



Descriptive statistics

Călinici Tudor

Objectives

- Present the data using tables
- Present the data using charts
- Compute the descriptive parameters

Research questions

- What do I have to do?
- Which are the variables that quantify the terms in the research question?
- Which other variables I have to collect?

Exposure to Parental Smoking in Childhood is Associated with High C-Reactive Protein in Adulthood:

- Research question: Is the exposure to parental smoking a risk factor for high C-reactive protein in adulthood?
- Variables for research question:
 - Exposure to parental smoking in childhood (yes/no)
 - Is the level of C-reactive Protein high in adulthood (yes/no)
- Other variables (may be relevant)
 - Gender (M/F), Age (Years), Occupation, Active smoker (Y/N), etc.

Descriptive statistics

- Collecting the data
- Presenting the data
- Compute descriptive parameters

Collecting the data

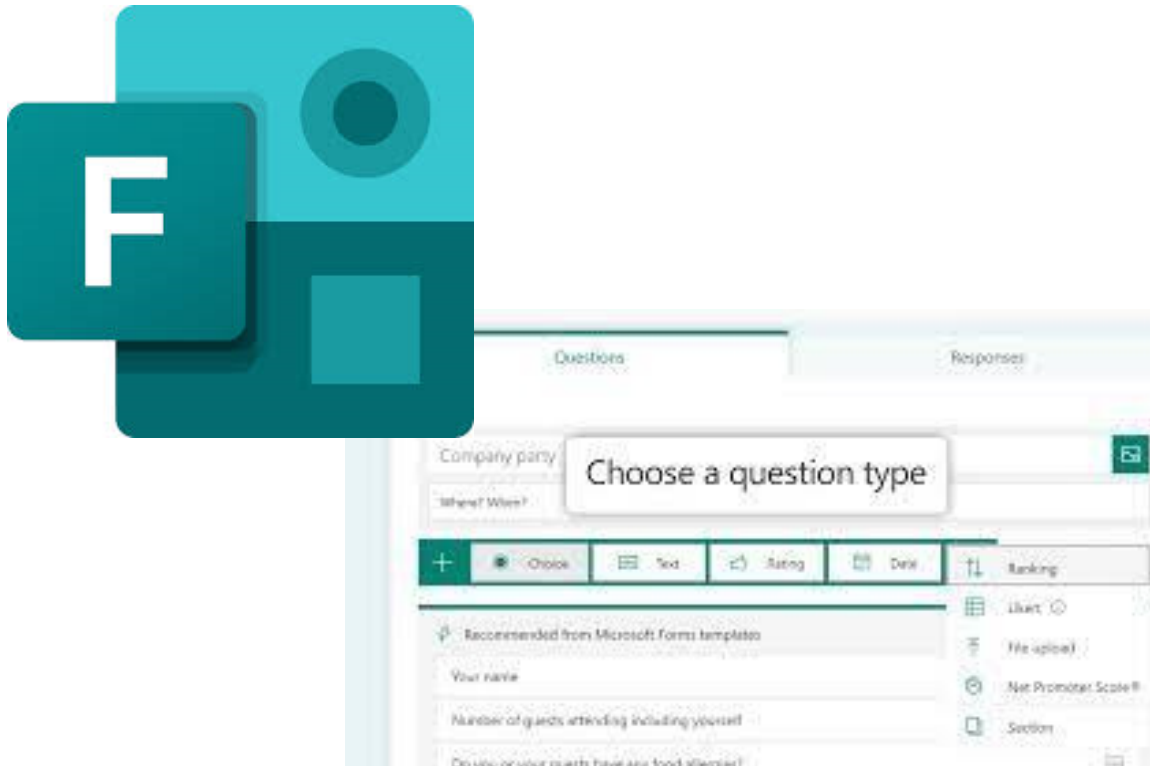
- First step in any research
- For medical research, this process depends of:
 - Objectives of the study
 - Type of the study
 - How the subjects are chosen
 - The human and financials available
 - Access to the data

HOW?

