# 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

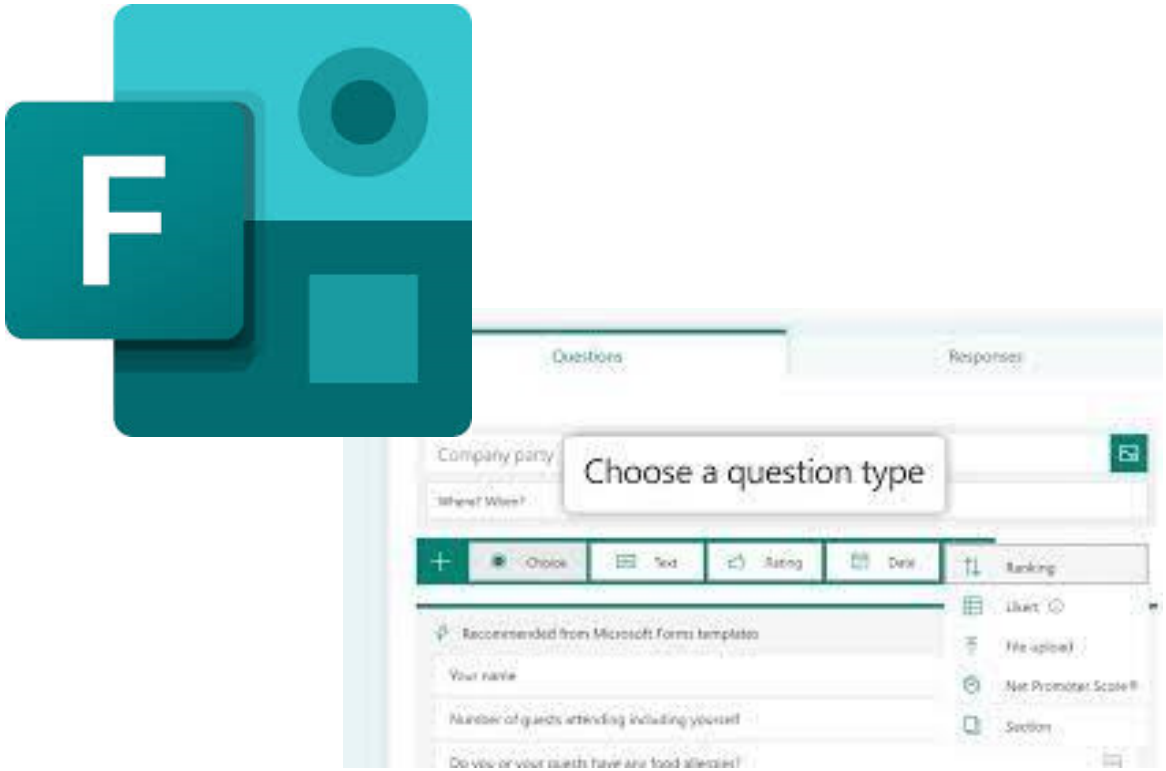
- For medical research, this process depends of:
  - Objectives of the study
  - Type of the study
  - How the subjects are chosen
  - The human and financials available
  - Access to the data

