

# HW9

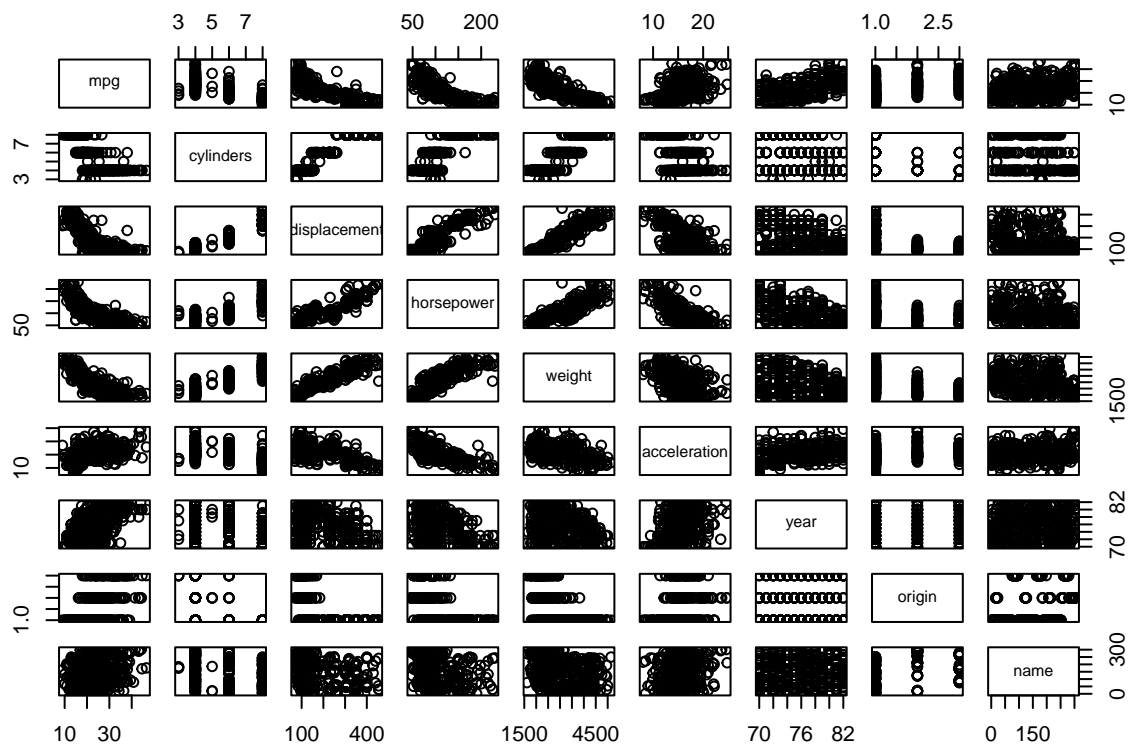
Getong Zhong

2023-04-03

## Chapter 3 Problem 9

(a)

```
pairs(Auto)
```



(b)

```
(cor_matrix <- cor(Auto[, -9]))
```

```
##          mpg  cylinders displacement horsepower      weight
## mpg          1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442
## cylinders    -0.7776175   1.0000000    0.9508233  0.8429834  0.8975273
## displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377
## weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392
## year         0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199
## origin       0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054
##          acceleration      year      origin
## mpg          0.4233285  0.5805410  0.5652088
## cylinders    -0.5046834 -0.3456474 -0.5689316
## displacement -0.5438005 -0.3698552 -0.6145351
## horsepower   -0.6891955 -0.4163615 -0.4551715
## weight       -0.4168392 -0.3091199 -0.5850054
## acceleration  1.0000000  0.2903161  0.2127458
## year         0.2903161  1.0000000  0.1815277
## origin       0.2127458  0.1815277  1.0000000
```

(c)

```
lr <- lm(mpg ~. - name , data = Auto)
summary(lr)
```

```
##
## 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
```

