Indira Gandhi
National Open University
School of Sciences

MST-004 STATISTICAL INFERENCE

Block

1

SAMPLING DISTRIBUTIONS

UNIT 1

Introduction to Sampling Distribution **7**

UNIT 2

Sampling Distribution(s) of Statistic(s) **29**

UNIT 3

Standard Sampling Distributions-I **53**

UNIT 4

Standard Sampling Distributions-II **69**

Appendix **91**

Curriculum and Course Design Committee

Block Preparation Team

Prof. G. K. Shukla **(Editor)**
Decision Science Group
Indian Institute of Management
Lucknow (UP)

Dr. Parmod Kumar **(Language Editor)**
School of Humanities, IGNOU

Mr. Prabhat Kumar Sangal **(Units 1 - 4)**
School of Sciences, IGNOU

Dr. Manish Trivedi **(Units 3 & 4)**
School of Sciences, IGNOU

Course Coordinator: Mr. Prabhat Kumar Sangal
Programme Coordinator: Dr. Manish Trivedi

Block Production

Mr. Sunil Kumar, AR (P),School of Sciences, IGNOU
CRC prepared by Mr. Prabhat Kumar Sangal, School of Sciences, IGNOU

Acknowledgement: I gratefully acknowledge my colleagues Dr. Manish Trivedi and Mr. Rajesh Kaliraman, Statistics Discipline, School of Sciences for their great support.

July, 2013

ISBN-978-81-266-

Further information on the Indira Gandhi National Open University may be obtained from University's Office at Maidan Garhi, New Delhi-110068 or visit University's website http://www.ignou.ac.in

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Director, School of Sciences.

Printed at:

STATISTICAL INFERENCE

In MST-003, the basic rules of probabilities and various probability distributions such as Bernoulli, Binomial, Poisson, Uniform, Normal, Exponential, etc. have been discussed. Now, the concepts of probability theory and probability distributions are used in drawing inference about the population parameters.

Generally, population parameters are unknown and when the population is too large or the units of the population are destructive in nature or there is a limited resources and manpower available then it is not possible practically to examine each and every unit of the population to obtain the population parameters. In such situations, one can draw sample from the population under study and utilize sample observations to draw reliable conclusions about the population parameters. The results obtained from the sample are projected in such a way that they are valid for the entire population. This technique is known as **statistical inference**. The statistical inference may be divided into two areas or parts:

(i) The population parameters are unknown and we may want to guess the true value of the unknown parameters on the basis of a random sample drawn from the population. This type of problem is known as **"Estimation"**.

(ii) Some information is available about the population and we may like to verify whether the information is true or not on the basis of a random sample drawn from the population. This type of problem is known as **"Testing of hypothesis"**.

This course is divided into four blocks and each block consists of four units.

In **Block 1**, we shall be discussing the concept of sampling distribution of a statistic. The sampling distributions of sample mean, sample proportion, sample variance, difference of two sample means, etc. are also discussed in this block. Here, we also discuss some important sampling distributions which are widely used in statistical inference and known as exact sampling distributions such as χ^2, t and F.

In **Block 2**, we discuss first part of statistical inference called estimation, through which we find the estimate of unknown parameter on the basis of sample data. Two types of estimation as point estimation and interval estimation are discussed in this block. The criteria of good estimator and different methods of estimation for the unknown parameters are also discussed in this block.

In **Block 3**, the second part of statistical inference, that is, testing of hypothesis is discussed. We discuss how the estimated or assumed or a hypothetical value of the population parameter can be tested and how we can take the decision about rejection or non-rejection of the hypothesis or assumed value. Here, different parametric tests as Z-test, t-test, F-test and χ^2-test for testing of hypothesis about population mean, difference of two population means, population proportion, difference of two population proportions, population variance, etc. are also discussed.

Most of the parametric tests are based on the assumptions that the parent population is normal and the data are measured in interval or ratio scale and concerned with testing the population mean, proportion, variance, etc. But if the assumption(s) is (are) not fulfilled then we cannot used the parametric test.

In such situations, non-parametric tests do the same job for us. In **Block 4,** we discussed different non-parametric tests i.e. sign test, Wilcoxon signed-rank test, chi-square test for goodness of fit and independence of attributes, Kolmogorov-Smirnor test for goodness of fit, Kruskal -Wallis test for one way analysis of variance with completed randomised design, Friedman test for randomised block design, etc.

