

統計學習 作業二

106070038 科管院學士班 杜葳葳

8. (a)

用 `read.csv()` 讀入 `Auto.csv`，或直接由 ISLR 讀入 `Auto` 資料集

將所有欄位的資料型別轉為 `numerical`，否則無法使用 `linear model`

以 `mpg` 為 `response`、`horsepower` 為 `predictor` 擬合 `linear model` 的結果如下：

```
Call:
lm(formula = mpg ~ horsepower, data = mydata)

Residuals:
    Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861    0.717499   55.66  <2e-16 ***
horsepower  -0.157845    0.006446  -24.49  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.906 on 390 degrees of freedom
(5 observations deleted due to missingness)
Multiple R-squared:  0.6059,    Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

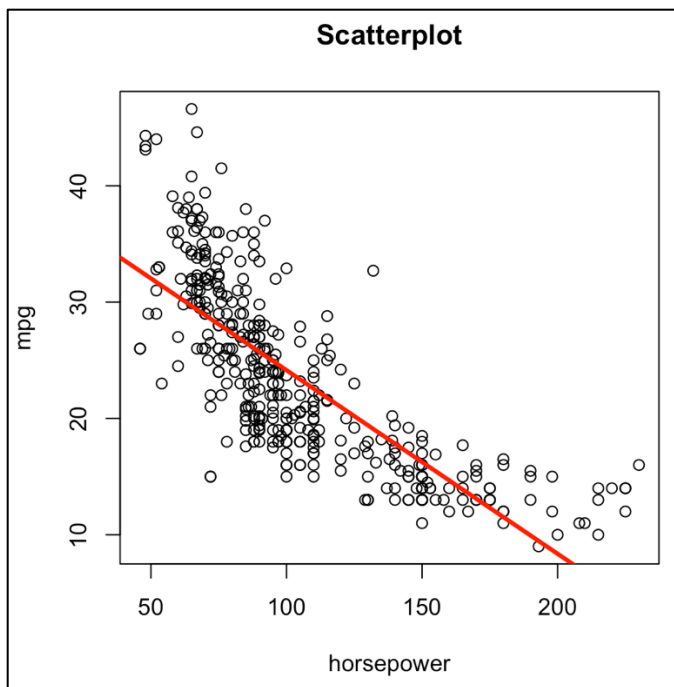
根據上面的結果，可初步得到 $\text{mpg} = 39.935861 - 0.157845 \times \text{horsepower}$ 的關係式

- 因為截距項與 `horsepower` 的 `p-value` 均遠小於 0.05，故推論在預測變數 `horsepower` 和反應變數 `mpg` 之間存在線性關係。
- `R` 平方是在評估和基線相比、擬合模型的好壞，`Multiple R-squared` 為 0.6059，代表 `mpg` 60.59% 的變化，可由模型中的自變量 `horsepower` 解釋。
- 預測變數的係數為 -0.157845，因為是負數，故可以推論預測變數與反應變數間存在負向關係。
- 在 95%信賴區間之下，當 `horsepower` = 98 時，`mpg` 擬合的值為 24.46708，信賴區間為 (23.97308, 24.96108)，預測區間為 (14.8094, 34.12476)。

	fit	lwr	upr	fit	lwr	upr
1	24.46708	23.97308	24.96108	1	24.46708	14.8094 34.12476

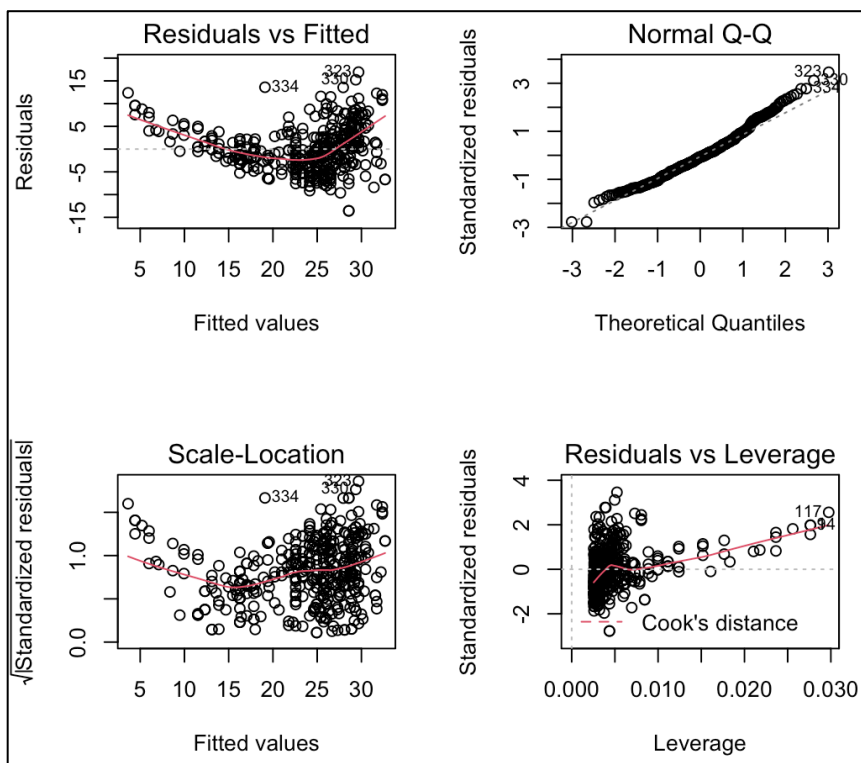
8. (b) 用 `plot()` 將剛剛擬合的線性模型畫成圖，接著用 `abline()` 將迴歸線畫出來

以 `horsepower` 為 `X`、`mpg` 為 `Y` 做散佈圖(`scatterplot`)。紅線為最小平方迴歸線，斜率為負，可觀察到 `horsepower` 和 `mpg` 大致呈現負向關係，當 `horsepower` 越大時、`mpg` 越小。



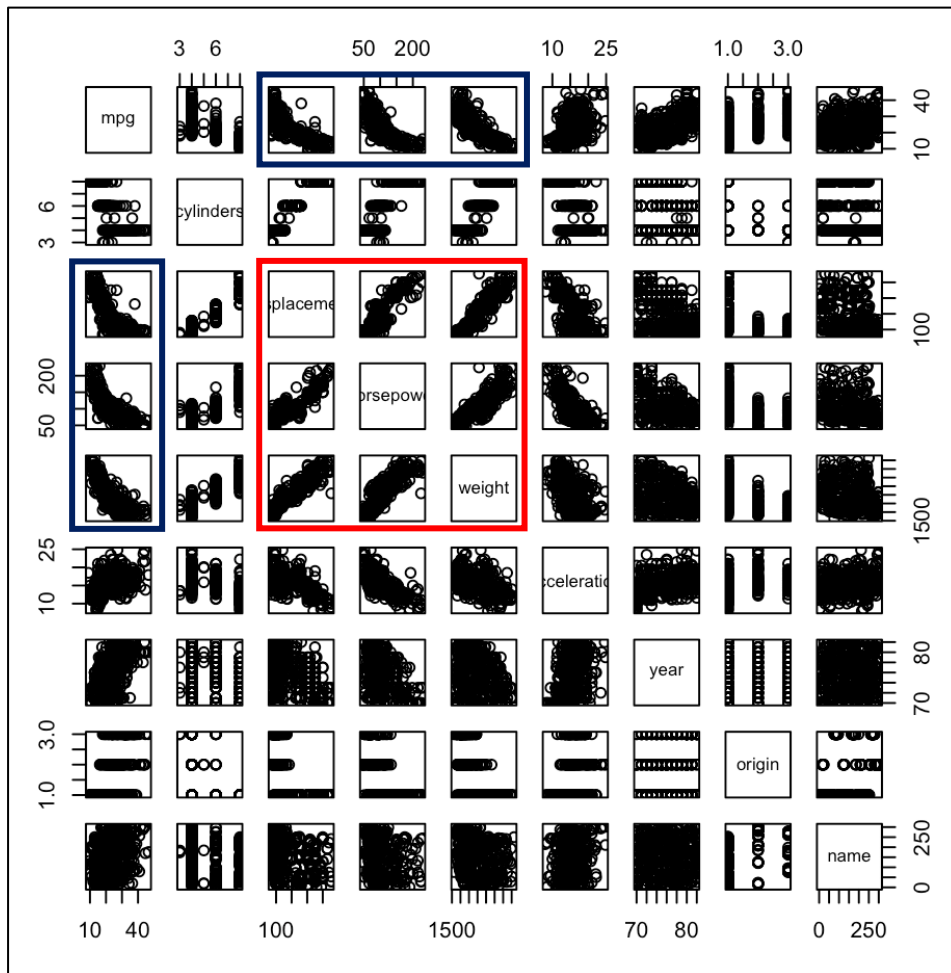
8. (c)

- 從 residuals vs fitted (左上圖) 可以觀察到殘差值與真實值略為呈現 U 型的關係，代表模型仍有可以改善的空間。在好的模型中，殘差和估計值基本是無關的，應均勻分佈在 0 的兩側。
- Normal Q-Q plot (右上圖) 是用來檢測殘差是否呈現常態分佈，從圖上可以觀察到有一些 outlier (於右上角的 323、330、334)，若要改進模型，也許可以將之刪除、或做特別處理。
- Scale Location (左下圖)顯示其具有異方差性。
