Reforming Master Programmes in Finance in Armenia and Moldova / REFINE

An Erasmus+ Capacity Building Project (2017-2020)

DATA ANALYSIS

Armenian State University of Economics

Teacher: Armen Ktoyan

BASIC INFORMATION

TITLE OF THE COURSE	Data Analysis
TEACHERS	Armen Ktoyan
YEAR OF THE COURSE	2nd Year
SEMESTER OF THE COURSE	3rd
LANGUAGE	Armenian
NUMBER OF ECTS CREDITS	5

LEARNING OUTCOMES

- Implement R in data analysis tasks
- implement T-tests for comparison of means and interpret the results
- analyse within-group and between-group Variations
- conduct ANOVA (using software) and interpret the results
- Implement factor analysis in solution of financial issues,
- understand the purpose and procedures of clustering
- assess the quality of classification of financial objects
- conduct agglomeration clusterization (using software) and interpret the results
- conduct K-means method (using software) and interpret the results,
- analyze cluster profiles,
- conduct logistic regression in R,
- implement ROC-analysis, built ROC-curve and interpret the results,
- Implement logistic regression results in financial risks assessment.

SYLLABUS OF THE COURSE

WEEK	TOPIC
1	Course Outline, Main Problems and Issued of Data Analysis
2	Functions and Simulations in R
3	T – tests and Data Visualization
4	Dispersion Analysis
5	Multi-factor Analysis of Variance
6	Factor analysis
7	Implementation of Factor Analysis
8	Presentation of Assignment 1: Homework
9	Cluster Analysis: Basics and Logics
10	Cluster Analysis: Implication in Finances
11	Discriminant Analysis: Basics and Logics
12	Discriminant Analysis: General Algorithm and Implementation
13	Logistic Regression
14	Assignment 2 -presentation

WEEK 1

Course Outline, Main Problems and Issues of Data Analysis

- Key Terms of Data Analysis
- Basic Stages of Data Analysis
- Application of Data Analysis in Finances
- Basics of R
- Installation of R
- Data Types, Data Structures in R
- Quantitative and Qualitative Data

WEEK 2

Functions and Simulations in R

- Functions in R

- Vectors, Vector Operations

- Functions for Statistical Distributions

- Simulation Programming in R

WEEK 3

T - tests and Data Visualization

- T-tests for Comparison of Means
- Methods for Multiple Comparisons
- Conducting T-tests in R
- Data Visualization in R
- Data Importing/Exporting in R

WEEK 4

Dispersion Analysis

- The Essence of Dispersion Analysis

- Dispersion Analysis Application Conditions

- One-factor ANOVA

- Within-group and Between-group Variations

- Conducting ANOVA in R

WEEK 5

Multi-factor Analysis of Variance

- Two-factor ANOVA

- Multi-factor ANOVA

- Conducting Multi-factor ANOVA in R

WEEK 6

Factor analysis

- The Purpose and Issues of Factor Analysis

- Factor Structure

- Components of Dispersion in Factor Analysis

- Principal Component Method

WEEK 7

Implementation of Factor Analysis

- Comparison of the Results of the Component and Factor Analysis.

- Commentary on the Results of Factor Analysis

- Implementation of Factor Analysis in Finances

- Conducting Factor Analysis in R

Task Assignment 1

WEEK 8

Presentation of Assignment 1: Homework

WEEK 9

Cluster Analysis: Basics and Logics

- The problem of Classification of Financial Objects.

- Cluster Analysis Methods.

- Distance Between Financial Objects

- Quality of Classification

- Procedures of Clustering

WEEK 10

Cluster Analysis: Implication in Finances

- Agglomeration Clusterization
- K-means method
- Two-Stages Clusterization
- Cluster Profiles and their Analysis.
- Conducting Cluster Analysis in R

WEEK 11

Discriminant Analysis: Basics and Logics

- Theoretical Backgrounds of Discriminant Analysis

- Areas and Methods of Application of Discriminant Analysis

- General Problems for Discriminant Analysis

WEEK 12

Discriminant Analysis: General Algorithm and Implementation

- Methods for Evaluating Indicator Information

- Stepping Algorithm

- Conducting Discriminant Analysis in R

WEEK 13

Logistic Regression

- Binary Logistic Regression Model
- Implementation of Logistic Regression in Financial Risks Assessment
- Conducting Logistic Regression in R
- ROC Analysis

WEEK 14

Assignment 2 -presentation

TEACHING METHODOLOGY

The following pedagogical approaches are used:

- Student-Centered Approach to Learning .
- High Tech Approach to Learning.

