

Assignment Guidelines: David Suda

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 Categorical variables, (2) at least 3 Covariates (one of which should be your dependent variable).
2. The data set may be downloaded from the internet, provided the website is clearly stated in the assignment.
3. Before committing yourself to the data you have chosen, contact the lecturer Dr. David Suda by email (david.suda@um.edu.mt). A reply will be given regarding whether the data is good for use, and whether or not it is being used by someone else. If a negative reply is given regarding the dataset, for whatever reason, a new dataset will need to be found.
4. When sending the dataset, identify yourselves by group and name.
5. 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>
You are not restricted to obtain data from just these websites.
6. 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 the study unit.
 - Name and surname of lecturer (Dr. David Suda).
 - A copy of the declaration of authorship which you will find in VLE (Ideally this should appear exactly after the front page.)
7. The write-up should be divided into the following chapters:
 - 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 dataset of interest to you, state clearly and define the fixed factors, covariates and the dependent variable/s.
 - 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.
 - 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. Furthermore, 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. Diagrams must be explained in detail. Diagrams must not be unnecessarily large. The maximum size of diagrams should be 9cm by 9cm.

- Parametric / Non parametric Tests: Before running any parametric / non-parametric test, 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_0 and H_1 hypothesis before displaying the results of any statistical test. (You should not attempt to perform all the tests found in the lecture notes. You must only apply those tests which can give you meaningful results.) 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.
 - Modelling: This can include – regression, general linear models and GLMs (logistic regression). You must check that all the assumptions related to the model are true. 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.
 - 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.
8. Important: Please note that you will not be given marks to copy and paste tables from SPSS or R to your assignment. You will be given marks when the statistical outputs are fully explained. Therefore, students who submit assignments containing statistical outputs without any explanation will not be given a pass mark. 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.
 9. 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 format before the deadline.
 10. Assignment Deadline: You will receive the assignment deadline via email from the departmental secretary, Ms. Stephanie Dimech.