

# MLR hw2

임지윤

## Q1. 매장별 유아 카시트 판매액 예측

dt.csv 데이터를 이용하여 회귀모형을 적합하려고 한다. 이는 매장별 유아 카시트 판매액(Sales)를 예측하기 위한 데이터이다. 다음 물음에 답하여라. (R을 이용하여 풀이)(검정에서는 유의수준 = 0.05 사용)

Variable name	Description
Sales	Unit sales (in thousands) at each location
CompPrice	Price charged by competitor at each location
Income	Community income level (in thousands of dollars)
Advertising	Local advertising budget for company at each location (in thousands of dollars)
Population	Population size in region (in thousands)
Price	Price company charges for car seats at each site
Age	Average age of the local population
Education	Education level at each location

```
library(tidyverse)
```

```
df = read_csv('./dt.csv')  
str(df)
```

Rows: 400 Columns: 8

— Column specification —

Delimiter: ","

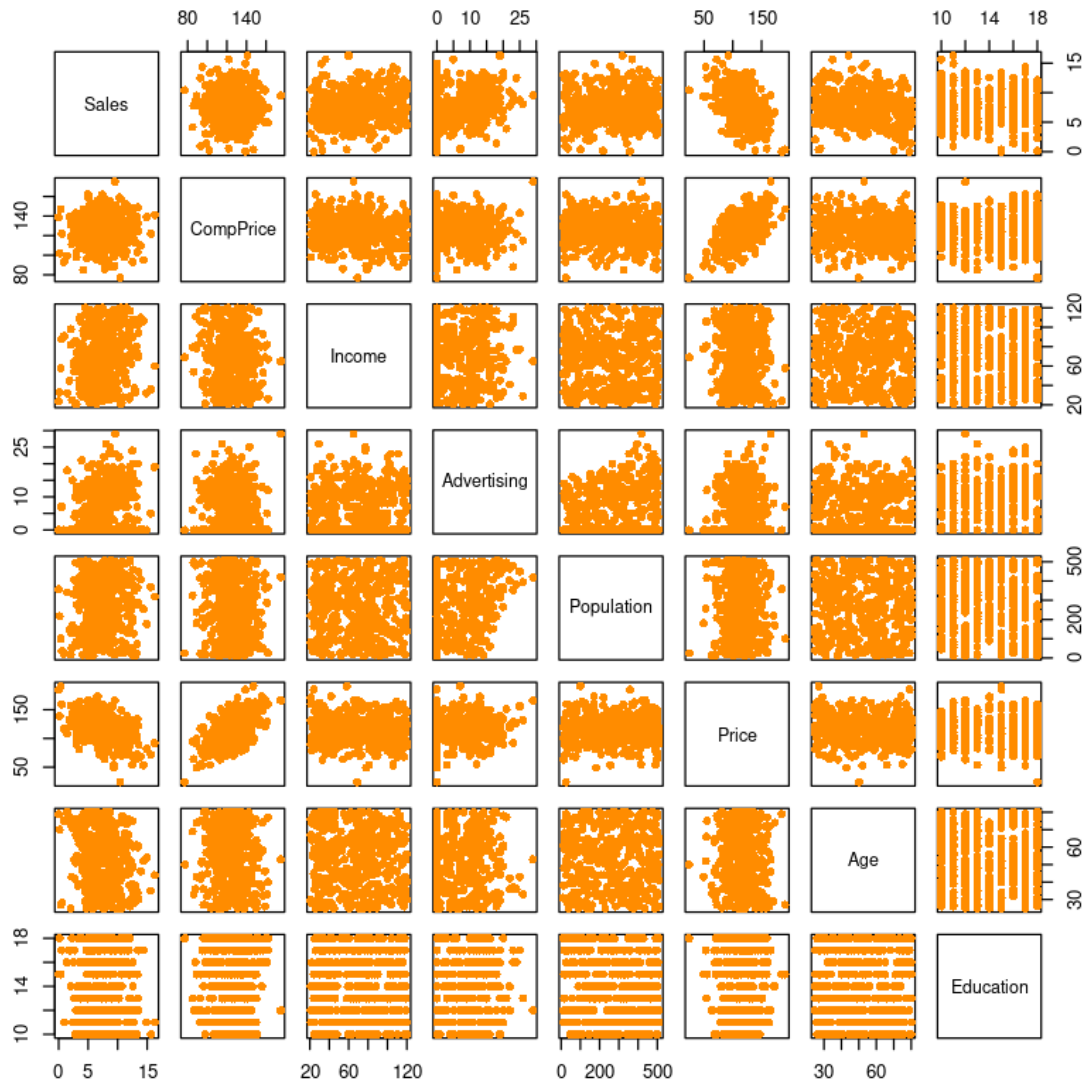
dbl (8): Sales, CompPrice, Income, Advertising, Population, Price, Age, Educ...

