LES FACULTÉS
DE L'UNIVERSITÉ
CATHOLIQUE DE LILLE

Models examples

EXPERIMENTAL DESIGN SAMPLING & TESTING

Baptiste Mokas

baptiste.mokas@gmail.com

weeki.io/dynamical-système

linktr.ee/baptistemokas

+33 7 69 08 54 19

KNOWLEDGE TREE

Part 1: Inference, Hypothesis Testing, Experimental Design and Application

KNOWLEDGE TREE

Part 1: Fundamentals of Hypothesis Testing

1.1 Introduction to Hypothesis Testing

- Definition and significance of hypothesis testing
- Role of hypothesis testing in research and decision-making

1.2 Types of Hypotheses

- Null hypothesis (H0) and alternative hypothesis (H1 or Ha)
- One-tailed vs. two-tailed tests
- Prediction of future outcomes

1.3 Errors in Hypothesis Testing

- Type I error (False Positive) and Type II error (False Negative)
- Relationship between significance level (α) and power of the test (β)

1.4 P-value and Significance Level

- Interpretation of the p-value
- Setting and understanding the significance level (commonly at 0.05)

Part 2: Steps and Considerations in Hypothesis Testing

2.1 Setting Up Hypotheses

- Formulating null and alternative hypotheses
- Identifying the appropriate test statistic

2.2 Choice of Significance Level

- Impact of α on decision making
- Considerations based on research context

2.3 Conducting the Test

- Calculation of the test statistic
- Determining the critical value or critical region

2.4 Making a Decision

- Rejecting or failing to reject the null hypothesis
- Interpretation in the context of the research question

Part 3: Assumptions and Limitations

3.1 Assumptions in Hypothesis Testing

- Common assumptions such as normality, independence, and homoscedasticity
- Implications of violating assumptions

3.2 Power and Sample Size

- The concept of statistical power
- Impact of sample size on hypothesis tests

3.3 Multiple Comparisons Problem

- Risk of Type I error with multiple tests
- Correction techniques (e.g., Bonferroni correction)

3.4 Limitations of p-values

- Misinterpretations and misconceptions
- Recent critiques and discussions in the statistical community

Part 4: Advanced Topics and Practical Application

4.1 Non-parametric Tests

- When and why to use non-parametric tests
- Key differences from parametric tests

4.2 Bayesian Hypothesis Testing

- Comparison with traditional (frequentist) hypothesis testing
- Bayes factor and its interpretation

4.3 Resampling Methods

- Introduction to bootstrapping and permutation tests
- Application in hypothesis testing

4.4 Reporting and Interpreting Results

- Importance of transparency in reporting
- Effect sizes, confidence intervals, and other essential metrics

OUR STORY LINE

Probability and Statistics

STEP -1 _ PROGRAM INTRODUCTION STEP 0 _ FOUNDATIONS STEP 1 _ THEORY OF SYSTEMS	STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY STEP 3 _ DATA OBSERVATION	STEP 4 _ INFERENCE & ESTIMATION THEORY	STEP 5 _ LINEAR MODEL EXAMPLES	STEP 6 _ OTHER MODEL EXAMPLES	STEP 7 _ NON LINEAR MODELS

OUR STORY LINE

Probability and Statistics

STEP -1 _ PROGRAM INTRODUCTION
- -1 - PROGRAM INTRODUCTION

STEP 0 _ FOUNDATIONS
- 0.1 - ELEMENTS OF CALCULUS & TOOLS
- 0.2 - EPISTEMOLOGY & THEORY OF KNOWLEDGE

STEP 1 _ THEORY OF SYSTEMS
- 1.1 - DYNAMICAL SYSTEMS
- 1.2 - COMPLEX ADAPTIVE SYSTEMS

STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY
- 2.1 - MEASURE THEORY
- 2.2 - PROBABILITY THEORY
- 2.3 - USUAL PROBABILITY DISTRIBUTIONS
- 2.4 - ASYMPTOTIC STATISTICS
- 2.5 - STOCHASTIC PROCESS & TIME SERIES
- 2.6 - INFORMATION GEOMETRY

