Quantitative analysis methods for public policies

Correlation Analysis

Co-funded by the Erasmus+ Programme of the European Union

Correlation

- It is one of the most useful concepts in Statistics since it examines the relationship between two or more variables, usually continuous.
- Examples include the relationship between hours of study and school achievement scores, level of stress and grades, etc.
- The researcher wishes to investigate two things:

(a) Is there any "meaningful" relationship between the two variables?

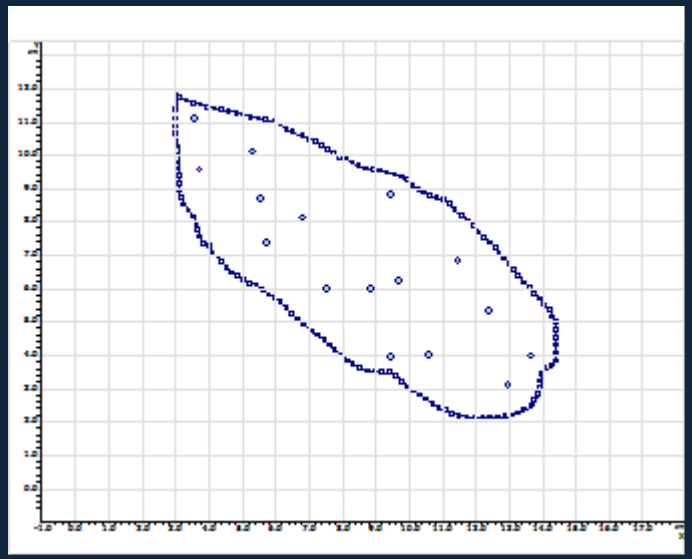
(b) If yes, how strong is this relationship?

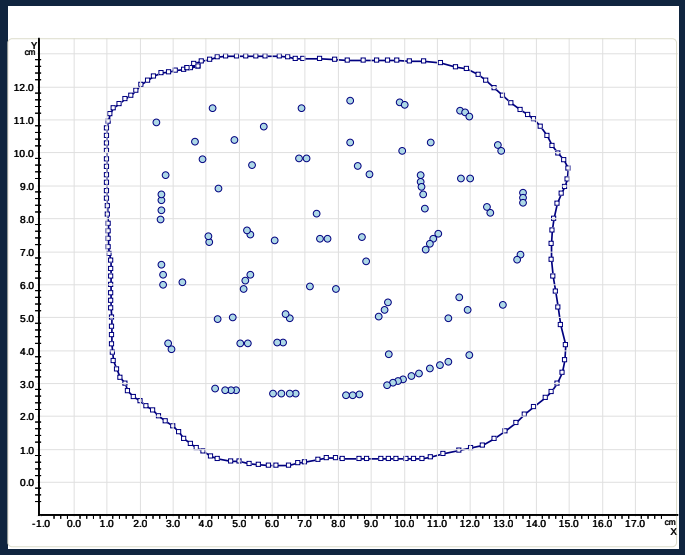
- Correlation analysis deals with those issues providing quantifiable answers.

Co-funded by the Erasmus+ Programme of the European Union

Using diagrams

Someone may claim that this information is provided by a simple diagram. This is partly true since diagrams answer mainly to (a) and not to (b). Below, there are two diagrams where a relationship seems more established in the left hand side than in the right hand side. However, not much more can be said.





Co-funded by the Erasmus+ Programme of the European Union

Pearson correlation coefficient

- It pertains to two continuous variables.
- Both variables have to follow (at least roughly) the normal distribution
- Also, they have to have a linear relationship. That means that a diagram between these two variables should display a line and not a curve.
- It takes values between -1 and +1.
- Values close to +1/-1 indicate positive/negative relationship.
- Thus, a correlation coefficient of 0.8 between hours of study and levels of stress tells us that the more someone studies, the more likely that she/he is nervous.
- Values close to 0 indicate no relationship

Co-funded by the Erasmus+ Programme of the European Union

Range of values

Below is a description of the range of values for a correlation coefficient:

- **0-0.20: non-existent or small positive relationship**
- **0.20-0.40: small to medium positive relationship**
- **0.40-0.60: medium to strong positive relationship**
- **0.60 and above: strong to very strong positive relationship**

In social sciences, most of the values will be under 0.50. It is highly unlikely to get values higher than 0.80.

Negative values have the same meaning as above in terms of the strength of the relationship but the direction is negative.

Co-funded by the Erasmus+ Programme of the European Union

Spearman correlation coefficient

- It takes the same range of values as Pearson
- Furthermore, its values have the same interpretation as before
- It is more flexible, since it does not require Pearson's assumptions of normality and linear relationship between the two variables
- For an applied practitioner it is often difficult to try to gauge the degree that two variables are normally distributed and have a linear relationship.
- Hence, we suggest that the researcher produce both coefficients through SPSS.
- In case of considerable discrepancy, Spearman is more appropriate since it is less strict than Pearson.

