19-09-2022 Experement-1

AIM: Estimation of population mean.

EXPERIMENT: A random sample X_1, X_2, X_3, X_4, X_5 of size 5 is drawn from a normal population with unknown mean μ. Consider the following estimator to estimate μ:

(i) $t_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

(ii) $t_2 = \frac{X_1 + X_2}{2} + X_3$

(iii) $t_3 = \frac{2X_1 + X_2 + \lambda X_3}{3}$

where λ is such that t_3 is an unbiased estimator of μ.

a) Find λ.

b) Are t_1 and t_2 unbiased?

c) State by giving reasons the estimator which is best among t_1, t_2 & t_3.

THEORY:

→ If the mean of the sampling distribution of a statistic is equal to the corresponding population parameter, then the statistic is said to be an unbiased estimator of the parameter, otherwise it is called a biased estimator.

Mathematically,

An estimator $\hat{\theta}$ of the parameter θ is said to be unbiased if its expected value is equal to θ.

→ Let for large samples, 2 consistent estimators t_n and t_n' of any parameter θ exist. and both be distributed asymptotically normal about the true value of the parameter θ with varience $\frac{\sigma^2}{n}$ and $\frac{\sigma'^2}{n}$, respectively.

Then, the estimator t_n (say) is said to be more efficient estimator than t_n' If

$$V(t_n) < V(t_n')$$
$$\text{or} \quad \sigma^2 < \sigma'^2 \quad \forall n.$$

CALCULATIONS:

Given that

$$X_i \sim N(\mu, \sigma^2)$$
$$\therefore E(X_i) = \mu$$

Varience $(X_i) = \sigma^2$

$\text{cov}(X_i, X_j) = 0 \quad , i, j = 1,2,3,4,5$

(i) $t_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

$$E[t_1] = E\left[\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right] = \frac{1}{5}\left[E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)\right] = \mu$$

Hence, t_1 is an unbiased estimator of μ.

(ii) $t_2 = \frac{X_1 + X_2}{2} + X_3$

$E(t_2) = E\left[\frac{X_1 + X_2 + 2X_3}{2}\right]$

$= \frac{1}{2}\left[E(X_1) + E(X_2) + 2E(X_3)\right]$

$= \frac{1}{2}[4\mu]$

$= 2\mu$

Hence, t_2 is not an unbiased estimator of μ.

(iii) Given that t_3 is an unbiased estimator of μ.

$\Rightarrow E[t_3] = \mu$

$E\left[\frac{2X_1 + X_2 + \lambda X_3}{3}\right] = \mu$

$\frac{1}{3}[2E(X_1) + E(X_2) + \lambda E(X_3)] = \mu$

$\frac{1}{3}[3\mu + \lambda\mu] = \mu$

$\lambda\mu = 0$

$\lambda = 0$

$Var(t_1) = Var\left(\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right)$

$= \frac{1}{25}\left(Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Var(X_5)\right)$

$= \frac{1}{25} \times 5\sigma^2$

$= \frac{\sigma^2}{5}$

$Var(t_2) = Var\left(\frac{X_1 + X_2 + 2X_3}{2}\right)$

$= \frac{1}{4}[Var(X_1) + Var(X_2) + 4Var(X_3)]$

$= \frac{3}{2}\sigma^2$

$Var(t_3) = Var\left(\frac{2X_1 + X_2}{3}\right)$

$= \frac{1}{9}[4Var(X_1) + Var(X_2)]$

$= \frac{5}{9}\sigma^2$

RESULTS:

a) The value of λ is 0

b) t_1 is an ubiased whereas t_2 is a biased estimator

c) Since $var(t_1)$ is the least, it is the best estimator

03-10-2022 Experiment-2.

AIM: To test and find the confidence intervals.

EXPERIMENT:

a) If a random sample of size $n=20$ from a normal population with the varience $\sigma^2=225$ has the mean $\bar{x}=64.3$. Construct a 95% confidence interval for the population mean μ.

b) An industrial designer wants to determine the average amount of time it takes an adult to assemble an "easy to assemble" toy. Use the following data (in minutes) of a random sample to construct a 95% confidence interval for the mean of the population sampled.

17, 13, 18, 19, 17, 21, 29, 22, 16, 28, 21, 15, 26, 23, 24, 20, 8, 17, 17, 21, 32, 18, 25, 22, 16, 10, 20, 22, 19, 14, 30, 22, 12, 24, 28, 11.

c) A paint manufacture wants to determine the average drying time of a new interior wall paint if for 12 test areas of equal size, he obtained a mean drying time of 66.3 minutes and a standard deviation of 8.4 minutes. Construct a 95% confidence interval for the mean μ.

THEORY:

→ If $\bar{x}$ is the value of the mean of a random sample of size n from a normal population with the known varience σ^2. Then,

$$\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is a $(1-\alpha)100\%$ confidence interval for the mean of the population.

→ If σ is unknown and normal population $n<30$ then reliability coefficient 't' is used instead of z.

