Compare Linear Regression Lines for the HP-67

by

Namir Shammas

This article presents an HP-67 program that calculates the linear regression statistics for two data sets and then compares their slopes and intercept.

Usage

CLREG	Clear the registers to initialize the program.
A	Add a data point for set 1.
[f] A	Delete a data point from set 1.
B	Add a data point for set 2.
[f] B	Delete a data point from set 2.
C	Calculate the slope, intercept, R^2 value, SSE, and standard error for the slope, for data set1.
D	Calculate the slope, intercept, R^2 value, SSE, and standard error for the slope, for data set2.
E	Calculate the student-t values for differences between the slopes and between the intercepts.

Example

Consider the following data for set 1:

X1	Y1
1	1
2	8
3	9
4	16
5	25

And the data for set 2:

X2	Y2
1	1.1
2	8
3	9
4	16
5	25

Using the above data, first calculate the regression coefficients for set 1 and set 2. The way the program is written, you should not disturb the stack when viewing intermediate results as they are needed for subsequent calculations. Wait until each set of calculations finishes before you can recall values from different registers into the stack. The Steps involved are (using a DSP 5 and FIX mode).

Step	Task	Command/Input	Output
1	Initialize the program.	[CL REG]	READY
2	Add the first data point for set 1.	1 [ENTER] 1 [A]	1.00000
3	Enter the second data point for set 1.	8 [ENTER] 2 [A]	2.00000
4	Repeat step 3 to enter the remaining data points in data set 1.	25 [ENTER] 5 [A]	5.00000
5	Add the first data point for set 2.	1.1 [ENTER] 1 [B]	1.00000
6	Enter the second data point for set 2.	4 [ENTER] 2 [B]	2.00000
7	Repeat step 6 to enter the remaining data points in data set 2.	25 [ENTER] 5 [B]	5.00000
8	Calculate the regression coefficients and other statistics for set 1. Calculate the slope.	[C]	5.60000
9	Calculate the intercept	[R/S]	–5.00000
10	Calculate the coefficient of determination.	[R/S]	0.94800
11	Calculate the SSE value.	[R/S]	17.20000
12	Calculate the standard error for the slope	[R/S]	0.75719
13	Complete the calculation run.	[R/S]	Blinking 0.00000
14	Calculate the regression coefficients and other statistics for set 2. Calculate the slope.	[D]	5.98000
15	Calculate the intercept	[R/S]	–6.92000
16	Calculate the coefficient of determination.	[R/S]	0.96128
17	Calculate the SSE value.	[R/S]	14.40400

Step	Task	Command/Input	Output
18	Calculate the standard error for the slope	[R/S]	0.69292
19	Complete the calculation run.	[R/S]	Blinking 0.00000

Next, calculate the student-t for the differences in the slope and the intercept:

Step	Task	Command/Input	Output
1	Calculate the student-t for the difference in slopes.	[E]	0.37023
2	Calculate the student-t for the difference in intercepts.	[R/S]	0.56402
3	Complete the calculation run.	[R/S]	Blinking 0.00000

The (absolute) values for the two calculated t statistics of 0.37023 and 0.56402 should be compared with $t_{0.05,6}$ = 2.447. As such, we cannot reject the hypothesis that the slopes of the two data sets are statistically different. Likewise, we cannot reject the hypothesis that the intercepts of the two data sets are statistically different.

Here are the regression results in tabular form (generated using Excel):

Regression Results for Data Set 1

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

Regression Results for Data Set 2

N	5
R-Sqr	0.961280403

ANOVA Table

Source of variation	SS	DF	MS	F
Regression	357.604	1	357.604	74.4801444
Residual	14.404	3	4.801333333	
Total	372.008	4		
	Coefficient	StdErr	95% Low Limit	95% Upper Limit
Intercept	-6.92	2.298144179	-14.23372045	0.39372045
Slope	5.98	0.692916541	3.774830313	8.185169687

Algorithms

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 $A = y_m - B\, x_m$
$R^2 = B (S_{xy} / S_{yy})$
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/Error	$SS_E = S_{yy} - B\, S_{xy}$	$n - 2$	MS_E	
Total	$SS_T = S_{yy}$	$n - 1$		

Comparing Slopes

$Sy.x^2 = (SSE_1 + SSE_2) / (n_1 + n_2 - 4)$

$SE_{slope.diff} = \sqrt{[Sy.x^2 (1/ SSE_1 + 1/SSE_2)]}$

$t_{slope} = (slope_1 - slope_2) / SE_{slope.diff}$

Comparing Intercepts

$SE_{intercept.diff} = \sqrt{[Sy.x^2 (1/n_1 + 1/n_2 + M1^2/SSE_1 + M2^2/SSE_2)]}$

