8.6: Prediction

Recall the third exam/final exam example. We examined the scatter plot 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 mean 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:

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

We predict that statistics 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 8.6.1

Recall the third exam/final exam example.

a. What would you predict the final exam score to be for a student who scored a 66 on the third exam?
b. What would you predict the final exam score to be for a student who scored a 90 on the third exam?

Answer

a. 145.27

b. The x values in the data are between 65 and 75. Ninety 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 90 into the equation for x and calculate a corresponding y value, the y value that you get will not be reliable.)

To understand really how unreliable the prediction can be outside of the observed x-values observed 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.

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

Exercise 8.6.1

Data are collected on the relationship between the number of hours per week practicing a musical instrument and scores on a math test. The line of best fit is as follows:

$$\hat{y} = 72.5 + 2.8x$$

What would you predict the score on a math test would be for a student who practices a musical instrument for five hours a week?

Answer

86.5

8.6.1 - Summary

After determining the presence of a strong correlation coefficient and calculating the line of best fit, you can use the least squares regression line to make predictions about your data.

8.6.2 - References

1. Data from the Centers for Disease Control and Prevention.
2. Data from the National Center for HIV, STD, and TB Prevention.

3. Data from the United States Census Bureau. Available online at www.census.gov/compendia/stat...atalities.html
4. Data from the National Center for Health Statistics.