```
# Introduction to R
# Copyright 2013 by Ani Katchova

# Set working directory to where csv file is located
setwd("C:/Econometrics/Data")

# Read the data
mydata<- read.csv("C:/Econometrics/Data/intro_auto.csv")
attach(mydata)

# List the variables
names(mydata)

# Show first lines of data
head(mydata)
mydata[1:10,]

# Descriptive statistics
summary(mpg)
sd(mpg)
length(mpg)
summary(price)
sd(price)

# Sort the data
sort(make)

# Frequency tables
table(make)
table (make, foreign)

# Correlation among variables
cor(price, mpg)

# T-test for mean of one group
t.test(mpg, mu=20)

# ANOVA for equality of means for two groups
anova(lm(mpg ~ factor(foreign)))

# OLS regression - mpg (dependent variable) and weight, length and foreign
(independent variables)
olsreg <- lm(mpg ~ weight + length + foreign)
summary(olsreg)
# summary(lm(mpg ~ weight + length + foreign))

# Plotting data
plot (mpg ~ weight)
olsreg1 <- lm(mpg ~ weight)
abline(olsreg1)

# Redefining variables
Y <- cbind(mpg)
X <- cbind(weight, length, foreign)
summary(Y)
```

```
summary(X)
olsreg <- lm(Y ~ X)
summary(olsreg)

# Install and use packages
# install.packages("plm")
# library(plm)
```

```
> # Introduction to R
> # Copyright 2013 by Ani Katchova
>
> # Set working directory to where csv file is located
> setwd("C:/Econometrics/Data")
>
> # Read the data
> mydata<- read.csv("C:/Econometrics/Data/intro_auto.csv")
> attach(mydata)
>
> # List the variables
> names(mydata)
[1] "make"    "price"   "mpg"     "repairs" "weight"  "length"  "foreign"
>
> # Show first lines of data
> head(mydata)
  make price mpg repairs weight length foreign
1  AMC  4099  22       3   2930    186       0
2  AMC  4749  17       3   3350    173       0
3  AMC  3799  22       3   2640    168       0
4 Audi  9690  17       5   2830    189       1
5 Audi  6295  23       3   2070    174       1
6  BMW  9735  25       4   2650    177       1
> mydata[1:10,]
    make price mpg repairs weight length foreign
1    AMC  4099  22       3   2930    186       0
2    AMC  4749  17       3   3350    173       0
3    AMC  3799  22       3   2640    168       0
4   Audi  9690  17       5   2830    189       1
5   Audi  6295  23       3   2070    174       1
6    BMW  9735  25       4   2650    177       1
7  Buick  4816  20       3   3250    196       0
8  Buick  7827  15       4   4080    222       0
9  Buick  5788  18       3   3670    218       0
10 Buick  4453  26       3   2230    170       0
>
> # Descriptive statistics
> summary(mpg)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.00   17.25   21.00   20.92   23.00   35.00
> sd(mpg)
[1] 4.757504
> length(mpg)
[1] 26
> summary(price)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3299    4466    5146    6652    8054   15910
> sd(price)
[1] 3371.12
>
> # Sort the data
> sort(make)
 [1] AMC       AMC       AMC       Audi      Audi      BMW       Buick
 [8] Buick     Buick     Buick     Buick     Buick     Buick     Cadillac
[15] Cadillac  Cadillac  Chevrolet Chevrolet Chevrolet Chevrolet Chevrolet
```

```
[22] Chevrolet Datsun    Datsun    Datsun    Datsun
Levels: AMC Audi BMW Buick Cadillac Chevrolet Datsun
>
> # Frequency tables
> table(make)
make
      AMC      Audi       BMW     Buick  Cadillac Chevrolet    Datsun
        3         2         1         7         3         6         4
> table (make, foreign)
           foreign
make        0 1
  AMC       3 0
  Audi      0 2
  BMW       0 1
  Buick     7 0
  Cadillac  3 0
  Chevrolet 6 0
  Datsun    0 4
>
> # Correlation among variables
> cor(price, mpg)
[1] -0.4384618
>
> # T-test for mean of one group
> t.test(mpg, mu=20)

       One Sample t-test

data:  mpg
t = 0.9893, df = 25, p-value = 0.332
alternative hypothesis: true mean is not equal to 20
95 percent confidence interval:
 19.00148 22.84467
sample estimates:
mean of x
 20.92308

>
> # ANOVA for equality of means for two groups
> anova(lm(mpg ~ factor(foreign)))
Analysis of Variance Table

Response: mpg
               Df Sum Sq Mean Sq F value Pr(>F)
factor(foreign)  1  90.69  90.688  4.5806 0.0427 *
Residuals       24 475.16  19.798
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # OLS regression - mpg (dependent variable) and weight, length and foreign
(independent variables)
> olsreg <- lm(mpg ~ weight + length + foreign)
> summary(olsreg)

Call:
```

```
lm(formula = mpg ~ weight + length + foreign)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3902 -1.2734 -0.2991  0.7241  8.5203

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 44.968582   9.322678   4.824 8.08e-05 ***
weight      -0.005008   0.002188  -2.289    0.032 *
length      -0.043056   0.076926  -0.560    0.581
foreign     -1.269211   1.632134  -0.778    0.445
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.917 on 22 degrees of freedom
Multiple R-squared: 0.6693,      Adjusted R-squared: 0.6242
F-statistic: 14.84 on 3 and 22 DF,  p-value: 1.673e-05

> # summary(lm(mpg ~ weight + length + foreign))
>
> # Plotting data
> plot (mpg ~ weight)
> olsreg1 <- lm(mpg ~ weight)
> abline(olsreg1)
>
> # Redefining variables
> Y <- cbind(mpg)
> X <- cbind(weight, length, foreign)
> summary(Y)
      mpg
 Min.   :14.00
 1st Qu.:17.25
 Median :21.00
 Mean   :20.92
 3rd Qu.:23.00
 Max.   :35.00
> summary(X)
     weight         length         foreign
 Min.   :2020   Min.   :163.0   Min.   :0.0000
 1st Qu.:2642   1st Qu.:173.2   1st Qu.:0.0000
 Median :3200   Median :191.0   Median :0.0000
 Mean   :3099   Mean   :190.1   Mean   :0.2692
 3rd Qu.:3610   3rd Qu.:203.0   3rd Qu.:0.7500
 Max.   :4330   Max.   :222.0   Max.   :1.0000
> olsreg <- lm(Y ~ X)
> summary(olsreg)

Call:
lm(formula = Y ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3902 -1.2734 -0.2991  0.7241  8.5203
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 44.968582   9.322678   4.824 8.08e-05 ***
Xweight     -0.005008   0.002188  -2.289    0.032 *
Xlength     -0.043056   0.076926  -0.560    0.581
Xforeign    -1.269211   1.632134  -0.778    0.445
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.917 on 22 degrees of freedom
Multiple R-squared: 0.6693,      Adjusted R-squared: 0.6242
F-statistic: 14.84 on 3 and 22 DF,  p-value: 1.673e-05

>
> # Install and use packages
> # install.packages("plm")
> # library(plm)
```