

# HW3

Patrick Gardocki

2023-06-09

## 6-30

```
octane = read.table("6-30.txt",header = TRUE)

octane=octane$Rating

o = sort(octane)

quantile(o, c(0.25,0.5,0.75),type=6)
```

```
##      25%      50%      75%
## 88.575 90.400 92.200
```

```
stem(octane, scale=2)
```

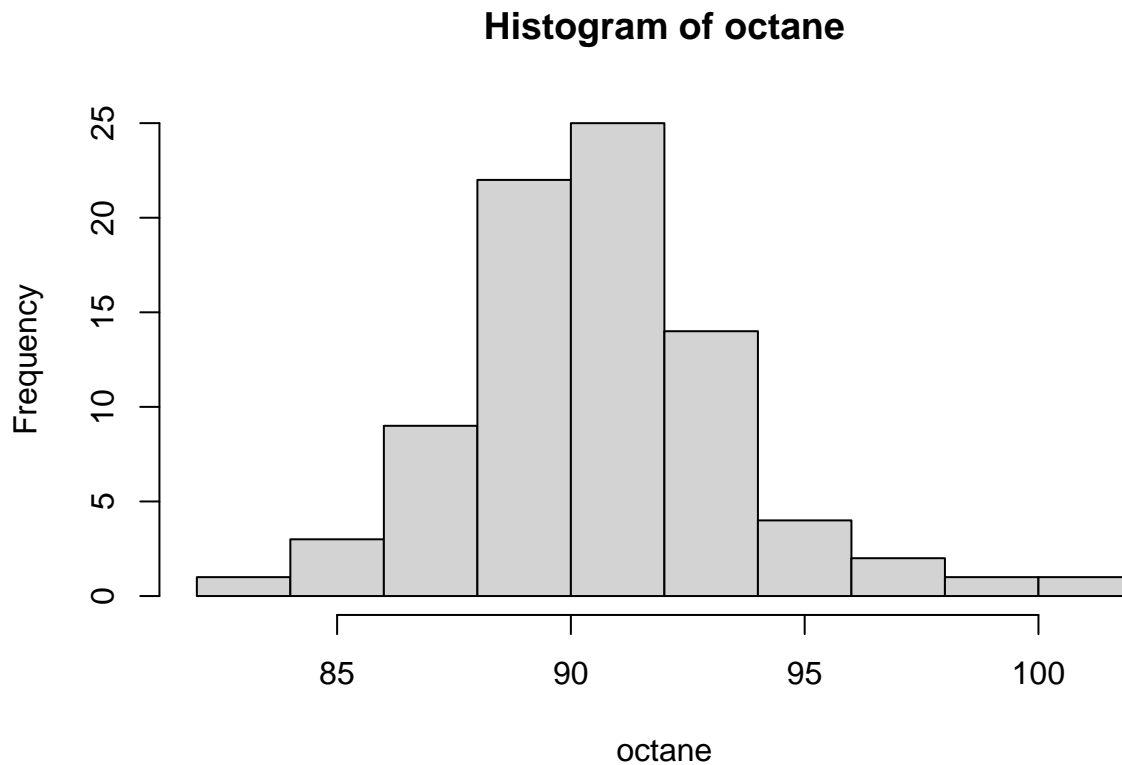
```
##
## The decimal point is at the |
##
##      83 | 4
##      84 | 33
##      85 | 3
##      86 | 777
##      87 | 456789
##      88 | 23334556679
##      89 | 0233678899
##      90 | 0111344456789
##      91 | 0001112256688
##      92 | 22236777
##      93 | 023347
##      94 | 2247
##      95 |
##      96 | 15
##      97 |
##      98 | 8
##      99 |
##     100 | 3
```

## 6-46 REL DENSity is $1/2$ ??????????

```
octane = read.table("6-30.txt",header = TRUE)

octane=octane$Rating

obj = hist(octane,breaks=8)
```



```
length(octane)
```

```
## [1] 82
```

```
library(knitr)
library(kableExtra)
df = data.frame(Frequency = obj$counts)

n=length(obj$breaks)
for(i in 1:(n-2)){
  df$Class[i]=paste(obj$breaks[i],"$\\le x <$", obj$breaks[i+1])
}

df$Class[n-1] = paste(obj$breaks[n-1],"$\\le x \\le$", obj$breaks[n])
```