Although the material is self contained and self explained in nature, even though, if the learners are interested to gain more and want to study the contents in more detail, you may consult the following books:

- Goon, A.M., Gupta, M.K. and Dasgupta, B.; *An Outline of Statistical Theory*, Vol II, World Press, Calcutta.
- Rohatgi, V.K.; *An Introduction to Probability and Statistics*, Wiley-India.
- Mood, A.M.; *Introduction to the Theory of Statistics*, Tata McGraw-Hill Book Company, New Delhi.
- Gupta, S. C. and Kapoor, V. K.; *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons.
- Makhopadhyay. P.; *Applied Statistics*, New Central Book Agency, Calcutta.
- Gibbons, J.D.; *Nonparametric Statistical Inference*, McGraw-Hill Book Company, New York.
- Daniel Wayne W.; *Applied Nonparametric Statistics*, Houghton Mifflin Company.
- Conover, W.J.; *Practical Nonparametric Statistics (3rd Ed.)*, Wiley-India.

SAMPLING DISTRIBUTIONS

One of the major objectives of statistical analysis is to draw reliable conclusions about the population on the basis of the analysis of the sample data. This process is known as statistical inference. For drawing the inference about the population parameter, we use a function of sample values such as sample mean, sample proportion, sample standard deviation, etc. which is known as statistic. If we draw all possible samples of same size and for each sample we calculate a statistic then the value of statistic may or may not be varying sample to sample. If we arrange all possible values of that statistic with its corresponding probabilities then this arrangement is known as sampling distribution of that statistic.

This is the first block of the course MST-004. The aim of this block is to put a foundation stone for the next blocks of this course and will provide a platform to the learners to understand the statistical inference. It comprises four units.

Unit 1: Introduction to Sampling Distribution

This unit explains the basis concept of sampling distribution of a statistic with examples. Here, we also discussed the most important theorem of Statistics "Central Limit Theorem" according to this theorem, the sampling distributions tend to be approximately normal for sample sizes greater than 30. In this unit, we also discussed the law of large numbers. According to law of large numbers, we come to know "how we draw a reliable inference about the population parameter by using a sample of finite size?"

Unit 2: Sampling Distribution(s) of Statistic(s)

This unit explores the sampling distributions of sample mean, sample proportion, sample variance, difference of two sample means, and difference of two sample proportions and ratio of two sample variances.

Unit 3: Standard Sampling Distributions-I

Sometimes, statistic may follow a particular sampling distribution such as χ^2 (read as chi-square), t, F, etc. which are known as exact sampling distributions. This unit provides the brief introduction of chi-square and t-distributions. Properties and applications of these distributions also explained in this unit.

Unit 4: Standard Sampling Distributions-II

Last unit of this block is devoted to the rest important sampling distribution which is widely used in Statistics i.e. F-distribution. Here, we discuss the properties and applications of this distribution. The standard sampling distributions are interacted with each other, therefore, the relations between them are also explored in this unit. In this unit, we also describe the procedure of finding the tabulated (critical) value of a variate which follows the exact sampling distribution such as χ^2, t and F.

Notations and Symbols

$X_1, X_2, \ldots, X_n$	: Random sample of size n
SRSWR	: Simple random sampling with replacement
SRSWOR	: Simple random sampling without replacement
$\bar{X}$	: Sample mean
S^2	: Sample variance
SE	: Standard error
$N(\mu, \sigma^2)$	: Normal distribution with mean μ and variance σ^2
$Z \sim N(0, 1)$	: Standard normal variate
P	: Population proportion
p	: Sample proportion
ν	: Degrees of freedom(df)
$\chi^2_{(\nu)}$	: Chi-square with ν df
Γ	: Gamma function
$B(a, b)$	: Beta function
log x	: Logarithm base e
α	: Area under the curve
$t_{(\nu), \alpha}$	: Tabulated value of t-statistic with ν df such that the area under the curve of t-distribution to its right (upper) tail is equal to α
$\chi^2_{(\nu), \alpha}$	: Tabulated value of χ^2-statistic with ν df such that the area under the curve of χ^2-distribution to its right (upper) tail is equal to α
$F_{(\nu_1, \nu_2), \alpha}$	: Tabulated value of F-statistic with (ν_1, ν_2) df such that the area under the curve of F-distribution to its right (upper) tail is equal to α