

Στατιστική Υπολογιστική

Εργασία 1

Αθανάσιος Θεοχάρης
AEM:2961

ENTER DAAAAAAAAAAAAAAAAAAAAATE

Καταχώρηση και προετοιμασία Δεδομένων

Καταχώρηση δεδομένων

Η διεύθυνση του αρχείου EXERCISE_1_WAGES.csv πρέπει να είναι σωστή.

```
EXERCISE_1_WAGES <- read.csv("~/R-course-projects/EXERCISE_1_WAGES.csv")
```

Αλλαγή δεδομένων στις κατηγορηματικές στήλες.

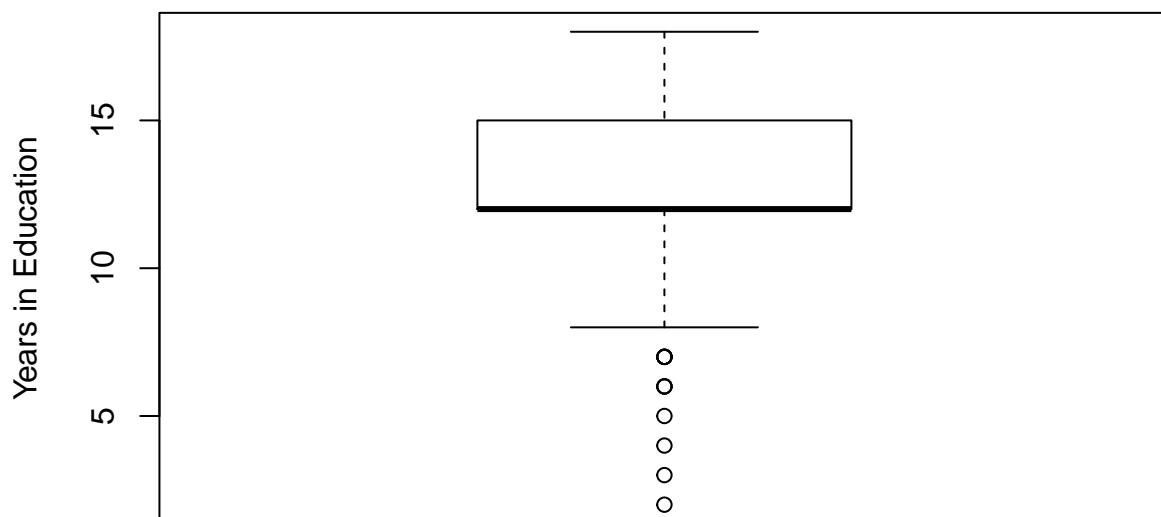
```
EXERCISE_1_WAGES$SEX = factor(EXERCISE_1_WAGES$SEX, levels=c(0,1), labels =c("Male","Female"))
EXERCISE_1_WAGES$RACE = factor(EXERCISE_1_WAGES$RACE, levels=c(1,2,3), labels =c("Other","Hispanic", "V"))
EXERCISE_1_WAGES$OCCUPATION = factor(EXERCISE_1_WAGES$OCCUPATION, levels=c(1,2,3,4,5,6), labels =c("Man"))
EXERCISE_1_WAGES$SECTOR = factor(EXERCISE_1_WAGES$SECTOR, levels=c(0,1,2), labels =c("Other","Manufact"))
EXERCISE_1_WAGES$MARR = factor(EXERCISE_1_WAGES$MARR, levels=c(0,1), labels =c("Unmarried","Married"))
```

Προβολή δεδομένων

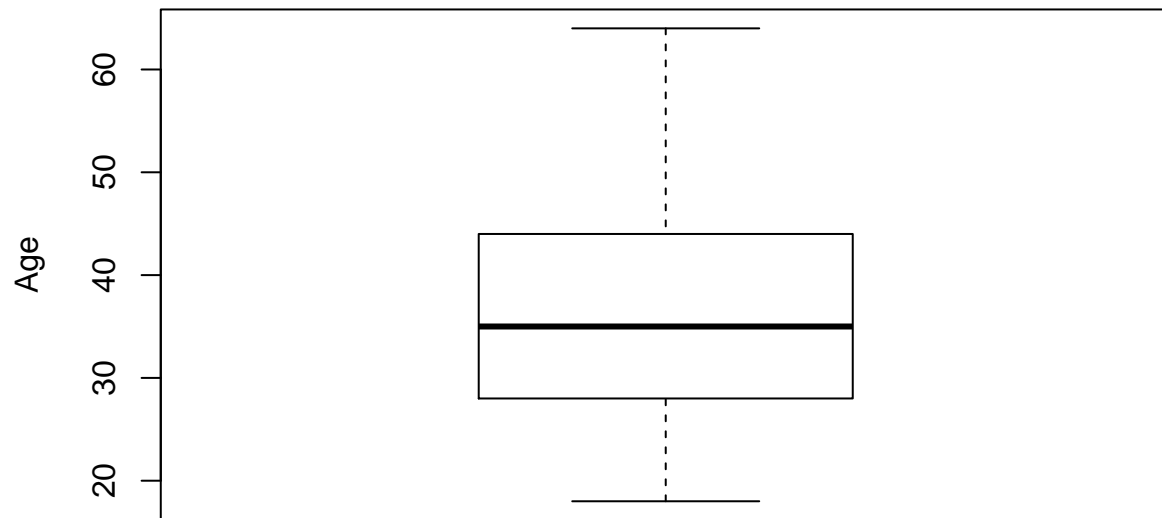
```
View(EXERCISE_1_WAGES)
```

Γραφικές παραστάσεις μεταβλητών

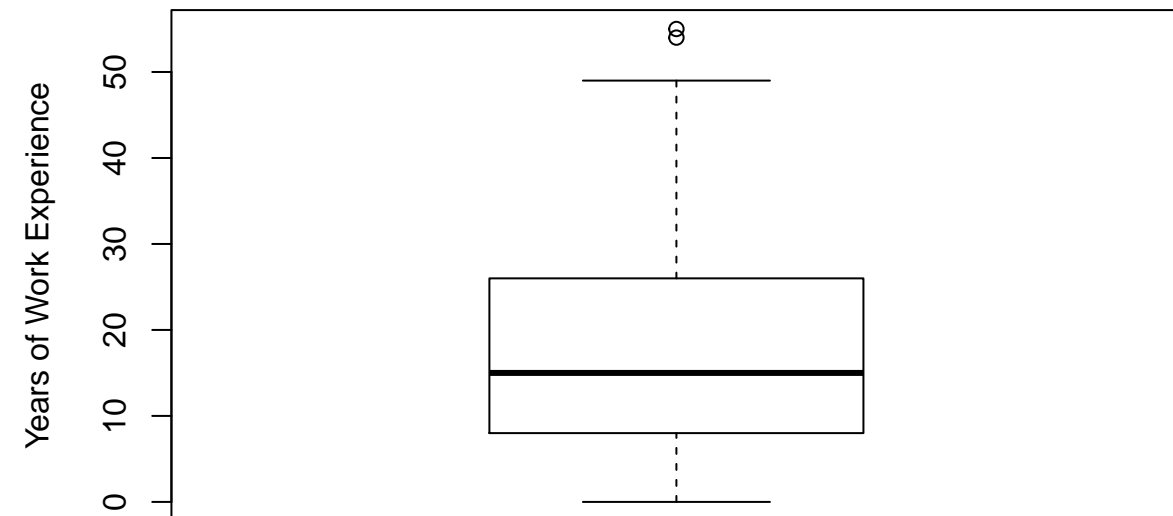
```
boxplot(EXERCISE_1_WAGES$EDUCATION, ylab="Years in Education")
```



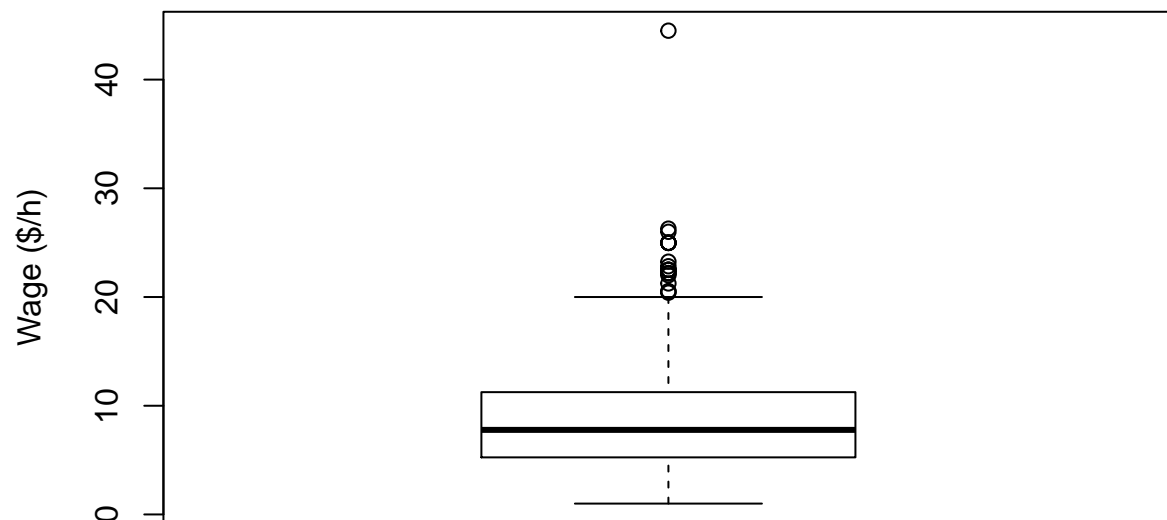
```
boxplot(EXERCISE_1_WAGES$AGE, ylab="Age")
```



```
boxplot(EXERCISE_1_WAGES$EXPERIENCE, ylab="Years of Work Experience")
```



