.10	2.64	2.47	2.74
	8.12	3.17	1.86	2.31

The output is:

3.	Row, tests, degrees of freedom
34.3	Row variance estimate
11.5	Row F value
3.	Column, laundry, degrees of freedom
99.4	Column variance estimate
33.3	Column F value
9.	Interaction degrees of freedom
2.98	Interaction variance estimate
5.16	Interaction F value
16.	Error degrees of freedom
0.578	Error variance estimate

(Time:18/22)

All of the factors are significant. The interaction effect indicates that the different laundries gave different results for the different tests. The high laundry and test F values indicate that even accounting for the different results that the laundries obtained with the different tests, there were significant differences among both laundries and tests.

6.13.4 For an example of the Latin square analysis of variance with program Variance-7, the data used for Example 6.13.2 will be tested: see Table 6.13. If it had been found that five different reactors had been used in the time-temperature study and that the reactors, designated A, B, C, D, and E, had been employed so that each reactor was used only once at each temperature and pressure in the following way:

$$\begin{matrix} A & B & C & D & E \\ B & C & D & E & A \\ C & D & E & A & B \\ D & E & A & B & C \\ E & A & B & C & D \end{matrix}$$

the configuration is seen to constitute a Latin square experimental design.

With Variance-7 in the calculator, the data from Example 6.13.2 are entered, first the number of rows, then the data by columns, all with KEY A. The data are then entered a second time by letters with KEY B.

5, KEY A, 16, KEY A, 30, KEY A, . . . , 70, KEY A, 88, KEY A,

completing the first entry of the data. The data are then reentered by letters: 16, KEY B, 92, KEY B, 95, KEY B, . . . , 30, KEY B for the letter A. 30, KEY B, 40, KEY B, 98, KEY B, . . . , 45, KEY B for the letter B, . . . , 90, KEY B, 80, KEY B, . . . , 15, KEY B . . . , completing the entry of the letter E data and the second entry of all of the data. When all of the data are entered the second time, the program goes immediately into the analysis-of-variance calculation. The output is:

4.0	Degrees of freedom for the row factor	
3775.	Row variance estimate	
32.3	Row F value	
4.0	Degrees of freedom for column factor	
798.8	Column variance estimate	
6.83	Column F value	
4.0	Degrees of freedom for the letter factor	
49.9	Letter factor variance estimate	
0.427	Letter F value	
12.0	Degrees of freedom for error	
117.0	Error variance estimate	(Time:20/23)

The row and column factors, the temperature and time, are significant, but the letter factor, the reactors, is not significant. The probability values for the F terms could be obtained from Variance-9, but the F for the letter factor is less than one and is obviously not significant, and the other two values, 32.3 and 6.83, will have probability values less than 0.01.

If the input of the letter values had not equaled the input of the row and column data, the program would have printed the two totals and stopped. The operator would then have a choice of continuing (the R/S KEY), or reentering the data. The program clears all flags when it stops, so it is ready for new data.

6.13.5 The example of a 2^n factorial experiment is for some electric furnace tests. The factors tested are a low and high furnace roof, a 750- and a 1000-lb furnace scrap charge, use of plate scrap and tubular scrap, and low-power input rate and high-power input rate. The response is the number of kilowatt hours per ton of melted product. The configuration of the design and the results are shown in Table 6.15. Alongside each result is the letter symbol for the response, based on the designations in the table footnote.

The results are entered in the program Variance-8 in some order conforming to the first column of Table 6.10. The data from Table 6.15 could be entered by rows or by columns. Either way would conform to the standard order. The calculated results are in the same order as the input. If the data are entered by rows, the first three values of the output would refer to the power effect, the roof effect, and the power by roof interaction. If the data were entered by columns, the first three calculated results would refer to the charge effect, the scrap effect, and the

Table 6.15 Results of 2^n Factorial Experiment

Furnace charge, lb	Low roof: Low-power input	Low roof: High-power input	High roof: Low-power input	High roof: High-power input
TUBE SCRAP				
700	715 (1)	862 (p)	785 (r)	843 (pr)
100	693 (c)	939 (cp)	670 (cr)	815 (cpr)
PLATE SCRAP				
700	775 (s)	970 (sp)	790 (sr)	987 (spr)
1000	727 (cs)	945 (csp)	741 (csr)	866 ($cspr$)

NOTE: R for roof factor — r designates high roof
C for charge factor — c designates 1000-lb charge
S for scrap factor — s designates plate scrap
P for power factor — p designates high-power input

charge by scrap interaction. Either way, the calculation would be correct and the final answers the same.

In this example the data are entered by columns: first the value of 2^n, 16, KEY A, and then the data: 715, KEY A, 693, KEY A, 775, KEY A, . . . , 815, KEY A, 987, KEY A, 866, KEY A. When all the data have been entered the calculator will start immediately on the calculation.

There are two groups of output, first the variance estimates and then the mean differences. The mean differences are calculated by subtracting the mean of the responses at the lower levels from the mean of the responses at the upper levels. A negative result indicates that the responses at the lower level are greater than the responses at the upper level.

From the data of Table 6.15 the output is:

Variance estimates		Mean differences
6848.	(C factor)	−41.38
14,340.	(S factor)	59.88
1502.	($C \times S$ interaction)	−19.38
110,723.	(P factor)	166.38
1173.	($C \times P$ interaction)	17.13
1208.	($S \times P$ interaction)	17.38
3452.	(CSP interaction)	−29.38
1040.	(R factor)	−16.13
5439.	($C \times R$ interaction)	−36.88
248.	($S \times R$ interaction)	7.88
638.	(CSR interaction)	12.63
4935.	($P \times R$ interaction)	−35.12
716.	(CPR interaction)	−13.38
613.	(SPR interaction)	12.38
431.	($CSPR$ interaction)	−10.38

The five values of variance estimate less than 1000 can be pooled to give another estimate of 529 with five degrees of freedom which can be used as an error variance estimate. With the pooled error variance estimate of 529, the variance estimates of the other factors may be tested by the F ratio. The C factor, the S factor, the P factor, and the $C \times R$ and the $P \times R$ interactions are significant. None of the others reach the 0.05 level of probability using program Variance-9.

The conclusions from the analysis are that the larger charge requires significantly less power per ton, the tube scrap requires less power per ton, and the lower power input requires less power per ton. The effects of higher charge and lower power input are reversed with the change from a high roof to a low roof, but the use of a high roof or low roof has no significant effect by itself.

6.14 THE F FUNCTION

The F function describes the distribution of the ratio of variance estimates calculated from samples drawn from normally distributed populations. The distribution depends on the number of degrees of freedom with which the variance estimates are calculated. If the variance estimates are calculated from samples from the same population, the ratio would be expected to fluctuate around the value of one. The fluctuation would depend on the size of the samples. The F function is a mathematical description of the variation of this ratio.

The F distribution is used similarly to the t distribution for means. A hypothesis is made that the variance estimates are estimates of the same variance. Their ratio should be unity. If a ratio is found that has a low probability of occurring if in fact the variances are the same, the hypothesis is rejected. The probability associated with the F value is the probability of a false rejection of the hypothesis.

The common practice is to look in an F table at the degrees of freedom of the numerator and denominator used in the F calculation for the F function value. If the calculated F exceeds the tabulated F value at the probability of the table, the hypothesis is rejected. Usually only 0.05 and 0.01 probability tables of the F function are available.

With the F probability program, Variance-9, the degrees of freedom and the F value are entered in the calculator. The program calculates the probability value. The degrees of freedom associated with the numerator must be entered first, and the degrees of freedom for the denominator of the F calculation entered second. The probability is not the same if the degrees of freedom are reversed.

It was mentioned in the text that the F distribution is related to both the χ^2 and the t distributions. F at one and ν degrees of freedom is equal to t^2 at ν degrees of freedom. And F at ν and infinite degrees of freedom equals χ^2/ν at the same probability level. The Variance-9 program cannot handle problems involving infinite degrees of freedom, but the correct probability can be obtained from the χ^2-5 program with a small additional calculation. The t-3 program is slightly faster than the F probability program so that in some cases it might be preferred for calculations where one degree of freedom is involved in the F calculation.

6.14.1 The F Probability Program, Variance-9

The input to the program is the numerator degrees of freedom, the denominator degrees of freedom, and the F value.

Data	HP	TI
ν_1	ENTER ↑	KEY A
ν_2	ENTER ↑	KEY A
F	KEY A	KEY A

The output is the probability value, which is the probability of an F as large as that observed if the variance estimates were made from the same population with samples having the degrees of freedom used in the calculation.

At the end of the program, in the Hewlett-Packard calculators, the input values are in Registers A, B, and C in the order of input. In the Texas Instrument calculators the values are in Registers 01, 02, and 03.

There is no restriction to the input for the F probability calculation, but there is a difference in calculation time when both degrees of freedom are odd numbers and when at least one of them is an even number. The difference in calculator time is only significant when the degrees of freedom are greater than about 10.

The calculation of the F probability is based on equations from the National Bureau of Standards handbook previously cited.[2] The equations are shown in Table 6.16. Equations (6.25) to (6.31) are those given in the reference. The square-bracket sections of the original equations are modified for use in the calculator, and the modifications are shown in Table 6.17. The modifications give identical results to those from the originals but permit the use of the same summation subroutines for all combinations of degrees of freedom input. The program flow is shown in Fig. 6.8.

Table 6.16 Equations for F Probability Calculation

ν_1 even

$$\Pr(\geqq F) = x^{\nu_2/2}\left[1 + \frac{\nu_2}{2}(1-x) + \frac{\nu_2(\nu_2+2)}{2\cdot 4}(1-x)^2 + \cdots + \frac{\nu_2(\nu_2+2)\cdots(\nu_2+\nu_1-4)}{2\cdot 4\cdots(\nu_1-2)}(1-x)^{(\nu_1-2)/2}\right] \tag{6.25}$$

ν_2 even

$$\Pr(\geqq F) = 1-(1-x)^{\nu_1/2}\left[1 + \frac{\nu_1}{2}(x) + \frac{\nu_1(\nu_1+2)}{2\cdot 4}x^2 + \cdots + \frac{\nu_1(\nu_1+2)\cdots(\nu_1+\nu_2-4)}{2\cdot 4\cdots(\nu_2-2)}x^{(\nu_2-2)/2}\right] \tag{6.26}$$

$$x = \frac{\nu_2}{\nu_2 + \nu_1 F} \tag{6.27}$$

ν_1, ν_2 odd

$$\Pr(\geqq F) = 1 - A + B \tag{6.28}$$

$$A = \begin{cases} \dfrac{2\theta}{\pi} \text{ for } \nu_2 = 1 \\[2ex] \dfrac{2}{\pi}\left[\theta + \sin\theta\left(\cos\theta + \dfrac{2}{3}\cos^3\theta + \cdots + \dfrac{2\cdot 4\cdots(\nu_2-3)}{3\cdot 5\cdots(\nu_2-2)}\cos^{\nu_2-2}\theta\right)\right] \end{cases} \tag{6.29}$$

$$B = \begin{cases} 0 \text{ for } \nu_1 = 1 \\[2ex] \dfrac{2}{\sqrt{\pi}}\dfrac{\left(\frac{\nu_2-1}{2}\right)!}{\left(\frac{\nu_2-2}{2}\right)!}\sin\theta\cos^{\nu_2}\theta\left[1 + \dfrac{\nu_2+1}{3}\sin^2 + \cdots + \dfrac{(\nu_2+1)(\nu_2+3)\cdots(\nu_2+\nu_1-4)}{3\cdot 5\cdots(\nu_1-2)}\sin^{\nu_1-3}\theta\right] \end{cases} \tag{6.30}$$

$$\theta = \arctan\sqrt{\frac{\nu_1}{\nu_2}F} \tag{6.31}$$

NOTE: ν_1 is the number of degrees of freedom for the numerator variance estimate.
ν_2 is the number of degrees of freedom for the denominator variance estimate.

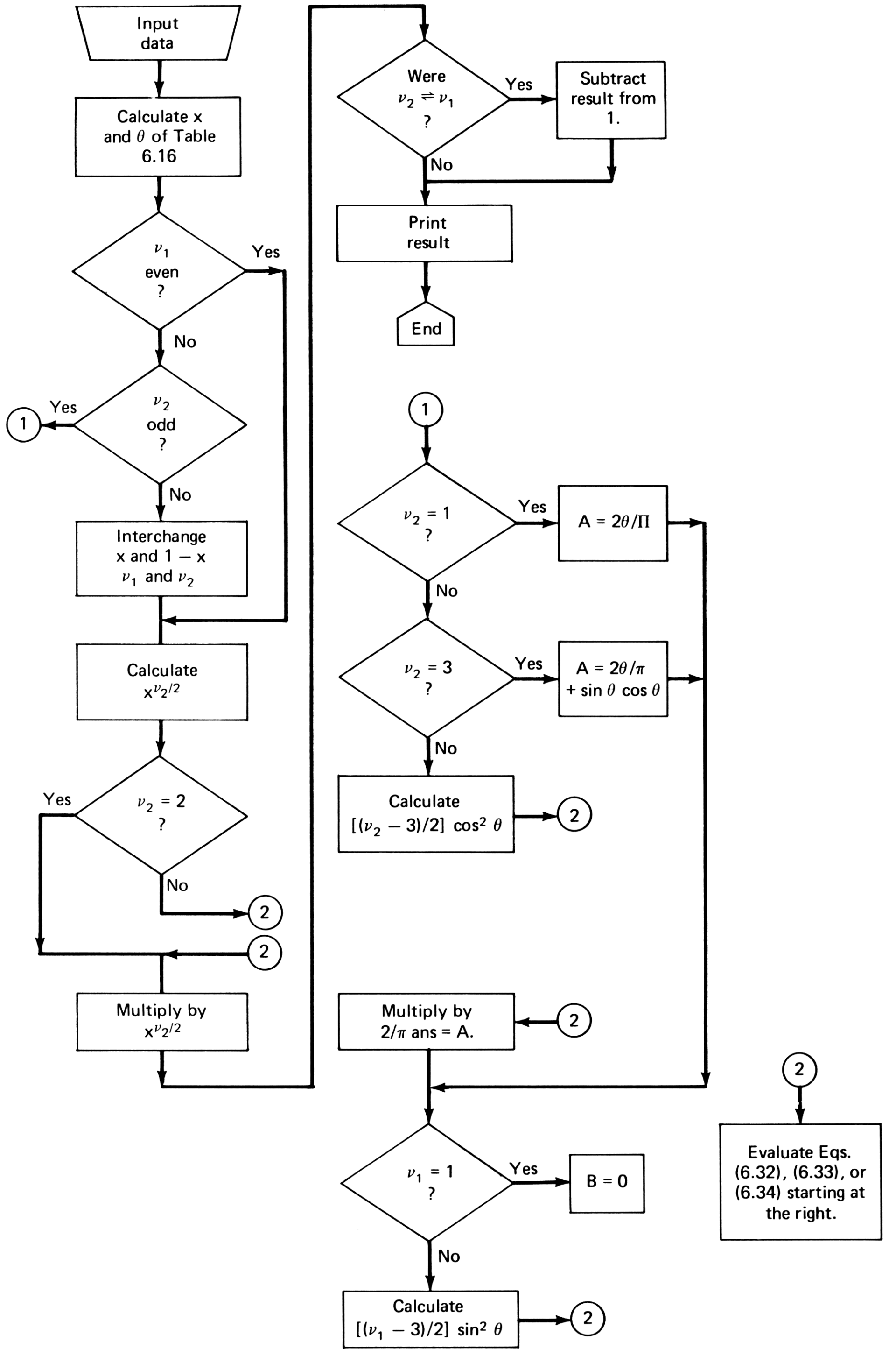


Figure 6.8 Flow diagram for the Variance-9 program: *F* factor probability.

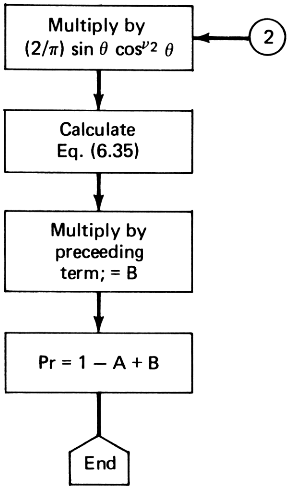


Figure 6.8 *(Cont.)*

Equation (6.30) has a term: $((\nu_2 - 2)/2)!$ which, when ν_2 is odd involves a fractional factorial calculation. Equation (6.35) in Table 6.17 shows an equivalent calculation that gives the same result. The form in Table 6.17 is used in the program Variance-9 to eliminate the necessity of calculating the fractional factorial.

After the input of the two degrees of freedom and the F value, the program tests the degrees of freedom to determine whether they are even or odd. If ν_1 is even, the program uses Eq. (6.25) with the modification of the square-bracket portion. If ν_1 is odd and ν_2 is even, the program uses the same equation but substitutes ν_1 for ν_2 and x for $(1 - x)$. The result is one minus the desired answer, and Flag 1 is used to indicate which result is correct.

If neither ν_1 nor ν_2 is even, the calculation is made with Eq. (6.28) with the

Table 6.17 Square-Bracket Equivalents for Equations in Table 6.16

$[\nu_1 \text{ even}]$
$$\left[1 + \frac{\nu_2}{2}(1-x)\left(1 + \frac{\nu_2+2}{4}(1-x)\left(1 + \cdots \left(1 + \frac{\nu_2+\nu_1-4}{(\nu_1-2)}(1-x)\right)\cdots\right)\right)\right] \tag{6.32}$$

$[A]$
$$\left[\theta + \sin\theta\cos\theta\left(1 + \frac{2}{3}\cos^2\theta\left(1 + \cdots\left(+\frac{\nu_2-3}{\nu_2-2}\cos^2\theta\right)\cdots\right)\right)\right] \tag{6.33}$$

$[B]$
$$\left[1 + \frac{\nu_2+1}{3}\sin^2\theta\left(1 + \frac{\nu_2+3}{5}\sin^2\theta\left(1 + \cdots\left(1 + \frac{\nu_2+\nu_1-4}{\nu_1-2}\sin^2\theta\right)\cdots\right)\right)\right] \tag{6.34}$$

$$\frac{2}{\sqrt{\pi}}\,\frac{\left(\frac{\nu_2-1}{2}\right)!}{\left(\frac{\nu_2-2}{2}\right)!} = \frac{2}{\pi}\,\frac{2\cdot 4\cdot 6\cdots(\nu_2-1)}{1\cdot 3\cdot 5\cdots(\nu_2-2)} \tag{6.35}$$

substitutions from Eqs. (6.33), (6.34), and (6.35) to simplify the programming. The calculation with two odd-numbered degrees of freedom requires the summation in the square-bracket terms to be made twice, and that is the reason for the additional calculator time.

The listing and detailed description of program Variance-9 are given in the Appendix.

7 REGRESSION

7.1 INTRODUCTION

The graphical presentation of the relation between two variables is a fairly common procedure. The stock market average is plotted against time. The frequency of divorce is plotted against income bracket. In mathematical analysis of the relation between the variables, two concepts are employed: correlation and regression. Correlation is the measure of the variation of one variable with the variation in the other. No attempt is made to define the relationship, merely to measure the degree of correspondence. Is there correlation between smoking and lung cancer? Is there correlation between IQ and income bracket? If there is perfect correlation, the correlation coefficient is 1.0. If there is no correlation, the correlation coefficient is 0.0.

With regression an attempt is made to define the relation between the variables, often by a linear equation:

$$y = b_0 + b_1 \cdot x \tag{7.1}$$

but sometimes by other algebraic expressions. The dependent variable y is sometimes related to more than one independent variable, x_1, x_2, . . . , etc. All types of relating equations are referred to as regressions. They are correlating expressions defining the relation between the variables.

The constant coefficients in the regression equations are regression coefficients. In the simple linear equation, the regression coefficient is the same as the slope of the line, and the b_0 term is the intercept of the line with the Y axis—the value of y when x is zero.

There is a mathematical relation between the correlation coefficient and the regression coefficient in a linear regression. If the two variables are normally distributed and if their values are put in standard normal form [as discussed in Chap. 3, Eq. (3.7)], the correlation coefficient and the regression coefficient will be identical. Sometimes the terms regression and correlation are loosely used interchangeably.

This chapter deals with forms of regression between two or more variables and uses the correlation coefficient to measure the agreement of the data with the regression equations.

7.2 THE LEAST-SQUARES CALCULATION

The usual (and recommended) procedure for finding the relation between variables is first to plot the data. If the data form a straight line then it is reasonable to try to fit a straight-line linear equation. If the data do not fall on a straight line, some transformation of the x variable might be called for. The transformation may be accomplished by plotting the data on semilog or log-log or some other graph paper in which the scales are not linear. Or the transformation may be accomplished by using some function of the x variable in place of x, and plotting the transformed values on linear paper.

If the data have considerable scatter and no trend is obvious, the next steps are not so clear. Figures 7.1 and 7.2 show some data where there appears to be a trend, but the shape of the curve that might fit the data is not immediately apparent. Section 7.4 deals with linear regressions that are not straight lines.

The relation most commonly used is a straight line. This relation is discussed in the next section. Linear relations that are not straight lines are just as easily calculated with a relatively simple program as is demonstrated later.

The equation fitted to the data is usually the "least-squares" fit, meaning that

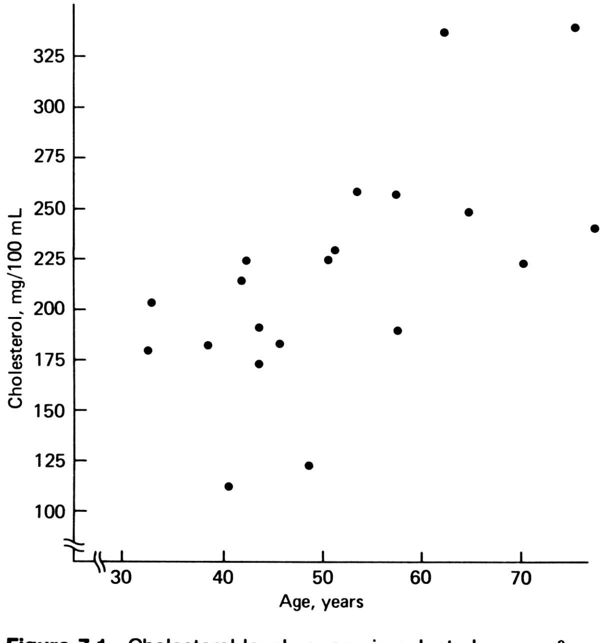


Figure 7.1 Cholesterol level vs. age in selected women.[9]

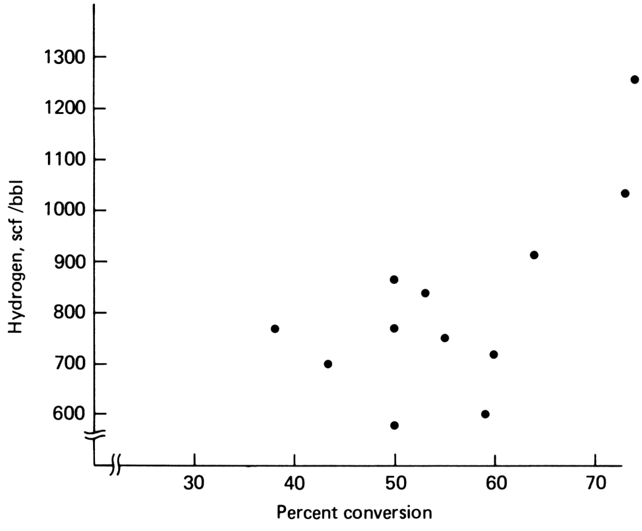


Figure 7.2 Hydrogen consumption vs. conversion[10] in heavy oil processing.

the sum of the squares of the deviations of the data from the regression line is a minimum. The constants in the equation, b_0 and b_1 for Eq. (7.1), are those that minimize the sum of squares of deviation for that particular equation. There may be other equations that give an even smaller sum of squares of deviation. The expression "least-squares" refers to the constants in the equation, not the form of the equation.

The calculation of the least-squares line for a straight-line linear equation is explained in the next section. The explanation will apply, with suitable modification, to any equation fitted by the same method. It is the method of calculating the "best" constants for the equation of your choice.

The question might be raised about minimizing the sum of squares of deviation between the data and the line as opposed to minimizing the sum of the deviations. Any straight line through $\bar{x}$ and $\bar{y}$ will have the sum of deviations equal to zero. This fact is inherent in the calculation of the mean, and was discussed in Chap. 2. There is no unique set of constants that minimize the sum of deviations.

7.3 CALCULATION OF THE LEAST-SQUARES STRAIGHT LINE

Given: a set of data: x_i, y_i; $i = 1$ to n
Find the equation $\hat{y} = b_0 + b_1 \cdot x$ such that $\Sigma(y - \hat{y})^2$ is a minimum.
Substitute for $\hat{y}$ the linear form: $b_0 + b_1 \cdot x$:

$$\Sigma(y - b_0 - b_1 \cdot x)^2 \tag{7.2}$$

This is the expression to be minimized.

In order to find the values of b_0 and b_1 that will minimize Eq. (7.2) it is differentiated with respect to both b_0 and b_1, and the differentials are equated with zero.

$$\Sigma y - \Sigma b_0 - b_1 \Sigma x = 0 \tag{7.3}$$

$$\Sigma xy - b_0 \Sigma x - b_1 \Sigma x^2 = 0 \tag{7.4}$$

These two equations are solved simultaneously to give:

$$b_1 = \frac{\Sigma xy - (\Sigma x \cdot \Sigma y)/n}{\Sigma x^2 - (\Sigma x \cdot \Sigma x)/n} \tag{7.5}$$

$$b_0 = \frac{\Sigma y - b_1 \cdot \Sigma x}{n} \tag{7.6}$$

Equations (7.5) and (7.6) give the values of the regression coefficient (the slope) and the intercept for the least-squares straight line through the data. The sum of squares of deviation of the y variable from this straight line is less than from any other straight line.

There may be other two-constant equations: $y = c_0 + c_1(1/x)$, or $y = d_0 + d_1 \cdot \log(x)$, that would give "better" correlation, that is, an even smaller sum of squares of deviation. The least-squares calculation gives the constants for the best equation of the form for which the calculation applies.

The measure of the goodness of fit is the correlation coefficient, usually designated r, and r^2 is the fraction of the sum of squares of deviation of the y variable from its mean that is accounted for by the regression line. If $\Sigma(y - \hat{y})^2$ is the sum of squares of deviation from the least-squares line:

$$r^2 = \frac{\Sigma(y - \bar{y})^2 - \Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2} \tag{7.7}$$

The correlation coefficient r may be calculated in several ways. The method used in the programs presented later is:

$$r = \sqrt{\frac{b_1 \cdot (\Sigma xy - (\Sigma x \cdot \Sigma y))/n}{\Sigma y^2 - (\Sigma y \cdot \Sigma y)/n}} \tag{7.8}$$

The larger the value of r, the better the correlation. An r value of unity indicates perfect correlation. An r value of zero indicates no correlation. Tables are available for establishing the significance of r for different amounts of data—different degrees of freedom for its calculation. These values, or the probability associated with these values, are the probabilities of obtaining a value of r as large as the tabulated values with the amount of data involved if there were no relation between the variables and each was taken randomly from normally distributed populations.

This type of statistical evaluation is similar to that discussed in Chap. 4 in the presentation of the t test. In fact there is a mathematical relation between the correlation coefficient r and the statistic t. With the t test a hypothesis was made and a t value was calculated from experimental data and a particular t equation. If the calculated t value corresponded to one of low probability, the hypothesis was rejected. With the correlation coefficient, the hypothesis is that there is no relation between the variables. If the calculated correlation coefficient is one that has a small probability of occurring when there is no relation between the variables, the hypothesis is rejected. Coefficient r can be converted to t, and the t tables may be used for the same type of judgment. The relation between r and t is:

$$t = \frac{r\sqrt{\nu}}{\sqrt{1 - r^2}} \tag{7.9}$$

where ν is the number of degrees of freedom associated with the regression equation. The number of degrees of freedom is equal to the number of sets of data n less the number of constants in the regression equation. With a linear equation, ν equals $n - 2$.

The significant values listed in probability tables for the correlation coefficient are identical with the two-tailed probability values associated with t, called α in this test. Programs were given in Chap. 4 for calculating the t probabilities if the degrees of freedom were known. The same t routines are used with the regression programs given later in this chapter, and the programs calculate t from the correlation coefficient and ν, using Eq. (7.9). The calculated probability α is the probability of a false rejection of the hypothesis that there is no relation between the variables.

In the discussion of the t test it was mentioned that the t value could be used to set a confidence range for the mean. If the measured mean was $\bar{x}$, the $1 - \alpha$ confidence range for the true mean was $\bar{x} \pm t_\alpha \cdot s(\bar{x})$, where $s(\bar{x})$ is the standard deviation of the mean. Similar confidence-range statements may be made about values calculated from the regression equations.

The estimated standard deviation of a measured variable was defined as:

$$s(x) = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \tag{2.10}$$

The following are equations for obtaining the estimated standard deviation of different quantities calculated by a least-squares regression equation.

The standard deviation of the regression estimate, sometimes called the standard error of estimate, is:

$$s(\hat{y}) = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} = \sqrt{\frac{(\Sigma y^2 - \Sigma y \cdot \Sigma y/n) - b_1(\Sigma xy - \Sigma x \cdot \Sigma y/n)}{n - 2}} \tag{7.10}$$

The standard deviation of the regression coefficient (the slope) is:

$$s(b_1) = \frac{s(\hat{y})}{\sqrt{\Sigma x^2 - \Sigma x \cdot \Sigma x/n}} \tag{7.11}$$

The standard deviation of $\bar{y}$ is:

$$s(\bar{y}) = \frac{s(\hat{y})}{\sqrt{n}} \tag{7.12}$$

The standard deviation of some value of x_i estimated from the regression equation is:

$$s(x_i) = s(\hat{y}) \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_i)^2}{\Sigma x^2 - \Sigma x \cdot \Sigma x/n}} \tag{7.13}$$

The standard deviation of the intercept b_0 is:

$$s(b_0) = s(\hat{y}) \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\Sigma x^2 - \Sigma x \cdot \Sigma x/n}} \tag{7.14}$$

The standard deviation of some predicted value Y_i, not associated with the data from which the regression was calculated, is:

$$s(Y_i) = s(\hat{y}) \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - X_i)^2}{\Sigma x^2 - \Sigma x \cdot \Sigma x/n}} \tag{7.15}$$

With any of these standard deviations, and with a t value corresponding to some probability α, a confidence range of $1 - \alpha$ may be calculated for the variable in question, equal to $\pm t_\alpha \cdot s$(variable). For example, the 90 percent confidence range for the slope would be $b_1 \pm t_{0.10} \cdot s(b_1)$, where $s(b_1)$ was calculated from Eq. (7.11), and $t_{0.10}$ could be obtained from the t-4 program.

7.3.1 Linear Regression Program: Regression-1

The calculation of the best straight line through a quantity of data is basically so easy, especially with a calculator, even one that is not programmable, that it is difficult to understand why it is not a standard procedure. If all that is desired is to draw a straight line through some data, a very simple program of about 26 lines will give the constants for the least-squares fit. Several calculators have built-in routines for calculating b_0 and b_1.

The additional features of Regression-1 are that it calculates the correlation coefficient r and the probability value to indicate how significant the correlation is. It also calculates the standard error of estimate [Eq. (7.10)], so that confidence ranges may be set for the various regression equation parameters.

With the program available on a magnetic card, the greatest amount of effort involved is in entering the data. The calculator running time for the calculation is about 10 s. However, if all that is wanted is the slope and the intercept to draw the best straight line, the program can be terminated after line 39 in the Hewlett-Packard version or after line 74 in the Texas Instruments version. These changes are pointed out in the detailed discussion of the programs in the Appendix.

A flow diagram of Regression-1 is shown in Fig. 7.3, and the listing and detailed description are given in the Appendix. The data are entered with KEY A: y_i, ENTER ↑, x_i, KEY A in the Hewlett-Packard program or y_i, KEY $x \rightleftharpoons t$, x_i, KEY A in the Texas Instruments program. When all the data have been entered, KEY C runs the program.

The order of entry is important. If the data are entered with the x variable first, the program will calculate the regression of x upon y instead of y upon x.

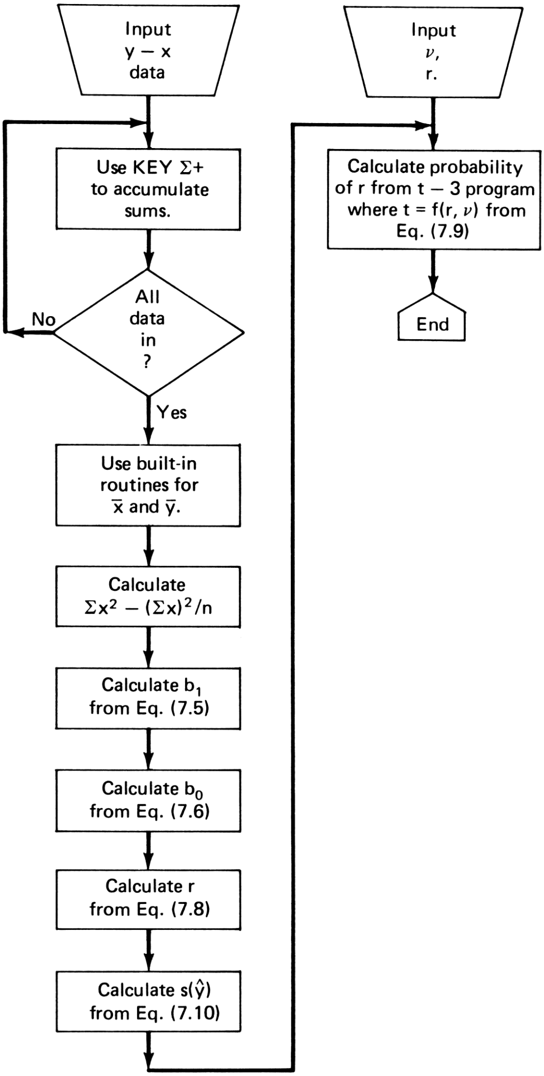


Figure 7.3 Flow diagram for the regression-1 program.

The program output is:

$\bar{x}$
$\bar{y}$
$(\Sigma x^2 - \Sigma x \cdot \Sigma x/n)$
b_1
b_0
r
α, probability associated with r
$s(\hat{y})$

The $\bar{x}$, $\bar{y}$ values are useful because the calculated line can be drawn through this point and the intercept b_0. The third term calculated is used in calculating the standard deviations in Eqs. (7.11), (7.13), (7.14), and (7.15).

The probability is calculated using the t-3 program which is made part of Regression-1 with the correlation coefficient converted to the equivalent t value by Eq. (7.9). This portion of the program can be used independently of the least-squares solution if a correlation coefficient is obtained from another source. Input of the degrees of freedom and the r value with KEY D will get the probability as output.

The program makes use of the built-in function of KEY $\Sigma+$ which calculates and stores Σx, Σx^2, Σy, Σy^2, Σxy, and n. If incorrect data have been entered, they may be deleted by reentering the same data with KEY $\Sigma-$. After the program has been run, data may be added or deleted to see their effect on the correlation. To add or delete data, KEY $\Sigma+$, or KEY $\Sigma-$ must be used. If additional data are added with KEY A after the program has been run, the input will be interpreted as the start of new data and all previous calculations will be cleared.

7.3.2 Example of Regression-1

Figure 7.4 shows some data[8] that appear to fit a straight line. The data for the figure are given in the Figure. With Regression-1 in the calculator, the data are entered:

HP	TI
459, ENTER ↑	459, KEY $x \rightleftharpoons t$
357, KEY A	357, KEY A
419, ENTER ↑	419, KEY $x \rightleftharpoons t$
392, KEY A	392, KEY A
⋮	⋮
114, ENTER ↑	114, KEY $x \rightleftharpoons t$
112, KEY A	112, KEY A

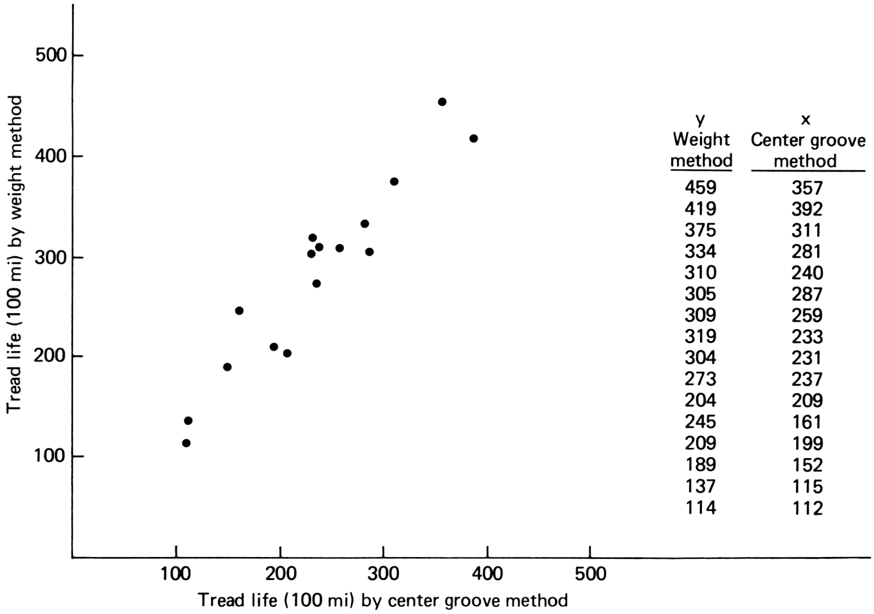
When all the data are entered, KEY C gives the following output:

236.0	$\bar{x}$	
281.6	$\bar{y}$	
94604.	$\Sigma x^2 - (\Sigma x)^2/n$	
1.137	b_1	
13.265	b_0	
0.9478	r	
2.5×10^{-8}	α, the probability	
31.435	$s(\hat{y})$	(Time:27/30)

The best-fitting straight-line equation is:

$$y = 13.265 + 1.137\ x$$

with a correlation coefficient of 0.948, which is highly significant.



y Weight method	x Center groove method
459	357
419	392
375	311
334	281
310	240
305	287
309	259
319	233
304	231
273	237
204	209
245	161
209	199
189	152
137	115
114	112

Figure 7.4 Tire-tread wear data.

The standard deviation of the slope, from Eq. (7.11) is:

$$s(b_1) = 31.435/\sqrt{94604} = 0.102$$

The standard deviation of the mean value of $\bar{y}$, from Eq. (7.12), with 16 sets of observations, is:

$$s(\bar{y}) = 31.435/\sqrt{16} = 7.859$$

Other values relating to the correlation could be calculated in a like manner from Eqs. (7.13) to (7.15).

7.4 REGRESSION LINES OTHER THAN STRAIGHT LINES

It was pointed out in the discussion of the least-squares equation that the least-squares calculation for a straight-line regression gives the best straight line, but it might not be the best linear equation. Figure 7.5 shows a number of curves in the form:

$$y = b_0 + b_1 \cdot F(x) \qquad (7.16)$$

where $F(x)$ is some function of x.

When a plot of the data shows an apparent trend of y with x, but when it is not evident that a straight line would be the best fit, a linear equation with some transformation of x might be better. The curves in Fig. 7.5 show the shape of a number of common transformations. If the y values increase as x increases and the data appear to be concave downward, the functions $1/x$, $\log(x)$, or $\sqrt{x}$ may

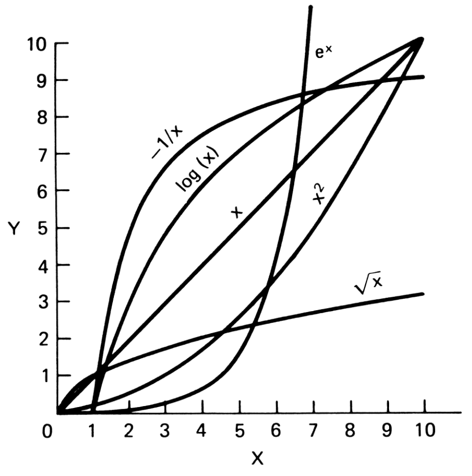


Figure 7.5 $y = b_0 + b_1 \cdot F(x)$. Representative curves of functions of x with y increasing with increases in x.

fit the data. If the data are concave upward, e^x, x^2, or x to some other power greater than one may be used.

If the y values decrease as x increases, the reverse is true. Curves of $1/x$, log (x), or $\sqrt{x}$ are concave upward, and the others are concave downward. This is illustrated in Fig. 7.6.

To calculate the constants in Eq. (7.16) that give the best fit (the least-squares fit) for the particular function of x selected, all the equations, (7.2) to (7.8), of the previous section apply with $F(x)$ substituted for x. The equations for the estimated standard deviations, (7.10) to (7.15) also apply with $F(x)$ substituted for x.

Calculator program Regression-2 makes the least-squares calculations for all of the functions shown in Fig. 7.5 and 7.6. It calculates the constants and correlation coefficient for each equation with one input of the data. The engineer may then select the one that gives the most satisfactory fit. It is not anticipated that all of these functions will often be of interest at the same time. The program offers a

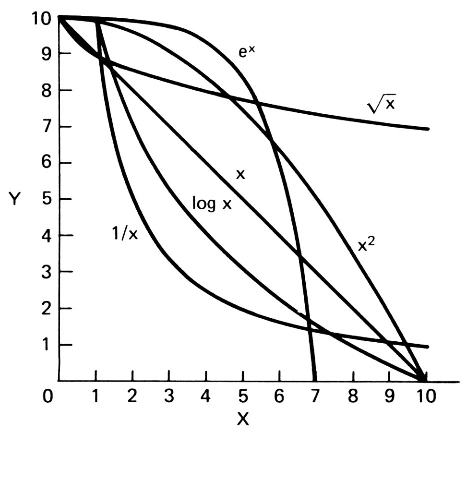


Figure 7.6 $y = b_0 + b_1 \cdot F(x)$. Representative curves of functions of x with y decreasing with increases in x.

technique for testing a number of forms of linear equations at the same time and the particular functions may be selected by the engineer and those not wanted can be easily deleted from the program.

7.4.1 Regression-2 Program

The listing and detailed description of the Regression-2 program are given in the Appendix. A flow diagram is shown in Fig. 7.7. The program as given provides for testing the functions x, x^2, $\sqrt{x}$, e^x, $1/x$, and $\log(x)$. Any other functions available in the calculator may be substituted for those listed in the program. Also, any function for which there is no interest may readily be deleted in the Hewlett-Packard program by simply eliminating the line in which that function

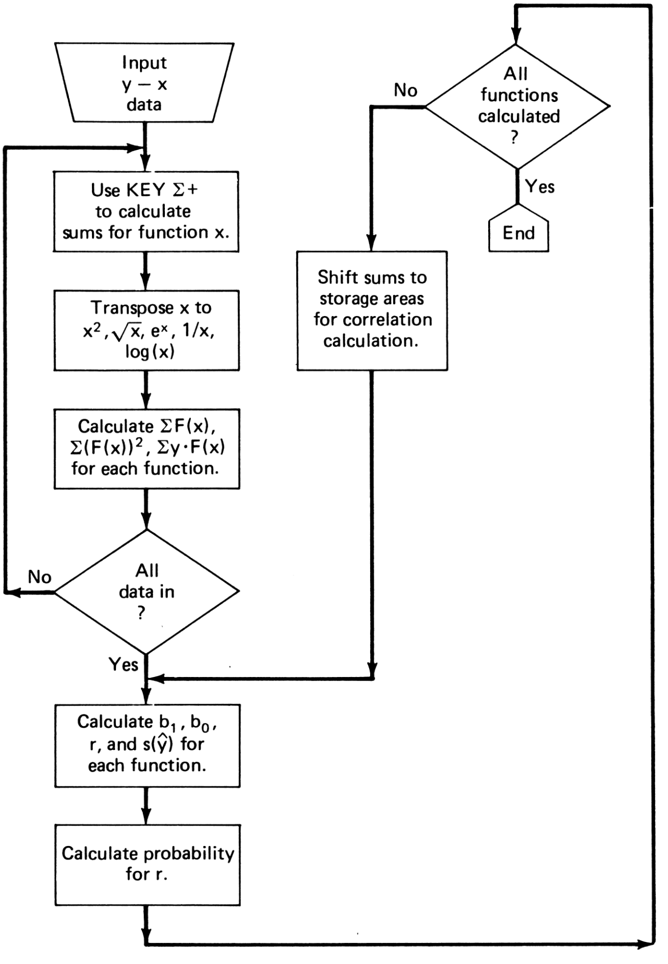


Figure 7.7 Flow diagram for the Regression-2 program: nonstraight line correlation.

is calculated. In the Texas Instruments program, two lines must be eliminated. If a function is deleted, the program uses the unchanged value of x and the simple linear calculation is repeated.

The calculation of the least-squares line for any function of x is the same as for a straight line with the function substituted for x in Eqs. (7.2) to (7.8). The program makes the transformation of the input values of x to the different functions and then carries out the same calculation for each transformation. For each function the output is b_1, b_0, r, α (the probability associated with r), and $s(\hat{y})$.

There are some restrictions to the use of the program as it is presented. Negative values of x may not be used with the $\sqrt{x}$ or the $\log(x)$ functions. Most calculators have a numerical limitation, and x values exceeding about ± 115 will exceed the limits for the e^x function. Zero values may not be used with the $1/x$ or $\log(x)$ functions. If the program is to be used a great deal, it is possible to include some control instructions to bypass the impossible calculations before they are encountered. However, it seems better to be aware of the restrictions and to make any changes in the program required for data that cannot be processed.

Some of the limits of the calculator may be overcome by coding the data: adding a constant to all the values to eliminate negative numbers; reducing all values by a constant to come within the e^x limits. Division by zero, with the $1/x$ function, may be avoided by substituting a very small number, 1×10^{-6}, which will have little effect on the calculation if there are a large number of data points. If the data cannot be coded to fit the limits of the calculator, the particular function that cannot handle the data must be eliminated or the program will terminate in the Hewlett-Packard system with an error message, or continue the calculation but produce an incorrect answer in the Texas Instruments system.

Regression-2 is run in a manner similar to Regression-1. Data are entered with KEY A: y_i, ENTER ↑, x_i, KEY A in the Hewlett-Packard calculators, and y_i, KEY $x \rightleftharpoons t$, x_i, KEY A in the Texas Instruments systems. When all the data are entered, KEY C starts the calculations. The output is b_1, b_0, r, α, and $s(\hat{y})$ for each function in the order: x, x^2, $\sqrt{x}$, e^x, $1/x$, and $\log(x)$.

The calculations follow Eqs. (7.3) to (7.8). The probability value α is based on the probability of t from Eq. (7.9), and the t-3 program which is merged into Regression-2. The calculation of $s(\hat{y})$ is by Eq. (7.10). The same calculations are made for each of the functions of x.

Incorrect data input cannot be corrected with KEY $\Sigma-$ as they can be in Regression-1. Also, after the program has been run, additional data cannot be added or data deleted as they can be in Regression-1 because the storage areas are changed during the calculation. After the data have been entered, and before the program has been run, the data may be saved on a data card and these data may be reentered and then additional data added for another run of the program.

7.4.2 Examples of Regression-2 Program

Figures 7.1 and 7.2 show some data where the shape of a correlating curve is not evident from the plot of the data. Figures 7.5 and 7.6 show the shape of different functions of x that might fit the data. In Fig. 7.2 the y values increase

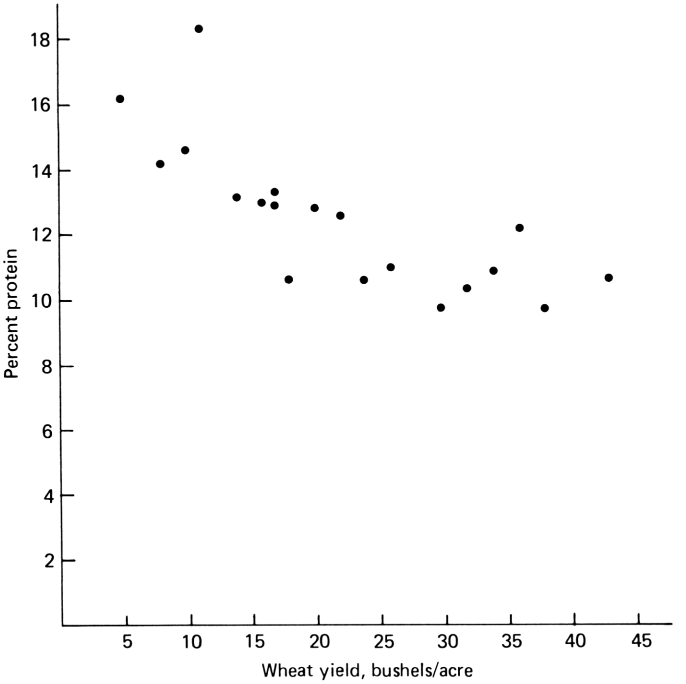


Figure 7.8 Percent protein content vs. yield of wheat (selected data).[9]

with increasing x values. Figure 7.5 might suggest fitting the data with e^x or x^2. The plot of Fig. 7.8 shows some y values decreasing with increase in x. Figure 7.6 would suggest fitting these data with $1/x$, $\log(x)$, or $\sqrt{x}$.

The first example uses the data of Fig. 7.2 as shown in Table 7.1.

7.4.2.1

Table 7.1 Data from Figure 7.2

Hydrogen consumption y, scf/bbl	Fraction conversion x
770	0.38
700	0.44
580	0.50
770	0.50
870	0.50
820	0.53
750	0.55
600	0.59
720	0.60
920	0.64
1040	0.73
1260	0.74

The Regression-2 program is used with all the functions. The correlation coefficient will indicate the best equation. None of the data is incompatible with any of the functions. The data are entered

	HP	TI
770	ENTER ↑	KEY $x \rightleftharpoons t$
0.38	KEY A	KEY A
700	ENTER ↑	KEY $x \rightleftharpoons t$
0.44	KEY A	KEY A
⋮		
1260	ENTER ↑	KEY $x \rightleftharpoons t$
0.74	KEY A	KEY A
	KEY C	KEY C

The output is:

	Function					
Variables	x	x^2	$\sqrt{x}$	e^x	$1/x$	$\log(x)$
b_1 (slope)	1183.8	1099.6	1700.6	691.11	−287.23	1385.1
b_0 (intercept)	115.71	462.09	−448.61	−397.78	1349.4	1177.7
r (correlation coefficient)	0.676	0.723	0.649	0.704	0.556	0.620
α (probability)	0.016	0.008	0.022	0.011	0.060	0.032
$s(\hat{y})$ (standard error of estimate)	146.4	137.2	151.1	141.1	165.1	155.9

(Time:115/209)

The best (largest) correlation coefficient is for the x^2 correlation, with e^x and x being next in order. All of the correlations are statistically significant except $1/x$, which has a probability level of 0.06. (Usually 0.05 is taken as the maximum acceptable probability, that is, a 5 percent chance of error in saying there is a correlation.)

The standard error of estimate is part of the output for each function. This is another measure of the reliability of the correlations. If it is anticipated that confidence ranges using Eqs. (7.11) to (7.14) will be useful, the value of $\Sigma x^2 - (\Sigma x)^2/n$ can be obtained for each function by placing a PRINT statement in the Hewlett-Packard program between lines 92 and 93, and in the Texas Instruments program between lines 204 and 205.

With one input of the data the six regression equations and evaluating parameters were calculated in 2 to 3 min of calculator time.

7.4.2.2 For this example, the data from Fig. 7.8 are used. They are given in Table 7.2.

Inasmuch as the y values decrease with increase in x and the curvature is upward, the functions e^x and x^2 cannot be expected to give good correlations. However since the data are compatible with these functions, it is simpler to leave the functions in the program than to delete them. A PRINT statement is inserted between lines 92 and 93 in the Hewlett-Packard program and between lines 204

Table 7.2 Data from Fig. 7.8

Percent protein y	Wheat yield x
10.7	43
9.8	38
12.2	36
10.9	34
10.4	32
9.8	30
11.0	26
11.6	24
12.6	22
12.8	20
10.6	18
13.4	17
13.0	17
13.0	16
13.2	14
18.3	11
14.6	10
14.2	8
16.2	5

and 205 in the Texas Instruments program to obtain the value of $\Sigma x^2 - (\Sigma x)^2/n$. The program is then run as in Example 7.4.2.1:

	HP	TI
10.7	ENTER ↑	KEY $x \rightleftharpoons t$
43	KEY A	KEY A
9.8	ENTER ↑	KEY $x \rightleftharpoons t$
38	KEY A	KEY A
⋮		
16.2	ENTER ↑	KEY $x \rightleftharpoons t$
5	KEY A	KEY A
	KEY C	KEY C

The output is:

Variables	Function: x	x^2	$\sqrt{x}$	e^x	$1/x$	$\log(x)$
$\Sigma x^2 - (\Sigma x)^2/n$	2140.5	5043754.4	26.09	2.1×10^{37}	0.0336	1.1201
b_1 (slope)	−0.1585	−0.0029	−1.489	-4.2×10^{-19}	38.66	−7.2649
b_0 (intercept)	16.05	14.30	19.33	12.65	10.16	21.88
r (correlation coefficient)	0.7814	0.6988	0.8105	0.2039	0.7552	0.8194
α (probability)	0.0001	0.0009	2.5×10^{-5}	0.4025	0.0002	1.8×10^{-5}
$s(\hat{y})$ (standard error of estimate)	1.4202	1.6280	1.3329	2.2280	1.4917	1.3045

(Time:180/285)

The log(x) function is best with a correlation coefficient of 0.8194. The e^x function is not significant, as expected. The x^2 function is significant although the curvature of the correlation equation is downward while the data appear to curve upward. The range of the y response is small and the scatter of the data is relatively wide so the x^2 function can give a statistically significant correlation in spite of the apparently incorrect curvature. However the log(x) and $\sqrt{x}$ functions are significantly better and have little to choose from between them.

There are 19 sets of data with 17 degrees of freedom. The t value for a 95 percent confidence range, from the t-2 program, is 2.11. The 95 percent confidence range for the slopes of any of the functions, from Eq. (7.11), would be:

$$\frac{\pm (2.11) \cdot s(y)}{\sqrt{\Sigma x^2 - (\Sigma x)^2/n}}$$

For the log(x) correlation, this would be

$$b_1 = -7.2649 \pm (2.11)(1.30)/\sqrt{1.1201} = -9.856 \text{ to } -4.674$$

Similar calculations may be made for the other equation parameters and for values estimated from the equations.

7.5 REGRESSION WITH TWO INDEPENDENT VARIABLES AND SECOND-DEGREE EQUATIONS

The correlating equations for two independent variables and for a second-degree function of one variable can be handled in a similar manner. The physical situations represented by the equations are significantly different, but the least-squares solution of the equations is the same. A linear equation with two independent variables would be:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 \tag{7.17}$$

A second-degree equation with one independent variable would be:

$$y = b_0 + b_1 \cdot x + b_2 \cdot x^2 \tag{7.17.1}$$

The first equation is used when it is expected that the dependent variable is affected independently by the two independent variables. Examples are: growth as a function of time and nutrients; conversion as a function of temperature and catalyst concentration; yield as a function of pressure and mixing rate.

The second equation is used when the relation between the variables is not linear and appears to change direction with increase in the independent variable. The curves illustrated in Fig. 7.5 and 7.6 showed a continuous directional trend of the y variable with increase in the x variable. When the trend of the y variable changes direction: reaches a maximum and then declines, or reaches a minimum and then starts to increase, the second-degree equation may provide a better fit than a single-term equation of the forms used in the Regression-2 program.

The two equations are discussed separately in the sections that follow, but the solution of the equations is handled in a similar manner with the Regression-3 program.

7.5.1 The Two Independent Variable Equation

The correlation of a dependent variable against two independent variables with the multiple linear equation (7.17) follows a similar procedure to that of the one independent linear equation solution of Sec. 7.3. The values of b_0, b_1, and b_2 are determined that will minimize the sum of the squares of the differences between the observed values of y and the values estimated from Eq. (7.17).

If $\hat{y}_i$ is the estimated value of y_i:

$$\hat{y}_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} \tag{7.18}$$

the value to be minimized is $\Sigma(y_i - \hat{y}_i)^2$ or $\Sigma(y_i - b_0 - b_1x_{1i} - b_2x_{2i})^2$. If the last expression is differentiated with respect to each of the constants and the three differentials are equated with zero and solved simultaneously, the solution will give the values of b_0, b_1, and b_2 that minimize the sum of squares of deviation. This is the least-squares solution and is similar to the solution for the simple linear equation given in Sec. 7.3.

The actual solution is illustrated in a number of the references previously cited and isn't repeated here. The final equations for the three constants are as follows:

$$b_1 = ((\Sigma yx_1 - (\Sigma y\Sigma x_1)/n)(\Sigma x_2{}^2 - (\Sigma x_2)^2/n) - (\Sigma yx_2 - (\Sigma y\Sigma x_2)/n)(\Sigma x_1x_2 - (\Sigma x_1\Sigma x_2)/n))/D \tag{7.19}$$

$$b_2 = ((\Sigma yx_2 - (\Sigma y\Sigma x_2)/n)(\Sigma x_1{}^2 - (\Sigma x_1)^2/n) - (\Sigma yx_1 - (\Sigma y\Sigma x_1)/n)(\Sigma x_1x_2 - (\Sigma x_1\Sigma x_2)/n))/D \tag{7.20}$$

$$D = (\Sigma x_1{}^2 - (\Sigma x_1)^2/n)(\Sigma x_2{}^2 - (\Sigma x_2)^2/n) - (\Sigma x_1x_2 - (\Sigma x_1\Sigma x_2)/n)^2 \tag{7.21}$$

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 \tag{7.22}$$

The values of the constants b_0, b_1, and b_2 calculated from Eqs. (7.19) to (7.22) will give the equation that produces the minimum sum of squares of deviation between the observed values of y and the values estimated from Eq. (7.18). This sum of squares will be equal to:

$$\text{Sum of squares} = (\Sigma y^2 - (\Sigma y)^2/n) - C \tag{7.23}$$

where $$C = b_1(\Sigma yx_1 - (\Sigma y\Sigma x_1)/n) + b_2(\Sigma yx_2 - (\Sigma y\Sigma x_2)/n) \tag{7.24}$$

The correlation coefficient R will equal:

$$R = (C/(\Sigma y^2 - (\Sigma y)^2/n)^{1/2} \tag{7.25}$$

The variance of estimate will be:

$$s^2(\hat{y}) = \text{Sum of squares}/(n - 3) \tag{7.26}$$

These equations are similar to Eqs. (7.8) and (7.10) for the single independent variable equation, and the terms have a similar role. The correlation coefficient is a measure of the goodness of fit of the equation to the data, and the probability of R can be obtained from tables or from the relation of R to t given in Eq. (7.9) where the number of degrees of freedom for the three-constant equation is $n-3$. The correlation coefficient for the multiple-term equation is written with uppercase R, and for the single-term equation with the lowercase r as a matter of convention.

