TWO YEAR POST GRADUATE PROGRAMME
(AS PER CBCS AND IN ACCORDANCE WITH NEP 2020)

Basic Ecoometrics

SEMESTER-III

ECONOMICS

KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

1. **Professor Madhurjya Prasad Bezbaruah**,
   Dept. of Economics, Gauhati University
2. **Professor Nissar Ahmed Barua**, Dept. of Economics, Gauhati University
3. **Dr. Gautam Mazumdar**, Dept. of Economics, Cotton University

Course Coordinator : **Parag Dutta,** KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTORS
1,2,3	**Nihal Jyoti Das**, Research Scholar,Gauhati University.
	Dr Rashmi Rekha Mahanta, Assistant Professor,
	Guwahati Commerce College.
4,5	**Dr Prasenjit Bujarbaruah**, Assistant Professor,
	Rajiv Gandhi University.
6,7	**Dr Rajib Sutradhar**, Assistant Professor,
	Christ University, Bangalore.
8	**Dr Nihal Jyoti Das**, Research Scholar,Gauhati University.
	Dr Parag Dutta, Assistant Professor, KKHSOU
9	**Dr Amiyo Sarma**, Assistant Professor, Gauhati University
10	**Dr Jublee Goswami**, Assistant Professor, LCB College
11	**Dr Anup Kumar Das**, Assistant Professor, Rajiv Gandhi University
12, 14	**Dr Shaveta Kohli**, Assistant Pofessor, Jammu University
13	**Bhaskar Bora**, Research Scholar, Gauhati University
	Dr Parag Dutta, Associate Professor, KKHSOU

Editorial Team

Content	:	**Professor M.P. Bezbaruah**, Gauhati University, (Unit 4,5,6,7,8,9,10,11,12,14,15)
		Dr Sagarika Dey, Assistant Professor, Assam University (Unit 1, 2,3,13)
Language	:	**Dr Parag Dutta**, Associate Professor, KKHSOU
		Dola Borkataki, Assistant Professor, KKHSOU (Unit 13)
Structure, Format & Graphics	:	**Dr Parag Dutta,** KKHSOU

March 2023

This Self Learning Material (SLM) of the Krishna Kanta Handiqui State Open University is made available under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 License (international): http://creativecommons.org/licenses/by-nc-sa/4.0/

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University

Headquarter : Patgaon, Rani Gate, Guwahati - 781017

City Office : Resham Nagar, Khanapara, Guwahati-781022; Website: www.kkhsou.in

The University acknowledges with thanks the financial support provided by the **Distance Education Bureau**, **UGC** for the preparation of this study material.

CONTENTS

		Pages
UNIT 1:	**Sampling and Estimation** Concept of Sampling Distribution and Standard Error of a Statistic; Principles of Estimation; Methods of Moments and Maximum Likelihood	5–19
UNIT 2:	**Statistical Inference-I** Testing of Hypothesis: Type I and Type II Errors; Test based on Standard Normal Distributions	20–37
UNIT 3:	**Statistical Inference-II** Small Sample Tests based on t and Chi-Square Distributions	38–56
UNIT 4:	**Linear Regression Model and Its Estimation** Two-variable Linear Regression Model – interpretations of the coefficients, the constant term and the random disturbance - OLS Estimation	57–66
UNIT 5:	**Classical Assumptions and the OLS Estimators** Classical Assumptions and their implications– Mean and Variance of OLS Estimators under the Classical Assumptions – the Best Linear Unbiased character	67–79
UNIT 6:	**Multiple Linear Regression Model and Its Estimation** Multiple Linear Regression Model and Its Estimation	80–92
UNIT 7:	**Inference from Linear Regression Estimation** The Coefficient of Determination – Adjusted R square - Test of Hypothesis about Regression Coefficients	93–114

UNIT 8: Forecasting with the Linear Regression Equation

Point and Interval Prediction with the Linear Regression Model

UNIT 9: Specification Errors and their Consequences

Specification Errors and their Consequences

UNIT 10: Problem of Multicollinearity

Multicollinearity: Effects, Detection and Remedies

UNIT 11: Qualitative Factors as Regressors

Nature of qualitative factors - use of Dummy Variables for capturing the effect of qualitative factors – interpretation of coefficients of dummy variables

UNIT 12: Introduction to Time Series

Components of a time series – Estimation of trend – Moving Average and Curve Fitting methods

UNIT 13: Stochastic Time Series

Idea of a stochastic time Series - Stationary and Non-stationary Time Series – Autocorrelation Function

UNIT 14: Introduction to Simultaneous Equation Model

Structural and Reduced Forms – Simultaneity, Indirect Least Squares and Two Stage least Squares

COURSE INTRODUCTION

Basic Econometrics is one of the course offered as optional course in the fourth semester of the M.A. in Economics programme of this University.This course deals with basics concepts of econometrics and enables the learners with various new knowledge and ideas of Basic Econometrics. This course comprises of 14 units.

The first unit discusses basic concepts of sampling and estimation such as concept of sampling distribution and standard error of a statistic; principles of estimation; methods of moments and maximum likelihood etc The second unit deals with testing of hypothesis, type I and type II error and test based on standard normal variates. Small sample tests based on t and chi square distributions are included in unit 3. Linear regression model and its estimation are included in unit 4. Classical assumptions and the OLS estimators are included in unit 5 such as mean and variance of the classical estimators and the best linear unbaised estimators. Unit 6 of the course includes multiple linear regression model and its estimation. The coffecient of determination, adjusted R square, test of hypothesis about regression cofficients are included in unit 7. This is the second block of the course and comprises of seven units. The 8th unit of this course discusses forcasting with the linear regression estimations. Both point and interval estimation are discussed in this chapter. Specification error and their consequences are discussed in chapter 9. The problem of multicolinearity, detection and remedies are included in chapter 10. Nature of qualitative factors, use of dummy variables for capturing the effect of qualitative factors and interpretation of coefficients of dummy variables are included in chapter 11.The chapter 12 of the course discusses differenet components of a time series, estimation of trend, moving average and curve Fitting methods. The concept and explanation of stochastic time series are given in Unit 13. The idea of stationary and non-stationary time Series and autocorrelation function are discussed in this unit. Unit 14 of the course analyses the simultaneous equation model. Both structural and reduced Forms of simultaneous equation model, the concept of simultaneity, indirect least squares and two stage least squares are included in this unit.

While going through a unit, you will notice some along-side boxes, which have been included to help you to know some of the difficult, unseen terms. Again, we have included some relevant questions in “CHECK YOUR PROGRESS”. These have been designed to self-check your progress of study. It will be better if you solve the problems put in these boxes immediately after you go through the sections of the units and then match your answers with “ANSWERS TO CHECK YOUR PROGRESS” given at the end of each unit.

While going through a unit, you will notice some along-side boxes, which have been included to help you to know some of the difficult, unseen terms. Again, we have included some relevant questions in “CHECK YOUR PROGRESS”. These have been designed to self-check your progress of study. It will be better if you solve the problems put in these boxes immediately after you go through the sections of the units and then match your answers with “ANSWERS TO CHECK YOUR PROGRESS” given at the end of each unit.

UNIT 1: SAMPLING AND ESTIMATION

UNIT STRUCTURE

1.1 Learning Objectives

1.2 Introduction

1.3 Concept of Sampling Distribution of a Statistic

1.3.1 Deriving a Sampling Distribution

1.4 Concept of Standard Error of a Statistic

1.4.1 Utility of Concept of Standard Error

1.5 Principles of Estimation

1.5.1 Properties of a Good Estimator

1.6 Methods of Estimation

1.6.1 Methods of Moments

1.6.2 Maximum Likelihood Method

1.7 Let us Sum up

1.8 Further Readings

1.9 Answers to Check Your Progress

1.10 Model Questions

1.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- explain sampling distribution
- discuss standard error of a statistic
- explain methods of moments
- describe the maximum likelihood method

1.2 INTRODUCTION

In statistics, a population is the entire pool from which a statistical sample is drawn. A population may refer to an entire group of people, objects, events, etc. But, a lot of data drawn and used by academicians, statisticians, researchers, marketers, analysts, etc. are actually samples, not populations. A finite subset of statistical individuals in

a population is called a sample, and the number of individuals or objects in a sample is called the sample size. Since it is difficult to study the entire population due to lack of time, money and man power, a sample is generally taken and we try to form an idea about the characteristics of the population on the basis of estimates obtained from the sample. Generally, our primary focus is in knowing the different measures of a variable (say '*x*') calculated on the basis of sample values of '*x*' and it is called a statistic. A corresponding measure calculated on the basis of population values is called a parameter. A sampling distribution is a statistic that is arrived out through repeated sampling from a larger population. Like any other distribution, a sampling distribution may have its mean, standard deviation and moments of higher order. Of particular importance is the standard deviation, which is designated as standard error of the statistic.

Whenever, we try to learn something about the population using a subset of it (i.e., a sample from the population), it is termed as statistical inference. There are two types of statistical inference– Estimation and hypothesis testing. In this chapter discussion is carried out about estimation part and hypothesis testing is discussed in the second chapter.

1.3 SAMPLING DISTRIBUTION OF A STATISTIC

Since statistics are random variable as their values can vary from sample to sample, therefore, it is customary to refer to their distribution as sampling distribution. A sampling distribution is a probability distribution of a statistic obtained from a larger number of samples drawn from a specific population. The sampling distribution of a statistic is the distribution of frequencies of a range of different outcomes that could possibly occur for a sample statistic. For example, if we consider a sample of size n from a given finite population, then the total number of possible samples is:

$$N_{C_n} = \frac{N!}{n!(N-n)!} = k, \text{ (say)}$$

For each of these *k* sample, we can compute some statistics *t* = *t(x1,x2,……,xn)*, in particular the mean $\bar{x}$, the variance *s2*, etc. The set of the values of the statistic so obtained, one for each sample, constitutes what is called the sampling distribution of the statistic. If, for example, the

particular statistic used is the sample mean, then the distribution is called the *sampling distribution of means.* Similarly, we could have sampling distribution of standard deviations, variances, medians, proportions etc. Moreover, sampling distribution of a statistic depends on the population distribution, the sample size *n* and the method of sampling.

1.3.1 Deriving a Sampling Distribution

Suppose random samples of size two are selected with replacement from a population consists of three numbers, i.e., 2, 4, 6. The population mean is μ = 4 and the total number of likely samples is 3^2 = 9.

Following is the list of samples with their respective means

Sample	2,2	2,4	2,6	4,2	4,4	4,6	6,2	6,4	6,6
Mean($\bar{x}$)	2	3	4	3	4	5	4	5	6

The sampling distribution of sample mean (is given below:

Value of Mean ($\bar{x}$)	2	3	4	5	6
Frequency *(f)*	1	2	3	2	1
Probability $p(\bar{x})$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

Now,

$$\mathrm{E}(\bar{x}) = \Sigma \bar{x}.p(\bar{x}) = 2\times\frac{1}{9} + 3\times\frac{2}{9} + 4\times\frac{3}{9} + 5\times\frac{2}{9} + 6\times\frac{1}{9} = 4$$

$$\mathrm{Var}(\bar{x}) \quad \Sigma \bar{x}^2.p(\bar{x}) \quad 2^2\times\frac{1}{9} \quad 3^2\times\frac{2}{9} \quad 4^2\times\frac{3}{9} \quad 5^2\times\frac{2}{9} \quad 6^2\times\frac{1}{9}$$

Above example illustrates how sampling distribution of a statistic is obtained by first listing all the equally likely probable samples. We can, however, obtain the sampling distribution of a statistic by analytical methods, without actually drawing all possible samples of the same size from the population. In our example the population was very small and sample size large relative to the population so that the number of possible samples was limited. In practice this would not be the case, and deriving the sampling distribution by repeated sampling would be almost impossible.

1.4 STANDARD ERROR OF A STATISTIC

Standard error is the standard deviation of the sampling distribution of a statistic and abbreviated as S.E. The standard deviation of a sample statistic is generally designated by the Greek letter *sigma (ó)*. It measures the sampling variability due to chance or random forces. Standard error plays a very crucial role in the large sample theory. It forms the basis for the testing of a hypothesis as the statistical inference based on construction of the confidence interval is mainly based on concept of standard error.

It is the magnitude of a typical or representative deviation between an estimate and the value of the true population parameter. If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into ó yields the estimated standard error (estimated standard deviation) of the estimator.

It can also be defined as the square root of the variance present in the sample. Symbolically,

$$\text{S.E.}(t) \quad \sqrt{\text{Var}(t)}$$

For example, the standard error of a sample mean will be

$$\text{S.E.}(\bar{x}) = \sqrt{\text{Var}(\bar{x})}$$

$$\text{S.E.}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

This shows that the standard deviation decreases as the sample size (n) increases. That means as the sample size increases we obtain more information from the sample and we can expect the value of sample mean to be closer to the population mean.

1.4.1 Utility of Concept of Standard Error

The concept of standard error is of great significance in statistical work as

- It is used as an instrument in testing a given hypothesis. It is used to test whether the sample statistic differs significantly from the corresponding hypothetical value about the

population.

- It provides idea about the reliability of a sample. The greater the standard error, the greater is the departure of actual frequency from expected ones and hence greater the unreliability. In order words, it is used to determine the precision of the sample statistic of some population parameter.
- It is used to test the significance of the difference between two independent sample statistics of the same population parameter.
- It helps in determining the limit within which the population parameters are expected to lie.



CHECK YOUR PROGRESS

Q 1: Explain the concept of sampling distribution of a statistic. (Answer in about 60 words)

.....

.....

.....

.....

.....

Q 2: What does Standard Error measure? (Answer in about 20 words)

.....

.....

Q 3: Standard error of a sample statistic is a measure of reliability of the sample. Explain. (Answer in about 50 words).

.....

.....

.....

.....

.....

.....

1.5 PRINCIPLES OF ESTIMATION

Estimation is a kind of statistical inference, which is concerned with making judgements about some unknown aspect of a given population on the basis of sample information. The unknown aspect may be the values of one or more of the population parameters. In estimation, we estimate the value of a particular parameter and later on try to conclude by combining our prior knowledge and assumptions about the population with the observations of the sample.

Thus, theory of Estimation is a branch of statistics that deals with estimating the values of parameters based on measured empirical data that has a random component. The parameters describe an underlying physical setting in such a way that their value affects the distribution of the measured data. An estimator attempts to approximate the unknown parameters using the measurements. The theory of estimation can be divided into two parts: Point estimation and interval estimation.

When we use the value of a statistic to estimate a population parameter, we call this point estimation, and we refer to the value of the statistic as point estimator of the parameter. More elaborately, if we have a population whose theoretical probability distribution is known but the parameter such as mean is unknown we can take a sample of size n from that population and obtain the sample mean which serves as an estimator of the population mean. That specific value of $\bar{x}$ the sample mean is called the point estimator. For example, if we use a value to estimate the mean of a population, or a value of S^2 to estimate population variance, we are in each case using a point estimate of the parameter in question. These are called point estimates because in each case a single number, or a single point on the real axis, is used to estimate the parameter. However, the point estimates provide no information about the magnitude of the sampling error that may occur. An alternative to point estimation is interval estimation. In interval estimation, the objective is to produce an interval within which the true value of the parameter may be considered to lie with some given level of probability. Thus, the interval estimator

gives a range of values within which the population parameter may lie. Most prevalent form of interval estimation is confidence intervals. When the interval estimate has a confidence level attached to it, it is called a confidence interval. Confidence level is a measure of the degree of reliability of the interval. A confidence level of 95% implies that 95% of all samples would give an interval that includes the population parameter is being estimated, and only 5% of all samples would yield an erroneous interval. Information about the precision of an interval estimate is conveyed by the width of the interval. If the confidence level is high and the resulting interval is quite narrow, our knowledge of the value of the parameter is reasonably precise. However, a very wide confidence interval gives the message that there is a great deal of uncertainty concerning the value of what we are estimating.

1.5.1 Properties of a Good Estimator

Generally, a good estimator is the one which is as close to the true value of the parameter as possible. However, its quality is to be evaluated in terms of the following properties:

- **Unbiasedness:** An estimator $\hat{\theta}$ = t(x) is said to be unbiased for a function θ if it equals θ in expectation: $E\theta\{t(X)\} = E(\hat{\theta}) = \theta$. That means an unbiased estimator is the one whose mean (or expected) value is equal to the value of the population parameter.
- **Consistency:** An estimator is said to be consistent if its value approaches the true value of the population parameter as the sample size increases. By tracing the behavior of the bias and of the variance of an estimator as the sample sizes approach infinity, we can test the consistency of the estimator. If the increase in sample size is accompanied by a reduction in the biasedness and the variance approaches zero, then the estimator is consistent.
- **Efficiency:** Two naturally desirable properties of estimators are for them to be unbiased and have minimal mean squared

error (MSE). These cannot in general both be satisfied simultaneously: a biased estimator may have lower mean squared error (MSE) than any unbiased estimator. Among unbiased estimators, there often exists one with the lowest variance, called the minimum variance unbiased estimator (MVUE). In some cases an unbiased efficient estimator exists, which, in addition to having the lowest variance among unbiased estimators, satisfies the Cramér–Rao bound, which is an absolute lower bound on variance for statistics of a variable.

- **Sufficiency:** An estimator is said to be a sufficient estimator of the parameter if it contains all the information in the sample regarding the parameter. This means that no other estimator can add any further information about the parameter. For example, between sample mean and median, sample median is not a sufficient estimator since it uses only ranking and not the values of the sample observation. In fact, an estimator cannot be efficient unless it makes use of all the sample information. Therefore, sufficiency is a necessary condition for efficiency of an estimator.

1.6 METHODS OF ESTIMATION

There are different methods of estimation which lead us to estimators that possess different properties. Some of the most commonly used methods of estimation are: method of moments, least square method and maximum likelihood method. Although, least square method is not part of this course, but it is a very popular method of estimation. In least square method, we choose that value of estimator which minimizes the sum of squares of deviation of the sample observations from the estimate.

1.6.1 Method of Moments

In statistics, the method of moments is a method of estimation of population parameters and probably the oldest method of

estimation. It starts by expressing the population moments as functions of the parameters of interest. Those expressions are then set equal to the sample moments. The number of such equations is the same as the number of parameters to be estimated. Those equations are then solved for the parameters of interest. The solutions are estimates of those parameters. That is, if we want to calculate the population mean and variance, then we have to calculate the sample mean and variance. The advantage and disadvantages of moment methods are discussed below:

- **Advantages**
  - Method of moments is simple (compared to other methods like the maximum likelihood method) and can be performed by hand.
- **Disadvantages**
  - The parameter estimates may be inaccurate. This is more frequent with smaller samples and less common with large samples.
  - The method may result insufficient statistics. In other words, it may not take into account all of the relevant information in the sample.

Let us discuss the estimation procedure of method of moments:

Let $f(x; \theta_1, \theta_2,.....,\theta_k)$ be the density function of the parent population with 'k' parameter, i.e., $\theta_1, \theta_2,.....,\theta_k$. If μ_r denotes the rth moment about origin, then

$$\mu_r = \int_{-\alpha}^{\alpha} x^r f(x; \theta_1, \theta_2,.....,\theta_k)dx \quad ;where\ r=1,2,.....,k \qquad (1)$$

In general, $\mu_1, \mu_2,.....,\mu_k$ will be function of the parameters $\theta_1, \theta_2,.....,\theta_k$.

Let x_i $i=1, 2,....., n$ be a random sample of size n from the given population. The method of moments consist in solving k-equations for $\theta_1, \theta_2,.....,\theta_k$ in terms of $\mu_1, \mu_2,.....,\mu_k$ and then replacing these moments $\mu_r, r=1,2,.....,k$ by sample moments, e.g.,

$$\widehat{\theta_i} = \theta_i(\widehat{\mu_1}, \widehat{\mu_2}, \ldots\ldots, \widehat{\mu_k}$$

$$= \theta_i(m'_1, m'_2, \ldots\ldots, m'_k \qquad \text{Where, } i=1,2,\ldots\ldots,k)$$

where, m'_i is the ith moment about the origin in the sample. Then by the method of moments, $\widehat{\theta_1}, \widehat{\theta_2}, \ldots\ldots\ldots, \widehat{\theta_k}$ are the required estimators of $\theta_1, \theta_2, \ldots\ldots, \theta_k$ respectively.



CHECK YOUR PROGRESS

Q 4: Differentiate Point and Interval estimation. (Answer in about 60 words)

.....

.....

.....

.....

.....

.....

Q 5: What do you mean by an unbiased estimator? (Answer in about 20 words)

.....

.....

1.6.2 Maximum Likelihood Estimation

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The logic of maximum likelihood is both intuitive and flexible, and as such the method has become a dominant means of statistical inference.

MLE, is by far, the most general method of estimation from the theoretical point of view. This method was initially formulated by Gauss, but later on developed by Prof. Fisher.

Before defining the MLE, let us define the Likelihood function. Let $x_1,x_2,x_3,\ldots\ldots,x_n$ be a random sample of size 'n' from a population with density function $f(x,\theta)$. Then the likelihood function of the sample values $x_1,x_2,x_3,\ldots\ldots,x_n$ usually denoted by L, is their joint density function, given by

$$L= f(x_1,\theta)\, f(x_2,\theta) \ldots\ldots f(x_n,\theta)$$

$$=\prod_{i=1}^{n} f(x_i,\theta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

L gives the relative likelihood that the random variable assume a particular set of values $x_1,x_2,x_3,\ldots\ldots,x_n$.

For a given sample $x_1,x_2,x_3,\ldots\ldots,x_n$ L becomes a function of the single variable θ, the parameter.

