Course 02402 Introduction to Statistics

Lecture 13: Summary of the course

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Overview - the 12 lectures

1. Chapter 1: Simple graphics and summary statistics
2. Chapter 2: Discrete distributions
3. Chapter 2: Continuous distributions
4. Chapter 3: One-sample confidence intervals
5. Chapter 3: One-sample hypothesis testing
6. Chapter 3: Two-sample statistics
7. Chapter 4: Statistics by simulation
8. Chapter 5: Simple linear regression analysis
9. Chapter 6: Multiple linear regression analysis
10. Chapter 7: Inference for proportions
11. Chapter 8: One-way analysis of variance
12. Chapter 8: Two-way analysis of variance

Overview

1. Chapter 1: Simple graphics and summary statistics
2. Chapter 2: Discrete distributions
3. Chapter 2: Continuous distributions
4. Chapter 3: One-sample confidence intervals
5. Chapter 3: One-sample hypothesis testing
6. Chapter 3: Two-sample statistics
7. Chapter 4: Statistics by simulation
8. Chapter 5: Simple linear regression analysis
9. Chapter 6: Multiple linear regression analysis
10. Chapter 7: Inference for proportions
11. Chapter 8: One-way analysis of variance
12. Chapter 8: Two-way analysis of variance

Chapter 1: Simple graphics and summary statistics

- Look at data as it is (descriptive statistics).
- Summary statistics:
  - Mean $\bar{x}$
  - Standard deviation s, variance s^2
  - Median, upper and lower quartiles
- Simple graphics:
  - Scatter plot (xy plot)
  - Histogram, cumulative distribution
  - Box plots, bar charts, pie charts

Overview

Chapter 2: Discrete distributions

- General concepts:
  - Definition of a stochastic variable
  - Density function
  - Distribution function
  - Mean and variance
- Specific distributions:
  - The binomial distribution
  - The hypergeometric distribution
  - The Poisson distribution

Overview

Chapter 2: Continuous distributions

- General concepts:
  - Density function, distribution function
  - Mean, variance
  - Calculation rules for stochastic variables
- Specific distributions:
  - Normal
  - Log-normal, uniform, exponential

Overview

Chapter 3: One-sample confidence intervals

- General concepts
  - Estimation, confidence intervals
  - Population and a random sample
  - Sampling distributions (t and χ^2)
  - Central Limit Theorem (CLT)
- Specific methods, one sample:
  - Confidence intervals for the mean
  - Confidence intervals for the variance (and standard deviation)

Overview

Chapter 3: One-sample hypothesis testing

- General concepts:
  - Hypotheses, p-value, significance level
  - Type I and Type II error, power
- Specific methods, one sample:
  - t-test for the mean
  - Sample size needed for desired power
  - Normal QQ-plot

Overview

Chapter 3: Two samples

- Specific methods, two samples:
  - Test and confidence interval for the mean difference (t-test)
- Specific methods, two PAIRED samples:
  - "Take difference" $\Rightarrow$ "One sample"
- Planning for precision and/or power
  - One-sample confidence interval: sample size needed for desired precision
  - One-sample hypothesis test: sample size needed for desired power (or other combinations)
  - Two-sample hypothesis test: sample size needed for desired power (or other combinations)

Overview

Chapter 4: Statistics by simulation

- Introduction to simulation
- Error propagation rules
- Bootstrapping
  - Parametric
  - Non-parametric
  - Confidence intervals (and hence also hypothesis testing)
- Specific situations (4 versions of confidence intervals):
  - One-sample and two-sample data
  - Parametric and non-parametric

Overview

Chapter 5: Simple linear regression analysis

- Two quantitative variables, x and y.
- Calculating least squares line.
- Inference for a simple linear regression model:
  - Statistical model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
  - Confidence intervals and tests for β_0 and β_1.
  - Confidence interval for the expected line.
  - Prediction interval.
- r and r^2
  - r describes the strength of a linear relation.
  - r^2 expresses the proportion of the y variability explained by the linear relation.

Overview

Chapter 6: Multiple linear regression analysis

- Many quantitative variables, $x_1, \cdots, x_p$ and y.
- Calculating least squares fit
- Inference for a multiple linear regression model
  - Statistical model: $Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i} + \varepsilon_i$
  - Confidence intervals and tests for β_i, $i = 1, \ldots, p$.
  - Confidence interval for the expected fit.
  - Prediction interval.
- r^2 expresses the proportion of the y variability explained by the linear relation.

Overview

Chapter 7: Inference for Proportions

- Specific methods, one, two and $k > 2$ samples
  - Binary/categorical response
- Estimation and confidence intervals for proportions
  - Large sample vs. small sample methods
- Hypotheses for one proportion
- Hypotheses for two proportions
- Analysis of contingency tables (χ^2-test) (all expected > 5)

Overview

Chapter 8: One-way analysis of variance

- Specific methods, k INDEPENDENT samples
- One-way analysis of variance
  - Compares the means of the groups
  - ANOVA-table: $SST = SS(Tr) + SSE$
  - F-test
  - Post hoc test: Pairwise t-test with/without Bonferroni correction.

Overview

Chapter 8: Two-way Analysis of Variance

- Block design - two-way analysis of variance
- ANOVA-table: $SST = SS(Tr) + SS(Bl) + SSE$
  - SST, $SS(Tr)$ and $SS(Bl)$ calculated like for one-way ANOVA
  - $SSE = SST - SS(Tr) - SS(Bl)$
- F-test
- Post hoc test: Pairwise t-test with/without Bonferroni correction.

Agenda