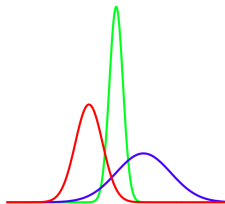




TECHNISCHE
UNIVERSITÄT
WIEN



Welcome to Statistics and Probability Theory

LV 107.254
Winter Semester 2023/24

Tijana Levajković

E105 Institute for Stochastics and Business Mathematics

Please send your comments to: tijana.levajkovic@tuwien.ac.at

Overview

- The lecture provides an **introduction** to
 - **Probability Theory** (Stochastics) = Theory of randomness
 - **Statistics** = Description of data
- The **main goal** of the course is to learn basic concepts of probability theory and statistics.
- It comprises of
 - **Relevant concepts of probability theory**
 - **Statistics**
 - **Descriptive** und **Inferential Statistics**
 - Implementation of statistical theory (basic ideas of **data analysis**)
- **TISS**: Course announcement, description and registration
- **TUWEL**: Course material and further course information

Accompanying exercises

- [UE Exercises](#) LV 107.369
 - Exercises will be on topics covered in lectures.
 - [TISS](#): Announcement and course description
 - [TUWEL](#): Course material, homework assignments and further information

Course dates/times

- Lectures

- LV Nr. 107.254
- **Mon 9:15-11:00** (10 min break), **Informatikhörsaal**

- Exercise sessions

- LV Nr. 107.369
- **Tue**, fourteen groups (please see TISS)
- The first exercise session is on **Tue 17.10.2023**.
- A brief **Introduction to R** is on **Tue 10.10.2023, 8:00-9:15** via zoom (**more details in TUWEL**).

- Consultation hours/ Problems & questions

- **Mon 11:30-12:30** and **by arrangement**
- `tijana.levajkovic@tuwien.ac.at`

Requirements

- Successful completion of **STEOP**.
- Basic knowledge of **Linear Algebra** and **Calculus**.
- **Registration** in TISS within the registration period.
- For **exercises**, a computer is needed on which
 - the statistics program **R** <https://cran.r-project.org>
 - the **RStudio** interface <https://www.rstudio.com/>

are installed. **R** is a **free open source software** and it is easy to install.

- **No prior knowledge of R is needed!**

Literature



L. Fahrmeir, C. Heumann, R. Kuünstler, I. Pigeot, G. Tutz

Statistik – Der Weg zur Datenanalyse.
Springer, Berlin Heidelberg
(available online via TU library)



J. Hedderich, L. Sachs

Angewandte Statistik – Methodensammlung mit R.
Springer, Berlin, Heidelberg
(available online via TU library)



N. Sharpe, R. De Veaux, P. Velleman

Business Statistics.
Pearson, Boston.



G. Kersting, A. Wakolbinger

Elementare Stochastik.
Birkhäuser, Basel
(available online via TU library)



M. Messer, G. Schneider

Statistik: Theorie und Praxis im Dialog.
Springer, Berlin
(available online via TU library)



W. Gurker

Statistik und Wahrscheinlichkeitstheorie.
TU-Verlag, Vienna
(available purchasable in Grafisches Zentrum)

Other books on the topics of the course can also be used!

Grading

- **One final comprehensive written exam** (based on the course material)
 - Multiple-choice questions
 - Four possible answers to each question.
Exactly **one** correct answer.
 - The exam will be **90 minutes** long
 - A **non-programmable** calculator and a **two-sided handwritten A4 sheet** may be used
 - Computers, smartphones, tablets, notes, books, etc., as well as discussions and consultations are **prohibited** during the exam!
- **Three** examination dates are offered:
 - 1st date **Mon, 29.01.2024 at 9:00**
 - 2nd date **Mon, 27.02.2024 at 9:00**
 - 3rd date **Mon, 07.05.2024 at 13:00**
- Registration via TISS

Note: Students who **do not deregister** and **do not attend** the exam will be **blocked** for registrations for the following examination term.

