Assignment 1

Graded

Student

James La Fontaine

Total Points

19 / 20 pts

Question 1

Question 1 **6** / 6 pts

1.1 1(a) **2** / 2 pts

Q: Give basic descriptive statistics for these data and produce a box plot. Briefly comment on the center, spread and shape of the distribution.

✔ **+ 2 pts** Perfect: central stat - e.g. summary(x)/mean/median AND spread stat - e.g. sd/IQR/var
mean=67.13/median=60
sd=40.54
AND boxplot
AND comment on centre, shape and spread

+ 1.5 pts Statistics, boxplot and comment provided, but minor mistake/s e.g. missing spread statistic (sd/IQR/var), slightly wrong numbers, no output from functions, comment incorrect (e.g. stating left/negative skew), etc.

+ 1 pt One or two of the following elements are present (but at least one is missing): descriptive statistics, boxplot, comment

+ 0 pts Wrong/no attempt

1.2 1(b) **1** / 1 pt

Q: Assuming a gamma distribution for these data, compute maximum likelihood esti-mates for the parameters.

✔ **+ 1 pt** Perfect e.g.

```
gammafit <- fitdistr(quiz, densfun = "gamma")
gammafit
```

```
##        shape          rate
##   3.041098377   0.045302228
##  (0.379119742) (0.006138772)
```

(alternatively, k=alpha=3.0411, theta=1/beta=22)
must include both working code (e.g. fitdistr or optim) and output numbers for full marks (or provide by-hand justification/working) .

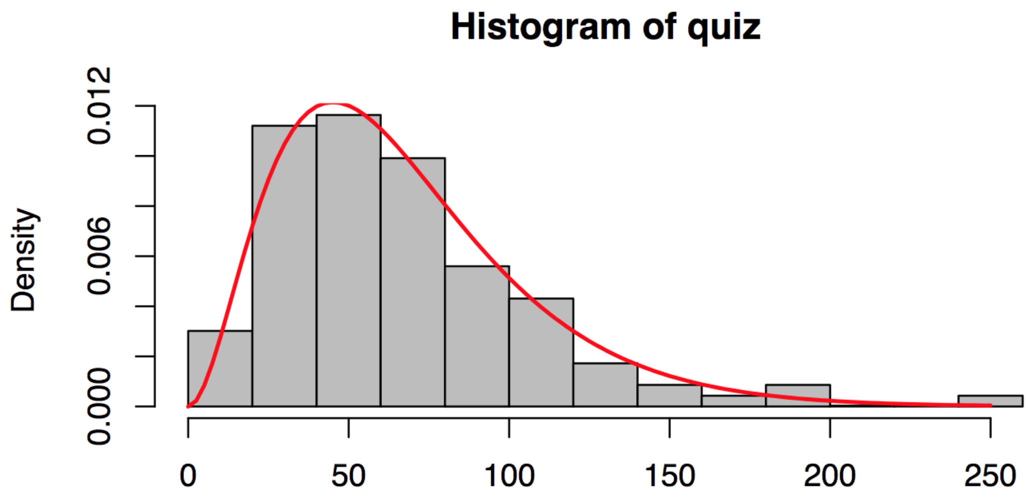
+ 0.5 pts Reasonable attempt but incomplete/some error:
e.g. correct values/output but no fitdistr/optim code, no output from code (or values incorrect), correct values with no explanation, incomplete/incorrect attempt at by-hand solution, fit gumbel instead of gamma, etc.

+ 0 pts Wrong/no attempt

1.3 **1(c)** **1** / 1 pt

Q: Draw a density histogram and superimpose a pdf for a gamma distribution using the estimated parameters.

✔ **+ 1 pt** Perfect: Correct graph (hist + density curve), different bins ok within reason (5 bins is too few)



