Econometrics

Course Calendar	
Week	Main Content
Week 1	**Introduction to Simple Regression**
Week 2	**Simple Regression**
Week 3	**Simple Regression: r^2 & Hands-on-Exercise**
Week 4	**Central Limit Theorem, Probability and Probability Density Function (PDF)**
Week 5	**Hypothesis Testing: Basics**
Week 6	**Simple Regression: Testing of Hypothesis**

Econometrics
Lecture 3: r^2 & Simple Regression
Hands on Exercise in Excel

Geetha Rani Prakasam, Ph.D.

Professor

Recap

- Simple Regression- Problem of Estimation
- Lease Square Criterion – OLS
- Assumptions of OLS
- Guass Markov theorem – BLUE
- Precision of the Estimated Parameters $\beta\hat{}_2$ & $\beta\hat{}_1$ given by their variance and SEs

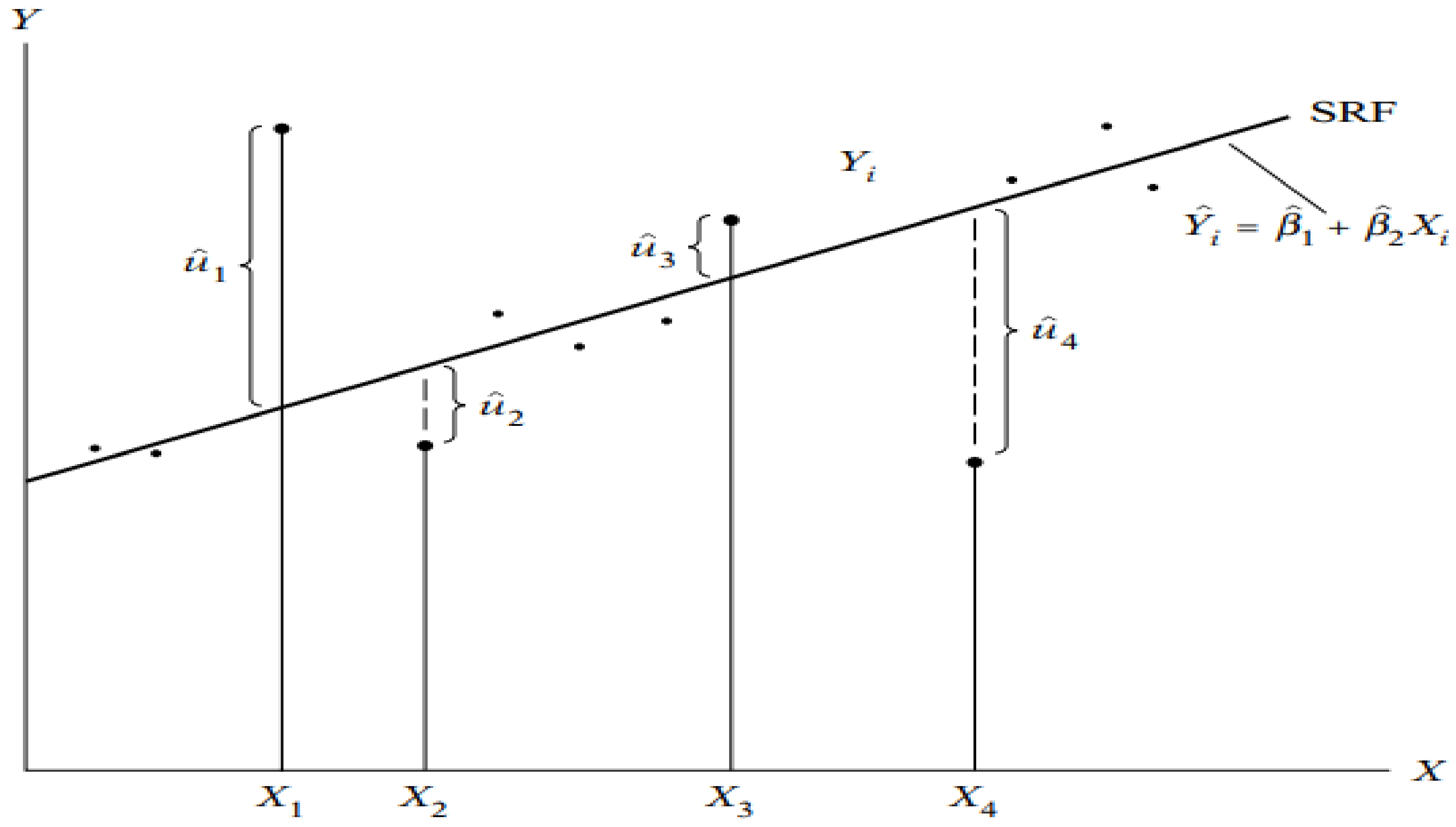
outline

- Co-efficient of determination r^2
- Recollecting the formula and estimating the simple regression parameters, intercept and slope
- Recollecting the formula and estimating the variance and standard errors of the simple regression parameters, intercept and slope
- The estimation done in excel with a numerical example
- Estimating r^2 in excel
- Normality of U

3-5. The coefficient of determination r^2: A measure of "Goodness of fit"

- How "well" the sample regression line fits the data
- Recall Fig 3.1: fitted line of SRF - there will be some positive $\hat{u}_i$ *and some negative* $\hat{u}_i$.