STEP 3 _ DATA OBSERVATION
- 3.1 - DESCRIPTIVE STATISTICS & DATAVIZUALISATION
- 3.2 - EXPLORATORY DATA ANALYSIS

STEP 4 _ INFERENCE & ESTIMATION THEORY
- 4.1 - PARAMETERS ESTIMATIONS & LEARNING
- 4.2 - EXPERIMENTAL DESIGN & HYPOTHESIS TESTING
- 4.4 - DECISION TREES & MODEL SELECTION
- 4.5 - BAYESIAN INFERENCES

STEP 5 _ LINEAR MODEL EXAMPLES
- 5.1 - SIMPLE LINEAR REGRESSION
- 5.2 - MULTIPLE LINEAR REGRESSION
- 5.3 - OTHER REGRESSIONS MODELS

STEP 6 _ OTHER MODEL EXAMPLES
- 6.1 - USUAL UNIVARIATE TESTING
- 6.2 - USUAL MULTIVARIATE TESTING
- 6.3 - NON PARAMETRIC STATISTICS

STEP 7 _ NON LINEAR MODELS
- 7.1 - PROBABILISTIC GRAPHICAL MODELS
- 7.2 - PERCOLATION THEORY
- 7.3 - SPATIAL STATISTICS
- 7.4 - EXTREM VALUE THEORY
- 7.5 - AGENT BASED MODELING
- 7.6 - NETWORK DYNAMICS

Null and Alternative Hypotheses

Part 2: Hypothesis Testing - Theory and Application

2.1 Foundations and General Definitions of Hypothesis Testing

Hypothesis testing is a fundamental concept in statistics, where you evaluate two complementary statements: the null hypothesis (HO) and the alternative hypothesis (H1). These hypotheses are used to make decisions about population parameters based on sample data. Here's the mathematical representation of the null and alternative hypotheses:

Null Hypothesis (H0):
The null hypothesis is a statement that there is no significant effect or relationship in the population. It often represents the status quo or a default assumption. Mathematically, the null hypothesis is often expressed as:

$$H0 : \theta = \theta_0$$

Where:
$H0$ represents the null hypothesis.
θ is the population parameter you're testing.
θ_0 is the specific value you're testing the parameter against. This value is often chosen based on the research question or as a baseline.
For example, if you're testing whether the average height of a population is 170 cm, the null hypothesis could be written as $H0 : \mu = 170$, where μ is the population mean.

Alternative Hypothesis (H1 or Ha) :
The alternative hypothesis is a statement that contradicts the null hypothesis and represents what you aim to demonstrate through your analysis. Mathematically, the alternative hypothesis can take several forms depending on the nature of your research question:
Two-Tailed Test ($\text{H1} : \theta \neq \theta_0$): This form is used when you want to test if the parameter is not equal to a specific value.

One-Tailed Test - Greater Than (H1: $\theta > \theta_0$): This form is used when you want to test if the parameter is greater than a specific value.
One-Tailed Test - Less Than ($\mathbf{H1} : \theta < \theta_0$): This form is used when you want to test if the parameter is less than a specific value.

Significance Level, p-Values, and their Interpretation

Significance Level (α) :

The significance level, denoted by α, is a predetermined threshold used in hypothesis testing to make decisions about whether to accept or reject the null hypothesis. It represents the probability of making a Type l error (false positive) when the null hypothesis is true. Common values for α include $0.05(5\%)$ and $0.01\ (1\%)$.

Mathematically, the significance level can be represented as:

$\alpha = P(\text{ Type I error }) = \text{Probability of rejecting } H0 \text{ when } H0 \text{ is true}$

p-Values:

The p-value is a statistical measure used to assess the strength of evidence against the null hypothesis. It quantifies the probability of observing a test statistic as extreme as, or more extreme than, the one obtained from your sample data, assuming the null hypothesis is true. A smaller p-value indicates stronger evidence against the null hypothesis.

Mathematically, the p-value can be defined as:

$p\text{-value} = P$ (observing data as or more extreme than the sample data $|H0$ is tr Interpretation:

p-Value Interpretation:

- If the p-value is less than or equal to the significance level $(p \leq \alpha)$, you reject the null hypothesis. This suggests that there is strong evidence against the null hypothesis in favor of the alternative hypothesis.
- If the p-value is greater than the significance level $(p > \alpha)$, you fail to reject the null hypothesis. This suggests that there is not enough evidence to reject the null hypothesis.