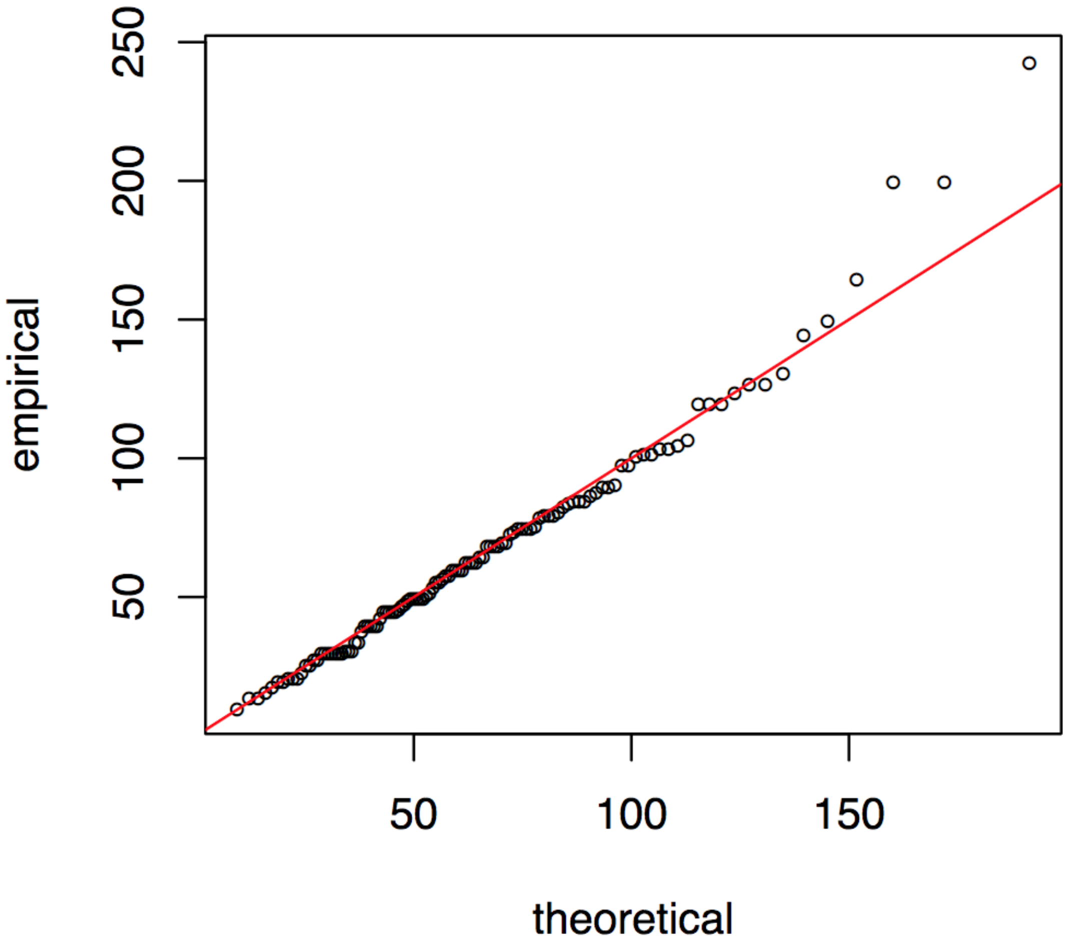
+ 0.5 pts Some sensible attempt but some mistake/shortcoming e.g. chosen bins make it hard to compare histogram with density curve (5 bins is too few), missing density curve (or dots instead of line) or missing histogram, correct code but no graph, etc

+ 0 pts Wrong/no attempt

1.4 **1(d)** **2** / 2 pts

Q: Draw a QQ plot to compare the data against the fitted gamma distribution. Include a reference line. Comment on the fit of the model to the data.

✔ **+ 2 pts** Perfect: Correct graph and sensible comment e.g. 'model is a good fit to data' and/or 'some potential deviation at right tail'



+ 2 pts Good, but used a line of best fit `abline(lm(y~x))` without checking whether or not the slope of this line is close to 1. The correct reference line to draw on this QQ plot is the 'y=x'line `abline(0,1)`.

+ 1 pt Some sensible attempt but some mistake e.g. comment missing, no line on plot, removed 'outliers', etc.

+ 0 pts Wrong (e.g. used qqnorm)/no attempt

Question 2

Question 2 **6** / 7 pts

2.1 2(a)i **2** / 2 pts

- ✔ **+ 1 pt** $E[x] = -2\theta + 3$
- ✔ **+ 0.5 pts** $E[x^2] = 2\theta^2 - 10\theta + 9$
- ✔ **+ 0.5 pts** $var(X) = 2\theta(1-\theta)$ or equivalent
- **+ 0.5 pts** $E(X)$, $E(X^2)$ and $var(X)$ all incorrect but some reasonable attempt made
- **+ 0 pts** Wrong/no attempt/no working shown

2.2 2(a)ii **1** / 1 pt

Q: Find the method of moments estimator and estimate of θ.

