

Categorical Predictor Variables

```
library(ISLR)  
data(Carseats)  
names(Carseats)
```

```
[1] "Sales"      "CompPrice"  "Income"     "Advertising" "Population"  
[6] "Price"     "ShelveLoc"  "Age"        "Education"   "Urban"  
[11] "US"
```

```
levels(Carseats$ShelveLoc)
```

```
[1] "Bad"      "Good"      "Medium"
```

Dummy Variables

```
levels(Carseats$ShelveLoc)
```

```
[1] "Bad"    "Good"   "Medium"
```

```
lm.fit=lm(Sales~.,data=Carseats)  
summary(lm.fit)
```

```
Call:  
lm(formula = Sales ~ ., data = Carseats)
```

```
Residuals:  
      Min       1Q   Median       3Q      Max  
-2.8692 -0.6908  0.0211  0.6636  3.4115
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)   5.6606231   0.6034487   9.380 < 2e-16 ***  
CompPrice     0.0928153   0.0041477  22.378 < 2e-16 ***  
Income        0.0158028   0.0018451   8.565 2.58e-16 ***  
Advertising    0.1230951   0.0111237  11.066 < 2e-16 ***  
Population     0.0002079   0.0003705   0.561  0.575  
Price        -0.0953579   0.0026711 -35.700 < 2e-16 ***  
ShelveLocGood  4.8501827   0.1531100  31.678 < 2e-16 ***  
ShelveLocMedium 1.9567148   0.1261056  15.516 < 2e-16 ***  
Age          -0.0460452   0.0031817 -14.472 < 2e-16 ***  
Education     -0.0211018   0.0197205  -1.070  0.285  
UrbanYes      0.1228864   0.1129761   1.088  0.277
```

```
USYes      -0.1840928  0.1498423  -1.229    0.220
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.019 on 388 degrees of freedom
Multiple R-squared:  0.8734,    Adjusted R-squared:  0.8698
F-statistic: 243.4 on 11 and 388 DF,  p-value: < 2.2e-16
```

```
contrasts(Carseats$ShelveLoc)
```

	Good	Medium
Bad	0	0
Good	1	0
Medium	0	1

Re-leveling dummy variables

```
contrasts(Carseats$ShelveLoc)
```

	Good	Medium
Bad	0	0
Good	1	0
Medium	0	1

```
Carseats$ShelveLoc <- relevel(Carseats$ShelveLoc, ref="Good")  
contrasts(Carseats$ShelveLoc)
```

	Bad	Medium
Good	0	0
Bad	1	0
Medium	0	1

```
summary(lm(Sales~.,data=Carseats)) # Check it out
```

```
Call:  
lm(formula = Sales ~ ., data = Carseats)
```

```
Residuals:  
    Min      1Q  Median      3Q     Max  
-2.8692 -0.6908  0.0211  0.6636  3.4115
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.5108058	0.6039582	17.403	< 2e-16	***
CompPrice	0.0928153	0.0041477	22.378	< 2e-16	***
Income	0.0158028	0.0018451	8.565	2.58e-16	***
Advertising	0.1230951	0.0111237	11.066	< 2e-16	***
Population	0.0002079	0.0003705	0.561	0.575	
Price	-0.0953579	0.0026711	-35.700	< 2e-16	***
ShelveLocBad	-4.8501827	0.1531100	-31.678	< 2e-16	***
ShelveLocMedium	-2.8934679	0.1308928	-22.106	< 2e-16	***
Age	-0.0460452	0.0031817	-14.472	< 2e-16	***
Education	-0.0211018	0.0197205	-1.070	0.285	
UrbanYes	0.1228864	0.1129761	1.088	0.277	
USYes	-0.1840928	0.1498423	-1.229	0.220	

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

Residual standard error: 1.019 on 388 degrees of freedom

Multiple R-squared: 0.8734, Adjusted R-squared: 0.8698

F-statistic: 243.4 on 11 and 388 DF, p-value: < 2.2e-16

Centered data

```
library(MASS)  
data(Boston)
```

```
Boston.centered = data.frame(scale(Boston, center=TRUE, scale=FALSE))
```

```
summary(Boston.centered)
```

crim	zn	indus	chas
Min. : -3.60720	Min. : -11.364	Min. : -10.677	Min. : -0.06917
1st Qu.: -3.53148	1st Qu.: -11.364	1st Qu.: -5.947	1st Qu.: -0.06917
Median : -3.35701	Median : -11.364	Median : -1.447	Median : -0.06917
Mean : 0.00000	Mean : 0.000	Mean : 0.000	Mean : 0.00000
3rd Qu.: 0.06356	3rd Qu.: 1.136	3rd Qu.: 6.963	3rd Qu.: -0.06917
Max. : 85.36268	Max. : 88.636	Max. : 16.603	Max. : 0.93083
nox	rm	age	dis
Min. : -0.1697	Min. : -2.72363	Min. : -65.675	Min. : -2.6654
1st Qu.: -0.1057	1st Qu.: -0.39913	1st Qu.: -23.550	1st Qu.: -1.6949
Median : -0.0167	Median : -0.07613	Median : 8.925	Median : -0.5876
Mean : 0.0000	Mean : 0.00000	Mean : 0.000	Mean : 0.0000
3rd Qu.: 0.0693	3rd Qu.: 0.33887	3rd Qu.: 25.500	3rd Qu.: 1.3934
Max. : 0.3163	Max. : 2.49537	Max. : 31.425	Max. : 8.3315
rad	tax	ptratio	black
Min. : -8.549	Min. : -221.24	Min. : -5.8555	Min. : -356.35
1st Qu.: -5.549	1st Qu.: -129.24	1st Qu.: -1.0555	1st Qu.: 18.70
Median : -4.549	Median : -78.24	Median : 0.5945	Median : 34.77
Mean : 0.000	Mean : 0.00	Mean : 0.0000	Mean : 0.00
3rd Qu.: 14.451	3rd Qu.: 257.76	3rd Qu.: 1.7445	3rd Qu.: 39.55
Max. : 14.451	Max. : 302.76	Max. : 3.5445	Max. : 40.23
lstat	medv		
Min. : -10.923	Min. : -17.533		

1st Qu.:	-5.703	1st Qu.:	-5.508
Median :	-1.293	Median :	-1.333
Mean :	0.000	Mean :	0.000
3rd Qu.:	4.302	3rd Qu.:	2.467
Max. :	25.317	Max. :	27.467

Model on centered data

```
summary(lm(medv~lstat+age,data=Boston))
```

Call:

```
lm(formula = medv ~ lstat + age, data = Boston)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.981	-3.978	-1.283	1.968	23.158

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	33.22276	0.73085	45.458	< 2e-16	***
lstat	-1.03207	0.04819	-21.416	< 2e-16	***
age	0.03454	0.01223	2.826	0.00491	**

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

Residual standard error: 6.173 on 503 degrees of freedom

Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495

F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

```
summary(lm(medv~lstat+age,data=Boston.centered))
```

Call:

```
lm(formula = medv ~ lstat + age, data = Boston.centered)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.981	-3.978	-1.283	1.968	23.158

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-8.697e-16	2.744e-01	0.000	1.00000	
lstat	-1.032e+00	4.819e-02	-21.416	< 2e-16	***
age	3.454e-02	1.223e-02	2.826	0.00491	**

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

Residual standard error: 6.173 on 503 degrees of freedom

Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495

F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

Standardized data

```
Boston.standardized = data.frame(scale(Boston, center=TRUE, scale=TRUE))  
summary(Boston.standardized)
```

```
      crim              zn              indus  
Min.   :-0.419367   Min.   :-0.48724   Min.   :-1.5563  
1st Qu.: -0.410563   1st Qu.: -0.48724   1st Qu.: -0.8668  
Median :-0.390280   Median :-0.48724   Median :-0.2109  
Mean   : 0.000000   Mean   : 0.00000   Mean   : 0.0000  
3rd Qu.: 0.007389   3rd Qu.: 0.04872   3rd Qu.: 1.0150  
Max.   : 9.924110   Max.   : 3.80047   Max.   : 2.4202  
  
      chas              nox              rm              age  
Min.   :-0.2723   Min.   :-1.4644   Min.   :-3.8764   Min.   :-2.3331  
1st Qu.: -0.2723   1st Qu.: -0.9121   1st Qu.: -0.5681   1st Qu.: -0.8366  
Median :-0.2723   Median :-0.1441   Median :-0.1084   Median : 0.3171  
Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000  
3rd Qu.: -0.2723   3rd Qu.: 0.5981   3rd Qu.: 0.4823   3rd Qu.: 0.9059  
Max.   : 3.6648   Max.   : 2.7296   Max.   : 3.5515   Max.   : 1.1164  
  
      dis              rad              tax              ptratio  
Min.   :-1.2658   Min.   :-0.9819   Min.   :-1.3127   Min.   :-2.7047  
1st Qu.: -0.8049   1st Qu.: -0.6373   1st Qu.: -0.7668   1st Qu.: -0.4876  
Median :-0.2790   Median :-0.5225   Median :-0.4642   Median : 0.2746  
Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000  
3rd Qu.: 0.6617   3rd Qu.: 1.6596   3rd Qu.: 1.5294   3rd Qu.: 0.8058  
Max.   : 3.9566   Max.   : 1.6596   Max.   : 1.7964   Max.   : 1.6372  
  
      black              lstat              medv  
Min.   :-3.9033   Min.   :-1.5296   Min.   :-1.9063  
1st Qu.: 0.2049   1st Qu.: -0.7986   1st Qu.: -0.5989  
Median : 0.3808   Median :-0.1811   Median :-0.1449  
Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
```

3rd Qu.:	0.4332	3rd Qu.:	0.6024	3rd Qu.:	0.2683
Max.:	0.4406	Max.:	3.5453	Max.:	2.9865

Standardized data

```
for(v in names(Boston.standardized)){  
  print(paste(v, as.character(round(sd(Boston[,v]),2)),  
as.character(round(sd(Boston.standardized[,v]),2))))  
}
```

```
[1] "crim 8.6 1"  
[1] "zn 23.32 1"  
[1] "indus 6.86 1"  
[1] "chas 0.25 1"  
[1] "nox 0.12 1"  
[1] "rm 0.7 1"  
[1] "age 28.15 1"  
[1] "dis 2.11 1"  
[1] "rad 8.71 1"  
[1] "tax 168.54 1"  
[1] "ptratio 2.16 1"  
[1] "black 91.29 1"  
[1] "lstat 7.14 1"  
[1] "medv 9.2 1"
```

Model on standardized data

```
summary(lm(medv~lstat+age,data=Boston))
```

Call:

```
lm(formula = medv ~ lstat + age, data = Boston)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.981	-3.978	-1.283	1.968	23.158

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	33.22276	0.73085	45.458	< 2e-16	***
lstat	-1.03207	0.04819	-21.416	< 2e-16	***
age	0.03454	0.01223	2.826	0.00491	**

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

Residual standard error: 6.173 on 503 degrees of freedom

Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495

F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

```
summary(lm(medv~lstat+age,data=Boston.standardized))
```

