МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЗАТВЕРДЖЕНО
на засіданні кафедри
статистики і економічного прогнозування
Протокол № 2 від 2.09.2024 р.

ПОГОДЖЕНО
Проректор з навчально-методичної роботи

Карина НЕМАШКАЛО

ОСНОВИ АНАЛІЗУ ДАНИХ
робоча програма навчальної дисципліни (РПНД)

Галузь знань	**12 Інформаційні технології**
Спеціальність	**122 Комп'ютерні науки**
Освітній рівень	**другий (магістерський)**
Освітня програма	**Комп'ютерні науки**

Статус дисципліни	**вибіркова**
Мова викладання, навчання та оцінювання	**англійська**

Розробники		
д.е.н., професор	підписано КЕП	Олена РАЄВНЄВА
к.е.н., доцент		Ольга БРОВКО
д.е.н., професор		Костянтин СТРИЖИЧЕНКО
викладач		Мар'яна СЕМКІВ
Завідувач кафедри статистики і економічного прогнозування	підписано КЕП	Олена РАЄВНЄВА
Гарант програми	підписано КЕП	Сергій МІНУХІН

Харків
2024

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

APPROVED
at the meeting of the department
statistics and economic forecasting
Protocol № 2 of 2.09.2024.

AGREED
Vice-rector for Educational and Methodical work

Karina NEMASHKALO

BASICS OF DATA ANALYSIS
Program of the course

Field of knowledge	**12 "Information technologies"**
Specialty	**122 Computer sciences**
Study cycle	**Second (master's**
Study programme	**"Computer sciences"**

Course status	**Elective**
Language	**English**

Developer:
Doctor of Economics, professor — digitally signed — Olena RAYEVNYEVA
PhD, associate professor — Olha BROVKO
Doctor of Economics, professor — Konstantin STRYZHYCHENKO
teacher — Mariana SEMKIV

Head of the Department of Statistics and economic forecasting — digitally signed — Olena RAYEVNYEVA

Head of Study Programme — digitally signed — Serhii MINUKHIN

Kharkiv
2024

INTRODUCTION

Data analytics is a science that studies important information and large amounts of data. It combines statistical methods of data analysis with the use of a number of IT technologies. The discipline offers a broad approach to data analysis and data retrieval methods and their application in practice. It combines cutting-edge research and practice in related fields and provides students with the necessary knowledge and skills to initiate and conduct data analysis that can transform company data into commercially valuable information.

The study of the course "Basics of data analysis" belongs to the cycle of selective disciplines of the "master's" educational degree for applicants of the specialty 122 "Computer science" of the educational and professional program "Computer science" and is aimed at the formation of basic competencies in the applicants regarding the ability to analyze, objectively evaluate phenomena and processes, establish the dependence of the performance indicator on a set of factors, identify changing trends in the studied phenomena and forecast their development and interpretation of the obtained research results, etc.

The purpose of the course: the formation of theoretical knowledge, applied skills and abilities regarding the organization of statistical observations, methods of statistical analysis and forecasting of socio-economic phenomena and processes.

The main tasks of studying the course "Basics of data analysis" are: determination of the principles of observation;

mastering data processing skills;

studying the methods of building indexes and their research;

formation of the sample and its analysis from the point of view of representativeness;

construction of data research indicators and their analysis.

The object of the course is socio-economic processes and components of social, cultural, economic, information and business systems.

The subject of the course is fundamental statistical methods of data research.

The learning outcomes and competencies formed by the course are defined in table 1.

Table 1

Learning outcomes and competencies formed by the course

Learning outcomes	Competencies
LO1. Have specialized conceptual knowledge that includes current scientific achievements in the field of computer science and is the basis for original thinking and conducting research, critical thinking of problems in the field of computer science and at the border of fields of knowledge.	SC02. The ability to formalize the subject area of a certain project in the form of an appropriate information model.
LO7. Develop and apply mathematical methods for the analysis of information models.	GC02. Ability to apply knowledge in practical situations
	SC03. Ability to use mathematical methods to analyze formalized models of the subject area.