Table 1: Frequency Distribution Table

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
$82 \leq x < 84$	1	0.0121951	1	0.0121951219512195
$84 \leq x < 86$	3	0.0365854	4	0.0487804878048781
$86 \leq x < 88$	9	0.1097561	13	0.158536585365854
$88 \leq x < 90$	22	0.2682927	35	0.426829268292683
$90 \leq x < 92$	25	0.3048780	60	0.731707317073171
$92 \leq x < 94$	14	0.1707317	74	0.902439024390244
$94 \leq x < 96$	4	0.0487805	78	0.951219512195122
$96 \leq x < 98$	2	0.0243902	80	0.975609756097561
$98 \leq x < 100$	1	0.0121951	81	0.98780487804878
$100 \leq x \leq 102$	1	0.0121951	82	1

```
df$"Relative Frequency"= obj$density*2
for(i in 1:(n-1)){
  df$"Cumulative Frequency"[i]=paste(cumsum(obj$counts)[i])
  df$"Cumulative Relative Frequency"[i]=paste(cumsum(obj$density*2)[i])
}

kable(df[c(2,1,3,4,5)], "latex", align="c", caption="Frequency Distribution Table", escape = FALSE)
```

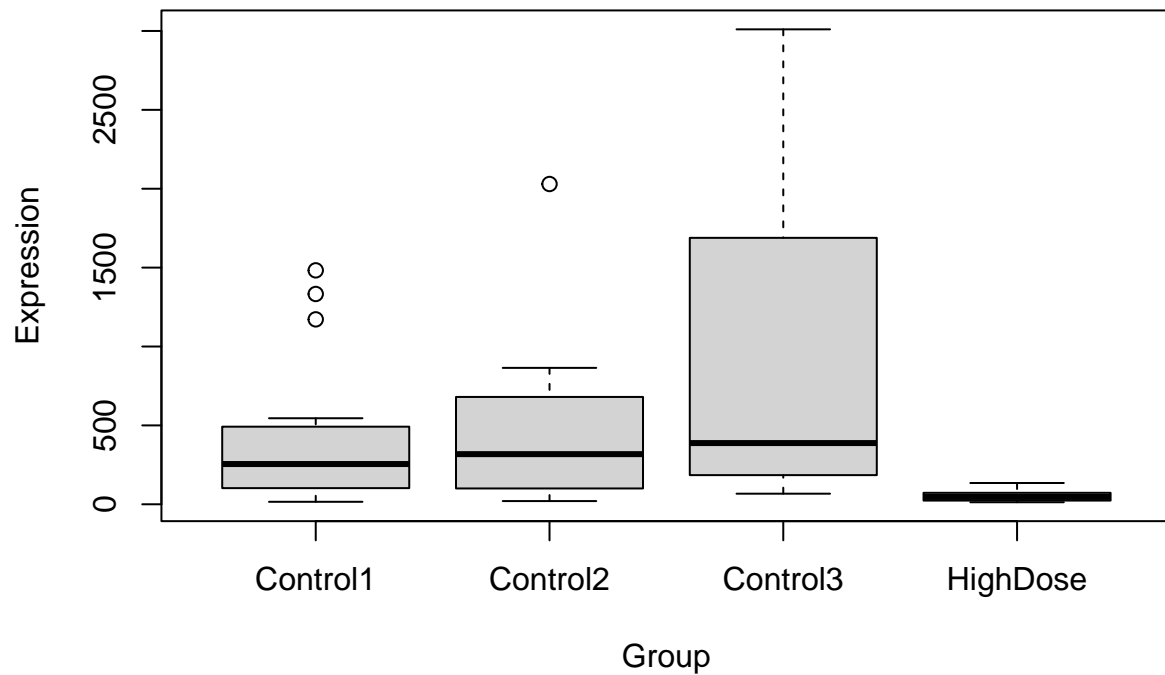
## 6-81

```
Expr = read.table("6-81.txt",header = TRUE)

attach(Expr)

boxplot(Expression~Group, main="Boxplot of 4 Groups",
  xlab="Group", ylab="Expression")
```

