Unit - I

Testing of Hypothesis

Population:

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.

Random sampling:

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

Symbols of population and samples:

Characteristic	Population	Sample
	Parameters	Statistics
Symbols	Population size = N	Sample size = n
	Population mean = μ	Sample mean = $\bar{x}$
	Population standard deviation = σ	Sample standard deviation = s
	Population proportion = p	Sample proportion = $\bar{p}$

Testing a hypothesis:

On the basis of sample information, we make decisions about the population. In taking such decisions, we make certain assumptions. These assumptions are known as statistical hypothesis. These hypothesis are tested. Assuming the hypothesis is correct, we calculate the probability of getting the observed sample. In this probability is less than a certain assigned value, the hypothesis is to be accepted, otherwise rejected.

Critical region:

A region, corresponding to a statistic t, in the sample space S which amounts to rejection of the null Hypothesis H_0 is called as critical region or region of rejection.

Type I Error

Reject H_0 when it is true

Type II Error

Accept H_0 when it is wrong.

Large Sample Tests: $(n > 30)$

1) Test for single mean
2) Test for two means
3) Test for single Proportion
4) Test for two proportions

Formula $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ (or) $\dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

where $\bar{x}$ - Sample mean

μ - Population mean

σ - Population SD

n - Sample size.

Assumption of null Hypothesis

$H_0 : \mu = \mu_0$ then the alternate hypothesis is * $H_1 : \mu \neq \mu_0$ (Two tailed)

* $H_1 : \mu > \mu_0$ (Right tailed)

* $H_1 : \mu < \mu_0$ (Left tailed).

Table for critical values on using normal probability:

Critical values	Level of Significance α		
	1 %	5 %	10 %
Two tailed test	$\|z\| = 2.58$	$\|z\| = 1.96$	$\|z\| = 1.645$
Right tailed test	$z = 2.33$	$z = 1.645$	$z = 1.28$
Left tailed test	$z = -2.33$	$z = -1.645$	$z = -1.28$

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 & SD 2.61 cms? (Test at 5% LOS. The value of z at 5% level is $|Z_\alpha| < 1.96$).

N/D'2016
A/M'2018

Answer:

Given $n = 900$ $\quad \bar{x} = 3.4$

$\mu = 3.25$

$s = 2.61 \quad \alpha = 5\%$

Assume $H_0: \bar{x} = \mu$

$H_0: \bar{x} \neq \mu$ (use two tailed test)

Calculation:

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3.4 - 3.25}{\left(\frac{2.61}{\sqrt{900}}\right)}$$

$$= 1.724$$

Table value

$$|Z| = 1.96$$

Conclusion $\quad$ cal Z < Tab Z

$\therefore$ Accept H_0

95% Confidence limits are

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

$$= 3.4 \pm 1.96 \left(\frac{2.61}{\sqrt{900}}\right)$$

$$= 3.4 \pm 0.1705 = 3.57 \text{ \& } 3.2295$$

Problem: 2

The mean life time of a sample of 100 light bulbs produced by a company is computed to be 1570 hours with a SD 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 $\alpha = 0.05$ & 0.01.

Answer:

Given $n = 100$, $\mu = 1600$, $s = 120$, $\bar{x} = 1570$

$\alpha = 0.05$ & $\alpha = 0.01$

Assume null hypothesis $H_0 : \mu = 1600$

Alternate hypothesis $H_1 : \mu \neq 1600$

Calculation

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1570 - 1600}{120/\sqrt{100}}$$

$$= -2.5$$

Take $|Z| = Z = 2.5$ Cal $Z = 2.5$

Table value

at $\alpha = 5\%$, $|Z| = 1.96$ Tab $Z = 1.96$

at $\alpha = 1\%$, $|Z| = 2.58$ Tab $Z = 2.58$

Conclusion:

5% LoS	1% LoS
Cal Z > Tab Z ∴ Reject H_0	Cal Z < Tab Z ∴ Accept H_0

Problem: 3

The mean breaking strength of the cables supplied by a manufacturer is 1800 with a SD of 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the 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%. LOS?