Significance Level (α) Interpretation:

- The choice of α is a trade-off between Type I and Type II errors. A lower α reduces the risk of Type I errors (false positives) but increases the risk of Type II errors (false negatives).
- A common choice for α is 0.05 , which corresponds to a 5% risk of making a Type I error. However, researchers can adjust α based on the specific requirements of the analysis and the field of study.

Types of errors

When performing hypothesis tests or statistical analyses, it's important to understand different types of errors and how to choose an appropriate test based on your research question and data. Two common types of errors in hypothesis testing are Type I and Type II errors.

Type I Error (False Positive):

A Type I error occurs when you reject the null hypothesis when it is actually true. In other words, you conclude that there is an effect or difference when there is none. The probability of making a Type I error is denoted as α, which is the significance level you choose.

Mathematically, a Type I error can be defined as:

$$P(\text{ Type I error }) = \alpha$$

Choosing a lower significance level (α) reduces the risk of Type I errors but may increase the risk of Type II errors.

Type II Error (False Negative):

A Type II error occurs when you fail to reject the null hypothesis when it is actually false. In this case, you miss detecting a real effect or difference that exists. The probability of making a Type II error is denoted as β. Mathematically, a Type II error can be defined as:

$$P(\text{ Type II error }) = \beta$$

The power of a statistical test is $1 - \beta$, which is the probability of correctly rejecting the null hypothesis when it is indeed false. Increasing the sample size or choosing a more powerful test can reduce the risk of Type II errors.

Factorial Design and Classic Experimental Plans

Part 1: Inference, Experimental Design and Application

1.2 Principles of Experimental Design

2×2 Factorial Design:

- In a 2×2 factorial design, there are two independent variables, each with two levels (conditions). This design is useful for studying the main effects of each variable and any interaction between them. It results in four treatment combinations.

3×2 Factorial Design:

- This design involves two independent variables. One variable has three levels, and the other has two levels. It allows for the examination of main effects and interactions but is more complex than the 2×2 design.

Latin Square Design:

- The Latin square design is used when there are three or more factors, and each factor has multiple levels. It ensures that each level of every factor occurs once in each row and column of the design. This design is suitable for controlling confounding effects in complex experiments.

Full Factorial Design:

- In a full factorial design, all possible combinations of factor levels are tested. This design is comprehensive but can become very complex as the number of factors and levels increases.

Fractional Factorial Design:

- When the full factorial design is impractical due to resource limitations, a fractional factorial design tests a subset of the possible combinations. This design provides a more efficient approach.

Factorial Design and Classic Experimental Plans

Part 1: Inference, Experimental Design and Application

1.2 Principles of Experimental Design

Randomized Controlled Trial (RCT):

- RCTs are commonly used in clinical and medical research. They involve randomizing participants into treatment and control groups to assess the impact of a treatment or intervention.

Crossover Design:

- Crossover designs are used when each participant receives multiple treatments in a specific order. This design helps control for individual variability.

Before-and-After Design:

- This design assesses the impact of an intervention by measuring a dependent variable before and after the treatment. It helps evaluate the effectiveness of the intervention over time.

Matched Pairs Design:

- In a matched pairs design, participants are matched based on certain characteristics, and then each participant in a pair receives a different treatment. This design controls for individual differences.

Repeated Measures Design:

- In repeated measures designs, the same group of participants is used for all treatment conditions. This design is suitable for studying changes within the same individuals over time.

Control Groups, Randomization, Replication, and Blocking

Control Groups:

A control group is a fundamental component of experimental design. It serves as a reference point for comparison with the treatment group(s) in an experiment. The control group is exposed to the same conditions as the treatment group(s) except for the specific factor (independent variable) being tested.

Key characteristics of control groups:
- They help researchers assess the impact of the treatment by providing a baseline for comparison.
- The control group should ideally be as similar as possible to the treatment group(s) in all aspects except for the treatment itself.
- In medical and clinical research, a placebo (inactive substance) is often used in control groups to eliminate the influence of expectations.

Randomization:

Randomization is the process of assigning subjects, treatments, or conditions to groups or experimental units in a random and unbiased manner. Randomization is a critical aspect of experimental design as it helps control for confounding variables and ensures that the groups are comparable.

Key principles of randomization:
- It reduces the risk of selection bias by ensuring that each subject has an equal chance of being in any group.
- Randomization can be applied in various ways, including simple random sampling, stratified random sampling, and block randomization.
- Randomization minimizes the influence of extraneous factors that could affect the results.