## Boxplot of 4 Groups

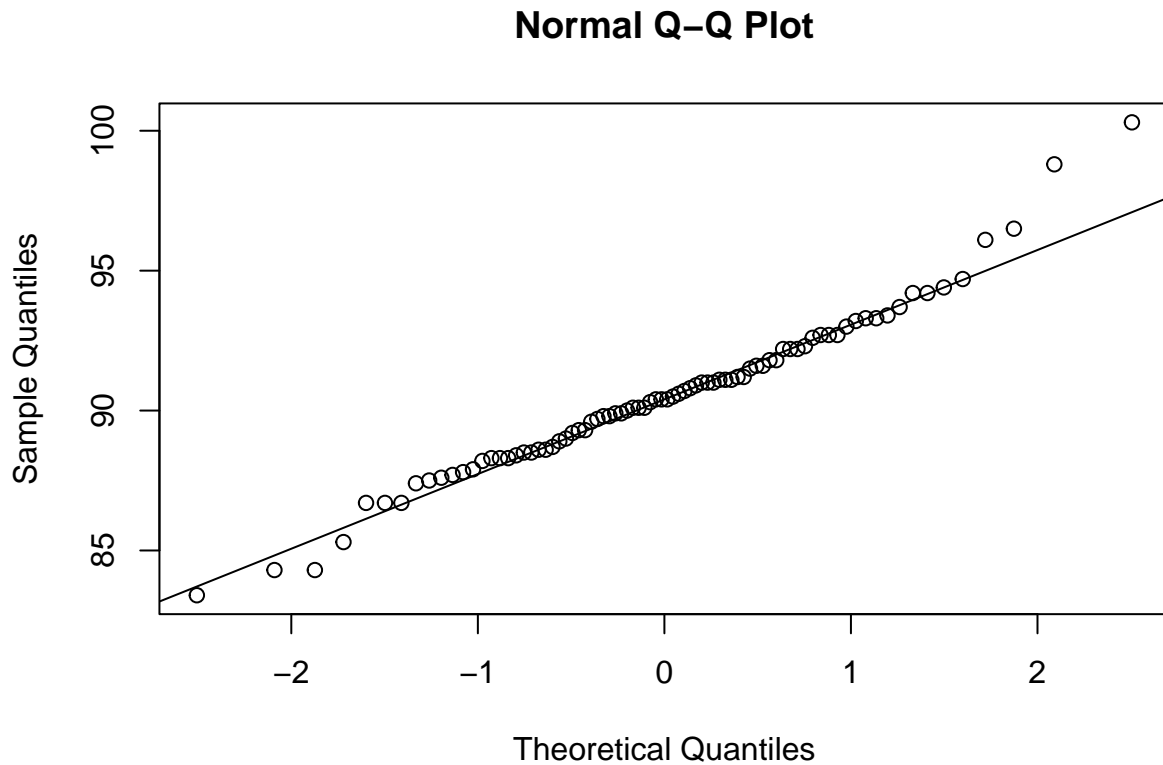


The control groups seem to have the same median value but drastically different variances. The 'high dose' group has a much smaller variance and slightly lower median expression as well.

## 6-98

```
octane = read.table("6-30.txt", header = TRUE)

octane=octane$Rating
qqnorm(octane)
qqline(octane)
```



The Normal Probability plot suggests that the data is normally distributed but seems to break down at both ends of the data range.

## 7-12

$\mu_X = 8.2 \text{ minutes}$ ,  $n = 49$ ,  $\sigma_X = 1.5 \text{ minutes}$ ,  $\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}} = 0.2143$  Under Central Limit Theorem,  $\bar{X}$  is approx. normally distributed.

(a)

$$P(\bar{X} < 10) = P(Z < \frac{10 - \mu}{\sigma_{\bar{X}}}) = P(Z < 8.4) = 1$$

(b)

$$P(5 < \bar{X} < 10) = P(\frac{5 - \mu}{\sigma_{\bar{X}}} < Z < \frac{10 - \mu}{\sigma_{\bar{X}}}) = P(Z < 8.4) - P(Z < -14.932) = 1 - 0 = 1$$

(c)

$$P(\bar{X} < 6) = P(Z < \frac{6 - \mu}{\sigma_{\bar{X}}}) = P(Z < -10.27) = 0$$

## 7-37

a)

if  $E(\hat{\Theta}) = \Theta$  then unbiased estimator

$$E(\hat{X}_1 - \hat{X}_2) = E(\hat{X}_1) - E(\hat{X}_2) = \mu_1 - \mu_2 \therefore \text{unbiased}$$

b)

$$\sigma_{\bar{X}} = \sqrt{V(\hat{\Theta})} = \sqrt{V(\hat{X}_1 - \hat{X}_2)} = \sqrt{V(\hat{X}_1) + V(\hat{X}_2) + 2COV(\hat{X}_1, \hat{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Estimating the standard error could be done by replacing the variance,  $\sigma$  with sample standard deviation, S.

