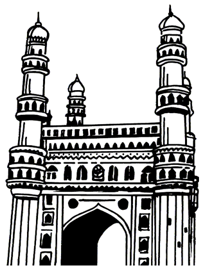
Rahul's ✔

Topper's Voice

AS PER CBCS SYLLABUS



B.Sc.

II Year IV Sem

Latest **2021-22** Edition

STATISTICS

PAPER - IV : INFERENCE

- **Study Manual**
- **FAQ's and Important Questions**
- **Short Question & Answers**
- **Choose the Correct Answers**
- **Fill in the blanks**
- **Exercise Problems**
- **Solved Previous Question Papers**
- **Solved Model Papers**

Price
` .149-00

- by -

WELL EXPERIENCED LECTURER

Rahul Publications™

Hyderabad. Ph : 66550071, 9391018098

B.Sc.

II Year IV Sem

STATISTICS

PAPER - IV : INFERENCE

Price ` 149-00

Sole Distributors : **☎ : 65500071, Cell : 9391018098**

VASU BOOK CENTRE

Shop No. 3, Beside Gokul Chat, Koti, Hyderabad.

Maternity Hospital Opp. Lane, Narayan Naik Complex, Koti, Hyderabad.

Near Andhra Bank, Subway, Sultan Bazar, Koti, Hyderabad -195.

CONTENTS

STATISTICS

PAPER - IV : INFERENCE

STUDY MANUAL

SOLVED PREVIOUS QUESTION PAPERS

SOLVED MODEL PAPERS

SYLLABUS

UNIT - I

Concepts of statistical hypotheses, null and alternative hypothesis, critical region, two types of errors, level of significance and power of a test. One and two tailed tests, test function (nonrandomized and randomized). Statement and Proof of Neyman-Pearson's fundamental lemma for Randomized tests. Examples in case of Binomial, Poisson, Exponential and Normal distributions and their powers.

UNIT - II

Large sample tests for single sample mean, difference of means, single sample proportion, difference of proportions and difference of standard deviations. Fisher's Z-transformation for population correlation coefficient(s) and testing the same in case of one sample and two samples. Definition of order statistics and statement of their distributions.

UNIT - III

Tests of significance based on χ^2 - χ^2 - test for specified variance, goodness of fit and test for independence of attributes (r x s, 2 × k and 2 × 2 contingency tables). Tests of significance based on student's - t – t-test for single sample specified mean, difference of means for independent and related samples, sample correlation coefficient. F - test for equality of population variances.

UNIT - IV

Non-parametric tests- their advantages and disadvantages, comparison with parametric tests. Measurement scale- nominal, ordinal, interval and ratio. Use of Central Limit Theorem in testing. One sample runs test, sign test and Wilcoxon-signed rank tests (single and paired samples). Two independent sample tests: Median test, Wilcoxon –Mann-Whitney U test, Wald Wolfowitz's runs test. Use of central limit theorem in testing.

Contents

Frequently Asked & Important Questions

UNIT - I

1. What is Hypothesis?

Ans : **(Feb.-21, June-19)**

Refer Unit-I, Q.No. 1

2. Explain briefly about test function.

Ans : **(June-19)**

Refer Unit-I, Q.No. 6

3. State and prove the Neymann Pearson's Lemma.

Ans : **(Feb.-21, June-19, June-18)**

Refer Unit-I, Q.No. 8

4. Let P the probability that a coin will fall head in a single toss in order to test $H_o : p = \frac{1}{2}$ against $H_1 : p = \frac{3}{4}$. The coin is tossed 5 times and H_o is rejected if more than 3 heads are obtained. Find the probability of Type I Error and power of the test.

Sol : **(June-18)**

Refer Unit-I, Prob. 7

5. Find the Best Critical region for testing $H_0 : \lambda = \lambda_0$ and against $H_1 : \lambda = \lambda_1$ for the poisson distribution.

Sol : **(Feb.-21)**

Refer Unit-I, Prob. 11

6. Obtain the most powerful test for testing the mean $\mu = \mu_0$ against $\mu = \mu_1 (\mu_1 = \mu_0)$ where $\sigma^2 = 1$ in normal population.

Sol : **(June-19)**

Refer Unit-I, Prob. 15

UNIT - II

1. Explain the Procedure for Testing of Hypothesis.

Ans : **(June-19, Imp.)**

Refer Unit-II, Q.No. 2

2. Explain the procedure for testing single sample mean.

Ans : **(Imp.)**

Refer Unit-II, Q.No. 3

3. Derive the large sample test procedure for difference of means.

Ans : **(Feb.-21, Imp.)**

Refer Unit-II, Q.No. 4

4. Data on days to maturity were recorded in two varieties of a pulse crop Deter-mine whether two means are significantly different.

	n	Mean	Variance
Variety A	60	60	8.20
Variety B	65	65	11.13

Sol : **(May/June-18)**

Refer Unit-II, Prob. 9

5. Describe the large sample test for single proportion.

Ans : **(Feb.-21)**

Refer Unit-II, Q.No. 6

6. Describe the test procedure for diffe-rence of proportions.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Refer Unit-II, Q.No. 7

7. Explain Fisher's Z-transformation. And its applications.

Ans : **(June-19, June-18, Imp.)**

Refer Unit-II, Q.No. 8

8. A random sample of 30 pairs of observation given a correlation coefficient 0.61 can it be regarded as drawn from a bivariate normal population having correlation coefficient is 0.7. Also compute the 95% confidence limits for the population correlation coefficient.

Sol : **(Imp.)**

Refer Unit-II, Prob. 25

9. Describe the test procedure for Fisher's Z-transformation for differences of Correlation Coefficient.

Ans : **(Feb.-21, Imp.)**

Refer Unit-II, Q.No. 10

10. Define order statistics and state their distributions.

Ans : **(June-19, June-18, Imp.)**

Refer Unit-II, Q.No. 11

UNIT - III

1. What is t-distribution? Explain the properties and applications of t-distribution.

Ans : **(June-21, Imp.)**

Refer Unit-III, Q.No. 2

2. Explain t-test for single mean.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Refer Unit-III, Q.No. 3

3. Explain t-test for difference of two means of Independent samples.

Ans : **(June-18)**

Refer Unit-III, Q.No. 4

4. Explain t-test for difference of two means of dependent samples.

Ans : **(Feb.-21, Imp.)**

Refer Unit-III, Q.No. 5

5. Explain t-test for correlation coefficient.

Ans : **(Feb.-21, Imp.)**

Refer Unit-III, Q.No. 6

6. Explain the concept of F-test for equality of population variance.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Refer Unit-III, Q.No. 7

7. What is χ^2 - test? State the features of χ^2 - test.

Ans : **(Imp.)**

Refer Unit-III, Q.No. 9

8. Explain the procedure for determining the value of χ^2.

Ans : **(Feb.-21, Imp.)**

Refer Unit-III, Q.No. 10

9. Write a short notes on χ^2 - test for goodness of fit.

Ans : **(June-19, June-18, Imp.)**

Refer Unit-III, Q.No. 14

10. Explain χ^2 - test for independent of two attribute.

Ans : **(Feb.-21, June-19, June-18)**

Refer Unit-III, Q.No. 15

UNIT - IV

1. What is meant by Non-parametric Tests. State their Advantages and Disadvantages.

Ans : **(June-19, June-18)**

Refer Unit-IV, Q.No. 1

2. Distinguish between Parametric Test and Non-parametric Test.

Ans : **(Feb.-21, June-19)**

Refer Unit-IV, Q.No. 2

3. Write a short note on Measurement of Scale.

Ans : **(June-19)**

Refer Unit-IV, Q.No. 3

4. State central limit theorem.

Ans : **(Feb.-21)**

Refer Unit-IV, Q.No. 4

5. Write a short notes on sign tests.

Ans : **(Feb.-21, June-18)**

Refer Unit-IV, Q.No. 6

6. Explain the median test procedure.

Ans : **(June-18)**

Refer Unit-IV, Q.No. 9

7. Explain Wilcoxon – Mann Whitney U Test procedure.

Ans : **(Feb.-21, June-19, June-18)**

Refer Unit-IV, Q.No. 10

8. Explain Wald - Wolfowitz Run test for one sample (or) test for Randomness Run.

Ans : **(Imp.)**

Refer Unit-IV, Q.No. 11

UNIT I	Concepts of statistical hypotheses, Null and Alternative hypothesis, Critical region, two types of errors, Level of significance and Power of a test. One and two tailed tests, test function (non-randomized and randomized). Statement and Proof of Neyman-Pearson's fundamental lemma for Randomized tests. Examples in case of Binomial, Poisson, Exponential and Normal distributions and their power of the test functions.

1.1 Concepts of Statistical Hypotheses

1.1.1 Null and Alternative Hypothesis, Critical Region

Q1. What is Hypothesis?

(OR)

Define the following terms :

(i) Hypothesis

(ii) Null Hypothesis

(iii) Alternative Hypothesis

(iv) Critical Region

Ans : **(Feb.-21, June-19)**

(i) Hypothesis :

Any statement about the population parameter is called hypothesis. If it is testing the hypothesis then it is called Testing of hypothesis

Ex : If we want to test the population mean $\mu = 45$ or not

(ii) Null Hypothesis :

It is denoted by H_0 and it is defined by a statement no differences is called Null hypothesis.

Ex : A sample of size $n = 30$ drawn from a population and it is found that the sample mean $\bar{x} = 15$. If we want to test the population mean is 10 (or) not ? then the null hypothesis is $H_0 : \mu = 10$

(iii) Alternative Hypothesis

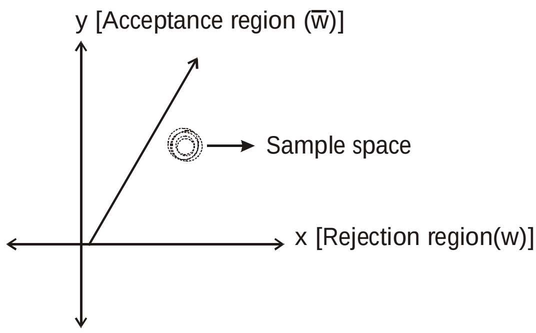
It is denoted by H_1 and it is defined by a statement against the Null hypothesis is called Alternative hypothesis.

Ex : A sample size $n = 30$ drawn from a population and it is found that the sample mean $(\bar{x}) = 15$. If we want to test the population mean is 10 (or) not ? then the alternative hypothesis is $H_1 : \mu \neq 10$.

(iv) Critical Region and Acceptance Region

Let x_1, x_2 ..... x_n are the samples of size 'n' drawn from the population the set of all points plotted between x-axis and y-axis; called sample space.

The division of the sample space in two exclusive regions i.e., Acceptance region and Rejection region [critical region]. The Acceptance region is denoted by '$\bar{w}$' and the Rejection region is denoted by 'w'



1.1.2 Two Types of Errors

Q2. Write short notes on Type-I and Type-II Error.

Ans : **(June-18)**

In Natural life we can see the following

1. Accept H_0, when it is true
2. Reject H_0, when it is false

3. Reject H_0, when it is true
4. Accept H_0, when it is false

In the above first two statements are correct statements and last two statements are Error statements.

TYPE -I ERROR : The Error statement reject H_0, when it is true is called Type -I Error.

TYPE -II ERROR : The Error statement Accept H_0, when it is false is called type - II Error

1.1.3 Level of Significance

Q3. Write a short notes on Level of Significance.

Ans :

The probability of Type- I Error is called level of significance which is denoted by 'α'

α = P [type - I Error]

α = P [Reject H_0/when it is true] [α = Alpha]

$\alpha = P\,[x \in w/\ H_0]$

$$\alpha = \int_{\omega} L_0 dx$$

β = P [type -II Error]

β = P [Accept H_0 / when it is false]

$\beta = P\,[x \in \bar{w}\ /H_1]$

$$\beta = \int_{\bar{w}} L_1 dx$$

Where L_0, L_1 are Liklihood functions

1.1.4 Power of a Test

Q4. Define the term Power of a test.

Ans : **(June-19, Feb.-21)**

For testing the hypothesis H_0 and H_1, the test with probability α and β of type -I and type -II. Errors respectively.

The quantity 1– β is called "power of the test"

The power of the test depends upon the difference between parameter value specified by H_0 , the actual value of parameter.

$1 - \beta$ = 1 – P [Type - II Error]

$1 - \beta$ = 1 – P [Accept H_0/ When it is false]

$$1 - \beta = 1 - P\left[x \in \bar{w} / H_1\right]$$

$$1 - \beta = 1 - \int_{\bar{w}} L_1 dx$$

$$1-\beta = \int_{w} L_1 dx$$

$$\begin{bmatrix} \therefore \bar{\omega}+\omega = 1 \\ \omega = 1-\bar{\omega} \end{bmatrix}$$

1.1.5 One and Two Tailed Tests

Q5. Write a short notes on One and two tailed tests.

Ans :

(i) One Tailed Test

In a statistical hypothesis test, if an alternative hypothesis is denoted by '>' (right-tailed) or '<' (left-tailed) symbol then this hypothesis test is referred to as a one-tailed test.

For example,

$H_1 : \mu > \mu_o$ refers to the right-tailed test and

$H_2 : \mu < \mu_o$ refers to the left-tailed test.

These two tests collectively are known as a single-tailed test.

When the alternative hypothesis contains a greater than (>) symbol then the critical region is completely located in the right tail of the distribution. But, when the alternative hypothesis contains a less than (<) symbol, the critical region is complex located in the left tail of the distribution. For example,

$H_1 : \mu \neq \mu_o$ (i.e., $\mu > \mu_o$ and $\mu < \mu_o$) refers to a two-tailed test.

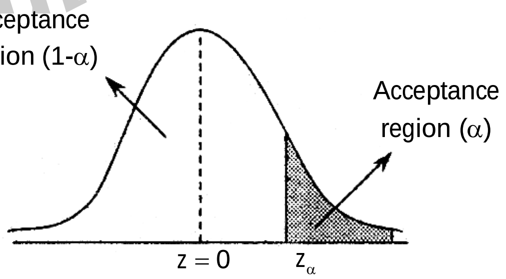


Fig (1) : Right - tailed Test Representation

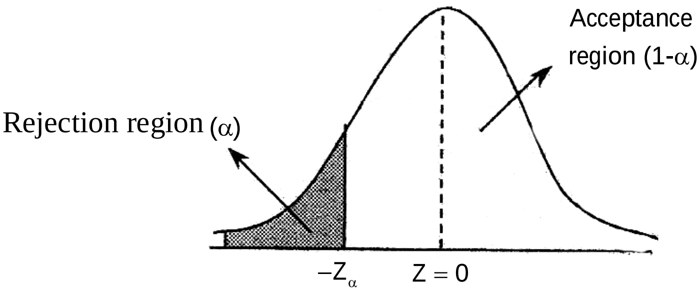


Fig (2) : Left-tailed Test Representation

(ii) Two Tailed Test

In a statistical hypothesis test, if an alternative hypothesis is represented by a 'not equal to $(\neq)$, symbol then this test is referred to as a two-tailed test because the critical region (which is divided into two parts) is placed in both (right-tailed and left-tailed) the tails of distribution.

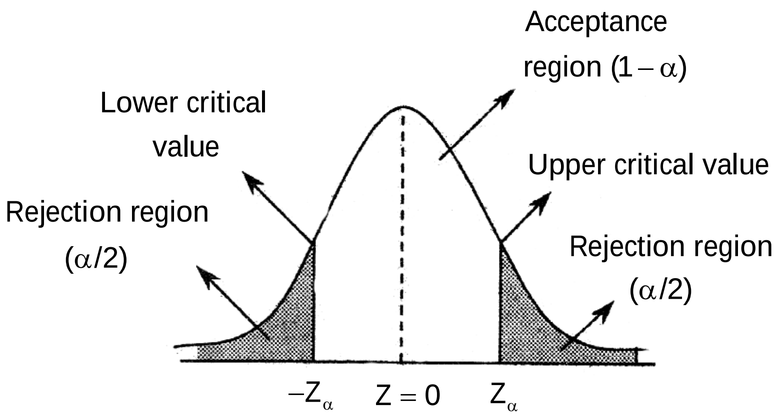


Fig (3) : Two-tailed Test Representation

PROBLEMS

1. Given Frequency function $f(x,\theta) = \frac{1}{\theta}$; $0 \le x \le \theta$ and that we have testing the Null Hypothesis $H_0 : \theta = 1$ and against alternative Hypothesis $H_1 : \theta = 2$ By means of a single observation x. Find the α and β, if critical region $x \ge 0.5$ and also we calculate the power of the Test.

Sol :

Given that,

$f(x, \theta) = \frac{1}{\theta}$; $0 \le x \le \theta$

and Null Hypothesis is $H_0 : \theta = 1$

Alternative Hypothesis is $H_1 : \theta = 2$

Critical region is $\omega : x \ge 0.5$

$\bar{\omega} : x < 0.5$

α = P [Type - I Error]

α = P[Reject H_0 / when it is true]

$\alpha = P[x \in \omega / H_0 : \theta = 1]$

$\alpha = P[\omega : x \ge 0.5 / H_0 : \theta = 1]$

$\alpha = \int_{\omega} L_0 dx$

$\alpha = \int_{0.5}^{\theta} f(x,\theta)dx$

$\alpha = \int_{0.5}^{\theta} \frac{1}{\theta}dx$ when $\theta = 1$

$\alpha = \int_{0.5}^{1} \frac{1}{1}dx$

$\alpha = [x]_{0.5}^{1}$

$\alpha = 1 - 0.5$

$\alpha = 0.5$

β = P[Type - II Error]

β = P[Accept H_0 /when it is False]

$\beta = P[x \in \bar{\omega} / H_1 : \theta = 2]$

$\beta = P[\bar{\omega} : x < 0.5 / H_1 : \theta = 2]$

$\beta = \int_{\bar{\omega}} L_1 dx$

$\beta = \int_{0}^{0.5} f(x,\theta)dx$

$\beta = \int_{0}^{0.5} \frac{1}{\theta}dx$ when $\theta = 2$

$\beta = \int_{0}^{0.5} \frac{1}{2}dx$

$\beta = \frac{1}{2}[x]_{0}^{0.5}$

$\beta = \frac{1}{2}[0.5 - 0]$

$\beta = \frac{1}{2}(0.5)$

$\beta = \frac{1}{4}$

Power of the test :

$1 - \beta = 1 - \frac{1}{4}$

$1 - \beta = \frac{3}{4}$

Conclusion :

$\alpha = \frac{1}{2}, \beta = \frac{1}{4}, 1 - \beta = \frac{3}{4}$

---

2. Given Frequency function $f(x,\theta) = \frac{1}{2\theta}$; $0 \le x \le \theta$ and given that the Null Hypothesis is $H_0 : \theta = 2$ and against alternative hypothesis $H_1 : \theta = 4$. Find the value of probability of Type -I and Type-II Error if critical region $x \ge 0.8$

Sol :

Given that,

$f(x,\theta) = \frac{1}{2\theta}$; $0 \le x \le \theta$

and Null Hypothesis $H_0 : \theta = 2$

Alternative Hypothesis $H_1 : \theta = 4$

Critical Region $\omega : x \ge 0.8$

$\bar{\omega} : x < 0.8$

α = P[Type - I Error]

α = P [Reject H_0 / when it is true]

$\alpha = P[x \in \omega / H_0 : \theta = 2]$

$\alpha = P[\omega : x \ge 0.8 / H_0 : \theta = 2]$

$\alpha = \int_{\omega} L_0 dx$

$\alpha = \int_{0.8}^{\theta} f(x,\theta)dx$

$$\alpha = \int_{0.8}^{\theta} \frac{1}{2\theta} dx \text{ when } \theta = 2$$

$$\alpha = \int_{0.8}^{2} \frac{1}{2(2)} dx$$

$$\alpha = \frac{1}{4}[x]_{0.8}^{2}$$

$$\alpha = \frac{1}{4}[2 - 0.8]$$

$$\alpha = \frac{1}{4}\left[\frac{6}{5}\right]$$

$$\alpha = \frac{\cancel{6}^{3}}{\cancel{20}_{10}}$$

$\alpha = 0.3$

β = P [Type -II Error]

β = P[Accept H_0/when it is false]

$$\beta = P[x \in \bar{\omega} / H_1 : \theta = 4]$$

$$\beta = P[\bar{\omega} : x < 0.8 / H_1 : \theta = 4]$$

$$\beta = \int_{\bar{w}} L_1 dx$$

$$\beta = \int_{0}^{0.8} f(x_1\theta) dx$$

$$\beta = \int_{0}^{0.8} \frac{1}{2\theta} dx \text{ when } \theta = 4$$

$$\beta = \int_{0}^{0.8} \frac{1}{2(4)} dx$$

$$\beta = \frac{1}{8}[x]_{0}^{0.8}$$

$$\beta = \frac{1}{8}[0.8 - 0]$$

$$\beta = \frac{1}{8}\left[\frac{8}{10}\right]$$

$$\beta = \frac{1}{10}$$

$\beta = 0.1$

Conclusion :

$\alpha = 0.3, \beta = 0.1$

3. Given Frequency function $f(x,\theta) = \frac{1}{\theta} e^{-x/\theta}$; $0 \le x \le \infty$ for testing the Null Hypothesis $H_0 : \theta = 3$ and alternative hypothesis $H_1 : \theta = 1$. Find the probability of Type -I and Type-II Error and power of the curve critical region $x \le 2$.

Sol :

Given that,

$$f(x,\theta) = \frac{1}{\theta} e^{-x/\theta}; \; 0 \le x \le \infty$$

and Null Hypothesis $H_0 : \theta = 3$

Alternative Hypothesis $H_1 : \theta = 1$

Critical region $\omega : x \le 2$

$\bar{w} : x > 2$

α = P[Type - I Error]

α = P[Reject H_0/ when it is true]

$$\alpha = P[x \in \omega / H_0 : \theta = 3]$$

$$\alpha = P[\omega : x \le 2 / H_0 : \theta = 3]$$

$$\alpha = \int_{\omega} L_0 dx$$

$$\alpha = \int_{0}^{2} f(x,\theta) dx$$

$$\alpha = \int_{0}^{2} \frac{1}{\theta} e^{-x/\theta} dx \quad \text{where } \theta = 3$$

$$\alpha = \int_0^2 \frac{1}{3} e^{-x/3} dx$$

$$\alpha = \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_0^2$$

$$\alpha = \frac{-\not{3}}{\not{3}}\left[e^{-2/3} - e^{-0/3}\right]$$

α = –1 [$e^{-2/3}$ –1]

α = 1– $e^{-2/3}$

α = 1 – 0.513

α = 0.4870

β = P [Type - II Error]

β = P[Accept H_0/when it is false]

$$\beta = P\left[x \in \overline{\omega} / H_1 : \theta = 1\right]$$

$$\beta = P\left[\overline{\omega} : x > 2 / H_1 : \theta = 1\right]$$

$$\beta = \int_{\overline{w}} L_1 dx$$

$$\beta = \int_2^\infty f(x, \theta) dx$$

$$\beta = \int_2^\infty \frac{1}{\theta} e^{-x/\theta} dx \quad \text{where } \theta = 1$$

$$\beta = \int_2^\infty \frac{1}{1} e^{-x} dx$$

$$\beta = -\left[e^{-x}\right]_2^\infty$$

$$\beta = -\left[e^{-\infty} - e^{-2}\right]$$

$$\beta = e^{-2} - e^{-\infty}$$

β = 0.1353 – 0

β = 0.135

Power of the test :

1 – β = 1 – 0.135

1 – β = 0.8650

Conclusion :

α = 0.4870, β = 0.135, 1 – β = 0.8650

4. One of the probability coin will fall Head in a single test in order to test $H_0 : P = \frac{1}{2}$ and $H_1 : P = \frac{3}{4}$. The coin is tassed 5 times and H_0 is Rejected More than 3 heads and obtained probability of Type-I and Type - II Error and power of the test.

Sol :

Given that,

$$H_0 : P = \frac{1}{2}$$

$$H_1 : P = \frac{3}{4}$$

It is follows to Binomial distribution then the probability mass function of Binomial distribution is P(x=x) = ${}^n c_x\, p^x q^{n-x}$; x = 0, 1, 2, ...n and also given that

critical region (ω) = x > 3

Acceptance Region ($\overline{\omega}$) = x $\leq$ 3 and n = 5

α = P[Type -I Error]

α = P [Reject H_0 / when it is true]

$$\alpha = P\left[x \in \omega / H_0 : P = \frac{1}{2}\right]$$

$$\alpha = P\left[\omega : x > 3 / H_0 : P = \frac{1}{2}\right]$$

α = P[x = 4 and x = 5]

$$\alpha = {}^5c_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4} + {}^5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$$

$$\alpha = {}^5c_4\left(\frac{1}{2}\right)^5 + {}^5c_5\left(\frac{1}{2}\right)^5$$

α = 5(0.0313) + (1) (0.0313)

$\alpha = 0.1565 + 0.0313$

$\alpha = 0.1878$

β = P[Type - II Error]

β = P[Accept H_0/when it is false]

$$\beta = P\left[x \in \overline{\omega} / H_1 : P = \frac{3}{4}\right]$$

$$\beta = P\left[\overline{\omega} : x \le 3 / H_1 : P = \frac{3}{4}\right]$$

$$\beta = [P(x=0) + P(x = 1) + P(x = 2) + P(x = 3)]$$

$$\beta = {}^5c_0\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^{5-0} + {}^5c_1\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^{5-1} + {}^5c_2\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^{5-2} + {}^5c_3\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^{5-3}$$

β = (1) (1)(0.0010) + 5(0.7500) (0.0039) + 10(0.5625) (0.0156) + 10(0.4219) (0.0625)

β = 0.0010 + 0.0146 + 0.0878 + 0.2637

β = 0. 3671

Power of the Test :

$1 - \beta = 1 - 0.3671$

$1 - \beta = 0.6329$

Conclusion :

$\alpha = 0.1878$, $\beta = 0.3671$, $1 - \beta = 0.6329$

5. If $x \le 2$ critical region for testing $\theta = 6$ against the alternative hypothesis $\theta = 4$ on the base of simple observation x in the population $f(x,\theta) = \theta.e^{-\theta x}$; $0 \le x \le \infty$. Find the power of the test.

Sol :

Given that, $H_0 : \theta = 6$

$H_1 : \theta = 4$

It follows the exponential distribution

$f(x,\theta) = \theta.e^{-\theta x}$; $0 \le x \le \infty$

Also given that,

Critical region (ω) = x > 2

Acceptance region ($\overline{\omega}$) = $x \le 2$

α = P[Type-I Error]

α = P[Reject H_0/when it is true]

$$\alpha = P\left[x \in \omega / H_0 : \theta = 6\right]$$

$$\alpha = P\left[\omega : x > 2 / H_0 : \theta = 6\right]$$

$$\alpha = \int_{\omega} L_0 dx$$

$$\alpha = \int_2^{\infty} f(x,\theta)dx$$

$$\alpha = \int_2^{\infty} \theta.e^{-\theta x}.dx$$

$$\alpha = \int_2^{\infty} 6.e^{-6x}dx$$

$$\alpha = 6\int_2^{\infty} e^{-6x}dx$$

$$\alpha = 6\left[\frac{e^{-6x}}{-6}\right]_2^{\infty}$$

$$\alpha = -\left[e^{-6x}\right]_2^{\infty}$$

$$\alpha = -[e^{-6\infty} - e^{-6(2)}]$$

$$\alpha = -[e^{-6\infty} - e^{-12}]$$

$$\alpha = e^{-12} - e^{-\infty}$$

$$\alpha = e^{-12}$$

β = P [Type - II Error]

β = P [Accept H_0/ when it is false]

$$\beta = P\left[x \in \overline{\omega} / H_1 : \theta = 4\right]$$

$$\beta = P\left[\overline{\omega} : x \le 2 / H_1 : \theta = 4\right]$$

$$\beta = \int_{\overline{w}} L_1 dx$$

$$\beta = \int_0^2 \theta . e^{-\theta x} dx$$

$$\beta = \theta \int_0^2 e^{-\theta x} dx$$

$$\beta = 4 \int_0^2 e^{-4x} dx$$

$$\beta = \not{4} \left[\frac{e^{-4x}}{-\not{4}} \right]_0^2$$

$$\beta = -\left[e^{-4x} \right]_0^2$$

$\beta = -[e^{-4(2)} - e^{-4(0)}]$

$\beta = -[e^{-8} - e^{-0}]$

$\beta = e^{+0} - e^{-8}$

$\beta = 1 - e^{-8}$

Power of the Test :

$1 - \beta = 1 - (1 - e^{-8})$

$1 - \beta = 1 - 1 + e^{-8}$

$1 - \beta = e^{-8}$

Conclusion :

$\alpha = e^{-12}$

$\beta = 1 - e^{-8}$

$1 - \beta = e^{-8}$

1.1.6 Test Function

Q6. Explain briefly about test function.

Ans : **(June-19)**

Most Powerful test [MP - Test]

A test procedure is said to be most powerful test for testing the null hypothesis $H_0 : \theta = \theta_0$ against simple Alternative Hypothesis $H_1 : \theta = \theta_1$ is called most powerful test.

Most Powerful Critical Region :

The critical region (w) is said to be most powerful critical region with size 'α' for testing Null hypothesis $H_0 : \theta = \theta_0$ and against simple Alternative Hypothesis $H_1 : \theta = \theta_1$ and satisfy the following conditions $P\{x \in \omega / H_1\} \geq P\{x \in \omega_1 / H_1\}$

ω	ω_1
$\alpha = P\{x \in \omega / H_0\}$	$\alpha_1 = P\{x \in \omega_1 / H_0\}$
$\beta = P\{x \in \bar{\omega} / H_1\}$	$\beta_1 = P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = 1-P\{x \in \bar{\omega} / H_1\}$	$1-\beta_1 = 1-P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = P\{x \in \omega / H_1\}$	$1-\beta_1 = P\{x \in \omega_1 / H_1\}$

Uniformly most powerful test (UMP - Test)

A test procedure is said to be uniformly most powerful test for testing the Null Hypothesis $H_0 : \theta = \theta_0$ and against composite Alternative Hypothesis $H_1 : \theta \neq \theta_1$ is called uniformly most powerful test.

Uniformly most powerful critical region:

The critical region (ω) is said to be uniformly most powerful critical region with size 'α' for testing Null hypothesis $H_0 : \theta = \theta_0$ and against composite Alternative Hypothesis $H_0 : \theta \neq \theta_1$ and satisfy the following condition

$P\{x \in \omega / H_1\} \geq P\{x \in \omega_1 / H_1\}$

ω	ω_1
$\alpha = P\{x \in \omega / H_0\}$	$\alpha_1 = P\{x \in \omega_1 / H_0\}$
$\beta = P\{x \in \bar{\omega} / H_1\}$	$\beta_1 = P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = 1-P\{x \in \bar{\omega} / H_1\}$	$1-\beta_1 = 1= P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = P\{x \in \omega / H_1\}$	$1-\beta_1 = P\{x \in \omega_1 / H_1\}$

1.1.7 Non-Randomized and Randomized

Q7. Write a short note on Non-Randomized and Randomized test function.

(OR)

Write short notes on test for randomness .

Ans : **(Feb.-21, June-18)**

(i) Randomized Test Function :

A function $\psi(.)$ is said to be randomized test function if its sample space values lie in the close interval [0,1] i.e., $0 \leq \psi(x) \leq 1$ for all x. The exception and rejection of null Hypothesis 'H_0' is decided with probabilities $\psi(x)$ and $1 - \psi(x)$. In case that value of a random variable x is equal to the observed value x, then two possible outcomes say p and $\bar{p}$ are considered with the above probabilities. The null hypothesis H_0 is rejected if P occurs in the experiment otherwise, H_0 is accepted. Here $\psi(x)$ is called critical function An example of randomized test function is,

$$\psi(x) = \begin{cases} 1 \ ; \quad \text{if } x = 3 \\ 1/8; \text{if } x = 2 \\ 0; \quad \text{otherwise} \end{cases}$$

(ii) Non-randomized Test function :

A function is said to be non-randomized test function if all sample values lie in the open interval {0,1} i.e., either 'O' or '1' when the null hypothesis 'H_0' is rejected , the sample is divided into two sets say Q and compliment of Q i.e., Q' in such a way that $Q \cup Q'$ = sample space and $Q \cap Q' = \phi$. Here Q is called rejection region because the 'H_0' is rejected it if x falls into Q.

A non-randomized test function can be written as,

$$\psi(x) = \begin{Bmatrix} 0; \text{ if } x \in Q \\ 1; \text{ otherwise} \end{Bmatrix}$$

1.2 Statement and Proof of Neyman-Pearson's Fundamental Lemma for Randomized Tests

Q8. State and prove the Neymann Pearson's Lemma.

Ans : **(Feb.-21, June-19, June-18)**

Statement : Let k > 0 be a constant and w is a critical region with size 'a' such that

$$w = \left\{ x \in s, \frac{f(x_1\theta_1)}{f(x_1\theta_0)} \geq k \right\} = \left\{ x \in s, \frac{L_1}{L_0 \geq k} \right\} \text{ and } \overline{w} = \left\{ x \in s, \frac{f(x_1\theta_1)}{f(x_1\theta_0)} < k \right\} = \left\{ x \in s, \frac{L_1}{L_0} < k \right\}$$

Where L_0, L_1 are Likilhood functions of x_1 , x_2 ....x_n under the Null hypothesis (H_0) and Alternative Hypothesis (H_1) then ω is the most powerful critical region for testing simple Null hypothesis $H_0 : \theta = \theta_0$ and against simple Alternative hypothesis $H_1 : \theta = \theta_1$.

Proof :

Given that,

$$\omega = \left\{ x \in s \frac{L_1}{L_0} \geq k \right\} \qquad ...... (1)$$

$$\overline{\omega} = \left\{ x \in s_1 \frac{L_1}{L_0} < k \right\} \qquad .... (2)$$

Let 'ω' be the critical region with size 'α'

$\alpha = P[x \in \omega / H_0]$

$$\alpha = \int_\omega L_0 dx$$

$$\dot{\beta} = P\left[x \in \overline{\omega} / H_1\right]$$

$\beta = \int_{\bar{\omega}} L_1 dx$ .... (4)

power of the test :

$1 - \beta = 1 - P\left[x \in \bar{\omega} / H_1\right]$

$1 - \beta = 1 - \int_{\bar{\omega}} L_1 dx$

$1 - \beta = \int_{\omega} L_1 dx$ .... (5)

And 'ω_1' be the Another critical region with size α_1

$\alpha_1 = P\left[x \in \omega_1 / H_0\right]$

$\alpha_1 = \int_{\omega_1} L_0 dx$ .... (6)

$\beta_1 = P\left[x \in \bar{\omega}_1 / H_1\right]$

$\beta_1 = \int_{\bar{\omega}_1} L_1 dx$ .... (7)

Power of the test :

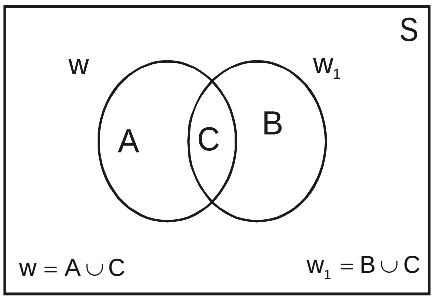
$1 - \beta_1 = 1 - P\left[x \in \bar{\omega}_1 / H_1\right]$

$1 - \beta_1 = 1 - \int_{\bar{w}_1} L_1 dx$

$1 - \beta_1 = \int_{\bar{\omega}_1} L_1 dx$ ...... (8)

Now, we show that $1 - \beta \geq 1 - \beta_1$

Let us assume that ω, ω_1 be the two regions i.e., $\omega = A \cup C, \omega_1 = B \cup C$



Now, Let $\alpha \leq \alpha_1$

$\int_{\omega} L_0 dx \geq \int_{\omega_1} L_0 dx$

$\int_{A \cup C} L_0 dx \geq \int_{B \cup C} L_0 dx$

$\int_{A} L_0 dx + \int_{C} L_0 dx \geq \int_{B} L_0 dx + \int_{C} L_0 dx$

$\int_{A} L_0 dx \geq \int_{B} L_0 dx$ ....(9)

From eq^n (1)

$\omega = \left\{ x \in s_1 \frac{L_1}{L_0} \geq k \right\}$

$\int_{\omega} L_1 dx \geq K . \int_{\omega} L_0 dx$

$\int_{A} L_1 dx \geq K \int_{A} L_0 dx$

$\int_{A} L_1 dx \geq K \int_{B} L_0 dx$ [From 9]

$K \int_{B} L_0 dx \leq \int_{A} L_1 dx$ ...... (10)

From eq^n (2)

$\bar{\omega} = \left\{ x \in s \frac{L_1}{L_0} < K \right\}$

$\int_{\bar{\omega}} L_1 dx < k \int_{\bar{\omega}} L_0 dx$

$\int_{B} L_1 dx < K \int_{B} L_0 dx$

$\int_{B} L_1 dx \leq \int_{A} L_1 dx$ .... (11)

[From 10]

Adding $\int_{C} L_1 dx$ on B.S

$$\int_B L_1 dx + \int_C L_1 dx \leq \int_A L_1 dx + \int_C L_1 dx$$

$$\int_{B\cup C} L_1 dx \leq \int_{A\cup C} L_1 dx$$

$$\int_{w_1} L_1 dx \leq \int_{w} L_1 dx$$

$1 - \beta_1 \leq 1 - \beta$

$1 - \beta \geq 1 - \beta_1$

$\therefore$ W is the most powerful critical region hence proved.

PROBLEMS

6. Find the best critical region for testing $H_0 : \lambda = \frac{2}{3}$ and against $H_1 : \lambda = \frac{4}{3}$ of the poisson distribution.

Sol :

Given that,

$$H_0 : \lambda = \frac{2}{3},\ H_1 : \lambda = \frac{4}{3}$$

and the probability mass function of poisson distribution is $P(x = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i, \lambda)$$

$$L = P(x_1, \lambda) . P(x_2, \lambda) .....P(x_n, \lambda)$$

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} . \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} ..... \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

By Neymann pearson's Lemma

$$\frac{L_1}{L_0} \geq K \quad ...... (1)$$

Here, $H_0 : \lambda = \frac{2}{3} \Rightarrow L_0 = \frac{e^{-n\left(\frac{2}{3}\right)}\left(\frac{2}{3}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$ ...(2)

$$H_1 : \lambda = \frac{4}{3} \Rightarrow L_1 = \frac{e^{-n\left(\frac{4}{3}\right)}\left(\frac{4}{3}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \quad ...(3)$$

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{\dfrac{e^{-\frac{4n}{3}}\left(\frac{4}{3}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}}{\dfrac{e^{-\frac{2n}{3}}\left(\frac{2}{3}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}} \geq K$$

$$\frac{e^{\frac{-4n}{3}}(2)^{\sum_{i=1}^{n} x_i}\left(\frac{2}{3}\right)^{\sum_{i=1}^{n} x_i}}{e^{\frac{-2n}{3}}\left(\frac{2}{3}\right)^{\sum_{i=1}^{n} x_i}} \geq K$$

$$e^{-\frac{4n}{3}} . e^{\frac{2n}{3}} (2)^{\sum_{i=1}^{n} x_i} \geq K$$

$$e^{-\frac{2n}{3}} (2)^{\sum_{i=1}^{n} xi} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[e^{-\frac{2n}{3}}(2)^{\sum_{i=1}^{n} x_i}\right] \geq \log K$$

$$\text{Log } e^{-\frac{2n}{3}} + \text{Log } (2)^{\sum_{i=1}^{n} x_i} \geq \text{Log } K$$

$$\frac{-2n}{3}\text{Log } e + \sum_{i=1}^{n} x_i \text{ Log } 2 \geq \text{Log } K$$

$$\sum_{i=1}^{n} x_i \geq \frac{\text{Log } k + \frac{2n}{3}\text{Log } e}{\text{Log} 2}$$

Divide by 'n'

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{\text{Log}\left(\frac{k}{n}\right) + \frac{2}{3}\text{Loge}}{\text{Log}\left(\frac{2}{n}\right)}$$

$$\bar{x} \geq \frac{\text{Log}\left(\frac{k}{n}\right) + \frac{2}{3}\text{Log } e}{\text{Log}\left(\frac{2}{n}\right)}$$

$$\text{Where } c = \frac{\text{Log}\left(\frac{k}{n}\right) + \frac{2}{3}\text{Log } e}{\text{Log}\left(\frac{2}{n}\right)}$$

$\therefore \quad \bar{x} \geq c$

The best critical region is $\omega = \{x : \bar{x} \geq c\}$

7. Let P the probability that a coin will fall head in a single toss in order to test $H_o : p = \frac{1}{2}$ against $H_1 : p = \frac{3}{4}$. The coin is tossed 5 times and H_o is rejected if more than 3 heads are obtained. Find the probability of Type I Error and power of the test.

Sol : **(June-18)**

Given that

$H_0 = P = \frac{1}{2}, H_1 : P = \frac{3}{4}$

And the probability mass function of Binomial Distribution is given by

$P(x = x) = {}^nC_x\, p^x\, q^{n-x}$

No of times a coin is tossed n = 5

Let x be the total no of heads that appear while tossing n times.

Then we have

$P(x = x) = {}^nc_x\, p^x (1-p)^{n-x}$

$[\therefore p + q = 1, q = 1 - p]$

$P(x = x) = {}^5c_x\, p^x (1-p)^{5-x}$

The critical region will be

$w = \{x : x \geq 4\}$

$\bar{w} = \{x : x \leq 3\}$

The probability of type-I error is given by

$\alpha = P(x \in w / H_o)$

$\alpha = P(x \geq 4 / P = \frac{1}{2}) \quad [\therefore P = \frac{1}{2} \text{ under } H_0]$

$$\alpha = {}^5c_4\left(\frac{1}{2}\right)^4\left(1-\frac{1}{2}\right)^{5-4} + {}^5c_5\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^{5-5}$$

$$\alpha = 5\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + 1\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$$

$$\alpha = 5\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^5 \quad \text{............ (1)}$$

$$\alpha = 5\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5$$

$$\alpha = 6\left(\frac{1}{2}\right)^5$$

$$\alpha = 6\left(\frac{1}{32}\right) = \frac{6}{32} = \frac{3}{16}$$

$\therefore$ Probability of type-I Error $(\alpha) = \frac{3}{16}$

let probability of type - II Error be,

$\beta = P(x \in \bar{w} / H_1)$

$$\beta = 1 - P(x \in w / H_1)$$

$$\beta = 1 - \left[P\left[x = 4 / P = \frac{3}{4}\right]\right] + P\left[x = 5 / P = \frac{3}{4}\right]$$

$$\left[\therefore P = \frac{3}{4} \text{ under } H_1\right]$$

$$\beta = 1 - \left[{}^5c_4\left(\frac{3}{4}\right)^4\left(1-\frac{3}{4}\right)^4 + {}^5c_5\left(\frac{3}{4}\right)^5\left(1-\frac{3}{4}\right)^{5-5}\right]$$

$$\beta = 1 - \left[{}^5c_4\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^1 + {}^5c_5\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)^0\right]$$

$\beta = 1 -$ [5(0.3164 (0.25) + (1) (0.2373)(1)]

$\beta = 1 -$[0.3935 + 0.2373)

$\Rightarrow \beta = 1 -$ [0.6328]

Probability of Type - II Error $(\beta) = \frac{459}{1250}$

8. Find the Best critical region for testing $H_0: \lambda = 2$ and against $H_1: \lambda = 4$ of the poisson distribution.