$t_{intercept} = (intercept_1 - intercept_2) / SE_{intercept.diff}$

The Inverse Student-t Table

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

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

Degrees of Freedom	α = 0.100	α = 0.050	α = 0.025	α = 0.010
100	1.660	1.984	2.276	2.626
Infinity	1.645	1.960	2.241	2.576

Memory Map

R0= Intercept2
R1= Slope2
R2= SE slope2
R3= SSE1
R4= sum of x2
R5= sum of x2 squared
R6= sum of y2
R7= sum of y2 squared
R8= sum of x2 * y2
R9= n2

SR0= Intercept1
SR1= Slope1
SR2= SE slope1
SR3= SSE1
SR4= sum of x1
SR5= sum of x1 squared
SR6= sum of y1
SR7= sum of y1 squared
SR8= sum of x1 * y1
SR9= n1

RA= x mean, Sy.x
RB= y mean, Sxx1
RC= Sxx, Sxx2
RD= Syy, std err of the difference between slopes
RE= Sxy, used
RI= used

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 a data point from the first data set
Σ+	

Program Step	***Comment***
RTN	
LBL a	remove a data point from the first data set
Σ–	
RTN	
LBL B	add a data point from the second data set
P<>S	
Σ+	
P<>S	
RTN	
LBL b	remove a data point from the second data set
P<>S	
Σ–	
P<>S	
RTN	
LBL 0	helper subroutine to calculate Sxx and Syy
X^2	
P<>S	
RCL 9	
P<>S	
1	
-	
*	
RTN	
LBL 1	helper subroutine to calculate x.mean^2 / Sxx + 1/n for either data set
MEAN	
X^2	
STO E	
SDEV	
GSB 0	
1/X	
RCL E	
*	
P<>S	
RCL 9	
P<>S	
1/X	

Program Step	Comment
+	
RTN	
LBL C	calculate linear regression coeffcients for 1st data sets
MEAN	
STO A	calculate and store x mean
X<>Y	
STO B	calculate and store y mean
SDEV	
GSB 0	
STO C	calculate and store Sxx
X<>Y	
GSB 0	
STO D	calculate and store Syy
MEAN	
*	
P<>S	
RCL 9	
*	
CHS	
RCL 8	
+	
STO E	calculate and store Sxy
RCL C	
/	
STO 1	calculate and store slope
R/S	display slope
RCL A	
*	
CHS	
RCL B	
+	
STO 0	calculate and store intercept
R/S	display intercept
RCL 1	
RCL E	
RCL D	
/	
*	
R/S	display R-Sqr (value is not stored!)
RCL D	

Program Step	Comment
RCL E	
RCL 1	
*	
–	
STO 3	calculate and store SSE
R/S	display SSE
RCL 9	
2	
–	
/	calculate MSE
RCL C	
/	
SQRT	
STO 2	calculate and store SE slope
P<>S	
R/S	
0	
-x-	signal end of regression
RTN	
LBL D	calculate linear regression coeffcients for 2nd data sets
P<>S	
GSB C	
P<>S	
RTN	
LBL E	compare slopes and intercepts
RCL 3	get SSE2
P<>S	
RCL 3	get SSE1
+	
RCL 9	get n1
P<>S	
RCL 9	get n2
+	
4	
–	
/	
STO A	calculate and store (Sy.x)^2
SDEV	
GSB 0	

Program Step	***Comment***
STO B	calculate and store Sxx1
P<>S	
SDEV	
GSB 0	
STO C	calculate and store Sxx2
P<>S	
RCL B	
1/X	
RCL C	
1/X	
+	
RCL A	
*	
SQRT	
STO D	calculate and store std err of the difference between slopes
1/X	
RCL 1	get slope2
P<>S	
RCL 1	get slope1
P<>S	
–	
*	
R/S	display student-t statistic for slope difference
GSB 1	calculate x1.mean^2 / Sxx1 + 1/n1
ST I	Store intermediate result in register I (using it as a regular register)
P<>S	
GSB 1	calculate x2.mean^2 / Sxx2 + 1/n2
P<>S	
RC I	Recall intermediate result from register I
+	
RCL A	
*	
SQRT	calculate pooled std dev for slopes
1/X	
P<>S	
RCL 0	get Intercept1
P<>S	
RCL 0	get Intercept2
–	calculate difference in slops
*	multiply by 1/s<intercept1 - intercept2>
R/S	calculate and display student-t for slope difference

Program Step	Comment
0	
-x-	
RTN	

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