R.M.K
GROUP OF ENGINEERING INSTITUTIONS

R.M.K
GROUP OF
INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS

Please read this disclaimer before proceeding:

This document is confidential and intended solely for the educational purpose of RMK Group of Educational Institutions. If you have received this document through email in error, please notify the system manager. This document contains proprietary information and is intended only to the respective group / learning community as intended. If you are not the addressee you should not disseminate, distribute or copy through e-mail. Please notify the sender immediately by e-mail if you have received this document by mistake and delete this document from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.

22MA401

PROBABILITY AND STATISTICS

DEPARTMENT	**Artificial Intelligence and Data Science**
BATCH/YEAR	**2022-2026/ II**
CREATED BY	**Dr. V. Banu Priya & Dr. S. Vidhya**
DATE	**26.02.2024**

Table of Contents

COURSE OBJECTIVE

S. No	TOPIC
1	To Provide the necessary basic concepts of random variables and to introduce some standard distributions.
2	To introduce the basic concepts of two dimensional random variables
3	To test the hypothesis for small and large samples.
4	To introduce the concepts of Analysis of Variances.
5	To understand the concept of statistical quality control

PREREQUISITES

S. No.	TOPICS	COURSE NAME WITH CODE
1	Basic Probability	Higher Secondary level
2	Basic Statistics	

Syllabus

22MA401	PROBABILITY AND STATISTICS (Theory Course with Laboratory Component)	L T P C 3 2 0 4

UNIT I ONE DIMENSIONAL RANDOM VARIABLES **15**

Basic probability definitions- Independent events- Conditional probability (revisit) - Random variable - Discrete and continuous random variables – Moments – Moment generating functions – Binomial, Poisson, Geometric, Uniform, Exponential and Normal distributions.

Experiments using R Programming:

1. Finding conditional probability.
2. Finding mean, variance and standard deviation.

UNIT II TWO DIMENSIONAL RANDOM VARIABLES **15**

Joint distributions – Marginal and conditional distributions – Covariance – Correlation and linear regression – Transformation of random variables.

Experiments using R Programming:

1. Finding marginal density functions for discrete random variables
2. Calculating correlation and regression

UNIT III TESTING OF HYPOTHESIS **15**

Sampling distributions - Estimation of parameters - Statistical hypothesis - Large sample tests based on Normal distribution for single mean and difference of means -Tests based on t and F distributions for mean and variance – Chisquare - Contingency table (test for independence) - Goodness of fit.

Experiments using R Programming:

1. Testing of hypothesis for given data using Z – test.
2. Testing of hypothesis for given data using t – test.

UNIT IV DESIGN OF EXPERIMENTS **15**

One way and Two way classifications - Completely randomized design – Randomized block design – Latin square design.

Experiments using R Programming:

1. Perform one- way ANOVA test for the given data.
2. Perform two-way ANOVA test for the given data.

UNIT V STATISTICAL QUALITY CONTROL **15**

Control charts for measurements (X and R charts) – Control charts for attributes (p, c and np charts) – Tolerance limits.

Experiments using R Programming:

1. Interpret the results for $\overline{X}$-Chart for variable data
2. Interpret the results for R-Chart for variable data

TOTAL: 75 PERIODS

COURSE OUTCOMES

Course Outcomes	Description	Knowledge Level
CO1	Understand the fundamental knowledge of modern probability theory and standard distributions.	K1, K2
CO2	Categorize the probability models and function of random variables based on one and two dimensional random variables.	K2
CO3	Employ the concept of testing the hypothesis in real life problems.	K3
CO4	Implement the analysis of variance for real life problems.	K3
CO5	Apply the statistical quality control in engineering and management problems.	K3

R.M.K GROUP OF INSTITUTIONS

CCO-PO/CO-PSO Mapping

CO's	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	1	-	-	-	-	-	1	-	-	-
CO2	3	2	1	-	-	-	-	-	1	-	-	-
CO3	3	2	1	-	-	-	-	-	1	-	-	-
CO4	3	2	1	-	-	-	-	-	1	-	-	-
CO5	3	2	1	-	-	-	-	-	1	-	-	-

CO's	PSO1	PSO2	PSO3
CO1	-	-	-
CO2	-	-	-
CO3	-	-	-
CO4	-	-	-
CO5	-	-	-

LECTURE PLAN

S.No.	Topics to be Recovered	No. of periods	Proposed Date	Actual Date	CO	Knowledge level	Mode of Delivery
1	Introduction	1	10.02.24		CO3	K2	PPT,Chalk, Board
2	Large sample test for single mean	1	12.02.24		CO3	K3	PPT,Chalk, Board
3	Large sample test for difference of means	1	13.02.24		CO3	K3	PPT,Chalk, Board
4	't' distribution for single mean	1	14.02.24		CO3	K3	PPT,Chalk, Board
5	't' distribution for difference of means	1	15.02.24		CO3	K3	PPT,Chalk, Board
6	Test based on Chi-square distribution	1	16.02.24		CO3	K3	PPT,Chalk, Board
7	Chi-square test for Goodness of fit	1	17.02.24		CO3	K3	PPT,Chalk, Board
8	Solving Problems	1	22.02.24		CO3	K3	PPT,Chalk, Board
9	Chi-square test for Independence of attributes	1	23.02.24		CO3	K3	PPT,Chalk, Board
10	Solving Problems	1	24.02.24		CO3	K3	PPT,Chalk, Board
11	F-distribution (Test for Variance)	1	26.02.24		CO3	K3	PPT,Chalk, Board
12	Solving Problems	1	27.02.24		CO3	K3	PPT,Chalk, Board

ACTIVITY BASED LEARNING

Activity based learning helps students express and embrace their curiosity. Once the students become curious, they tend to explore and learn by themselves. To evoke curiosity in students, Practice quiz is designed for all the five units.

Quiz – Testing of Hypothesis

https://quizizz.com/admin/quiz/58d2bf6d47b3a35403c4a42a/hypothesis-test

https://quizlet.com/73300415/hypothesis-testing-flash-cards/

LECTURE NOTES:
UNIT III TESTING OF HYPOTHESIS

IMPORTANT TERMINOLOGIES:

POLULATION:

A Population in statistics means a set of objects or mainly the set of numbers which are measurements or observations pertaining to the objects. The population is finite or infinite according to the number of elements of the set is finite or infinite.

SAMPLING:

A part selected from the population is called a sample. The process of selection of a sample is called sampling.

RANDOM SAMPLING:

A random sampling is one in which each number of population has an equal chance of being included in it. There are NC_n different samples of size n that can be picked up from a population size N

PARAMETERS:

The statistical constants of the population are called Parameters.

Population size = N

Population mean = μ

Population Standard deviation = σ

Population proportion = p

STATISTICS:

The statistical constants of the sample are called statistics.

Sample size = n

Sample mean = $\bar{x}$

Sample standard deviation = s

Sample proportion = p'

TESTING A HYPOTHESIS:

On the basis of a sample information, we make certain decisions about the population. In taking such decisions, we make certain assumptions. These assumptions are known as statistical hypothesis. There hypothesis are tested.

Assuming the hypothesis correct, we calculate the probability of getting the observed sample. If this probability is less than a certain assigned value , the hypothesis is to be accepted, otherwise rejected.

NULL HYPOTHESIS:

Null hypothesis is based on analysing the problem, it is the hypothesis of no difference. Thus we shall presume that there is no significant difference between the observed value and expected value. Null hypothesis is denoted by H_o.

ALTERNATE HYPOTHESIS:

Any hypothesis which is complementary to the null hypothesis is called an Alternative hypothesis and it is denoted by H_1.

CRITICAL REGION:

A region corresponding to a statistic t in the sample space S which amounts to rejection of Null hypothesis is called as critical region or rejection region.

The region of the sample space S which amounts to the acceptance of Null hypothesis is called acceptance region.

CRITICAL VALUE OR SIGNIFICANT VALUE:

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

LEVEL OF SIGNIFICANCE:

The probability that the value of the statistic lies in the critical region is called the level of significance. In general the levels are 1% and 5%.

ERRORS:

In sampling theory to draw valid inferences about the population parameter on the basis of the sample results, we decide to accept or to reject H_o after examining a sample from it.

TYPE I ERROR:

IF H_o is rejected while it should have been accepted.

TYPE II ERROR:

IF H_o is accepted while it should have been rejected.

GENERAL PROCEDURE OF HYPOTHESIS TESTS:

(i). From the problem context identify the parameter of interest.

(ii). State the Null hypothesis H_o.

(iii). Specify an appropriate alternate hypothesis H_1.

(iv). Choose a significance level α.

(v). Determine an appropriate test statistic.

(vi). State the rejection region for the statistic.

(vii). Compute any necessary sample quantities, substitute these into the equation of the test statistic and compute the value.

(viii). Conclusion: Decide whether or not H_o should be rejected and report that in the problem context.

Critical Values	Level of Significance		
	1%	5%	10%
Two tailed test	|z|=2.58	|z|=1.96	|z|=1.28
Right tailed test	Z = 2.33	Z = 1.645	Z = 1.28
Left tailed test	Z = -2.33	Z = -1.645	Z= -1.298

Large sample test for single mean – Z Test

Note:

For large sample, s nearly equal to σ the test statistic Z is given by

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ or $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

Problems:

1). A sample of 900 members has a mean 3.4 cm and standard deviation 2.61 cm. Is the sample from a large population of mean 3.25 cms and standard deviation of 2.61 cms. Test at 5% level of significance. The value of Z at 5% level of significance is $|Z_\alpha| < 1.96$

Solution:

Given that $n = 900$, $\mu = 3.25$, $s = 2.61$, $\bar{x} = 3.4$, $\alpha = 5\%$

(i). $H_o: \mu = 3.25$

(ii). $H_1: \mu \neq 3.25$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $-1.96 < Z < 1.96$

(v). The Test statistic: $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}} = 1.724$

(vi). Conclusion: Since $Z = 1.724$ lies in the interval $-1.96 < Z < 1.96$, we accept H_o

2). A normal population has a mean of 6.48 and standard deviation of 1.5. In a sample of 400 members mean is 6.75. Is the difference significant?

