4. Classical Normal Regression Model (CNLRM)

STATA WORKSHOP
THU 22 FEB 17.00 ONWARDS
@ COMPUTER LAB

We know that the classical theory of statistical inference consists of:

1. Estimation

We have covered this topic since we were able to estimate the parameters β_1, β_2, and σ^2 by using the method of OLS.

We also proved that these estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ satisfy several desirable statistical properties, such as unbiasedness, minimum variance, and linearity (BLUE property).

However, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ change their values from sample to sample. The following tables show the two different sets of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}$ depending on the two different sample data.

SO FAR,

(1) we got $\hat{\beta}_1$, $\hat{\beta}_2$, $var(\hat{\beta}_1)$, $var(\hat{\beta}_2)$, $se(\hat{\beta}_1)$, $se(\hat{\beta}_2)$, $cov(\hat{\beta}_1, \hat{\beta}_2)$

(2) WE PROVE BLUE PROPERTY OF OUR OLS ESTIMATORS: $\hat{\beta}_1$, $\hat{\beta}_2$

Table 4.1: Estimating the expenditure of the household with income

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$ (u_i)	Residual squared $(Y-\hat{Y})^2$ (u_i^2)
1	390	500	394.95	-4.95	24.53
2	425	600	454.24	-29.24	854.87
3	560	700	513.52	46.48	2160.04
4	575	800	572.81	2.19	4.80
5	630	900	632.10	-2.10	4.39
6	679	1000	691.38	-12.38	153.29
Sum	3259	4500	0	0	3201.90

CHAPTER 4's HIGHLIGHT: NORMALITY ASSUMPTION OF OUR u_i

is RD — MEAN; SPREAD → σ^2, σ

If we use this sample data. We can estimate:

$\beta_1 \leftarrow \hat{\beta}_1 = 98.524$ ✓

$\beta_2 \leftarrow \hat{\beta}_2 = 0.593$ ✓

UNBIASED ESTIMATOR FOR σ_u^2

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{3201.90}{6-2} = 800.476$$

Why do we divide by n-2 instead of n or n-1?

Table 4.2: Estimating the expenditure of the household with income with another sample data

Family (i)	Actual Y_i	Income X_i	Regression Estimate $\hat{Y}$	Residual $Y-\hat{Y}$	Residual squared $(Y-\hat{Y})^2$
1	360	500	325.71	64.29	4132.65
2	390	600	406.43	18.57	344.90
3	440	700	487.14	72.86	5308.16
4	575	800	567.86	7.14	51.02
5	670	900	648.57	-18.57	344.90
6	730	1000	729.29	-50.29	2528.65

2	390	600	406.43	18.57	344.90
3	440	700	487.14	72.86	5308.16
4	575	800	567.86	7.14	51.02
5	670	900	648.57	-18.57	344.90
6	730	1000	729.29	-50.29	2528.65
Sum	3165	4500	0	0	12710.29

If we use this sample data. We can estimate:

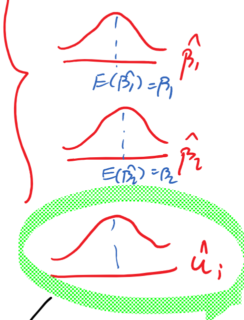
$$\hat{\beta}_1 = -77.857$$

$$\hat{\beta}_2 = 0.807$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{12710.29}{6-2} = 3177.571$$

TRUE
BUT UNKOWN
β_1
β_2

SAMPLE1 : $\hat{\beta}_1, \hat{\beta}_2$
SAMPLE2 : $\hat{\beta}_1, \hat{\beta}_2$
⋮
SAMPLE n : $\hat{\beta}_1, \hat{\beta}_2$



- What type of distribution?
- Characteristics / patterns of $\hat{u}_i$

From the example, you can easily see that these estimators are **RANDOM VARIABLES**. Therefore, we need to learn another part of statistical inference which is called **Hypothesis Testing.**

2 **Hypothesis Testing**

The main objective is to find out how close of $\hat{\beta}_1$ and $\hat{\beta}_2$ to the true β_1 and the true β_2, respectively. Also, we would like to see how close of $\hat{\sigma}^2$ compared to the true σ^2.

