UNIVERSITY OF WESTERN MACEDONIA

Advanced Research Methods

Statistics:

Course Presentation

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Lies, Damn Lies and Statistics

- Cynics sarcastically say that you can prove anything with statistics. Others argue that you can't do anything with statistics. Some remind us that Statistics is a way of lying.

An example:

- The vast majority of Greeks have a higher than average number of lower limbs.

- Indeed: There are 500 of our compatriots who have no lower limbs (amputations, etc.) Two and a half thousand have one leg. The rest (10 997 000) have two legs.

- Average lower limbs of Greeks = $500\times0+2.500\times1+10.997.000\times2$ / $11.000.000 = 1{,}999682$.

- The vast majority of Greeks have 2 legs, $2>1.999682$

M. Manolopoulos

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

The Importance of Good Statistical Knowledge

- Why Statistics is more important than mathematics

- Why psychology students hate statistics:
    - Breaking News - Psychology Students Hate Statistics!
    - ...and why you should not be afraid of it!

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Variables and their measurement

- Variable: Any quantity/property/attribute that can take variable values.
- Any event or characteristic of phenomena, objects or organisms that can be changed and measured.
- Different ways of categorising variables
  - Quantitative - Qualitative
  - Continuous - Discretre
  - Independent - Dependent
- Importance of the measuring instrument

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Independent and Dependent Variables

Independent Variable

The variable that the researcher changes in order to determine the effect of the change on another variable. In a cause-and-effect relationship, the independent variable plays the role of cause

Dependent Variable (Dependent Variable)

The variable affected by the changes the researcher makes to the independent variable. In a **cause-effect** relationship, the dependent variable plays the role of the outcome

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

IV or DV, depending on the type of problem under consideration.

Independent Variable		***Dependent Variable***
Student attitude towards school	→	Student's performance at school

Independent Variable		***Dependent Variable***
Educational level of parents	→	Student attitude towards school

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

4 levels of measurement

- Categorical or Nominal variables
- Ordinal variables
- Equal interval variables
- Ratio variables

NON-PARAMETRIC - No Distribution Assumptions

PARAMETERS - Normal Distribution

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Test - Examples

The state of health of a patient (very serious, serious, moderate, normal)

The level of education (primary, secondary, tertiary, etc.)

The degree of one's satisfaction with a product (very much / much / a little / not at all)

Nationality (Greek, French, German, Finnish, French, Korean)

The finish time in a marathon race.

Age → *RATIO*

Age group (10-15, 15-20, 20-25, ...) → *ORDINAL*

The evaluation system for the university's courses (inadequate = 1, moderate = 2, good = 3, very good = 4, excellent = 5)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Measurement of dependent variable

Ideally:

- Objective
- Quantitative

➢ Common ways of measuring reactions:

- Reaction accuracy
- Delay or reaction time: the time between the presentation of a stimulus and the beginning of the execution of a reaction
- Reaction rate: time necessary to complete a reaction
- Frequency of reaction: how many times in a given period of time
- Reaction intensity

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Types of Hypotheses

Declarative Hypothesis - H1 (Directional Hypothesis)

The researcher makes a prediction of the result he expects to find in the survey, according to the theoretical background

Examples:

- *There is a positive association between students' attitudes towards school and their performance*
- *Children with medium anxiety compared to children with low or high anxiety have the highest performance in written tests*

Null hypothesis - H0

This hypothesis indicates the absence of an effect or an association between the variables

Example:

There is no association between students' attitudes towards school and their performance

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

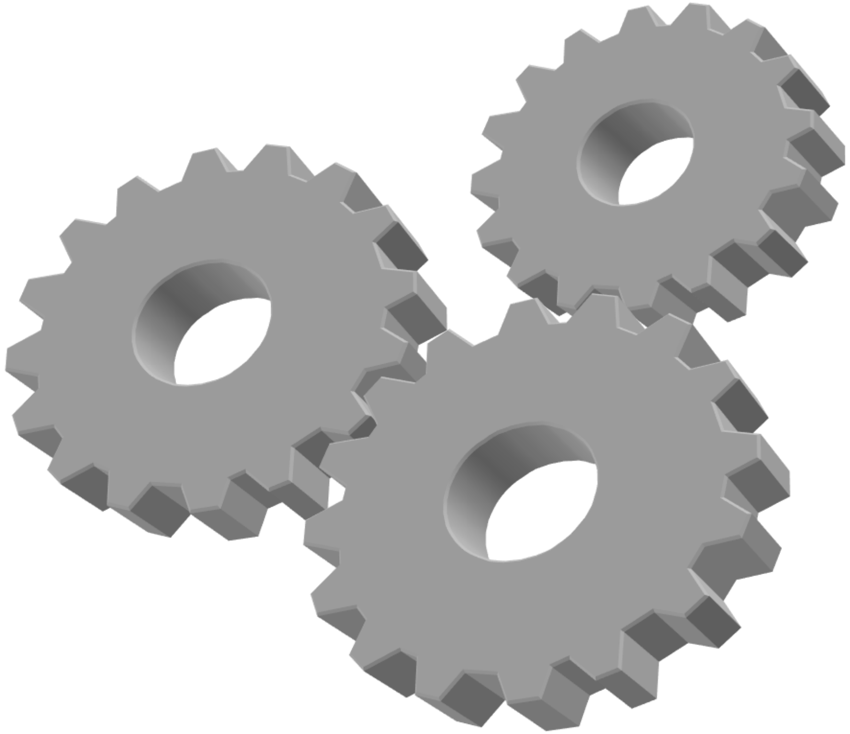
Useful Resources

- A brief introduction - a reminder of important methodological and statistical points
  - http://www.jolley-mitchell.com/Appendix/Introduction%20to%20Statistics.pdf

- Useful resources:
  - https://edge.sagepub.com/field5e
  - http://stackexchange.com/
  - Youtube tutorials (e.g. Psychology Research Hub)
  - Google searches...

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Sampling, Representativeness

Randomization

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Structure (Parts) of Psychological Measurements - Noise Levels

1. "**Actual score**" of the quantity we want to measure
2. "**Scores of other irrelevant quantities**" intruding in our measurement
3. **Systematic (non-accidental) bias**
   [Problem if it affects some participants more than others, e.g. experimenter bias]
4. **Random (non-systematic/occasional) error** (e.g. fatigue, boredom, ambiguity of questions, external distractions, noises) - eliminated by a large number of participants

❖We want as much of 1 and as little of the rest → <u>Critical role of sampling</u>

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SAMPLING:
SAMPLE - POPULATION

- Nature of the problem determines the type of participants
- The ideal sample should be representative

- **Sample**: The subset selected from a larger set of individuals of the same type (e.g. schizophrenia patients, schoolchildren, individuals with dyslexia, etc.) to represent it for the purposes of the research
- **Population**: broadest set of individuals from which the sample is drawn
  - A basic requirement in any sample-based research is that empirical findings should only be generalised to **populations** that have the same characteristics as the **sample** (external validity).
  - **Sample representativeness** is determined by:
    - Samplinh method / selection of included participants
    - Sample size

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Qualitative methods

- Targeting wealth of information
  - Intentional sampling
  - Maximum Variance Sampling
  - Outlier sampling
  - Average / typical case sampling
  - Theoretical or conceptual sampling
  - Homogeneous sampling
  - Critical sampling
  - Opportunistic sampling (after initiation of the study)
  - Snowball sampling
  - Confirmation/rejection sampling

- Sample/space size varies

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sampling Types

<u>Non-random sampling (without probability)</u>

- Convenient sampling
- Snowball sampling

<u>Random sampling (with probability)</u>

- Simple random sampling
- Stratified sampling
- Multi-stage sampling
- Cluster sampling

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Random Sampling as Generalizability

- Random sampling contributes to representativeness, and thus to generalisability
- Every member of the population has the same chance of being included in the sample
  - With this assumption we can statistically calculate the accuracy of the research sample
- The opposite is the *biased sample*, e.g. over-representation or under-representation (c.f. bias error)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sampling and noise

- **Random sampling error**: sampling inevitably leads to a difference between our measurement and the actual measurement of the population. It depends on
  - The degree of precision of the sampling we wish to have in our generalisations
  - The heterogeneity of the population
  - The magnitude of the difference between the sample value and the population value that is tolerable for the investigator
  - *Relation: error inversely proportional to the square root of the sample size, e.g. ½ error = 4x sample*

- **'Bias error'**: Biased sample selection (systematic error, fixed direction). Negligence/practical weakness of researcher.
  - Related to circumvention of random - unbiased sampling. E.g. children's IQ and self-selection

- *Increasing sample size primarily helps to reduce random error*

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Bias error and sample size

- Example: the US presidential election, 1936. Literary Digest magazine.
  - 10.000.000 voters
  - Sample selection through the telephone directory and car registration numbers
  - Systematic error: Not everyone had a phone/car
  - Non-representative sample: Differences in age, income, etc.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

The importance of randomization

- Measurement of a variable involves natural variation, e.g. repeated measurement of the same variable in the same individual or measurement of the same variable in several different individuals
    - Statistics s (or σ) → Standard deviation

- Score = True Score + error = TS + e
    - e = Noise (Experimenter effects, Instrument measurement error, Variable definition error/validity, Reliability problems)

- What we are interested in experimentally is whether our experimental manipulation causes a variation around the mean that is statistically more significant than this natural variation.
- **Randomization** separates the TS from the e
    - Because e is random, the independent variable has an individual effect on the variance of the TS

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

10 Factors That Play a Role in Determining Sample Size

- 1) Number of subgroups of the population to be compared
  - If we move on to further group comparisons, e.g. gender, socio-economic status
- Alternative approach: keeping these factors constant
  - pairwise equation (matching)
  - repeated measurements
- 2) *The size of the population*
  - Numerically small population can be represented by a smaller sample
  - But influence limited, required sample size does not increase proportionally to population size, but gradually, and up to a point

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

10 Factors That Play a Role in Determining Sample Size

- 3) Population heterogeneity with respect to the variable under study
  - The greater the heterogeneity, the larger the sample we need
- Alternative approach: limiting the scope of the population/population of interest
  - E.g. IQ 50 - 200 vs. 100 - 120, left-handed vs. right-handed
  - ❖Sample size depends not only on population size, but on how different the members are from each other
- 4) Are mediating factors controlled for?
  - The more the 3rd intervening factors, the greater the heterogeneity/dispersion of the population

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

10 Factors That Play a Role in Determining Sample Size

- 5) Desired degree of accuracy in our estimates (**confidence intervals**)
  - The narrower the desired range of the confidence interval (higher precision) [confidence level], the larger the sample required
- 6) Size of expected differences
  - The smaller the expected differences, the larger the sample size necessary to avoid being inconclusive, however small
- 7) How "novel" is the result of our research expected to be?
  - Previous research, theories vs. innovation

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

10 Factors That Play a Role in Determining Sample Size

- 8) The "leakage - loss" of participants
  - E.g. longitudinal study. Sample as large as possible.
  - Continuous representativeness check. Better 90% of a low initial sample than 30% of a large sample
- 9) Number of information we get from each participant
  - The less information - measurements, the more the sample.
  - E.g. case study vs. questionnaires
- 10) Degree of data reliability
  - The more accurate the measurements, the more "tidy" and controlled the data collection and analysis process, the fewer the necessary participants

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sample size calculation

- Approximate estimate:
  - Sizes in similar studies by other researchers
  - The more people willing/available
  - Depending on the expected study costs
- Statistical methods for estimating the necessary sample (e.g. power analysis)
  - ** Must be determined prior to conducting the survey, based on the desired degree of accuracy of the results, and the maximum measurement error that would be tolerated
  - ** **Acceptable sampling error rate**
    ($p < .05$; $p < .01$; $p < .001$)

- Required size depends on:
  - Type of research (e.g. psychophysiology vs. personality)
  - Experimental design (e.g. independent samples vs. repeated measures)
  - Number of independent variables and their levels
  - Recommended 10-20% above minimum required size → losses (can also damage representativeness)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Ideal Sample - General Recommendations

- Size sufficient to avoid:
  - Type I error: rejecting a null hypothesis when it is correct
  - Type II error: accepting a null hypothesis when it is false
- Pilot survey for pre-testing both the sampling procedure and the sampling result

1. **Calculation of correlation indicators (correlation)**: 100 participants, not less than 50, minimum 30 per variable
2. **Comparison of means/standard deviations**: 30-50 per group, minimum 15 per group
3. **Comparison of percentages** (e.g. control x^2): 10 (5) per facsimile - combined group
4. **Sampling**: fewer/more people for individual research questions (e.g. 350 questionnaires)
5. **Parametric (30^+) - Non-parametric (30^-)**

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

3 Research objectives: reliability, validity, generalisability

- **Validity**: Measuring what we were interested in measuring.
- **Reliability**: Re-executability by us or someone else
  - Indicates the degree of stability of measurements/agreement between repeated measurements under the same conditions.
  - It is expressed by a numerical value - index: **Reliability coefficient** (0-1)
- **Generalisability (external validity)**: Applicability beyond the specific participants, under the specific circumstances

- **+Importance**: Missing in the absence of all the above

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

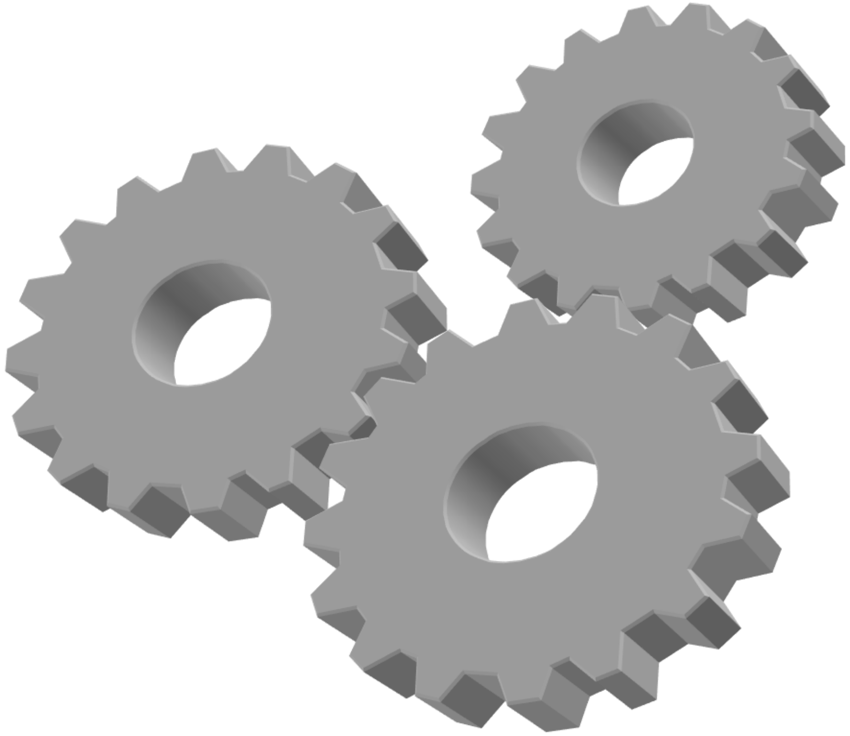
Structure (Parts) of Psychological Measurements

1. "**Actual score**" of the quantity we want to measure
2. "**Scores of other irrelevant quantities**" intruding in our measurement
3. **Systematic (non-accidental) bias**
   [Problem if it affects some participants more than others, e.g. experimenter bias]
4. **Random (non-systematic/occasional) error** (e.g. fatigue, boredom, ambiguity of questions, external distractions, noises) - eliminated by a large number of participants

❖We want as much of 1 and as little of the rest (***2 - validity, 4 - reliability***)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

<u>Types of research</u>: Correlational, Experimental, Observational methods

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Typical Quantitative Research Designs

- Experimental designs
  - Comparison of interventions/practices or groups
  - Interference in one or more groups, no interference in another group
- Correlational designs
  - Cases of difficulty of intervention / classification of people in the groups we want
  - Checking correlation (association or covariation) between variables
- Survey designs
  - Description of trends in a population
  - Sample
  - Questionnaire/scale
  - Attitudes, opinions, behaviours, characteristics
  - E.g. gallop

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Research methods

- <u>Correlational</u>
  - Associations between variables
- <u>Observation methods (Questionnaire, Interview, Observation, Focus Groups)</u>
  - No direct handling of variables
  - Systematic/comprehensive recording of behaviours to get the best generalisability (trends)
- <u>Quasi-experimental</u>
  - Intermediate solution to problems that prevent us from performing a (fully) experimental study
- <u>Experimental</u>
  - Handling one or more variables
  - Effective way of measuring them
  - ✓ <u>Causality</u>

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Research Types I: The Experimental Method

- Causality isolation, Direct variable handling, Random variable control
  - Distribution of participants in experimental and control groups (e.g. Clinical Psychology: placebo - no drug)

- <u>Experimental designs</u>
  - **Independent measurements** design (Between-groups design)
    - 2 or more groups for each condition
    - Each participant participates 1 time, in 1 of the conditions
  - **Within-subjects / Repeated measures** design
    - Each participant participates in all conditions
  - **Mixed designs**

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Randomization

- Between-groups design
  - Randomisation between experimental groups absolutely necessary
  - Avoids allocating participants in a biased way, which can lead to systematic differences

- Within-subjects design
  - Randomization (or counterbalancing) of the order of participation in each experimental condition
  - One experimental condition can affect the next one/s!

➢ Randomization Table/Computer Randomization Generator/Excel

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Independent Measurements. Between-groups Design

+

- Simplicity (e.g. counterbalancing)
- Avoiding the possibility of carry-over effects
- *Practical difficulties of each participant's participation in all experimental conditions*

-

- Costly in terms of time, effort, number of subjects
- Lack of sensitivity to experimental manipulations (statistical reasons - statistical power, noise, systematic differences between groups, etc.)
  - ✓ C.f. Group equation (matching)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Repeated Measures. Within-Subjects Design

+

- Economy
- Sensitivity - less "noise"
  - ✓ The ultimate equation (matching)!

❖ **Wherever possible, preferable to comparing different groups!!**

-

- Carry-over effects from participation in previous conditions, systematic differentiation
  - ✓ Fatigue, boredom, practice
  - ❖ Randomization of the order of conditions
  - ❖ Counterbalancing - can be a variable in statistical analysis e.g. A-B, B-A
- Independence between belonging to each condition

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Experimental Design

- Experimental group VS control group
- 1 independent variable with 2 levels (presence - absence of experimental manipulation)
    - Frequently also covariate analysis

❖ More complex experimental designs:

- Multiple levels of independent variables
- Latin square designs
- Multi-factorial designs
- Multi-factorial designs with repeated measurements

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple levels of independent variables

- Example: drug dosage.

- *Important*: Control of order effects
  - **Latin square**. Counterbalancing: each possible set of conditions is shown only once.
  - **Weighted Latin square** (when we have an even number of conditions, e.g. A before B more times than B before A)

- 2, 3 or more independent variables OK for correlation, difficult more than 3 for experimental design

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Latin Square Designs (Latin Square Designs)

- *Problem*: They do not completely eradicate serial effects, e.g. A before B more times than B before A
- *Solution:* weighted Latin Square design (when we have an even number of conditions)

- A B C D
- B D A C
- D C B A
- C A D B

- With more independent variables, it can be calculated automatically: http://www.edgarweb.org.uk/choosedesign.htm

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multivariate Experimental Designs

- 2, 3 or more independent variables
- OK for correlation, difficult with more than 3 for experimental design
- Interactions of independent variables: More universal/generalizable research
- Between subjects design: the number of necessary participants is greatly increased

- E.g. Air traffic controllers
- Shifts: Morning, Noon, Evening
- Room temperature: Cold, warm

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Descriptive Statistics

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

What is Statistics Used For?

- Exploratory Data Analysis (EDA) – Descriptive Statistics:
  - Summary
  - Finding patterns
  - Identifying potential causes and underlying structure (forming hypotheses)

- Confirmatory Statistics: Verification (or testing) of hypotheses
  - Inferential Statistics: Checking if a hypothesis is justified based on our data

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Exploratory Data Analysis

- Pre-analysis Checks: Verification for potential errors/omissions before starting statistical analysis
  - Omissions/Errors – e.g., data copying to the computer
  - Missing Data – we can estimate them based on the remaining data if we want to avoid losses
  - Outliers – if not the result of an error, they can undergo mathematical transformation
- Descriptive Tables – graphs
- Checking for patterns

Co-funded by the European Union

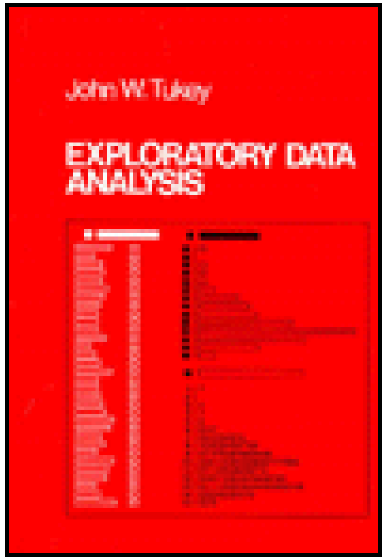
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

EDA - Slogan

- It all started with a scientist/book:
  - Tukey,J.W. (1977) Exploratory data analysis Addison-Wesley

“Exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone—as the first step.”

“It is important to understand what you CAN DO before you learn to measure how WELL you seem to have DONE it.”



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

A Research Example

- Language Comprehension Ability
- ΣΤ` elementary school grade, 50 students
- Language ability test
- Scale 1 – 10 (1 = lowest, 10 = highest)

- 5 8 7 4 6 10 5 8 6 7 6 7 5 9 4 7 4 4 5 8 6 1 6 7 5 3 5 9 5 10 4 8 6 3 6 1 5 7 4 8 9 6 3 6 5 6 7 5 8 6

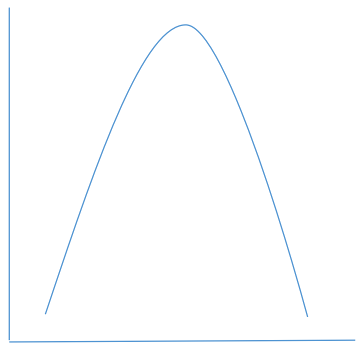
Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Descriptive Statistics: Numeric - Graphic Methods

- Descriptive Statistics:
  - Summary presentation of the sample and values checking

- Numeric methods
  - Measures of central tendency
  - Measures of dispersion

- Graphic methods
  - (e.g., frequency distribution)
  - Selection based on the type of variable to be presented



Numeric Methods: Mode

- A central value to represent the sample.

- Mode: Observation with the highest frequency (typical score).
  - Suitable for measurements with many repetitions (e.g., discrete data), even for nominal scale data.
  - Not influenced by extreme values.
  - More than one mode possible? (bimodal – multimodal)

- **EXAMPLE**
- 10 10 8 1 3 5 7 2 3 10 6 10

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mode

- 5 8 7 4 6 10 5 8 6 7 6 7 5 9 4 7 4 4 5 8 6 1 6 7 5 3 5 9 5 10 4 8 6 3 6 1 5 7 4 8 9 6 3 6 5 6 7 5 8 6

Values	Frequency (f)	Values	Frequency (f)
1	2	6	11
2	0	7	7
3	3	8	6
4	6	9	3
5	10	10	2

- Simple frequency count.
  - Average the 2 values in case of tie in frequencies.
  - Probability of a bi-modal / multi-modal distribution.
  - Probability of no mode.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Advantages / Disadvantages of Mode

Advantages	• **Shows the most frequent value in the distribution.** • **Unaffected by extreme values.** • **Can be calculated even with unknown outliers.** • **Provides more information about the distribution than the mean, when the distribution has a U shape.**
Disadvantages	• Does not take into account the exact value of each measurement. • Cannot be used to calculate population parameters. • Not useful with a small number of data points with similar frequencies (e.g., 1, 1, 2, 3, 3, 4). • Cannot be accurately calculated when dealing with grouped distributions (e.g., age groups).

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Numeric Methods: Median

1. **Median**
   The number in the middle of a series of a sample observations, after arranging them in ascending or descending order, is the $(n+1/2)^{th}$ observation. If the total number of values is even, then it is the average of the 2 middle observations.
   - Not influenced by extreme values and non-normally distributed data.
   - Suitable for ordinal/interval/continuous variables.
   - Sensitive to sampling fluctuations. Does not take into account all data points in the sample.

❖**EXAMPLE**

✓10 8 1 3 5 7 2 3 10 6 10

✓1 2 3 3 5 6 7 8 10 10 10

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Median

- 18 25 21 4 13 15 17 22 28

4	13	15	17	18	21	22	25	28
1	2	3	4	5	6	7	8	9

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Median

- 18 25 21 4 13 15 17 22

4	**13**	**15**	**17**	**18**	**21**	**22**	**25**
1	2	3	4	5	6	7	8

17 + 18 / 2 = 17.5

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Median

- 13 21 17 4 13 21 17 21 28

4	13	13	17	17	21	21	21	28
1	**2.5**	**2.5**	**4.5**	**4.5**	**7**	**7**	**7**	**9**
1	2	3	4	5	6	7	8	9

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Median

- 13 21 17 4 13 21 17 21 28 30

4	13	13	17	17	21	21	21	28	30
1	**2.5**	**2.5**	**4.5**	**4.5**	**7**	**7**	**7**	**9**	**10**
1	2	3	4	5	6	7	8	9	10

17 + 21 / 2 = 19

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Advantages/ Disadvantages for Median

Advantages	• **Usually easier to compute than the mean (not always).** • **Not affected by extreme values, better for skewed distribution.** • **Can be calculated even with unknown outlier values.**
Disadvantages	• Does not consider the precise value of each observation. • Cannot be used to calculate population parameters. • If the distribution has few observations, it may not represent them accurately (e.g. 2, 5, 8, 67, 110 → δ = 8).

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Numeric Methods: Mean

3. Mean.

$\bar{x} = \sum x_i / n$

- Affected by extreme values / non-normal distributions.
- For ordinal/interval/continuous variables.
- Utilizes all values in the sample.
- Mathematically useful, insensitive to sampling fluctuations.

❖ **EXAMPLE**

✓ 9 7 6 6 7 - 4 3 9 9 10

✓ $\bar{x} = 7$

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

Advantages/ Disadvantages for Mean

Advantages	• **Easy to calculate.** • **Represents more accuratelly the central value of the distribution than other indicators.** • **Can be used to calculate population parameters (parametric tests).**
Disadvantages	• Sensitive to the values in the distribution. • Since it is calculated algebraically, its value may not be part of the distribution's values. • Very sensitive to outliers.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Choosing Central Tendency Measurements

- Mean.
  - Symmetric distribution
  - Equal intervals / ratio scale
  - Parametric analyses
- Median
  - Asymmetric distribution – Outliers
  - Ordinal scale
  - Equal intervals / ratio scale
- Mode
  - Fast assessment
  - Nominal / ordinal scale

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Representativeness of the Mean

- The mean as a statistical model representing our data
- Representativeness evident from the homogeneity of the data, how much individual values differ from the mean $(x_i-\bar{x})$
- Model validity: What interests us is the dispersion (variability) of values around the mean $\sum(x_i-\bar{x})$
    - Hence, the need for measures of dispersion

- Variance: $\sum \sum (x_i-\bar{x})^2 / N-1$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Numerical Methods - Measures of Dispersion

1. **Variance [s^2] —— Standard deviation [s]**.
   Variance (Dispersion): The spread of values from their mean. Influenced by outliers.

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} \qquad s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

2. **Range**
   The difference between the largest and smallest observation.
   Influenced by outliers.

- **Interquartile range**
  The difference between the third quartile (Q3, representing the upper 25% of data) and the first quartile (Q1, representing the lower 25% of data).
  Not influenced by outliers.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Advantages / Disadvantages of Range

Advantages	• **Easy to calculate.** • **Includes the extreme values of the distribution.**
Disadvantages	• Affected by extreme values. • Does not provide information about the dispersion of values between the extreme values.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Advantages / Disadvantages of Interquartile Range

Advantages	• **Not affected by outliers.** • **Relatively easy to compute.** • **Representative of the central values of the distribution.**
Disadvantages	• Is not influenced by outliers. • Does not allow precise interpretation of a specific value in the distribution. • Inaccurate with data grouped with large intervals. • Does not describe population parameters needed in inductive statistics.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Advantages / Disadvantages of Standard Deviation

Advantages	• **Can be used to calculate population parameters needed in inductive statistics.** • **Takes into account all observations.** • **The most sensitive measure of dispersion.**
Disadvantages	• Relatively complicated calculation • Extremely sensitive to outliers

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Choosing Dispersion Measures

- Outliers?
  - Avoid standard deviation

- Ordinal scale
  - Interquartile range (and median)
  - Mode / range, complimentarily

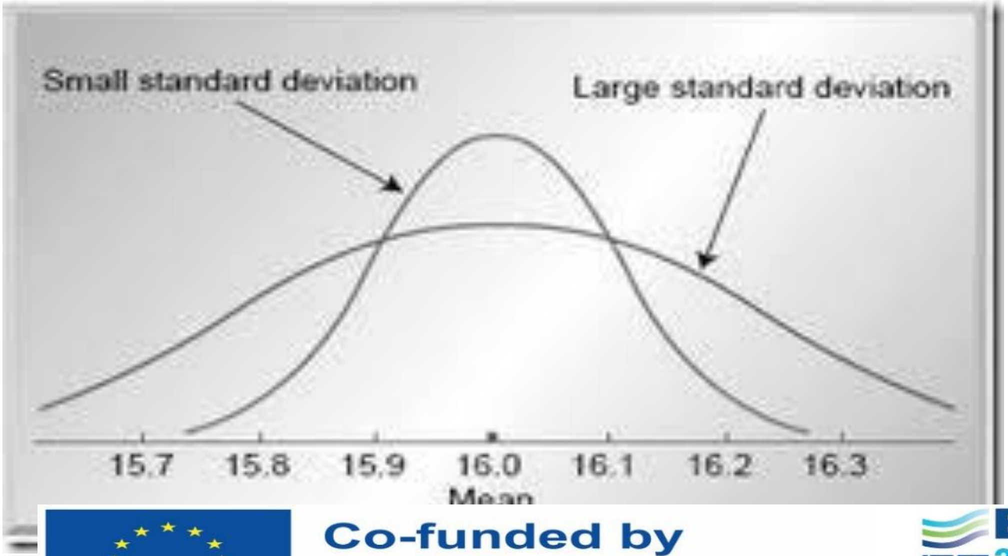
- Interval / ratio scale
  - Standard deviation (and mean)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Standard Deviation and the Shape of the Distribution

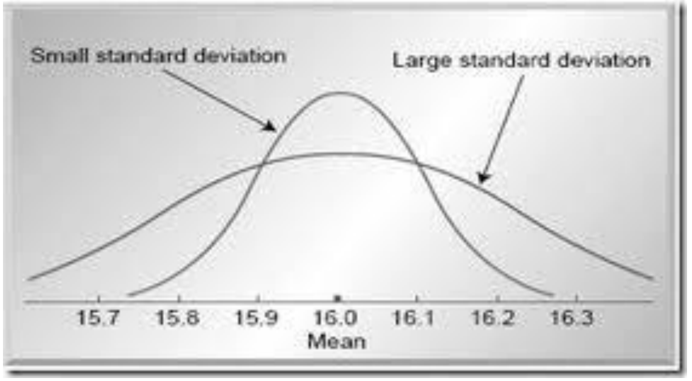
- Sample Variance and Standard Deviation indicate:
  - The accuracy of the mean as a model representing our data.
    - 9 7 6 6 7 || 4 3 9 9 10 → $\bar{x} = 7$
  - the shape of the distribution



Co-funded by the European Union

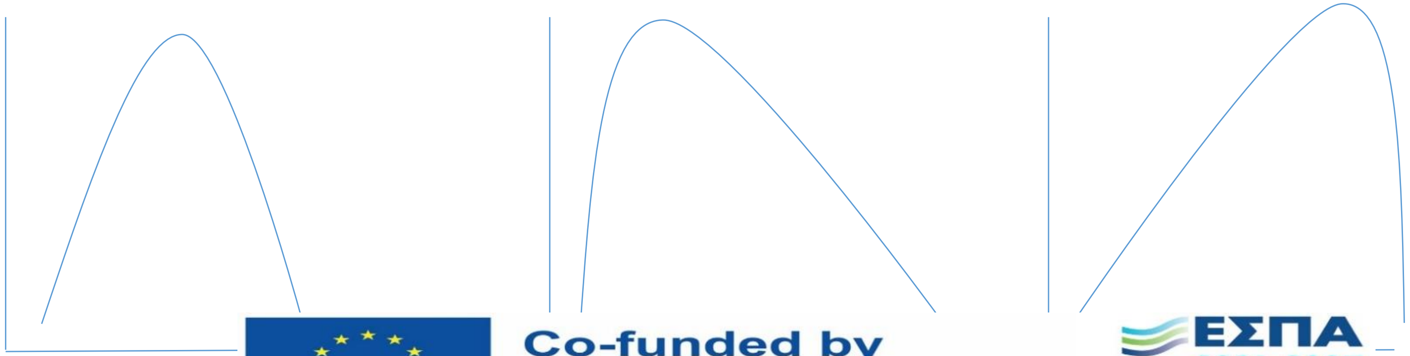
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Measures of Central Tendency & Types of Distributions



- Kurtosis – Broad / Narrow

- Normal Distribution (α)
  - As the sample size increases, it tends toward a normal distribution.
  - Mode, median, and mean coincide.
- Positive Skewness (β)
  - Mode → Median → Mean.
- Negative Skewness (γ):
  - Mean → Median → Mode.



α

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

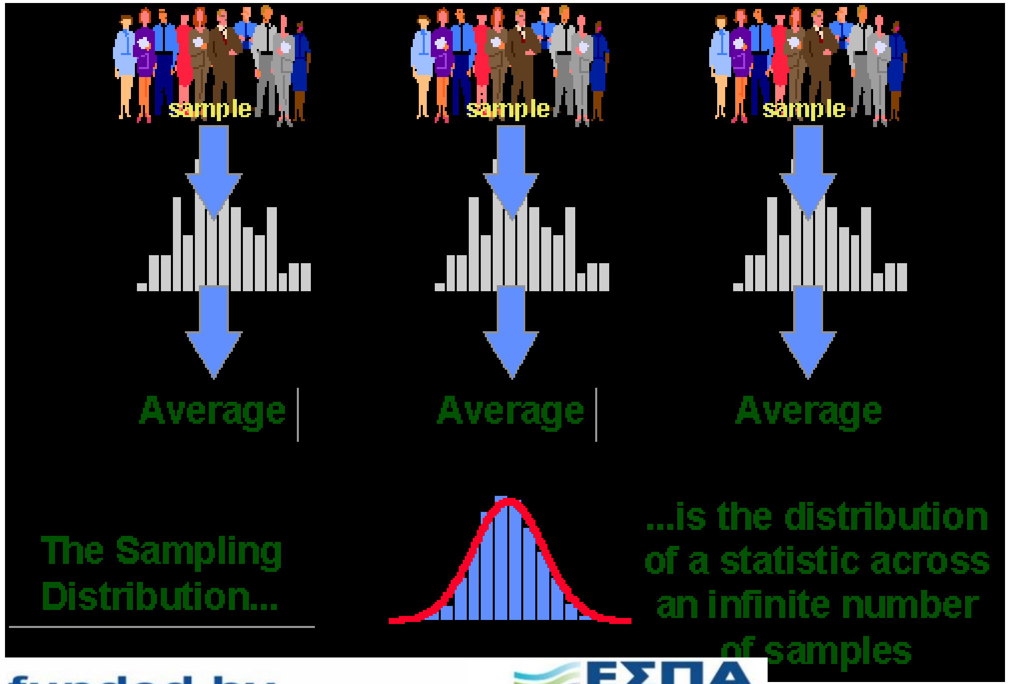
Sample Distribution – Standard Error

- How well does our sample represent the population of interest?
- The average among various samples from the same population will differ slightly.

How representative is the mean of our sample?

Standard error: The "standard deviation" among samples

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$



Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Confidence Intervals (C.I.)

- Common Confidence Level in Psychology: 95%
  - o How confident are we about the upper and lower bounds that the mean of any sample will fall within?
- In psychology, with a 95% confidence level:
  - o 95 out of 100 different samples we take will have their means between the upper and lower confidence limits.

- Top C.I. = M + 2 * SE
- Low C.I. = M – 2 * SE

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Descriptive Statistics: Graphical Methods

- Various graphical methods of representation
  - (Measurement scale of data is considered in the selection)

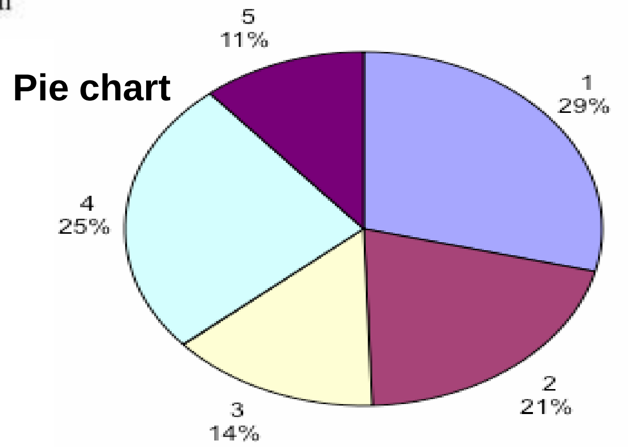
- Histogram
- Stem and leaf plot
- Box plot
- Bubble chart
- Scatter plot
- Run chart

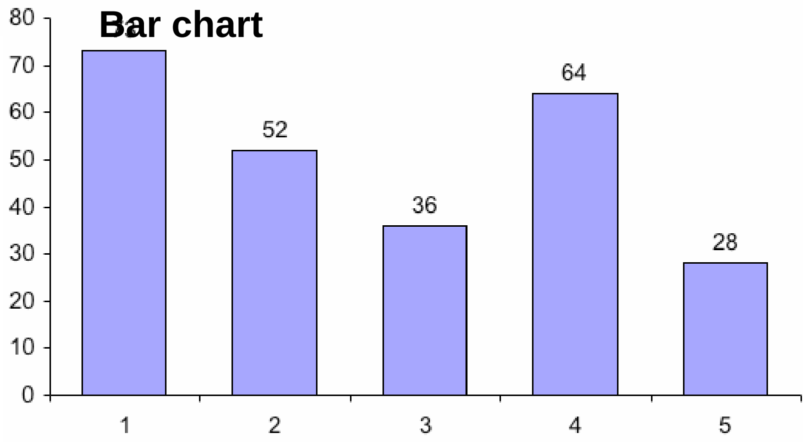
Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Frequency Distribution / Bar Chart: Nominal Data

Studies	Frequency	Relative Frequency
Accounting	73	28.9%
Finance	52	20.6
General management	36	14.2
Marketing/Sales	64	25.3
Other	28	11.1
Total	[illegible]	100

All three convey the same **information**
(based on the same data: frequencies).
Simply different **presentation**.





Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

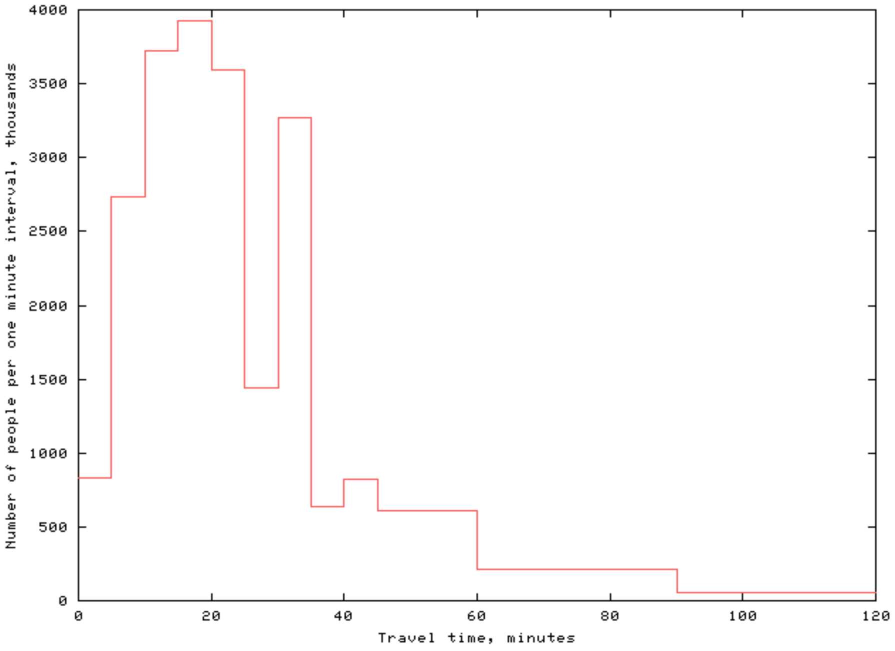
Quantitative Data: Frequency Distribution

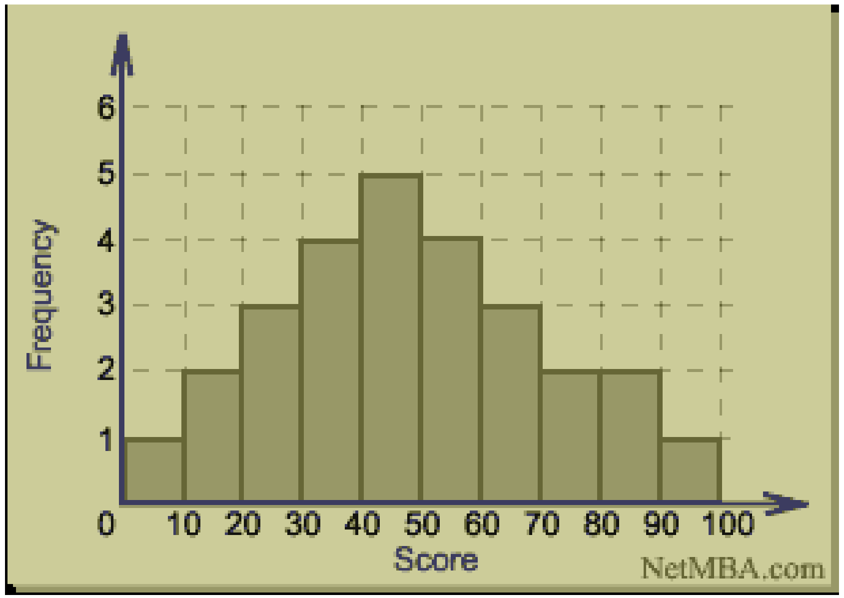
- Pros of grouping: Better management for extensive data, easier estimation
- Cons of grouping: Loss of information
- Decision on grouping range is subjective

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Histogram: Frewquency Distribution – Quantitative Data





- ✓ How wide should the columns (bins) be?
- ✓ Where should they start?
- ✓ Are the apparent structures real (especially double peaks - bimodality)?
    - Choices on these matters can dramatically impact the apparent interpreta

Stem and Leaf Plots

- Data representation in a type of historgam
  - 54 56 57 59 63 64 66 68 68 72 72 75 76 81 84 88 106
  - Tens on the stem and units on the leaf
- **Advantage**:
  - Impression of the general structure of the data
  - Includes all the data
  - Can be used for comparing two groups of data

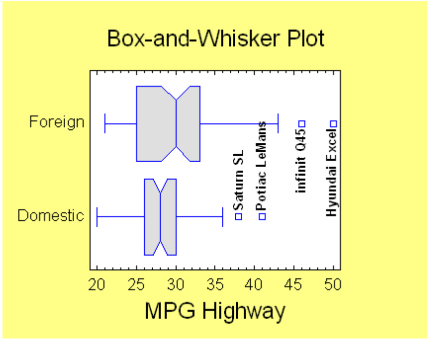
5	4 6 7 9
6	3 4 6 8 8
7	2 2 5 6
8	1 4 8
9	
10	6

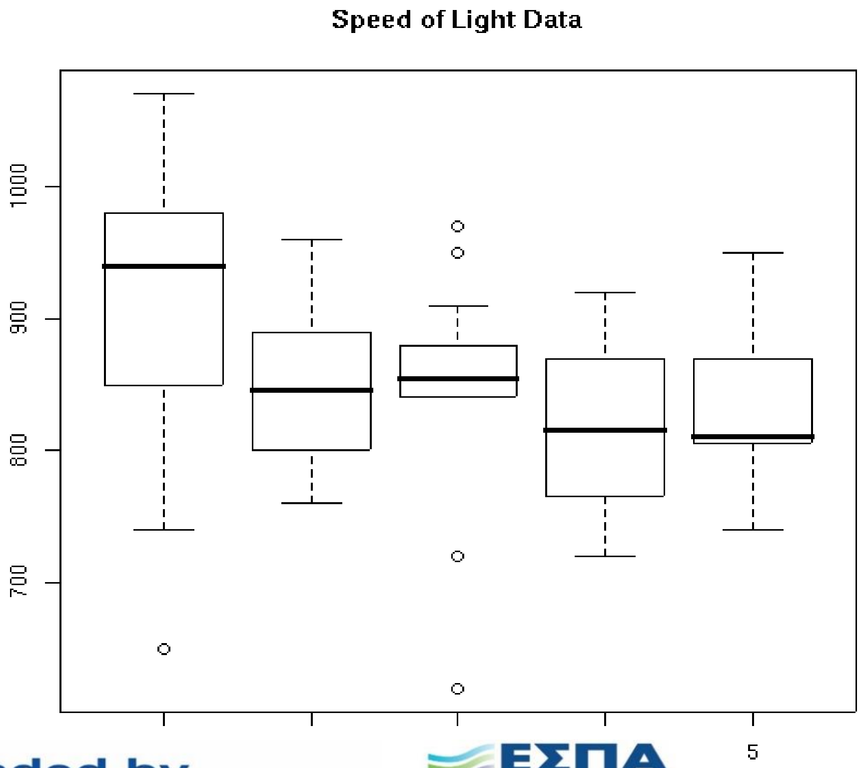
Set A		Set B
6 5 3	4	
8 7 6 5	5	4 6 7 9
7 3 2	6	3 4 6 8 8
4 2	7	2 2 5 6
6	8	1 4 8
	9	
	10	6

Box plot

- Reporting only the mean can hide important details about the data structure.

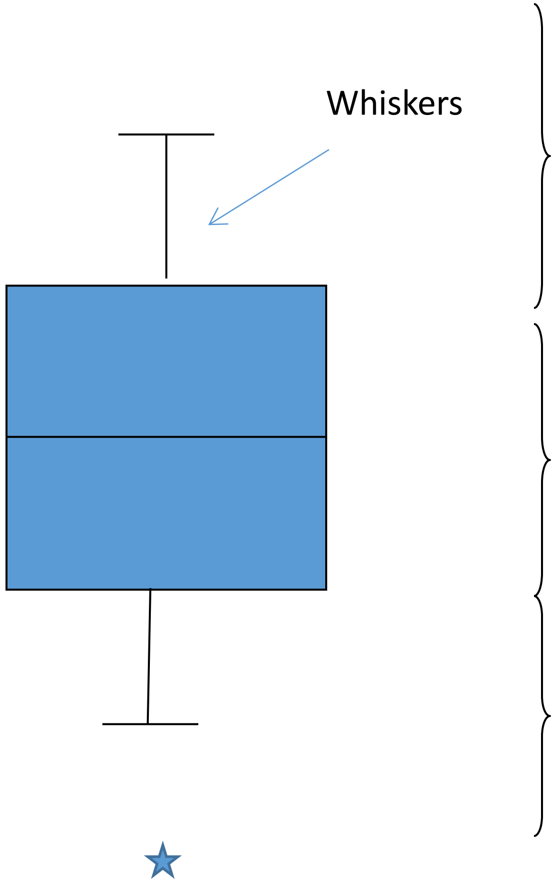
- Mean does not give the full picture regarding:
    - Variance
    - Symmetry
    - Floor – ceiling effects
    - Outliers





The 5 Number Summary

Boxplots

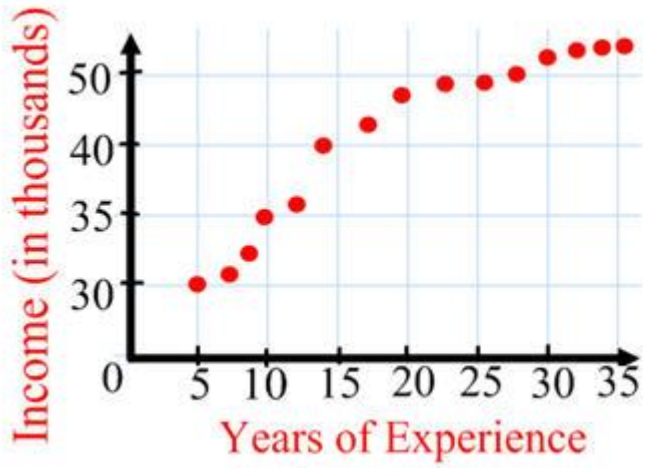


Interquartile range describes the spread

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Scatter plot

- Graph illustrating the relationship between two ratio variables. We are interested in understanding how the two variables correlate.
- To examine this relationship, we create a scatterplot, a diagram which plots one variable against the other.
- The independent variable, denoted as X, is usually placed on the horizontal axis, while the dependent variable, represented by Y, is positioned on the vertical axis.

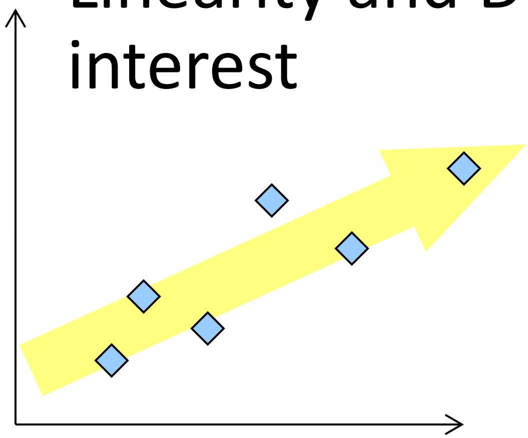


Co-funded by the European Union

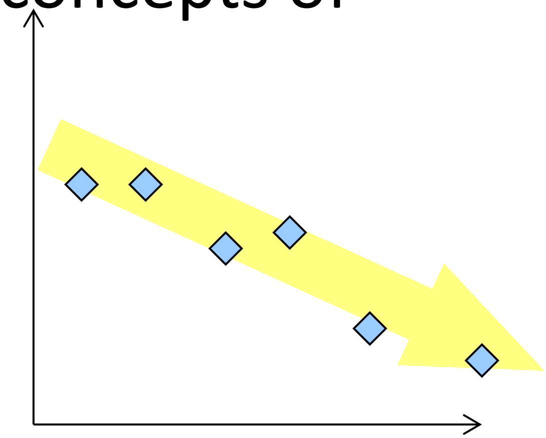
ΕΣΠΑ 2021-2027
Partnership Agreement 2021-2027

Scatter plot

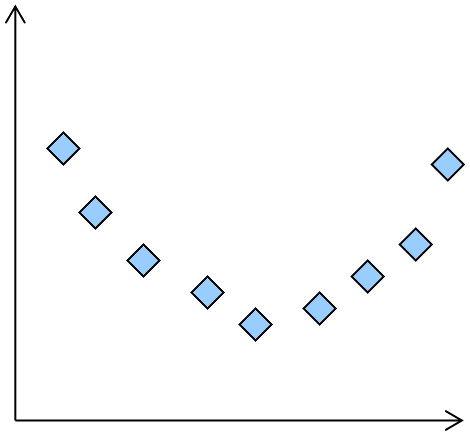
- Linearity and Directionality are two concepts of interest



Positive Linear Association



Negative Linear Association

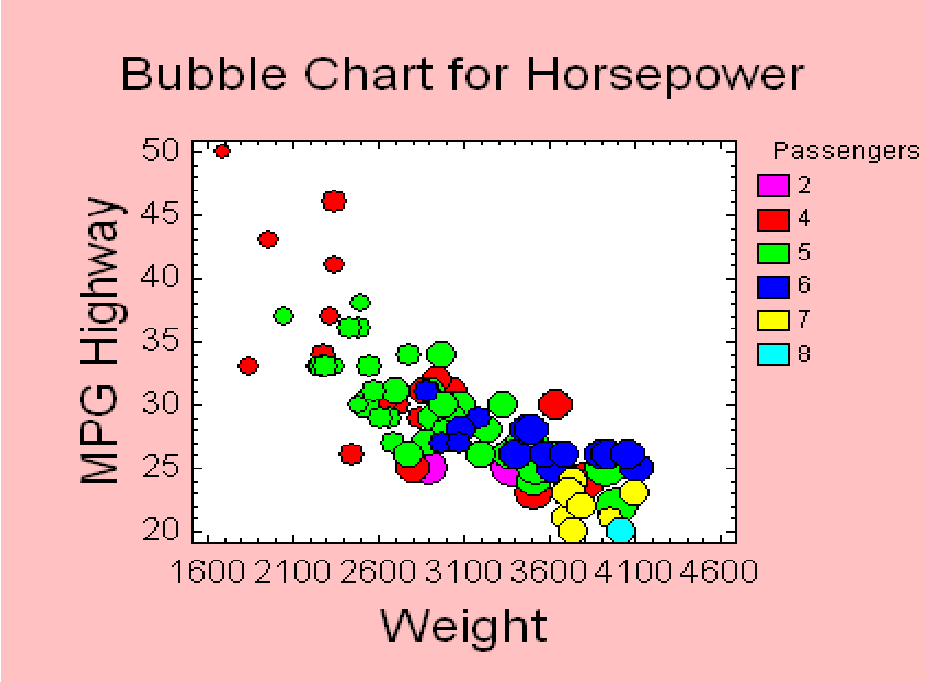


Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Bubble chart

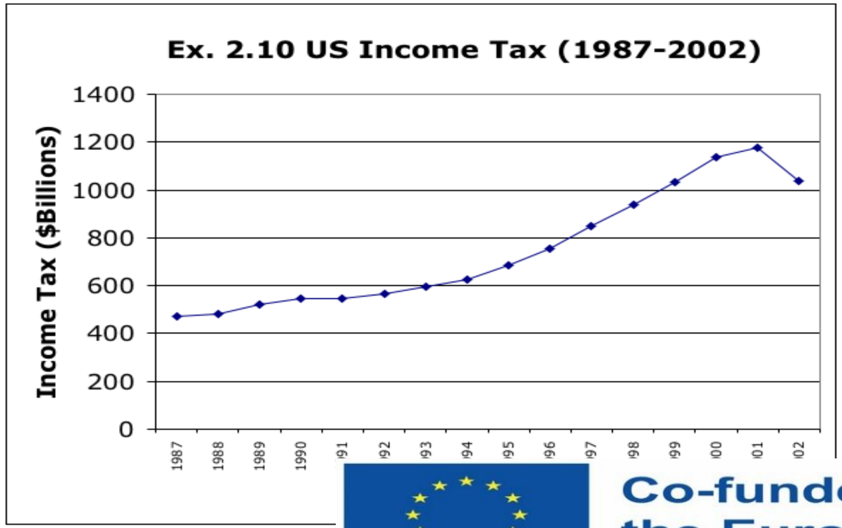
- The bubble chart can be used to represent four variables simultaneously:
  - X Axis
  - Y Axis
  - Bubble Size
  - Bubble Color



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Timeseries Data

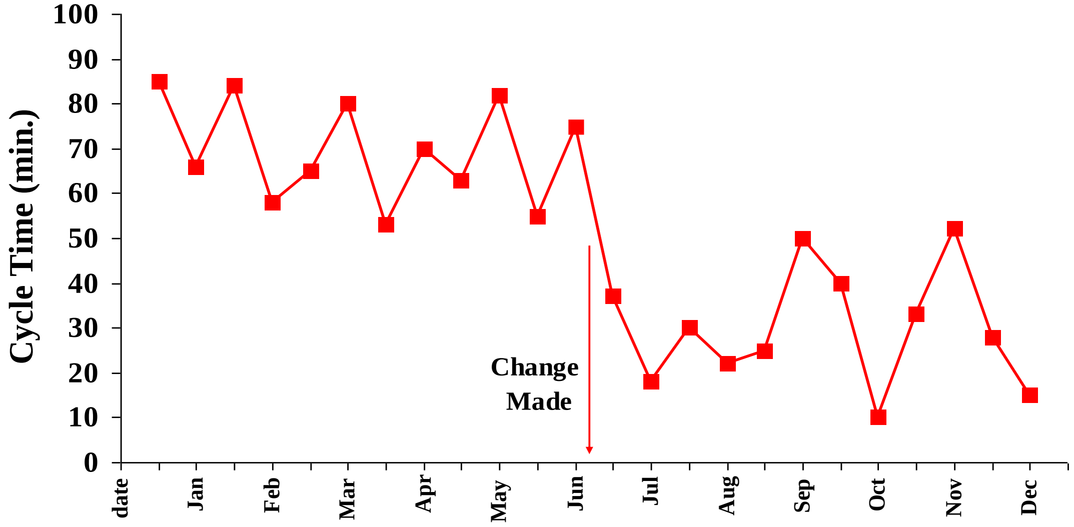
- Observations taken at the same moment in time are called cross-sectional data.
- Observations taken at successive time points are referred to as time-series data.
- Time-series data are presented using line charts, which plot the variable's value on the vertical axis in relation to time periods on the horizontal axis.



Visualization of total income tax in USA for the years 1987 to 2002...

Cycling Time Improvement

Run Chart – Graphical representation of a characteristic measured over time.



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Check List for Graphical Methods

- Does a graphical representation contribute to:
  - Maximizing the understanding of a set of data?
  - Revealing the underlying structure of the data?
  - Extracting significant variables?
  - Detecting outliers and anomalies?
  - Checking underlying assumptions?
  - Developing economical (parsimonious) models?

- If yes, include it!

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Elements of Good Representation

- Systematic organization for better understanding
- Semantic autonomy
- Accompanied by clear and supportive titles (graph, table, axes, etc.)
- Pay attention to scale of measurement

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Presenting Descriptive Statistics

- In writing (in text or in tables) or graphically
  - Reporting means (and standard deviations) for each experimental group / control group
  - Standard error / confidence intervals

- APA guidelines. Use 2 decimals
  - M = Mean
  - Mdn = Median
  - SE = Standard Error
  - SD = Standard Deviation

- EXAMPLE: The mean number of pints a tourist will drink before starting to walk around Faliraki naked is 12 pints (SD= 2.34)
- Women require a significantly smaller number of drinks (M=8.4, SD=2.53) compared to men (M=12, SD=1.22) before experiencing difficulties in walking.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Summary: Which Method to Use?

	Numerical Data	Nominal Data
Data from one Variable.	Histogram, Boxplot, or Stem and Leaf Plot	Frequency Tables and Relative Frequencies, Bar Charts, and Pie Charts
Association Between two Variables	Scatterplot	Association Tables, or Bar Charts

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS: Calculating Measures of Central Tendency & Dispersion

- Analyze → Descriptive Statistics → Frequencies
- Statistics
  - Choose Mean, Median, Mode

- Analyze → Descriptive Statistics → Explore
  - Dependent: Performance
  - Factor: Class
- Statistics
  - Choose B[...]

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS – Descriptive Methods

- Creating Frequency Tables
  - Analyze → Descriptives → Frequencies
- Graphs
  - In the same menu, choose "Charts"
  - Bar chart, Pie chart, Histogram
    - Chart values: "Frequencies" or "Percentages"
- Diagrams
  - Analyze → Descriptive Statistics → Explore
  - "Display" menu, option "Plots"
  - Stem and Leaf, Histogram, Boxplot
  - Line Chart: Graph → Legacy Dialogs → Line (Simple Line Chart)
- Boxplot
  - Graphs → Legacy Dialogs → Boxplot

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sources

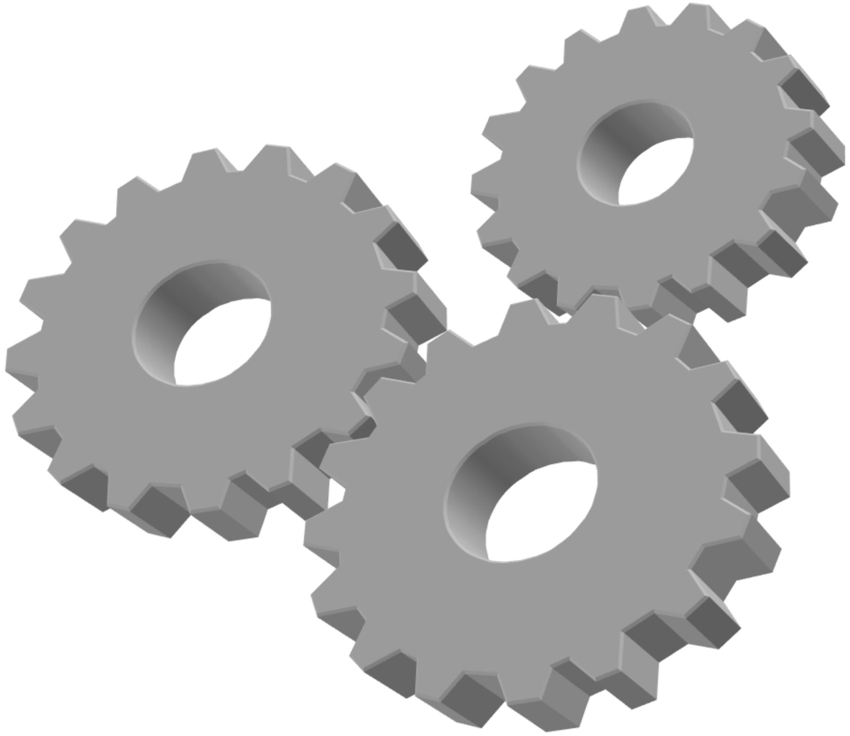
- Roland Baddeley, University of Bristol (2006)
- Brooks/Cole, Thomson Learning Inc. (2005)

- Tukey (1977) Exploratory data analysis. Addison-Wesley.
  - The book that started all about EDA.
- Hogan,D.C. Mosteller,F. and Tukey,J.W. (1983) Understanding robust and exploratory data analysis
  - Good in the cases where we need transformations.

- Web sites:
  - http://www.statgraphics.com/eda.htm
  - http://www.itl.nist.gov/div898/handbook/eda/eda.htm
  - http://en.wikipedia.org/wiki/Exploratory_data_analysis

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

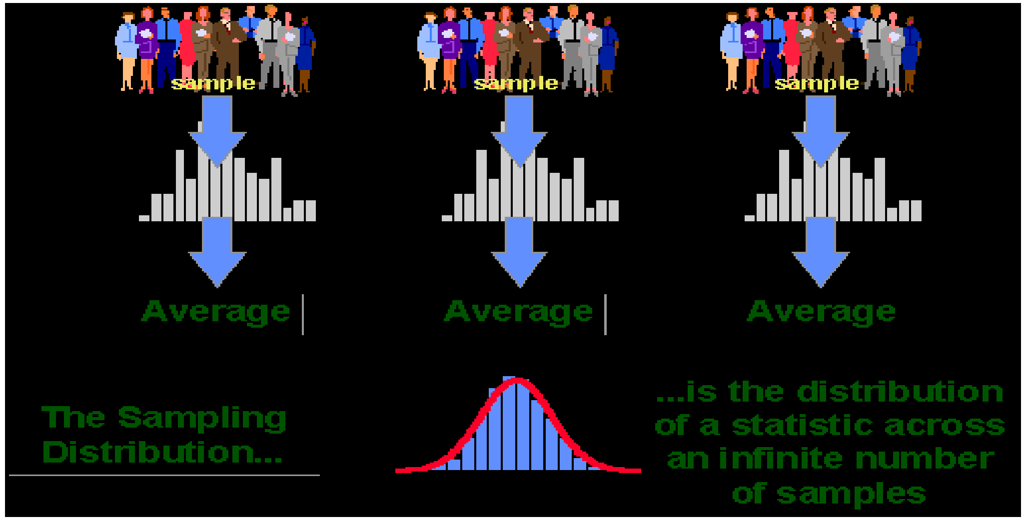
Inferential Statistics
Hypothesis testing
Correlation, T test

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Differences And the Central Limit Theorem

- **Sampling error**
  - Two experimental groups sampling the same population are expected to have similar means, but there will be little difference between them (due to individual differences)
- **Central Limit Theorem**
  - Averages of many different samples will probably be close to the population average



- Inferential Statistics helps us to estimate the **probability of** error in terms of the representation of a population by characteristics of a sample (e.g. mean, standard deviation, etc.)
- More generally, it defines the probability of error contained in the

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

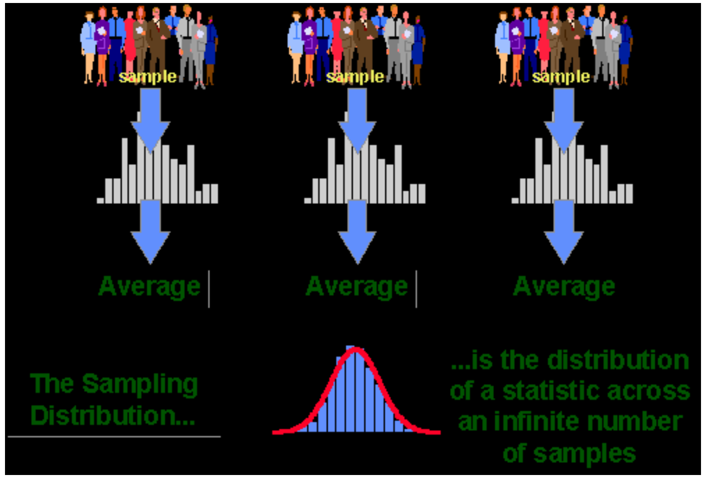
Sample classification - Typical error

- The sampling distribution is a normal distribution of different samples, so we can make predictions based on this distribution
  - For example, of the infinite possible samples, $\bar{X}$ has a 68.26% chance of being 1 standard deviation from m.
  - 95% → Z = +-1.96. 99% → Z =+-2.58. 99.73% → Z = +-3
    - Z values are a way of expressing a measurement in terms of how many standard deviations it is from the mean

- E.g. $\bar{X}$ = 75, s = 18, N = 36
  - $\sigma_{\bar{X}}$ = 18 / √36 = 3 (standard error)
  - $CI_{95\%}$ → Z = 75 +-1.96 * 3 → 69.12 to 80.88
  - $CI_{99\%}$ → Z = 75 +- 2.58 * 3 → 67.26 to 82.74

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Standard error (σ): the standard deviation of the samples



Co-funded by the European Union

Inferential Statistics & Hypothesis Formulation

- The aim is to describe a phenomenon, to draw conclusions about a function
  - Hypothesis testing to investigate scientific questions

- **Experimental hypothesis (H1)**: hypothesis that the experimenter's manipulation will have some effect
- **Null hypothesis (H0)**: the experimental manipulation will have no effect - no difference between two conditions
- E.g. Drinking alcohol increases the likelihood of falling
  H1: Alcohol users will fall more often than non-users
  H0: Falls are as frequent regardless of alcohol consumption

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Inferential Statistics

- Inferential statistics tests the probability that our experimental prediction is right or wrong
  - Assessing whether the data we have collected is the result of randomness
  - Reducing the probability of random effects increases the likelihood that they are due to our experimental manipulation
    - For example, the greater the difference in the means of 2 samples, the more likely they represent different populations

- **Rejection of H0 – Acceptance of H1**
  - We accept our findings as true only when there is 95% confidence that our result is not a product of chance
  - If the probability of it being random does not exceed 5%, then we can accept our findings
- 95% ($p < .05$) has simply become accepted as a norm in psychology because of Ronald Fisher's (1890-1962) proposal

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Increasing Confidence in Research Findings

- Experimental manipulation may lead to differentiation of the mean of 1 sample, with two possible explanations:
  - Experimental manipulation changed what we measure (two different populations)
  - Samples come from the same population, but we selected individuals with opposite properties from within that population (outliers from the normal sample distribution), so experimental manipulation did not make a difference
- The greater the difference between the means of 2 samples, the more likely they represent different populations

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Hypothesis Testing

- Calculating the probability that two samples come from the same population
  - High: No effect of experimental manipulation (H0)
  - Low: Effect of experimental manipulation (H1)
  - Exactly how much probability is low?
    Fisher says 5%!
- How exactly do we calculate this probability?
  - Depends on the experimental design and method of data analysis
- However, we can review some general principles...

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Calculating the probability that the samples come from the same population

- Sources of volatility:
  - Systematic differences due to experimental manipulation
  - Non-systematic differences due to individual differences between samples (e.g. intelligence, motivation)

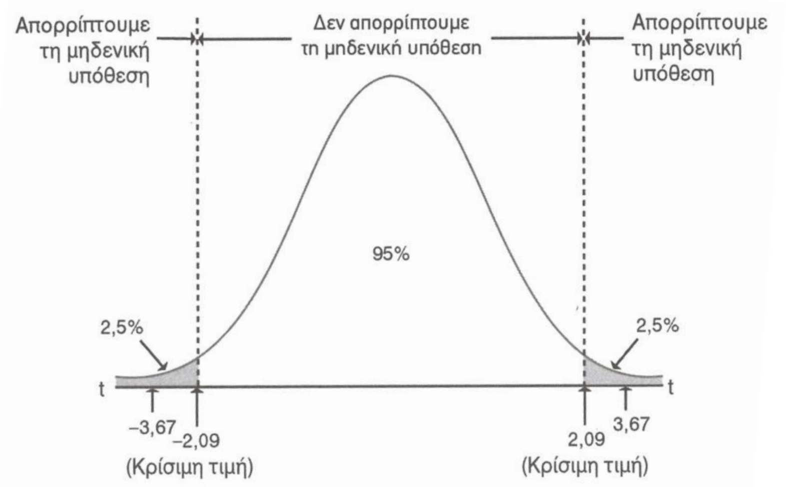
- We test the difference between two samples using statistical tests whose characteristics we know
  - We know the distribution, and therefore we can calculate the probability of getting any value in that distribution - example age of death!
- For a comparison of two different sample means, statistical tests represent the following fraction: Systematic Differentiation / Non-Systematic Differentiation
  - Comparison of variability resulting from experimental manipulation with naturally occuring random variability

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Hypothesis Testing procedure

1. Formulation of hypotheses
   - H0, H1
2. Selection of an appropriate statistical criterion
   - Type of variables
   - Measurement scale
   - E.g. parametric - non-parametric
3. Definition of significance level (a)
   - Usually 5% or 1%
4. Comparison with critical value
   - Rejection/acceptance area
   - Equal or greater = Statistical significance



Σχήμα 7.2. Εξετάζοντας τη μηδενική υπόθεση (αμφίπλευρος έλεγχος).

Two-tailed (two-sided)
VS.
One-tailed hypothesis

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

What does the probabilistic nature of statistical tests mean?

- Two objective possibilities:
  - Our experimental manipulation had an effect on the 2 samples
  - Our experimental manipulation had absolutely no effect
- Statistical tests tell us which of the two alternatives is more **likely**.
- We want to be as confident as possible, so we use a strict Fisher **significance level** of α=5%

❖ Because the process is probabilistic, the conclusion may not hold, even if we have 95% confidence that the difference is not random

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

7.

Box 5.2: What we can and can't conclude from a significant test statistic

1. **The importance of an effect**: We've seen already that the basic idea behind hypothesis testing involves us generating an experimental hypothesis (the means of our experimental conditions will differ) and a null hypothesis (the means of our experimental conditions will be the same). If the probability of obtaining the value of our test statistic by chance is less than .05 then we generally accept the experimental hypothesis as true: our means are indeed different. Normally we say 'our experiment had a *significant* effect'. However, don't be fooled by that word 'significant', because even if the probability of our effect being a chance result is small (less than .05) it doesn't necessarily follow that the effect is important. Very small and unimportant effects can turn out to be statistically significant just because huge numbers of people have been used in the experiment (see page 152).
2. **Non-significant results**: Once you've calculated your test statistic, you calculate the probability of that test statistic occurring by chance; if this probability were greater than .05 you reject your experimental hypothesis. However, this does *not* mean that the null hypothesis is true. Remember that the null hypothesis is that the means in different groups are identical, and all that a non-significant result tells us is that the means are not different enough to be anything other than a chance finding. It doesn't tell us that the means are the same; as Cohen (1990) points out, a non-significant result should never be interpreted (despite the fact it often is) as 'no difference between means' or 'no relationship between variables'. Cohen also points out that the null hypothesis is *never* true because we know from sampling distributions (see page 132) that two random samples will have slightly different means, and even though these differences can be very small (e.g. one mean might be 10 and another might be 10.00000000000000000001) they are nevertheless different. In fact, even such a small difference would be deemed as statistically significant if a big enough sample were used (see page 152). So, significance testing can never tell us that the null hypothesis is true, because it never is!
3. **Significant results**: OK, we may not be able to accept the null hypothesis as being true, but we can at least conclude that it is false when our results are significant, right? Wrong! A significant test statistic is based on probabilistic reasoning, which severely limits what we can conclude. Again, Cohen (1994), who was an incredibly lucid writer on statistics, points out that formal reasoning relies on an initial statement of fact followed by a statement about the current state of affairs, and an inferred conclusion. This syllogism illustrates what I mean:

➢ If a man has no legs then h
  ○ This man plays football.
  ○ Therefore, this man has

The syllogism starts with a statement of fact that allows the end conclusion to be reached because you can deny the man has no legs (the antecedent) by denying that he can't play football (the consequent). A comparable statement of the null hypothesis would be:

➢ If the null hypothesis is correct, then this test statistic can not occur.
  ○ This test statistic has occurred.
  ○ Therefore, the null hypothesis is false.

This is all very nice except that the null hypothesis is not represented in this way because it is based on probabilities. Instead it should be stated as follows:

➢ If the null hypothesis is correct, then this test statistic is highly unlikely.
  ○ This test statistic has occurred.
  ○ Therefore, the null hypothesis is highly unlikely.

If we go back to a football example we could get a similar statement:

➢ If a man plays football then he probably doesn't play for England (this is true because there are thousands of people who play football and only a handful make it to the dizzy heights of the England squad!).
  ○ Phil Neville plays for England.
  ○ Therefore, Phil Neville probably doesn't play football.

Now although at first glance this seems perfectly logical (Phil Neville certainly doesn't play football in the conventional sense of the term) it is actually completely ridiculous – the conclusion is wrong because Phil Neville is a professional footballer (despite his best attempts to prove otherwise!). This illustrates a common fallacy in hypothesis testing. In fact hypothesis testing allows us to say very little about the null hypothesis.

would occur when, in the real world our experimental manij ulation has no effect, yet we have got a large test statist because we coincidentally selected two very dissimilar sample So, the sample means were very different but not because of th experimental manipulation. If we use Fisher's criterion then th probability of this error is .05 (or 5%) when the experiment ha no effect – this value is known as the α-level. Assuming ou experiment has no effect, if we replicated the experiment 10 times we could expect that on five occasions the sample mean would be different enough to make us believe that the exper ment had been successful, even though it hadn't.

➢ *Type II error*: This is when we believe that our experimenta reality it hasn't. This woul e experimental manipulatio a small test statistic (perhap

Two types of errors

- **Type I error**: we accept H1, i.e. that our experimental manipulation caused the differentiation of the 2 samples, when in fact it is not true.
  - E.g. We randomly selected two very different samples, so they also differed in the mean, regardless of our experimental manipulation
  - This result is one of the 5 times we could be misled if we repeated the experiment 100 times (α=.05)
- **Type II error**: we reject H1 (and thus we accept H0), i.e. we assume that our experimental manipulation had no effect, when in fact it did.
  - E.g. Our intervention worked but the statistical test gave us a low value, perhaps because there is a high degree of natural variability in the 2 samples (c.f. the equation mentioned earlier)
  - Ideally this error should have a small probability of occurring ($1-\beta = .2$)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Effect Size

- Measuring the magnitude of the difference in our measurement can add complementary information to the likelihood of this difference being random
  - Such effects may be: experimental manipulation, magnitude of association between variables
- Objective and standardized measurement of the magnitude of the observed phenomenon
- It can be compared between different studies, measurement units and so on.
- *Pearson correlation coefficient (r)*
  - $r = 0.10$ - small
  - $r = 0.30$ - moderate
  - $r = 0.50$ - large
- r^2 = The total percentage of variability that can be explained by the experiment

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Statistical Power (Statistical Power)

- The effect size is affected by:
  1. Sample size
  2. Level of significance (a)
  3. Power of the test to discover a phenomenon of this magnitude
- If we know the first two, we can also calculate the third

- 1 - b (type II error) : The probability of finding an effect that exists

1. After the experiment is done we know: 1) significance level (e.g. α = .05), 2) effect size of our sample, 3) sample size
   → we can calculate β (the power of our test), we want it to be above 0.8
2. Calculating the sample size in advance is necessary to discover a certain effect size:
   a and *b* are known, effect size estimation from previous surveys
   → from these we can deduce the number of participants in each sample

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sample size calculation

- Complex statistical calculations, but nowadays there are corresponding programs (e.g. nQuery Adviser)
- Cohen tables

- Guidelines:
- $\alpha = .05$, $\beta = .8$
  - $r = .1 \rightarrow 783$
  - $r = .3 \rightarrow 85$
  - $r = .5 \rightarrow 28$

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Confidence Intervals (C.I.)

- Confidence level commonly used in Psychology: 95% (99%, 90%...)
  - How confident are we about the upper and lower limits in which the average of any sample will be included?
  - 95 of the 100 different samples we take will fall between the upper and lower limits of the confidence levels

- **Confidence interval**: The range of values within which the actual value of the parameter of interest lies (e.g. mean)
- Top C.I. = X + 2 * SE
- Low C.I. = X - 2 * SE

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric VS Non-Parametric Analyses: when are they used?

- Important decision. Wrong choice can lead to:
  - Using the wrong statistical analysis (breach of conditions)
  - Using statistical analysis of lower power (loss of statistical power)
- Parametric analyses use information on the mean and variance.
- Non-parametric statistics do not make assumptions about the distribution of data, so they have less statistical power because they use less information
  - E.g. Parametric correlation will use information on the mean and variance, while non-parametric correlation only takes into account the serial position of data pairs (scores).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric Distribution assumptions

- Parametric Distribution Assumptions
  - Measurement independence
  - Measurements come from a population that follows a normal distribution
  - Populations (e.g. in the comparison of 2 groups or 2 experimental groups) have the same variance (variance homogeneity hypothesis)

- Non-Parametric Distribution Hypotheses
  - Measurement independence
  - The variable of interest has some continuity (can be ranked)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Points of differentiation

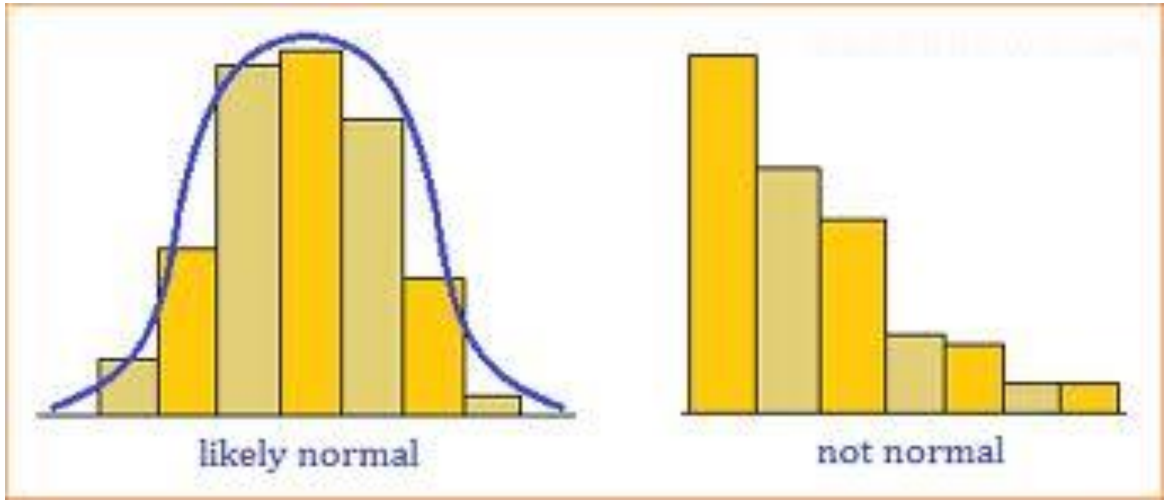
- Variable characteristics:
  - If our measurements are on a nominal/ordinal scale, then we use non-parametric statistical analysis
  - If our measurements are on an equal intervals/ratio scale, then we use parametric statistical analysis
- A normal distribution in the data is necessary / prerequisite for parametric analysis, as mentioned above
- Contrary, the non-parametric distribution makes no assumptions about the distribution
  - Ranking of variables and analysis of the position of each measurement in the hierarchical ranking

- As the distribution moves away from normality, the risk of inaccuracy of the parametric analysis increases, so it is preferable to use a non-parametric alternative
  - Small samples → non-parametric tests

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Assessing Normality of Distribution



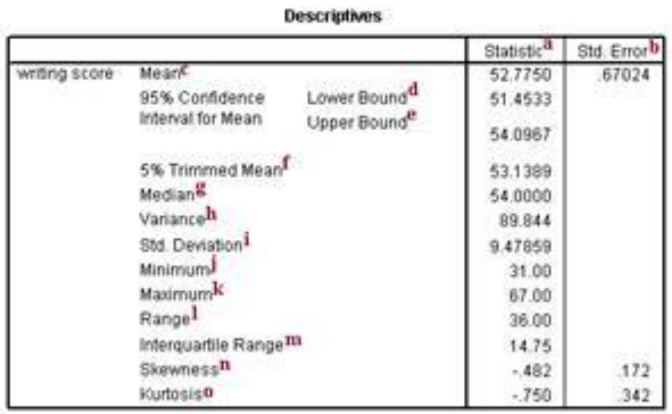
Visual check by histograms!