The square root of the variance of estimate can be used to set confidence limits to values calculated from the equation in a manner similar to that discussed in Sec. 7.3.

The Regression-3 program carries out all the calculations indicated in the preceding equations. In addition, it gives the results for the correlation of the dependent variable with each of the independent variables separately.

It is often the case where a response is thought to be dependent upon two factors: time and temperature, organic and inorganic sulfur; pressure and mixing rate, that the correlation with one factor is overriding. If the correlation with two factors is tested and found to be significant, it is seldom that the time is taken to determine whether the correlation with just one factor is satisfactory. The usual situation requires three correlation calculations and a comparison of the correlation coefficients.

The Regression-3 program makes all three calculations with one input of the data. In addition to calculating the three correlation coefficients, the Regression-3 program makes a statistical comparison of the best two.

In making the calculation of the least-squares solution to Eq. (7.17) all the terms required for the least-squares solutions for the single variable equations are obtained. In Regression-3, before the two independent variable constants are calculated, the best-fitting constants for each of the single variable equations:

$$y = b_0 + b_1 \cdot x_1 \tag{7.27}$$

and

$$y = b_0 + b_1 \cdot x_2 \tag{7.28}$$

are calculated along with the correlation coefficient for each. No additional data input is required and the additional calculator time is insignificant.

Each of the three correlation equations has a correlation coefficient that is a measure of the fit of the equation. By the nature of the arithmetic the correlation coefficient for the two-independent-variable equation cannot be less than the coefficient for the equation with one of the variables. However, the two-variable-equation calculation has one less degree of freedom than the equation with one variable and the correlation coefficient may not be significantly larger statistically. There is a t test for measuring the difference between two correlation coefficients. The equation is:

$$t = \sqrt{\frac{\nu(r_2{}^2 - r_1{}^2)}{1 - r_2{}^2}} \tag{7.29}$$

where r_1 is the smaller correlation coefficient
r_2 is the larger correlation coefficient
ν is the degrees of freedom associated with r_2

The probability value associated with t at ν degrees of freedom is the probability of a false rejection of the hypothesis that the two correlations are equivalent. If t is large and the probability is small, r_2 is accepted as being better than r_1. The usual criterion is a probability of 0.05.

The Regression-3 program selects the larger of the correlation coefficients for the single-variable correlations and calculates t from Eq. (7.29) for a comparison with the correlation coefficient of the two-variable equation. The t value is printed, and the probability can be obtained with the t-3 program. The HP-67 and HP-97 systems do not have sufficient program memory for the two programs to be merged. With the TI-59 system, the programs could be merged with a change in the partition between program and storage memory. However, in the interests of keeping all the programs consistent, both programs stop with the output of the t value.

7.5.2 Regression-3 Program

The program takes the original y, x_1, and x_2 data and makes all the calculations discussed in the preceeding paragraphs. The data are entered, in the Hewlett-Packard program: y_i, ENTER ↑, x_{1i}, ENTER ↑, x_{2i}, KEY A. In the Texas Instruments program all the data are entered with KEY A: y_i, KEY A, x_{1i}, KEY A, x_{2i}, KEY A. When all the data have been entered, KEY C runs the program.

The output is b_1, b_0, and r for Eq. (7.27); b_1, b_0, and r for Eq. (7.28); and b_1, b_2, b_0, and R for Eq. (7.17). The program then prints the t value for the comparison of the third R value with the larger of the first two, following Eq. (7.29). If the probability for the t value is desired, it can be obtained with the t-3 program by entering the number of degrees of freedom, $n - 3$, and the t value.

A flow diagram for Regression-3 is given in Fig. 7.9. Regression-3 also handles a second-degree polynomial correlation, and the flow for that portion of the program is included in Fig. 7.9. A discussion of the second-degree polynomial follows in Sec. 7.5.4.

All of the sums, sums of squares, and sums of cross products are accumulated as the data are entered. In the Hewlett-Packard program the built-in functions of KEY $\Sigma+$ are used for all the sums except $\Sigma x_1 x_2$. In the Texas Instruments program the KEY $\Sigma+$ is used only for the y-x_1 data, and the other sums are calculated directly. The reason for this difference is that the Hewlett-Packard system has a storage register shift facility that shifts 10 storage areas at one time.

The calculation of b_1, b_0, and r for the two single-variable equations directly follows Eqs. (7.5) and (7.6); r is calculated from Eq. (7.8).

The calculation of b_1, b_2, and b_0 for the two-variable equation follows Eqs. (7.19) to (7.22). The D value of Eq. (7.20) is calculated first and it is used for both b_1 and b_2. The correlation coefficient is calculated from Eq. (7.25). This

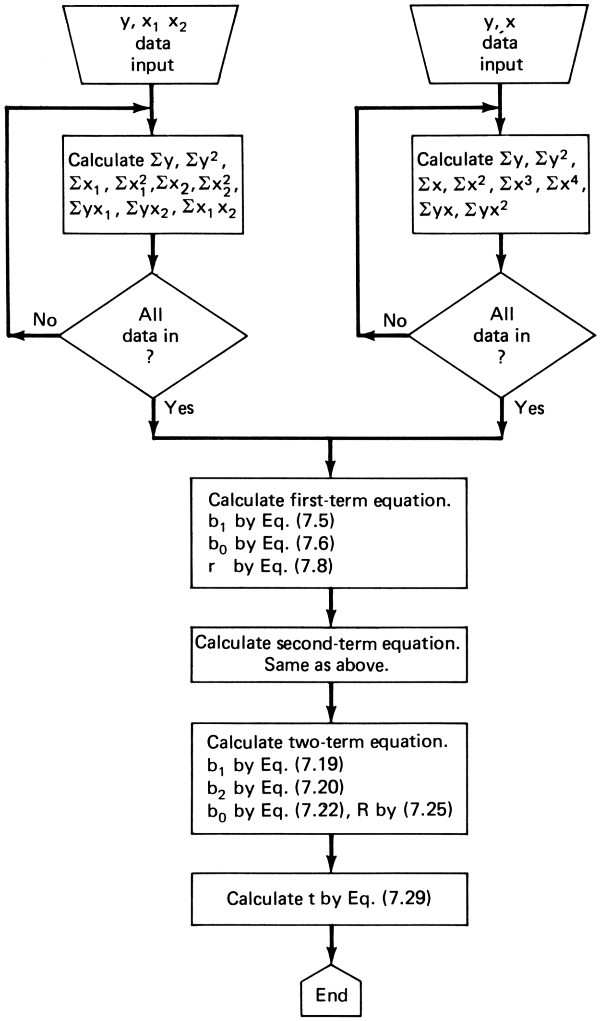


Figure 7.9 Flow diagram for the Regression-3 program: two-variable and polynomial correlation.

value is compared with the larger of the two correlation coefficients from the one-variable equations, and t is calculated from Eq. (7.29).

At the completion of the calculation the different values are stored in the registers shown in Table 7.3. These values are available if confidence limits or further calculations are desired.

7.5.3 Examples of Regression-3 with Two Independent Variables

7.5.3.1 The data[11] in Table 7.4 are the plant-available phosphorus as dependent variable and the organic and inorganic phosphorus in the soil as independent variables.

Table 7.3 Storage Values after Regression-3

Value	Register HP	Register TI
Σx_1	4′	01
Σx_2	4	08
Σy	6′	04
$\Sigma x_1^2 - (\Sigma x_1)^2/n$	E	02
$\Sigma x_2^2 - (\Sigma x_2)^2/n$	5	09
$\Sigma x_1 x_2 - (\Sigma x_1 \Sigma x_2)/n$	0	10
$\Sigma y^2 - (\Sigma y)^2/n$	7	05
$\Sigma y x_1 - (\Sigma y \Sigma x_1)/n$	D	06
$\Sigma y x_2 - (\Sigma y \Sigma x_2)/n$	8	07
n	9	03
b_1 Eq. (7.19)	2	11
b_2 Eq. (7.20)	3	12
R Eq. (7.25)	6	13
r_1 Eq. (7.8)	A	20
r_2 Eq. (7.8)	C	22
D Eq. (7.21)	1	21
C Eq. (7.24)	B	14

NOTE: Prime indicates secondary registers.

Table 7.4 Data for Plant-Available Phosphorus and Soil Phosphorus

Plant-available phosphorus y	Inorganic phosphorus x_1	Organic phosphorus x_2
64	0.4	53
60	0.4	23
71	3.1	19
61	0.6	34
54	4.7	24
77	1.7	65
81	9.4	44
93	10.1	31
93	11.6	29
51	12.6	58
76	10.9	37
96	23.1	46
77	23.1	50
93	21.6	44
95	23.1	56
54	1.9	36
168	26.8	58
99	29.9	51

The first assumption might be that the plant-available phosphorus is a function of both the organic and inorganic phosphorus in the soil.

A correlation of the total data gives the result:

$$y = 56.25 + 1.79\,x_1 + 0.09\,x_2 \tag{7.30}$$

with a correlation coefficient of 0.695, which is statistically significant, and probability of 0.002. The study might end here with a satisfactory correlating equation. However, with Regression-3 some additional interesting information is revealed.

If the data of Table 7.4 are entered in the Regression-3 program:

		HP	TI
	64	ENTER ↑	KEY A
	0.4	ENTER ↑	KEY A
	53	KEY A	KEY A
	60	ENTER ↑	KEY A
	0.4	ENTER ↑	KEY A
	23	KEY A	KEY A
	⋮		
	99	ENTER ↑	KEY A
	29.9	ENTER ↑	KEY A
	51	KEY A	KEY A
With all the data entered:		KEY C	KEY C

The output is as follows (the additional information is explanatory, the programs produce only the numerical results):

1.8434	b_1 for the equation $y = b_0 + b_1 \cdot$ inorganic phosphorus
59.2590	b_0 for the same equation
0.6934	r, correlation coefficient for the same equation
0.7023	b_1 for the equation $y = b_0 + b_1 \cdot$ organic phosphorus
51.7013	b_0 for the same equation
0.3545	r, correlation coefficient for the same equation
1.7898	b_1 for the equation $y = b_0 + b_1 \cdot$ inorganic phosphorus $+ b_2 \cdot$ organic phosphorus
0.0866	b_2 for the same equation
56.2510	b_0 for the same equation
0.6945	R, correlation coefficient for the same equation
0.2088	t, comparison of the third equation with the first

(Time:17/18)

These results show that the correlation for the third equation is almost the same as for the first. The very low t value (which has a probability value of 0.837 from the t-3 program) indicates there is no significant difference between the first equation with one variable and the third which uses two. The correlation

Table 7.5 Data for Example 7.5.3.2

Octane number y	Catalyst purity x_1, %	Carbon level x_2, wt. %
88.6	99.8	3.0
86.4	99.7	10.0
87.2	99.6	7.0
88.4	99.5	2.0
87.2	99.4	5.0
86.8	99.3	6.0
86.1	99.2	8.0
87.3	99.1	3.0
86.4	99.0	5.0
86.6	98.9	4.0
87.1	98.8	2.0

of plant-available phosphorus is almost entirely with the inorganic phosphorus, and the organic phosphorus contributes nothing significant to the correlation. Without the calculation of the single-variable correlations, the more involved two-variable equation might be used. Regression-3 calculates and evaluates all three possibilities with one entry of the data.

7.5.3.2 Of course, sometimes two variables are better than one. The data[10] in Table 7.5 are for some gasoline octane ratings of the products of feedstock conversion tests with the catalyst purity and the level of carbon on the catalyst as independent variables.

The data are entered with the Regression-3 program in the calculator:

	HP	TI
88.6	ENTER ↑	KEY A
99.8	ENTER ↑	KEY A
3.0	KEY A	KEY A
86.4	ENTER ↑	KEY A
99.7	ENTER ↑	KEY A
10.0	KEY A	KEY A
⋮		
87.1	ENTER ↑	KEY A
98.8	ENTER ↑	KEY A
2.0	KEY A	KEY A
When all the data are entered:	KEY C	KEY C

The output is:

1.1273 $\quad b_1$ for equation $y = b_0 + b_1 \cdot$ catalyst purity

−24.8382	b_0 for the same equation
0.4703	r: correlation coefficient for same equation
−0.2030	b_1 for equation $y = b_0 + b_1 \cdot$ carbon level
88.1152	b_0 for the same equation
0.6561	r: correlation coefficient for same equation
1.9189	b_1 for equation $y = b_0 + b_1 \cdot$ catalyst purity $+ b_2 \cdot$ carbon level
−0.2903	b_2 for the same equation
−101.9923	b_0 for the same equation
0.9959	R: correlation coefficient for same equation
23.5382	t for comparison of third equation with second (Time:17/18)

The correlation coefficient for the second equation is statistically significant (probability value of 0.028 from the Regression-1 program, KEY D), and might be accepted as a satisfactory correlation. But the correlation with both variables, with a correlation coefficient of 0.9959, is so much better, that it is probably worth using the two-variable equation to correlate the octane number.

When data for two independent variables are available, Regression-3 provides a means for testing the correlation with either or both variables with no more effort than for testing with one variable.

7.5.4 Second-Degree Polynomial Correlation

It was mentioned in Sec. 7.4 and illustrated in Figs. 7.5 and 7.6, that the second-degree equation in the form $y = b_0 + b_1 x^2$ produced a curve that was concave either upward or downward depending on the sign of b_1. If a plot of y against x shows a change in direction, decreasing with increase in x and then increasing, or vice versa, the best fit could still be with a second-degree equation, but with a first-degree x term of a different sign than the x^2 term. Figure 7.10 illustrates some second-degree curves.

Actual experimental data seldom have such obvious configurations. Figure 7.11 shows a more realistic situation where there is an obvious relationship between the variables but the shape of the best-fitting curve is not clear. The data[12] for Fig. 7.11 are the coke consumption and air-to-steam ratio in a water-gas plant. The actual data are given in Table 7.6. Three curves are drawn in Fig. 7.11, a straight line, a second-degree equation with only the square term, and a second-degree equation with two terms. It is not clear from the figure which, if any, of these curves gives a good fit or which gives the best fit. The curves are drawn from best-fitting equations as will be illustrated in the example that follows.

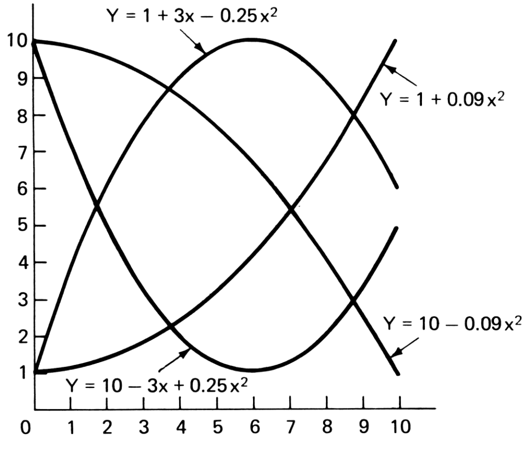


Figure 7.10 Representative curves of functions of x^2: $y = b_0 + b_1 \cdot x$ and $y = b_0 + b_1 \cdot x + b_2 \cdot x^2$

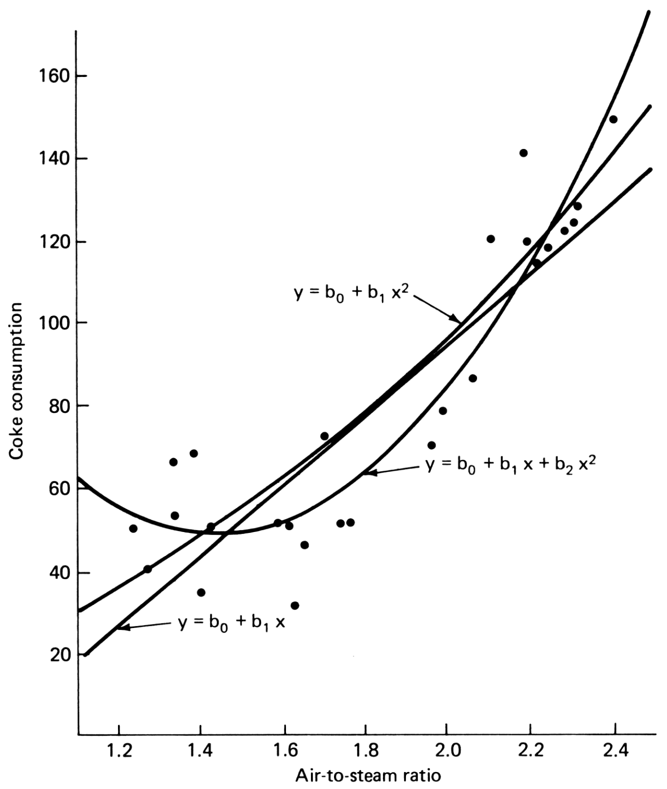


Figure 7.11 Coke consumption vs. air-to-steam ratio in a water-gas plant.

Table 7.6 Data for Points on Fig. 7.11

Coke	Air-to-steam	Coke	Air-to-steam
120	2.11	51	1.76
122	2.29	53	1.33
128	2.32	50	1.23
124	2.31	34	1.40
118	2.25	68	1.38
114	2.22	70	1.96
119	2.20	49	1.47
149	2.41	50	1.42
141	2.19	66	1.33
86	2.06	46	1.65
78	1.99	40	1.26
31	1.62	51	1.61
51	1.59	51	1.74
72	1.70		

The calculation of the least-squares fit for a second-degree polynomial is identical with the calculation of the two-variable equation except that x^2 is substituted for the second independent variable x_2. All the equations, (7.18) to (7.26), are applicable with x^2 substituted for x_2. There are a number of programs available to calculators for the solution of a second-degree polynomial. The additional feature of Regression-3 is that in addition to producing the constants for the second-degree polynomial it also gives the constants for the simple linear equation and for the equation with only the x^2 term, and gives the correlation coefficients for all three.

The program follows the same calculation that is used for the two-variable solution, except that it is not necessary to enter the x^2 term as that value is readily calculated from the input value of x.

To run Regression-3 for the polynomial solution, the data are entered with KEY B to differentiate from the two-variable case. With the Hewlett-Packard program the data are entered: y_i, ENTER ↑, x_i, KEY B. With the Texas Instruments program the data are entered: y_i, KEY $x \rightleftharpoons t$, x_i, KEY B. When all the data have been entered, KEY C initiates the calculation, just as with the two-variable data.

The output is the constants and correlation coefficient for the three equations:

$$y = b_0 + b_1 x$$
$$y = b_0 + b_1 x^2$$

and

$$y = b_0 + b_1 x + b_2 x^2$$

The program then gives the t value for the comparison of the third equation with the better of the other two. The following example illustrates the operation of the program.

7.5.5 Example of Regression-3, Second-Degree Polynomial

The data from Table 7.6 are entered:

	HP	TI
120	ENTER ↑	KEY $x \rightleftharpoons t$
2.11	KEY B	KEY B
122	ENTER ↑	KEY $x \rightleftharpoons t$
2.29	KEY B	KEY B
⋮		
51	ENTER ↑	KEY $x \rightleftharpoons t$
1.74	KEY B	KEY B
With all the data entered:	KEY C	KEY C

The output is:

84.3605	b_1 for the equation: $y = b_0 + b_1x$
−73.5109	b_0 for the same equation
0.8874	r: correlation coefficient
23.7383	b_1 for the equation: $y = b_0 + b_1x^2$
−1.9736	b_0 for the same equation
0.9100	r: correlation coefficient
−323.7970	b_1 for the equation: $y = b_0 + b_1x + b_2x^2$
112.3223	b_2 for the same equation
281.2296	b_0 for the same equation
0.9467	R: correlation coefficient
3.9679	t for the comparison of R with the r value from the second equation (Time:17/18)

All of the equations are statistically significant, and the t value for the comparison of the third equation with the second is also significant—probability 0.006. The experimenter would have to decide whether the improvement of the second-degree polynomial over the single-term equation is sufficient to warrant the extra complexity. The statistical analysis indicates it is a significant improvement.

The Regression-3 program provides for the calculation of the three equations. The interpretation of the results must depend on some familiarity with the experimental situation.

Table 7.7 summarizes the Regression-1, Regression-2, and Regression-3 programs.

Table 7.7 Summary of Regression-1, Regression-2, and Regression-3 Programs

	Program							
Data	**Regression-1**		**Regression-2**		**Regression-3**			
Equations	$y = b_0 + b_1 x$		$y = b_0 + b_1 \cdot F(x)$		$y = a_0 + a_1 x_1$ $y = c_0 + c_1 x_2$ $y = b_0 + b_1 x_1 + b_2 x_2$		$y = a_0 + a_1 x$ $y = c_0 + c_1 x^2$ $y = b_0 + b_1 x + b_2 x^2$	
	HP	TI	HP	TI	HP	TI	HP	TI
Input	y_i, ENTER ↑ x_i, KEY A	KEY $x \rightleftharpoons t$ KEY A	y_i, ENTER ↑ x_i, KEY A	KEY $x \rightleftharpoons t$ KEY A	y_i, ENTER ↑ x_{1i}, ENTER ↑ x_{2i}, KEY A	KEY A KEY A KEY A	y_i, ENTER ↑ x_i, KEY B	KEY $x \rightleftharpoons t$ KEY B
Run	KEY C	KEY C	KEY C	KEY C	KEY C	KEY C	KEY C	KEY C
Output	$\bar{x}$ $\bar{y}$ $\Sigma x^2 - \frac{(\Sigma x)^2}{n}$ b_1 b_0 r α $s(y)$		b_1 b_0 r α $s(y)$ $F(x) = x$ x^2 $\sqrt{x}$ e^x $1/x$ $\log(x)$		a_1 a_0 r_1 c_1 c_0 r_2 b_1 b_2 b_0 R t*		a_1 a_0 r_1 c_1 c_0 r_2 b_1 b_2 b_0 R t*	
Input	ν, ENTER ↑ r, KEY D	KEY $x \rightleftharpoons t$ KEY D						
Output	α							

* Comparing R and the larger of r_1 and r_2.

7.6 REGRESSION WITH THREE INDEPENDENT VARIABLES

The discussion of Sec. 7.5 dealing with the correlation of one dependent variable with two independent variables may be expanded to three or more independent variables. The number of possible linear correlating equations increases rapidly as the number of independent variables increase. The total number is $2^n - 1$ with n independent variables. The discussion that follows deals with three independent variables, which is the practical limit for the programmable calculators currently (1980) available. Program Regression-3, and the program of this section, Regression-4, establish a pattern which can be readily expanded to a larger number of variables if sufficient storage and program area are available.

There are programs available in larger computer systems for handling data from several independent variables. The majority of these programs correlate the dependent variable against each of the independent variables and select the one that gives the best correlation. The programs then calculate the correlation of the dependent variable against pairs of independent variables including the best of the single variables in each pair. It then goes on to using three independent variables, including the best two in each set of three, and so on. This procedure does not always give the best correlation because the best correlation with two variables may not include the variable that gave the best single variable correlation. And the best three might not include both of the best two.

The mathematics supporting this statement is beyond the scope of material covered in this book, but it involves the correlation of the independent variables with each other. An explanation can be found in Ref. 9, 10, or 13.

The Regression-4 program calculates the correlation of a dependent variable with three independent variables, each separately, with each pair, and with all three, and gives the correlation coefficient for each correlation. The same program may be expanded to more than three independent variables.

The solution of a correlation equation for one and two independent variables has been discussed. The calculation for three or more follows a similar pattern. In the solution given in the following paragraphs a general formulation is indicated. As previously stated, the Regression-4 program can handle only three independent variables.

An equation estimating the response for three or more variables is written:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots \tag{7.31}$$

The least-squares solution is obtained by minimizing the sum of squares of deviation between the estimated response of Eq. (7.31) from the observed responses:

$$\Sigma(y - b_0 - b_1x_1 - b_2x_2 - b_3x_3 - \cdots)^2 = \text{minimum} \tag{7.32}$$

The value of the constants giving the minimum value for Eq. (7.32) is obtained by differentiating with respect to each of the constants and equating the differential equations to zero. The resulting equations, after differentiating and equating to zero are as follows:

$$\begin{aligned}
b_0 n &+ b_1\Sigma x_1 &+ b_2\Sigma x_2 &+ b_3\Sigma x_3 &+ \cdots &= \Sigma y \\
b_0\Sigma x_1 &+ b_1\Sigma x_1^2 &+ b_2\Sigma x_1x_2 &+ b_3\Sigma x_1x_3 &+ \cdots &= \Sigma yx_1 \\
b_0\Sigma x_2 &+ b_1\Sigma x_1x_2 &+ b_2\Sigma x_2^2 &+ b_3\Sigma x_2x_3 &+ \cdots &= \Sigma yx_2 \\
b_0\Sigma x_3 &+ b_1\Sigma x_1x_3 &+ b_2\Sigma x_2x_3 &+ b_3\Sigma x_3^2 &+ \cdots &= \Sigma yx_3 \\
& & \cdots
\end{aligned} \tag{7.33}$$

The solution of Eq. (7.33) will give the least-squares constants for Eq. (7.31).

There are a number of ways of solving Eq. (7.33). In the Regression-4 program, the equations are modified by substituting b_0 from the first equation in each of the others, and then solving the remaining three. After the solution of the three equations for the values of b_1, b_2, and b_3, b_0 is obtained from the relation:

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - b_3\bar{x}_3 \tag{7.34}$$

which is obtained by dividing the first equation by n.

The correlation coefficient and the variance of estimate are calculated by direct extension of Eqs. (7.23) to (7.26).

7.6.1 The Regression-4 Program

Regression-4 takes input of the observations of one dependent variable and three independent variables and calculates all combinations of linear correlations of the dependent variable with the independent variables. It gives the equation constants and the correlation coefficient for each correlation. The statistical significance, the probability values, for any of the correlation coefficients can be obtained from the KEY D portion of the Regression-1 program, or from the equivalent t test of the t-3 program.

The program is given in two parts. The first part calculates the one-variable and the two-variable correlations and stores the sums for the calculation of the three-variable correlation; the second part continues the calculation without any input of data and gives the constants for the three-variable equation. The total program is too large for the HP-67 and HP-97 systems. With each part stored on a separate magnetic card, the only interruption in the output is the time needed to insert the second card and press KEY A.

The program could be placed in the TI-59 calculator as one program, but it would require two magnetic cards for storage, and would necessitate a change in the partition between storage and program memory. No time is saved in the Texas Instruments system by inserting the two program cards at the start of the calculation or by inserting the second card after the calculation is partially completed. And with the second method, more storage area is available if it is wanted. Therefore the Texas Instruments program is also presented in two parts.

The programs are run as follows: with Part 1 in the calculator, the data are entered:

	HP	TI
y_i	ENTER ↑	KEY A
x_{1i}	ENTER ↑	KEY A
x_{2i}	ENTER ↑	KEY A
x_{3i}	KEY A	KEY A
⋮		
y_n	ENTER ↑	KEY A
x_{1n}	ENTER ↑	KEY A
x_{2n}	ENTER ↑	KEY A
x_{3n}	KEY A	KEY A
When all the data are in:	KEY C	KEY C

The output from the calculation is b_1, b_0, and r for each of the single-variable equations; b_1, b_2, b_0, and R for each of the two-variable equations in the order y vs. x_1x_2, y vs. x_1x_3, and y vs. x_2x_3. The second part of the program is then put in the calculator, and KEY A gives b_1, b_2, b_3, b_0, and R for the three-variable equation.

The correlation coefficients may be compared by the t calculation of Eq. (7.29), and the probability value may be calculated from the t-1 program, KEY D; or the probability values of the individual r values can be calculated from the Regression-1 program, KEY D. This portion of the Regression-1 program could be added to the second part of Regression-4, and the probability values would be obtained directly if desired.

The calculations of Regression-4 for the one- and two-variable correlations are similar to those in Regression-3. The sums required for all the calculations are accumulated as the data are entered. The calculation of the constants for the three one-variable equations, and for the three two-variable equations are made in subroutines in the program, and the sums are shifted to the correct storage areas before each calculation.

The calculation of the three-variable equation (7.31) is the solution of the simultaneous equations (7.33). This solution is carried out in the program by the standard Doolittle method of the systematic elimination of a variable and reducing the number of equations until only one equation with one variable remains. The Texas Instruments system has a built-in matrix manipulation routine to handle up to a 9×9 matrix, but the method used in Regression-4 is faster when it is used with three simultaneous equations. The four equations of (7.33) are reduced to three by substituting the value of b_0 from the first into each of the other three.

The total calculator time for the entire program, including inserting and running the second card is 60 to 70 s, of which only about 20 are for the second card. Thus, very little time could be saved by a faster solution to the three or four simultaneous equations.

Figure 7.12 shows the flow of the program. Details and a listing are given in the Appendix.

Table 7.8 shows the location of the calculated values at the end of part two of the calculation.

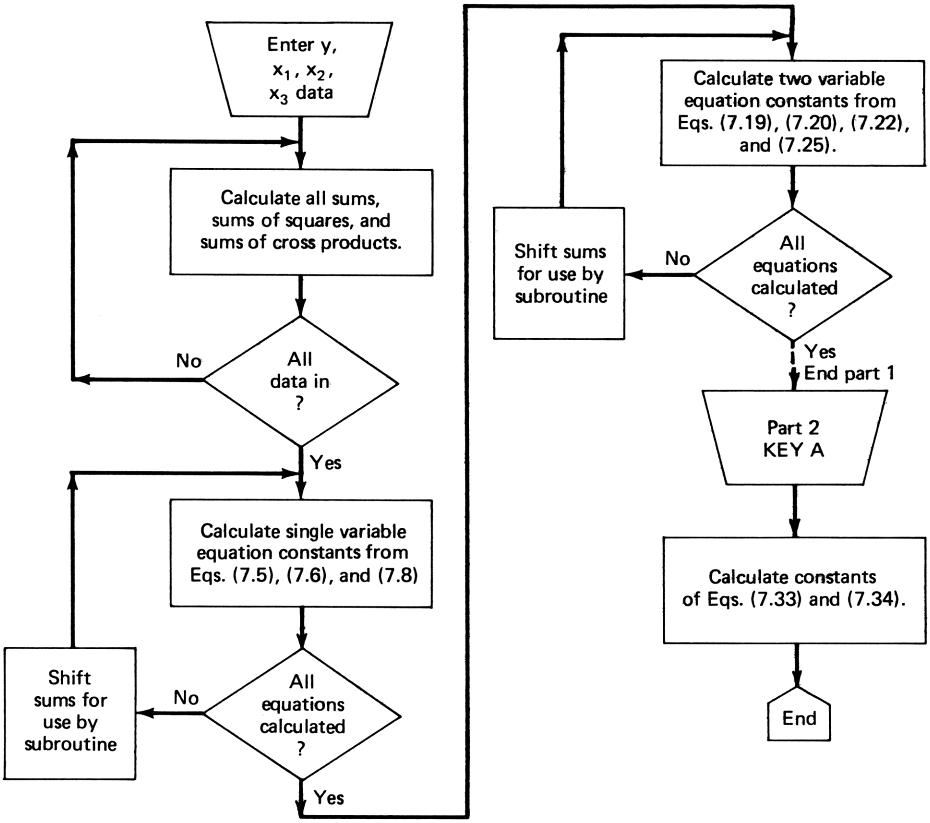


Figure 7.12 Flow diagram for the Regression-4 program.

Table 7.8 Storage Register Values after Regression-4, Part 2

	Register			Register	
Value	**HP**	**TI**	**Value**	**HP**	**TI**
Σx_1	4′	13	$\Sigma x_2x_3 - (\Sigma x_2\Sigma x_3)/n$	3	16
Σx_2	4	07	$\Sigma y^2 - (\Sigma y)^2/n$	17	05
Σx_3	0′	01	$\Sigma yx_1 - (\Sigma y\Sigma x_1)/n$	18	06
Σy	6′	04	$\Sigma yx_2 - (\Sigma y\Sigma x_2)/n$	8	09
$\Sigma x_1^2 - (\Sigma x_1)^2/n$	5′	02	$\Sigma yx_3 - (\Sigma y\Sigma x_3)/n$	12	12
$\Sigma x_2^2 - (\Sigma x_2)^2/n$	5	08	n	19	03
$\Sigma x_3^2 - (\Sigma x_3)^2/n$	1′	11	b_1	1	00
$\Sigma x_1x_2 - (\Sigma x_1\Sigma x_2)/n$	B	25	b_2	0	10
$\Sigma x_1x_3 \quad (\Sigma x_1\Sigma x_3)/n$	C	17	b_3	6	14

NOTE: Prime indicates secondary registers.

7.6.2 Example of Regression-4

Table 7.9 gives some flow rate data for a crystal line product together with moisture content, dimension ratio, and impurity content as possible variables affecting the flow rate. The data in Table 7.9 are 18 randomly selected values from a total of 48 points in the original data.[12]

Table 7.9 Data for Example 7.6.2

Flow rate y, g/s	Percent moisture x_1	Length/ breadth x_2	Percent impurity x_3
3.21	0.12	3.2	0.01
3.25	0.12	2.7	0.00
4.00	0.17	2.7	0.00
3.62	0.24	2.8	0.00
3.76	0.10	2.6	0.00
4.55	0.11	2.0	0.02
5.32	0.10	2.0	0.07
4.39	0.10	2.0	0.02
4.59	0.17	2.2	0.03
5.00	0.17	2.4	0.04
3.68	0.15	2.4	0.02
3.18	0.23	2.2	0.10
5.00	0.21	1.9	0.04
0.00	0.37	2.3	0.14
3.70	0.28	2.4	0.05
3.40	0.32	3.3	0.08
0.00	0.28	3.5	0.12
2.33	0.22	3.0	0.06

With the first part of the Regression-4 program in the calculator, the data are entered:

	HP	TI
3.21	ENTER ↑	KEY A
0.12	ENTER ↑	KEY A
3.2	ENTER ↑	KEY A
0.01	KEY A	KEY A
3.25	ENTER ↑	KEY A
0.12	ENTER ↑	KEY A
2.7	ENTER ↑	KEY A
0.00	KEY A	KEY A
⋮		
2.33	ENTER ↑	KEY A
0.22	ENTER ↑	KEY A
3.0	ENTER ↑	KEY A
0.06	KEY A	KEY A
With all the data in:	KEY C	KEY C

The output is:

$$\left.\begin{array}{rl} -11.6935 & b_1 \\ 5.7466 & b_0 \\ 0.6457 & r \end{array}\right\} \text{For correlation against moisture}$$

$$\left.\begin{array}{rl} -1.7319 & b_1 \\ 7.8863 & b_0 \\ 0.5591 & r \end{array}\right\} \text{For correlation against dimension ratio}$$

$$\left.\begin{array}{rl} -22.0747 & b_1 \\ 4.4800 & b_0 \\ 0.6378 & r \end{array}\right\} \text{For correlation against percent impurity}$$

$$\left.\begin{array}{rl} -9.3762 & b_1 \\ -1.2038 & b_2 \\ 8.3507 & b_0 \\ 0.7426 & R \end{array}\right\} \text{For correlation against moisture and dimension ratio}$$

$$\left.\begin{array}{rl} -7.0051 & b_1 \\ -12.4066 & b_2 \\ 5.3968 & b_0 \\ 0.6916 & R \end{array}\right\} \text{For correlation against moisture and percent impurity}$$

$$\left.\begin{array}{rl} -1.5382 & b_1 \\ -20.2368 & b_2 \\ 8.2950 & b_0 \\ 0.8065 & R \end{array}\right\} \text{For correlation against dimension ratio and percent impurity}$$

(Time:35/43)

That completes the output for the first part of the program. The interesting point of these data is that the best of the single-variable correlations is with moisture, a correlation coefficient of 0.6457. The best correlation with two variables does not include moisture, but is with the other two: dimension ratio and percent impurity with a correlation coefficient of 0.8065. If the moisture term had been retained, the best correlation that would be obtained would have a correlation coefficient of 0.7426.

If the three-variable correlation is to be calculated, the second part of the Regression-4 program is put into the calculator, and KEY A gives the following output:

$$\left.\begin{array}{rl} -2.6198 & b_1 \\ -1.4239 & b_2 \\ -16.7576 & b_3 \\ 8.3545 & b_0 \\ 0.8119 & R \end{array}\right\} \text{For the three-variable correlation}$$

(Time:60/70)

The significance of the different correlation coefficients can be tested with KEY D and the Regression-1 program. All of the correlations are significant. The improvement in going from one variable to two, or from two to three can be tested with the t-1 program and KEY D. The correlation coefficients of the two equations to be compared are converted to the t variable by Eq. (7.29). If this test is made, the three-variable equation is not significantly better than the best of the two, giving the probability value of 0.559. However the best of the two-variable equations is significantly better than the best single-variable equation—probability 0.006.

It is very unusual to find data in which the independent variable that gives the best single-variable correlation is not included in the best correlation with two variables. However the situation does occur, and these data were used to demonstrate such an example. The Regression-4 program permits a check on this possibility without any significant increase in calculating time.

7.7 THIRD-DEGREE POLYNOMIAL CORRELATION

Regression-4 can be used to obtain a third-degree polynomial correlation of the form:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 \tag{7.35}$$

where x^2 and x^3 are substituted for x_2 and x_3 in all the equations of the preceding sections. A relatively simple change in Regression-4 will permit the use of that program for the solution of Eq. (7.35) with the input of just the $y - x$ data. The changes in the programs are as follows:

In the Hewlett-Packard program the lines 009 to 022 are changed to the following:

```
009   STO B
010   Σ+
011   ↓
012   RCL B
013   x²
014   STO C
015   P⇌S
016   Σ+
017   RCL B
018   RCL C
019   ×
020   STO +3
```

021	RCL D
022	RCL C

Lines 023 to 025 are deleted.

In the Hewlett-Packard program it is important to delete the old lines before adding the new lines. The program as it exists fills the program memory space, and if lines are added before lines are deleted, the lines at the end of the program will be lost.

In the Texas Instruments program the changes are as follows:

Delete lines 059 to 065

Replace lines 052 to 058 with:

052	(
053	RCL
054	07
055	×
056	RCL
057	08
058	)

Delete lines 031 to 035

Replace lines 022 to 030 with:

022	STO
023	07
024	Σ+
025	INV
026	STF
027	01
028	RCL
029	07
030	$\times^2$

The changes in the TI program are given for the higher-numbered lines first because changes in the lower-numbered lines would affect the line numbering in the later part of the program and the instructions might not be clear.

The changes are all made in the first part of Regression-4. The operation of the second part is not affected by the changes. Engineers finding frequent use for the third-degree polynomial correlation could save two versions of part 1 of the Regression-4 program and use the same part 2 for both.

7.8 CORRELATION FOR A FAMILY OF CURVES

When data fall in a family of lines such as illustrated in Fig. 7.13, and the differences among the lines form some quantitative parameter, it is sometimes possible to

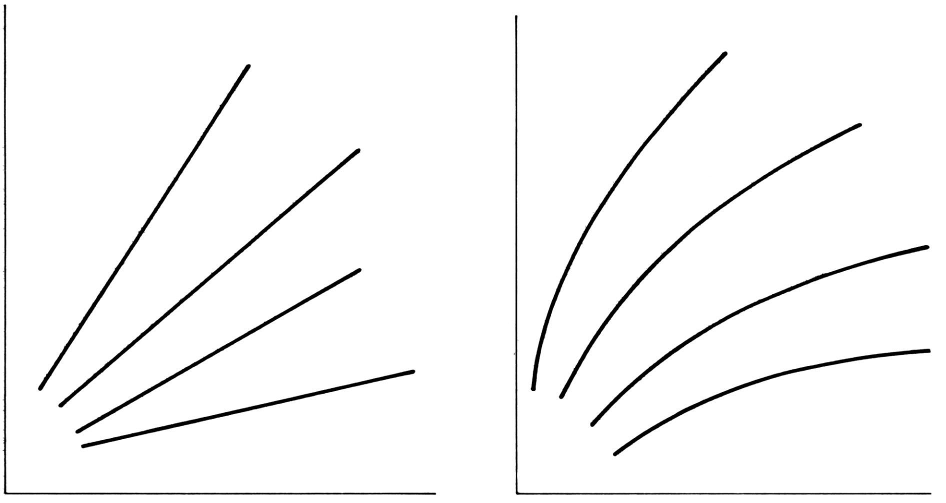


Figure 7.13 Typical curves for correlation with the Regression-5 program.

correlate the entire family of curves with a two-variable linear equation. If the individual lines can be fitted with an equation of the form:

$$y = b_0 + s \cdot x_1 \qquad \text{or} \qquad y = b_0 + s \cdot F_1(x_1) \tag{7.36}$$

and the difference in the slopes or regression coefficients s is a function of some second variable, the parameter of the family of curves, which can also be expressed in a linear form:

$$s = b_1 + b_2 \cdot x_2 \qquad \text{or} \qquad s = b_1 + b_2 \cdot F_2(x_2) \tag{7.37}$$

the two equations can be combined to give a linear equation with two variables:

$$y = b_0 + b_1 \cdot x_1 + b_2(x_1 x_2)$$

or

$$y = b_0 + b_1 \cdot F_1(x_1) + b_2 \cdot F_1(x_1) \cdot F_2(x_2) \tag{7.38}$$

The constants in Eq. (7.38) can be obtained by the same calculation used for solving the two-variable correlations in Regression-3.

In examining a family of curves, the shape of the individual curves may be fairly evident and the equation used to fit them may be selected by inspection. If there is a question about the best fit for the individual lines in the family, the Regression-2 program can be used for the data of one line. Whichever function from Regression-2 gives the best fit for one set of data may readily be substituted for x in the Regression-5 program and used for the entire family of curves. This substitution is illustrated in Example 7.8.2.2 that follows.

Selecting the function that best describes the change in slope, or the displacement from line to line in the family, is more difficult than fitting a curve to one set of data. If the parameter values increase by powers of ten, or some other constant, and the curves are evenly spaced, a log function would appear appropriate. If the slopes of a family of straight lines appear to increase in an orderly manner with a constant change in the value of the parameter, a linear relation might apply. Inasmuch as this function is difficult to judge, Regression-5 provides for testing two functions at the option of the engineer. In the program as presented, the two functions of F_2 of Eq. (7.38) are x and $\log(x)$. Either of these may be readily changed, and the second function may be omitted entirely if desired.

An alternative approach to a family of curves is to test each set of data with Regression-2 and find the best equation for each member of the family. The constants of these equations can then be correlated against the parameter, also with Regression-2. The best equation for the parameter can then be combined algebraically with the best form for the family of curves to get an equation to fit the total data. Inasmuch as all of the data must be entered at least once with either method, and entering the data is much more time-consuming than running the programs, either method should provide satisfactory results. If the shape of the family of curves is reasonably clear, Regression-5 will be faster and provide less superfluous information. If the shape of the family of curves is not evident, Regression-2 might provide useful information for comparing the member curves among themselves.

Regression-5 is given both as a useful program for correlating a family of curves, and as an example of the correlating techniques available in the programmable calculators.

7.8.1 Regression-5 Program

Regression-5 takes input of $y - x_1$ data and a parameter x_2 for a family of curves. It fits Eq. (7.38) in the least-squares manner for the functions F_1 and F_2 selected by the engineer. As presented, F_1 is simply x_1, and F_2 is x_2 and $\log(x_2)$. The function F_1 is at line 11 and the function F_2 at lines 9 and 47 in the Hewlett-Packard program and at lines 23, 31, and 90 in the Texas Instruments program. The program has facility for testing two values of F_2, but the calculation is made for only the first value unless Flag 1 is set before the program is started.

The output from the program is the values of b_1, b_2, b_0 of Eq. (7.38), and R the correlation coefficient, for one or two values of F_2, depending on whether or not Flag 1 has been set.

The calculations are the same least-squares solution used in Regression-3 for a two-variable correlation and discussed in Sec. 7.5.1. The probability associated with the correlation coefficient can be obtained with KEY D of the Regression-1 program, or the probability routine of that program could be merged with Regression-5 and the probability made part of the output.

Figure 7.14 gives the flow of the program, and a listing and detailed description are given in the Appendix.

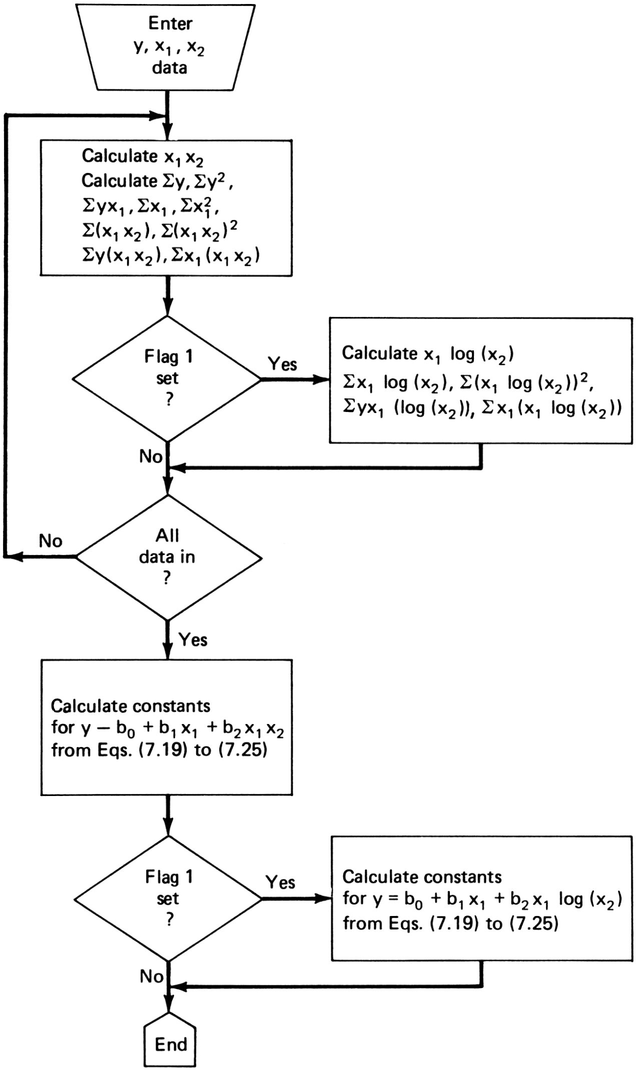


Figure 7.14 Flow diagram for the Regression-5 program.

7.8.2 Examples of Regression-5

7.8.2.1 Figure 7.15 shows a family of curves for some solids transport data.[14] The ordinate is solids transport rate, lb/(s)(ft^2), and the abscissa is the pressure drop per foot of vertical rise, lb/(ft^2)(ft). A second parameter, the gas velocity, ft/s, appears to separate the data into a series of different lines. The data are shown in Table 7.10.

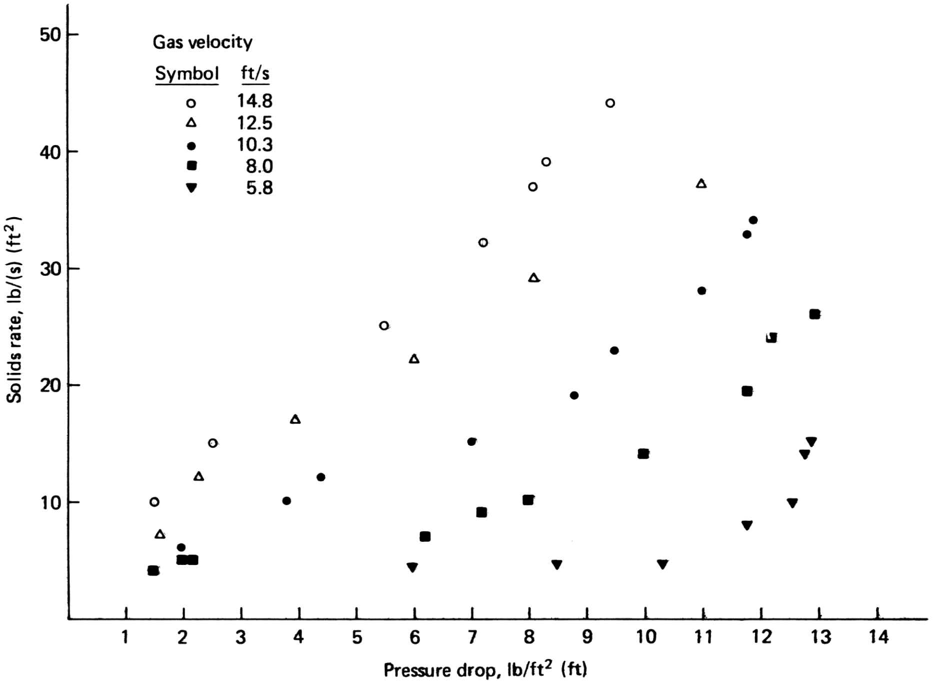


Figure 7.15 Solids transport rate as functions of pressure drop and gas velocity.

Table 7.10 Solids Transport Data of Fig. 7.15

Solids rate y, lb/(s)(ft)²	Pressure drop x_1, lb/(ft²)(ft)	Gas velocity x_2, ft/s	Solids rate y, lb/(s)(ft²)	Pressure drop x_1, lb/(ft²)(ft)	Gas velocity x_2, ft/s
10	1.5	14.8	15	2.5	14.8
25	5.5	14.8	32	7.2	14.8
37	8.1	14.8	39	8.3	14.8
44	9.4	14.8	7	1.6	12.5
12	2.3	12.5	17	4.0	12.5
22	6.0	12.5	29	8.1	12.5
37	11.0	12.5	6	2.0	10.3
10	3.8	10.3	12	4.4	10.3
15	7.0	10.3	19	8.8	10.3
23	9.5	10.3	28	11.0	10.3
33	11.8	10.3	34	11.9	10.3
4	1.5	8.0	5	2.0	8.0
5	2.2	8.0	7	6.2	8.0
9	7.2	8.0	10	8.0	8.0
14	10.0	8.0	19.5	11.8	8.0
24	12.2	8.0	26	13.0	8.0
4.5	6.0	5.8	4.5	8.5	5.8
4.5	10.3	5.8	8	11.8	5.8
14	12.8	5.8	10	12.6	5.8
			15	12.9	5.8

The data in Fig. 7.15 appear to form straight lines; therefore, Regression-5 will not be modified. Also, the parameter values increase in equal increments so the simple linear function of x_2 will be used without change. The data are entered in Regression-5:

	HP	TI
10	ENTER ↑	KEY A
1.5	ENTER ↑	KEY A
14.8	KEY A	KEY A
25	ENTER ↑	KEY A
5.5	ENTER ↑	KEY A
14.8	KEY A	KEY A
⋮		
15	ENTER	KEY A
12.9	ENTER	KEY A
5.8	KEY A	KEY A
With all the data entered:	KEY C	KEY C

The output is:

−1.6805	b_1 for equation	
0.4128	b_2 (7.38)	
0.8415	b_0	
0.9760	R: correlation coefficient	(Time:7/10)

The correlating equation is:

$$\text{Solids rate} = 0.8415 - 1.6805\,(\text{pressure drop}) + 0.4128\,(\text{pressure drop})(\text{gas velocity})$$

Or, separating the variables:

$$\text{Slope} = -1.6805 + 0.4128\,(\text{gas velocity})$$
$$\text{Solids rate} = 0.8415 + (\text{slope})(\text{pressure drop})$$

with a correlation coefficient of 0.9760—highly significant.

7.8.2.2 One final example with the family of data in Fig. 7.16 which obviously cannot be fitted with a set of straight lines. The data[15] are for treating cellulose with sulfuric acid to reduce the chain structure. The breakdown of the cellulose is determined by measuring the viscosity of the slurry. Viscosity, in centipoises, is plotted against time with the acid concentration, in mol/100 g, as the parameter. The data are given in Table 7.11.

The trend of the data is y decreasing as x increases, and the data are concave upward. From Fig. 7.6, x, $1/x$, $\log(x)$, or $\sqrt{x}$ are suggested as possible correlating

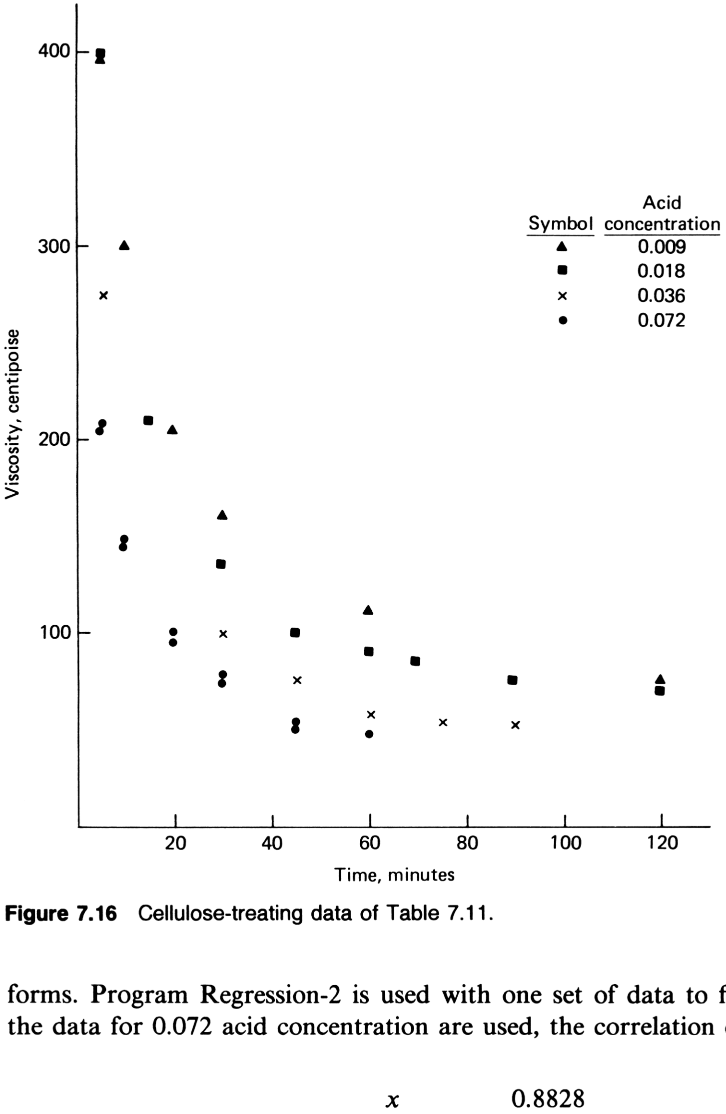


Figure 7.16 Cellulose-treating data of Table 7.11.

forms. Program Regression-2 is used with one set of data to find the best fit. If the data for 0.072 acid concentration are used, the correlation coefficients are:

x	0.8828
$1/x$	0.9872
$\log(x)$	0.9852
$\sqrt{x}$	0.9440

There is not much to choose from between the $1/x$ and the $\log(x)$ functions. If one of these functions is selected for F_1 in Eq. (7.38), that function is added to the Regression-5 program between lines 10 and 11 in the Hewlett-Packard program, or between lines 23 and 24 in the Texas Instruments program.

The selection of the function for F_2 is more a matter of judgment. The sets of

Table 7.11 Cellulose Treating Data

Viscosity y, cP	Time x_1, min	Acid concentration x_2, mol/100g	Viscosity y, cP	Time x_1, min	Acid concentration x_2, mol/100g
215	5	0.072	210	5	0.072
145	10	0.072	140	10	0.072
100	20	0.072	90	20	0.072
80	30	0.072	75	30	0.072
55	45	0.072	53	45	0.072
45	60	0.072	275	5	0.036
100	30	0.036	80	45	0.036
65	60	0.036	60	75	0.036
55	90	0.036	400	5	0.018
210	15	0.018	135	30	0.018
100	45	0.018	90	60	0.018
85	70	0.018	75	90	0.018
70	120	0.018	400	5	0.009
300	10	0.009	205	20	0.009
160	30	0.009	110	60	0.009
75	120	0.009			

data appear to be evenly spaced, and the parameter doubles for each increment indicating a log function might be suitable. Setting Flag 1 before the program is run calls in the log function for x_2 as well as the simple value of x_2. If an additional function for F_2 was wanted in place of x_2, that function could be added between lines 8 and 9 in the Hewlett-Packard program and between lines 30 and 31 in the Texas Instruments program. (Note that changing line numbers in the beginning section of a program changes all the subsequent numbers so that any changes should be made to the higher numbers first.)

The function $1/x$ is inserted in the program for F_1 as suggested above, and Flag 1 is set. The data are entered:

	HP	TI
215	ENTER ↑	KEY A
5	ENTER ↑	KEY A
0.072	KEY A	KEY A
145	ENTER ↑	KEY A
10	ENTER ↑	KEY A
0.072	KEY A	KEY A
⋮		
110	ENTER ↑	KEY A
60	ENTER ↑	KEY A
0.009	KEY A	KEY A
With all the data in:	KEY C	KEY C

The output is:

1997.	b_1 for equation: Viscosity = $b_0 + s$ (reciprocal of time)
−18072.	b_2 for equation: $s = b_1 + b_2$(acid concentration)
61.63	b_0
0.9691	*R:* correlation coefficient
−835.	b_1 for equation: Viscosity = $b_0 + s$ (reciprocal of time)
−1379.	b_2 for equation: $s = b_1 + b_2$ (log(acid concentration))
61.47	b_0
0.9730	*R:* correlation coefficient (Time:17/22)

The log function for the second variable is slightly better than the direct value. The experimenter would have to decide which function to use; both correlation coefficients are highly significant. The program provides for testing a number of possibilities, but it does not make any decisions.

7.9 FINAL REMARKS

The calculation time for most of the programs is a matter of seconds. The mean time for all of the examples in the text is 28/42 s. It usually takes much longer to enter the data than to run the programs. If data are entered incorrectly, the answers will be incorrect. The programs will process the data entered. Each of the programs has a data input routine before the calculation is started except Variance-8. It is possible to write a correction routine for each of the programs by simply reversing the signs of the input routines. No correction routine is included in any of the programs listed, but the use of correction routines is recommended if the calculations are to be made on a systematic basis.

After the data are entered but before the programs are run, the original data are stored in preassigned registers. After the programs are run, the data are usually transformed or relocated to other storage areas. It is possible to save the original data on a magnetic card before starting the calculation. If the data are saved on a magnetic card and there are doubts about the results after the program has been run, the original data may be reentered from the magnetic card and either modified or checked; additional data may be added, and the program can be run again.

The programs given provide for a variety of tests, but they give only numerical results. Numerical correlations and statistical significance calculations require interpretation and understanding.

APPENDIXES

The Appendixes contain the listings of all the calculator programs discussed in the text as well as line-by-line descriptions of their operation. Appendix A has all the Hewlett-Packard programs, and Appendix B has all the Texas Instruments programs. All examples and explanations given in the text apply to both programs.