Solution:

Given that $n = 400$, $\mu = 6.48$, $s = 1.5$, $\bar{x} = 6.75$

(i). $H_o: \mu = \bar{x}$ [No significant difference]

(ii). $H_1: \mu \neq \bar{x}$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $-1.96 < Z < 1.96$

(v). The Test statistic: $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{6.75-6.48}{1.5/\sqrt{400}} = 3.6$

(vi). Conclusion: Since $Z = 3.6$ does not lie in the interval $-1.96 < Z < 1.96$, we reject H_o

Hence there is a significance difference.

3). The mean breaking through of the cables supplied by a manufacture is 1800 with a S.D of 100. By a new technique in the manufacturing process , it is claimed that the breaking strength of a cable has increased. In order to test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?

Solution:

Given that $n = 50$, $\mu = 1800$, $s = 100$, $\dot{x} = 1850$, $\alpha = 1\%$

(i). $H_o: \mu = 1800$

(ii). $H_1: \mu > 1800$

(iii). $\alpha = 1\%$

(iv). Acceptance Region: $Z < 2.33$

(v). The Test statistic: $Z = \frac{\dot{x}-\mu}{s/\sqrt{n}}$

$$= \frac{1850-1800}{100/\sqrt{50}}$$

$$= 3.54$$

(vi). Conclusion: Since the test statistic $Z = 3.54$ does not satisfy $Z < 2.33$,

We reject H_o

4). An insurance agent has claimed that the average age of policy holders who insure through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had insured through him reveal that the mean and S.D are 28.8 years and 6.35 years respectively. Test his claim at 5% level of significance

Solution:

Given that $n = 100$, $\mu = 30.5$, $s = 6.35$, $\dot{x} =28.8$, $\alpha = 5\%$

(i). $H_o: \mu = 30.5$

(ii). $H_1: \mu < 30.5$

R.M.K GROUP OF INSTITUTIONS

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $Z > -1.645$

(v). The Test statistic: $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{28.8-30.5}{6.35/\sqrt{100}} = -2.68$

(vi). Conclusion: Since the test statistic $Z = -2.68$ does not satisfy $Z > -1.645$,

We reject H_o

5). The mean life time of a sample of 100 light bulbs produced by a company is computed to be 1570 hours with a S.D of 120 hours. If μ is the mean life time of all the bulbs produced by the company, test the hypothesis $\mu = 1600$ hours, against the alternative hypothesis $\mu \neq 1600$ hours with $\alpha = 0.05$ and $\alpha = 0.01$

Solution:

Given that $n = 100$, $\mu = 1600$, $s = 120$, $\bar{x} = 1570$, $\alpha = 5\%$, 1%

(i). $H_o: \mu = 1600$

(ii). $H_1: \mu \neq 1600$

(iii). (a). $\alpha = 5\%$ (b). $\alpha = 1\%$

(iv). Acceptance Region: (a). $-1.96 < Z < 1.96$ (b). $-2.58 < Z < 2.58$

(v). The Test statistic: $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{1570-1600}{120/\sqrt{100}} = -2.5$

(vi). Conclusion:

(a). Since $Z = -2.5$ does not lie in the interval $-1.96 < Z < 1.96$, we reject H_o at 5% level

(b). Since $Z = -2.5$ lies in the interval $-2.58 < Z < 2.58$, we accept H_o at 1% level.

Large sample test based on difference of means

Note:

The test statistic Z is given by $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ (or) $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$

Problems:

1). The means of two large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches?

Solution:

Given that $n_1 = 1000, n_2 = 2000$, $\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$, $\sigma_1 = 2.5 = \sigma_2$

(i). $H_o: \mu_1 = \mu_2$[No significant difference]

(ii). $H_1: \mu_1 \neq \mu_2$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $-1.96 < Z < 1.96$

(v). The Test statistic: $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

$$Z = \frac{67.5-68}{\sqrt{\frac{2.5^2}{1000}+\frac{2.5^2}{2000}}} = -5.16$$

(vi). Conclusion:

Since $Z = -5.16$ does not lie in the interval $-1.96 < Z < 1.96$, we reject H_o

2). Random samples drawn from two countries give the following data relating to the heights of adult males. Is the difference between standard deviation significant?

Solution:

	Country A	Country B
Mean Height (inches)	67.42	67.25
S.D (inches)	2.58	2.5
Number in samples	1000	1200

Solution:

Given that $n_1 = 1000, n_2 = 1200$, $\dot{x}_1 = 67.42$, $\dot{x}_2 = 67.25$, $s_1 = 2.58$, $s_2 = 2.5$

(i). $H_o: \mu_1 = \mu_2$[No significant difference]

(ii). $H_1: \mu_1 \neq \mu_2$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $-1.96 < Z < 1.96$

(v). The test statistic: $Z = \frac{\dot{x}_1 - \dot{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$Z = \frac{67.42 - 67.25}{\sqrt{\frac{2.58^2}{1000} + \frac{2.5^2}{1200}}}$$

$$Z = 1.567$$

(vi). Conclusion:

Since $Z = 1.567$ lies in the interval $-1.96 < Z < 1.96$, we accept H_o

3). The average marks scored by 32 boys is 72 with a S.D of 8, while that for 36 girls is 70 with a S.D of 6. Test at 1% level of significance whether the boys perform better than girls

Solution:

Given that $n_1 = 32$, $n_2 = 36$, $\dot{x}_1 = 72$, $\dot{x}_2 = 70$, $s_1 = 8$, $s_2 = 6$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 > \mu_2$

(iii). $\alpha = 1\%$

(iv). Acceptance Region: $Z < 2.33$

(v). The test statistic: $Z = \frac{\dot{x}_1 - \dot{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$Z = \frac{72 - 70}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}}$$

$$Z = 1.15$$

(vi). Conclusion: Since the test statistic $Z = 1.15$ satisfies $Z < 2.33$,

We accept H_o.

4). A simple sample of 6400 Englishmen has a mean of 170 inches and a standard deviation of 6.4 inches, while a simple sample of heights of 1600 Australians has a mean of 172 inches and a standard deviation of 6.3 inches. Do the data indicate that Australians are on the average taller than English men?

Solution:

Given that $n_1 = 6400$, $n_2 = 1600$, $\bar{x}_1 = 170$, $\bar{x}_2 = 172$, $s_1 = 6.4$, $s_2 = 6.3$

μ_1 indicates that mean height of the population of Englishmen

μ_2 indicates that mean height of the population of Australian

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 < \mu_2$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $Z > -1.645$

(v). The test statistic: $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$Z = \frac{170 - 172}{\sqrt{\frac{6.4^2}{6400} + \frac{6.3^2}{1600}}}$$

$$Z = -11.32$$

(vi). Conclusion: Since the test statistic $Z = -11.32$ does not satisfy $Z > -1.645$,

We reject H_o.

5). The mean height of 50 male students who showed above average participation in college athletics was 68.2 inches with a standard deviation of 2.5 inches, while 50 male students who showed no interest in such participation had a mean height of 67.5 inches with a standard deviation of 2.8 inches

(a). Test the hypothesis that male students who participate in college athletics are taller than other male students.

(b). By how much should the sample size of each of the two groups be increased in order that the observed difference of 0.7 inches in the mean heights be significant at the 5% level of significance?

Solution:

Given that $n_1 = 50$, $n_2 = 50$, $\bar{x}_1 = 68.2$, $\bar{x}_2 = 67.5$, $s_1 = 2.5$, $s_2 = 2.8$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 > \mu_2$

(iii). $\alpha = 5\%$

(iv). Acceptance Region: $Z < 1.645$

(v). The test statistic: $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$Z = \frac{68.2 - 67.5}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}}$$

$$Z = 1.32$$

(vi). Conclusion: Since the test statistic $Z = 1.32$ satisfies $Z < 1.645$,

We accept H_o.

Therefore, we conclude that the college athletics are not taller than other male students.

(b). The difference between the mean heights of two groups, each of size n will be significant at 5% level of significance if $Z > 1.645$

$$\frac{68.2 - 67.5}{\sqrt{\frac{2.5^2}{n} + \frac{2.8^2}{n}}} > 1.645$$

$$\frac{0.7}{\sqrt{14.09/n}} > 1.645$$

$$\frac{0.7\sqrt{n}}{3.754} > 1.645$$

$$\frac{0.7^2 n}{3.754^2} > 1.645^2$$

$$n > 77.83$$

Which implies $n = 78$.

Hence, the sample size of the two groups should be increased by at least 78-50=28, in order that the difference between the mean heights of the two groups is significant.

Testing of hypothesis- t Distribution

Testing of hypothesis for Mean using t-distribution

Applications of t-distribution:

(i). It is used to test the significance of the difference of the mean of a random sample and the mean of the population.

(ii). It is used to test the significance of the difference between two sample means.

(iii). To test the significance of an observed sample correlation coefficient and sample regression coefficient.

(iv). To test the significance of observed partial and multiple correlation coefficients.

Note:

For the samples $(n < 30)$, σ known, decision is based on the t-distribution with $v = n - 1$ degrees of freedom.

The test statistic for t-distribution is $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

Problems based on t-test using single mean

1). A mechanist is making engine parts with axle diameters of 0.7 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.04 inch. Compute the statistic you would use to test, whether the work is meeting the specification.

Solution:

Given that $n = 10$, $\bar{x} = 0.742$, $s = 0.04$, $\mu = 0.7$

(i). H_0: $\mu = 0.7$

(ii). H_1: $\mu \neq 0.7$

(iii). $\alpha = 5\%$, $d.f = n - 1 = 10 - 1 = 9$

(iv). Acceptance Region: $-2.262 < t < 2.262$

(v). The Test Statistic: $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$t = \frac{0.742 - 0.7}{0.04/\sqrt{9}}$$

$$t = 3.15$$

(vi). Conclusion: Since $t = 3.15$ does not lie in the interval $-2.262 < t < 2.262$, we reject H_o. i.e., the product not confirming the specification.

2). A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviation from this mean equal to 135 square inches. Show that the assumption of a mean of43.5 inches for the population is reasonable under 1% level of significance. Also obtain 99% fiducial limit for the same.