Disadvantages:

- ✓ Sample distribution may not be representative of the population
- ✓ You can't get a clear picture of the distribution for N<30
- ✓ Subjectivity! How much should it differ from the normal distribution?

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

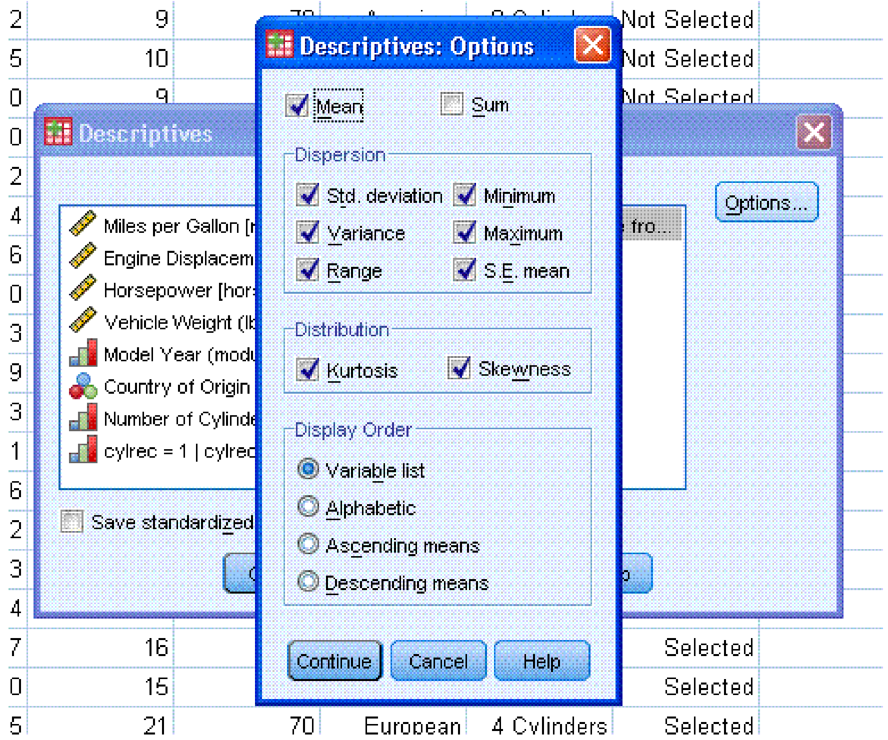
Assessing Normality of Distribution



Descriptives

			Statistic[a]	Std. Error[b]
writing score	Mean[c]		52.7750	.67024
	95% Confidence Interval for Mean	Lower Bound[d]	51.4533	
		Upper Bound[e]	54.0967	
	5% Trimmed Mean[f]		53.1389	
	Median[g]		54.0000	
	Variance[h]		89.844	
	Std. Deviation[i]		9.47859	
	Minimum[j]		31.00	
	Maximum[k]		67.00	
	Range[l]		36.00	
	Interquartile Range[m]		14.75	
	Skewness[n]		-.482	.172
	Kurtosis[o]		-.750	.342

- Easy method of assessing the distribution in the SPSS frequency table: Symmetry (skewness ~= 1), Curvature (kurtosis ~= 1) → distribution width
- Can be tested with statistical tests (e.g. Kolmogorov-Smirnov, Shapiro-Wilks)
- **Warning:** statistical power for these testers also depends on sample size



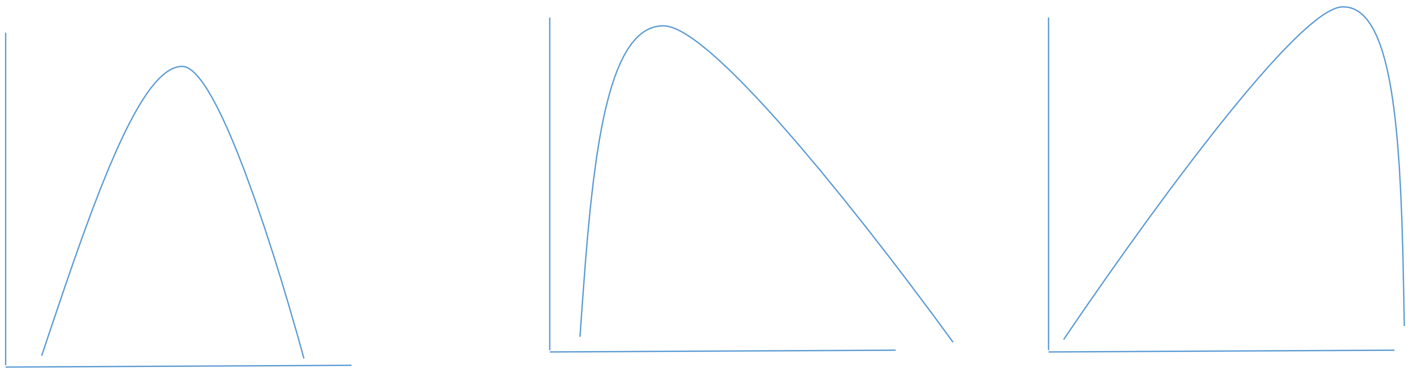
Andy Field: Discovering Statistics Using SPSS

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Partnership Agreement 2021-2027

Types of Distributions

- Normal distribution
- Positive Assymetry
- Negative Assymetry

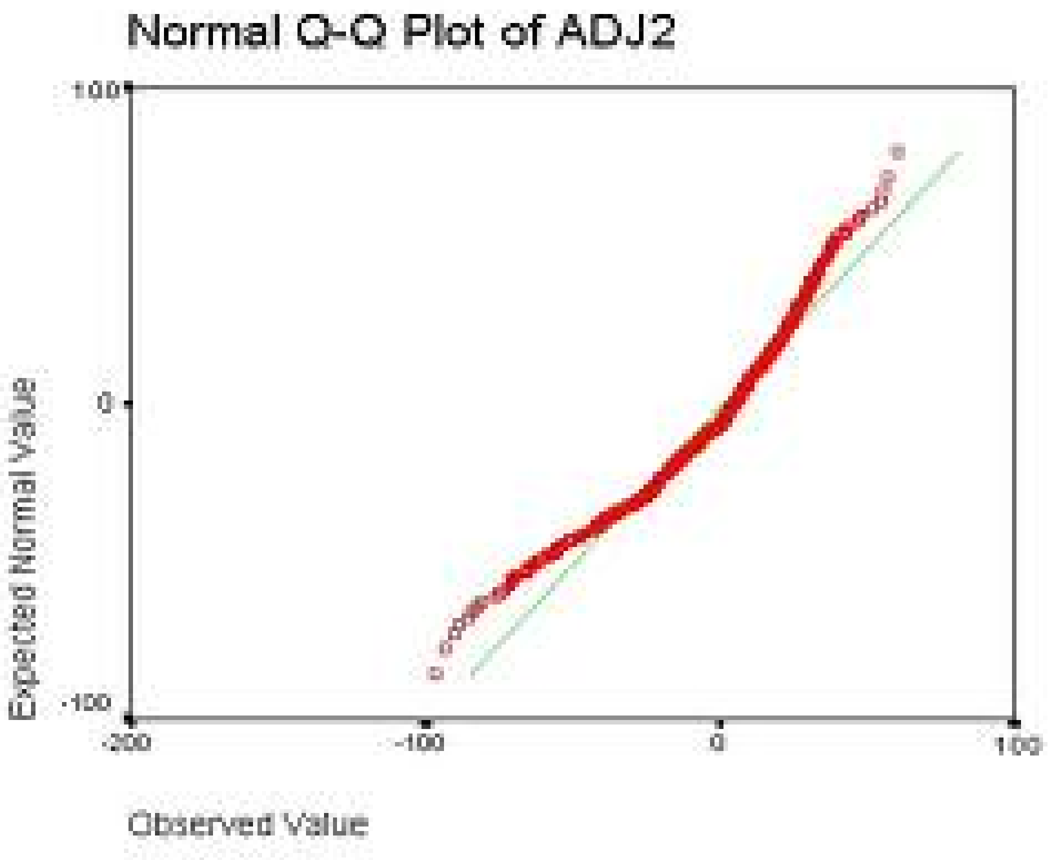
- Distribution curvature (kurtosis) - Wide / Thin



Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Another Way: Q-Q Plot



Expected normal distribution values vs. observed values in our sample

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Choosing Between Parametric or Non-Parametric Analysis

- Parametric when we are sure that **the population** where our sample comes from follows a normal distribution, with respect to the variable of interest. Non-parametric when:
  - Measurement is coded as a hierarchical rank, e.g. rank of student performance in a class, star rating of a movie
  - There are values off the measurement scale, e.g. very low/high. In non-parametric analysis we replace them with random low/high values, since it is not the exact measurement that matters but the hierarchical ranking
  - Population does not have a normal distribution and there is no possibility of transformation to a normal distribution (e.g. logarithmic)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Difficulties in Choosing Between Parametric or Non-Parametric Analysis

- When there are few data, it is more difficult to decide (e.g. by visual inspection) whether they follow a normal distribution. Formal normality tests (e.g. Kolmogorov-Smirnoff test) may have reduced power to detect normal and non-normal distributions.
  - Often we choose non-parametric methods in cases of small samples ($n < 6$). Choice does not affect the result so much if the sample is large enough (but subjective just how large is enough!)
- It is useful to check previous data (from other surveys). The important thing is the distribution of the population, not the sample.
- Subjectivity about the most appropriate choice in case of doubt.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Types of Statistical tests

- There is at least 1 corresponding non-parametric test for each parametric test

- Categories:
  1. Tests comparing groups (independent samples)
  2. Tests comparing variables (dependent samples)
  3. Tests of relationship between variables (correlation)

- **Note**: Despite the advantages of non-parametric analysis (such as cases where the distribution in the population is unknown), there are also disadvantages:
  - Lower statistical power (1-b). e.g. we would need a larger sample to achieve the same power with a parametric analysis
  - Difficulties of interpretation (e.g. what is the meaning of a difference between ranks versus a difference in the unit of meas[illegible]

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Choice of Statistical Test

	Type of Data	
Goal	**Measurement (from Gaussian Population)**	**Rank, Score, or Measurement (from Non- Gaussian Population)**
Describe one group	Mean, SD	Median, interquartile range
Compare one group to a hypothetical value	one-sample *t* test	Wilcoxon test
Compare two unpaired groups	Unpaired *t* test	Mann-Whitney test
Compare two paired groups	Paired *t* test	Wilcoxon test
Compare three or more unmatched groups	one-way ANOVA	Kruskal-Wallis test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test
Quantify association between two variables	Pearson correlation	Spearman correlation
Predict value from another measured variable	Simple linear regression or nonlinear regression	Nonparametric regression**
Predict value from several measured or binomial variables	Multiple linear regression* or	

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

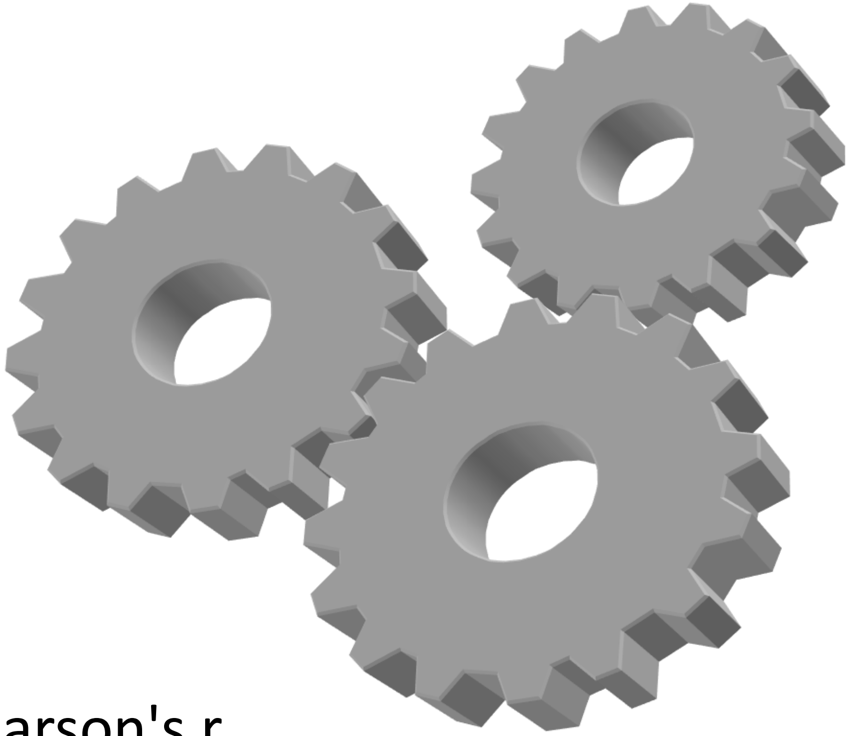
Testing Normality of Distribution in SPSS

- Frequency tables:
  - Analyze → Descriptive Statistics → Frequencies

- Histograms, Boxplots, QQ-plots, PP-plots
  - Analyze → Descriptive Statistics → Explore (plots)
  - Analyze → Descriptive Statistics → Q-Q Plots

- Goodness of fit checks
  - Kolmogorov-Smirnov (K-S)
    - Analyze → Nonparametric Tests → Legacy Dialogues → 1-Sample K-S
    - We want $p > .05$ (not significant)
  - Shapiro-Wilk (S-W)
    - We want a value close to 1

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Correlation

Pearson's r

Spearman's rho (ρ)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

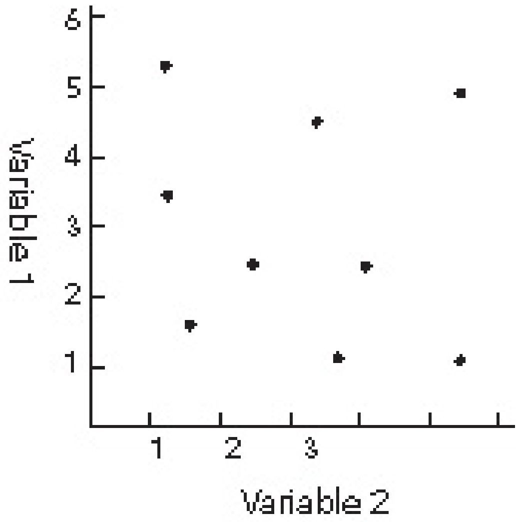
Correlation

- Objective: The association between variables. What happens to 1 or more variables when a 2nd variable changes?
- Correlation coefficient
  - Measuring the degree of association between 2 variables
  - -1 to +1
- Positive correlation
  - As one variable increases, so does the other
  - E.g. The more years of education, the higher the annual salary
- Negative correlation
  - As one variable increases, the other decreases
  - E.g. As the running speed increases, the endurance reserves decrease

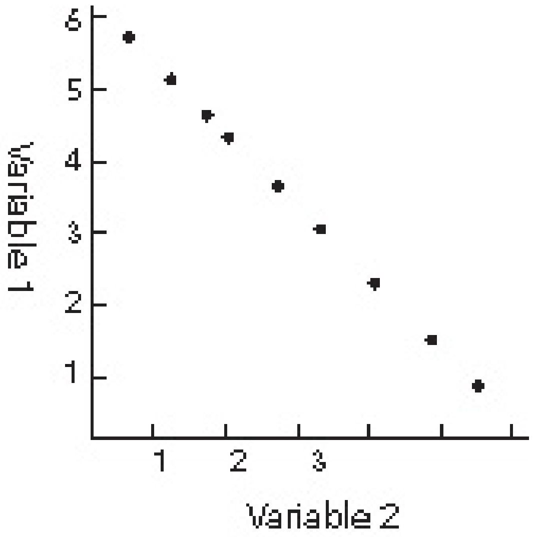
➢ The higher the correlation coefficient, the stronger the variable between the 2 variables (regardless of the sign)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

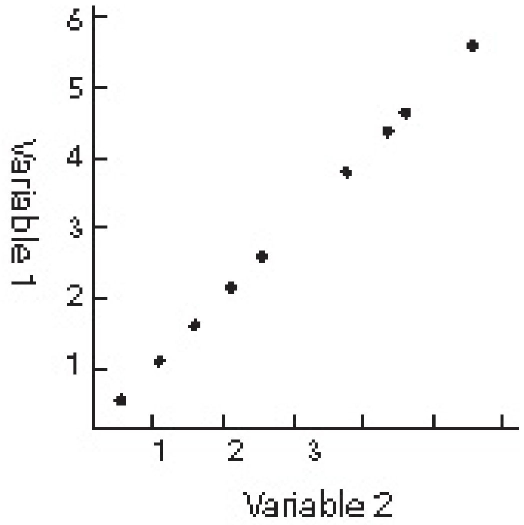
Type of Association



No correlation
Correlation coefficient = 0



Perfect negative correlation
Correlation coefficient = –1

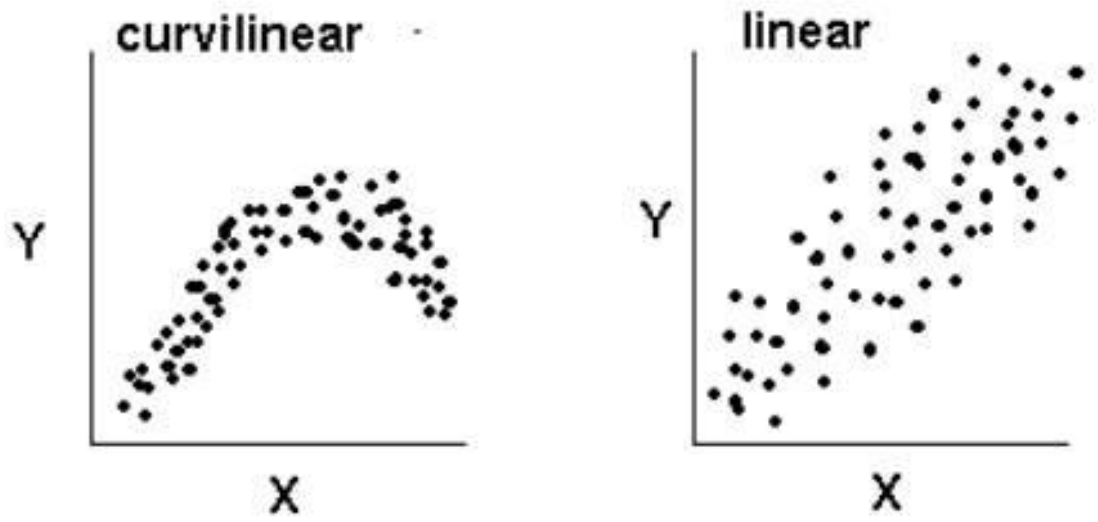


Perfect positive correlation
Correlation coefficient = +1

- Illustrate correlation with scatterplot
- Notional or real regression line plotting

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Curves Correlations



- Positive curvilinear relationship
- Negative curvilinear relationship

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Correlation

- <u>Correlation coefficient (r)</u>
  - Measuring the degree of association between 2 variables

➢The higher the correlation coefficient, the stronger the association between the 2 variables (regardless of the sign)

➢The sign gives information on the direction of the correlation, while the absolute numerical value (0-1) gives the strength of the correlation (distance of points from the regression line)

  - r > .30 → Low correlation
  - r > .50 → Moderate correlation
  - r > .70 → High correlation
  - r = .80 → Very high correlation

Co-funded by the European Union

ΕΣΠΑ 2021-2027 Βιώσιμη Ανάπτυξη για Όλους Partnership Agreement 2021-2027

Correlation and Covariance

- **Covariance**: covariation of the 2 variables
- $COV_{XY} = [\Sigma\ (X - \bar{X}) \times (Y - \bar{Y})] / N$
- $r = COV_{XY} / s_X \times s_Y$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factors Affecting Correlation & Interpretation

- Factors affecting the correlation coefficient
  - Group homogeneity
  - Curvilinear relationship between variables
  - Outliers

- Correlation interpretation
  - Co-variation, <u>not causality</u>
  - Causality implies a time series
  - Absence of cause, absence of effect
  - X cause Y, Y cause X, Z cause X & Y

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Types of Correlation Coefficients

- Pearson's r
  - **Parametric**
  - Equal intervals / Ratio scale
  - Linear relationship

$$r = \frac{N\Sigma(XY) - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}},$$

όπου

r = ο συντελεστής συσχέτισης Pearson r,
N = ο αριθμός των ατόμων που παίρνουν μέρος στην έρευνα,
X = η κάθε τιμή της πρώτης μεταβλητής,
Y = η κάθε τιμή της δεύτερης μεταβλητής, και
Σ = το άθροισμα των...

- Spearman's rho
  - **Non-parametric**
  - Ordinal scale
  - Linear relationship

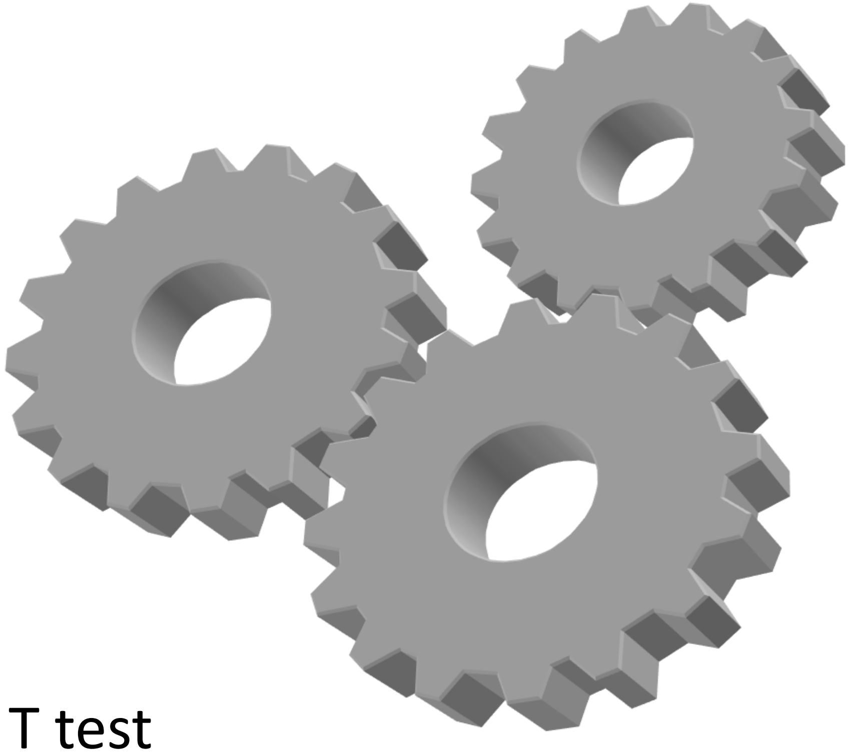
$$rho = 1 - \frac{6\Sigma d^2}{N\left(N^2 - 1\right)},$$

όπου

d = η διαφορά (απόκλιση) των δύο ιεραρχικών τιμών κάθε ατόμου,
N = ο αριθμός των ατόμων που συμμετέχουν στην έρευνα, και
Σ = το άθροισμα των...

○ **SPSS: Analyze → Correlate → Bivariate**

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



T test

Group Comparisons

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

Group Comparisons: Calculating the Probability that the Samples Come From the Same Population

- Sources of volatility:
  - Systematic differences due to experimental manipulation
  - Non-systematic differences due to individual differences between samples (e.g. intelligence, motivation)

- We assess the difference between two samples using statistical tests whose characteristics we know (*t test*)
  - We know the distribution, and therefore we can calculate the probability of getting any value in that distribution - example age of death!
- For a conditional comparison of two different samples, statistical tests represent the following fraction: Systematic Differentiation/Non-Systematic Differentiation (*ANOVA*)
  - Comparison of variability resulting from experimental manipulation with normal variability

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-Tet

- Comparing 2 groups
  - $t = (m - m_{12})/SE$ (estimate of the standard error of the difference between the 2 means)
  - Standard error reveals how well the average of a sample represents the population
  - Standard error of the difference of 2 means reveals the random (natural, non-systematic) variation that we might expect between the means of 2 samples
  - Similarly, different samples from the same population should have quite similar means and standard errors, and so should their differences ($m_1 – m_2$)

- **Therefore,** t-test represents the difference between 2 means corrected (denominator) by the degree of completely random variation of these samples

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-test for One Sample - known Standard Deviation

- Conditions
  - Equal interval or ratio scale
  - One sample of participants
  - Data must meet the conditions for the use of parametric criteria
- t-test - Standard deviation (σ) known
- **Sample distribution of the mean**
  - $\sigma_M = \sigma / \sqrt{N}$ (SE, standard deviation of population means)
  - E.g. $M = 58$, $N = 8$
  - $\mu = 50$, $\sigma = 10$
  - H0: $m = 50$. H1: μ =/= 50

- **$z = (M - \mu) / \sigma_M$** or **$z = (M - \mu) / (\sigma / \sqrt{N})$**
  - $\sigma_M = \sigma / \sqrt{N} = 10 / \sqrt{8} = 3.53$
  - $z = (M - \mu) / \sigma_M = (58 - 50) / 3.53 = 2.27$ < .05 so we reject H0
  - Compare to Z table of standardized values

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Z Table (Two-TAiled Hypothesis)

Two tails of Z

Entries in the table represent two-tailed P values for z statistics

	hundredths									
tenths	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	1.00000	0.99202	0.98404	0.97607	0.96809	0.96012	0.95216	0.94419	0.93624	0.92829
0.1	0.92034	0.91241	0.90448	0.89657	0.88866	0.88076	0.87288	0.86501	0.85715	0.84931
0.2	0.84148	0.83367	0.82587	0.81809	0.81033	0.80259	0.79486	0.78716	0.77948	0.77182
0.3	0.76418	0.75656	0.74897	0.74140	0.73386	0.72634	0.71885	0.71138	0.70395	0.69654
0.4	0.68916	0.68181	0.67449	0.66720	0.65994	0.65271	0.64552	0.63836	0.63123	0.62413
0.5	0.61708	0.61005	0.60306	0.59611	0.58920	0.58232	0.57548	0.56868	0.56191	0.55519
0.6	0.54851	0.54186	0.53526	0.52869	0.52217	0.51569	0.50925	0.50286	0.49650	0.49019
0.7	0.48393	0.47770	0.47152	0.46539	0.45930	0.45325	0.44725	0.44130	0.43539	0.42953
0.8	0.42371	0.41794	0.41222	0.40654	0.40091	0.39533	0.38979	0.38430	0.37886	0.37347
0.9	0.36812	0.36282	0.35757	0.35237	0.34722	0.34211	0.33706	0.33205	0.32709	0.32217
1.0	0.31731	0.31250	0.30773	0.30301	0.29834	0.29372	0.28914	0.28462	0.28014	0.27571
1.1	0.27133	0.26700	0.26271	0.25848	0.25429	0.25014	0.24605	0.24200	0.23800	0.23405
1.2	0.23014	0.22628	0.22246	0.21870	0.21498	0.21130	0.20767	0.20408	0.20055	0.19705
1.3	0.19360	0.19020	0.18684	0.18352	0.18025	0.17702	0.17383	0.17069	0.16759	0.16453
1.4	0.16151	0.15854	0.15561	0.15272	0.14987	0.14706	0.14429	0.14156	0.13887	0.13622
1.5	0.13361	0.13104	0.12851	0.12602	0.12356	0.12114	0.11876	0.11642	0.11411	0.11183
1.6	0.10960	0.10740	0.10523	0.10310	0.10101	0.09894	0.09691	0.09492	0.09296	0.09103
1.7	0.08913	0.08727	0.08543	0.08363	0.08186	0.08012	0.07841	0.07673	0.07508	0.07345
1.8	0.07186	0.07030	0.06876	0.06725	0.06577	0.06431	0.06289	0.06148	0.06011	0.05876
1.9	0.05743	0.05613	0.05486	0.05361	0.05238	0.05118	0.05000	0.04884	0.04770	0.04659
2.0	0.04550	0.04443	0.04338	0.04236	0.04135	0.04036	0.03940	0.03845	0.03753	0.03662
2.1	0.03573	0.03486	0.03401	0.03317	0.03235	0.03156	0.03077	0.03001	0.02926	0.02852
2.2	0.02781	0.02711	0.02642	0.02575	0.02509	0.02445	0.02382	0.02321	0.02261	0.02202
2.3	0.02145	0.02089	0.02034	0.01981	0.01928	0.01877	0.01827	0.01779	0.01731	0.01685
2.4	0.01640	0.01595	0.01552	0.01510	0.01469	0.01429	0.01389	0.01351	0.01314	0.01277
2.5	0.01242	0.01207	0.01174	0.01141	0.01109	0.01077	0.01047	0.01017	0.00988	0.00960
2.6	0.00932	0.00905	0.00879	0.00854	0.00829	0.00805	0.00781	0.00759	0.00736	0.00715
2.7	0.00693	0.00673	0.00653	0.00633	0.00614	0.00596	0.00578	0.00561	0.00544	0.00527
2.8	0.00511	0.00495	0.00480	0.00465	0.00451	0.00437	0.00424	0.00410	0.00398	0.00385
2.9	0.00373	0.00361	0.00350	0.00339	0.00328	0.00318	0.00308	0.00298	0.00288	0.00279
3.0	0.00270	0.00261	0.00253	0.00245	0.00237	0.00229	0.00221	0.00214	0.00207	0.00200
3.1	0.00194	0.00187	0.00181	0.00175	0.00169	0.00163	0.00158	0.00152	0.00147	0.00142
3.2	0.00137	0.00133	0.00128	0.00124	0.00120	0.00115	0.00111	0.00108	0.00104	0.00100
3.3	0.00097	0.00093	0.00090	0.00087	0.00084	0.00081	0.00078	0.00075	0.00072	0.00070
3.4	0.00067	0.00065	0.00063	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050	0.00048
3.5	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00037	0.00036	0.00034	0.00033
3.6	0.00032	0.00031	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024	0.00023	0.00022

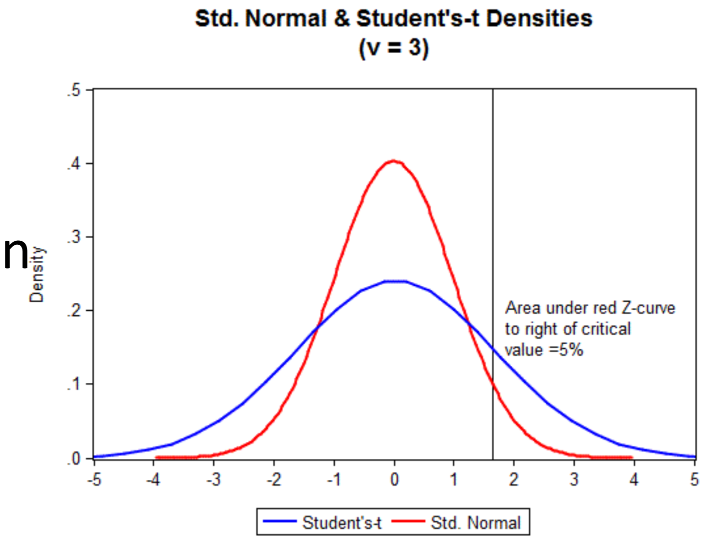
Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-test for One Sample - Unknown Standard Deviation

- Conditions
  - Equal intervals or ratio scale
  - One sample of participants
  - Data must meet the conditions for the use of parametric criteria

- Variation in population unknown (σ_M/σ)
- **Use sample standard deviation (s)**
  - $s_M = s / \sqrt{N}$
- So, we cannot refer to a normal distribution (z), we rather use a t distribution
  - Similarities: symmetrical
  - Differences: **Different distribution for each sample size.** We need to check the appropriate **degrees of freedom** (calculated based on the sample size)
  - Larger rejection area than normal distribution



Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-test for One Sample - Unknown Standard Deviation

- Personality Traits Test
- Does our sample tested in the number of lies per week differ from the population?
  - μ = 7.5
  - H0: μ = M. H1: μ =/= M
- We calculate M = 8.7
- We calculate s
  - s = 1.70
- We calculate t
  - **t = (M - μ) / (s / √N)** = (8.7 - 7.5) / (1.70 / √10) = 2.22
- **Comparison with table of standard values (t)**
  - If |t| > critical value, reject. H0
  - t < 2.26, we accept H0
  - t(9) = 2.22

Participant	Score
1	9
2	8
3	10
4	8
5	7
6	8
7	12
8	9
9	6
10	10

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-test for Independent Samples

- Independent t-test
  - 2 groups, different participants in each group
  - E.g. 2 groups of 10 people, one group read a popular psychology book, another group read Marie Claire. Dependent measurement: degree of happiness from their relationship (assume that the measurement is objective - "happiness meter"

Group Statistics

	Book Read	N	Mean	Std. Deviation	Std. Error Mean
Relationship Happiness	Women are from Bras, Men are from Penis	10	20.0000	4.10961	1.29957
	Marie Claire	10	24.2000	4.70933	1.48922

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Relationship Happiness	Equal variances assumed	.491	.492	-2.125	18	.048	-4.2000	1.97653	-8.35253	-.04747
	Equal variances not assumed			-2.125	17.676	.048	-4.2000	1.97653	-8.35800	-.04200

$Df = n_1 + n_2 - 2$

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Assessing Effect Size

- Conversion of t to r (Pearson coefficient)

- $r = \sqrt{t^2 / (t^2 + df)} = .45$

- **Interpretation and reporting of the t-test result**:
  - On average, the degree of happiness from the relationship after reading Marie Claire (M = 24.20, SE = 1.49) was statistically greater than the degree of happiness after reading the popular psychology book (M = 20.00, SE = 1.30), *t*(18) = 2.12, p = .048, r = .45)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T test of Independent Measures in SPSS

- Analyze → Compare Means → Independent-Samples T Test
  - Test variable (dependent variable)
  - Grouping Variable → Define Groups (independent variable)

  - t, df, p value (sig.)
  - Equal variances assumed (unless Levene test < .05)
  - descriptives to check which group is larger, in case of statistically significant difference

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T-test for Paired Samples / Repeated Measures)

- Dependent (matched pairs) t-test
  - 2 groups, same participants in each group, repeated measurements
  - E.g. 1 group of 500 people, each read a book on pop psychology and a book on Research Methods by Andy Field.
  - Counterbalancing the order of participation in each condition, 6 months apart
  - Dependent measurement: degree of happiness from their relationship (assume that the measurement is objective - "happiness meter").

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Women are from Bras, Men are from Penis	20.0180	500	9.98123	.44637
	Field & Hole	18.4900	500	8.99153	.40211

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Women are from Bras, Men are from Penis & Field & Hole	500	.117	.009

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Women are from Bras, Men are from Penis - Field & Hole	1.5280	12.62807	.56474	.4184	2.6376	2.706	499	.007

SPSS Output 6.3

- df = n - 1

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Effect Size Control (Effect Size)

- Conversion of t to r (Pearson coefficient)
- $r = t^2 / (t^2 + df) = .12$

- **Interpretation and reporting of the t-test result**:
  - On average, the degree of happiness from the relationship after reading the popular psychology book (M = 20.02, SE = .45) was statistically greater than the degree of happiness after reading the research methods book (M = 18.49, SE = .40), $t(499) = 2.70$, p = .007, r = .12)
  - While p is important, the coefficient r indicates a very small effect size!

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

T test of dependent measures in SPSS

- Analyze → Compare Means → Paired-Samples T Test
  - t, df, p value (sig.)
  - descriptives to check which measurement is larger, in case of a statistically significant difference (increase or decrease)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Retrieved from

- Comparison with population mean, variance (s) known
  - **z = (M - μ) / (σ / √N)**
- Comparison with population mean, variance (s) unknown
  - **t = (M - μ) / (s / √N)**

- Comparison of the means of two independent samples

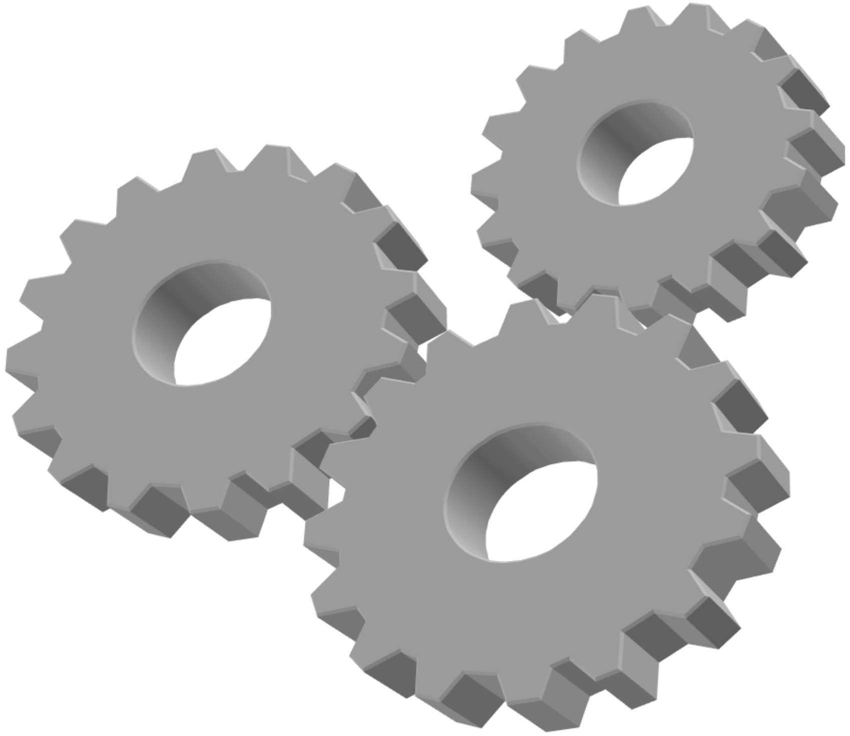
$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\left[\dfrac{\left(\Sigma X_A^2 - \dfrac{(\Sigma X_A)^2}{N_A}\right) + \left(\Sigma X_B^2 - \dfrac{(\Sigma X_B)^2}{N_B}\right)}{(N_A + N_B - 2)}\right]\left[\dfrac{N_A + N_B}{(N_A)(N_B)}\right]}},$$

- Comparison of the means of two dependent samples, or means from repeated measures in the same subjects

$$t = \frac{\Sigma d}{\sqrt{\left[\dfrac{N\Sigma d^2 - (\Sigma d)^2}{N-1}\right]}},$$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Group comparisons II:

Univariate Analysis of Variance (ANOVA) with one or more factors

Analysis of Covariance (ANCOVA)

Multivariate Analysis of Variance (MANOVA)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Analysis Of Variance - ANOVA

- ANOVA (Analysis of Variance) essentially serves as an extension of the t-test. Its primary function is to compare more than two groups, resolving the issue of multiple comparisons (family-wise error) by conducting a unified comparison across all groups from the outset.
  - In essence, ANOVA offers a robust analytical framework, ensuring a comprehensive and precise assessment when dealing with complex research designs involving multiple groups and independent variables.

- Handling Multiple Independent Variables
  - One of ANOVA's significant strengths lies in its ability to handle multiple independent variables. It not only investigates the main effect of each variable but also explores the interactions between these variables, providing a nuanced understanding of their combined impact on the observed phenomena.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

The Logic Behind ANOVA

- **Variation in Independent Measurement Groups** (for example, 2 groups):
  1. **Differences in Individual Values (Noise)**.
     - Individual Differences.
     - Random Error.
  2. **Different Means per Group:**
     - Effect of the Independent Variable Manipulated by the Experimenter.
  - Overall variability: composite effect of all these sources.

