

## Using Your TI-83/84 Calculator: Linear Correlation and Regression

### Elementary Statistics

Dr. Laura Schultz

This handout describes how to use your calculator for various linear correlation and regression applications. For illustration purposes, we will work with a data set consisting of the winning men's Olympic high jump heights (in inches) paired with the years those heights were attained. To simplify the regression equation, I have coded the Olympic year to be zero in 1900. You can find this data set in the Appendix at the end of this handout.

1. Before we can get started, you will need to enter the men's Olympic high jump data into your calculator. We will be using this data set for several different applications, so it will be helpful to enter the data into named lists.

Press **[STAT]****[ENTER]**. Insert a list by highlighting **L<sub>1</sub>** and pressing **[2nd]****[DEL]**. Name this list **OLYR** and proceed to enter the year-code data into this list. Insert another list named **JUMP** and enter the high-jump height data into this list. Check over your lists to make sure you didn't enter any incorrect data values. Note that each **OLYR** value must be paired with the corresponding high **JUMP** height for that year.

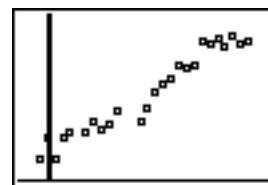
OLYR	JUMP	L1	9
4	71.25	-----	
0	74.8		
4	71		
8	75		
12	76		
20	76.25		
24	78		

OLYR(1) = -4

2. **Generating a Scatterplot.** To get a sense of the data, start by generating a scatterplot. Press **[2nd]****[Y=]** to access the **[STAT PLOT]** menu. Make sure all the plots except **Plot1** are turned off, and then press **[1]**. Select the first **Type** of plot. At the **Xlist** prompt, enter the name of the list containing the predictor variable; this variable will be assigned to the *x*-axis of the plot. For this example, the **Xlist** is **OLYR**. Enter the name of the response variable at the **Ylist** prompt; the values in this list will be plotted on the *y*-axis. The **Ylist** for this example is **JUMP**.



3. Press **[ZOOM]****[9]** to view the scatterplot. There should be no line drawn through the points on your plot; if there is a line, you will need to press **[Y=]** and make sure all the equations are empty for **Plot1**. (Select any equation you need to clear and press **[CLEAR]**.) What can you tell from the scatterplot? Does there appear to be a linear correlation between **OLYR** and **JUMP**? If so, is it positive or negative? How strong does the correlation appear to be?



4. The next step is to find the linear correlation coefficient (*r*) and determine whether there is a significant linear correlation between our two variables. The **LinRegTTest** function on your calculator provides “one-stop shopping” for answering these and other questions relating to linear correlation and regression. Press **[STAT]** and scroll right to the **TESTS** menu. Scroll down to **LinRegTTest** and press **[ENTER]**. (Note: This is menu-item **F** on a **TI-84** calculator, but it is **E** on a **TI-83** calculator.)



5. You will be prompted for the following information:

- **Xlist:** Enter the name of the list containing the predictor ( $x$ ) variable. For this example, type **OLYR** and press **ENTER**.

- **Ylist:** Enter the name of the list containing the response ( $y$ ) variable. For this example, type **JUMP** and press **ENTER**.

- **$\beta$  &  $\rho$ :** Select the sign that appears in your alternative hypothesis. (Consult your lecture notes for more details regarding the hypothesis testing procedure for linear correlation.) We are only concerned with  **$\rho$**  (rho), which is the population linear correlation coefficient. For the purposes of this course, we will always be using  **$\rho \neq 0$**  as our alternative hypothesis when testing whether there is a significant linear correlation between  $x$  and  $y$ . Select  **$\neq 0$**  and press **ENTER**.

- **RegEQ:** Here you can specify one of the built-in  $Y=$  functions as a place to store the regression equation that is generated. Doing so will allow you to add the regression line to your scatterplot later. (Note that this same line will be drawn on all subsequent plots, too, unless you press **Y=** and **CLEAR** to clear the equation from memory when you are finished working with this data set.) I generally store my regression equations as  $Y_1$ . Press **VAR** and scroll right to **Y-VARS** and press **ENTER** to select **1:Function**. Then, press **ENTER** again to select **1:Y<sub>1</sub>**. These keystrokes will enter  $Y_1$  into the **RegEQ** field.

