














## Document Information

<b>Analyzed document</b>	2.2 Quantitative Methods Block 2.pdf (D165799781)
<b>Submitted</b>	2023-05-03 19:21:00
<b>Submitted by</b>	Satyaraj
<b>Submitter email</b>	cwplan@icfaiuniversity.in
<b>Similarity</b>	4%
<b>Analysis address</b>	cwplan.ibsh@analysis.urkund.com

## Sources included in the report

<b>W</b>	URL: <a href="https://www.lokniti.org/POST_POLL_ANALYSIS_2021">https://www.lokniti.org/POST_POLL_ANALYSIS_2021</a> Fetched: 2023-05-03 19:21:00		<b>1</b>
<b>W</b>	URL: <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a> Fetched: 2021-10-22 08:53:26		<b>14</b>
<b>W</b>	URL: <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a> Fetched: 2023-01-18 07:05:28		<b>13</b>
<b>W</b>	URL: <a href="https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gede...">https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gede...</a> Fetched: 2021-11-11 19:43:01		<b>4</b>
<b>W</b>	URL: <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a> Fetched: 2022-01-18 21:08:19		<b>6</b>
<b>W</b>	URL: <a href="https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf">https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf</a> Fetched: 2020-12-10 00:20:42		<b>2</b>
<b>W</b>	URL: <a href="https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf">https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf</a> Fetched: 2021-11-17 17:03:33		<b>3</b>
<b>W</b>	URL: <a href="https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf">https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf</a> Fetched: 2020-09-14 09:31:07		<b>3</b>
<b>W</b>	URL: <a href="http://www.math.umd.edu/~wolfe/241Hf11m3.pdf">http://www.math.umd.edu/~wolfe/241Hf11m3.pdf</a> Fetched: 2021-10-07 11:25:47		<b>1</b>
<b>W</b>	URL: <a href="https://www.hksccm.org/images/stories/ActivityReports/Biostatistics%20I.pdf">https://www.hksccm.org/images/stories/ActivityReports/Biostatistics%20I.pdf</a> Fetched: 2022-01-10 01:13:30		<b>2</b>

<b>W</b>	URL: <a href="https://en.wikipedia.org/wiki/Continued_fraction">https://en.wikipedia.org/wiki/Continued_fraction</a> Fetched: 2019-10-29 11:56:34	 <b>1</b>
<b>W</b>	URL: <a href="https://erho.weebly.com/uploads/2/7/8/4/27841631/4.1_-_correlation_and_slrm.pdf">https://erho.weebly.com/uploads/2/7/8/4/27841631/4.1_-_correlation_and_slrm.pdf</a> Fetched: 2022-09-07 06:06:55	 <b>1</b>
<b>W</b>	URL: <a href="http://fin.bus.ku.ac.th/135512%Economic%20Environment%20for%20Finance/Lecture%20Slides/Linea...">http://fin.bus.ku.ac.th/135512%Economic%20Environment%20for%20Finance/Lecture%20Slides/Linea...</a> Fetched: 2021-12-09 11:15:24	 <b>1</b>

## Entire Document

Statistical Relations and Hypothesis Testing Quantitative Methods Quantitative Methods Block-II  
Quantitative Methods Block II STATISTICAL RELATIONS AND HYPOTHESIS TESTING UNIT 5 Statistical Inference and Hypothesis Testing 01-30 UNIT 6 Correlation and Linear Regression 31-54  
ii Editorial Team Prof. R. Prasad IFHE (Deemed-to-be-University), Hyderabad Prof. K. Seethapathi IFHE (Deemed-to-be-University), Hyderabad Dr. Vishal Mishra IFHE (Deemed-to-be-University), Hyderabad Dr. V.S.S.N. Narasimha Murty Kadiyala IFHE (Deemed-to-be-University), Hyderabad Dr. Sanjay.Fuloria IFHE (Deemed-to-be-University), Hyderabad Prof. Muthu Kumar IFHE (Deemed-to-be-University), Hyderabad Content Development Team Dr. Kaustov Chakraborty Prof. Muthu Kumar IFHE (Deemed-to-be-University), Hyderabad IFHE (Deemed-to-be-University), Hyderabad Dr. Sashikala P Dr. Y. V. Subrahmanyam IFHE (Deemed-to-be-University), Hyderabad IFHE (Deemed-to-be-University), Hyderabad Proofreading, Language Editing and Layout Team Ms.Jayashree Murthy Mr. Venkateswarlu IFHE (Deemed-to-be-University), Hyderabad IFHE (Deemed-to-be-University), Hyderabad Mr. Prasad Sistla IFHE (Deemed-to-be-University), Hyderabad © The ICFAI Foundation for Higher Education (IFHE), Hyderabad. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means – electronic, mechanical, photocopying or otherwise – without prior permission in writing from The ICFAI Foundation for Higher Education (IFHE), Hyderabad. Ref. No. QM SLM 102021B2 For any clarification regarding this book, the students may please write to The ICFAI Foundation for Higher Education (IFHE), Hyderabad specifying the unit and page number. While every possible care has been taken in type-setting and printing this book, The ICFAI Foundation for Higher Education (IFHE), Hyderabad welcomes suggestions from students for improvement in future editions. Our E-mail id: cwfeedback@icfaiuniversity.in Center for Distance and Online Education (CDOE) The ICFAI Foundation for Higher Education (Deemed-to-be-University Under Section 3 of UGC Act, 1956) Donthanapally, Shankarapalli Road, Hyderabad-501203.

iii BLOCK II: STATISTICAL RELATIONS AND HYPOTHESIS TESTING Any management or business decision is based on data and its analysis. Any management decision cannot be based on unproven hypothesis. Hence, hypothesis testing is very much required for management decision making and implementation. Statistical techniques are used in hypothesis testing in management or business problems. Usually the hypotheses prove or disprove the existence of relationships, impacts and effects between different variables. To find out the relationship between two different variables correlation is used. To find out the impact of one variable on other, linear regression is used. In this block, statistical inference and hypotheses testing, correlation and linear regression are discussed.

Unit-5 Statistical Inference and Hypothesis Testing discusses about the various types of samples, sampling distribution, central limit theorem, and testing of hypothesis.

Drawing conclusions on population based on study of a sample taken from population is known as statistical inference. A population means a collection of all the data points under study. Different statistical indices can be used to test the hypothesis. Unit-6 Correlation and Linear Regression deals with the study of linear regression and coefficient of correlation. Linear regression constructs linear models while correlation coefficient gives a measure of the relevance of the model. The

correlation analysis determines the degree to which the variables are related, regression analysis develops the relationship between the variables, and coefficient of

correlation indicates the strength of a linear relationship between them. Regression analysis will bring cause and effect relationships between two or more variables.

Unit 5 Statistical Inference and Hypothesis Testing Structure 5.1 Introduction 5.2 Objectives 5.3 Population and Sample 5.4 Types of Sampling 5.5 Statistical Errors 5.6

Sampling Distribution 5.7 Central Limit Theorem 5.8 Estimation 5.9 Testing of Hypothesis 5.10 Large Sample Test for Mean 5.11 Small Sample Test for Mean 5.12 Hypotheses Involving Means of Two Populations (

Large Sample) 5.13 Hypotheses Involving Means of Two Populations (Small Sample) 5.14 Type I and Type II Errors 5.15 Large Sample Test for Proportion 5.16 Hypotheses Involving Proportions of Two Populations 5.17

Summary 5.18 Glossary 5.19 Suggested Readings/Reference Material 5.20 Self-Assessment Questions 5.21 Answers to Check Your Progress Questions 5.1 Introduction In the last unit of previous block, you studies different types of

probability distributions and how to make decisions under certainty as well as under uncertainty. In this unit, you will study, sampling, standard error and test of hypothesis.

Many a time, it is not possible to study each and every component of the entire data points under study due to time and cost constraints. In such cases, a few components are observed and statistical measures are applied. This process of inferring something about a large group called 'population', by studying a small part (sample) of it is called 'sampling'. Population can be studied by enumeration and sampling. Due to constraints such as time, money and energy, most managers prefer the sampling method when the population is too large. This method may lead to errors due to the element of chance. Often, managers believe on some result called 'hypothesis statement'. But sample data sometimes does not support the result obtained. The process that tests the possibility of the difference occurring due to chance elements is known as 'hypothesis testing'.

Block II: Statistical Relations and

Hypothesis Testing 2 5.2

Objectives After going through the unit, you should be able to: ?

List the advantages and disadvantages of sampling; ? Recall various types of sampling; ? Define sampling distribution; ?

Explain central limit theorem; ? Define point estimates and interval estimates; ? Generalize testing of hypothesis; and ?

Explain test of significance for difference between two means. 5.3 Population and Sample The population is the set of data to which findings are to be generalized.

For example, if we want to study about people living below poverty line in India, then the data point which represents such people is population. The basic idea of statistics is to extrapolate from the data you have collected to make general conclusions about the larger population from which the data sample was derived. For example, it is very difficult to measure the average life of the tubelights manufactured by a company. But, we can collect some pieces of tubelights (called 'sample') and based on the average life of this sample we can take the decision about population. So, sample is a collection of data from the total population as: (a) All people below poverty line in Hyderabad, (b) All people below poverty line in South India. Samples are a subset of the population and always lesser in size. Characteristics of population are known as 'parameters', whereas characteristics of sample are known as 'statistics'. The following exhibit will facilitate us to understand the population and sample. Exhibit 5.1: Post-poll Surveys by CSDS Post-poll surveys were conducted by many research institutions. One such study was by the Centre for the Study of Developing Societies (CSDS), an Indian research institute through its research programme Lokniti along with The Hindu, in 4 different states - Assam, Kerala, Tamil Nadu and West Bengal in 2021 where the Assembly elections were conducted. The surveys were conducted from March 28, 2021 through May 1, 2021 in different phases in the states in the local languages. Contd....

Unit 5: Statistical Inference and Hypothesis Testing 3

The survey was conducted with 3473 voters in Assam, 3424 voters in Kerala, 4354 voters in Tamil Nadu and 4223 voters in West Bengal. The study report presented many tables and charts based on the responses to various questions. The study concluded that the people verdicts in these states were based on the respective local issues in each state.

Source:

**100% MATCHING BLOCK 1/52**

**W**

<https://www.thehindu.com/opinion/op-ed/local-factors-determine-electoral-outcomes-in-states/article34475075.ece>

and [https://www.lokniti.org/POST\\_POLL\\_ANALYSIS\\_2021](https://www.lokniti.org/POST_POLL_ANALYSIS_2021)

As given in the Exhibit 5.1, sample data was collected by CSDS, then the data was classified and analysed. Population: All eligible voters. Population parameter: The proportion of all eligible voters who favor a particular candidate. Sample: The eligible voters surveyed - 3473 voters (Assam), 3424 voters (Kerala), 4354 voters (Tamil Nadu) and 4223 (West Bengal). Sample statistics: The proportion of eligible surveyed voters in each state who favor a particular candidate's win. The following notations are used to denote population parameters and sample statistics in this unit. Population Sample Size  $N$  Mean  $\mu$  Standard Deviation  $\sigma$  Proportion  $p$

The two methods of enumeration are – complete enumeration method or the census method and selective method or sample method. Both methods have their own advantages and disadvantages. 5.3.1 Advantages and Limitations of Sampling Advantages Sampling has the following advantages over a census (study of the entire population): ? Cost: The total cost of a sample will be much less than that of the population. ? Time-lag: With smaller number of observations it is possible to provide results much faster as compared to the total number of observations.

Block II: Statistical Relations and Hypothesis Testing 4 ?

Less Work: Sampling involves less work as compared to a census. Hence, the chances of errors while processing the data are less. ? Destructive Testing: Sometimes we cannot select the whole population for a test. For example, if a manufacturer wants to find out the average life of the tubelights he cannot use every tubelight for testing, as no tubelights would be left for the customers. Hence, sampling is used to determine the average life of a tubelight in such a case. ? Greater Scope: Sampling has a greater scope regarding the variety of information by virtue of its flexibility and adaptability. Finally, if population is too large it is not possible to test the entire population except using sampling.

Limitation of Sampling ? Possibility errors due to sampling may be high for small administrative areas. ? It may not be feasible for problems that require very high accuracy. 5.4 Types of Sampling There are various methods of sampling.

Based on the selection technique, it may be: (i) Non-random or Judgmental Sampling, (ii) Random or Probability Sampling. Non-random or Judgmental Sampling: In this case, sample is normally selected on the personal judgment or convenience of the enumerator. Judgment sampling avoids the statistical analysis that is necessary to make probability samples. For example, the population under study for interview may be people living in Hyderabad. The investigator may choose as a sample, a certain number of his friends living in Hyderabad. He may exercise all care in attempting to make the sample representative of the population. Random or Probability Sampling: Every element of the population has a known chance of being included in the sample. Random sampling is widely used in statistics as it is a more convenient and valid method of sampling.

It may be of different types as illustrated by Figure 5.1. Figure 5.1: Random Sampling

Random Sampling

Simple Random Sampling Stratified Random Sampling Systematic Sampling Cluster Sampling

Unit 5: Statistical Inference and Hypothesis Testing 5 5.4.1

Simple Random Sampling In simple random sampling, we select a certain number of elements randomly. Here each possible sample has an equal chance of being selected and each item in the entire population also has an equal chance of being selected. The selection is usually made with the help of random numbers. Example 1 Suppose there are  $N = 500$  students in a school from which a sample of  $n = 10$  students is to be taken. The students are numbered from 1 to 100. Since our data runs into three digits we use random numbers that contain three digits. All numbers exceeding 500 are ignored because they do not correspond to any serial numbers in the data. No number is repeated twice. In this way, each possible sample has an equal chance of being selected. Further, each item in the entire population also has an equal chance of being selected. 5.4.2 Systematic Sampling Systematic sampling is the most commonly used method of sampling. It is also called simple random sampling. Elements are selected from the population at a fixed interval (at  $k$ th interval). Each element has an equal chance of being selected but each sample does not have an equal chance of being selected. Example 2 Suppose the HAL colony is divided into  $N = 450$  blocks numbered consecutively. A 10 percent sample of blocks is to be taken, which gives a sampling interval of  $k = 10$ . If the random number between 1 and 10 is 5, the blocks with the numbers 05, 15, 20, 25, 30... 450 may be selected as the sample. 5.4.3 Stratified Sampling Stratified sampling is more complex than systematic sampling. The whole population is first divided into homogenous mutually exclusive groups called strata. The segments are based on some predetermined criteria such as geographic location, size or demographic characteristic. After forming the strata, we can either select

at random from each stratum a specified number of elements corresponding to the proportion of that stratum in the population as a whole, or draw an equal number of elements from each stratum and give weight to the results according to the stratum's proportion of total population.

Example 3 The Population of Hyderabad may be divided into different strata such as students, commuters, senior citizens, etc. 5.4.4 Cluster Sampling Cluster sampling divides the population into groups or clusters, and then selects a random sample from these clusters. This approach overcomes the constraints

Block II: Statistical Relations and Hypothesis Testing 6

of costs and time associated with a much dispersed population. Cluster sampling differs from stratified sampling as in the latter, the elements of each stratum are homogeneous (there are relatively minor variations within them). In cluster sampling, the elements of each cluster are not homogeneous but representative of the population. Example 4 We may divide the city of Hyderabad into 4-5 sectors say Zone I, Zone II, Zone III, Zone IV and Zone V. From the addresses on the bills we could classify the customers in the following clusters: Zone I people Zone II people Zone III people Zone IV people Zone V people 5.5 Statistical Errors Estimation based on sample or even population sometimes leads to error.

These errors may be due to the error in collection process, or in process of analysis. For example, if we want to estimate the average savings of workers for the year 2004-2005, we would have to question every worker in our population or some sample of such workers. In either case, the average savings calculated may not be the true average savings for the population. Based on the types of errors, they may be classified into – non-sampling errors and sampling errors. 5.5.1 Non-sampling Errors Non-sampling error occurs at the time of collection, and editing and is common to both sampling and census. Such errors may occur in a sample or in a census. For example, missing some people and double counting others, respondents giving incorrect answers or not answering some questions, imprecise questions, interviewers leading the respondent's answer or giving incorrect information, interviewing the wrong unit, and not capturing or coding the responses correctly. 5.5.2 Sampling Errors Sampling error

**52%**

**MATCHING BLOCK 2/52**

**W**

is the difference between the value of sample statistics and the value of the corresponding population parameter. For example, in case

of mean sample error it can be written as  

$$\bar{X} - \mu$$

Sampling errors occurs because of chance. Normally, sampling error originates at the time of collecting samples.

