LiMo WiSe 16/17 Sheet 5: Ex 1

Task:

We have obtained the following matrices as a result of a regression analysis (with intercept β_0).

$$X'X = \begin{pmatrix} 9 & 136 & 269 & 260 \\ 136 & 2114 & 4176 & 3583 \\ 269 & 4176 & 8257 & 7104 \\ 260 & 3583 & 7104 & 12276 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 45 \\ 648 \\ 1283 \\ 1821 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 9.610932 & 0.0085878 & -0.2791475 & -0.0445217 \\ 0.0085878 & 0.5099641 & -0.2588636 & 0.0007765 \\ -0.2791475 & -0.2588636 & 0.1395 & 0.0007369 \\ -0.0445217 & 0.0007765 & 0.0007396 & 0.0003698 \end{pmatrix}$$

$$(X'X)^{-1}X'y = \begin{pmatrix} -1.163461 \\ 0.135270 \\ 0.019950 \\ 0.121954 \end{pmatrix}$$

$$y'y = 285$$

a. Calculate the test statistic for the test of overall regression and illustrate your intermediate steps in an ANOVA table. Interpret the result.

b. Calculate $\hat{\beta}$ and the diagonal elements of $\widehat{Cov}(\hat{\beta})$ and test the hypothesis whether each regression coefficient equals 0.

c. Define the matrix C for testing the hypothesis $H_0 : \beta_0 = 0, \beta_1 = \beta_3, \beta_2 = 0$.

d. Find the model equation for the reduced model in c).

Answers

a)

ANOVA-Table:

Source	DF	Sum of squares	Mean squares	F
Model	3	57.97	19.32	47.7
Error	5	203	0.405	
Total	8	60		

Obtaining degrees of freedom (DF)

- $DF_{\text{model}} = p$ and we can get $p+1$ from the dimension of $X'X \Rightarrow 4 \Rightarrow p = 3$.
- $DF_{\text{total}} = n - 1$; $n = (X'X)_{11} = 9$.
- $DF_{\text{error}} = n - p - 1 = 5$.

Obtaining Sum of Squares

Total Sum of Squares

$$
\begin{aligned}
SST &= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \\
&= y'y - n\left(\frac{(X'y)_{11}}{n}\right)^2 \\
&= 285 - 9\left(\frac{45}{9}\right)^2 \\
&= 60
\end{aligned}
$$

Residual Sum of Squares

$$
\begin{aligned}
SSR &= \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2 \\
&= \hat{y}'\hat{y} - n\bar{y}^2 \\
&\Rightarrow \text{Using } \hat{y} = X(X'X)^{-1}X'y \\
&= (X(X'X)^{-1}X'y)'X(X'X)^{-1}X'y - n\bar{y}^2 \\
&= y'X(X'X)^{-1}X'X(X'X)^{-1}X'y - n\bar{y}^2 \\
&= y'X(X'X)^{-1}X'y - n\bar{y}^2 \\
&= (X'y)'X(X'X)^{-1}X'y - n\bar{y}^2 \\
&= \begin{pmatrix} 45 & 648 & 1283 & 1821 \end{pmatrix} \begin{pmatrix} -1.16 \\ 0.4 \\ 0.02 \\ 0.12 \end{pmatrix} - n\bar{y}^2 \\
&= 57.97
\end{aligned}
$$

Error Sum of Squares

$$
SSE = SST - SSR = 60 - 57.93 = 2.03
$$

Mean Squres

- Residual mean square $MSR = \frac{SSR}{p} = 19.32$
- Error mean square $MSE = \frac{SSE}{n-p-1} = 0.405$

F-value

$$F = \frac{MSR}{MSE} = 47.7 \overset{H_0}{\sim} F(3,5)$$

To obtain the (two-sided) p-value:

```
(1 - pt(47, 3, 5)) * 2
```

```
## [1] 0.003679194
```

We reject $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ because $p < 0.05$.

b)

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} -1.163461 \\ 0.135270 \\ 0.019950 \\ 0.121954 \end{pmatrix}$$

$$\widehat{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{\epsilon_i}^2}{n-p-1} = MSE = 0.405$$

We need the diagonal elements of $diag(\widehat{Cov}(\hat{\beta})) = (3.8924, 0.2065, 0.0565, 0.00015)$.

Testing wether each regression coefficient equals 0:

$$T = \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_k} \sim t(n-p-1, 0)$$

In our case: $H_0 : \beta_k = 0$ hence

$$T = \frac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}}$$

with

$$\hat{\sigma}_{\beta_k} = \sqrt{\widehat{Cov}(\hat{\beta})_{(k+1)(k+1)}}$$

This results in

	$\hat{\beta}_k$	$\hat{\sigma}_k$	t	p
β_0	-1.16	1.97	-0.59	0.709
β_1	0.14	0.45	0.30	0.389
β_2	0.02	0.24	0.08	0.468
β_3	0.12	0.01	9.97	0.0009

c) $H_0 : \beta_0 = 0, \beta_1 = \beta_3, \beta_2 = 0; \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$

In matrix notation:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

d)

reduced model

$$y_i = (\underbrace{x_{i1} + x_{i3}}_{\tilde{x_i}} \boldsymbol{\beta} + \boldsymbol{\varepsilon_i})$$

$$\tilde{x_i} = \begin{pmatrix} x_{11} + x_{13} \\ \dots \\ x_{11} + x_{13} \end{pmatrix} = \mathbf{x}_1 + \mathbf{x}_3$$

$$\hat{\beta} = (\tilde{x_1}'\tilde{x})^{-1}\tilde{x_1}y$$

$$x = \begin{pmatrix} 1 & x_{11} & \dots & x_{13} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{n3} \end{pmatrix}$$

$$x'x = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & \sum x_{i3} \\ \sum x_{i1} & \sum \mathbf{x_{i1}^2} & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} \\ \sum x_{i2} & \dots & \sum x_{i2}^2 & \sum \mathbf{x_{i2}x_{i3}} \\ \dots & \dots & \dots & \sum \mathbf{x_{i3}^2} \end{pmatrix}$$

$$\begin{aligned} \tilde{x}'\tilde{x} &= (x_1 + x_3)'(x_1 + x_3) \\ &= x_1x_1 + 2x_1x_3 + x_3x_3 \\ &= 2114 + 2 * 3583 + 12276 = \underline{21556} \end{aligned}$$

$$\begin{aligned} \tilde{x}'y &= (x_1 + x_3)'y \\ &= x_1y + x_3y \\ &= 648 + 1821 = \underline{2469} \end{aligned}$$

$$\hat{\beta} = \frac{2469}{21556} = \underline{\underline{0.11454}}$$

Reduced model: $\hat{y_i} = 0.114 * (x_{i1} + x_{i3})$