Chapter 3

Multiple Linear Regression

Chapter 3f

$$H_0 : \beta_1 = \beta_2 = \dots \beta_x$$

$$y'y = SS_{Res} + \hat{\beta}'x'y$$

$$y'y - n\bar{y}^2 = SS_{Res} + \hat{\beta}'x'y - SS_R(\beta_0)$$

$$SS_T = SS_{Res} + \underbrace{SS_R(\beta_1, \dots \beta_x \mid \beta_0)}_{SS_R}$$

Testing general linear hypothesis about β

1. Suppose we have a full model

$$\boldsymbol{y_{n\times 1}} = \boldsymbol{X_{n\times p}\beta_{p\times 1}} + \boldsymbol{\epsilon_{n\times 1}}$$

where $\boldsymbol{\beta'} = (\beta_0, \beta_1, ..., \beta_k)$, $p = k + 1$, $\boldsymbol{E(\epsilon)} = \boldsymbol{0}$ and $\boldsymbol{Var(\epsilon)} = \sigma^2 \boldsymbol{I_n}$.

Assume $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

We want to test $H_0 : \boldsymbol{T\beta} = \boldsymbol{0}$ where $\boldsymbol{T}$ is an $r \times p$ matrix such that all the r equations in $\boldsymbol{T\beta} = \boldsymbol{0}$ are independent.

For example, suppose we consider the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon,$$

and we want to test $H_0 : \underbrace{\beta_1 = \beta_3}_{\text{affect response in similar manner}}, \beta_4 = 0, \beta_5 = 0$, then

$$\boldsymbol{T} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}.$$

$$T\beta = \begin{pmatrix} \beta_1 - \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

Hence, $H_0 : \boldsymbol{T\beta} = \boldsymbol{0}$ is exactly the same as $H_0 : \beta_1 = \beta_3, \beta_4 = 0, \beta_5 = 0$

2. By applying $\boldsymbol{T\beta} = \mathbf{0}$ to the full model, we will obtain a reduced model

$$\boldsymbol{y}_{n\times 1} = \boldsymbol{Z}_{n\times(p-r)}\boldsymbol{\Gamma}_{(p-r)\times 1} + \boldsymbol{\epsilon}_{n\times 1}.$$

where $\boldsymbol{\Gamma} = (\gamma_0, \gamma_1, ..., \gamma_{p-r-1})$.

For the example, by applying $\beta_1 = \beta_3$, $\beta_4 = 0$, $\beta_5 = 0$, we get the reduced model

originally $\beta_3 x_3$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1 x_3 + 0x_4 + 0x_5 + \epsilon$$

$$y = \beta_0 + \beta_1(x_1 + x_3) + \beta_2 x_2 + \epsilon$$ reduced model

and the model can be written as

$$y = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \epsilon \quad \text{where } z_1 = x_1 + x_3 \text{ and } z_2 = x_2.$$

$= Z\Gamma + \epsilon$

3. To test $H_0 : \boldsymbol{T\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{T\beta} \neq \mathbf{0}$, we consider the following models:

   Full model (FM): $\quad \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$

   Reduced model (RM): $\quad \boldsymbol{y} = \boldsymbol{Z\Gamma} + \boldsymbol{\epsilon}$

   For the example,

   Full model (FM): $\quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$

   Reduced model (RM): $\quad y = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \epsilon \quad$ where $z_1 = x_1 + x_3$ and $z_2 = x_2$

4. The regression sum of squares and its degrees of freedom for testing $H_0 : \boldsymbol{T\beta} = \mathbf{0}$ are given as

   $SS_H = SS_{Res}(RM) - SS_{Res}(FM)$

   Degrees of freedom for $SS_H = r$

5. The F statistic for testing $H_0 : \boldsymbol{T\beta} = \mathbf{0}$ is given as $F = \dfrac{SS_H/r}{SS_{Res}(FM)/(n-p)}$.

   Reject H_0 if $F > F_{\alpha,r,n-p}$

6. An example based on the pr2103 data is displayed here. The p-value is 0.9822 which is large, so we do not reject $H_0 : \beta_1 = \beta_3, \beta_4 = 0, \beta_5 = 0$.