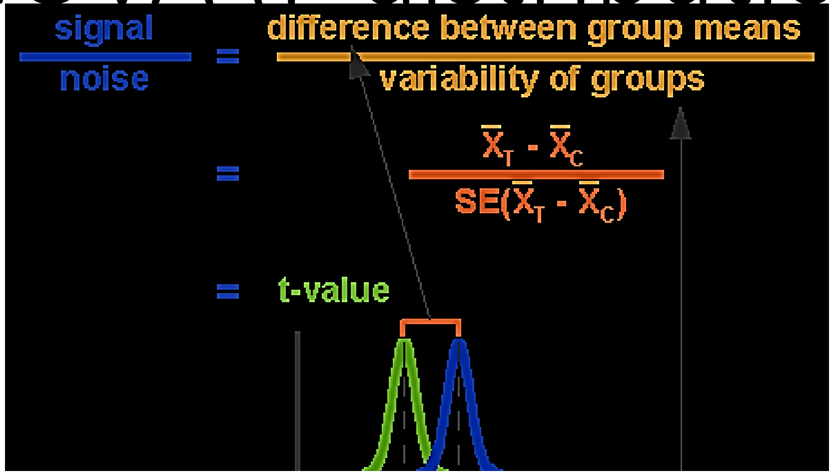
- Reducing Sampling Error / Noise → appropriate sampling method important, matching noise in groups (randomization is helpful)
- In the Case of Statistically Significant Difference: 2 > 1
  - Can be expressed as a ratio, the more significant the larger:

$$\frac{\text{Variance due to Group Differences}}{\text{Random Variance (error)}}$$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Differences Between T-Test (T distribution) and ANOVA (F distribution)



- Compares two samples by assessing the difference betweem two means
- Compares 1-2 samples
- Assesses only one IV

$$F = \frac{\text{estimate of } \sigma^2 \text{ from means}}{\text{estimate of } \sigma^2 \text{ from individuals}}$$

$$F = \frac{\text{Variance between Treatments}}{\text{Variance within Treatments}}$$

$$F = \frac{\text{Variance of Treatments}}{\text{Variance of Error}}$$

- Compares samples by assessing the variance among all samples
- Compares 2+ samples
- Assesses two or more IVs
- Assesses the **interaction** between IVs

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

ANOVA Assumptions

- Independence of measurements of independent groups – samples (random sampling)
- Scale of the Dependent Variable: interval or ratio (not nominal or hierarchical)
- Symmetric (Normal) Distribution of the Dependent Variable in every group/sample (or at least similarly skewed in all groups)
- Homogeneity of Variances of all groups/samples

- A good overview of ANOVA:
  - http://old.psych.uoa.gr/~vpavlop/index.files/pdf/ANOVA_models.pdf

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Anova Structure

- Null hypothesis:
  - all (three or more) groups have similar means and, therefore, systematic differentiation (variance) will be similar to nonsystematic differentiation if they originate from the same population.

- ANOVA based on F ratio: compares systematic variability (SS_M numerator) with non-systematic (random variability– SS_R denominator) in our data
  - However, since it approximates the overall differentiation of all groups, it cannot answer the question of which specific groups differ. Its function is to reject the null hypothesis, as stated above.
  - For this goal we conduct "planned comparisons" or "post-hoc tests"

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Comparisons

- Which levels of the variable (factor) differ?

❖Correction of family-wise error (*decided before conducting the analysis*): Planned VS. Posthoc comparisons

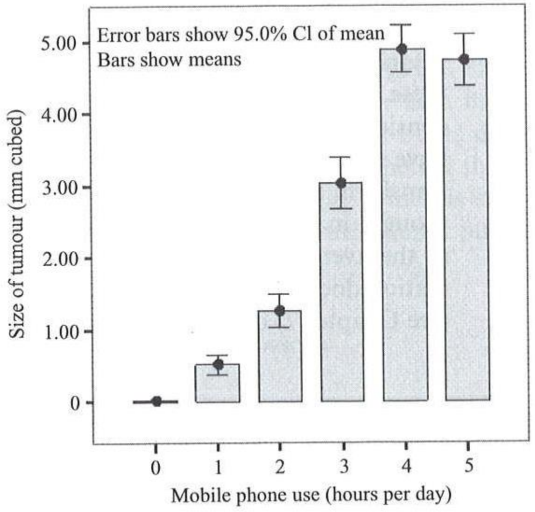
  - Separate **t-tests** (*few planned comparisons*)
  - **LSD** (lenient, small number of conditions)
  - **Sidak**
  - **Bonferonni** (stringent) (*multiple planned comparisons*)
    - $\alpha/n_{t:}$ alpha divided by number of comparisons (e.g. α = .05 / 6 = .0083)
  - **Tukey HSD** (very stringent) (*all possible combinations*)
  - **Scheffe** (quite stringent)

  o **Games Howell** (stringent) *(can be used with unequal variances, *or in case of unequal sample sizes in groups*)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

One-Way Avova, Bet. Subj.s: Preliminary Info.

- Assumption of Homogeneity of Variances
- Levene test: Is the variance equal in all groups? If p<.05, then variances are not equal, therefore we have a violation of the Assumption of Homogeneity of Variances → transformation or … arametric method



Fi…
da…

Descriptives

Size of Tumour (MM cubed)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
0	20	.0175	.01213	.00271	.0119	.0232	.00	.04
1	20	.5149	.28419	.06355	.3819	.6479	.00	.94
2	20	1.2614	.49218	.11005	1.0310	1.4917	.48	2.34
3	20	3.0216	.76556	.17118	2.6633	3.3799	1.77	4.31
4	20	4.8878	.69625	.15569	4.5619	5.2137	3.04	6.05
5	20	4.7306	.78163	.17478	4.3648	5.0964	2.70	6.14
Total	120	2.4056	2.02662	.18500	2.0393	2.7720	.00	6.14

SPSS Output 6.4

Test of Homogeneity of Variances

Size of Tumour (MM cubed)

Levene Statistic	df1	df2	Sig.
10.245	5	114	.000

Anova Results

ANOVA

Size of Tumour (MM cubed)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	450.664	5	90.133	269.733	.000
Within Groups	38.094	114	.334		
Total	488.758	119			

- Between groups: SS_M (M=Model) – systematic variability
- Within groups: SS_R (R=Random) – random (expected) variability
- Mean Squares (MS) = Sum of Squares (SS)/df

- **Some useful Post hoc tests**, at increasing degrees of strictness: Bonferroni, Games-Howell, Tukey HSD

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Comparisons

Dependent Variable: Size of Tumour (MM cubed)
Games-Howell

(I) Mobile Phone Use (Hours Per Day)	(J) Mobile Phone Use (Hours Per Day)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	-.4973*	.18280	.000	-.6982	-.2964
	2	-1.2438*	.18280	.000	-1.5916	-.8960
	3	-3.0040*	.18280	.000	-3.5450	-2.4631
	4	-4.8702*	.18280	.000	-5.3622	-4.3783
	5	-4.7130*	.18280	.000	-5.2653	-4.1608
1	0	.4973*	.18280	.000	.2964	.6982
	2	-.7465*	.18280	.000	-1.1327	-.3603
	3	-2.5067*	.18280	.000	-3.0710	-1.9424
	4	-4.3729*	.18280	.000	-4.8909	-3.8549
	5	-4.2157*	.18280	.000	-4.7908	-3.6406
2	0	1.2438*	.18280	.000	.8960	1.5916
	1	.7465*	.18280	.000	.3603	1.1327
	3	-1.7602*	.18280	.000	-2.3762	-1.1443
	4	-3.6264*	.18280	.000	-4.2017	-3.0512
	5	-3.4692*	.18280	.000	-4.0949	-2.8436
3	0	3.0040*	.18280	.000	2.4631	3.5450
	1	2.5067*	.18280	.000	1.9424	3.0710
	2	1.7602*	.18280	.000	1.1443	2.3762
	4	-1.8662*	.18280	.000	-2.5607	-1.1717
	5	-1.7090*	.18280	.000	-2.4429	-.9751
4	0	4.8702*	.18280	.000	4.3783	5.3622
	1	4.3729*	.18280	.000	3.8549	4.8909
	2	3.6264*	.18280	.000	3.0512	4.2017
	3	1.8662*	.18280	.000	1.1717	2.5607
	5	.1572	.18280	.984	-.5455	.8599
5	0	4.7130*	.18280	.000	4.1608	5.2653
	1	4.2157*	.18280	.000	3.6406	4.7908
	2	3.4692*	.18280	.000	2.8436	4.0949
	3	1.7090*	.18280	.000	.9751	2.4429
	4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

*. Th

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Reporting Results

- The Levene test indicated a violation of the homogeneity assumption, $F(5, 114) = 10.25$, $p < .001$. Despite attempting data transformation, the problem persisted. Therefore, the ANOVA findings are reported. The analysis revealed a statistically significant difference in participants' brain volume sizes due to mobile phone usage, $F(5, 114) = 269.73$, $p < .001$. Post hoc comparisons (Games-Howell) showed significant differences in each group ($p < .001$ for each comparison) except for the 4 and 5-hour usage groups.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

One-Way ANOVA: One Factor, Repeated Measures

- Factor = variable
- Comparison of multiple means from numerous (>2) measurements in a sample (repeated measures)
  - Each participant contributes to more than one groups/conditions

- One-way ANOVA, repeated measures.
- Assumptions:
  - Difference of Means
  - Interval or ratio scale
  - Repeated measures or matched samples
  - Typical assumptions for parametric analyses

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

One-way Avova – Repeated Measures

- <u>Additional Assumption</u>: *Sphericity*
  - Not only groups, differences between groups should also be similarly distributed
  - Difference between every pair of groups/conditions (paired scores, e.g. weight before and after training) have similar variance

- SPSS: Mauchly’s test of sphericity
  - Statistical significance means violation of sphericity

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS EXAMPLE: 1-Way Repeated

- IV: Number of drinks consumed (1-4 drinks)
- DV: Number of people notices/flirting
- Is this bevavior influenced by alcohol consumption?

Within-Subjects Factors

Measure: MEASURE_1

ALCOHOL	Dependent Variable
1	PINT1
2	PINT2
3	PINT3
4	PINT4

Descriptive Statistics

	Mean	Std. Deviation	N
1 Pint	11.7500	4.31491	20
2 Pints	11.7000	4.65776	20
3 Pints	15.2000	5.80018	20
4 Pints	14.9500	4.67327	20

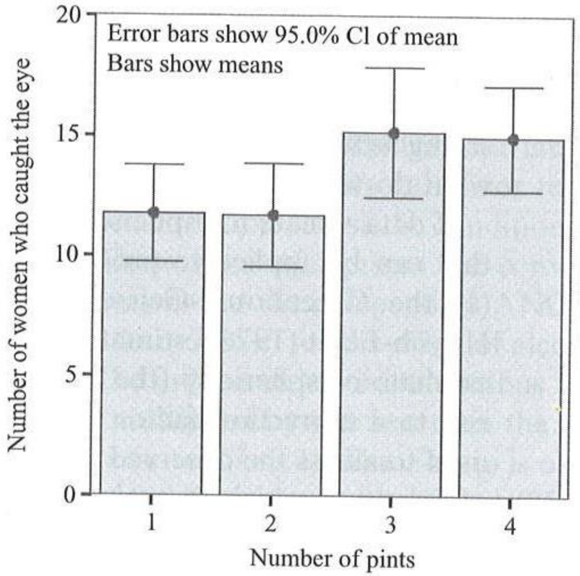


Figure 6.4 Error bar chart of the mean number of women eyed-up after

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Partnership Agreement 2021-2027

Sphericity & Within-Subjects Effects

- <u>Testing for sphericity</u>

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
ALCOHOL	.477	13.122	5	.022	.745	.849	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: ALCOHOL

- If violated, we will use corrected values (correction on degre[...]
  - Greenhouse-Geisser (stricter)
  - Huynh-Feldt
  - Lower Bound

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ALCOHOL	Sphericity Assumed	225.100	3	75.033	4.729	.005
	Greenhouse-Geisser	225.100	2.235	100.706	4.729	.011
	Huynh-Feldt	225.100	2.547	88.370	4.729	.008
	Lower-bound	225.100	1.000	225.100	4.729	.042
Error(ALCOHOL)	Sphericity Assumed	904.400	57	15.867		
	Greenhouse-Geisser	904.400	42.469	21.296		

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Post hocs

Descriptive Statistics

	Mean	Std. Deviation	N
1 Pint	11.7500	4.31491	20
2 Pints	11.7000	4.65776	20
3 Pints	15.2000	5.80018	20
4 Pints	14.9500	4.67327	20

Pairwise Comparisons

Measure: MEASURE_1

(I) ALCOHOL	(J) ALCOHOL	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	5.000E-02	.742	1.000	-2.133	2.233
	3	-3.450	1.391	.136	-7.544	.644
	4	-3.200	1.454	.242	-7.480	1.080
2	1	-5.000E-02	.742	1.000	-2.233	2.133
	3	-3.500*	1.139	.038	-6.853	-.147
	4	-3.250	1.420	.202	-7.429	.929
3	1	3.450	1.391	.136	-.644	7.544
	2	3.500*	1.139	.038	.147	6.853
	4	.250	1.269	1.000	-3.485	3.985
4	1	3.200	1.454	.242	-1.080	7.480
	2	3.250	1.420	.202	-.929	7.429
	3	-.250	1.269	1.000	-3.985	3.485

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

- The violation of Mauchly's sphericity test ($\chi2 = 13.12$, $p = .022$) necessitated the use of Huynh-Feldt correction for degrees of freedom. The results revealed that men's perception of women was influenced by alcohol consumption [$F(2.55, 48.40) = 4.73$, $p = .008$, $\eta^2 = .20$]. Post hoc comparisons with Bonferroni correction indicated a significant difference between the consumption of 2 and 3 drinks (M2 vs M3: $p = .038$). This implies that men's perception significantly differed when they consumed 2 drinks compared to 3 drinks. The effect size ($r=.40$) suggests a moderate effect.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

One-way ANOVA in SPSS

- 1 Factor, Independent Measures
  - Analyze → Compare Means → One-Way ANOVA
  - Dependent: DV
  - Factor: IV

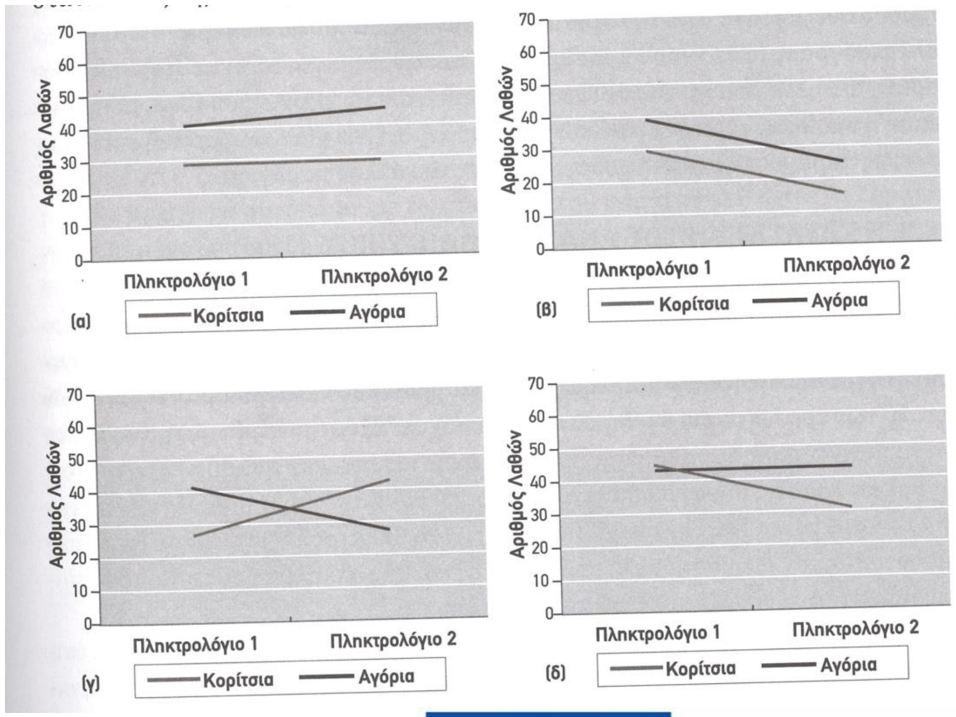
- 1 Factor, Repeated Measures
  - Analyze → General Linear Model → Repeated Measures
  - Within-Subject Factor Name: Name your IV and fill in the number of levels
  - Measure Name: Name your DV

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

ANOVA: Two or More Factors: Main Effects & Interactions

- All potential combinations of the IVs
  - 2 × 2, 3 × 6 etc.
  - E.g. 2 keyboards, 2 genders (2 × 2)



- Main effect
  - The separate effect of each IV on the DV
    - What is the effect of gender (difference between men and women)
    - What is the effect of keyboard (difference between the two keyboards)
- Interaction
  - The effect of each level of an IV on each level of the other IV
    - How does each keyboard affect independently men and women?
    - Alternatively, is there any difference between the two keyboards only in men, or only in women?

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Two-way ANOVA – Independent Measures

- Two IVs, independent samples / groups

❑EXAMPLE:

- A. Age (< 40, > 40), B. Music Genre (3 levels)
- DV: Rating for preference at each music genre (-100 έως +100)

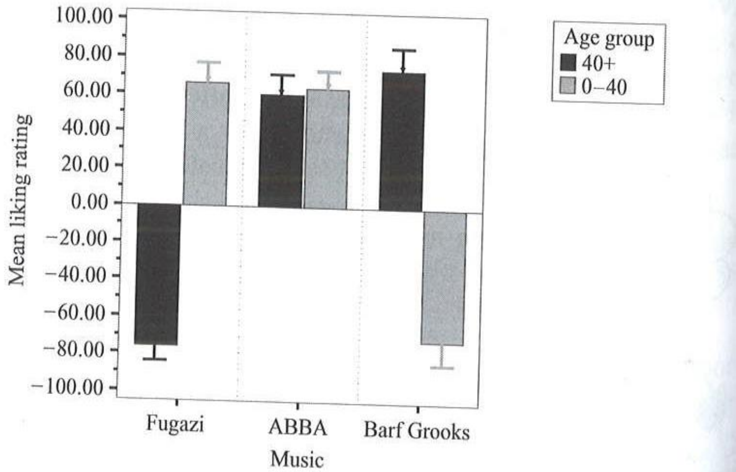


Figure 6.5 Error bar chart of the mean ratings of different types of music for two different age groups

Descriptive Statistics

Dependent Variable: Liking Rating

Music	Age Group	Mean	Std. Deviation	N
Fugazi	40+	-75.8667	14.37193	15
	0-40	66.2000	19.90406	15
	Total	-4.8333	74.23406	30
Abba	40+	59.9333	19.98380	15
	0-40	64.1333	16.99524	15
	Total	62.0333	18.35189	30
Barf Grooks	40+	74.2667	22.29499	15
	0-40	-71.4667	23.17901	15
	Total	1.4000	77.40783	30
Total	40+	19.4444	70.93164	45
	0-40	19.6222	68.06257	45
	Total	19.5333	69.12035	90

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Avova Results

- Testing homogeneity of variance

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Liking Rating

F	df1	df2	Sig.
1.189	5	84	.322

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+MUSIC+AGE+MUSIC * AGE

Tests of Between-Subjects Effects

Dependent Variable: Liking Rating

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	392654.933[a]	5	78530.987	202.639	.000
Intercept	34339.600	1	34339.600	88.609	.000
MUSIC	**81864.067**	**2**	**40932.033**	105.620	**.000**
AGE	**.711**	**1**	**.711**	.002	**.966**
MUSIC * AGE	**310790.156**	**2**	**155395.078**	400.977	**.000**
Error	32553.467	84	387.541		
Total	459548.000	90			
Corrected Total	425208.400	89			

a. R Squared = .923 (Adjusted R Squared = .919)

Multiple Comparisons

Dependent Variable: Liking Rating

	(I) Music	(J) Music	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Games-Howell	Fugazi	Abba	-66.8667*	5.08292	.000	-101.1477	-32.5857
		Barf Grooks	-6.2333	5.08292	.946	-53.3343	40.8677
	Abba	Fugazi	66.8667*	5.08292	.000	32.5857	101.1477
		Barf Grooks	60.6333*	5.08292	.001	24.9547	96.3119
	Barf Grooks	Fugazi	6.2333	5.08292	.946	-40.8677	53.3343
		Abba	-60.6333*	5.08292	.001	-96.3119	-24.9547

Based on observed means.

*. The mean differe

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Reporting Results

- The results indicate that music genre significantly influences the ratings [$F(2, 84) = 105.62$, $p < .001$]. Subsequent pairwise comparisons (adjusted with Games-Howell correction) reveal that ABBA received higher ratings compared to Fugazi and Barf Grooks ($ps < .001$).

- There was no statistically significant difference in ratings between age groups.

- The interaction between age and music genre was statistically significant [$F(2, 84) = 400.98$, $p < .001$), suggesting that different music genres were rated differently by individuals of different ages. Specifically, Fugazi was rated more positively by younger individuals ($M = 66.20$, $SD = 19.90$) compared to older individuals ($M = -75.87$, $SD = 14.37$); ABBA received similar ratings from both younger ($M = 64.13$, $SD = 16.99$) and older individuals ($M = 59.93$, $SD = 19.98$); Barf Grooks had less positive ratings from younger individuals ($M = -71.47$, $SD = 23.17$) compared to older individuals ($M = 74.27$, $SD = 22.29$).

- These findings emphasize the nuanced influence of both age and music genre on individuals' preferences, highlighting the need for targeted analyses when exploring the complexities of factors affecting subjective evaluations.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Two-way ANOVA – Repeated Measures

- Two IVs, repeated measure

- EXAMPLE (sample of 4 patients and 4 conditions – 2 × 2):
  - A. Drug (antidepressant, placebo)
  - B. Therapy (None, Cognitive-Behavioral)
  - DV: Number of suicidal thoughts in the last week of each month

Drug	Placebo		Anti-depressant	
Therapy	None	CBT	None	CBT
Α.Χ.	70	60	81	52
Λ.Μ.	66	52	70	40
Φ.Δ.	56	41	60	31
Β.Κ.	68	59	77	49
Average	**65**	**53**	**72**	**43**

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Two-way ANOVA – Repeated Measures

- Descriptives

Within-Subjects Factors

Measure: MEASURE_1

Drug	Therapy	Dependent Variable
1	1	NoPlacebo
	2	CBTPlacebo
2	1	NoAntidepr
	2	CBTAntidepr

Descriptive Statistics

	Mean	Std. Deviation	N
NoPlacebo	65.0000	6.21825	4
		595	4
		145	4
		683	4

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Drug	1.000	.000	0	.	1.000	1.000	1.000
Therapy	1.000	.000	0	.	1.000	1.000	1.000
Drug * Therapy	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Drug + Therapy + Drug * Therapy

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Testing similarity of variances

(sphericity)

Avova Results (Significance)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Drug	Sphericity Assumed	9.000	1	9.000	1.459	.314	.327	1.459	.139
	Greenhouse-Geisser	9.000	1.000	9.000	1.459	.314	.327	1.459	.139
	Huynh-Feldt	9.000	1.000	9.000	1.459	.314	.327	1.459	.139
	Lower-bound	9.000	1.000	9.000	1.459	.314	.327	1.459	.139
Error(Drug)	Sphericity Assumed	18.500	3	6.167					
	Greenhouse-Geisser	18.500	3.000	6.167					
	Huynh-Feldt	18.500	3.000	6.167					
	Lower-bound	18.500	3.000	6.167					
Therapy	Sphericity Assumed	1681.000	1	1681.000	530.842	.000	.994	530.842	1.000
	Greenhouse-Geisser	1681.000	1.000	1681.000	530.842	.000	.994	530.842	1.000
	Huynh-Feldt	1681.000	1.000	1681.000	530.842	.000	.994	530.842	1.000
	Lower-bound	1681.000	1.000	1681.000	530.842	.000	.994	530.842	1.000
Error(Therapy)	Sphericity Assumed	9.500	3	3.167					
	Greenhouse-Geisser	9.500	3.000	3.167					
	Huynh-Feldt	9.500	3.000	3.167					
	Lower-bound	9.500	3.000	3.167					
Drug * Therapy	Sphericity Assumed	289.000	1	289.000	192.667	.001	.985	192.667	1.000
	Greenhouse-Geisser	289.000	1.000	289.000	192.667	.001	.985	192.667	1.000
	Huynh-Feldt	289.000	1.000	289.000	192.667	.001	.985	192.667	1.000
	Lower-bound	289.000	1.000	289.000	192.667	.001	.985	192.667	1.000
Error(Drug*Therapy)	Sphericity Assumed	4.500	3	1.500					
	Greenhouse-Geisser	4.500	3.000	1.500					
	Huynh-Feldt	4.500	3.000	1.500					
	Lower-bound	4.500	3.000	1.500					

a. Computed using alpha = .05

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Avova Results (Means)

- Drug

Estimates

Measure: MEASURE_1

Drug	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	59.000	3.725	47.146	70.854
2	57.500	4.668	42.644	72.356

- Therapy

Estimates

Measure: MEASURE_1

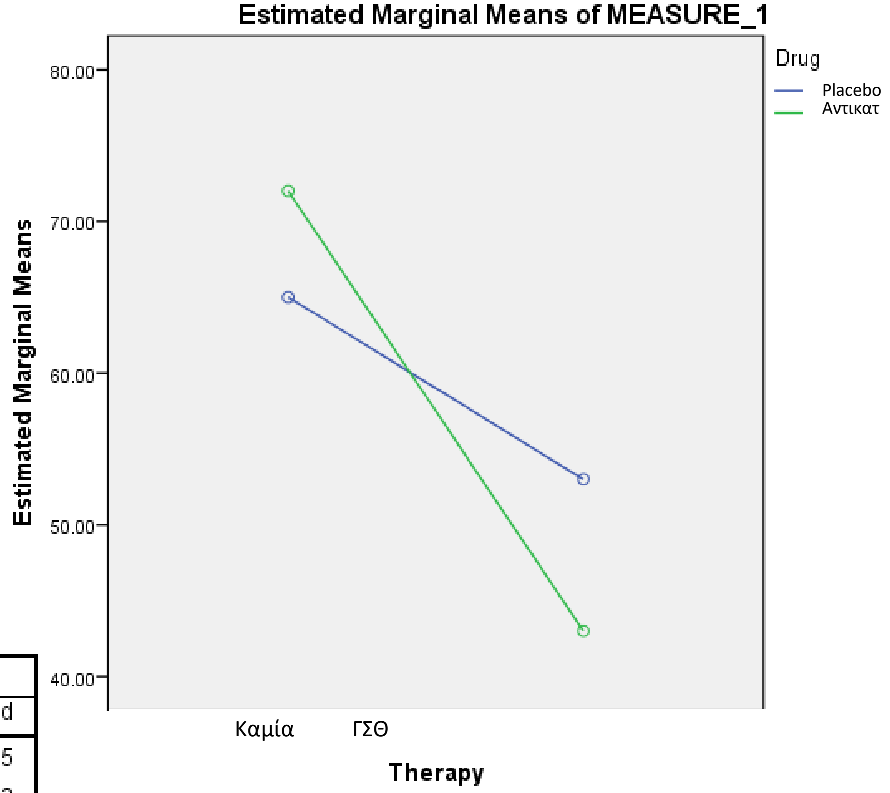
Therapy	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	68.500	3.824	56.329	80.671
2	48.000	4.546	33.532	62.468

- Interaction

4. Drug * Therapy

Measure: MEASURE_1

Drug	Therapy	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	65.000	3.109	55.105	74.895
	2	53.000	4.378	39.067	66.933
2	1	72.000	4.601	57.358	86.642
	2	43.000	4.743	27.904	58.096



CONCLUSION?

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

ANCOVA (Analysis Of Covariance)

- Adapting the results of ANOVA based on the linear relationship (c.f., regression analysis) of the DV and the covariate.
  - The covariate intervenes the relationship between the independent and dependent variables.
  - Useful in cases of non-experimental approaches where handling important confounding variables is challenging.
- Same logic and process as ANOVA, with the only difference being the use of adjusted means.
- Same assumptions, but an additional requirement: homogeneity of regression.
  - Absence of interaction between the covariate and the independent variable.
  - Also, lack of correlation between them, especially when multiple covariates are involved.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

ANOVA in SPSS: 2 Factors

- 2 Factors, Independent Measures
  - Analyze → General Linear Model → Univariate
  - Dependent Variable: DV
  - Fixed Factor(s): IV(s)

- 2 Factors, Repeated Measures
  - Analyze → General Linear Model → Repeated Measures
  - Within-Subject Factor Name: Name the IVs and register the number of levels for each
  - Between Subjects Factor: Independent measures factor(s) (**<u>mixed design</u>**)
  - Measure Name: Name the DV

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

MANOVA / MANCOVA

- Multivariate Analysis of Variance
  - Multiple DVs, one or more IVs
  - Do the population means on a set of dependent variables vary across the levels of a factor or factors?
- More than one DVs
  - Should be moderately correlated with one another
- MANCOVA
  - Same logic with ANCOVA: what would happen if all cases scored equally on the covariates, so that the effect of the factors over and beyond the covariates can be isolated.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

STEPS

- MANOVA statistics, testing main multivariate hypothesis (that the population means on the multiple dependent variables are equal across groups)
  - *Wilks' Lambda*, *Pillai's Trace*, *Hotelling's Trace* (*T*), *Roy's Largest Root*
  - Also report effect sizes (*Partial eta-square* η^2)
- If statistically significant, test separate ANOVA results per DV (with Bonferroni corections)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

MANOVA ASSUMPTIONS

- Multivariate normality*
  - dependent variables are multivariately normally distributed for each population
  - each variable is normally distributed ignoring the other variables and each variable is normally distributed at every combination of values of the other variables
- Homogeneity of variance-covariances matrices*
  - variances for each dependent variable are approximately equal in all groups & covariances between pairs of dependent variables are approximately equal for all groups
- Independence of measurements
  - Random sampling of participants, independent measurements

*robust to violations when groups sized (Ns) are roughly equal

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example

- The pupils at a high school come from three different primary schools.
- The head teacher wanted to know whether there were academic differences between the pupils from the three different primary schools.

- 20 students randomly selected from each School
- DV: marks in end-of-year English and Maths exams. ("English score" and "Maths score“)
- IV: "School", three levels (A, B, C)

Co-funded by the European Union

ΕΣΠΑ 2021-2027 Βιώσιμη Ανάπτυξη για Όλους Partnership Agreement 2021-2027

SPSS Output

Descriptive Statistics

	School	Mean	Std. Deviation	N
English_Score	School A	75.6000	8.22960	20
	School B	62.7000	9.10234	20
	School C	61.5500	7.14124	20
	Total	66.6167	10.30401	60
Maths_Score	School A	43.9000	8.46603	20
	School B	40.7500	8.16201	20
	School C	30.7500	7.71789	20

- Statistically significant difference in ac. perf. among schools
- F (4, 112) = 13.74, p < .0005; Wilk's Λ = 0.450, partial n2 = .33.

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.989	2435.089[a]	2.000	56.000	.000	.989
	Wilks' Lambda	.011	2435.089[a]	2.000	56.000	.000	.989
	Hotelling's Trace	86.967	2435.089[a]	2.000	56.000	.000	.989
	Roy's Largest Root	86.967	2435.089[a]	2.000	56.000	.000	.989
School	Pillai's Trace	.616	12.681	4.000	114.000	.000	.308
	Wilks' Lambda	.450	13.735[a]	4.000	112.000	.000	.329
	Hotelling's Trace	1.075	14.782	4.000	110.000	.000	.350
	Roy's Largest Root	.915	26.072[c]	2.000	57.000	.000	.478

a. Exact statistic

b. Computed using alpha = .05

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + Schoo

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Univariate ANOVAs

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	English_Score	2434.233[a]	2	1217.117	18.114	.000	.389
	Maths_Score	1885.633[c]	2	942.817	14.295	.000	.334
Intercept	English_Score	266266.817	1	266266.817	3962.769	.000	.986
	Maths_Score	88781.067	1	88781.067	1346.134	.000	.959
School	English_Score	2434.233	2	1217.117	18.114	.000	.389
	Maths_Score	1885.633	2	942.817	14.295	.000	.334
Error	English_Score	3829.950	57	67.192			
	Maths_Score	3759.300	57	65.953			
Total	English_Score	272531.000	60				
	Maths_Score	94426.000	60				
Corrected Total	English_Score	6264.183	59				
	Maths_Score	5644.933	59				

a. R Squared = .389 (Adjusted R Squared = .367)

b. Computed using alpha = .05

c. R Squared = .334 (Adjusted R Squared = .311)

- English ($F (2, 57) = 18.11$; $p < .0005$; partial $\eta 2 = .39$)
- Maths ($F (2, 57) = 14.30$; $p < .0005$; partial $\eta 2 = .33$).
- Bonferroni corrected critical α: $p < .025$.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Comparisons

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) School	(J) School	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
English_Score	School A	School B	12.9000*	2.59214	.000	6.6622	19.1378
		School C	14.0500*	2.59214	.000	7.8122	20.2878
	School B	School A	-12.9000*	2.59214	.000	-19.1378	-6.6622
		School C	1.1500	2.59214	.897	-5.0878	7.3878
	School C	School A	-14.0500*	2.59214	.000	-20.2878	-7.8122
		School B	-1.1500	2.59214	.897	-7.3878	5.0878
Maths_Score	School A	School B	3.1500	2.56812	.443	-3.0300	9.3300
		School C	13.1500*	2.56812	.000	6.9700	19.3300
	School B	School A	-3.1500	2.56812	.443	-9.3300	3.0300
		School C	10.0000*	2.56812	.001	3.8200	16.1800
	School C	School A	-13.1500*	2.56812	.000	-19.3300	-6.9700
		School B	-10.0000*	2.56812	.001	-16.1800	-3.8200

Based on observed means.
The error term is Mean Square(Error) = 65.953.

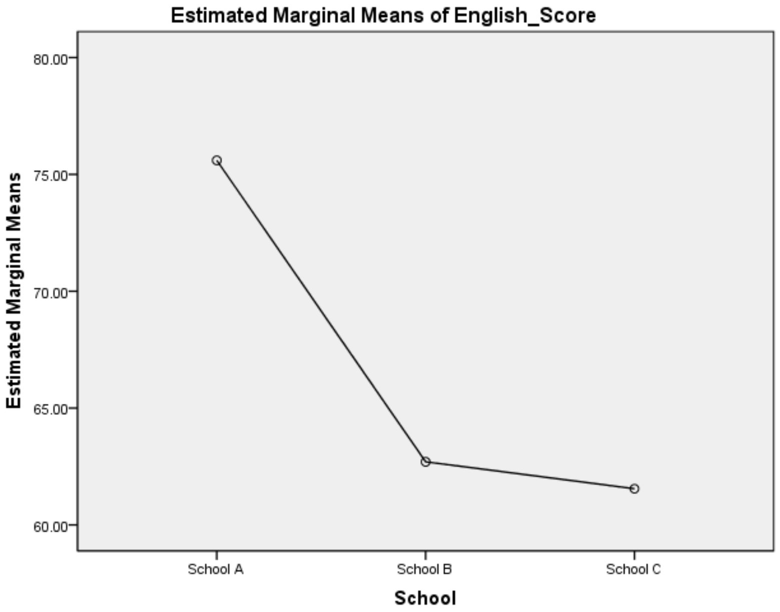
*. The mean difference is significant at the .05 level.

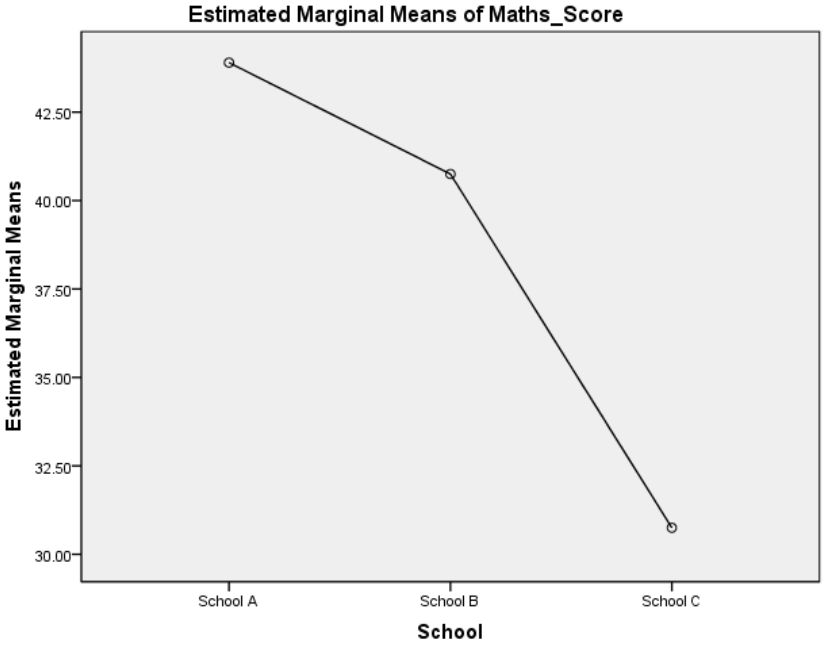
- English: Sig. A vs B ($p < .0005$), A vs C ($p < .0005$), Non-sig. B vs C ($p = .897$).
- Math: Sig. A vs C ($p < .0005$), B vs C ($p = .001$), Non-sig. A vs B ($p = .443$).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

PLOTS

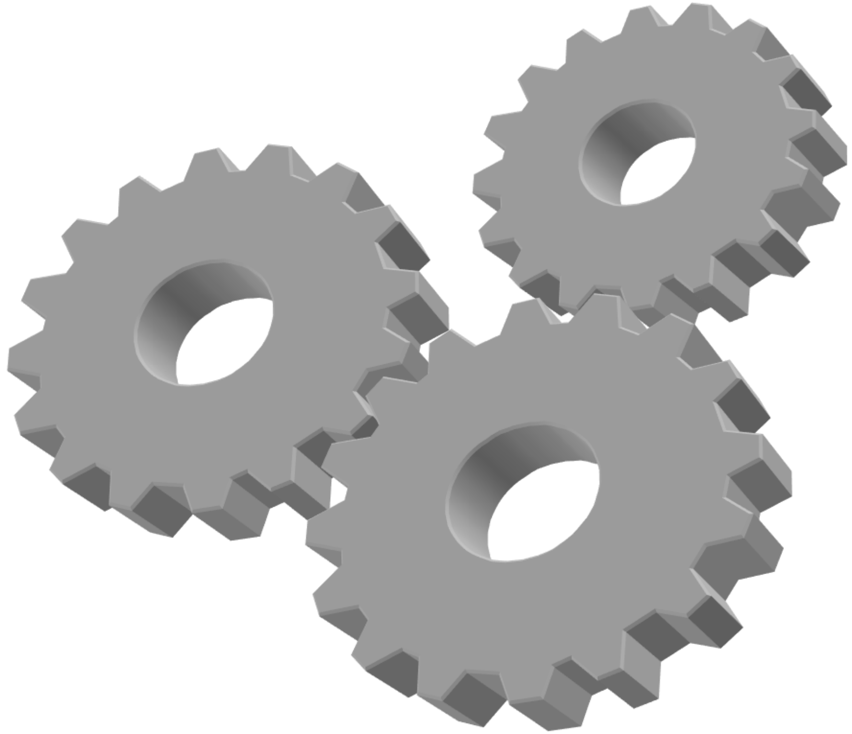




- MANOVA / MANCOVA @ SPSS:
  Analyze → General Linear Model → Multivariate

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Non parametric tests

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric vs. Non-Parametric Analyses: When to Use

- **Key Decision:** Incorrect choice can lead to:
  - Incorrect statistical analysis (violation of assumptions)
  - Lower statistical power (loss of statistical power)
- **Parametric Statistics:** Utilize information about the mean and variance.
- **Non-Parametric Statistics:** Make no assumptions about the data distribution, hence have lower statistical power due to less information used.
  - **Example:** Parametric correlation uses information about the mean and variance, while non-parametric correlation considers only the sequential position of data pairs (ranks).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric vs. Non-Parametric Analyses

Parametric:

- Data in Numeric Scores (Interval/Ratio Scale)
- Use of Means & Variance
- Involves assumptions about the population distribution's shape.

Non-parametric:

- Data in Categories or Hierarchical Order (Nominal/Ordinal Scale)
- Use of Frequencies
- Doesn't assume any specific population distribution shape.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Assumptions of Parametric & Non-Parametric Distributions

- **Assumptions of Parametric Distribution:**
  - Independence of measurements.
  - Measurements are derived from a population following a normal distribution.
  - Populations (e.g., in comparing 2 groups or 2 experimental conditions) have equal variances (homogeneity of variance assumption).

