Statistical data analysis - Course ID: T510028101

Content

A first introduction to probability and statistics, this course provides background to understand and produce statistical analysis, including design of experiments, estimation, confidence intervals and hypothesis testing. Applicability and limitations of these methods will be illustrated in the light of modern data sets and manipulation using statistical software. Precepts are based on real data analysis.

Learning objectives - Knowledge

- explain relevant data types and their representation for statistical analysis
- explain probabilities and random variables
- explain distributions of random variables
- explain inference and hypothesis testing
- explain how data may be collected from experiments involving randomness

Learning objectives - Skills

- choose an appropriate experimental design in respect to a given task
- perform statistical analyzes on data collected
- use a statistical tool for analysis and visualization of data

Learning objectives - Competences

- use statistical methods and tools to interpret experimental data

Recommended literature

OpenIntro Statistics (David Diez, Mine Cetinkaya-Rundel, Christopher Barr).
It can be downloaded for free here: https://www.openintro.org/book/os/

Other recommended books:

- Applied statistics and probability for engineers / Douglas C. Montgomery, George C. Runger —3rd ed. ISBN 0-471-20454-4
- A Handbook of Statistical Analyses Using R / Brian S. Everitt, Torsten Hothorn , ISBN 1420079336

Course organization

Thursdays 10-12: Main lecture
Thursdays 12-14: Exercise time (independent work on exercises with help of instructors).

Lecture planning

Lesson	Week	Date	TOPICS
1	36	9/Sep	Introduction to the course Descriptive statistics – part I
2	37	16/sep	Descriptive statistics – part II
3	38	23/Sep	Probability distributions
4	39	30/Sep	Hypothesis testing (one sample)
5	40	7/Oct	Hypothesis testing (two samples)
-	41	14/Oct	NO CLASS
-	42	21/Oct	NO CLASS (Autum holidays)
6	43	28/Oct	ANOVA one-way
7	44	4/Nov	R class (hypothesis testing + ANOVA)
8	45	11/Nov	ANOVA two-way Notions of experimental design
9	46	18/Nov	Regression analysis
10	47	25/Nov	Multiple regression
11	48	2/Dec	Logistic regression
12	49	9/Dec	Recap of statistical concepts, questions' time, etc

Lesson 1 and 2: Descriptive statistics

1.1. Statistics
- What is statistics?
- Definitions: Variable, data, population and sample
- Descriptive and Inferential statistics

1.2. Variables and Types of Data

Qualitative, quantitative, categorial, ordinal, discrete, continuous

1.3. Data organization, histograms and bar charts

1.4. Measures of:
- Central Tendency (Location):
    - Mean
    - Median
    - Mode
    - Midrange
- Variation (Dispersion)
    - Range
    - Variance
    - Standard Deviation
    - Coefficient of Variation
- Position
    - Z-score
    - Percentile
    - Decile and Quartile
    - Outlier

1.4. Data representation
- Box-plots

- Frequency distributions
- Graphs: Time series, Pie graphs, Scatter plots

1.5. Shapes of frequency distributions
- Shapes of distributions
- Skewness
- Kurtosis

Lesson 3: Probability distributions

2.1. Probability
- Probability experiment, outcome and sample space
- Classical and empirical probability
- Law of large numbers

2.2. Random variables

2.3. Probability distributions and cumulative probability distributions

2.4. Types of probability distributions
- Discrete: Bernoulli, Binomial, Negative Binomial, Poisson, Geometric and Multinomial Distribution
- Continuous

2.5. Shapes of distributions: Skewness and kurtosis

2.6. Some distributions used in inferential statistics
- Normal distribution
- Chi square distribution
- F (Fisher) distribution
- t-Student distribution

Lesson 4: Hypothesis testing – One sample

- An example
- Hypothesis testing for the mean (n<30)
- Confidence Intervals for the mean (n < 30)
- Confidence Intervals for Standard Deviations
- Type error I and type error II
- Large samples

Lesson 5: Hypothesis testing – Two samples

- An example
- Comparison of means (independent samples)
- Comparison of variances
- Paired data test (dependent samples)

Lesson 6: One-way ANOVA

- Analysis of variance: Example

- One-Way Analysis of Variance
- The Least significant difference intervals

Lesson 7: Statistics in R

- Using R in statistics
- Hypothesis testing in R
- One-way ANOVA in R

Lesson 8: Two-way ANOVA

- Two-Way Analysis of Variance
- Interactions
- Experimental design
- Sample size
- Randomization

Lesson 9: Regression analysis

- Regression analysis
- Correlation
- Determining the regression line
- Influential points
- Types of variation
- Coefficient of determination
- Prediction interval

Lesson 10: Multiple regression

- Simple regression vs. Multiple regression
- Covariates
- Variable selection
- Interpreting results

Lesson 11: Logistic regression

- Logistic regression vs. linear regression
- Logit function
- Coefficients and odds ratio
- Interpreting results

Lesson 12: Experimental design and recap of statistical concepts

- Recap of everything we learned during the semester
- Miscellaneous exercises
- Questions about exam, etc.