IIIrd SEMESTER B.SC PSYCHOLOGY

CALICUT UNIVERSITY

# PROBABILITY DISTRIBUTIONS AND PARAMETRIC TESTS

Prepared by

SAJITHA.K.S

Asst. Prof. in Department of Psychology

# **SYLLABUS**

# **SEMESTER III**

#### **STA 3C 02-** PROBABILITY DISTRIBUTIONS AND PARAMETRIC TESTS

Contract Hours per week:5Number of credits:3Number of Contact Hours:90Course Evaluation:External 60 Marks+ Internal 15 MarksDuration of Exam:2 Hours

#### **Question Paper Pattern**

Type of	Question number	
Questions	(From To)	Marks
	01 to 12	Short answer type carries 2 marks each - 12 questions
		(Maximum Marks 20)
Short Answer		
	13 to 19	Paragraph/ Problem type carries 5 marks each – 7
Paragraph/		questions
Problems		(Maximum Marks 30)
	20 to 21	Essay type carries 10 marks (1 out of 2)
Essay		(Maximum Marks 10)
Total	01 to 21	60

Question Paper setter has to give equal importance to both theory and problems in sections B and C.

#### Objectives

- 1. To get a general understanding on various probability distributions
- 2. To familiarize the uses of Statistical test.

**Module 1:** *Distribution Theory*- Binomial, Poisson and Normal Distributions, Mean and Variance (without derivations), Numerical Problems, Fitting, Importance of Normal Distribution, standard normal distribution, simple problems using standard normal tables, Central Limit Theorem (Concepts only)

#### 25 Hours

**Module2:** *Methods of Sampling*- Random Sampling, Simple Random Sampling, Stratified, Systematic and Cluster Sampling, Non Random sampling, Subjective sampling, Judgment sampling and convience sampling

#### 20 Hours

**Module 3:** *Fundamentals of Testing*- Type-I & Type-II Errors, Critical Region, Level of Significance, Power, *p* value, Tests of Significance

**Module 4:** Large Sample Tests – Test of a Single, Mean Equality of Two Means, Test of a Single Proportion, and Equality of Two Proportions

#### 10

**Hours Module 5:** *Small Sample tests*-Test of a Single Mean, Paired and Unpaired t-Test, Chi- Square Test of Variance, F-Test for the Equality of Variance, Tests of Correlation

#### 20

#### Hours

# References

- 1. Gupta, S.P. Statistical Methods. Sultan Chand and Sons: New Delhi.
- 2. Gupta, S.C., &Kapoor, V.K. *Fundamentals of Applied Statistics*. New Delhi: Sultan Chand and Sons.
- 3. Garret, H.E., &Woodworth, R.S. *Statistics in Psychology and Education*. Bombay: Vakila, Feffex and Simens Ltd.
- 4. Mood, A.M., Graybill, F.A and Boes, D.C. *Introduction to Theory of Statistics*. 3rd Edition Paperback International Edition.
- 5. Mukhopadhyay, P. *Mathematical Statistics*. New central Book Agency (P) Ltd: Calcutta.

# **DISTRIBUTION THEORY**

# **INTRODUCTION**

We have seen that a random variable is a variable that is subject to random variations so that it can take on different values, each with an associated probability. We have seen that a probability distribution links each outcome of a random experiment to a value of the random variable or process with its probability of occurrence.

#### **Probability distribution**

It is also known as theoritical distribution can be defined as a distribution of frequencies which is not based on actual experiments or observations but is constructed through expected frequencies obtained by mathematical computation based on certain hypothesis.

#### **Binomial distribution**

<u>It</u> can be thought of as simply the probability of success or failure outcome inn an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes.

#### → Mean and variance of B

For a binomial distribution *Mean=np and variance=npq* 

If X-B(8,1\3), Mean=8\3 and variance =8 $3 \times 2/3 = 16/9$ 

# Fitting of a Binomial Distribution

The probabilities of the binomial distribution to be fitted is calculated using the

# formula,

 $P(X = x) = (n/x) p^{x}q^{n-x}, x=0,1,2,...n$ 

Binomial distribution must meet the following three criteria:

- 1. The number of observations or trial is fixed
- 2. Each observation or trial is independent.
- 3. The probability of success is exactly the same from one trial to another.