RM FM

```
> z1 <- x1+x3
> z2 <- x2
>
> anova(lm(y~z1+z2),lm(y~x1+x2+x3+x4+x5))
Analysis of Variance Table

Model 1: y ~ z1 + z2
Model 2: y ~ x1 + x2 + x3 + x4 + x5
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     80 8915.2
2     77 8895.7  3    19.585 0.0565 0.9822
```

RM

FM

subtract

get this

p value for testing hypothesis

do not reject H_0 if p value is large

7. The extra sum of squares method is a special case of testing the general linear hypothesis about $\boldsymbol{\beta}$. For example, if we test $H_0 : \beta_4 = \beta_5 = 0$, then

$$\boldsymbol{T} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underbrace{\beta_4 x_4 + \beta_5 x_5}_{\text{test if } 0} + \varepsilon$$

and $SS_H = SS_R(\beta_4, \beta_5 | \beta_0, \beta_1, \beta_2, \beta_3)$.

Example - pr2103 data

Consider the multiple linear regression model:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon$

$y =$ systolic blood pressure in mmHg

$x_1 =$ diastolic blood pressure in mmHg

$x_2 =$ number of heart beats per minute

$x_3 =$ weight in kg

$x_4 =$ height in m

$x_5 =$ age in years

$x_6 =$ exam score

(i) Test $H_0 : \beta_4 = \beta_6 = 0$.

(ii) Test $H_0 : \beta_1 = \beta_2 = \beta_3, \beta_4 = \beta_6 = 0$.

(iii) Test $H_0 : \beta_1 = (\beta_2 + \beta_3)/2, \beta_4 = \beta_6 = 0$.

(i) Test $H_0 : \beta_4 = \beta_6 = 0$.

Full model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon$$

Reduced model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_5 x_5 + \epsilon$$

$\downarrow$

set $\beta_4 = \beta_6 = 0$

```r
# to test beta4=beta6=0
anova(lm(y~x1+x2+x3+x5),lm(y~x1+x2+x3+x4+x5+x6))
```

29:1 (Top Level)

Console | Background Jobs

R 3.4.1 · ~/

```
> # to test beta4=beta6=0
> anova(lm(y~x1+x2+x3+x5),lm(y~x1+x2+x3+x4+x5+x6))
Analysis of Variance Table

Model 1: y ~ x1 + x2 + x3 + x5
Model 2: y ~ x1 + x2 + x3 + x4 + x5 + x6
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     78 8898.5
2     76 8723.6  2     174.9 0.7619 0.4703
```

Do not reject Ho

$r=2$

$\Leftrightarrow$ 2 equations (2 β to test)

(ii) Test $H_0 : \beta_1 = \beta_2 = \beta_3, \beta_4 = \beta_6 = 0.$

Full model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon$$

Reduced model:

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \beta_1 x_3 + \beta_5 x_5 + \epsilon$$

$$y = \beta_0 + \beta_1 (x_1 + x_2 + x_3) + \beta_5 x_5 + \epsilon$$

use constraint

```r
# to test beta1=beta2=beta3, beta4=beta6=0
x123 <- x1+x2+x3
anova(lm(y~x123+x5),lm(y~x1+x2+x3+x4+x5+x6))
```

33:1 (Top Level)

Console | Background Jobs

R 3.4.1 · ~/

```
> x123 <- x1+x2+x3
> anova(lm(y~x123+x5),lm(y~x1+x2+x3+x4+x5+x6))
Analysis of Variance Table

Model 1: y ~ x123 + x5
Model 2: y ~ x1 + x2 + x3 + x4 + x5 + x6
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     80 9381.7
2     76 8723.6  4    658.13 1.4334 0.2311
```

