



Computer analysis of social information and data visualization

Syllabus

Short description of the discipline

Level of higher education	<i>Master's level</i>
Field of science	<i>05 Social and behavioral sciences</i>
Speciality	<i>054 Sociology</i>
Educational program	<i>Social Data Analytics</i>
Status of the discipline	<i>Mandatory</i>
Forms of study	<i>full-time</i>
Year of study, semester	<i>5th year of study, first semester</i>
Number of credits / hours	<i>5 ECTS credits /150 hrs: lectures – 18 hrs, seminars – 36 hrs, independent work – 96 hrs.</i>
Form of semester control	<i>Exam, calculation and graphic work, module test.</i>
Schedule	<i>https://schedule.kpi.ua/</i>
Language of instruction	<i>English</i>
Information about lecturers / instructors	<i>Lecturer: Hennadii Korzhov, Associate Professor, PhD in Sociology, korzhovga@gmail.com</i>
Course placement	<i>Google Classroom: https://classroom.google.com/c/ODAYmzY5NDgzNTU5?cjc=uukng5tq</i>

Program of the discipline

1. Description of the academic discipline, its purpose, subject of study and learning outcomes

In the modern world, we are daily confronted with huge amounts of information about various aspects of social life, a significant part of which is presented in numerical form. Understanding what is behind these numbers, whether they are really an objective reflection of social realities, is a task that everyone must solve to make the right choice, to plan their lives wisely and not to become the object of sophisticated or cynical manipulations. The ability to find, process and analyze information is becoming a vital competence for every conscious citizen and educated person. As a result of mastering the discipline, students will learn practical skills and abilities to work with various numerical information that characterizes social structures, processes and phenomena; the ability to use specialized programs to create their own databases and work with ready-made large amounts of social information that are available on the Internet. Students will have the opportunity to acquire basic skills and abilities in processing and analyzing statistical and sociological data, as well as the ability to work critically and creatively with large amounts of information, discover

hidden social connections and patterns, and to be able to see individual facts of social reality through the lenses of broader socio-cultural and socio-political processes.

In accordance with the requirements of the OPP, the **purpose** of the discipline is to form the following **abilities** in students:

- Ability to abstract thinking, analysis and synthesis (ZK 1)
- Ability to present the results of scientific research and prepare scientific reports and publications (FK 10)

The task of the discipline is to form the following **learning outcomes**:

- Analyze social phenomena and processes using empirical data and modern concepts and theories of sociology (PR01)
- Apply scientific knowledge, sociological and statistical methods, digital technologies, specialized software to solve complex problems of sociology and related fields of knowledge (PR04)
- Generalize the results of their own scientific research and present them in scientific reports, publications (PR11)

As a result of mastering the discipline, students will have the opportunity to creatively apply various methods of processing and analyzing quantitative sociological information; to formulate, substantiate and empirically verify research hypotheses of a descriptive and explanatory type; to create and modify databases (arrays), perform various transformations of variables in order to deepen the analysis; and also to discover latent, hidden social connections and patterns through quantitative analysis of sociological research data.

In the modern world of information and communication technologies, every highly qualified specialist deals with large volumes of various information. Students will be able to more effectively use various numerical information about various spheres of life (economic, financial, social, demographic, political processes, views, attitudes and values of people, consciousness and behavior of certain groups, etc.) in their professional activities and everyday life. This will contribute not only to a more conscious and creative use of information, improve the skills of collecting, processing, transforming, modifying, analyzing and interpreting information sources, but also to improving existing and acquiring new skills and competencies that will help you feel more confident in the labor market and compete more effectively for a better, meaningful, creative, and prestigious work.

2. Prerequisites and postrequisites of the discipline (place in the structural and logical scheme of study according to the relevant educational program)

Prerequisites: it is desirable to have basic skills in working with specialized software for processing and analyzing quantitative social information (e.g., R), with sociological data archives (e.g., National Bank of Sociological Data "Kyiv Archive", GESIS, etc.), basic level of English not lower than A2.

In parallel, the following educational components have to be mastered: "Computer analysis of social information and data visualization" with the following educational components: "Analytical sociology and

social behavior: modern approaches", "Methodology and methods of sociological research in the digital age".

The knowledge, skills and abilities obtained during training in this discipline can be successfully applied in mastering other special sociological disciplines.

3. Content of the academic discipline

Section 1. Basic computer analysis of sociological information

Topic 1. Fundamentals of statistics for the analysis of sociological data

Topic 2. Univariate analysis of sociological data

Topic 3. Data selection and transformation

Topic 4. Data visualization: Graphs

Topic 5. Analysis of two-dimensional tables

Topic 6. Analysis of multiple responses

Topic 7. Nonparametric criteria. Comparison measures

Section 2. Principles and procedures of multivariate analysis of sociological data

Topic 8. Regression analysis

Topic 9. Logistic regression

Topic 10. Discriminant analysis

Topic 11. Analysis of variance

Topic 12. Factor analysis

Topic 13. Path analysis

Topic 14. Cluster analysis

Topic 15. Data visualization: Interactive modes of working with graphs

4. Educational materials and resources

4.1. Basic literature:

1. Паніна Н.В. Технологія соціологічного дослідження: Курс лекцій. – Київ, 2000.
2. Паніотто В.І., Максименко В.С., Харченко Н.М. Статистичний аналіз соціологічних даних. – Київ: КМА, 2004.
3. Практикум з математичної статистики в психології. Приклади у системі SPSS: навчальний посібник / Р.М. Білоус, В.П. Черненко. – Кременчук: ПП Щербатих О.В., 2020.

