

Coursework Assignment - 5CCM242A/6CCM242B

This document details the content of the Coursework Assignment that you will need to submit on the KEATS page of the module by **4pm on April 5th, 2022**. Please submit up to **2 (two) A4 pages (not including the department coversheet)** in **.pdf/.docx or equivalent formats with minimum font size of 12pt**. This must include figures and R code whenever required. If your submission includes more than two pages, only the first two pages of your submission will be marked.

Your submission must be anonymous. Your submission will be marked on a scale up to 100 with the marking scheme below and it will contribute 20% of the final mark for the module.

The datasets to be used in the exercises below can be found on the KEATS page of the module in the Coursework section.

Exercise 1

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

- **Name.** Dart point type: Darl, Ensor, Pedernales, Travis, Wells
 - **Length.** Maximum Length (mm)
 - **Width.** Maximum Width (mm)
 - **Thickness.** Maximum Thickness (mm)
 - **Weight.** Weight (g).
- (a) Using an appropriate model selection strategy and variables transformation if necessary, choose and fit the the best model for the variable Weight. This includes checking the model assumptions and fixing obvious issues. Comment on potential issues that you were not able to fix (if any). **[40 Marks]**
- (b) Obtain a 90% prediction interval for a new observation of Weight for a Dart of type Pedernales with Length = 50, Width = 20 and Thickness = 6. **[15 Marks]**

Exercise 2

Download from the KEATS page of the module and import in R the dataset `wheat.txt`. This dataset is a redacted version of the `seeds` dataset available at <https://archive.ics.uci.edu/ml/datasets/seeds>. The `wheat.txt` dataset contains the measurements of the variables `area`, `perimeters`, `compactness` and an `asymmetry` coefficient for seeds from 2 species of wheat (Kama and Rosa, as denoted by the `species` variable).

- (a) Select a generalised linear model to predict the species of the wheat seed based on the measurements of area, perimeter, compactness and asymmetry. **[30 Marks]**
- (b) What is the probability that a seed with area = 13, perimeter=10, compactness=0.75, asymmetry=2 is of species Rosa? Is there a reason why we should not trust this prediction? **[15 Marks]**