The two calculators use different logic systems. Hewlett-Packard uses the RPN system which puts the operation after the operator and carries out each operation as it is encountered in the program. The Texas Instruments calculators use an algebraic operating system which puts the operation between the operand and the operator and which follows a hierarchy of operations ending with an equals sign. In this hierarchy the higher-order operations, logarithms, exponentiation, and trigonometric functions, are carried out first. Multiplication and division are next, and addition and subtraction are last. The entire operation up to the equals sign is evaluated by the calculator. Operations in parentheses are dealt with separately.

For example, to convert from degrees Celsius to degrees Fahrenheit, F = C (9/5) + 32, the program would be very much the same in both systems since the formula conforms to the hierarchy of operations. Starting with 100°C, the programs would look like the following:

HP	TI
100	100
ENTER ↑	×
9	9
×	÷
5	5
÷	+
32	32
+	=

However, for the reverse conversion, Fahrenheit to Celsius, the formulation is $C = (F - 32)(5/9)$, and the formula does not conform to the hierarchy. In this case the Texas Instruments is slightly different from the Hewlett-Packard version. Starting with 100°F, the programs would be:

HP	TI
100	(
ENTER ↑	100
32	−
−	32
5	)
×	×
9	5
÷	÷
	9
	=

If the parentheses were not included in the TI program, the constant 32 would be multiplied by 5/9 before the subtraction.

The data input is slightly different in the two calculators. The Hewlett-Packard can automatically store the first four data input values in what is called a "stack." The Texas Instruments calculator can store two values. Where there are more than two input values, the Texas Instruments programs store the data in registers as it is entered and call it out later in the program. The Texas Instruments calculators have more storage areas available and can readily be programmed to handle more data than the Hewlett-Packard calculators.

In moving from one location in a program to another, the Hewlett-Packard calculators search down the program until the designated label is found. With the Hewlett-Packard programs it is possible to use the same label more than once provided the one sought is further on in the program. The Texas Instruments start each search at the start of the program, and so, a label in that system cannot be used more than once. This situation offers no restriction on the programming, since any one of 72 keys can be used as a label; and the Texas Instruments program can also be directed to other sections of the program by line numbers.

In the programs in this text, with the Hewlett-Packard system, odd-number labels are used for designations further on in the program, and even-number and letter designations are used when the program is directed to a section earlier in the operation. With the Texas Instruments program, the labels are used in order from the keyboard, and no references are made to program line numbers.

The Hewlett-Packard calculators have an automatic storage-area shift key that moves 10 storage areas at a time. This operation is used in some of the programs where the same calculation is done more than once on values that have been stored. When the similar operations are carried out in the Texas Instruments programs, the values are shifted individually before the routine calculation is repeated.

In the programs given in this text there are no operations in one system that cannot be readily translated into the other system.

The programs that follow are:

1. Statistical parameter program
2. Normal probability program
3. Binomial probability program
4. Poisson probability program
5. t-1 program: testing hypotheses about means
6. t-2 program: confidence intervals and sample size
7. χ^2-1 program: ratio test
8. χ^2-2 program: equal expectation test
9. χ^2-3 program: unequal expectation test
10. χ^2-4 program: contingency tables
11. χ^2-5 program: χ^2 probability
12. Variance-1: comparison of variance with a standard
13. Variance-2: compare two variances
14. Variance-3: Bartlett χ^2 test for variance homogeneity
15. Variance-4: one-factor analysis of variance
16. Variance-5: two-factor analysis of variance, no replication
17. Variance-6: two-factor analysis of variance, with replication
18. Variance-7: Latin square analysis of variance
19. Variance-8: 2^n factorial analysis of variance
20. Variance-9: F probability calculation
21. Regression-1: straight-line regression
22. Regression-2: curved-line regression
23. Regression-3: two-variable regression
24. Regression-4: multiple-variable regression
25. Regression-5: family of curves

A

HEWLETT-PACKARD CALCULATOR PROGRAMS

Table A.1 Statistical Parameter Program (HP-97)

For individual values: x_i, KEY A   Run: KEY C
For weighted values: f_i, ENTER ↑, x_i, KEY B   Run: KEY C

Output: Arithmetic mean, midrange, geometric mean, harmonic mean, log mean, estimated standard deviation, data standard deviation, estimated variance, data variance, range, minimum value, maximum value, Σx (or Σfx), Σx^2 (or Σfx^2), number of groups if applicable, total number of data points

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLB	21 12	036	F1?	16 23 01	071	×	-35
002	SF0	16 21 00	037	GTO3	22 03	072	ST+0	35-55 00
003	*LBLA	21 11	038	RCL1	36 01	073	*LBL3	21 03
004	F2?	16 23 02	039	1/X	52	074	ISZI	16 26 46
005	GTO1	22 01	040	ST+8	35-55 08	075	RTN	24
006	CLRG	16-53	041	RCLA	36 11	076	*LBL2	21 02
007	P⇄S	16-51	042	X<0?	16-45	077	CF0	16 22 00
008	CLRG	16-53	043	GTO3	22 03	078	RCL7	36 07
009	STOA	35 11	044	RCL1	36 01	079	RCL5	36 05
010	STOB	35 12	045	LN	32	080	RCL9	36 09
011	*LBL1	21 01	046	ST+0	35-55 00	081	P⇄S	16-51
012	SF2	16 21 02	047	GTO3	22 03	082	STO9	35 09
013	STO1	35 01	048	*LBL1	21 01	083	R↓	-31
014	R↓	-31	049	RCLC	36 13	084	STO5	35 05
015	STOC	35 13	050	ST+9	35-55 09	085	R↓	-31
016	RCL1	36 01	051	RCL1	36 01	086	STO4	35 04
017	X=0?	16-43	052	×	-35	087	RCLI	36 46
018	SF1	16 21 01	053	STO6	35 06	088	STOE	35 15
019	RCLA	36 11	054	ST+7	35-55 07	089	P⇄S	16-51
020	X>Y?	16-34	055	RCL1	36 01	090	*LBLC	21 13
021	GTO1	22 01	056	×	-35	091	F0?	16 23 00
022	RCLB	36 12	057	ST+5	35-55 05	092	GTO2	22 02
023	RCL1	36 01	058	F1?	16 23 01	093	RCLB	36 12
024	X≤Y?	16-35	059	GTO3	22 03	094	RCLA	36 11
025	GTO3	22 03	060	RCL1	36 01	095	-	-45
026	RCL1	36 01	061	1/X	52	096	STO9	35 09
027	STOB	35 12	062	RCLC	36 13	097	RCLB	36 12
028	GTO3	22 03	063	×	-35	098	RCLA	36 11
029	*LBL1	21 01	064	ST+8	35-55 08	099	X=0?	16-43
030	RCL1	36 01	065	RCLA	36 11	100	GTO1	22 01
031	STOA	35 11	066	X<0?	16-45	101	X<0?	16-45
032	*LBL3	21 03	067	GTO3	22 03	102	GTO1	22 01
033	F0?	16 23 00	068	RCL1	36 01	103	÷	-24
034	GTO1	22 01	069	LN	32	104	LN	32
035	Σ+	56	070	RCLC	36 13	105	÷	-24

Table A.1 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
106	STO4	35 04	124	RCL8	36 08	142	X^2	53
107	*LBL1	21 01	125	RCLI	36 46	143	STO7	35 07
108	RCLB	36 12	126	÷	-24	144	RCLI	36 46
109	RCLA	36 11	127	1/X	52	145	1	01
110	+	-55	128	STO3	35 03	146	-	-45
111	2	02	129	RCLA	36 11	147	×	-35
112	÷	-24	130	X<0?	16-45	148	RCLI	36 46
113	STO1	35 01	131	GTO1	22 01	149	÷	-24
114	P⇄S	16-51	132	RCL0	36 00	150	STO8	35 08
115	RCL9	36 09	133	RCLI	36 46	151	√X	54
116	STOI	35 46	134	÷	-24	152	STO6	35 06
117	RCL4	36 04	135	e^x	33	153	CF1	16 22 01
118	STOC	35 13	136	STO2	35 02	154	CF2	16 22 02
119	RCL5	36 05	137	*LBL1	21 01	155	PREG	16-13
120	STOD	35 14	138	$\bar{x}$	16 53	156	RTN	24
121	P⇄S	16-51	139	STO0	35 00	157	R/S	51
122	F1?	16 23 01	140	S	16 54			
123	GTO1	22 01	141	STO5	35 05			

A.1 DETAILS OF STATISTICAL PARAMETER PROGRAM

Lines 1 to 10 initiate the program. Flag 0 is set at line 2 if weighted data are to be calculated as indicated by starting with KEY B. Flag 2 is tested at line 4 to establish whether the data entry is the first. If so, all the registers are cleared. Lines 8 and 9 store the first data entry in Registers A and B.

Lines 11 to 18 set Flag 2 at line 12 after the first data entry and store the data. The entry is checked to determine if it is zero. If so Flag 1 is set.

Lines 19 to 31 check the input data against the previous minimum and maximum values stored in Registers A and B. If the new data is less than the first or greater than the second, a switch of values is made and the new minimum or maximum is stored.

Lines 32 to 47 carry out the accumulation of the data if single values are entered. If the data are weighted, the program goes to line 48. Line 35 uses KEY $\Sigma+$ to calculate Σx, Σx^2, and the number of data entries. Line 36 checks whether Flag 1 is set indicating a zero entry. If so, the balance of the calculations are bypassed. If not, $1/x$ is calculated at line 39 and accumulated in Register 8. Lines 41 to 43 check whether the data entry is negative. If so, the balance of the calculations are bypassed. If not, lines 44 to 46 calculate the natural log of x and accumulate the sum in Register 0.

Lines 48 to 72 accumulate the values for the weighted data calculations. Σf is stored in Register 9 at line 50, Σfx is stored in Register 7 at line 54, and Σfx^2 is stored in Register 5 at line 57. A check is made at line 58 for a zero data entry. If there was one, the balance of the calculations are bypassed. If not $\Sigma f/x$ is stored in Register 8, and if the data is not negative, $\Sigma f \cdot \ln(x)$ is stored in Register 0.

Lines 73 to 75 increment Register I for each calculation to count the number of groups for the weighted data calculation. The program then stops for further data entry.

Lines 76 to 89 start the calculation for the statistical parameters for the weighted data case. This section of the program shifts the accumulated sums from the weighted calculation to the same storage areas that are used for the individual data calculation so the same routines may be used for both. Lines 76 to 89 are not used if the data entries were individual values.

Lines 90 to 92 are the start of the calculation portion of the program. If Flag 0 has been set, indicating weighted data entry, the program goes back to line 76, if not, it goes on to line 93.

Lines 93 to 96 calculate the range and store the result in Register 9.

Lines 97 to 106 calculate the log mean and store the value in Register 4. A check is first made to determine if the minimum value is zero or less. If so, the log mean is not calculated.

Lines 107 to 113 calculate the midrange and store the value in Register 1.

Lines 114 to 121 recover the values of n (or Σf), Σx (or Σfx), and Σx^2 (or Σfx^2) and store these values in Registers I, C, and D.

Lines 122 to 136 calculate the harmonic mean and the geometric mean. First a check is made at line 122 to determine if there was a zero value entered. If so, the program goes on to line 137. The harmonic mean is calculated at line 127 and stored in Register 3. A check is made at line 130 to determine if a negative value was entered. If so, the program bypasses the geometric mean calculation. If not, the geometric mean is calculated at line 135 and stored in Register 2.

Lines 137 to 152 use the built-in routines to calculate the arithmetic mean and the estimated standard deviation at lines 138 and 140. These routines work for the weighted data because of the shifting done back at lines 76 to 89. The estimated variance is calculated by squaring the estimated standard deviation at line 142. The data variance is calculated from the estimated variance by multiplying by $(n - 1)/n$ at line 149. The data standard deviation is the square root of the data variance, line 151.

Lines 153 to 156 The flags are cleared at lines 153 and 154 and all the results are printed at line 155. The program ends at line 156.

Table A.2 Normal Probability Program (HP-97)

Input	Output
Mean, ENTER ↑, Std. Dev., ENTER ↑, x_1, ENTER ↑, x_2 KEY A:	Probability ($x_1 \leqq x \leqq x_2$)
Mean, ENTER ↑, Std. Dev., ENTER ↑, x, KEY C:	Probability ($\geqq x$)
Mean, ENTER ↑, Std. Dev., ENTER ↑, x, KEY E:	Probability ($\leqq x$)
Mean, ENTER ↑, Std. Dev., ENTER ↑, P, KEY c:	x such that Pr ($\geqq x$) = P
Mean, ENTER ↑, Std. Dev., ENTER ↑, P, KEY e:	x such that Pr ($\leqq x$) = P

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	035	GTOB	22 12	069	2	02
002	SF1	16 21 01	036	GTO7	22 07	070	7	07
003	STOD	35 14	037	*LBLB	21 12	071	×	-35
004	R↓	-31	038	F0?	16 23 00	072	1	01
005	GTOC	22 13	039	GTOD	22 14	073	+	-55
006	*LBLE	21 15	040	SF0	16 21 00	074	1/X	52
007	SF2	16 21 02	041	STO3	35 03	075	STOE	35 15
008	*LBLC	21 13	042	RCLD	36 14	076	.	-62
009	GSBa	23 16 11	043	STO4	35 04	077	9	09
010	*LBLb	21 16 12	044	RCLA	36 11	078	3	03
011	CHS	-22	045	GTOb	22 16 12	079	7	07
012	RCL4	36 04	046	*LBLD	21 14	080	3	03
013	+	-55	047	ST-3	35-45 03	081	×	-35
014	STO9	35 09	048	RCL3	36 03	082	.	-62
015	RCLB	36 12	049	ABS	16 31	083	1	01
016	÷	-24	050	GTO7	22 07	084	2	02
017	GSB4	23 04	051	*LBL4	21 04	085	0	00
018	RCL9	36 09	052	STO1	35 01	086	2	02
019	X>0?	16-44	053	X^2	53	087	CHS	-22
020	GTO2	22 02	054	2	02	088	+	-55
021	F2?	16 23 02	055	÷	-24	089	RCLE	36 15
022	GTO3	22 03	056	CHS	-22	090	×	-35
023	*LBL1	21 01	057	e^x	33	091	.	-62
024	1	01	058	Pi	16-24	092	4	04
025	RCL0	36 00	059	2	02	093	3	03
026	-	-45	060	×	-35	094	6	06
027	STO0	35 00	061	√X	54	095	2	02
028	GTO3	22 03	062	÷	-24	096	+	-55
029	*LBL2	21 02	063	STO2	35 02	097	RCLE	36 15
030	F2?	16 23 02	064	RCL1	36 01	098	×	-35
031	GTO1	22 01	065	ABS	16 31	099	RCL2	36 02
032	*LBL3	21 03	066	.	-62	100	×	-35
033	RCL0	36 00	067	3	03	101	STO0	35 00
034	F1?	16 23 01	068	3	03	102	RTN	24

Table A.2 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
103	*LBLe	21 16 15	142	2	02	181	1	01
104	SF2	16 21 02	143	.	-62	182	+	-55
105	*LBLc	21 16 13	144	5	05	183	÷	-24
106	GSBa	23 16 11	145	1	01	184	CHS	-22
107	RCL4	36 04	146	5	05	185	RCL1	36 01
108	.	-62	147	5	05	186	+	-55
109	5	05	148	1	01	187	F2?	16 23 02
110	X>Y?	16-34	149	7	07	188	GTO8	22 08
111	GTO5	22 05	150	+	-55	189	F0?	16 23 00
112	1	01	151	RCL1	36 01	190	GTO9	22 09
113	RCL4	36 04	152	.	-62	191	*LBL6	21 06
114	-	-45	153	0	00	192	RCLB	36 12
115	STO4	35 04	154	0	00	193	×	-35
116	SF0	16 21 00	155	1	01	194	RCLA	36 11
117	*LBL5	21 05	156	3	03	195	+	-55
118	RCL4	36 04	157	0	00	196	GTO7	22 07
119	X^2	53	158	8	08	197	*LBL8	21 08
120	1/X	52	159	×	-35	198	F0?	16 23 00
121	LN	32	160	.	-62	199	GTO6	22 06
122	√X	54	161	1	01	200	*LBL9	21 09
123	STO1	35 01	162	8	08	201	RCLB	36 12
124	.	-62	163	9	09	202	×	-35
125	0	00	164	2	02	203	RCLA	36 11
126	1	01	165	6	06	204	X⇄Y	-41
127	0	00	166	9	09	205	-	-45
128	3	03	167	+	-55	206	*LBL7	21 07
129	2	02	168	RCL1	36 01	207	PRTX	-14
130	8	08	169	×	-35	208	CF0	16 22 00
131	×	-35	170	1	01	209	CF1	16 22 01
132	.	-62	171	.	-62	210	RTN	24
133	8	08	172	4	04	211	*LBLa	21 16 11
134	0	00	173	3	03	212	STOC	35 13
135	2	02	174	2	02	213	STO4	35 04
136	8	08	175	7	07	214	R↓	-31
137	5	05	176	8	08	215	STOB	35 12
138	3	03	177	8	08	216	R↓	-31
139	+	-55	178	+	-55	217	STOA	35 11
140	RCL1	36 01	179	RCL1	36 01	218	RTN	24
141	×	-35	180	×	-35	219	R/S	51

A.2 DETAILS OF NORMAL PROBABILITY PROGRAM

The normal probability program is in two parts. Lines 1 to 102 form the first part which calculates the probability associated with a given mean, standard deviation, and variable value. The second part, lines 103 to 205, take input of a probability value and calculate the corresponding normal deviate. Lines 206 to 218 are common to both and are the output routine and the data storage routine.

Lines 1 to 8 establish which type of probability is to be calculated: Pr($\geqq x$), Pr($\leqq x$), or Pr($x_1 \leqq x \leqq x_2$). If the last probability is wanted, KEY A, Flag 1 is set. If the second probability is wanted, KEY E, Flag 2 is set. If the first probability is wanted, no flags are set and the program proceeds to line 9.

Line 9 directs the program to Subroutine a at the end of the program to store the data input. This same subroutine is used for the second part of the program.

Lines 10 to 16 standardize the input variable using Eq. (3.7) and store in Register 9 the information whether x is greater than or less than the mean.

Line 17 directs the program to the probability calculation routine starting at line 51. The resulting probability is stored in Register 0.

Lines 18 to 36 check whether Flag 2 is set for a probability $\leqq x$, and whether x is greater than or less than the mean. The output is either α, the calculated probability, or $1 - \alpha$ depending on the problem. The following table shows the possibilities. Lines 18 to 36 select the proper one. Line 34 checks whether Flag 1 was set, indicating that two probabilities are to be calculated. If so the program goes to line 37. If not, the program goes to the output routine at line 206.

Probability	Pr($\leqq x$) Flag 2 set	Pr($\geqq x$)
$m > x$	α	$1-\alpha$
$m < x$	$1-\alpha$	α

Lines 37 to 45 are used when two probabilities are calculated for the solution required with KEY A. The first time around, the probability is stored in Register 3, the second variable value is recalled from Register D, and the program returns to line 10 for a second calculation. The second time around, indicated by Flag 0 being set, the program goes on to line 46.

Lines 46 to 50 obtain the difference between the two probabilities and direct the program to the output routine. The absolute value is used.

Lines 51 to 102 carry out the probability calculation following Eqs. (3.14), (3.11), and (3.12).

Lines 103 to 105 establish whether the input probability is associated with a minimum value or with a maximum value: KEY e for the latter, which sets Flag 2, or KEY c for the former.

Line 106 directs the program to Subroutine a at the end to store the data input.

Lines 107 to 116 establish whether the input probability was greater than 0.5. If so, the value is changed to unity minus the input value and Flag 0 is set.

Lines 117 to 186 calculate the standard deviate value associated with the input probability. Equations (3.17) and (3.16) are used.

Lines 187 to 205 check whether Flag 2 is set for a probability $\leqq x$, and whether Flag 0 is set for an input probability greater than 0.5. The program then makes the selection, following Table 3.3, for the proper solution.

Lines 206 to 210 are the output (print) routine for both types of calculation.

Lines 211 to 218 are Subroutine a for the input data storage used for both types of solution.

At the completion of the program, the input values are in Registers A, B, C, and D in order of input. This information may be recalled for verification if desired.

Table A.3 Binomial Probability Program (HP-97)

n	ENTER ↑	KEY A gives Pr(r)	n	ENTER ↑	KEY c gives r such that Pr(≧r) = P
p	ENTER ↑	KEY B gives Pr(>r)	p	ENTER ↑	
r		KEY C gives Pr(≧r)	P		KEY e gives r such that Pr(≦r) = P
		KEY D gives Pr(<r)			
		KEY E gives Pr(≦r)			

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	037	SF0	16 21 00	073	RCL8	36 08
002	GSB1	23 01	038	*LBLe	21 16 15	074	-	-45
003	RCL1	36 01	039	SF1	16 21 01	075	X<0?	16-45
004	PRTX	-14	040	X<0?	16-45	076	GTO9	22 09
005	RTN	24	041	GTO9	22 09	077	STO7	35 07
006	*LBLB	21 12	042	1	01	078	RCLA	36 11
007	SF2	16 21 02	043	X⇄Y	-41	079	Y^x	31
008	GSB1	23 01	044	X>Y?	16-34	080	STO1	35 01
009	RCL2	36 02	045	GTO9	22 09	081	X=0?	16-43
010	PRTX	-14	046	X⇄Y	-41	082	GTOa	22 16 11
011	RTN	24	047	R↓	-31	083	STO5	35 05
012	*LBLC	21 13	048	*LBL1	21 01	084	RCLE	36 15
013	GSB1	23 01	049	CLRG	16-53	085	RCLI	36 46
014	RCL3	36 03	050	STOC	35 13	086	X=Y?	16-33
015	PRTX	-14	051	STO9	35 09	087	GTO6	22 06
016	RTN	24	052	R↓	-31	088	*LBL4	21 04
017	*LBLD	21 14	053	STOB	35 12	089	RCL5	36 05
018	GSB1	23 01	054	STO8	35 08	090	STO4	35 04
019	RCL4	36 04	055	R↓	-31	091	RCL1	36 01
020	PRTX	-14	056	STOA	35 11	092	RCL8	36 08
021	RTN	24	057	*LBL2	21 02	093	×	-35
022	*LBLE	21 15	058	RCL9	36 09	094	RCL7	36 07
023	GSB1	23 01	059	STOE	35 15	095	÷	-24
024	RCL5	36 05	060	RCLA	36 11	096	RCLA	36 11
025	PRTX	-14	061	FRC	16 44	097	RCLI	36 46
026	RTN	24	062	X≠0?	16-42	098	-	-45
027	*LBL1	21 01	063	GTO9	22 09	099	×	-35
028	X<0?	16-45	064	LSTX	16-63	100	ISZI	16 26 46
029	GTO9	22 09	065	RCLE	36 15	101	RCLI	36 46
030	FRC	16 44	066	X>Y?	16-34	102	÷	-24
031	X>0?	16-44	067	GTO9	22 09	103	STO1	35 01
032	GTO9	22 09	068	F2?	16 23 02	104	ST+5	35-55 05
033	R↓	-31	069	GSB7	23 07	105	F1?	16 23 01
034	LSTX	16-63	070	F1?	16 23 01	106	GSB3	23 03
035	GTO1	22 01	071	GSB1	23 01	107	1	01
036	*LBLc	21 16 13	072	1	01	108	RCL5	36 05

Table A.3 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
109	X>Y?	16-34	144	RCL5	36 05	179	*LBL5	21 05
110	GTO5	22 05	145	PRTX	-14	180	RCLI	36 46
111	RCLE	36 15	146	RCLI	36 46	181	PRTX	-14
112	RCLI	36 46	147	PRTX	-14	182	GTO6	22 06
113	X=Y?	16-33	148	SPC	16-11	183	*LBL7	21 07
114	GTO6	22 06	149	1	01	184	X=Y?	16-33
115	GTO4	22 04	150	-	-45	185	GTO9	22 09
116	*LBL6	21 06	151	RCL4	36 04	186	RTN	24
117	1	01	152	PRTX	-14	187	*LBLa	21 16 11
118	RCL4	36 04	153	R↓	-31	188	1	01
119	-	-45	154	GTO3	22 03	189	RCL8	36 08
120	STO3	35 03	155	*LBL1	21 01	190	-	-45
121	RCL1	36 01	156	1	01	191	STO8	35 08
122	-	-45	157	RCL9	36 09	192	RCLA	36 11
123	STO2	35 02	158	-	-45	193	RCL9	36 09
124	RTN	24	159	STO9	35 09	194	-	-45
125	*LBL1	21 01	160	1	01	195	STO9	35 09
126	F0?	16 23 00	161	RCL5	36 05	196	GSB2	23 02
127	GSB1	23 01	162	-	-45	197	RCL3	36 03
128	RCLA	36 11	163	PRTX	-14	198	RCL5	36 05
129	STOE	35 15	164	RCLI	36 46	199	STO3	35 03
130	RTN	24	165	1	01	200	R↓	-31
131	*LBL1	21 01	166	+	-55	201	STO5	35 05
132	1	01	167	PRTX	-14	202	RCL4	36 04
133	RCL9	36 09	168	SPC	16-11	203	RCL2	36 02
134	-	-45	169	1	01	204	STO4	35 04
135	STO9	35 09	170	RCL4	36 04	205	R↓	-31
136	RTN	24	171	-	-45	206	STO2	35 02
137	*LBL3	21 03	172	PRTX	-14	207	RTN	24
138	RCL9	36 09	173	RCLI	36 46	208	*LBL9	21 09
139	RCL5	36 05	174	*LBL3	21 03	209	CF0	16 22 00
140	X≤Y?	16-35	175	PRTX	-14	210	CF1	16 22 01
141	RTN	24	176	CF0	16 22 00	211	0	00
142	F0?	16 23 00	177	CF1	16 22 01	212	÷	-24
143	GTO1	22 01	178	R/S	51	213	R/S	51

A.3 DETAILS OF THE BINOMIAL PROBABILITY PROGRAM

Lines 1 to 26 establish which type of solution is requested with variable value input. The calculation is the same in all cases and is carried out by a subroutine so that after the calculation is completed the program returns to the starting location. Each section of lines 1 to 26 recalls a different storage register to print the specified output. At line 7, Flag 2 is set if Pr($>r$), is requested. Later in the program a check is made to ensure that r is not equal to n.

Lines 27 to 35 check the variable value input. If it is negative, or not an integer, an error message is printed.

Lines 36 to 47 start the program if there is probability input and a value of the variable is to be determined. Flag 1 is set to indicate that there is probability input, and Flag 0 is set if a "greater than" probability is desired. Lines 40 to 45 check the input probability. If it is negative or greater than 1, an error message is printed.

Lines 48 to 56 start the calculation in all cases. The storage registers are cleared from any previous calculation results, and the input values are stored. The actual calculation starts at line 57.

Lines 57 to 69 are further checks on the input data; if n is not an integer, or if r is greater than n, or if $r = n$ when a "greater than" probability is sought, an error message is printed.

Lines 70 and 71 check whether there is probability input. If so, the program goes to line 125 to make an adjustment.

Lines 72 to 82 check whether the value of p is greater than unity. If so, an error message is printed. If not, Pr(0) is calculated, to start the summation. Line 81 checks whether the answer is zero, that is, below the limits of the calculator. If so, the program goes to line 187 to change $1 - p$ for p and $n - r$ for r; it then returns to line 57.

Lines 83 to 115 are the summation calculation routine following Eq. (3.20). Each individual calculation is stored in Register 1 as it is made, and the summation is stored in Register 5. At line 105, a check is made to establish whether the probability calculated has to be checked against an input probability (Flag 1). If so, the program goes to line 137 to make the check. **Lines 107 to 110** check to prevent the probability from exceeding 1.0, due to calculator rounding errors. If the cumulative probability exceeds 1.0, the program goes to line 179. **Lines 111 to 113** check to see if the summation calculation has reached the value of the input variable r; if so, the program goes to line 116; if not, the program goes back to line 88 for the calculation of another term.

Lines 116 to 124 are the end operation. The different probabilities are stored in the proper registers, and the program returns to the starting section to print the output.

Lines 125 to 136 are the routine that adjusts the calculation for probability input. With r input, the summation continues until the probability for r is reached. With probability input, n is substituted for r so the summation will continue until the input P value is reached. The calculation, in all cases, is the probability from zero to some value of r. If the probability beyond r is desired, $1 - P$ is substituted for P and the calculation is made in the usual way. The substitution is made at lines 131 to 136 if it is called for.

Lines 137 to 141 are the routine to check whether the input probability has been reached. If not, the program returns for the calculation of another term. If so, the program goes on to

Lines 142 to 154 to print the results if a "less than" probability was asked for, or to

Lines 155 to 178 if a "greater than" probability was sought. The calculation was continued until the input probability was first exceeded. This value, with the corresponding value of r, and also the probability for the preceding term, which had been stored in Register 4, together with the value of $r - 1$, are printed.

Lines 179 to 182 If the cumulative probability has exceeded 1.0, due to calculator rounding errors, the program prints the value of r up to this point, and then goes on to continue the calculation.

Lines 183 to 186 are the routine to check if $r = n$ when Pr($>r$) was called for: KEY B.

Lines 187 to 207 are the routine for switching $1 - p$ for p and $n - r$ for r if the calculation of Pr(0) was zero. After the calculation is completed, lines 197 to 207 return the original values so the output corresponds to the original input.

Lines 208 to 212 are the routine to generate the error message by dividing by zero.

At the end of the program, the input and the calculated probabilities are stored as shown in Table 3.6.

Table A.4 Poisson Probability Program (HP-97)

m	ENTER ↑	KEY A gives Pr(r)	m	ENTER ↑	KEY c gives r such that Pr(≧r) = P
		KEY B gives Pr(>r)			
r		KEY C gives Pr(≧r)	P		
		KEY D gives Pr(<r)			KEY e gives r such that Pr(≦r) = P
		KEY E gives Pr(≦r)			

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	037	GSB5	23 05	073	STO1	35 01
002	GSB1	23 01	038	2	02	074	ST+5	35-55 05
003	RCL1	36 01	039	2	02	075	F1?	16 23 01
004	PRTX	-14	040	7	07	076	GSB7	23 07
005	RTN	24	041	STO9	35 09	077	RCLI	36 46
006	*LBLB	21 12	042	RCLA	36 11	078	RCLB	36 12
007	GSB1	23 01	043	X≤Y?	16-35	079	X=Y?	16-33
008	RCL2	36 02	044	GTO2	22 02	080	GTO1	22 01
009	PRTX	-14	045	RCL9	36 09	081	RCL5	36 05
010	RTN	24	046	÷	-24	082	1	01
011	*LBLC	21 13	047	1	01	083	EEX	-23
012	GSB1	23 01	048	-	-45	084	9	09
013	RCL3	36 03	049	STO8	35 08	085	6	06
014	PRTX	-14	050	RCL9	36 09	086	X≤Y?	16-35
015	RTN	24	051	*LBL2	21 02	087	GSB3	23 03
016	*LBLD	21 14	052	CHS	-22	088	GTO4	22 04
017	GSB1	23 01	053	e^x	33	089	*LBL1	21 01
018	RCL4	36 04	054	STO1	35 01	090	RCL8	36 08
019	PRTX	-14	055	STO5	35 05	091	X>0?	16-44
020	RTN	24	056	RCLB	36 12	092	GSB3	23 03
021	*LBLE	21 15	057	X=0?	16-43	093	1	01
022	GSB1	23 01	058	GTO1	22 01	094	RCL4	36 04
023	RCL5	36 05	059	F1?	16 23 01	095	-	-45
024	PRTX	-14	060	GTO7	22 07	096	STO3	35 03
025	RTN	24	061	FRC	16 44	097	RCL1	36 01
026	*LBLc	21 16 13	062	X>0?	16-44	098	-	-45
027	SF0	16 21 00	063	GTO9	22 09	099	STO2	35 02
028	*LBLe	21 16 15	064	R↓	-31	100	RTN	24
029	SF1	16 21 01	065	*LBL4	21 04	101	*LBL5	21 05
030	*LBL1	21 01	066	ISZI	16 26 46	102	1	01
031	CLRG	16-53	067	RCLA	36 11	103	RCLE	36 15
032	STOB	35 12	068	RCLI	36 46	104	-	-45
033	STOE	35 15	069	÷	-24	105	STOE	35 15
034	R↓	-31	070	RCL1	36 01	106	RTN	24
035	STOA	35 11	071	ST+4	35-55 04	107	*LBL7	21 07
036	F0?	16 23 00	072	×	-35	108	RCLB	36 12

Table A.4 *(Cont.)*

Step	Key	Code	Step	Key	Code	Step	Key	Code
109	1	01	132	-	-45	155	X>0?	16-44
110	X≤Y?	16-35	133	PRTX	-14	156	GTO1	22 01
111	GTO9	22 09	134	RCLI	36 46	157	X=0?	16-43
112	RCLE	36 15	135	1	01	158	GTO1	22 01
113	RCL5	36 05	136	+	-55	159	1	01
114	X≤Y?	16-35	137	PRTX	-14	160	+	-55
115	GTO4	22 04	138	SPC	16-11	161	STO7	35 07
116	F0?	16 23 00	139	1	01	162	*LBL1	21 01
117	GTO1	22 01	140	RCL4	36 04	163	RCL7	36 07
118	RCL5	36 05	141	-	-45	164	RCL9	36 09
119	PRTX	-14	142	PRTX	-14	165	×	-35
120	RCLI	36 46	143	RCLI	36 46	166	CHS	-22
121	PRTX	-14	144	*LBL5	21 05	167	e^x	33
122	SPC	16-11	145	PRTX	-14	168	ST×1	35-35 01
123	1	01	146	CF0	16 22 00	169	ST×4	35-35 04
124	-	-45	147	CF1	16 22 01	170	ST×5	35-35 05
125	RCL4	36 04	148	R/S	51	171	RTN	24
126	PRTX	-14	149	*LBL3	21 03	172	*LBL9	21 09
127	R↓	-31	150	RCL8	36 08	173	CF0	16 22 00
128	GTO5	22 05	151	1	01	174	CF1	16 22 01
129	*LBL1	21 01	152	STO7	35 07	175	0	00
130	1	01	153	-	-45	176	÷	-24
131	RCL5	36 05	154	STO8	35 08	177	R/S	51

A.4 DETAILS OF THE POISSON PROBABILITY PROGRAM

Lines 1 to 29 establish the type of output requested. The calculation is the same in all cases, and all the probabilities are stored in the calculator. The calculation is run as a subroutine and the program returns to the starting area when the calculation is completed and the proper probability is recalled and printed.

Lines 30 to 50 clear the storage registers, store the input data, and check whether the expected value m is greater than 227, which is the maximum value that the calculator can handle. If m is less than 227, the program goes on to line 51 to start the calculation. If m is greater than 227, the following adjustments are made.

Value m is reduced to 227, and the remainder is retained in Register 8 for use further on in the program. As the summation increases toward the upper limit of the calculator, the balance of the exponent is retrieved and used to reduce the value of the summation by multiplying by e^{-q}, where q is either the remainder or another portion if the remainder is still greater than 227.

Line 36 is a check on whether with input probability a "greater than" probability is to be matched. If so, the program goes to line 101 to substitute $1 - P$ for P.

Lines 51 to 64 make the initial calculation of e^{-m} for the probability of zero events. This value is stored in Registers 1 and 5 for further use. A check is made at line 57 to determine if this was the probability called for. If so, the program goes to line 89 to terminate. Flag 1 is checked at line 59 to determine if the probability calculated is to be matched against an input probability. This is done at line 107.

Lines 65 to 88 are the main summation calculation. Line 66 increments the index in Register I, and the next integer probability is calculated by Eq. (3.27). Flag 1 is checked at line 75, and the index value is checked against the value of r at line 79. When the index reaches r, the calculation is complete. Lines 81 to 86 determine whether the summation is approaching the upper limit of the calculator. If so, another increment of the factored e^{-m} value is used at line 149.

Lines 89 to 100 are the termination routine. The calculated probabilities are stored in Registers 1 to 5 and the program returns to the starting section for the selection of the one to print.

Lines 101 to 106 are the routine to substitute $1 - P$ for P if a "greater than" probability is to be calculated.

Lines 107 to 115 check whether the input probability has been exceeded by the summation calculation. If not, the program returns to line 65 for the calculation of another term. If so, the program continues to line 116. Lines 108 to 111 check the input P value. If it is greater than unity, an error message is obtained.

Lines 116 to 148 are the termination routine for the calculation to match an input probability. The calculation is continued until the input probability is first exceeded. The output is the probability and r value for that term, and the probability and the $r - 1$ value for the preceding term which had been stored in Register 4. If a "greater than" probability has been asked for (line 116), the program goes to line 129 where unity minus the calculated probabilities are calculated and printed.

Lines 149 to 171 are the routine for returning the portion of the e^{-m} term that was removed if m was greater than 227. The portion not used was stored in Register 8, and it is returned at this point. The probabilities are accumulated in three storage registers, the probability of the current value of i in Register 1, the summation up to $i - 1$ in Register 4, and the summation up to i in Register 5. Each of these registers is multiplied by the value of e^{-s}, where s is the factor of the original m value removed back at lines 45 to 49.

Lines 172 to 176 are the error-message-generating routine by division by zero.

Table 3.10 shows the location of the input items and the different probabilities at the termination of the program.

Table A.5 t-1 Program for Hypotheses about Means (HP-97)

Hypothesis: $\bar{x} = m$	Hypothesis: $m_x = m_y$	Find α
x_i, KEY A	x_i, KEY A	t, ENTER ↑
m, KEY C	y_i, KEY B Run KEY C	ν, KEY D
Output: $\bar{x}$, $s(x)$, t, ν, α	Output: $\bar{x}$, $s(x)$, $\bar{y}$, $s(y)$, t, ν, α	Output: α

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	036	÷	-24	071	-	-45
002	F2?	16 23 02	037	P⇄S	16-51	072	STO3	35 03
003	GTO1	22 01	038	RCL9	36 09	073	×	-35
004	CLRG	16-53	039	√X	54	074	STO4	35 04
005	P⇄S	16-51	040	×	-35	075	S	16 54
006	CLRG	16-53	041	PRTX	-14	076	X²	53
007	*LBL1	21 01	042	RCL9	36 09	077	RCLE	36 15
008	SF2	16 21 02	043	1	01	078	×	-35
009	Σ+	56	044	-	-45	079	RCL4	36 04
010	RTN	24	045	PRTX	-14	080	+	-55
011	*LBLB	21 12	046	CF2	16 22 02	081	RCL3	36 03
012	F2?	16 23 02	047	GTOD	22 14	082	RCLE	36 15
013	GTO1	22 01	048	*LBL1	21 01	083	+	-55
014	GTO3	22 03	049	P⇄S	16-51	084	STO5	35 05
015	*LBL1	21 01	050	x̄	16 53	085	÷	-24
016	P⇄S	16-51	051	PRTX	-14	086	RCL9	36 09
017	*LBL3	21 03	052	S	16 54	087	1/X	52
018	Σ+	56	053	PRTX	-14	088	RCLD	36 14
019	RTN	24	054	SPC	16-11	089	1/X	52
020	*LBLC	21 13	055	P⇄S	16-51	090	+	-55
021	RCL9	36 09	056	x̄	16 53	091	×	-35
022	X≠0?	16-42	057	PRTX	-14	092	√X	54
023	GTO1	22 01	058	STOA	35 11	093	STOB	35 12
024	R↓	-31	059	RCL9	36 09	094	x̄	16 53
025	STOE	35 15	060	STOD	35 14	095	RCLA	36 11
026	x̄	16 53	061	1	01	096	-	-45
027	PRTX	-14	062	-	-45	097	RCLB	36 12
028	RCLE	36 15	063	STOE	35 15	098	÷	-24
029	-	-45	064	S	16 54	099	PRTX	-14
030	STOC	35 13	065	PRTX	-14	100	RCL5	36 05
031	S	16 54	066	SPC	16-11	101	PRTX	-14
032	PRTX	-14	067	X²	53	102	*LBLD	21 14
033	SPC	16-11	068	P⇄S	16-51	103	STO0	35 00
034	RCLC	36 13	069	RCL9	36 09	104	R↓	-31
035	X⇄Y	-41	070	1	01	105	STO1	35 01

Table A.5 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
106	R↑	16-31	134	RCLI	36 46	162	RCLA	36 11
107	√X	54	135	1	01	163	COS	42
108	÷	-24	136	X=Y?	16-33	164	×	-35
109	ABS	16 31	137	GTO3	22 03	165	RCLA	36 11
110	RAD	16-22	138	RCL8	36 08	166	+	-55
111	TAN⁻¹	16 43	139	RCLA	36 11	167	GSBa	23 16 11
112	STOA	35 11	140	COS	42	168	GTO6	22 06
113	SIN	41	141	X²	53	169	*LBL4	21 04
114	STOB	35 12	142	RCLI	36 46	170	RCLA	36 11
115	RCL0	36 00	143	÷	-24	171	GSBa	23 16 11
116	FRC	16 44	144	RCLI	36 46	172	GTO6	22 06
117	0	00	145	1	01	173	*LBL5	21 05
118	X=Y?	16-33	146	-	-45	174	RCLB	36 12
119	GTO1	22 01	147	×	-35	175	*LBL6	21 06
120	0	00	148	×	-35	176	CHS	-22
121	÷	-24	149	1	01	177	1	01
122	*LBL1	21 01	150	+	-55	178	+	-55
123	RCL0	36 00	151	STO8	35 08	179	PRTX	-14
124	2	02	152	DSZI	16 25 46	180	RTN	24
125	X>Y?	16-34	153	DSZI	16 25 46	181	*LBLa	21 16 11
126	GTO4	22 04	154	GTO2	22 02	182	2	02
127	X=Y?	16-33	155	RCLB	36 12	183	Pi	16-24
128	GTO5	22 05	156	×	-35	184	÷	-24
129	-	-45	157	GTO6	22 06	185	×	-35
130	STOI	35 46	158	*LBL3	21 03	186	RTN	24
131	1	01	159	RCL8	36 08	187	R/S	51
132	STO8	35 08	160	RCLB	36 12			
133	*LBL2	21 02	161	×	-35			

A.5 DETAILS OF THE *t*-1 PROGRAM

Lines 1 to 6 clear all registers before the first data are entered.

Lines 7 to 10 make use of the built-in calculator routine of KEY $\Sigma+$ to obtain Σx, Σx^2, and a count of the data input.

Lines 11 to 19 are used if a second set of data is entered for comparing two means.

Lines 20 to 23 establish which type of hypothesis is being tested. If Register 9, recalled at line 21, is not zero, it indicates two sets of data were entered, and the program goes to line 48. If it is zero, the program continues to

Lines 24 to 47 The program uses the built-in routine to obtain $\bar{x}$ and $s(x)$, and calculates t from Eq. (4.4). The degrees of freedom are calculated as $n - 1$, and the program goes on to line 102 to calculate the probability.

Lines 48 to 65 obtain the means and the standard deviation estimates for the two sets of data from the built-in functions of $\bar{x}$ and s.

Lines 66 to 85 calculate the expression under the square root sign in Eq. (4.7).

Lines 86 to 92 calculate the denominator of Eq. (4.6).

Lines 93 to 99 calculate and print the t value from Eq. (4.6).

Lines 100 to 101 print the degrees of freedom and take the program to line 102 to start the probability calculation.

Lines 102 to 114 start the probability calculation following the rearranged Eq. (4.17) or (4.18). Lines 102 to 114 calculate θ and $\sin \theta$, and store the value.

Lines 115 to 132 check whether the number of degrees of freedom is one or two. If ν is 1, the probability is $1 - 2\theta/\pi$; if ν is 2, the probability is $\sin \theta$. If ν is not an integer, an error message is generated by dividing by zero.

Lines 133 to 154 The expansion sections of the rearrangement of Eqs. (4.17) and (4.18) are the same for even and odd values of the degrees of freedom. The multiplying factors are different. Lines 133 to 154 carry out the calculation of the section inside the multiple parentheses of the rearranged equations. If ν is odd, when the calculation is completed the program goes to line 158. If the value of ν is even, the program goes on to line 155. Whether ν is even or odd is determined by decreasing ν in increments of 2 for each term of the calculation and determining whether the final term is 1 or 0.

Lines 155 to 157 multiply the calculation by $\sin \theta$ if ν is even and then go on to the end routine at line 175.

Lines 158 to 168 multiply the calculation by $\sin \theta \cos \theta$, add θ, and multiply the sum by $2/\pi$ at line 181; then the program goes on to the end routine.

Lines 169 to 174 supply the values for the result when ν equals 1 or 2.

Lines 175 to 180 are the end routine which subtract the calculated probability from 1.0 to get the α value used in this text.

Lines 181 to 186 calculate $2/\pi$.

The program section starting with line 102 is the routine for calculating the probability associated with t and ν. This section of the program can be lifted entirely from the t-1 program and used elsewhere if the α calculation is wanted. Text references to the t-3 program refer to this portion of the program.

Table A.6 ***t*-2 Program for Calculating Confidence Intervals and Sample Size (HP-97)**

Confidence intervals	Sample size	Find t
x_i, KEY a P (confidence = $1-\alpha$), KEY b Output: $\bar{x}$, x_l, x_u	$s(x)$, ENTER ↑ α, ENTER ↑ β, ENTER ↑ Δ ($= m - m_1$), KEY c Output: n	α, ENTER ν, KEY d Output: t

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLa	21 16 11	036	×	-35	071	1	01
002	F2?	16 23 02	037	RCLB	36 12	072	-	-45
003	GTO1	22 01	038	X⇄Y	-41	073	X<0?	16-45
004	CLRG	16-53	039	-	-45	074	1	01
005	P⇄S	16-51	040	PRTX	-14	075	STOI	35 46
006	CLRG	16-53	041	LSTX	16-63	076	*LBL8	21 08
007	*LBL1	21 01	042	RCLB	36 12	077	ISZI	16 26 46
008	SF2	16 21 02	043	+	-55	078	RCL1	36 01
009	Σ+	56	044	PRTX	-14	079	GSB9	23 09
010	RTN	24	045	CF2	16 22 02	080	STO3	35 03
011	*LBLb	21 16 12	046	RTN	24	081	RCL0	36 00
012	STOD	35 14	047	*LBLc	21 16 13	082	GSB9	23 09
013	1	01	048	STOA	35 11	083	RCL3	36 03
014	-	-45	049	R↓	-31	084	+	-55
015	CHS	-22	050	2	02	085	RCLE	36 15
016	P⇄S	16-51	051	×	-35	086	×	-35
017	RCL9	36 09	052	STOB	35 12	087	X²	53
018	1	01	053	R↓	-31	088	RCLI	36 46
019	-	-45	054	STOC	35 13	089	1	01
020	STOI	35 46	055	R↓	-31	090	+	-55
021	X⇄Y	-41	056	STOD	35 14	091	X≤Y?	16-35
022	GSB7	23 07	057	RCLA	36 11	092	GTO8	22 08
023	GSB9	23 09	058	÷	-24	093	PRTX	-14
024	STOA	35 11	059	STOE	35 15	094	RTN	24
025	P⇄S	16-51	060	SF2	16 21 02	095	*LBLd	21 16 14
026	x̄	16 53	061	RCLC	36 13	096	STOI	35 46
027	PRTX	-14	062	GSB7	23 07	097	R↓	-31
028	STOB	35 12	063	RCLB	36 12	098	STOA	35 11
029	S	16 54	064	GSB7	23 07	099	GSB7	23 07
030	P⇄S	16-51	065	RCL1	36 01	100	GSB9	23 09
031	RCL9	36 09	066	+	-55	101	PRTX	-14
032	√X	54	067	RCLE	36 15	102	RTN	24
033	÷	-24	068	×	-35	103	*LBL7	21 07
034	STOC	35 13	069	X²	53	104	2	02
035	RCLA	36 11	070	INT	16 34	105	÷	-24

Table A.6 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
106	X²	53	144	1	01	182	×	-35
107	1/X	52	145	+	-55	183	1	01
108	LN	32	146	÷	-24	184	6	06
109	√X	54	147	CHS	-22	185	+	-55
110	STO0	35 00	148	RCL0	36 00	186	RCL2	36 02
111	.	-62	149	+	-55	187	×	-35
112	2	02	150	F2?	16 23 02	188	3	03
113	7	07	151	STO1	35 01	189	+	-55
114	0	00	152	STO0	35 00	190	RCL8	36 08
115	6	06	153	RTN	24	191	×	-35
116	1	01	154	*LBL9	21 09	192	4	04
117	×	-35	155	STO8	35 08	193	×	-35
118	2	02	156	X²	53	194	RCLI	36 46
119	.	-62	157	STO2	35 02	195	×	-35
120	3	03	158	RCLI	36 46	196	+	-55
121	0	00	159	3	03	197	RCL2	36 02
122	7	07	160	Yˣ	31	198	3	03
123	5	05	161	3	03	199	×	-35
124	3	03	162	8	08	200	1	01
125	+	-55	163	4	04	201	9	09
126	.	-62	164	×	-35	202	+	-55
127	9	09	165	STO7	35 07	203	RCL2	36 02
128	9	09	166	RCL8	36 08	204	×	-35
129	2	02	167	×	-35	205	1	01
130	2	02	168	RCL8	36 08	206	7	07
131	9	09	169	3	03	207	+	-55
132	RCL0	36 00	170	Yˣ	31	208	RCL2	36 02
133	×	-35	171	RCL8	36 08	209	×	-35
134	.	-62	172	+	-55	210	1	01
135	0	00	173	9	09	211	5	05
136	4	04	174	6	06	212	-	-45
137	4	04	175	×	-35	213	RCL8	36 08
138	8	08	176	RCLI	36 46	214	×	-35
139	1	01	177	X²	53	215	+	-55
140	RCL0	36 00	178	×	-35	216	RCL7	36 07
141	X²	53	179	+	-55	217	÷	-24
142	×	-35	180	RCL2	36 02	218	RTN	24
143	+	-55	181	5	05	219	R/S	51

A.6 DETAILS OF THE *t*-2 PROGRAM

Lines 1 to 10 clear the calculator registers with the first data entry and use the built-in summation routine of KEY $\Sigma+$ to calculate Σx, Σx^2, and the number of data entries.

Lines 11 to 21 take the input probability entry, which is for a confidence interval, and convert it to an α value for the t calculation. They also calculate the degrees of freedom from the amount of data entered, $n - 1$.

Lines 22 and 23 direct the program to the t calculation routine that starts at line 103.

Lines 24 to 34 After the t calculation, the program returns to line 24. The built-in routines for calculating $\bar{x}$ and $s(x)$ are used, and $s(\bar{x})$ is calculated by dividing $s(x)$ by the square root of n.

Lines 35 to 46 calculate the lower limit of the confidence range and the upper limit from Eqs. (4.10) and (4.11), print the results, and terminate the program.

Lines 47 to 61 are the start of the sample size calculation. The β input probability is a one-sided probability, and the t calculation based on a probability input is for a two-sided probability. The β value is therefore multiplied by 2 at line 51. Quantity $s(x)/(m_1 - m)$, of Eq. (4.16), is used in the calculation as a single value and is stored in Register E at line 59. Flag 2 is set for a purpose that will be discussed at line 103.

Line 62 directs the program to the start of the t calculation, but only the normal deviate value is calculated if Flag 2 is set.

Lines 63 to 75 recover the value of 2β for a second calculation of the normal deviate with this probability value. The sample size is calculated from Eq. (4.16) using the two normal deviate values instead of the t values. The value obtained from this calculation is the minimum sample that would be required if $s(x)$ was not an estimate but was the true standard deviation of the population. The actual sample size calculation is by trial and error. Starting the trial-and-error calculation with the minimum sample size minimizes the number of trials.

Lines 76 to 92 are the trial-and-error routine; t_α and t_β are calculated by the routine starting at line 154, and the equality:

$$n = \left((t_\alpha + t_\beta) \frac{s(x)}{m_1 - m} \right)^2 \tag{A.1}$$

is tested starting with the value obtained using the normal deviates. n is incremented by one for each successive trial, and when $n + 1$ exceeds the right-hand side of Eq. (A.1) the calculation is terminated. $n + 1$ is used, since the t calculations are based on degrees of freedom, and the number of degrees of freedom equals $n - 1$.

Lines 93 and 94 print the value of n and terminate the calculation.

Lines 95 to 101 are the input routine and end routine for calculating t from input of α and degrees of freedom. The routine simply directs the program to the same routines used for the confidence interval of sample size calculations.

Lines 103 to 153 calculate the normal deviate from Eq. (4.21); u, of Eq. (4.21), is evaluated in lines 104 to 110, and the main calculation is carried out in lines 111 to 149. Since the calculation is used twice when the sample size calculation is made, Flag 2 is used with lines 150 to 153 to store one value in Register 1 and the other in Register 0.

Lines 154 to 218 calculate the t value from the normal deviate and the degrees of freedom following Eq. (4.24). If the g term was to be added to obtain greater accuracy, it would be added after line 217.

Table A.7 Programs χ^2-1, χ^2-2, and χ^2-3 (HP-97)

χ^2-1	χ^2-2	χ^2-3
$a/b = p/q$?	$x_i = \bar{x}$?	$x_i/f_i = \Sigma x/\Sigma f$?
a, ENTER ↑ b, ENTER ↑ p, ENTER ↑ q, KEY A	x_i, KEY B KEY C	f_i, ENTER ↑ x_i, KEY D KEY E
Output:	Output:	Output:
Degrees of freedom χ^2	$\bar{x}$ Degrees of freedom χ^2	χ^2 for each $f_i x_i$ Degrees of freedom χ^2 total

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	001	*LBLB	21 12	001	*LBLD	21 14
002	÷	-24	002	F0?	16 23 00	002	F0?	16 23 00
003	STOC	35 13	003	GTO1	22 01	003	GTO1	22 01
004	R↓	-31	004	CLRG	16-53	004	CLRG	16-53
005	STOB	35 12	005	P⇄S	16-51	005	P⇄S	16-51
006	R↓	-31	006	SF0	16 21 00	006	CLRG	16-53
007	STOA	35 11	007	*LBL1	21 01	007	*LBL1	21 01
008	RCLC	36 13	008	Σ+	56	008	SF0	16 21 00
009	RCLB	36 12	009	RTN	24	009	STOi	35 45
010	×	-35	010	*LBLC	21 13	010	RCLE	36 15
011	-	-45	011	$\bar{x}$	16 53	011	+	-55
012	ABS	16 31	012	PRTX	-14	012	STOE	35 15
013	RCLC	36 13	013	P⇄S	16-51	013	ISZI	16 26 46
014	1	01	014	RCL9	36 09	014	R↓	-31
015	PRTX	-14	015	1	01	015	STOi	35 45
016	+	-55	016	-	-45	016	RCLD	36 14
017	2	02	017	PRTX	-14	017	+	-55
018	÷	-24	018	RCL5	36 05	018	STOD	35 14
019	-	-45	019	RCL9	36 09	019	ISZI	16 26 46
020	X²	53	020	×	-35	020	RTN	24
021	RCLA	36 11	021	RCL4	36 04	021	*LBLE	21 15
022	RCLB	36 12	022	÷	-24	022	RCLE	36 15
023	+	-55	023	RCL4	36 04	023	RCLD	36 14
024	RCLC	36 13	024	-	-45	024	÷	-24
025	×	-35	025	PRTX	-14	025	STOE	35 15
026	÷	-24	026	CF0	16 22 00	026	RCLI	36 46
027	PRTX	-14	027	RTN	24	027	STOD	35 14
028	1	01				028	0	00
029	X⇄Y	-41				029	STOI	35 46
030	RTN	24				030	*LBL2	21 02

Table A.7 Program χ^2-3 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
031	RCLi	36 45	042	RCLi	36 45	053	RCLI	36 46
032	0	00	043	÷	-24	054	2	02
033	STOi	35 45	044	PRTX	-14	055	÷	-24
034	X⇄Y	-41	045	ST+0	35-55 00	056	1	01
035	ISZI	16 26 46	046	ISZI	16 26 46	057	-	-45
036	RCLi	36 45	047	RCLD	36 14	058	PRTX	-14
037	RCLE	36 15	048	RCLI	36 46	059	RCL0	36 00
038	×	-35	049	X≠Y?	16-32	060	PRTX	-14
039	STOi	35 45	050	GTO2	22 02	061	SPC	16-11
040	-	-45	051	SPC	16-11	062	RTN	24
041	X²	53	052	CF0	16 22 00	063	R/S	51

A.7 DETAILS OF χ^2-1 PROGRAM: RATIO TEST

Lines 1 to 7 obtain the ratio p/q and store it and the values of a and b.

Lines 8 to 12 calculate the value of $|a - (p/q)b|$ of Eq. (5.5).

Lines 13 to 20 print the value 1 as the degrees of freedom and complete the calculation of the numerator of Eq. (5.5).

Lines 21 to 27 complete the calculation of Eq. (5.5) and print the resulting χ^2 value.

Lines 28 to 30 place the degrees of freedom and χ^2 in the proper sequence to be used in the χ^2-5 program, if the probability is desired, and end the program.

If χ^2-5 is merged with χ^2-1, it starts at line 30, replacing the "end" statement. With the two programs merged, the calculation would continue until the probability value was obtained and printed.

If the programs were not merged and the probability value was desired, χ^2-5 could be entered into the calculator and run with KEY e without any data entry inasmuch as the data were arranged in the proper order at the end of χ^2-1.

A.8 DETAILS OF χ^2-2 PROGRAM: EQUAL EXPECTATION

Lines 1 to 6 This program compares a number of counted observations with their mean. All the data are entered in the same manner, with KEY B. Lines 1 to 6 establish whether the number being entered is the first by checking Flag 0. If Flag 0 is not set, indicating that the value is the first, the registers are cleared before the value is entered. If Flag 0 is set, the program goes to line 7.

Lines 7 to 9 process the data using KEY Σ+ which calculates Σx^2, Σx, and counts the data entries.

Line 10 starts the calculation of χ^2. The last term of Eq. (5.9) is used, and all the terms are obtained by built-in routines in the calculator.

Lines 11 to 12 obtain the value of $\bar{x}$ and print it.

Lines 13 to 17 obtain the value of the degrees of freedom, equal to $n - 1$, and print it.

Lines 18 to 25 calculate χ^2 by Eq. (5.9) and print the result.

Lines 26 and 27 clear Flag 0 for the next calculation and end the program. The degrees of freedom and the χ^2 are in the proper sequence to use the probability program, χ^2-5, so no adjustment is necessary.

If program χ^2-5 was merged with the χ^2-2 program, it would start at line 27, replacing the "end" statement. If program χ^2-5 was not merged with program χ^2-2 and the probability was desired, it could be entered into the calculator and run when the χ^2-2 program results were obtained without any additional data entry.

A.9 DETAILS OF χ^2-3 PROGRAM: UNEQUAL EXPECTATION

Lines 1 to 6 establish whether the data entry are the first. If they are the first, the registers are cleared before the data are processed. If they are not the first, the program goes to line 7.