LO8. Develop mathematical models and data analysis methods (including large ones).	SC 03. Ability to use mathematical methods to analyze formalized models of the subject area.
LO9. Develop algorithmic and software for data analysis (including large data).	GC02. Ability to apply knowledge in practical situations
	SC05. Ability to develop, describe, analyze and optimize architectural solutions of information and computer systems for various purposes.
LO19. To analyze the current state and global trends in the development of computer sciences and information technologies.	SC08. The ability to develop and implement software development projects, including in unpredictable conditions, with unclear requirements and the need to apply new strategic approaches, use software tools to organize teamwork on the project.
	SC12. Ability to develop, apply and integrate data processing and analysis technologies in high-performance systems and cloud platforms to ensure efficient use of computing resources of computer systems.

COURSE CONTENT

Content module 1. Methodological foundations of data analysis

Topic 1. Nature of observation of socio-economic objects

1.1. The notion of statistical observation. Formation of the information base of statistical observation.

1.2. Method of analytical groupings.

1.3. Provision of statistical data: tables, graphs, maps.

Topic 2. Selection: its construction and verification of its representativeness

2.1. Basis of selective observation, reasons and conditions of its application.

2.2. The theoretical basis of a selective observation. Types and methods of selecting units in selective population, providing representativeness of the sample.

2.3. Classification of sampling errors. The procedure for determining sampling errors (representativeness) with different methods of selection.

2.4. Methods of distributing the results of selective observation to the general totality.

2.5. Determination of the necessary sample size.

Topic 3. Grouping data and study of groups

3.1. Grouping is the basis for scientific data

3.2. Methods of visualization for grouped data

Content module 2. Methods of research and analysis of components of socio-economic systems

Topic 4. Indicators research of data

4.1. The essence of research indicators.

4.2. Types and forms of indicators (indicators) in data analysis.

4.3. Ways and methods of calculating research indicators.

Topic 5. World indexes and their research

5.1. Indexs and their classification

5.2. General index of quantitative indicators

5.3. General index of qualitative indicators

5.4. Two-factor systems of interconnected indices

5.5. Index method for analyzing the dynamics of the average level

5.6. Chain and basic indexes

The list of laboratory studies in the course is given in table 2.

Table 2

The list of laboratory studies

Name of the topic and/or task	Content
Topic 1.	Laboratory work on topic 1. Formation of a representative sample.
Topic 2.	Laboratory assignment on topic 2. Means of visualization of statistical data
Topic 3.	Laboratory work on topic 3. Data grouping and group research
Topic 4.	Laboratory work on topic 4. Statistical means of studying the dynamics of phenomena and processes.
Topic 5.	Laboratory work on topic 5. Indicators of the study of world trends.

The list of self-studies in the course is given in table 3.

Table 3

List of self-studies

Name of the topic and/or task	Content
Topic 1-5	Studying lecture material
Topic 1-2	Essay writing Solving a situational (case) task
Topic 1-5	Performing laboratory work

The number of hours of lecture and laboratory classes and hours of independent work is given in the work plan (technological map) for the course.

TEACHING METHODS

In the process of teaching course, in order to acquire certain learning outcomes, to activate the educational process, it is envisaged to use such teaching methods as:

group work (Topic 1-2), problem lectures (Topic 1), creating situations of cognitive novelty (topics 4-5.

in person (demonstration (Topic 1-5)).

practical (laboratory work (Topic 2-5), essay (Topic 1), test tasks (Topic 1-5), etc.).

FORMS AND METHODS OF ASSESSMENT

The University uses a 100-point cumulative system for assessing the learning outcomes of students.

Current control is carried out during lectures and laboratory classes and is aimed at checking the level of readiness of the student to perform a specific job and is evaluated by the amount of points scored:

− for disciplines with a form of semester control examination (exam): the maximum amount is 60 points; the minimum amount that allows a student of higher education to pass an exam is 35 points

The **final control** includes current control and an exam.

Semester control is conducted in the form of a semester exam (exam). The semester exam (exam) is taken during the exam session.

The maximum number of points that a student of higher education can receive

during the examination (examination) is 40 points. The minimum amount for which the exam is considered passed is 25 points.

The **final grade i**n the course is determined:

– for disciplines with a form of exam, the final grade is the amount of all points received during the current control and the exam grade.

During the teaching of the academic discipline, the following control measures are used:

Current control: Laboratory works (25 points), essay in the form of a presentation (10 points), test control (15 points), written control works (10 points).

