

# Chapter 11

## Simple Linear Regression

### Fitting the Model: The Least Squares Approach

#### Exercise 11.28 Sweetness of Orange Juice

**Problem:** The quality of the orange juice produced by a manufacturer is constantly monitored. Is there a relationship between the sweetness index and a chemical measure such as the amount of water-soluble pectin (parts per million) in the orange juice? Data collected on these two variables during 24 production runs at a juice-manufacturing plant are shown in Table 11 – 1. Suppose a manufacturer wants to use simple linear regression to predict the sweetness ( $y$ ) from the amount of pectin ( $x$ ). Find the least squares regression line for the data. Predict the sweetness index if the amount of pectin in the orange juice is 300 ppm.

Table 11 – 1  
Sweetness of  
Orange Juice

Sweetness Index	Pectin (ppm)
5.2	220
5.5	227
6.0	259
5.9	210
5.8	224
6.0	215
5.8	231
5.6	268
5.6	239
5.9	212
5.4	410
5.6	256
5.8	306
5.5	259
5.3	284
5.3	383
5.7	271
5.5	264
5.7	227
5.3	263
5.9	232
5.8	220
5.8	246
5.9	241

11-2 Chapter 11

**Solution:** We begin by entering the data into the calculator. In this example, the pectin level is the independent variable so it should be entered into list 1 and the sweetness index is the dependent variable so it should be entered into list 2, since the sweetness index should be based upon the pectin level and not vice-versa. In general, the x's go into List 1 and the y's into List 2.

After you have entered all the data your screen will look like Figure 11 – 1.

L1	L2	L3	Z
227	5		
269	7		
282	9		
232	10		
220	11		
246	12		
241	13		
-----			
L2(25)=			

Figure 11 – 1

1. Press **[STAT]** and arrow over to the CALC menu.

Notice that the TI-83/84 Plus has two different linear regressions **4:LinReg(ax + b)** and **8:LinReg(a + bx)**. Number 8 most closely matches the statistics form so we will use it in this example. However, both are acceptable.

2. Press **[8]** or arrow down to **8:LinReg(a + bx)**. See Figure 11 – 2. Press **[ENTER]**. **LinReg (a+bx)** will appear at the top of your home screen. See Figure 11 – 3.

EDIT	TESTS
2:2-Var Stats	
3:Med-Med	
4:LinReg(ax+b)	
5:QuadReg	
6:CubicReg	
7:QuartReg	
8:LinReg(a+bx)	

Figure 11 – 2

LinReg(a+bx)
--------------

Figure 11 – 3

3. The format for this command is **LinReg(a+bx) xlist, ylist, Y-vars** where the regression equation should be stored. Here we are using Lists 1 and 2 so press **[2nd] L1** **[,]** **[2nd] L2** for the lists. To get the Y where the equation will be stored, press **[,]** **[VAR]** arrow to Y-vars, press **[ENTER]** and then **[ENTER]** again to choose **Y<sub>1</sub>**. See Figure 11 – 4.

LinReg(a+bx) L1,
L2, Y1

Figure 11 – 4

4. Press **[ENTER]**. Your screen will appear as in Figure 11 – 5.

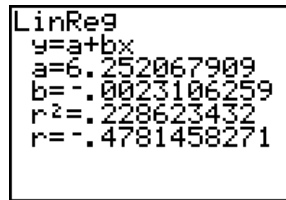


Figure 11 – 5

As can be seen from the screen in Figure 11 – 5 the slope is -0.0023 and the y-intercept is 6.2521 when rounded to 4 decimal places. Thus, the Least Squares Line is  $\hat{y} = 6.2521 - 0.0023x$

5. Press **[ZOOM]** **[9]** and the graph will appear as in Figure 11 – 6.

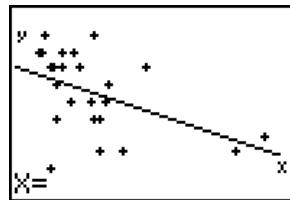


Figure 11 – 6

Once you have the graph displayed you can use the **CALC** feature of the calculator to predict specific y-values. See Figure 11- 7.

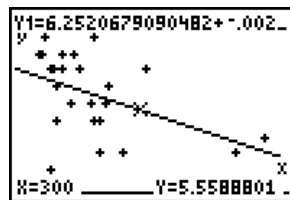


Figure 11 – 7

As can be seen from the screen in Figure 11 – 7, if  $x = 300$  we predict  $y = 5.559$  when rounded to four decimal places.

## The Coefficients of Correlation and Determination

### Example 11.4 Relating Crime Rate & Casino Employment

**Problem:** Legalized gambling is available on several riverboat casinos operated by a city in Mississippi. The mayor of the city wants to know the correlation between the number of casino employees and the yearly crime rate. The records for the past 10 years are examined, and the results are listed in Table 11 – 2. Determine the coefficient of correlation,  $r$ .

