Stats 100C – Linear Models

University of California, Los Angeles

Duc Vu

Fall 2021

This is stats 100C – Linear Models taught by Professor Christou. There is not an official textbook used for the course. Instead, handouts and reference materials are distributed and can be accessed through the class website. You can find other math/stats lecture notes through my personal blog. Let me know through my email if you notice something mathematically wrong/concerning. Thank you!

Contents

§1 Lec 1: Sep 27, 2021

§1.1 Simple Linear Regression Models

Consider

$$Y_i = \mu + \varepsilon_i$$

with $\varepsilon_i \overset{\text{i.i.d}}{\sim} N(0,\sigma)$; specifically, $Y_1,\ldots,Y_n \overset{\text{i.i.d}}{\sim} N(\mu,\sigma)$. We want to estimate μ and σ^2 using least squares or method of maximum likelihood (MML).

Method of Least Squares (OLS – Ordinary Least Squares):

$$\begin{aligned}
\min Q &= \sum_{i=1}^n (Y_i-\mu)^2 \\
\frac{\partial Q}{\partial \mu} &= -2\sum (Y_i-\mu) = 0 \\
\sum Y_i - n\hat{\mu} &= 0 \\
\implies \hat{\mu} &= \overline{Y}
\end{aligned}$$

Method of Maximum Likelihood (MML):

$$\begin{aligned}
f(y_i) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
&= \left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
L &= f(y_1)\ldots f(y_n) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2} \\
\ln L &= -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum (y_i-\mu)^2 \\
\frac{\partial \ln L}{\partial \mu} &= 0, \qquad \frac{\partial \ln L}{\partial \sigma^2} = 0
\end{aligned}$$

Solve the above, we obtain the MLE of μ and σ^2

$$\hat{\mu} = \hat{y}, \qquad \hat{\sigma}^2 = \frac{\sum (y_i-\hat{\mu})^2}{n} = \frac{\sum (y_i-\overline{y})^2}{n}$$

Notice that $\hat{\sigma}^2$ is biased and we adjust it to be unbiased as follows

$$S^2 = \frac{\sum (y_i-\overline{y})^2}{n-1}$$

§1.2 Prediction Problem

Given $Y_1,\ldots,Y_n$, we want to predict a new Y, e.g., Y_0. An educated guess here is

$$\hat{Y}_0 = \overline{Y}$$

1. Predictor assumption: $\hat{Y}_0 = \sum_{i=1}^n a_i Y_i$
2. We want $\hat{Y}_0$ to be unbiased, i.e., $E\hat{Y}_0 = \mu$

$$\begin{aligned}
E\sum a_i Y_i &= \mu \\
\sum a_i E Y_i &= \mu \\
\implies \sum a_i &= 1
\end{aligned}$$

3. Minimize the mean square error of prediction, i.e.,

$$E\left(Y_0 - \hat{Y}_0\right)^2 \qquad \text{s.t.} \qquad \sum a_i = 1$$

Notice that this is a constraint optimization problem, we use the method of Lagrange multiplier to obtain

$$\min Q = E\left(Y_0 - \hat{Y}_0\right)^2 - 2\lambda\left(\sum a_i - 1\right)$$

<u>*Note*</u>: $EW^2 = \text{var}(W) + (EW)^2$

$$\begin{aligned}
\min Q &= \text{var}\left(Y_0 - \hat{Y}_0\right) - 2\lambda\left[\sum a_i - 1\right] \\
&= \text{var}(Y_0) + \text{var}(\hat{Y}_0) - 2\,\text{cov}\left(Y_0, \hat{Y}_0\right) - 2\lambda\left[\sum a_i - 1\right] \\
&= \sigma^2 + \sigma^2 \sum a_i^2 - 2\lambda\left[\sum a_i - 1\right] \\
\frac{\partial Q}{\partial a_i} &= 2\sigma^2 a_i - 2\lambda = 0 \\
a_i &= \frac{\lambda}{\sigma^2}
\end{aligned}$$

Notice that $a_1 = a_2 = \ldots = a_n = \frac{\lambda}{\sigma^2}$. So

$$\sum a_i = \frac{n\lambda}{\sigma^2} = 1 \implies \lambda = \frac{\sigma^2}{n}$$

Thus, we can see that

$$a_i = \frac{1}{n}$$

and therefore since $\hat{Y}_0 = \sum a_i Y_i$, it follows that $\hat{Y}_0 = \overline{Y}$.

<u>Prediction Interval</u>:

$$Y_0 - \hat{Y}_0 \sim N\left(0, \sigma\sqrt{1 + \frac{1}{n}}\right)$$

Recall from 100B

$$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$$

So,

$$\frac{\frac{Y_0 - \hat{Y}_0 - 0}{\sigma\sqrt{1+\frac{1}{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{Y_0 - \hat{Y}_0}{S\sqrt{1+\frac{1}{n}}} \sim t_{n-1}$$

We can now construct the prediction interval for Y_0 as follows

$$P\left(-t_{\frac{\alpha}{2};n-1} \le \frac{Y_0 - \hat{Y}_0}{S\sqrt{1+\frac{1}{n}}} \le t_{\frac{\alpha}{2};n-1}\right) = 1 - \alpha$$

Finally, $Y_0 \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-1} S\sqrt{1 + \frac{1}{n}}$.

Remark 1.1. Compare this to the confidence interval for μ : $\mu \in \overline{Y} \pm t_{\frac{\alpha}{2};n-1}\frac{S}{\sqrt{n}}$.

§2 Lec 2: Sep 29, 2021

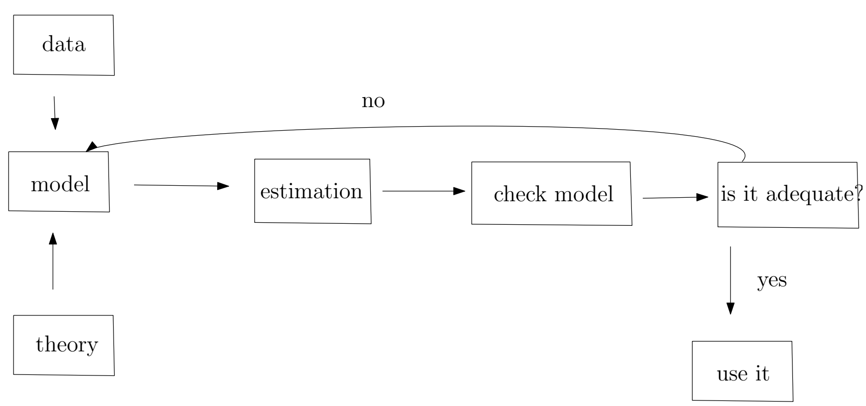
§2.1 Linear Regression

Consider a simple regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$
$$\text{or } Y_i = \beta_1 X_i + \varepsilon_i$$

Data:

y	x
y_1	x_1
$\vdots$	$\vdots$
y_n	x_n



where the parameters are

$$\begin{cases} \beta_0: & \text{intercept} \\ \beta_1: & \text{slope} \end{cases}$$

and $X_1, \ldots, X_n$ are predictors that are not random; $\varepsilon_1, \ldots, \varepsilon_n$ are random error terms/disturbance/stochastic terms, and $Y_1, \ldots, Y_n$ are random response variable.
Assumption (Gauss-Markov Conditions):

$$E(\varepsilon_i) = 0, \quad \text{var}(\varepsilon_i) = \sigma^2$$

$\varepsilon_1, \ldots, \varepsilon_n$ are independent. Using the Gauss-Markov conditions,

$$\begin{aligned} EY_i &= \beta_0 + \beta_1 X_i \\ \text{var}(Y_i) &= \sigma^2 \\ \min Q &= \sum \varepsilon_i^2 \\ \min Q &= \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \\ \frac{\partial Q}{\partial \beta_0} &= -2 \sum (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum (Y_i - \beta_0 - \beta_1 X_i) X_i = 0 \end{aligned}$$

So,

$$\begin{cases} \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \\ \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0 \end{cases}$$

$$\implies \begin{cases} n\beta_0 + \beta_1 \sum x_i = \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 = \sum x_i y_i \end{cases} \quad \text{– normal equations}$$

We can solve the above to get $\hat{\beta}_0, \hat{\beta}_1$.

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Determinant of the matrix:

$$\begin{aligned} n \sum x_i^2 - \left(\sum x_i\right)^2 &= n \left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right] \\ &= n \sum (x_i - \overline{x})^2 \geq 0 \end{aligned}$$

If $x_1 = x_2 = \ldots = x_n = \overline{x}$ then $\sum (x_i - \overline{x})^2 = 0$. From normal equations we get

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \quad \text{from (1)}$$

and plug (1) into (2) to obtain

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

or

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

or

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} \tag{*}$$

or

$$\hat{\beta}_1 = \frac{\sum (y_i - \overline{y}) x_i}{\sum (x_i - \overline{x})^2}$$

or

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

<u>*Note*</u>: From (*), we have

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} \\ &= \frac{(x_1 - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} + \ldots + \frac{(x_n - \overline{x}) y_n}{\sum (x_i - \overline{x})^2} \\ &= k_1 y_1 + \ldots + k_n y_n = \sum_{i=1}^{n} k_i y_i \end{aligned}$$

where $k_i = \frac{x_i - \overline{x}}{\sum(x_i-\overline{x})^2}$. Notice that

$$\begin{aligned}\sum k_i &= 0\\ \sum k_i^2 &= \frac{1}{\sum(x_i-\overline{x})^2}\\ \sum k_i x_i &= \frac{\sum(x_i-\overline{x})x_i}{\sum(x_i-\overline{x})^2} = 1\end{aligned}$$

Properties of $\hat{\beta}_1$:

$$\begin{aligned}E\hat{\beta}_1 &= E\sum k_i y_i = \sum k_i E y_i\\ &= \sum k_i\left(\beta_0+\beta_1 x_i\right)\\ &= \beta_0\sum k_i + \beta_1 \sum k_i x_i\\ &= \beta_1 - \text{unbiased}\end{aligned}$$

For the variance,

$$\begin{aligned}\operatorname{var}(\hat{\beta}_1) &= \operatorname{var}\left(\sum k_i y_i\right)\\ &= \sum k_i^2 \operatorname{var}(Y_i)\\ &= \frac{\sigma^2}{\sum(x_i-\overline{x})^2}\end{aligned}$$

Properties of $\hat{\beta}_0$:

$$\begin{aligned}\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1\overline{x}\\ &= \sum \frac{y_i}{n} - \overline{x}\sum k_i y_i\\ &= \sum\left(\frac{1}{n} - \overline{x}k_i\right) y_i\\ &= \sum_{i=1}^n l_i y_i\end{aligned}$$

where $l_i = \frac{1}{n} - \overline{x}k_i$ and the properties of l_i are

$$\begin{aligned}\sum l_i &= 1\\ \sum l_i^2 &= \sum\left(\frac{1}{n} - \overline{x}k_i\right)^2 = \sum\left(\frac{1}{n^2} + \overline{x}^2 k_i^2 - \frac{2}{n}\overline{x}k_i\right)\\ &= \frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i-\overline{x})^2}\\ \sum l_i x_i &= 0\end{aligned}$$

Now, we can easily show that $\hat{\beta}_0$ is unbiased

$$\begin{aligned}E\hat{\beta}_0 &= E\sum l_i y_i = \sum l_i E y_i\\ &= \sum l_i\left(\beta_0+\beta_1 x_i\right) = \beta_0 \sum l_i + \beta_1 \sum l_i x_i\\ &= \beta_0\end{aligned}$$

Thus,

$$\operatorname{var}\left(\hat{\beta}_0\right) = \operatorname{var}\left(\sum l_i y_i\right) = \sigma^2 \sum l_i^2 = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum (x_i - \overline{x})^2}\right)$$

The fitted value is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \overline{y} + \hat{\beta}_1 \left(x_i - \overline{x}\right)$$

and the residual is defined as

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

with properties

$$\begin{aligned}\sum e_i &= 0\\ \sum e_i x_i &= 0\\ \sum e_i \hat{y}_i &= 0\end{aligned}$$

<u>Estimation Using MML:</u>

Assume $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. Then $Y_i \sim N\left(\beta_0 + \beta_1 X_i, \sigma\right)$. The log-likelihood function is

$$\ln L = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

So, we need to solve

$$\frac{\partial \ln L}{\partial \beta_0} = 0, \quad \frac{\partial \ln L}{\partial \beta_1} = 0$$

to get $\hat{\beta}_0, \hat{\beta}_1$ which are the same as least squares method.

$$\begin{aligned}\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - \beta_0 - \beta_1 x_i)^2 = 0\\ \hat{\sigma}^2 &= \frac{\sum e_i^2}{n}\end{aligned}$$

Then,

$$\sum (y_i - \overline{y})^2 = \sum \left(\underbrace{y_i - \hat{y}_i}_{e_i} + \hat{y}_i + \overline{y}\right)^2$$

in which we expand to get

$$\underbrace{\sum (y_i - \overline{y})^2}_{\text{SST}} = \underbrace{\sum e_i^2}_{\text{SSE}} + \underbrace{\sum (\hat{y}_i - \overline{y})^2}_{\text{SSR}}$$

in which

$$\begin{cases}\text{SST: sum of squares total}\\ \text{SSE: sum of squares error}\\ \text{SSR: sum of squares regression}\end{cases}$$

§3 Lec 3: Oct 1, 2021

§3.1 Gauss-Markov Theorem

Recall

$$\hat{\beta}_1 = \sum k_i Y_i$$

where $k_i = \frac{x_i - \overline{x}}{\sum (x_i - \overline{x})^2}$. Consider now

$$b_1 = \sum a_i Y_i$$

which is another unbiased estimator of β_1. Then $Eb_1 = \beta_1$ or $E \sum a_i Y_i = \beta_1$. So

$$\begin{aligned} \beta_1 &= \sum a_i E Y_i \\ &= \sum a_i \left(\beta_0 + \beta_1 X_i\right) \\ &= \beta_0 \sum a_i + \beta_1 \sum a_i X_i \end{aligned}$$

Thus,

$$\begin{cases} \sum a_i = 0 \\ \sum a_i x_i = 1 \end{cases}$$

and we know that

$$\text{var}(b_1) = \text{var}\left(\sum_{i=1}^n a_i Y_i\right) = \sigma^2 \sum a_i^2$$

and

$$\text{var}(\hat{\beta}_1) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum (x_i - \overline{x})^2}$$

Now let $a_i = k_i + d_i$. Then,

$$\begin{aligned} \text{var}(b_1) &= \sigma^2 \sum (k_i + d_i)^2 \\ &= \sigma^2 \sum k_i^2 + \sigma^2 \sum d_i^2 + 2\sigma^2 \sum k_i d_i \end{aligned}$$

We need to show $\sum k_i d_i = 0$.

$$\begin{aligned} \sum k_i (a_i - k_i) &= \sum k_i a_i - \sum k_i^2 \\ &= \frac{\sum (x_i - \overline{x}) a_i}{\sum (x_i - \overline{x})^2} - \frac{1}{\sum (x_i - \overline{x})^2} \\ &= \frac{\sum x_i a_i}{\sum (x_i - \overline{x})^2} - \frac{\overline{x} \sum a_i}{\sum (x_i - \overline{x})^2} - \frac{1}{\sum (x_i - \overline{x})^2} \\ &= 0 \end{aligned}$$

So $\text{var}(b_1) \geq \text{var}(\hat{\beta}_1)$ and therefore $\hat{\beta}_1$ is the best linear unbiased estimator (BLUE).

§3.2 Estimation of Variance

Using MML

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n}$$

Is it unbiased?

$$E\hat{\sigma}^2 = \frac{\sum E e_i^2}{n} = \frac{\sum \left[\text{var}(e_i) + (E e_i)^2\right]}{n}$$

<u>*Note*</u>: $e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$. So

$$Ee_i = E\left[Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right] = (\beta_0 + \beta_1 X_i) - (\beta_0 + \beta_1 X_i) = 0$$

Then,

$$E\hat{\sigma}^2 = \frac{\sum \operatorname{var}(e_i)}{n}$$

Notice that

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

or

$$e_i = Y_i - \overline{Y} - \hat{\beta}_1 (X_i - \overline{X})$$

where $\hat{Y}_i = \overline{Y} + \hat{\beta}_1(X_i - \overline{X})$. Substitute in and we get

$$\begin{aligned}
\operatorname{var}(e_i) &= \operatorname{var}\left[Y_i - \overline{Y} - \hat{\beta}_1(X_i - \overline{X})\right] \\
&= \operatorname{var}(Y_i) + \operatorname{var}(\overline{Y}) + (X_i - \overline{X})^2 \operatorname{var}(\hat{\beta}_1) - 2\operatorname{cov}(Y_i, \overline{Y}) - 2(X_i - \overline{X})\operatorname{cov}(Y_i, \hat{\beta}_1) \\
&\quad + 2(X_i - \overline{X})\operatorname{cov}(\overline{Y}, \hat{\beta}_1)
\end{aligned}$$

Let's compute each term there.

$$\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\
\operatorname{var}(Y_i) &= \sigma^2 \\
\overline{Y} &= \beta_0 + \beta_1 \overline{X} + \frac{\sum \varepsilon_i}{n} \\
\operatorname{var}(\overline{Y}) &= \frac{\sigma^2}{n} \\
\operatorname{cov}(Y_i, \overline{Y}) &= \operatorname{cov}\left(Y_i, \frac{Y_1 + \ldots + Y_i + \ldots + Y_n}{n}\right) \\
&= \frac{1}{n}\operatorname{cov}(Y_i, Y_1) + \ldots + \frac{1}{n}\operatorname{cov}(Y_i, Y_i) + \ldots + \frac{1}{n}\operatorname{cov}(Y_i, Y_n) \\
&= \frac{\sigma^2}{n} \\
\operatorname{cov}(Y_i, \hat{\beta}_1) &= \operatorname{cov}(Y_i, \sum k_i Y_i) \\
&= \operatorname{cov}(Y_i, k_1 Y_1) + \ldots + \operatorname{cov}(Y_i, k_i Y_i) + \ldots + \operatorname{cov}(Y_i, k_n Y_n) \\
&= k_1 \operatorname{cov}(Y_i, Y_1) + \ldots + k_i \operatorname{cov}(Y_i, Y_i) + \ldots + k_n \operatorname{cov}(Y_1, Y_n) \\
&= \sigma^2 k_i = \sigma^2 \frac{x_i - \overline{x}}{\sum (x_i - \overline{x})^2}
\end{aligned}$$

<u>*Note*</u>: A property of covariance

$$\operatorname{cov}(aY, bQ) = ab\operatorname{cov}(Y, Q)$$

And for the last term,

$$\begin{aligned}
\operatorname{cov}(\overline{Y}, \hat{\beta}_1) &= \operatorname{cov}\left(\frac{Y_1 + \ldots + Y_n}{n}, k_1 Y_1 + \ldots + k_n Y_n\right) \\
&= \operatorname{cov}(\frac{Y_1}{n}, k_1 Y_1 + \ldots + k_n Y_n) + \ldots + \operatorname{cov}(\frac{Y_n}{n}, k_1 Y_1 + \ldots + k_n Y_n) \\
&= \frac{\sigma^2}{n} k_1 + \frac{\sigma^2}{n} k_2 + \ldots + \frac{\sigma^2}{n} k_n \\
&= \frac{\sigma^2}{n} \sum k_i = 0
\end{aligned}$$

Now, we're ready to compute the variance

$$\begin{aligned}\operatorname{var}(e_i) &= \sigma^2 + \frac{\sigma^2}{n} + \frac{\sigma^2(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2} - \frac{2\sigma^2}{n} - \frac{2\sigma^2(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2} \\ &= \sigma^2\left(1-\frac{1}{n}-\frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)\end{aligned}$$

Therefore,

$$\begin{aligned}E\hat{\sigma}^2 &= \frac{\sum \operatorname{var}(e_i)}{n} = \sigma^2\frac{\sum_{i=1}^n\left(1-\frac{1}{n}-\frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)}{n} \\ &= \frac{(n-2)}{n}\sigma^2\end{aligned}$$

It follows that the unbiased estimator of σ^2 is

$$S_e^2 = \frac{n}{n-2}\sigma^2 = \frac{\sum e_i^2}{n-2}$$

§3.3 Distribution Theory

Let $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and we assume $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0,\sigma)$

$$\begin{aligned}\hat{\beta}_1 &= \sum k_i Y_i \implies \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum(x_i-\overline{x})^2}}\right) \\ \hat{\beta}_0 &= \sum l_i Y_i \implies \hat{\beta}_0 \sim N\left(\beta_0, \sigma\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{\sum(x_i-\overline{x})^2}}\right)\end{aligned}$$

We will show $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$ in the next lecture.

§4 Lec 4: Oct 4, 2021

§4.1 Centered Model

Consider the model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1, \dots, n$ and Gauss-Markov conditions hold, i.e.,

$$E[\varepsilon_i] = 0$$
$$\text{var}[\varepsilon_i] = \sigma^2$$

for $i = 1, \dots, n$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent (we assume $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$). This is non-centered model. Let's look at a centered model

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i \pm \beta_1 \overline{X} + \varepsilon_i \\ Y_i &= \beta_0 + \beta_1 \overline{X} + \beta_1 (X_i - \overline{X}) + \varepsilon_i \\ Y_i &= \gamma_0 + \beta_1 Z_i + \varepsilon_i \quad - \text{centered model} \end{aligned}$$

where $\gamma_0 = \beta_0 + \beta_1 \overline{X}$ and $Z_i = X_i - \overline{X}$.
Note: $\sum z_i = \sum (x_i - \overline{x}) = 0$ and $\overline{z} = 0$. So,

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (z_i - \overline{z}) y_i}{\sum (z_i - \overline{z})^2} = \frac{\sum z_i y_i}{\sum z_i^2} = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} \text{ – same as non-centered model} \\ \hat{\gamma}_0 &= \overline{y} - \hat{\beta}_1 \overline{z} = \overline{y} \end{aligned}$$

Notice $\hat{y}_i = \overline{y} + \hat{\beta}_1 (x_i - \overline{x})$ which is the same as $\hat{y}_i$ of the non-centered model.

