

## Subject card

|                                             |                                                                                                                                                                                                     |                                                          |                                         |                                     |                                                                                                                      |            |     |
|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|-----------------------------------------|-------------------------------------|----------------------------------------------------------------------------------------------------------------------|------------|-----|
| Subject name and code                       | Data Preparation, PG_00157248                                                                                                                                                                       |                                                          |                                         |                                     |                                                                                                                      |            |     |
| Field of study                              | Informatics and Econometrics                                                                                                                                                                        |                                                          |                                         |                                     |                                                                                                                      |            |     |
| Date of commencement of studies             | October 2024                                                                                                                                                                                        |                                                          | Academic year of realisation of subject |                                     | 2024/2025                                                                                                            |            |     |
| Education level                             | Master's studies                                                                                                                                                                                    |                                                          | Subject group                           |                                     | Obligatory subject group in the field of study<br>Subject group related to scientific research in the field of study |            |     |
| Mode of study                               | part-time studies                                                                                                                                                                                   |                                                          | Mode of delivery                        |                                     | at the university                                                                                                    |            |     |
| Year of study                               | 1                                                                                                                                                                                                   |                                                          | Language of instruction                 |                                     | Polish                                                                                                               |            |     |
| Semester of study                           | 2                                                                                                                                                                                                   |                                                          | ECTS credits                            |                                     | 3.0                                                                                                                  |            |     |
| Learning profile                            | academic                                                                                                                                                                                            |                                                          | Assessment form                         |                                     |                                                                                                                      |            |     |
| Conducting unit                             | Department of Statistics -> Faculty of Management -> Rector                                                                                                                                         |                                                          |                                         |                                     |                                                                                                                      |            |     |
| Name and surname of lecturer (lecturers)    | Subject supervisor                                                                                                                                                                                  |                                                          | dr Arkadiusz Kozłowski                  |                                     |                                                                                                                      |            |     |
|                                             | Teachers                                                                                                                                                                                            |                                                          | dr hab. Krzysztof Najman                |                                     |                                                                                                                      |            |     |
| Lesson types                                | Lesson type                                                                                                                                                                                         | Lecture                                                  | Tutorial                                | Laboratory                          | Project                                                                                                              | Seminar    | SUM |
|                                             | Number of study hours                                                                                                                                                                               | 0.0                                                      | 15.0                                    | 0.0                                 | 0.0                                                                                                                  | 0.0        | 15  |
|                                             | E-learning hours included: 0.0                                                                                                                                                                      |                                                          |                                         |                                     |                                                                                                                      |            |     |
| Learning activity and number of study hours | Learning activity                                                                                                                                                                                   | Participation in didactic classes included in study plan |                                         | Participation in consultation hours |                                                                                                                      | Self-study | SUM |
|                                             | Number of study hours                                                                                                                                                                               | 15                                                       |                                         | 15.0                                |                                                                                                                      | 45.0       | 75  |
| Subject objectives                          | To familiarize students with the theoretical and practical aspects of the data preparation stage.<br>To familiarize students with the sources of errors occurring at the stage of data preparation. |                                                          |                                         |                                     |                                                                                                                      |            |     |

| Learning outcomes | Course outcome                                                                                                                                                                                                                                                                                                                                                   | Subject outcome                                                                                         | Method of verification            |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|-----------------------------------|
|                   | [liEMU2_W04] The student has an in-depth knowledge of advanced mathematical, statistical, econometric and IT methods that enable the acquisition, processing and analysis of data reflecting the functioning and growth of the national economy and its components, as well as the phenomena and processes occurring in their environment.                       | The student knows and understands the essence of preparing data for analysis.                           | [SW4] test/exam - oral or written |
|                   | [liEMU2_K06] The student knows the need for ethical, sustainable and socially responsible behavior in professional and social life. Initiates and organises activities for the social environment and public interest.                                                                                                                                           | Is able to design the data analysis process, indicating potential problems and methods of solving them. | [SK4] test/exam - oral or written |
|                   | [liEMU2_K02] The student can communicate freely with the public on specialised topics in the field of computer science and econometrics in and outside the workplace, communicate his knowledge and share his skills through various media. Culturally participates in discussions, is not afraid to ask questions and knows how to give constructive criticism. | The student knows and understands the sources of problems in the data preparation process.              | [SK4] test/exam - oral or written |
|                   | [liEMU2_K01] The student understands the need for continuous completion and deepening of acquired knowledge. The student inspires and organizes others' learning processes.                                                                                                                                                                                      | The student is able to critically explain the usefulness of the data preparation process.               | [SK4] test/exam - oral or written |
|                   | [liEMU2_U05] The student is able to make a statistical description of basic and detailed economic categories at an advanced level, formulate and verify hypotheses about their formation.                                                                                                                                                                        | Is able to indicate and critically evaluate the analyzes needed in the data preparation process.        | [SU4] test/exam - oral or written |
|                   | [liEMU2_W05] The student has an in-depth knowledge of socio-economic data sources, their databases and how to create them.                                                                                                                                                                                                                                       | The student knows and understands the essence of preparing data for analysis.                           | [SW4] test/exam - oral or written |
|                   | [liEMU2_U07] The student is able to build advanced formal models of complex economic phenomena and processes, estimate them, carry out their verification and apply them to modeling, forecasting and optimization of resources of economic institutions of varying degrees of complexity, their structure and the course of processes in them.                  | The student knows and understands the sources of problems in the data preparation process.              | [SU4] test/exam - oral or written |
|                   | [liEMU2_K03] The student is able to communicate freely with the public inside and outside the workplace, transfer his knowledge and share his skills through various media.                                                                                                                                                                                      | Is able to design the data analysis process, indicating potential problems and methods of solving them. | [SK4] test/exam - oral or written |

