Assignment 1

James La Fontaine

26/08/2021

## Question 1

### 1a)

quiz = read.delim("quiz.txt", header = FALSE, sep = "")  
responses = quiz[,1]  
summary(responses)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.00 40.00 60.00 67.13 85.00 243.00

sd(responses)

## [1] 40.54038

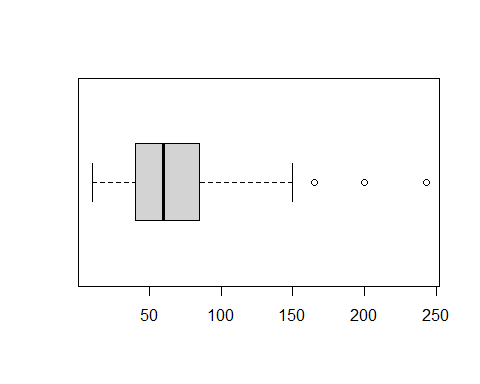
IQR(responses)

## [1] 45

quantile(responses, type = 7)

## 0% 25% 50% 75% 100%   
## 10 40 60 85 243

boxplot(responses, horizontal = TRUE)



This distribution is asymmetrical and positively skewed with the centre lying roughly below the median 60. This distribution has a relatively large spread with a range of 233 due to an outlier, and even still has a range of approximately 140 when ignoring outliers. The distribution additionally has a relatively loose IQR of 45 and relatively large standard deviation of 40.54. The responses appear to spread out further and further as they become larger.

### 1b)

# starting values for the parameters determined via the moment of methods  
alpha.hat = mean(responses)^2/var(responses)  
theta.hat = var(responses)/mean(responses)  
  
gamma.fit = fitdistr(responses, densfun = "gamma", start = list(shape = alpha.hat, scale = theta.hat))   
  
shape.hat = gamma.fit$estimate[[1]]  
scale.hat = gamma.fit$estimate[[2]]  
shape.hat

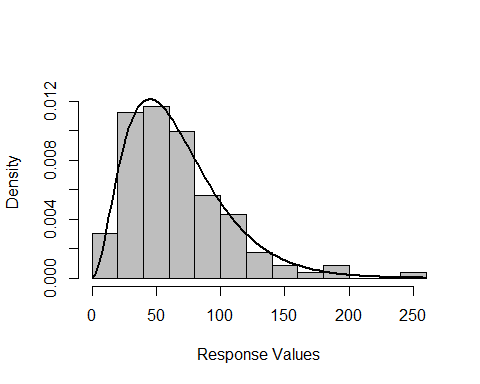
## [1] 3.040857

scale.hat

## [1] 22.07282

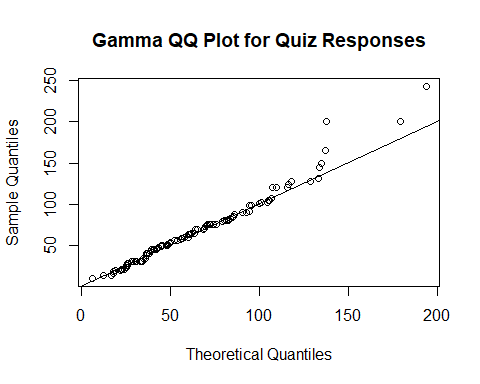
### 1c)

hist(responses, freq = FALSE, col = "gray", main = NULL, xlab = "Response Values", nclass = 10, ylim = c(0,0.013))  
curve(dgamma(x, shape = shape.hat, scale = scale.hat), lwd = 2, add = TRUE)



### 1d)

gamma.sample = rgamma(116, shape = shape.hat, scale = scale.hat)  
  
  
qqplot(gamma.sample, responses, main = "Gamma QQ Plot for Quiz Responses", xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")  
  
abline(a=0, b=1)



The model fits the data quite well and thus this QQ Plot demonstrates that the gamma distribution with the parameters estimated earlier is a good approximation for the distribution of the data.

## Question 4

Let Damjan’s average of the sample minimum and maximum be Estimator 1. Let Julia’s sample median be Estimator 2. Let Martina’s sample mean be Estimator 3.

numberofsimulations = 20000  
N = numberofsimulations  
estimator1 = 1:N  
estimator2 = 1:N  
estimator3 = 1:N  
for (i in 1:N) {  
 normal.sample = rnorm(10)  
 estimator1[i] = (max(normal.sample) + min(normal.sample)) / 2  
 estimator2[i] = median(normal.sample)  
 estimator3[i] = mean(normal.sample)  
}  
  
# subtract 0 as this is the true mean of the standard normal distribution  
bias.estimator1 = mean(estimator1) - 0  
bias.estimator2 = mean(estimator2) - 0  
bias.estimator3 = mean(estimator3) - 0  
  
variance.estimator1 = var(estimator1)  
variance.estimator2 = var(estimator2)  
variance.estimator3 = var(estimator3)

## Bias of Estimator 1: -0.003434098

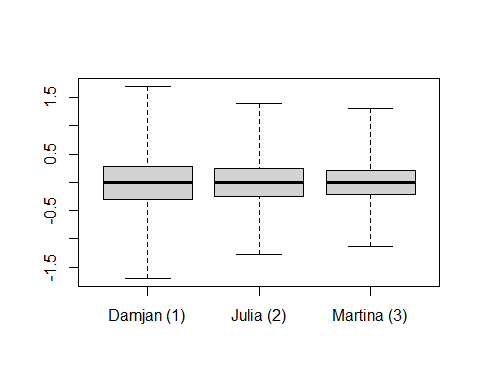
## Bias of Estimator 2: -0.002656974

## Bias of Estimator 3: -0.003047521

## Variance of Estimator 1: 0.1855932

## Variance of Estimator 2: 0.1414125

## Variance of Estimator 3: 0.1003291



Note that the boxplot whiskers have been extended out to the maximums and minimums of each data set for improved visual clarity.