Unit 5: Statistical Inference and Hypothesis Testing 7 5.6

Sampling Distribution As discussed earlier,

the value of population parameter (population mean, population standard deviation, population proportion, etc.,) is always constant, but the same is not true for sample statistics (sample mean, sample standard deviation, sample proportion of defectives, etc). Different samples from a population may have different sample statistics; they depend on the element selected in the sample. Further, sampling is applicable to random samples only, whose mean can express probability distribution. This distribution is called sampling distribution. Sampling theory helps us to determine whether the differences between two samples are actually due to chance variation or whether they are really significant.

In general, probability distribution of sample statistics is called sampling distribution. 5.6.1

Population

and Sample Proportion Population proportion is the ratio of number of elements in a population with the specific characteristics. In statistics it is denoted by  $p$ . The sample proportion is the ratio of number of elements in a sample with the specific characteristics. So,  $p = \frac{N}{X}$  and  $\hat{p} = \frac{n}{x}$

$x$

Where,  $N$  = Total elements in the population,  $n$  = Total elements in the sample,  $X$  = Total elements in the population that possess that specific character,  $x$  = Total elements in the sample that possess that specific character. 5.6.2 Standard Error of Statistics Standard error of the statistics in the sampling distribution is standard deviation of the collected sample. For example, standard error of mean is equal to the standard deviation of sample distribution. Mean of all collected samples and mean of population is always equal but standard deviation of a sample mean may not be equal to the standard deviation of population, because variability of sample mean may not be same as the variability of population. The standard deviation or standard error of sample mean can be calculated by using formula:  $\sigma_x = \sigma / \sqrt{n}$  This formula is useful when: (i) Sample is selected from finite population and without replacement. (ii) Sample is selected from infinite population with or

Block II: Statistical Relations and Hypothesis Testing 8

without

replacement. If sample is selected from finite population and with replacement, the formula will be as follows:  $\sigma_x = \sigma / \sqrt{n}$

**75%**

**MATCHING BLOCK 3/52**

**W**

$N$  and  $n$  ? ? ? , Where,  $N$  is the size of population and  $n$  is the size of sample.

Sampling is always done without replacement. While sampling  $n$  different items, if  $n$  is small with respect to  $N$ , we may regard "sampling without replacement" to be the same as "sampling with replacement". This is similar to substituting the Hypergeometric Distribution with the Binomial Distribution if  $(n/N)$  is small enough. Here, this will be the case when  $(n/N) \leq 0.05$ . Following are the two important observations regarding sampling distribution: (i) Standard deviation of sampling distribution decreases if the sample size increases. (ii) Population standard deviation is more than the sample standard deviation. The formulae for the Standard Errors (S.E.) of some well-known statistics for random samples are given below:

a. Sample mean

$\sigma_x = \sigma / \sqrt{n}$  b. Sample proportion  $p$ :  $\sigma_p = \sqrt{pq/n}$ , where,  $q = 1 - p$  c. Difference of two sample means  $\bar{X}_1$  and  $\bar{X}_2$ : i.e.,  $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$  Where,  $\bar{X}_1$  and  $\bar{X}_2$  are the means of two random samples of size  $n_1$  and  $n_2$  drawn from two population with standard deviation  $\sigma_1$  and  $\sigma_2$  respectively.

d. Difference of two proportions:  $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$  Where,  $\hat{p}_1$  and  $\hat{p}_2$  are the proportions of two random samples of size  $n_1$  and  $n_2$  drawn from two population and  $\hat{p} = (n_1 \hat{p}_1 + n_2 \hat{p}_2) / (n_1 + n_2)$ .

e. For a finite population of size  $N$ , when a sample is drawn without replacement,  $\sigma_x = \sigma / \sqrt{n} \cdot \sqrt{(N-n)/N}$

If the value of the sampling fraction  $n/N$  is less than 0.05, then finite population correction factor  $\sqrt{(N-n)/N}$  need not be used.

5.6.3 Applications of Standard Error Standard error is used in various applications of statistics. Standard error is a useful tool when it is used in tests of hypotheses or tests of significance. It gives an idea about the reliability and precision of a sample. It helps in determining the limits (confidence limits) within which the parameters are expected to lie.

Unit 5: Statistical Inference and Hypothesis Testing 9

Example 5 A simple random sample of size 81 is drawn from a finite population consisting of 225 units. If the population Standard Deviation is 14, find the standard error of sample mean when the sample is drawn (a) with replacement, (b) without replacement.

a. If the sample is drawn with replacement, then Standard Error of sample mean:  $\sigma_x = \sigma / \sqrt{n} = 14 / \sqrt{81} = 14/9$

b. If the sample is drawn without replacement, then Standard Error of sample mean:  $\sigma_x = \sigma / \sqrt{n} \cdot \sqrt{(N-n)/N} = 14 / \sqrt{81} \cdot \sqrt{(225-81)/225}$

Here,  $\sigma = 14$ ,

$n = 81$ ,  $N = 225$   $\times \sigma = (14/9)$ . 1)  $225 / 81 (225 \times ? = 1.55 \times 0.801 = 1.24$ . Example 6 The following details are available with regard to a hypothesis test on difference between means of two populations:  $n_1 = 16$   $n_2 = 9$  The samples are independently collected and it is assumed that the two populations have the same variance. The pooled estimate of the underlying common population variance is 217.04. What is the estimated standard error of difference between the means? Estimated standard error of difference between the means:  $= \sqrt{\frac{217.04}{16} + \frac{217.04}{9}} = 6.138$   $n_1 = 16$   $n_2 = 9$  ? ? ? ? ?

5.6.4 Shape of the Sampling Distribution Shape

**70%**

**MATCHING BLOCK 4/52**

**W**

of the sampling distribution depends on the shape of the population from which the sample

**70%**

**MATCHING BLOCK 5/52**

**W**

of the sampling distribution depends on the shape of the population from which the sample

has been taken. If the population is normally distributed, then the sample will also be normally distributed. In this case, – Mean of sample

$\bar{x} - \mu = \mu$  population mean – Standard deviation of  $\bar{x} = \sigma / \sqrt{n}$  Figure 5.2: Population Distribution and Sampling Distribution of the Mean

Block II: Statistical Relations and Hypothesis Testing 10

In the normal situation most of the distribution is not exactly normal and the shape of distribution of the sample mean from which population is taken is also different. But for simplicity of decision, we consider most of the distribution as approximate to normal distribution. The approximation of normal distribution can be inferred by the central limit theorem. 5.7 Central Limit Theorem

**73%**

**MATCHING BLOCK 6/52**

**W**

According to the central limit theorem, sampling distribution of the sample mean  $\bar{X}$  is approximately a normal

**79%**

**MATCHING BLOCK 7/52**

**W**

sampling distribution of the sample mean  $\bar{X}$  is approximately a normal distribution with mean  $\mu$  and standard deviation

which is equal to standard error

$\bar{X}$ ,

provided that the sample size  $n$  is sufficiently large. The mean and standard deviation of the sampling mean

$\bar{X}$  are:  $\mu$  Mean of sample mean  $\bar{x} = \mu$  population mean  $\sigma / \sqrt{n}$  Standard deviation of  $\bar{x} = \sigma / \sqrt{n}$  Further, if  $p$  be the proportion of defectives in a random sample of size  $n$  drawn from a population having the proportion of defectives  $p$ , then

**54%**

**MATCHING BLOCK 8/52**

**W**

the sampling distribution of the sample proportion of defectives  $p$  is approximately a normal distribution with mean  $p$  and standard deviation

is equal to standard error of

$p$ ,



provided the sample size  $n$  is sufficiently large. The sample size is said to be large if  $n \geq 30$ , again remember that for  $x = \sigma/\sqrt{n}$ ,  $n/N$  must be less than or equal to 0.05. According to the central limit theorem if  $n$  approaches a large number, the distribution of sample mean approaches normality. Therefore, for large sample sizes ( $n \geq 30$ ), regardless of the original population distribution, the normal distribution is a good approximation to the sampling distribution of  $\bar{x}$  when  $\sigma$  is known, and this can be used as practical application of central limit theorem. Sampling distribution of sample proportion  $\hat{p}$  is approximately normal, if the sample size is sufficiently large. Sample size is considered to be large if  $np \geq 5$  and  $nq \geq 5$ . Example 7 The time that it takes to find a taxi when leaving a restaurant follows a left skewed distribution with a mean of 12 minutes and a standard deviation of 3 minutes. If 100 restaurant patrons

are randomly sampled and the average time that it takes for them to find a taxi is calculated, then what type of sampling distribution will be obtained? By central limit theorem for large samples, the sample mean is approximately normally distributed with mean = population mean, and standard deviation  $\frac{\sigma}{\sqrt{n}}$

n ? ? ? ?  $E_x = ? = 12 \text{ min}$

Unit 5: Statistical Inference and Hypothesis Testing  $11 \times \frac{1}{10} = 1.1$ 

min ? Approximately normal distribution with  $E(x) = 12$  min and  $x ? = 0.30$

min. Example 8 A sample of size 96 has been taken from a population and the estimated standard error of proportion is found to be 0.05. What is the sample proportion?

$p^* = \hat{p} - pq/n = 0.05$  or  $\hat{p} + pq/n = 0.05$  ? ? ?  $\hat{p} - p/p? = 96(0.05)/2$  or  $\hat{p} + p/p? = 0.24$  or  $\hat{p} - p/p 0.24$  ? ? = 0 ?  
 $2(1)(1)(4/0.24) \hat{p} - 2/1$  ? ? ? ? ? ? ? ? ? ? =  $10.20/2 = 0.40.58$

Estimation The value assigned to population parameter based on sample is called estimate. Estimate can be either point estimate or an interval estimate. Point estimate is a single number that estimates the exact value of the population parameter of interest whereas, an interval estimate includes a range of possible values that are likely to include the actual population parameter.

5.8.1 Point Estimates A manager is sometimes interested in single value estimator than in interval. In point estimates normally we select a single value to a population parameter. The basic characteristic in choosing a sample statistic as a point estimate of

**66%**      **MATCHING BLOCK 9/52**

W

the value of the population parameter is that the sample statistic be an unbiased estimator.

Normally a statistician is interested in selecting unbiased and small variance. 5.8.2 Characteristics of a Good Estimator

The sample statistics (mean, median etc.), used for estimating a population parameter is called an estimator. The unbiased estimator should have the following characteristics: Unbiased: In unbiased estimates of a population the expected value is equal to that of the parameter.

Block II: Statistical Relations and Hypothesis Testing 12

Consistency: Unbiased estimators are called as

**97%**      **MATCHING BLOCK 10/52**

W

consistent if the difference between the estimator and parameter grows smaller as the sample size grows larger.

**97%      MATCHING BLOCK 11/52**

W

consistent if the difference between the estimator and parameter grows smaller as the sample size grows larger.

Efficiency:



63%

**MATCHING BLOCK 12/52**

**W**

If there are two estimators then the estimator whose variance is less is said to be relatively efficient.

63%

**MATCHING BLOCK 13/52**

**W**

If there are two estimators then the estimator whose variance is less is said to be relatively efficient.

We can show the standard error of the sample median as greater than  $X \sigma$ . Therefore  $X$  is more efficient. Sufficiency: This is the degree to which all possible information in the sample is used, to estimate the corresponding parameter. 5.8.3 Point Estimator of Population Mean The best estimator of population mean is the sample mean as it possesses the characteristics of a good estimator. Suppose, we want to find out the average starting salary offered by a software finance company to its employees in Hyderabad. As the data in the population may be too large, we may decide to study a random sample of such companies and thereby draw conclusions about the population mean. In this case, the sample mean  $\bar{X}$  is a point estimator of the population mean  $\mu$ . 5.8.4 Point Estimator of Population Proportion Sample proportion possesses desirable characteristics of the estimator; therefore the sample proportion  $p$  is an estimator of the population proportion. Often we are interested in population proportions i.e., the ratio of observations in a population which can either have or do not have a certain characteristic. For example, (i) The proportion of students in class V whose age is below 10. (ii) The proportion of companies which have profit margin ratio of more than 20. In this case, we can study a random sample. If a certain percentage in a sample has a specific character we can use this as a rough estimate for whole proportion. 5.8.5 Point Estimator of Population Variance The sample variance can be used as a point estimator

70%

**MATCHING BLOCK 14/52**

**W**

of the population variance, and the sample standard deviation is a point estimator of the

population standard deviation. 5.8.6 Interval Estimates Point estimator is often insufficient and gives wrong idea about population. To overcome this problem we normally use interval estimate, which can give more clear idea about the population. Terminology for Interval Estimate An interval estimate is a range of values within which the actual value of the population parameter may fall.

Unit 5: Statistical Inference and Hypothesis Testing 13

Interval limits are the lower and upper values of the interval estimate. A confidence interval is an interval estimate for which there is a specified degree of confidence that the actual value of the population parameter will fall within the interval. These are normally probability statements. A confidence coefficient is the proportion of intervals that would include the population value in the long run if the process leading to the creation of the interval were repeated many times. It can be expressed as a fraction or a decimal. A confidence level like the confidence coefficient, expresses the degree of certainty that the interval will include the actual value of the population parameter. It is expressed as a percentage for example, a 95 confidence coefficient is the equivalent of a 95 percent confidence level. Accuracy is

66%

**MATCHING BLOCK 15/52**

**W**

the difference between the observed sample statistic and the actual value of the population parameter

66%

**MATCHING BLOCK 16/52**

**W**

the difference between the observed sample statistic and the actual value of the population parameter

being estimated. This may also be referred to as estimation error or sampling error. Confidence Interval for a Population Mean (σ known) When the sample size is more than 30, i.e.,  $n \geq 30$  as per central limit theorem, we can assume that underlying population is normally distributed and interval can be estimated using the formula given below:  $\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$  where,  $Z$  = the Z value corresponding to the level of confidence. 5.8.7 The Student t Distribution When the sample size is small the normal distribution cannot give the appropriate result. So we can use t distribution to find the interval when: (a) Sample size is less than 30 ( $n < 30$ ) (b) Population standard deviation is not known. To find the t value in the table, perform the following steps: (i) Determine the confidence level. (For this example, assume 95%). (ii) Subtract this decimal from 1.00 to calculate the significance level (

$\alpha$ ). (For this example  $\alpha = 1.00 - 0.95 = 0.05$ ). (

iii) Divide the result in half. (For this example, we would get  $0.05/2 = 0.025$ ). (iv) Determine the number of degrees of freedom,  $d.f = n - 1$ . (For example, for a sample size of 20, the  $d.f = 20 - 1 = 19$ ). (v) Now we would look in the column headed for = 0.025 and row for  $d.f = 19$  and find the respective value 2.0930.

Percentiles of  
the Student

t Distribution Single Tail Probabilities of a Type I Error  $d.f/\alpha$

0.100	0.050	0.025	0.010	0.005
19	1.3277	1.7291	2.0930	2.5395
2.8609	In this case, the confidence interval is			

$\bar{X} \pm t(s/\sqrt{n})$

Block II: Statistical Relations and Hypothesis Testing 14 x 7

Example 9 A random sample of 64 vegetable vendors in a city was taken. The average price of potatoes was found to be Rs.7.50 per kg with a standard deviation of 48 paise per kg. You are required to find out the interval within which the true mean price per kg of potatoes for the entire city will fall 95% of the time. Given:  $n = 64$ ,  $\bar{X} = 7.50$  (Rs./kg.),  $s = 48$

(Paise/kg) = 0.48 (Re/kg) Since the standard deviation of population is not known, the sample standard deviation will be used as an estimate of the population standard deviation. Estimated standard error of mean,  $s/\sqrt{n} = 48/\sqrt{64} = 6$  95%

confidence interval is required.  $Z$  (from normal tables) = 1.96 Lower confidence limit =  $\bar{X} - 1.96 s/\sqrt{n} = 7.50 - 1.96(6) = Rs.7.38/kg$ . Upper confidence limit =  $\bar{X} + 1.96 s/\sqrt{n} = 7.50 + 1.96(6) = Rs.7.62/kg$ . Check Your Progress - 1 1. According to

if  $n$  approaches a large number, the distribution of sample mean approaches normality 2.

A simple random sample of size 100 is drawn from a finite population consisting of 400 units. If the population Standard Deviation is 18, find the standard error of sample mean when the sample is drawn without replacement. a. 1.65 b. 1.42 c. 1.68 d. 1.56 e. 1.98 3. If the population variance and the sample size are 26 and 5 respectively, then the standard error of the sample mean will be a. 0.192 b. 0.438 c. 0.980 d. 2.280 e. 11.62.

x ? x ?

Unit 5: Statistical Inference and Hypothesis Testing 15 4.