Use `spec()` to retrieve the full column specification for this data.  
Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
spc_tbl_ [400 × 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Sales      : num [1:400] 9.5 11.22 10.06 7.4 4.15 ...
 $ CompPrice  : num [1:400] 138 111 113 117 141 124 115 136 132 132 ...
 $ Income     : num [1:400] 73 48 35 100 64 113 105 81 110 113 ...
 $ Advertising: num [1:400] 11 16 10 4 3 13 0 15 0 0 ...
 $ Population : num [1:400] 276 260 269 466 340 501 45 425 108 131 ...
 $ Price      : num [1:400] 120 83 80 97 128 72 108 120 124 124 ...
 $ Age        : num [1:400] 42 65 59 55 38 78 71 67 76 76 ...
 $ Education  : num [1:400] 17 10 12 14 13 16 15 10 10 17 ...
- attr(*, "spec")=
 .. cols(
 ..   Sales = col_double(),
 ..   CompPrice = col_double(),
 ..   Income = col_double(),
 ..   Advertising = col_double(),
 ..   Population = col_double(),
 ..   Price = col_double(),
 ..   Age = col_double(),
 ..   Education = col_double()
 .. )
- attr(*, "problems")=<externalptr>
```

(1) 이 데이터의 산점도 행렬을 그리시오.

```
pairs(df, pch=16, col='darkorange')
```



(2) Sales를 예측하기 위한 중회귀분석을 하려고 한다. 이를 위한 모형을 설정하시오.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_6 x_{i6} + \epsilon_i, \quad i = 1, 2, \dots, 400$$

- 설명변수(독립변수):

$X_1 = \text{CompPrice}, X_2 = \text{Income}, X_3 = \text{Advertising}, X_4 = \text{Population}, X_5 = \text{Price}, X_6 = \text{Education}$

- 반응변수(종속변수):  $y = Sales$
- 오차항:  $\epsilon_1, \epsilon_2, \dots, \epsilon_{400}, \quad (\sim_{i.i.d} N(0, \sigma^2))$

(3) 최소제곱법의 의한 회귀직선을 적합시키시키고, 모형 적합 결과를 설명하시오.

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots \beta_6 x_{i6}))^2$$

$$\text{최소제곱추정량: } (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_6) = \underset{(\beta_0, \beta_1, \dots, \beta_6) \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_{i1} + \dots \beta_6 x_{i6})\}^2$$

모형적합 (함수이용)

```
# df[,which(!(names(df) %in% c('Age')))]
```

```
fit_Sales <- lm(Sales ~ .-Age, data=df)
# fit_Sales <- lm(Sales ~ CompPrice + Income + Advertising + Population + Price + Education)
summary(fit_Sales)
```

Call:

```
lm(formula = Sales ~ . - Age, data = df)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-5.9921	-1.4004	-0.1145	1.2378	5.0112

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.5970091	1.1077522	4.150	4.08e-05	***
CompPrice	0.0971849	0.0083574	11.629	< 2e-16	***
Income	0.0132566	0.0037106	3.573	0.000397	***
Advertising	0.1295888	0.0161449	8.027	1.17e-14	***
Population	0.0001374	0.0007334	0.187	0.851487	
Price	-0.0905544	0.0053869	-16.810	< 2e-16	***
Education	-0.0407910	0.0396398	-1.029	0.304093	

---

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

Residual standard error: 2.06 on 393 degrees of freedom

Multiple R-squared: 0.4761, Adjusted R-squared: 0.4681

F-statistic: 59.54 on 6 and 393 DF, p-value: < 2.2e-16

## (결과해석)

$$\begin{aligned}\hat{y} &= \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5 + \hat{\beta}_6 x_6 \\ &= 4.5970 + 0.09718x_1 + 0.01325x_2 + 0.1295x_3 + 0.0001x_4 - 0.0905x_5 - 0.0407x_6\end{aligned}$$

