Statistics
Course booklet

Statistics for college and university students. Contains descriptive statistics, probability theory, estimation, hypothesis testing, regression analysis, and more.

SOWISO

About SOWISO

SOWISO offers:
- a homework, practice, and **learning environment**;
- personalised **feedback** on all answer attempts;
- different **testing and assessment** tools;
- customisable **mathematics courses** with explanations, examples, and endless **randomised practice exercises**;
- an authoring tool to **create original material**;
- **learning analytics** giving detailed insight into student performance;
- **integration** with your LMS/VLE.

Our learning environment guides students along as they solve problems. When doing exercises, students can enter open answer calculations or mathematical formulas. The software will analyse their answer and provide feedback and hints helping the student understand the next step in the solution process, and/or highlight any mistakes they made.

SOWISO increases student engagement and saves teachers time checking and grading!

Pricing

SOWISO partners with higher education institutions on a SAAS licensing basis.

The cost for the platform starts at € 5.50 per student per year, with an additional per student per year fee of € 7.50 per course.

Alternatively students can pay for their own license in our webshop.

Our digital courses are a fully interactive alternative to paper books and offer a personalised and adaptive learning experience that fits today's generation of students.

How are courses structured?

The courses are structured in chapters and subchapters consisting of units. The unit subjects are listed in more detail on the following pages.

Each unit consists of (at least) one theory page and one package of exercises.

Theory pages contain explanations, (randomised) examples and visualisations and (interactive) graphs.

The course contains over 500 **exercises**, of which a lot are randomised, allowing for endless practicing. Exercises include hints and feedback for the students while solving the exercises.

COURSE CONTENT

Chapter 1: Descriptive statistics (21 topic)

1. *Types of data and measurement (6 topics)*
   - a. Qualitative and quantitative variables
   - b. The hierarchy of measurement scales
   - c. Nominal scale
   - d. Ordinal scale
   - e. Interval scale
   - f. Ratio scale

2. *Frequency distributions (5 topics)*
   - a. Frequency distributions
   - b. Frequency distribution tables
   - c. Frequency distribution graphs
   - d. Shape of a distribution
   - e. Measures of location I: Quantiles

3. *Measures of central tendency (5 topics)*
   - a. Mode
   - b. Median
   - c. Mean
   - d. Central tendency and the shape of a distribution
   - e. Sensitivity to outliers

4. *Measures of variability (4 topics)*
   - a. Range, interquartile range , and the five-number summary
   - b. Interquartile range rule for identifying outliers
   - c. Deviation from the mean and the sum of squares
   - d. Variance and standard deviation

5. *Measures of location II: Z-scores (1 topic)*
   - a. Z-scores

Chapter 2: Correlation (4 topics)

6. *Correlation (4 topics)*
   a. Displaying the relationship between two variables
   b. Measuring the relationship between two variables
   c. Direction of a linear relationship: Covariance
   d. Strength of a linear relationship: Pearson Correlation Coefficient

Chapter 3: Probability (19 topic)

7. *Randomness (5 topics)*
   a. Sets, subsets and elements
   b. Random experiments
   c. Sample space
   d. Events
   e. Complement of an event

8. *Relationship between events (4 topics)*
   a. Mutual exclusivity
   b. Difference
   c. Intersection
   d. Union

9. *Probability (9 topics)*
   a. Definition of probability
   b. Probability of the complement
   c. Conditional probability
   d. Independence
   e. Probability of the intersection
   f. Probability of the union
   g. Probability of the difference
   h. Law of total probability
   i. Bayes' theorem

10. *Contingency tables (1 topic)*
   a. Interpreting contingency tables

THEORY EXAMPLES

📄 Setting the Criteria for a Decision

Once the hypotheses of the test have been formulated, the next step is to set the criteria for a decision. This should always be done <u>before</u> the sample data is collected.

Specifically, we need to determine what values of the sample statistic will lead to the rejection of the null hypothesis. Because a sample provides an incomplete picture of a population, some discrepancy between a sample statistic and its corresponding population parameter is to be expected.

How much discrepancy is reasonable to expect can be derived from the sampling distribution of the sample statistic under the null hypothesis. If the null hypothesis is true, it is likely that the sample statistic will be relatively close in value to the mean of the sampling distribution.

As the difference between a sample statistic and the mean of the hypothesized sampling distribution increases, your confidence in the null hypothesis being true should decrease. If you observe a sample statistic that is *extremely unlikely* to occur given that the null hypothesis is true, this should lead to the rejection of the null hypothesis.In order to formalize what constitutes as an extremely unlikely result, the *significance level* of the test needs to be set.

High-probability
outcomes

Extreme,
low-probability
outcomes

Extreme,
low-probability
outcomes
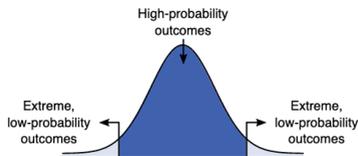
Simple Random Sampling

Definition	Simple random sampling

Simple random sampling is a basic sampling method where the individuals of the sample are selected from the population as a whole, usually with the help of a random number generator.

Each individual is chosen entirely by chance and each member of the population has an *equal probability* of being selected.

