



**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(ФИНАНСОВЫЙ УНИВЕРСИТЕТ)**

Департамент математики

О.Е. Пыркина

PROBABILITY THEORY AND MATHEMATICAL STATISTIC FOR APPLICATIONS IN DATA ANALYSIS

Учебное пособие

по дисциплине «Анализ данных» (на английском языке)

Направление подготовки: 38.03.01 Экономика

профили:

Международные финансы

Международная торговля и налогообложение,

Мировая экономика и международный бизнес,

Мировые финансы,

Международный бизнес энергетических компаний,

Программы подготовки бакалавра

С реализацией или частичной реализацией образовательных
программ на английском языке



**МОСКВА
2023**



**FEDERAL STATE-FUNDED EDUCATIONAL INSTITUTION
OF HIGHER EDUCATION
«FINANCIAL UNIVERSITY UNDER THE GOVERNMENT
OF RUSSIAN FEDERATION»
Department of Mathematics**

O.E. Pyrkina

PROBABILITY THEORY AND MATHEMATICAL STATISTIC FOR APPLICATIONS IN DATA ANALYSIS

Textbook
(in English)

BSc in Economics (38.03.01)

Concentration:
International Finance
International Trade and Taxation
Accounting and Financial Analysis
World Economy and International Business
World Finance
Economics and Finance of the Fuel and Energy Sector



МОСКВА
2023

УДК 519.2
ББК 22.17
П95

Рецензенты:

С.А. Зададаев, кандидат физико-математических наук, доцент, руководитель департамента математики;

В.В. Булатов, доктор физико-математических наук, доктор экономических наук, профессор, действительный член РАН, ведущий научный сотрудник Института проблем механики им. А.Ю. Ишлинского РАН

Пыркина О.Е.

П95 Теория вероятностей и математическая статистика для применения в анализе данных: Учебное пособие / О.Е. Пыркина. — М.: Прометей, 2023. — 582 с.

ISBN 978-5-00172-475-9

Учебное пособие "Probability Theory and Mathematical Statistic for Applications in Data Analysis" на английском языке (русский вариант названия — «Теория вероятностей и математическая статистика для применения в анализе данных») готовит читателей к успешной работе с информацией в рамках современной науки о данных (Data Science). Продуктивное развитие цифровой экономики невозможно без умения специалистов грамотно и эффективно оперировать непрерывно поступающим потоком цифровых данных статистического характера. Для обработки таких данных и принятия управленческих решений на основе данных необходимы умения и навыки как технического, так и теоретического уровня, позволяющие проводить обобщения и делать выводы на основе поступившей информации.

В пособии последовательно рассматриваются традиционные темы курсов теории вероятностей и математической статистики как теоретической основы анализа данных. Все вопросы курса рассматриваются с применением статистических функций и пакета анализа данных Excel. Курс дополнен примерами, задачами и тестовыми вопросами для самопроверки. Пособие состоит из 20 глав, введения и заключения.

Учебное пособие может быть использовано студентами и преподавателями университетов (в частности, Финансового университета при Правительстве РФ) в курсе дисциплины «Анализ данных» (дисциплины базовой части математического цикла дисциплин по направлению подготовки 38.03.01 «Экономика», профили: «Международные финансы» (на английском языке), «Международная торговля и налогообложение» (на английском языке), «Мировая экономика и международный бизнес» (с частичной реализацией на английском языке), «Мировые финансы» (с частичной реализацией на английском языке), «Международный бизнес энергетических компаний» (с частичной реализацией на английском языке), программы подготовки бакалавра.

Одобрено советом департамента математики, протокол №03 от 23 сентября 2022 года.

© Пыркина О.Е., 2023

ISBN 978-5-00172-475-9

© Издательство «Прометей», 2023

UDC 519.2
BBC 22.17
P95

Readers:

S.A. Zadadaev, Ph.D Physics & Mathematics, associate professor, Head of Mathematics Department
V.V. Bulatov, Doctor of Science Physics & Mathematics, Doctor of Science Economics, professor, Member of Russian Academy of Natural Science, Leading Research Scientist in the Ishlinsky Institute for Problems in Mechanics of the Russian Academy of Sciences

Pyrkina O.E.
P95 Probability Theory and Mathematical Statistics
for Applications in Data Analysis: Textbook /
O.E. Pyrkina. — M.: Prometej, 2023. — 582 pages.

ISBN 978-5-00172-475-9

The textbook "Probability Theory and Mathematical Statistics for Applications in Data Analysis" prepares readers for successful operation with information as a part of contemporary data science. The productive formation and development of the digital economy is impossible without the ability of specialists to operate competently and effectively with a continuously incoming stream of digital statistical data. To process such data and to make management decisions based on the data, skills and abilities of both technical and theoretical levels are required, that allows to carry out generalizations and make conclusions based on the information received.