# Poisson Distribution

It was first introduced by Simeon Denis Poisson . Poisson distribution is a tool that helps to predict the probability of certain events from happening when you know how often the events has occurred. It gives us the probability of a given number of events happening in a fixed interval of time.

Formula to find Poisson distribution is given below:

 $P(x) = (e^{-\lambda} * \lambda^{x}) / x!$ 

# **Normal Distribution**

\_\_\_\_\_ Normal Distribution is a continuous symmetrical probability distribution in which frequencies are distributed even about the mean of distribution.

# Properties of Normal Distribution:

- 1. The mean, median, .and mode are all equal
- 2. The curve is symmetric at the centre
- **3.** Exactly half of the values are to the left of center and exactly half the values are to the right.
- 4. The total area under the curve is 1
- 5. A standard normal model is a normal distribution with a mean of 1 and SD of 1

# SHAPE OF NORMAL DISTRIBUTION



#### **Standard Normal Distribution**

The **standard normal distribution**, also called the *z*-**distribution**, is a special <u>normal distribution</u> where the <u>mean</u> is 0 and the SD is 1.

Any normal distribution can be standardized by converting its values into *z*-scores. *Z*-scores tell you how many standard deviations from the mean each value lies.



#### **Importance of z scores:**

The absolute value of z scores indicates how far the scores lies from the mean when measured in standard deviation.

#### **CENTRAL LIMIT THEOREM(CLT)**

The CLT states that if you have a population with mean and SD and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed.

# MODULE 2 METHODS OF SAMPLING

# Population and sample

- **1.** The set of all units from which the required information has to be collected is called **population.**
- 2. The selected representative subset of population is called **sample**.
- 3. Sampling is the process of selecting this subset, on which we focus our study.

#### Advantages of sampling

- 1. Less time
- 2. Less cost
- 3. Only technique if the testing process is destructive
- 4. Only technique when there is practical infeasibility.
- 5. Only technique when the population is infinite
- 6. Enough reliability of inference based on sampling
- 7. Quality of data collected

#### **Disadvantages of sampling**

- 1. Chances of bias
- 2. Difficulties in selecting truly representative samples
- 3. Inadequate knowledge in the subject
- 4. Changeability of units
- 5. Impossibility of sampling

#### **TECHNIQUES OF SAMPLING**

It may be classified as probability and nonprobability sampling.

- 1. <u>Probability sampling</u>:- It is based on the theory of probability. It includes
  - a) <u>Simple random sampling:-</u> It can be defined as each item of the population has equal chance of being selected. It use lottery method and table random numbers for selecting samples.
  - b) <u>Systematic sampling:</u> Systematic sampling is a type of probability sampling method in which sample members from a larger population are selected according to a random starting point but with a fixed, periodic interval. This interval, called the sampling interval, is calculated by dividing the population size by the desired sample size

- c) \_Stratified sampling:- Under this, dividing the heterogeneous population into homogeneous called stratum, from which the sample is selected.
- d) Cluster sampling:- population divided into groups called cluster. Then usin simple random method selecting one cluster. From that cluster, selecting all the samples.

#### Merits and Demerits of Simple random sampling

- 1. Lack of bias
- 2. Simplicity
- 3. Less knowledge required

#### **Demerits:-**

- **1.** Time consuming
- 2. Cost involved
- **3.** Difficulty to get all list of the population

# Merits and Demerits of systematic sampling

# <u>Merits</u>

- Easy to Execute and Understand.
- Control and Sense of Process.
- Clustered Selection Eliminated.
- Low Risk Factor.
- Assumes Size of Population Can Be Determined.
- Need for Natural Degree of Randomness.
- Greater Risk of Data Manipulation.