- ✔ **+ 1 pt** $\tilde{\theta} = \frac{3-\bar{X}}{2} = \frac{3-1.75}{2} = 0.625$ (or 5/8)
- **+ 0.5 pts** Wrong/no number but made an attempt to setup and solve $E(X) = \bar{X}$ OR correct number but no working shown
- **+ 0 pts** Wrong/no attempt

2.3 2(a)iii **1** / 1 pt

Find the standard error of this estimate.

- ✔ **+ 1 pt** $var(\bar{X}) = \frac{1}{n}var(x) = \frac{2\theta-2\theta^2}{n}$ and $var(\tilde{\theta}) = \frac{\theta-\theta^2}{2n}$ so $se(\tilde{\theta}) = \sqrt{\frac{2\theta-2\theta^2}{n}} \approx \sqrt{(0.625-0.625^2)/(2\times 20)} = 0.0765$ (equivalent answers e.g. $\frac{\sqrt{6}}{32}$ or $\frac{1}{16}\sqrt{\frac{3}{2}}$ are also ok)
- **+ 1 pt** Alternative acceptable solution: $se(\tilde{\theta}) \approx \frac{1}{2}\frac{s}{\sqrt{20}} = 0.0879$. Note: a more precise solution exists.
- **+ 0.5 pts** Some reasonable attempt but some mistake such as estimate value incorrect, estimate value wasn't calculated, forgot square root etc.
- **+ 0 pts** Wrong/no attempt

2.4 2(b)i **1** / 1 pt

Q: Find the likelihood function

- ✔ **+ 1 pt** Correct expression for the likelihood function $\theta^{2F_1}\,(2\theta(1-\theta))^{F_2}\,(1-\theta)^{2F_3}$ or $2^{F_2}\,\theta^{2F_1+F_2}\,(1-\theta)^{F_2+2F_3}$ (or equivalent)
- **+ 0.5 pts** Incorrect but reasonable attempt was made (e.g. final answer has + instead of x between terms)
- **+ 0 pts** Wrong/no attempt

1 Do not need to show log likelihood function in this question

2.5 **2(b)ii** **1** / 1 pt

Q: Find that the maximum likelihood estimator and estimate of θ.

✔ **+ 1 pt** MLE: $\theta = \frac{2F_1+F_2}{2n} = 0.625$ or 5/8 (need to show working)

+ 0.5 pts Incorrect/incomplete but reasonable attempt was made

+ 0 pts Wrong/no attempt

2.6 **2(b)iii** **0** / 1 pt

Q: Find the variance of this estimator.

+ 1 pt $var(\hat{\theta}) = \frac{\theta-\theta^2}{2n}$ with clear explanation e.g.

Since $F_1 + F_2 + F_3 = n$ and $n\bar{X} = \sum X_i = F_1 + 2F_2 + 3F_3$, we can obtain $2F_1 + F_2 = 3n - n\bar{X}$. Therefore, $\hat{\theta} = \frac{2F_1+F_2}{2n} = \frac{3-\bar{X}}{2} = \tilde{\theta}$, i.e. the MLE is the same as the method of moments estimator. So we have $\mathrm{var}(\hat{\theta}) = \mathrm{var}(\tilde{\theta}) = \frac{\theta-\theta^2}{2n}$.

(and/or estimated value=0.00586)

+ 0.5 pts Some reasonable attempt but incomplete or lacking explanation or some mistakes in the mathematical derivation of the expression for $var(\hat{\theta})$

✔ **+ 0 pts** Wrong/no attempt

Question 3

Question 3 **0.5** / 0.5 pts

3.1 **3(a)i** **0.5** / 0.5 pts

✔ **+ 0.5 pts** Any attempt on any part/s of Q3

+ 0 pts No attempt Q3

Question 4

Question 4 **0.5** / 0.5 pts

✔ **+ 0.5 pts** Any attempt on Q4

+ 0 pts No attempt on Q4

Question 5

Question 5 **6** / 6 pts

5.1 5(a) **4** / 4 pts

✔ **+ 2 pts** Method mark: any reasonable attempt at all/any parts

✔ **+ 1 pt** All $E(X)$ correct (don't necessarily need to get to final most refined answers here to prove that they're not equal to mu, but check for any logical flaws/mistakes, especially for T_4 ... need to at least provide justification that $E(X_4^2) > 0$):
$E(T_1) = \mu$
$E(T_2) = \frac{5}{3}\mu$
$E(T_3) = \mu$
$E(T_4) = \mu + \frac{1}{4}(\sigma^2 + \mu^2)$

✔ **+ 1 pt** Concluding that T_1 and T_3 are unbiased or award for any claim that matches to the student's findings and only award if student attempted all four expected values (even if some mistakes).

