

COURSE OUTLINE

(1) GENERAL

SCHOOL	SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY		
ACADEMIC UNIT	DEPARTMENT OF STATISTICS		
LEVEL OF STUDIES	1st Cycle (UNDERGRADUATE)		
COURSE CODE	6248	SEMESTER	8th
COURSE TITLE	STSP: STATISTICAL ADVANCED MODELLING		
INDEPENDENT TEACHING ACTIVITIES		WEEKLY TEACHING HOURS	CREDITS
Lectures		4	7
Labs			
COURSE TYPE	Elective – Specific Background		
PREREQUISITE COURSES:	Knowledge of topics related to Linear Models and Generalized Linear Models would be useful.		
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	GREEK		
IS THE COURSE OFFERED TO ERASMUS STUDENTS	NO		
COURSE WEBSITE (URL)	https://www.dept.aueb.gr/en/stat/content/stsp-statistical-advanced-modeling-7-ects		

(2) LEARNING OUTCOMES

Learning outcomes
Upon successful completion the students will: • Formulate appropriate statistical models for complex data • Understand and apply statistical models • Use generalized, hierarchical, and latent models • Evaluate model fit quality and predictive performance • Check and interpret model assumptions • Implement advanced models in R • Present and document statistical conclusions
General Competences

(3) SYLLABUS

The course focuses on modern methods of statistical modeling beyond classical linear models. Emphasis is placed on model formulation, parameter estimation, hypothesis testing, model diagnostics and evaluation, as well as the interpretation and
--

communication of results. The aim is to develop critical thinking regarding model assumptions, data structure, and the adequacy of conclusions, with applications to real data using statistical software.

- **Principles of Statistical Modeling:** Statistical models as data-generating mechanisms; likelihood function and identifiability; the logic of modeling and algorithms (model-based vs algorithmic thinking); interpretation versus prediction; goodness of fit versus predictive performance and cross-validation; bias–variance trade-off; overfitting.

- **Generalized Linear Models (GLMs) and Extensions:** Review of generalized linear models (exponential family of distributions and link functions); models for binary data (logistic regression); models for count data (Poisson regression); overdispersion in count models; hurdle and zero-inflated models; models for positive continuous data (log-normal, Gamma, and inverse Gamma models); models for proportions using beta regression; multinomial models; ordinal multinomial regression.

- **Mixed Effects Models (LMM):** Random effects and hierarchical structure; random intercepts (LMM as an extension of linear models with random effects); interpretation of random intercept models; within-group/individual correlation; variance components; prediction and shrinkage; BLUPs (Best Linear Unbiased Predictions); random slope models and covariance structures (random slopes: individual trajectories, heterogeneity in rates of change; fixed vs random effects: interpretation of “average” trend and “individual deviation”; selection of random effects structure).

- **Estimation and Inference in Mixed Effects Models:** Estimation via Maximum Likelihood and Restricted Maximum Likelihood (ML and REML)—what is estimated and when each method is preferred; model testing and comparison (Likelihood Ratio Test with notes on boundary issues, and AIC/BIC criteria); interpretation of coefficients and quantification of uncertainty.

- **GLMM overview and logistic regression with random intercept:** link functions; non-normal distributions; conditional likelihood; logistic regression with random intercepts: interpretation of random heterogeneity in probabilities/odds; computational methods: Laplace approximation / adaptive quadrature; comparison with standard logistic regression.

- **Longitudinal and Repeated Measures Data:** Correlated observations; random intercepts and slopes; marginal versus conditional models; introduction to Generalized Estimating Equations (GEE).

- **Latent Variable Models:** Measurement error models; generalized factor analysis models with latent variables; introduction to item response theory (IRT) models; identifiability issues; psychometric models.

- **Missing Data Handling:** Missing at random (MAR); missing not at random (MNAR); imputation methods and MICE approach.

- **Case Studies:** Integrated data analyses; applications in economic, social, and medical data; model criticism and scientific conclusions.

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY <i>Face-to-face, Distance learning, etc.</i>	Face-to-face	
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY	YES, email, eclass	
TEACHING METHODS	Activity	Semester workload
	Lectures	
	Course Total	175
STUDENT PERFORMANCE EVALUATION	Written examination at the end of the semester Announced in class and at eclass.	

(5) ATTACHED BIBLIOGRAPHY

• Faraway J.J (2006). <i>Extending the Linear Model with R</i>. Chapman & Hall/CRC Taylor & Francis Group. • Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279 • Pinheiro & Bates – <i>Mixed-Effects Models in S and S-PLUS</i> • Moustaki, I., Steele, F., Chen, Y., & Bartholomew, D. (2026). <i>Analysis of Multivariate Social Science Data: Statistical Machine Learning Methods</i> (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003483342 • van Buuren, S. (2018). Flexible Imputation of Missing Data, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429492259 • Pinheiro, José C., and Douglas M. Bates. Mixed-effects models in S and S-PLUS. New York, NY: Springer New York, 2000. • Hambleton, Ronald K., Hariharan Swaminathan, and H. Jane Rogers. Fundamentals of item response theory. Vol. 2. Sage, 1991. • Hagenaars, Jacques A., and Allan L. McCutcheon, eds. Applied latent class analysis. Cambridge University Press, 2002. • Hagenaars, Jacques A., and Allan L. McCutcheon, eds. Applied latent class analysis. Cambridge University Press, 2002.