Co-funded by the Erasmus+ Programme of the European Union

Further comments

- Spearman can be used also in case of at least one discrete (ordinal) variable although other techniques (t-test, ANOVA, chi-square) or coefficients (Kendall) may be more appropriate.
- Correlation DOES NOT imply causation, just association. To establish causation, more is required mainly the use of an experimental design.
- Quite often, a correlation coefficient is associated with a hypothesis testing whether it is statistically significantly different from zero. Since, many studies have large sample sizes thus distorting the meaning of the test, we suggest that these results not be considered and stick to the interpretation of the values.

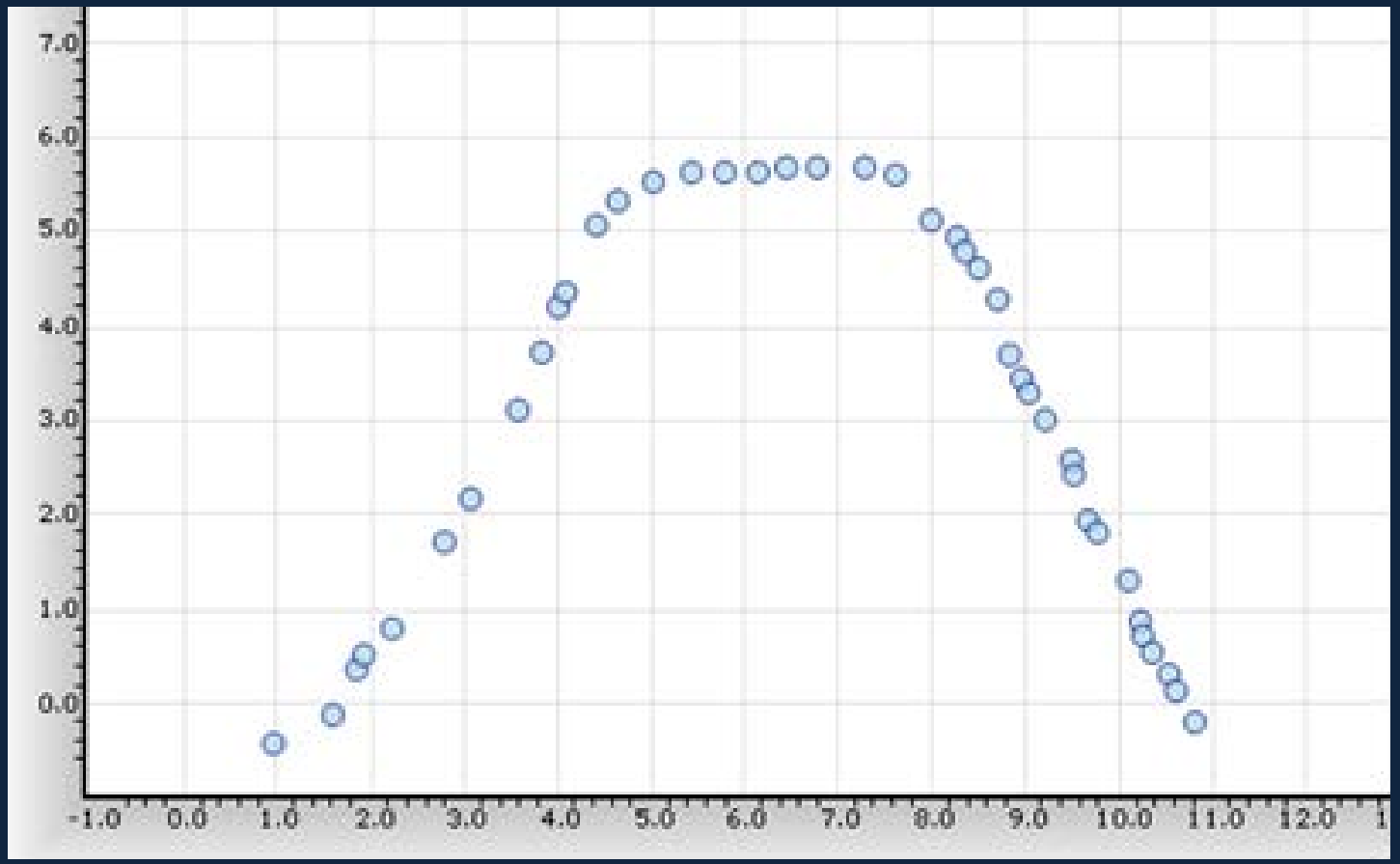
Co-funded by the Erasmus+ Programme of the European Union

Partial correlation coefficient

- Person and Spearman analyze only two variables each time. However, most of the problems have more than that.
- An extension is the partial correlation coefficient where two variables are correlated adjusted for the effect of other variable(s).
- Example: The relationship between hours of study and student grade may be incomplete due to the involvement of other variables such as socioeconomic status, IQ, etc.
- Hence, we need to have a procedure where the relationship of these two variables is measured on people of the same say, socioeconomic status.
- Since this is not easily done in the initial stages of sampling the partial correlation coefficient adjusts the effect of external variable(s).
- Its values and interpretation are the same as before with the note that the relationship of the two variables is adjusted for the effect of other variables(s).

Co-funded by the Erasmus+ Programme of the European Union

Interpret this diagram!



Co-funded by the Erasmus+ Programme of the European Union