Statistics

Lecture 1

Introduction

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

David Bartl
Statistics
INM/BASTA

Outline of the lecture

- Reading list
- What is statistics?
- Why statistics “lie”
- Some history of statistics in ancient times
- Why statistics is useful in business
- How to use a computer in statistics

Reading list

Compulsory:

- KELLER, Gerald: *Statistics for Management and Economics.* 11th Edition. Cengage Learning, 2017. ISBN 978-1-337-09345-3

- SIEGEL, Andrew: *Practical Business Statistics.* 7th Edition. Academic Press, 2016. ISBN 978-0-12-804250-2

Recommended:

- **www.statsoft.com/textbook**
- http://**onlinestatbook.com**/

Reading list

Free Online Textbooks:

- Many textbooks on statistics and other disciplines can be found at https://**freetextbook.org**/
- Online Statistics Education: An Interactive Multimedia Course of Study http://**onlinestatbook.com**/
- The Electronic Statistics Textbook by StatSoft, Inc. (2013) **www.statsoft.com/textbook**
- The printed version of the latter textbook: HILL, T. & LEWICKI, P. (2007). *STATISTICS: Methods and Applications.* StatSoft, Tulsa, OK.

Reading list

Recommended:

- ÖZDEMIR, Durmuş: *Applied Statistics for Economics and Business.* 2nd Edition. Springer, 2016. ISBN 978-3-319-26495-0 (hardcover). ISBN 978-3-319-79962-9 (softcover).
- UBØE, Jan: *Introductory Statistics for Business and Economics: Theory, Exercises and Solutions.* 1st Edition. Springer, 2017. ISBN 978-3-319-70935-2 (hardcover). ISBN 978-3-319-89016-6 (softcover).
- QUIRK, Thomas: *Excel 2016 for Business Statistics: A Guide to Solving Practical Problems.* 1st Edition. Springer, 2016. ISBN 978-3-319-38958-5 (softcover).
- HERKENHOFF, Linda, FOGLI, John: *Applied Statistics for Business and Management using Microsoft Excel.* 1st Edition. Springer, 2013. ISBN 978-1-4614-8422-6 (softcover).

Reading list

Optional:

- ANDERSON, D. R., SWEENEY, D. J., WILLIAMS, Th. A., FREEMAN, J., SHOESMITH, E.: *Statistics for Business and Economics.* Cengage Learning, 2017. ISBN 978-1-4737-2656-7
- DANIEL, W. W., TERREL, J.: *Business Statistics for Management and Economics.* Houghton Mifflin, 1995. ISBN 0-395-73717-6
- WOOLDRIDGE, J. M.: *Introductory Econometrics: A Modern Approach.* Mason, OH: Thomson/South-Western, 2006. ISBN 0-324-28978-2
- VAN MATRE, J. G., GILBREATH, G. H.: *Statistics for Business and Economics.* BPI/IRWIN, Homewood, 1997. ISBN 0-256-03719-1

What is statistics?

The word “statistics” has two meanings:

- Statistics is a table, graph, or any numerical information

- **Statistics is a collection of methods and procedures dealing with information, and with numerical information in particular**

The word “statistic” has a special meaning

- Statistic is a random variable:
  - a function of the random sample
  - a formula or an algebraic expression

What is statistics for us?

The statistics is a collection, or a system, of methods and procedures dealing with numerical (quantitative) and non-numerical (qualitative) information. In particular, statistics deals with:

- **collection** of the information (census, poll, questionnaires, interviews)
- **description** of the information (structuration, storage in the computer)
- **analysis** of the information (by using statistical methods)
- **evaluation** of the information (explanation, interpretation and presentation)

Statistics lie…

"*Statistics is a particularly cunning form of a lie.*"

— an unknown English lord

"*There are three kinds of lies: lies, damned lies, and statistics.*"

— of unknown origin / Mark Twain (?)