A random sample of 100 vegetable vendors in a city was taken. The average price of potatoes was found to be Rs.36.50 per kg with a standard deviation of 90 paise per kg. You are required to find out the interval within which the true mean price per kg of potatoes for the entire city will fall 95% of the time. a. 36.32,36.67 b. 36,38 c. 36.5,37.4 d. 35.8,37.4 e. 36.32,37.4

5. The set of data for which findings are to be generated is called..... 6. Characteristics of sample are called a.

Parameters b. Statistics c. Probability d. Hypothesis e. Inference 7. Which of the following is not an advantage of sampling? a. Less cost b. Faster data c. Lesser work d. Destructive testing e. Highly accurate data 8. Which of the following is not a type of random sampling? a. Simple Random sampling b. Stratified random sampling c. Systematic sampling d. Cluster sampling e. Judgmental sampling 9. Fill in the blank in the following statement. Non sampling error occurs at the time of collection and ..... 5.9

Testing of Hypothesis Testing of hypothesis is a new concept for a new reader, but it's quite a familiar aspect of a real life problem. For example, a placement consultant recruiting fresh graduates from colleges for various companies observed that the mean

Block II: Statistical Relations and Hypothesis Testing 16

annual salary offered to the candidates placed by him in the retailing firms was Rs.480000 in 2019. From January 2020 to March 2020, the consultant placed as many as 150 fresh graduates in the retailing firms. The consultant is of the opinion that the mean annual salary offered by the retailing firms to the candidates placed by him from January 2020 to March 2020, is more than the mean salary offered by the same employers in 2019. He has collected a sample of 36 candidates placed in 2020 in the retailing firms and has observed that the mean annual salary offered to them is Rs.504000 with a standard deviation of Rs.18000. It is to be determined whether the candidates placed by the consultant from January 2020 to March 2020, in the retailing firms, have been offered a mean annual salary greater than the mean annual salary offered to them in 2019. How can we solve this type of problem? Normally, the following steps are required in testing hypothesis. (a) State null and alternative hypothesis; (b) State decision rule; (c) Compute test statistic; (d) Draw a picture showing acceptance region(s), rejection region(s), and position of the decision rule; (e) Make the decision; and (f) Relate the decision to the context of the problem. The first step in hypothesis testing is to make an assumption about the population. In this example, the consultant opinion is the assumption about the population. The next step is to gather sample and determine the sample statistic. There are actually two hypotheses that are tested. One is called null hypothesis and other is called alternative hypothesis.

The null hypothesis asserts that there is no (significant) difference between the statistic and the

population parameter. Alternative hypothesis contradicts the null hypothesis  $H_0$ , and is denoted by the symbol  $H_1$ . In general, the Null Hypothesis is  $H_0$ :

$\mu = 0$  ,  $H_0 : \mu \geq 0$  or

**87%**

**MATCHING BLOCK 17/52**

**W**

$H_0 : \mu \geq 0$  and the Alternative Hypothesis may be: i.  $H_1 : \mu < 0$  or

$\mu < 0$  or  $\mu \geq 0$  or ii.  $H_1 : \mu > 0$  ,

or iii.  $H_1 : \mu \geq 0$  . The

corresponding tests of hypotheses are called two-tailed (or two-sided), right-tailed (one-sided) and left-tailed (one-sided) tests respectively. The Alternative Hypothesis in (i) is known as a two-tailed alternative and in (ii) and (iii) is known as right-tailed and left-tailed alternatives respectively. In our example, Null hypothesis  $H_0$  : population mean  $\mu = 480000$  Alternative hypothesis  $H_1$  : population mean  $\mu < 480000$ . This is one tail right hand test. Normally, to test the validity of our hypothesis the difference between the hypothesized value and the actual value of the sample statistic will be determined. If the difference between the hypothesized population parameter and the actual value is large then we automatically reject our hypothesis. If it is small, we accept it.

Unit 5: Statistical Inference and Hypothesis Testing 17 5.9.1 Level of Significance:  $\alpha$

Level of significance defines unlikely values of sample statistic if null hypothesis is true i.e., the percentage cases that would be rejecting the hypothesis when it is actually true and is symbolized as  $\alpha$ . For example, if the significant level is 5%, it means out of 100 we would reject about 5 cases when they should be actually accepted.  $(1 - \alpha)$  is the probability of accepting a true null hypothesis, and is referred to as the confidence level. If the significant level is 5%, we can say that we are 95% confident that we have made the right decision. Acceptance and rejection region can be drawn in two ways: Scale of the Original Variable Method: In this method, we can compute limit(s) and determine where the sample statistic lies in relation to the limit(s). Standardized Scale Method: In this method, we can compute a standardized statistic based on the sample data and compare where the standardized statistic lies in relation to the critical value as determined by the test's significance level. 5.9.2 Two-Tailed and One-Tailed Tests As we discussed earlier, hypothesis test may be one-tail test or two-tail test. In one-tail test, rejection region is located in only one-tail of the sampling distribution whereas in two-tail test rejection region is located in both the tails of the sampling distribution. If the test is two-tailed, then we can write  $H_1 : \mu \neq 0$  and if it is one-tailed (i.e., either right-tailed or left-tailed), then we can write  $H_1 : \mu > 0$  or  $H_1 : \mu < 0$ .

Figure 5.3: A Picture Showing Acceptance Region, Rejection Region(s) Finding Critical 'z' Values for Two-Tailed Test If the significant level is given as 5%, how can we find the value of Z in a one-tail test? Calculate the value of  $\alpha/2$

and deduct from 0.50 in the above example; it would be  $0.50 - 0.025 = 0.475$ . Look for the value of 0.475 in the Z-table.

Z 0.05 0.06 0.07 1.8 0.4678 0.4686 0.4693 1.7 0.4744 0.4750 0.4756

Block II: Statistical Relations and Hypothesis Testing 18 5.10

Large Sample Test for Mean (one-tailed test) Example 10 Let us solve the following

example: A placement consultant recruiting fresh graduates from colleges for various companies observed that the mean annual salary offered to the candidates placed by him in the retailing firms was Rs.480000 in 2019. From January 2020 to March 2020, the consultant placed as many as 150 fresh graduates in the retailing firms. The consultant is of the opinion that the mean annual salary offered by the retailing firms to the candidates placed by him from January 2020 to March 2020, is more than the mean salary offered by the same employers in 2019. He has collected a sample of 36 candidates placed in 2020 in the retailing firms and has observed that the mean annual salary offered to them is Rs.504000 with a standard deviation of Rs.18000. It is to be determined whether the candidates placed by the consultant from January 2020 to March 2020, in the retailing firms, have been offered a mean annual salary greater than the mean annual salary offered to them in 2019.

First Method: Standardized Scale Method Here,  $H_0 : \mu = 480000$   $H_1 : \mu > 480000$  Sampling fraction =  $n/N = 36/800 = 0.045 > 0.05$  Hence, this should be treated as sampling from infinite population. Standard error of mean,  $\hat{\sigma}_x = \sigma/\sqrt{n} =$

Since the sample size is  $36 < 30$  the normal distribution can be used, Standardized value of sample mean,  $Z = (\bar{x} - \mu) / (\sigma/\sqrt{n}) =$   
 $=$  At the significance level of 5%, the critical value for the test = 1.64 Standardized value (8) < Critical value (1.64).

Figure 5.4: Accept/Reject  $H_0$  ? Examples using Critical Values 480000 480492

Unit 5: Statistical Inference and Hypothesis Testing 19

Hence, the sample statistic falls in the rejection region. So, the null hypothesis is rejected and alternative hypothesis is accepted. Hence, at 5% significance level, it can be concluded

that the mean annual salary offered to the candidates placed in the retail firms from January 2020 up to March 2020, is more than Rs.480000. Second Method: Scale of the Original Variable Method Here,  $H_0 : \mu = 480000$   $H_1 : \mu > 480000$  Sampling fraction =

$n/N = 36/800 = 0.045 > 0.05$

Hence, this should be treated as sampling from infinite population. Standard error of mean,  $\hat{\sigma}_x = \sigma/\sqrt{n} =$

Since the sample size is  $36 < 30$  the normal distribution can be used. At the significance level of 5%, the critical value for the test = 1.64 Standardized value (8) < Critical value (1.64).  $(\bar{x} - \mu) / (\sigma/\sqrt{n}) < 1.64$

Since,  $\bar{x} = 500000$  Hence, the sample statistic falls in the rejection region. So, the null hypothesis is rejected and alternative hypothesis is accepted. Hence, at 5% significance level, it can be concluded

that the mean annual salary offered to the candidates placed in the retail firms from January 2020 up to March 2020, is more than Rs.480000.

Example 11 The mean lifetime of 100 electric bulbs produced by a manufacturing company is estimated to be 1570 hours with a standard deviation of 120 hours.

If  $\mu$  is the mean lifetime of all the bulbs produced by the company, test the hypothesis  $\mu = 1600$  hrs, using a level of significance of 0.05. This is a two-tailed test as we want to test if the hypothesis for mean lifetime is exactly 1600 hrs.  $N = 100$ ;  $\bar{x} = 1570$ ;  $S = 120$ ;  $\mu = 1600$   $H_0 : (\mu = 1600)$ ;  $H_1 : (\mu \neq 1600)$

Block II: Statistical Relations and Hypothesis Testing 20 The test statistic is:  $Z = (\bar{x} - \mu) / (S/\sqrt{n}) = (1570 - 1600) / (120/\sqrt{100}) = -2.5$

The Critical Region at 5% level is  $[Z \leq -1.96]$ ;

calculated value falls in C.R. Hence, the null hypothesis  $H_0$  is rejected. Hence, it can be concluded that the mean lifetime is not 1600 Hrs. 5.11

Small Sample Test for Mean For a small sample, where the size is less than 30 and population standard deviation is not known, we can use, t-distribution as test statistics.

Example 12 A commodity merchant knows that the mean retail price of a specific variety of rice three months ago was Rs.14.50 per kg. In the current month the merchant has collected the information on the price charged for the same variety of rice by 16 randomly selected merchants in the same city. It was found from the sample that the mean retail price was Rs.15.00 per kg and the standard deviation was Rs.1.25 per kg. It is to be tested whether the mean retail price of the rice in the current month is more than Rs.14.50 per kg. At a significance level of 5 percent, what is the conclusion?  $H_0: 14.50 \leq \mu$ ;  $H_1: 14.50 < \mu$ ;  $\alpha = 0.05$  Estimated standard error of mean:

**87%**

**MATCHING BLOCK 18/52**

**W**

$\bar{x} \pm 1.25 \sqrt{0.3125}$   $n = 16$  Since the population standard deviation is not known and the sample size is

less than 30, the t-distribution will be used. Degrees of freedom =  $n - 1 = 16 - 1 = 15$  Since this is a right tailed test with 0.05, the rejection area is 0.05 under the right tail. Hence, we should look under the 0.10 column in the t-distribution (0.10 area in both tails combined). Critical t-value = 1.753 Standardized sample statistic,  $t = 0$

$\bar{x} = 15.00$ ,  $\mu_0 = 14.50$ ,  $s = 1.25$ ,  $n = 16$  Figure 5.5: The Graph of pdf for Standard Normal Distribution

Unit 5: Statistical Inference and Hypothesis Testing 21

We find that the sample statistic falls in the acceptance region. Hence, we accept the null hypothesis,  $H_0$ . It is concluded that the mean price of the rice is not more than Rs.14.50 per kg. 5.12 Hypotheses Involving Means of Two Populations (Large Sample) Two samples are said to be independent if they are drawn from two different populations. This section discusses how the hypothesis test can be conducted to calculate the difference between two means of large and independent populations. Suppose for two independent populations, population means are  $\mu_1$  and  $\mu_2$

respectively. Two samples are taken, one from each population, of sizes  $n_1$  and  $n_2$  respectively. We need to test whether  $\mu_1 = \mu_2$ .  $\bar{x}_1$  = The mean of the sample drawn from population 1  $\bar{x}_2$  = The mean of the sample drawn from population 2. From the central limit theorem,  $\bar{x}_1$  and  $\bar{x}_2$

are approximately normally distributed with standard deviation of  $\frac{\sigma_1}{\sqrt{n_1}}$  and  $\frac{\sigma_2}{\sqrt{n_2}}$ .

Using this, we can calculate the difference between the two means of large and independent populations. Standard error:  $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Null and alternative hypothesis are any of the following:  $H_0: \mu_1 = \mu_2$   $H_1: \mu_1 \neq \mu_2$  (two tail)  $H_0: \mu_1 \geq \mu_2$   $H_1: \mu_1 < \mu_2$  (right tail)  $H_0: \mu_1 \leq \mu_2$   $H_1: \mu_1 > \mu_2$  (left tail) This can also be written as follows:  $H_0: \mu_1 - \mu_2 = 0$   $H_1: \mu_1 - \mu_2 \neq 0$  (two tail)  $H_0: \mu_1 - \mu_2 \geq 0$   $H_1: \mu_1 - \mu_2 < 0$  (right tail)  $H_0: \mu_1 - \mu_2 \leq 0$   $H_1: \mu_1 - \mu_2 > 0$  (left tail)  $\bar{x}_1 - \bar{x}_2 \pm (x_{\alpha/2}) (\sigma_{\bar{x}_1 - \bar{x}_2})$  Test statistic =  $\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$

Because the difference between the normal distributes and random variables is also normal, we can take sample standard deviation in place of population standard deviation if the sample is sufficiently large; in this case, it is,  $\bar{s}_1$  and  $\bar{s}_2$

$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Block II: Statistical Relations and Hypothesis Testing 22

Example 13 Two medical research laboratories, Alpha Pharma Ltd. (APL) and Orient Laboratories Ltd. (OLL), have been independently conducting research for producing drugs that will provide relief to those suffering from asthma. Very recently they have produced drugs for providing relief to the asthma patients. The drug produced by APL was tested on a group of 80 asthma patients and it provided an average of 10.5 hours of relief, with a standard deviation of 1.5 hours. The drug produced by OLL was tested on a group of 64 asthma patients and it provided an average of 9.6 hours of relief, with a standard deviation of 2.5 hours. It is to be tested whether the

mean hours of relief provided by the drug produced by APL is more than the mean hours of relief provided by the drug produced by OLL. Let the following notations be used:  $\mu_1$  = Population mean hours of relief provided by the drug of APL.  $\mu_2$  = Population mean hours of relief provided by the drug of OLL.  $H_0: \mu_1 = \mu_2$ ;  $H_1: \mu_1 > \mu_2$  Estimated standard

error of the difference between the two means:  $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.5)^2}{80} + \frac{(2.5)^2}{64}} = 0.3547$  hours Standardization of difference between sample means:  $Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{10.5 - 9.6 - 0}{0.3547} = 2.54$

$H_0: \mu_1 = \mu_2$   $H_1: \mu_1 > \mu_2$   $\alpha = 0.05$   $Z_{\alpha} = 1.645$

Since  $Z = 2.54 > Z_{\alpha} = 1.645$ , we reject  $H_0$  and conclude that the mean hours of relief provided by the drug produced by APL is more than the mean hours of relief provided by the drug produced by OLL.

Critical value at 4 percent level of significance = 1.75 Figure 5.6: Critical Resin for Z Test So, we can see that the difference between the sample mean falls in the rejection region. Hence, null hypothesis is rejected and alternative hypothesis is accepted. Therefore, at a significance level of 4%, we can conclude that the mean number of hours of relief provided by the drug produced by APL is significantly higher than OLL.

Unit 5: Statistical Inference and Hypothesis Testing 23 5.13

Hypotheses Involving Means of Two Populations (Small Sample) For small sample, the pooled estimate of  $\sigma^2$  and  $\sigma^2$  can be calculated using the following equation:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad s_p = \sqrt{s_p^2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Example 14 The following details are available with regard to a hypothesis test on difference between means of two populations:  $n_1 = 16$ ,  $s_1 = 16$ ,  $n_2 = 9$ ,  $s_2 = 12$  The samples are independently collected and it is assumed that the two populations have the same variance. What is the estimated standard error of difference between means? Pooled estimate of the underlying common population variance,  $s_p^2 = \frac{(16-1)16^2 + (9-1)12^2}{16+9-2} = 217.04$

Estimated standard error of difference between means:  $s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 217.04 \times 6.138 = 1331.4$

Type I and Type II Errors In the testing of hypothesis, the null hypothesis is rejected or accepted on the basis of the sample statistic. Thus, the decision is always error prone. The possible errors in the testing of hypothesis are: Type I Errors – The probability of Type I error is also called the significance level of the test. If we use higher significance level for testing hypothesis, the higher is the probability of committing Type I error. Type II Errors occurs in Hypothesis testing when the Null Hypothesis is false and accepted. It is also represented by  $\beta$  or power of test.

Actual Decision Accept  $H_0$  Reject  $H_0$   $H_0$  is True Correct Decision Wrong Decision (Type I error)  $H_0$  is False Wrong Decision (Type II error) Correct Decision

Block II: Statistical Relations and Hypothesis Testing 24

There exists a tradeoff between the two types of

errors. The probability of making one type of error can be reduced only if we are willing to increase the probability of making the other type of error.