Sol :

Given that,

$H_0: \lambda = 2$, $H_1: \lambda = 4$

and probability mass function of poisson distribution is $P(x = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i, \lambda)$$

$$L = P(x_i, \lambda).P(x_2, \lambda) \ldots\ldots P(x_n,\lambda)$$

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}.\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\ldots\ldots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}$$

By Neymann pearson's Lemma

$$\frac{L_1}{L_2} \geq K \quad \ldots\ldots.(1)$$

Here, $H_0 : \lambda = 2$

$$\Rightarrow L_0 = \frac{e^{-n(2)}(2)^{\sum_{i=1}x_i}}{\prod_{i=1}^{n}x_i!} \ldots\ldots.. (2)$$

$$H_1 : \lambda = 4 \Rightarrow L_1 = \frac{e^{-n(4)}(4)^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!} \ldots..(3)$$

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{\dfrac{e^{-4n}(4)^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}}{\dfrac{e^{-2n}(2)^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}} \geq K$$

$$\frac{e^{-4n}(2)^{\sum_{i=1}^{n}x_i}(2)^{\sum_{i=1}^{n}x_i}}{e^{-2n}(2)^{\sum_{i=1}^{n}x_i}} \geq K$$

$$e^{-4n}e^{2n}(2)^{\sum_{i=1}^{n}x_i} \geq K$$

$$e^{-2n}(2)^{\sum_{i=1}^{n}x_i} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[e^{-2n}(2)^{\sum_{i=1}^{n}x_i}\right] \geq \log K$$

$$\text{Log } e^{-2n} + \text{Log } (2)^{\sum_{i=1}^{n}x_i} \geq \text{Log } K$$

$$-2n \text{ Log } e + \sum_{i=1}^{n}x_i \log 2 \geq \text{Log } K$$

$$\sum_{i=1}^{n} x_i \text{ Log } 2 \geq \text{Log K} + 2n \text{ Log e}$$

$$\sum_{i=1}^{n} x_i \geq \frac{\text{Log k} + 2n\text{ Loge}}{\text{Log}2}$$

Divide by 'n'

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{\text{Log}\left(\frac{k}{n}\right) + 2\text{ Log e}}{\text{Log}\left(\frac{2}{n}\right)}$$

$$\bar{x} \geq \frac{\text{Log}\left(\frac{k}{n}\right) + 2\text{ Log e}}{\text{Log}\left(\frac{2}{n}\right)}$$

$\therefore\ \bar{x} \geq C$

The best critical region is $\omega = \{x : \bar{x} \geq c\}$

9. Find the best critical region for testing $H_0 : \theta = \frac{3}{2}, H_1 : \theta = \frac{4}{3}$ of the Exponential distribution.

Sol :

Given that,

$H_0 : \theta = \frac{3}{2}, H_1 : \theta = \frac{4}{3}$

The probability density function of Exponential distribution is $f(x) = \theta.e^{-\theta x}$

Then the Liklihood function is

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

$$L = f(x_1,\theta).f(x_2,\theta) \text{ ....} f(x_n, \theta)$$

$$L = \theta. e^{-\theta x_1}. \theta. e^{-\theta x_2} \text{ ....} \theta\ e^{-\theta x_n}$$

$$L = \theta^n e^{-\theta(x_1 + x_2 + ...+ x_n)}$$

$$L = \theta^n . e^{-\theta \sum_{i=1}^{n} x_i}$$

By Neymam pearson's lemma

$$\frac{L_1}{L_0} \geq K \quad \text{....... (1)}$$

Here, $H_0 : \theta = \frac{3}{2}$

$$\Rightarrow L_0 = \left(\frac{3}{2}\right)^n e^{-\left(\frac{3}{2}\right)\sum_{i=1}^{n} x_i} \quad \text{........ (2)}$$

$$H_1 : \theta = \frac{4}{3} \Rightarrow L_1 = \left(\frac{4}{3}\right)^n e^{-\left(\frac{4}{3}\right)\sum_{i=1}^{n} x_i} \quad \text{.... (3)}$$

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{\left(\frac{4}{3}\right)^n e^{-\left(\frac{4}{3}\right)\sum_{i=1}^{n} x_i}}{\left(\frac{3}{2}\right)^n e^{-\left(\frac{3}{2}\right)\sum_{i=1}^{n} x_i}} \geq K$$

$$\left(\frac{4}{3}\right)^n \left(\frac{2}{3}\right)^n e^{-\left(\frac{4}{3}\right)\sum_{i=1}^{n} x_i} . e^{\left(\frac{3}{2}\right)\sum_{i=1}^{n} x_i} \geq K$$

$$\left(\frac{8}{9}\right)^n e^{\sum_{i=1}^{n} x_i \left(\frac{3}{2} - \frac{4}{3}\right)} \geq K \quad [\therefore a^n b^n = (ab)^n]$$

$$\left(\frac{8}{9}\right)^n e^{\sum_{i=1}^{n} x_i \left(\frac{9-8}{6}\right)} \geq K$$

$$\left(\frac{8}{9}\right)^n e^{\sum_{i=1}^{n} x_i \left(\frac{1}{6}\right)} \geq K$$

Takine 'Log' on b.s

$$\text{Log}\left[\left(\frac{8}{9}\right)^n e^{\sum_{i=1}^{n} x_i \left(\frac{1}{6}\right)}\right] \geq \text{Log K}$$

$$\text{Log}\left(\frac{8}{9}\right)^n + \text{Log}\, e^{\sum_{i=1}^{n} x_i \left(\frac{1}{6}\right)} \geq \text{log K}$$

$$n.\text{Log}\left(\frac{8}{9}\right) + \frac{1}{6}\sum_{i=1}^{n} x_i \text{ Loge} \geq \text{Log K}$$

$$\frac{1}{6}\sum_{i=1}^{n} x_i \text{ Log e} \geq \text{Log K} - n \text{ Log}\left(\frac{8}{9}\right)$$

$$\sum_{i=1}^{n} x_i \text{ Loge} \geq 6\left[\text{Log k} - n\log\left(\frac{8}{9}\right)\right]$$

$$\sum_{i=1}^{n} x_i \geq \frac{6\left[\text{Log k} - n \text{ Log}\left(\frac{8}{9}\right)\right]}{\text{Log e}}$$

Divide by 'n'

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{6\left[\log\left(\frac{k}{n}\right) - \text{Log}\left(\frac{8}{9}\right)\right]}{\text{Log}\left(\frac{e}{n}\right)}$$

$$\bar{x} \geq \frac{6\left[\text{Log}\left(\frac{k}{n}\right) - \text{Log}\left(\frac{8}{9}\right)\right]}{\text{Log}\left(\frac{e}{n}\right)}$$

Where $c = \frac{6\left[\text{Log}\left(\frac{k}{n}\right) - \text{Log}\left(\frac{8}{9}\right)\right]}{\text{Log}\left(\frac{e}{n}\right)}$

$\therefore \ \bar{x} \geq c$

The best critical region is $\omega = \{x : \bar{x} \geq c\}$

10. Find the best critical region for testing $H_0 : \theta = \theta_0$ and against $H_1 : \theta = \theta_1$ ($> \theta_0$).

for a random variables $x_1, x_2 \ldots x_n$ from Exponential distribution.

Sol :

Given that,

$H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1 (> \theta_0)$

The probability density function of Exponential distribution is $f(x) = \theta . e^{-\theta x}$

Then the Liklihood function is

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

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

$$L = \theta^n e^{-\theta x_1} . \ \theta e^{-\theta x_2} \ldots\ldots \theta . e^{-\theta x_n}$$

$$L = \theta^n \ e^{-\theta(x_1 + x_2 + \ldots\ldots + x_n)}$$

$$L = \theta^n \ e^{-\theta \sum_{i=1}^{n} xi}$$

By Neymann pearson's lemma

$$\frac{L_1}{L_0} \geq K \qquad \ldots\ldots.(1)$$

Here, $H_0 : \theta = \theta_0 \Rightarrow L_0 = (\theta_0)^n \ e^{-\theta_0 \sum_{i=1}^{n} x_i}$ .... (2)

$H_0 : \theta = \theta_1 \Rightarrow L_1 = (\theta_1)^n \ e^{-\theta_1 \sum_{i=1}^{n} x_i}$ ..... (3)

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{(\theta_1)^n e^{-\theta_1} \sum_{i=1}^{n} x_i}{(\theta_0)^n e^{-\theta_0} \sum_{i=1}^{n} x_i} \geq K$$

$$\frac{(\theta_1)^n}{(\theta_0)^n} . \frac{e^{-\theta_1 \sum_{i=1}^{n} x_i}}{e^{-\theta_0 \sum_{i=1}^{n} x_i}} \geq K$$

$$\left(\frac{\theta_1}{\theta_0}\right)^n \ e^{-\theta_1 \sum_{i=1}^{n} x_i} \cdot e^{\theta_0 \sum_{i=1}^{n} x_i} \geq K$$

$$\left(\frac{\theta_1}{\theta_0}\right)^n \ e^{\sum_{i=1}^{n} x_i (\theta_0 - \theta_1)} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[\left(\frac{\theta_1}{\theta_0}\right)^n e^{\sum_{i=1}^{n} x_i (\theta_0 - \theta_1)}\right] \geq \text{Log K}$$

$$\text{Log}\left(\frac{\theta_1}{\theta_0}\right)^n + \text{Log } e^{\sum_{i=1}^{n} x_i(\theta_0-\theta_1)} \geq \text{Log K}$$

$$n\,\text{Log}\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{n} x_i\,(\theta_0 - \theta_1)\text{ Log e} \geq \text{Log K}$$

$$\sum_{i=1}^{n} x_i\,(\theta_0 - \theta_1)\text{ Log e} \geq \text{Log k} - n\,\text{Log}\left(\frac{\theta_1}{\theta_0}\right)$$

$$\sum_{i=1}^{n} x_i \geq \frac{\text{Log k} - n\,\text{Log}\left(\frac{\theta_1}{\theta_0}\right)}{\text{Log } e^{(\theta_0-\theta_1)}}$$

Divide by 'n'

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{\text{Log}\left(\frac{k}{n}\right) - \text{Log}\left(\frac{\theta_1}{\theta_0}\right)}{\text{Loge}\left(\theta_0 - \theta_1\right)/n}$$

$$\text{Where } c = \frac{\text{Log}\left(\frac{k}{n}\right) - \text{Log}\left(\frac{\theta_1}{\theta_0}\right)}{\text{Loge}\left(\theta_0 - \theta_1\right)/n}$$

The best critical region is $w = \{x : \bar{x} \geq c\}$

11. Find the Best Critical region for testing $H_0 : \lambda = \lambda_0$ and against $H_1 : \lambda = \lambda_1$ for the poisson distribution.

Sol : **(Feb.-21)**

Given that,

$H_0 : \lambda = \lambda_0$, $H_1 : \lambda = \lambda_1$

The probability mass function of poisson distribution is $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i, \lambda)$$

$$L = P(x_1, \lambda)\,P(x_2, \lambda)......P(x_n, \lambda)$$

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} ...... \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L = \frac{e^{-\lambda(n)}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

By Neymann pearson's Lemma

$$\frac{L_1}{L_0} \geq K \quad .... (1)$$

Here $H_0 : \lambda = \lambda_0$

$$\Rightarrow L_0 = \frac{e^{-n(\lambda_0)}\left(\lambda_0\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \quad .... (2)$$

$H_1 : \lambda = \lambda_1$

$$\Rightarrow L_1 = \frac{e^{-n(\lambda_1)}\left(\lambda_1\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \quad ...... (3)$$

Eq^n(2) & (3) sub in Eq^n (1)

$$\frac{\dfrac{e^{-n\lambda_1}\left(\lambda_1\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}}{\dfrac{e^{-n\lambda_0}\left(\lambda_0\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}} \geq K$$

$$\frac{e^{-n\lambda_1}\left(\lambda_1\right)^{\sum_{i=1}^{n} x_i}}{e^{-n\lambda_0}\left(\lambda_0\right)^{\sum_{i=1}^{n} x_i}} \geq K$$

$$e^{-n\lambda_1}e^{n\lambda_0}\left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n} x_i} \geq K$$

$$e^{n(\lambda_0-\lambda_1)}\left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n}x_i} \geq K$$

Taking ' Log' on b.s

$$\text{Log}\left[e^{n(\lambda_0-\lambda_1)}\left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n}x_i}\right] \geq \text{Log } K$$

$$\text{Log } e^{n(\lambda_0-\lambda_1)} + \text{Log}\left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^{n}x_i} \geq \text{Log } K$$

$$n(\lambda_0-\lambda_1)\text{ Log } e + \sum_{i=1}^{n}x_i \text{ Log}\left(\frac{\lambda_1}{\lambda_0}\right) \geq \text{Log } K$$

$$\sum_{i=1}^{n}x_i\left(\text{Log}\left(\frac{\lambda_1}{\lambda_0}\right)\right) \geq \text{Log } K - n(\lambda_0-\lambda_1)\text{ Log } e$$

$$\sum_{i=1}^{n}x_i \geq \frac{\text{Log } k - n(\lambda_0-\lambda_1)\text{Log } e}{\text{Log}\left(\frac{\lambda_1}{\lambda_0}\right)}$$

Divide by 'n' on b.s

$$\frac{\sum_{i=1}^{n}x_i}{n} \geq \frac{\text{Log}\left(k/n\right) - \text{Log } e(\lambda_0-\lambda_1)}{\text{Log}\left(\frac{\lambda_1/\lambda_0}{n}\right)}$$

$$\bar{x} \geq \frac{\text{Log}\left(k/n\right) - (\lambda_0-\lambda_1)\text{Log } e}{\text{Log}\left(\frac{\lambda_1/\lambda_0}{n}\right)}$$

$$\text{Where } C = \frac{\text{Log}\left(k/n\right) - (\lambda_0-\lambda_1)\text{Log } e}{\text{Log}\left(\frac{\lambda_1/\lambda_0}{n}\right)}$$

$\therefore\ \bar{x} \geq c$

$\therefore$ The Best critical region is $\omega = \{x : \bar{x} \geq c\}$

12. Find the Best critical region for testing $H_0 : \theta = 3$ and against $H_1 : \theta = 4$ the population whose density function is $f(x) = \theta(1-\theta)^x$

Sol :

Given that,

$H_0 : \theta = 3$, $H_1 : \theta = 4$

and $f(x) = \theta(1-\theta)^x$

Then the Liklihood function is

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

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

$$L = \theta(1-\theta)^{x_1} \,.\, \theta(1-\theta)^{x_2} \ldots.. \theta(1-\theta)^{x_n}$$

$$L = \theta^n (1-\theta)^{\sum_{i=1}^{n}x_i}$$

By Neymann pearson's lemma

$$\frac{L_1}{L_0} \geq K \quad \ldots\ldots. (1)$$

Here, $H_0 : \theta = 3$

$$\Rightarrow L_0 = (3)^n (1-3)^{\sum_{i=1}^{n}x_i} = (3)^n (-2)^{\sum_{i=1}^{n}x_i} \quad \ldots.(2)$$

$$H_1 : \theta = 4 \Rightarrow L_1 = (4)^n (1-4)^{\sum_{i=1}^{n}x_i}$$

$$\Rightarrow (4)^n (-3)^{\sum_{i=1}^{n}x_i} \quad \ldots.. (3)$$

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{(4)^n (-3)^{\sum_{i=1}^{n}x_i}}{(3)^n (-2)^{\sum_{i=1}^{n}x_i}}$$

$$\left(\frac{4}{3}\right)^n \left(\frac{3}{2}\right)^{\sum_{i=1}^{n}x_i} \geq K$$

$$\left(\frac{4}{3}\right)^n \left(\frac{3}{2}\right)^{\sum_{i=1}^{n}x_i} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[\left(\frac{4}{3}\right)^n\left(\frac{3}{2}\right)^{\sum_{i=1}^{n}x_i}\right] \geq \text{Log K}$$

$$\text{Log}\left(\frac{4}{3}\right)^n + \text{Log}\left(\frac{3}{2}\right)^{\sum_{i=1}^{n}x_i} \geq \text{Log K}$$

$$n\,\text{Log}\left(\frac{4}{3}\right) + \sum_{i=1}^{n}x_i\ \text{Log}\left(\frac{3}{2}\right) \geq \text{Log K}$$

$$\sum_{i=1}^{n}x_i\ \text{Log}\left(\frac{3}{2}\right) \geq \text{Log k} - n\,\text{Log}\left(\frac{4}{3}\right)$$

$$\sum_{i=1}^{n}x_i \geq \frac{\text{Log}\left(k/n\right) - \text{Log}\left(\frac{4}{3}\right)}{\text{Log}\left(\frac{\frac{3}{2}}{n}\right)}$$

$$\bar{x} \geq \frac{\text{Log}\left(k/n\right) - \text{Log}\left(\frac{4}{3}\right)}{\text{Log}\left(\frac{3n}{2}\right)}$$

Where $C = \dfrac{\text{Log}\left(k/n\right) - \text{Log}\left(4/3\right)}{\text{Log}\left(3n/2\right)}$

$\therefore\ \bar{x} \geq c$

The Best critical region is $w = \{x : \bar{x} \geq c\}$

13. Find the Best critical region for testing $H_0{:}\theta = 1$ and against $H_1{:}\theta = 2$. whose probability density function is $f(x) = \dfrac{1}{\theta}e^{-x/\theta}$

Sol :

Given that,

$H_0 : \theta = 1,\ H_1 : \theta = 2$

and the probability density function is

$$f(x) = \frac{1}{\theta}e^{-x/\theta}$$

Then the Liklihood function is

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

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

$$L = \frac{1}{\theta}e^{-x_1/\theta}.\ \frac{1}{\theta}e^{-x_2/\theta}\ \ldots\ldots\ \frac{1}{\theta}e^{-x_n/\theta}$$

$$L = \frac{1}{\theta^n}\ e^{-\sum_{i=1}^{n}x_i/\theta}$$

By Neymann pearson's lemma

$$\frac{L_1}{L_0} \geq K \quad \ldots\ldots\ (1)$$

Here, $H_0 : \theta = 1$

$$\Rightarrow L_0 = \frac{1}{(1)^n}\ e^{-\sum_{i=1}^{n}x_i/1} = e^{-\sum_{i=1}^{n}x_i} \quad \ldots.\ (2)$$

$$H_1 : \theta = 2 \Rightarrow L_1 = \frac{1}{2^n}\ e^{-\sum_{i=1}^{n}x_i/2} \quad \ldots.\ (3)$$

Eq^n (2) & (3) sub in Eq^n (1)

$$\frac{\frac{1}{2^n}e^{-\sum_{i=1}^{n}x_i/2}}{e^{-\sum_{i=1}^{n}x_i}} \geq K$$

$$2^{-n}\ e^{-\sum_{i=1}^{n}x_i/2}\ e^{\sum_{i=1}^{n}x_i} \geq K$$

$$2^{-n}\ e^{\sum_{i=1}^{n}x_i\left(1-\frac{1}{2}\right)} \geq K$$

$$2^{-n}\ e^{\sum_{i=1}^{n}x_i\left(1/2\right)} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[2^{-n}e^{\sum_{i=1}^{n}x_i\left(1/2\right)}\right] \geq \text{Log K}$$

$$\text{Log } 2^{-n} + \text{Log } e^{\sum_{i=1}^{n} x_i (1/2)} \geq \text{Log K}$$

$$-n \text{ Log } 2 + \frac{1}{2} \sum_{i=1}^{n} x_i \text{ Log e} \geq \text{Log K}$$

$$\frac{1}{2} \sum_{i=1}^{n} x_i \text{ Log e} \geq \text{Log K} + n \text{ Log } 2$$

$$\sum_{i=1}^{n} x_i \geq \frac{\text{Log k} + n \text{ Log } 2}{\frac{1}{2}\text{Log e}}$$

$$\bar{x} \geq \frac{\text{Log}(k/n) + \text{Log } 2}{\frac{1}{2}\text{Log}(e/n)}$$

Where $c = \dfrac{\text{Log}(k/n) + \text{Log } 2}{\frac{1}{2}\text{Log}(e/n)}$

$\therefore \bar{x} \geq c$

$\therefore$ The best critical region is $w = \{x : \bar{x} \geq c\}$

14. Find the Best critical region for testing $H_0 : P = \frac{1}{2}$ and against $H_1 : P = \frac{1}{4}$ for the Geometric distribution.

Sol :

Given that,

$H_0 : P = \frac{1}{2}, H_1 : P = \frac{1}{4}$

The probability mass function of Geomtric distribution is

$$P(x = x) = p.q^x$$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_1) . P(x_2) \ldots\ldots P(x_n)$$

$$L = p.q^{x_1} . pq^{x_2} \ldots\ldots p.q^{x_n}$$

$$L = p^n q^{\sum_{i=1}^{n} x_i}$$

By Neymann pearson's lemma

$$\frac{L_1}{L_0} \geq K \quad \ldots\ldots. (1)$$

Here, $H_0 : P = \frac{1}{2}$

$$\Rightarrow L_0 = \left(\frac{1}{2}\right)^n q^{\sum_{i=1}^{n} x_i} = \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{\sum_{i=1}^{n} x_i} \quad \ldots.. (2)$$

$$H_1 : P = \frac{1}{4}$$

$$\Rightarrow L_1 = \left(\frac{1}{4}\right)^n q^{\sum_{i=1}^{n} x_i} = \left(\frac{1}{4}\right)^n \left(\frac{3}{4}\right)^{\sum_{i=1}^{n} x_i} \quad \ldots.. (3)$$

Sub Eq^n (2) & (3) in Eq^n (1)

$$\frac{\left(\frac{1}{4}\right)^n \left(\frac{3}{4}\right)^{\sum_{i=1}^{n} x_i}}{\left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{\sum_{i=1}^{n} x_i}} \geq K$$

$$\left(\frac{\frac{1}{4}}{\frac{1}{2}}\right)^n \left(\frac{\frac{3}{4}}{\frac{1}{2}}\right)^{\sum_{i=1}^{n} x_i} \geq K$$

$$\left(\frac{2}{4}\right)^n \left(\frac{6}{4}\right)^{\sum_{i=1}^{n} x_i} \geq K$$

$$\left(\frac{1}{2}\right)^n \left(\frac{3}{2}\right)^{\sum_{i=1}^{n} x_i} \geq K$$

Taking 'Log' on b.s

$$\text{Log}\left[\left(\frac{1}{2}\right)^n \left(\frac{3}{2}\right)^{\sum_{i=1}^{n} x_i}\right] \geq \text{Log K}$$

$$\text{Log}\left(\frac{1}{2}\right)^n + \text{Log}\left(\frac{3}{2}\right)^{\sum_{i=1}^{n} x_i} \geq \text{Log K}$$

$$n\,\text{Log}\left(\frac{1}{2}\right) + \sum_{i=1}^{n} x_i\ \text{Log}\left(\frac{3}{2}\right) \geq \text{Log K}$$

$$\sum_{i=1}^{n} x_i\ \text{Log}\left(\frac{3}{2}\right) \geq \text{Log K} - n\,\text{Log}\left(\frac{1}{2}\right)$$

$$\sum_{i=1}^{n} x_i \geq \frac{\text{Log k} - n\,\text{Log}\left(\frac{1}{2}\right)}{\text{Log}\left(\frac{3}{2}\right)}$$

Divide by 'n' on b.s

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \frac{\text{Log}\left(k/n\right) - \text{Log}\left(\frac{1}{2}\right)}{\text{Log}\left(\frac{3/2}{n}\right)}$$

$$\overline{x} \geq \frac{\text{Log}\left(k/n\right) - \text{Log}\left(\frac{1}{2}\right)}{\text{Log}\left(\frac{3n}{2}\right)}$$

Where $c = \dfrac{\text{Log}\left(k/n\right) - \text{Log}\left(\frac{1}{2}\right)}{\text{Log}\left(\frac{3n}{2}\right)}$

$\therefore\ \overline{x} \geq c$

$\therefore$ The Best critical region is $\omega = \{x : \overline{x} \geq c\}$

15. Obtain the most powerful test for testing the mean $\mu = \mu_0$ against $\mu = \mu_1 (\mu_1 = \mu_0)$ where $\sigma^2 = 1$ in normal population.

Sol : **(June-19)**

Given that,

Null hypothesis H_0 is,

$H_0 : \mu = \mu_0$

Alternative hypothesis H_i is,

$H_1 : \mu = \mu_1$

The probability function of a normal distribution is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{1}{2\sigma^2}.(x-\mu)^2}$$

where x lie in range $-\infty < x < \infty$

$\therefore\ \sigma^2 = 1$ we get,

$$f(x) = \frac{1}{\sqrt{2\pi}}.e^{-1}(x-\mu)^2$$

where x lie in range $-\infty < x < \infty$

The likehood function of the normal distribution is,

$L = f(x_1,\mu), f(x_2, \mu), f(x_3,\mu), \ldots\ldots., f(x_n,\mu)$

$$L = \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(x_1-\mu)^2}, \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(x_2-\mu)^2},$$

$$\frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(x_3-\mu)^2}, \ldots\ldots., \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(x_n-\mu)^2}$$

$$L = \left(\frac{1}{\sqrt{2\pi}}\right)^n.e^{-\frac{1}{2}\Sigma(x_i-\mu)^2}$$

Under H_0 we have $\mu = \mu$ then we get,

$$L_0 = \left(\frac{1}{\sqrt{2\pi}}\right)^n.e^{-\frac{1}{2}\Sigma(x_i-\mu_0)^2} \qquad \ldots\ldots (1)$$

Under H_1 we have $\mu = \mu_1$ then we get,

$$L_1 = \left(\frac{1}{\sqrt{2\pi}}\right)^n.e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2} \qquad \ldots. (2)$$

According to Neyman - Pearson lemma the best critical region is calculated by using the following relation.

$$\frac{L_1}{L_0} > k$$

Substituting the value of L_0 and L_1 from equations (1) and (2) respectively, we get,

$$\Rightarrow \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n .e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n .e^{-\frac{1}{2}\Sigma(x_i-x_0)^2}} > k$$

$$\Rightarrow \frac{e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2}}{e^{-\frac{1}{2}\Sigma(x_i-\mu_0)^2}} > k$$

Applying log on both sides we get,

$$\Rightarrow \log\left[\frac{e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2}}{e^{-\frac{1}{2}\Sigma(x_i-\mu_0)^2}}\right] > \log k$$

$$\Rightarrow \log\left(e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2}\right) - \log\left(e^{-\frac{1}{2}\Sigma(x_i-\mu_0)^2}\right) > \log k$$

$$\Rightarrow -\frac{1}{2}\Sigma(x_i - \mu_1)^2 \log e + \frac{1}{2}\Sigma(x_i - \mu_0)^2 \log e > \log k$$

$$\Rightarrow -\frac{1}{2}\Sigma(x_i - \mu_1)^2 + \frac{1}{2}\Sigma(x_i - \mu_0)^2 > \log k$$

$[\because \log e = 1]$

$$\Rightarrow -\frac{1}{2}(\Sigma x_i^2 - 2\Sigma x_i\mu_1 + \Sigma\mu_1^2) + \frac{1}{2}\left(\Sigma x_i^2 - 2\Sigma x_i\mu_0 + \Sigma\mu_0^2\right) > \log k$$

$$\Rightarrow \frac{1}{2}\left[-\Sigma x_i^2 + 2\Sigma x_i\mu_1 - \Sigma\mu_1^2\right] + \frac{1}{2}\left[\Sigma x_i^2 - 2\Sigma x_i\mu_0 + \Sigma\mu_0^2\right] > \log k$$

$$\Rightarrow \frac{1}{2}\left[-\Sigma x_i^2 + 2\Sigma x_i\mu_1 - \Sigma\mu_i^2 + \Sigma x_i^2 - 2\Sigma x_i\mu_0 + \Sigma\mu_0^2\right] > \log k$$

$$\Rightarrow \frac{1}{2}\left[2\Sigma x_i(\mu_1 - \mu_o) - \Sigma\mu_1^2 + \Sigma\mu_0^2\right]$$

$$\Rightarrow \frac{1}{2}\left[\sum_{i=1}^{n}\left(\mu_0^2 - \mu_1^2\right) + 2\sum_{i=1}^{n}x_i(\mu_1 - \mu_0)\right] > \log k$$

Applying summation we get,

$$\Rightarrow \frac{1}{2}\left[n\left(\mu_0^2 - \mu_1^2\right) + nx_i(\mu_1 - \mu_-)\right] > \log k$$

$$\Rightarrow \frac{1}{2}\left[n\left(\mu_0^2 - \mu_1^2\right)\right] + \frac{1}{2}\left[(2nx_i(\mu_1 - \mu_0)\right] > \log k$$

$$\Rightarrow \frac{1}{2}n\left(\mu_0^2 - \mu_1^2\right) + nx_i(\mu_1 - \mu_0) > \log k$$

Multiplying the above equation by 2 we get,

$$\Rightarrow \frac{2}{2}\left[n\left(\mu_0^2 - \mu_1^2\right) + nx_i(\mu_1 - \mu_0)\right] > 2\log k$$

$$\Rightarrow n\left(\mu_0^2 - \mu_1^2\right) + nx_i(\mu_1 - \mu_0) > 2\log k$$

$$\Rightarrow nx_i\,[m_1 - m_0] > 2\log k - n\left[\mu_0^2 - \mu_1^2\right]$$

$$\Rightarrow x_i > \frac{2\log k - n\left[\mu_0^2 - \mu_1^2\right]}{n\left[\mu_1 - \mu_0\right]}$$

$$\Rightarrow \bar{x} > \frac{2\log k + n\left[\mu_1^2 - \mu_0^2\right]}{n\left[\mu_1 - \mu_0\right]}$$

Hence, if $\mu_1 > \mu_0$

Then the best critical region will be,

$W = (x : \bar{x} > P)$ where,

$$P = \left[\frac{2\log k + n[\mu_1^2 - \mu_0^2]}{n\left[\mu_1 - \mu_0\right]}\right]$$

If $\mu_1 < \mu_0$ then the best critical region will be,

$W = (x : \bar{x} < Q)$ where,

$$Q = \frac{2\log k + n\left[\mu_0^2 - \mu_1^2\right]}{n\left[\mu_0 - \mu_1\right]}$$

1.3 Examples in Case of Binomial, Poisson, Exponential and Normal Distributions and their Power of the Test Functions

Q9. Explain briefly about Liklihood Functions.

Ans :

Let x_1, x_2 ...... x_n are 'n' dimensional random variables with probability density function $f(xi, \theta)$ and probability mass function $p(xi, \theta)$ then the Liklihood function of the random variables x_1, x_2 ..... x_n are given by

$$L = \prod_{i=1}^{n} f(x_i, \theta) \Rightarrow L = f(x_1, \theta)$$

$$= f(x_2, \theta) .... f(x_n, \theta)$$

$$L = \prod_{i=1}^{n} P(x_i, \theta)$$

$$\Rightarrow L = P(x_1, \theta).P(x_2, \theta).....P(x_n, \theta)$$

PROBLEMS

16. Find the Likilhood function of Binomial distribution.

Sol :

The probability mass function of Binomial distribution is $P(x) = {}^n c_x p^x q^{n-x}$; x = 0, 1, 2, ... n Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i, \theta)$$

$$L = P(x_1, \theta).P(x_2, \theta) ..... P(x_n, \theta)$$

$$L = {}^n c_{x_1} p^{x_1} q^{n-x_1} . {}^n c_{x_2} p^{x_2} q^{n-x_2} ...... {}^n c_{x_n} p^{x_n} q^{n-x_n}$$

$$L = \binom{n}{x_1}\binom{n}{x_2} ...... \binom{n}{x_n} . P^{x_1+x_2+......+x_n} q^{n-(x_1+x_2+.....+x_n)}$$

$$L = \prod_{i=1}^{n} \binom{n}{x_i} P^{\sum_{i=1}^{n} x_i} q^{n-\sum_{i=1}^{n} x_i}$$

Conclusion :

The Liklihood function of Binomial distribution is $L = \prod_{i=1}^{n} \binom{n}{x_i} P^{\sum_{i=1}^{n} x_i} q^{n-\sum_{i=1}^{n} x_i}$

17. Find the Liklihood function of

(i) Poisson distribution

(ii) Geometric distribution

(iii) Negative Binomial distribution

(iv) Exponential distribution

(v) Normal distribution

Sol :

(i) Poisson distribution :

The probability mass function of poisson distribution is

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}; x = 0, 1, 2, ...\infty$$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_i) . P(x_2)..... P(x_n)$$

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} .... \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L = \frac{\left(e^{-\lambda}\right)^n \lambda^{x_1+x_2+.......+x_n}}{\prod_{i=1}^{n} x_i!}$$

$$L = \frac{\left(e^{-\lambda}\right)^n \lambda \sum_{i=1}^{n} x_i}{\prod_{i=1}^{n} x_i!}$$

(ii) Geometric Distribution :

The probability mass function of Geometric distribution is

$P(x) = p.q^x \; ; \; x = 0, 1, 2, ... \infty$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_1). P(x_2) ..... P(x_n)$$

$$L = p.q^{x_1}.p.q^{x_2}......p.q^{x_n}$$

$$L = p^n q^{x_1+x_2+.....+x_n}$$

$$L = p^n q^{\sum_{i=1}^{n} x_i}$$

(iii) Negative Binomial distribution :

The probability mass function of Negative Binomial distribution is

$$P(x) = \binom{x+r-1}{x} p^r q^x$$

Then the liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_1) . P(x_2) ..... P(x_n)$$

$$L = \binom{x_1+r-1}{x_1}.p^r q^{x_1}.\binom{x_2+r-1}{x_2}p^r q^{x_2}....\binom{x_n+r-1}{x_n}p^r q^{x_n}$$

$$L = \binom{x_1+r-1}{x_1}\binom{x_2+r-1}{x_2}.....\binom{x_n+r-1}{x_n}$$

$$(p^r)^n q^{x_1+x_2+.....+x_n}$$

$$L = \prod_{i=1}^{n}\binom{x_i+r-1}{x_i} p^{nr} q^{\sum_{i=1}^{n} x_i}$$

(iv) Exponential distribution :

The probability density function of Exponential distribution is

$f(x) = \theta .e^{-\theta x}$

Then the Liklihood function is $L = \prod_{i=1}^{n} f(x_i)$

$L = f(x_1) .f(x_2) ...... f(x_n)$

$L = \theta.e^{-\theta x_1}.\theta.e^{-\theta x_2}......\theta e^{-\theta x_n}$

$L = \theta^n e^{-\theta(x_1+x_2+.....+x_n)}$

$L = \theta^n e^{-\theta\sum_{i=1}^{n} x_i}$

(v) Normal distribution : The probability density function of normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Then the Liklihood function is

$L = \prod_{i=1}^{n} f(x_i)$

$L = f(x_1).\ f(x_2)\ .....\ f(x_n)$

$$L = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x_1-\mu}{\sigma}\right)^2} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2}$$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{20^{-2}}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Conclusion :

(i) The Liklihood function of poisson distribution is $L = \dfrac{(e-\lambda)^n \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$

(ii) The Liklihood function of Geometric distribution is $L = p^n q^{\sum_{i=1}^{n} x_i}$

(iii) The Liklihood function of Negative Binomial distribution is $L = \prod_{i=1}^{n}\binom{x_i+r-1}{x_i} p^{nr}.q^{\sum_{i=1}^{n} x_i}$

(iv) The Liklihood function of Exponential distribution is $L = \theta^n.e^{-\theta}\sum_{i=1}^{n} x_i$

(v) The Liklihood function of Normal distribution is $L = \left(\dfrac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{20^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

Short Question and Answers

1. Null Hypothesis

Ans :

It is denoted by H_0 and it is defined by a statement no differences is called Null hypothesis.

Ex : A sample of size n = 30 drawn from a population and it is found that the sample mean $\bar{x}$ = 15. If we want to test the population mean is 10 (or) not ? If then the null hypothesis is $H_0 : \mu = 10$

2. Alternative Hypothesis.

Ans :

It is denoted by H_1 and it is defined by a statement against the Null hypothesis is called Alternative hypothesis.

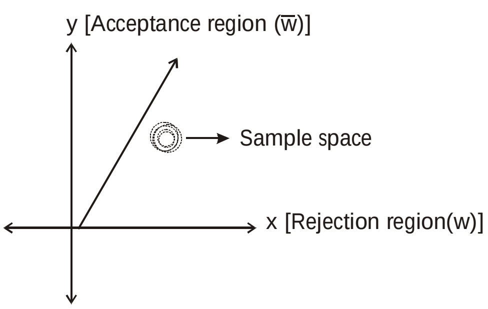
Ex : A sample size n=30 drawn from a population and it is found that the sample mean $(\bar{x})$ = 15. If we want to test the population mean is 10 (or) not ? then the alternative hypothesis is $H_1 : \mu \neq 10$.

3. Critical Region

Ans :

Let x_1, x_2 ..... x_n are the samples of size 'n' drawn from the population the set of all points plotted between x-axis and y-axis; called sample space.

The division of the sample space in two exclusive regions i.e., Acceptance region and Rejection region [critical region]. The Acceptance region is denoted by '$\bar{w}$'and the Rejection region is denoted by 'w'



4. Define the term Power of a test.

Ans :

For testing the hypothesis H_0 and H_1, the test with probability α and β of type -I and type - II. Errors respectively.

The quantity $1-\beta$ is called "power of the test"

The power of the test depends upon the difference between parameter value specified by H_0 , the actual value of parameter.

$$1-\beta = 1 - P\text{ [Type - II Error]}$$

$$1-\beta = 1 - P\text{ [Accept } H_0\text{/ When it is false]}$$

$$1-\beta = 1 - P\left[x \in \bar{w} / H_1\right]$$

$$1-\beta = 1-\int_{\bar{w}} L_1 dx$$

$$1-\beta = \int_{w} L_1 dx$$

$$\begin{bmatrix} \therefore \bar{\omega}+\omega = 1 \\ \omega = 1-\bar{\omega} \end{bmatrix}$$

5. Write a short note on Non-Randomized and Randomized test function.

Ans :

(i) Randomized Test Function :

A function $\psi(.)$ is said to be randomized test function if its sample space values lie in the close interval [0,1] i.e., $0 \le \psi(x) \le 1$ for all x. The exception and rejection of null Hypothesis 'H_0' is decided with probabilities $\psi(x)$ and $1 - \psi(x)$. In case that value of a random variable x is equal to the observed value x, then two possible outcomes say p and $\bar{p}$ are considered with the above probabilities. The null hypothesis H_0 is rejected if P occurs in the experiment otherwise, H_0 is accepted. Here $\psi(x)$ is called critical function An example of randomized test function is,

$$y(x) = \begin{cases} 1\ ; & \text{if } x = 3 \\ 1/8; & \text{if } x = 2 \\ 0; & \text{otherwise} \end{cases}$$

(ii) Non-randomized Test function :

A function is said to be non-randomized test function if all sample values lie in the open interval $\{0,1\}$ i.e., either 'O' or '1' when the null hypothesis 'H_0' is rejected , the sample is divided into two sets say Q and compliment of Q i.e., Q' in such a way that $Q \cup Q'$ = sample space and $Q \cap Q' = \phi$. Here Q is called rejection region because the 'H_0' is rejected it if x falls into Q.

A non-randomized test function can be written as,

$$\psi(x) = \begin{Bmatrix} 0; \text{ if } x \in Q \\ 1; \text{ otherwise} \end{Bmatrix}$$

6. Write short notes on Type-I and Type-II Error.

Ans :

In Natural life we can see the following

1. Accept H_0, when it is true
2. Reject H_0, when it is false
3. Reject H_0, when it is true
4. Accept H_0, when it is false

In the above first two statements are correct statements and last two statements are Error statements.

TYPE -I ERROR : The Error statement reject H_0, when it is true is called Type -I Error.

TYPE -II ERROR : The Error statement Accept H_0, when it is false is called type - II Error

7. Write a short notes on Level of Significance.

Ans :

The probability of Type- I Error is called level of significance which is denoted by 'α'

$\propto$ = P [type - I Error]

$\propto$ = P [Reject H_0/when it is true]

$\propto$ = P [$x \in w / H_0$]

$$\propto = \int_{\omega} L_0 dx$$

β = P [type -II Error]

β = P [Accept H_0 / when it is false]

β = P [$x \in \bar{w}$ /H_1]

$\beta = \int_{\bar{w}} L_1 dx$

Where L_0, L_1 are Liklihood functions

---

8. Most Powerful test.

Ans :

A test procedure is said to be most powerful test for testing the null hypothesis $H_0 : \theta = \theta_0$ against simple Alternative Hypothesis $H_1 : \theta = \theta_1$ is called most powerful test.

Most Powerful Critical Region :

The critical region (w) is said to be most powerful critical region with size 'α' for testing Null hypothesis $H_0 : \theta = \theta_0$ and against simple Alternative Hypothesis $H_1 : \theta = \theta_1$ and satisfy the following conditions $P\{x \in \omega / H_1\} \geq P\{x \in \omega_1 / H_1\}$

ω	ω_1
$\alpha = P\{x \in \omega / H_0\}$	$\alpha_1 = P\{x \in \omega_1 / H_0\}$
$\beta = P\{x \in \bar{\omega} / H_1\}$	$\beta_1 = P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = 1-P\{x \in \bar{\omega} / H_1\}$	$1-\beta_1 = 1-P\{x \in \bar{\omega}_1 / H_1\}$
$1-\beta = P\{x \in \omega / H_1\}$	$1-\beta_1 = P\{x \in \omega_1 / H_1\}$

---

9. Find the Liklihood function of Poisson distribution.

Ans :

The probability mass function of poisson distribution is

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0, 1, 2, ...\infty$$

Then the Liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_i) . P(x_2)..... P(x_n)$$

$$L = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} .... \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$L = \frac{\left(e^{-\lambda}\right)^n \lambda^{x_1+x_2+.......+x_n}}{\prod_{i=1}^{n} x_i!}$$

$$L = \frac{\left(e^{-\lambda}\right)^n \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

10. Negative Binomial distribution.

Ans :

The probability mass function of Negative Binomial distribution is

$$P(x) = \binom{x+r-1}{x} p^r q^x$$

Then the liklihood function is

$$L = \prod_{i=1}^{n} P(x_i)$$

$$L = P(x_1) \,.\, P(x_2) \,.....\, P(x_n)$$

$$L = \binom{x_1+r-1}{x_1}.p^r q^{x_1}.\binom{x_2+r-1}{x_2}p^r q^{x_2}....\binom{x_n+r-1}{x_n}p^r q^{x_n}$$

$$L = \binom{x_1+r-1}{x_1}\binom{x_2+r-1}{x_2}.....\binom{x_n+r-1}{x_n}$$

$$(p^r)^n q^{x_1+x_2+.....+x_n}$$

$$L = \prod_{i=1}^{n}\binom{x_i+r-1}{x_i} p^{nr} q^{\sum_{i=1}^{n} x_i}$$

11. Exponential distribution.

Ans :

The probability density function of Exponential distribution is

$$f(x) = \theta \,.e^{-\theta x}$$

Then the Liklihood function is $L = \prod_{i=1}^{n} f(x_i)$

$$L = f(x_1) \,.f(x_2) \,......\, f(x_n)$$

$$L = \theta.e^{-\theta x_1}.\theta.e^{-\theta x_2}......\theta e^{-\theta x_n}$$

$$L = \theta^n e^{-\theta(x_1+x_2+.....+x_n)}$$

$$L = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

Exercise Problems

1. Let p be the probability that a coin will fall head in a single toss in order to test $H_0 : \rho = \frac{1}{2}$ against $H_1 : \rho = \frac{3}{4}$. The coin is tossed 5 times and H_0 is rejected if more than 3 heads are obtained. Find the probability of type I error and power test.

   [Ans : $\frac{3}{16}, \frac{81}{128}$]

---

2. A dice was thrown 400 times and 'sinx' resulted 80 times. Do the data justify the hypothesis of an unbiased dice?

   [Ans: 1.77]

---

3. In a big city 325 men out of 600 were found to be smokers. Does this information support the conclusion support the conclusion that the majority of men in this city are smokers? (State the hypothesis clearly).

   [Ans : 1.08]

---

5. It is desired to test a hypothesis is $H_0 : P - P_0 = \frac{1}{2}$ against $H_1 : P = \frac{3}{4}$ on the basic of tossing a coin once, where P is the probability of getting a head in a single trial and agreeing to accept H_0 if a tail appears and to accept H_1 otherwise. Find the value of α and β.

   [Ans : $\frac{1}{2}, \frac{1}{4}$]

---

6. For testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$, Where

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

   Obtain a and b for critical region $x \geq 5$. Also obtain power function of the test.