Call:

```
lm(formula = medv ~ lstat + age, data = Boston.standardized)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.7376	-0.4325	-0.1396	0.2140	2.5180

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.615e-16	2.984e-02	0.000	1.00000
lstat	-8.013e-01	3.742e-02	-21.416	< 2e-16 ***
age	1.057e-01	3.742e-02	2.826	0.00491 **

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

Residual standard error: 0.6712 on 503 degrees of freedom

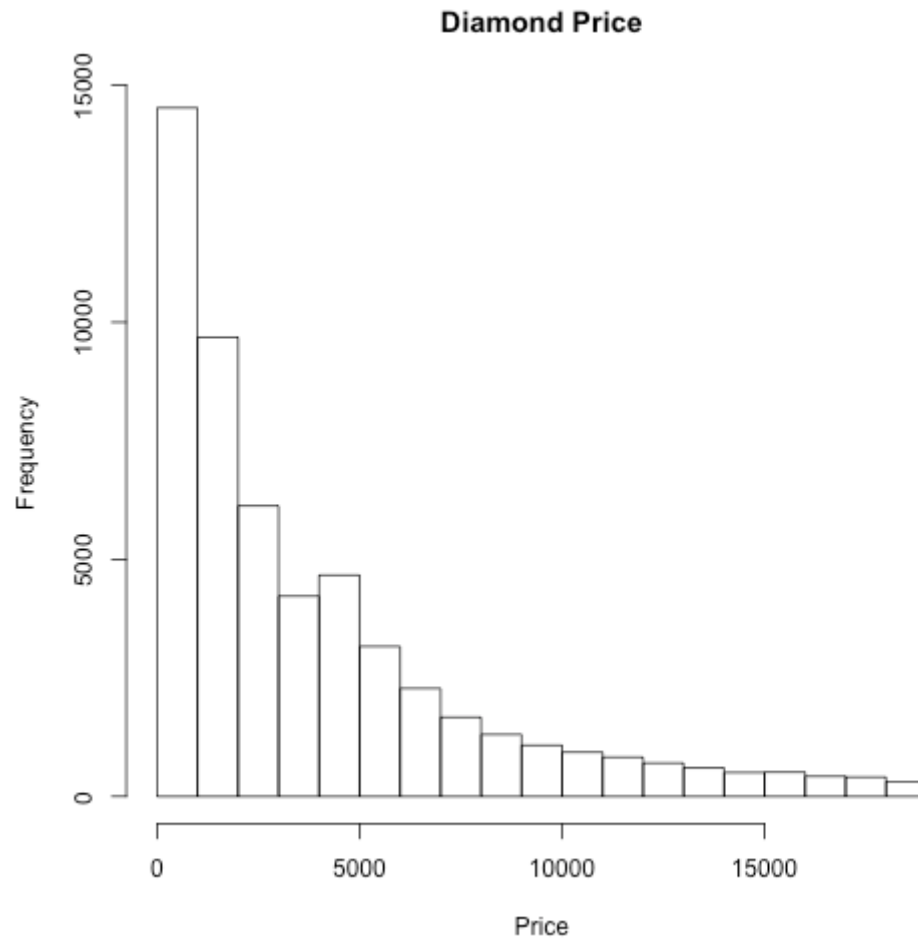
Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495

F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

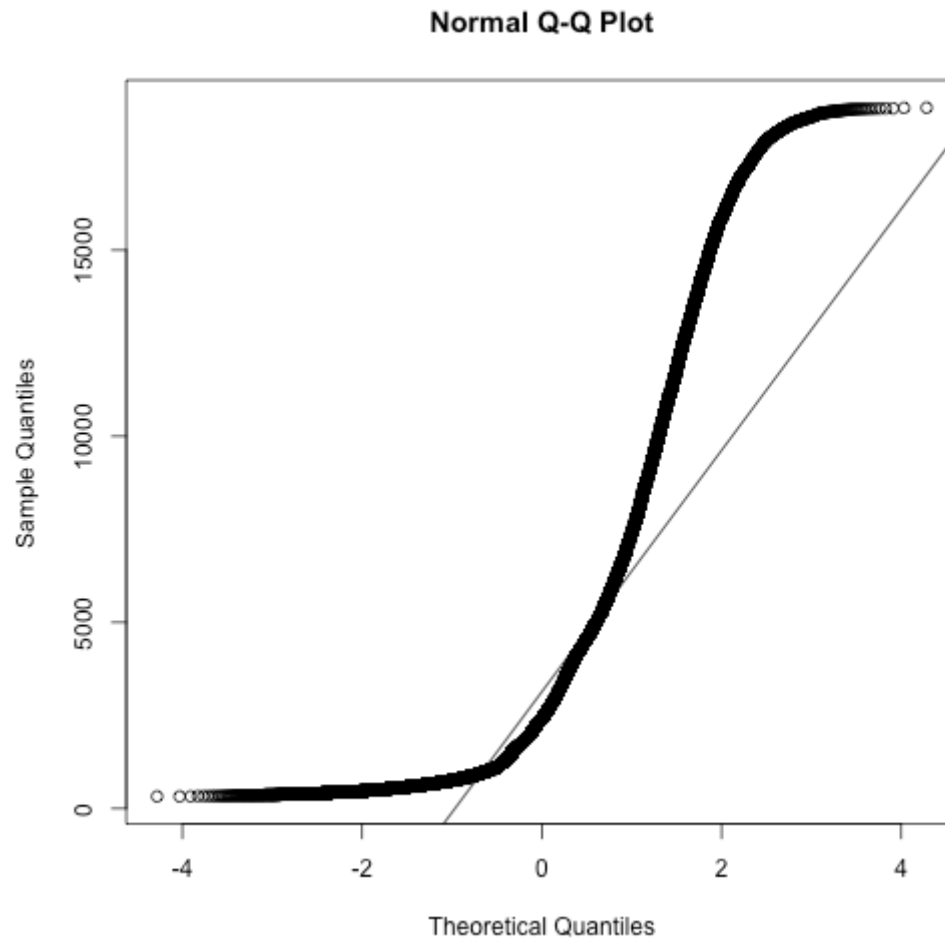
Log() may make data more normal

```
require(ggplot2)
attach(diamonds)

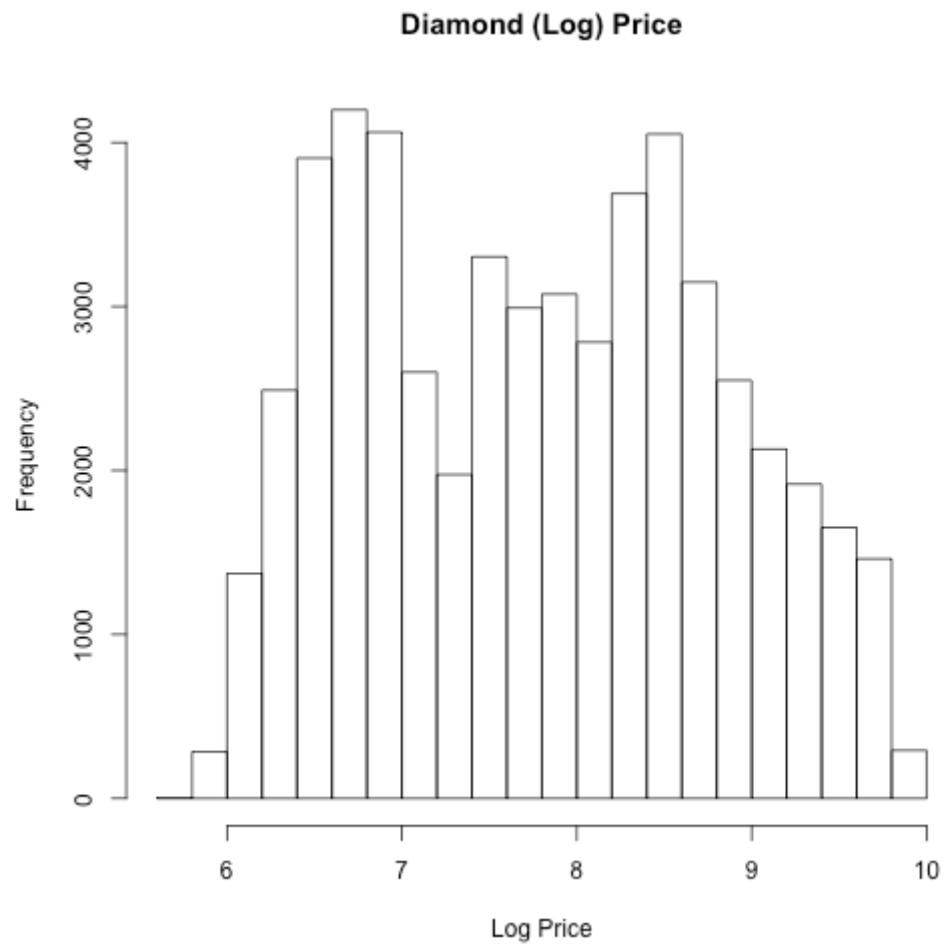
hist(diamonds$price, main="Diamond Price", xlab="Price")
```