Lines 7 to 20 process the input data. A maximum of 11 pairs of weighting factors and observations can be handled. These require 22 storage areas. Three additional storage areas are used for intermediate calculations, making a total of 25. If more than 11 pairs of data are to be entered, more storage area would have to be made available.

The storage areas are used sequentially in pairs, starting with Register 0 for the first number of observations, and Register 1 for the first weighting factor. Line 9 stores the number of observations, line 13 increments the storage area number, and line 15 stores the weighting factor. The number of observations is summed in Register E, which is also Register 24; and Register I, which is Register 25, is used to index the register numbers to locate the individual pairs of data input. The weighting factors are summed in Register D, which is Register 23 at lines 16 to 18.

Lines 19 and 20 increment the index register after the weighting factor is stored and end the data input routine.

Line 21 starts the χ^2 calculation which follows Eq. (5.7).

Lines 22 to 29 divide the total observations by the total of the weighting factors and store this value in Register E. The total data input from Register I is stored in Register D for later use, and Register I is reset to zero at line 29.

Lines 30 to 50 calculate the χ^2 for each pair of data and accumulate the sum of the χ^2 values in Register 0. Each value is printed as it is calculated (line 44).

The number of pairs of data is checked against the total input at line 49. If all the data have not been processed, the program returns to line 30 to calculate the next pair. If all have been processed, the program goes on to line 51.

Lines 51 to 62 calculate the number of degrees of freedom—the total data input divided by two, less one—print this result, print the total χ^2 accumulated in Register 0, and end the program.

The program ends with the degrees of freedom and χ^2 in the proper order to go directly to program χ^2-5 for a probability calculation. Program χ^2-5 can either be merged with program χ^2-3, starting at line 62 for the direct calculation of the probability, or it can be put into the calculator when program χ^2-3 is completed, and the probability may then be calculated with KEY e.

Table A.8 χ^2**-4 Program:** $r \times c$ **Contingency Tables (HP-97)**

Input: Number of rows, KEY a, data by columns, KEY a; KEY b

Output:
χ^2 for each row
Total degrees of freedom
χ^2 total

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLa	21 16 11	036	+	-55	071	X⇄Y	-41
002	F0?	16 23 00	037	STOI	35 46	072	RCLi	36 45
003	GTO1	22 01	038	R↓	-31	073	÷	-24
004	CLRG	16-53	039	RCL0	36 00	074	ST+0	35-55 00
005	P⇄S	16-51	040	STOD	35 14	075	GSB3	23 03
006	CLRG	16-53	041	÷	-24	076	R↓	-31
007	STOA	35 11	042	ST+i	35-55 45	077	STOi	35 45
008	STOI	35 46	043	RCLI	36 46	078	DSZI	16 25 46
009	SF0	16 21 00	044	RCLA	36 11	079	GTO1	22 01
010	RTN	24	045	2	02	080	RCLA	36 11
011	*LBL1	21 01	046	x	-35	081	STOI	35 46
012	STOi	35 45	047	-	-45	082	*LBLc	21 16 13
013	ST+0	35-55 00	048	STOI	35 46	083	RCLi	36 45
014	RCLI	36 46	049	DSZI	16 25 46	084	RCLB	36 12
015	RCLA	36 11	050	GTO6	22 06	085	x	-35
016	+	-55	051	0	00	086	RCLI	36 46
017	STOI	35 46	052	STO0	35 00	087	RCLA	36 11
018	R↓	-31	053	RCLA	36 11	088	+	-55
019	ST+i	35-55 45	054	STOI	35 46	089	STOI	35 46
020	GSB3	23 03	055	RTN	24	090	R↓	-31
021	DSZI	16 25 46	056	*LBLb	21 16 12	091	RCLi	36 45
022	RTN	24	057	RCL1	36 01	092	-	-45
023	1	01	058	STOE	35 15	093	PRTX	-14
024	RCLC	36 13	059	RCLA	36 11	094	GSB3	23 03
025	+	-55	060	3	03	095	DSZI	16 25 46
026	STOC	35 13	061	x	-35	096	GTOc	22 16 13
027	RCLA	36 11	062	*LBL8	21 08	097	SPC	16-11
028	STOI	35 46	063	STOI	35 46	098	RCLA	36 11
029	*LBL6	21 06	064	RCLi	36 45	099	1	01
030	RCLi	36 45	065	GSB3	23 03	100	-	-45
031	X²	53	066	R↓	-31	101	RCLC	36 13
032	RCLA	36 11	067	RCLi	36 45	102	1	01
033	2	02	068	RCLB	36 12	103	-	-45
034	x	-35	069	+	-55	104	x	-35
035	RCLI	36 46	070	STOB	35 12	105	PRTX	-14

Table A.8 *(Cont.)*

Step	Key	Code	Step	Key	Code	Step	Key	Code
106	STOI	35 46	122	5	05	138	X⇄Y	-41
107	1	01	123	-	-45	139	RTN	24
108	X≠Y?	16-32	124	RCLE	36 15	140	*LBL1	21 01
109	GTO7	22 07	125	÷	-24	141	RCLI	36 46
110	RCLE	36 15	126	X^2	53	142	RCLA	36 11
111	RCLD	36 14	127	*LBL7	21 07	143	2	02
112	RCL3	36 03	128	RCL0	36 00	144	×	-35
113	×	-35	129	1	01	145	+	-55
114	RCLB	36 12	130	-	-45	146	GTO8	22 08
115	÷	-24	131	RCLB	36 12	147	*LBL3	21 03
116	-	-45	132	×	-35	148	RCLI	36 46
117	ABS	16 31	133	×	-35	149	RCLA	36 11
118	X=0?	16-43	134	*LBL9	21 09	150	-	-45
119	GTO9	22 09	135	PRTX	-14	151	STOI	35 46
120	STOE	35 15	136	CF0	16 22 00	152	RTN	24
121		-62	137	RCLI	36 46	153	R/S	51

A.10 DETAILS OF χ^2-4 PROGRAM: CONTINGENCY TABLES

Program χ^2-4 is listed in Table A-8. Contingency table data are in rows and columns, and the designation is arbitrary. Program χ^2-4 can process up to six rows of data with no limitation on the number of columns.

Lines 1 to 10 All the data are entered with KEY a. The first entry is the number of rows. The data for each column are processed as a unit. Having the number of rows as an index, the program "knows" when the data from one column have been entered. With the first data input all registers are cleared. Flag 0, at line 2, indicates whether the input is the first.

Lines 11 to 22 After the entry of the number of rows, the balance of the data is entered by columns. The program uses three storage areas for each row of data: one for the current entry, one for the row total, and one for $\Sigma x_{ij}^2/C_j$ for each ith row. C_j is the total of the jth column. The current entry is stored at line 12 and the column total is stored in Register 0 at line 13. The register index is changed at line 17, and the row total is accumulated at line 19. The program uses a routine at line 147 to shift the index back for the next calculation. The number of entries is checked at line 21. If they have not equaled the number of rows, stored in Register A, the calculation stops for the next entry. If a total column has been entered, the program goes on to line 23.

Lines 23 to 28 count the number of columns, store the value in Register C, and reset the index with the number of rows.

Lines 29 to 50 square each column entry and divide by the column total, stored in Register 0. The column total is stored in Register D (line 40) to be available if the data are from a 2×2 table with one degree of freedom and the χ^2 calculation needs to be corrected for the one degree of freedom. The $\Sigma x_{ij}^2/C_j$ for each column entry is stored in a different register for each row for later division by the row total. The registers are changed at lines 32 to 37 and changed back at lines 43 to 48. The calculation is repeated for each column entry and a check is made at line 49 with a return to line 29 from line 50 if the column is not completed, and a continuation to line 51 if it is.

Lines 51 to 55 reset Registers 0, A, and I for the start of the next column of data.

Line 56 (KEY b) starts the χ^2 calculation when all the data have been entered.

Lines 57 and 58 save the last data entry in Register E to be available if the correction for one degree of freedom is required.

Lines 59 to 63 set the index register for the start of the χ^2 calculation.

Lines 64 to 74 recall the value of $\Sigma x_{ij}^2/C_j$ and divide by R_i. The quotient is summed in Register 0; ΣR_i is accumulated in Register B to give the total of all the data.

Lines 75 to 79 establish whether all of the rows have been calculated. If not, the program goes to line 140 to reset the index register and then return to line 62 to repeat the calculation for the next row. If all the rows have been included in the calculation, the program goes on to line 80.

Lines 80 to 97 calculate the χ^2 value for each row by calculating $T\sum_{j=i}^{c}(x_{ij}^2/R_iC_j) - T$, which equals χ^2 for the ith row. T is the sum of all the data. R_i is the sum of the ith row, C_j is the sum of the jth column, and the summation is carried out for all columns from 1 to c, where there are c columns.

The result is printed at line 93. The program goes to line 147 to reset the index register, and a check is made at line 95 to determine if all the rows have been calculated. If not, the program returns to line 82 to repeat the calculation for the next row. If all the rows have been calculated, the program goes on to line 98.

Lines 98 to 106 calculate the degrees of freedom: $(r\text{-}1)(c\text{-}1)$. The result is printed and stored in Register I.

Lines 107 to 109 establish whether the degrees of freedom are more than 1. If so, the program goes on to line 127. If not, the program continues with line 110.

Lines 110 to 126 make the correction to χ^2 for a calculation of one degree of freedom. In a 2×2 contingency table, all the differences between observed and expected values are equal. χ^2 without correction is the square of this difference, Δ^2 multiplied by the sum of the reciprocals of the expected values. χ^2 corrected is equal to χ^2 uncorrected multiplied by $(\Delta - 0.5)^2/\Delta^2$. This is the calculation made in lines 110 to 126. The difference Δ is calculated from the last data entry saved in Register E. Lines 118 and 119 are for the rare case when the 2×2 table has a zero χ^2 and prevent the calculator from having to divide by zero.

Lines 127 to 132 calculate χ^2 (uncorrected) by using Eq. (5.8) and the summation of x_{ij}^2/R_iC_j stored in Register 0 at lines 62 to 74.

Line 133 multiplies the uncorrected χ^2 either by the correction term calculated at lines 110 to 126, or by unity if the number of degrees of freedom was greater than unity.

Lines 134 to 139 print the answer, clear Flag 0 for the next calculation, place the degrees of freedom and the χ^2 value in the proper order for use by program χ^2-5 for the probability calculation, and end.

Lines 140 to 146 are a routine to reset the index register when the row totals are being calculated.

Lines 147 to 152 are a routine to shift the index register. The same operation is used four times in the program, and some program space is saved by using a subroutine.

Table A.9 χ^2-5 Program: χ^2 Probability (HP-97)

Input: Degrees of freedom, ENTER ↑
χ^2, KEY e

Output: Probability of $\geqq \chi^2$

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLe	21 16 15	035	2	02	069	*LBL3	21 03
002	STO2	35 02	036	÷	-24	070	Pi	16-24
003	R↓	-31	037	Y^x	31	071	√X	54
004	STO1	35 01	038	STOE	35 15	072	ST÷7	35-24 07
005	STO3	35 03	039	RCLD	36 14	073	*LBL1	21 01
006	2	02	040	CHS	-22	074	2	02
007	×	-35	041	e^x	33	075	ST+3	35-55 03
008	1	01	042	×	-35	076	*LBL4	21 04
009	-	-45	043	STO7	35 07	077	RCL2	36 02
010	√X	54	044	RCL3	36 03	078	RCL3	36 03
011	5	05	045	2	02	079	÷	-24
012	+	-55	046	+	-55	080	ST×5	35-35 05
013	RCL2	36 02	047	2	02	081	2	02
014	2	02	048	÷	-24	082	ST+3	35-55 03
015	×	-35	049	STO4	35 04	083	RCL5	36 05
016	√X	54	050	INT	16 34	084	ST+6	35-55 06
017	X≤Y?	16-35	051	LSTX	16-63	085	1	01
018	GTO1	22 01	052	X≠Y?	16-32	086	EEX	-23
019	1	01	053	GTO0	22 00	087	CHS	-22
020	EEX	-23	054	1	01	088	4	04
021	CHS	-22	055	-	-45	089	X≤Y?	16-35
022	5	05	056	N!	16 52	090	GTO4	22 04
023	PRTX	-14	057	ST÷7	35-24 07	091	RCL6	36 06
024	RTN	24	058	GTO1	22 01	092	1	01
025	*LBL1	21 01	059	*LBL0	21 00	093	+	-55
026	0	00	060	.	-62	094	RCL7	36 07
027	STO6	35 06	061	5	05	095	×	-35
028	1	01	062	X=Y?	16-33	096	1	01
029	STO5	35 05	063	GTO3	22 03	097	X⇄Y	-41
030	RCL2	36 02	064	X⇄Y	-41	098	-	-45
031	2	02	065	1	01	099	PRTX	-14
032	÷	-24	066	-	-45	100	RTN	24
033	STOD	35 14	067	ST÷7	35-24 07	101	R/S	51
034	RCL3	36 03	068	GTO0	22 00			

A.11 DETAILS OF χ^2-5 PROGRAM: χ^2 PROBABILITY

The χ^2-5 program can be run independently with input of degrees of freedom and the χ^2 value. The output is the probability of an equal or greater χ^2. The program can also be merged with any of the other χ^2 programs, and after the χ^2 value has been calculated by that program, the probability value will also be calculated. The labels for the different routines and the storage areas used have been selected so there will be no interference with any of the programs with which χ^2-5 might be merged.

The calculation of the χ^2 probability follows Eq. (5.10).

Lines 1 to 4 store the values of χ^2 and the degrees of freedom in Registers 2 and 1.

Lines 5 to 24 test to determine whether the input value of χ^2 exceeds its mean from a normal approximation by more than five standard deviations. If not, the program goes to line 25. If so, the program prints a probability value of 1×10^{-5} and ends.

Lines 25 to 29 start the probability calculation by storing values of 0 and 1 in Registers 6 and 5.

Lines 30 to 43 calculate $(\chi^2/2)^{\nu/2} e^{-\chi^2/2}$ and store the value in Register 7.

Lines 44 to 53 establish whether $(\nu + 2)/2$ is an integer. If not, the program goes to line 59 to calculate the gamma value. If so, the program goes on to line 54.

Lines 54 to 58 calculate the gamma value of $(\nu + 2)/2$, when ν is even, by the factorial function, and divide the value in Register 7 by the result. This is the first variable term of Eq. (5.10). (*Note:* If the number of degrees of freedom is even and greater than 136, it will exceed the capacity of the calculator. It would be possible to put a control routine in the program to handle large numbers of degrees of freedom, but it is so seldom that χ^2 is calculated for degrees of freedom exceeding even 100 that the precaution, which would enlarge the program and slow down the calculation, is not deemed worthwhile. If the χ^2-5 program will not handle the data in a special case, it is suggested that the normal approximation be used.)

Lines 59 to 72 calculate the gamma function for odd numbers of degrees of freedom, using the relations shown below Eq. (5.10).

Lines 73 to 75 calculate the value of $(\nu + 2)$ for the start of the summation calculation of Eq. (5.10).

Lines 76 to 90 carry out the summation calculation. The individual terms of the summation are calculated in Register 5, and the summation is accumulated in Register 6. When the last term calculated is less than 0.0001, the calculation is terminated. If greater accuracy is desired, a smaller value can be used at lines 85 to 88.

Lines 91 to 100 complete the calculation of Eq. (5.10) by adding 1 to the summation term and multiplying the sum by the value stored in Register 7, and then subtracting the product from 1. The result is printed at line 99, and the program ends.

Table A.10 Variance-1 Program: Compare Variance with a Standard (HP-97)

Input: x_i, KEY A; σ^2, KEY B
Output: $s^2(x)$, degrees of freedom, F, probability

Input: Degrees of freedom, ENTER ↑, $s^2(x)$, ENTER ↑, σ^2, KEY C
Output: F, probability

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	037	RCLC	36 13	073	2	02
002	F0?	16 23 00	038	÷	-24	074	÷	-24
003	GTO1	22 01	039	PRTX	-14	075	STOD	35 14
004	CLRG	16-53	040	STOC	35 13	076	RCL3	36 03
005	P⇄S	16-51	041	RCLB	36 12	077	2	02
006	CLRG	16-53	042	×	-35	078	÷	-24
007	SF0	16 21 00	043	CF0	16 22 00	079	Y^x	31
008	*LBL1	21 01	044	*LBL1	21 01	080	STOE	35 15
009	Σ+	56	045	STO2	35 02	081	RCLD	36 14
010	RTN	24	046	R↓	-31	082	CHS	-22
011	*LBLC	21 13	047	STO3	35 03	083	e^x	33
012	STOC	35 13	048	2	02	084	×	-35
013	R↓	-31	049	×	-35	085	STO7	35 07
014	STOA	35 11	050	1	01	086	RCL3	36 03
015	R↓	-31	051	-	-45	087	2	02
016	STOB	35 12	052	√X	54	088	+	-55
017	RCLA	36 11	053	5	05	089	2	02
018	RCLC	36 13	054	+	-55	090	÷	-24
019	÷	-24	055	RCL2	36 02	091	INT	16 34
020	PRTX	-14	056	2	02	092	LSTX	16-63
021	RCLB	36 12	057	×	-35	093	X≠Y?	16-32
022	×	-35	058	√X	54	094	GTO2	22 02
023	GTO1	22 01	059	X≤Y?	16-35	095	1	01
024	*LBLB	21 12	060	GTO1	22 01	096	-	-45
025	STOC	35 13	061	1	01	097	N!	16 52
026	S	16 54	062	EEX	-23	098	ST÷7	35-24 07
027	X^2	53	063	CHS	-22	099	GTO1	22 01
028	PRTX	-14	064	5	05	100	*LBL2	21 02
029	STOA	35 11	065	PRTX	-14	101	.	-62
030	P⇄S	16-51	066	RTN	24	102	5	05
031	RCL9	36 09	067	*LBL1	21 01	103	X=Y?	16-33
032	1	01	068	0	00	104	GTO3	22 03
033	-	-45	069	STO6	35 06	105	X⇄Y	-41
034	PRTX	-14	070	1	01	106	1	01
035	STOB	35 12	071	STO8	35 08	107	-	-45
036	RCLA	36 11	072	RCL2	36 02	108	ST÷7	35-24 07

Table A.10 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
109	GTO2	22 02	121	ST×8	35-35 08	133	1	01
110	*LBL3	21 03	122	2	02	134	+	-55
111	Pi	16-24	123	ST+3	35-55 03	135	RCL7	36 07
112	√X	54	124	RCL8	36 08	136	×	-35
113	ST÷7	35-24 07	125	ST+6	35-55 06	137	1	01
114	*LBL1	21 01	126	1	01	138	X⇄Y	-41
115	2	02	127	EEX	-23	139	-	-45
116	ST+3	35-55 03	128	CHS	-22	140	PRTX	-14
117	*LBL4	21 04	129	4	04	141	STOD	35 14
118	RCL2	36 02	130	X≤Y?	16-35	142	P⇄S	16-51
119	RCL3	36 03	131	GTO4	22 04	143	RTN	24
120	÷	-24	132	RCL6	36 06	144	R/S	51

A.12 DETAILS OF VARIANCE-1 PROGRAM: COMPARE VARIANCE WITH A STANDARD

Lines 1 to 10 take the input of data and calculate Σx, Σx^2, and n with KEY $\Sigma+$. The first data entry clears all registers.

Lines 11 to 23 are for input of a variance estimate already calculated for which a comparison test is desired. The F ratio $s^2(x)/\sigma^2$ is calculated and multiplied by the number of degrees of freedom, since the probability value will be calculated for χ^2.

Lines 24 to 44 calculate $s^2(x)$ using the built-in routine of KEY s, calculate the degrees of freedom, and the F ratio. The program prints these three results and multiplies F by the number of degrees of freedom for the probability calculation.

Lines 45 to 144 are the same as χ^2-5 for the probability calculation of χ^2; χ^2 at ν degrees of freedom has the same probability as F at ν, and ∞ degrees of freedom when χ^2 is equal to $F \cdot \nu$.

Table A.11 Variance-2 Program: Comparison of Two Variances (HP-97)

Input: x_i, KEY A; y_i, KEY B; KEY C

Output: $s^2(x)$, ν_x, $s^2(y)$, ν_y, F value (probability can be obtained with Variance-9)

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	022	X²	53	043	X>Y?	16-34
002	F0?	16 23 00	023	STOE	35 15	044	GTO1	22 01
003	GTO1	22 01	024	PRTX	-14	045	÷	-24
004	CLRG	16-53	025	RCL9	36 09	046	PRTX	-14
005	P⇄S	16-51	026	1	01	047	RCLA	36 11
006	CLRG	16-53	027	-	-45	048	X⇄Y	-41
007	SF0	16 21 00	028	STOA	35 11	049	RCLB	36 12
008	GTO1	22 01	029	P⇄S	16-51	050	X⇄Y	-41
009	*LBLB	21 12	030	RCL9	36 09	051	RTN	24
010	F0?	16 23 00	031	1	01	052	*LBL1	21 01
011	GTO3	22 03	032	-	-45	053	RCLE	36 15
012	GTO1	22 01	033	PRTX	-14	054	RCLD	36 14
013	*LBL3	21 03	034	STOB	35 12	055	÷	-24
014	CF0	16 22 00	035	S	16 54	056	PRTX	-14
015	P⇄S	16-51	036	X²	53	057	RCLB	36 12
016	*LBL1	21 01	037	STOD	35 14	058	X⇄Y	-41
017	Σ+	56	038	PRTX	-14	059	RCLA	36 11
018	RTN	24	039	RCLA	36 11	060	X⇄Y	-41
019	*LBLC	21 13	040	PRTX	-14	061	RTN	24
020	P⇄S	16-51	041	RCLD	36 14	062	R/S	51
021	S	16 54	042	RCLE	36 15			

A.13 DETAILS OF VARIANCE-2 PROGRAM: COMPARISON OF TWO VARIANCES

Lines 1 to 8 accept the input of the first set of data. With the entry of the first data value, the registers are all cleared.

Lines 9 to 15 accept the input of the second set of data. With the first entry of the second set of data, the storage registers are shifted, and both sets of data are processed by the next lines.

Lines 16 to 18 use the KEY Σ+ to store both Σx, Σx^2, Σy, Σy^2, and the number of points in each set of data.

Lines 19 to 29 use KEY s to calculate the variance estimate of the first set of data. They also calculate the degrees of freedom of the second set and store the result. The variance estimate $s^2(x)$ is printed and the storage areas are shifted for the calculation of the second set of data.

Lines 30 to 40 use KEY s to calculate the variance estimate of the second set of data. They calculate the degrees of freedom of the first set, and print the results.

Lines 41 to 51 check whether the first variance estimate is greater than the second. If so, the program goes to line 52. If not, the program divides the second by the first to obtain the F value. It prints this value and then shifts the degrees of freedom to the proper order for using program Variance-9 to calculate the probability.

Lines 52 to 61 are used if the first variance estimate is larger than the second. In this case, the first is divided by the second to obtain the F value, which is printed. The degrees of freedom and the F value are then put in the proper order for Variance-9.

Table A.12 Variance-3 Program: Bartlett χ^2 Test for Variance Homogeneity (HP-97)

Input: x_i, KEY A, . . . , KEY B KEY C	Input: ν_i, ENTER ↑; $s^2(x_i)$, KEY D KEY C
Output: $s^2(x)$, ν, for each set, ν total, χ^2, probability	Output: ν total, χ^2, probability

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	039	0	00	077	1/X	52
002	F0?	16 23 00	040	STO4	35 04	078	-	-45
003	GTO1	22 01	041	STO5	35 05	079	x	-35
004	CLRG	16-53	042	STO6	35 06	080	1	01
005	P⇄S	16-51	043	STO7	35 07	081	+	-55
006	CLRG	16-53	044	STO8	35 08	082	÷	-24
007	SF0	16 21 00	045	STO9	35 09	083	PRTX	-14
008	*LBL1	21 01	046	P⇄S	16-51	084	RCLD	36 14
009	Σ+	56	047	RTN	24	085	X⇄Y	-41
010	RTN	24	048	*LBLD	21 14	086	CF0	16 22 00
011	*LBLB	21 12	049	F0?	16 23 00	087	*LBLe	21 16 15
012	S	16 54	050	GTO1	22 01	088	CLRG	16-53
013	X²	53	051	CLRG	16-53	089	STO9	35 09
014	STOB	35 12	052	*LBL1	21 01	090	R↓	-31
015	PRTX	-14	053	SF0	16 21 00	091	STO1	35 01
016	P⇄S	16-51	054	STOB	35 12	092	2	02
017	RCL9	36 09	055	R↓	-31	093	x	-35
018	1	01	056	STOA	35 11	094	1	01
019	-	-45	057	GTO6	22 06	095	-	-45
020	STOA	35 11	058	*LBLC	21 13	096	√X	54
021	PRTX	-14	059	RCL5	36 05	097	5	05
022	SPC	16-11	060	RCL4	36 04	098	+	-55
023	P⇄S	16-51	061	÷	-24	099	RCL9	36 09
024	*LBL6	21 06	062	LN	32	100	2	02
025	ST+4	35-55 04	063	RCL4	36 04	101	x	-35
026	RCLB	36 12	064	x	-35	102	√X	54
027	x	-35	065	RCL9	36 09	103	X≤Y?	16-35
028	ST+5	35-55 05	066	-	-45	104	GTO1	22 01
029	RCLB	36 12	067	3	03	105	1	01
030	LN	32	068	RCL0	36 00	106	EEX	-23
031	RCLA	36 11	069	1	01	107	CHS	-22
032	1/X	52	070	-	-45	108	5	05
033	ST+6	35-55 06	071	PRTX	-14	109	PRTX	-14
034	÷	-24	072	STOD	35 14	110	RTN	24
035	ST+9	35-55 09	073	x	-35	111	*LBL1	21 01
036	1	01	074	1/X	52	112	1	01
037	ST+0	35-55 00	075	RCL6	36 06	113	STO5	35 05
038	P⇄S	16-51	076	RCL4	36 04	114	RCL9	36 09

Table A.12 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
115	2	02	140	1	01	165	÷	-24
116	÷	-24	141	-	-45	166	ST×5	35-35 05
117	STOA	35 11	142	N!	16 52	167	2	02
118	RCL1	36 01	143	ST÷3	35-24 03	168	ST+1	35-55 01
119	4	04	144	GTO1	22 01	169	RCL5	36 05
120	÷	-24	145	*LBL2	21 02	170	ST+6	35-55 06
121	Y^x	31	146	.	-62	171	1	01
122	STOB	35 12	147	5	05	172	EEX	-23
123	RCLA	36 11	148	X=Y?	16-33	173	CHS	-22
124	CHS	-22	149	GTO3	22 03	174	4	04
125	e^x	33	150	X⇄Y	-41	175	X≤Y?	16-35
126	×	-35	151	1	01	176	GTO4	22 04
127	STO3	35 03	152	-	-45	177	RCL6	36 06
128	RCLB	36 12	153	ST÷3	35-24 03	178	1	01
129	ST×3	35-35 03	154	GTO2	22 02	179	+	-55
130	RCL1	36 01	155	*LBL3	21 03	180	RCL3	36 03
131	2	02	156	Pi	16-24	181	×	-35
132	+	-55	157	√X	54	182	1	01
133	2	02	158	ST÷3	35-24 03	183	X⇄Y	-41
134	÷	-24	159	*LBL1	21 01	184	-	-45
135	STO4	35 04	160	2	02	185	PRTX	-14
136	INT	16 34	161	ST+1	35-55 01	186	RTN	24
137	LSTX	16-63	162	*LBL4	21 04			
138	X≠Y?	16-32	163	RCL9	36 09			
139	GTO2	22 02	164	RCL1	36 01			

A.14 DETAILS OF VARIANCE-3 PROGRAM: BARTLETT χ^2 TEST

Lines 1 to 10 take the individual values for each set of data and accumulate the sum and sum of squares with KEY $\Sigma+$.

Lines 11 to 47 use the built-in functions to obtain the variance estimate for each set of data, and then the sums required for the evaluation of Eqs. (6.11) and (6.12) are accumulated. The storage areas used for the built-in functions are then cleared for the next set of data.

Lines 48 to 57 take the input data if the variance estimates and degrees of freedom are already available. The program is directed back to line 24 to accumulate sums for the χ^2 calculation.

Lines 58 to 86 calculate χ^2 from Eqs. (6.12) and (6.11) and place the results in proper sequence for use of χ^2-5 to calculate the probability.

Lines 87 to 186 are the same as the χ^2-5 program for calculating χ^2 probability.

Table A.13 Variance-4 Program: One-Factor Analysis of Variance, with Replicates (HP-97)

Input: x_i, KEY A for each group of data
KEY B when data for group have been entered
KEY C after KEY B for the last group

Output: $\bar{x}$ for each group
Group degrees of freedom
Group variance estimate
Error degrees of freedom
Error variance estimate
F value

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	027	1	01	053	RCLI	36 46
002	F0?	16 23 00	028	ST+3	35-55 03	054	-	-45
003	GTO1	22 01	029	RTN	24	055	RCL3	36 03
004	CLRG	16-53	030	*LBLC	21 13	056	1	01
005	P⇄S	16-51	031	P⇄S	16-51	057	-	-45
006	CLRG	16-53	032	RCL4	36 04	058	PRTX	-14
007	SF0	16 21 00	033	X²	53	059	STOA	35 11
008	*LBL1	21 01	034	RCL9	36 09	060	÷	-24
009	ST+1	35-55 01	035	÷	-24	061	PRTX	-14
010	ISZI	16 26 46	036	STOI	35 46	062	RCL5	36 05
011	Σ+	56	037	RCL9	36 09	063	÷	-24
012	RTN	24	038	RCL5	36 05	064	STOB	35 12
013	*LBLB	21 12	039	P⇄S	16-51	065	SPC	16-11
014	RCL1	36 01	040	STO5	35 05	066	RCL4	36 04
015	X²	53	041	R↓	-31	067	PRTX	-14
016	RCLI	36 46	042	STO9	35 09	068	RCL5	36 05
017	ST+4	35-55 04	043	RCL5	36 05	069	PRTX	-14
018	÷	-24	044	RCL2	36 02	070	CF0	16 22 00
019	ST+2	35-55 02	045	-	-45	071	RCLA	36 11
020	RCL1	36 01	046	RCL4	36 04	072	RCL4	36 04
021	÷	-24	047	RCL3	36 03	073	RCLB	36 12
022	PRTX	-14	048	-	-45	074	SPC	16-11
023	SPC	16-11	049	STO4	35 04	075	PRTX	-14
024	0	00	050	÷	-24	076	RTN	24
025	STOI	35 46	051	STO5	35 05	077	R/S	51
026	STO1	35 01	052	RCL2	36 02			

A.15 DETAILS OF VARIANCE-4 PROGRAM: ONE-FACTOR ANALYSIS OF VARIANCE

Lines 1 to 12 process the individual data values using KEY $\Sigma+$ to obtain Σx and Σx^2 and the total amount of data to calculate the total sum of squares.

Lines 13 to 29 calculate the group total squared divided by the number in each group, for the group sum of squares. The mean for each group is also calculated and printed.

Lines 30 to 45 calculate the error sum of squares by subtracting the group sum of squares from the total sum of squares.

Lines 46 to 51 calculate the error degrees of freedom and the error variance estimate and store the results.

Lines 52 to 61 divide the group sum of squares by the group degrees of freedom, to obtain the group variance estimate, and print the latter two values.

Lines 62 to 65 calculate and store the F value by dividing the group variance estimate by the error variance estimate.

Lines 66 to 76 print the error degrees of freedom, the error variance estimate, and the F value. The group degrees of freedom, the error degrees of freedom and the F value are then put in the proper sequence for the probability calculation by the Variance-9 program if it is desired.

Table A.14 Variance-5 Program: Two-Factor Analysis of Variance, No Replicates (HP-97)

Input: Number of rows r, KEY A
Data by columns, KEY A
When all data are entered, KEY B

Output: Row factor degrees of freedom
Row factor variance estimate
Row factor F value

Column factor degrees of freedom
Column factor variance estimate
Column factor F value

Error degrees of freedom
Error variance estimate

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	031	STOB	35 12	061	-	-45
002	F0?	16 23 00	032	RCLA	36 11	062	STOD	35 14
003	GTO1	22 01	033	STOI	35 46	063	RCLB	36 12
004	CLRG	16-53	034	*LBLa	21 16 11	064	RCLI	36 46
005	P⇄S	16-51	035	RCLi	36 45	065	-	-45
006	CLRG	16-53	036	X^2	53	066	STOB	35 12
007	STOA	35 11	037	RCLD	36 14	067	RCL0	36 00
008	STOI	35 46	038	+	-55	068	RCLB	36 12
009	SF0	16 21 00	039	STOD	35 14	069	-	-45
010	RTN	24	040	DSZI	16 25 46	070	RCLD	36 14
011	*LBL1	21 01	041	GTOa	22 16 11	071	-	-45
012	ST+i	35-55 45	042	P⇄S	16-51	072	RCL9	36 09
013	ST+0	35-55 00	043	RCL5	36 05	073	RCLA	36 11
014	Σ+	56	044	RCL4	36 04	074	-	-45
015	DSZI	16 25 46	045	X^2	53	075	RCL9	36 09
016	RTN	24	046	RCL9	36 09	076	RCLA	36 11
017	RCL0	36 00	047	÷	-24	077	÷	-24
018	X^2	53	048	STOI	35 46	078	-	-45
019	RCLB	36 12	049	-	-45	079	1	01
020	+	-55	050	RCL9	36 09	080	+	-55
021	STOB	35 12	051	P⇄S	16-51	081	STO4	35 04
022	0	00	052	STO9	35 09	082	÷	-24
023	STO0	35 00	053	R↓	-31	083	STO5	35 05
024	RCLA	36 11	054	STO0	35 00	084	RCLD	36 14
025	STOI	35 46	055	RCLD	36 14	085	RCLA	36 11
026	RTN	24	056	RCL9	36 09	086	1	01
027	*LBLB	21 12	057	RCLA	36 11	087	-	-45
028	RCLB	36 12	058	÷	-24	088	PRTX	-14
029	RCLA	36 11	059	÷	-24	089	÷	-24
030	÷	-24	060	RCLI	36 46	090	PRTX	-14

Table A.14 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
091	RCL5	36 05	100	-	-45	109	PRTX	-14
092	÷	-24	101	PRTX	-14	110	RCL5	36 05
093	PRTX	-14	102	÷	-24	111	PRTX	-14
094	SPC	16-11	103	PRTX	-14	112	CF0	16 22 00
095	RCLB	36 12	104	RCL5	36 05	113	SPC	16-11
096	RCL9	36 09	105	÷	-24	114	RTN	24
097	RCLA	36 11	106	PRTX	-14	115	R/S	51
098	÷	-24	107	SPC	16-11			
099	1	01	108	RCL4	36 04			

A.16 DETAILS OF THE VARIANCE-5 PROGRAM: TWO-FACTOR ANALYSIS OF VARIANCE

Lines 1 to 10 clear all the registers and store the number of rows.

Lines 11 to 16 accumulate the data for each column. The column total is accumulated in Register 0, and the value for each row is accumulated in a different register, set by the index *i*. The total for all the data is accumulated by KEY Σ+.

Lines 17 to 26 calculate the square of the column total and accumulate the sum of the squares in Register B. They reset zero in Register 0 for the next column total, and reset the *i* index to accumulate the next row totals. Lines 17 to 26 are used after the program has counted data entries equal to the number of rows.

Lines 27 to 33 calculate $\Sigma(\text{column total})^2/\text{number in column}$, store the result in Register B and reset index *i*.

Lines 34 to 41 calculate the square of the row totals stored in the registers determined by index *i*, sum the total, and store the value in Register D.

Lines 42 to 54 calculate $\Sigma x^2/n$ and store the value in Register I, and calculate the total sum of squares and store the value in Register 0.

Lines 55 to 62 calculate the row sum of squares and store the result in Register D.

Lines 63 to 66 calculate the column sum of squares and store the result in Register B.

Lines 67 to 83 calculate the error sum of squares, the error degrees of freedom, and the error variance estimate. The last two values are stored in Registers 4 and 5.

Lines 84 to 94 calculate and print the row degrees of freedom, the row variance estimate, and the row F value.

Lines 95 to 107 calculate and print the same values for the column factor.

Lines 108 to 114 print the error degrees of freedom and the error variance estimate, and clear Flag 0 for the start of another calculation.

All the calculations follow the formulas of Table 6.6.

If the column total or column means are wanted as part of the output, the instructions could be inserted between lines 17 and 18. The column total is recalled at line 17 to be squared for the analysis of variance calculation. Before it is squared, it could be printed or divided by the number of rows to obtain the column mean.

If the row total or row means were wanted the instructions could be inserted between lines 35 and 36. The row totals are recalled at line 35. The number of columns is the total data stored in secondary register 9 divided by the number of rows. This value could be used to get the row mean.

The program can handle up to 13 rows of data with no limit on the number of columns.

Table A.15 Variance-6 Program: Two-Factor Analysis of Variance with Replicates (HP-97)

Input: Number of rows, ENTER ↑; number of replicates, KEY A
Data by columns, x_i, KEY A
When all the data are entered, KEY B

Output: Row degrees of freedom / Row variance estimate / Row F value
Column degrees of freedom / Column variance estimate / Column F value
Interaction degrees of freedom / Interaction variance estimate / Interaction F value
Error degrees of freedom / Error variance estimate

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	032	P⇄S	16-51	063	DSZI	16 25 46
002	F0?	16 23 00	033	0	00	064	GTOa	22 16 11
003	GTO1	22 01	034	STO0	35 00	065	P⇄S	16-51
004	CLRG	16-53	035	STOE	35 15	066	RCL5	36 05
005	P⇄S	16-51	036	DSZI	16 25 46	067	RCL4	36 04
006	CLRG	16-53	037	RTN	24	068	X²	53
007	STOC	35 13	038	RCLD	36 14	069	RCL9	36 09
008	R↓	-31	039	X²	53	070	÷	-24
009	STOA	35 11	040	RCLB	36 12	071	STOI	35 46
010	STOI	35 46	041	+	-55	072	-	-45
011	SF0	16 21 00	042	STOB	35 12	073	RCL9	36 09
012	RTN	24	043	0	00	074	RCL3	36 03
013	*LBL1	21 01	044	STOD	35 14	075	P⇄S	16-51
014	ST+0	35-55 00	045	RCLA	36 11	076	STO3	35 03
015	Σ+	56	046	STOI	35 46	077	R↓	-31
016	1	01	047	RTN	24	078	STO9	35 09
017	RCLE	36 15	048	*LBLB	21 12	079	R↓	-31
018	+	-55	049	RCLB	36 12	080	STO0	35 00
019	STOE	35 15	050	RCLA	36 11	081	RCLD	36 14
020	RCLC	36 13	051	RCLC	36 13	082	RCL9	36 09
021	X≠Y?	16-32	052	×	-35	083	RCLA	36 11
022	RTN	24	053	÷	-24	084	÷	-24
023	RCL0	36 00	054	STOB	35 12	085	÷	-24
024	ST+i	35-55 45	055	RCLA	36 11	086	RCLI	36 46
025	RCLD	36 14	056	STOI	35 46	087	-	-45
026	+	-55	057	*LBLa	21 16 11	088	STOD	35 14
027	STOD	35 14	058	RCLi	36 45	089	RCLB	36 12
028	RCL0	36 00	059	X²	53	090	RCLI	36 46
029	X²	53	060	RCLD	36 14	091	-	-45
030	P⇄S	16-51	061	+	-55	092	STOB	35 12
031	ST+3	35-55 03	062	STOD	35 14	093	RCL3	36 03

Table A.15 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
094	RCLC	36 13	120	RCL3	36 03	146	×	-35
095	÷	-24	121	RCLD	36 14	147	÷	-24
096	RCLI	36 46	122	-	-45	148	1	01
097	-	-45	123	RCLB	36 12	149	-	-45
098	STO3	35 03	124	-	-45	150	PRTX	-14
099	RCL0	36 00	125	RCL1	36 01	151	÷	-24
100	RCL3	36 03	126	RCL9	36 09	152	PRTX	-14
101	-	-45	127	RCLA	36 11	153	RCL7	36 07
102	RCL9	36 09	128	÷	-24	154	÷	-24
103	1	01	129	RCLC	36 13	155	PRTX	-14
104	RCLC	36 13	130	÷	-24	156	SPC	16-11
105	1/X	52	131	1	01	157	RCL8	36 08
106	-	-45	132	-	-45	158	PRTX	-14
107	×	-35	133	×	-35	159	RCL7	36 07
108	STO4	35 04	134	STO8	35 08	160	PRTX	-14
109	÷	-24	135	÷	-24	161	RCL5	36 05
110	STO5	35 05	136	STO7	35 07	162	÷	-24
111	RCLD	36 14	137	RCL2	36 02	163	PRTX	-14
112	RCLA	36 11	138	RCL7	36 07	164	SPC	16-11
113	1	01	139	÷	-24	165	RCL4	36 04
114	-	-45	140	PRTX	-14	166	PRTX	-14
115	PRTX	-14	141	SPC	16-11	167	RCL5	36 05
116	STO1	35 01	142	RCLB	36 12	168	PRTX	-14
117	÷	-24	143	RCL9	36 09	169	CF0	16 22 00
118	PRTX	-14	144	RCLA	36 11	170	RTN	24
119	STO2	35 02	145	RCLC	36 13			

A.17 DETAILS OF THE VARIANCE-6 PROGRAM: TWO-FACTOR ANALYSIS OF VARIANCE WITH REPLICATES

Lines 1 to 12 clear all registers and store the number of rows and the number of replicates.

Lines 13 to 22 process the replicate data. The replicate total is accumulated in Register 0. The total of all the data is accumulated with KEY Σ+.

Lines 23 to 37 process the column data. The replicate totals are accumulated in separate registers, designated by index i to obtain the row totals. The replicate total is squared and the sum of the squares of the replicate totals is stored in Register 13. The column total is accumulated in Register D.

Lines 38 to 47 process the data when a total column has been entered. The column total is squared and the sum is accumulated in Register B.

Lines 48 to 56 calculate the Σ(column total)$^2/rk$ of Table 6.7.

Lines 57 to 64 calculate the sum of the squares of the row totals obtained from index i.

Lines 65 to 80 calculate $\Sigma x^2/n$ and the total sum of squares and store the results in Registers I and 0.

Lines 81 to 92 calculate the row sum of squares and the column sum of squares and store the results in Registers D and B.

Lines 93 to 98 calculate the term called "sub-total" in Table 6.7 and store the result in Register 3.

Lines 99 to 110 calculate the error degrees of freedom and the error variance estimate, following the last line in Table 6.7, and store the results in Registers 4 and 5.

Lines 111 to 118 calculate and print the row degrees of freedom and the row variance estimate.

Lines 119 to 140 calculate the interaction degrees of freedom and the interaction variance estimate, and store the values. The row variance estimate is divided by the interaction variance estimate to obtain the row F value, which is printed.

Lines 141 to 156 calculate the column degrees of freedom, the column variance estimate and the column F value, and print the results.

Lines 157 to 170 recall and print the interaction degrees of freedom and variance estimate, and calculate and print the interaction F value. They then recall and print the error degrees of freedom and the error variance estimate.

All of the calculations follow Table 6.7.

The program can handle up to 12 rows of data with no limit on the number of replicates or on the number of columns.

Table A.16 Variance-7 Program: Latin Square Analysis of Variance (HP-97)

Input: Number of rows, KEY A
Data by columns, x_i, KEY A
Data by letters, x_i, KEY B

Output: Row degrees of freedom
Row variance estimate
Row F value

Column degrees of freedom
Column variance estimate
Column F value

Letter degrees of freedom
Letter variance estimate
Letter F value

Error degrees of freedom
Error variance estimate

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	033	X²	53	065	RCLA	36 11
002	F0?	16 23 00	034	RCLD	36 14	066	STOI	35 46
003	GTO1	22 01	035	+	-55	067	*LBL6	21 06
004	CLRG	16-53	036	STOD	35 14	068	RCLi	36 45
005	P⇄S	16-51	037	DSZI	16 25 46	069	ST+0	35-55 00
006	CLRG	16-53	038	GTO2	22 02	070	X²	53
007	STOA	35 11	039	SF1	16 21 01	071	RCLC	36 13
008	STOI	35 46	040	RCLA	36 11	072	+	-55
009	SF0	16 21 00	041	STOI	35 46	073	STOC	35 13
010	RTN	24	042	*LBL4	21 04	074	DSZI	16 25 46
011	*LBL1	21 01	043	0	00	075	GTO6	22 06
012	ST+i	35-55 45	044	STOi	35 45	076	RCL0	36 00
013	ST+0	35-55 00	045	DSZI	16 25 46	077	P⇄S	16-51
014	Σ+	56	046	GTO4	22 04	078	RCL4	36 04
015	DSZI	16 25 46	047	RCLA	36 11	079	X=Y?	16-33
016	RTN	24	048	STOI	35 46	080	GTO1	22 01
017	RCL0	36 00	049	RCLE	36 15	081	PRTX	-14
018	X²	53	050	*LBL1	21 01	082	R↓	-31
019	RCLB	36 12	051	ST+i	35-55 45	083	PRTX	-14
020	+	-55	052	1	01	084	RTN	24
021	STOB	35 12	053	RCLC	36 13	085	*LBL1	21 01
022	0	00	054	+	-55	086	RCL5	36 05
023	STO0	35 00	055	STOC	35 13	087	RCL4	36 04
024	RCLA	36 11	056	RCLA	36 11	088	X²	53
025	STOI	35 46	057	X≠Y?	16-32	089	RCL9	36 09
026	RTN	24	058	RTN	24	090	÷	-24
027	*LBLB	21 12	059	0	00	091	STOI	35 46
028	F1?	16 23 01	060	STOC	35 13	092	-	-45
029	GTO1	22 01	061	DSZI	16 25 46	093	RCL9	36 09
030	STOE	35 15	062	RTN	24	094	P⇄S	16-51
031	*LBL2	21 02	063	CF0	16 22 00	095	STO9	35 09
032	RCLi	36 45	064	CF1	16 22 01	096	R↓	-31

Table A.16 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
097	STO1	35 01	119	RCLD	36 14	141	RCL5	36 05
098	RCLD	36 14	120	-	-45	142	PRTX	-14
099	RCLA	36 11	121	RCLC	36 13	143	SPC	16-11
100	÷	-24	122	-	-45	144	RTN	24
101	RCLI	36 46	123	RCL9	36 09	145	*LBL1	21 01
102	-	-45	124	RCLA	36 11	146	RCLA	36 11
103	STOD	35 14	125	3	03	147	1	01
104	RCLB	36 12	126	x	-35	148	-	-45
105	RCLA	36 11	127	-	-45	149	PRTX	-14
106	÷	-24	128	2	02	150	÷	-24
107	RCLI	36 46	129	+	-55	151	PRTX	-14
108	-	-45	130	STO4	35 04	152	RCL5	36 05
109	STOB	35 12	131	÷	-24	153	X=0?	16-43
110	RCLC	36 13	132	STO5	35 05	154	GTO1	22 01
111	RCLA	36 11	133	RCLD	36 14	155	÷	-24
112	÷	-24	134	GSB1	23 01	156	PRTX	-14
113	RCLI	36 46	135	RCLB	36 12	157	*LBL1	21 01
114	-	-45	136	GSB1	23 01	158	SPC	16-11
115	STOC	35 13	137	RCLC	36 13	159	RTN	24
116	RCL1	36 01	138	GSB1	23 01			
117	RCLB	36 12	139	RCL4	36 04			
118	-	-45	140	PRTX	-14			

A.18 DETAILS OF THE VARIANCE-7 PROGRAM: LATIN SQUARE ANALYSIS OF VARIANCE

Lines 1 to 10 clear the registers and take input of the number of rows which value is stored in Registers A and I as a control.

Lines 11 to 16 take the data by columns. The sum of the entry for each row is accumulated in a different register, determined by the index *i*. The column total is accumulated in Register 0.

Lines 17 to 26 square the column total and accumulate the sum in Register B, and reset the indices for the next column of data.

Lines 27 to 49 recall the sum of each row, square the row total, and sum the squares for the calculation of the row variance estimate. The value is stored in Register D. The index *i* is reset to 0 for the letter totals, and the first value of the letter data is put in position for the next calculation.

Lines 50 to 62 accumulate the totals for each letter in a different register, controlled by index *i*. The number of entries for each letter are checked with Register C, and when the value equals the number of rows, the indices are reset for the next letter data.

Lines 63 to 66 clear the flags used and start the calculation of the analysis of variance.

Lines 67 to 75 accumulate the square of the totals for each letter, and add the totals in Register 0.

Lines 76 to 84 compare the total of the letter data with the total of the data originally entered. If the totals are different, the program prints the two values and waits. The program can either be continued with KEY R/S, or started over if the difference is too large.

Lines 85 to 97 calculate $(\Sigma x)^2/n$, and the total sums of squares and store the results in Registers I and 1.

Lines 98 to 115 calculate the row sum of squares, the column sum of squares, and the letter sum of squares, following Table 6.9, and store the results in Registers D, B, and C.

Lines 116 to 132 calculate the error sum of squares, the error degrees of freedom, and the error variance estimate. The latter two values are stored in Registers 4 and 5.

Lines 133 to 138 use the same subroutine at line 145, to calculate the row, column, and letter degrees of freedom (which are all the same), and variance estimates, and to print the results. The subroutine also divides the variance estimates by the error estimate to obtain the F value, which is printed.

Lines 139 to 144 recall and print the error degrees of freedom and the error variance and end the program.

Lines 145 to 159 are the subroutine to print the row, column, and letter degrees of freedom, variance estimates, and F values. Line 153 checks the error variance for a value of zero. If it is zero, the F value is not calculated.

All of the calculations follow the formulas in Table 6.3.

Table A.17 Variance-8 Program: 2^n Factorial Analysis of Variance (HP-97)

Input: 2^n (4, 8, or 16), KEY A
Data in standard order: (1), *a*, *b*, *ab*, *c*, *ac*, . . . , KEY A

Output: Variance estimates in order of data input
Mean differences in order of data input

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	037	X≠Y?	16-32	073	STOD	35 14
002	F0?	16 23 00	038	GTO1	22 01	074	RCL5	36 05
003	GTO1	22 01	039	RCL2	36 02	075	STO3	35 03
004	STOA	35 11	040	RCL3	36 03	076	RCL6	36 06
005	STOI	35 46	041	STO2	35 02	077	STOE	35 15
006	SF0	16 21 00	042	R↓	-31	078	RCL7	36 07
007	RTN	24	043	STO3	35 03	079	STO4	35 04
008	*LBL1	21 01	044	GTO3	22 03	080	RCL9	36 09
009	STOi	35 45	045	*LBL1	21 01	081	STO5	35 05
010	DSZI	16 25 46	046	8	08	082	RCLC	36 13
011	RTN	24	047	X≠Y?	16-32	083	STO9	35 09
012	RCLA	36 11	048	GTO1	22 01	084	RCL8	36 08
013	STOI	35 46	049	RCL2	36 02	085	STOC	35 13
014	LN	32	050	RCL3	36 03	086	P⇄S	16-51
015	2	02	051	STO2	35 02	087	RCL2	36 02
016	LN	32	052	R↓	-31	088	RCLC	36 13
017	÷	-24	053	RCL4	36 04	089	STO2	35 02
018	STOB	35 12	054	RCL5	36 05	090	R↓	-31
019	*LBL2	21 02	055	STO3	35 03	091	STOC	35 13
020	RCLi	36 45	056	R↓	-31	092	RCL0	36 00
021	DSZI	16 25 46	057	RCL7	36 07	093	RCLD	36 14
022	RCLi	36 45	058	STO4	35 04	094	STO0	35 00
023	+	-55	059	R↓	-31	095	R↓	-31
024	RCLi	36 45	060	RCL6	36 06	096	STOD	35 14
025	ISZI	16 26 46	061	STO7	35 07	097	RCL1	36 01
026	RCLi	36 45	062	R↓	-31	098	RCLE	36 15
027	-	-45	063	STO6	35 06	099	STO1	35 01
028	X⇄Y	-41	064	R↓	-31	100	R↓	-31
029	STOi	35 45	065	STO5	35 05	101	STOE	35 15
030	DSZI	16 25 46	066	GTO3	22 03	102	RCL5	36 05
031	R↓	-31	067	*LBL1	21 01	103	RCL4	36 04
032	STOi	35 45	068	RCL2	36 02	104	STO5	35 05
033	DSZI	16 25 46	069	STOC	35 13	105	RCLC	36 13
034	GTO2	22 02	070	RCL3	36 03	106	STO4	35 04
035	4	04	071	STO2	35 02	107	R↓	-31
036	RCLA	36 11	072	RCL4	36 04	108	R↓	-31

Table A.17 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
109	STOC	35 13	126	1	01	143	CF0	16 22 00
110	RCL3	36 03	127	-	-45	144	SPC	16-11
111	RCLD	36 14	128	STOB	35 12	145	RCLA	36 11
112	STO3	35 03	129	X>0?	16-44	146	1	01
113	R↓	-31	130	GTO2	22 02	147	-	-45
114	STOD	35 14	131	RCLA	36 11	148	STOI	35 46
115	P⇄S	16-51	132	1	01	149	*LBL6	21 06
116	RCLC	36 13	133	-	-45	150	RCLi	36 45
117	STO8	35 08	134	STOI	35 46	151	RCLA	36 11
118	RCLD	36 14	135	*LBL4	21 04	152	2	02
119	STO7	35 07	136	RCLi	36 45	153	÷	-24
120	RCLE	36 15	137	X²	53	154	÷	-24
121	STO6	35 06	138	RCLA	36 11	155	PRTX	-14
122	*LBL3	21 03	139	÷	-24	156	DSZI	16 25 46
123	RCLA	36 11	140	PRTX	-14	157	GTO6	22 06
124	STOI	35 46	141	DSZI	16 25 46	158	RTN	24
125	RCLB	36 12	142	GTO4	22 04			

A.19 DETAILS OF THE VARIANCE-8 PROGRAM: 2^n FACTORIAL CALCULATION

Lines 1 to 7 take input of the size of the experiment, 2^n, and store the value in Registers A and I.

Lines 8 to 11 take the experimental data input and store each value in a different register, determined by index *i*. The program counts the input and when 2^n items have been entered, proceeds to the next operation.

Lines 12 to 18 calculate the value of n from the 2^n input, and store the value in Register B.

Lines 19 to 34 make the first calculation of Table 6.10, adding the data in pairs and then subtracting the first from the second of each pair. The results are stored in the registers set by index *i*. This calculation will be carried out n times.

Lines 35 to 44 shift the values from the calculation of lines 19 to 34 back to the original registers for the next calculation if $n = 2$.

Lines 45 to 66 do the same thing if $n = 3$.

Lines 67 to 121 shift the results from the calculation of lines 19 to 34 back to the original registers if $n = 4$.

Lines 122 to 130 decrease the value in Register B by 1, and return the program to line 19 for the next calculation of Table 6.10. The operation is carried out n times, controlled by the value stored in Register B at line 18.

Lines 131 to 142 calculate the variance estimates from the last column of Table 6.10, by squaring and dividing by 2^n. The results are printed in the order in which the data were entered.

Lines 143 to 158 clear Flag 0, set at the start of the program, and calculate the mean differences from the last column of Table 6.10 by dividing that value by 2^{n-1}. The results are also printed in the order of data input.

Table A.18 Variance-9 Program: *F* Probability Calculation (HP-97)

Input: ν_1, numerator degrees of freedom, ENTER ↑
ν_2, denominator degrees of freedom, ENTER ↑
F, KEY A
Output: Probability associated with *F* at ν_1 and ν_2 degrees of freedom

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	037	R↓	-31	073	2	02
002	SF1	16 21 01	038	STO1	35 01	074	×	-35
003	CLRG	16-53	039	*LBL1	21 01	075	RCL0	36 00
004	STOC	35 13	040	RCL1	36 01	076	+	-55
005	R↓	-31	041	2	02	077	÷	-24
006	STO2	35 02	042	GSB9	23 09	078	RCL6	36 06
007	R↓	-31	043	1	01	079	×	-35
008	STO1	35 01	044	RCL4	36 04	080	ST×7	35-35 07
009	RCL2	36 02	045	-	-45	081	1	01
010	RCL2	36 02	046	STO6	35 06	082	ST+7	35-55 07
011	RCL1	36 01	047	RCL2	36 02	083	DSZI	16 25 46
012	RCLC	36 13	048	STO8	35 08	084	GTO2	22 02
013	×	-35	049	2	02	085	*LBLe	21 16 15
014	+	-55	050	STO9	35 09	086	RCL7	36 07
015	÷	-24	051	1	01	087	RCLD	36 14
016	STO4	35 04	052	STO7	35 07	088	×	-35
017	RCL1	36 01	053	STOD	35 14	089	RCLE	36 15
018	2	02	054	RCL4	36 04	090	+	-55
019	÷	-24	055	RCL2	36 02	091	RCL5	36 05
020	FRC	16 44	056	2	02	092	×	-35
021	X=0?	16-43	057	÷	-24	093	F0?	16 23 00
022	GTO1	22 01	058	Y^x	31	094	GTO5	22 05
023	CF1	16 22 01	059	STO5	35 05	095	F1?	16 23 01
024	RCL2	36 02	060	RCL1	36 01	096	GTO4	22 04
025	2	02	061	2	02	097	CHS	-22
026	÷	-24	062	X=Y?	16-33	098	1	01
027	FRC	16 44	063	GTOe	22 16 15	099	+	-55
028	X≠0?	16-42	064	*LBL2	21 02	100	*LBL4	21 04
029	GTO3	22 03	065	RCL8	36 08	101	PRTX	-14
030	1	01	066	RCLI	36 46	102	SPC	16-11
031	RCL4	36 04	067	2	02	103	CF0	16 22 00
032	-	-45	068	×	-35	104	CF1	16 22 01
033	STO4	35 04	069	+	-55	105	CF2	16 22 02
034	RCL2	36 02	070	RCL9	36 09	106	R/S	51
035	RCL1	36 01	071	-	-45	107	*LBL3	21 03
036	STO2	35 02	072	RCLI	36 46	108	SF0	16 21 00

Table A.18 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
109	RCL1	36 01	145	GTO2	22 02	181	÷	-24
110	RCL2	36 02	146	RCLE	36 15	182	ST×9	35-35 09
111	÷	-24	147	RCL5	36 05	183	DSZI	16 25 46
112	RCLC	36 13	148	×	-35	184	DSZI	16 25 46
113	×	-35	149	*LBL5	21 05	185	GTOd	22 16 14
114	$\sqrt{X}$	54	150	F2?	16 23 02	186	RCL9	36 09
115	RAD	16-22	151	GTO1	22 01	187	ST×5	35-35 05
116	TAN^{-1}	16 43	152	STOA	35 11	188	RCL4	36 04
117	STO3	35 03	153	SF2	16 21 02	189	SIN	41
118	STO4	35 04	154	RCL4	36 04	190	ST×5	35-35 05
119	STOE	35 15	155	SIN	41	191	RCL4	36 04
120	COS	42	156	X^2	53	192	COS	42
121	X^2	53	157	STO6	35 06	193	RCL2	36 02
122	STO6	35 06	158	0	00	194	Y^x	31
123	RCL4	36 04	159	STOE	35 15	195	ST×5	35-35 05
124	SIN	41	160	1	01	196	1	01
125	RCL4	36 04	161	STO3	35 03	197	STO9	35 09
126	COS	42	162	STOD	35 14	198	RCL1	36 01
127	×	-35	163	STO7	35 07	199	3	03
128	STOD	35 14	164	STO9	35 09	200	GSB9	23 09
129	2	02	165	RCL1	36 01	201	GTO2	22 02
130	ENT↑	-21	166	X≠Y?	16-32	202	*LBL1	21 01
131	Pi	16-24	167	GTO3	22 03	203	1	01
132	÷	-24	168	0	00	204	+	-55
133	STO5	35 05	169	GTO1	22 01	205	RCLA	36 11
134	RCL2	36 02	170	*LBL3	21 03	206	-	-45
135	3	03	171	RCL2	36 02	207	GTO4	22 04
136	GSB9	23 09	172	STO8	35 08	208	*LBL9	21 09
137	0	00	173	1	01	209	-	-45
138	STO8	35 08	174	-	-45	210	2	02
139	STO9	35 09	175	STOI	35 46	211	÷	-24
140	1	01	176	*LBLd	21 16 14	212	STOI	35 46
141	STO0	35 00	177	RCLI	36 46	213	RTN	24
142	STO7	35 07	178	RCLI	36 46			
143	RCL2	36 02	179	1	01			
144	X≠Y?	16-32	180	-	-45			

A.20 DETAILS OF THE VARIANCE-9 PROGRAM: *F* PROBABILITY

Lines 1 to 16 take the input data: ν_1, ν_2, and F. They store these values and calculate x of Table 6.16, Eq. (6.27).

