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

List of Theorems

List of Definitions

§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

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

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

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

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 $\mu :\ \mu \in \overline{Y} \pm t_{\frac{\alpha}{2};n-1}\frac{S}{\sqrt{n}}$.

§2 Lec 2: Sep 29, 2021

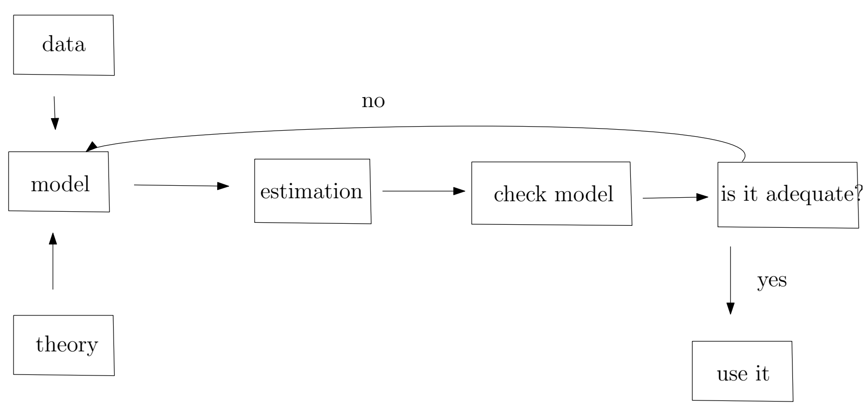
§2.1 Linear Regression

Consider a simple regression model

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ \text{or } Y_i &= \beta_1 X_i + \varepsilon_i \end{aligned}$$

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):

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

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

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

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

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

<u>Properties of $\hat{\beta}_1$</u>:

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

For the variance,

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

<u>Properties of $\hat{\beta}_0$</u>:

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

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

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

Thus,

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

and we know that

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

and

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

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

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

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 $\operatorname{var}(b_1) \geq \operatorname{var}(\hat{\beta}_1)$ 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?

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

<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

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

Therefore,

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

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 $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$

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

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,

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

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,

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

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 $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. Then, $Y_i \sim N\left(\beta_0 + \beta_1 X_i, \sigma\right)$. 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

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

<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

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

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

§5.2 A Note on Gamma Distribution

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

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

where

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

is the Gamma function.
<u>Property</u>:

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

If α is an integer, then

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

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

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

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 $0 \le R^2 \le 1$. We have

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

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

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

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

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

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,

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

<u>Pivot Method</u>:

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

and after some manipulation we get

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

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,

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

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,

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

and we know

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

Therefore, we can form the F statistics

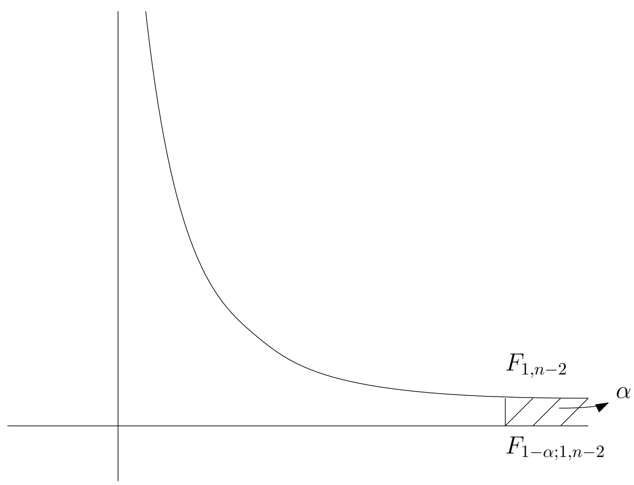
$$\frac{\frac{\hat{\beta}_1^2 \sum(x_i-\overline{x})^2}{\sigma^2}/1}{\frac{(n-2)S_e^2}{\sigma^2}/(n-2)} = \frac{\hat{\beta}_1^2 \sum(x_i-\overline{x})^2}{S_e^2} \sim F_{1,\,n-2}$$

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,

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



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 $\hat{\beta}_0$,

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

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

$$Y_i = \beta_1 X_i + \varepsilon_i$$
$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

We know

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

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):

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

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:

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

So,

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

Notice that

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

So,

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

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.,

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

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,

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

<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,

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

§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

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

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,

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

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

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

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.