4.2. Additional literature

1. Горбачик А.П., Сальнікова С.А. Аналіз даних соціологічних досліджень засобами SPSS: Навч. пос. – Луцьк: Вежа; Волин. нац. ун-т ім. Лесі Українки, 2008. – 164 с.
2. Нг А., Су К. Опануй числа! Наука про дані для нефахівців. Харків: Фабула, 2024.
3. Пінкер С. Раціональність. Київ: Лабораторія, 2024.
4. Шпігельгальтер Д. Мистецтво статистики: Прийняття аргументованих рішень на підставі даних. Київ: КМ-Букс, 2023.

5. Albers M. Introduction to Quantitative Data Analysis in the Behavioral and Social Sciences. Hoboken, N.J.: Wiley, 2017.
6. Antonius R. Interpreting Quantitative Data with SPSS. London: Sage, 2003.
7. Druckman D. Doing Research: Methods of Inquiry for Conflict Analysis / Daniel Druckman/ SAGE Publication. 2005. 387 p.
8. Greasley P. Quantitative Data Analysis Using SPSS An Introduction for Health & Social Science. Maidenhead: Open University Press, 2008.
9. Hallsworth M., Kirkman E. Behavioural Insights. Cambridge: MIT Press, 2020.
10. Ho, R. Handbook of univariate and multivariate data analysis and interpretation with SPSS. Boca Raton: Chapman & Hall, 2006.
11. Welch S., Comer S. Quantitative methods for public administration: Techniques and applications. 3rd ed. Long Grove: Waveland Press, 2006.

4.3. Educational online resources

1. <http://www.socio-journal.kpi.kiev.ua> – Bulletin of Igor Sikorsky Kyiv Polytechnic Institute. Sociology. Political Science. Law.
2. <http://i-soc.com.ua/journal/content.php> – Sociology: theory, methods, marketing.
3. <http://www.nbu.gov.ua> – National Library of Ukraine named after V.I. Vernadsky.
4. <https://prometheus.org.ua> – Online courses of Ukraine and the world.
5. <https://coursera.org> – Online courses.
6. www.ec.europa.eu/eurostat – Statistical information on the countries of the European Union.
7. www.stat.gov.ua – Statistical Service of Ukraine
8. <http://www.gesis.org> – Archive of data of international comparative sociological research.
9. <http://ukraine.survey-archive.com> – National Bank of Sociological Data “Kyiv Archive”.

All of the above sources are available online or on the discipline page in Google Classroom, as well as in the methodological office of the Department of Sociology (room 503, building 7). Required reading passages will be indicated during the course of the relevant topics.

Educational content

5. Methodology for mastering the academic discipline (educational component)

Lectures

Lecture 1. Fundamentals of statistics for analyzing sociological data

Basic concepts of the methodology of sociological research. Three main paradigms of sociological research. Quantitative (positivist, statistical) paradigm. Qualitative (understanding, interpretative) paradigm. Critical social science. Comparative characteristics of the three paradigms: the purpose of the study, the nature of social reality and human beings, the relationship between science and common sense, the essence of explanation and theory, criteria for the truth of knowledge, the place of values in scientific research. Statistical (quantitative) and humanitarian (qualitative) approaches in sociology.

Tasks for the students' self-education (SSE):

1. Compile a table "Comparison of three paradigms of sociological research".
2. Find an article in a scientific sociological journal that presents the results of a quantitative scientific study on a topic close to your master's thesis. Determine the type of this study. Determine the essence of the scientific problem, object and subject, goal and objectives of the study, hypotheses and operational definitions of basic concepts, methodology and main results.

Lecture 2. Univariate analysis of sociological data

Data matrix. Frequency, proportion. Statistical series and their ordering. Univariate distribution tables. The problem of "compression" of social information, selection of the most informative features. Use of descriptive statistics in the analysis.

Measures of central tendency. Arithmetic mean, its properties and calculations. Measures of central tendency for qualitative features: median, mode. Measures of variation. Dispersion and its properties. Coefficient of variation.

Tasks for the students' self-education (SSE):

1. Consider the procedures for ordering statistical series.
2. Get acquainted with the requirements for forming intervals in distribution series.
3. Calculate the mean, mode, and median for the proposed metric variables, in which individual values are combined into categories with ranges.

Lecture 3. Data selection and transformation. Data visualization: Graphs.

Observation selection. Classification of operators. Extraction of a random sample. Sorting observations. Distribution of observations into groups. Creation of a new variable using various procedures (Compute, Recode, Count). Conversion of variables from one type to another.

Logical expressions and functions. Indices in sociological research. Indirect measurement. Construction of indices as a measurement technique and as an integral part of the analysis of empirical information. Logical indices in sociology. Indices for comparing groups. Creating indices of total ratings (R. Likert scales). Basics of working with the syntax editor. Graphical representation of the behavior of a variable. Building graphs and editing them. Types of graphs. The value of graphical presentation of data.

