Title

Exploratory data analysis with statistical software

Scope

A proposal has been made to strengthen the training offered by the School of Technology of Tomar (ESTT) at the Polytechnic of Tomar (IPT) by providing a microcredential course in exploratory data analysis using statistical software such as Excel and IBM SPSS. The course is intended for use in an operational context involving data processing. This introductory course covers descriptive statistics in a natural way by analysing different data sets, some of which are real, using the provided software.

We propose the creation of a basic course in exploratory data analysis using statistical software, worth 1 ECTS credit and comprising 27 hours in total (17 contact hours and 10 hours of self-directed study).

Context

Exploratory data analysis has evolved due to the ease with which statistical results can now be produced quickly, thanks to increasingly effective computational resources. However, it is important to interpret these results correctly. On the other hand, organisations currently have access to large amounts of variables and data, which they wish to synthesise and transform quickly and easily into knowledge. It is therefore essential to have human resources with statistical skills in order to make efficient, differentiated decisions. In various fields of knowledge: Engineering, biology, health, psychology, agronomy, economics, management, sociology, education/teaching, sports and tourism, to name a few, it is currently essential to use software that assists in the analysis and statistical processing of available data when carrying out different activities. In a developed society and an ever-changing global world, statistical literacy is increasingly important for understanding the world around us.

Target Audience

Technicians and senior managers in public and private organisations, as well as researchers and teachers from a variety of disciplines. Social and Behavioural Sciences (such as Sociology and Human Resources), Business Sciences (namely Economics and Management); Life and Health Sciences (such as Medicine, Pharmacy, Nursing, Psychology, Genetics, Biomedical Engineering, Agriculture and Forestry); Sports Sciences; Engineering (Chemical, Computer Science, Computing and Information Technology and Systems, etc.); Geography, Archaeology, among others.

Student Places

The minimum number of places is ten (10) and the maximum is twenty (20).

Course Structure

The training will cover the following subject content:

1 **Descriptive Statistics**

1.1 Introduction

1.2 Univariate data analysis and treatment

1.2.1 Frequency distributions

1.2.2 Location measures

1.2.3 Dispersion measures

1.2.4 Shape measures

1.2.5 Graphical representation

1.3 Analysis and processing of bivariate data.

1.3.1 Cross-tabulation of variables (contingency tables).

1.3.2 Graphical representation

2 **Correlation and simple linear regression**

2.1 Introduction

2.2 Scatter diagram

2.3 Correlation coefficient

2.4 Simple linear regression model Least squares method

2.5 Coefficient of determination

Methodology

This course is based on a pedagogical model designed for classroom teaching, with the option of live videoconference streaming. Following a theoretical introduction to the statistical content, the software will be used to demonstrate its application, and the resulting outputs will be interpreted and discussed.

Participants should have some knowledge of mathematics at secondary school level and basic computer skills.

Participants with a laptop can install IBM SPSS to use during the course.

Assessment Method

Project involving knowledge application of knowledge in data analysis using software (100%).