Fig 3.1 Least-squares criterion.



Source: Basic Econometrics by Damodar Gujarati, Page 59.

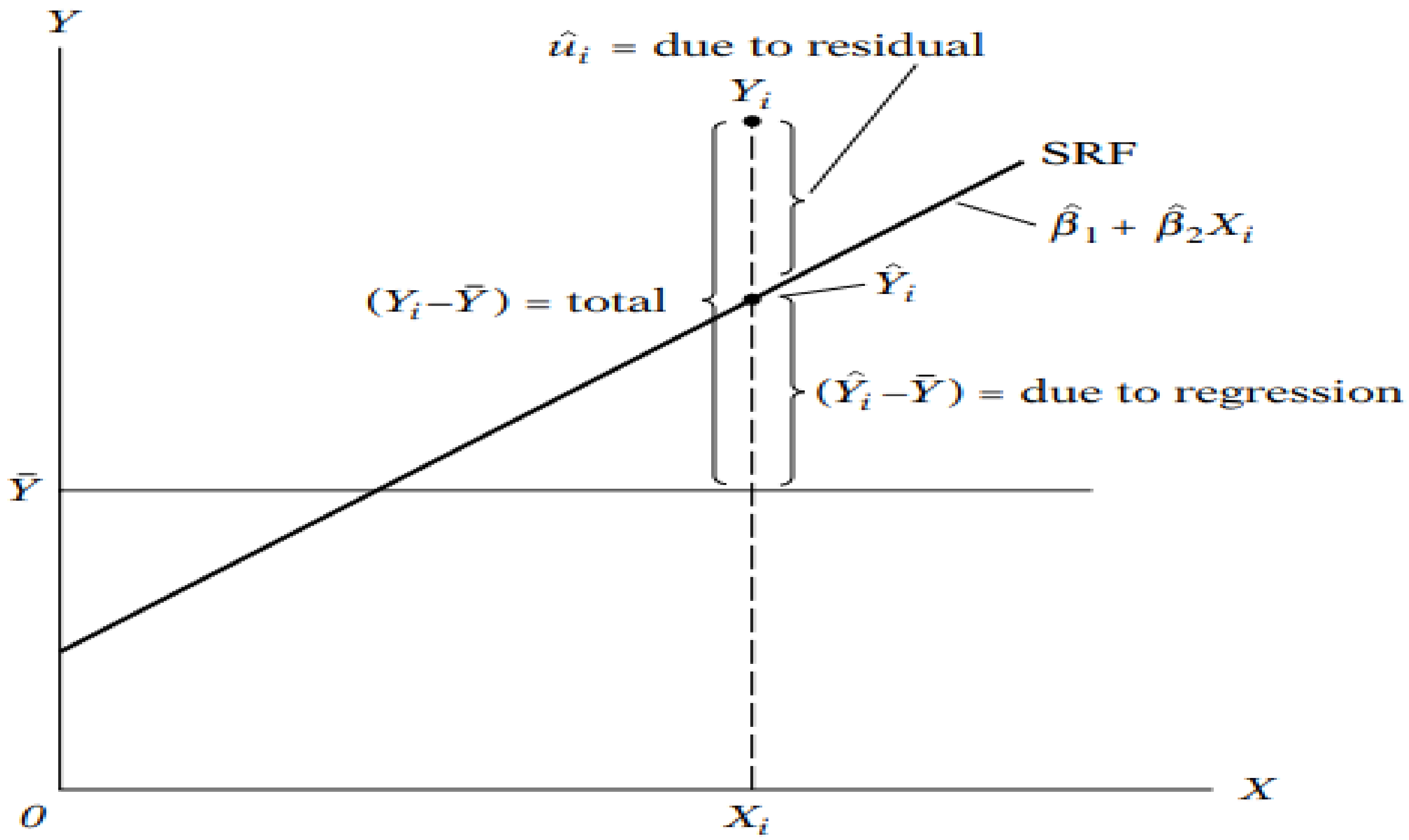
The coefficient of determination r^2

- What we hope for is that these residuals around the regression line are as small as possible.
- The **coefficient of determination** r^2 **(two-variable case) or** R^2 (multiple regression) is a summary measure that tells how well the sample regression line fits the data.
- To compute this r^2, we proceed as follows: Recall that
- $Y_i = \hat{Y}_i - \hat{u}_i$ -----------(2.6.3)
- or in the deviation form: $y_i = \hat{y}_i - \hat{u}_i$ -----------(3.5.1)

3-5. The coefficient of determination r^2: A measure of "Goodness of fit"

- Squaring Equation 3.5.1 on both sides and summing over the sample, we obtain
- $\sum y_i^2 = \sum \hat{y}_i^2 + \sum u_i^2 + 2\sum \hat{y}_i \hat{u}_l$
- $= \sum \hat{y}_i^2 + \sum u_i^2$
- $= \hat{\beta}_2 \sum x_i^2 + \sum u_i^2$ -----------(3.5.2)
- TSS = ESS + RSS ----------------(3.5.3)
- shows that the total variation in the observed Y values about their mean value can be partitioned into two parts, one attributable to the regression line and the other to random forces because not all actual Y observations lie on the fitted line. Geometrically, we have Figure 3.9.