CALCULATIONS:

(a) Given: $n=20$, $\bar{x}=64.3$, $\sigma=15$ and $z_{0.025}=1.96$

put all these values in-

$$\bar{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow 64.3-1.96\times\frac{15}{\sqrt{20}} < \mu < 64.3+1.96\times\frac{15}{\sqrt{20}}$$

$$\Rightarrow 57.7 < \mu < 21.79.$$

(b) $n = 36$, $\bar{x} = 19.92$, $z_{0.025} = 1.96$, $\sigma = 5.73$

$$\Rightarrow 19.92 - 1.96 \times \frac{5.73}{\sqrt{36}} < \mu < 19.92 + \frac{5.73}{\sqrt{36}} \times 1.96$$

$$\Rightarrow 18.05 < \mu < 21.79$$

(c) $\bar{x} = 66.3$, $s = 8.4$ and $t_{0.025} = 2.201$

$$66.3 - 2.201 \cdot \frac{8.4}{\sqrt{12}} < \mu < 66.3 + 2.201 \cdot \frac{8.4}{\sqrt{12}}$$

or $$61.0 < \mu < 71.6$$

RESULTS:

a) A 95% confidence interval for the population mean μ lies between 57.7 and 70.9

b) The 95% confidence limits are 18.05 and 21.79 minutes.

c) We can assert with 95% confidence that the interval from 61.0 minutes to 71.6 minutes contains the true average drying time of the paint.

01/4/22

11-10-2022. Experiment-3

AIM: To test and find the confidence interval for difference between means of population.

EXPERIMENT:

a) A study has been made to compare the nicotine contents of two brands of cigarettes. 10 cigarettes of Brand A had an average nicotine content of 3.1 mg with a standard deviation of 0.5 mg, while 8 cigarettes of Brand B had an average nicotine content of 2.7 mg with a standard deviation of 0.7 mg.
Assuming that the two sets of data are independent random sample from normal population with equal varience. Construct a 95% confidence interval for the difference between the mean nicotine contents of the two brands of cigrattes.

b) Construct a 94% confidence interval for the difference between the mean lifetimes of two kinds of light bulbs. Given that a random sample of 40 lightbulbs of the first kind lasted on the average 418 hours of continuous use and and 50 lightbulbs of the second kind lasted on the average 402 hours of continuous use. The population standard deviations are known to be $\sigma_1 = 26$ and $\sigma_2 = 22$.

c) In a random sample 136 of 400 persons given a flu vaccine experienced some discomfort. Construct a 95% confidence interval for the true proportion of person who will experience some discomfort from the vaccine.

THEORY:

→ If $\bar{x}_1, \bar{x}_2$, s_1 and s_2 are the values of the means and the Standard deviation of Independent random sample of sizes n_1 and n_2 from normal population with equal variences. Then,

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 + \bar{x}_2) + t_{\alpha/2, n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

is a $(1-\alpha)100\%$ confidence interval for difference between two population means.

→ If $\bar{x}_1$ and $\bar{x}_2$ are the values of means of independent random sample of sizes n_1 and n_2 from normal population with the known varience σ_1^2 & σ_2^2. Then,

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

is $(1-\alpha)100\%$ confidence interval for the difference between the two population means.

→ If X is a binomial random variable with the parameters n and θ. n is large and $\hat{\theta} = x/n$ then

$$\hat{\theta} - z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta + \hat{\theta} + z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

is an approximate $(1-\alpha)100\%$ confidence interval for θ.

CALCULATIONS:

a) $n_1 = 10,\ n_2 = 8,\ s_1 = 0.5,\ s_2 = 0.7,\ \alpha = 0.05$

$S_p = \sqrt{\dfrac{9(0.25) + 7(0.49)}{16}} = 0.596$

$t_{0.025,16} = 2.120$

$\Rightarrow (3.1 - 2.7) - 2.120(0.596)\sqrt{\dfrac{1}{10} + \dfrac{1}{8}} < \mu_1 - \mu_2 < (3.1 - 2.7) + 2.120(0.596)\sqrt{\dfrac{1}{10} + \dfrac{1}{8}}$

$\Rightarrow -0.20 < \mu_1 - \mu_2 < 1.00$

b) $\alpha = 0.06$, $z_{0.03} = 1.88$, Confidence interval for $\mu_1 - \mu_2$?

$\Rightarrow (418 - 402) - 1.88\sqrt{\dfrac{262}{40} + \dfrac{22^2}{50}} < \mu_1 - \mu_2 < (418 - 402) + \sqrt{\dfrac{26^2}{40} + \dfrac{22^2}{50}}$

$\Rightarrow 6.3 < \mu_1 - \mu_2 < 25.7$

c) $n = 400$, $\hat{\theta} = \dfrac{136}{400} = 0.34$ and $z_{0.025} = 1.96$

$\Rightarrow 0.34 - 1.96\sqrt{\dfrac{(0.34)(0.66)}{400}} < \theta < 0.34 + 1.96\sqrt{\dfrac{(0.34)(0.66)}{400}}$

$\Rightarrow 0.294 < \theta < 0.386$

or

$\Rightarrow 0.29 < \theta < 0.39$

RESULTS:

a) The 95% confidence limits are -0.20 and 1.00 milligram but observe that since this includes $\mu_1 - \mu_2 = 0$. we cannot conclude that there is a real difference between the nicotine contents of the brands of cigarettes.

b) we are 94% confident that the interval from 6.3 to 25.7 hours contains the actual difference between the mean lifetimes of two kinds of light bulbs. The fact that both confidence limits are positive suggests that on the average the first kind of light bulb is superior to the second kind.

c) Confidence interval for the binomial parameter θ is $0.29 < \theta < 0.39$.