The following methods and forms of study are used in the course:

- lectures (2 hours a week).
- seminars and laboratory works (2 hours per two weeks: interviews, practical exercises, conducting analysis on statistical package (R, in some cases - SPSS), based on real data sets, discussions of results, assignments and case presentation).
- group work.
- self study.
- current control and grading include: participation in interviews, laboratory work, group work results and case presentations.
- intermediate control includes mid-term exam (individual analysis of a financial data, using a software).
- final exam (a short research, based on implementation of data analysis methods).

LABOUR MARKET RELEVANCE

Some surveys haves been implemented within a group employers from RA market. The employer's list covered the financial institutions, non financial corporations and state governmental bodies.

The employers have outlined the following skills and competences, which have relations to data analysis and that are of great importance.

- select and apply the necessary tools for data analysis and research
- be able to learn new methods of data analysis on their own,
- analyse financial processes as a system, describe the links between parts of a system,
- predict trends of development of processes,
- interpret results in a meaningful way,
- choose and use modern software packages for data analysis,
- be able to search and find the necessary information according to the analytical issue, critically analyse different sources of data,
- identify the basics and the reasons of specific situations,
- apply mathematical, statistical, financial analysis methods and models, identify their strengths and weaknesses, and propose ways to improve them.

ASSESSMENT AND GRADING

Assessment		
	Performance	Grade ratio
Class attendance and participation **Interviews (brief, private, 10-minute interview of student), discussions, activity of laboratory works.**	40-100	20%
Group works (slide presentations)	40-100	20%
Case study	40-100	10%
Mid-term exam	40-100	10%
Final exam (short research)	40-100	40%

ASSESSMENT AND GRADING

Performance and grade			
Percentage	Criteria	Grading numbers	Grade
0-39%		D	Insufficient
40-70%	basic criteria met	C- C C+	Sufficient
71-85%	average performance with some errors	B- B B+	Good
86-95%	above average performance with minor errors	A- A	Very good
96-100%	outstanding performance	A+	Outstanding

REFERENCES

1. Matloff N. The Art of R Programming. A Tour of Statistical Software Design, No Starch Press, Inc. San Francisco (http://diytranscriptomics.com/Reading/files/The%20Art%20of%20R%20Programming.pdf)

2. Agresti, A. (2010). *Analysis of Ordinal Categorical Data.* 2nd ed. New York: Wiley.

3. Johnson, R.A. & D. W. Wichern (2012). *Applied multivariate statistical analysis* Sixth Edition, PHI.

4. Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed). Boston, MA: Pearson.

5. *Методы статистического моделирования для экономистов: учебное пособие* / Г.А. Соколов, Н.А. Чистякова, Рос. экон. ун-им. Г.В. Плеханова. – М.: Изд-во РЭУ им. Г.В. Плеханова, 2011.

COURSE ASSIGNMENT 1

Homework: Analysing Data Sets

- compose a data base from a set of financial-economic indicators of a sample of companies or objects, on Excel basic,

- implement ANOVA and factor analysis,

- comment the results,

- assess the trends and main features of development of a sample.

COURSE ASSIGNMENT 2

Group work: Analysing Data Sets

- The students will be divided into groups (4-6 students in each group). They will be asked to implement analysis on a big data base and according to it results answer the following questions.

- what kind of latent variables can be produced from initial indicators,

- which is the optimal way of classifying the objects,

- how can be classified the new objects,

- how can be analyzed the cluster profiles,

- are their significant differences between means of the groups,

- what kind of financial and managerial decisions can be formed based on the results of data analysis.

THANK FOR ATTENTION

(arm.ktoyan@gmail.com)