```
qqnorm(diamonds$price); qqline(diamonds$price)
```

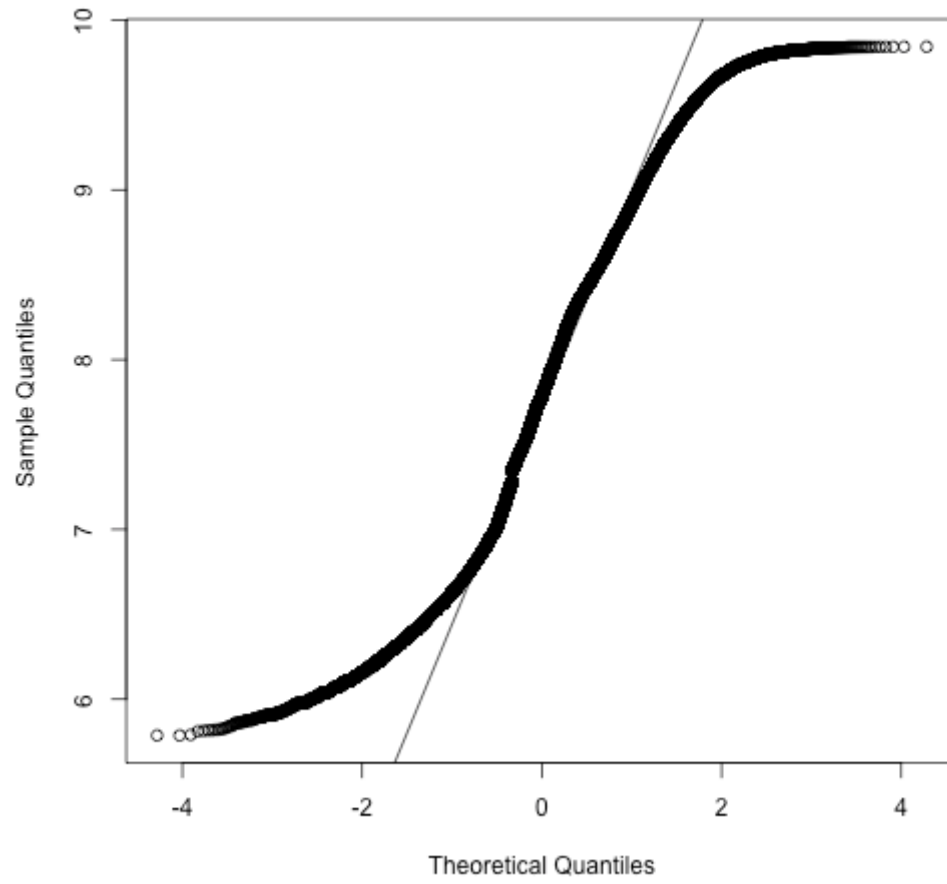


