Time Series Analysis - 478 - HW #3

Alex Towell (atowell@siue.edu)

Contents

Problem 1

Regression through the origin: We will consider a special case of simple linear regression where the intercept is assumed to be zero from the outset. Let

$$Yi = \beta x_i + \epsilon_i \,,$$

where $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$.

Part (A)

$$\boxed{\text{Define } Q(\beta) \coloneqq \sum_{i=1}^n (Y_i - \beta x_i)^2 \text{. the minimizer of } Q(\beta) \text{ is } \hat\beta = \frac{\sum x_i Y_i}{\sum x_i^2}.}$$

Proof. The function to minimize is given by

$$Q(\beta) = \sum_{i=1}^n (Y_i - \beta x_i)^2 \,, \tag{1}$$

which has a minimum when its derivative is zero,

$$\begin{aligned} \left.\frac{\mathrm{d}Q}{\mathrm{d}\beta}\right|_{\hat\beta} &= 0 \\ -2\sum (Y_i - \hat\beta x_i) x_i &= 0 \\ \sum Y_i x_i &= \hat\beta \sum x_i^2 \,. \end{aligned}$$

Solving for $\hat\beta$,

$$\hat\beta = \frac{\sum Y_i x_i}{\sum x_i^2} \,. \tag{2}$$

□

Part (B)

Show $\mathrm{E}(\hat{\beta}) = \beta$.

Proof. The expectation of $\hat{\beta}$ is given by

$$\begin{aligned} \mathrm{E}(\hat{\beta}) &= \mathrm{E}\left(\frac{\sum Y_i x_i}{\sum x_i^2}\right) \\ &= \frac{\sum \mathrm{E}(Y_i x_i)}{\sum x_i^2} \\ &= \frac{\sum x_i \, \mathrm{E}(\beta x_i + \epsilon_i)}{\sum x_i^2} \\ &= \frac{\beta \sum x_i^2}{\sum x_i^2}\,, \end{aligned}$$

which finally can be simplified to

$$\mathrm{E}(\hat{\beta}) = \beta\,. \tag{3}$$

□

Part (C)

Show $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2}$.

Proof. The variance of $\hat{\beta}$ is given by

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \left(\sum x_i^2\right)^{-2} \mathrm{Var}\left(\sum Y_i x_i\right) \\ &= \left(\sum x_i^2\right)^{-2} \sum x_i^2 \, \mathrm{Var}(Y_i) \\ &= \left(\sum x_i^2\right)^{-2} \sum x_i^2 \, \mathrm{Var}(\beta x_i + \epsilon_i) \\ &= \left(\sum x_i^2\right)^{-2} \sum x_i^2 \, \mathrm{Var}(\epsilon_i) \\ &= \sigma^2 \left(\sum x_i^2\right)^{-2} \sum x_i^2 \end{aligned}$$

which finally can be simplified to

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 \left(\sum x_i^2\right)^{-1}\,. \tag{4}$$

□

Part (D)

Write the model as $\mathbf{Y} = \mathbf{X}\beta + \epsilon$.

The equation $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ can be rewritten as

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}\,. \tag{5}$$

Proof. The original model is given by

$$Y_i \coloneqq \beta x_i + \epsilon_i \,.$$

When we perform the calculation $\mathbf{Y} = \mathbf{X}\beta$, the i-th element of $\mathbf{Y}$ is $Y_i = \beta x_i + \epsilon_i$. □

Part (E)

Verify $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is equivalent to the minimizer in Part (A).

Proof. To show that $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{\sum x_i Y_i}{\sum x_i^2}$, we apply the matrix operations to demonstrate equivalence to Part (A).