- Online – Microsoft Forms
- Google Forms

Microsoft Forms

- Use your authentication credentials
 - @elearn.umfcluj.ro



Google Forms

- Use your Google Account



Google Forms

A screenshot of a Google Form titled "Brand Awareness" in a web browser. The form contains two multiple-choice questions. The first question is "Which of the following brands have you heard of?" with options: Adidas, Nike, Puma, Converse, and Reebok. The second question is "Which of the following brands have you purchased?" with the option: Adidas. A sign-in overlay is positioned in the foreground, featuring the Google logo, the text "Sign in with your Google Account", a grey person icon, an "Email" input field, a "Password" input field, a blue "Sign in" button, and a "Stay signed in" checkbox with a "Need help?" link.

Or you may have the data already collected

- In medical files
- In electronic format



Check data consistency!



Describe the variables

- Frequency tables – charts
- Contingency tables – charts
- Distribution of the quantitative variables—
Histogram, Boxplot

Frequency table

Name of the variable	Frequency (value,%)
Val l	F _l , f% _l
...	
...	
Total	n 100%

Contingency table

	D+	D-	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	a+b+c+d

PIVOT TABLE

The screenshot displays an Excel spreadsheet with a PivotTable and its corresponding Field List. The PivotTable summarizes sales data by color. The Field List shows 'Color' is selected for the Values area and 'Date' is selected for the Row Labels area.

Row Labels	Sum of Sales
Blue	180
Green	190
Red	144
Silver	476
Grand Total	990

The Field List on the right shows the following configuration:

- Row Labels: Date
- Values: Sum of Sales

Charts

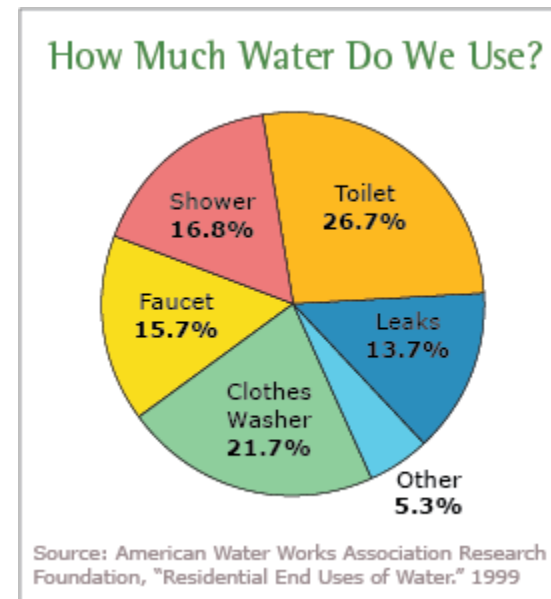
- Chart – graphical representation of a characteristic
- Type of the variable and the goal of the researcher => type of the chart

How to do - Excel

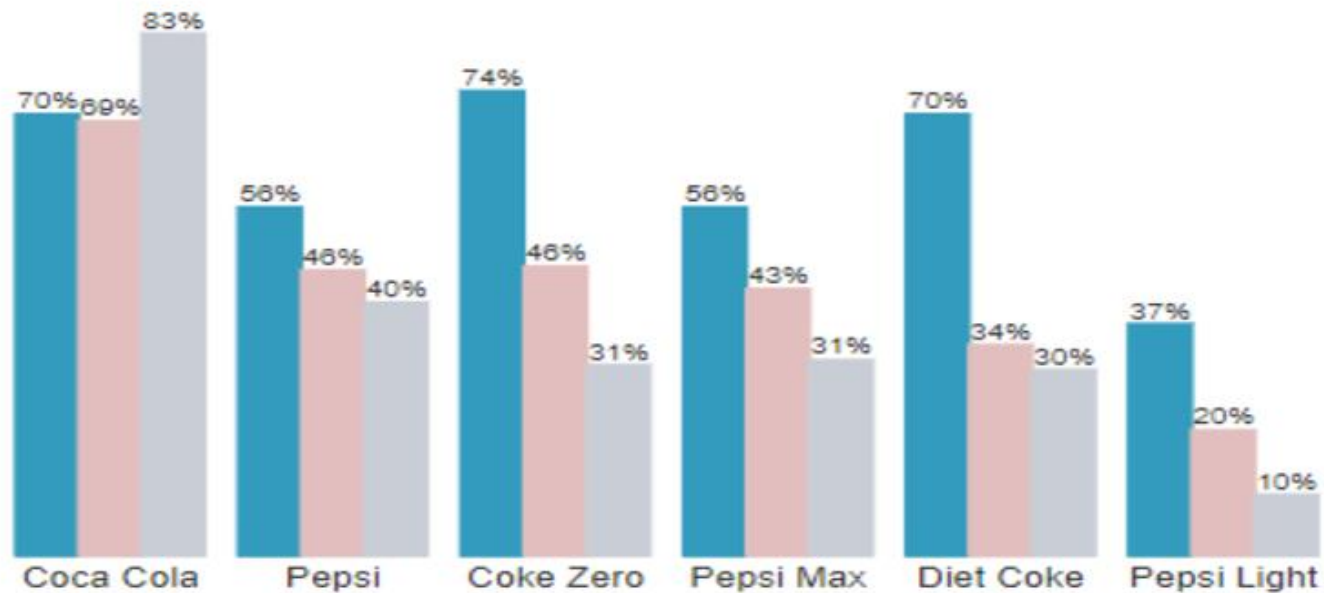
- Select the data
- Choose the correct chart
- Add the rest of the elements