c)

$$S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$E(S_P^2) = E\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right) = \frac{1}{n_1+n_2-2} \left((n_1-1)E(S_1^2) + (n_2-1)E(S_2^2)\right) = \frac{1}{n_1+n_2-2} \left((n_1-1) * \sigma_1^2 + (n_2-1) * \sigma_2^2\right) = \sigma^2 * \frac{n_1+n_2-2}{n_1+n_2-2} = \sigma^2$$

## 7-44 CHECK!!!!!!!!!!

$$f(x) = p(1-p)^{x-1}$$

$$L(p) = \prod_{i=1}^n p(1-p)^{x_i-1} = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

$$l(p) = n \ln(p) + \left(\sum_{i=1}^n x_i - n\right) \ln(1-p)$$

$$\frac{\partial l}{\partial p} = 0 = \frac{n}{p} - \frac{\sum_{i=1}^n x_i - n}{1-p} = \frac{n(1-p) - p \sum_{i=1}^n x_i - n}{p(1-p)} = n - p \sum_{i=1}^n x_i \therefore$$

$$p = \frac{n}{\sum_{i=1}^n x_i}$$

## 8-1

a)

$$z_0 = 2.14 \rightarrow P(Z < 2.14) = 0.9838 \therefore \alpha = 2 \times (1 - 0.9838) \rightarrow CI = 100 \times (1 - \alpha) = 96.76\%$$

b)

$$z_0 = 2.49 \rightarrow P(Z < 2.49) = 0.9963 \therefore \alpha = 2 \times (1 - 0.9963) \rightarrow CI = 100 \times (1 - \alpha) = 98.72\%$$

c)

$$z_0 = 1.85 \rightarrow P(Z < 1.85) = 0.9678 \therefore \alpha = 2 \times (1 - 0.9678) \rightarrow CI = 100 \times (1 - \alpha) = 93.56\%$$

d)

$$z_0 = 2.00 \rightarrow P(Z < 2.00) = 0.9772 \therefore \alpha = (1 - 0.9772) \rightarrow CI = 100 \times (1 - \alpha) = 97.72\%$$

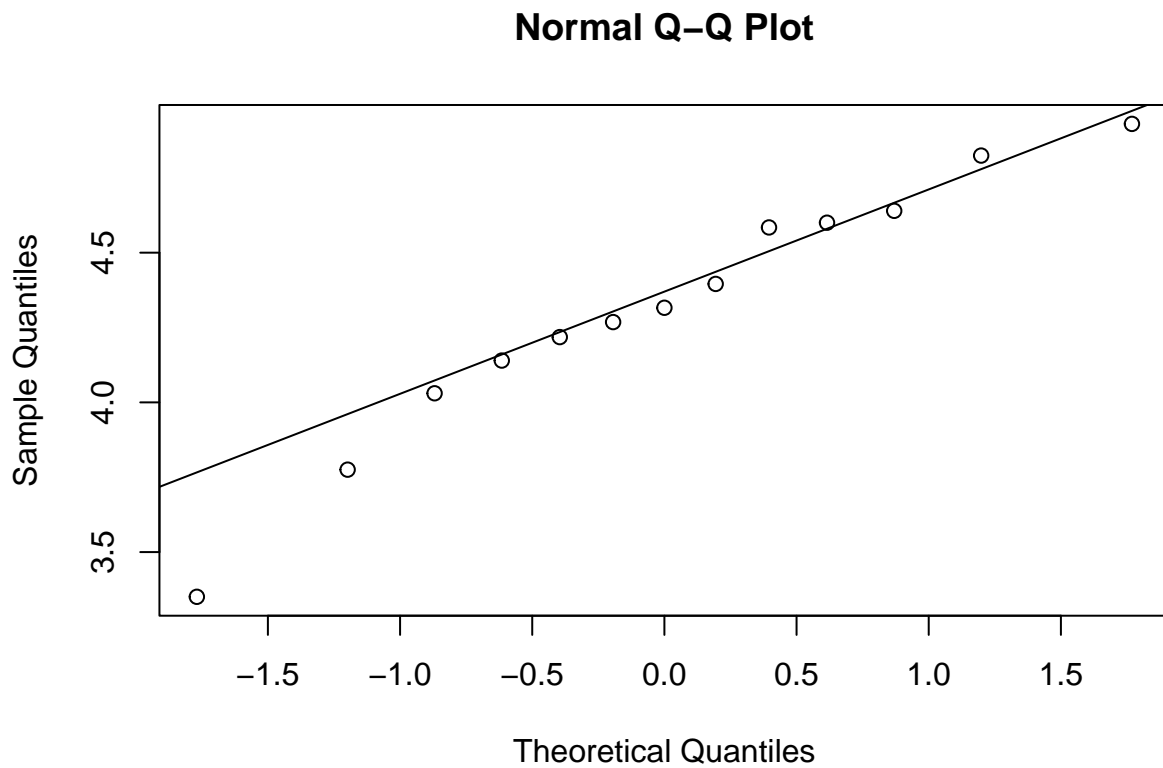
e)