The transpose of $\mathbf{X}$ is $\mathbf{X}' = (x_1\, x_2\, \cdots\, x_n)$, a row vector. Making the appropriate substitutions, we get

$$\hat{\beta} = \left((x_1\, x_2\, \cdots\, x_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right)^{-1} (x_1\, x_2\, \cdots\, x_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \,. \tag{6}$$

We see that $X'X$ is just the sum of squares $x_1^2 + \cdots + x_n^2$ and $X'Y$ is just $x_1 Y_1 + \cdots + x_n Y_n$, resulting in

$$\hat{\beta} = (x_1^2 + x_2^2 + \cdots + x_n^2)^{-1}(x_1 Y_1 + x_2 Y_2 + \cdots + x_n Y_n) \,.$$

Since the inverse operation $^{-1} \colon \mathbb{R} \mapsto \mathbb{R}$ is just the reciprocal, we may rewrite the above as

$$\hat{\beta} = \frac{x_1 Y_1 + x_2 Y_2 + \cdots + x_n Y_n}{x_1^2 + x_2^2 + \cdots + x_n^2} \,.$$

which may finally rewritten as

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2} \,. \tag{7}$$

□

Part (F)

Show that $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ is equivalent to the scalar form in Part (C).

Proof. The variance of $\hat{\beta}$ is given by

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) \,. \tag{8}$$

Note that the variance of a random vector $\mathbf{T}$ multiplied on the left by a vector $\mathbf{U}$ is given by

$$\mathrm{Var}(\mathbf{U}\mathbf{T}) = \mathbf{U}\,\mathrm{Var}(\mathbf{T})\mathbf{U}' \,. \tag{9}$$

Thus,

$$\mathrm{Var}(\hat{\beta}) = \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \mathrm{Var}(\mathbf{Y}) \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' \,. \tag{10}$$

By the property of transposition, $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$, we may rewrite the above as

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-T} \,.$$

By the assumption that $\{\epsilon_i\}$ are i.i.d., the variance of $\mathbf{Y}$ is $\sigma^2\mathbf{I}_p$,

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-T} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-T} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-T}\,.\end{aligned}$$

Given a symmetric matrix $\mathbf{A}$, $\mathbf{A}' = \mathbf{A}$. Since $\mathbf{X}'\mathbf{X}$ is "symmetric" (a scalar value, being viewed as a matrix, is by definition symmetric), we can simplify the above to

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\,.$$

□

Observe that since $\hat{\beta} = c\mathbf{Y}$ where c is constant and $\mathbf{Y}$ is a vector of normals, $\hat{\beta}$ is normally distributed $\hat{\beta} \sim \mathcal{N}(\beta, \mathrm{Var}(\hat{\beta}))$.

Problem 2

Burple, Stephens, and Gloopshire (2014) report on a study in the Journal of Questionable Research. Data were collected on the number of minutes Y_i it took $n = 237$ Glippers to learn how to drive a small, motorized car. Two predictors of interest are the estimated age of the glipper in months x_{i1} and the Glippers Maladaptive Score (GMS) x_{i2}, a number from 50 to 100 that summarizes how poor the glipper's vision is. Consider the following multiple regression output from R.

Table 1: Regression output from R

Coefficients	Estimate	Std. Error
(Intercept)	-182.57923	7.41169
Age	8.56069	0.31150
GMS	0.28066	0.04621

Table 2: Anaylysis of Variance Table

type	df	ss	ms
regression			
error		23565	
total	236	104682	

Common code

The following R code is used to produce many of the answers to this problem set.

```
n <- 237
p <- 3
ss.e <- 23565
ss.t <- 104682

# ss.e + ss.r = ss.t => ss.r = ss.t - ss.e
ss.r <- ss.t - ss.e

df.r <- p-1
```

```
df.e <- n-p
df.t <- n-1

ms.r <- ss.r / df.r
ms.e <- ss.e / df.e

# sample variance
ms.t <- ss.t / df.t

B0.est <- -182.57923
B1.est <- 8.56069
B2.est <- 0.28066

B0.se <- 7.41169
B1.se <- 0.31150
B2.se <- 0.04621
```

Part (A)

Complete the ANOVA table.

There are $p = 3$ parameters in the model and it is given that there are $n = 237$ observations with a sum of squared error $\mathrm{SS_E} = 23565$ and sum of squared total $\mathrm{SS_T} = 104682$.