Answer: Given $n = 50$, $\mu = 1800$, $s = 100$, $\bar{x} = 1850$

$\alpha = 1\%$

Assume null hypothesis $H_0: \bar{x} = \mu$

Alternate hypothesis $H_1: \bar{x} > \mu$

Calculation: $Z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{1850 - 1800}{(100/\sqrt{50})}$

$= 3.54$

Table value of Z = 2.33 at 1%.

Conclusion:

Cal Z > Tab Z

$\therefore$ we reject H_0

Problem: 4

The guaranteed average life of a certain type of electric light bulbs is 1000 hours with a SD of 125 hours. It is decided to sample the output so as to ensure that 90 percent of the bulbs do not fall short of the guaranteed average by more than 2.5%. what must be the minimum size of the sample?

Answer:

Given $\mu = 1000$ hours

$\sigma = 125$ hours.

Since, we do not want the sample mean to be less than the guaranteed average mean ($\mu = 1000$) by more than 2.5%.

$\therefore \bar{x} > 1000 - 2.5\%$ of 1000

$\Rightarrow \bar{x} > 1000 - 25$

$\Rightarrow \bar{x} > 975$

Formula $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

we want $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{975 - 1000}{125/\sqrt{n}} > -\dfrac{\sqrt{n}}{5}$

According to the given condition

$$P(Z > -\sqrt{n}/5) = 0.90$$

$$P(0 < Z < \sqrt{n}/5) = 0.40$$

$\therefore \sqrt{n}/5 = 1.28$ (from the normal prob. tables)

$$\Rightarrow n = 25 \times (1.28)^2$$

$= 41$ approximately.

Large sample test (Normal distribution) for difference of means:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (\text{or}) \quad \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Problem : 1

Examine whether the difference in the variability in yields is significant at 5% LOS for the following.

	set of 40 plots	set of 60 plots
mean yield per plot	1258	1243
SD per plot	34	28

N/D'2017

Answer: Given $n_1 = 40$, $\bar{x}_1 = 1258$, $s_1 = 34$
$n_2 = 60$, $\bar{x}_2 = 1243$, $s_2 = 28$ } $\alpha = 5\%$ LOS

Assume null Hypothesis H_0: $\bar{x}_1 = \bar{x}_2$
There is no difference b/w means of samples.

Alternate Hypothesis H_1: $\bar{x}_1 \neq \bar{x}_2$

Calculation

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1258 - 1243}{\sqrt{\frac{1156}{40} + \frac{784}{60}}}$$

$$= 2.32$$

Table value : $Z = 1.96$ at 5% LOS

Conclusion:
Here Cal Z > Tab Z

$\therefore$ Reject H_0

Problem : 2

The means of two large samples of 1000 & 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the Same population of SD 2.5 inches?

Answer:

Given $n_1 = 1000$, $\bar{x}_1 = 67.5$, $s_1 = s_2 = 2.5$

$n_2 = 2000$, $\bar{x}_2 = 68$, $\alpha = 5\%$ LOS

Assume null hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

Alternate hypothesis $H_1: \bar{x}_1 \neq \bar{x}_2$

Calculation:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{67.5 - 68}{\sqrt{\frac{2.5^2}{1000} + \frac{2.5^2}{2000}}}$$

$$= -5.16$$

$|Z| = 5.16$

Table value of $Z = 1.96$, 5% LOS

Conclusion

Here Cal Z > Tab Z

$\therefore$ Reject H_0

Problem: 3

A sample heights of 6400 indians has a mean of 67.85 inches & SD 2.56 inches, while a sample heights of 1600 Australians has a mean of 68.55 inches & SD 2.52 inches. Do the data indicate that Australians are on the average taller than indians?

Answer:

Given $n_1 = 6400$, $\bar{x}_1 = 67.85$ $s_1 = 2.56$

$n_2 = 1600$, $\bar{x}_2 = 68.55$ $s_2 = 2.52$

$\alpha = 5\%$ LOS.

Null Hypothesis H_0:

There is no difference b/w the average heights indians & Australians.

ie. $H_0: \bar{x}_1 = \bar{x}_2$

Alternate hypothesis $H_1: \bar{x}_1 < \bar{x}_2$

Calculation:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{67.85 - 68.55}{\sqrt{\frac{2.56^2}{6400} + \frac{2.52^2}{1600}}}$$

$$= -9.906$$

$$|Z| = 9.906$$

Table value $Z = 1.645$ (5% LOS)

Conclusion:

Cal Z > Tab Z

∴ Reject H_0

Problem: 3

In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400, the mean is 15. Could the samples have been drawn from the sample population with SD 4?

Answer:

$n_1 = 500$ $\quad \bar{x}_1 = 20$ $\quad \sigma_1 = \sigma_2 = 4$

$n_2 = 400$ $\quad \bar{x}_2 = 15$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$ (There is no difference b/w the mean values of the samples).

Alternate Hypothesis $H_1 : \bar{x}_1 \neq \bar{x}_2$

Calculation

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{20-15}{\sqrt{\frac{4^2}{500} + \frac{4^2}{400}}}$$

$$= 18.6$$

Table value of $Z = 2.58$ (at 1% LOS)

Conclusion: Cal Z > Tab Z

$\therefore$ Reject H_0

Problem: 4

The sales manager of a large company conducted a sample survey in states A & B taking 400 samples in each case. The results were in the following table. Test whether the average sale is same in the 2 states at 1% Level.

A/M'2017

	State A	State B
Average sales	RS. 2500	RS. 2200
(SD)	RS. 400	RS. 550

Answer:

Given $n_1 = 400$, $\bar{x}_1 = 2500$, $s_1 = 400$

$n_2 = 400$, $\bar{x}_2 = 2200$, $s_2 = 550$

$\alpha = 1\%$

Null Hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

There is no difference between the average sales of two states.

Alternate Hypothesis $H_1: \bar{x}_1 \neq \bar{x}_2$

Calculation:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2500 - 2200}{\sqrt{\frac{400^2}{400} + \frac{550^2}{400}}}$$

$$= 8.82$$

Table value of $Z = 2.58$

Conclusion:

Here Cal Z > Tab Z

$\therefore$ Reject H_0

Problem: 5

Test the significance of the difference b/w the means of the samples, drawn from the two normal populations with the same SD from the following data.

	Size	Mean	SD
Sample I	100	61	4
Sample II	200	63	6

Answer:

$n_1 = 100$, $\bar{x}_1 = 61$, $s_1 = 4$ $\alpha = 5\%$ LOS

$n_2 = 200$, $\bar{x}_2 = 63$, $s_2 = 6$

Null Hypothesis H_0 : $\bar{x}_1 = \bar{x}_2$

Alternate Hypothesis H_1 : $\bar{x}_1 \neq \bar{x}_2$

Calculation: $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{61-63}{\sqrt{\frac{4^2}{100} + \frac{6^2}{200}}}$$

$$= -3.43$$

$$|Z| = 3.43$$

Table value of $Z = 1.96$

Conclusion: Cal Z > Table Z

$\therefore$ Reject H_0

Large Sample test for single proportion

Formula $$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

Problem: 1

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

Answer: Given $n = 1000$, $X = 540$, $p = \frac{X}{n} = \frac{540}{1000} = 0.54$

P = Population proportion of rice eaters in Maharastra $= \frac{1}{2} = 0.5$

$Q = 1 - P = 1 - \frac{1}{2} = 0.5$

Null Hypothesis H_0: Both rice & wheat eaters are equally popular. $(P = 0.5)$

Alternate Hypothesis H_1: $P \neq 0.5$

Calculation $Z = \dfrac{p - P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.54 - 0.50}{\sqrt{\frac{(0.5)(0.5)}{1000}}} = 2.53$

Table value of $Z = 2.58$

Conclusion:

Cal $Z <$ Tab Z

Accept H_0

Problem: 2

Twenty people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease, is 85% in favour of the hypothesis that it is more, at 5% Level. (use large Sample Test).

Answer: Given $n = 20$

X = Number of persons who survived after attack by a disease $= 18$

$p = \frac{x}{20} = 0.90$

$P = 0.85 \quad \Rightarrow Q = 1 - P = 0.15$

Null Hypothesis H_0: $P = 0.85$

Alternate hypothesis H_1: $P > 0.85$

Calculation: $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.90 - 0.95}{\sqrt{\frac{(0.85)(0.15)}{20}}}$

$= 0.626$

Table value of $Z = 1.645$

Conclusion:

Cal Z < Tab Z

Accept H_0

Problem: 3

A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a Sample of 200 pieces of equipments revealed that 18 were faculty. Test his claim at a significance level of 5% & 1%.

Answer: Given $n = 200$

x = Number of pieces conforming to specifications in the samples $= 200 - 18 = 182$

p = Sample proportion conforming to specifications

$= \frac{182}{200} = 0.91$

$P = 0.95, \quad Q = 1 - P = 0.05, \quad \alpha = 0.05 \ \& \ 0.01$ LOS.

Null Hypothesis H_0: The proportion of pieces conforming to specification in the population is 95%.

ie. $P = 0.95$

Alternate Hypothesis H_1: $P < 0.95$

Calculation:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} = -2.6$$

$$|Z| = 2.6$$

Table value of $Z = 1.645$ (5% LOS)

$Z = 2.33$ (1% LOS)

Conclusion:

(i) 5% LOS	(ii) 1% LOS
Cal Z > Tab Z	Cal Z > Tab Z
Reject H_0	Reject H_0

Problem: 4

A coin is tossed 144 times and a Person gets 80 heads. Can we say that the coin is unbiased one?

Answer: Given $n = 144$

P = probability of getting head $= \frac{1}{2}$

$Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$

x = number of successes = number of getting heads

$= 80$

$$p = \frac{X}{n} = \frac{80}{144} = \frac{5}{9}$$

Null Hypothesis H_0: Coin is unbiased.

Alternate hypothesis H_1: Coin is biased.

Calculation:

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{\left(\frac{5}{9}\right) - \left(\frac{1}{2}\right)}{\sqrt{\frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{144}}} = 1.33$$

Table value of $Z = 1.96$

Conclusion: Cal Z < Tab Z

Accept Z_0.

Problem - 5

In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers.

Answer: Given $n = 600$

$$p = \frac{325}{600} = 0.5417$$

P = Population proportion of smokers in the city $= 0.5$

$\therefore Q = 1 - P = 0.5$

Assume $H_0: p = 0.5$

$H_1: p > 0.5$

Calculation:

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 2.04$$

Table value of $Z = 2.33$ (1 % LOS)

Conclusion:

Cal $Z <$ Tab Z

$\therefore$ Accept H_0.

Large Sample test for difference of Proportions

Formula $$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Problem: 1

During a Country wide investigation the incidence of TB was found to be 1%. In a college of 400 strength 5 were reported to be affected whereas in another college of 1200 strength 10 were reported to be affected. Does this indicate any significant difference.

Answer:

Given $P = 1\% = \frac{1}{100} = 0.01$, $Q = 1-P = 0.99$

$n_1 = 400$, $n_2 = 1200$, $p_1 = \frac{5}{400} = 0.0125$, $p_2 = \frac{10}{1200} = 0.0083$

Null Hypothesis H_0: There is no difference b/w the Proportions

ie. $P_1 = P_2$

Alternate Hypothesis H_1: $P_1 \neq P_2$

Calculation:

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.0125 - 0.0083}{\sqrt{(0.01)(0.99)\left\{\frac{1}{400}+\frac{1}{1200}\right\}}}$$

$$= 0.7368$$

Table value of $Z = 1.96$ at 5% LoS

Conclusion:

Cal Z < Tab Z

∴ Accept H_0.

Problem: 2

Random Samples of 400 men & 600 women were asked whether they are like to have a residence near flyover. 200 men & 325 women were in favour of the proposals. Test the hypothesis that proportions of men & women infavour of the proposal, are same against that they are not, at 5% LoS.

Answer:

Given

$n_1 = 400$ x_1 = No. of men favouring the proposal $= 200$

$n_2 = 600$

x_2 = No. of women favouring the proposal $= 325$.

$p_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5$, $p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.541$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(400)(0.5) + (600)(0.541)}{400 + 600}$$

$$= 0.525$$

$$Q = 1 - P = 1 - 0.525 = 0.475$$

Null Hypothesis $H_0: P_1 = P_2$

There is no difference b/w the Proportions.

