

**State University –  
Higher school of economics**

*International college of economics and finance*

**Syllabus**

**Probability Theory and Introductory Statistics**

for Bachelors degree in Economics 080100.62

Author: Katyshev P.K.

Moscow, 2012

## **1. Course description**

The course PROBABILITY THEORY AND INTRODUCTORY STATISTICS is two semesters course for the first year students. This is introductory course for the students who are specialized in economics. The course is taught in Russian and English.

The basic concepts and notions of statistics are taught in this course: descriptive statistics, population and sample, parameter estimation, hypothesis testing, etc. Moreover, some topics of probability theory that are necessary for the main course are taught as well.

## **2. Course objectives**

The purpose of the course is to give the students basic knowledge of the introductory Probability Theory and Mathematical and Applied Statistics. At the end of the course they should know the subjects and basic methods of statistical analysis. During the process of study the students are acquiring the skills of primary data processing viz. to find median, mean and other descriptive statistics of the distribution; they should learn to graphic presentation of data (histogram, stem-and-leaf-plots, dot-plots, box-plots).

Data must be collected according to a well-developed plan, and students should get the initial knowledge what is well designed and well conducted experiment. They must detect the sources of bias.

Students should understand the difference between population and sample and respectively between population and sample characteristics. It seems inexpediently to teach statistics without teaching the elements of Probability Theory therefore some basic concepts and notions of the Probability Theory are taught in the course, namely, probability space, random events, probability of the random event. Students should calculate the probabilities of the complex events, solve combinatorial problems, use Bayes formula. They should know the notion of random variable and its distribution, find the characteristics of random variable (mean, variance, standard deviation, etc.).

At the end of the course students have to formulate and to solve the traditional problems of introductory statistics, viz. parameter estimation, hypothesis testing, correlation analysis, analysis of variances. One of the purposes of the course is to prepare the students for the learning of simple and multiple regressions.

The course exposition is not mathematically rigorous therefore the proves of many theorems and even the strict formulations of some results are skipped.

During the course students should solve a lot of problems including computer exercises with the real data.

At the end of the course the students have to know the main theoretical results of introductory statistics, should be able to make the required calculations and to use the standard statistical methods for applied problems.

One of the purposes of the course INTRODUCTORY STATISTICS is to prepare the first year students for Advanced Placement Test (APT).

## **3. Distribution of hours by topics and types of work**

N	Names of sections and topics	TOTAL (hours)	Auditorium studies (hours)		Self-study
			Including		
			Lectures	Classes	
1	Primary data analysis	24	6	6	12
2	Data collection; planning and conducting experiments	22	6	6	10
3	Elements of Probability Theory	24	6	6	12
4	Discrete random variables	16	4	4	8
5	Continuous random variables	16	4	4	8
6	Joint distribution of two random variables	16	4	4	8
7	Limit Theorems	16	4	4	8
8	Sampling distribution	16	4	4	8
9	Point parameter estimates	24	6	6	12
10	Interval estimates	32	8	8	16
11	Hypothesis testing	40	10	10	20
12	Simple regression and sampling correlation	24	6	6	12
	<b>TOTAL</b>	<b>270</b>	<b>68</b>	<b>68</b>	<b>134</b>

#### 4. The methods

The following methods and forms of study are used in the course:

- lectures (2 hours a week)
- classes (2 hours a week)
- weekly written home assignments
- teachers' consultations
- self-study.

In total the course includes 42 hours of lectures and 42 hours of classes for the first year students.

#### 5. Forms of control

- home assignments
- control tasks
- fall mid-term exam (180 min.)
- exam at the end of the second module (180 min.)
- spring mid-term exam (180 min.)
- AP test

#### 6. Grades determination

The first years students sit for

- fall mid–term exam
- exam at the end of the second module
- spring mid–term exam
- APT

These exams consist of multiple choice questions and free response questions.

The grade for the first and the second modules is determined as the weighted sum of

- grade for the fall mid–term exam (30%)
- grade for the exam at the end of the second module (60%)
- average grade for home assignments and control tasks in the first and the second modules (10%)

The final grade for the course is determined as the weighted sum of

- grade for the first and the second modules (25%)
- grade for the spring mid–term exam (35%)
- APT grade (35%)
- average grade for home assignments and control tasks in the third and the fourth modules (5%)

## **7. Main reading**

1. Wonnacott R.J. and Wonnacott T.H. Introductory statistics for business and economics. John Wiley & Sons, fourth edition, 1985 (for the first year students, WW).

## **8. Supplementary reading**

1. Aivazian S.A., Mhitarin V.S. Probability Theory and Applied Statistics. Moscow, UNITY, 2001 (in Russian).
2. Gmurman V.E. Probability Theory and Mathematical Statistics. Moscow, “Higher School”, 1998 (in Russian).
3. Gmurman V.E. Guide to problems on Probability Theory and Mathematical Statistics. Moscow, “Higher School”, 1998 (in Russian).
4. Shvedov A.S. Probability Theory and Mathematical Statistics. Moscow, HSE, 1995 (in Russian).
5. Hogg R.V. and Tanis E.A., Probability and Statistical Inference, Prentice Hall, 1993.