Fig 3.9: Breakdown of the variation of Yi into two components



Source: Basic Econometrics by Damodar Gujarati, Page 83.

3-5. The coefficient of determination r^2: A measure of "Goodness of fit"

- Fig 3.9 - Breakdown of the variation of Yi into two components
- The r^2 can be interpreted as the fraction of sample variation in *y* that is explained by *x*

Note 1: in the simple regression analysis, r^2 is the same as the square of the correlation coefficient, *r*.

$$= \frac{\sum\left(Y_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} + \frac{\sum \hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

3-5. The coefficient of determination r^2: A measure of "Goodness of fit"

- Note 2:
- TSS = $\sum y_i^2$ = Total Sum of Squares
- ESS = $\sum \hat{Y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$ = Explained Sum of Squares
- RSS = $\sum \hat{u}_i^2$ = Residual Sum of Squares
- Now dividing Equation 3.5.3 by TSS on both sides, we obtain

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \text{ ; or}$$

$$1 = r^2 + \frac{RSS}{TSS} \text{ ;} \quad \text{or} \quad r^2 = 1 - \frac{RSS}{TSS} \qquad \text{----(3.5.4)}$$

3-5. The coefficient of determination r^2: A measure of "Goodness of fit"

- We now define $r^2 = ESS/TSS$

•r^2 is coefficient of determination, it measures the proportion or percentage of the total variation in Y explained by the regression Model

- Two properties of r^2 *may be noted:*
- It is a nonnegative quantity
- $0 \leq r^2 \leq 1$;
- A quantity closely related to but conceptually very much different from r^2 *is the* **coefficient of correlation;**
- $r = \pm\sqrt{r^2}$ is sample correlation coefficient

Summary and Conclusions

- The basic framework of regression analysis is the **CLRM.**
- The CLRM is based on a set of assumptions.
- Based on these assumptions, the least-squares estimators take on certain properties summarized in the Gauss–Markov theorem, that in the class of linear unbiased estimators, the least-squares estimators have minimum variance.
- In short, they are BLUE.

Summary and Conclusions

- The precision of OLS estimators is measured by their **standard errors:** we shall see how the standard errors enable one to draw inferences on the population parameters, the β coefficients – Test of Significance.
- The overall goodness of fit of the regression model is measured by the **coefficient of determination, r^2.**
- **It tells what proportion of the variation in the dependent variable,** or regressand, is explained by the explanatory variable, or regressor. This r^2 lies between 0 and 1; the closer it is to 1, the better is the fit.

Estimating the Regression Parameters

- $\beta\hat{}_2 = \sum x_i\, y_i / \sum x_i^2$ -------------(3.1.6)
  - Small / lower case letter denote the deviation from mean values
- $\beta\hat{}_1 = \bar{y} - \beta\hat{}_2\, \bar{x}$ -------------(3.1.7)

- Estimating the regression parameters $\beta\hat{}_2$ & $\beta\hat{}_1$ Using MSExcel
- Steps to be completed before estimating $\beta\hat{}_2$
- $\sum x_i\, y_i$ → the product of the deviations from the mean values of $\bar{x}$ and $\bar{y}$
- $\sum x_i^2$ → the product of the deviations from the mean values of $\bar{x}$

Data on the Following Table on Y and X Variables, N=15

PRICE OF GOLD (Y)	X= Consumer Price Index(CPI) 1[982-84 = 100]
147.98	60.6
193.44	65.2
307.62	72.6
612.51	82.4
459.61	90.9
376.01	96.5
423.83	99.6
360.29	103.9
317.3	107.6
367.87	109.6
446.5	113.6
436.93	118.3
381.28	124
384.08	130.7
362.04	136.2

	PRICE OF GOLD (Y)	X= Consumer Price Index(CPI) 1[982-84 = 100]	Deviation from x̄ = x	Deviation from ȳ = Y	xy	x^2
	147.98	60.6	-40.18	-223.84	8993.86	1614.43
	193.44	65.2	-35.58	-178.38	6346.74	1265.94
	307.62	72.6	-28.18	-64.20	1809.14	794.11
	612.51	82.4	-18.38	240.69	-4423.89	337.82
	459.61	90.9	-9.88	87.79	-867.37	97.61
	376.01	96.5	-4.28	4.19	-17.94	18.32
	423.83	99.6	-1.18	52.01	-61.37	1.39
	360.29	103.9	3.12	-11.53	-35.97	9.73
	317.3	107.6	6.82	-54.52	-371.82	46.51
	367.87	109.6	8.82	-3.95	-34.83	77.79
	446.5	113.6	12.82	74.68	957.41	164.35
	436.93	118.3	17.52	65.11	1140.74	306.95
	381.28	124	23.22	9.46	219.68	539.17
	384.08	130.7	29.92	12.26	366.84	895.21
	362.04	136.2	35.42	-9.78	-346.38	1254.58
TOTAL (**Σ**)	5577.29	1511.7	0.00	0.00	13674.81	7423.92
AVERAGE	371.82	100.78				