Semester control: Grading including Exam (40 points).

More detailed information about the evaluation system is provided in the work plan (technological map) for the course.

An example of an examination paper

Semyon Kuznets Kharkiv National University of Economics
Second (master's) level of higher education
Specialty 122 "Computer sciences"
Educational and professional program "Computer sciences"
Course "Basics of data analysis"

EXAM CARD № 1

Stereotype task (tests). (20 points)

1	Indicators characterizing volumes and dimensions of socio-economic phenomena are expressed in units of measurement: a) in kind, labor and cost; b) coefficients, percentages, inches; c) all answers are correct.
2	The population census refers to the observation of: a) selective; b) monographic; c) continuous; d) main massif.
3	Marketing research on the adaptation of a new product brand on the market refers to the observation of: a) current; b) simultaneous; c) periodic.
4	The seasonality index is: a) the percentage ratio of individual levels to the average level of a given series of dynamics; b) the ratio of individual levels to the total sum of levels of the dynamics series; c) moving average of individual levels of the series.
5	The criterion of homogeneity of the population is considered to be: a) decile coefficient of differentiation; b) quartile coefficient of variation; c) root mean square coefficient of variation.
6	The possible deviation of the indicators of the sample population from the indicators of the general population is measured by: a) sampling error; b) dispersion; c) mean square deviation.

7	The statistical population is: a) phenomena and processes of social life; b) a set of indicators studied by statistics; c) a set of phenomena that are homogeneous in at least one characteristic.
8	During the survey of the population, residents of every fourth apartment were interviewed about the total income of families. Every fourth apartment had four rooms. As a result, the results of the family survey were incorrect. What type of error occurred during this examination? a) random error of representativeness; b) systematic registration error; c) random error of representativeness; d) systematic registration error.
9	Equipment inventory refers to monitoring: a) documentary observation; b) direct; c) survey; d) expert assessments.
10	The study of product quality at the enterprise refers to the observation of: a) selective; b) monographic; c) continuous; d) main massif.
11	How is observation carried out: a) direct; b) documented; c) survey; d) expert assessments.
12	Structural grouping is a) type of grouping, which allows dividing the group into groups, characterizing its structure by a variable feature; b) division of the studied population into socio-economic types or classes; c) a type of grouping that allows you to identify the presence and directionality of the connection of features.
13	The examination was carried out: 1) every 10th agricultural enterprise out of 150 that irrigates land at the expense of its own funds in order to study the efficiency of the use of irrigated areas; 2) agricultural firm "Koblevo" in order to study the reserves of increasing the efficiency of irrigation in this farm. Which of the examinations are selective? a) the first; b) the second; c) both the first and the second; d) does not have a correct answer.
14	How does a 4-fold increase in variance affect the sampling error? a) the sampling error will increase by 4 times; b) the sampling error will decrease by a factor of 2; c) the sampling error will not change; d) the sampling error will increase by 2 times.
15	Statistical observation is a) systematic, scientifically organized collection of data on mass social processes and phenomena with the help of registration of their essential features; b) registration of social phenomena and processes in special forms of statistical observation; c) collection of observation materials and their registration.

16	Grouping is: a) typology of socio-economic phenomena; b) distribution of the population into groups of essential characteristics; c) presentation, processing and calculation of group and general summaries characterizing the phenomenon d) classification of socio-economic phenomena and processes.
17	Which of the following composite index formulas is an index of the physical volume of product exports: a) $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$; b) $\frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}$; c) $\frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}$. d) $\frac{\Sigma q_0 p_0}{\Sigma q 1 p 0}$.
18	Setting the start and end date of data collection is a) objective observation time; b) subjective observation time; c) critical moment of observation.
19	The median in the distribution series is: a) the highest frequency (or value of the characteristic); b) the value of the sign that occurs most often; c) the value of the sign , mentally divides the distribution series into two equal parts.
20	The general price index is equal to $I_P = 0.92$. How has the cost of sales changed due to prices? Answers: 1) increased by 8%; 2) decreased by 8%; 3) changed by 0.92%. a) increased by 8%; b) decreased by 8%; c) changed by 0.92%.

Diagnostic task 1 (calculation test). (3 points)

The share of highly liquid assets in the sum of current assets is 18%. Find the dispersion of the share of highly liquid assets. Explain the calculations.