$H_0: \beta_2 = 0$ $H_0: \beta_1 = 0$
$H_1: \beta_2 \neq 0$ $H_1: \beta_2 \neq 0$

To achieve this goal, we need to know the probability distributions of $\hat{\beta}_1$, $\hat{\beta}_2$,and $\hat{\sigma}^2$. Consider the estimator of β_2:

$$\hat{\beta}_2 = \sum k_i Y_i$$

ex: $\log(\text{wage}) = \beta_1 + \beta_2 \log(\text{EXPERIENCE}) + \beta_3 \log(\text{EDUCATION}) + u_i$

We can write the above equation as:

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i)$$

$H_0: \beta_2 = 0$ $H_0: \beta_3 = 0$
$H_1: \beta_2 \neq 0$ $H_1: \beta_3 \neq 0$

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$

From this equation, the probability distribution of $\hat{\beta}_2$ will depend on the assumption made about the probability distribution of u_i

4.1 The Normality Assumption for u_i

In the classical normal linear regression model (CNLRM), we assume that each u_i is distributed normally :

$$u_i \sim N(0, \sigma^2)$$

u_i is normally distributed with mean equals to 0 and variance equals to σ^2

where
Mean:

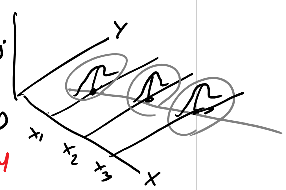
$$\boxed{E(u_i) = 0}$$

Variance:

$$\boxed{E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2}$$ constant variance

$$\boxed{cov(u_i, u_j) = E\{[u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i U_j) = 0}$$

no correation between u_i and u_j



Therefore,

IN EVIEWS, WE HAVE A TEST CALLED "NORMALITY TEST" ON ERROR TERM

$$u_i \sim N(0, \sigma^2)$$

Also, u_i and u_j are not only uncorrelated but also independently distributed.

we can then write the above equation as:

u_i is normally and independently distributed

$$\boxed{u_i \sim NID(0, \sigma^2)}$$

where NID stands for normally and independently distributed.

4.2 Properties of OLS estimators under the normality assumption of u_i

1. They are unbiased. ✓ : $E(\hat{\beta}_1) = \beta_1$ & $E(\hat{\beta}_2) = \beta_2$
2. They have minimum variance.
3. By 1+2 properties, they are minimum-variance unbiased, or $\boxed{\text{efficient estimators.}}$ $\hat{\beta}_1$ & $\hat{\beta}_2$
4. $\hat{\beta}_1$ is normally distributed with:

$$\text{Mean:} E(\hat{\beta}_1) = \beta_1$$

$$\text{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

Therefore,

$$\boxed{\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})}$$

~~By the properties of the normal distribution, we can:~~

5. $\hat{\beta}_2$ is normally distributed with

$$\text{Mean:} E(\hat{\beta}_2) = \beta_2$$

$$\mathrm{var}(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2}$$

or more compactly

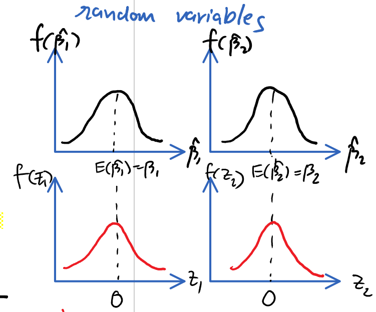
$$\boxed{\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})}$$

then we can define the standard normal distribution as

$$Z_1 = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \qquad Z_2 = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}}$$

$$Z_1 \sim N(0,1) \qquad Z_2 \sim N(0,1)$$

Z_1 & Z_2 are standardized random variables



standardized normal distribution

6. $(n-2)(\hat{\sigma}^2/\sigma^2)$ is distributed as the χ^2 (chi-square) distribution with (n-2) df.

7. $(\hat{\beta}_1, \hat{\beta}_2)$ are distributed independently of $\hat{\sigma}^2$.

σ^2 variance of u_i

8. $\hat{\beta}_1$ and $\hat{\beta}_2$ have the minimum variance in the entire class of unbiased estimators, whether linear or not.

it say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are BUE

9. we can find out the probability distribution of Y_i as following:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

since $u_i \sim N(0, \sigma^2)$

then

$E(Y) = ?$

$\mathrm{var}(Y) = ?$

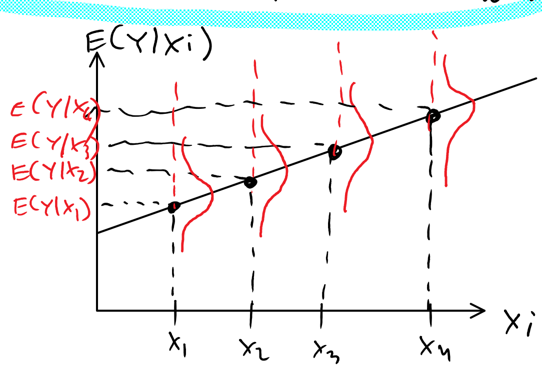
$E(Y) = \beta_1 + \beta_2 X_i$ (since $E(u_i) = 0$)

$\mathrm{var}(Y_i) = \mathrm{var}[u_i]$!!! $= \sigma_u^2$

So $Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma_u^2)$

* W/O NORMALITY ASSUMPTION : $\hat{\beta}_1$ & $\hat{\beta}_2$ ARE BLUE

W/ NORMALITY ASSUMPTION : $\hat{\beta}_1$ & $\hat{\beta}_2$ ARE BUE.



$E(Y|X_i) = \beta_1 + \beta_2 X_i$

• $= E(Y|X_i)$