SHAM-REG 3

SHAMMAS SOFTWARE SERVICES
1533F HONEY GROVE DRIVE
RICHMOND, VIRGINIA 23229

DISCLAIMER OF ALL WARRANTY AND LIABILITY

The author of the program makes no warranties, either expressed of implied, with respect to this manual or with respect to the software described in this manual, its quality, performance, merchantability or fitness for any particular purpose. SHAM-REG 3 is sold "As Is". While every effort is made to make SHAM-REG 3 versatile, in no event will Shammas Software Services and/or the author be liable for direct, indirect, incidental or consequential damages resulting from any defect in the software.

TABLE OF CONTENTS

FORWARD

SHAM-REG 3 is a linear regression pac that will run on the Hewlett-Packard HP 41CV or the HP 41C with a quad memory. The HP-IL interface module (model # 82160A) and the digital cassette drive (model # 82161A) are also needed. All of the hardware products mentioned above are manufactured by the Hewlett-Packard Co.

SHAM-REG 3

This program package performs linear regression between two variables, one called the independent variable whose variations affect a "response" variable, called the dependent variable. The purpose of regression analysis as a whole is to test the validity of an assumed "model" equation that tries to explain the relationship between the dependent variable and the independent variable. The main feature of SHAM-REG 3 is that it is possible to use weights or frequencies with the input data. Once the model is satisfactory, regression goes further to estimate the parameters (coefficients) involved in the model.

Often some transformations are needed to "linearize" the model so that the latter will lend itself to the methods used for "true" linear models. SHAM-REG 3 will carry out a variety of transformations to provide for a large number of models.

What do the programs do? The answer can be explained as follow :

(1) Input/Output : The user can key in data from the keyboard and store them on mass storage file for retrieval later. Data entered from the keyboard cannot exceed the calculator's memory allocated to store numbers. For large files, the user should use DATA BASE 1 to generate and edit such files. The routines in SHAM-REG 3 will only read large numeric files.

(2) Regression phase 1 : The programs regard the "raw" data as

merely a set of records. Each record contains two or more variables. The program has no preconceived notion of which variable is the independent or dependent variable, or the weight or frequency (should the last two be involved), especially when more than two variables are in each record. Since the regression will be carried out between two variables only and the data may contain MORE THAN TWO VARIABLES, the user will select the independent variable and dependent variable by number (they are seen as X1, X2, X3,...etc.) as well as their ptional weights (or frequencies) and transformations. This will permit to examine multi-variable data and shorten data input. The program will then carry out the regression and present the type of model, the coefficient of determination, r^2, slope and intercept. There are two routines that will present the regression results for a number of popular curves.

(3) Regression phase 2 : For the last selected model, the user can obtain an ANOVA table, standard errors for the regression coefficients, confidence intervals for the regression coefficients and projections for the dependent variable with its confidence intervals.

A word about transformations. Most of the available programs today use routines that take the values of the independent and dependent variables, tranform them IN THE SAME ROUTINE to the desired forms. This means that for 'n' linearizing models there should be 'n' routines. Not so with SHAM-REG 3 ! We have chosen

to carry out these transformations such that the transformation of the dependent variable and the independent variable are done SEPERATELY ! For example :

Model : X2 = aX1^b

transformation for X2 is logarithmic.

transformation for X1 is logarithmic.

linearized model : LN X2 = LN(a) + b LN(X1).

Model : X2 = a e^(b/X1)

transformation for X2 is logarithmic.

transformation for X1 is reciprocal.

Linearized model : LN(X2) = LN(a) + b(1/X1)

Model : X2 = a + b/X1

no transformation for X2.

transformation for X1 is reciprocal.

linearized model : X2 = a + b(1/X1)

This means that for 'n' transformation routines there can be n^2 models ! The transformation routines are the same for both variables involved in the regression.

A WORD OF CAUTION : The transformation routines expect the data to give valid results, so if the data has negative numbers do not use square root and logarthmic transformations. Also avoid reciprocal and logarithmic transformations when zeros exist in your data.