- $x_1 = \text{CompPrice}$ ,  $x_2 = \text{Income}$ ,  $x_3 = \text{Advertising}$ ,  $x_4 = \text{Population}$ ,  $x_5 = \text{Price}$ ,  $x_6 = \text{Education}$
- 다른 변수값이 일정할 때, 경쟁사 가격(CompPric)이 1단위 증가하면 총 판매액(Sales)이 평균적으로 0.09718 만큼증가한다.
- 다른 변수값이 일정할 때, 소득수준(Income)이 1단위(1000달러) 증가하면 총 판매액(Sales)이 평균적으로 0.01325 만큼 증가한다.
- 다른 변수값이 일정할 때, 광고비(Advertising)가 1단위(1000달러) 증가하면 총 판매액(Sales)이 평균적으로 0.1295 만큼 증가한다.
- 다른 변수값이 일정할 때, 사이트 별 카시트 가격(Price)이 1단위 증가하면 총 판매액(Sales)이 평균적으로 0.0905 만큼 감소한다.
- $R^2_{\text{Adj}} = 0.4681$  \$로 적합된 회귀모형이 총 변동의 46.81% 설명하고 있다.
- 회귀모형의 유의성 검정(F-test) 결과  $p\text{-value} < 2.2e-16$  으로 매우 작으므로 회귀모형이 유의하지 않다는 귀무가설을 기각할 수 있다.
- 개별 회귀계수의 유의성 검정(t-test) 결과 유의수준 5% 하에서 CompPrice, Income, Advertising, Price 가 유의하다.

## Matrix

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} \Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

```
n = nrow(df)
X = cbind(rep(1,n), df$CompPrice, df$Income, df$Advertising, df$Population, df$Price, df$Education)
y = df$Sales
```

```
beta_hat = solve(t(X)%*%X)%*% t(X) %*% y
beta_hat
```

A matrix: 7 × 1 of type dbl

```
4.5970090863 |
0.0971849195 |
0.0132566345 |
0.1295887568 |
```

```
0.0001373898 |
-0.0905544248 |
-0.0407909986 |
```

- lm을 이용하나 행렬을 이용해서 하나 동일한 결과를 얻음을 확인할 수 있다.

(4) 회귀직선의 유의성 검정을 위한 가설을 설정하고, 분산분석표를 이용하여 가설 검정을 수행하시오.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_6 = 0$$

$$H_1 : \text{not } H_0 \quad (\text{최소한 하나의 } \beta_i \neq 0, \quad i = 1, 2, \dots, 6)$$

```
df2 = df[,which(!(names(df) %in% c('Age')))]
colnames(df2)
```

1. 'Sales'
2. 'CompPrice'
3. 'Income'
4. 'Advertising'
5. 'Population'
6. 'Price'
7. 'Education'

```
anova(fit_Sales)
```

A anova: 7 × 5

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
CompPrice	1	13.0666859	13.0666859	3.08043028	8.001871e-02
Income	1	79.0733616	79.0733616	18.64129736	1.999655e-05
Advertising	1	219.3512681	219.3512681	51.71137454	3.269539e-12
Population	1	0.3824026	0.3824026	0.09015021	7.641450e-01
Price	1	1198.8668836	1198.8668836	282.62911342	3.569783e-48
Education	1	4.4918032	4.4918032	1.05892853	3.040926e-01
Residuals	393	1667.0422928	4.2418379	NA	NA

$$H_0 : y = \beta_0 \cdot 1,$$

$$H_1 : \beta_0 \cdot 1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

```

null_model <- lm(Sales ~ 1, data=df2) # H0
fit_Sales <- lm(Sales ~ ., data=df2) # H1

anova(null_model, fit_Sales)

```

A anova: 2 × 6

	Res.Df			Sum of Sq		Pr(>F)
	<dbl>	RSS <dbl>	Df <dbl>	<dbl>	F <dbl>	<dbl>
1	399	3182.275	NA	NA	NA	NA
2	393	1667.042	6	1515.232	59.53522	3.000393e-52

p-value가 매우작은 값으로 유의수준 5% 하에서 적합한 회귀모형은 유의하다.

```

400-1 # null model: 잔차항의 자유도 (n-p)
400-6-1 # Full model: 잔차항의 자유도 (n-p-1)

```

399

393

(5) 오차의 분산에 대한 추정량을 구하시오.

$$SSE = \sum (y_i - \hat{y}_i)^2, \quad RMSE = \sqrt{\frac{SSE}{n - p - 1}} = \hat{\sigma}$$

```

summary(fit_Sales)$sigma # RMSE
(summary(fit_Sales)$sigma)^2 # MSE

