# Homework1 Solutions

Fangyi Peng

**1.(a)**

the sample mean is:

$$x = {}_{i=1}^{n} x_i/n = \frac{29.36}{12} \;\; 2.45$$

the sample variance is:

$$s^2 = \frac{{}_{i=1}^{n}(x_i \;\; x)^2}{n \;\; 1} = \frac{3.27}{11} \;\; 0.297$$

the sample standard deviation is:

$$s = s^2 \;\; 0.545$$

**1.(b)**

Steps:

- sort data: the data is sorted as follows:

$$1.75, 1.84, 1.91, 1.92, 2.12, 2.35, 2.53, 2.62, 2.83, 3.09, 3.15, 3.25$$

- calculate the five summary numbers:
  min=1.75,
  the $25^{th}$ percentile($Q_1$, the data which is less than $Q_1$ accounts for 25 %)=1.915(in fact any number between 1.91 and 1.92 is fine),
  median= $\frac{2.35 + 2.53}{2} = 2.44$,
  the $75^{th}$ percentile($Q_3$)=2.96,
  max=3.25.

- calculate the InterQuartile Range(IQR):

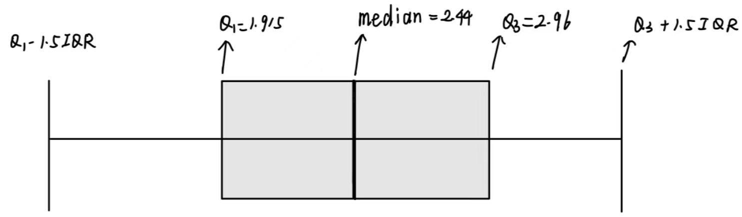
$$IQR = Q_3 \;\; Q_1 = 2.96 \;\; 1.915 = 1.045$$

- outline:

$$\text{lower outline} = Q_1 \;\; 1.5 \text{ Œ } IQR = 1.915 \;\; 1.5 \text{ Œ } 1.045 = 1.915 \;\; 1.5675 = 0.3475$$

$$\text{upper outline} = Q_3 \;\; 1.5 \text{ Œ } IQR = 2.96 + 1.5 \text{ Œ } 1.045 = 2.96 + 1.5675 = 4.5275$$

From the min and max we know all data is between lower outline and upper outline , so there is no outlier.The boxplot is drawn as follows:




**2.Part1**

read data:

```
smiles=read.table("smiles.txt")
names(smiles)=c("groups","scores")
attach(smiles)
```

Construct histograms and stem-and-leaf plots for each of the four categories:

```
tapply(scores,groups,stem)
```

```
##
##   The decimal point is at the |
##
##   2 | 55
##   3 | 0000555
##   4 | 0055
##   5 | 00555
##   6 | 000005555
##   7 | 05
##   8 | 0000
##   9 | 0
##
##
##   The decimal point is at the |
##
##   2 | 55
##   3 | 000555555
##   4 | 005555
##   5 | 00000055
##   6 | 055
##   7 | 0005
##   8 | 5
```

```
##   9 | 0
##
##
##   The decimal point is at the |
##
##   2 | 5
##   3 | 005555
##   4 | 0000000555
##   5 | 0000555555
##   6 | 005
##   7 | 5
##   8 | 000
##
##
##   The decimal point is at the |
##
##   2 | 00055555
##   3 | 000055
##   4 | 00000555555
##   5 | 05
##   6 | 000055
##   7 |
##   8 | 0
```

```
## $false
## NULL
##
## $felt
## NULL
##
## $miserable
## NULL
##
## $neutral
## NULL
```

Comments:

```
splitgroup=split(scores,groups) # splits the data by group
attach(splitgroup)
par(mfrow=c(2,2))
hist(false,main="Histogram of false smile")
hist(felt,main="Histogram of felt smile")
```

```
hist(miserable,main="Histogram of miserable smile")
hist(neutral,main="Histogram of neutral smile")
```


