

# Statistics

## Lecture 11

### Correlation analysis



By Prof.Chelli

---

### Lecture objective

**Introducing correlational analysis and how to compute Pearson correlation coefficient.**



# 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  $r$ .



## **Experimental 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 versus quasi-experimental research

---

These two research designs 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.



## Correlational research

---

Unlike experimental research, **correlational research is descriptive** because there is **no manipulation** of the independent variable; its aim is just to find if there is a relationship between two variables or more.

**Correlational research** can be used in different ways:

- to test a relationship between variables and to make predictions.
- It establishes **a statistically** relationship between them.



## Types of correlations

---

Correlation can be:

1. **Positive:** there is an increase or decrease in in the variables.
2. **Negative:** variables are opposite, for example, when a variable increases, the other one decreases.
3. **No correlation** : zero correlation: variables are not stastically correlated.

## Correlational statistical tests

- **The chi-square test** is used to test relationship between **categorical variables**.
- ***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,2005, p.218)
- ***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).
- ***Spearman rank correlation coefficient*** is a non-parametric test used to find relationship when the variables are measured on an ordinal scale



## Pearson r correlation

---

This kind of correlation analysis is used to quantify the association between two continuous 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 correlation between first exam scores and final exam scores.
  - **Alternative hypothesis:** there is a positive correlation between first exam scores and final scores exam.

# 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 suggests no association between the two continuous variables

# continued

---

<b>Correlation Strength vs. Axes</b>	
<b>Perp. Distance</b>	<b>Conclusion</b>
distance = 1	Perfect Correlation
$0.9 \leq \text{distance}$	Very Strong Correlation
$0.7 \leq \text{distance} < 0.9$	Strong Correlation
$0.5 \leq \text{distance} < 0.7$	Moderate Correlation
$0.3 \leq \text{distance} < 0.5$	Weak Correlation
$0 < \text{distance} < 0.3$	Very Weak Correlation
distance = 0	No Correlation



# Pearson correlation coefficient

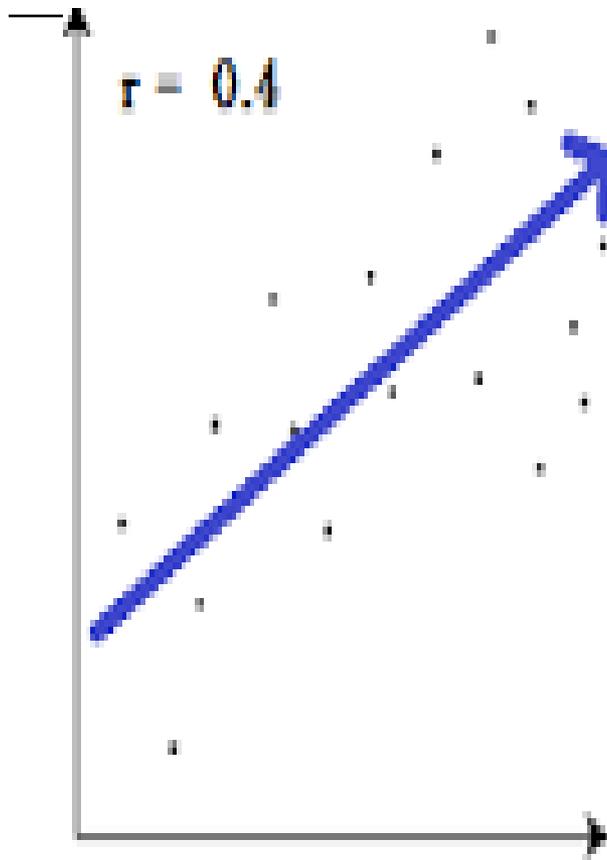
---

The Pearson correlation coefficient denoted as  $r$  is one of many types of coefficients in the field of statistics.

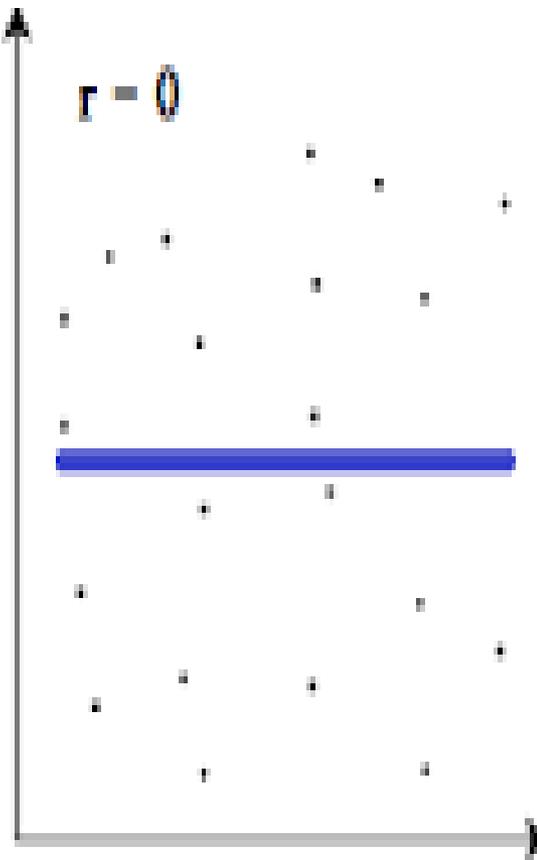
It is a helpful statistical formula that measures the strength and the direction of the relationships between variables.

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.