USER'S INSTRUCTIONS

Throughout the instructions "R/S*" indicates to omit the "R/S" when a printer is used.

(1) Here, first, is a list of the programs and their sizes in the SHAM-REG 3 package.

Name	Filename	Size (reg.)
R31	R31	45
R32	R32	86
R33	R33	89
LIB1	LIB1	18
	Total =	238

The above will allow the user to decide the SIZE for the registers. Program R31 is only needed for Input/Output operations. Program R32 and R33 perform the regression and its detailed statistics respectively. Their co-residence is really up to the user's convenience. Program LIB1 is the transformation functions library and MUST be resident in the calculator's memory (i.e. RAM) while R32 and/or R33 are used. Therefore the number of memory space dedicated for programs should be between 107 and

238 registers. Go to step (2) after setting up the proper SIZE.

(2) To load program LIB1.

INPUT : [ALPHA] LIB1 [ALPHA]

FUNCTION : XEQ READP

FUNCTION : GTO . .

The latter will ensure that the library of functions will be available at all times and not be overwritten while reading other programs, as long as the next READP is not issued while the program pointer is in the LIB1 program.

(3) To carry out data Input/Output.

(3.1) (Optional) Load program R31 if not in RAM already.

INPUT : [ALPHA] R31 [ALPHA]

FUNCTION : XEQ READP

(3.2) To initialize and link to the program.

FUNCTION : XEQ R31

(3.3) To input data from the keyboard. It is important to key in all the data in this stage.

(3.3.1) FUNCTION : XEQ A

DISPLAY : NO. VARS?

INPUT : # of variables per record (i.e. data point).

FUNCTION : R/S

(3.3.2) A loop will start to display each variable and the data point number (record#).

DISPLAY : Xi,j? (i= var#, j= record#)

INPUT : Xi,j

FUNCTION : R/S

(3.3.3) For more data input.

FUNCTION : R/S

Go to step (3.3.2)

(3.3.4) Repeat step (3.3.3) untill all the data are entered.

(3.3.5) (Optional) To delete the last 'k' data points entered.

INPUT : k

FUNCTION : STO - 07

FUNCTION : STO - 08

(3.4) To Save the data keyed in.

(3.4.1) FUNCTION : XEQ C

DISPLAY : FILENAME?

INPUT : filename

FUNCTION : R/S

(3.4.2) DISPLAY : STATUS?

INPUT : (Optional) Status filename of a that which already exists on tape and matches the current status. If no input is made the system creates a new status file under the name "S(data filename)"

FUNCTION : R/S

Note : If flag 01 is set the above step cannot be done, because it needs not to be.

(3.5) To input data from a file. This may be the data keyed in

earlier and stored or the data prepared and edited using DATA BASE 1.

(3.5.1) FUNCTION : XEQ B

DISPLAY : FILENAME?

INPUT : filename

FUNCTION : R/S

(3.5.2) DISPLAY : SIZE?

DISPLAY : Calculator's SIZE memory partition.

FUNCTION : R/S

(3.5.3) A beep will signal the end of the data input.

(4) To go to the regression phase 1.

(4.1) (Optional) Load program R32 if not in RAM already.

INPUT : [ALPHA] R32 [APLHA]

FUNCTION : XEQ READP

(4.2) To link to the program and reset default condition (no weights or frequencies).

FUNCTION : XEQ R32

(4.3) To select regression recipients.

(4.3.1) (Optional) To select "weighed" regression.

FUNCTION : XEQ a

(4.3.2) (Optional) To select regression with frequencies.

FUNCTION : XEQ b

(4.3.3) To select the independent and dependent variables.

FUNCTION : XEQ A

DISPLAY : IND VAR?

INPUT : independent variable by number.

FUNCTION : R/S

DISPLAY : DPND VAR?

INPUT : dependent variable by number.

