Assignment 1

Name: James La Fontaine

Student Number: 1079860

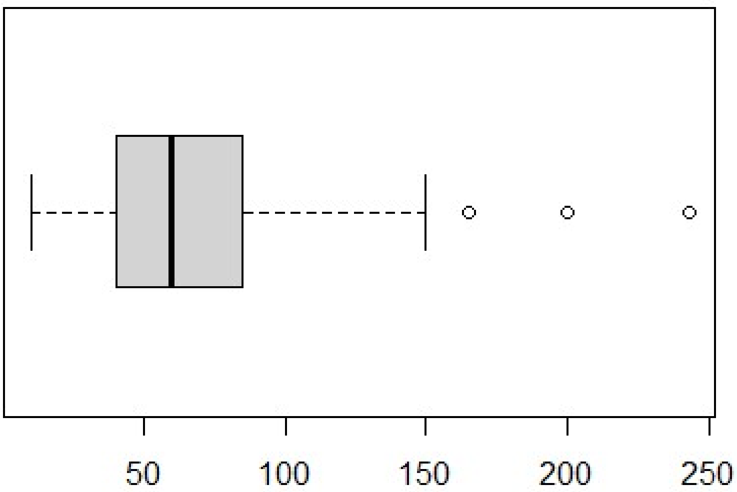
Tutorial Day and Time: Friday 2:15 PM – 4:15 PM

Tutor's Name: Haoyu Yang

Question 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)
```



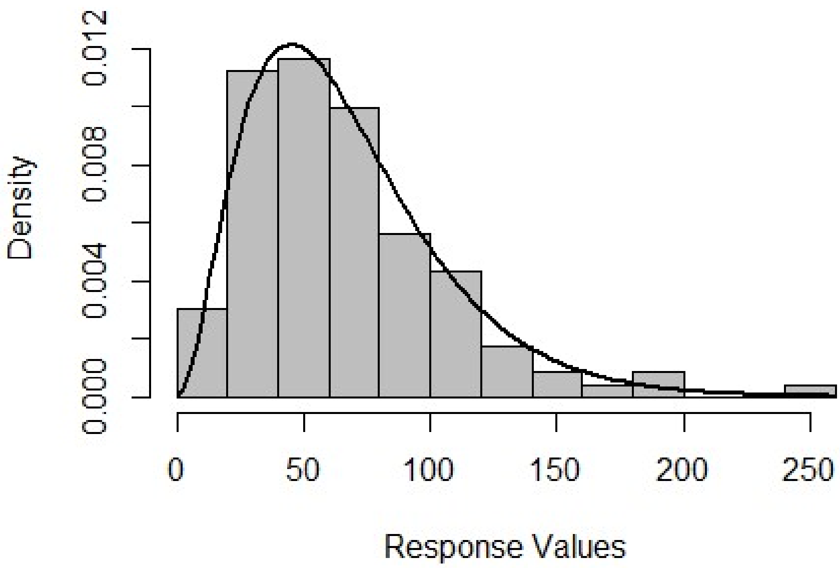
This distribution is asymmetrical and positively skewed with the centre lying roughly below the median 60. This distribution has a relatively large spread with a range of 233 due to an outlier, and even still has a range of approximately 140 when ignoring outliers. The distribution additionally has a 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.

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
```