   [Ans : 0.75]

Choose the Correct Answers

1. Statistical hypothesis are statements about the probability distributions of the ________ [c]
   (a) Mean (b) Standard deviation
   (c) Population (d) Hypothesis
2. ________ involves rejection of null hypothesis when it is the true. [d]
   (a) Null hypothesis (b) Alternative hypothesis
   (c) Type II error (d) Type I error
3. The most commonly used levels of significance during the testing of hypothesis are ________ [c]
   (a) 1% (b) 5%
   (c) Both (a) and (b) (d) None of the above
4. The test statistic under the null hypothesis 2 = ________ [a]
   (a) $\frac{t-E(t)}{S.\ E(t)}$ (b) $\frac{t+E(t)}{S.E(t)}$
   (c) $\frac{t}{S.E(t)}$ (d) $\frac{t}{E(t)}$
5. ________ helps in determining the critical region of size α. [b]
   (a) Neyman-pearson (b) Neyman-pearson lemma
   (c) Neyman-pearson factorization (d) Neyman factorization
6. The probability function of poisson distribution is________ [d]
   (a) $f(x,\lambda)=\frac{-e^{\lambda}\lambda^{x}}{2x!}$ (b) $f(x,\lambda)=\frac{-e^{-x}\lambda^{x}}{x!}$
   (c) $f(x,\lambda)=\frac{-e^{\lambda}\lambda^{x}}{x!}$ (d) $f(x,\lambda)=\frac{e^{\lambda}\lambda^{x}}{x!}$
7. The condition $H_1=\mu<\mu_0$ denotes ________ test . [a]
   (a) Lelt tailed (b) Right tailed
   (c) Two tailed (d) One tailed
8. The value of test statistic depends upon the level of significance and ________. [b]
   (a) Null hypothesis (b) Alternative hypothesis
   (c) Type I error (d) Type II error
9. If an alternative hypothesis is denoted by '>' or '<' then the hypothesis test is referred to as ________ [c]
   (a) Type I error (b) Type II error
   (c) One tailed (d) Two tailed
10. A function y(.) is said to randomized test function if its sample space values cie in ________ [d]
    (a) (0,1) (b) (1,2)
    (c) [1,2] (d) [0,1]

Fill in the blanks

1. A hypothesis that contradict to the given null hypothesis is called an ________.
2. ________ involves accptance of the null hypothesis when it is false and should be rejected.
3. The Level of significance also refers to the ________.
4. ________ refers to the probability of containing a statistics random value 't' in the critical region.
5. An alternative hypothesis that is represented by a $\neq$ symbol then the rest is riferred to as ________.
6. ________ helps in testing the null hypothesis H_0 against the alternative hypothesis (H_1).
7. The area under the probability cure of critical region is ________ where the null hypothesis is accepted.
8. $\psi(x) = \begin{cases} 1 & \text{if} & x = 3 \\ 1/8 & \text{if} & x = 2 \\ 0 & 0 & \text{otherwise} \end{cases}$ is an example of ________.
9. ________ is defined as the function when the sample values lies in open internal (0,1)
10. The significance level of test (α) is calcuted using ________ .

ANSWERS

1. Alternative hypothesis
2. Type II error
3. Size of type I error
4. Level of significance
5. Two tailed test
6. Power of test
7. Acceptance region
8. Randomized test function
9. Non-Randomized test function
10. $\alpha = P(x \in w / H_0)$

Very Short Questions and Answers

1. Define statistical hypothesis.

Ans :

Statistical hypothesis are statement about the probability distributions of the populations.

2. What is alternative hypothesis ?

Ans :

A hypothesis which is complementary to the null hypothesis is called an alternative hypothesis and it is denoted by it.

3. Define level of significance.

Ans :

The size of the type 1 error is called as level if significantly and it is denoted by α.

4. What is type II error ?

Ans :

Type II error involves acceptance of the null hypothesis when it is false and should be rejected .

5. What is one-tailed test ?

Ans :

If an alternative hypothesis is denoted by ' > ' (or) ' < ' symbol, in a statistical hypothesis test then this test is referred to be as one-tailed test.

UNIT II	Large sample tests for single sample mean, difference of means, single sample proportion, difference of proportions and difference of standard deviations. Fisher's Z-transformation for population correlation coefficient(s) and testing the same in case of one sample and two samples. Definition of order statistics and statement of their distributions.

2.1 Large Sample Tests

Q1. Define Large Sample Test.

Ans :

The sample size $n \geq 30$ then it is called large sample test. The different large sample tests are :

(i) Test for single mean

(ii) Test for two means

(iii) Test for standard deviation

(iv) Test for single proportion

(v) Test for two proportions

(vi) Test for correlation coefficient.

2.1.1 Procedure for Testing of Hypothesis

Q2. Explain the Procedure for Testing of Hypothesis.

Ans : **(June-19, Imp.)**

The various steps in testing of statistical hypothesis in a systematic manner.

Null Hypothesis (H_0)

Set up the Null hypothesis H_0.

Alternative Hypothesis (H_1)

Set up the alternative hypothesis H_1. This test is decides whether we have to use one tailed test or two tailed test.

Level of Significance (α)

Choose the appropriate level of significance (α) depending on the reliability of the estimator values.

Test Statistic

Compute the test statistic $Z = \left|\frac{t - E(t)}{SE(t)}\right| \sim N(0,1)$ under H_0.

Conclusion

Compare the calculated value of z in step 4 with the significant value (tabulated value z) at given level of significance if $|z_{cal}| < z_{tab}$ then we accept H_0 otherwise we reject H_0.

One Tailed Test

A Test of statistical hypothesis is said to be one tailed test if alternative hypothesis is lessthan type or greater than type.

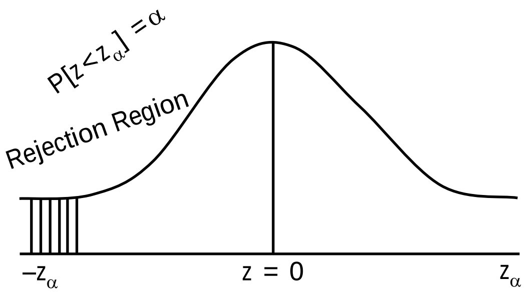
The one tailed test are 2 types.

(i) Left tailed test.

(ii) Right tailed test

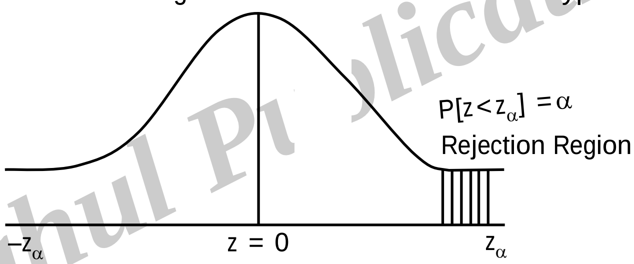
(i) Left Tailed Test

The test procedure is said to be left tailed test if the alternative hypothesis is lessthan type.



(ii) Right Tailed Test

The test procedure is said to be right tailed test if the alternative Hypothesis is greaterthan type.



Two Tailed Test

A test procedure is said to be two tailed test if alternative hypothesis is not equal types.

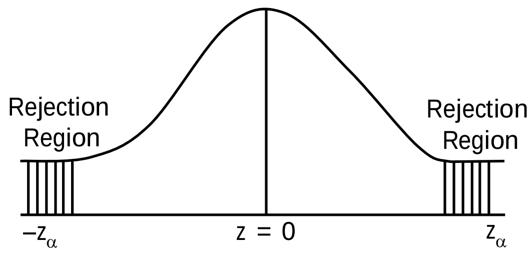


Table Values of Large Samples

Critical value (a)	Level of significance		
	1%	5%	10%
Two Tailed Test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Right Tailed Test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left Tailed Test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

2.1.2 Single Sample Mean

Q3. Explain the procedure for testing single sample mean.

Ans : **(Imp.)**

If $x_1, x_2 ..... x_n$ be a random samples drawn from a normal population with mean (μ) and variance (σ^2) of size 'n' then the sample mean is normally distributed with mean (μ) and variance (σ^2/n).

i.e., $\bar{x} \sim N(\mu, \sigma^2/n)$

Null Hypothesis (H_0)

There is no significance difference between population mean and sample mean.

(or)

The samples has been drawn from a population.

Alternative Hypothesis (H_1)

There is significance difference between population mean and sample mean.

(or)

The samples has not been drawn from population.

Level of Significance (α)

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null hypothesis then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0,1) \text{ under } H_0$$

$$z = \left|\frac{\bar{x} - E(\bar{x})}{S.E(\bar{x})}\right| \sim N(0,1)$$

$$z = \left|\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}\right| \sim N(0,1)$$

$\bar{x}$ = Sample mean

μ = Population mean

σ = Population standard deviation

n = Sample size

Remarks

If the population standard deviation σ is unknown then we use its estimate provided by sample variance is given by,

$$s^2 = \hat{\sigma}^2$$

$$s = \hat{\sigma}$$

Then the test statistic is given by

$$z = \left|\frac{\bar{x} - \mu}{s / \sqrt{n}}\right| \sim N(0,1) \text{ Under } H_0$$

Confidence Limits

(i) The 95% confidence limits for population mean μ are:

$$\bar{x} \pm 1.96 \text{ S.E } (\bar{x})$$

(or)

$$\bar{x} \pm 1.96 \ \sigma/\sqrt{n}$$

(ii) The 98% confidence limits for population mean μ are

$$\bar{x} \pm 2.33 \text{ S.E } (\bar{x})$$

(or)

$$\bar{x} \pm 2.33 \ \sigma/\sqrt{n}$$

(iii) The 99% confidence limits for population mean μ are

$$\bar{x} \pm 2.58 \text{ S.E } (\bar{x})$$

(or)

$$\bar{x} \pm 2.58 \ \sigma/\sqrt{n}$$

Incase of finite population size 'N' then the corresponding 95%, 98%, 99% confidence limits for population mean μ are respectively.

(i) The 95% confidence limits for population mean μ are:

$$\bar{x} \pm 1.96 \ \sigma/\sqrt{n} \sqrt{\frac{N-n}{n-1}}$$

(ii) The 98% confidence limits for population mean μ are

$$\bar{x} \pm 2.33 \text{ S.E } (\bar{x})$$

(or)

$$\bar{x} \pm 2.33 \ \sigma/\sqrt{n} \sqrt{\frac{N-n}{n-1}}$$

(iii) The 99% confidence limits for population mean μ are

$$\bar{x} \pm 2.58 \text{ S.E } (\bar{x})$$

(or)

$$\bar{x} \pm 2.58 \ \sigma/\sqrt{n} \sqrt{\frac{N-n}{n-1}}$$

Conclusion

If the calculated value of $z(z_{cal})$ is less than the Tabulated value of $z(z_{tab})$ at certain level of significance then we accept H_0 otherwise we reject H_0.

PROBLEMS

1. A sample of 900 members as a mean 3.5 cm and standard deviation 2.61 cm is the sample from a large population mean 3.25 cm and also calculate the 95% confidence limits for the population mean.

Sol :

Null Hypothesis (H_0)

The samples has been drawn from a population mean is 3.25 cm

i.e., $H_0: \mu = 3.25$ cm

Alternative Hypothesis (H_1)

The samples have not drawn from a population mean is 3.25 cm.

i.e., $H_1: \mu \neq 3.25$ cm [Two Tailed Test]

Given that,

$n = 900, \bar{x} = 3.5$ cm, $s = 2.61$ cm, $\mu = 3.25$ cm

Under the Null Hypothesis, the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{3.5-3.25}{2.6/\sqrt{900}}\right| \sim N(0,1)$$

$$z = \left|\frac{0.25}{0.087}\right| \sim N(0,1)$$

$$z = 2.873$$

$$|z_{cal}| = 2.873$$

$|z_{Tab}|$ value at 5% level of significance of two tailed test is 1.96.

$\therefore$ The calculated value of z is greaterthan the tabulated value of z,

i.e., $|z_{cal}| > |z_{tab}|$

$2.873 > 1.96$

We reject H_0

Conclusion

The samples are not drawn from the population mean 3.25 cm.

The 95% confidence limits for population mean μ are $\bar{x} \pm 1.96 \ s/\sqrt{n}$.

$$[\bar{x} + 1.96 \ s/\sqrt{n}, \ \bar{x} - 1.96 \ s/\sqrt{n}]$$

$$\left[3.5+1.96\left(\frac{2.61}{\sqrt{900}}\right), 3.5-1.96\left(\frac{2.61}{\sqrt{9000}}\right)\right]$$

[3.67, 3.32]

$\therefore$ The 95% confidence limits for population mean μ are [3.67, 3.32].

2. An Ambulance services claims best critical region that is taken on average 8.9 min its destination in emergency call to check on this claims the agency which license ambulance services as timed on 50 emergency calls getting a mean of 9.3 min with standard deviation 1.6 min. What can they conclude of the level of significance $\alpha = 0.05$.

Sol :

Null Hypothesis (H_0)

The ambulance service to reach the destination in 8.9 min average.

i.e., $H_0 : m = 8.9$

Alternative Hypothesis (H_1)

The ambulance service does not reach the destination in 8.9 min average.

i.e., $H_1: \mu > 8.9$ [Right Tailed Test]

Given that,

$n = 50, \bar{x} = 9.3$ Min, $S = 1.6$ Min

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{9.3-8.9}{1.6/\sqrt{50}}\right| \sim N(0,1)$$

$$z = \left|\frac{0.400}{0.226}\right| \sim N(0,1)$$

$$z = 1.769$$

$$|z_{cal}| = 1.769$$

$|z_{Tab}|$ value at 5% level of significance of right tailed test is 1.645.

∴ The calculated value of z is greaterthan the tabulated value of z,

∴ $z_{cal} > z_{tab}$

i.e., $1.769 > 1.645$

∴ We reject H_0

Conclusion

The ambulance does not reach the destination of population mean 8.9 min average.

3. The mean breaking strength of cables supplied by a manufactured is 1800 with standard deviation 100 by a new technique manufacturing is claimed that the breaking strength as increasing in order to test these claimed a sample of 50 cables is tested. It is found to be that the mean is 1850. We can support claim at 0.01 level of significance.

Sol :

Null Hypothesis (H_0)

The breaking strength of cables is 1800 i.e., H_0: m = 1800.

i.e., $H_0 : \mu = 8.9$

Alternative Hypothesis (H_1)

The breaking strength of cables in increasing in order.

i.e., $H_1: \mu > 1800$ [Right Tailed Test]

Given that,

$\mu = 1800$, S.D(σ) = 100, n = 50,

$\bar{x} = 1850$

Under the Null Hypothesis, the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{1850-1800}{100/\sqrt{50}}\right| \sim N(0,1)$$

$$z = \left|\frac{50}{14.1421}\right| \sim N(0,1)$$

$$z = 3.53$$

$$|z_{cal}| = 3.53$$

$|z_{Tab}|$ value at 1% level of significance of right tailed test is 2.33.

∴ The calculated value of $z(z_{cal})$ is greaterthan the tabulated value of $z(z_{Tab})$

∴ $z_{cal} > z_{Tab}$

i.e., $3.53 > 2.33$

∴ We reject H_0

Conclusion

The breaking strength of cables is increasing.

4. A sample of 400 observations has mean 95 and standard deviation 12. Let it be a random sample for a population mean 98. What are the maximum values of population mean.

Sol :

Null Hypothesis (H_0)

The samples are drawn from the population mean 98 i.e., H_0; $\mu = 98$

Alternative Hypothesis (H_1)

The samples are not drawn from the population mean.

i.e., $H_1: \mu \neq 98$ [Two Tailed Test]

Given that,

$n = 400$, $\bar{x} = 95$, $\mu = 98$, $\sigma = 12$

Under H_0 the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{95-98}{12/\sqrt{400}}\right| \sim N(0,1)$$

$$z = \left|\frac{-3}{12/20}\right| \sim N(0,1)$$

$$z = \left|\frac{-3}{0.6}\right|$$

$$z = |-5| = 5$$

$|z_{cal}| = 5$

$|z_{Tab}|$ value at 5% level of significance of right tailed test is 1.96.

∴ The calculated value of z is greaterthan the tabulated value of z

i.e., 5 > 1.96

∴ We reject H_0

Conclusion

The samples are not drawn from the population mean 98.

5. A sample of 150 items is taken from a population whose mean is 58 and standard deviation 10. If it be a random sample from a population with mean 60 and also calculate 95% confidence limits?

Sol :

Null Hypothesis (H_0)

The samples are drawn from the population mean 60,

i.e., $H_0 : \theta = 60$

Alternative Hypothesis (H_1)

The samples are not drawn from the population mean 60

i.e., $H_1 : \theta \neq 60$ [Two Tailed Test]

Given that,

$n = 150,\ \bar{x} = 58,\ \sigma = 10,\ \mu = 60$

Under H_0, the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{58-60}{10/\sqrt{150}}\right| \sim N(0,1)$$

$$z = \left|\frac{-2}{0.816}\right| \sim N(0,1)$$

$$z = 2.45$$

$|z_{cal}| = 2.45$

$|z_{tab}|$ value at 5% level of significance of two tailed test is 1.96.

∴ The calculated value of z is greaterthan the tabulated value of z.

∴ $z_{cal} > z_{tab}$

i.e., 2.45 > 1.96

∴ We reject H_0

Conclusion

The samples are not drawn from the population mean 60.

The 95% confidence limits for μ are:

$$\bar{x} \pm 1.96\ s/\sqrt{n}$$

$$[\bar{x} + 1.96(10/\sqrt{150}),\ \bar{x} - 1.96\ s/\sqrt{n}]$$

$$[58 + 1.96(10/\sqrt{150}),\ 58 - 1.96\ (10/\sqrt{150})]$$

$$[58 + 1.96(0.816),\ 58 - 1.96\ (0.816)]$$

$$[58 + 1.599,\ 58 - 1.599]$$

$$[59.599,\ 56,401]$$

∴ 95% confidence limits are [59.599, 56.401].

6. A sample of 450 meters as a mean 4.5 cm standard deviation 3.61 cm is a sample from a large population mean 4.25 cm as a standard deviation 3.61 cm.

Sol :

Null Hypothesis (H_0)

The samples are drawn from the population mean 4.25 cm

i.e., $H_0 : \mu = 4.25$ cm

Alternative Hypothesis (H_1)

The samples has not been drawn from the population mean

i.e., $H_1 : \mu > 4.25$ cm [Right Tailed Test]

Given that,

$n = 450,\ \bar{x} = 4.5,\ \mu = 4.25,\ \sigma = 3.61$

Under H_0, the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{4.5-4.25}{3.61/\sqrt{450}}\right| \sim N(0,1)$$

$$z = \left|\frac{0.25}{0.17}\right| \sim N(0,1)$$

$z = 1.47$

$|z_{cal}| = 1.47$

$|z_{tab}|$ value at 5% level of significance of right tailed test is 1.645.

$\therefore$ The calculated value of z is lessthan the tabulated value of z.

$\therefore\ |z_{cal}| < |z_{tab}|$

i.e., $1.47 < 1.645$

$\therefore$ We Accept H_0

Conclusion

The samples has been drawn from the population mean 4.25 cm.

---

7. A random sample of 100 articles are selected from batch of 2000 articles show that the average diagramatical of the article 0.354 with standard deviation 0.048. Find the 95% confidence limits for the average of these batch 2000 articles. Test whether the samples has been drawn from the population mean is 0.421.

Sol :

Null Hypothesis (H_0)

The samples has been drawn from the population mean is 0.421

i.e., $H_0 : \mu = 0.421$

Alternative Hypothesis (H_1)

The samples has not been drawn from the population mean,

i.e., $H_1 : \mu \neq 0.421$ [Tow Tailed Test]

Given that,

$n = 100,\ \bar{x} = 0.354,\ s = 0.048,\ \mu = 0.421$

Under H_0, the test statistic is given by,

$$z = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$z = \left|\frac{0.354-0.421}{0.048/\sqrt{100}}\right| \sim N(0,1)$$

$$z = \left|\frac{-0.0670}{0.0048}\right| \sim N(0,1)$$

$|z_{cal}| = 13.958$

$|z_{tab}|$ value at 5% level of significance of two tailed test is 1.96.

$\therefore$ The calculated value of z is greaterthan the tabulated value of z.

$\therefore\ |z_{cal}| > |z_{tab}|$

i.e., $13.958 > 1.96$

$\therefore$ We Reject H_0

Conclusion

The samples has not been drawn from the population mean

The 95% confidence limits for μ are

$$\bar{x} \pm 1.96 / s / \sqrt{n}$$

$$[\bar{x} + 1.96\ s/\sqrt{n}\ ,\ \bar{x} - 1.96\ s/\sqrt{n}\ ]$$

$$[\bar{x} + 1.96\left(\frac{0.048}{\sqrt{100}}\right),\ \bar{x} - 1.96\left(\frac{0.048}{\sqrt{100}}\right)]$$

$$[0.354 + 1.96(0.0048,\ 0.354 - 1.96(0.0048)]$$

$$[0.354 + 1.96(0.0048)\ \ - 0.354\ -0.0094]$$

$$[0.3634\ ,\ 0.3446]$$

$\therefore$ 95% confidence limits are [0.3634, 0.3446].

2.1.3 Difference of Means

Q4. Derive the large sample test procedure for difference of means.

Ans : **(Feb.-21, Imp.)**

Let x_1 be the mean of random sample of size n_1 from the population with mean μ_1 and variance σ_1^2.

Let x_2 be the mean of random sample of size n_2 from the population with mean μ_2 and variance σ_2^2.

i.e., $x_1 \sim N(\mu_1, \sigma_1^2)$ then $\bar{x}_1 \sim N(\mu_1, \sigma_1^2/n_1)$

and

$x_2 \sim N(\mu_2, \sigma_2^2)$ then $\bar{x}_2 \sim N(\mu_2, \sigma_2^2/n_2)$

Null Hypothesis (H_0)

There is no significance difference between two means,

i.e., $H_0: \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means.

i.e., $H_1: \mu_1 \neq \mu_2$

Level of Significance

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0,1) \text{ under } H_0$$

$$z = \left|\frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)}\right| \sim N(0, 1) \text{ under } H_0 \quad \text{... (1)}$$

Now consider,

$$E(\bar{x}_1 - \bar{x}_2) = E_1(\bar{x}_1) - E(\bar{x}_2)$$

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

$$E(\bar{x}_1 - \bar{x}_2) = 0 \quad \text{... (2)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{v(\bar{x}_1 - \bar{x}_2)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{v(\bar{x}_1) + v(\bar{x}_2)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{... (3)}$$

Sub Equation (2) and (3) in Equation (1)

$$z = \left|\frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

Remarks

1. If σ_1 and σ_2 are unknown, then we use the sample standard deviation s_1 and s_2 as $s_1 = \hat{\sigma}_1$, $s_2 = \hat{\sigma}_2$ then the test statistics is given by

$$z = \left|\frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

2. If $\sigma_1 = \sigma_2 = \sigma$ then the test statistics is given by,

$$z = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_2}{n_1} + \frac{\sigma_2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

Confidence Limits

(i) The 95% confidence limits for difference of two means is $(\bar{x}_1 - \bar{x}_2) \pm 1.96 \text{ S.E } (\bar{x}_1 - \bar{x}_2)$.

(ii) The 98% confidence limits for difference of two means is $(\bar{x}_1 - \bar{x}_2) \pm 2.33 \text{ S.E} (\bar{x}_1 - \bar{x}_2)$.

(iii) The 99% confidence limits for difference of two means is $(\bar{x}_1 - \bar{x}_2) \pm 2.58 \text{ S.E} (\bar{x}_1 - \bar{x}_2)$.

Conclusion

If the calculated value of $z(z_{cal})$ is lessthan the tabulated value of $z(z_{tab})$ at certain level of significance then we accept H_0 otherwise we reject H_0.

PROBLEMS

7. The average hourly wage of a sample of 150 workers in plant A was 2.56 Rs. With the standard deviation is 1.08 Rs. The average hourly wage of a sample of 200 workers in plant B was 2.87 Rs with the standard deviation is 1.28 Rs can an applicant safely assumed that the hourly wages paid by plant B are higher than the paid by plant A.

Sol :

Null Hypothesis (H_0)

There is no significance difference between plant A and plant B.

i.e., $H_0: \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

The average wage paid by plant B are higher than the plant A.

i.e., $H_1 : \mu_2 > \mu_1$ [Right Tailed Test]

Test Statistic

Given that,

$\bar{x}_1 = 2.56$ $\bar{x}_2 = 2.87$

$s_1 = 1.08$ $s_2 = 1.28$

$n_1 = 150$ $n_2 = 200$

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{2.56 - 2.87}{\sqrt{\frac{(1.08)^2}{150} + \frac{(1.28)^2}{200}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{-0.3100}{\sqrt{0.0078 + 0.0082}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-0.3100}{0.1265} \right| \sim N(0, 1)$$

$|z_{cal}| = 2.4506$

The tabulated value at 5% level of significance of right tailed test is 1.645.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $2.4506 > 1.645$

We reject H_0

Conclusion

The average wage paid by plant B are higher than the plant A.

8. A store keeper wanted to buy a large quantity of light bulbs from two brands A and B. A brought 100 bulbs from each brand and found to be testing that mean life time of 1120 hrs and standard deviation 75 hrs and the brand B as mean lifetime of 1062 hrs and standard deviation 80 hrs. Test whether difference of mean is significant.

Sol :

Null Hypothesis (H_0)

There is no significance difference between two brands i.e., $H_0 : \sigma_1 = \sigma_2$.

Alternative Hypothesis (H_1)

There is significance difference between two brands i.e., H_0: $\mu_1 \neq \mu_2$. [Two Tailed Test].

Test Statistic

Given That,

$\bar{x}_1 = 1120$, $\bar{x}_2 = 1062$

$s_1 = 75$, $s_2 = 80$

$n_1 = 100$, $n_2 = 100$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{1120 - 1062}{\sqrt{\frac{(75)^2}{100} + \frac{(80)^2}{100}}}\right| \sim N(0, 1)$$

$$z = \left|\frac{58}{\sqrt{56.2 + 64}}\right| \sim N(0, 1)$$

$$z = \left|\frac{58}{10.9636}\right| \sim N(0, 1)$$

$|z_{cal}| = 5.29$

The tabulated value at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $5.29 > 1.96$

We reject H_0

Conclusion

There is significance difference between two brands.

9. Data on days to maturity were recorded in two varieties of a pulse crop Determine whether two means are significantly different.

	n	Mean	Variance
Variety A	60	60	8.20
Variety B	65	65	11.13

Sol : **(May/June-18)**

Null Hypothesis (H_0)

There is no significance difference between two means.

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means.

i.e., $H_1 : \mu_1 \neq \mu_2$

Level of Significance (α)

Consider the appropriate at 5% level of significance.

Given that

$\bar{x}_1 = 60$ $\bar{x}_2 = 60$

$n_1 = 60$ $n_2 = 60$

$s_1^2 = 8.20$ $s_2^2 = 11.13$

$s_1 = 2.86$ $s_2 = 3.3362$

Test Statistics

Under the Null Hypothesis then the test statistics is given by,

$$z = \left| \frac{\bar{x}_1 + \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0.$$

$$z = \left| \frac{60 - 65}{\sqrt{\frac{(8.20)^2}{60} + \frac{(11.13)^2}{65}}} \right| \sim N(0, 1) \text{ under } H_0.$$

$$z = \left| \frac{-5}{\sqrt{(1.1207) + (1.9058)}} \right| \sim N(0, 1) \text{ under } H_0.$$

$$z = \left| \frac{-5}{1.7397} \right|$$

$$z = |-2.8741|$$

$$z = 2.8741$$

$$z_{cal} = 2.8741$$

The tabulated value at 5% L.O.S for two tailed test is 1.96.

The calculated value of z is greater than the tabulated value of z.

$\therefore \quad z_{cal} > z_{tab}$

$2.8741 > 1.96$

We reject H_0

Conclusion

There is significance difference between two means.

10. A random sample of 1200 men from one state gives the average pay as 400 Rs. per month with standard deviation 60 and 1000 men of another state gives the average pay as 500 RS. Per month with standard deviation 80. Test whether the mean levels of pay of the men from the two states different significance.

Sol :

Null Hypothesis (H_0)

There is no significance difference between pay of two states i.e., $H_0 : \mu_1 = \mu_2$.

Alternative Hypothesis (H_1)

There is significance difference between pay of two states i.e., $H_1 : \mu_1 \neq \mu_2$. [Two Tailed Test].

Test Statistic

Given That,

$\bar{x}_1 = 400$, $\bar{x}_2 = 500$

$s_1 = 60$, $s_2 = 80$

$n_1 = 1200$, $n_2 = 1000$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{400 - 500}{\sqrt{\frac{(60)^2}{1200} + \frac{(80)^2}{1000}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-100}{\sqrt{3 + 6.4}} \right| \sim N(0, 1)$$

$$z = \left| \frac{100}{3.0659} \right| \sim N(0, 1)$$

$$|z_{cal}| = 32.6$$

The tabulated value at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $32.6 > 1.96$

We reject H_0

Conclusion

There is significance difference between pay of two States.

11. A random sample of 500 the mean is found to be 20 another independent sample of 400 mean is 15 with the population standard deviation 4.

Sol : **(June-19)**

Null Hypothesis (H_0)

There is no significance difference between two mean,

i.e., $H_0 : \mu_1 = \mu_2$.

Alternative Hypothesis (H_1)

There is a significance difference between two mean,

i.e., $H_1 : \mu_1 \neq \mu_2$. [Two Tailed Test]

Test Statistic

Given That,

$\bar{x}_1 = 20$, $\bar{x}_2 = 15$

$n_1 = 500$, $n_2 = 400$

$\sigma = 4$,

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{20 - 15}{4\sqrt{\frac{1}{500} + \frac{1}{400}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{5}{4\sqrt{0.0045}} \right| \sim N(0, 1)$$

$$z = \left| \frac{5}{0.2683} \right| \sim N(0, 1)$$

$|z_{cal}| = 18.6$

The tabulated value at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $18.6 > 1.96$

We reject H_0

Conclusion

There is a significance difference between two mean.

---

12. A sample of heights of 6400 English man as mean of 67.85 and standard deviation 2.56 by the another sample of height of 1600 Australia man as mean 68.55 and standard deviation 2.52. The data integrate that the Australian man average is taller than Englishmen.

Sol :

Null Hypothesis (H_0)

There is no significance difference between heights of English men and Australian men,

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

The average height of Australian men is taller than English men.

i.e., $H_1 : \mu_2 > \mu_1$. [Right Tailed Test]

Test Statistic

Given That,

$\bar{x}_1 = 67.85$, $\bar{x}_2 = 68.55$

$s_1 = 2.56$, $s_2 = 2.52$

$n_1 = 6400$, $n_2 = 1600$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-0.7000}{\sqrt{0.0010 + 0.0040}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.7000}{0.0707} \right| \sim N(0, 1)$$

$|z_{cal}| = 9.90$

The tabulated value at 5% level of significance for Right Tailed Test is 1.645.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $9.90 > 1.645$

We reject H_0

Conclusion

The Average height of Australian men is taller than English men.

---

13. The mean of two single large samples are 1000 and 2000 numbers are 67.5 and 68.0 respectively then the sample is recorded is drawn from the samples, population standard deviation 2.5: Test 5% level of significance.

Sol :

Null Hypothesis (H_0)

There is no significance difference between heights of English men and Australian men,

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two large, sample,

i.e., $H_1 : \mu_1 > \mu_2$. Two Tailed Test

Test Statistic

Given That,

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68.0$

$n_1 = 100$, $n_2 = 2000$

$\sigma = 2.5$,

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{67.5 - 68.0}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-0.5}{2.5\sqrt{0.001 + 0.0005}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.5}{2.5\sqrt{0.0015}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.5}{0.0015} \right| \sim N(0, 1)$$

$|z_{cal}| = 5.165$

The tabulated value at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., $5.165 > 1.96$

We reject H_0

Conclusion

There is significance difference between two large samples.

Q5. Explain the procedure test for standard deviation.

Ans :

If s_1 and s_2 are standard deviation of two independent samples then.

Null Hypothesis (H_0)

There is no significance difference between two standard deviation.

i.e., $H_0 : \sigma_1 = \sigma_2$

Alternative Hypothesis (H_1)

There is significance difference between two standard deviation,

i.e., $H_1 : \sigma_1 \neq \sigma_2$

Level of Significance (α)

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left|\frac{t - E(t)}{S.E\,(t)}\right| \sim N(0, 1) \text{ Under H0}$$

$$z = \left|\frac{(s_1 - s_2) - E(s_1 - s_2)}{S.E\,(s_1 - s_1)}\right| \sim N(0, 1) \quad \ldots (1)$$

Now consider,

$$E(s_1 - s_2) = E(s_1) - E(s_2)$$
$$= \sigma_1 - \sigma_2$$
$$= \sigma_1 - \sigma_1 \quad (\therefore \sigma_2 = \sigma_1)$$
$$E(s_1 - s_2) = 0 \quad \ldots (2)$$

and $S.E(s_1 - s_2) = \sqrt{v(s_1 - s_2)}$

$$= \sqrt{v(s_1) + v(s_2)}$$

$$S.E(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}} \quad \ldots (3)$$

Sub Equation (2) and (3) in equation (1)

$$Z = \sqrt{\frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}} \sim N(0, 1) \text{ Under } H_0$$

Remarks

1. If population standard deviation σ_1 and σ_2 are unknown then we use sample standard deviation s_1 and s_2 as

$$s_1 = \hat{\sigma}_1 \quad , \quad s_2 = \hat{\sigma}_2$$

Then the test statistic is given by

$$z = \left|\frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

2. If the population standard deviations are known i.e., $\sigma_1 = \sigma_2 = \sigma$ then the test statistic is given by,

$$z = \left|\frac{s_1 - s_2}{\sqrt{\frac{\sigma^2}{2n_1} + \frac{\sigma^2}{2n_2}}}\right| \sim N(0, 1) \text{ Under } H_0$$

$$z = \left|\frac{s_1 - s_2}{\sigma\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

Confidence Limits

(i) The 95% confidence limits for standard deviation is $(s_1 - s_2) \pm 1.96$ S.E $(s_1 - s_2)$.

(ii) The 98% confidence limits for standard deviation is $(s_1 - s_2) \pm 2.33$ S.E. $(s_1 - s_2)$.

(iii) The 99% confidence limits for standard deviation is $(s_1 - s_2) \pm 2.58$ S.E $(s_1 - s_2)$.

Conclusion

If the calculated value of $z(z_{cal})$ is lessthan the tabulated value of $z(z_{tab})$ at certain level of significance then we accept H_0 otherwise we reject H_0.

PROBLEMS

14. A random samples are drawn from two countries given the following data relative to the.

Heights	Country-A	Country-B
Mean (height inches)	67.42	67.25
S.D (inches)	2.58	2.50
No. of Samples	1000	1200

Sol : **(Imp.)**

Given that,

$\bar{x}_1 = 67.42$, $\bar{x}_2 = 67.25$

$s_1 = 2.58$, $s_2 = 2.50$

$n_1 = 1000$, $n_2 = 1200$

Procedure 1

Test for two means

Null Hypothesis (H_0)

There is no significance difference between two means,

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means i.e., $H_1 : \mu_1 \neq \mu_2$ two tailed test.

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under H0}$$

$$z = \left| \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.50)^2}{1200}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.50)^2}{1200}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.1700}{\sqrt{0.0067 + 0.0052}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.1700}{0.0119} \right|$$

$|z_{cal}| = 1.56$

The tabulated value of 5% level of significance for two tailed test is 1.96.

The calculated value of z is lessthan the tabulated value of z.

$\therefore$ zcal < ztab

i.e., 1.56 < 1.96

We accept H_0

Conclusion

There is no significance difference between two means.

Procedure 2

Test for standard deviation

Null Hypothesis (H_0)

There is no significance difference between two standard deviation,

i.e., $H_0 : \sigma_1 = \sigma_2$

Alternative Hypothesis (H_1)

There is significance difference between two standard deviation.

i.e., $H_1 : \sigma_1 \neq \sigma_2$ [two tailed test].

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{2.58 - 2.50}{\sqrt{\frac{(2.58)^2}{2(1000)} + \frac{(2.50)^2}{2(1200)}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.08}{\sqrt{0.0033 + 0.0026}} \right| \sim N(0, 1)$$

$$z = \left| \frac{0.08}{0.07} \right| \sim N(0, 1)$$

$|z_{cal}| = 1.14$

The tabulated value of 5% level of significance for two tailed test is 1.96.

The calculated value of z is lessthan the tabulated value of z.

$\therefore z_{cal} < z_{tab}$

i.e., $1.14 < 1.96$

We accept H_0

Conclusion

There is no significance difference between two standard deviation.

15. A random samples drawn from two universities given the following data to the weights.

Weight	University-A	University-B
Mean	55	57
Standard deviation	10	15
No. of Samples	400	100

Samples of students we drawn from two universities and their weight in kgs, mean and standard deviation calculated. Test whether there is significance difference between mean and standard deviation.

Sol : **(Imp.)**

Given that,

$\bar{x}_1 = 55$, $\bar{x}_2 = 57$

$s_1 = 10$, $s_2 = 15$

$n_1 = 400$, $n_2 = 100$

Procedure 1

Test for two means

Null Hypothesis (H_0)

There is no significance difference between two means,

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means i.e., $H_1 : \mu_1 \neq \mu_2$ [two tailed test].

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{55 - 57}{\sqrt{\frac{(10)^2}{400} + \frac{(15)^2}{100}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-2}{\sqrt{0.25 + 2.25}} \right| \sim N(0, 1)$$

$$z = \left| \frac{2}{1.5811} \right| \sim N(0, 1)$$

$|z_{cal}| = 1.26$

The tabulated value of 5% level of significance for two tailed test is 1.96.

The calculated value of z is lessthan the tabulated value of z.

$\therefore z_{cal} < z_{tab}$

i.e., $1.26 < 1.96$

We accept H_0

Conclusion

There is no significance difference between two mean.

Procedure 2

Test for Standard Deviation

Null Hypothesis (H_0)

There is no significance difference between two standard deviation.

i.e., $H_0 : \sigma_1 = \sigma_2$

Alternative Hypothesis (H_1)

There is no significance difference between two standard deviation.

i.e., $H_1 : \sigma_1 \neq \sigma_2$

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right| \sim N(0, 1) \text{ under H0}$$

$$z = \left| \frac{10 - 15}{\sqrt{\frac{(10)^2}{2(400)} + \frac{(15)^2}{2(100)}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-5}{\sqrt{0.1250 + 1.1250}} \right| \sim N(0, 1)$$

$$z = \left| \frac{5}{1.11} \right| \sim N(0, 1)$$

$|z_{cal}| = 4.50$

The tabulated value of 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore z_{cal} > z_{tab}$

i.e., 450 > 1.96

We reject H_0

Conclusion

There is significance difference between two standard deviation.

16. In survey of income of two classes of workers to random samples given the following details. Test whether difference between mean and standard deviation.

	Sample - 1	Sample - 2
Mean	582	546
Standard deviation	24	28
No. of Samples	100	100

Sol :

Given that,

$\bar{x}_1 = 582$, $\bar{x}_2 = 546$

$s_1 = 24$, $s_2 = 28$

$n_1 = 100$, $n_2 = 100$

Procedure 1

Test for two means

Null Hypothesis (H_0)

There is no significance difference between two mean,

i.e., $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means

i.e., $H_1 : \mu_1 \neq \mu_2$ [Two Tailed Test].

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{582 - 546}{\sqrt{\frac{(24)^2}{100} + \frac{(28)^2}{100}}} \right| \sim N(0,\ 1)$$

$$z = \left| \frac{36}{\sqrt{15.7600 + 7.8400}} \right| \sim N(0,\ 1)$$

$$z = \left| \frac{36}{3.68} \right| \sim N(0,\ 1)$$

$|z_{cal}| = 9.78$

The tabulated value at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore \quad z_{cal} > z_{tab}$

i.e., $9.78 > 1.96$

We Reject H_0

Conclusion

There is significance difference between two mean.

Procedure 2

Test for Standard Deviation

Null Hypothesis (H_0)

There is no significance difference between two standard deviation.

i.e., $H_0: \sigma_1 = \sigma_2$

Alternative Hypothesis (H_1)

There is no significance difference between two standard deviation.

i.e., $H_1 : \sigma_1 \neq \sigma_2$ [Two Tailed Test]

Level of Significance (α)

Consider the appropriate at 5% level of significance for two tailed test is 1.96.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right| \sim N(0,\ 1) \text{ under } H_0$$

$$z = \left| \frac{24 - 28}{\sqrt{\frac{(24)^2}{200} + \frac{(28)^2}{200}}} \right| \sim N(0,\ 1)$$

$$z = \left| \frac{-4}{\sqrt{2.8800 + 3.9200}} \right| \sim N(0,\ 1)$$

$$z = \left| \frac{4}{2.6077} \right| \sim N(0,\ 1)$$

$|z_{cal}| = 1.53$

The tabulated value of at 5% level of significance for two tailed test is 1.96.

The calculated value of z is lessthan the tabulated value of z.

$\therefore z_{cal} < z_{tab}$

i.e., $1.53 < 1.96$

We Accept H_0

Conclusion

There is no significance difference between two standard deviation.

2.1.4 Single Sample Proportion

Q6. Write short notes on large sample test of single proportion.

(OR)

Describe the large sample test for single proportion.

Ans : **(Feb.-21)**

If 'x' is a number of success in 'n' independent trails, the constant probability of success (p) for each trail then,

$E(x) = np, \qquad V(x) = npQ$

where

$Q = 1 - p$ is the probability of failure.

It has been prove that the large sample size n the binomial distribution tends to normal distribution.

Hence for large sample size $X \sim N(np, npQ)$ then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{x - E(x)}{S.E(x)}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{x - np}{\sqrt{npQ}}\right| \sim N(0, 1) \text{ under } H_0$$

Remarks

In a sample of size 'n', let 'x' be the number of persons possessing the given attribute then the proportion of success is $p = \frac{x}{n}$ then the test statistic is given by.

$$z = \left|\frac{p - E(p)}{S.E(p)}\right| \sim N(0, 1) \text{ Under } H_0$$

$$z = \left|\frac{p - E\left(\frac{x}{n}\right)}{S.E\left(\frac{x}{n}\right)}\right| \sim N(0, 1) \text{ under } H_0 \quad ...(1)$$

Now consider,

$$E\left(\frac{x}{n}\right) = \frac{1}{n} E(x) \qquad V\left(\frac{x}{n}\right) = \frac{1}{n^2} V(x)$$

$$E\left(\frac{x}{n}\right) = \frac{1}{\not{n}} (\not{n}p) \qquad V\left(\frac{x}{n}\right) = \frac{1}{n^{\not{2}}} (\not{n}pQ)$$

$$E\left(\frac{x}{n}\right) = p \; ... (2) \qquad V\left(\frac{x}{n}\right) = \frac{pQ}{n}$$

$$S.E\left(\frac{x}{n}\right) = \sqrt{\frac{pQ}{n}} \; ... (3)$$

Sub Equation (2) and (3) in Equation (1)

$$z = \left|\frac{p - P}{\sqrt{\frac{PQ}{n}}}\right| \sim N(0, 1) \text{ Under } H_0$$

Confidence Limits

(i) The 95% confidence limits for the single proportion is,

$$nP \pm 1.96 \sqrt{npQ}$$

and

$$P \pm 1.96 \sqrt{\frac{PQ}{n}}.$$

(ii) The 98% confidence limits for the single proportion is,

$$nP \pm 2.33 \sqrt{npQ}.$$

and

$$P \pm 2.33 \sqrt{\frac{pQ}{n}}$$

(iii) The 99% confidence limits for the single proportion is,

$$nP \pm 2.58 \sqrt{npQ}.$$

and

$$P \pm 2.58 \sqrt{\frac{pQ}{n}}$$

Conclusion

If the calculated value of z is lessthan the tabulated value of z at certain level of significance then we accept H_0, otherwise we reject H_0.

PROBLEMS

17. A die is rolling 9000 times and rolling of a 3 and 4 is observed 3,240 times. Test whether the die is an unbiased and find the limits between which is the probability of throw 3 and 4 dies (z_{tab} = 3).

Sol :

Null Hypothesis (H_0)

The die is unbiased

$$H_0 : p(3) + p(4) = \frac{1}{6} + \frac{1}{6} = \frac{\not{2}^1}{\not{6}_3}$$

$$H_0 : p = \frac{1}{3}$$

Alternative Hypothesis (H_1)

The die is biassed

H_1: $p \neq \frac{1}{3}$ [Two Tailed Test]

Given that,

$n = 9,000$, $x = 3,240$

$p = \frac{1}{3}$, $Q = \frac{2}{3}$

$[\, Q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}\,]$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{x - np}{\sqrt{npQ}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{3,240 - 9,000\left(\frac{1}{3}\right)}{\sqrt{(9,000)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}}\right|$$

$$z = \left|\frac{3240 - 3000}{\sqrt{2000}}\right|$$

$$z = \left|\frac{240}{44.7213}\right|$$

$|z_{cal}| = 5.366$

The calculated value of z is greater than the tabulated value of z.

$\therefore$ $z_{cal} > z_{tab}$

i.e., $5.366 > 3$

$\therefore$ we reject H_0

Conclusion

The die is biased

Confidence Limits

$np \pm 3\sqrt{nPQ}$

$[nP + 3\sqrt{nPQ}\ ,\ np - 3\sqrt{nPQ}\,]$

$$\left[(9000)\left(\frac{1}{3}\right) + 3\sqrt{(9000)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}\ ,\ (9000)\left(\frac{1}{3}\right) - 3\sqrt{(9000)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}\right]$$

[3000 + 3(44.7214) , 3000 – 3 (44.7214)]

[3000 + 134.1642 , 3000 – 134.1642]

[3134.1642 , 2865.8358]

18. A coin is tossed 10,000 times and it turns of 5195 times. Test whether the coin may be regarded as unbiased one.

Sol :

Null Hypothesis (H_0)

The coin is regarded as unbiased.