07/11/22

01-11-22 Experiment - 4

AIM :

Testing of hypothesis

EXPERIMENT :

(i) A sample of 900 members has a mean 3.4 cm & S.D 2.61 cm. If the population is normal then test whether the sample has been drawn from a large population of mean 3.25 cm & S.D at 5% LOS

(ii) An insurance agent has claimed that the average age of policy holders who ensured through him is less than the average of all the agents which is 30.5 years. A random sample of 100 policy holders who ensured through him, given the following age distribution -

Age	Frequencies
16-20	12
21-25	22
26-30	20
31-35	30
36-40	16

Calculate mean & SD of the distribution & use these values to test his claims at 5% LOS.

(iii) Assume that IQ scores for certain population are approximately with mean μ & varience 100. To test $H_0 : \mu = 100$ against the one-sided alternative hypothesis $H_1 : \mu > 100$ we take a random sample of size 16 from this population & observed $\bar{x} = 113.5$. Test whether we accept or reject H_0 at 5% level of significance?

THEORY :

Testing of hypothesis —

1) Formulate the null hypothesis $H_0 : \mu = \mu_0$ & Formulate the alternative hypothesis

Case-1 → $\mu_1 : \mu \neq \mu_0$ (Two tailed)
Case-2 → $\mu_1 : \mu < \mu_0$ (left tailed)
Case-3 → $\mu_1 : \mu > \mu_0$ (Right-tailed)

2) choose the level of significance (α)

$\alpha = 0.05$ (5%) or $\alpha = 0.01$ (1%)

3) After this find the test statistic value :

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

→ For two-tailed test if

$$|Z_{cal}| \geq Z_{\alpha/2}$$

Then, reject H_0 otherwise accept H_0

→ For right-tailed test if

$$Z_{cal} \geq Z_{\alpha}$$

Then, reject H_0 otherwise accept H_0.

→ For left-tailed if

$$Z_{cal} \leq Z_{\alpha}$$

Then, reject H_0 otherwise accept H_0.

CALCULATIONS :

(i) $\mu_0 : \mu = 3.25$
$\mu_0 : \mu \neq 3.25$

$\bar{x} = 3.4$
$\sigma = 2.61$
$n = 900$
$\alpha = 0.05$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{———} \quad (*)$$

putting all the values in $(*)$,

$$Z = \frac{3.4 - 3.25}{2.61/30} = \frac{0.15}{0.087} = 1.724$$

$$Z_{\alpha/2} = Z_{0.025} = 1.96$$

$Z_{cal} < Z_{\alpha/2}$ (Using two-tailed test)

$\Rightarrow$ Accept H_0

(ii)

Age	f	x_i	x_i^2	f_ix_i	$f_ix_i^2$
15.5 - 20.5	12	18	324	216	3888
20.5 - 25.5	22	23	529	506	11638
25.5 - 30.5	20	28	784	560	15680
30.5 - 35.5	30	33	1089	990	32670
35.5 - 40.5	16	38	1444	608	23104
	$\Sigma f_i = 100$			$\Sigma f_ix_i = 2880$	$\Sigma f_ix_i^2 = 86980$

$$\text{Mean } (\bar{x}) = \frac{\Sigma f_ix_i}{\Sigma f_i} = \frac{2880}{100} = 28.8$$

$$\text{Varience } (\sigma^2) = \frac{\Sigma f_ix_i^2}{\Sigma f_i} - \left[\frac{\Sigma f_ix_i}{\Sigma f_i}\right]^2$$

$$= \frac{86980}{100} - (28.8)^2$$

$$= 40.56$$

$$\sigma = 6.3529$$

$\mu_0 : \mu = 30.5$ $\quad \sigma = 6.3529$ $\quad \bar{x} = 28.8$
$\mu_1 : \mu < 30.5$ $\quad n = 100$ $\quad \alpha = 0.05$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{28.8 - 30.5}{6.35/\sqrt{100}} = -2.677$$

$Z_\alpha = Z_{0.05} = -1.645$

For left tailed test,

$Z_{cal} < Z_\alpha$

but $Z_{cal} = -2.677 < -1.645$

$\Rightarrow$ Reject H_0

(iii) $H_0 : \mu = 100$ $\bar{x} = 113.5$ $n = 16$

$H_1 : \mu > 100$ $\sigma = 10$

Test statistic, $Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

$= \dfrac{113.5 - 100}{10/\sqrt{16}}$

$Z = 5.4$

$\alpha = 0.05$

$\therefore Z_\alpha = Z_{0.05} = 1.645$

For right tailed test,

$Z_{cal} = 5.4 > Z_\alpha = 1.645$

$\Rightarrow$ Reject H_0

RESULT:

(i) we will accept the null hypothesis (H_0) at 5% level of significance.

(ii) we will reject H_0 & the average age of policy holders who ensured through him is less than the average age of all the agents which is 30.5 years, is not true.

(iii) we will reject the null hypothesis (H_0) at 5% level of significance.

09-11-22

Experiment-5

AIM:

Testing of hypothesis concerning the varience of population.

EXPERIMENT:

(i) Suppose that metal of a uniformly thickness of a part is used in a semi-conductor is critical & that measurements of the thickness of a random sample of 18 such parts have the varience (S^2)=0.68. The process is considered to be under control if the variation of the thickness is given by $\sigma^2 > 0.36$. Assuming that the measurement constitute a random sample from a normal population. Test $H_0: \sigma^2 = 0.36$ against the alternative hypothesis $\sigma^2 > 0.36$ at 5% LOS.