Example 15 An electronic equipment manufacturer uses hypothesis testing for the inspection of incoming raw material. The null hypothesis is that the number of defective items in the lot is equal to the acceptable limit. The alternative hypothesis is that the number of defective items in the lot is more than the acceptable limit. If the manufacturer performs Type I error (i.e., Reject null hypothesis when it is true), then he will reject the good lot. If he performs Type II error (i.e., Accept null hypothesis when it is false), then he will accept the lot with number of defectives more than the acceptable limit. In this particular test, Type II error is more serious, so the  $\beta$  of the test should be very low. 5.15 Large Sample Test for Proportion Often we conduct the test of hypothesis about a population proportion. For example, 40% of the employees included among them those working effectively for less than 8 hours a day. The Managing Director of the company wants to know whether this percentage holds good. To test the hypothesis that the proportion  $p$  in the population has a specified value  $p_0$ , we have to first check the size of sample. The procedure includes these steps: For a large sample size  $n$  ( $n \geq 30$ ) sampling distribution of  $p$  is approximately normal. The Null Hypothesis is  $H_0 : p = p_0$ . The Alternative Hypothesis may be (i)  $H_1 : p \neq p_0$  or (ii)  $H_1 : p > p_0$ , or (iii)  $H_1 : p < p_0$ . The test statistic  $z = \frac{p - p_0}{\sigma_p}$  where, Standard error of  $p = \sqrt{p_0 q_0 / n}$ .

Example 16 Western Electricals Company manufactures heavy duty voltage stabilizers. The General Manager of the company wants to compare the performance of the stabilizers manufactured by the company with the industry standards. He knows that 18% of all the stabilizers sold in the industry require repairs during the first year of sale. The company sampled 144 customers and found that in case of 36 customers the stabilizers required repairs in the first year of sale. Using normal distribution and a significance level of 5% you are required to determine if the performance of the stabilizers manufactured by the company is different from the industry standards. Sample size,  $n = 144$

Unit 5: Statistical Inference and Hypothesis Testing 25 Sample proportion,  $p = 36/144 = 0.25$   $H_0 : p = 0.18$  ( $= p_0$ );  $H_1 : p \neq 0.18$ ;  $\alpha = 0.05$  Standard error of proportion:  $\sigma_p = \sqrt{p_0 q_0 / n} = \sqrt{0.18(1-0.18) / 144} = 0.032$  Standardizing the sample proportion:  $z = \frac{p - p_0}{\sigma_p} = \frac{0.25 - 0.18}{0.032} = 2.1875$



or 2.19 Figure 5.7: Critical Resin for Test of Proportion From the normal distribution curve, we can find out that the sample proportion falls in the rejection region. Hence, we infer that the performance of the stabilizers manufactured by the company significantly differs from the industry standards. Example 17 Consider the above example, and determine if the performance of the stabilizers manufactured by the company is less than the industry standards.  $H_0 : p = 0.18$  ( $= p_0$ );  $H_1 : p < 0.18$  This is left hand test so

$\alpha = 0.05$  ?

and the rejection level will be  $Z > -1.96$ , but  $Z = 2.19$ , so it falls outside the rejection region. Hence, we infer that the performance of the stabilizers manufactured by the company is not less than the industry standards. 5.16 Hypotheses Involving Proportions of Two Populations For any of the following hypotheses: 1 2 1 2 0

**28% MATCHING BLOCK 20/52**

**W**

a H : p p H : p p ? ? (Two tail) 1 2 1 2 0 a H : p p H : p p ? ? (Right tail) 1 2 1 2 0 a H : p p H : p p ? ? (Left tail) Or 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Two tail) 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Right tail) 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (

**19% MATCHING BLOCK 21/52**

**W**

p p H : p p ? ? (Two tail) 1 2 1 2 0 a H : p p H : p p ? ? (Right tail) 1 2 1 2 0 a H : p p H : p p ? ? (Left tail) Or 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Two tail) 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Right tail) 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (

**30% MATCHING BLOCK 19/52**

**W**

p H : p p ? ? (Two tail) 1 2 1 2 0 a H : p p H : p p ? ? (Right tail) 1 2 1 2 0 a H : p p H : p p ? ? (Left tail) Or 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Two tail) 1 2 1 2 0 a H : p p 0 H : p p 0 ? ? ? ? (Right tail) 1 2 1 2 0 a H : p p 0 H : p

Left tail)

Block II: Statistical Relations and Hypothesis Testing 26

To estimate the overall proportion of successes, we have to calculate the combined proportions from both the samples.

1 1 2 2 1 2

$n_1 p_1 n_2 p_2$  ? ? ? Standard error =  $2 \sqrt{p_1 q_1 p_2 q_2}$  ? ? ? The test statistic is:  $1 2 1 2 1 1 1 1 1 2 (p_1) (p_2) z = \frac{\hat{p} - p}{\sqrt{\hat{p} \hat{q}}}$  p q p q n n ? ? ? ?

which has the standard normal distribution. Example 18 A survey firm conducts door-to-door surveys on a variety of issues. Some individuals cooperate with the interviewer and complete the interview questionnaire while others do not.

The following data are available. Respondents Sample Size Number Not Cooperating Men 200 110 Women 300 210

It is to be tested at a significance level of 0.05, whether the proportion of men and women who cooperate with the interviewer are equal. Which of the following can be inferred from the test?  $p_1$  : proportion of men cooperating  $p_2$  :

proportion of women cooperating. Null hypothesis:  $p_1 = p_2$  Alternative hypothesis:  $p_1 \neq p_2$  1 90 p 0.45 200 ? ? ; 22 90 p 0.30 300 ? ? Estimated value of

**61% MATCHING BLOCK 22/52**

**W**

the population proportion,  $p = 1 1 1 2 1 2 (n_1 p_1) (n_2 p_2) (200 \times 0.45) (300 \times 0.30) 180 0.36 n_1 n_2 200 300 500 ? ? ? ? ? ? 1 2 p$  p ? ? (estimated) =  $1 2 p(1 p) p(1 p) n_1 n_2 ? ? ? ? ? ? ? ? ? ? ? ? = 0.36(1 - 0.36) 0.36(1 - 0.36) 200 300 ? ? ? ? ? ? ? ? ? ? ? ? ? ?$  ? ?

Test statistic:  $(0.45 - 0.30) / 0.342 = 0.04382$  ? ? ? Critical values:  $\pm 1.96$

Unit 5: Statistical Inference and Hypothesis Testing 27

The test statistic goes beyond the critical value in the right tail. Hence, we reject null hypothesis. We can conclude that the proportion of men and women who cooperate with the interviewer are not equal. Check Your Progress - 2 10.

Type II Errors occurs in Hypothesis testing when the Null Hypothesis is ----- 11.



A sample of size 100 has been taken from a population and the estimated standard error of proportion is found to be 0.04. What is the sample proportion?

a. 0.4 b. 0.3 c. 0.1 d. 0.2 e. 0.6 12. The mean lifetime of 1000 electric bulbs produced by a manufacturing company is estimated to be 1450 hours with a standard deviation of 80 hours.

If  $\mu$  is the mean lifetime of all the bulbs produced by the company, and we wish to test the hypothesis  $\mu = 1500$  hrs, compute the Z statistic. a. 19.76 b. -19.76 c. 21.12 d. -21.12 e. 26.76 13. The drug produced by CIPLA was tested on a

group of 80 asthma patients and it provided an average of 12 hours of relief, with a standard deviation of 0.5 hours. The drug produced by ABOTT was tested on a group of 64 asthma patients and it provided an average of 10 hours of relief, with a standard deviation of 0.4 hours. It is to be tested whether the

mean hours of relief provided by the drug produced by CIPLA is more than the mean hours of relief provided by the drug produced by ABOTT. Compute estimated standard error of the difference between the two means a. 0.25 b. 0.15 c. 0.075 d. 0.75 e. 0.015

Block II: Statistical Relations and Hypothesis Testing 28 5.17

Summary ? Sampling methods are widely used in statistics for estimating character of population. They are used due to various characteristics like low cost and time etc. ? Non-random and Random sampling are the two main types of sampling used in estimation.

Simple random sampling, Stratified random sampling, Systematic sampling, Cluster sampling are the different types of random sampling.

Estimation based on sample or even population sometimes leads to error.

Sampling distribution helps to determine whether the differences between two samples are actually due to chance variation or whether they are really significant. In general, probability distribution of sample statistics is called sampling distribution. ?

The standard deviation of the collected sample is normally termed as standard error of the statistics, which is the most useful tool used in tests of hypotheses.

It helps in determining the limits (confidence limits) within which the parameters are expected to lie.

The most common test is to test the significance of the mean and proportion. Some other advanced tests are test of difference of mean and test of difference of proportion. 5.18

Glossary Alpha: The probability of a Type I error. Alternate Hypothesis: The conclusion we accept when the data fails to support the null hypothesis. Beta: The probability of a Type II error. Central Limit Theorem: A result assuring that

**46% MATCHING BLOCK 23/52**

**W**

the sampling distribution of the mean approaches normality as the sample size increases, regardless of the shape of the population distribution from which the sample is

selected.

Cluster Sampling: The population is divided into groups, or clusters, to select a random sample of these clusters.

Confidence Limits: The upper and lower boundaries of a confidence interval. Estimate: A specific observed value of

**87% MATCHING BLOCK 24/52**

**W**

an estimator. Estimator: A sample statistic used to estimate a population parameter.

Hypothesis: An assumption made about population parameter. Null Hypothesis: The hypothesis, or assumption, about a population parameter we wish to test, usually an assumption of the status quo. One-Tailed Test: A hypothesis test in which there is only one rejection region. Population: The set of data to which findings are to be generalized. Sample: A collection of data from population.

Unit 5: Statistical Inference and Hypothesis Testing 29

**Stratified Sampling:** The population is divided into relatively homogeneous groups, called strata. Then either at random from each stratum a specified number of elements corresponding to the proportion of that stratum in the population as a whole are selected or an equal number of elements from each stratum are drawn and given weight to the results according to the stratum's proportion of total population. **Systematic Sampling:** Elements are selected from the population at a uniform interval that is measured in time, order, or space.

**Two-Tailed Test:** A hypothesis test in which

the null hypothesis

is rejected if the sample value is significantly higher

or lower than the hypothesized value of the population parameter. **Type I Error:**

Rejecting a null hypothesis when it is true. **Type II Error:** Accepting a null hypothesis when it is false. 5.19

**Suggested Readings/Reference Material** 1. Gupta, S. P. Statistical Methods. 46th Revised ed. New Delhi: Sultan Chand & Sons. 2021. 2. I. Levin Richard, H. Siddiqui Masood, S. Rubin David, Rastogi Sanjay. Statistics for Management. Pearson Education; Eighth edition, 2017. 3. Gerald Keller. Statistics for Management and Economics. Cengage, 2017. 4. Arora, P. N., and Arora, S. CA Foundation Course Statistics. 6th ed. S Chand Publishing, 2018. 5. Mario F Triola. Elementary Statistics. 13th ed., 2018. 6. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. Statistics for Business and Economics. 13th Edition, Cengage Learning India Pvt. Ltd., 2019. 7. S D Sharma. Operations Research. Kedar Nath Ram Nath, 2018. 8. Hamdy A. Taha. Operations Research: An Introduction. 10th ed., Pearson, 2016. 9. Malhotra, N. (2012), Marketing Research: An Applied Orientation, 7th ed., Pearson, 2019. 10. Cooper, D.R. and Schindler, P.S. and J. K. Sharma (2018), Business Research Methods, 12th edition, McGraw-Hill Education. 5.20

**Self-Assessment Questions** 1. Describe the process of hypothesis testing. 2. Describe the null and alternate hypothesis typically represented in the hypothesis testing process. 3. Define the significant level.

**Block II: Statistical Relations and Hypothesis Testing** 30 4.

State the condition under which it is appropriate to use one-tailed test and two-tailed test. 5. What do you mean by power of test? 5.21 **Answers to Check Your Progress Questions** 1. The central limit theorem 2. (d) 1.56 3. (c). 980 4. (a) 36.32, 36.67 5. Population 6. (b) Statistics 7. (e) Highly accurate data is not possible in sampling 8. (e) Judgmental sampling 9. Editing. 10. False and accepted. 11. (d) 0.2 12. (b) -19.76 13. (c) 0.075

**Unit 6 Correlation and Linear Regression Structure** 6.1 Introduction 6.2 Objectives 6.3 Correlation 6.4 Rank Correlation 6.5 Simple Linear Regression 6.6 Coefficient of Determination 6.7 The Standard Error 6.8 Making Inference about Population Parameter 6.9 Geometry of Regression 6.10 Misuses and Caveats 6.11 Applications in Finance 6.12 Regression using Microsoft-Excel 6.13

**Summary** 6.14 **Glossary** 6.15 **Suggested Readings/Reference Material** 6.16 **Self-Assessment Questions** 6.17 **Answers to Check Your Progress Questions** 6.1 **Introduction** In the previous unit, you learnt different types of sampling methods, inferences and tests of hypothesis. In this unit, you will learn the correlation between two variables and its significance. In business and finance, it may be observed that many variables are related, change in one variable may bring about change in other variables; or change in some variables may bring change in one variable. For example, the change in manufacturing cost of the product will show its effect on selling price of the product, sales, and profitability of the business firm. The study of the relationship in variables is very important for decision-making. 6.2 **Objectives** After going through the unit, you should be able to: ?

Define correlation between two variables; ? Explain how to Use Scatter diagrams to visualize the relationship between two variables; ? Outline the use of regression analysis to estimate the relationship between two variables; ? Explain Least squares estimating equation;

**Block II: Statistical Relations and Hypothesis Testing** 32 ?

Analyze

**71%**

**MATCHING BLOCK 25/52**

**W**

coefficient of determination as a measure of strength of the relationship between two variables; ?

Indicate limitations of regression and correlation analyzes and caveats about their use; and ? Explain the use of regression and correlation in finance. 6.3 Correlation There are many ways to measure the relationship between two or more variables. Two useful methods often used are scatter diagram and correlation. 6.3.1 The Scatter Diagram The scatter diagram indicates the nature of the potential relationship between the variables in two ways. First, it indicates whether the variables are related and if they are related then the kind of line, or estimating equation that describes this relationship. Possible relationship between two variables X and Y is shown in figure 6.1. Figure 6.1: Possible Relationships between Two Variables X and Y in Scatter Diagrams

Possible Relationships Between two variable X and Y in Scatter Diagrams Y X (a) Direct linear Y X (b) Inverse linear Y X (c) Direct curvilinear Y X (d) Inverse curvilinear Y X (e) Inverse linear with more scattering Y X (f) No relationship 6.3.2

Correlation Analysis As seen earlier, a scatter diagram shows the types of relationship using the graphical representation. The same result can be arrived at with a single number called correlation coefficient, which measures the degree of linear relationship among the variables. The direction of change is indicated by + or – signs; the former refers to the movement in the same direction and the latter in the opposite direction; an absence of correlation is indicated by zero. This coefficient ranges between –1 and 1. Further note that in figure 1 (a) the slope of line is positive and in 1 (e) the slope of line is negative. 6.3.3 Calculating Coefficient of Correlation Pearson's Coefficient of Correlation is the most widely used method for calculating coefficient of correlation and is expressed as:  $r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$  Where,  $\sum (X - \bar{X})^2$  and  $\sum (Y - \bar{Y})^2$  are the standard deviation of X and Y.

#### Unit 6: Correlation and Linear Regression 33

The covariance denoted by  $\text{Cov}(X, Y)$  of two random variables

**54% MATCHING BLOCK 26/52**

**W**

X and Y is defined by:  $\text{Cov}(X, Y) = \frac{1}{n} \sum (X - \bar{X})(Y - \bar{Y})$  or,  $r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$

**73% MATCHING BLOCK 27/52**

**W**

$\text{Cov}(X, Y) = \frac{1}{n} \sum (X - \bar{X})(Y - \bar{Y})$  or,  $r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$

Sometimes, coefficient of determination is also used to measure the strength of a relationship between two variables. It is explained by  $r^2$  i.e.,

**41% MATCHING BLOCK 28/52**

**W**

the square of the correlation coefficient. It explains the extent of the variation of a dependent variable and is expressed by the independent variable.