Do not reject Ho

big then don't reject

4 constrains

(iii) Test $H_0 : \beta_1 = (\beta_2 + \beta_3)/2, \beta_4 = \beta_6 = 0.$

Full model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon$$

Reduced model:

$$y = \beta_0 + \frac{\beta_2 + \beta_3}{2} x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_5 x_5 + \epsilon$$

$$y = \beta_0 + \left(\frac{x_1}{2} + 1\right)\beta_2 + \left(\frac{x_1}{2} + 1\right)\beta_3 + \beta_5 x_5 + \epsilon$$

$$y = \beta_0 + \beta_2 z_1 + \beta_3 z_2 + \beta_5 x_5 + \epsilon$$

$$y = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 x_5 + \epsilon$$

where $z_1 = \frac{x_1}{2} + 1$ and $z_2 = \frac{x_1}{2} + 1$

```r
# to test beta1=(beta2+beta3)/2, beta4=beta6=0
z1 <- x1/2+1
z2 <- x2/2+1
anova(lm(y~z1+z2+x5),lm(y~x1+x2+x3+x4+x5+x6))
```

39:46 (Top Level)

Console **Background Jobs**

R 3.4.1 · ~/

RM FM

```
> # to test beta1=(beta2+beta3)/2, beta4=beta6=0
> z1 <- x1/2+1
> z2 <- x2/2+1
> anova(lm(y~z1+z2+x5),lm(y~x1+x2+x3+x4+x5+x6))
Analysis of Variance Table

Model 1: y ~ z1 + z2 + x5
Model 2: y ~ x1 + x2 + x3 + x4 + x5 + x6
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     79 10883.8
2     76  8723.6  3    2160.2 6.2733 0.0007331 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reject Ho

coz very small

Tests and confidence intervals on individual regression coefficients

$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$

1. Suppose we have the multiple linear regression model

$$\boldsymbol{y}_{\boldsymbol{n}\times\boldsymbol{1}} = \boldsymbol{X}_{\boldsymbol{n}\times\boldsymbol{p}}\boldsymbol{\beta}_{\boldsymbol{p}\times\boldsymbol{1}} + \boldsymbol{\epsilon}_{\boldsymbol{n}\times\boldsymbol{1}}$$

$e_i = y_i - \hat{y}_i$ random variable

$\varepsilon_i \sim N(0, \sigma^2)$ random variable

where $\boldsymbol{\beta'} = (\beta_0, \beta_1, ..., \beta_k)$, $p = k+1$, $\boldsymbol{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\boldsymbol{Var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I_n}$.

Assume $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

$e_i = \hat{\varepsilon}_i$

$$\boldsymbol{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$\hat{\beta} = (X'X)^{-1}X'y$

$$\boldsymbol{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X'X})^{-1} = \sigma^2\boldsymbol{C}$$

$$Var\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ .\\ .\\ \hat{\beta}_k\end{bmatrix} = \sigma^2\begin{bmatrix}C_{00} & C_{01} & C_{02} & ... & C_{0k}\\ C_{10} & C_{11} & C_{12} & ... & C_{1k}\\ . & & & & \\ . & & & & \\ C_{k0} & C_{k1} & C_{k2} & ... & C_{kk}\end{bmatrix}.$$

$\hat{\sigma}^2 = MS_{Res}$

2. Linear model theory:

$T = \dfrac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$ follows the t distribution with $n - p$ degrees of freedom.

$Var(\hat{\beta}_j) = \hat{\sigma}^2 C_{jj}$

3. For testing $H_0 : \beta_j = c$ versus $H_1 : \beta_j \neq c$, we reject H_0 if

$$\frac{\hat{\beta}_j - c}{\sqrt{\hat{\sigma}^2 C_{jj}}} < -t_{\alpha/2,n-p} \text{ or } \frac{\hat{\beta}_j - c}{\sqrt{\hat{\sigma}^2 C_{jj}}} > t_{\alpha/2,n-p}$$

4. A $100(1-\alpha)\%$ confidence interval for β_j is given as

$$\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}}$$

5. For example, to perform tests and construct confidence intervals on individual regression coefficients, the following R codes can be used for the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$,

```
summary(lm(y~x1+x2+x3))
confint(lm(y~x1+x2+x3), level=0.95)
```

6. An example based on the pr2103 data is displayed in here.

