Course : Research statistics **Lecturer : Pr. Saliha CHELLI**

Level : Master

Lecture : 11

Pearson correlation test

Objective : Introducing correlational analysis and how to compute Pearson correlation coefficient

Lecture objectives : Introducing the independent-samples t-test and how to compute it.

Introduction

Quantitative research can be descriptive, correlational, experimental or quasi-experimental. The goal of correlational research (associational research) is to determine whether a relationship exists between variables and if so, the strength of that relationship and its direction. This is often tested statistically through correlations, which allow the researcher to determine how closely two variables are related (DÖrneiy, 2007). This lecture introduces one way of statistically testing correlation called the Pearson's correlation r.

1. Experimental research versus correlational research

In Experimental research studies, researchers manipulate one or more variables (independent variables) to determine the effect on another variable (dependent variable). This manipulation is described as a treatment and the researchers' goal is to determine whether there is a causal relationship. Experimental and quasi-experimental reseach are two research designs which differ in the use of participants. According to the experimental design, participants are randomly assigned to either the treatment group or the control group, whereas they are not assigned randomly in the quasi-experimental design. In general, intact groups or already existing ones are used**.** Unlike experimental research, correlational reseach is descriptive because there is no manipulation of the independent variable; its aim is just to find if there is a realtionship between two variables or more.

Correlational research can be used in different ways:

-to test a relationship between variables and to make predictions.

- It estabishes a statistically relationship between them.

2. Types of correlation

a .**Positive**: there is an incresase or decrease in bithe variables.

b.**Negative**: variables are opposite, for example, when a variable increases, the other one decresases.

c. **No correlation** : zero correlation: variables are not stastically correlat

3. Correlational statistical tests

a.The chi-square test is non-parametric used to test relationship between categorical variables.

b.Pearson r correlation test is a parametric test , the most widely used correlation statistic to measure the degree of the relationship between linearly related variables , measured on interval or ratio scale (Marczyk, DeMatteo& Festinger, p.218).

c.Kendall rank correlation **is a** non-parametric test . It investigates the significance of the correlation between two series of observations obtained in pairs (Kanji-Gopal, 2006, p. 79).

***d.Spearman rank correlation* coefficient is** a non-parametric test used to find realtionship when the variables are measured on an ordinal scale.

3 .Pearson r correlation

This kind of correlation analysis is used to quantify the association between two continious variables (an independent and dependent variable . In this kind of analysis, we estimate a sample correlation coefficient, denoted as r. Example: Is there a positive correlation between exam scores and time allotted to the exam? Is there a positive correlation between first exam scores and final exam scores.

Null hypothesis: There is no correaltion between first exam scores and final exam scores.

Alternative hypothesis: there is a positive correlation between first exam scores and final scores exam