**32% MATCHING BLOCK 29/52**

**W**

of the variation of a dependent variable and is expressed by the independent variable. A high value of  $r^2$  shows a good linear relationship between the two variables. If  $r = 1$  and

$r^2 = 1$ , it indicates a perfect relationship between the variables. The detailed discussion about coefficient of determination is given in succeeding units. Example 1 The following data shows the sales and operating costs of a firm over a period of eight years: Year 2013 2014 2015 2016 2017 2018 2019 2020 Sales (Rs. in lakh) 106 140 144 155 136 150 146 143 Operating costs (Rs. in lakh) 100 114 120 125 105 124 122 118 What is the coefficient of correlation between sales and operating costs of the firm? Solution Let the following notations be used: X : Sales (Rs. in lakh); Y : Operating costs (Rs. in lakh) A

**55%**

**MATCHING BLOCK 30/52**

W

X 106 140 144 155 136 150 146 143 ? X = 1,120 B Y 100 114 120 125 105 124 122 118 ? Y = 928 C X - X -34 0 4 15 -4 10  
6 3 D Y - Y -16 -2 4 9 -11 8 6 2 E (X - X) 2 1156 0 16 225 16 100 36 9 ? (X - X) 2 = 1,558 F (Y - Y) 2 256 4 16 81 121  
64 36 4 ? (

**33%**

**MATCHING BLOCK 33/52**

W

X 106 140 144 155 136 150 146 143 ?X = 1,120 B Y 100 114 120 125 105 124 122 118 ?Y = 928 C X - X -34 0 4 15 -4 10  
6 3 D Y - Y -16 -2 4 9 -11 8 6 2 E (X - X) 2 1156 0 16 225 16 100 36 9 ?(X - X) 2 = 1,558 F (Y - Y) 2 256 4 16 81 121  
64 36 4 ?(Y - Y) 2 = 582 G C ? D 544 0 16 135 44 80 36 6 ?(X - X) (Y - Y) = 861 X = X n ? = 1,120 8 = 140 ; Y = Y n ? =  
928 8 = 116 r = 2 2 (X X) (Y Y) (X X) (Y Y) ? ? ? ? ? ? = 861 1,558 x 582 = 0.9042

**84%**

**MATCHING BLOCK 31/52**

W

$$\frac{(X - \bar{X})(Y - \bar{Y})}{1,558} = \frac{861}{1,558} \quad \bar{X} = \frac{\sum X}{n} = \frac{1,120}{8} = 140; \quad \bar{Y} = \frac{\sum Y}{n} = \frac{928}{8} = 116 \quad r = \frac{2 \sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{861}{1,558}$$

**84%**

**MATCHING BLOCK 32/52**

W

$$\frac{(X - \bar{X})(Y - \bar{Y})}{1,558} = \frac{861}{1,558} \quad \bar{X} = \frac{\sum X}{n} = \frac{1,120}{8} = 140; \quad \bar{Y} = \frac{\sum Y}{n} = \frac{928}{8} = 116 \quad r = \frac{2 \sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{861}{1,558}$$

**52%**

**MATCHING BLOCK 34/52**

W

$X(Y - Y) = 861$   $X = X_n ? = 1,120$   $8 = 140$  ;  $Y = Y_n ? = 928$   $8 = 116$   $r = 2$   $2(X X)(Y Y)(X X)(Y Y) ? ? ? ? ? ? = 861$   $1,558$   
 $\times 582 = 0.9042$

Block II: Statistical Relations and Hypothesis Testing 34

Interpretation from the value of  $r = 0.9042$  Positive value of  $r$  represents positive correlation between X and Y.

### 6.4 Rank Correlation

Rank correlation is also a technique used to test the direction and strength of the relationship between two variables. It is used when both the underlying variables are ordinal or when the data are available in the ordinal form irrespective of the type of variable. In other words, it's a tool to show whether any one set of numbers has an effect on another set of numbers. Spearman's rank correlation coefficient can be calculated using the following formula:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Where,  $N$  = Number of individuals ranked  $D_i$  = Difference in the ranks of the  $i$ th individual. Example 2 Ten competitors in a singing contest are ranked by three judges in the following order: 1st Judge 1 6 5 10 3 2 4 9 7 8 2nd Judge 3 5 8 4 7 10 2 1 6 9 3rd Judge 6 4 9 8 1 2 3 10 5 7 Use the rank correlation coefficient to determine which pair of judges has the nearest approach to common tastes in vocal or singing. Solution In order to find out which pair of judges has the nearest approach to common tastes in singing, we compare Rank Correlation between the judgments of: (i) 1st judge and 2nd judge; (ii) 2nd judge and 3rd judge; (iii) 1st judge and 3rd judge. Rank by 1st Judge  $R_1$  Rank by 2nd Judge  $R_2$  Rank by 3rd Judge  $R_3$   $(R_1 - R_2)$   $(R_2 - R_3)$   $(R_1 - R_3)$   $(R_1 - R_2)^2$   $(R_2 - R_3)^2$   $(R_1 - R_3)^2$  2 2 2 1 3 6 4 9 25 6 5 4 1 1 4 5 8 9 9 1 16 10 4 8 36 16 4

Unit 6: Correlation and Linear Regression 35 3 7 1 16 36 4 2 10 2 64 64 0 4 2 3 4 1 1 9 1 10 64 81 1 7 6 5 1 1 4 8 9 7 1 4 1  $N = 10$

N = 10 N = 10 200 D 2 ? ? 2 D 214 ? ? 60 D 2 ? ? Rank correlation between the judgment of 1st and 2nd judges:  $R = 1 - \frac{2 \times 36}{200}$

$$\Sigma D N N ? ? R (1 \text{ and } 2) = 1 - 3.6 \times 10^{-10} = 1 - 1.200990 = 1 - 1.212 = -0.212$$

Rank correlation between the judgments of 2nd and 3rd judges:  $r(2 \text{ and } 3) = 1 - \frac{3 \times 6 \times 214}{10 \times 10} = 1 - 1.297 = -0.297$   
 Rank correlation between judgments of 1st and 3rd judges:  $r(1 \text{ and } 3) = 1 - \frac{3 \times 6 \times 60}{10 \times 10} = 1 - 3.60 = -0.636$ . Since coefficient of correlation  $r(1 \text{ and } 3) = 0.636$  is maximum in the judgments of the first and third judges, we conclude that 1st judge and 3rd judge have the nearest approach to common tastes in singing. Check Your Progress - 1.1. A relationship model was developed between two variables X and Y, with X as the independent variable. After computations,  $\sum(X - \bar{X})(Y - \bar{Y}) = 682$ ,  $\sum(X - \bar{X})^2 = 484$ ,  $\sum(Y - \bar{Y})^2 = 1454$ . The coefficient of correlation between X and Y is a. 0.2717 b. 0.8129 c. 0.3731 d. 0.2563 e. 0.7863.

Block II: Statistical Relations and Hypothesis Testing 36 2.

Two judges ranked 10 people in a contestant as given below: Judge 1: 6 4 3 1 2 7 9 8 10 5 Judge 2: 4 1 6 7 5 8 10 9 3 2  
 Calculate the coefficient of correlation. a. 0.987 b. 0 c. 1 d. 0.5 e. 0.685 3. Which of the following technique shows the negative relationship between two variables in a diagram? a. Correlation b. Scatter Diagram c. Frequency diagram d. Rank correlation e. Regression 4. A statistician plotted the relationship between two variables from the collected data, which was shown as a horizontal line in the scatter diagram. Which of the following relationship exists between them? a. Direct linear b. Inverse linear c. Direct curvilinear d. Inverse curvilinear e. No relationship 5. If the correlation coefficient is above 0.9. The strength of relationship between the variables is \_\_\_\_\_ 6.5 Simple Linear Regression As a financial analyst one often tries to find out the relationship between the variables to draw conclusions about the future. For example, the financial analyst may be interested in finding how stock price will move with the movement in economic indication like, GDP, Inflation, export, etc. In the earlier section, we have measured only the strength of linear relationship. Using the simple linear regression, we can establish the linear relationship between two variables. Regression analysis is related with study of the dependent of one variable, the dependent variable, on one or more variables, the independent variable or Unit 6: Correlation and Linear Regression 37

explanatory variable, with a view to estimating and/or predicting the (population) mean or average value of the dependent variable in terms of the known or fixed (in repeated sampling) values of the independent variable. In the two variables X and Y, the variable we are seeking to explain (to find) is called Dependent Variable and the other variable is called Independent Variable. Consider, a total population of 100 families in a hypothetical community and their weekly income (X) and weekly expenditure (Y) both in dollars. The 100 families are divided into some income groups. The conditional expected value of the weekly expenditure with respect to weekly income i.e.  $E(Y|X)$  depends on the given values of the variable X.  $E(Y|X_i)$  is a function of  $X_i$  or  $E(Y|X_i) = f(X_i)$ . Now the question is how we define the function  $f(X_i)$ . Because in real situation we don't have a clear idea about the entire population. We assume that  $E(Y|X_i)$  is a linear function of  $X_i$  i.e.  $E(Y|X_i) = f(X_i) = A + B_1 X_i$ . The above equation represents the Population Regression Function (PRF), where the A and  $B_1$  are unknown and fixed parameters, also known as the intercept and slope respectively. In regression analysis our objective is to estimate the PRF i.e. estimating the value of unknown parameters (A and  $B_1$ ) on the basis of observation on Y and X. If we look into the above example, the next question that comes to our mind is what about the consumption expenditure of an individual family related to a fixed weekly income? The individual family weekly expenditure does not necessarily increase or decrease as the income level increase or decrease. Given an income level

$X_i$ ,

an individual family income expenditure is clustered around the conditional expected value i.e.  $E(Y|X_i)$ . Thus, the deviation of the individual  $Y_i$  from the  $E(Y|X_i)$  is represented as  $E_i = Y_i - E(Y|X_i)$  or,  $Y_i = E(Y|X_i) + E_i$  or,  $Y_i = A + B_1 X_i + E_i$ .  $E_i$  is defined as the error term and it's a random component. Our objective is to estimate the PRF on the basis of sample information. Sample Regression Function (SRF) is used to sample regression line. Where,  $\hat{E(Y|X_i)}$  = Estimator of  $E(Y|X_i)$ ,  $\hat{a}$  = Estimator of A, and  $\hat{b}_1$  = Estimator of  $B_1$ . The following regression equation explains that relation:  $Y = \hat{a} + \hat{b}_1 X + ?$  or

Block II: Statistical Relations and Hypothesis Testing 38

Slope  $Y = a + bX$  Independent Variable Dependent Variable  $Y$  Intercept Where,  $\epsilon$  are random variables with mean 0.  $b$  is called slope or regression coefficient of the equation. For example, the regression equation  $Y = 10 + 5X$ , slope is  $b = 5$ .  $a = 10$  is the  $Y$  intercept or constant term.

6.5.1 Assumptions in Regression If we consider the population Regression Function  $Y_i = A + B_1X_i + E_i$ . it shows that the value of  $Y_i$  depends on the value of  $X_i$  and  $E_i$ . If we don't specific about how  $X_i$  and  $E_i$  are generated, then we can't make any statistical inference about  $Y_i$ ,  $A$ , and  $B_1$ . The following assumptions in regression are very important to find out the statistical inferences about the regression parameters:

- The relationship between the variables  $X$  and  $Y$  is linear, which implies the formula  $E(Y|X = x) = a + bx$  at any given value of  $X = x$ .
- $X$  values are fixed in repeated sampling. Values taken by the regressor  $X$  are considered fixed in repeated samples.
- Zero mean value of error  $E_i$ . Given the value of  $X$ , the mean or expected value of the error term  $E_i$  is zero.  $E(E_i | X_i) = 0$ .
- Equal variance of error  $E_i$ . Given the value of  $X$ , the variance of the error term  $E_i$  is  $\sigma^2$ .  $Var(E_i | X_i) = \sigma^2$ .
- Given any two  $X$  values such as  $X_i$  and  $X_j$  ( $i$  and  $j$  are not same), the correlation between the corresponding error terms  $E_i$  and  $E_j$  is zero.
- Zero covariance between  $E_i$  and  $X_i$ .
- At each  $X$ , the distribution of  $Y$  is normal, and the variances  $2 \times \sigma^2$  are equal. This implies that  $X \epsilon'$  have the same variance  $2 \times \sigma^2$ .
- The  $Y$ -values are independent of each other.
- The independent variable  $X$  is not random.
- The expected value of error term is zero.
- No assumption is made regarding the distribution of  $X$ .

#### Unit 6: Correlation and Linear Regression 39

The above assumptions related to error terms  $E_i$  can be stated as

$E_i \sim$

$N(0, \sigma^2)$ . Also, under the normality assumptions, the error terms are not only correlated but also independently distributed i.e.  $E_i \sim NID(0, \sigma^2)$ .

Direct and Indirect Relation If a causal relationship is established, the relationship may be direct or inverse too. In case of direct relationship the slope of the line will be positive, whereas in the case of indirect relationship slope of the line will be negative.

6.5.2 Calculating Regression Constants The main objective of regression analysis is to find a line that provides the best possible description of the relationship of dependent and independent variables. Alternatively, we can say how the SRF should be constructed so that  $a$  is "as close as" possible to the true  $A$  and  $b$  is "as close as" possible to true  $B_1$ .

The method of Ordinary Least square (OLS) is one of the popular methods of finding unknown parameters. According to the Method of Least Squares, the regression should be established in a manner such that when we take sum of squares values of the deviation between predicted values and observed values. Differentiate the above equation with respect to  $a$  and  $b$ , we get the two normal equations (showing below).

$$\sum (Y - a - bX)^2$$

We can obtain the following formulae to calculate the coefficients ' $a$ ' and ' $b$ ' in the regression line:  $\hat{Y} = a + bX$

$$b = \frac{\sum X Y - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

$$a = \frac{\sum Y - b \sum X}{n}$$

Regression coefficient can also be found with the help of correlation coefficient. The relation between two is summarized below: Regression coefficient of  $Y$  on  $X$ :

$$b_{YX} = r \frac{\sigma_Y}{\sigma_X}$$

Regression coefficient of  $X$  on  $Y$ :  $b_{XY} = r \frac{\sigma_X}{\sigma_Y}$  Where, ' $r$ ' is the coefficient of correlation between  $X$  and  $Y$   $\sigma_X$  is the population standard deviation of  $X$   $\sigma_Y$

is the population standard deviation of  $Y$ . The coefficient of correlation can also be expressed as geometric mean of the two regression coefficients.

$$r = \sqrt{b_{YX} \times b_{XY}}$$

Block II: Statistical Relations and Hypothesis Testing 40 6.5.3.



Statistical Property of the Estimators a and b  
 1 a. Both a and b are the unbiased estimators of A and B. It means that  $E(a) = A$  and  $E(b) = B$ .  
 b. Both a and b can be represented as the linear combination of y.  
 c. The variance of a and b are as followed.  
 Example 3 To formulate the relationship between the quantity that can be sold and the price for one of its new milk products, the marketing manager of a company has collected the following information.  
 Price (Rs.) Quantity that can be sold ('000 units)  
 28.0 130 26.0 145 25.0 170 24.0 190 21.0 215 19.0 225  
 i. To formulate the relationship between the price and the quantity that can be sold with the help of the least square method the following table will be useful.  
 ii. Use this relationship to predict the quantity that can be sold if the price is Rs.22.  
 Total X 28 26 25 24 21 19 143  
 Y 130 145 170 190 215 225 1,075  
 XY 3,640 3,770 4,250 4,560 4,515 4,275 25,010  
 X<sup>2</sup> 784 676 625 576 441 361 3,463  
 In order to determine the regression line for the plot of main exam points vs model exam points, we need to calculate the coefficients a and b.

Unit 6: Correlation and Linear Regression  
 41  $\sum XY = 25,010$ ;  $\sum X^2 = 3,463$ ;  $\sum Y = 1,075$   $\sum X = 143$   $\sum Y = 1,075$   $\sum X^2 = 3,463$   $\sum Y^2 = 1,155,625$   $n = 6$  ? ? ?  
 $\sum Y = 1,075$   $\sum Y^2 = 1,155,625$   $n = 6$  ? ? ?

$b = \frac{25,010 - 6(23.83)(179.17)}{10.89 - 6(23.83)(23.83)} = 179.17 - (-10.89)(23.83) = 438.68$   
 The relationship between price (X) and quantity sold (Y) is:  $Y = 438.68 - 10.89X$   
 From the slope of this regression (-10.89) it can be stated that for every unit increase in price, it is estimated that the average demand will decrease by 10.89 points.  
 Suppose we want to use this relationship to predict the quantity that can be sold. If the price is Rs.22, it becomes  
 For price (X) = 22, quantity sold (Y) =  $438.68 - 10.89 \times 22 = 199.10$   
 Since the quantity is given in 1000s, actual quantity is 199,100 units.  
 6.6 Coefficient of Determination  
 Coefficient of determination ( $R^2$ ) is measured by the formula:  $R^2 = \frac{\sum (Y - \hat{Y})^2}{\sum Y^2 - n(\bar{Y})^2}$   
 The numerator is termed as total variation in Y and the denominator is called variation explained by the model. In other words, we can say that

**77%**

**MATCHING BLOCK 35/52**

**W**

the coefficient of determination is the ratio of the explained variation to the total variation. The coefficient of determination is

a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The coefficient of determination can be written in terms of correlation also.  $R^2 = r^2$ , the coefficient of correlation measures the strength of the linear association between x and y. Therefore  $0 \leq R^2 \leq 1$ , and greater the value of  $R^2$ , better it fits the data on line.  
 6.6.1 Regression Sum of Square and Error Sum of Square  
 Regression Sum of Square (RSS) is the portion of total sum square that is explained by the regression line; and ESS is the portion of sum square that is not explained by the use of the regression model.