From this given information, we may compute the remaining values in the ANOVA table by observing the following set of relationships:

$$\begin{aligned} \mathrm{SS_R} &= \mathrm{SS_T} - \mathrm{SS_E} \\ \mathrm{MS_R} &= \mathrm{SS_R}/\mathrm{df_R} \\ \mathrm{MS_E} &= \mathrm{SS_E}/\mathrm{df_E} \\ \mathrm{MS_T} &= \mathrm{SS_T}/\mathrm{df_T} \end{aligned}$$

where $\mathrm{df_R} = \mathrm{p} - 1$, $\mathrm{df_E} = \mathrm{n} - \mathrm{p}$, and $\mathrm{df_T} = \mathrm{n} - 1$.

We output the ANOVA table with the following R code:

```
anova_table <- data.frame(
   type = c("regression","residual","total"),
   ss = c(ss.r,ss.e,ss.t),
   dof = c(df.r,df.e,df.t),
   ms = c(ms.r,ms.e,ms.t)
)
knitr::kable(anova_table, digits=2, caption = "ANOVA table.")
```

Table 3: ANOVA table.

type	ss	dof	ms
regression	81117	2	40558.50
residual	23565	234	100.71
total	104682	236	443.57

Part (B)

Calculate the F-statistic from the ANOVA table and use it to test

$$H_0\colon \beta_1 = \beta_2 = 0\,.$$

What does this imply about β_1 and β_2?

Consider the model

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon_i \tag{11}$$

where x_1 and x_2 are the independent predictor variables.

If we wish to do an overall test on the model, then we are interested in the hypothesis test

$$\begin{aligned} H_0 &\colon \beta_1 = \beta_2 = 0 \\ H_1 &\colon \beta_1 \neq 0 \vee \beta_2 \neq 0\,. \end{aligned}$$

Rejection of H_0 implies that one or more predictors is signficant in the model.

If H_0 is true and the model errors are normal and independent with constant variance, then the test statistic for significance of regression is

$$F_0 = \frac{\mathrm{SS_R}/(\mathrm{p}-1)}{\mathrm{SS_E}/(\mathrm{n}-\mathrm{p})} = \frac{\mathrm{MS_R}}{\mathrm{MS_E}}\,, \tag{12}$$

where $F_0 \sim F(p-1, n-p)$.

We compute the statistic using the following R code:

```
F.ratio=ms.r/ms.e
p_value=1-pf(F.ratio,p,n-p) #p-value from the F test
cat("test statistic F0 =",F.ratio, "with p-value", p_value, "\n")
```

```
## test statistic F0 = 402.7451 with p-value 0
```

We see that $\Pr(F_0 > 3) \approx 0.05$. We have a much larger value and we see that the p-value is essentially 0. Thus, we have very strong evidence to reject the null hypothesis, i.e., either β_1 or β_2 (or both) are not zero and we conclude that one or more of the predictors in the model is significant.

Part (C)

Report each of $\hat{\beta}_1$ and $\hat{\beta}_2$. Construct t-tests $H_0\colon \beta_1 = 0$ and $H_0\colon \beta_2 = 0$ individually. Can either predictor be dropped in the presence of the other?

Observe $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 \mathbf{C})$ and therefore $\hat{\beta}_j \sim (\beta_j, \sigma^2 C_{jj})$. Thus, $\frac{\hat{\beta}_j - \beta_j}{\sqrt{(\sigma^2 C_{jj})}} \sim \mathcal{N}(0,1)$. Since we do not know σ^2, we estimate it with $\hat{\sigma}^2 = \mathrm{MS_E}$, in which case

$$\frac{\hat{\beta}_j - \beta_j}{\mathrm{SE}(\hat{\beta}_j)} \sim t(\mathrm{df} = \mathrm{n} - \mathrm{p})\,,$$

where $\mathrm{SE}(\hat{\beta}_j) \coloneqq \sqrt{\hat{\sigma}^2 C_{jj}}$ is denoted the standard error and $\mathbf{C} \coloneqq (\mathbf{X}'\mathbf{X})^{-1}$.