Qualitative variables - frequency

- PIE chart
- The data – the frequency table, not the statistical series

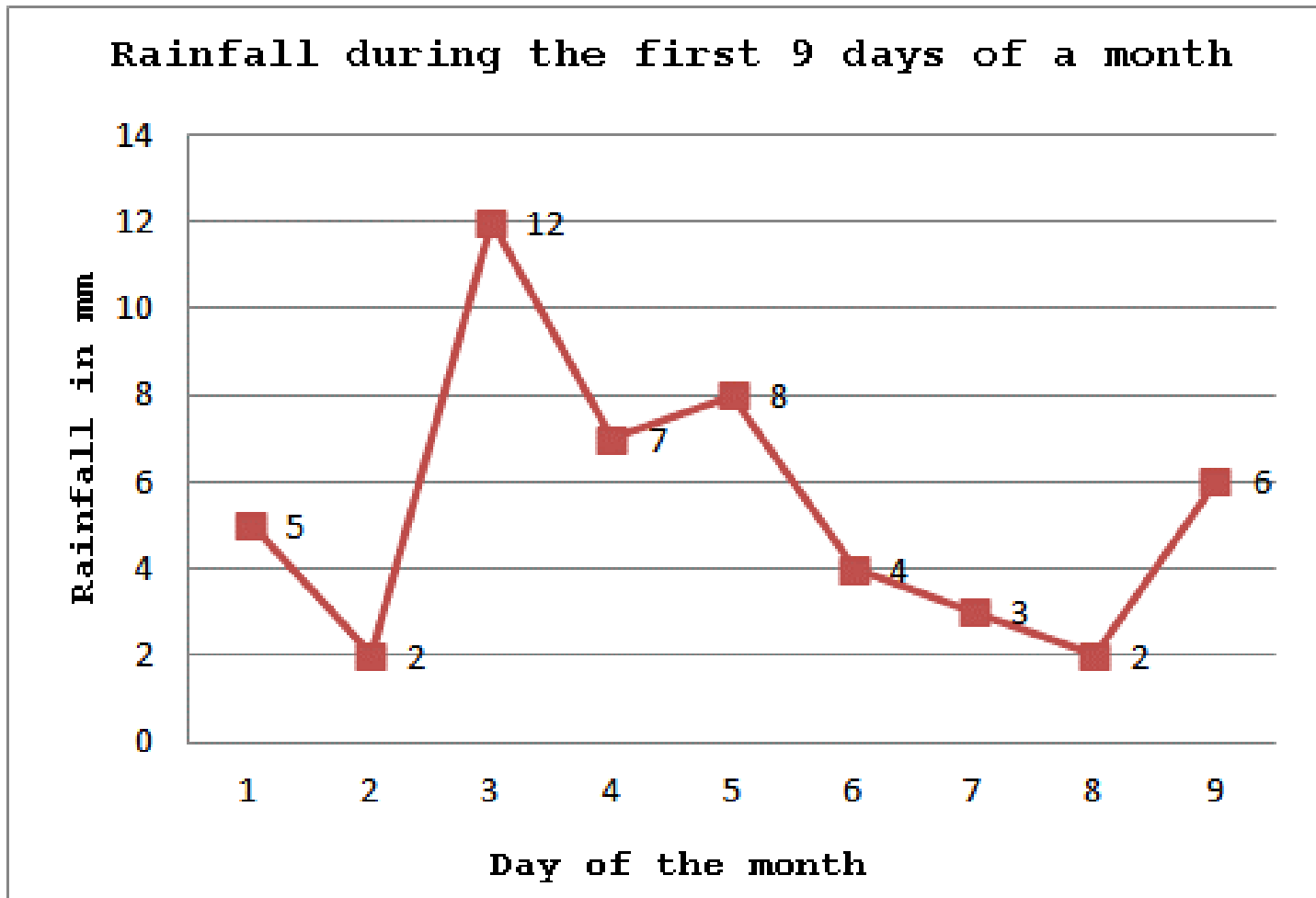


The Column (Bar) Chart



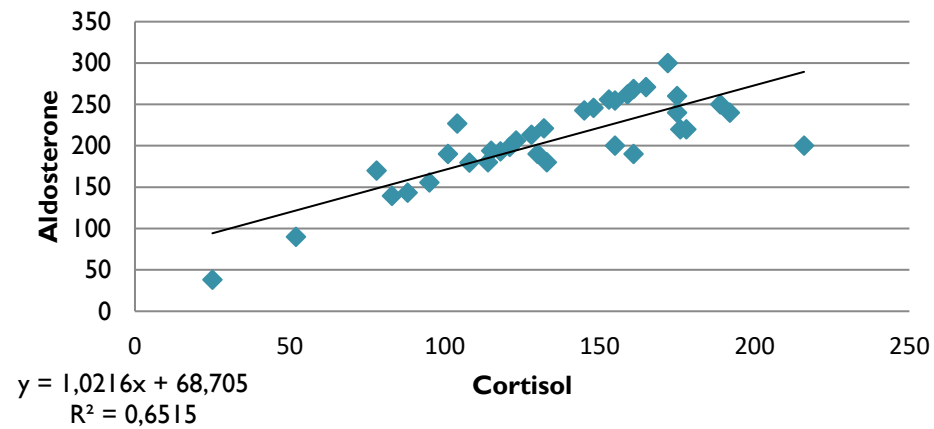
- I am on a diet, so I tend to watch what I eat and drink
- I tend watch what I eat and drink, but don't consider myself
- I typically eat and drink whatever I feel like