Table 11 – 2  
Crime Rates and Casino  
Employment

Year	Number x of Casino Employees (in thousands)	Crime Rate (Number of Crimes per 1,000 population)
2001	15	1.35
2002	18	1.63
2003	24	2.33
2004	22	2.41
2005	25	2.63
2006	29	2.93
2007	30	3.41
2008	32	3.26
2009	35	3.63
2010	38	4.15

**Solution:** Begin by entering the data into the calculator.

Remember to enter the data in the exact order it is given. Since the Crime Rate is the dependent variable, it will be placed into List 2 and since the Number of Casino Employees is the independent variable, it will be placed into List 1. See Figure 11 – 8.

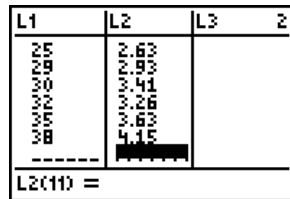


Figure 11 – 8

1. Press **[STAT]** and arrow over to the CALC menu.

Notice that the TI-83/84 Plus has two different linear regressions **4:LinReg(ax + b)** and **8:LinReg(a + bx)**. Number 8 most closely matches the statistics form so we will use it in this example. However, both are acceptable.

2. Press **[8]** or arrow down to **8:LinReg(a + bx)**. See Figure 11 – 9. Press **[ENTER]**. **LinReg (a+bx)** will appear at the top of your home screen. See Figure 11 – 10.

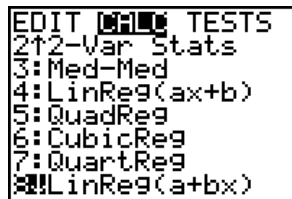


Figure 11 – 9

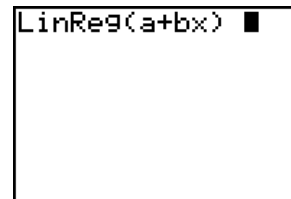


Figure 11 – 10

3. The format for this command is  $\text{LinReg}(a+bx)$  xlist, ylist, Y-vars where the regression equation should be stored. Here we are using Lists 1 and 2 so press  $\text{2nd}$  L1  $\text{,}$   $\text{2nd}$  L2 for the lists. To get the Y where the equation will be stored, press  $\text{,}$   $\text{VARS}$  arrow to Y-vars, press  $\text{ENTER}$  and then  $\text{ENTER}$  again to choose  $Y_1$ . See Figure 11 – 11.

```
LinReg(a+bx) L1,
L2, Y1
```

Figure 11 – 11

4. Press  $\text{ENTER}$ . Your screen will appear as in Figure 11 – 12.

```
LinReg
y=a+bx
a=-.3979654036
b=.1183196046
r2=.9742277817
r=.9870297775
```

Figure 11 – 12

If you do not see the  $r$  and  $r^2$  you will need to follow these steps:

Press:  $\text{2nd}$  [CATALOG] to enter the catalog

Next **Arrow Down** to “**DiagnosticsOn**”

Press:  $\text{ENTER}$   $\text{ENTER}$

You only have to turn on the diagnostics once.

As can be seen from the screen in Figure 11 – 12,  $r = 0.9870$  and  $r^2 = 0.9742$  when rounded to four decimal places. The linear correlation coefficient is  $r = 0.9870$  which indicates a very strong positive linear relationship between number of casino employees and crime rate. The coefficient of determination is  $r^2 = 0.9742$  which indicates that 97.42% of the variability in crime rates can be explained by the number of casino employees.

**Exercise 11.121 Feeding Habits of Fish**

**Problem:** Zoologists recorded the number of aggressive strikes of two black-bream fish feeding at the bottom of an aquarium in the 10-minute period following the addition of food. Table 11 – 3 lists the weekly number of strikes and the age of the fish (in days). Fit the equation of a straight line model relating the number of strikes ( $y$ ) to age of fish ( $x$ ). Then test the null hypothesis that the population slope is zero against the alternative that the population slope is less than 0.

Table 11 – 3  
Feeding Habits of  
Fish

Week	Number of Strikes	Age of Fish (days)
1	85	120
2	63	136
3	34	150
4	39	155
5	58	162
6	35	169
7	57	178
8	12	184
9	15	190

**Solution:** Begin this problem by entering the data into the calculator. Put the  $x$ 's into list 1 and the  $y$ 's into list 2. The problem states that  $y$  is the number of strikes and that  $x$  is the age of the fish. After entering the data your screen should appear as in Figure 11 – 13.

L1	L2	L3	Z
120	85		
136	63		
150	34		
155	39		
162	58		
169	35		
178	57		
L2(1)=85			

Figure 11 – 13

1. Press **[STAT]** and arrow over to the CALC menu.

Notice that the TI-83/84 Plus has two different linear regressions **4:LinReg(ax + b)** and **8:LinReg(a + bx)**. Number 8 most closely matches the statistics form so we will use it in this example. However, both are acceptable.