- 從 Residuals vs Leverage (右下圖)殘差與槓桿圖可發現，第 94 和 117 筆資料可能是高槓桿點。



9. (a)

將所有變數（九個欄位）用 `plot()` 兩兩做出散佈圖，從圖中可以發現，displacement、horsepower、weight 兩兩之間有明顯的正向關係，而 mpg 與三者呈現反向關係



9. (b)

用 `cor()` 計算兩兩變數之間的相關係數，如 9.(a) 所觀察到的，cylinders、displacement、horsepower 兩兩之間的相關係數皆大於 0.8，呈現高度正相關，mpg 與 displacement、weight 的相關係數小於 -0.8，屬於高度負相關，mpg 和 horsepower 的相關係數為 -0.7784268，呈現中度負相關。

此外，cylinders 和 displacement、horsepower、weight 之間的相關係數皆大於 0.8，顯示其具有高度正相關。

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

9.(c) 以 mpg 對除了 name 以外的變數做多變量線性迴歸，結果如下圖：

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

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
(5 observations deleted due to missingness)
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

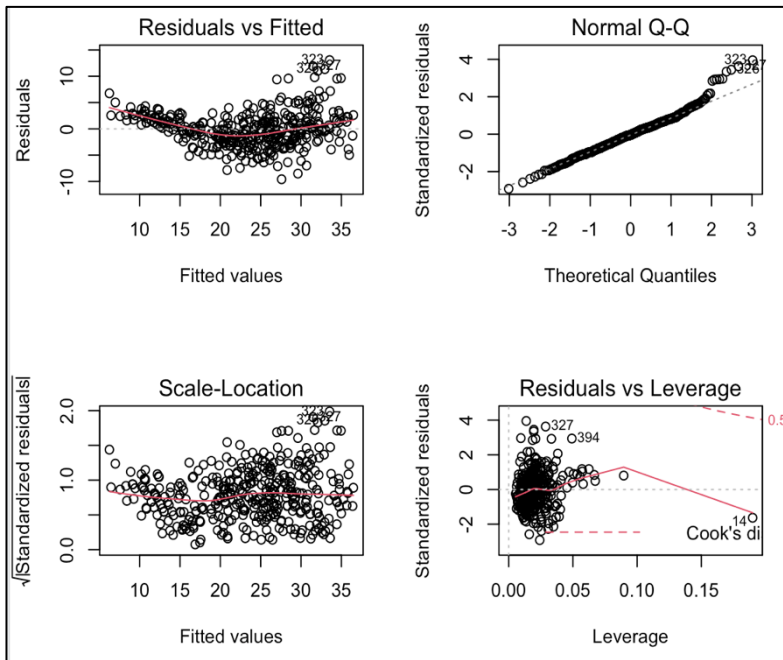
i. 有，以 P-value 來判斷，displacement、year、origin 對 mpg 具有正向關係，weight 對 mpg 則具有負向關係。

ii. 從 P-value < 0.05 判斷，origin、year、displacement、weight 具有統計顯著性。

iii. year 變數的係數具有統計顯著性，其係數為 0.750773，每增加 1 year、mpg 會增加 0.750773。

9.(d) 圖貼在下頁

- 從 residuals vs fitted (左上圖) 可以觀察到在擬合值 20-35 之間的殘差值變動較大，可能不太符合常態分佈的假設。
- Normal Q-Q plot(右上圖) 從圖上可以觀察到有一些 outlier(於右上角的 323、326、327)，不符合常態分佈的假設。
- Scale Location (左下圖) 顯示擬合值 20-35 之間具有較大的異方差性。
- 從 Residuals vs Leverage(右下圖)殘差與槓桿圖可發現，第 14 筆資料可能是高槓桿，另外，第 327、394 筆資料的 Cook's Distance 偏大。



9. (e)

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

Residuals:

	Min	1Q	Median	3Q	Max
	-7.6303	-1.4481	0.0596	1.2739	11.1386

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.548e+01	5.314e+01	0.668	0.50475
cylinders	6.989e+00	8.248e+00	0.847	0.39738
displacement	-4.785e-01	1.894e-01	-2.527	0.01192 *
horsepower	5.034e-01	3.470e-01	1.451	0.14769
weight	4.133e-03	1.759e-02	0.235	0.81442