(ii) Weights in kg of 10 students are given below-

38, 40, 45, 53, 47, 43, 55, 48, 52, 49.

Can we say that varience of the distribution of weights of all students from which the above sample was taken is equal to 20 square kg at 5% level of significance?

THEORY:

Testing the equality of population varience.

To test the H_0 that the varience of a normal population equals to a given constant given a random sample of size 'n' from a normal population. We shall test H_0,

$$H_0: \sigma^2 = \sigma_0^2 \text{ against}$$
$$H_1: \sigma^2 \neq \sigma_0^2 \text{ or } \sigma^2 > \sigma_0^2 \text{ or } \sigma^2 < \sigma_0^2$$

Here, we use the theorem that if x_1 & x_2 are independent r.v x_1 has a chi-square distribution with V_1 degree of freedom & (x_1+x_2) has a chi-square distribution with $V > V_1$ degree of freedom then x_2 has a chi-square distribution with $V-V_1$ degree of freedom. Thus, the critical region for testing the H_0 against the two one-sided alternatives as

$$\chi^2 \geq \chi^2_{\alpha}(n-1)$$
$$\& \ \chi^2 \leq \chi^2_{(1-\alpha),(n-1)}$$

is obtained by using the test statistic $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

CALCULATION:

(i) Given:

$S^2 = 0.68$

$n = 18$

$\alpha = 5\% = 0.05$

$\sigma_0^2 = 0.36$

$H_0 : \sigma^2 = 0.36$

$H_1 : \sigma^2 > 0.36$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = 32.111$

Tabulated value, $\chi^2_{0.05,17} = 27.587$

$\chi^2_{cal} > \chi^2_{0.05,17}$

(ii)

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
38	-9	81
40	-7	49
45	-2	4
53	6	36
47	0	0
43	-4	16
55	8	64
48	1	1
52	5	25
49	2	4
470		280

$\bar{x} = \frac{470}{10} = 47$

$s^2 = \frac{1}{n-1}(\Sigma(x_i - \bar{x})^2)$

$s^2 = \frac{1}{9} \times 280$

$s^2 = 31.11$

$\alpha = 0.05 = 5\%$

$n = 10$

$s^2 = 31.11$

$\sigma_0^2 = 20$

$H_0 : \sigma^2 = 20$

$H_1 : \sigma^2 > 20$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)\,31.11}{20}$

$\chi^2 = 14$

Tabulated value,

$\chi^2_{0.05,9} = 16.919$

$\chi^2_{cal} = 14 < \chi^2_{0.05,19}$

RESULT:

(i) we may reject H_0 & conclude that the varience of a given sample is not equal to 0,36 i.e the process used in manufacturing of the part must be revised.

(ii) we may accept H_0 & conclude that the varience of the distribution of weights of all the students is equal to 20 square kg.

10-11-22

Experiment - 6

AIM:

Testing the equality of two population variences.

EXPERIMENT:

Two samples are drawn from the two normal population from the data collected. Test whether the two samples have the same varience at 5% level of significance.

Sample 1 : 60, 65, 71, 74, 76, 82, 85, 87
Sample 2 : 61, 66, 67, 85, 78, 63, 85, 86, 88, 91

THEORY:

F distribution is used for asserting the equality of population variences.
$x_1, x_2, ---, x_n$ be a sample of size n_1
$y_1, y_2, ---, y_n$ be a sample of size n_2
from $N(\mu_1, \sigma_1^2)$ & $N(\mu_2, \sigma_2^2)$ respectively.

Here, we test the $H_0: \sigma_1^2 = \sigma_2^2$ & the test statistic is -

$$F = \frac{S_1^2}{S_2^2}, \quad (S_1^2 > S_2^2) \text{ ——— } (*)$$

where
$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$$
$$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\bar{y})^2$$

The sampling distribution of F-statistic in (*) follows F-distribution with

$v_1 = (n_1-1)$ degree of freedom
& $V_2 = (n_2-1)$ degree of freedom

As we know the larger the varience is taken in the numerator of (*) & the degree of freedom corresponding to this varience is taken as V_1.

$$F = \frac{S_2^2}{S_1} \quad (\because S_2^2 > S_1^2)$$
$$\sim F_{V_1=(n_2-1),\ V_2=(n_2-1)}(\alpha)$$

CALCULATION:

$H_0: \sigma_1^2 = \sigma_2^2$
against $H_1: \sigma_1^2 > \sigma_2^2$, to test H_0 Test statistic is

$$F = \frac{S_1^2}{S_2^2} \quad (\text{if } S_1^2 > S_2^2)$$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
60	-15	225
65	-10	100
71	-4	16
74	-1	1
76	1	1
82	7	49
85	10	100
87	12	144
600		636

y_i	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
61	-16	256
66	-11	121
67	-10	100
85	8	64
78	1	1
63	-14	296
85	8	64
86	9	81
88	11	121
91	14	296
770		1079

$\bar{x} = \frac{600}{8} = 75$

$S_1^2 = \frac{636}{7} = 90.85$

$S_1^2 = 90.85$

$\bar{y} = \frac{770}{10} = 77$

$S_2^2 = \frac{1079}{9}$

$S_2^2 = 133.33$

$F = \frac{S_2^2}{S_1^2} \quad (S_2^2 > S_1^2)$

$= \frac{133.33}{90.85} = 1.4676$

Also, tabulated $F_{9,7}(0.05) = 3.68$

$\Rightarrow$ Cal F < tab F

$\therefore$ we accept the null hypothesis.

