**MATH10282 Introduction to Statistics**
**Semester 2, 2021/2022**
**Coursework assignment using R**

The deadline for submitting this coursework is 11:00am on Friday 20 May 2022. Please submit your work to Blackboard.

This coursework comprises 10% of the overall marks for the module.

## Instructions

(a) You can include code and numerical results from R by copying and pasting into your document. Comments and discussion of the results should be added as required. You can save any plots created in R, for example as a PDF, and import these into your final report. If you wish to include handwritten section (e.g. with mathematical notation), you may do so by including a scan or photo of the handwritten part in your report.

Include in your report all R commands used to generate results.

(b) Please include your name and ID in your report. If you have any queries or problems, please contact me as soon as possible.

(c) Your report should not be longer than 5 printed pages. Longer reports will be penalised.

Sugar is a major ingredient in many breakfast cereals. The data below give the sugar content as a percentage of weight for 19 brands of cereal for children and 19 brands of cereal for adults.

| Children Cereal | Adult Cereal |
|:---:|:---:|
| 40.3 | 20 |
| 55 | 30.2 |
| 45.7 | 2.2 |
| 43.3 | 7.5 |
| 50.3 | 4.4 |
| 45.9 | 22.2 |
| 53.5 | 16.6 |
| 43 | 14.5 |
| 44.2 | 21.4 |
| 44 | 3.3 |
| 47.4 | 6.6 |
| 44 | 7.8 |
| 33.6 | 10.6 |
| 55.1 | 16.2 |
| 48.8 | 14.5 |
| 50.4 | 4.1 |
| 37.8 | 15.8 |
| 60.3 | 4.1 |
| 46.6 | 2.4 |

(a) Compute the mean, median, standard deviation, maximum and minimum of the data on sugar content of the cereal for children.

[1]

(b) Compute the mean, median, standard deviation, maximum and minimum of the data on sugar content of the cereal for adults. [1]

(c) Compare the results in parts (a) and (b) and comment. [1]

(d) Draw boxplots of the data on sugar content of the cereal for children and the data on sugar content of the cereal for adults. Plot them side by side and comment on how they compare. [1]

(e) Fit a normal distribution to the data on sugar content of the cereal for children. Comment on the adequacy of the fit of the normal distribution. You may use the Kolmogorov Smirnov test by using the R command

ks.test (x, "pnorm", mean, sd)

where x contains the data, mean is the sample mean of the data, and sd is the sample standard deviation of the data. If the p-value returned by the command is above 0.05 you may consider the distribution as providing an adequate fit. [1]

(f) Fit a normal distribution to the data on sugar content of the cereal for adults and comment on the adequacy of its fit. [1]

(g) Construct a 95% confidence interval for the difference between the population mean of the data on sugar content of the cereal for children and the population

mean of the data on sugar content of the cereal for adults. You may assume that the sample standard deviations computed in parts (a) and (b) are the true population standard deviations. Comment on the confidence interval. Does the confidence interval suggest a significant difference between the sugar contents? [2]

(h) Let $\mu_X$ = the population mean of the data on sugar content of the cereal for children and $\mu_Y$ = the population mean of the data on sugar content of the cereal for adults. Test the hypothesis $H_0 : \mu_X = \mu_Y$ versus $H_1 : \mu_X > \mu_Y$ at the five percent level of significance. You may assume that the sample standard deviations computed in parts (a) and (b) are the true population standard deviations. Report conclusions of the test. [2]

[Total 10 marks]