FUNCTION : R/S

If no weights or frequencies are to be used (default condition) go to step (4.4).

If weights are involved.

DISPLAY : W?

INPUT : variable number of that which will provide the weights.

FUNCTION : R/S

If frequencies are involved.

DISPLAY : F?

INPUT : variable number of that which will provide the frequencies.

(4.4) To select the transformation for each variable, use the following table.

Transformation	Alpha input
None	none
1/var	1/
log(var)	LN
var^2	SQR
var^(1/2)	SQRT

var^(3)	CUBE
var^(-1/2)	1/SQRT
var^(-2)	1/SQR

Remember that the same alpha code applies to both variables.

FUNCTION : XEQ B

DISPLAY : TRSF Xi?

INPUT : Alpha code for the transformation of the dependent variable. If none make no input.

FUNCTION : R/S

DISPLAY : TRSF Xj?

INPUT : Alpha code for the transformation of the independent variable. If none make no input.

FUNCTION : R/S

(4.5) To carry out the regression for the currently selected variables and transformations.

FUNCTION : XEQ C

DISPLAY : curve equation.

FUNCTION : R/S*

DISPLAY : R2= (coefficient of determination)

FUNCTION : R/S*

DISPLAY : a= (intercept)

FUNCTION : R/S*

DISPLAY : b= (slope)

FUNCTION : R/S*

DISPLAY : an audible beep.

(4.6) To obtain the four "standard" curves (linear, logarithmic, exponential and power)

FUNCTION : XEQ c

DISPLAY : Same results as in step (4.5) for each of the four curves.

(4.7) To obtain the four standard curves plus others of the type :

$$LN(Y) = a + b\ (1/X)$$

$$1/Y = a + b\ (1/X)$$

$$1/Y = a + b\ LN(X)$$

FUNCTION : XEQ d

DISPLAY : results similar to those in step (4.5)

NOTE : In both step (4.6) and (4.7) the curve types shown last at the end of the routines or the ones at the time the routines are interrupted, are the currently selected curves, for which program R33 would perform additional stat. calculations.

(4.8) To calculate the estimates for the dependent variables, based on the currently chosen curve.

FUNCTION : XEQ E

DISPLAY : Xi?

INPUT : independent variable value.

FUNCTION : R/S

DISPLAY : Xj.= (estimated dependent variable)

(5) To go to regression phase 2.

(5.1) (Optional) To protect program R32 from being erased.

FUNCTION : GTO . .

(Optional) Load program R33 if not in RAM already.

INPUT : [ALPHA] R33 [ALPHA]

FUNCTION : XEQ READP

(5.2) To run the program program.

FUNCTION : XEQ R33

DISPLAY : ANOVA

FUNCTION : R/S*

DISPLAY : TOTAL CORR

FUNCTION : R/S*

DISPLAY : D.F.=(degree of freedom)

FUNCTION : R/S*

DISPLAY : SS=(corrected sum of squares)

FUNCTION : R/S*

DISPLAY : REGRESSION

FUNCTION : R/S*

DISPLAY : D.F. = 1

FUNCTION : R/S*

DISPLAY : SS= (sum of squares of regression)

FUNCTION : R/S*

DISPLAY : RESIDUAL

FUNCTION : R/S*

DISPLAY : D.F.= (degree of freedom)

FUNCTION : R/S*

DISPLAY : S^2 = (residual mean square)

FUNCTION : R/S*

DISPLAY : F = (F-statistic)

FUNCTION : R/S* (IMPORTANT : IF NO PRINTER IS USED DO NOT OMIT PRESSING THE R/S KEY)

DISPLAY : audible beep.

(5.3) To examine the standard errors for the regression coefficients.

FUNCTION : XEQ A

DISPLAY : SEa= (s.e.[a])

FUNCTION : R/S*

DISPLAY : SEb= (s.e.[b])

(5.4) To calculate the student-t statistics needed for interval estimations.

