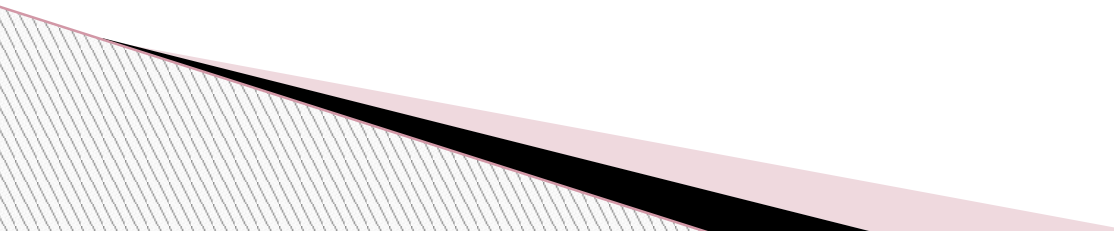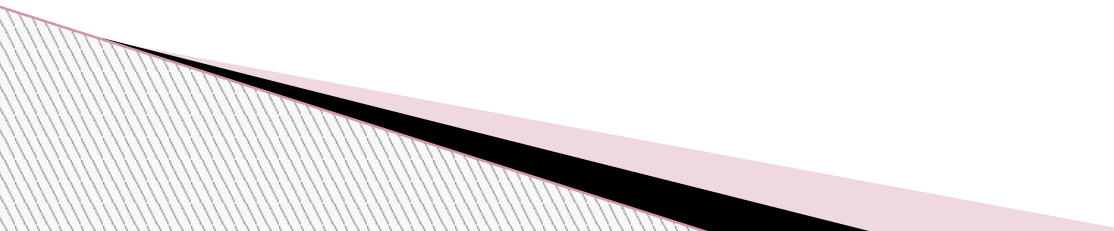# STATISTICS I BA

- WINTER TERM 2021/2022

- Zuzana Pacáková
- pacakova@pef.czu.cz
- FEM - E429
- consultations: Wed 9:30-11:00

Seminars and Lectures

- seminars:
- Wednesday – face-to-face seminars at FEM
- Thursday 12:15 – online in MS Teams
- link: seminar Thu 12:15

- lectures
- Thursday 8:45 – online in MS Teams
- link: lecture Thu 8:45

CREDIT (PASS)

○ <u>attendance at seminars and lectures</u>

○ <u>class tests</u>
  ● three credit tests

To pass the credit, the student has to receive at least 51 points out of 100 (the sum for all tests).

Students are given ONE ATTEMPT for each test and have to take the test in the seminar they are registered for in UIS!

<u>Example:</u>
  Student1: TEST1 – 16p, TEST2 – 20p, TEST3 – 15p → 51 points → passed
  Student2: TEST1 – 0p, TEST2 – 25p, TEST3 – 30p → 55 points → passed
  Student3: TEST1 – 0p, TEST2 – 15p, TEST3 – 20p → 35 points → did not pass

EXAMINATION

- <u>oral – project defence</u>

Reading

- any basic level textbook on statistical methods in English

- **Sonia Taylor: *Business Statistics for Non-mathematicians*, Palgrave MacMillan, 2007 *(available at the International Relations Office, FEM)***

- Field, A. *Discovering Statistics Using SPSS*. London: SAGE Publications, 2005

- KhanAcademy courses
  - https://www.khanacademy.org/math/statistics-probability
- http://cast.massey.ac.nz
  - CAST must be downloaded and installed on your computer.