	PRICE OF GOLD (Y)	X= Consumer Price Index(CPI) 1[982-84 = 100]	Deviation from x̄; x̄ - Xi = xi	Deviation from ȳ; ȳ -Yi= yi	xiyi	xi^2
	(1)	(2)	(3)	(4)	(5)	(6)

- Deviation from x̄ means the difference between x̄ from each of the Xi denoted by lower case *xi* in column 3 and is calculated by =C3-C19;
- =C3-C19 - "$" is used to fix the cell – which contains the x̄
- Deviation from ȳ means the difference between ȳ from each of the Yi and noted as lower case *yi* in column 4 & is calculated by =B3-B19;
- =B3-B19 – "$" is used to fix the cell – that contains the ȳ
- xiyi = multiplication of values in column 3 and 4
- xi^2 = square the values in column 3
- Apply these obtained values in the formula for $\beta\hat{}_2 = \sum x_i\, y_i / \sum x_i^2$

Slope co-efficient: $\beta\hat{}_2 = \sum x_i y_i / \sum x_i^2$

- $\beta\hat{}_2 = \sum x_i y_i / \sum x_i^2$
- $\sum x_i y_i = 13674.81$
- $\sum x_i^2 = 7423.92$
- 13674.81 /7423.92 = **1.841993**

Estimating Intercept co-efficient: $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$

- $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$
- $\bar{y} = 371.82$
- $\bar{x} = 100.78$
- $\hat{\beta}_2 =$ **1.841993**
- $\hat{\beta}_1 =$ 371.82 - **1.841993 *** 100.78
- $\hat{\beta}_1 =$ **186.183**

Variance and SE of least-squares estimates

- $\text{var}(\hat{\beta}_2) = \sigma^2 / \sum x^2_i$ (3.3.1)
- $\text{se}(\hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_2)}$ (3.3.2)
- $\text{var}(\hat{\beta}_1) = \sigma^2 \sum X^2_i / n \sum x^2_i$ (3.3.3)
- $\text{se}(\hat{\beta}_1) = \sqrt{\text{Var}(\hat{\beta}_1)}$ (3.3.4)
- Where
- $\hat{\sigma}^2 = \sum \hat{u}^2_i / (n - 2)$

Reference

Chapter 3:Two-Variable Regression Model: The Problem of Estimation in Basic Econometrics by Domodar Gujarati

Chapter 4: THE NORMALITY ASSUMPTION:

Classical Normal Linear Regression Model

(CNLRM), in Basic Econometrics by Domodar Gujarati

OLS_Reg_Ex_March_30_2022_Stud - Excel

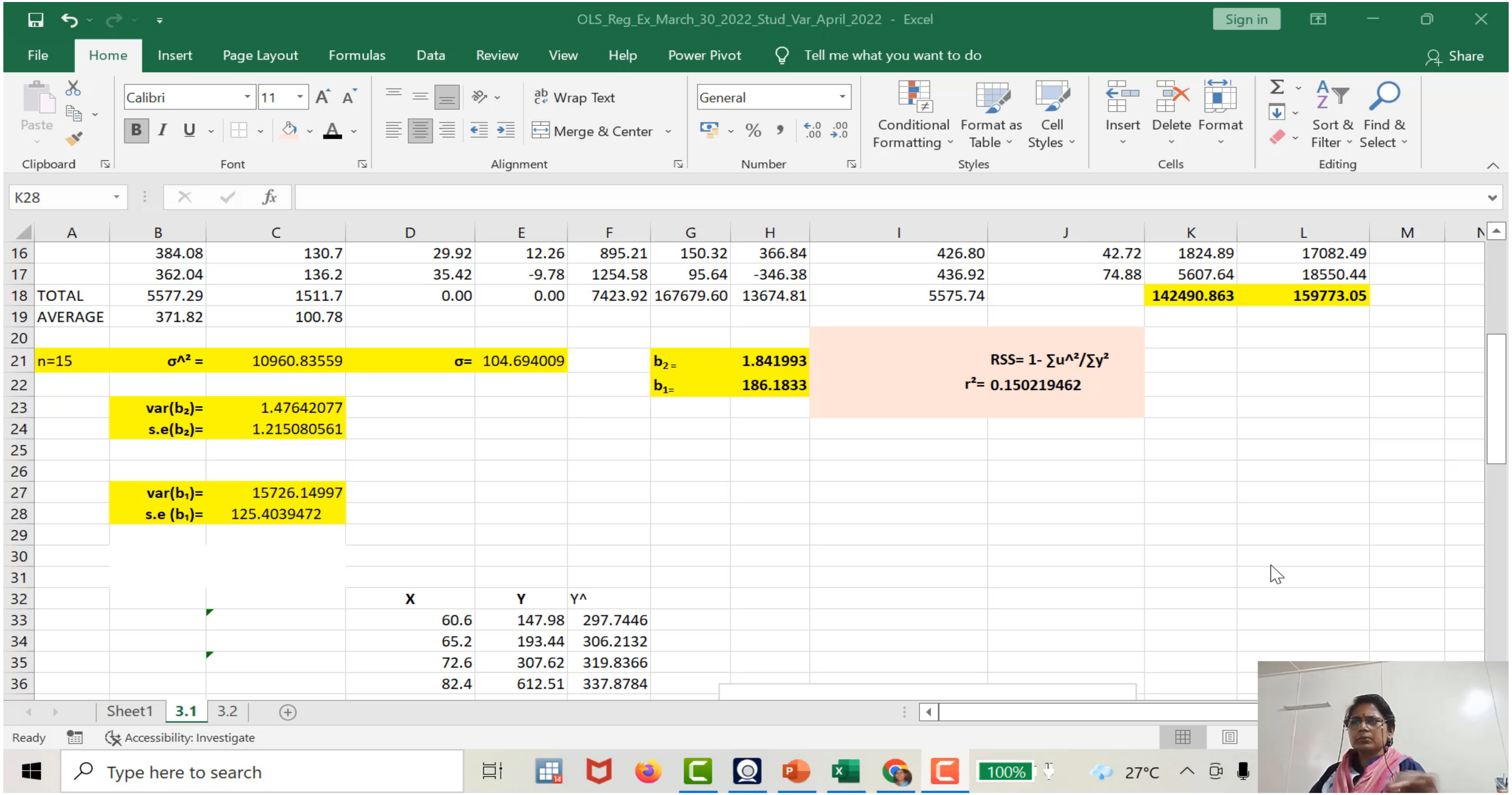
	A	B	C	D	E	F	G	H	I	J	K
1											
2		PRICE OF GOLD	X= Consumer Price Index(CPI) 1[982-84 = 100]	xi	yi	y²	x²	xiyi	Y^ = 186.18 + 1.841 X	u^	u²
3		147.98	60.6								
4		193.44	65.2								
5		307.62	72.6								
6		612.51	82.4								
7		459.61	90.9								
8		376.01	96.5								
9		423.83	99.6								
10		360.29	103.9								
11		317.3	107.6								
12		367.87	109.6								
13		446.5	113.6								
14		436.93	118.3								
15		381.28	124								
16		384.08	130.7								
17		362.04	136.2								

OLS_Reg_Ex_March_30_2022_Stud_Var_April_2022 - Excel

G40

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		PRICE OF GOLD	X= Consumer Price Index(CPI) 1[982-84 = 100]	x	y	x^2	y^2	xy	Y^ = 186.18 + 1.841 X	u^	u^2	X^2
3		147.98	60.6	-40.18	-223.84	1614.43	50104.05	8993.86	297.74	149.76	22429.44	3672.36
4		193.44	65.2	-35.58	-178.38	1265.94	31819.19	6346.74	306.21	112.77	12717.79	4251.04
5		307.62	72.6	-28.18	-64.20	794.11	4121.55	1809.14	319.84	12.22	149.25	5270.76
6		612.51	82.4	-18.38	240.69	337.82	57932.00	-4423.89	337.88	-274.63	75422.52	6789.76
7		459.61	90.9	-9.88	87.79	97.61	7707.20	-867.37	353.53	-106.08	11253.62	8262.81
8		376.01	96.5	-4.28	4.19	18.32	17.56	-17.94	363.84	-12.17	148.19	9312.25
9		423.83	99.6	-1.18	52.01	1.39	2705.11	-61.37	369.54	-54.29	2947.01	9920.16
10		360.29	103.9	3.12	-11.53	9.73	132.93	-35.97	377.46	17.17	294.81	10795.21
11		317.3	107.6	6.82	-54.52	46.51	2972.36	-371.82	384.27	66.97	4485.20	11577.76
12		367.87	109.6	8.82	-3.95	77.79	15.60	-34.83	387.95	20.08	403.35	12012.16
13		446.5	113.6	12.82	74.68	164.35	5577.20	957.41	395.32	-51.18	2619.64	12904.96
14		436.93	118.3	17.52	65.11	306.95	4239.40	1140.74	403.97	-32.96	1086.34	13994.89
15		381.28	124	23.22	9.46	539.17	89.50	219.68	414.46	33.18	1101.18	15376
16		384.08	130.7	29.92	12.26	895.21	150.32	366.84	426.80	42.72	1824.89	17082.49
17		362.04	136.2	35.42	-9.78	1254.58	95.64	-346.38	436.92	74.88	5607.64	18550.44
18	TOTAL	5577.29	1511.7	0.00	0.00	7423.92	167679.60	13674.81	5575.74		142490.863	159773.05
19	AVERAGE	371.82	100.78									