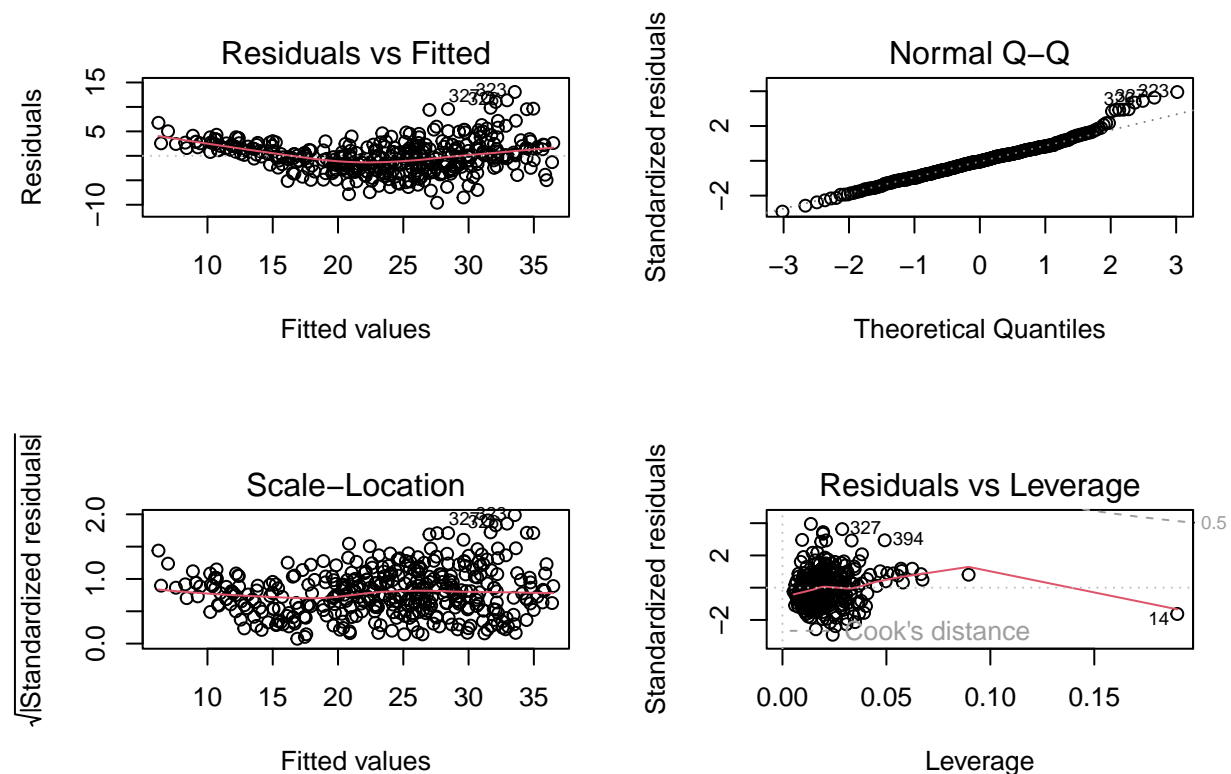
Above is the summary of the model. From the summary we can see that P-value is  $< 2.2e-16$ , which means the p-value is less than the significant level of 0.05, therefore it suggested that there is at least one relationship between the predictors and the response variable. Displacement, weight, year, and origin have a

statistically significant relationship with mpg at the 0.05 significance level. From the coefficient of year, it has the value of 0.750773, which suggest one unit increase in year will have 0.750773 unit increase in mpg.

(d)

Points like 327, 394 seems like placed outside the 2 units of the standard residuals, and should be regarded as bad leverage points. From the Normal QQ plot, we didn't observe extreme underfit of the normal distribution line. However, in the residuals plot, there is a slightly unequal variance as fitted values increases. In the standard residual plots, we didn't observe any usually large outliers.

```
par(mfrow = c(2, 2))
plot(lr)
```



(e)

weight & origin and displacement & year are statistically significant since the p-value for both interactions are less than the significant level of 0.05.

```
lr2 <- lm(mpg ~. - name + origin * weight + year * displacement + origin * year +
          origin * displacement, data = Auto)
summary(lr2)
```

##

```
## Call:
## lm(formula = mpg ~ . - name + origin * weight + year * displacement +
##      origin * year + origin * displacement, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.2492 -1.8674 -0.1427  1.5114 12.9025
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -7.156e+01  1.571e+01  -4.555 7.06e-06 ***
## cylinders     -1.138e-01  3.176e-01  -0.358  0.72038
## displacement   2.417e-01  4.793e-02   5.044 7.09e-07 ***
## horsepower    -2.571e-02  1.341e-02  -1.917  0.05599 .
## weight        -1.767e-03  1.389e-03  -1.272  0.20411
## acceleration   7.533e-02  9.259e-02   0.814  0.41637
## year          1.373e+00  2.023e-01   6.787 4.40e-11 ***
## origin         8.061e+00  5.494e+00   1.467  0.14313
## weight:origin  -2.801e-03  9.402e-04  -2.979  0.00308 **
## displacement:year -3.432e-03  6.156e-04  -5.575 4.69e-08 ***
## year:origin    -2.721e-02  7.074e-02  -0.385  0.70073
## displacement:origin 1.523e-02  1.338e-02   1.138  0.25569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.089 on 380 degrees of freedom
## Multiple R-squared:  0.8477, Adjusted R-squared:  0.8433
## F-statistic: 192.3 on 11 and 380 DF, p-value: < 2.2e-16
```

