# Applied linear models

## Prof. Lucia Paci

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course focuses on normal linear models and generalized linear models employed to explain how a response variable relates to one or more explanatory variables. The aim of the course is to provide students with an advanced knowledge in linear and generalized linear models from both a theoretical and a practical perspective.

By the end of the course, students will be able to:

1. achieve advanced knowledge of regression models used to investigate relationships between response and covariates, under the linear or generalized linear framework (*knowledge and understanding*);
2. provide a suitable description of the data; specify a normal linear regression model, estimate its parameters and test their significance; specify a generalized linear model, by combining a random component with a linear predictor using a proper link function; estimate and test the significance of the parameters of a generalized linear model (*applying knowledge and understanding*);
3. perform variable selection procedures and identify relevant predictors; evaluate the goodness of fit of a model and to detect violations of the model assumptions; interpret the results of an empirical analysis (*making judgements*);
4. describe with an appropriate statistical language the model assumptions, and communicate the results of empirical findings using suitable tools (*communication skills*);
5. independently develop and implement linear and generalized linear models to analyse real data (*learning skills*).

***COURSE CONTENT***

* *Exploratory data analysis and linear regression*: data sources; plotting data; simple linear regression;
* *Multiple linear regression*: model specification and assumptions; parameter estimation (least squares, maximum likelihood and Bayesian inference); statistical inference for regression; inclusion of qualitative regressors;
* *Regression diagnostics*: definitions and properties of residuals; influential observations and leverage points; transformations; multicollinearity; generalized least squares;
* *Variable selection procedures:* effects of model misspecification; stepwise methods; cross-validation;
* *Generalized linear models*: linear predictor and link functions; maximum likelihood estimators; logistic and probit regression; Poisson regression; goodness of fit.

***READING LIST***

Class notes, papers, coding and further material will be posted on the University platform Blackboard.

Useful readings are:

* Faraway, *Linear Models with R*, 2nd edition, Chapman & Hall, 2014 (Chp 1-8).
* J. Fox, *Applied Regression Analysis and Generalized Linear Models*, 3rd edition, Sage, 2016 (Chp 2-15).
* Gelman A., Hill J. and Vehtari A. (2020) *Regression and other stories*, Cambridge University Press (Chp 1-2; 6-13).
* James G., Witten D., Hastie T. and Tibshirani R. (2013) *An Introduction to Statistical Learning with Applications in R*, Springer (Chp 3-6).
* S. Weisberg, *Applied Linear Regression,* 4rd edition, Wiley, 2014.

***TEACHING METHOD***

Frontal lectures, practical classes, discussion of case studies and lab sessions using R (60 hours lessons + 15 hours exercises). Attending the lectures with an active participation and ongoing individual study are strongly recommended.

***ASSESSMENT METHOD AND CRITERIA***

The assessment is based on two parts:

1. Written exam: multiple-choice and open-ended questions, theoretical and lab exercises involving the entire content of the course and aiming at evaluating student’s knowledge according to the learning outcomes of the course;
2. Oral interview: theoretical questions and lab exercises involving the content of the course and aiming at evaluating student’s knowledge according to the learning outcomes of the course. Only students that pass the written exam are admitted to the oral interview.

The final grade is equally based on the two parts.

Assignment (optional): work-home data analysis project implemented in R, carried out individually by the student and with a final oral discussion that replaces the oral interview.

***NOTES AND PREREQUISITES***

Students enrolling in this course are expected to know foundations of algebra, probability, data analysis, and statistical inference, i.e., the topics covered in the courses “Mathematical methods and probability” and “Statistical Inference” taught in the first term. Moreover, students are expected to be familiar with R statistical software.

*office hours*

The office hours is online on Thursday morning upon request via email.