STA302/1001H1S - Method of Data Analysis
Assignment 1

Due Fibruary 7th, 23:59

Instructions: This is individual assignment. It is worth 100 points. Please use Rmarkdown to write your solutions and submit your solutions with relevant R code included as a pdf file via **Crowdmark**.

Question

The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year can be predicted from their ACT test score. The data stored in the data file GPA.txt can be found on the course webside. You can read the data file using the following R code:

```r
Dat=read.table("GPA.txt", header=F)
# change column names
names(Dat)<-c("GPA","ACT")
# print the six first observations
head(Dat)
```

```
##      GPA ACT
## 1 3.897  21
## 2 3.885  14
## 3 3.778  28
## 4 2.540  22
## 5 3.028  21
## 6 3.865  31
```

Answer the following questions, showing all your work where appropriate.

(a) Compute the mean and variance of both ACT score and GPA.

(b) Compute the correlation between ACT score and GPA. Comment on the strength and direction of the linear relationship between the variables.

(c) Fit a simple linear regression using ACT score as the explanatory variable, and GPA as the response variable.

(d) What is the estimated intercept and slope of the regression line?

(e) Write in words the interpretation of the slope.

(f) What is the standard deviation around the regression line, i.e. estimate σ^2 ?

(g) Use a t-test to determine whether or not there is a linear relationship between ACT score and GPA.

(h) Construct a scatter plot of the length of the tibia against the length of the humerus. Superimpose the regression line.

(i) Compute the ANOVA table corresponding to the model.

(j) From the table determine the mean square error(MSE).

(k) Use the ANOVA F-test to determine whether or not there is a linear relationship between ACT score and GPA.

(l) How do the results in (k) compare to those in (g)?

(m) What proportion of the variation in GPA is explained by the regression model?

(n) Construct a 95% confidence interval for the estimated mean GPA of students whose ACT test score was 28.

(o) Construct a 95% prediction interval for a particular student whose ACT test score was 20.