**Statistical Learning**

Dates: 24 – 26 January 2018
 (Starting on Wednesday 13.00 hrs and closing on Friday 13.00hrs.)
 Lecturers: Marjolein Fokkema, Tom Wilderjans, Mark de Rooij
Venue: Leiden University
ECTS: 2

**Requirements**
Please take a laptop with you to the course with R installed, since we will do much practical work. Also take your R-skills! (if you do not have any R skills, please take an R-course first). Further information with respect to preparation for the course will be send to the students in December. Students must be aware that the course requires quite a lot of a priori preparation in terms of watching online lectures and reading chapters of the book.

**Description and Goals**
Statistical learning refers to a vast set of tools for understanding data. Two classes of such tools can be distinguished: “supervised” and “unsupervised”. Supervised statistical learning involves building a statistical model for predicting an output (response, dependent) variable based on one or more input (predictor) variables. There are many areas of psychology where such a predictive question is of interest. For example, finding early markers for Alzheimer’s or other diseases, selection studies for personnel or education, or prediction of treatment outcomes. In unsupervised statistical learning, there are only input variables but no supervising output (dependent) variable; nevertheless we can learn relationships and structures from such data using cluster analysis and methods for dimension reduction. In this course we aim to give the student a firm theoretical basis for understanding and evaluating statistical learning techniques and teach the students skills to apply statistical learning techniques in empirical research.

Upon completion of this course, students will:

  Have knowledge about the difference between explanation and prediction, about the bias-variance trade-off, and about “learners”
  Have a good understanding of several important classes of learning techniques and be able to apply them in R to data: linear regression and classification methods, nonlinear models (splines, GAM), ensemble methods (regression/classification trees, bagging, random forest, boosting), the kernel-trick and unsupervised learning methods (dimension reduction and clustering).
 Know how to evaluate the performance of a statistical learning method by using resampling methods (validation approach, cross-validation, bootstrap) and are able to apply these methods with R to empirical data.

**Literature**
The course is based on the following book: James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. New York: Springer. A free copy is available online at [http://www-bcf.usc.edu/~gareth/ISL/](http://www-bcf.usc.edu/~gareth/ISL/#_blank)

**Registration**
IOPS members can register by sending an email to secretariaat.iops@rug.nl This course is not meant for ReMa students. ReMa students have the possibility to participate in the regular master course at Leiden University. Please contact Tom Wilderjans: t.f.wilderjans@fsw.leidenuniv.nl
The minimum number of participants is 8.

**Venue**Faculty of Social and Behavioural Sciences Leiden University
Pieter de la Court Building, **room 1A37**
Wassenaarseweg 52
2333 AK Leiden
The Netherlands
[Directions/route](https://www.google.com/maps/dir/Wassenaarseweg%2B52%2C%2B2333%2BAK%2BLeiden%2C%2BNederland/%4052.1696888%2C4.4737743%2C16z/data%3D%213m1%214b1%214m8%214m7%211m0%211m5%211m1%211s0x47c5c6e7f9bc7075%3A0xc080a570f9677d14%212m2%211d4.4781517%212d52.1696889#_blank)