Probability vs. Statistics

Different subjects: both about **random processes**

- **Deterministic/Non-probabilistic**

- We are **certain** of the value of a variable (the outcome of a variable is fixed)
- tomorrow's date, the cost to hire the venue

- **Probability theory**

- A quantity takes values without **certainty**
 - We know the price of a stock today, do we know its price tomorrow, in a week, in a month? Due to a variety of (partially unknown) factors affecting stock prices, there will always be **uncertainty** in its future value.
 - If the random mechanism that generates the stock price were known **exactly**, calculating probabilities of its possible values is what **probability theory** does.

- **Statistics**

- The random mechanism is **unknown**. In order to **learn and understand the random mechanism**, one can collect data and analyze them.
- Three stages in statistical studies:
 - Data collection with the aim of drawing probabilistic conclusions
 - Data analysis (initial descriptive statistics, fitting an appropriate statistical model and estimating its parameters)
 - Statistical inference

Probability vis-a-vis Statistics

- Probability

- the population (the whole) is known and the probabilities about members of the population (subsets) are computed.
 - logically self-contained
 - some rules for calculating probabilities
 - one correct answer

- Statistics

- members of the population (data, subsets) are observed and we infer about the population (whole).
 - messier and more of an art
 - collecting/observing experimental data with the aim of drawing probabilistic conclusions
 - no single correct answer

Probability vs. Statistics: Examples

1. Probability example

We toss a **fair** coin (equal probability of heads or tails) 100 times. What is the probability of getting 60 or more heads?

- There is only one answer (about 0.028444) and we will learn how to compute it.
- ★ The random process is known (probability of heads is 0.5)
- ★ The aim is to determine the probability of a certain result (at least 60 heads) from the random process

2. Statistics example

We have a coin of **unknown** origin. To investigate whether it is fair, we toss it 100 times and count the number of heads. Suppose we count 60 heads. As statisticians, we have to draw a conclusion (inference) from this data.

- There are many ways to proceed. Different statisticians can draw different conclusions.
- ★ The outcome is known (60 heads).
- ★ The aim is to explain the unknown random process (the probability of heads).

Probability Theory and Statistics

- Wide range of applications
 - medicine, physical sciences, engineering, the social sciences, the life sciences, economics and computer science
 - testing one medical treatment versus another (or a placebo)
 - measures of genetic linkage
 - the search for elementary particles
 - machine learning for vision or language
 - gambling probabilities and strategies
 - climate modeling
 - economic forecasting
 - epidemiology
 - marketing
 - ...
 - We are going to study toy models like coins and dice.
 - The coin toss is a realistic model for all situations with two possible outcomes: success or failure of a treatment, bet, etc.

Topics

I Probability theory

- Basic probability theory
 - counting (multiplication rules, permutations, combinatorics), probability space, the calculus with probabilities, conditional probability and independence, Law of total probability, the Bayes theorem
- Random variables
 - discrete and continuous random variables, distribution functions, probability mass functions (pmfs), probability density functions (pdfs), common families of distributions
 - moments (expectation and variance), properties, transformations, independence, covariation and correlation
- Random sample
 - properties of a random sample, sum of random variables from random sample, sampling from normal distribution, properties of sample mean and sample variance, order statistics
 - Law of Large Numbers, Central Limit Theorem, Normal approximations.

Topics

II Statistics

- Descriptive statistics
 - numerical summaries
 - elementary statistics, empirical distribution
 - graphical representations (histograms, tables, diagrams, boxplots)
- Inferential statistics
 - interval estimation (confidence intervals)
 - hypothesis testing, one sample t -test, two sample t -test, p -value for these tests
 - nonparametric tests, analysis of variance and multiple testing
- Linear regression
 - correlation
 - simple linear model
 - regression line, coefficient of determination

Alternative in summer semester

- Statistics and Probability Theory
 - LV 107.273
 - Sommer semester 2024
 - The course description in TISS



Best wishes for a successful semester!