- **Assumptions of Non-Parametric Distribution:**
  - Independence of measurements.
  - The variable of interest has some form of continuity (can be ranked or ordered).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Paranetric vs. Non-Parametric Analyses

Parametric method	Equivalent Non-Parametric
One sample t-test	Chi-square test (test of good fit)
t-test independent measures	Mann-Whitney's U test
t-test repeated measures	Wilcoxon's Signed Rank Test
One-way ANOVA, independent measures	Kruskal-Wallis test
One-way ANOVA, repeated measures	Friedman test

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric vs. Non-Parametric Analyses: When to Use

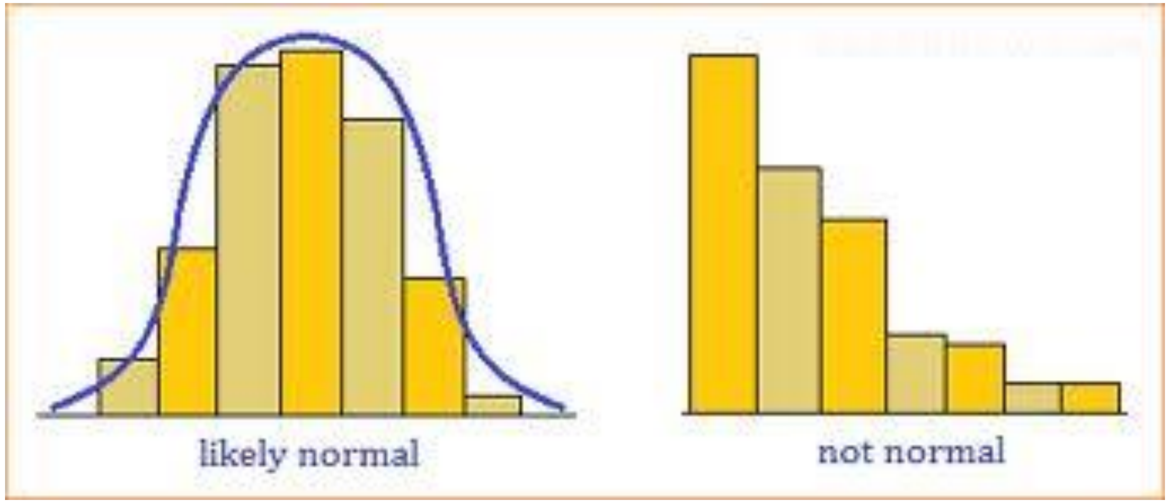
- We have already discussed the basic differences between parametric and non-parametric analyses.
  - Parametric analyses use information about the mean and variance, whereas non-parametric statistics make no assumptions about the data distribution, leading to lower statistical power due to the utilization of fewer pieces of information.

- This emphasizes the importance of descriptive analysis of our data, especially checking for the normality of the distribution. This can be verified in various ways using software like SPSS…

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

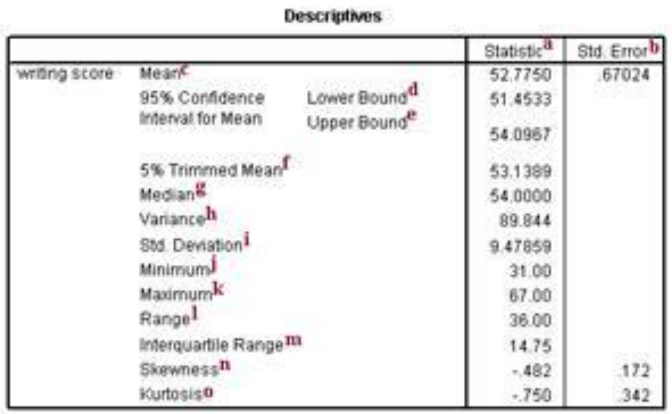
Checking For Normal Distribution



Visual Inspection Using Histograms!

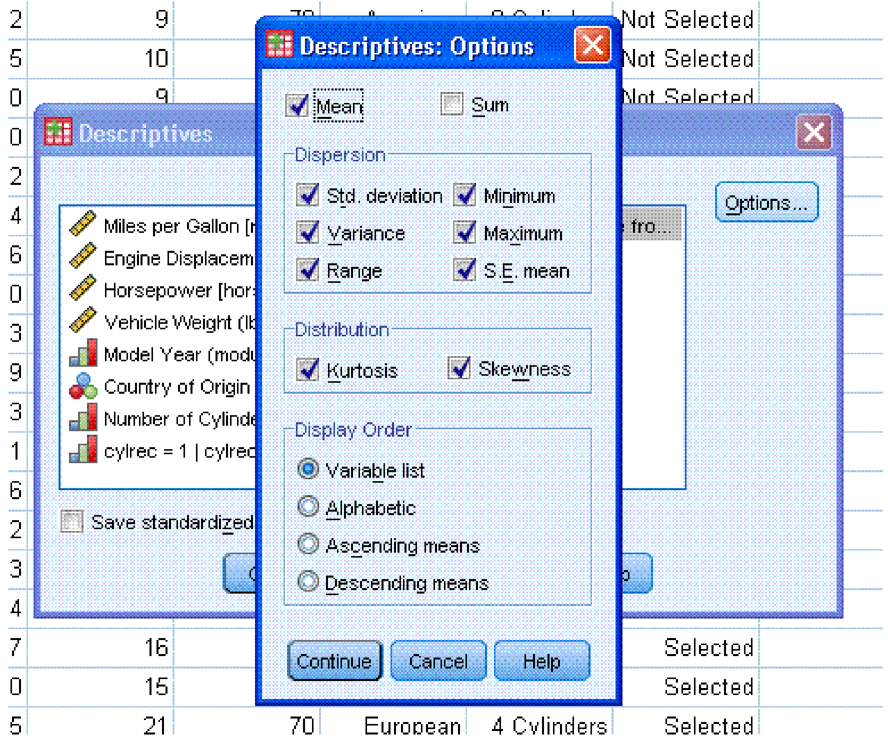
Disadvantages:
- The sample distribution might not be representative of the population.
- It's challenging to have a clear picture of the distribution for N<30.
- Subjectivity! How much deviation from normality is significant?

Checking For Normal Distribution



Descriptives

			Statistic[a]	Std. Error[b]
writing score	Mean[c]		52.7750	.67024
	95% Confidence Interval for Mean	Lower Bound[d]	51.4533	
		Upper Bound[e]	54.0967	
	5% Trimmed Mean[f]		53.1389	
	Median[g]		54.0000	
	Variance[h]		89.844	
	Std. Deviation[i]		9.47859	
	Minimum[j]		31.00	
	Maximum[k]		67.00	
	Range[l]		36.00	
	Interquartile Range[m]		14.75	
	Skewness[n]		-.482	.172
	Kurtosis[o]		-.750	.342



Easy Distribution Check Method in SPSS Frequency Table:

Symmetry (skewness ≈ 1), Kurtosis (kurtosis ≈ 1) → Distribution Spread

It can also be assessed through statistical tests (e.g., Kolmogorov-Smirnov, Shapiro-Wilks).

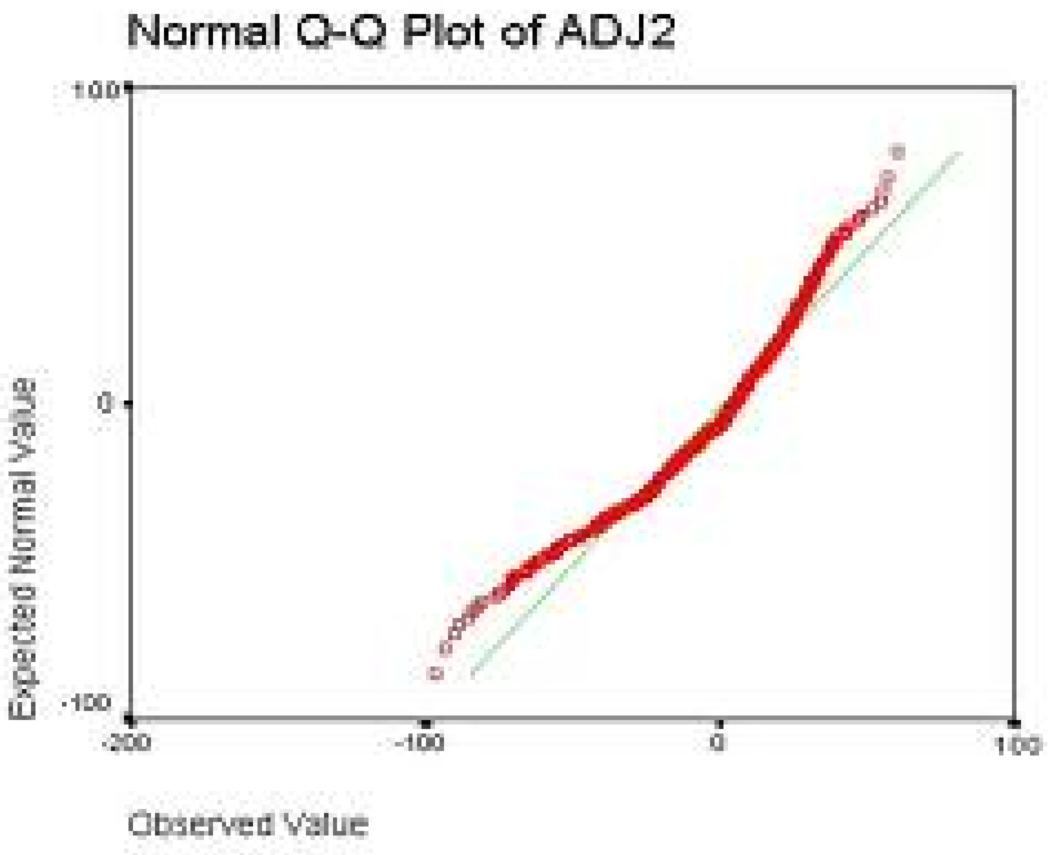
Caution: Statistical power for these tests also depends on the sample size.

Andy Field: Discovering Statistics Using SPSS

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Another Way: Q-Q Plot



Expected (based on Normal Distribution) VS Observed Values in Our Sample

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Checking For Normal Distribution in SPSS

- Frequency Tables:
  - Analyze → Descriptive Statistics → Frequencies

- Histograms, Box-plots, QQ-plots, PP-plots
  - Analyze → Descriptive Statistics → Explore (Plots)
  - Analyze → Descriptive Statistics → Q-Q Plots

- Tests of good fit
  - Kolmogorov-Smirnov (K-S)
    - Analyze → Nonparametric Tests → Legacy Dialogs → 1-Sample K-S
    - Required $p > .05$ (non significant)
  - Shapiro-Wilk (S-W)
    - Value should be around 1

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

What Do We Do in Case of Parametric Assumptions Violation?

- Note: Despite the advantages of non-parametric analysis (such as in cases where we don't know the population distribution), there are drawbacks:
- Lower statistical power ($1-\beta$) – increased risk of Type II errors. For example, we might need a larger sample to achieve the same power as a parametric analysis.
- **Do We Settle for the Loss of Statistical Power?** Not necessarily. We can attempt an intermediate solution: make our data parametric!
  - Addressing skewness in the distribution.
  - Handling outliers.
  - Checking for linearity.
  - Addressing unequal variances.
- This approach can potentially preserve the advantages of parametric analysis while dealing with the challenges posed by non-parametric situations.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Data Transformations

- **Data Trimming: Handling Outliers**
  - E.g., 2-3 standard deviations above/below the mean.
- **Data Transformations: statistically acceptable**
  - The only thing to watch out for is the interpretation of the data!
- Transformation is the process where a mathematical formula is applied to our data (to each score) to address issues like non-normality.
  - Caution: Only in cases of models involving the relationship between variables (e.g., regression analysis), can transformation be applied to a single variable. In variable comparisons, all variables must be transformed.
  - Feel free to try multiple alternatives and choose the best one!

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Popular Transformations

❖ **<u>Use “Compute” function in SPSS</u>**

- **Logarithmic Transformation (positive skewness, non-linearity)**
  - Cannot be applied to zero or negative values (unless a constant is added to all values).
- **Square Root Transformation (Positive Skewness, Lack of Linearity)**
  - Same issue with negative values.
- **Reciprocal Transformation (1/X)**
  - Minimum becomes 0 (for very large scores).
  - Inversion of small and large scores*.
  - Cannot be applied to zero values.
- **Inverse Transformation (Using any of the previous methods: Negative Skewness)**
  - Feel free to try these transformations and choose the one that best fits your data!

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Suggested Transformations

- As suggested by Tabachnick and Fidell (2007) and Howell (2007), the following guidelines (including SPSS compute commands) should be used when transforming data.

- If your data distribution is… Use this transformation method.
  - Moderately positive skewness Square-Root
    - → NEWX = SQRT(X)
  - Substantially positive skewness Logarithmic (Log 10)
    - → NEWX = LG10(X)
  - Substantially positive skewness (with zero values) Logarithmic (Log 10)
    - → NEWX = LG10(X + C)
  - Moderately negative skewness Square-Root
    - → NEWX = SQRT(K – X)
  - Substantially negative skewness Logarithmic (Log 10)
    - → NEWX = LG10(K – X)

- C = a constant added to each score so that the smallest score is 1. K = a constant from which each score is subtracted so that the smallest score is 1; usually equal to the largest score + 1.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS Exercise Data

- Festival data
- Compute variable
- Transforms: Ln, SQRT
- Histograms

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

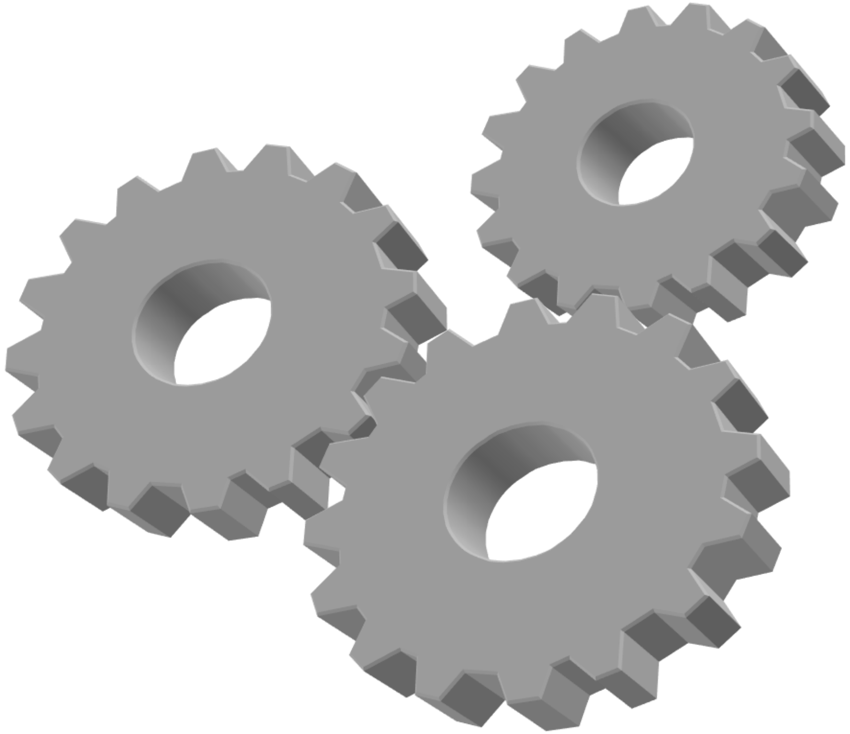
Non Parametric Tests

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027



Chi-Square test (χ^2)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Parametric vs. Non-Parametric Analyses

Parametric	Non-Parametric
Single sample t-test	Chi-square test for goodness of fit

- Chi-square test – testing whether Observed Frequency distribution differs in statistically significant way from Expected Frequency

- 1 variable: Criterion for good fit/appropriateness.
- 2 variables: Independence.

- Requirements:
  - Qualitative data - categorical scale.
  - Frequencies.
  - Prediction of the number of participants in each category.
  - 20 participants per variable, 5 per cell.
  - Each participant contributes to only one cell.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

X^2: 1 variable - goodness of fit

- Instead of a One-Sample t-Test
- To test the differences between observed (actual) and expected (random) variables - Deviation from a random model:
- $\mathbf{X^2 = \Sigma [(O - E)^2 / E]}$
  - $\chi^2 = \Sigma$ [(Observed - Expected)^2 / Expected]

❖ <u>Example</u>

Students' Study Method			
Regular	Irregular	Combined	Total
51	27	42	120

- **A = N / k** = 120 / 3 = 40
- **df = k - 1** = 3 - 1 = 2
- x^2= [(51-40)^2 / 40] + [(27 - 40)^2 / 40] + [(42 - 40)^2 / 40] = 7.35 > 5.99 (critical value from table)
- $x^2(2) = 7.35$, $p = .025$

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

X^2 - Independence of Measurement

- Differences between Observed and Expected Variables (When the Two Variables are Independent)
- Test for Independence of the Two Variables
- Double-Entry Cross Table
  - Use absolute frequencies, not percentages (convert if N is known).

- *Example*

Teaching Method	***Low***	***Average***	***High***	**Total**
New	6	15	23	**44**
Old	10	8	24	**42**
Total	**16**	**23**	**47**	**T = 86**

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Hypotheses

- **H0**: the two variables (performance and teaching method) are independent.

  or

- The number of students with low, moderate, and high performance will be equal for both teaching methods.

- **H1**: The two variables are dependent (related to each other).

  or

- The number of students with low, moderate, and high performance will be different for the two teaching methods.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

X^2 - Calculation

- Calculate x^2 with the same formula
- **$x^2 = \Sigma [(O - E)^2 / E]$**
- $x^2 = [(6 - 8.19)^2 / 8.19] + [(15 - 11.77)^2 / 11.77] + [(23 - 24.05)^2 / 24.05] + [(10 - 7.81)^2 / 7.81] + [(8 - 11.23)^2 / 11.23] + [(24 - 22.95)^2 / 22.95] = 3.11$

- **df = (Σ - 1) * (Γ - 1)** = 2
- 3.11 > 5.99 (critical value from table)
- Therefore we accept H0: x^2 (2) = 3.11, ns
  - If we aceepted H1 we should test in pairs: low - average, low - high, average - high
  - Bonferroni correction: α = .05/3 = .017

- **Calculation for A**
  - **A = (Γ * Σ) / T**

- A1 = (44 * 16) / 86 = 8.19
- A2 = (44 * 23) / 86 = 11.77
- A3 = (44 * 47) / 86 = 24.05
- A4 = (42 * 16) / 86 = 7.81
- A5 = (42 * 23) / 86 = 11.23
- A6 = (42 * 47) / 86 = 22.95

- Ά1 + A2 + ... + Ai = T (here 86)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS: x^2

- 1 qualitative variable
- Analyze --> Nonparametric tests --> legacy dialogs --> chi-square

- 2 qualitative variables
- Analyze --> Descriptive statistics --> Crosstabs
- Statistics: choose x^2 or other test

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mann-Whitney U-Test

- Instead of t-test - independent samples
- Using data on an ordinal scale:
  - H0 Correct: The hierarchical order in Group A will not be systematically higher/lower than Group B.
  - H0 Incorrect: The hierarchical order in Group A will be systematically higher/lower than Group B.

- It tests whether the hierarchical scores of the two samples are randomly mixed or systematically clustered at different extremes of the scale.

- Procedure:
  - Rank groups A & B (from lower to higher).
  - Compute UA and UB: How many scores from the other group follow the sequence of this specific score.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Calculation for Mann-Whitney U

	Ordered Scores		Points for	Points for
Rank	**Score**	**Sample**	**Sample A**	**Sample B**
1	2	(A)	6 points	
2	6	(A)	6 points	
3	8	(B)		4 points
4	9	(A)	5 points	
5	15	(A)	5 points	
6	18	(B)		2 points
7	27	(A)	4 points	
8	48	(A)	4 points	
9	63	(B)		0 points
10	68	(B)		0 points
11	71	(B)		0 points
12	94	(B)		0 points
			$U_A = 30$	$U_B = 6$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mann-Whitney U Assumptions

- NOT required:
- Normal distribution
- Homogeneity of Variance

- Assumes:
- DV is continuous
- Few equal rank scores (hierarchical order)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mann-Whitney U-Test

- EXAMPLE: Comparing men to dogs regarding the number of behaviors typically exhibited by dogs. Sample of 20 "individuals" from each group. Natural observation for 24 hours and recording of behaviors.
- Checking the normality of the distribution:

Tests of Normality

	Species	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Dog-Like Behaviour	Dog	.244	20	.003	.899	20	.039
	Man	.175	20	.109	.933	20	.176

a. Lilliefors Significance Correction

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mann-Whitney U Calculation

http://faculty.vassar.edu/lowry/utest.html

Ranks

	Species	N	Mean Rank	Sum of Ranks
Dog-Like Behaviour	Dog	20	20.77	415.50
	Man	20	20.23	404.50
	Total	40		

- Descriptive information using our ranked data (rankings)
- Informs us about which group has the lowest and highest scores.

Test Statistics[b]

	Dog-Like Behaviour
Mann-Whitney U	194.500
Wilcoxon W	404.500
Z	-.150
Asymp. Sig. (2-tailed)	.881
Exact Sig. [2*(1-tailed Sig.)]	.883[a]

a. Not corrected for ties.

b. Grouping Variable: Species

✓ There is no statistically significant difference in "dog-like" behaviors between men and dogs.

Mann-Whitney U-Test Results

- Men (Median = 27) did not significantly differ from dogs (Median = 24) in the number of "dog-like" behaviors exhibited ($U = 194.5$, $p = .881$).

- *Can also calculate effect size:*
  $r = Z / \sqrt{N} = -0.15/\sqrt{40} = -.02$

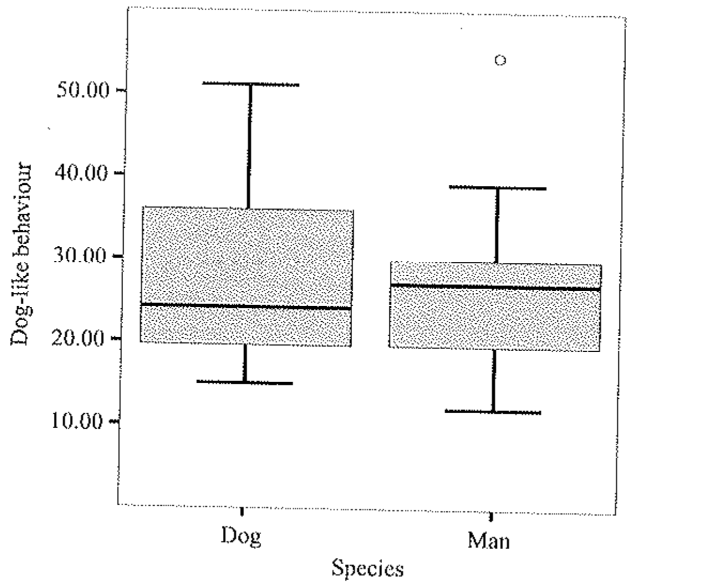


Figure 7.1 Boxplot for the dog-like behaviour in dogs and men

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Wilcoxon Signed-Ranks Test

- Alternative to t-test – repeated measures
- Ordinal scale data (can be assigned ranks)

Α/α	**Treatments**			
Συμ	**1**	**2**	**Διαφορά**	**Ιεραρχ. Σειρά**
1	18	43	+25	6 (largest)
2	9	14	+5	2
3	21	20	-1	1 (smallest)
4	30	48	+18	5
5	14	21	+7	3
6	12	4	-8	4

- Summary of rankings whose difference is positive: 6 + 2 + 5 + 3 = 16
- Summary of rankings whose difference is negative: 1 + 4 = 5
- Wilcoxon T smallest of the two values: 5

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Assumptions of Wilcoxon Signed-Ranks

- NO distribution normality assumed

Assumptions:

- DV continuous
- Few identical/tied scores (rankings)
  - If there several scores have 0 ranking, we distribute the tied ranks equally between the groups of values with negative and positive signs.
  - If two scores have the same difference, assign them the average of the tied ranks (e.g., a tie between the 3rd and 4th rank will result in a rank of 3.5 for both scores).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Wilcoxon Signed-Ranks Test

- EXAMPLE:
  - Do hidden messages in TV ads affect beverage consumption?
  - 32 pps, half see ad without hidden message first, after 6 months same spot with hidden message. The other half followed the reverse order.
  - DV: How many beverages consumed the week after watching the ads.

- If H0 correct: the difference between rankings will not be systematically positive or negative
  - Uniform mixture of positive and negative score differences.
  - No difference between conditions
- If H0 incorrect: Systematic difference in scores, more positive or negative
  - Sign. difference between conditions

Co-funded by the European Union
ΕΣΠΑ 2021-2027 Βιώσιμη Ανάπτυξη για Όλους Partnership Agreement 2021-2027

Wilcoxon Signed-Ranks Test: Results

- Assessing normality of distribution:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Message	.177	32	.012	.961	32	.296
No Message	.236	32	.000	.914	32	.014

a. Lilliefors Significance Correction

- Test Re

Ranks

		N	Mean Rank	Sum of Ranks
No Message - Message	Negative Ranks	11[a]	10.14	111.50
	Positive Ranks	17[b]	17.32	294.50
	Ties	4[c]		
	Total	32		

a. No Message < Message

b. No Message > Message

c. Message = No Message

Smaller of the 2, compared to critical value from table

T-

T+

Co-funded by the European Union

ΕΣΠΑ 2021-2027 Βιώσιμη Ανάπτυξη για Όλους Partnership Agreement 2021-2027

Wilcoxon Signed-Ranks: Results

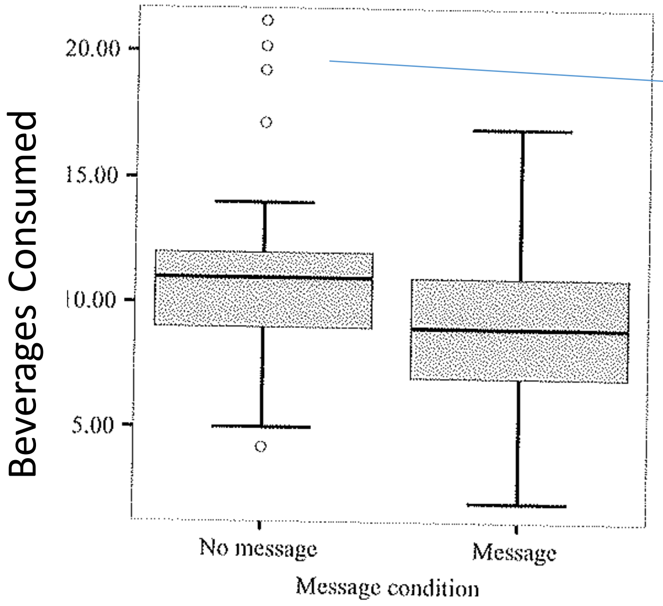
- Transformation to Z score → Follows normal distribution ($\mu = 0, \sigma = 1$)

Test Statistics[b]

	No Message - Message
Z	-2.094[a]
Asymp. Sig. (2-tailed)	.036

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test



Non parametric test is not influenced by outliers!

Conclusion: The number of beverages consumed after the subconscious message (Mdn = 9) is significantly lower than beverages consumed after the regular advertisement(Mdn = 11),
$T = 111.50, p = .036$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- Alternative to One-Way ANOVA, independent groups
- Ordinal scale data

1. Original Data (numerical scores)

Treatments (raw scores)		
I	II	III
14	2	26
3	14	8
21	9	14
5	12	19
16	5	20

2. Find the rank of each score

Numerical Score	Ordinal Rank
2	1
3	2
5	3.5
5	3.5
8	5
9	6
12	7
14	9
14	9
14	9
16	11
19	12
20	13
21	14
26	15

3. Ordinal Data (ranked scores)

Treatments (ranked scores)		
I	II	III
9	1	15
2	9	5
14	6	9
3.5	7	12

4. Ranks in each treatment summed to give T

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- If H0 is true: Hierarchy under a condition (T) will not be systematically higher or lower than others. No difference between the conditions.

- If H0 is false: Hierarchy under at least one condition (T) will be systematically higher or lower than another condition. There is a difference between the conditions.

- Does NOT assume normal distribution

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- In case of a statistically significant result in the Kruskal-Wallis Test – which groups differ?
- We employ Mann-Whitney U tests between pairs of groups.
- Critical value might require Bonferroni correction (= α / number of comparisons)
- Not all possible comparisons, only theoretically meaningful ones.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- EXAMPLE: Treatment of fear of clowns in children. 3 groups (15 children/group), different positive information about clowns.
  - A) McDonald's Advertisements
  - B) Story
  - C) Real Clown
- Also, 1 control group

- Dependent variable: Clown phobia assessment (ascending scale, 0-5).

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- Assessing normality of distribution

Tests of Normality

	Format of Information	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Fear beliefs	Advert	.173	15	.200*	.897	15	.086
	Story	.217	15	.056	.855	15	.020
	Exposure	.230	15	.032	.867	15	.030
	None	.419	15	.000	.603	15	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

- Test Res

Ranks

	Format of Information	N	Mean Rank
Fear beliefs	Advert	15	45.03
	Story	15	21.87
	Exposure	15	23.77
	None	15	31.33
	Total	60	

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test

- Test Results:

Ranks

	Format of Information	N	Mean Rank
Fear beliefs	Advert	15	45.03
	Story	15	21.87
	Exposure	15	23.77
	None	15	31.33
	Total	60	

Test Statistics[a,b]

	Fear beliefs
Chi-Square	17.058
df	3
Asymp. Sig.	.001

a. Kruskal Wallis Test

b. Grouping Variable: Format of Information

Following x^2 distribution, df = k -1
k = number of groups

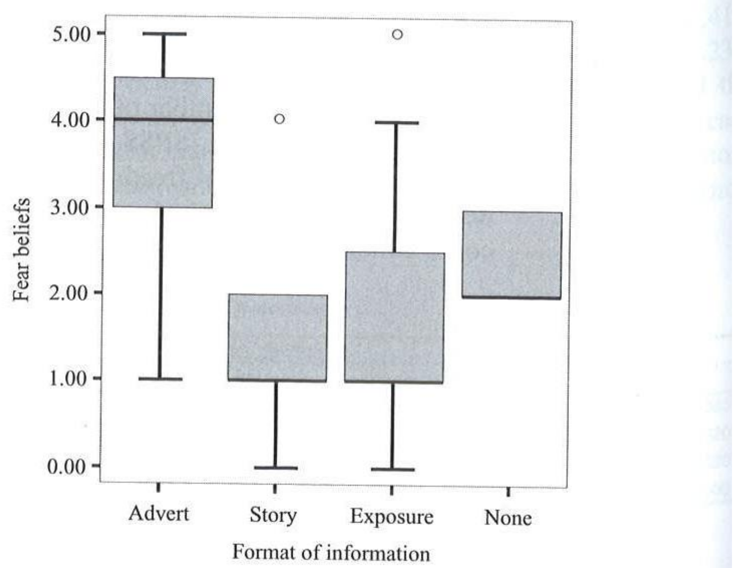


Figure 7.3 Boxplot for the fear beliefs about clowns after exposure to different formats of information (adverts, stories, a real clown or nothing)

Conclusion: Different presentation methods of positive information about clowns affect children's fear ratings.

Which groups differ though?

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Kruskal-Wallis Test - Mann-Whitney Post-Hocs

- Use Mann-Whitney U test between group pairs. Not all possible comparisons, but only theoretically meaningful ones.
- ritical value may need Bonferroni correction (= α / number of comparisons) → .05/3 = .0167

Advert vs. control:

Test Statistics[b]

	Fear beliefs
Mann-Whitney U	37.500
Wilcoxon W	157.500
Z	-3.261
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001[a]

a. Not corrected for ties.

b. Grouping Variable: Format of Information

Story vs. control:

Test Statistics[b]

	Fear beliefs
Mann-Whitney U	65.000
Wilcoxon W	185.000
Z	-2.091
Asymp. Sig. (2-tailed)	.037
Exact Sig. [2*(1-tailed Sig.)]	.050[a]

a. Not corrected for ties.

b. Grouping Variable: Format of Information

Exposure vs. control:

Test Statistics[b]

	Fear beliefs
Mann-Whitney U	72.500
Wilcoxon W	192.500
Z	-1.743
Asymp. Sig. (2-tailed)	.081
Exact Sig. [2*(1-tailed Sig.)]	.098[a]

a. Not corrected for ties.

b. Grouping Variable: Format of Information

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

Kruskal-Wallis Test – Reporting Results

Children's fear of clowns was significantly influenced by the way information was presented ($H(3) = 17.06$, $p = .001$). This result was further explored with Mann-Whitney tests, using Bonferroni correction with a significance level of .0167.

Children's fear significantly increased after watching advertisements, compared to the control group ($U = 37.50$, $p = .001$). However, children's fear did not significantly change after hearing stories or being exposed to a real clown ($U = 65$ and $U = 72.5$ respectively).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Friedman Test

- Alternative to one-way ANOVA, repeated measures
  - Data on a hierarchical (ordinal) scale
  - Does NOT assume normality
  - Requires continuous DV

- If H0 is true:
  - The ranking of a group or time-point measurement is not systematically greater or smaller than the ranking of another time point.
  - No difference between conditions.
- If H0 is false:
  - The rankings of at least one group or time-point will be systematically greater or smaller than the hierarchical order of another time point.
  - There is a significant difference between conditions.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Friedman Test - Παράδειγμα

- EXAMPLE: Television Program Influences on Daily Life of Couples. 54 couples watch 3 types of programs, in random order.

- Dependent Variable: How many times they argue within the next hour after watching.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Friedman Test

- Assessment of normality of distribution:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Eastenders	.137	54	.013	.914	54	.001
Friends	.150	54	.004	.943	54	.012
National Geographic	.121	54	.046	.943	54	.012

a. Lilliefors Significance Correction

- Test Re:

Ranks

	Mean Rank
Eastenders	2.29
Friends	1.81
National Geographic	1.91

Test Statistics[a]

N	54
Chi-Square	7.586
df	2
Asymp. Sig.	.023

a. Friedman Test

Following x^2 distribution,
df = k -1
k = number of conditions

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Friedman Test – Wilcoxon Post-Hocs

- Using the Wilcoxon Signed-Ranks test among pairs of conditions. Not all possible comparisons, only theoretically significant ones.
- Critical value might need Bonferroni correction (= α / number of comparisons).

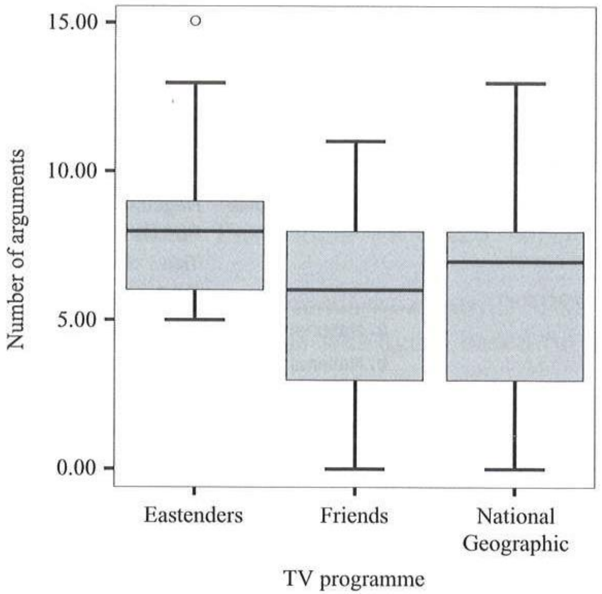


Figure 7.4 Boxplot for the number of arguments had after watching Eastenders, Friends or a National Geographic programme about whales

Ranks

		N	Mean Rank	Sum of Ranks
National Geographic - Eastenders	Negative Ranks	31[a]	28.85	894.50
	Positive Ranks	18[b]	18.36	330.50
	Ties	5[c]		
	Total	54		
National Geographic - Friends	Negative Ranks	21[d]	22.00	462.00
	Positive Ranks	24[e]	23.88	573.00
	Ties	9[f]		
	Total	54		

a. National Geographic < Eastenders
b. National Geographic > Eastenders
c. Eastenders = National Geographic
d. National Geographic < Friends
e. National Geographic > Friends
f. Friends = National Geographic

Test Statistics[c]

	National Geographic - Eastenders	National Geographic - Friends
Z	-2.813[a]	-.629[b]
Asymp. Sig. (2-tailed)	.005	.530

a. Based on positive ranks.
b. Based on negative ranks.
c. Wilcoxon Signed Ranks Test

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Friedman Test – Reporting Results

The number of arguments between couples significantly varied depending on the TV program they watched ($\chi^2 = 7.59$, $p = .023$). This result was further explored using Wilcoxon tests, for which Bonferroni correction was applied – the significance level used was .025.

Watching the soap opera led to a significant increase in arguments compared to watching the documentary ($Z = 330.50$, $p = .005$). In contrast, arguments were not significantly affected by watching the sitcom 'Friends,' compared to watching the documentary ($Z = 462$, $p > .05$).

In conclusion, it appears that watching the soap opera results in a significant increase in arguments between couples compared to watching the documentary, while watching the sitcom does not cause any differentiation compared to the control condition.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

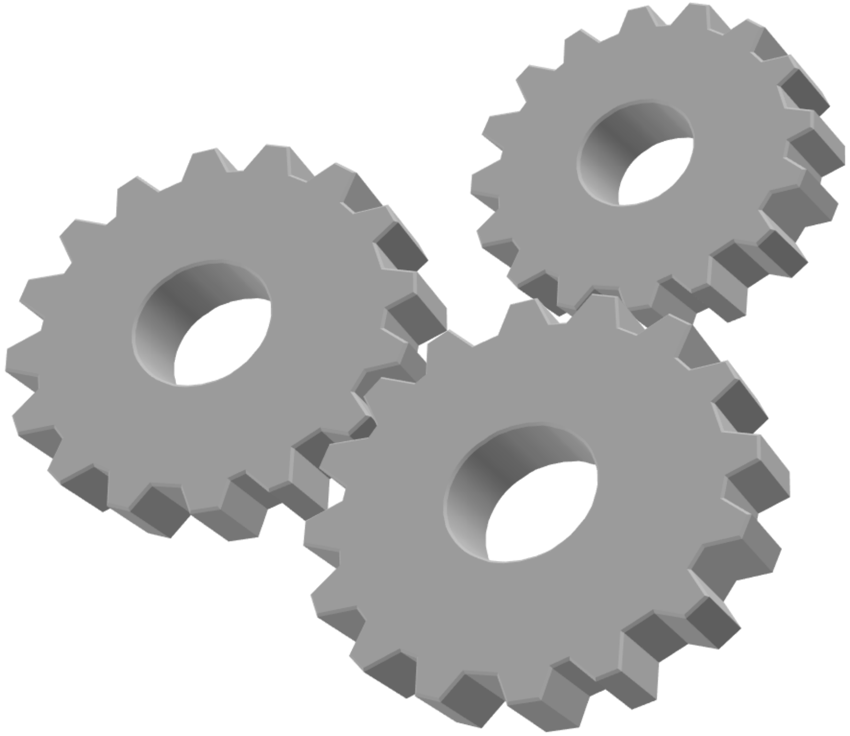
Summary - SPSS

- Nominal or Ordinal Variables
- Non-parametric tests have fewer assumptions but also less power – higher chance for Type II error

- **Chi-square test (one-sample t testt)**
    - 1 qualitative variable: Analyze → Nonparametric tests → legacy dialogs → chi-square
    - 2 qualitative variables: Analyze → Descriptive statistics → Crosstabs
- **Mann-Whitney U (t-test independent measures)**
    - Analyze→ Non-parametric tests → Legacy Dialogs → 2 Independent Samples
- **Wilcoxon Signed Rank (t-test repeated measures)**
    - Analyze→ Non-parametric tests → Legacy Dialogs → 2 Related Samples
- **Kruskal-Wallis (One-way ANOVA – independent measures)**
    - Analyze→ Non-parametric tests → Legacy Dialogs → K Independent Samples
- **Friedman (One-way ANOVA – repeated measures)**
    - Analyze→ Non-parametric tests → Legacy Dialogs → K Related Samples

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Principal Components Analysis (PCA)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factor Analysis (FA) & Principal Components Analyis (PCA)

Suitable to reduce our data by using less variables

- Especially in Psychology, we often deal with difficult to measure phenomena or constructs (unlike direct measurement in positive sciences)
- Utilizing various measurements (e.g., survey questions).
  - All these parts might be components of a common factor, hidden behind them.

- *Alternatively:* describe the association between variables and confirm theories about these associations.

- **Objective:** Discover these common patterns, the underlying (hidden) factors expressing our various measurements.
- Find the smallest number of factors expressing the largest possible percentage of the initial variance.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Basic Principles

❑Two highly correlated variables likely represent the same phenomenon.