```

2.05957226026746

4.24183789526319

## matrix

```
n = nrow(df2)
X = as.matrix(cbind(rep(1,n), df2 |> select(2:7)))
#names(X)[names(X)=='rep(1, n)'] <- c('1')
y = df2$Sales

beta_hat = solve(t(X) %*% X) %*% t(X) %*% y
y_hat = X %*% beta_hat
sse <- sum((y-y_hat)^2)
sqrt(sse/(n-6-1)) # p=6, RMSE
sse/(n-6-1)
```

2.05957226026746

4.2418378952632

(6) 결정계수와 수정된 결정계수를 구하시오.

```
summary(fit_Sales)
```

Call:

```
lm(formula = Sales ~ ., data = df2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-5.9921	-1.4004	-0.1145	1.2378	5.0112

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.5970091	1.1077522	4.150	4.08e-05	***
CompPrice	0.0971849	0.0083574	11.629	< 2e-16	***
Income	0.0132566	0.0037106	3.573	0.000397	***
Advertising	0.1295888	0.0161449	8.027	1.17e-14	***
Population	0.0001374	0.0007334	0.187	0.851487	
Price	-0.0905544	0.0053869	-16.810	< 2e-16	***
Education	-0.0407910	0.0396398	-1.029	0.304093	

---

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



Residual standard error: 2.06 on 393 degrees of freedom  
 Multiple R-squared: 0.4761, Adjusted R-squared: 0.4681  
 F-statistic: 59.54 on 6 and 393 DF, p-value: < 2.2e-16

- $R^2 = 0.4761$ ,  $\text{adj-}R^2 = 0.4681$

(7) 개별 회귀계수의 유의성검정을 수행하시오.

```
summary(fit_Sales)$coef
```

A matrix: 7 × 4 of type dbl

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.5970090863	1.1077522094	4.1498532	4.081124e-05
CompPrice	0.0971849195	0.0083574142	11.6285872	4.587518e-27
Income	0.0132566345	0.0037106386	3.5726019	3.973365e-04
Advertising	0.1295887568	0.0161449482	8.0265824	1.171428e-14
Population	0.0001373898	0.0007333539	0.1873444	8.514873e-01
Price	-0.0905544248	0.0053869071	-16.8100958	3.622042e-48
Education	-0.0407909986	0.0396397597	-1.0290425	3.040926e-01

유의수준 5% 하에서 개별 회귀계수의 유의성 검정 결과 CompPrice, Income, Advertising, Price 변수가 유의하게 나타났다.

(8) 회귀계수에 대한 90% 신뢰구간을 구하시오.

```
confint(fit_Sales, level=0.90)
```

A matrix: 7 × 2 of type dbl

	5 %	95 %
(Intercept)	2.770613583	6.42340459
CompPrice	0.083405716	0.11096412
Income	0.007138756	0.01937451
Advertising	0.102969933	0.15620758
Population	-0.001071720	0.00134650
Price	-0.099436035	-0.08167281

	5 %	95 %
Education	-0.106146660	0.02456466

**(9) CompPrice = 100, Income = 70, Advertising = 20, Population = 300, Price = 80, Education = 12**

인 지역에 위치한 매장의 평균 판매액을 예측하고, 95% 신뢰구간을 구하시오.

```
new_df = data.frame(CompPrice=100, Income=70, Advertising = 20, Population = 300, Price = 80, Education = 12)
```

A data.frame: 1 × 6

CompPrice <dbl>	Income <dbl>	Advertising <dbl>	Population <dbl>	Price <dbl>	Education <dbl>
100	70	20	300	80	12

```
predict(fit_Sales,
        newdata = new_df,
        interval = c('confidence'),
        level = 0.95) ## 평균반응
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	10.14261	9.537128	10.7481

**(10) 위 매장에 대하여 개별 판매액 예측하고, 95% 신뢰구간을 구하시오.**

```
predict(fit_Sales,
        newdata = new_df,
        interval = c('prediction'),
        level = 0.95)
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	10.14261	6.048434	14.23679

(11) 잔차에 대한 산점도를 그리고, 결과를 설명하여라.

