Assignment Guidelines

1. Choosing the data set

Choose a data set which is large enough (not less than 40 individuals).

The data set must contain:

1. at least 2 Fixed Factors,
2. at least 3 Covariates (one of which should be your dependent variable).

Important

1. The data set may be downloaded from the internet, provided the website is clearly stated in the assignment.

2. Before committing yourself to the data you have chosen, contact the lecturer and discuss the data set with him.
   Name: Mark Anthony Caruana
   Room: 508
   email address is mark.caruana@um.edu.mt

3. If you wish to send the data set via email to the lecturer make sure that you specify clearly the dependent variable, the covariates and the fixed factors. Furthermore make sure you identify yourself by specifying your name, surname and by giving the course details.

Some Websites which contain data sets:

http://www.statsci.org/data

http://www.statsci.org/datasets.html

http://www.assda.edu.au/

http://www.sci.usq.edu.au/staff/dunn/Datasets/

http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/DataSets

Note:

1. There can be NO group assignments.
2. All students will have different data sets.

3. If a data set you choose has already been assigned to another student you will be asked to choose another data set. (Data sets are allocated to students on a first come first served basis.)

Example of a Data Set:

Covariate	Covariate	Covariate	FF	FF	FF
Stress level	Age	Hours of Work	Edu Level	Gender	Drive
3.67	6	0	0	1	N
4.58	20	0	2	1	N
10.11	34	39	4	2	Y
12.46	59	45	2	2	Y
7.9	23	10	3	1	Y
13.6	34	41	1	2	Y
...	...	...	...	...	...

2 - The Write-Up

The main sections of the write-up must not be less than 25 pages in length. (There is no limit on the maximum number of pages). The font should be Times New Roman, size 12, with 1.5 line spacing and the text should be justified. Furthermore each section and subsection should be numbered.
The title of each chapter should be 18pt bold. The title of each section should be 14pt bold while the title of each subsection should be 12pt bold.

In addition to the above mentioned 25 pages, the assignment must have a front page, table of contents an appendix which contains the data set and a list of references.

The front page must contain the following information:
Title of your assignment (Choose an original title.)
Your name, surname, id card number and course which you are following.
Name and code of this credit (SOR0230 or SOR1231).
Name and surname of lecturer (Mark A. Caruana).

Together with your assignment you are to submit a copy of the **declaration of authorship** which you will find in VLE. Ideally this should appear exactly after the front page.

The write-up should be divided into the following chapters:

1. Introduction

   In this section specify why you chose this data set, from which website (or magazine / thesis / book) you got the data set, why is the data set of interest to you, state clearly which are the fixed factors, covariates and the dependent variable/s.

2. Aims and Objectives

   Here you specify what you would like to do with your data set: Give a general description of the sample being studied, see which fixed factors/ covariates influence the

dependent variable.

3. Descriptive Statistics & Illustrations

   In this section you should present and discuss the measures of location and dispersion of a number of the variables in your data set. Further more you should include at least one bar chart, pie chart, histogram, box plot and scatter diagram to display graphically some of the variables contained within the data set you chose. This section should note be longer than 7 pages. Diagrams must be explained in detail. Diagrams must not be too large! The maximum size of diagrams should be 9cm by 9cm.

4. Parametric / Non parametric Tests

   Before running any parametric / non-parametric text, it is important to mention why this test is being conducted. It is also important to explain not only your results but also the repercussions that these results will have on the analysis of the data set.

   Don't forget to write the H_o and H_1 hypothesis before displaying the results of any statistical test.

   (Ideally this section should have at least three parametric or non-parametric tests.)

   You should also include the chi-squared test and the table of correlations. The latter contains important information which will be used in the Regression section.

5. Regression

   Create a Linear Regression model and state how this model can be improved.

   Create a Multiple Regression Model.

   In each case do not forget to write the equation of the model and to perform an analysis of residuals, Cook's distance and Leverages.

   If possible create a non-linear regression model.

6. General Linear Models

   Create an ANOVA and ANCOVA model or Logistic regression Model (binary or multinomial.

   It is important to carefully remove any interactions and variables which are not

significant.

Do not forget to write the equation of the model and to perform an analysis of residuals, Cook's distance and Leverages.

7. Conclusion

Outline the most important results of your research, point out any possible improvements to your study and clearly indicate any limitations that you encountered.

Important

Please note that you are NOT being given marks to copy and paste tables from SPSS to your assignment. You will be given marks when the SPSS outputs are fully explained. Therefore no marks will be given to students who submit assignments containing SPSS outputs only.

Keep in mind that if a table or diagram is not relevant to the analysis of your data set then it should not be included in the assignment.

3. Submission of Assignment

Students are required to submit a printed and bound version of the assignment to the secretary of the department of statistics and O.R. before the deadline. Spiral binding is most frequently used.

Another copy of the assignment is to be sent via email to the lecturer in .pdf, .doc or .docx format before the deadline.

4. Assignment Deadline

THE ASSIGNMENT DEADLINE IS AT NOON ON THURSDAY 30TH JUNE 2016 . LATE SUBMISSIONS WILL NOT BE ACCEPTED.