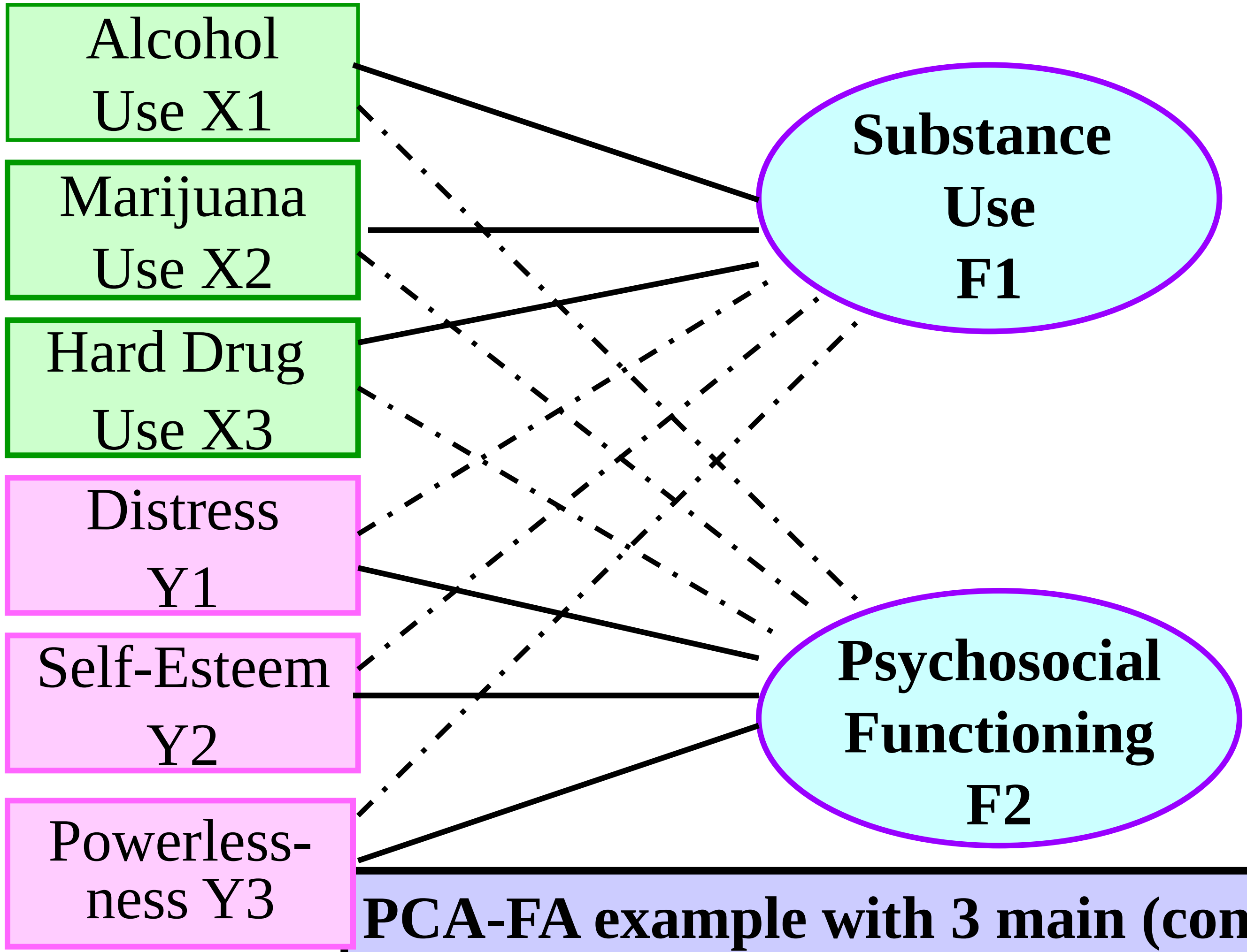
- Combining them into a factor simplifies the studied phenomenon and reduces measurement errors/noise.
  - E.g. years of service in a job (e.g., mail carrier) and the number of delivered letters
  - → Professional Experience: The type of "factor" or "principal component" we aim to discover.

➢Different phrasing of our objective::

  - **To summarize** relationships among a large number of variables concisely and accurately, so as to **make a concept or quality more understandable**.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



PCA-FA example with 3 main (continuous line) loadings and 3 non significant (dotted lines) loadings for each factor.

PCA

- ❑Explains the Total Variance in our dataset (Common, Unique, & Error)
  - *Extracts the maximum possible percentage of variance using the minimum number of components.*
  - Good method to decrease the number of our factors
    - principal components are linear combinations of the original factors and can be thought of as "new" factors
    - Initial factors need to be correlated and have similar variances

- The first component accounts for the highest percentage of variance.
- Each subsequent component accounts for as much remaining variance as possible.
- All variance is fully accounted for if all components are retained.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

PCA Assumptions

- There needs to be a **linear relationship between all variables**.
  - The reason for this assumption is that a PCA is based on Pearson correlation coefficients, and as such, there needs to be a linear relationship between the variables
- You should have **sampling adequacy**
  - for PCA to produce a reliable result, large enough sample sizes are required.
- Your data should be **suitable for data reduction**.
  - Effectively, you need to have adequate correlations between the variables to be reduced to a smaller number of components.
- There should be **no significant outliers**.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Independence Between Factors

- ***Note***: Principal Components (PCs) are **not correlated** to one another
- We want PCs to be independent from one another
  - c.f correlation matrix
- Lack of correlation guarantees independence of PCs <u>only</u> when our data are normally distributed

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Markers of Principal Components

- Communalities (based on Eigenvalues): Percentage of variance accounted for by all factors
    - Similar to R^2 in Regression Analysis

- Loadings:
  Percentage of variance of each (initial) factor that is accounted for by each factor
    - Similar to β
        - > .5 → strong
        - > .3 → medium
        - < .3 → small

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Rotated Component Matrix

	Component					
	1	2	3	4	5	6
ghq25	,824	-,002	,157	,050	-,011	,142
ghq27	,779	,131	,194	,034	,137	,162
ghq28	,768	,206	,120	,011	,076	,203
ghq24	,722	,244	,089	,169	,220	,126
ghq23	,588	,362	,146	,309	,214	,003
ghq22	,531	,340	,227	,350	,188	-,092
ghq26	,463	,283	,264	,317	,140	-,143
ghq18	,127	,734	,246	-,040	,062	,183
ghq19	,287	,705	-,054	,171	,135	-,025
ghq20	,203	,686	,129	,024	,228	,035
ghq17	,128	,685	,261	,133	-,025	,127
ghq15	,104	,577	,012	,284	,050	-,052
ghq16	,020	,535	,110	-,022	,159	,280
ghq2	,154	,059	,681	,150	,188	-,217
ghq1	,153	,209	,658	,011	,126	,066
ghq3	,187	,055	,657	,180	,244	,177
ghq21	,120	,319	,534	,014	-,026	,261
ghq10	,116	,051	,470	,370	,100	,173
ghq12	,179	,141	-,004	,759	,147	,042
ghq14	,087	,130	,369	,612	,136	,146
ghq11	,030	,027	,174	,608	,010	,409
ghq13	,176	,170	,462	,465	,152	,126
ghq5	,102	,175	,174	,062	,854	,055
ghq6	,202	,170	,235	,167	,768	,077
ghq7	,142	,192	,104	,146	,440	,438
ghq4	,297	,032	,350	,230	,426	,135
ghq8	,273	,094	,156	,205	,204	,656
ghq9	,259	,330	,042	,274	-,015	,501

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

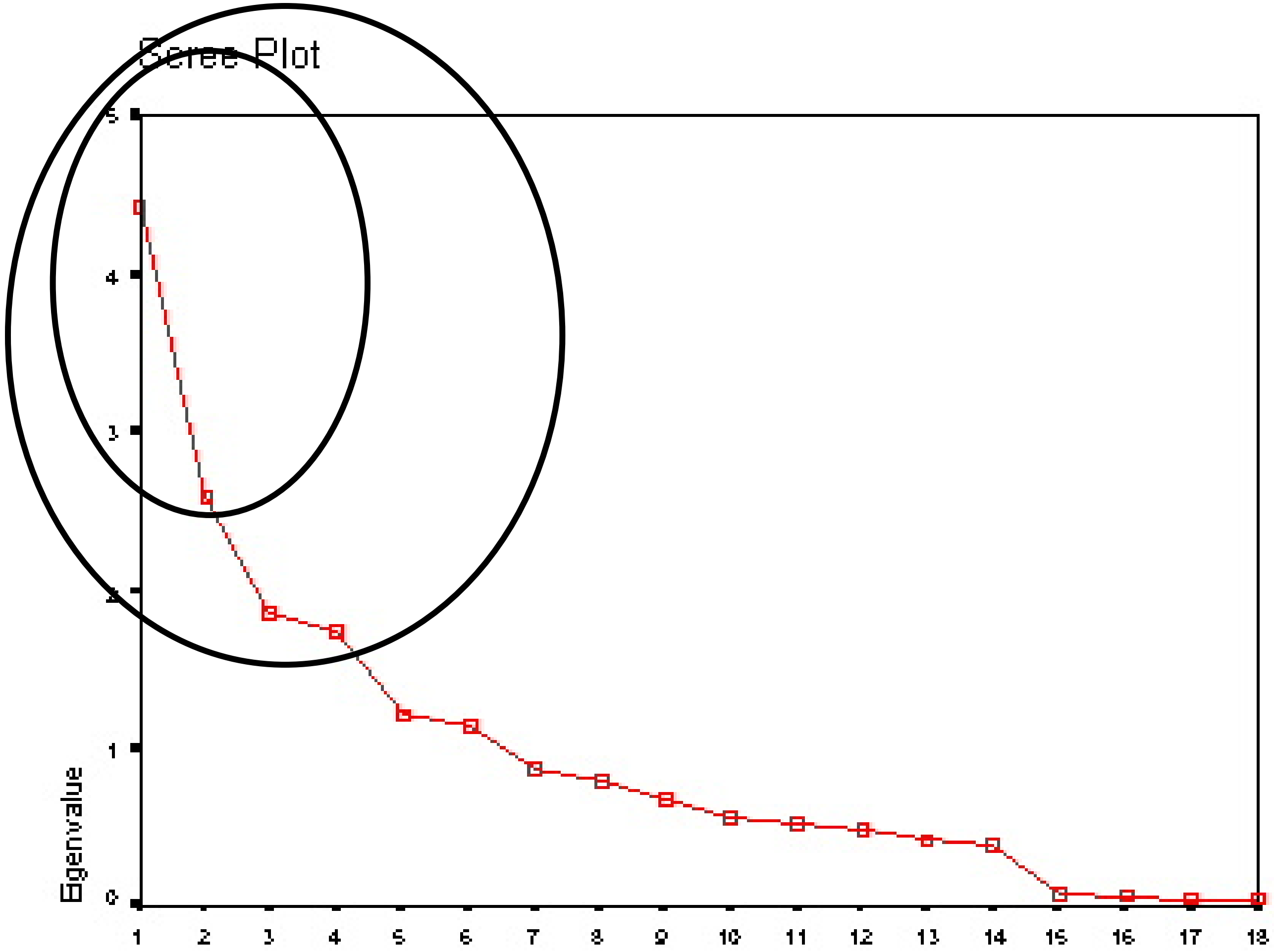
1st Important Question: How Many Principal Components?

- Retaining all factors (e.g., 12 distinct survey questions) perfectly expresses all our data - but we did not actually summarize.
- With very few variables (e.g., 1), significant pieces of information hidden within the total variance might be lost.

- So, which is the optimal number of principal components needed to summarize our data adequately? Alternative ways to decide:
  - Kaiser criterion: Statistical test (eigenvalues > 1)
  - Scree Plot: Where we have a decrease of the steepness of change in the gragh
  - As many needed to account for 70-80% of the total variance

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Scree Plot: One Method To Assess The Optimal Number of PCs

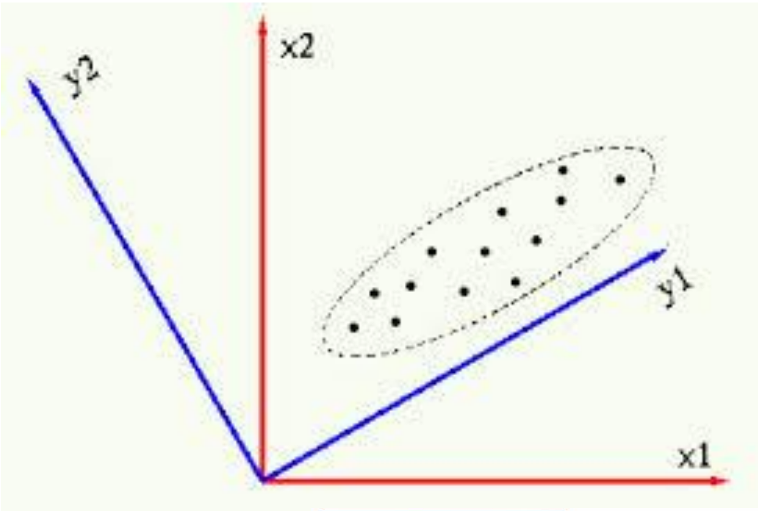


Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

2^{nd} Important Question: Interpretation of PCs

- Most often the next step is the **rotation** of the principle components
  - Rotations of axes XY
- Might significantly improve the interpretation



- ➢ Does not change the percentage of total variance accounted for,
- ➢ BUT the percentage of variance accounted for by each factor, therefore the identity/nature of each factor as well

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Usual Rotation Methods

- **No rotation**
  - Even though default option in SPSS, rarely selected. It is more or less always good to include a rotation.

- **Varimax rotation**
  - Simplifies PCs, maximizes loadings.
  - The most usual choice.
- **Quartimax rotation**
  - Simplifies PCs, minimizes the number of PCs needed to summarize each factor.
- **Equimax rotation**
  - Balanced intermediate of Varimax and Quartimax.

- Non orthogonal rotations… (e.g. “Direct Oblimin”, “Promax”)
  - More complicated
  - Pcs are not independent, might be partially correlated

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

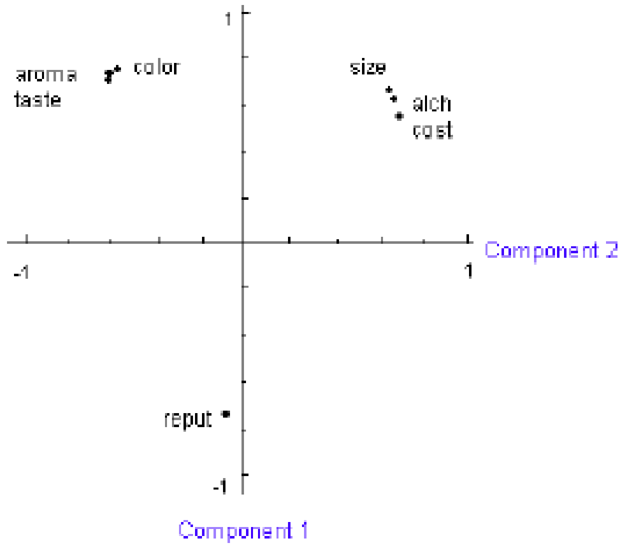
Example: Unrotated solution

Component Matrix[a]

	Component 1	Component 2
COLOR	.760	-.576
AROMA	.736	-.614
REPUTA	-.735	-.071
TASTE	.710	-.646
COST	.550	.734
ALCOHO	.632	.699
SIZE	.667	.675

Extraction Method: Principal Compone

a.2 components extracted.



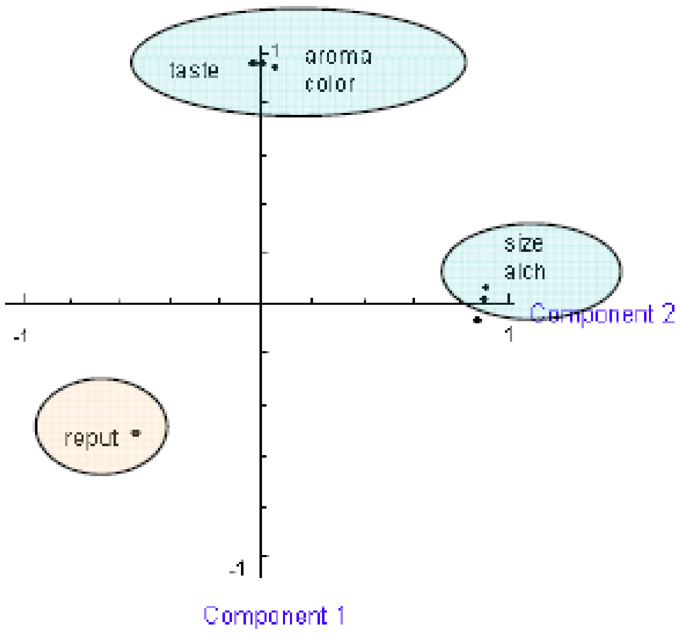
Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example: Solution after Rotation

Rotated Component Matrix[a]

	Component	
	1	2
TASTE	.960	-.028
AROMA	.958	1.E-02
COLOR	.952	6.E-02
SIZE	7.E-02	.947
ALCOHOL	2.E-02	.942
COST	-.061	.916
REPUTAT	-.512	-.533

Extraction Method: Principal Component Analy:
Rotation Method: Varimax with Kaiser Normaliz
a. Rotation converged in 3 iteratioı



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Rotations: Main Points

- No rotation needed with only one factor
- Orthogonal rotations are all mathematically similar, some create a result more easy to interpret
- Varimax: usually the best performing and easier to interpret, therefore the most extensively used

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Steps for PCA

1. **Compute Correlation Matrix**: Calculate correlations among variables.
2. **Build Full Model**: Include all factors initially.
3. **Decide on Retained Factors**: Determine how many factors to keep based on:
    - Total Explained Variance
    - Percentage of Variance Explained per Variable
    - Interpretability
    - Replicability
4. **PCs "rotation" and naming (interpretation)** (can be quite tricky sometimes)
5. **Computing "factor scores"** (replacing initial datapoints with the new datapoins of the reduced number of factors - PCs)
6. **"Apply" the selected solution** (number of PCs)
    - Theoretical implications – understanding of the reduced dataset
    - Statistical implications – we can use the new composite principle component scores in further statistical analyses

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

To Conclude…

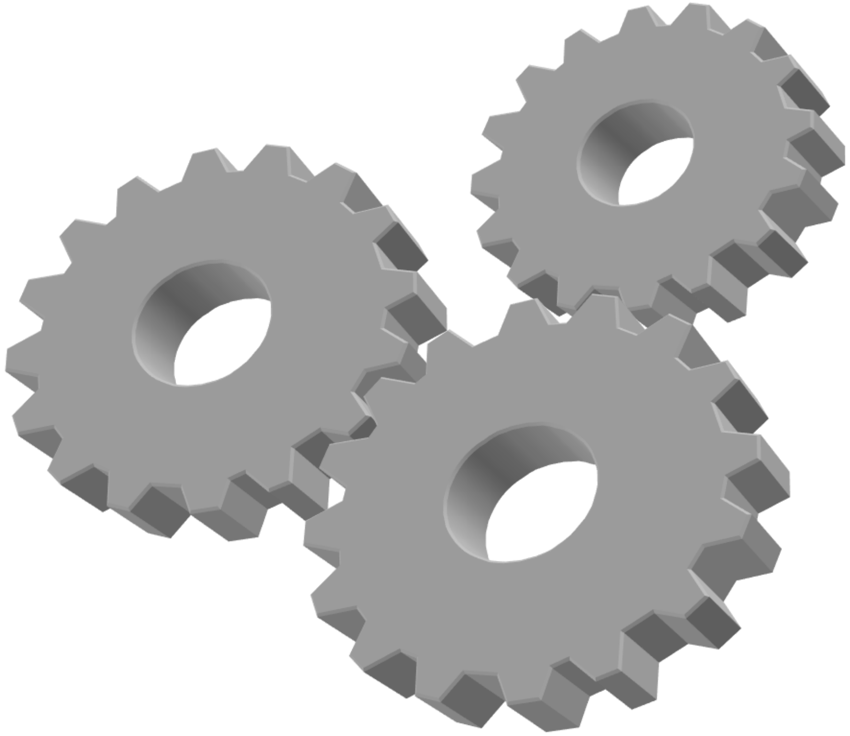
- We often have more quantity and depth in our data than we can intuitively comprehend

- Both FA and PCA summarize our data

- This option proves invaluable, enhancing our understanding and enabling clearer insights.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Factor Analysis (FA)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factor Analysis (FA) & Principal Components Analyis (PCA)

Suitable to reduce our data by using less variables

- Especially in Psychology, we often deal with difficult to measure phenomena or constructs (unlike direct measurement in positive sciences)
- Utilizing various measurements (e.g., survey questions).
  - All these parts might be components of a common factor, hidden behind them.

- *Alternatively:* describe the association between variables and confirm theories about these associations.

- **Objective:** Discover these common patterns, the underlying (hidden) factors expressing our various measurements.
- Find the smallest number of factors expressing the largest possible percentage of the initial variance.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Basic Principles

❑Two highly correlated variables likely represent the same phenomenon.

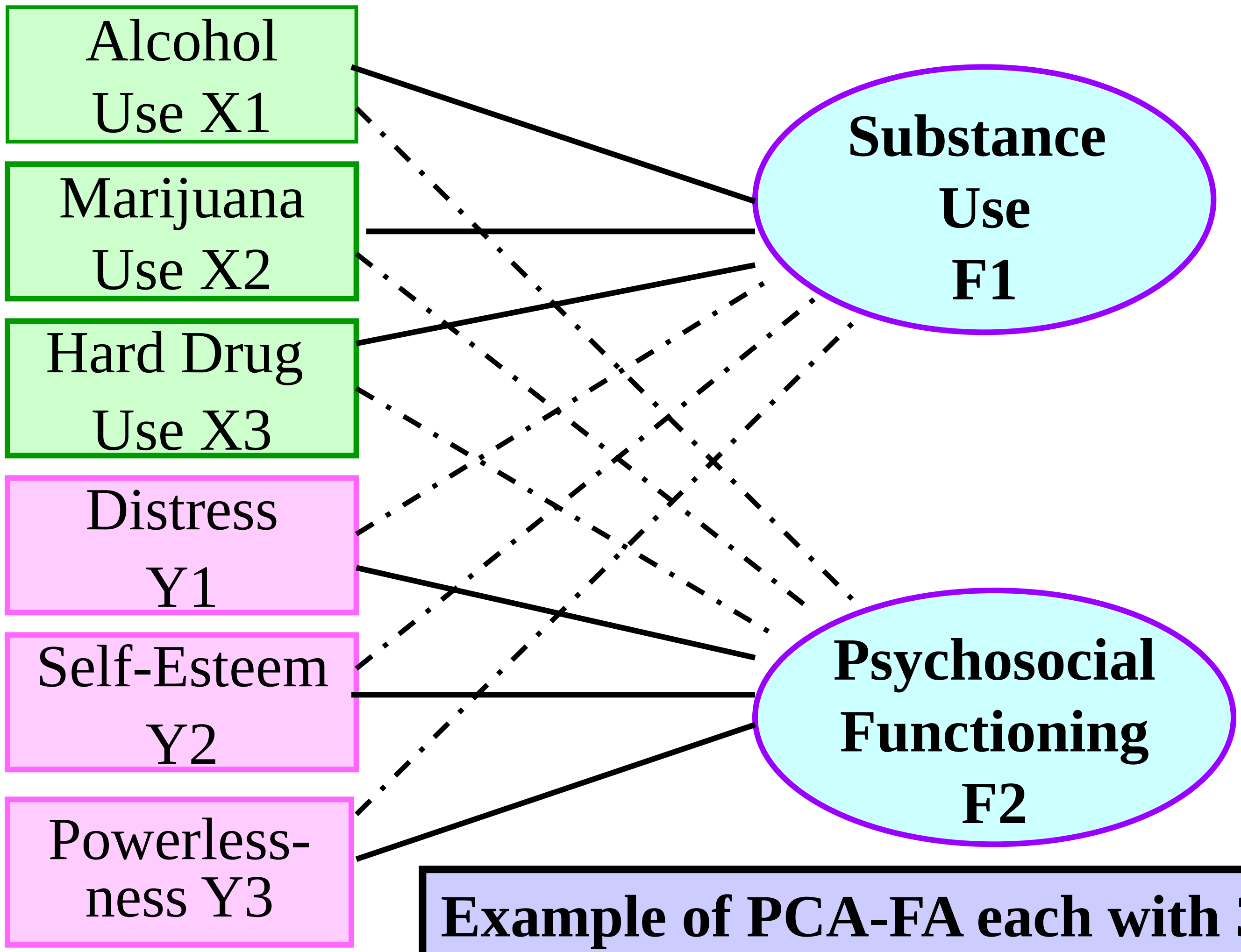
- Combining them into a factor simplifies the studied phenomenon and reduces measurement errors/noise.
  - E.g. years of service in a job (e.g., mail carrier) and the number of delivered letters
  - → Professional Experience: The type of "factor" or "principal component" we aim to discover.

➢Different phrasing of our objective::

- **To summarize** relationships among a large number of variables concisely and accurately, so as to **make a concept or quality more understandable**.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Example of PCA-FA each with 3 main (bold-faced) loadings and each with 3 inconsequential (dashed-line) loadings.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factor Analysis

❑Types:

- Exploratory Factor Analysis (EFA)
- Confirmatory Factor Analysis (CFA)

- Assumptions
  - Normally distributed data
  - Variables measured in ratio scale
  - Correlation among variables sufficient ($r > .20$) but not too high ($r < .80$)
  - Linear relationship between the variables (linearity)
  - No outliers

- Big enough sample necessary (Comrey & Lee, 1992)
  - 50 cases is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or more is excellent.
  - As a rule of thumb, a bare <u>minimum of 10 observations per variable</u> is necessary to avoid computational difficulties.
- Variables = (3x to 5x) × Factors

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Exploratory Factor Analysis (EFA)

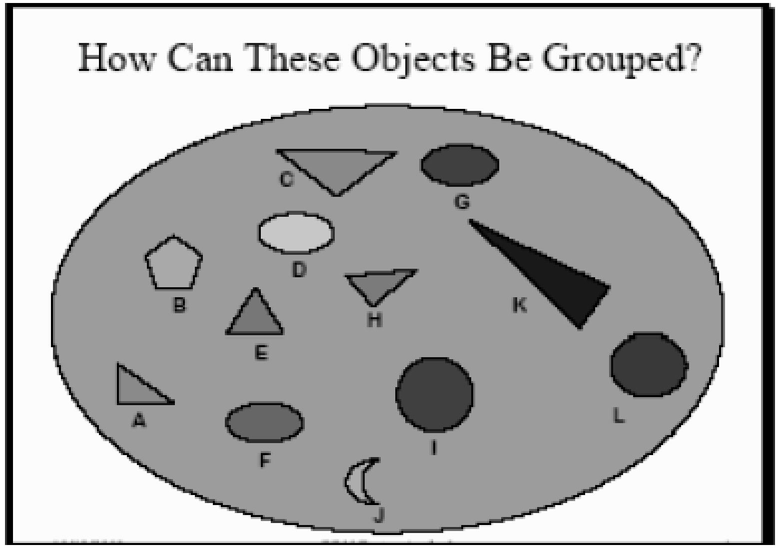
- Used only for the initial exploration and summary of a set of variables, achieved through their grouping in composite factors
  - Goal is to analyze only the percentage of common variance among the variables
  - Useful when our goal is to *create/construct* factors

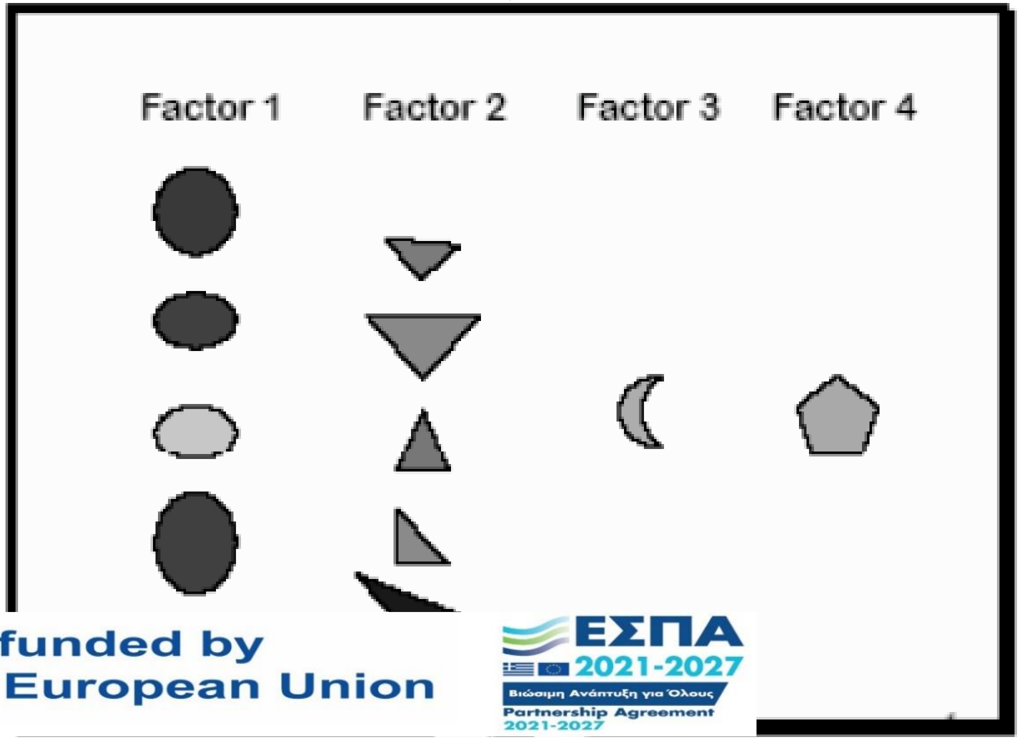
- Introduced by Charles Spearman at the beginning of the 20th century

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example





Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Confirmatory Factor Analysis (CFA)

- Test theory-based hypotheses

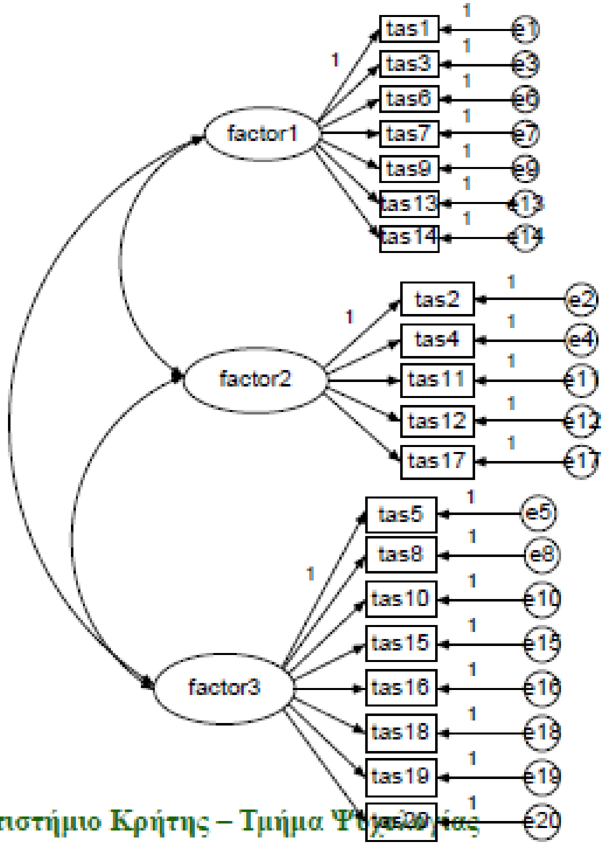
- Useful for examining whether a **hypothesized pattern/matrix of associations** between some variables (correlations) **is verified/supported** by our data

- Example…

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

EXAMPLE



Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Step to Conduct FA (Similar to PCA)

1. **Compute Correlation Matrix**
   - Note: The main difference is that here we only take into account common variance among the variables
2. **Choose the extraction method**
   - Principal factor analysis (also called Principal axis FA), maximum likelihood, generalized least squares, unweighted least squares
3. **Decide the Number of Factors to keep**, based on:
      - Total Explained Variance
      - Percentage of Variance Explained per Variable
      - Interpretability
      - Replicability
4. **Factor “rotation” and interpretation**
5. **Computing “factor scores”** (replacing initial datapoints with the new datapoins of the reduced number of factors)
6. **“Apply” the selected solution** (number of factors)
   - Theoretical implications – understanding of the reduced dataset
   - Statistical implications – we can use the new composite factor scores in further statistical analyses

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

FA in SPSS –1st Step

- The quality of our data (relevant to the adherence to the assumptions presented earlier) can be assessed in SPSS using:
  - **Keiser-Meyer-Olkin** value
    - Sample sufficiency (**>.50**), the closer to 1 the better
  - **Bartlett's Test of Sphericity** value
    - Do the correlations between variables allow the use of FA? (**p <0.05**)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,951
Bartlett's Test of Sphericity	Approx. Chi-Square	44323,516
	df	1540
	Sig.	,000

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

FA in SPSS – 2nd STEP

- **Correlation Matrix**

	1	2	3	4	5
1	1.00				
2	**0.87**	1.00			
3	0.04	0.11	1.00		
4	0.06	0.10	**0.51**	1.00	
5	0.14	0.08	**0.61**	**0.49**	1.00

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

FA & Differences to PCA

- In many books/articles, PCA (Principal Component Analysis) and EFA (Exploratory Factor Analysis) are described together.
- Even in SPSS, they are included under the Factor Analysis section.

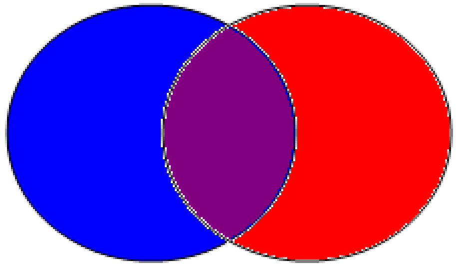
- However, they are not the same; their logic is different.
  - **PCA**: attempts to account for *<u>all the variance</u>*
  - **EFA**: accounts *<u>only for the common variance</u>* among the variables

Co-funded by the European Union

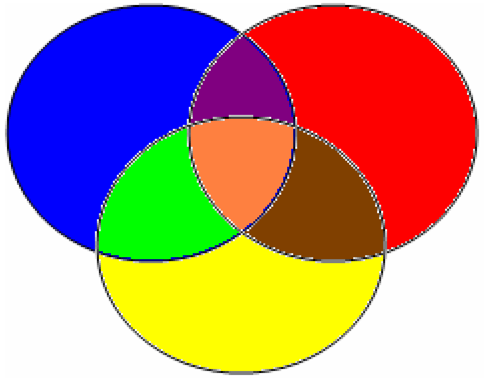
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

PCA vs. EFA

- PCA
  - data = variance (+error)
  - Error is considered to be equal for all measurements



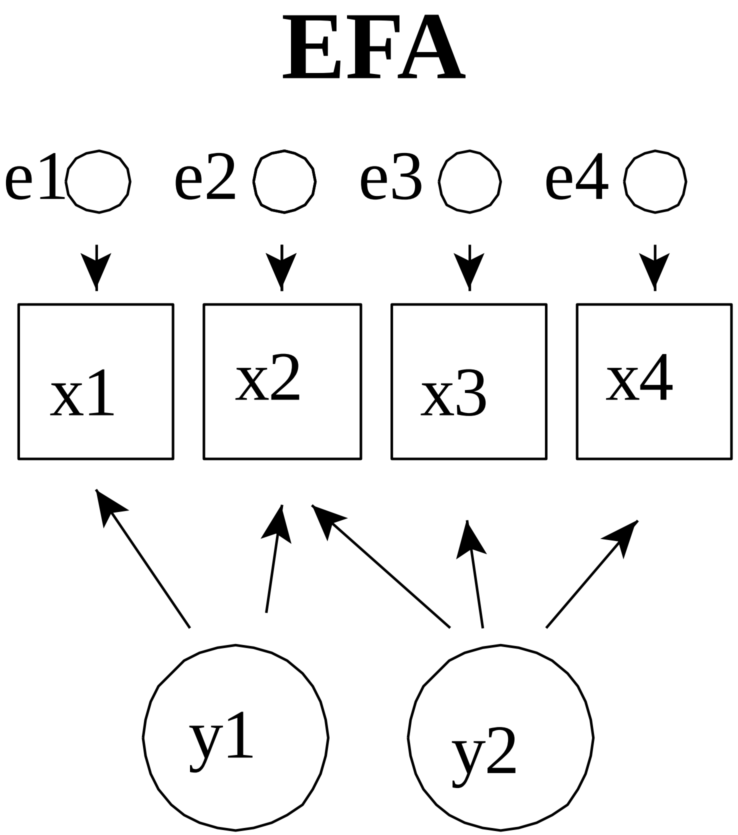
- EFA
  - data = common variance + specific variance + error
  - The specific variance (that is NOT common with other variables) might differ from variable to variable

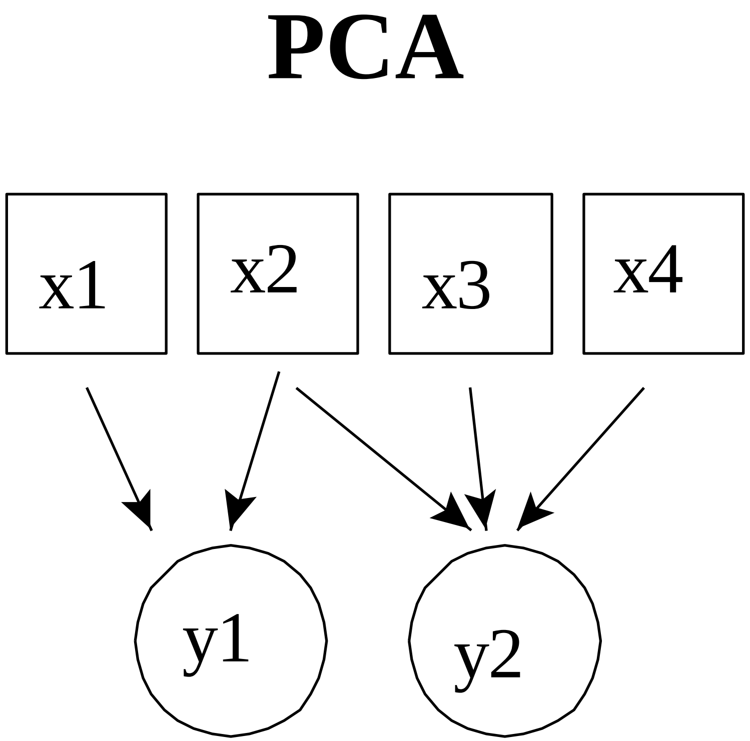


Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

PCA vs. EFA Path Diagram: Exploratory factor analysis





Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factors VS. Principal Compenents

- **Factors** are real/existing **hidden variables**, that cause the covariance of the measured variables

- **Principal Components** are **empirically defined sums of variables**, without the necessity of a theoretical framework supporting their hypothesized existence.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

PCA and EFA

- EFA or PCA?