Block II: Statistical Relations and Hypothesis Testing  
 42  $RSS$  (Regression Sum of Squares) =  $\sum (\hat{Y} - \bar{Y})^2$   
 $ESS$  (Error Sum of Squares) =  $\sum (Y - \hat{Y})^2$   
 We can show that  $TSS = RSS + ESS$   
 $TSS$  (Total Sum of Squares) =  $\sum (Y - \bar{Y})^2$   
 And coefficient of determination =

$TSS = RSS$  Coefficient of determination is also calculated by  $R^2 = \frac{RSS}{TSS}$  or  $R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$  6.7

The Standard Error  
 Standard error measures the reliability of the linear equation. It tells about the width of the error and hence the value of Y spread for given X. It's a measure of dispersion like standard deviation and measures the variability or scatter of the observed values around the regression line. The standard error of the estimate for a regression equation is given by

$s_e = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}}$  Where, Y = Values of the dependent variable  $\hat{Y}$  =

Estimated values from the estimating equation that corresponds to each Y value  
 n = Number of data points used to fit the regression line. In the above equation, you can observe that the sum of the squared deviations is divided by n - 2 and not by n. This is because we have lost 2 degrees of freedom in estimating the regression line. We used the sample to compute a and b.  
 6.7.1 Shortcut Method to Compute Standard Error ( $s_e$ )  
 Equation for calculating standard error of estimate is also given by the formula:

$s_e = \sqrt{\frac{\sum Y^2 - n(\bar{Y})^2 - \frac{(\sum XY)^2}{n}}{n - 2}}$

Where, X = Values of the independent variable Y = Values of the dependent variable a = Y-intercept of the estimating equation



Unit 6: Correlation and Linear Regression 43  $b$  = Slope of the estimating equation  $n$  = Number of data points. 6.8 Making Inference about Population Parameter The sample regression line can be used as a true estimator of population regression line. Let the population regression line be  $Y = A + BX$  and sample regression equation be  $Y = a + bx$ . Now we can make inferences about the slope  $B$  of the true regression line, with the given sample regression line slope  $b$ . 6.8.1 Prediction Interval Standard error can be used to predict interval around

$\hat{Y}$

which would contain the actual  $Y$ . For a large sample where  $n \geq 30$ ,

$\hat{Y} \pm Z s_e$

would be the interval, where  $Z$  is the appropriate Standard Normal Value. For a small sample where  $n < 30$ ,

$\hat{Y} \pm t s_e$

would be the interval, where  $t$  is the appropriate  $t$  value (with  $n - 2$  degrees of freedom). 6.8.2 Accepting the Model We know  $b$  is an estimator of  $B$ . The random variable  $b$  is a normal distribution with mean  $B$ . Now, we can calculate the test statistics  $t$  with  $n - 2$  degrees of freedom using the formula given below: 0

$H_0: b = B$   $t = \frac{b - B}{s_b} \dots$  (1) Where, standard error of regression coefficient ( $s_b$ ) can be calculated using the following formula:  $s_b = \frac{s_e}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}}$

$s_e = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n - 2}}$  ... (2) Remember that  $s_e = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n - 2}}$

Finally, we can find the value of  $t$  for the given significant level and for  $n - 2$  degree of freedom and compare with the calculated  $t$  statistics. In equation (2), if  $t$  statistics falls within the acceptance region, we can accept the null hypothesis; and if  $t$  statistics fall within the rejection region, we reject the null hypothesis.

Block II: Statistical Relations and Hypothesis Testing 44

Example 4 Consider our previous example of sample linear equation  $Y = 60 + 5X$ . Suppose we want to check whether the computed model is spurious or not at 5% significant level, we can set up the hypothesis as given here.  $H_0: B = 0$  (There is no linear relationship);  $H_1: B > 0$  (There is a positive relationship)

$s_e = 13.83$ ;  $s_b = \frac{13.83}{\sqrt{2528 - 10 \times 14 \times 14}} = \frac{13.83}{83.23} = 0.166$  Therefore,  $t = \frac{5 - 0}{0.166} = 30.12$

At  $\alpha = 5\%$ , the rejection region is  $t > 2.132$  But the calculated  $t$  value = 30.12; As 30.12  $> 2.132$  Therefore we reject  $H_0$ ; that means there is a linear relationship. 6.8.3. F Test An F-test can also be used to test for significance in regression. In

linear regression, we have only one independent variable. Thus, the F-test will provide the same result as the t-test. The reason behind the use of the F test for determining whether the regression relationship is statistically significant and is based on the development of two independent estimates of  $\sigma^2$ . The next question is to perform the F-test, for which we need to find out the F statistic. To find the F statistic, we require the concept of mean square. The mean square can be evaluated dividing the sum of square by the corresponding degree of freedom. In simple linear regression, the variation occurred by the regression line has degree of freedom equals to the number of independent variables i.e. 1, whereas the degree of freedom of the variation due to error is equal to  $(n-2)$ . If the null hypothesis is true, MSR and MSE are two independent estimates of  $\sigma^2$  and the sampling distribution of MSR/MSE follows an F distribution with numerator df equal to one and denominator df equal to  $n-2$ . The summary of F-test is as follows: Test statistics:  $F = \frac{MSR}{MSE}$  Rejection Rule:

Reject  $H_0$  if  $p \text{ value} \leq \alpha$  Reject  $H_0$  if

$F_{cal} \geq F_{\alpha}$

is based on F-distribution with 1 degree of freedom in the numerator and  $(n-2)$  degrees of freedom in the denominator.

Unit 6: Correlation and Linear Regression 45 6.9 Geometry of Regression In this section, we will interpret the shape of regression line for different values of  $a$  and  $b$ . As discussed earlier,

**83% MATCHING BLOCK 36/52**

**W**

$a$  is the intercept and  $b$  is the slope of the line.

**91% MATCHING BLOCK 37/52**

**W**

$a$  is the intercept and  $b$  is the slope of the line.

63%

**MATCHING BLOCK 38/52**

W

a is the intercept and b is the slope of the line. Figure 6.2: Linear Equations 6.9.1 Slope of Regression Line

Slope is basically the inclination of a line and

71%

**MATCHING BLOCK 39/52**

W

it represents the change in the dependent variable for every unit change in the

independent variable. Suppose there are two points on the straight line  $(X_1, Y_1)$  and  $(X_2, Y_2)$ , then slope of the line will be  $b = \frac{Y_2 - Y_1}{X_2 - X_1}$

$X_1, Y_1, X_2, Y_2$  ?

We have seen that the slope of a line may be positive, negative, zero or undefined. The sign of the slope indicates whether it is rising or falling. The magnitude (absolute value) of the slope indicates the relative steepness of the line. 6.10 Misuses and Caveats Misuses ? Estimating line can be applied beyond the range of data (extrapolation). But estimating equation is valid only over the same range as the one from which the sample was taken initially. ? Usually regression and correlation are used to determine a cause and effect relationship, which is another mistake, because regression and correlation do not determine the cause and effect. In other words, if there is correlation between X and Y it does not mean one causes the other. ? Sometimes the regression line fails to recognize some variables that can be dependent on time. ? Misrepresenting  $r^2$  and r values is another mistake. For example,  $r = 0.4$  does not mean that 40% variation is explained by the line. But it means that only 16% of variation is explained by the line, because the value of  $r^2 = 0.16$ . ? Regression analysis mistakenly calculates relationships when they do not exist.

Block II: Statistical Relations and Hypothesis Testing 46

Caveats ? Before using the regression and correlation know the limitations of the technique. ? Use common intelligence while applying the regression and correlation. Check Your Progress - 2 6. Two variables X and Y have the regression relationship

$$\hat{Y} = 5 + 4X$$

X. Then which of the following is correct? a. For a given value of  $Y = 22$ , the estimated value of X is 4 b. For a given value of  $Y = 25$ , the estimated value of X is 5 c. The slope of the line is 5 d. The intercept of the above line is 4 e. For a given value of  $Y = 20$ , the estimated value of X is 2 7. If the variances of X and Y are 2.3 and 4.1, and the covariance between them is 2.1 then the coefficient of correlation between X and Y is a. 0.098 b. 0.222 c. 0.408 d. 0.684 e. 0.748. 8. The slope and the intercept on Y-axis for a simple regression line are 3 and 5 respectively. If the independent variable has a value of 5 then the estimated value of the dependent variable is a. 3 b. 5 c. 20 d. 27 e. 28. 6.11 Applications in Finance Correlation and regression have various applications in finance as discussed below. 6.11.1 Risk of a Portfolio Correlation is used to calculate risk of portfolio. Suppose an investor has a portfolio consisting of two stocks A and B whose return and risk are as follows:

Unit 6: Correlation and Linear Regression 47

Value (Rs. Thousand) Return Standard Deviation Security A 30 30% 1.5 Security B 70 20% 2.5 Portfolio Value 100 The return on the portfolio of the above securities will be the weighted average return.  $R_p = W_1 R_1 + W_2 R_2$  i.e., return of the portfolio will be  $0.3 \times 30 + 0.7 \times 20 = 22\%$  But the risk of portfolio will not necessarily be the same as weighted average risk of securities, because the risk of the portfolio is dependant on the coefficient of correlation between them. For  $-1$  correlation coefficient, the two returns move exactly opposite to each other and the portfolio risk will be minimum and for  $+1$  correlation coefficient, return of two securities will move in the same direction, so the risk of portfolio would be maximum. For a 2 asset portfolio, the standard deviation  $\sigma_p$  is given by

$$\sigma_p = \sqrt{W_1^2 \sigma_1^2 + W_2^2 \sigma_2^2 + 2W_1 W_2 \rho \sigma_1 \sigma_2}$$

Where,  $W_1$  and  $W_2$  are the weights for the assets in the portfolio.  $\rho_{1,2}$  is the coefficient of correlation between the returns of the two assets.  $\sigma_1$  and  $\sigma_2$  are the standard deviations of the returns of the two assets respectively. Note:  $\rho_{1,2} \sigma_1 \sigma_2 = \text{Covariance between returns of the two assets}$ . 6.11.2 Characteristic Line Financial analysts often talk about beta of share which measures the sensitivity of the stock price with the market index. Bombay Stock Exchange (BSE), Sensitive Index (Sensex), BSE National Index, NSE (nifty) are some popular market indices. As they represent the movement of the whole market, they are called as market index. Beta equals to 1.2 means if the market index moves by 10%, the securities of beta 1.2 will move by  $10 \times 1.2$  or 12%. Beta of securities more than one represents aggressive securities and beta of less than one represents defensive securities. Beta is used to measure the Securities.

#### Block II: Statistical Relations and Hypothesis Testing 48

The total risk of a security is measured in terms of the variance ( $\sigma^2$ ) [or standard deviation ( $\sigma$ )] of its returns. Total risk comprises both systematic and unsystematic components. How do we split the total risk, namely, the variance, into the systematic and unsystematic risk components? This is relatively simple. Total Risk of a Security  $i = \sigma_i^2$ ; Systematic Risk of a Security  $i = \sigma_m^2 \rho_{i,m}^2$  Where,  $\rho_{i,m}$  is the beta of security  $i$  and  $\sigma_m^2$

is the variance of the market portfolio. But,  $\rho_{i,m} = \sigma_{i,m} / \sigma_i \sigma_m$ ; Substituting for  $\rho_{i,m}$  in above equation we have Systematic Risk of Security  $i = \sigma_m^2 \rho_{i,m}^2$

$\sigma_{i,m} = \rho_{i,m} \sigma_i \sigma_m$ ; since  $\rho_{i,m} = \sigma_{i,m} / \sigma_i \sigma_m$ ,  $\sigma_{i,m} = \rho_{i,m} \sigma_i \sigma_m$

where  $\rho_{i,m}$  is the correlation coefficient between the security return and the market return. Therefore, Unsystematic Risk of a Security

$$\sigma_i^2 = \sigma_m^2 \rho_{i,m}^2 + \sigma_{i,m}^2$$

In terms of percentage of total risk (security variance): Systematic Risk =  $(\sigma_m^2 \rho_{i,m}^2 / \sigma_i^2) \times 100 = 100 \rho_{i,m}^2$

Unsystematic Risk =  $(\sigma_{i,m}^2 / \sigma_i^2) \times 100 = 100 (1 - \rho_{i,m}^2)$ . 6.11.3

Cost-Volume-Profit Analysis The Cost-Volume-Profit (CVP) analysis is used widely in management and cost account to determine the relationship governing sales, profit and cost. One of the basic assumptions in CVP analysis is that total cost can be segregated into fixed, variable cost and semi-variable cost. Simple linear regression is commonly used to segregate the fixed and variable components of semi-fixed or semi-variable costs. 6.11.4 Time Series and Demand Forecasting Forecasting of future demand and sales is an important aspect of budgeting. Many techniques are used to forecast the future demand and sales. In the trend projection method, the time series of demand for a product over the past years is studied and the trend observed is projected into the future. A linear relationship (if it exists) is normally established using historical data. 6.11.5 Tests in Investments Investors and analysts question if the price of stock is dependent on historical price, because if the market is not efficiently working then the investor having

#### Unit 6: Correlation and Linear Regression 49

more information can earn higher profit than the investor who has less information. Efficient Market Hypothesis holds that security prices fully reflect all available information any time. Statistical tests like auto correlation test and run tests have been applied to examine whether the market is efficient or not, and if it is efficient, then what is the level of information reflected in the market price of share. Based on the level of information market can be in weak efficiency, semi-strong efficiency or strong efficiency form. Weak form of efficiency says that current prices fully reflect all historical information. According to the semi-strong form, current stock prices reflect all publicly available information such as earnings, stock and cash dividends, splits, mergers and takeovers, interest rate changes, etc. According to the strong form, prices of securities fully reflect all available information – both public and private run tests and serial correlation test are used to test the weak form of efficiency. 6.11.6 Serial Correlation Tests Serial correlation has been used to find the correlation coefficient between current changes and past changes in a share's price. The test selects a certain number of stocks. The changes in the prices of these stocks are observed for a particular period. The same is done for another period. Correlation analysis is conducted on these changes. If the correlation between these changes is near or equal to zero, it implies that the price changes are independent of each other. 6.11.7 Run Tests A run is defined as "

**100% MATCHING BLOCK 40/52**

**W**

a sequence of identical occurrences preceded and followed by different occurrences or by none at all."

For testing the randomness of share prices, we take a series of stock prices. Starting with the first price, each price change is denoted by a plus (+) or a minus (–) sign. Plus (+) sign indicates that the price under consideration has increased compared to its preceding price and a minus (–) sign indicates that the price under consideration has decreased compared to the preceding price. In case, the price under consideration is same as its preceding price, we indicate it by a zero. In case, the sign has changed from plus to minus or from minus to plus, a new run is counted to have begun. To test the independence of the prices, we require: Total number of runs :  $r$  Number of positive price changes :  $n_1$  Number of negative price changes :  $n_2$  Once we have this data, the mean and the standard deviation of the mean, are calculated by using the formulae given below:  $\sigma_r = \sqrt{\frac{2n_1n_2}{n}}$

**75% MATCHING BLOCK 41/52**

**W**

$n_1 n_2 n_3 n_4 n_5$  and  $r = 1 2 1 2 1 2 2 1 2 2 n_1 n_2 (2n_1 n_2 n_1) (n_1 n_2) ($

**81% MATCHING BLOCK 44/52**

**W**

$n_1 n_2 n_3 n_4 n_5$  and  $r = 1 2 1 2 1 2 2 1 2 2 n_1 n_2 (2n_1 n_2 n_1) (n_1 n_2) (n_1 n_2) ? ? ? ? ?$

**81% MATCHING BLOCK 45/52**

**W**

$n_1 n_2 n_3 n_4 n_5$  and  $r = 1 2 1 2 1 2 2 1 2 2 n_1 n_2 (2n_1 n_2 n_1) (n_1 n_2) (n_1 n_2) ? ? ? ? ?$

**100% MATCHING BLOCK 42/52**

**W**

$r = 1 2 1 2 1 2 2 1 2 2 n_1 n_2 (2n_1 n_2 n_1) (n_1 n_2) (n_1 n_2) ? ? ? ? ?$

**100% MATCHING BLOCK 43/52**

**W**

$n_1 n_2 (2n_1 n_2 n_1) (n_1 n_2) (n_1 n_2) ? ? ? ? ?$

#### Block II: Statistical Relations and Hypothesis Testing 50

At a given level of significance, we calculate the upper and lower limits and check whether the number of runs observed from the test falls within the limits or not. If it is between the limits, we conclude that the prices are random or independent of each other. 6.12 Regression using Microsoft-Excel A company is in the process of deciding whether to purchase a maintenance contract for its new computer wheel alignment and balancing machine. Managers feel that maintenance expense should be related to usage, and they collected the following information on weekly usage (hours) and annual maintenance expense (in hundreds of dollars). Weekly Usage (Hours) Annual Maintenance Expense 13 17 10 22 20 30 28 37 32 47 17 30 24 32.5 31 39 40 51.5 38 40 How to perform the regression analysis of the above example in EXCEL is explained below. Step1: Go to Data tab in Excel. Step 2: Under the Data tab, go to Data Analysis toolbox. Step 3: Select the Regression from the Data Analysis Toolbox. Step 4: After that a new window will open.

