> #################################

> # R session 1.2: Extra exercise #

> #################################

>

> setwd("...")

>

> # Select the file "TBAD.txt" and check the variables

> tbad<-read.table("tbad.txt",header=T,sep="\t",dec=".")

> str(tbad)

'data.frame': 186 obs. of 9 variables:

$ Specialty : Factor w/ 6 levels "fp","gp","im",..: 6 6 6 6 3 5 3 1 6 6 ...

$ Secondary.specialty : int 0 1 0 0 0 0 0 0 0 0 ...

$ Certification.level : int 1 1 2 1 1 0 0 1 1 1 ...

$ Gender : int 0 0 0 1 0 0 0 0 0 1 ...

$ Medical.school : int 0 1 1 1 1 1 1 0 0 1 ...

$ Residence : int 0 0 0 0 0 1 1 1 1 0 ...

$ Years.of.experience : int 37 35 0 21 8 0 38 22 23 0 ...

$ Total.average.costs.per.patient.per.month: num 47.6 25.9 53.7 74.1 18.2 ...

$ Total.patients.per.month : int 2108 18 430 255 12 707 71 88 ...

>

> # Secondary.specialty, Certification.level, Gender,

> # Medical.school and Residence should be factors

> # and add labels

> tbad$Secondary.specialty<-factor(tbad$Secondary.specialty,

+ levels=0:1,labels=c("no","yes"))

> tbad$Certification.level<-factor(tbad$Certification.level)

> tbad$Gender=factor(tbad$Gender,

+ levels=0:1,labels=c("man","woman"))

> tbad$Medical.school<-factor(tbad$Medical.school,

+ levels=0:1,labels=c("USA","Foreign"))

> tbad$Residence<-factor(tbad$Residence,

+ levels=0:1,labels=c("USA","Foreign"))

>

**> # Frequency tables and bar plot for Certification level**

> tab<-table(tbad$Certification.level)

> tabp<-prop.table(tab)

> ft <- rbind(tab, tabp, cumsum(tab), cumsum(tabp))

> ft <- t(ft)

> colnames(ft) <- c("freq","percent",

+ "cumul freq","cumul percent")

> round(ft, 2)

freq percent cumul freq cumul percent

0 27 0.15 27 0.15

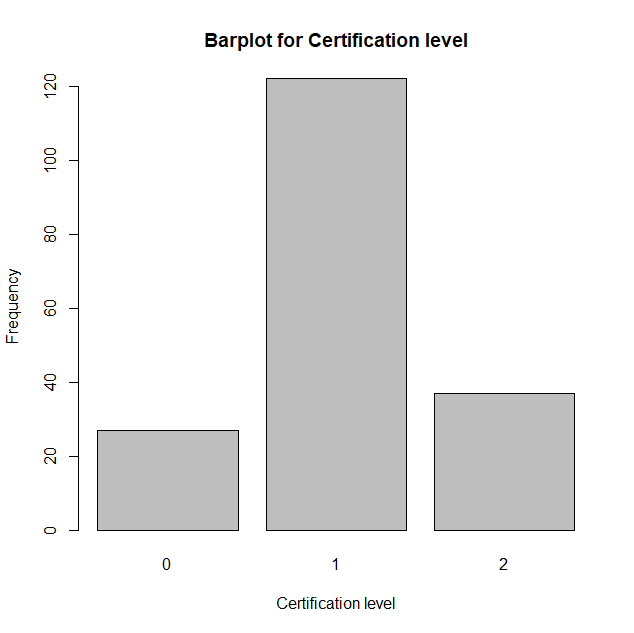
1 122 0.66 149 0.80

2 37 0.20 186 1.00

>

> barplot(tab,main="Barplot for Certification level",

+ xlab="Certification level", ylab="Frequency")



>

**> # Descriptive measures for number of patients per month**

> summary(tbad$Total.patients.per.month)

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

1.0 74.5 323.0 621.9 762.0 7424.0

> sd(tbad$Total.patients.per.month)

[1] 994.8756

>

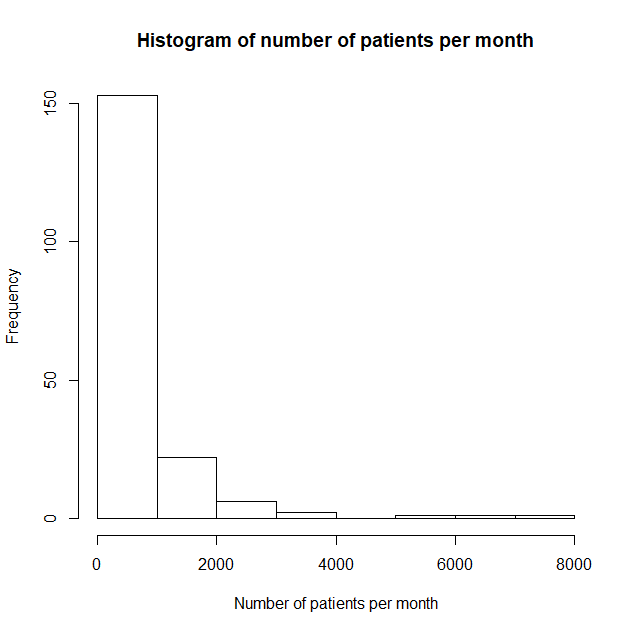
**> # Hist & boxplot for number of patients per month**

> hist(tbad$Total.patients.per.month,

+ main="Histogram of number of patients per month",

+ xlab="Number of patients per month",

+ ylab="Frequency")