# HOW?

- Online – Microsoft Forms
- Google Forms

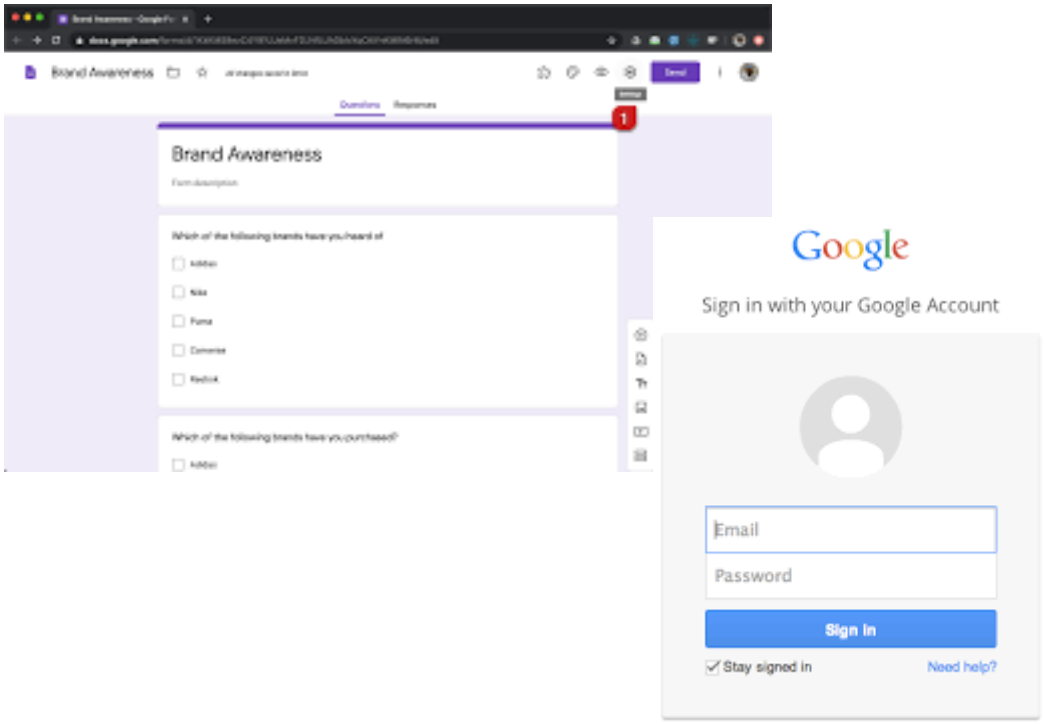
# Microsoft Forms

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



# Google Forms

- Use your Google Account




Google Forms

# Or you may have the data already collected

- In medical files
- In electronical format

# Check data consistency!

# Describe the variables

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

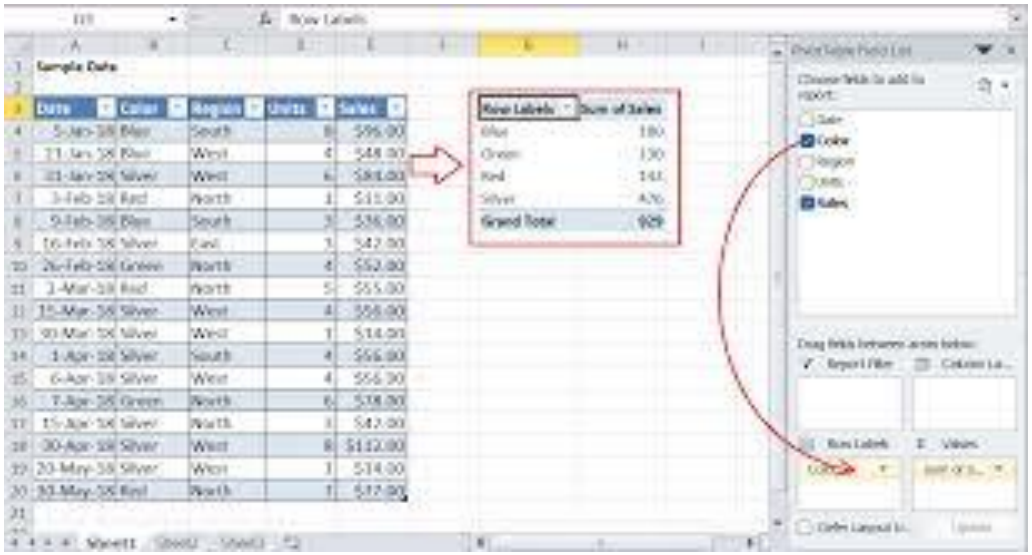
# Frequency table

| Name of the variable | Frequency (value,%) |
|---|---|
| Val 1 | F1, f%1 |
| … | |
| … | |
| 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


Sample Data
Row Labels
Sum of Sales
Blue
Green
Red
Silver
Grand Total
Color
Sales