```
hist(log(diamonds$price), main="Diamond (Log) Price", xlab="Log Price")
```



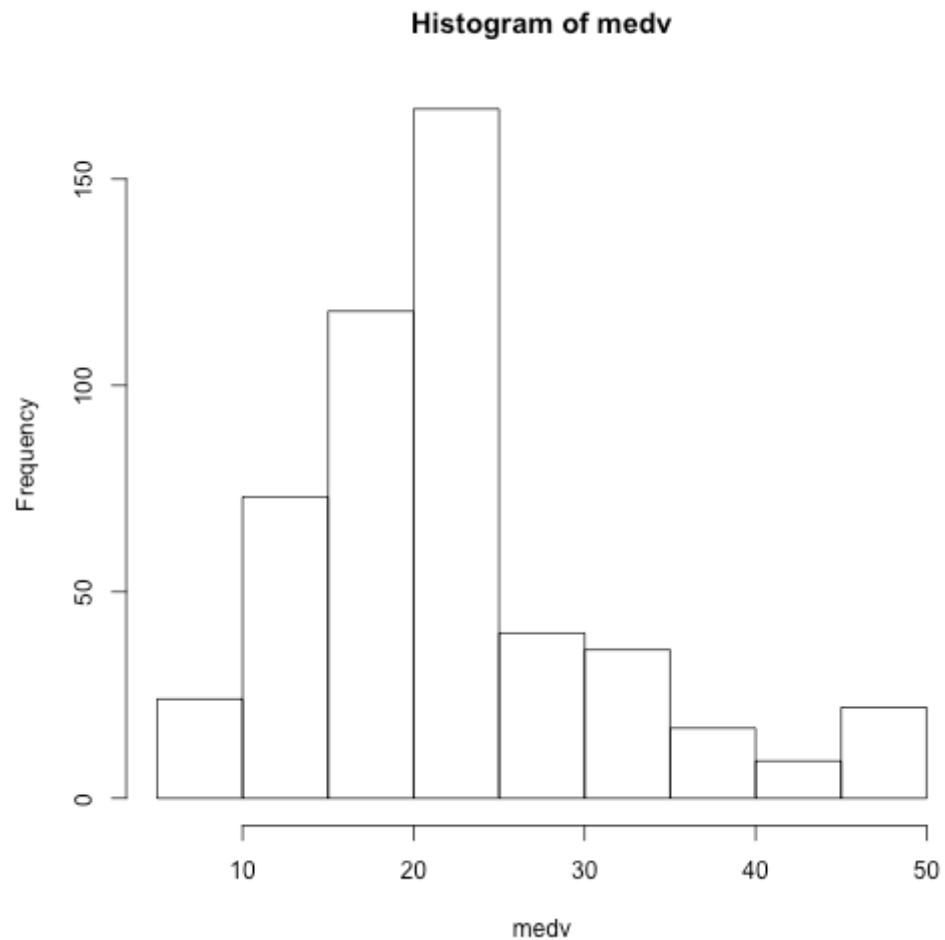
```
qqnorm(log(diamonds$price)); qqline(log(diamonds$price))
```

Normal Q-Q Plot

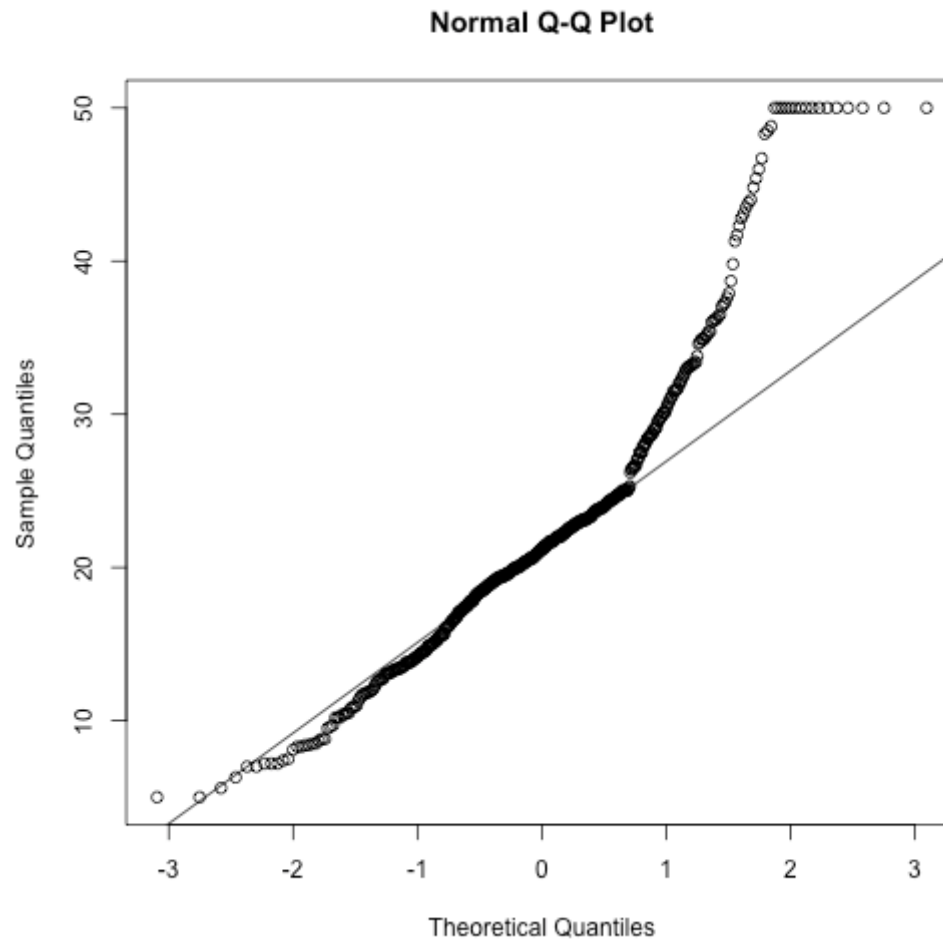


Log() may make data more normal