+ 0 pts Wrong/no attempt/simply stating expected values or that T1&T3 are unbiased with no working shown

5.2 5(b) **2** / 2 pts

✔ **+ 1 pt** T_3 has a smaller variance than T_1 (or conclusion consistent with variances calculated, even if one or both variances wrong).

✔ **+ 0.5 pts** $Var(T_1) = \frac{5}{18}\sigma^2 = 0.2778\sigma^2$ (ok as fraction) with working shown

✔ **+ 0.5 pts** $Var(T_3) = \frac{1}{4}\sigma^2 = 0.25\sigma^2$ (ok as fraction) with working shown

+ 0 pts Wrong/no attempt

Question 6

Neatness and code **0** / 0 pts

Neatness

✔ **– 0 pts** Assignment presentation is ok and working code present (if relevant). No messiness deduction.

No questions assigned to the following page.

Assignment 1

Name: James La Fontaine

Student Number: 1079860

Tutorial Day and Time: Friday 2:15 PM – 4:15 PM

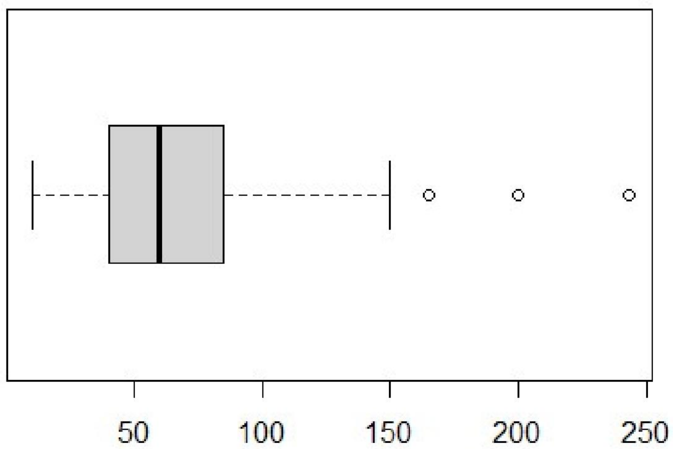
Tutor's Name: Haoyu Yang

Question assigned to the following page: 1.1

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 and are mostly concentrated between 10 and 100.

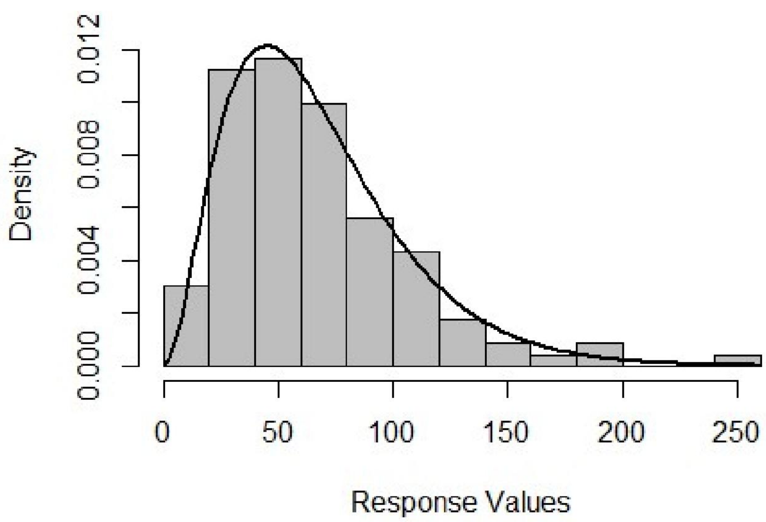
Questions assigned to the following page: 1.3 and 1.2