Solution:

Given that $n = 16$ $\bar{x} = 41.5$, $\mu = 43.5$,

$\sum(x - \bar{x})^2 = 135$

$s^2 = \frac{\sum(x-\bar{x})^2 = 135}{n} = \frac{135}{16} = 8.44$

$s = 2.9$

(i). H_0: $\mu = 43.5$

(ii). H_1: $\mu \neq 43.5$

(iii). $\alpha = 1\%$, $d.f = n - 1 = 16 - 1 = 15$

(iv). Acceptance Region: $-2.947 < t < 2.947$

(v). The Test Statistic: $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$t = \frac{41.5-43.5}{2.9/\sqrt{15}}$$

$$t = -2.67$$

(vi). Conclusion: Since $t = -2.67$ lies in the interval $-2.947 < t < 2.947$, we accept H_o. Hence, we conclude that the assumption is reasonable.

99% fiducial limits for μ is given by

$\bar{x} \pm t_{0.01} \frac{s}{\sqrt{n}}$

$= 41.5 \pm 2.947(\frac{2.9}{4})$

$= 43.71 \ \& \ 39.29$

i.e., $39.29 < \mu < 43.71$

3). A certain injection administered to each of 12 patients resulted in the following increases of blood pressure:

5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4

Can it be concluded that the injection will be in general accompanied by an increase in B.P?

Solution:

Given data is 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4

So, $n = 12$

$\bar{x} = \frac{\Sigma x}{n} = \frac{31}{12} = 2.58$

$s^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 = \frac{185}{12} - 2.58^2 = 8.761$

$s = 2.96$

(i). H_0: $\mu = 0$[No increase of blood pressure]

(ii). H_1: $\mu > 0$

(iii). $\alpha = 5\%$, $d.f = n - 1 = 12 - 1 = 11$

(iv). Acceptance Region: $t < 1.796$

(v). The Test Statistic: $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$$t = \frac{2.58 - 0}{2.96/\sqrt{11}}$$

$$t = 2.89$$

(vi). Conclusion: Since $t = 2.89$ does not satisfy $t < 1.796$, we reject H_o. Hence, we conclude that the injection is accompanied by an increase in B.P.

4). The mean life time of a sample of 25 bulbs is found as 1550 hours with a S.D of 120 hours. The company manufacturing the bulbs claims that the average life of their bulbs is 1600 hours. Is the claim acceptable at 5% level of significance?

Solution:

Given that $n = 25$ $\bar{x} = 1550$, $\mu = 1600$, $s = 120$

(i). H_0: $\mu = 1600$

(ii). H_1: $\mu < 1600$

(iii). $\alpha = 5\%$, $d.f = n - 1 = 25 - 1 = 24$

(iv). Acceptance Region: $t > -1.711$

(v). The Test Statistic: $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$$t = \frac{1550 - 1600}{120/\sqrt{24}}$$

$$t = -2.04$$

R.M.K GROUP OF INSTITUTIONS

(vi). Conclusion: Since $t = -2.04$ does not satisfy $t > -1.711$, we reject H_o. Hence, we conclude that the claim of the company can not be accepted at 5% level of significance.

5). A random sample of 10 boys had the following I.Qs:

70, 120, 110, 101, 88, 83, 95, 98, 107, 100

Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie.

Solution:

Given data is 70, 120, 110, 101, 88, 83, 95, 98, 107, 100

So, $n = 10$

$\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$

$s^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 = \frac{96312}{10} - 97.2^2 = 183.36$

$s = 13.5$

(i). $H_0: \ \mu = 100$

(ii). $H_1: \ \mu \neq 100$

(iii). $\alpha = 5\%$, $d.f = n - 1 = 10 - 1 = 9$

(iv). Acceptance Region: $-2.262 < t < 2.262$

(v). The Test Statistic: $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$t = \frac{97.2-100}{13.5/\sqrt{9}}$$

$$t = -0.62$$

(vi). Conclusion: Since $t = -0.62$ lies in the interval $-2.262 < t < 2.262$, we accept H_o.

i.e., we may conclude that the data are consistent with the assumption of mean I.Q of 100 in the population.

95% confidence limits are given by

$\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$

$= 97.2 \pm 2.26(\frac{13.5}{\sqrt{10}})$

$= 106.85$ & 87.55 Hence, 95% confidence limits within which the mean I.Q values of samples of 10 boys will lie is (87.55, 106.85)

T-distribution for difference of means

Assumption:

(i). Parent populations, from which the samples have been drawn are normally distributed.

(ii). Population variances are equal and unknown.

(iii). The two samples are random and independent.

Note:

If s_1, s_2 are the standard deviations and $\dot{x}_1$,$\dot{x}_2$ are the means of two samples n_1, n_2 then the statistic t to be tested is

$$t = \frac{\dot{x}_1 - \dot{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where, $S^2 = \frac{n_1 s_1{}^2 + n_2 s_2{}^2}{n_1+n_2-2}$

Problems

1). The following are the number of sales which a sample of 9 salespeople of industrial chemicals in Gujarath and a sample of 6 salespeople of industrial chemicals in Maharashtra made over a certain fixed period of time:

Gujarath	59	68	44	71	63	46	69	54	48
Maharashtra	50	36	62	52	70	41			

Assuming that the population samples can be approximated closely with normal distribution having the same variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 \neq 0$ at 0.01 level of significance.

Solution:

Given that $n_1 = 9, n_2 = 6$

$$\dot{x}_1 = \frac{\Sigma x_1}{n} = \frac{522}{9} = 58$$

$$\dot{x}_2 = \frac{\Sigma x_2}{n} = \frac{311}{6} = 51.83$$

$$s_1{}^2 = \frac{\Sigma x_1{}^2}{n} - \dot{x}_1{}^2 = \frac{31149}{9} - 58^2 = 96.89$$

$$s_2{}^2 = \frac{\Sigma x_2{}^2}{n} - \dot{x}_2{}^2 = \frac{16925}{6} - 51.83^2 = 134.48$$

Now,

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{9(96.89) + 6(134.48)}{9+6-2} = 129.15$$

$$S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 129.15\left(\frac{1}{9} + \frac{1}{6}\right) = 35.88$$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 \neq \mu_2$

(iii). $\alpha = 1\%$, $d.f = n_1 + n_2 - 2 = 9 + 6 - 2 = 13$

(iv). Acceptance Region: $-3.012 < t < 3.012$

(v). The test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{58 - 51.83}{\sqrt{35.88}}$$

$$t = 1.03$$

(vi). Conclusion:

Since t = 1.03 lies in the interval $-3.012 < t < 3.012$, we accept H_o

2). The following table gives the values of protein from Kangeyan cow's milk and buffalo's milk. Examine if these differences are significant

Cow's Milk	1.9	1.95	2	2.02	1.85	1.8
Buffalo's Milk	2.12	2	2.2	2.45	2.2	2.1

Solution:

Given that $n_1 = 6, n_2 = 6$

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{11.52}{6} = 1.92$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{13.07}{6} = 2.178$$

$$s_1^2 = \frac{\Sigma x_1^2}{n} - \bar{x}_1^2 = \frac{22.1554}{6} - 1.92^2 = 0.00616$$

$$s_2^2 = \frac{\Sigma x_2^2}{n} - \bar{x}_2^2 = \frac{25.5869}{6} - 2.17^2 = 0.01936$$

Now,

$$S^2 = \frac{n_1 s_1{}^2 + n_2 s_2{}^2}{n_1+n_2-2} = \frac{6(0.00616)+6(0.01936)}{6+6-2} = 0.0153$$

$$S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = 0.0153\left(\frac{1}{6}+\frac{1}{6}\right) = 0.0051$$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 \neq \mu_2$

(iii). $\alpha = 5\%,\ d.f = n_1 + n_2 - 2 = 6 + 6 - 2 = 10$

(iv). Acceptance Region: $-2.228 < t < 2.228$

(v). The test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$t = \frac{1.92-2.18}{\sqrt{0.0051}}$$

$$t = -3.641$$

(vi). Conclusion:

Since t = -3.641 does not lie in the interval $-2.228 < t < 2.228$, we reject H_o.

3). Two horses A and B were tested according to the time(in seconds)

To run a particular race with the following results:

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether Horse A is running faster than Horse B at 5% level.

Solution:

Given that $n_1 = 7, n_2 = 6$

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{219}{7} = 31.29$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{169}{6} = 28.17$$

$$s_1{}^2 = \frac{\Sigma x_1{}^2}{n} - \bar{x}_1{}^2 = \frac{6883}{7} - 31.29^2 = 4.23$$

$$s_2{}^2 = \frac{\Sigma x_2{}^2}{n} - \bar{x}_2{}^2 = \frac{4787}{6} - 28.17^2 = 4.28$$

Now,

$$S^2 = \frac{n_1 s_1^{\ 2} + n_2 s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{7(4.23)+6(4.28)}{7+6-2} = 5.03$$

$$S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 5.03\left(\frac{1}{7} + \frac{1}{6}\right) = 1.5569$$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 > \mu_2$

(iii). $\alpha = 5\%\ d.f = n_1 + n_2 - 2 = 7 + 6 - 2 = 11$

(iv). Acceptance Region: $t < 1.796$

(v). The test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{31.29 - 28.17}{\sqrt{1.5569}}$$

$$t = 2.498$$

(vi). Conclusion:

Since t $= 2.498$ does not satisfy $t < 1.796$, we reject H_o.

4). A group of 10 rats fed on diet A and another group of 8 rats fed on diet B, recorded the following increase in weight(gms)

Diet A: 5, 6, 8, 1, 12, 4, 3, 9, 6, 10

Diet B: 2, 3, 6, 8, 10, 1, 2, 8

Does it show the inferiority of Diet A over Diet B?

Solution:

Given that $n_1 = 10, n_2 = 8$

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{64}{10} = 6.4$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{40}{8} = 5$$

$$s_1^{\ 2} = \frac{\Sigma x_1^{\ 2}}{n} - \bar{x}_1^{\ 2} = \frac{512}{10} - 6.4^2 = 10.24$$

$$s_2^{\ 2} = \frac{\Sigma x_2^{\ 2}}{n} - \bar{x}_2^{\ 2} = \frac{282}{8} - 5^2 = 10.25$$

R.M.K GROUP OF INSTITUTIONS

Now,

$$S^2 = \frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1+n_2-2} = \frac{10(10.24)+8(10.25)}{10+8-2} = 11.525$$

$$S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = 11.525\left(\frac{1}{10}+\frac{1}{8}\right) = 2.5931$$

(i). $H_o: \mu_1 = \mu_2$

(ii). $H_1: \mu_1 < \mu_2$

(iii). $\alpha = 5\%$ $d.f = n_1 + n_2 - 2 = 10 + 8 - 2 = 16$

(iv). Acceptance Region: $t > -1.746$

(v). The test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$t = \frac{6.4-5}{\sqrt{2.5931}}$$

$$t = 0.869$$

(vi). Conclusion:

Since t = 0.869 satisfies $t > -1.746$, we accept H_o.

