Compare Linear Regression Lines for the HP-41C

by

Namir Shammas

This article presents an HP-41C program that calculates the linear regression statistics for two data sets and then compares their slopes and intercept. The program runs on an HP41CX without any additional modules. In the case of the HP-41C and HP-41CV you need the X-Functions module to access the REGSWAP command which plays a central role in repeatedly swapping registers that store data related to two sets of observations.

Usage

XEQ CMPLR	Initialize the program.
A	Add a data point for set 1.
A	Delete a data point from set 1. When flag 2 is set, the program deletes the last input.
B	Add a data point for set 2.
B	Delete a data point from set 2. When flag 2 is set, the program deletes the last input.
C	Calculate the slope, intercept, R^2 value, SSE, and standard error for the slope, for data set1.
C	Toggle flag 1. When flag 1 is set, it suppresses the data prompt and processes the data already in the stack. When clear, the program prompts you first to enter the data.
D	Calculate the slope, intercept, R^2 value, SSE, and standard error for the slope, for data set2.
D	Toggle flag 2. When flag 2 is set, the program automatically deletes the last data input for either set. When clear, you must enter a data point to delete.
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 FIX 5 display mode).

Step	Task	Command/Input	Output
1	Initialize the program.	[XEQ] [ALPHA] CMPLR [ALPHA]	READY
2	Toggle flag 01 to set it, in order to bypass the prompts.	C	Flag 01 is on
3	Add the first data point for set 1.	1 [ENTER] 1 [A]	1.00000
4	Enter the second data point for set 1.	8 [ENTER] 2 [A]	2.00000
5	Repeat step 4 to enter the remaining data points in data set 1.	25 [ENTER] 5 [A]	5.00000
6	Add the first data point for set 2.	1.1 [ENTER] 1 [B]	1.00000
7	Enter the second data point for set 2.	4 [ENTER] 2 [B]	2.00000
8	Repeat step 7 to enter the remaining data points in data set 2.	25 [ENTER] 5 [B]	5.00000

Step	Task	Command/Input	Output
9	Calculate the regression coefficients and other statistics for set 1.	[C]	DATA SET 1
10	Calculate the slope.	[R/S]	SLOPE=5.60000
11	Calculate the intercept	[R/S]	INTERC=–5.00000
12	Calculate the coefficient of determination.		R2=0.94800
13	Calculate the SSE value.	[R/S]	SSE=17.20000
14	Calculate the standard error for the slope	[R/S]	SE SLP=0.75719
15	Complete the calculation run.	[R/S]	DONE SET1
16	Calculate the regression coefficients and other statistics for set 2.	[D]	DATA SET 2
17	Calculate the slope.	[R/S]	SLOPE=5.98000
18	Calculate the intercept	[R/S]	INTERC=–6.92000
19	Calculate the coefficient of determination.		R2=0.961280
20	Calculate the SSE value.	[R/S]	SSE=14.40400
21	Calculate the standard error for the slope	[R/S]	SE SLP=0.692916
22	Complete the calculation run.	[R/S]	DONE SET2

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

Step	Task	Command/Input	Output
1	Calculate the student-t for the difference in slopes.	[E]	T SLP=0.37023
2	Calculate the student-t for the difference in intercepts.	[R/S]	T INT=–0.56402

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

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

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

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

Compare the absolute value of t_{slope} with $t_{\alpha/2,n1+n2-4}$.

Comparing Intercepts

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

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

Compare the absolute value of $t_{intercept}$ with $t_{\alpha/2,n1+n2-4}$.

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
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

Degrees of Freedom	α = 0.100	α = 0.050	α = 0.025	α = 0.010
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= Intercept1
R01= Slope1
R02= SE slope1
R03= SSE1
R04= sum of x1
R05= sum of x1 squared
R06= sum of y1
R07= sum of y1 squared
R08= sum of x1 * y1
R09= n1
R10= Intercept2
R11= Slope2
R12= SE slope2
R13= SSE2
R14= sum of x2
R15= sum of x2 squared
R16= sum of y2
R17= sum of y2 squared
R18= sum of x2 * y2
R19= n2
R20= x mean, Sy.x
R21= y mean, Sxx1
R22= Sxx, Sxx2
R23= Syy, std err of the difference between slopes
R24= Sxy, used
R25= used
R26= last X input for data set 1
R27= last Y input for data set 1

R28= last X input for data set 2
R29= last Y input for data set 2

Flags

Flag	Use
00	When set, program processes data set2. When clear, program processes data set1
01	When set, program suppresses data input prompt. When clear, program displays input prompt.
02	When set, labels a and b delete the last input values (which you can do only one). When clear, user must enter values to delete.

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 "CMPLR"	initialize the program
SUMREG 04	clear the summation registers
CLΣ	
XEQ "PS"	
ClΣ	
XEQ "PS"	
CF 00	clear flags 1 to 3
CF 01	
CF 02	
"READY"	
PROMPT	
RTN	
LBL c	toggle flag 01
FC?C 01	
SF 01	
RTN	
LBL d	toggle flag 02
FC?C 02	
SF 02	
RTN	

Program Step	Comment
LBL A	add a data point from the first data set
"Y↗X? SET1"	
FC? 01	
PROMPT	
STO 26	store X
X<>Y	
STO 27	store Y
X<>Y	
Σ+	
RTN	
LBL a	remove a data point from the first data set
"Y↗X? SET1"	
FC? 01	
PROMPT	
FS? 02	push the last X entered?
RCL 27	
FS?C 02	push the last Y entered?
RCL 26	
Σ–	
RTN	
LBL B	add a data point from the second data set
"Y↗X? SET2"	
FC? 01	
PROMPT	
STO 28	store X
X<>Y	
STO 29	store Y
X<>Y	
XEQ "PS"	
Σ+	
XEQ "PS"	
RTN	
LBL b	remove a data point from the second data set
"Y↗X? SET2"	
FC? 01	
PROMPT	
FS? 02	push the last X entered?

Program Step	***Comment***
RCL 28	
FS?C 02	push the last Y entered?
RCL 29	
XEQ "PS"	
Σ–	
XEQ "PS"	
RTN	
LBL 00	helper subroutine to calculate Sxx and Syy
X^2	
RCL 09	
1	
–	
*	
RTN	
LBL 01	helper subroutine to calculate x.mean^2 / Sxx + 1/n for either data set
MEAN	
X^2	
STO 24	
SDEV	
XEQ 00	
1/X	
RCL 24	
*	
RCL 09	
1/X	
+	
RTN	
LBL "PS"	swap registers 0 to 9 and 10 to 19. This subroutine is based on the HP-67 P<>S command which swaps the same range of registers in the HP-67.
0.010010	push swap control value
REGSWAP	swap register (HP-41CX specific or use X-Function module in HP-41C and HP-41CV)
R↓	
RTN	
LBL C	calculate linear regression coefficients for 1st data sets
"DATA SET2"	
FS? 00	

Program Step	***Comment***
"DATA SET1"	
PROMPT	display data set name
MEAN	
STO 20	calculate and store x mean
X<>Y	
STO 21	calculate and store y mean
SDEV	
XEQ 00	
STO 22	calculate and store Sxx
X<>Y	
XEQ 00	
STO 23	calculate and store Syy
MEAN	
*	
RCL 09	
*	
CHS	
RCL 08	
+	
STO 24	calculate and store Sxy
RCL 22	
/	
STO 01	calculate and store slope
"SLOPE="	
ARCL X	
PROMPT	display slope
RCL 20	
*	
CHS	
RCL 21	
+	
STO 00	calculate and store intercept
"INTRC="	
ARCL X	
PROMPT	display intercept
RCL 01	
RCL 24	
RCL 23	
/	
*	
"R2="	

Program Step	***Comment***
ARCL X	
PROMPT	display R-Sqr (value is not stored!)
RCL 23	
RCL 01	
RCL 24	
*	
–	
STO 03	calculate and store SSE
"SSE="	
ARCL X	
PROMPT	display SSE
RCL 09	
2	
–	
/	calculate MSE
RCL 22	
/	
SQRT	
STO 02	calculate and store SE slope
"SE SLP="	
ARCL X	
PROMPT	
FS?C 00	Exit routine if flag 00 is set, because label D called label C.
RTN	
"DONE SET1"	
PROMPT	
RTN	
LBL D	calculate linear regression coefficients for 2nd data sets
SF 00	Flag calculations for data set 2
XEQ "PS"	Swap registers 00 to 09 with registers 10 to 19
XEQ C	Call label C to perform the calculations
XEQ "PS"	Swap registers 00 to 09 with registers 10 to 19
"DONE SET2"	
PROMPT	
RTN	
LBL E	compare slopes and intercepts
RCL 03	get SSE1
RCL 13	get SSE2
+	

Program Step	***Comment***
RCL 09	get n2
RCL 19	get n1
+	
4	
–	
/	
STO 20	calculate and store (Sy.x)^2
SDEV	
XEQ 00	
STO 21	calculate and store Sxx1
XEQ "PS"	
SDEV	
XEQ 00	
STO 22	calculate and store Sxx2
XEQ "PS"	
RCL 21	
1/X	
RCL 22	
1/X	
+	
RCL 20	
*	
SQRT	
STO 23	calculate and store std err of the difference between slopes
1/X	
RCL 01	get slope1
RCL 11	get slope2
–	
*	
"T SLP="	
ARCL X	
PROMPT	display student-t statistic for slope difference
XEQ 01	calculate x1.mean^2 / Sxx1 + 1/n1
STO 25	
XEQ "PS"	
XEQ 01	calculate x2.mean^2 / Sxx2 + 1/n2
XEQ "PS"	
RCL 25	
+	
RCL 20	
*	

Program Step	Comment
SQRT	calculate pooled std dev for intercept
1/X	
RCL 00	get Intercept1
RCL 10	get Intercept2
–	calculate difference in intercepts
*	multiply by 1/s<intercept1 - intercept2>
"T INT="	
PROMPT	calculate and display student-t for intercept difference
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.

Here is the code for an alternate version (and very slow) of subroutine PS when REGSWAP command is not available. This version uses register 30 to 33 to store the stack.

Program Step	Comment
LBL PS	
STO 30	store current stack in regs 30 to 33
R↓	
STO 31	
R↓	
STO 32	
R↓	
STO 33	
0.009	push register swap control variables
10.019	
LBL 09	start loop to swap registers
RCL IND Y	
X<> IND Z	old register Y is now in register Z
STO IND Y	old register X is now in register Y
R↓	drop the stack to restore the two indices back in registers Y and X
ISG Y	
STO X	
ISG X	
GTO 09	end of register swap loop
RCL 33	restore stack before the swap
RCL 32	
RCL 31	
RCL 30	

RTN	