The principle of MLE consists of finding an estimator (of the parameter) which maximizes L for variations in the parameter. Thus, if there exists a function = $\hat{\theta}= \hat{\theta}\,(x_1,x_2,x_3,\ldots\ldots,x_n)$ of the sample values which maximizes L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called Maximum Likelihood Estimator (MLE). Thus, $\hat{\theta}$ is the solution, if any, of

$$\frac{\delta L}{\delta \theta} = 0 \text{ and } \frac{\delta^2 L}{\delta \theta^2} \leq 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Since L>0, so is log L, which shows that L and log L attain their extreme values (Maxima or minima) at the same value of $\hat{\theta}$. From equation (2), we can now rewrite as

$$\frac{1}{L}\frac{\delta L}{\delta \theta} = 0 \Rightarrow \frac{\delta Log L}{\delta \theta} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where equation (3) is usually referred to as the Likelihood Equation. If the likelihood function is differentiable, the derivative test for determining maxima can be applied. In some cases, the first-order conditions of the likelihood function can be solved explicitly; for instance, the ordinary least squares estimator maximizes the likelihood of the linear regression model. Under most circumstances, however, numerical methods will be necessary to find the maximum of the likelihood function.

➢ **Properties of Maximum Likelihood Estimator:**

- ML estimators are consistent estimators for unknown parameters.
- ML estimators are most efficient estimators, i.e., they are minimum variance bound estimators.
- ML estimators are asymptotically normally distributed about the true value of the parameter.
- ML estimators need not be unbiased estimates of the corresponding population parameters.
- ML estimators possess the property of invariance. That is, if $\hat{\theta}$ is the ML estimator of θ and $g(\hat{\theta})$ is a continuous function of $\hat{\theta}$, then $g(\hat{\theta})$ is an ML estimator of $g(\theta)$.



CHECK YOUR PROGRESS

Q 6: Interpret the joint probability density function of the likelihood function. (Answer in about 60 words)

.....

.....

.....

.....

.....

.....

Q 7: What is the reason behind using log transformation of likelihood function while satisfying the maximizing condition of ML estimation method? (Answer in about 50 words)

.....

.....

.....

.....

.....

.....

.....



1.7 LET US SUM UP

- Sampling distribution of a statistic refers to the distribution of frequencies of a range of different outcomes that could possibly occur for a sample statistic.
- The standard deviation of the sampling distribution of a statistic is known as its Standard error.
- Standard error is used to determine the precision of the sample statistic of some population parameter.
- Theory of estimation tells us about how to estimate numerical values for the population parameter from the information provided by the sample.
- In point estimation, we estimate a single value; whereas, in interval estimation, we estimate an interval within which the parameter is expected to lie with some probability.
- Main properties of a good estimator are unbiasedness, consistency, efficiency and sufficiency.
- Under moment method, we calculate the corresponding moments of the sample to estimate the moments of population distribution.
- Maximum likelihood method estimates the unknown population parameters in such a way that the likelihood of observing the given sample is maximum.



1.8 FURTHER READINGS

1) S.P. Singh, A. K. Parashar, and H.P. Singh, *Econometrics and Mathematical Economics,* 5th edition, S. Chand & Company Ltd.
2) S.P. Gupta, *Statistical Methods,* 4th edition, Sultan Chand & Sons.
3) S.P. Gupta and M.P. Gupta, *Business Statistics*
4) S.C. Gupta, *Fundamentals of Statistics,* 7th edition, Himalaya Publishing House.

5) J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th edition, Cengage Learning

6) Miller and M. Miller, *Mathematical Statistics*, 7th Edition, Pearson.

7) G.M.K. Madnani, *Introduction to Econometrics Principles and Applications*, 8th edition, Oxford & IBH Publishing Co. Pvt. Ltd.



1.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: if we consider a sample of size n from a given finite population, then the total number of possible samples is:

$$N_{C_n} = \frac{N!}{n!(N-n)!} = \text{k, (say)}$$

For each of these *k* sample, we can compute some statistics $t = t(x_1, x_2, \ldots\ldots, x_n)$, in particular the mean $\bar{x}$, the variance s^2, etc. The set of the values of the statistic so obtained, one for each sample, constitutes what is called the sampling distribution of the statistic.

Ans to Q No 2: Standard error measures the sampling variability due to chance or random forces.

Ans to Q No 3: Standard error provides idea about the reliability of a sample. The greater the standard error, the greater is the departure of actual frequency from expected ones and hence greater the unreliability. On the contrary, the smaller the standard error, the more reliable the sample will be.

Ans to Q No 4: When we use the value of a statistic to estimate a population parameter, we call this point estimation. These are called point estimates because in each case a single number, or a single point on the real axis, is used to estimate the parameter. On the other hand, in interval estimation, the objective is to produce an interval within which the true value of the parameter may be considered to lie with some given level of probability. So it provides two bound, i.e., lower bound and upper bound, within which the

true population parameter is expected to lie.

Ans to Q No 5: An unbiased estimator is the one whose mean (or expected) value is equal to the value of the population parameter.

Ans to Q No 6: The joint probability density function of the likelihood function can be interpreted in two ways:

a) Given the population parameter (θ), it indicates the probability of a set of sample observations.

b) Secondly, it may be interpreted as a function of population parameter (θ), conditional on a set of sample outcomes. This interpretation is known as Likelihood function.

Ans to Q No 7: Practically, in most cases, we do not maximise the likelihood function directly. Instead, we maximise the log likelihood function with respect to the parameter, è. This gives the same results. It means that the set of values for è that maximises the log-likelihood function also maximises the likelihood function. Because the log-likelihood function is a (positive) monotonic transformations of likelihood function. Besides, it is mathematically easier to maximise the log-likelihood function than to maximise the likelihood function directly.



1.10 MODEL QUESTIONS

Short Questions (Answer each question in about 150 words)

Q 1: Write a short note on Standard Error of a Statistic.

Q 2: Elaborate the Moment method of estimation.

Long Questions (Answer each question in about 300-500 words)

Q 1: Discuss the Maximum likelihood method of estimation.

Q 2: A population consists of four numbers 0,3,6,9. Draw all possible samples of size two with replacement. Find the sampling distribution of the sample mean. Also find mean and variance of the sample mean.

*** ***** ***

UNIT 2: STATISTICAL INFERENCE-I

UNIT STRUCTURE

2.1 Learning Objectives

2.2 Introduction

2.3 Concept of Testing of Hypothesis

2.4 Type I and Type II Errors

2.5 Test based on Standard Normal Distributions

2.6 Let us Sum up

2.7 Further Readings

2.8 Answers to Check Your Progress

2.9 Model Questions

2.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- know about the concept of hypothesis testing.
- distinguish between type I error and type II error.
- learn about the various steps involved in testing a statistical hypothesis.
- know about the specifications of a standard normal distribution.
- test the hypothesis related to samplings of variables.
- test the hypothesis related to significance of difference in standard deviations.
- test the hypothesis related to samplings of attributes.

2.2 INTRODUCTION

Statistical inference is that branch of statistics which is concerned with using probability concepts to deal with uncertainty in decision making. It deals with two different classes of problems: Estimation and Hypothesis testing. In the previous chapter, we have discussed about estimation. In the present chapter, we will be discussing about hypothesis testing and some related concepts. In the later sections, we will be discussing about

standard normal distribution and the various tests which are based on it.

2.3 CONCEPT OF TESTING OF HYPOTHESIS

A hypothesis is a theoretical proposition that is capable of empirical verification. It may be viewed as an explanation of some event (or events) and which may be true or false explanation. So, hypothesis testing is used to test some hypothesis about population from which the sample is drawn. In other words, hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. It is used to assess the plausibility of a hypothesis by using sample data. Such data may come from a larger population, or from a data-generating process. Thus, hypothesis testing is a procedure in inferential statistics that assesses two mutually exclusive theories about the properties of a population.

Generally, there are two types of hypotheses as found in statistics: Simple hypothesis and Composite hypothesis. If the hypothesis completely specifies the population, then it is known as simple hypothesis. For example, marks of students in mathematics examination are normally distributed with variance ($ó^2$) 12 and the hypothesis that it's mean (μ) is 76. So, here we have stated a simple hypothesis as mean and variance together specifies a normal distribution. However, a hypothesis in which not all the parameters are specified is called a composite hypothesis. For example, if we hypothesize that mean (μ) is 76 and $ó^2<10$ or $\mu>68$ and $ó^2=12$, the hypothesis becomes a composite hypothesis because we cannot know the exact distribution of the population in either case. Again, all the assumptions about population are not testable. Those assumptions that are not liable to confront to any test is called maintained hypothesis, and the remaining hypothesis are testable hypothesis. The hypothesis which is to be tested is known as null hypothesis. According to Prof. R.A. Fisher, Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true. It is denoted by H_0. Since null hypothesis is a testable hypothesis, there must be counter proposition to test the proposed null hypothesis. This counter proposition is called alternative hypothesis. It is usually denoted by H_1. It is very important to

explicitly state the alternative hypothesis (H_1) in respect of any null hypothesis (H_0). Again, alternative hypothesis can be of two types: Two-tailed alternative hypothesis and One-tailed (right tailed or left-tailed) alternative hypothesis respectively.



CHECK YOUR PROGRESS

Q 1: Define a statistical hypothesis. (Answer in about 30 words)

.....

.....

.....

Q 2: State the link between null and alternative hypothesis? (Answer in about 60 words)

.....

.....

.....

.....

.....

.....

2.3.1 Characteristics of a Good Hypothesis

- **Hypothesis should be simple -** A simple hypothesis contains one predictor and one outcome variable, whereas a complex hypothesis contains more than one predictor variable or more than one outcome variable. Complex hypothesis like this cannot be easily tested with a single statistical test and should always be separated into 2 or more simple hypotheses.
- **Hypothesis should be specific -** A specific hypothesis leaves no ambiguity about the subjects and variables, or about how the test of statistical significance will be applied. It uses concise operational definitions that summarize the nature and source of the subjects and the approach to measuring variables.

- **Hypothesis should be stated in advance -** The hypothesis must be stated in writing during the proposal state. This will help to keep the research effort focused on the primary objective and create a stronger basis for interpreting the study's results as compared to a hypothesis that emerges as a result of inspecting the data.

2.4 TYPES OF ERRORS IN HYPOTHESIS TESTING

Statisticians define two types of errors in hypothesis testing.

- **Type 1 Error**: When the hypothesis is true, but our test rejects it, i.e., Reject H_0 when it is true.
- **Type 2 Error**: When the hypothesis is false but our tests accepts it, i.e. Accept H_0 when it is wrong.

In a statistical hypothesis testing experiment, a Type 1 error is committed by rejecting the null hypothesis when it is true. The probability of committing a Type 1 error is denoted by α (alpha), where

α = *Prob. (Type 1 error)*

= *Prob. (Rejecting null hypothesis when it is true)*

On the other hand, a Type 2 error is committed by accepting the null hypothesis when it is false. The probability of committing a Type 2 error is denoted by β (beta) where,

β = *Prob. (Type 2 error)*

= *Prob. (Accepting null hypothesis when it is false)*

The table summarizes the four possible outcomes for a hypothesis test.

	Accept Null (H_0)	**Reject Null (H_1)**
Null is True	Correct Decision	Type 1 Error
Null is False	Type 2 Error	Correct Decision

Example - Fire alarm analogy for the types of errors

A fire alarm provides a good analogy for the types of hypothesis testing errors. Preferably, the alarm rings when there is a fire and does not ring in the absence of a fire. However, if the alarm rings when there is no fire, it is a false positive, or a Type I error in statistical terms.

Conversely, if the fire alarm fails to ring when there is a fire, it is a false negative, or a Type II error.

While testing hypothesis, the aim is to reduce both the types of errors. But, due to fixed sample size, it is not possible to control both the errors simultaneously. Thus, a trade-off between these two types of errors has to be made: the probability of making one type of error can only be reduced if we are willing to increase the probability of making the other type of error. In order to get a low β, we have to put up with a high α. To deal with this situation, business managers decide the appropriate level of significance by examining the cost attached to both types of errors.

It is more dangerous to reject a true hypothesis (Type I error) than to accept a false one (Type II error). Hence, we keep the probability of committing Type I error at a certain level, called the level of significance (also known as the size of rejection region or size of the critical region or size of test) and subsequently try to minimize type II error as far as possible. In most statistical tests, the level of significance is generally fixed at 5%, which means that the probability of accepting a true hypothesis is 95%.

2.4.1 Power of Test

Power of test means the probability of rejecting a null hypothesis when it is false or the probability of rejecting null hypothesis when alternative hypothesis is true. W can write,

$$P[\text{Accept } H_0 \text{ when is } H_0 \text{ wrong}] + P[\text{Accept } H_0 \text{ when is } H_0 \text{ true}] = 1$$

Or $$P[\text{Accept } H_0 \text{ when is } H_0 \text{ true}] = 1 - P[\text{Accept } H_0 \text{ when is } H_0 \text{ wrong}]$$

Or $$\text{Power of Test} = 1 - \beta$$

The value of power of test lies between zero (0) and one (1). As the power of test increases, the probability of committing type II error decreases. Therefore, the usual practice in hypothesis is to fix α (the size of type I error), and then try to obtain a criterion which minimizes β (the size of type II error) or maximizes $1-\beta$ (the power of test).

2.5 PROCEDURE FOR TESTING OF A HYPOTHESIS

The various steps involves in testing of a statistical hypothesis are summarized below in a systematic manner:

Step 1: Formulate the null and alternative hypotheses.

Step 2: Choose an appropriate level of significance (α) depending on the permissible risk.

Step 3: Identify the sample statistic to be used and its sampling distribution.

Step 4: Define and Compute the test statistic under null hypothesis (H_0). Some of commonly used distributions in obtaining the test statistic are normal, *t*, chi-square, F and tests based on them.

Step 5: Obtain the critical value and the critical (rejection) region of the test statistic from the appropriate tables.

Step 6: If the computed value of the test statistic lies outside the critical region, we reject H_0 at level of significance (α). However, if it lies outside the critical region, we fail reject H_0.



CHECK YOUR PROGRESS

Q 3: Explain the concepts of Type I and Type II errors. (Answer in about 100 words)

.....

.....

.....

.....

.....

.....

.....

.....

2.6 TEST BASED ON STANDARD NORMAL DISTRIBUTIONS

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that

data near the mean are more frequent in occurrence than data far from the mean. Thus, in probability theory, a normal (or Gaussian) distribution is a type of continuous probability distribution for a real-valued random variable.

A continuous random variable x is said to follow normal distribution with Mean (ì) and standard deviation (ó) with the probability and its probability density function (p.d.f.) is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\,.\sigma}.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad -\infty < x < \infty \qquad (1)$$

The varietal $\frac{x-\mu}{\sigma}$ is known as Standard Normal Variate and the distribution of it is known a Standard Normal Distribution which is given by:

$$p(z) = \frac{1}{\sqrt{2\pi}\,.\sigma}.e^{-\frac{z^2}{2}}; \quad -\infty < z < \infty \qquad (2)$$

Standard normal distribution is the simplest case of a normal distribution, where the mean μ= 0 and standard deviation σ= 1. Again, the tests based on standard normal distribution are known as Z-test. In other words, a Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Since many test statistics are approximately normally distributed for large samples (due to the central limit theorem), many statistical tests can be performed as approximate Z-tests if the sample size is not too small. Some of the large sample tests which are based on standard normal distribution are listed below:

2.6.1 Test of Significance for a Single Mean

Suppose, we are asked to test whether the sample mean is significantly different from the population mean. This may also ask as whether the mean of the population has a specified value μ_0 ($\mu=\mu_0$) or whether the given random sample has been drawn from a population with mean (μ_0) and variance (σ^2). All these problems are basically the same. Therefore, the null hypothesis corresponding to the initial problem is

$$H_0: \bar{x} = \mu_0$$

And the alternative hypothesis is

$$H_1: \bar{x} \neq \mu_0$$

While testing the above null hypothesis, we can have two different situations: (i) when population variance is known; and (ii) when population variance is unknown. For situation (i), the test statistics is

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

For situation (ii) when population variance is unknown, then the estimated sample variance (s^2) is used instead of population variance, i.e.,

$$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

At last, we compare the calculated value of Z with 1.96 or 2.58 (for two-tailed test) and reject or retain the null hypothesis at 5% or 1% level of significance respectively. That means, if the calculated Z-value is lower than the critical value, then we fail to reject the null hypothesis. However, if the calculated value is greater than the critical value, we reject the null hypothesis and accept the alternative hypothesis.

Example 1: It is claimed that a random sample of 100 tyres with a mean life of 15269 kilometers is drawn from a population which has a mean life of 15200 kilometers and standard deviation of 1248 kilometers. Test the validity of the claim at 5% level of significance.

Solution: In usual notations, we are given with

$$n = 100$$

$$\bar{x} = 15269\text{kms}$$

$$\mu = 15200\text{kms}$$

$$\sigma = 1248\text{kms}$$

The null hypothesis to be test is

$$H_0: \bar{x} = \mu$$

i.e., there is no significant difference between the sample mean and population mean.

Alternative hypothesis is

$$H_1: \bar{x} \neq \mu$$

Since the sample size is large (n>30), under the null hypothesis the test statistic is

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$= \frac{15269 - 15200}{1248/\sqrt{100}}$$

$$= \frac{69}{124.8}$$

$$= 0.5529$$

Since the computed Z value is less than 1.96, it is not significant at 5% level of significance. Therefore, we cannot reject the null hypothesis, which may be accepted a 5% level. This implies that the sample has been drawn from the population.

2.6.2 Test of Significance for Difference of Means

Sometimes we need to test the significance of the difference between two independent sample means. The objective may be to ascertain whether both samples are drawn from the same population or to check whether a particular characteristic is the same in two populations. So, the underlying null hypothesis is that population means of both the populations are the same, i.e.,

$$H_0: \mu_1 = \mu_2$$

and the alternative hypothesis is

$$H_1: \mu_1 \neq \mu_2$$

The similar situations, as observed in the previous test, can also be faced in this test. That means, we may face with two situations: when the population variances are known and when the population variances are unknown. For the earlier situation,

the test statistic for the underlying null hypothesis is

$$Z=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

If the population variances are unknown, then their estimates provided by the corresponding sample variance s_1^2 and s_2^2 respectively used, i.e.,

$$Z=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

Example 2: A man buys 50 electric bulbs of Phillips and 50 bulbs of Panasonic. He finds that Phillips bulbs gave an average life of 1512hours with standard deviation 80hours and Panasonic bulbs gave an average life of 1500hours with standard deviation of 60hours. Is there any significant difference between the mean life of bulbs of the two companies at 1% level of significance?

Solution: Let x_1 and x_2 be life bulbs of Phillips and Panasonic respectively. In usual notation, we are given with

$n_1=50$, $\bar{x}_1= 1512$, $s_1= 80$; therefore $s_1^2= 6400$

$n_2=50$, $\bar{x}_2= 1500$, $s_2= 60$; therefore $s_2^2= 3600$

H_0: there is no significant difference between the mean life of bulbs of the two companies.

H_1: there is significant difference between the mean life of bulbs of the two companies.

Applying Z test,

$$Z=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$=\frac{1512-1500}{\sqrt{\frac{6400}{50}+\frac{3600}{50}}}$$

$$= 0.85$$

Since the calculated Z value is less than 2.58, we accept the null hypothesis and conclude that there is no significant difference between mean life of bulbs of the two companies.

2.6.3 Test of Significance for Difference of Standard Deviation

Suppose we need to test the significance of the difference between standard deviations (say, σ_1 and σ_2) of two independent random samples of size n_1 and n_2 from two populations. The underlying null hypothesis is that the sample standard deviations do not differ significantly, i.e.,

$$H_0: \sigma_1 = \sigma_2$$

and the alternative hypothesis is

$$H_1: \sigma_1 \neq \sigma_2$$

To test the null hypothesis, the test statistic is given by

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

Usually, σ_1^2 and σ_2^2 are unknown. In such a situation, we use their estimates given by the corresponding sample variances. Hence, the test statistic reduces to

$$Z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

Example 3: In a sample of 1000, mean is 17.5 and standard deviation is 2.5. In another sample of 800, the mean is 18 and standard deviation is 2.7. Is there any significant difference between standard deviation of the two samples at 1% level of significance?

Solution: In usual notation,

$$n_1 = 1000,\ \bar{x}_1 = 17.5,\ s_1 = 2.5$$

$$n_2 = 800,\ \bar{x}_2 = 18,\ s_2 = 2.7$$

Here,

$$H_0: s_1 = s_2$$

$$H_1: s_1 \neq s_2$$

The test statistic is given by

$$|Z| = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

$$= \frac{2.5 - 2.7}{\sqrt{\frac{2.5}{2(1000)} + \frac{2.7}{2(800)}}}$$

$$= 2.28$$

Since |Z| is less than 2.58, we accept the null hypothesis at 1% level of significance. Therefore, we can conclude that there is significant difference between standard deviation of the two samples.

2.6.4 Test For Single Proportion

Let *p* denote the proportion of individuals or objects in a population who possess a specified attribute. If an individual or object with the property is labeled a success (*S*), then *p* is the population proportion of successes. Tests concerning *p* will be based on a random sample of size *n* from the population. If n is large, then both *X* (number of S's in the sample) and $p = \frac{X}{n}$ are approximately normally distributed. When *p* is unbiased, i.e., E(*p*)= P (where P is the population proportion of successes) and its standard deviation is

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}},$$

then it follows that the test statistic

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

has approximately a standard normal distribution.

Example 4: A dice is thrown 9000 times, out of which 3220 times yield 3 or 4. Is this consistent with the hypothesis that the dice is unbiased?