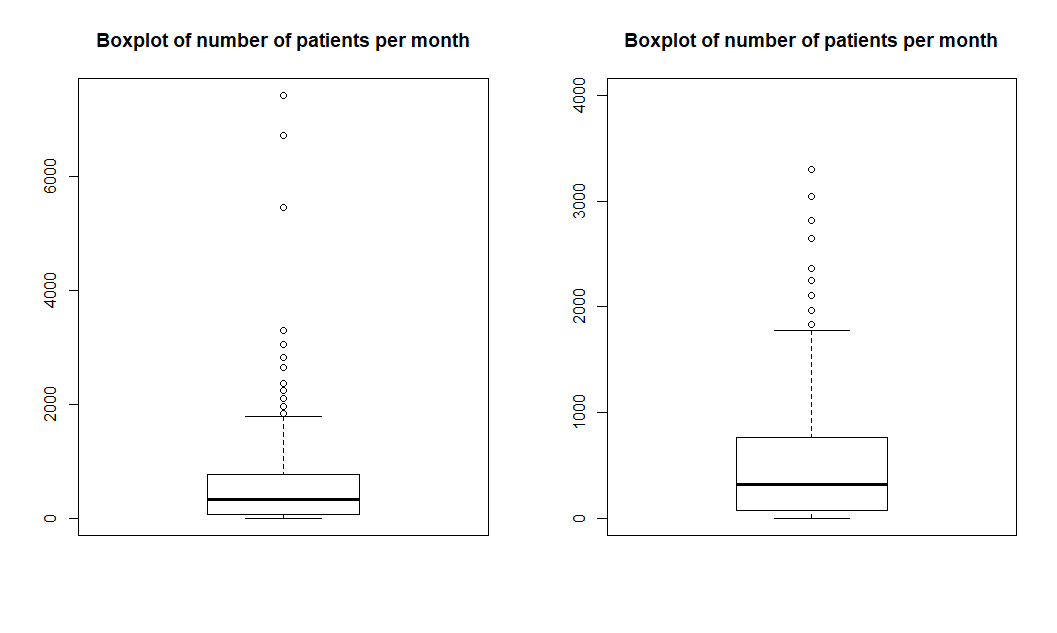
> boxplot(tbad$Total.patients.per.month,

+ main="Boxplot of number of patients per month")

> boxplot(tbad$Total.patients.per.month,

+ ylim=c(0,4000),

+ main="Boxplot of number of patients per month")



>

**> #descr & boxplot for experience by gender**

> tapply(tbad$Years.of.experience,

+ tbad$Gender, summary)

$`man`

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

0.00 8.00 15.50 17.49 28.00 47.00

$woman

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

0.00 6.00 14.00 13.83 18.75 38.00

> boxplot(tbad$Years.of.experience ~ tbad$Gender,

+ main="Boxplot of Years of experience,\n by gender")

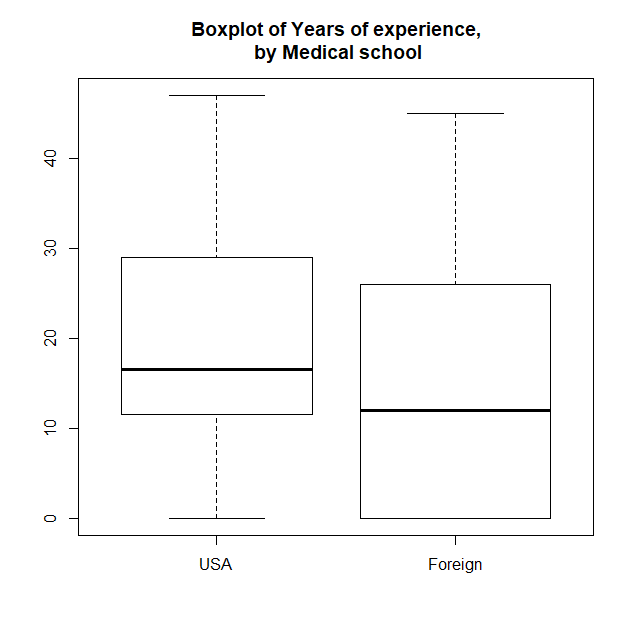


>

**> #experience by medical school**

> boxplot(tbad$Years.of.experience ~ tbad$Medical.school,

+ main="Boxplot of Years of experience,\n by Medical school")



> YoeUSA = tbad$Years.of.experience[tbad$Medical.school=="USA"]

> YoeFor = tbad$Years.of.experience[tbad$Medical.school=="Foreign"]

> var.test(YoeUSA,YoeFor)

F test to compare two variances

data: YoeUSA and YoeFor

F = 0.74047, num df = 99, denom df = 85, p-value = 0.1496

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.4886081 1.1148233

sample estimates:

ratio of variances

0.7404664

> t.test(YoeUSA,YoeFor,var.equal=T)

Two Sample t-test

data: YoeUSA and YoeFor

t = 3.4068, df = 184, p-value = 0.0008071

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

2.684002 10.069951

sample estimates:

mean of x mean of y

19.97000 13.59302

>

**> #cost less than 60, for pediatrics**

> summary(tbad$Total.average.costs.per.patient.per.month[tbad$Specialty=="pd"])

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

2.10 23.10 34.95 51.05 61.10 225.90

> length(tbad$Total.average.costs.per.patient.per.month[tbad$Specialty=="pd"])

[1] 56

> t.test(tbad$Total.average.costs.per.patient.per.month[tbad$Specialty=="pd"],

+ mu=60,alternative="less")

One Sample t-test

data: tbad$Total.average.costs.per.patient.per.month[tbad$Specialty == "pd"]

t = -1.5303, df = 55, p-value = 0.06583

alternative hypothesis: true mean is less than 60

95 percent confidence interval:

-Inf 60.83489

sample estimates:

mean of x

51.04643

>