Multiple Linear Regression for the HP–41C

by

Namir Shammas

This article presents an HP–41C program that performs multiple linear regression for two independent variables. The results include the regression ANOVA table and the confidence intervals for the regression slopes and intercept.

Usage

XEQ MRAT	Initialize the program
A	Add a data point
E	Clear statistical registers
D	Enter coefficients to calculate the inverse student–t statistic
B	Delete the last data point you entered
A	Delete a data point
B	Calculate regression coefficients and ANOVA table
C	Calculate the confidence intervals for the regression slopes and intercept. This option also calculates the inverse student–t statistic used in obtaining the confidence intervals.
C	Calculate the confidence intervals for the regression slopes and intercept. This option allows the user to supply the inverse student–t statistic used in obtaining the confidence intervals.

Example

Consider the following data:

X	Y	Z
1	1	10
2	4	8
3	9	6
4	16	5
5	25	4

Using the above data, calculate the regression coefficients and the ANOVA table. The Steps involved are (results are based on the FIX 5 and FIX 0 display formats):

Step	Task	Command/Input	Output
1	Initialize the program.	[XEQ] [ALPHA] MRAT [ALPHA]	0.00000
2	Add the first data point.	[A]	Z↗Y↗X?
3	Enter the first data point.	1 [ENTER] 1 [ENTER] 10 [R/S]	1.00000
4	Add the second data point.	[A]	Z↗Y↗X?
5	Enter the second data point.	2 [ENTER] 4 [ENTER] 8 [R/S]	2.00000
6	Repeat steps 4 and 5 to enter the remaining data points.		
7	Calculate the regression coefficients and the ANOVA table. Start with calculating the slope for variable X.	[B]	SLPX=–2.78571
8	Obtain the slope for variable Y.	[R/S]	SLPY=0.21429
9	Calculate the intercept.	[R/S]	INTRC=12.60000
10	Calculate the coefficient of determination.	[R/S]	R–SQR=0.99754
11	Obtain the number of observations.	[R/S]	N=5.00000
12	Obtain the sum of squares for the regression (SSR).	[R/S]	SSR=23.14286
13	Obtain the sum of squares for the residuals (i.e. errors) (SSE).	[R/S]	SSE=0.05714
14	Obtain the sum of squares for the total variation (SST).	[R/S]	SST=23.20000
15	Obtain the regression degrees of freedom.	[R/S]	DF REG=2.

Step	Task	Command/Input	Output
16	Obtain the residuals degrees of freedom.	[R/S]	DF ERR=2.
17	Obtain the total variation degrees of freedom.	[R/S]	DF TOT=4.
18	Obtain the mean regression sum of square (MSR).	[R/S]	MSR=11.57143
19	Obtain the mean residuals sum of square (MSE).	[R/S]	MSE=0.02857
20	Obtain the F statistic.	[R/S]	F=405.00005
21		[R/S]	(audible beep)

Next, calculate the standard error and confidence intervals for the regression coefficients. The Steps involved are:

Step	Task	Command/Input	Output
1	Calculate standard error for the slope (using student–t = 4.30265 to obtain the confidence interval at 95% confidence and 2 degrees of freedom).	[C]	T?
2	Enter the inverse student-t value.	4.30265 [R/S]	SLX SE=0.27627
3	Calculate the lower limit for the slope of variable X.	[R/S]	SLX LL=-3.97442
4	Calculate the upper limit for the slope of variable X.	[R/S]	SLX UL=-1.59701
5	Calculate the standard error for the slope of variable Y.		SLY SE=0.04518
6	Calculate the lower limit for the slope or variable Y.	[R/S]	SLY LL=0.01991
7	Calculate the upper limit for the slope or variable X.	[R/S]	SLY UL=0.40866
8	Calculate the standard error for the intercept.	[R/S]	INT SE=0.36253
9	Calculate the lower limit for the intercept.	[R/S]	INT LL=11.04016
10	Calculate the upper limit for the intercept.	[R/S]	INT UL=14.15984
11	End the output sequence.	[R/S]	(audible beep)

Finally, recalculate the standard error and confidence intervals for the regression coefficients using the built–in approximation for the inverse student–t statistic. The Steps involved are:

Step	Task	Command/Input	Output
1	Enter the empirical coefficients for the built–in approximation for the inverse student–t statistic	[D]	A↗B↗C?
2	Enter the actual values for coefficients A, B, and C for the 95% confidence level (same as the 0.05 significance level).	0.672951400 [ENTER] 1.208789 [ENTER] 0.734348 [R/S]	0.67295
3	Calculate standard error for the slope of variable X.	[C]	SLX SE=0.27627
4	Calculate the lower limit for the slope of variable X.	[R/S]	SLX LL=-3.97644
5	Calculate the upper limit for the slope or variable X.	[R/S]	SLX UL=–1.59499
6	Calculate the standard error for the slope of variable Y.		SLY SE=0.04518
7	Calculate the lower limit for the slope or variable Y.	[R/S]	SLY LL=0.01958
8	Calculate the upper limit for the slope or variable X.	[R/S]	SLY UL=0.40899
9	Calculate the standard error for the intercept.	[R/S]	INT SE=0.36253
10	Calculate the lower limit for the intercept.	[R/S]	INT LL=11.03750
11	Calculate the upper limit for the intercept.	[R/S]	INT UL=14.16250
12	End the output sequence.	[R/S]	(audible beep)

Here are the regression results in a table (generated using Excel):

Regression Results

N	5			
R-Sqr	0.997536946			
ANOVA Table				
Source of variation	SS	DF	MS	F
Regression	23.14285714	2	11.57142857	405
Residual	0.057142857	2	0.028571429	
Total	23.2	4		
	Coefficient	StdErr	95% Low Limit	95% Upper Limit
Intercept	12.6	0.362530787	11.04015592	14.15984408
SlopeX	-2.785714286	0.276272566	-3.974419195	-1.597009376
SlopeY	0.214285714	0.045175395	0.019911677	0.408659752

Algorithms

Statistical Summations

$\sum x$ = sum of x
$\sum x^2$ = sum of x^2
$\sum y$ = sum of y
$\sum y^2$ = sum of y^2
$\sum z$ = sum of z
$\sum z^2$ = sum of z^2
$\sum xy$ = sum of $x \cdot y$
$\sum xz$ = sum of $x \cdot z$
$\sum yz$ = sum of $y \cdot z$
n = number of observations

Regression Coefficients

$xm = \sum x / n$
$ym = \sum y / n$
$zm = \sum z / n$
$Sxx = \sum x2 - (\sum x)\ 2 / n = \sum x2 - n\ (xm)2$
$Syy = \sum y2 - (\sum y)\ 2 / n = \sum y2 - n\ (ym)2$
$Szz = \sum z2 - (\sum z)\ 2 / n = \sum z2 - n\ (zm)2$
$Sxy = \sum xy - (\sum x)\ (\sum y)/ n = \sum xy - n\ xm\ ym$
$Sxz = \sum xz - (\sum x)\ (\sum z)/ n = \sum xz - n\ xm\ zm$
$Syz = \sum yz - (\sum y)\ (\sum z)/ n = \sum yz - n\ ym\ zm$
$D = Sxx\ Syy - (Sxy)2$

Slope b = [Sxz Syy – Syz Sxy] / [Sxx Syy – (Sxy)2] = [Sxz Syy – Syz Sxy] / D
Slope c = [Syz Sxx – Sxz Sxy] / [Sxx Syy – (Sxy)2] = [Syz Sxx – Sxz Sxy] / D
Intercept a = zm – b xm – c ym
Model is Z = a + b X + c Y
R2 = [a ∑z + b ∑xz + c ∑yz – (∑z) 2 / n] / [∑z2 – (∑z) 2 / n]
= [a ∑z + b ∑xz + c ∑yz – (∑z) 2 / n] / Szz
MSR = SSR / 2
MSE = SSE / (n – 3)

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F0
Regression	SSR = b Sxz + c Syz	2	MSR	MSR / MSE
Residual	SSE = Szz – b Sxz - c Syz	n – 3	MSE	
Total	SST = Szz	n – 1		