- https://moodle.czu.cz/

- IBM SPSS Statistics
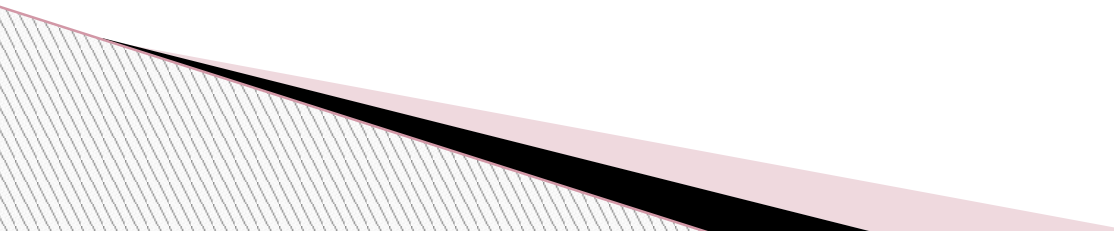
Introduction to Statistics
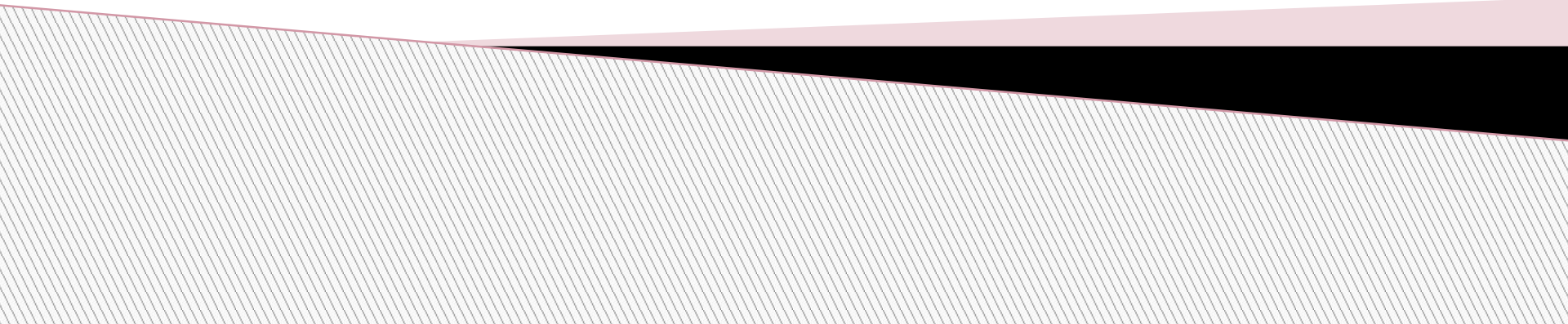
What is Statistics I
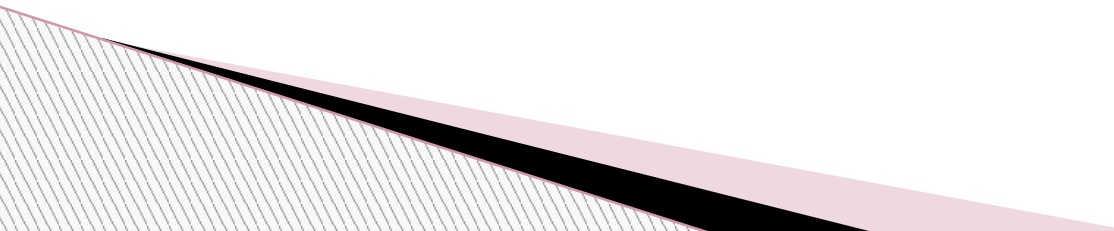
Many people understand statistics as
- *a collection of numerical facts expressed as a summarizing statement*
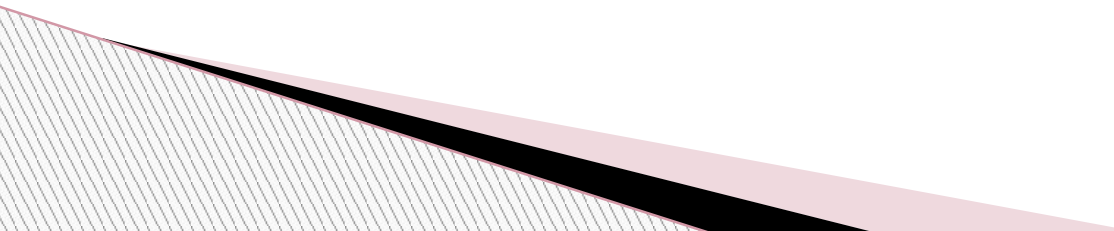
For example
- seven out of ten doctors recommend to eat ice cream when having a sore throat
- Jaromír Jágr scored the goal for 54 times in the period 2005/2006

What is Statistics II

The way we are going to understand statistics is more complex

- Statistics is a *method for dealing with data*

- *Statistics is a science of collecting, organizing, summarizing, and analyzing information (data) to draw conclusions or answer questions.*

Definition of Terms I

- **Population** – complete set of individuals, objects, or measurements having same common observable characteristic

