STAT537: Statistics for Research I: HW#1

Due on Sep. 6, 2016

Dr. Schmidhammer TR 3:40 pm - 4:40 pm

Wenqiang Feng

Contents

Problem 1

Ex 3.11

Solution. (a) **Relative frequency histogram plot:**

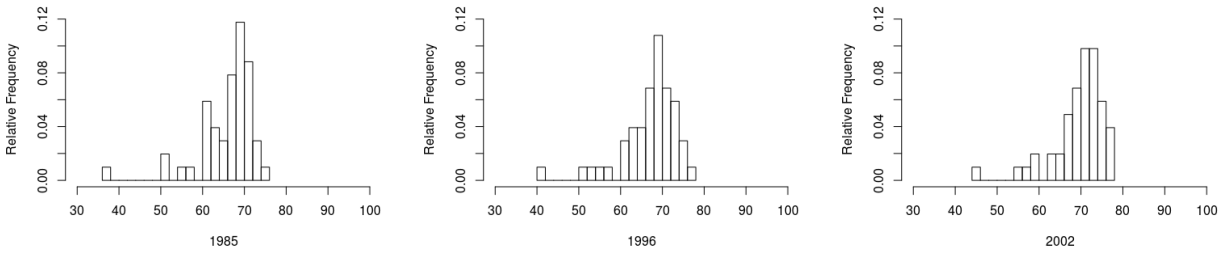

Figure 1: Relative frequency histogram plot for the homeownership data for the years 1985, 1996, and 2002.

(b) The major difference is the mean of the homeownship which increased with time elapsed from 1985-2002.

(c) In my opinion, the economic growth promoted the incomes of the families and then more people have money to buy house.

```
> mean(data1985)
[1] 65.87647
> mean(data1996)
[1] 66.84314
> mean(data2002)
[1] 69.44902
```

(d) From the mean in (c), we can see that the rate increased from 65.87% to 66.84% in the first 10 years and the rate increase from 66.84% to 69.45% in second 6 years. Although, the rate in increasing, it is still a slow increase. So, the Congress should use the information in these plots for writing tax laws that allow major tax deductions for homeownership.

□

Problem 2

Ex 3.12

Solution. • **stem-and-leaf plots for 1985:**

```
The decimal point is 1 digit(s) to the right of the |

3 | 7
4 |
4 |
5 | 014
5 | 7
6 | 1111122344
```

```
6 | 5667777888888899999
7 | 0000000011122234
7 | 6
```

- **stem-and-leaf plots for 1996:**

```
The decimal point is 1 digit(s) to the right of the |

4 | 0
4 |
5 | 13
5 | 57
6 | 1222333
6 | 5555777778888888999999
7 | 000112233333344
7 | 57
```

- **stem-and-leaf plots for 2002:**

```
The decimal point is 1 digit(s) to the right of the |

4 | 4
4 |
5 |
5 | 578
6 | 034
6 | 666777789999
7 | 000000000222222333444444
7 | 55566777
```

□

Problem 3

Ex 3.13

Solution. The stem-and-leaf plots and histograms indicate that the plots are unimodal, left skewed and asymmetrical . □

Problem 4

Ex 3.36

Solution. (a) The boxplots in Figure.2 indicate that distributions for each of the three years are all asysmmetric and left skewed.

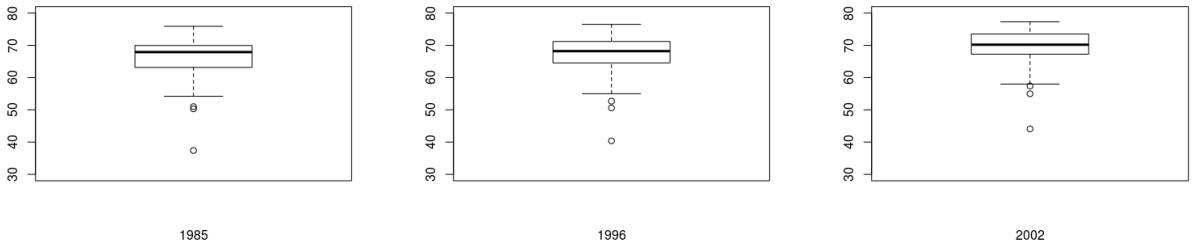

Figure 2: Boxplots for the years 1985, 1996, and 2002. .

(b) Compared the descriptions given in part (a) to the descriptions given in Exercise 3.11, they consistent with each other:asysmmetric and left skewed.

□

Problem 5

Ex 3.37

Solution. (a) The mean and median for the three years of data can be found in Table.1. The median value is most appropriate for these data sets, since our data is left skewed.

Table 1: The mean and median for the three years of data

	1985	1996	2002
mean (%)	65.87647	66.84314	69.44902
median(%)	67.9	68.2	70.2
std(%)	6.734407	6.688027	6.162901

(b) From the standard deviation of the data in Table.1, we may conclude that the degree of variability in rates over the three years has decreased from 1985 to 2002.

□

Problem 6

Ex 3.38

Solution. (a) The boxplots in Figure.3 indicate that median hownownership rate continuously increased with the time elapsed from 1985 to 2002.

(b) The boxplots in Figure.3 indicate that variation of hownownership rate continuously decreased with the time elapsed from 1985 to 2002.

(c) Yes. Since there exist outliers in bottom of the boxoplots. More precisely: state 9, 33, 12 are extremely low in homeownership rate in 1985; state 9, 12, 33 are extremely low in homeownership rate in 1996; state 9, 33, 12 are extremely low in homeownership rate in 2002.

(d) No, since no outliers at the top of the boxplots.

