WORKSHOP Part I
TYPES OF DATA

Dr., Professor **JELENA TITKO**
EKA University of Applied Sciences
Jelena.titko@eka.edu.lv

eka
EKONOMIKAS UN KULTŪRAS
AUGSTSKOLA

Types of Data

QUANTITATIVE (*Numerical*):

- Discrete
- Continuous

QUALITATIVE (*Categorical*):

- Nominal
- Ordinal

Quantitative data

Discrete data contain the values that fall under integers or whole numbers.

Examples: number of students in a class; number of children in a family, and etc.

Continuous data are in the form of fractional numbers.

Examples: height of a person; stock price

Qualitative data

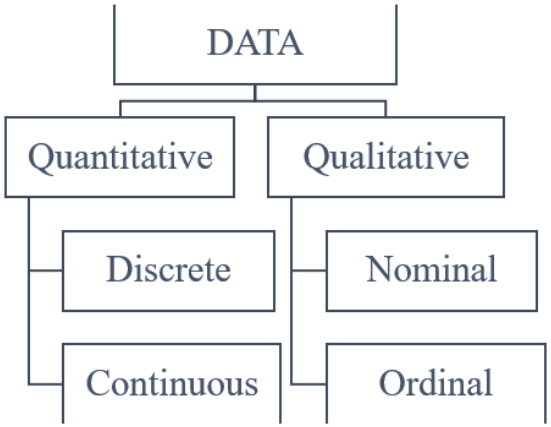
Nominal data is used to label variables without any order.

Examples: marital status; nationality

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale.

Examples: customer satisfaction (1-absolutely not satisfied; 5 – absolutely satisfied); education level (secondary education; Bachelor, Master; Dr.)

Training 1.1



1. What is your age?

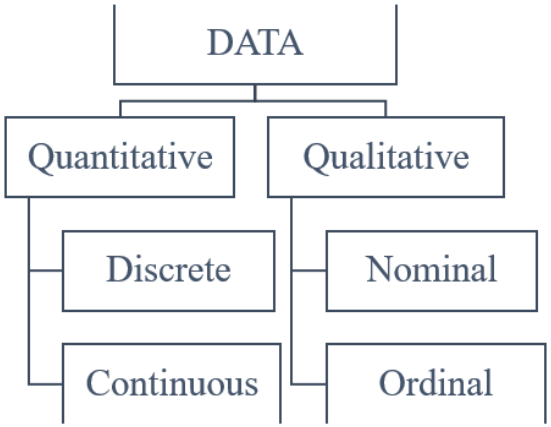
Answer here _____

2. How many study courses do you have per semester?

3. What is your gender?

- Male
- Female
- Other

Training



1. Are you satisfied with your job?

- Strongly disagree
- Disagree
- Neither disagree, nor agree
- Agree
- Strongly agree

Likert scale

A *Likert scale* is commonly used to measure attitudes, knowledge, perceptions, values, and behavioural changes.

Level of Agreement	Amount of Use	Level of difficulty
1 – Strongly disagree	1 – Never use	1 – Very difficult
2 – Disagree	2 – Almost never	2 – Difficult
3 – Neither agree, nor disagree	3 – Sometimes	3 – Neutral
4 – Agree	4 – Almost every time	4 – Easy
5 – Strongly agree	5 – Frequently use	5 – Vey easy

Question from a Forum

«I am analyzing a data from survey. ... Since the respondents can only answer the questions with a number between 1 to 5 (or 1 to 10), does it mean that the data are discrete and cannot be normally distributed? If it is not normal distribution and some are skewed, can I still do the two-way anova analysis or what should I do??»

«I used Likert scale(1 to 5) in my survey and collected 169 responses. I was wondering if I have to test normality before analysing my data or not and if yes? what kind of test is matched with my data?»

WHY they are asking?

«Reviewers of research reports frequently criticize the choice of statistical methods. ...frequently the use of various parametric methods such as analysis of variance, regression, correlation are faulted because: (a) the sample size is too small, (b) the data may not be normally distributed, or (c) the data are from Likert scales, which are ordinal, so parametric statistics cannot be used.» (*Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics*)

WHY we should check whether the data is normally distributed?