R.M.K GROUP OF INSTITUTIONS

Test Based on χ^2 Distribution

(A). χ^2 Test based on Population Variance:

Let $x_1, x_2, \ldots x_n$ be a random sample from a normal population with variance σ^2. Set the null hypothesis H_0: $\sigma^2 = \sigma_0{}^2$. Then the test statistic is $\chi^2 = \frac{ns^2}{\sigma_0{}^2}$ where s^2 is the variance of the sample is said to follow with $n-1$ degrees of freedom.

Note:

(i). If calculated $\chi^2 <$ table χ^2 , then we accept H_0. Otherwise reject H_0.

(ii). If the sample size $n > 30$, we can follow Fisher's approximation $Z = \sqrt{2\chi^2} - \sqrt{2n-1}$ and if $|Z| < 3$ for any level of significance , we can accept H_0. Otherwise reject H_0

Problems:

1). A random sample of size 25 from a population gives the sample standard deviation 8.5. Test the hypothesis that the population standard deviation is 10

Solution:

Given that $n = 25, s = 8.5, \ \sigma = 10$

(i). H_o: $\sigma = 10$

(ii). H_1: $\sigma \neq 10$

(iii). $\alpha = 5\%$ $d.f = n - 1 = 25 - 1 = 24$

(iv). Table Value of $\chi^2 = 36.415$

(v). Test Statistic:

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

$$\chi^2 = \frac{25(8.5^2)}{10^2}$$

$$\chi^2 = 18.06$$

(vi). Conclusion:

Since $18.06 < 36.415$, we accept H_o.

2). It is believed that the precision(as measured by variance) of an instrument is no 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

Solution:

Given that $n = 11, \quad \sigma^2 = 0.16$

$\bar{x} = \frac{\Sigma x}{n} = \frac{27.6}{11} = 2.51$

$s^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{0.1891}{11} = 0.0172$

(i). $H_o: \sigma^2 = 0.16$

(ii). $H_1: \sigma^2 \neq 0.16$

(iii). $\alpha = 1\% \; d.f = n - 1 = 11 - 1 = 10$

(iv). Table Value of $\chi^2 = 23.2$

(v). Test Statistic:

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

$$\chi^2 = \frac{11(0.0172)}{0.16}$$

$$\chi^2 = 1.182$$

(vi). Conclusion:

Since $1.182 < 23.2$, we accept H_o.

We conclude that the data are consistent with the hypothesis that the precision of the instrument is 0.16.

3). Test the hypothesis that $\sigma = 10$, given that s=15 for a random sample of size 50 from a normal population.

Solution:

Given that $n = 50, \; s = 15, \; \sigma = 10$

(i). H_o: $\sigma = 10$

(ii). $H_1: \sigma \neq 10$

(iii). α is not given, choose $\alpha = 5\%$ and $\alpha = 1\%$

(v). Test Statistic:

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

$$\chi^2 = \frac{50(15^2)}{100}$$

$$\chi^2 = 112.5$$

Since $n = 50(> 30)$, we use the test statistic $Z = \sqrt{2\chi^2} - \sqrt{2n-1}$

$$Z = \sqrt{2(112.5)} - \sqrt{2(50) - 1}$$
$$Z = 5.05$$

(vi). Conclusion:

Since $Z = 5.05$ does not lie in the interval $|Z| < 3$, we reject H_o. Hence it is significant at all levels of significance. Hence, we conclude that $\sigma \neq 10$.

(B). χ^2 Test for Goodness of fit:

χ^2 Test for Goodness of fit is a test to find if the deviation of the experiment from theory is just by chance or it is due to inadequacy of the theory to fit the observed data.

By this test, we test whether differences between observed and expected frequencies are significant or not.

In χ^2 Test, the statistic of goodness of fit is given by

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

Where, O denotes Observed frequency

E denotes Expected frequency

Applications of χ^2 distribution:

(i). To test the "goodness of fit"

(ii). To test the "independence of attributes"

(iii). To test if the hypothetical value of the population variance is σ^2.

(iv). To test the homogeneity of independent estimates of the population variance.

(v). To test the homogeneity of independent estimates of the population correlation coefficient.

Conditions for the application of χ^2 Test:

(i). The sample observations should be independent.

(ii). Constraints on the cell frequencies, if any must be linear. i.e., $\sum O_i = \sum E_i$

(iii). N, the total frequency should be atleast 50

(iv). No theoretical cell frequency should be less than 5

R.M.K GROUP OF INSTITUTIONS

Note:

(i). In the case of

fitting a Binomial distribution , $d.f = n - 1$

fitting a Poisson distribution , $d.f = n - 2$

fitting a Normal distribution , $d.f = n - 3$

(ii). If $\chi^2 = 0$, all observed and expected frequencies coincide.

(iii). For, χ^2 distribution, $mean = v, variance = 2v$.

(iii). If calculated $\chi^2 <$ table χ^2 , then we accept H_0. Otherwise reject H_0.

Problems:

1). Four coins are tossed 160 times and the following results were obtained

No. of Heads	0	1	2	3	4
Observed frequencies	17	52	54	31	6

Under the assumptions that coins are unbiased, find the expected frequencies of getting 0, 1, 2, 3, 4 heads and test the goodness of fit.

Solution:

Given that n= the number of data=5, N =the total number of frequencies=160

(i). H_0: The coins are unbiased

(ii). H_1: The coins are biased

(iii). $\alpha = 5\% d.f = n - 1 = 5 - 1 = 4$

(iv). Table value of $\chi^2 = 9.488$

(v). The test statistic is

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

Probability of getting head = $p = 1/2$

Probability of getting tail = $q = 1/2$

Then the expected frequencies are $p(x) = nCxp^x q^{n-x}$ $\quad x = 0, 1, 2, 3, 4$

$$p(0\ head) = 0.0625$$
$$p(1\ head) = 0.25$$
$$p(2\ \ head) = 0.375$$

R.M.K GROUP OF INSTITUTIONS

$p(3\ head) = 0.25$

$p(4\ head) = 0.0625$

$\sum O_i = 160$

χ^2 value is calculated from the following table

No. of heads(x_i)	O	$p(x_i)$	$E = N\,p(x_i)$	$\frac{(O-E)^2}{E}$
0	17	0.0625	10	4.9
1	52	0.25	40	3.6
2	54	0.375	60	0.6
3	31	0.25	40	2.025
4	6	0.0625	10	1.6
	$\sum O_i = 160$		$\sum E_i = 160$	$\chi^2 = 12.725$

(vi). Conclusion:

Since 12.725 is not less than 9.488 we reject H_0.

Hence the coins are biased.

2). The table below gives the number of accidents that occurred during the various day of the week. Test whether the accidents are uniformly distributed over the week.

Days	Mon	Tue	Wed	Thu	Fri	Sat
No. of accidents	14	18	12	11	15	14

Solution:

The parameter of interest is to test the accidents are uniformly distributed.

Given that n= the number of data=6, N =the total number of accidents=84

(i). H_0: The accidents are uniformly distributed over the week.

(ii). H_1: The accidents are not uniformly distributed.

(iii). $\alpha = 5\%$ $d.f = n - 1 = 6 - 1 = 5$

(iv). Table value of $\chi^2 = 11.070$

(v). The test statistic is

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

On the assumption of H_0, the expected number of accidents on any day is 84/6=14

χ^2 value is calculated from the following table

Days	$\boldsymbol{O}$	$\boldsymbol{E}$	$\boldsymbol{O - E}$	$\boldsymbol{\frac{(O-E)^2}{E}}$
Monday	14	14	0	0
Tuesday	18	14	4	1.143
Wednesday	12	14	-2	0.286
Thursday	11	14	-3	0.643
Friday	15	14	1	0.071
Saturday	14	14	0	0
	$\sum O_i = 84$	$\sum E_i = 84$		$\chi^2 = 2.143$

(vi). Conclusion:

Since $2.143 < 11.070$, we accept H_0.

Hence we conclude that the accidents are uniformly distributed over the week.

7). In 120 throws of a single die, the following distributions of faces was observed.

Face	1	2	3	4	5	6
Frequency	30	25	18	10	22	15

Can you say that the die is biased?

Solution:

Given that n= the number of data=6, N =the total number of frequencies=120

(i). H_0: The coins are unbiased

(ii). H_1: The coins are biased

(iii). $\alpha = 5\%\ d.f = n - 1 = 6 - 1 = 5$

(iv). Table value of $\chi^2 = 11.07$

(v). The test statistic is

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

On the assumption of H_0, the expected number of frequencies for each face is 120/6=20

χ^2 value is calculated from the following table

Face	O	E	$O-E$	$\frac{(O-E)^2}{E}$
1	30	20	10	5
2	25	20	5	1.25
3	18	20	-2	0.20
4	10	20	-10	5
5	22	20	2	0.20
6	15	20	-5	1.25
	$\sum O_i = 120$	$\sum E_i = 120$		$\chi^2 = 12.9$

(vi). Conclusion:

Since 12.9 is not less than 11.07 we reject H_0.

Hence the die is biased.

4). A sample analysis of examination results of 500 students was made. It was found that 220 students was failed, 170 have secured a third class, 90 have secured a second class and the rest a first class. So do these figures support the general belief that the above categories are in the ratio 4:3:2:1 respectively?

Solution:

Given that $n = 4$, N =Total number of students=500

(i). H_0: The results in the four categories are in the ratio 4:3:2:1

(ii). H_1: The results in the four categories are not in the ratio 4:3:2:1

(iii). $\alpha = 5\%$ $d.f = n - 1 = 4 - 1$ =3

(iv). Table value of $\chi^2 = 7.815$

(v). The test statistic is

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

On the assumption of H_0, the expected frequencies of the four categories are (4/10)500, (3/10)500, (2/10)500, (1/10)500

i.e., 200, 150, 100, 50

χ^2 value is calculated from the following table

Categories	O	E	$O - E$	$\frac{(O-E)^2}{E}$
Failures	220	200	20	2
III Class	170	150	20	2.667
II Class	90	100	-10	1
I Class	20	50	-30	18
	$\sum O_i = 500$	$\sum E_i = 500$		$\chi^2 = 23.667$

R.M.K GROUP OF INSTITUTIONS

(vi). Conclusion:

Since 23.667 is not less than 7.815 we reject H_0.