```
> summary(lm(y~x1+x2+x3))

Call:
lm(formula = y ~ x1 + x2 + x3)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.1586  -7.3682  -0.5432   5.8787  29.8728 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  12.6946    11.7860   1.077  0.28472    
x1            0.6383     0.1193   5.352 8.30e-07 ***
x2            0.3684     0.1092   3.375  0.00115 ** 
x3            0.6186     0.1362   4.542 1.97e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.62 on 79 degrees of freedom
Multiple R-squared:  0.5686,    Adjusted R-squared:  0.5522 
F-statistic: 34.71 on 3 and 79 DF,  p-value: 2.056e-14
```

$H_0: \beta_0 = 0$

```
> confint(lm(y~x1+x2+x3), level=0.95)
                 2.5 %     97.5 %
(Intercept) -10.7648547 36.1540141
x1            0.4008855  0.8756886
x2            0.1511581  0.5856838
x3            0.3475071  0.8896978
```

Estimation of mean response

1. Suppose we have the multiple linear regression model

   $$\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

   where $\boldsymbol{\beta}' = (\beta_0, \beta_1, ..., \beta_k)$, $p = k+1$, $\boldsymbol{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\boldsymbol{Var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_{\boldsymbol{n}}$.
   Assume $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$.

2. The model can also be written as

   $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i,\ i = 1, 2, ..., n.$

   At the point $\boldsymbol{x}_{\mathbf{0}}' = [1\ x_{01}\ x_{02}\ ...\ x_{0k}]$, the response y_0 is

   $y_0 = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + ... + \beta_k x_{0k} + \epsilon_0.$

$E(y|x_0)$ mean of response given this set of regressor values

3. The mean response at $\boldsymbol{x_0}$ is

scalar quantity

$$E(y|\boldsymbol{x_0}) = E(y_0) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + ... + \beta_k x_{0k} + \underbrace{E(\epsilon_0)}_{0}$$

$$E(y|\boldsymbol{x_0}) = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + ... + \beta_k x_{0k} = \boldsymbol{x_0'}\boldsymbol{\beta}$$

4. The fitted value at $\boldsymbol{x_0'}$ is

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + ... + \hat{\beta}_k x_{0k} = \boldsymbol{x_0'}\hat{\boldsymbol{\beta}}$$

function of y

$$= \begin{bmatrix} 1 & x_{01} & x_{02} & ... & x_{0k} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ \hat{\beta}_k \end{bmatrix}.$$

5. The fitted value $\hat{y}_0$ is a function of y_i's and hence it is normally distributed.

6. The mean $E(y|\boldsymbol{x_0})$ can be estimated by $\hat{y}_0$.

unbiased estimator for mean

7. The mean of $\hat{y}_0$ is

   $E(\hat{y}_0) = E(\boldsymbol{x_0'}\hat{\boldsymbol{\beta}}) = \boldsymbol{x_0'}E(\hat{\boldsymbol{\beta}}) = \boldsymbol{x_0'}\boldsymbol{\beta} = E(y|\boldsymbol{x_0})$

8. The variance of $\hat{y}_0$ is

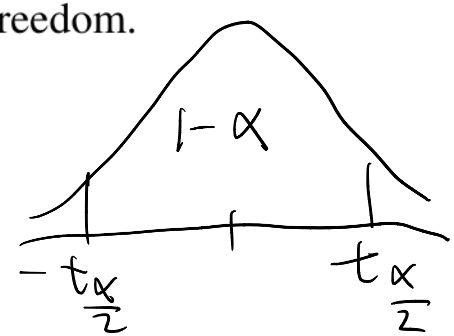
   $Var(\hat{y}_0) = Var(\boldsymbol{x_0'}\hat{\boldsymbol{\beta}}) = \boldsymbol{x_0'}\, Var(\hat{\boldsymbol{\beta}})\, \boldsymbol{x_0} = \sigma^2\boldsymbol{x_0'}\, (\boldsymbol{X'X})^{-1}\, \boldsymbol{x_0}$

9. The variance of $\hat{y}_0$ can be estimated as

   $\widehat{Var(\hat{y}_0)} = \hat{\sigma}^2\boldsymbol{x_0'}\, (\boldsymbol{X'X})^{-1}\, \boldsymbol{x_0}$

10. Linear model theory:

$\dfrac{\hat{y}_0 - E(y|\boldsymbol{x_0})}{\sqrt{\widehat{Var}(\hat{y}_0)}}$ follows the t distribution with $n-p$ degrees of freedom.



$$P\left(-t_{\alpha/2,n-p} \le \frac{\hat{y}_0 - E(y|\boldsymbol{x_0})}{\sqrt{\widehat{Var}(\hat{y}_0)}} \le t_{\alpha/2,n-p}\right) = 1-\alpha$$

$$P\left(\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\widehat{Var}(\hat{y}_0)} \le E(y|\boldsymbol{x_0}) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\widehat{Var}(\hat{y}_0)}\right) = 1-\alpha$$

parameter

$$P\left(\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x_0'}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{x_0}} \le E(y|\boldsymbol{x_0}) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x_0'}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{x_0}}\right) = 1-\alpha$$

A $100(1-\alpha)\%$ confidence interval for $E(y|\boldsymbol{x_0})$ is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x_0'}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{x_0}} \le E(y|\boldsymbol{x_0}) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x_0'}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{x_0}}$$

11. For the delivery time data, a 95% confidence interval for the mean response for $x_1 = 8$ cases and $x_2 = 275$ feet is displayed in here. The R codes

