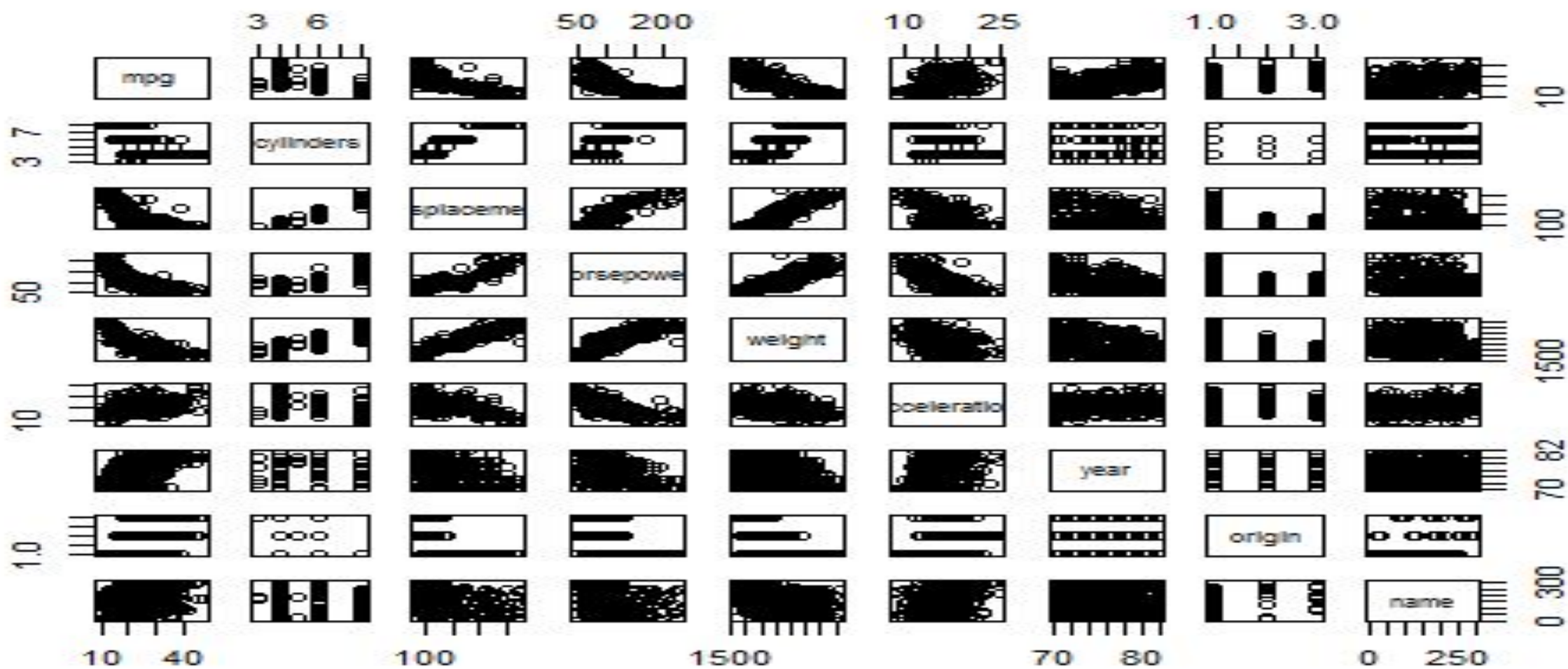


Assignment #2

20161137 복영규

#1



#2

```
> cor(Auto[,names(Auto)!="name"])
```

	mpg	cylinders	displacement	horsepower
mpg	1.0000000	-0.7776175	-0.8051269	-0.7784268
cylinders	-0.7776175	1.0000000	0.9508233	0.8429834
displacement	-0.8051269	0.9508233	1.0000000	0.8972570
horsepower	-0.7784268	0.8429834	0.8972570	1.0000000
weight	-0.8322442	0.8975273	0.9329944	0.8645377
acceleration	0.4233285	-0.5046834	-0.5438005	-0.6891955
year	0.5805410	-0.3456474	-0.3698552	-0.4163615
origin	0.5652088	-0.5689316	-0.6145351	-0.4551715

	weight	acceleration	year	origin
mpg	-0.8322442	0.4233285	0.5805410	0.5652088
cylinders	0.8975273	-0.5046834	-0.3456474	-0.5689316
displacement	0.9329944	-0.5438005	-0.3698552	-0.6145351
horsepower	0.8645377	-0.6891955	-0.4163615	-0.4551715
weight	1.0000000	-0.4168392	-0.3091199	-0.5850054
acceleration	-0.4168392	1.0000000	0.2903161	0.2127458
year	-0.3091199	0.2903161	1.0000000	0.1815277
origin	-0.5850054	0.2127458	0.1815277	1.0000000

