Linear Regression for the HP-67

by

Namir Shammas

This article presents an HP-67 program that performs linear regression. The results include the regression ANOVA table and the confidence intervals for the regression slope and intercept.

Usage

A	Add a data point
E	Clear statistical registers
D	Enter coefficients to calculate the inverse student-t statistic
f A	Delete a data point
B	Calculate regression coefficients and ANOVA table
C	Calculate the confidence intervals for the regression slope and intercept. This option also calculates the inverse student-t statistic used in obtaining the confidence intervals.
f C	Calculate the confidence intervals for the regression slope 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
1	1
2	4
3	9

X	Y
4	16
5	25

Using the above data, calculate the regression coefficients and the ANOVA table. The Steps involved are:

Step	Task	Command/Input	Output
1	Initialize the program.	[E]	(not relevant)
2	Add the first data point.	1 [ENTER] 1 [A]	1.00000
3	Add the second data point.	4 [ENTER] 2 [A]	2.00000
4	Repeat step 3 to enter the remaining data points.	9 [ENTER] 3 [A] 16 [ENTER] 4 [A] 25 [ENTER] 5 [A]	5.00000
5	Calculate the regression coefficients and the ANOVA table. Start with calculating the slope.	[B]	6.00000
6	Calculate the intercept.	[R/S]	–7.00000
7	Calculate the coefficient of determination.	[R/S]	0.96257
8	Obtain the number of observations.	[R/S]	5.00000
9	Obtain the sum of squares for the regression (SSR).	[R/S]	360.00000
10	Obtain the sum of squares for the residuals (i.e. errors) (SSE).	[R/S]	14.00000
11	Obtain the sum of squares for the total variation (SST).	[R/S]	374.00000
12	Obtain the regression degrees of freedom.	[R/S]	1.00000
12	Obtain the residuals degrees of freedom.	[R/S]	3.00000
13	Obtain the total variation degrees of freedom.	[R/S]	4.00000
14	Obtain the mean regression sum of square (MSR).	[R/S]	360.00000
15	Obtain the mean residuals sum of square (MSE).	[R/S]	4.66667
16	Obtain the F statistic.	[R/S]	77.14286
17	End the calculations.	[R/S]	77.14286 (blinks)

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 = 3.18245 to obtain the confidence interval at 95% confidence and 3 degrees of freedom).	3.18245 [f] [C]	0.68313
2	Calculate the lower limit for the slope.	[R/S]	3.82598
3	Calculate the upper limit for the slope.	[R/S]	8.17402
4	Calculate the standard error for the intercept.	[R/S]	2.26569
5	Calculate the lower limit for the intercept.	[R/S]	–14.21042
6	Calculate the upper limit for the intercept.	[R/S]	0.21042
7	End the calculations.	[R/S]	0.21042 (blinks)

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 A, B, and C for the built-in approximation for the inverse student-t statistic. In this case, enter the set for the 0.05 significance level.	0.672951400 [ENTER] 1.208789 [ENTER] 0.734348 [D]	0.67295
2	Start the statistical calculations for the standard errors and confidence intervals for the regression coefficients. Start by obtaining the standard error for the slope.	[C]	0.68313
3	Calculate the lower limit for the slope.	[R/S]	3.82636
4	Calculate the upper limit for the slope.	[R/S]	8.17363

Step	Task	Command/Input	Output
5	Calculate the standard error for the intercept.	[R/S]	2.26569
6	Calculate the lower limit for the intercept.	[R/S]	–14.20916
7	Calculate the upper limit for the intercept.	[R/S]	0.20916
8	End the calculations.	[R/S]	0.20916 (blinks)

Here are the regression results in a table:

Regression Results

N	5
R-Sqr	0.962566845

ANOVA Table

Source of variation	SS	DF	MS	F
Regression	360	1	360	77.14285714
Residual	14	3	4.666666667	
Total	374	4		
	Coefficient	StdErr	95% Low Limit	95% Upper Limit
Intercept	-7	2.265686062	-14.21042424	0.210424237
Slope	6	0.683130051	3.825975293	8.174024707

Algorithms and Equations

Statistical Summations

$\sum x$ = sum of x

$\sum x^2$ = sum of x^2

$\sum y^2$ = sum of y

$\sum y^2$ = sum of y^2

$\sum xy$ = sum of $x \cdot y$

n = number of observations

Regression Coefficients

$x_m = \sum x / n$

$y_m = \sum y / n$

$S_{xx} = \sum x^2 - (\sum x)^2 / n = \sum x^2 - n\,(x_m)^2$

$S_{yy} = \sum y^2 - (\sum y)^2 / n = \sum y^2 - n\,(y_m)^2$

$S_{xy} = \sum xy - (\sum x)(\sum y)/ n = \sum xy - n\,x_m\,y_m$

Slope $B = S_{xy} / S_{xx} = (\sum xy - n\, x_m\, y_m) / (\sum x^2 - n\, (x_m)^2)$

Intercept $A = y_m - B\, x_m$

For line: $y = A + B\, x$

ANOVA Table

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F_0*
Regression	$SS_R = B\, S_{xy}$	1	MS_R	MS_R / MS_E
Residual	$SS_E = S_{yy} - B\, S_{xy}$	$n - 2$	MS_E	
Total	$SS_T = S_{yy}$	$n - 1$		