1b)

```r
# 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)

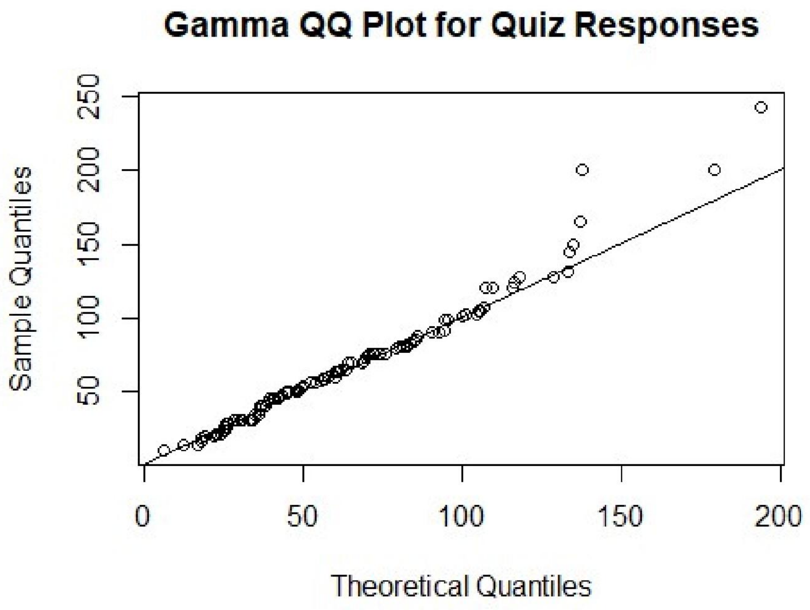
```r
hist(responses, freq = FALSE, col = "gray", main = NULL, xlab = "Response Values", nclass
= 10, ylim = c(0,0.013))
curve(dgamma(x, shape = shape.hat, scale = scale.hat), lwd = 2, add = TRUE)
```



Question assigned to the following page: 1.4

1d)

```
gamma.sample = rgamma(116, shape = shape.hat, scale = scale.hat)


qqplot(gamma.sample, responses, main = "Gamma QQ Plot for Quiz Responses", xlab = "Theore
tical 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.

Questions assigned to the following page: 2.1, 2.3, and 2.2

Question 2

a)

i) $E(X) = \sum x\, p(x)$

$= \theta^2 + 2(2\theta(1-\theta)) + 3(1-\theta)^2$

$= \theta^2 + 4\theta - 4\theta^2 + 3 - 6\theta + 3\theta^2$

$= 3 - 2\theta$

$var(X) = E(X^2) - E(X)^2$

$= \theta^2 + 4(2\theta(1-\theta)) + 9(1-\theta)^2 - \mu^2$

$= \theta^2 + 8\theta - 8\theta^2 + 9 - 18\theta + 9\theta^2 - \mu^2$

$= 2\theta^2 - 10\theta + 9 - (3-2\theta)^2$

$= 2\theta^2 - 10\theta + 9 - (9 - 12\theta + 4\theta^2)$

$= 2\theta - 2\theta^2$

ii) $\bar{X} = E(X)$

$\bar{X} = 3 - 2\theta$

$\bar{X} - 3 = -2\theta$

$\frac{3-\bar{X}}{2} = \tilde{\theta}$

$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = 1.75$

$\Rightarrow \tilde{\theta} = \frac{3-1.75}{2} = 0.625$

iii) $se(\tilde{\theta}) = \sqrt{\hat{var}(\tilde{\theta})}$

$= \sqrt{\hat{var}\left(\frac{3-\bar{X}}{2}\right)}$

$= \sqrt{\frac{1}{4}\hat{var}(\bar{X})}$

$= \frac{1}{2}\sqrt{\hat{var}(\bar{X})}$

$= \frac{1}{2}\sqrt{\frac{\hat{\sigma}^2}{20}}$

$= \frac{1}{\ldots}\sqrt{2\tilde{\theta} - 2\tilde{\theta}^2} = 0.\ldots$

$= 2\sqrt{20}\sqrt{20} - 20 \approx 0.077$

Question 2

b)

i) $$L(\theta) = \theta^{2F_1} \cdot (2\theta - 2\theta^2)^{F_2} \cdot (1-\theta)^{2F_3}$$

$$\ln L(\theta) = \ln(\theta^{2F_1}) \cdot \ln((2\theta - 2\theta^2)^{F_2}) \cdot \ln((1-\theta)^{2F_3})$$

$$\ln L(\theta) = 2F_1 \ln(\theta) \cdot F_2 \ln(2\theta - 2\theta^2) \cdot 2F_3 \ln(1-\theta)$$

ii) $$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{2F_1}{\theta} + \frac{F_2(2-4\theta)}{2\theta - 2\theta^2} - \frac{2F_3}{1-\theta} = 0$$

$$\Rightarrow \frac{2F_1}{\theta} + \frac{F_2(1-2\theta)}{\theta - \theta^2} - \frac{2F_3}{1-\theta} = 0$$