Hence we conclude that the results in the four categories are not in the ratio 4:3:2:1

5). The following is the distribution of the hourly number of trucks arriving at a company's warehouse

Trucks arriving per hour	0	1	2	3	4	5	6	7	8
Frequency	52	151	130	102	45	12	5	1	2

Find the mean of the distribution and using its mean(rounded tom one decimal) as the parameter λ fit a Poisson distribution. Test for the goodness of fit at 5% level of significance.

Solution:

The parameter of interest is λ(mean).

Given that n =the number of data=7, $N = 500$ =Total frequencies.

[Clubbing the last three frequencies to single one.]

(i). H_0: Poisson fit is a good fit

(ii). H_1: Poisson fit is not a good fit

(iii). $\alpha = 5\%$ $d.f = n - 2 = 7 - 2 = 5$

(iv). Table value of $\chi^2 = 11.07$

(v). The test statistic is

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

Mean of the distribution is $\lambda = \frac{\sum f_i x_i}{\sum f_i} = \frac{1040}{500} = 2.1$

By Poisson distribution the frequency of r success is

$$N(r) = \frac{Ne^{-\lambda}\lambda^r}{r!} \qquad r = 0, 1, 2, 3, \ldots, 8$$

$$N(r) = \frac{500\, e^{-2.1} 2.1^r}{r!} \qquad r = 0, 1, 2, 3, \ldots, 8$$

$N(0) = 61$

$N(1) = 129$

$N(2) = 135$

$N(3) = 94$

$N(4) = 50$

$N(5) = 21$

$N(6) = 7$

$N(7) = 2$

$N(8) = 1$

$\boldsymbol{O}$	$\boldsymbol{E}$	$\boldsymbol{O-E}$	$\boldsymbol{\frac{(O-E)^2}{E}}$
52	61	-9	1.33
151	129	22	3.75
130	135	-5	0.185
102	94	8	0.68
45	50	-5	0.5
12	21	-9	3.86
8	10	-2	0.4

$\chi^2 = 10.705$

(vi). Conclusion:

Since $10.705 < 11.070$, we accept H_0.

Hence, we conclude that Poisson fit is a good fit.

(C). χ^2 Test to test the independence of attributes:

Let us consider two attributes A and B. A divided into r classes $A_1, A_2, \ldots, A_r$ and B divided into s classes $B_1, B_2, \ldots, B_s$. If this is expressed as a rxs matrix, the matrix is called a rxs contingency table.

Note:

(i). Expected frequency of each cell = $\frac{Product\ of\ corresponding\ row\ total\ \ and\ coloumn\ total}{Grand\ total}$

(ii). In χ^2 Test, the statistic value is calculated by

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

Where, O denotes Observed frequency

E denotes Expected frequency

(iii). If calculated χ^2 < table χ^2 , then we accept H_0. Otherwise reject H_0.

(iv). For a 2 x 2 contingency table

a	b
c	d

The statistic is given by

$$\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

(v). $d.f = (r-1)(s-1)$

Problems:

1). Two researchers A and B adopted different techniques while rating the students level. Can you say that the techniques adopted by them are significant?

Researchers	Below average	Average	Above average	Genius	Total
A	40	33	25	2	100
B	86	60	44	10	200
Total	126	93	69	12	300

Solution:

The parameter of interest is χ^2.

(i). H_0: There is no significant difference between the techniques of two researchers

(ii). H_1: There is a significant difference between the techniques of two researchers

(iii). $\alpha = 5\%$ $d.f = (r-1)(s-1) = (2-1)(4-1) = 3$

(iv). Table value of $\chi^2 = 7.815$

(v). The test statistic is

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

Expected frequency of each cell = $\frac{Product\ of\ corresponding\ row\ total\ and\ coloumn\ total}{Grand\ total}$

Expected frequency for 40 = $\frac{(100)\ (126)}{300} = 42$

Expected frequency for 33 = $\frac{(100)\ (93)}{300} = 31$

Expected frequency for 25 = $\frac{(100)\ (69)}{300} = 23$

Expected frequency for 2 = $\frac{(100)\ (12)}{300} = 4$

Expected frequency for 86 = $\frac{(200)\ (126)}{300} = 84$

Expected frequency for 60 = $\frac{(200)\ (93)}{300} = 62$

Expected frequency for 44 = $\frac{(200)\ (69)}{300} = 46$

Expected frequency for 10 = $\frac{(200)\ (12)}{300} = 8$

χ^2 value is calculated from the following table

O	E	$O-E$	$\frac{(O-E)^2}{E}$
40	42	-2	0.0952
33	31	2	0.1290
25	23	2	0.1739
2	4	-2	1
86	84	2	0.0476
60	62	-2	0.0645
44	46	-2	0.0869
10	8	2	0.5

$$\chi^2 = 2.0971$$

(vi). Conclusion:

Since $2.0971 < 7.815$, we accept H_0.

Hence we conclude that there is no significant difference between the techniques of two researchers.

2). Theory predicts that the proportion of beans in four groups A, B, C, D should be 9:3:3:1. In an experiment among 1600 beans the numbers in four groups were 882, 313, 287, 118. Does the experiment support the theory

Solution:

Given that $n = 4$, N =Total number of students=1600

(i). H_0: The experiment supports the theory

(ii). H_1: The experiment does not support the theory.

(iii). $\alpha = 5\%$ $d.f = n - 1 = 4 - 1 = 3$

(iv). Table value of $\chi^2 = 7.815$

(v). The test statistic is

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

On the assumption of H_0, the expected frequencies of the four categories are (9/16)1600, (3/16)1600, (3/16)1600, (1/16)1600

i.e., 900, 300, 300, 100

χ^2 value is calculated from the following table

0	E	$O - E$	$\frac{(O-E)^2}{E}$
882	900	-18	0.36
313	300	13	0.1878
287	300	-13	0.1878
118	100	18	0.36

$$\chi^2 = 4.73$$

(vi). Conclusion:

Since $4.73 < 7.815$, we accept H_0

Hence we conclude that the experiment supports the theory.

3). Two sample polls of votes for two candidates A and B for a public office are taken one from among residents of rural and urban areas. The results are given below. Examine whether the nature of the area is related to voting preference in this election.

Area/Votes for	A	B	Total
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

R.M.K GROUP OF INSTITUTIONS

Solution:

The parameter of interest is χ^2

(i). H_0: The nature area is independent of voting preference in the election.

(ii). H_1: The nature area is dependent of voting preference in the election.

(iii). $\alpha = 5\%\ d.f = (r-1)(s-1) = (2-1)(2-1) = 1$

(iv). Table value of $\chi^2 = 5.991$

(v). The test statistic is

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

Expected frequency of each cell = $\frac{Product\ of\ corresponding\ row\ total\ and\ coloumn\ total}{Grand\ total}$

Expected frequency for 620 = $\frac{(1000)\ (1170)}{2000} = 585$

Expected frequency for 380 = $\frac{(1000)\ (830)}{2000} = 415$

Expected frequency for 550 = $\frac{(1000)\ (1170)}{2000} = 585$

Expected frequency for 450 = $\frac{(1000)\ (830)}{2000} = 415$

χ^2 value is calculated from the following table

O	E	$O-E$	$\frac{(O-E)^2}{E}$
620	585	35	2
380	415	-35	3
550	585	-35	2
450	415	35	3

$\chi^2 = 10$

(vi). Conclusion:

Since 10 is not less than 5.991, we reject H_0.

Hence, we conclude that The nature area is dependent of voting preference in the election.

. 1000 students at college were graded according to their I.Q. and their economic conditions. What conclusion can you draw from the following data?

Economy/I.Q. Level	High	Low	Total
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

Solution:

The parameter of interest is χ^2

(i). H_0: The given attributes are independent.

(ii). H_1: The given attributes are not independent.

(iii). $\alpha = 5\%$ $d.f = (r-1)(s-1) = (2-1)(2-1) = 1$

(iv). Table value of $\chi^2 = 3.841$

(v). The test statistic is

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

Expected frequency of each cell = $\frac{Product\ of\ corresponding\ row\ total\ and\ coloumn\ total}{Grand\ total}$

Expected frequency for 460 = $\frac{(600)\,(700)}{1000} = 420$

Expected frequency for 140 = $\frac{(600)\,(300)}{1000} = 180$

Expected frequency for 240 = $\frac{(400)\,(700)}{1000} = 280$

Expected frequency for 160 = $\frac{(400)\,(300)}{1000} = 120$

χ^2 value is calculated from the following table

R.M.K GROUP OF INSTITUTIONS

χ^2 value is calculated from the following table

O	E	$O-E$	$\frac{(O-E)^2}{E}$
460	420	40	3.81
140	180	-40	8.88
240	280	-40	5.714
160	120	40	13.33

$$\chi^2 = 31.7373$$

(vi). Conclusion:

Since 31.7373 is not less than 3.841, we reject H_0.

Hence, we conclude that the attributes I.Q. as Economic conditions are not independent.

5). Find if there is any association between extravagance in fathers and extravagance in sons from the following data

	Extravagant father	Miserly father
Extravagant son	327	741
Miserly son	545	234

Determine the coefficient of association also.

Solution:

The parameter of interest is χ^2

(i). H_0: Namely that the extravagance in sons and fathers are not significant.

(ii). H_1: Namely that the extravagance in sons and fathers are significant.

(iii). $\alpha = 5\%$ $d.f = (r-1)(s-1) = (2-1)(2-1) = 1$

(iv). Table value of $\chi^2 = 3.841$

(v). The test statistic is

$$\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$a = 327, b = 741, c = 545, d = 234$

$$\chi^2 = \frac{(327+741+545+234)[(327)(234)-(545)(741)]^2}{(1068)(779)(872)(975)}$$

$$\chi^2 = 230.24$$

(vi). Conclusion:

Since, 230.24 is not less than 3.841 we reject H_0.