# Charts

- Chart – graphical representation of a characteristic

- Type of the variable and the goal of the researcher => type of the chart

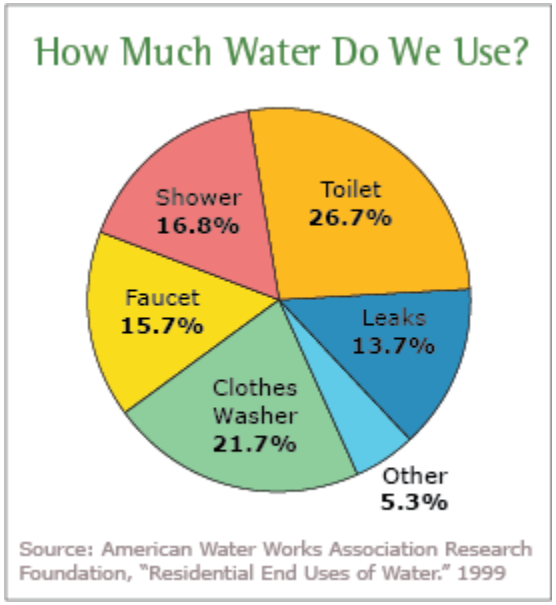
# How to do - Excel

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

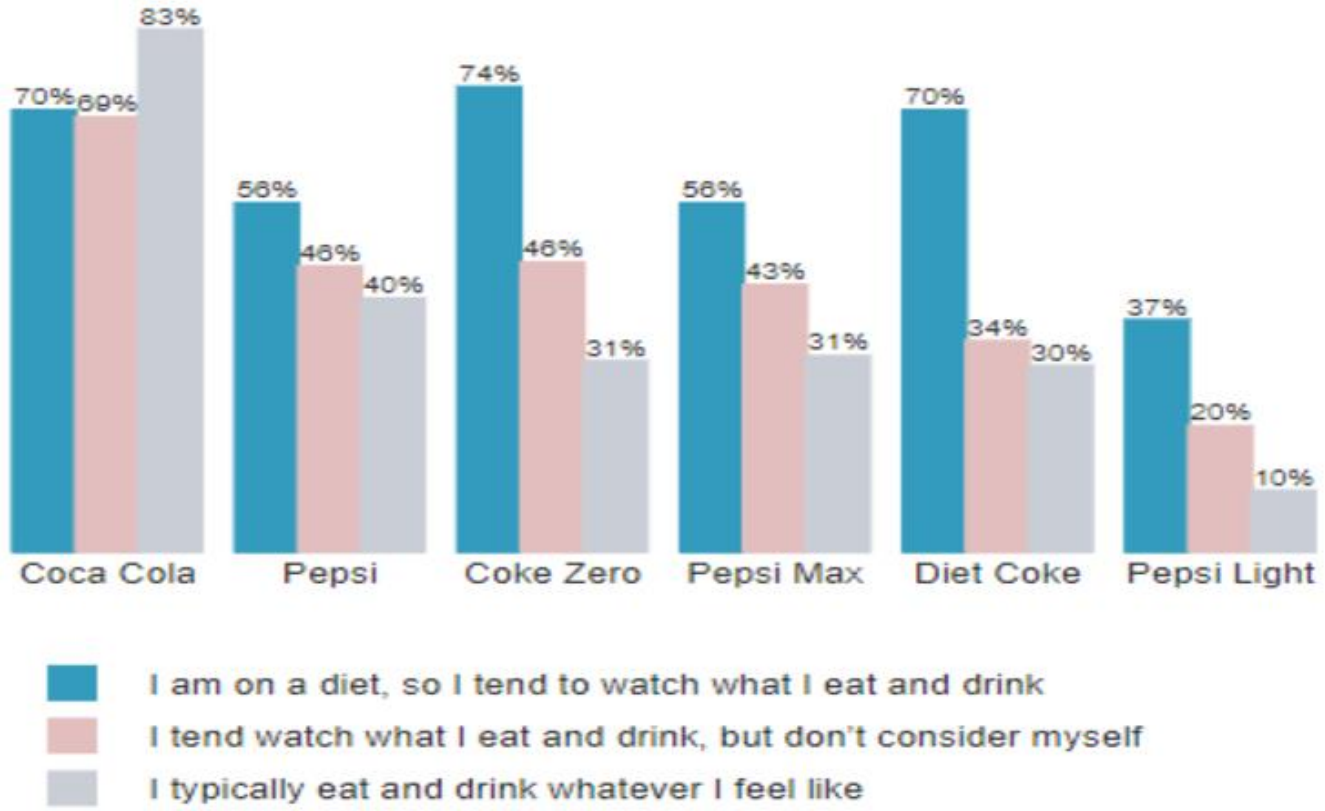
# Qualitative variables - frequency

- PIE chart

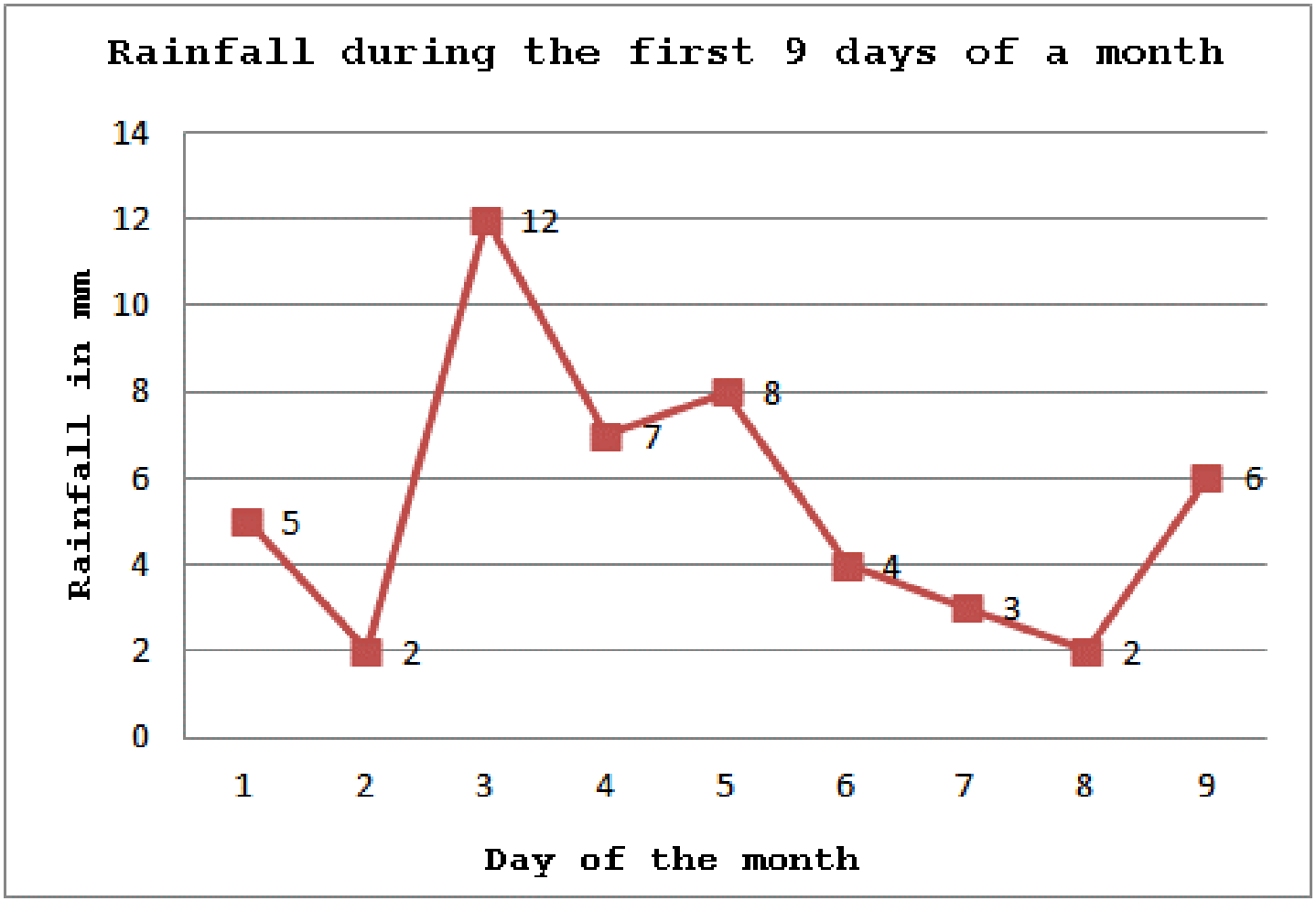
- The data – the frequency table, not the statistical seria