We consider the hypothesis test for testing the significance of β_j,

$$\begin{aligned} H_0 &: \beta_j = 0\,, \\ H_1 &: \beta_j \neq 0\,, \end{aligned}$$

with test statistic

$$t_0 = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)} \sim t(\text{df} = \text{n} - \text{p})\,.$$

We reject H_0 if t_0 has a sufficiently small p-value, where

$$p\text{-value} = 2\,\Pr(t < -|t_0|)\,.$$

We specify that we reject H_0 if p-value $\leq \alpha = 0.05$ (or equivalently $|t_0| > t_{\alpha/2,n-p}$).

```
B1.t0 <- B1.est / B1.se
B2.t0 <- B2.est / B2.se
cat("B1.t0 =",B1.t0,"\n")
```

```
## B1.t0 = 27.48215
```

```
cat("B2.t0 =",B2.t0,"\n")
```

```
## B2.t0 = 6.073577
```

```
#p-value from the T test
B1.pval=2*pt(-1*abs(B1.t0),df=n-p)
B2.pval=2*pt(-1*abs(B2.t0),df=n-p)
cat("B1 pvalue =",B1.pval,",B2 pvalue =", B2.pval)
```

```
## B1 pvalue = 3.321975e-75 ,B2 pvalue = 5.019091e-09
```

It is given that $\hat{\beta}_1 = 8.56$ and $\hat{\beta}_2 = 0.28$ respectively with standard errors 0.31150 and 0.04621 and p-values of $3.321975e^{-75}$ and $5.019091e^{-09}$. These p-values are practically *zero* and thus in the presence in the presence of the other, neither predictor should be dropped.

Part (D)

Interpret both coefficients.

The estimated coefficients in the model are given by $\hat{\beta}_1 = 8.56$ and $\hat{\beta}_2 = 0.28$, thus

$$\hat{Y}_i = \hat{\beta}_0 + 8.56x_{i1} + 0.28x_{i2} + \epsilon_i \tag{13}$$

which has an estimated expectation

$$E(\hat{Y}_i) = \hat{\beta}_0 + 8.56x_{i1} + 0.28x_{i2}\,, \tag{14}$$

where x_{i1} is the estimated age of the glipper in months, x_{i2} is the Glippers Maladaptive Score (GMS), a number between 50 and 100 that summarizes how poor the glipper's vision is, and $\hat{Y}_i$ is the number of minutes it takes a glipper to learn how to drive a small, motorized car with those predictors.

According to the model, the rate of increase with respect to age and GMA is expected to increase the number of minutes to learn by 8.56 and 0.28 minutes respectively.

Part (E)

Report R^2; how is it interpretted here?

Recall that $R^2 = SS_R/SS_T = 1 - SS_E/SS_T$. Thus, it is the percentage of the variability explained by the regression model versus the total variability. All things else being equal, the more of the variability we can explain by the regression model the better. Since R^2 is a number between 0 and 1, the closer to 1 the better.

We run the following R code:

```r
#R2 <- 1 - ss.e / ss.t
R2 <- ss.r / ss.t
cat("R2 =",R2,"\n")
```

```
## R2 = 0.7748897
```

We see that $R^2 \approx 0.77$, and thus the model explains around 77% of the variability in the data.

Note that a large value of R^2 does not necessarily imply the model is good, since we can always capture more variability in a model by adding more parameters to it, which contributes to *over-fitting* the particulars of the observe data.

