7CCMMS61T Statistics for Data Analysis Coursework

The deadline to submit your assignment on the KEATS page of the module is on **16th December 2019 at 5pm**.

Please submit a report which answers the questions below, including an appropriate description of the methodology you used (with code snippets when appropriate) and the graphics you produced.

Question 1

The file `protein.csv` contains data from several European countries in the 1980s on consumption of different categories of food.

Exploratory data Analysis:

(a) For each variable, calculate appropriate summary statistics to show the level and spread of the data (one statistic for each is enough).

   Marks: 3

(b) For each variable, plot the data in a suitable way to illustrate the level and the spread.

   Marks: 3

(c) Calculate a summary statistic to show the association of the consumption of fruit and vegetables with each of the other food categories.

   Marks: 3

(d) Show a plot illustrating the association of the consumption of fruit and vegetables with each of the other food categories.

Marks: 3

Inference:

(e) Provide confidence intervals at level 95% for the mean consumption of each category of food.

Marks: 3

(f) Carry out the appropriate test of hypothesis to check if the average consumption of starch is larger than the average consumption of nuts. Also check if the assumptions behind this test are reasonable in this case.

Marks: 5

Question 2

The file `DartPoints.csv` contains data on 91 Archaic dart points recovered during surface surveys at Fort Hood, Texas. These data have been extracted from the R package `archdata`. The dataset contains the following variables:

- `Name`. Dart point type: Darl, Ensor, Pedernales, Travis, Wells

- `Length`. Maximum Length (mm)

- `Width`. Maximum Width (mm)

- `Thickness`. Maxmimum Thickness (mm)

- `B.Width`. Basal width (mm)

- `J.Width`. Juncture width (mm)

- `H.Length`. Haft element length (mm)

- `Weight`. Weight (gm)

- `Blade.Sh` Blade shape: E - Excurvate, I - Incurvate, R - Recurvate, S - Straight.

- `Base.Sh` Base shape: E - Excurvate, I - Incurvate, R - Recurvate, S - Straight.

- `Should.Sh` Shoulder shape: E - Excurvate, I - Incurvate, S - Straight, X - None.

- `Should.Or` Shoulder orientation: B - Barbed, H - Horizontal, T - Tapered, X - None.

- `Haft.Sh` Shape lateral haft element A - Angular, E - Excurvate, I - Incurvate, R - Recurvate, S - Straight.

- `Haft.Or` Orientation lateral haft element: C - Concave, E - Expanding, P - Parallel, T - Contracting, V - Convex.

and `NA` denote missing data (You can ignore them for the purpose of the exercise).

Exploratory data Analysis:

(a) State the scaling of each of the above variables.

   Marks: 3

(b) First consider the variable `Length`. Represent graphically the relationship between `Length` and the other variables and describe any interesting patterns.

   Marks: 5

(c) For the variables which seems to be associated with `Length` calculate a summary statistic which will describe the strength of the association, if possible.

   Marks: 3

(d) Compute and represent graphically the relative frequency distribution of `Weight` conditionally to the various types of blade shape.

   Marks: 3

Multiple linear regression:

(e) Select an appropriate multiple regression model, which can be used to predict the weight of the dart, using some or all (after appropriate selection) of the variables listed above as explanatory variables (with the exclusion of the weight itself, of course).

Marks: 10

(f) Check and describe the fit of your model using whatever graphical or numerical methods seem appropriate.

Marks: 3

(g) Interpret the fitted model in practical terms. What does it tell you about predicting the dart weight?

Marks: 3

(d) Predict the expected dart weight for a dart point of type Travis, with maximum length 70 mm, H.Length 60mm, Thickness 50 mm, B.Width 50 mm, J.Width 50 mm, Width 60 mm and with both blade shape and base shape recurvate, straight shoulder shape, barbed shoulder orientation, excurvate shape for the lateral haft element and parallel orientation of the lateral haft element. Give a 95% confidence interval for this expected weight. Is there any reason to be cautious about your estimate?

Marks: 5