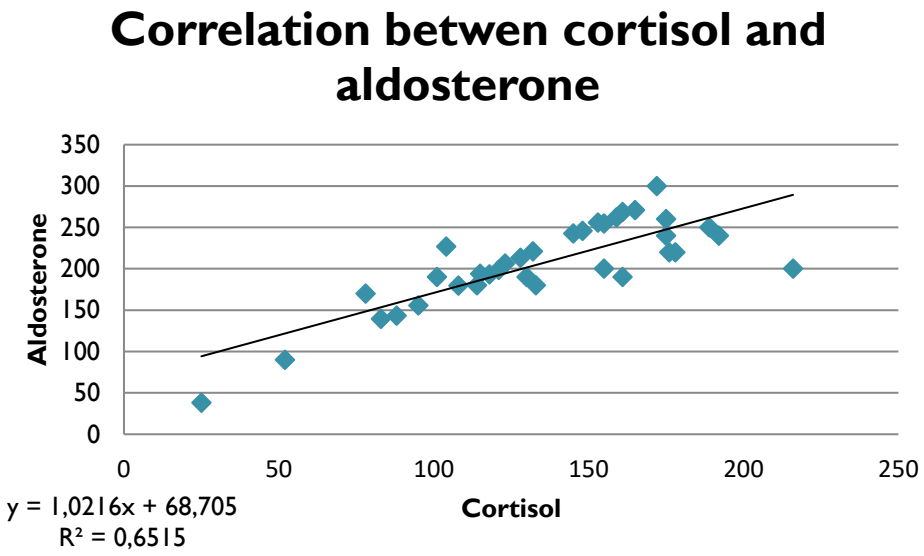
# The Column (Bar) Chart



# Line chart

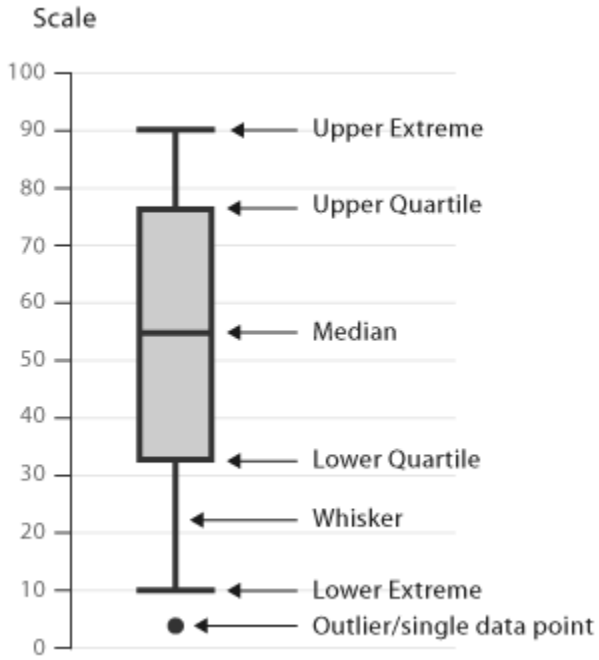

Rainfall during the first 9 days of a month
Rainfall in mm
14
12
10
8
6
4
2
0
5
2
12
7
8
4
3
2
6
1
2
3
4
5
6
7
8
9
Day of the month

# Correlation chart

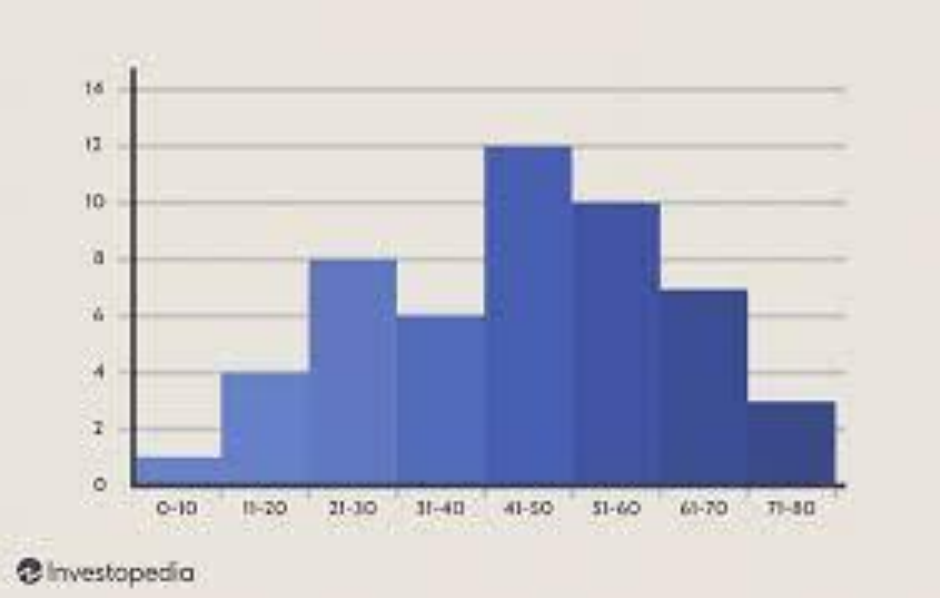


# Independence chart

# Box and wisker



# Histogram


14
12
10
8
6
4
2
0
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
Investopedia

# 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