Line chart

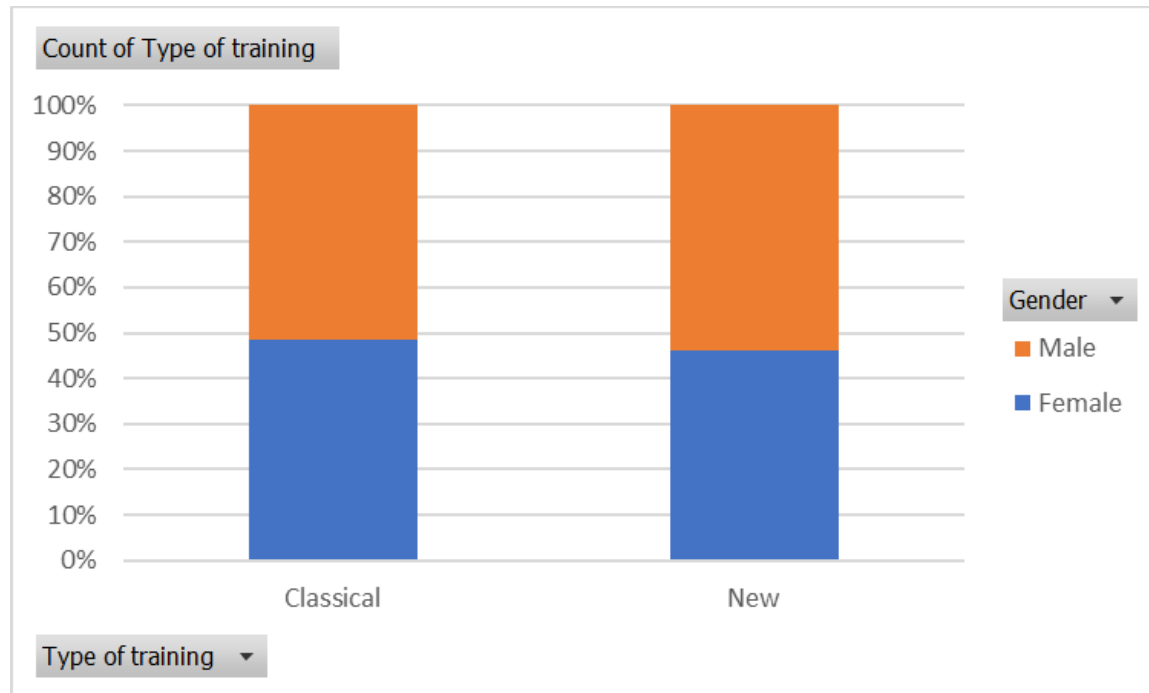


Correlation chart

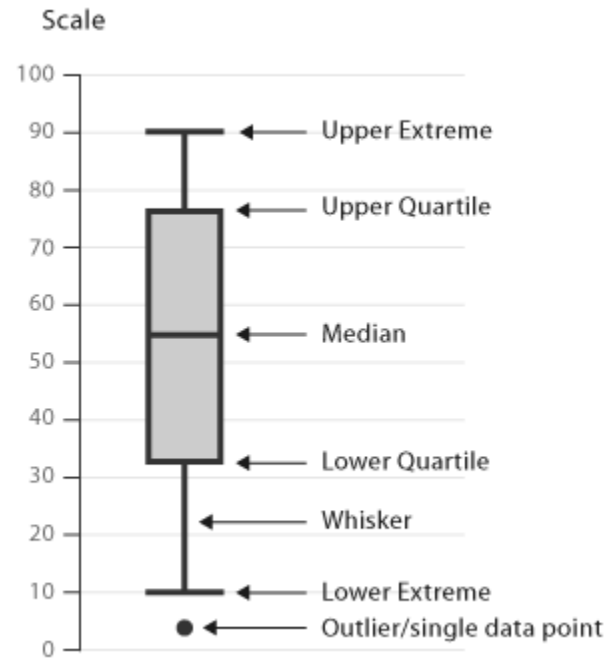
Correlation between cortisol and aldosterone



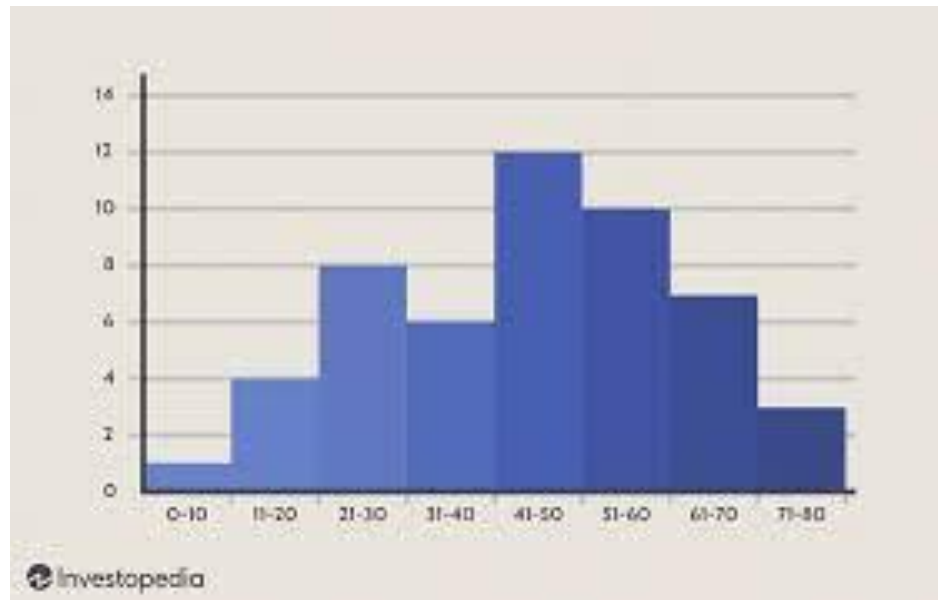
Independence chart



Box and whisker



Histogram



Parameters

- For qualitative variables
- For quantitative variables

If a pathology is present

- Prevalence
- If is possible: incidence
- If the sample is representative CI
- <https://www.graphpad.com/quickcalcs/confInterval2/>

If you study a diagnostic test

- Se, Sp, PPV, NPV – value, CI

When is needed

- OR
- RR, RD
 - Value and CI



Use it!

- <https://statpages.info/ctab2x2.html>

Quantitative data

- Mean (SD) for normal distribution
- Median (Q3-Q1) otherwise
- Data analysis