Solution: In usual notation, we are given

$$n = 9000$$

$$p = \frac{3220}{9000} = 0.36$$

$$P = \text{probability of getting 3 or 4}$$

$$= \frac{2}{6} = 0.33$$

Now,

H_0: the dice is unbiased.

H_1: the dice is biased.

Test statistic is given by

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

$$= \frac{0.36 - 0.33}{\sqrt{\frac{0.33(1-0.33)}{9000}}}$$

$$= 7.5$$

Since the calculated Z value is greater than 1.96, we reject the null hypothesis and conclude that the dice is biased.

2.6.5 Test Of Significance For Difference Of Proportions

If two samples are drawn from two different populations, we may be asked to find whether the difference between the proportion of the success is significant or not. In such a case, the null hypothesis is that there is no significant difference between p_1 (the proportion of success in sample 1) and p_2 (the proportion of success in sample 2). The difference, if any, is due to fluctuations of random sampling. Under the null hypothesis

$$H_0: p_1 = p_2$$

the test statistic for the difference of proportions is

$$Z=\frac{p_1-p_2}{\sqrt{P(1-P)(\frac{1}{n_1}+\frac{1}{n_2})}}$$

In general, *P,* the common population proportion is not known. So, we use its unbiased estimate provided by both the samples taken together which is given by

$$\hat{P}=\frac{n_1p_1+n_2p_2}{n_1+n_2}$$

By substituting into the above mentioned test statistic, we get the required test statistic for testing the underlying null hypothesis.

Example 5: A company has the head office at Kalkata and a branch at Mumbai. A personal director wanted to know if the worker at the two places would like the introduction of a new plan of work. And a survey was conducted for this purpose. Out of a sample of 500 workers at Kalkata, 62% favored the plan. At Mumbai, out of a sample of 400 workers 41% were against it. Is there any significant difference between the two groups in their attitude towards the new plan at 5% level of significance?

Solution: In usual notation,

$$n_1=500$$

$$n_2=400$$

$$p_1=0.62$$

$$q_1=1-p_1=1-0.62=0.38$$

$$q_2=0.41$$

$$p_2=1-q_2=1-0.41=0.59$$

Here,

$$H_0: p_1=p_2$$

$$H_1: p_1\neq p_2$$

Since, population proportion is not known, therefore

$$\hat{P}=\frac{n_1p_1+n_2p_2}{n_1+n_2}$$

$$=\frac{500(0.62)+400(0.59)}{500+400}$$

$$=0.394$$

The test statistic for the difference of proportions is

$$Z=\frac{p_1-p_2}{\sqrt{P(1-P)(\frac{1}{n_1}+\frac{1}{n_2})}}$$

$$=\frac{0.62-0.59}{\sqrt{0.606(1-0.606)(\frac{1}{500}+\frac{1}{400})}}$$

$$=0.94$$

Since calculated Z value is less than 1.96, we fail to reject the null hypothesis. So we accept the null hypothesis and conclude that there is significant difference between the two groups in their attitude towards the new plan at 5% level of significance.



CHECK YOUR PROGRESS

Q 4: State the conditions of a standard normal distribution. (Answer in about 100 words)

.....

.....

.....

.....

.....

.....

.....

.....

Q 5: what is a test of significance? (Answer in about 30 words)

.....

.....

.....

.....



2.7 LET US SUM UP

- Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter.
- A null hypothesis is a definitive statement about the population parameter(s), which is under test. Any counter hypothesis to the null hypothesis is called an alternative hypothesis.
- Two types of errors need to be deal while testing a statistical hypothesis.
- Practically, consequences of type II error are likely to be more seriously than the consequences of type I error. So, we fix the size of type I error and then try to obtain a criterion which minimizes the size of type II error.
- The statistics that leads to the rejection of a null hypothesis gives us a region called critical region or rejection region. While those lead to the acceptance of null hypothesis gives us a region called acceptance region.
- Standard normal distribution is the simplest case of a normal distribution, where the mean $\mu = 0$ and standard deviation $\sigma = 1$.



2.8 FURTHER READING

1) S.P. Singh, A. K. Parashar, and H.P. Singh, Econometrics and Mathematical Economics, 5th edition, S. Chand & Company Ltd.
2) S.P. Gupta, Statistical Methods, 4th edition, Sultan Chand & Sons.
3) S.P. Gupta and M.P. Gupta, Business Statistics
4) S.C. Gupta, Fundamentals of Statistics, 7th edition, Himalaya Publishing House.
5) J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th edition, Cengage Learning

2.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: A statistical hypothesis is some assumption or statement, which may or may not be true about a population which we wanted to test on the basis of the evidence from a random sample.

Ans to Q No 2: Since any counter hypothesis to the null hypothesis is called an alternative hypothesis; therefore it is important to explicitly state alternative hypothesis in respect of the null hypothesis, because the acceptance or rejection of any null hypothesis is meaningful only if it is being tested against a rival hypothesis (or alternative hypothesis).

Ans to Q No 3: Go through section 21.4 and answer.

Ans to Q No 4: A continuous random variable x is said to follow normal distribution with Mean (μ) and standard deviation (σ) with the probability and its probability density function (p.d.f.) is given by

$$p(x) = \frac{1}{\sqrt{2\pi}.\sigma}.e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}; \quad -\infty < x < \infty \qquad (1)$$

The varietal is known as Standard Normal Variate and the distribution of it is known a Standard Normal Distribution which is given by:

$$p(z) = \frac{1}{\sqrt{2\pi}.\sigma}.e^{-\frac{z^2}{2}}; \quad -\infty < z < \infty \qquad (2)$$

That means, a standard normal distribution is the simplest case of a normal distribution, where the mean μ= 0 and standard deviation σ= 1.

Ans to Q No 5: The process of assessing the significance of a statistic or difference between two independent statistics is known as a test of significance. Insignificance of the test can be attributed to the fluctuations of random sampling.



2.10 MODEL QUESTIONS

Short Questions (Answer each question in about 150 words)

Q 1: Explain the following terms:

(a) Level of significance,

(b) Critical values,

(c) Acceptance and Rejection Regions.

Q 2: What is a statistical hypothesis? Mention the different characteristics of a good hypothesis?

Long Questions (Answer each question in about 300-500 words)

Q 1: Explain the various steps involved in the testing a statistical hypothesis.

Q 2: In a survey of buying habits of 400 women shoppers are chosen at random in Supermarket A. Their average monthly food expenditure is 5000 with standard deviation of 800. For 400 women shoppers chosen at random in supermarket B, the average monthly food expenditure is 4400 with standard deviation 1100. Test whether average monthly food expenditures of the two groups of shoppers are equal?

Q 3: A machine puts out 16 imperfect articles in sample of 500. After repairing, it puts out 3 imperfect articles in a batch of 100. Has the machine improved?

*** ***** ***

UNIT 3 : STATISTICAL INFERENCE-II

UNIT STRUCTURE

3.1 Learning Objectives

3.2 Introduction

3.3 Student's '*T*' Distribution

3.3.1 Properties of Student's '*T*' Distribution

3.4 Comparison of Standard Normal Distribution and Student's '*T*' Distribution

3.5 Tests Based on Student's '*T*' Distribution

3.5.1 '*T*' Test for Significance of the Mean of a Random Sample

3.5.2 '*T*' Test for Difference between Means

3.5.3 Paired '*T*' Test for Difference of Means

3.5.4 '*T*' Test for Significance of an Observed Sample Correlation Co-Efficient

3.6 Chi-Square (X^2) Test

3.6.1 Properties of Chi-Square Test Statistic

3.6.2 Conditions for Applying X^2 Test

3.7 Degrees of Freedom

3.8 Yates' Corrections for Continuity

3.9 Applications of Chi-Square (X^2) Test

3.9.1 Chi-Square Test of Goodness of Fit

3.9.2 Chi-Square Test for Independence of Attributes

3.9.3 Chi-Square Test for the Population Variance

3.10 Let us Sum up

3.11 Further Readings

3.12 Answers to Check Your Progress

3.13 Model Questions

3.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- learn about *t*-distribution and chi-square distribution;

- distinguish between standard normal distribution and *t*-distribution;
- perform various tests based on *t*-distribution and chi-square distribution;
- know importance of the concept of degrees of freedom;
- understand how to measure goodness of fit of a test;

3.2 INTRODUCTION

In the previous chapter, we have discussed about the tests based on standard normal distribution which assumes a normally distributed population. However, if sample size n is small, then the distribution of standardized statistics is far from normality and consequently normal test cannot be applied. Therefore, When the size of the sample is small (less than 30), we use small sample tests. While dealing with small samples our main interest is not to estimate the population values as is true in large samples; rather our interest lies in testing a given hypothesis, i.e. in ascertaining whether observed values could have arisen by sampling fluctuations from some value given in advance.

3.3 STUDENT'S 't' DISTRIBUTION

One of the greatest contributors in the development of theory of small samples is Sir William S. Gosset. He has done the theoretical work on *'t'* distribution. He used a pen name *'student'* to publish his works. Therefore, the *'t'* distribution is commonly known as student's *'t'* distribution. The *'t'* distribution is used when sample size is less than 30. The mathematical derivation of *'t'* distribution is based on the following assumptions:

a) The population is normally distributed from which the sample(s) is drawn.
b) The sample(s) is random and independent of each other.
c) Population variance is unknown.

If $x_1, x_2, \ldots\ldots, x_n$ is a random sample of size n from a normally distributed population with mean (μ) and variance σ^2, then *'t'* statistic s defined as

$$t=\frac{\bar{x}-\mu}{s}\sqrt{n}$$

where, $\bar{x}$ is the sample mean and *S* is the unbiased estimator of σ, i.e.,

$$S=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

't' statistic follows student's *'t'* distribution with degrees of freedom v=(n–1) and its probability density function (p.d.f.) is given by

$$p(t)= C[1+\frac{t^2}{v}]^{-(v+1)/2}$$

Where C is a constant required to make the area under the curve equal to unity.

3.3.1 Properties of Student's 't' Distribution

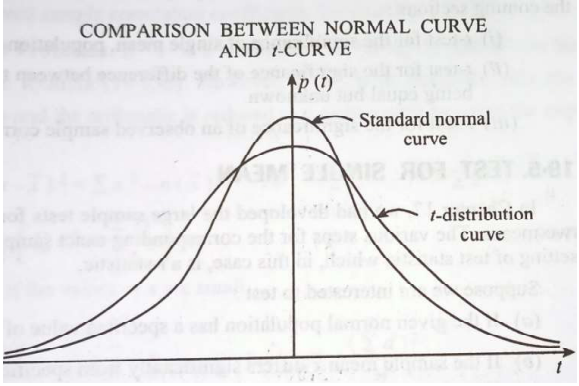
- The variable t-distribution ranges from minus infinity ($-\infty$) to plus infinity (∞).
- The constant ***c*** is actually a function of ***v***. So that for a particular value of ***v***, the distribution of f (t) is completely specified. Thus, f (t) is a family of functions, one for each value of ***v***.
- Like the standard normal distribution, the t-distribution is symmetrical and has a mean zero.
- The variance of the t-distribution is greater than one, but approaches one as the number of degrees of freedom and therefore, the sample size becomes large. Thus, the variance of the t-distribution approaches the variance of the standard normal distribution as the sample size increases. It can be demonstrated that for an infinite number of degrees of freedom ***(v*** $=\infty$***)***, the t-distribution and normal-distribution are exactly equal.

3.4 COMPARISON BETWEEN STANDARD NORMAL DISTRIBUTION AND 't' DISTRIBUTION

The diagram 1 provides a comparison of standard normal distribution and student's *'t'* distribution. It shows that the tails of the student's *'t'* distribution have greater area (or probability) than the tails of the standard normal distribution. Therefore, interval widths from student's *'t'* distribution are wider than those based on the standard normal distribution. However, student's *'t'* distribution has less area (probability) in the center than the

standard normal distribution. Again, as the number of degrees of freedom (d.f) increases, i.e., for large *v* (practically ***v*** > 30), student's *'t'* distribution tends to be standard normal distribution.

Diagram : 1





CHECK YOUR PROGRESS

Q 1: Mention the specific condition for using t-test. (Answer in about 20 words)

.....

.....

Q 2: Explain the relationship between Standard Normal Distribution and Student's 't' distribution. (Answer in about 80 words)

.....

.....

.....

.....

3.5 TESTS BASED ON STUDENT'S 't' DISTRIBUTION

The following are some of the applications or tests based on Student's *'t'* distribution:

3.5.1 *'t'* Test For Significance Of The Mean Of A Random Sample

In determining whether the sample mean ($\bar{x}$) differs significantly from specified values of population mean (μ_0) when

population variance is unknown, we test the following null hypothesis

$$H_0: \bar{x} = \mu_0$$

And the alternative hypothesis is

$$H_1: \bar{x} \neq \mu_0$$

Under the given null hypothesis, the test statistic is

$$t = \frac{\bar{x} - \mu_0}{S}\sqrt{n}$$

where, $\bar{x} = \frac{\sum x}{n}$ and $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$ and it follows Student's *'t'* distribution with (n–1) degrees of freedom.

At the end, we compare the computed test statistic with the tabulated value of *t* for (n–1) degrees of freedom at the given level of significance. If the calculated | t | exceeds tabulated *'t',* we reject the null hypothesis, i.e., there is significant difference between sample mean ($\bar{x}$) and population mean (μ_0). Otherwise, if calculated | t | is less than tabulated *'t',* the null hypothesis can be accepted at the level of significance.

Example 1: The mean weekly sales of the chocolate bar in candy stores were 146.3 bars per store. After an advertising campaign, the mean weekly sales in 21 stores increased to 153.7 bars and showed a standard deviation of 17.2. Was the advertising campaign successful?

Solution: Given,

$$n = 21$$

$$\bar{x} = 153.7$$

$$S = 17.2$$

Here,

H_0: the advertising campaign was not successful

H_1: the advertising campaign was successful (i.e., Right tailed)

Under the null hypothesis, the test statistic is

$$t = \frac{153.7 - 146.3}{17.2}\sqrt{21}$$

$$= 1.9716$$

The tabulated value of *t* for 20 degrees of freedom at 5% level of significance for one tailed test is 1.725. Since our calculated value is greater than the tabulated value, we reject the null hypothesis. Hence, the advertising campaign was successful.

3.5.2 't' Test For Difference Between Means

Suppose we are test whether the two independent samples (n_1 and n_2) have been drawn from the same normal population. Let $x_1, x_2, \ldots\ldots, x_n$ and $y_1, y_2, \ldots\ldots, y_n$ be the two independent random samples from the given normal population. Here the null hypothesis is that the samples have been drawn from the same normal population. In other words, the sample means $\bar{x}$ and $\bar{y}$ do not differ significantly.

$$H_0: \mu_x = \mu_y$$

and alternative hypothesis is

$$H_0: \mu_x \neq \mu_y$$

Under the null hypothesis, the test statistic is

$$t = \frac{\bar{x}-\bar{y}}{S}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

where,

$$S = \sqrt{\frac{\sum(x-\bar{x})^2+\sum(y-\bar{y})^2}{n_1+n_2-2}}$$

Degrees of freedom= n_1+n_2-2

Example 2: The average numbers of articles produced by two machines per day are 200 and 250 with standard deviations 20 and 25 respectively on the basis of records of 25 day's production. Can we regard both the machines equally efficient at 1% level of significance?

Solution: In usual notations,

$$n_1=n_2=25$$

$$\bar{x}_1=200 \quad ; \quad \bar{x}_2=250$$

$$s_1=20 \quad ; \quad s_2=25$$

Here,

H_0: both the machines are equally efficient

H_1: both the machines are not equally efficient

Under the null hypothesis, the test statistic is

$$t=\frac{\bar{x}_1-\bar{x}_2}{S}\sqrt{\frac{n_1n_2}{n_1+n_2}}$$

$$\text{where, } S=\sqrt{\frac{25\times20^2+25\times25^2}{25+25-2}}=23.105$$

now,

$$t=\frac{200-250}{23.105}\sqrt{\frac{25\times25}{25+25}}$$

$$t=-7.65$$

$$|t|=7.65$$

The tabulated value of t for 48 degrees of freedom at 1% level of significance is 2.58.

Since the |t| > 2.58, therefore it is highly significant. Hence, we reject the null hypothesis and conclude that both the machines are not equally efficient at 1% level of significance.

3.5.3 Paired 'T' Test For Difference Of Means

Unlike the previous test, we may sometimes face with a situation where two samples x_1, x_2,……., x_n and y_1, y_2,……, y_n drawn from the same population, are dependent on each other in pairs, i.e., the pair of observations (x_1,y_1), (x_2, y_2),....., (xn,yn) corresponds to the same 1st, 2nd, ..., nth unit respectively. Suppose we want to test the difference in efficacy of two drugs (one is domestically produced and another is imported) used for the same illness. Now, in order to test the significance of the difference

between the two drugs (x and y), we cannot use the test of difference of mean, instead we apply the paired 't' test given below:

Let, the difference in the observation for the ith unit be $d_i = x_1 - y_1$; where i=1, 2,.....,n

Under the null hypothesis that difference between the two drug is due to fluctuations in sampling, i.e., $H_0: \mu_x = \mu_y$, the test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

Where, $\bar{d} = \frac{\sum d}{n}$ and $S = \sqrt{\frac{\sum (d-\bar{d})^2}{n-1}} = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}$

follows student's 't' distribution with (n-1) degrees of freedom.

Example 3: A drug is given to 10 patients and the increments in their blood pressure were recorded to be 3, 6, -2, 4, -3, 4, 6, 0, 0, 2. Is it reasonable to believe that the drug has no effect on the change of blood pressure at 5% level of significance?

Solution: Here, the null hypothesis is that the drug has no effect on the change of blood pressure.

To test this hypothesis, the test statistic is given by

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

$$\text{where, } S = \sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}$$

We have

Patient	A	B	C	D	E	F	G	H	I	J	Summation
d	3	6	–2	4	–3	4	6	0	0	2	20
d^2	9	36	4	16	9	16	36	0	0	4	130

Now,

$$S = \sqrt{\frac{130 - 10(2)^2}{10-1}} = 3.162$$

$$t = \frac{2 \times \sqrt{10}}{3.162}$$

$$= 2$$

Here, the tabulated value of t for 9 degrees of freedom at 5% level

of significance is 2.26.

Since the calculated value is less than the tabulated value, we accept the null hypothesis. Hence, we conclude that it is reasonable to believe that the drug has no effect on the change of blood pressure

3.5.4 't' Test For Significance Of An Observed Sample Correlation Co-Efficient

Suppose r be the observed correlation coefficient in a sample of n pairs of observations from a bivariate normal population. Now, the underlying hypothesis to test is that the variables in the population are uncorrelated, i.e., H_0: ρ=0 (where ? is population correlation coefficient) and the test statistic for the hypothesis is

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

i.e., t follows student's '*t*' distribution with (n-2) degrees of freedom.

Example 4: A random sample of 27 pairs of observations from a normal population gives a correlation coefficient of 0.42. Is it likely that the variables in the population are uncorrelated?

Solution: Given,

n= 27

r= 0.42

The null hypothesis to be testes is

H_0: the variables in the population are uncorrelated

H_1: the variables in the population are correlated

Under the null hypothesis, the test statistic is given by

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

$$= \frac{0.42}{1-(0.42)^2}\sqrt{27-2}$$

$$= 2.31$$

The tabulated value for t for 25 degrees of freedom at 5% level of significance for a two-tailed test is 2.06. Since calculated t is greater than tabulated value, it is significant. Hence, null hypothesis is rejected at 5% level of significance and conclude that the variables in the population are correlated.



CHECK YOUR PROGRESS

Q 3: Explain the t test for testing the significance of an observed sample correlation. (Answer in about 60 words)

.....

.....

.....

.....

.....

3.6 CHI-SQUARE (χ^2) TEST

The chi-square test is one of the simplest and most widely used non-parametric tests in statistical work. It is denoted by χ^2, where the lower case Greek letter χ (Chi). The chi-square test is first used by Karl Pearson to compare the observed frequencies with the expected frequencies. It describes the magnitude of the discrepancy between theory and observation. It is defined as

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where, O refers to the observed frequencies and E refers to the expected frequencies.

An expected frequency can be calculated using the following equation:

$$\mathrm{E} = \frac{RT \times CT}{N}$$

Where, RT is the row total; CT is the column total and N is the total number of observations.

The chi-square statistic has *(r -l)(c-1)* degrees of freedom; where r and c

refers to number of rows and columns respectively. For example, if there are 4 rows and 5 columns, then degrees of freedom is (4-1)(5-1) = 12. It should be noted that the value of χ^2 is always positive and it ranges between zero (0) to infinity (∞). Again, χ^2 is derived from observations; therefore it is a statistic and not a parameter. Therefore, the χ^2 termed as Non-Parametric Test. It is also advantageous to use the χ^2 test as it does not made any assumptions about the parameters of the population.

3.6.1 Properties of Chi-Square Test Statistic

- Summation of observed frequencies and expected frequencies always becomes zero, i.e.,

$$\sum(O - E) = \sum O - \sum E = N - N = 0$$

- The χ^2 test depends on the following two factors: first, set of observed and expected frequencies; and second, number of degrees of freedom.
- The χ^2 distribution is a limiting case of the multinomial distribution.
- The χ^2 test can also be applied to discrete random variables whose frequencies can be counted and tabulated with or without grouping.

3.6.2 Conditions For Applying χ^2 Test

- The χ^2 test statistic can be used only if the following conditions are satisfied:
- Total frequency must be reasonably large.
- No theoretical frequencies should be less than 5 in any case, though 10 is better. When theoretical frequencies are less than 10 (specifically less than 5) then we cannot use χ^2 test.
- The constraints on the cell frequencies, if any, should not be linear.