```
attach(Boston)  
hist(medv)
```

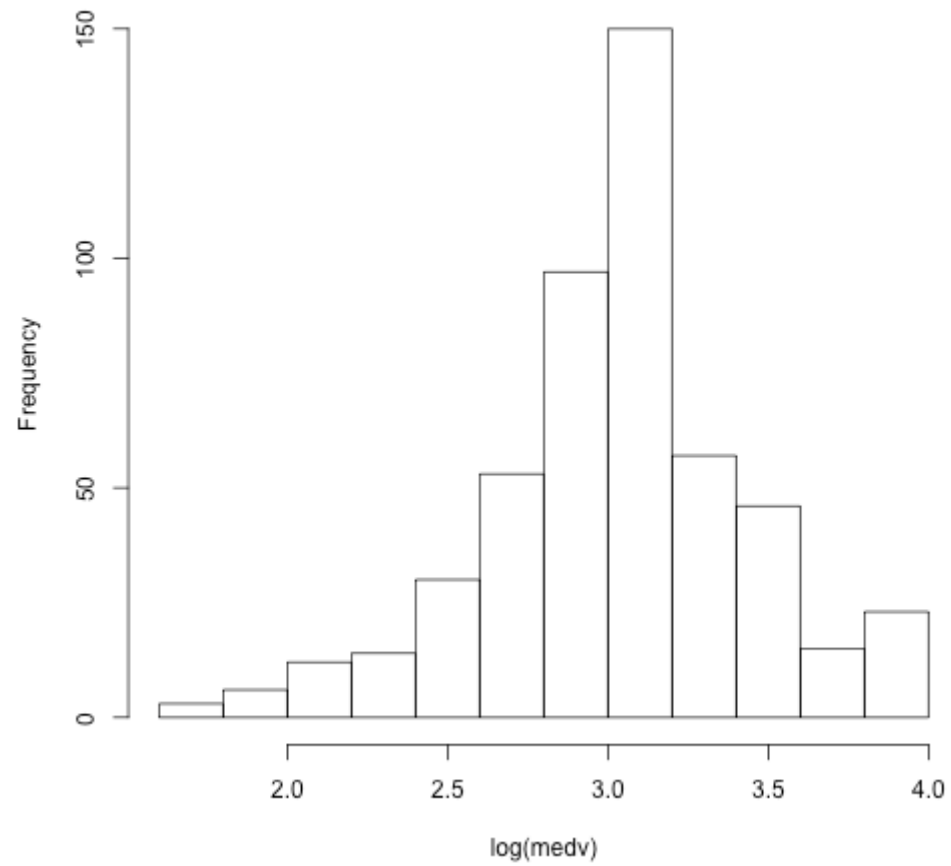


```
qqnorm(medv); qqline(medv)
```



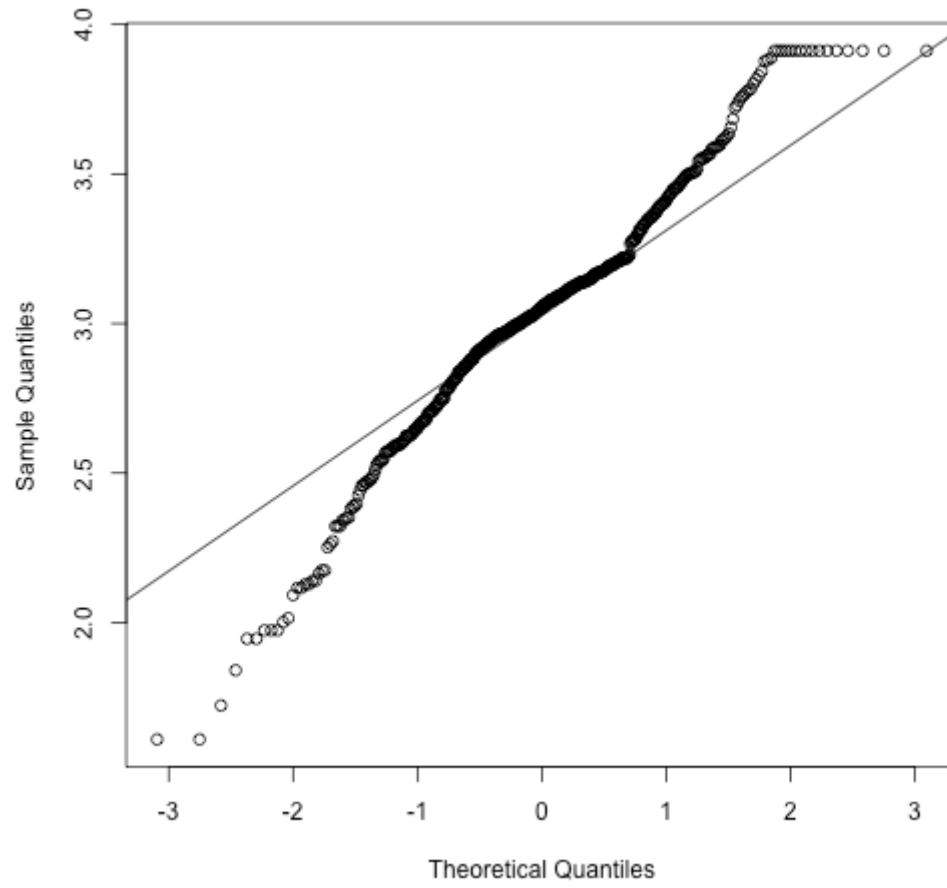
```
hist(log(medv))
```

Histogram of log(medv)



```
qqnorm(log(medv)); qqline(log(medv))
```

Normal Q-Q Plot



$$Y = \text{Log}(X) * B$$

```
summary(lm(medv~log(rm),data=Boston))
```

Call:

```
lm(formula = medv ~ log(rm), data = Boston)
```

Residuals:

Min	1Q	Median	3Q	Max
-19.487	-2.875	-0.104	2.837	39.816

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-76.488	5.028	-15.21	<2e-16	***
log(rm)	54.055	2.739	19.73	<2e-16	***

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

Residual standard error: 6.915 on 504 degrees of freedom

Multiple R-squared: 0.4358, Adjusted R-squared: 0.4347

F-statistic: 389.3 on 1 and 504 DF, p-value: < 2.2e-16

$$\text{Log}(Y) = X * B$$

```
summary(lm(log(medv)~rm,data=Boston))
```

Call:

```
lm(formula = log(medv) ~ rm, data = Boston)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.20386	-0.09419	0.06416	0.16991	1.36088

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.72374	0.12699	5.699	2.05e-08	***
rm	0.36769	0.02008	18.309	< 2e-16	***

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

Residual standard error: 0.3171 on 504 degrees of freedom

Multiple R-squared: 0.3995, Adjusted R-squared: 0.3983

F-statistic: 335.2 on 1 and 504 DF, p-value: < 2.2e-16

$$\text{Log}(Y) = \text{Log}(X)*B$$

```
summary(lm(log(medv)~log(rm),data=Boston))
```

Call:

```
lm(formula = log(medv) ~ log(rm), data = Boston)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.21784	-0.08913	0.05841	0.17311	1.52771

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.0348	0.2356	-4.392	1.37e-05	***
log(rm)	2.2214	0.1284	17.302	< 2e-16	***

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

Residual standard error: 0.3241 on 504 degrees of freedom

Multiple R-squared: 0.3726, Adjusted R-squared: 0.3714

F-statistic: 299.4 on 1 and 504 DF, p-value: < 2.2e-16

Count Data: $\text{Log}(Y) = X*B$

```
College <- read.table("../../data/College.csv", header=TRUE, sep=",")
options(scipen="2")
summary(lm(Apps~Outstate+PhD+S.F.Ratio, data=College))
```

```
Call:
lm(formula = Apps ~ Outstate + PhD + S.F.Ratio, data = College)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-5194.2 -1526.9  -625.3   462.0 17678.3
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-6017.33972	1164.55442	-5.167	4.10e-07	***
Outstate	0.09953	0.05149	1.933	0.0541	.
PhD	62.11394	10.55757	5.883	9.77e-09	***
S.F.Ratio	211.54423	51.85450	4.080	5.65e-05	***

```
---
```

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

```
Residual standard error: 2865 on 334 degrees of freedom
(1 observation deleted due to missingness)
```

```
Multiple R-squared:  0.1634,    Adjusted R-squared:  0.1559
```

```
F-statistic: 21.74 on 3 and 334 DF,  p-value: 6.927e-13
```

```
summary(glm(Apps~Outstate+PhD+S.F.Ratio, family=poisson(link="log"),
data=College))
```

```
Call:
glm(formula = Apps ~ Outstate + PhD + S.F.Ratio, family = poisson(link =
"log"),
     data = College)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-113.785	-30.019	-13.824	9.366	209.813

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.466289123	0.009928967	349.11	<2e-16	***
Outstate	0.000030619	0.000000355	86.26	<2e-16	***
PhD	0.035274463	0.000095444	369.58	<2e-16	***
S.F.Ratio	0.093560077	0.000349618	267.61	<2e-16	***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 877566 on 337 degrees of freedom
Residual deviance: 621594 on 334 degrees of freedom
(1 observation deleted due to missingness)
AIC: 624672

Number of Fisher Scoring iterations: 5

Box-Cox