2. Press **[8]** or arrow down to **8:LinReg(a + bx)**. See Figure 11 – 14. Press **[ENTER]**. **LinReg (a+bx)** will appear at the top of your home screen. See Figure 11 – 15.

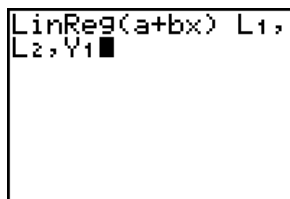
EDIT	TESTS
2:2-Var Stats	
3:Med-Med	
4:LinReg(ax+b)	
5:QuadReg	
6:CubicReg	
7:QuartReg	
8:LinReg(a+bx)	

Figure 11 – 14

LinReg(a+bx)

Figure 11 – 15

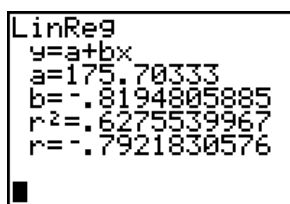
3. The format for this command is LinReg(a+bx) xlist, ylist, Y-vars where the regression equation should be stored. Here we are using Lists 1 and 2 so press  $\boxed{2\text{nd}}$  L1  $\boxed{,}$   $\boxed{2\text{nd}}$  L2 for the lists. To get the Y where the equation will be stored, press  $\boxed{,}$   $\boxed{\text{VARs}}$  arrow to Y-vars, press  $\boxed{\text{ENTER}}$  and then  $\boxed{\text{ENTER}}$  again to choose Y<sub>1</sub>. See Figure 11 – 16.



```
LinReg(a+bx) L1,
L2, Y1
```

Figure 11 – 16

4. Press  $\boxed{\text{ENTER}}$ . Your screen will appear as in Figure 11 – 17.



```
LinReg
y=a+bx
a=175.70333
b=-.8194805885
r^2=.6275539967
r=-.7921830576
```

Figure 11 – 17

If you do not see the  $r$  and  $r^2$  you will need to follow these steps:

Press:  $\boxed{2\text{nd}}$  [CATALOG] to enter the catalog

Next **Arrow Down** to “DiagnosticsOn”

Press:  $\boxed{\text{ENTER}}$   $\boxed{\text{ENTER}}$  You only have to turn the diagnostics on once.

As can be seen from the screen in Figure 11 – 17, the slope of the regression line is -0.8195 and the y-intercept is 175.7033 when rounded to 4 decimal places. Therefore the linear regression equation is  $\hat{y} = 175.7033 - 0.8195x$ . From Figure 11 – 17 we also see  $r = 0.7922$  and  $r^2 = 0.6276$ .

To test the null hypothesis that the population slope is zero against the alternative that the population slope is less than 0, the null and alternative hypotheses are:

$$H_0 : \beta_1 = 0 \text{ (there is no linear relationship)}$$

$$H_a : \beta_1 < 0 \text{ (there is a negative linear relationship)}$$

Where  $\beta_1$  denotes the slope of the population regression line that relates number of strikes to age.

1. Press **[STAT]**, arrow over to TESTS, and press **[ALPHA]** E or arrow down to **E:LinRegTTest** and press **[ENTER]**. See Figure 11 – 18

```

EDIT CALC TESTS
0:1-2-SampTInt...
A:1-PropZInt...
B:2-PropZInt...
C:χ²-Test...
D:2-SampFTest...
E:LinRegTTest...
F:ANOVA(
    
```

Figure 11 – 18

2. Enter Xlist as List 1, Ylist as List 2, and Freq as 1. See Figure 11 – 19.
3. Highlight the correct alternative hypothesis and press **[ENTER]**. Here we need to highlight  $< 0$ . See Figure 11 – 19.

```

LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
μ & σ: ≠ 0 < 0 > 0
RegEQ:Y1
Calculate
    
```

Figure 11 – 19

4. The RegEQ: line allows you to store the regression equation to a particular Y variable. We will store it to  $Y_1$  by pressing **[VARS]**, arrowing to Y-VARS, pressing **[ENTER]** and then **[ENTER]** again. Your screen should appear as in Figure 11 – 19.
5. Highlight Calculate and press **[ENTER]**. Your screen will appear as in Figure 11 – 20. If you scroll down you will also see Figure 11-21.

```

LinRegTTest
y=a+bx
μ < 0 and σ < 0
t=-3.434337016
p=.0054603746
df=7
↓a=175.70333
    
```

Figure 11 – 20

```

LinRegTTest
y=a+bx
μ < 0 and σ < 0
tb=-.8194805885
s=15.4348755
r²=.6275539967
r=-.7921830576
    
```

Figure 11 – 21

From Figure 11 – 20, we see the t-value is given as  $t = -3.4343$  when rounded to four decimal places. The  $p$ -value is approximately 0.0055. Since the  $p$ -value is less than the significance level, we reject  $H_0$ . At the 10% significance level, the data provide sufficient evidence to conclude that the slope of the population regression line is less than 0 and hence that number of strikes is useful as a predictor of age for black-bream fish.