ETC5242 Assignment 1

Group 42

Khuyen Nguyen: kngu0086@student.monash.edu, Evan Ginting: , Rachitha Werake:

REMEMBER TO ADD STUDENT E-MAIL ABOVE^ (remove this line once done)

```r
library(MASS)
library(tidyverse)
grades <- read_csv("data/GradesData.csv")
```

Task 1 - fit a distribution [30 marks]

The lecturer wants to examine "genuine" (non-zero) attempts only. Check and modify the data (if necessary) to ensure that only genuine attempts are analysed. [1 mark]

```r
grades <- grades %>%
  filter(Total != 0)
```

Traditionally Normal distributions have been used to model grade distributions. Plot your data and explain whether you think that this is good idea in this instance. [3 marks]

```r
dist <- grades %>%
  ggplot(aes(x = Total, y = after_stat(density))) +
  geom_histogram(colour = "blue",
                 fill = "blue",
                 alpha = 0.2) +
  geom_density(colour = "blue",
               fill = "blue",
               alpha = 0.2) +
  ggtitle("Distribution of Total Grades") +
  xlab("Total Score") +
  ylab("Density") +
  theme_bw()

dist
```

Distribution of Total Grades
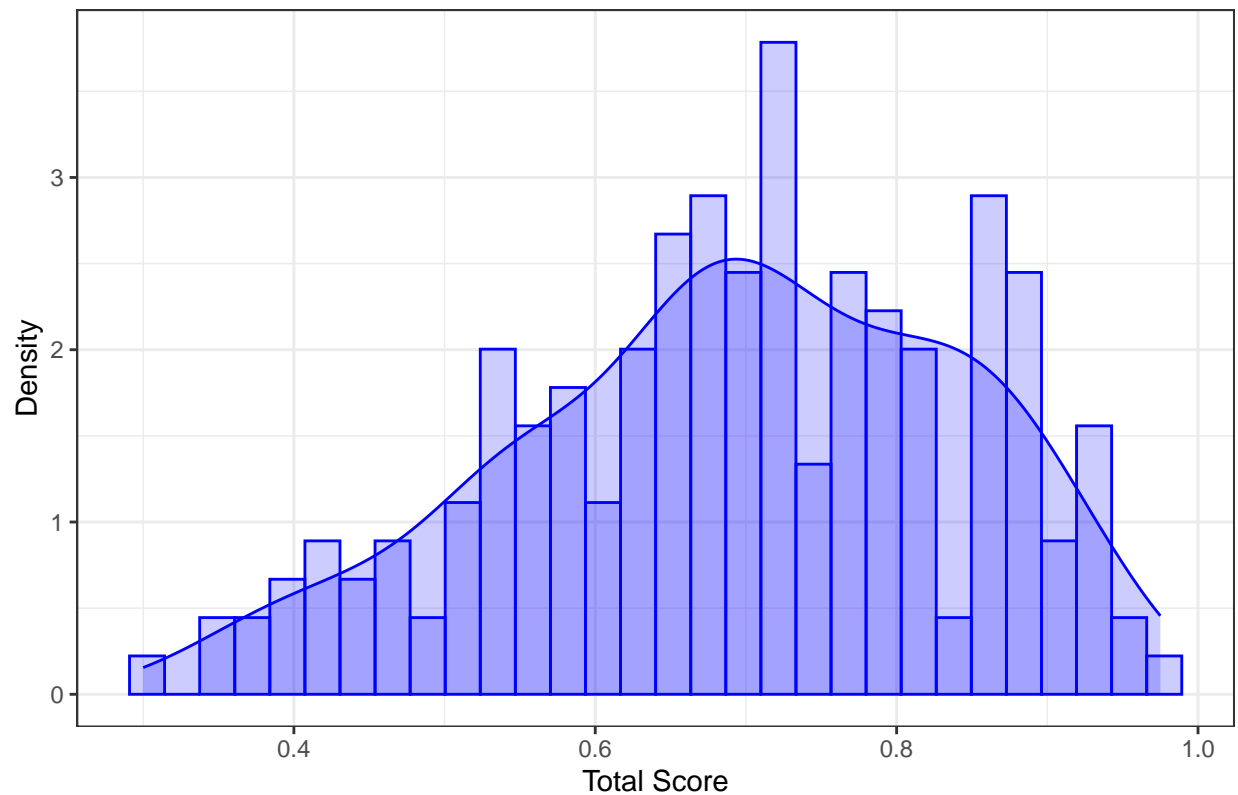
A normal distribution may be a good idea in this instance, as the visualisation does resemble a bell-shape. It is generally symmetric and unimodal.

An alternative is the beta distribution. Explain why this may be a valid alternative in this case. [2 marks]

The beta distribution could work given we give the correct parameter estimates of α and β. We can see the distribution does tend slightly to the right and perhaps the beta distribution can assist here but we would need to fit the data using Maximum Likelihood.

Use Maximum Likelihood to fit both a Normal and Beta distribution to the grades. Use a Bootstrap QQplot to assess the fit of each. Which distribution do you recommend? Be sure to briefly explain your reasoning. [8 marks]

```
fit_norm <- fitdistr(grades$Total, "normal")
params_norm <- fit_norm$estimate
```

```
fit_beta <- fitdistr(grades$Total, "beta", start = list(shape1 = 1, shape2 = 1))
params_beta <- fit_beta$estimate
```

The lecturer is interested in trying the Beta distribution. Use your MLE's to report the mean and the median of the grade distribution (recall that the mean is a function of the shape and rate parameters). Be sure to interpret these values. [4 marks]

Plot and interpret a 99% parametric bootstrap of the mean of the beta distribution (HINT: set warning=FALSE in your code chunk). Did the average quiz mark match the lecturer's goal? [4 marks]

Using the MLE's, what is the estimated proportion of students within 15% of the average? According to the lecturer's benchmark, what proportion would have failed? How many would get HD's? [4 marks]

Overall, do you think that the quiz achieved the lecturer's aims? [4 marks]

Task 2 - Are Postgrad students better? [30 marks]

Task 3 - Bayesian Analysis [33 marks]