§4.2 Distribution Theory Using the Centered Model

Have

$$\begin{aligned} Y_i &\sim N\left(\gamma_0 + \beta_1 \left(X_i - \overline{X}\right), \sigma\right) \\ \hat{\beta}_1 &\sim \left(\beta_1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right) \\ \hat{\gamma}_0 = \overline{Y} &\sim N\left(\gamma_0, \frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

Now, let's show that $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$. We have

$$\begin{aligned} \frac{Y_i - \gamma_0 - \beta_1 (X_i - \overline{X})}{\sigma} &\sim N(0, 1) \\ \frac{\left[Y_i - \gamma_0 - \beta_1 (X_i - \overline{X})\right]^2}{\sigma^2} &\sim \mathcal{X}_1^2 \end{aligned}$$

It follows that

$$\frac{\sum_{i=1}^n \left[Y_i - \gamma_0 - \beta_1 (X_i - \overline{X})\right]^2}{\sigma^2} \sim \mathcal{X}_n^2$$

Notice that $\frac{(n-2)S_e^2}{\sigma^2} = \frac{\sum e_i^2}{\sigma^2}$. Let's manipulate this expression. First, let

$$L = \frac{\sum \left[Y_i - \gamma_0 - \beta_1 (X_i - \overline{X}) \pm \hat{\gamma}_0 \pm \hat{\beta}_1 (X_i - \overline{X})\right]^2}{\sigma^2}$$

Then,

$$\begin{aligned} L &= \frac{\sum \left[y_i - \hat{\gamma}_0 - \hat{\beta}_1(x_i - \overline{x}) + (\hat{\gamma}_0 - \gamma_0) + (\hat{\beta}_1 - \beta_1)(x_i - \overline{x})\right]^2}{\sigma^2} \\ &= \frac{\sum \left[e_i + (\hat{\gamma}_0 - \gamma_0) + (\hat{\beta}_1 - \beta_1)(x_i - \overline{x}\right]^2}{\sigma^2} \\ &= \frac{\sum e_i^2}{\sigma^2} + \frac{n\left(\hat{\gamma}_0 - \gamma_0\right)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2 \sum (x_i - \overline{x})^2}{\sigma^2} + \frac{2(\hat{\gamma}_0 - \gamma_0)\sum e_i}{\sigma^2} \\ &+ \frac{2(\hat{\beta}_1 - \beta_1)\sum e_i(x_i - \overline{x})}{\sigma^2} + \frac{2(\hat{\gamma}_0 - \gamma_0)(\hat{\beta}_1 - \beta_1)\sum(x_i - \overline{x})}{\sigma^2} \end{aligned}$$

So far,

$$\underbrace{\frac{\sum \left[y_i - \gamma_0 - \beta_1(x_i - \overline{x})\right]^2}{\sigma^2}}_{\mathcal{X}_n^2} = \underbrace{\frac{(n-2)S_e^2}{\sigma^2}}_{?} + \underbrace{\frac{\hat{\gamma}_0 - \gamma_0}{\sigma/\sqrt{n}}}_{\mathcal{X}_1^2} + \underbrace{\left[\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{\sum(x_i - \overline{x})^2}}\right]^2}_{\mathcal{X}_1^2}$$

$$Q = Q_1 + Q_2 + Q_3$$

Let's use moment generating function to find "?". Notice that Q_1, Q_2, Q_3 are independent why?

$$\begin{aligned} M_Q(t) &= M_{Q_1+Q_2+Q_3} \\ M_Q(t) &= M_{Q_1}(t) \cdot M_{Q_2}(t) \cdot M_{Q_3}(t) \end{aligned}$$

We have

$$\begin{aligned} Q \sim \mathcal{X}_n^2 &\implies M_Q(t) = (1-2t)^{-\frac{n}{2}} \\ Q_2 \sim \mathcal{X}_1^2 &\implies M_{Q_2}(t) = (1-2t)^{-\frac{1}{2}} \\ Q_3 \sim \mathcal{X}_1^2 &\implies M_{Q_3}(t) = (1-2t)^{-\frac{1}{2}} \\ &\implies M_{Q_1}(t) = (1-2t)^{\frac{-n+2}{2}} \\ &\implies Q_1 = \frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2 \end{aligned}$$

<u>*Note*</u>: If $Y \sim \Gamma(\alpha, \beta)$ then

$$M_Y(t) = (1 - \beta t)^{-\alpha}$$

and

$$M_{cY}(t) = M_Y(ct)$$

Let's now find the distribution of s_e^2.

$$\begin{aligned} S_e^2 &= \frac{\sigma^2}{n-2} Q_1 \\ M_{S_e^2}(t) &= M_{\frac{\sigma^2}{n-2}Q_1}(t) = M_{Q_1}\left(\frac{\sigma^2}{n-2}t\right) \\ M_{S_e^2}(t) &= \left(1 - \frac{2\sigma^2}{n-2}t\right)^{\frac{-n+2}{2}} \end{aligned}$$

Therefore,

$$\begin{aligned} S_e^2 &\sim \Gamma\left(\frac{n-2}{2}, \frac{2\sigma^2}{n-2}\right) \\ ES_e^2 &= \sigma^2, \quad \text{var}(S_e^2) = \frac{2\sigma^4}{n-2} \end{aligned}$$

Another way to show this result is to use the non-centered model

$$\frac{\sum\left(Y_i-\beta_0-\beta_1 X_i \pm \hat{\beta}_0 \pm \hat{\beta}_1 X_i\right)^2}{\sigma^2}$$

§5 Lec 5: Oct 6, 2021

§5.1 Distribution Theory Using Non-Centered Model

Recall that we want to show $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$ using the non-centered model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ for $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. Then, $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma)$. Let

$$M = \frac{\sum \left(Y_i - \beta_0 - \beta_1 X_i \pm \hat{\beta}_0 \pm \hat{\beta}_1 X_i\right)^2}{\sigma^2} \sim \mathcal{X}_n^2$$

Then,

$$\begin{aligned} M &= \frac{\sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i + (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) x_i\right)^2}{\sigma^2} \\ &= \frac{\sum e_i^2}{\sigma^2} + \frac{n(\hat{\beta}_0 - \beta_0)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2 \sum x_i^2}{\sigma^2} + \frac{2(\hat{\beta}_0 - \beta_0)\sum e_i}{\sigma^2} + \frac{2(\hat{\beta}_1 - \beta_1)\sum e_i x_i}{\sigma^2} \\ &\quad + \frac{2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)\sum x_i}{\sigma^2} \\ &= \underbrace{\frac{\sum e_i^2}{\sigma^2}}_{\frac{(n-2)S_e^2}{\sigma^2}} + \underbrace{\frac{n(\hat{\beta}_0 - \beta_0)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2 \sum x_i^2}{\sigma^2} + \frac{2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)\sum x_i}{\sigma^2}}_{?} \qquad (**) \end{aligned}$$

Let $D = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} = \overline{Y}$ and consider

$$\frac{(\hat{\beta}_1 - \beta_1)^2}{\text{var}(\hat{\beta}_1)} + \frac{(D - (\beta_0 + \beta_1 \overline{x}))^2}{\text{var}(D)} \qquad (*)$$

<u>*Note*</u>: $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right)$ and

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ \overline{Y} &= \frac{\sum Y_i}{n} = \beta_0 + \beta_1 \overline{X} + \frac{\sum \varepsilon_i}{n} \end{aligned}$$

So $\overline{Y} \sim N\left(\beta_0 + \beta_1 \overline{X}, \frac{\sigma}{\sqrt{n}}\right)$ and thus $\frac{D - (\beta_0 + \beta_1 \overline{X})}{\sigma/\sqrt{n}} \sim N(0,1)$. It follows that each term in (*) follows chi-square distribution with 1 degree of freedom. Now, we have

$$\begin{aligned} (*) &= \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2} \sum (x_i - \overline{x})^2 + \frac{n(\hat{\beta}_0 - \beta_0)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2}{\sigma^2} n\overline{x}^2 + \frac{2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)}{\sigma^2} \sum x_i \\ &= \frac{(\hat{\beta}_1 - \beta_1)^2 \left(\sum x_i^2 - n\overline{x}^2\right)}{\sigma^2} + \frac{n(\hat{\beta}_0 - \beta_0)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2 n\overline{x}^2}{\sigma^2} + \frac{2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)\sum x_i}{\sigma^2} \end{aligned}$$

which is equivalent to the last three terms of (**). We just need to show that

$$\begin{aligned} \text{cov}(\overline{Y}, \hat{\beta}_1) &= 0 \\ \text{cov}(\overline{Y}, e_i) &= 0 \\ \text{cov}(\hat{\beta}_1, e_i) &= 0 \end{aligned}$$

> **Remark 5.1.** Under normality, zero covariance implies independence.

§5.2 A Note on Gamma Distribution

Let $Q \sim \Gamma(\alpha, \beta)$. Then

$$\begin{aligned} EQ &= \alpha\beta \\ \text{var}(Q) &= \alpha\beta^2 \\ M_Q(t) &= (1-\beta t)^{-\alpha} \\ EQ^k &= \frac{\Gamma(\alpha+k)\beta^k}{\Gamma(\alpha)} \end{aligned}$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

is the Gamma function.
Property:

$$\begin{aligned} \Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-1) \\ \Gamma(\alpha+1) &= \alpha\Gamma(\alpha) \end{aligned}$$

If α is an integer, then

$$\Gamma(\alpha) = (\alpha-1)!$$

Recall that $S_e^2 \sim \Gamma\left(\frac{n-2}{2}, \frac{2\sigma^2}{n-2}\right)$

$$ES_e^2 = \sigma^2, \quad \text{var}(S_e^2) = \frac{2\sigma^4}{n-2}$$

Is S_e unbiased estimator of σ?

$$\begin{aligned} ES_e &= E\left[S_e^2\right]^{\frac{1}{2}} \\ &= \frac{\Gamma\left(\frac{n-2}{2}+\frac{1}{2}\right)\left(\frac{2\sigma^2}{n-2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{n-2}{2}\right)} \\ &= \sigma\sqrt{\frac{2}{n-2}}\Gamma\left(\frac{n-1}{2}\right)/\Gamma\left(\frac{n-2}{2}\right) \\ &= \sigma A \end{aligned}$$

Thus, it's biased and we can adjust the result to be unbiased, i.e., $\frac{S_e}{A}$.
If $Y \sim \mathcal{X}_n^2$, then

$$M_Y(t) = (1-2t)^{-\frac{n}{2}}$$

which is $\Gamma\left(\frac{n}{2}, 2\right)$.

§5.3 Coefficient of Determination

Recall

$$\underbrace{\sum(y_i-\overline{y})^2}_{\text{SST}} = \underbrace{\sum e_i^2}_{\text{SSE}} + \underbrace{\sum(\hat{y}_i-\overline{y})^2}_{\text{SSR}}$$

where $\hat{Y}_i = \overline{y} + \hat{\beta}_1(x_i-\overline{x})$. We define R^2 as

$$R^2 = \frac{\text{SSR}}{\text{SST}} \quad \text{or} \quad R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

and $0 \leq R^2 \leq 1$. We have

$$\begin{aligned} \text{var}(\hat{Y}_i) &= \text{var}\left(\overline{y} + \hat{\beta}_1(x_i - \overline{x})\right) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right) \end{aligned}$$

Another way to show this is to express $\hat{Y}_i$ as a linear combination of $Y_1, \ldots, Y_n$.

$$\begin{aligned} \hat{Y}_i &= \overline{y} + \hat{\beta}_1(x_i - \overline{x}) \\ &= \frac{\sum y_j}{n} + (x_i - \overline{x}) \sum k_j y_j \\ &= \sum \left[\frac{1}{n} + (x_i - \overline{x})k_j\right] y_j \\ \text{var}(\hat{Y}_i) &= \sigma^2 \sum \left[\frac{1}{n} + (x_i - \overline{x})k_j\right]^2 \\ &= \sigma^2 \sum \left[\frac{1}{n^2} + (x_i - \overline{x})^2 k_j^2 + \frac{2}{n}(x_i - \overline{x})k_j\right] \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right) \end{aligned}$$

Consider

$$e_i = y_i - \hat{y}_i = y_i - \overline{y} - \hat{\beta}_1(x_i - \overline{x}) = \sum a_l y_l - \frac{\sum y_l}{n} - (x_i - \overline{x}) \sum k_l y_l = \sum \left[a_l - \frac{1}{n} - (x_i - \overline{x})k_l\right] y_l$$

where

$$a_l = \begin{cases} 1, & \text{if } l = i \\ 0, & \text{otherwise} \end{cases}$$

§6 Lec 6: Oct 8, 2021

§6.1 Variance & Covariance Operations

Have

$$\operatorname{cov}\left(\sum a_i Y_i, \sum b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \operatorname{cov}(Y_i, Y_j) = \sum a_i b_i \operatorname{cov}(Y_i, Y_i) = \sigma^2 \sum a_i b_i$$

because $Y_1, \ldots, Y_n$ are independent.

Example 6.1

Consider $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\begin{aligned}\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) &= \operatorname{cov}\left(\sum l_i Y_i, \sum k_i Y_j\right)\\ &= \sigma^2 \sum l_i k_i \\ &= \sigma^2 \sum \left[\left(\frac{1}{n} - k_i \overline{x}\right) k_i\right] \\ &= \sigma^2 \frac{1}{n} \sum k_i - \sigma^2 \overline{x} \sum k_i^2 \\ &= -\frac{\sigma^2 \overline{x}}{\sum (x_i - \overline{x})^2}\end{aligned}$$

Or

$$\begin{aligned}\operatorname{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) &= \operatorname{cov}\left(\overline{Y} - \hat{\beta}_1 \overline{X}, \hat{\beta}_1\right)\\ &= \operatorname{cov}\left(\overline{Y}, \hat{\beta}_1\right) - \overline{X} \operatorname{var}(\hat{\beta}_1) \\ &= \frac{-\overline{x}\sigma^2}{\sum (x_i - \overline{x})^2}\end{aligned}$$

Example 6.2

Consider $\hat{Y}_i$ and $\hat{Y}_j$

$$\begin{aligned}\operatorname{cov}\left(\hat{Y}_i, \hat{Y}_j\right) &= \operatorname{cov}\left(\overline{y} + \hat{\beta}_1 (x_i - \overline{x}), \overline{y} + \hat{\beta}_1 (x_j - \overline{x})\right)\\ &= \frac{\sigma^2}{n} + 0 + 0 + \frac{(x_i - \overline{x})(x_j - \overline{x})}{\sum (x_i - \overline{x})^2} \sigma^2 \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \overline{x})(x_j - \overline{x})}{\sum (x_i - \overline{x})^2}\right)\end{aligned}$$

When $i = j$,

$$\operatorname{var}(\hat{Y}_i) = \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right)$$

Example 6.3 (Cont'd)
Notice that

$$\begin{aligned}\hat{Y}_i &= \overline{y} + \hat{\beta}_1(x_i - \overline{x}) = \frac{\sum y_l}{n} + (x_i - \overline{x}) \sum k_l y_l \\ &= \sum \left[\frac{1}{n} + (x_i - \overline{x})k_l\right] y_l = \sum a_l y_l \\ \hat{Y}_j &= \ldots = \sum b_v y_v \\ \operatorname{cov}\left(\hat{Y}_i, \hat{Y}_j\right) &= \sigma^2 \sum a_l b_l \\ &= \sigma^2 \sum \left[\frac{1}{n} + (x_i - \overline{x})k_l\right]\left[\frac{1}{n} + (x_j - \overline{x})k_l\right] \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \overline{x})(x_j - \overline{x})}{\sum (x_i - \overline{x})^2}\right)\end{aligned}$$

§6.2 Inference

Construct a confidence interval $1 - \alpha$ for β_1

$$P\left(L \leq \beta_1 \leq U\right) = 1 - \alpha$$

Know

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right)$$

and

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$$

Consider

$$\operatorname{cov}\left(\hat{\beta}_1, e_i\right) = 0$$

Under normality, since their covariance is 0, $\hat{\beta}_1$ and S_e^2 are independent. Thus,

$$\frac{\frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \overline{x})^2}}}{\sqrt{\frac{(n-2)S_e^2}{\sigma^2} / (n-2)}} = \frac{\hat{\beta}_1 - \beta_1}{S_e / \sqrt{\sum (x_i - \overline{x})^2}} \sim t_{n-2}$$

Pivot Method:

$$P\left(-t_{\frac{\alpha}{2}; n-2} \leq \frac{\hat{\beta}_1 - \beta_1}{S_e / \sqrt{\sum (x_i - \overline{x})^2}} \leq t_{\frac{\alpha}{2}; n-2}\right) = 1 - \alpha$$

and after some manipulation we get

$$P\left(\hat{\beta}_1 - t_{\frac{\alpha}{2}; n-2} \cdot \frac{S_e}{\sqrt{\sum (x_i - \overline{x})^2}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\frac{\alpha}{2}; n-2} \cdot \frac{S_e}{\sqrt{\sum (x_i - \overline{x})^2}}\right) = 1 - \alpha$$

We are $1 - \alpha$ confident that

$$\beta_1 \in \left[\hat{\beta}_1 \pm t_{\frac{\alpha}{2}; n-2} \cdot \frac{S_e}{\sqrt{\sum (x_i - \overline{x})^2}}\right]$$

For $\hat{\beta}_0$,

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}}\right)$$

and we proceed similarly to obtain

$$\beta_0 \in \left[\hat{\beta}_0 \pm t_{\frac{\alpha}{2};n-2} \cdot S_e\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}}\right]$$

Say if we want to construct a confidence interval for $\beta_0 - 2\beta_1$:

$$\begin{aligned}
\operatorname{var}(\hat{\beta}_0 - 2\hat{\beta}_1) &= \operatorname{var}(\hat{\beta}_0) + 4\operatorname{var}(\hat{\beta}_1) - 4\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) \\
&= \sigma^2\left[\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2} + \frac{4}{\sum(x_i - \overline{x})^2} + \frac{4\overline{x}}{\sum(x_i - \overline{x})^2}\right] \\
&= \sigma^2\left[\frac{1}{n} + \frac{(\overline{x}+2)^2}{\sum(x_i - \overline{x})^2}\right]
\end{aligned}$$

So,

$$\hat{\beta}_0 - 2\hat{\beta}_1 \sim N\left(\beta_0 - 2\beta_1, \sigma\sqrt{\frac{1}{n} + \frac{(\overline{x}+2)^2}{\sum(x_i - \overline{x})^2}}\right)$$

Thus, the C.I. is

$$\beta_0 - 2\beta_1 \in \left[\hat{\beta}_0 - 2\hat{\beta}_1 \pm t_{\frac{\alpha}{2};n-2} \cdot S_e\sqrt{\frac{1}{n} + \frac{(\overline{x}+2)^2}{\sum(x_i - \overline{x})^2}}\right]$$

§6.3 Prediction Interval

Prediction interval for Y_0 when $X = X_0$. Let's begin with error of prediction: $Y_0 - \hat{Y}_0$. We know

- $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- $Y_0 = \beta_0 + \beta_1 X_0 + \varepsilon_0$
- $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$

So

$$\begin{aligned}
E(Y_0 - \hat{Y}_0) &= 0 \\
\operatorname{var}(Y_0 - \hat{Y}_0) &= \operatorname{var}(Y_0) + \operatorname{var}(\hat{Y}_0) - 2\operatorname{cov}(Y_0, \hat{Y}_0) \\
&= \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)
\end{aligned}$$

We apply the same procedure in the inference section

$$\left.\begin{aligned}
&Y_0 - \hat{Y}_0 \sim N\left(0, \sigma\sqrt{1 + \tfrac{1}{n} + \tfrac{(x_0 - \overline{x})^2}{\sum(x_i - \overline{x})^2}}\right) \\
&\tfrac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2
\end{aligned}\right\} \implies Y_0 \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-2} S_e\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum(x_i - \overline{x})^2}}$$

C.I. for EY_0 for a given $X = X_0$

$$\hat{Y}_0 \sim N\left(\beta_0 + \beta_1 X_0, \sigma\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum(x_i - \overline{x})^2}}\right)$$

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$$

$$\implies EY_0 \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-2} \cdot S_e\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum(x_i - \overline{x})^2}}$$

§7 Lec 7: Oct 11, 2021

§7.1 Hypothesis Testing

Consider the model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Example 7.1

Hypothesis testing examples

$$\begin{aligned} H_0 : \beta_1 = 0, &\quad H_a : \beta_1 \neq 0 \\ H_0 : \beta_1 = 1, &\quad H_a : \beta_1 \neq 1 \\ H_0 : \beta_0 = 0, &\quad H_a : \beta_0 \neq 0 \\ H_0 : \beta_0 + \beta_1 = 0, &\quad H_a : \beta_0 + \beta_1 \neq 0 \\ H_0 : \left.\begin{matrix} \beta_0 = \beta_0^* \\ \beta_1 = \beta_1^* \end{matrix}\right\}, &\quad H_a : \text{ not true} \end{aligned}$$

Let's consider the following two-sided test

$$\begin{aligned} H_0 :&\ \beta_1 = 0 \\ H_a :&\ \beta_1 \neq 0 \end{aligned}$$

Recall under H_0,

$$\left.\begin{matrix} \hat{\beta}_1 \sim N\left(0, \frac{\sigma}{\sqrt{\sum(x_i - \overline{x})^2}}\right) \\ \frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2 \end{matrix}\right\} \implies t = \frac{\hat{\beta}_1}{S_e / \sqrt{\sum(x_i - \overline{x})^2}} \sim t_{n-2}$$

We reject H_0 if $t > t_{\frac{\alpha}{2};n-2}$ or $t < -t_{\frac{\alpha}{2};n-2}$. Using a $1-\alpha$ C.I.

$$\beta_1 \in \hat{\beta}_1 \pm t_{\frac{\alpha}{2};n-2} \frac{S_e}{\sqrt{\sum(x_i - \overline{x})^2}}$$

For example, for $-2 \leq \beta_1 \leq 2$, we do not reject H_0.

$$p - \text{value} = 2P\left(t > t^*\right)$$

We reject H_0 if p-value $< \alpha$.
Test $H_0 :\ \beta_1 = 0$ using the F statistics. Under H_0,

$$\hat{\beta}_1 \sim N\left(0, \frac{\sigma}{\sqrt{\sum(x_i - \overline{x})^2}}\right)$$

$$\frac{\hat{\beta}_1 - 0}{\sigma / \sqrt{\sum(x_i - \overline{x})^2}} \sim N(0,1)$$

Then,

$$\frac{\hat{\beta}_1^2 \sum(x_i - \overline{x})^2}{\sigma^2} \sim \mathcal{X}_1^2$$

and we know

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$$

Therefore, we can form the F statistics

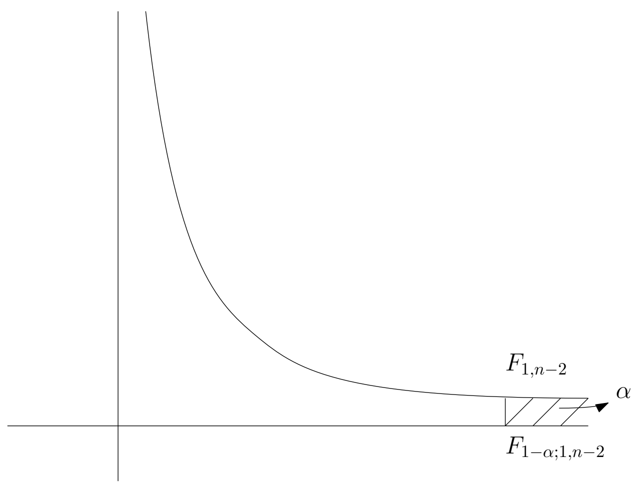
$$\frac{\frac{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}{\sigma^2}/1}{\frac{(n-2)S_e^2}{\sigma^2}/(n-2)} = \frac{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}{S_e^2} \sim F_{1,\,n-2}$$

Definition 7.2 (F Distribution) — Let $U \sim \mathcal{X}_n^2$ and $V \sim \mathcal{X}_m^2$ and U, V are independent. Then,

$$\frac{\frac{U}{n}}{\frac{V}{m}} \sim F_{n,m}$$

We can observe that $t_{n-2}^2 = F_{1,\,n-2}$. In general,

$$\begin{gathered} Z \sim N(0,1) \\ U \sim \mathcal{X}_n^2 \\ Z,\ U \text{ are independent} \\ \frac{Z}{\sqrt{U/n}} \sim t_n \implies \frac{Z^2/1}{U/n} \sim F_{1,\,n} \end{gathered}$$



Let's find the expected value of the F statistics.

- Denominator:
$$ES_e^2 = \sigma^2$$
- Numerator:
$$\begin{aligned} E\hat{\beta}_1^2 \sum (x_i - \overline{x})^2 &= \sum (x_i - \overline{x})^2 E\hat{\beta}_1^2 \\ &= \sum (x_i - \overline{x})^2 \left(\text{var}(\hat{\beta}_1 + (E\hat{\beta}_1)^2\right) \\ &= \sum (x_i - \overline{x})^2 \left(\frac{\sigma^2}{\sum (x_i - \overline{x})^2} + \beta_1^2\right) \\ &= \sigma^2 + \beta_1^2 \sum (x_i - \overline{x})^2 \end{aligned}$$

Under H_0 the ratio is approximately equal to 1. If H_0 is not true the ratio is greater than 1.