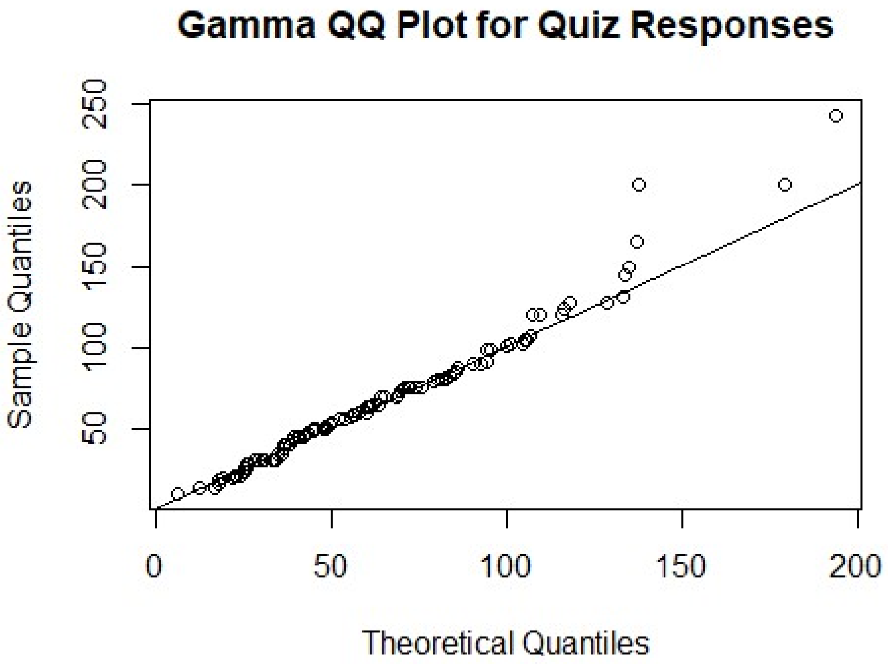
1c)

```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)
```



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

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

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

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

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

$= \frac{1}{2\sqrt{20}}\sqrt{2\tilde{\theta} - 2\tilde{\theta}^2} = 0.077$

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

$\Rightarrow \frac{2F_1}{\theta} + \frac{f_2(1-2\theta)}{\theta - \theta^2} - \frac{2f_3}{1-\theta} = 0$

$\Rightarrow \frac{2f_1}{\theta} + \frac{F_2(1-2\theta)}{\theta(1-\theta)} - \frac{2f_3}{1-\theta} = 0$

$\Rightarrow 2f_1(1-\theta) + F_2(1-2\theta) - 2f_3\theta = 0$

$\Rightarrow 2f_1 - 2F_1\theta + F_2 - 2F_2\theta - 2F_3\theta = 0$

$\Rightarrow -2f_1\theta - 2F_2\theta - 2F_3\theta = -2f_1 - F_2$

$\Rightarrow \theta(-2f_1 - 2F_2 - 2F_3) = -2f_1 - F_2$

$\Rightarrow \hat{\theta} = \frac{2F_1 + f_2}{2(F_1 + F_2 + F_3)} = \frac{2 \cdot 9 + 7}{2 \cdot (9+7+4)} = 5/8 = 0.625$

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 3

a) $X \sim Unif(0, \theta)$

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

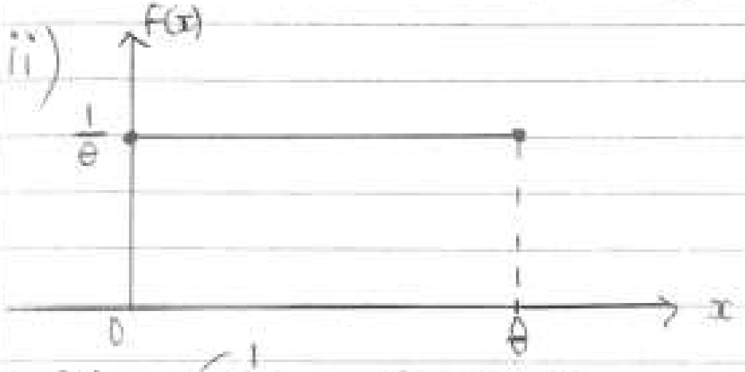
$\bar{X} = \frac{\theta}{2}$

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

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

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

ii)



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

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

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

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

$$Var(X_{(n)}) = E(X_{(n)}^2) - E(X_{(n)})^2$$
$$= \frac{\theta^2}{1+2} - \frac{\theta^2}{4} = \frac{\theta^2}{3} - \frac{\theta^2}{4}$$
$$= \frac{\theta^2}{12}$$

Question 3

b)

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

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

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

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

$\square$

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

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

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

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

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

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

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

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

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

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

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

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

Question 3

c)

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

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

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

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

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

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

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

Question 3

c)