Interval for Slope b

At 100(1 – α) confidence:
SlopeStdErr = √ [MSE (Syy / D)]
Δb = $t_{\alpha/2,n-3}$ · SlopeStdErr
Confidence Interval = b ± $t_{\alpha/2,n-3}$ · √ [MSE (Syy / D)]

Interval for Slope c

At 100(1 – α) confidence:
SlopeStdErr = √ [MSE (Sxx / D)]
Δc = $t_{\alpha/2,n-3}$ · SlopeStdErr
Confidence Interval = c ± $t_{\alpha/2,n-3}$ · √ [MSE (Sxx / D)]

Interval for Intercept a

At 100(1 – α) confidence:
IntStdErr = √ (MSE [1/n + (Syy / D)(xm) 2 + (Sxx / D)(ym) 2 – 2(Sxy / D) xm ym])
= √ (MSE [1/n +{Syy (xm) 2 + Sxx (ym) 2 – 2(Sxy xm ym)}/D])
Δa = $t_{\alpha/2,n-3}$ · IntStdErr
Confidence Interval = a ± $t_{\alpha/2,n-3}$ · √ (MSE [1/n +{Syy (xm) 2 + Sxx (ym) 2 – 2(Sxy xm ym)}/D])

The Inverse Student–t Probability Distribution Function

You can also calculate the inverse two–tailed Student–t value using well known approximations. Recently, I was able to obtain a set of approximations that fits the inverse two–tailed Student–t and the degrees of freedom using the following model:

$$Tinv = \exp(A + B / df + C / df^2)$$

The following table shows the values for the constant A, B, and C for different values of the significance level, α:

Significance Level α	A	B	C
0.200	0.248069936	0.660674	0.226537
0.150	0.364320592	0.767873	0.308868
0.100	0.497661825	0.925738	0.445297
0.050	0.672951400	1.208789	0.734348
0.025	0.807141675	1.503440	1.093993

Label D of the program LRAT allows you to enter the values for A, B, and C. you select a particular set of these coefficients that corresponds to the significance level for the two–sided inverse student–t statistic.

Following a more traditional approach, here is the table for the inverse two–tailed Student–t probability distribution function. The last row of the table contains values for the inverse normal probability distribution function.

Degrees of Freedom	α = **0.100**	α = **0.050**	α = **0.025**	α = **0.010**
1	6.314	12.706	25.452	63.657
2	2.920	4.303	6.205	9.925
3	2.353	3.182	4.177	5.841
4	2.132	2.776	3.495	4.604
5	2.015	2.571	3.163	4.032
6	1.943	2.447	2.969	3.707
7	1.895	2.365	2.841	3.499
8	1.860	2.306	2.752	3.355
9	1.833	2.262	2.685	3.250
10	1.812	2.228	2.634	3.169
11	1.796	2.201	2.593	3.106
12	1.782	2.179	2.560	3.055
13	1.771	2.160	2.533	3.012
14	1.761	2.145	2.510	2.977
15	1.753	2.131	2.490	2.947
16	1.746	2.120	2.473	2.921
17	1.740	2.110	2.458	2.898
18	1.734	2.101	2.445	2.878
19	1.729	2.093	2.433	2.861
20	1.725	2.086	2.423	2.845

Degrees of Freedom	α = 0.100	α = 0.050	α = 0.025	α = 0.010
21	1.721	2.080	2.414	2.831
22	1.717	2.074	2.405	2.819
23	1.714	2.069	2.398	2.807
24	1.711	2.064	2.391	2.797
25	1.708	2.060	2.385	2.787
26	1.706	2.056	2.379	2.779
27	1.703	2.052	2.373	2.771
28	1.701	2.048	2.368	2.763
29	1.699	2.045	2.364	2.756
30	1.697	2.042	2.360	2.750
31	1.696	2.040	2.356	2.744
32	1.694	2.037	2.352	2.738
33	1.692	2.035	2.348	2.733
34	1.691	2.032	2.345	2.728
35	1.690	2.030	2.342	2.724
36	1.688	2.028	2.339	2.719
37	1.687	2.026	2.336	2.715
38	1.686	2.024	2.334	2.712
39	1.685	2.023	2.331	2.708
40	1.684	2.021	2.329	2.704
50	1.676	2.009	2.311	2.678
60	1.671	2.000	2.299	2.660
70	1.667	1.994	2.291	2.648
80	1.664	1.990	2.284	2.639
90	1.662	1.987	2.280	2.632
100	1.660	1.984	2.276	2.626
Infinity	1.645	1.960	2.241	2.576