|                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                               |
|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| Subject contents                                               | Quality of statistical data: essence, definition, characteristics of a high-quality data set: relevance to users' needs, timeliness and punctuality, availability and transparency, comparability, consistency, accuracy. Basic sources of errors in statistical data: classification of random and non-random errors, methods of minimizing and correcting them. Stages of the data preparation process for Big Data analyses: registration, coding, control, equalization of moments/time periods, creation of aggregated units/features (micro aggregates), imputation of missing data, transformation of features and cases, selection (reduction) of features and units for analysis. Essence, purpose, definition, basic techniques and their properties: data coding: measurement scales, weakening and strengthening scales, dichotomization, equalizing moments/time periods: chronological r-me averages, creating aggregated units/features (micro-aggregates), imputation of missing data: models of the formation of gaps (MAR, MCAR, MNAR), selected methods of deductive imputation, hot-deck, means, multiple, transformations of features and cases, the impact of transformation procedures on the statistical properties of the data set: standardization, unitarization, zero unitarization, ratio transformations, selection ( reduction) of units: unit sampling procedures, feature selection process: full set of features, generation of feature subsets, evaluation, testing of stop conditions, validation, final set of features, feature selection methods: full, heuristic, random, feature weighting, replacing individual features with artificial ones ( including correlation analysis, Principal Component Analysis). Outliers data analysis: essence, definition, basic techniques for detecting and testing outliers, consequences for data analysis, application examples. Data anonymization: essence, purpose, legal aspects (Polish, EU and international law, GIODO opinions), basic data anonymization techniques. Connecting and integrating data sets: essence, purpose, basic data integration techniques: deterministic and stochastic, data pairing. |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                               |
| Prerequisites and co-requisites                                | Multivariate Analysis, Sampling Methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                               |
| Assessment methods and criteria                                | Subject passing criteria                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Passing threshold                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Percentage of the final grade |
|                                                                | test                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 51.0%                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 100.0%                        |
| Recommended reading                                            | Basic literature                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Barnett V., Lewis T., Outliers in Statistical Data. John Wiley, 1994.<br><br>Balicki A., Migdał Najman K., Najman K., Normalizacja odpowiedzi w badaniach opinii a problem grupowania, Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdanskiego, nr 3, 2008, s. 53-62.<br><br>Migdał Najman K., Najman K., Data Analysis, Machine Learning and Applications, Applying the Kohonen Self-organizing Map Networks to Selecting Variables, Studies in Classification, Data Analysis and Knowledge Organization, Presisach C., Burkhardt H., Schmidt-Thieme L., Decker R., Springer Verlag Berlin Heidelberg, 2008, s. 45-54.<br><br>Walesiak M., Problemy selekcji i ważenia zmiennych w zagadnieniu klasyfikacji, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 1076, Taksonomia, 12, 2005, s. 106-118. |                               |
|                                                                | Supplementary literature                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Kordos J., Jakość danych statystycznych, PWE, 1988.<br><br>Szymkowiak M., Imputacja i kalibracja - nowe możliwości estymacji w badaniach statystycznych z brakami odpowiedzi, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Poznaniu, 2009 nr 116, s. 90-105.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                               |
|                                                                | eResources addresses                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Adresy na platformie eNauczanie:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                               |
| Example issues/<br>example questions/<br>tasks being completed |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                               |
| Work placement                                                 | Not applicable                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                               |

Document generated electronically. Does not require a seal or signature.