**2.Part2**

Obtain the 5-numerical summaries and the corresponding boxplots for all four categories:

```
summary(false)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.500   3.625   5.500   5.368   6.500   9.000
```

```
var(false)
```

```
## [1] 3.338012
```

```
sd(false)
```

```
## [1] 1.827023
```

repeat on felt smile:

```
summary(felt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##    2.500    3.500    4.750    4.912    5.875    9.000
```

```
var(felt)
```

```
## [1] 2.825312
```

```
sd(felt)
```

```
## [1] 1.680866
```

repeat on miserable smile:

```
summary(miserable)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.500   4.000   4.750   4.912   5.500   8.000
```

```
var(miserable)
```

```
## [1] 2.113191
```

```
sd(miserable)
```

```
## [1] 1.453682
```

repeat on neural smile:

```
summary(neutral)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.000   3.000   4.000   4.118   4.875   8.000
```

```
var(neutral)
```

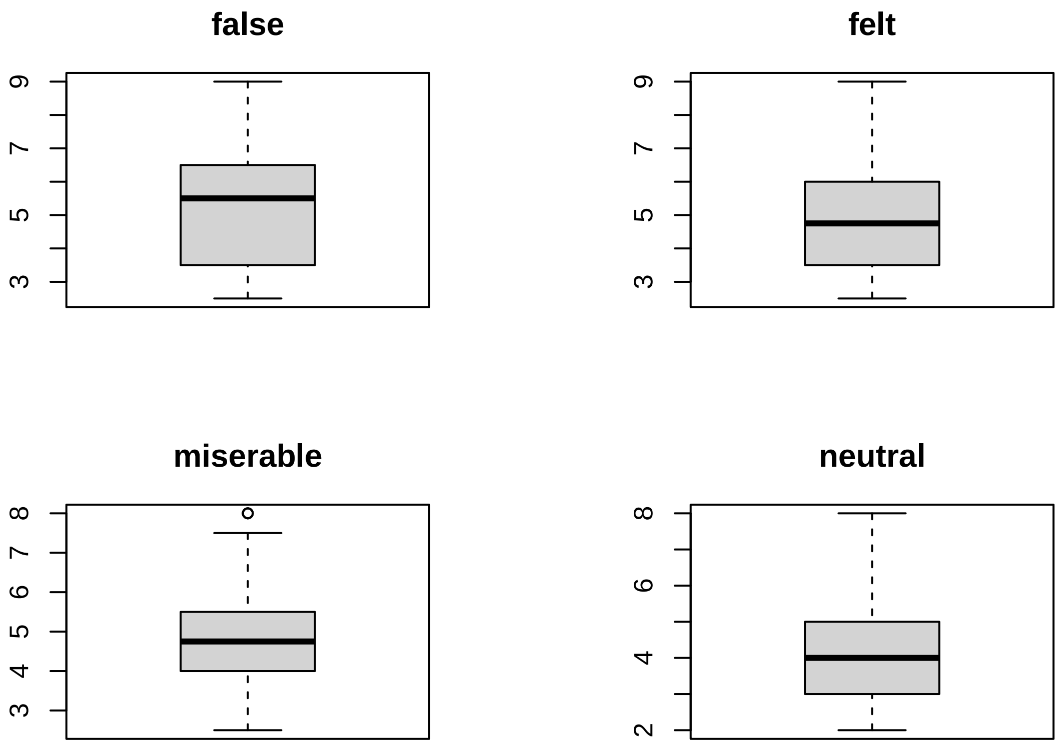
```
## [1] 2.319073
```

```
sd(neutral)
```

```
## [1] 1.52285
```