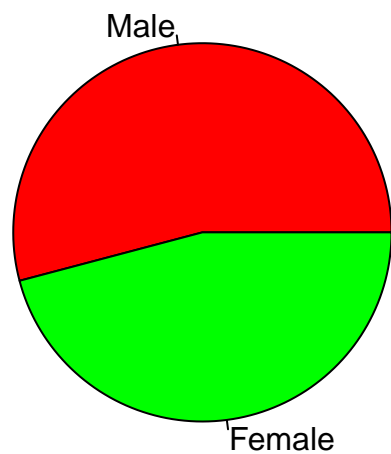
```
boxplot(EXERCISE_1_WAGES$WAGE, ylab="Wage ($/h)")
```



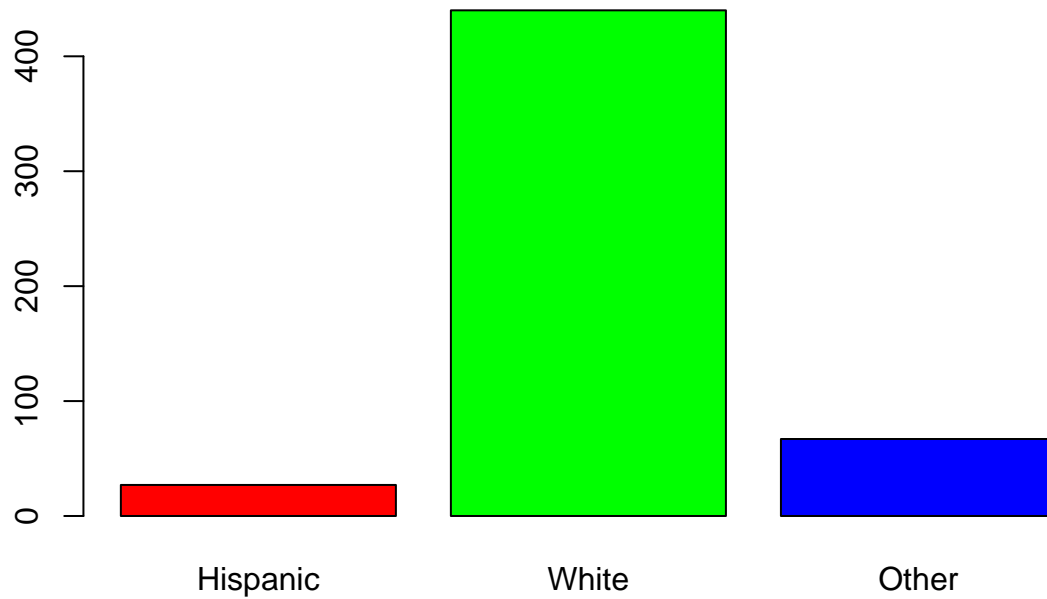
```
table = table(EXERCISE_1_WAGES$SEX)
barplot(c(table["Male"], table["Female"]), col = c("red", "green"))
```



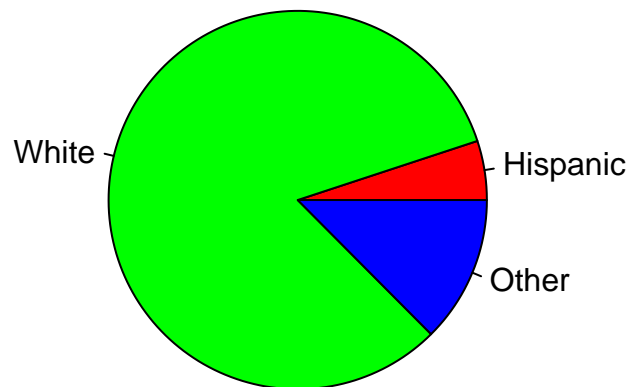
```
pie(c(table["Male"], table["Female"]), col = c("red", "green"))
```



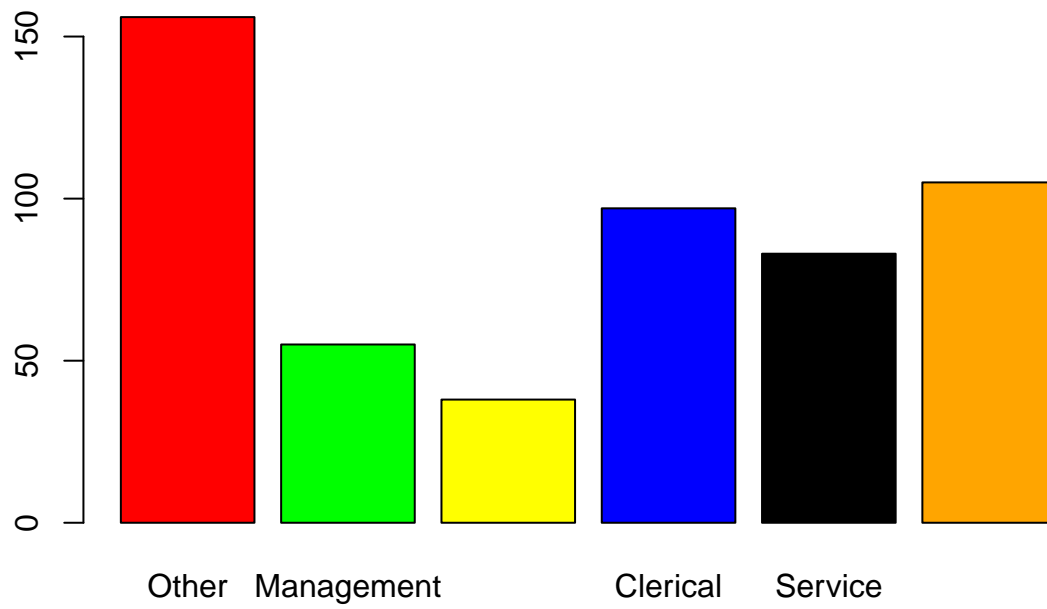
```
table = table(EXERCISE_1_WAGES$RACE)
barplot(c(table["Hispanic"], table["White"], table["Other"]), col = c("red", "green", "blue"))
```



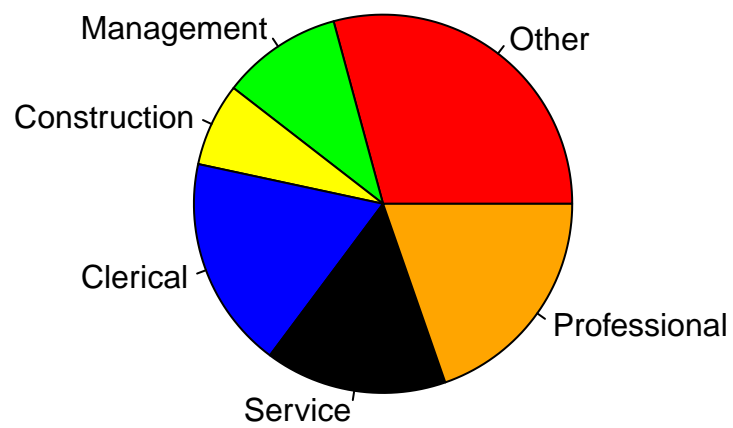
```
pie(c(table["Hispanic"], table["White"], table["Other"]), col = c("red", "green", "blue"))
```



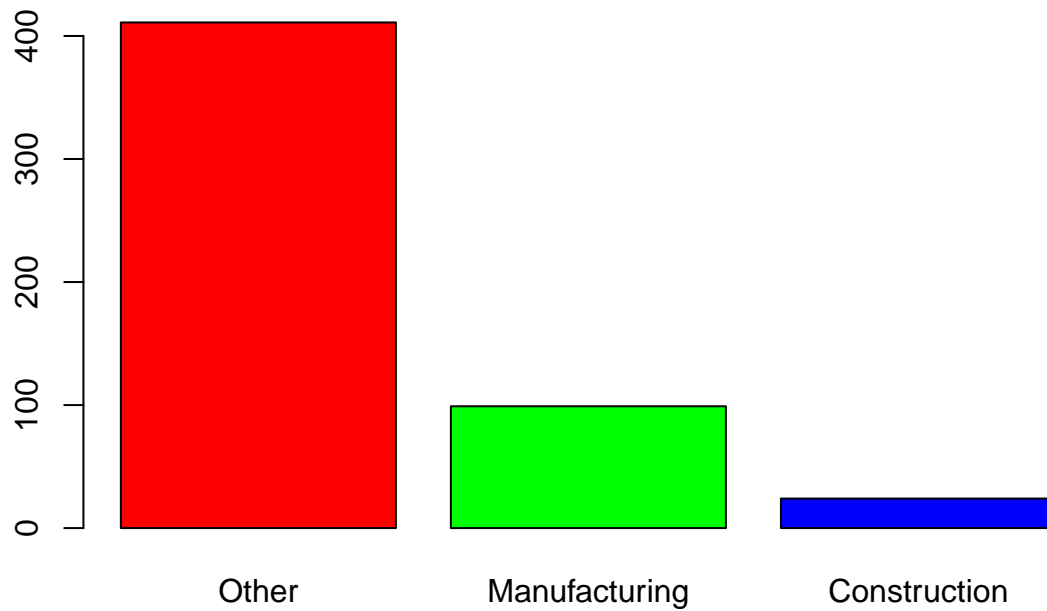
```
table = table(EXERCISE_1_WAGES$OCCUPATION)
barplot(c(table["Other"], table["Management"], table["Construction"], table["Clerical"], table["Service"])
```