Sheet1 | 3.1 | 3.2



	A	B	C	D	E	F	G	H	I	J	K	L
16		384.08	130.7	29.92	12.26	895.21	150.32	366.84	426.80	42.72	1824.89	17082.49
17		362.04	136.2	35.42	-9.78	1254.58	95.64	-346.38	436.92	74.88	5607.64	18550.44
18	TOTAL	5577.29	1511.7	0.00	0.00	7423.92	167679.60	13674.81	5575.74		142490.863	159773.05
19	AVERAGE	371.82	100.78									
20												
21	n=15	σ^2 =	10960.83559	σ=	104.694009		b_2 =	1.841993	RSS= 1- ∑u^2/∑y^2			
22							b_1 =	186.1833	r^2= 0.150219462			
23		var(b_2)=	1.47642077									
24		s.e(b_2)=	1.215080561									
25												
26												
27		var(b_1)=	15726.14997									
28		s.e (b_1)=	125.4039472									
29												
30												
31												
32				X	Y	Y^						
33				60.6	147.98	297.7446						
34				65.2	193.44	306.2132						
35				72.6	307.62	319.8366						
36				82.4	612.51	337.8784						

Interpretation of Regression Coefficients

- Y^ = 186.18 + 1.841 X

  (125.404) (1.215)

- The intercept coefficient in the first row is 186.18 indicate when X takes value zero, what is the price of Gold Price.
- The slope coefficient (parameter estimate) is 1.84.
- So, for every unit increase in CPI (X), a 1.84 unit increase in Gold Price (Y) is predicted.
- The second row represent the standard errors (se) associated with the intercept and slope coefficients.
- The standard error is used for testing whether the parameter is significantly different from 0 by dividing the parameter estimate by the standard error to obtain a t value (will learn later).

THE NORMALITY ASSUMPTION:
Classical Normal Linear Regression Model (CNLRM)

- Classical theory of statistical inference consists of two branches, namely, **estimation and hypothesis testing.**
- So far, we covered estimation of the parameters of the (two-variable) linear regression model. Using the method of OLS we were able to estimate the parameters β_1 , β_2 , and σ2.
- Under the assumptions of the CLRM, we show that the estimators of these parameters, $\beta\hat{}_1$, $\beta\hat{}_2$ and σˆ2, satisfy several desirable statistical properties, such as unbiasedness, minimum variance, etc (**the BLUE property**)
- Since these are estimators, their values will change from sample to sample. Therefore, these estimators are random variables.

Classical Normal Linear Regression Model

- But estimation is half the battle. Hypothesis testing is the other half.
- Our objective is not only to estimate the SRF, but also to use it to draw inferences about the PRF.
- Thus, we would like to find out how close $\hat{\beta}_1$ *is to the true* β_1 or how close $\hat{\sigma}^2$ *is to the true* σ^2.
- Therefore, since $\hat{\beta}_1$, $\hat{\beta}_2$, *and* $\hat{\sigma}^2$ *are random variables, we need to find out their probability* distributions, for without that knowledge we will not be able to relate them to their true values.

4.1 The Probability Distribution of Disturbances *ui*

- Since the method of OLS does not make any assumption about the probabilistic nature of ui , it is of little help for the purpose of drawing inferences about the PRF from the SRF, the Gauss–Markov theorem notwithstanding.
- This void can be filled if we are willing to assume that the u's follow some probability distribution. In the regression context it is usually assumed that the u's follow the normal distribution.
- Adding the normality assumption for ui to the assumptions of the classical linear regression model (CLRM) discussed earlier, we obtain what is known as the **classical normal linear regression model (CNLRM).**

4-2.The normality assumption

- CNLR assumes that each u_i is distributed normally $u_i \sim N(0, \sigma^2)$ with:

  Mean = $E(u_i) = 0$ (Assn. 3)--- (4.2.1)

  Variance = $E(u^2_i) = \sigma^2$ (Assn. 4)-----(4.2.2)

  $Cov(u_i, u_j) = E(u_i, u_j) = 0$ (i#j) (Assn. 5) (4.2.3)

- The assumptions given above can be more compactly stated as $ui \sim N(0, \sigma 2)$ -------- (4.2.4)
- Note: For two normally distributed variables, the zero covariance or correlation means independence of them, so u_i and u_j are not only uncorrelated but also independently distributed. Therefore
- **$u_i \sim NID(0, \sigma^2)$** -------(4.2.5)
- *is Normal and Independently Distributed*

4-2. Why the normality assumption?

1. With a few exceptions, the distribution of sum of a large number of independent and identically distributed random variables tends to a normal distribution as the number of such variables increases indefinitely. It is the **Central Limit Theorem (CLT)** that provides a theoretical justification for the assumption of normality of *ui* .
2. A variant of the **CLT** states that, if the number of variables is not very large or they are not strictly independent, their sum may still be normally distributed
3. Under the normality assumption for u_i , the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are also normally distributed.

4-2. Why the normality assumption?