Tasks for the students' self-education (SSE):

1. Get acquainted with one of the examples of an innovative project for collecting and analyzing sociological information.
2. Find an Internet resource with which you can create web questionnaires. Get acquainted with its capabilities and limitations.
3. Select one study on the website of the National Bank of Sociological Data "Kyiv Archive", download the array and methodological documentation. Get acquainted with the topic and variables of the selected study.
4. For the metric variable (age, length of service), perform recoding in order to create combined categories. Check the correctness of the modification.
5. Create an index based on 2-3 primary variables. Explain the logic of your actions. Present the results obtained.
6. Construct different types of one-dimensional data distribution graphs for variables of nominal, ordinal and metric levels. Explain the choice of a particular type of graph.
7. Edit the obtained graphs.

Lecture 4. Analysis of two-dimensional tables.

A two-dimensional table as a tool for studying the relationship between two features: structure and reading rules. Marginal column and marginal row. Situations in which the use of association tables is inappropriate, and ways to get out of them. The concept of statistical association. Analysis of the degree of relationship between two nominal variables.

Chi-square criterion: purpose, description and limitations. Calculation and interpretation of the chi-square coefficient. Number of degrees of freedom. Application of tables to calculate the level of significance of the chi-square. Coefficients built on chi-square.

Refinement method in the analysis of the relationship between features. The method of "spurious relationship". Model with an mediated variable.

Tasks for the students' self-education (SSE):

1. Calculate relative indicators from the absolute indicators given in the association table.
2. Based on tables of two-dimensional distributions, analyze the attitude of the population of Ukraine to government bodies, to various events of public life depending on age, gender, social status, region.

Lecture 5. Analysis of multiple responses. Nonparametric criteria. Comparison measures

Dichotomous method. Definition of sets. Frequency tables for dichotomous sets. Tables of connectivity with dichotomous sets. Categorical method. Definition of sets. Frequency tables for categorical sets. Tables of connectivity with categorical sets. Comparison of dichotomous and categorical methods. Parametric and nonparametric criteria. Characteristics of comparison measures. Means. T-test for one sample, for comparing two independent samples, for paired samples. One-factor analysis of variance.

Tasks for the students' self-education (SSE):

1. Compare dichotomous and categorical methods.
2. Practical application of the analysis of variance.

Lecture 6. Regression analysis. Logistic regression.

Simple linear regression. Regression line, regression coefficient and their construction. Methods of using multiple regression: stepwise and hierarchical. Features of including independent variables in the regression equation. Problems caused by multicollinearity and options for their solution.

Nonlinear regression. Difference between logistic regression and linear regression. Concepts of probability, chance and natural logarithm of chance and their use in the logistic equation. Log likelihood. Distribution of predicted probabilities. Ordinal regression. Probit analysis.

Tasks for the students' self-education (SSE):

1. Graphical representation of regression.
2. Interpretation of logistic regression results.

Lecture 7. Discriminant analysis. Analysis of variance

The concept of discriminant analysis. Methods of introducing predictors into discriminant analysis (standard, hierarchical, statistical). Determining the best discriminators based on average values and standard deviations. The procedure and features of constructing discriminant functions. Referring to the coefficients of the discriminant equation and intragroup correlations between predictors. Assessing the effectiveness of discriminant analysis.

The concept of analysis of variance. The difference between analysis of variance and Student's t-test. Calculation scheme in analysis of variance (calculation of intragroup and intergroup sum of squares). Assessing statistical significance in analysis of variance. Methods of multiple comparisons. Kruskal-Wallis analysis of variance. Characteristics of different comparison measures: rules of application. Analysis of variance using repeated measurements.

Tasks for the students' self-education (SSE):

1. Interpretation of the coefficients of the discriminant equation and intragroup correlations between predictors.
2. Assessing the effectiveness of discriminant analysis.
3. Practical application of the Kruskal-Wallis analysis of variance.

Lecture 8. Factor analysis. Path analysis.

The concept of factor analysis. The purpose of factor analysis. The sequence of stages in factor analysis. Problem formulation. Correlation matrix analysis. The method of principal components. Criteria for determining the number of factors (Kaiser, scree). The rotation method.

The relationship between the dependent, independent variable and the residual. The essence of path analysis. The use of multiple equations. Types of relationship models between three variables. Stages of path analysis (construction of a causal scheme, establishment of the strength of relationships, construction of a path diagram, calculation of path coefficients, quality check, interpretation of results).

Tasks for the students' self-education (SSE):

1. Compare the results of factor analysis using two different methods - principal components and rotation.
2. Using the available data from one of the studies, develop a model of path analysis and conduct its empirical verification.

Lecture 9. Cluster analysis. Data visualization: Interactive modes of working with graphs.

The concept of cluster analysis. The purpose of cluster analysis. Stages of cluster analysis. Selection of variables. The Euclidean distance method. The need for data normalization. Hierarchical and non-hierarchical clustering methods. Methods of presenting results: dendrogram and accumulation diagram. Methods of determining the number of clusters: theoretical considerations, cluster union distance, cluster size. Operations for checking the reliability of clusters: using different methods of measuring distance, using different clustering approaches, splitting data into two parts, discarding individual variables. Transferring statistical results and diagrams to Word. Exporting pivot tables and diagrams as HTML documents.

Column diagrams. Line diagrams. Area, pie, box diagrams. Histograms. Scatter diagrams.

Interactive modes of working with graphs. Correction of interactive graphs. Construction of a diagram based on the data of the pivot table.