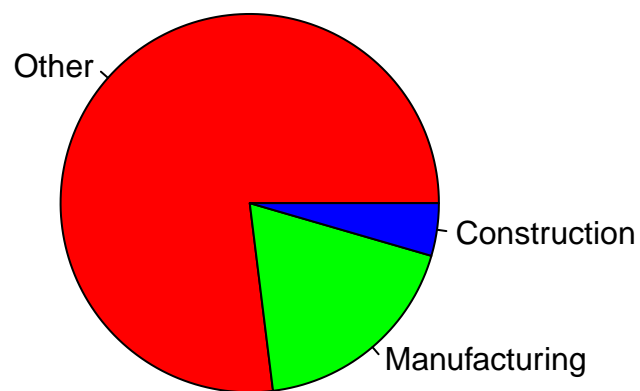
```
pie(c(table["Other"], table["Management"], table["Construction"], table["Clerical"], table["Service"],
```



```
table = table(EXERCISE_1_WAGES$SECTOR)
barplot(c(table["Other"], table["Manufacturing"], table["Construction"]), col = c("red", "green", "blue"))
```



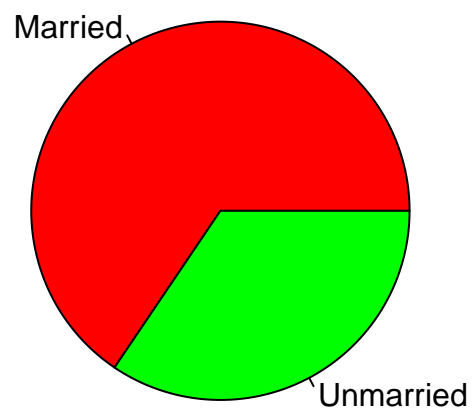
```
pie(c(table["Other"], table["Manufacturing"], table["Construction"]), col = c("red", "green", "blue"))
```



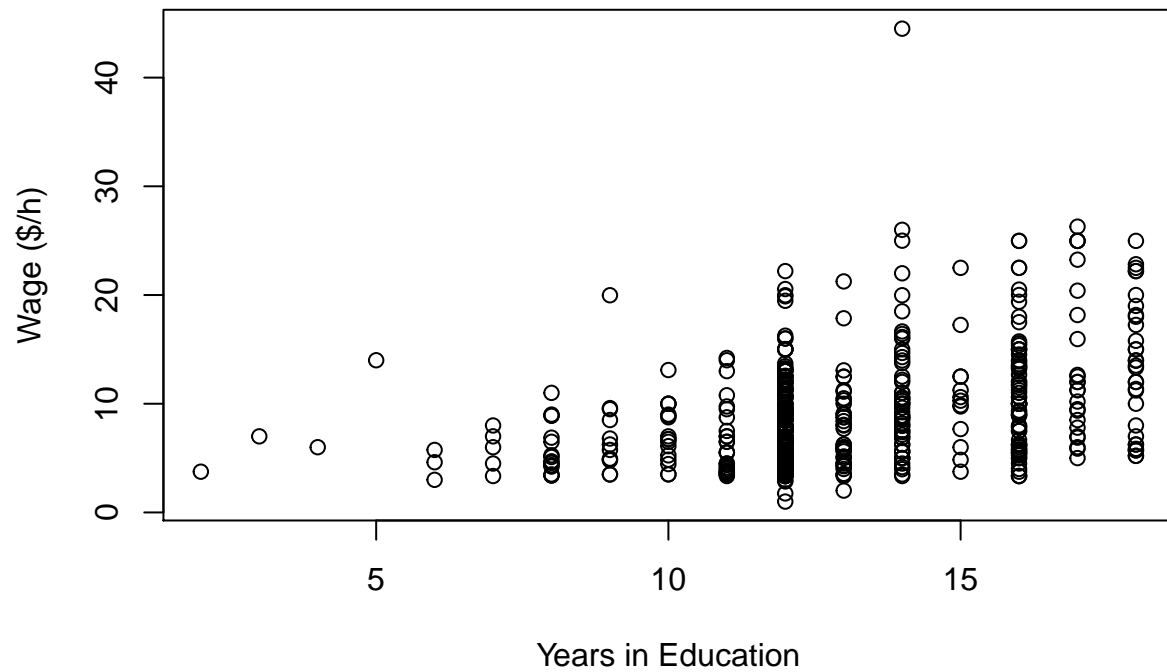
```
table = table(EXERCISE_1_WAGES$MARR)
barplot(c(table["Married"], table["Unmarried"]), col = c("red", "green"))
```



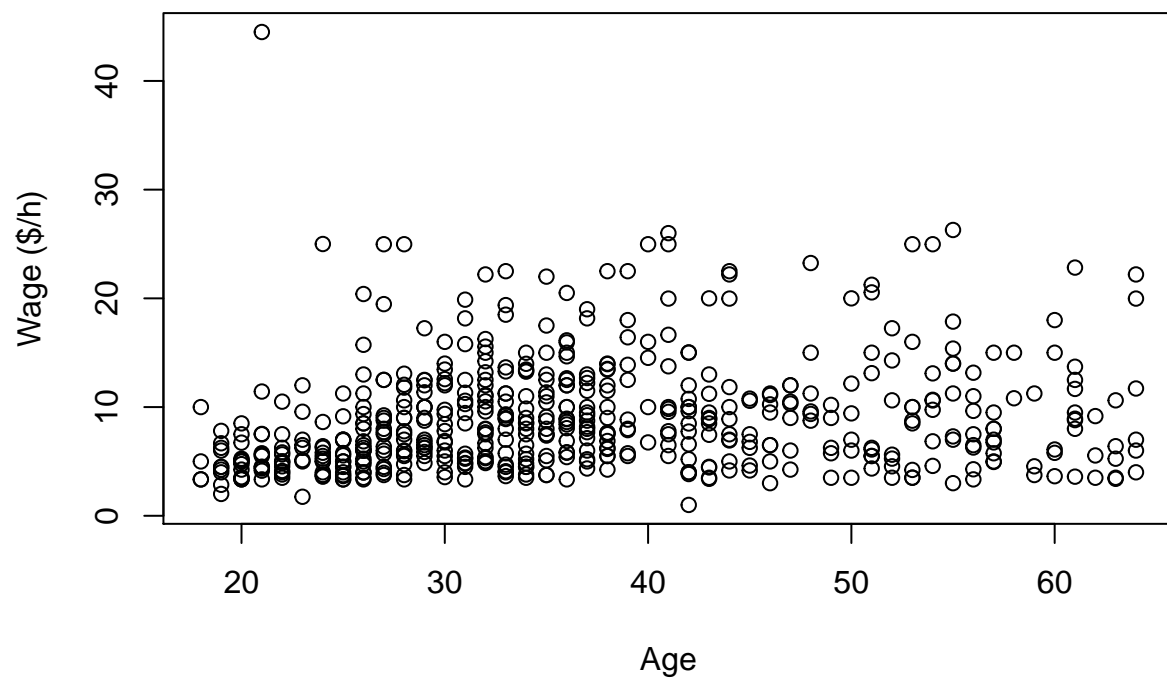
```
pie(c(table["Married"], table["Unmarried"]), col = c("red", "green"))
```



```
plot(EXERCISE_1_WAGES$EDUCATION, EXERCISE_1_WAGES$WAGE, xlab = "Years in Education", ylab = "Wage ($/h)
```



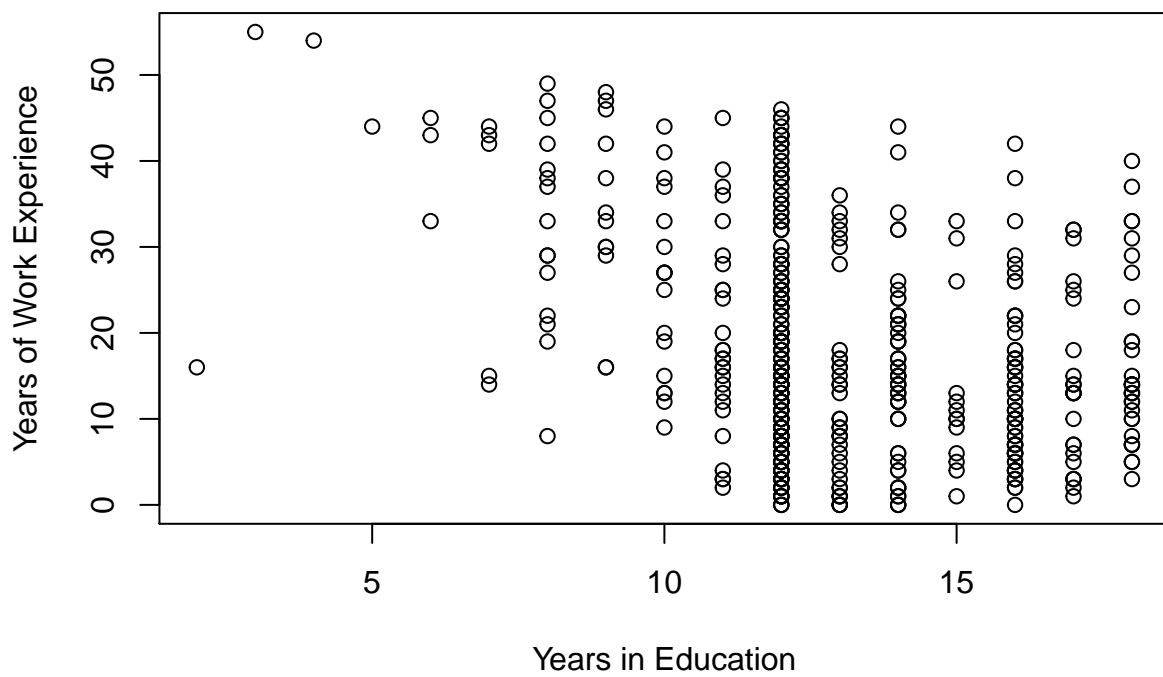
```
plot(EXERCISE_1_WAGES$AGE, EXERCISE_1_WAGES$WAGE, xlab = "Age", ylab = "Wage ($/h)")
```