Control Groups, Randomization, Replication, and Blocking

Part 1: Inference, Experimental Design and Application

1.2 Principles of Experimental Design

Replication:

Replication involves conducting the same experiment or study multiple times to validate the results and increase the reliability of findings. Replication helps ensure that observed effects are consistent and not due to random variation.

Key points regarding replication:
- Replication is particularly important in scientific research to confirm the validity of experimental results.
- It allows for the generalization of findings to a broader population.
- The number of replications required depends on the specific research question and the degree of certainty needed in the results.

Blocking:

Blocking is a technique used to control for the influence of a known extraneous variable that might affect the results. In blocking, subjects are divided into groups or blocks based on a specific characteristic (e.g., age, gender) that is known to influence the outcome.

Key features of blocking:
- It helps ensure that the groups being compared are more homogeneous with respect to the blocking variable.
- Blocking is often used in experiments to reduce the variability within each group and increase the precision of the study.
- The choice of which variable to block on depends on the researcher's understanding of the problem and the goal of the study.

KEYWORDS (NEW)

- Sondage
- Étude de cohorte
- Hypothèse nulle vs hypothèse alternative
- Niveau de signification p-value
- Biais expérimental
- F de fisher
- Hypothèse alternative (H1)
- Invariance du maximum de vraisemblance
- Principe d’un test de probabilité
- Collecte des données
- Test bilatéral
- Fiabilité des données
- Hypothèse nulle (H0)
- Risque de première espèce (alpha)
- Plan d’expérience
- Risque de deuxième espèce (beta)
- Étude cas-témoins
- Test d'hypothèse
- Puissance du test
- Test d’homogénéité sur la variance
- Statistique de test
- Théorème de Kolmogorov-Smirnov
- Test à deux queues
- Test à une queue

Use Case: Hypothesis Testing in Clinical Trials

In the context of the course on Inference and Estimation Theory, covering topics on statistical inference, point estimation, interval estimation, hypothesis testing, experimental design, and advanced topics, let's explore a use case related to hypothesis testing in clinical trials. This use case involves applying hypothesis testing techniques to determine the efficacy of a new medical treatment.

Description:

In this use case, we will focus on using hypothesis testing to assess the effectiveness of a new medical treatment compared to an existing standard treatment. This scenario is common in clinical trials when researchers need to make decisions about the adoption of a new treatment protocol based on statistical evidence.

Key Components:

Introduction to Inference and Estimation Theory: Understanding the basics of statistical inference, point and interval estimation, and the role of probability in making informed decisions.

Hypothesis Testing: Applying hypothesis testing principles to compare the effectiveness of medical treatments, including defining null and alternative hypotheses, setting significance levels, and interpreting p-values.

Experimental Design: Planning a clinical trial with control groups, randomization, and replication. Deciding on the appropriate sample size and considering ethical considerations in experimental design.

Advanced Topics in Inference and Estimation: Exploring Bayesian inference as an alternative approach to traditional hypothesis testing, nonparametric estimation for flexibility in data analysis, and the importance of statistical power and effect size in clinical trials.

Python Code Example (Hypothesis Testing in Clinical Trials):

```python
import numpy as np
import scipy.stats as stats

# Generate synthetic data for two treatment groups (e.g., patient recovery times)
np.random.seed(42)
treatment_group = np.random.normal(10, 2, size=30)  # New treatment
standard_group = np.random.normal(12, 2, size=30)  # Standard treatment

# Perform a two-sample t-test for hypothesis testing
t_stat, p_value = stats.ttest_ind(treatment_group, standard_group)

# Define significance level (alpha)
alpha = 0.05

# Determine whether to reject the null hypothesis
if p_value < alpha:
    result = "Reject Null Hypothesis"
else:
    result = "Fail to Reject Null Hypothesis"

# Print results
print("Hypothesis Testing in Clinical Trial")
print("Null Hypothesis: The new treatment is as effective as the standard treatment.")
print("Alternative Hypothesis: The new treatment is more effective than the standard treatment.")
print(f"t-statistic: {t_stat:.2f}")
print(f"P-value: {p_value:.4f}")
print(f"Result: {result}")
```

In this code, we generate synthetic recovery time data for two treatment groups and perform a two-sample t-test to compare the effectiveness of the new treatment to the standard treatment. The code calculates the t-statistic, p-value, and determines whether to reject the null hypothesis based on a predefined significance level (alpha).