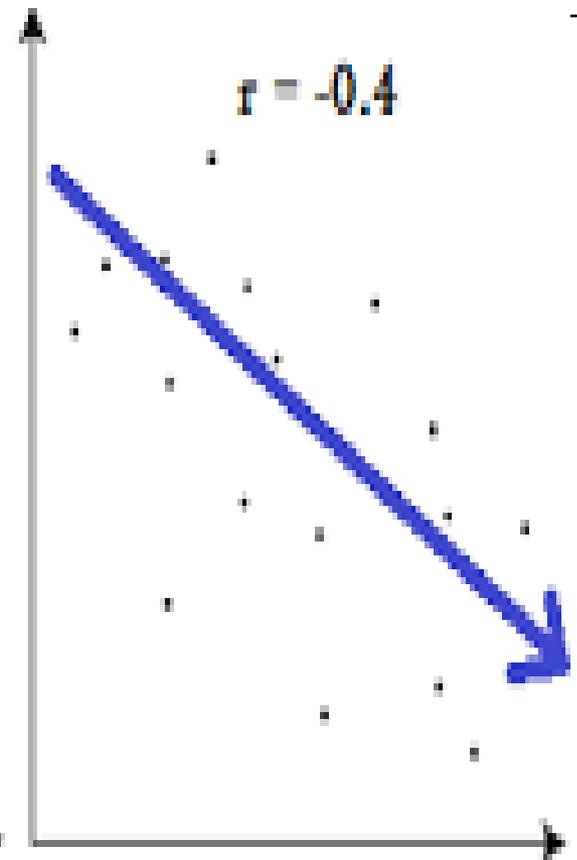
**Correlation is displayed by scatter plots as shown in these figures**



Positive Correlation



No correlation



Negative

# Calculating the value of Pearson correlation coefficient (r) manually

- **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  $y^2$  like this.
- **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					
<b>Sum</b>					

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

---

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

Where:

- N = number of pairs of scores
- $\sum xy$  = sum of the products of paired scores
- $\sum x$  = sum of x scores
- $\sum y$  = sum of y scores
- $\sum x^2$  = sum of squared x scores
- $\sum y^2$  = sum of squared y scores

The value found for  $r$  indicates the magnitude of the correlation as shown in the diagram below

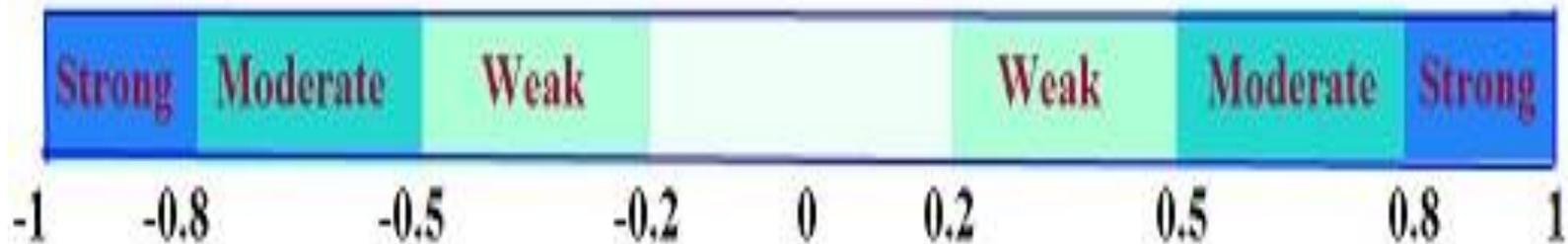
---

## Pearson Correlation Coefficient

Negative Correlation

No Correlation

Positive Correlation





## **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 less ( $<$ ) than the negative critical value or greater ( $>$ ) than the positive critical value, then the correlation is statistically significant.
  - The **critical value** is obtained **from a statistical table**, specifically from statistics books

# Example

---

## Suppose

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

We calculated  $r = \mathbf{0.872}$

We choose  $\alpha = \mathbf{0.05}$

**Hypothesis= two-tailed**

**Critical value of df= 10 at alpha level 0.05 is 0.576**

$\mathbf{0.842} > \mathbf{0.576} \longrightarrow$  the results show a strong positive correlation statistically significant at 0.05

## Example with SPSS Values

---

A Pearson correlation analysis was performed using SPSS to investigate the relationship between students' motivation and academic achievement. The results showed a **moderate positive correlation**,  $r = .45$ , with  $df = 28$ , which was **statistically significant at the 0.05 level (two-tailed)** ( $p = .012$ ). This suggests that higher levels of motivation are associated with higher academic achievement.

Simply  $r = .45 > P = .012$   $\longrightarrow$  the results show a moderate positive correlation statistically significant at 0.05 level.

### If the Result are NOT Significant

The Pearson correlation analysis revealed a **weak negative correlation**,  $r = -.18$ , with  $df = 28$ , which was **not statistically significant** ( $p = .34$ ). This indicates that no statistically meaningful relationship exists between the two variables.

Simply  $r = -.18 < P = .012$   $\longrightarrow$  the results show a weak negative correlation, not statistically significant at 0.05 level.



## Key SPSS Elements You Should Always Mention

---

- ✓ Type of correlation (Pearson / Spearman)
- ✓ Value of  $r$
- ✓ **p-value (Sig. 2-tailed)**
- ✓ Direction (positive / negative)
- ✓ Strength (weak / moderate / strong)
- ✓ Conclusion about significance

# References

---

- Dörnyei, Z. (2007). Research methods in applied linguistics: Quantitative, qualitative & mixed methodologies. Oxford: Oxford University Press.
- Kanji-Gopal, K. ( 2006). 100 statistical tests 3<sup>rd</sup> 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.
- Pearson correlation formula : <https://www.questionpro.com/blog/pearson-correlation-coefficient/>
-