

Practice Assignment: Scatterplot

To complete this practice assignment, go to the *Relationship Between Two Quantitative Variables* tool at https://lumen-learning.shinyapps.io/association_quantitative/.

- 1) Using the broadband and Gross Domestic Product (GDP) dataset, create a scatterplot that describes the relationship between GDP in billions of U.S. dollars (USD) and the number of broadband subscribers. Answer the following:

Part A: What is the mean GDP in billions of USD?

Answer: 1,909 USD

Part B: Does the dataset contain any outliers? If so, which countries?

Answer: Yes, China is an outlier.

Part C: Describe the relationship between the number of broadband subscribers and GDP. Include a complete description of the shape.

Answer: There is a positive relationship between GDP (in billions of USD) and the number of broadband subscribers. This relationship seems to be linear with the exception of a few atypical points.

Part D: Calculate and interpret the value of the correlation coefficient, r .

Answer: $r = 0.771$; there is a moderate to strong positive relationship between GDP (in billions of USD) and the number of broadband subscribers.

- 2) Continue using the broadband and GDP dataset, but now, turn on the option “Enable Dragging/Deleting of Points.” Remove the observation for China by dragging the dot representing China off of the plot.

Part A: What is the new mean GDP in billions of USD?

Answer: 1,706 USD

Part B: How did the relationship between the number of broadband subscribers and GDP change?

Answer: Now, there is a really strong positive relationship between GDP (in billions of USD) and the number of broadband subscribers. This relationship seems to be linear with no recognizable outliers.

Part C: Calculate and interpret the value of the correlation coefficient, r .

Answer: $r = 0.980$; there is an almost perfect, positive linear relationship between GDP (in billions of USD) and the number of broadband subscribers.

Part D: What effect did removing the extreme observation have on the value of the correlation coefficient, r ?

Answer: The correlation coefficient increased.

The formula to calculate the correlation coefficient r is:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Part E: In In-Class Activity 4.C, you learned that the mean is affected by outliers. Based on this formula, why do you think the value of r changed?

Answers will vary.

Sample answer: Since r is calculated using deviations from the mean, r is similarly sensitive to outliers that fall far from the invisible line.