Lines 17 to 23 check whether ν_1 is even or odd. If even, the program is directed to line 39. If odd, the program goes on to line 24.

Lines 24 to 29 check whether ν_2 is even or odd. If odd, the program goes to line 107. If even, the program goes on to line 30.

Lines 30 to 38 substitute $(1 - x)$ for x and interchange ν_1 and ν_2.

Lines 39 to 63 start the calculation of the probability if either ν_1 or ν_2 is even. The expressions in square brackets in Eqs. (6.32), (6.33), and (6.34) are all similar and may be evaluated by the same summation program. The evaluation of the probability involves obtaining the value of the expression in the square brackets and multiplying it by different terms depending on the value of the degrees of freedom. These lines in the program set the values for the calculation when either ν_1 or ν_2 is even.

Lines 64 to 84 evaluate the square bracket expressions of Eqs. (6.32), (6.33), and (6.34).

Lines 85 to 99 multiply the square bracket term by the terms shown in Eq. (6.25) or (6.26).

Lines 100 to 106 print the result and clear all flags.

Lines 107 to 145 start the calculation if both degrees of freedom are odd. θ of Table 6.16 is calculated, and the values are set for the evaluation of the square bracket term for the A solution to the two-part calculation involved. The program returns to line 64 for the evaluation.

Lines 146 to 185 calculate the factorial expression of Eq. (6.35).

Lines 186 to 201 set the values for the evaluation of the square bracket term for the B part of the solution, returning again to line 64.

Lines 202 to 207 carry out the calculation of Eq. (6.28), and return the program to line 100 to print the result.

Lines 208 to 213 are a routine used several times in the program and are made a subroutine to conserve program memory.

The program keeps track of the type of solution by the use of several flags, which are all cleared at the end so the program may be used again without interference from a previous solution.

Table A.19 Regression-1 Program: Straight-Line Linear Correlation (HP-97)

Equation: $y = b_0 + b_1 \cdot x$

Input: Experimental data—
y_i, ENTER ↑; x_i, KEY A
When all data are in, KEY C

Output: $\bar{x}$, $\bar{y}$
$(\Sigma x^2 - (\Sigma x)^2/n)$, b_1, b_0,
r, probability of r, $s(\hat{y})$

Input: Correlation results—
Degrees of freedom, ν, ENTER ↑
Correlation coefficient r, KEY D

Output: Probability value

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	035	RCL6	36 06	069	÷	-24
002	F0?	16 23 00	036	+	-55	070	-	-45
003	GTO1	22 01	037	RCL9	36 09	071	RTN	24
004	CLRG	16-53	038	÷	-24	072	*LBLD	21 14
005	P⇄S	16-51	039	PRTX	-14	073	ENT↑	-21
006	SF0	16 21 00	040	RCLB	36 12	074	X^2	53
007	*LBL1	21 01	041	RCLA	36 11	075	CHS	-22
008	Σ+	56	042	×	-35	076	1	01
009	RTN	24	043	STOE	35 15	077	+	-55
010	*LBLC	21 13	044	RCL7	36 07	078	$\sqrt{X}$	54
011	CF0	16 22 00	045	RCL6	36 06	079	÷	-24
012	$\bar{x}$	16 53	046	X^2	53	080	RAD	16-22
013	PRTX	-14	047	GSB1	23 01	081	TAN^{-1}	16 43
014	R↓	-31	048	STOD	35 14	082	STOA	35 11
015	PRTX	-14	049	÷	-24	083	SIN	41
016	SPC	16-11	050	$\sqrt{X}$	54	084	STOB	35 12
017	P⇄S	16-51	051	PRTX	-14	085	R↓	-31
018	RCL8	36 08	052	STOA	35 11	086	2	02
019	RCL4	36 04	053	RCLD	36 14	087	X>Y?	16-34
020	RCL6	36 06	054	RCLE	36 15	088	GTO4	22 04
021	×	-35	055	-	-45	089	X=Y?	16-33
022	GSB1	23 01	056	RCL9	36 09	090	GTO5	22 05
023	STOA	35 11	057	2	02	091	-	-45
024	RCL5	36 05	058	-	-45	092	STOI	35 46
025	RCL4	36 04	059	STO0	35 00	093	1	01
026	X^2	53	060	÷	-24	094	STO2	35 02
027	GSB1	23 01	061	$\sqrt{X}$	54	095	*LBL2	21 02
028	PRTX	-14	062	STOC	35 13	096	RCLI	36 46
029	÷	-24	063	SF2	16 21 02	097	1	01
030	STOB	35 12	064	RCL0	36 00	098	X=Y?	16-33
031	PRTX	-14	065	RCLA	36 11	099	GTO3	22 03
032	RCL4	36 04	066	GTOD	22 14	100	RCL2	36 02
033	×	-35	067	*LBL1	21 01	101	RCLA	36 11
034	CHS	-22	068	RCL9	36 09	102	COS	42

Table A.19 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
103	X²	53	121	RCL2	36 02	139	1	01
104	RCLI	36 46	122	RCLB	36 12	140	+	-55
105	÷	-24	123	x	-35	141	PRTX	-14
106	RCLI	36 46	124	RCLA	36 11	142	F2?	16 23 02
107	1	01	125	COS	42	143	GTO1	22 01
108	-	-45	126	x	-35	144	RTN	24
109	x	-35	127	RCLA	36 11	145	*LBL1	21 01
110	x	-35	128	+	-55	146	RCLC	36 13
111	1	01	129	GSBa	23 16 11	147	PRTX	-14
112	+	-55	130	GTO6	22 06	148	P⇄S	16-51
113	STO2	35 02	131	*LBL4	21 04	149	RTN	24
114	DSZI	16 25 46	132	RCLA	36 11	150	*LBLa	21 16 11
115	DSZI	16 25 46	133	GSBa	23 16 11	151	2	02
116	GTO2	22 02	134	GTO6	22 06	152	Pi	16-24
117	RCLB	36 12	135	*LBL5	21 05	153	÷	-24
118	x	-35	136	RCLB	36 12	154	x	-35
119	GTO6	22 06	137	*LBL6	21 06	155	RTN	24
120	*LBL3	21 03	138	CHS	-22			

A.21 DETAILS OF THE REGRESSION-1 PROGRAM: STRAIGHT-LINE LINEAR CORRELATION

Lines 1 to 9 take input of the y-x data. The dependent variable is entered first. With the first data entry, the program clears the registers. KEY $\Sigma+$ is used to accumulate the sums, the sums of squares, and the sum of cross-products required for the calculation.

Lines 10 to 17 start the calculation after all the data have been entered. The $\bar{x}$ and $\bar{y}$ values are obtained directly with the $\bar{x}$ KEY.

Lines 18 to 39 calculate $(\Sigma x^2 - (\Sigma x)^2/n)$, b_1, and b_0 from Eqs. (7.5) and (7.6) and print the results. The first term is used, as discussed in the text, to place a confidence band on the regression coefficient. [See Eq. (7.11).]

Lines 40 to 66 calculate $s(\hat{y})$ from Eq. (7.10) and the correlation coefficient r from Eq. (7.8) and print the r value. The degrees of freedom and the correlation coefficient are put in the proper sequence for the probability calculation.

Lines 67 to 71 are a routine that is used several times in the calculations: dividing by n and subtracting the result from some previous value.

Lines 72 to 155 are the same as the t-3 program for calculating the t probability. The correlation coefficient is related to t as shown in Eq. (7.9). The proper adjustment of the input degrees of freedom and r value to obtain the corresponding t

value is made in lines 73 to 79. The balance of the program is identical with t-3. The probability value and the value of $s(\hat{y})$ are printed at the end of the probability calculation.

The portion of the Regression-1 program used to calculate the probability can be used directly with input of ν and r with KEY D. The output will be the probability value.

Table A.20 Regression-2 Program: Linear Correlations other than Straight Lines (HP-97)

Equation: $y = b_0 + b_1 F(x)$

Input: Experimental data—
y_i, ENTER ↑; x_i, KEY A
KEY C

Output: b_1, b_0, r, probability of r, $s(\hat{y})$ for each $F(x)$

F functions:
x (line 10)
x^2 (lines 38–45)
$\sqrt{x}$ (line 36)
e^x (line 26)
$1/x$ (line 23)
$\log(x)$ (line 14)

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	034	P⇄S	16-51	067	P⇄S	16-51
002	F0?	16 23 00	035	RCLA	36 11	068	STO5	35 05
003	GTO3	22 03	036	√X	54	069	GSB1	23 01
004	CLRG	16-53	037	GSB7	23 07	070	GSB8	23 08
005	P⇄S	16-51	038	RCLA	36 11	071	P⇄S	16-51
006	CLRG	16-53	039	X²	53	072	GSB1	23 01
007	SF0	16 21 00	040	RCLB	36 12	073	GSB8	23 08
008	*LBL3	21 03	041	×	-35	074	RCL0	36 00
009	STOA	35 11	042	ST+0	35-55 00	075	STO8	35 08
010	Σ+	56	043	RCLA	36 11	076	RCL7	36 07
011	R↓	-31	044	X²	53	077	STO5	35 05
012	STOB	35 12	045	X²	53	078	RCL9	36 09
013	RCLA	36 11	046	P⇄S	16-51	079	STO4	35 04
014	LOG	16 32	047	ST+6	35-55 06	080	GSB1	23 01
015	ST+9	35-55 09	048	RTN	24	081	RTN	24
016	X²	53	049	*LBLC	21 13	082	*LBL1	21 01
017	ST+7	35-55 07	050	CF0	16 22 00	083	RCL8	36 08
018	LSTX	16-63	051	P⇄S	16-51	084	RCL4	36 04
019	RCLB	36 12	052	RCL9	36 09	085	RCLC	36 13
020	×	-35	053	STOD	35 14	086	×	-35
021	ST+0	35-55 00	054	RCL7	36 07	087	GSB9	23 09
022	RCLA	36 11	055	RCL6	36 06	088	STOA	35 11
023	1/X	52	056	STOC	35 13	089	RCL5	36 05
024	GSB7	23 07	057	X²	53	090	RCL4	36 04
025	RCLA	36 11	058	GSB9	23 09	091	X²	53
026	eˣ	33	059	STOE	35 15	092	GSB9	23 09
027	ST+4	35-55 04	060	GSB1	23 01	093	÷	-24
028	X²	53	061	RCL5	36 05	094	PRTX	-14
029	ST+5	35-55 05	062	STO4	35 04	095	STOB	35 12
030	LSTX	16-63	063	RCL0	36 00	096	RCL4	36 04
031	RCLB	36 12	064	STO8	35 08	097	×	-35
032	×	-35	065	P⇄S	16-51	098	CHS	-22
033	ST+8	35-55 08	066	RCL6	36 06	099	RCLC	36 13

Table A.20 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
100	+	-55	141	COS	42	182	RCL6	36 06
101	RCLD	36 14	142	X^2	53	183	-	-45
102	÷	-24	143	RCLI	36 46	184	RCLD	36 14
103	PRTX	-14	144	÷	-24	185	2	02
104	RCLB	36 12	145	RCLI	36 46	186	-	-45
105	RCLA	36 11	146	1	01	187	÷	-24
106	×	-35	147	-	-45	188	√X	54
107	STO6	35 06	148	×	-35	189	PRTX	-14
108	RCLE	36 15	149	×	-35	190	SPC	16-11
109	÷	-24	150	1	01	191	RTN	24
110	√X	54	151	+	-55	192	*LBL7	21 07
111	PRTX	-14	152	STO8	35 08	193	ST+1	35-55 01
112	ENT↑	-21	153	DSZI	16 25 46	194	X^2	53
113	X^2	53	154	DSZI	16 25 46	195	ST+2	35-55 02
114	CHS	-22	155	GTO2	22 02	196	LSTX	16-63
115	1	01	156	RCLB	36 12	197	RCLB	36 12
116	+	-55	157	×	-35	198	×	-35
117	√X	54	158	GTO6	22 06	199	ST+3	35-55 03
118	÷	-24	159	*LBL3	21 03	200	RTN	24
119	RAD	16-22	160	RCL8	36 08	201	*LBL8	21 08
120	TAN^{-1}	16 43	161	RCLB	36 12	202	RCL1	36 01
121	STOA	35 11	162	×	-35	203	STO4	35 04
122	SIN	41	163	RCLA	36 11	204	RCL2	36 02
123	STOB	35 12	164	COS	42	205	STO5	35 05
124	RCLD	36 14	165	×	-35	206	RCL3	36 03
125	4	04	166	RCLA	36 11	207	STO8	35 08
126	X>Y?	16-34	167	+	-55	208	GSB1	23 01
127	GTO4	22 04	168	GSBa	23 16 11	209	RTN	24
128	X=Y?	16-33	169	GTO6	22 06	210	*LBL9	21 09
129	GTO5	22 05	170	*LBL4	21 04	211	RCLD	36 14
130	-	-45	171	RCLA	36 11	212	÷	-24
131	STOI	35 46	172	GSBa	23 16 11	213	-	-45
132	1	01	173	GTO6	22 06	214	RTN	24
133	STO8	35 08	174	*LBL5	21 05	215	*LBLa	21 16 11
134	*LBL2	21 02	175	RCLB	36 12	216	2	02
135	RCLI	36 46	176	*LBL6	21 06	217	Pi	16-24
136	1	01	177	CHS	-22	218	÷	-24
137	X=Y?	16-33	178	1	01	219	×	-35
138	GTO3	22 03	179	+	-55	220	RTN	24
139	RCL8	36 08	180	PRTX	-14			
140	RCLA	36 11	181	RCLE	36 15			

A.22 DETAILS OF THE REGRESSION-2 PROGRAM: LINEAR CORRELATION OTHER THAN STRAIGHT LINES

Lines 1 to 12 take the input data and calculate the sums for the linear correlation with KEY $\Sigma+$. With the first data entry the registers are cleared.

Lines 13 to 48 transform the x variable into the five other functions that will be tested, and accumulate the sums, the sums of squares, and the sums of cross products for these functions. All are stored in specific storage areas.

Lines 49 to 81 All of the calculations of the regression coefficients, correlation coefficient, and probability value are done by the same routine. Program lines 49 to 81 shift the sums and sums of squares to the same storage areas that are used by the calculation routine.

Lines 82 to 111 calculate the regression coefficients by Eqs. (7.5) and (7.6), and the correlation coefficient by Eq. (7.8). These results are printed.

Lines 112 to 180 are the t-3 probability calculation used in the Regression-1 program with one modification. In this program, the probability calculation uses the value of n in place of the degrees of freedom, and an adjustment is made in the program at line 125. The probability value is printed at line 180.

Lines 181 to 191 calculate and print the $s(\hat{y})$ term for each correlation, and end the program.

Lines 192 to 200 are a routine for storing two sets of sums, which in the Hewlett-Packard program can be stored in the same numbered storage registers by previously shifting the registers with KEY P-S.

Lines 201 to 209 are a subroutine used in shifting the sums before the calculation of the equation constants. The subroutine is used because the shifting for two of the calculations is identical.

Lines 210 to 214 are a small subroutine that is used several times in the calculation to divide a quantity by n and subtract the result from a previous value.

Lines 215 to 220 are part of the probability calculation.

Note that the probability routine in this program cannot be used with entry of ν and r because of the modifications to the program.

If any of the functions are to be deleted or changed, the location of the first formation of the functions is shown at the top of the table. The function x^2 cannot be easily changed, as its calculation uses the same Σx^2 as is used for the straight-line calculation.

Table A.21 Regression-3 Program: Two-Variable and Second-Degree Correlations (HP-97)

Two-variable: $y = b_0 + b_1x_1$
$y = b_0 + b_1x_2$
$y = b_0 + b_1x_1 + b_2x_2$

Input: y_i, ENTER ↑; x_{1i}, ENTER ↑; x_{2i}, KEY A
KEY C

Second-degree: $y = b_0 + b_1x$
$y = b_0 + b_1x^2$
$y = b_0 + b_1x + b_2x^2$

Input: y_i, ENTER ↑; x_i, KEY B
KEY C

Output: b_1, b_0, r for first equation
b_1, b_0, r for second equation
b_1, b_2, b_0, R for third equation
t comparing R of third equation with better of other two

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	033	Σ+	56	065	STOB	35 12
002	F0?	16 23 00	034	R↓	-31	066	P⇄S	16-51
003	GTO1	22 01	035	STOA	35 11	067	RCL0	36 00
004	CLRG	16-53	036	RCLB	36 12	068	RCL4	36 04
005	P⇄S	16-51	037	X²	53	069	RCLB	36 12
006	CLRG	16-53	038	P⇄S	16-51	070	×	-35
007	SF0	16 21 00	039	Σ+	56	071	RCL9	36 09
008	*LBL1	21 01	040	P⇄S	16-51	072	÷	-24
009	STOC	35 13	041	RCLB	36 12	073	-	-45
010	R↓	-31	042	3	03	074	STO0	35 00
011	STOB	35 12	043	Y^x	31	075	X²	53
012	Σ+	56	044	ST+0	35-55 00	076	CHS	-22
013	R↓	-31	045	RTN	24	077	RCLE	36 15
014	STOA	35 11	046	*LBLC	21 13	078	RCL5	36 05
015	P⇄S	16-51	047	CF0	16 22 00	079	×	-35
016	RCLC	36 13	048	P⇄S	16-51	080	+	-55
017	Σ+	56	049	GSB1	23 01	081	STO1	35 01
018	RCLB	36 12	050	STOA	35 11	082	RCLD	36 14
019	RCLC	36 13	051	STOI	35 46	083	RCL5	36 05
020	×	-35	052	P⇄S	16-51	084	×	-35
021	P⇄S	16-51	053	GSB1	23 01	085	RCL8	36 08
022	ST+0	35-55 00	054	STOC	35 13	086	RCL0	36 00
023	R/S	51	055	RCLI	36 46	087	×	-35
024	*LBLB	21 12	056	X≤Y?	16-35	088	-	-45
025	F0?	16 23 00	057	R↓	-31	089	RCL1	36 01
026	GTO1	22 01	058	STOI	35 46	090	÷	-24
027	CLRG	16-53	059	P⇄S	16-51	091	PRTX	-14
028	P⇄S	16-51	060	RCL8	36 08	092	STO2	35 02
029	CLRG	16-53	061	STOD	35 14	093	RCL8	36 08
030	SF0	16 21 00	062	RCL5	36 05	094	RCLE	36 15
031	*LBL1	21 01	063	STOE	35 15	095	×	-35
032	STOB	35 12	064	RCL4	36 04	096	RCLD	36 14

Table A.21 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
097	RCL0	36 00	129	SPC	16-11	161	RCL9	36 09
098	×	-35	130	RCL6	36 06	162	÷	-24
099	-	-45	131	X²	53	163	-	-45
100	RCL1	36 01	132	RCLI	36 46	164	STO8	35 08
101	÷	-24	133	X²	53	165	RCL5	36 05
102	PRTX	-14	134	-	-45	166	RCL4	36 04
103	STO3	35 03	135	RCL6	36 06	167	X²	53
104	RCL4	36 04	136	X²	53	168	RCL9	36 09
105	×	-35	137	CHS	-22	169	÷	-24
106	RCL2	36 02	138	1	01	170	-	-45
107	RCLB	36 12	139	+	-55	171	STO5	35 05
108	×	-35	140	÷	-24	172	÷	-24
109	+	-55	141	RCL9	36 09	173	PRTX	-14
110	CHS	-22	142	3	03	174	RCL4	36 04
111	RCL6	36 06	143	-	-45	175	×	-35
112	+	-55	144	×	-35	176	CHS	-22
113	RCL9	36 09	145	√X	54	177	RCL6	36 06
114	÷	-24	146	PRTX	-14	178	+	-55
115	PRTX	-14	147	SPC	16-11	179	RCL9	36 09
116	RCL2	36 02	148	R/S	51	180	÷	-24
117	RCLD	36 14	149	*LBL1	21 01	181	PRTX	-14
118	×	-35	150	RCL7	36 07	182	RCL8	36 08
119	RCL3	36 03	151	RCL6	36 06	183	X²	53
120	RCL8	36 08	152	X²	53	184	RCL5	36 05
121	×	-35	153	RCL9	36 09	185	÷	-24
122	+	-55	154	÷	-24	186	RCL7	36 07
123	STOB	35 12	155	-	-45	187	÷	-24
124	RCL7	36 07	156	STO7	35 07	188	√X	54
125	÷	-24	157	RCL8	36 08	189	PRTX	-14
126	√X	54	158	RCL4	36 04	190	SPC	16-11
127	PRTX	-14	159	RCL6	36 06	191	RTN	24
128	STO6	35 06	160	×	-35			

A.23 DETAILS OF THE REGRESSION-3 PROGRAM: TWO-VARIABLE AND POLYNOMIAL CORRELATION

Lines 1 to 14 take the y, x_1, x_2 input data and calculate the sums of squares and cross-products of y and x_1 with the built-in function of KEY $\Sigma+$.

Lines 15 to 23 shift the storage areas and make the same summation calculations of y and x_2 with KEY $\Sigma+$. Summation x_1x_2 is made at lines 18 to 20 and stored in Register 0.

Lines 24 to 37 take input for the polynomial equation calculation and carry out the same operations for this solution as lines 1 to 14 for the two-variable equation.

Lines 38 to 45 do the same as lines 15 to 23, but for the polynomial equation. Two different routines are used for the two similar sets of calculations because the second routine has input of only one x variable, and it is necessary to square this value for the second variable.

Lines 46 to 53 start the calculation of the least-squares equations. The calculations of the two single-term equations are carried out in a subroutine at line 149. The calculation returns to the main program with the value of the correlation coefficient, which is stored.

Lines 54 to 58 compare the two correlation coefficients and store the larger in Register I for later comparison with the correlation coefficient of the two-variable equation.

Lines 59 to 81 start the calculation of the two-variable least-squares equation. The denominator of Eqs. (7.19) and (7.20)—the D value of Eq. (7.21)—is calculated and stored in Register 1.

Lines 82 to 91 calculate b_1 from Eq. (7.19).

Lines 92 to 102 calculate b_2 from Eq. (7.20).

Lines 103 to 115 calculate b_0 from Eq. (7.22).

Lines 116 to 127 calculate the correlation coefficient from Eq. (7.25).

Lines 128 to 148 recall the larger of the correlation coefficients of the single-variable equations and calculate t from Eq. (7.29). The program ends at line 148.

Lines 149 to 191 are the subroutine for the calculation of the single-variable equations. Lines 150 to 173 calculate b_1 from Eq. (7.5), and lines 174 to 181 calculate b_0 from Eq. (7.6). Lines 182 to 191 calculate the correlation coefficient from Eq. (7.8) and return the operation to the main program.

Table A.22 Regression-4 Program: Three-Variable Equation Correlation (Part 1) (HP-97)

Equations: $y = b_0 + b_1x$ for all variables
$y = b_0 + b_1x_i + b_2x_j$ for all pairs
$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$

Input: y_i, ENTER ↑; x_{1i}, ENTER ↑; x_{2i}; ENTER ↑; x_{3i}, KEY A

When all data are in, KEY C

Output: b_1, b_0, r for single-variable equations in order x_1, x_2, x_3
b_1, b_2, b_0, R for two-variable equations in order x_1x_2, x_1x_3, x_2x_3
b_1, b_2, b_3, b_0, R for three-variable equation

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	034	RCLA	36 11	067	RCL0	36 00
002	F0?	16 23 00	035	x	-35	068	GSB2	23 02
003	GTO1	22 01	036	ST+2	35-55 02	069	RCL2	36 02
004	CLRG	16-53	037	RCLD	36 14	070	X²	53
005	P⇄S	16-51	038	RCLB	36 12	071	RCL1	36 01
006	CLRG	16-53	039	x	-35	072	÷	-24
007	SF0	16 21 00	040	ST+3	35-55 03	073	RCL7	36 07
008	*LBL1	21 01	041	RTN	24	074	÷	-24
009	STOD	35 14	042	*LBLC	21 13	075	√X	54
010	R↓	-31	043	CF0	16 22 00	076	PRTX	-14
011	STOC	35 13	044	P⇄S	16-51	077	SPC	16-11
012	R↓	-31	045	GSB1	23 01	078	RCL3	36 03
013	STOB	35 12	046	RCL2	36 02	079	RCLD	36 14
014	Σ+	56	047	STOA	35 11	080	RCL0	36 00
015	R↓	-31	048	RCL3	36 03	081	x	-35
016	STOA	35 11	049	STOB	35 12	082	GSB3	23 03
017	RCLC	36 13	050	RCL4	36 04	083	STOC	35 13
018	P⇄S	16-51	051	STOD	35 14	084	RCLA	36 11
019	Σ+	56	052	P⇄S	16-51	085	RCL4	36 04
020	RCLB	36 12	053	GSB1	23 01	086	RCL0	36 00
021	RCLC	36 13	054	RCL2	36 02	087	x	-35
022	x	-35	055	RCL0	36 00	088	GSB3	23 03
023	ST+3	35-55 03	056	RCL6	36 06	089	STOA	35 11
024	RCLC	36 13	057	x	-35	090	RCLB	36 12
025	RCLD	36 14	058	GSB3	23 03	091	RCL4	36 04
026	x	-35	059	STO2	35 02	092	RCLD	36 14
027	ST+2	35-55 02	060	RCL1	36 01	093	x	-35
028	P⇄S	16-51	061	RCL0	36 00	094	GSB3	23 03
029	RCLD	36 14	062	X²	53	095	STOB	35 12
030	ST+0	35-55 00	063	GSB3	23 03	096	RCLA	36 11
031	X²	53	064	STO1	35 01	097	STO3	35 03
032	ST+1	35-55 01	065	÷	-24	098	RCL5	36 05
033	RCLD	36 14	066	PRTX	-14	099	RCL8	36 08

Table A.22 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
100	RCL4	36 04	142	CHS	-22	184	GSB2	23 02
101	P⇄S	16-51	143	RCL6	36 06	185	STOE	35 15
102	STO0	35 00	144	+	-55	186	RCL2	36 02
103	R↓	-31	145	RCL9	36 09	187	RCL5	36 05
104	STO2	35 02	146	÷	-24	188	×	-35
105	R↓	-31	147	PRTX	-14	189	RCL8	36 08
106	STO1	35 01	148	RTN	24	190	GSB2	23 02
107	RCLB	36 12	149	*LBL3	21 03	191	STOI	35 46
108	STO3	35 03	150	RCL9	36 09	192	RCL0	36 00
109	GTO4	22 04	151	÷	-24	193	×	-35
110	*LBL1	21 01	152	-	-45	194	RCLE	36 15
111	RCL8	36 08	153	RTN	24	195	RCL4	36 04
112	RCL4	36 04	154	*LBL4	21 04	196	×	-35
113	RCL6	36 06	155	GSB1	23 01	197	+	-55
114	×	-35	156	P⇄S	16-51	198	CHS	-22
115	GSB3	23 03	157	RCL0	36 00	199	RCL6	36 06
116	STO8	35 08	158	RCL1	36 01	200	+	-55
117	RCL5	36 05	159	RCL2	36 02	201	RCL9	36 09
118	RCL4	36 04	160	P⇄S	16-51	202	÷	-24
119	X^2	53	161	STO2	35 02	203	PRTX	-14
120	GSB3	23 03	162	R↓	-31	204	RCLE	36 15
121	STO5	35 05	163	STO1	35 01	205	RCL8	36 08
122	÷	-24	164	R↓	-31	206	×	-35
123	PRTX	-14	165	STO0	35 00	207	RCLI	36 46
124	RCL4	36 04	166	RCLC	36 13	208	RCL2	36 02
125	GSB2	23 02	167	STO3	35 03	209	×	-35
126	RCL8	36 08	168	GSB1	23 01	210	+	-55
127	X^2	53	169	P⇄S	16-51	211	RCL7	36 07
128	RCL7	36 07	170	GSB1	23 01	212	÷	-24
129	RCL6	36 06	171	RTN	24	213	√X	54
130	X^2	53	172	*LBL1	21 01	214	PRTX	-14
131	GSB3	23 03	173	RCL1	36 01	215	SPC	16-11
132	STO7	35 07	174	RCL5	36 05	216	RTN	24
133	÷	-24	175	×	-35	217	*LBL2	21 02
134	RCL5	36 05	176	RCL3	36 03	218	RCL3	36 03
135	÷	-24	177	X^2	53	219	×	-35
136	√X	54	178	-	-45	220	-	-45
137	PRTX	-14	179	STOD	35 14	221	RCLD	36 14
138	SPC	16-11	180	RCL8	36 08	222	÷	-24
139	RTN	24	181	RCL1	36 01	223	PRTX	-14
140	*LBL2	21 02	182	×	-35	224	RTN	24
141	×	-35	183	RCL2	36 02			

A.24 DETAILS OF THE REGRESSION-4 PROGRAM: THREE-VARIABLE CORRELATION (PART 1)

Lines 1 to 14 take the input data and calculate the sums of y with x_1 by means of KEY $\Sigma+$.

Lines 15 to 19 shift the storage areas and calculate the sums of y and x_2, also with KEY $\Sigma+$.

Lines 20 to 41 calculate and store the balance of the sums and sums of cross products.

Lines 42 to 45 start the calculation of the least-squares equations. The single-variable equations are calculated with a subroutine at line 110. Line 45 directs the program to this subroutine for the y-x_1 calculation.

Lines 46 to 53 shift the storage areas and redirect the program to the subroutine at line 110 for the y-x_2 calculation.

Lines 54 to 76 calculate the y-x_3 correlation, following Eqs. (7.5) to (7.8), without using the subroutine, inasmuch as no program space would be saved by shifting all the sums necessary for the calculation.

Lines 77 to 109 calculate the sums of squares of cross products required for the two-variable least-squares equations. These values are calculated and stored, and the program is directed to line 154 where the constants for the equations are calculated.

Lines 110 to 139 are the subroutine for calculating b_1 and the correlation coefficient for the single-variable equations, following Eqs. (7.5) and (7.8)

Lines 140 to 148 are a subroutine for calculating b_0 for the one-variable equation following Eq. (7.6).

Lines 149 to 153 are a small subroutine that is used several times in the calculation to divide a quantity by n and subtract the result from a previous value.

Lines 154 to 171 start the calculation of the three two-variable equations. The actual calculation is done by a subroutine starting at line 172. The same subroutine is used three times, and the sums required for the calculation are shifted to the proper storage areas by the program lines 154 to 171.

Lines 172 to 216 calculate the two-variable least-squares equation following Eqs. (7.19) to (7.25).

Lines 217 to 224 calculate the regression coefficients by Eqs. (7.19) and (7.20) for all the equations.

Table A.23 Regression-4 Program (Part 2) (HP-97)

At the completion of Part 1 of program Regression-4, only the output for the one- and two-variable equations has been made. Part 2 produces the output for the three-variable equation. The program is run with KEY A.

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	039	÷	-24	077	×	-35
002	RCL2	36 02	040	RCL2	36 02	078	CHS	-22
003	RCL1	36 01	041	-	-45	079	RCL0	36 00
004	÷	-24	042	STO0	35 00	080	+	-55
005	STO2	35 02	043	R↓	-31	081	PRTX	-14
006	RCLC	36 13	044	RCL1	36 01	082	STO0	35 00
007	RCL1	36 01	045	÷	-24	083	RCLE	36 15
008	÷	-24	046	RCLD	36 14	084	×	-35
009	STOD	35 14	047	-	-45	085	RCLD	36 14
010	RCLA	36 11	048	STOA	35 11	086	RCL1	36 01
011	RCL1	36 01	049	RCLB	36 12	087	×	-35
012	÷	-24	050	RCL1	36 01	088	+	-55
013	STOE	35 15	051	÷	-24	089	CHS	-22
014	RCL8	36 08	052	RCLE	36 15	090	RCL2	36 02
015	RCLA	36 11	053	-	-45	091	+	-55
016	STO1	35 01	054	STO1	35 01	092	PRTX	-14
017	÷	-24	055	RCL0	36 00	093	STO6	35 06
018	RCL2	36 02	056	RCL1	36 01	094	P⇄S	16-51
019	-	-45	057	÷	-24	095	RCL0	36 00
020	STOI	35 46	058	STO0	35 00	096	RCL4	36 04
021	RCLB	36 12	059	RCLA	36 11	097	RCL8	36 08
022	RCL1	36 01	060	RCL1	36 01	098	RCL2	36 02
023	÷	-24	061	÷	-24	099	P⇄S	16-51
024	RCLD	36 14	062	STOA	35 11	100	STO2	35 02
025	-	-45	063	RCLI	36 46	101	R↓	-31
026	STO6	35 06	064	RCL9	36 09	102	STOD	35 14
027	RCL5	36 05	065	÷	-24	103	R↓	-31
028	RCL1	36 01	066	RCL0	36 00	104	STOE	35 15
029	÷	-24	067	-	-45	105	R↓	-31
030	RCLE	36 15	068	RCL6	36 06	106	RCL6	36 06
031	-	-45	069	RCL9	36 09	107	×	-35
032	STO9	35 09	070	÷	-24	108	RCLE	36 15
033	P⇄S	16-51	071	RCLA	36 11	109	RCL1	36 01
034	RCL5	36 05	072	-	-45	110	×	-35
035	RCL8	36 08	073	÷	-24	111	+	-55
036	P⇄S	16-51	074	PRTX	-14	112	RCL4	36 04
037	RCLC	36 13	075	STO1	35 01	113	RCL0	36 00
038	STO1	35 01	076	RCLA	36 11	114	×	-35

Table A.23 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
115	+	-55	124	RCL1	36 01	133	×	-35
116	CHS	-22	125	RCLD	36 14	134	+	-55
117	P⇄S	16-51	126	×	-35	135	RCL7	36 07
118	RCL6	36 06	127	RCL0	36 00	136	÷	-24
119	+	-55	128	RCL8	36 08	137	√X	54
120	RCL9	36 09	129	×	-35	138	PRTX	-14
121	÷	-24	130	+	-55	139	SPC	16-11
122	P⇄S	16-51	131	RCL6	36 06	140	RTN	24
123	PRTX	-14	132	RCL2	36 02			

A.25 DETAILS OF THE REGRESSION-4 PROGRAM (PART 2)

This portion of the program is the solution to three simultaneous equations derived from Eqs. (7.33) with b_0 eliminated from the four equations shown.

Lines 1 to 54 divide each equation by the coefficient of the third term and subtract the quotients to eliminate b_3.

Lines 55 to 70 divide each term in the remaining two equations by the coefficient of b_2 and subtract to obtain one equation with only b_1 to be obtained.

Lines 71 to 74 calculate b_1.

Lines 75 to 81 substitute b_1 in the previous equation and obtain b_2.

Lines 82 to 92 substitute b_1 and b_2 in one of the previous equations to obtain b_3.

Lines 93 to 105 shift some of the values for the later calculation of the correlation coefficient.

Lines 106 to 123 calculate b_0 from Eq. (7.34).

Lines 124 to 140 calculate the correlation coefficient from Eqs. (7.24) and (7.25) expanded to include a third term.

Table A.24 Regression-5 Program: Correlation of Family of Curves (HP-97)

Equation: $y = b_0 + b_1x_1 + b_2(x_1x_2)$
And if Flag 1 is set: $y = b_0 + b_1x_1 + b_2(x_1 \cdot \log(x_2))$

Input: y_i, ENTER ↑; x_{1i}, ENTER ↑; x_{2i}, KEY A
KEY C

Output: b_1, b_2, b_0, R; and if Flag 1 is set: b_1, b_2, b_0, R of second equation

Step	Key	Code	Step	Key	Code	Step	Key	Code
001	*LBLA	21 11	036	RCLA	36 11	071	÷	-24
002	F0?	16 23 00	037	RCLE	36 15	072	-	-45
003	GTO1	22 01	038	×	-35	073	STOB	35 12
004	CLRG	16-53	039	ST+2	35-55 02	074	RCL8	36 08
005	P⇄S	16-51	040	RCLB	36 12	075	RCL6	36 06
006	CLRG	16-53	041	RCLE	36 15	076	RCL4	36 04
007	SF0	16 21 00	042	×	-35	077	×	-35
008	*LBL1	21 01	043	ST+3	35-55 03	078	RCLD	36 14
009	STOC	35 13	044	RTN	24	079	÷	-24
010	R↓	-31	045	*LBL2	21 02	080	-	-45
011	STOB	35 12	046	RCLC	36 13	081	STOC	35 13
012	Σ+	56	047	LOG	16 32	082	P⇄S	16-51
013	R↓	-31	048	STOD	35 14	083	GSB3	23 03
014	STOA	35 11	049	RCLB	36 12	084	F1?	16 23 01
015	RCLC	36 13	050	×	-35	085	GTO4	22 04
016	RCLI	36 46	051	STOE	35 15	086	R/S	51
017	+	-55	052	P⇄S	16-51	087	*LBL3	21 03
018	STOI	35 46	053	RTN	24	088	RCL3	36 03
019	RCLC	36 13	054	*LBLC	21 13	089	RCLI	36 46
020	RCLB	36 12	055	CF0	16 22 00	090	RCL0	36 00
021	×	-35	056	P⇄S	16-51	091	×	-35
022	STOE	35 15	057	RCL7	36 07	092	RCLD	36 14
023	GSB1	23 01	058	RCL6	36 06	093	÷	-24
024	F1?	16 23 01	059	STOE	35 15	094	-	-45
025	GSB2	23 02	060	X²	53	095	STO3	35 03
026	F1?	16 23 01	061	RCL9	36 09	096	RCL2	36 02
027	GSB1	23 01	062	STOD	35 14	097	RCL0	36 00
028	F1?	16 23 01	063	÷	-24	098	RCLE	36 15
029	P⇄S	16-51	064	-	-45	099	×	-35
030	R/S	51	065	STOA	35 11	100	RCLD	36 14
031	*LBL1	21 01	066	RCL5	36 05	101	÷	-24
032	RCLE	36 15	067	RCL4	36 04	102	-	-45
033	ST+0	35-55 00	068	STOI	35 46	103	STO2	35 02
034	X²	53	069	X²	53	104	RCL1	36 01
035	ST+1	35-55 01	070	RCLD	36 14	105	RCL0	36 00

Table A.24 (*Cont.*)

Step	Key	Code	Step	Key	Code	Step	Key	Code
106	X^2	53	128	STO6	35 06	150	÷	-24
107	RCLD	36 14	129	RCL2	36 02	151	PRTX	-14
108	÷	-24	130	RCLB	36 12	152	RCL6	36 06
109	-	-45	131	×	-35	153	RCLC	36 13
110	STO1	35 01	132	RCLC	36 13	154	×	-35
111	RCLB	36 12	133	RCL3	36 03	155	RCL1	36 01
112	RCL1	36 01	134	×	-35	156	RCL2	36 02
113	×	-35	135	-	-45	157	×	-35
114	RCL3	36 03	136	RCL8	36 08	158	+	-55
115	X^2	53	137	÷	-24	159	RCLA	36 11
116	-	-45	138	PRTX	-14	160	÷	-24
117	STO8	35 08	139	STO1	35 01	161	√X	54
118	RCLC	36 13	140	RCL0	36 00	162	PRTX	-14
119	RCL1	36 01	141	×	-35	163	SPC	16-11
120	×	-35	142	CHS	-22	164	RTN	24
121	RCL2	36 02	143	RCLE	36 15	165	*LBL4	21 04
122	RCL3	36 03	144	+	-55	166	CF1	16 22 01
123	×	-35	145	RCL6	36 06	167	P⇄S	16-51
124	-	-45	146	RCLI	36 46	168	GSB3	23 03
125	RCL8	36 08	147	×	-35	169	P⇄S	16-51
126	÷	-24	148	-	-45	170	R/S	51
127	PRTX	-14	149	RCLD	36 14			

A.26 DETAILS OF THE REGRESSION-5 PROGRAM: CORRELATION OF FAMILY OF CURVES

Lines 1 to 12 take the input data and calculate the sums of y-x_1 with KEY $\Sigma+$. The value of x_1 enters the program at line 11. If some function of x_1 is to be used in place of the actual value, the function must be added between lines 10 and 11.

Lines 13 to 22 calculate Σx_2 and the product $x_1 \cdot x_2$.

Line 23 directs the program to a subroutine starting at line 31 where Σx_1x_2, $\Sigma(x_1x_2)^2$, $\Sigma y(x_1x_2)$, and $\Sigma x_1(x_1x_2)$ are calculated.

Lines 24 to 30 check whether Flag 1 is set. If so, the calculations mentioned just above are repeated with $\log(x_2)$ in place of x_2, and the storage areas are shifted so the sums of the second set are distinguished from the first.

Lines 31 to 44 are the subroutine for calculating the sums mentioned at line 23.

Lines 45 to 53 calculate the log function and $x_1(\log(x_2))$ if it is called for by Flag 1.

Lines 54 to 82 start the calculation of the least-squares solution which follows Eqs. (7.19) to (7.25) and is similar to the Regression-3 program. The terms that are the same for both the direct value of x_1 and $\log(x_2)$ are calculated in lines 54 to 82. The balance of the calculation is made in the subroutine starting at line 87.

Lines 83 to 86 check whether Flag 1 is set, indicating the $\log(x_2)$ calculation is called for. If not, the program ends at line 86. If Flag 1 is set, the program goes to line 165 to shift the storage areas and then goes to the subroutine at line 87 to repeat the calculation.

Lines 87 to 164 follow Eqs. (7.19) to (7.25) for the calculation of the two-variable equation, where $(x_1 x_2)$, or $(x_1 \cdot \log(x_2))$ is the second variable. This portion of the program is the same as the Regression-3 program.

Lines 165 to 170 shift the storage areas for the log function and call on the subroutine for the second calculation.

B
TEXAS INSTRUMENTS CALCULATOR PROGRAMS

Table B.1 Statistical Parameter Program (TI-59)

For individual values: x_i, KEY A, Run: KEY C
For weighted values: f_i, KEY $x \rightleftharpoons t$; x_i, KEY B Run: KEY C

Output: Arithmetic mean, midrange, geometric mean, harmonic mean, log mean, estimated standard deviation, data standard deviation, estimated variance, data variance, range, minimum value, maximum value, Σx (or Σfx), Σx^2 (or Σfx^2), number of groups if applicable, total number of data points

Step	Code	Key
000	91	R/S
001	76	LBL
002	12	B
003	86	STF
004	00	00
005	76	LBL
006	11	A
007	87	IFF
008	02	02
009	22	INV
010	47	CMS
011	42	STO
012	10	10
013	42	STO
014	11	11
015	76	LBL
016	22	INV
017	86	STF
018	02	02
019	42	STO
020	09	09
021	32	X:T
022	42	STO
023	07	07
024	00	0
025	22	INV
026	67	EQ
027	23	LNX
028	86	STF
029	01	01
030	76	LBL
031	23	LNX
032	43	RCL
033	10	10
034	32	X:T
035	43	RCL
036	09	09

Step	Code	Key
037	77	GE
038	24	CE
039	42	STO
040	10	10
041	76	LBL
042	24	CE
043	32	X:T
044	43	RCL
045	11	11
046	77	GE
047	25	CLR
048	43	RCL
049	09	09
050	42	STO
051	11	11
052	76	LBL
053	25	CLR
054	87	IFF
055	00	00
056	32	X:T
057	43	RCL
058	09	09
059	78	Σ+
060	87	IFF
061	01	01
062	33	X²
063	43	RCL
064	09	09
065	35	1/X
066	44	SUM
067	08	08
068	00	0
069	32	X:T
070	43	RCL
071	10	10
072	22	INV
073	77	GE

Step	Code	Key
074	33	X²
075	43	RCL
076	09	09
077	23	LNX
078	44	SUM
079	23	23
080	61	GTO
081	33	X²
082	76	LBL
083	32	X:T
084	53	(
085	43	RCL
086	07	07
087	44	SUM
088	13	13
089	65	×
090	43	RCL
091	09	09
092	54	)
093	42	STO
094	14	14
095	44	SUM
096	15	15
097	53	(
098	43	RCL
099	14	14
100	65	×
101	43	RCL
102	09	09
103	54	)
104	44	SUM
105	16	16
106	87	IFF
107	01	01
108	33	X²
109	53	(
110	43	RCL

Table B.1 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
111	09	09	155	02	02	199	10	10
112	35	1/X	156	43	RCL	200	54	)
113	65	×	157	15	15	201	23	LNX
114	43	RCL	158	42	STO	202	54	)
115	07	07	159	01	01	203	54	)
116	54	)	160	43	RCL	204	42	STO
117	44	SUM	161	17	17	205	20	20
118	08	08	162	42	STO	206	76	LBL
119	00	0	163	14	14	207	35	1/X
120	32	X:T	164	76	LBL	208	53	(
121	43	RCL	165	34	ΓX	209	53	(
122	10	10	166	53	(	210	43	RCL
123	22	INV	167	43	RCL	211	11	11
124	77	GE	168	11	11	212	85	+
125	33	X²	169	75	-	213	43	RCL
126	53	(	170	43	RCL	214	10	10
127	43	RCL	171	10	10	215	54	)
128	09	09	172	54	)	216	55	÷
129	23	LNX	173	42	STO	217	02	2
130	65	×	174	09	09	218	54	)
131	43	RCL	175	00	0	219	42	STO
132	07	07	176	32	X:T	220	17	17
133	54	)	177	43	RCL	221	43	RCL
134	44	SUM	178	10	10	222	03	03
135	23	23	179	67	EQ	223	42	STO
136	76	LBL	180	35	1/X	224	15	15
137	33	X²	181	22	INV	225	43	RCL
138	01	1	182	77	GE	226	01	01
139	44	SUM	183	35	1/X	227	42	STO
140	17	17	184	53	(	228	12	12
141	91	R/S	185	53	(	229	43	RCL
142	76	LBL	186	43	RCL	230	02	02
143	13	C	187	11	11	231	42	STO
144	22	INV	188	75	-	232	13	13
145	87	IFF	189	43	RCL	233	87	IFF
146	00	00	190	10	10	234	01	01
147	34	ΓX	191	54	)	235	42	STO
148	43	RCL	192	55	÷	236	53	(
149	13	13	193	53	(	237	43	RCL
150	42	STO	194	53	(	238	03	03
151	03	03	195	43	RCL	239	55	÷
152	43	RCL	196	11	11	240	43	RCL
153	16	16	197	55	÷	241	08	08
154	42	STO	198	43	RCL	242	54	)

Table B.1 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	42	STO	273	42	STO	303	22	22
244	21	21	274	07	07	304	42	STO
245	00	0	275	69	OP	305	02	02
246	32	X:T	276	11	11	306	43	RCL
247	43	RCL	277	42	STO	307	00	00
248	10	10	278	08	08	308	99	PRT
249	22	INV	279	34	√X	309	01	1
250	77	GE	280	42	STO	310	42	STO
251	42	STO	281	06	06	311	00	00
252	53	(	282	43	RCL	312	76	LBL
253	43	RCL	283	17	17	313	65	×
254	23	23	284	42	STO	314	73	RC*
255	55	÷	285	01	01	315	00	00
256	43	RCL	286	43	RCL	316	99	PRT
257	03	03	287	18	18	317	01	1
258	54	)	288	42	STO	318	05	5
259	22	INV	289	05	05	319	32	X:T
260	23	LNX	290	43	RCL	320	43	RCL
261	42	STO	291	20	20	321	00	00
262	22	22	292	42	STO	322	67	EQ
263	76	LBL	293	04	04	323	75	-
264	42	STO	294	43	RCL	324	01	1
265	79	$\bar{X}$	295	19	19	325	44	SUM
266	42	STO	296	42	STO	326	00	00
267	19	19	297	00	00	327	61	GTO
268	22	INV	298	43	RCL	328	65	×
269	79	$\bar{X}$	299	21	21	329	76	LBL
270	42	STO	300	42	STO	330	75	-
271	18	18	301	03	03	331	81	RST
272	33	X^2	302	43	RCL	332	91	R/S

B.1 DETAILS OF THE STATISTICAL PARAMETER PROGRAM

Lines 0 to 14 initiate the program. Flag 0 is set at line 3 if weighted data are to be calculated as indicated by starting with KEY B. Flag 2 is tested at line 7 to establish whether the data entry is the first. If so, all registers are cleared. Lines 11 to 14 store the first data entry in Registers 10 and 11.

Lines 15 to 29 set Flag 2 at line 17 after the first data entry and store the data. The entry is checked to determine if it is zero. If so, Flag 1 is set.

Lines 30 to 81 check the input data against the previous minimum and maximum values stored in Registers 10 and 11. If the new data is less than the first or greater than the second a switch is made. The balance of this portion of the program calculates the sums for individual data entries. A check is made for zero or negative data input. $\Sigma\, 1/x$ and Σ log (x) are bypassed if the calculation is impossible.

Lines 82 to 135 calculate the sums for grouped data. Again, a check is made for zero and negative values, and the impossible calculations are bypassed.

Lines 136 to 141 count the number of groups of data.

Lines 142 to 163 start the calculation of the various parameters. If the input data are individual values, lines 148 to 163 are bypassed. If the data are grouped data, the sums for the calculation of the mean and standard deviation are shifted to the registers used by KEY $\Sigma+$ so the built-in functions of KEY $\bar{x}$, and KEY s can be used.

Lines 164 to 262 carry out most of the mean calculations. The range is calculated at line 174, the log mean at line 204, the midrange at line 220, the harmonic mean at line 244, and the geometric mean at line 262. A check is made at line 175, and if the minimum value is equal or less than zero, the log mean calculation is bypassed. If Flag 1 is set (line 233), indicating a zero data input, the harmonic mean calculation is bypassed. If the minimum value is equal to or less than zero, the geometric mean calculation is bypassed (line 248).

Lines 263 to 281 use the built-in functions to calculate the arithmetic mean, the standard deviation, and the variance.

Lines 282 to 305 shift the calculated results to the storage areas shown in Table 2.4. The shifting is required because some of these areas are those used by the built-in functions and may not be used until the mean and variance have been calculated.

Lines 306 to 328 are the print routine.

Lines 329 to 331 reset all flags so the program can be run again.

Table B.2 Normal Probability Program (TI-59)

Input	Output
Mean, KEY A; Std. Dev., KEY A; x_1, KEY A, x_2, KEY A	Probability ($x_1 \leqq x \leqq x_2$)
Mean, KEY C; Std. Dev., KEY C; x, KEY C	Probability ($\geqq x$)
Mean, KEY E; Std. Dev., KEY E; x, KEY E	Probability ($\leqq x$)
Mean, KEY C'; Std. Dev., KEY C'; P, KEY C'	x such that $\Pr(\geqq x) = P$
Mean, KEY E'; Std. Dev., KEY E'; P, KEY E'	x such that $\Pr(\leqq x) = P$

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	037	05	05	074	77	GE
001	76	LBL	038	42	STO	075	33	X²
002	11	A	039	03	03	076	87	IFF
003	86	STF	040	86	STF	077	02	02
004	01	01	041	03	03	078	34	√X
005	61	GTO	042	22	INV	079	76	LBL
006	13	C	043	87	IFF	080	35	1/X
007	76	LBL	044	01	01	081	53	(
008	15	E	045	25	CLR	082	01	1
009	86	STF	046	91	R/S	083	75	-
010	02	02	047	76	LBL	084	43	RCL
011	76	LBL	048	24	CE	085	00	00
012	13	C	049	42	STO	086	54	)
013	87	IFF	050	04	04	087	42	STO
014	05	05	051	76	LBL	088	00	00
015	22	INV	052	25	CLR	089	61	GTO
016	42	STO	053	43	RCL	090	34	√X
017	01	01	054	01	01	091	76	LBL
018	86	STF	055	94	+/-	092	33	X²
019	05	05	056	85	+	093	87	IFF
020	91	R/S	057	43	RCL	094	02	02
021	76	LBL	058	05	05	095	35	1/X
022	22	INV	059	95	=	096	76	LBL
023	87	IFF	060	42	STO	097	34	√X
024	04	04	061	09	09	098	43	RCL
025	23	LNX	062	55	÷	099	00	00
026	42	STO	063	43	RCL	100	87	IFF
027	02	02	064	02	02	101	01	01
028	86	STF	065	95	=	102	42	STO
029	04	04	066	42	STO	103	61	GTO
030	91	R/S	067	07	07	104	65	×
031	76	LBL	068	71	SBR	105	76	LBL
032	23	LNX	069	32	X↕T	106	42	STO
033	87	IFF	070	00	0	107	87	IFF
034	03	03	071	32	X↕T	108	00	00
035	24	CE	072	43	RCL	109	44	SUM
036	42	STO	073	09	09	110	86	STF

Table B.2 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
111	00	00	155	65	×	199	93	.
112	42	STO	156	89	π	200	09	9
113	06	06	157	54	)	201	03	3
114	43	RCL	158	34	√X	202	07	7
115	04	04	159	35	1/X	203	03	3
116	42	STO	160	54	)	204	54	)
117	05	05	161	65	×	205	54	)
118	43	RCL	162	53	(	206	54	)
119	01	01	163	43	RCL	207	54	)
120	61	GTO	164	07	07	208	42	STO
121	25	CLR	165	33	X^2	209	00	00
122	76	LBL	166	55	÷	210	92	RTN
123	44	SUM	167	02	2	211	76	LBL
124	22	INV	168	54	)	212	10	E'
125	44	SUM	169	94	+/-	213	86	STF
126	06	06	170	22	INV	214	02	02
127	43	RCL	171	23	LNX	215	76	LBL
128	06	06	172	54	)	216	18	C'
129	50	|X|	173	65	×	217	87	IFF
130	61	GTO	174	53	(	218	05	05
131	65	×	175	43	RCL	219	45	Y^X
132	76	LBL	176	10	10	220	42	STO
133	32	X⇌T	177	65	×	221	01	01
134	43	RCL	178	53	(	222	86	STF
135	07	07	179	93	.	223	05	05
136	53	(	180	04	4	224	91	R/S
137	50	|X|	181	03	3	225	76	LBL
138	65	×	182	06	6	226	45	Y^X
139	93	.	183	02	2	227	87	IFF
140	03	3	184	85	+	228	04	04
141	03	3	185	43	RCL	229	52	EE
142	02	2	186	10	10	230	42	STO
143	07	7	187	65	×	231	02	02
144	85	+	188	53	(	232	86	STF
145	01	1	189	93	.	233	04	04
146	54	)	190	01	1	234	91	R/S
147	35	1/X	191	02	2	235	76	LBL
148	42	STO	192	00	0	236	52	EE
149	10	10	193	02	2	237	42	STO
150	53	(	194	94	+/-	238	05	05
151	53	(	195	85	+	239	42	STO
152	53	(	196	43	RCL	240	03	03
153	53	(	197	10	10	241	32	X⇌T
154	02	2	198	65	×	242	93	.

Table B.2 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	05	5	287	08	8	331	09	9
244	77	GE	288	05	5	332	02	2
245	53	(	289	03	3	333	06	6
246	53	(	290	85	+	334	09	9
247	01	1	291	53	(	335	85	+
248	75	-	292	43	RCL	336	43	RCL
249	43	RCL	293	07	07	337	07	07
250	05	05	294	65	×	338	65	×
251	54	)	295	93	.	339	93	.
252	42	STO	296	00	0	340	00	0
253	05	05	297	01	1	341	00	0
254	86	STF	298	00	0	342	01	1
255	00	00	299	03	3	343	03	3
256	76	LBL	300	02	2	344	00	0
257	53	(	301	08	8	345	08	8
258	43	RCL	302	54	)	346	54	)
259	05	05	303	54	)	347	54	)
260	33	X^2	304	54	)	348	54	)
261	35	1/X	305	54	)	349	54	)
262	23	LNX	306	55	÷	350	94	+/-
263	34	√X	307	53	(	351	85	+
264	42	STO	308	01	1	352	43	RCL
265	07	07	309	85	+	353	07	07
266	53	(	310	53	(	354	54	)
267	53	(	311	43	RCL	355	42	STO
268	53	(	312	07	07	356	12	12
269	02	2	313	65	×	357	87	IFF
270	93	.	314	53	(	358	02	02
271	05	5	315	01	1	359	54	)
272	01	1	316	93	.	360	87	IFF
273	05	5	317	04	4	361	00	00
274	05	5	318	03	3	362	55	÷
275	01	1	319	02	2	363	76	LBL
276	07	7	320	07	7	364	61	GTO
277	85	+	321	08	8	365	53	(
278	53	(	322	08	8	366	53	(
279	43	RCL	323	85	+	367	43	RCL
280	07	07	324	43	RCL	368	12	12
281	65	×	325	07	07	369	65	×
282	53	(	326	65	×	370	43	RCL
283	93	.	327	53	(	371	02	02
284	08	8	328	93	.	372	54	)
285	00	0	329	01	1	373	85	+
286	02	2	330	08	8	374	43	RCL

Table B.2 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
375	01	01	385	55	÷	395	85	+
376	54	)	386	53	(	396	43	RCL
377	61	GTO	387	53	(	397	01	01
378	65	×	388	43	RCL	398	54	)
379	76	LBL	389	12	12	399	76	LBL
380	54	)	390	65	×	400	65	×
381	87	IFF	391	43	RCL	401	99	PRT
382	00	00	392	02	02	402	81	RST
383	61	GTO	393	54	)	403	91	R/S
384	76	LBL	394	94	+/-			

B.2 DETAILS OF NORMAL PROBABILITY PROGRAM

The program is in two parts: lines 1 to 210 calculate a probability for a given variable, and lines 211 to 398 calculate the variable value with an input of probability. Both portions use the same print routine at lines 399 to 402.