CHECK YOUR PROGRESS

Q 4: What are different characteristics of Chi-square test statistic? (Answer in about 60 words)

.....

.....

.....

.....

Q 5: What are the determining factors of Chi-square test? (Answer in about 20 words)

.....

.....

Q 6: Mention the different conditions for applying Chi-square test. (Answer in about 60 words).

.....

.....

.....

.....

.....

.....

3.7 DEGREES OF FREEDOM

By the term *"degrees of freedom"*, we mean the number of classes to which the values can be assigned arbitrarily or at will without violating the restrictions or limitations placed. While comparing the calculated value of a test statistic (say χ^2) with the table value, we have to determine the degrees of freedom. Symbolically, the degrees of freedom are denoted by the symbol *v* (nu) or by d.f. and are obtained as follows:

$$v = n\text{-}k$$

Where, n and k represents the number of independent observations and independent constraints respectively.

Let us understand this with an example. Suppose we are to choose any four numbers whose total is 60, we can exercise our independent

choice for any three numbers only, the fourth number fixed by the virtue of the total being 60. Thus, there was only one restriction on our freedom. In this case, the degrees of freedom were only three. If more restrictions are placed, our freedom to choose will be curtailed.

3.8 YATES' CORRECTIONS FOR CONTINUITY

One of the conditions of applying χ^2 test is the cell frequencies should not be less than 5. However, when cell frequency in 2×2 table is less than 5, we apply the correction method introduced by F. Yates; popularly known a Yates' correction for continuity. This method involves reduction in the deviation of observed frequencies from theoretical frequencies, which leads to reduction in the value of χ^2. In a 2×2 table, it adjusts the observed frequencies in each cell in such a way as to reduce absolute deviation of the observed frequency from the theoretical frequency for that cell $\text{by}\frac{1}{2}$. The adjustments for all the cells are to be made without changing the marginal totals.



CHECK YOUR PROGRESS

Q 7: What do you mean by degrees of freedom? (Answer in about 30 words)

.....

.....

.....

3.9 APPLICATIONS OF CHI-SQUARE (X^2) TEST

The chi-square tests can be applicable to a wide range of problems in practice, some of which are mentioned in this section:

3.9.1 Chi-Square Test Of Goodness of Fit

Chi-square test is popularly known as test of goodness of fit. It helps us to known whether the theoretical distribution has fitted the empirical distribution or not. In this Chi-square test, we test the

null hypothesis that there is no significant difference between the observed and the theoretical or expected values. The test statistic for the underlying null hypothesis is given by

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

and it follows χ^2 distribution with (n-1) degrees of freedom.

If the computed χ^2 value is less than the corresponding tabulated value for degrees of freedom (n-1), then we fail to reject the null hypothesis at the required level of significance. This leads to the inference that the discrepancy between observed and expected values is due to fluctuations in sampling. However, if computed χ^2 value is greater than the tabulated value, we reject the null hypothesis which implies that the discrepancy between observed and expected values is not due to fluctuations in sampling.

Example 5: A die is rolled 100 times with the following distribution:

Number	1	2	3	4	5	6
Observed Frequency	17	14	20	17	17	15

At 1% level of significance, determine whether the die is uniform.

Solution: We are given,

Number of categories= 6

Total frequency=100

Here, null hypothesis is he die is uniform.

Under null hypothesis, expected frequency for each face of the die $= \frac{100}{6} = 16.67$

Number	O	E	O – E	$(O-E)^2$	$(O-E)^2/E$
1	17	16.67	0.33	0.1089	0.0065
2	14	16.67	– 2.67	7.1289	0.4276
3	20	16.67	3.33	11.0889	0.6652
4	17	16.67	0.33	0.1089	0.0065
5	17	16.67	0.33	0.1089	0.0065
6	15	16.67	–1.67	2.7889	0.1673
Summation			0		1.2796

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 1.2796$$

The tabulated value of chi-square for degrees of freedom 5 at 1%

level of significance is 15.086. Since the calculated value is less than tabulated values, it is not significant. Therefore, we fail to reject the null hypothesis at 1% level of significance. Hence, accept the null hypothesis and conclude that the die is uniform.

3.9.2 Chi-Square Test For Independence of Attributes

To find out whether two or more attributes are associated or not, we can use Chi-square test. Suppose, we have N observations in a population, which are divided into r mutually exclusive and exhaustive classes with respect to the attribute A (i.e., A_1, A_2,....., A_r). Again, the same population divided into s mutually exclusive and exhaustive classes with respect to another attribute B (i.e., B_1, B_2,....., B_s). Now, we are to test whether the attributes are related or independent. So the underlying null hypothesis would be that the two attributes are independent.

If (A_i, B_j) be the number of observations possessing both the attributes (i.e., A and B), so under the null hypothesis, the expected frequency for (A_i, B_j) is given by

$$E[(A_i,B_j)] = N.P[A_i,B_j] = N.P[A_i \cap B_j] = N.P[A_i].\ P[B_j] = N\frac{A_i}{N}\frac{B_j}{N} = \frac{A_iB_j}{N}$$

If $(A_i,B_j)_0$ denotes the expected frequencies of (A_i,B_j) then

$$(A_i,B_j)_0 = \frac{A_iB_j}{N} \qquad \text{(where, i=1,2,.....,r and j=1,2,.....,s)}$$

Since we have both observed frequencies (A_i,B_j) and the corresponding expected frequencies $(A_i,B_j)_0$ so to test the null hypothesis, the test statistic is given by

$$\chi^2 = \sum_i \sum_j \left[\frac{[(A_iB_j) - (A_iB_j)_0]^2}{(A_iB_j)_0}\right]$$

which follows χ^2 distribution with (r -1) (s -1) degrees of freedom.

3.9.3 Chi-Square Test For The Population Variance

Suppose we are asked to test whether a given normally distributed population has a specified variance or not. To test that

we set the following null hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$

And the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma_0^2}$$

which follows χ^2 distribution with (n -1) degrees of freedom, where s^2 denotes sample variance and it can be calculated using following formula:

$$s^2 = \frac{\sum(x-\bar{x})^2}{n}$$

Example 6: A random sample of size 20 from a normal population gives the sample standard deviation of 6. Test the hypothesis that the population standard deviation is 9.

Solution: In usual notation, we are given

n= 20 ; s=6; therefore s^2= 36

The null hypothesis is that the population standard deviation (σ) = 9. Under the null hypothesis,

$$\chi^2 = \frac{ns^2}{\sigma_0^2}$$

$$= \frac{20\times36}{9^2} = 8.89$$

The tabulated value of chi-square for 19 degrees of freedom at 5% level of significance is 30.144

Since, the calculated value is less than the tabulated value, it is not significant. Hence, we accept the null hypothesis 5% level of significance and conclude that population standard deviation is 9.



CHECK YOUR PROGRESS

Q 8: Mention the different uses of Chi-square test. (Answer in about 40 words)

.....

.....

.....



3.10 LET US SUM UP

- In small samples tests, our main interest is not to estimate the population values as is true in large samples; rather our interest lies in testing a given hypothesis.
- The *'t'* distribution is used when sample size is less than or equal to 30.
- *t'* distribution assumes that the population is normally distributed and its variance is unknown.
- As the number of degrees of freedom (d.f) increases, student's *'t'* distribution approaches to be standard normal distribution.
- *t*-test is used to test the significance of the mean of a random sample, difference between means of two samples (both independent samples and dependent samples) and significance of an observed correlation co-efficient.
- χ^2 test is a non-parametric test as it is derived from observations; therefore it is a statistic and not a parameter.
- χ^2 test is used to test the discrepancy between theory and observation.
- The χ^2 test depends on the following two factors: first, set of observed and expected frequencies; and second, number of degrees of freedom.
- No theoretical frequencies should be less than 5 in any case, though 10 is better. Otherwise, we cannot use χ^2 test.
- The term "degrees of freedom" refers to the number of classes to which the values can be assigned arbitrarily or at will without violating the restrictions or limitations placed.
- When cell frequency in 2×2 table is less than 5, we apply Yates' correction method.
- Yates' correction method involves reduction in the deviation of observed frequencies from theoretical frequencies, which leads to reduction in the value of χ^2.

- The goodness of fit test helps us to known how well the theoretical distribution has fitted the empirical distribution.



3.11 FURTHER READINGS

1) S.P. Gupta, Statistical Methods, 4th edition, Sultan Chand & Sons.
2) S.P. Gupta and M.P. Gupta, Business Statistics
3) S.C. Gupta, Fundamentals of Statistics, 7th edition, Himalaya Publishing House.
4) J. L. Devore, Probability and Statistics for Engineering and the Sciences, 8th edition, Cengage Learning



3.12 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: when sample size is small, i.e., less than or equal to 30, we apply *t*- test.

Ans to Q No 2: The relationship between *'t'* distribution and standard normal distribution is as follows: the tails of the *'t'* distribution have greater area (or probability) than the tails of the standard normal distribution. Therefore, interval widths from *'t'* distribution are wider than those based on the standard normal distribution. However, *'t'* distribution has less area (probability) in the center than the standard normal distribution. Again, as the number of degrees of freedom (d.f) increases, i.e., for large v (practically $v > 30$), *'t'* distribution tends to be standard normal distribution.

Ans to Q No 3: Go through section 3.5.4 and answer.

Ans to Q No 4: Different characteristics of chi-square test statistic are:

- The value of χ^2 is always positive and it ranges between zero (0) to infinity (∞).
- χ^2 is derived from observations; therefore it is a statistic and not a parameter.
- It is also advantageous to use the χ^2 test as it does not made

any assumptions about the parameters of the population.

Ans to Q No 5: Determinants of χ^2 test are following:

Set of observed and expected frequencies; and

Number of degrees of freedom.

Ans to Q No 6: Go through section 3.6.2 and answer.

Ans to Q No 7: By the term "degrees of freedom", we mean the number of classes to which the values can be assigned arbitrarily or at will without violating the restrictions or limitations placed.

Ans to Q No 8: Different uses of Chi-square test are mentioned below:

- Chi-square test as a test of independence
- Chi-square test as a test of homogeneity.
- Chi-square test as a test of goodness fit.



3.13 MODEL QUESTIONS

Short Questions (Answer each question in about 150 words)

Q 1: Compare Student's *t*-distribution with Standard normal distribution using a suitable diagram.

Q 2: Write a short note on the concept of degrees of freedom.

Q 3: In the past the standard deviation of weights of certain 40kg packages filled by a machine was 0.25kg. A random sample of 20 packages showed a standard deviation of 0.32kg. Is the apparent increase in variability significant at 5% level of significance?

Long Questions (Answer each question in about 300-500 words)

Q 1: Discuss various applications of Chi-square test.

Q 2: Explain Student's t-distribution and its properties.

Q 3: The IQs (intelligence quotients) of 16 students from one area of a city showed a mean of 107 with a standard deviation of 10, while the IQs of 14 students from another area of the city showed a mean of 112 with a standard deviation of 8. Is there a significant difference between the IQs of the two groups at 1% and 5% level of significance?

*** ***** ***

UNIT 4 : LINEAR REGRESSION MODEL AND ITS ESTIMATION

UNIT STRUCTURE

4.1 Learning Objectives

4.2 Introduction

4.3 Two variable Linear regression model

4.3.1 Introduction of the Disturbance Term

4.4 Ordinary Least Square Method

4.5 Let Us Sum Up

4.6 Further Readings

4.7 Answers to Check your Progress

4.8 Model Questions

4.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- Know about the linear regression model.
- Discuss about the necessity of the introduction of the disturbance term
- Know about the estimation of the regression coefficients.

4.2 INTRODUCTION

One of the objectives of studies in economics and other social sciences is to determine whether a change in one variable (X) causes a change in another variable (Y) or not. For example, it is frequently debated whether decrease in corporate tax rate leads to an increase in private sector investment or not. Similarly, some researcher may be interested to empirically test Keynes' theory of consumption, which states that household's current consumption expenditure (Y) is the function of current income (X). Generally, the correlation analysis is used to test the relationship between two variables. However, correlation coefficient between two variables do not necessarily indicate the cause effect relationship.

Correlation coefficient between two variables, shows the actual effect of one variable on another, if all other things remain constant. However, in reality we do not have such a situation when all other factors remain constant. For example, although current income is one of the factors determining the consumption expenditure of a household there are many factors other than current income, those affect the consumption expenditure of a household; such as number of family members, location of the households, level of education of the family members etc. In regression analysis we try to find out the variation in consumption expenditure of a household in terms of variation in income of the household; while controlling the effect of all the other factors. Thus, regression analysis is defined as a statistical tool that explains the variation in one variable in terms of the variation in one or more than one variable (while controlling the effect of some other variables on the former).

4.3 TWO-VARIABLE LINEAR REGRESSION MODEL

In reality, an econometric model is very complicated that contains a number of variables and a numbers of equations. However, for simplicity we may start with a two variable model. Let's assume that monthly consumption expenditure (Y) of a household is the function of household income. It is written as:

$$Y = f(X)$$

Now, the relationship between monthly consumption expenditure and household income may be given by the following equation:

$$Y_t = \alpha + \beta X_t \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In the equation (1) monthly consumption expenditure (Y_t) is the dependent variable and household income (X_t) is the explanatory or independent variable. Both and b are parameters those explains the relationship between the dependent and independent variables.

4.3.1 Introduction of the Disturbance Term

Equation (1) states that given the value of the parametersa α and β, for a given value of X we have a definite value of Y.

However, such a relationship between the variables is possible to have in physical sciences (physics or chemistry), not in social sciences or economics. For example, equation (1) states that for given values of α and β, all the households with a definite level of income will have equal level of consumption expenditure. However, it does not happen in reality. In reality, two households with the equal level of income generally have different levels of consumption expenditure. To, explain this phenomenon we introduce another term in the regression model, which is known as random disturbance term indicated by U. Finally,we have the regression model:

$$Y_t = \alpha + \beta X_t + U_t$$

Noted that the disturbance term U_t is a random variable and it is not observable. We can explain its presence in the model, but do not know about its magnitude.

Rational for the Introduction of the Disturbance Term

Here, queries arise regarding the necessity of the introduction of the disturbance term in the regression model. Three points are given regarding the necessity of the disturbance term in the regression model, although they are not mutually exclusive.

- There are many variables other than the explanatory variable(s) mentioned in the model which affect the dependent variable. It is not possible to introduce those variables in the model because:
  - Some of the variables are not quantifiable.
  - Although, some of the variables are quantifiable, it is difficult to collect data for those variable.
  - Sometimes, it may be possible to quantify some variable and collect data of those variables.However, their impact on the dependent variable may be so limited that their statistical estimation will be difficult and uncertain.

To solve this problem the average effect of all such variables is given by the intercept term (α) of the regression model. And the

variation in the effect of these variables from their average effect is given by the disturbance term.

- Over and above the total effect of the relevant factors, there is a basic and unpredictable element of randomness in human responses. Although in the economic theory we assume human being to be rational, it is not always true. Many a times our responses are not rational. Such randomness in human behavior can be adequately characterized only by the introduction of a random variable term.
- The third source of error lies is the errors of observation and measurement.

  One important point is to be noted that the necessity of the introduction of the random disturbance term in the regression model leads to the existence of econometrics. Whether we may say it the devil or god, its existence only gives us the opportunity to learn econometrics.



CHECK YOUR PROGRESS

Q 1: Is it possible to observe the random disturbance term?

.....

.....

Q 1: What is regression analysis?

.....

.....

Q 3: Why some of the explanatory variables cannot be introduced in the regression model?

.....

.....

4.4 ORDINARY LEAST SQUARE METHOD

Given the regression model $Y_t = \alpha + \beta X_t + U_t$, our objective is to find out the values of the parameters α and β. As the parameter is a population

characteristics, to find out the actual values of α and β we need information relating to all the units in the population. But collecting information about the entire population is costly both in terms of time and money. So, we may estimate their values from sample data. There are different methods of estimating the values of α and β. One of the popular and widely used method is Ordinary Least Square (OLS). The OLS estimators of α and β are written as $\hat{\alpha}$ and $\hat{\beta}$. And the estimated regression model is written as : $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$. The difference between the actual/ observed value (Y_t) and estimated value ($\hat{Y}_t$) of the dependent variable is termed as the residual value, which is generally indicated by e.

$$e = Y_t - \hat{Y}_t$$

$$\text{or, } e = Y_t - \hat{\alpha} - \hat{\beta} X_t$$

Under ordinary least square (OLS) method we try to estimate such values of the estimators ($\hat{\alpha}$ and $\hat{\beta}$) that minimize the overall magnitude of the residual term. Now, the point is how to estimate the overall magnitude of the residuals.

- One of the possibilities is to sum up the values of the residuals. But the problem here is that as some of the residuals are positives and some other are negative, the positive and negative values may mutually have cancelled out and the sum total may become very small.
- Another, alternative is sum of the absolute values of the residuals ignoring their signs. But ignoring their signs make them unsuitable for further algebraic manipulation. Moreover, it gives equal weights to the smaller and larger values. But in reality, the larger values should get more weight than the smaller one.
- To overcome these problems, the overall magnitude of the residuals is estimated as sum of squares of the residuals i.e. Σe^2 . In other words, we try to estimate such values of $\hat{\alpha}$ and $\hat{\beta}$, which make Σe^2 minimum. Mathematically,

Minimise,

$$S = \Sigma e^2 = \Sigma (Y_t - \hat{Y}_t)^2, \text{ Subject to } \hat{\alpha} \text{ and } \hat{\beta}$$

We know that for minimization there are two conditions to be satisfied.

The first order condition states that,

$$\frac{\partial S}{\partial \hat{\alpha}} = 0 \qquad \frac{\partial S}{\partial \hat{\beta}} = 0$$

The second order condition for minimization states,

$$[H1] = \left|\frac{\partial^2 S}{\partial \hat{\alpha}^2}\right| > 0 \qquad \text{and } [H2] = \begin{vmatrix} \frac{\partial^2 S}{\partial \hat{\alpha}^2} & \frac{\partial^2 S}{\partial \hat{\alpha}\partial \hat{\beta}} \\ \frac{\partial^2 S}{\partial \hat{\beta}\partial \hat{\alpha}} & \frac{\partial^2 S}{\partial \hat{\beta}^2} \end{vmatrix} >$$

Now,

$$\frac{\partial S}{\partial \hat{\alpha}} = 0$$

Or, $\frac{\partial}{\partial \hat{\alpha}} \Sigma (Y_t - \hat{Y}_t)^2 = 0$

or, $\frac{\partial}{\partial \hat{\alpha}} \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2 = 0$

Or, $2 \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t) \frac{\partial}{\partial \hat{\alpha}} (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$

Or, $2 \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)(-1) = 0$

Or, $\Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$ 2

Or, $\Sigma Y_t - n.\hat{\alpha} - \hat{\beta} \Sigma X_t = 0$

Or, $n.\hat{\alpha} = \Sigma Y_t - \hat{\beta} \Sigma X_t$

Or, $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$

Similarly,

$$\frac{\partial S}{\partial \hat{\beta}} = 0$$

Or, $\frac{\partial}{\partial \hat{\beta}} \Sigma (Y_t - \hat{Y}_t)^2 = 0$

Or, $\frac{\partial}{\partial \hat{\beta}} \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2 = 0$

Or, $2\Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t) \frac{\partial}{\partial \hat{\beta}} (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$

Or, $2\Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)(-X_t) = 0$

Or, $\Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)(-X_t) = 0$

Or, $-\Sigma X_t Y_t + \hat{\alpha} \Sigma X_t + \hat{\beta} \Sigma X_t^2 = 0$

Or, $-\Sigma X_t Y_t + (\bar{Y} - \hat{\beta} \bar{X}) \Sigma X_t + \hat{\beta} \Sigma X_t^2 = 0$

Or, $-\Sigma X_t Y_t + \bar{Y} \Sigma X_t - \hat{\beta} \bar{X} \Sigma X_t + \hat{\beta} \Sigma X_t^2 = 0$

Or, $\hat{\beta} \Sigma X_t^2 - \hat{\beta} \bar{X} \Sigma X_t = \Sigma X_t Y_t - \bar{Y} \Sigma X_t$

Or, $\hat{\beta} (\Sigma X_t^2 - \bar{X} \Sigma X_t) = \Sigma X_t Y_t - \bar{Y} \Sigma X_t$

Or, $\hat{\beta} = \frac{\Sigma Xt\, Yt - \bar{Y} \Sigma Xt}{(\Sigma Xt2 - \bar{X} \Sigma Xt)}$ …….. (3)

Now, $\Sigma (X_t - \bar{X})(Y_t - \bar{Y})$

$= \Sigma (X_t Y_t - \bar{Y} X_t - \bar{X} Y_t + \bar{X}\bar{Y})$

$= \Sigma X_t Y_t - \bar{Y} \Sigma X_t - \bar{X} \Sigma Y_t + \Sigma \bar{X}\bar{Y}$

$= \Sigma X_t Y_t - \bar{Y} \Sigma X_t - \bar{X} n.\bar{Y} + n.\bar{X}\bar{Y}$

$= \Sigma X_t Y_t - \bar{Y} \Sigma X_t$ ……..(4)

Similarly, $\Sigma (X_t - \bar{X})^2$

$= \Sigma (X_t^2 - 2\bar{X} X_t - \bar{X}^2)$

$= \Sigma X_t^2 - 2\bar{X} \Sigma X_t + \Sigma \bar{X}^2$

$= \Sigma X_t^2 - 2\bar{X}.n.\bar{X} + n.\bar{X}^2$

$= \Sigma X_t^2 - 2n.\bar{X}^2 + n.\bar{X}^2$

$$=\Sigma X_t^2 - n.\bar{X}^2$$

$$=\Sigma X_t^2 - \bar{X} n.\bar{X}$$

$$= \Sigma X_t^2 - \bar{X}\Sigma X_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

Considering equations (4) and (5), equation (3) may be rewritten as

$$\hat{\beta} = \frac{\Sigma(Xt - \bar{X})(Yt - \bar{Y})}{\Sigma(Xt - \bar{X})^2}$$

$$\text{Or, } \hat{\beta} = \frac{\Sigma xy}{\Sigma x^2}$$

[The interested students may check the second order conditions for minimization].