- **Sample** – subset or part of population

- **Unit** – single member of a population

Three parts of practical statistics

- I. Data Collection
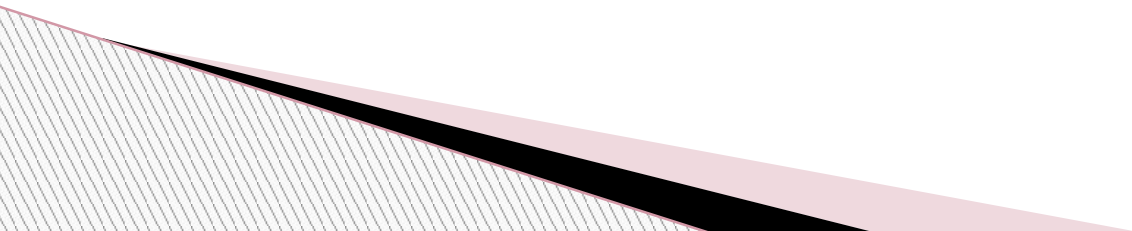
- II. Descriptive Statistics
  - consists of organizing and summarizing the information collected
  - graphical and numerical description
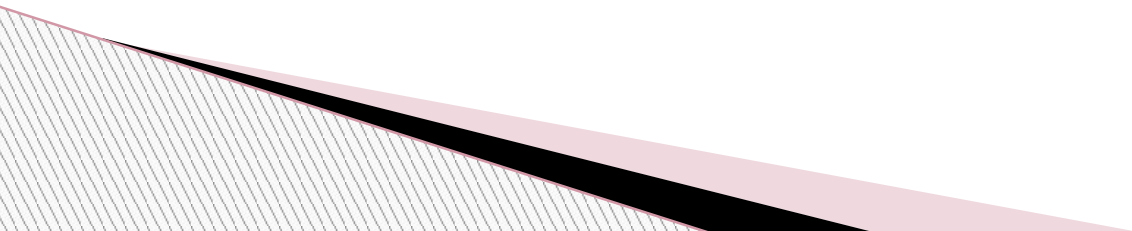
- III. Statistical Inference
  - generalizing conclusions and its evaluation using probability terms
  - sample → population

1. Data Collection

- a census

- survey samples
- designed experiments
- existing data sources

Sampling

- simple random sample

- stratified sample

- systematic sample

- cluster sample

Population and sample

population of the USA

Is it a random sample?

Note on population and sample

A population can be
- very general (all human beings)

 OR

- very narrow (all male ginger cats called Bob)

 BUT

- in praxis we collect data from samples and use these data to infer about
the population as a whole

- *e.g. election survey, medical research survey, biological experiments, computer literacy survey*

Example I

- **Population** – CULS students

- **Sample** – students of statistical course

  Is it a *random sample?*

- **Unit** – a concrete student

- **Variables** – age, height, number of siblings, hair colour, …

1. Descriptive Statistics

- methods used to describe and graph the data depend on the type of a variable

Definition of Terms II

- **Variable** – any characteristic of a person, group, or environment (it means a statistical unit) that can vary or denote a difference

(e.g. age, political ideology, pollution count)

- **Data** – numbers collected as a result of observations, interviews, this is set of information for a sample of units

- **Statistic** – number describing a characteristic of a sample (e.g. average age of a sample of CULS students, percentage of students successfully passing the exam)

Types of Variables

- **Quantitative**

    → **continuous**

    → **discrete**

- **Qualitative**

    → **nominal**

    → **ordinal**
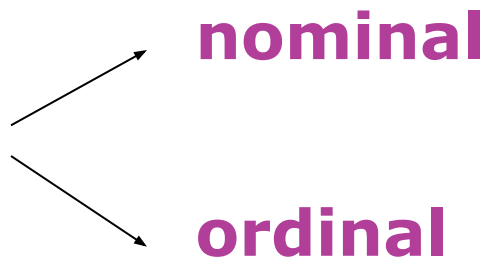
Example – types of variables

- height
- vital capacity
- number of siblings
- hair colour
- level of education

Qualitative Variables

- How to handle with qualitative variables?

- We are usually not working with original values (words), but we use variable coding.

- variable **GENDER**
  - values – female, male
  - coding – female → 1
    – male → 2

Variables Coding

- with **ordinal** variables it is recommended to use a scale that reflects the order of the values