```
library(ggplot2)
library(lmtest) ## bptest, dwtest
```

Loading required package: zoo

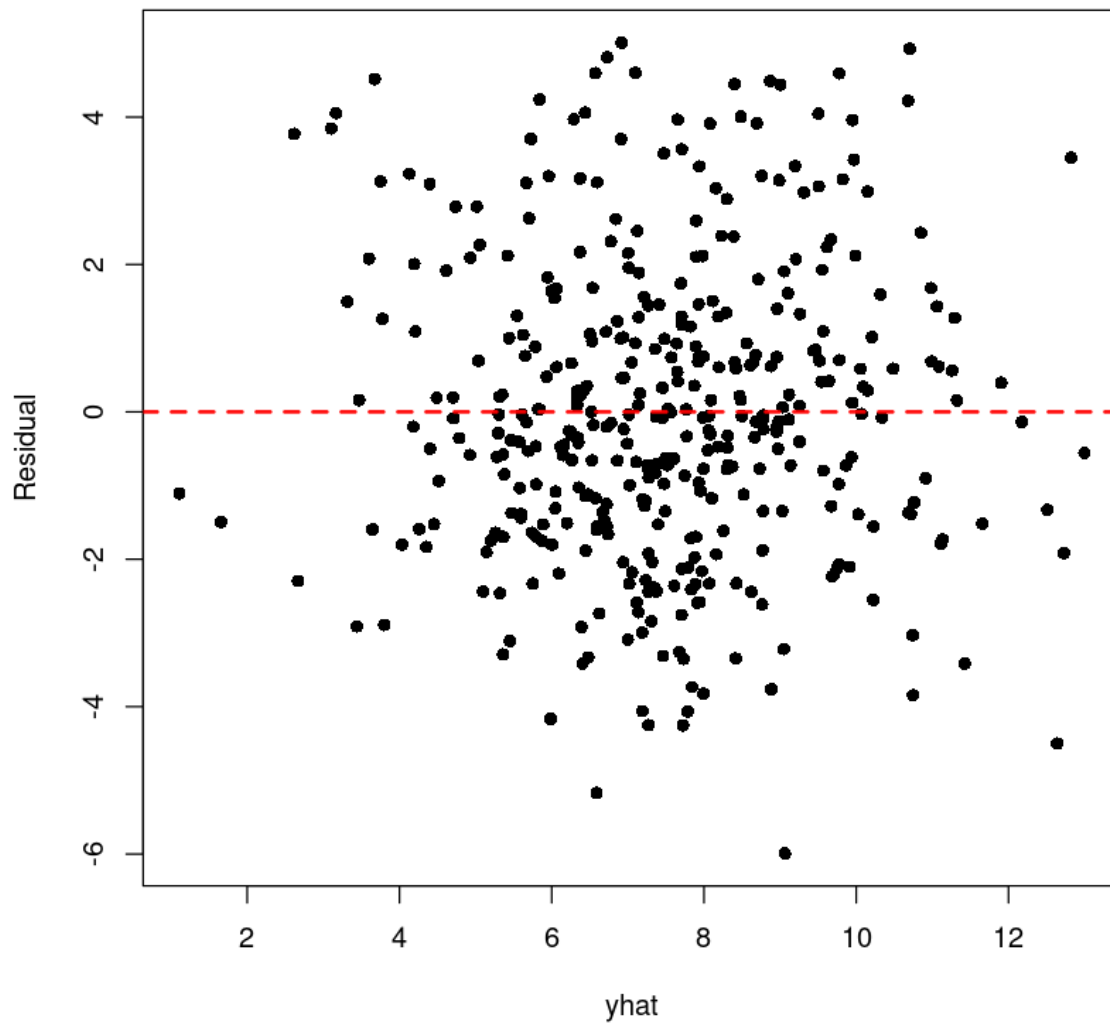
Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

```
yhat <- fit_Sales$fitted
resid <- fit_Sales$residuals
```

```
plot(resid ~ yhat, pch=16, ylab='Residual')
abline(h=0, lty=2, lwd=2, col='red')
```



- 0에 대해 대칭형태를 보이며, 등분산성을 만족하는 것으로 보인다.

(12) 잔차에 대한 등분산성 검정을 수행하여라.