- Highlight **Calculate** and press **ENTER**.

```
LinRegTTest
Xlist:OLYR
Ylist:JUMP
Freq:1
 $\beta$  &  $\rho$ : $\neq 0$  <0 >0
RegEQ:Y1
Calculate
```

```
VAR Y-VARS
1:Function...
2:Parametric...
3:Polar...
4:On/Off...
```

```
FUNCTION
1:Y1
2:Y2
3:Y3
4:Y4
5:Y5
6:Y6
7:Y7
```

6. Your calculator will return two screens full of output; use the **↓** and **↑** keys to scroll through all of the output. Let's start by finding the **linear correlation coefficient ( $r$ )** for our data. You will need to scroll down to the bottom of the second screen to find  $r$ . For our example,  $r = 0.972$  (Always round  $r$  to 3 decimal places.) What does  $r$  tell us? First of all, its sign tells us that there likely is a positive correlation between Olympic year and the winning men's high jump height. The slope of the regression line will also be positive. Second, because  $r$  is very close to 1, we can expect that there is a near-perfect positive correlation.

```
LinRegTTest
y=a+bx
 $\beta \neq 0$  and  $\rho \neq 0$ 
t=19.73392442
P=6.473904E-16
df=23
 $\downarrow$ a=71.90716494
```

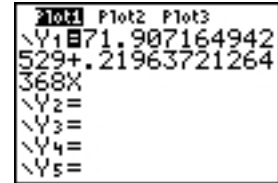
```
LinRegTTest
y=a+bx
 $\beta \neq 0$  and  $\rho \neq 0$ 
tb=.2196372126
s=1.857064887
r2=.9442326596
r=.9717163473
```

7. Is there a statistically significant linear correlation? That is, can we conclude that the population linear correlation coefficient ( **$\rho$** ) is not equal to 0? The **linear regression  $t$  test** addresses this question. Given how large  $r$  was for our data, it is a safe bet that we have a significant linear correlation between **OLYR** and **JUMP**. You would report the results of the  $t$  test for this example as  $t_{23} = 19.7339$ ,  $P = 6.47 \times 10^{-16}$  (two-tailed). Note that I reported the degrees of freedom as a subscript ( $df = n - 2$ ). Round the  $t$ -test statistic to 4 decimal places and the  $P$ -value to 3 significant figures. Given that the  $P$ -value is less than the significance level of  $\alpha = .05$ , we can conclude that there is a significant linear correlation between the

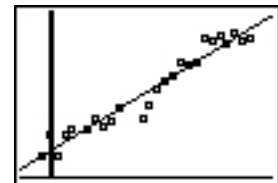
Olympic year and the winning men's high jump height. (I present the formal hypothesis test at the end of this handout.)

8. The **coefficient of determination ( $r^2$ )** tells us how much of the variability in  $y$  can be explained by the linear correlation between  $x$  and  $y$ . By convention,  $r^2$  is reported as a percentage. For our example,  $r^2 = 94.4\%$ . Round  $r^2$  to 3 decimal places. *What does this mean?* 94.4% of the variation in men's winning high jump heights can be attributed to the linear relationship between the Olympic year and the winning high jump height.

9. Given that we had a significant linear correlation, it is appropriate to construct the **linear regression equation** for our data. There are two methods for doing so. First, note that the previous calculator displays indicate that  $y = a + bx$ . Your calculator reports values for both  $a$  (the  $y$ -intercept) and  $b$  (the slope). The second method is to press  $\boxed{Y=}$  and record the equation given for  $Y_1$ . In either case, round the  $y$ -intercept and slope values to 3 significant figures each when you report the linear regression equation. The linear regression equation for our sample data is  $\hat{y} = 71.9 + 0.220x$ . In situations where there is not a significant linear correlation, do not bother constructing a linear regression equation. (Without a significant linear correlation coefficient, we cannot make predictions from a regression equation. The best estimate of  $\hat{y}$  for any corresponding  $x$  value is simply  $\bar{y}$  if there is no linear relationship between  $x$  and  $y$ .)

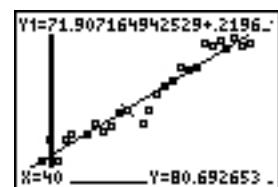
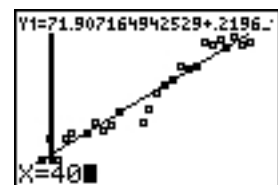


10. Let's plot the data again and see what it looks like with the **regression line**. Press  $\boxed{\text{ZOOM}}\boxed{9}$ . Your calculator will return the scatterplot with the regression line in place. Note how well the regression line fits our data. The stronger the linear correlation, the closer the data points will cluster along the regression line.



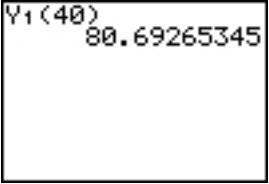
11. What is the marginal change in men's high jump performance per year? **Marginal change** is simply the slope of the regression line. Hence, the marginal change for our example is 0.220 inches/year. In other words, men's high jump performance improves by 0.220 inches per year.