Memory Map

R00 = R^2
R01 = Z, intercept
R02 = X, slope X
R03 = Y, slope Y
R04 = ΣX
R05 = ΣX^2
R06 = ΣY
R07 = ΣY^2
R08 = ΣXY
R09 = n
R10 = ΣZ
R11 = ΣZ^2

R12 = ΣXZ
R13 = ΣYZ
R14 = x-mean
R15 = y-mean
R16 = z-mean
R17 = Sxx
R18 = Syy
R19 = Szz
R20 = Sxy
R21 = Sxz
R22 = Syz
R23 = D
R24 = SSR, MSR
R25 = SSE, MSE
R26 = StdErr for regression coefficient
R27 = A
R28 = B
R29 = C
R30 = inverse student–t
R31 = n – 3
R32 = n – 1

Source Code

The source code for the program appears below. Please note the following:

- Text appearing in a pair of double quotes represents characters in the Alpha register.
- The blank lines are intentionally inserted to separate logical blocks of commands:

Program Step	***Comment***
LBL MRAT	
LBL E	Reset statistical registers location, clear registers, clear the stack, and clear flag 00
ΣREG 04	
CLREG	
CLST	
CF 00	
RTN	
LBL A	add a data point
X↗Y↗Z?	
PROMPT	
STO 01	

Program Step	Comment
R↓	
STO 03	
R↓	
STO 02	
R↓	
R↓	
ST+ 10	
ST* Y	
ST* Z	
X^2	
ST+ 11	
R↓	
ST+ 13	
R↓	
ST+ 12	
RCL 03	
RCL 02	
Σ+	
SF 00	
RTN	
LBL b	delete last entry
FC? 00	flag 0 clear—>last entry has already been deleted
GTO a	prompt for data point
RCL 02	recall last X, Y, and Z to the stack
RCL 03	
RCL 01	
GTO 02	
LBL a	remove a data point
X↗Y↗Z?	
PROMPT	
STO 01	
R↓	
STO 03	
R↓	
STO 02	
R↓	
R↓	
LBL 02	

Program Step	***Comment***
ST– 10	
ST* Y	
ST* Z	
X^2	
ST– 11	
R↓	
ST– 13	
R↓	
ST– 12	
RCL 03	
RCL 02	
Σ–	
CF 00	
RTN	
LBL D	Prompt used to enter coefficients for the inverse student-t approximation
A↗B↗C?	
PROMPT	
STO 29	
R↓	
STO 28	
R↓	
STO 27	
RTN	
LBL 01	display content of X register
ARCL X	
PROMPT	
RTN	
LBL B	calculate regression coefficients and ANOVA table
CF 00	
RCL 09	
1	
–	
STO 32	calculate and store n–1
2	
–	
STO 31	calculate and store n–3
MEAN	

Program Step	Comment
STO 14	store xmean
X<>Y	
STO 15	store ymean
RCL 10	
RCL 09	
/	
STO 16	calculate and store zmean
SDEV	
X^2	
X<>Y	
X^2	
RCL 32	
ST* Z	calculate Sxx
*	
STO 18	calculate and store Syy
X<>Y	
STO 17	store Sxx
RCL 11	
RCL 10	
X^2	
RCL 09	
/	
–	
STO 19	calculate and store Szz
RCL 08	
STO 20	
MEAN	
*	
RCL 09	
*	
ST– 20	calculate and store Sxy
RCL 12	
STO 21	
MEAN	
RCL 10	
*	
ST– 21	calculate and store Sxz
RCL 13	
STO 22	
MEAN	
X<>Y	

Program Step	***Comment***
RCL 10	
*	
ST– 22	calculate and store Syz
RCL 17	
RCL 18	
*	
RCL 20	
X^2	
–	
STO 23	calculate and store D
RCL 21	
RCL 18	
*	
RCL 22	
RCL 20	
*	
–	
RCL 23	
/	
STO 02	calculate and store slope X
"SLP X="	
XEQ 01	
RCL 22	
RCL 17	
*	
RCL 21	
RCL 20	
*	
–	
RCL 23	
/	
STO 03	calculate and store slope Y
"SLP Y="	
XEQ 01	
RCL 16	
RCL 02	
RCL 14	
*	
–	
RCL 03	
RCL 15	