Problem 3

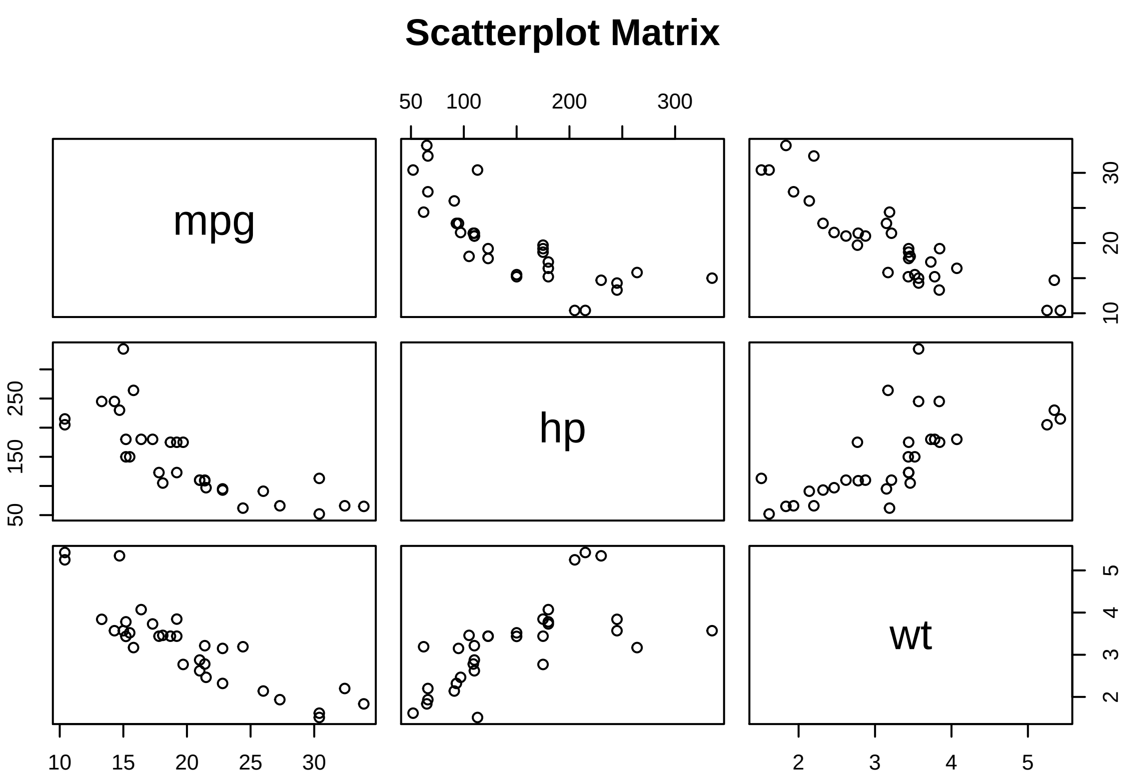
Consider the "mtcars" data in R. You can load and view the dataset by typing "mtcars" in R console. Consider the response variable mileage per hour ($Y = \text{mpg}$), and two predictors, horsepower and weight ($X_1 = \text{hp}, \text{X}_2 = \text{wt}$). Ignore other variables for now.

Part (A)

Obtain and report the scatterplot matrix; what does it tell you about the relationship between mpg and each of the predictors, horsepower and weight?

We print out the scatterplot matrix with the following R code.

```r
pairs(mpg~hp+wt,mtcars, main="Scatterplot Matrix")        #construct scatterplots
```



The scatter plot visually indicates that the miles per gallon (mpg) is negatively correlated with both the horsepower (hp) and the weight (wt).

Part (B)

Fit the regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. Report the ANOVA table and the table of regression coefficients.

We use the following R code to fit the model and then report the ANOVA table.

```
#### Regression Analysis Using lm() ####

# fit a multiple regression model
mpg_hp_wt.model=lm(mpg~hp+wt, data=mtcars)

# get details from the regression output
summary(mpg_hp_wt.model)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
## hp          -0.03177    0.00903  -3.519  0.00145 **
## wt          -3.87783    0.63273  -6.129 1.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

```
# get the anova table
anova(mpg_hp_wt.model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
hp	1	678.3729	678.372874	100.86152	0.0e+00
wt	1	252.6266	252.626559	37.56091	1.1e-06
Residuals	29	195.0478	6.725785	NA	NA

Part (C)

Comment on the significance of the overall model.

The p-value $9.109e^{-12}$ for the overall test is extremely small (very large F_0 statistic), thus we have very strong evidence that either β_1 or β_2 (or both) are not zero. In other words, we conclude that one or more of the predictors in the model are significant.

Part (D)

Comment of the significance of each predictor. Can either predictor be dropped in the presence of the other?