(f)

```
loglr <- lm (mpg ~. - name, data = Auto)
summary(loglr)
```

```
##
## 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
```

```
sqrtrlr <- lm(sqrt(mpg) ~. - name, data = Auto)
summary(sqrtrlr)
```

```
##
## Call:
## lm(formula = sqrt(mpg) ~ . - name, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98891 -0.18946  0.00505  0.16947  1.02581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.075e+00  4.290e-01   2.506   0.0126 *
## cylinders     -5.942e-02  2.986e-02  -1.990   0.0474 *
## displacement  1.752e-03  6.942e-04   2.524   0.0120 *
## horsepower    -2.512e-03  1.274e-03  -1.972   0.0493 *
## weight        -6.367e-04  6.024e-05 -10.570 < 2e-16 ***
## acceleration  2.738e-03  9.131e-03   0.300   0.7644
## year           7.381e-02  4.709e-03  15.675 < 2e-16 ***
## origin        1.217e-01  2.569e-02   4.735 3.09e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3074 on 384 degrees of freedom
## Multiple R-squared:  0.8561, Adjusted R-squared:  0.8535
## F-statistic: 326.3 on 7 and 384 DF,  p-value: < 2.2e-16
```

```
sqrllr <- lm(mpg^2 ~. - name, data = Auto)
summary(sqrllr)
```

```
##
## Call:
## lm(formula = mpg^2 ~ . - name, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -483.45 -141.87  -19.62  103.58 1042.84
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.878e+03  2.928e+02  -6.412 4.22e-10 ***
## cylinders    -1.436e+01  2.038e+01  -0.704  0.48157
## displacement  1.328e+00  4.738e-01   2.802  0.00534 **
## horsepower   -3.587e-01  8.693e-01  -0.413  0.68009
## weight       -3.522e-01  4.111e-02  -8.567 2.62e-16 ***
```

```
## acceleration  9.278e+00  6.232e+00   1.489  0.13740
## year          4.081e+01  3.214e+00  12.698  < 2e-16 ***
## origin        9.509e+01  1.754e+01   5.422  1.04e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.8 on 384 degrees of freedom
## Multiple R-squared:  0.7292, Adjusted R-squared:  0.7243
## F-statistic: 147.8 on 7 and 384 DF,  p-value: < 2.2e-16
```

From the above summaries for each model, we can find all of them are statistically significant as they have p-value that smaller than 0.05. In all the model summaries, variable acceraltions are not significant in all the transformations. On the other hands, Year, other, origin are all significant in all transformations.

## Chapter 4

### Question 1

From the logistic function:  $p(X) = 1 / (1 + \exp(-[\text{beta1} + \text{beta2X1} + \dots \text{betaXp}]))$  we can get:  $p(X) = 1 - p(X) * \exp(-[\text{beta1} + \text{beta2X1} + \dots \text{betaXp}])$  and then rearrange the equation we got:  $p(X) / (1 - p(X)) = \exp(\text{beta1} + \text{beta2X1} + \dots \text{betaXp})$  finally we take the log of the both sides:  $\log[p(X) / (1 - p(X))] = \text{beta1} + \text{beta2X1} + \dots \text{betaXp}$  that is the logit regression.

### Question 6

(a)  $P(Y = 1 \mid X_1 = 40, X_2 = 3.5) = 1 / (1 + \exp(-(-6 + 0.05 * 40 + 1 * 3.5))) = 0.731$  Therefore, the possibility for students who studies for 40h and has undergrad GPA of 3.5 gets an A in the class is 73.1%.

(b) set the function :  $0.5 = 1 / (1 + \exp(-(-6 + 0.05 * h + 1 * 3.5)))$  and solve for h we got,  $h = 50$ . Therefore, students in part (a) need to study 50h to have a 50 % chance of getting an A in the class.

### Question 9

(a) Recall the function:  $\text{Odd} = P(\text{default}) / (1 - P(\text{default}))$ . Given the odd is 0.37, solve for the  $P(\text{default})$  we got that  $P(\text{default}) + 0.37 * P(\text{default}) = 0.37$ ,  $P(\text{default}) = 0.37 / 1.37 = 0.27$ . Therefore, on average, 27% of people with an odds of 0.37 of defaulting on their credit card payment will in fact default.

(b) If  $P(\text{default}) = 0.16$ ,  $\text{Odd} = 0.16 / (1 - 0.16) = 0.191$ . Therefore, the odds she will be default is 0.191.