$H_0 : p = \frac{1}{2}$

Alternative Hypothesis (H_0)

The coin is regarded as biassed.

$H_1 : p \neq \frac{1}{2}$ [Two Tailed Test]

Given that,

$x = 5195,\ n = 10,000,\ P = \frac{1}{2},\ Q = \frac{1}{2}$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{x - np}{\sqrt{npQ}}\right| \sim N(0, 1) \text{ Under } H_0$$

$$z = \left|\frac{5195 - (10000)\left(\frac{1}{2}\right)}{\sqrt{(10,000)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}}\right|$$

$$z = \left|\frac{5195 - 5000}{\sqrt{2500}}\right|$$

$$z = \left|\frac{195}{50}\right|$$

$|z_{cal}| = 3.9$

The calculated value of z is greaterthan the tabulated value of z.

$\therefore \quad z_{cal} > z_{tab}$

$3.9 > 3$

We reject H_0

Conclusion

The coin is regarded as biassed.

19. The die is thrown 3000 times and throw of 1 or 2 or 3 is observed 1240 times. Test whether the die is unbiased or not.

Sol :

Null Hypothesis (H_0)

The die is unbiased

$$H_0: p(1) + p(2) + p(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

$$= \frac{\cancel{3}^{1}}{\cancel{6}^{2}}$$

$$H_0: P = \frac{1}{2}$$

Alternative Hypothesis (H_1)

The die is biassed

$$H_1 : p \neq \frac{1}{2}$$

Given that,

$$x = 1240,\ n = 3000,\ p = \frac{1}{2},\ Q = \frac{1}{2}$$

Under the null Hypothesis then the test statistic is given by,

$$z = \left|\frac{x - np}{\sqrt{npQ}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{1240 - (3000)\left(\frac{1}{2}\right)}{\sqrt{(3000)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}}\right|$$

$$z = \left|\frac{1240 - 1500}{\sqrt{750}}\right|$$

$$z = \left|\frac{-260}{27.3861}\right|$$

$|z_{cal}| = 9.4939$

The calculated value of z is greaterthan the tabulated value of z

$\therefore \quad z_{cal} > z_{tab}$

i.e., $9.4939 > 3$

$\therefore$ we reject H_0

Conclusion

The die is biassed.

20. A random sample of 500 pineapples was taken from a large consignment and 65 5 they found to be bad. Test whether the hypothesis the promotion of bad one in a sample of the size is 0.015.

Sol :

Null Hypothesis (H_0)

The proportion of bad pineapples is,

$$H_0 : P = 0.015$$

Alternative Hypothesis (H_1)

The proportion of good pineapples is,

$$H_1 : P \neq 0.015$$

Given that,

$$x = 65,\ n = 500,\ P = 0.015,\ Q = 0.985$$

Under the null Hypothesis then the test statistic is given by,

$$z = \left|\frac{x - nP}{\sqrt{nPQ}}\right| \sim N(0, 1) \text{ Under } H_0$$

$$z = \left|\frac{65 - (500)(0.015)}{\sqrt{(500)(0.015)(0.985)}}\right|$$

$$z = \left|\frac{65 - 7.5000}{\sqrt{7.3875}}\right|$$

$$z = \left|\frac{57.5000}{\sqrt{2.7180}}\right|$$

$|z_{cal}| = 21.1553$

The calculated value of z is greaterthan the tabulated value of z.

$\therefore \quad z_{cal} > z_{tab}$

i.e., 21.1553 > 3

$\therefore$ We Reject H_0

Conclusion

The proportion of pineapples is good.

2.1.5 Difference of Proportions

Q7. Describe the test procedure for difference of proportions.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Let x_1, x_2 are the number of persons possessing given attribute A.

In a random sample of sizes n_1 and n_2 from the two populations respectively then the sample proportions are given by,

$$p_1 = \frac{x_1}{n_1} \quad , \quad p_2 = \frac{x_2}{n_2}$$

If p_1 and p_2 are the population proportion then,

$E(p_1) = P_1$, $E(p_2) = P_2$

and

$$V(p_1) = \frac{P_1Q_1}{n_1} \quad , \quad v(p_2) = \frac{P_2Q_2}{n_2}$$

Null Hypothesis (H_0)

There is no significance difference between two proportions.

$H_0: P_1 = P_2$

Alternative Hypothesis (H_1)

There is significance difference between two proportion.

$H_1: P_1 \neq P_2$

Level of Significance

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{(p_1 - p_2) - E(p_1 - p_2)}{S.E\,(p_1 - p_2)}\right| \sim N(0, 1)$$

under H_0 ... (1)

Now consider,

$E(p_1 - p_2) = E(p_1) - E(p_2)$

$= P_1 - P_2$

$= P_1 - P_1$

$E(p_1 - p_2) = 0$ ... (2)

$S.E(p_1 - p_2) = \sqrt{V(p1 - p2)}$

$= \sqrt{V(p_1) + V(p_2)}$

$$S.E\,(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{p_2Q_2}{n_2}} \quad \text{... (3)}$$

Sub equations eq (2) and (3) in (1)

$$z = \left|\frac{(p_1 - p_2) - 0}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{p_1 - p_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}\right| \sim N(0, 1) \text{ under } H_0$$

Remarks

1. If $p_1 = p_2 = P$, $Q_1 = Q_2 = Q$ then the test statistics is given by,

$$z = \left|\frac{p_1 - p_2}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}}\right|$$

$$z = \left| \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim N(0, 1) \text{ under } H_0$$

2. In general we don't have any information for the population proportion then we calculate the estimated value of $\hat{P}$ and $\hat{Q}$ then the test statistic is given by,

$$z = \left| \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim N(0, 1) \text{ under } H_0$$

where $\hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$ and $\hat{Q} = 1 - \hat{P}$

Confidence Limits

(i) The 95% confidence limits for the difference of two proportions are;

$(p_1 - p_2) \pm 1.96 \text{ S.E}(p_1 - p_2)$

(ii) The 98% confidence limits for the difference of two proportions are;

$(p_1 - p_2) \pm 2.33 \text{ S.E } (p_1 - p_2)$

(iii) The 99% confidence limits for the difference of two proportion are

$(p_1 - p_2) \pm 2.58 \text{ S.E } (p_1 - p_2)$

Conclusion

If the calculated value of z is lessthan the tabulated value of z at certain level of significance then we accept H_0 otherwise we reject H_0.

PROBLEMS

21. A random sample of 400 men and 600 women where asked whether they would like to have flyover near these residence 200 men and 325 women were in the favour of the proposal. Test the hypothesis that the proportions of men and women were in favour of the proposal are same significance that they are not.

Sol :

Null Hypothesis (H_0)

There is no significance difference between the men and women in favour of the proposal.

i.e., $H_0 : P_1 = P_2$

Alternative Hypothesis (H_1)

There is significance difference between the men and women in favour of proposal.

i.e., $H_1 : P_1 \neq P_2$

Level of Significance

Consider the appropriate at α% level of significance.

Test Statistic

Given that,

$n_1 = 400$, $n_2 = 600$

$x_1 = 200$, $x_2 = 325$

then $p_1 = \frac{x_1}{n_1} = \frac{200}{400} = \frac{1}{2}$

$p_1 = 0.5$

$p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.541$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim N(0,1)$$

Where,

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\hat{p} = \frac{(400)(0.5) + (600)(0.541)}{400 + 600}$$

$$\hat{p} = \frac{200+324}{1000}$$

$$\hat{p} = \frac{524}{1000}$$

$$\hat{p} = 0.524$$

$$\hat{Q} = 1 - \hat{p}$$

$$= 1 - 0.524$$

$$\hat{Q} = 0.476$$

Sub $\hat{p}, \hat{Q}$ values in – z

$$z = \left| \frac{0.5 - 0.541}{\sqrt{(0.524)(0.476)\left(\frac{1}{400}+\frac{1}{600}\right)}} \right| \sim N(0, 1)$$

$$z = \left| \frac{-0.0410}{\sqrt{(0.2494)\,(0.0042)}} \right|$$

$$z = \left| \frac{0.0410}{\sqrt{0.0010}} \right|$$

$$z = \left| \frac{0.0410}{0.0316} \right|$$

$|z_{cal}| = 1.2975$

The tabulated value of z at 5% level of significance for two tailed test is 1.96.

The calculated value of z is lessthan the tabulated value of z.

$\therefore \quad z_{cal} < ztab$

i.e., $1.2975 < 1.96$

$\therefore$ We Accept H_0

Conclusion

There is no significance difference between the men and women in favour of the proposal.

22. 1000 apples kept under the one type of the storage were found to be showing rating to the extent of 4%. 1500 apples kept under the another type of storage showed 3% rating can it be reasonable to conclude that time second type of storage, is superior that the first type of storage?

Sol :

Null Hypothesis (H_0)

There is no significance difference between two types of storages i.e.,

$$H_0 : P_1 = P_2$$

Alternative Hypothesis (H_1)

The second type of storage is superior than the first type of storage.

i.e., $H_1 : P_2 > P_1$ [Right Tailed Test]

Level of Significance

Consider the appropriate at 5% level of significance.

Test Statistic

Given that,

$n_1 = 1000$, $n_2 = 1500$

$p_1 = 4\% \Rightarrow p_1 = 4/100$

$p_1 = 3\% \Rightarrow p_2 = 3/100$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \right| \sim N(0,1) \text{ under } H_0$$

Where,

$$\hat{p} = \frac{n_1p_2 + n_2p_2}{n_1 + n_2}$$

$$\hat{p} = \frac{(1000)\left(\frac{4}{100}\right)+(1500)\left(\frac{3}{100}\right)}{1000+1500}$$

$$\hat{p} = \frac{40+45}{2500}$$

$$\hat{p} = \frac{85}{2500}$$

$$\hat{p} = 0.034$$

$$\hat{Q} = 1 - \hat{p}$$

$$\hat{Q} = 1 - 0.034$$

$$\hat{Q} = 0.966$$

Sub $\hat{p}, \hat{Q}$ values in z

$$z = \left|\frac{\frac{4}{100}-\frac{3}{100}}{\sqrt{(0.034)(0.966)\left(\frac{1}{1000}+\frac{1}{1500}\right)}}\right| \sim N(0, 1)$$

$$z = \left|\frac{0.0100}{\sqrt{(0.034)\,(0.0017)}}\right| \sim N(0, 1)$$

$$z = \left|\frac{0.0100}{0.0075}\right|$$

$$|z_{cal}| = 1.33$$

The tabulated value of z at 5% level of significance for Right tailed test is 1.645.

$\therefore \quad z_{cal} < z_{tab}$

i.e., $1.33 < 1.645$

$\therefore$ We Accept H_0

Conclusion

There is no significance different between two types of storages.

23. Before increases and excess duty on a tea 800 persons out of a sample of 1000 persons where found to be tea drinkers. After increasing and excess duty of 800 persons in significant decreases the consumption of the tea after increasing excess duty.

Sol :

Null Hypothesis (H_0)

There is no significance difference between the consumption of tea after and before increasing the excess duty i.e.,

$$H_0 : P_1 = P_2$$

Alternative Hypothesis (H_1)

There is significance difference between the consumption of tea after and before increasing the excess duty.

i.e., $H_1 : P_1 \neq P_2$ [Two Tailed Test]

Level of Significance

Consider the appropriate at 5% level of significance.

Test Statistic

Given that,

$n_1 = 1000$, $n_2 = 1200$

$x_1 = 800$, $x_2 = 800$

$$\text{Then } p_1 = \frac{x_1}{n_1} = \frac{800}{1000} = 0.8$$

$$p_2 = \frac{x_2}{n_2} = \frac{800}{1200} = 0.66$$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{p_1 - p_2}{\sqrt{\hat{p}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\right| \sim N(0,1) \text{ under } H_0$$

Where,

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

$$\hat{p} = \frac{(1000)\,(0.8)+(1200)(0.66)}{1000+1200}$$

$$\hat{p} = \frac{800+792}{2200}$$

$$\hat{p} = \frac{1592}{2200}$$

$$\hat{p} = 0.7236$$

$$\hat{Q} = 1 - \hat{p}$$

$$\hat{Q} = 1 - 0.7236$$

$$\hat{Q} = 0.2764$$

Sub $\hat{p}, \hat{Q}$ values in z

$$z = \left|\frac{0.8-0.66}{\sqrt{(0.7236)(0.2764)\left(\frac{1}{1000}+\frac{1}{1200}\right)}}\right| \sim N(0, 1)$$

$$z = \left|\frac{0.1400}{\sqrt{(0.2000)(0.0018)}}\right| \sim N(0, 1)$$

$$z = \left|\frac{0.1400}{0.0190}\right|$$

$$|z_{cal}| = 7.3684$$

The tabulated value of z at 5% level of significance for two tailed test is 1.96.

The calculated value of z is greaterthan the tabulated value of z.

$\therefore \quad z_{cal} > z_{tab}$

i.e., $7.3684 > 1.96$

$\therefore$ We Reject H_0

Conclusion

There is significance difference between the consumption of tea after and before increasing the excess duty.

24. A company has the head office at calculate at a branch of bombay the personal director wanted to know if the workers at the two places would like the introduction of the new plan of a survey was conducted for this purpose out of 1500 workers at calcutta 62% favoured the new plan and at bombay out of 400 workers 59% were against the new plan is there any significance difference between the two groups in their attributes towards the new plan.

Sol :

Null Hypothesis (H_0)

There is no significance difference between the two groups in their attributes.

i.e., $H_0 : P_1 = P_2$

Alternative Hypothesis (H_1)

There is significance difference between the two groups in their attributes.

i.e., $H_1 : P_1 \neq P_2$ [Two Tailed Test]

Level of Significance

Consider the appropriate at 5% level of significance.

Test Statistic

Given that,

$n_1 = 1500 \quad , \quad n_2 = 400$

$p_1 = 62\% \quad \Rightarrow \quad p_1 = \frac{62}{100}$

$p_2 = 59\% \quad \Rightarrow \quad p_2 = \frac{59}{100}$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\right| \sim N(0,1) \text{ under } H_0$$

Where,

$$\hat{p} = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$\hat{P} = \frac{\left(\frac{62}{100}\right)(1500)+\left(\frac{59}{100}\right)(400)}{500+400}$$

$$\hat{P} = \frac{930+236}{1900}$$

$$\hat{P} = \frac{1166}{1900}$$

$$\hat{P} = 0.6137$$

$$\hat{Q} = 1 - \hat{p}$$

$$\hat{Q} = 1 - 0.6137$$

$$\hat{Q} = 0.3863$$

Sub $\hat{p}, \hat{Q}$ values in z

$$z = \left|\frac{\frac{62}{100}-\frac{59}{100}}{\sqrt{(0.6137)(0.3863)\left(\frac{1}{1500}+\frac{1}{400}\right)}}\right| \sim N(0, 1)$$

$$z = \left|\frac{0.0300}{\sqrt{(0.2371)\,(0.0032)}}\right|$$

$$z = \left|\frac{0.0300}{0.0275}\right|$$

$$|z_{cal}| = 1.0909$$

The tabulated value of z at 5% level of significance for two tailed test is 1.96.

$\therefore \quad z_{cal} < z_{tab}$

i.e., $1.0409 < 1.96$

$\therefore$ We Accept H_0

Conclusion

There is no significance difference between the two groups in their attributes.

2.2 Fisher's Z-transformation

Q8. Explain Fisher's Z-transformation. And its applications.

Ans : **(June-19, June-18, Imp.)**

The t-test is used for testing the significance of r(observed sample correlation coefficient) from bi-variate normal population with p = 0 (i.e., uncorrelated). If it is used when $p \neq 0$ which is,

$$z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)$$

It can be expressed in a trigonometric form as,

$$z = \tan h^{-1} r.$$

With this, the distribution of 'Z' is approximately constant.

$$\varepsilon = \frac{1}{2}\log_e\left(\frac{1+p}{1-p}\right)$$

$$= \tan h^{-1} p$$

In case of large sample size (e.g.: $n > 50$) the distribution is approximately normal and the variance is given by $\frac{1}{n-3}$.

Applications of Z-transformation

- To test whether there exist significant difference between two sample coefficients or not.
- To test whether there exist significant difference between two sample coefficients or not.
- Different coefficient of correlation from various sample can be combined using z-transformation.

2.2.1 Test for Single Correlation Coefficient

Q9. Derive the test procedure for Single Correlation for Large Sample.

Ans :

Let $(x_1, y_1), (x_2, y_2)$ ..... (x_n, y_n) be a 'n' observations drawn from a bivariate normal population with population correlation coefficients and sample correlation coefficient 'r'.

Null Hypothesis (H_0)

There is no significance difference between population correlation coefficient and sample correlation coefficient

i.e., $H_0 : \rho = \rho_0$

Alternative Hypothesis (H_1)

There is significance difference between population correlation coefficient and sample correlation coefficient i.e., $H_1 : \rho \neq \rho_0$

Level of Significance (α)

Consider the appropriate at α% level of significance.

Fisher suggested the z-transformations for sample correlation coefficient is

$$z_r = \frac{1}{2} \log_e \left[\frac{1+r}{1-r}\right] = \text{Tanh}^{-1}(r)$$

and population correlation coefficient is,

$$z_\rho = \frac{1}{2} \log_e \left[\frac{1+\rho}{1-\rho}\right] = \text{Tanh}^{-1}(\rho)$$

and variance $= \frac{1}{n-3}$ [for large samples].

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0, 1)$$

$$z = \left|\frac{z_r - z_\rho}{\sqrt{1/n-3}}\right| \sim N(0, 1)$$

Confidence Limits

(i) The 95% confidence limits for single correlation coefficient is $z_r \pm 1.96 \sqrt{\frac{1}{n-3}}$.

(ii) The 98% confidence limits for single correlation coefficient is $z_r \pm 2.33 \sqrt{\frac{1}{n-3}}$.

(iii) The 99%confidence limits for single correlation coefficient is $z_r \pm 2.58 \sqrt{\frac{1}{n-3}}$.

Conclusion

The calculated value of z is lessthan the tabulated value of z at certain level of significance then we accept H_0 otherwise we reject H_0.

PROBLEMS

25. A random sample of 30 pairs of observation given a correlation coefficient 0.61 can it be regarded as drawn from a bivariate normal population having correlation coefficient is 0.7. Also compute the 95% confidence limits for the population correlation coefficient.

Sol : **(Imp.)**

Null Hypothesis (H_0)

The population correlation coefficient is 0.7 i.e., $H_0 : \rho = 0.7$.

Alternative Hypothesis (H_1)

The population correlation coefficient is not 0.7 i.e., $H_1 : \rho \neq 0.7$ [Two Tailed Test.]

Level of Significance

Consider the appropriate at 5% level of significance for two-tailed test is 1.96.

Test Statistic

Given that,

$n = 30, \rho = 0.7, r = 0.61$

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{z_r - z_\rho}{\sqrt{\frac{1}{n-3}}}\right| \sim N(0, 1) \text{ under } H_0$$

Where,

$$z_r = \frac{1}{2} \log_e \left(\frac{1+r}{1-r}\right)$$

$$z_r = \frac{1}{2}\log_e\left(\frac{1+0.61}{1-0.61}\right)$$

$$z_r = \frac{1}{2}\log_e\left(\frac{1.6100}{0.3900}\right)$$

$$z_r = \frac{1}{2}\log_e(4.128)$$

$$z_r = \frac{1}{2}(0.6157).$$

$$z_r = 0.3079$$

$$z_\rho = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)$$

$$= \frac{1}{2}\log_e\left(\frac{1+0.7}{1-0.7}\right)$$

$$= \frac{1}{2}\log_e\left(\frac{1.7000}{0.3000}\right)$$

$$= \frac{1}{2}\log_e(5.6667)$$

$$= \frac{1}{2}(0.7533)$$

$$z_\rho = 0.376$$

Sub z_r, z_ρ values in z

$$z = \left|\frac{0.307-0.376}{\sqrt{\frac{1}{30-3}}}\right|$$

$$z = \left|\frac{0.0690}{0.1925}\right|$$

$$|z_{cal}| = 0.3584$$

The tabulated value at 5% level of significance is 1.96.

$\therefore z_{cal} < ztab$

i.e., $0.3584 < 1.96$

$\therefore$ we accept H_0

Conclusion

The population correlation coefficient is 0.7.

2.2.2 Test for Two Correlation Coefficient

Q10. Describe the test procedure for Fisher's Z-transformation for differences of Correlation Coefficient.

Ans : **(Feb.-21, Imp.)**

Let r_1, r_2 be the sample correlation coefficient observed in two independent samples of sizes n_1 and n_2 respectively then

$$z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_1}\right)$$

$$z_2 = \frac{1}{2}\log_e\left(\frac{1+r_2}{1-r_2}\right)$$

Null Hypothesis (H_0)

There is no significance difference between two correlation coefficient.

i.e., $H_0 : \rho_1 = \rho_2$

Alternative Hypothesis (H_1)

There is significance difference between two correlation coefficient.

i.e., $H_1 : \rho_1 \neq \rho_2$

Level of Significance

Consider the appropriate at a% level of significance.

Test Statistic

Under the null hypothesis then the test statistic is given by

$$z = \left|\frac{(z_1-z_2)-E(z_1-z_2)}{S.E(z_1-z_2)}\right| \sim N(0, 1) \text{ under } H_0 \quad \text{... (1)}$$

Now consider,

$$E(z_1 - z_2) = E(z_1) - E(z_2)$$
$$= \rho_1 - \rho_2$$
$$= \rho_1 - \rho_1$$
$$E(z_1 - z_2) = 0 \quad \text{... (2)}$$

$$S.E(z_1 - z_2) = \sqrt{v(z_1-z_2)}$$
$$= \sqrt{v(z_1)+v(z_2)}$$

$$S.E(z_1 - z_2) = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad \text{... (3)}$$

Sub Equation (2) and (3) in (1)

$$z = \left| \frac{(z_1 - z_2) - 0}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \right| \sim N(0, 1)$$

$$z = \left| \frac{(z_1 - z_2)}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \right| \sim N(0, 1) \text{ under } H_0$$

Confidence Limits

(i) The 95% confidence limits for two correlation coefficient is,

$$(z_1 - z_2) \pm 1.96 \, S.E(z_1 - z_2)$$

(or)

$$(z_1 - z_2) \pm 1.96 \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

(ii) The 98% confidence limits for two correlation coefficient is,

$$(z_1 - z_2) \pm 2.33 \, S.E \, (z_1 - z_2)$$

(or)

$$(z_1 - z_2) \pm 2.33 \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

(iii) The 99% confidence limits for two correlation coefficient is,

$$(z_1 - z_2) \pm 2.58 \, S.E(z_1 - z_2)$$

(or)

$$(z_1 - z_2) \pm 2.58 \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

Conclusion

The calculated value of z is lessthan the tabulated value of z then we accept H_0 otherwise we reject H_0.

2.3 Definition of Order Statistics and Statement of their Distributions

Q11. Define order statistics and state their distributions.

Ans : **(June-19, June-18, Imp.)**

If a_1, a_2, a_3 .... a_n represents the n independent and identical distributed variates of cummulative distribution F(a), then the r[th] order statistics of a denoted by ar is given by,

$$a_r = a_{(1)} < a_{(2)} < a_{(3)} < a_{(4)} < \ldots\ldots a_{(n)}$$

Where,

r = 1, 2, 3, 4, 5 .... n is the order statistic of $a_{(r)}$.

Distribution

i) If $b_1 = a_{(1)}$ represents the smallest order statistics of a_1, a_2, a_3 ..... a_n then its cummulative distribution function $f_n(a)$ is given as,

$$F_1(a) = 1 - [1 - F(a)]^n$$

ii) If $b_r = a_{(r)}$ represents the r[th] order statistic (ar) of a_1, a_2, a_3......a_n then its cummulative distribution $F_r(a)$ is given as,

$$F_r(a) = P[x_{(a)} \le x]$$

$$= \sum_{j=r}^{n} \binom{n}{j} F_{(a)}^{j} [1 - F(a)^{n-1}]$$

iii) If $b_n = a_{(n)}$ represents the largest order statistic $(x_{(n)})$ of a_1, a_2, a_3, ... a_n then its cummulative distribution $F_n(a)$ is given as,

$$F_n(a) = [F(a)]^n.$$

Short Question and Answers

1. Define Large Sample Test.

Ans :

The sample size $n \geq 30$ then it is called large sample test. The different large sample tests are :

(i) Test for single mean

(ii) Test for two means

(iii) Test for standard deviation

(iv) Test for single proportion

(v) Test for two proportions

(vi) Test for correlation coefficient.

2. Explain the Procedure for Testing of Hypothesis.

Ans :

The various steps in testing of statistical hypothesis in a systematic manner.

Null Hypothesis (H_0)

Set up the Null hypothesis H_0.

Alternative Hypothesis (H_1)

Set up the alternative hypothesis H_1. This test is decides whether we have to use one tailed test or two tailed test.

Level of Significance (α)

Choose the appropriate level of significance (α) depending on the reliability of the estimator values.

Test Statistic

Compute the test statistic,

$$Z = \left|\frac{t - E(t)}{SE(t)}\right| \sim N(0,1) \text{ under } H_0.$$

Conclusion

Compare the calculated value of z in step 4 with the significant value (tabulated value z) at given level of significance if $|z_{cal}| < z_{tab}$ then we accept H_0 otherwise we reject H_0.

3. Describe the test procedure for difference of proportions.

Ans :

Let x_1, x_2 are the number of persons possessing given attribute A.

In a random sample of sizes n_1 and n_2 from the two populations respectively then the sample proportions are given by,

$$P_1 = \frac{x_1}{n_1} \quad , \quad P_2 = \frac{x_2}{n_2}$$

If p_1 and p_2 are the population proportion then,

$$E(p_1) = P_1 \quad , \quad E(p_2) = P_2$$

and

$$V(p_1) = \frac{P_1Q_1}{n_1} \quad , \quad v(p_2) = \frac{P_2Q_2}{n_2}$$

Null Hypothesis (H_0)

There is no significance difference between two proportions.

$$H_0 : P_1 = P_2$$

Alternative Hypothesis (H_1)

There is significance difference between two proportion.

$$H_1 : P_1 \neq P_2$$

Level of Significance

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{t - E(t)}{S.E(t)}\right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left|\frac{(p_1 - p_2) - E(p_1 - p_2)}{S.E\,(p_1 - p_2)}\right| \sim N(0, 1)$$

under H_0 ... (1)

Now consider,

$$E(p_1 - p_2) = E(p_1) - E(p_2)$$
$$= P_1 - P_2$$
$$= P_1 - P_1$$
$$E(p_1 - p_2) = 0 \quad \text{... (2)}$$
$$S.E(p_1 - p_2) = \sqrt{V(p1 - p2)}$$
$$= \sqrt{V(p_1) + V(p_2)}$$
$$S.E\,(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{p_2Q_2}{n_2}} \quad \text{... (3)}$$

Sub equations eq (2) and (3) in (1)

$$z = \left| \frac{(p_1 - p_2) - 0}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{p_1 - p_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

4. Derive the large sample test procedure for difference of means.

Ans :

Let x_1 be the mean of random sample of size n_1 from the population with mean μ_1 and variance σ_1^2.

Let x_2 be the mean of random sample of size n_2 from the population with mean μ_2 and variance σ_2^2.

i.e., $x_1 \sim N(\mu_1, \sigma_1^2)$ then $\bar{x}_1 \sim N(\mu_1,\ \sigma_1^2/n_1)$

and

$x_2 \sim N(\mu_2, \sigma_2^2)$ then $\bar{x}_2 \sim N(\mu_2,\ \sigma_2^2/n_2)$

Null Hypothesis (H_0)

There is no significance difference between two means,

i.e., $H_0: \mu_1 = \mu_2$

Alternative Hypothesis (H_1)

There is significance difference between two means.

i.e., $H_1: \mu_1 \neq \mu_2$

Level of Significance

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left| \frac{t - E(t)}{S.E(t)} \right| \sim N(0,1) \text{ under } H_0$$

$$z = \left| \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.E(\bar{x}_1 - \bar{x}_2)} \right| \sim N(0,\ 1)$$

under H_0 ... (1)

Now consider,

$$E(\bar{x}_1 - \bar{x}_2) = E_1(\bar{x}_1) - E(\bar{x}_2)$$
$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$
$$E(\bar{x}_1 - \bar{x}_2) = 0 \quad \text{... (2)}$$
$$S.E\,(\bar{x}_1 - \bar{x}_2) = \sqrt{v(\bar{x}_1 - \bar{x}_2)}$$
$$S.E\,(\bar{x}_1 - \bar{x}_2) = \sqrt{v(\bar{x}_1) + v(\bar{x}_2)}$$
$$S.E\,(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{... (3)}$$

Sub Equation (2) and (3) in Equation (1)

$$z = \left| \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| \sim N(0, 1) \text{ under } H_0$$

5. Write short notes on large sample test of single proportion.

Ans :

If 'x' is a number of success in 'n' independent trails, the constant probability of success (p) for each trail then,

$$E(x) = np, \qquad V(x) = npQ$$

where

$$Q = 1 - p \text{ is the probability of failure.}$$

It has been prove that the large sample size n the binomial distribution tends to normal distribution.

Hence for large sample size X ~ N (np, npQ) then the test statistic is given by,

$$z = \left| \frac{t - E(t)}{S.E(t)} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{x - E(x)}{S.E(x)} \right| \sim N(0, 1) \text{ under } H_0$$

$$z = \left| \frac{x - np}{\sqrt{npQ}} \right| \sim N(0, 1) \text{ under } H_0$$

6. Explain the procedure test for standard deviation.

Ans :

If s_1 and s_2 are standard deviation of two independent samples then.

Null Hypothesis (H_0)

There is no significance difference between two standard deviation.

i.e., $H_0 : \sigma_1 = \sigma_2$

Alternative Hypothesis (H_1)

There is significance difference between two standard deviation,

i.e., $H_1: \sigma_1 \neq \sigma_2$

Level of Significance (α)

Consider the appropriate at α% level of significance.

Test Statistic

Under the Null Hypothesis then the test statistic is given by

$$z = \left| \frac{t - E(t)}{S.E\,(t)} \right| \sim N(0, 1) \text{ Under H0}$$

$$z = \left| \frac{(s_1 - s_2) - E(s_1 - s_2)}{S.E\,(s_1 - s_1)} \right| \sim N(0, 1) \qquad \text{... (1)}$$

Now consider,

$$E(s_1 - s_2) = E(s_1) - E(s_2)$$
$$= \sigma_1 - \sigma_2$$
$$= \sigma_1 - \sigma_1 \quad (\because \sigma_2 = \sigma_1)$$
$$E(s_1 - s_2) = 0 \qquad \text{... (2)}$$

and $S.E(s_1 - s_2) = \sqrt{v(s_1 - s_2)}$

$$= \sqrt{v(s_1) + v(s_2)}$$

$$S.E(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}} \qquad \text{... (3)}$$

Sub Equation (2) and (3) in equation (1)

$$Z = \sqrt{\frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}} \sim N(0, 1) \text{ Under } H_0$$

7. Explain Fisher's Z-transformation. And its applications.

Ans :

The t-test is used for testing the significance of r(observed sample correlation coefficient) from bi-variate normal population with p = 0 (i.e., uncorrelated). If it is used when p $\neq$ 0 which is,

$$z = \frac{1}{2} \log_e \left(\frac{1 + r}{1 - r} \right)$$

It can be expressed in a trigonometric form as,

$$z = \tan h^{-1} r.$$

With this, the distribution of 'Z' is approximately constant.

$$\varepsilon = \frac{1}{2}\log_e\left(\frac{1+p}{1-p}\right)$$

$$= \tan h^{-1} p$$

In case of large sample size (e.g.: n>50) the distribution is approximately normal and the variance is given by $\frac{1}{n-3}$.

8. Applications of Z-transformation

Ans :

- To test whether there exist significant difference between two sample coefficients or not.
- To test whether there exist significant difference between two sample coefficients or not.
- Different coefficient of correlation from various sample can be combined using z-transformation.

9. Derive the test procedure for Single Correlation for Large Sample.

Ans :

Let $(x_1, y_1), (x_2, y_2) ..... (x_n, y_n)$ be a 'n' observations drawn from a bivariate normal population with population correlation coefficient 's' and sample correlation coefficient 'r'.

Null Hypothesis (H_0)

There is no significance difference between population correlation coefficient and sample correlation coefficient

i.e., $H_0 : \rho = \rho_0$

Alternative Hypothesis (H_1)

There is significance difference between population correlation coefficient and sample correlation coefficient i.e., $H_1: \rho \neq \rho_0$

Level of Significance (α)

Consider the appropriate at α% level of significance.

Fisher suggested the z-transformations for sample correlation coefficient is

$$z_r = \frac{1}{2}\log_e\left[\frac{1+r}{1-r}\right] = \text{Tanh}^{-1}(r)$$

and population correlation coefficient is,

$$z_\rho = \frac{1}{2}\log_e\left[\frac{1+\rho}{1-\rho}\right] = \text{Tanh}^{-1}(\rho)$$

and variance = $\frac{1}{n-3}$ [for large samples].

Test Statistic

Under the Null Hypothesis then the test statistic is given by,

$$z = \left|\frac{t-E(t)}{S.E(t)}\right| \sim N(0, 1)$$

$$z = \left|\frac{z_r - z_\rho}{\sqrt{1/n-3}}\right| \sim N(0, 1)$$

Confidence Limits

(i) The 95% confidence limits for single correlation coefficient is $z_r \pm 1.96\sqrt{\frac{1}{n-3}}$.

(ii) The 98% confidence limits for single correlation coefficient is $z_r \pm 2.33\sqrt{\frac{1}{n-3}}$.

(iii) The 99%confidence limits for single correlation coefficient is $z_r \pm 2.58\sqrt{\frac{1}{n-3}}$.

Conclusion

The calculated value of z is lessthan the tabulated value of z at certain level of significance then we accept H_0 otherwise we reject H_0.

10. Describe the test procedure for Fisher's Z-transformation for differences of Correlation Coefficient.

Ans :

Let r_1, r_2 be the sample correlation coefficient observed in two independent samples of sizes n_1 and n_2 respectively then

$$z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_1}\right)$$

$$z_2 = \frac{1}{2}\log_e\left(\frac{1+r_2}{1-r_2}\right)$$

Null Hypothesis (H_0)

There is no significance difference between two correlation coefficient.

i.e., $H_0 : \rho_1 = \rho_2$

Alternative Hypothesis (H_1)

There is significance difference between two correlation coefficient.

i.e., $H_1 : \rho_1 \neq \rho_2$

Level of Significance

Consider the appropriate at a% level of significance.

Test Statistic

Under the null hypothesis then the test statistic is given by

$$z = \left|\frac{(z_1 - z_2) - E(z_1 - z_2)}{S.E(z_1 - z_2)}\right| \sim N(0, 1) \text{ under } H_0 \quad \ldots (1)$$

Now consider,

$$E(z_1 - z_2) = E(z_1) - E(z_2)$$
$$= \rho_1 - \rho_2$$
$$= \rho_1 - \rho_1$$
$$E(z_1 - z_2) = 0 \quad \ldots (2)$$

$$S.E(z_1 - z_2) = \sqrt{v(z_1 - z_2)}$$
$$= \sqrt{v(z_1) + v(z_2)}$$

$$S.E(z_1 - z_2) = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad \ldots (3)$$

Sub Equation (2) and (3) in (1)

$$z = \left|\frac{(z_1 - z_2) - 0}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}\right| \sim N(0, 1)$$

$$z = \left|\frac{(z_1 - z_2)}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}\right| \sim N(0, 1) \text{ under } H_0$$

11. Define order statistics.

Ans :

If $a_1, a_2, a_3 \ldots a_n$ represents the n independent and identical distributed variates of cummulative distribution F(a), then the r^{th} order statistics of a denoted by ar is given by,

$$a_r = a_{(1)} < a_{(2)} < a_{(3)} < a_{(4)} < \ldots\ldots a_{(n)}$$

Where,

r = 1, 2, 3, 4, 5 .... n is the order statistic of $a_{(r)}$.

Exercise Problems

1. A sample of 900 members as a mean 3.4 cms and S.D 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms. If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of true mean.

   (Ans : 3.40 ± 0.1705, 3.40 ± 0.2027)

2. The means of two single large samples of 1,000 and 2,000 members are 67.5 inches and 68.0 inches respectively can the samples be regarded as drawn from the same population of standard deviation 2.5 inches. (Test at 5% level of significance).

   (Ans : - 5.1)

3. A die is thrown 9,000 times and a throw of 3 or 4 is observed 3,240 times. Show that the die can not be regarded as an unbiased one and find the limits between which the probability of a throw of 3 or 4 lies.

   (Ans : 0.345, 0.375)

4. Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same against that they are not, at 5% level.

   (Ans:- 1.269)

5. Two populations have their means equal, but S.D of one is twice the other. Show that,

   (i) In the samples of size 2,000 from each drawn under simple that of sampling conditions, the difference of means will, in all probability, not exceed 0.15o, Where *a* is the smaller S.D.

   (ii) What is the probability that the difference will exceed half this amount?

   (Ans : 0.1336)

6. A random sample of 28 pairs of observations shows a correlation coefficient of 0.74. Is it reasonable to believe that the sample comes from a bivariate normal population with correlation coefficient 0.6?

   (Ans : 0.949,1.285)

Choose the Correct Answer

1. The exception and rejection of null hypothesis H_0 is decided with probabilities ψ (x) and ______ [a]

   (a) $1 - \psi(x)$ (b) $1 + \psi(x)$

   (c) $\psi(x) - 1$ (d) $2\psi(x)$

2. 95% confidence limits for μ is given by ______ [b]

   (a) $\bar{x} \pm 1.56 \frac{\sigma}{\sqrt{n}}$ (b) $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

   (c) $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ (d) $\bar{x} + 2.33 \frac{\sigma}{\sqrt{n}}$

3. The test statistic corresponding to $\bar{x}$ for a large sample of a sample of a single mean is given by ______ . [c]

   (a) $z = \frac{\bar{x} + \mu}{\frac{\sigma}{\sqrt{n}}}$ (b) $z = \frac{\bar{x}}{\sigma / \sqrt{n}}$

   (c) $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ (d) $z = \frac{\bar{x} - \mu}{\sigma}$

4. The test statistics for large samples for differences of means is given by ______ . [d]

   (a) $z = \frac{(\bar{x}_1 - \bar{x}_2)}{s. \in (\bar{x}_1 + \bar{x}_2)}$ (b) $z = \frac{(\bar{x}_1 - \bar{x}_2) - \in}{s. \in (\bar{x}_1 - \bar{x}_2)}$

   (c) $z = \frac{(\bar{x}_1 + \bar{x}_2)}{s. \in (\bar{x}_1 - \bar{x}_2)^2}$ (d) $z = \frac{(\bar{x}_1 - \bar{x}_2) - \in (\bar{x}_1 - \bar{x}_2)}{s. \in (\bar{x}_1 - \bar{x}_2)}$

5. If a sample is drawn from a finite population of size N, then the observed proportion of P is ______ [a]

   (a) S.E. (P) = $\sqrt{\left(\frac{N-n}{N-1}\right)\frac{PQ}{n}}$ (b) S.E. (P) = $\sqrt{\left(\frac{N-n}{N-1}\right)PQ}$

   (c) S.E (P) = $\left(\frac{N-n}{N-1}\right)\frac{PQ}{n}$ (d) S.E (P) = $\sqrt{\frac{N-n}{N-1}}$

6. ______ can be defined as the ratio of numbers of successes to the total no.of members in sample. [b]

   (a) Population proportion (b) Proportion

   (c) Sample proportion (d) Difference of proportions

7. Population proportion (P.P) is given by ______ [c]

(a) POP : P = $\dfrac{\text{Number of successes in sample}}{\text{Total no. of members in a sample}}$

(b) P.P = $\dfrac{\text{Number of members in a population}}{\text{Total numbers of members in a sample}}$

(c) P.P = $\dfrac{\text{Number of members in a population}}{\text{Total numberof members in a population}}$

(d) None of the above

8. Fisher's z = transformation is give by ______ [d]

(a) $z = \log_e\left(\dfrac{1+r}{1-r}\right)$ (b) $z = \log_e\left(\dfrac{1-r}{1+r}\right)$

(c) $z = \dfrac{1}{2}\log_e\left(\dfrac{1-r}{1+r}\right)$ (d) $z = \dfrac{1}{2}\log_e\left(\dfrac{1+r}{1-r}\right)$

9. If $b_1 = a_{(1)}$ represents the smallest order statistics of $a_1, a_2, ...a_n$ then its cummulative dist. Function f_n (a) is given as ______. [a]

(a) $f_1(a) = 1 - [1 - f(a)]^n$ (b) $f_n(a) = [F(a)]^n$

(c) $f_1(a) = 1 + [1 + f(a)]^n$ (d) $f_n(a) = F(a)$

10. 99% confidence limits for population proportion is given by ______ [b]

(a) $(p \pm 1.96\sqrt{\dfrac{pq}{n}})$ (b) $p \pm 2.58\sqrt{\dfrac{pq}{n}}$

(c) $(p \pm 2.33\sqrt{\dfrac{pq}{n}})$ (d) $(p \pm 1.58\sqrt{\dfrac{pq}{n}})$

Fill in the blanks

1. The 98% confidence limits for μ is ________.
2. The difference between two independent normal variater $(\bar{x}_1 - \bar{x}_2)$ is also said to be ________.
3. The 95% confidence Limits for population proportion is given by ________.
4. Sample proportion is given by ________.
5. ________ of finite population can be defined as the ratio of the number of members in the population with some atlribute to the total number of numbers.
6. Different coefficient of correlation from various sample can be combined using ________.
7. If $b_n = a_{(n)}$ represents the largest order statistic (x_n) of a_1, a_2, a_3 ...... a_n then its cummulative dist f_n (a) is given as ________.
8. The test statistic for difference between proportions $(p_1 - p_2)$ is given by ________.
9. The S.E for the difference between two standard deviations for a large sample is ________.
10. In case of large sample siges the distribution of z is approximately normal and the variance is given by ________.

ANSWERS

1. $\bar{x} \pm 2.33 \frac{\sigma}{\sqrt{n}}$
2. Normally distributed
3. $(p \pm 1.96 \sqrt{\frac{pq}{n}})$
4. $\frac{\text{Number of successes in sample}}{\text{Total number of members in a sample}}$
5. Population proportion
6. z - transformation
7. $F_{(n)}(a) = [F(a)]^n$
8. $z = \frac{(p_1 - p_2) - e(p_1 - p_2)}{\sqrt{var(p_1 - p_2)}}$
9. S.E $(S_1 - S_2) = \sqrt{\left(\frac{\sigma_1^2}{2n_1}\right) + \left(\frac{\sigma_2^2}{2n_2}\right)}$
10. $\frac{1}{n-3}$

Very Short Questions and Answers

1. Give one Example for randomized test function.

Ans :

An example of randomized test function is

$$P(x) = \begin{cases} 1 & \text{if} \quad x = 3 \\ 1/8 & \text{if} \quad x = 2 \\ 0 & \text{Otherwise} \end{cases}$$

2. Write the formula for the difference of properties.

Ans :

The difference of proportion for a large sample test is given by

$$z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{\text{var}\ (p_1 - p_2)}} \sim N(0,1)$$

3. What is sample proportion ?

Ans :

Sample proportion can be defined as the ratio of number of successes to the total number of members in the sample.

4. What is the equation of fisher's z- transformation ?

Ans :

Fisher's z-transformation is given by

$$z = \frac{1}{2} \log_e \left(\frac{1+r}{1-r} \right)$$

5. Write any one application of z-transformation.

Ans :

z -transformation is used to test whether there exist significant difference between two sample coefficients or not.

UNIT III	Tests of significance based on χ^2 - χ^2-test for specified variance, goodness of fit and test forindependence of attributes (rxs, 2 × k and 2 × 2 contingency tables). Tests of significance basedon student's - t – t-test for single sample specified mean, difference of means for independentand related samples, sample correlation coefficient. F- test for equality of populationvariances.

3.1 Small Sample Test

Q1. Define small sample test.

Ans :

The sample size n < 30 then it is called small sample test. We will discuss the following tests.

i) t-test

(a) t-test for single mean

(b) t-test for difference of two mean (Independent samples)

(c) t-test for difference of two means (Dependent samples)

(or)

Paired t-test.

(d) t-test for correlation coefficient

ii) F - test

(a) F-test for equality of population variance

iii) χ^2 - test

(a) χ^2 -test for goodness of fit

(b) χ^2 -test for independent of two attributes.

3.1.1 Tests of Significance Based on Student's - t

Q2. What is t-distribution? Explain the properties and applications of t-distribution.

Ans : **(June-21, Imp.)**

When population standard deviation (s_p) is not known and the sample is of small size (*i.e.*, n £ 30).

Properties of Student's t-Distribution

(i) The probability curve of t is symmetric, like in standard normal distribution (z).

(ii) The distribution ranges from – α to a just as does a normal distribution.

(iii) The t-distribution is bell shaped and symmetrical around mean zero, like normal distribution.

(iv) The shapes of the t-distribution changes as the sample size changes (the number of degrees of freedom changes) whereas it is same for all sample sizes in z-distribution.

(v) The variance of t-distribution is always greater than one and is defined only when n > 3.

(vi) The t-distribution is more of platykurtic (less packed at centre and higher in tails) than the normal distribution.

(vii) The t-distribution has a greater dispersion than the normal distribution. As n becomes larger, the t-distribution approaches the standard normal distribution.

Applications of t-Distribution

The following are some important applications of t-distribution.

(i) Test of hypothesis about the population mean.

(ii) Test of hypothesis about the difference between two means.

(iii) Test of hypothesis about the difference between two means with dependent samples.

(iv) Test of hypothesis about coefficient of correlation.

3.1.2 t-test for Single Sample Specified Mean

Q3. Explain t-test for single mean.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Let $x_1, x_2.....x_n$ be a random samples drawn from the normal population with specified mean and variance then

Null Hypothesis (H0)

There is no significance difference between population mean and sample mean

(or)

The samples have been drawn from a population.

i.e., $H_0 : \mu = \mu_0$

Alternative Hypothesis(H_1)

There is significance difference between population mean and sample mean.

(or)

The samples have not been drawn from population

i.e., $H_1 : \mu_1 \neq \mu_0$

Level of Significance

Consider the appropriate at α% level of significance.

Test statistic

The test statistic is given by

$$t = \left|\frac{\bar{x} - \mu}{s / \sqrt{n-1}}\right| \sim t_{(n-1)} \text{ d.f}$$

where,

$\bar{x}$ = sample mean

μ = population

Remarks

if S is unknown then the test statistic is given by

$$t = \left|\frac{\bar{x} - \mu}{s\sqrt{n}}\right| \sim t_{(n-1)} \text{ d.f}$$

Confidence limits

(i) The 95% confidence limits for t-test for single mean is $\bar{x} \pm t_{0.05}\ s/\sqrt{n}$.

(ii) The 99% confidence limits for t-test for single mean is $\bar{x} \pm t_{0.01}\ s/\sqrt{n}$.

Conclusion

If the calculated value of t is lessthan the tabulated value of t at certain level of significance for (n – 1) degrees of freedom the we accept H_0 otherwise we reject H_0

Applications

(i) To test the sample mean ($\bar{x}$) differ significantly form the population mean.

i.e., t-test for single mean.

(ii) To test the significance difference between two means.

i.e., t-test for two mean

(iii) To test the significance on observed sample correlation coefficient.

i.e., t-test for correlation coefficient

(iv) To test the significance observed partial correlation coefficient

PROBLEMS

1. The mean weakly sales of soap bars in department stores was 146.3 bars per stores. After a advertising campaign the mean weakly sales in 22 stores for a typical week increased to 153.7 and show a standard deviation 17.2 was advertising successful or not ?

Sol :

Null Hypothesis(H_0)

There is no significance difference between the sales of soap bars before and after advertising campaign.

i.e., $H_0 : \mu = 146.3$

Alternative Hypothesis(H_1)

After advertising campaign the mean weekly sales increases.

i.e., $H_1 : \mu > 146.3$ [Right tailed test]

Level of significance (a)

Consider the appropriate at 5% level of significance for right tailed test with (n – 1) degrees of freedom is 1.72

Test statistic

Given that,

n = 22, $\bar{x}$ = 153.7, μ = 146.3, S = 17.2

Then the test statistic is given by

$$t = \left|\frac{x-\mu}{s\sqrt{n-1}}\right| \sim t_{(n-1)} \text{ d.f}$$

$$t = \left|\frac{153.7-146.3}{17.2/\sqrt{22-1}}\right| \sim t_{(22-1)} \text{ d.f}$$

$$t = \left|\frac{7.4}{3.7533}\right| \sim t_{(21)} \text{ d.f}$$

$$t = \left|\frac{7.4}{3.7533}\right| \sim t_{(21)} \text{ d.f}$$

t_{cal} = 1.97

The calculated value of t is 1.97

The tabulated value at 5% level of significance with (22 – 1) degree of freedom is 1.72.

The t calculated value is greater than the t tabulated value.

i.e., $t_{cal} > t_{tab}$

∴ 1.97 > 1.72

we reject H_0

Conclusion

After advertising campaign the mean weekly sales of soap bars is increased.

2. A random sample of 10 boys they follow IQ is 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do this support the assumption of the population mean I.Q of 100 and find a responsible range in which most of the mean I.Q values of sample of 10 boys.

Sol :

Null Hypothesis (H_0)

The population mean of I.Q is 100

i.e., $H_0 : \mu = 100$

Alternative Hypothesis(H_1)

The population mean of I.Q is not 100.

i.e., $H_1 : \mu \neq 100$ [Two Tailed Test]

Level of Signification ($\propto$)

Consider the appropriate at 5% level of significance

Test statistic :

Given that,

n = 10, μ = 100

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
70	– 27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	–9.2	84.64
83	– 14.2	201.64
95	– 2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
$\Sigma x_i = 972$		$\Sigma(x_i - \bar{x})^2 = 1833.6$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{972}{10} = 97.2$$

and $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S^2 = \frac{1}{10-1}(1833.6)$$

$$S^2 = \frac{1833.6}{9}$$

$$S^2 = 203.73$$

$$S = 14.273$$

Then the test statistics is given by

$$t = \left|\frac{\bar{x} - \mu}{S/\sqrt{n-1}}\right| \sim t_{(n-1)} \text{ differ}$$

$$t = \left|\frac{97.2 - 100}{14.278/\sqrt{10-1}}\right| \sim t_{(10-1)} \text{ differ}$$

$$t = \left|\frac{2.8}{14.272/\sqrt{9}}\right| \sim t_9$$

$$t = \left|\frac{2.8}{4.757}\right| \sim t_9 \text{ differ}$$

$t_{cal} = 0.5886$

The tabulated value of t at 5% level of significance with (10 – 1) degrees of freedom is 2.262

The t calculate value is lessthan the t tabulated value.

i.e., $t_{cal} < t_{tab}$

$\therefore$ 0.5886 < 2.262

$\therefore$ we accept H_0

Conclusion

The population mean if I.Q is 100.

3. The height of 10 males of the given location are and to be 70, 67, 62, 68, 61, 70, 64, 64, 66, 68 inches it is responsible to below that average height is greater 64 in inches.

Sol :

Null Hypothesis(H_0)

The average height of male is lessthan 64.

i.e., $H_0 : \mu < 64$

Alternative Hypothesis(H_1)

The average height of male is greaterthan 64.

i.e., $H_1 : \mu > 64$

Level of significance :

Consider the appropriate at 5% level of significance.

Test statistics

Given that n = 10

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
70	4	16
67	1	1
62	– 4	16
68	2	4
61	– 5	25
70	4	16
64	– 2	4
64	– 2	4
66	0	0
68	2	4
$\Sigma x_i = 660$		$\Sigma(x_i - \bar{x})^2 = 90$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{660}{10} = 66$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S^2 = \frac{1}{10-1}(90)$$

$$S^2 = \frac{90}{9}$$

$S^2 = 10$

$S = 3.16$

Then the test statistic is given by

$$t = \left|\frac{\bar{x} - \mu}{S/\sqrt{n}}\right| \sim t_{(n-1)} \text{ differ}$$

$$t = \left|\frac{66 - 64}{3.16/\sqrt{10}}\right| \sim t_{(10-1)} \text{ differ}$$

$$t = \left|\frac{2}{3.16/3.16}\right| \sim t_9$$

$t_{cal} = 2$

The tabulated value of at 5% level of significance with (10 – 1) degree of freedom is 2.262.

The calculated value of t is lessthan the tabulated value of t

$\therefore \quad t_{cal} < t_{tab}$

i.e., $2 < 2.262$

$\therefore$ we accept H_0

Conclusion

The average height of male is lessthan 64.

3.1.3 Difference of means for Independent and Related Samples

Q4. Explain t-test for difference of two means of Independent samples.

Ans : **(June-18)**

Suppose we want to test the independent samples x_i : $i = 1, 2, ...n_1$ and y_i : $i = 1, 2, ...n_2$ have been drawn from the population with same mean .

(or)

The two sample means $(\bar{x} + \bar{y})$ are differ significance or not.

Null Hypothesis(H_0)

The samples are drawn from the population with same mean.

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The samples are not drawn from the population with same.

i.e., $H_1 : \mu_x \neq \mu_y$

Level of significance

Consider the appropriate at α% level of significance for $(n_1 + n_2 - 2)$ degree of freedom.

Test statistic

The sample mean $\bar{x}$ and $\bar{y}$ do not differ significance then the test statistic is given by

$$t = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim t_{(n_1+n_2-2)} \text{ differ}$$

where,

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2\right]$$

(sample S.D. unknown)

and

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \text{ (sample S – D. known)}$$

Confidence limits

(i) The 95% confidence limits for difference of two means is $(\bar{x} - \bar{y}) \pm t_{0.05}\, S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

(ii) The 99% confidence limits for difference of two means is $(\bar{x} - \bar{y}) \pm t_{0.01}\, S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Conclusion

The calculated value of t is lessthan the tabulated value of t at certain α% level of significance for $(n_1 + n_2 - 2)$ degree of freedom then we accept H_0 otherwise we reject H_0.

PROBLEMS

4. A sample of two types of electrical bulbs where tested for length of lifetime and the following data are obtained.

	Type-I	Type-II
Sample No	8	7
Sample mean	1235 hrs	1063 hrs
Sample S.D	36 hrs	40 hrs

If the difference in the mean sufficient to warranty the type-I is superior than type-II regarding the length of life.

Sol :

Null Hypothesis

There is no significance difference between the two types of electrical bulbs.

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The electrical bulb of types - I is superior than type-II

i.e., $H_1 : \mu_x > \mu_y$ [Right tailed test]

Level of significance

Consider the appropriate at 5% level of significance for ($n_1 + n_2 - 2$) degree of freedom for right tailed test.

Test statistic

Given that, $n_1 = 8$, $n_2 = 7$

$\bar{x} = 1235$ hrs, $\bar{y} = 1036$ hrs

$S_1 = 36$ hrs, $S_2 = 40$ hrs

The test statistics is given by

$$t = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim t_{(n_1+n_2-2)} \text{ d.f}$$

where,

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{8(36)^2 + 7(40)^2}{8+7-2}$$

$$S^2 = \frac{8(1296) + 7(1600)}{15-2}$$

$$S^2 = \frac{10368 + 11200}{13}$$

$$S^2 = \frac{21568}{13}$$

$S^2 = 1659.07$

$S = 40.73$

'S' value sub in t

$$t = \left| \frac{1235 - 1036}{40.73\sqrt{\frac{1}{8} + \frac{1}{7}}} \right| \sim t_{(8+7-2)} \text{ differ}$$

$$t = \left| \frac{199}{40.73(0.5175)} \right| \sim t_{13} \text{ differ}$$

$$t = \left| \frac{199}{21.07} \right| \sim t_{13} \text{ differ}$$

$t_{cal} = 9.47$

The tabulated value of t at 5% level of significance with (13) degrees of freedom is 1.771.

The calculated value of t is greaterthan the tabulated value of t.

$\therefore$ $t_{cal} > t_{tab}$

i.e., $9.47 > 1.771$

$\therefore$ we reject H_0

Conclusion

The electric bulb of type-I is superior than type-II.

---

5. The height of 6 randomly choice of the soldiers are 63, 65, 68, 69, 71, 72. Those 9 randomly choosen soldiers 61, 62, 65, 66, 69, 70, 71, 72, 73 discuss the height i.e., the data throw on the suggestion that soliders are on the average taller than soliders of type-II.

Sol :

Null Hypothesis(H_0)

There is no significance difference between two heights.

i.e., $h_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The first type of heights are taller than second type.

I.e., $H_1 : \mu_x > \mu_y$ [Right Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance for $(n_1 + n_2 - 2)$ degrees of freedom

Test statistic

x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
63	61	– 5	25	– 6.6	43.56
65	62	– 3	9	– 5.6	31.36
68	65	0	0	– 2.6	6.76
69	66	1	1	– 1.6	2.56
71	69	3	9	1.4	1.96
72	70	4	16	2.4	5.76
	71			3.4	11.56
	72			4.4	19.36
	73			5.4	29.16
$\Sigma x_i = 408$	$\Sigma y_i = 609$		$\Sigma(x_i - x)^2 = 60$		$(y_1 - \bar{y})^2 = 152.4$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{408}{6} = 68$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{609}{9} = 67.6$$

Given that, $n_1 = 6, \; n_2 = 9$

The test statistic is given by

$$t = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim (n_1 + n_2 - 2) \text{ d.f}$$

where,

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_i - \bar{y})^2\right]$$

$$S^2 = \frac{1}{6 + 9 - 2}[60 + 152.04]$$

$$S^2 = \frac{212.04}{13}$$

$S^2 = 16.31$

$S = 4.038$

Sub 'S' value in 't'

$$t = \left|\frac{68 - 67.6}{4.038\sqrt{\frac{1}{6}+\frac{1}{9}}}\right| \sim t_{(6+9-2)} \text{ d.f}$$

$$t = \left|\frac{0.4}{4.038(0.527)}\right| \sim t_{13} \text{ d.f}$$

$$t = \left|\frac{0.4}{2.128}\right| \sim t_{13} \text{ d.f}$$

$t_{cal} = 0.187$

The tabulated value of t at 5% level of significance with 13 degrees of freedom is 1.771.

The calculated value of t is lessthan the tabulated value of t.

$t_{cal} < t_{tab}$

i.e., $0.187 < 1.771$

$\therefore$ we accept H_0

Conclusion

There is no significance difference between two heights.

---

6. In a certain experiment contain two types of pig food A and B. The following results of increased in their weights we observed.

Pig No	1	2	3	4	5	6	7	8
Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	55	50	54	54	53

Calculated the food B is better than food A.

Sol :

Null Hypothesis (H_0)

There is no significance difference between two types of food A and B

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The food B is better than food A

i.e., $\mu_y > \mu_x$ [Right Tailed Test]

Level of Significance

Consider the appropriate at 5% level of significance for $(n_1 + n_2 - 2)$ degrees of freedom.

Test statistic

x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
49	52	– 1.87	3.49	– 1.12	1.25
53	55	2.13	4.53	1.88	3.53
51	52	0.13	0.01	– 1.12	1.25
52	55	1.13	1.27	1.88	3.53
47	50	– 3.87	14.97	– 3.12	9.73
50	54	– 0.87	0.75	0.88	0.77
52	54	1.13	1.27	0.88	0.77
53	53	2.13	4.53	– 0.12	0.01
Σx_i = 407	Σy_i = 425		$\Sigma(x_i - x)^2$ = 30.82		$\Sigma x_i = (y_1 - \bar{y})^2$ = 20.84

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{407}{8} = 50.87$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{425}{8} = 53.12$$

Given that, $n_1 = 8, \; n_2 = 8$

The test statistic is given by

$$t = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim t_{(n_1 + n_2 - 2)} \text{ d.f}$$

where,

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_i - \bar{y})^2\right]$$

$$S^2 = \frac{1}{8 + 8 - 2}[30.82 + 20.84]$$

$$S^2 = \frac{51.66}{14}$$

$$S^2 = 3.69$$

$$S^2 = 1.92$$

Sub 's' value in 't'

$$t = \left|\frac{50.87 - 53.12}{(1.92)_1\sqrt{\frac{1}{8}+\frac{1}{8}}}\right| \sim t_{(8+8-2)} \text{ d.f}$$

$$t = \left|\frac{-2.25}{1.92(0.5)}\right| \sim t_{(14)} \text{ d.f} \Rightarrow \left|\frac{2.25}{0.96}\right| \sim t_{(14)} \text{ d.f}$$

$t_{cal} = 2.34$

The tabulated value of t at 5% level of significance with degree of freedom is 2.571.

The calculated value of t is lessthan the tabulated value of t.

$\therefore\ t_{cal} < t_{tab}$

i.e., $2.34 < 2.571$

we accept H_0

Conclusion

There is no significance difference between two types of food A and B.

---

7. To test the claim that the resistance of an electrical wire can be reduced by atlest 0.05 allowing 25 value so obtained for each standard wire allowed protected by the following results.

Particulars	Mean	S.D
Allowed wire	0.083	0.003
Standard wire	0.136	0.002

Test at 5% level of significance

Sol :

Null Hypothesis(H_0)

The is no significance difference between two types of wires.

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

There is significance difference between

Two types of wires

i.e., $H_1 : \mu x \neq \mu y$ [Two Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance for $(n_1 + n_2 - 2)$ degree of freedom

Given that $n_1 = 25, n_2 = 25$ $[n_1 = n_2]$

Given mean of allowed wire $\bar{x} = 0.083$

Given mean of standard wire $\bar{y} = 0.136$

Standard deviations of allowed wire

$S_1 = 0.003, \quad S_2 = 0.002$

Test statistic

The test statistic is given by

$$t = \left| \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \right| \sim t_{(n_1+n_2-2)}$$

$$t = \left| \frac{0.083 - 0.136}{\frac{(0.003)^2}{25} + \frac{(0.002)^2}{25}} \right| \sim t_{(25+25-2)}$$

$$t = \left| \frac{-0.053}{\sqrt{0.0000052}} \right| \sim t_{48}$$

$$t = \left| \frac{0.053}{0.000721} \right|$$

$t = 73.5$

$t_{cal} = 73.5$

The tabulated value of t at 5% level of significance with (48) degrees of freedom is 3.

The calculated value of t is greaterthan the tabulated value of t

$\therefore \quad t_{cal} > t_{tab}$

$73.6 > 3$

$\therefore$ we reject H_0

Conclusion

There is significance difference between two types of wires.

Q5. Explain t-test for difference of two means of dependent samples.

Ans : **(Feb.-21, Imp.)**

Let us consider in this case whether sample sizes are equal i.e., $n_1 = n_2 = n$ and the two samples are non independent (dependent) but the sample observations are paired together.

Then the problem is to test the sample means are differ significance (or) not .

Now consider $d_i = x_i - y_i$ and take set of d_i values as a single sample and apply t-test for single mean.

Null Hypothesis(H_0)

There is no significance difference between two sample means

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

There is significance difference between two sample means

i.e., $H_0 : \mu_x \neq \mu_y$

Level of significance

Consider the appropriate at α% level of significance for two tailed test.

Test statistic

The test statistic is given by

$$t = \left|\frac{\bar{d}}{S/\sqrt{n}}\right| \sim t_{(n-1)} \text{ d.f}$$

where,

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(d_i - \bar{d})^2\right] \quad \text{(or)}$$

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}d_i^2 - n(\bar{d})^2\right]$$

It follows t-distribution with (n – 1) degrees of freedom.

Conclusion

If the calculated value of t is lessthan the tabulated value of t at certain level of significance for two tailed test for (n – 1) degrees of freedom then we accept H_0 otherwise we reject H_0.

PROBLEMS

8. The scores of 10 candidates before and after training are given below :

Before	**84**	**48**	**36**	**37**	**54**	**69**	**83**	**96**	**90**	**95**
After	**90**	**58**	**56**	**49**	**62**	**81**	**84**	**86**	**84**	**75**

Is the training effective ?

Sol :

Null Hypothesis(H_0)

There is no significance difference between the scores of 10 candidate before and after training.

(or)

The training is not effective

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The training is effective

i.e., $H_0 : \mu_y > \mu_x$ [Right tailed test]

Level of significance

Consider the appropriate at 5% level of significance for right tailed test with (10 – 1 = 9) degrees of freedom.

Test statistic

Given that n = 10

x_i	y_i	$d_i = x_i - y_i$	d_1^2
84	90	– 6	36
48	58	– 10	100
36	56	– 20	400
37	49	– 12	144
54	62	– 8	64
69	81	– 12	144
83	84	– 1	1
96	86	10	100
90	84	6	36
95	75	20	400
		$\Sigma d_i = -33$	$\Sigma d_i^2 = 1425$

$$\bar{d} = \frac{\Sigma d_i}{n}$$

$$\bar{d} = \frac{-33}{10}$$

$$\bar{d} = -3.3$$

$$(\bar{d})^2 = 10.89$$

The test statistic is given by

$$t = \left|\frac{\bar{d}}{S/\sqrt{n}}\right| \sim t_{(n-1)} \text{ d.f}$$

where,

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} d_i^2 - n(d)^2\right]$$

$$S^2 = \frac{1}{10-1}[1425 - 10(10.89)]$$

$$S^2 = \frac{1}{9}\,[1425 - 108.9]$$

$$S^2 = \frac{1361.1}{9}$$

$$S^2 = 146.23$$

$$S = 12.092$$

Sub 's' value in t

$$t = \left|\frac{-3.3}{12.092/\sqrt{10}}\right| \sim t_{(10-1)} \text{ differ}$$

$$t = \left|\frac{3.3}{3.823}\right| \sim t_9 \text{ differ}$$

$$t = 0.86$$

$$t_{cal} = 0.86$$

The calculated value of t is 0.86

The tabulated value of t at 5% level of significance with (9) degree of freedom is 1.833

The calculated value of t is lessthan the tabulated value of t

$\therefore\ t_{cal} < t_{tab}$

i.e., $0.86 < 1.833$

$\therefore$ we accept H_0

Conclusion

The training is not effective

9. In a certain experiment to compare two animal foods A and B the following results of increases in weights where observed in animals.

Animal No	1	2	3	4	5	6	7	8
Food A	49	53	51	52	47	50	52	53
Food B	52	55	52	53	50	54	54	53

Also test whether this case when the camp of set of 8 animals where used in both the foods.

Sol :

Null Hypothesis(H_0)

There is no significance difference between two animal foods.

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

There is significance difference between two animal foods

i.e., $H_1 : \mu_x \neq \mu_y$

Level of significance

Consider the appropriate at 5% level of significance for two tailed test with (8 – 1 = 7) degrees of freedom.

Test statistic

Given that n = 8

x_i	y_i	$d_i = x_i - y_i$	d_1^2
49	52	– 3	9
53	55	– 2	4
51	52	– 1	1
52	53	– 1	1
47	50	– 3	9
50	54	– 4	16
52	54	– 2	4
53	53	0	0
		$\Sigma d_i = -16$	$\Sigma d_i^2 = 44$

$$\bar{d} = \frac{\Sigma d_i}{n}$$

$$\bar{d} = \frac{-16}{8}$$

$$\bar{d} = -2 \Rightarrow (\bar{d})^2 = 4$$

The test statistic is given by

$$t = \left|\frac{\bar{d}}{S/\sqrt{n}}\right| \sim t_{(n-1)}$$

where,

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} d_1^2 - n(\bar{d})^2\right]$$

$$S^2 = \frac{1}{8-1}[44 - 8(4)]$$

$$S^2 = \frac{1}{7}[44 - 32]$$

$$S^2 = \frac{12}{7}$$

$$S^2 = 1.71$$

$$S = 1.30$$

Sub 's' value in 't'

$$t = \left|\frac{-2}{1.30/\sqrt{8}}\right| \sim t_{(8-1)} \text{ d.f}$$

$$t = \left|\frac{2}{0.459}\right| \sim t_7 \text{ d.f}$$

$$t = 4.35$$

$$t_{cal} = 4.35$$

The calculated value of t is 4.35

The tabulated value of t at 5% level of significance with 7 degrees of freedom is 1.895

The calculated value of t is greaterthan the tabulated value of t

$\therefore$ $t_{cal} > t_{tab}$

i.e., $4.35 > 1.895$

$\therefore$ we reject H_0

Conclusion

There is significance difference between two animal foods.

10. A certain stimulus administrated to each of 12 patients in the following increasing blood pressures 5, 2, 8, –1, 3, 0, –2, 1, 5, 0, 4, 6 can it be conclude that the stimulus will general be accompanied by an increases in blood pressure.

Sol :

Null Hypothesis(H_0)

There is no significance difference between blood pressure reading before and after drug

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

The patients blood pressure is increased after the drugs

i.e., $H_1 : \mu_y > \mu_x$ [Right Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance for right tailed test with (12 – 1 = 11) degrees of freedom.

Test statistic

Given that n = 12

d_i	d_i^2
5	25
2	4
8	64
– 1	1
3	9
0	0
– 2	4
1	1
5	25
0	0
4	16
6	36
$\Sigma d_i = -31$	$\Sigma d_i^2 = 185$

$$\bar{d} = \frac{\Sigma d_i}{n}$$

$$\bar{d} = \frac{31}{12}$$

$$\bar{d} = 2.58$$

$$\Sigma d_i^2 = 185$$

The test statistic is given by

$$t = \left|\frac{\bar{d}}{S/\sqrt{n}}\right| \sim t_{(n-1)} \text{ differ}$$

where,

$$S^2 = \frac{1}{n-1}\left[\sum_{n-1}^{n} d_i^2 - n(\bar{d})^2\right]$$

$$S^2 = \frac{1}{12-1}[185.12(2.58)^2]$$

$$S^2 = \frac{1}{11}[185 - 12(6.65)]$$

$$S^2 = \frac{185 - 79.8}{11}$$

$$S^2 = \frac{105.2}{11}$$

$$S^2 = 9.56$$

$$S = 3.09$$

Sub 's' value in t

$$t = \left|\frac{2.58}{3.09/\sqrt{12}}\right| \sim t_{(12-1)} \text{ d.f}$$

$$t = \left|\frac{2.58}{0.89}\right| \sim t_{(11)} \text{ d.f}$$

$$t_c = 2.898$$

$$t_{cal} = 2.898$$

The calculated value of t is 2.898

The tabulated value of t at 5% level of significance with 11 degrees of freedom is 1.89.

The calculated value of t is greaterthan the tabulated value of t

$\therefore \quad t_{cal} > t_{tab}$

i.e., $2.898 > 1.89$

$\therefore$ we reject H_0

Conclusion

The patients blood pressure is increased after the drugs.

11. Two laboratories carry out independent estimate of a particular chemicals in a medicine produced by a certain form a sample is taken from each batch in halved and the separate halved send to the two laboratories. The following data is obtained number of samples 10 mean values of different estimates 0.6 sm of the squares of difference 20 is the different significant.

Sol :

Null Hypothesis(H_0)

There is no significance difference between two laboratories.

i.e., $H_0 : \mu_x = \mu_y$

Alternative Hypothesis(H_1)

There is significance difference between two laboratories.

i.e., $H_1 : \mu_x \neq \mu_y$ [Right Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance for right tailed test with (10 – 1 = 9) degrees of freedom.

Test statistic

Given that n = 10

Mean value of different estimates i.e., $\bar{d} = 0.6$ sum of the squares of the difference is

$$\Sigma(d_i - \bar{d})^2 = 20$$

Then the test statistic is given by

$$t = \left|\frac{d}{S/\sqrt{n}}\right| \sim t_{(n-1)} \text{ differ}$$

where,

$$S^2 = \frac{1}{n-1}\left[\Sigma(d_i - \bar{d})^2\right]$$

$$S^2 = \frac{1}{10-1}[20]$$

$$S^2 = \frac{20}{9}$$

$$S^2 = 2.2222$$

$$S = 1.4907$$

sub 's' value in 't'

$$t = \left|\frac{0.6}{1.4907/\sqrt{10}}\right| \sim t_{(10-1)} \text{ d.f}$$

$$t = \left|\frac{0.6}{0.471}\right| \sim t_9 \text{ d.f}$$

$$t = 1.2738$$

$$t_{cal} = 1.2738$$

The calculated value of t is is 1.2738

The tabulated value of t at 5% level of significance with 9 degrees of freedom is 2.262

The calculated value of t is lessthan the tabulated value of t.

∴ $t_{cal} < t_{tab}$

i.e., 1.2738 < 2.262

∴ we accept H_0

Conclusion

There is no significance difference between two laboraties.

3.1.4 Sample Correlation Coefficient

Q6. Explain t-test for correlation coefficient.

(OR)

Explain the test procedures for sample correlation coefficient based on students.

Ans : **(Feb.-21, Imp.)**

It 'r' is the observed sample correlation coefficient in a sample of size 'n' pairs of observations from a bivariate normal population.

The R.A. Fisher proved that under the

Null Hypothesis(H_0)

The population correlation coefficient is not significant.

i.e., $H_0 : \int = \int_0$

Alternative Hypothesis(H_1)

The population correlation coefficient is significant

i.e., $H_1 : \int \neq \int_0$

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} \text{ d.f}$$

i.e., It follows t-distribution with (n – 2) degrees of freedom

Conclusion

The calculated value of t is lessthan the tabulated value of t at α% level of significance with (n – 2) degree of freedom then we accept H_0 otherwise we reject H_0.

<u>PROBLEMS</u>

12. A random sample of 24 pairs of observation from a normal population given a correlation coefficient is 0.6 is the significant of correlation in the population.

Sol :

Null Hypothesis(H_0)

The correlation coefficient is not significance in the population

i.e., $H_0 : \int = \int_0$

Alternative Hypothesis(H_1)

The correlation coefficient is significant in the population.

i.e., $H_1 : \int \neq \int_0$ [Two Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance with 25 degrees of freedom is 0.06

Test statistic

Given that,

$r = 0.6, \; n = 27$

The test satistic is given by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} \text{ d.f}$$

$$t = \frac{(0.6)\sqrt{27-2}}{\sqrt{1-(0.6)^2}} \sim t_{(27-2)} \text{ d.f}$$

$$t = \frac{(06)\sqrt{25}}{\sqrt{1-0.36}} \sim t_{25} \text{ d.f}$$

$$t = \frac{(06)5}{0.8}$$

$$t = \frac{3}{0.8}$$

$$t_{cal} = 3.75$$

The calculated value of t is 3.75

The tabulated value of t at 5% level of significance with degree of freedom is 0.06.

The calculated value of t is greaterthan the tabulated value of t

$\therefore \quad t_{cal} > t_{tab}$

i.e., $3.75 > 0.06$

$\therefore$ we reject H_0

Conclusion

The correlation coefficient is significant in the population.

13. A restaurant owner ranked is 17 waiters interns of their speed and efficiency on job. The correlated their ranks with the total amount of tips each of there waiter received for one week period. The obtained value of correlation coefficient is 0.438. What do you conclude.

Sol :

Null Hypothesis(H_0)

There is no significance difference between sample correlation coefficient and population correlation coefficient.

i.e., $H_0 : \int = \int_0$

Alternative Hypothesis(H_1)

There is significance difference between sample correlation coefficient and population correlation coefficient.

i.e., $H_1 : \int \neq \int_0$ [Two Tailed Test]

Level of significance

Consider the appropriate at 5% level of significance with (17-2) degrees of freedom is 2.131

Test statistic

Given that n = 17, r = 0.438

The test statistic is given by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} \text{ d.f}$$

$$t = \frac{(0.4328)\sqrt{17.2}}{\sqrt{1-(0.438)^2}} \sim t_{(17-2)} \text{ d.f}$$

$$t = \frac{(0.438)\sqrt{15}}{\sqrt{1-0.1918}} \sim t_{15} \text{ d.f}$$

$$t = \frac{1.696}{\sqrt{0.8082}}$$

$$t = \frac{1.696}{0.898}$$

$$t_{cal} = 1.88$$

The calculated value of t is 1.88

The tabulated value of t at 5% level of significance with 15 degrees of freedom is 2.131

The calculated value of t is lessthan the tabulated value of t

$\therefore$ $t_{cal} < t_{tab}$

i.e., $1.88 < 2.131$

$\therefore$ we accept H_0

Conclusion

There is no significance difference between sample correlation coefficient, and population correlation coefficient.

3.2 F-Test for Equality of Population Variances

Q7. Explain the concept of F-test for equality of population variance.

Ans : **(Feb.-21, June-19, June-18, Imp.)**

Suppose we want to test whether two independent sample x_i ; $i = 1, 2.....n_1$ and y_i ; $i = 1, 2 .....n_2$ these samples has been drawn from the normal population with the same variance σ^2

Objective

In this procedure to test the two independent estimators of the population variance are equal or not.

Null Hypothesis(H_0)

The population variance are equal

i.e., $H_0 : \sigma_x^2 = \sigma_y^2 = \sigma_0^2$

Alternative Hypothesis(H_1)

The population variance are not equal

i.e., $H_1 : \sigma_x^2 = \sigma_y^2 = \sigma_0^2$

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ d.f; If } (S_x^2 > S_y^2)$$

$$F = \frac{S_y^2}{S_x^2} \sim F(v_1, v_2) \text{ d.f; If } (S_y^2 > S_x^2)$$

where,

$v_1 = (n_1 - 1)$ d.f

$v_2 = (n_2 - 1)$ d.f

and $$S_x^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1}(x_i - \overline{x})^2\right]$$

$$S_x^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n} x^2 - n(\overline{x})^2\right]$$

$$S_y^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2}(y_i - \bar{y})^2\right]$$

$$S_y^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2}(y_i)^2 - n(\bar{y})^2\right]$$

Conclusion

If the calculated value of F is lessthan the tabulated value of F at certain level of significance with $v_1 = (n_1 - 1)$ and $v_2 = (n_2 - 1)$ degrees of freedom then we accept H_0 otherwise we reject H_0.

Q8. Explain the properties and applications of F-Test.

Ans :

Properties

1. The F-distribution curve lies in only first quadrant (Q1) and is unimodal.
2. The F-distribution is independent (free) of population parameter and depends only on the degree of freedom (i.e., V_1 and V_2) according to its order.
3. The F-distribution mode is less than unity (i.e., mode < 1).

Applications

1. It is used for testing the equality of many population means.
2. It is used for comparing he sample variances.
3. It is used for performing analysis of variance.
4. It is used for testing the significance of regression equation.
5. It is used for determining whether the ratio incrementally changes from unity at any level chosen randomly.

PROBLEMS

14. Two independent samples given the following results

Sample	Size	Sample Mean	Sum of squares deviation from mean
1	10	15	90
2	12	14	108

Test whether the samples comes from the normal population at 5% level of significance.

Sol :

Null Hypothesis(H_0)

The samples are drawn from same normal population.

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis(H_1)

The samples are not drawn from same normal population.

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$ [Two tailed test]

Level of significance

Consider the appropriate at 5% level of significance

Test statistic

Given that,

$n_1 = 10, \quad n_2 = 12$

$\bar{x} = 15, \quad \bar{y} = 14$

$\Sigma(x_i - \bar{x})^2 = 90, \quad \Sigma(y_i - \bar{y})^2 = 108$

$S_x^2 = \frac{1}{n_1 - 1} = \Sigma(x_i - \bar{x})^2 \qquad S_y^2 = \frac{1}{n_2 - 1}\Sigma(y_i - \bar{y})^2$

$S_x^2 = \frac{1}{10-1}(90) \qquad S_y^2 = \frac{1}{12-1}(108)$

$S_x^2 = \frac{90}{9} \qquad S_y^2 = \frac{108}{11}$

$S_x^2 = 10 \qquad S_y^2 = 9.8$

$\therefore \quad S_x^2 > S_y^2$

Then the test statistic is given by

$F = \frac{S_x^2}{S_y^2} \sim F(9, 11)$ d.f

$F = \frac{10}{9.8} \sim F(9, 11)$ d.f

$F_{cal} = 1.02$

The tabulated value of F at 5% level of significance with (9, 11) degrees of freedom is 2.90.

The calculated value of F is lessthan the tabulated value of F

$\therefore \quad F_{cal} < F_{tan}$

i.e., $1.02 < 2.90$

$\therefore$ we accept H_0

Conclusion

The samples are drawn from same normal population

15. Two random samples of 11,9 observations show the sample standard deviation of their weights 0.8 and 0.5 respectively. Assuming the weights distributions are normal. Test the hypothesis that there variances are gains the alternative hypothesis that they are not equal at 5% level of significance.

Sol :

Null Hypothesis(H_0)

The population variance are equal

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis(H_1)

The population variance are not equal

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$

Level of significance

Consider the appropriate at 5% level of significance.

Test statistic

Given that,

$n_1 = 11, \quad n_2 = 9$

$S_x = 0.8, \quad S_y = 0.5$

Now, $S_x^2 = (0.8)^2, \quad S_y^2 = (0.5)^2$

$S_x^2 = 0.64, \quad S_y^2 = 0.25$

$\therefore \; S_x^2 > S_y^2$

The test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ differ}$$

$$F = \frac{0.64}{0.25} \sim F(10, 8) \text{ differ}$$

$F_{cal} = 2.56$

The tabulated value of F at 5% level of significance with 10, 8 degree of freedom is 3.35

$\therefore$ The cal value of F is lessthan the tabulated value of F.

$\therefore \; F_{cal} < F_{tab}$

i.e., $2.56 < 3.35$

$\therefore$ we accept H_0

Conclusion

The population variance are equal.

16. Two independent samples of 8,7 items respectively as the following values of the variables.

Sample 1	Sample 2
9	10
11	12
13	10
11	14
15	9
19	8
12	10
14	

To estimate the population variance differ significantly.

Sol :

Null Hypothesis(H_0)

Population variances are equal

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis(H_1)

Population variances are not equal

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$

Level of significance

Consider the appropriate at 5% level of significance

Test statistic

Given that,

$n_1 = 8, \quad n_2 = 7$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
9	10	– 4	– 0.42	16	0.17
11	12	– 2	1.58	4	2.43
13	10	0	– 0.42	0	0.17
11	14	– 2	3.58	4	12.81
15	9	2	– 1.42	4	2.01
19	8	6	– 2.42	36	5.85
12	10	– 1	– 0.42	1	0.17
14		1		1	
$\Sigma x_i = 104$	$\Sigma y_i = 73$	$\Sigma(x_i - \bar{x}) = 0$	$\Sigma(y_i - \bar{y})^2 = 0.06$	$\Sigma(x_i - \bar{x})^2 = 66$	$\Sigma(y_i - \bar{y})^2 = 23.61$

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \bar{y} = \frac{\Sigma y_i}{n}$$

$$\bar{x} = \frac{104}{8} \qquad \bar{y} = \frac{73}{7}$$

$$\bar{x} = 13 \qquad \bar{y} = 10.42$$

$$S_x^2 = \frac{1}{n_1 - 1}[\Sigma(x_i - \bar{x})^2]$$

$$S_x^2 = \frac{1}{8-1}[66]$$

$$S_x^2 = \frac{66}{7}$$

$$S_x^2 = 9.42$$

$$S_y^2 = \frac{1}{n_2 - 1}[\Sigma(y_i - \bar{y})^2]$$

$$S_y^2 = \frac{1}{7-1}[23.61]$$

$$S_y^2 = \frac{23.61}{6}$$

$$S_y^2 = 3.93$$

$$\therefore S_x^2 > S_y^2$$

The test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ d.f}$$

$$F = \frac{9.42}{3.93} \sim F(7, 6) \text{ d.f}$$

$F_{cal} = 2.39$

The tabulated value of F at 5% level of significance with 7,6 degrees of freedom is 4.21

The calculated value of F is lessthan the tabulated value of F

$\therefore \quad F_{cal} < F_{tab}$

i.e., $\quad 2.39 < 4.21$

$\therefore$ we accept H_0

Conclusion

Population variances are equal

17. In one sample of 8 observations the sum of squares of deviations of the some values from the sample mean is 84.4 and other sample of 10 observations it was 102.6. Test whether significant different or not at 5% level of significance.

Sol :

Null Hypothesis(H_0)

Population variances are equal

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis(H_1)

Population variances are not equal

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$

Level of significance

Consider the appropriate at 5% level of significance

Test statistic

Given that,

$n_1 = 8, \quad n_2 = 10$

$\Sigma(x_i - \bar{x})^2 = 84.4, \quad \Sigma(y_i - \bar{y})^2 = 102.6$

$$S_x^2 = \frac{1}{n_1 - 1}[\Sigma(x_i - \bar{x})^2]$$

$$= \frac{1}{8-1}(84.4)$$

$$= \frac{84.4}{7}$$

$$S_x^2 = 12.05$$

$$S_y^2 = \frac{1}{n_2 - 1}[\Sigma(y_i - \bar{y})^2]$$

$$= \frac{1}{10-1}[102.6]$$

$$= \frac{102.6}{9}$$

$$S_y^2 = 11.4$$

$$\therefore \quad S_x^2 > S_y^2$$

Then the test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ d.f}$$

$$F = \frac{12.05}{11.4} \sim F(7, 9) \text{ d.f}$$

$$F_{cal} = 1.05$$

The tabulated value of F at 5% level of significance with 7, 9 degrees of freedom is 3.29.

The calculated value of F is lessthan the tabulated value of F

$\therefore \quad F_{cal} < F_{tab}$

i.e., $1.05 < 3.29$

$\therefore$ we accept H_0

Conclusion

Population variance are equal

18. The following table shows the yield of crops per 20 plots half of which are treated with fertilizers.

Treated	**5**	**0**	**8**	**3**	**6**	**1**	**0**	**3**	**3**	**1**
Untreated	**1**	**4**	**1**	**2**	**3**	**2**	**5**	**0**	**2**	**0**

whether the treatment by

(i) Change the variability of the plant crop yields

(ii) Improve the average yield of crops

Sol :

Null Hypothesis(H_0)

There is no significance difference between two variances.

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis(H_1)

There is significance differences between two variances

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$

Level of significance

Consider the appropriate at 5% level of significance

Test statistic

Given that,

$n_1 = 10, \quad n_2 = 10$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
5	1	2	– 1	4	1
0	4	– 3	2	9	4
8	1	5	– 1	25	1
3	2	0	0	0	0
6	3	3	1	9	1
1	2	– 2	0	4	0
0	5	– 3	3	9	9
3	0	0	– 2	0	4
3	2	0	0	0	0
1	0	– 2	– 2	4	4
$\Sigma x_i = 30$	$\Sigma y_i = 20$	$\Sigma(x_i - \bar{x}) = 0$	$\Sigma(y_i - \bar{y})^2 = 0$	$\Sigma(x_i - \bar{x})^2 = 64$	$\Sigma(y_i - \bar{y})^2 = 24$

$\bar{x} = \frac{\Sigma xi}{N} = \frac{30}{10}$ $\qquad$ $\bar{y} = \frac{\Sigma y_i}{N} = \frac{20}{10}$

$\bar{x} = 3$ $\qquad$ $\bar{y} = 2$

$S_x^2 = \frac{1}{n_1 - 1}[\Sigma(x_i - \bar{x})^2]$ $\qquad$ $S_y^2 = \frac{1}{n_2 - 1}[\Sigma(y_i - \bar{y})^2]$

$S_x^2 = \frac{1}{10-1}[64]$ $S_y^2 = \frac{1}{10-1}[24]$

$S_x^2 = \frac{64}{9}$ $S_y^2 = \frac{24}{9}$

$S_x^2 = 7.11$ $S_y^2 = 2.66$

$\therefore \quad S_x^2 > S_y^2$

Then the test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ differ}$$

$$F = \frac{7.11}{2.66} \sim F(9, 9) \text{ differ}$$

$F_{cal} = 2.67$

The tabulated value of F at 5% level of significance with 9, 9 degrees of freedom is 3.18.

The calculated value of F is lessthan the tabulated value of F

$\therefore \quad F_{cal} < F_{tab}$

i.e., $2.67 < 3.18$

$\therefore$ we accept H_0

Conclusion

There is no significance difference between two variances.

3.3 Tests of Significance based on χ^2

Q9. What is χ^2 - test? State the features of χ^2 - test.

Ans : **(Imp.)**

Chi-square (χ^2) test is defined as the quantity used to describe the magnitude of discrepancy or difference between observed and expected (theoretical) frequencies. It is also known as distribution free test.

The chi-square (χ^2) test derives its name from the Greek alphabet chi (χ), pronounced as kye/ki. Its use was first initiated by Karl Pearson in the year 1900, hence it is also known as Pearson's χ^2 test. It is very simple and the most frequently used non-parametric test. It is very beneficial when the parametric tests like Z, F and t-tests fails to obtain the results. In contrast to the parametric tests which deal with the quantitative information, χ^2 test deals with the qualitative aspects. It is especially used in genetical studies for determining whether the recorded information conforms to the hypothesis or not.

Features of Chi-square (χ^2) Test

1. Chi-square (χ^2) test is a common test which has its main use in research (sociological, psychological and business), behavioral sciences and genetics.
2. This test is mainly applied in determining the hypothesis in order to obtain inferences but has no use in estimation.
3. This test does not depend on mean, variance, standard deviation, etc. It mainly focuses on frequencies.
4. The test can be applied in between the whole range of frequencies, both observed and expected or (theoretical), while parametric tests are used for a single value.
5. With every rise in the number of degrees of freedom (*df*), a new chi-square value is obtained.

Q10. Explain the procedure for determining the value of χ^2.

Ans : **(Feb.-21, Imp.)**

The year 1990 χ^2 quantity % describes the magnitude of the discrepancy between theory and observation. It is defined as :

$$x^2 = \Sigma \frac{(O - E)^2}{E}$$

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

Steps

To determine the value of y^2. the steps required are:

(i) Calculate the expected frequencies. In general the expected frequency for any call can be calculated from the following equation :

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

E = Expected frequency

RT = The row total for the row containing the cell

CT = The column total for the column containing the cell

N = The total number of observations.

(ii) Take the difference between observed and expected frequencies and obtain the squares of these differences, i.e., obtain the values of $(O - E)^2$.

(iii) Divide the values of $(O - E)^2$ obtained in step (ii) by the respective expected frequency and obtain the total $\Sigma\,[(O - E)^2/E]$. This gives the value of χ^2 which can range from zero to infinity. If χ^2 is zero it means that the observed and expected frequencies completely coincide. The greater the discrepancy between the observed and expected frequencies, the greater shall be the value of χ^2.

The calculated value of χ^2 is compared with the table value of χ^2 for given degrees of freedom at a certain specified level of significance. If at the stated level (generally 5% level is selected), the calculated value of χ^2 is more than the table value of χ^2, the difference between theory and observation is considered to be significant, *i.e.*, it could not have arisen due to fluctuations of simple sampling. If, on the other hand, the calculated value of χ^2 is less than the table value, the difference between theory and observation is not considered as significant, i.e., it is regarded as due to fluctuations of simple sampling and hence ignored.

The computed value of χ^2 is a random variable which takes on different values from sample to sample. That is χ^2 has a sampling distribution just as do the other test statistics discussed in earlier chapter.

It should be noted that the value of χ^2 (is always positive and its upper limit is infinity. Also since χ^2 is derived from observations, it is a statistic and not a parameter (there is no parameter corresponding to it). The *y* test is. Therefore, termed non-parametric. It is one of the great advantages of this test that it involves no assumption about the form of the original distribution from which the observations come.

Q11. Explain the conditions for applying χ^2 test.

Ans :

The following conditions should be satisfied before applying the χ^2 test

1. In the first place N must be reasonably large to ensure the similarity between theoretically correct distribution and our sampling distribution of χ^2 the chi-square statistic. It is difficult to say what constitutes largeness, but as a general rule χ^2 test should not be used when N is less than 50, however few the cells.

2. No theoretical cell frequency should be small when the expected frequencies are too small, the value of χ^2 will be overestimated and will result in too many rejections of the null hypothesis. To avoid making incorrect inferences, a general rule is followed that expected frequency of less than 5 in one cell of a contingency table is too small to use. When the table contains more than one cell with an expected frequency of less than 5 we **"pool"** the frequencies which are less than 5 with the preceding or succeeding frequency so that the resulting sum is 5 or more. However, in doing so, we reduce the number of categories of data and will gain less information from contingency table.

3. The constraints on the cell frequencies if any should be linear, *i.e.,* they should not involve square and higher powers of the frequencies such as $\Sigma O = \Sigma E = N$.

Q12. Explain the uses of χ^2 test.

Ans : **(Imp.)**

The χ^2 test is one of the most popular statistical inference procedures today. It is applicable to a very large number of problems in practice which can be summed up under the following heads:

i) Test as a test of independence: With the help of χ^2 test we can find out whether two or more attributes are associated or not. Suppose we have N observations classified according to some attributes we may ask whether the attributes are related or independent. Thus, we can find out whether quinine is effective in controlling fever or not, whether there is any association between marriage and failure, or eye colour of husband and wife. In order to test whether or not the attributes are associated we take the null hypothesis that there is no association in the attributes under study or, in other words, the two attributes are independent. If the calculated value of χ^2 is less than the table value at a certain level of significance (generally 5% level), we say that the results of the experiment provide no evidence for doubting the hypothesis or, in other words the hypothesis that the attributes are not associated holds good. On the other hand, if the calculated value of χ^2 is greater than the table value at a certain level of significance, we say that the results of the experiment do not support the hypothesis or, in other words, the attributes are associated. It should be noted that χ^2 is not a measure of the degree or form of relationship, it only tells us whether two principles of classification are or are not significantly related, without reference to any assumptions concerning the form of relationship.

(ii) χ^2 test as a test of goodness of fit*, χ^2 test is very popularly known as test of goodness of fit for the reason that is enables us to ascertain how appropriately the theoretical distributions such as Binomial, Poisson, Normal, etc., fit empirical distributions, i.e., those obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested in finding out how well this curve Its with the observed facts. A test of the concordance (goodness of fit) of the two can be made just by inspection, but such a test is obviously inadequate. Precision can be secured by applying the γ^2 test.

3.3.1 χ^2 - Test for Specified Variance

Q13. Discuss χ^2 - test for Specified Variance.

Ans : **(June-18, Imp.)**

χ^2-test is used to determine whether the specified variance and the sample variance differs significantly. Here, the value of specified variance is the historical value or target value. This test is similar to t-test for a single sample mean.

The chi-square test(χ^2) for the specified variance is given by,

$$\chi^2 = (n-1)\frac{\sigma^2}{s^2}$$

Where,

n = Sample size

s^2 = Sample variance

σ^2 = Specified variance.

3.3.2 Goodness of Fit

Q14. Write a short notes on χ^2 - test for goodness of fit.

Ans : **(June-19, June-18, Imp.)**

The most powerful test for testing the significance difference between the theory and experiment is known as 'χ^2 - test for goodness of fit'.

It was proposed by 'proof karl pearson's in the year 1990.

Objective

It is used to find the significance difference between the observed (form experiment) and expected (from theory) frequencies.

Let 0, 0_2....0_n, $\in_1, \in_2$ .... $\in_n$ be the 'n' independent observed and expected frequencies.

Null Hypothesis(H_0)

The fitted distribution for the given data is the best fit.

Alternative Hypothesis(H_1)

The fitted distribution for the given data is not best fit.

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(0_i - \in_i)^2}{\in_i}\right] \sim \chi^2(\alpha, n-1)$$

Since χ^2 calculation is to be calculated and it should be compared with χ^2 tabulated value for (n – 1) degrees of freedom at the decided level of significance.

Conclusion

χ^2 calculation lessthan $\chi^2(\alpha, n-1)$ degrees of freedom then we accept H_0. Other wise we reject H_0.

PROBLEMS

19. The following table given no. of air craft accidents that occurred during the 6 days of the week. Find test whether the accidents are uniformly distributed over the week.

Day	Mon	Tue	Wed	Thu	Fri	Sat
Accidents	14	18	12	11	15	14

Sol :

Null Hypothesis(H_0)

The accident are uniformly distributed over the week.

Alternative Hypothesis(H_1)

The accidents are not uniformly distributed over the week.

Level of significance

Consider the appropriate at 5% level of significance for (6 – 1) = 5 degrees of freedom

Test statistic

Days	Accidents 0_i	$\in_i$	$(0_i - \in_i)^2$	$(0_i - \in_i)^2/\in_i$
Mon	14	14	0	0
Tue	18	14	16	1.14
Wed	12	14	4	0.28
Thu	11	14	9	0.64
Fri	15	14	1	0.07
Sat	14	14	0	0
	$\Sigma 0_i = 84$			2.13

Expected frequency $\in_i = \dfrac{\Sigma 0_i}{n}$

$$\in_i = \frac{84}{6}$$

$$\in_i = 14$$

χ^2 calculation $\sum_{i=1}^{n}\left[\dfrac{(0_i - \in_i)^2}{\in_i}\right] \sim \chi^2(\alpha, n-1)$

$\therefore$ χ^2 calculation = 2.13

The tabulated value of χ^2 at 5% level of significance with (6 – 1) = 5 degrees of freedom is 11.07.

The calculated value of χ^2 is lessthan the tabulate value of χ^2.

i.e., 2.13 < 11.07

$\therefore$ we accept H_0

Conclusion

The accidents are uniformly distributed over the week.

20. The die is thrown 6 times

Phase	1	2	3	4	5	6
Dies	8	7	12	8	14	11

To test the 5% level of significance of the die is unbiased.

Sol :

Null Hypothesis(H_0)

The die is unbiased

Alternative Hypothesis(H_1)

The die is biased

Level of significance

Consider the appropriate at 5% level of significance for (6 – 1) = 5 degrees of freedom

Test statistic

Phase	0_i	$\in_i$	$(0_i - \in_i)^2$	$(0_i - \in_i)^2/\in_i$
1	8	10	4	0.4
2	7	10	9	0.9
3	12	10	4	0.4
4	8	10	4	0.4
5	14	10	16	1.6
6	11	10	1	0.1
	$\Sigma 0_i = 60$			3.8

Expected frequency $\in_i = \dfrac{\Sigma 0_i}{n}$

$$\in_i = \frac{60}{6}$$

$$\in_i = 10$$

χ^2 calculation $\sum_{i=1}^{n}\left(\frac{(0_i - \in_i)^2}{\in_i}\right) \sim \chi^2\ (\alpha, n-1)$

$\therefore$ χ^2 calculation = 3.8

The tabulated value of χ^2 at 5% level of significance with (6 _ 1) = 5 degrees of freedom is 11.07.

The calculated value of χ^2 is lessthan the tabulated value of χ^2

i.e., 3.8 < 11.07

$\therefore$ we accept H_0

Conclusion

The die is unbiased

21. The theory predict proportion of be answer in the four graphs A, B, C, D should be 9:3:3:1. In an experiment among 1600 be answer the number of four graphs where 882, 313, 287,313, 287, 118. Thus the experiment result support the theory.

Sol :

Null Hypothesis(H_0)

The experiment result support the theory.

Alternative Hypothesis(H_1)

The experiment result does not support the theory.

Level of significance

Consider the appropriate at 5% level of significance.

Test statistic

Given observed frequency are 882, 313, 287, 118.

$$E(882) = \frac{9}{16} \times 1600 = 900$$

$$E(313) = \frac{3}{16} \times 1600 = 300$$

$$E(287) = \frac{3}{16} \times 1600 = 300$$

$$E(118) = \frac{1}{16} \times 1600 = 100$$

0_i	$\in_i$	$(0_i - \in_i)^2$	$(0_i - \in_i)^2 / \in_i$
882	900	324	0.36
313	300	169	0.56
287	300	169	0.56
118	100	324	3.24
			4.72

χ^2 calculation $\sum_{i=1}^{n}\left[\frac{(0_i - \in_i)^2}{\in_i}\right] \sim \chi^2 (\alpha, n - 1)$

$\therefore$ χ^2 calculation = 4.72

The tabulated value of χ^2 at 5% level of significance is 7.81

The calculated value of χ^2 is lessthan the tabulated values of χ^2

i.e., 4.72 < 7.81

$\therefore$ we accept H_0

Conclusion

The experiment result support the theory.

3.3.3 Test for Independence of Attributes

Q15. Explain χ^2 - test for independent of two attribute.

Ans : **(Feb.-21, June-19, June-18)**

Let us consider two attributes A and B. Let the attribute A is divided into 'r' classes i.e., A_1, A_2....A_r and B is divided into 'S' classes i.e., B_1, B_2.....B_S

Then the following rxs contingency table express the self frequencies of the attributes A_i and B_j; i = 1, 2,......S

A \ B	B_1	B_2	....	B_j	....	B_S	Total
A_1	(A_1B_1)	(A_1B_2)	....	(A_1B_j)	....	(A_1B_S)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	....	(A_2B_j)	....	(A_2B_S)	(A_2)
⋮	⋮	⋮		⋮		⋮	⋮
A_i	(A_iB_1)	(A_iB_2)	....	(A_iB_j)	....	$A_iB_S)$	(A_i)
⋮	⋮	⋮		⋮		⋮	⋮
A_r	(A_rB_1)	(A_rB_2)	....	(A_rB_j)	....	B_S	(A_r)
Total	(B_1)	(B_2)	....	(B_j)	....	(B_S)	N

where,

(A_i) = Number of persons having the attribute A_i

(B_j) = Number of persons having the attribute B_j

(A_iB_j) = Number of persons having both the attributes A_i and B_j

and $$\sum_{i=1}^{r} A_i = \sum_{j=1}^{s} B_j = N$$

and the probabilities are

$$P(A_i) = \frac{(A_i)}{N}$$

$$P(B_j) = \frac{(B_j)}{N}$$

$$P(A_iB_j) = P(A_i).P(B_j)$$

$$P(A_iB_j) = \frac{(A_i)}{N} = \frac{(B_j)}{N}$$

The expected frequency are

$$E(A_iB_j) = N.P(A_iB_j)$$

$$E(A_iB_j) = \cancel{N} \cdot \frac{(A_i)}{\cancel{N}} \cdot \frac{(B_j)}{N}$$

$$E(A_iB_j) = \frac{(A_i)\cdot(B_j)}{N}$$

Null Hypothesis(H_0)

The attributes A and B are independent.

Alternative Hypothesis(H_1)

The attribute A and B are not independent.

Level of significance

Consider the appropriate at α% level of significance.

Test statistic

Under the Null hypothesis then the test statistic is given by

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{(0_{ij} - \epsilon_{ij})^2}{\epsilon_{ij}}\right] \sim \chi^2[\alpha, (r-1)(s-1) \text{ differ}$$

where, $0_{ij} = (A_iB_j)$

$$\epsilon_{ij} = \frac{(A_i)(B_j)}{N}; \qquad i = 1, 2....r$$

$$j = 1, 2....s$$

Since χ^2 calculation is to be calculate and it should be compared with χ^2 tabulated value for [(r–1)(s–1)] degrees of freedom at the decided level of significance.

Conclusion

It χ^2 calculated value is lessthan the χ^2 tabulated value then we accept H_0 otherwise we reject H_0.

Theorem

For a 2 × 2 contingency table

Attribute - B

Attribute - A		
	a	**b**
	c	d

Then prove that $\chi^2 = \dfrac{N[ad - bc]^2}{(a+b)(a+c)(c+d)(b+d)}$

where

N = a + b + c + d

Proof :

Null Hypothesis(H_0)

The attributes A and B are independent.

Alternative Hypothesis(H_1)

The attribute A and B are not independent

2 × 2 contingency table.

A \ B	B_1	B_2	Total
A_1	$0_{11} = a$	$0_{12} = b$	$(A_1) = a+b$
A_2	$0_{21} = c$	$0_{22} = d$	$(A_2) = c+d$
Total	$(B_1) = a+c$	$(B_2) = b+d$	$N = a+b+c+d$

Expected Frequency

$$\in_{ij} = \frac{(A_i)(B_j)}{N}$$

$$\in_{11} = \frac{(A_1)(B_1)}{N} = \frac{(a+b)(a+c)}{N}, \qquad \in_{21} = \frac{(A_2)(B_1)}{N} = \frac{(c+d)(a+c)}{N}$$

$$\in_{12} = \frac{(A_1)(B_2)}{N} = \frac{(a+b)(b+d)}{N}, \qquad \in_{22} = \frac{(A_2)(B_2)}{N} = \frac{(c+d)(b+d)}{N}$$

By the definition of χ^2 - test for independent of two attributes is

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{(0_{ij} - \in_{ij})^2}{\in_{ij}}\right] \sim \chi^2[(r-1)\ (s-1)]\ \text{d.f}$$

$$\chi^2 = \left[\frac{\left(a - \frac{(a+b)(a+c)}{N}\right)^2}{\frac{(a+b)(a+c)}{N}}\right] + \left[\frac{\left(b - \frac{(a+b)(b+d)}{N}\right)^2}{\frac{(a+b)(b+d)}{N}}\right] + \left[\frac{\left(c - \frac{(c+d)(a+c)}{N}\right)^2}{\frac{(c+d)(a+c)}{N}}\right]$$

$$+ \left[\frac{\left(d - \frac{(c+d)(b+d)}{N}\right)^2}{\frac{(c+d)(b+d)}{N}}\right]$$

$$\chi^2 = \left[\frac{\frac{[a(N)-(a+b)(a+c)]^2}{N^2}}{\frac{(a+b)(a+c)}{\cancel{N}}}\right] + \left[\frac{\frac{[b(N)-(a+b)(b+d)]^2}{N^2}}{\frac{(a+b)(b+d)}{\cancel{N}}}\right] + \left[\frac{\frac{[c(N)-(c+d)(a+c)]^2}{N^2}}{\frac{(c+d)(a+c)}{\cancel{N}}}\right]$$

$$+ \left[\frac{\frac{[d(N)-(c+d)(b+d)]^2}{N^2}}{\frac{(c+d)(b+d)}{\cancel{N}}}\right]$$

$$\chi^2 = \frac{1}{N}\left[\frac{[a(a+b+c+d)-(a+b)(a+c)]^2}{(a+b)(a+c)} + \frac{[b(a+b+c+d)-(a+b)(b+d)]^2}{(a+b)(b+d)}\right.$$

$$+\frac{\left[c\left(a+b+c+d\right)-(c+d)(a+c)\right]^2}{(c+d)(a+c)}+\frac{\left[d(a+b+c+d)-(c+d)(b+d)\right]^2}{(a+b)(b+d)}\Bigg]$$

$$\chi^2 = \frac{1}{N}\left[\frac{\left[(a^2+ab+ac+ad)-(a^2+ac+ab+bc)\right]^2}{(a+b)(a+c)}+\frac{\left[(ab+b^2+bc+cd)-(ab+ad+b^2+bd)\right]^2}{(a+b)(b+d)}\right.$$

$$\left.\frac{\left[(ac+bc+c^2+cd)-(ac+c^2+ad+cd)\right]^2}{(c+d)(a+c)}+\frac{\left[(ad+bd+cd+d)^2-(bc+cd+bd+d^2)\right]^2}{(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{1}{N}\left[\frac{\left[\cancel{a^2}+ab+\cancel{ac}+ad-\cancel{a^2}-\cancel{ac}-\cancel{ab}-bc\right]^2}{(a+b)(a+c)}+\frac{\left[\cancel{ab}+\cancel{b^2}+bc+bd-\cancel{ab}-ad-\cancel{b^2}-\cancel{bd}\right]^2}{(a+b)(b+d)}\right.$$

$$\chi^2 = \frac{1}{N}\left.\frac{\left[\cancel{ac}+bc+\cancel{c^2}+\cancel{cd}-\cancel{ac}-\cancel{c^2}-ad-\cancel{cd}\right]^2}{(c+d)(a+c)}+\frac{\left[ad+\cancel{bd}+\cancel{cd}+\cancel{d^2}-bc-\cancel{cd}-\cancel{bd}-\cancel{d^2}\right]^2}{(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{1}{N}\left[\frac{(ad-bc)^2}{(a+b)(a+c)}+\frac{(-1)^2(ad-bc)^2}{(a+b)(b+d)}+\frac{(-1)^2(ad-bc)^2}{(c+d)(a+c)}+\frac{(ad-bc)^2}{(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{(ad-bc)^2}{N}\left[\frac{1}{(a+b)(a+c)}+\frac{1}{(a+b)(b+d)}+\frac{1}{(c+d)(a+c)}\ \frac{1}{(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{(ad-bc)^2}{N}\left[\frac{(c+d)(b+d)+(a+c)(c+d)+(a+b)(b+d)+(a+b)(a+c)}{(a+b)(a+c)(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{(ad-bc)^2}{N}\left[\frac{bc+cd+bd+d^2+ac+ad+c^2+cd+ab+ad+b^2+bd+a^2+ac+ab+bc}{(a+b)(a+c)(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{(ad-bc)^2}{N}\left[\frac{(a+b+c+d)^2}{(a+b)(a+c)(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{(ad-bc)^2}{\cancel{N}}\left[\frac{N^2}{(a+b)(a+c)(c+d)(b+d)}\right]$$

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b(a+c)(c+d)(b+d)}$$

$\therefore$ Hence proved

Yate's Correction

In a 2×2 contingency table if any self frequency is lessthan 5 is

$$\chi^2 = \frac{N\left[(ad - bc) - \frac{N}{2}\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

PROBLEMS

22. Two treatments have applied on 500 agricultural plots and the given below

Treatment - I \ Treatment - II		
	208	92
	32	168

Test whether the treatments are independent or not.

Sol :

Null Hypothesis(H_0)

The treatments are independent

Alternative Hypothesis(H_1)

The treatments are not independent

Level of significance

Consider the appropriate at 5% level of significance with (1) degrees of freedom.

Test statistic

Under the Null hypothesis then the test statistic is given by

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{(0_{ij} - \in_{ij})^2}{\in_{ij}}\right] \sim \chi^2\ [(r-1)\ (s-1)]\ d.f$$

Instead of this we can use for 2×2 contingency table then

$$\chi^2 = \frac{N[ad - bc)]^2}{(a+c)(b+d)(a+b)(c+d)}$$

Let The Treatment - I is A

The Treatment - II is B

The attribute A and B are not independent 2 × 2 contingency table.

A \ B	B_1	B_2	Total
A_1	a = 208	b = 92	a+b = 300
A_2	c = 32	d = 168	c +d = 200
Total	a+c = 240	b+d = 260	N = a+b+c+d = 500

$$\chi^2 = \frac{500[208(168) - 92(32)]^2}{(240)(260)(300)(200)}$$

$$\chi^2 = \frac{500[34944 - 2944]^2}{3744000000}$$

$$\chi^2 = \frac{500[32000]^2}{3744000000}$$

$$\chi^2 = \frac{5.12 \times 10^{11}}{3744000000}$$

$$\chi^2 = 136.75$$

The tabulated value of χ^2 at 5% level of significance for (1) degrees of freedom is 1.84.

χ^2 Calculation is greaterthan the χ^2 tabulated

$\therefore \quad \chi^2_{cal} > \chi^2_{tab}$

i.e., 136.75 > 1.84

$\therefore$ we reject H_0

Conclusion

The two treatments are not independent.

Short Question and Answers

1. Explain t-test for single mean.

Ans :

Let x_1, x_2.....x_n be a random samples drawn from the normal population with specified mean and variance then

Null Hypothesis(H_0)

There is no significance difference between population mean and sample mean

(or)

The samples have been drawn from a population.

i.e., $H_0 : \mu = \mu_0$

Alternative Hypothesis(H_1)

There is significance difference between population mean and sample mean.

(or)

The samples have not been drawn from population

i.e., $H_1 : \mu_1 \neq \mu_0$

Level of Significance

Consider the appropriate at α% level of significance.

Test statistic

The test statistic is given by

$$t = \left|\frac{\bar{x} - \mu}{s / \sqrt{n-1}}\right| \sim t_{(n-1)} \text{ d.f}$$

where,

$\bar{x}$ = sample mean

μ = population

Remarks

if S is unknown then the test statistic is given by

$$t = \left|\frac{\bar{x} - \mu}{s\sqrt{n}}\right| \sim t_{(n-1)} \text{ d.f}$$

Confidence limits

(i) The 95% confidence limits for t-test for single mean is $\bar{x} \pm t_{0.05}\ s\sqrt{n}$.

(ii) The 99% confidence limits for t-test for single mean is $\bar{x} \pm t_{0.01}\ s\sqrt{n}$.

Conclusion

If the calculated value of t is lessthan the tabulated value of t at certain level of significance for (n – 1) degrees of freedom the we accept H_0 otherwise we reject H_0

2. Explain χ^2 - test for independent of two attribute.

Ans :

Let us consider two attributes A and B. Let the attribute A is divided into 'r' classes i.e., $A_1, A_2....A_r$ and B is divided into 'S' classes i.e., $B_1, B_2.....B_S$

Then the following rxs contingency table express the self frequencies of the attributes A_i and B_j; i = 1, 2,......S

A \ B	B_1	B_2	....	B_j	....	B_S	Total
A_1	(A_1B_1)	(A_1B_2)	....	(A_1B_j)	....	(A_1B_S)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	....	(A_2B_j)	....	(A_2B_S)	(A_2)
⋮	⋮	⋮		⋮		⋮	⋮
A_i	(A_iB_1)	(A_iB_2)	....	(A_iB_j)	....	$A_iB_S)$	(A_i)
⋮	⋮	⋮		⋮		⋮	⋮
A_r	(A_rB_1)	(A_rB_2)	....	(A_rB_j)	....	B_S	(A_r)
Total	(B_1)	(B_2)	....	(B_j)	....	(B_S)	N

where,

(A_i) = Number of persons having the attribute A_i

(B_j) = Number of persons having the attribute B_j

(A_iB_j) = Number of persons having both the attributes A_i and B_j

and $\sum_{i=1}^{r} A_i = \sum_{j=1}^{S} B_j = N$

and the probabilities are

$$P(A_i) = \frac{(A_i)}{N}$$

$$P(B_j) = \frac{(B_j)}{N}$$

$$P(A_iB_j) = P(A_i).P(B_j)$$

$$P(A_iB_j) = \frac{(A_i)}{N} = \frac{(B_j)}{N}$$

The expected frequency are

$$E(A_iB_j) = N.P(A_iB_j)$$

$$E(A_iB_j) = \cancel{N}\cdot\frac{(A_i)}{\cancel{N}}\cdot\frac{(B_j)}{N}$$

$$E(A_iB_j) = \frac{(A_i)\cdot(B_j)}{N}$$

3. Write a short notes on χ^2 - test for goodness of fit.

Ans :

The most powerful test for testing the significance difference between the theory and experiment is known as 'χ^2 - test for goodness of fit'.

It was proposed by 'proof karl pearson's in the year 1990.

Objective

It is used to find the significance difference between the observed (form experiment) and expected (from theory) frequencies.

Let 0, 0_2....0_n, $\in_1, \in_2$ ....$\in_n$ be the 'n' independent observed and expected frequencies.

Null Hypothesis(H_0)

The fitted distribution for the given data is the best fit.

Alternative Hypothesis(H_1)

The fitted distribution for the given data is not best fit.

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(0_i - \in_i)^2}{\in_i}\right] \sim \chi^2(\alpha, n-1)$$

Since χ^2 calculation is to be calculated and it should be compared with χ^2 tabulated value for (n – 1) degrees of freedom at the decided level of significance.

Conclusion

χ^2 calculation lessthan $\chi^2(\alpha, n-1)$ degrees of freedom then we accept H_0. Other wise we reject H_0.

4. Explain the concept of F-test for equality of population variance.

Ans :

Suppose we want to test whether two independent sample x_i ; i = 1, 2.....n_1 and y_i ; i = 1, 2 .....n_2 these samples has been drawn from the normal population with the same variance σ^2

Objective

In this procedure to test the two independent estimators of the population variance are equal or not.

Null Hypothesis(H_0)

The population variance are equal

$$\text{i.e.,}\quad H_0 : \sigma_x^2 = \sigma_y^2 = \sigma_0^2$$

Alternative Hypothesis(H_1)

The population variance are not equal

$$\text{i.e.,}\quad H_1 : \sigma_x^2 = \sigma_y^2 = \sigma_0^2$$

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$F = \frac{S_x^2}{S_y^2} \sim F(v_1, v_2) \text{ d.f; If } (S_x^2 > S_y^2)$$

$$F = \frac{S_y^2}{S_x^2} \sim F(v_1, v_2) \text{ d.f; If } (S_y^2 > S_x^2)$$

where,

v_1 = $(n_1 - 1)$ d.f

v_2 = $(n_2 - 1)$ d.f

and $$S_x^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1}(x_i - \bar{x})^2\right]$$

$$S_x^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n}x^2 - n(\bar{x})^2\right]$$

$$S_y^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2}(y_i - \bar{y})^2\right]$$

$$S_y^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2}(y_i)^2 - n(\bar{y})^2\right]$$

Conclusion

If the calculated value of F is lessthan the tabulated value of F at certain level of significance with $v_1 = (n_1 - 1)$ and $v_2 = (n_2 - 1)$ degrees of freedom then we accept H_0 otherwise we reject H_0.

5. Properties of Student's t-Distribution

Ans :

(i) The probability curve of t is symmetric, like in standard normal distribution (z).

(ii) The distribution ranges from $-\alpha$ to a just as does a normal distribution.

(iii) The t-distribution is bell shaped and symmetrical around mean zero, like normal distribution.

(iv) The shapes of the t-distribution changes as the sample size changes (the number of degrees of freedom changes) whereas it is same for all sample sizes in z-distribution.

(v) The variance of t-distribution is always greater than one and is defined only when $n > 3$.

(vi) The t-distribution is more of platykurtic (less packed at centre and higher in tails) than the normal distribution.

(vii) The t-distribution has a greater dispersion than the normal distribution. As n becomes larger, the t-distribution approaches the standard normal distribution.

6. Define small sample test.

Ans :

The sample size $n < 30$ then it is called small - small sample test. We will discuss the following tests.

i) t-test

(a) t-test for single mean

(b) t-test for difference of two mean (Independent samples)

(c) t-test for difference of two means (Dependent samples)

(or)

Paired t-test.

(d) t-test for correlation coefficient

ii) F - test

(a) F-test for equality of population variance

iii) χ^2 - test

(a) χ^2 -test for goodness of fit

(b) χ^2 -test for independent of two attributes.

7. Applications of t-Distribution

Ans :

The following are some important applications of t-distribution.

(i) Test of hypothesis about the population mean.

(ii) Test of hypothesis about the difference between two means.

(iii) Test of hypothesis about the difference between two means with dependent samples.

(iv) Test of hypothesis about coefficient of correlation.

8. Explain t-test for correlation coefficient.

Ans :

It 'r' is the observed sample correlation coefficient in a sample of size 'n' pairs of observations from a bivariate normal population.

The R.A. Fisher proved that under the

Null Hypothesis(H_0)

The population correlation coefficient is not significant.

i.e., $H_0 : \int = \int_0$

Alternative Hypothesis(H_1)

The population correlation coefficient is significant

i.e., $H_1 : \int \neq \int_0$

Level of significance

Consider the appropriate at α% level of significance

Test statistic

The test statistic is given by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} \text{ d.f}$$

i.e., It follows t-distribution with (n – 2) degrees of freedom

Conclusion

The calculated value of t is lessthan the tabulated value of t at α% level of significance with (n – 2) degree of freedom then we accept H_0 otherwise we reject H_0.

9. Applications of F-Test.

Ans :

i) It is used for testing the equality of many population means.

ii) It is used for comparing he sample variances.

iii) It is used for performing analysis of variance.

iv) It is used for testing the significance of regression equation.

v) It is used for determining whether the ratio incrementally changes from unity at any level chosen randomly.

10. What is χ^2 - test?

Ans :

Chi-square (χ^2) test is defined as the quantity used to describe the magnitude of discrepancy or difference between observed and expected (theoretical) frequencies. It is also known as distribution free test.

The chi-square (χ^2) test derives its name from the Greek alphabet chi (χ), pronounced as kye/ki. Its use was first initiated by Karl Pearson in the year 1900, hence it is also known as Pearson's χ^2 test. It is very simple and the most frequently used non-parametric test. It is very beneficial when the parametric tests like Z, F and t-tests fails to obtain the results. In contrast to the parametric tests which deal with the quantitative information, χ^2 test deals with the qualitative aspects. It is especially used in genetical studies for determining whether the recorded information conforms to the hypothesis or not.

11. Features of Chi-square (χ^2) Test

Ans :

i) Chi-square (χ^2) test is a common test which has its main use in research (sociological, psychological and business), behavioral sciences and genetics.

ii) This test is mainly applied in determining the hypothesis in order to obtain inferences but has no use in estimation.

iii) This test does not depend on mean, variance, standard deviation, etc. It mainly focuses on frequencies.

iv) The test can be applied in between the whole range of frequencies, both observed and expected or (theoretical), while parametric tests are used for a single value.

v) With every rise in the number of degrees of freedom (*df*), a new chi-square value is obtained.

12. Explain the conditions for applying χ^2 test.

Ans :

The following conditions should be satisfied before applying the χ^2 test

1. In the first place N must be reasonably large to ensure the similarity between theoretically correct distribution and our sampling distribution of χ^2 the chi-square statistic. It is difficult to say what constitutes largeness, but as a general rule χ^2 test should not be used when N is less than 50, however few the cells.

2. No theoretical cell frequency should be small when the expected frequencies are too small, the value of χ^2 will be overestimated and will result in too many rejections of the null hypothesis. To avoid making incorrect inferences, a general rule is followed that expected frequency of less than 5 in one cell of a contingency table is too small to use. When the table contains more than one cell with an expected frequency of less than 5 we **"pool"** the frequencies which are less than 5 with the preceding or succeeding frequency so that the resulting sum is 5 or more. However, in doing so, we reduce the number of categories of data and will gain less information from contngency table.

3. The constraints on the cell frequencies if any should be linear, *i.e.,* they should not involve square and higher powers of the frequencies such as $\Sigma O = \Sigma E = N$.

13. Explain the uses of χ^2 test.

Ans :

The χ^2 test is one of the most popular statistical inference procedures today. It is applicable to a very large number of problems in practice which can be summed up under the following heads:

i) Test as a test of independence: With the help of χ^2 test we can find out whether two or more attributes are associated or not. Suppose we have N observations classified according to some attributes we may ask whether the attributes are related or independent. Thus, we can find out whether quinine is effective in controlling fever or not, whether there is any association between marriage and failure, or eye colour of husband and wife. In order to test whether or not the attributes are associated we take the null hypothesis that there is no association in the attributes under study or, in other words, the two attributes are independent. If the calculated value of χ^2 is less than the table value at a certain level of significance (generally 5% level), we say that the results of the experiment provide no evidence for doubting the hypothesis or, in other words the hypothesis that the attributes are not associated holds good. On the other hand, if the calculated value of χ^2 is greater than the table value at a certain level of significance, we say that the results of the experiment do not support the hypothesis or, in other words, the attributes are associated. It should be noted that χ^2 is not a measure of the degree or form of relationship, it only tells us whether two principles of classification are or are not significantly related, without reference to any assumptions concerning the form of relationship.

(ii) **χ^2 test as a test of goodness of fit*, χ^2** test is very popularly known as test of goodness of fit for the reason that is enables us to ascertain how appropriately the theoretical distributions such as Binomial, Poisson, Normal, etc., fit empirical distributions, i.e., those obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested in finding out how well this curve Its with the observed facts. A test of the concordance (goodness of fit) of the two can be made just by inspection, but such a test is obviously inadequate. Precision can be secured by applying the γ^2 test.

14. Discuss χ^2 - test for Specified Variance.

Ans :

χ^2-test is used to determine whether the specified variance and the sample variances differs significantly. Here, the value of specified variance is the historical value or target value. This test is similar to t-test for a single sample mean.

The chi-square test(χ^2) for the specified variance is given by,

$$\chi^2 = (n - 1)\frac{\sigma^2}{s^2}$$

Where,

n = Sample size

s^2 = Sample variance

σ^2 = Specified variance.

Exercise Problems

1. The following figures show the distribution of digits in numbers chosen at random from a telephone directory:

Digits	0	1	2	3	4	5	6	7	8	9	Total
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory.

(Ans : 58.54)

2. It is believed that the precision of an instrument is not more than 0.16. Write down the null and alternative hypothesis for testing this belief. Carry out the test at 1% level given 11 measurements of the same subject on the instrument:

2.5, 2.3, 2.4, 2.3, 2.5, 2.7, 2.5, 2.6, 2.6, 2.7, 2.5.

(Ans : 1.182)

3. Two sample pools of votes for two candidates A and B for a public office are taken, one from among the residents of rural areas. The results are given in the adjoining table. Examine whether the nature of the area is related to voting preference in this election.

	Votes for		
Area	**A**	**B**	**Total**
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

(Ans : 10.08)

4. Output of 8,000 graduates in a town 800 are females, out of 1,600 graduate employees 120 are females. Use χ^2 to determine if any distinction is made in appointment on the basis of sex. Value of χ^2 at 5% level for one degree of freedom is 3.84.

(Ans : 13.89)

5. The following results were obtained when two sets of times were subjected to two different treatments x and y to enhance their tensile strength.

Treatment X was applied on 400 items and 80 were found to have gained in strength. Treatment y was applied on 400 items and 20 were found to have gashed in strength. Is treatment y superior to treatment x? Use the χ^2-test.

(Ans : x^2 = 41.14)

6. Two independent random samples, each of 8 individuals provide the following data. Estimate the variance ratio and test the significance.

Sample-I	63	64	65	65	66	66	67	68
Sample-II	69	66	67	67	66	68	69	69

The value of F at 5% level for 7 degrees of freedom is 3.80 approximately.

(Ans: 1.52)

Choose the Correct Answer

1. _______ is defined as the quantity used to describe the magnitude of discrepancy or difference between observed and expected frequencies. [d]

   (a) F - test (b) t - test

   (c) z - test (d) chi-square test

2. _______ is used for performing analysis of variance [a]

   (a) f - distribution (b) t - distribution

   (c) z - distribution (d) chi-square test

3. The t-distribution ranges from _______ just as does a normal distribution. [b]

   (a) 0 to ∞ (b) $-\infty$ to ∞

   (c) $-\infty$ to 0 (d) $-\infty$ to 1

4. If the calculated value 06 $|t|$ is greater than the tabulated value for (n –1) d. fat 5% level then _______ the hypothesis. [b]

   (a) Accepted (b) Rejected

   (c) Both (a) & (b) (d) None of the above

5. Based on the null hypothesis, $H_0 = P = 0$. Which means population correlation coefficient is zero, its test statistic is given as _______ [c]

   (a) $t = \frac{r}{(1-r^2)} \times \sqrt{n-2}$ (b) $t = \frac{r}{\sqrt{(1-r^2)}} \times n-2$

   (c) $t = \frac{r}{\sqrt{(1-r^2)}} \times \sqrt{n-2}$ (d) $t = \frac{r}{\sqrt{(1-r^2)}} \times (n-2)^2$

6. _______ is used for determining whether the ratio incrementally changes from unity at any level chosen randomly [d]

   (a) t - distribution (b) z-distribution

   (c) shi square test (d) F-distribution

7. The chisquare (χ^2) test is given by _______. [a]

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

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

8. The object of the ______ is used to find out whether two independent estimates of population variance differ significantly. [c]

   (a) t - test (b) z-test

   (c) f - test (d) chi square test

9. Which is a valid condition for the χ^2 test ? [b]

   (a) The number of observations N must be very small

   (b) The no. of classes 'p' must be neither too large nor small

   (c) Idividual frequencies must be large

   (d) None of above

10. If $S_1^2 =$ 666.69 and $S_2^2 = 1109.33$ The F = ______. [a]

   (a) 1.66 (b) 1.5

   (c) 1.52 (d) 2.0

Fill in the blanks

1. ________ is defined as the quantity used to describe the magnitude of discrepancy (or) difference between observed and expected frequencies.
2. Chi-square test is also known as ________
3. In ________, the numerator variance must be always greater than the denominator variance.
4. The ________ has a greater dispersions than the standard normal distribution.
5. The degrees of freedom for a sample of size n is ________.
6. The sample observations should be ________ of each other in a chi square test.
7. The total number of observation used for chisquare test must greater than ________.
8. The t-test for the difference of means for independent samples is given by________.
9. ________ is a continuous probability distribution used when two different normal population are sample.
10. t-distribution is used to test the ________ and ________.

ANSWERS

1. Chi-square test
2. Distribution free test
3. F - test
4. t - distribution
5. n – 1
6. Independent
7. 50
8. $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d}{Sp\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
9. f - distribution
10. Population mean, Difference between two means.

Very Short Questions and Answers

1. Define chi square (χ^2) test

Ans :

Chi square test is defined as the quantity used to describe the magnitude of discrepancy or difference between observed and expected frequencies.

2. Write the formula for t-distribution.

Ans :

The formula for t-distribution is given by, $t = \dfrac{(\bar{x}-\mu)}{\sigma_s / \sqrt{n}}$

3. Write the formula for F ratio.

Ans :

The formula for F ratio is given by, $F = \dfrac{S_1^{\,2}}{S_2^{\,2}}$

Where $S_1^{\,2} = \dfrac{1}{n_1-1}\Sigma(x-\bar{x})^2$ and $S_2^{\,2} = \dfrac{1}{n_1-1}\Sigma(y-\bar{y})^2$

4. Write any one property of t-distribution.

Ans :

The probability curve of t is symmetric, Like in standard normal distribution.

5. Write any two application of F- distribution.

Ans :

- F-distribution is used for comparing the sample variances.
- It is used for performing analysis of variances.

UNIT IV	Non-parametric tests - their advantages and disadvantages, comparison with parametric tests.Measurement scale - nominal, ordinal, interval and ratio. Use of Central Limit Theorem intesting. One sample runs test, sign test and Wilcoxon-signed rank tests (single and pairedsamples). Two independent sample tests: Median test, Wilcoxon –Mann-Whitney U test,Wald Wolfowitz's runs test. Use of central limit theorem in testing

4.1 Non-parametric Tests

4.1.1 Their Advantages and Disadvantages

Q1. What is meant by Non-parametric Tests. State their Advantages and Dis-advantages.

Ans : **(June-19, June-18)**

A statistical test which does not depend on the particular form of frequency function from which the samples are drawn are not related with estimating the parameters is called Non-parametric test.

Assumption

1. Sample observations are independent.
2. The variable under study us continuous.
3. The probability density function is continuous.
4. Lower order moments exists.

Advantages

(i) Non-parametric methods are very simple and easy to apply.

(ii) Non-parametric methods do not required much complicated sampling theory.

(iii) No assumption is made about the form of frequency function of the parent population from which the samples have drawn.

(vi) Generally the socio-economic data are normally distributed, non-parametric methods have found applications in sociology.

(v) If the data is given in ranks (or) signs (or) grade (A, A^+, A^-, B, B^+, B^-....) then the Non-parametric methods are more useful and easy to calculate but parametric methods cannot be apply.

(vi) No parametric technique will apply to the data which are measured in nominal scale while Non-parametric methods exist to deal with such data.

(vii) It the Non-parametric methods can be used even if the sample size 6 (or) more.

(viii) Non-parametric test may be quiet powerful even if the sample size is small.

Disadvantages of Non-parametric test

(i) Non-parametric test can we used only the measurement are nominal (or) ordinal even in that case if a parametric test exists. It is more powerful than the Non-parametric test.

(ii) Non-parametric test are designed to test the statistical hypothesis only and not for estimating parameter.

(iii) All Non-parametric methods are not as simple as they are to be claimed.

4.1.2 Comparison with Parametric Tests

Q2. Distinguish between Parametric Test and Non-parametric Test.

Ans : **(Feb.-21, June-19)**

S.No.	Parametric test	S.No.	Non-parametric test
1.	The sample observations are independent.	1.	The sample observations are independent
2.	Variable in study need not be continuous (may be discrete or continuous).	2.	Variable in study is continuous.
3.	It is based on the assumption that the parent population is normal.	3.	Non-parametric test do not have assumption regarding the parent population.
4.	In this method, we can estimate the population parameter which are not known.	4.	In this method we cannot estimate the population parameter.
5.	This method cannot be applied for the data which is in ranks and grades.	5.	Only Non-parametric methods can be applied for the data of ranks and grades.
6.	Mean and variance are more general statistical measures used in this method.	6.	In this method median is usual measure for testing the hypothesis.
7.	To test two independent sample we use t-test.	7.	To test two independent samples we use run test (or) mann-whitney 'U' test.
8.	These methods are more efficient and general power.	8.	These methods are less efficient and low power.
9.	These methods are strong.	9.	These methods are weak.
10.	In this method we can draw more conclusion.	10.	In this method we cannot draw more Conclusion.

4.2 Measurement Scale - Nominal, Ordinal, Interval and Ratio

Q3. Write a short note on Measurement of Scale.

Ans : **(June-19)**

The most widely used classification of measurement scales are

(i) Nominal scale

(ii) Ordinal scale

(iii) Interval scale

(iv) Ratio scale

(i) Nominal Scale

Nominal scale is simply a system of assigning number symbols to events inorder to label them. For example, the assignment of no. of basket ball players inorder to identity then nomial scale is the least powerful level of measurement. It indicates no order (or) distance relationship and has no arithmetic origin. A nominal scale simply describes difference between them, by assigning them to categories. Thus, nominal data are thus counted data.

(ii) Ordinal Scale

The lowest level of the ordinal scale i.e., commonly used is the ordinal scale. Rank orders represents ordinal scale and are frequently used in research related to positive phenomenon a student ranks in graduation class involves the use of an ordinal scale.

(iii) Interval scale

In case of Interval scale the intervals are adjusted in terms of some rules that has been established as the basis called making the units equal The units are equal only in so for as one accept the assumptions on which the rule is based.

The Forenheit scale is an example of interval scale are provided more powerful measurements than ordinal scale.

(iv) Ratio Scale

Ratio scale represents the actual amount of variables measures of physical dimensions such as distance, height, weight etc are examples. Generally all statistical techniques are useful, will ratio test and all manipulations that can also be carried out with ratio scales.