Hence, we conclude that the extravagance in sons and fathers are significant.

(vii). Coefficient of attributes$=\frac{ad-bc}{ad+bc} = \frac{-327330}{480363} = -0.6814$

F-distribution [Test for Variance]

Suppose that two independent normal populations are of interest, when the population means and variances say μ_1, μ_2 and σ_1^2, σ_2^2 are unknown. We wish to test the hypothesis about the equality of the two variances, say $H_0: \sigma_1^2 = \sigma_2^2$. Assume that two random samples of size n_1 from population 1 and of size n_2 from population 2 are available and let s_1^2, s_2^2 be the sample variances. We wish to test the hypothesis

(i). $H_o: \sigma_1^2 = \sigma_2^2$

(ii). $H_1: \sigma_1^2 \neq \sigma_2^2$

The development of a test procedure for these hypothesis requires a new probability distribution, the F distribution.

The test statistic is

$F = \frac{S_1^2}{S_2^2}$ if $S_1^2 > S_2^2$ and

$F = \frac{S_2^2}{S_1^2}$ if $S_2^2 > S_1^2$

Where $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$ with $d.f$ $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$

If calculated F is less than table F, we accept H_o otherwise reject H_o.

Applications of F Test:

F Test is used to test (i). Whether two independent samples have been drawn from the normal populations with the same variance σ^2, or (ii). Whether two independent estimates of the population variance are homogeneous or not.

Properties of the F distribution:

(i). The square of the t variate with n degrees of freedom follows a F distribution with l and n degrees of freedom.

(ii). The mean of the F distribution is $\frac{v_2}{v_2 - 2}$ if $v_2 > 2$

(iii). The variance of the F distribution is $\frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}$ if $v_2 > 4$

Problems:

1). A group of 10 rats fed on diet A and another group of 8 rats fed on diet B, recorded the following increase in weight(gms)

Diet A: 5, 6, 8, 1, 12, 4, 3, 9, 6, 10

Diet B: 2, 3, 6, 8, 10, 1, 2, 8

Find if the variances are significantly different.

Solution:

Given that $n_1 = 10, n_2 = 8$

$\dot{x}_1 = \frac{\Sigma x_1}{n} = \frac{64}{10} = 6.4$

$\dot{x}_2 = \frac{\Sigma x_2}{n} = \frac{40}{8} = 5$

$s_1{}^2 = \frac{\Sigma x_1{}^2}{n} - \dot{x}_1{}^2 = \frac{512}{10} - 6.4^2 = 10.24$

$s_2{}^2 = \frac{\Sigma x_2{}^2}{n} - \dot{x}_2{}^2 = \frac{282}{8} - 5^2 = 10.25$

$S_1{}^2 = \frac{n_1 s_1{}^2}{n_1 - 1} = \frac{10(10.24)}{9} = 11.3777$

$S_2{}^2 = \frac{n_2 s_2{}^2}{n_2 - 1} = \frac{8(10.25)}{7} = 11.7143$

We have $S_2{}^2 > S_1{}^2$

The parameter of the interest is $\sigma_1{}^2$ and $\sigma_2{}^2$

(i). $H_o: \sigma_1{}^2 = \sigma_2{}^2$ [The difference of variance is not significant]

(ii). $H_1: \sigma_1{}^2 \neq \sigma_2{}^2$

(iii). $\alpha = 5\% d.f \;\; v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ i.e., $v_1 = 9$ and $v_2 = 7$

(iv). Table value of $F = 3.29$

(v). The test statistic is

$F = \frac{S_2{}^2}{S_1{}^2} = \frac{11.7143}{11.3777} = 1.02958$

(vi). Conclusion:

Since 1.02958 < 3.29, we accept H_o.

Hence we conclude that the two samples have come from populations with equal variances.

2). Two independent samples of sizes 9 and 7 from a normal population had the following values of the variables.

Sample I: 18, 13, 12, 15, 12, 14, 16, 14, 15

Sample II: 16, 19, 13, 16, 18, 13, 15

Do the estimates of the population variance differ significantly at 5% level?

Solution:

Given that $n_1 = 9, n_2 = 7$

$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{129}{9} = 14.3333$

$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{110}{7} = 15.7143$

$s_1{}^2 = \frac{\Sigma x_1{}^2}{n} - \bar{x}_1{}^2 = \frac{1879}{9} - 14.3333^2 = 3.3342$

$s_2{}^2 = \frac{\Sigma x_2{}^2}{n} - \bar{x}_2{}^2 = \frac{1760}{7} - 15.7143^2 = 4.4894$

$S_1{}^2 = \frac{n_1 s_1{}^2}{n_1 - 1} = \frac{9(3.3342)}{8} = 3.751$

$S_2{}^2 = \frac{n_2 s_2{}^2}{n_2 - 1} = \frac{7(4.4894)}{6} = 5.2376$

We have $S_2{}^2 > S_1{}^2$

The parameter of the interest is $\sigma_1{}^2$ and $\sigma_2{}^2$

(i). $H_o: \sigma_1{}^2 = \sigma_2{}^2$ [The difference of variance is not significant]

(ii). $H_1: \sigma_1{}^2 \neq \sigma_2{}^2$

(iii). $\alpha = 5\% d.f\ \ v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ i.e., v_1 =8 and $v_2 = 6$

(iv). Table value of $F = 3.58$

(v). The test statistic is

$F = \frac{S_2{}^2}{S_1{}^2} = \frac{5.2376}{3.751} = 1.3963$

(vi). Conclusion:

Since 1.3963 < 3.58, we accept H_o.

Hence we conclude that the differences is not significant.

3). In one sample of 10 observations , the sum of the squares of the deviations of the sample values from the sample mean was 120 and in another sample of 12 observations it was 314. Test whether the differences is significant at 5% level of significance.

Solution:

Given that $n_1 = 10,\ n_2 = 12$

R.M.K GROUP OF INSTITUTIONS

$\sum(x_1-\overline{x_1})^2 = 120$ and $\sum(x_2-\overline{x_2})^2 = 314$

$$s_1^{\,2} = \frac{\sum(x_1-\overline{x_1})^2=120}{n_1} = \frac{120}{10} = 12$$

$$s_2^{\,2} = \frac{\sum(x_2-\overline{x_2})^2=314}{n_2} = \frac{314}{12} = 26.1667$$

$$S_1^{\,2} = \frac{n_1 s_1^{\,2}}{n_1-1} = \frac{10(12)}{9} = 13.33$$

$$S_2^{\,2} = \frac{n_2 s_2^{\,2}}{n_2-1} = \frac{12(26.1667)}{11} = 28.55$$

We have $S_2^{\,2} > S_1^{\,2}$

The parameter of the interest is $\sigma_1^{\,2}$ and $\sigma_2^{\,2}$

(i). $H_o: \sigma_1^{\,2} = \sigma_2^{\,2}$ [The difference of variance is not significant]

(ii). $H_1: \sigma_1^{\,2} \neq \sigma_2^{\,2}$

(iii). $\alpha = 5\% d.f\ \ v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ i.e., $v_1 = 9$ and $v_2 = 11$

(iv). Table value of $F = 3.11$

(v). The test statistic is

$$F = \frac{S_2^{\,2}}{S_1^{\,2}} = \frac{28.55}{13.33} = 2.14$$

(vi). Conclusion:

Since 2.14 < 3.11, we accept H_o.

Hence we conclude that the differences is not significant.

1. A sociologist focusing on popular culture and media believes that the average number of hours per week (hrs/week) spent using social media is greater for women than for men. Examining two independent simple random samples of 100 individuals each, the researcher calculates sample standard deviations of 2.3 hrs/week and 2.5 hrs/week for women and men respectively. If the average number of hrs/week spent using social media for the sample of women is 1 hour greater than that for the sample of men, what conclusion can be madefrom a hypothesis testwhere:

$$H_0: \mu W - \mu M = 0$$
$$H_1: \mu W - \mu M > 0$$

   a. The observed difference in average number of hrs/week spent using social media is notsignificant
   b. The observed difference in average number of hrs/week spent using social media is significant
   c. A conclusion is not possible without knowing the average number of hrs/week spent using social media in eachsample
   d. A conclusion is not possible without knowing the populationsizes

2. A 99% t_based confidence interval for the mean price for a gallon of gasoline (dollars) is calculated using a simple random sample of gallon gasoline prices for 50 gas stations.Given that the 99% confidence interval is $\$3.32 < \mu < \$3.98$, what is the sample mean price for a gallon of gasoline(dollars)?
Please select the best answer of those provided below.

   a. $\$0.33$
   b. $\$3.65$
   c. Not Enough Information; we would need to know the variation in the sampleof gallon gasoline prices
   d. Not Enough Information; we would need to know the variation in the populationof llon gasoline prices.