(4) The normal distribution is a comparatively simple distribution involving only two parameters (mean and variance)

(5) If we are dealing with a small, or finite, sample size, (less than 100 observations), the normality assumption assumes a critical role. It not only helps us to derive the exact probability distributions of OLS estimators but also enables us to use the *t, F, and* $\chi 2$ statistical tests for regression models.

(6) Finally, in *large samples, t and F statistics have approximately the t and F probability* distributions so that the *t and F tests that are based on the assumption that the error term* is normally distributed can still be applied validly.

4-3. Properties of OLS estimators under the normality assumption

- **With the normality assumption the OLS estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}^2$ have the following properties:**

1. They are unbiased

2. They have minimum variance. Combined 1 and 2, they are efficient estimators.

3. Consistency, that is, as the sample size increases indefinitely, the estimators converge to their true population values.

4-3. Properties of OLS estimators under the normality assumption

4. β_1 *(being a linear function of ui) is normally distributed with*

Mean : $E(\beta\hat{}_1) = \beta_1$ ------------ (4.3.1)

$var(\beta\hat{}_1) = \sigma^2 \sum X^2_i / n \sum x^2_i$ ------------ (3.3.3) 4.3.2)

Or more compactly, it can be written as

$\beta\hat{}_1 \sim N(\beta_1, \sigma\hat{}_{\beta 1}{}^2)$; ------------ (4.3.3)

Then by the properties of the normal distribution, the variable *Z, which is defined as*

$E(\beta\hat{}_2) = \beta_2$ ------------ (4.3.4)

$var(\beta\hat{}_2) = \sigma^2 / \sum x^2_i$ -----------(3.3.1) (4.3.5)

$Z = (\beta\hat{}_1 - \beta_1) / \sigma\hat{}_{\beta 1}$ is $\sim N(0,1)$ ------------ (4.3.6)

Fig 4.1 Probability distribution of Betas

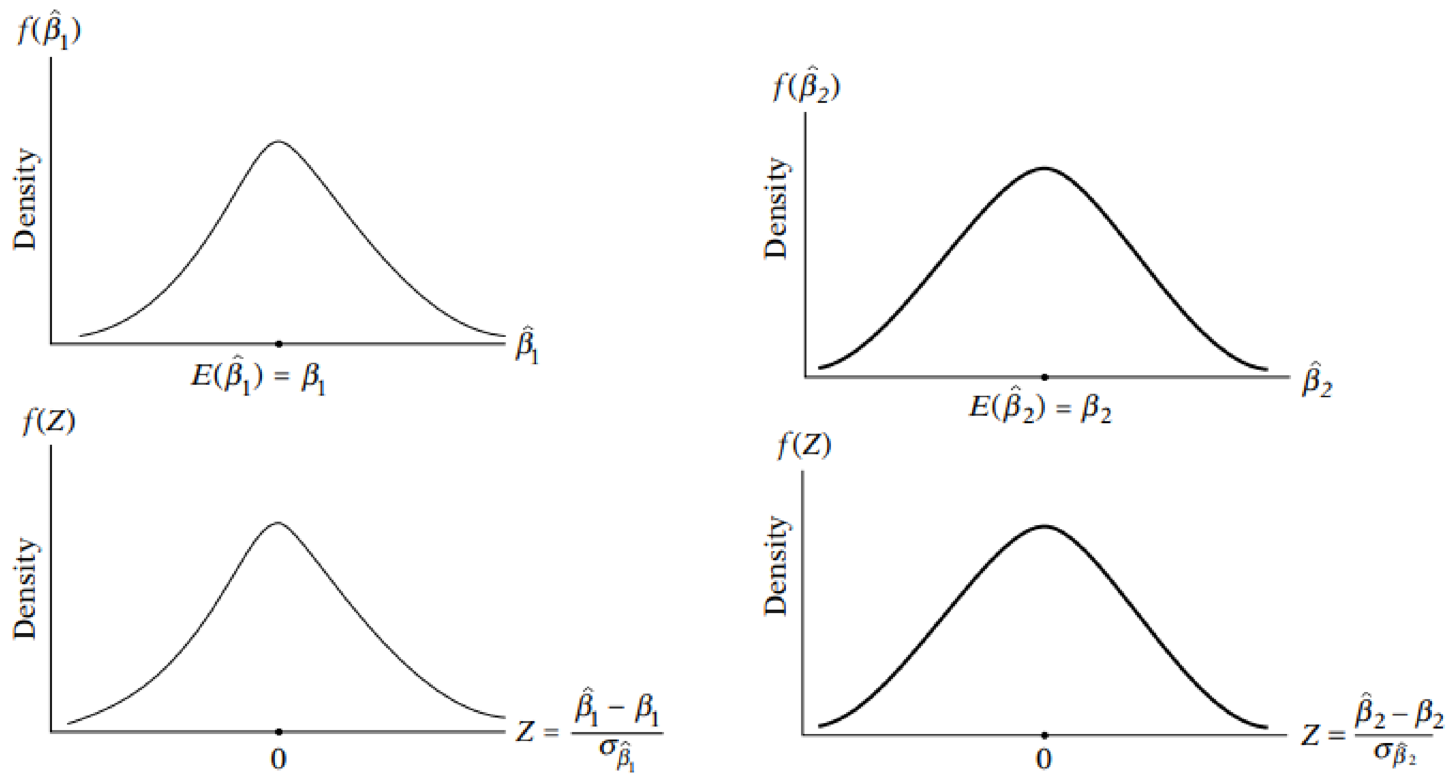


FIGURE 4.1 Probability distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$.