$$\Rightarrow \frac{2F_1}{\theta} + \frac{F_2(1-2\theta)}{\theta(1-\theta)} - \frac{2F_3}{1-\theta} = 0$$

$$\Rightarrow 2F_1(1-\theta) + F_2(1-2\theta) - 2F_3\theta = 0$$

$$\Rightarrow 2F_1 - 2F_1\theta + F_2 - 2F_2\theta - 2F_3\theta = 0$$

$$\Rightarrow -2F_1\theta - 2F_2\theta - 2F_3\theta = -2F_1 - F_2$$

$$\Rightarrow \theta(-2F_1 - 2F_2 - 2F_3) = -2F_1 - F_2$$

$$\Rightarrow \hat{\theta} = \frac{2F_1 + F_2}{2(F_1 + F_2 + F_3)} = \frac{2 \cdot 9 + 7}{2 \cdot (9+7+4)} = \frac{5}{8} = 0.625$$

Question assigned to the following page: 2.6

Question 2

iii) $var(\hat{\theta}) = var\left(\frac{2F_1 + F_2}{2(F_1 + F_2 + F_3)}\right)$

$\bar{X} = \frac{F_1 + 2F_2 + 3F_3}{n}$

Question assigned to the following page: 3.1

Question 3

a) $X \sim \text{Unif}(0, \theta)$

i) $E(X) = \frac{\theta}{2}$, $var(X) = \frac{\theta^2}{12}$

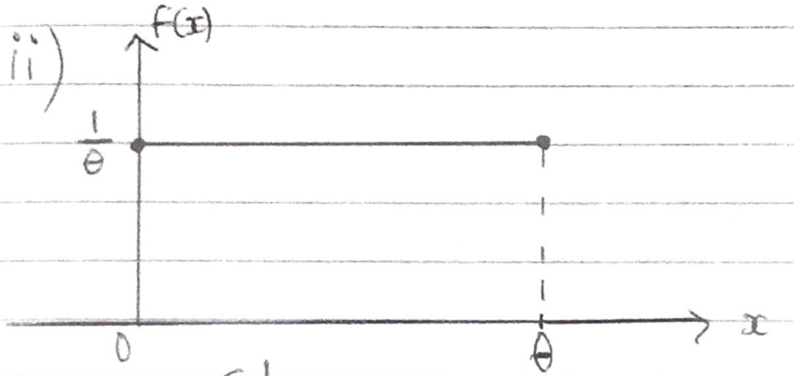
$\bar{X} = \frac{\theta}{2}$

$\tilde{\theta} = 2\bar{X}$

$E(2\bar{X}) = 2E(\bar{X}) = \theta$

$var(2\bar{X}) = 4var(\bar{X}) = \frac{4\sigma^2}{1} = \frac{\theta^2}{3}$

ii)



$$L(\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$\frac{1}{\theta}$ is highest when θ is lowest and $\theta \ge X_{(n)}$

$\Rightarrow \hat{\theta} = X_{(n)} = X_{(1)}$ as $n=1$

$E(X_{(n)}) = \frac{\theta}{1+1} = \frac{\theta}{2}$

$Var(X_{(n)}) = E(X_{(n)}^2) - E(X_{(n)})^2$

$= \frac{\theta^2}{1+2} - \frac{\theta^2}{4} = \frac{\theta^2}{3} - \frac{\theta^2}{4}$

$= \frac{\theta^2}{12}$

No questions assigned to the following page.

Question 3

b)

i) $var(\hat{\theta}-\theta) = E[(\hat{\theta}-\theta)^2] - (E(\hat{\theta}-\theta))^2$

$Var(\hat{\theta}) = MSE(\hat{\theta}) - (E(\hat{\theta})-\theta)^2$

$var(\hat{\theta}) = MSE(\hat{\theta}) - bias(\hat{\theta})^2$

$\Rightarrow MSE(\hat{\theta}) = var(\hat{\theta}) + bias(\hat{\theta})^2$

$\square$

ii) MME: $\tilde{\theta} = 2\bar{X}$

$MSE(\tilde{\theta}) = var(\tilde{\theta}) + bias(\tilde{\theta})^2$

$= \frac{\theta^2}{3} + 0 = \frac{\theta^2}{3}$

MLE: $\hat{\theta} = X_{(n)}$

$MSE(\hat{\theta}) = var(\hat{\theta}) + bias(\hat{\theta})^2$

$= \frac{\theta^2}{12} + \left(\frac{\theta}{2} - \theta\right)^2$

$= \frac{\theta^2}{12} + \frac{\theta^2}{4}$

$= \frac{\theta^2}{3} = MSE(\tilde{\theta})$

iii) $T_3 = X_{(n)} - \frac{\theta}{2}$

$MSE(T_3) = var\left(X_{(n)} - \frac{\theta}{2}\right) + bias\left(X_{(n)} - \frac{\theta}{2}\right)^2$