Example: Let's illustrate with an example. Suppose marks obtained in the examination(Y) is a linear function of the number of hours studied by the student weekly(X). The regression model is written as $Y_t = \alpha + \beta X_t + U_t$. Given the following data on X and Y, estimate the regression model.

X	2	3	4	5	6	7	8
Y	5	10	10	15	20	20	25

Solution:

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2
2	5	-3	-10	30	9
3	10	-2	-5	10	4
4	10	-1	-5	5	1
5	15	0	0	0	0
6	20	1	5	5	1
7	20	2	5	10	4
8	25	3	10	30	9
ΣX =35	ΣY=105			Σxy = 90	Σx^2= 28

$$\bar{X} = \frac{\Sigma X}{n} = \frac{35}{7} = 5$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{105}{7} = 15$$

$$\hat{\beta} = \frac{\Sigma xy}{\Sigma x^2} = \frac{90}{28} = 3.21$$

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ = 15 – (3.21). 5 = -1.05

Therefore, the estimated model is : $\hat{Y}_t$= -1.05 + 3.21X_t

Suppose a student Arijit studies 12 hours in a week, so the expected marks he get in the examination is $\hat{Y}_{Arijit}$= -1.05 + 3.21(12) = -1.05 +38.52 = 40.45



CHECK YOUR PROGRESS

Q 4: Define residual.

.....

.....

Q 5: What is a parameter?

.....

.....

Q 6: Define ordinary least square method.

.....

.....



4.5 LET US SUM UP

- The above discussion gives an idea about the two variable linear regression model. Further, we have discussed about the ordinary least square (OLS) method to estimate the parameters in the model. However, the OLS method is constraint by certain assumptions. The OLS estimators exhibit the characteristics of a good estimator only when such assumptions are fulfilled. The OLS assumptions and properties of the OLS estimators are discussed in the next chapter.



4.6 FURTHER READING

1) Gujrati, D.N. (1995): Basic Econometrics, Tata McGraw Hill, New Delhi.

2) Koutsoyiannis, A. (1977): Theory of Econometrics, the MacMillan Press Ltd. , London.

3) Wooldridge, J. F. (2008): Introductory Econometrics: A Modern

Approach, Third edition, South- Western Cengage Learning India.

4) Johnston, J. (1991): Econometric Methods, MacGraw Hill Book Co. , London.



4.7 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: It is not possible to observe the existence of random disturbance term.

Ans to Q No 2: Regression is a statistical tool that explains the variation in onevariable in terms of the variation in some other variable.

Ans to Q No 3: It is not possible to introduce some of the explanatory variables in the model because those variables may not be quantifiable, data may not be available for those variable, or their impact on the dependent variable may be so limited that their statistical estimation will be difficult and uncertain.

Ans to Q No 4: Residual is the difference between the observed and estimated value of the dependent variable.

Ans to Q No 5: Parameter is a population characteristic. For example, population mean is a parameter.

Ans to Q No 6: Ordinary least square method estimates such values of the parameters which minimises the sum of the square of the residuals.



4.8 MODEL QUESTIONS

Q 1: What is the rational for the introduction of the disturbance term?

Q 2: Discuss how the values of the parameters of a regression model is estimated using ordinary least square method.

Q 3: Given the regression model $Y_t = \alpha + \beta X_t + U_t$ and the following data set, find out the estimated regression function

X	5	10	12	13	20
Y	50	65	70	75	80

UNIT 5 : CLASSICAL ASSUMPTIONS AND THE OLS ESTIMATORS

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Classical Assumptions and their implications

5.4 Properties of OLS estimators

5.5 Unbiased estimator of the variance of disturbance term

5.6 Let Us Sum Up

5.7 Further Readings

5.8 Answers to Check your Progress

5.9 Model Questions

5.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- know about the standard OLS assumptions.
- prove that the OLS estimators are best, linear and unbiased estimators.
- derive the unbiased estimator of the variance of the disturbance term.
- know about the estimation of the regression coefficients.

5.2 INTRODUCTION

In the previous unit, we have discussed about the two variable linear regression model. The regression model in our analysis was $Y_t = \alpha + \beta X_t + U_t$. As mentioned in that unit our objective is to find out the values of the parameters α and β. Then we discussed the ordinary least square (OLS) method and derived the OLS estimators of the parameters:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta} = \frac{\Sigma xtyt}{\Sigma xt^2}$$

The OLS estimator $\hat{\alpha}$ is expected to represent α and $\hat{\beta}$ is expected to represent β. An estimator will be the representative of the parameter only if it satisfies certain characteristics such as linearity, unbiasedness and efficiency (minimum variance). If an estimator satisfies all these characteristics, it is known as the best, linear, unbiased estimator (BLUE). These characteristics are known as the characteristics of a good estimator. In other words, OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ will have to satisfy these characteristics in order to represent α and β. In this unit, we will discuss the properties of the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ to know whether they are best, linear, and unbiased estimator (BLUE) or not.

5.3 CLASSICAL ASSUMPTIONS AND THEIR IMPLICATIONS

There are certain assumptions that the regression model has to satisfy and only then, the ordinary least square (OLS) method can be applied to estimate the parameters. These are:

- The random disturbance term has zero mean across different observations.

  $$E(U_t) = 0$$

  It does not mean that the value of disturbance term is zero in all observations. It may take any positive or negative value in a particular observation. However, their expected (average) value is zero.

- The disturbance term in different observations has constant variance. The variance of the disturbance term is defined as

  $E\{U_t - E(U_t)\}^2$

  $=E(U_t - 0)^2$

  $=E(U_t^2)$

  $= s^2$

  This assumption is known as the assumption of homoscedasticity or absence of heteroscedasticity. If this assumption is violated the OLS estimator is unbiased, but the property of minimum variance is no longer fulfilled. In other words, the OLS estimator will not be the best estimator.

- The disturbance terms in different observations are independently distributed. This means that the value taken by U in the observation t is not related to the value taken by U in observation s.

$$Cov(U_tU_s) = E(U_t - EU_t)(U_s - EU_s) = 0$$

This assumption is known as the absence of auto-correlation or serially auto-correlation.

- Each disturbance term is normally distributed; i.e. $U_t \sim N(0, \sigma^2)$.
- There is no randomness in the explanatory variables. This means that each X_t value is fixed and there is no question of X_t having different possible values with assigned probabilities.

$$Cov(X_t\ U_t) = 0$$

This is known as the absence of endogeneity problem.



CHECK YOUR PROGRESS

Q 1: What is heteroscedasticity?

.....

.....

Q 2: Define auto-correlation.

.....

.....

Q 3: Define the problem of endogeneity.

.....

.....

5.4 PROPERTIES OF OLS ESTIMATORS

Given the classical assumptions, which are also known as the standard OLS assumptions, now we may proceed to examine the properties of OLS estimators. In other words, we will check whether OLS estimators satisfies the BLUE properties of an estimator or not. The OLS estimator of β is given:

$$\hat{\beta} = \frac{\Sigma xtyt}{\Sigma xt^2}$$

$$\hat{\beta} = \frac{\Sigma xt(Yt-\bar{Y})}{\Sigma xt^2}$$

$$\hat{\beta} = \frac{\Sigma xtYt}{\Sigma xt^2} - \bar{Y}\frac{\Sigma xt}{\Sigma xt^2}$$

$$\hat{\beta} = \frac{\Sigma xt}{\Sigma xt^2} Y_t \quad , \text{as } \Sigma x_t = 0$$

$$\hat{\beta} = \Sigma\left(\frac{xt}{\Sigma xt^2}\right) Y_t$$

$$\hat{\beta} = \Sigma W_t Y_t \quad , \text{where } Wt = \frac{xt}{\Sigma xt^2}$$

Thus, $\hat{\beta}$ is a linear combination of sample Y values. So, $\hat{\beta}$ is a linear estimator of β .

Before, examining the other properties of $\hat{\beta}$, it will be wise to note a few results about Wt.

Given, $Wt = \frac{xt}{\Sigma xt^2}$

$$\Sigma W_t = \frac{\Sigma xt}{\Sigma xt^2} = \frac{0}{\Sigma xt^2} = 0$$

$$\Sigma W_t X_t = \frac{\Sigma xt}{\Sigma xt^2} . x_t = \frac{\Sigma xt^2}{\Sigma xt^2} = 1$$

$$\Sigma W_t^2 = \Sigma \left(\frac{xt}{\Sigma xt^2}\right)^2 = \frac{1}{(\Sigma xt^2)}$$

Now,

$$\hat{\beta} = \Sigma W_t Y_t$$

$$\hat{\beta} = \Sigma W_t (\alpha + \beta X_t + U_t)$$

$$\hat{\beta} = \alpha \Sigma W_t + \beta \Sigma W_t X_t + \Sigma W_t U_t$$

Given, $\Sigma W_t = 0$ and $\Sigma W_t X_t = 1$

$$\hat{\beta} = \beta + \Sigma W_t U_t$$

$$E(\hat{\beta}) = E(\beta + \Sigma W_t U_t)$$

$$E(\hat{\beta}) = E(\beta) + \Sigma W_t E(U_t)$$

Given, Wt is a constant and $E(U_t) = 0$

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

It states that $\hat{\beta}$ is an unbiased estimator of β.

Now, we have to show that the variance of $\hat{\beta}$ is lower than the variance of any other estimator of $\tilde{\beta}$ First we will have to find out the variance of $\hat{\beta}$; which is given as:

$Var(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2$

$Var(\hat{\beta}) = E[\beta + \Sigma W_t U_t - \beta]^2$

$Var(\hat{\beta}) = E[\Sigma W_t U_t]^2$

Where, $[\Sigma W_t U_t]^2 = [W_1U_1 + W_2U_2 + W_3U_3 + \ldots\ldots\ldots\ldots\ldots + W_nU_n]^2$

Or, $[\Sigma WtU_t]^2 = W_1^2U_1^2 + W_2^2U_2^2 + W_3^2U_3^2 + \ldots\ldots\ldots + W_n^2U_n^2 + 2\,W_1U_1\,W_2U_2 + 2\,W_1U_1\,W_3U_3 + \ldots\ldots\ldots\ldots\ldots$

Or, $[\Sigma W_tU_t]^2 = [\Sigma W_t^2U_t^2 + \Sigma W_tU_tW_sU_s]$ where, $t \neq s$

Therefore,

$Var(\hat{\beta}) = E[\Sigma W_t^2U_t^2 + \Sigma W_tU_tW_sU_s]$

$Var(\hat{\beta}) = \Sigma W_t^2\,EU_t^2 + \Sigma W_tW_sE(U_tU_s)$ where, $E(U_tU_s) = 0$

$Var(\hat{\beta}) = \Sigma W_t^2 \cdot \sigma^2$

$Var(\hat{\beta}) = \sigma^2\Sigma W_t^2$

$Var(\hat{\beta}) = \sigma^2 \cdot \frac{1}{(\Sigma xt^2)}$ As, $\Sigma W_t^2 = \frac{1}{(\Sigma xt^2)}$

Now, to prove that the variance of $\tilde{\beta}$ is lower than the any other estimator of β, let 'b' is another liner estimator (other than OLS estimator) of β.

Where, $b = \Sigma C_tY_t$

The difference between C_t and Wt is d_t. or $C_t = W_t + d_t$

Now, expanding b we have:

$b = \Sigma C_t (\alpha + \beta X_t + U_t)$

$b = \alpha\,\Sigma C_t + \beta\Sigma C_t X_t + \Sigma C_t U_t$

Now,

$E(b) = E(\alpha\,\Sigma C_t + \beta\Sigma C_t X_t + \Sigma C_t U_t)$

Or, $E(b) = \alpha\,\Sigma C_t + \beta\Sigma C_t X_t + \Sigma C_t E(U_t)$

Or, $E(b) = \alpha\,\Sigma C_t + \beta.\Sigma C_t X_t$ As, $E(U_t) = 0$

'b' will be the unbiased estimator of β, if E(b) = β

And E(b) =β, if and only if $\Sigma C_t = 0$ and $\Sigma C_t X_t = 1$

In that case, $\Sigma W_t + \Sigma d_t = 0$ and $\Sigma W_t X_t + \Sigma d_t X_t = 1$

Or, $\Sigma d_t = 0$ and $\Sigma d_t X_t = 0$; as we already know $\Sigma W_t = 0$ and $\Sigma W_t X_t = 1$

In such situation,

$$b = \alpha \Sigma C_t + \beta \Sigma C_t X_t + \Sigma C_t U_t$$

$$b = \alpha . 0 + \beta . 1 + \Sigma C_t U_t$$

$$b = \beta + \Sigma C_t U_t$$

Now, the variance of 'b' is given as:

$$\text{Var (b)} = E [b - E(b)]^2$$

$$\text{Or Var (b)} = E [\beta + \Sigma C_t U_t - \beta]^2$$

$$\text{Or Var (b)} = E (\Sigma C_t U_t)^2$$

$$\text{or, Var (b)} = E (\Sigma C_t^2 U_t^2 + \Sigma C_t C_s U_t U_s)$$

$$\text{or, Var (b)} = \Sigma C_t^2 E (U_t^2) + \Sigma C_t C_s E (U_t U_s)$$

$$\text{or, Var (b)} = \Sigma C_t^2 . \sigma^2$$

$$\text{or, Var (b)} = \sigma^2 . \Sigma C_t^2$$

$$\text{or, Var (b)} = \sigma^2 \Sigma (W_t + d_t)^2$$

$$\text{or, Var (b)} = \sigma^2 \Sigma (W_t^2 + d_t^2 + 2 W_t d_t)$$

$$\text{or, Var (b)} = \sigma^2 \Sigma W_t^2 + \sigma^2 \Sigma d_t^2 + 2\sigma^2 \Sigma W_t d_t$$

Here, $\Sigma W_t d_t = \Sigma \frac{xt}{\Sigma xt^2} . d_t$

Or, $\Sigma W_t d_t = \frac{1}{\Sigma xt^2} \Sigma (X_t - \bar{X}) d_t$

Or, $\Sigma W_t d_t = \frac{1}{\Sigma xt^2} (\Sigma X_t d_t - \bar{X} \Sigma d_t) = 0$

As, $\Sigma X_t d_t = 0 = \bar{X} \Sigma d_t$, for b to be an unbiased estimator of β.

Thus,

$$\text{Var (b)} = \sigma^2 \Sigma W_t^2 + \sigma^2 \Sigma d_t^2$$

$$\text{Or, Var (b)} = \text{Var} (\hat{\beta}) + \sigma^2 \Sigma d_t^2$$

Here, σ^2 is always positive and Σd_t^2 being sum of squared terms is always non- negative. In other words, $\sigma^2 \Sigma d_t^2$ is always greater than equal to zero, i.e. $\sigma^2 \Sigma d_t^2 \geq 0$. It means that variance of b is greater than equal to the variance of $\hat{\beta}$; i.e. Var (b) $\geq$ Var ($\hat{\beta}$).

And Var (b) = Var ($\hat{\beta}$), if and only if $s^2 \Sigma d_t^2 = 0$, or $\Sigma d_t^2 = 0$. It will be the case when d_t are all individually equal to 0. And it is possible, when $C_t = W_t$ or $b = \hat{\beta}$. Hence it is proved that $\hat{\beta}$ is the best linear unbiased estimator of $\hat{\beta}$. In other words, under the standard OLS assumptions, the OLS estimators are best, linear, unbiased estimator. In econometric literature it is known as the Gauss Markov theorem.

Properties of $\hat{\alpha}$

The OLS estimator of α is $\hat{\alpha}$.

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

Where, $\hat{\beta} = \Sigma W_t Y_t$

So, the $\hat{\alpha}$ may be written as:

$\hat{\alpha} = \bar{Y} - \Sigma W_t Y_t . \bar{X}$

Or, $\hat{\alpha} = \frac{1}{n}\Sigma Y_t - \bar{X} . \Sigma W_t Y_t$

Or, $\hat{\alpha} = \Sigma(\frac{1}{n} - \bar{X} . W_t) Y_t$

Now, denoting $(\frac{1}{n} - \bar{X} . W_t)$ by R_t we have

$\hat{\alpha} = \Sigma R_t Y_t$

Therefore, like $\hat{\beta}$, $\hat{\alpha}$ is also linear in Y_t.

Where,

$\Sigma R_t = \Sigma(\frac{1}{n} - \bar{X} . W_t) = n . \frac{1}{n} - \bar{X} . \Sigma W_t = 1 - 0 = 1$

And, $\Sigma R_t X_t = \Sigma(\frac{1}{n} - \bar{X} . W_t) X_t = \Sigma X_t \frac{1}{n} - \bar{X} . \Sigma W_t X_t = \bar{X} - \bar{X} = 0$

Now,

$\hat{\alpha} = \Sigma R_t Y_t$

Or, $\hat{\alpha} = \Sigma R_t(\alpha + \beta X_t + U_t)$

Or, $\hat{\alpha} = \alpha \Sigma R_t + \beta \Sigma X_t R_t + \Sigma R_t U_t$

Or, $\hat{\alpha} = \alpha + \Sigma R_t U_t$

Or, $E(\hat{\alpha}) = E(\alpha + \Sigma R_t U_t)$

Or, $E(\hat{\alpha}) = \alpha + \Sigma R_t E U_t$

Or, $E(\hat{\alpha}) = \alpha$

Thus, $\hat{\alpha}$ is an unbiased ~~unbiased~~ estimator of α.

Variance of $\hat{\alpha}$:

$Var(\hat{\alpha}) = E[\hat{\alpha} - E(\hat{\alpha})]^2$

$= E[\alpha + \Sigma R_t U_t - \alpha]^2$

$= E[\Sigma R_t U_t]^2$

$= E[\Sigma R_t U_t]^2$

$= E(\Sigma R_t^2 U_t^2 + \Sigma R_t R_s U_t U_s)$ where, $t \neq s$

$= E(\Sigma R_t^2 U_t^2)$

$= \Sigma R_t^2 E U_t^2$

$= \Sigma R_t^2 \sigma^2$

$= \Sigma(\frac{1}{n} - \bar{X}. W_t)^2 \sigma^2$

Simplifying, it we will have

$$Var(\hat{\alpha}) = [\frac{1}{n} + \frac{1}{(\Sigma x t^2)} \bar{X}^2]. \sigma^2$$

Interested students may prove that the variance of $\hat{\alpha}$ is lower than any other estimator of α.



CHECK YOUR PROGRESS

Q 4: Define the BLUE property of an OLS estimator.

.....

.....

Q 5: What is meant by an efficient estimator?

.....

.....

5.5 ESTIMATION OF THE VARIANCE OF THE DISTURBANCE TERM (σ^2)

In the above discussion regarding the BLUE properties of the OLS estimators, the variance of $\hat{\alpha}$ and $\hat{\beta}$ contains an unknown parameter i.e. σ^2. Thus, if we have to estimate the variance of $\hat{\beta}$, the estimation of σ^2 is necessary. Here, σ^2 is the variance of the disturbance term. As discussed above, the disturbance term is unobservable. So, σ^2 cannot be estimated from the disturbance term. It is assumed that the OLS residuals (e) give some idea about the disturbance term, and we estimate σ^2 with the OLS residuals.