## 9. Internet resources

1. <http://www.stat.yale.edu/Courses/1997-98/101/expdes.htm>
2. <http://www.itl.nist.gov/div898/handbook/pri/section3/pri3.htm>
3. [http://www.cas.lancs.ac.uk/glossary\\_v1.1/main.html](http://www.cas.lancs.ac.uk/glossary_v1.1/main.html)
4. <http://home.xnet.com/~fidler/triton/math/review/mat170/mat170.htm>
5. <http://www.statsoft.com/textbook/esc.html>
6. <http://mail.pittstate.edu/~winters/tutorial/>

## 10. Course outline

### 1. Primary data analysis (*WW Ch. 2*)

- 1.1. Graphical representation of the data. Dotplot. Steam– and leaf–plot. Histogram.
- 1.2. Outliers. Clusters. The shape of histogram.
- 1.3. Elementary statistics. Characteristics of the center of the distribution: arithmetic mean, median, mode, geometric mean. Characteristics of the dispersion of the distribution: range, mean quadratic deviation, interquartile range, mean absolute deviation, mean relative deviation. Box–plot.
- 1.4. Linear transformation of the data and transformation of the elementary statistics.
- 1.5. Characteristics of observations' positions: quartiles, percentiles, z-score.
- 1.6. Grouped observations and their statistics.

### 2. Data collection; planning and conducting experiments (*WW, Ch. 1*)

- 2.1. The methods of data collection: census, sample survey, experiment, observational study.
- 2.2. Population, samples and random selection
- 2.3. Sources of bias in sampling and surveys.
- 2.4. Sampling methods: simple random sampling, stratified random sampling, cluster sampling.
- 2.5. Planning and conducting experiments. Characteristics of well-designed experiment.
- 2.6. Treatments, control groups, random assignments and replication.
- 2.7. Sources of bias and confounding, including placebo effect and blinding.
- 2.8. Completely randomized design, block design

### 3. Elements of Probability Theory (*WW, Ch. 3, 4*)

- 3.1. Experiment with random outcomes. Interpretation probability including long–run relative frequency interpretation. “Law of Large Numbers” concept.
- 3.2. The space of elementary outcomes as a mathematical model of the experiment with random outcomes. Algebra of events. Disjoint events.
- 3.3. Probability in the space of elementary outcomes. The notion of classic probability. Elementary combinatorics. Addition rule.
- 3.4. Conditional probability. Multiplication rule. Independent events.
- 3.5. Formula of full probability. Bayes formula.

- 4. Discrete random variables (WW, Ch. 3, 4)**
  - 4.1. The examples of discrete random variables. Distributions of discrete random variables. Relative frequencies and cumulative frequencies.
  - 4.2. Expectation and variance of a discrete random variable. Standard deviation.
  - 4.3. Binomial distribution. Geometric distribution.
  
- 5. Continuous random variables (WW, Ch. 3, 4)**
  - 5.1. The examples of Continuous random variables. Cumulative distribution function. Density function. Expectation and variance. Standard deviation.
  - 5.2. Normal (Gaussian) distribution and its properties. Using tables of the normal distribution.
  - 5.3. Linear transformation of the random variables.
  
- 6. Joint distribution of two random variables (WW, Ch. 5)**
  - 6.1. Joint distribution of two discrete random variables. Marginal distributions. Conditional distributions. Conditional expectation.
  - 6.2. Independent random variables.
  - 6.3. Covariance. Correlation as a measure of linear relationship of two random variables. Uncorrelated and independent random variables. Expectation and variance of the linear combination of two random variables.
  
- 7. Limit Theorems (WW, Ch. 6)**
  - 7.1. The law of large numbers.
  - 7.2. Central Limit Theorem. Normal approximation for the binomial distribution.
  - 7.3. Joint distribution of two random variables.
  
- 8. Sampling distribution (WW, Ch. 6)**
  - 8.1. Distribution of sampling mean and sampling proportion.
  - 8.2. Sampling distribution of a difference between two independent sample means. Sampling distribution of a difference between two independent sample proportions.
  - 8.3.  $t$ -distribution. Chi-square distribution.
  
- 9. Point parameter estimates (WW, Ch. 7)**
  - 9.1. Point parameter estimators. Examples: sampling mean and sampling variance. Expectation and variance of the sampling mean. Estimation of the proportion.
  - 9.2. Properties of estimators: unbiasedness, efficiency, consistency.
  - 9.3. Estimators of the population mean and variance.
  - 9.4. Estimator of the population proportion.
  
- 10. Interval estimates (WW, Ch. 8)**
  - 10.1. Confidence intervals. Confidence interval for population mean. Normal approximation for large samples. Small samples and  $t$ -distribution.

- 10.2. Confidence intervals for difference between two population means (independent and matched samples).
- 10.3. Confidence intervals for difference between two proportions.
- 10.4. One- and two-sided confidence intervals.

**11. Hypothesis testing (WW, Ch. 9, 17)**

- 11.1. Hypothesis and statistical test. Test for a population mean. Using confidence intervals and test-statistics.
- 11.2. One-sided and two-sided tests.  $P$ -values.
- 11.3. Type I error and type II error. Significance and power of a test.
- 11.4. Standard tests for
  - 11.4.1. population mean
  - 11.4.2. population proportion
  - 11.4.3. difference between means for independent and matched samples
  - 11.4.4. difference between proportions
- 11.5. Chi-square test (goodness of fit test). Contingency tables.

**12. Simple regression and sampling correlation (WW, Ch.11, 12, 15)**

- 12.1.  $X$ - $Y$  graph. Fitting a line. Ordinary least squares.
- 12.2. Transformations of nonlinear models into linear models.
- 12.3. Outliers.
- 12.4. Fitted values.
- 12.5. Errors and residuals of regression model.
- 12.6. Statistical properties of regression estimates. Confidence interval for the slope. Hypothesis testing.
- 12.7. Population and sampling correlation.
- 12.8. Correlation and simple regression. Coefficient of determination.