4.3 Use of Central Limit Theorem in Testing

Q4. State central limit theorem.

Ans : **(Feb.-21)**

Central Limit Theorem

This theorem states that the distribution of the sum of i.i.d (identically and independently distribution) random variables will be normal asymptotically under general conditions with mean $\mu = \sum_{i=1}^{n} \mu_1$ and standard deviation σ where

$$\left[\sigma^2 = \sum_{i=1}^{n} \sigma_1^2\right]$$

Therefore the distributions $S_n = x_1 + x_2 + x_3 + ... + x_n$ i.e., $S_n = x_i$ where S_n is the random variable whose mean $\mu_i = E(xi)$ and variance $\sigma_1^2 = v(xi)$.

Following are some of the cases of central limit theorem (C.L.T)

1. De-Moivre's Laplace Theorem

It is the first case of central Limit theorem which was stated by Laplace. According to this theorem, the distribution of random variables with respect to the probability of success (p) is asymptotically normal as n tends to infinity. It can be written as if a random various

$$x_i = \begin{cases} 1; & \text{if probability y is p} \\ 0; & \text{ifprobability is q} \end{cases}$$

where i = 1, 2, 3 ....

Then the distribution $S_n = x_1 + x_2 + x_3 ....x_n$ is normal as $n \to \infty$

2. Lindeberg-Levy Theorem :

This theorem was proposed by lindeberg and Levy by considering two assumptions.

(i) The distribution of random variables is independent and identical.

(ii) Variance (σ^2) must be finite.

This theorem states that under the above assumptions if the random variables are distributed with $E(xi) = \mu_1$ and $v(xi) = \sigma^2$ then the sum $S_n = x_1 + x_2 + x_3 + ... + x_n$ follows normal distribution where $\mu = n\mu_1$(mean) $\sigma^2 = n\sigma_1^2$ (variance).

3. Liapounoff's central limit theorem

This is a generalized case of central limit theorem where the distribution of random variables is not identical. In this case third absolute moment (ρ^3) is considered whose distribution can be given as,

$$\rho^3 = \sum_{i=1}^{n} \rho_i^3$$

Then under general conditions, if E(xi) = μ_i and v(xi) = σ_1^2 and $\underset{n\to\infty}{Lt} = \frac{p}{\sigma} = 0$, then the sum $x = x_1 + x_2 + x_3 + ...x_n$ follows normal distribution at $N(\mu_1, \sigma^2)$ with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

4.4 One Sample Runs Test

Q5. Explain the test procedure for One Sample Runs Test.

Ans :

A sequential set of homogeneous letters which surrounds another sequential set of homogeneous letters is referred as run.

Step by Step Procedure for Calculating One Sample Run Test

Step 1

Initially define the hypothesis (H_0), Alternate hypothesis (H_1) and Level of Significance (LOS) for the given sample data.

Step 2

Calculate the statistic R (i.e., the number of runs observed from the given series of item) for given sample using the formula.

$$\boxed{R = n_1 + n_2}$$

Where,

n_1 = Represents the total number of successful terms in given sequence of data

n_2 = Represents the total number of unsuccessful terms in given sequence of data.

Step 3

Calculate the sampling distribution of R using the following formula.

$$Z = \frac{R - \mu_R}{\sigma_R}$$

Where,

$R = n_1 + n_2$

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Step 4

According to run test the given hypothesis (H_0) is rejected if |Z| > critical value.

The critical value $Z_{\frac{\alpha}{2}}$ (for two tailed test) and Z_α (for one tailed test) can be obtained from Z-table.

4.4.1 Sign Test

Q6. Derive the sign test.

(OR)

Write a short notes on sign tests.

Ans : **(Feb.-21, June-18)**

Sign test of one sample data (or) Sign test

Let $x_1, x_2...x_n$ be the 'n' independent random samples drawn from the population.

Let 'μ' be the median of the given data.

Objective

The main objective of this test is to check the median divides the data equally (or) not

Hence the Null Hypothesis and alternative Hypothesis of this test will be as follows.

Null Hypothesis (H_0)

The median divides the data equally

i.e., $H_0 : \mu = \mu_0$

Alternative Hypothesis (H_1)

The median divides the data not equally.

i.e., $H_1 : \mu \neq \mu_0$

Under the null hypothesis we calculate $x_1 - m_0, x_2 - m_0 .....x_n - m_0$ and observed the (positive) +ve and –ve signs. If there is any $(x_i - m_0)$ will be zero then omit the observation and do the remaining calculation for the reduced sample size.

Since m_0 is median

$$\therefore \quad p(x > m_0) = p(x < m_0) = \frac{1}{2}$$

Let 'u' be the number of signs which are minimum of +ve and –ve signs.

If we consider getting a sign minimum in +ve (or) –ve in a success.

Then u follows binomial distribution and it is given by

$$p(u) = n_{c_u}\left(\frac{1}{2}\right)^u\left(\frac{1}{2}\right)^{n-u}; \; u = 0, 1, 2.....n$$

$$p(u) = n_{c_u}\left(\frac{1}{2}\right)^{\cancel{u}+n-\cancel{u}}$$

$$p(u) = n_{cu}\left(\frac{1}{2}\right)^n$$

From the cumulative binomial table obtain the value of p as follows.

$$p = \sum_{r=0}^{u} n_{c_r}\left(\frac{1}{2}\right)^n = p' \text{ (say)}$$

Since p is to be calculated and it should be compared with level of significance (α)

i.e., $(p' > \alpha)$

Conclusion

If p' greater than 'α' then we accept H_0 otherwise we reject H_0

If the sample size 'n' is large ($n \geq 25$) then the statistic u follows to normal distribution then

i.e., mean = E(u) = np and

variance = v(u) = npq

$\therefore$ E(u) = np v(u) = npq

$$E(u) = n\left(\frac{1}{2}\right) \qquad v(u) = n\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

$$E(u) = \frac{n}{2} \qquad v(u) = \frac{n}{4}$$

$$S.E(u) = \sqrt{v(u)}$$

$$S.E(u) = \sqrt{\frac{n}{4}}$$

Therefore the statistic z is given by

$$z = \left|\frac{u - E(u)}{S.E(u)}\right| \sim N(0, 1)$$

$$z = \left|\frac{u - \frac{n}{2}}{\sqrt{\frac{n}{4}}}\right| \sim N(0, 1)$$

Conclusion : The calculated values of z is compared with the tabulated values of z at certain level of significance

If $z_{cal} < z_{tab}$ then we reject H_0 otherwise we accept H_0.

PROBLEMS

1. A typing school chain that in a 6 feet intensive course it can training student of type and the average at least 60 words per minute. A random sample of 15 graduates in a given typing test and the median number of words per minute by each of the student is given below. Test the hypothesis that the median typing speed of graduates is atleast 60 words per minute.

Students	1	2	3	4	5	6	7	8
Words per Minute	81	76	53	71	66	59	88	73
	9	10	11	12	13	14	15	
	80	66	58	70	60	56	55	

Sol :

Null Hypothesis (H_0)

The median no.of words per minute is 60

i.e., $H_0 : u = 60$

Alternative Hypothesis (H_1)

The median no.of words is atleast 60 per minute

i.e., $H_1 : \mu > 60$

Compare the each value with median,

write the +ve (or) –ve zero signs.

Students	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Words per minute	81	76	53	71	66	59	88	73	80	66	58	70	60	56	55
Sign	+ve	+ve	–ve	+ve	+ve	–ve	+ve	+ve	+ve	+ve	–ve	+ve	0	–ve	–ve

No.of +ve signs = 9

No.of –ve signs = 5

No.of zeros = 1

Here one of the different result is zero then omit the sample observation so the sample size is 14.

$u = \min(u^+, u^-)$

$u = \min(9, 5)$

$u = 5$

$$P = \sum_{r=0}^{u} {}^nC_r \left(\frac{1}{2}\right)^n \qquad \begin{bmatrix} n = 14 \\ u = 5 \end{bmatrix}$$

$$P = \sum_{r=0}^{5} {}^{14}C_r \left(\frac{1}{2}\right)^{14}$$

$$P = \frac{1}{(2)^{14}} \left[\sum_{r=0}^{5} {}^{14}C_r\right]$$

$$P = \frac{1}{16384} \left[{}^{14}C_0 + {}^{14}C_1 + {}^{14}C_2 + {}^{14}C_3 + {}^{14}C_4 + {}^{14}C_5\right]$$

$$P = \frac{1}{16384}[1 + 14 + 91 + 364 + 1001 + 2002]$$

$$P = \frac{3473}{16384} = 0.21$$

$P^1 = 0.21$ (say)

$P^1 >$ 5% level of significance is 0.05

$\therefore$ 0.23 > 0.05

$\therefore$ we accept H_0

Conclusion

The median no.of words per minute is 60.

2. The following data interms are the amount by large industrial plan in 40 days 17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26, 19, 23, 28, 19, 16, 22, 24, 17, 20, 13, 19, 10, 23, 18, 31, 13, 20, 17, 24, 14.

Sol :

Given the data,

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26, 19, 23, 28, 19, 16, 22, 24, 17, 20, 13, 19, 10, 23, 18, 31, 13, 20, 17, 24, 14.

Assending order

6, 9, 10, 13, 13, 14, 14, 15, 15, 16, 17, 17, 17, 17 18, 18, 19, 19, 19, 19, 20, 20, 20, 20, 22, 22, 23, 23, 24, 24, 24, 24, 24, 25, 26, 27, 28, 29, 31.

$\therefore$ Median $= \frac{19+20}{2} = \frac{39}{2} = 19.5$ [$\therefore$ sum of the middle numbers divided by 2]

Null Hypothesis (H_0)

$H_0 : u = 19.5$

Alternative Hypothesis (H_1)

$H_1 : \mu \neq 19.5$ [Two tailed test]

Compare the each value with median, write the +ve (or) –ve (or) zero signs.

Days	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Amount	6	9	10	13	13	14	14	15	15	16	17	17	17	17	18	18	19	19	19	19	20	20	20	20	22	22	23	23	23	24	24	24	24	24	25	26	27	28	29	31
Sign	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

No.of +ve sign s = 20

No.of –ve sign s = 20

No.of zero's = 0

$u = \min(u^+, u^-)$

$u = \min(20, 20)$

$u = 20$

Let $u = 20.5$

Since $n = 40 \geq 25 \sim N.D$

$$Z = \left|\frac{u - E(u)}{S.E(u)}\right| \sim N(0, 1)$$

$$Z = \left|\frac{u - n/2}{\sqrt{n/4}}\right| \sim N(0, 1)$$

$$Z = \left|\frac{20.5 - 40/2}{\sqrt{40/4}}\right| \sim N(0, 1)$$

$$Z = \left|\frac{20.5 - 20}{\sqrt{10}}\right|$$

$$Z = \left|\frac{0.5}{3.16}\right|$$

$$Z_{cal} = 0.15$$

The calculated value of z is 0.15

The tabulated value of z at 5% level of significance is 1.96

$\therefore$ $z_{tab} = 1.96$

$\therefore$ $z_{cal} < z_{tab}$

i.e., $0.15 < 1.96$

$\therefore$ we reject H_0

Conclusion

The median is not equal to 19.5

Q7. Explain Sign Test for Two Samples Data (or) Paired Samples.

Ans :

Let (x_1, y_1) (x_2, y_2).....(x_n, y_n) be the 'n' paired observations collected from the two population with probability density function is $f_1(x)$ and $f_2(y)$

Objective

The main objective of this test is to check the difference between two samples are significant or not.

Null Hypothesis (H_0)

There is no significance difference between two samples

i.e., $H_0 : f_1(x) = f_2(y)$

Alternative Hypothesis (H_1)

There is significance difference between two samples.

i.e., $H_1 : f_1(x) \neq f_2(y)$

under the Null Hypothesis hence we calculate $d_i = (x_i - y_i)$; i = 1,2.....n and observe the +ve and –ve signs.

If there is any value of d_i is zero then unit the pair of observations from the samples and we know that

$$p[x_i - y_i > 0] = p[x_i - y_i < 0] = \frac{1}{2}$$

let 'u' be the statistic as the no.of +ve signs and it follows to binomial distribution with parameters (n,p)

Then ui ~ B(n, p)

Let k = Total no.of +ve deviations,

Now we calculate

$$P(u \leq k) = \sum_{r=0}^{k} \binom{n}{r} p^r q^{n-r}$$

$$P(u \leq k) = \sum_{r=0}^{k} \binom{n}{r} \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{n-r}$$

$$P(u \leq k) = \sum_{r=0}^{k} \binom{n}{r} \left(\frac{1}{2}\right)^{r+n-r}$$

$$P(u \leq k) = \sum_{r=0}^{k} \binom{n}{r} \left(\frac{1}{2}\right)^{n} = p' \text{ (say)}$$

Conclusion

If we conclude p' is greater than level of significance (α). then we accept H_0 otherwise we reject H_0

If 'n' is large ($n \geq 25$) it follows to normal distribution

i.e., mean = E(u) = np and

variance = v(u) = npq

$\therefore$ E(u) = np v(u) = npq

$$E(u) = n\left(\frac{1}{2}\right) \quad v(u) = n\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

$$E(u) = \frac{n}{2} \quad v(u) = \frac{n}{4}$$

$$S.E(u) = \sqrt{v(u)}$$

$$S.E(u) = \sqrt{\frac{n}{4}}$$

Therefore the statistic z is given by

$$Z = \left|\frac{u - E(u)}{S.E(u)}\right| \sim N(0, 1)$$

$$Z = \left|\frac{u - n/2}{\sqrt{n/4}}\right| \sim N(0, 1)$$

The calculated value of z is compared with the tabulated value of z at certain level of significance.

If $z_{cal} < z_{tab}$ then we reject H_0 otherwise we accept H_0

PROBLEMS

3. The following data is the set of observations drawn from the two population.

Sample 1	**8**	**10**	**12**	**14**	**8**	**11**	**6**	**13**
Sample 2	**10**	**8**	**10**	**9**	**6**	**12**	**7**	**15**

Test whether the two samples have been drawn from population or not

Sol :

Null Hypothesis (H_0)

Two samples have been drawn from population.

i.e., $H_0 : f_1(x) = f_2(y)$

Alternative

Two samples have not drawn from population.

i.e., $H_1 : f_1(x) \neq f_2(y)$

x_i	y_i	$d_i = x_i - y_i$	Signs
8	10	– 2	–
10	8	2	+
12	10	2	+
14	9	5	+
8	6	2	+
11	12	– 1	–
6	7	– 1	–
13	15	– 2	–

k = no.of +ve deviation signs = 4

n = 8

Now, we calculate

$$P(u \leq k) = \sum_{r=0}^{k} \binom{n}{r} \left(\frac{1}{2}\right)^n$$

$$P(u \leq k) = \binom{8}{r}\left(\frac{1}{2}\right)^8$$

$$P(u \leq k) = \frac{1}{2^8}\left[\sum_{r=0}^{4}\binom{8}{r}\right]$$

$$P(u \leq k) = \frac{1}{256}\left[\binom{8}{0}+\binom{8}{1}+\binom{8}{2}+\binom{8}{3}+\binom{8}{4}\right]$$

$$P(u \leq k) = \frac{1}{256}[1 + 8 + 28 + 56 + 70]$$

$$P(u \leq k) = \frac{163}{256}$$

$P(u \leq k) = 0.6367 = p^1$ (say)

$P^1 > 5\%$ Level of significance

$0.6367 > 0.05$

$\therefore$ we accept H_0

Conclusion

The two samples have been drawn from population

4. To determine the percent shrinkage of a synthetic fibre, tests were made at two different temperatures. 10 tests were at lower temperature and higher temperature. The resulting data are given below:

Shrinkage	1	2	3	4	5	6	7	8	9	10
Low temperature	4.7	2.3	4.78	4.14	3.63	3.66	3.92	4.54	4.3	3.8
High temperature	3.1	4.2	4.1	4.4	4.05	3.5	3.75	4.04	4.1	5.3

Use an appropriate non parametric test to type whether the percent shrinkage of a synthetic fiber were made at two different temperature are equal OR not.

Sol :

Null Hypothesis (H_0) : The two different temperature are equal

$$H_0 : f_1(x) = f_2(y)$$

Alternative Hypothesis (H_1) : The two different temperature are not equal

$$H_1 : f_1(x) \neq f_2(y)$$

x_i	y_i	$d_i = x_i - y_1$	Signs
4.7	3.1	1.6	+
2.3	4.2	-1.9	–
4.78	4.1	0.68	+
4.14	4.4	– 0. 26	–
3.63	4.05	–0.42	–
3.66	3.05	0.16	+
3.92	3.75	0.17	+
4.54	4.04	0.5	+
4.3	4.1	0.2	+
3.8	5.3	– 1.5	–

K = no. of +ve deviation signs = 5

n = 10

Now we calculate

$$P(u \le k) = \sum_{r=0}^{k} \binom{n}{r} \left(\frac{1}{2}\right)^n$$

$$P(u \le k) = \sum_{r=0}^{5} \binom{10}{r} \left(\frac{1}{2}\right)^{10}$$

$$P(u \le 5) = \frac{1}{2^{10}} \left[\sum_{r=0}^{5} \binom{10}{r}\right]$$

$$P(u \le 5) = \frac{1}{1024}$$

$$\left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} + \binom{10}{4} + \binom{10}{5}\right]$$

$$P(u \le 5) = \frac{1}{1024}$$

[1 + 10 + 45 + 120 + 210 + 252]

$$P(u \le 5) = \frac{638}{1024}$$

$P(u \leq 5) = 0.6230 = P^1$ (say)

$P^1 > 5\%$ L.O.S

$0.6230 > 0.05$

We accept H_0

Conclusion : The two different temperature are equal.

4.4.2 Wilcoxon-Signed Rank Tests

Q8. Explain briefly about Wilcoxon Sign Ranked Test.

Ans : **(Feb.-21, June-18)**

In the sign test we consider signs only. If we consider signs and ranks also then the test is known as "Signed Ranked Test" It was first introduced by "Frank Wilcoxon". Hence it is known as "Wilcoxon signed ranked test".

Let $(x_1, y_1), (x_2, y_2)$ ..... (x_n, y_n) be the 'n' paired samples collected from the two populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

Objective :

The main objective of this test is to test the difference between two samples significant or not.

In other words, to test the two samples are coming from same distribution or not.

Null Hypothesis (H_0) :

There is no significance difference between two samples.

(or)

The two samples are drawn from same distribution.

i.e., $H_0 : f_1(x) = f_2(y)$

Alternative Hypothesis (H_1) :

There is significance difference between two samples.

(or)

The two samples are not drawn from same distribution.

i.e., $H_1 : f_1(x) \neq f_2(y)$

Level of Significance :

Consider the appropriate at α.1. Level of Significance.

Under the Null Hypothesis :

(i) We calculate $d_i = x_i - y_i$; i = 1, 2 ..... n

(ii) In any one of d_i value is zero then remove that paired observations and calculate Ranks for $|d_i|$.

(iii) Calculate sum of ranks of (positive) +ve signs and –ve signs separately.

(iv) Obtain T = Min (sum of ranks of +ve signs, sum of ranks of –ve signs)

(v) Obtain critical value of T for the corresponding value of n at specified level of significance.

$\therefore$ If T calculation value is greater than T Tabulation value then we accept H_0 otherwise we reject H_0

If 'n' is large ($n \geq 25$) it follows normal distribution then the test statistic is given by

$$z = \left|\frac{T - E\ (T)}{S.E\ (T)}\right|$$

Where, $E\ (T) = \dfrac{n(n+1)}{4}$

$$V(T) = \frac{n\left[(n+1)(2n+1)\right]}{24}$$

Since the z calculated value should be compared with the z critical value at decided level of significance.

Inference :

If z_{cal} value is greaterthan z_{tab} value then we accept H_0 otherwise we reject H_0 .

PROBLEMS

5. The following are the data of a random samples drawn from a continuous population 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2, 1.7 use wilcoxon sign ranked test to test the hypothesis H_0 : MD = 1.8 vs H_1 : MD $\neq$ 1.8.

Sol :

Given that,

Median = 1.8

d_i	1.5	2.2	0.9	1.3	2.0	1.6	1.8	1.5	2.0	1.2	1.7
d_i – Md	–0.3	0.4	–0.9	–0.5	0.2	–0.2	0	–0.3	0.2	–0.6	–0.1
Ranks	5.5	7	10	8	3	3	0	5.5	3	9	1

T^+ = Sum of the ranks of the +ve deviation

T^+ = 3 + 3 + 7

T^+ = 13

T^- = sum of the ranks of the –ve dis

T^- = 5.5 + 10 + 8 + 3 + 5.5 + 9 + 1

T^- = 42

T = min (T^+, T^-)

T = min (13, 42)

T = 13

The calculated value of T = 13

The tabulated value of T since n = 10

from wilcoxon table is 8

i.e., $T_{cal} > T_{Tab}$

$13 > 8$

$\therefore$ We Accept H_0

Conclusion :

The Median is 1.8

6. The following examination scores regarded

Pairs	1	2	3	4	5	6	7	8	9	10
with sample problem	531	621	663	579	451	660	591	719	543	575
With out sample problem	509	540	688	502	424	683	568	748	530	524

Test the Null Hypothesis at 0.05 level of significance. The sample problem increase the scores by 50 points against alternative hypothesis that the increase is less than 50 points.

Sol :

Given that,

Median = 50

Null Hypothesis (H_0) :

H_0 : Median = 50

Alternative Hypothesis (H_1) :

H_1 : median < 50

Let x_i = scores with sample problem.

y_i = scores without sample problem.

Pairs	1	2	3	4	5	6	7	8	9	10
di	22	81	–25	77	27	–23	23	–29	13	51
di – MD	–28	31	–75	27	–23	–73	–27	–79	–37	1
Ranks	5	6	9	3.5	2	8	3.5	10	7	1

T^+ = Sum of the ranks of +ve dis

$T^+ = 6 + 3.5 + 1$

$T^+ = 10.5$

T^- = Sum of the ranks of –ve dis

$T^- = 5 + 9 + 2 + 8 + 3.5 + 10 + 7$

$T^- = 44.5$

$T = \min(T^+, T^-)$

$T = \min(10.5, 44.5)$

$T = 10.5$

The calculated value of T = 10.5

The tabulated value of T since n = 10 from wilcoxon table is 10.

i.e., $T_{cal} > T_{tab}$

$10.5 > 10$

$\therefore$ we Accept H_0

Conclusion :

The median is 50.

4.5 Two Independent Sample Tests

4.5.1 Median Test

Q9. Explain the median test procedure.

Ans : **(June-18)**

Median test for two samples :

If the sizes of the two samples are different then we cannot apply sign and wilcoxon sign rank test. In such a case the median test is relevant (suitable) to test the hypothesis.

Let $x_1, x_2 ...... x_{n1}$ and $y_1, y_2 .... y_{n2}$ be the two samples of sizes n_1 and n_2 collected from the two populations with probability functions $f_1(x)$ & $f_2(y)$ respectively.

Objective:

The objective of this test is to test the difference between two samples significant (or) not.

Hence the null hypothesis of this test will be as follows

Null Hypothesis (H_0) :

There is no significance difference between two samples.

i.e., $H_0 : f_1(x) = f_2(y)$

$H_1 : f_1(x) \neq f_2(y)$

under H_0

1. First we arrange the two samples in an ascending order.
2. Calculate the Median (M) for this arrangement.
3. Find out the number of values greater than (or) equal to M in first and second samples.
4. Let it be m_1 and m_2 respectively.
5. Find out the number of values less than M in first and second samples.

   i.e., $(n_1 - m_1)$ and $(n_2 - m_2)$ respectively. This can be represented in the following 2×2 contigency table.

	Sample I	Sample II	Total
No. of observations $\geq$ m	m_1	m_2	$m_1 + m_2$
No. of observations < m	$n_1 + m_2$	$n_2 - m_2$	$n_1 + n_2 - (m_1 + m_2)$
Total	n_1	n_2	$n_1 + n_2$

The mean weekly sales of soap bars in department stores was 146.3 bars per stores. After a advertising compaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and show a standard deviation 17.2 was advertising successful (or) not ?

If either n_1 and n_2 (or) both is less than 10 then we apply hyper geometric distribution

$$P(M_1) = \frac{{}^{n_1}C_{m_1} \cdot {}^{n_2}C_{m_2}}{{}^{n_1+n_2}C_{m_1+m_2}}$$

We calculate P and it should be compare with α

If P > α we accept H_0 otherwise we reject H_0

If Both n_1 & n_2 > 10 we apply χ^2 statistic for 2 × 2 contigency table given by

$$\chi^2 = \frac{N[ad - bc]^2}{(a+c)(a+b)(b+d)(c+d)} \sim \chi^2_1 \text{ d.f}$$

Where N = a + b + c + d

It any one of the frequency is less than 5 then we apply yate's correction given by

$$\chi^2 = \frac{N\left[(ad - bc) - \frac{N}{2}\right]^2}{(a+c)(a+b)(c+d)(b+d)}$$

Since χ^2 calculation is to be calculated and it should be compare with χ^2 tabulation value for 1 degree of freedom at α.1 level of significance.

Inference : If $\chi^2_{cal} < \chi^2_{(1,1)}$

Order Statistics :

Let $x_1, x_2 \ldots\ldots x_n$ be n independent identically distributed variates, each with commutative distribution function F(x). If these variables are arranged in ascending order of magnitude and then written as $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \ldots\ldots < x_{(n)}$. we all $x_{(r)}$ as the r^{th} order statistics, r = 1, 2 .....n. The $x_{(r)}$ because of inequality relations. Among then r^{th} necessary independent. If we write $y_1 = x_{(1)}$ = the smallest of $x_1, x_2 \ldots\ldots x_n$.

$y_r = x_{(r)}$ = the smallest of $x_1, x_2 \ldots x_n$

$y_n = x_{(n)}$ = the Largest of $x_1, x_2 \ldots x_n$

Statement :

1. The cumulative distribution function of the largest ordered statistic x_n is given by $F_{(n)}(x) = [F(x)]^n$

2. The cumulative distribution function of smallest ordered statistic x_1 is given by $F_1(x) = 1-[1-F(x)]^n$

3. The cumulative distribution function of the r^{th} order statistic x_r is given by

$F_r(x) = \phi\,[x_r \leq x]$

$F_r(x) = \phi$ [Atleast r of x_i 's are $\leq x$]

$$= \sum_{j=r}^{n} \begin{bmatrix} n \\ j \end{bmatrix} F(x)\,[1-F(x)]^{n-j}$$

4.5.2 Wilcoxon – Mann Whitney U Test

Q10. Explain Wilcoxon – Mann Whitney U Test procedure.

Ans : **(Feb.-21, June-19, June-18)**

This is one of the most widely used two samples non-parametric test and it is most useful alternative to the t-test when the assumption of T-test are not satisfied.

Let $x_1, x_2 .....x_{n1}$ and $y_1, y_2, ... y_{n2}$ be the two samples of sizes n_1 & n_2 collected from two populations with probability density functions $f_1(x)$ & $f_2(y)$ respectively.

Objective : The objective of this test is to test the difference between two samples significant (or) not.

Hence the null hypothesis of this test will be as follows.

Null Hypothesis (H_0) :

There is no significance difference between two samples

i.e., $H_0 : F_1(x) = F_2(y)$

under H_0

1. First we combine the two samples and arrange them in an ascending order and identity to which samples which observation belongs.

2. Find out the ranks of ordered sequence.

3. Calculate the sum of ranks of second sequence and denote it with T.

$\therefore$ The Mann whitney U-test statistics is defined as

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T$$

To find the distribution T is very difficult but Mann & whitney have observed and obtained mean & variance for T.

If T is small or large follows normal distribution with mean & variance as

$$E(u) = \frac{n_1 n_2}{2}$$

$$V(u) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

The test statistic z is given by

$$z = \left|\frac{U - E(u)}{S.E(u)}\right| \sim N(0,1)$$

$$z = \left|\frac{u - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}\right| \sim N(0,1)$$

Since $|z|$ is to be calculated and it should be compared with z critical value at decided level of significance.

Inference :

If $|z|_{cal} < 1.96$ (or) 2.58

We reject H_0 at 5% Level of significance other wise we Accept H_0.

4.5.3 Wald Wolfowitz's Runs Test

Q11. Explain Wald - Wolfowitz Run test for one sample (or) test for Randomness Run.

Ans :

A run is defined as sequence of letters of one kind surrounded by a sequence of letters of another kind.

The no. of elements in a run is usually referred to as the "Length (L)"

Ex : AA BBA A BB AA BB AAA

Total no. of runs = 7

Runs in A series = 4

Runs in B series = 3

Let x_1 x_2 ...... x_n be the 'n' independent samples drawn from the population.

Objective :

The main objective of this test is to test the collected samples are random or not.

Hence the Null Hypothesis and alternative Hypothesis of the test will be as follows.

Null Hypothesis (H_0) :

The samples are drawn at random.

Alternative Hypothesis (H_1)

The samples are not drawn at random.

Under the Null Hypothesis

(i) First we arrange the sample observations in an ascending order and obtain the median (M)

(ii) If $x_i \leq M$ denote A and If $x_i > M$ denote B then we will get some arrangement, Let us say AA BBB A BB AA BB AAA

(iii) Obtain total no. of runs and it is denoted by 'r'.

(iv) Obtain total no. of letters in series A and B it denoted by 'r_1' and 'r_2' respectively.

If $n_1 \leq 20$, $n_2 \leq 20$ we consider it small samples and r_1, r_2 be the critical values obtained from the run tables corresponding to the sample sizes n_1 and n_2 at α% level of significance .

Conclusion :

If $r_1 < r < r_2$ then we reject H_0 otherwise we accept H_0.

If any one of sample sizes n_1 and n_2 is greaterthan 20 then the distribution is follows to normal distribution .

$\therefore$ The test statistic is given by

$$z = \left|\frac{r - E(r)}{S.E(r)}\right| \sim N(0,1)$$

where, $E(r) = \frac{n+2}{2}$

$$V(r) = \frac{n}{4}\left[\frac{n-2}{n-1}\right]$$

Conclusion:

1. If z_{cal} is lessthan z_{tab} value ($z_{cal} < z_{tab}$) then we reject H_0 otherwise we accept H_0 at 5% or 1% are level of significance.

PROBLEMS

7. In a Industrial production line items are inspected periodically for defective. The following is a sequence of defective items 'D' and Non-defective items 'N' produced by the production line

D D D N N N D N N D D N N N N D D D N N D N N N N D N D

Determine whether the defective items are occure in random or not at 5% level of significance.

Sol :

Null Hypothesis (H_0) :

The production line of defective and Non-defective items is random order.

Alternative Hypothesis (H_1) :

The production line of defective and Non-defective items is does not random order.

The combined order sample defective and Non defective items produced by the produced by the production line are

DDD NNN D NN DD NNNN DDD NN D NNNN DND

n_1 = No. of defectives in the sequence = 12

n_2 = No. of Non - defectives in the sequence = 16

$\therefore n = n_1 + n_2$

$n = 28$

Let r_1 = No. of runs with defectives = 7

r_2 = No. of runs with Non-defectives = 6

$\therefore r = r_1 + r_2$

$r = 7 + 6$

$r = 13$

$\therefore$ Total no. of runs r = 13

The lower critical value for $n_1 = 12$ and $n_2 = 16$ at 5% Level of significance is 9.

$\therefore$ The upper critical value for $n_1 = 12$ and $n_2 = 16$ at 5% level of significance is 20.

$\therefore r_1 < r < r_2$

$9 < 13 < 20$

$\therefore$ We Reject H_0

Conclusion:

The production line of defective and Non- defective items is does not random order.

8. The win and Lose of recorded of a certain cricket team for their last 50 matches was follows a WWWWWW L WWWWWWLWL WWWLL WWWWL WWW LL WWWWWWLLWWLLLWWL WWW.

Sol :

Null Hypothesis (H_0)

The symbols of win and lose of cricket team for their last 50 matches are random order.

Alternative Hypothesis (H_1)

The symbols of win and lose of cricket team for their last 50 matches are not in random order.

The combined order sample of win and lose are

WWWWWW L WWWWWW L WL WWW LL WWWW L WWW LL WWWWWW LL WW LLL WWL WWW

n_1 = No. of wins in 50 matches = 36

n_2 = No. of lose in 50 matches = 14

$\therefore n = n_1 + n_2$

$n = 36 + 14$

$n = 50$

Let r_1 = No. of runs of win = 10

r_2 = No. of runs of lose = 9

$\therefore r = r_1 + r_2$

$r = 10 + 9$

$r = 19$

$\therefore$ The test statistic is given by

$$z = \left|\frac{r.E(r)}{S.E(r)}\right| \sim N(0,1)$$

Where, $E(r) = \frac{n+2}{2}$

$$E(r) = \frac{50+2}{2}$$

$$E(r) = \frac{52}{2}$$

$$E(r) = 26$$

$$V(r) = \frac{n}{4}\left[\frac{n-2}{n-1}\right]$$

$$V(r) = \frac{50}{4}\left[\frac{50-2}{50-1}\right]$$

$$= \frac{50}{4}\left[\frac{48}{49}\right]$$

$$= 12.5\ [0.97]$$

$$V(r) = 12.12.$$

$$S.E(r) = \sqrt{12.12}$$

$$S.E(r) = 3.48$$

Then $z = \left|\frac{19-26}{3.48}\right| \sim N(0,1)$

$$z = \left|\frac{-7}{3.48}\right|$$

$$z = \left|\frac{7}{3.48}\right|$$

$$z_{cal} = 2.011$$

z_{tab} at 5% Level of significance is 1.645.

$\therefore z_{cal} > z_{tab}$

i.e., $2.011 > 1.645$

$\therefore$ We accept H_0

Conclusion :

The symbols of win and lose of cricket team for their last 50 matches are in random order.

Q12. Explain the procedure for Wald-wolfowitz Run test for two samples.

Ans :

Let x_1, x_2 ..... x_n and y_1 , y_2 ...... y_n be the two samples collected from two different population with density functions $f_1(x)$ and $f_2(y)$ respectively

Objective :

The main objective of this test is to test the two samples are drawn from the population at random or not.

In other words, to test the two samples are coming from the same population or not .

Hence the Null Hypothesis and alternative Hypothesis of this test will be as follows.

Null Hypothesis

The two samples are drawn from the two populations at random.

Alternative Hypothesis

The two samples are drawn from the two populations at not random.

Under the Null Hypothesis

(i) First we combine the two samples and arrange the sample observation in the order of their Magnitude. Now total no. of elements are $n = n_1 + n_2$. In these arrangement denote "A" is the elements belongs to first sample and denote "B" if the elements belongs to second sample.

(ii) Obtain the total no. of runs and it is denoted by "r"

(iii) Obtain the total no. of letters in "A" and it is denoted by "r_1".

(iv) Obtain the total no. of letters in "B" and it is denoted by "r_2".

If $n_1 \leq 20$, $n_2 \leq 20$ it consider the small samples and r_1 and r_2 be the critical values obtained from run tables corresponding to n_1 and n_2 at α% level of significance.

If $r_1 < r < r_2$ then we reject H_0 otherwise we accept H_0.

If any one of n_1 and n_2 is greater than 20 it follows to normal distribution then the test statistic is given by

$$z = \left| \frac{r - E(r)}{S.E(r)} \right| \sim N(0,1)$$

Where $$E(r) = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$V(r) = \frac{2n_1n_2\left[2n_1n_2 - n_1 - n_2\right]}{\left(n_1 + n_2\right)^2\left(n_1 + n_2 - 1\right)}$$

Conclusion :

If the z calculation value is lessthan the z tabulated value then we reject H_0 other wise we accept H_0 at 5% (or) 1% level of significance.

PROBLEMS

9. Two quality controlled laboratories independently collected the samples of 25 articles from A no. of sales deposits and tested then the no. of defectives per sales for deposits were as follows.

Lab A :	**9**	**3**	**1**	**3**	**0**	**7**	**2**	**11**
Lab B :	**12**	**6**	**6**	**4**	**8**	**5**	**4**	

Test the hypothesis is that two Laboratories have samples from same Lot (population) by wald-wolfowitz test.

Sol :

Null Hypothesis

These two laboratories samples have been drawn from the same population.

i.e., $f_1(x) = f_2(y)$

Alternative Hypothesis :

These two Laboratories samples have not drawn from the same population

i.e., $f_1(x) \neq f_2(y)$

Given that,

$n_1 = 8$, $n_2 = 7$

Arranging the given samples into one combined order samples.

Let "Lab A " symbol is "A"

"Lab B" symbol is "B"

Ascending order :

0 1 2 3 3 4 4 5 6 6 7

A A A A A B B B B B A

8 9 11 12

B A A B

The combined order sample is AAAAABBBBBABAAB

Given that, $n_1 = 8$, $n_2 = 7$

$\therefore n = n_1 + n_2$

$n = 8 + 7$

$n = 15$

Let r_1 = No. of runs of A series = 3

r_2 = No. of runs of B series = 3

$\therefore r = r_1 + r_2$

$r = 3 + 3$

$r = 6$ = Total no. of runs

The lower critical value for $n_1 = 8, n_2 = 7$ at 5% level of significance is 4.

The upper critical value for $n_1 = 8, n_2 = 7$. at 5% level of significance is 13

$r_1 < r < r_2$

$4 < 6 < 13$

$\therefore$ We Reject H_0

Conclusion :

These two laboratories samples have not drawn from same population.

10. The following 20 members are taken from 2 digited tables 51, 68, 30, 81, 90, 46, 99, 98, 11, 6, 19, 43, 95, 82, 65, 85, 65, 81, 0, 50. Test the random numbers by the run test on the basis of up and down runs.

Sol :

Null Hypothesis: The above numbers are in random order.

Alternative Hypothesis :

The above numbers are not in random order.

The ascending order of above numbers is 0, 6, 11, 19, 30, 43, 46, 50, 51, 65, 65, 68, 81, 81, 82, 85, 90, 95, 98, 99.

n = 20 (even)

Median = avg of $\left(\frac{n}{2}, \frac{n}{2}+1\right)$ terms

Median = avg of (10, 11) terms

Median = average of (65, 65)

Median = 65

Use symbal 'A' for an observation less than Median, use symbol 'B' for an observation greaterthan Median. To get a combined order samples the sequence of symbol is A B A BB A BB AAAA BB O B O B AA AB A BB A BB AAAA BBBB AA

n_1 = no. of A symbols = 9

n_2 = no. of B symbols = 9

$\therefore n = n_1 + n_2$

$n = 9 + 9$

$n = 18$

r_1 = No. of runs of A symbols = 5

r_2 = No. of runs of B symbols = 4

$\therefore$ r = $r_1 + r_2$

r = 5 + 4

r = 9 = Total no. of runs

The lower critical value for n_1 = 9, n_2 = 9 at 5% level of significance is 5.

The upper critical value for n_1 = 9, n_2 = 9 at 5% level of significance is 15.

$r_1 < r < r_2$

$5 < 9 < 15$

$\therefore$ We Reject H_0

Conclusion :

The bove numbers are not in random order.

11. Given the following ten - paired sample observations.

Sample	**42.7**	**50.4**	**44.2**	**49.7**	**49.4**	**55.3**	**46.1**	**49.8**	**48.7**	**45.9**
Sample	**54.2**	**43.7**	**34.0**	**46.1**	**55.4**	**58.3**	**53.4**	**46.3**	**57.0**	**61.9**

Test the hypothesis that the two samples have come from identical populations by,

(i) Sign test

(ii) Wilcoxon signed rank test.

Sol :

(i) Sign Test

Let,

The null hypothesis be H_0 : Two samples belongs to same population

Alternative hypothesis be H_1 : Two samples belongs to different population

Sample 1 (X_i)	**Sample 2 (Y_i)**	**Difference (d) ($d = X_i - Y_i$)**	**Sign**
42.7	54.2	-11.5	–
50.4	43.7	6.7	+
44.2	34.0	+10.2	+
49.7	46.1	+ 3.6	+
49.4	55.4	– 6	–
55.3	58.3	– 3	–
46.1	53.4	– 7.3	–
49.8	46.3	+ 3.5	+
48.7	57.0	– 8.3	–
45.9	61.9	– 16	–

If can be inferred from above table that,

Number of sample observation with positive (+) sign = 4

Number of sample observation with negative (–) sign = 6

$\because$ Number of +ve terms < Number of – ve terms = 6

X = Total number of – ve terms = 6

Now,

Check whether X < np or X > np

We know that,

$X = 6$

n = 10 (number of sample observation)

$P = \frac{1}{2}$

$\therefore\ 6 > 10.\frac{1}{2}$

$6 > 5$

$\because$ X > np the following test statistics is considered.

$$Z = \frac{(6-0.5)-10.\frac{1}{2}}{\sqrt{10.\frac{1}{2}.\frac{1}{2}}} \qquad \left(\because q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}\right)$$

$$= \frac{5.5-2}{\sqrt{\frac{5}{2}}} = \frac{0.5}{\sqrt{2.5}}$$

$$= \frac{0.5}{1.58} = 0.316$$

The calculated value of Z i.e., $Z_{cal} = 0.316$

Decision

The tabulated value of Z at 5% level of significance is $Z_{tab} = 1.96$ (Refer Z table)

$\because Z_{cal} < Z_{tab}$ ($\because 0.316 < 1.96$)

The null hypothesis H_0 is accepted.

Hence, the two samples belongs to sample population

(ii) Wilcoxon Signed Rank Test

Null hypothesis is H_0 : Two samples belongs to same population.

Alternate hypothesis H_1 : Two samples belongs to different population.

Sample A (X_i)	Sample B (Y_i)	$X_i - Y_i$ (Sample A - Sample B)	$\lvert X_i - Y_i \rvert$	Rank
42.7	54.2	– 11.5	$\lvert -11.5 \rvert = 11.5$	2
50.4	43.7	6.7	$\lvert 6.7 \rvert = 6.7$	6
44.2	34	10.2	$\lvert 10.2 \rvert = 10.2$	3
49.7	46.1	3.6	$\lvert 3.6 \rvert = 3.6$	9
49.4	55.4	– 6	$\lvert -6 \rvert = 6$	8
55.3	58.3	– 3	$\lvert -3 \rvert = 3$	10
46.1	53.4	– 7.3	$\lvert 7.3 \rvert = 7.3$	5
49.8	46.3	+ 3.5	$\lvert 3.5 \rvert = 3.5$	7
48.7	57.0	–8.3	$\lvert -8.3 \rvert = 8.3$	4
45.9	61.9	– 16	$\lvert -16 \rvert = 16$	1

Ranks in the above table are assigned based on the difference. (Ignoring the sing).

Now,

Calculate the T(Sum of + ve rank) and T^- (sum of – ve ranks)

$$\therefore\ T^+ = \sum_{i=1}^{n} \text{sum of + ve ranks}$$

$$= 6 + 3 + 9 + 7 = 25$$

Since T^+ is smaller than T^-

$$T = T^+ = 25$$

Test Statistic

The test statistic is calculated using the formula,

$$|Z| = \frac{\left|T - \frac{n(n+1)}{4}\right|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{\left|25 - \frac{10(10+1)}{4}\right|}{\sqrt{\frac{10(10+1)(2(10)+1)}{24}}}$$

$$= \frac{\left|25 - 5\left(\frac{11}{2}\right)\right|}{\sqrt{\frac{10 \times 11 \times 21}{24}}} = \frac{\left|25 - \frac{55}{2}\right|}{\sqrt{\frac{2310}{24}}}$$

$$= \frac{25 - 27.5}{\sqrt{96.25}}$$

$$= \frac{-2.5}{9.81} = -0.254$$

$$|Z| = 0.254$$

The calculated value of $|Z|$ is $Z_{cal} = 0.254$

Decision

The tabulated value of Z at 5% level of significance is $Z_{tab} = 1.96$

Since,

$$Z_{cal} > Z_{tab} \ (0.254 < 1.96)$$

The null hypothesis H_0 is accepted.

$\therefore$ The samples belongs to different population.

12. The following are the weights in pounds, before and after a dieting programme of 16 person for one month,

Before	After
147.0	137.9
183.5	176.2
232.1	219.0
161.6	163.8
197.5	193.5
206.3	201.4
177.0	180.6
215.4	203.2
147.7	149.0
208.1	195.4
166.8	158.5
131.9	134.4
150.3	149.3
197.2	189.1
159.8	159.1
171.7	173.2

Use the signed rank test to test the dieting programme is effective or not 5% level of significance.

Sol :

Let,

The null hypothesis be $H_0 : \mu_1 = \mu_2$

The alternate hypothesis be $H_1 : \mu_1 > \mu_2$

Now, calculate difference (d) and Rank (R) for the given data.