Now, for $\hat{\beta}_0$,

$$\left.\begin{aligned} &\hat{\beta}_0 \sim N\left(0, \sigma\sqrt{\tfrac{1}{n} + \tfrac{\overline{x}^2}{\sum (x_i - \overline{x})^2}}\right) \\ &\tfrac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2 \end{aligned}\right\} \implies t = \frac{\hat{\beta}_0}{S_e\sqrt{\tfrac{1}{n} + \tfrac{\overline{x}^2}{\sum (x_i - \overline{x})^2}}} \sim t_{n-2}$$

and consider $H_0 : \beta_1 = 1 (\beta_1 - 1 = 0)$ and $H_a : \beta_1 \neq 1\ (\beta_1 - 1 \neq 0)$. Then under H_0,

$$\frac{\hat{\beta}_1 - 1}{\sigma / \sqrt{\sum (x_i - \overline{x})^2}} \sim N(0, 1)$$

Test Statistics:

$$\frac{\hat{\beta}_1 - 1}{S_e / \sqrt{\sum (x_i - \overline{x})^2}} \sim t_{n-2}$$

Using F statistics

$$\frac{(\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2}{\sigma^2} \sim \mathcal{X}_1^2$$

and thus

$$\frac{(\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2}{S_e^2} \sim F_{1,\, n-2}$$

§8 Lec 8: Oct 13, 2021

§8.1 Likelihood Ratio Test

Consider

$$\begin{aligned} Y_i &= \beta_1 X_i + \varepsilon_i \\ H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned}$$

We know

$$\begin{aligned} \hat{\beta}_1 &\sim N\left(0, \frac{\sigma}{\sqrt{\sum x_i^2}}\right) \\ \frac{(n-1)S_e^2}{\sigma^2} &\sim \mathcal{X}_{n-1}^2 \end{aligned}$$

So t_{test}: $\frac{\hat{\beta}_1}{S_e/\sqrt{\sum x_i^2}} \sim t_{n-1}$ and F_{test}: $\frac{\hat{\beta}_1^2 \sum x_i^2}{S_e^2} \sim F_{1,n-1}$.

Likelihood Ratio Test (LRT):

$$\begin{aligned} \text{For testing: } & H_0 : \beta_1 = 0 \\ \text{For the model: } & Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \end{aligned}$$

Show that this LRT is equivalent to the F statistic.

We reject H_0 if

$$\Lambda = \frac{L(\hat{w})}{L(\hat{\omega})} < k$$

where $L(\hat{w})$ is the maximized likelihood function under H_0 and $L(\hat{\omega})$ is maximized likelihood function under no restrictions. Under $H_0 : \beta_1 = 0$, we have $Y_i = \beta_0 + \varepsilon_i$. The likelihood function is

$$\begin{aligned} L &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum (y_i - \beta_0)^2} \\ \ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (y_i - \beta_0)^2 \\ \hat{\beta}_0 &= \overline{y} \\ \hat{\sigma}_0^2 &= \frac{\sum (y_i - \overline{y})^2}{n} \end{aligned}$$

Under no restriction, the estimates are the MLEs of $\beta_0, \beta_1, \sigma^2$ which are $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}_1^2 = \frac{\sum e_i^2}{n}$. Back to LRT, we have

$$\begin{aligned} \Lambda &= \frac{L(\hat{w})}{L(\hat{\omega})} \\ &= \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_0^2}\sum (y_i - \overline{y})^2}}{(2\pi\sigma_1^2) e^{-\frac{1}{2\sigma_1^2}\sum e_i^2}} < k \end{aligned}$$

Note:

$$\begin{aligned} \sum (y_i - \overline{y})^2 &= n\sigma_0^2 \\ \sum e_i^2 &= n\sigma_1^2 \end{aligned}$$

So,

$$\frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}}{(2\pi\hat{\sigma}_1^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}} < k$$
$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} < k^{\frac{2}{n}}$$
$$\frac{\sum e_i^2/n}{\sum(y_i-\overline{y})^2/n} < k^{\frac{2}{n}}$$

Notice that

$$\sum(y_i-\overline{y})^2 = \sum e_i^2 + \sum(\hat{y}_i-\overline{y})^2$$
$$\sum(y_i-\overline{y})^2 = \sum e_i^2 + \hat{\beta}_1^2\sum(x_i-\overline{x})^2$$

So,

$$\frac{\sum e_i^2}{\sum e_i^2 + \hat{\beta}_1^2\sum(x_i-\overline{x})^2} < k^{\frac{2}{n}}$$
$$\frac{1}{1+\frac{\hat{\beta}_1^2\sum(x_i-\overline{x})^2}{\sum e_i^2}} < k^{\frac{2}{n}}$$
$$\frac{\hat{\beta}_1^2\sum(x_i-\overline{x})^2}{(n-2)S_e^2} > k^{-\frac{2}{n}} - 1$$
$$\frac{\hat{\beta}_1\sum(x_i-\overline{x})^2}{S_e^2} > (n-2)\left(k^{-\frac{2}{n}}-1\right) = k'$$

We reject H_0 if

$$\frac{\hat{\beta}_1^2\sum(x_i-\overline{x})^2}{S_e^2} > k'$$

Recall we stated that we reject H_0 if $\Lambda = \frac{L(\hat{w})}{L(\hat{\omega})} < k$. Let's find k. First, we need α (type I error). Before that, we know

$$\frac{\hat{\beta}_1^2\sum(x_i-\overline{x})^2}{S_e^2} \sim F_{1,n-2}$$

So,

$$P\left(F_{1,n-2} > k' \middle| H_0 \text{ is true}\right) = \alpha$$

§8.2 Power Analysis in Simple Regression

Using the non-central t distribution

Definition 8.1 (Non-central t) — Let $Z \sim N(\delta,1)$ and $U \sim \mathcal{X}_n^2$ and Z and U are independent. Then,

$$\frac{Z}{\sqrt{U/n}} \sim t_n \text{ (NCP} = \delta)$$

Back to the t ratio. If H_0 is true,

$$\frac{\frac{\hat{\beta}_1}{\sigma/\sqrt{\sum(x_i-\overline{x})^2}}}{\sqrt{\frac{(n-2)S_e^2}{\sigma^2}/(n-2)}}$$

follows central t_{n-2} in which the numerator follows standard normal distribution. If H_0 is not true, then the numerator follows $N\left(\frac{\beta_1\sqrt{\sum(x_i-\overline{x})^2}}{\sigma}, 1\right)$. Thus, the ratio follows t_{n-2} (NCP = $\frac{\beta_1\sqrt{\sum(x_i-\overline{x})^2}}{\sigma}$). Finally, the power is

$$1-\beta = P\left(t_{n-2}(\text{NCP}) > t_{\frac{\alpha}{2};n-2}\right) + P\left(t_{n-2})\text{NCP}) < -t_{\frac{\alpha}{2};n-2}\right)$$

§9 Lec 9: Oct 15, 2021

§9.1 Extra Sum of Squares Method

So far, we have learnt several ways for hypothesis testing for, e.g.,

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned}$$

which are

1. t statistics
2. F statistics
3. Likelihood ratio test
4. Extra sum of square principle (reduced and full model)

$$\frac{(SSE_R - SSE_F)/(df_R - df_F)}{SSE_F/df_F} \sim F_{1,n-2}$$

$$\left.\begin{aligned} SSE_F &= \sum e_i^2 \\ df_F &= n-2 \end{aligned}\right\}$$

Under H_0: $\beta_1 = 0$ we have a reduced model

$$Y_i = \beta_0 + \varepsilon_i \implies \hat{\beta}_0 = \overline{y}$$

Therefore $SSE_R = \sum (y_i - \overline{y})^2$ and $df_R = n - 1$. Thus,

$$\frac{\left(\sum (y_i - \overline{y})^2 - \sum e_i^2\right) / (n - 1 - (n-2))}{\sum e_i^2/(n-2)}$$

<u>*Note*</u>:

$$\underbrace{\sum (y_i - \overline{y})^2}_{\text{SST}} = \underbrace{\sum e_i^2}_{\text{SSE}} + \underbrace{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}_{\text{SSR}}$$

So,

$$\frac{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}{S_e^2} \sim F_{1,n-2}$$

$$\left(\frac{\hat{\beta}_1}{S_e/\sqrt{\sum (x_i - \overline{x})^2}}\right)^2 \sim t_{n-2}^2$$

Example 9.1

Use the extra sum of squares method to test

$$H_0 : \beta_1 = 1$$
$$H_a : \beta_1 \neq 1$$

Reduced model: $Y_i = \beta_0 + x_i + \varepsilon_i$

$$\begin{aligned}
Y_i - x_i &= \beta_0 + \varepsilon_i \\
\hat{\beta}_0 &= \overline{y} - \overline{x} \\
SSE_R &= \sum \left(y_i - x_i - (\overline{y} - \overline{x})\right)^2 \\
&= \sum \left(y_i - \overline{y} - (x_i - \overline{x})\right)^2 \\
&= \sum (y_i - \overline{y})^2 + \sum (x_i - \overline{x})^2 - 2\sum (x_i - \overline{x})(y_i - \overline{y}) \qquad (*)
\end{aligned}$$

<u>*Note*</u>:

$$\begin{aligned}
\sum (y_i - \overline{y})^2 &= \sum e_i^2 + \hat{\beta}_1^2 \sum (x_i - \overline{x})^2 \\
\hat{\beta}_1 &= \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} \\
\implies \sum (x_i - \overline{x})(y_i - \overline{y}) &= \hat{\beta}_1 \sum (x_i - \overline{x})^2
\end{aligned}$$

So, we have

$$\begin{aligned}
(*) &= \sum e_i^2 + \hat{\beta}_1^2 \sum (x_i - \overline{x})^2 + \sum (x_i - \overline{x})^2 - 2\hat{\beta}_1 \sum (x_i - \overline{x})^2 \\
SSE_R &= \sum e_i^2 + (\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2
\end{aligned}$$

Test statistics:

$$\frac{(SSE_R - SSE_F)/(df_R - df_F)}{SSE_F/df_F}$$

$$\frac{\left(\sum e_i^2 + (\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2 - \sum e_i^2\right) / (n - 1 - (n-2))}{\sum e_i^2/(n-2)}$$

$$\frac{(\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2}{S_e^2} \sim F_{1,n-2}$$

Proof. Under H_0,

$$\begin{cases} \hat{\beta}_1 \sim N\left(1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right) \\ \frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2 \end{cases}$$

So,

$$\frac{\left[\frac{(\hat{\beta}_1 - 1)}{\sigma/\sqrt{\sum (x_i - \overline{x})^2}}\right]^2 /1}{\frac{(n-2)S_e^2}{\sigma^2}/(n-2)}$$

$$\frac{(\hat{\beta}_1 - 1)^2 \sum (x_i - \overline{x})^2}{S_e^2} \sim F_{1,n-2}$$ □

§9.2 Power Analysis Using Non-Central F Distribution

Definition 9.2 — 1. $Y \sim N(\mu, 1)$ then $Y^2 \sim \mathcal{X}_1^2 \ (\theta = \mu^2)$

2. Suppose $Y \sim N(\mu, \sigma)$

$$\frac{Y}{\sigma} \sim N\left(\frac{\mu}{\sigma}, 1\right)$$
$$\frac{Y^2}{\sigma^2} \sim \mathcal{X}_1^2 \ (\theta = \frac{\mu^2}{\sigma^2})$$

MGF of $Y \sim \mathcal{X}_1^2$ (NCP $= \theta$). Then

$$M_Y(t) = (1-2t)^{-\frac{1}{2}} e^{\theta \frac{t}{1-2t}}$$

If $\theta = 0 \implies M_Y(t) = (1-2t)^{-\frac{1}{2}}$.
Consider now

$$Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$$

Find distribution of $Q = \frac{Y_1^2}{\sigma^2} + \ldots + \frac{Y_n^2}{\sigma^2}$.

$$\begin{aligned} M_Q(t) &= \left[(1-2t)^{-\frac{1}{2}} e^{\frac{\mu^2}{\sigma^2} \frac{t}{1-2t}}\right]^n \\ &= (1-2t)^{-\frac{n}{2}} e^{\frac{n\mu^2}{\sigma^2} \frac{t}{1-2t}} \\ Q &= \frac{\sum Y_i^2}{\sigma^2} \sim \mathcal{X}_n^2 \ \left(\theta = \frac{n\mu^2}{\sigma^2}\right) \end{aligned}$$

<u>Non-Central F Distribution</u>: Let $U \sim \mathcal{X}_n^2$ (NCP $= \theta$) and $V \sim \mathcal{X}_m^2$. If U, V are independent, then

$$\frac{U/n}{V/m} \sim F_{n,m} \ (\text{NCP} = \theta)$$

Back to simple regression:

$$\begin{aligned} \hat{\beta}_1 &\sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right) \\ \frac{\hat{\beta}_1}{\sigma / \sqrt{\sum (x_i - \overline{x})^2}} &\sim N\left(\frac{\beta_1}{\sigma / \sqrt{\sum (x_i - \overline{x})^2}}, 1\right) \\ \frac{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}{\sigma^2} &\sim \mathcal{X}_1^2 \ \left(\theta = \frac{\beta_1^2 \sum (x_i - \overline{x})^2}{\sigma^2}\right) \\ \frac{(n-2) S_e^2}{\sigma^2} &\sim \mathcal{X}_{n-2}^2 \\ \frac{\frac{\hat{\beta}_1^2 \sum (x_i - \overline{x})^2}{\sigma^2} / 1}{\frac{(n-2) S_e^2}{\sigma^2} / (n-2)} &\sim F_{1,n-2} \ \left(\theta = \frac{\beta_1^2 \sum (x_i - \overline{x})^2}{\sigma^2}\right) \end{aligned}$$

Thus,

$$1 - \beta = P\left(F_{1,n-2}(\theta) > F_{1-\alpha;1,n-2}\right)$$

§10 Lec 10: Oct 18, 2021

§10.1 Multiple Regression

Consider:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, \ldots, n$$

where we have k predictors

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{12} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\begin{aligned} \mathbf{Y} &: n \times 1 \text{ response vector} \\ \mathbf{X} &: n \times (k+1) \text{ regression matrix} \\ \boldsymbol{\beta} &: (k+1) \times 1 \text{ parameter vector} \\ \boldsymbol{\varepsilon} &: n \times 1 \text{ error vector} \end{aligned}$$

Assumption: Gauss-Markov conditions

$$\left.\begin{array}{l} E[\varepsilon_i] = 0, \quad i = 1, \ldots, n \\ \operatorname{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \ldots, n \\ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \text{ are independent} \end{array}\right\} \implies E[\boldsymbol{\varepsilon}] = \mathbf{0}, \ \operatorname{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ be a random vector with mean vector

$$\boldsymbol{\mu} = E[\mathbf{Y}] = E \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} EY_1 \\ \vdots \\ EY_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

and variance covariance matrix

$$\boldsymbol{\Sigma} = E[\mathbf{Y} - \boldsymbol{\mu}][\mathbf{Y} - \boldsymbol{\mu}] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix}$$

$$E \begin{bmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \\ \vdots \\ Y_n - \mu_n \end{bmatrix} \begin{bmatrix} Y_1 - \mu_1 & Y_2 - \mu_2 & \ldots & Y_n - \mu_n \end{bmatrix}$$

Properties: Let $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ be a vector of constants and let $\mathbf{a}'\mathbf{Y}$ be a linear combination $\mathbf{Y}$. Then

$$\begin{aligned} E\left[\mathbf{a}'\mathbf{Y}\right] &= \mathbf{a}'E\mathbf{Y} = \mathbf{a}'\boldsymbol{\mu} = \sum a_i\mu_i \\ \operatorname{var}\left(\mathbf{a}'\mathbf{Y}\right) &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \end{aligned}$$

Let $\mathbf{A}$ be an $m \times n$ matrix of constant and consider $\mathbf{AY}$ ($m \times 1$ vector). Then

$$\begin{aligned} E\left[\mathbf{AY}\right] &= \mathbf{A}E\mathbf{Y} = \mathbf{A}\boldsymbol{\mu} \\ \operatorname{var}\left(\mathbf{AY}\right) &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' \end{aligned}$$

Using the Gauss-Markov conditions

$$\begin{aligned} E\mathbf{Y} &= E\left[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right] = \mathbf{X}\boldsymbol{\beta} \\ \operatorname{var}(\mathbf{Y}) &= \operatorname{var}\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{I} \end{aligned}$$

Estimation of β using Least Squares:

1. Geometric interpretation of least squares – orthogonal projection

$$\begin{aligned} \mathbf{X}'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) &= \mathbf{0} \\ \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= \mathbf{0} \\ \mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}'\mathbf{Y} \\ \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

   which is the least squares estimator of β.

2. Minimize the error sum of squares

$$\min Q = \sum \varepsilon_i^2$$

   or $\min Q = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ but $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Or

$$\min Q = \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)$$

   Then,

$$\begin{aligned} \min Q &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ \frac{\partial Q}{\partial \boldsymbol{\beta}} &= \mathbf{0} \qquad (*) \end{aligned}$$

Note: Matrix and vector differentiation:

Let $\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$ and $g(\boldsymbol{\theta})$ be a function of $\boldsymbol{\theta}$. Then

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial g(\boldsymbol{\theta})}{\partial \theta_p} \end{pmatrix}$$

Let $g(\boldsymbol{\theta}) = \mathbf{c}'\boldsymbol{\theta}$. Then,

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{c}$$

Let $\mathbf{A}$ be a symmetric matrix and consider $g(\boldsymbol{\theta}) = \boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}$. Then,

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$$

So apply these result to (*), we obtain

$$\begin{gathered} 2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \\ \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \end{gathered}$$

which is known as the normal equations. Notice that

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

which is OLS estimator of $\boldsymbol{\beta}$.

§11 Lec 11: Oct 20, 2021

§11.1 Multiple Regression (Cont'd)

Recall that

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$E\left[\boldsymbol{\varepsilon}\right] = \mathbf{0}$$
$$\text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

Least squares:

$$\min \sum \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Normal Equations:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \implies \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

Note that $\mathbf{X}$ is not a square matrix, so $\mathbf{X}'\mathbf{X}$ has to go together in order for it to be invertible.

$$\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_k' \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} n & \mathbf{1}'\mathbf{x}_1 & \mathbf{1}'\mathbf{x}_2 & \dots & \mathbf{1}'\mathbf{x}_k \\ \mathbf{x}_1'\mathbf{1} & \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \dots & \mathbf{x}_1'\mathbf{x}_k \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{x}_k'\mathbf{1} & \mathbf{x}_k'\mathbf{x}_1 & \dots & \mathbf{x}_k'\mathbf{x}_k & \end{bmatrix}$$

which is a symmetric $(k+1) \times (k+1)$ matrix. We have

$$\mathbf{x_1}\mathbf{x_1} = \sum x_{i1}^2$$
$$\mathbf{x_1}'\mathbf{x_2} = \sum x_{i1}x_{i2}$$

Partition $\mathbf{X}$ and $\boldsymbol{\beta}$

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{X_{(0)}} \end{pmatrix}$$
$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta_{(0)}} \end{pmatrix}$$

Model:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{1} & \mathbf{X_{(0)}} \end{pmatrix} \begin{pmatrix} \beta_0 & \boldsymbol{\beta_{(0)}} \end{pmatrix} + \boldsymbol{\varepsilon}$$
$$\mathbf{Y} = \beta_0 \mathbf{1} + \mathbf{X_{(0)}}\boldsymbol{\beta_{(0)}} + \boldsymbol{\varepsilon}$$

Then,

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{X_{(0)}}' \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{X_{(0)}} \end{pmatrix}$$
$$= \begin{pmatrix} n & \mathbf{1}'\mathbf{X_{(0)}} \\ \mathbf{X_{(0)}}'\mathbf{1} & \mathbf{X_{(0)}}'\mathbf{X_{(0)}} \end{pmatrix}$$

So

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta_{(0)}}} \end{bmatrix} = \begin{bmatrix} n & \mathbf{1}'\mathbf{X_{(0)}} \\ \mathbf{X_{(0)}}'\mathbf{1} & \mathbf{X_{(0)}}'\mathbf{X_{(0)}} \end{bmatrix} \begin{bmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{X_{(0)}}\mathbf{Y} \end{bmatrix}$$

Also,

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \mathbf{1}' \\ \mathbf{X_{(0)}}' \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{X}'_{(\mathbf{0})}\mathbf{Y} \end{bmatrix}$$

Fitted Values:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

or

$$\hat{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

or

$$\hat{Y} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

where $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ which is $n \times n$ "hat" matrix.
Properties of $\mathbf{H}$:

1. $\mathbf{H}' = \mathbf{H}$ symmetric
$$\left(\mathbf{ABC}\right)' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$$
2. $\mathbf{HH} = \mathbf{H}$ – idempotent
$$\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X} = \mathbf{H}$$
3. $\operatorname{tr}\mathbf{H} = \operatorname{tr}\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] = \operatorname{tr}\left[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\right] = \operatorname{tr}\mathbf{I}_{k+1} = k+1$. Notice that the property of trace is
$$\operatorname{tr}\left(\mathbf{ABC}\right) = \operatorname{tr}\left(\mathbf{BCA}\right) = \operatorname{tr}\left(\mathbf{CAB}\right) \neq \operatorname{tr}\left(\mathbf{BAC}\right)$$
4. $\mathbf{HX} = \mathbf{X}$ or $\mathbf{H}\left(\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_k}\right) = \left(\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_k}\right)$

Residuals:

$$\begin{aligned} e_i &= y_i - \hat{y}_i \quad i = 1, \ldots, n \\ \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} \\ \mathbf{e} &= \mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}} \\ \mathbf{e} &= \mathbf{Y} - \mathbf{HY} \\ \mathbf{e} &= \left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y} = \left(\mathbf{I} - \mathbf{H}\right)\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\ &= \left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon} \\ &= \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon} \end{aligned}$$

Overall, we have two expressions for $\mathbf{e}$

$$\begin{aligned} \mathbf{e} &= \left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y} \\ \mathbf{e} &= \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon} \end{aligned}$$

Notice that the error sum of squares

$$\text{SSE} = \sum e_i^2 = \mathbf{e}'\mathbf{e} = \left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y}\right]'\left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y}\right] = \mathbf{Y}'\left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y}$$

or

$$\text{SSE} = \left[\left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}\right]'\left[\left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}\right] = \boldsymbol{\varepsilon}'\left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}$$

Properties of $\hat{\beta}$:

$$E\hat{\boldsymbol{\beta}} = E\left[\left(\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\mathbf{Y}\right)\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\underbrace{E\mathbf{Y}}_{=\boldsymbol{\beta}} = \boldsymbol{\beta}$$

which is unbiased.

$$\begin{aligned}\operatorname{var}(\boldsymbol{\beta}) &= \operatorname{var}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\end{aligned}$$

which is variance covariance matrix of $\hat{\boldsymbol{\beta}}$. Specifically,

$$\begin{aligned}\operatorname{var}\left(\hat{\boldsymbol{\beta}}\right) &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2\begin{bmatrix} v_{00} & v_{01} & \dots & v_{0k} \\ v_{10} & v_{11} & \dots & v_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k0} & v_{k1} & \dots & v_{kk}\end{bmatrix} \\ \operatorname{var}\left(\hat{\beta}_0\right) &= \sigma^2 v_{00} \\ \operatorname{var}(\hat{\beta}_1) &= \sigma^2 v_{11} \\ \operatorname{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) &= \sigma^2 v_{12}\end{aligned}$$

where

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \{v_{ij}\}_{i=1,\dots,n;j=1,\dots,n}$$

§12 Lec 12: Oct 22, 2021

§12.1 Gauss-Markov Theorem in Multiple Regression

Let $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ be the least squares estimator of $\boldsymbol{\beta}$ and let $\mathbf{b} = \mathbf{M}^*\mathbf{Y}$ be an unbiased estimator of $\boldsymbol{\beta}$ (not the least squares). Let's define $\mathbf{M}^* = \mathbf{M} + \left(\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\right)$.
$\mathbf{b}$ is unbiased

$$E\mathbf{b} = \beta$$

because

$$E\mathbf{M}^*\mathbf{Y} = \boldsymbol{\beta}$$

or

$$\begin{aligned} E\left[\mathbf{M} + \left(\mathbf{X}'\mathbf{X}^{-1}\right)\mathbf{X}'\right]\mathbf{Y} &= \boldsymbol{\beta} \\ \left(\mathbf{M} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{X}\boldsymbol{\beta} &= \boldsymbol{\beta} \\ \mathbf{M}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta} &= \boldsymbol{\beta} \\ \mathbf{M}\mathbf{X} &= 0 \end{aligned}$$

Check var($\mathbf{b}$).

$$\text{var}(\mathbf{b}) = \text{var}\left(\mathbf{M}^*\mathbf{Y}\right) = \text{var}\left[\mathbf{M} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{Y}$$

Note:

$$\text{var}(\mathbf{A}\mathbf{Y}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$$

where $\text{var}(\mathbf{Y}) = \sigma^2\mathbf{I}$. Then,

$$\begin{aligned} \text{var}(\mathbf{b}) &= \sigma^2\left[\mathbf{M} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\mathbf{M}' + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\ &= \sigma^2\mathbf{M}\mathbf{M}' + \sigma^2\mathbf{M}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}' \\ &+ \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\mathbf{M}\mathbf{M}' + \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\mathbf{M}\mathbf{M}' + \text{var}(\hat{\beta}_1) \end{aligned}$$

A matrix $\mathbf{B}$ is positive definite if for a non zero vector $\mathbf{a}$

$$\mathbf{a}'\mathbf{B}\mathbf{a} > 0$$

Aside Note:

$$\text{var}\left(\mathbf{a}\mathbf{Y}'\right) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} > 0$$

Now, let $\mathbf{a}$ be a non zero vector

$$\begin{aligned} \mathbf{a}'\mathbf{M}\mathbf{M}'\mathbf{a} &= (\mathbf{M}'\mathbf{a})'\left(\mathbf{M}'\mathbf{a}\right) \\ &= \mathbf{q}'\mathbf{q} \\ &= \sum q_i^2 > 0 \end{aligned}$$

Therefore, $\mathbf{M}\mathbf{M}'$ is a positive definite matrix and thus $\text{var}(\mathbf{b}) \geq \text{var}\left(\hat{\boldsymbol{\beta}}\right)$.