Tasks for the students' self-education (SSE):

1. Perform operations to check the reliability of clusters.
2. Construct a diagram based on the data of the pivot table.

Seminar (practical) classes

The main goals of practical classes are to master the lecture material and the results of students' independent work, deepen the acquired knowledge, develop skills in preparing speeches, formulating and defending one's own position, as well as acquiring skills in analyzing sociological information using specialized software.

Practical class 1. Formulating research hypotheses and substantiating the methods of their empirical verification.

1. Based on the proposed set of empirical indicators, formulate different types of hypotheses (descriptive and explanatory, primary and secondary, basic hypotheses and consequential hypotheses, alternative hypotheses). Give a logical (theoretical) justification for the proposed hypotheses.

2. Suggest methods for empirical verification of hypotheses.

3. Indicate the conditions under which hypotheses can be considered fully, partially confirmed, and under which they should be rejected.

Write a report.

Practical exercise 2. Designing different types and types of scales for measuring individual social indicators.

1. Based on the hypotheses put forward (see practical work No. 1), identify 10-15 variables that will be used to test the hypotheses.

2. Suggest 2-3 (if possible) measurement methods for each variable.

3. Divide the variables into three groups depending on the type of measurement scales.

4. For 3 ordinal and 3 metric variables, perform a transformation (modification) into scales of a lower level of measurement. What are the advantages and disadvantages of this type of transformation? When should it be used?
5. Write a report.

Practical exercise 3. Practicing data management skills and their transformation in the PSPP software environment.

1. Creating a new variable using various procedures (Compute, Recode, Count).
2. Recoding variables.
3. Converting variables from one type to another.
4. Basics of working with the syntax editor.
5. Report.

Practical lesson 4. Univariate analysis of sociological data.

1. Construct frequency and percentage distributions of data.
2. Describe the obtained frequencies (valid, missing values, absolute, relative and accumulated frequencies). Construct graphs of univariate distribution.
3. Use statistical characteristics to analyze univariate distributions.
4. Standardize indicators (Save z-values).
5. Write a report.

Practical lesson 5. Data visualization: Graphs.

1. Construct a univariate distribution of the three proposed variables. Make a summary table.
2. Present the properties of variables in graphical form.
3. Construct different types of graphs. Explain the advantages and disadvantages of each of them.
4. Edit the graphs. Describe the resulting graphs.
5. Write a report, paying special attention to the functions and significance of graphical presentation of data.

Practical lesson 6. Analysis of two-dimensional tables

1. From the proposed list of variables, create 5 pairs by combining the independent and dependent variables. Formulate a hypothesis about the relationship between the variables.
2. Create a two-dimensional table to study the relationship between two characteristics. Consider its structure, describe the results obtained. Describe the marginal column and marginal row.
3. Calculate the chi-square coefficient. Conduct statistical examination and meaningful interpretation.

4. Taking into account the number of degrees of freedom, apply the tables to calculate the chi-square significance level. Compare with the results obtained in PSPP.

Practical lesson 7. Analysis of two-dimensional tables

1. Apply coefficients based on chi-square for variables of different levels of measurement.
2. Analyze the relationship between two nominal variables using a two-dimensional table, chi-square, and the main coefficients of pairwise association between nominal variables.
3. Add a third variable to test the hypothesis. Describe the three-dimensional table.
4. Write a report, focusing on the heuristic possibilities of using the chi-square coefficient, the procedure for its application, the calculation algorithm, and the model for interpreting the results.

Practical lesson 8. Analysis of multiple responses

1. Conduct a frequency distribution of data using the multiple dichotomy method.
2. Conduct a frequency distribution of data using the multiple category method.
3. Construct two-dimensional tables using dichotomous sets.
4. Construct two-dimensional tables using categorical sets.
5. Compare dichotomous and categorical methods. Draw conclusions.

Practical lesson 9. Comparison measures

1. Compare means using the (Compare means) command.
2. Apply a T-test for one sample, describe the results.
3. Apply a T-test to compare two independent samples, describe the results.
4. Apply a T-test for paired samples, describe the results, draw conclusions.
5. Conduct a one-way analysis of variance. Interpret the coefficients, draw meaningful conclusions.
6. Write a report, drawing conclusions about the features and limitations of the application of various comparison measures.

Practical lesson 10. Linear regression

1. Calculate the equation of simple linear regression.
2. Construct a regression line.
3. Create a theoretical model of multiple linear regression.
4. Conduct multiple linear regression using PSPP. Calculation of the multiple linear regression equation.
5. Interpretation of multiple linear regression coefficients.
6. Report.

Practical lesson 11. Logistic regression

1. Characteristics of the differences between logistic and linear regression.
2. Use of probability, chance and natural logarithm of chance in the logistic equation.
3. Log likelihood.
4. Distribution of predicted probabilities.
5. Application of ordinal regression.
6. Brief report on the results.

Practical lesson 12. Discriminant analysis

1. Application of the forced inclusion method.
2. Statistics
3. Classification.
4. Analysis of the obtained results. Calculation of the forecast
5. Assessment of the quality of the model.
6. Application of the stepwise method.
7. Report.

Practical lesson 13. Analysis of variance

1. Mastering the calculation scheme in the analysis of variance (calculation of the intragroup and intergroup sum of squares).
2. Assessment of statistical significance in the analysis of variance.
3. Application of multiple comparison methods.
4. Conducting the Kruskal-Wallis analysis of variance.
5. Application of different comparison measures: rules and interpretation of results.
6. Conducting the analysis of variance using repeated measurements.
7. Report.