#### Unit 6: Correlation and Linear Regression 51

Step 5: Enter A1:A11 in the input X range Step 6: Enter B1: B11 in the output Y range Step 7: Select the labels Step 8: Select the confidence level and make it 95% Step 9: Select the output range and choose any cell reference in the box. In this case, we choose F3 cell Step 10: Click OK From the Excel Output, we can observe that the regression equation is  $= 10.3669 + 0.9578 x$ . The value of coefficient of determination is 0.8603 i.e. 86.03 percent of the variability in the dependent variable can be explained by the linear relationship between annual maintenance and weekly usage. In the Anova table, the p value is 0.00011 which is less than  $\alpha$ . So, we can reject the null hypothesis and conclude that the relationship between annual maintenance and weekly usage is significant. Check Your Progress - 3 9. The coefficient of determination is \_\_\_\_\_ of the coefficient of correlation 10. The shares advisory consultant was analyzing the volatility or risk of a portfolio (Y) compared to the market as a whole (X). Which will help him measure this value better? a. Rank correlation b. Coefficient of correlation c. Beta d. Standard deviation e. Standard error

Block II: Statistical Relations and Hypothesis Testing 52 6.13

Summary ? Correlation and regression are the statistical measures used to describe the nature and strength of the relationship between two or more variables. While correlation analysis determines the degree to which the variables are related, regression analysis develops the relationship between the variables. ? The coefficient of determination  $r^2$  is used to analyze how good the fit is and gives the variation explained by the model. Regression and correlation methods have wide application in finance. They are used to measure risk in portfolio, establishing characteristic lines and measuring cost volume relationship. 6.14 Glossary Coefficient of Correlation: It indicates the direction of the relationship between two variables, direct or inverse. It is also measured as the square root

**68%**

**MATCHING BLOCK 51/52**

**W**

of the coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the regression line,

**58%**

**MATCHING BLOCK 48/52**

**W**

coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the

**52%**

**MATCHING BLOCK 49/52**

**W**

coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the

**57%**

**MATCHING BLOCK 46/52**

**W**

Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by

**57%**

**MATCHING BLOCK 47/52**

**W**

Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by

**62%**

**MATCHING BLOCK 50/52**

**W**

Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the

that is, by Y's relationship with the independent variable. Correlation Analysis: A technique used to determine the degree of linear relationship between two variables. Dependent Variable: The variable be predicted in the regression analysis. Direct Relationship: If

**73%**

**MATCHING BLOCK 52/52**

**W**

the value of the independent variable increases, so does the value of the dependent variable, and this

relationship is known as direct relationship. Independent Variables: The known variable, or variables, in regression analysis. Inverse Relationship: A linear relationship between two variables in which, as the independent variable increases, the dependent variable decreases. Regression Line: A line fitted to a set of data points to estimate the relationship between two variables. Regression: The general process of predicting one variable from another by statistical means, using previous data. Scatter Diagram: A graph of points on a rectangular grid; the X and Y co- ordinates of each point correspond to the two measurements made on some particular sample element, and the pattern of points illustrates the relationship between the two variables. Standard Error: The standard deviation of the sampling distribution of a statistic.

Unit 6: Correlation and Linear Regression 53 6.15

Suggested Readings/Reference Material 1. Gupta, S. P. Statistical Methods. 46th Revised ed. New Delhi: Sultan Chand & Sons. 2021 2. I. Levin Richard, H. Siddiqui Masood, S. Rubin David, Rastogi Sanjay. Statistics for Management. Pearson Education; Eighth edition, 2017 3. Gerald Keller. Statistics for Management and Economics. Cengage, 2017. 4. Arora, P. N., and Arora, S. CA Foundation Course Statistics. 6th ed. S Chand Publishing, 2018. 5. Mario F Triola. Elementary Statistics. 13th ed., 2018. 6. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. Statistics for Business and Economics. 13th Edition, Cengage Learning India Pvt. Ltd., 2019. 7. S D Sharma. Operations Research. Kedar Nath Ram Nath, 2018. 8. Hamdy A. Taha. Operations Research: An Introduction. 10th ed., Pearson, 2016. 9. Malhotra, N. (2012), Marketing Research: An Applied Orientation, 7th ed., Pearson, 2019. 10. Cooper, D.R. and Schindler, P.S. and J. K. Sharma (2018), Business Research Methods, 12th edition, McGraw-Hill Education. 6.16

Self-Assessment Questions 1. Correlation and Regression are not same. Explain. 2. What are the merits and demerits of correlation? 3. Define direct and inverse relationships. 4. Explain the reason for construction scatter diagram. 5. What do you mean by curvilinear relationship? Explain the difference between liner and curvilinear relationship. 6.17 Answers to Check Your Progress Questions 1. (b) 8129 ( $682/\sqrt{484 \times 454}$ ) 2. (a) 987 3. (b) Scatter Diagram 4. (e) No relationship- A horizontal line in the scatter diagram indicates no relationship between the variables since dependent variable becomes constant for any value of independent variable. 5. Strong/very strong as the value is very nearer to ONE

Block II: Statistical Relations and Hypothesis Testing 54 6. (

b) For a given value of Y = 25, the estimated value of X is 5 7. (b) 0.222 8. (c) 20 9. The coefficient of determination is square of the coefficient of correlation. 10. (c) Beta It measures the volatility or risk of a portfolio compared to the market value. While the securities with beta value below 1 show less volatility but provide lower, but more stable returns.

Quantitative Methods Course Structure Block I: Introduction to Statistics and Probability 1. Arranging Data 2. Central Tendency and Dispersion 3. Probability 4. Probability Distribution and Decision Theory Block II: Statistical Relations and Hypothesis Testing 5 Statistical Inference and Hypothesis Testing 6 Correlation and Linear Regression Block III: Statistical Regression and Quality Control 7 Multiple Regression 8 Time Series Analysis 9 Quality Control Block V: Statistical Distributions, Variations and IT 10 Chi-Square Test and Analysis of Variance 11 Role of IT in Modern Business Enterprise 12 Statistical Software Tools Block V: Advanced Statistics 13 Index Numbers 14 Simulation 15 Linear Programming Block VI: Business Research 16 Introduction to Business Research Methods 17 Questionnaire Design 18 Report Writing

## Hit and source - focused comparison, Side by Side

<b>Submitted text</b>	As student entered the text in the submitted document.
<b>Matching text</b>	As the text appears in the source.

<b>1/52</b>	<b>SUBMITTED TEXT</b>	2 WORDS	<b>100% MATCHING TEXT</b>	2 WORDS
	<p><a href="https://www.thehindu.com/opinion/op-ed/local-factors-determine-electoral-outcomes-in-states/article34475075.ece">https://www.thehindu.com/opinion/op-ed/local-factors-determine-electoral-outcomes-in-states/article34475075.ece</a></p> <p><b>W</b> <a href="https://www.lokniti.org/POST_POLL_ANALYSIS_2021">https://www.lokniti.org/POST_POLL_ANALYSIS_2021</a></p>		<p><a href="https://www.thehindu.com/opinion/op-ed/local-factors-determine-electoral-outcomes-in-states/article34475075.ece">https://www.thehindu.com/opinion/op-ed/local-factors-determine-electoral-outcomes-in-states/article34475075.ece</a></p>	
<b>2/52</b>	<b>SUBMITTED TEXT</b>	22 WORDS	<b>52% MATCHING TEXT</b>	22 WORDS
	<p>is the difference between the value of sample statistics and the value of the corresponding population parameter. For example, in case</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p>is the difference between the expected value of the estimator and the value of the parameter. In this case,</p>	
<b>3/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>75% MATCHING TEXT</b>	19 WORDS
	<p><math>N/n</math> ? ? ? , Where, N is the size of population and n is the size of sample.</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p><math>N/n</math>), where N is the total number of units in the population and n is the size of the sample.</p>	
<b>4/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>70% MATCHING TEXT</b>	16 WORDS
	<p>of the sampling distribution depends on the shape of the population from which the sample</p> <p><b>W</b> <a href="https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt...">https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt ...</a></p>		<p>of the sampling distribution of is the mean of the population from which the sample</p>	
<b>5/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>70% MATCHING TEXT</b>	16 WORDS
	<p>of the sampling distribution depends on the shape of the population from which the sample</p> <p><b>W</b> <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a></p>		<p>of the sampling distribution of is the mean of the population from which the sample</p>	
<b>6/52</b>	<b>SUBMITTED TEXT</b>	18 WORDS	<b>73% MATCHING TEXT</b>	18 WORDS
	<p>According to the central limit theorem, sampling distribution of the sample mean <math>\bar{X}</math> is approximately a normal</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>According to the central limit theorem, for a large sample size, sampling distribution of the sample mean <math>\bar{x}</math> is approximately normal,</p>	



<b>7/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>79% MATCHING TEXT</b>	19 WORDS
	<p>sampling distribution of the sample mean <math>\bar{X}</math> is approximately a normal distribution with mean <math>\mu</math> and standard deviation</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p>sampling distribution of the sample mean <math>\bar{X}</math> tends to a normal distribution with mean <math>\mu</math> and standard deviation</p>	
<b>8/52</b>	<b>SUBMITTED TEXT</b>	22 WORDS	<b>54% MATCHING TEXT</b>	22 WORDS
	<p>the sampling distribution of the sample proportion of defectives <math>p</math> is approximately a normal distribution with mean <math>p</math> and standard deviation</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p>the sampling distribution of the sample proportion <math>p</math> approaches to a normal distribution with proportion <math>p</math> and standard deviation</p>	
<b>9/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>66% MATCHING TEXT</b>	16 WORDS
	<p>the value of the population parameter is that the sample statistic be an unbiased estimator.</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p>the value of the corresponding population parameter, the sample statistic is said to be an unbiased estimator.</p>	
<b>10/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>97% MATCHING TEXT</b>	19 WORDS
	<p>consistent if the difference between the estimator and parameter grows smaller as the sample size grows larger.</p> <p><b>W</b> <a href="https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt...">https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt ...</a></p>		<p>consistent if the difference between the estimator and the parameter grows smaller as the sample size grows larger.</p>	
<b>11/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>97% MATCHING TEXT</b>	19 WORDS
	<p>consistent if the difference between the estimator and parameter grows smaller as the sample size grows larger.</p> <p><b>W</b> <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a></p>		<p>consistent if the difference between the estimator and the parameter grows smaller as the sample size grows larger.</p>	

<b>12/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>63% MATCHING TEXT</b>	19 WORDS
<p>If there are two estimators then the estimator whose variance is less is said to be relatively efficient.</p> <p><b>W</b> <a href="https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt...">https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt ...</a></p>		<p>If there are two unbiased estimators of a parameter, the one whose variance is smaller is said to be relatively more efficient. 10.2</p>		
<b>13/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>63% MATCHING TEXT</b>	19 WORDS
<p>If there are two estimators then the estimator whose variance is less is said to be relatively efficient.</p> <p><b>W</b> <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a></p>		<p>If there are two unbiased estimators of a parameter, the one whose variance is smaller is said to be relatively more efficient.</p>		
<b>14/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>70% MATCHING TEXT</b>	16 WORDS
<p>of the population variance, and the sample standard deviation is a point estimator of the</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>of the population variance and the sample standard deviation is the square root of the</p>		
<b>15/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>66% MATCHING TEXT</b>	16 WORDS
<p>the difference between the observed sample statistic and the actual value of the population parameter</p> <p><b>W</b> <a href="https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt...">https://www.worldsupporter.org/nl/chapter/41414-samenvatting-managerial-statistics-keller-gedeelt ...</a></p>		<p>the difference between the sample statistic Figuur 1 and the hypothesized value of the parameter</p>		
<b>16/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>66% MATCHING TEXT</b>	16 WORDS
<p>the difference between the observed sample statistic and the actual value of the population parameter</p> <p><b>W</b> <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a></p>		<p>the difference between the sample statistic Figuur 1 and the hypothesized value of the parameter</p>		
<b>17/52</b>	<b>SUBMITTED TEXT</b>	24 WORDS	<b>87% MATCHING TEXT</b>	24 WORDS
<p><math>H_0 : \mu \neq 0</math> and the Alternative Hypothesis may be: i. <math>H_1 : \mu = 0</math></p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p><math>H_0 : \mu = 0</math> and the alternative hypothesis will be <math>H_1 : \mu \neq 0</math></p>		

<b>18/52</b>	<b>SUBMITTED TEXT</b>	21 WORDS	<b>87% MATCHING TEXT</b>	21 WORDS
	<p><math>\bar{X} \pm 1.25 \hat{\sigma} / \sqrt{n}</math> Since the population standard deviation is not known and the sample size is</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p><math>\bar{x} \pm z_{\alpha/2} s / \sqrt{n}</math> When the population standard deviation is not known and the sample size is</p>	
<b>19/52</b>	<b>SUBMITTED TEXT</b>	74 WORDS	<b>30% MATCHING TEXT</b>	74 WORDS
	<p><math>P(H) : P(p) ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Left tail) Or <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math></p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p><math>P(H/A) = 285 P(A/H) \cdot P(H) P(A/H) \cdot P(H) + P(A/M) \cdot P(M) + P(A/L) \cdot P(L) (0.70)(0.30) = (0.70)(0.30) + (0.40)(0.50) + (0.20)(0.20) = 0.467</math> <math>P(H/A) = (b) P(M/A) P(A/M) \cdot P(M) P(A/H) \cdot P(H) + P(A/M) \cdot P(M) + P(A/L) \cdot P(L) (0.40)(0.50) = (0.70)(0.30) + (0.40)(0.50) + (0.20)(0.20) = 0.444</math> <math>P(M/A) = (c) P(L/A) P(A/L) \cdot P(L) P(A/H) \cdot P(P)</math></p>	
<b>20/52</b>	<b>SUBMITTED TEXT</b>	81 WORDS	<b>28% MATCHING TEXT</b>	81 WORDS
	<p><math>a(H) : P(p) : P(p) ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Left tail) Or <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math></p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p><math>a) P(H/P(A/H)) \cdot P(H) P(A/H) \cdot P(H) + P(A/M) \cdot P(M) + P(A/L) \cdot P(L) (0.70)(0.30) = (0.70)(0.30) + (0.40)(0.50) + (0.20)(0.20) = 0.467</math> <math>P(H/A) = (b) P(M/A) P(A/M) \cdot P(M) P(A/H) \cdot P(H) + P(A/M) \cdot P(M) + P(A/L) \cdot P(L) (0.40)(0.50) = (0.70)(0.30) + (0.40)(0.50) + (0.20)(0.20) = 0.444</math> <math>P(M/A) = (c) P(L/A) P(A/L) \cdot P(L) P(A/H) \cdot P(P)</math></p>	
<b>21/52</b>	<b>SUBMITTED TEXT</b>	80 WORDS	<b>19% MATCHING TEXT</b>	80 WORDS
	<p><math>P(p) : P(p) ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ?</math> (Left tail) Or <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Two tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math> (Right tail) <math>1/2</math> <math>1/2</math> <math>0</math> <math>a(H) : P(p) : P(p) ? ? ? ?</math></p> <p><b>W</b> <a href="https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf">https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf</a></p>		<p><math>P(6) P(6) P(6) - -</math> Business Support <math>P(5) P(7) P(7) - -</math> Eating &amp; Drinking Establishments <math>P(4)(16) P(16) P(16) P(5,16) P(5,16) P(5,16) P(16) -</math> Financial Institutions <math>P(5,16) P(5,16) P(5,16) P(16) -</math> Funeral &amp; Mortuary Services - <math>P(5,16) P(5,16) P(5,16) P(16) -</math> Instructional Studios <math>P(5,16) P(5,16) P(5,16) P(16) -</math></p>	