Lines 0 to 10 establish which type of probability is to be calculated: $\Pr(\geqq x)$, $\Pr(\leqq x)$, or $\Pr(x_1 \leqq x \leqq x_2)$. If the last, Flag 1 is set; if the second, Flag 2 is set. If the first probability is wanted, neither flag is set.

Lines 11 to 50 store the input data.

Lines 51 to 67 standardize the input variable using Eq. (3.7) and store in Register 09 the information whether x is greater than or less than the mean.

Lines 68 and 69 direct the program to the probability calculation routine starting at line 132. The resulting probability is stored in Register 00.

Lines 70 to 97 check whether Flag 2 is set for a probability $\leqq x$, and whether x is greater than or less than the mean. The output is either α, the calculated probability, or $1 - \alpha$, depending on the problem. The following table shows the possibilities. Lines 70 to 97 select the proper one.

Mean	$\Pr(\leqq x)$ Flag 2 set	$\Pr(\geqq x)$
$m > x$	α	$1 - \alpha$
$m < x$	$1 - \alpha$	α

Lines 98 to 104 check whether two probabilities are to be calculated (Flag 1 being set). If so, the program goes to line 105 to arrange the values for the second calculation. If not, the program goes to the print routine at the end.

Lines 105 to 121 store the first probability value in Register 06, recall the second value of the variable, and locate it for the probability calculation. The program then returns to line 51 for the second calculation.

Lines 122 to 131 calculate the difference between the two probabilities, and direct the program to the print routine with the absolute value.

Lines 132 to 210 carry out the probability calculation following Eqs. (3.14), (3.11), and (3.12). The result is stored in Register 00.

Lines 211 to 240 start the program for probability input. If a "less than" probability is to be matched, Flag 2 is set at line 213. The balance of the lines store the input data.

Lines 241 to 255 establish whether the input probability was greater than 0.5. If so, the value is changed to 1 minus the input value and Flag 0 is set.

Lines 256 to 356 calculate the standard deviate value associated with the probability input. Equations (3.17) and (3.16) are used.

Lines 357 and 362 check whether Flag 2 is set for a probability $\leqq x$ and whether Flag 0 is set for an input probability greater than 0.5. The program then makes the selection, following Table 3.3, for the proper solution.

Lines 363 to 398 carry out the selection mentioned just above.

Lines 339 to 402 are the print routine for all types of operation. These lines also clear all flags so the program may be run again without interference from previous operations.

At the completion of the program, the input values are in Registers 01, 02, 03, and 04 in the order of input.

Table B.3 Binomial Probability Program (TI-59)

n, KEY A; *p*, KEY A; *r*, KEY A gives Pr(*r*)
n, KEY B; *p*, KEY B; *r*, KEY B gives Pr(>*r*)
n, KEY C; *p*, KEY C; *r*, KEY C gives Pr($\geqq r$)
n, KEY D; *p*, KEY D; *r*, KEY D gives Pr(<*r*)
n, KEY E; *p*, KEY E; *r*, KEY E gives Pr($\leqq r$)

n, KEY C'; *p*, KEY C'; *P*, KEY C' gives *r* such that $P = \Pr(\geqq r)$

n, KEY E'; *p*, KEY E'; *P*, KEY E' gives *r* such that $P = \Pr(\leqq r)$

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	038	91	R/S	076	05	05
001	76	LBL	039	76	LBL	077	91	R/S
002	11	A	040	15	E	078	76	LBL
003	71	SBR	041	71	SBR	079	23	LNX
004	22	INV	042	22	INV	080	87	IFF
005	43	RCL	043	43	RCL	081	04	04
006	04	04	044	08	08	082	24	CE
007	99	PRT	045	99	PRT	083	42	STO
008	81	RST	046	81	RST	084	02	02
009	91	R/S	047	91	R/S	085	42	STO
010	76	LBL	048	76	LBL	086	10	10
011	12	B	049	18	C'	087	32	X:T
012	86	STF	050	86	STF	088	00	0
013	02	02	051	00	00	089	77	GE
014	71	SBR	052	76	LBL	090	95	=
015	22	INV	053	10	E'	091	01	1
016	43	RCL	054	86	STF	092	22	INV
017	05	05	055	01	01	093	77	GE
018	99	PRT	056	76	LBL	094	95	=
019	81	RST	057	22	INV	095	86	STF
020	91	R/S	058	87	IFF	096	04	04
021	76	LBL	059	05	05	097	91	R/S
022	13	C	060	23	LNX	098	76	LBL
023	71	SBR	061	47	CMS	099	24	CE
024	22	INV	062	42	STO	100	42	STO
025	43	RCL	063	01	01	101	03	03
026	06	06	064	29	CP	102	42	STO
027	99	PRT	065	22	INV	103	13	13
028	81	RST	066	59	INT	104	42	STO
029	91	R/S	067	22	INV	105	14	14
030	76	LBL	068	67	EQ	106	32	X:T
031	14	D	069	95	=	107	43	RCL
032	71	SBR	070	43	RCL	108	01	01
033	22	INV	071	01	01	109	22	INV
034	43	RCL	072	22	INV	110	77	GE
035	07	07	073	77	GE	111	95	=
036	99	PRT	074	95	=	112	87	IFF
037	81	RST	075	86	STF	113	01	01

Table B.3 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
114	25	CLR	157	53	(	200	75	-
115	43	RCL	158	53	(	201	43	RCL
116	03	03	159	01	1	202	10	10
117	22	INV	160	75	-	203	54	)
118	59	INT	161	43	RCL	204	54	)
119	32	X:T	162	10	10	205	65	×
120	00	0	163	54	)	206	53	(
121	22	INV	164	45	Y^X	207	53	(
122	67	EQ	165	43	RCL	208	43	RCL
123	95	=	166	01	01	209	01	01
124	61	GTO	167	54	)	210	75	-
125	32	X:T	168	22	INV	211	43	RCL
126	76	LBL	169	77	GE	212	09	09
127	25	CLR	170	55	÷	213	85	+
128	43	RCL	171	42	STO	214	01	1
129	03	03	172	04	04	215	54	)
130	32	X:T	173	44	SUM	216	55	÷
131	00	0	174	08	08	217	43	RCL
132	77	GE	175	43	RCL	218	09	09
133	95	=	176	13	13	219	54	)
134	01	1	177	32	X:T	220	54	)
135	32	X:T	178	43	RCL	221	54	)
136	77	GE	179	09	09	222	42	STO
137	95	=	180	67	EQ	223	04	04
138	43	RCL	181	33	X^2	224	43	RCL
139	01	01	182	87	IFF	225	08	08
140	42	STO	183	01	01	226	42	STO
141	13	13	184	35	1/X	227	07	07
142	22	INV	185	76	LBL	228	43	RCL
143	87	IFF	186	34	ΓX	229	04	04
144	00	00	187	69	OP	230	44	SUM
145	32	X:T	188	29	29	231	08	08
146	53	(	189	53	(	232	87	IFF
147	01	1	190	43	RCL	233	01	01
148	75	-	191	04	04	234	35	1/X
149	43	RCL	192	65	×	235	43	RCL
150	14	14	193	53	(	236	13	13
151	54	)	194	53	(	237	32	X:T
152	42	STO	195	43	RCL	238	43	RCL
153	14	14	196	10	10	239	09	09
154	76	LBL	197	55	÷	240	67	EQ
155	32	X:T	198	53	(	241	33	X^2
156	29	CP	199	01	1	242	61	GTO

Table B.3 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	34	√X	286	65	×	329	07	07
244	76	LBL	287	87	IFF	330	54	)
245	33	X²	288	00	00	331	99	PRT
246	53	(	289	42	STO	332	43	RCL
247	43	RCL	290	43	RCL	333	09	09
248	08	08	291	08	08	334	99	PRT
249	75	-	292	99	PRT	335	98	ADV
250	43	RCL	293	43	RCL	336	81	RST
251	04	04	294	09	09	337	91	R/S
252	54	)	295	99	PRT	338	76	LBL
253	42	STO	296	98	ADV	339	95	=
254	07	07	297	43	RCL	340	25	CLR
255	53	(	298	07	07	341	53	(
256	53	(	299	99	PRT	342	01	1
257	01	1	300	53	(	343	55	÷
258	75	-	301	43	RCL	344	00	0
259	43	RCL	302	09	09	345	54	)
260	08	08	303	75	-	346	91	R/S
261	54	)	304	01	1	347	76	LBL
262	42	STO	305	54	)	348	55	÷
263	05	05	306	99	PRT	349	25	CLR
264	85	+	307	81	RST	350	53	(
265	43	RCL	308	76	LBL	351	01	1
266	04	04	309	42	STO	352	75	-
267	54	)	310	53	(	353	43	RCL
268	42	STO	311	01	1	354	10	10
269	06	06	312	75	-	355	54	)
270	87	IFF	313	43	RCL	356	42	STO
271	06	06	314	08	08	357	10	10
272	75	-	315	54	)	358	53	(
273	92	RTN	316	99	PRT	359	43	RCL
274	76	LBL	317	53	(	360	01	01
275	35	1/X	318	43	RCL	361	75	-
276	43	RCL	319	09	09	362	43	RCL
277	14	14	320	85	+	363	13	13
278	32	X⇌T	321	01	1	364	54	)
279	43	RCL	322	54	)	365	42	STO
280	08	08	323	99	PRT	366	13	13
281	77	GE	324	98	ADV	367	86	STF
282	65	×	325	53	(	368	06	06
283	61	GTO	326	01	1	369	61	GTO
284	34	√X	327	75	-	370	32	X⇌T
285	76	LBL	328	43	RCL	371	76	LBL

Table B.3 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
372	75	-	380	32	X⇌T	388	42	STO
373	43	RCL	381	42	STO	389	07	07
374	06	06	382	08	08	390	32	X⇌T
375	32	X⇌T	383	43	RCL	391	42	STO
376	43	RCL	384	07	07	392	05	05
377	08	08	385	32	X⇌T	393	92	RTN
378	42	STO	386	43	RCL			
379	06	06	387	05	05			

B.3 DETAILS OF THE BINOMIAL PROBABILITY PROGRAM

Lines 0 to 47 establish which type of solution is requested with variable-value input. The calculation is the same in all cases and is carried out by a subroutine so that after the calculation is completed the program returns to the starting location. Each section of lines 1 to 47 recalls a different storage register to print the specified output. At line 12 Flag 2 is set if Pr(r) is requested. Later in the program, a check is made if Flag 2 is set to ensure that r is not equal to or greater than n.

Lines 48 to 55 start the program if there is probability input. Flags 0 and 1 are set to indicate the type of probability to be matched.

Lines 56 to 97 store and check the input data. A check is made at lines 64 to 69 to ensure that n is an integer. A check is made at lines 88 to 94 to ensure that the p value is greater than zero and less than one.

Lines 98 to 114 store the last input value, either r or P. A check is made to ensure that this value is not greater than n. At line 113 a check is made on Flag 1. If it is set, it indicates a P value input and the program goes on to line 126.

Lines 115 to 125 check the input r value to ensure it is an integer and greater than zero. If any of the checks are negative, the program goes to a routine at line 338 which calls for division by zero and causes an "error" signal. At line 125 the program is directed to the main calculation routine.

Lines 126 to 153 check the input P value to ensure it is greater than zero and less than one. The probability calculation is a summation of probabilities until the value of r is reached, following Eq. (3.20). When an input probability is to be matched, the value of n is substituted for the terminating value of the summation so the input probability will be reached. This substitution is made at lines 138 to 141.

The calculation finds the probability up to some value of r. If a probability beyond some value of r is desired, KEY C', the calculation is made for 1 minus the input probability value. The substitution is made at lines 146 to 153.

Lines 154 to 184 start the calculation by finding the probability of zero events, Pr(0), Eq. (3.23). If the probability of zero events is below the lower limits of the calculator, the program is directed to a routine at line 347 to clear the error message from the calculator and to substitute $1 - p$ for p and $n - r$ for r.

Lines 185 to 243 are the summation calculation routine following Eq. (3.20). Each individual calculation is stored in Register 04 as it is made, and the summation is stored in Register 08. After each calculation, a check is made to determine whether the value of r has been reached, and if there was probability input, whether the probability value was exceeded. If not, the program returns to line 185 for another calculation. Register 09 is used to index the calculations.

Lines 244 to 273 are the end operations for variable-value input. The different probabilities are stored in the registers shown in Table 3.6, and the program returns to the starting area to print the requested value.

Lines 274 to 284 are the routine to check whether the calculated probability has exceeded the input value of P. If not, the program returns to line 185 for the calculation of another term. If so, the program goes on to line 285.

Lines 285 to 337 are the print routine for a calculation to match input probability. Lines 285 to 307 print the results if a "less than" probability was to be matched. Lines 308 to 337 print the results if a "greater than" probability is required.

Lines 338 to 346 are the division by zero routine to generate an error message if impossible data are entered.

Lines 347 to 370 are the routine to substitute $1 - p$ for p and $n - r$ for r if the probability of zero with the input data is beyond the lower limit of the calculator. Flag 6 is set at line 367, and when the calculation is completed, the original values are replaced in the next routine.

Lines 371 to 393 are the replacement routine before the program is returned to the starting areas for printing the results.

Table B.4 Poisson Probability Program (TI-59)

m, KEY A; *r*, KEY A gives Pr(*r*)
m, KEY B; *r*, KEY B gives Pr(>*r*)
m, KEY C; *r*, KEY C gives Pr(≧*r*)
m, KEY D; *r*, KEY D gives Pr(<*r*)
m, KEY E; *r*, KEY E gives Pr(≦*r*)

m, KEY C'; *P*, KEY C' gives *r* such that *P* = Pr(≧*r*)

m, KEY E'; *P*, KEY E' gives *r* such that *P* = Pr(≦*r*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	038	15	E	076	43	RCL
001	76	LBL	039	71	SBR	077	11	11
002	11	A	040	22	INV	078	75	-
003	71	SBR	041	43	RCL	079	01	1
004	22	INV	042	07	07	080	54	)
005	43	RCL	043	99	PRT	081	42	STO
006	03	03	044	81	RST	082	12	12
007	99	PRT	045	91	R/S	083	91	R/S
008	81	RST	046	76	LBL	084	76	LBL
009	91	R/S	047	18	C'	085	24	CE
010	76	LBL	048	86	STF	086	43	RCL
011	12	B	049	00	00	087	01	01
012	71	SBR	050	76	LBL	088	42	STO
013	22	INV	051	10	E'	089	11	11
014	43	RCL	052	86	STF	090	91	R/S
015	04	04	053	01	01	091	76	LBL
016	99	PRT	054	76	LBL	092	23	LNX
017	81	RST	055	22	INV	093	42	STO
018	91	R/S	056	87	IFF	094	02	02
019	76	LBL	057	05	05	095	42	STO
020	13	C	058	23	LNX	096	10	10
021	71	SBR	059	86	STF	097	87	IFF
022	22	INV	060	05	05	098	01	01
023	43	RCL	061	47	CMS	099	25	CLR
024	05	05	062	42	STO	100	22	INV
025	99	PRT	063	01	01	101	59	INT
026	81	RST	064	32	X:T	102	32	X:T
027	91	R/S	065	02	2	103	00	0
028	76	LBL	066	02	2	104	22	INV
029	14	D	067	07	7	105	67	EQ
030	71	SBR	068	42	STO	106	95	=
031	22	INV	069	11	11	107	32	X:T
032	43	RCL	070	77	GE	108	61	GTO
033	06	06	071	24	CE	109	32	X:T
034	99	PRT	072	53	(	110	76	LBL
035	81	RST	073	43	RCL	111	25	CLR
036	91	R/S	074	01	01	112	22	INV
037	76	LBL	075	55	÷	113	87	IFF

Table B.4 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
114	00	00	157	28	28	200	00	0
115	55	÷	158	53	(	201	77	GE
116	53	(	159	43	RCL	202	35	1/X
117	01	1	160	01	01	203	86	STF
118	75	-	161	55	÷	204	04	04
119	43	RCL	162	43	RCL	205	61	GTO
120	10	10	163	08	08	206	44	SUM
121	54	)	164	65	×	207	76	LBL
122	42	STO	165	43	RCL	208	35	1/X
123	10	10	166	03	03	209	53	(
124	76	LBL	167	44	SUM	210	43	RCL
125	55	÷	168	06	06	211	07	07
126	32	X:T	169	54	)	212	75	-
127	00	0	170	42	STO	213	43	RCL
128	77	GE	171	03	03	214	03	03
129	95	=	172	44	SUM	215	54	)
130	01	1	173	07	07	216	53	(
131	22	INV	174	87	IFF	217	42	STO
132	77	GE	175	01	01	218	06	06
133	95	=	176	42	STO	219	94	+/-
134	76	LBL	177	43	RCL	220	85	+
135	32	X:T	178	08	08	221	01	1
136	43	RCL	179	32	X:T	222	54	)
137	11	11	180	43	RCL	223	53	(
138	94	+/-	181	02	02	224	42	STO
139	22	INV	182	67	EQ	225	05	05
140	23	LNX	183	34	√X	226	75	-
141	42	STO	184	43	RCL	227	43	RCL
142	03	03	185	07	07	228	03	03
143	42	STO	186	32	X:T	229	54	)
144	07	07	187	52	EE	230	42	STO
145	43	RCL	188	09	9	231	04	04
146	02	02	189	06	6	232	22	INV
147	32	X:T	190	22	INV	233	57	ENG
148	00	0	191	77	GE	234	92	RTN
149	67	EQ	192	44	SUM	235	76	LBL
150	34	√X	193	61	GTO	236	42	STO
151	87	IFF	194	33	X²	237	43	RCL
152	01	01	195	76	LBL	238	07	07
153	42	STO	196	34	√X	239	32	X:T
154	76	LBL	197	43	RCL	240	43	RCL
155	33	X²	198	12	12	241	10	10
156	69	OP	199	32	X:T	242	77	GE

Table B.4. (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	33	X^2	277	01	1	311	53	(
244	87	IFF	278	85	+	312	43	RCL
245	00	00	279	43	RCL	313	12	12
246	53	(	280	08	08	314	75	-
247	43	RCL	281	54	)	315	01	1
248	07	07	282	99	PRT	316	42	STO
249	99	PRT	283	98	ADV	317	03	03
250	43	RCL	284	53	(	318	54	)
251	08	08	285	01	1	319	42	STO
252	99	PRT	286	75	-	320	12	12
253	98	ADV	287	43	RCL	321	76	LBL
254	43	RCL	288	06	06	322	52	EE
255	06	06	289	54	)	323	53	(
256	99	PRT	290	99	PRT	324	43	RCL
257	53	(	291	43	RCL	325	13	13
258	43	RCL	292	08	08	326	65	×
259	08	08	293	99	PRT	327	43	RCL
260	75	-	294	98	ADV	328	11	11
261	01	1	295	81	RST	329	54	)
262	54	)	296	91	R/S	330	94	+/-
263	99	PRT	297	76	LBL	331	22	INV
264	98	ADV	298	44	SUM	332	23	LNX
265	81	RST	299	01	1	333	49	PRD
266	91	R/S	300	32	X⇌T	334	03	03
267	76	LBL	301	43	RCL	335	49	PRD
268	53	(	302	12	12	336	06	06
269	53	(	303	77	GE	337	49	PRD
270	01	1	304	45	Y^X	338	07	07
271	75	-	305	42	STO	339	87	IFF
272	43	RCL	306	13	13	340	04	04
273	07	07	307	61	GTO	341	35	1/X
274	54	)	308	52	EE	342	61	GTO
275	99	PRT	309	76	LBL	343	33	X^2
276	53	(	310	45	Y^X	344	91	R/S

B.4 DETAILS OF THE POISSON PROBABILITY PROGRAM

Lines 0 to 53 establish the type of output requested. The calculation is the same in all cases, and all the probabilities are stored in the calculator. The starting area is returned to at the end of the calculation and the desired probability is recalled and printed.

Lines 54 to 96 store the input values and check whether the expected value m exceeds 227, which is the limit the calculator can process. The first probability calculation is Pr(0), which is e^{-m}, and if m is larger than 227, the answer is below the lower limit of the calculator. If m is less than 227, the program skips lines 72 to 83. If not, m is reduced to 227 and the remainder is stored in Register 12 to be called up later.

Lines 97 to 99 check whether an input probability value is to be matched. If so, the program goes to line 110.

Lines 100 to 109 check whether the input value of r was an integer. If not, the program generates an error message by being directed to a nonexistent routine.

Lines 110 to 133 check whether the input probability to be matched is "less than" or "greater than." Inasmuch as the probability calculation is from zero up to some value of r, if a "greater than" probability is to be matched, a value 1 minus the input value is substituted before the calculation is started. These lines also check whether the input probability value was less than zero or greater than 1. If so, the error message is generated.

Lines 134 to 153 start the probability calculation with the determination of e^{-m}, lines 136 to 140.

Lines 154 to 194 are the main summation calculation, following Eqs. (3.26) and (3.27). After each calculation a check is made to determine whether the value of r has been reached, or with probability input, whether the value of P has been exceeded. If not, the calculation is repeated with Register 09 serving as index.

If the original value of m exceeded 227 and had been reduced earlier in the program, the possibility exists that the calculated probability will exceed the upper limit of the calculator before the balance of the original m value is returned. Lines 185 to 192 check the calculated probability against 1×10^{96}. If it exceeds this value, the program goes to a routine to return some of the original m value.

Lines 194 to 206 are a check, after the probability calculation has been completed, whether all the original m value has been used. If not, the program goes to the routine at line 297 to add the remainder.

Lines 207 to 234 are the end routine to distribute the calculated probabilities in the registers listed in Table 3.10 before returning to the starting location for printing out the desired result.

Lines 235 to 296 are the end routine if probability input was to be matched. If the probability has not been matched, line 242, the program returns to the calculation routine. If the probability has been matched, the program either prints the results, if a "less than" probability was asked for, lines 247 to 265; or it makes the necessary adjustment and prints the "greater than" probabilities at lines 267 to 296.

Lines 297 to 344 are the routine for returning the portion of the e^{-m} term that was removed if m was greater than 227. The portion not used was stored in Register 12, and it is returned at this point. The probabilities that have been accumulated in Registers 03, 06, and 07 are adjusted. The program then returns to the calculation routine at line 154.

Table B.5 t-1 Program for Hypotheses about Means (TI-59)

Hypothesis: $\bar{x} = m$	Hypothesis: $m_x = m_y$	Find α
x_i, KEY A m, KEY C	x_i, KEY A y_i, KEY B Run: KEY C	t, KEY $x \rightleftharpoons t$ ν, KEY D
Output: $\bar{x}$, $s(x)$, t, ν, α	Output: $\bar{x}$, $s(x)$, $\bar{y}$, $s(y)$, t, ν, α	Output: α

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	036	99	PRT	072	75	-
001	76	LBL	037	98	ADV	073	01	1
002	11	A	038	43	RCL	074	54	)
003	87	IFF	039	03	03	075	42	STO
004	02	02	040	42	STO	076	19	19
005	22	INV	041	17	17	077	61	GTO
006	47	CMS	042	29	CP	078	24	CE
007	76	LBL	043	43	RCL	079	76	LBL
008	22	INV	044	13	13	080	23	LNX
009	86	STF	045	22	INV	081	43	RCL
010	02	02	046	67	EQ	082	11	11
011	78	Σ+	047	23	LNX	083	42	STO
012	91	R/S	048	53	(	084	01	01
013	76	LBL	049	53	(	085	43	RCL
014	12	B	050	53	(	086	12	12
015	44	SUM	051	43	RCL	087	42	STO
016	11	11	052	14	14	088	02	02
017	33	X²	053	75	-	089	43	RCL
018	44	SUM	054	43	RCL	090	13	13
019	12	12	055	07	07	091	42	STO
020	01	1	056	54	)	092	03	03
021	44	SUM	057	55	÷	093	79	x̄
022	13	13	058	43	RCL	094	99	PRT
023	91	R/S	059	15	15	095	42	STO
024	76	LBL	060	54	)	096	13	13
025	13	C	061	65	×	097	22	INV
026	42	STO	062	43	RCL	098	79	x̄
027	07	07	063	03	03	099	99	PRT
028	79	x̄	064	34	√X	100	98	ADV
029	99	PRT	065	54	)	101	42	STO
030	42	STO	066	42	STO	102	12	12
031	14	14	067	16	16	103	53	(
032	22	INV	068	99	PRT	104	53	(
033	79	x̄	069	53	(	105	53	(
034	42	STO	070	43	RCL	106	43	RCL
035	15	15	071	03	03	107	15	15

Table B.5 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
108	33	X²	151	14	14	194	34	√X
109	65	×	152	75	-	195	54	)
110	53	(	153	43	RCL	196	50	I×I
111	43	RCL	154	13	13	197	70	RAD
112	17	17	155	54	)	198	22	INV
113	75	-	156	55	÷	199	30	TAN
114	01	1	157	43	RCL	200	42	STO
115	54	)	158	18	18	201	10	10
116	54	)	159	54	)	202	38	SIN
117	85	+	160	55	÷	203	42	STO
118	53	(	161	53	(	204	11	11
119	43	RCL	162	43	RCL	205	29	CP
120	12	12	163	17	17	206	43	RCL
121	33	X²	164	35	1/X	207	00	00
122	65	×	165	85	+	208	22	INV
123	53	(	166	43	RCL	209	59	INT
124	43	RCL	167	03	03	210	22	INV
125	03	03	168	35	1/X	211	67	EQ
126	75	-	169	54	)	212	95	=
127	01	1	170	34	√X	213	02	2
128	54	)	171	54	)	214	32	X:T
129	54	)	172	99	PRT	215	43	RCL
130	54	)	173	42	STO	216	00	00
131	55	÷	174	16	16	217	22	INV
132	53	(	175	76	LBL	218	77	GE
133	43	RCL	176	24	CE	219	42	STO
134	03	03	177	43	RCL	220	67	EQ
135	85	+	178	16	16	221	43	RCL
136	43	RCL	179	32	X:T	222	53	(
137	17	17	180	43	RCL	223	43	RCL
138	75	-	181	19	19	224	00	00
139	02	2	182	99	PRT	225	75	-
140	54	)	183	76	LBL	226	02	2
141	42	STO	184	14	D	227	54	)
142	19	19	185	42	STO	228	42	STO
143	54	)	186	00	00	229	09	09
144	34	√X	187	32	X:T	230	01	1
145	42	STO	188	53	(	231	42	STO
146	18	18	189	42	STO	232	08	08
147	53	(	190	01	01	233	76	LBL
148	53	(	191	55	÷	234	35	1/X
149	53	(	192	43	RCL	235	43	RCL
150	43	RCL	193	00	00	236	09	09

Table B.5 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
237	32	X:T	273	97	DSZ	309	76	LBL
238	01	1	274	09	09	310	42	STO
239	67	EQ	275	35	1/X	311	43	RCL
240	44	SUM	276	53	(	312	11	11
241	53	(	277	43	RCL	313	42	STO
242	53	(	278	08	08	314	14	14
243	43	RCL	279	65	×	315	61	GTO
244	08	08	280	43	RCL	316	65	×
245	65	×	281	11	11	317	76	LBL
246	53	(	282	54	)	318	43	RCL
247	43	RCL	283	61	GTO	319	43	RCL
248	10	10	284	25	CLR	320	11	11
249	39	COS	285	76	LBL	321	76	LBL
250	33	X²	286	44	SUM	322	25	CLR
251	55	÷	287	53	(	323	94	+/-
252	43	RCL	288	53	(	324	85	+
253	09	09	289	53	(	325	01	1
254	54	)	290	43	RCL	326	95	=
255	65	×	291	08	08	327	99	PRT
256	53	(	292	65	×	328	81	RST
257	43	RCL	293	43	RCL	329	91	R/S
258	09	09	294	11	11	330	76	LBL
259	75	-	295	65	×	331	65	×
260	01	1	296	43	RCL	332	53	(
261	54	)	297	10	10	333	43	RCL
262	54	)	298	39	COS	334	14	14
263	85	+	299	54	)	335	65	×
264	01	1	300	85	+	336	02	2
265	54	)	301	43	RCL	337	55	÷
266	42	STO	302	10	10	338	89	π
267	08	08	303	54	)	339	54	)
268	97	DSZ	304	54	)	340	61	GTO
269	09	09	305	42	STO	341	25	CLR
270	52	EE	306	14	14	342	91	R/S
271	76	LBL	307	61	GTO			
272	52	EE	308	65	×			

B.5 DETAILS OF THE t-1 Program

Lines 0 to 12 take the input of data if one mean is to be tested and use the built-in function of KEY Σ+ to accumulate the data.

Lines 13 to 23 take the input of the second set of data if two means are to be compared. The sums required for the calculation of the standard deviation are accumulated by the program, and the number of data entries is counted.

Lines 24 to 47 use the built-in routines to calculate the mean and standard deviation of the first set of data. A check is made on Register 13 to determine whether a second set of data was entered. If so, the program goes on to line 79. If not, the program continues with

Lines 48 to 78 The t value is calculated from Eq. (4.4), and the program goes on to the probability routine at line 175.

Lines 79 to 100 shift the sums of the second variable to the storage registers used by the built-in routines for the calculation of the mean and the standard deviation. These are then calculated.

Lines 101 to 174 calculate the t value from Eq. (4.6) for the comparison of two means.

Lines 175 to 182 arrange the t value and the degrees of freedom in the proper sequence for the calculation of the probability.

Lines 183 to 204 start the calculation of the probability following Eqs. (4.17) and (4.18). Lines 189 to 204 calculate and store the terms θ and sin θ.

Lines 205 to 221 check whether the number of degrees of freedom is one or two. If ν is 1, the probability is $1 - 2\theta/\pi$; if ν is 2, the probability is sin θ. If ν is not an integer, an error message is generated by directing the program to a nonexistent label.

Lines 222 to 275 The expansion sections of the rearrangement of Eqs. (4.17) and (4.18) are the same for even and odd values of the ν. The multiplying factors are different. Lines 233 to 275 carry out the calculation of the section inside the multiple parentheses of the rearranged equations. If ν is odd when the calculation is completed, the program goes to line 285. If the value of ν is even, the program goes to line 276. Whether ν is odd or even is determined by decreasing ν in increments of 2 for each step in the calculation and determining whether the final value is 1 or zero.

Lines 276 to 284 multiply the calculation by sin θ if ν is even and then go on to the end routine at line 321.

Lines 285 to 308 multiply the calculation by sin θ cos θ, add θ, and multiply the sum by $2/\pi$ and then go to the end routine.

Lines 309 to 320 supply the values for the probability when ν is one or two.

Lines 321 to 329 are the end routine to subtract the calculated probability from 1 and print the result.

Lines 330 to 341 are the routine to multiply by $2/\pi$.

The section of the program starting with line 183 is the routine for calculating the probability with input of t and ν. This entire section of the program can be lifted from t-1 and used elsewhere if the α calculation is wanted. References in the text to t-3 refer to this portion of the program.

Table B.6 t-2 Program for Calculating Confidence Intervals and Sample Size (TI-59)

Confidence interval	Sample size	Find t
x_i, KEY A' P (confidence = $1 - \alpha$), KEY B'	$s(x)$, KEY C' α, KEY C' β, KEY C' Δ (= $m - m_1$), KEY C'	α, KEY $x \rightleftharpoons t$ ν, KEY D'
Output: $\bar{x}$, x_l, x_u	Output: n	Output: t

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	035	24	CE	070	65	×
001	76	LBL	036	42	STO	071	43	RCL
002	16	A'	037	12	12	072	12	12
003	87	IFF	038	79	X̄	073	54	)
004	02	02	039	99	PRT	074	99	PRT
005	22	INV	040	42	STO	075	98	ADV
006	47	CMS	041	13	13	076	81	RST
007	76	LBL	042	53	(	077	91	R/S
008	22	INV	043	22	INV	078	76	LBL
009	86	STF	044	79	X̄	079	18	C'
010	02	02	045	55	÷	080	87	IFF
011	78	Σ+	046	43	RCL	081	03	03
012	91	R/S	047	03	03	082	25	CLR
013	76	LBL	048	34	√X	083	47	CMS
014	17	B'	049	54	)	084	42	STO
015	53	(	050	53	(	085	10	10
016	42	STO	051	53	(	086	86	STF
017	10	10	052	42	STO	087	03	03
018	75	-	053	14	14	088	91	R/S
019	01	1	054	65	×	089	76	LBL
020	54	)	055	43	RCL	090	25	CLR
021	94	+/-	056	12	12	091	87	IFF
022	42	STO	057	54	)	092	04	04
023	11	11	058	75	-	093	32	X:T
024	53	(	059	43	RCL	094	42	STO
025	43	RCL	060	13	13	095	11	11
026	03	03	061	54	)	096	42	STO
027	75	-	062	94	+/-	097	21	21
028	01	1	063	99	PRT	098	86	STF
029	54	)	064	53	(	099	04	04
030	42	STO	065	43	RCL	100	91	R/S
031	09	09	066	13	13	101	76	LBL
032	71	SBR	067	85	+	102	32	X:T
033	23	LNX	068	43	RCL	103	87	IFF
034	71	SBR	069	14	14	104	05	05

Table B.6 (*Cont.*)

Step	Code	Key
105	33	X²
106	42	STO
107	13	13
108	53	(
109	43	RCL
110	13	13
111	65	×
112	02	2
113	54	)
114	42	STO
115	13	13
116	86	STF
117	05	05
118	91	R/S
119	76	LBL
120	33	X²
121	53	(
122	42	STO
123	12	12
124	55	÷
125	43	RCL
126	10	10
127	54	)
128	42	STO
129	15	15
130	86	STF
131	00	00
132	43	RCL
133	11	11
134	71	SBR
135	23	LNX
136	22	INV
137	86	STF
138	00	00
139	43	RCL
140	13	13
141	42	STO
142	11	11
143	71	SBR
144	23	LNX
145	53	(
146	53	(
147	53	(
148	43	RCL
149	00	00
150	85	+
151	43	RCL
152	07	07
153	54	)
154	55	÷
155	43	RCL
156	15	15
157	54	)
158	33	X²
159	59	INT
160	75	−
161	01	1
162	54	)
163	42	STO
164	17	17
165	00	0
166	32	X⇌T
167	43	RCL
168	17	17
169	77	GE
170	34	√X
171	01	1
172	42	STO
173	09	09
174	61	GTO
175	35	1/X
176	76	LBL
177	34	√X
178	43	RCL
179	17	17
180	42	STO
181	09	09
182	76	LBL
183	35	1/X
184	69	OP
185	29	29
186	43	RCL
187	07	07
188	71	SBR
189	24	CE
190	42	STO
191	08	08
192	00	0
193	42	STO
194	14	14
195	43	RCL
196	00	00
197	71	SBR
198	24	CE
199	42	STO
200	16	16
201	00	0
202	42	STO
203	14	14
204	53	(
205	53	(
206	43	RCL
207	16	16
208	85	+
209	43	RCL
210	08	08
211	54	)
212	55	÷
213	43	RCL
214	15	15
215	54	)
216	33	X²
217	32	X⇌T
218	53	(
219	43	RCL
220	09	09
221	85	+
222	01	1
223	54	)
224	22	INV
225	77	GE
226	35	1/X
227	99	PRT
228	43	RCL
229	21	21
230	42	STO
231	14	14
232	98	ADV
233	81	RST

Table B.6 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
234	91	R/S	277	85	+	320	76	LBL
235	76	LBL	278	93	.	321	44	SUM
236	19	D'	279	02	2	322	22	INV
237	47	CMS	280	07	7	323	86	STF
238	42	STO	281	00	0	324	00	00
239	09	09	282	06	6	325	42	STO
240	32	X:T	283	01	1	326	07	07
241	42	STO	284	65	×	327	92	RTN
242	11	11	285	43	RCL	328	76	LBL
243	42	STO	286	00	00	329	24	CE
244	12	12	287	54	)	330	42	STO
245	71	SBR	288	55	÷	331	18	18
246	23	LNX	289	53	(	332	33	X^2
247	71	SBR	290	01	1	333	42	STO
248	24	CE	291	85	+	334	19	19
249	99	PRT	292	93	.	335	53	(
250	98	ADV	293	09	9	336	53	(
251	81	RST	294	09	9	337	53	(
252	91	R/S	295	02	2	338	53	(
253	76	LBL	296	02	2	339	43	RCL
254	23	LNX	297	09	9	340	19	19
255	53	(	298	65	×	341	65	×
256	43	RCL	299	43	RCL	342	03	3
257	11	11	300	00	00	343	85	+
258	55	÷	301	85	+	344	01	1
259	02	2	302	93	.	345	09	9
260	54	)	303	00	0	346	54	)
261	33	X^2	304	04	4	347	65	×
262	35	1/X	305	04	4	348	43	RCL
263	23	LNX	306	08	8	349	19	19
264	34	$\sqrt{X}$	307	01	1	350	85	+
265	53	(	308	65	×	351	01	1
266	42	STO	309	43	RCL	352	07	7
267	00	00	310	00	00	353	54	)
268	75	-	311	33	X^2	354	65	×
269	53	(	312	54	)	355	43	RCL
270	02	2	313	54	)	356	19	19
271	93	.	314	87	IFF	357	75	-
272	03	3	315	00	00	358	01	1
273	00	0	316	44	SUM	359	05	5
274	07	7	317	42	STO	360	54	)
275	05	5	318	00	00	361	65	×
276	03	3	319	92	RTN	362	43	RCL

Table B.6 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
363	18	18	391	09	09	419	08	8
364	54	)	392	54	)	420	04	4
365	44	SUM	393	44	SUM	421	65	×
366	14	14	394	14	14	422	53	(
367	53	(	395	53	(	423	43	RCL
368	53	(	396	53	(	424	09	09
369	53	(	397	43	RCL	425	45	YX
370	43	RCL	398	19	19	426	03	3
371	19	19	399	65	×	427	54	)
372	65	×	400	43	RCL	428	54	)
373	05	5	401	18	18	429	42	STO
374	85	+	402	85	+	430	20	20
375	01	1	403	43	RCL	431	65	×
376	06	6	404	18	18	432	43	RCL
377	54	)	405	54	)	433	18	18
378	65	×	406	65	×	434	54	)
379	43	RCL	407	09	9	435	44	SUM
380	19	19	408	06	6	436	14	14
381	85	+	409	65	×	437	53	(
382	03	3	410	43	RCL	438	43	RCL
383	54	)	411	09	09	439	14	14
384	65	×	412	33	X²	440	55	÷
385	43	RCL	413	54	)	441	43	RCL
386	18	18	414	44	SUM	442	20	20
387	65	×	415	14	14	443	54	)
388	04	4	416	53	(	444	92	RTN
389	65	×	417	53	(			
390	43	RCL	418	03	3			

B.6 DETAILS OF THE t-2 PROGRAM

Lines 0 to 12 take the input variable data for the confidence interval calculation and calculate the sums using the built-in function KEY Σ+.

Lines 13 to 31 convert the input probability to an α value, and calculate the degrees of freedom. The program then goes to the t routine at line 253.

Lines 32 to 36 call the two subroutines. The normal probability z value is calculated first, and the t value is calculated from the normal value.

Lines 37 to 44 After the t value has been calculated, the program uses the built-in functions to obtain the mean and the standard deviation.

Lines 45 to 77 calculate the lower limit and the upper limit of the confidence range, and print the results. Equations (4.10) and (4.11) are followed for the calculations.

Lines 78 to 133 are the start of the sample size calculation. The input values are stored, and the β value is doubled, since the t calculation is based on a two-sided probability value and the β value is for one tail of the distribution; $s(x)/(m_1 - m)$ of Eq. (4.16) is used in the calculation of the sample size as a single value. It is calculated at line 127 and stored in Register 15. Flag 00 is set for a purpose to be described later.

Lines 134 and 135 direct the program to the start of the t calculation, but only the normal deviate value is calculated if Flag 00 is set.

Lines 136 to 144 calculate the normal deviate for the 2β probability value.

Lines 145 to 164 calculate the sample size from the normal deviate values corresponding to α and 2β. The value obtained from this calculation is the minimum sample that would be required if $s(x)$ was not an estimate but was the true standard deviation of the population. The actual sample-size calculation is by trial and error, and starting with the minimum sample size, rather than starting with zero or one, minimizes the number of trials.

Lines 165 to 181 check that the minimum sample size is greater than 1. If not, the initial calculation is made with a sample size of 1.

Lines 182 to 198 call on the t subroutine for the calculation of t_α and t_β.

Lines 199 to 226 are the trial-and-error routine; n is calculated from the following relation:

$$n = \left((t_\alpha + t_\beta)\frac{s(x)}{m_1 - m}\right)^2 \tag{B.1.}$$

and tested against the value used for the two t calculations, which is stored in Register 09. The value is incremented by 1 for each trial until the values are in agreement.

Lines 227 to 234 are the print routine and a shifting of some of the values to conform with the storage given on page 61.

Lines 235 to 252 are the start of the calculation of t, if the calculation is commenced from the keyboard of the calculator rather than from one of the other routines.

Lines 253 to 327 calculate the normal deviate from Eq. (4.21); u of Eq. (4.21) is evaluated in lines 255 to 264, and the main calculation is carried out in lines 265 to 313. Inasmuch as the calculation is used twice when the sample-size calculation is made, Flag 00 is used at lines 314 to 326 to store one value in Register 00 and the other in Register 07.

Lines 328 to 444 calculate the t value from the normal deviate and the degrees of freedom, following Eq. (4.24). If the g term mentioned in the text were to be added to obtain greater accuracy, it would be added after line 443.

Table B.7 χ^2**-1 Program (TI-59)**

a/*b* = *p*/*q*?

a, KEY A
b, KEY A
p, KEY A
q, KEY A

Output:
Degrees of freedom
χ^2

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	033	24	CE	066	01	1
001	76	LBL	034	42	STO	067	54	)
002	11	A	035	04	04	068	55	÷
003	87	IFF	036	53	(	069	02	2
004	01	01	037	43	RCL	070	54	)
005	22	INV	038	03	03	071	54	)
006	86	STF	039	55	÷	072	33	X²
007	01	01	040	43	RCL	073	55	÷
008	47	CMS	041	04	04	074	53	(
009	42	STO	042	54	)	075	53	(
010	01	01	043	42	STO	076	43	RCL
011	91	R/S	044	03	03	077	03	03
012	76	LBL	045	53	(	078	65	×
013	22	INV	046	53	(	079	53	(
014	87	IFF	047	53	(	080	43	RCL
015	02	02	048	43	RCL	081	01	01
016	23	LNX	049	01	01	082	85	+
017	86	STF	050	75	−	083	43	RCL
018	02	02	051	53	(	084	02	02
019	42	STO	052	43	RCL	085	54	)
020	02	02	053	03	03	086	54	)
021	91	R/S	054	65	×	087	54	)
022	76	LBL	055	43	RCL	088	54	)
023	23	LNX	056	02	02	089	42	STO
024	87	IFF	057	54	)	090	05	05
025	03	03	058	54	)	091	01	1
026	24	CE	059	50	I×I	092	99	PRT
027	86	STF	060	75	−	093	43	RCL
028	03	03	061	53	(	094	05	05
029	42	STO	062	53	(	095	99	PRT
030	03	03	063	43	RCL	096	98	ADV
031	91	R/S	064	03	03	097	81	RST
032	76	LBL	065	85	+	098	91	R/S

B.7 DETAILS OF χ^2-1 PROGRAM: RATIO TEST

Lines 0 to 35 store the input data in Registers 01 to 04.

Lines 36 to 44 calculate the value of p/q and store it in Register 03.

Lines 45 to 59 calculate the value of $|a - (p/q)b|$ of Eq. (5.5).

Lines 60 to 72 complete the calculation of the numerator of Eq. (5.5).

Lines 73 to 90 complete the calculation of Eq. (5.5) and store the result, which is the χ^2 value, in Register 05.

Lines 91 to 97 print the value 1 for the number of degrees of freedom, then print the χ^2 value, and place the degrees of freedom and the χ^2 value in the proper sequence for the calculation of the probability by the χ^2-5 program.

The χ^2-5 program may be appended immediately following line 96, and the probability value will be part of the output.

Table B.8 χ^2**-2 Program (TI-59)**

$x_i = \bar{x}$?

x_i, KEY B

KEY C

Output:

$\bar{x}$
Degrees of freedom
χ^2

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	014	13	C	028	65	×
001	76	LBL	015	79	$\bar{x}$	029	43	RCL
002	12	B	016	99	PRT	030	03	03
003	87	IFF	017	53	(	031	55	÷
004	04	04	018	43	RCL	032	43	RCL
005	25	CLR	019	03	03	033	01	01
006	47	CMS	020	75	-	034	75	-
007	86	STF	021	01	1	035	43	RCL
008	04	04	022	54	)	036	01	01
009	76	LBL	023	99	PRT	037	54	)
010	25	CLR	024	32	X⇌T	038	99	PRT
011	78	Σ+	025	53	(	039	98	ADV
012	91	R/S	026	43	RCL	040	81	RST
013	76	LBL	027	02	02	041	91	R/S

B.8 DETAILS OF χ^2-2 PROGRAM: EQUAL EXPECTATION

Lines 0 to 12 take the input x_i values. The built-in functions of KEY Σ+ are used to accumulate the sums required for the χ^2 calculation.

Lines 13 to 16 obtain and print the value of $\bar{x}$.

Lines 17 to 24 calculate the degrees of freedom, print the result, and hold the value in the t register.

Lines 25 to 40 complete the calculation of χ^2, using the last equivalent form of Eq. (5.9), and print the result.

The degrees of freedom and the χ^2 value are in the proper sequence to be used with the χ^2-5 program to calculate the probability. The χ^2-5 program may be appended directly after line 39 and the probability value will be part of the output.

Table B.9 χ^2-3 Program (TI-59)

$x_i/f_i = \Sigma x/\Sigma f$? f_i, KEY $x \rightleftharpoons t$, x_i, KEY D, . . . , KEY E

Output:

χ^2 for each $f_i x_i$
Degrees of freedom
Total χ^2

Step	Code	Key
000	91	R/S
001	76	LBL
002	14	D
003	87	IFF
004	05	05
005	32	X:T
006	86	STF
007	05	05
008	47	CMS
009	42	STO
010	07	07
011	32	X:T
012	42	STO
013	08	08
014	01	1
015	01	1
016	42	STO
017	00	00
018	43	RCL
019	07	07
020	72	ST*
021	00	00
022	69	OP
023	20	20
024	44	SUM
025	01	01
026	43	RCL
027	08	08
028	72	ST*
029	00	00
030	44	SUM
031	02	02
032	69	OP
033	20	20
034	91	R/S
035	76	LBL
036	32	X:T
037	72	ST*
038	00	00
039	44	SUM
040	01	01
041	69	OP
042	20	20
043	32	X:T
044	72	ST*
045	00	00
046	44	SUM
047	02	02
048	69	OP
049	20	20
050	91	R/S
051	76	LBL
052	15	E
053	53	(
054	43	RCL
055	01	01
056	55	÷
057	43	RCL
058	02	02
059	54	)
060	42	STO
061	01	01
062	43	RCL
063	00	00
064	42	STO
065	02	02
066	01	1
067	01	1
068	42	STO
069	00	00
070	76	LBL
071	33	X²
072	73	RC*
073	00	00
074	42	STO
075	04	04
076	69	OP
077	20	20
078	53	(
079	53	(
080	43	RCL
081	04	04
082	75	-
083	53	(
084	43	RCL
085	01	01
086	65	×
087	73	RC*
088	00	00
089	54	)
090	72	ST*
091	00	00
092	54	)
093	33	X²
094	55	÷
095	73	RC*
096	00	00
097	54	)
098	99	PRT
099	44	SUM
100	05	05
101	69	OP
102	20	20
103	43	RCL
104	00	00
105	32	X:T
106	43	RCL
107	02	02
108	67	EQ
109	34	ΓX
110	61	GTO

Table B.9 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
111	33	X²	119	01	1	127	98	ADV
112	76	LBL	120	01	1	128	99	PRT
113	34	√X	121	54	)	129	32	X⇌T
114	53	(	122	55	÷	130	43	RCL
115	53	(	123	02	2	131	05	05
116	43	RCL	124	75	-	132	99	PRT
117	00	00	125	01	1	133	81	RST
118	75	-	126	54	)	134	91	R/S

B.9 DETAILS OF THE χ^2-3 PROGRAM: UNEQUAL EXPECTATION

Lines 0 to 34 take the first input of f_1 and x_1 and prepare the calculator for the balance of the data. The individual values are stored until all of the data are entered, so that twice as many storage areas are required as pairs of data. The first 11 storage areas are reserved for additional calculations so that 24 pairs of data may be entered without changing the standard partitioning of the calculator. If the χ^2-5 program is appended to χ^2-3 to include the probability calculation, there is still sufficient storage capacity for an additional five pairs of data. If χ^2-5 is not included, the program can handle up to 40 pairs of data. These increases in the data capacity would be achieved by exchanging program memory space for storage memory. No change in the program would be required, as the storage areas are selected by an index in the program.

Register 00 is used as the index register, and the input data are stored in the registers indicated by the number in the index which is incremented by two after each pair of data is entered.

Lines 35 to 50 take the input of data after the first pair; Σf is accumulated in Register 02, and Σx is accumulated in Register 01.

Lines 51 to 69 calculate and store the value of $\Sigma x/\Sigma f$.

Lines 70 to 111 recover the individual values of x_i and f_i and calculate the individual χ^2 values for each pair of entries, following Eq. (5.7). The sum is accumulated in Register 05. Each χ^2 value is printed as it is obtained—line 98.

Lines 112 to 134 calculate the degrees of freedom: 1 fewer than the total number of pairs of data, indicated by the last number in the index register. This value is printed and the total χ^2 value is printed. The degrees of freedom and the χ^2 value are put in the proper sequence for use by the χ^2-5 program for the probability calculation.

If χ^2-5 is appended to χ^2-3, it is started at line 133 and the probability calculation will follow directly after the calculations for χ^2-3 have been completed.

Table B.10 χ^2-4 Program: $r \times c$ Contingency Tables (TI-59)

Input: Number of rows, r, KEY A'; data by columns, x_{ij}, KEY A'; KEY B'

Output:

χ^2 for each row
Total degrees of freedom
χ^2 Total

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	037	74	SM*	074	53	(
001	76	LBL	038	06	06	075	32	X:T
002	16	A'	039	71	SBR	076	55	÷
003	87	IFF	040	43	RCL	077	43	RCL
004	00	00	041	69	OP	078	00	00
005	42	STO	042	36	36	079	42	STO
006	86	STF	043	43	RCL	080	04	04
007	00	00	044	06	06	081	54	)
008	47	CMS	045	32	X:T	082	74	SM*
009	42	STO	046	01	1	083	06	06
010	01	01	047	00	0	084	53	(
011	42	STO	048	67	EQ	085	43	RCL
012	06	06	049	44	SUM	086	01	01
013	01	1	050	91	R/S	087	65	×
014	00	0	051	76	LBL	088	02	2
015	44	SUM	052	44	SUM	089	54	)
016	06	06	053	01	1	090	22	INV
017	91	R/S	054	44	SUM	091	44	SUM
018	76	LBL	055	03	03	092	06	06
019	42	STO	056	43	RCL	093	69	OP
020	42	STO	057	01	01	094	36	36
021	02	02	058	44	SUM	095	43	RCL
022	72	ST*	059	06	06	096	06	06
023	06	06	060	76	LBL	097	32	X:T
024	44	SUM	061	45	Y^X	098	01	1
025	00	00	062	73	RC*	099	00	0
026	53	(	063	06	06	100	67	EQ
027	43	RCL	064	33	X^2	101	52	EE
028	06	06	065	32	X:T	102	61	GTO
029	85	+	066	53	(	103	45	Y^X
030	43	RCL	067	43	RCL	104	76	LBL
031	01	01	068	01	01	105	52	EE
032	54	)	069	65	×	106	00	0
033	42	STO	070	02	2	107	42	STO
034	06	06	071	54	)	108	00	00
035	43	RCL	072	44	SUM	109	43	RCL
036	02	02	073	06	06	110	01	01

Table B.10 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
111	42	STO	154	43	RCL	197	06	06
112	06	06	155	32	X⇌T	198	75	-
113	01	1	156	72	ST*	199	32	X⇌T
114	00	0	157	06	06	200	54	)
115	44	SUM	158	69	OP	201	94	+/-
116	06	06	159	36	36	202	99	PRT
117	91	R/S	160	43	RCL	203	71	SBR
118	76	LBL	161	06	06	204	43	RCL
119	17	B'	162	32	X⇌T	205	69	OP
120	43	RCL	163	01	1	206	36	36
121	11	11	164	00	0	207	43	RCL
122	42	STO	165	67	EQ	208	06	06
123	05	05	166	54	)	209	32	X⇌T
124	53	(	167	53	(	210	01	1
125	43	RCL	168	43	RCL	211	00	0
126	01	01	169	01	01	212	67	EQ
127	65	×	170	65	×	213	61	GTO
128	02	2	171	02	2	214	61	GTO
129	54	)	172	54	)	215	55	÷
130	76	LBL	173	61	GTO	216	76	LBL
131	53	(	174	53	(	217	61	GTO
132	44	SUM	175	76	LBL	218	53	(
133	06	06	176	54	)	219	53	(
134	73	RC*	177	43	RCL	220	43	RCL
135	06	06	178	01	01	221	01	01
136	42	STO	179	44	SUM	222	75	-
137	02	02	180	06	06	223	01	1
138	71	SBR	181	76	LBL	224	54	)
139	43	RCL	182	55	÷	225	65	×
140	53	(	183	53	(	226	53	(
141	73	RC*	184	73	RC*	227	43	RCL
142	06	06	185	06	06	228	03	03
143	44	SUM	186	65	×	229	75	-
144	07	07	187	43	RCL	230	01	1
145	55	÷	188	07	07	231	54	)
146	43	RCL	189	54	)	232	54	)
147	02	02	190	32	X⇌T	233	98	ADV
148	54	)	191	43	RCL	234	99	PRT
149	35	1/X	192	01	01	235	42	STO
150	44	SUM	193	44	SUM	236	06	06
151	00	00	194	06	06	237	32	X⇌T
152	32	X⇌T	195	53	(	238	01	1
153	71	SBR	196	73	RC*	239	42	STO

Table B.10 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
240	01	01	263	05	5	286	01	01
241	22	INV	264	54	)	287	54	)
242	67	EQ	265	55	÷	288	99	PRT
243	35	1/X	266	43	RCL	289	32	X:T
244	53	(	267	08	08	290	43	RCL
245	53	(	268	54	)	291	06	06
246	53	(	269	33	X²	292	32	X:T
247	43	RCL	270	42	STO	293	81	RST
248	05	05	271	01	01	294	92	RTN
249	75	-	272	76	LBL	295	76	LBL
250	43	RCL	273	35	1/X	296	43	RCL
251	13	13	274	53	(	297	43	RCL
252	65	×	275	53	(	298	06	06
253	43	RCL	276	43	RCL	299	75	-
254	04	04	277	00	00	300	43	RCL
255	55	÷	278	75	-	301	01	01
256	43	RCL	279	01	1	302	54	)
257	07	07	280	54	)	303	42	STO
258	54	)	281	65	×	304	06	06
259	42	STO	282	43	RCL	305	92	RTN
260	08	08	283	07	07	306	91	R/S
261	75	-	284	65	×			
262	93	.	285	43	RCL			

B.10 DETAILS OF χ^2-4 PROGRAM: CONTINGENCY TABLES

Contingency table data are in rows and columns, and the designation is arbitrary. Program χ^2-4 as listed in Table B.10 can handle up to 16 rows of data with no limit on the number of columns. If χ^2-5, for the probability calculation, is appended to χ^2-4, the standard partition between storage and program memory must be changed to accommodate the total program. In this case 10 storage areas must be shifted to program availability and the total number of rows that can be accommodated will be 13.

Lines 0 to 17 take the input of the number of rows and arrange the storage areas for the main bulk of data. Register 06 is used to index the input.

Lines 18 to 50 process the data until a complete column has been entered. The data for each column are processed as a unit. Having the number of rows as an index, the program "knows" when the data from one column have been entered. Three storage areas are used for each row of data: one for the row total, one for $\Sigma x_{ij}^2/C_j$, where C_j is the total of the jth column, and one for the current column total.

Lines 51 to 117 are used when the total for one column has been entered. The number of columns is totaled in Register 03. Each column entry is squared and divided by the column total and stored. The column total has been accumulated in Register 00, and the current value is stored in Register 04 for later use if the data are from a 2 × 2 table with only one degree of freedom. The $\Sigma x_{ij}^2/C_j$ for each column is stored in a different register for each row for later division by the row total. After the calculation has been made for each column entry, the registers are set at lines 104 to 116 for the next column entry.

Lines 118 to 151 start the calculation of χ^2, following Eq. (5.8). The last data entry is stored in Register 05 for later use if the data are from a 2 × 2 table. The values of each $\Sigma x_{ij}^2/C_j$ are recalled and divided by the row total and summed in Register 00; ΣR_i is accumulated in Register 07 to obtain the total of all the data.

Lines 152 to 202 calculated χ^2 for each row from the term $T \cdot \Sigma(x_{ij}^2/R_i C_j) - T$, which equals χ^2 for the ith row, where T is the total of all the data. This value is printed.

Lines 203 to 215 direct the program to the subroutine that resets the index register and checks whether all the rows' χ^2's have been calculated. If not, the program returns to line 181 for the next calculation. If so, the program proceeds to line 216.

Lines 216 to 236 calculate the total number of degrees of freedom, equal to $(r - 1)(c - 1)$, and print this result. If the number of degrees of freedom is one, a correction is made to the χ^2 value.

Lines 237 to 271 make the correction for one degree of freedom. In a 2 × 2 contingency table all the differences between observed and expected values are equal. χ^2 without correction is the square of this difference Δ^2 multiplied by the sum of the reciprocals of the expected values. χ^2 corrected is equal to χ^2 uncorrected multiplied by $(\Delta - 0.5)^2/\Delta^2$. This is the correction made in lines 244 to 271. The difference Δ is calculated from the last data entry saved in Register 05. The correction is stored in Register 01, which had a value of 1.0 for use if there is no correction.

Lines 272 to 287 calculate χ^2 uncorrected and multiply by the value in Register 01.