- When variance is similar for all variables, both methods produce similar results
- If we analyze a large number of variables, both methods will produce similar results

- If every variable has a different level of noise (**unequal variances**) it is preferable to use **Factor Analysis (EFA)**

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

FA Στο SPSS

- Choose:
  - Analyze → Data reduction → Factor

- Input the variables of interest
  - Extraction method: Principal Components (or other…)
  - Define the method of Extraction (e.g. “Based on Eigenvalue”)
  - Tick: Correlation matrix, scree plot
  - Select rotation method (usually Varimax)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example

- Rate how much you enjoy the following music genres:
    - Country
    - Blues
    - Classical
    - Folk
    - Jazz
    - Opera
    - Rap
    - Heavy Metal

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Correlation matrix

Correlation Matrix

		Country Western Music	Blues or R & B Music	Classical Music	Folk Music	Jazz Music	Opera	Rap Music	Heavy Metal Music
Correlation	Country Western Music	1.000	.035	-.096	.215	-.107	-.014	-.034	-.070
	Blues or R & B Music	.035	1.000	.209	.167	.556	.217	.177	.107
	Classical Music	-.096	.209	1.000	.410	.283	.600	.016	.002
	Folk Music	.215	.167	.410	1.000	.111	.324	-.058	-.048
	Jazz Music	-.107	.556	.283	.111	1.000	.246	.176	.098
	Opera	-.014	.217	.600	.324	.246	1.000	.104	-.005
	Rap Music	-.034	.177	.016	-.058	.176	.104	1.000	.346
	Heavy Metal Music	-.070	.107	.002	-.048	.098	-.005	.346	1.000
Sig. (1-tailed)	Country Western Music		.110	.000	.000	.000	.315	.112	.007
	Blues or R & B Music	.110		.000	.000	.000	.000	.000	.000
	Classical Music	.000	.000		.000	.000	.000	.282	.469
	Folk Music	.000	.000	.000		.000	.000	.020	.046
	Jazz Music	.000	.000	.000	.000		.000	.000	.000
	Opera	.315	.000	.000	.000	.000		.000	.423
	Rap Music	.112	.000	.282	.020	.000	.000		.000
	Heavy Metal Music	.007	.000	.469	.046	.000	.423	.000	

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Commonalities

Communalities

	Initial	Extraction
Country Western Music	.105	.381
Blues or R & B Music	.337	.565
Classical Music	.449	.841
Folk Music	.251	.384
Jazz Music	.358	.602
Opera	.386	.448
Rap Music	.162	.515
Heavy Metal Music	.129	.232

Extraction Method: Principal Axis Factoring.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Variance explained

Total Variance Explained

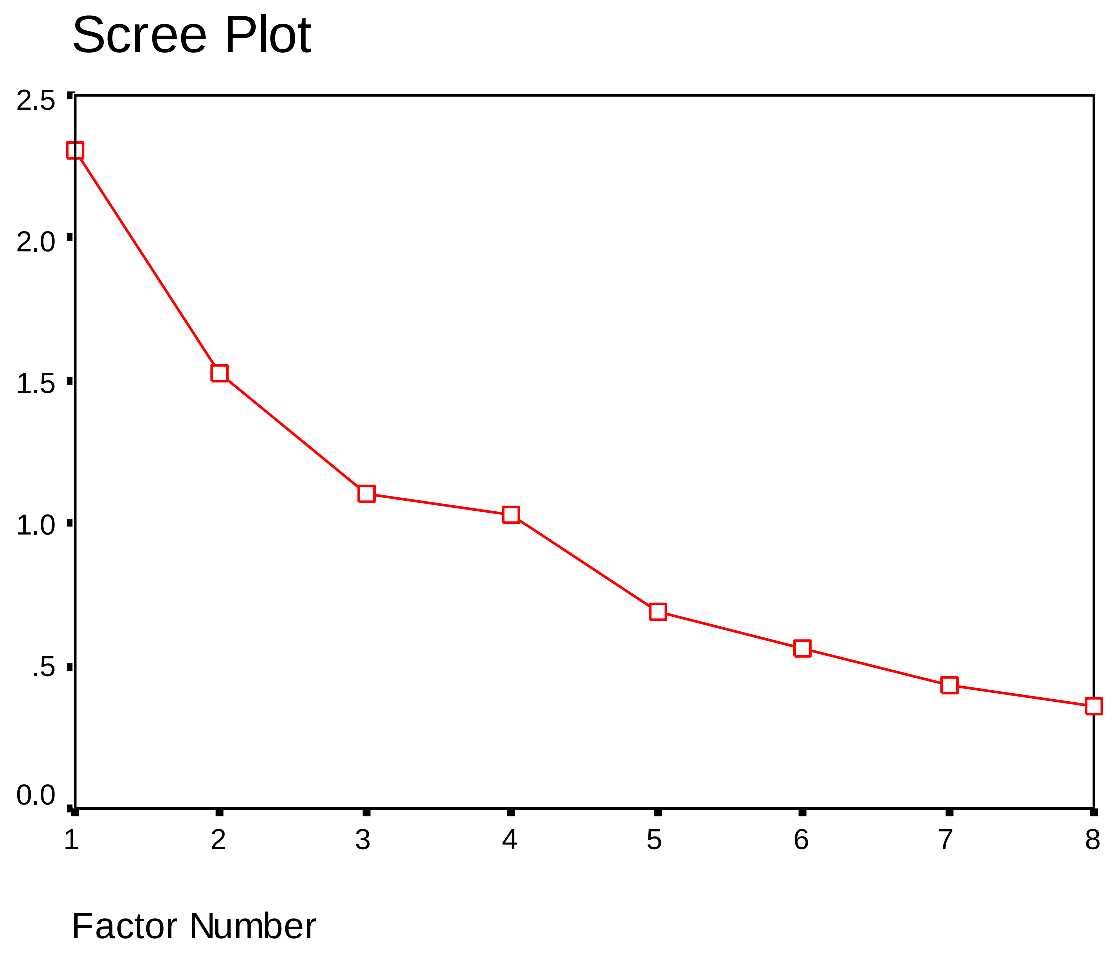
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.308	28.854	28.854	1.909	23.868	23.868	1.514	18.930	18.930
2	1.522	19.027	47.881	.997	12.461	36.329	1.131	14.139	33.069
3	1.099	13.742	61.623	.571	7.140	43.469	.772	9.647	42.716
4	1.029	12.867	74.490	.493	6.159	49.627	.553	6.911	49.627
5	.688	8.594	83.084						
6	.564	7.055	90.139						
7	.428	5.353	95.492						
8	.361	4.508	100.000						

Extraction Method: Principal Axis Factoring.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Scree Plot



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

The raw factors extracted

Factor Matrix[a]

	Factor			
	1	2	3	4
Country Western Music	-.033	-.170	.427	.412
Blues or R & B Music	.558	.351	.350	-.089
Classical Music	.786	-.377	-.284	-.030
Folk Music	.432	-.338	.184	.223
Jazz Music	.599	.370	.196	-.262
Opera	.617	-.208	-.137	.074
Rap Music	.209	.520	-.240	.379
Heavy Metal Music	.102	.371	-.193	.217

Extraction Method: Principal Axis Factoring.

a. Attempted to extract 4 factors. More than 25 iterations required. (Convergence=.006). Extraction was terminated.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

And when they have been tidied up by varimax

Rotated Factor Matrix[a]

	Factor			
	1	2	3	4
Country Western Music	-.019	-.028	-.046	.615
Blues or R & B Music	.139	.719	.134	.103
Classical Music	.898	.139	-.021	-.123
Folk Music	.485	.104	-.094	.360
Jazz Music	.187	.734	.113	-.126
Opera	.646	.162	.066	.011
Rap Music	.025	.118	.708	-.009
Heavy Metal Music	-.019	.064	.474	-.058

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

And how you convert between raw and cleaned up factors

Factor Transformation Matrix

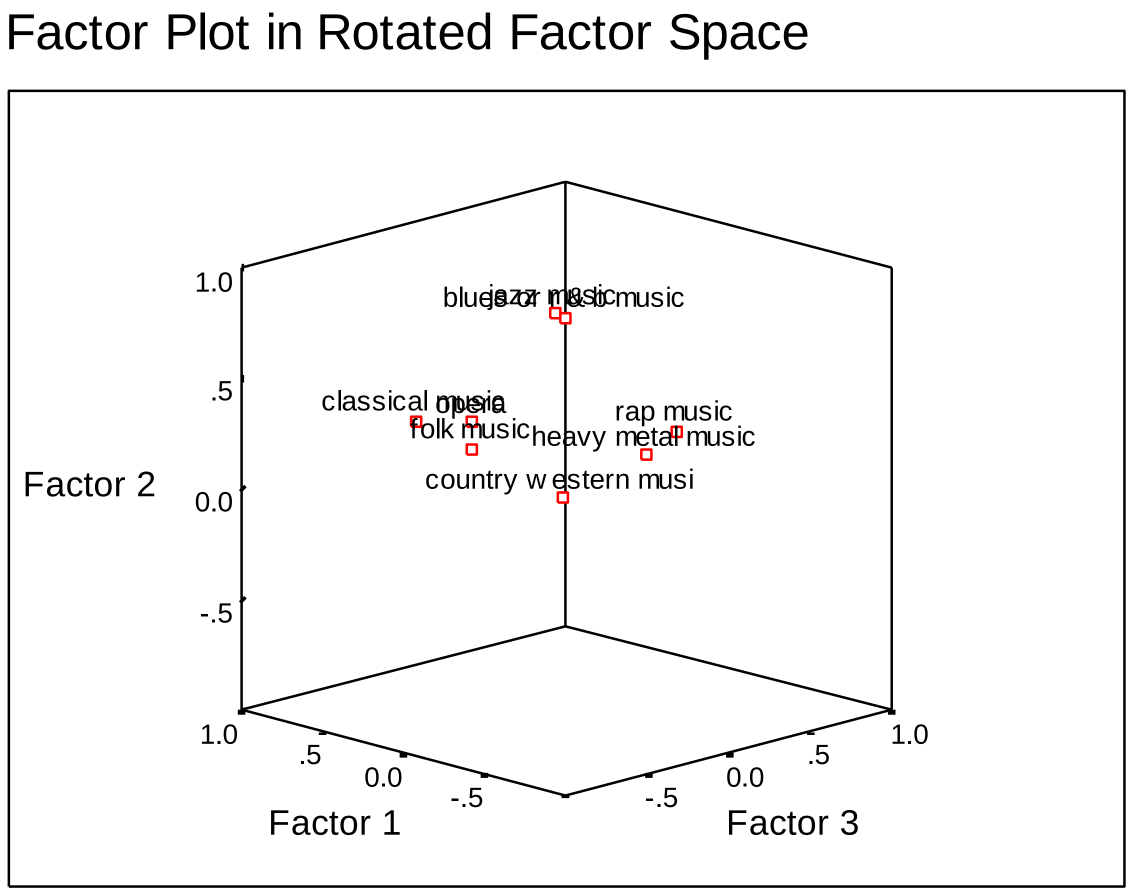
Factor	1	2	3	4
1	.789	.590	.169	.010
2	-.511	.495	.669	-.218
3	-.314	.525	-.407	.678
4	.133	-.362	.599	.702

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Factor Plot (useful?? …but can be rotated)



Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

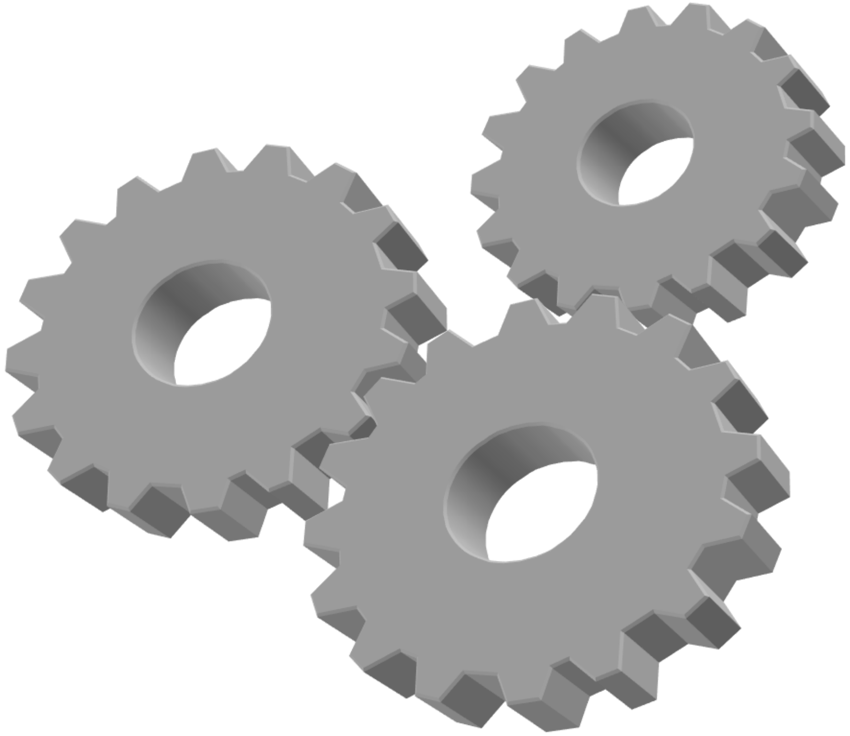
To CONCLUDE

- We often have more quantity and depth in our data than we can intuitively comprehend

- Both FA and PCA summarize our data
  - There are significant similarities but also differences between the two methods
  - Choice of the most appropriate option depending on our goals
  - The difference is smaller when we have large datasets and big numbers of variables

- This option proves invaluable, enhancing our understanding and enabling clearer insights.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Cluster Analysis (CA)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

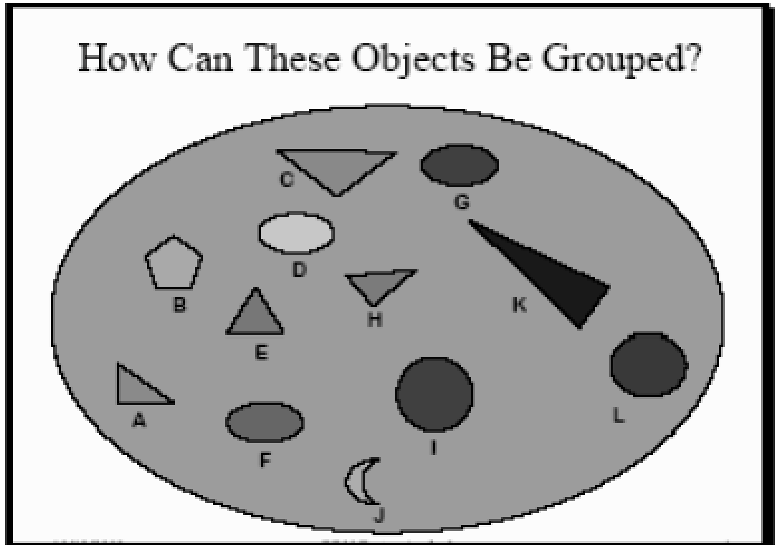
What is Cluster Analysis (CA)?

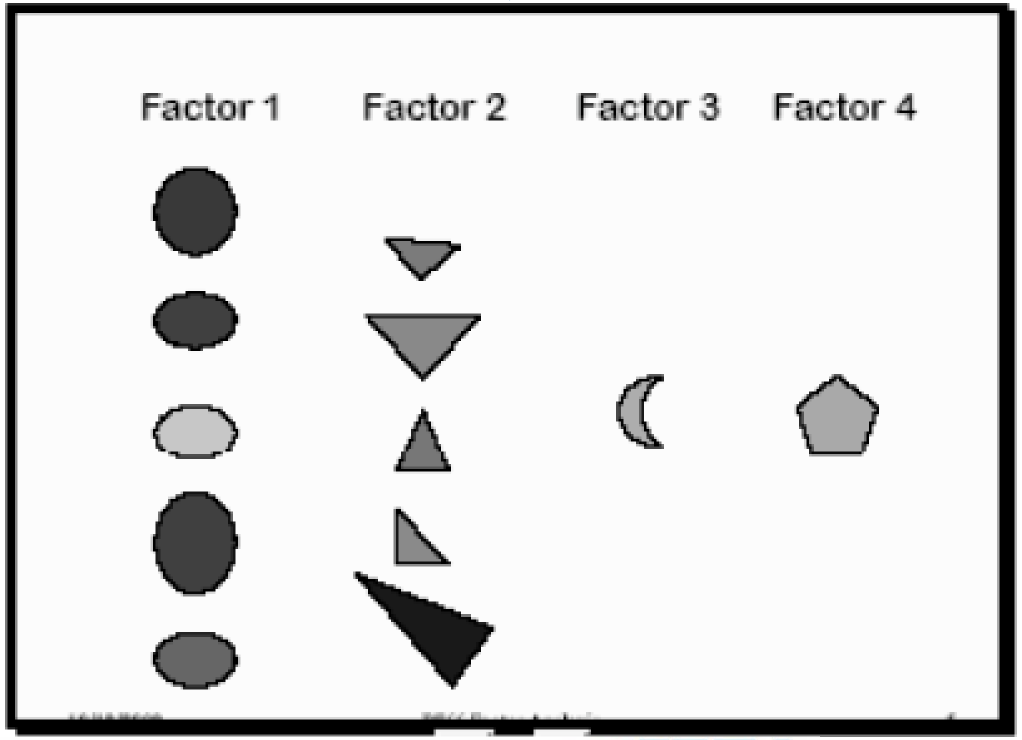
- Another exploratory method aiming to identify homogenous groups of cases, for groupings not previously known (a posteriori grouping, on the basis of how closely associated they are)
  - Typically used when no assumption about the likely relationships within the data
  - Can handle binary, nominal, ordinal, and interval and ratio scale data
  - Often used in conjunction with other analyses (such as discriminant analysis).
  - Recently often used in machine learning, data mining, and big data analysis.
- The primary value of cluster analysis lies in the classification of data, as suggested by "natural" groupings of the data themselves.
  - Cluster analysis is comparable with factor analysis in its objective of assessing structure; cluster analysis differs from factor analysis in that cluster analysis groups objects, whereas factor analysis is primarily concerned with grouping variables.

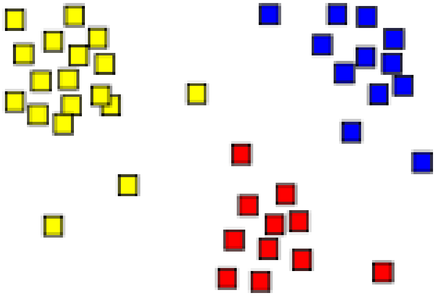
Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example







Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Use / Interpretation of Cluster Analysis (CA)

- Interpretation of the researcher is important, to decide if the result is meaningful
  - An process of knowledge discovery or interactive multi-objective optimization that involves trial and failure
  - “Clustering is in the eye of the beholder" (Estivill-Castro, V., 2002)

- Typical applications
  - As a stand-alone tool to get insight into data distribution
  - As a preprocessing step for other algorithms

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

CA Applications in Social Sciences

- Cluster analysis was originated in anthropology by Driver and Kroeber in 1932[1] and introduced to psychology by Joseph Zubin in 1938[2] and Robert Tryon in 1939[3] and famously used by Cattell beginning in 1943[4] for trait theory classification in personality psychology.
  (Source: Wikipedia)

- EXAMPLE: Identify student groups that need special attention
  - Researchers may measure psychological, aptitude, and achievement characteristics. A cluster analysis then may identify what homogeneous groups exist among students

  - Potential Clusters:
    - high achievers in all subjects
    - students that excel in certain subjects but fail in others

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Different Types (Methods) for CA

- 3 methods for CA:
  - ***K-means cluster*** - quickly cluster large data sets.
    - Define the number of clusters in advance. Useful to test different models with a different assumed number of clusters.
  - ***Hierarchical cluster*** - the most common.
    - Series of models: *1* (all cases in one cluster) to *n* (each case is an individual cluster). Can also cluster variables together, like FA. Can handle nominal, ordinal, and scale data; however it is not recommended to mix different levels of measurement.
  - ***Two-step cluster* analysis** 1) pre-clustering and 2) hierarchical methods. Combination of the above 2 methods.
    - Can handle large data sets that would take a long time to compute with hierarchical cluster methods. Can handle scale and ordinal data in the same model, automatically selects the number of clusters.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

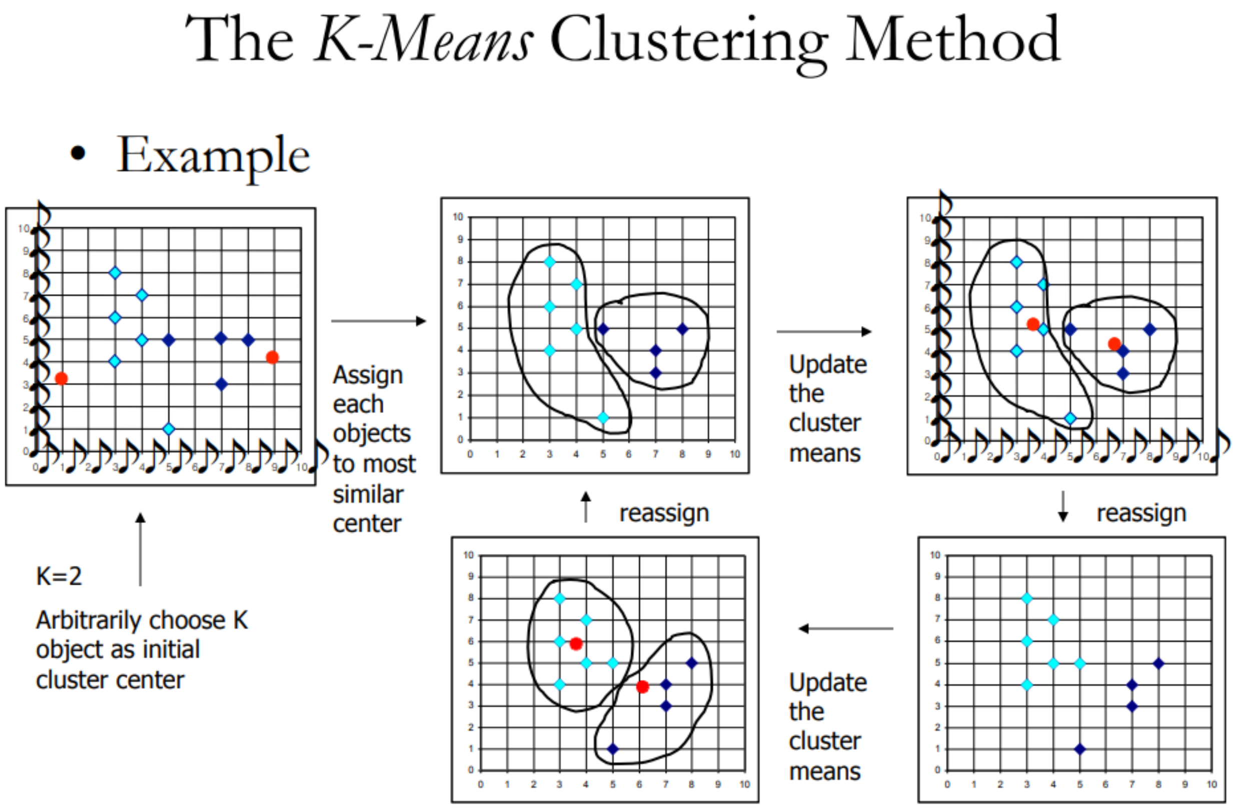
Organizational Principles of Different Types (Methods) for CA

- 3 methods for CA:
  - ***K-means cluster***.
    - Centroid-based clustering: each cluster represented by a central vector, not necessarily a member of the data set.
    - *k* cluster centers, squared distances from the cluster are minimized
    - Usually require *k* to be specified in advance
    - Prefer clusters of approximately similar size, might incorrectly cut borders of clusters (which is not surprising since the algorithm optimizes cluster centers, not cluster borders).
  - ***Hierarchical cluster***.
    - Connectivity-based clustering: objects more related to nearby objects than to objects farther away
    - Extensive hierarchy of clusters that merge with each other at certain distances.
    - Potential for "chaining phenomenon": outliers will either show up as additional clusters or even cause other clusters to merge
  - ***Two-step cluster* analysis**
    - Combination of the above 2 methods.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

K-Means Clustering



***K-means cluster*.**

+

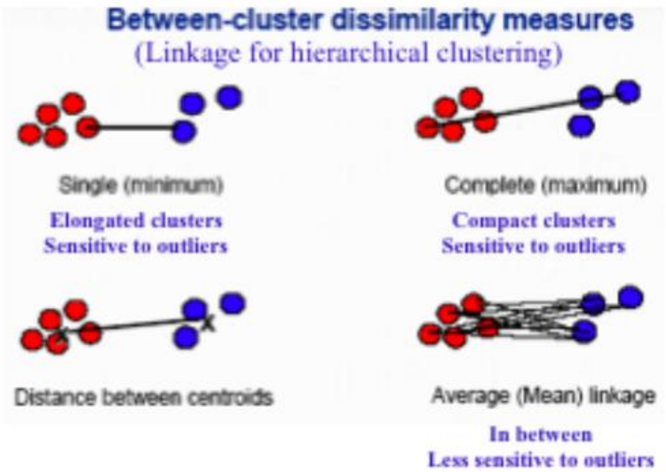
- relatively efficient and easy

-

- *not suitable for categorical data*
- *k* needs to be specified in advance
- bad with noise / outliers

Hierarchical Clustering

Distance Between Clusters



- Hierarchical algorithms can be agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms begin with each element as a separate cluster and merge them in successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

➢ Iterative Procedure

1. Randomly generate *k* clusters and determine the cluster centers, or directly generate *k* seed points as cluster centers.
2. Assign each point to the nearest cluster center.
3. Recompute the new cluster centers.
4. Repeat until some convergence criterion is met (usually that the assignment hasn't changed).

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Single Linkage Method: Nearest Neighbor

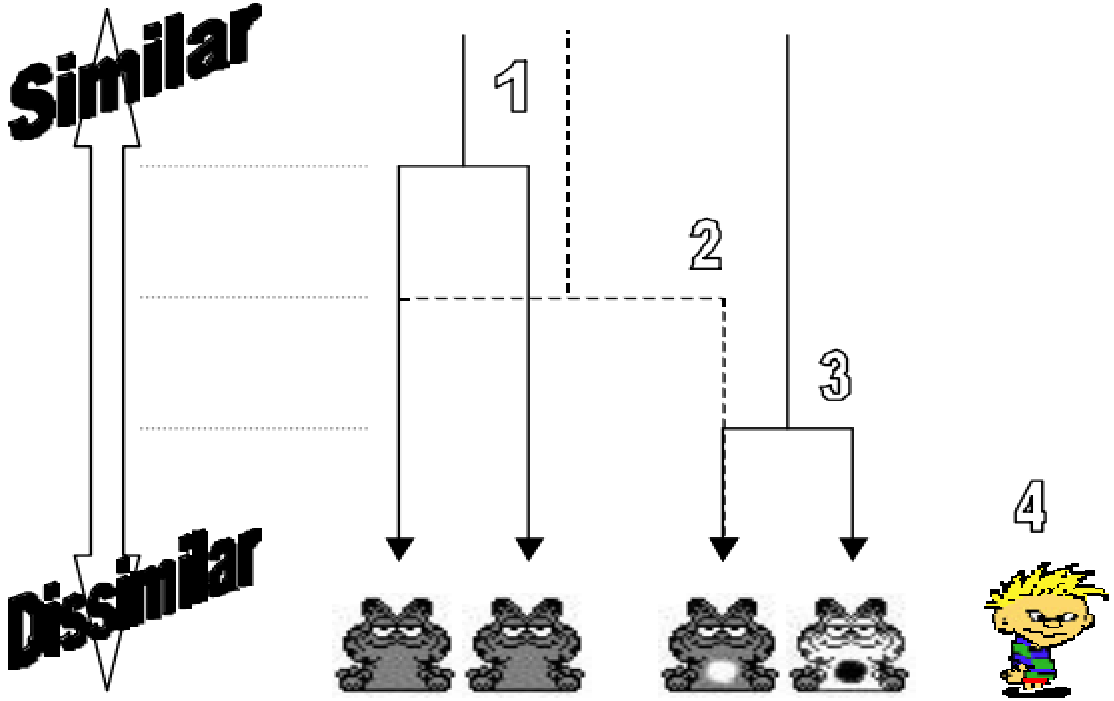


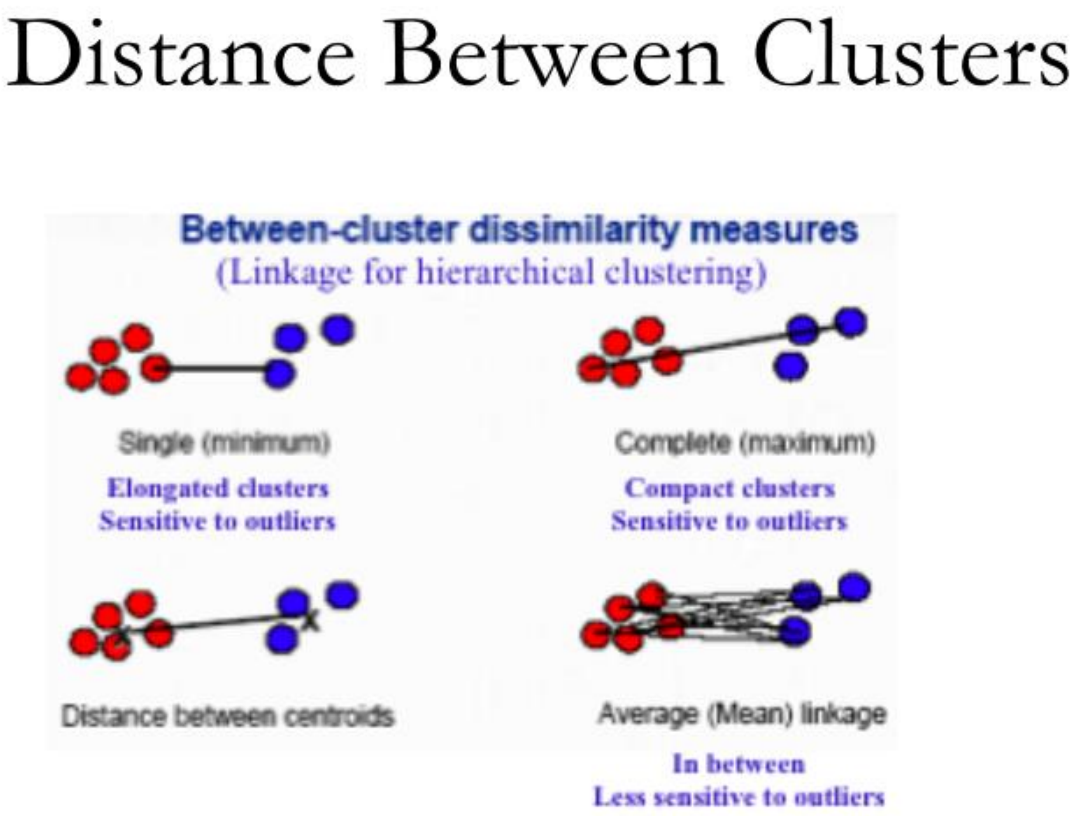
Figure 3

Figure 3 shows how the simple linkage method works. If we measured 5 animals on their physical characteristics (colour, number of legs, eyes etc.) and wanted to cluster these animals based on these characteristics we would start with the two most similar animals. First, imagine

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Hierarchical Clustering



Hierarchical cluster.

+

- Simple and fast
- Can handle nominal, ordinal, and scale data
- Apart from cases, it can also cluster variables together, like FA.

-

- Does not yield the same result with each run, since the resulting clusters depend on the initial random assignments.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

CA in SPSS

- Choose:
  - Analyze → Classify→
    - Two-step cluster
    - K-Means Cluster
    - Hierarchical Cluster

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example

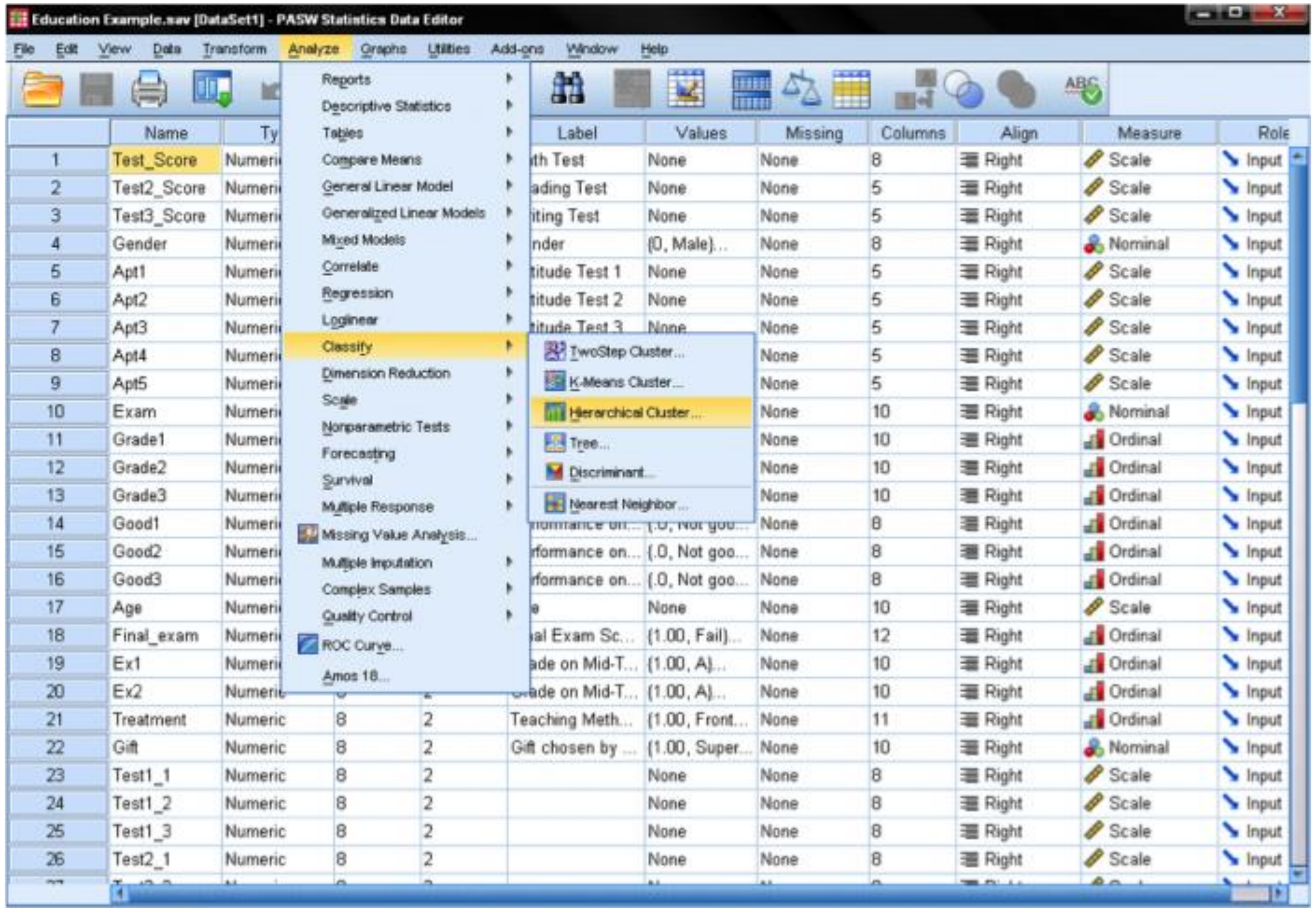
- Clusters of students emerging based on standardized test scores in mathematics, reading, and writing.

- Use hierarchical cluster analysis.
- 3 basic steps:
  1. calculate the distances,
  2. link the clusters, and
  3. choose a solution by selecting the right number of clusters.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

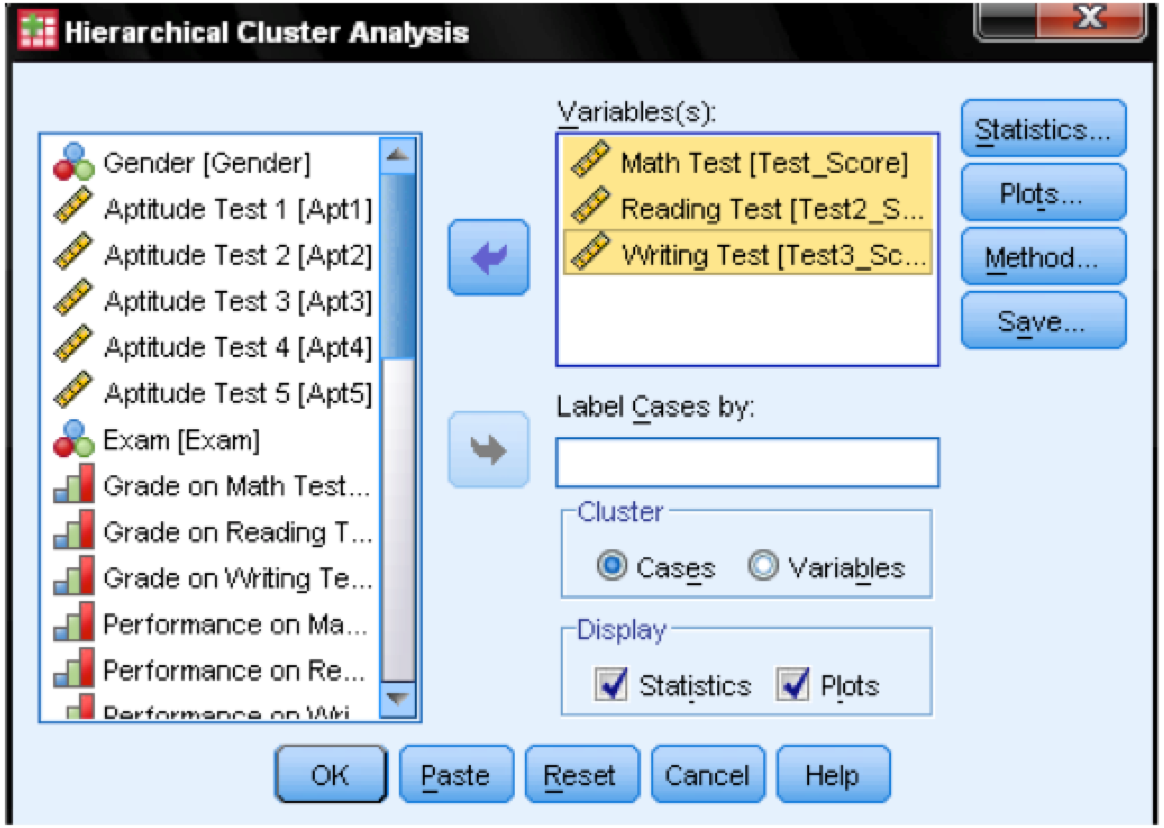
CA in SPSS



Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

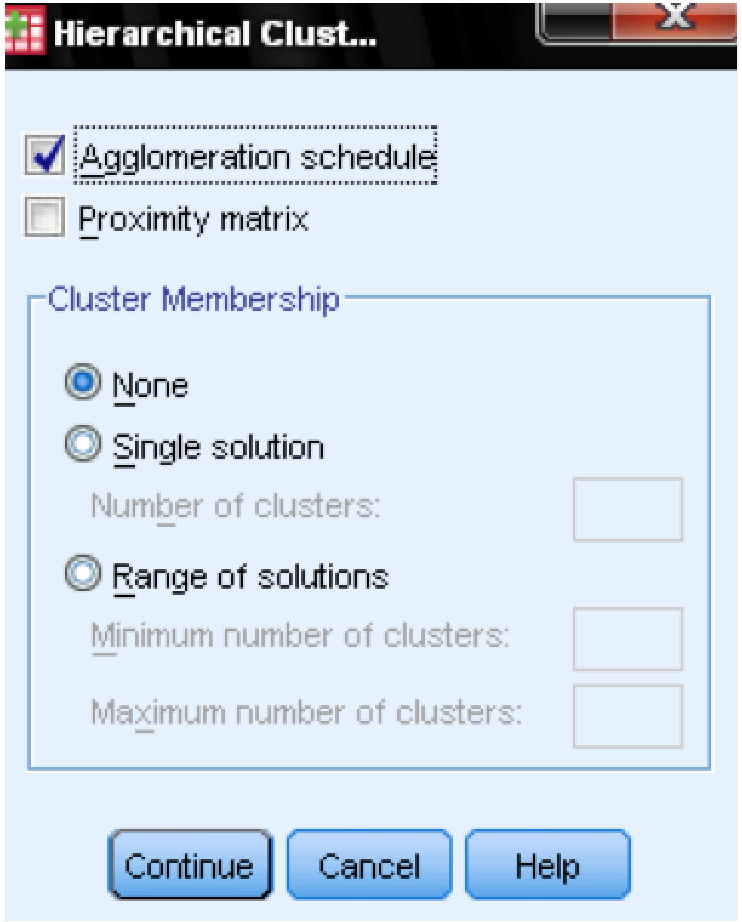
Input Variables of Interest



- Note
- Choice of Clustering:
  - Cases
  - Variable

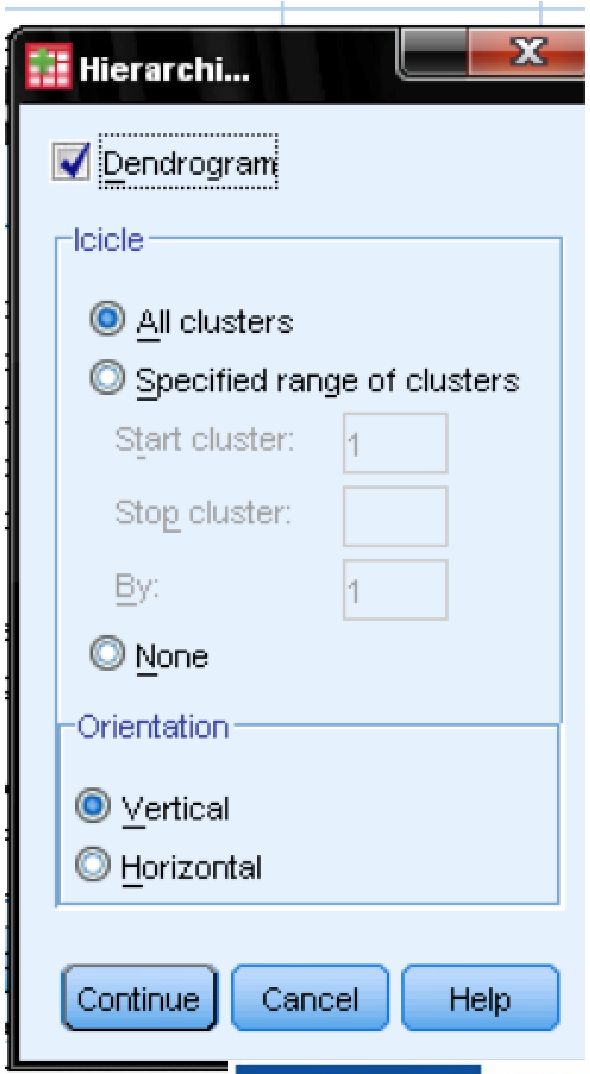
Co-funded by the European Union
ΕΣΠΑ 2021-2027
Partnership Agreement 2021-2027

Sub-Menu: Statistics



- Proximity matrix = the distances calculated in the first step of the analysis
- Predicted cluster membership of the cases in our observations.
- Can also ask for specific number of clusters or a range of solutions