```
### 등분산성
## H0 : 등분산 vs. H1 : 이분산 (Heteroscedasticity)
bptest(fit_Sales)
```

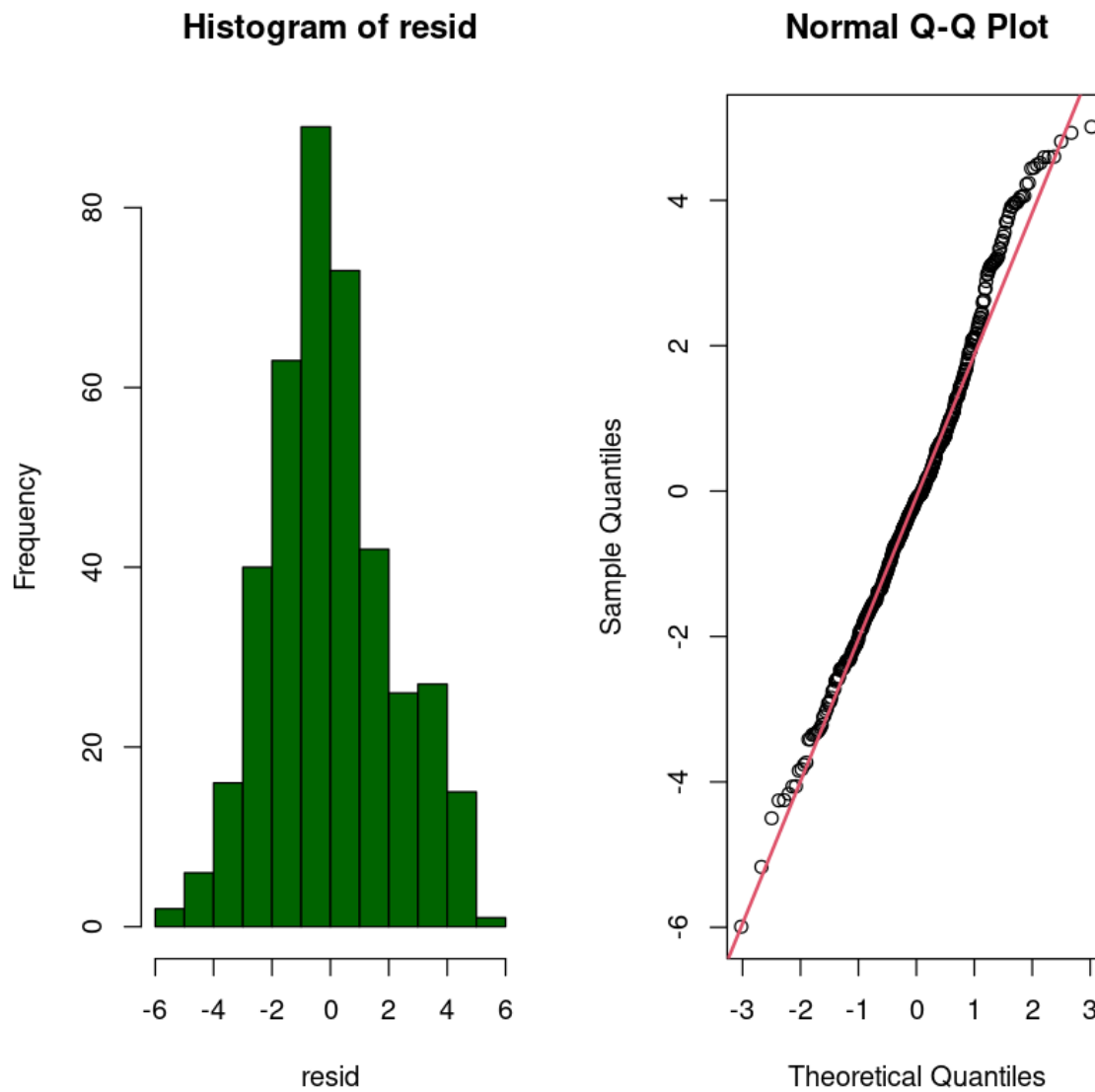
studentized Breusch-Pagan test

```
data: fit_Sales
BP = 2.6835, df = 6, p-value = 0.8474
```

유의수준 5% 하에서 Breusch-Pagan 검정 결과 p-value=0.8474로 유의수준 0.05보다 크므로 귀무가설을 기각할 수 없다. 따라서 잔차가 등분산성을 만족한다고 판단한다.

(13) 잔차에 대한 히스토그램, QQ plot을 그리고, 정규성 검정을 수행하여라.

```
par(mfrow=c(1,2))
hist(resid, col='darkgreen')
qqnorm(resid)
qqline(resid, col=2, lwd=2)
```



- Q-Q Plot의 우상단의 잔차들이 qqline에서 살짝 벗어나 있는 걸로보아 오른쪽으로 살짝 꼬리가 긴 형태를 띠는 것으로 보이지만 그림으로 판단하기에 애매한 것 같다.

