Addinsoft
40 rue Damrémont
75018 Paris
Email: training@xlstat.com
Tel: +33(0)170060488 / (646) 412 3348

ADDINSOFT DATA SCIENCE MADE SIMPLE

PREDICTIVE MODELS: LINEAR, LOGISTIC, PLS AND ANCOVA REGRESSIONS

This course is intended for people who wish to implement modeling methods.
Different types of modeling will be covered:

- Multiple linear regression
- ANCOVA-type general linear model
- Logistic regression (binary or multimodal response)
- PLS Regression

The objective of this course is to give participants the methodological know-how to do these analyses:

- Context and objectives
- Conditions for use
- Model quality measurement
- Implementation and interpretation of results, etc.

Duration: 28 hours (4 days)
Location: The course is available in-person at your premises or online via videoconference.
Price (excluding VAT) :

- Inter-company online training : n/a
- Intra-company online training : $7650,00 per training
- Intra-company on site training : $7650,00 per training + trainer's travel expenses

Trainee profiles:

- Technicians – Employees – Executives – Researchers – Students
- Anyone wishing to implement and interpret the results of a predictive statistical analysis method of the regression type or general linear model type (PLS, ANCOVA , etc.).

Required experience :

- A good understanding of basic statistical tools: descriptive statistics, confidence intervals, p-value, alpha risk, hypothesis testing…
- Some knowledge of correlation and linear regression

Training objectives

- Define the context and objectives of the different methods
- Implement and interpret the results of the following methods:
  - Multiple linear regression
  - General linear model (ANCOVA, MANOVA...)
  - PLS regression
  - Logistic regression
- Check the conditions of implementation for the different methods
- Obtain a broad understanding of the mathematical concepts inherent to these methods.

Addinsoft
40 rue Damrémont
75018 Paris
Email: training@xlstat.com
Tel: +33(0)170060488 / (646) 412 3348

ADDINSOFT DATA SCIENCE MADE SIMPLE

- Identify and take into account multicollinearity problems on linear models: variable selection, PLS, etc.
- Measure the goodness of fit and predictive quality (of coefficient estimation) of the model
- Detect and analyze outliers/influential points

Training syllabus

- Reviewing the concept of correlation
  - Defining the correlation coefficient
  - Interpreting the value of the correlation coefficient
  - Sources of confusion: correlation, causation, slope, etc.
  - The different correlation coefficients: Pearson's coefficient - Spearman's coefficient

- Simple linear regression-type modeling
  - Mathematical principles and concepts inherent to simple linear regression
  - Hypothesis testing of the significance of the model
  - Quality of the model
  - Coefficient of determination R^2, adjusted R^2, R^2 Prev
  - Use of the model: Prediction of individual values - Confidence intervals of predictions
  - Graphical treatment of the results
  - Mathematical principles and concepts inherent to multiple linear regression
  - Model inference, variable inference (Fisher statistics)
  - Residual analysis:
    - Residual calculations
    - Physical and statistical significance
    - Homogeneity
    - Distribution, Normality
    - Suspect values
    - Graphical analysis

- Suspect values and influential points
  - Residuals: Studentized residuals, Leverage, Cook distance
  - Model quality: Goodness of fit, coefficient of determination R^2, adjusted R^2 - Prediction quality, estimation error
  - Use of the model: Prediction (forecast) of individual values, Confidence intervals of predictions (forecasts)
  - Graphic illustration of results

- Multiple regression model
  - Significance of the coefficients
  - Hierarchy of the coefficients
  - Problems related to multicollinearity
  - Measures of collinearities: Correlation coefficient, VIF
  - Solving multicollinearity problems

- Analysis of multicollinearity problems through variable selection

- Detection of collinearity: Harmful effects of collinearity between explanatory variables - Detection tools: correlation, VIF, sign consistency
- Proposed solutions: Structured experimentation, Variable selection, PLS

- Treatment of collinearity - Selection of variables:
  - Selection by optimization. R^2, adjusted R^2, AIC and BIC criteria
  - Stepwise selection algorithms: Forward selection, Backward selection, Stepwise regression

- Implementation and Interpretation of results for PLS regressions
  - Context and objectives
  - Introducing the different regression methods on collinear data: PCR, Ridge regression, and PLS
  - Mathematical principles and concepts of PCR and PLS Regression
  - Present the different versions of the PLS regression
  - Implementation and interpretation of results: graphs, model coefficients etc.
  - Choosing the number of components (cross-validation)
  - Components and regression coefficients
  - Quality of fit, quality of prediction
  - Q^2 and R^2 coefficients
  - Importance of explanatory variables for prediction: Standardized coefficients - VIP
  - Selection of variables

- Implementation and Interpretation of results for and ANCOVA (general linear models)
  - Context and objectives
  - The notion of interaction between qualitative and quantitative explanatory variables
  - Combined lines model
  - Model complete
  - Implementation and interpretation of results for all different models
  - Reading and using the model
  - Significance tests of the different terms (Fisher's F)
  - Model cleaning (selection of influential terms and variables)
  - Conditions for using ANCOVA

- Logistic regression-type modeling
  - Context and objectives
  - Differences between linear and logistic regression
  - Definition of the Logit model
  - Implementation and interpretation of results
  - Classification of quantitative explanatory variables
  - Estimation and interpretation of model coefficients
  - Tests of the contribution of a variable (Wald test, likelihood ratio tests)
  - Interpretation of Wald's Chi-square
  - Odds ratios
  - Parallel odds ratios et relative risks

- Analysis of the grading table: success rate, failure rate - true positives, true negatives, false positives, false negatives
- Fitted probabilities and use of the model for prediction purposes
- Conditions for use

Training organization

Arkesys teaching team:

- **Thierry Anthouard** is the head of the Arkesys Group's statistical training program and has always been passionate about the field of statistics. In 1992, he launched the development of the Arkesys Group's statistics training program. His "by example" pedagogical approach allows him to popularize statistics and to make it accessible to all learners. As a consultant supporting of key accounts, he adapts to all types of contexts and learning issues.

- **Jérôme-Philippe Garsi** is a statistical instructor with 13 years of experience in the training field. Since his doctorate on clinical issues, his work is mainly focused on the interest of populations, their health and well-being. At ease with any audience, he makes pedagogy and the simplification of scientific knowledge a priority. To do so, he always takes the greatest care to be clear in his written documents as well as in his oral presentations.

Teaching techniques:

- Learners are welcomed in a dedicated training room
- Theoretical explanations followed by guided practice and autonomous learning
- Additional support documents are posted online after the course
- Supplemental exercises can be loaded onto a USB key

Follow-up: evaluating training participation and results

- Attendance sheets
- Review and evaluation exercises
- Evaluation at the end of the course
- Certificate of completion of the course

Accessibility for disabled people:
People with disabilities who wish to take this course can contact us directly so we can examine together the best way to proceed.

Amount of time the course materials will be available before it starts: 2 weeks

Online training: The link to the virtual classroom will be sent by email the week before the course starts.

Contact: For further information you can contact us by email at: training@xlstat.com or by phone at (646) 412 3348.

Addinsoft
40 rue Damrémont
75018 Paris
Email: training@xlstat.com
Tel: +33(0)170060488 / (646) 412 3348

ADDINSOFT DATA SCIENCE MADE SIMPLE