Interval for Slope

At $100(1 - \alpha)$ confidence:

$SlopeStdErr = \sqrt{(MS_E / S_{xx})}$

Confidence interval for slope $= B \pm \Delta B\, t_{\alpha/2,n-2} \cdot \sqrt{(MS_E / S_{xx})}$

Interval for Intercept

At $100(1 - \alpha)$ confidence:

$IntStdErr = \sqrt{(MS_E\, [1/n + (x_m)^2 / S_{xx}])}$

Confidence interval for slope $= A \pm t_{\alpha/2,n-2} \cdot \sqrt{(MS_E\, [1/n + (x_m)^2 / S_{xx}])}$

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
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

Degrees of Freedom	α = 0.100	α = 0.050	α = 0.025	α = 0.010
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

R0 = xmean
R1 = ymean
R2 = Sxx
R3 = Syy
R4 = Sxy
R5 = Slope
R6 = Intercept
R7 = Student-t t(alpha/2,n-2)
R8 = StdErr for slope, StdErr for Intercept
R9 = used
SR0 = coefficient A used to calculate student-t
SR1 = coefficient B used to calculate student-t
SR2 = coefficient C used to calculate student-t
SR3 =
SR4 = $\sum x$
SR5 = $\sum x^2$
SR6 = $\sum y$
SR7 = $\sum y^2$
SR8 = $\sum xy$
SR9 = n
RA = SSR, MR
RB = SSE, ME
RC = used
RD = n

RE = n-1, n-2

Source Code

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

- The blank lines are intentionally inserted to separate logical blocks of commands:

Program Step	*Comment*
LBL A	Add data
Σ+	Can insert data transformation steps after LBL A
RTN	
LBL a	Remove data
Σ−	Can insert data transformation steps after LBL a
RTN	
LBL E	Clear the registers
CLRREG	
RTN	
LBL D	Store coefficients for the approximation of student−t
P<>S	
STO 2	
R↓	
STO 1	
R↓	
STO 0	
P<>S	
−x−	
RTN	
LBL B	Calculate regression statistics and ANOVA table
MEAN	get the means
STO 0	
X<>Y	
STO 1	
P<>S	
RCL 9	
P<>S	
STO D	Store n
1	
−	
STO E	Calculate and store n−1
SDEV	Get the standard deviation

Program Step	*Comment*
X^2	
RCL E	
*	
STO 2	Calculate and store Sxx
X<>Y	
X^2	
RCL E	
*	
STO 3	Calculate and store Syy
P<>S	
RCL 8	
RCL 4	
RCL 6	
*	
RCL 9	
/	
-	
P<>S	
STO 4	Calculate and store Sxy
RCL 2	
/	
STO 5	Calculate and store the slope
R/S	Display the slope
MEAN	
RCL 5	
*	
-	
STO 6	Calculate and store the intercept
R/S	Display the intercept
RCL 5	
RCL 4	
*	
STO A	Calculate and store SSR
RCL 3	
/	
R/S	Calculate and display R-Sqr
RCL D	
R/S	Display n
RCL A	
R/S	Display SSR
CHS	
RCL 3	
+	
STO B	Calculate and store SSE

Program Step	*Comment*
R/S	Display SSE
RCL 3	Display SST
R/S	
1	
R/S	Display DF for regression
RCL D	
2	
-	
STO E	Store n-2
R/S	Display DF for errors
+	
R/S	Display DF for total variation
RCL A	
R/S	Display MSR
RCL B	
RCL E	
/	
STO B	Calculate and store MSE
R/S	
RCL A	
X<>Y	
/	
R/S	Calculate and display F
-x-	Blink to indicate end of calculations
RTN	
LBL C	Calculate confidence intervals for slope and intercept
P<>S	
RCL 9	
2	
-	
STO C	
RCL 0	
RCL 1	
RCL C	
/	
+	
RCL 2	
RCL C	
X^2	
/	
+	
EXP	Calculate approximation to inverse student-t

Program Step	*Comment*
P<>S	
LBL c	
STO 7	Store the inverse student-t value
RCL B	
RCL 2	
/	
SQRT	
STO 8	Calculate and store standard error for the slope
R/S	Display standard error for the slope
RCL 7	
*	
STO 9	Calculate and store confidence limit difference
CHS	
RCL 5	
+	
R/S	Display lower limit for confidence interval for slope
RCL 5	
RCL 9	
+	
R/S	Display lower limit for confidence interval for slope
RCL D	
1/X	
RCL 0	
X^2	
RCL 2	
/	
+	
RCL B	
*	
SQRT	
STO 8	Calculate and store standard error for the intercept
R/S	Display standard error for the intercept
RCL 7	
*	
STO 9	Calculate and store confidence limit difference
CHS	
RCL 6	
+	
R/S	Display lower limit for confidence interval for intercept
RCL 6	

Program Step	*Comment*
RCL 9	
+	
R/S	**Display lower limit for confidence interval for intercept**
-x-	**Blink to indicate end of calculations**
RTN	

Note: You can insert additional code in labels A and a to transform the X and Y values before the Σ+ or Σ- command. Keep in mind that in such case, the regression results, ANOVA table, and other statistics describe the transformed variables and not the original data.