Before	**After**	**Difference (d)**	**Rank (Assigning rank in ascending order)**
147.0	137.9	9.1	13
183.5	176.2	7.3	10
232.1	219.0	13.1	16
161.6	163.8	–2.2	5
197.5	193.5	4.0	8
206.3	201.4	4.9	9
177.0	180.6	– 3.6	7
215.4	203.2	12.2	14
147.7	149.0	– 1.3	3
208.1	195.4	12.7	15
166.8	158.5	8.3	12
131.9	134.4	– 2.5	6
150.3	149.3	1.0	2
197.2	189.1	8.1	11
159.8	159.1	0.7	1
171.7	173.2	–1.5	4

Ranks in the above table are assigned based on the difference. That is the weight with smallest difference is assigned rank 1 and so on.

The Wilcoxon signed rank test is calculated with the help of following formula.

$$Z = \frac{\left| T - \frac{n(n+1)}{4} \right|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad \text{...... (1)}$$

Now, calculate the rank sum of positive term (T^+) and rank sum of negative sum (T^-)

T = Sum of all positive terms

= 13 + 10 + 16 + 8 + 9 + 14 + 15 + 12 + 2 + 11 + 1

= 111

T = Sum of all negative ranks

= 5 + 7 + 3 + 6 + 4

= 25

Since the wilcoxon test is for paired sample we consider T as T^+

$\therefore$ T = 111

n = 16

Substituting the above value in equation (1) we get,

$$Z = \frac{\left|T^{+} - \frac{n(n+1)}{4}\right|}{\frac{n(n+1)(2n+1)}{4}}$$

$$= \frac{\left|111 = \frac{16(16+1)}{4}\right|}{\sqrt{\frac{16(16+1)(2(16)+1)}{24}}}$$

$$= \frac{\left|111 - \frac{16(17)}{4}\right|}{\sqrt{\frac{16 \times 17 \times 33}{24}}}$$

$$= \frac{|111 - 68|}{\sqrt{\frac{8976}{24}}}$$

$$= \frac{43}{\sqrt{374}}$$

$$= \frac{43}{19.33} = 2.22$$

$\therefore$ |Z| or the calculated value of Z = 2.22.

Decision

The calculated value of Z_{cal} i.e., |Z| = 2.22

The calculated value of Z_{tab} at 5% level of significance is $Z_{0.05}$ = 1.645 (Z table).

$\because Z_{cal} > Z_{tab}$

$\because H_0$ is rejected

Short Question and Answers

1. State central limit theorem.

Ans :

Central Limit Theorem

This theorem states that the distribution of the sum of i.i.d (identically and independently distribution) random variables will be normal asymptotically under general conditions with mean $\mu = \sum_{i=1}^{n} \mu_1$ and standard deviation σ where $\left[\sigma^2 = \sum_{i=1}^{n} \sigma_1^2\right]$

Therefore the distributions $S_n = x_1 + x_2 + x_3 + \ldots\ldots + x_n$ i.e., $S_n = x_i$ where S_n is the random variable whose mean $\mu_i = E(xi)$ and variance $\sigma_1^2 = v(xi)$.

Following are some of the cases of central limit theorem (C.L.T)

1. De-Moivre's Laplace Theorem

It is the first case of central Limit theorem which was stated by Laplace. According to this theorem, the distribution of random variables with respect to the probability of success (p) is asymptotically normal as n tends to infinity. It can be written as if a random various

$$x_i = \begin{cases} 1; & \text{if probability y is p} \\ 0; & \text{ifprobability is q} \end{cases}$$

where i = 1, 2, 3 ....

Then the distribution $S_n = x_1 + x_2 + x_3 \ldots\ldots x_n$ is normal as $n \to \infty$

2. Lindeberg-Levy Theorem :

This theorem was proposed by lindeberg and Levy by considering two assumptions.

(i) The distribution of random variables is independent and identical.

(ii) Variance (σ^2) must be finite.

This theorem states that under the above assumptions if the random variables are distributed with $E(xi) = \mu_1$ and $v(xi) = \sigma^2$ then the sum $S_n = x_1 + x_2 + x_3 + \ldots\ldots + x_n$ follows normal distribution where $\mu = n\mu_1$(mean) $\sigma^2 = n\sigma_1^2$ (variance).

3. Liapounoff's central limit theorem

This is a generalized case of central limit theorem where the distribution of random variables is not identical. In this case third absolute moment (ρ^3) is considered whose distribution can be given as,

$$\rho^3 = \sum_{i=1}^{n} \rho_i^3$$

Then under general conditions, if $E(xi) = \mu_i$ and $v(xi) = \sigma_1^2$ and $\underset{n\to\infty}{Lt} = \frac{p}{\sigma} = 0$, then the sum $x = x_1 + x_2 + x_3 + \ldots\ldots x_n$ follows normal distribution at $N(\mu_1, \sigma^2)$ with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

---

2. Write a short notes on sign tests.

Ans :

Sign test of one sample data (or) Sign test

Let $x_1, x_2 \ldots x_n$ be the 'n' independent random samples drawn from the population.

Let 'μ' be the median of the given data.

Objective

The main objective of this test is to check the median divides the data equally (or) not

Hence the Null Hypothesis and alternative Hypothesis of this test will be as follows.

Null Hypothesis (H_0)

The median divides the data equally

i.e., $H_0 : \mu = \mu_0$

Alternative Hypothesis (H_1)

The median divides the data not equally.

i.e., $H_1 : \mu \neq \mu_0$

Under the null hypothesis we calculate $x_1 - m_0$, $x_2 - m_0$ .....$x_n - m_0$ and observed the (positive) +ve and –ve signs. If there is any $(x_i - m_0)$ will be zero then omit the observation and do the remaining calculation for the reduced sample size. Since m_0 is median

$$\therefore \quad p(x > m_0) = p(x < m_0) = \frac{1}{2}$$

Let 'u' be the number of signs which are minimum of +ve and –ve signs.

If we consider getting a sign minimum in +ve (or) –ve in a success.

Then u follows binomial distribution and it is given by

$$p(u) = n_{c_u}\left(\frac{1}{2}\right)^{u}\left(\frac{1}{2}\right)^{n-u};$$

$$u = 0, 1, 2.....n$$

$$p(u) = n_{c_u}\left(\frac{1}{2}\right)^{\cancel{u}+n-\cancel{u}}$$

$$p(u) = n_{cu}\left(\frac{1}{2}\right)^{n}$$

From the cumulative binomial table obtain the value of p as follows.

$$p = \sum_{r=0}^{u} n_{c_r}\left(\frac{1}{2}\right)^{n} = p' \text{ (say)}$$

Since p is to be calculated and it should be compared with level of significance (α)

i.e., $(p' > \alpha)$

3. Discuss the advantages and disadvatages of Non-parametric tests.

Ans :

(i) Non-parametric methods are very simple and easy to apply.

(ii) Non-parametric methods do not required much complicated sampling theory.

(iii) No assumption is made about the form of frequency function of the parent population from which the samples have drawn.

(vi) Generally the socio-economic data are normally distributed, non-parametric methods have found applications in sociology.

(v) If the data is given in ranks (or) signs (or) grade (A, A^+, A^-, B, B^+, B^-....) then the Non-parametric methods are more useful and easy to calculate but parametric methods cannot be apply.

(vi) No parametric technique will apply to the data which are measured in nominal scale while Non-parametric methods exist to deal with such data.

(vii) It the Non-parametric methods can be used even if the sample size 6 (or) more.

(viii) Non-parametric test may be quiet powerful even if the sample size is small.

Disadvantages of Non-parametric test

(i) Non-parametric test can we used only the measurement are nominal (or) ordinal even in that case if a parametric test exists. It is more powerful than the Non-parametric test.

(ii) Non-parametric test are designed to test the statistical hypothesis only and not for estimating parameter.

(iii) All Non-parametric methods are not as simple as they are to be claimed.

4. Write a short note on Measurement of Scale.

Ans :

(i) Nominal Scale

Nominal scale is simply a system of assigning number symbols to events inorder to label them. For example, the assignment of no. of basket ball players inorder to identity then nomial scale is the least powerful level of measurement. It indicates no order (or) distance relationship and has no arithmetic origin. A nominal scale simply describes difference between them, by assigning them to categories. Thus, nominal data are thus counted data.

(ii) Ordinal Scale

The lowest level of the ordinal scale i.e., commonly used is the ordinal scale. Rank orders represents ordinal scale and are frequently used in research related to positive phenomenon a student ranks in graduation class involves the use of an ordinal scale.

(iii) Interval scale

In case of Interval scale the intervals are adjusted in terms of some rules that has been established as the basis called making the units equal The units are equal only in so for as one accept the assumptions on which the rule is based.

The Forenheit scale is an example of interval scale are provided more powerful measurements than ordinal scale.

(iv) Ratio Scale

Ratio scale represents the actual amount of variables measures of physical dimensions such as distance, height, weight etc are examples. Generally all statistical techniques are useful, will ratio test and all manipulations that can also be carried out with ratio scales.

5. Define Run and Length of Run.

Ans :

A sequential set of homoteneous letters which surrounds another sequential set of homogeneous letters is referred as run.

Length of a run is defined as the total number of values in a run.

6. What is meant by Non-parametric Tests.

Ans :

A statistical test which does not depend on the particular form of frequency function from which the samples are drawn are not related with estimating the parameters is called Non-parametric test.

Assumption

1. Sample observations are independent.
2. The variable under study us continuous.
3. The probability density function is continuous.
4. Lower order moments exists.

7. Explain briefly about Wilcoxon Sign Ranked Test.

Ans :

In the sign test we consider signs only. If we consider signs and ranks also then the test is known as “Signed Ranked Test” It was first introduced by “Frank Wilcoxon”. Hence it is known as “Wilcoxon signed ranked test”.

Let $(x_1, y_1), (x_2, y_2) ..... (x_n, y_n)$ be the ‘n’ paired samples collected from the two populations with probability density functions $f_1(x)$ and $f_2(y)$ respectively.

Objective :

The main objective of this test is to test the difference between two samples significant or not.

In other words, to test the two samples are coming from same distribution or not.

Null Hypothesis (H_0) :

There is no significance difference between two samples.

(or)

The two samples are drawn from same distribution.

i.e., $H_0 : f_1(x) = f_2(y)$

Alternative Hypothesis (H_1) :

There is significance difference between two samples.

(or)

The two samples are not drawn from same distribution.

i.e., $H_1 : f_1(x) \neq f_2(y)$

Level of Significance :

Consider the appropriate at α.1. Level of Significance.

Under the Null Hypothesis :

(i) We calculate $d_i = x_i - y_i$; $i = 1, 2 ..... n$

(ii) In any one of d_i value is zero then remove that paired observations and calculate Ranks for $|d_i|$.

(iii) Calculate sum of ranks of (positive) +ve signs and –ve signs separately.

(iv) Obtain T = Min (sum of ranks of +ve signs, sum of ranks of –ve signs)

(v) Obtain critical value of T for the corresponding value of n at specified level of significance.

$\therefore$ If T calculation value is greater than T Tabulation value then we accept H_0 otherwise we reject H_0

If 'n' is large ($n \geq 25$) it follows normal distribution then the test statistic is given by

$$z = \left|\frac{T - E\ (T)}{S.E\ (T)}\right|$$

$$\text{Where, } E\ (T) = \frac{n(n+1)}{4}$$

$$V(T) = \frac{n\left[(n+1)(2n+1)\right]}{24}$$

Since the z calculated value should be compared with the z critical value at decided level of significance.

Inference :

If z_{cal} value is greaterthan z_{tab} value then we accept H_0 otherwise we reject H_0 .

8. Explain the median test procedure.

Ans :

Median test for two samples :

If the sizes of the two samples are different then we cannot apply sign and wilcoxon sign rank test. In such a case the median test is relevant (suitable) to test the hypothesis.

Let $x_1, x_2 ...... x_{n1}$ and $y_1, y_2 .... y_{n2}$ be the two samples of sizes n_1 and n_2 collected from the two populations with probability functions $f_1(x)$ & f_2 (y) respectively.

Objective:

The objective of this test is to test the difference between two samples significant (or) not.

Hence the null hypothesis of this test will be as follows

Null Hypothesis (H_0) :

There is no significance difference between two samples.

i.e., $H_0 : f_1(x) = f_2(y)$

$H_1 : f_1(x) \neq f_2(y)$

under H_0

1. First we arrange the two samples in an ascending order.
2. Calculate the Median (M) for this arrangement.
3. Find out the number of values greater than (or) equal to M in first and second samples.
4. Let it be m_1 and m_2 respectively.
5. Find out the number of values less than M in first and second samples.

i.e., $(n_1 - m_1)$ and $(n_2 - m_2)$ respectively. This can be represented in the following 2 × 2 contigency table.

---

9. Explain the procedure for Wald-wolfowitz Run test for two samples.

Ans :

Let x_1, x_2 ..... x_n and y_1 , y_2 ...... y_n be the two samples collected from two different population with density functions $f_1(x)$ and $f_2(y)$ respectively

Objective :

The main objective of this test is to test the two samples are drawn from the population at random or not.

In other words, to test the two samples are coming from the same population or not .

Hence the Null Hypothesis and alternative Hypothesis of this test will be as follows.

Null Hypothesis

The two samples are drawn from the two populations at random.

Alternative Hypothesis

The two samples are drawn from the two populations at not random.

Under the Null Hypothesis

(i) First we combine the two samples and arrange the sample observation in the order of their Magnitude. Now total no. of elements are $n = n_1 + n_2$. In these arrangement denote “A” is the elements belongs to first sample and denote “B” if the elements belongs to second sample.

(ii) Obtain the total no. of runs and it is denoted by “r”

(iii) Obtain the total no. of letters in “A” and it is denoted by “r_1”.

(iv) Obtain the total no. of letters in “B” and it is denoted by “r_2”.

If $n_1 \leq 20$, $n_2 \leq 20$ it consider the small samples and r_1 and r_2 be the critical values obtained from run tables corresponding to n_1 and n_2 at α% level of significance.

If $r_1 < r < r_2$ then we reject H_0 otherwise we accept H_0.

If any one of n_1 and n_2 is greater than 20 it follows to normal distribution then the test statistic is given by

$$z = \left|\frac{r - E(r)}{S.E(r)}\right| \sim N(0,1)$$

Where $E(r) = \frac{2n_1n_2}{n_1 + n_2} + 1$

$$V(r) = \frac{2n_1n_2\left[2n_1n_2 - n_1 - n_2\right]}{\left(n_1 + n_2\right)^2\left(n_1 + n_2 - 1\right)}$$

10. Distinguish between Parametric Test and Non-parametric Test.

Ans :

S.No.	Parametric test	S.No.	Non-parametric test
1.	The sample observations are independent.	1.	The sample observations are independent
2.	Variable in study need not be continuous (may be discrete or continuous).	2.	Variable in study is continuous.
3.	It is based on the assumption that the parent population is normal.	3.	Non-parametric test do not have assumption regarding the parent population.
4.	In this method, we can estimate the population parameter which are not known.	4.	In this method we cannot estimate the population parameter.
5.	This method cannot be applied for the data which is in ranks and grades.	5.	Only Non-parametric methods can be applied for the data of ranks and grades.
6.	Mean and variance are more general statistical measures used in this method.	6.	In this method median is usual measure for testing the hypothesis.
7.	To test two independent sample we use t-test.	7.	To test two independent samples we use run test (or) mann-whitney 'U' test.
8.	These methods are more efficient and general power.	8.	These methods are less efficient and low power.
9.	These methods are strong.	9.	These methods are weak.
10.	In this method we can draw more conclusion.	10.	In this method we cannot draw more Conclusion.

Exercise Problems

1. The following data represents the number of hours that a portable car vacuum cleaner operates before recharging is required.

Operating	1.4	2.3	0.8	1.4	1.8	1.5
Time (h)	1.9	1.4	2.1	1.1	1.6	

Use the Wilcoxon signed-rank test to test the hypothesis, at a 5% level of significance, that this

[Ans : 22.5]

2. After a two - week study program to acquaint company personnel with certain tax regulations, the manager of the international business division of a large firm constructed a 20-questions, true/false examination to test the effectiveness of the study program. The examination was constructed with correct answers running in the following sequence :

T FF T F T F TT F T FF T F T F TT F

Does this sequence indicate a departure from randomness in the arrangement of T and F answers? Perform a runs test at $\alpha = 0.05$

[Ans : 2.30]

3. The data of 10 plots each, under two treatments are as given below,

(Treat.1)X :	46	45	32	42	39	48	49	30	51	34
(Treat.2)Y :	44	40	59	47	55	50	47	71	43	55

The hypothesis of equality of median response under two treatments can be tested by the median test.

[Ans : 0.078]

4. An engineer is investigating two different types of metering devices. A and B for an electronic fuel injection system to determine if they differ in their fuel mileage performance. The system is installed on 12 different cars, and a test is run with each metering system in turn on each car. The observed fuel mileage data (in miles/ gallon) is shown below :

A	18.7	20.3	20.8	18.3	16.4	16.8
B	17.6	21.2	19.1	17.5	16.9	16.4

A	17.2	19.1	17.9	19.8	18.2	19.1
B	17.7	19.2	17.5	21.4	17.6	18.8

Use the sign test at a level of significance of 5% to determine whether there is any difference between the two systems.

[Ans : 4]

Choose the Correct Answers

1. ________ scale which specify rank order of a particular event. [b]
   (a) Nominal (b) Ordinal
   (c) Ratio (d) Internal
2. ________ scale that gives actual amount of variables in terms of some specific units [c]
   (a) Nominal (b) Ordinal
   (c) Ratio (d) Internal
3. ________ scale which is constructed by adjusting equal interval of units based on some standard rules [d]
   (a) Nominal (b) Ordinal
   (c) Ratio (d) Internal
4. In wald-walfowitz run test the mean value is ________ [b]
   (a) $\frac{m+n}{2}$ (b) $\frac{2mn}{m+n}+1$
   (c) $\frac{m+n}{2mn}+1$ (d) None
5. No ________ scale which specifies order or distance relationship of an object [a]
   (a) Nominal (b) Ordinal
   (c) Ratio (d) Internal
6. ________ deals with types of data which are measured using nominal scale [b]
   (a) Parametric methods (b) Non-parametric method
   (c) Nominal (d) Ordinal
7. ________ test is depending on their signs [c]
   (a) Run (b) Parametric
   (c) Sign (d) Signed rank
8. Non-parametric tests does not ________ on f(x) [c]
   (a) Independent (b) Dependent
   (c) Depend (d) None
9. If the measurement scale of data is nominal or ordinal ________ methods can be used [b]
   (a) Parametric (b) Non-parametric
   (c) Measurement of scale (d) Run
10. Run is a sequence of ________ [a]
   (a) Letters (b) Words
   (c) Paragraphs (d) None

Fill in the blanks

1. In non-parametric method __________ function of sample is unknown.
2. In non-parametric method the order of moments must be __________.
3. Non-parametric methods are applicable only for __________ measurements.
4. Nominal scale specifies order (or) distance relationship of an __________.
5. Median test for large samples is __________.
6. Variance value is median test is __________.
7. Wald-Wolfowitz run test for large samples is __________.
8. Mann-Whitney U-test for large samples is __________.
9. A __________ is defined as a sequence of letters of one kind surrounded by sequence of letters of the other kind.
10. The number of elements in a run is known as __________.

ANSWERS

1. Frequency
2. Low
3. Nominal
4. Object
5. $\dfrac{M - E(M)}{\sqrt{v(M)}}$
6. $\dfrac{n_1 n_2 (N+1)}{4N^2}$
7. $\dfrac{R - \left(\dfrac{2mn}{m+n} + 1\right)}{(m+n)^2 (m+n-1)}$
8. $\dfrac{U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$
9. Run
10. Length of run

Very Short Questions and Answers

Q1. State any two differences of parametric and non-parametric tests.

Ans :

Parametric

1. Information about population is completely known.
2. No parametric test exist for nominal scale data

Non - Parametric

1. No information about the population is available
2. It is applied both variable and attributes.

Q2. Define ordinal scale.

Ans :

Ordinal scale is a lowest level of measurement which specify rank order of a particular event.

Q3. Define run.

Ans :

A run is defined as a sequence of letters of one kind, surrounded by sequence of letters of the other kind.

Q4. Define length of a run.

Ans :

The number of elements in a run is known as length as length of a run.

Q5. Define Ratio Scale.

Ans :

It gives actual amount of variables in terms of some specific units.

FACULTIES OF SCIENCE

B.Sc. (CBCS) (IV – Semester) Examination

January / February - 2021

STATISTICS

Time : 2 Hours] **[Max. Marks : 80**

PART - A (4 × 5 = 20 Marks)

Note: Answer any Four questions.

		Answers
1.	Define null - hypothesis and alternate hypothesis, critical region and power of test. Give one example for each.	**(Unit-I, SQA-1, 2, 3, 4)**
2.	Define randomized and non randomized test functions. Give two examples for each.	**(Unit-I, SQA-5)**
3.	Describe large sample test for single proposition.	**(Unit-II, SQA-5)**
4.	Explain Fisher's z-transformations for two samples and associated test procedure.	**(Unit-II, SQA-7)**
5.	Explain t-test for single mean.	**(Unit-III, SQA-1)**
6.	Explain χ^2 - test for 2 × 2 contingency table for independence of attributes.	**(Unit-III, SQA-2)**
7.	Explain use of central limit theorem in testing. Give two examples.	**(Unit-IV, SQA-1)**
8.	Describe one sample sign test.	**(Unit-IV, SQA-2)**

PART - B (3 × 20 = 60 Marks)

Note: Answer any three questions.

9.	Let x_1, x_2 .... x_n be a random sample from poisson population with parameter λ. Obtain the best critical region for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda > \lambda_0$ at α% level of significance.	**(Unit-I, Prob. 11)**
10.	State and prove Neyman - Pearson lemma.	**(Unit-I, Q.No. 8)**
11.	Explain the large sample test procedure for difference of standard deviations.	**(Unit-II, Q.No. 7)**
12.	Explain the large sample test procedure for difference of means.	**(Unit-II, Q.No. 4)**
13.	Describe the test procedures based on Snedecor's F-distribution for homogeneity of population variances and z^2 – test procedure for population variance.	**(Unit-III, Q.No. 7, 10)**
14.	Explain the test procedures for sample correlation coefficient based on students t-distribution and paired t-test	**(Unit-III, Q.No. 2, 5)**
15.	Explain two sample signed rank test for small samples along with its large sample approximation Compare the same with sign test.	**(Unit-IV, Q.No. 7)**
16.	(i) Describe Mann-Whitney u-test for small samples. Also give its large sample approximation.	**(Unit-IV, Q.No. 11)**
	(ii) Compare parametric and non-parametric test.	**(Unit-IV, Q.No. 2)**

FACULTY OF SCIENCE

B.Sc. IV-Semester (CBCS) Examination

May / June - 2019

Paper - IV : Inference (DSC)

STATISTICS

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

[Short Answer Type]

Note : Answer any FIVE of the following questions.

1. Define Null and Alternative Hypothesis. **(Unit-I, SQA-1, 2)**
2. Explain Most powerful Test. **(Unit-I, SQA-4)**
3. Write the General test procedure of Large sample test. **(Unit-II, SQA-2)**
4. Define order Statistic. **(Unit-II, SQA-11)**
5. State central limit theorem. **(Unit-IV, SQA-1)**
6. Write the assumptions of t-test. **(Unit-III, SQA-7)**
7. Define Run and Length of Run. **(Unit-IV, SQA-5)**
8. Define Non parametric test and write its any two advantages over parametric tests. **(Unit-IV, SQA-6, 3)**

PART - B (4 × 15 = 60 Marks)

[Essay Answer Type]

Note : Answer ALL from the questions.

9. (a) State and prove Neymann-Pearson Lemma. Explain its importance in testing the hypothesis. **(Unit-I, Q.No. 8)**

OR

(b) Obtain Most powerful critical region for testing $H_0 : \mu = \mu_0$ Vs $H_1 : \mu = \mu_1$ for $N(\mu, \sigma^2)$ population. **(Unit-I, Prob. 15)**

10. (a) (i) Explain the large sample procedure for testing the equality of two population standard deviations. **(Unit-II, Q.No. 7)**

(ii) In a random sample of 500, the mean is found to be 20. In another independent sample of 400, the mean is 15. Could the samples have been drawn from the same population with SD 4. **(Unit-II, Prob. 11)**

OR

(b) (i) Explain Fisher's Z transformation. Explain its applications. **(Unit-II, Q.No. 8)**

(ii) Describe the large sample test procedure for equality of two proportions. **(Unit-II, Q.No. 7)**

11. (a) (i) Describe chi-square test for testing goodness of fit. **(Unit-III, Q.No. 14)**

(ii) Derive the chi-square test for 2 x 2 contingency table. **(Unit-III, Q.No. 15)**

OR

(b) (i) Explain the small sample test procedure for testing the equality of population variances. **(Unit-III, Q.No. 7)**

(ii) Explain t-test of significance of single mean. **(Unit-III, Q.No. 3)**

12. (a) (i) Explain Wald Wolfowitz run test for testing the equality of two population distribution functions. **(Unit-IV, Q.No. 12)**

(ii) To determine the percent shrinkage of a synthetic fibre, tests were made at two different temperatures. 10 tests were at lower temperature and higher temperature. The resulting data are given below:

Shrinkage	1	2	3	4	5	6	7	8	9	10
Low temperature	4.7	2.3	4.78	4.14	3.63	3.66	3.92	4.54	4.3	3.8
High temperature	3.1	4.2	4.1	4.4	4.05	3.5	3.75	4.04	4.1	5.3

Use an appropriate non parametric test to type whether the percent shrinkage of a synthetic fiber were made at two different temperature are equal OR not. **(Unit-IV, Prob. 4)**

OR

(b) (i) Explain Parametric tests Vs Non Non Parametric tests. **(Unit-IV, Q.No. 2)**

(ii) Explain Mann-Whitney U-test procedure. **(Unit-IV, Q.No. 11)**

FACULTIES OF SCIENCE

B.Sc. (CBCS) (IV – Semester) Examination

May/ June - 2018

STATISTICS

Time : 2 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer any five of the following questions.

ANSWER

1. Explain two types of error in testing of hypothesis. **(Unit-I, SQA-6)**
2. Explain about non-randomized test. **(Unit-I, SQA-5)**
3. Describe the large sample test procedure for difference of standard deviations. **(Unit-II, SQA-6)**
4. Explain Fisher's Z-transformation for population correlation coefficient. **(Unit-II, SQA-7)**
5. Explain the χ^2 - (Chi-square) test for goodness of Fit. **(Unit-III, SQA-3)**
6. Explain about F- test for equality of population veriances. **(Unit-III, SQA-4)**
7. Discuss advantages and disadvantages of non-parametric tests. **(Unit-IV, SQA-3)**
8. Describe nominal, ordinal, interval and ratio scales. **(Unit-IV, SQA-4)**

PART - B (4 × 15 = 60 Marks)

Note: Answer all questions from the following.

9. (a) State and prove Neymans Pearson Lemms. **(Unit-I, Q.No. 8)**

OR

(b) Let P the probability that a coin will fall head in a single toss in order to test $H_o : p = \frac{1}{2}$ against $H_1 : p = \frac{3}{4}$. The coin is tossed 5 times and H_o is rejected if more than 3 heads are obtained. Find the probability of Type I Error and power of the test. **(Unit-I, Prob. 7)**

10. (a) (i) Describe the test of significance of difference of proportions for large samples. **(Unit-II, Q.No. 7)**

(ii) Data on days to maturity were recorded in two varieties of a pulse crop Deter-mine whether two means are signifi-cantly different.

	n	Mean	Variance
Variety A	60	60	8.20
Variety B	65	65	11.13

(Unit-II, Prob. 9)

OR

(b) (i) Define order statistics and state their distributions. **(Unit-II, Q.No. 11)**

(ii) Explain run test procedure and its purpose for two sample case. **(Unit-IV, Q.No. 5)**

11. (a) Describe the χ^2 - test for independence of attributes and χ^2 - test for specified variance. **(Unit-III, Q.No. 13, 15)**

(b) Explain :

(i) t - test for difference of means **(Unit-III, Q.No. 14)**

(ii) t - test for specified mean. **(Unit-III, Q.No. 3)**

12. (a) (i) Explain the Median Test procedure. **(Unit-IV, Q.No. 10)**

(ii) Describe test procedure for Wilcoxon - Mann - Whitney U test **(Unit-IV, Q.No. 11)**

(b) (i) Explain Wilcoxon - Signed Rank Test for paired samples. **(Unit-IV, Q.No. 9)**

(ii) Describe the test procedure for sign test. **(Unit-IV, Q.No. 6)**

MAHATMA GANDHI UNIVERSITY

FACULTIES OF SCIENCE

B.Sc. (CBCS) (IV – Semester) Examination

May/ June - 2019

STATISTICS

Time : 3 Hours] **[Max. Marks : 80**

PART - A (5 × 4 = 20 Marks)

Note: Answer all the following questions.

ANSWER

1. Define type - I and type - II errors. **(Unit - I, SQA 6)**

2. For a sample of 400 observations from normal population with mean 95 and SD 12. Find 95% confidence limits for the population mean.

Ans :

Given that n = 400, $\bar{x}$ = 95, s = 12.

Then the 95% confidence limits for the population mean is

$[\bar{x} + 1.96\, s/\sqrt{n},\ \bar{x} - 1.96\, s/\sqrt{n}]$

$[95 + 1.96 \frac{12}{\sqrt{400}}, 95 - 1.96 \frac{12}{\sqrt{400}}]$

[95 + 1.96 (0.8485) , 95 – 1.96 (0.8485)]

[97.66, 93.33]

∴ **Conclusion :** The 95% confidence limits for population mean (μ) are [97.66, 93.33]

3. Explain t-test for single mean. **(Unit - III, SQA 1)**

4. Define sign test. **(Unit - IV, SQA 2)**

PART - B (4 × 15 = 60 Marks)

Note: Answer ALL the questions

5. (a) Obtain the best critical region for testing $H_0 : \theta = \theta_1$ against $H_1 : \theta = \theta_1$, for the exponential distribution. **(Unit - I, Prob.10)**

OR

(b) State and prove the Neymann Pearson's Lemma. **(Unit - I, Q.No.8)**

6. (a) (i) Explain the test procedure for single proportion. **(Unit - II, Q.No.6)**

(ii) A random sample of 64 items produced by a machine contained 14 defectives. Find 99% confidence limits for proportion defective items of population.

Ans :

Sample size (n) = 64

No. of defectives (x) = 14

P = sample proportion of defectives

$$P = \frac{x}{n} = \frac{14}{64} = 0.21875$$

p = 0.21875

q = 1 – p

q = 1 –0.21875

q = 0.7815

The 99% limits of proporation defective item is $p \pm z_r \sqrt{pq/n}$ [α = 1% L.O.S]

$$\left[p + z_\alpha \sqrt{pq/n} \,,\, p - z_\alpha \sqrt{pq/n} \right] \quad [\therefore z_\alpha = 2.58]$$

$$\left[0.21875 + 2.58\sqrt{\frac{0.21875 \times 1815}{64}}, 0.21875 - 2.58\sqrt{\frac{0.21875 \times 0.1815}{64}} \right]$$

[0.21875 + 2.58 (0.0515), 0.21875 – 2.58 (0.0515)]

[0.21875 + 0.1329 , 0.21875 –0.1329]

[0.3517 , 0.0859]

Conclusion : The 99% confidence limits for proportion defective item is (0.3517, 0.0859)

OR

(b) (i) Explain the test procedure for difference of proportions. **(Unit - II, Q.No.7)**

(ii) The correlation coefficient of a bivariate random sample of size 50 is 0.36 can this be regarded as drawn from a normal population with correlation coefficient 0.5 at 1% loss.

Ans :

Null Hypothesis (H_0) : The population correlation coefficient is 0.5 i.e H_0 = 0.5

Alternative Hypothesis (H_1) : The population correlation coefficient is not 0.5

i.e., $H_1 : \neq 0.5$ [Two Tailed Test]

Level of signifitance :

Consider the oppropriate at 1% level of significance for two Tailed test is 2.58.

Test statistics : Given that

n =50, s = 0.5, r = 0.36

under the Null Hypothesis then the test statistics is given by

$$Z = \left|\frac{z_r - z_s}{\sqrt{1/n-3}}\right| \sim N(0,1) \text{ under } H_0$$

Where

$$z_r = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)$$

$$z_r = \frac{1}{2}\log_e\left(\frac{1+0.36}{1-0.36}\right)$$

$$z_r = \frac{1}{2}\log_e\left(\frac{1.36}{0.64}\right)$$

$$z_r = \frac{1}{2}\log_e(2.1250)$$

$$z_r = \frac{1}{2}(0.3274)$$

$$\boxed{z_r = 0.1657}$$

And $z_s = \frac{1}{2}\log_e\left(\frac{1+s}{1-s}\right)$

$$z_s = \frac{1}{2}\log_e\left(\frac{1+0.5}{1-0.5}\right)$$

$$z_s = \frac{1}{2}\log_e\left(\frac{1+0.5}{1-0.5}\right)$$

$$z_s = \frac{1}{2}\log_e\left(\frac{1.5}{0.5}\right)$$

$$z_s = \frac{1}{2}\log_e(3)$$

$$z_s = \frac{1}{2}(0.4771)$$

$$\boxed{z_s = 0.2386}$$

Substitute z_r, z_y values in z

$$z = \left| \frac{z_r - z_s}{1/\sqrt{n-3}} \right| \sim N(0,1) \text{ under } H_0$$

$$z = \left| \frac{0.1637 - 0.2386}{1/\sqrt{50-3}} \right| \sim N(0,1) \text{ under } H_0$$

$$z = \left| \frac{0.1637 - 0.2386}{1/\sqrt{47}} \right|$$

$$z = \left| \frac{-0.0749}{0.1459} \right|$$

$z = |-0.5134|$

$z_{cal} = 0.5134$

The tabulated value at 1% L.O.S is 2.58

$\therefore\ z_{cal} < z_{tab}$

$0.1534 < 2.58$

we accept H_0

Conclusion : The population correlation coefficient is 0.5.

7. (a) (i) Explain paired t-test for difference of means. **(Unit - III, Q.No.5)**

(ii) Explain the test procedures for sample correlation coefficient based on students. **(Unit - III, Q.No.6)**

OR

(b) (i) Explain $\chi^2 =$ test for Indepence of Attributes. **(Unit - III, Q.No.15)**

(ii) Two random samples of sizes 9 and 12 have the SD's 2.9 and 2.6. Test the significance difference between the variances at 5% loss. ($F_{0.5',\ 3,\ 11} = 2.95$)

Ans :

Null Hypothesis (H_0): The samples are drawn from same normal population.

i.e., $H_0 : \sigma_x^2 = \sigma_y^2$

Alternative Hypothesis (H_1) : The samples are not drawn from same normal population

i.e., $H_1 : \sigma_x^2 \neq \sigma_y^2$ [Two Tailed Test]

level of significance (α) : Consider the appropriate at 5% L.O.S.

Test statistics :

Given that

$n_1 = 9$ $n_2 = 12$

$s_x = 2.9$ $s_y = 2.6$

$s_x^2 = 8.41$ $s_y^2 = 6.76$

$\therefore\ s_x^2 + s_y^2$

Then the test statistics is given by $F = \dfrac{s_x^2}{s_y^2} \sim F(8,11)$ d.f

$$F = \frac{8.41}{6.76} \sim F(8,11) \text{ d.f}$$

$F_{cal} = 1.2441$

The tabulated value of F at 5% L.O.S with (8,11) d.f is 2.95

The calculated value of F is lessthan the tabulated value of F.

$\therefore\ F_{cal} < F_{tab}$

i.e., $1.2441 < 2.95$

$\therefore$ we accept H_0

Conclusion : The samples are drawn from same normal population.

8. (a) Define Non - parametric tests. Distinguish between parametric and non-parametric test. **(Unit - IV, Q.No.1,2)**

OR

(b) Explain Wald - Wolfowitz Run test. **(Unit - IV, Q.No.11)**

FACULTY OF SCIENCE

B.Sc. IV-Semester (CBCS) Examination

Model Paper - I

Subject : STATISTICS

Paper - IV : Inference

Time : 3 Hours] **[Max. Marks : 80**

PART - A (8 × 4 = 32 Marks)

[Short Answer Type]

Note : Answer any EIGHT of the following questions.

1. Critical Region **(Unit-I, SQA-3)**
2. Write a short note on Non-Randomized and Randomized test function. **(Unit-I, SQA-5)**
3. Negative Binomial distribution. **(Unit-I, SQA-10)**
4. Define Large Sample Test. **(Unit-II, SQA-1)**
5. Describe the test procedure for diffe-rence of proportions. **(Unit-II, SQA-3)**
6. Applications of Z-transformation **(Unit-II, SQA-8)**
7. Explain χ^2 - test for independent of two attribute. **(Unit-III, SQA-2)**
8. Explain t-test for single mean. **(Unit-III, SQA-1)**
9. Applications of F-Test. **(Unit-III, SQA-9)**
10. Write a short notes on sign tests. **(Unit-IV, SQA-2)**
11. Write a short note on Measurement of Scale. **(Unit-IV, SQA-4)**
12. What is meant by Non-parametric Tests. **(Unit-IV, SQA-6)**

PART - B (4 × 12 = 48 Marks)

[Essay Answer Type]

Note : Answer ALL from the questions.

13. One of the probability coin will fall Head in a single test in order to test $H_0 : P = \frac{1}{2}$ and $H_1 : P = \frac{3}{4}$. The coin is tassed 5 times and H_0 is Rejected More than 3 heads and obtained probability of Type-I and Type - II Error and power of the test. **(Unit-I, Prob. 4)**

(OR)

14. Write a short notes on One and two tailed tests. **(Unit-I, Q.No. 5)**
15. Explain the Procedure for Testing of Hypothesis. **(Unit-II, Q.No. 2)**

(OR)

16. A random sample of 100 articles are selected from batch of 2000 articles show that the average diagramatical of the article 0.354 with standard deviation 0.048. Find the 95% confidence limits for the average of these batch 2000 articles. Test whether the samples has been drawn from the population mean is 0.421. **(Unit-II, Prob. 7)**

17. What is t-distribution? Explain the properties and applications of t-distribution. **(Unit-III, Q.No. 2)**

(OR)

18. Two random samples of 11,9 observations show the sample standard deviation of their weights 0.8 and 0.5 respectively. Assuming the weights distributions are normal. Test the hypothesis that there variances are gains the alternative hypothesis that they are not equal at 5% level of significance. **(Unit-III, Prob. 15)**

19. (a) What is meant by Non-parametric Tests. State their Advantages and Disadvantages. **(Unit-IV, Q.No. 1)**

(OR)

20. (a) The following examination scores regarded

Pairs	1	2	3	4	5	6	7	8	9	10
with sample problem	531	621	663	579	451	660	591	719	543	575
With out sample problem	509	540	688	502	424	683	568	748	530	524

Test the Null Hypothesis at 0.05 level of significance. The sample problem increase the scores by 50 points against alternative hypothesis that the increase is less than 50 points. **(Unit-IV, Prob. 6)**

FACULTY OF SCIENCE

B.Sc. IV-Semester (CBCS) Examination

Model Paper - II

Subject : STATISTICS

Paper - IV : Inference

Time : 3 Hours] **[Max. Marks : 80**

PART - A (8 × 4 = 32 Marks)

[Short Answer Type]

Note : Answer any EIGHT of the following questions.

1. Write short notes on Type-I and Type-II Error. **(Unit-I, SQA-6)**
2. Most Powerful test. **(Unit-I, SQA-8)**
3. Exponential distribution. **(Unit-I, SQA-11)**
4. Derive the large sample test procedure for difference of means. **(Unit-II, SQA-4)**
5. Explain the Procedure for Testing of Hypothesis. **(Unit-II, SQA-2)**
6. Derive the test procedure for Single Correlation for Large Sample. **(Unit-II, SQA-9)**
7. Explain the concept of F-test for equality of population variance. **(Unit-III, SQA-4)**
8. Write a short notes on χ^2 - test for goodness of fit. **(Unit-III, SQA-3)**
9. What is χ^2 - test? **(Unit-III, SQA-10)**
10. State central limit theorem. **(Unit-IV, SQA-1)**
11. Discuss the advantages and disadvatages of Non-parametric tests. **(Unit-IV, SQA-3)**
12. Define Run and Length of Run. **(Unit-IV, SQA-5)**

PART - B (4 × 12 = 48 Marks)

[Essay Answer Type]

Note : Answer ALL from the questions.

13. Obtain the most powerful test for testing the mean $\mu = \mu_0$ against $\mu = \mu_1(\mu_1 = \mu_0)$ where $\sigma^2 = 1$ in normal population. **(Unit-I, Prob. 15)**

(OR)

14. State and prove the Neymann Pearson's Lemma. **(Unit-I, Q.No. 8)**
15. Derive the large sample test procedure for difference of means. **(Unit-II, Q.No. 4)**

(OR)

16. A random sample of 400 men and 600 women where asked whether they would like to have flyover near these residence 200 men and 325 women were in the favour of the proposal. Test the hypo-thesis that the proportions of men and women were in favour of the proposal are same significance that they are not. **(Unit-II, Prob. 21)**

17. Explain t-test for single mean. **(Unit-III, Q.No. 3)**

(OR)

18. The die is thrown 6 times

Phase	1	2	3	4	5	6
Dies	8	7	12	8	14	11

To test the 5% level of significance of the die is unbiased. **(Unit-III, Prob. 20)**

19. Distinguish between Parametric Test and Non-parametric Test. **(Unit-IV, Q.No. 2)**

(OR)

20. In a Industrial production line items are inspected periodically for defective. The following is a sequence of defective items 'D' and Non-defective items 'N' produced by the production line

D D D N N N D N N D D N N N N D D D N N D N N N N N D N D

Determine whether the defective items are occure in random or not at 5% level of significance. **(Unit-IV, Prob. 7)**

FACULTY OF SCIENCE

B.Sc. IV-Semester (CBCS) Examination

Model Paper - III

Subject : STATISTICS

Paper - IV : Inference

Time : 3 Hours] **[Max. Marks : 80**

PART - A (8 × 4 = 32 Marks)

[Short Answer Type]

Note : Answer any EIGHT of the following questions.

1. Write a short notes on Level of Significance. **(Unit-I, SQA-7)**
2. Find the Liklihood function of Poisson distribution. **(Unit-I, SQA-9)**
3. Find the best critical region for testing $H_0 : \lambda = \frac{2}{3}$ and against $H_1 : \lambda = \frac{4}{3}$ of the poisson distribution. **(Unit-I, Prob. 6)**
4. Write short notes on large sample test of single proportion. **(Unit-II, SQA-5)**
5. Define order statistics. **(Unit-II, SQA-11)**
6. Explain the procedure test for standard deviation. **(Unit-II, SQA-6)**
7. Define small sample test. **(Unit-III, SQA-6)**
8. Properties of Student's t-Distribution **(Unit-III, SQA-5)**
9. Features of Chi-square (χ^2) Test **(Unit-III, SQA-11)**
10. Explain briefly about Wilcoxon Sign Ranked Test. **(Unit-IV, SQA-7)**
11. Explain the procedure for Wald-wolfowitz Run test for two samples. **(Unit-IV, SQA-9)**
12. Distinguish between Parametric Test and Non-parametric Test. **(Unit-IV, SQA-10)**

PART - B (4 × 12 = 48 Marks)

[Essay Answer Type]

Note : Answer ALL from the questions.

13. Find the best critical region for testing $H_0 : \theta = \frac{3}{2}, H_1 : \theta = \frac{4}{3}$ of the Exponential distribution. **(Unit-I, Prob. 9)**

(OR)

14. Given Frequency function $f(x,\theta) = \frac{1}{\theta}e^{-x/\theta};\ 0 \le x \le \infty$ for testing the Null Hypothesis $H_0 : \theta = 3$ and alternative hypothesis $H_1 : \theta = 1$. Find the probability of Type -I and Type-II Error and power of the curve critical region $x \le 2$. **(Unit-I, Prob. 3)**

15. Write short notes on large sample test of single proportion. **(Unit-II, Q.No. 6)**

(OR)

16. 1000 apples kept under the one type of the storage were found to be showing rating to the extent of 4%. 1500 apples kept under the another type of storage showed 3% rating can it be reasonable to conclude that time second type of storage, is superior that the first type of storage? **(Unit-II, Prob. 22)**

17. Explain the concept of F-test for equality of population variance. **(Unit-III, Q.No. 7)**

(OR)

18. Explain χ^2 - test for independent of two attribute. **(Unit-III, Q.No. 15)**

19. Explain the median test procedure. **(Unit-IV, Q.No. 9)**

(OR)

20. Explain the procedure for Wald-wolfowitz Run test for two samples. **(Unit-IV, Q.No. 12)**