**Government of Russian Federation**

**Federal State Autonomous Educational Institution of High Professional Education**

**«National Research University Higher School of Economics»**

National Research University

High School of Economics

Faculty of Psychology

**Syllabus for the course**

**«Probability Theory and**

**Mathematical Statistics»**

(Теория вероятностей и математическая статистика)

030300.68 «Cognitive sciences and technologies: from neuron to cognition», Master of Science

Authors:

Alexey A. Neznanov, associate professor, aneznanov@hse.ru

Ilya A. Makarov, senior lecturer, iamakarov@hse.ru

Approved by: Head of Data Analysis and Artificial Intelligence Department, Sergey O. Kuznetsov

Recommended by:

Moscow, 2014

# Teachers

**Authors, lecturers, tutors**:

Alexey A. Neznanov, National Research University Higher School of Economics, Department of Data Analysis and Artificial Intelligence, associate professor

Ilya A. Makarov, National Research University Higher School of Economics, Department of Data Analysis and Artificial Intelligence, Deputy Head, senior lecturer

**Teaching assistant**: Dmitry Zotikov (to be confirmed)

# Scope of Use

The program establishes minimum demands of students’ knowledge and skills, and determines content of the course.

The present syllabus is aimed for the department teaching the course, their teaching assistants, and students of the Master of Science program 030300.68 «Cognitive sciences and technologies: from neuron to cognition».

This syllabus meets the standards required by:

* Educational standards of National Research University Higher School of Economics;
* Educational program «Psychology» of Federal Master’s Degree Program 030300.68, 2011;
* University curriculum of the Master’s program in psychology (030300.68) for 2014.

# Summary

The course “Probability Theory and Mathematical Statistics” (in English) covers basic methods of computational probability theory, and learn to apply principles of mathematical statistics at experiment design and validation of obtained results. This course, together with the two other mathematical courses, provides sufficient background for all quantitative and computational modeling disciplines, which are studied at the Master’s program 030300.68 «Cognitive sciences and technologies: from neuron to cognition». Students study discrete and continuous probability measures, conditional and independent probabilities, distributions and statistical hypotheses; learn how to design, compute, analyze and evaluate data from experiment; apply their knowledge to psychological problems and their formalization in terms of computational methods for real experiments.

# Learning Objectives

Learning objectives of the course «Probability Theory and Mathematical Statistics» are to familiarize students with experiment design, its foundation and connections to mathematical modelling and extraction of knowledge from experiment:

* Basic elements of probability theory;
* Common univariate and multivariate probability distributions and their meaning;
* Confidence intervals and hypotheses testing;
* Contingency tables;
* Analysis of Variance (ANOVA);
* Asymptotic normal approximations;
* Power analysis;
* Experimental design.

# Learning outcomes

 After completing the study of the discipline «Probability Theory and Mathematical Statistics», the student should:

* Know the basic concepts of probability, discrete and continuous random variables, probability distributions, expected values, joint probability distributions, independence, and its connections with other sciences.
* Be able to translate a real-world problem into mathematical terms.
* Be able to formulate hypotheses that can be tested using experiments.
* Be able to understand and interpret specific experimental design techniques.
* Be able to construct samples, use randomization,
* Possess main definitions of probability theory, mathematical statistics and experimental design; build experiment schemes; select appropriate tests and statistical analysis methods.
* Possess basic statistical packages in MatLab
* Possess techniques of proving theorems and thinking out counter-examples.
* Learn to develop complex mathematical reasoning.

After completing the study of the discipline « Probability Theory and Mathematical Statistics» the student should have the following competences:

| **Competence**  | **Code** | **Code (UC)** | **Descriptors (indicators of achievement of the result)** | **Educative forms and methods aimed at generation and development of the competence** |
| --- | --- | --- | --- | --- |
| The ability to reflect developed methods of activity. | SC-1 | SC-М1 | The student is able to reflect developed mathematical methods to psychological fields and problems. | Lectures and tutorials, group discussions, presentations, paper reviews. |
| The ability to propose a model to invent and test methods and tools of professional activity | SC-2 | SC-М2 | The student is able to improve and develop research methods of sample generation, experimental design, randomization, post-hoc analysis, etc. | Classes, home works.Computational part of hypotheses testing in MatLab. |
| Capability of development of new research methods, change of scientific and industrial profile of self-activities | SC-3 | SC-М3 | The student obtain necessary knowledge in probability theory and mathematical statistics, which is sufficient to develop new methods on other sciences | Home tasks, paper reviews |
| The ability to describe problems and situations of professional activity in terms of humanitarian, economic and social sciences to solve problems which occur across sciences, in allied professional fields. | PC-5 | IC-M5.3\_5.4\_5.6\_2.4.1 | The student is able to describe psychological problems in terms of computational mathematics. | Lectures and tutorials, group discussions, presentations, paper reviews. |
| The ability to detect, transmit common goals in the professional and social activities | PC-8 | SPC-M3 | The student is able to identify mathematical aspect in psychological researches; evaluate correctness of the used methods, and their applicability in each current situation | Discussion of paper reviews; cross discipline lectures |