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

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

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

$$\begin{aligned}\mathbf{x_1}\mathbf{x_1} &= \sum x_{i1}^2 \\ \mathbf{x_1}'\mathbf{x_2} &= \sum x_{i1}x_{i2}\end{aligned}$$

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

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

Model:

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

Then,

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

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
$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$$

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
$$\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{BCA}) = \operatorname{tr}(\mathbf{CAB}) \neq \operatorname{tr}(\mathbf{BAC})$$

4. $\mathbf{HX} = \mathbf{X}$ or $\mathbf{H}(\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_k}) = (\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_k})$

Residuals:

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

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

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

Notice that the error sum of squares

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

or

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

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.

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

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($\mathbf{b}$).

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

Note:

$$\mathrm{var}(\mathbf{AY}) = \mathbf{A\Sigma A}'$$

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

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

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

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

Aside Note:

$$\mathrm{var}\left(\mathbf{aY}'\right) = \mathbf{a}'\mathbf{\Sigma a} > 0$$

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

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

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

§12.2 Gauss-Markov Theorem For a Linear Combination

We have

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

or

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

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

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

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

Special Case:

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

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

§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

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

Joint MGF: 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

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

Notice that

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

§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

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

□

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

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

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

Theorem 13.2

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

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

Note that

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

Then,

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

Proof. Use the above theorem

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

Then,

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

□

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

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

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 $\iff$ $\operatorname{cov}(\mathbf{Q}_1, \mathbf{Q}_2) = \mathbf{0}$.

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

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

Theorem 13.4

We have

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

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

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

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

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

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

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

<u>*Note*</u>:

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

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 $\mathbf{e}$.

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

<u>*Note*</u>: $\mathbf{HX} = \mathbf{X}$.

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

For $\mathbf{e}$,

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

§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

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

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

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

Then,

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}_{12} + \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_1'\mathbf{Y} \tag{1}$$
$$\mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}}_{12} + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\mathbf{Y} \tag{2}$$

From (1),

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}_{12} = \mathbf{X}_1'\mathbf{Y} - \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21}$$

So,

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

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

$$\mathbf{X}_2'\mathbf{X}_1\left[\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y} - \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21}\right] + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\mathbf{Y}$$
$$\mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y} - \mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} + \left(\mathbf{X}_2'\mathbf{X}_2\right)\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2{}'\mathbf{Y}$$

Then,

$$\left(\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21}\right) - \mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\mathbf{Y} - \mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{Y}$$
$$\mathbf{X}_2'\left[\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\left[\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{Y}$$
$$\mathbf{X}_2'\left[\mathbf{I} - \mathbf{H}_1\right]\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\left[\mathbf{I} - \mathbf{H}_1\right]\mathbf{Y}$$
$$\mathbf{X}_2'\left(\mathbf{I} - \mathbf{H}_1\right)\left(\mathbf{I} - \mathbf{H}_1\right)\mathbf{X}_2\hat{\boldsymbol{\beta}}_{21} = \mathbf{X}_2'\left(\mathbf{I} - \mathbf{H}_1\right)\left(\mathbf{I} - \mathbf{H}_1\right)\mathbf{Y}$$
$$\left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_2\right]'\left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_2\right]\hat{\boldsymbol{\beta}}_{21} = \left[\left(\mathbf{I} - \mathbf{H}\right)\mathbf{X}_2\right]'\left[\left(\mathbf{I} - \mathbf{H}_1\right)\mathbf{Y}\right]$$

Note:

$$\left(\mathbf{I} - \mathbf{H}_1\right)\mathbf{Y} = \mathbf{Y}^*$$

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,

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

So,

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

and thus

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

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

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