§12.2 Gauss-Markov Theorem For a Linear Combination

We have

$$\begin{aligned}\text{var}\left(\mathbf{a}'\hat{\boldsymbol{\beta}}\right) &= \mathbf{a}'\,\text{var}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{a} \\ &= \sigma^2\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}\end{aligned}$$

or

$$\begin{aligned}var\left(a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2\right) &= a_0^2\,\text{var}(\hat{\beta}_0) + a_1^2\,\text{var}(\hat{\beta}_1) + a_2^2\,\text{var}(\hat{\beta}_2) + 2a_0a_1\,\text{cov}(\hat{\beta}_0,\hat{\beta}_1) \\ &\quad + 2a_0a_2\,\text{cov}(\hat{\beta}_0,\hat{\beta}_2) + 2a_1a_1\,\text{cov}(\hat{\beta}_1,\hat{\beta}_2)\end{aligned}$$

Let's compare it to $\text{var}(\mathbf{a}'\mathbf{b})$.

$$\begin{aligned}\text{var}\left(\mathbf{a}'\mathbf{b}\right) &= \mathbf{a}'\,\text{var}(\mathbf{b})\mathbf{a} \\ &= \sigma^2\mathbf{a}'\left[\mathbf{M}\mathbf{M}' + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]\mathbf{a} \\ &= \sigma^2\mathbf{a}'\mathbf{M}\mathbf{M}'\mathbf{a} + \sigma^2\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a} \\ &= \sigma^2\mathbf{a}'\mathbf{M}\mathbf{M}'\mathbf{a} + \text{var}\left(\mathbf{a}'\hat{\boldsymbol{\beta}}\right)\end{aligned}$$

Thus, $\text{var}(\mathbf{a}'\mathbf{b}) \geq \text{var}\left(\mathbf{a}'\hat{\boldsymbol{\beta}}\right)$.

<u>Special Case</u>:

$$\mathbf{a} = \begin{pmatrix}0\\0\\\vdots\\1\\\vdots\\0\\0\end{pmatrix}$$

$$\text{var}(b_i) \geq \text{var}(\hat{\beta}_i)$$

§12.3 Review of Multivariate Normal Distribution

Normality assumption: $\varepsilon_1,\ldots,\varepsilon_n \overset{\text{i.i.d}}{\sim} N(0,\delta)$

$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0},\sigma^2\mathbf{I}\right)$$

Let $\mathbf{Y} \sim N_n\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

Consider

$$\left.\begin{aligned}f(\boldsymbol{\varepsilon}) &= f(\varepsilon_1)\cdot f(\varepsilon_2)\ldots f(\varepsilon_n) \\ f(\varepsilon_i) &= \tfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\Sigma_i^2}\end{aligned}\right\} = \frac{1}{(2\pi)^{\frac{n}{2}}}\left|\sigma^2\mathbf{I}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{\varepsilon}(\sigma^2\mathbf{I})^{-1}\boldsymbol{\varepsilon}}$$

So

$$f(\boldsymbol{\varepsilon}) = (2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}} \implies \boldsymbol{\varepsilon} \sim N_n(\mathbf{0},\sigma^2\mathbf{I})$$

<u>Joint MGF</u>: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Then

$$M_{\mathbf{Y}}(\mathbf{t}) = Ee^{\mathbf{t}'\mathbf{Y}} = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$$

where $\mathbf{t} = \begin{pmatrix}t_1\\\vdots\\t_n\end{pmatrix}$.

Theorem 12.1

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\mathbf{A}$ be $m \times n$ matrix of constant and $\mathbf{c}$ $m \times 1$ vector of constants. Using the joint mgf

$$\mathbf{AY} \sim N_m\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}'\right)$$
$$\mathbf{AY} + \mathbf{c} \sim N_m\left(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A\Sigma A}'\right)$$

Notice that

$$\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\left.\begin{array}{l} \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}) \\ E\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} \\ \text{var}(\mathbf{Y}) = \sigma^2\mathbf{I} \end{array}\right\} \implies \mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

§13 Lec 13: Oct 25, 2021

§13.1 Theorems in Multivariate Normal Distribution

Consider: $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}} (\mathbf{y}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\boldsymbol{\mu})$$
$$M_{\mathbf{y}}(\mathbf{t}) = Ee^{\mathbf{t}'\mathbf{y}} = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$$

Proof. Let $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$. Then, the spectral decomposition of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}', \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$
$$\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{P}'$$

So,

$$\begin{aligned} M_{Z_i}(t_i) &= Ee^{t_i z_i} = e^{\frac{1}{2}t_i^2} \\ M_{\mathbf{Z}}(\mathbf{t}) &= Ee^{\mathbf{t}'\mathbf{z}} = Ee^{t_1 z_1 + \ldots + t_n z_n} \\ &= Ee^{t_1 z_1} \cdot Ee^{t_2 z_2} \ldots Ee^{t_n z_n} \\ &= e^{\frac{1}{2}\mathbf{t}'\mathbf{t}} \\ M_{\mathbf{Y}}(\mathbf{t}) &= M_{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}+\boldsymbol{\mu}}(\mathbf{t}) \\ &= Ee^{\mathbf{t}'\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}+\boldsymbol{\mu}\right)} \\ &= e^{\mathbf{t}'\boldsymbol{\mu}} Ee^{\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{t}\right)'\mathbf{Z}} \end{aligned}$$

Let $t^* = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{t}$. Then

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\mathbf{t}'\boldsymbol{\mu}} Ee^{\mathbf{t}^{*'}\mathbf{Z}} \\ &= e^{\mathbf{t}'\boldsymbol{\mu}} M_{\mathbf{Z}}(\mathbf{t}^*) = e^{\mathbf{t}'\boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{t}^{*'}TBA} \\ &= e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \end{aligned}$$

□

Theorem 13.1

Let $\mathbf{A}$ be $m \times n$ matrix of constants and $\mathbf{C}$ be $m \times 1$ vector of constants. Then

$$\mathbf{AY} + \mathbf{C} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{C}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$
$$\mathbf{AY} \sim N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

Proof. We have

$$\begin{aligned} M_{\mathbf{AY}+\mathbf{C}}(\mathbf{t}) &= Ee^{\mathbf{t}'(\mathbf{AY}+\mathbf{C})} \\ &= e^{\mathbf{t}'\mathbf{C}} \cdot Ee^{(\mathbf{A}'\mathbf{t})'\mathbf{Y}} \end{aligned}$$

Let $\mathbf{t}^* = \mathbf{A}'\mathbf{t}$. Then

$$\begin{aligned} M_{\mathbf{AY}+\mathbf{C}}(\mathbf{t}) &= e^{\mathbf{t}'\mathbf{C}} \cdot M_{\mathbf{Y}}(\mathbf{t}^*) \\ &= e^{\mathbf{t}'\mathbf{C}} \cdot e^{\mathbf{t}^{*\prime}\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^{*\prime}\boldsymbol{\Sigma}\mathbf{t}^*} \\ &= e^{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}+\mathbf{C})+\frac{1}{2}\mathbf{t}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{t}} \end{aligned}$$

Thus, $\mathbf{AY} + \mathbf{C} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{C}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. □

Theorem 13.2

Let $\mathbf{Y}\ N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

Note that

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \hline Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}$$

Then,

$$\begin{aligned} \mathbf{Q}_1 &\sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \\ \mathbf{Q}_2 &\sim N_{n-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) \end{aligned}$$

Proof. Use the above theorem

$$\mathbf{Q}_1 = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \end{pmatrix} \mathbf{Y} = \mathbf{AY}$$

Then,

$$\begin{aligned} E\mathbf{Q}_1 &= E\mathbf{AY} \\ &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \\ &= \boldsymbol{\mu}_1 \\ \operatorname{var}(\mathbf{Q}_1) &= \operatorname{var}(\mathbf{AY}) \\ &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ \mathbf{0}' \end{pmatrix} \\ &= \boldsymbol{\Sigma}_{11} \end{aligned}$$

□

If $\mathbf{A} = \mathbf{a}'$ (row vector), then $\mathbf{a}'\mathbf{Y} \sim N\left(\mathbf{a}'\boldsymbol{\mu}, \sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}\right)$.

Theorem 13.3

Independence for $\mathbf{Y} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix}$

$$\mathbf{Y} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$
$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

Then, the MGF is

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \\ &= e^{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2} \end{aligned}$$

If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, then

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\mathbf{t}_1'\boldsymbol{\mu}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1} e^{\mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2} \\ &= M_{\mathbf{Q}_1}(t_1) \cdot M_{\mathbf{Q}_2}(t_2) \end{aligned}$$

Thus, $\mathbf{Q}_1, \mathbf{Q}_2$ are independent $\iff \operatorname{cov}\left(\mathbf{Q}_1, \mathbf{Q}_2\right) = \mathbf{0}$.

For $\mathbf{AY}, \mathbf{BY}$, we have

$$\operatorname{cov}\left(\mathbf{AY}, \mathbf{BY}\right) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}'$$

Theorem 13.4

We have

$$\mathbf{Q}_1 \Big| \mathbf{Q}_2 \sim N_p\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{Q}_2 - \boldsymbol{\mu}_2\right), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

Back to multiple regression

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$
$$\mathbf{Y} \sim N\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

Then, the likelihood function is

$$L = f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}}\left(\sigma^2\mathbf{I}\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\left(\sigma^2\mathbf{I}\right)^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}$$

or

$$L = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2}\sigma^2(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}$$
$$\ln L = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right)$$

Thus for β,

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \mathbf{0} \implies \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

and estimation for σ^2

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{\mu}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right) = 0$$
$$\hat{\sigma}^2 = \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{n} = \frac{\mathbf{e}'\mathbf{e}}{n}$$

Now, $\mathbf{e} = (\mathbf{I} - \mathbf{H})\,\mathbf{Y} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$. Therefore,

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= \mathbf{Y}'\,(\mathbf{I} - \mathbf{H})\,\mathbf{Y} = \boldsymbol{\varepsilon}'\,(\mathbf{I} - \mathbf{H})\,\boldsymbol{\varepsilon} \\
\hat{\sigma}^2 &= \frac{\mathbf{e}'\mathbf{e}}{n} \\
&= \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}}{n} \\
&= \frac{\boldsymbol{\varepsilon}'\,(\mathbf{I} - \mathbf{H})\,\boldsymbol{\varepsilon}}{n}
\end{aligned}$$

§14 Lec 14: Oct 27, 2021

§14.1 Mean and Variance in Multivariate Normal Distribution

Consider

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)\\ \implies \mathbf{Y} &\sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)\end{aligned}$$

Joint pdf of $\mathbf{Y}$ is

$$f(\mathbf{y}) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}$$

Using the method of maximum we obtain the MLEs of $\boldsymbol{\beta}$ and σ^2

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

which is the same as the least squares estimator. And

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)}{n} = \frac{\mathbf{e}'\mathbf{e}}{n}$$

Note that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ or $\mathbf{e} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$. Therefore,

$$\mathbf{e}'\mathbf{e} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \quad \text{or} \quad \mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$$

So

$$\begin{aligned}E\hat{\sigma}^2 &= \frac{1}{n}E\mathbf{e}'\mathbf{e}\\ &= \frac{1}{n}E\left[\underbrace{\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}}_{\text{scalar}}\right]\\ &= \frac{1}{n}E\left[\operatorname{tr}(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right]\\ &= \frac{1}{n}\operatorname{tr}\left[E\left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right]\\ &= \frac{1}{n}\operatorname{tr}\left[\left(\mathbf{I} - \mathbf{H}\right)E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\right]\end{aligned}$$

Note:

$$\begin{aligned}\boldsymbol{\Sigma} &= E\left(\mathbf{Y} - \boldsymbol{\mu}\right)\left(\mathbf{Y} - \boldsymbol{\mu}\right)'\\ E\left[\mathbf{Y}\mathbf{Y}'\right] &= \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'\end{aligned}$$

where

$$\begin{aligned}E(\boldsymbol{\varepsilon}) &= \mathbf{0}\\ \operatorname{var}(\boldsymbol{\varepsilon}) &= \sigma^2\mathbf{I}\end{aligned}$$

Then,

$$\begin{aligned}E\hat{\sigma}^2 &= \frac{1}{n}\operatorname{tr}\left[\left(\mathbf{I} - \mathbf{H}\right)\left(\sigma^2\mathbf{I} + \mathbf{0}\mathbf{0}'\right)\right]\\ &= \frac{1}{n}\operatorname{tr}\left(\mathbf{I} - \mathbf{H}\right)\sigma^2\mathbf{I}\\ &= \frac{\sigma^2}{n}\operatorname{tr}\left(\mathbf{I} - \mathbf{H}\right)\end{aligned}$$

Let's compute $\operatorname{tr}(\mathbf{I}-\mathbf{H})$.

$$\begin{aligned}\operatorname{tr}(\mathbf{I}-\mathbf{H}) &= \operatorname{tr}(\mathbf{I}) - \operatorname{tr}(\mathbf{H}) \\ &= \operatorname{tr}(\mathbf{I}) - \operatorname{tr}\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] \\ &= \operatorname{tr}(\mathbf{I}) - \operatorname{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\right] \\ &= \operatorname{tr}(\mathbf{I}_n) - \operatorname{tr}(\mathbf{I}_{k+1}) \\ &= n-k-1\end{aligned}$$

So,

$$E\hat{\sigma}^2 = \sigma^2 \frac{n-k-1}{n}$$

which is biased. Therefore, the unbiased estimator of σ^2 is

$$S_e^2 = \hat{\sigma}^2 \frac{n}{n-k-1} = \frac{\mathbf{e}'\mathbf{e}}{n}\frac{n}{n-k-1} = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$$

In simple regression ($k=1$ – one predictor)

$$S_e^2 = \frac{\mathbf{e}'\mathbf{e}}{n-2} = \frac{\sum e_i^2}{n-2}$$

Now, let's find the mean and variance of $\hat{\mathbf{Y}}$ and $\mathbf{e}$.

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{H}\mathbf{Y} \\ E\hat{\mathbf{Y}} &= \mathbf{H}E\mathbf{Y} \\ &= \mathbf{H}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta}\end{aligned}$$

Note: $\mathbf{H}\mathbf{X} = \mathbf{X}$.

$$\begin{aligned}\operatorname{var}\left(\hat{\mathbf{Y}}\right) &= \operatorname{var}(\mathbf{H}\mathbf{Y}) \\ &= \sigma^2\mathbf{H}\end{aligned}$$

For $\mathbf{e}$,

$$\begin{aligned}E\mathbf{e} &= E\left[(\mathbf{I}-\mathbf{H})\mathbf{Y}\right] \\ &= E\left[\mathbf{Y}-\mathbf{H}\mathbf{Y}\right] \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{0} \\ \operatorname{var}(\mathbf{e}) &= \operatorname{var}\left[(\mathbf{I}-\mathbf{H})\mathbf{Y}\right] \\ &= \sigma^2(\mathbf{I}-\mathbf{H})\end{aligned}$$

§14.2 Independent Vectors in Multiple Regression

If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{AY}$ and $\mathbf{BY}$ are independent iff

$$\operatorname{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$$

Apply this result for multiple regression

$$\operatorname{cov}\left(\hat{\mathbf{Y}}, \mathbf{e}\right), \quad \operatorname{cov}\left(\hat{\boldsymbol{\beta}}, \mathbf{e}\right)$$

or use

$$\begin{pmatrix} \hat{\mathbf{Y}} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{HY} \\ (\mathbf{I}-\mathbf{H})\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{H} \\ \mathbf{I}-\mathbf{H} \end{pmatrix} \mathbf{Y} = \mathbf{AY}$$

$\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$.

$$\begin{aligned} \operatorname{var}(\mathbf{AY}) &= \mathbf{A}\operatorname{var}(\mathbf{Y})\mathbf{A}' \\ &= \sigma^2 \begin{pmatrix} \mathbf{H} \\ \mathbf{I}-\mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{H} & \mathbf{I}-\mathbf{H} \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}-\mathbf{H} \end{pmatrix} \end{aligned}$$

$\hat{\mathbf{Y}}$ and $\mathbf{e}$ are independent. Similarly, we can show that $\hat{\boldsymbol{\beta}}$ and $\mathbf{e}$ are independent.

§14.3 Partial Regression

Consider

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}$$

with the following three models

$$\begin{aligned} \mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} &\implies \hat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y} \\ \mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} &\implies \hat{\boldsymbol{\beta}}_2 = \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{Y} \end{aligned}$$

and

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{or} \quad \mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

§15 Lec 15: Oct 29, 2021

§15.1 Partial Regression (Cont'd)

Normal equation:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

using

$$\mathbf{X}' = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{12} \\ \hat{\boldsymbol{\beta}}_{21} \end{pmatrix}$$

Then,

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}$$

and

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{X}_1'\mathbf{Y} \\ \mathbf{X}_2'\mathbf{Y} \end{pmatrix}$$

and the normal equations are

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{12} \\ \hat{\boldsymbol{\beta}}_{21} \end{pmatrix} = \begin{pmatrix} \mathbf{X_1'Y} \\ \mathbf{X_2'Y} \end{pmatrix}$$

Then,

$$\mathbf{X_1'X_1}\hat{\boldsymbol{\beta}}_{12} + \mathbf{X_1'X_2}\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_1'Y} \tag{1}$$

$$\mathbf{X_2'X_1}\hat{\boldsymbol{\beta}}_{12} + \mathbf{X_2'X_2}\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'Y} \tag{2}$$

From (1),

$$\mathbf{X_1'X_1}\hat{\boldsymbol{\beta}}_{12} = \mathbf{X_1'Y} - \mathbf{X_1'X_2}\hat{\boldsymbol{\beta}}_{21}$$

So,

$$\hat{\boldsymbol{\beta}}_{12} = \underbrace{(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Y}}_{\hat{\boldsymbol{\beta}}_1} - (\mathbf{X_1'X_1})^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} \tag{3}$$

Let's find $\hat{\boldsymbol{\beta}}_{21}$ by substitute (3) into (2).

$$\mathbf{X_2'X_1}\left[(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Y} - (\mathbf{X_1'X_1})^{-1}\mathbf{X_1'X_2}\hat{\boldsymbol{\beta}}_{21}\right] + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'Y}$$

$$\mathbf{X_2'X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Y} - \mathbf{X_2'X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'X_2}\hat{\boldsymbol{\beta}}_{21} + (\mathbf{X_2'X_2})\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2{}'Y}$$

Then,

$$\left(\mathbf{X_2'X_2}\hat{\boldsymbol{\beta}}_{21}\right) - \mathbf{X_2'X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'X_2}\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'Y} - \mathbf{X_2'X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Y}$$

$$\mathbf{X_2'}\left[\mathbf{I} - \mathbf{X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'}\right]\mathbf{X_2}\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'}\left[\mathbf{I} - \mathbf{X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'}\right]\mathbf{Y}$$

$$\mathbf{X_2'}[\mathbf{I} - \mathbf{H_1}]\mathbf{X_2}\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'}[\mathbf{I} - \mathbf{H_1}]\mathbf{Y}$$

$$\mathbf{X_2'}(\mathbf{I} - \mathbf{H}_1)(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X_2'}(\mathbf{I} - \mathbf{H}_1)(\mathbf{I} - \mathbf{H}_1)\mathbf{Y}$$

$$[(\mathbf{I} - \mathbf{H})\mathbf{X_2}]'[(\mathbf{I} - \mathbf{H})\mathbf{X_2}]\hat{\boldsymbol{\beta}}_{21} = [(\mathbf{I} - \mathbf{H})\mathbf{X_2}]'[(\mathbf{I} - \mathbf{H_1})\mathbf{Y}]$$

Note:

$$(\mathbf{I} - \mathbf{H_1})\mathbf{Y} = \mathbf{Y}^*$$

which is residuals from regression of $\mathbf{Y}$ on $\mathbf{X_1}$. Suppose

$$\mathbf{X_2} = \begin{pmatrix} \mathbf{x_3} & \mathbf{x_4} & \mathbf{x_5} \end{pmatrix}$$

Here $k = 5$ and

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{x_1} & \mathbf{x_2} & | & \mathbf{x_3} & \mathbf{x_4} & \mathbf{x_5} \end{pmatrix}$$

where

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 \end{pmatrix}$$

Then,

$$\begin{aligned} (\mathbf{I} - \mathbf{H_1})\,\mathbf{X_2} &= (\mathbf{I} - \mathbf{H_1})\begin{pmatrix} \mathbf{x_3} & \mathbf{x_4} & \mathbf{x_5} \end{pmatrix} \\ &= \left[(\mathbf{I} - \mathbf{H}_1)\,\mathbf{x_3} \quad (\mathbf{I} - \mathbf{H_1})\,\mathbf{x_4} \quad (\mathbf{I} - \mathbf{H_1})\,\mathbf{x_5}\right] \\ &= \mathbf{X_2}^* \end{aligned}$$

So,

$$\left(\mathbf{X}_2^{*\prime}\mathbf{X}_2^*\right)\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2^{*\prime}\mathbf{Y}^*$$

and thus

$$\hat{\boldsymbol{\beta}}_{21} = \left(\mathbf{X}_2^{*\prime}\mathbf{X}_2^*\right)^{-1}\mathbf{X}_2^{*\prime}\mathbf{Y}^*$$

Special Case 1:

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} \mathbf{1} & \mathbf{X}_{(0)} \end{pmatrix} \\ \boldsymbol{\beta} &= \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_{(0)} \end{pmatrix} \end{aligned}$$

Now, let's use partial regression to find $\hat{\boldsymbol{\beta}}_{(0)}$.
Regression $\mathbf{Y}$ on $\mathbf{1}$: $\mathbf{Y} = \beta_0\mathbf{1} + \boldsymbol{\varepsilon}$ and

$$\mathbf{Y}^* = (\mathbf{I} - \mathbf{H}_1)\,\mathbf{Y} = \left[\mathbf{I} - \mathbf{1}\,(\mathbf{1}'\mathbf{1})^{-1}\,\mathbf{1}'\right]\mathbf{Y} = \left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right]\mathbf{Y} = \begin{bmatrix} y_1 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{bmatrix}$$

$\mathbf{X}^*_{(0)}$regress $\mathbf{X}_{(0)}$ on $\mathbf{1}$

$$\begin{aligned} \mathbf{X}^*_{(0)} &= (\mathbf{I} - \mathbf{H}_1)\,\mathbf{X}_{(0)} \\ &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)} \\ &= \left[\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{x_1}, \dots, \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)x_k\right] \\ &= \begin{bmatrix} x_{11} - \overline{x_1} & \dots & x_{1k} - \overline{x_k} \\ x_{21} - \overline{x_1} & \dots & x_{2k} - \overline{x_k} \\ \vdots & & \vdots \\ x_{n1} - \overline{x_1} & \dots & x_{nk} - \overline{x_k} \end{bmatrix} \end{aligned}$$

Finally, to estimate the vector of the slopes $\boldsymbol{\beta}_{(0)} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$

$$\text{We regress} \begin{bmatrix} y_1 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{bmatrix} \quad \text{on} \quad \begin{bmatrix} x_{11} - \overline{x_1} & \dots & x_{1k} - \overline{x_k} \\ x_{21} - \overline{x_1} & \dots & x_{2k} - \overline{x_k} \\ \vdots & & \vdots \\ x_{n1} - \overline{x_1} & \dots & x_{nk} - \overline{x_k} \end{bmatrix}$$

to get $\hat{\boldsymbol{\beta}}_{(0)} = \left(\mathbf{X}^{*'}_{(0)}\mathbf{X}^{*}_{(0)}\right)^{-1}\mathbf{X}^{*'}_{(0)}\mathbf{Y}^{*}$ where

$$\mathbf{X}^{*}_{(0)} = \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}$$
$$\mathbf{Y}^{*} = \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{Y}$$

§16 Lec 16: Nov 1, 2021

§16.1 Partial Regression (Cont'd)

Consider:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Then,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

The partial regression of $\mathbf{Y}^*$ on $\mathbf{X}_2^*$

$$\hat{\boldsymbol{\beta}}_{21} = \left(\mathbf{X}_2^{*\prime}\mathbf{X}_2^*\right)^{-1}\mathbf{X}_2^{*\prime}\mathbf{Y}^*$$

i.e., $\mathbf{Y}^* = \mathbf{X}_2^*\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$.
Special Case 2: Begin with

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with k predictors. Then, we add an extra predictor $\mathbf{Z}$. The new model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + c\mathbf{Z} + \boldsymbol{\varepsilon}$$

Use partial regression to estimate c.