```
newdata <- data.frame(x1=8, x2=275)
predict.lm(lm(y~x1+x2),newdata,interval="confidence",
           level=0.95)
```

can be used to construct a 95% confidence interval for the mean response.

Example - Delivery time data

95% CI on the mean delivery time for an outlet requiring $x_1 = 8$ cases

and where the distance $x_2 = 275$ feet

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix} \qquad \hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = [1 \quad 8 \quad 275]\begin{bmatrix} 2.34123 \\ 1.61591 \\ 0.01438 \end{bmatrix} = 19.22 \text{ minutes}$$

The variance of $\hat{y}_0$ is estimated by

MS Res

$$\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 10.6239 [1 \quad 8 \quad 275]$$

$$\times \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$

$$= 10.6239(0.05346) = 0.56794$$

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \le E(y|x_0) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

$$19.22 - 2.074\sqrt{0.56794} \le E(y|x_0) \le 19.22 + 2.074\sqrt{0.56794}$$

$$17.66 \le E(y|x_0) \le 20.78$$

```
> #95% confidence interval of mean of y at x0
> newdata <- data.frame(x1=8, x2=275)
> predict.lm(lm(y~x1+x2), newdata, interval="confidence",level=0.95)
       fit     lwr      upr
1 19.22432 17.6539 20.79474
```

Prediction of new observation

1. Suppose we have the multiple linear regression model

$$\boldsymbol{y_{n\times 1}} = \boldsymbol{X_{n\times p}\beta_{p\times 1}} + \boldsymbol{\epsilon_{n\times 1}}$$

where $\boldsymbol{\beta'} = (\beta_0, \beta_1, ..., \beta_k)$, $p = k+1$, $\boldsymbol{E(\epsilon) = 0}$ and $\boldsymbol{Var(\epsilon)} = \sigma^2\boldsymbol{I_n}$. Assume $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

2. The model can also be written as

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i,\ i = 1, 2, ..., n.$

3. Suppose we want to predict a future response y_0 at the point $\boldsymbol{x_0'} = \begin{bmatrix} 1 & x_{01} & x_{02} & ... & x_{0k} \end{bmatrix}$,

$$y_0 = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + ... + \beta_k x_{0k} + \epsilon_0.$$

4. The predicted response at $\boldsymbol{x_0'}$ is

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \ldots + \hat{\beta}_k x_{0k} = \boldsymbol{x_0'}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & x_{01} & x_{02} & \ldots & x_{0k} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ \hat{\beta}_k \end{bmatrix}.$$

5. The mean of $y_0 - \hat{y}_0$ is zero because $E(y_0) = E(\hat{y}_0) = \boldsymbol{x_0'}\boldsymbol{\beta}$.

6. The variance of $y_0 - \hat{y}_0$ is

$$\begin{aligned} Var(y_0 - \hat{y}_0) &= Var(y_0) + Var(\hat{y}_0) - 2\overset{=0}{Cov(y_0, \hat{y}_0)} \\ &= Var(y_0) + Var(\hat{y}_0) \\ &\qquad \because \text{ the future response } y_0 \text{ is independent of the data used to predict } y_0 \\ &= \sigma^2 + \sigma^2 \boldsymbol{x_0'} \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{x_0} \\ &= \sigma^2 [1 + \boldsymbol{x_0'} \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{x_0}] \end{aligned}$$

