

LINEAR REGRESSION AND CORRELATION: PRACTICE*

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

1 Student Learning Outcomes

- The student will explore the properties of linear regression.

2 Given

The data below are real. Keep in mind that these are only reported figures. (*Source: Centers for Disease Control and Prevention, National Center for HIV, STD, and TB Prevention, October 24, 2003*)

*Version 1.8: Feb 19, 2009 2:29 pm US/Central

†<http://creativecommons.org/licenses/by/2.0/>



Adults and Adolescents only, United States

Year	# AIDS cases diagnosed	# AIDS deaths
Pre-1981	91	29
1981	319	121
1982	1,170	453
1983	3,076	1,482
1984	6,240	3,466
1985	11,776	6,878
1986	19,032	11,987
1987	28,564	16,162
1988	35,447	20,868
1989	42,674	27,591
1990	48,634	31,335
1991	59,660	36,560
1992	78,530	41,055
1993	78,834	44,730
1994	71,874	49,095
1995	68,505	49,456
1996	59,347	38,510
1997	47,149	20,736
1998	38,393	19,005
1999	25,174	18,454
2000	25,522	17,347
2001	25,643	17,402
2002	26,464	16,371
Total	802,118	489,093

Table 1

NOTE: We will use the columns “year” and “# AIDS cases diagnosed” for all questions unless otherwise stated.

3 Graphing

Graph “year” vs. “# AIDS cases diagnosed.” **Plot the points on the graph located below in the section titled "Plot"** . Do not include pre-1981. Label both axes with words. Scale both axes.

Source URL: <http://cnx.org/content/col10522/latest/>

Saylor URL: <http://www.saylor.org/courses/ma121/>

<http://cnx.org/content/m17088/1.8/>

Attributed to: Barbara Illowsky and Susan Dean



4 Data

Exercise 1

Enter your data into your calculator or computer. The pre-1981 data should not be included. Why is that so?

5 Linear Equation

Write the linear equation below, rounding to 4 decimal places:

Exercise 2

(Solution on p. 5.)

Calculate the following:

- a. $a =$
- b. $b =$
- c. corr. =
- d. $n =$ (# of pairs)

Exercise 3

(Solution on p. 5.)

equation: $\hat{y} =$

6 Solve

Exercise 4

(Solution on p. 5.)

Solve.

- a. When $x = 1985$, $\hat{y} =$
- b. When $x = 1990$, $\hat{y} =$

7 Plot

Plot the 2 above points on the graph below. Then, connect the 2 points to form the regression line.



Obtain the graph on your calculator or computer.



8 Discussion Questions

Look at the graph above.

Exercise 5

Does the line seem to fit the data? Why or why not?

Exercise 6

Do you think a linear fit is best? Why or why not?

Exercise 7

Hand draw a smooth curve on the graph above that shows the flow of the data.

Exercise 8

What does the correlation imply about the relationship between time (years) and the number of diagnosed AIDS cases reported in the U.S.?

Exercise 9

Why is “year” the independent variable and “# AIDS cases diagnosed.” the dependent variable (instead of the reverse)?

Exercise 10

(Solution on p. 5.)

Solve.

- When $x = 1970$, $\hat{y} =$
- Why doesn't this answer make sense?

Solutions to Exercises in this Module

Solution to Exercise 2 (p. 3)

- a. $a = -3,448,225$
- b. $b = 1750$
- c. $\text{corr.} = 0.4526$
- d. $n = 22$

Solution to Exercise 3 (p. 3)

$$\hat{y} = -3,448,225 + 1750x$$

Solution to Exercise 4 (p. 3)

- a. 25082
- b. 33,831

Solution to Exercise 10 (p. 4)

- a. -1164