acceleration	-5.859e+00	2.174e+00	-2.696	0.00735 **
year	6.974e-01	6.097e-01	1.144	0.25340
origin	-2.090e+01	7.097e+00	-2.944	0.00345 **
cylinders:displacement	-3.383e-03	6.455e-03	-0.524	0.60051
cylinders:horsepower	1.161e-02	2.420e-02	0.480	0.63157
cylinders:weight	3.575e-04	8.955e-04	0.399	0.69000
cylinders:acceleration	2.779e-01	1.664e-01	1.670	0.09584 .
cylinders:year	-1.741e-01	9.714e-02	-1.793	0.07389 .
cylinders:origin	4.022e-01	4.926e-01	0.816	0.41482
displacement:horsepower	-8.491e-05	2.885e-04	-0.294	0.76867
displacement:weight	2.472e-05	1.470e-05	1.682	0.09342 .
displacement:acceleration	-3.479e-03	3.342e-03	-1.041	0.29853
displacement:year	5.934e-03	2.391e-03	2.482	0.01352 *
displacement:origin	2.398e-02	1.947e-02	1.232	0.21875
horsepower:weight	-1.968e-05	2.924e-05	-0.673	0.50124
horsepower:acceleration	-7.213e-03	3.719e-03	-1.939	0.05325 .
horsepower:year	-5.838e-03	3.938e-03	-1.482	0.13916
horsepower:origin	2.233e-03	2.930e-02	0.076	0.93931
weight:acceleration	2.346e-04	2.289e-04	1.025	0.30596
weight:year	-2.245e-04	2.127e-04	-1.056	0.29182
weight:origin	-5.789e-04	1.591e-03	-0.364	0.71623
acceleration:year	5.562e-02	2.558e-02	2.174	0.03033 *
acceleration:origin	4.583e-01	1.567e-01	2.926	0.00365 **
year:origin	1.393e-01	7.399e-02	1.882	0.06062 .

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

Residual standard error: 2.695 on 363 degrees of freedom
Multiple R-squared: 0.8893, Adjusted R-squared: 0.8808
F-statistic: 104.2 on 28 and 363 DF, p-value: < 2.2e-16

以除了 name 以外的變數為單一變量、和兩兩相乘為變數做多項是回歸分析，結果如上頁，Multiple R-square 為 0.8893。

根據 $P\text{-value} < 0.05$ ，具有統計顯著性的變量有 displacement、acceleration、origin、displacement:year、acceleration:year、acceleration:origin。

9. (f)

- 將除了 name 以外的預測變數取平方，對 mpg 做迴歸