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

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

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

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

$var(X_{(n)}) = E(X^2_{(n)}) - E(X_{(n)})^2$

$= \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}$

$= \frac{n\theta^2(n+1)^2}{(n+2)(n+1)^2} - \frac{n^2\theta^2(n+2)}{(n+1)^2(n+2)}$

$= \frac{n\theta^2(n+1)^2 - n^2\theta^2(n+2)}{(n+2)(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}$

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

$= \frac{n\theta^2}{(n+2)(n+1)^2} + \left(\frac{n\theta}{n+1} - \theta\right)^2$

$= \frac{n\theta^2}{(n+2)(n+1)^2} + \left(\frac{n\theta - \theta(n+1)}{n+1}\right)^2$

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

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

$= \frac{\theta^2(n+n+2)}{(n+2)(n+1)^2} = \frac{\theta^2 \cdot 2(n+1)}{(n+2)(n+1)^2} = \frac{2\theta^2}{(n+2)(n+1)} < MSE(\tilde{\theta})$

Question 3

c)

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

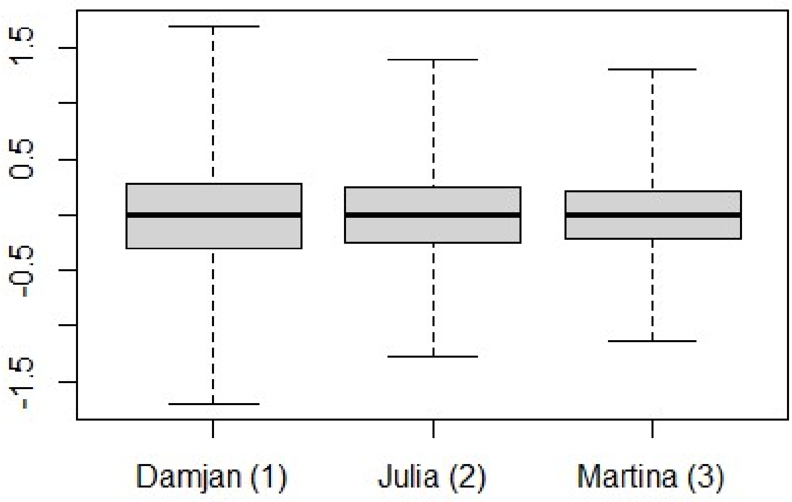
$$\text{bias}(a\hat{\theta}) = 0$$
$$E(a\hat{\theta}) - \theta = 0$$
$$\frac{an\theta}{n-1} = \theta$$
$$an\theta = \theta(n-1)$$
$$an = n-1$$
$$a = \frac{n-1}{n}$$

Question 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 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 5

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

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

$\Rightarrow T_1$ is unbiased

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

$\Rightarrow T_2$ is biased

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

$\Rightarrow T_3$ is unbiased

Question 5

a)

$\text{bias}(T_4) = E(T_4) - \mu$

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

$= E[\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 + \frac{1}{4}X_4^2] - \mu$

$= \frac{1}{3}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{3}E(X_3) + \frac{1}{4}E(X_4^2) - \mu$

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

$\Rightarrow T_4$ is biased

b) $\text{var}(T_1) = \text{var}(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3 + \frac{1}{6}X_4)$

$= \frac{1}{9}\text{var}(X_1) + \frac{1}{9}\text{var}(X_2) + \frac{1}{36}\text{var}(X_3) + \frac{1}{36}\text{var}(X_4)$

$= \frac{\sigma^2}{9} + \frac{\sigma^2}{9} + \frac{\sigma^2}{36} + \frac{\sigma^2}{36} = \frac{5\sigma^2}{18}$

$\text{Var}(T_3) = \text{var}(\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3 + \frac{1}{4}X_4)$

$= \frac{1}{16}\text{var}(X_1) + \frac{1}{16}\text{var}(X_2) + \frac{1}{16}\text{var}(X_3) + \frac{1}{16}\text{var}(X_4)$

$= \frac{\sigma^2}{4}$

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

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

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