FUNCTION : XEQ B

DISPLAY : PROB.?

INPUT : error probability, fraction

FUNCTION : R/S

DISPLAY : T= (student-t)

(5.5) To calculate the confidence intervals, based on the most recent error probability supplied in step (5.4), for the regression coefficients.

FUNCTION : XEQ C

DISPLAY : aUL= (a, upper limit)

FUNCTION : R/S*

DISPLAY : a= (a)

FUNCTION : R/S*

DISPLAY : aLL= (a, lower limit)

FUNCTION : R/S*

DISPLAY : bUL= (b, upper limit)

FUNCTION : R/S*

DISPLAY : b= (b)

FUNCTION : R/S*

DISPLAY : bLL= (b, lower limit)

(5.6) To calculate the confidence interval for the projected values of the dependent variable, based on the most recent error probability supplied in step (5.4).

FUNCTION : XEQ D

DISPLAY : X?

INPUT : value of the independent variable

FUNCTION : R/S

DISPLAY : Y.UL= (upper limit of estimate)

FUNCTION : R/S*

DISPLAY : Y.= (Value of estimate)

FUNCTION : R/S*

DISPLAY : Y.LL= (lower limit of estimate)

MEMORY CONFIGURATION

REG #	CONTENT
00	sum of X
01	sum of X^2
02	sum of Y
03	sum of Y^2
04	sum of X*Y
05	number of data points.
06	data filename
07	overall # of data points
08	# of data points in RAM
09	# of variables.
10	used
11	depnd. var. #
12	indep. var. #
13	transf. code for depnd. var.
14	transf. code for indep. var.
15	a, intercept, used
16	b, slope, used
17	student-t, used
18	residual mean square, used
19-25	used

26	weight/freq. pointer
27	used
28	X1,1
29	X2,1
30	etc...

Flags 00 to 04 are used.

EXAMPLE

Given the following data tables,

X1	X2	X3	X4
1	.5	1	3
2	1	.25	5
5	5	.04	7
10	10	.01	4

Carry out the following :

(1) Set the size equal to 50.

(2) Load program LIB1 that contains the functions' library. Perform a GTO [.][.] to protect this program from being erased while loading other programs.

(3) Load program R31, with which you will key in the above data. Use (LBL A).

(4) Save the data on mass storage medium under the filename "F/W" and using a status filename of "SFW". Use (LBL B).

(5) Load program R32.

(6) Select a weighed regression. Use (LBL a).

(7) Select X1 as the independent variable, X2 as the dependent

variable and X3 as the weights.

(8) Inform the program that no transformations will be performed on the data (i.e fitting a straight line). Use (LBL B) but make no input.

(9) Perform the regression calculations . Use (LBL C).

(10) Now we want to perform a regression with no weights (default conditions). Reset to default conditions using (LBL R32).

(11) As before select X1 and X2 to be the independent and dependent variables, respectively.

(12) Since we are not going to transform the data this time either, we can go to the regression calculation step directly. Use (LBL C).

(13) Obtain an estimate for X2 when X1=3. Use (LBL E).

(13) We want to fit, with frequencies, an exponential curve. Set the regression type to use frequencies. Use (LBL b).

(14) Select X1, X2 and X4 as the independent, dependent and frequency variables, respectively.

(15) The X2 data will need to be transformed by taking their logarithms, while the X1 data will remain unchanged. Use (LBL B).

(16) Perform the regression calculations.

(17) Perform regression calculations using the four standard curves routine (LBL c).

(18) We want to obatin the ANOVA calculations for the last curve type obtained. Load program R33 and run.

(19) Obtain the student-t statistic at 95% confidence (or 5% risk). Use (LBL B).

(20) Calculate the confidence interval for the regression coefficients at 95% confidence. Use (LBL C).

(21) Obtain the standard errors for the regression coefficients. Use (LBL A).

