MATH10282 Introduction to Statistics
Semester 2, 2019/20
Solutions to the R Coursework assignment

The data for this question are contained in the following table:

Year	Life expectancy of men	Life expectancy of women
1920	53.6	54.6
1930	58.1	61.6
1940	60.8	65.2
1950	65.6	71.1
1960	66.6	73.1
1970	67.1	74.7
1980	70.0	77.5
1990	71.8	78.8

The data are life expectancy of men and women by year.

(a) The commands

```
x=c(53.6, 58.1, 60.8, 65.6, 66.6, 67.1, 70.0, 71.8)
c(min(x),median(x),mean(x),max(x),sd(x))
```

give the following

Minimum = 53.60, Median = 66.10, Mean = 64.20, Maximum = 71.80, SD = 6.195159

[1]

(b) The commands

```
y=c(54.6, 61.6, 65.2, 71.1, 73.1, 74.7, 77.5, 78.8)
c(min(y),median(y),mean(y),max(y),sd(y))
```

give the following

Minimum = 54.60, Median = 72.10, Mean = 69.58, Maximum = 78.80, SD = 8.416947

[1]

(c) The minimum, median, mean, maximum and standard deviation all appear to be larger for women. [1]

(d) The command

```
boxplot(x,y,names=c("Men","Women"),ylab="Life expectancy")
```

gives

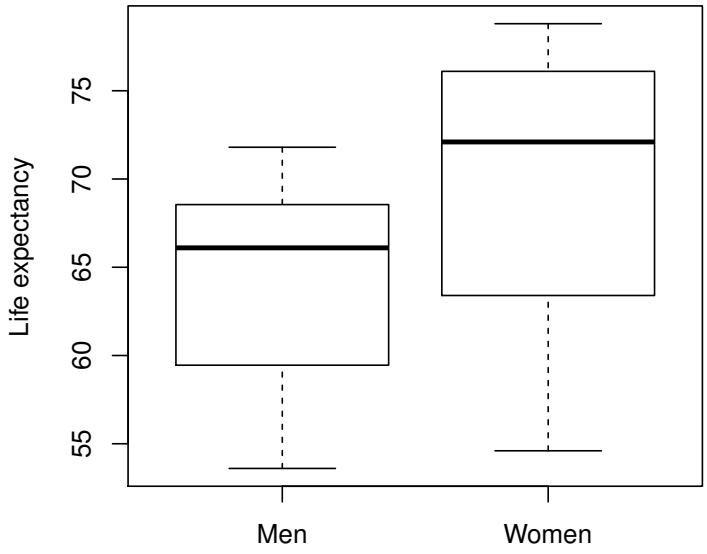

Figure 1: The boxplots of the life expectancies of men and women.

The boxplots show that the median, quartiles and the inter quartile range are larger for women. [1]

(e) The estimates of the two parameters of the normal distribution are

$$\widehat{\mu} = \frac{1}{8}\sum_{i=1}^{8} x_i = 64.20$$

and

$$\widehat{\sigma} = \sqrt{\frac{1}{8}\sum_{i=1}^{8} (x_i - \widehat{\mu})^2} = 5.795041.$$

The command

```
ks.test(x,"pnorm",64.2,5.795041)
```

gives the output

```
        One-sample Kolmogorov-Smirnov test

data:  x
D = 0.22045, p-value = 0.7571
alternative hypothesis: two-sided
```

Hence, the fit of the normal distribution is adequate. [1]

(f) The estimates of the two parameters of the normal distribution are

$$\widehat{\mu} = \frac{1}{8}\sum_{i=1}^{8} y_i = 69.575$$

and

$$\widehat{\sigma} = \sqrt{\frac{1}{8}\sum_{i=1}^{8} (y_i - \widehat{\mu})^2} = 7.873333.$$

The command

```
ks.test(y,"pnorm",69.575,7.873333)
```

gives the output

```
        One-sample Kolmogorov-Smirnov test

data:  y
D = 0.20179, p-value = 0.841
alternative hypothesis: two-sided
```

Hence, the fit of the normal distribution is adequate. [1]

(g) The following commands

```
up<-mean(x)-mean(y)+qnorm(0.975)*sqrt(var(x)/8+var(y)/8)
low<-mean(x)-mean(y)-qnorm(0.975)*sqrt(var(x)/8+var(y)/8)
```

give the 95% confidence interval as $[-12.61709, 1.867093]$. Since this interval does contain zero, the mean life expectancies of men and women are not significantly different. [2]

(h) The following commands

```
plot(year,x,xlab="Year",ylab="Life expectancy",xlim=c(1920,1990),
       ylim=c(50,80),pch="o")
par(new=TRUE)
plot(year,y,xlab="",ylab="",xlim=c(1920,1990),ylim=c(50,80),pch="+")
legend(1960,60,legend=c("Men","Women"),pch=c("o","+"))
```

give the output

Figure 2: Life expectancies of men and women versus year.

Life expectancies for women are consistently larger. Life expectancies generally increase with year. The increase appears more sharper for women. [2]

[Total 10 marks]