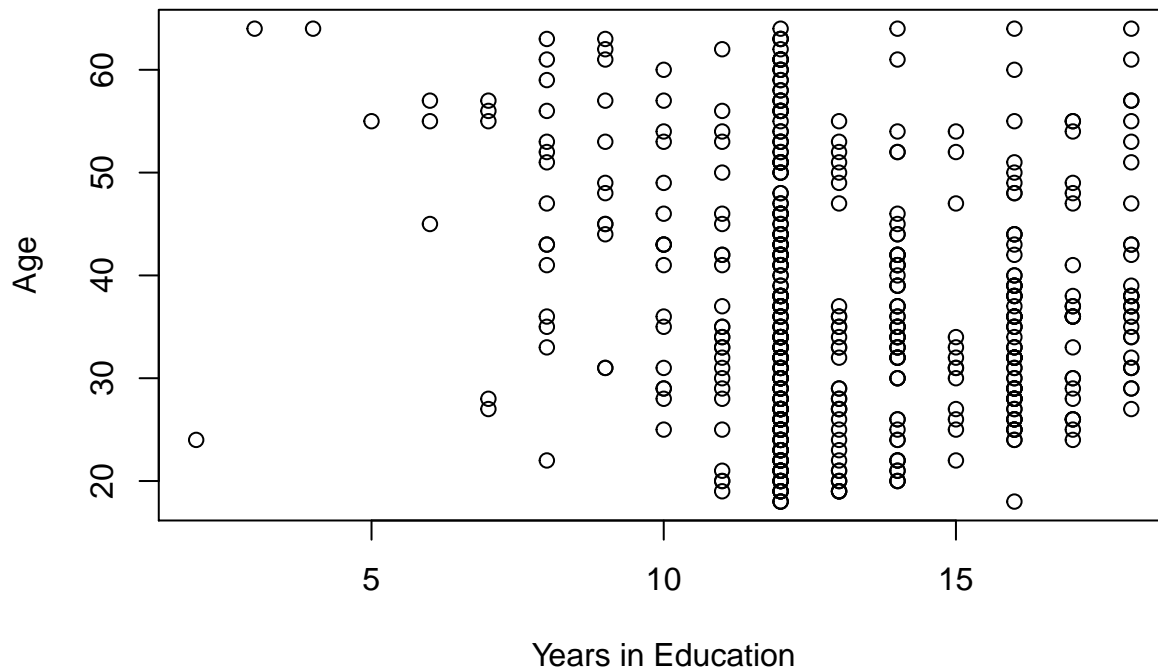
```
plot(EXERCISE_1_WAGES$EXPERIENCE, EXERCISE_1_WAGES$WAGE, xlab = "Years of Work Experience", ylab = "Wage ($/h)")
```




```
plot(EXERCISE_1_WAGES$EDUCATION, EXERCISE_1_WAGES$EXPERIENCE, xlab = "Years in Education", ylab = "Years
```



```
plot(EXERCISE_1_WAGES$EDUCATION, EXERCISE_1_WAGES$AGE, xlab = "Years in Education", ylab = "Age")
```



```
plot(EXERCISE_1_WAGES$EXPERIENCE, EXERCISE_1_WAGES$AGE, xlab = "Years of Work Experience", ylab = "Age")
```

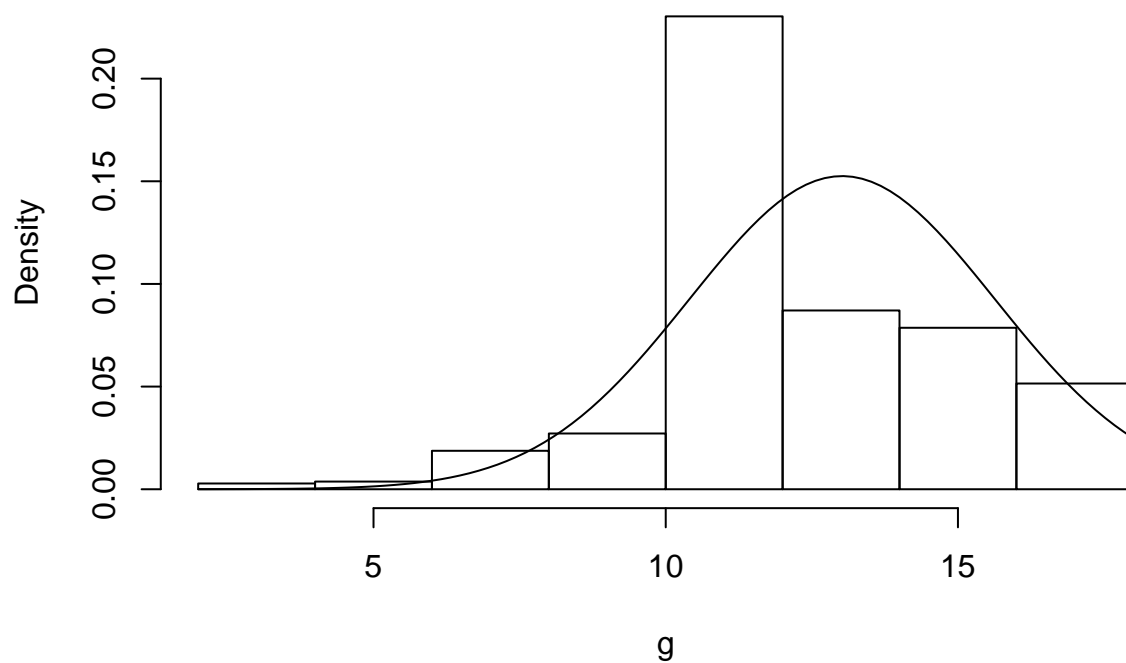


##

Έλεγχος κανονικής κατανομής

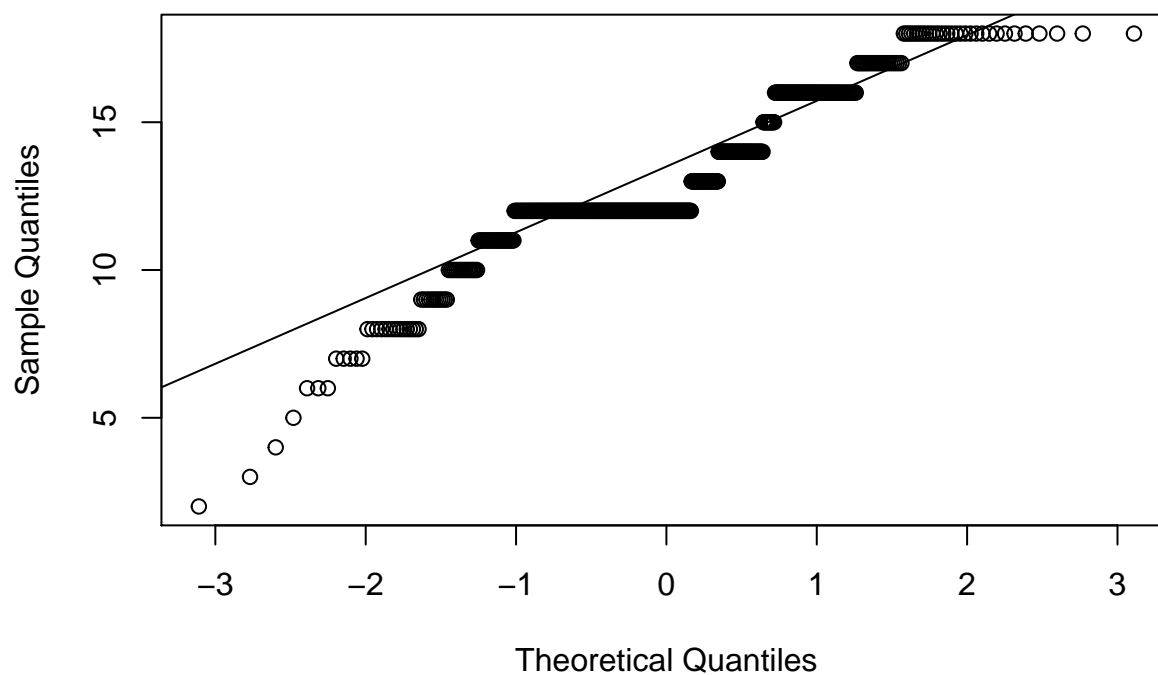
```
g<-EXERCISE_1_WAGES$EDUCATION
h<-hist(g, freq = FALSE, main = "Histogram of Years of Education")
x<-seq(min(g), max(g), by=0.02)
curve(dnorm(x, mean = mean(g), sd = sd(g)), add=TRUE)
```

Histogram of Years of Education



```
qqnorm(g, main = "QQPlot - Years of Education")
qqline(g)
```

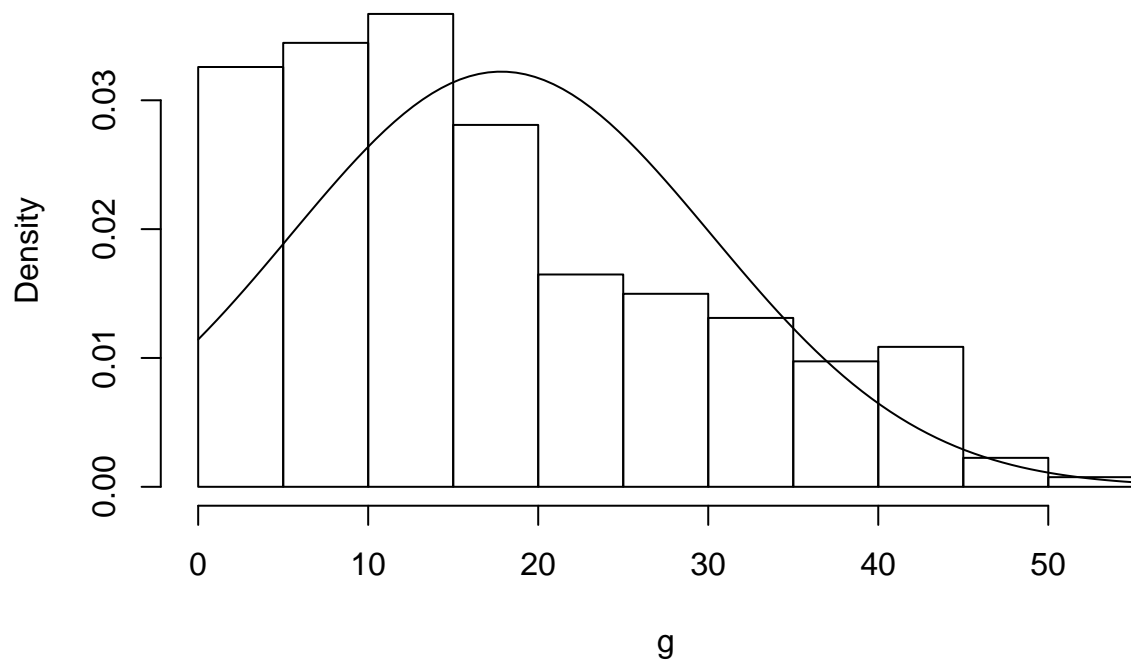
QQPlot - Years of Education



```
g<-EXERCISE_1_WAGES$EXPERIENCE
h<-hist(g, freq = FALSE, main = "Histogram of Years of Experience")
```