<b>22/52</b>	<b>SUBMITTED TEXT</b>	70 WORDS	<b>61% MATCHING TEXT</b>	70 WORDS
	<p>the population proportion, <math>p = \frac{1}{2}</math> (1) (1) 1.96 1.96 p p  <math>(200 \times 0.45) (300 \times 0.30) 180 0.36 n n 200 300 500 ? ? ? ?</math>  <math>? ? 1 2 p p ? ?</math> (estimated) = <math>1 2 p(1 p) p(1 p) n n ? ? ? ? ? ?</math>  <math>? ? ? ? ? ? ? = 0.36(1 0.36) 0.36(1 0.36) 200 300 ? ? ? ? ?</math>  <math>? ? ? ? ? ? ? ? ?</math></p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>the Population Proportion P 69 95% (1) (1) 1.96 1.96 p p  <math>p p p P p n n - - - \leq \leq + 99\% (1) (1) 2.58 2.58 p p p p p P</math>  <math>p n n - - - \leq \leq +</math></p>	
<b>23/52</b>	<b>SUBMITTED TEXT</b>	27 WORDS	<b>46% MATCHING TEXT</b>	27 WORDS
	<p>the sampling distribution of the mean approaches normality as the sample size increases, regardless of the shape of the population distribution from which the sample is</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>the sampling distribution of the mean can be approximated by the normal distribution. This is regardless of the distribution of the population from which the random sample is</p>	
<b>24/52</b>	<b>SUBMITTED TEXT</b>	13 WORDS	<b>87% MATCHING TEXT</b>	13 WORDS
	<p>an estimator. Estimator: A sample statistic used to estimate a population parameter.</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>An estimator is a sample statistic used to estimate a population parameter.</p>	
<b>25/52</b>	<b>SUBMITTED TEXT</b>	16 WORDS	<b>71% MATCHING TEXT</b>	16 WORDS
	<p>coefficient of determination as a measure of strength of the relationship between two variables; ?</p> <p><b>W</b> <a href="https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf">https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf</a></p>		<p>Coefficient of Correlation (r) is a measure of the strength of the relationship between two variables.</p>	
<b>26/52</b>	<b>SUBMITTED TEXT</b>	26 WORDS	<b>54% MATCHING TEXT</b>	26 WORDS
	<p>X and Y is defined by: <math>(X X)(Y Y) n 1 ? ? ? ?</math> or, <math>r = ? ? ? ? ? ?</math>  <math>? 2 2 ) Y (Y ) X (X ) Y (Y) X (X</math></p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>X and Y. That is <math>X = 7</math> and <math>Y = 10</math>. Likewise, <math>( ) ( ) X X Y Y \sum -</math>  <math>- = 83, 2 ( ) X X \sum - = 58</math> and <math>2 ( ) Y Y \sum - = 124 r = 2 2 ( ) ( )</math>  <math>( ) X X Y Y X X</math></p>	

<b>27/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>73% MATCHING TEXT</b>	19 WORDS
	<p>X X)(Y Y) n 1 ? ? ? ? or, r = ? ? ? ? ? ? 2 2 )Y (Y )X (X )Y (Y)X (X</p> <p>(X</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p><math display="block">\frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{1}{N} \frac{\sum (Y - \bar{Y}) \sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}</math></p>	
<b>28/52</b>	<b>SUBMITTED TEXT</b>	25 WORDS	<b>41% MATCHING TEXT</b>	25 WORDS
	<p>the square of the correlation coefficient. It explains the extent of the variation of a dependent variable and is expressed by the independent variable.</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p>the square root of the correlation coefficient to get the Coefficient of Determination; this gives the percent of variation in the dependent variable that is 'explained' by the independent variable. •</p>	
<b>29/52</b>	<b>SUBMITTED TEXT</b>	37 WORDS	<b>32% MATCHING TEXT</b>	37 WORDS
	<p>of the variation of a dependent variable and is expressed by the independent variable. A high value of <math>r^2</math> shows a good linear relationship between the two variables. If <math>r = 1</math> and</p> <p><b>W</b> <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a></p>		<p>of the variation in the dependent variable Y is explained by the variation in the independent variable X. If coefficient of correlation is 0, then there is no linear relationship between the two variables, <math>R^2 = 0</math>, and</p>	
<b>30/52</b>	<b>SUBMITTED TEXT</b>	92 WORDS	<b>55% MATCHING TEXT</b>	92 WORDS
	<p>X 106 140 144 155 136 150 146 143 ? X = 1,120 B Y 100 114 120 125 105 124 122 118 ? Y = 928 C X - X -34 0 4 15 -4 10 6 3 D Y - Y -16 -2 4 9 -11 8 6 2 E (X - X) 2 1156 0 16 225 16 100 36 9 ? (X - X) 2 =1,558 F (Y - Y) 2 256 4 16 81 121 64 36 4 ? (</p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p><math display="block">\frac{\sum X Y - \bar{X} \bar{Y}}{\sqrt{\sum X^2 - \bar{X}^2} \sqrt{\sum Y^2 - \bar{Y}^2}} = \frac{83}{\sqrt{(58)(124)}} = 0.9787</math></p>	
<b>31/52</b>	<b>SUBMITTED TEXT</b>	46 WORDS	<b>84% MATCHING TEXT</b>	46 WORDS
	<p><math display="block">\frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{861}{\sqrt{1,120 \cdot 8}} = \frac{861}{\sqrt{8960}} = \frac{861}{94.603} = 9.101</math></p> <p><b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a></p>		<p><math display="block">\frac{\sum X Y - \bar{X} \bar{Y}}{\sqrt{\sum X^2 - \bar{X}^2} \sqrt{\sum Y^2 - \bar{Y}^2}} = \frac{83}{\sqrt{(58)(124)}} = 0.9787</math></p>	

<b>32/52</b>	<b>SUBMITTED TEXT</b>	46 WORDS	<b>84% MATCHING TEXT</b>	46 WORDS
	$X - X)(Y - Y) = 861 X = X n ? = 1,120 8 = 140 ; Y = Y n ? = 928 8 = 116 r = 2 2 (X X) (Y Y) (X X) (Y Y) ? ? ? ? ? ? = 861 1,558$ <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		$X - X)(Y - Y) \sum (X - X) 2 1 N \sum (Y - Y) \sum (X - X)(Y - Y) \sum (X - X) \sum (Y - Y) 2 2 .....(4.3) 2 105$	
<b>33/52</b>	<b>SUBMITTED TEXT</b>	163 WORDS	<b>33% MATCHING TEXT</b>	163 WORDS
	$X 106 140 144 155 136 150 146 143 ? X = 1,120 B Y 100 114 120 125 105 124 122 118 ? Y = 928 C X - X -34 0 4 15 -4 10 6 3 D Y - Y -16 -2 4 9 -11 8 6 2 E (X - X) 2 1156 0 16 225 16 100 36 9 ? (X - X) 2 = 1,558 F (Y - Y) 2 256 4 16 81 121 64 36 4 ? (Y - Y) 2 = 582 G C ? D 544 0 16 135 44 80 36 6 ? (X - X) (Y - Y) = 861 X = X n ? = 1,120 8 = 140 ; Y = Y n ? = 928 8 = 116 r = 2 2 (X X) (Y Y) (X X) (Y Y) ? ? ? ? ? ? = 861 1,558 x 582 = 0.9042$ <p><b>W</b> <a href="https://www.mathcentre.ac.uk/resources/uploaded/hobsonajustthematics20021296smcetp.pdf">https://www.mathcentre.ac.uk/resources/uploaded/hobsonajustthematics20021296smcetp.pdf</a></p>		$x^3 - 5x^2y + 6y^3$ . Solution $\partial z / \partial x = 21x^2 - 10xy$ ; $\partial z / \partial y = -5x^2 + 18y^2$ ; $\partial^2 z / \partial x^2 = 42x - 10y$ ; $\partial^2 z / \partial y^2 = 36y$ ; $\partial^2 z / \partial x \partial y = -10x$ ; $\partial^2 x \partial y = -10x$ . 2. $y \sin x + x \cos y$ . Solution $\partial z / \partial x = y \cos x + \cos y$ ; $\partial z / \partial y = x - \sin y$ ; $\partial^2 x \partial y = -$	
<b>34/52</b>	<b>SUBMITTED TEXT</b>	49 WORDS	<b>52% MATCHING TEXT</b>	49 WORDS
	$X)(Y - Y) = 861 X = X n ? = 1,120 8 = 140 ; Y = Y n ? = 928 8 = 116 r = 2 2 (X X) (Y Y) (X X) (Y Y) ? ? ? ? ? ? = 861 1,558 x 582 = 0.9042$ <p><b>W</b> <a href="http://www.math.umd.edu/~wolfe/241Hf11m3.pdf">http://www.math.umd.edu/~wolfe/241Hf11m3.pdf</a></p>		$x,y) y^*(x.^3)/3; fx=@(x,y) y^* x.^2; fy=@(x,y) x; x=*2:.05:2; y=x; %$	
<b>35/52</b>	<b>SUBMITTED TEXT</b>	21 WORDS	<b>77% MATCHING TEXT</b>	21 WORDS
	<p>the coefficient of determination is the ratio of the explained variation to the total variation. The coefficient of determination is</p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p>the coefficient of determination gives the ratio of the explained variance to the total The coefficient of determination is</p>	
<b>36/52</b>	<b>SUBMITTED TEXT</b>	13 WORDS	<b>83% MATCHING TEXT</b>	13 WORDS
	<p>a is the intercept and b is the slope of the line.</p> <p><b>W</b> <a href="https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf">https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf</a></p>		<p>a is the Y-intercept • b is the slope of the line •</p>	

<b>37/52</b>	<b>SUBMITTED TEXT</b>	13 WORDS	<b>91% MATCHING TEXT</b>	13 WORDS
a is the intercept and b is the slope of the line.		a' is the X-intercept and b' is the slope of the straight line.		
<b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a>				
<b>38/52</b>	<b>SUBMITTED TEXT</b>	24 WORDS	<b>63% MATCHING TEXT</b>	24 WORDS
a is the intercept and b is the slope of the line. Figure 6.2: Linear Equations 6.9.1 Slope of Regression Line		a is the y intercept and b is the slope of the line. ) Example „ Find the equation of the regression line		
<b>W</b> <a href="https://www.hkscem.org/images/stories/ActivityReports/Biostatistics%20l.pdf">https://www.hkscem.org/images/stories/ActivityReports/Biostatistics%20l.pdf</a>				
<b>39/52</b>	<b>SUBMITTED TEXT</b>	15 WORDS	<b>71% MATCHING TEXT</b>	15 WORDS
it represents the change in the dependent variable for every unit change in the		it represents the change in the quantity ( $\omega t + \alpha$ ) for every unit of change in the		
<b>W</b> <a href="https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf">https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf</a>				
<b>40/52</b>	<b>SUBMITTED TEXT</b>	17 WORDS	<b>100% MATCHING TEXT</b>	17 WORDS
a sequence of identical occurrences preceded and followed by different occurrences or by none at all."		a sequence of identical occurrences preceded and followed by different occurrences or by none at all.		
<b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a>				
<b>41/52</b>	<b>SUBMITTED TEXT</b>	28 WORDS	<b>75% MATCHING TEXT</b>	28 WORDS
n n 1 n n ? ? and ? r = 1 2 1 2 1 2 2 1 2 1 2 2 n n (2n n n n ) (n n ) (		N N N N Revolving Projecting N N N N N N N N		
<b>W</b> <a href="https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf">https://docs.sandiego.gov/municode/MuniCodeChapter13/Ch13Art01Division05.pdf</a>				
<b>42/52</b>	<b>SUBMITTED TEXT</b>	30 WORDS	<b>100% MATCHING TEXT</b>	30 WORDS
r = 1 2 1 2 1 2 2 1 2 1 2 2 n n (2n n n n ) (n n ) (n n 1) ? ? ? ? ?		r n n n n n n n n n n		
<b>W</b> <a href="https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1">https://googlegroups.com/group/isbrb22011/attach/362cdc29500933a0/bstats.pdf?part=0.1</a>				



<b>43/52</b>	<b>SUBMITTED TEXT</b>	19 WORDS	<b>100% MATCHING TEXT</b>	19 WORDS
	<p><math>n n (2n n n n) (n n) (n n 1) ? ? ? ? ?</math></p> <p><b>W</b> <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a></p>		<p><math>n1n2)/(n1 + n2)] + 1 = [(2 \times 13 \times 15)/(13 + 15)] + 1 = (390/28) + 1 = 14.93</math></p> <p><math>2n1n2 (2n1n2 - n1 - n2)</math></p>	
<b>44/52</b>	<b>SUBMITTED TEXT</b>	37 WORDS	<b>81% MATCHING TEXT</b>	37 WORDS
	<p><math>n n 1 n n ? ?</math> and <math>? r = 1 2 1 2 1 2 2 1 2 1 2 2 n n (2 n n n n) (n n) (n n 1) ? ? ? ? ?</math></p> <p><b>W</b> <a href="https://en.wikipedia.org/wiki/Continued_fraction">https://en.wikipedia.org/wiki/Continued_fraction</a></p>		<p><math>n, 15 n, 17 n, 19 n, \dots ] \{\displaystyle \backslash n\}=[0;</math> <math>n,3n,5n,7n,9n,11n,13n,15n,17n,19</math></p>	
<b>45/52</b>	<b>SUBMITTED TEXT</b>	37 WORDS	<b>81% MATCHING TEXT</b>	37 WORDS
	<p><math>n n 1 n n ? ?</math> and <math>? r = 1 2 1 2 1 2 2 1 2 1 2 2 n n (2 n n n n) (n n) (n n 1) ? ? ? ? ?</math></p> <p><b>W</b> <a href="https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf">https://www.mathcentre.ac.uk/resources/uploaded/hobsonajjustthemaths20021296smcetp.pdf</a></p>		<p><math>n(n + 1) 2 + 3 1 6 n(n + 1)(2 n + 1) + 2 1 2 n(n + 1) = 1 4</math> <math>n(n + 1)[n(n + 1) + 4n + 2 + 4] = 1 4 n(n + 1)[</math></p>	
<b>46/52</b>	<b>SUBMITTED TEXT</b>	20 WORDS	<b>57% MATCHING TEXT</b>	20 WORDS
	<p>Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by</p> <p><b>W</b> <a href="https://erho.weebly.com/uploads/2/7/8/4/27841631/4.1_-_correlation_and_slrm.pdf">https://erho.weebly.com/uploads/2/7/8/4/27841631/4.1_-_correlation_and_slrm.pdf</a></p>		<p>coefficient of determination is the portion of the total variation in the dependent variable that is explained by</p>	
<b>47/52</b>	<b>SUBMITTED TEXT</b>	20 WORDS	<b>57% MATCHING TEXT</b>	20 WORDS
	<p>Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by</p> <p><b>W</b> <a href="http://fin.bus.ku.ac.th/135512%20Economic%20Environment%20for%20Finance/Lecture%20Slides/Linear%20...">http://fin.bus.ku.ac.th/135512%20Economic%20Environment%20for%20Finance/Lecture%20Slides/Linear%20...</a></p>		<p>coefficient of determination is the portion of the total variation in the dependent variable that is explained by</p>	
<b>48/52</b>	<b>SUBMITTED TEXT</b>	24 WORDS	<b>58% MATCHING TEXT</b>	24 WORDS
	<p>coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the</p> <p><b>W</b> <a href="https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf">https://www.rjerz.com/c/bas/baug/lectures/WK09-Correlation_and_Regression_6.pdf</a></p>		<p>Coefficient of Determination coefficient of determination (<math>r^2</math>) is the proportion of the total in the dependent variable (Y) that is explained or accounted for the</p>	

<b>49/52</b>	<b>SUBMITTED TEXT</b>	24 WORDS	<b>52% MATCHING TEXT</b>	24 WORDS
coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the		COEFFICIENT OF DETERMINATION Coefficient of determination gives the percentage variation in the dependent variable that is accounted for the		
W <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a>				
<b>50/52</b>	<b>SUBMITTED TEXT</b>	21 WORDS	<b>62% MATCHING TEXT</b>	21 WORDS
Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the		coefficient of determination measures the amount of variation in the dependent variable that is explained by the		
W <a href="https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller">https://www.worldsupporter.org/en/chapter/41133-samenvatting-managerial-statistics-keller</a>				
<b>51/52</b>	<b>SUBMITTED TEXT</b>	28 WORDS	<b>68% MATCHING TEXT</b>	28 WORDS
of the coefficient of determination. Coefficient of Determination: A measure of the proportion of variation in Y, the dependent variable that is explained by the regression line,		of Estimate „ The coefficient of coefficient of determination, by $r^2$ , is a measure of the variation of the dependent variable that is explained by the regression line		
W <a href="https://www.hkscem.org/images/stories/ActivityReports/Biostatistics%20I.pdf">https://www.hkscem.org/images/stories/ActivityReports/Biostatistics%20I.pdf</a>				
<b>52/52</b>	<b>SUBMITTED TEXT</b>	18 WORDS	<b>73% MATCHING TEXT</b>	18 WORDS
the value of the independent variable increases, so does the value of the dependent variable, and this		the value of the independent variable we can obtain the value of the dependent variable and this		
W <a href="https://pdfcoffee.com/objective-2-pdf-free.html">https://pdfcoffee.com/objective-2-pdf-free.html</a>				