1. Regress $\mathbf{Y}$ on $\mathbf{X} \to \mathbf{e}$ residuals
2. Regress $\mathbf{Z}$ on $\mathbf{X} \to \mathbf{Z}^*$ residuals.
3. Regress $\mathbf{e}$ on $\mathbf{Z}^*$ to get

$$\hat{c} = \left(\mathbf{Z}^{*\prime}\mathbf{Z}^*\right)^{-1}\mathbf{Z}^{*\prime}\mathbf{e}$$

or

$$\hat{c} = \frac{\mathbf{Z}^{*\prime}\mathbf{e}}{\mathbf{Z}^{*\prime}\mathbf{Z}^*} = \frac{\mathbf{e}'\mathbf{Z}^{*\prime}}{\mathbf{Z}^{*\prime}\mathbf{Z}^*}$$

Change in the error sum of squares when a new predictor is added in the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + c\mathbf{Z} + \boldsymbol{\varepsilon} \quad (2)$$

Residuals using (1)

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

Residuals using (2)

$$\mathbf{u} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\delta}} - \hat{c}\mathbf{Z}$$

Now, we need to find $\hat{\boldsymbol{\delta}}$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + c\mathbf{Z} + \boldsymbol{\varepsilon}$$
$$\mathbf{Y} = \begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta} \\ c\end{pmatrix} + \boldsymbol{\varepsilon}$$
$$\mathbf{Y} = \mathbf{w}\begin{pmatrix}\boldsymbol{\beta} \\ c\end{pmatrix} + \boldsymbol{\varepsilon}$$
$$\mathbf{Y} = \mathbf{w}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

Normal equations:

$$\mathbf{w}'\mathbf{w}\boldsymbol{\eta} = \mathbf{w}'\mathbf{Y}$$

or

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\delta}} + \mathbf{X}'\mathbf{Z}\hat{c} = \mathbf{X}'\mathbf{Y} \tag{1}$$
$$\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\delta}} + \mathbf{Z}'\mathbf{Z}\hat{c} = \mathbf{Z}'\mathbf{Y} \tag{2}$$

From (1)

$$\hat{\boldsymbol{\delta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Z}\hat{c}\right]$$

or

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\hat{c}$$

Now back to $\mathbf{u}$

$$\begin{aligned}
\mathbf{u} &= \mathbf{Y} - X\hat{\boldsymbol{\beta}} + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\hat{c} - \hat{c}\mathbf{Z} \\
\mathbf{u} &= \mathbf{e} - \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\mathbf{Z}\hat{c} \\
&= \mathbf{e} - \left[\mathbf{I} - \mathbf{H}\right]\mathbf{Z}\hat{c} \\
&= \mathbf{e} - \mathbf{Z}^*\hat{c}
\end{aligned}$$

The SSE is

$$\begin{aligned}
\text{SSE}_{\mathbf{XZ}} &= \mathbf{u}'\mathbf{u} \\
&= \left(\mathbf{e} - \mathbf{Z}^*\hat{c}\right)'\left(\mathbf{e} - \mathbf{Z}^*\hat{c}\right) \\
&= \mathbf{e}'\mathbf{e} - 2\mathbf{Z}^{*'}\mathbf{e}\hat{c} + \mathbf{Z}^{*'}\mathbf{Z}^*\hat{c}^2 \\
&= \mathbf{e}'\mathbf{e} - 2\hat{c}^2\mathbf{Z}^{*'}\mathbf{Z}^* + \mathbf{Z}^{*'}\mathbf{Z}^*\hat{c}^2 \\
&= \mathbf{e}'\mathbf{e} - \mathbf{Z}^{*'}\mathbf{Z}^*\hat{c}^2
\end{aligned}$$

Thus, we can conclude that adding a new predictor would never increase SSE, i.e., $\mathbf{u}'\mathbf{u} \le \mathbf{e}'\mathbf{e}$. Note that the new R^2 is

$$\begin{aligned}
R^2_{\mathbf{XZ}} &= 1 - \frac{\mathbf{u}'\mathbf{u}}{\text{SST}} \\
&= 1 - \frac{\mathbf{e}'\mathbf{e}}{\text{SST}} + \frac{\mathbf{Z}^{*'}\mathbf{Z}^*\hat{c}^2}{\text{SST}} \\
&= R^2_{\mathbf{X}} + \frac{\mathbf{Z}^{*'}\mathbf{Z}^*\hat{c}^2}{\text{SST}}
\end{aligned}$$

So, $R^2_{\mathbf{XZ}} \ge R^2_{\mathbf{X}}$.

§16.2 Partial Correlation

Consider

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

where

$$\begin{cases} Y_i: & \text{income} \\ X_{i1}: & \text{age} \\ X_{i2}: & \text{number of years of education} \end{cases}$$

- Regress $\mathbf{Y}$ on $\mathbf{X}_1 \to \mathbf{Y}^*$ residuals.
- Regress $\mathbf{X}_2$ on $\mathbf{X}_1 \to \mathbf{X}_2^*$ residuals.

$$
\begin{aligned}
r^2_{\mathbf{YX_2}|\mathbf{X_1}} &= \frac{\operatorname{cov}^2\left(\mathbf{Y}^*, \mathbf{X}_2^*\right)}{\operatorname{var}(\mathbf{X_2}^*)\operatorname{var}(\mathbf{Y}^*)} \\
&= \frac{\left[\sum\left(Y_1^* - \overline{Y}^*\right)\left(X_{i2}^* - \overline{X}_2^*/(n-1)\right)\right]^2}{\frac{\sum\left(X_{i2}^* - \mathbf{X_2}^*\right)^2}{n-1}\frac{\sum\left(Y_i^* - \overline{Y}^*\right)^2}{n-1}} \\
&= \frac{\left(\sum Y_i^* X_{i2}\right)^2}{\left(\sum X_2^{*2}\right)\left(\sum Y_i^{*2}\right)} \\
&= \frac{\left(\mathbf{Y}^{*\prime}\mathbf{X}_2^*\right)^2}{\left(\mathbf{X}_2^{*\prime}\mathbf{X}_2^*\right)\left(\mathbf{Y}^{*\prime}\mathbf{Y}^*\right)}
\end{aligned}
$$

Another method:

- Regress $\mathbf{Y}$ on $X_1, X_2, \ldots, X_{k-1} \to \mathbf{Y}^*$.
- Regress X_k on $X_1, X_2, \ldots, X_{k-1} \to \mathbf{X}_k^2$.

$$
r^2_{YX_k|X_1,\ldots,x_{k-1}} = \frac{\operatorname{SSE}\left(Y \text{ on } X_1, \ldots, X_{k-1}\right) - \operatorname{SSE}\left(Y \text{ on } X_1, \ldots, X_k\right)}{\operatorname{SSE}\left(Y \text{ on } X_1, \ldots, X_{k-1}\right)}
$$

§17 Lec 17: Nov 3, 2021

§17.1 Constrained Least Squares

Consider

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

We want to estimate $\boldsymbol{\beta}$ subject to a set of linear constraints of the form $\mathbf{c}\boldsymbol{\beta} = \boldsymbol{\gamma}$ where $\mathbf{C} : m \times k+1$, $\boldsymbol{\beta} : k+1 \times 1$ and $\boldsymbol{\gamma} : m \times 1$.
Suppose $k = 4$

$$\begin{cases} \beta_0 + 2\beta_1 - 3\beta_2 + 5\beta_3 - \beta_4 = 5 \\ 2\beta_0 - \beta_1 + \beta_2 + 3\beta_3 = 10 \end{cases}$$

or

$$\begin{pmatrix} 1 & 2 & -3 & 5 & -1 \\ 2 & -1 & 1 & 3 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$

We still minimize $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ but now subject to $\mathbf{c}\boldsymbol{\beta} = \boldsymbol{\gamma}$.
Method of Lagrange Multipliers:

$$\min Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}' (\mathbf{c}\boldsymbol{\beta} - \boldsymbol{\gamma})$$

So,

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + 2\mathbf{c}\boldsymbol{\lambda} = \mathbf{0}$$

Solve for $\boldsymbol{\beta}$ to get $\hat{\boldsymbol{\beta}}\mathbf{c}$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_c &= (\mathbf{X}'\mathbf{X})^{-1} [\mathbf{X}'\mathbf{Y} - \mathbf{c}'\boldsymbol{\lambda}] \\ \hat{\boldsymbol{\beta}}_c &= \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\boldsymbol{\lambda} \end{aligned}$$

Now, we need to find $\boldsymbol{\lambda}$. So

$$\begin{aligned} \mathbf{c}\hat{\boldsymbol{\beta}}_c &= \mathbf{c}\hat{\boldsymbol{\beta}} - \mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\boldsymbol{\lambda} \\ \boldsymbol{\gamma} &= \mathbf{c}\hat{\boldsymbol{\beta}} - \mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\boldsymbol{\lambda} \\ \boldsymbol{\lambda} &= \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right) \end{aligned}$$

Therefore,

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

Fitted values:

$$\hat{\mathbf{Y}}_c = \mathbf{X}\hat{\boldsymbol{\beta}}_c = \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

Residuals:

$$\mathbf{e}_c = \mathbf{Y} - \hat{\mathbf{Y}}_c = \mathbf{e} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

Error sum of squares:

$$\begin{aligned} \text{SSE}_c &= \mathbf{e}_c^{'}\mathbf{e}_c \\ &= \left[\mathbf{e} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)\right]' \left[\mathbf{e} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)\right] \\ &= \mathbf{e}'\mathbf{e} + \mathbf{e}'\mathbf{X}[\ldots] + [\ldots]\mathbf{X}'\mathbf{e} \\ &\quad + \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}' \left[\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}'\right]^{-1} \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right) \end{aligned}$$

Finally,

$$\mathbf{e}_c^{'}\mathbf{e}_c = \mathbf{e}'\mathbf{e} + \left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{c}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{c}'\right]^{-1}\left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

We can deduce that $\text{SSE}_c \geq \text{SSE}$.
MLE of σ^2

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n}$$

For the constrained model

$$\hat{\sigma}_c^2 = \frac{\mathbf{e}_c^{'}\mathbf{e}_c}{n}$$

and

$$E\hat{\sigma}_c^2 = \frac{(n-k-1+m)\sigma^2}{n}$$

Method Using the Canonical Form of the Model:

$$\mathbf{c}\boldsymbol{\beta} = \boldsymbol{\gamma}$$

$$\mathbf{c} = \begin{pmatrix}\mathbf{c}_1 & \mathbf{c}_2\end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2\end{pmatrix}$$

$$\mathbf{c}_1\boldsymbol{\beta}_1 + \mathbf{c}_2\boldsymbol{\beta}_2 = \boldsymbol{\gamma}$$

$$\begin{pmatrix}1 & 2 \\ 2 & -1\end{pmatrix}\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} + \begin{pmatrix}-3 & 5 & -1 \\ 1 & 3 & 0\end{pmatrix}\begin{pmatrix}\beta_2 \\ \beta_3 \\ \beta_4\end{pmatrix} = \begin{pmatrix}5 \\ 10\end{pmatrix}$$

Back to the model using the same partition we get

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

Then,

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}_1\mathbf{c}_1^{-1}\left[\boldsymbol{\gamma} - \mathbf{c}_2\boldsymbol{\beta}_2\right] + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\ \mathbf{Y} - \mathbf{X}_1\mathbf{c}_1^{-1}\boldsymbol{\gamma} &= \left(\mathbf{X}_2 - \mathbf{X}_1\mathbf{c}_1^{-1}\mathbf{c}_2\right)\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\ \mathbf{Y}_r &= \mathbf{X}_{2r}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}\end{aligned}$$

which is the same form as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Thus,

$$\hat{\boldsymbol{\beta}}_{2c} = \left(\mathbf{X}_{2r}^{'}\mathbf{X}_{2r}\right)^{-1}\mathbf{X}_{2r}^{'}\mathbf{Y}_r$$

and therefore,

$$\hat{\beta}_{1c} = \mathbf{c}_1^{-1}\left(\boldsymbol{\gamma} - \mathbf{c}_2\hat{\boldsymbol{\beta}}_{2c}\right)$$

Overall,

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{c}'\left[\mathbf{c}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{c}'\right]^{-1}\left(\mathbf{c}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

or

$$\hat{\boldsymbol{\beta}}_c = \begin{bmatrix}\hat{\boldsymbol{\beta}}_{1c} \\ \hat{\boldsymbol{\beta}}_{2c}\end{bmatrix}$$

which is from canonical form. Next, let's find the mean and variance of $\hat{\boldsymbol{\beta}}_c$.

$$E\hat{\boldsymbol{\beta}}_c = \boldsymbol{\beta}$$

Notice that

$$\hat{\boldsymbol{\beta}}_c = \left[\mathbf{I} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{c}'\left[\mathbf{c}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{c}'\right]^{-1}\mathbf{c}\right]\hat{\boldsymbol{\beta}} + \text{const} = \mathbf{A}\hat{\boldsymbol{\beta}}$$

So

$$\text{var}(\hat{\boldsymbol{\beta}}_c) = \sigma^2 \mathbf{A} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{A}'$$

or using the canonical model

$$\text{var}(\hat{\boldsymbol{\beta}}_c) = \begin{pmatrix} \text{var}(\hat{\beta}_{1c}) & \text{cov}\left(\hat{\boldsymbol{\beta}}_{1c}, \hat{\boldsymbol{\beta}}_{2c}\right) \\ \text{cov}\left(\hat{\boldsymbol{\beta}}_{1c}, \hat{\boldsymbol{\beta}}_{2c}\right) & \text{var}(\hat{\boldsymbol{\beta}}_{2c}) \end{pmatrix}$$

§18 Lec 18: Nov 5, 2021

Consider:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$$
$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$$

Then, $\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right)$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}$$
$$\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$
$$\hat{\boldsymbol{\beta}}_1 \sim N\left(\beta_1, \sigma\sqrt{v_{11}}\right)$$
$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \begin{pmatrix} v_{00} & v_{01} & \ldots & v_{0k} \\ v_{10} & v_{11} & \ldots & v_{1k} \\ \vdots & & \ddots & \vdots \\ v_{k1} & v_{k2} & \ldots & v_{kk} \end{pmatrix}$$

§18.1 Quadratic Forms of Normally Distributed Random Variables

We have

a) $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$

$$Z_1, \ldots, Z_n \overset{\text{i.i.d}}{\sim} N(0, 1)$$
$$Z_i^2 \sim \mathcal{X}_1^2$$
$$\sum Z_i^2 \sim \mathcal{X}_n^2$$
$$\mathbf{Z}'\mathbf{Z} \sim \mathcal{X}_n^2$$

b) $\mathbf{Z} \sim N_n\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$. Then,

$$Z_i \sim N(0, \sigma)$$
$$\frac{Z_i}{\sigma} \sim N(0, 1)$$
$$\frac{Z_i^2}{\sigma^2} \sim \mathcal{X}_1^2$$
$$\frac{\sum Z_i^2}{\sigma^2} \sim \mathcal{X}_n^2$$
$$\frac{\mathbf{Z}}{\sigma} \sim N_n(\mathbf{0}, \mathbf{I})$$
$$\frac{\mathbf{Z}'\mathbf{Z}}{\sigma^2} \sim \mathcal{X}_n^2$$

In multiple regression,

$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$$
$$\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma^2} \sim \mathcal{X}_n^2$$

c) $\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}\right)$

$$Y_i \sim N\left(\mu_i, \sigma\right) \implies \frac{Y_i - \mu_i}{\sigma} \sim N(0,1)$$
$$\left(\frac{Y_i - \mu_i}{\sigma}\right)^2 \sim \mathcal{X}_1^2$$
$$\sum \left(\frac{Y_i - \mu_i}{\sigma}\right)^2 \sim \mathcal{X}_n^2$$
$$\frac{\left(\mathbf{Y} - \boldsymbol{\mu}\right)' \left(\mathbf{Y} - \boldsymbol{\mu}\right)}{\sigma^2} \sim \mathcal{X}_n^2$$

In multiple regression

$$\mathbf{Y} \sim N_n\left(X\boldsymbol{\beta}, \sigma^2 \mathbf{I}\right)$$
$$\frac{\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)' \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)}{\sigma^2} \sim \mathcal{X}_n^2$$

d) $\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, use $\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{Y} - \boldsymbol{\mu}\right)$. $\boldsymbol{\Sigma}$ is symmetric matrix

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$$
$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

where $|\boldsymbol{\Sigma} - \lambda \mathbf{I}| = 0$. If $\mathbf{x}$ is a new zero vector such that $\boldsymbol{\Sigma}\mathbf{x} = \lambda \mathbf{x}$, we say that $\mathbf{x}$ is an eigenvector of $\boldsymbol{\Sigma}$. Normalize the eigenvectors so that they have length 1

$$(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_n, \mathbf{e}_n)$$
$$\mathbf{e}_i'\mathbf{e}_j = 0, \quad \mathbf{e}_i'\mathbf{e}_i = 1$$
$$\mathbf{P} = \begin{pmatrix} \mathbf{e}_1 & \ldots & \mathbf{e}_n \end{pmatrix}$$
$$\mathbf{P}\mathbf{P}' = \mathbf{I}$$
$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \ldots + \lambda_n \mathbf{e}_n \mathbf{e}_n'$$

Result:

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} = \mathbf{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{P}'$$

Properties:

$$\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)' = \boldsymbol{\Sigma}^{-\frac{1}{2}}$$
$$\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}} = \boldsymbol{\Sigma}^{-1}$$
$$\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{P}'$$
$$\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{P}'$$
$$\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)' = \boldsymbol{\Sigma}^{\frac{1}{2}}$$
$$\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{\Sigma}$$

Back to the transformation

$$\begin{aligned}\mathbf{V} &= \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\mu}) \\ E\mathbf{V} &= \boldsymbol{\Sigma}^{-\frac{1}{2}}E(\mathbf{Y}-\boldsymbol{\mu}) = \mathbf{0} \\ \text{var}(\mathbf{V}) &= \text{var}\left[\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\mu})\right] \\ &= \text{var}\left[\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}-\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\right] \\ &= \text{var}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}\right) \\ &= \boldsymbol{\Sigma}^{-\frac{1}{2}}\text{var}(\mathbf{Y})\boldsymbol{\Sigma}^{-\frac{1}{2}} \\ &= \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}} \\ &= \mathbf{I}\end{aligned}$$

So, $\mathbf{V} \sim N_n(\mathbf{0},\mathbf{I})$. Then $\mathbf{V}'\mathbf{V} \sim \mathcal{X}_n^2$ and because $\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\mu})$, it follows that

$$\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)(\mathbf{Y}-\boldsymbol{\mu})'\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)(\mathbf{Y}-\boldsymbol{\mu}) = (\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\mu}) \sim \mathcal{X}_n^2$$

Therefore,

$$(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu}) \sim \mathcal{X}_n^2$$

In multiple regression

$$\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

We want to create a X^2 random variable using the distribution of $\hat{\boldsymbol{\beta}}$. Let $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$.

$$\begin{aligned}E\mathbf{V} &= (\mathbf{X}'\mathbf{X})^{\frac{1}{2}}E\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) = \mathbf{0} \\ \text{var}(\mathbf{V}) &= \text{var}\left[(\mathbf{X}'\mathbf{X})^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right] \\ &= (\mathbf{X}'\mathbf{X})^{\frac{1}{2}}\text{var}(\hat{\boldsymbol{\beta}})(\mathbf{X}'\mathbf{X})^{\frac{1}{2}} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{\frac{1}{2}}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})^{\frac{1}{2}} \\ &= \sigma^2\mathbf{I}\end{aligned}$$

We have so far

$$\begin{gathered}\mathbf{V} \sim N_{k+1}(\mathbf{0},\sigma^2\mathbf{I}) \\ \frac{\mathbf{V}'\mathbf{V}}{\sigma^2} \sim \mathcal{X}_{k+1}^2 \\ \frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma} \sim \mathcal{X}_{k+1}^2\end{gathered}$$

Summary:

$$\begin{cases}\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2} \sim \mathcal{X}_n^2 \\ \frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^2} \sim \mathcal{X}_{k+1}^2\end{cases}$$

Problem 18.1. Show that $\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2$

Proof. Have

$$\frac{\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\pm\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\pm\mathbf{X}\hat{\boldsymbol{\beta}}\right)}{\sigma^2} \sim \mathcal{X}_n^2$$

Rearrange and expand

$$\begin{aligned}\frac{\left(\mathbf{e}+\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right)'\left(\mathbf{e}+\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right)}{\sigma^2} &= \frac{\mathbf{e}'\mathbf{e}}{\sigma^2}+\frac{\mathbf{e}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma^2}+\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{e}}{\sigma^2}\\ &\quad+\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma^2}\\ &= \frac{\mathbf{e}'\mathbf{e}}{\sigma^2}+\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma^2}\end{aligned}$$

Note: $S_e^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k-1} \implies \mathbf{e}'\mathbf{e} = (n-k-1)S_e^2$

$$\underbrace{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})/\sigma^2}_{\sim\mathcal{X}_n^2} = \frac{(n-k-1)S_e^2}{\sigma^2}+\underbrace{\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma^2}}_{\sim\mathcal{X}_{k+1}^2}$$

We know $\text{cov}(\hat{\boldsymbol{\beta}},\mathbf{e}) = \mathbf{0}$.

$$\begin{aligned}Q &= Q_1+Q_2\\ M_Q(t) &= M_{Q_1}(t)\cdot M_{Q_2}(t)\\ M_{Q_1}(t) &= \frac{M_Q(t)}{M_{Q_2}(t)}\\ &= \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{k+1}{2}}}\\ &= (1-2t)^{-\frac{n-k-1}{2}}\end{aligned}$$

So, $Q_1 = \frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2$. □

In simple regression, $k=1$,

$$\begin{aligned}\frac{(n-2)S_e^2}{\sigma^2} &\sim \mathcal{X}_{n-2}^2\\ S_e^2 &= \frac{\sigma^2}{n-k-1}Q_1\end{aligned}$$

So,

$$\begin{aligned}M_{S_e^2}(t) &= M_{\frac{\sigma^2}{n-k-1}Q_1}(t)\\ &= M_{Q_1}\left(\frac{\sigma^2 t}{n-k-1}\right)\\ &= \left(1-\frac{2\sigma^2 t}{n-k-1}\right)^{-\frac{n-k-1}{2}}\end{aligned}$$

Thus, $S_e^2 \sim \Gamma\left(\frac{n-k-1}{2}, \frac{2\sigma^2}{n-k-1}\right)$

$$\begin{aligned}ES_e^2 &= \sigma^2\\ \text{var}(S_e^2) &= \frac{2\sigma^4}{n-k-1}\end{aligned}$$

§19 Lec 19: Nov 8, 2021

§19.1 Quadratic Forms and Their Distribution – Overview

1. $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$
$$\mathbf{Z}'\mathbf{Z} \sim \mathcal{X}_n^2$$
2. $\mathbf{Z} \sim N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$
$$\frac{\mathbf{Z}'\mathbf{Z}}{\sigma^2} \sim \mathcal{X}_n^2$$
and
$$\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma^2} \sim \mathcal{X}_n^2$$
3. $\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}\right)$
$$\frac{(\mathbf{Y}-\boldsymbol{\mu})'(\mathbf{Y}-\boldsymbol{\mu})}{\sigma^2} \sim \mathcal{X}_n^2$$
or
$$\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2} \sim \mathcal{X}_n^2$$
4. $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. From the spectral decomposition,
$$\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\mu})$$
Then,
$$\mathbf{V} \sim N_n(\mathbf{0}, \mathbf{I})$$
From 1), $\mathbf{V}'\mathbf{V} \sim \mathcal{X}_n^2$ or
$$(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu}) \sim \mathcal{X}_n^2$$
$\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\right)$
$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$$
$$\mathbf{V} \sim N_{k+1}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$$
From 2),
$$\frac{\mathbf{V}'\mathbf{V}}{\sigma^2} \sim \mathcal{X}_{k+1}^2$$
Finally,
$$\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)}{\sigma^2} \sim \mathcal{X}_{k+1}^2$$

Also, recall that we showed in last lecture
$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2$$

§19.2 Another Proof of Quadratic Forms and Their Distribution

1. Let $\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{I})$ and $\mathbf{Z} = \mathbf{P}'\mathbf{Y}$ where $\mathbf{P}$ is an orthogonal matrix where $\mathbf{P}'\mathbf{P} = \mathbf{I}$. Then, $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$.