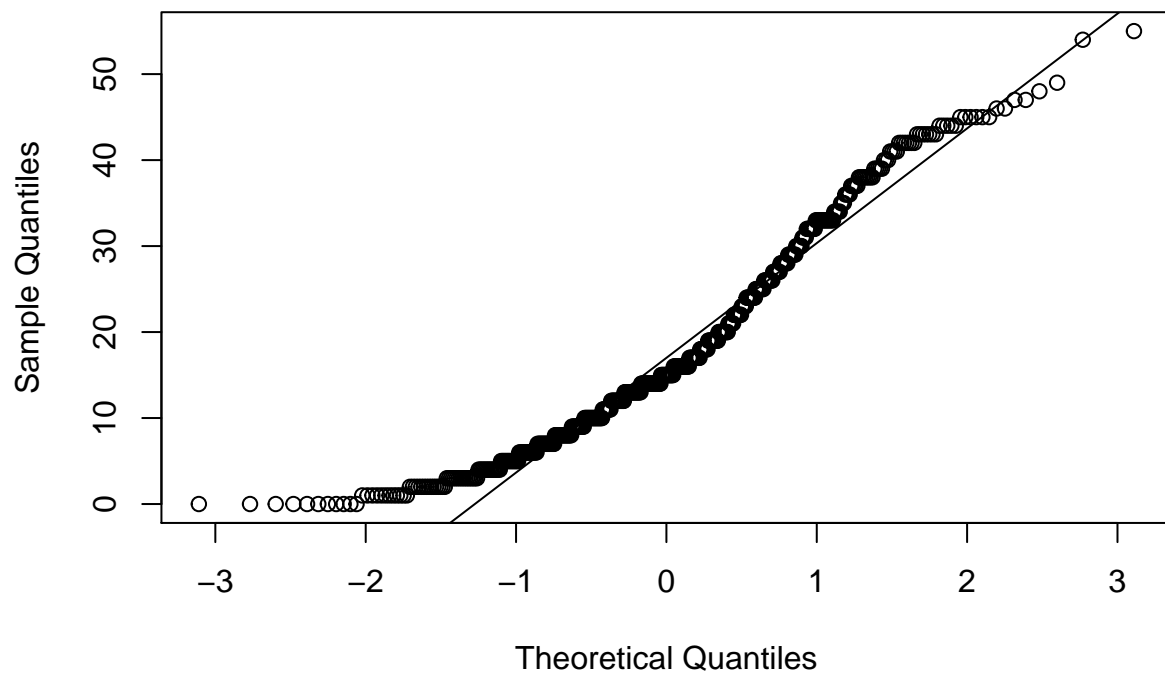
```
x<-seq(min(g), max(g), by=0.02)
curve(dnorm(x, mean = mean(g), sd = sd(g)), add=TRUE)
```

Histogram of Years of Experience



```
qqnorm(g, main = "QQPlot - Years of Experience")
qqline(g)
```

QQPlot – Years of Experience

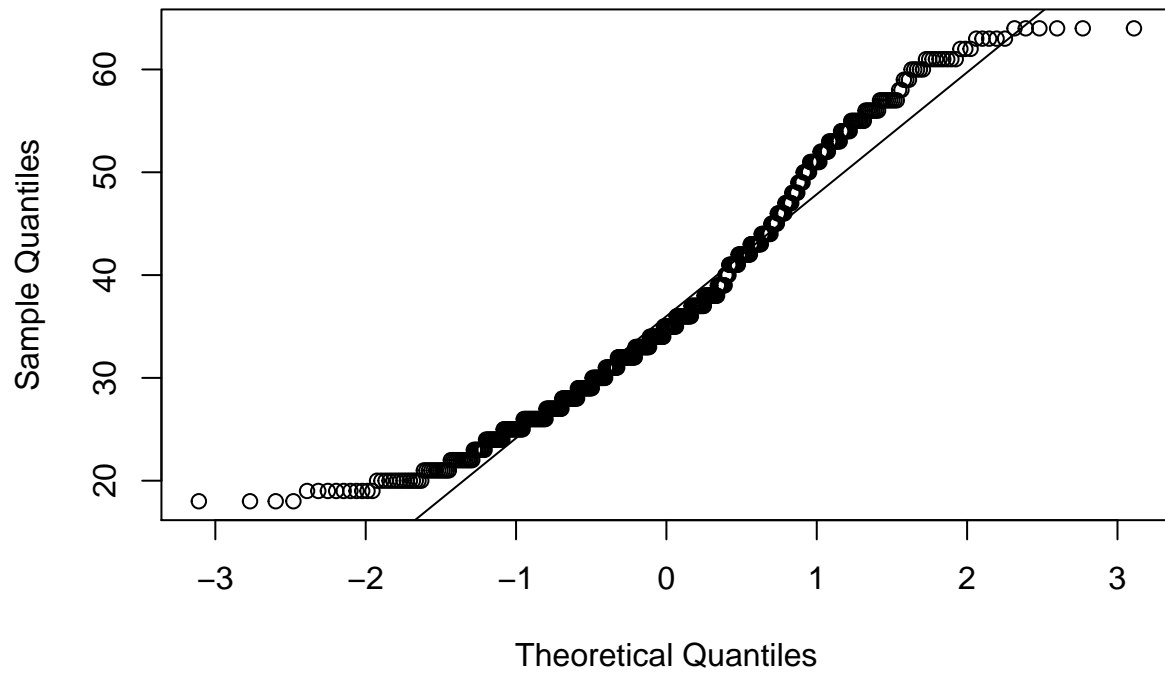


```
g<-EXERCISE_1_WAGES$AGE
h<-hist(g, freq = FALSE, main = "Histogram of Age")
x<-seq(min(g), max(g), by=0.02)
curve(dnorm(x, mean = mean(g), sd = sd(g)), add=TRUE)
```



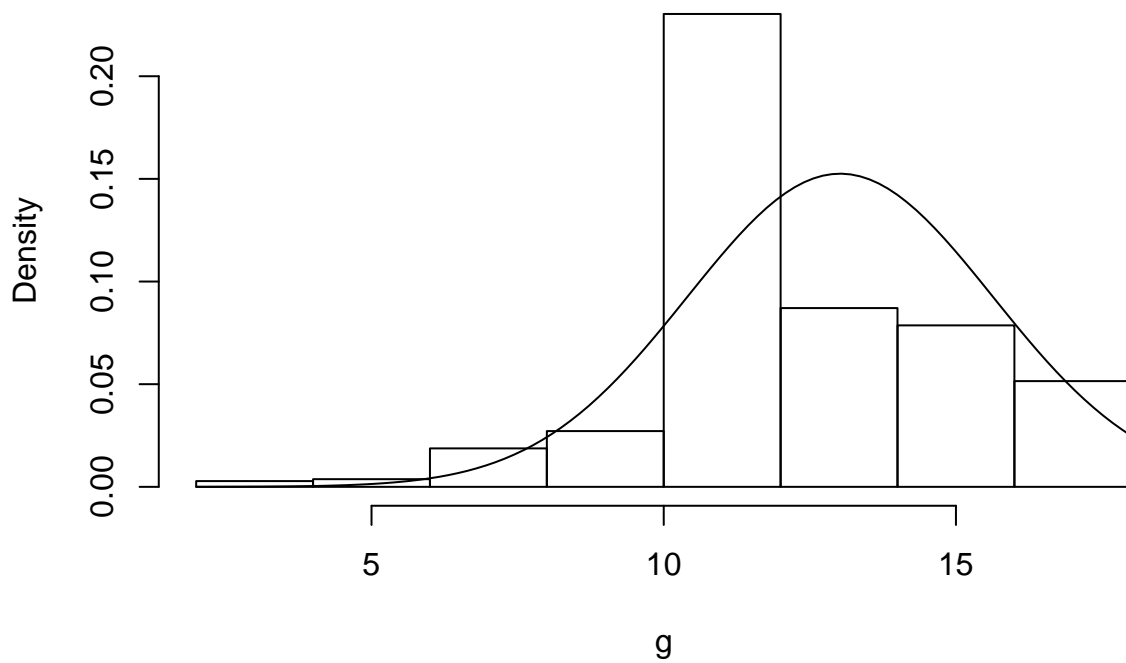
```
qqnorm(g, main = "QQPlot - Years of Age")
qqline(g)
```

