HW 02

Group 10

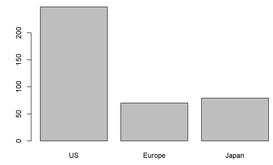
2022-10-03

Homework 2

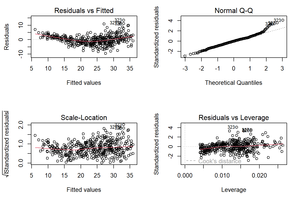
Read csv

```
auto = read.csv("Auto.csv", na.strings = "NA")
```

1(a)

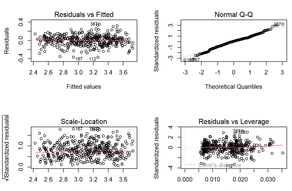


1(b)



The plots indicate that: - The error term is not normally distributed - There is variance is not constant, thus heteroskedasticity is present

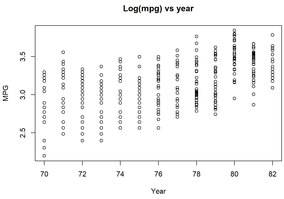
1(c)




```
##
## Call:
## lm(formula = log(mpg) ~ origin + log(weight) + year + I(year^2),
##     data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37408 -0.06782  0.00899  0.06903  0.35766
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.4693014  2.6833895   6.883 2.34e-11 ***
## originEurope  0.0668291  0.0176293   3.791 0.000174 ***
## originJapan   0.0319711  0.0179382   1.782 0.075477 .
## log(weight) -0.8750305  0.0270390 -32.362  < 2e-16 ***
## year        -0.2559684  0.0712094  -3.595 0.000366 ***
## I(year^2)    0.0019051  0.0004687   4.065 5.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1136 on 391 degrees of freedom
## Multiple R-squared:  0.8898, Adjusted R-squared:  0.8884
## F-statistic: 631.7 on 5 and 391 DF,  p-value: < 2.2e-16
```

The model assumptions seem to have been roughly satisfied now.

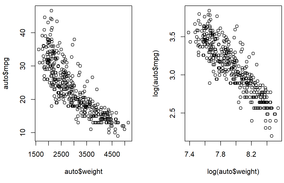
1(d)



The relationship appears U-shaped based on the plot above. The minimum is 67.1781178.

1(e)

```
##
## Call:
## lm(formula = log(mpg) ~ origin + log(weight) + year + I(year^2),
##     data = auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37408 -0.06782  0.00899  0.06903  0.35766
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.4693014  2.6833895   6.883 2.34e-11 ***
## originEurope  0.0668291  0.0176293   3.791 0.000174 ***
## originJapan   0.0319711  0.0179382   1.782 0.075477 .
## log(weight) -0.8750305  0.0270390 -32.362  < 2e-16 ***
## year        -0.2559684  0.0712094  -3.595 0.000366 ***
## I(year^2)    0.0019051  0.0004687   4.065 5.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1136 on 391 degrees of freedom
## Multiple R-squared:  0.8898, Adjusted R-squared:  0.8884
## F-statistic: 631.7 on 5 and 391 DF,  p-value: < 2.2e-16
```



It tells us that as you increase the weight the mpg falls. The relationship for the unlogged version is similar, less linear, but still negative.

2(a)

$$
\begin{gathered}
y_i = \gamma_0 + \gamma_1(x_i - \bar{x}) + \gamma_2(x_i - \bar{x})^2 + \epsilon_i = \\
\gamma_0 + y_i x_i - y_i \bar{x} + y_2 x_i^2 - 2y_2 x_i \bar{x} + \gamma_2 \bar{x}^2 + \epsilon_i = \\
(\gamma_0 - \gamma_1 \bar{x} + \gamma_2 \bar{x}^2) + (\gamma_1 - 2\gamma_2 \bar{x}) x_i + \gamma_2 x_i^2 + \epsilon_i \\
\therefore \beta_0 = \gamma_0 - \gamma_1 \bar{x} + \gamma_2 \bar{x}^2 \\
\beta_1 = \gamma_1 - 2\gamma_2 \bar{x} \\
\beta_2 = \gamma_2
\end{gathered}
$$

2(b)

```
##
## Call:
## lm(formula = mpg ~ year + year_squared, data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.349  -5.109  -0.878   4.587  18.196
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  577.25230  146.67144   3.936 9.81e-05 ***
## year         -15.84090    3.86508  -4.098 5.05e-05 ***
## year_squared   0.11230    0.02542   4.419 1.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.23 on 394 degrees of freedom
## Multiple R-squared:  0.3694, Adjusted R-squared:  0.3662
## F-statistic: 115.4 on 2 and 394 DF,  p-value: < 2.2e-16
```

2(c)

The correlation between year and year squared is 0.999759.

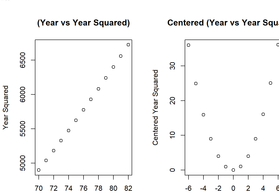
2(d)

The mean of year is 75.9949622.

2(e)

The correlation between centered year and centered year squared is 0.014414

2(f)



2(g)

```
##
## Call:
## lm(formula = mpg ~ year_centered + year_centered_squared, data = auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.349  -5.109  -0.878   4.587  18.196
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           21.99061    0.46577  47.214  < 2e-16 ***
## year_centered         -1.22778    0.08486 -14.469  < 2e-16 ***
## year_centered_squared  0.11230    0.02542   4.419 1.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.23 on 394 degrees of freedom
## Multiple R-squared:  0.3694, Adjusted R-squared:  0.3662
## F-statistic: 115.4 on 2 and 394 DF,  p-value: < 2.2e-16
```

2(h)

$$
\begin{gathered}
\beta_0 = \gamma_0 - \gamma_1 \bar{x} + \gamma_2 \bar{x}^2 \\
\beta_1 = \gamma_1 - 2\gamma_2 \bar{x} \\
\beta_2 = \gamma_2 \\
\\
\beta_0 = 577.2522975 \\
\beta_1 = -15.8409008 \\
\beta_2 = 0.1123014
\end{gathered}
$$