Level of education – variable coding

correct

- primary education → 1
- apprenticeship → 2
- secondary → 3
- higher post-secondary schools → 4
- university → 5
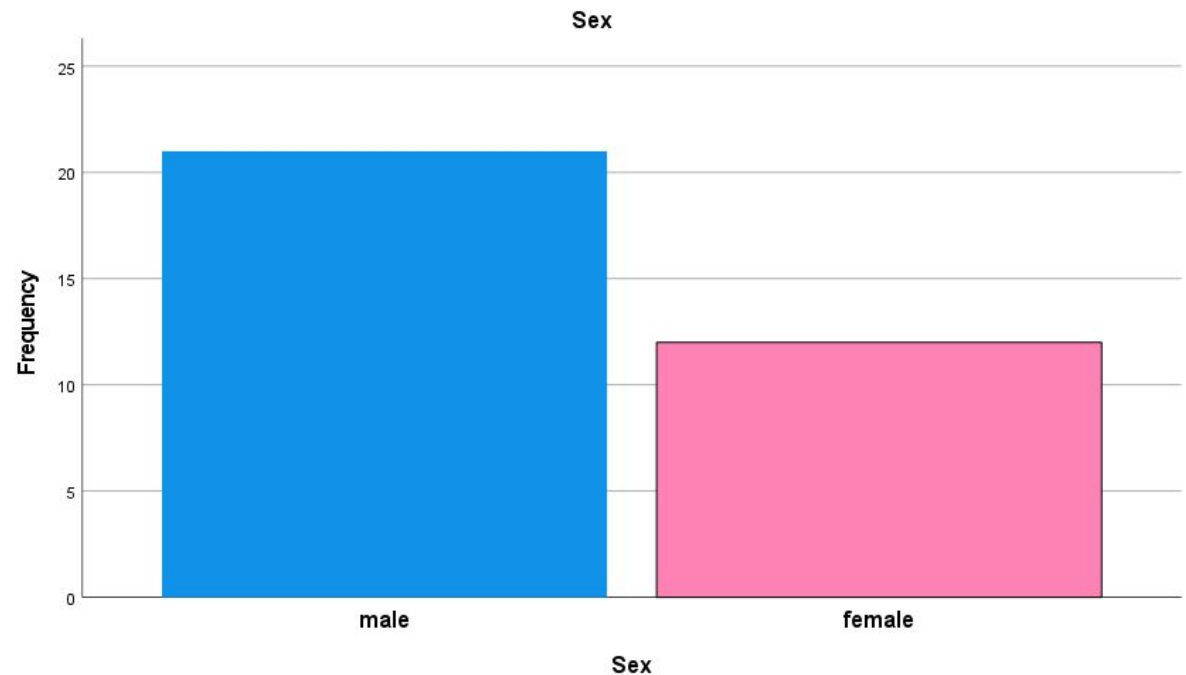
incorrect

- primary education → 2
- apprenticeship → 5
- secondary → 1
- higher post-secondary schools → 4
- university → 3

Example: Guide Dogs

qualitative variable

Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	21	63,6	63,6	63,6
	female	12	36,4	36,4	100,0
	Total	33	100.0	100.0	

Example: Guide Dogs

Scatter Plot of Points from the exam by Sex
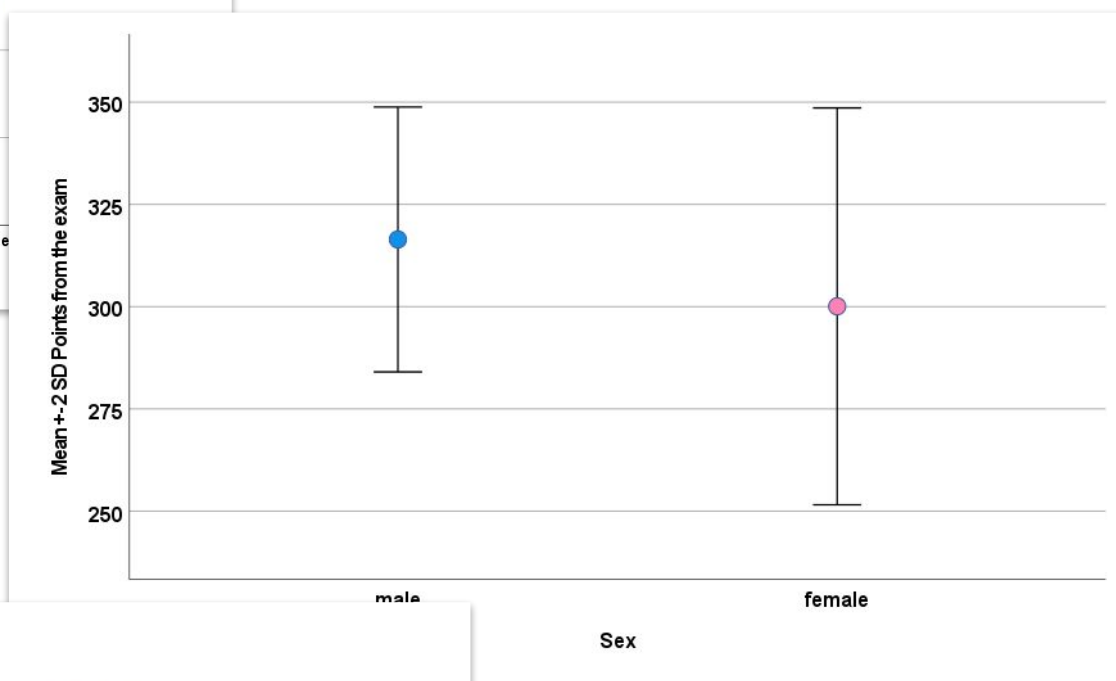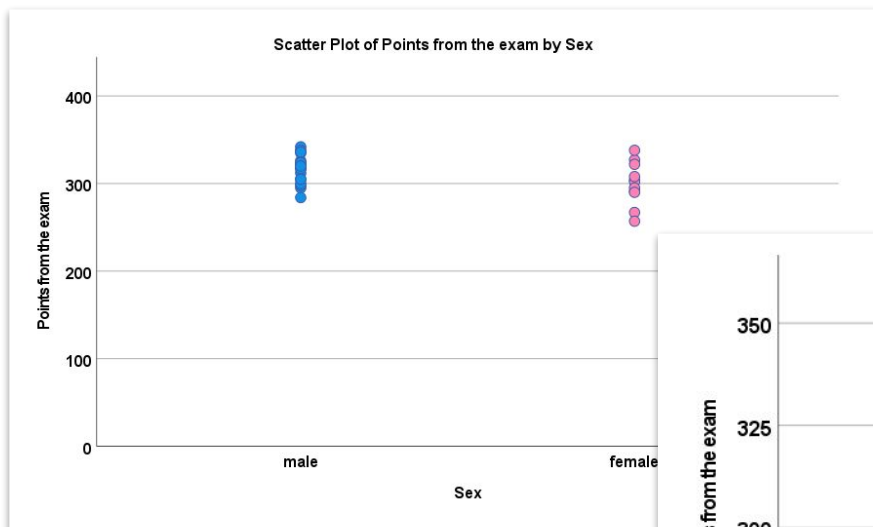
Descriptives