RESULT:

The two samples have the same variance at 5% LOS.

Experiment-7

AIM: Testing of hypothesis using test of propotions.

EXPERIMENT:

(a) In a sample of 1000 people in Maharashtra, 540 are rice eaters. Can we assume that the rest wheat eaters and rice eaters are equally popular in this state at 1% level of significance?

(b) 20 people are attacked by a disease and only 18 survive. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of hypothesis that it is more at 5% level of significance?

THEORY:

In a sample of size 'n', let x be the no. of persons or items possessing the attribute. Then, the proportion of attribute in the sample will be -

$$\hat{p} = \frac{x}{n}$$

For testing $H_0: p = p_0$, we calculate the test statistic:—

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \quad \text{, where } q_0 = 1 - p_0 \quad \text{———} \quad (*)$$

CALCULATIONS:

a) Here, $x = 540$, $n = 1000$, $\alpha = 0.01$, $p_0 = \frac{1}{2}$ & $q_0 = \frac{1}{2}$
$H_0: p = p_0$, $H_1: p \neq p_0$
Putting all these values in $(*)$,

$$Z = \frac{0.54 - 0.5}{\sqrt{\frac{0.25}{1000}}} = \frac{0.04}{\sqrt{0.00025}} = 0.25$$

...so, $Z_{0.005} = 2.575$
Hence, $Z_{cal} < \text{tabulated } Z_{\alpha/2}$

...ere, $x = 18$, $n = 20$, $\alpha = 0.05$, $p_0 = 85$, $\hat{p} = \frac{x}{n} = \frac{18}{20} = 0.9$

...$_0: p = p_0 = 0.85$
...inst $H_1: p > p_0$
...tting all these values in $(*)$, $Z = \frac{0.9 - 0.85}{\sqrt{\frac{0.12}{20}}} = \frac{0.05}{0.077} = 0.64$

$Z_{0.05} = 1.645$
Hence, $Z_{cal} < \text{tabulated } Z_{\alpha}$.

RESULT:

(a) Since the calculated test statistic is less than tabulated $Z_{\alpha/2}$. Then wheat eaters and rice eaters may be equal in Maharashtra.

(b) Since the calculated test statistic is less than tabulated $Z_{\alpha/2}$ The survival rate if attacked by this disease is 85%.

14-11-22

Experiment-8

AIM: To use the level of significance for the difference of proportion and find the confidence interval for the mean of the population.

EXPERIMENT:

(a) 500 eggs were taken from a large consignment and 60 were found to be Bad. Assign the limit within which percentage probability lies.

(b) In a random sample of 1000 persons from town A, 400 were found to be consumer of rice. In another sample of 800 persons from Town B, 400 found to be consumer of rice. Test the hypothesis whether data reveals any significant difference town A & B as far as proportion of rice consumptions is concerned.

THEORY:

(b) In general, we do not have any information as the proportion of A's in the population from which the sample has been taken. Under $H_0: p_1 = p_2 = p$ (say) an unbiased estimate of population proportion P based on both the samples is

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Test statistic under $H_0: p_1 = p_2$ for difference of proportion

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \text{where, } Q = 1-P$$

(a) The probable limit for a normal variate X are:—

$$E(X) \pm 3\sqrt{V(X)}$$

$$\Rightarrow E(X) \pm 3.SE(X)$$

CALCULATIONS:

(a) We have, $n = 500$, $p = \frac{60}{500} = 0.12$

$$q = 1-p = 1-0.12 = 0.88, \quad E(X) = p = 0.12$$

$$V(X) = \frac{pq}{n} = \frac{0.12 \times 0.88}{500}$$

$$= 0.0002112$$

$$S.E(X) = 0.014532721\mathit{7}$$

Now, limit within which percentage lies, i.e, normal variate $\bar{x}$ is

$$E(X) \pm 3\sqrt{V(X)}$$

or $E(X) \pm 3\ S.E(X)$

$= 0.12 \pm 3(0.014532)$

$= [-0.076404, 0.163596]$

Thus, percentage probability is,

$[7.6404\%, 16.3598\%]$

(b) Here we have,

$n_1 = 1000$, $n_2 = 800$

$p_1 = 0.4 = \left(\frac{400}{1000}\right)$, $p_2 = 0.5\left(\frac{400}{800}\right)$

$q_1 = 1-0.4 = 0.6$, $q_2 = 1-0.5 = 0.5$

$$\hat{p} = \frac{n_1p_1 - n_2p_2}{n_1+n_2}$$

$$= \frac{0.14 \times 1000 - 0.5 \times 800}{1800}$$

$$= \frac{8}{18}$$

$$= \frac{4}{9} = 0.444$$

$\hat{q} = 1 - 0.444 = 0.5556 \approx 0.556$

Now,

$$z = \frac{0.4 - 0.5}{\sqrt{0.444 \times 0.556\left(\frac{1}{1000} + \frac{1}{800}\right)}} = -4.243$$

$|z| = 4.243$

Tabulated value of $z_{0.05} = 1.645$

$\Rightarrow$ Cal Z > Tab Z

$\therefore$ we may reject H_0 and accept H_1 at 5% LOS.