# Place of the discipline in the Master’s program structure

The course «Probability Theory and Mathematical Statistics» is facultative coursetaught in the first year of the Master’s program «Cognitive sciences and technologies». It is recommended for all students of the Master’s program who do not have fundamental knowledge in advanced mathematics at their previous bachelor/specialist program.

##  Prerequisites

No special knowledge is required, but all students are advised to prepare for studying mathematical discipline, even if they had previous education only with humanitarian profile.

However it is strongly recommended, that all students have listened to the adaptation courses “Calculus” and “Linear Algebra” from the first two modules. Classes are supported by using MatLab packages, so some basic knowledge of computational methods in MatLab are welcomed.

The following knowledge and competence are needed to study the discipline:

* A good command of the English language.
* A basic knowledge in mathematics.
* A basic programming or MatLab experience.

Main competences developed after completing the study this discipline can be used to learn the following disciplines:

* Computational modelling.
* Qualitative and quantitative methods in psychology.
* Digital signal processing.
* Mathematical statistics and experimental design (MAGoLEGO).

 **Comparison with the other courses at HSE**

 One can only find the course “Probability Theory and Introductory Statistics”

by Mikhail Zhitlukhin <http://www.hse.ru/edu/courses/130683673.html>

at [www.hse.ru/edu/courses](http://www.hse.ru/edu/courses) educational portal.

 The main differences between ICEF course and this discipline are the following aspects:

* ICEF course is bachelor basis course for sophomore students, providing wide range of knowledge for economists with strong mathematical background. We adapted this master course for students with possible humanitarian background. All in all, it is simpler, but more practically oriented.
* Our course is made in close contact with colleagues from the Department of Psychology, who gave us ideas, examples and methods to connect “Probability Theory and Mathematical Statistics” with their experience in computational psychology.
* ICEF course does not have interactive model for combining practical and theoretical model of classes. Our course goal is to provide sufficient mathematical background for Research seminar and Course project of the 1st year master students so they successfully design an experiment, make a hypothesis and know how to test it and the whole scheme of experiment as well.

 There are many “Probability Theory and Mathematical Statistics” courses in the world, as it lies in foundation of many applied methods for neuroscience. This course is devoted to practical aspects of experimental design and common errors avoidance. We took one of the famous books by R. Montgomery and use O. Anderson in addition to core literature. However, we develop our own methodical book, which takes into consideration specifics of our course oriented on psychological master program and integration of MatLab computational section with experimental design.

# Schedule

One pair consists of 1 academic hour for lecture and 1 academic hour for classes after lecture.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Topic** | **Total hours** | **Contact hours** | **Self-study** |
| **Lectures** | **Seminars** |
| 1 | Foundations of probability theory. MatLab introduction. | 20 | 4 | 4 | 12 |
| 2 | Measurements and scales. Main statistical concepts. Data preparation and preprocessing in MatLab. | 20 | 4 | 4 | 12 |
| 3 | ANOVA basics. | 12 | 2 | 2 | 8 |
| 4 | Experimental design concepts. Visualization of experiment results. | 12 | 2 | 2 | 8 |
| 5 | Power analysis and confidence intervals. | 12 | 2 | 2 | 8 |
| 6 | Multiple regressions and interaction. | 12 | 2 | 2 | 8 |
| 7 | Solving the problem of multiple comparisons and post-hoc analysis. | 12 | 2 | 2 | 8 |
| 8 | Common errors in experimental design. | 8 | 2 | 2 | 4 |
|  | **Total** | **108** | **20** | **20** | **68** |

# Requirements and Grading

|  |  |  |
| --- | --- | --- |
| **Type of grading** | **Type of work** | **Characteristics** |
| 1 | 2 |  |
| Test | 1 |  | Enter test |
| Homework | 1 | 1 | Solving homework tasks and examples. |
| Special homework - paper review |  | 1 | Description of mathematical methods applied in psychological research paper |
| Exam |  | 1 | Oral exam. Preparation time – 120 min. |
| Final |  |  |  |  |

# 9. Assessment

*The assessment* consists of classwork and homework, assigned after each lecture. Students have to demonstrate their knowledge in each lecture topic concerning both theoretical facts, and practical tasks’ solving including MatLab exercises. All tasks are connected through the discipline and have increasing complexity. All weekly tasks should be verified during classes or by teaching assistant.