```
## Shapiro-Wilk Test
## H0 : normal distribution vs. H1 : not H0
shapiro.test(resid)
```

#### Shapiro-Wilk normality test

```
data: resid
W = 0.98887, p-value = 0.003863
```

유의수준 5% 하에서 Shapiro wilk 정규성 검정 결과  $p - value = 0.003863$  으로 유의수준 0.05 보다 작으므로 정규성을 만족한다는 귀무가설을 기각할 수 있다. 따라서 정규성을 따르지 않는 것으로 판단한다.

#### (14) 잔차에 대한 독립성 검정을 수행하시오.

```
dwtest(fit_Sales, alternative = "two.sided") #H0 : uncorrelated vs H1 : rho != 0
```

#### Durbin-Watson test

```
data: fit_Sales
DW = 1.9677, p-value = 0.7479
alternative hypothesis: true autocorrelation is not 0
```

더빈왓슨 통계량이 1.9677로 2에 가까우며,  $p - value = 0.7479$ 로 유의수준 5% 하에서 잔차가 독립이라는 귀무가설을 기각할 수 없다. 따라서 잔차들간의 correlation이 없다(즉, 독립이다)고 판단한다.

#### Q2. 다음 물음에 답하여라.

위 데이터에 대하여 다음 물음에 답하여라. (R을 이용하여 풀이)(검정에서는 유의수준 = 0.05 사용)

#### (1) 위에서 적합한 모형에서 개별 회귀계수의 유의성 검정 결과 유의하지 않은 변수는 무엇인가?

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.5970091   1.1077522   4.150 4.08e-05 ***
CompPrice    0.0971849   0.0083574  11.629 < 2e-16 ***
Income       0.0132566   0.0037106   3.573 0.000397 ***
Advertising  0.1295888   0.0161449   8.027 1.17e-14 ***
Population   0.0001374   0.0007334   0.187 0.851487
Price        -0.0905544   0.0053869 -16.810 < 2e-16 ***
Education    -0.0407910   0.0396398  -1.029 0.304093
```

- 5% 하에서 개별 회귀계수의 유의성 결과 유의수준 Population과 Education이 유의하지 않았다.

(2) 위에서 유의하지 않았던 변수를 제외한 모형을 축소모형(Reduced Model)으로 하는 부분 F검정을 수행

하여라. 검정에 필요한 가설을 설정하고, 검정 결과를 설명하시오.

$$H_0 : \beta_4 = \beta_6 = 0 \quad (RM)$$

$$H_1 : \text{not } H_0$$

(3) 1번에서 설정한 모형과, 축소모형 중 어느 모형이 이 데이터에 대한 설명을 잘 하고 있는지를 비교하시오.

```
reduced_model = lm(Sales ~ .-Population - Education, data=df2) ## q=4, r=2
# reduced_model = lm(Sales ~ CompPrice + Income + Advertising + Price, data=df2) ## q=4, r=2
full_model = lm(Sales ~ ., data=df2) # 1번에서 설정한 모델. (=fit_Sales) ## p=6
anova(reduced_model, full_model)
```

A anova: 2 × 6

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	395	1671.898	NA	NA	NA	NA
2	393	1667.042	2	4.856036	0.5723976	0.5646408

```
p <- full_model$rank - 1
q <- reduced_model$rank-1
SSE_FM <- anova(full_model)$Sum[p+1] #SSE_FM
SSE_RM <- anova(reduced_model)$Sum[q+1] #SSE_RM

F0 <- ((SSE_RM-SSE_FM)/(p-q))/(SSE_FM/(nrow(df2)-p-1))
F0
```

0.572397599860576

```
## 기각역
qf(0.95,p-q,nrow(df2)-p-1)
```

3.01868441743385

$F_0 < F_{0.05}(2, 393)$  이므로 귀무가설을 기각할 수 없다. 따라서 축소모형이 이 데이터에 대한 설명을 잘 하고 있다고 할 수 있다.



```
summary(reduced_model)
```

Call:

```
lm(formula = Sales ~ . - Population - Education, data = df2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-5.8438	-1.4519	-0.1742	1.1659	5.1762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.077797	0.910848	4.477	9.92e-06 ***
CompPrice	0.096881	0.008309	11.660	< 2e-16 ***
Income	0.013439	0.003699	3.633	0.000317 ***
Advertising	0.130868	0.015559	8.411	7.50e-16 ***
Price	-0.090507	0.005378	-16.830	< 2e-16 ***

---

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

Residual standard error: 2.057 on 395 degrees of freedom

Multiple R-squared: 0.4746, Adjusted R-squared: 0.4693

F-statistic: 89.21 on 4 and 395 DF, p-value: < 2.2e-16