$$z_0 = 1.96 \rightarrow P(Z < 1.96) = 0.9750 \therefore \alpha = (1 - 0.9750) \rightarrow CI = 100 \times (1 - \alpha) = 97.50\%$$

8-41

a)

```
speed = c(3.775302, 3.350679, 4.217981, 4.030324, 4.639692, 4.139665, 4.395575, 4.824257, 4.268119, 4.511119)
qqnorm(speed)
qqline(speed)
```



```
library(nortest)
ad.test(speed)
```

```
##
## Anderson-Darling normality test
##
## data: speed
## A = 0.23339, p-value = 0.7448
```

b)

$$\bar{x} - t_{0.025,12} \left( \frac{s}{\sqrt{n}} \right) \leq \mu \leq \bar{x} + t_{0.025,12} \left( \frac{s}{\sqrt{n}} \right)$$

$\mu = 4.313$ ;  $s = 0.4328$ ;  $n = 13$ ;  $t_{0.025,12} = 2.179$

$$4.313 \pm 2.179 \left( \frac{0.4328}{\sqrt{13}} \right) \rightarrow 4.051 \leq \mu \leq 4.575$$

```
mean(speed)
```

```
## [1] 4.313222
```

```
sd(speed)
```

```
## [1] 0.4328017
```

```
length(speed)
```

```
## [1] 13
```

```
qt(.025,12)
```

```
## [1] -2.178813
```

```
qt(.05,12)
```

```
## [1] -1.782288
```

c)

$$\bar{x} - t_{0.05,12} \left( \frac{s}{\sqrt{n}} \right) \leq \mu$$

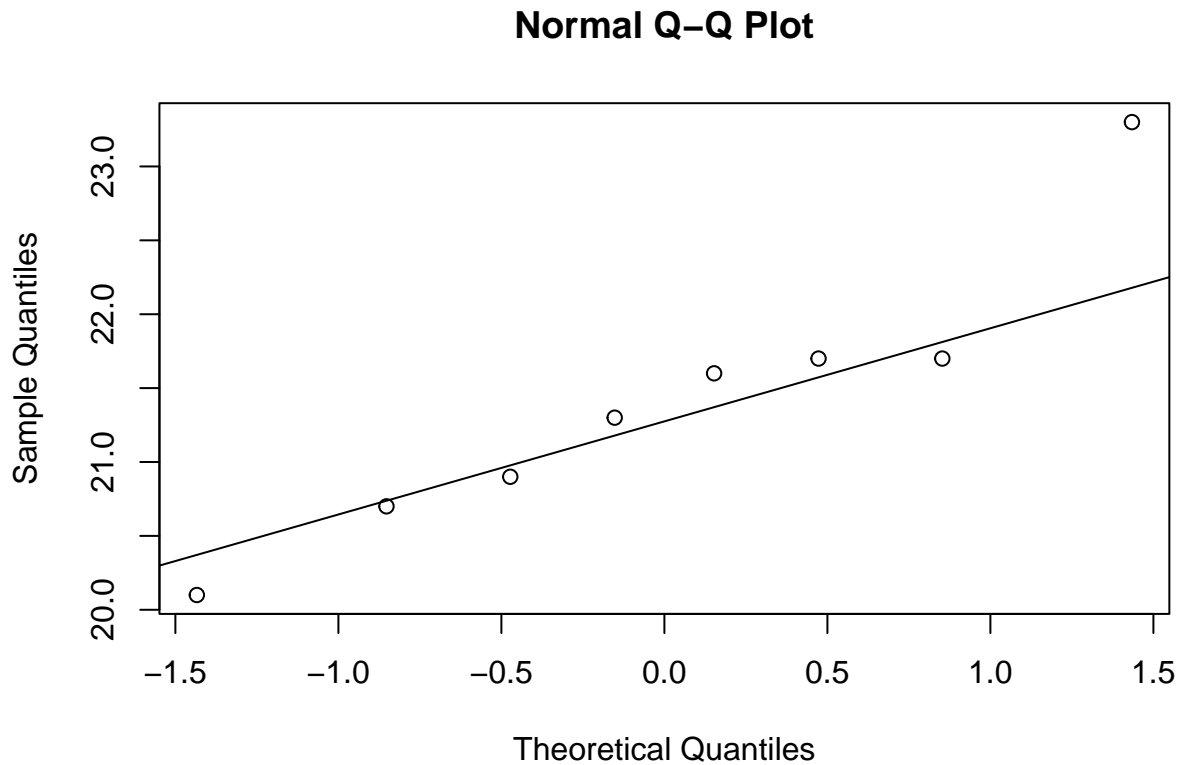
$\mu = 4.313$ ;  $s = 0.4328$ ;  $n = 13$ ;  $t_{0.05,12} = 1.782$

