

3.29: Assignment- Linear Relationships

In this activity we will:

- Learn how to compute the correlation.
- Practice interpreting the value of the correlation.
- See an example of how including an outlier can *increase* the correlation.

Recall the following example: The average gestation period, or time of pregnancy, of an animal is closely related to its longevity—the length of its lifespan. Data on the average gestation period and longevity (in captivity) of 40 different species of animals have been recorded.

Instructions

Click on the link corresponding to your statistical package to see instructions for completing the activity, and then answer the questions below.

[R](#) | [StatCrunch](#) | [Minitab](#) | [Excel 2007](#) | [TI Calculator](#)

Remember that the correlation is only an appropriate measure of the **linear** relationship between two quantitative variables. First produce a scatterplot to verify that gestation and longevity are nearly linear in their relationship.

Instructions

Click on the link corresponding to your statistical package to see instructions for completing the activity, and then answer the questions below.

[R](#) | [StatCrunch](#) | [Minitab](#) | [Excel 2007](#) | [TI Calculator](#)

Observe that the relationship between gestation period and longevity is linear and positive. Now we will compute the correlation between gestation period and longevity.

Instructions

Click on the link corresponding to your statistical package to see instructions for completing the activity, and then answer the questions below.

[R](#) | [StatCrunch](#) | [Minitab](#) | [Excel 2007](#) | [TI Calculator](#)

Question 1:

Report the correlation between gestation and longevity and comment on the strength and direction of the relationship. Interpret your findings in context.

Now return to the scatterplot that you created earlier. Notice that there is an outlier in both longevity (40 years) and gestation (645 days). Note: This outlier corresponds to the longevity and gestation period of the elephant.

What do you think will happen to the correlation if we remove this outlier?

Instructions

Click on the link corresponding to your statistical package to see instructions for completing the activity, and then answer the questions below.

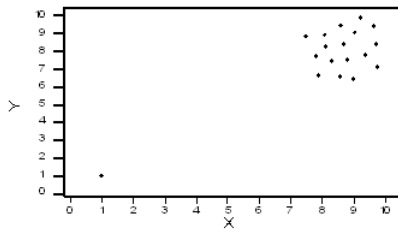
[R](#) | [StatCrunch](#) | [Minitab](#) | [Excel 2007](#) | [TI Calculator](#)

Question 2:

Report the new value for the correlation between gestation and longevity and compare it to the value you found earlier when the outlier was included. What is it about this outlier that results in the fact that its inclusion in the data causes the correlation to increase? (Hint: look at the scatterplot.)

Comment

In the last activity, we saw an example where there was a positive linear relationship between the two variables, and including the outlier just “strengthened” it. Consider the hypothetical data displayed by the following scatterplot:



In this case, the low outlier gives an “illusion” of a positive linear relationship, whereas in reality, there is no linear relationship between X and Y.

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

This page titled [3.29: Assignment- Linear Relationships](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Bill Pelz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.