The textbook discusses step by step traditional topics of courses in probability theory and mathematical statistics as a theoretical foundation for data analysis. All course questions are considered with application of statistical functions and the Excel data analysis package. The course is supplemented with examples, tasks and test questions for self-examination. The textbook includes 20 chapters, an introduction and conclusion.

The textbook can be used by students and lecturers of universities (in particular, the Financial University under the Government of the Russian Federation) in the course of "Data Analysis" (disciplines of the basic part of the mathematical cycle of disciplines, for a field of study 38.03.01 "Economics", study programs (concentrations): "International Finance" (in English), "International Trade and Taxation" (in English), "World Economy and International Business" (with partial implementation in English), "World Finance" (with partial implementation in English), "International Business of Energy companies" (with partial implementation in English), level of study: bachelor's degree programs.

© Pyrkina O.E., 2023
ISBN 978-5-00172-475-9 © Prometheus publishing house, 2023

ОГЛАВЛЕНИЕ

Chapter 1. Event Algebra. Basic Concepts.....	12
1.1. Introduction: What is Probability?.....	12
1.2. Random experiment	13
1.3. Events.....	15
<i>Self-testing questions</i>	27
Chapter 2. Probability and Its Postulates. Probability Rules.....	29
2.1. How Could We Define Probability?	29
2.2. Formalism: Postulates and Consequences	29
2.3. Introduction into Combinatorial Calculus:	
Permutation and Combinations	35
<i>Self-testing questions</i>	42
<i>Self-testing questions: answers</i>	47
Chapter 3. Conditional Probability. Statistical Independence....	48
3.1. The notion of conditional probability	48
3.2. The multiplication rule of probabilities	
and statistical independence	50
<i>Self-testing questions</i>	58
<i>Self-testing questions: answers</i>	63
Chapter 4. Bayes' Theorem and Total Probability Formula.	
Bivariate probabilities.....	64
4.1. Bayes' Theorem	64
4.2. Total Probability Formula	66
4.3. Bivariate probabilities: general setup.....	70
<i>Self-testing questions</i>	77
<i>Self-testing questions: answers</i>	79
Chapter 5. Random variables. Probability distributions for	
discrete random variables	80
5.1. Random variables	80
5.2. Probability distributions for discrete random	
variables	83
5.3. Expectations for Discrete Random Variables	89

5.4. Variances for Discrete Random Variables.....	93
5.5. The Linear Function of Discrete Random Variable.	95
<i>Self-testing questions</i>	99
<i>Self-testing questions: answers</i>	103
Chapter 6. Jointly Distributed Discrete Random Variables.....	105
6.1. The Joint probability Function and Marginal Probability Functions	105
6.2. The Conditional Probabilities and Independence of Discrete Random Variables	108
6.3. The Joint Cumulative Probability Function.....	111
6.4. The Covariance and Correlation Coefficient	112
<i>Self-testing questions</i>	120
<i>Self-testing questions: answers</i>	122
Chapter 7. Bernoulli Trials and Binomial Distribution. The Hypergeometric Distribution. The Geometric Distribution. The Poisson Distribution.....	125
7.1. Bernoulli Trials	125
7.2. The Binomial Distribution	128
7.3. The Hypergeometric Distribution	138
7.4. The Geometric Distribution	142
7.5. The Poisson Distribution	145
<i>Self-testing questions</i>	153
<i>Self-testing questions: answers</i>	155
Chapter 8. Continuous Random Variables.....	157
8.1. Continuous random variables: Statement of a Problem ...	157
8.2. Probability Distributions for continuous random variables.	158
8.3. Numerical characteristics for continuous random variables	167
8.4. Jointly Distributed Continuous Random Variables	171
8.5. Uniform Distribution: general view.....	176
8.6. Normal Distribution: general view	180
8.7. The central limit theorem.	197
8.8. The normal distribution as an approximation to the binomial and Poisson distributions.....	202

