

Biostatistics I: Descriptive Statistics

Correlation

Eleni-Rosalina Andrinopoulou

Department of Biostatistics, Erasmus Medical Center

✉ e.andrinopoulou@erasmusmc.nl

🐦 [@erandrinopoulou](https://twitter.com/erandrinopoulou)

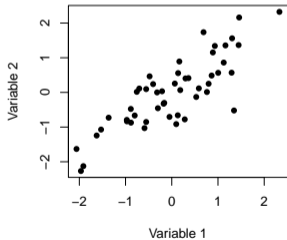
In this Section

- ▶ Correlation coefficients
- ▶ Examples

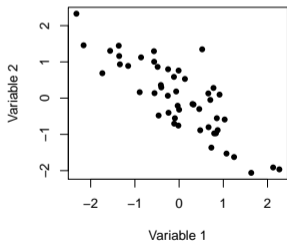
Correlation

Correlation is a measure that describes the strength of the association between two variables. Let's assume that we have two continuous variables, we can get the following relationships:

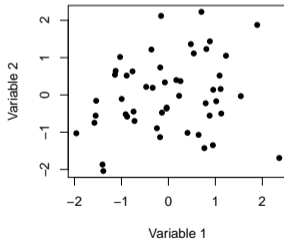
Positive correlation



Negative correlation



No correlation



Pearson Correlation

- ▶ magnitude of association
- ▶ linear association
- ▶ direction of the relationship

A relationship is linear when a change in one variable is associated with a proportional change in the other variable

Pearson Correlation: $corr(X, Y) = \frac{cov(X, Y)}{sd(X)sd(Y)}$,

where $cov(X, Y)$ is the covariance and $sd(X)$, $sd(Y)$ are the standard deviations

Spearman Correlation

- ▶ direction of the relationship
- ▶ monotonic relationship

In a monotonic relationship, the variables tend to change together, but not always at a constant rate (as in the linear case)

The Spearman correlation coefficient is based on the ranked values:

$$\text{corr}_R(X, Y) = \frac{\text{cov}(R_X, R_Y)}{\text{sd}(R_X)\text{sd}(R_Y)}$$

What is rank?

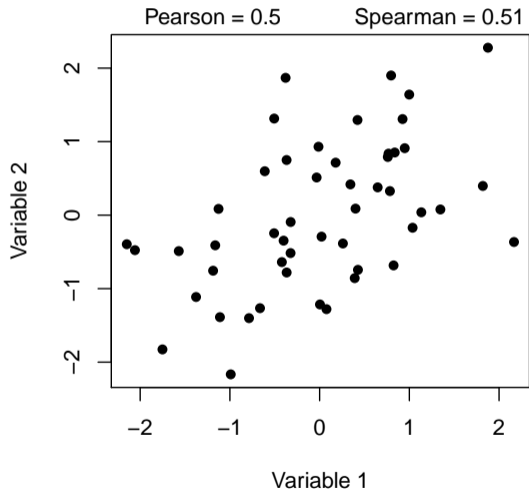
Ranks are integers indicating the rank of some values. E.g. the rank of 3, 10, 16, 6, 2 is 2, 4, 5, 3, 1:

```
rank(c(3, 10, 16, 6, 2))
```

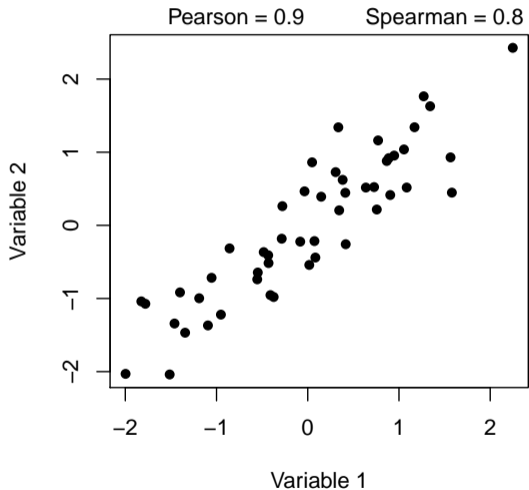
```
[1] 2 4 5 3 1
```

Difference between Pearson and Spearman

Weak positive correlation:

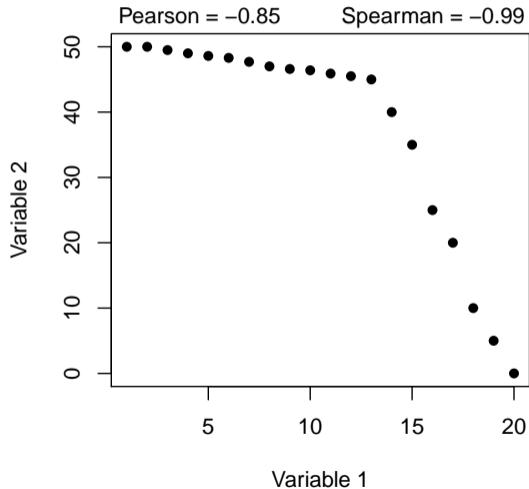


Strong positive correlation:

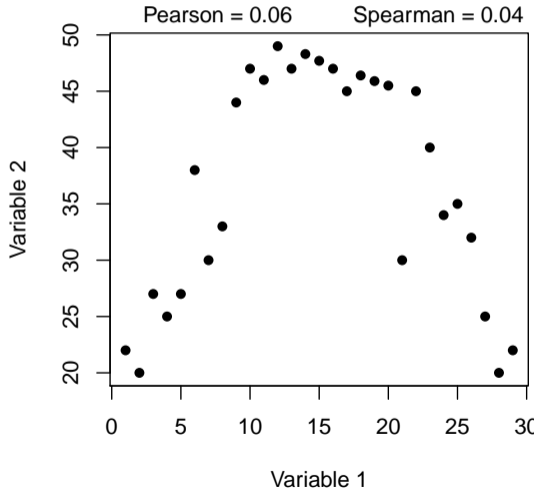


Difference between Pearson and Spearman

Linear VS monotonic relationship:



What if there is a correlation but this is not linear?



Other Correlation Measures

- ▶ **Point-Biserial:** It evaluates the association between a continuous variable with a categorical (dichotomous) variable
- ▶ **Intraclass:** It evaluates the association between two continuous variables that are structured in groups

Note

- ▶ Correlation must not be confused with causality
- ▶ If two variables are correlated, it does not imply that one variable causes the changes in another variable