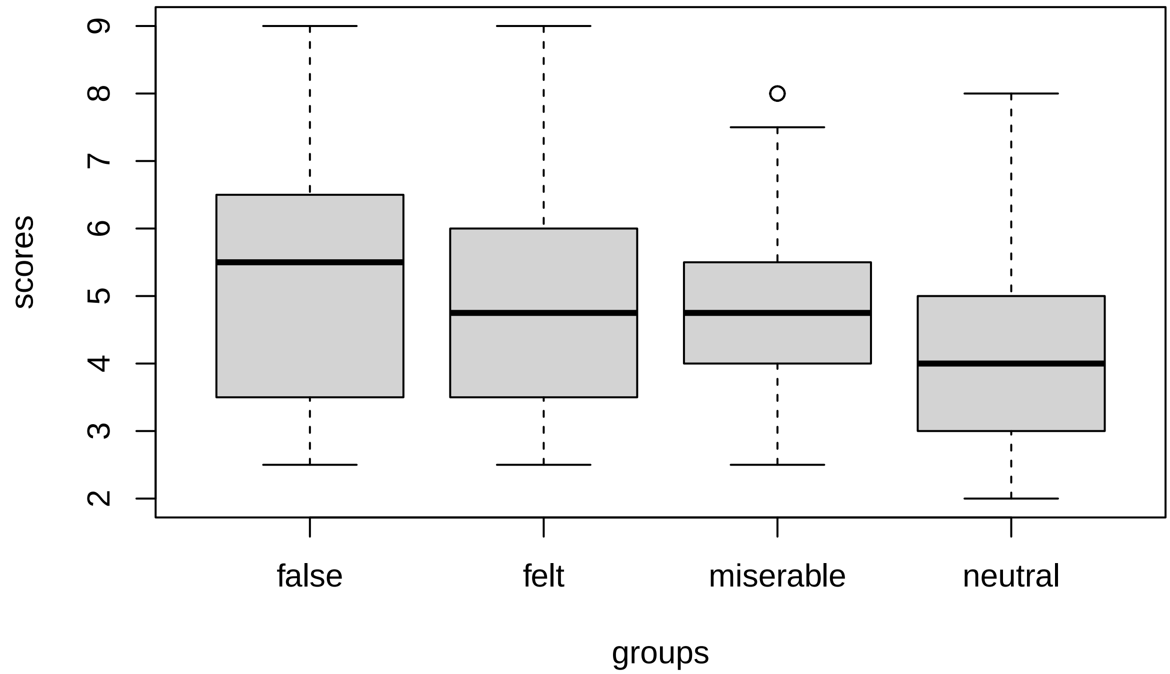
The box plot:

```
par(mfrow=c(2,2),mar = c(3, 4, 3, 4) + 0.1)
boxplot(false)
title("false")
boxplot(felt)
title("felt")
boxplot(miserable)
title("miserable")
boxplot(neutral)
title("neutral")
```




Alternatively:

```
par(mfrow=c(1,1))
boxplot(scores~groups)
```




**2.Part3**

Summarize your findings for the data analysis in the context of the problem:

**3.**

$$p_1 = \frac{x_1}{n_1} = 150/200 = 0.75, p_2 = \frac{x_2}{n_2} = 185/250 = 0.74$$

According to CLT,

$$\frac{x_1}{n_1} \;\; N\left(p_1, \frac{p_1(1 \;\; p_1)}{n_1}\right) \quad \frac{x_2}{n_2} \;\; N\left(p_2, \frac{p_2(1 \;\; p_2)}{n_2}\right)$$

Then

$$\frac{x_1}{n_1} \;\; \frac{x_2}{n_2} \;\; N\left(p_1 \;\; p_2, \frac{p_1(1 \;\; p_1)}{n_1} + \frac{p_2(1 \;\; p_2)}{n_2}\right)$$

$$\frac{x_1}{n_1} \;\; \frac{x_1}{n_2} \;\; N\left(0.01, 1.707 \text{ Œ } 10^3\right)$$

**4.(a)**

$E(x_1 \;\; x_2) = E(x_1) \;\; E(x_2) = u_1 \;\; u_2$, thus $x_1 \;\; x_2$ is the unbiased estimator of $u_1 \;\; u_2$. $\mathrm{Var}(x_1 \;\; x_2) = \mathrm{Var}(x_1) +$

$\mathrm{Var}\,(x_2\ \ )$ (since the two population is assumed independent) with CLT, $x_1\ \ N\,({}_1, \frac{{}^2_1}{n_1})$, $\quad x_2\ \ N\,({}_2, \frac{{}^2_2}{n_2})$, then

$$\mathrm{var}\,(x_1\ \ x_2) = \frac{{}^2_1}{n_1} + \frac{{}^2_2}{n_2}, \quad sd\,(x_1\ \ x_2) = \frac{{}^2_1}{n_2} + \frac{{}^2_2}{n_2}.$$

(Pay attention: the estimator is a statistics which shouldn't involve params.)