2. Let $\mathbf{A}$ be a symmetric and idempotent matrix. Then the eigenvalues are 0 or 1.

   Proof. Have $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Multiply both sides by $\mathbf{x}'$

   $$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \lambda\mathbf{x}'\mathbf{x} \\ \mathbf{x}'\mathbf{A}\mathbf{A}\mathbf{x} &= \lambda\mathbf{x}'\mathbf{x} \\ (\mathbf{A}\mathbf{x})'(\mathbf{A}\mathbf{x}) &= \lambda\mathbf{x}'\mathbf{x} \\ \lambda^2\mathbf{x}'\mathbf{x} &= \lambda\mathbf{x}'\mathbf{x} \end{aligned}$$

   Therefore, $\lambda = 0$ or $\lambda = 1$.

   Question 19.1. How many 1's?

   Using the trace of $\mathbf{A}$,

   $$\begin{aligned} \operatorname{tr}\mathbf{A} &= \operatorname{tr}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}') \\ &= \operatorname{tr}(\boldsymbol{\Lambda}\mathbf{P}\mathbf{P}') \\ &= \operatorname{tr}\boldsymbol{\Lambda} \end{aligned}$$

   □

3. Let $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{I})$ and suppose $\mathbf{A}$ is a symmetric and idempotent matrix. Then $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \mathcal{X}_1^2$ where $r = \operatorname{tr}(\mathbf{A})$ (number of eigenvalues equal to 1).

   $$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' \implies \mathbf{Y}'\mathbf{A}\mathbf{Y} = \mathbf{Y}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{Y} = \mathbf{Z}'\boldsymbol{\Lambda}\mathbf{Z} \quad \text{from 1)}$$

   Then,

   $$\mathbf{Y}'\mathbf{A}\mathbf{Y} = z_1^2 + z_2^2 + \ldots + z_r^2 \sim \mathcal{X}_r^2$$

   where $\mathbf{Z} \sim N_n(\mathbf{O}, \mathbf{I}) \implies z_i \sim N(0,1)$, and so $z_i^2 \sim \mathcal{X}_1^2$

4. Use the previous theorem (3.) to show that $\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2$

   $$S_e^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k-1} \implies \mathbf{e}'\mathbf{e} = (n-k-1)S_e^2$$

   WTS: $\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2$

   Proof. Have

   $$\left.\begin{aligned} \mathbf{e} &= (\mathbf{I} - \mathbf{H})\mathbf{Y} \\ \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \end{aligned}\right\} \implies \mathbf{e} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$$

   Therefore,

   $$\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}}{\sigma}(\mathbf{I} - \mathbf{H})\frac{\boldsymbol{\varepsilon}}{\sigma} = \boldsymbol{\varepsilon}^{*'}(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}^*$$

   where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I})$. Using the theorem above (3.), we conclude that

   $$\boldsymbol{\varepsilon}^{*'}(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}^* = \frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{\operatorname{tr}(\mathbf{I}-\mathbf{H})} = \mathcal{X}_{n-k-1}^2$$

   □

§19.3 Efficiency of Least Squares Estimators

Let $\hat{\theta}$ be an unbiased estimator of θ. Then,

$$\operatorname{var}\left(\hat{\theta}\right) \geq \frac{1}{nI(\theta)}$$

This is known as the Cramer-Rao Lower Bound. Recall the score function

$$S = \frac{\partial \ln f(x;\theta)}{\partial \theta}$$

and the information matrix

$$I(\theta) = E\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2 = -\frac{E\partial^2 f(x;\theta)}{\partial \theta^2} = \operatorname{var}(S)$$

and $nI(\theta)$ is the information in the sample. An estimator is efficient if

- It is unbiased
- its variance is equal to the Cramer-Rao lower bound.

Also,

$$I(\theta) = -E\frac{\partial^2 \ln L}{\partial \theta^2}$$

for $Y_1, \ldots, Y_n$ i.i.d

§20 Lec 20: Nov 10, 2021

§20.1 Information Matrix and Efficient Estimator

Let $Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Is $\overline{y}$ an efficient estimator for μ where

$$E\overline{y} = \mu$$
$$\text{var}(\overline{y}) = \frac{\sigma^2}{n}$$

Consider the pdf

$$\begin{aligned}
f(y_i) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
L &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2} \\
\ln L &= -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i-\mu)^2 \\
\frac{\partial \ln L}{\partial \mu} &= \frac{2}{2\sigma^2}\sum(y_i-\mu) = \frac{1}{\sigma^2}\left(\sum y_i - n\mu\right) \\
\frac{\partial^2 \ln L}{\partial \mu^2} &= -\frac{\mu}{\sigma^2}
\end{aligned}$$

Cramer-Rao Lower Bound:

$$\frac{1}{-E\frac{\partial^2 \ln L}{\partial \mu^2}} = \frac{1}{-\left(-\frac{n}{\sigma^2}\right)} = \frac{\sigma^2}{n}$$

Thus, $\overline{y}$ is an efficient estimator for μ.
Let $\hat{\boldsymbol{\theta}}$ be the estimator of $\boldsymbol{\theta}$.

1. $E\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$
2. Find $\text{var}(\hat{\boldsymbol{\theta}})$ and compared it with the inverse of the information matrix $\mathbf{I}^{-1}(\boldsymbol{\theta})$ where

$$\mathbf{I}(\boldsymbol{\theta}) = -E\begin{pmatrix}
\frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_p} \\
\frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_p} \\
\vdots & \vdots & & \vdots \\
\frac{\partial^2 \ln L}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_p^2}
\end{pmatrix}$$

In multiple regression: $\beta_0, \beta_1, \ldots, \beta_k, \sigma^2$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$
$$\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$
$$\implies \ln L = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)$$
$$\ln L = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left(\mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)$$

Then

$$\mathbf{I}(\boldsymbol{\theta}) = -E\begin{pmatrix}
\frac{\partial^2 \ln L}{\partial \beta_0^2} & \frac{\partial^2 \ln L}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 \ln L}{\partial \beta_0 \partial \beta_k} & \frac{\partial^2 \ln L}{\partial \beta_0 \partial \sigma^2} \\
\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \ln L}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_k} & \frac{\partial^2 \ln L}{\partial \beta_1 \partial \sigma^2} \\
\vdots & \vdots & & \vdots & \vdots \\
\frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_0} & \frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_1} & \cdots & \frac{\partial^2 \ln L}{\partial \beta_k^2} & \frac{\partial^2 \ln L}{\partial \beta_k \partial \sigma^2} \\
\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta_0} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta_1} & \cdots & \frac{\partial^2 \ln L}{\partial \beta_k^2} & \frac{\partial^2 \ln L}{\partial (\sigma^2)^{(2)}}
\end{pmatrix} = -E\begin{pmatrix}
\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial \ln L}{\partial \boldsymbol{\beta} \partial \sigma^2} \\
\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \boldsymbol{\beta}'} & \frac{\partial^2 \ln L}{\partial (\sigma^2)^{(2)}}
\end{pmatrix}$$

Then,

$$\begin{aligned}
\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2}\left(-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) \\
\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= -\frac{1}{2\sigma^2}\left(2\mathbf{X}'\mathbf{X}\right) = -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \\
\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\partial \sigma^2} &= -\frac{1}{2\sigma^4}\left(-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) \\
\frac{\partial^2 \ln L}{\partial (\sigma^2)^{(2)}} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) \\
E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^{(2)}}\right] &= \frac{n}{2\sigma^4} - \frac{n}{\sigma^4} = -\frac{n}{2\sigma^4}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{I}(\boldsymbol{\theta}) &= \begin{pmatrix} -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & \mathbf{0} \\ \mathbf{0}' & -\frac{n}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & \mathbf{0} \\ \mathbf{0}' & \frac{n}{2\sigma^4} \end{pmatrix} \\
\mathbf{I}^{-1}(\boldsymbol{\theta}) &= \begin{pmatrix} \sigma^2(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{n} \end{pmatrix}
\end{aligned}$$

Notice that $E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. So $\hat{\boldsymbol{\beta}}$ is an efficient estimator of $\boldsymbol{\beta}$.

$$\begin{aligned}
S_e^2 &= \frac{\mathbf{e}'\mathbf{e}}{n-k-1}, \quad ES_e^2 = \sigma^2 \\
\text{var}(S_e^2) &= \frac{2\sigma^4}{n-k-1}
\end{aligned}$$

§20.2 Centered Model

Consider $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\begin{aligned}
\mathbf{X} &= \begin{pmatrix} \mathbf{1} & \mathbf{X}_{(0)} \end{pmatrix} \\
\boldsymbol{\beta} &= \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_{(0)} \end{pmatrix}
\end{aligned}$$

Then,

$$\mathbf{Y} = \beta_0\mathbf{1} + \mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon} \pm \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)}$$

Rearrange this expression and we obtain

$$\begin{aligned}
\mathbf{Y} &= \beta_0\mathbf{1} + \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} + \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon} \\
&= \mathbf{1}\left(\beta_0 + \frac{1}{n}\mathbf{1}'\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)}\right) + \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon} \\
&= \gamma_0\mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon}
\end{aligned}$$

Estimate the centered model

$$\begin{pmatrix} \hat{\gamma}_0 \\ \hat{\boldsymbol{\beta}}_{(0)} \end{pmatrix} = \begin{pmatrix} n & \mathbf{1}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} n & 0 \\ 0 & \mathbf{Z}'\mathbf{Z} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{pmatrix}$$

Thus,

$$\begin{aligned}\hat{\gamma}_0 &= \overline{y} \\ \hat{\boldsymbol{\beta}}_{(0)} &= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y} \\ &= \left(\mathbf{X}'_{(0)}\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}\right)^{-1}\mathbf{X}'_{(0)}\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{Y} \\ &= \left(\mathbf{X}^{*'}_{(0)}\mathbf{X}^{*}_{(0)}\right)'\mathbf{X}^{*'}_{(0)}\mathbf{Y}^{*}\end{aligned}$$

Observe that $\mathbf{Y} \sim N_n\left(\gamma_0\mathbf{1}+\mathbf{Z}\boldsymbol{\beta}_{(0)}, \sigma^2\mathbf{I}\right)$. Then,

$$\frac{\left(\mathbf{Y}-\gamma_0\mathbf{1}-\mathbf{Z}\boldsymbol{\beta}_{(0)}\right)'\left(\mathbf{Y}-\gamma_0\mathbf{1}-\mathbf{Z}\boldsymbol{\beta}_{(0)}\right)}{\sigma^2} \sim \mathcal{X}_n^2$$

- Fitted values: $\hat{\mathbf{Y}} = \mathbf{1}\hat{\gamma}_0 + \mathbf{Z}\hat{\boldsymbol{\beta}}_{(0)}$
- $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{1}\hat{\gamma}_0 - \mathbf{Z}\hat{\boldsymbol{\beta}}_{(0)}$

Note: Fitted values and residuals are the same for both models.

§21 Lec 21: Nov 12, 2021

§21.1 Confidence Intervals in Multiple Regression

Consider

$$\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

Let's find a $1-\alpha$ confidence interval for β_1.

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma\sqrt{v_{11}}\right)$$

$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1}$$

$$\frac{\frac{\hat{\beta}_1-\beta_1}{\sigma\sqrt{v_{11}}}}{\sqrt{\frac{(n-k-1)S_e^2}{\sigma^2}/n-k-1}} = \frac{\hat{\beta}_1-\beta_1}{S_e\sqrt{v_{11}}} \sim t_{n-k-1}$$

$$P\left(-t_{\frac{\alpha}{2};n-k-1} = \frac{\hat{\beta}_1-\beta_1}{S_e\sqrt{v_{11}}} \le t_{\frac{\alpha}{2};n-k-1}\right) = 1-\alpha$$

Finally,

$$\beta_1 \in \hat{\beta}_1 \pm t_{\frac{\alpha}{2};n-k-1} \cdot S_e\sqrt{v_{11}}$$

In general, to construct a confidence interval for $\mathbf{a}'\boldsymbol{\beta}$

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \sim N\left(\mathbf{a}'\boldsymbol{\beta}, \sigma\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}\right)$$

Then,

$$\frac{\frac{\mathbf{a}'\hat{\boldsymbol{\beta}}-\mathbf{a}'\boldsymbol{\beta}}{\sigma\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}}{\sqrt{\frac{(n-k-1)S_e^2}{\sigma^2}/(n-k-1)}} \sim t_{n-k-1}$$

$$\frac{\mathbf{a}'\hat{\boldsymbol{\beta}}-\mathbf{a}'\boldsymbol{\beta}}{S_e\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}} \sim t_{n-k-1}$$

Finally,

$$\mathbf{a}'\boldsymbol{\beta} \in \mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2};n-k-1} \cdot S_e\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}$$

If $\mathbf{a} = \begin{pmatrix}0 & 1 & 0 & 0 & \dots & 0\end{pmatrix}$ then $\mathbf{a}'\boldsymbol{\beta} = \beta_1$.
Prediction Interval for Y_0: For a given $\mathbf{X}_0' = \begin{pmatrix}1 & x_{01} & x_{02} & \dots x_{0k}\end{pmatrix}$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{pmatrix}Y_1\\ \vdots \\ Y_n\end{pmatrix} = \begin{pmatrix}1 & x_{01} & x_{02} & \dots & x_{0k}\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_n\end{pmatrix} + \begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix}$$

So the predictor is

$$\hat{Y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$$

Error of the prediction is $Y_0 - \hat{Y}_0$ with $E\left(Y_0 - \hat{Y}_0\right) = EY_0 - E\hat{Y}_0 = \mathbf{X}_0'\boldsymbol{\beta} - \mathbf{X}_0'\boldsymbol{\beta} = 0$. Note that $Y_0 = \hat{X}_0'\boldsymbol{\beta} + \varepsilon_0$

$$\begin{aligned}\operatorname{var}\left(Y_0 - \hat{Y}_0\right) &= \operatorname{var}(Y_0) + \operatorname{var}(\hat{Y}_0)\\ &= \sigma^2 + \sigma^2 \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0\\ &= \sigma^2\left(1 + \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0\right)\end{aligned}$$

Then,

$$Y_0 - \hat{Y}_0 \sim N\left(0, \sigma\sqrt{1 + \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}\right)$$

$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1}$$

With this, we can construct a t ratio

$$\frac{\frac{Y_0 - \hat{Y}_0}{\sigma\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}}}{\sqrt{\frac{(n-k-1)S_e^2}{\sigma^2}/(n-k-1)}} = \frac{Y_0 - \hat{Y}_0}{S_e\sqrt{1 + \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}} \sim t_{n-k-1}$$

and the prediction interval for Y_0 is

$$Y_0 \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-k-1} \cdot S_e\sqrt{1 + \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}$$

For a given $\mathbf{x}_0' = \begin{pmatrix}1 & x_{01} & x_{02} & \dots & x_{0k}\end{pmatrix}$, $\hat{Y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ and

$$\hat{Y}_0 \sim N\left(\mathbf{x}_0'\boldsymbol{\beta}, \sigma\sqrt{\mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}\right)$$

$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1}$$

SO the confidence interval for EY_0 is

$$EY_0 = \mathbf{x}_0'\boldsymbol{\beta} \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-k-1} \cdot S_e\sqrt{\mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0}$$

§21.2 Hypothesis Testing

Suppose $k = 5$ then

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \varepsilon_i$$

Suppose we want to test

1. $H_0 : \beta_1 = 0$, $H_a : \beta_1 \neq 0$
2. $H_0 : \beta_3 = 2$, $H_a : \beta_3 \neq 2$
3. $H_0 : \beta_2 - \beta_5 = 0$, $\beta_2 - \beta_5 \neq 0$
4. $H_0 : \beta_2 = \beta_5 = 0$, H_a : not true
5. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$, H_a : not true or $\beta_{(0)} = \mathbf{0}$, $\beta_{(0)} \neq \mathbf{0}$

As the above can be expressed using

$$\begin{aligned}H_0 &: \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}\\ H_a &: \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma}\end{aligned}$$

1. $\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$, $\gamma = 0$
2. $\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$, $\gamma = 2$
3. $\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix}$, $\gamma = 0$
4. $\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$, $\boldsymbol{\gamma} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Check:
$$\mathbf{C}\boldsymbol{\beta} = \begin{pmatrix} \beta_2 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
5. $$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$
or $\mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \end{pmatrix}$

In general, $\mathbf{C}$ is $m \times k+1$ matrix.

$$\begin{aligned} H_0 &: \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma} \implies \mathbf{c}\boldsymbol{\beta} - \boldsymbol{\gamma} = \mathbf{0} \\ H_a &: \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma} \implies \mathbf{c}\boldsymbol{\beta} - \boldsymbol{\gamma} \neq \mathbf{0} \end{aligned}$$

Consider $\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}$ and find its distribution under H_0.

$$\begin{aligned} E\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right) &= 0 \\ \operatorname{var}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right) &= \sigma^2 \mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}' \end{aligned}$$

Therefore,

$$\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma} \sim N_m\left(\mathbf{0}, \sigma^2 \mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)$$

and let

$$\mathbf{V} = \left[\mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right]^{-\frac{1}{2}} \left[\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right]$$

Then $E\mathbf{V} = \mathbf{0}$

$$\operatorname{var}(\mathbf{V}) = \sigma^2 \mathbf{I}_{m \times m}$$

So, $\mathbf{V} \sim N_m\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$ and $\frac{\mathbf{V}'\mathbf{V}}{\sigma^2} \sim \mathcal{X}^2_m$

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left(\mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{\sigma^2} \sim \mathcal{X}^2_m$$
$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1}$$

$\hat{\boldsymbol{\beta}}$ and S_e^2 are independent. Therefore,

$$\frac{\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left(\mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{\sigma^2} / m}{\frac{(n-k-1)S_e^2}{\sigma^2} / n-k-1} \sim F_{m,n-k-1}$$

or

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left(\mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{mS_e^2} \sim F_{m,n-k-1}$$

§22 Lec 22: Nov 15, 2021

§22.1 F Test for the General Linear Hypothesis

Consider:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$$
$$H_a : \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\gamma}$$

Under H_0 : $\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma} \sim N_m \left(\mathbf{0}, \sigma^2 \mathbf{C} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)$

$$\left.\begin{array}{l} \frac{(\mathbf{C}\hat{\boldsymbol{\beta}}-\boldsymbol{\gamma})'\left(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right)^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}-\boldsymbol{\gamma})}{\sigma^2} \sim \mathcal{X}^2_m \\ \frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1} \end{array}\right\}$$

$$\implies \frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left(\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{mS_e^2} \sim F_{m,n-k-1}$$

Reject H_0 if $F > F_{1-\alpha;m,n-k-1}$
Note: $EmS_e^2 = m\sigma^2$ – expected value of the denominator. Expected value of the numerator (using properties of trace) is

$$m\sigma^2 + (\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})' \left(\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{C}'\right)^{-1} (\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})$$

If H_0 is true, the second term becomes 0 and the expected value is approximately equal to 1.

§22.2 F Statistics and t statistics in Multiple Regression

Suppose $H_0 : \beta_1 = 0$, $H_a : \beta_1 \neq 0$, $k = 5$ and $m = 1$. Then,

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma = 0$$

and $\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma} = \hat{\beta}_1$. Then the F statistics is

$$\frac{\hat{\beta}_1^2}{S_e^2 v_{11}} \sim F_{1,n-k-1}$$

Now we test $H_0 : \beta_1 = 0$ using the t statistics

$$\left.\begin{array}{l} \hat{\beta}_1 \sim N\left(0, \sigma\sqrt{v_{11}}\right) \\ \frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1} \end{array}\right\} \implies \frac{\hat{\beta}_1}{S_e\sqrt{v_{11}}} \sim t_{n-k-1}$$

Thus, $t^2_{n-k-1} = F_{1,n-k-1}$.
Suppose

$$H_0 : \mathbf{a}'\boldsymbol{\beta} = 0$$
$$H_a : \mathbf{a}'\boldsymbol{\beta} \neq 0$$

t-statistics: $\mathbf{a}'\hat{\boldsymbol{\beta}} \sim N\left(0, \sigma\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}\right)$ and

$$\frac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-k-1}$$

Then,

$$\frac{\mathbf{a}'\hat{\boldsymbol{\beta}}}{S_e\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}} \sim t_{n-k-1}$$

§22.3 Power Analysis in Multiple Regression

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{I})$. Then $\mathbf{Y}'\mathbf{Y} \sim \mathcal{X}_n^2(\text{NCP} = \boldsymbol{\mu}'\boldsymbol{\mu})$. Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. Then,

$$\frac{\mathbf{Y}}{\sigma} \sim N_n\left(\frac{\boldsymbol{\mu}}{\sigma}, \mathbf{I}\right)$$
$$\frac{\mathbf{Y}'\mathbf{Y}}{\sigma^2} \sim \mathcal{X}_n^2\left(\text{NCP} = \frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{\sigma^2}\right)$$

Let $Q \sim \mathcal{X}_n^2(\text{NCP} = \theta)$

$$M_Q(t) = (1-2t)^{-\frac{n}{2}} e^{\theta \frac{t}{1-2t}}$$

When H_0 is no true,

$$\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma} \sim N_m\left(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma}, \sigma^2 \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right)$$

Let $\mathbf{V} = \frac{\left(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right)^{-\frac{1}{2}}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})}{\sigma}$. Then,

$$\mathbf{V} \sim N_m\left(\frac{\left(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right)^{-\frac{1}{2}}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})}{\sigma}, \mathbf{I}\right)$$

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{\sigma^2} \sim \mathcal{X}_m^2\left(\text{NCP} = \frac{(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})}{\sigma^2}\right)$$

and the non-central F distribution

$$\left.\begin{array}{l} U \sim \mathcal{X}_n^2(\text{NCP} = \theta) \\ V \sim \mathcal{X}_m^2 \end{array}\right\} \implies \frac{U/n}{V/m} \sim F_{n,m}(\text{NCP} = \theta)$$

where U, V are independent. Apply this for the power analysis

$$\frac{\frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}}{\sigma^2}/m}{\frac{(n-k-1)S_e^2}{\sigma^2}/(n-k-1)} = \frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{mS_e^2} \sim F_{m,n-k-1}$$

with NCP $= \frac{(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})}{\sigma^2}$

figure here

The power is

$$1 - \beta = P\left(F_{m,n-k-1}(\text{NCP} = \theta) > F_{1-\alpha;m,n-k-1}\right)$$

§22.4 F Statistics Using the Extra Sum of Squares

Under $H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$, we have a constrained least squares problem with

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

and

$$\text{SSE}_c = \mathbf{e_c}'\mathbf{e}_c = \mathbf{e}'\mathbf{e} + \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

Using extra sum of squares

$$\frac{(\text{SSE}_c - \text{SSE}_F)/(df_R - df_F)}{\text{SSE}_F/df_F} \sim F_{df_R - df_F, df_F}$$

where $df_F = n-k-1$ and $df_R = n-(k-m)-1$ so $df_R - df_F = m$, e.g.,

$$\begin{gathered} k = 5 \\ H_0 : \beta_1 = \beta_2 = 0 \\ \text{Full} : n-5-1 \\ \text{Reduced} : n-3-1 \\ \implies m = 2 \end{gathered}$$

Thus,

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{mS_e^2} \sim F_{m,n-k-1}$$

which is the same as method 1.

