

LINEAR REGRESSION AND CORRELATION: PREDICTION*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0[†]

Abstract

Linear Regression and Correlation: Prediction is a part of Collaborative Statistics collection (coll0522)
by Barbara Illowsky and Susan Dean with contributions from Roberta Bloom.

Recall the third exam/final exam example.

We examined the scatterplot and showed that the correlation coefficient is significant. We found the equation of the best fit line for the final exam grade as a function of the grade on the third exam. We can now use the least squares regression line for prediction.

Suppose you want to estimate, or predict, the final exam score of statistics students who received 73 on the third exam. The exam scores (*x-values*) range from 65 to 75. **Since 73 is between the *x-values* 65 and 75**, substitute $x = 73$ into the equation. Then:

$$\hat{y} = -173.51 + 4.83(73) = 179.08 \quad (1)$$

We predict that statistic students who earn a grade of 73 on the third exam will earn a grade of 179.08 on the final exam, on average.

Example 1

Recall the third exam/final exam example.

Problem 1

What would you predict the final exam score to be for a student who scored a 66 on the third exam?

Solution

145.27

Problem 2

(Solution on p. 2.)

What would you predict the final exam score to be for a student who scored a 90 on the third exam?

**With contributions from Roberta Bloom

*Version 1.8: Jun 27, 2012 6:44 am -0500

[†]<http://creativecommons.org/licenses/by/3.0/>

Solutions to Exercises in this Module

Solution to Example 1, Problem 2 (p. 1)

The x values in the data are between 65 and 75. 90 is outside of the domain of the observed x values in the data (independent variable), so you cannot reliably predict the final exam score for this student. (Even though it is possible to enter x into the equation and calculate a y value, you should not do so!)

To really understand how unreliable the prediction can be outside of the observed x values in the data, make the substitution $x = 90$ into the equation.

$$\hat{y} = -173.51 + 4.83(90) = 261.19$$

The final exam score is predicted to be 261.19. The largest the final exam score can be is 200.

NOTE: The process of predicting inside of the observed x values in the data is called **interpolation**. The process of predicting outside of the observed x values in the data is called **extrapolation**.