Linear Regression and Correlation: Regression Lab I*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

Abstract

This module provides a lab of Linear Regression and Correlation as a part of Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

Class Time:
Names:

1 Student Learning Outcomes:

- The student will calculate and construct the line of best fit between two variables.
- The student will evaluate the relationship between two variables to determine if that relationship is significant.

2 Collect the Data

Use 8 members of your class for the sample. Collect bivariate data (distance an individual lives from school, the cost of supplies for the current term).

1. Complete the table.

Distance from school	Cost of supplies this term

*Version 1.11: Jun 26, 2012 6:49 am -0500
†http://creativecommons.org/licenses/by/3.0/

Table 1

2. Which variable should be the dependent variable and which should be the independent variable? Why?
3. Graph "distance" vs. "cost." Plot the points on the graph. Label both axes with words. Scale both axes.

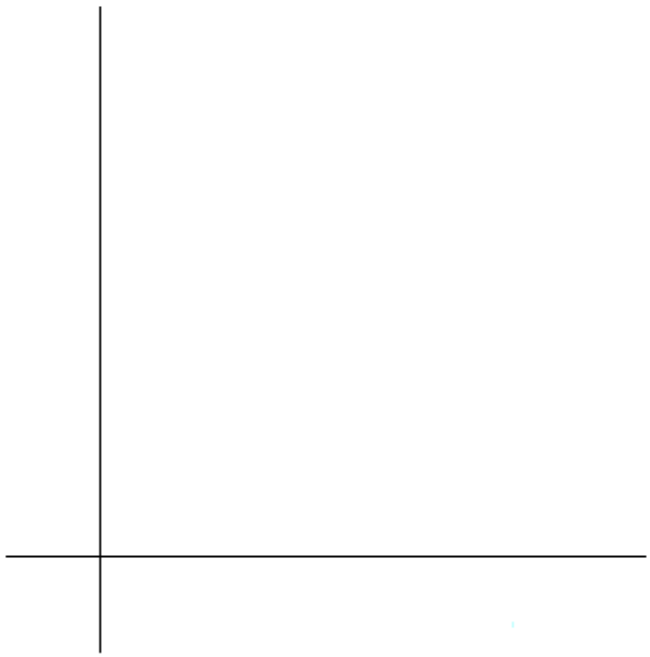

Figure 1

3 Analyze the Data

Enter your data into your calculator or computer. Write the linear equation below, rounding to 4 decimal places.

1. Calculate the following:

   a. $a =$
   b. $b=$
   c. correlation $=$
   d. $n =$
   e. equation: $\hat{y} =$
   f. Is the correlation significant? Why or why not? (Answer in 1-3 complete sentences.)

2. Supply an answer for the following senarios:
   a. For a person who lives 8 miles from campus, predict the total cost of supplies this term:
   b. For a person who lives 80 miles from campus, predict the total cost of supplies this term:
3. Obtain the graph on your calculator or computer. Sketch the regression line below.

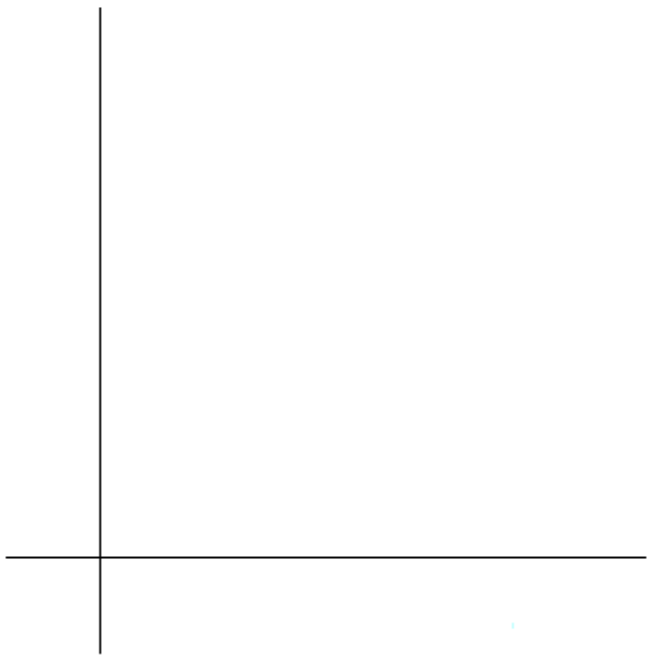

Figure 2

4 Discussion Questions

1. Answer each with 1-3 complete sentences.
   a. Does the line seem to fit the data? Why?
   b. What does the correlation imply about the relationship between the distance and the cost?
2. Are there any outliers? If so, which point is an outlier?
3. Should the outlier, if it exists, be removed? Why or why not?