The regression model is given,

$Y_t = \alpha + \beta X_t + U_t$ …….. (5.1)

Or, $\bar{Y} = \alpha + \beta \bar{X} + \bar{U}$ ……...(5.2)

Now, (5.1) – (5.2) gives

$Y_t - \bar{Y} = \alpha + \beta X_t + U_t - (\alpha + \beta \bar{X} + \bar{U})$

$y_t = \alpha + \beta X_t + U_t - \alpha - \beta \bar{X} - \bar{U}$

$y_t = \beta(X_t - \bar{X}) + (U_t - \bar{U})$

$y_t = x_t\beta + (U_t - \bar{U})$ ……(5.3)

Again, the residual is given,

$e = Y_t - \bar{Y}$

$Y_t = \bar{Y} + e_t$

$Y_t = \hat{\alpha} + \hat{\beta} X_t + e_t$ …….. (5.4)

Or , $\Sigma Y_t = \Sigma\hat{\alpha} + \hat{\beta}\Sigma X_t + \Sigma e_t$

As, Σe_t is equal to to zero,

$\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ ……..(5.5)

Now, (5.4) – (5.5) gives

$Y_t - \bar{Y} = \hat{\alpha} + \hat{\beta} X_t + e_t - \hat{\alpha} - \hat{\beta}\bar{X}$

$y_t = \hat{\beta} x_t + e_t$ ……(5.6)

Now, substituting equation (5.6) in equation (5.3) we have

$\hat{\beta} x_t + e_t = x_t\beta + (U_t - \bar{U})$

Or, $e_t = x_t\beta + (U_t - \bar{U}) - \hat{\beta} x_t$

Or, $e_t = (U_t - \bar{U}) - \hat{\beta} x_t + x_t\beta$

Or, $e_t = (U_t - \bar{U}) - (\hat{\beta} - \beta) x_t$

Or, $\Sigma e_t^2 = \Sigma[(U_t - \bar{U}) - (\hat{\beta} - \beta)x_t]^2$

Or, $\Sigma e_t^2 = \Sigma(U_t - \bar{U})^2 + (\hat{\beta} - \beta)^2 \Sigma x_t^2 - 2(\hat{\beta} - \beta)\Sigma x_t(U_t - \bar{U})$

Now taking expected value of both sides, we have

$E(\Sigma e_t^2) = E[\Sigma(U_t - \bar{U})^2 + (\hat{\beta} - \beta)^2 \Sigma x_t^2 - 2(\hat{\beta} - \beta)\Sigma x_t(U_t - \bar{U})]$

$E(\Sigma e_t^2) = E\Sigma(U_t - \bar{U})^2 + \Sigma x_t^2 E(\hat{\beta} - \beta)^2 - 2E[(\hat{\beta} - \beta)\Sigma x_t(U_t - \bar{U})]$ ……..(5.7)

In equation (5.7),

$E(\hat{\beta} - \beta)^2 = E[\hat{\beta} - E(\hat{\beta})]^2 = Var(\hat{\beta}) = \sigma^2 . \frac{1}{(\Sigma x t^2)}$ …….. (5.8)

And,

$E\Sigma(U_t - \bar{U})^2$

$= E\Sigma (U_t^2 + \bar{U}^2 - 2 U_t . \bar{U})$

$= E (\Sigma U_t^2 + \Sigma \bar{U}^2 - 2\Sigma U_t . \bar{U})$

$= E[\Sigma U_t^2 + n.(\frac{\Sigma Ut}{n})^2 - 2\Sigma U_t.(\frac{\Sigma Ut}{n})]$

$= E[\Sigma U_t^2 + \Sigma U_t^2 \frac{1}{n} - 2\Sigma U_t^2 . \frac{1}{n}]$

$= E [\Sigma U_t^2 - \Sigma U_t^2 . \frac{1}{n}]$

$= E [n.U_t^2 - n.U_t^2 . \frac{1}{n}]$

$= n. E(U_t^2) - E (U_t^2)$

$= n. \sigma^2 - \sigma^2$

$= \sigma^2 (n-1)$ …….. (5.9)

Again,

$2E[(\hat{\beta} - \beta)\Sigma x_t(U_t - \bar{U})$

$= 2E\Sigma W_t x_t U_t^2$ ……. (5.10)

[As, $(\hat{\beta} - \beta) = \Sigma W_t U_t$ and $\Sigma x_t (U_t - \bar{U}) = \Sigma x_t U_t$]

Now, substituting equations (5.8), (5.9) and (5.10) in equation (5.7), we have

$E(\Sigma e_t^2) = \sigma^2 (n-1) + \Sigma x_t^2 \sigma^2 . \frac{1}{(\Sigma x t^2)} - 2E\Sigma W_t x_t U_t^2$

$E(\Sigma e_t^2) = \sigma^2 (n-1) + \sigma^2 - 2.EU_t^2$ As, $\Sigma W_t x_t = 1$

$E(\Sigma e_t^2) = n. \sigma^2 - \sigma^2 + \sigma^2 - 2. \sigma^2$

$E(\Sigma e_t^2) = n. \sigma^2 - 2.\sigma^2$

$E(\Sigma e_t^2) = \sigma^2 (n - 2)$

$\sigma^2 = \frac{1}{(n-2)} . E(\Sigma e_t^2)$

$\sigma^2 = E [\frac{1}{(n-2)} . \Sigma e_t^2]$

That will be the unbiased estimator of σ^2, whose expected value is equal to σ^2. Here the expected value of $(\frac{1}{(n-2)} . \Sigma e_t^2)$ is equal to σ^2. So, we may conclude that the unbiased estimator of? the variance of the disturbance term (σ^2) is

$$\hat{\sigma}^2 = .e_t^2$$

The value of $\hat{\sigma}^2$ can be used for estimating the variance of $\hat{\beta}$.



5.6 LET US SUM UP

- The above discussion gives a clear idea about the standard OLS assumptions and also about the properties of the OLS estimator. It is proved that the OLS estimators are the best, linear and unbiased estimator of the parameters; and also the unbiased estimator of the variance of the disturbance term is derived. However, the two variable linear regression model is the simplest form of regression model. In reality, the changes in the dependent variable can be adequately explained in terms of more than one explanatory variable. The regression model that includes more than one explanatory variable is known as the general linear regression model, which will be discussed in the next unit.



5.7 FURTHER READINGS

1) Gujrati, D.N. (1995): *Basic Econometrics*, Tata McGraw Hill, New Delhi.
2) Koutsoyiannis, A. (1977): *Theory of Econometrics*, the MacMillan Press Ltd. , London.
3) Wooldridge, J. F. (2008): *Introductory Econometrics*: A Modern Approach, Third edition, South- Western Cengage Learning India.
4) Johnston, J. (1991): *Econometric Methods*, McGraw Hill Book Co., London.



5.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: One of the standard OLS assumptions is that the disturbance term has constant variance. If this assumption is violated, the problem arises is known as heteroscedasticity.

Ans to Q No 2: If the disturbance terms in different observations are serially correlated to each other, then the problem of auto correlation arises.

Ans to Q No 3: One of the standard OLS assumptions is that there is no randomness in the explanatory variables. In other words, explanatory variable is not related to the disturbance term. The violation of this assumption, i.e. if there is any randomness in the explanatory variable(s), then the endogeneity problem arises.

Ans to Q No 4: The BLUE property of the OLS estimator means that given the standard OLS assumptions, the OLS estimator is the linear and unbiased estimator of the parameter, and it has minimum variance compared to any other estimator of that parameter.

Ans to Q No 5: An efficient estimator means that ~~if~~ it has minimum variance compared to any other estimator of the parameter.



5.9 MODEL QUESTIONS

Q 1: What are the standard classical (OLS) assumptions? Explain their consequences.

Q 2: Given the two variable regression model $Y_t = \alpha + \beta X_t + U_t$

- Find out the unbiased estimator of β .
- Derive the variance of the OLS estimator of β .

Q 3: Show that OLS estimator is the best, linear and unbiased estimator (BLUE).

Q 4: Given the regression model $Y_t = \alpha + \beta X_t + U_t$ and the following data set:

X	5	10	12	13	20
Y	50	65	70	75	80

- Estimate the unbiased estimator of the variance of the disturbance term.

*** ***** ***

UNIT 6 : MULTIPLE REGRESSION ANALYSIS

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Multiple Regression Examples

6.4 Mechanics and Interpretation of Ordinary Least Squares

6.5 Interpreting the OLS Regression Equation

6.6 Let Us Sum Up

6.7 Further Readings

6.8 Answers to Check Your Progress

6.9 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- The importance between Multiple Regression Analysis
- The difference between Simple and Multiple Regression Analysis
- Estimation of Multiple Regression Coefficients
- Interpretation of Multiple Regression Coefficients

6.2 INTRODUCTION

In simple regression analysis, we treat a dependent variable, *y*, as a function of a single independent variable, *x*. The primary drawback in using simple regression analysis for empirical work is that it is very difficult to analyse how *x* affects *y*, assuming ceteris paribus conclusions, or other things remaining same. In the real world, y may be affected by multiple factors, besides x.

Multiple regression analysis is more amenable to ceteris paribus analysis because it allows us to explicitly control for many other factors that simultaneously affect the dependent variable. This is important both for testing economic theories and for evaluating policy effects when we must rely on non-experimental data. Because multiple regression models

can accommodate many explanatory variables that may be correlated, we can hope to infer causality in cases where simple regression analysis would be misleading. Naturally, if we add more factors to our model that are useful for explaining *y*, then more of the variation in *y* can be explained. Thus, multiple regression analysis can be used to build better models for predicting the dependent variable.

An additional advantage of multiple regression analysis is that it can incorporate fairly general functional form relationships. In the simple regression model, only one function of a single explanatory variable can appear in the equation. As we will see, the multiple regression model allows for much more flexibility.

6.3 MULTIPLE REGRESSION EXAMPLE

The Model with Two Independent Variables

We begin with some simple examples to show how multiple regression analysis can be used to solve problems that cannot be solved by simple regression.

$$wage = b_0 + b_1 educ + u, \qquad (6.1)$$

The first example is a simple variation of the wage equation in (6.1) for obtaining the effect of education on hourly wage:

$$wage = b_0 + b_1 educ + b_2 exper + u, \qquad (6.2)$$

where *exper* is years of labor market experience. Thus, *wage* is determined by the two explanatory or independent variables, education and experience, and by other unobserved factors, which are contained in *u*. We are still primarily interested in the effect of *educ* on *wage*, holding fixed all other factors affecting *wage*; that is, we are interested in the parameter b_1.

Compared with a simple regression analysis relating *wage* to *educ* in eq (6.1), equation (6.2) effectively takes *exper* out of the error term and puts it explicitly in the equation. Because *exper* appears in the equation, its coefficient, b_2, measures the ceteris paribus effect of *exper* on *wage*, which is also of some interest.

Not surprisingly, just as with simple regression in eq (6.1), we will have to make assumptions about how *u* in (6.2) is related to the independent

variables, *educ* and *exper*. However, as we will see in subsequent section, there is one thing of which we can be confident: because (6.2) contains experience explicitly, we will be able to measure the effect of education on wage, holding experience fixed. In a simple regression analysis—which puts *exper* in the error term—we would have to assume that experience is uncorrelated with education, a tenuous assumption.

In the previous example, we have shown how observable factors other than the variable of primary interest *educ* in equation (6.2) and can be included in a regression model. Generally, we can write a model with two independent variables as

$$y = b_0 + b_1 x_1 + b_2 x_2 + u, \quad \text{——————————} \quad (6.3)$$

where

b_0 is the intercept.

b_1 measures the change in y with respect to x_1, holding other factors fixed.

b_2 measures the change in y with respect to x_2, holding other factors fixed.

Multiple regression analysis is also useful for generalizing functional relationships between variables. As an example, suppose family consumption (*cons*) is a quadratic function of family income (*inc*):

$$cons = b_0 + b_1 inc + b_2 inc^2 + u, \quad \text{——————————} \quad (6.4)$$

where u contains other factors affecting consumption. In this model, consumption depends on only one observed factor, income; so it might seem that it can be handled in a simple regression framework. But the model falls outside simple regression because it contains two functions of income, *inc* and inc^2 (and therefore three parameters, b_0, b_1, and b_2).

Nevertheless, the consumption function is easily written as a regression model with two independent variables by letting $x_1 = inc$ and $x_2 = inc^2$.

In the model with two independent variables, the key assumption about how u is related to x_1 and x_2 is

$$E(u|x_1, x_2) = 0. \quad \text{——————————} \quad (6.5)$$

The interpretation of condition (6.5) is that, for any values of $x1$ and $x2$ in the population, the average of the unobserved factors is equal to zero. As with simple regression, the important part of the assumption is that the expected value of u is the same for all combinations of x_1 and x_2; that this

common value is zero is no assumption at all as long as the intercept b_0 is included in the model.

How can we interpret the zero conditional mean assumption in the previous examples? In equation (6.2), the assumption is E(*u*| *educ*, *exper*) = 0. This implies that other factors affecting *wage* are not related on average to *educ* and *exper*. Therefore, if we think innate ability is part of *u*, then we will need average ability levels to be the same across all combinations of education and experience in the working population. This may or may not be true, but, as we will see later, this is the question we need to ask in order to determine whether the method of ordinary least squares produces unbiased estimators.

When applied to the quadratic consumption function in (4), the zero conditional mean assumption has a slightly different interpretation. Written literally, equation (5) becomes E(u|inc,inc^2) = 0. Since inc^2 is known when *inc* is known, including inc^2 in the expectation is redundant:

E(u|inc,inc2) = 0 is the same

as E(u|inc) = 0. Nothing is wrong with putting inc2 along with *inc* in the expectation when stating the assumption, but E(u/|inc) = 0 is more concise.

The Model with *k* Independent Variables : Once we are in the context of multiple regression, there is no need to stop with two independent variables. Multiple regression analysis allows many observed factors to affect *y*. In the wage example, we might also include amount of job training, years of tenure with the current employer, measures of ability, and even demographic variables like the number of siblings or mother's education. In the school funding example, additional variables might include measures of teacher quality and school size.

The general **multiple linear regression model** (also called the *multiple regression model*) can be written in the population as

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k + u, \quad \text{——————————(6.6)}$$

where

b_0 is the intercept.

b_1 is the parameter associated with x_1.

b_2 is the parameter associated with x_2, and so on.

Since there are k independent variables and an intercept, equation (6.6) contains k 1 (unknown) population parameters. For shorthand purposes, we will sometimes refer to the parameters other than the intercept as **slope parameters**, even though this is not always literally what they are. [See equation (6.4), where neither b_1 nor b_2 is itself a slope, but together they determine the slope of the relationship between consumption and income.] Just as in simple regression, the variable u is the **error term** or **disturbance**. It contains factors other than $x_1, x_2, \ldots, x_k$ that affect y. No matter how many explanatory variables we include in our model, there will always be factors we cannot include, and these are collectively contained in u.

The term "linear" in a multiple linear regression model means that equation (6.6) is linear in the *parameters*, bj. Many applications of multiple linear regressions involve nonlinear relationships among the underlying variables.

The key assumption for the general multiple regression model is easy to state in terms of a conditional expectation:

$$E(u|x_1, x_2, \ldots, x_k) = 0. \qquad (6.7)$$

At a minimum, equation (6.7) requires that all factors in the unobserved error term be uncorrelated with the explanatory variables. It also means that we have correctly accounted for the functional relationships between the explained and explanatory variables. Any problem that causes u to be correlated with any of the independent variables causes (6.7) to fail.

6.4 MECHANICS AND INTERPRETATION OF ORDINARY LEAST SQUARES

We now summarize some computational and algebraic features of the method of ordinary least squares as it applies to a particular set of data. We also discuss how to interpret the estimated equation.

Obtaining the OLS Estimates

We first consider estimating the model with two independent variables. The estimated OLS equation is written in a form similar to the simple regression case:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2, \qquad (6.8)$$

where

$\hat{b}_0$ = the estimate of b_0.

$\hat{b}_1$ = the estimate of b_1.

$\hat{b}_2$ = the estimate of b_2.

But how do we obtain $\hat{b}_0$, $\hat{b}_1$, and $\hat{b}_2$? The method of **ordinary least squares** chooses the estimates to minimize the sum of squared residuals. That is, given n observations on y, x_1, and x_2, $\{(xi_1, xi_2, yi): i = 1, 2, \ldots, n\}$, the estimates $\hat{b}_0$, $\hat{b}_1$, and $\hat{b}_2$ are chosen simultaneously to make

$$\sum_{i=1}^{n} (yi - \hat{b}_0 - \hat{b}_1 xi_1 - \hat{b}_2 xi_2)^2 \qquad \text{——————————(6.9)}$$

as small as possible.

To understand what OLS is doing, it is important to master the meaning of the indexing of the independent variables in (6.9). The independent variables have two subscripts here, *i* followed by either 1 or 2. The *i* subscript refers to the observation number. Thus, the sum in (9) is over all $i = 1$ to n observations. The second index is simply a method of distinguishing between different independent variables. In the example relating *wage* to *educ* and *exper*, $xi_1 = educ_i$ is education for person *i* in the sample, and $xi_2 = experi$ is experience for person *i*. The sum of squared residuals in equation (6.9) is

$$\sum_{i=1}^{n} (wage_i - \hat{b}_0 - \hat{b}_1 educ_i - \hat{b}_2 exper_i)^2$$

In what follows, the *i* subscript is reserved for indexing the observation number. If we write x_{ij}, then this means the *i*th observation on the *j*th dependent variable. (Some authors prefer to switch the order of the observation number and the variable number, so that x_{1i} is observation *i* on variable one. But this is just a matter of notational taste.)

In the general case with k independent variables, we seek estimates $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$ in the equation

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \ldots + \hat{b}_k x_k. \qquad \text{——————————} (6.10)$$

The OLS estimates, $k + 1$ of them, are chosen to minimize the sum of squared residuals:

$$\sum_{i=1}^{n} (yi - \hat{b}_0 - \hat{b}_1 - xi_1 - \ldots - \hat{b}_k x_{ik})^2 \qquad (6.11)$$

This minimization problem can be solved using multivariable calculus. This leads to $k + 1$ linear equations in $k + 1$ unknowns $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$:

$$\sum_{i=1}^{n}(yi - \hat{b}_0 - \hat{b}_1 xi_1 - \ldots - \hat{b}_k xi_k) = 0$$

$$\sum_{i=1}^{n} xi\ 1(yi - \hat{b}_0 - \hat{b}_1 xi_1 - \ldots - \hat{b}_k xi_k) = 0$$

$$\sum_{i=1}^{n} xi_2(yi - \hat{b}_0 - \hat{b}_1 xi_1 - \ldots - \hat{b}_k xi_k) = 0 \qquad (6.12)$$

$$\sum_{i=1}^{n} xi_k(yi - \hat{b}_0 - \hat{b}_1 xi_1 - \ldots - \hat{b}_k xi_k) = 0.$$

These are often called the OLS **first order conditions**. As with the simple regression model, the OLS first order conditions can be obtained by the method of moments: under assumption (6.7), $E(u) = 0$ and $E(x_j u) = 0$, where $j = 1, 2, \ldots, k$. The equations in (6.12) are the sample counterparts of these population moments, although we have omitted the division by the sample size n.

For even moderately sized n and k, solving the equations in (6.12) by hand calculations is tedious. Nevertheless, modern computers running standard statistics and econometrics software can solve these equations with large n and k very quickly.

There is only one slight caveat: we must assume that the equations in (6.12) can be solved *uniquely* for the $\hat{b}_j$. For now, we just assume this, as it is usually the case in well specified models. In Section 3.3, we state the assumption needed for unique OLS estimates to exist (see Assumption MLR.3).

As in simple regression analysis, equation (10) is called the **OLS regression line** or the

sample regression function (**SRF**). We will call $\hat{b}_0$ the **OLS intercept estimate** and $\hat{b}_1, \ldots, \hat{b}_k$ the **OLS slope estimates** (corresponding to the independent variables $x_1, x_2, \ldots, x_k$).

To indicate that an OLS regression has been run, we will either write out equation (10) with y and $x_1, \ldots, x_k$ replaced by their variable names (such as *wage*, *educ*, and *exper*), or we will say that "we ran an OLS regression of y on $x_1, x_2, \ldots, x_k$" or that "we regressed y on $x_1, x_2, \ldots, x_k$." These are shorthand for saying that the method of ordinary least squares was used to obtain the OLS equation (10). Unless explicitly stated otherwise, we always estimate an intercept along with the slopes.

6.5 INTERPRETING THE OLS REGRESSION EQUATION

More important than the details underlying the computation of the $\hat{b}_j$ is the *interpretation* of the estimated equation. We begin with the case of two independent variables:

$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2$. - ——————————————————— (6.13)

The intercept $\hat{b}_0$ in equation (6.13) is the predicted value of y when x_1 =0 and $x_2 = 0$.

Sometimes, setting x_1 and x_2 both equal to zero is an interesting scenario; in other cases, it will not make sense. Nevertheless, the intercept is always needed to obtain a prediction of y from the OLS regression line, as (6.13) makes clear.

The estimates $\hat{b}_1$ and $\hat{b}_2$ have **partial effect**, or **ceteris paribus**, interpretations. From

equation (6.13), we have

$\Delta\hat{y} = \hat{b}_1 \Delta x_1 + \hat{b}_2 \Delta x_2$,

so we can obtain the predicted change in y given the changes in x_1 and x_2. (Note how the intercept has nothing to do with the changes in *y.*) In particular, when x_2 is held fixed, so that $\Delta x_2 = 0$, then $\Delta\hat{y} = \hat{b}_1 \Delta x_1$, holding x_2 fixed. The key point is that, by including x_2 in our model, we obtain a coefficient

on x_1 with a ceteris paribus interpretation. This is why multiple regression analysis is

so useful. Similarly,

$\Delta\hat{y} = \hat{b}_2 \Delta x_2$, holding x_1 fixed.

The case with more than two independent variables is similar. The OLS regression line is

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \ldots + \hat{b}_k x_k. \quad (6.14)$$

Written in terms of changes,

$$\Delta\hat{y} = \hat{b}_1 \Delta x_1 + \hat{b}_2 \Delta x_2 + \ldots + \hat{b}_k \Delta x_k. \quad (6.15)$$

The coefficient on $x1$ measures the change in $\hat{y}$ due to a one-unit increase in $x1$, holding all other independent variables fixed. That is,

$$\Delta\hat{y} = \hat{b}_1 \Delta x_1, \quad (6.16)$$

holding x_2, x_3, …, x_k fixed. Thus, we have controlled for the variables x_2, x_3, …, x_k when estimating the effect of x_1 on y. The other coefficients have a similar interpretation.

Expected Value of OLS estimators : We now turn to the statistical properties of OLS for estimating the parameters in an underlying population model. In this section, we derive the expected value of the OLS estimators. In particular, we state and discuss four assumptions, which are direct extensions of the simple regression model assumptions, under which the OLS estimators are unbiased for the population parameters. We also explicitly obtain the bias in OLS when an important variable has been omitted from the regression. The first assumption we make simply defines the multiple linear regression (MLR) model.

Assumption MLR1 Linear in Parameters : The model in the population can be written as

$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + u$

where b_0, b_1, …, b_k are the unknown parameters (constants) of interest and u is an unobserved random error or disturbance term.

Assumption MLR.2: Random Sampling : We have a random sample of n observations, $\{(xi_1, xi_2, \ldots, xi_k, yi): i = 1, 2, \ldots, n\}$, following

the population model in Assumption MLR.1.

Assumption MLR.3: No Perfect Collinearity : In the sample (and therefore

in the population), none of the independent variables is constant, and there are no *exact linear* relationships among the independent variables.