QQPlot – Years of Age

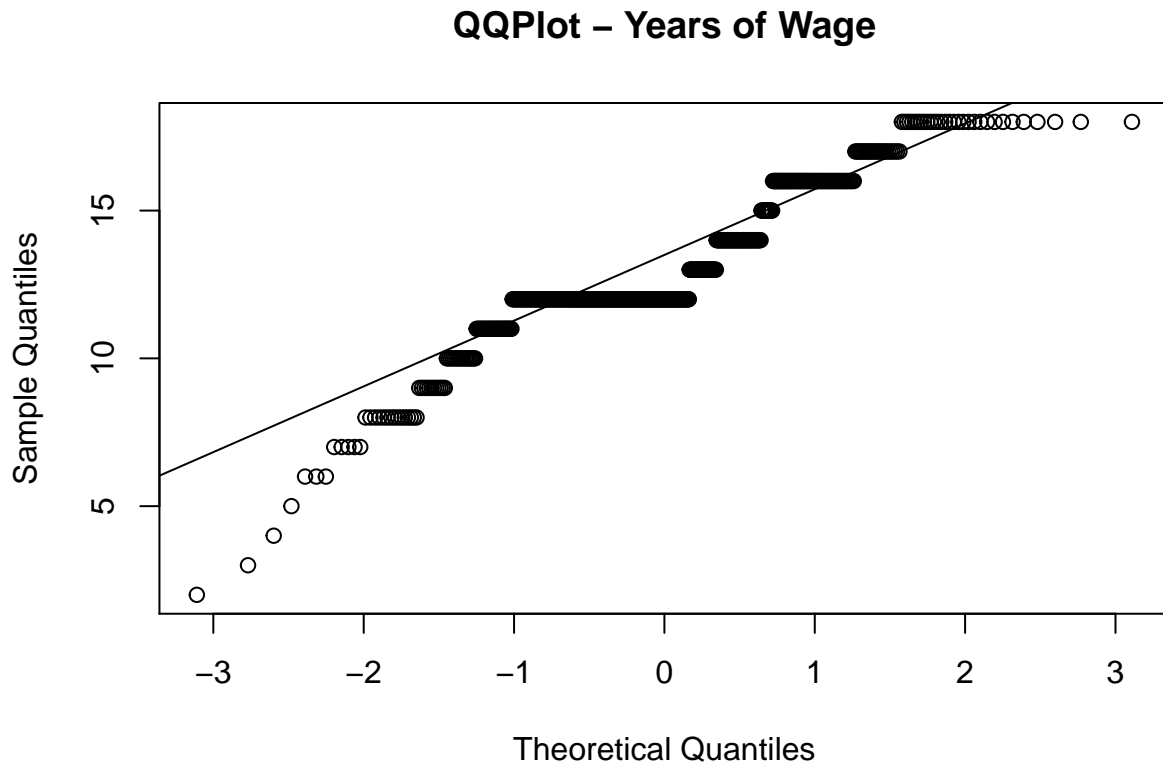


```
g<-EXERCISE_1_WAGES$EDUCATION
h<-hist(g, freq = FALSE, main = "Histogram of Wage")
x<-seq(min(g), max(g), by=0.02)
curve(dnorm(x, mean = mean(g), sd = sd(g)), add=TRUE)
```

Histogram of Wage



```
qqnorm(g, main = "QQPlot - Years of Wage")
qqline(g)
```

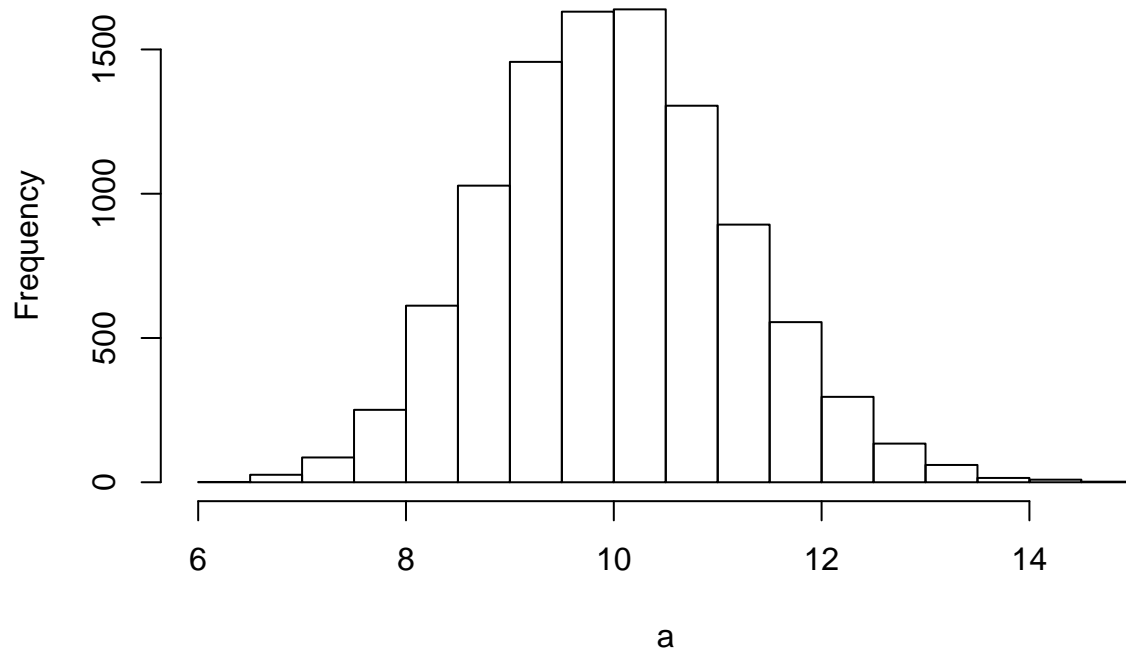


Εό-

ρρηση δ.ε. μέσου όρου και διαμέσου μισθού για άντρες

```
x = c()
k<- 1:534
for (i in k)
  if (EXERCISE_1_WAGES$SEX[i] == "Male")
    x<- c(x,EXERCISE_1_WAGES$WAGE[i])
a<-numeric(10000)
for(i in 1:10000) a[i]<-mean(sample(x, replace=T, 20))
hist(a, main = "Histogram of means")
```

Histogram of means



```
mean <- mean(a)
se <- sd(a) / sqrt(length(a))
print(mean - se)
```

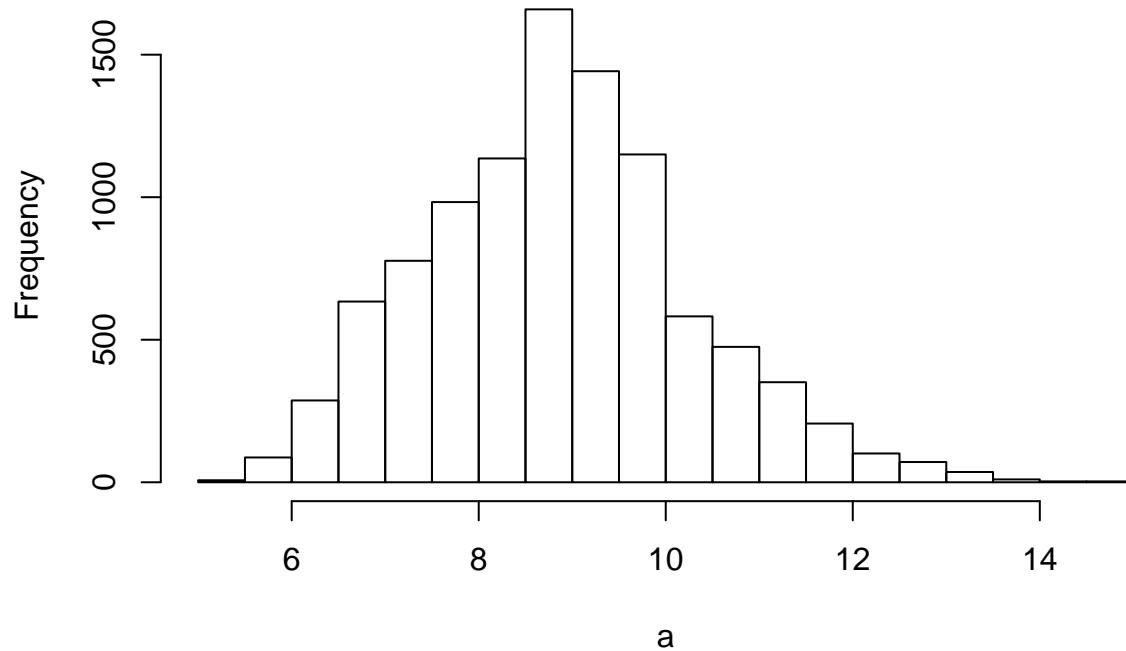
```
## [1] 9.994305
```

```
print(mean + se)
```

```
## [1] 10.01791
```

```
for(i in 1:10000) a[i]<-median(sample(x, replace=T, 20))
hist(a, , main = "Histogram of medians")
```


Histogram of medians



```
mean <- mean(a)
se <- sd(a) / sqrt(length(a))
print(mean - se)
```

```
## [1] 8.885209
```