Diagnostic task 2 (essay). (7 points)

Make a conclusion about the localization and concentration of employees at enterprises according to the specified types of economic activity: the coefficient of localization at industrial enterprises is 3.4; at agricultural enterprises - 5.7; at trade enterprises – 7.8. The concentration factor is 85%. Explain the answer.

Heuristic task (calculation). (10 points)

We have data on the volume of merchandise turnover in the region, million hryvnias:

2017	2018	2019	2020	2021	2022	2023
135	120	110	130	145	156	167

Analyze this dynamic series using base and chain methods. Submit the obtained indicators in tabular form. Calculate the average annual absolute growth, average annual growth rates and average annual growth rates of turnover for 2017-2023. Explain the calculation and results. Draw conclusions based on the obtained indicators.

Approved at the meeting of the Department of Statistics and Economic Forecasting
protocol No. ____dated "___"_____20___year.

Examiner Doctor of Economics, Prof. O. Rayevnyeva

Chief Department PhD, Associate professor O. Brovko

Assessment criteria

Final scores for the exam consist of the sum of points for completing all tasks, rounded to a whole number according to the rules of mathematics.

The algorithm for solving each task includes separate stages that differ in complexity, time-consumingness, and importance for solving the task. Therefore, individual tasks and stages of their solution are evaluated separately from each other in this way:

stereotyped task in the form of closed tests – 20 tests, the maximum score per test is 1 point (the maximum number of points is 20);

diagnostic task 1 (calculation test) – maximum score 3 points:

1 point – selection of the calculation formula;

1 point – implementation of calculation actions;

1 point – formulation of the answer to the question.

diagnostic task 2 (essay)– maximum score 7 points:

1 point – understanding the essence of the task;

2 points – selection of calculation formulas or algorithm for task performance;

2 points – making a calculation or formulating an algorithm of actions to perform the task;

2 points - the presence of reasoned conclusions based on the results of analytical work;

heuristic task – maximum score 10 points:

2 points – understanding the essence of the task;

2 points – selection of calculation formulas that correspond to the essence of the questions;

2 points – calculation of indicators according to separate formulas related to the research topic;

2 points – calculation of indicators in full in accordance with the questions posed in the task;

2 points – providing reasoned conclusions based on the results of calculation work.

RECOMMENDED LITERATURE

Main

1. Rayevnyeva O. Statistics [Electronic resource] : textbook / O. Rayevnyeva, I. Aksonova, O. Brovko [et al.]; Simon Kuzdnets Kharkiv national university of economics. - E-text data (3,53 МБ). - Kharkiv : S. Kuznets KhNUE, 2020. - 376 p. : il. - The title screen. - Referenc.: p. 356-362. http://repository.hneu.edu.ua/handle/123456789/25678

2. Andrew Bruce, Peter Bruce, Peter Gedeck. Practical Statistics for Data Scientists. - O'Reilly Media, Inc., 2020 - 350 p.

3. Cheryl L. Jennings, Douglas C. Montgomery, Murat Kulahchi. Introduction to Time Series Analysis and Forecasting. - Wiley. John Wiley & Sons, LTD, 2015- 672 p.

4. Scott E. Page. The Model Thinker: What You Need to Know to Make Data Work for You. - Basic Books, 2019. - 448 p.

5. Alfonso Zamora Sais, Carlos Quesada Gonzalez, Diego Mondejar Ruiz, Luis Hurtado Gil. An Introduction to Data Analysis in R: Hands-on Coding, Data Mining, Visualization and Statistics from Scratch. - Springer, 2020. - 276 p.

6. Галущак М. П., Галущак О. Я., Кужда Т. І. Прогнозування соціально-економічних процесів: навчальний посібник для економічних спеціальностей. – Тернопіль: ФОП Паляниця, 2021. – 160 с.

7. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.

8. Литвин В.В., Пасічник В.В. Аналіз даних та знань: навчальний посібник. К.: Магнолія 2006, 2023. - 276 с.

9. Муляр В. П. Візуалізація даних та інфографіка. Харків: ФОП Панов А. М. 2020. 200 с.