The variance of $y_0 - \hat{y}_0$ can be estimated as

$$\widehat{Var(y_0 - \hat{y}_0)} = \hat{\sigma}^2 [1 + \boldsymbol{x_0'} \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{x_0}] \, .$$

7. Linear model theory:

$$\frac{y_0 - \hat{y}_0}{\sqrt{\widehat{Var(y_0 - \hat{y}_0)}}}$$ follows the t distribution with $n-p$ degrees of freedom.

$$P\left(-t_{\alpha/2,n-p} \leq \frac{y_0 - \hat{y}_0}{\sqrt{\widehat{Var(y_0 - \hat{y}_0)}}} \leq t_{\alpha/2,n-p}\right) = 1 - \alpha$$

$$P\left(\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\widehat{Var(y_0 - \hat{y}_0)}} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\widehat{Var(y_0 - \hat{y}_0)}}\right) = 1 - \alpha$$

$$P\left(\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1 + \boldsymbol{x_0'}\ (\boldsymbol{X'X})^{-1}\ \boldsymbol{x_0}]} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1 + \boldsymbol{x_0'}\ (\boldsymbol{X'X})^{-1}\ \boldsymbol{x_0}]}\right)$$
$$= 1 - \alpha$$

A $100(1-\alpha)\%$ **prediction** interval for y_0 is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1 + \boldsymbol{x_0'}\ (\boldsymbol{X'X})^{-1}\ \boldsymbol{x_0}]} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1 + \boldsymbol{x_0'}\ (\boldsymbol{X'X})^{-1}\ \boldsymbol{x_0}]}$$

8. A confidence interval is used for estimating a constant, for example $E(y|\boldsymbol{x_0})$. Here, we are predicting a future observation y_0 which is not a constant but a random variable. The interval used for prediction is known as a prediction interval.

9. For the delivery time data, a 95% prediction interval of a future response for $x_1 = 8$ cases and $x_2 = 275$ feet is displayed here. The R codes

```
newdata <- data.frame(x1=8, x2=275)
predict.lm(lm(y~x1+x2),newdata,interval="prediction",
                  level=0.95)
```

can be used to construct the 95% prediction interval for a future response.

Example - Delivery time data

95% prediction interval of the delivery time for an outlet requiring $x_1 = 8$ cases

and where the distance $x_2 = 275$ feet

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix} \qquad \hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = [1 \quad 8 \quad 275]\begin{bmatrix} 2.34123 \\ 1.61591 \\ 0.01438 \end{bmatrix} = 19.22 \text{ minutes}$$

$$\begin{aligned}\hat{\sigma}^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 &= 10.6239[1 \quad 8 \quad 275] \\ &\times \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix} \\ &= 10.6239(0.05346) = 0.56794\end{aligned}$$

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2[1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]}$$

$$19.22 - 2.074\sqrt{10.6239(1+0.05346)} \le y_0 \le 19.22 + 2.074\sqrt{10.6239(1+0.05346)}$$

$$12.28 \le y_0 \le 26.16$$

```
> #95% prediction interval of mean of y at x0
> newdata <- data.frame(x1=8, x2=275)
> predict.lm(lm(y~x1+x2), newdata, interval="prediction",level=0.95)
       fit      lwr      upr
1 19.22432 12.28456 26.16407
```

The End