8.9. The Exponential Distribution.....	208
8.10. The Lognormal distribution	212
<i>Self-testing questions</i>	214
<i>Self-testing questions . Answers</i>	220
Chapter 9. Laws of large numbers.....	221
9.1. Chebyshev inequality	221
9.2. Laws of large numbers.	224
9.3. Bernoulli's theorem	230
<i>Self-testing questions</i>	232
<i>Self-testing questions. Answers</i>	235
Chapter 10. Moments of a single random variable and jointly distributed continuous random variables	236
10.1. Moments and higher-order moments of probability distribution	236
10.2. Moments of two or more random variables	243
10.3. Conditional distributions.....	244
10.4. Moment generating functions	250
<i>Self-testing questions</i>	256
<i>Self-testing questions. Answers</i>	258
Chapter 11. Jointly Distributed Continuous Random Variables.....	260
11.1. Joint density functions	260
11.1. Function of two random variables.....	267
11.2. Bivariate normal distribution	277
<i>Self-testing questions</i>	285
<i>Self-testing questions. Answers</i>	287
Chapter 12. Introduction in the theory of Markov Chains	288
12.1. The main notions	288
12.2. Specifying a Markov Chain.....	290
12.3. Long-term behavior of a Markov chain	294
12.4. Absorbing Markov Chains.....	297
<i>Self-testing questions</i>	305
<i>Self-testing questions. Answers</i>	306

Chapter 13. Summarizing Numerical Information.....	307
13.1. Population and samples	307
13.2. Distinction between two types of data sets	308
13.3. Numerical Summary: Measures of Central Tendency ...	309
13.4. Numerical Summary: Measures of Dispersions	314
<i>Self-testing questions</i>	323
<i>Self-testing questions . Answers</i>	324
Chapter 14. Summarizing Numerical Information for Grouped Data	326
14.1. Grouping the observations	326
14.2. Numerical summary of grouped data	335
<i>Self-testing questions</i>	348
<i>Self-testing questions. Answers</i>	350
Chapter 15. Sampling and Sampling Distributions	351
15.1. Sampling from a population	351
15.2. Sampling distribution of the sample mean.	355
15.3. Sampling distribution of the sample proportion	362
15.4. Sampling distribution of the sample variance	366
<i>Self-testing questions</i>	375
<i>Self-testing questions. Answers</i>	377
Chapter 16. Point Estimations and Methods of its' Creation	379
16.1. Introduction: main definitions.....	379
16.2. Unbiased estimators, their efficiency and consistency	383
16.3. Method of moments	392
16.4. Method of Maximum Likelihood Estimation	394
<i>Self-testing questions</i>	403
<i>Self-testing questions. Answers</i>	404
Chapter 17. Confidence intervals	405
17.1. Interval estimation: introduction	405
17.2. Interval estimation: the center and boundaries	407
17.3. Confidence intervals for the mean of a normal distribution: population variance known	410

17.4. Confidence intervals for the mean of a normal distribution: population variance unknown, large sample size	419
17.5. The Student's t Distribution	421
17.6. Confidence intervals for the mean of a normal distribution: population variance unknown, small sample size	425
17.7. Confidence intervals for the population proportion (large samples)	428
17.8. Confidence intervals for the variance of a normal population.....	431
17.9. Estimating the sample size.....	435
<i>Self-testing questions</i>	440
<i>Self-testing questions. Answers</i>	442
Chapter 18. Hypothesis Testing	443
18.1. The concept of statistical hypothesis testing.....	443
18.2. Tests of the mean of a normal distribution: simple null, population variance known	454
18.3. What is meant by the rejection of a null hypothesis? P-value of the test	457
18.4. Tests of the mean of a normal distribution: population variance known. Composite null and alternative hypothesis.....	459
18.5. Test of the mean of a normal distribution, population variance unknown: large sample sizes.	464
18.6. Test of the mean of a normal distribution, population variance unknown.....	467
18.7. Test of the variance of a normal distribution	470
18.8. Test of the population proportions (large samples). ...	474
18.9. Tests for the differences between two means. Test based on matched pairs. Test based on independent samples.....	477
18.10. Tests for the differences between two population proportions (large samples).....	487
18.11. Testing the equality of the variances of two normal populations. F-distribution	491

18.12. Measuring the power of a test	495
18.13. Some comments of hypothesis testing	503
18.14. Test of normality	505
18.15. Goodness-of-fit tests	508
18.16. A test of association in contingency tables.....	517
<i>Self-testing questions</i>	521
<i>Self-testing questions. Answers</i>	528
Chapter 19. Some nonparametric tests.....	531
19.1. Introduction. The sign test	531
19.2. The Wilcoxon test	537
19.3. The Mann-Whitney test.....	543
19.4. Discussion	549
<i>Self-testing questions</i>	551
<i>Self-testing questions. Answers</i>	554
Chapter 20. ANOVA (Analysis of variance).....	554
20.1. Comparison of several population means.....	554
20.2. One-way analysis of variance	561
20.3. The Kruskal-Wallis test.....	573
<i>Self-testing questions</i>	577
<i>Self-testing questions. Answers</i>	579
Bibliography	580