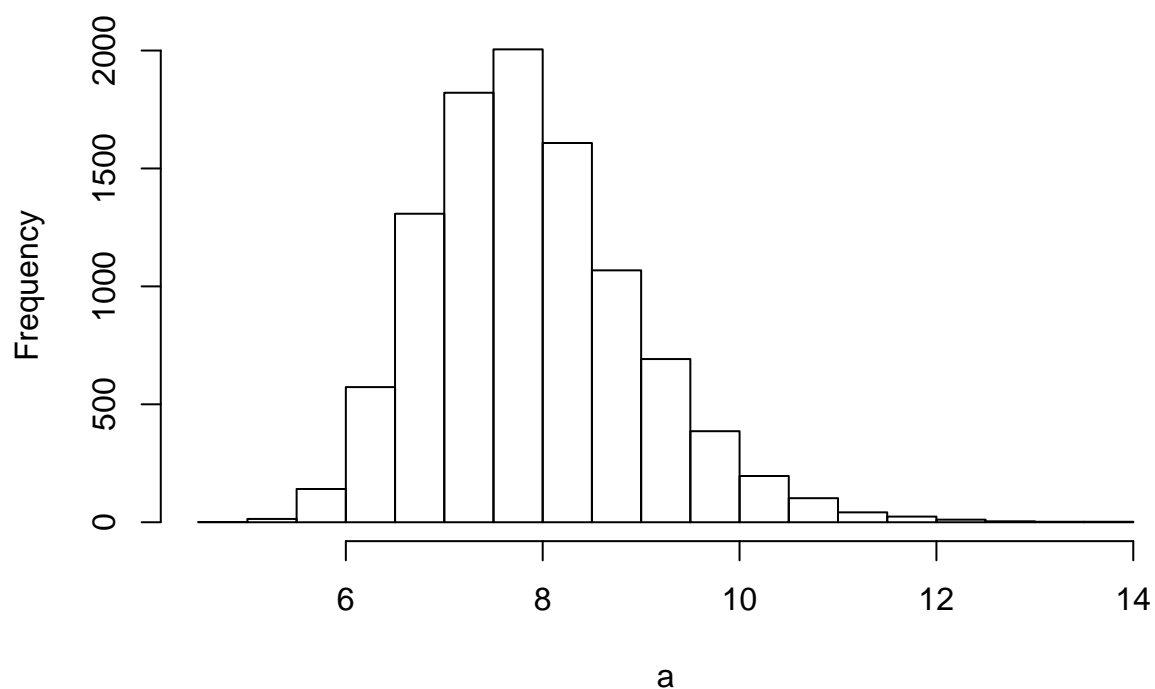
```
print(mean + se)
```

```
## [1] 8.913619
```

Εύρεση δ.ε. μέσου όρου και διαμέσου μισθού για γυναίκες

```
x = c()
k<- 1:534
for (i in k)
  if (EXERCISE_1_WAGES$SEX[i] == "Female")
    x<- c(x,EXERCISE_1_WAGES$WAGE[i])
a<-numeric(10000)
for(i in 1:10000) a[i]<-mean(sample(x, replace=T, 20))
hist(a, main = "Histogram of means")
```

Histogram of means



```
mean <- mean(a)
se <- sd(a) / sqrt(length(a))
mean - se
```

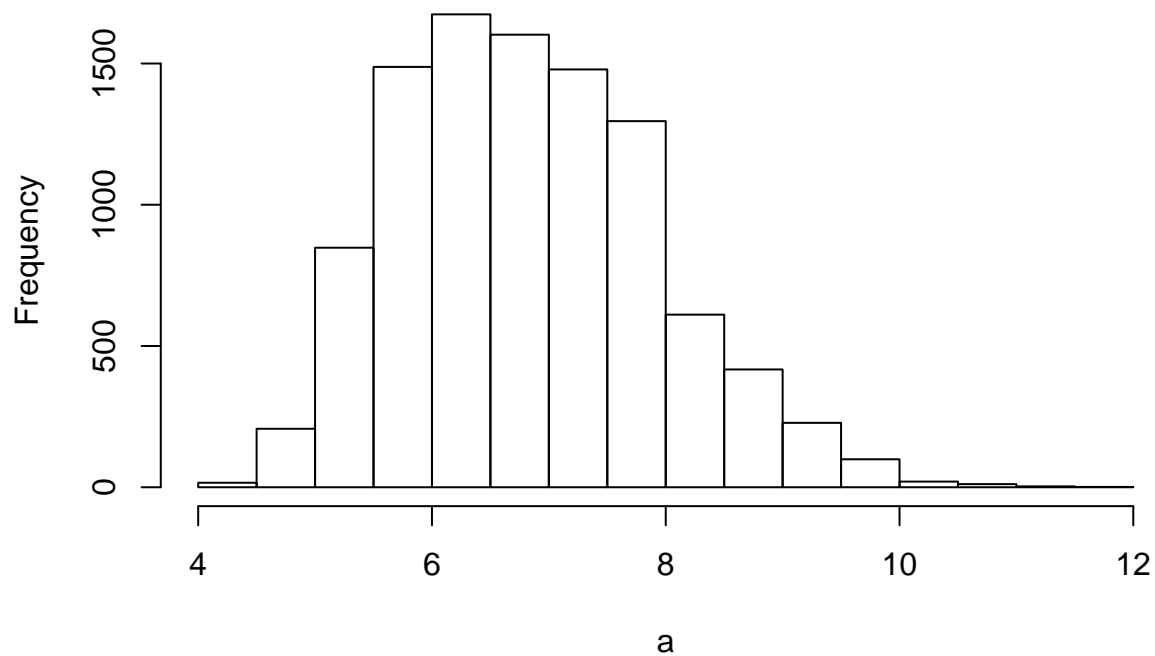
```
## [1] 7.882582
```

```
mean + se
```

```
## [1] 7.903906
```

```
for(i in 1:10000) a[i]<-median(sample(x, replace=T, 20))
hist(a, main = "Histogram of medians")
```

Histogram of medians



```
mean <- mean(a)
se <- sd(a) / sqrt(length(a))
print(mean - se)
```

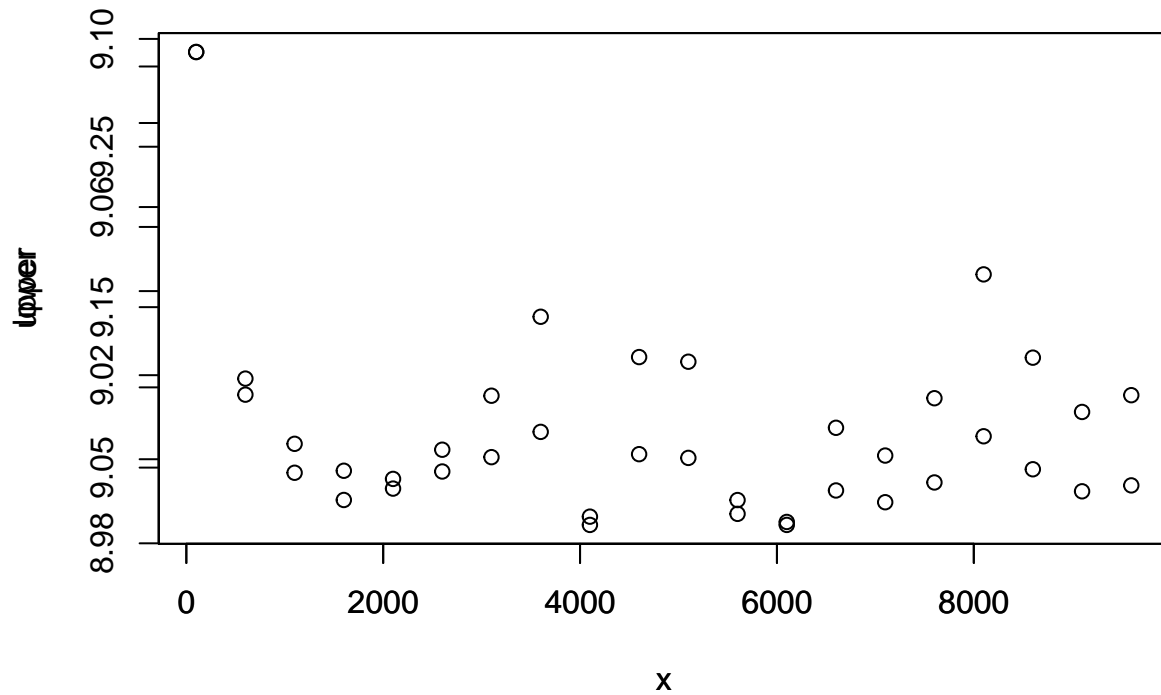
```
## [1] 6.846698
```

```
print(mean + se)
```

```
## [1] 6.868896
```

Εύρεση δ.ε. μέσου όρου μισθού

```
x <- seq(100,10000,by = 500)
lower <- c()
upper <- c()
for (j in x) {
  a<-numeric(j)
  for(i in 1:j) a[i]<-mean(sample(EXERCISE_1_WAGES$WAGE, replace=T, 20))
  mean <- mean(a)
  se <- sd(a) / sqrt(length(a))
  lower<- c(lower, mean - se)
  upper <- c(upper, mean + se)
}
plot(x, upper)
par(new = TRUE)
plot(x, lower)
```