$$4.313 - 1.782 \left( \frac{0.4328}{\sqrt{13}} \right) \rightarrow 4.099 \leq \mu$$

**8-53**



```
temp = c(23.3, 21.7, 21.6, 21.7, 21.3, 20.7, 20.9, 20.1)
qqnorm(temp)
qqline(temp)
```



```
library(nortest)
ad.test(temp)
```

```
##
##  Anderson-Darling normality test
##
## data:  temp
## A = 0.35247, p-value = 0.3666
```

```
sd(temp)
```

```
## [1] 0.9463275
```

```
length(temp)
```

```
## [1] 8
```

$n = 8; s = 0.9463$

$\chi^2_{\alpha/2, n-1} = \chi^2_{0.025, 7} = 16.012; \chi^2_{0.975, 7} = 1.689$

$\frac{7(0.9463)^2}{16.012} \leq \sigma^2 \leq \frac{7(0.9463)^2}{1.689} = 0.392 \leq \sigma^2 \leq 3.709 \rightarrow 0.626 \leq \sigma \leq 1.926$

## Q2

a)

for  $\mu_1$ :  $E(\frac{\sum X_i}{n} + \frac{\sum Y_j}{m}) = \frac{1}{2} \left( E(\frac{\sum X_i}{n}) + E(\frac{\sum Y_j}{m}) \right) = \frac{1}{2} \left( \frac{1}{n} E(\sum X_i) + \frac{1}{m} E(\sum Y_j) \right) = \frac{1}{2}(\mu + \mu) = \mu \therefore$   
unbiased

for  $\mu_2$ :  $E(\frac{\sum X_i + \sum Y_j}{n+m}) = \frac{1}{n+m} (E(\sum X_i) + E(\sum Y_j)) = \frac{1}{n+m} (n\mu + m\mu) = \frac{n+m}{n+m}(\mu) = \mu \therefore$  unbiased

b)

for  $\mu_1$ :  $Var(\frac{\sum X_i}{n} + \frac{\sum Y_j}{m}) = \frac{1}{2} \left( \frac{1}{n^2} Var(\sum X_i) + \frac{1}{m^2} Var(\sum Y_j) \right) = \frac{1}{2} \left( \frac{1}{n^2} n\sigma^2 + \frac{1}{m^2} m\sigma^2 \right) = \frac{1}{2} \left( \frac{\sigma^2}{n} + \frac{\sigma^2}{m} \right) =$   
 $\frac{1}{2} \left( \frac{m\sigma^2}{nm} + \frac{n\sigma^2}{nm} \right) = \frac{\sigma^2}{2nm} (n+m)$

for  $\mu_2$ :  $Var(\frac{\sum X_i + \sum Y_j}{n+m}) = \frac{1}{n+m} (Var(\sum X_i) + Var(\sum Y_j)) = \frac{1}{n+m} (n\sigma^2 + m\sigma^2) = \sigma^2$

c)

Relative Efficiency:  $\frac{MSE(\hat{\mu}_1)}{MSE(\hat{\mu}_2)}$

$$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) - Bias^2 \therefore Relative Efficiency = \frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\frac{\sigma^2(n+m)}{2nm}}{\sigma^2} = \frac{n+m}{2nm}$$

d)

Based on relative efficiency being less than 1,  $\hat{\mu}_1$  is the better estimator.