#3-1

```
> model = lm(mpg ~ cylinders, data = Auto)
> summary(model)

Call:
lm(formula = mpg ~ cylinders, data = Auto)

Residuals:
    Min       1Q   Median       3Q      Max
-14.2413  -3.1832  -0.6332   2.5491  17.9168

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.9155     0.8349   51.40  <2e-16 ***
cylinders    -3.5581     0.1457  -24.43  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.914 on 390 degrees of freedom
Multiple R-squared:  0.6047,    Adjusted R-squared:  0.6037
F-statistic: 596.6 on 1 and 390 DF,  p-value: < 2.2e-16
```

- # Simple linear regression model (mpg ~ cylinders)
- # Cylinders have statistically significant relationship to mpg in this model.
- # $\text{mpg} = -3.56 \text{ cylinders} + 42.92$
- # 60.37% of variability of the mpg can be explained by this model.

#3-2

```
> model = lm(mpg ~ . - name, data = Auto)
> summary(model)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.218435    4.644294  -3.707  0.00024 ***
cylinders     -0.493376    0.323282  -1.526  0.12780
displacement  0.019896    0.007515   2.647  0.00844 **
horsepower    -0.016951    0.013787  -1.230  0.21963
weight        -0.006474    0.000652  -9.929  < 2e-16 ***
acceleration  0.080576    0.098845   0.815  0.41548
year          0.750773    0.050973  14.729  < 2e-16 ***
origin        1.426141    0.278136   5.127  4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

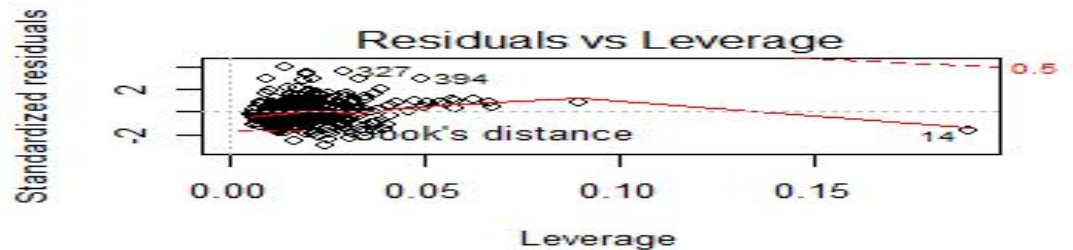
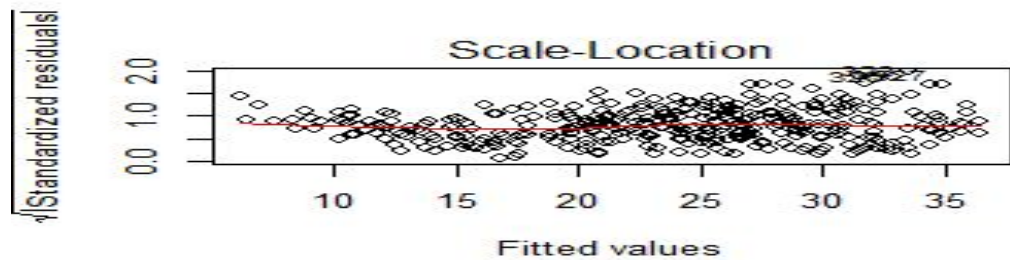
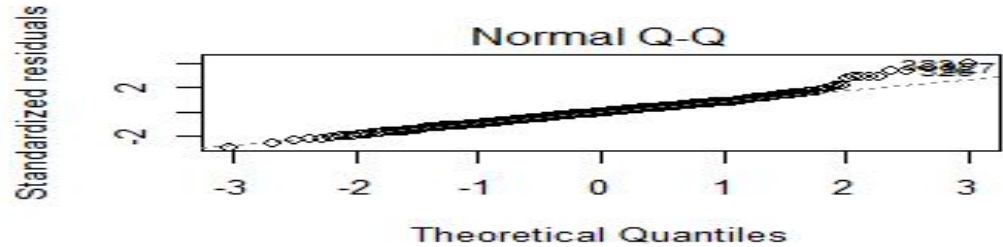
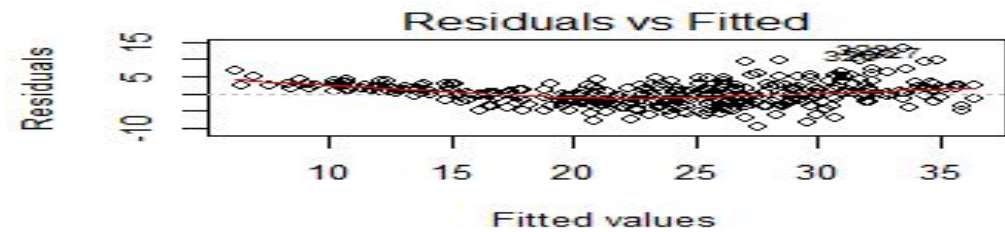
Multiple linear regression model (mpg ~ . - name)

Displacement, weight, year and origin have statistically significant relationship to mpg in this model.

$\text{mpg} = -0.49 \text{ cylinders} + 0.02 \text{ displacement} - 0.02 \text{ horsepower} - 0.01 \text{ weight} + 0.08 \text{ acceleration} + 0.75 \text{ year} + 1.43 \text{ origin} - 17.22$

81.82% of variability of the mpg can be explained by this model

#4



Residuals versus Fitted graph seems like U-shaped, that means mpg has non-linear relationship with other variables.

Residuals versus Leverage graph indicates unusually large outliers which standardized residuals value is higher than 2 or lower than -2 and unusually high leverage point which is labeled 14.

#My code

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