3. The statement "If there is sufficient evidence to reject a null hypothesis at the 10% significance level, then there is sufficient evidence to reject it at the 5% significance level"is: Please select the best answer of those providedbelow.
   a. AlwaysTrue
   b. NeverTrue
   c. Sometimes True; the p-value for the statistical test needs to be provided for a

conclusion

d. Not Enough Information; this would depend on the type of statistical test used

4. Green sea turtles have normally distributed weights, measured in kilograms, with a mean of 134.5 and a variance of 49.0. A particular green sea turtle's weight has a $z\ score\ of\ -2.4$. What is the weight of this green sea turtle? Round to the nearest whole number.
   a. $17\ kg$
   b. $151\ kg$
   c. $118\ kg$
   d. $252\ kg$

5. What percentage of measurements in a dataset fall above the median?
   a. 49%
   b. 50%
   c. 51%
   d. Cannot Be Determined

6. What is the mean of a Chi Square distribution with 6 degrees of freedom?
   a) 4
   b) 12
   c) 6
   d) 8

7. A bag contains 80 chocolates. This bag has 4 different colors of chocolates in it. If all four colors of chocolates were equally likely to be put in the bag, what would be the expected number of chocolates of each color?
   a) 12
   b) 11
   c) 20
   d) 9

8. Find Variance for an $F\_Distribution$ with $v_1 = 5$ and $v_2 = 9$. Where v_1 and v_2 are the degrees of freedoms
   a) 1.587
   b) 1.378
   c) 1.578
   d) 1.498

9. Calculate the value of $f\_statistic$ having a cumulative probability of 0.95.
a) 0.55
b) 0.5
c) 0.05
d) 0.05

10. Find the Expectation for a $F\_\ Distribution$ variable with $v_1 = 7$ and $v_2 = 8$. Where v_1 and v_2 are the degrees of freedoms

a) $4/7$
b) $4/6$
c) $4/3$
d) $4/5$

11. A t_test sample has 7 pairs of samples. The distribution should contain ____________
a) 16 degrees of freedom
b) 15 degrees of freedom
c) 5 degrees of freedom
d) 6 degrees of freedom

12. Which of the following p_values will lead us to reject the null hypothesis if the significance level of the test is 5%
a) 0.10
b) 0.15
c) 0.20
d) 0.025

13. Suppose that we reject a null hypothesis at 5% level of significance. For which of the following level of significance do we also reject the null hypothesis?
a) 6%
b) 4%
c) 3%
d) 2%

14. A hypothesis test is conducted to test whether the mean age of clients at a certain health spa is equal to 25 or not. It is known that the population standard deviation of clients at the spa is 10.36 clients are randomly selected, and their ages recorded, with the sample mean age being 22.8. What is your decision, at the 5% level of significance, regarding the null hypothesis that the mean age is equal to 25?

a). reject the null hypothesis at the 5% level of significance and conclude that the mean age of clients at the spa is less than 25

R.M.K GROUP OF INSTITUTIONS

b). reject the null hypothesis at the 5% level of significance and conclude that the

mean age of clients at the spa is not equal to 25

c). reject the null hypothesis at the 5% level of significance and conclude that the mean

age of clients at the spa is more than 25

d).do not reject the null hypothesis at the 5% level of significance and conclude that the mean age of clients at the spa is 25

15. A hypothesis test is conducted to test whether the mean age of clients at a certain health spa is equal to 25 or not. It is known that the population standard deviation of clients at the spa is 10. 36 clients are randomly selected, and their ages recorded, with the sample mean age being 20.8. What is your decision, at the 5% level of significance, regarding the null hypothesis that the mean age is equal to 25?

a). reject the null hypothesis at the 5% level of significance and conclude that the mean

age of clients at the spa is less than 25

b). reject the null hypothesis at the 5% level of significance and conclude that the mean

age of clients at the spa is not equal to 25

c). reject the null hypothesis at the 5% level of significance and conclude that the mean

age of clients at the spa is more than 25

d). do not reject the null hypothesis at the 5% level of significance and conclude that the mean age of clients at the spa is 25

16. According to a certain TV broadcast station, the average number of violent incidents shown per episode of a TV series is 7. A researcher believes that this has increased in the last few years. A random sample of 16 recent episodes is selected which produced a sample mean of 7.2 violent incidents. Assume that the number of violent incidents follows a normal distribution and that the population standard deviation is 1.2. What would be the conclusion of a hypothesis test, if we were to perform a hypothesis test at a 5% level of significance in order to test whether the researcher's belief is accurate or not (assume that the null hypothesis states that there is no change in the average number of violent incidents shown per episode)?

a).We cannot reject the null hypothesis and conclude that the mean number of violent incidents per episode has not increased

b). We reject the null hypothesis and conclude that the mean number of violent incidents per episode has indeed increased

c).We reject the null hypothesis since the p-value is greater than 0.05

d).We cannot reject the null hypothesis since the p-value is less than 0.05

17). A social scientist claims that the average adult watches less than 26 hours of television per week. He collects data on 25 individuals' television viewing habits and finds that their mean number of hours watching television was 22.4 hours. Assume the population standard deviation is known to be eight hours, and the significance level adopted is 1%. What is the conclusion based on the data above?

a). Since $z < -z_{0.01}$, we reject the null hypothesis and conclude that the social scientist is right

b). Since $< -z_{0.01}$, , we fail to reject the alternate hypothesis and conclude that the social scientist is right

c). Since $> -z_{0.01}$, we fail to reject the null hypothesis and conclude that the social scientist's claim cannot be proved

d). Since $z < -z_{0.01}$, we fail to reject the null hypothesis and conclude that the social scientist's claim cannot be prove

18). The nine items of a sample had the following values $45, 47, 50, 52, 48, 49, 47, 53, 51$ and the population mean is 47.5. Find the calculated t_value.

a) 1.84

b) 2.81

c) 2.10

d)3.40

19). The recommended daily dietary allowance for zinc among males older than age 50 years is day. A study undertaken on a sample of 115 males aged between 65 and 74 years ports the average daily intake as 11.3 mg with a standard deviation of 6.43 mg. searchers with to test whether the actual average daily zinc intake of males aged tween 65 and 74 years falls below the recommended allowance. What is the value of e test statistic in this case?

a) $t = -6.17$

b) $z = -6.17$

c) $t = -4.50$

d) $z = -4.50$

). The mean life of a battery used in a digital clock is 305 days. The lives of the tteries follow a normal distribution. The battery was recently modified to last longer. sample of 20 of the modified batteries had a mean life of 308 days with a standard viation of 12 days. A hypothesis test is undertaken to determine whether the odification increased the battery life. What is the value of the test statistic for the pothesis test?

a) $z = 2.24$

b) $t = 2.24$

c) $z = 1.12$

d) $t = 1.12$

Answers:

1	2	3	4	5	6	7	8	9	10
b	b	c	c	d	c	c	a	d	c
11	12	13	14	15	16	17	18	19	20
d	d	a	d	b	a	c	a	a	d

R.M.K GROUP OF INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS

ASSIGNMENTS: UNIT III

LEVEL 1

1. The table below gives the number of accidents that took place in an industry during the various days of the week. Test whether the accidents are uniformly distributed over the week.

Days	Mon	Tues	Wed	Thurs	Fri	Sat
No. of accidents	14	18	12	11	15	14

2. 4 coins were tossed at a time and this operation is repeated 160 times. It is found that 4 heads occur 6 times, 3 heads occur 43 times, 2 heads occur 69 times and one head occurs 34 times. Discuss whether the coins may be regarded as unbiased?

3. In the accounting department of a bank, 100 accounts are selected at random and estimated for errors .The following results were obtained.

No. of errors	0	1	2	3	4	5	6
No. of accounts	35	40	19	2	0	2	2

Does this information verify that the errors are distributed according to the Poisson law.

LEVEL 2

1. Among 64 offspring's of a certain cross between guinea pigs, 34 were red, 10 were black and 20 were white. According to the genetic model, these numbers should be in the ratio 9 : 3 : 4. Do the data consistent with the model at 5 % level?

2. Can vaccination be regarded as preventive measure of small pox as evidenced by the following data of 1482 persons exposed to small pox in a locality? 368 in all were attacked of these 1482 persons and 343 were vaccinated and of these only 35 were attacked.
3. Calculate the expected frequencies for the following data presuming the two attributes, condition of home and condition of child are independent

		Condition of home	
		Clean	Dirty
Condition of child	Clean	70	50
	Family Clean	80	20
	Dirty	35	45

Use χ^2 at 5% level of significance to test whether the two attributes are independent.

R.M.K GROUP OF INSTITUTIONS

LEVEL 3

1. A manufacturer claims that only 4 % of his products supplied by him are defective. A random sample of 600 products contained 36 defectives. Test the claim of the manufacturer.
2. A die is thrown 180 times. Getting 3or 4 is considered as success. If the number of successes in any particular experiment is 80, find whether this excess is due to fluctuations of sampling.
3. A certain cubical die was thrown 9000 times and 5 or 6 was obtained 3240 times. Test the hypothesis that the die is unbiased.

LEVEL 4

1. In two large populations there are 30% and 25% respectively are fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?
2. In a referendum submitted to the students' body at a university, 850 men and 560 women voted. 500 men and 320 women voted yes. Does this indicate a significant difference of opinion between men and women on this matter?
3. The mean weight obtained from a random sample of size 100 is 64 gms. The S.D of the weight distribution of the population is 3 gms. Test the statement that the mean weight of the population is 67 gms at 5% level of significance. Also set up 95 confidence limits of the mean weight of the population.

LEVEL 5

1. An examination was given to two classes consisting of 40 and 50 students respectively. In the 1st class the mean mark was 74 with a S.D of 8, while in the 2nd class the mean mark was 78 with a S.D of 7. Is there a significant difference between the performance of the two classes at the levels of 0.05, 0.01 ?
2. A sample of heights of 6400 soldiers has a mean of 67.85 inches and a S.D of 2.56 inches. While another sample of heights of sailors has a mean of 68.55 inches with a S.D of 2.52 inches. Do the data indicate that the sailors are on the average taller than soldiers?
3. Two independent samples of sizes 7 and 6 had the following values:

Sample A	28	30	32	33	31	29	34
Sample B	29	30	30	24	27	28	

Examine whether the samples have been drawn from normal populations having same variance.

PART A QUESTIONS AND ANSWERS

1. **Distinguish between parameters and statistics.**

   Solution : Statistical constants of the population such as mean (μ) and S.D (σ) are usually known as parameters.

   Statistical measures computed from sample observations alone such as mean ($\bar{x}$) and S.D (s) are usually known as statistics.

2. **Explain statistical hypothesis.**

   Solution: Assumptions made about a parameter of a statistical population are called statistical hypothesis.

3. **What are Type I and Type II errors?**

   Solution: Type I Error: Rejecting a null hypothesis when it is true.

   Type II Error: Accepting a null hypothesis when it is wrong.

4. **Critical values in hypothesis testing – Explain.**

   Solution: The value that separates the rejection region from the acceptance region is called the critical value.

5. **What do you mean by a Null Hypothesis?**

   Solution: A definite statement about population parameters is called a Null hypothesis. Null hypothesis is a hypothesis of no difference. It is denoted by H_0.

   A statement which is complementary to the null hypothesis is called an Alternative hypothesis. It is denoted by H_1.

6. **Define the term "Degrees of freedom".**

   Solution: Degrees of freedom is the number of independent observations in a set. If 'n' is the sample size then the degrees of freedom for variance is n-1

7. **Write down the confidence limits for the mean at 1%, 5% levels of large sample.**

   Solution: $\bar{x} \pm 2.58\dfrac{\sigma}{\sqrt{n}}$, $\bar{x} \pm 1.96\dfrac{\sigma}{\sqrt{n}}$

8. **Write down the test statistic for single mean for large sample.**

   Solution: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, where $\bar{x}$ - Sample mean μ - Population mean,

   σ - Population S.D. and n - Sample size.