```
mpg cylinders displacement horsepower weight acceleration year origin name
1 18      8         307         130   3504         12.0   70    1 chevrolet chevelle malibu
2 15      8         350         165   3693         11.5   70    1 buick skylark 320
3 18      8         318         150   3436         11.0   70    1 plymouth satellite
4 16      8         304         150   3433         12.0   70    1 amc rebel sst
5 17      8         302         140   3449         10.5   70    1 ford torino
6 15      8         429         198   4341         10.0   70    1 ford galaxie 500
> # list apply
> Auto_2[2:8] <- lapply(Auto_2[2:8], function(x) x^2)
> colnames(Auto_2) <- c("mpg", "cylinders^2", "displacement^2", "horsepower^2", "weight^2", "acceleration^2", "year^2", "origin^2", "name")
> head(Auto_2)
```

	mpg	cylinders^2	displacement^2	horsepower^2	weight^2	acceleration^2	year^2	origin^2	name
1	18	64	94249	16900	12278016	144.00	4900	1	chevrolet chevelle malibu
2	15	64	122500	27225	13638249	132.25	4900	1	buick skylark 320
3	18	64	101124	22500	11806096	121.00	4900	1	plymouth satellite
4	16	64	92416	22500	11785489	144.00	4900	1	amc rebel sst
5	17	64	91204	19600	11895601	110.25	4900	1	ford torino
6	15	64	184041	39204	18844281	100.00	4900	1	ford galaxie 500

結果如下：