	Sex	Mean	Minimum	Maximum	Range
			Statistic		
Points from the exam	male	316,44	284	342	58
	female	300,09	257	338	81

ID	name	date of birth	degree	lenght of praxis (years)
3407	Zora N...		assistant	6
3698	Beata Nová	12.5.1945	docent	35
1135	Jan Spurný	3.2.1978	assistant professor	8
1247	Petra Šulcová	13.11.1962	assistant professor	23
3408	Jana Rychlá	5.4.1975	assistant	9
1266	Jaroslav Tichý	3.3.1963	docent	23
1354	Helena Veselá	16.9.1964	docent	24
1058	Petr Krátký	23.10.1956	professor	31
3245	Květa Pešková	6.4.1973	assistant professor	12

VARIABLES

ID	name	date of birth	degree	lenght of praxis (years)
3407	Zora Mlčochová	12.6.1980	**2**	6
			3	35
			2	8
1247			**2**	23
340			**1**	9
126			**3**	23
1354	Helena Veselá	16.9.1964	**3**	24
1058	Petr Krátký	23.10.1956	**4**	31
3245	Květa Pešková	6.4.1973	**2**	12

VARIABLES
with coding for variable „degree"

1 – assistant
2 –assistant professor
3 – docent
4 - professor

ID	name	date of birth	degree	lenght of praxis (years)
3407	Zora Mlčochová	12.6.1980	assistant professor	6
3698	Beata Nová	12.5.1945	docent	35
1135	Jan Spurný		assistant	8
1247	Petra Šul...			23
3408	Jana Rychlá	3.4.1973	assistant	9
1266	Jaroslav Tichý	3.3.1963	docent	23
1354	Helena Veselá	16.9.1964	docent	24
1058	Petr Krátký	23.10.1956	professor	31
3245	Květa Pešková	6.4.1973	assistant professor	12

each row represents one UNIT

ID	name	date of birth	degree	lenght of praxis (years)
3407	Zora Mlčochová	12.6.1980	assistant professor	6
3698	Beata Nová	12.5.1945	docent	35
1135	Jan Spurný		assistant	8
1247	Petra Šulcová			23
3408	Jana Rychlá	5.4.1975	assistant	9
1266	Jaroslav Tichý	3.3.1963	docent	23
1354	Helena Veselá	16.9.1964	docent	24
1058	Petr Krátký	23.10.1956	professor	31
3245	Květa Pešková	6.4.1973	assistant professor	12

DATA

- average length of praxis is 19 years

- 3 out of 9 lecturers (30%) are „docents"

STATISTICS

1. Statistical Inference

- *Inferential statistics uses methods that take the results obtained from a sample, extend them to the population, and measures the reliability of the result.*

How to draw good conclusions….