RESULT:

(a) The probability limit lies between 7.6404% to 16.3598%.

(b) There is a significant difference in the proportion of wheat consumers in the two cities.

21~~4~~-11-22

Experiment-9

AIM: To test the hypothesis using χ^2 test.

EXPERIMENT:

(a) 300 digits were chosen from a table of numbers and the following frequency distribution of numbers was obtained:

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	26	28	33	32	28	37	33	30	30	23

Test the hypothesis the digits are uniformly distributed over the table.

(b) A sample analysis of examination results of 200 MBA students was made. It was found that 46 students had failed, 68 secured 3rd division, 62 secured 2nd division and the rest were placed in 1st division. Do these figures commensurate with general exam results which is in the ratio 2:3:3:2 for various categories resp.?

(c) It is believed that the varience of an instrument is not equal to 0.16. write down the null and alternative hypothesis for testing this belief. Carry out test at 1% level of significance given 11 measurements of same subject on students. The observations are:
8.5, 3.3, 2.4, 2.3, 2.5, 2.7, 2.5, 2.6, 2.6, 2.7, 8.5.

THEORY:

(a) If fi or Oi are observed and expected Ei are frequency, the rare Pearson's χ^2 is given by

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

(b) For test of varience

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma_0^2}$$

(c) Critical region for testing the null hypothesis against the two one-sided alternative are

$$\chi^2 \geq \chi^2_{\alpha, n-1}$$
$$\& \ \chi^2 \leq \chi^2_{\alpha, n-1}$$

where the test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

As for a two sided alternative is concerned, we reject the null hypothesis if

$$\chi^2 \geq \chi^2_{\alpha/2, n-1}$$
$$\text{or } \chi^2 \leq \chi^2_{1-\alpha/2, n-1}$$

where $(n-1)$ is the degree of freedom.

CALCULATIONS:

(a) H_0: digits are uniformly distributed

Digits	Oi	Ei	$(Oi-Ei)^2$	$(Oi-Ei)^2/Ei$
0	26	30	16	0.53
1	28	30	4	0.133
2	33	30	9	0.3
3	32	30	4	0.133
4	28	30	4	0.133
5	37	30	49	.633
6	33	30	9	0.3
7	30	30	0	0
8	30	30	0	0
9	23	30	49	1.633
Total	300	300	144	4.735 ≈ 4.8

Tabulated value of χ_α, $\alpha = 0.01$ and $(n-1)$ df $= (10-1) = 9$ d.f

$\Rightarrow \chi^2_{0.01, 9} = 21.67$

Also, $\text{Cal } \chi^2 = \sum_{i=0}^{9} \frac{(Oi-Ei)^2}{Ei}$

$= 4.8$

Hence, $\text{Cal } \chi^2 < \chi^2_{0.01, 9}$

∴ H_0 is accepted.

(b) Assumption: The data must meet the chi-square goodness of fit.

H_0: The figures commensurate the general exam result in the ratio 2:3:3:2.

H: The result is not distributed in the ratio 2:3:3:2

X	Oi	Ei	$(Oi-Ei)^2$	$(Oi-Ei)^2/Ei$
Failed	46	$\frac{2}{10} \times 200$	36	0.9
3rd div.	60	$\frac{3}{10} \times 200$	64	1.066
2nd div.	62	$\frac{3}{10} \times 200$	64	0.066
1st div.	24	$\frac{2}{10} \times 200$	256	6.4
Total	200	200		8.432

$\Rightarrow \text{Cal } \chi^2 = 8.432$.

Tab $\chi^2_{0.05, 2}$ ⇒ critical value at 5% LOS & 2df is 5.9

$\Rightarrow \text{Cal } \chi^2 = 8.432 > \text{tab } \chi^2 = 5.991$

∴ H_0 is rejected.

(C) Let us assume

H : variance not equal to 0.16 i.e $\sigma^2 \neq 0.16$

$H_0 : \sigma^2 = 0.16$

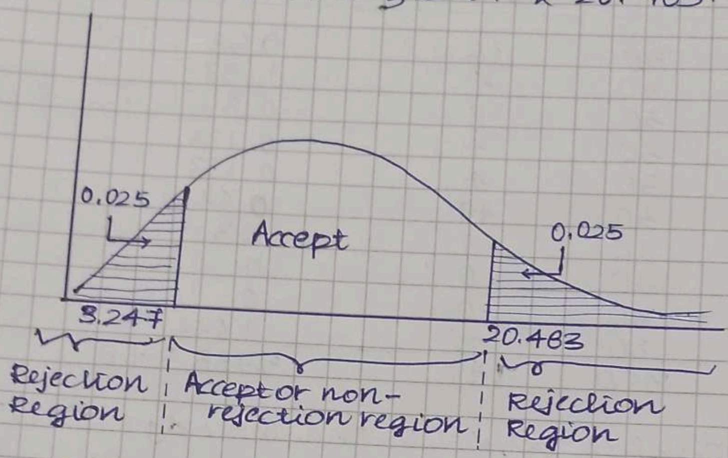
Decision rule :—

at $\alpha = 0.05$, critical value of χ^2 are

$\chi_{\frac{0.05}{2}, 10} = \chi_{0.025, 10} = 20.483$

$\chi_{1-\frac{0.05}{2}, 10} = \chi_{0.975, 10} = 3.247$

Reject Ho unless the computed value of the test statistic is between 3.247 & 20.483.