Lines 288 to 294 print the result and place the total number of degrees of freedom and the total χ^2 in the proper sequence for use with the χ^2-5 program to calculate the probability.

Lines 295 to 305 are a subroutine for shifting the value in the index register after the storage of the different columns of data.

If χ^2-5 is to be appended to χ^2-4 for the probability calculation, it would be added after line 305, and lines 293 and 294 would be changed to GO TO E'. E' is the label of the probability program.

Table B.11 χ^2-5 Program: χ^2 Probability (TI-59)

Input: Degrees of freedom, KEY $x \rightleftharpoons t$
χ^2, KEY E'

Output: Probability of $\geqq \chi^2$

Step	Code	Key
000	91	R/S
001	76	LBL
002	10	E'
003	42	STO
004	02	02
005	32	X:T
006	42	STO
007	01	01
008	53	(
009	53	(
010	42	STO
011	03	03
012	65	×
013	02	2
014	75	-
015	01	1
016	54	)
017	34	√X
018	85	+
019	05	5
020	54	)
021	42	STO
022	04	04
023	53	(
024	43	RCL
025	02	02
026	65	×
027	02	2
028	54	)
029	34	√X
030	32	X:T
031	43	RCL
032	04	04
033	77	GE
034	65	×
035	01	1
036	52	EE
037	94	+/-
038	05	5
039	99	PRT
040	81	RST
041	92	RTN
042	76	LBL
043	65	×
044	00	0
045	42	STO
046	06	06
047	01	1
048	42	STO
049	05	05
050	53	(
051	53	(
052	43	RCL
053	02	02
054	55	÷
055	02	2
056	54	)
057	42	STO
058	08	08
059	45	YX
060	53	(
061	43	RCL
062	03	03
063	55	÷
064	02	2
065	54	)
066	54	)
067	42	STO
068	07	07
069	53	(
070	43	RCL
071	08	08
072	94	+/-
073	22	INV
074	23	LNX
075	49	PRD
076	07	07
077	54	)
078	53	(
079	53	(
080	43	RCL
081	03	03
082	85	+
083	02	2
084	54	)
085	55	÷
086	02	2
087	54	)
088	42	STO
089	04	04
090	53	(
091	43	RCL
092	04	04
093	75	-
094	01	1
095	54	)
096	42	STO
097	04	04
098	59	INT
099	32	X:T
100	43	RCL
101	04	04
102	22	INV
103	67	EQ
104	71	SBR
105	76	LBL
106	75	-
107	22	INV
108	49	PRD
109	07	07
110	69	OP
111	34	34
112	00	0
113	32	X:T
114	43	RCL
115	04	04
116	67	EQ

Table B.11 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
117	81	RST	142	07	07	167	01	1
118	61	GTO	143	76	LBL	168	52	EE
119	75	-	144	81	RST	169	94	+/-
120	76	LBL	145	02	2	170	04	4
121	71	SBR	146	44	SUM	171	22	INV
122	43	RCL	147	03	03	172	77	GE
123	04	04	148	76	LBL	173	95	=
124	22	INV	149	95	=	174	53	(
125	49	PRD	150	53	(	175	01	1
126	07	07	151	43	RCL	176	75	-
127	32	X:T	152	02	02	177	53	(
128	93	.	153	55	÷	178	43	RCL
129	05	5	154	43	RCL	179	06	06
130	67	EQ	155	03	03	180	85	+
131	85	+	156	54	)	181	01	1
132	69	OP	157	49	PRD	182	54	)
133	34	34	158	05	05	183	65	×
134	61	GTO	159	02	2	184	43	RCL
135	71	SBR	160	44	SUM	185	07	07
136	76	LBL	161	03	03	186	54	)
137	85	+	162	43	RCL	187	22	INV
138	89	π	163	05	05	188	52	EE
139	34	FX	164	44	SUM	189	99	PRT
140	22	INV	165	06	06	190	81	RST
141	49	PRD	166	32	X:T			

B.11 DETAILS OF χ^2-5 PROGRAM: χ^2 PROBABILITY

The χ^2-5 program can be run with input of degrees of freedom and a χ^2 value. The output is the probability of an equal or greater χ^2. The program can also be appended to any of the χ^2 programs that calculate a χ^2 from input of data. The labels for the different routines and the storage areas used have been selected so that there will be no interference with any of the χ^2 programs. If χ^2-5 is appended to χ^2-4 the partition between storage area and program area will have to be adjusted. The shifting of 10 storage areas will be sufficient. No change is required for any of the other programs.

Lines 0 to 7 store the χ^2 and degrees of freedom input.

Lines 8 to 41 test to determine whether the input value of χ^2 exceeds its mean from a normal approximation by more than five standard deviations. If not, the program goes to line 42. If so, the program prints a probability value of 1×10^{-5} and ends.

Lines 42 to 77 start the χ^2 calculation using Eq. (5.10). Values of 0 and 1 are stored in Registers 06 and 05 for the start of summation calculations; $(\chi^2/2)^{(\nu/2)} \cdot e^{-(\chi^2/2)}$ is calculated and stored.

Lines 78 to 104 establish whether $(\nu + 2)/2$ is an integer. If not, the program goes to line 120 to calculate the gamma value. If so, the program goes on to line 105.

Lines 105 to 119 calculate the first term of Eq. (5.10) if ν is even.

Lines 120 to 142 calculate the value of the first term of Eq. (5.10) if ν is odd.

Lines 143 to 173 carry out the summation calculation of Eq. (5.10). The individual terms of the summation are calculated in Register 05, and the summation is accumulated in Register 06. When the last term calculated is less than .0001, the calculation is terminated. If greater accuracy is desired, a smaller value can be used at lines 167 to 170.

Lines 174 to 190 complete the calculation of χ^2 by multiplying the summation term by the value stored in Register 07 and subtracting the product from 1.0. The calculator units are then returned to decimal form and the result is printed.

Table B.12 Variance-1 Program: Compare Variance with Standard (TI-59)

Input: x_i, KEY A; σ^2, KEY B
Output: $s^2(x)$, degrees of freedom, F, probability of F

Input: Degrees of freedom, KEY C; $s^2(x)$, KEY C, σ^2, KEY C
Output: F, probability of F

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	038	12	12	076	53	(
001	76	LBL	039	53	(	077	43	RCL
002	11	A	040	53	(	078	10	10
003	87	IFF	041	43	RCL	079	55	÷
004	01	01	042	10	10	080	43	RCL
005	22	INV	043	55	÷	081	12	12
006	47	CMS	044	43	RCL	082	54	)
007	86	STF	045	12	12	083	99	PRT
008	01	01	046	54	)	084	42	STO
009	76	LBL	047	99	PRT	085	12	12
010	22	INV	048	65	×	086	65	×
011	78	Σ+	049	43	RCL	087	43	RCL
012	91	R/S	050	11	11	088	21	21
013	76	LBL	051	54	)	089	54	)
014	13	C	052	61	GTO	090	76	LBL
015	87	IFF	053	25	CLR	091	25	CLR
016	02	02	054	76	LBL	092	42	STO
017	23	LNX	055	12	B	093	22	22
018	42	STO	056	42	STO	094	53	(
019	11	11	057	12	12	095	53	(
020	42	STO	058	22	INV	096	43	RCL
021	21	21	059	79	$\bar{x}$	097	21	21
022	86	STF	060	33	X²	098	65	×
023	02	02	061	99	PRT	099	02	2
024	91	R/S	062	42	STO	100	75	-
025	76	LBL	063	10	10	101	01	1
026	23	LNX	064	53	(	102	54	)
027	87	IFF	065	43	RCL	103	34	√X
028	03	03	066	03	03	104	85	+
029	24	CE	067	75	-	105	05	5
030	42	STO	068	01	1	106	54	)
031	10	10	069	54	)	107	42	STO
032	86	STF	070	99	PRT	108	09	09
033	03	03	071	42	STO	109	53	(
034	91	R/S	072	11	11	110	43	RCL
035	76	LBL	073	42	STO	111	22	22
036	24	CE	074	21	21	112	65	×
037	42	STO	075	53	(	113	02	2

Table B.12 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
114	54	)	157	43	RCL	200	32	X:T
115	34	√X	158	08	08	201	43	RCL
116	32	X:T	159	94	+/-	202	09	09
117	43	RCL	160	22	INV	203	67	EQ
118	09	09	161	23	LNX	204	81	RST
119	77	GE	162	49	PRD	205	61	GTO
120	65	×	163	07	07	206	75	-
121	01	1	164	54	)	207	76	LBL
122	52	EE	165	53	(	208	71	SBR
123	94	+/-	166	53	(	209	43	RCL
124	05	5	167	43	RCL	210	09	09
125	99	PRT	168	21	21	211	22	INV
126	22	INV	169	85	+	212	49	PRD
127	52	EE	170	02	2	213	07	07
128	81	RST	171	54	)	214	32	X:T
129	76	LBL	172	55	÷	215	93	.
130	65	×	173	02	2	216	05	5
131	00	0	174	54	)	217	67	EQ
132	42	STO	175	42	STO	218	85	+
133	26	26	176	09	09	219	69	OP
134	01	1	177	53	(	220	39	39
135	42	STO	178	43	RCL	221	61	GTO
136	25	25	179	09	09	222	71	SBR
137	53	(	180	75	-	223	76	LBL
138	53	(	181	01	1	224	85	+
139	43	RCL	182	54	)	225	89	π
140	22	22	183	42	STO	226	34	√X
141	55	÷	184	09	09	227	22	INV
142	02	2	185	59	INT	228	49	PRD
143	54	)	186	32	X:T	229	07	07
144	42	STO	187	43	RCL	230	76	LBL
145	08	08	188	09	09	231	81	RST
146	45	YX	189	22	INV	232	02	2
147	53	(	190	67	EQ	233	44	SUM
148	43	RCL	191	71	SBR	234	21	21
149	21	21	192	76	LBL	235	76	LBL
150	55	÷	193	75	-	236	95	=
151	02	2	194	22	INV	237	53	(
152	54	)	195	49	PRD	238	43	RCL
153	54	)	196	07	07	239	22	22
154	42	STO	197	69	OP	240	55	÷
155	07	07	198	39	39	241	43	RCL
156	53	(	199	00	0	242	21	21

Table B.12 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	54	)	256	94	+/-	269	54	)
244	49	PRD	257	04	4	270	65	×
245	25	25	258	22	INV	271	43	RCL
246	02	2	259	77	GE	272	07	07
247	44	SUM	260	95	=	273	54	)
248	21	21	261	53	(	274	22	INV
249	43	RCL	262	01	1	275	52	EE
250	25	25	263	75	-	276	99	PRT
251	44	SUM	264	53	(	277	42	STO
252	26	26	265	43	RCL	278	13	13
253	32	X:T	266	26	26	279	81	RST
254	01	1	267	85	+	280	91	R/S
255	52	EE	268	01	1			

B.12 DETAILS OF VARIANCE-1 PROGRAM: COMPARISON OF VARIANCE WITH A STANDARD

Lines 0 to 12 take the input data and use the built-in function of KEY $\Sigma+$. (Erroneous input can be corrected directly with KEY $\Sigma-$.)

Lines 13 to 38 are for the second portion of the program which takes input of an estimated variance. This portion of the program stores the input data.

Lines 39 to 53 calculate and print the F value from the input of an estimated variance and the standard variance and direct the program to the probability calculation.

Lines 54 to 63 use the built-in function to calculate the variance estimate from the input data, print and store the result. (If the mean of the input data is wanted, it could be obtained after line 57 with KEY $\bar{x}$ and PRINT.)

Lines 64 to 72 calculate the degrees of freedom and print and store the result.

Lines 73 to 89 calculate the F value, print the result, and arrange the data in the proper sequence for the calculation of the probability.

Lines 90 to 279 are the same as the χ^2-5 program for the probability calculation of χ^2; χ^2 at ν degrees of freedom has the same probability as F at ν, and ∞ degrees of freedom when χ^2 is equal to $F \cdot \nu$.

Table B.13 Variance-2 Program: Comparison of Two Variances (TI-59)

Input: x_i, KEY A; y_i, KEY B; KEY C

Output: $s^2(x)$, ν_x, $s^2(y)$, ν_y, F value (probability of F can be obtained with the Variance-9 program)

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	037	54	)	074	43	RCL
001	76	LBL	038	42	STO	075	13	13
002	11	A	039	11	11	076	54	)
003	87	IFF	040	99	PRT	077	32	X:T
004	01	01	041	98	ADV	078	01	1
005	22	INV	042	43	RCL	079	77	GE
006	86	STF	043	07	07	080	23	LNX
007	01	01	044	42	STO	081	32	X:T
008	47	CMS	045	01	01	082	99	PRT
009	76	LBL	046	43	RCL	083	42	STO
010	22	INV	047	08	08	084	03	03
011	78	Σ+	048	42	STO	085	43	RCL
012	91	R/S	049	02	02	086	11	11
013	76	LBL	050	43	RCL	087	42	STO
014	12	B	051	09	09	088	01	01
015	44	SUM	052	42	STO	089	43	RCL
016	07	07	053	03	03	090	03	03
017	33	X^2	054	22	INV	091	81	RST
018	44	SUM	055	79	$\bar{X}$	092	76	LBL
019	08	08	056	33	X^2	093	23	LNX
020	01	1	057	42	STO	094	32	X:T
021	44	SUM	058	13	13	095	35	1/X
022	09	09	059	99	PRT	096	99	PRT
023	91	R/S	060	53	(	097	42	STO
024	76	LBL	061	43	RCL	098	03	03
025	13	C	062	03	03	099	43	RCL
026	22	INV	063	75	-	100	02	02
027	79	$\bar{X}$	064	01	1	101	42	STO
028	33	X^2	065	54	)	102	01	01
029	42	STO	066	99	PRT	103	43	RCL
030	14	14	067	42	STO	104	11	11
031	99	PRT	068	02	02	105	42	STO
032	53	(	069	98	ADV	106	02	02
033	43	RCL	070	53	(	107	43	RCL
034	03	03	071	43	RCL	108	03	03
035	75	-	072	14	14	109	81	RST
036	01	1	073	55	÷			

B.13 DETAILS OF VARIANCE-2 PROGRAM: COMPARISON OF TWO VARIANCES

Lines 0 to 12 Take the input of the x_i data and use the built-in function of KEY Σ+. (Erroneous data may be deleted with KEY Σ−.)

Lines 13 to 23 take the input of the second set of data, store the sum and the sum of squares in Registers 07 and 08, and count the data in Register 09.

Lines 24 to 40 obtain the variance estimate and the degrees of freedom from built-in functions, and print and store these results. (If the mean of the first set of data is wanted, instructions KEY $\bar{x}$ and PRINT may be inserted after line 25.)

Lines 41 to 53 shift the sums of the second set of data into the storage areas used by the built-in variance estimate routine.

Lines 54 to 69 use the built-in routine to obtain the variance estimate and degrees of freedom of the second set of data. These values are printed and stored. (If the mean of the second set of data is wanted, instructions KEY $\bar{x}$ and PRINT may be inserted between lines 53 and 54. Note that if instructions for the first set were changed to obtain the mean, the line numbers for the second set of data would have been changed.)

Lines 70 to 80 check the two variance estimates so that F is calculated from the ratio of the larger to the smaller. If the first is larger than the second, the calculation is made at

Lines 81 to 91 The value of F is printed and the data are arranged for calculation of the probability by the Variance-9 program.

Lines 92 to 109 calculate and print the F value if the second variance estimate is larger than the first. The results are then arranged in the proper sequence for calculation of the probability by the Variance-9 program.

Table B.14 Variance-3 Program: Bartlett χ^2 Test for Variance Homogeneity (TI-59)

Input: x_i, KEY A, . . . , KEY B
KEY C

Output: $s^2(x)$, ν, for each set
ν total, χ^2, probability

Input: ν_i, KEY $x \rightleftharpoons t$; $s^2(x_i)$, KEY D
KEY C

Output: ν total, χ^2, probability

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	039	43	RCL	078	23	LNX
001	76	LBL	040	08	08	079	86	STF
002	11	A	041	65	×	080	00	00
003	87	IFF	042	43	RCL	081	47	CMS
004	00	00	043	09	09	082	76	LBL
005	22	INV	044	54	)	083	23	LNX
006	47	CMS	045	44	SUM	084	42	STO
007	86	STF	046	12	12	085	08	08
008	00	00	047	53	(	086	32	X:T
009	76	LBL	048	43	RCL	087	61	GTO
010	22	INV	049	08	08	088	24	CE
011	78	Σ+	050	23	LNX	089	76	LBL
012	91	R/S	051	65	×	090	13	C
013	76	LBL	052	43	RCL	091	53	(
014	12	B	053	09	09	092	53	(
015	22	INV	054	54	)	093	53	(
016	79	$\bar{x}$	055	44	SUM	094	43	RCL
017	33	X^2	056	13	13	095	12	12
018	99	PRT	057	01	1	096	55	÷
019	42	STO	058	44	SUM	097	43	RCL
020	08	08	059	07	07	098	10	10
021	53	(	060	00	0	099	54	)
022	43	RCL	061	42	STO	100	23	LNX
023	03	03	062	01	01	101	65	×
024	75	-	063	42	STO	102	43	RCL
025	01	1	064	02	02	103	10	10
026	54	)	065	42	STO	104	75	-
027	99	PRT	066	03	03	105	43	RCL
028	98	ADV	067	42	STO	106	13	13
029	76	LBL	068	04	04	107	54	)
030	24	CE	069	42	STO	108	55	÷
031	42	STO	070	05	05	109	53	(
032	09	09	071	42	STO	110	53	(
033	44	SUM	072	06	06	111	01	1
034	10	10	073	91	R/S	112	85	+
035	35	1/X	074	76	LBL	113	53	(
036	44	SUM	075	14	D	114	53	(
037	11	11	076	87	IFF	115	53	(
038	53	(	077	00	00	116	43	RCL

Table B.14 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
117	07	07	161	53	(	205	42	STO
118	75	-	162	43	RCL	206	07	07
119	01	1	163	02	02	207	53	(
120	54	)	164	65	×	208	43	RCL
121	99	PRT	165	02	2	209	08	08
122	42	STO	166	54	)	210	94	+/-
123	01	01	167	34	√X	211	22	INV
124	54	)	168	32	X⇌T	212	23	LNX
125	65	×	169	43	RCL	213	49	PRD
126	03	3	170	07	07	214	07	07
127	54	)	171	77	GE	215	54	)
128	35	1/X	172	65	×	216	53	(
129	65	×	173	01	1	217	53	(
130	53	(	174	52	EE	218	43	RCL
131	43	RCL	175	94	+/-	219	03	03
132	11	11	176	05	5	220	85	+
133	75	-	177	99	PRT	221	02	2
134	43	RCL	178	81	RST	222	54	)
135	10	10	179	92	RTN	223	55	÷
136	35	1/X	180	76	LBL	224	02	2
137	54	)	181	65	×	225	54	)
138	54	)	182	00	0	226	42	STO
139	54	)	183	42	STO	227	04	04
140	54	)	184	06	06	228	53	(
141	99	PRT	185	01	1	229	43	RCL
142	42	STO	186	42	STO	230	04	04
143	02	02	187	05	05	231	75	-
144	43	RCL	188	53	(	232	01	1
145	01	01	189	53	(	233	54	)
146	53	(	190	43	RCL	234	42	STO
147	53	(	191	02	02	235	04	04
148	42	STO	192	55	÷	236	59	INT
149	03	03	193	02	2	237	32	X⇌T
150	65	×	194	54	)	238	43	RCL
151	02	2	195	42	STO	239	04	04
152	75	-	196	08	08	240	22	INV
153	01	1	197	45	Y^X	241	67	EQ
154	54	)	198	53	(	242	71	SBR
155	34	√X	199	43	RCL	243	76	LBL
156	85	+	200	03	03	244	75	-
157	05	5	201	55	÷	245	22	INV
158	54	)	202	02	2	246	49	PRD
159	42	STO	203	54	)	247	07	07
160	04	04	204	54	)	248	69	OP

Table B.14 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
249	34	34	276	89	π	303	06	06
250	00	0	277	34	√x	304	32	X⇌T
251	32	X⇌T	278	22	INV	305	01	1
252	43	RCL	279	49	PRD	306	52	EE
253	04	04	280	07	07	307	94	+/-
254	67	EQ	281	76	LBL	308	04	4
255	81	RST	282	81	RST	309	22	INV
256	61	GTO	283	02	2	310	77	GE
257	75	-	284	44	SUM	311	95	=
258	76	LBL	285	03	03	312	53	(
259	71	SBR	286	76	LBL	313	01	1
260	43	RCL	287	95	=	314	75	-
261	04	04	288	53	(	315	53	(
262	22	INV	289	43	RCL	316	43	RCL
263	49	PRD	290	02	02	317	06	06
264	07	07	291	55	÷	318	85	+
265	32	X⇌T	292	43	RCL	319	01	1
266	93	.	293	03	03	320	54	)
267	05	5	294	54	)	321	65	×
268	67	EQ	295	49	PRD	322	43	RCL
269	85	+	296	05	05	323	07	07
270	69	OP	297	02	2	324	54	)
271	34	34	298	44	SUM	325	22	INV
272	61	GTO	299	03	03	326	52	EE
273	71	SBR	300	43	RCL	327	99	PRT
274	76	LBL	301	05	05	328	81	RST
275	85	+	302	44	SUM	329	92	RTN

B.14 DETAILS OF VARIANCE-3 PROGRAM: BARTLETT χ^2 TEST

Lines 0 to 12 take the individual values for each set of data and accumulate the sum and sum of squares with KEY $\Sigma+$.

Lines 13 to 73 use the built-in functions to obtain the variance estimate for each set of data and then accumulate the sums required for the evaluation of Eq. (6.11) and (6.12). The storage areas used for the built-in functions are then cleared for the next set of data.

Lines 74 to 88 take the input data if the variance estimates and degrees of freedom are already available. The program is directed back to line 29 to accumulate the sums for the χ^2 calculation.

Lines 89 to 141 calculate χ^2 from Eqs. (6.11) and (6.12) and place the results in proper sequence for use of the χ^2-5 program to calculate the probability.

Lines 142 to 329 are the same as the χ^2-5 program for calculating the χ^2 probability.

Table B.15 Variance-4 Program: One-Factor Analysis of Variance, with Replicates (TI-59)

Input: x_i, KEY A for each group of data
KEY B after each group
KEY C after KEY B of the last group

Output: $\bar{x}$ for each group
Group degrees of freedom
Group variance estimate
Error degrees of freedom
Error variance estimate
F value

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	037	43	RCL	074	54	)
001	76	LBL	038	09	09	075	54	)
002	11	A	039	54	)	076	53	(
003	87	IFF	040	99	PRT	077	42	STO
004	00	00	041	98	ADV	078	11	11
005	22	INV	042	00	0	079	55	÷
006	47	CMS	043	42	STO	080	53	(
007	86	STF	044	09	09	081	43	RCL
008	00	00	045	42	STO	082	07	07
009	76	LBL	046	00	00	083	75	-
010	22	INV	047	91	R/S	084	01	1
011	44	SUM	048	76	LBL	085	54	)
012	00	00	049	13	C	086	99	PRT
013	78	Σ+	050	53	(	087	54	)
014	01	1	051	43	RCL	088	99	PRT
015	44	SUM	052	02	02	089	42	STO
016	09	09	053	75	-	090	01	01
017	91	R/S	054	43	RCL	091	98	ADV
018	76	LBL	055	01	01	092	53	(
019	12	B	056	33	X^2	093	53	(
020	01	1	057	55	÷	094	43	RCL
021	44	SUM	058	43	RCL	095	10	10
022	07	07	059	03	03	096	75	-
023	53	(	060	54	)	097	43	RCL
024	43	RCL	061	42	STO	098	11	11
025	00	00	062	10	10	099	54	)
026	33	X^2	063	53	(	100	55	÷
027	55	÷	064	43	RCL	101	53	(
028	43	RCL	065	08	08	102	43	RCL
029	09	09	066	75	-	103	03	03
030	54	)	067	53	(	104	75	-
031	44	SUM	068	43	RCL	105	43	RCL
032	08	08	069	01	01	106	07	07
033	53	(	070	33	X^2	107	54	)
034	43	RCL	071	55	÷	108	99	PRT
035	00	00	072	43	RCL	109	54	)
036	55	÷	073	03	03	110	99	PRT

Table B.15 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
111	42	STO	116	01	01	121	99	PRT
112	02	02	117	55	÷	122	81	RST
113	98	ADV	118	43	RCL	123	91	R/S
114	53	(	119	02	02			
115	43	RCL	120	54	)			

B.15 DETAILS OF THE VARIANCE-4 PROGRAM: ONE-FACTOR ANALYSIS OF VARIANCE

Lines 0 to 17 process the input data for each group of observations. The total sum and sum of squares of all the data are accumulated with the built-in functions of KEY $\Sigma+$. The group sum and the number in each group is calculated separately.

Lines 18 to 47 calculate the group total squared divided by the number in each group, for the group sum of squares. The mean for each group is calculated and printed. The groups are counted, and the registers used to obtain the group total are cleared.

Lines 48 to 62 calculate and store the error sum of squares.

Lines 63 to 88 calculate and print the group degrees of freedom and the group variance estimate, following the formulae in Table 6.4.

Lines 89 to 110 calculate the error degrees of freedom and the error variance estimate and print the values.

Lines 111 to 122 divide the group variance estimate by the error variance estimate for the F value and print the result.

The probability value associated with F could be obtained by making the Variance-9 program a part of the Variance-4 program, or by simply running Variance-9 after Variance-4 results are obtained.

Table B.16 Variance-5 Program: Two-Factor Analysis of Variance, No Replicates (TI-59)

Input: Number of rows *r*, KEY A
Data by columns, KEY A
When all data are entered, KEY B

Output: Row factor degrees of freedom
Row factor variance estimate
Row factor F value

Column factor degrees of freedom
Column factor variance estimate
Column factor F value

Error degrees of freedom
Error variance estimate

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	034	91	R/S	068	24	CE
001	76	LBL	035	76	LBL	069	61	GTO
002	11	A	036	23	LNX	070	12	B
003	87	IFF	037	43	RCL	071	76	LBL
004	00	00	038	08	08	072	24	CE
005	22	INV	039	33	X²	073	53	(
006	47	CMS	040	44	SUM	074	43	RCL
007	86	STF	041	07	07	075	02	02
008	00	00	042	00	0	076	75	-
009	42	STO	043	42	STO	077	53	(
010	00	00	044	08	08	078	43	RCL
011	42	STO	045	43	RCL	079	01	01
012	09	09	046	00	00	080	33	X²
013	01	1	047	44	SUM	081	55	÷
014	01	1	048	09	09	082	43	RCL
015	44	SUM	049	01	1	083	03	03
016	09	09	050	44	SUM	084	54	)
017	91	R/S	051	10	10	085	42	STO
018	76	LBL	052	91	R/S	086	02	02
019	22	INV	053	76	LBL	087	54	)
020	74	SM*	054	12	B	088	42	STO
021	09	09	055	73	RC*	089	01	01
022	44	SUM	056	09	09	090	53	(
023	08	08	057	33	X²	091	53	(
024	78	Σ+	058	44	SUM	092	43	RCL
025	69	OP	059	08	08	093	08	08
026	39	39	060	69	OP	094	55	÷
027	43	RCL	061	39	39	095	43	RCL
028	09	09	062	43	RCL	096	10	10
029	32	X⇌T	063	09	09	097	75	-
030	01	1	064	32	X⇌T	098	43	RCL
031	01	1	065	01	1	099	02	02
032	67	EQ	066	01	1	100	54	)
033	23	LNX	067	67	EQ	101	42	STO

Table B.16 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
102	04	04	135	42	STO	168	15	15
103	55	÷	136	12	12	169	53	(
104	53	(	137	54	)	170	43	RCL
105	43	RCL	138	42	STO	171	11	11
106	00	00	139	13	13	172	55	÷
107	75	-	140	53	(	173	43	RCL
108	01	1	141	53	(	174	15	15
109	54	)	142	43	RCL	175	54	)
110	99	PRT	143	01	01	176	99	PRT
111	54	)	144	75	-	177	98	ADV
112	99	PRT	145	43	RCL	178	43	RCL
113	42	STO	146	04	04	179	12	12
114	11	11	147	75	-	180	99	PRT
115	53	(	148	43	RCL	181	53	(
116	53	(	149	05	05	182	43	RCL
117	43	RCL	150	54	)	183	13	13
118	07	07	151	55	÷	184	99	PRT
119	55	÷	152	53	(	185	55	÷
120	43	RCL	153	43	RCL	186	43	RCL
121	00	00	154	03	03	187	15	15
122	75	-	155	75	-	188	54	)
123	43	RCL	156	43	RCL	189	99	PRT
124	02	02	157	00	00	190	98	ADV
125	54	)	158	75	-	191	43	RCL
126	42	STO	159	43	RCL	192	14	14
127	05	05	160	10	10	193	99	PRT
128	55	÷	161	85	+	194	43	RCL
129	53	(	162	01	1	195	15	15
130	43	RCL	163	54	)	196	99	PRT
131	10	10	164	42	STO	197	81	RST
132	75	-	165	14	14	198	91	R/S
133	01	1	166	54	)			
134	54	)	167	42	STO			

B.16 DETAILS OF THE VARIANCE-5 PROGRAM: TWO-FACTOR ANALYSIS OF VARIANCE

Lines 0 to 17 take the input of the number of rows and set the storage areas to receive the data.

Lines 18 to 34 take the input of one column of data and accumulate the sum for each row in a different register. The sum and sum of squares of all the data are accumulated with KEY $\Sigma+$.

Lines 35 to 52 calculate the square of the column total and accumulate the sum of the squares in Register 07. The indices are reset for the input of the next column data.

Lines 53 to 70 calculate the square of the row totals, sum the squares, and store the value in Register 08.

Lines 71 to 89 calculate $\Sigma x^2/n$ and store the value in Register 02 and calculate the total sum of squares and store the value in Register 01.

Lines 90 to 114 calculate the row degrees of freedom and the row variance estimate and print the results.

Lines 115 to 139 calculate the column variance estimate and the column degrees of freedom and store the results.

Lines 140 to 168 calculate and store the error variance estimate and the error degrees of freedom.

Lines 169 to 177 divide the row variance estimate by the error variance estimate to obtain the row F value.

Lines 178 to 190 print the column degrees of freedom, the column variance estimate, and divide the column variance estimate by the error variance estimate to obtain the column F value.

Lines 191 to 197 print the error degrees of freedom and the error variance estimate.

All of the calculations follow the formulae in Table 6.6.

If the column and/or row total or means are wanted as part of the output, minor changes to the program will produce these values. The column totals are recalled at line 38. The total could be printed, or the mean calculated by dividing by the number of rows. The row totals are recalled at line 56. These values could be printed or divided by the number of columns to obtain the row means. The number of columns is the total number of data entries, stored in Register 03, divided by the number of rows.

Table B.17 Variance-6 Program: Two-Factor Analysis of Variance with Replicates (TI-59)

Input: Number of rows *r*, KEY A; number of replicates *k*, KEY A;
Data by columns, KEY A
When all data are entered, KEY B

Output: Row degrees of freedom
Row variance estimate
Row *F* value

Column degrees of freedom
Column variance estimate
Column *F* value

Interaction degrees of freedom
Interaction variance estimate
Interaction *F* value

Error degrees of freedom
Error variance estimate

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	034	78	Σ+	068	44	SUM
001	76	LBL	035	01	1	069	13	13
002	11	A	036	44	SUM	070	43	RCL
003	87	IFF	037	11	11	071	13	13
004	00	00	038	43	RCL	072	32	X⇌T
005	22	INV	039	11	11	073	43	RCL
006	47	CMS	040	32	X⇌T	074	00	00
007	86	STF	041	43	RCL	075	67	EQ
008	00	00	042	10	10	076	25	CLR
009	42	STO	043	67	EQ	077	91	R/S
010	00	00	044	24	CE	078	76	LBL
011	42	STO	045	91	R/S	079	25	CLR
012	09	09	046	76	LBL	080	53	(
013	01	1	047	24	CE	081	43	RCL
014	05	5	048	53	(	082	08	08
015	44	SUM	049	43	RCL	083	33	X²
016	09	09	050	07	07	084	55	÷
017	91	R/S	051	74	SM*	085	53	(
018	76	LBL	052	09	09	086	43	RCL
019	22	INV	053	33	X²	087	00	00
020	87	IFF	054	55	÷	088	65	×
021	01	01	055	43	RCL	089	43	RCL
022	23	LNX	056	10	10	090	10	10
023	86	STF	057	54	)	091	54	)
024	01	01	058	44	SUM	092	54	)
025	42	STO	059	12	12	093	44	SUM
026	10	10	060	69	OP	094	14	14
027	91	R/S	061	39	39	095	00	0
028	76	LBL	062	00	0	096	42	STO
029	23	LNX	063	42	STO	097	13	13
030	44	SUM	064	07	07	098	42	STO
031	08	08	065	42	STO	099	08	08
032	44	SUM	066	11	11	100	43	RCL
033	07	07	067	01	1	101	00	00

Table B.17 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
102	44	SUM	146	14	14	190	75	-
103	09	09	147	53	(	191	43	RCL
104	91	R/S	148	43	RCL	192	01	01
105	76	LBL	149	03	03	193	54	)
106	12	B	150	55	÷	194	42	STO
107	00	0	151	43	RCL	195	04	04
108	42	STO	152	00	00	196	53	(
109	04	04	153	55	÷	197	53	(
110	53	(	154	43	RCL	198	43	RCL
111	43	RCL	155	10	10	199	04	04
112	01	01	156	54	)	200	55	÷
113	33	X^2	157	42	STO	201	53	(
114	55	÷	158	06	06	202	43	RCL
115	43	RCL	159	76	LBL	203	00	00
116	03	03	160	32	X⇌T	204	75	-
117	54	)	161	53	(	205	01	1
118	42	STO	162	73	RC*	206	54	)
119	01	01	163	09	09	207	99	PRT
120	53	(	164	33	X^2	208	54	)
121	43	RCL	165	55	÷	209	99	PRT
122	02	02	166	43	RCL	210	55	÷
123	75	-	167	06	06	211	53	(
124	43	RCL	168	55	÷	212	53	(
125	01	01	169	43	RCL	213	43	RCL
126	54	)	170	10	10	214	12	12
127	42	STO	171	54	)	215	75	-
128	02	02	172	44	SUM	216	43	RCL
129	53	(	173	04	04	217	04	04
130	43	RCL	174	69	OP	218	75	-
131	12	12	175	39	39	219	43	RCL
132	75	-	176	43	RCL	220	14	14
133	43	RCL	177	09	09	221	54	)
134	01	01	178	32	X⇌T	222	42	STO
135	54	)	179	01	1	223	05	05
136	42	STO	180	05	5	224	55	÷
137	12	12	181	67	EQ	225	53	(
138	53	(	182	33	X^2	226	53	(
139	43	RCL	183	61	GTO	227	43	RCL
140	14	14	184	32	X⇌T	228	00	00
141	75	-	185	76	LBL	229	75	-
142	43	RCL	186	33	X^2	230	01	1
143	01	01	187	53	(	231	54	)
144	54	)	188	43	RCL	232	65	×
145	42	STO	189	04	04	233	53	(

Table B.17 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
234	43	RCL	263	43	RCL	292	06	06
235	06	06	264	16	16	293	65	×
236	75	-	265	54	)	294	53	(
237	01	1	266	99	PRT	295	43	RCL
238	54	)	267	98	ADV	296	10	10
239	54	)	268	43	RCL	297	75	-
240	42	STO	269	15	15	298	01	1
241	15	15	270	99	PRT	299	54	)
242	54	)	271	43	RCL	300	54	)
243	42	STO	272	16	16	301	42	STO
244	16	16	273	53	(	302	18	18
245	54	)	274	99	PRT	303	54	)
246	99	PRT	275	55	÷	304	54	)
247	98	ADV	276	53	(	305	99	PRT
248	53	(	277	53	(	306	98	ADV
249	53	(	278	43	RCL	307	53	(
250	43	RCL	279	02	02	308	43	RCL
251	14	14	280	75	-	309	17	17
252	55	÷	281	43	RCL	310	55	÷
253	53	(	282	12	12	311	43	RCL
254	43	RCL	283	54	)	312	18	18
255	06	06	284	42	STO	313	99	PRT
256	75	-	285	17	17	314	54	)
257	01	1	286	55	÷	315	99	PRT
258	54	)	287	53	(	316	98	ADV
259	99	PRT	288	43	RCL	317	81	RST
260	54	)	289	00	00	318	91	R/S
261	99	PRT	290	65	×			
262	55	÷	291	43	RCL			

B.17 DETAILS OF THE VARIANCE-6 PROGRAM: TWO-FACTOR ANALYSIS OF VARIANCE WITH REPLICATES

Lines 0 to 17 take input of the number of rows, store the value, and set the registers for input of the column data.

Lines 18 to 27 take input of the number of replicates and store the value.

Lines 28 to 45 use KEY Σ+ to sum all of the data and obtain the replicate total in Register 07 and the column total in Register 08.

Lines 46 to 77 accumulate the replicate totals for each row in a different register, and accumulate the column total.

Lines 78 to 104 process the data when a complete column has been entered. The program counts the entries and goes to this portion of the program when $r \cdot k$ entries have been made. The square of the column total is divided by $r \cdot k$ and the sum accumulated in Register 14. The column total and replicate total registers are reset to zero, and the program returns to accept the next column's data.

Lines 105 to 158 start the calculation when all the data have been entered. The total sum of squares is calculated in lines 110 to 128. The sum of squares called "subtotal" in Table 6.7 is calculated in lines 129 to 137. The column sum of squares is calculated in lines 139 to 146. The product of rows $\times$ columns is calculated in lines 148 to 158.

Lines 159 to 184 recover the row totals and calculate Σ (row total)$^2/ck$ of Table 6.7.

Lines 185 to 195 calculate the row sum of squares.

Lines 196 to 209 calculate the row variance estimate and the row degrees of freedom and print the results.

Lines 210 to 247 calculate the interaction variance estimate and interaction degrees of freedom and store the results. This portion then divides the row variance estimate by the interaction variance estimate to obtain the row F value.

Lines 248 to 267 calculate and print the column degrees of freedom, variance estimate, and F value.

Lines 268 to 305 print the interaction degrees of freedom and the interaction variance estimate and then calculate the error variance estimate and error degrees of freedom. The interaction variance estimate is then divided by the error variance estimate to obtain the interaction F value.

Lines 306 to 318 print the error degrees of freedom and the error variance estimate.

All of the calculations follow the formulae in Table 6.7.

Table B.18 Variance-7 Program: Latin-Square Analysis of Variance (TI-59)

Input: Number of rows, KEY A; data by columns, KEY A; data by letters, KEY B

Output: Row degrees of freedom — Letter degrees of freedom
Row variance estimate — Letter variance estimate
Row F value — Letter F value

Column degrees of freedom — Error degrees of freedom
Column variance estimate — Error variance estimate
Column F value

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	035	76	LBL	070	33	X^2
001	76	LBL	036	23	LNX	071	55	÷
002	11	A	037	53	(	072	43	RCL
003	87	IFF	038	43	RCL	073	00	00
004	00	00	039	07	07	074	54	)
005	22	INV	040	33	X^2	075	44	SUM
006	47	CMS	041	55	÷	076	10	10
007	42	STO	042	43	RCL	077	00	0
008	00	00	043	00	00	078	42	STO
009	42	STO	044	54	)	079	05	05
010	09	09	045	44	SUM	080	72	ST*
011	86	STF	046	08	08	081	09	09
012	00	00	047	00	0	082	69	OP
013	01	1	048	42	STO	083	39	39
014	00	0	049	07	07	084	43	RCL
015	44	SUM	050	43	RCL	085	09	09
016	09	09	051	00	00	086	32	X:T
017	91	R/S	052	44	SUM	087	01	1
018	76	LBL	053	09	09	088	00	0
019	22	INV	054	91	R/S	089	67	EQ
020	74	SM*	055	76	LBL	090	25	CLR
021	09	09	056	12	B	091	61	GTO
022	44	SUM	057	87	IFF	092	43	RCL
023	07	07	058	01	01	093	76	LBL
024	78	Σ+	059	24	CE	094	25	CLR
025	69	OP	060	42	STO	095	86	STF
026	39	39	061	04	04	096	01	01
027	43	RCL	062	00	0	097	43	RCL
028	09	09	063	42	STO	098	00	00
029	32	X:T	064	06	06	099	44	SUM
030	01	1	065	76	LBL	100	09	09
031	00	0	066	43	RCL	101	43	RCL
032	67	EQ	067	53	(	102	04	04
033	23	LNX	068	73	RC*	103	76	LBL
034	91	R/S	069	09	09	104	24	CE

Table B.18 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
105	44	SUM	148	67	EQ	191	02	02
106	06	06	149	34	ΓX	192	75	-
107	74	SM*	150	99	PRT	193	43	RCL
108	09	09	151	32	X:T	194	01	01
109	01	1	152	99	PRT	195	54	)
110	44	SUM	153	91	R/S	196	42	STO
111	05	05	154	76	LBL	197	02	02
112	43	RCL	155	34	ΓX	198	53	(
113	05	05	156	53	(	199	43	RCL
114	32	X:T	157	73	RC*	200	05	05
115	43	RCL	158	09	09	201	75	-
116	00	00	159	33	X²	202	43	RCL
117	67	EQ	160	55	÷	203	01	01
118	32	X:T	161	43	RCL	204	54	)
119	91	R/S	162	00	00	205	42	STO
120	76	LBL	163	54	)	206	05	05
121	32	X:T	164	44	SUM	207	53	(
122	00	0	165	05	05	208	43	RCL
123	42	STO	166	69	OP	209	08	08
124	04	04	167	39	39	210	75	-
125	42	STO	168	43	RCL	211	43	RCL
126	05	05	169	09	09	212	01	01
127	69	OP	170	32	X:T	213	54	)
128	39	39	171	01	1	214	42	STO
129	43	RCL	172	00	0	215	08	08
130	09	09	173	67	EQ	216	53	(
131	32	X:T	174	35	1/X	217	43	RCL
132	01	1	175	61	GTO	218	10	10
133	00	0	176	34	ΓX	219	75	-
134	67	EQ	177	76	LBL	220	43	RCL
135	33	X²	178	35	1/X	221	01	01
136	91	R/S	179	53	(	222	54	)
137	76	LBL	180	43	RCL	223	42	STO
138	33	X²	181	01	01	224	10	10
139	43	RCL	182	33	X²	225	53	(
140	00	00	183	55	÷	226	43	RCL
141	44	SUM	184	43	RCL	227	02	02
142	09	09	185	03	03	228	75	-
143	43	RCL	186	54	)	229	43	RCL
144	06	06	187	42	STO	230	05	05
145	32	X:T	188	01	01	231	75	-
146	43	RCL	189	53	(	232	43	RCL
147	01	01	190	43	RCL	233	08	08

Table B.18 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
234	75	-	260	07	07	286	06	06
235	43	RCL	261	54	)	287	99	PRT
236	10	10	262	42	STO	288	98	ADV
237	54	)	263	06	06	289	81	RST
238	42	STO	264	53	(	290	91	R/S
239	06	06	265	53	(	291	76	LBL
240	53	(	266	43	RCL	292	42	STO
241	43	RCL	267	10	10	293	55	÷
242	03	03	268	71	SBR	294	53	(
243	75	-	269	42	STO	295	43	RCL
244	53	(	270	53	(	296	00	00
245	03	3	271	53	(	297	75	-
246	65	×	272	43	RCL	298	01	1
247	43	RCL	273	08	08	299	54	)
248	00	00	274	71	SBR	300	99	PRT
249	54	)	275	42	STO	301	54	)
250	85	+	276	53	(	302	99	PRT
251	02	2	277	53	(	303	55	÷
252	54	)	278	43	RCL	304	43	RCL
253	42	STO	279	05	05	305	06	06
254	07	07	280	71	SBR	306	54	)
255	53	(	281	42	STO	307	99	PRT
256	43	RCL	282	43	RCL	308	98	ADV
257	06	06	283	07	07	309	92	RTN
258	55	÷	284	99	PRT			
259	43	RCL	285	43	RCL			

B.18 DETAILS OF THE VARIANCE-7 PROGRAM: LATIN-SQUARE ANALYSIS OF VARIANCE

Lines 0 to 17 take the input of the number of rows and store the value. They then set the storage registers to calculate the row totals.

Lines 18 to 34 take the data for each column, store each successive value in a different register to accumulate the row totals, and accumulate the column total in Register 07. The total data are summed with KEY $\Sigma+$.

Lines 35 to 54 are utilized when one column of data has been entered. The program keeps track of the number of entries and when the number for one column has been entered, the column total is squared and divided by the number in the column. The sum is accumulated in Register 08. The registers used for the column total are then reset to zero and the program returns for the next column of data.

Lines 55 to 64 store the first entry for the letter data.

Lines 65 to 102 calculate the row sum of squares before the processing of the letter data to free the storage areas for the latter calculation. Lines 67 to 76 calculate Σ (row total)$^2/k$ and accumulate the sum in Register 10. Lines 82 to 89 check whether all the rows have been included in the calculation. If not, the program returns to line 65; if so, the program goes on to line 93. Lines 93 to 102 clear the registers used for the row totals so they may be used for the letter totals.

Lines 103 to 119 accumulate the letter totals in one of the registers assigned by the index in Register 09, depending on the original number of rows.

Lines 120 to 136 change the index value in Register 09 when one-letter data have been entered. The count is made in Register 05.

Lines 137 to 153 check the total entered for the letter factors against the total entered for the column data. If the totals are different, the program prints the two totals and waits. The operator may then proceed with KEY R/S if the difference is small, or may start over. If the totals are the same, the program proceeds to line 154.

Lines 154 to 176 calculate the Σ (letter total)$^2/k$.

Lines 177 to 197 calculate the total sum of squares.

Lines 198 to 224 calculate the sum of squares for the letter, the column, and the row factors: 198 to 206 for the letters, 207 to 215 for the columns, and 216 to 224 for the rows.

Lines 225 to 263 calculate the error sum of squares, degrees of freedom, and variance estimate.

Lines 264 to 281 calculate the variance estimate for the row factor, the column factor, and the letter factor. All the calculations are done in the subroutine at line 291 which divide the factor sum of squares by the factor degrees of freedom, which are all equal to k-1, and then divide by the error variance estimate to obtain the factor F value.

Lines 282 to 290 print the error variance estimate and the error degrees of freedom.

Lines 291 to 309 are the subroutine to calculate the factor variance estimates and F values and to print the results.

All of the calculations follow the formulae in Table 6.9.

Table B.19 Variance-8 Program: 2^n Factorial Analysis of Variance (TI-59)

Input: 2^n (either 4, 8, or 16), KEY A
Data in standard order: (1), *a*, *b*, *ab*, *c*, *ac*, . . . , KEY A

Output: Variance estimates in order of data input
Mean differences in order of data input

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	038	67	EQ	076	44	SUM
001	76	LBL	039	23	LNX	077	01	01
002	11	A	040	91	R/S	078	76	LBL
003	87	IFF	041	76	LBL	079	32	X:T
004	00	00	042	23	LNX	080	53	(
005	22	INV	043	43	RCL	081	73	RC*
006	47	CMS	044	00	00	082	01	01
007	53	(	045	44	SUM	083	94	+/-
008	42	STO	046	01	01	084	69	OP
009	00	00	047	76	LBL	085	31	31
010	85	+	048	24	CE	086	85	+
011	03	3	049	53	(	087	73	RC*
012	54	)	050	73	RC*	088	01	01
013	42	STO	051	01	01	089	54	)
014	01	01	052	69	OP	090	72	ST*
015	53	(	053	31	31	091	02	02
016	43	RCL	054	85	+	092	69	OP
017	00	00	055	73	RC*	093	32	32
018	65	×	056	01	01	094	69	OP
019	02	2	057	54	)	095	31	31
020	85	+	058	72	ST*	096	43	RCL
021	03	3	059	02	02	097	01	01
022	54	)	060	69	OP	098	32	X:T
023	42	STO	061	32	32	099	03	3
024	02	02	062	69	OP	100	67	EQ
025	86	STF	063	31	31	101	33	X²
026	00	00	064	43	RCL	102	61	GTO
027	91	R/S	065	01	01	103	32	X:T
028	76	LBL	066	32	X:T	104	76	LBL
029	22	INV	067	03	3	105	33	X²
030	72	ST*	068	67	EQ	106	43	RCL
031	01	01	069	25	CLR	107	00	00
032	69	OP	070	61	GTO	108	44	SUM
033	31	31	071	24	CE	109	01	01
034	43	RCL	072	76	LBL	110	44	SUM
035	01	01	073	25	CLR	111	02	02
036	32	X:T	074	43	RCL	112	76	LBL
037	03	3	075	00	00	113	34	√X

Table B.19 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
114	73	RC*	153	00	00	192	44	SUM
115	02	02	154	44	SUM	193	98	ADV
116	72	ST*	155	01	01	194	43	RCL
117	01	01	156	44	SUM	195	00	00
118	69	OP	157	02	02	196	44	SUM
119	32	32	158	61	GTO	197	01	01
120	69	OP	159	24	CE	198	73	RC*
121	31	31	160	76	LBL	199	01	01
122	43	RCL	161	42	STO	200	69	OP
123	01	01	162	43	RCL	201	31	31
124	32	X⇄T	163	00	00	202	76	LBL
125	03	3	164	44	SUM	203	45	Y^X
126	67	EQ	165	01	01	204	53	(
127	35	1/X	166	73	RC*	205	73	RC*
128	61	GTO	167	01	01	206	01	01
129	34	√X	168	69	OP	207	55	÷
130	76	LBL	169	31	31	208	43	RCL
131	35	1/X	170	76	LBL	209	00	00
132	01	1	171	43	RCL	210	65	×
133	44	SUM	172	53	(	211	02	2
134	03	03	173	73	RC*	212	54	)
135	53	(	174	01	01	213	99	PRT
136	43	RCL	175	33	X^2	214	69	OP
137	00	00	176	55	÷	215	31	31
138	23	LNX	177	43	RCL	216	43	RCL
139	55	÷	178	00	00	217	01	01
140	02	2	179	54	)	218	32	X⇄T
141	23	LNX	180	99	PRT	219	03	3
142	54	)	181	69	OP	220	67	EQ
143	52	EE	182	31	31	221	52	EE
144	22	INV	183	43	RCL	222	61	GTO
145	52	EE	184	01	01	223	45	Y^X
146	59	INT	185	32	X⇄T	224	76	LBL
147	32	X⇄T	186	03	3	225	52	EE
148	43	RCL	187	67	EQ	226	98	ADV
149	03	03	188	44	SUM	227	81	RST
150	67	EQ	189	61	GTO	228	91	R/S
151	42	STO	190	43	RCL			
152	43	RCL	191	76	LBL			

B.19 DETAILS OF THE VARIANCE-8 PROGRAM: 2^n FACTORIAL CALCULATION

Lines 0 to 27 take the input of the size of the experiment, 2^n, and set the storage area size to receive the balance of the data.

Lines 28 to 40 store the experimental data in registers assigned by the index of Register 01. When all of the data are entered, as indicated by the value of Register 01, the program goes on to line 41.

Lines 41 to 46 reset the index register.

Lines 47 to 71 sum the data in pairs and store the results in registers assigned by the index of Register 02. When all the pairs have been added, the program goes on to line 72.

Lines 72 to 103 subtract the first of each pair of data from the second and store the difference. When each pair has been differenced, the program goes on to line 104.

Lines 104 to 111 reset the index registers.

Lines 112 to 129 place the sums and differences in the original registers.

Lines 130 to 159 check whether the sums and differences have been calculated n times. If so, the calculation is completed and the program goes on to line 160. If not, the indices are reset and the program returns to line 47 for another calculation of the sums and differences by pairs.

Lines 160 to 190 calculate the variance estimate of each term of the factorial by squaring the last sum and difference calculation and dividing by 2^n. The results are printed.

Lines 191 to 227 calculate the mean difference of each term of the factorial by dividing the last sum and difference calculation by 2^{n-1}. These results are printed.

Table B.20 Variance-9 Program: *F* Probability Calculation (TI-59)

Input: ν, numerator degrees of freedom, KEY A; ν_2 denominator degrees of freedom, KEY A; F, KEY A

Output: Probability associated with *F* at ν_1 and ν_2 degrees of freedom

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	039	03	03	078	43	RCL
001	76	LBL	040	54	)	079	02	02
002	11	A	041	54	)	080	32	X:T
003	87	IFF	042	54	)	081	43	RCL
004	00	00	043	42	STO	082	01	01
005	22	INV	044	04	04	083	42	STO
006	86	STF	045	29	CP	084	02	02
007	00	00	046	53	(	085	32	X:T
008	47	CMS	047	43	RCL	086	42	STO
009	42	STO	048	01	01	087	01	01
010	01	01	049	55	÷	088	76	LBL
011	91	R/S	050	02	2	089	24	CE
012	76	LBL	051	54	)	090	53	(
013	22	INV	052	22	INV	091	53	(
014	87	IFF	053	59	INT	092	43	RCL
015	01	01	054	67	EQ	093	01	01
016	23	LNX	055	24	CE	094	75	-
017	86	STF	056	22	INV	095	02	2
018	01	01	057	86	STF	096	71	SBR
019	42	STO	058	01	01	097	95	=
020	02	02	059	53	(	098	53	(
021	91	R/S	060	43	RCL	099	01	1
022	76	LBL	061	02	02	100	75	-
023	23	LNX	062	55	÷	101	43	RCL
024	42	STO	063	02	2	102	04	04
025	03	03	064	54	)	103	54	)
026	53	(	065	22	INV	104	42	STO
027	43	RCL	066	59	INT	105	06	06
028	02	02	067	22	INV	106	43	RCL
029	55	÷	068	67	EQ	107	02	02
030	53	(	069	25	CLR	108	42	STO
031	43	RCL	070	53	(	109	08	08
032	02	02	071	01	1	110	02	2
033	85	+	072	75	-	111	42	STO
034	53	(	073	43	RCL	112	11	11
035	43	RCL	074	04	04	113	01	1
036	01	01	075	54	)	114	42	STO
037	65	×	076	42	STO	115	07	07
038	43	RCL	077	04	04	116	42	STO

Table B.20 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
117	13	13	160	02	2	203	53	(
118	53	(	161	85	+	204	94	+/-
119	43	RCL	162	43	RCL	205	85	+
120	04	04	163	00	00	206	01	1
121	45	Y^X	164	54	)	207	54	)
122	53	(	165	54	)	208	76	LBL
123	43	RCL	166	65	×	209	35	1/X
124	02	02	167	43	RCL	210	99	PRT
125	55	÷	168	06	06	211	98	ADV
126	02	2	169	54	)	212	81	RST
127	54	)	170	49	PRD	213	91	R/S
128	54	)	171	07	07	214	76	LBL
129	42	STO	172	01	1	215	25	CLR
130	05	05	173	44	SUM	216	86	STF
131	43	RCL	174	07	07	217	03	03
132	01	01	175	97	DSZ	218	53	(
133	32	X:T	176	09	09	219	43	RCL
134	02	2	177	33	X^2	220	01	01
135	67	EQ	178	76	LBL	221	55	÷
136	32	X:T	179	32	X:T	222	43	RCL
137	76	LBL	180	53	(	223	02	02
138	33	X^2	181	53	(	224	65	×
139	53	(	182	53	(	225	43	RCL
140	53	(	183	43	RCL	226	03	03
141	53	(	184	07	07	227	54	)
142	43	RCL	185	65	×	228	34	$\sqrt{X}$
143	08	08	186	43	RCL	229	70	RAD
144	85	+	187	13	13	230	22	INV
145	53	(	188	54	)	231	30	TAN
146	43	RCL	189	85	+	232	42	STO
147	09	09	190	43	RCL	233	12	12
148	65	×	191	10	10	234	42	STO
149	02	2	192	54	)	235	04	04
150	54	)	193	65	×	236	42	STO
151	75	-	194	43	RCL	237	10	10
152	43	RCL	195	05	05	238	39	COS
153	11	11	196	54	)	239	33	X^2
154	54	)	197	87	IFF	240	42	STO
155	55	÷	198	03	03	241	06	06
156	53	(	199	34	$\sqrt{X}$	242	53	(
157	43	RCL	200	87	IFF	243	43	RCL
158	09	09	201	01	01	244	04	04
159	65	×	202	35	1/X	245	38	SIN

Table B.20 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
246	65	×	289	05	05	332	42	STO
247	43	RCL	290	54	)	333	08	08
248	04	04	291	76	LBL	334	75	−
249	39	COS	292	34	√x	335	01	1
250	54	)	293	87	IFF	336	54	)
251	42	STO	294	04	04	337	42	STO
252	13	13	295	42	STO	338	09	09
253	53	(	296	42	STO	339	76	LBL
254	02	2	297	14	14	340	44	SUM
255	55	÷	298	86	STF	341	53	(
256	89	π	299	04	04	342	43	RCL
257	54	)	300	43	RCL	343	09	09
258	42	STO	301	04	04	344	55	÷
259	05	05	302	38	SIN	345	53	(
260	53	(	303	33	x²	346	43	RCL
261	53	(	304	42	STO	347	09	09
262	43	RCL	305	06	06	348	75	−
263	02	02	306	00	0	349	01	1
264	75	−	307	42	STO	350	54	)
265	03	3	308	10	10	351	54	)
266	71	SBR	309	01	1	352	49	PRD
267	95	=	310	42	STO	353	11	11
268	00	0	311	12	12	354	97	DSZ
269	42	STO	312	42	STO	355	09	09
270	08	08	313	13	13	356	45	yˣ
271	42	STO	314	42	STO	357	76	LBL
272	11	11	315	07	07	358	45	yˣ
273	01	1	316	42	STO	359	97	DSZ
274	42	STO	317	11	11	360	09	09
275	00	00	318	32	x⇌t	361	44	SUM
276	42	STO	319	43	RCL	362	43	RCL
277	07	07	320	01	01	363	11	11
278	32	x⇌t	321	22	INV	364	49	PRD
279	43	RCL	322	67	EQ	365	05	05
280	02	02	323	43	RCL	366	43	RCL
281	22	INV	324	00	0	367	04	04
282	67	EQ	325	61	GTO	368	38	SIN
283	33	x²	326	42	STO	369	49	PRD
284	53	(	327	76	LBL	370	05	05
285	43	RCL	328	43	RCL	371	53	(
286	10	10	329	53	(	372	43	RCL
287	65	×	330	43	RCL	373	04	04
288	43	RCL	331	02	02	374	39	COS

Table B.20 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
375	45	YX	388	75	-	401	54	)
376	43	RCL	389	03	3	402	61	GTO
377	02	02	390	71	SBR	403	35	1/X
378	54	)	391	95	=	404	76	LBL
379	49	PRD	392	61	GTO	405	95	=
380	05	05	393	33	X²	406	54	)
381	01	1	394	76	LBL	407	55	÷
382	42	STO	395	42	STO	408	02	2
383	11	11	396	85	+	409	54	)
384	53	(	397	01	1	410	42	STO
385	53	(	398	75	-	411	09	09
386	43	RCL	399	43	RCL	412	92	RTN
387	01	01	400	14	14			

B.20 DETAILS OF THE VARIANCE-9 PROGRAM: *F* PROBABILITY

Lines 0 to 44 take the input data: ν_1, ν_2, and F. They store these values and calculate x of Table 6.16.