Source: Basic Econometrics by Damodar Gujarati, Page 111.

4-3. Properties of OLS estimators under the normality assumption

5. $\hat{\beta}_2$ is normally distributed $\sim N(\beta_2, \hat{\sigma}_{\beta 2}^2)$

   And $Z = (\hat{\beta}_2 - \beta_2)/\hat{\sigma}_{\beta 2}$ is $\sim N(0,1)$

6. $(n-2)\hat{\sigma}^2/\sigma^2$ is distributed as the

   $\chi^2_{(n-2)}$

7. $\hat{\beta}_1$ and $\hat{\beta}_2$ are distributed independently of $\hat{\sigma}^2$. They have minimum variance in the entire class of unbiased estimators, whether linear or not. They are best unbiased estimators (BUE)

8. Let u_i is $\sim N(0, \sigma^2)$ then Y_i is $\sim$

   $N[E(Y_i); Var(Y_i)] = N[\beta_1 + \beta_2 X_i; \sigma^2]$

To sum up:

- The important point to note is that the normality assumption enables us to derive the probability, or sampling, distributions of $\beta\hat{}_1$ and $\beta\hat{}_2$ (both normal) and $\hat{}\sigma 2$ (related to the chi square).
- This simplifies the task of establishing confidence intervals and testing (statistical) hypotheses.
- Note that, with the assumption that ui $\sim$ N(0, σ2), Yi , being a linear function of ui, is itself normally distributed with the mean and variance given by
- $E(Yi) = \beta_1 + \beta_2 Xi$ -------(4.3.7)
- var (Yi) = σ2 ------ (4.3.8)
- More neatly, we can write
- $Yi \sim N(\beta_1 + \beta_2 Xi , \sigma 2)$ -----(4.3.9)

4-4. The method of Maximum likelihood (ML)

- ML is point estimation method with some stronger theoretical properties than OLS (Appendix 4.A in the Basic Econometrics book)
- The estimators of coefficients β's by OLS and ML are identical. They are true estimators of the β's
- (ML estimator of σ^2) = $\Sigma \hat{u}_i^2/n$ (is biased estimator)
- (OLS estimator of σ^2) = $\Sigma \hat{u}_i^2/n-2$ (is unbiased estimator)
- When sample size (n) gets larger the two estimators tend to be equal
- **Probability distributions related to the Normal Distribution: The t, χ^2, and F distributions** (details in Appendix A in the Basic Econometrics book)

Summary and Conclusions

1. Classical *normal linear regression model (CNLRM).*
2. This model differs from CLRM in that it specifically assumes that the disturbance term *ui entering the regression model is normally distributed.* The CLRM does not require any assumption about the probability distribution of *ui ; it only requires that the mean value of ui is zero and its variance is a finite constant.*
3. The theoretical justification for the normality assumption is the **central limit theorem.**
4. Without the normality assumption, under the other assumptions discussed in Chapter 2, the Gauss–Markov theorem showed that the OLS estimators are BLUE.

Summary and Conclusions

- 5. With the additional assn. of normality, the OLS estimators are not only best unbiased estimators (BUE) but also follow well-known probability distributions. The OLS estimators of the intercept and slope are themselves normally distributed and the OLS estimator of the variance of u_i *(=* $\hat{\sigma}^2$*) is related to the chi-square distribution.*
- 6. This knowledge is important and useful in drawing inferences about the values of the population parameters.
- 7.An alternative to the least-squares method is the method of **ML.** To use this method, however, one must make an assumption about the probability distribution of the disturbance term u_i *. In the regression context, the assumption* most popularly made is that u_i *follows the normal distribution.*

Summary and Conclusions

- 8. Under the normality assn., the ML and OLS estimators of the intercept and slope parameters are identical. However, the OLS and ML estimators of the variance of ui are different. In large samples these two estimators converge.
- 9. Thus the ML method is generally called a large-sample method. The ML method is of broader application, that regression models that are nonlinear in the parameters.
- 10. Largely rely on the OLS method for reasons: (a) Compared to ML, the OLS is easy to apply; (b) the ML and OLS estimators of β1 and β2 are identical; and (c) even in moderately large samples the OLS and ML estimators of σ2 do not differ vastly.

Reference

- Chapter 4: Classical Normal Linear Regression Model (CNLRM)**, in Basic Econometrics Book by Domodar Gujarati**

What next?

- Normality Assumption
- Testing of hypothesis
- Prior to that recollecting the basic statistics for understanding Testing of hypothesis
- Central Limit Theorem