Assumption MLR.4: Zero Conditional Mean : The error *u* has an expected value of zero given any values of the independent variables.

In other words,

$E(u|x_1, x_2, ..., x_k) = 0$.

Theorem 1: Unbiasedness of OLS

Under Assumptions MLR.1 through MLR.4,

$E(\hat{b}_j) = b_j$, $j = 0, 1, \ldots, k$,

for any values of the population parameter *bj*. In other words, the OLS estimators are unbiased estimators of the population parameters.

Assumption MLR.5: Homoskedasticity : The error *u* has the same variance given any values of the explanatory variables. In other words, $Var(u|x_1, \ldots, x_k) = \sigma^2$

Theorem 2: Sampling Variances of the OLS Slope Estimators : Under Assumptions MLR.1 through MLR.5, conditional on the sample values of the Independent variables,

$Var(\hat{b}_j) = \sigma^2 / SST_j(1-R^2_j)$

for $j = 1, 2, \ldots, k$, where $SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$

is the total sample variation in *xj*, and R_j^2 is the *R*-squared from regressing *xj* on all other independent variables (and including an intercept).

Theorem 3: Unbiased estimation of σ^2

Under the Gauss Markov assumptions MLR1 through MLR 4, $E(\hat{\sigma}^2) = \sigma^2$

Theorem 4: Gauss-Markov Theorem

Under Assumptions MLR.1 through MLR.5, $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$ are the best linear unbiased

Estimators (BLUEs) of $b_0, b_1, \ldots, b_k$, respectively.

It is because of this theorem that Assumptions MLR.1 through MLR.5 are known as the

Gauss-Markov assumptions (for cross-sectional analysis).

The importance of the Gauss-Markov Theorem is that, when the standard

set of assumptions holds, we need not look for alternative unbiased estimators, none will be better than OLS. Equivalently, if we are presented with an estimator that is both linear and unbiased, then we know that the variance of this estimator is at least as large as the OLS variance; no additional calculation is needed to show this. For our purposes, Theorem 3 justifies the use of OLS to estimate multiple regression models. If any of the Gauss-Markov assumptions fail, then this theorem no longer holds.



CHECK YOUR PROGRESS

Q 1: Define multiple regression analysis

.....

.....

Q 2: State one drawback simple regression analysis for doing empirical work.

.....

.....

Q 3: State one assumption of multiple regression analysis

.....

.....

.....



6.6 LET US SUM UP

- Multiple regression analysis is allows us to explicitly control for many other factors that simultaneously affect the dependent variable
- The advantage of multiple regression analysis is that it can incorporate fairly general functional relationships
- Multiple regression analysis is also useful for generalizing functional relationships between variables.
- But multiple regression analysis is based on certain assumptions



6.7 FURTHER READINGS

1) Gujrati, D.N. (1995): *Basic Econometrics*, Tata McGraw Hill, New Delhi.

2) Koutsoyiannis, A. (1977): *Theory of Econometrics*, the MacMillan Press Ltd. , London.

3) Wooldridge, J. F. (2008): *Introductory Econometrics*: A Modern Approach, Third edition, South- Western Cengage Learning India.

4) Johnston, J. (1991): *Econometric Methods*, MacGraw Hill Book Co. , London.



6.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Multiple regression analysis predicts the value a dependent variable based on the values of two or more independent variables.

Ans to Q No 2: The one drawback in using simple regression analysis for empirical work is that it is very difficult to analyse how *x* affects *y*, assuming ceteris paribus conclusions, or other things remaining same.

Ans to Q No 3: One assumption of the multiple regression analysis is that variance of the residuals is constant.



6.9 MODEL QUESTIONS

Q 1: Briefly explain the significance of the stochastic disturbance term.

Q 2: Explain the difference between population regression function and sample regression function

Q 3: Explain the difference between Simple and Multiple Linear Regression Model

Q 4: Explain the properties of least square estimators.

Q 5: Calculate the regression lines and the coefficient of correlation from the following data.

X	40	38	35	42	30
Y	30	35	40	36	29

Q 6: Find the regression lines Y on X and X on Y. Predict the average value of Y When X = 90.

X	30	60	50	40	40	60	70	50
Y	30	20	30	50	30	60	60	40

*** ***** ***

UNIT 7 : COEFFICIENT OF DETERMINATION AND HYPOTHESIS TESTING

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Coefficient of Determination as a Measurement of Goodness of Fit

7.4 Adjusted R2

7.5 Hypothesis Testing

7.6 Let Us Sum Up

7.7 Further Reading

7.8 Answers to Check Your Progress

7.9 Model Questions

7.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- explain coefficient of determination as a measurement of goodness of fit
- define adjusted R^2
- explain hypothesis testing

7.2 INTRODUCTION

Let us assume that we can write a model with two independent variables as

$$y = b_0 + b_1 x_1 + b_2 x_2 + u, \text{———————————} (7.1)$$

where

b_0 is the intercept.

b_1 measures the change in y with respect to x_1, holding other factors fixed.

b_2 measures the change in y with respect to x_2, holding other factors fixed.

Estimation of regression equation (7.1) through OLS will yield the following equation

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 \quad \text{——————————(7.2)}$$

Property 1: The sum, and therefore the sample average of the OLS residuals, is zero.

$$\sum_{i=1}^{n} \hat{u}i = 0$$

Property 2: The sample covariance between the regressors and OLS residuals is Zero.

$$\sum_{i=1}^{n} xi\hat{u}i = 0.$$

Writing each y_i as its fitted value $\hat{y}_i$ (which is estimated through OLS) plus its residual provides another way to interpret OLS regression

$$y_i = \hat{y}_i + \hat{u}_i \quad \text{——————————} (7.3)$$
$$= (\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2) + \hat{u}_i$$

Thus, we can view OLS as decomposing into two parts, a fitted value and a residual.

Define the total sum of squares (SST), the explained sum of squares (SSE), and the residual sum of squares (SSR) (also known as the sum of squared residuals), as follows:

$$\textbf{Total Sum of Squares (SST)} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\textbf{Explained Sum of Squares (SSE)} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$\textbf{Residual Sum of Squares (SSR)} = \sum_{i=1}^{n} \hat{u}_i^2$$

SST=SSE +SSR ——————————————— (7.4)

7.3 COEFFICIENT OF DETERMINATION AS A MEASURE OF GOODNESS-OF-FIT

So far, we have no way of measuring how well the explanatory or independent variable, *x*, explains the dependent variable, *y*. It is often useful to compute a number that summarizes how well the OLS regression line fits the data. In the following discussion, be sure to remember that we assume that an intercept is estimated along with the slope.

Assuming that the total sum of squares, SST, is not equal to zero—which is true except in the very unlikely event that all the *yi* equal the same value—we can divide (7.4) by SST to get

1= SSE/SST + SSR/SST. The ***R*-squared** of the regression, sometimes, called the **coefficient of determination**, is defined as R^2 ;

SSE/SST = 1 - SSR/SST**————————————————————(7.5)**

R^2 is the ratio of the explained variation compared to the total variation; thus, it is Interpreted as the *fraction of the sample variation in y that is explained by x*. The second equality in (7.5) provides another way for computing R^2.

From (7.4), the value of R^2 is always between zero and one, because SSE can be no greater than SST. When interpreting R^2, we usually multiply it by 100 to change it into a percent: 100 R^2 is the *percentage of the sample variation in y that is explained by x*. If the data points all lie on the same line, OLS provides a perfect fit to the data. In this case, $R^2 = 1$. A value of R^2 that is nearly equal to zero indicates a poor fit of the OLS line: very little of the variation in the *yi* is captured by the variation in the $\hat{y}_i$ (which all lie on the OLS regression line). In fact, it can be shown that R^2 is equal to the *square* of the sample correlation coefficient between *yi* and $\hat{y}_i$. This is where the term "*R*-squared" came from. (The letter *R* was traditionally used to denote an estimate of a population correlation coefficient, and its usage has survived in regression analysis.)

In the CEO salary regression, we obtain the following:

Salary= 963.191 + 18.501 *roe* ————————————————(7.6)

n =209, R^2 = 0.0132.

We have reproduced the OLS regression line and the number of observations for clarity. Using the *R*-squared (rounded to four decimal places) reported for this equation, we can see how much of the variation in salary is actually explained by the return on equity. The answer is: not much. The firm's return on equity explains only about 1.3% of the variation in salaries for this sample of 209 CEOs. That means that 98.7% of the salary variations for these CEOs is left unexplained! This lack of explanatory power may not be too surprising because many other characteristics of both the firm and the individual CEO should influence salary; these factors are necessarily included in the errors in a simple regression analysis.

In the social sciences, low *R*-squareds in regression equations are not uncommon, especially for cross-sectional analysis. but it is worth emphasizing now that a seemingly low *R*-squared does not necessarily mean that an OLS regression equation is useless. It is still possible that (7.6) is a good estimate of the ceteris paribus relationship between *salary* and *roe*; whether or not this is true does *not* depend directly on the size of *R* squared. Students who are first learning econometrics tend to put too much weight on the size of the *R*-squared in evaluating regression equations. For now, be aware that using *R*-squared as the main gauge of success for an econometric analysis can lead to trouble. Sometimes, the explanatory variable explains a substantial part of the sample variation in the dependent variable



CHECK YOUR PROGRESS

Q 1: State true or false

(a) The sample average of the OLS residuals is zero.

.....

.....

(b) The sample covariance between the regressors and OLS residuals is zero.

.....

.....

7.4 ADJUSTED R^2

To see how the usual *R*-squared might be adjusted, it is usefully written as

$R^2 = 1 - (SSR/n)/(SST/n)$, ——————————————(7.7)

where SSR is the sum of squared residuals and SST is the total sum of squares; compared with equation (7.5), all we have done is divide both SSR and SST by *n*. This expression reveals what R^2 is actually estimating. Define σ^2_y as the population variance of *y* and let σ^2_u denote the population variance of the error term, *u*.

The **population *R* squared** is defined as $\sigma^2 = 1 - \sigma^2_{u/}\sigma^2_y$ ——- (7.8) this is the proportion of the variation in *y* in the population explained by the independent variables. This is what R^2 is supposed to be estimating.

*R*2 estimates σ^2_u by SSR/*n*, which we know to be biased. So why not replace SSR/*n* with SSR/(*n* -*k* -1)? Also, we can use SST/(*n*-1) in place of SST/*n*, as the former is the unbiased estimator of σ^2_y

Using these estimators, we arrive at the adjusted *R*squared:

$$R^2 = 1 - [SSR/(n - k - 1)]/[SST/(n - 1)]$$
$$= 1 - \sigma\hat{}^2/[SST/(n - 1)], \text{ ----------(7.9)}$$

because $\sigma\hat{}^2$ = SSR/(*n* - *k* - 1). Because of the notation used to denote the adjusted *R* squared, it is sometimes called *R-bar squared*.

The adjusted *R*-squared is sometimes called the *corrected R-squared*, but this is not a good name because it implies that adj -R^2 is somehow better than R^2 as an estimator of the population *R*-squared. The primary attractiveness of R^2 is that it imposes a penalty for adding additional Independent variables to a model. We know that R^2 can never fall when a new independent variable is added to a regression equation: this is because SSR never goes up (and usually falls) as more independent variables are added. But the formula for R^2 shows that it depends explicitly on *k*, the number of independent variables. If an independent variable is added to a regression, SSR falls, but so does the *df* in the regression, *n* - *k* - 1. SSR/(*n* - *k* 2-1) can go up or down when a new independent variable is added to a regression.

7.5 HYPOTHESIS TESTING

Up to this point, we have formed a set of assumptions under which OLS is unbiased; we have also derived and discussed the bias caused by omitted variables. In the previous unit, we obtained the variances of the OLS estimators under the Gauss-Markov assumptions. We also know that this variance is smallest among linear unbiased estimators. Knowing the expected value and variance of the OLS estimators is useful for describing the precision of the OLS estimators. However, in order to perform statistical inference, we need to know more than just the first two moments of $b\hat{}_j$; we need to know the full sampling distribution of the $b\hat{}_j$. Even under the Gauss-Markov assumptions, the distribution of $b\hat{}_j$ can have virtually any shape.

When we condition on the values of the independent variables in our sample, it is clear that the sampling distributions of the OLS estimators depend on the underlying distribution of the errors. To make the sampling distributions of the $b\hat{}_j$ tractable, we now assume that the unobserved error is *normally distributed* in the population. We call this the **normality assumption**.

Assumption MLR.6 Normality : The population error u is *independent* of the explanatory variables $x_1, x_2, \ldots, x_k$ and is normally distributed with zero mean and variance o^2: $u \sim$ Normal $(0, o^2)$.

Assumption MLR.6 is much stronger than any of our previous assumptions. In fact, since u is independent of the xj under MLR.6, E(u|$x1$, …, xk) 5= E(u) = 0 and Var(u|x_1, …, x_k) = Var(u) = o^2. Thus, if we make Assumption MLR.6, then we are necessarily assuming MLR.4 and MLR.5. To emphasize that we are assuming more than before, we will refer to the full set of Assumptions MLR.1 through MLR.6.

For cross-sectional regression applications, Assumptions MLR.1 through MLR.6 are called the **classical linear model** (**CLM**) **assumptions**. Thus, we will refer to the model under these six assumptions as the **classical linear model**. It is best to think of the CLM assumptions as containing all of

the Gauss-Markov assumptions *plus* the assumption of a normally distributed error term. Under the CLM assumptions, the OLS estimators $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$ have a stronger efficiency property than they would under the Gauss-Markov assumptions. It can be shown that the OLS estimators are the **minimum variance unbiased estimators**, which means that OLS has the smallest variance among unbiased estimators;

A succinct way to summarize the population assumptions of the CLM is

$y|\mathbf{x} \sim \text{Normal}(b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k, o^2)$, ——————————(7.10)

where **x** is again shorthand for $(x_1, \ldots, x_k)$. Thus, conditional on **x**, *y* has a normal distribution with mean linear in $x_1, \ldots, x_k$ and a constant variance.

Theorem 7.1 Normal Sampling Distributions : Under the CLM assumptions MLR.1 through MLR.6, conditional on the sample values of the independent variables,

$\hat{b}j \sim \text{Normal}[bj, \text{Var}(\hat{b}j)]$,

Testing Hypotheses about a Single Population

Parameter: The *t* Test

This section covers very important topic of testing hypotheses about any single parameter in the population regression function. The population model can be written as

$y = b_0 + b_1x_1 + \ldots + b_kx_k + u$, ——————————————————(7.11)

and we assume that it satisfies the CLM assumptions. We know that OLS produces unbiased estimators of the *bj* . In this section, we study how to test hypotheses about a particular *bj.* For a full understanding of hypothesis testing, one must remember that the *bj* are unknown features of the population, and we will never know them with certainty. Nevertheless, we can *hypothesize* about the value of *bj* and then use statistical inference to test our hypothesis. In order to construct hypotheses tests, we need the following result:

Theorem 7.2 *t* Distribution for the Standardized Estimators : Under the CLM assumptions MLR.1 through MLR.6,

$(\hat{b}_j - b_j)/\text{se}(\hat{b}_j) \sim t_{n-k-1} = t_{df}$, ——————————————————-(7.12)

where $k + 1$ is the number of unknown parameters in the population model $y = b_0 + b_1x_1 + \ldots + b_kx_k + u$ (*k* slope parameters and the intercept b_0) and *n* -

k- 1 is the degrees of freedom (*df*).

Theorem 7.2 is important in that it allows us to test hypotheses involving the *bj*. In most applications, our primary interest lies in testing the **null hypothesis**

H_0: $b_j = 0$, ——————————————————————— (7.13)

where *j* corresponds to any of the *k* independent variables. It is important to understand what (7.13) means and to be able to describe this hypothesis in simple language for a particular application. Since *bj* measures the partial effect of x_j on (the expected value of) *y*, after controlling for all other independent variables, (7.13) means that, once x_1, x_2, …, x_{j-1}, x_{j+1}, …, x_k have been accounted for, x_j has *no effect* on the expected value of *y*. We cannot state the null hypothesis as "x_j does have a partial effect on *y*" because this is true for any value of *bj* other than zero. Classical testing is suited for testing *simple hypotheses* like (7.13).

As an example, consider the wage equation

$\log(wage) = b_0 + b_1 educ + b_2 exper + b_3 tenure + u.$ ——————————(7.14)

The null hypothesis H_0: $b2 = 0$ means that, once education and tenure have been accounted for, the number of years in the workforce (*exper*) has no effect on hourly wage. This is an economically interesting hypothesis. If it is true, it implies that a person's work history prior to the current employment does not affect wage. If $b_2 = 0$, then prior work experience contributes to productivity, and hence to wage.

The mechanics of testing (7.13) in the multiple regression context are very similar. The hard part is obtaining the coefficient estimates, the standard errors, and the critical values, but most of this work is done automatically by econometrics software. Our job is to learn how regression output can be used to test hypotheses of interest.

The statistic we use to test (7.13) (against any alternative) is called "the" ***t* statistic** or "the" ***t* ratio** of $\hat{b}_j$ and is defined as

$t_{\hat{b}_j} = \hat{b}_j / se(\hat{b}_j)$. ——————————————————————(7.15)

We have put "the" in quotation marks because, as we will see shortly, a more general form of the *t* statistic is needed for testing other hypotheses about *bj*. For now, it is important to know that (7.12) is suitable only for

testing (7.13). For particular applications, it is helpful to index *t* statistics using the name of the independent variable; for example, t_{educ} would be the *t* statistic for $b\hat{}_{educ}$. The *t* statistic for $b\hat{}_j$ is simple to compute given $b\hat{}j$ and its standard error. In fact, most regression packages do the division for you and report the *t* statistic along with each
coefficient and its standard error.

Before discussing how to use (7.15) formally to test H0: *bj* = 0, it is useful to see why $t_{b\hat{}j}$ has features that make it reasonable as a test statistic to detect *bj* = 0. First, since se($b\hat{}_j$) is always positive, $t_{b\hat{}j}$ has the same sign as $b\hat{}_j$: if $b\hat{}j$ is positive, then so is $t_{b\hat{}j}$, and if $b_{\hat{}j}$ is negative,
so is $t_{b\hat{}j}$. Second, for a given value of se($b\hat{}_j$), a larger value of $b\hat{}j$ leads to larger values of $t_{b\hat{}j}$
If $b\hat{}_j$ becomes more negative, so does $t_{b\hat{}j}$.

Since we are testing H_0: *bj* = 0, it is only natural to look at our unbiased estimator of *bj* , $b\hat{}j$, for guidance. In any interesting application, the point estimate $b\hat{}j$ will *never* exactly be zero, whether or not H_0 is true. The question is: How far is $b\hat{}j$ from zero? A sample value of $b\hat{}j$ very far from zero provides evidence against H0: *bj* = 0. However, we must recognize that there is a sampling error in our estimate $b\hat{}j$, so the size of $b\hat{}j$ must be weighed against its sampling error. Since the standard error of $b\hat{}j$ is an estimate of the standard deviation of $b\hat{}j$, $t_{b\hat{}j}$ measures how many estimated standard deviations $b\hat{}j$ is away from zero. This is precisely what we do in testing whether the mean of a population is zero, using the standard *t* statistic from introductory statistics. Values of $t_{b\hat{}j}$ sufficiently far from zero will result in a rejection of H_0. The precise rejection rule depends on the alternative hypothesis and the chosen significance level of the test.

Determining a rule for rejecting (7.13) at a given significance level—that is, the Probability of rejecting H_0 when it is true—requires knowing the sampling distribution of $t_{b\hat{}j}$ when H_0 is true. From Theorem 7.2, we know this to be t_{n-k-1}. This is the key theoretical result needed for testing (7.13).

Before proceeding, it is important to remember that we are testing hypotheses about the *population* parameters. We are *not* testing hypotheses about the estimates from a particular sample. Thus, it never makes sense

to state a null hypothesis as "H0: $b\hat{}_1 = 0$" or, even worse, as "H0: .237 = 0" when the estimate of a parameter is .237 in the sample. We are testing whether the unknown population value, b_1, is zero.

Some treatments of regression analysis define the *t* statistic as the *absolute value* of (4.5), so that the *t* statistic is always positive. This practice has the drawback of making testing against one-sided alternatives clumsy. Throughout this text, the *t* statistic always has the same sign as the corresponding OLS coefficient estimate.

Testing against One-Sided Alternatives : To determine a rule for rejecting H_0, we need to decide on the relevant **alternative hypothesis**.

First, consider a **one-sided alternative** of the form

H_1: $bj = 0$. ——————————————————————————(7.16)

When we state the alternative as in equation (7.16), we are really saying that the null hypothesis is H_0: $bj = 0$. For example, if b_j is the coefficient on education in a wage regression, we only care about detecting that b_j is different from zero when b_j is actually positive. You may remember from introductory statistics that the null value that is hardest to reject in favor of (7.16) is $b_j = 0$. In other words, if we reject the null hypothesis $b_j = 0$ then we automatically reject $bj = 0$. Therefore, it suffices to act as if we are testing H_0: $bj = 0$ against H_1: $bj > 0$, effectively ignoring $bj < 0$,

How should we choose a rejection rule? We must first decide on a **significance level** ("level" for short) or the probability of rejecting H_0 when it is in fact true. For concreteness, suppose we have decided on a 5% significance level, as this is the most popular choice. Thus, we are willing to mistakenly reject H_0 when it is true 5% of the time. Now, while $t_{b\hat{}j}$ has a *t* distribution under H_0—so that it has zero mean—under the alternative $bj > 0$, the expected value of $t_{b\hat{}j}$ is positive. Thus, we are looking for a "sufficiently large" positive value of $t_{b\hat{}j}$ in order to reject H_0: $bj = 0$ in favor of H_1: $bj > 0$. Negative values of $t_{b\hat{}j}$ provide no evidence in favor of H_1.