Practical lesson 14. Factor analysis

1. Problem formulation.
2. Analysis of the correlation matrix.
3. Application of the principal components method.
4. Criteria for determining the number of factors (Kaiser, rocky scree).
5. Application of the rotation method.
6. Writing a report on the results obtained.

Practical lesson 15. Path Analysis

1. Construct a cause-and-effect diagram.
2. Determine the strength of relationships.

3. Construct a path diagram.
4. Calculate path coefficients.
5. Quality check.
6. Interpret results.
7. Write a report.

Practical lesson 16. Cluster Analysis

1. Selection of variables for cluster analysis.
2. Application of the Euclidean distance method. Data normalization.
3. Application of hierarchical and non-hierarchical clustering methods.
4. Use of different methods of presenting results (dendrograms and accumulation diagrams).
5. Application of different methods of determining the number of clusters: (theoretical considerations, cluster union distance, cluster size).
6. Operations for checking the reliability of clusters: use of different methods of measuring distance, use of different clustering approaches, splitting data into two parts, discarding individual variables.
7. Short report.

Practical lesson 17. Data visualization: Interactive modes of working with graphs

1. Conducting one-dimensional and two-dimensional data analysis. Construction of diagrams.
2. Transferring the obtained results and diagrams to Word.
3. Exporting pivot tables and charts as HTML documents.
4. Getting to know interactive modes of working with graphs. Correcting interactive graphs.
5. Building a chart based on pivot table data.
6. Writing a report.

Practical lesson 18. Modular test

6. Students' self-education (SSE):

The students' self-education includes:

- preparation for classroom lessons – 52 hours;
- preparation for calculation and graphic work – 10 hours;
- preparation for modular control work – 4 hours;
- preparation for the exam – 30 hours.

Total – 96 hours.

7. Politics of the course

Attendance at classes and completion of assignments

Attendance at lectures is desirable. During their conduct, theoretical approaches to data analysis within the quantitative paradigm, mathematical and statistical methods of processing, generalizing and analyzing sociological data will be considered. In addition, the lecturer will present numerous and diverse examples of the practical application of each individual method with a demonstration of the algorithm in specialized software. An important component of the lecture will be the interpretation of the results obtained as a result of quantitative analysis, the presentation and justification of substantive conclusions based on statistical examination. Thus, the lecture combines both the theoretical and methodological and methodological and practical components of the analytical process with quantitative sociological information. For students who wish to achieve excellent or good results in their studies, active work in lectures is a necessary prerequisite. However, making up for missed lectures is not required.

Attendance at practical (seminar) classes is mandatory. Missed practical (seminar) classes should be completed independently and worked out during consultations.

The student's rating will be largely formed by the results of his work in practical (seminar) classes. Each missed practical class (regardless of the reasons for the absence) reduces the student's final rating in the discipline. A student who missed practical classes may receive a low rating, which will not allow such a student to take the exam. In this case, the topics from the missed seminar classes must be studied, and the practical tasks completed by the student. The student's knowledge (understanding) of the missed topics (task completion) will be checked during communication with the teacher according to the consultation schedule available on the website of the Department of Sociology, or during a break in the lesson. A student who completes the relevant tasks (answers the questions) will receive the corresponding points for the rating depending on the quality of the answers (task completion).

Forms of work

The lectures cover the content of the main theoretical and methodological principles of conducting a quantitative analysis of various social processes, including social conflicts, consider various case studies and examples of the use of specialized software to solve a specific analytical problem, provide algorithms for applying a particular method, and also offer interpretations of the results obtained. During the lectures, the teacher will provide a holistic overview of the diverse methods of processing and analyzing quantitative social information, focusing on the most popular and effective methods that have gained recognition within the framework of empirical research - both fundamental and applied. Using examples of well-known research programs and projects, students will have the opportunity to get acquainted with the best modern examples of combining theory and empirical research aimed at verifying hypotheses based on conceptual provisions. The topic of the lectures is covered in the working program of the discipline. Questions from students to the teacher are welcome during the lecture. The lecturer may ask questions to individual students or the audience as a whole. Dialogue between students and the teacher is welcome during the lecture.

Students are expected to be able to apply theoretical provisions in practice. In practical classes, problematic aspects of using various methods of computer analysis of sociological information will be discussed. Practical classes are aimed at developing practical skills and abilities in using specialized computer software to process and analyze sociological data. During practical classes, the teacher will summarize and analyze the errors and shortcomings of the work carried out by students on processing and analyzing quantitative sociological information, answer students' questions, and students will work on each other's errors and shortcomings. In classes, various case studies, work with various databases, and active methods of learning activities in pairs and microgroups will be used. During practical and laboratory

classes, students will work with various sources of quantitative sociological information, learn to process and analyze it using specialized computer software. In laboratory classes, students will learn to use the basic methods of descriptive statistics, two-dimensional and multidimensional analysis of sociological information.

During the semester, each student on the chosen topic prepares and writes a modular test paper in the form of their own research project using available data sets. In the process of working on the modular test paper, students will acquire competencies in writing analytical sociological texts based on their own analysis of sociological data, using descriptive statistics, one-dimensional, two-dimensional and basic multidimensional methods.

Rules of conduct in classes

During classes, it is recommended to turn off phones in order to achieve a higher level of attention and focus on educational activities. At the same time, on the recommendation of the teacher, you can use communication tools to search for information on the discipline's Google disk or on the Internet.

During classes - both seminar and lecture, students are expected to be active, participate in discussing issues, complete practical tasks, exercises and tasks, ask questions, contribute to collective discussions, etc.

During answers in practical classes, you should not read from a smartphone, tablet or laptop. It is worth using notes made by the student, summaries of the read educational material and data analysis conducted independently at home or in pairs.

Deadline and retake policy

Each written homework assignment, which the teacher will notify about in advance, must be completed before the start of the corresponding practical (seminar) class. The completed assignment must be submitted the day before the date of the corresponding class (posted on the discipline page in Google Classroom - in the case of distance (online) learning, sent to the teacher's email - in the case of face-to-face learning). Points will be deducted for untimely submitted homework (see University Policy on Academic Integrity).

The policy and principles of academic integrity are defined in Section 3 of the Code of Honor of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". More information: <https://kpi.ua/code>.

Standards of Ethical Behavior

The standards of ethical behavior of students and employees are defined in Section 2 of the Code of Honor of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". More information: <https://kpi.ua/code>.

Policy for the Use of Artificial Intelligence

The policy for the use of artificial intelligence and its principles are regulated by the order "Policy for the Use of Artificial Intelligence for Academic Activities at Igor Sikorsky Kyiv Polytechnic Institute". More information: <https://osvita.kpi.ua/node/1225>.

8. Types of control and rating system for assessing learning outcomes

The student's rating consists of the points for:

- 1) answers in seminar classes
- 2) module test
- 3) calculation and graphic assignment
- 4) exam.

1. Answers in seminar classes.

Weighted score – 5 points for answering one question. The maximum number of points for all seminar classes is $r_{sem} = 5 \text{ points} * 5 = 25 \text{ points}$.

When answering each question, the student receives:

- “excellent”, a complete answer (not less than 90% of the required information) if the student demonstrates deep knowledge of the material when answering, presents it logically and consistently, makes well-founded conclusions, freely operates with specific data, easily and convincingly answers the questions posed – 5 points;
- “good”, a sufficiently complete answer (not less than 75% of the required information), or a complete answer with minor inaccuracies, answers most of the questions posed – 4 points;
- “satisfactory”, incomplete answer (at least 60% of the required information) and significant errors, answers the questions poorly or does not answer at all – 3 points;
- “unsatisfactory”, no work at the seminar – 0-2 points.

2. Modular test.

Weighted score – 5.

Evaluation criteria:

- “excellent”, the topic is fully disclosed (at least 90% of the required information), the work is written independently, meets the established requirements and is submitted on time – 5 points;
- “good”, the topic is not sufficiently disclosed (at least 75% of the required information), there are minor deviations from the established requirements, submitted on time – 4 points;
- “satisfactory”, the topic is poorly disclosed and/or there are significant deviations from the established requirements and/or the work is submitted with a significant delay – 3 points;
- “unsatisfactory”, the work does not meet the requirements – 0-2 points.

3. Calculation and graphic assignment.

Weighted score – 20.

The work is assessed according to the completeness of the task, the correctness of the description of the results, the theoretical and factual saturation, the depth of the analysis.

Evaluation criteria:

- “excellent”, the work fully meets the task, is completed independently, provides a correct and clear description of the results, contains a sufficient theoretical and factual basis, demonstrates a deep analysis, all requirements are met – 18-20 points;
- “good”, the work generally meets the task, minor deviations from the requirements or minor inaccuracies in the description of the results are possible, the level of analysis is sufficient – 15-17 points;
- “satisfactory”, the work contains only part of the necessary elements of the task or is performed superficially, there are significant deviations from the requirements, insufficient theoretical or factual basis, limited or weak analysis – 7-12 points;
- “unsatisfactory”, the work does not meet the established requirements: key parts of the task are missing, the content is incomplete or not independent, the analysis is missing or grossly erroneous – 0-11 points.

4. Exam.

Weighted score – 50.

The exam involves oral answers to 2 questions, the list of which is given in Appendix 2.

Evaluation criteria for each question:

- 40-50 points – the student answers almost all exam questions, demonstrates deep knowledge of the material, presents it logically and consistently, makes well-founded conclusions, freely operates with specific data, expresses his own position on debatable issues, demonstrates signs of theoretical thinking and sociological imagination;
- 30-39 points – the student answers most exam questions, demonstrates a good level of knowledge of the material;
- 20-29 points – the student answers approximately half of the exam questions, demonstrates rather superficial knowledge;

– 0-19 points – the student answers only some exam questions, does not have his own position, allows significant inaccuracies.

Incentive points (no more than 10 points for all types of work):

- for scientific and research activities (participation in conferences, "Days of Science of the Faculty of Social Sciences", student work competitions, publications);
- participation in faculty Olympiads in the academic discipline and all-Ukrainian Olympiads.

Conditions for a positive intermediate certification:

To receive a "pass" from the first intermediate certification, the student must have at least 12 points, to receive a "pass" from the second intermediate certification, the student must have at least 24 points.

Conditions for admission to the exam:

The condition for a student to be admitted to the exam is to receive a preliminary rating of at least 24 points.

Table of correspondence of rating points to grades on the university scale:

Rating of the student (points)	University scale of assessments of the level of acquired competencies (learning outcomes)
100-95	Excellent
94-85	Very good
84-75	Good
74-65	Satisfactory
64-60	Sufficient
Less than 60	Failed

Possible marks in the semester control report:

Not admitted	Failure to meet the conditions for admission to semester control
Fixed	Violation of the principles of academic integrity or moral and ethical norms of behavior
Applicant did not appear	Applicant was admitted, but did not appear for the exam

Review of the rating assessment system (RAS) during the semester

The RAS can be reviewed upon a motivated application from an applicant, a student government body or a student trade union committee, submitted to the head of the providing department. The review procedure is defined in Section 7 of the Regulations on the system of assessment of learning outcomes at Igor Sikorsky Kyiv Polytechnic Institute https://osvita.kpi.ua/sites/default/files/downloads/Pologennia_RSO_2025.pdf

9. Additional information on the discipline (educational component)

Requirements for calculation and graphic work (CGW)

The work is performed using the statistical program for processing social data PSPP. The volume of the work should be 7-10 pages of text, 1.5 spacing, 14 font Times New Roman (excluding tabular and graphic material). Both the CGW itself and its defense during one of the practical classes according to the schedule of performances are evaluated.

1. Formulate a scientific problem that you would like to investigate using secondary analysis of empirical research data. It is recommended to choose a topic that is related to the research issues that are

planned to be conducted as part of the work on the future bachelor's thesis. Review the available datasets of sociological research and choose the one that will allow you to investigate the problem you have chosen. The dataset can be obtained from open archives of sociological information known to you, e.g., the National Bank of Sociological Data "Kyiv Archive". To make the right choice, you should be well-versed in the content and nature of the data of this study. Briefly justify the scientific problem, formulate a research question. It is necessary to describe in detail the research methodology, the data set, and also indicate the data source (website). The tools and the data set should be added to the CGW.

2. Select several variables from the data set (e.g., 8-10), with the help of which you can conduct a small independent secondary study. Some of the variables should correspond to the main topic of your research, and the other part should contain socio-demographic characteristics of the respondents.

3. Come up with several hypotheses (e.g., 3-5) that would reflect the scientific problem that your mini-study will be devoted to studying. The hypotheses and the variables you selected should correspond to each other. Among your hypotheses, there should be at least one causal (explanatory).

4. Think over a model for empirical verification of your hypotheses using methods of both descriptive statistics and statistical analysis.

5. Test your hypotheses using various analysis methods, including one-dimensional and two-dimensional, correlation, regression, variance, factor, cluster, etc. Provide a statistical and meaningful interpretation of the results obtained. Draw conclusions about the validity of your hypotheses. Summarize.

As a result of completing the task, you must submit 3 files:

- 1) text of the assignment (Word file) (7-10 p.),
- 2) presentation (PowerPoint file) (15+ slides),
- 3) PSPP calculation file (output) exported in Word file.

List of questions for the module test:

1. Define the three main paradigms of sociological research.
2. Explain the differences between the quantitative and qualitative paradigms of social data analysis.
3. Describe the view of critical social science on the nature of social reality.
4. Compare the approaches of the three paradigms to the criteria for the truth of knowledge.
5. Explain the relationship between scientific knowledge and common sense in different paradigms.
6. Indicate the role of values in the positivist, interpretive, and critical traditions.
7. Give an example of a sociological problem and define its object and subject in the context of quantitative research.
8. Identify the type of quantitative sociological research by brief description (exploratory, descriptive, analytical).
9. Explain the concept of a 'data matrix' and give an example of its structure.
10. Describe the procedure for calculating the frequency and proportion for a univariate distribution of a variable.
11. Construct a univariate distribution table based on the proposed data set.
12. Explain the problem of "compression" of social information and the role of selecting informative features.
13. Calculate the arithmetic mean for the given values of the variable.
14. Calculate the mode and median for data combined into interval categories.
15. Calculate the variance and coefficient of variation for a given statistical series.
16. Explain the difference between measures of central tendency for quantitative and qualitative features.
17. Explain the principles of forming intervals in distribution series.
18. Give an example of a situation in which it is advisable to use the median rather than the mean.
19. Describe the procedure for selecting a random sample from available data.
20. Formulate a logical expression for filtering observations according to given conditions.

21. Explain the difference between the Compute, Recode, and Count procedures in data processing.
22. Construct a new variable using type conversion (for example, from text to numeric format).
23. Give an example of an index in sociological research and explain its purpose.
24. Explain the essence of logical indices in sociology and give an example of their use.
25. Explain the principle of constructing an index of total ratings (Likert scale).
26. Compile an index for comparing two groups based on the proposed ratings.
27. Describe the value of sorting observations in the process of data analysis.
28. Give an example of a variable that needs to be recoded and justify why.
29. Explain the differences between univariate, bivariate, and multivariate analysis (briefly).
30. Explain the role of a syntax editor in working with data and the benefits of using it.

List of questions submitted for semester control (exam):

1. Quantitative analysis as modeling of social processes.
2. Factors for choosing a specific method of analyzing quantitative data. Univariate and multivariate analysis.
3. Logic and main stages of computer analysis of sociological data. Levels of sociological analysis of empirical data.
4. Procedures for interpreting and operationalizing concepts.
5. The concept of measurement errors, their types.
6. Point and interval estimation.
7. Sociological indices, methods of their calculation, scope of application.
8. Integral indices and their role in quantitative analysis of sociological data.
9. Creating a new variable using various procedures.
10. Converting variables from one type to another.
11. Basics of working with the syntax editor.
12. Data management. Merging and splitting files. Division of cases into groups. Selection and sorting of cases.
13. Arithmetic mean as a statistical indicator. Interpretation of means when analyzing social data.
14. Median and mode as statistical indicators; methods of determination, interpretation.
15. Measures of variation: calculation, interpretation, consideration in the analysis of social processes. Coefficient of variability of categories.
16. Reliable probability and confidence interval: concept and interpretation. Interval estimation for binomial distribution.
17. Using statistical characteristics for the analysis of univariate distributions. Descriptive statistics.
18. Standardization of indicators.
19. Conversion of z-scores into normalized scales.
20. The concordance table as a tool for studying the relationship between two characteristics.
21. Chi-square test: purpose, description, limitations, calculation and interpretation.
22. Chi-squared association coefficients.
23. Refinement method in the analysis of the relationship between features. The method of "false relationship". Model with a mediated variable.
24. The concept of statistical dependence. Types of relationship and strength of relationship.
25. Relationship coefficients for features measured on a nominal scale: calculation and interpretation of values.
26. Relationship coefficients for features measured on an ordinal scale: calculation and interpretation of values.
27. Relationship coefficients for features measured on a metric scale: calculation and interpretation of values.

28. The essence of statistical hypothesis testing procedures.
29. Characteristics of comparison measures. Means. T-test for one sample.
30. T-test for comparing two independent samples.
31. T-test for paired samples.
32. Procedure for analysis of variance.
33. One-way analysis of variance.
34. Methods of multiple comparisons.
35. Kreskel-Wallis analysis of variance.
36. General algorithm for analyzing the relationship between variables. Functional and correlation relationship.
37. Linear and nonlinear relationship. Covariance. The concept of statistical dependence. Strength of relationship.
38. Pearson's pair correlation coefficient: formula, calculation rules, range of values, conditions of application, interpretation, level of significance. Typical errors in the use and interpretation of the correlation coefficient.
39. Rank correlation. Spearman's rank correlation coefficient: rank covariance, calculation rules, range of values, interpretation.
40. Kendall's rank correlation coefficient: calculation and interpretation.
41. Graphical representation of the behavior of a variable. Construction of graphs and their editing. Types of graphs. The value of graphical presentation of data.
42. Comparison of dichotomous and categorical methods of multiple response analysis.
43. Dichotomous method of multiple response analysis: definition of sets, frequency and two-dimensional tables for dichotomous sets.
44. Categorical method of multiple response analysis: definition of sets, frequency and two-dimensional tables for categorical variables.
45. Types of the analysis of variance.
46. The essence of covariance analysis.
47. Interpretation of results in the analysis of variance.
48. Factor analysis and its purpose.
49. Sample size in factor analysis.
50. Determining the number of factors in factor analysis.
51. Regression and prediction.
52. Simple linear regression.
53. Regression line, regression coefficient and their construction. Methods of using multiple regression: stepwise and hierarchical.
54. Nonlinear regression.
55. Sequence of inclusion of independent variables in the equation in stepwise regression.
56. Multicollinearity in multiple regression.
57. Features of discriminant analysis.
58. Discriminant functions.
59. Checking the effectiveness of discriminant analysis.
60. The essence of path analysis.
61. Basic model in path analysis.
62. Path diagram in path analysis.
63. Stages of cluster analysis.
64. Interpretation and profiling of clusters.
65. Ways of presenting results in cluster analysis.
66. Logistic regression.

67. Exporting data.
68. Interactive tables.

Informal distance and online courses

There is a possibility of crediting certificates for distance or online courses on the relevant topic. In particular, it is recommended to study the online course "Data Analysis. Applied tasks of statistical data analysis: relationships, trends, forecasts, classifications". The course is posted on the Coursera educational platform at the link: <https://www.coursera.org/specializations/analiz-dannykh?#courses>. Other available options: English-language course "Survey data collection and analytics" (<https://www.coursera.org/specializations/data-collection>), Ukrainian-language courses "Data visualization" (https://courses.prometheus.org.ua/courses/IRF/DV101/2016_T3/about), "Practical sociology for media professionals" (MEDIASOC101) (https://courses.prometheus.org.ua/courses/course-v1:Prometheus+MEDIASOC101+2020_T3/about), "Sociology and social research: what, how, why?" (SOC101) (https://courses.prometheus.org.ua/courses/IRF/SOC101/2015_T1/about). It is possible to use other massive open online courses (in whole or in part) subject to agreement on their topics and content with the lecturer, as well as other educational resources posted online, e.g., PSPP for beginners (<https://www.youtube.com/watch?v=ZpwZS3XnEZA>). It is possible to use other open online courses subject to agreement on their topics and content with the lecturer.

Inclusive learning is allowed.

The working program of the academic discipline (syllabus) is:

Compiled by: Associate Professor of the Department of Sociology, Candidate of Social Sciences (PhD in Sociology), Associate Professor Korzhov Hennadii Oleksandrovych

Approved by: Department of Sociology (protocol No. 12 dated 05/24/2024)

Approved by: Methodological Commission of the Federal State Educational Service (protocol No. 9 dated 06/26/2024)