Lines 45 to 55 check whether ν_1 is even or odd. If even, the program goes to line 88. If odd, the program goes on to line 56.

Lines 56 to 69 check whether ν_2 is even or odd. If odd, the program goes to line 214. If even, the program goes on to line 70.

Lines 70 to 87 substitute $(1 - x)$ for x and interchange ν_1 and ν_2.

Lines 88 to 136 start the calculation of the probability if either ν_1 or ν_2 is even. The expressions in square brackets in Eqs. (6.32), (6.33), and (6.34) are all similar and may be evaluated by the same summation program. The evaluation of the probability involves obtaining the value of the expression in the square brackets and multiplying it by different terms depending on the value of the degrees of freedom. These lines in the program set the values for the calculation when either ν_1 or ν_2 is even.

Lines 137 to 177 evaluate the square bracket expressions for all three equations: (6.32), (6.33), and (6.34).

Lines 178 to 207 multiply the square bracket term by the terms shown in Eq. (6.25) or (6.26), depending on whether ν_1 or ν_2 is even.

Lines 208 to 213 print the results and clear all flags for another calculation.

Lines 214 to 283 start the calculation if both degrees of freedom are odd; θ of Table 6.16 is calculated, and the values are set for the evaluation of the square bracket term for the A solution to the two-part calculation involved for odd degrees of freedom. The program returns to line 137 for the evaluation.

Lines 284 to 338 evaluate the terms of the square bracket solution for the B part of the calculation.

Lines 339 to 361 calculate the factorial expression of Eq. (6.35).

Lines 362 to 391 calculate the terms outside the square brackets of Eq. (6.30).

Lines 392 to 393 direct the program to line 137 for the solution of the square bracket expression for the B solution.

Lines 394 to 403 carry out the calculation of Eq. (6.28) and direct the program back to line 208 to print the result.

Lines 404 to 412 are a calculation that is used several times and is in the program as a subroutine to conserve program memory.

Table B.21 Regression-1 Program: Straight-Line Linear Correlation (TI-59)

Equation: $y = b_0 + b_1 \cdot x$

Input: Experimental data—
y_i, KEY $x \rightleftarrows t$, x_i, KEY A
When all data are in, KEY C

Output: $\bar{x}$, $\bar{y}$,
$(\Sigma x^2 - (\Sigma x)^2/n)$, b_1, b_0,
r, probability of r, $s(\hat{y})$

Input: Correlation results—
Degrees of freedom, ν, KEY $x \rightleftarrows t$
Correlation coefficient r, KEY D

Output: Probability value

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	035	42	STO	070	55	÷
001	76	LBL	036	11	11	071	43	RCL
002	11	A	037	53	(	072	03	03
003	87	IFF	038	43	RCL	073	54	)
004	00	00	039	02	02	074	99	PRT
005	22	INV	040	75	-	075	53	(
006	47	CMS	041	43	RCL	076	43	RCL
007	86	STF	042	01	01	077	03	03
008	00	00	043	33	X²	078	75	-
009	76	LBL	044	55	÷	079	02	2
010	22	INV	045	43	RCL	080	54	)
011	78	Σ+	046	03	03	081	32	X⇌T
012	91	R/S	047	54	)	082	53	(
013	76	LBL	048	42	STO	083	53	(
014	13	C	049	12	12	084	43	RCL
015	86	STF	050	99	PRT	085	13	13
016	00	00	051	53	(	086	65	×
017	79	x̄	052	53	(	087	43	RCL
018	99	PRT	053	43	RCL	088	11	11
019	32	X⇌T	054	04	04	089	54	)
020	99	PRT	055	75	-	090	42	STO
021	98	ADV	056	53	(	091	16	16
022	53	(	057	43	RCL	092	55	÷
023	43	RCL	058	11	11	093	53	(
024	06	06	059	55	÷	094	43	RCL
025	75	-	060	43	RCL	095	05	05
026	43	RCL	061	12	12	096	75	-
027	01	01	062	54	)	097	53	(
028	65	×	063	42	STO	098	43	RCL
029	43	RCL	064	13	13	099	04	04
030	04	04	065	99	PRT	100	33	X²
031	55	÷	066	65	×	101	55	÷
032	43	RCL	067	43	RCL	102	43	RCL
033	03	03	068	01	01	103	03	03
034	54	)	069	54	)	104	54	)

Table B.21 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
105	54	)	148	54	)	191	52	EE
106	42	STO	149	42	STO	192	76	LBL
107	07	07	150	09	09	193	52	EE
108	54	)	151	01	1	194	97	DSZ
109	34	√X	152	42	STO	195	09	09
110	99	PRT	153	08	08	196	35	1/X
111	76	LBL	154	76	LBL	197	53	(
112	14	D	155	35	1/X	198	43	RCL
113	53	(	156	43	RCL	199	08	08
114	42	STO	157	09	09	200	65	×
115	15	15	158	32	X⇌T	201	43	RCL
116	55	÷	159	01	1	202	11	11
117	53	(	160	67	EQ	203	54	)
118	01	1	161	44	SUM	204	61	GTO
119	75	-	162	53	(	205	25	CLR
120	43	RCL	163	53	(	206	76	LBL
121	15	15	164	43	RCL	207	44	SUM
122	33	X²	165	08	08	208	53	(
123	54	)	166	65	×	209	53	(
124	34	√X	167	53	(	210	53	(
125	54	)	168	43	RCL	211	43	RCL
126	70	RAD	169	10	10	212	08	08
127	22	INV	170	39	COS	213	65	×
128	30	TAN	171	33	X²	214	43	RCL
129	42	STO	172	55	÷	215	11	11
130	10	10	173	43	RCL	216	65	×
131	38	SIN	174	09	09	217	43	RCL
132	42	STO	175	54	)	218	10	10
133	11	11	176	65	×	219	39	COS
134	02	2	177	53	(	220	54	)
135	32	X⇌T	178	43	RCL	221	85	+
136	42	STO	179	09	09	222	43	RCL
137	00	00	180	75	-	223	10	10
138	22	INV	181	01	1	224	54	)
139	77	GE	182	54	)	225	54	)
140	42	STO	183	54	)	226	42	STO
141	67	EQ	184	85	+	227	14	14
142	43	RCL	185	01	1	228	61	GTO
143	53	(	186	54	)	229	65	×
144	43	RCL	187	42	STO	230	76	LBL
145	00	00	188	08	08	231	42	STO
146	75	-	189	97	DSZ	232	43	RCL
147	02	2	190	09	09	233	11	11

Table B.21 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
234	42	STO	253	91	R/S	272	07	07
235	14	14	254	76	LBL	273	75	-
236	61	GTO	255	65	×	274	43	RCL
237	65	×	256	53	(	275	16	16
238	76	LBL	257	43	RCL	276	54	)
239	43	RCL	258	14	14	277	55	÷
240	43	RCL	259	65	×	278	53	(
241	11	11	260	02	2	279	43	RCL
242	76	LBL	261	55	÷	280	03	03
243	25	CLR	262	89	π	281	75	-
244	94	+/-	263	54	)	282	02	2
245	85	+	264	61	GTO	283	54	)
246	01	1	265	25	CLR	284	54	)
247	95	=	266	91	R/S	285	34	√X
248	99	PRT	267	76	LBL	286	99	PRT
249	87	IFF	268	95	=	287	81	RST
250	00	00	269	53	(	288	91	R/S
251	95	=	270	53	(			
252	81	RST	271	43	RCL			

B.21 DETAILS OF THE REGRESSION-1 PROGRAM: STRAIGHT-LINE LINEAR CORRELATION

Lines 0 to 12 take the input y-x data. The dependent variable y is entered first. The necessary sums and sums of squares are accumulated with the built-in functions of KEY $\Sigma+$. If incorrect data are entered, they may be deleted by reentering them and using KEY $\Sigma-$.

Lines 13 to 21 print the values of $\bar{x}$ and $\bar{y}$ by using the built-in functions for obtaining these values.

Lines 22 to 50 calculate the values to obtain b_1 from Eq. (7.5); $\Sigma\ x^2 - (\Sigma x)^2/n$ is printed.

Lines 51 to 65 calculate and print b_1 and start the calculation of b_0 from Eq. (7.6).

Lines 66 to 74 complete the calculation and print b_0.

Lines 75 to 81 calculate the degrees of freedom and store the value in the t register for later use in the probability calculation.

Lines 82 to 110 calculate the correlation coefficient from Eq. (7.8) and print the result. The degrees of freedom and the correlation coefficient are now in the proper sequence for the probability calculation.

Lines 111 to 266 are the probability calculation by the same program as t-3 for the t probability calculation. Lines 113 to 125 convert the correlation coefficient to the equivalent t value by Eq. (7.9).

Lines 267 to 287 calculate $s(\hat{y})$, following Eq. (7.10), and print the result. This portion of the program is not used if the input data are the number of degrees of freedom and the correlation coefficient rather than actual y-x data. Flag 0, set at line 15, directs the program to this routine if it is needed.

Table B.22 Regression-2 Program: Linear Correlations Other Than Straight Lines (TI-59)

Equation: $y = b_0 + b_1 \cdot F(x)$

Input: Experimental data:
y_i, KEY $x \rightleftharpoons t$, x_i, KEY A
When all data are in, KEY C

Output: b_1, b_0, r, probability of r, $s(\hat{y})$ for each function

F functions:
x line 17
x^2 lines 20 and 29
$\sqrt{x}$ lines 38 and 47
e^x lines 56/57 and 66/67
$1/x$ lines 76 and 84
log (x) lines 93 and 102

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	034	44	SUM	068	65	×
001	76	LBL	035	09	09	069	43	RCL
002	11	A	036	43	RCL	070	23	23
003	87	IFF	037	22	22	071	54	)
004	00	00	038	34	√X	072	44	SUM
005	22	INV	039	44	SUM	073	15	15
006	47	CMS	040	10	10	074	43	RCL
007	86	STF	041	33	X²	075	22	22
008	00	00	042	44	SUM	076	35	1/X
009	76	LBL	043	11	11	077	44	SUM
010	22	INV	044	53	(	078	16	16
011	42	STO	045	43	RCL	079	33	X²
012	22	22	046	22	22	080	44	SUM
013	32	X⇌T	047	34	√X	081	17	17
014	42	STO	048	65	×	082	43	RCL
015	23	23	049	43	RCL	083	22	22
016	32	X⇌T	050	23	23	084	35	1/X
017	78	Σ+	051	54	)	085	65	×
018	43	RCL	052	44	SUM	086	43	RCL
019	22	22	053	12	12	087	23	23
020	33	X²	054	43	RCL	088	54	)
021	44	SUM	055	22	22	089	44	SUM
022	07	07	056	22	INV	090	18	18
023	33	X²	057	23	LNX	091	43	RCL
024	44	SUM	058	44	SUM	092	22	22
025	08	08	059	13	13	093	28	LOG
026	53	(	060	33	X²	094	44	SUM
027	43	RCL	061	44	SUM	095	19	19
028	22	22	062	14	14	096	33	X²
029	33	X²	063	53	(	097	44	SUM
030	65	×	064	43	RCL	098	20	20
031	43	RCL	065	22	22	099	53	(
032	23	23	066	22	INV	100	43	RCL
033	54	)	067	23	LNX	101	22	22

Table B.22 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
102	28	LOG	145	01	01	188	53	(
103	65	×	146	43	RCL	189	61	GTO
104	43	RCL	147	14	14	190	25	CLR
105	23	23	148	42	STO	191	55	÷
106	54	)	149	02	02	192	53	(
107	44	SUM	150	43	RCL	193	43	RCL
108	21	21	151	15	15	194	02	02
109	91	R/S	152	42	STO	195	75	-
110	76	LBL	153	06	06	196	53	(
111	13	C	154	71	SBR	197	43	RCL
112	71	SBR	155	23	LNX	198	01	01
113	23	LNX	156	43	RCL	199	33	X²
114	43	RCL	157	16	16	200	55	÷
115	07	07	158	42	STO	201	43	RCL
116	42	STO	159	01	01	202	03	03
117	01	01	160	43	RCL	203	54	)
118	43	RCL	161	17	17	204	54	)
119	08	08	162	42	STO	205	54	)
120	42	STO	163	02	02	206	42	STO
121	02	02	164	43	RCL	207	00	00
122	43	RCL	165	18	18	208	99	PRT
123	09	09	166	42	STO	209	53	(
124	42	STO	167	06	06	210	53	(
125	06	06	168	71	SBR	211	43	RCL
126	71	SBR	169	23	LNX	212	04	04
127	23	LNX	170	43	RCL	213	75	-
128	43	RCL	171	19	19	214	43	RCL
129	10	10	172	42	STO	215	00	00
130	42	STO	173	01	01	216	65	×
131	01	01	174	43	RCL	217	43	RCL
132	43	RCL	175	20	20	218	01	01
133	11	11	176	42	STO	219	54	)
134	42	STO	177	02	02	220	55	÷
135	02	02	178	43	RCL	221	43	RCL
136	43	RCL	179	21	21	222	03	03
137	12	12	180	42	STO	223	54	)
138	42	STO	181	06	06	224	99	PRT
139	06	06	182	71	SBR	225	53	(
140	71	SBR	183	23	LNX	226	53	(
141	23	LNX	184	81	RST	227	71	SBR
142	43	RCL	185	91	R/S	228	25	CLR
143	13	13	186	76	LBL	229	75	-
144	42	STO	187	23	LNX	230	53	(

Table B.22 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
231	43	RCL	274	75	-	317	53	(
232	00	00	275	43	RCL	318	53	(
233	65	×	276	00	00	319	43	RCL
234	71	SBR	277	33	X^2	320	00	00
235	24	CE	278	54	)	321	65	×
236	54	)	279	34	√X	322	53	(
237	54	)	280	54	)	323	43	RCL
238	55	÷	281	70	RAD	324	01	01
239	53	(	282	22	INV	325	39	COS
240	43	RCL	283	30	TAN	326	33	X^2
241	03	03	284	42	STO	327	55	÷
242	75	-	285	01	01	328	43	RCL
243	02	2	286	38	SIN	329	06	06
244	54	)	287	42	STO	330	54	)
245	54	)	288	02	02	331	65	×
246	34	√X	289	02	2	332	53	(
247	42	STO	290	32	X⇌T	333	43	RCL
248	24	24	291	42	STO	334	06	06
249	53	(	292	00	00	335	75	-
250	43	RCL	293	22	INV	336	01	1
251	03	03	294	77	GE	337	54	)
252	75	-	295	42	STO	338	54	)
253	02	2	296	67	EQ	339	85	+
254	54	)	297	43	RCL	340	01	1
255	32	X⇌T	298	53	(	341	54	)
256	53	(	299	43	RCL	342	42	STO
257	43	RCL	300	00	00	343	00	00
258	00	00	301	75	-	344	97	DSZ
259	65	×	302	02	2	345	06	06
260	71	SBR	303	54	)	346	52	EE
261	24	CE	304	42	STO	347	76	LBL
262	55	÷	305	06	06	348	52	EE
263	71	SBR	306	01	1	349	97	DSZ
264	25	CLR	307	42	STO	350	06	06
265	54	)	308	00	00	351	35	1/X
266	34	√X	309	76	LBL	352	53	(
267	99	PRT	310	35	1/X	353	43	RCL
268	53	(	311	43	RCL	354	00	00
269	42	STO	312	06	06	355	65	×
270	00	00	313	32	X⇌T	356	43	RCL
271	55	÷	314	01	1	357	02	02
272	53	(	315	67	EQ	358	54	)
273	01	1	316	44	SUM	359	61	GTO

Table B.22 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
360	45	YX	392	65	×	424	43	RCL
361	76	LBL	393	76	LBL	425	06	06
362	44	SUM	394	43	RCL	426	75	-
363	53	(	395	43	RCL	427	53	(
364	53	(	396	02	02	428	43	RCL
365	53	(	397	76	LBL	429	01	01
366	43	RCL	398	45	YX	430	65	×
367	00	00	399	94	+/-	431	43	RCL
368	65	×	400	85	+	432	04	04
369	43	RCL	401	01	1	433	55	÷
370	02	02	402	95	=	434	43	RCL
371	65	×	403	99	PRT	435	03	03
372	43	RCL	404	43	RCL	436	54	)
373	01	01	405	24	24	437	54	)
374	39	COS	406	99	PRT	438	92	RTN
375	54	)	407	98	ADV	439	76	LBL
376	85	+	408	92	RTN	440	25	CLR
377	43	RCL	409	76	LBL	441	53	(
378	01	01	410	65	×	442	43	RCL
379	54	)	411	53	(	443	05	05
380	54	)	412	43	RCL	444	75	-
381	42	STO	413	06	06	445	53	(
382	06	06	414	65	×	446	43	RCL
383	61	GTO	415	02	2	447	04	04
384	65	×	416	55	÷	448	33	X²
385	76	LBL	417	89	π	449	55	÷
386	42	STO	418	54	)	450	43	RCL
387	43	RCL	419	61	GTO	451	03	03
388	02	02	420	45	YX	452	54	)
389	42	STO	421	76	LBL	453	54	)
390	06	06	422	24	CE	454	92	RTN
391	61	GTO	423	53	(			

B.22 DETAILS OF THE REGRESSION-2 PROGRAM: LINEAR CORRELATIONS OTHER THAN STRAIGHT LINES

Lines 0 to 17 take the input data and calculate the sums for the function x with KEY $\Sigma+$.

Lines 18 to 109 convert the input value of x to x^2, $\sqrt{x}$, e^x, $1/x$, and log (x); for each function they calculate $\Sigma F(x)$, $\Sigma(F(x))^2$, and $\Sigma y \cdot F(x)$, and store the sums in Registers 07 to 21 in order.

Lines 110 and 111 start the calculation of the correlation with each of the functions. All of the calculations are carried out in the same manner, and are done by a subroutine starting at line 186.

Lines 112 to 185 shift the sums calculated for each function to the storage registers used by the built-in function of KEY $\Sigma+$ so that the correlation calculation can call up the values from the same registers.

Lines 186 to 208 calculate b_1 from Eq. (7.5). The numerator of Eq. (7.5) is calculated by a subroutine starting at line 421, since this same calculation will be made to calculate both r and $s(\hat{y})$.

Lines 209 to 224 calculate b_0 from Eq. (7.6).

Lines 225 to 248 calculate $s(\hat{y})$ and store the value in Register 24. The term $\Sigma y^2 - (\Sigma y)^2/n$ is calculated by a subroutine starting at line 439 since this calculation will be made again for the calculation of r.

Lines 249 to 267 calculate the correlation coefficient r.

Lines 268 to 403 form the probability calculation and are identical with the t-3 program, with the conversion of r to t, following Eq. (7.9).

Lines 404 to 408 recall the value of $s(\hat{y})$, print the result, and then return the program for the calculation of the next correlation function.

Lines 409 to 420 are part of the probability calculation and are called upon when the degrees of freedom are odd.

Lines 421 to 438 are the subroutine for calculating the numerator of Eq. (7.5).

Lines 439 to 454 are the subroutine for calculating $\Sigma y^2 - (\Sigma y)^2/n$.

Table B.23 Regression-3 Program: Two-Variable and Second-Degree Correlations (TI-59)

Two-variable
$y = b_0 + b_1x_1$
$y = b_0 + b_1x_2$
$y = b_0 + b_1x_1 + b_2x_2$

Input: y_i, Key A; x_{1i}, KEY A; x_{2i}, KEY A
KEY C

Second-degree
$y = b_0 + b_1x$
$y = b_0 + b_1x^2$
$y = b_0 + b_1x + b_2x^2$

Input: y_i, KEY $x \rightleftharpoons t$; x_i, KEY B
KEY C

Output: b_1, b_0, r for first equation
b_1, b_0, r for second equation
b_1, b_2, b_0, R for third equation
t comparing R of third equation with better of the other two

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	033	22	INV	066	12	12
001	76	LBL	034	86	STF	067	54	)
002	11	A	035	01	01	068	44	SUM
003	87	IFF	036	22	INV	069	10	10
004	00	00	037	86	STF	070	91	R/S
005	22	INV	038	02	02	071	76	LBL
006	47	CMS	039	43	RCL	072	12	B
007	86	STF	040	11	11	073	87	IFF
008	00	00	041	32	X⇌T	074	00	00
009	76	LBL	042	43	RCL	075	25	CLR
010	22	INV	043	12	12	076	47	CMS
011	87	IFF	044	78	Σ+	077	86	STF
012	01	01	045	53	(	078	00	00
013	23	LNX	046	43	RCL	079	76	LBL
014	42	STO	047	11	11	080	25	CLR
015	11	11	048	65	×	081	42	STO
016	86	STF	049	43	RCL	082	12	12
017	01	01	050	13	13	083	32	X⇌T
018	91	R/S	051	54	)	084	42	STO
019	76	LBL	052	44	SUM	085	11	11
020	23	LNX	053	07	07	086	32	X⇌T
021	87	IFF	054	43	RCL	087	78	Σ+
022	02	02	055	13	13	088	53	(
023	24	CE	056	44	SUM	089	43	RCL
024	42	STO	057	08	08	090	12	12
025	12	12	058	33	X²	091	33	X²
026	86	STF	059	44	SUM	092	44	SUM
027	02	02	060	09	09	093	08	08
028	91	R/S	061	53	(	094	65	×
029	76	LBL	062	43	RCL	095	43	RCL
030	24	CE	063	13	13	096	12	12
031	42	STO	064	65	×	097	54	)
032	13	13	065	43	RCL	098	44	SUM

Table B.23 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
099	10	10	142	43	RCL	185	54	)
100	43	RCL	143	03	03	186	42	STO
101	12	12	144	54	)	187	07	07
102	33	X^2	145	42	STO	188	53	(
103	33	X^2	146	06	06	189	43	RCL
104	44	SUM	147	53	(	190	10	10
105	09	09	148	43	RCL	191	75	-
106	53	(	149	02	02	192	43	RCL
107	43	RCL	150	75	-	193	08	08
108	12	12	151	43	RCL	194	65	×
109	33	X^2	152	01	01	195	43	RCL
110	65	×	153	33	X^2	196	01	01
111	43	RCL	154	55	÷	197	55	÷
112	11	11	155	43	RCL	198	43	RCL
113	54	)	156	03	03	199	03	03
114	44	SUM	157	54	)	200	54	)
115	07	07	158	42	STO	201	42	STO
116	91	R/S	159	02	02	202	10	10
117	76	LBL	160	53	(	203	53	(
118	13	C	161	43	RCL	204	43	RCL
119	53	(	162	09	09	205	06	06
120	43	RCL	163	75	-	206	55	÷
121	05	05	164	43	RCL	207	43	RCL
122	75	-	165	08	08	208	02	02
123	43	RCL	166	33	X^2	209	54	)
124	04	04	167	55	÷	210	99	PRT
125	33	X^2	168	43	RCL	211	42	STO
126	55	÷	169	03	03	212	11	11
127	43	RCL	170	54	)	213	53	(
128	03	03	171	42	STO	214	53	(
129	54	)	172	09	09	215	43	RCL
130	42	STO	173	53	(	216	04	04
131	05	05	174	43	RCL	217	75	-
132	53	(	175	07	07	218	43	RCL
133	43	RCL	176	75	-	219	11	11
134	06	06	177	43	RCL	220	65	×
135	75	-	178	08	08	221	43	RCL
136	43	RCL	179	65	×	222	01	01
137	01	01	180	43	RCL	223	54	)
138	65	×	181	04	04	224	55	÷
139	43	RCL	182	55	÷	225	43	RCL
140	04	04	183	43	RCL	226	03	03
141	55	÷	184	03	03	227	54	)

Table B.23 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
228	99	PRT	271	43	RCL	314	21	21
229	53	(	272	11	11	315	54	)
230	43	RCL	273	65	×	316	99	PRT
231	11	11	274	43	RCL	317	42	STO
232	65	×	275	07	07	318	11	11
233	43	RCL	276	55	÷	319	53	(
234	06	06	277	43	RCL	320	53	(
235	55	÷	278	05	05	321	43	RCL
236	43	RCL	279	54	)	322	07	07
237	05	05	280	34	√X	323	65	×
238	54	)	281	99	PRT	324	43	RCL
239	34	√X	282	98	ADV	325	02	02
240	99	PRT	283	42	STO	326	75	-
241	98	ADV	284	22	22	327	43	RCL
242	42	STO	285	53	(	328	06	06
243	20	20	286	43	RCL	329	65	×
244	53	(	287	02	02	330	43	RCL
245	43	RCL	288	65	×	331	10	10
246	07	07	289	43	RCL	332	54	)
247	55	÷	290	09	09	333	55	÷
248	43	RCL	291	75	-	334	43	RCL
249	09	09	292	43	RCL	335	21	21
250	54	)	293	10	10	336	54	)
251	99	PRT	294	33	X²	337	99	PRT
252	42	STO	295	54	)	338	42	STO
253	11	11	296	42	STO	339	12	12
254	53	(	297	21	21	340	53	(
255	53	(	298	53	(	341	53	(
256	43	RCL	299	53	(	342	43	RCL
257	04	04	300	43	RCL	343	04	04
258	75	-	301	06	06	344	75	-
259	43	RCL	302	65	×	345	43	RCL
260	11	11	303	43	RCL	346	11	11
261	65	×	304	09	09	347	65	×
262	43	RCL	305	75	-	348	43	RCL
263	08	08	306	43	RCL	349	01	01
264	54	)	307	07	07	350	75	-
265	55	÷	308	65	×	351	43	RCL
266	43	RCL	309	43	RCL	352	12	12
267	03	03	310	10	10	353	65	×
268	54	)	311	54	)	354	43	RCL
269	99	PRT	312	55	÷	355	08	08
270	53	(	313	43	RCL	356	54	)

Table B.23 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
357	55	÷	382	34	√X	407	43	RCL
358	43	RCL	383	99	PRT	408	00	00
359	03	03	384	98	ADV	409	54	)
360	54	)	385	42	STO	410	55	÷
361	99	PRT	386	13	13	411	53	(
362	53	(	387	43	RCL	412	01	1
363	53	(	388	20	20	413	75	-
364	43	RCL	389	32	X:T	414	43	RCL
365	11	11	390	43	RCL	415	13	13
366	65	×	391	22	22	416	33	X²
367	43	RCL	392	77	GE	417	54	)
368	06	06	393	42	STO	418	54	)
369	85	+	394	32	X:T	419	65	×
370	43	RCL	395	76	LBL	420	53	(
371	12	12	396	42	STO	421	43	RCL
372	65	×	397	33	X²	422	03	03
373	43	RCL	398	42	STO	423	75	-
374	07	07	399	00	00	424	03	3
375	54	)	400	53	(	425	54	)
376	42	STO	401	53	(	426	54	)
377	14	14	402	53	(	427	34	√X
378	55	÷	403	43	RCL	428	99	PRT
379	43	RCL	404	13	13	429	81	RST
380	05	05	405	33	X²	430	91	R/S
381	54	)	406	75	-			

B.23 DETAILS OF THE REGRESSION-3 PROGRAM: TWO-VARIABLE AND POLYNOMIAL CORRELATION

Lines 0 to 44 take the input y, x_1, and x_2 data and calculate the sums relating to y and x_1 with the built-in functions of KEY $\Sigma+$.

Lines 45 to 70 calculate the sums relating to y and x_2, and summation x_1x_2, and store the values in selected registers.

Lines 71 to 116 do the same as the above except with the input y and x data for the polynomial equation solution. The two different routines are used because the second includes the squaring of the x value for the second term of the polynomial.

Lines 117 to 202 start the calculation of the least-squares equations by calculating and storing the terms in Eqs. (7.19), (7.20), and (7.21).

Lines 203 to 240 calculate b_1, b_o, and r for the first equation from Eqs. (7.5), (7.6), and (7.8).

Lines 241 to 281 store the correlation coefficient from the first equation and then calculate b_1, b_0, and r for the second equation.

Lines 282 to 316 store the correlation coefficient from the second equation and then calculate b_1 for the two-variable equation from Eq. (7.19).

Lines 317 to 337 calculate b_2 for the two-variable equation from Eq. (7.20).

Lines 338 to 361 calculate b_0 for the two-variable equation from Eq. (7.22).

Lines 362 to 383 calculate the correlation coefficient for the two-variable equation from Eq. (7.25).

Lines 384 to 394 select the larger of the correlation coefficients from the first two equations.

Lines 395 to 429 calculate t from a comparison of the correlation coefficients by Eq. (7.29), and end the program.

Table B.24 Regression-4 Program: Three-Variable Correlation Equation (Part 1) (TI-59)

Equations: $y = b_0 + b_1x$ for all variables
$y = b_0 + b_1x_i + b_2x_j$ for all pairs
$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$

Input: y_i, KEY A; x_{1i}, KEY A, x_{2i}, KEY A, x_{3i}, KEY A
When all data are in: Key C

Output: b_1, b_0, r for single-variable equations in order x_1, x_2 x_3
b_1, b_2, b_0, R for two-variable equations in order x_1x_2, x_1x_3, x_2x_3
b_1, b_2, b_3, b_0, R for three-variable equation

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	034	03	03	068	44	SUM
001	76	LBL	035	25	CLR	069	13	13
002	11	A	036	42	STO	070	33	X²
003	87	IFF	037	08	08	071	44	SUM
004	00	00	038	44	SUM	072	14	14
005	22	INV	039	10	10	073	53	(
006	47	CMS	040	33	X²	074	43	RCL
007	86	STF	041	44	SUM	075	00	00
008	00	00	042	11	11	076	65	×
009	76	LBL	043	53	(	077	43	RCL
010	22	INV	044	43	RCL	078	09	09
011	87	IFF	045	00	00	079	54	)
012	01	01	046	65	×	080	44	SUM
013	23	LNX	047	43	RCL	081	15	15
014	42	STO	048	08	08	082	53	(
015	00	00	049	54	)	083	43	RCL
016	32	X⇌T	050	44	SUM	084	07	07
017	86	STF	051	12	12	085	65	×
018	01	01	052	86	STF	086	43	RCL
019	91	R/S	053	03	03	087	08	08
020	76	LBL	054	91	R/S	088	54	)
021	23	LNX	055	76	LBL	089	44	SUM
022	87	IFF	056	25	CLR	090	16	16
023	02	02	057	22	INV	091	53	(
024	24	CE	058	86	STF	092	43	RCL
025	42	STO	059	01	01	093	07	07
026	07	07	060	22	INV	094	65	×
027	78	Σ+	061	86	STF	095	43	RCL
028	86	STF	062	02	02	096	09	09
029	02	02	063	22	INV	097	54	)
030	91	R/S	064	86	STF	098	44	SUM
031	76	LBL	065	03	03	099	17	17
032	24	CE	066	42	STO	100	53	(
033	87	IFF	067	09	09	101	43	RCL

Table B.24 (*Cont.*)

Step	Code	Key
102	08	08
103	65	×
104	43	RCL
105	09	09
106	54	)
107	44	SUM
108	18	18
109	91	R/S
110	76	LBL
111	13	C
112	43	RCL
113	01	01
114	42	STO
115	07	07
116	53	(
117	43	RCL
118	05	05
119	75	-
120	53	(
121	43	RCL
122	04	04
123	33	X²
124	55	÷
125	43	RCL
126	03	03
127	54	)
128	54	)
129	42	STO
130	05	05
131	71	SBR
132	32	X⇌T
133	43	RCL
134	10	10
135	42	STO
136	01	01
137	43	RCL
138	02	02
139	42	STO
140	08	08
141	43	RCL
142	11	11
143	42	STO
144	02	02
145	43	RCL
146	06	06
147	42	STO
148	09	09
149	43	RCL
150	12	12
151	42	STO
152	06	06
153	71	SBR
154	32	X⇌T
155	43	RCL
156	02	02
157	42	STO
158	11	11
159	43	RCL
160	06	06
161	42	STO
162	12	12
163	43	RCL
164	13	13
165	42	STO
166	01	01
167	43	RCL
168	14	14
169	42	STO
170	02	02
171	43	RCL
172	15	15
173	42	STO
174	06	06
175	71	SBR
176	32	X⇌T
177	53	(
178	43	RCL
179	16	16
180	75	-
181	53	(
182	43	RCL
183	07	07
184	65	×
185	43	RCL
186	10	10
187	55	÷
188	43	RCL
189	03	03
190	54	)
191	54	)
192	42	STO
193	16	16
194	53	(
195	43	RCL
196	17	17
197	75	-
198	53	(
199	43	RCL
200	07	07
201	65	×
202	43	RCL
203	13	13
204	55	÷
205	43	RCL
206	03	03
207	54	)
208	54	)
209	42	STO
210	17	17
211	53	(
212	43	RCL
213	18	18
214	75	-
215	53	(
216	43	RCL
217	10	10
218	65	×
219	43	RCL
220	13	13
221	55	÷
222	43	RCL
223	03	03
224	54	)
225	54	)
226	42	STO
227	18	18
228	71	SBR
229	33	X²
230	43	RCL

Table B.24 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
231	10	10	274	42	STO	317	53	(
232	42	STO	275	06	06	318	43	RCL
233	22	22	276	43	RCL	319	06	06
234	43	RCL	277	24	24	320	75	-
235	13	13	278	42	STO	321	43	RCL
236	42	STO	279	09	09	322	01	01
237	10	10	280	43	RCL	323	65	×
238	43	RCL	281	08	08	324	43	RCL
239	11	11	282	42	STO	325	04	04
240	42	STO	283	02	02	326	55	÷
241	23	23	284	43	RCL	327	43	RCL
242	43	RCL	285	23	23	328	03	03
243	02	02	286	42	STO	329	54	)
244	42	STO	287	08	08	330	42	STO
245	11	11	288	43	RCL	331	06	06
246	43	RCL	289	16	16	332	55	÷
247	12	12	290	42	STO	333	43	RCL
248	42	STO	291	17	17	334	02	02
249	24	24	292	43	RCL	335	54	)
250	43	RCL	293	18	18	336	99	PRT
251	06	06	294	42	STO	337	42	STO
252	42	STO	295	16	16	338	19	19
253	12	12	296	71	SBR	339	53	(
254	43	RCL	297	33	X^2	340	53	(
255	16	16	298	81	RST	341	43	RCL
256	42	STO	299	76	LBL	342	04	04
257	25	25	300	32	X⇌T	343	75	-
258	43	RCL	301	53	(	344	53	(
259	17	17	302	43	RCL	345	43	RCL
260	42	STO	303	02	02	346	19	19
261	16	16	304	75	-	347	65	×
262	71	SBR	305	53	(	348	43	RCL
263	33	X^2	306	43	RCL	349	01	01
264	43	RCL	307	01	01	350	54	)
265	07	07	308	33	X^2	351	54	)
266	42	STO	309	55	÷	352	55	÷
267	13	13	310	43	RCL	353	43	RCL
268	43	RCL	311	03	03	354	03	03
269	22	22	312	54	)	355	54	)
270	42	STO	313	54	)	356	99	PRT
271	07	07	314	42	STO	357	53	(
272	43	RCL	315	02	02	358	43	RCL
273	09	09	316	53	(	359	19	19

Table B.24 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
360	65	×	398	16	16	436	43	RCL
361	43	RCL	399	54	)	437	07	07
362	06	06	400	55	÷	438	75	-
363	55	÷	401	43	RCL	439	43	RCL
364	43	RCL	402	19	19	440	21	21
365	05	05	403	54	)	441	65	×
366	54	)	404	99	PRT	442	43	RCL
367	34	√X	405	42	STO	443	10	10
368	99	PRT	406	20	20	444	54	)
369	98	ADV	407	53	(	445	55	÷
370	92	RTN	408	53	(	446	43	RCL
371	76	LBL	409	43	RCL	447	03	03
372	33	X^2	410	12	12	448	54	)
373	53	(	411	65	×	449	99	PRT
374	43	RCL	412	43	RCL	450	53	(
375	11	11	413	08	08	451	53	(
376	65	×	414	75	-	452	43	RCL
377	43	RCL	415	43	RCL	453	20	20
378	08	08	416	09	09	454	65	×
379	75	-	417	65	×	455	43	RCL
380	43	RCL	418	43	RCL	456	09	09
381	16	16	419	16	16	457	85	+
382	33	X^2	420	54	)	458	43	RCL
383	54	)	421	55	÷	459	21	21
384	42	STO	422	43	RCL	460	65	×
385	19	19	423	19	19	461	43	RCL
386	53	(	424	54	)	462	12	12
387	53	(	425	42	STO	463	54	)
388	43	RCL	426	21	21	464	55	÷
389	09	09	427	99	PRT	465	43	RCL
390	65	×	428	53	(	466	05	05
391	43	RCL	429	53	(	467	54	)
392	11	11	430	43	RCL	468	34	√X
393	75	-	431	04	04	469	99	PRT
394	43	RCL	432	75	-	470	98	ADV
395	12	12	433	43	RCL	471	92	RTN
396	65	×	434	20	20			
397	43	RCL	435	65	×			

B.24 DETAILS OF THE REGRESSION-4 PROGRAM: THREE-VARIABLE CORRELATION (PART 1)

Lines 0 to 27 take the input data y, x_1 and use the built-in functions of KEY $\Sigma+$ to accumulate the sums and sums of cross products.

Lines 28 to 54 take the input of x_2 and accumulate the sums of terms including that variable.

Lines 55 to 109 take the input of x_3 and accumulate all the sums including that variable.

Lines 110 to 132 start the calculation of the least-squares equations. All of the one-variable equations are calculated with the same subroutine, starting at line 299. Lines 117 to 130 calculate $\Sigma y^2 - (\Sigma y)^2/n$, which is used in all of the equations, and then direct the program to the calculation subroutine at line 299.

Lines 133 to 154 shift the sums involving the second variable to the proper storage areas for use by the calculation subroutine for the next single-variable equation.

Lines 155 to 176 shift the sums of the third variable and then call on the calculation subroutine.

Lines 177 to 229 calculate the cross product terms of Eqs. (7.19) and (7.20) that are required for the two-variable least-squares solution, and then call on a subroutine at line 371 which calculates all the two-variable equations.

Lines 230 to 263 shift the terms for the second two-variable equation to the proper storage registers for the subroutine solution.

Lines 264 to 297 do the same for the third two variable-equation.

Line 298 ends the calculations of Part 1 of the program.

Lines 299 to 370 are the subroutine for the single-variable equation solution. The solution follows Eqs. (7.5) to (7.8).

Lines 371 to 471 are the subroutine for the calculation of a two-variable least-squares solution. The calculation is the same as in the Regression-3 program and follows Eqs. (7.19) to (7.25).

All the sums required for the three-variable equation solution are stored in preparation for Part 2 of the program.

Table B.25 Regression-4 Program (Part 2) (TI-59)

Part 1 of Regression-4 gives the output for the one- and two-variable equations. Part 2 gives the output for the three-variable equation. Part 2 is run with KEY A.

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	76	LBL	041	55	÷	082	23	23
001	11	A	042	43	RCL	083	53	(
002	53	(	043	16	16	084	43	RCL
003	43	RCL	044	54	)	085	15	15
004	06	06	045	42	STO	086	75	-
005	55	÷	046	19	19	087	43	RCL
006	43	RCL	047	53	(	088	21	21
007	17	17	048	43	RCL	089	54	)
008	54	)	049	08	08	090	42	STO
009	42	STO	050	55	÷	091	24	24
010	00	00	051	43	RCL	092	53	(
011	53	(	052	16	16	093	43	RCL
012	43	RCL	053	54	)	094	19	19
013	02	02	054	42	STO	095	75	-
014	55	÷	055	20	20	096	43	RCL
015	43	RCL	056	53	(	097	22	22
016	17	17	057	43	RCL	098	54	)
017	54	)	058	12	12	099	42	STO
018	42	STO	059	55	÷	100	22	22
019	10	10	060	43	RCL	101	53	(
020	53	(	061	11	11	102	43	RCL
021	43	RCL	062	54	)	103	20	20
022	25	25	063	42	STO	104	75	-
023	55	÷	064	21	21	105	43	RCL
024	43	RCL	065	53	(	106	23	23
025	17	17	066	43	RCL	107	54	)
026	54	)	067	17	17	108	42	STO
027	42	STO	068	55	÷	109	23	23
028	14	14	069	43	RCL	110	53	(
029	53	(	070	11	11	111	43	RCL
030	43	RCL	071	54	)	112	00	00
031	09	09	072	42	STO	113	75	-
032	55	÷	073	22	22	114	43	RCL
033	43	RCL	074	53	(	115	15	15
034	16	16	075	43	RCL	116	54	)
035	54	)	076	16	16	117	42	STO
036	42	STO	077	55	÷	118	15	15
037	15	15	078	43	RCL	119	53	(
038	53	(	079	11	11	120	43	RCL
039	43	RCL	080	54	)	121	10	10
040	25	25	081	42	STO	122	75	-

Table B.25 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
123	43	RCL	166	19	19	209	10	10
124	19	19	167	55	÷	210	99	PRT
125	54	)	168	43	RCL	211	53	(
126	42	STO	169	20	20	212	53	(
127	19	19	170	54	)	213	43	RCL
128	53	(	171	42	STO	214	06	06
129	43	RCL	172	21	21	215	75	-
130	14	14	173	53	(	216	43	RCL
131	75	-	174	53	(	217	00	00
132	43	RCL	175	43	RCL	218	65	×
133	20	20	176	23	23	219	43	RCL
134	54	)	177	75	-	220	02	02
135	42	STO	178	43	RCL	221	75	-
136	20	20	179	10	10	222	43	RCL
137	53	(	180	54	)	223	25	25
138	43	RCL	181	55	÷	224	65	×
139	24	24	182	53	(	225	43	RCL
140	55	÷	183	43	RCL	226	10	10
141	43	RCL	184	21	21	227	54	)
142	23	23	185	75	-	228	55	÷
143	54	)	186	43	RCL	229	43	RCL
144	42	STO	187	22	22	230	17	17
145	10	10	188	54	)	231	54	)
146	53	(	189	54	)	232	42	STO
147	43	RCL	190	42	STO	233	14	14
148	22	22	191	00	00	234	99	PRT
149	55	÷	192	99	PRT	235	53	(
150	43	RCL	193	53	(	236	53	(
151	23	23	194	53	(	237	43	RCL
152	54	)	195	43	RCL	238	04	04
153	42	STO	196	15	15	239	75	-
154	22	22	197	75	-	240	43	RCL
155	53	(	198	43	RCL	241	00	00
156	43	RCL	199	19	19	242	65	×
157	15	15	200	65	×	243	43	RCL
158	55	÷	201	43	RCL	244	13	13
159	43	RCL	202	00	00	245	75	-
160	20	20	203	54	)	246	43	RCL
161	54	)	204	55	÷	247	10	10
162	42	STO	205	43	RCL	248	65	×
163	23	23	206	20	20	249	43	RCL
164	53	(	207	54	)	250	07	07
165	43	RCL	208	42	STO	251	75	-

Table B.25 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
252	43	RCL	265	43	RCL	278	14	14
253	14	14	266	00	00	279	65	×
254	65	×	267	65	×	280	43	RCL
255	43	RCL	268	43	RCL	281	12	12
256	01	01	269	06	06	282	54	)
257	54	)	270	85	+	283	55	÷
258	55	÷	271	43	RCL	284	43	RCL
259	43	RCL	272	10	10	285	05	05
260	03	03	273	65	×	286	54	)
261	54	)	274	43	RCL	287	34	√X
262	99	PRT	275	09	09	288	99	PRT
263	53	(	276	85	+	289	91	R/S
264	53	(	277	43	RCL			

B.25 DETAILS OF THE REGRESSION-4 PROGRAM, PART 2

This portion of the program is the solution of three simultaneous equations derived from Eq. (7.33). The term b_0 is eliminated from the four equations shown.

Lines 0 to 136 divide each term of the three equations by the coefficient of the last term and then subtract the corresponding pairs to eliminate the third variable.

Lines 137 to 172 divide each term of the two equations by the coefficient of the last term to derive two equations which have a coefficient of 1.0 for the last term.

Lines 173 to 192 subtract the two equations remaining at line 172 and divide the first term by the coefficient of the second term to obtain the value of b_1.

Lines 193 to 210 substitute the value of b_1 into one of the equations remaining at line 172 to obtain the value of b_2.

Lines 211 to 234 substitute b_1 and b_2 into one of the equations remaining at line 136 to obtain b_3.

Lines 235 to 262 calculate b_0 from an extension of Eq. (7.22) to three variables.

Lines 263 to 289 calculate the correlation coefficient by an extension of Eqs. (7.24) and (7.25) to three variables, and end the calculation.

Table B.26 Regression-5 Program: Correlation of Family of Curves (TI-59)

Equation: $y = b_0 + b_1 x_1 + b_2(x_1 x_2)$
and if Flag 1 is set: $y = b_0 + b_1 x_1 + b_2(x_1 \cdot \log(x_2))$

Input: y_i, KEY A; x_{1i}, KEY A; x_{2i}, KEY A; . . . ; with all data in: KEY C

Output: b_1, b_2, b_0, R; and if Flag 1 is set: b_1, b_2, b_0, R for second equation.

Step	Code	Key	Step	Code	Key	Step	Code	Key
000	91	R/S	038	78	Σ+	076	76	LBL
001	76	LBL	039	53	(	077	32	X:T
002	11	A	040	43	RCL	078	22	INV
003	87	IFF	041	13	13	079	86	STF
004	02	02	042	44	SUM	080	04	04
005	22	INV	043	16	16	081	22	INV
006	47	CMS	044	65	×	082	86	STF
007	86	STF	045	43	RCL	083	03	03
008	02	02	046	12	12	084	91	R/S
009	76	LBL	047	54	)	085	76	LBL
010	22	INV	048	42	STO	086	25	CLR
011	87	IFF	049	15	15	087	43	RCL
012	03	03	050	44	SUM	088	13	13
013	23	LNX	051	00	00	089	53	(
014	42	STO	052	33	X^2	090	28	LOG
015	11	11	053	44	SUM	091	42	STO
016	86	STF	054	21	21	092	14	14
017	03	03	055	53	(	093	65	×
018	91	R/S	056	43	RCL	094	43	RCL
019	76	LBL	057	11	11	095	12	12
020	23	LNX	058	65	×	096	54	)
021	87	IFF	059	43	RCL	097	42	STO
022	04	04	060	15	15	098	15	15
023	24	CE	061	54	)	099	44	SUM
024	42	STO	062	44	SUM	100	24	24
025	12	12	063	22	22	101	33	X^2
026	86	STF	064	53	(	102	44	SUM
027	04	04	065	43	RCL	103	25	25
028	91	R/S	066	12	12	104	53	(
029	76	LBL	067	65	×	105	43	RCL
030	24	CE	068	43	RCL	106	11	11
031	42	STO	069	15	15	107	65	×
032	13	13	070	54	)	108	43	RCL
033	43	RCL	071	44	SUM	109	15	15
034	11	11	072	23	23	110	54	)
035	32	X:T	073	87	IFF	111	44	SUM
036	43	RCL	074	01	01	112	26	26
037	12	12	075	25	CLR	113	53	(

Table B.26 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
114	43	RCL	157	75	−	200	01	01
115	12	12	158	43	RCL	201	65	×
116	65	×	159	01	01	202	43	RCL
117	43	RCL	160	33	X²	203	00	00
118	15	15	161	55	÷	204	55	÷
119	54	)	162	43	RCL	205	43	RCL
120	44	SUM	163	03	03	206	03	03
121	27	27	164	54	)	207	54	)
122	61	GTO	165	42	STO	208	42	STO
123	32	X⇌T	166	02	02	209	23	23
124	76	LBL	167	53	(	210	71	SBR
125	13	C	168	43	RCL	211	33	X²
126	53	(	169	21	21	212	87	IFF
127	43	RCL	170	75	−	213	01	01
128	05	05	171	43	RCL	214	34	√X
129	75	−	172	00	00	215	81	RST
130	43	RCL	173	33	X²	216	91	R/S
131	04	04	174	55	÷	217	76	LBL
132	33	X²	175	43	RCL	218	34	√X
133	55	÷	176	03	03	219	53	(
134	43	RCL	177	54	)	220	43	RCL
135	03	03	178	42	STO	221	25	25
136	54	)	179	21	21	222	75	−
137	42	STO	180	53	(	223	43	RCL
138	05	05	181	43	RCL	224	24	24
139	53	(	182	22	22	225	33	X²
140	43	RCL	183	75	−	226	55	÷
141	06	06	184	43	RCL	227	43	RCL
142	75	−	185	04	04	228	03	03
143	43	RCL	186	65	×	229	54	)
144	04	04	187	43	RCL	230	42	STO
145	65	×	188	00	00	231	21	21
146	43	RCL	189	55	÷	232	53	(
147	01	01	190	43	RCL	233	43	RCL
148	55	÷	191	03	03	234	26	26
149	43	RCL	192	54	)	235	75	−
150	03	03	193	42	STO	236	43	RCL
151	54	)	194	22	22	237	04	04
152	42	STO	195	53	(	238	65	×
153	06	06	196	43	RCL	239	43	RCL
154	53	(	197	23	23	240	24	24
155	43	RCL	198	75	−	241	55	÷
156	02	02	199	43	RCL	242	43	RCL

Table B.26 (*Cont.*)

Step	Code	Key	Step	Code	Key	Step	Code	Key
243	03	03	287	14	14	331	53	(
244	54	)	288	53	(	332	43	RCL
245	42	STO	289	53	(	333	04	04
246	22	22	290	43	RCL	334	75	-
247	53	(	291	06	06	335	43	RCL
248	43	RCL	292	65	×	336	11	11
249	27	27	293	43	RCL	337	65	×
250	75	-	294	21	21	338	43	RCL
251	43	RCL	295	75	-	339	01	01
252	24	24	296	43	RCL	340	75	-
253	65	×	297	22	22	341	43	RCL
254	43	RCL	298	65	×	342	12	12
255	01	01	299	43	RCL	343	65	×
256	55	÷	300	23	23	344	43	RCL
257	43	RCL	301	54	)	345	00	00
258	03	03	302	55	÷	346	54	)
259	54	)	303	43	RCL	347	55	÷
260	42	STO	304	14	14	348	43	RCL
261	23	23	305	54	)	349	03	03
262	43	RCL	306	99	PRT	350	54	)
263	24	24	307	42	STO	351	99	PRT
264	42	STO	308	11	11	352	53	(
265	00	00	309	53	(	353	53	(
266	71	SBR	310	53	(	354	43	RCL
267	33	X^2	311	43	RCL	355	11	11
268	22	INV	312	22	22	356	65	×
269	86	STF	313	65	×	357	43	RCL
270	01	01	314	43	RCL	358	06	06
271	81	RST	315	02	02	359	85	+
272	91	R/S	316	75	-	360	43	RCL
273	76	LBL	317	43	RCL	361	12	12
274	33	X^2	318	06	06	362	65	×
275	53	(	319	65	×	363	43	RCL
276	43	RCL	320	43	RCL	364	22	22
277	02	02	321	23	23	365	54	)
278	65	×	322	54	)	366	55	÷
279	43	RCL	323	55	÷	367	43	RCL
280	21	21	324	43	RCL	368	05	05
281	75	-	325	14	14	369	54	)
282	43	RCL	326	54	)	370	34	√X
283	23	23	327	99	PRT	371	99	PRT
284	33	X^2	328	42	STO	372	98	ADV
285	54	)	329	12	12	373	92	RTN
286	42	STO	330	53	(			

B.26 DETAILS OF THE REGRESSION-5 PROGRAM: CORRELATION OF FAMILY OF CURVES

Lines 0 to 28 take input of y and x_1 data. The value of x_1 is first used at line 24. If the function of x_1 is to be changed, the change must be made between lines 23 and 24.

Lines 29 to 72 take input of the x_2 data and calculate the various sums required for the least-squares calculation. The y-x_1 sums are accumulated with KEY $\Sigma+$, and the balance with program instructions.

Lines 73 to 84 check whether Flag 1 is set. If so, the program goes to line 85 to calculate log (x_2). If not, the program is arranged for more data input.

Lines 85 to 123 calculate the sums involving log (x_2) if that function is used. If some other function than log is desired, the change should be made at line 90.

Lines 124 to 211 start the calculation of the least-squares solution. The calculation is carried out by a subroutine starting at line 273. Lines 126 to 209 calculate and store the various terms used in the solution of Eqs. (7.19) to (7.25). At line 210 the program is directed to the calculation subroutine.

Lines 212 to 216 check whether Flag 1 is set. If not, the program ends. If so, the program goes on to line 217.

Lines 217 to 267 calculate the terms necessary for the second solution of the least-squares equations, using the second function of x_2. These are stored in the same registers used by the first function so that the calculation can be made by the same subroutine.

Lines 268 to 272 clear Flag 1 for the next use of the program and end the calculation.

Lines 273 to 308 start the calculation of the least-squares solution of a two-variable correlation. The same calculations are made as in Regression-3. Equations (7.19) to (7.25) are followed; b_1 is calculated at line 305.

Lines 309 to 329 calculate b_2 from Eq. (7.20).

Lines 330 to 351 calculate b_0 from Eq. (7.22).

Lines 352 to 373 calculate R from Eq. (7.25) and end the calculations.

REFERENCES

1. *ASTM Manual on Quality Control of Materials:* Special Technical Publication 15-C, American Society for Testing and Materials, Philadelphia, Pa.
2. *Handbook of Mathematical Functions:* Applied Mathematics Series—55, U.S. Department of Commerce, 1964.*
3. *1979 World Almanac:* Newspaper Enterprise Associates, New York.
4. Bartlett, M. S.: *Supplement to Journal of the Royal Statistical Society,* vol. 2, 1935, p. 248.
5. Cochran, W. G., and G. M. Cox: *Experimental Design,* John Wiley & Sons, New York, 1957.
6. Davies, O. L. (ed): *Design and Analysis of Industrial Experiments,* Hafner Publishing Co., New York, 1956.
7. Scheffe, H.: *Analysis of Variance,* John Wiley and Sons, New York, 1959.
8. Natrella, M. G.: *Experimental Statistics,* National Bureau of Standards Handbook 91, U.S. Government Printing Office, Washington, D.C., 1963.†
9. Snedecor, G. W., and W. G. Cochran: *Statistical Methods,* Iowa State College Press, Ames, Iowa, 1967.
10. Volk, W.: *Applied Statistics for Engineers,* Robert E. Krieger Publishing Co., New York, 1980.
11. Eid, M. T., C. A. Black, and O. Kempthorne: Iowa Agricultural Experiment Station Research Bulletin 406, Ames, Iowa, February, 1954.
12. Davies, O. L. (ed): *Statistical Methods in Research and Production.* Hafner Publishing Co., New York, 1957.

13. Mosteller, F., and J. W. Tukey: *Data Analysis and Regression,* Addison-Wesley Publishing Co., Inc., Reading, Mass., 1977.

14. Yerushalmi, J., D. H. Turner, and A. M. Squires: *Industrial Engineering Chemistry, Process Design and Development,* vol. 15, no. 1., 1976.

15. Malm, C. J., K. T. Barkey, et al.: *Industrial & Engineering Chemistry,* vol. 44, no. 12, 1952.

* Some of the equations in this reference were adapted from *Approximations for Digital Computers,* by Cecil Hastings, Jr., assisted by Jeanne T. Hayward and James P. Wong, Jr. (copyright 1955 by the Rand Corporation) pp. 167 and 192. These adaptations are used with permission of Princeton University Press.

† Data from this reference were from "Measurement of Treadwear of Commercial Tires," G. G. Richey and J. Mandel in *Rubber Age,* vol. 73, no. 2, May 1953.

INDEX

About the Author

William Volk holds a masters degree in mathematical statistics from Rutgers University and an undergraduate degree in chemical engineering from New York University. He has worked for thirty-five years in chemical engineering research and development, the last fifteen of these years in statistical analysis and computer programming as applied to research data. He has taught statistics at both Polytechnic Institute of Brooklyn and Rutgers.

He holds thirteen U.S. patents on process development inventions and has written numerous articles for technical publications on data handling and process development. He is also author of Applied Statistics for Engineers, *a book that is now in its third edition. He is presently a consultant in data processing.*

(continued from front flap)

The subjects covered are:

- Statistical parameters
- Probability distributions
- *t* Function
- χ^2 Function
- Analysis of variance
- Regression and correlation

All you have to do to use the programs is enter your data. The output is relevant statistics and probabilities. And the programs are sufficiently detailed so you can store completed information, retrieve it, and use it for subsequent calculations.

Clearly, what all this means is—you *don't* have to have a thoroughgoing knowledge of statistical probability to apply statistical analysis to your problems. You simply need this thoroughgoing guide to direct you to the right procedures, the right answers, and the right interpretations of these answers.

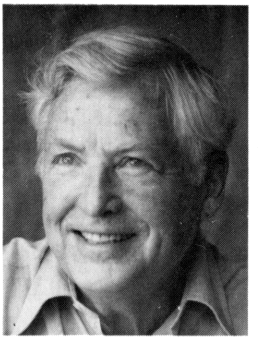


Photo: Elaine Miller

About the Author

William Volk holds a masters degree in mathematical statistics from Rutgers University and an undergraduate degree in chemical engineering from New York University. He has worked for thirty-five years in chemical engineering research and development, the last fifteen of these years in statistical analysis and computer programming as applied to research data. He has taught statistics at both Polytechnic Institute of Brooklyn and Rutgers.

He holds thirteen U.S. patents on process development inventions and has written numerous articles for technical publications on data handling and process development. He is also author of *Applied Statistics for Engineers,* a book that is now in its third edition. He is presently a consultant in data processing.

Photo: Karen Leeds
Design: Jehaber & Peace Inc.

Engineering
Statistics
Research

A unique new tool for engineers who want to apply statistical analysis to their problems but *aren't* experts on mathematical statistics . . .

ENGINEERING STATISTICS with a PROGRAMMABLE CALCULATOR

by William Volk

- Puts statistical programs that deliver not only the statistics but also the probability associated with the answers—right at your fingertips
- Provides programs in both RPN and AOS languages and gives examples from typical engineering problems for every program
- Gives you a handle on the magnitude of risks and size of errors in your conclusions
- Helps you make decisions based on objective probability calculations—on the spot

It is the *only* tool you need to give you the statistical data you need to solve your engineering problems!

McGraw-Hill Book Company
Serving the Need for Knowledge
1221 Avenue of the Americas
New York, NY 10020

ISBN 0-07-067552-X