This use case demonstrates how hypothesis testing techniques from the course on Inference and Estimation Theory can be applied to make critical decisions in clinical trials, ultimately impacting healthcare outcomes and treatment protocols.

REFERENCES V2

- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for Experimenters. Wiley.
- Fisher, R. A. (1935). The Design of Experiments. Oliver & Boyd.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). Mathematical Statistics with Applications. Cengage Learning.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). Contrasts and Effect Sizes in Behavioral Research: A correlational approach. Cambridge University Press.
- Montgomery, D. C. (2017). Design and Analysis of Experiments. Wiley.
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Routledge.
- Casella, G., & Berger, R. L. (2002). Statistical Inference. Duxbury Advanced Series.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis. CRC Press.
- Lehmann, E. L. (1997). Testing Statistical Hypotheses. Springer.
- Hogg, R. V., McKean, J. W., & Craig, A. T. (2018). Introduction to Mathematical Statistics. Pearson.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association.
- Efron, B., & Tibshirani, R. J. (1994). An Introduction to the Bootstrap. CRC Press.
- Kirk, R. E. (2018). Experimental Design: Procedures for the Behavioral Sciences. Sage Publications.
- Agresti, A., & Franklin, C. (2012). Statistics: The Art and Science of Learning from Data. Pearson.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. CRC Press.
- Zar, J. H. (2010). Biostatistical Analysis. Pearson.
- Keppel, G., & Wickens, T. D. (2004). Design and Analysis: A Researcher's Handbook. Pearson.
- Sheskin, D. J. (2011). Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press.
- Siegel, S., & Castellan, N. J. (1988). Nonparametric Statistics for The Behavioral Sciences. McGraw-Hill.
- Kish, L. (1965). Survey Sampling. Wiley.

TEXTE DE DESCRIPTION DU COURS

"Introduction to Inference and Estimation Theory" is your gateway to the fascinating world of statistical inference, where you'll uncover the core principles and techniques that empower you to draw meaningful conclusions from data. Whether you're a statistician, data analyst, or researcher, this comprehensive course equips you with the knowledge and tools to make informed decisions based on evidence.

The journey begins with "Basics of Inference," unraveling the essence of statistical inference, the role of probability, and the different types of statistical inference. You'll grasp the fundamentals of "Point Estimation," exploring estimators, their properties (bias, consistency, efficiency), and the powerful concept of Maximum Likelihood Estimation (MLE). Next, "Interval Estimation" comes into focus, where you'll dive into confidence intervals, learn how to construct them, and gain the expertise to interpret these intervals effectively.

"Hypothesis Testing" takes center stage in Part 2, beginning with the "Foundations of Hypothesis Testing." You'll delve into null and alternative hypotheses, significance levels (alpha), and the interpretation of p-values. Parametric hypothesis tests, such as Z-tests, T-tests, and ANOVA, are demystified in "Parametric Hypothesis Tests," including their applications in various scenarios. The course then explores "Nonparametric Hypothesis Tests," unveiling tests like the Wilcoxon Signed-Rank Test, Mann-Whitney U Test, and Kruskal-Wallis Test, and how they extend your analytical toolkit.

"Experimental Design" takes the spotlight in Part 3, starting with the "Principles of Experimental Design." Learn about control groups, randomization, replication, and the power of factorial design. Dive into the art of "Planning Experiments," from choosing experimental variables to determining sample sizes and considering ethical aspects of design. Discover "Designing Surveys and Observational Studies," where you'll uncover survey design principles, sampling methods, and the nuances of bias and error in observational studies.

In "Advanced Topics in Inference and Estimation" (Part 4), you'll embark on a deeper exploration. "Bayesian Inference" introduces you to Bayes' Theorem, posterior probability, and the integration of prior information into estimation and hypothesis testing. "Nonparametric Estimation" showcases kernel density estimation, nonparametric regression, and the power of bootstrap resampling. Finally, "Statistical Power and Effect Size" reveals the critical importance of understanding statistical power, along with effect size measures like Cohen's d and Eta-squared, and how they drive power analysis in experimental design.

By the end of this course, you'll be equipped with a robust understanding of inference and estimation theory, ready to navigate the complexities of data analysis, experimentation, and hypothesis testing with confidence.