```
Call:
lm(formula = mpg ~ (. - name), data = Auto_2)

Residuals:
    Min       1Q   Median       3Q      Max
-9.6786 -2.3227 -0.0582  1.9073 12.9807

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.208e+00  2.356e+00   0.513  0.608382
`cylinders^2` -8.829e-02  2.521e-02  -3.502  0.000515 ***
`displacement^2` 5.680e-05  1.382e-05   4.109  4.87e-05 ***
`horsepower^2` -3.621e-05  4.975e-05  -0.728  0.467201
`weight^2` -9.351e-07  8.978e-08 -10.416 < 2e-16 ***
`acceleration^2` 6.278e-03  2.690e-03   2.334  0.020130 *
`year^2` 4.999e-03  3.530e-04  14.160 < 2e-16 ***
`origin^2` 4.129e-01  6.914e-02   5.971  5.37e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.539 on 384 degrees of freedom
Multiple R-squared:  0.7981,    Adjusted R-squared:  0.7944
F-statistic: 216.8 on 7 and 384 DF,  p-value: < 2.2e-16
```

與 9.(c)相比，Multiple R-squared 和 Adjusted R-squared 都下降了，因此推論此資料集不適合取平方後的迴歸模型。

● 將除了 name 以外的預測變數取 log，對 mpg 做迴歸

	mpg	log(cylinders)	log(displacement)	log(horsepower)	log(weight)	log(acceleration)	log(year)	log(origin)	name
1	18	2.079442	5.726848	4.867534	8.161660	2.484907	4.248495	0	chevrolet chevelle malibu
2	15	2.079442	5.857933	5.105945	8.214194	2.442347	4.248495	0	buick skylark 320
3	18	2.079442	5.762051	5.010635	8.142063	2.397895	4.248495	0	plymouth satellite
4	16	2.079442	5.717028	5.010635	8.141190	2.484907	4.248495	0	amc rebel sst
5	17	2.079442	5.710427	4.941642	8.145840	2.351375	4.248495	0	ford torino
6	15	2.079442	6.061457	5.288267	8.375860	2.302585	4.248495	0	ford galaxie 500

結果如下：

```
Call:
lm(formula = mpg ~ (. - name), data = Auto_3)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5987 -1.8172 -0.0181  1.5906 12.8132