```
library(MASS)
attach(Boston)

lm.fit=lm(medv~lstat, data=Boston)
summary(lm.fit)
```

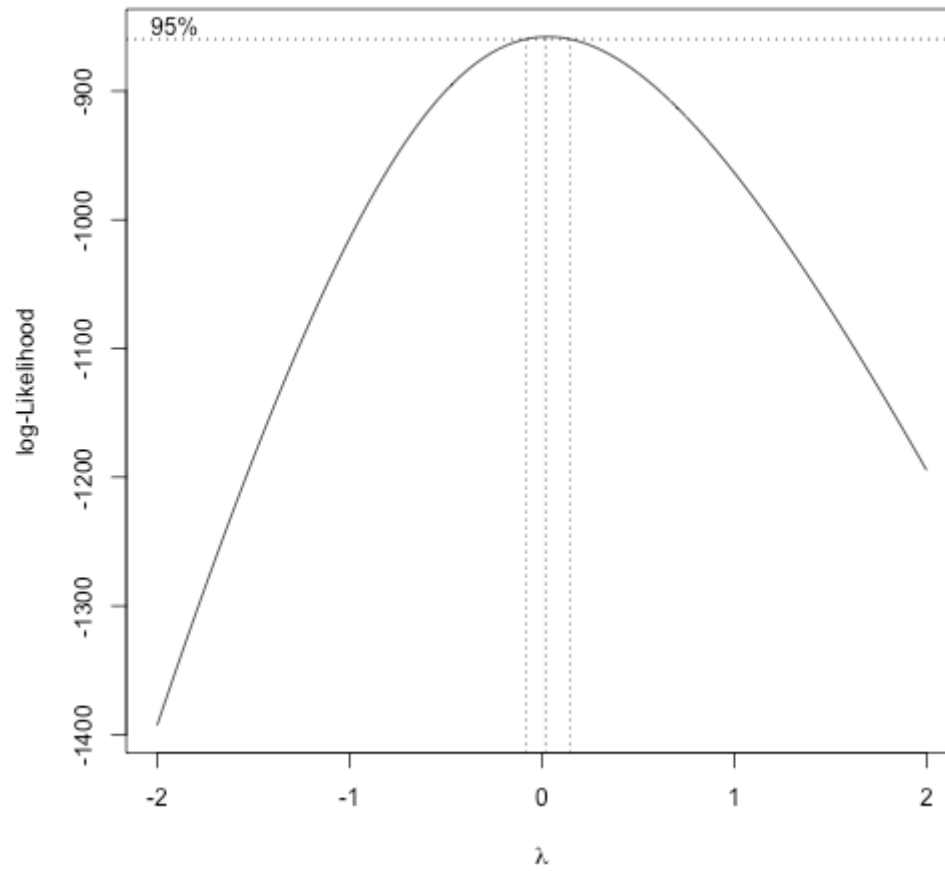
```
Call:
lm(formula = medv ~ lstat, data = Boston)

Residuals:
    Min       1Q   Median       3Q      Max
-15.168  -3.990  -1.318   2.034  24.500

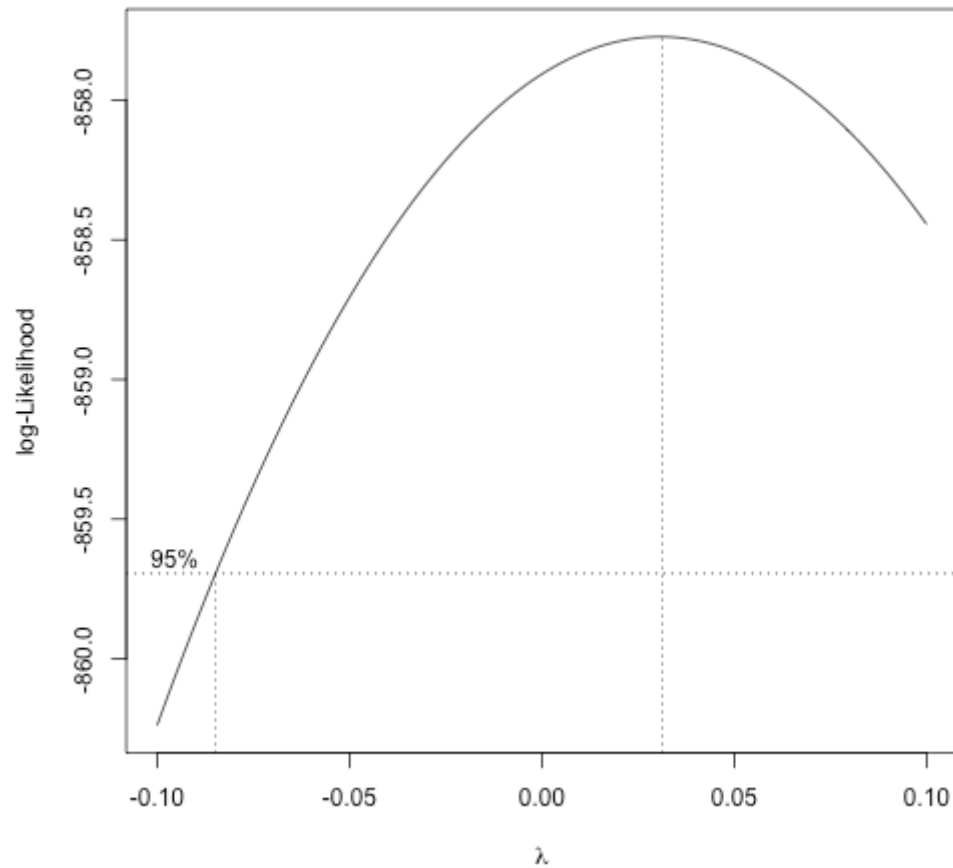
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 34.55384    0.56263   61.41  <2e-16 ***
lstat       -0.95005    0.03873  -24.53  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.216 on 504 degrees of freedom
Multiple R-squared:  0.5441,    Adjusted R-squared:  0.5432
F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
boxcox(lm.fit)
```



```
boxcox(lm.fit, lambda = seq(-0.1, 0.1, 0.01))
```



```
medv.box=medv^0.04  
lm.fit.box.cox=lm(medv.box~lstat, data=Boston)  
summary(lm.fit.box.cox)
```

```
Call:  
lm(formula = medv.box ~ lstat, data = Boston)
```

```
Residuals:
```


Min	1Q	Median	3Q	Max
-0.041296	-0.006786	-0.000995	0.005111	0.040816

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.15544153	0.00099201	1164.75	<2e-16	***
lstat	-0.00207355	0.00006829	-30.36	<2e-16	***

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

Residual standard error: 0.01096 on 504 degrees of freedom

Multiple R-squared: 0.6465, Adjusted R-squared: 0.6458

F-statistic: 921.9 on 1 and 504 DF, p-value: < 2.2e-16

Polynomials (1st Degree vs. 2nd Degree)

```
lm.fit2=lm(medv~lstat+I(lstat^2), data=Boston)
summary(lm.fit2)
```

```
Call:
lm(formula = medv ~ lstat + I(lstat^2), data = Boston)

Residuals:
    Min       1Q   Median       3Q      Max
-15.2834  -3.8313  -0.5295   2.3095  25.4148