§23 Lec 23: Nov 17, 2021

§23.1 Testing the Overall Significance of the Model

Consider

$$H_0 : \boldsymbol{\beta}_{(0)} = \mathbf{0}$$
$$H_a : \boldsymbol{\beta}_{(0)} \neq \mathbf{0}$$

We can test this hypothesis using the F test for the general linear hypothesis

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{kS_e^2} \sim F_{k,n-k-1}$$

where $m = k$ in this case. We can also use the following test statistic

$$\frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)}$$

<u>*Note*</u>: $\text{SSR} = \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$ where $\mathbf{C} = \begin{pmatrix}\mathbf{0} & \mathbf{I}_k\end{pmatrix}$.

§23.2 Likelihood Ratio Test

Consider:

$$H_0 : \boldsymbol{\beta}_{(0)} = \boldsymbol{\gamma}$$
$$H_a : \boldsymbol{\beta}_{(0)} \neq \boldsymbol{\gamma}$$

We reject H_0 if

$$\Lambda = \frac{L\left(\hat{w}\right)}{L\left(\hat{\omega}\right)} < k$$

where

- $L\left(\hat{w}\right)$: maximized likelihood function under H_0
- $L\left(\hat{\omega}\right)$: maximized likelihood function under no restriction

Note that

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \implies \mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

Thus, the likelihood function is

$$L = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}$$

Without any restrictions

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

and

$$\hat{\sigma}_1^2 = \frac{\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)}{n} = \frac{\mathbf{e}'\mathbf{e}}{n}$$

Under H_0,

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

and

$$\hat{\sigma}_0^2 = \frac{\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_c\right)' \left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_c\right)}{n} = \frac{\mathbf{e}_c'\mathbf{e}_c}{n}$$

Back to LRT, we have

$$\Lambda = \frac{\left(2\pi\hat{\sigma}_0^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma}_0^2}\mathbf{e}_c'\mathbf{e}_c}}{\left(2\pi\hat{\sigma}_1^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma}_1^2}\mathbf{e}'\mathbf{e}}} < k$$

Replace

$$\mathbf{e}_c'\mathbf{e}_c = n\hat{\sigma}_0^2$$
$$\mathbf{e}'\mathbf{e} = n\hat{\sigma}_1^2$$

We obtain

$$\frac{\mathbf{e}'\mathbf{e}}{\mathbf{e}_c'\mathbf{e}_c} < k^{\frac{2}{n}}$$

Also,

$$\mathbf{e}_c'\mathbf{e}_c = \mathbf{e}'\mathbf{e} + \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)$$

Thus,

$$\frac{1}{1 + \frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{\mathbf{e}'\mathbf{e}}} < k^{\frac{2}{n}}$$

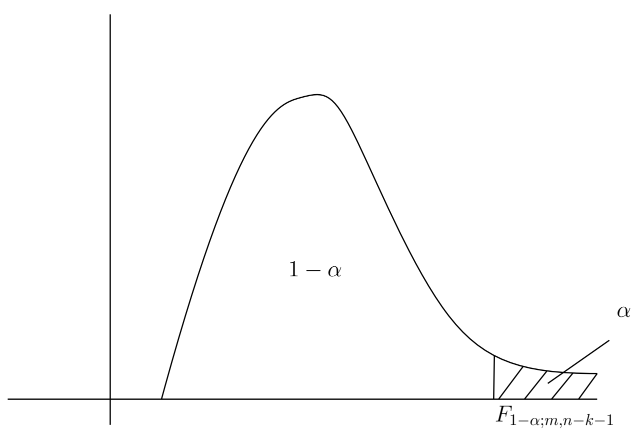
We see that if H_0 is true then $\mathbf{C}\hat{\boldsymbol{\beta}} \approx \boldsymbol{\gamma}$ and therefore the ratio above is approximately equal to 1. If H_0 is no true then the ratio above is less than 1. Manipulating the above expression, we have

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)' \left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}\right)}{mS_e^2} > \left(k^{-\frac{2}{n}} - 1\right)\frac{n-k-1}{m} = k'$$

Use significance level α (type I error) to find k' (rejection region).

$$P\left(F_{m,n-k-1} > k'\right) = \alpha$$

Therefore, $k' = F_{1-\alpha;m,n-k-1}$.



§23.3 Multi-Collinearity

This is problem when some predictors are highly correlated with other predictors.

Example 23.1

Suppose $k = 2$.

$$\left.\begin{array}{l} H_0 : \beta_1 = \beta_2 = 0 \\ H_a : \text{at least on } \beta_i \neq 0 \end{array}\right\} \text{Use F statistic}$$

Suppose we reject H_0 (at least one $b_i \neq 0$). Then test $\beta_1 = 0$ and $\beta_2 = 0$ individually.

$$\begin{array}{ll} H_0 : \beta_1 = 0 & H_0 : \beta_2 = 0 \\ H_a : \beta_1 \neq 0 & H_a : \beta_2 \neq 0 \end{array}$$

Suppose we don't reject H_0 is both tests. This contradiction between the F statistic and the t statistics is a problem caused by multi-collinearity.

Multi-collinearity inflates the variance of $\hat{\beta}_i$ and therefore the corresponding t statistics will be small. To explain this we will use the centered and scaled model

$$\mathbf{Y} = \gamma_0 \mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon}$$

where $\mathbf{Z} = \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}$. Or

$$Y_i = \gamma_0 + \beta_1 \left(X_{i1} - \overline{X}_1\right) + \beta_2 \left(X_{i2} - \overline{X}_2\right) + \ldots + \beta_k \left(X_{ik} - \overline{X}_k\right) + \varepsilon_i$$

where $i = 1, 2, \ldots, n$. Or

$$Y_i = \gamma_0 + \beta_1 z_{i1} + \beta_2 Z_{i2} + \ldots + \beta_k Z_{ik} + \varepsilon_i$$

where $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ are the centered predictors.

§24 Lec 24: Nov 19, 2021

§24.1 Centered and Scaled Model in Matrix/Vector Form

Consider the centered model

$$\begin{aligned}
\mathbf{Y} &= \gamma_0 \mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon} \\
\gamma_0 &= \beta_0 + \frac{1}{n}\mathbf{1}'\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} \\
&= \beta_0 + \beta_1 \overline{\mathbf{x}}_1 + \ldots + \beta_k \overline{\mathbf{x}}_k \\
\mathbf{Z} &= \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X}_{(0)}
\end{aligned}$$

or

$$\begin{aligned}
Y_i &= \gamma_0 + \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \ldots + \beta_k(x_{ik} - \overline{x}_k) + \varepsilon_i \\
&= \gamma_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \ldots + \beta_k Z_{ik} + \varepsilon_i
\end{aligned}$$

Centering and scaling: multiply and divide each centered prediction by $\sqrt{\sum(x_{ij} - \overline{x}_j)^2}$. Then,

$$Y_i = \gamma_0 + \beta_1\sqrt{\sum(x_{i1} - \overline{x}_1)^2}\frac{(x_{1i} - \overline{x}_1}{\sqrt{\sum(x_{i1} - \overline{x}_1)^2}} + \ldots + \beta_k\sqrt{\sum(x_{ik} - \overline{x}_k)^2}\frac{(x_{ik} - \overline{x}_k)}{\sqrt{\sum(x_{ik} - \overline{x}_k)^2}} + \varepsilon_i$$

or

$$Y_i = \gamma_0 + \delta_1 Z_{si1} + \ldots + \delta_k Z_{sik} + \varepsilon_i$$

where

$$\begin{aligned}
\delta_j &= \beta_j\sqrt{\sum(x_{ij} - \overline{x}_j)^2} \\
Z_{sij} &= \frac{x_{ij} - \overline{x}_j}{\sqrt{\sum(x_{ij} - \overline{x}_j)^2}}
\end{aligned}$$

From the centered model

$$\mathbf{Y} = \gamma_0\mathbf{1} + \underbrace{\mathbf{Z}\mathbf{D}^{-1}}_{\mathbf{Z}_s}\underbrace{\mathbf{D}\boldsymbol{\beta}_{(0)}}_{\boldsymbol{\delta}_{(0)}} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{D} = \begin{pmatrix} \sqrt{\sum(x_{i1} - \overline{x}_1)^2} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\sum(x_{ik} - \overline{x}_k)^2} \end{pmatrix}$$

Then

$$\mathbf{Y} = \gamma_0\mathbf{1} + \mathbf{Z}_s\boldsymbol{\delta}_{(0)} + \boldsymbol{\varepsilon}$$

Note:

1. $\mathbf{1}'\mathbf{Z}_s = \mathbf{1}'\mathbf{Z}\mathbf{D}^{-1} = \mathbf{1}'\left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right]\mathbf{X}_{(0)}\mathbf{D}^{-1} = \mathbf{0}'$
2. $\mathbf{Z}_s'\mathbf{1} = \mathbf{0}$
3. $\mathbf{Z}_s'\mathbf{Z}_s = \begin{bmatrix} \mathbf{Z}_{s1}' \\ \vdots \\ \mathbf{Z}_{sk}' \end{bmatrix}\begin{bmatrix} \mathbf{Z}_{s1} & \ldots & \mathbf{Z}_{sk} \end{bmatrix}$

Let's examine $\mathbf{Z}_{s1}'\mathbf{Z}_{s1}$ and $\mathbf{Z}_{s1}'\mathbf{Z}_{s2}$

$$\mathbf{Z}_{s1}'\mathbf{Z}_{s1} = \begin{bmatrix} \frac{x_{i1}-\overline{x}_1}{\sqrt{\sum(x_{i1}-\overline{x}_1)^2}} & \cdots & \frac{x_{n1}-\overline{x}_1}{\sqrt{\sum(x_{i1}-x_1)^2}} \end{bmatrix} \begin{pmatrix} \frac{x_{i1}-\overline{x}_1}{\sqrt{\sum(x_{i1}-\overline{x}_1)^2}} \\ \vdots \\ \frac{x_{n1}-\overline{x}_1}{\sqrt{\sum(x_{i1}-\overline{x}_1)^2}} \end{pmatrix} = \frac{\sum(x_{i1}-\overline{x}_1)^2}{\sum(x_{i1}-\overline{x}_1)^2} = 1$$

Similarly for $\mathbf{Z}_{s1}'\mathbf{Z}_{s2}$,

$$\mathbf{Z}_{s1}'\mathbf{Z}_{s2} = \frac{\sum(x_{i1}-\overline{x}_1)(x_{i2}-\overline{x}_2)/n-1}{\sqrt{\sum(x_{i1}-\overline{x}_1)^2}\sqrt{\sum(x_{i2}-\overline{x}_2)^2/n-1}} = v_{12}$$

Then,

$$\mathbf{Z}_s'\mathbf{Z}_s = \mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}$$

Estimation of $\mathbf{Y} = \gamma_0 \mathbf{1} + \mathbf{Z}_s \delta_{(0)} + \boldsymbol{\varepsilon}$

$$\begin{pmatrix} \hat{\gamma}_0 \\ \hat{\boldsymbol{\delta}}_{(0)} \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z}_s \\ \mathbf{Z}_s'\mathbf{1} & \mathbf{Z}_s'\mathbf{Z}_s \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{Z}_s'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} n & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_s'\mathbf{Z}_s \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{Z}_s'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & \mathbf{0}' \\ \mathbf{0} & (\mathbf{Z}_s'\mathbf{Z}_s)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}'\mathbf{Y} \\ \mathbf{Z}_s'\mathbf{Y} \end{pmatrix}$$

So

$$\hat{\gamma}_0 = \overline{y}$$

which is the same as the estimate of γ_0 of the centered model. And

$$\hat{\boldsymbol{\delta}}_{(0)} = (\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{Y}$$

Properties:

$$\begin{aligned} E\hat{\boldsymbol{\delta}}_{(0)} &= (\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s' E\mathbf{Y} \\ &= (\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\left(\gamma_0\mathbf{1} + \mathbf{Z}_s\boldsymbol{\delta}_{(0)}\right) \\ &= \mathbf{0} + (\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{Z}_s\boldsymbol{\delta}_{(0)} \\ &= \boldsymbol{\delta}_{(0)} \\ \operatorname{var}(\hat{\boldsymbol{\delta}}_{(0)}) &= \operatorname{var}\left[(\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{Y}\right] \\ &= \sigma^2\mathbf{R}^{-1} \end{aligned}$$

Non-Centered Model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ or } \mathbf{Y} = \beta_0\mathbf{1} + \mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} + \boldsymbol{\varepsilon}$$

Centered Model:

$$\mathbf{Y} = \gamma_0\mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{(0)} + \varepsilon$$

Centered/Scaled Model:

$$\mathbf{Y} = \gamma_0\mathbf{1} + \mathbf{Z}_s\boldsymbol{\delta}_{(0)} + \varepsilon$$

where

$$\begin{aligned} \gamma_0 &= \beta_0 + \frac{1}{n}\mathbf{1}'\mathbf{X}_{(0)}\boldsymbol{\beta}_{(0)} \\ \boldsymbol{\delta}_{(0)} &= \mathbf{D}\boldsymbol{\beta}_{(0)} \end{aligned}$$

So

$$\begin{aligned}\hat{\beta}_0 &= \hat{\gamma}_0 - \frac{1}{n}\mathbf{1}'\mathbf{X}_{(0)}\hat{\boldsymbol{\beta}}_{(0)} \\ &= \mathbf{y} - \frac{1}{n}\mathbf{1}'\mathbf{X}_{(0)}\mathbf{D}^{-1}\hat{\boldsymbol{\delta}}_{(0)} \\ \hat{\boldsymbol{\beta}}_{(0)} &= \mathbf{D}^{-1}\hat{\boldsymbol{\delta}}_{(0)}\end{aligned}$$

§25 Lec 25: Nov 22, 2021

§25.1 Multi-Collinearity

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ be predictors. Some predictors are highly correlated with other predictors. Earlier we saw that $\delta_1 = \beta_1 \sqrt{\sum (x_{i1} - \overline{x}_1)^2}$

$$\begin{aligned}
\hat{\delta}_1 &= \hat{\beta}_1 \sqrt{\sum (x_{i1} - \overline{x}_1)^2} \\
\hat{\beta}_1 &= \frac{\hat{\delta}_1}{\sqrt{\sum (x_{i1} - \overline{x}_1)^2}} \\
\operatorname{var}(\hat{\beta}_1) &= \frac{\operatorname{var}(\hat{\delta}_1)}{\sum (x_{i1} - \overline{x}_1)^2}
\end{aligned}$$

So let's find the variance of $\hat{\delta}_1$ using the centered and scaled model.

$$\operatorname{var}(\hat{\boldsymbol{\delta}}_{(0)}) = \sigma^2 \mathbf{R}^{-1} = \sigma^2 \begin{pmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1k} \\ r_{21} & 1 & r_{23} & \ldots & r_{2k} \\ \vdots & & \ddots & & \vdots \\ r_{k1} & r_{k2} & r_{k3} & & 1 \end{pmatrix} = *$$

Therefore, $\operatorname{var}(\hat{\delta}_1) = \sigma^2 \mathbf{R}^{-1}[1,1]$. Using the inverse of a partitioned matrix

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_{11}^{-1} & -\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \\ -\mathbf{C}_{21}\mathbf{C}_{11}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \end{pmatrix}$$

Here $A_{11} = 1$, $\mathbf{A}_{12} = \mathbf{r}'$, $\mathbf{A}_{21} = r$, $\mathbf{A}_{22} = \mathbf{R}_{22}$, $\mathbf{C}_{11} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$. Therefore,

$$\operatorname{var}(\hat{\delta}_1) = \frac{\sigma^2}{1 - \mathbf{r}'\mathbf{R}_{22}^{-1}\mathbf{r}}$$

We will sow that $\operatorname{var}(\hat{\delta}_1) = \frac{\sigma^2}{1 - R_1^2}$ where R_1^2 is the R-square from the regression of X_1 on $X_2, X_3, \ldots, X_k$. Instead, we can regress Z_{s_1} on $Z_{s_2}, Z_{s_3}, \ldots, Z_{s_k}$. Because we have seen that the three models:

- Non-centered
- Centered
- Centered/Scaled

Here is the model

$$\begin{aligned}
Z_{si1} &= \alpha_0 + \alpha_1 Z_{si2} + \alpha_2 Z_{si3} + \ldots + \alpha_{k-1} Z_{sik} + \varepsilon_i \\
R_1^2 &= \frac{\text{SSR}}{\text{SST}} \\
\text{SST} &= \sum \left(Z_{si1} - \overline{Z}_{s1} \right)^2 = \sum Z_{si1}^2
\end{aligned}$$

But $\sum Z_{i1}^2 = \mathbf{Z}_{s1}'\mathbf{Z}_{s1} = 1$. So far, we have $R^2 = \text{SSR} = \sum \left(\hat{Z}_{si1} - \underbrace{\overline{Z}_{s1}}_{\to 0} \right)^2$ or $R_1^2 = \sum \hat{Z}_{si1}^2 = \hat{\mathbf{Z}}_{s1}'\hat{\mathbf{Z}}_{s1}$. Here, $\hat{\mathbf{Z}}_{s1} = \mathbf{H}\mathbf{Z}_{s1}$ where $\mathbf{H}$ is the hat matrix using $\mathbf{Z}_{s_2}, \mathbf{Z}_{s_3}, \ldots, \mathbf{Z}_{s_k}$. Therefore, $R_1^2 = \mathbf{Z}_{s1}'\mathbf{H}\mathbf{Z}_{s_1}$ or

$$R_1^2 = \mathbf{Z}_{s_1}'\mathbf{Z}_s^* \left(\mathbf{Z}_s^{*\prime}\mathbf{Z}^* \right)^{-1} \mathbf{Z}_s^{*\prime}\mathbf{Z}_{s_1} = \mathbf{r}'\mathbf{R}_{22}^{-1}\mathbf{r}$$

Earlier we found that

$$\text{var}(\hat{\delta}_1) = \frac{\sigma^2}{1 - \mathbf{r}'\mathbf{R}_{22}^{-1}\mathbf{r}} = \frac{\sigma^2}{1 - R_1^2}$$

Now back to the variance of $\hat{\beta}_1$:

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}(\hat{\delta}_1)}{\sum(x_{i1} - \overline{x}_1)^2}$$

Replace $\text{var}(\hat{\delta}_1) = \frac{\sigma^2}{1-R_1^2}$ we obtain

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{(1 - R_1^2)\sum(x_{i1} - \overline{x}_1)^2}$$

Therefore, if R_1^2 is close to 1, then $\text{var}(\hat{\beta}_1)$ is large.
Detection of Multi-Collinearity: use variance inflation factor (VIF). For each predictor j compute VIF_j

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R-square from the regression of predictor x_j on the other predictors. For example, if VIF > 10, then $R_j^2 > 0.90$ which means x_j is highly correlated with the other predictors.

§25.2 Generalized Least Squares

Consider the model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

So far we assumed the Gauss-Markov conditions. Suppose now

$$E\boldsymbol{\varepsilon} = \mathbf{0}$$
$$\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$$

where $\mathbf{V}$ is a symmetric matrix of constants. If we use the ordinary least squares (OLS) estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. We still get $E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ because $E\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ but $\text{var}(\hat{\boldsymbol{\beta}}) = \text{var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Therefore, $\hat{\boldsymbol{\beta}}$ is not BLUE because the Gauss-Markov conditions do not hold. We transform the model as follows: let $\mathbf{V}^{-\frac{1}{2}}$ be the inverse square root matrix of $\mathbf{V}$. Multiply the model on both sides by $\mathbf{V}^{-\frac{1}{2}}$

$$\mathbf{V}^{-\frac{1}{2}}\mathbf{Y} = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\varepsilon}$$

or

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$$
$$E\boldsymbol{\varepsilon}^* = E\left(\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\varepsilon}\right) = 0$$
$$\text{var}(\boldsymbol{\varepsilon}^*) = \text{var}\left(\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{V}^{-\frac{1}{2}}\mathbf{V}\mathbf{V}^{-\frac{1}{2}} = \sigma^2\mathbf{I}$$

with this transformation we see that the Gauss-Markov conditions hold. Therefore, we estimate $\boldsymbol{\beta}$ using

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\mathbf{X}^{*'}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*'}\mathbf{Y}^*$$

Replace $\mathbf{X}^* = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}$ and $\mathbf{Y}^* = \mathbf{V}^{-\frac{1}{2}}\mathbf{Y}$ to get

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

Then, the mean is

$$E\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}E\mathbf{Y} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

which is unbiased. And

$$\operatorname{var}\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = \operatorname{var}\left(\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\right) = \sigma^2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$$

§26 Lec 26: Nov 24, 2021

§26.1 Generalized Least Squares (Cont'd)

Estimate $\boldsymbol{\beta}$ by direct minimization of the error sum of squares using the transformed model

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$$

$\min \boldsymbol{\varepsilon}^{*'}\boldsymbol{\varepsilon}^*$ or $\min\left(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}\right)'\left(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}\right)$. Replace $\mathbf{Y}^* = \mathbf{V}^{-\frac{1}{2}}\mathbf{Y}$ and $\mathbf{X}^* = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}$. We minimize

$$\begin{aligned}
\min Q &= \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)' \mathbf{V}^{-1}\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) \\
&= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - 2\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} \\
\frac{\partial Q}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = 0 \\
\hat{\boldsymbol{\beta}}_{GLS} &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}
\end{aligned}$$

Assume now $\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2\mathbf{V}\right)$. Then, $\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\right)$ and

$$\begin{aligned}
L &= \frac{1}{(2\pi)^{\frac{n}{2}}}\left|\sigma^2\mathbf{V}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})'(\sigma^2\mathbf{V})^{-1}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})} \\
\ln L &= -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2}\ln|V| - \frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right) \\
\frac{\partial \ln L}{\partial \boldsymbol{\theta}} &= \mathbf{0} \\
\hat{\boldsymbol{\beta}}_{GLS} &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}
\end{aligned}$$

Estimation of σ^2

$$\begin{aligned}
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\right) = 0 \\
\hat{\sigma}^2 &= \frac{\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_{GLS}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_{GLS}\right)}{n} = \frac{\left(\mathbf{y}^* - \mathbf{x}^*\hat{\boldsymbol{\beta}}_{GLS}\right)'\left(\mathbf{y}^* - \mathbf{x}^*\hat{\boldsymbol{\beta}}_{GLS}\right)}{n} \\
&= \frac{\mathbf{e}_{GLS}'\mathbf{e}_{GLS}}{n}
\end{aligned}$$

Use the properties of trace to find $E\hat{\sigma}^2$

$$\begin{aligned}
E\hat{\sigma}^2 &= \frac{1}{n}E\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_{GLS}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}_{GLS}\right) \\
&= \frac{1}{n}\operatorname{tr}\mathbf{V}^{-1}\left[\operatorname{var}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}\right) + \mathbf{0}\mathbf{0}'\right]
\end{aligned}$$

because $E\left[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}\right] = \mathbf{0}$. So

$$\begin{aligned}
\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS} &= \mathbf{Y} - \mathbf{X}\left[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \\
&= \left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{Y}
\end{aligned}$$

So

$$\begin{aligned}
\operatorname{var}\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}\right) &= \sigma^2\left[\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\mathbf{V}\left[\mathbf{I} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\right] \\
&= \sigma^2\mathbf{V} - \sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'
\end{aligned}$$

Back to expectation

$$
\begin{aligned}
E\hat{\sigma}^2 &= \frac{1}{n}\operatorname{tr}\mathbf{V}^{-1}\left[\sigma^2\mathbf{V} - \sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\right] \\
&= \frac{1}{n}\left[\sigma^2\operatorname{tr}\mathbf{I}_n - \sigma^2\operatorname{tr}\mathbf{I}_{k+1}\right] \\
&= \frac{n-k-1}{n}\sigma^2
\end{aligned}
$$

Thus, the unbiased estimator of σ^2 is $S^2_{\mathbf{e}_{GLS}} = \frac{\mathbf{e}'_{GLS}\mathbf{e}_{GLS}}{n-k-1}$.