Sub-Menu: Plots

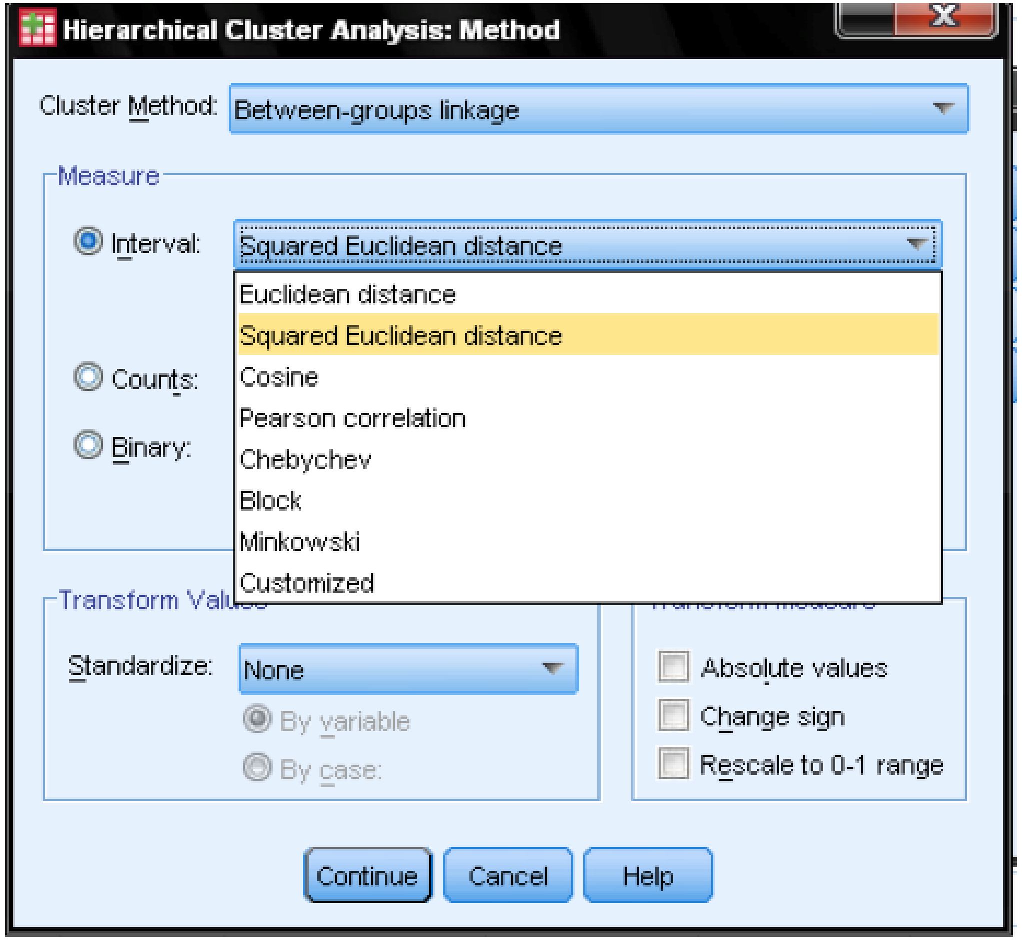


- Choose *Dendrogram*
  - graphically shows how the clusters are merged
  - allows us to identify what the appropriate number of clusters is

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sub-Menu: Method I

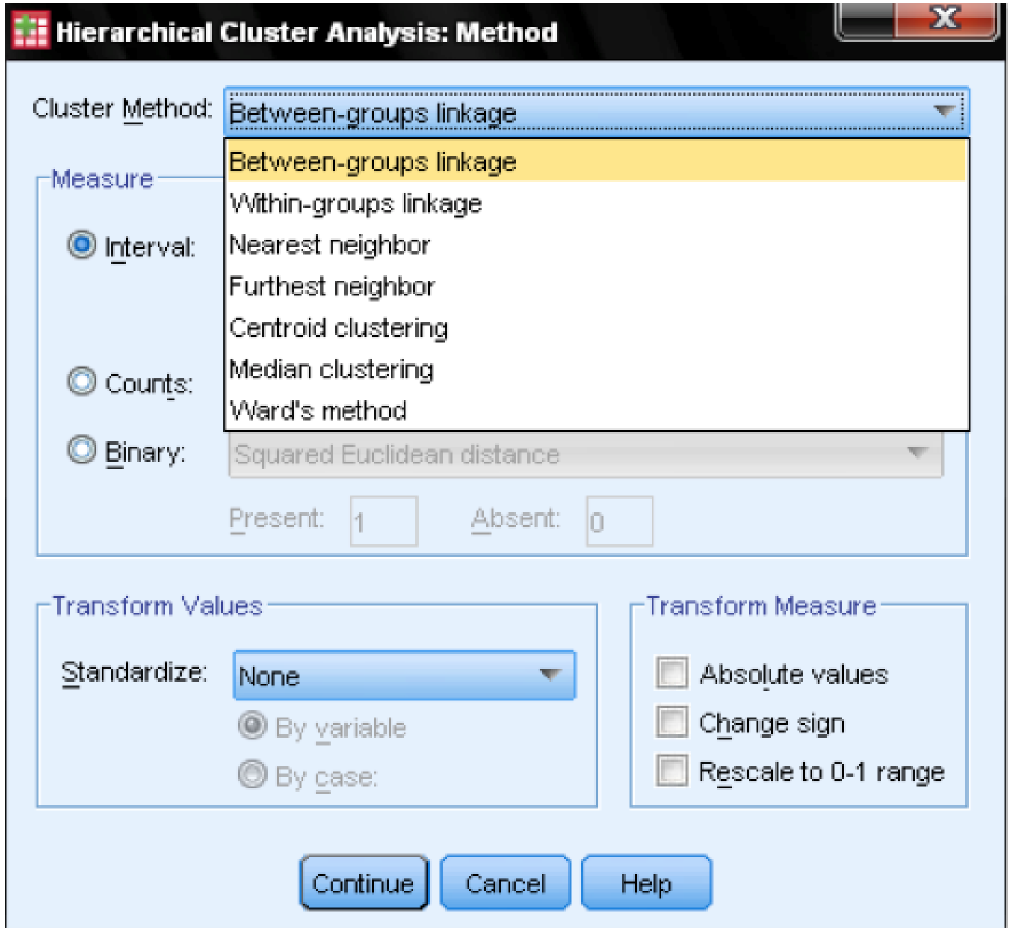


- Choose distance measure
  - *Interval* (scale)
  - *Counts* (ordinal)
  - *Binary* (nominal)
- And estimation method
  - Squared Euclidean is a popular choice for interval and binary data
  - Chi-squared / Phi-Squared (standardized version of Chi-Squared) is a popular choice for counts data

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sub-Menu: Method II



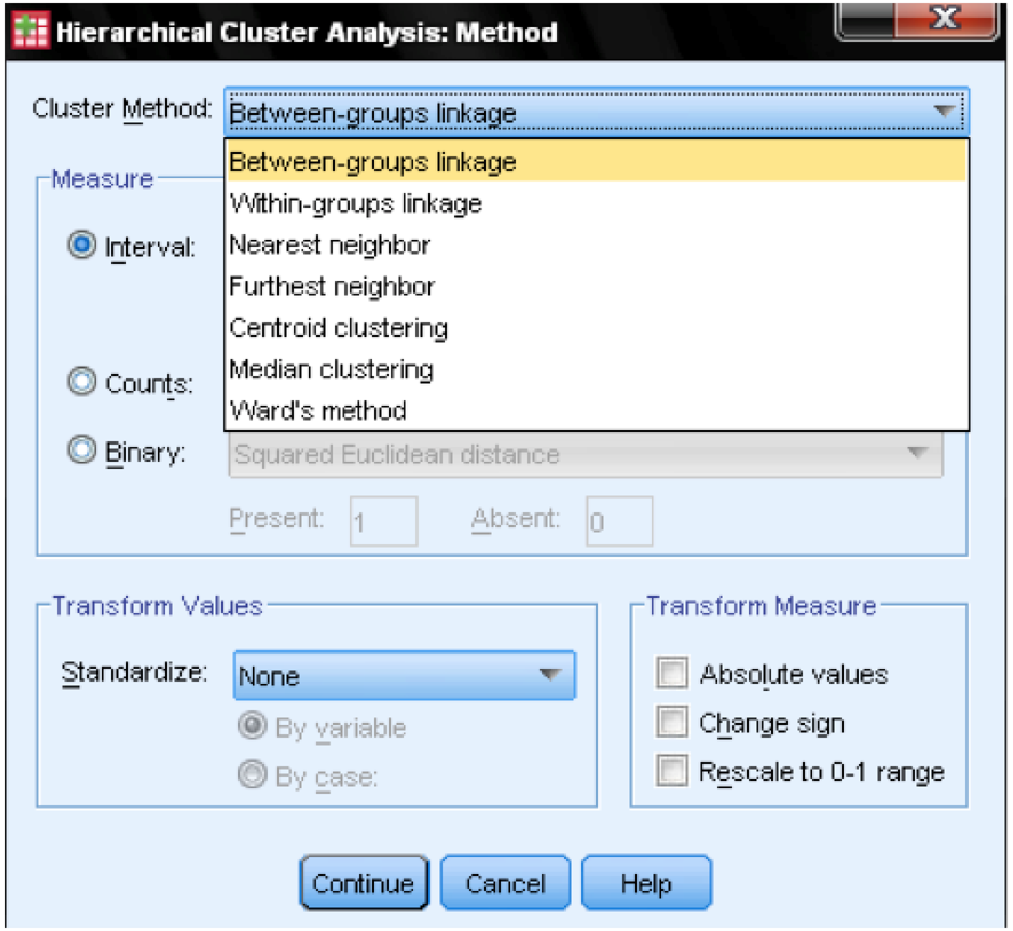
Choose clustering method. Common options:

- Between-groups linkage (use the average distance of all data points within these clusters),
- Nearest neighbor (single linkage: use the smallest distance between two data points in the clusters),
- Furthest neighbor (complete linkage: use the largest distance between two data points in the clusters),
- Ward's method (distance of all clusters to the grand average of the sample).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Sub-Menu: Method III



- Single linkage works best with long chains of clusters,
- Complete linkage works best with dense blobs of clusters.
- Between-groups linkage works with both cluster types.
  1. Use single linkage first - helps in identifying outliers.
  2. Exclude these outliers.
  3. Then use Ward's method - uses the *F* value (like in ANOVA) to maximize the significance of differences between clusters.

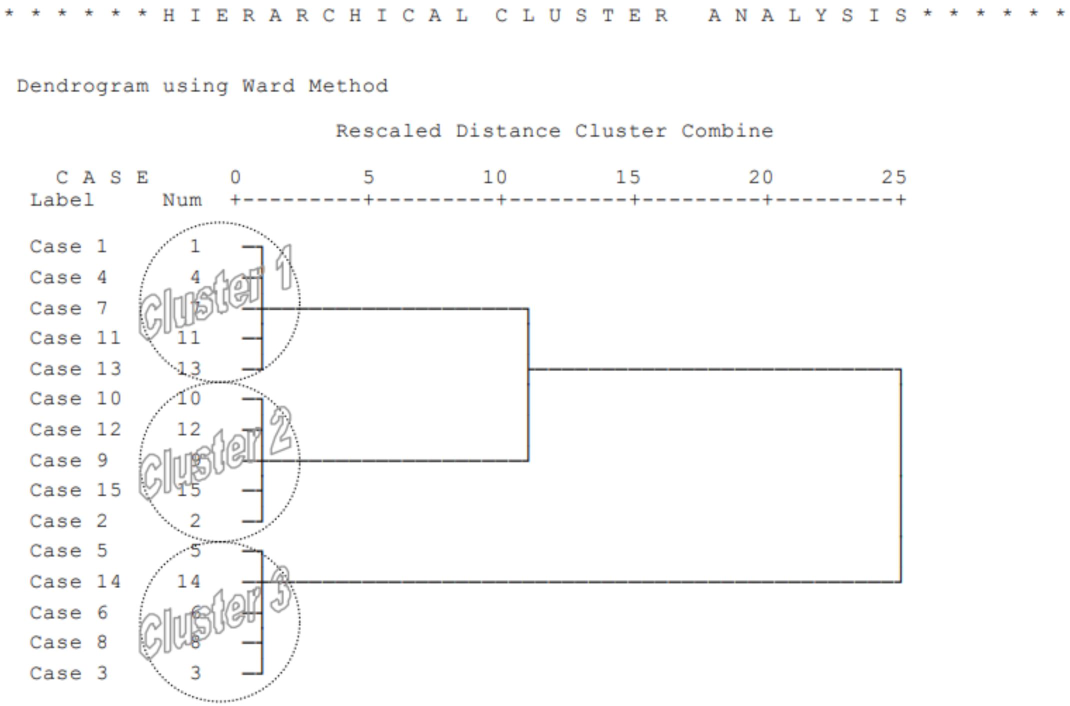
Standardization: to Z scores or by centering the scale.

Can also transfor

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

SPSS Output – Andy Field Clustering Symptoms Example



Case Summaries [a]

	DSMIV Classification	Ward Method
1	GAD	1
2	Depression	2
3	OCD	3
4	GAD	1
5	OCD	3
6	OCD	3
7	GAD	1
8	OCD	3
9	Depression	2
10	Depression	2
11	GAD	1
12	Depression	2
13	GAD	1
14	OCD	3
15	Depression	2
Total N	15	15

a. Limited to first 100 cases.

Having eyeballed the dendrogram and decided how many clusters are p analysis asking SPSS to save a new variable in which cluster codes are assigned to cases (with the researcher specifying the number of clusters in the data).

Here, we ask cluster group codings for three clusters. The resulting codes for each case in this analysis map exactly onto the DSM-IV classifications.

Although this example is very simplistic it shows you how useful cluster analysis can be in developing and validating diagnostic tools, or in establishing natural clusters of symptoms for certain disorders.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Other Examples

- Example of K-Means Cluster Analysis
  - https://www.youtube.com/watch?v=e27G-UCju0E
  - https://www.youtube.com/watch?v=yWwHi8RTYnQ

- Example of 2 step Cluster Analysis
  - https://www.youtube.com/watch?v=DpucueFsigA&list=PLnMJlbz3seflgZdXeXxL8QgKiDLyH6Q-w
  - https://www.youtube.com/watch?v=BrmfYtT98W0
  - https://www.youtube.com/watch?v=Odk0kLuUGvY

- Other resources
  - http://calcnet.mth.cmich.edu/org/spss/staprocclassification.htm

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Limitations of CA

- There are several things to be aware of when conducting cluster analysis:

1. The different methods of clustering usually give very different results. This occurs because of the different criterion for merging clusters (including cases). It is important to think carefully about which method is best for what you are interested in looking at.
2. With the exception of simple linkage, the results will be affected by the way in which the variables are ordered.
3. The analysis is not stable when cases are dropped: this occurs because selection of a case (or merger of clusters) depends on similarity of one case to the cluster. Dropping one case can drastically affect the course in which the analysis progresses.
4. The hierarchical; nature of the analysis means that early ‘bad judgements’ cannot be rectified.

Co-funded by the European Union

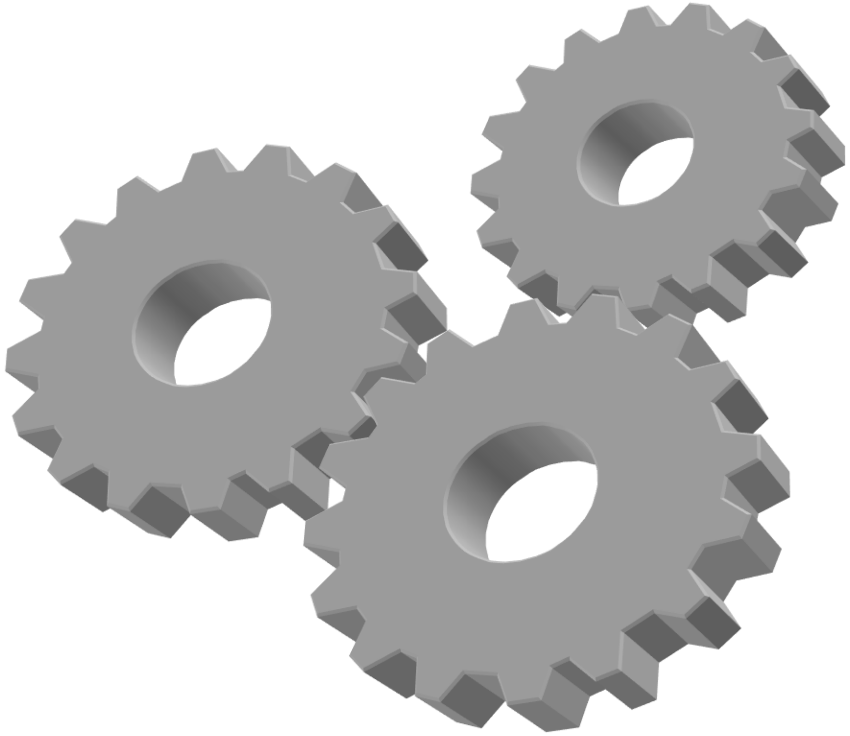
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

To CONCLUDE

- We often have more quantity and depth in our data than we can intuitively comprehend
- It is sometimes helpful to discover clusters in our data, without an a-priori expectation of how many or which clusters will be discovered (data-driven clustering)

- CA, FA and PCA summarize our data
  - There are significant similarities but also differences between the three methods
  - CA and FA/PCA can be combined, e.g. first conduct FA and cluster factor data → you can reduce messiness and complexity in your data and arrive more quickly at a manageable number of clusters.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Statistics

Regression Analysis

(Linear, Simple, Multiple, Mixed-Model)

Co-funded by the European Union

ΕΣΠΑ 2021-2027

Βιώσιμη Ανάπτυξη για Όλους

Partnership Agreement 2021-2027

Regression Analysis

- **Regression**
- Predicting Values of a Variable from Values of One (Simple Regression) or Multiple Other Variables (Multiple Regression)
  - Extension of Correlation (r) and ANOVA
  - Statistical Model of the Relationship between Variables
  - Prediction: One of the Key Goals in Science

- **Independent / Predictive Variable**
- **Dependent Variable / Criterion Variable**

- **Types of Regression**
  - Simple / Multiple
  - Linear / Non-linear (e.g., Curvilinear, Sigmoid, Logarithmic [Logistic Regression])

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Simple Linear Regression

- Assumptions:
  - Sufficient correlation between the 2 variables
  - Linear relationship

- <u>EXAMPLE</u>: Predicting Job Satisfaction (Y) based on Conscientiousness (X) (personality trait)

Οι επιδόσεις των εργαζομένων στην κλίμακα ευσυνειδησίας και στο τεστ επαγγελματικής ικανοποίησης

Άτομο	Ευσυνειδησία	Επαγγελματική ικανοποίηση
1	74	56
2	57	58
3	45	43
4	81	75
5	56	61
6	62	60
7	33	42
8	83	93
9	72	59
10	53	64
Περιγραφικοί στατιστικοί δείκτες για το δείγμα		
$\bar{X}$	61,6	61,1
s	16,04	14,73
COV	189,27	
N	10	10

- Testing correlation →

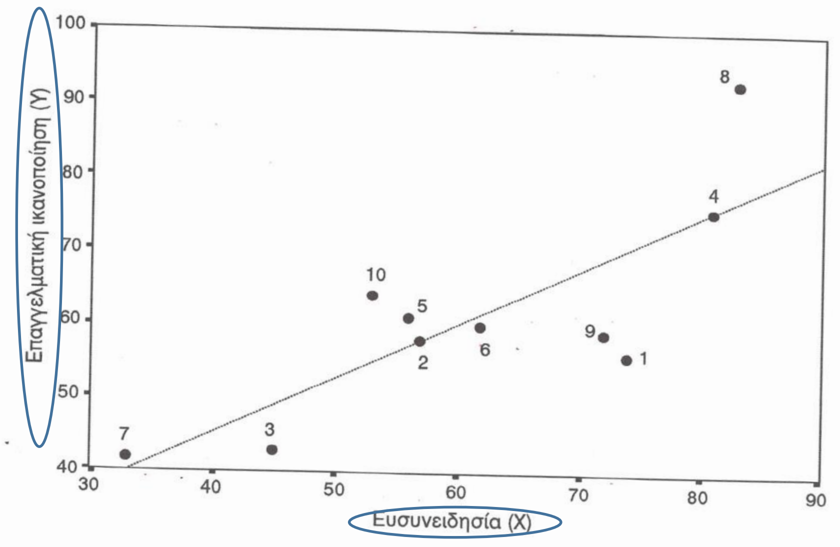
$$r = \frac{COV_{XY}}{s_X s_Y} = \frac{189,27}{236,27} = 0,80$$

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Descriptive Analysis

- Before the statistical analysis, we can create a scatter plot to examine the relationship between the variables.



Σχήμα 10.1. Το διάγραμμα σκεδασμού και η γραμμή παλινδρόμησης για τις μεταβλητές της ευσυνειδησίας (X) και της επαγγελματικής ικανοποίησης (Y).

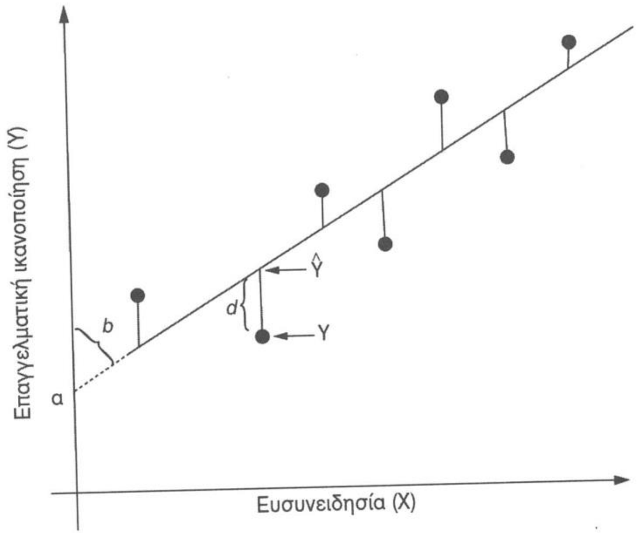
- ***Regression Line:***
  - *Passes as Close as Possible to All Data Points*
- ***Regression Coefficients:***
  - *a: Value where the line intersects the X-axis (constant)*
  - *b: Slope of the line (significance matters)*
- ***Equation:***
  - *Y = a + bX + error*
  - *This is our statistical model.*

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Calculating the Regression Line (Least Squares Method)

- Difference (d) between observed (Y) and predicted value (Ŷ): **Prediction Error**
  - Mathematical formula for minimizing the sum of these squared errors.



Σχήμα 10.3. Η γραμμή παλινδρόμησης για τις επιδόσεις των εργαζομένων στην κλίμακα ευσυνειδησίας (X) και στο τεστ επαγγελματικής ικανοποίησης (Y).

$$\hat{Y} = \alpha + bX,$$

όπου

$\hat{Y}$ = η προβλεπόμενη τιμή Y,

α = η τιμή του Y όταν το X=0,

b = η κλίση της ευθείας, δηλαδή η γωνία που σχηματίζει η ευθεία με τον άξονα των X, και

X = κάθε τιμή του X.

$$b = \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{N(\Sigma X^2) - (\Sigma X)^2},$$

όπου

N = ο αριθμός των ατόμων του δείγματος,

X = οι τιμές της μεταβλητής X,

Y = οι τιμές της μεταβλητής Y, και

Σ = το άθροισμα των...

$$a = \bar{Y} - b\bar{X},$$

όπου

$\bar{X}$ = ο μέσος όρος των τιμών της μεταβλητής X,

$\bar{Y}$ = ο μέσος όρος των τιμών της μεταβλητής Y, και

Assessing the Predictive Value of Regression

- The predictive value of each model is expressed by the coefficient of determination (R^2).
  - It represents the percentage of the variance in the dependent variable Y explained by the independent variable X (simple regression).
  - It also represents the percentage of the total variance in the dependent variable (DV) explained by the set of independent variables (IVs) (multiple regression).

- Simple Linear Regression
  - $Y = \alpha + \beta X + \varepsilon$
- Multiple Linear Regression
  - $Y = \alpha + \beta X + \beta_2 Z + \ldots + \beta_i \Omega + \varepsilon$
  - $\beta \ldots \beta_i$: beta coefficients
  - In multiple regression, high correlation between predictor variables is undesirable (multicollinearity).

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Simple Linear Regression

$$b = \frac{10(39341) - (616)(611)}{10(40262) - (379456)} = \mathbf{0.74}$$

a = 61.1 – (0.74 × 61.6) =
= 61.1 – 45.6 = **15.5**

- **Ŷ = 15.5 + 0.74 × X**
- To draw the regression line (scatterplot) we need at least two pairs of values.
  - E.g. (40, 45.1), (50, 52.5)
  - (10, ?)

$$b = \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{N(\Sigma X^2) - (\Sigma X)^2},$$

$$a = \bar{Y} - b\bar{X},$$

όπου

$\bar{X}$ = ο μέσος όρος των τιμών της μεταβλητής X,

$\bar{Y}$ = ο μέσος όρος των τιμών της μεταβλητής Y, και

b = η κλίση της ευθείας.

Οι επιδόσεις τ… …κής ικανοποίησης

Άτομο	Ευσυνειδησία	Επαγγελματική ικανοποίηση
1	74	56
2	57	58
3	45	43
4	81	75
5	56	61
6	62	60
7	33	42
8	83	93
9	72	59
10	53	64
Περιγραφικοί στατιστικοί δείκτες για το δείγμα		
$\bar{X}$	61,6	61,1
s	16,04	14,73
COV	189,27	
N	10	10

- Coefficient of determination (r^2)
  - The shared percentage of variation explained by the 2 variables
  - r^2 = Σ(Ŷ – Y) / Σ(Y

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Linear Regression

- Assumptions:
  - 2 or more independent variables
  - Satisfactory correlation between each independent variable and the dependent variable
    - BUT not between the independent variables (multicollinearity)
  - Linear relationship

- <u>EXAMPLE</u>: Predict vocational sucess (Y) based on age (X_1), education (X_2), and professional experience (X_3)

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n,$$

όπου

$\hat{Y}$ = η προβλεπόμενη τιμή Υ,

α = η τιμή του Υ όταν το Χ=0,

$b_1, b_2, b_3, \ldots b_n$ = η κλίση της ευθείας για κάθε προβλεπτική μεταβλητή X_1, X_2, X_3, X_n, και

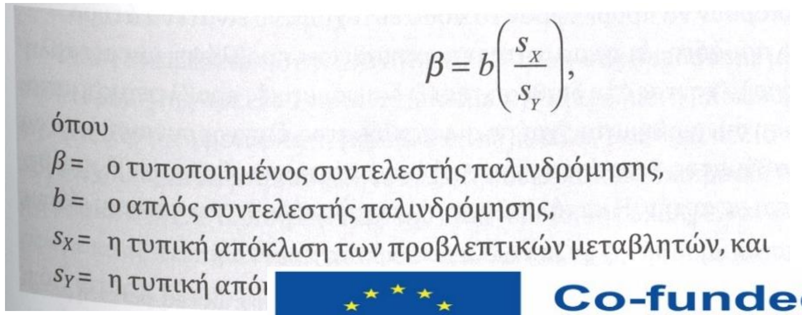
$X_1, X_2, X_3, \ldots X_n$ = η τιμή της κάθε προβλεπτικής μεταβλητής.

- **Multiple correlation coefficient (R):** (R): the overall correlation of the DV with all the IVs
- R = COV(Y, Ŷ) / $s_Y s_{\hat{Y}}$
  - COV = covariance
  - s = standard deviation
- **R^2**: the percentage of total variance in the dependent variable explained by the set of independent variables

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Regression Coefficients in Multiple Linear Regression

- Regression coefficients (b) for every IV
  - Calculated independently, similar to simple regression
- Assessing the predictive value of each independent variable through b
- **Standardized regression coefficient (β)**
  - obtained by converting to z-scores, equalizing the variances of all coefficients
  - Makes comparison more «fair»
  - R^2 calculated for each β separately



$$\beta = b\left(\frac{s_X}{s_Y}\right),$$

όπου

$\beta =$ ο τυποποιημένος συντελεστής παλινδρόμησης,

$b =$ ο απλός συντελεστής παλινδρόμησης,

$s_X =$ η τυπική απόκλιση των προβλεπτικών μεταβλητών, και

$s_Y =$ η τυπική από

Reporting Regression Results

- The model created describes the factors that significantly predict professional success ($R^2 = .82$). Professional success is influenced by work experience [$t(196) = 12.26$, $p < .001$], education [$t(196) = 12.12$, $p < .001$], and employee's age [$t(196) = 4.55$, $p < .001$].
- Work experience explains 51% of the variance in measuring professional success, education explains 51%, and the employee's age explains 19% of the variance.

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Simple Linear Regression in SPSS

- Predict health levels (Y) based on optimism (X)
- Analyze → Regression → Linear

Adjusted based on sample size

Correlations

		health	optimism
health	Pearson Correlation	1	,756**
	Sig. (2-tailed)		,000
	N	179	179
optimism	Pearson Correlation	,756**	1
	Sig. (2-tailed)	,000	
	N	179	179

**. Correlation is significant at the 0.01 level (2-tailed).

Εικόνα 10.1. Πίνακας συντελεστών συσχέτισης των μεταβλητών «health» και «optimism»

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,756[a]	,571	,569	7,233

a. Predictors: (Constant), optimism

Εικόνα 10.3. Σύνοψη μοντέλου της ανάλυσης απλής παλινδρόμησης

How successful is the model's prediction

Asesses the assumption of linear association

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12347,875	1	12347,875	235,997	,000[a]
	Residual	9261,030	177	52,322		
	Total	21608,905	178			

a. Predictors: (Constant), optimism
b. Dependent Variable: health

Εικόνα 10.4. Πίνακας ANOVA στην ανάλυση απλής παλινδρόμησης

a

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-14,751	4,707		-3,134	,002
	optimism	,359	,023	,756	15,362	,000

a. Dependent Variable: health

Εικόνα 10.5. Πίνακας συντελεστών του μοντέλου στην ανάλυση απλής παλινδρόμησης

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Linear Regression in SPSS

- Predicting the level of happiness (Y) based on emotional intelligence scale and emotional expression scale (emotion recognition, emotion use, emotion control, emotion understanding, positive & negative expression: X_1-X_6)

- Analyze → Regression → Linear
  - Method: ?
  - CAUTION! (in some input methods the order of input is meaningful)
  - Statistics, Plots (ZRESID * ZPRED) and other options

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Multiple Linear Regression in SPSS

- Analyze → Regression → Linear

Model Summary[e]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	,661[a]	,437	,434	13,358	,437	137,378	1	177	,000
2	,715[b]	,511	,506	12,481	,074	26,764	1	176	,000
3	,755[c]	,570	,562	11,744	,058	23,791	1	175	,000
4	,768[d]	,589	,580	11,506	,020	8,295	1	174	,004

a. Predictors: (Constant), eq_use
b. Predictors: (Constant), eq_use, emot_po
c. Predictors: (Constant), eq_use, emot_po, emot_ne
d. Predictors: (Constant), eq_use, emot_po, emot_ne, eq_perc
e. Dependent Variable: happiness

Εικόνα 10.9. Σύνοψη μοντέλου στην ανάλυση πολλαπλής παλινδρόμησης

Predictive success of the model

Percentage of model improvement (compares to previous)

Significance of improvement

Coefficients[a]

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	t	Sig.	Collinearity Statistics: Tolerance	VIF
1	(Constant)	64,999	5,282		12,305	,000		
	eq_use	1,065	,091	,661	11,721	,000	1,000	1,000
2	(Constant)	43,259	6,482		6,674	,000		
	eq_use	,747	,105	,464	7,131	,000	,656	1,524
	emot_po	1,069	,207	,336	5,173	,000	,656	1,524
3	(Constant)	68,272	7,968		8,568	,000		
	eq_use	,532	,108	,330	4,925	,000	,547	1,829
	emot_po	1,178	,196	,371	6,020	,000	,648	1,544
	emot_ne	-,763	,156	-,269	-4,878	,000	,811	1,233
4	(Constant)	56,167	8,867		6,335	,000		
	eq_use	,567	,107	,352	5,326	,000	,540	1,853
	emot_po	1,057	,196	,333	5,383	,000	,618	1,618
	emot_ne	-,661	,157	-,233	-4,206	,000	,770	1,298
	eq_perc	,327	,114	,147	2,880	,004	,907	1,103

a. Dependent Variable: hap

Εικόνα 10.10. Πίνακας συ

a

b_{1-4}

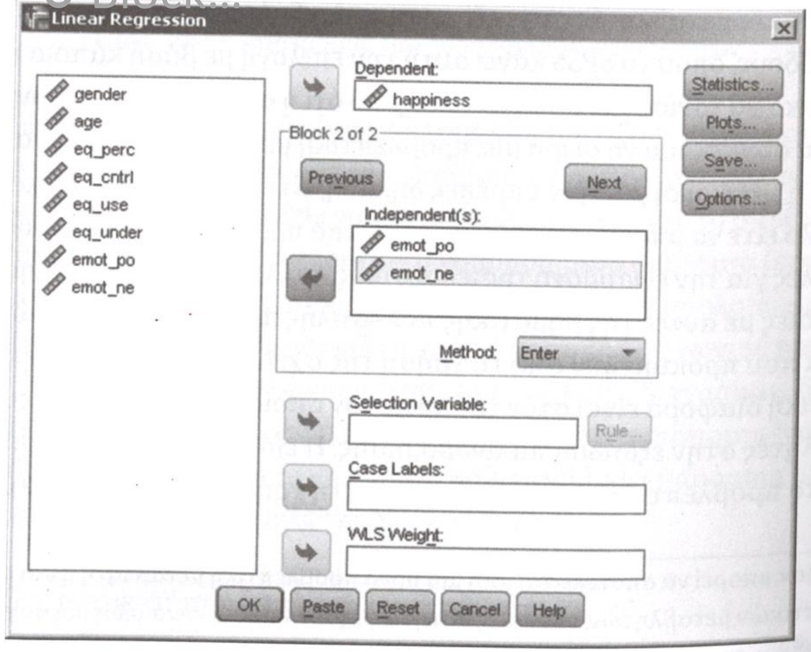
Comparison with standardized values (β)

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Hierarchical Multiple Linear Regression in SPSS

- Researcher decides on the order in which the independent variables will be imput (based on theoretical/empirical criteria).
- Analyze → Regression → Linear
- Block...



Εικόνα 10.14. Πλαίσιο διαλόγου Linear Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,705[a]	,496	,485	12,742	,496	42,880	4	174	,000
2	,769[b]	,592	,578	11,535	,096	20,159	2	172	,000

a. Predictors: (Constant), eq_under, eq_cntrl, eq_perc, eq_use
b. Predictors: (Constant), eq_under, eq_cntrl, eq_perc, eq_use, emot_ne, emot_po

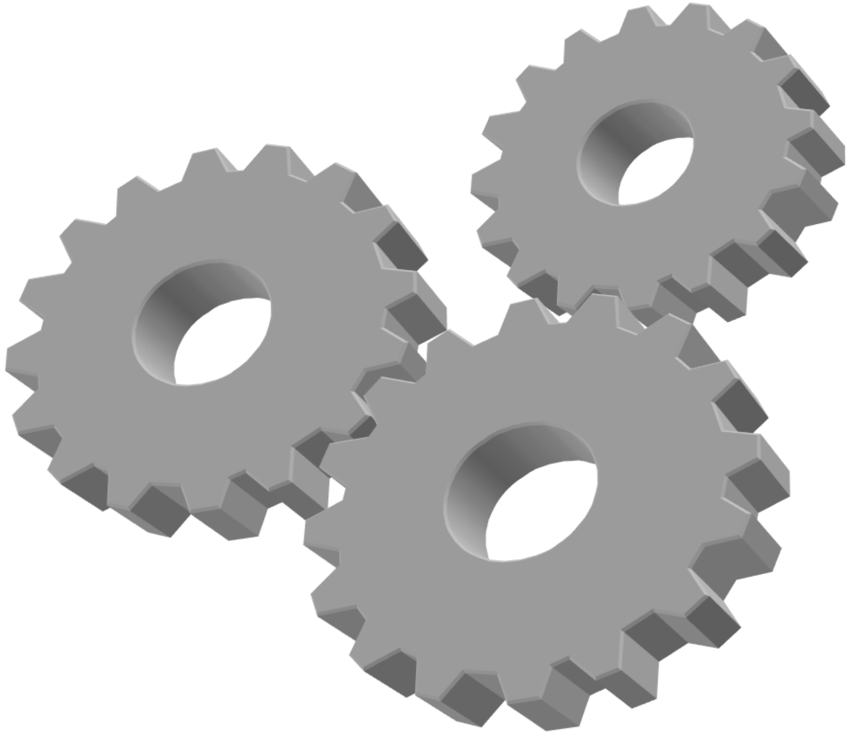
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41,477	9,450		4,389	,000
	eq_perc	,512	,126	,230	4,055	,000
	eq_cntrl	-,014	,108	-,008	-,130	,897
	eq_use	1,006	,098	,624	10,265	,000
	eq_under	,133	,142	,054	,937	,350
2	(Constant)	59,174	12,176		4,860	,000
	eq_perc	,337	,118	,151	2,862	,005
	eq_cntrl	-,102	,108	-,056	-,938	,349
	eq_use	,588	,112	,365	5,245	,000
	eq_under	,043	,129	,018	,334	,739
	emot_po	1,020	,200	,321	5,101	,000
	emot_ne	-,723	,172	-,255	-4,205	,000

a. Dependent Variable: happiness

Εικόνα 10.16. Πίνακας συντελεστών του μοντέλου στην ιεραρχική πολλαπλή παλινδρόμηση

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027



Mixed Models

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mixed Models Advantages

- In simple linear regression models: Y = a + bX + error
- The percentage of the variance that is unrelated to the independent variable(s) is accounted for by the error.

- In mixed models, one or more significant variables unrelated to experimental manipulation (e.g., stimulus type - such as word/image, grammatical structure, condition, etc.) explain part of the remaining variance.
  - Essentially, our model better accounts for the data, minimizing the residual variance due to error.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Mixed Models Advantages II

- The mixed model:

  - It can handle cases of correlated variables or violations of measurement independence (e.g., similarities in students' mathematical abilities within a class, multiple measurements from each participant - such as performance in a video game).

  - It can handle heterogeneous variances.

    - ML / REML (maximum likelihood) vs ANOVA

  - It can handle hierarchical data (e.g., groups of students selected from a cohort of selected schools).

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Criterion Variables in Mixed Models

- Fixed effects
  - IVs such as age, gender, years of experience etc.
- Random effects
  - Variables related to the sample and our data (participants, word in a list, image in the set of experimental visual stimuli, school class, etc.).
- <u>Model</u>:
  - $DV = IV_1 + IV_2 + \ldots + IV_\nu + RV_1 + \ldots + RV_\nu + Err$

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Example

- Winter & Grawunder (2012)
- Predicting voice pitch based on the appropriate politeness of the situation (polite vs. informal, e.g., asking a favor from a teacher or a friend).

- DV = voice pitch (continuous measure)
- IV_1 = appropriate politeness (2 levels)
- IV_2 = gender (men < women)
- RV_1 = individual differences (participant)
- RV_2 = differences between items

- http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Fixed VS. Random Effects

- So, a random effect is generally something that can be expected to have a nonsystematic, idiosyncratic, unpredictable, or "random" influence on your data. In experiments, that's often "subject" and "item", and you generally want to generalize over the idiosyncrasies of individual subjects and items.
- Fixed effects on the other hand are expected to have a systematic and predictable influence on your data.
- But there's more to it. One definition of fixed effects says that fixed effects "exhaust the population of interest", or they exhaust "the levels of a factor". Think back of sex. There's only "male" or "female" for the variable "gender" in our study, so these are the only two levels of this factor. Our experiment includes both categories and thus exhausts the category sex. With our factor "politeness" it's a bit trickier. You could imagine that there are more politeness levels than just the two that we tested. But in the context of our experiment, we *operationally defined* politeness as the difference between these two categories – and because we tested both, we fully "exhaust" the factor politeness (as defined by us).
- In contrast, random effects generally sample from the population of interest. That means that they are far away from "exhausting the population" … because there's usually many many more subjects or items that you could have tested. The levels of the factor in your experiment is a tiny subset of the levels "out there" in the world.

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Type of Variance Analysis is Dependent on Scale of Measurement of the DV

- Continuous DV:
  - Linear regression model with mixed effects

- Dual-type DV (e.g. 0-1)
  - Logistic regression model with mixed effects)

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Conducting Mixed Model Analysis

- SPSS (using Syntax)
  - Linear Mixed-Effects Modeling in SPSS: An introduction to the mixed procedure
  - Mixed models in SPSS short guide

- Using open access software R
  - http://www.r-project.org/
  - Needs familiarization / basic programming skills, but it is nothing too scary or difficult…
  - Lme4 package for mixed models
  - Mixed Effects Models in R

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Q & A

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement

Additional Resources

- Please find activities, readings, video and websites on e-class!

Co-funded by the European Union

ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027

Thank you!

Co-funded by the European Union
ΕΣΠΑ 2021-2027
Βιώσιμη Ανάπτυξη για Όλους
Partnership Agreement 2021-2027