```
Assignment #2.R *
1 #1
2 install.packages("ISLR")
3 library("ISLR")
4 pairs(Auto)
5
6 #2
7 cor(Auto[,names(Auto)!="name"])
8
9 #3
10 model = lm(mpg ~ cylinders, data = Auto)
11 summary(model)
12 model = lm(mpg ~. -name, data = Auto)
13 summary(model)
14
15 #4
16 par(mfrow = c(2,2))
17 plot(model)
18
```

18:1 (Top Level) R Script

Console C:/Users/YoungKyu/Desktop/경명통계분석/ ↗

```
-9.5903 -2.1565 -0.1169 1.8690 13.0604
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-17.218435	4.644294	-3.707	0.00024 ***
cylinders	-0.493376	0.323282	-1.526	0.12780
displacement	0.019896	0.007515	2.647	0.00844 **
horsepower	-0.016951	0.013787	-1.230	0.21963
weight	-0.006474	0.000652	-9.929	< 2e-16 ***
acceleration	0.080576	0.098845	0.815	0.41548
year	0.750773	0.050973	14.729	< 2e-16 ***
origin	1.426141	0.278136	5.127	4.67e-07 ***

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

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8182
F-statistic: 252.4 on 7 and 384 DF, p-value: < 2.2e-16

```
> par(mfrow = c(2,2))
> plot(model)
>
```

Environment History

Global Environment

Values

model List of 13

Files Plots Packages Help Viewer

Zoom Export Publish

The figure displays four diagnostic plots for a linear model:

- Residuals vs Fitted:** A scatter plot showing residuals on the y-axis (ranging from -10 to 15) against fitted values on the x-axis (ranging from 10 to 35). The points are scattered around a horizontal red line at zero, indicating a good fit.
- Normal Q-Q:** A plot of standardized residuals on the y-axis (ranging from -2 to 2) against theoretical quantiles on the x-axis (ranging from -3 to 3). The points follow a straight line, suggesting the residuals are normally distributed.
- Scale-Location:** A plot of the square root of absolute standardized residuals on the y-axis (ranging from 0.0 to 2.0) against fitted values on the x-axis (ranging from 10 to 35). The points are scattered around a horizontal red line, indicating constant variance.
- Residuals vs Leverage:** A plot of standardized residuals on the y-axis (ranging from -2 to 2) against leverage on the x-axis (ranging from 0.00 to 0.15). The plot shows a dense cluster of points with a red line indicating the expected behavior. Points 27, 394, and 14 are highlighted.