$= var(X_{(n)}) + \left[E(X_{(n)}) - \frac{\theta}{2}\right]^2$

$= \frac{\theta^2}{12} < MSE(\hat{\theta}) < MSE(\tilde{\theta})$

No questions assigned to the following page.

Question 3

c)

i) $E(X_i) = \frac{\theta}{2}$ $\quad$ $var(X_i) = \frac{\theta^2}{12}$

$\bar{X} = \frac{\theta}{2}$

$\tilde{\theta} = 2\bar{X}$

$E(2\bar{X}) = 2E(\bar{X}) = \theta$

$var(2\bar{X}) = 4var(\bar{X}) = \frac{4\sigma^2}{n} = \frac{\theta^2}{3n}$

$MSE(\tilde{\theta}) = var(\tilde{\theta}) + bias(\tilde{\theta})^2$

$= \frac{\theta^2}{3n} + 0 = \frac{\theta^2}{3n}$

No questions assigned to the following page.

Question 3

c)

ii) $L(\theta) = \begin{cases} \frac{1}{\theta^n} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$

$\frac{1}{\theta^n}$ is highest when θ is lowest and $\theta \ge X_{(n)}$

$\Rightarrow \hat{\theta} = X_{(n)}$

$E(X_{(n)}) = \frac{n\theta}{n+1}$

$$\begin{aligned} var(X_{(n)}) &= E(X^2_{(n)}) - E(X_{(n)})^2 \\ &= \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} \\ &= \frac{n\theta^2(n+1)^2}{(n+2)(n+1)^2} - \frac{n^2\theta^2(n+2)}{(n+1)^2(n+2)} \\ &= \frac{n\theta^2(n+1)^2 - n^2\theta^2(n+2)}{(n+2)(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2} \end{aligned}$$

$$\begin{aligned} MSE(\hat{\theta}) &= var(\hat{\theta}) + bias(\hat{\theta})^2 \\ &= \frac{n\theta^2}{(n+2)(n+1)^2} + \left(\frac{n\theta}{n+1} - \theta\right)^2 \\ &= \frac{n\theta^2}{(n+2)(n+1)^2} + \left(\frac{n\theta - \theta(n+1)}{n+1}\right)^2 \\ &= \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2(n+2)}{(n+2)(n+1)^2} \\ &= \frac{n\theta^2 + \theta^2(n+2)}{(n+2)(n+1)^2} \\ &= \frac{\theta^2(n+n+2)}{(n+2)(n+1)^2} = \frac{\theta^2 \cdot 2(n+1)}{(n+2)(n+1)^2} = \frac{2\theta^2}{(n+2)(n+1)} < MSE(\tilde{\theta}) \end{aligned}$$

No questions assigned to the following page.

Question 3

c)

iii) MSE is minimised for $\hat{\theta}$ when the bias$(\hat{\theta}) = 0$

bias$(a\hat{\theta}) = 0$

$E(a\hat{\theta}) - \theta = 0$

$\frac{an\theta}{n-1} = \theta$

$an\theta = \theta(n-1)$

$an = n-1$

$a = \frac{n-1}{n}$

Question assigned to the following page: 4

Question 4

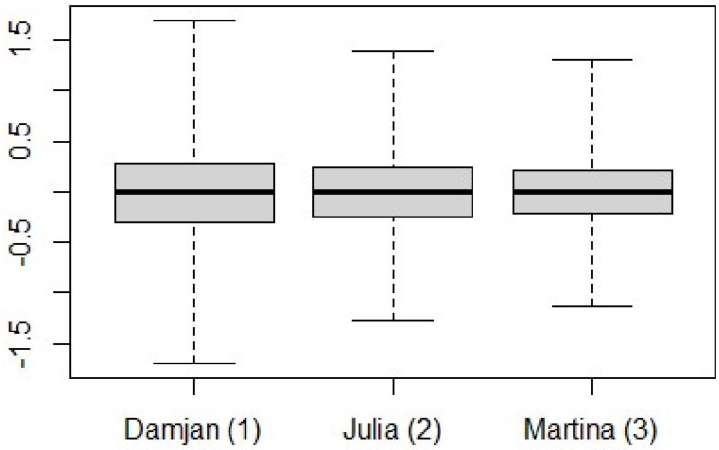
Let Damjan's average of the sample minimum and maximum be Estimator 1. Let Julia's sample median be Estimator 2. Let Martina's sample mean be Estimator 3.

```
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
```

Question assigned to the following page: 4

Question 4 (cont.)