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -66.5643    17.5053  -3.803 0.000167 ***
cylinders      1.4818     1.6589   0.893 0.372273
displacement  -1.0551     1.5385  -0.686 0.493230
horsepower    -6.9657     1.5569  -4.474 1.01e-05 ***
weight       -12.5728     2.2251  -5.650 3.12e-08 ***
acceleration  -4.9831     1.6078  -3.099 0.002082 **
year          54.9857     3.5555 15.465 < 2e-16 ***
origin        1.5822     0.5083   3.113 0.001991 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.069 on 384 degrees of freedom
Multiple R-squared:  0.8482,    Adjusted R-squared:  0.8454
F-statistic: 306.5 on 7 and 384 DF,  p-value: < 2.2e-16
```

與 9.(c)相比，Multiple R-squared 和 Adjusted R-squared 都上升了，因此推論若先取 log 再做迴歸模型，會比直接一次方做迴歸模型更加適配。

● 除了 name 以外的預測變數開根號，對 mpg 做迴歸

	mpg	sqrt(cylinders)	sqrt(displacement)	sqrt(horsepower)	sqrt(weight)	sqrt(acceleration)	sqrt(year)	sqrt(origin)	name
1	18	2.828427	17.52142	11.40175	59.19459	3.464102	8.3666	1	chevrolet chevelle malibu
2	15	2.828427	18.70829	12.84523	60.77006	3.391165	8.3666	1	buick skylark 320
3	18	2.828427	17.83255	12.24745	58.61740	3.316625	8.3666	1	plymouth satellite
4	16	2.828427	17.43560	12.24745	58.59181	3.464102	8.3666	1	amc rebel sst
5	17	2.828427	17.37815	11.83216	58.72819	3.240370	8.3666	1	ford torino
6	15	2.828427	20.71232	14.07125	65.88627	3.162278	8.3666	1	ford galaxie 500

結果如下：

```
Call:
lm(formula = mpg ~ (. - name), data = Auto_4)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5250 -1.9822 -0.1111  1.7347 13.0681

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -49.79814    9.17832  -5.426 1.02e-07 ***
`sqrt(cylinders)`  -0.23699    1.53753  -0.154  0.8776
`sqrt(displacement)`  0.22580    0.22940   0.984  0.3256
`sqrt(horsepower)` -0.77976    0.30788  -2.533  0.0117 *
`sqrt(weight)`    -0.62172    0.07898  -7.872 3.59e-14 ***
`sqrt(acceleration)` -0.82529    0.83443  -0.989  0.3233
`sqrt(year)`      12.79030    0.85891  14.891 < 2e-16 ***
`sqrt(origin)`     3.26036    0.76767   4.247 2.72e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.21 on 384 degrees of freedom
Multiple R-squared:  0.8338,    Adjusted R-squared:  0.8308
F-statistic: 275.3 on 7 and 384 DF,  p-value: < 2.2e-16
```

與 9.(c)相比，**Multiple R-squared** 和 **Adjusted R-squared** 都上升了，因此推論若先開根號再做迴歸模型，會比直接一次方做迴歸模型更加適配。

總結來說，三個方法相比，適配性由佳到差：取 log > 開根號 > 一次方 > 平方

10.(a)

以 Price、Urban、US 做為預測變數，對 Sales 做的多變量線性迴歸模型結果如下：

```
Call:
lm(formula = Sales ~ Price + Urban + US, data = Carseats)

Residuals:
    Min       1Q   Median       3Q      Max
-6.9206 -1.6220 -0.0564  1.5786  7.0581

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.043469    0.651012  20.036 < 2e-16 ***
Price        -0.054459    0.005242 -10.389 < 2e-16 ***
UrbanYes     -0.021916    0.271650  -0.081  0.936
USYes        1.200573    0.259042   4.635 4.86e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.472 on 396 degrees of freedom
Multiple R-squared:  0.2393,    Adjusted R-squared:  0.2335
F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

Multiple R-squared 為 0.2393，模式可解釋反應變數的比例僅 23.93%。

10.(b)

Urban 和 US 的資料型別為 qualitative。

- 觀察 P-value 可發現，Price 與 US 具有統計顯著性，Price 的係數為 -0.054459，表示當 Price 增加一單位，Sales 便會下降 0.054459 單位。
- US 的係數為 1.200573，表示當商店開設在美國，Sales 便會上升 1.200573 單位。
- Urban 並無統計顯著性，係數為 -0.021916，與 Sales 呈現負向關係。