§26.2 Comparing Regression Equations

Suppose we have two data sets on the same variables

$$
\begin{aligned}
\mathbf{Y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\
\mathbf{Y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2
\end{aligned}
$$

Let $\boldsymbol{\beta}_1 = \begin{pmatrix}\boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}_1^{(2)}\end{pmatrix}$, $\boldsymbol{\beta}_2 = \begin{pmatrix}\boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}_2^{(2)}\end{pmatrix}$. Note that

$$
\boldsymbol{\beta}^{(1)} : p \times 1, \quad \boldsymbol{\beta}_1^{(2)} : (k+1-p) \times 1, \quad \boldsymbol{\beta}_2^{(2)} : (k+1-p) \times 1
$$

Suppose we want to test $\boldsymbol{\beta}_1^{(2)} = \boldsymbol{\beta}_2^{(2)}$ (assume that the first p elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the same). We can construct one model as follows

$$
\begin{pmatrix}\mathbf{Y}_1 \\ \mathbf{Y}_2\end{pmatrix} = \begin{pmatrix}\mathbf{X}_1^{(1)} & \mathbf{X}_1^{(2)} & \mathbf{0} \\ \mathbf{X}_2^{(1)} & \mathbf{0} & \mathbf{X}_2^{(2)}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}_1^{(2)} \\ \boldsymbol{\beta}_2^{(2)}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2\end{pmatrix}
$$

Therefore, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and the hypothesis $\boldsymbol{\beta}_1^2 = \boldsymbol{\beta}_2^2$ can be tested using

$$
\begin{aligned}
H_0 &: \mathbf{C}\boldsymbol{\beta} = \mathbf{0} \\
H_a &: \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}
\end{aligned}
$$

§27 Lec 27: Nov 29, 2021

§27.1 Comparing Regression Equations (Cont'd)

We can use the F test for the general linear hypothesis

$$\frac{\left(\mathbf{C}\hat{\boldsymbol{\beta}}-\boldsymbol{\gamma}\right)'\left[\mathbf{C}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{C}'\right]^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}}-\boldsymbol{\gamma}\right)}{mS_e^2} \sim F_{m,n-k-1}$$

Example 27.1

Suppose $k=5$ and $p=3$. We want to test

$$\begin{aligned} &H_0: \beta_3^{(1)}=\beta_3^{(2)}, \quad \beta_4^{(1)}=\beta_4^{(2)}, \quad \beta_5^{(1)}=\beta_5^{(2)} \\ &H_a: \text{ not true} \\ &\mathbf{C}=\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

and

$$\boldsymbol{\beta}=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3^{(1)} \\ \beta_4^{(1)} \\ \beta_5^{(1)} \\ \beta_3^{(2)} \\ \beta_4^{(2)} \\ \beta_5^{(2)} \end{pmatrix}$$

In general,

$$\mathbf{C}=\left(\mathbf{0}_{k+1-p,p} \quad \mathbf{I}_{k+1-p} \quad -\mathbf{I}_{k+1-p}\right)$$

Therefore,

$$\mathbf{C}:(k+1-p)\times(2(k+1)-p)$$

We can also test the hypothesis using the extra sum of squares principle. Under the null hypothesis, the model is expressed as follows

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}=\begin{pmatrix} \mathbf{X}_1' & \mathbf{X}_1^2 \\ \mathbf{X}_2' & \mathbf{X}_2^2 \end{pmatrix}\begin{pmatrix} \boldsymbol{\beta}' \\ \boldsymbol{\beta}^* \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

where $\boldsymbol{\beta}^*$ is the common beta subvector under H_0. Therefore,

$$\begin{aligned} \frac{(\mathrm{SSE}_R-\mathrm{SSE}_F)/(df_R-df_F)}{\mathrm{SSE}_F/df_F} &\sim F_{df_R-df_F,df_F} \\ df_F=n-p-2(k+1-p) &= n+p-2(k+1) \\ df_R=n-p-(k+1-p) &= n-k-1 \end{aligned}$$

Example 27.2

Suppose $k = 5$, $p = 4$

$$\begin{aligned} &H_0 : \beta_4^{(1)} = \beta_5^{(1)}, \quad \beta_5^{(1)} = \beta_5^{(2)} \\ &H_a : \text{ not true} \end{aligned}$$

Formulation:

$$\begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11}^{(1)} & x_{12}^{(1)} & x_{13}^{(1)} & x_{14}^{(1)} & x_{15}^{(1)} & 0 & 0 \\ 1 & x_{21}^{(1)} & x_{22}^{(1)} & x_{23}^{(1)} & x_{24}^{(1)} & x_{25}^{(1)} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1}^{(1)} & x_{n2}^{(1)} & x_{n3}^{(1)} & x_{n4}^{(1)} & x_{n5}^{(1)} & 0 & 0 \\ 1 & x_{11}^{(2)} & x_{12}^{(2)} & x_{13}^{(2)} & 0 & 0 & x_{14}^{(2)} & x_{15}^{(2)} \\ 1 & x_{21}^{(2)} & x_{22}^{(2)} & x_{23}^{(2)} & 0 & 0 & x_{24}^{(2)} & x_{25}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1}^{(2)} & x_{n2}^{(2)} & x_{n3}^{(2)} & 0 & 0 & x_{n4}^{(2)} & x_{n5}^{(2)} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4^{(1)} \\ \beta_5^{(1)} \\ \beta_4^2 \\ \beta_5^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{n1} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{n2} \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

$$df_F = n - 8, \quad df_R = n - 6$$

§27.2 Deleting a Single Point in Multiple Regression

We want to explore the effect of deleting a single point in multiple regression

- Effect on $\hat{\boldsymbol{\beta}}$
- Effect on S_e^2
- Effect on fitted values

We can delete one point at a time and run a new regression each time to see the effect. But this will require $n + 1$ regressions (one on the full data set and n regressions when we delete data point i, $i = 1, \ldots, n$). There is a more automated way and the result is based on the residuals, e_i, and leverage values h_{ii} from the regression of the full data set.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Suppose we want to delete data point i

$$\begin{pmatrix} \mathbf{Y}_{(i)} \\ Y_i \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{(i)} \\ x_i' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon}_{(i)} \\ \varepsilon_i \end{pmatrix}$$

where $\mathbf{Y}_{(i)}$, $\mathbf{X}_{(i)}$ are the vectors $\mathbf{Y}$ and matrix $\mathbf{X}$ after deleting point i from the data set. We are working now with the model

$$\begin{aligned} \mathbf{Y}_{(i)} &= \mathbf{X}_{(i)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(i)} \\ \hat{\boldsymbol{\beta}}_{(i)} &= \left(\mathbf{X}_{(i)}'\mathbf{X}_{(i)}\right)^{-1} \mathbf{X}_{(i)}'\mathbf{Y}_{(i)} \end{aligned}$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ is the estimator of the vector $\boldsymbol{\beta}$ after deleting data point i.

§28 Lec 28: Dec 1, 2021

§28.1 Deleting a Single Point in Multiple Regression (Cont'd)

Let's find expressions for $\mathbf{X}_{(i)}'\mathbf{X}_{(i)}$ and $\mathbf{X}_{(i)}'\mathbf{Y}_{(i)}$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}_{(i)}' & \mathbf{x}_i \end{pmatrix} \begin{pmatrix} \mathbf{X}_{(i)} \\ \mathbf{x}_i' \end{pmatrix} = \mathbf{X}_{(i)}'\mathbf{X}_{(i)} + \mathbf{x}_i\mathbf{x}_i'$$
$$\mathbf{X}_{(i)}'\mathbf{X}_{(i)} = \mathbf{X}'\mathbf{X} - \mathbf{x}_i\mathbf{x}_i'$$

<u>Result</u>: Let $\mathbf{A}$ be a square invertible matrix and $\mathbf{b}$ be a vector. Then

$$\left[\mathbf{A} - \mathbf{b}\mathbf{b}'\right]^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\mathbf{b}\mathbf{b}'\mathbf{A}^{-1}}{1 - \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}}$$

<u>*Note*</u>: $\mathbf{b}'\mathbf{A}^{-1}\mathbf{b} \neq 1$. We can verify this as follows: multiply both sides by $\left(\mathbf{A} - \mathbf{b}\mathbf{b}'\right)$ to get identity matrix both sides. In our problem $\mathbf{A}$ is $\mathbf{X}'\mathbf{X}$ and $\mathbf{b}$ is $\mathbf{x}_i$. Using the result we can find $\left(\mathbf{X}_{(i)}'\mathbf{X}_{(i)}\right)^{-1}$. Now let's find $\mathbf{X}_{(i)}'\mathbf{Y}_{(i)}$. From $\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \mathbf{X}_{(i)}' & \mathbf{X}_{(i)} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{(i)} \\ y_i \end{pmatrix} = \mathbf{X}_{(i)}'\mathbf{Y}_{(i)} + \mathbf{x}_i y_i$. We get $\mathbf{X}_{(i)}'\mathbf{Y}_{(i)} = \mathbf{X}'\mathbf{Y} - \mathbf{x}_i y_i$. Back to $\hat{\boldsymbol{\beta}}_{(i)} = \left(\mathbf{X}_{(i)}'\mathbf{X}_{(i)}\right)^{-1}\mathbf{X}_{(i)}'\mathbf{Y}_{(i)}$. To find

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i}{1 - h_{ii}} e_i$$

Then

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i e_i}{1 - h_{ii}}$$

This is the difference in the estimator of $\boldsymbol{\beta}$ before and after deleting data point i.
Effect on fitted values:

$$\begin{aligned} \hat{Y}_i - \hat{Y}_{i(i)} &= \mathbf{x}_i'\hat{\boldsymbol{\beta}} - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{(i)} \\ &= \mathbf{x}_i'\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right) \\ &= \mathbf{x}_i'\frac{\left(\mathbf{X}'\mathbf{X}\right)^{-1} x_i e_i}{1 - h_{ii}} \\ &= \frac{h_{ii}}{1 - h_{ii}} e_i \end{aligned}$$

Finally, we can show that the error sum of square after deleting data point are connected with the error sum of squares of the full data set as follows

$$(n - k - 2)S_{e(i)}^2 = (n - k - 1)S_e^2 - \frac{e_i^2}{1 - h_{ii}}$$

- $S_{e(i)}^2$ is the unbiased estimator of σ^2 after deleting data point i
- S_e^2 is unbiased estimator of σ^2 using full data set
- e_i, h_{ii} are residual i and leverage i using full data set

$$\begin{aligned} \mathbf{e} &= \left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y} \\ e_i &= \left(\mathbf{I} - \mathbf{H}\right)_i \mathbf{Y} \\ h_{ii} &= \mathbf{x}_i'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i \end{aligned}$$

Adding a data point in multiple regression

$$\begin{pmatrix} \mathbf{Y} \\ y_0 \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_0' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \varepsilon_0 \end{pmatrix}$$
$$\mathbf{Y}_{\text{new}} = \mathbf{X}_{\text{new}} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{\text{new}}$$
$$\begin{pmatrix} \mathbf{X}_{\text{new}}' & \mathbf{X}_{\text{new}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}' & \mathbf{x}_0 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_0' \end{pmatrix} = \mathbf{X}'\mathbf{X} + \mathbf{x}_0 \mathbf{x}_0'$$

<u>Result</u>: Let $\mathbf{A}$ be a square invertible matrix and $\boldsymbol{\beta}$ a vector s.t. $1 + \mathbf{b}'\mathbf{A}^{-1}\mathbf{b} \neq 1$. Then

$$\left(\mathbf{A} + \mathbf{b}\mathbf{b}'\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{b}\mathbf{b}'\mathbf{A}^{-1}}{1 + \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}}$$

Here $\mathbf{A}$ is $\mathbf{X}'\mathbf{X}$ and $\mathbf{b}$ is $\mathbf{x}_0$. Also,

$$\mathbf{X}_{\text{new}}' \mathbf{Y}_{\text{new}} = \begin{pmatrix} \mathbf{X}' & \mathbf{x}_0 \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ y_0 \end{pmatrix} = \mathbf{X}'\mathbf{Y} + \mathbf{x}_0 y_0$$

Finally,

$$\hat{\boldsymbol{\beta}}_{\text{new}} = \hat{\boldsymbol{\beta}} + \frac{\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0 e_0}{1 + h_{00}}$$

where $e_0 = y_0 - \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ and $h_{00} = \mathbf{x}_0' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0$.

§28.2 Influential Analysis

Internally studentized residuals

$$\left.\begin{aligned} & e_i \sim N\left(0, \sigma\sqrt{1-h_{ii}}\right) \\ & \tfrac{(n-k-1)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-k-1}^2 \end{aligned}\right\} \implies \frac{\frac{e_i}{\sigma\sqrt{1-h_{ii}}}}{\sqrt{\frac{(n-k-1)S_e^2}{\sigma^2}/(n-k-1)}} = \frac{e_i}{S_e\sqrt{1-h_{ii}}}$$

This ratio does not follow a t distribution because e_i and S_e are not independent. Let $v_i = \frac{e_i}{S_e\sqrt{1-h_{ii}}}$. Show that

$$\frac{v_i^2}{n-k-1} \sim \text{beta}\left(\frac{1}{2}, \frac{1}{2}(n-k-2)\right)$$

So we need to show that

$$\frac{e_i^2}{\text{SSE}(1-h_{ii})} \sim \text{beta}\left(\frac{1}{2}, \frac{1}{2}(n-k-2)\right)$$

1. $\mathbf{e} = \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}$, thus

   $$e_i = \mathbf{c}_i' \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}$$

   where $\mathbf{c}_i' = \begin{pmatrix} 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \end{pmatrix}$ in which 1 is at the ith position. So

   $$e_i^2 = e_i e_i = \boldsymbol{\varepsilon}' \left(\mathbf{I} - \mathbf{H}\right) c_i c_i' \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}$$

2. $\text{SSE} = \boldsymbol{\varepsilon}' \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}$

§29 Lec 29: Dec 3, 2021

§29.1 Influential Analysis (Cont'd)

Let's express $\frac{e_i^2}{\text{SSE}(1-h_{ii})}$ as follows

$$\frac{\boldsymbol{\varepsilon}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\left(\mathbf{I}-\mathbf{H}\right)\boldsymbol{\varepsilon}(1-h_{ii})}$$

Divide by σ^2

$$\frac{\frac{\boldsymbol{\varepsilon}'}{\sigma}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\frac{\boldsymbol{\varepsilon}}{\sigma}}{\frac{\boldsymbol{\varepsilon}'}{\sigma}\left(\mathbf{I}-\mathbf{H}\right)\frac{\boldsymbol{\varepsilon}}{\sigma}(1-h_{ii})} = \frac{\mathbf{Z}'\mathbf{Q}\mathbf{Z}}{\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Z}}$$

Here $\mathbf{Z} = \frac{\boldsymbol{\varepsilon}}{\sigma} \sim N\left(\mathbf{0},\mathbf{I}\right)$ and $\mathbf{Q} = \frac{(\mathbf{I}-\mathbf{H})\mathbf{c}_i\mathbf{c}_i'(\mathbf{I}-\mathbf{H})}{1-h_{ii}}$

3. Have

$$\begin{aligned}\mathbf{QQ} &= \frac{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)}{(1-h_{ii})^2}\\ &= \frac{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)}{(1-h_{ii})^2}\\ &= \frac{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)}{1-h_{ii}}\\ &= \mathbf{Q}\end{aligned}$$

Note $\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i = 1-h_{ii}$. Thus, $\mathbf{Q}$ is symmetric and idempotent matrix. Because $\mathbf{Z} \sim N\left(\mathbf{0},\mathbf{I}\right)$, it follows that $\mathbf{Z}'\mathbf{QZ} \sim \mathcal{X}^2_{\text{tr}(\mathbf{Q})}$

$$\text{tr}\left(\mathbf{Q}\right) = \text{tr}\left(\frac{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)}{1-h_{ii}}\right) = \text{tr}\left(\frac{\mathbf{c}_i'\left(\mathbf{I}-\mathbf{H}\right)\left(\mathbf{I}-\mathbf{H}\right)\mathbf{c}_i}{1-h_{ii}}\right) = 1$$

Thus, $\mathbf{Z}'\mathbf{QZ} \sim \mathcal{X}^2_1$

4. Back to the ratio

$$\frac{\mathbf{Z}'\mathbf{QZ}}{\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Z}} = \frac{\mathbf{Z}'\mathbf{QZ}}{\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\mathbf{Z}+\mathbf{Z}'\mathbf{QZ}}$$

$$\left.\begin{aligned}&\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\mathbf{Q} = \mathbf{Q}-\mathbf{HQ}-\mathbf{QQ}\\ &\mathbf{HQ} = \frac{\mathbf{H}(\mathbf{I}-\mathbf{H})\mathbf{c}_i\mathbf{c}_i'(\mathbf{I}-\mathbf{H})}{1-h_{ii}} = \mathbf{0}\end{aligned}\right\}$$

Thus, $\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\mathbf{Z}$ and $\mathbf{Z}'\mathbf{QZ}$ are independent.

5. Consider

$$\begin{aligned}\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right) &= \mathbf{I}-\mathbf{H}-\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Q}-\mathbf{Q}\left(\mathbf{I}-\mathbf{H}\right)+\mathbf{QQ}\\ &= \mathbf{I}-\mathbf{H}-\mathbf{Q}+\mathbf{0}-\mathbf{Q}+\mathbf{0}+\mathbf{Q}\\ &= \mathbf{I}-\mathbf{H}-\mathbf{Q}\end{aligned}$$

Therefore, $\mathbf{I}-\mathbf{H}-\mathbf{Q}$ is symmetric and idempotent matrix. It follows that

$$\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\mathbf{Z} \sim \mathcal{X}^2_{\text{tr}(\mathbf{I}-\mathbf{H}-\mathbf{Q})}$$

or

$$\mathbf{Z}'\left(\mathbf{I}-\mathbf{H}-\mathbf{Q}\right)\mathbf{Z} \sim \mathcal{X}^2_{n-k-2}$$

Result: Let $X \sim \Gamma(\alpha_1, \beta)$, $Y \sim \Gamma(\alpha_2, \beta)$ and X, Y are independent. Let $U = X + Y$ and $V = \frac{X}{X+Y}$. Then, U, V are independent and

$$U \sim \Gamma(\alpha_1 + \alpha_2, \beta)$$
$$V \sim \text{beta}(\alpha_1, \alpha_2)$$

Therefore,

$$\frac{\mathbf{Z}'\mathbf{Q}\mathbf{Z}}{\mathbf{Z}'(\mathbf{I} - \mathbf{H} - \mathbf{Q})\mathbf{Z} + \mathbf{Z}'\mathbf{Q}\mathbf{Z}} \sim \text{beta}\left(\frac{1}{2}, \frac{1}{2}(n-k-2)\right)$$

$\mathbf{Z}'\mathbf{Q}\mathbf{Z} \sim \mathcal{X}_1^2$ or $\mathbf{Z}'\mathbf{Q}\mathbf{Z} \sim \Gamma\left(\frac{1}{2}, 2\right)$. Also,

$$\mathbf{Z}'(\mathbf{I} - \mathbf{H} - \mathbf{Q})\mathbf{Z} \sim \mathcal{X}_{n-k-2}^2 \text{ or } \sim \Gamma\left(\frac{1}{2}(n-k-2), 2\right)$$

We can conclude that $\frac{r_i^2}{n-k-1} \sim \text{beta}\left(\frac{1}{2}, \frac{1}{2}(n-k-2)\right)$

§29.2 Externally Studentized Residual

Consider the ratio

$$t_i = \frac{e_i}{S_{e(i)}\sqrt{1 - h_{ii}}}$$

where $S_{e(i)}^2$ is the unbiased estimator of σ^2 after data point i is deleted from the data set. Notice that

$$\begin{aligned} t_i^2 &= \frac{e_i^2}{S_{e(i)}^2(1 - h_{ii})} \\ &= \frac{e_i^2(n-k-2)}{(n-k-2)S_{e(i)}^2(1 - h_{ii})} \end{aligned}$$

But $(n-k-2)S_{e(i)}^2 = (n-k-1)S_e^2 - \frac{e_i^2}{1-h_{ii}}$. Then,

$$\begin{aligned} t_i^2 &= \frac{e_i^2(n-k-2)}{\left[(n-k-1)S_e^2 - \frac{e_i^2}{1-h_{ii}}\right](1 - h_{ii})} \\ &= \frac{e_i^2(n-k-2)}{(n-k-1)S_e^2(1-h_{ii}) - e_i^2} \end{aligned}$$

Note: $r_i^2 = \frac{e_i^2}{S_e^2(1-h_{ii})} \implies e_i^2 = r_i^2 S_e^2(1 - h_{ii})$. Then,

$$\begin{aligned} t_i^2 &= \frac{r_i^2 S_e^2(1-h_{ii})(n-k-2)}{(n-k-1)S_e^2(1-h_{ii}) - r_i^2 S_e^2(1-h_{ii})} \\ &= \frac{r_i^2(n-k-2)}{n-k-1-r_i^2} = \frac{B}{1-B}(n-k-2) \end{aligned}$$

where $B = \frac{r_i^2}{n-k-1}$ with $\frac{r_i^2}{n-k-1} \sim \text{beta}\left(\frac{1}{2}, \frac{1}{2}(n-k-2)\right)$. From homework #10, exercise #5: If $B \sim \text{beta}\left(\frac{1}{2}\alpha, \frac{1}{2}\beta\right)$, then

$$\frac{\beta B}{\alpha(1-B)} \sim F_{\alpha,\beta}$$

Here $\alpha = 1$, $\beta = n - k - 2$. It follows that

$$t_i^2 = \frac{B}{1-B}(n-k-2) \sim F_{1,n-k-2}$$

and therefore

$$t_i = \frac{e_i}{S_{e(i)}\sqrt{1-h_{ii}}} \sim t_{n-k-2}$$

§29.3 A Note on Valuable Selection

Effect on the regression when predictors are removed from the model

a) Effect on $\hat{\boldsymbol{\beta}}$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

is the correct model or $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. Suppose we decided to use $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$. Then $\hat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y}$ and therefore

$$\begin{aligned} E\hat{\boldsymbol{\beta}}_1 &= \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2\right) \\ &= \boldsymbol{\beta}_1 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2 \end{aligned}$$

which is not unbiased.

b) Effect on the variance covariance matrix of $\hat{\boldsymbol{\beta}}$:

- From short regression: $\text{var}(\hat{\boldsymbol{\beta}}_1) = \sigma^2\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}$
- Long regression:

$$\text{var}\left(\hat{\boldsymbol{\beta}}_{1\cdot 2}\right) = \sigma^2\left(\mathbf{X}_1^{*'}\mathbf{X}_1^*\right)^{-1} = \sigma^2\left(\mathbf{X}_1'(\mathbf{I}-\mathbf{H})\mathbf{X}_1\right)^{-1} = \sigma^2\left(\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{H}_2\mathbf{X}_1\right)^{-1}$$

<u>Result</u>: If $\mathbf{A}^{-1} \geq \mathbf{B}^{-1}$ then $\mathbf{A} \leq \mathbf{B}$. Then

$$\begin{gathered} \left[\text{var}\left(\hat{\boldsymbol{\beta}}_1\right)\right]^{-1} - \left[\text{var}\left(\hat{\boldsymbol{\beta}}_{1\cdot 2}\right)\right]^{-1} = \frac{1}{\sigma^2}\mathbf{X}_1'\mathbf{H}_2\mathbf{X}_1 \geq 0 \\ \text{var}\left(\hat{\boldsymbol{\beta}}_1\right) \leq \text{var}\left(\hat{\boldsymbol{\beta}}_{1\cdot 2}\right) \end{gathered}$$