Population

Random Number
Generator
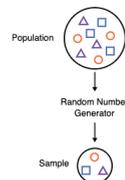
Sample

Stratified random sample

Definition	Stratified sampling

When taking a **stratified random sample**, the population of interest is first divided into groups of individuals that share at least one common characteristic. These subgroups are called **strata**.

A simple random sample is then drawn from each of the strata and these samples are combined into a final sample

Population

Strata

SRS   SRS   SRS
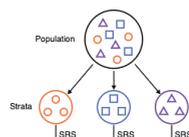
Chapter 4: Probability distributions (11 topics)

11. *Probability models (2 topics)*
    a. Discrete probability models
    b. Continuous probability models

12. *Random variables (5 topics)*
    a. Random variables
    b. Probability distributions
    c. Expected value of the random variable
    d. Variance of a random variable
    e. Sums of random variables

13. *Common distributions (4 topics)*
    a. The binomial distribution
    b. Expected value and variance of a binomial random variable
    c. The normal distribution
    d. The normal probability distribution

Chapter 5: Sampling (5 topics)

14. *Sampling and sampling methods (2 topics)*
    a. Sampling and unbiased sampling methods
    b. Biased sampling methods

15. *Sampling distributions (3 topics)*
    a. Sampling distributions
    b. Sampling distribution of the sample mean
    c. Sampling distribution of the sample proportion

Chapter 6: Parameter estimation and confidence intervals (4 topics)

16. *Parameter estimation and confidence intervals (4 topics)*
    a. Parameter estimation

b. Constructing a 95% confidence interval for the population mean
c. Confidence interval for the population mean
d. Confidence interval for the population proportion

Chapter 7: Hypothesis testing (26 topics)

17. Hypothesis testing (10 topics)
   a. Hypothesis testing procedure
   b. Formulating the research hypotheses
   c. Two-tailed vs one-tailed testing
   d. Setting the criteria for a decision
   e. Computing the test statistic
   f. Computing the p-value and making a decision
   g. Assumptions of the Z-test
   h. Connection between hypothesis testing and confidence intervals
   i. Errors in decision making
   j. Statistical power

18. Hypothesis test for a population proportion (4 topics)
   a. Hypotheses of a population proportion test
   b. Large-sample proportion test: Test statistic and p-value
   c. Small-sample proportion test: Test statistic and p-value
   d. Hypothesis test for a proportion and confidence intervals

19. One-sample t-test (3 topics)
   a. One-sample t-test: Purpose, hypotheses, and assumptions
   b. One-sample t-test: Test statistic and p-value
   c. Confidence interval for μ when σ is unkown

EXERCISE EXAMPLES

✎ Finding the Regression Equation

Consider the following 5 pairs of data points:

X	Y
1	1
3	7
10	9
2	3
4	5

Find the regression equation for this dataset. Give your answer in the form $\hat{Y} = aX + b$ or $\hat{Y} = aX - b$.

Round the regression coefficient and intercept to 2 decimal places.
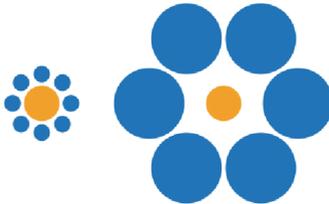
$\hat{Y} =$ ___ X (▾) ___

✔ Check   ▣ Theory   ⚒ Solution

✎ One-sample t-test: Test Statistic and p-value

The image below is an example of the Ebbinghaus illusion.

Although both orange circles are exactly the same size, the circle on the right appears to be smaller.

In an attempt to prove the effect of this illusion, a researcher prepares an image where both circles are exactly 22 cm in diameter. The example is then shown to individual participants who are told that the left circle is 22 cm in diameter, after which they are asked to estimate the diameter of the right circle.

The researcher plans on using a *one-sample t-test* to prove that people significantly underestimate the diameter of the right circle.

sowiso.com   info@sowiso.com

Chapter 8: Testing for differences in mean and proportion (9 topics)

20.   *Paired samples t-test (3 topics)*
  a. Paired samples t-test: Purpose, hypotheses and assumptions
  b. Paired samples t-test: Test statistic and p-value
  c. Confidence interval for a mean difference

21. *Independent samples t-test (3 topics)*
  a. Independent samples t-test: Purpose, hypotheses and assumptions
  b. Independent samples t-test: Test statistic and p-value
  c. Confidence interval for the difference between two independent means

22. *Independent proportions z-test (3 topics)*
  a. Independent proportion z-test: Purpose, hypotheses and assumptions
  b. Independent proportion z-test: Test statistic and p-value
  c. Confidence interval for the difference between two independent proportions

Chapter 9: Regression analysis (7 topics)

23. *Simple linear regression (4 topics)*
  a. Introduction to regression analysis
  b. Residuals and total squared error
  c. Finding the regression equation
  d. The coefficient of determination
  e. Regression analysis and causality

24. *Multiple linear regression (3 topics)*
  a. Multiple linear regression
  b. Overfitting and multicollinearity
  c. Dummy variables

Chapter 10: Categorical association (4 topics)

Missing something? SOWISO allows teachers to create their own content in our authoring environment.

SOWISO