**STAT 40001/ MA 59800 Statistical Computing Fall 2017**

**Lab-9**

1. The data “vacation” data provided in the link below describe a sample of 200 Chicago households regarding their vacation. The data includes the following variables

1. miles miles traveled per year

2. income annual income in $1000's

3. age average age of adult members of household

4. kids number of children in household

<http://www.principlesofeconometrics.com/poe4/poe4stata.htm>

1. Import the data in R (Note the format of the data)  
   > library(foreign)

> data = read.dta("C:\\Users\\wu1114\\Desktop\\vacation.dta")  
> head(data)

miles income age kids

1 902 41 26 0

2 491 31 38 3

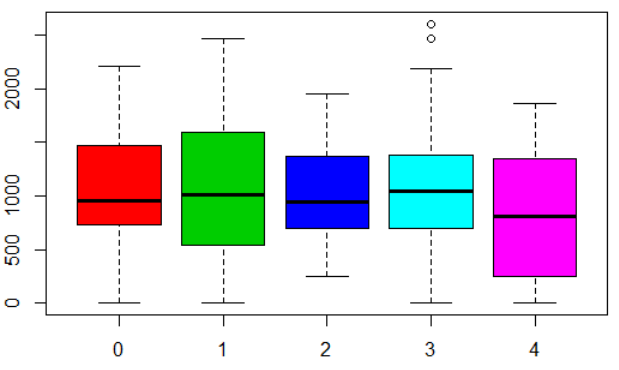
3 1841 87 40 2

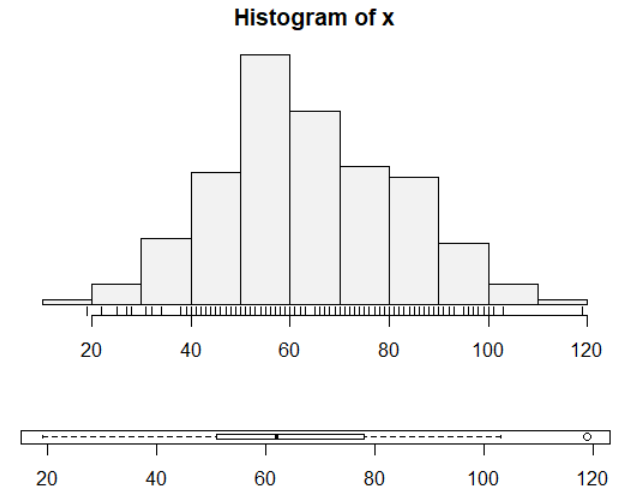
4 406 54 48 4

5 0 77 43 4

6 1899 70 55 2

1. Display the miles distribution based on the number of kinds by drawing parallel box-> attach(data)

> boxplot(miles~factor(kids),col = c(2,3,4,5,6,7))  


1. Draw histogram along with boxplot of the income data. (You will need to use UsingR packages and simple.hist.and.boxplot(your data)  
   simple.hist.and.boxplot(income)  
   
2. We wish to estimate the average number of heartbeats per minute for a certain population. The average number of heartbeats per minute for a sample of 49 subjects was found to be 90. Assume that these 49 patients constitute a random sample, and that the population is normally distributed with a standard deviation of 10.
3. Construct a 95% confidence interval for the average number of heartbeats per minute for the entire population  
   library(BSDA)  
   zsum.test(mean.x=90,sigma.x=10,n.x=49)$conf.int  
   [1] 87.20005 92.79995

attr(,"conf.level")

[1] 0.95

1. Construct a 90% confidence interval for the average number of heartbeats per minute for the entire population  
   > zsum.test(mean.x=90,sigma.x=10,n.x=49,conf.level = 0.9)$conf.int

[1] 87.65021 92.34979

attr(,"conf.level")

[1] 0.9

1. The following are the head circumferences (centimeters) at birth of 15 infants in a local hospital

33.38 32.15 33.99 34.10 33.97 34.34 33.95 33.85

34.23 32.73 33.46 34.13 34.45 34.19 34.05

Construct a 95% confidence interval for the head circumferences (centimeters) at birth of all infants born at the local hospital.  
> head = scan()

1: 33.38 32.15 33.99 34.10 33.97 34.34 33.95 33.85

9: 34.23 32.73 33.46 34.13 34.45 34.19 34.05

16:

Read 15 items

> head

[1] 33.38 32.15 33.99 34.10 33.97 34.34 33.95 33.85 34.23 32.73 33.46 34.13

[13] 34.45 34.19 34.0

> t.test(head)$conf.int

[1] 33.44895 34.14705

1. The data are from a national sample of 6000 households with a male head earning less than $15,000 annually in 1966. The data were classified into 39 demographic groups for analysis. The study was undertaken in the context of proposals for a guaranteed annual wage (negative income tax). At issue was the response of labor supply (average hours) to increasing hourly wages. The study was undertaken to estimate this response from available data.

Data are available on the DASL web site

<http://lib.stat.cmu.edu/DASL/Datafiles/wagesdat.html>

Import the Data set in R-readable format (You may first save and then import it using read.table). It appears that missing values are indicated by “\*”.

1. Create a clean dataset by replacing “\*” by NA.  
   > data = read.table("C:\\Users\\wu1114\\Desktop\\lab9.txt",header = T, na.strings = "\*")

head(data)

HRS RATE ERSP ERNO NEIN ASSET AGE DEP RACE SCHOOL

1 2157 2.905 1121 291 380 7250 38.5 2.340 32.1 10.5

2 2174 2.970 1128 301 398 7744 39.3 2.335 31.2 10.5

3 2062 2.350 1214 326 185 3068 40.1 2.851 NA 8.9

4 2111 2.511 1203 49 117 1632 22.4 1.159 27.5 11.5

5 2134 2.791 1013 594 730 12710 57.7 1.229 32.5 8.8

6 2185 3.040 1135 287 382 7706 38.6 2.602 31.4 10.7

1. What is the Average yearly earnings of spouse ($).  
   > mean(data$ERSP,na.rm = T)

[1] 1101.703