12. **Making predictions from a regression equation.** Let's use the regression equation to predict what the winning high jump height would have been in 1940 had the Olympics been held that year. Once again, there are several approaches you can use. Let's start by working off the scatterplot display. Press  $\boxed{\text{TRACE}}$  and then  $\boxed{\downarrow}$  to hop from the data points onto the regression line. Type in the  $x$  value you want to plug into the regression equation and press  $\boxed{\text{ENTER}}$ . For this example, recall that the Olympic year is coded as the number of years since 1900. Therefore, type  $40$  and press  $\boxed{\text{ENTER}}$ . Your calculator will return the display shown to the right. The predicted  $y$  value for an  $x$  value of 40 is reported on the lower right-hand corner of the display. Round your predictions to 3 significant figures. Hence, we predict that the



winning men's high jump would have been 80.7 inches had the Olympics been held in 1940.

13. Another approach is to work from the home screen. Press  $\boxed{2\text{nd}}\boxed{\text{MODE}}$  to exit the plot display. Then, press  $\boxed{\text{VAR}}\boxed{\rightarrow}\boxed{1}\boxed{1}$  to paste  $Y_1$  to the home screen and type  $\boxed{(}\boxed{4}\boxed{0}\boxed{)}\boxed{\text{ENTER}}$ . Doing so will tell your calculator to plug  $x = 40$  into the regression equation stored for  $Y_1$ . Your calculator will return the output screen shown to the right. Once again, you would round to 3 significant figures and report that you predict that the winning Olympic men's high jump height in 1940 would have been 80.7 inches.



A calculator screen showing the result of a regression calculation. The display shows "Y1(40)" on the top line and "80.69265345" on the bottom line.

14. The final approach is to write down the equation given for  $Y_1$  and plug 40 into the equation manually. If you choose to the third approach, make sure you use all the given digits and round only at the end. I strongly advise that you adopt one of the first two approaches for making predictions.
15. When you are finished working with this data set, don't forget to press  $\boxed{\text{Y=}}$  and  $\boxed{\text{CLEAR}}$  to clear the regression equation for  $Y_1$  from memory. Forgetting to do so will cause you problems the next time you try to generate a stat plot.

Here is the formal hypothesis test for this example:

**Claim:** There is a linear correlation between the Olympic year and the winning men's high jump height.

Let  $\rho$  = the population linear correlation coefficient

$H_0: \rho = 0$  (There is no linear correlation between the Olympic year and the winning men's high jump height.)

$H_1: \rho \neq 0$  (There is a linear correlation between the Olympic year and the winning men's high jump height.)

Conduct a two-tailed linear regression t test with a significance level of  $\alpha = .05$

Reject  $H_0$  because our P-value is less than  $\alpha$  ( $6.47 \times 10^{-16} \lll \lll .05$ )

**Conclusion:** There is a significant linear correlation between the Olympic year and the winning men's high jump height ( $t_{23} = 19.7339$ ,  $P = 6.47 \times 10^{-16}$ , two-tailed).

## Appendix: Data Sets

oly_gold		
	Year	High_Jump
1	-4	71.2500
2	0	74.8000
3	4	71.0000
4	8	75.0000
5	12	76.0000
6	20	76.2500
7	24	78.0000
8	28	76.3750
9	32	77.6250
10	36	79.9375
11	48	78.0000
12	52	80.3200
13	56	83.2500
14	60	85.0000
15	64	85.7500
16	68	88.2500
17	72	87.7500
18	76	88.5000
19	80	92.7500
20	84	92.5000
21	88	93.5000
22	92	92.0000
23	96	94.0000
24	100	92.5000
25	104	92.9000

The Olympic year has been coded as the number of years since 1900. Men's winning high-jump heights are given in inches.

Spring 2007 Class Data		
	Shoe_size	Height
1	12.0	72
2	7.0	64
3	4.5	62
4	11.0	70
5	8.5	69
6	5.0	65
7	12.0	72
8	7.5	65
9	8.5	65
10	5.5	65
11	9.5	68
12	5.5	61
13	10.5	69
14	12.0	77
15	14.0	73
16	7.0	65
17	7.0	67
18	12.0	71
19	12.0	73
20	7.0	64
21	13.0	71
22	11.0	71
23	12.0	72
24	4.5	61
25	10.5	71
26	10.0	66
27	10.0	67
28	13.0	73
29	7.5	69
30	4.5	61
31	8.5	70
32	14.0	75
33	10.0	72
34	6.5	66

Heights are given in inches. Women's shoe sizes have been converted to men's sizes by subtracting 1.5 from the reported shoe sizes.