The test statistic is

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma^2} \qquad \text{———— } (*)$$

$n = 11$, $\sigma^2 = 0.16$

For s^2, use population varience formula

$$s^2 = \frac{\Sigma(X_i - \bar{X})}{n-1}$$

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
8.5	-4.8	23.04
3.3	-0.4	0.16
2.4	-1.3	1.69
2.3	-1.4	1.69
2.5	-1.2	1.44
2.7	1	1
2.5	-1.2	1.44
2.6	-1.1	1.21
2.6	-1.1	1.21
2.7	-1	1
8.5	4.8	23.04
Total		57.19

$\bar{X} = \frac{40.6}{11} \approx 3.7$

$$s^2 = \frac{57.19}{10} = 5.719 \approx 5.72$$

Now, $\chi^2 = \frac{(10)(5.72)}{0.16} = 357.5$ put this is (*).

$\chi^2 = 3.57$ which lies between 3.247 & 20.483

We donot reject Ho.

RESULTS:

(a) The digits are uniformly distributed since Ho is accepted.

(b) We conclude that this figure do not commensurate with the general exam results which is in the ratio 2:3:3:2 for various categories respectively.

(c) Based on the data we are unable to conclude that the population variance is not 16.

30-11-22

Experiment-10.

AIM: To test the hypothesis using signed rank test

EXPERIMENT:

The following are 15 measurements:
97.5, 95.2, 97.3, 96.0, 96.8, 100.3, 97.4, 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5, 94.9.
Use the signed rank test at 5% LOS to test whether mean equals 98.5.

CALCULATIONS:

The null hypothesis under the test is $H_0: \mu = 98.5$ against the alternative hypothesis $H_1: \mu \neq 98.5$

Measurements (x_i)	$x_i - 98.5$	Ranks	Signed Rank
97.5	-1.0	4	-4
95.2	-3.3	12	-12
97.3	-1.2	6	-6
96.0	-2.5	10	-10
96.8	-1.7	7	-7
100.3	1.8	8	8
97.4	-1.1	5	-5
95.3	-3.2	11	-11
93.2	-5.3	14	-14
99.1	0.6	2	2
96.1	-2.4	9	-9
97.6	-0.9	3	-3
98.2	-0.3	1	-1
94.9	-3.6	13	-13

$T^+ = 8+2 = 10$

$T^- = 4+12+6+10+7+5+11+14+9+3+1+13 = 95$

$\therefore$ The test statistic is $T = \min(T^+, T^-) = 10$

Now, Table value $T_{0.05} = 21$ for $n = 14$

$T = 10 < T_{0.05} = 21$

The null hypothesis may be rejected.

RESULTS:

The mean is not equal to 98.5.

05-12-2022. Experiment-11

AIM: Testing of hypothesis using analysis of variance (ANOVA).

EXPERIMENT:

The average number of days by mice inoculated with 5 strains of typhoid organisms along with the SD and number of mice involved in each experiment is given below. On the basis of this data, what will be your conclusion regarding the strain of typhoid organisms?

Strains of typhoid	A	B	C	D	E
No. of mice (n_i)	10	6	8	11	5
Average ($\bar{y}_i$)	10.9	13.5	11.5	11.2	15.4
Standard deviation (s_i)	12.72	5.96	3.24	5.65	3.64

THEORY:

The ANOVA table is given by —

Source of variation	Sum of squares	Degree of freedom	Mean sum of squares	Variance Ratio (F)
Treatment	SST	$K-1$	$\frac{SST}{K-1} = MST$	$F = \frac{MST}{MSE} \sim F(K-1, N-K)$
Error	SSE	$N-K$	$\frac{SSE}{N-K} = MSE$	(If MST > MSE)
Total	TSS	$N-1$		

Where,
SST = sum of squares due to treatement
SSE = sum of squares due to error
MST = mean sum of squares due to treatment
MSE = mean sum of squares due to error.
TSS = Total sum of squares.

Now, $\bar{y}_i = \frac{T_i}{n_i} \Rightarrow T_i = n_i \bar{y}_i$

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}^2 - \bar{y}_i^2$$

$$\Rightarrow \sum_{j=1}^{n_i} y_{ij}^2 = n_i(s_i^2 + \bar{y}_i^2)$$

Also, correction factor $= \frac{G^2}{N} = CF$

$\therefore RSS = \sum_i \sum_j y_{ij}^2$

$= \sum_i n_i (s_i^2 + \bar{y}_i)^2$

Thus, $TSS = RSS - CF$

$SST = \sum_i \frac{T_i^2}{n_i} - CF$

$SSE = TSS - SST$

CALCULATIONS:

Here, our null hypothesis is:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

i.e H_0: different strains of typhoid organisms are homogenous.

Against Alternative hypothesis,

H_1: atleast two of the means are different.

Let T_{i0} be the total of the ith strain of typhoid and $G = \sum_i T_{i0}$ be the grand total.