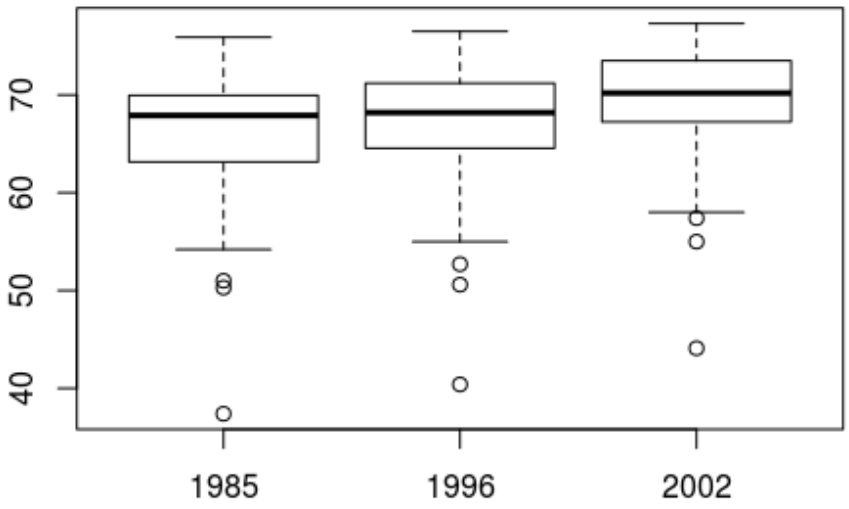

Figure 3: Boxplots for the years 1985, 1996, and 2002. .

□

Problem 7

Ex 3.76

Solution. (a) The mean, median, and standard deviation can be found in Table.2.

Table 2: the mean, median, and standard deviation

mean	median	std
61667.95	13956.5	117539.3

(b) The 25th ,50th , and 75th percentiles can be found in Table.3.

Table 3: Find the 25th ,50th , and 75th percentiles.

25%	50%	75%
9967.00	13956.50	56975.75

(c) Plot the data in a boxplot and a histogram.

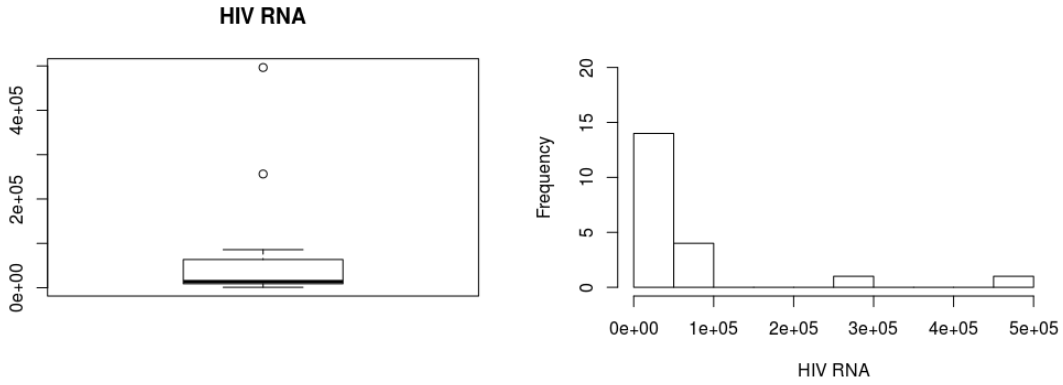

Figure 4: Boxplot and a histogram. Left: Boxplot; Right:histogram

(d) According to the boxplot and a histogram in Figure.4, the distribution of this data set is unimodal and right skewed.

□

Appendix

R code for HW#1

Listing 1: R Source code for HW#1

```
rm(list = ls())
# set the path or enverionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Hw#1/data")

#install.packages("readxl") # CRAN version
library(readxl)

rawdata = read_excel("ex3-11.xlsx",sheet = 1)
attach(rawdata)

home =as.matrix(rawdata)

data1985 = home[c(Year==1985),2]
data1996 = home[c(Year==1996),2]
data2002 = home[c(Year==2002),2]

#par(mfrow=c(3,1))
hist(data1985, freq = F, right = F, ylim = c(0, 0.12),
     xlim = c(30, 100), xlab = "1985", ylab = "Relative Frequency",
     main = "", breaks = 20)
hist(data1996, freq = F, right = F, ylim = c(0, 0.12),
     xlim = c(30, 100), xlab = "1996", ylab = "Relative Frequency",
     main = "", breaks = 20)
hist(data2002, freq = F, right = F, ylim = c(0, 0.12),
     xlim = c(30, 100), xlab = "2002", ylab = "Relative Frequency",
     main = "", breaks = 20)
```

```
#
stem(data1985)
stem(data1996)
stem(data2002)


#

boxplot(data1985, xlab="1985", ylim=c(30, 80))
boxplot(data1996, xlab="1996", ylim=c(30, 80))
boxplot(data2002, xlab="2002", ylim=c(30, 80))

#
mean(data1985)
mean(data1996)
mean(data2002)
#
median(data1985)
median(data1996)
median(data2002)
#
sd(data1985)
sd(data1996)
sd(data2002)


##
y = cbind(data1985,data1996,data2002)
cnames=c("1985", "1996", "2002")
boxplot(y,names=cnames)
```

Listing 2: R Source code for HW#1

```
rm(list = ls())
# set the path or enverionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Hw#1/data")

#install.packages("readxl") # CRAN version
library(readxl)

rawdata = read_excel("ex3-76.xlsx",sheet = 1)
attach(rawdata)

mean(rawdata$HIV_RNA)
median(rawdata$HIV_RNA)
sd(rawdata$HIV_RNA)

quantile(rawdata$HIV_RNA)[2:4]

boxplot(rawdata$HIV_RNA, main="HIV RNA")
hist(rawdata$HIV_RNA, xlab="HIV RNA", main="",
     ylim=c(0,20), breaks=10)
```