```
summary(full_model)
```

Call:

```
lm(formula = Sales ~ ., data = df2)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-5.9921	-1.4004	-0.1145	1.2378	5.0112

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.5970091	1.1077522	4.150	4.08e-05 ***
CompPrice	0.0971849	0.0083574	11.629	< 2e-16 ***
Income	0.0132566	0.0037106	3.573	0.000397 ***

```
Advertising 0.1295888 0.0161449 8.027 1.17e-14 ***
Population 0.0001374 0.0007334 0.187 0.851487
Price -0.0905544 0.0053869 -16.810 < 2e-16 ***
Education -0.0407910 0.0396398 -1.029 0.304093
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.06 on 393 degrees of freedom
Multiple R-squared: 0.4761, Adjusted R-squared: 0.4681
F-statistic: 59.54 on 6 and 393 DF, p-value: < 2.2e-16
```

- 축소모형의 RMSE가 좀 더 줄어들었으며,  $R^2_{Adj}$  값 역시 0.4693으로 full모델의 값인 0.4681 보다 좀 더 좋아졌다.

### Q3. 일반 선형 가설검정 (General Linear Hypothesis Test)을 수행하여라.

1번에서 설정한 모형에 대하여 아래의 일반 선형 가설검정(General Linear Hypothesis Test)을 수행하시오. (R을 이용하여 풀이)(검정에서는 유의수준 = 0.05 사용)(회귀계수는  $\beta_i$ 로 표현해야 하지만, 각자 설정이 다를 수가 있기 때문에 회귀계수 대신 변수 이름을 사용하겠음. 예  $\beta_1 = \text{CompPrice}$ )

```
library(car)
```

```
Loading required package: carData
```

```
Attaching package: 'car'
```

```
The following object is masked from 'package:dplyr':
```

```
recode
```

```
The following object is masked from 'package:purrr':
```

```
some
```

```
fit_Sales ## 1에서 설정한 모델
```

Call:

```
lm(formula = Sales ~ ., data = df2)
```

Coefficients:

```
(Intercept)    CompPrice      Income Advertising  Population      Price  
  4.5970091    0.0971849    0.0132566    0.1295888    0.0001374   -0.0905544  
  Education  
 -0.0407910
```

(1)  $H_0 : \text{CompPrice} = \text{Income}$  vs.  $H_1 : \text{not } H_0$

```
# CompPrice-Income = 0 => (0,1,-1,0,0,0,0)*beta  
linearHypothesis(fit_Sales, c(0,1,-1,0,0,0,0),0)
```

A anova:  $2 \times 6$

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	2040.234	NA	NA	NA	NA
2	393	1667.042	1	373.1914	87.97871	5.357559e-19

```
linearHypothesis(fit_Sales, c(0,-1,1,0,0,0,0),0)
```

A anova:  $2 \times 6$

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	2040.234	NA	NA	NA	NA
2	393	1667.042	1	373.1914	87.97871	5.357559e-19

(2)  $H_0 : \text{CompPrice} = -\text{Price}$  vs.  $H_1 : \text{not } H_0$

```
# H_0: Comprice + Price = 0
linearHypothesis(fit_Sales, c(0,1,0,0,0,-1,0),0)
```

A anova: 2 × 6

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	2653.572	NA	NA	NA	NA
2	393	1667.042	1	986.5296	232.5713	1.403571e-41

```
linearHypothesis(fit_Sales, c(0,-1,0,0,0,1,0),0)
```

A anova: 2 × 6

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	2653.572	NA	NA	NA	NA
2	393	1667.042	1	986.5296	232.5713	1.403571e-41

(3)  $H_0$ 를 기각할 수 있는 제약조건을 만들어 보시오.(단 2개 이상의 변수 사용)

```
# H_0: Income = Advertising
linearHypothesis(fit_Sales, c(0,0,1,-1,0,0,0),0)
```

A anova: 2 × 6

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	1870.519	NA	NA	NA	NA
2	393	1667.042	1	203.4765	47.96894	1.776524e-11

```
# H_0: Income = Advertising+2
linearHypothesis(fit_Sales, c(0,0,1,-1,0,0,0),2)
```

A anova:  $2 \times 6$

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	394	69008.501	NA	NA	NA	NA
2	393	1667.042	1	67341.46	15875.54	7.549323e- 320