2.3 Correlation (Part 1)

In the previous lesson, we used direction, form, and strength to describe the association between two quantitative variables.

To quantify the strength of a *linear* relationship between two quantitative variables, we use the **correlation *r***.

You will learn how to calculate the correlation in the next lesson.

In this lesson, we will focus on how to understand and interpret the correlation.

The **correlation *r*** is the measure of strength and direction of a linear relationship between two quantitative variables.

The correlation r is a value between -1 and 1 ($-1 \leq r \leq 1$).

If the relationship is negative, then $r < 0$.
If the relationship is positive, then $r > 0$.

If $r = 1$ or $r = -1$, then there is a perfect linear relationship. In other words, all of the points will be exactly on a line.

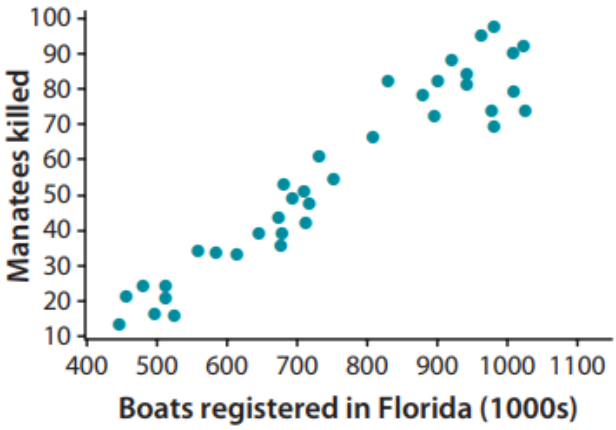
If there is very little scatter from the linear form, then r is closer to 1 or -1.
The more scatter from the linear form, the closer r is to 0.

Example: Manatees are large, gentle, slow-moving creatures found along the coast of Florida. Many manatees are injured or killed by boats. Here is a scatterplot showing the relationship between the number of boats registered in Florida (in thousands) and the number of manatees killed by boats for the years 1977 to 2013.

r > 0

Is the correlation positive or negative? Is it closer to 1 or -1? Explain your reasoning.

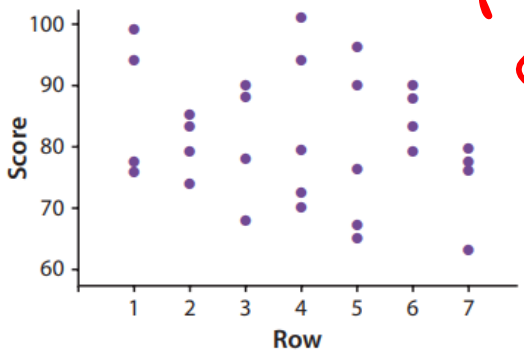
0



The correlation is closer to 1 because the data is closer together in a linear form.

Example: To see if seat location affects test scores, a statistics teacher randomly assigned students to seat locations in his classroom for a particular chapter and recorded the test score for each student at the end of the chapter. The explanatory variable in this experiment is which row the student was assigned, where Row 1 is closest to the front and Row 7 is the farthest away. Here is the scatterplot showing the relationship between row and test score.

r < 0



Is the correlation positive or negative? Is it closer to 0 or -1? Explain your reasoning.

0

The correlation is closer to 0 since the data is more spread out.