The hypothesis for testing the significance of β_j,

$$H_0 : \beta_j = 0\,,$$

is given by the individual t-test

$$t_0 = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)}\,, \tag{15}$$

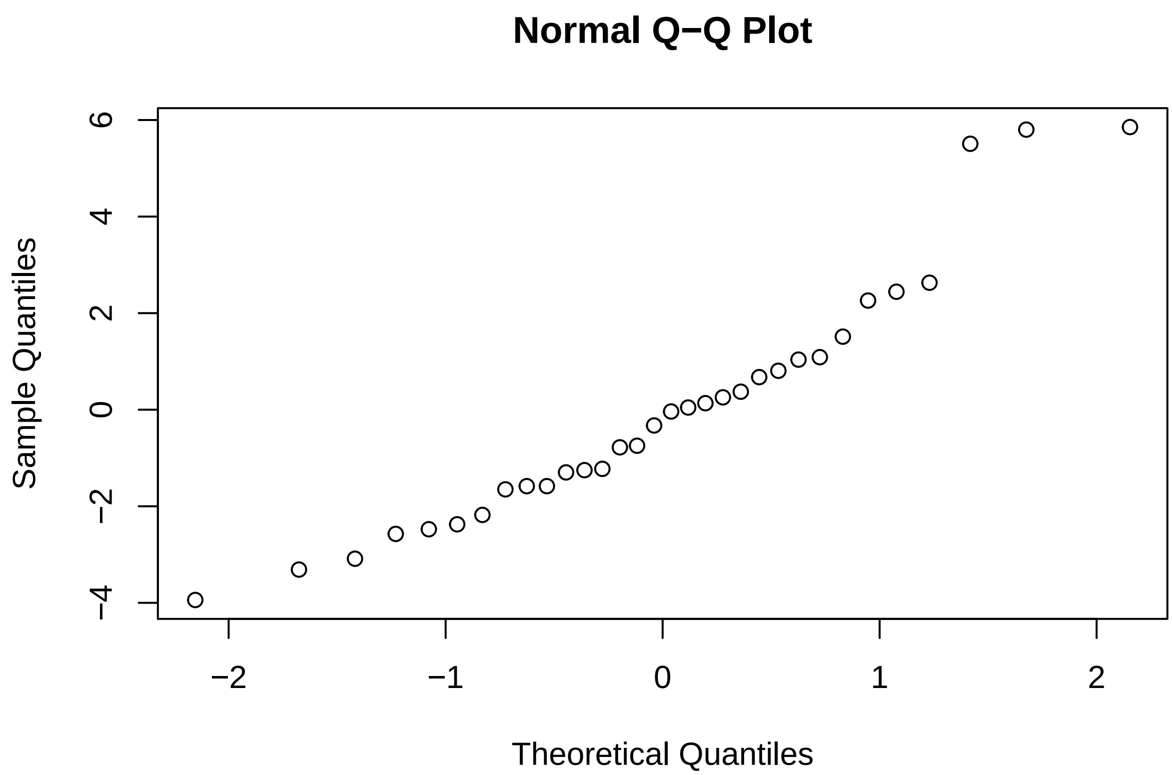
which are reported in the regression summary in Part (B). Both have p-values that are approximately *zero* so in the presence of the other neither should be dropped.

Part (E)

Obtain the normal probability plot and a histogram of the residuals. What do these plots tell you?

A QQ-normal probability plot is shown next.

```
# construct qq plot of the residuals
qqnorm(mpg_hp_wt.model$residual)
```



The QQ-plot of the residuals is not quite a 45 degree line, suggesting that the assumption of normality on the error term may not be reasonable under some circumstances.

Next, we show the historgram of the residuals

```
# construct a histogram plot of the residuals
hist(mpg_hp_wt.model$residual)
```

This is not terribly symmetric, confirming our earlier suspicions.

Part (F)

Obtain SSR(x_1), SSR($x_2|x_1$), and verify SSR(x_1, x_2) = SSR(x_1) + SSR($x_2|x_1$).