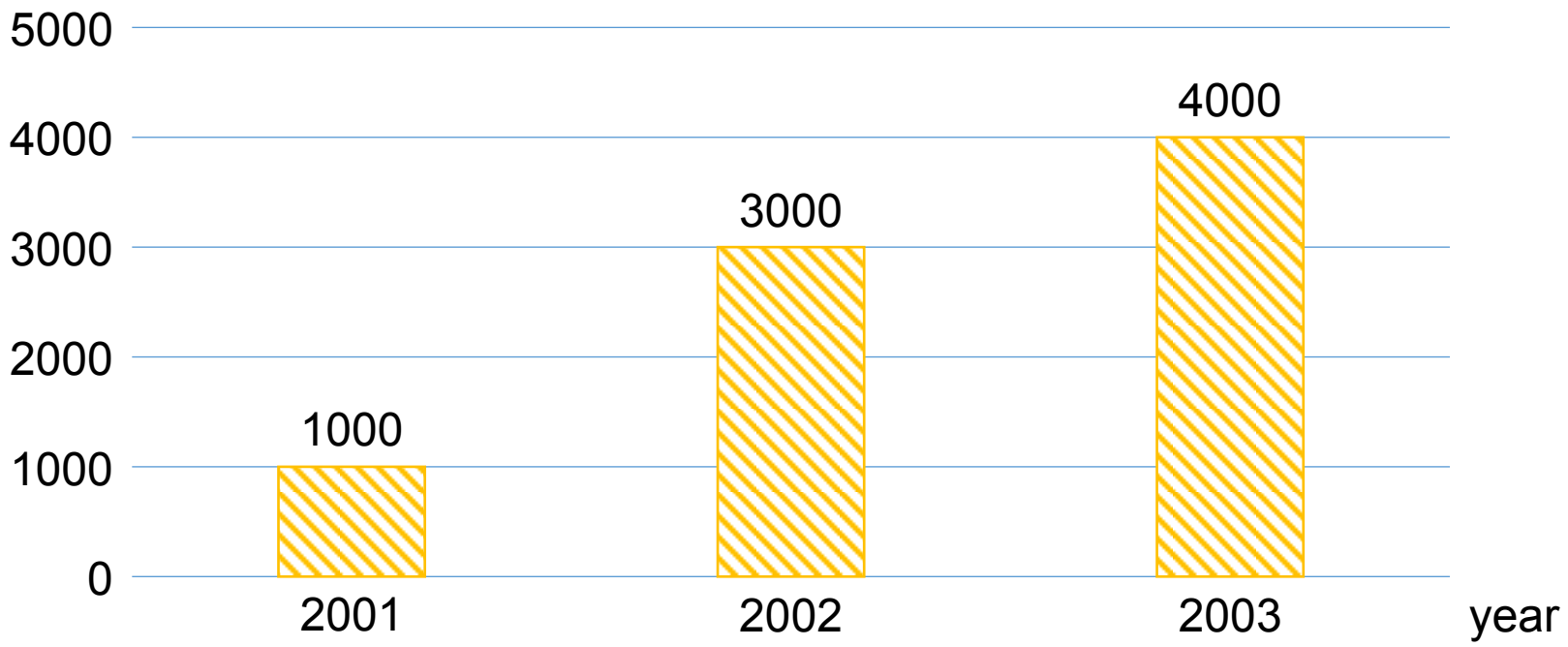
"*The only statistics you can trust are the ones you have falsified yourself.*"
"*I only believe in statistics that I doctored myself.*"

— Sir Winston Churchill

"*The statistics is boring, but provides valuable information.*"

— a song by Zdeněk Svěrák and Jaroslav Uhlíř

Example: The number of crimes in the City of XYZ



Bad news: The number of crimes increased by 300 % and by 400 %

Good news: The crime growth rate decreased by 50 %

Median salaries in various professions

profession	median salary (in CZK per month)
physicians	43 174
lawyers	41 725
programmers	41 164
scientists	34 342
teachers	26 168

Some history of statistics in ancient times

- “Statistics” in ancient Egypt, Mesopotamia, China

- The oldest “statistics” – the description of a state – the depiction of the given geographical, economic, and political state (situation)

- One of the first works on the theory of the state:

  Francesco Sansovino:

  „*del Governo et Administratione di diversi Regni, et Republichi*“

  Italy, 1583

Some modern history of statistics

Adolphe Quételet (1796–1874), a Belgian astronomer, mathematician, statistician, and sociologist:

- introduced the concept of “homme moyen”, an average man, a prototype the Nature strives for, but is unreal
- the foundation of modern statistics: the concept of the normal distribution, mean and variance

Some modern history of statistics

18th and 19th century – foundations for further development of statistics:

- Italians (three brothers)
  - Jacob Bernoulli
  - Daniel Bernoulli
  - Nicolas Bernoulli

- the French
  - Joseph-Louis Lagrange, comte de l'Empire
  - Pierre-Simon de Laplace