# **Demerits**

- 1. Greater risk
- 2. Assumes size of the population can be determined

# Merits and demerits of stratified sampling merits

- The population consists of *N* elements.
- The population is divided into *H* groups, called **strata**.
- Each element of the population can be assigned to one, and only one, stratum.
- The researcher obtains a probability sample from each stratum.

#### **Demerits**

stratified sampling has two main disadvantages. It may require more administrative effort than a simple random sample. And the analysis is computationally more complex.

#### 2. Non probability sampling

Non-probability sampling is a sampling method in which not all members of the population have an equal chance of participating in the study, unlike probability sampling. It includes

- a) <u>Convenience sampling:-</u> under this, the choice of the sample let completely to the convenience of the interviewer.
- b) <u>Purposive/judgement sampling</u>: Purposive sampling, also known as judgmental, selective, or subjective sampling, is a form of non-probability sampling in which researchers rely on their own judgment when choosing members of the population to participate in their surveys.
- c) <u>Quota sampling</u>: Under this, we divide the population into different subpopulation and from which we select units according to our convenience.
- d) <u>Snowball sampling:</u>- In which existing study subjects recruit future subject from among their acquaintances.

# <u>Sampling and nonsampling errors</u>

Sampling error is a type of error, occurs due to the sample selected does not perfectly represents the population of interest.

An error occurs due to sources other than sampling, while conducting survey activities is known as nonsampling errors. Important nonsampling errors are

- Faulty planning
- Errors in response
- Errors in the design of survey
- Errors in compilation
- Publication errors

# **FUNDAMENTALS OF TESTING**

\_\_\_\_\_To differentiate between real systematic patterns and random, chance occurrence, researchers rely on statistical techniques known as hypothesis testing, which is introduced in this module.

# **Test of Hypothesis**

It is a process or procedure under which a statistical hypothesis is laid down and it is accepted or rejected on the basis of a random sample drawn from the population. Hypothesis must be

- Clear
- Testable
- Related to the existing body or theory
- Logical unity and comprehensive
- Capable of verification
- Operationisable

# **Null and Alternative Hypothesis**

- **A null hypothesis** is that hypothesis which is tested for its possible rejection under the assumption that it is true
- The hypothesis which is tested against the null hypothesis is called *alternative hypothesis*

# **Errors in Hypothesis testing**

- **Type I error :-** It occurs when a researcher rejects null hypothesis that is actually true.
- **Type II error:** It occurs when a researcher fails to reject a null hypothesis that is really false.

# Level of significance

The upper limit for the probability of type I error, fixed by the researcher is called level of significance.

# **Critical region**

In a test procedure we calculate a test statistic oon which we base our decision. The range of variation of this statistics is divided into two regions, acceptance region and rejection region. If the computed value of the test statistics falls in the rejection region we reject the null hypothesis. The rejection region is also known as critical region.

# **Critical value**

The value of the test statistic which separates the rejection region from the acceptance region is called the critical value.

#### One tailed and two tailed test

- In **one tailed** test, the rejection region will be located in only one tail which may be either right or left.
- A *two tailed test* is one in which we reject the null hypothesis if the computed value of the test static is significantly greater than or lower than the critical value of the test static.

# Power of a test

Probability for rejecting the null hypothesis when the alternative hypothesis is true is called power of a test.

Power of a test = 1-P[type II error]

#### **Sampling Distribution**

Sampling distribution of a statistic can be thought of as the distribution of values obtained for that statistic over repeated sampling

#### Sampling distribution of sample mean

Sample mean is a random variable, as it changes from one sample to another with a particular probability. Therefore, sample mean has a probability distribution known as sampling distribution of sample mean.

# **T** distribution

The T distribution, also known as the Student's t-distribution. It is a type of probability distribution that is similar to the normal distribution which has bell shape but has heavier tails.

#### F distribution

The *F*-distribution is a particular parametrization, which is also called the beta distribution of the second kind. The letter F is used to represent a test statistic that follows the F distribution.