*Final assessment* is the final exam. Students have to demonstrate knowledge of theoretical facts, but the most of tasks would evaluate their ability to solve practical examples, present straight operation, and recognition skills to solve them. **Final exam** would consist of two parts, having equal weights 50% of exam mark.

# The grade formula:

***Final course mark*** *is obtained from the following formula*:

*Оfinal = 0,6\*О cumulative+0,4\*Оexam* .

The grades are rounded in favour of examiner/lecturer with respect to regularity of class and home works. Grades, having a fractional part greater than 0.5, are rounded up only with positive exam mark and homework regularity greater then 75%.

**Table of Grade Accordance**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ten-point****Grading Scale** |

|  |  |
| --- | --- |
|  |  |

**Five-point****Grading Scale** |  |
| 1 - very bad2 – bad3 – no pass | Unsatisfactory – 2 | **FAIL** |
| 4 – pass5 – highly pass | Satisfactory – 3 | **PASS** |
| 6 – good7 – very good | Good – 4 |
| 8 – almost excellent9 – excellent10 – perfect | Excellent – 5 |

# Course Description

The following list describes main mathematical definitions, properties and objects, which we would consider in the course in correspondence with lecture order.

**Topic 1. Foundations of probability theory. MatLab introduction.**

**Content**:

Discrete and continuous random variables. Common univariate and multivariate probability

distributions and their meaning. Normal distribution and central limit theorem. Expected values. Joint probability distributions, independence. Law of large numbers.

Main tools and data formats of MatLab. MatLab package “Statistics toolbox”.

**Further guidance**:

 Additional MatLab packages. Different formalizations of probability theory.

**Topic 2. Measurements and scales. Main statistical concepts. Data preparation and**

**preprocessing in MatLab.**

**Content**:

Measurements, scales and errors: physical sense and formalization.

Population, sample, random sample.

Methods of data collection: census, sample survey, experiment, observational study.

Relations between variables: magnitude, reliability. Strange of relation and size of sample. Levels of statistical significance. Importance of normal distribution.

Frequency tables: interpretations.

Sources of bias in sampling.

Descriptive statistics. Mean and confidence interval. Shape of the Distribution. Normality.

Correlations. Linear correlation. Significance of correlations. Interpretation of the values of correlations. Outliers. Correlation matrices.

Simple groups comparison and t-test.

Types of sampling: simple random sampling, stratified random sampling, cluster sampling.

Preparing sample data.

**Further guidance**:

 Non-homogeneous groups in samples.

Measuring nonlinear relations.

Identifying biases and mean substitution.

Crosstabulation.

**Topic 3. ANOVA basics.**

**Content**:

Analysis of variance. ANOVA and t-test. Controlling for factors.

1-way ANOVA. 2-way ANOVA

**Further guidance**:

 N-way analysis of variance.

Analysis of covariance.

**Topic 4. Experimental design concepts. Visualization of experiment results.**

**Content**:

Planning and conducting an experiment.

Control groups, random assignments, replication.

Sources of bias in experiments. Mixing factors, placebo effect, blinding.

Completely randomized design. Block design.

Visualizing of experiment schemes.

**Further guidance**:

 Complexdesignes.

**Topic 5. Power analysis and confidence intervals.**

**Content**:

Again about sampling theory: deep into hypothesis testing logic and increasing statistical

power. Types of significance testing. Calculating required sample size.

Inadequacies of the hypothesis testing approach and advantages of interval estimation.

Selecting proper confidence limits. Power analysis in experimental design.

**Further guidance**:

 Confidence intervals in ANOVA. Standardized experimental effects.

**Topic 6. Multiple regressions, interaction.**

**Content**:

Regression models. Main regression equations. Coefficient of determination (R-square).

Interpreting the correlation coefficient. Limitations. Assumption of linearity and normality.

Choice of the number of variables.

Residual analysis in case of extreme outliers.

**Further guidance**:

 Factoranalysis.

**Topic 7. Solving the problem of multiple comparisons and post-hoc analysis.**

**Content**:

P value and alpha in multiple comparisons. Post-hoc analysis.