Some modern history of statistics

18th and 19th century – foundations for further development of statistics:

- Swiss
  - Leonhard Euler

- German
  - Carl Friedrich Gauss (Johann Carl Friedrich Gauß)

The catchword of statistics:

POPULATION = the collection of everything

Some modern history of statistics

The beginning of the 20th and 19th century – inductive statistics:

- earlier: a description of every detail
- now: conclusions about the population based on the sample

The catchword of modern statistics:

SAMPLE

Some modern history of statistics

Founders of modern statistics:

- Russians
  - Pafnuty Lvovich Chebyshev
  - Aleksandr Mikhailovich Lyapunov
  - Andrey Andreyevich Markov

- British / English
  - Ronald Fisher
  - Karl Pearson

- Polish
  - Jerzy Neyman

Historical conclusion

Correct understanding of statistical concepts and methods

is a prerequisite for successful work of any specialist in economy.

Statistics and computers

Czech Statistical Office
→ https://www.czso.cz/

Eurostat
→ https://ec.europa.eu/eurostat/

Electronic textbooks of statistics
→ **www.statsoft.com/textbook**
→ http://**onlinestatbook.com**/
→ https://**freetextbook.org**/

Software:
— Excel

Specialized statistical software:
— SPSS
— Statgraphics
— Statistica

— **gretl** = Gnu Regression, Econometrics and Time-series Library

Statistics

The purpose of statistics is to present data in a comprehensive form.

The goal is to analyse the information and reveal relations hidden in the data.

There are two approaches:

- Descriptive statistics (categorization, characteristics)
  - we shall deal with it now

- Inductive statistics (assumptions about the origin of the data, probability distributions)
  - we shall deal with it later

Data — Data unit — Data item — Observation — Dataset

Data — (plural) — measurements and observations

Data unit — one entity (e.g. a person) in the *population*, under study, about which the data are collected

Data item — a characteristics (an attribute) of a data unit (e.g. the date of birth, gender, income, …), also called a **variable**

Observation — an occurrence of a specific data item recorded about a data unit, also called a **datum** (singular of “data”)

Dataset — a complete collection of all observations

Statistical unit

Examples of statistical units:

- inhabitants of a country
- houses in a country
- flats in a country
- customers of a supermarket
- employers
- employees of a company
- organizations of a given type (such as supermarkets)
- students of a university
- electors
- products
- events (accidents, coin tosses, rolling a dice)

Statistical unit

A statistical unit is determined from three points of view at least:

- merit viewpoint (e.g. a male university student)
- spatial viewpoint (e.g. a university student in Karviná)
- time viewpoint (e.g. this year a first-year student)

Statistical unit

A census example:

- merit viewpoint: all persons
- spatial viewpoint: who are present in the territory of the Czech republic
- time viewpoint: at the crucial moment

  (the midnight between
  Friday 25th March 2011 and
  Saturday 26th March 2011)

Population — Sample — Data item

Population — a collection of all data units of the same (merit, spatial and time) specification

Sample — a selected subset of the population

Data item — a property or an attribute of a data unit of the population

Data items – **statistical variables** – are:

- **qualitative** (categorical), such as the gender, colour, taste, satisfaction
- **quantitative** (numerical), such as the revenue, price, number of customers

Qualitative data items

Qualitative (categorical) data items – **qualitative statistical variables** – are:

- nominal – only the name, such as:
  - gender (male, female)
  - colour (blue, red, yellow, green, white, black, …)

- ordinal – the values can be compared and ordered, such as
  - satisfaction:

    terrible < poor < not bad < good < excellent

  - knowledge:

    basic < advanced < expert

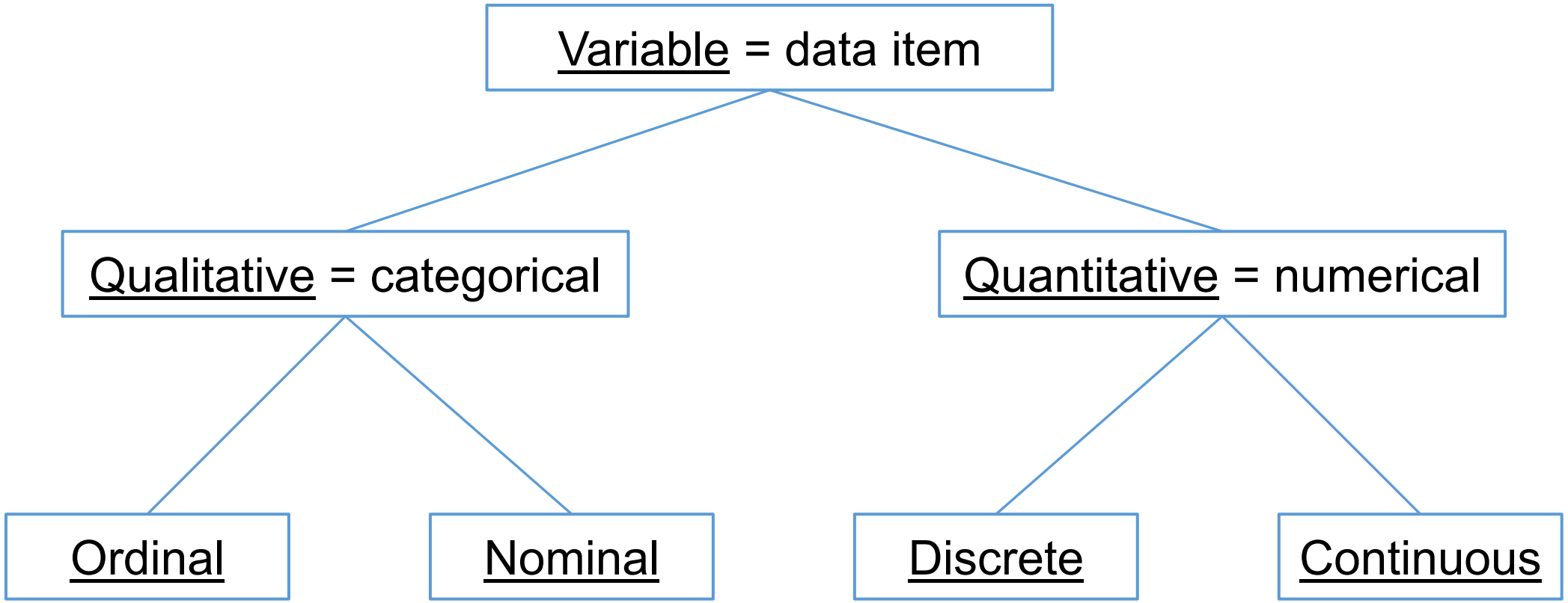
Quantitative data items

Quantitative (numerical) data items – **quantitative statistical variables** – are:

- discrete – only finite or countably infinite distinct values, such as:
  - the number of customers per day (0, 1, 2, 3, …)
  - the result of rolling a dice (1, 2, 3, 4, 5, 6)

- continuous – values from an interval (bounded or unbounded), such as
  - time between two events $(t \in (0, +\infty))$
  - unit price of goods $(p \in (0, +\infty))$
  - the proportion of indefectible products $(\rho \in [0, 1])$

Data items = Variables



Example: a Dataset where Statistical units = employees

ID	Gender	Age	Marital Status	Education	Position	Salary per Year	Evaluation
5060	M	65	divorced	secondary	worker	258800	4
1030	M	60	divorced	university	manager	630000	2
3049	M	60	married	primary	operator	436600	5
5047	M	60	widowed	primary+vocational	worker	240600	3
5061	M	60	widowed	primary+vocational	worker	241800	1
5087	M	60	widowed	secondary	worker	239500	—
5133	F	60	married	secondary	worker	241100	4
5177	F	60	widowed	secondary	worker	239600	4
3030	F	58	widowed	primary	operator	422600	1
3014	F	56	widowed	university	operator	303600	3
5012	F	56	widowed	primary+vocational	worker	223100	4
5056	M	56	divorced	primary	worker	225200	5
5101	M	56	unmarried	primary+vocational	worker	224600	4
5106	M	56	married	primary+vocational	worker	226100	7
5146	F	56	married	primary+vocational	worker	224900	3
5153	M	56	divorced	secondary	worker	224500	4
5189	M	56	married	primary+vocational	worker	224600	1
5196	M	56	widowed	primary+vocational	worker	222800	3
1031	M	55	married	university	manager	429000	—
5016	M	55	divorced	secondary	administrative officer	259000	5
5021	F	55	married	primary+vocational	worker	220200	—
5062	F	55	widowed	primary+vocational	worker	221400	5
5107	M	55	divorced	primary+vocational	worker	220500	4
5154	F	55	widowed	primary+vocational	worker	219200	5
5195	M	55	married	primary+vocational	worker	219400	6