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.862007   0.872084   49.15  <2e-16 ***
lstat       -2.332821   0.123803  -18.84  <2e-16 ***
I(lstat^2)   0.043547   0.003745   11.63  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.524 on 503 degrees of freedom
Multiple R-squared:  0.6407,    Adjusted R-squared:  0.6393
F-statistic: 448.5 on 2 and 503 DF,  p-value: < 2.2e-16
```

```
lm.fit=lm(medv~lstat, data=Boston) # Model withouth quadratic term
summary(lm.fit)
```

```
Call:
```

```
lm(formula = medv ~ lstat, data = Boston)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-15.168	-3.990	-1.318	2.034	24.500

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	34.55384	0.56263	61.41	<2e-16	***
lstat	-0.95005	0.03873	-24.53	<2e-16	***

```
---
```

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

```
Residual standard error: 6.216 on 504 degrees of freedom
```

```
Multiple R-squared:  0.5441,    Adjusted R-squared:  0.5432
```

```
F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
anova(lm.fit,lm.fit2) # Tests if quadratic model is superior
```

```
Analysis of Variance Table
```

```
Model 1: medv ~ lstat
```

```
Model 2: medv ~ lstat + I(lstat^2)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	504	19472					
2	503	15347	1	4125.1	135.2	< 2.2e-16	***

```
---
```

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

Polynomials (1st Degree vs. 5th Degree)

```
lm.fit5=lm(medv~poly(lstat,5), data=Boston)
summary(lm.fit5) # See the higher R-Squared
```

Call:

```
lm(formula = medv ~ poly(lstat, 5), data = Boston)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.5433	-3.1039	-0.7052	2.0844	27.1153

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	22.5328	0.2318	97.197	< 2e-16	***
poly(lstat, 5)1	-152.4595	5.2148	-29.236	< 2e-16	***
poly(lstat, 5)2	64.2272	5.2148	12.316	< 2e-16	***
poly(lstat, 5)3	-27.0511	5.2148	-5.187	0.00000031	***
poly(lstat, 5)4	25.4517	5.2148	4.881	0.00000142	***
poly(lstat, 5)5	-19.2524	5.2148	-3.692	0.000247	***

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

Residual standard error: 5.215 on 500 degrees of freedom

Multiple R-squared: 0.6817, Adjusted R-squared: 0.6785

F-statistic: 214.2 on 5 and 500 DF, p-value: < 2.2e-16

```
anova(lm.fit,lm.fit5) # Test if the polynomial model is superior
```

Analysis of Variance Table

Model 1: medv ~ lstat

Model 2: medv ~ poly(lstat, 5)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	504	19472				
2	500	13597	4	5875.3	54.013	< 2.2e-16 ***

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

Lagged Models

```
HousingStarts=read.csv("../../data/HousingStarts.csv",header=T,na.strings="?")  
HousingStarts=na.omit(HousingStarts) # Removes NA's  
head(HousingStarts)
```

	Month	T	KUnits	S.P	Q1	Q2	Q3	Q4	
1	Jan	1990	1	99.2	329.08	1	0	0	0
2	Feb	1990	2	86.9	331.89	1	0	0	0
3	Mar	1990	3	108.5	339.94	1	0	0	0
4	Apr	1990	4	119.0	330.80	0	1	0	0
5	May	1990	5	121.1	361.23	0	1	0	0
6	Jun	1990	6	117.8	358.02	0	1	0	0

```
lm.KUnits = lm(KUnits~T+S.P+Q2+Q3+Q4, data=HousingStarts)  
summary(lm.KUnits)
```

```
Call:  
lm(formula = KUnits ~ T + S.P + Q2 + Q3 + Q4, data = HousingStarts)
```

```
Residuals:  
      Min       1Q   Median       3Q      Max  
-57.739 -19.253  -2.681   20.302   71.428
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  71.191003   6.033686  11.799 < 2e-16 ***
```

T	-0.290596	0.043991	-6.606	2.53e-10	***
S.P	0.072582	0.008377	8.664	6.82e-16	***
Q2	30.888546	5.327939	5.797	2.11e-08	***
Q3	28.531099	5.392824	5.291	2.75e-07	***
Q4	8.374365	5.393472	1.553	0.122	

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

Residual standard error: 29.89 on 240 degrees of freedom

Multiple R-squared: 0.3431, Adjusted R-squared: 0.3294

F-statistic: 25.07 on 5 and 240 DF, p-value: < 2.2e-16

Lagged Models (DW Test)

```
require(lmtest)  
dwtest(lm.KUnits)
```

Durbin-Watson test

```
data: lm.KUnits  
DW = 0.30838, p-value < 2.2e-16  
alternative hypothesis: true autocorrelation is greater than 0
```


Lagged Models

```
lm.KUnits.all = lm(KUnits~T+S.P+Q2+Q3+Q4+
                   KUnits.L1+KUnits.L2+KUnits.L3+KUnits.L4,
                   data=HousingStarts)
summary(lm.KUnits.all)
```

Call:

```
lm(formula = KUnits ~ T + S.P + Q2 + Q3 + Q4 + KUnits.L1 + KUnits.L2 +
    KUnits.L3 + KUnits.L4, data = HousingStarts)
```

Residuals:

Min	1Q	Median	3Q	Max
-27.182	-8.056	-1.276	6.263	45.123

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.121072	3.593364	2.817	0.005271	**
T	-0.019198	0.019127	-1.004	0.316561	
S.P	0.002363	0.003914	0.604	0.546513	
Q2	-0.239829	4.065405	-0.059	0.953009	
Q3	-9.567059	2.868058	-3.336	0.000991	***
Q4	-17.224210	2.434830	-7.074	1.76e-11	***
KUnits.L1	0.921704	0.074281	12.408	< 2e-16	***
KUnits.L2	0.150916	0.087870	1.717	0.087222	.
KUnits.L3	-0.263768	0.086529	-3.048	0.002568	**
KUnits.L4	0.161815	0.070622	2.291	0.022843	*

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

Residual standard error: 11.86 on 232 degrees of freedom

```
(4 observations deleted due to missingness)
Multiple R-squared:  0.8995,    Adjusted R-squared:  0.8956
F-statistic: 230.7 on 9 and 232 DF,  p-value: < 2.2e-16
```

```
dwtest(lm.KUnits.all)
```

Durbin-Watson test

```
data:  lm.KUnits.all
DW = 2.3881, p-value = 0.997
alternative hypothesis: true autocorrelation is greater than 0
```