10.(c)

$$\text{Sales} = 13.043469 - 0.054459 \times \text{Price} - 0.021916 \times \text{Urban} + 1.200573 \times \text{US}$$

其中 Urban 與 US 分別以 1 和 0 代表商店是與不是來自 Urban 或 US。

10.(d)

Urban 的 P-value 為 0.936，在 95% 的信心水準下，不具有統計顯著性，故 reject H_0 。

10.(e)

將 Urban 移除重新跑迴歸模型，新模型中所有的變數均具有統計顯著性，結果如下：

```
Call:
lm(formula = Sales ~ Price + US, data = Carseats)

Residuals:
    Min       1Q   Median       3Q      Max
-6.9269 -1.6286 -0.0574  1.5766  7.0515

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.03079    0.63098   20.652 < 2e-16 ***
Price       -0.05448    0.00523  -10.416 < 2e-16 ***
USYes        1.19964    0.25846   4.641 4.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.469 on 397 degrees of freedom
Multiple R-squared:  0.2393,    Adjusted R-squared:  0.2354
F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

10.(f)

將(a)與(e)兩個模型相比，Multiple R-squared 皆相同，然而，(e)的 Adjusted R-squared 較(a)為大，故(e)的適配程度較高。

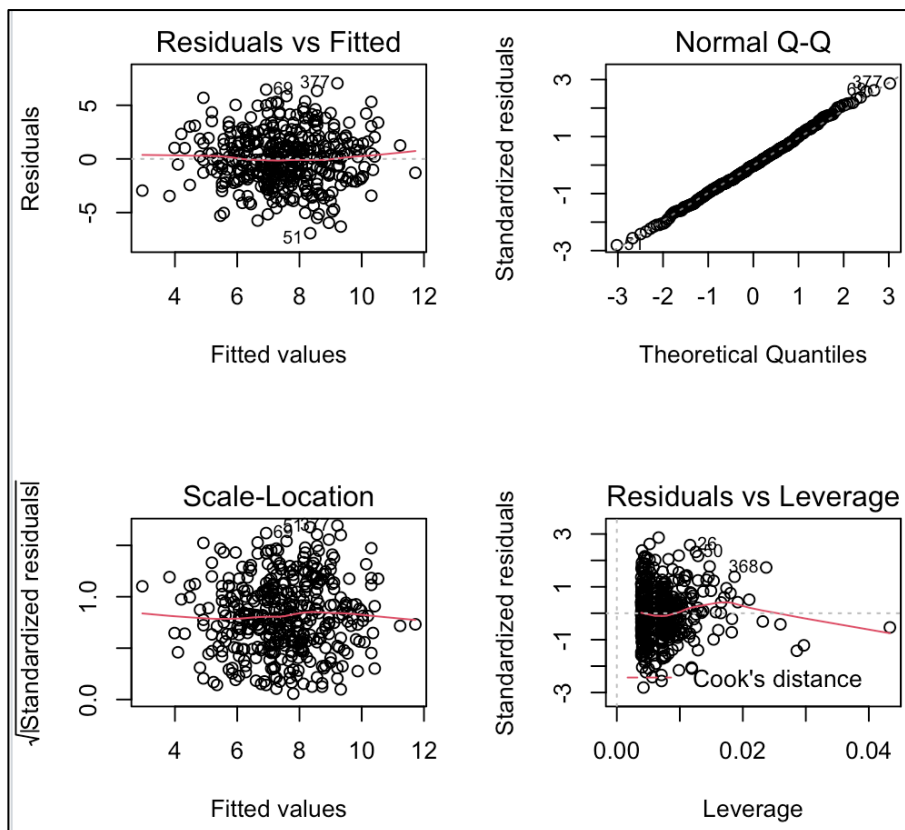
10.(g)

使用 `confint()` 來計算 95%信賴區間，截距項的信賴區間為 $[11.79032020, 14.27126531]$ ，Price 的信賴區間為 $[-0.06475984, -0.04419543]$ ，US 的信賴區間為 $[0.69151957, 1.70776632]$ 。

	2.5 %	97.5 %
(Intercept)	11.79032020	14.27126531
Price	-0.06475984	-0.04419543
USYes	0.69151957	1.70776632

10.(h)

- 從 residuals vs fitted (左上圖) 可以觀察殘差大致上至常態分佈，該組資料具有離群值第 51、69、377 筆資料。
- 從 Normal Q-Q plot (右上圖) 顯示出整個模式大致上符合常態假設，離群值為第 51、69、377 筆資料。
- Scale Location (左下圖) 方差應大致符合定值的假設。
- 從 Residuals vs Leverage (右下圖) 殘差與槓桿圖可發現，整體而言並無資料槓桿值偏大，但第 26、50、368 筆資料可能有較大的 Cook's Distance。



附錄：R 語言程式碼