9. **What is the standard error of the sample proportion when the population is (i) known (ii) not known**

   Solution : (i) $SE = \sqrt{\dfrac{PQ}{n}}$ (ii) $SE = \sqrt{\dfrac{pq}{n}}$

10. **What is the standard error of the difference between two sample proportions when the population proportion is (i) known (ii) not known**

    Solution : (i) $SE = \sqrt{PQ(\dfrac{1}{n1} + \dfrac{1}{n_2})}$

    (ii) If P is not known ,an unbiased estimator of P based on both samples

    Given by $\hat{p} = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

11.A random sample of 400 mangoes was taken from a big consignment and 40 were found to be bad. Prove that the percentage of bad mangoes in the consignment will, in all probability, lie between 5.5. & 14.5.

Solution: Given $p = \frac{40}{400} = 0.1 \Rightarrow q = 0.9$

The extreme limit of confident is $P = p \pm 3\sqrt{\frac{pq}{n}} = 0.1 \pm 3\sqrt{\frac{(0.1)(0.9)}{400}}$ (i.e)

(0.055, 0.145)

The percentage of bad mangoes in the consignment will, in all probability lie between (5.5, 14.5).

12. **The mean value of a random sample of 60 items was found to be 145 with a S.D of 40. Find the 95% confidence limits for the population mean.**

Solution: 95% confidence limits for the population mean is

$\left(\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}\right)$ (i.e) $\left(145 - \frac{(1.96)(40)}{\sqrt{60}}, 145 + \frac{(1.96)(40)}{\sqrt{60}}\right)$ (i.e) $(134.9, 155.1)$.

13.The wages of a factory's workers are assumed to be normally distributed with mean μ and variance Rs. 25. A random sample of 36 workers gives the total wages equal to Rs. 1800. Test the hypothesis $\mu = 52$, against the alternative $\mu = 49$ at 1% level of significance.

Solution: Given $n = 36,\ \bar{x} = \frac{1800}{36} = 50,\ \mu = 52 \ \&\ \sigma = 5.$

Let $H_0 : \mu = 52$ and $H_1 : \mu = 49 < 52$ $[\text{left tailed test}]$

Test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{50 - 52}{5/\sqrt{36}} = -2.4.$ At 1 % level, the table value of z

$\therefore$ calculated value of z > table value of z. Hence H_0 is rejected at 1% level of significance.

$\therefore$ The average $\mu = 49$ is accepted.

14. **State the applications of t – distribution.**

Solution : t - distribution is used

(i) to test if the sample mean $\bar{x}$ differs significantly from the population mean (μ),

(ii) to test the significance of the difference between two sample means,

(iii)to test the significance of an observed sample correlation coefficient and sample regression coefficient.

(iv)to test the significance of observed partial and multiple correlation coefficients.

15.Write the applications of F-test.

Solution : F – test is used to test,

(i) whether the two independent samples have been drawn from the normal populations with the same variance σ^2.

(ii) whether the two independent estimates of population variance differ significantly or not.

16.Write down any two uses of χ^2- test.

Solution: The following are the important applications of χ^2 distribution

i. χ^2 distribution is used to test the goodness of fit.

ii. It is also used to test the independence of attributes.

PART B QUESTIONS

1. A manufacturer claims that only 4 % of his products supplied by him are defective. A random sample of 600 products contained 36 defectives. Test the claim of the manufacturer.
   Ans: Z = 2.5, H_0 rejected.
2. A die is thrown 180 times. Getting 3or 4 is considered as success. If the number of successes in any particular experiment is 80, find whether this excess is due to fluctuations of sampling.
   Ans: Z = 2.58, H_0 rejected.

3. A certain cubical die was thrown 9000 times and 5 or 6 was obtained 3240 times. Test the hypothesis that the die is unbiased.
   Ans: Z = 0.035, **H_0 accepted.**
4. In two large populations there are 30% and 25% respectively are fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?
   Ans: Z = 2.553, **H_0 rejected.**

5. In a referendum submitted to the students' body at a university, 850 men and 560 women voted. 500 men and 320 women voted yes. Does this indicate a significant difference of opinion between men and women on this matter?
   Ans: **H_0 accepted.**
6. 600 articles from a factory are examined and found to be 2% defective. 800 similar articles from another factory are found to have 1.6% defectives. Can it be concluded that the products of the first factory are inferior to those of the second at 5% level.
   Ans : Z=0.67, H_0 accepted.

7. The mean weight obtained from a random sample of size 100 is 64 gms. The S.D of the weight distribution of the population is 3 gms. Test the statement that the mean weight of the population is 67 gms at 5% level of significance. Also set up 95 confidence limits of the mean weight of the population.
   Ans: **H_0 rejected. Limits: 63.22, 64.77.**

8. An examination was given to two classes consisting of 40 and 50 students respectively. In the 1^{st} class the mean mark was 74 with a S.D of 8, while in the 2^{nd} class the mean mark was 78 with a S.D of 7. Is there a significant difference between the performance of the two classes at the levels of 0.05, 0.01 ?
   Ans: **H_0 accepted at 1% and is rejected at 5%.**

R.M.K
GROUP OF
INSTITUTIONS

9.A sample of heights of 6400 soldiers has a mean of 67.85 inches and a S.D of 2.56 inches. While another sample of heights of sailors has a mean of 68.55 inches with a S.D of 2.52 inches. Do the data indicate that the sailors are on the average taller than soldiers?

Ans: H0 rejected.

10.Two independent samples of sizes 7 and 6 had the following values:

Sample A	28	30	32	33	31	29	34
Sample B	29	30	30	24	27	28	

Examine whether the samples have been drawn from normal populations having same variance.

Ans: F = 1.116, H0 accepted.

R.M.K GROUP OF INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS

SUPPORTIVE ONLINE CERTIFICATION COURSES

The following NPTEL and Coursera courses are the supportive online certification courses for the Unit Testing of Hypothesis

https://nptel.ac.in/content/storage2/courses/103106120/LectureNotes/Lec3_4.pdf

https://www.coursera.org/lecture/basic-statistics/7-01-hypotheses-N1Klj

REAL TIME APPLICATIONS

View the lecture on YouTube:

1. https://youtu.be/5k2MPALfmz8

2. https://youtu.be/kx-pcQAPvoc

CONTENT BEYOND THE SYLLABUS

Basic Ideas for the unit: Testing of Hypothesis

View the lecture on YouTube:

1. Probability and Statistics by Dr. Somesh Kumar,

   Department of Mathematics, IIT Kharagpur

   https://youtu.be/14PQawp_rjk

2. Hypothesis Testing by Dr. J Maiti,

   Department of Management, IIT Kharagpur.

   https://youtu.be/3Tsiw84O_7o

MINI PROJECT

- https://www.analyticsvidhya.com/blog/2021/07/t-test-performing-hypothesis-testing-with-python/

Write a Python Program for t-test for the following problems and compute the result.

LEVEL 1

A group of 10 rats fed on diet A and another group of 8 rats fed on diet B, recorded the following increase in weights (gms).

Diet A: 5, 6, 8, 1, 12, 4, 3, 9, 6, 10

Diet B: 2, 3, 6, 8, 10, 1, 2, 8

Does it show superiority of diet A over dies B?

LEVEL 2

Two horses A and B were tested according to the time(in seconds)

To run a particular race with the following results

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether Horse A is running faster than Horse B at 5 level.

LEVEL 3

A mechanist is making engine parts with axle diameters of 0.7 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.04 inch. Compute the statistic you would use to test, whether the work is meeting the specification.

LEVEL 4

Write a Python Program for t-test for the following problem and compute the result.

A random sample of 10 boys had the following I.Qs:

70,120,110,101,88,83,95,98,107,100

Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of samples of10 boys lie.

LEVEL 5

Write a Python Program for t-test for the following problem and compute the result.

A random sample of 10 boys had the following IQ's: 70, 120, 110, 101, 88, 83, 95, 98, 107 & 100. Does the data support the assumption of a population mean IQ of 100?

ASSESSMENT SCHEDULE

S. NO.	ASSESSMENT DETAILS	PROPOSED DATE
1	MCQ TEST – I	15.03.2024
2	MCQ TEST – II	16.03.2024
3	CYCLE TEST – II	18.03.2024
4	INTERNAL ASSESSMENT TEST – II	11.04.2024

R.M.K GROUP OF INSTITUTIONS

R.M.K GROUP OF INSTITUTIONS

PRESCRIBED TEXT BOOKS & REFERENCE BOOKS

TEXT BOOKS:

1. Grewal. B.S. and Grewal. J.S., “Numerical Methods in Engineering and Science ", 10th Edition, Khanna Publishers, New Delhi, 2015.

2. Johnson, R.A., Miller, I and Freund J., “Miller and Freund’s Probability and Statistics for

Engineers", Pearson Education, Asia, 8th Edition, 2015.

REFERENCES:

1. Burden, R.L and Faires, J.D, "Numerical Analysis”, 9th Edition, Cengage Learning, 2016.

2. Devore. J.L., "Probability and Statistics for Engineering and the Sciences”, Cengage Learning, New Delhi, 8th Edition, 2014.

3. Gerald. C.F. and Wheatley. P.O. "Applied Numerical Analysis” Pearson Education, Asia, New Delhi, 2006.

4. Spiegel. M.R., Schiller. J. and Srinivasan. R.A., "Schaum’s Outlines on Probability and Statistics ", Tata McGraw Hill Edition, 2004.

5. Walpole. R.E., Myers. R.H., Myers. S.L. and Ye. K., “Probability and Statistics for Engineers and Scientists", 8th Edition, Pearson Education, Asia, 2007.

Thank you

Disclaimer:

This document is confidential and intended solely for the educational purpose of RMK Group of Educational Institutions. If you have received this document through email in error, please notify the system manager. This document contains proprietary information and is intended only to the respective group / learning community as intended. If you are not the addressee you should not disseminate, distribute or copy through e-mail. Please notify the sender immediately by e-mail if you have received this document by mistake and delete this document from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited.