

```

# Linear Regression in R
# Copyright 2013 by Ani Katchova

mydata<- read.csv("C:/Econometrics/Data/regression_auto.csv")
attach(mydata)

# Define variables
Y <- cbind(mpg)
X1 <- cbind(weight1)
X <- cbind(weight1, price, foreign)

# Descriptive statistics
summary(Y)
summary(X)

# Correlation among variables
cor(Y, X)

# Plotting data on a scatter diagram
plot(Y ~ X1, data = mydata)

# Simple linear regression
olsreg1 <- lm(Y ~ X1)
summary(olsreg1)
confint(olsreg1, level=0.95)
anova(olsreg1)

# Plotting regression line
abline(olsreg1)

# Predicted values for dependent variable
Ylhat <- fitted(olsreg1)
summary(Ylhat)
plot(Ylhat ~ X1)

# Regression residuals
elhat <- resid(olsreg1)
summary(elhat)
plot(elhat ~ X1)

# Multiple linear regression
olsreg2 <- lm(Y ~ X)
summary(olsreg2)
confint(olsreg2, level=0.95)
anova(olsreg2)

# Predicted values for dependent variable
Yhat <- fitted(olsreg2)
summary(Yhat)

# Regression residuals
ehat <- resid(olsreg2)
summary(ehat)

```

```

> # Linear Regression in R
> # Copyright 2013 by Ani Katchova
>
> mydata<- read.csv("C:/Econometrics/Data/regression_auto.csv")
> attach(mydata)
>
> # Define variables
> Y <- cbind(mpg)
> X1 <- cbind(weight1)
> X <- cbind(weight1, price, foreign)
>
> # Descriptive statistics
> 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)
      weight1      price      foreign
Min.   :2.020   Min.   : 3299   Min.   :0.0000
1st Qu.:2.643   1st Qu.: 4466   1st Qu.:0.0000
Median :3.200   Median : 5146   Median :0.0000
Mean   :3.099   Mean   : 6652   Mean   :0.2692
3rd Qu.:3.610   3rd Qu.: 8054   3rd Qu.:0.7500
Max.   :4.330   Max.   :15906   Max.   :1.0000
>
> # Correlation among variables
> cor(Y, X)
      weight1      price      foreign
mpg -0.8081609 -0.4384618  0.4003376
>
> # Plotting data on a scatter diagram
> plot(Y ~ X1, data = mydata)
>
> # Simple linear regression
> olsreg1 <- lm(Y ~ X1)
> summary(olsreg1)

Call:
lm(formula = Y ~ X1)

Residuals:
    Min       1Q   Median       3Q      Max
-5.4123 -1.6073 -0.1043  0.9261  8.1072

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.0665     2.6112   14.578 2.02e-13 ***
X1          -5.5315     0.8229   -6.722 5.93e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 2.86 on 24 degrees of freedom
Multiple R-squared: 0.6531, Adjusted R-squared: 0.6387
F-statistic: 45.19 on 1 and 24 DF, p-value: 5.935e-07

```
> confint(olsreg1, level=0.95)
              2.5 %      97.5 %
(Intercept) 32.677256 43.455664
X1          -7.229797 -3.833196
> anova(olsreg1)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)    
X1      1  369.57   369.57   45.189 5.935e-07 ***
Residuals 24  196.28     8.18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> # Plotting regression line
> abline(olsreg1)
>
> # Predicted values for dependent variable
> Ylhat <- fitted(olsreg1)
> summary(Ylhat)
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 14.12   18.10   20.37   20.92   23.45   26.89
> plot(Ylhat ~ X1)
>
> # Regression residuals
> elhat <- resid(olsreg1)
> summary(elhat)
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.4120 -1.6070 -0.1043  0.0000  0.9261  8.1070
> plot(elhat ~ X1)
>
> # Multiple linear regression
> olsreg2 <- lm(Y ~ X)
> summary(olsreg2)
```

Call:

```
lm(formula = Y ~ X)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-4.6942	-1.1857	-0.0452	0.6433	8.6895

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	42.1661962	4.2647533	9.887	1.48e-09 ***
Xweight1	-7.1211114	1.6046735	-4.438	0.000207 ***
Xprice	0.0002258	0.0002654	0.851	0.404002
Xforeign	-2.5071265	2.0565685	-1.219	0.235723

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.89 on 22 degrees of freedom
Multiple R-squared: 0.6752, Adjusted R-squared: 0.6309
F-statistic: 15.25 on 3 and 22 DF, p-value: 1.374e-05

```
> confint(olsreg2, level=0.95)
```

```
                2.5 %          97.5 %  
(Intercept)  3.332164e+01 51.0107531780  
Xweight1     -1.044900e+01 -3.7932221856  
Xprice       -3.245229e-04  0.0007760878  
Xforeign     -6.772188e+00  1.7579354345
```

```
> anova(olsreg2)
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	3	382.08	127.360	15.247	1.374e-05 ***
Residuals	22	183.77	8.353		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
>
```

```
> # Predicted values for dependent variable
```

```
> Yhat <- fitted(olsreg2)
```

```
> summary(Yhat)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
13.90	17.91	20.46	20.92	23.99	27.89

```
>
```

```
> # Regression residuals
```

```
> ehat <- resid(olsreg2)
```

```
> summary(ehat)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-4.69400	-1.18600	-0.04524	0.00000	0.64330	8.68900



