

# Statistical Business Analysis : Regression & Modelling

Nom : **BANULESCU-RADU** Prénom : **Denisa** Année : **M1** Semestre : **8**

Nature : **CM** Volume horaire : **12 H** ECTS / Coef : **2**

Prérequis	<ul style="list-style-type: none"><li>✓ Cours de statistiques et probabilités</li><li>✓ Cours d'économétrie linéaire</li><li>✓ Cours d'économétrie des variables qualitatives</li><li>✓ Algèbre linéaire</li></ul>
Résumé	<p>Ce cours est une présentation et un rappel des notions et des méthodes statistiques/économétriques qui font l'objet du programme de préparation pour la certification SAS® Certified Statistical Business Analyst Using SAS®9: Regression and Modeling.</p> <p>Nous rappelons la distinction entre statistique descriptive et statistique inférentielle, abordons ensuite les techniques d'estimation des paramètres, notamment les estimateurs ponctuels et les intervalles de confiance. Les tests statistiques usuels portant sur des paramètres inconnus sont revus. La question de la proximité entre deux ou plusieurs populations (en termes de moyenne, variance, etc.) est aussi examinée à l'aide de tests paramétriques. Nous rappelons les principales caractéristiques du modèle de régression linéaire simple/multiple et de la régression logistique. Nous analysons également les mesures du pouvoir prédictif de ces différents modèles.</p>
Objectifs	<ul style="list-style-type: none"><li>- Distinction entre la statistique descriptive et la statistique inférentielle</li><li>- Estimation des paramètres : estimateurs ponctuels et intervalles de confiance</li><li>- Présentation de la procédure de test statistique : test d'hypothèses, niveau de risque/confiance, statistique de test, valeur critique, p-value, règle de décision</li><li>- ANOVA<ul style="list-style-type: none"><li>o Vérification des hypothèses ANOVA</li><li>o Etude de la proximité de deux ou plusieurs populations en utilisant les procédures GLM et TTEST</li></ul></li><li>- Régression linéaire<ul style="list-style-type: none"><li>o Estimation des modèles de régression simple/multiple au moyen des procédures REG, GLM, PLM et interprétation des sorties</li><li>o Comparaison/Sélection des modèles à l'aide des procédures REG et GLMSELECT, et analyse de la validité des modèles de régression linéaire</li></ul></li><li>- Régression Logistique :<ul style="list-style-type: none"><li>o Estimation à l'aide de la procédure LOGISTIC et interprétation des sorties</li></ul></li><li>- Analyse du pouvoir prédictif des modèles</li><li>- Calcul des mesures de performance des modèles<ul style="list-style-type: none"><li>o Réalisation et interprétation des graphs (ROC, LIFT, etc.)</li><li>o Méthodes de sélection du cut-off optimal dans le cadre des modèles de scoring</li></ul></li></ul>
Bibliographie	<ul style="list-style-type: none"><li>✓ Documentation SAS : SAS® Certified Statistical Business Analyst Using SAS®9: Regression and Modeling</li><li>✓ Greene, W., Econometric Analysis, 7th edition, Pearson, 2011</li><li>✓ Hurlin C. et Mignon V., Statistique et Probabilité en Economie Gestion, éditions Dunod, collection Open Book, 2015</li><li>✓ Newbold, P., Carlson, W.L. and B.M. Thorne, Statistics for Business and Economics, Pearson, 8th edition. 2013</li></ul>

---

# PLAN

---

## **PART I - Statistics : Introduction to ANOVA, Regression, and Logistic Regression**

- **Chapter 1 - Course Overview and Review of Concepts**
  - 1.1 Course Overview
  - 1.2 Quick Review of Statistical Concepts
  - 1.3 One-Sample t-Tests 1-39
  - 1.4 Two-Sample t-Tests
- **Chapter 2 - ANOVA and Regression**
  - 2.1 Graphical Analysis
  - 2.2 One-Way ANOVA
  - 2.3 ANOVA Post Hoc Tests
  - 2.4 Pearson Correlation
  - 2.5 Simple Linear Regression
- **Chapter 3 - More Complex Linear Models**
  - 3.1 Two-Way ANOVA and Interactions
  - 3.2 Multiple Regression
- **Chapter 4 - Model Building and Effect Selection**
  - 4.1 Stepwise Selection Using Significance Level
  - 4.2 Information Criterion and Other Selection Options
- **Chapter 5 - Model Post-Fitting for Inference**
  - 5.1 Examining Residuals
  - 5.2 Influential Observations
  - 5.3 Collinearity
- **Chapter 6 - Model Building and Scoring for Prediction**
  - 6.1 Brief Introduction to Predictive Modeling
  - 6.2 Scoring Predictive
- **Chapter 7 - Categorical Data**
  - 7.1 Describing Categorical Data
  - 7.2 Tests of Association
  - 7.3 Introduction to Logistic Regression.
  - 7.4 Logistic Regression with Categorical Predictors
  - 7.5 Stepwise Selection with Interactions and Predictions

## **PART II - Predictive Modeling Using Logistic Regression**

- **Chapter 1 - Predictive Modeling**
  - 1.1 Introduction
  - 1.2 Analytical Challenges
- **Chapter 2 - Fitting the Model**
  - 2.1 The Model
  - 2.2 Adjustments for Oversampling
  - 2.3 Chapter Summary
- **Chapter 3 - Preparing the Input Variables**
  - 3.1 Missing Values
  - 3.2 Categorical Inputs
  - 3.3 Variable Clustering
  - 3.4 Variable Screening
  - 3.5 Subset Selection
- **Chapter 4 - Measuring Classifier Performance**
  - 4.1 Honest Assessment
  - 4.2 Misclassification
  - 4.3 Allocation Rules
  - 4.5 Model Selection Plots