The following R code generates the sum of squared regressions for the models.

```
#### Testing on group of coefficients
result.full=lm(mpg~hp+wt, data=mtcars)
result.red_x1=lm(mpg~hp, data=mtcars)
result.red_x2=lm(mpg~wt, data=mtcars)
summary(result.full)    #get details from the regression output
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
## hp          -0.03177    0.00903  -3.519  0.00145 **
## wt          -3.87783    0.63273  -6.129 1.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
```

```
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

```
anova(result.full)       #ANOVA table for the full model
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
hp	1	678.3729	678.372874	100.86152	0.0e+00
wt	1	252.6266	252.626559	37.56091	1.1e-06
Residuals	29	195.0478	6.725785	NA	NA

```
anova(result.red_x1)         #ANOVA table for the reduced model
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
hp	1	678.3729	678.37287	45.4598	2e-07
Residuals	30	447.6743	14.92248	NA	NA

```
anova(result.red_x2)         #ANOVA table for the reduced model
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
wt	1	847.7252	847.725250	91.37533	0
Residuals	30	278.3219	9.277398	NA	NA

```
anova(result.red_x1,result.full,test = "F") #F-test for the effect of act and year togother
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	447.6743	NA	NA	NA	NA
29	195.0478	1	252.6266	37.56091	1.1e-06

The sum of squares for the regression for the full model is $\text{SSR}(x_1, x_2) = 930.996$. The sum of squares for the regression when we remove x_1 is $\text{SSR}(x_2) = 847.7252$. The sum of squares for the regression when we remove x_2 is $\text{SSR}(x_1) = 678.3729$. The extra sum of squares when we add x_2 given x_1 is already in the model is $\text{SSR}(x_2|x_1) = \text{SSR}(x_1, x_2) - \text{SSR}(x_1) = 930.996 - 678.3729 = 252.6266$. Thus, plugging them in, $930.996 = 678.3729 + 252.6266$.

Part (G)

Obtain and interpret an 95% interval estimate of $\text{E}(Y_h)$ when $x_{h1} = 100$ and $x_{h2} = 4$.

Look in the next section, Alternative approach to Problem 3, where I compute the intervals. There's a lot of other stuff there, also, but the intervals are near the end of the R code block.

I get the result $[16.66102, 20.41629]$.

Alternative approach to Problem 3

We use the matrix approach. Here is the R code.

```r
#### Regression analysis using matrix methods ####
p3.Y <- mtcars[c("mpg")]
rownames(p3.Y) <- NULL
colnames(p3.Y) <- NULL
p3.Y <- data.matrix(p3.Y)
p3.n <- nrow(p3.Y)
p3.X <- mtcars[c("hp","wt")]
rownames(p3.X) <- NULL
colnames(p3.X) <- NULL
p3.X <- cbind(rep(1,p3.n),data.matrix(p3.X))

p3.XtX=t(p3.X)%*%p3.X       #get X'X
p3.C=solve(p3.XtX)

p3.coeffs=p3.C%*%t(p3.X)%*%p3.Y #Least Squares result

print(p3.coeffs)
```

```
##             [,1]
## [1,] 37.22727012
## [2,] -0.03177295
## [3,] -3.87783074
```

```r
# number of parameters p
p3.p <- 3

# degrees of freedom for SSR,SSE,and SST respectively
p3.df.r <- p3.p-1
p3.df.e <- p3.n-p3.p
p3.df.t <- p3.n-1

# SSE = y'y - B'x'y
p3.ss.e=t(p3.Y)%*%p3.Y-t(p3.coeffs)%*%t(p3.X)%*%p3.Y       # calculate SSE
p3.ms.e=p3.ss.e/p3.df.e # calculate MSE

# SSR = B'X'y - n avg(y)^2
p3.ss.r=(t(p3.coeffs)%*%t(p3.X)%*%p3.Y-p3.n*mean(p3.Y)^2)     # calculate SSR
p3.ms.r=p3.ss.r/p3.df.r                                      # calculate MSR

# SST = sum(y[i] - avg(y))
#     = y'y - n*avg(y)
p3.ss.t=t(p3.Y)%*%p3.Y-p3.n*mean(p3.Y)^2                       # calculate SSTO
p3.ms.t=p3.ss.t/p3.df.t

# estimate of sigma^2 is mean squared error (MSE)
p3.sigma2_hat = p3.ms.e
print(p3.sigma2_hat)
```

```
##          [,1]
## [1,] 6.725785
```

```r
# covariance matrix for coefficient estimators
p3.cov=p3.sigma2_hat[1]*p3.C
print(p3.C)
```

```
##                 [,1]          [,2]          [,3]
## [1,]  3.800481e-01  2.207476e-05 -0.1094214557
## [2,]  2.207476e-05  1.212285e-05 -0.0005595912
## [3,] -1.094215e-01 -5.595912e-04  0.0595249024
```

```
p3.r.sq=(t(p3.coeffs)%*%t(p3.X)%*%p3.Y-p3.n*mean(p3.Y)^2)/p3.ss.t  # calculate R-square
print(p3.r.sq)
```

```
##           [,1]
## [1,] 0.8267855
```

```
p3.F.ratio=p3.ms.r/p3.ms.e      #F test for overall model significance
p3.pval=1-pf(p3.F.ratio,2,p3.n-3)   #p-value from the F test

p3.anova <- data.frame(
   type = c("regression","residual","total"),
   ss = c(p3.ss.r,p3.ss.e,p3.ss.t),
   dof=c(p3.df.r,p3.df.e,p3.df.t),
   ms = c(p3.ms.r,p3.ms.e,p3.ms.t)
)
knitr::kable(p3.anova, digits=2, caption = "ANOVA table.")
```

Table 9: ANOVA table.

type	ss	dof	ms
regression	931.00	2	465.50
residual	195.05	29	6.73
total	1126.05	31	36.32

```
p3.B0.est = p3.coeffs[1]
p3.B0.se = sqrt(p3.ms.e*p3.C[1,1])

p3.B1.est = p3.coeffs[2]
p3.B1.se = sqrt(p3.ms.e*p3.C[2,2])

p3.B2.est = p3.coeffs[3]
p3.B2.se = sqrt(p3.ms.e*p3.C[3,3])

p3.B0.t0 <- p3.B0.est / p3.B0.se
p3.B1.t0 <- p3.B1.est / p3.B1.se
p3.B2.t0 <- p3.B2.est / p3.B2.se


cat("overall F0 =",p3.F.ratio,"pval =",p3.pval)
```

```
## overall F0 = 69.21121 pval = 9.109047e-12
```

```
cat("B0 =",p3.B0.est,"se(B0) =",p3.B0.se,"t0 =",p3.B0.t0)
```

```
## B0 = 37.22727 se(B0) = 1.598788 t0 = 23.28469
```

```
cat("B1 =",p3.B1.est,"se(B1) =",p3.B1.se,"t0 =",p3.B1.t0)
```

```
## B1 = -0.03177295 se(B1) = 0.00902971 t0 = -3.518712
```

```r
cat("B2 =",p3.B2.est,"se(B2) =",p3.B2.se,"t0 =",p3.B2.t0)
```

```
## B2 = -3.877831 se(B2) = 0.6327335 t0 = -6.128695
```

```r
##### code for confidence intervals
x0 <- c(1,100,4)
yhat=t(p3.coeffs)%*%x0
print(yhat)
```

```
##          [,1]
## [1,] 18.53865
```

```r
se.yhat=sqrt(p3.ms.e*t(x0)%*%p3.C%*%x0)      #std.error for the mean response
CI.l=yhat-qt(0.975,n-3)*se.yhat
CI.u=yhat+qt(0.975,n-3)*se.yhat      #the lower/upper limit for a 95% CI
print(CI.l)
```

```
##          [,1]
## [1,] 16.66102
```

```r
print(CI.u)
```

```
##          [,1]
## [1,] 20.41629
```

We collect the results and summarize the results in the following table:

Table 10: Coefficient statistics

	$\hat{\beta}_j$	SE($\hat{\beta}_j$)	t_0	F_0
β_0	37.22727	1.598788	23.28469	69.21121
β_1	-0.03177295	0.00902971	-3.518712	
β_2	-3.877831	0.6327335	-6.128695	