The problem with multiple comparisons: assumptions and limitations. Examples of

inadequate results.

Controlling the familywise error rate: Bonferroni correction.

Controlling the false discovery rate: Benjamini–Hochberg procedure.

Post-hoc analysis, planned comparisons and experimental design.

**Further guidance**:

 Methods of Tukey, Dunnett, Dunn, Holm and others.

**Topic 8. Common errors in experimental design.**

**Content**:

Spurious correlations and samples sizes. Correlations and dependencies.

Hunting for the statistical significance. Changing the definition of the outcome. Using a

different time scale. Changing criteria for including or excluding a subject. Removing

arbitrarily points as outliers. Changing ways to clump or separate subgroups, ways to

normalize the data, statistical tests, and algorithms for computing statistical tests.

Special multiple comparisons traps.

**Further guidance**:

 Robust statistical methods. Methodologies for robust experiments.

1. **Term Educational Technology**

The following educational technologies are used in the study process:

* Discussion and analysis of the results of the home task during office hours;
* Individual education methods, which depend on the progress of each student;
* Analysis of skills to formulate common problem in terms of mathematics and solve it;
* Practicing in software tools.
1. **Recommendations for course lecturer**

Course lecturer is advised to use interactive learning methods, which allow participation of the majority of students, such as slide presentations, combined with writing materials on board, and usage of interdisciplinary papers to present connections between mathematics and psychology.

The course is intended to be average, but it is normal to differentiate tasks in a group if necessary, and direct fast learners to solve more complicated tasks.

There are necessary requirements to prepare MatLab guideline and examples.

Lecture may advise to take additional MAGoLEGO course “Mathematical statistics and experiment design” to pay more attention to experiment design problems.

1. **Recommendations for students**

#  The course is interactive. Lectures are supposed to have standalone format, although classes will take both using board and laptops before lecture starts. Such chosen order of classes and lectures will help students to understand theoretical material since they already know about computational part.

# There will be special office hours for students, who would like to get more precise understanding of each topic. Teaching assistant will also help you. All tutors are ready to answer your questions online by official e-mails that you can find in the “contacts” section.

There are comprehensive list of additional Internet-sources on statistical analysis and MatLab best practices.

# Final exam questions

1. Probability. Probability space.
2. Common univariate and multivariate probability distributions.
3. Normal distribution.
4. Descriptive statistics.
5. Type I and type II errors.
6. Maximum likelihood estimators.
7. Confidence interval.
8. Classification of scales.
9. Measurements errors.
10. Sampling methods.
11. Randomization methods.
12. Methods of data collection.
13. Statistical significance.
14. Frequency tables.
15. Linear correlation.
16. Outliers methods.
17. T-test.
18. Analysis of variance (ANOVA).
19. Linear regression.
20. Statistical power analysis.
21. Experiment design and basic experiments schemes. Runs and blocks.
22. Hierarchical experiment design.
23. Basic methods of Ad-Hoc analysis.

#  Reading and Materials

The facultative course is intended to present a brief overview of basic statistics methods and provide necessary background for experimental design to accompany course project. We present list of recommended literature as well as MatLab tutorials. The course has a lot of real-life problems and examples and contains sufficient knowledge for both, verification of existing experiment model and its comparison with the other possible schemes of experimental design. The reference list of books might be useful for students as they provide additional exercises and theoretical information.

## Required Reading

Montgomery D.C., Runger G.C. Applied Statistics and Probability for Engineers. - 6th ed. - Wiley, 2014. - 836 p.

Montgomery D.C. Design and Analysis of Experiments. - 8th ed. - Wiley, 2012. - 752 p.

## Recommended Reading

 Anderson O. Experiment! Planning, implementing and interpreting, - Wiley, 2012. - 288 p.

 Giesbrecht F.G., Gumpertz M.L. Planning, Construction, and Statistical Analysis of Comparative Experiments. - Wiley, 2004. - 705 p.

 Jeff Wu C.F., Hamada M. Experiments: Planning, Analysis, and Optimization. - 2nd Edition. - Wiley, 2009.

## List of papers for review

Currently in work with lecturers from Psychology department.

 The first pack contains permutation methods and hypotheses testing.

## Course telemaintenance

All material of the discipline are posted in informational educational site at NRU HSE portal [www.hse.ru](http://www.hse.ru). Students are provided with links on psychological papers, tests, electronic books, articles, etc.

1. Equipment

The course requires a laptop, projector, and acoustic systems.

The syllabus is prepared by Alexey A. Neznanov and Ilya A. Makarov.