10. Статистичний аналіз даних : навчальний посібник / Т. М. Паянок, Т. М. Задорожня. – Ірпінь : Університет державної фіскальної служби України, 2020. – 312 с.

11. Чекотовський Е. В. Статистичні методи на основі Microsoft Excel 2016: навч. посібник. Київ: Знання, 2018. 408 с.

Additional

12. Rayevnyeva O. The Impact of a Sensitivity of Economic Activities on the Economic Behaviour of Enterprise / O. Rayevnyeva, S. Filip, I. Aksonova and other // Economics of Development. - 2022. - Vol. 21. - No. 3. - P. 27-39. http://repository.hneu.edu.ua/handle/123456789/29607

13. Miller, J. A., Aldosari, M., Saeed, F., Barna, N. H., Rana, S., Arpinar, I. B., & Liu, N. (2024). A survey of deep learning and foundation models for time series forecasting. arXiv preprint arXiv:2401.13912.

14. Iqbal, A., & Amin, R. (2024). Time series forecasting and anomaly detection using deep learning. Computers & Chemical Engineering, 182, 108560.

15. Аналіз даних та багатовимірна статистика : робоча програма навчальної дисципліни для здобувачів вищої освіти спеціальності 076 «Підприємництво та торгівля» освітньо-наукової програми «Підприємництво, торгівля та біржова діяльність» третього (освітньо-наукового) рівня: [Електронне видання] / укл. І. В. Аксьонова. – Харків : ХНЕУ ім. С. Кузнеця, 2024. – 8 с. (Укр. мов.) http://repository.hneu.edu.ua/handle/123456789/32313

16. Горват А.А., Молнар О.О., Мінькович В.В. Обробка, візуалізація та аналіз експериментальних даних з використанням пакету Origin: Навчальний посібник. Ужгород: Видавництво УжНУ “ Говерла”, 2020. – 64 с.:

17. Доценко С. І. Організація та системи керування базами даних: Навч. посібник. – Харків: УкрДУЗТ, 2023. – 117 с.

18. Корват О. В. Перспективи використання Microsoft Power BI у статистичному аналізі даних / О. В. Корват. // Розвиток бухгалтерського обліку,

оподаткування і контролю в умовах інтеграційних процесів : матеріали Всеукр-ої наук.-практ. інтернет-конф., 22-23 жовт. 2020 р. – Херсон : ДВНЗ «ХДАУ», 2020. - С. 307-308http://repository.hneu.edu.ua/bitstream/123456789/24849/1/%D0%9A%D0%BE%D1%80%D0%B2%D0%B0%D1%82_%D0%A5%D0%94%D0%90%D0%A3_2020.pdf

19. Прогнозування соціально-економічних процесів : конспект лекцій / уклад.: О. В. Шебаніна та ін. Миколаїв : МНАУ, 2022. - 95 с.

20. Статистичні методи та моделі оцінювання й прогнозування поведінки соціально-економічних систем в умовах інформаційної економіки : звіт кафедри статистики і економічного прогнозування за 2019-2020 навч. рік / викон. Раєвнєва О. В., Аксьонова І. В., Бровко О. І. та ін. – Х.: ХНЕУ ім. С. Кузнеця, 2020. - 107 с. http://repository.hneu.edu.ua/handle/123456789/23792

21. Управління бізнес-процесами підприємства. Комплексний тренінг: навч. посіб. / П.Г. Банщиков, В.М. Гордієнко, О.О. Кизенко, Г.С. Скитьова. – К.: КНЕУ, 2018. – 283 с.

Information resources

22. Газін А. Візуалізація даних як навичка // А. Газін [Електронний ресурс]. – Режим доступу: https://aiukraine.com/wp-content/uploads/2017/10/1_4-Gazin.pdf

23. Офіційний сайт Всесвітнього Банку [Електронний ресурс]. – Режим доступу: http://www.worldbank.org/

24. Офіційний сайт Європейського центрального банку [Електронний ресурс]. – Режим доступу: https://www.ecb.europa.eu/stats/html/index.en.html

25. Офіційний сайт департаменту статистики Організації Об'єднаних Націй [Електронний ресурс]. – Режим доступу: http://unstats.un.org/unsd/default.htm

26. Course page on the Moodle platform (personal learning system). - Access mode: https://pns.hneu.edu.ua/course/view.php?id=4705