1. REGRESSION assumption

The error term is **normally distributed** with zero mean value (positive and negative error values compensate each other).

2. T-TEST *(statistic method used to determine if there is a significant difference between the means of two groups based on a sample of data).* Assumption: data variables must follow a **normal distribution**

Likert data

The general question centers on whether you should use a *parametric* or *nonparametric* test to analyze Likert data.

Likert data are ordinal, discrete, and have a limited range. These properties violate the assumptions of most **parametric tests**.

- **Parametric tests** assume that the data are continuous and follow a normal distribution.
- **Nonparametric tests** are accurate with ordinal data and do not assume a normal distribution.

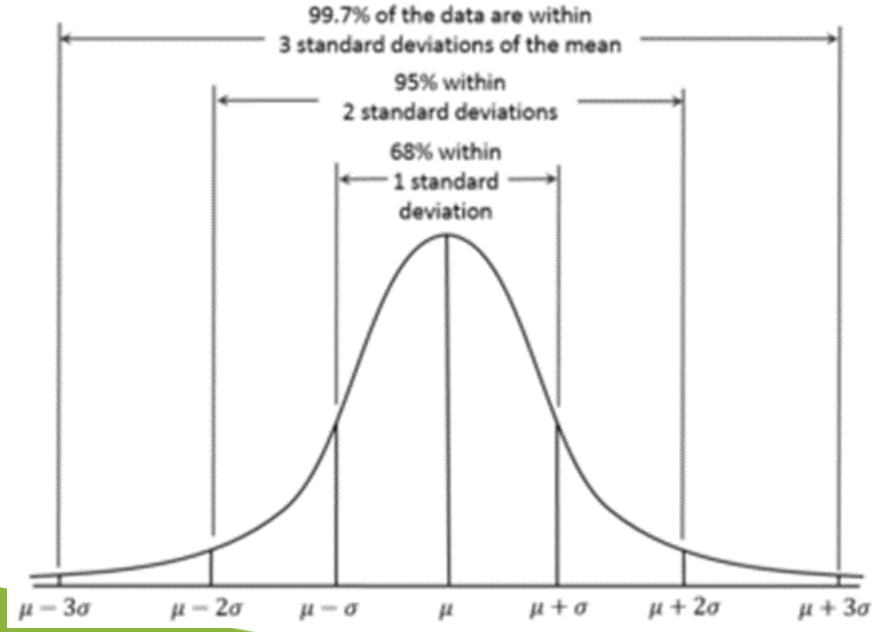
Normal distribution

We say the data is "normally distributed«, if:

- mean = median = mode
- there is a symmetry about the center
- 50% of values less than the mean
- and 50% greater than the mean

Examples of normal distribution

- height of the population
- IQ of a particular population
- birth weight of a newborn



Training 1.2

Test for normal distribution in SPSS

STEP 1. Use the data from the file «2021_Training1.2_DATA»

About the research:

Goal: to investigate the relationship between bank value and bank-specific indicators

Method: Multiple linear regression

Bank Value (calculated by NASDAQ) is regressed against bank specific indices from the financial reports of Latvian banks (2012-2019)

Training 1.2

Test for normal distribution in SPSS

Variables:

X1 – Operational Profit (000'EUR)

X2 – EBITDA

X3 – Assets (000'EUR)

X4 - Legal customers

X5 - Deposits (000'EUR)

X6 - Loans (000'EUR)

Y – Bank value

Training 1.2

Test for normal distribution in SPSS

STEP 2. Place the data to SPSS

STEP 3. Label the data

STEP 4. "Analyze" --> "Nonparametric Tests" --> "One sample"