EXAMPLE PRINTOUT

```
                SIZE 050

LIB1
                  READP
                 GTO ..
PACKING

R31
                  READP

                XEQ "R31"

                   XEQ A
NO. VARS?
                 4    RUN
X1,1?
                 1    RUN
X2,1?
                .5    RUN
X3,1?
                 1    RUN
X4,1?
                 3    RUN
                      RUN
X1,2?
                 2    RUN
X2,2?
                 1    RUN
X3,2?
               .25    RUN
X4,2?
                 5    RUN
                      RUN
X1,3?
                 5    RUN
X2,3?
                 5    RUN
X3,3?
               .04    RUN
X4,3?
                 7    RUN
                      RUN
X1,4?
                10    RUN
X2,4?
                10    RUN
X3,4?
               .01    RUN
X4,4?
                 4    RUN
                                    XEQ C
FILENAME?
F/W                                   RUN
STATUS?
SFW                                   RUN

R32
                                  READP

                                  XEQ a

                                   XEQ A
IND VAR?
                                 1    RUN
DPND VAR?
                                 2    RUN
W?
                                 3    RUN

                                   XEQ B
TRSF X2?
                                      RUN
TRSF X1?
                                      RUN

                                   XEQ C

 X2=a+b  X1

R2= 0.95985
a= -6.12245E-1
b= 1.02551E0
```

EXAMPLE PRINTOUT

```
                XEQ "R32"

                   XEQ A

IND VAR?

                 1   RUN

DPND VAR?

                 2   RUN

                   XEQ C

 X2=a+b  X1

R2= 0.99452
a= -7.65306E-1
b= 1.08673E0


                   XEQ E

X1?

                 3   RUN

X2.= 2.49490E0

                   XEQ b

                   XEQ A

IND VAR?

                 1   RUN

DPND VAR?

                 2   RUN

F?

                 4   RUN

                   XEQ B

TRSF X2?
LN
                     RUN
TRSF X1?

                     RUN
```

```
                                    XEQ C

LNX2=a+b  X1

R2= 0.85491
a= -5.19014E-1
b= 3.21116E-1

                                    XEQ c

 X2=a+b  X1

R2= 0.99261
a= -7.94638E-1
b= 1.09771E0


LNX2=a+b LNX1

R2= 0.98222
a= -7.91154E-1
b= 1.39624E0


 X2=a+b LNX1

R2= 0.90493
a= -1.06806E0
b= 4.25163E0


LNX2=a+b  X1

R2= 0.85491
a= -5.19014E-1
b= 3.21116E-1
```

EXAMPLE PRINTOUT

```
R33
                    READF

                XEQ "R33"
    ANOVA

TOTAL CORR
D.F. =3.
SS= 2.29679E1

REGRESSION
D.F. = 1
SS= 1.96354E1

RESIDUAL
D.F. =2.
SS= 3.33252E0

St2= 1.66626E0
F = 1.17841E1

                    XEQ B
PROB. ?
               .05    RUN
T= 2.11040E0

                    XEQ C
aUL= 5.88508E-1
a = -5.19014E-1
aLL= -1.62654E0

bUL= 5.18531E-1
b = 3.21116E-1
bLL= 1.23702E-1

                    XEQ A
SE a= 5.24793E-1
SE b= 9.35436E-2
```

SUMMARY OF COMMANDS

LBL R31 : I/O routines.

LBL A : Input data from keyboard.

LBL B : Load data from files.

LBL C : Save keyboard inputed data on file.

LBL R32 : Regression phase 1.

set default regression.

LBL A : Select variables.

LBL B : Variable transformations.

LBL C : Obtain basic results for one curve.

LBL E : Projections.

LBL a : set regression with weights.

LBL b : set regression with frequencies.

LBL c : Obtain basic results for the four "standard" curves.

LBL d : Same as above but for more curves

LBL R33 : ANOVA table.

LBL A : Standard errors.

LBL B : Calculate student-t.

LBL C : Confidence intervals for slope and intercept.

LBL D : Confidence intervals for projections.