All 3 estimators appear to have negligible bias which can likely be explained by the randomness of the samples in the simulation. Martina's sample mean estimator appears to have achieved the lowest variance out of all 3 of the estimators and so we would expect it to be the most accurate estimator to use for the population mean. Note that the boxplot whiskers have been extended out to the maximums and minimums of each boxplot for improved visual clarity.

Question assigned to the following page: 5.1

Question 5

a) $E(X_i) = \mu \quad var(X_i) = \sigma^2 > 0$

$$bias(T_1) = E(T_1) - \mu$$
$$= E\left[\frac{1}{3}(X_1 + X_2) + \frac{1}{6}(X_3 + X_4)\right] - \mu$$
$$= E\left[\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3 + \frac{1}{6}X_4\right] - \mu$$
$$= \frac{1}{3}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{6}E(X_3) + \frac{1}{6}E(X_4) - \mu$$
$$= \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu + \frac{1}{6}\mu - \mu = 0$$

$\Rightarrow T_1$ is unbiased

$$bias(T_2) = E(T_2) - \mu$$
$$= E\left[\frac{1}{6}(X_1 + 2X_2 + 3X_3 + 4X_4)\right] - \mu$$
$$= E\left[\frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3 + \frac{2}{3}X_4\right] - \mu$$
$$= \frac{1}{6}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{2}E(X_3) + \frac{2}{3}E(X_4) - \mu$$
$$= \frac{1}{6}\mu + \frac{1}{3}\mu + \frac{1}{2}\mu + \frac{2}{3}\mu - \mu = \frac{2}{3}\mu$$

$\Rightarrow T_2$ is biased

$$bias(T_3) = E(T_3) - \mu$$
$$= E\left[\frac{1}{4}(X_1 + X_2 + X_3 + X_4)\right] - \mu$$
$$= E\left[\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + \frac{1}{4}X_4\right] - \mu$$
$$= \frac{1}{4}E(X_1) + \frac{1}{4}E(X_2) + \frac{1}{4}E(X_3) + \frac{1}{4}E(X_4) - \mu$$
$$= \frac{1}{4}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu - \mu = 0$$

$\Rightarrow T_3$ is unbiased

Questions assigned to the following page: 5.1 and 5.2

Question 5

a)

$$\text{bias}(T_4) = E(T_4) - \mu$$
$$= E\left[\frac{1}{3}(X_1 + X_2 + X_3) + \frac{1}{4}X_4^2\right] - \mu$$
$$= E\left[\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 + \frac{1}{4}X_4^2\right] - \mu$$
$$= \frac{1}{3}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{3}E(X_3) + \frac{1}{4}E(X_4^2) - \mu$$
$$= \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{4}\mu^2 - \mu = \frac{1}{4}\mu^2$$

$\Rightarrow T_4$ is biased

b)
$$\text{var}(T_1) = \text{var}\left(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3 + \frac{1}{6}X_4\right)$$
$$= \frac{1}{9}\text{var}(X_1) + \frac{1}{9}\text{var}(X_2) + \frac{1}{36}\text{var}(X_3) + \frac{1}{36}\text{var}(X_4)$$
$$= \frac{\sigma^2}{9} + \frac{\sigma^2}{9} + \frac{\sigma^2}{36} + \frac{\sigma^2}{36} = \frac{5\sigma^2}{18}$$

$$\text{Var}(T_3) = \text{var}\left(\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + \frac{1}{4}X_4\right)$$
$$= \frac{1}{16}\text{var}(X_1) + \frac{1}{16}\text{var}(X_2) + \frac{1}{16}\text{var}(X_3) + \frac{1}{16}\text{var}(X_4)$$
$$= \frac{\sigma^2}{4}$$

$$\Rightarrow \text{var}(T_3) < \text{var}(T_1)$$

$$\frac{\sigma^2}{4} < \frac{5\sigma^2}{18}$$

We would expect T_3 to be a more accurate estimator of μ than T_1.