Results of the test for normal distribution

H0: the population is normally distributed

H1: the population is not normally-distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of VAR00001 is normal with mean 280 259,35 and standard deviation 286 158,19.	One-Sample Kolmogorov-Smirnov Test	,128	Retain the null hypothesis.
2	The distribution of VAR00002 is normal with mean 35 794,40 and standard deviation 39 602,84.	One-Sample Kolmogorov-Smirnov Test	,160	Retain the null hypothesis.
3	The distribution of VAR00003 is normal with mean 335 853,42 and standard deviation 337 485,96.	One-Sample Kolmogorov-Smirnov Test	,159	Retain the null hypothesis.
4	The distribution of VAR00004 is normal with mean 2 537 356,30 and standard deviation 1 992 422,63.	One-Sample Kolmogorov-Smirnov Test	,053	Retain the null hypothesis.
5	The distribution of VAR00005 is normal with mean 25 597,45 and standard deviation 25 510,35.	One-Sample Kolmogorov-Smirnov Test	,064	Retain the null hypothesis.
6	The distribution of VAR00006 is normal with mean 1 856 156,45 and standard deviation 1 493 543,23.	One-Sample Kolmogorov-Smirnov Test	,124	Retain the null hypothesis.
7	The distribution of VAR00007 is normal with mean 1 488 333,65 and standard deviation 1 379 866,29.	One-Sample Kolmogorov-Smirnov Test	,051	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Training 1.3

Ideal case *vs.* reality

STEP 1. Use the data from the file «2021_Training1.3_DATA»

STEP 2. Place the data to SPSS

~~**STEP 3.** Label the data~~

STEP 4. "Analyze" --> "Nonparametric Tests" --> "One sample"

Results of the test for normal distribution

The distribution of data of some variables is NOT NORMAL (null hypothesis is rejected)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of VAR00008 occur with equal probabilities.	One-Sample Chi-Square Test	,000	Reject the null hypothesis.
2	The distribution of VAR00001 is normal with mean 280 259,35 and standard deviation 286 158,19.	One-Sample Kolmogorov-Smirnov Test	,128	Retain the null hypothesis.
3	The distribution of VAR00002 is normal with mean 35 794,40 and standard deviation 39 602,84.	One-Sample Kolmogorov-Smirnov Test	,160	Retain the null hypothesis.
4	The distribution of VAR00003 is normal with mean 335 853,42 and standard deviation 337 485,96.	One-Sample Kolmogorov-Smirnov Test	,159	Retain the null hypothesis.
5	The distribution of VAR00004 is normal with mean 2 537 356,30 and standard deviation 1 992 422,63.	One-Sample Kolmogorov-Smirnov Test	,053	Retain the null hypothesis.
6	The distribution of VAR00005 is normal with mean 345 594,60 and standard deviation 358 641,17.	One-Sample Kolmogorov-Smirnov Test	,037	Reject the null hypothesis.
7	The distribution of VAR00006 is normal with mean 25 597,45 and standard deviation 25 510,35.	One-Sample Kolmogorov-Smirnov Test	,064	Retain the null hypothesis.
8	The distribution of VAR00007 is normal with mean 319 997,15 and standard deviation 333 226,24.	One-Sample Kolmogorov-Smirnov Test	,038	Reject the null hypothesis.
9	The distribution of VAR00009 is normal with mean 157,98 and standard deviation 151,67.	One-Sample Kolmogorov-Smirnov Test	,017	Reject the null hypothesis.
10	The distribution of VAR00010 is normal with mean 239 078,50 and standard deviation 287 733,38.	One-Sample Kolmogorov-Smirnov Test	,050	Reject the null hypothesis.
11	The distribution of VAR00011 is normal with mean 371 003,20 and standard deviation 375 122,57.	One-Sample Kolmogorov-Smirnov Test	,027	Reject the null hypothesis.
12	The distribution of VAR00012 is normal with mean 1 856 156,45 and	One-Sample Kolmogorov-	,124	Retain the null

So what?

Use nonparametric tests when your data don't meet the assumption about normally distributed data!

	Parametric Procedure	**Nonparametric Procedure**
Compare means between two distinct/independent groups	Two-sample t-test	Wilcoxon rank sum test
Estimate the degree of association between two quantitative variables	Pearson coefficient of correlation	Spearman's rank Correlation