

Subject card

Subject name and code	Data Preparation, PG_00157334						
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 Subject group related to scientific research in the field of study		
Mode of study	full-time studies		Mode of delivery		at the university		
Year of study	1		Language of instruction		Polish		
Semester of study	2		ECTS credits		2.0		
Learning profile	academic		Assessment form		credit		
Conducting unit	Department of Statistics -> Faculty of Management -> Rector						
Name and surname of lecturer (lecturers)	Subject supervisor		dr hab. Krzysztof Najman				
	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		20.0	50
Subject objectives	Familiarizing students with theoretical and practical aspects of the data preparation stage for socio-economic analyses..Familiarizing students with sources of errors occurring at the data preparation stage.						

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[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.	Is able to discuss, communicate and share knowledge regarding the preparation of data for statistical analysis.	[SK1] oral statement/conversation/discussion
	[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.	Is able to take into account problems related to data preparation in the process of building advanced models, forecasting and optimizing economic phenomena.	[SU1] oral statement/conversation/discussion [SU5] implementation of a problem task
	[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 communicate and transfer knowledge regarding the data preparation process.	[SK1] oral statement/conversation/discussion [SK5] implementation of a problem task
	[liEMU2_W05] The student has an in-depth knowledge of socio-economic data sources, their databases and how to create them.	Has in-depth knowledge of the sources of socio-economic data and typical problems with their preparation for statistical analysis.	[SW1] oral statement/conversation/discussion [SW5] implementation of a problem task
	[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 take into account data preparation problems in the process of statistical description.	[SU1] oral statement/conversation/discussion [SU5] implementation of a problem task
	[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.	Has in-depth knowledge of advanced mathematical, statistical, econometric and IT methods needed to properly prepare data for analysis.	[SW1] oral statement/conversation/discussion [SW5] implementation of a problem task
	[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 aware of the need for ethical and socially responsible behavior in preparing data for statistical analysis.	[SK1] oral statement/conversation/discussion
	[liEMU2_K01] The student understands the need for continuous completion and deepening of acquired knowledge. The student inspires and organizes others' learning processes.	Understands the need to constantly update knowledge regarding the process of preparing data for statistical analysis.	[SK1] oral statement/conversation/discussion [SK5] implementation of a problem task
	[liEMU2_U06] The student is able to forecast complex economic phenomena and processes at an advanced level and check the properties of the obtained forecasts.	Is able to take into account the specific conditions for forecasting complex economic phenomena in the process of preparing data.	[SU1] oral statement/conversation/discussion [SU5] implementation of a problem task

Subject contents	Quality of statistical data: essence, definition, characteristics of a high-quality data set: relevance to user needs, timeliness and punctuality, accessibility and transparency, comparability, consistency, accuracy. Basic sources of errors in statistical data: classification of random and non-random errors, methods of their minimization and correction.Stages of the data preparation process for Big Data analyses: registration, coding, control, equalization of moments/time periods, creation of aggregated units/characteristics (micro aggregates), imputation of missing data, transformations 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, equalization of moments/time periods: chronological average r-me, creation of aggregated units/features (microaggregates), imputation of missing data: models of missingness formation (MAR, MCAR, MNAR), selected methods of deductive imputation, hot-deck, average, multiple, transformations of features and cases, influence of transformation procedures on statistical properties of a data set: standardization, unitarization, zero unitarization, ratio transformations, selection (reduction) of units: unit sampling procedures, feature selection process: full feature set, generating feature subsets, evaluation, testing stopping conditions, validation, final feature set, feature selection methods: full, heuristic, random, feature weighting, replacing single features with artificial ones (including correlation analysis, Principal Component Analysis). Analysis of anomalous data (outliers): essence, definition, basic techniques for detecting and testing anomalous data, consequences for data analysis, application examples. Data anonymization: essence, purpose, legal aspects (Polish, EU and international law, GIODO opinions), basic data anonymization techniques. Combining and integrating data sets: essence, purpose, basic data integration techniques: deterministic and stochastic, data pairing.		
Prerequisites and co-requisites	Fundamentals of statistics, econometrics, multivariate statistical analysis, sampling survey, programming in R or Python.		
Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	final project	51.0%	100.0%
Recommended reading	Basic literature	Barnett V., Lewis T., Outliers in Statistical Data. John Wiley, 1994. 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. 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. 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	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. Kordos J., Jakość danych statystycznych, PWE, 1988.	
	eResources addresses		
	Example issues/ example questions/ tasks being completed		
Work placement	Not applicable		

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