**4.(b)**

$$\begin{aligned} E\,(x_1^2\ \ x_2^2) &= E\,(x_1^2)\ \ E\,(x_2^2) = \mathrm{Var}\,(x_1) + E\,(x_1)^2\ \ (\mathrm{Var}\,(x_2) + E\,(x_2))^2 \\ &= \frac{{}^2_1}{n_1} + u_1^2\ \ \frac{{}^2_2}{n_2}\ \ u_2^2 \\ \text{Bias}\ \ &= \frac{{}^2_1}{n_1}\ \ \frac{{}^2_2}{n_2}.\ \text{When } n_1, n_2\ \ ,\ \text{Bias}\ \ 0. \end{aligned}$$

**4.(c)**

$$\frac{(n_1\ \ 1)\,s_1^2}{{}^2}\ \ {}^2\,(n_1\ \ 1)\,, \frac{(n_2\ \ 1)\,s_2^2}{{}^2}\ \ {}^2\,(n_2\ \ 1)$$

Since the two populations are independent,

$$\begin{aligned} &\frac{(n_1\ \ 1)\,s_1^2}{{}^2} + \frac{(n_2\ \ 1)\,s_2^2}{{}^2}\ \ {}^2\,(n_1 + n_2\ \ 2)\,,\ \text{then} \\ &E\,(\frac{(n_1\ \ 1)\,s_1^2 + (n_2\ \ 1)\,s_2^2}{{}^2}) = n_1 + n_2\ \ 2 \\ &E\,(\frac{(n_1\ \ 1)\,s_1^2 + (n\ \ 1)s_2^2}{n_1 + n_2\ \ 2}) = E\,(S_p^2) = {}^2 \end{aligned}$$

**5.(a)**

For an exponential distribution, $E(x) = \frac{1}{}$. Consider the first order moment is $x$,

$$X = \frac{1}{{}_{mom}}\ \ {}_{mom} = \frac{1}{X}$$

**5.(b)**

$$\begin{aligned} &l(;x) = \log \prod_{i=1}^{n} f\,(x_i;) = \sum_{i=1}^{n} \log f\,(x_i;) = \sum_{i=1}^{n} \log e^{x_i} = n \log\ \ \sum_{i=1}^{n} x_i \\ &l(;x) = \frac{n}{}\ \ \sum_{i=1}^{n} x_i = 0\ \ {}_{mle} = \frac{1}{x}. \end{aligned}$$

**5.(c)**

$$_{\text{mom}} = _{\text{ml e}} = \frac{1}{x} = \frac{1}{18.76/6} \quad 0.32$$

**5.(d)**

$$x_i \quad \text{Exp}()((1,)) \quad T = \sum_{i=1}^{n} x_i \quad (n,) \text{ with } f_T(x) = \frac{{}^{n} x^{n1} e^{x}}{(n)}, x > 0.$$

$$E\left(\frac{1}{x}\right) = E\left(\frac{n}{T}\right) = \int_0 \frac{n}{x} f_T(x)dx = \int_0 \frac{n}{n\ 1} \underline{\frac{{}^{n1} x^{n2} e^{x}}{(n\ 1)}} dx = \frac{n}{n\ 1} \quad .$$

**6.(a)**

$$E(x) = \int_0^1 x f(x) dx = \int_0^1 x(+1)x dx = \frac{+1}{+2} x^{+2}\Big|_0^1 = \frac{+1}{+2}$$

Let $x = \frac{+1}{+2} \quad _{\text{MOM}} = \frac{1\ 2x}{x\ 1}$

**6.(b)**

$$L(;x) = \log f(x_i;) = \sum_{i=1}^{n} \log(+1)x_i = n\log(+1) + \sum_{i=1}^{n} \log x_i$$

$$l(;x) = \frac{n}{+1} + \sum_{i=1}^{n} \log x_i = 0 \quad _{mu} = \frac{n}{\sum_{i=1}^{n} \log x_i} \quad 1$$

**7.**

$$x_i \quad u(a,b), \quad f(x) = \begin{cases} \frac{1}{ba} & a \quad x \quad b, \\ 0 & \text{else} \end{cases}$$

$$u_1 = E(x) = \frac{a+b}{2}, u_2 = E(x^2) = \text{var}(x) + E(x)^2 = \frac{(b\ a)^2}{12} + \left(\frac{a+b}{2}\right)$$

$$m_1 = x, m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, u_1 = m_1 \& u_2 = m_2$$

$$m_1 = \frac{a+b}{2}, \quad m_2 = \frac{(b\ a)^2}{12} + \left(\frac{a+b}{2}\right)^2$$

$$\begin{cases} a+b = 2m_1 \\ b\ a = 12\left(m_2\ m_1^2\right) \end{cases} \qquad \begin{matrix} a = m_1 \quad 3\left(m_2\ m_1^2\right) \\ b = m_1 + 3\left(m_2\ m_1^2\right) \end{matrix}$$