

# Chapter 13

## Simple Linear Regression and Correlation: Inferential Methods

### 13.0 Introduction

The TI-83 does a superb job of supporting inference in the context of simple linear regression. From easy entry of data, to checking assumptions, to performing the actual inference, simplicity is the order of the day. We have already discussed most of the skills of plotting the graphs, and you are encouraged to consult that material at the end of Chapter 3 and throughout Chapter 5.

### 13.1 Estimating the Population Regression Line

#### Example 13.2: Mother's Age and Babies Birth Weight

Medical researchers have noted that adolescent females are much more likely to deliver low birth weight babies than are adult females. Because low birth weight babies have higher mortality rates there have been a number of studies examining the relationship between birth weight and mother's age for babies born to young mothers. One such study is described in the article "The Risk of Teen Mothers Having Low Birth Weight Babies: Implications of Recent Medical Research for School Health Personnel" (*J. of School Health* (1998): 271–274). The accompanying data on  $x$  = maternal age (years) and  $y$  = birth weight of baby (grams) is consistent with summary values given in the referenced article and also with data published by the National Center for Health Statistics.

	Observation									
	1	2	3	4	5	6	7	8	9	10
$x$	15	17	18	15	16	19	17	16	18	19
$y$	2289	3393	3271	2648	2897	3327	2970	2535	3138	3573

Enter these data in your calculator, and estimate the population regression line using the sequence below. Remember from Section 5.2 that these keystrokes will result in a calculated least squares regression line, and paste that line into the calculator function,  $Y_1$ , for graphing purposes. Also recall that to get the  $Y_1$  you must execute sequence (a) within sequence (b):

(a) **VAR**S > **Y-VARS** > **1:Function** > **1:Y<sub>1</sub>** > **Enter...**

(b) **Stat** > **Calc** > **8:LinReg(a+bx)** > **L<sub>1</sub>, L<sub>2</sub>, Y<sub>1</sub>** > **Enter.**

After performing these, your screen should look like this:

```
LinReg
y=a+bx
a=-1163.45000
b=245.15000
r2=.78091
r=.88369
```

What we would like to add at this point is the capability of acquiring a point estimate of the average  $y$  value (birth weight) for a particular  $x$  value (age of mother). On paper, we substitute the  $x$  value into the regression equation:

$$\begin{aligned}\hat{y} &= -1163.45 + 245.15x \\ \hat{y}(18) &= -1163.45 + 245.15(18) \\ &= 3249.25\end{aligned}$$

On the TI-83, we can duplicate this effort by evaluating the function  $Y_1$  when  $x = 18$ . Here are the keystrokes:

**VAR**S > **Y-VARS** > **1:Function** > **1:Y<sub>1</sub>** > **(18)** > **Enter...**

```
LinReg
y=a+bx
a=-1163.45000
b=245.15000
r2=.78091
r=.88369
Y1(18)
      3249.25000
```

## 13.2 Inference for the slope of a Population Regression Line

### Example 13.4: Athletic Performance and Cardiovascular Fitness

Is cardiovascular fitness (as measured by time to exhaustion running on a treadmill) related to an athlete's performance in a 20-km ski race? The accompanying data on  $x$  = treadmill time to exhaustion (min.) and  $y$  = 20-km ski time (min) was taken from the article "Physiological Characteristics and Performance of Top U.S. Biathletes" (*Medicine and Science in Sports and Exercise* (1995): 1302–1310):

$x$	7.7	8.4	8.7	9.0	9.6	9.6	10.0	10.2	10.4	11.0	11.7
$y$	71.0	71.4	65.0	68.7	64.4	69.4	63.0	64.6	66.9	62.6	61.7

We will test the hypothesis,  $H_0: \beta = 0$ , and construct a 95% confidence interval for  $\beta$  based on the sample data we have. The TI-83 will not calculate the confidence interval automatically, but we can use the information from the hypothesis test to construct the confidence interval.