#### Chi square distribution

In probability theory, the chi-squared distribution (also chi-square or  $\chi^2$ -distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables.

# LARGE SAMPLE TESTS

# Meaning

A test is said to be a large sample test if the sample size is greater than 30.

The following is the test procedure for any simple null hypothesis tested against a simple alternative.

- Set up the null and alternative hypothesis
- Choose the suitable test statistic
- Calculate the value of the test statistic
- Compare the calculated value with the corresponding table value
- Check whether the value falls in the acceptance region or rejection region
- Decide to accept or reject H<sub>0</sub> based on the previous step
- Arrive at the inferences

# Important large sample test

- 1. Test of a single mean
- 2. Test of equality of two means
- **3.** Test of a single proportion
- 4. Test of equality of two proportion

#### 1. Test of a single mean

This test is intended to test whether the given population mean is true.that is, to test whether the difference between sample mean and population mean is significant or it is only due to sampling fluctuations. For this type of problems we can apply Z test or t test. The test procedure is

- First set the hypothesis
- > Secondly decide the test criterion ie Z or t test
- > Using the formula  $(\bar{x} \mu)$  / SE

 $\bar{\mathbf{x}} =$ Mean of Sample

- $\mu$  = Mean of Population
- $\sigma$  = Standard Deviation of Population
- **n** = Number of Observation
- ➤ For Z test , the degree of freedom is infinity while for t test it is n-1.
- Then get the table value of the test statistic, for the degree of freedom and level of significance.
- Finally take a decision either to accept or to reject the hypothesis originally set in step I.
- 2. Testing equality of two population means:

# Procedure

- > Set up a null hypothesis that there is no significant difference between the two means ie  $\mu 1 = \mu 2$
- > When population SD is known or when the samples are large the test applied is Ztest.

Otherwise t test

- ➢ Use Test statistic
- Degree of freedom is (n1+n2-2)
- Obtain the table value
- When the calculated value is less than the table value numerically the test statistic falls in the acceptance region and so we accept the null hypothesis.

#### 3. Testing population proportion

Here the null hypothesis is there is no significant difference between sample proportion and population proportion or  $P=P_0$ . We apply Z test

#### 4. Testing equality of two proportion

Here the hypothesis is there is no significant difference between two population proportions or  $P_1=P_2$ . We apply Z test

#### SMALL SAMPLE TEST

#### Meaning

A test is said to be a small sample test if the sample size is less than 30.

**Test of a single mean;** The single mean (or one-sample) t-test is used to compare the mean of a variable in a sample of data to a (hypothesized) mean in the population from which our sample data are drawn. This is important because we seldom have access to data for an entire population.

**Test of equality of two population means:** Test if two population means are equal. The two-sample t-test is **used to determine if two population means are equal**. A common application is to test if a new process or treatment is superior to a current process or treatment.

**Paired t test :** Paired T-Test. The paired sample t-test, sometimes called the dependent sample t-test, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample t-test, each subject or entity is measured twice, resulting in pairs of observations.

**Unpaired t test** : An unpaired t-test (also known as an independent t-test) is a statistical procedure that compares the averages/means of two independent or unrelated groups to determine if there is a significant difference between the two.

**Test of equality of two variance:** An F-test is used to test if the variances of two populations are equal. This test can be a two-tailed test or a one-tailed test. The two-tailed version tests against the alternative that the variances are not equal.

#### Chi square test of variance

The chi-square test for variance is a non-parametric statistical procedure with a chi-square-distributed test statistic that is used for determining whether the variance of a variable obtained from a particular sample has the same size as the known population variance of the same variables.

#### F test for the equality of variance

In statistics, an F-test of equality of variances is a test for the null hypothesis that two normal populations have the same variance.

#### Test of correlation

Correlation test is **used to evaluate the association between two or more variables**. For instance, if we are interested to know whether there is a relationship between the heights of fathers and sons, a correlation coefficient can be calculated to answer this question.