STATISTICS

[module, 2022–2023]

Professor: Alexander (Sasha) Shapoval
University of Lodz
ashapoval@nes.ru

Course information

Course Website:

Instructor's Office Hours:

Class Time:

Room Number:

TAs: [Names and contact information]

Course description

This is a "practical" course that covers a few selected concepts and tools that are most needed for working with models in Economics and Finance. Although students may be familiar with the introductory concepts from the course, we will soon jump into topics that do not get enough attention in the curriculum of any undergraduate math programs, including the best ones.
In this course, we will rarely discuss proofs. Methods and results will be described in a loose manner — with the goal of starting to think like economists and getting intuitive and practical understanding of how to solve problems. This will lead us to important insights about the world around us.

Course requirements, grading, and attendance policies

Student's achievements will be evaluated based on problem sets and the final exam. The exam and the problem sets are graded on the scale of 0 to 100 points. The final score is also computed on the 100-point scale and then converted to transcript grades (from 2 to 5+).
The weights related to the problem sets and the exam are 0.4 and 0.6. This yields the following equation for the final score:
Final_Score = 0.4 * (PS1 + PS2 + PS3 + PS4) / 4 + 0.6 * Exam

Course contents

In the course, we will discuss the following topics.

- Formulation of a problem in statistics. Example: Does first baby come earlier? (ThinkStats)
- Bayesian approach. Example: spam filter (PSE, PYTHON)
- Basic knowledge of probabilities. Random variables, probability distribution function, cumulative distribution functions. Transformation of random variables. Exponential distribution. The sum of independent random variables. Convolution. The sum of exponential distributions and the gamma distribution. Waiting time paradox. Persistence of bad luck. Ratios. Order statistics and Cauchy distribution. (PROB)
- Uniform distribution and beta distribution as order statistics. Empirical distributions. Kolmogorov-Smirnov statistics. Independence on the true distribution (PROB). Monte-Carlo method. Pseudorandom numbers. (SDA)
- Statistical tests. Hypotheses, test statistics, significance level, power. Linear test statistics, the Fisher discriminant function. Nonlinear test statistics, neural networks. Goodness-of-fit tests. Pearson chi-square test. (SDA)
- Inverse cumulative distribution function. Bootstrap (SDA).
- General concepts of parameter estimation. Point estimators. Estimators for mean, variance, covariance. Bias. Method of moments. The method of maximum likelihood. Maximum likelihood with binned data. Relationship between ML and Bayesian estimators (PSE, SDA).
- The method of least squares. Connection with maximum likelihood Testing goodness-of-fit with chi-squared. Statistical errors and confidence intervals. (PSE, SDA)
- Regressions. Option to interpolate and extrapolate data. Option to verify dependencies. Unbiasedness of the regression coefficients. The variance and the confidence interval of the slope of the response variable. Hypothesis: the slope is zero. Discussion on an appropriate model choice. (ECON)
- Information and entropy. Application to rational inattention. Example: optimal costly observation of AR-1 processes. General equilibrium with uncertainty (INFO).

Description of course methodology

All course materials will be presented during lectures and seminars. I strongly recommend taking notes in class. The relevant chapters in handbooks are mentioned next to the topics. However, I am going to present my understanding of these topics rather than the explicit content of the handbooks.

Sample tasks for course evaluation

Two random samples X_1, …, X_{20} and Y_1, …, Y_{100} are obtained independently from two normal distributions $N(\mu,\sigma^2)$ and $N(\mu,4\sigma^2)$ that share common parameters.
(a) Suggest estimators for these two parameters
(b) Is your estimator for μ consistent? Provide a brief argument.
(c) Propose a confidence interval for μ with the confidence level $1 - \alpha$

Your boss asked you to test some hypothesis at the 4% level. You designed a test, collected a random sample and obtained the p-value of 0.0439. Then, you and your colleague learn that the boss was absent-minded and unintentionally gave both of you the same assignment. As you learn, your colleague also collected a random sample independently from you and obtained the p-value of 0.0478. Now it is almost the deadline to report. You do not have

time to inspect each other's test design, data, and calculation, but must report your conclusion of whether to accept or reject.
You are considering the following potential plans.
1. The 4% level does not make sense. At the standard 5%, the hypothesis is rejected by your sample and hence you should answer "reject."
2. Although each of the samples does not warrant immediate rejection, the two p-values are near the borderline of 4%, so it is more appropriate to coordinate between you and your colleague and reject the hypothesis.
3. Although the p-values are above the 4% threshold, it seems that the data does not really support the hypothesis. To convey your doubts, the best action is to coordinate between you and your colleague and for one of you to say "reject" and for the other one to say "accept." With this, the boss will make the decision that suits him the best (or flip a coin to resolve the conflicting evidence).
4. You and your colleague should be honest, and both say "accept."

Comment on the validity of each of the four reasonings and pick one that you think is the best.

Course materials

Required textbooks and materials

[This section must include literature sources that have been published in the last 5 years, as required by Russian State educational standards. Please make sure that the materials you list on the syllabus are ordered by the library]
Bruce E. Hansen, Probability and Statistics for Economists (PSE)
Glen Cowan, Statistical Data Analysis (SDA)
Jeffrey M. Wooldridge, Introductory Econometrics: A Modern Approach (looking forward to the course of econometrics, ECON)
William Feller, An Introduction to Probability Theory and Its Applications (PROB)
John S. Conery. Explorations in Computing: An Introduction to Computer Science and Python Programming (elementary introduction to Python, PYTHON)

Additional materials

Allen B. Downey Think Stats: Probability and Statistics for Programmers (the handbook can be considered as elementary practical introduction to various topics of our discipline, Think Stats)
Donald B. Rubin, Basic Concepts of Statistical Inference for Causal Effects in Experiments and Observational Studies
Thomas M. Cover, Elements of Information Theory (INFO)
Hogg, R. V., J. W. McKean, and A. T. Craig, Introduction to Mathematical Statistics

Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.