```
library (graphics)
library (MASS)
library (ISLR)
data (Auto)
summary (Auto)

# mydata <- read.csv (file = "/Users/wei/Desktop/Auto.csv")
# mydata$origin = as.numeric (mydata$origin)
# mydata$horsepower = as.numeric (mydata$horsepower)

lm.fit = lm (mpg~horsepower, data=Auto)
summary (fit)
## 8.(a) iv.
predict (lm.fit, newdata = data.frame (horsepower = c(98)), interval = "confidence")
predict (lm.fit, newdata = data.frame (horsepower = c(98)), interval = "prediction")
# 8.(b)
plot (Auto$horsepower, Auto$mpg, main = "Scatterplot", ylab = "mpg", xlab = "horsepower")
abline (lm.fit, lwd = 3, col = "red")
# 8.(c)
par (mfrow = c(2,2))
plot (lm.fit)
## 9.(a)
plot (Auto)
# 9.(b)
cor (Auto[, !(names (Auto) == "name" )])
# 9.(c)
lm.fit = lm (mpg~.-name, data = Auto)
summary (lm.fit)
# 9.(d)
par (mfrow = c(2,2))
plot (lm.fit)
# 9.(e)
lm.fit.inter = lm (mpg ~ (.-name) * (.-name), data = Auto)
summary (lm.fit.inter)
# 9.(f)
Auto_2 <- Auto
head (Auto_2)
# 平方
Auto_2[2:8] <- lapply (Auto_2 [2:8], function(x) x^2)
colnames (Auto_2) <- c ("mpg", "cylinders^2", "displacement^2", "horsepower^2", "weight^2",
"acceleration^2", "year^2", "origin^2", "name")
head (Auto_2)
lm.fit = lm (mpg ~ (.-name), data = Auto_2)
summary (lm.fit)
```

```

Auto_3 <- Auto
# 取 log
Auto_3[2:8] <- lapply (Auto_3[2:8], function(x) log(x))
colnames (Auto_3) <- c ("mpg", "log(cylinders)", "log(displacement)", "log(horsepower)", "log(weight)",
"log(acceleration)", "log(year)", "log(origin)","name")
head (Auto_3)
lm.fit = lm (mpg ~ (.-name), data = Auto_3)
summary (lm.fit)

Auto_4 <- Auto
# 開根號
Auto_4[2:8] <- lapply (Auto_4[2:8], function(x) sqrt(x))
colnames (Auto_4) <- c ("mpg", "sqrt(cylinders)", "sqrt(displacement)", "sqrt(horsepower)", "sqrt(weight)",
"sqrt(acceleration)", "sqrt(year)", "sqrt(origin)","name")
head (Auto_4)
lm.fit = lm (mpg ~ (.-name), data = Auto_4)
summary (lm.fit)

## 10.(a)
library (ISLR)
summary (Carseats)
lm.fit = lm (Sales ~ Price + Urban + US, data = Carseats)
summary (lm.fit)
# 10.(e)
lm.fit = lm (Sales ~ Price + US, data = Carseats)
summary (lm.fit)
# 10.(g)
confint (lm.fit)
# 10.(h)
plot (lm.fit)

```