Program Step	Comment
*	
–	
STO 01	calculate and store intercept
"INTC="	
XEQ 01	
RCL 01	
RCL 10	
*	
RCL 02	
RCL 12	
*	
+	
RCL 03	
RCL 13	
*	
+	
RCL 10	
X^2	
RCL 09	
/	
STO 00	store value of $(\Sigma z)^2$ / n for reuse next
–	
RCL 11	
RCL 00	
–	
/	
STO 00	calculate and store R^2
"R2="	
XEQ 01	
RCL 09	recall the number of observations
"N="	
XEQ 01	
RCL 02	
RCL 21	
*	
RCL 03	
RCL 22	
*	
+	
STO 24	calculate and store SSR
"SSR="	

Program Step	***Comment***
XEQ 01	
CHS	
RCL 19	
+	
STO 25	calculate and store SSE
"SSE="	
XEQ 01	
RCL 19	
"SST="	
XEQ 01	
FIX 0	
2	
"DF REG="	
XEQ 01	
RCL 31	
"DF ERR="	
XEQ 01	
RCL 32	
"DF TOT="	
XEQ 01	
FIX 5	reset the display. You can customize this step
RCL 24	
2	
/	
STO 24	calculate and store MSR
"MSR="	
XEQ 01	
RCL 25	
RCL 31	
/	
STO 25	calculate and store MSE
"MSE="	
XEQ 01	
/	
"F="	
XEQ 01	
BEEP	
RTN	
lbl c	prompt user for two–sided inverse student–t
"T?"	

Program Step	Comment
PROMPT	
GTO 00	resume at calculations for standard error and confidence interval for slopes and intercept
LBL C	calculate standard error and confidence interval for slopes and intercept
RCL 27	start by calculating approximation for two–sided inverse student–t
RCL 28	
RCL 29	
RCL 31	
ST/ Z	
X^2	
/	
+	
+	
EXP	
Lbl 00	
STO 30	store the two–sided inverse student–t
RCL 18	
RCL 23	
/	
RCL 25	
*	
SQRT	
STO 26	calculate and store std error for slope X
"SLX SE="	
XEQ 01	
RCL 30	
*	
CHS	
RCL 02	
+	calculate lower limit for the slope X confidence interval
"SLX LL="	
XEQ 01	
RCL 02	
RCL 26	
RCL 30	
*	
+	calculate upper limit for the slope X confidence interval
"SLX UL="	
XEQ 01	

Program Step	Comment
RCL 17	
RCL 23	
/	
RCL 25	
*	
SQRT	
STO 26	calculate and store std error for slope Y
"SLY SE="	
XEQ 01	
RCL 30	
*	
CHS	
RCL 03	
+	calculate lower limit for the slope Y confidence interval
"SLY LL="	
XEQ 01	
RCL 03	
RCL 26	
RCL 30	
*	
+	calculate upper limit for the slope Y confidence interval
"SLY UL="	
XEQ 01	
MEAN	
*	
RCL 20	
*	
–2	
*	
RCL 18	
RCL 14	
X^2	
*	
+	
RCL 17	
RCL 15	
X^2	
*	
+	
RCL 23	
/	

Program Step	Comment
RCL 09	
1/X	
+	
RCL 25	
*	
SQRT	
STO 26	calculate and store std error for intercept
"INT SE=	
XEQ 01	
RCL 30	
*	
CHS	
RCL 01	
+	calculate lower limit for the intercept's confidence interval
"INT LL="	
XEQ 01	
RCL 01	
RCL 26	
RCL 30	
*	
+	calculate upper limit for the intercept's confidence interval
"INT UL="	
XEQ 01	
BEEP	
RTN	

Note: You can insert additional code in labels A and a to transform the input values for the X, Y and Z variables before updating the statistical summations. Keep in mind that in such case, the regression results, ANOVA table, and other statistics describe the transformed variables and not the original data.