Bookmark this page

Statistics Package Exercise: Computing a Correlation

Learning Objective: Interpret the value of the correlation coefficient, and be aware of its limitations as a numerical measure of the association between two quantitative variables.

In this activity we will:

- learn how to compute the correlation (r).
- 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.

- R | StatCrunch | TI Calculator | Minitab | Excel

R Instructions
To open R with the dataset preloaded, right-click here and choose "Save Target As" to download the file to your computer. Then find the downloaded file and double-click it to open it in R.

The data have been loaded into the data frame

```
a
```

. Enter the command

```
a
```

to see the data. The variables in

```
a
```

are

```
animal
```

,

```
gestation
```

, and

```
longevity
```

.

- `animal` : the name of the animal species
- `gestation` : the average gestation period of the species, in days
- `longevity` : the average longevity of the species, in years

Notice that the correlation between gestation and longevity has changed.

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.

- R StatCrunch TI Calculator Minitab Excel

R Instructions
To do this in R, copy the command:

```
plot(a$longevity,a$gestation,xlab="Average Longevity of Species (years)", ylab="Average Gestation Period of Species (days)")
```

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

- R StatCrunch TI Calculator Minitab Excel

R Instructions
To do that in R, copy the command:

```
cor(a$longevity,a$gestation)
```

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.

Learn By Doing (1/1 point)

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

Your Answer:

The r is 0.6632397. That means it's a mildly strong positive correlation -- so the longer the longevity of an animal, the longer the gestation period.

Our Answer:

The correlation is r = 0.663. The relationship is moderately strong and positive. The positive direction of the relationship means that, in general, animals that have a longer lifespan also have longer pregnancies. The moderate strength of the relationship tells us that there is some considerable variation around the relationship's linear form.

Resubmit Reset

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?

- R StatCrunch TI Calculator Minitab Excel

R Instructions

To do this in R, copy the following command:

```
cor(a$longevity[a$animal!="elephant"],a$gestation[a$animal!="elephant"])
```

Notice that the correlation between

```
gestation
```

and

```
longevity
```

has changed.

Learn By Doing (1/1 point)

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.)

Your Answer:

The new r value is 0.5190389; the outlier skewed the correlation, making the relation between longevity and gestation appear stronger than it was.

Our Answer:

After removing the outlier, the correlation drops from .663 to .519. The elephant is an outlier, however, it is consistent with the linear form of the relationship between longevity and gestation period. In this case, it would be most informative to report both correlations.

Resubmit Reset