$T_{i0} = n_i \bar{y}_i$

$T_1 = 10 \times 10.9 = 109$

$T_2 = 6 \times 13.5 = 81$

$T_3 = 8 \times 11.5 = 92$

$T_4 = 11 \times 11.2 = 123.2$

$T_5 = 5 \times 15.4 = 77$

$G = \sum_i T_i = 109 + 81 + 92 + 123.2 + 77 = 482.2$

$CF = \frac{G^2}{N} = \frac{(482.2)^2}{40} = 5812.921$

$\sum_{j=1}^{n_i} y_{ij}^2 = n_i (s_i^2 + \bar{y}_i^2)$

$= 10[12.72^2 + 10.9^2] + 6[5.96^2 + 13.5^2] + 8[3.24^2 + 11.5^2] + 11[5.65^2 + 11.2^2] + 5[3.64^2 + 15.4^2]$

$RSS = 8237.7299$

$TSS = RSS - CF$

$= 8237.7299 - 5812.921$

$= 2424.8089$

$SST = \sum_i \frac{T_i^2}{n_i} - CF$

$= \left[\frac{(109)^2}{10} + \frac{(81)^2}{6} + \frac{(92)^2}{8} + \frac{(123.2)^2}{11} + \frac{(77)^2}{5}\right] - 5812.921$

$= 92.313$

$SSE = TSS - SST$

$= 2424.8089 - 92.313$

$= 2332.4959$

ANOVA TABLE :

Sources of variation	Sum of squares	Degree of Freedom	Mean sum of squares	Varience Ratio(F)
Treatment	92.313	5-1 = 4	MST = 23.07	$F = \frac{MSE}{MST} = 2.8876$
Error	2332.4959	40-5=35	MSE = 66.64	$\sim F(4,35)$

Also, we have from the table, $F(4,35) = 2.65$
Since Cal F > Tab $F(4,35)$

we may reject or may not accept the null hypothesis at 5% LOS.

RESULT :

Atleast two of the means are different.

21-11-22

Experiment-12.

AIM: To test the goodness of fit for the given data using chi-square test.

EXPERIMENT :

(a) Two samples polls of votes for two candidates A & B for a public office are taken, 2 from among the residence of rural areas, the results are given in the adjoining table. Examine whether the nature of the area is related to voting preference in the election.

AREA	Votes for A	Votes for B	TOTAL
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

(b) A survey of 800 families with 4 children revealed the following distribution -

No. of boys	0	1	2	3	4
No. of girls	4	3	2	1	0
No. of families	32	178	290	236	64

Is this result consistent with the hypotheses that males and female births are equally probable?

THEORY :

To test the null hypothesis that the two attributes are independent, we use χ^2-test.
First we setup the null hypothesis,

H_0 : Two attributes A and B are independent.
H_1 : Two attributes are dependent or associated.

To test H_0, we calculated the test statistic,

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\left[\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right] \sim \chi^2_{(r-1)(s-1)}(\;)$$

where, E_{ij} = expected frequency = $\frac{\Sigma(A_i) \times \Sigma(B_j)}{N}$

& O_{ij} = observed frequency.

(a) If O_i are observed frequencies and E_i are expected frequencies then Karl Pearson's χ^2 is given by —

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

(b) If x follows binomial distribution, then its probability distribution is given by —

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

p = probability of success

q = probability of families

n = Total

CALCULATIONS:

(a)

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

Comparing with general 2X2 table, here

$a = 620$, $b = 380$, $c = 550$, $d = 450$

$N = 2000$

H_0 : Nature of area is not related to voting preferences.

H_1 : Nature of area is related to voting prefrences.

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

$$= \frac{2000[(620)(450) - (380)(550)]^2}{1000 \times 1000 \times 830 \times 1170}$$

$$= \frac{2 \times 70000 \times 70000}{1000 \times 830 \times 1170}$$

$$= 10.091$$

$\alpha = 5\% = 0.05$

Tab $\chi^2 = \chi^2_{(0.05,1)} = 3.841$

Since, Cal χ^2 > Tab χ^2 our H_0 is rejected.

(b) H_0: Data is consistent with hypothesis of equal probability of male and female birth.

Then, $p = q = \frac{1}{2}$ with $n = 4$ and $N = 800$. Thus, expected frequency will be calculated using binomial distribution.

Let $p(r_i$ = probability of r male births in a family of 4)

Then, $f(r) = N \times p(r) = N \times {}^4C_r\, p^r q^{4-r}$

putting $r = 0, 1, 2, 3, 4$

$f(1) = 800 \times {}^4C_1 \left(\frac{1}{2}\right)^4 = 800 \times 4 \times \frac{1}{16} = 200$

$f(0) = 800 \times {}^4C_0 \left(\frac{1}{2}\right)^4 = 800 \times \frac{1}{16} = 50$

$f(2) = 800 \times {}^4C_2 \left(\frac{1}{2}\right)^4 = 800 \times 6 \times \frac{1}{16} = 300$

$f(3) = 800 \times {}^4C_3 \left(\frac{1}{2}\right)^4 = 200$

$f(4) = 800 \times {}^4C_4 \left(\frac{1}{2}\right)^4 = 50$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
32	50	324	6.48
178	200	484	2.42
290	300	100	0.33
236	200	1296	6.48
64	50	196	3.92
800	800		19.63

$$\text{Cal } \chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right] = 19.63$$

$$\text{Tab } \chi^2 = \chi^2_{(0.05,\, 4+1-1)} = \chi^2_{(0.05,\, 4)} = 9.488$$

Since Cal χ^2 > Tab χ^2, we reject our H_0.

RESULTS:

(a) Yes, the nature of area is related to voting preferences in the election.

(b) The given data is not consistent with the hypothesis that male & female birth are equally probable.