The definition of "sufficiently large," with a 5% significance level, is the 95th percentile in a *t* distribution with $n - k - 1$ degrees of freedom; denote this by *c*. In other words, the **rejection rule** is that H_0 is rejected in favor of H_1 at the 5% significance level if $t_{b\hat{}j} > c$. ——————————(7.17)

By our choice of the **critical value** *c*, rejection of H_0 will occur for 5% of all random samples when H_0 is true.

The rejection rule in (7.17) is an example of a **one-tailed test**. To obtain *c*, we only need the significance level and the degrees of freedom. For example, for a 5% level test and with *n* - *k* - 1 = 28 degrees of freedom, the critical value is *c* = 1.701. If $t_{\hat{b}_j}$ <= 1.701, then we fail to reject H_0 in favor of (7.17) at the 5% level. Note that a negative value for $t_{\hat{b}_j}$, no matter how large in absolute value, leads to a failure in rejecting H_0 in favor of (7.17).

The same procedure can be used with other significance levels. For a 10% level test and if *df* = 21, the critical value is *c* = 1.323. For a 1% significance level and if *df* = 21, *c* = 2.518. All of these critical values are obtained directly from t distribution Table. You should note a pattern in the critical values: As the significance level falls, the critical value increases, so that we require a larger and larger value of $t_{\hat{b}_j}$ in order to reject H_0. Thus, if H_0 is rejected at, say, the 5% level, then it is automatically rejected at the 10% level as well. It makes no sense to reject the null hypothesis at, say, the 5% level and then to redo the test to determine the outcome at the 10% level.

As the degrees of freedom in the *t* distribution get large, the *t* distribution approaches the standard normal distribution. For example, when *n* -*k* - 1 = 120, the 5% critical value for the one-sided alternative (7.17) is 1.658, compared with the standard normal value of 1.645. These are close enough for practical purposes; for degrees of freedom greater than 120, one can use the standard normal critical values.

Two-Sided Alternatives : In applications, it is common to test the null hypothesis H_0: *bj* = 0 against a **two-sided alternative**; that is,

H_1: b_j ~= 0. ——————————————— (7.18)

Under this alternative, *xj* has a ceteris paribus effect on *y* without specifying whether the effect is positive or negative. This is the relevant alternative when the sign of b_j is not well determined by theory (or common sense). Even when we know whether b_j is positive or negative under the

alternative, a two-sided test is often prudent. At a minimum, using a two-sided alternative prevents us from looking at the estimated equation and then basing the alternative on whether $\hat{b}_j$ is positive or negative. Using the regression estimates to help us formulate the null or alternative hypotheses is not allowed because classical statistical inference presumes that we state the null and alternative about the population before looking at the data. For example, we should not first estimate the equation relating math performance to enrollment, note that the estimated effect is negative, and then decide the relevant alternative is H_1: *b enroll* = 0.

When the alternative is two-sided, we are interested in the *absolute value* of the *t* statistic.

The rejection rule for H0: *bj* = 0 against 7.18) is

$$|t_{\hat{b}_j}| > c, \qquad \text{(7.19)}$$

where | . | denotes absolute value and *c* is an appropriately chosen critical value. To find *c*, we again specify a significance level, say 5%. For a **two-tailed test**, *c* is chosen to make the area in each tail of the *t* distribution equal 2.5%. In other words, *c* is the 97.5th percentile in the *t* distribution with *n* - *k* - 1 degrees of freedom. When *n* - *k* - 1 = 25, the 5% critical value for a two-sided test is *c* = 2.060.

When a specific alternative is not stated, it is usually considered to be two-sided. In the remainder of this text, the default will be a two-sided alternative, and 5% will be the default significance level. When carrying out empirical econometric analysis, it is always a good idea to be explicit about the alternative and the significance level. If H_0 is rejected in favor of (7.19) at the 5% level, we usually say that "x_j is **statistically significant**, or statistically different from zero, at the 5% level." If H_0 is not rejected, we say that "x_j is **statistically insignificant** at the 5% level."

Testing Other Hypotheses about *bj*

Although H_0: $b_j = 0$ is the most common hypothesis, we sometimes want to test whether b_j is equal to some other given constant. Two common examples are $b_j = 1$ and *bj* = 21. Generally, if the null is stated as H_0: $b_j = a_j$, — (7.20)

where a_j is our hypothesized value of b_j, then the appropriate *t* statistic is $t = (\hat{b}_j - a_j)/\text{se}(\hat{b}_j)$.

As before, *t* measures how many estimated standard deviations $\hat{b}_j$ is away from the Hypothesized value of b_j. The general *t* statistic is usefully written as

t = (*estimate - hypothesized value*)/ *standard error* ——————— (7.21)

Under (7.20), this *t* statistic is distributed as t_{n-k-1} from Theorem 7.2. The usual *t* statistic is obtained when $a_j = 0$. We can use the general *t* statistic to test against one-sided or two-sided alternatives. For example, if the null and alternative hypotheses are H_0: $b_j = 1$ and H_1: $bj > 1$, then we find the critical value for a one-sided alternative *exactly* as before: the difference is in how we compute the *t* statistic, not in how we obtain the appropriate *c*. We reject H_0 in favor of H_1 if $t > c$. In this case, we would say that "$\hat{b}_j$ is statistically greater than one" at the appropriate significance level.

Testing of multiple linear restrictions : F test

The t statistic associated with any OLS coefficient can be used to test whether the corresponding unknown parameter in the population is equal to any given constant (which is usually, but not always, zero). We have just shown how to test hypotheses about a single linear combination of the b_j by rearranging the equation and running a regression using transformed variables. But so far, we have only covered hypotheses involving a single restriction. Frequently, we wish to test multiple hypotheses about the underlying parameters $b_0, b_1, \ldots, b_k$. We begin with the leading case of testing whether a set of independent variables has no partial effect on a dependent variable.

Testing Exclusion Restrictions: We already know how to test whether a particular variable has no partial effect on the dependent variable: use the t statistic. Now we want to test whether a group of variables has no effect on the dependent variable. More precisely, the null hypothesis is that a set of variables has no effect on y, once another set of variables has been controlled.

As an illustration of why testing significance of a group of variables is useful, we consider the following model that explains major IPL salaries:

log (salary) = $b_0 + b_1$ years + $b_2$gamesyr + $b_3$bavg + $b_4$hrunsyr + $b_5$rbisyr + u, ——————— (7.22)

where salary is the 2018 total salary, years is years in the IPL league, gamesyr is average games played per year, bavg is career batting average (for example, bavg = 30), hrunsyr is runs made at home ground per year, and rbisyr is runs scored per year. Suppose we want to test the null hypothesis that, once years in the league and games per year have been controlled for, the statistics measuring performance—bavg, hrunsyr, and rbisyr—have no effect on salary. Essentially, the null hypothesis states that productivity as measured by cricket statistics has no effect on salary

In terms of the parameters of the model, the null hypothesis is stated as

$H_0: b_3 = 0, b_4 = 0, b_5 = 0$ ---------- (7.23)

The null (7.23) constitutes three exclusion restrictions: if (7.23) is true, then bavg, hrunsyr, and rbisyr have no effect on log(salary) after years and gamesyr have been controlled for and therefore should be excluded from the model. This is an example of a set of multiple restrictions because we are putting more than one restriction on the parameters in (7.23); we will see more general examples of multiple restrictions later. A test of multiple restrictions is called a multiple hypotheses test or a joint hypotheses test. What should be the alternative to (7.23)? If what we have in mind is that "performance statistics matter, even after controlling for years in the league and games per year," then the appropriate alternative is simply

H1: H_0 is not true ——————————————(7.24)

The alternative (7.23) holds if at least one of b_3, b_4, or b_5 is different from zero. (Any or all could be different from zero.) The test we study here is constructed to detect any violation of H_0. It is also valid when the alternative is something like H1: $b_3 > 0$, $b_4 > 0$, or $b_5 > 0$ but it will not be the best possible test under such alternatives.

How should we proceed in testing (7.23) against (7.24)? It is tempting to test (7.23) by using the t statistics on the variables bavg, hrunsyr, and rbisyr to determine whether each variable is individually significant. This option is not appropriate. A particular t statistic tests a hypothesis that puts no restrictions on the other parameters. Besides, we would have three outcomes to contend with—one for each t statistic. What would constitute rejection of (7.23) at, say, the 5% level? Should all three or only one of the

three t statistics be required to be significant at the 5% level? These are hard questions, and fortunately we do not have to answer them. Furthermore, using separate t statistics to test a multiple hypothesis like (7.23) can be very misleading. We need a way to test the exclusion restrictions jointly. To illustrate these issues, we estimate equation (7.22) using a hypothetical database consisting of IPL leagues over the past decade.

$$\log(\text{salary}) = 11.10 + .0689\ \text{years} + .0126\ \text{gamesyr} + .00098\ \text{bavg} + .0144\ \text{hrunsyr} + .0108\text{rbisyr}$$
$$(0.29)\quad (0.0121)\quad (0.0026)\quad (0.00110)\quad (0.0161)\quad (0.0072)$$
$$N = 353,\ SSR = 183.186,\ R^2 = .6278 \qquad \text{----------(7.25)}$$

where SSR is the sum of squared residuals. We have left several terms after the decimal in SSR and R-squared to facilitate future comparisons. Equation (7.25) reveals that, while years and gamesyr are statistically significant, none of the variables bavg, hrunsyr, and rbisyr has a statistically significant t statistic against a two-sided alternative, at the 5% significance level. (The t statistic on rbisyr is the closest to being significant; its two-sided p-value is 0.134.) Thus, based on the three t statistics, it appears that we cannot reject H_0.

This conclusion turns out to be wrong. In order to see this, we must derive a test of multiple restrictions whose distribution is known and tabulated. The sum of squared residuals now turns out to provide a very convenient basis for testing multiple hypotheses. We will also show how the R-squared can be used in the special case of testing for exclusion restrictions. Knowing the sum of squared residuals in (7.25) tells us nothing about the truth of the hypothesis in (7.23). However, the factor that will tell us something is how much the SSR increases when we drop the variables bavg, hrunsyr, and rbisyr from the model. Remember that, because the OLS estimates are chosen to minimize the sum of squared residuals, the SSR always increases when variables are dropped from the model; this is an algebraic fact. The question is whether this increase is large enough, relative to the SSR in the model with all of the variables, to warrant rejecting the null hypothesis.

The model without the three variables in question is simply

$$\log(\text{salary}) = b_0 + b_1\,\text{years} + b_2\,\text{gamesyr} + u. \qquad \text{————————(7.26)}$$

In the context of hypothesis testing, equation (7.26) is the restricted model

for testing (7.23); model (7.22) is called the unrestricted model. The restricted model always has fewer parameters than the unrestricted model When we estimate the restricted model using the database of IPL leagues , we obtain

Log(saˆlary) = 11.22 + .0713 years + .0202 gamesyr ——————(7.27)

(0.11) (.0125) (.0013)

n = 353, SSR =198.311, R^2 = .5971.

As we surmised, the SSR from (7.27) is greater than the SSR from (7.23), and the R squared from the restricted model is less than the R-squared from the unrestricted model. What we need to decide is whether the increase in the SSR in going from the unrestricted model to the restricted model (183.186 to 198.311) is large enough to warrant rejection of (7.23). As with all testing, the answer depends on the significance level of the test. But we cannot carry out the test at a chosen significance level until we have a statistic whose distribution is known, and can be tabulated, under H_0. Thus, we need a way to combine the information in the two SSRs to obtain a test statistic with a known distribution under H_0.

Since it is no more difficult, we might as well derive the test for the general case. Write the unrestricted model with k independent variables as

$$y = b_0 + b_1x_1 + \ldots.. + b_k x_k + u; \text{ ——————(7.28)}$$

the number of parameters in the unrestricted model is k + 1. (Remember to add one for the intercept.) Suppose that we have q exclusion restrictions to test: that is, the null hypothesis states that q of the variables in (7.28) have zero coefficients. For notational simplicity, assume that it is the last q variables in the list of independent variables: x_{k-q+1}, …, x_k. (The order of the variables, of course, is arbitrary and unimportant.) The null hypothesis is stated as

$$H_0 = b_{k-q+1} = 0, \ldots\ldots\ldots\ldots\ldots b_k = 0 \text{ ----------(7.29)}$$

which puts q exclusion restrictions on the model (7.28). The alternative to (7.29) is simply that it is false; this means that at least one of the parameters listed in (7.29) is different from zero. When we impose the restrictions under H_0, we are left with the restrictions model

$$y = b_0 + b_1x_1 + \ldots.. + b_{k-q} x_{k-q} + u \text{ ----------(7.30)}$$

In this subsection, we assume that both the unrestricted and restricted models contain an intercept, since that is the case most widely encountered in practice. Now for the test statistic itself. Earlier, we suggested that looking at the relative increase in the SSR when moving from the unrestricted to the restricted model should be informative for testing the hypothesis (7.29). The F statistic (or F ratio) is defined by

$F= (SSR_r - SSR_{ur})/ q/ SSR_{ur}/(n -k -1)$ ----------(7.31)

where SSR_r is the sum of squared residuals from the restricted model and SSR_{ur} is the sum of squared residuals from the unrestricted model. You should immediately notice that, since SSR_r can be no smaller than SSR_{ur}, the F statistic is always nonnegative (and almost always strictly positive). Thus, if you compute a negative F statistic, then something is wrong; the order of the SSR_s in the numerator of F has usually been reversed. Also, the SSR in the denominator of F is the SSR from the unrestricted model. The easiest way to remember where the SSR_s appear is to think of F as measuring the relative increase in SSR when moving from the unrestricted to the restricted model.

The difference in SSR_s in the numerator of F is divided by q, which is the number of restrictions imposed in moving from the unrestricted to the restricted model (q independent variables are dropped). Therefore, we can write

q = numerator degrees of freedom = $df_r - df_{ur}$, ----------(7.32)

which also shows that q is the difference in degrees of freedom between the restricted and unrestricted models. (Recall that df = number of observations - number of estimated parameters.) Since the restricted model has fewer parameters—and each model is estimated using the same n observations—df_r is always greater than df_{ur}. The SSR in the denominator of F is divided by the degrees of freedom in the unrestricted model

$n-k-1$ = denominator degrees of freedom = df_{ur} ——————(7.33)

In fact, the denominator of F is just the unbiased estimator of σ^2 = Var(u) in the unrestricted model. In particular application, computing the F statistic is easier than wading through the somewhat cumbersome notation used to describe the general case. We first obtain the degrees of freedom

in the unrestricted model, df_{ur}. Then, we count how many variables are excluded in the restricted model; this is q. The SSRs are reported with every OLS regression, and so forming the F statistic is simple.

In the major league baseball salary regression, n = 353, and the full model (7.21) contains six parameters. Thus, n -k -1 = df_{ur} = 353 - 6 = 347. The restricted model (7.26) contains three fewer independent variables than (7.21), and so q = 3. Thus, we have all of the ingredients to compute the F statistic; we hold off doing so until we know what to do with it.

In order to use the F statistic, we must know its sampling distribution under the null in order to choose critical values and rejection rules. It can be shown that, under H0 (and assuming the CLM assumptions hold), F is distributed as an F random variable with (q,n-k -1) degrees of freedom. We write this as

$$F \sim F_{q,n-k-1}.$$

The distribution of Fq,n-k-1 is readily tabulated and available in statistical tables and, even more importantly, in statistical software. We will not derive the F distribution because the mathematics is very involved. Basically, it can be shown that equation (7.31) is actually the ratio of two independent chi-square random variables, divided by their respective degrees of freedom. The numerator chi-square random variable has q degrees of freedom, and the chi-square in the denominator has n - k -1 degrees of freedom. This is the definition of an F distributed random variable. It is pretty clear from the definition of F that we will reject H_0 in favor of H_1 when F is sufficiently “large.” How large depends on our chosen significance level. Suppose that we have decided on a 5% level test. Let c be the 95th percentile in the Fq,n-k-1 distribution. This critical value depends on q (the numerator df) and n -k -1 (the denominator df). It is important to keep the numerator and denominator degrees of freedom straight.

The 10%, 5%, and 1% critical values for the F distribution. The rejection rule is simple. Once c has been obtained, we reject H_0 in favor of H_1 at the chosen significance level if

$$F > c. \quad (4.40)$$

With a 5% significance level, q = 3, n - k -1 = 60, and the critical value is c=

2.76. We would reject H0 at the 5% level if the computed value of the F statistic exceeds 2.76. For the same degrees of freedom, the 1% critical value is 4.13. In most applications, the numerator degrees of freedom (q) will be notably smaller than the denominator degrees of freedom (n - k - 1). Applications where n - k - 1 is small are unlikely to be successful because the parameters in the null model will probably not be precisely estimated. When the denominator df reaches about 120, the F distribution is no longer sensitive to it. (This is entirely analogous to the t distribution being well-approximated by the standard normal distribution as the df gets large.) Thus, there is an entry in the table for the denominator df, and this is what we use with large samples (since n - k - 1 is then large). A similar statement holds for a very large numerator df, but this rarely occurs in applications.

If H_0 is rejected, then we say that x_{k-q+1}, …, x_k are jointly statistically significant (or just jointly significant) at the appropriate significance level. We are now in a position to test the hypothesis that we began this section with: after controlling for years and gamesyr, the variables bavg, hrunsyr, and rbisyr have no effect on players' salaries. In practice, it is easiest to first compute (SSRr SSRur)/SSRur and to multiply the result by (n- k - 1)/q; the reason the formula is stated as in (7.31) is that it makes it easier to keep the numerator and denominator degrees of freedom straight. Using the SSRs in (7.25) and (7.27), we have

F= (198.11-183.186)/183.16 *347/3=9.55

This number is well above the 1% critical value in the F distribution with 3 and 347 degrees of freedom, and so we soundly reject the hypothesis that bavg, hrunsyr, and rbisyr have no effect on salary.

The outcome of the joint test may seem surprising in light of the insignificant t statistics for the three variables. What is happening is that the two variables hrunsyr and rbisyr are highly correlated, and this multicollinearity makes it difficult to uncover the partial effect of each variable; this is reflected in the individual t statistics. The F statistic tests whether these variables (including bavg) are jointly significant, and multicollinearity between hrunsyr and rbisyr is much less relevant for testing this hypothesis. The F statistic is often useful for testing exclusion of a group of variables

when the variables in the group are highly correlated. For example, suppose we want to test whether firm performance affects the salaries of chief executive officers. There are many ways to measure firm performance, and it probably would not be clear ahead of time which measures would be most important. Since measures of firm performance are likely to be highly correlated, hoping to find individually significant measures might be asking too much due to multicollinearity. But an F test can be used to determine whether, as a group, the firm performance variables affect salary.



CHECK YOUR PROGRESS

Q 2: Define population R square

.....

.....



7.6 LET US SUM UP

- The coefficient of determination is a measurement of explaining how much variability of one factor can be caused by its relationship to another related factor.
- The modified version of R-squared is known as adjusted R squared. Adjusted R-squared has been adjusted for the number of predictors in the model.
- The process of assessing the plausibility of a hypothesis by using sample data is known as hypothesis testing



7.7 FURTHER READINGS:

1) Dougherty, C. (2007), *Introduction to Econometrics*, 3rd Edition., Oxford University Press, New Delhi.

2) Greene, William H. 2005. Econometric Analysis (5th ed.). Upper Saddle River, NJ: Prentice Hall.

3) Gujarati, Damodar. 2003. Basic Econometrics (4th ed.). New York: McGraw-Hill.

4) Johnston, J. 1984. Econometric Methods (3rd ed.). New York: McGraw-Hill. (Note: pages for Johnston, J. and J. DiNardo's Econometric Methods 4th ed. (in parentheses)).

5) Kmenta, Jan. 1986. *Elements of Econometrics* (2nd ed.). New York: Macmillan

6) Maddala, G.S. (Ed.) (1993), *Econometric Methods and Applications* (2 Vols.) Aldershot U.K.

7) Woolwridge Jefferry (2009): *Introductory Econometrics: A Modern Approach*, 5th Edition, South-Western Cengage Learnin



7.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (a) True

(c) True

Ans to Q No 2: Population R squared is the proportion of the variation in *y* in the population explained by the independent variables.



7.9 MODEL QUESTIONS

Q 1: Define the following terms :

a) Hypothesis Testing

b) Level of Significance

c) Null Hypothesis and alternative hypothesis testing

d) Adj R^2

Q 2: The mean lifetime of a sample of 400 fluorescent light produced by a company is found to be 1600 hours with a standard deviation of 150 hours.

a) Test the hypothesis that the mean life time of the bulbds produced in general is different from the mean life of 1570 hours at α=0.01 level of significance (Critical value of Z_α=±2.33).

b) Test the hypothesis that the mean life time of the bulbs produced

in general is higher than the mean life of 1570 hours at α=0.01 level of significance (Critical value of Z_{α}=±2.33).

Q 3: Are rent rates influenced by the student population in a college town? Let *rent* be the average monthly rent paid on rental units in Guwahati. Let *pop* denote the total city population, *avginc* the average city income, and *pctstu* the student population as a percentage of the total population. One model to test for a relationship is

log(*rent*) = *b*0 + *b*1log(*pop*) + *b*2log(*avginc*) + *b*3*pctstu* + *u*.

(i) State the null hypothesis that size of the student body relative to the population has

no ceteris paribus effect on monthly rents. State the alternative that there is an effect.

(ii) What signs do you expect for *b*1 and *b*2?

*** ***** ***