### The hypothesis test

With the treadmill time in List1 and the ski time in List2, execute these keystrokes:

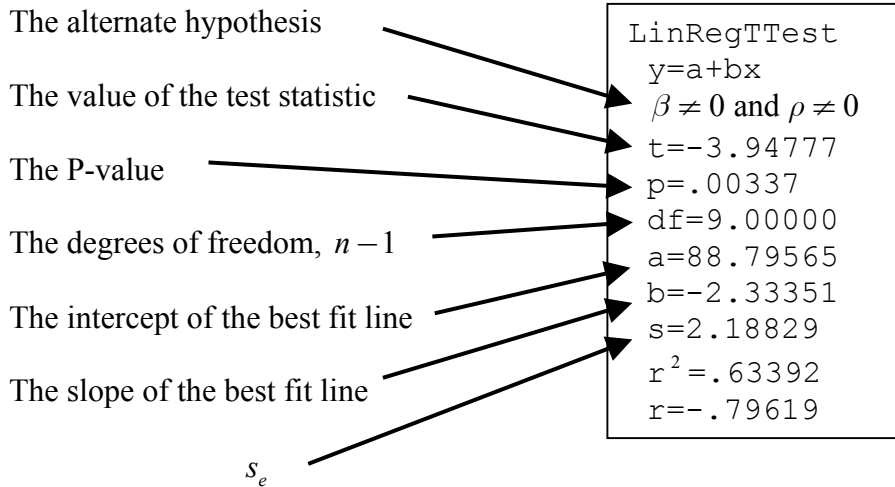
**STAT > TESTS > LinRegTTest > ENTER**

The resulting screen of options for this hypothesis test is presented below, complete with the elements you should choose for this example:

```
LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
β & ρ:≠0 <0 >0
RegEQ:
Calculate
```

These options are for the most part nothing new. If you calculated the best fit line and put the formula in  $Y_1$  you could paste that in **RegEQ**, but it is just as easy to let the calculator work from the actual data in List1 and List2. The key choice in this screen is the selection of the alternate hypothesis, and this is almost invariably a two-tailed test in elementary statistics. To get the usual (two-sided) confidence interval, you must select

the two-tailed alternate hypothesis. Arrow down to **Calculate** and press **ENTER**. Here are the results of these selections, annotated:



Compare these results with the computer output in POD and once again we have duplicated the correct answers.

### The confidence interval

From the output for the hypothesis test, a table look-up, and a little algebra we can construct the 95% confidence interval for the slope of the population regression line,  $\beta$ . In order to successfully do this, we need to find the value of  $s_b$ . The bad news from our perspective is that the TI-83 does not directly calculate  $s_b$ . The good news is that it does calculate enough so we can figure out the rest. To justify the procedure we will present below, here's a little algebra starting with a formula from POD.

$$\begin{aligned}
 s_b &= \frac{s_e}{\sqrt{S_{xx}}} \\
 &= \frac{s_e}{\sqrt{\sum (x - \bar{x})^2}} \\
 &= \frac{s_e}{\sqrt{(n-1)} \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}} \\
 &= \frac{s_e}{S_x \sqrt{n-1}}
 \end{aligned}$$

This would appear to be algebra that only a math teacher could love, and we are only interested in the result. The calculator does in fact evaluate all the quantities on the right-hand side of this formula, and we can take advantage of that fact. It is technically possible to perform all these calculations with a single sequence of keystrokes, but there are an awful lot of those **VARS** sequences embedded in the calculation. Unless you are writing a TI-83 program, it is easier just to do the calculations by calculating or looking up each value separately and then calculating the final result. The 95% confidence interval for  $\beta$  is given by this formula:

$$b \pm (t \text{ critical value}) \cdot s_b$$

Substituting the critical  $t$  value from the table, calculating  $s_b$  from  $s_e$ ,  $S_x$ , and  $n$ , all of which are provided during the calculations for the hypothesis test, we find the 95% confidence interval:

$$\begin{aligned} & b \pm (t \text{ critical value}) \cdot \frac{s_e}{S_x \sqrt{n-1}} \\ & -2.33351 \pm (2.26) \cdot \frac{2.18829}{1.1707 \sqrt{11-1}} \\ & (-3.66939, \quad -.99763) \end{aligned}$$

### 13.3 Afterword

It turned out to be a bit more difficult to make a confidence interval for the slope of the regression line, but nevertheless – it still looks a lot better than calculating it by hand.