

**FACULTY OF PURE AND APPLIED MATHEMATICS
SUBJECT CARD**

Name of subject in Polish: **Zaawansowane zastosowania pakietów statystycznych**

Name of subject in English: **Advanced Applications of Statistical Packages**

Main field of study: **Applied Mathematics**

Specialization: **Data Engineering**

Profile: **Academic**

Level and form of studies: **2nd level, full-time**

Kind of subject: **Optional**

Subject code:

Group of courses: **YES**

	Lecture	Classes	Laboratory	Project	Seminar
Number of hours of organized classes in University (ZZU)	30		30		
Number of hours of total student workload (CNPS)	90		60		
Form of crediting	Examination-/crediting with grade*	Examination /crediting with grade*	Examination-/crediting with grade*	Examination /crediting with grade*	Examination /crediting with grade*
For group of courses mark (X) final course	X				
Number of ECTS points	3		2		
including number of ECTS points for practical classes (P)	2		2		
including number of ECTS points corresponding to classes that require direct participation of lecturers and other academics (BU)	1.5		1.5		

*delete as not necessary

PREREQUISITES RELATING TO KNOWLEDGE, SKILLS AND OTHER COMPETENCES

1. Probability theory and mathematical statistics mastered at the level corresponding to the first degree of mathematical studies.
2. Knowledge of theoretical foundations and the ability to fit a linear regression model.
3. Good knowledge of any procedural programming language, basics of S language are recommended.
4. Familiarity with any math or statistics package that has a programming language built-in in addition to a graphical user interface.

SUBJECT OBJECTIVES

- C1 A review of advanced applications of statistical packages.
C2 Improving the skills of using the R statistical package.

SUBJECT EDUCATIONAL EFFECTS

relating to knowledge:

PEU_W01 The student knows the difference between the predictive and confirmatory approach to modeling.

relating to skills:

PEU_U01 The student is able to use the R package utilizing the tidyverse and tidymodels package ecosystem.

PEU_U02 The student is able to analyze data using advanced statistical models.

relating to social competences:

PEU_K01 The student is prepared to acquire new competences and independently analyze data in search of answers to the questions he is interested in.

PROGRAMME CONTENT

Lecture		Number of hours
Lec1	Introduction to the lecture; Determining topics for the 1st and 2nd thematic blocks for lectures 7–14; Discussion about differences between predictive and confirmatory approaches to modeling.	2
Lec2	Introduction to generalized linear models; Exponential family of distributions and its properties.	2
Lec3	Estimation and reasoning about parameters in generalized linear models.	2
Lec4	Comparing generalized linear models; Variable selection methods; Verification of assumptions of a generalized linear model.	2
Lec5	Special cases of the generalized linear model for discrete data – logistic, Poisson and negative binomial regression.	2
Lec6	Regularization methods: Lasso, ridge and Elastic-Net; Zero-inflated Poisson and negative binomial regression.	2
Lec7- Lec10	1st thematic block – one of the topics illustrating advanced applications of statistical packages, e.g.: analysis of interval-censored data, Bayesian models, bootstrap methods, generalized linear mixed models, statistical tools of meta-analysis, spatial statistics, statistics of finite populations.	8
Lec11- Lec14	2nd thematic block – one of the topics illustrating advanced applications of statistical packages from the same pool, but different from selected in the first block.	8
Lec15	Lecture summary.	2
	Total hours	30

Laboratory		Number of hours
Lab1	Reading and processing data in tidyverse.	2
Lab2	Data visualization in ggplot2.	2
Lab3	Fitting models using tidymodels.	2
Lab4- Lab7	Exercises illustrating fitting and statistical inference for generalized linear models in confirmatory and predictive approaches.	8

Lab8- Lab11	Exercises illustrating the topic presented in the first thematic block (lectures 7–10).	8
Lab12- La15	Exercises illustrating the topic presented in the second thematic block (lectures 11–14).	8
	Total hours	30

TEACHING TOOLS USED

N1. Computer presentation and traditional methods.
N2. Computer laboratory.
N3. Student's self work.

EVALUATION OF SUBJECT LEARNING OUTCOMES ACHIEVEMENT

Evaluation (F – forming during semester), P – concluding (at semester end)	Learning outcomes code	Way of evaluating learning outcomes achievement
F1	PEU_U01 PEU_U02 PEU_K01	Three laboratory reports performed in small student groups.
F2	PEU_W01 PEU_U02	A test during lecture.
$P = 5/7 * F1 + 2/7 * F2$		

PRIMARY AND SECONDARY LITERATURE

PRIMARY LITERATURE:

- [1] H. Wickham, M. Çetinkaya-Rundel, G. Grolemund, *R for Data Science*, O'Reilly 2023, wydanie 2.
- [2] M. Kuhn, J. Silge, *Tidy Modeling with R*, O'Reilly 2022, wydanie 1.
- [3] E. Jones, S. Harden, M. J. Crawley, *The R Book*, Wiley 2022, wydanie 3.
- [4] A. Agresti, *Foundations of Linear and Generalized Linear Models*, Wiley 2015, wydanie 1.

SECONDARY LITERATURE:

- [1] E. Gentle, W. K. Härdle, Y. Mori, *Handbook of Computational Statistics*, Springer 2012, wydanie 2.
- [2] K. Bogaerts, A. Komarek, E. Lesaffre, *Survival Analysis with Interval-Censored Data*, Chapman & Hall 2018, wydanie 1.
- [3] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Chapman and Hall 2020, wydanie 2.
- [4] B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall 1993, wydanie 1.
- [5] W. W. Stroup, *Generalized linear mixed models: modern concepts, methods and applications*, Chapman and Hall 2012, wydanie 1.
- [6] G. Schwarzer, *Meta-Analysis with R*, Springer 2015, wydanie 1.

- [7] R. S. Bivand, E. Pebesma, V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, Springer 2008, wydanie 1.
- [8] Y. Tillé, *Sampling and Estimation from Finite Populations*, Wiley 2020, wydanie 1.

SUBJECT SUPERVISOR (NAME AND SURNAME, E-MAIL ADDRESS)

Andrzej Giniewicz (Andrzej.Giniewicz@pwr.edu.pl)