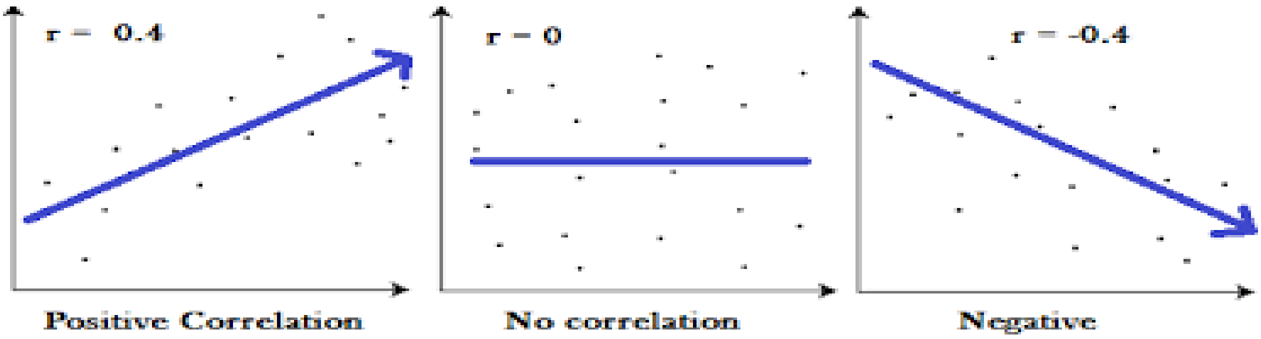
3.1. Pearson correlation coefficient

The correlation coefficient ranges between – 1 and + 1 and quantifies the direction and strength between the two variables. The correlation may be positive or negative (Marczyk, DeMatteo& Festinger,2005, p.216) The sign of the correlation coefficient indicates the direction of the association. The magnitude of the correlation coefficient indicates the strength of the association. For example, a correlation of r = 0.8 reveals a strong , positive association while r = -0. 3 shows a weak negative association. A correlation close to zero suggest no association between the two continious variables. Correlation strength is shown below.

Correlation Strength vs. Axes	
Perp. Distance	Conclusion
distance = 1	Perfect Correlation
0.9 ≤ distance	Very Strong Correlation
0.7 ≤ distance < 0.9	Strong Correlation
0.5 ≤ distance < 0.7	Moderate Correlation
0.3 ≤ distance < 0.5	Weak Correlation
0 < distance < 0.3	Very Weak Correlation
distance = 0	No Correlation

The Pearson correlation coefficient is one of many types of coefficients in the field of statistics.It is a helpful statistical formula that measures the strength between variables and relationships. As stated previously the value of r ranges between – 1.00 and + 1.00. If the value is in the positive range, this means that the correlation is positive. But if it is in the negative range, it is negative.

Correlation is displayed by scatter plots as shown in these figures



4. Calculating the value of Pearson correlation coefficient (r)

Step one: Make a table with your data for two variables, label the variable (x) and (y) and add three more columns labeled (xy), and (x^2) and y^2 as shown in the contingency table.

Step two: Complete the chart using x and y values, multiple x and y and find x^2 and y^2.

Step 3: After that find the sums.

Sample	x	y	xy	x^2	y^2
1					
2					
3					
Sum					

Step 4: Use the formula to find Pearson correlation value (r)

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

N = number of pairs of scores
Σxy = sum of the products of paired scores
Σx = sum of x scores
Σy = sum of y scores
Σx^2 = sum of squared x scores
Σy^2 = sum of squared y scores

5. Testing the significance of the correlation coefficient

Testing the significance is to decide whether the relationship is strong enough to be used to model relationship in the population. One way of making the decision is comparing the value of r to the appropriate critical value based on the sample; in this case we need to have :

- the degree of freedom: **df= n-2**
- alpha : **0.05**
- and whether the hypothesis **is one tailed or two-tailed**

If the value of r is **<** negative critical value or r > positive critical value, the r is significant.

Suppose We have a sample of 12 **df**= 12-2= **10**

We calculated r= **0.872**

We choose alpha=**0.05**

The Critical value of df= 10 at alpha level 0.05 is **0. 497** (refer to the table in the appendix)

0.842>**0.497** → r is significant.

References

DÖrneiy, Z. (2007). Research methods in applied linguistics: Quantitative, qualitative & mixed methodologies. Oxford: Oxford University Press.

Kanji-Gopal, K. (2006). 100 statistical tests (3rd ed). London: Sage Publications.

Marczyk, G, DeMatteo, D, & Festinger,D. (2005).Essentials of research design.and methodology. John Wiley & Sons, Inc.

Recommended references

Larson-Hall, J. (2010). Doing statistical analysis in second language research using SPSS.

Pearon correlation formula : https://www.questionpro.com/blog/pearson-correlation-coefficient/

Appendix : Critical value table of Pearson correlation

df =N-2	Level of significance for a one-tailed test			
	.05	.025	.01	.005
	Level of significance for a two-tailed test			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
20	.360	.423	.492	.537
30	.296	.349	.409	.449
40	.257	.304	.358	.393
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.256	.284
90	.173	.205	.242	.267
100	.164	.195	.230	.254
∞	.073	.087	.103	.114