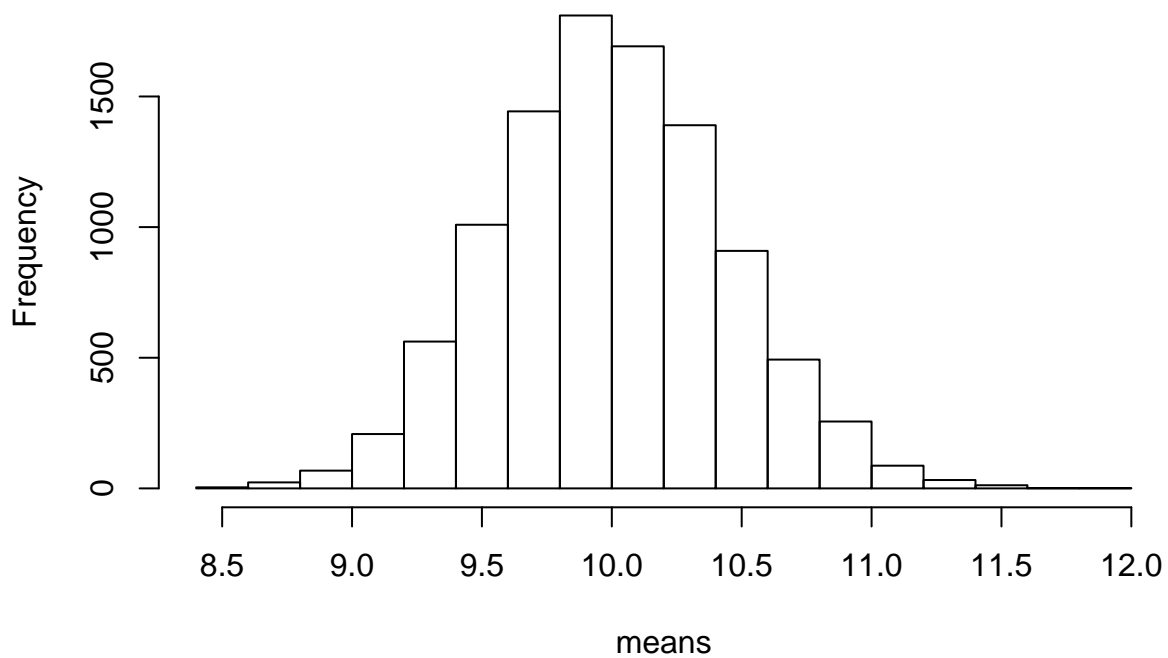


βεβαίωση K00

Επι-

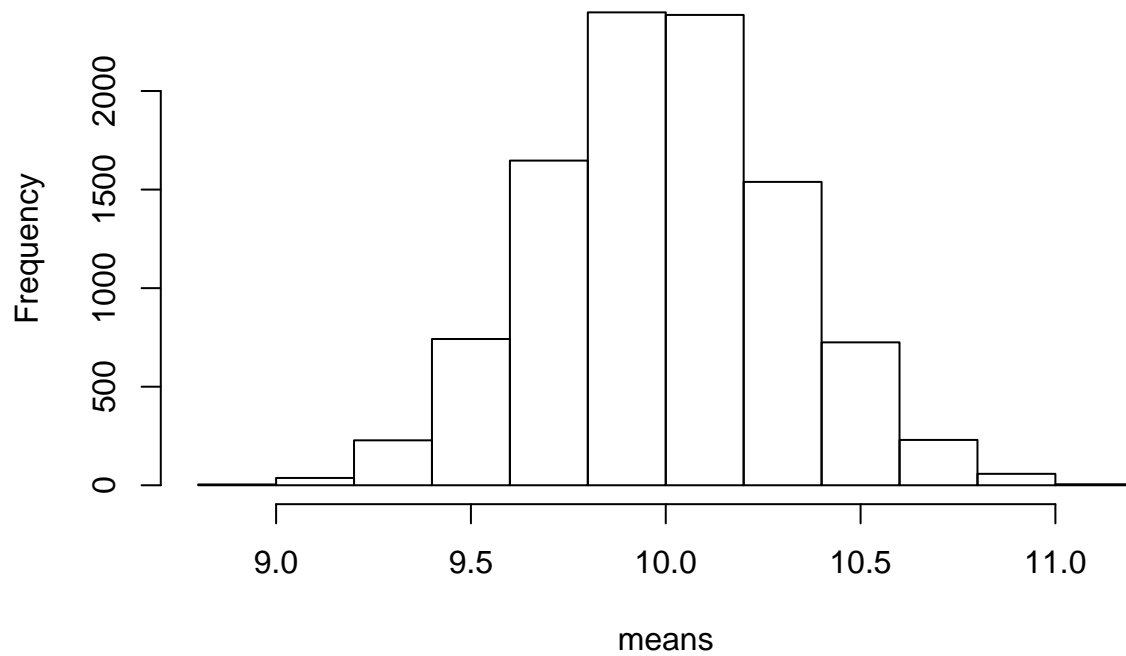
```
means<-numeric(10000)
for (i in 1:10000) {means[i]<-mean(rchisq(1000, df=1)*10)}
hist(means, main = "Chi-squared")
```

Chi-squared



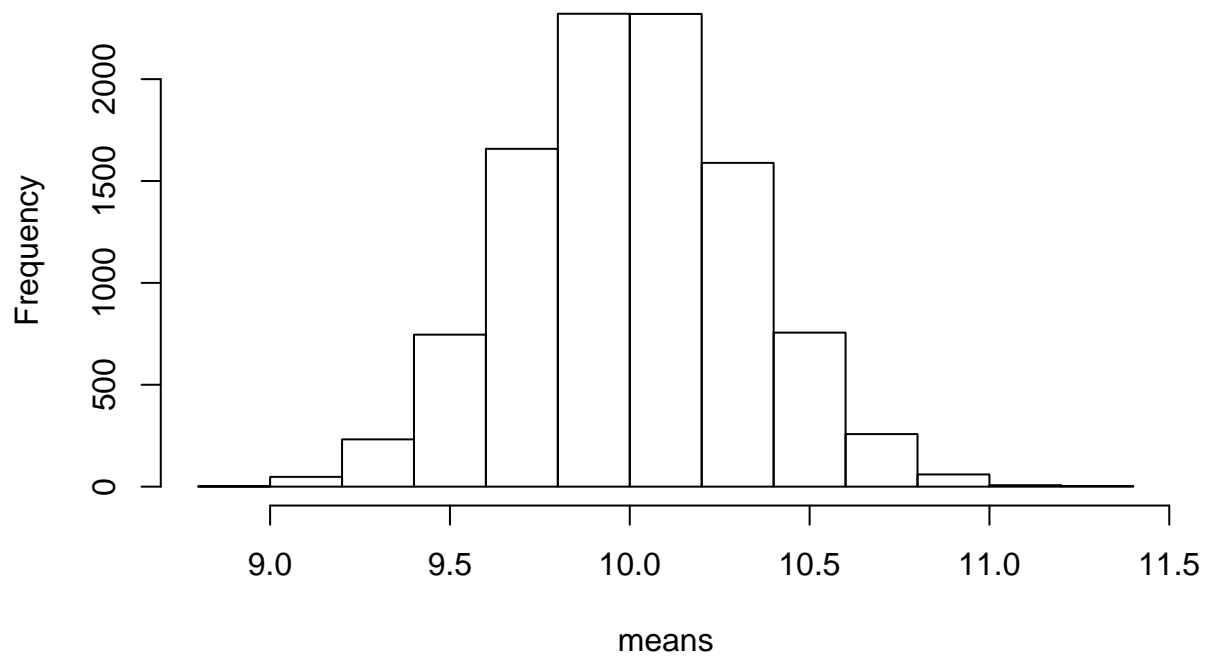
```
means<-numeric(10000)
for (i in 1:10000) {means[i]<-mean(rexp(1000)*10)}
hist(means, main = "Exponential")
```

Exponential



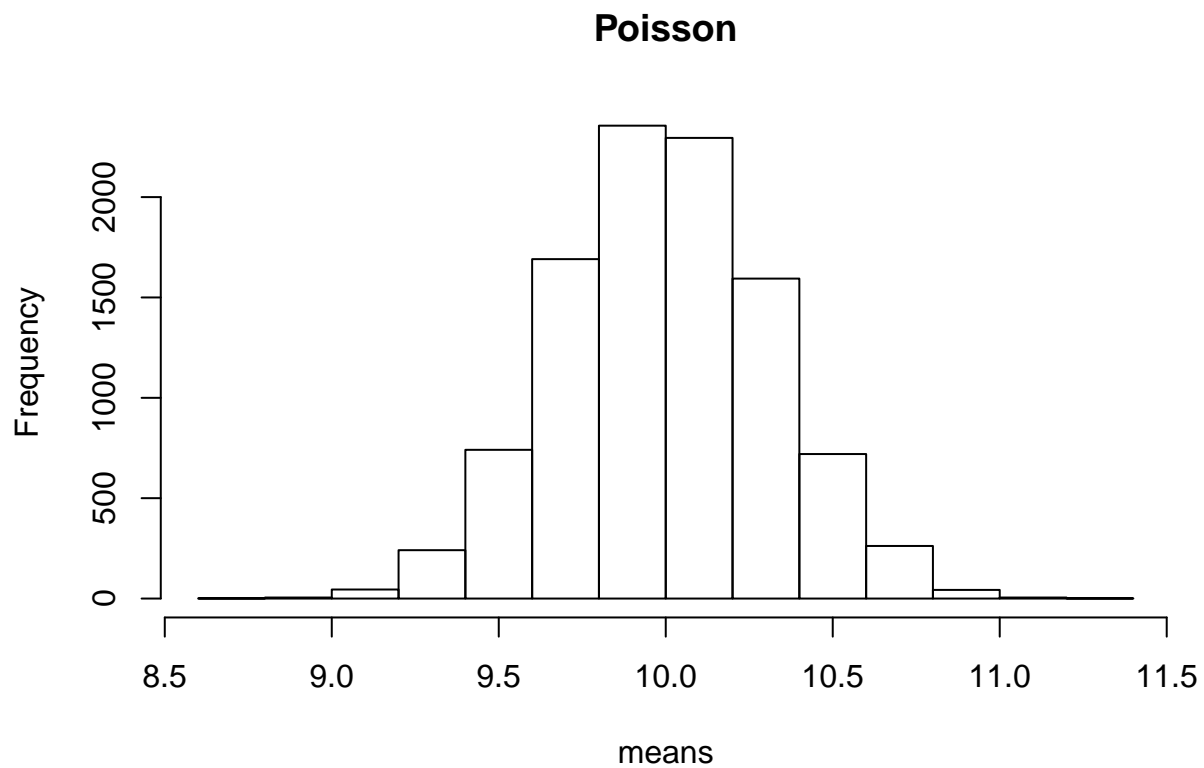
```
means<-numeric(10000)
for (i in 1:10000) {means[i]<-mean(rgamma(1000, shape = 1)*10)}
hist(means, main = "Gamma")
```

Gamma



```
means<-numeric(10000)
for (i in 1:10000) {means[i]<-mean(rpois(1000, lambda = 1)*10)}
```

```
hist(means, main = "Poisson")
```



```
means<-numeric(10000)  
for (i in 1:10000) {means[i]<-mean(rweibull(1000, shape = 1)*10)}  
hist(means, main = "Weibull")
```