Alternate Hypothesis $H_1: P_1 \neq P_2$

<u>Calculation</u>: $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.500 - 0.541}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$

$$= -1.269$$

$$|Z| = 1.269$$

Table value of $Z = 1.96$

<u>Conclusion</u>: Cal Z < Tab Z

$\therefore$ Accept H_0

Problem : 3

A machine produced 20 defective units in a sample of 400. After overhauling the machine, it produced 10 defective units in a batch of 300. Has the machine improved in production due to overhauling. Test it at 5% LoS.

Given $n_1 = 400$, $n_2 = 300$

$$p_1 = \frac{20}{400} = 0.05$$

$$p_2 = \frac{10}{300} = 0.033$$

$$\therefore P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(400)(0.05) + (300)(0.033)}{400+300}$$

$$= 0.0427$$

$$\therefore Q = 1 - P = 0.9573$$

Null Hypothesis H_0: $P_1 = P_2$

There is no difference b/w the Proportions

Alternate hypothesis H_1: $P_1 > P_2$

Calculation: $Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.050 - 0.033}{\sqrt{(0.0427)(0.9573)\left(\frac{1}{400}+\frac{1}{300}\right)}}$

$$= 1.1$$

Table value of $Z = 1.96$

Conclusion:

Cal Z < Tab Z

Accept H_0.

Problem: 4

In two large populations, there are 30 & 25 percent respectively of blue-eyed people. In this difference likely to hidden in samples of 1200 & 900 respectively from the two populations?

Answers:

Given $n_1 = 1200$, $n_2 = 900$

P_1 = Proportion of blue eyed people in the first Population $= 30\% = 0.30$

P_2 = Proportion of blue eyed people in the second Population $= 25\% = 0.25$

$Q_1 = 1 - P_1 = 1 - 0.30 = 0.70$

$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$

Null Hypothesis H_0: There is no difference b/w the sample proportions $P_1 \neq P_2$

Alternate Hypothesis H_1: $P_1 \neq P_2$

Calculation:

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} = \frac{0.30 - 0.25}{\sqrt{\frac{(0.3)(0.7)}{1200} + \frac{(0.25)(0.75)}{900}}}$$

$$= 2.56$$

Table value of $Z = 1.96$

Conclusion:

Cal Z > Tab Z

Reject H_0

∴ we conclude that the difference in the population proportions is unlikely to be hidden in sampling.

Hence the problem.

Small Sample Test ($n < 30$)

t- test

Type1 : t-test for single mean

Type 2: t- test for two means.

Formule $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$ where $\bar{x}$ - Sample mean

μ - Population mean.

s - SD

n - number of Samples.

Degrees of freedom $= n-1$

Problem: 1

Given a Sample mean of 83, a sample SD of 12.5 & a sample size of 22, test the hypothesis that the value of the population mean is 70 against the alternative that it is more than 70. Use the 0.025 significance level.

A/M 2018

Given $n = 22$, $\mu = 70$

$s = 12.5$, $\bar{x} = 83$

$\alpha = 0.025$

$= \frac{25}{1000} \times 100 = 2.5\%$.

Null Hypothesis H_0:

There is no difference b/w the Sample mean & population mean.

ie. $\bar{x} = \mu$

Alternate hypothesis H_1: $\bar{x} > \mu$

Degrees of freedom $df = 22-1 = 21$

Calculation $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$= \frac{83-70}{(12.5/\sqrt{21})} = 4.77$$

Table value of $t = 2.080$ (2.5 LOS , Df = 21)

Conclusion:

Cal t > Tab t

Reject H_0

Problem:2

A machinist 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 SD of 0.04 inch. Compute the statistic you would use to test, whether the work is meeting the specification.

A/M 2019

Answer: Given $n = 10$

$\bar{x} = 0.742$

$s = 0.04$

$\mu = 0.7$

$\alpha = 5\%$ LOS

$Df = n-1 = 10-1 = 9$

(assume) Null Hypothesis H_0: There is no difference b/w the sample mean & the population mean

$\bar{x} = \mu$

Alternate Hypothesis H_1: $\bar{x} \neq \mu$

Calculation: $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$= \frac{0.742 - 0.7}{(0.04/\sqrt{9})} = 3.15$$

Table value of $Z = 2.262$ (5% LOS, Df = 9)

Conclusion:

Cal t > Tab t

$\therefore$ Reject H_0

Problem: 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?

Answer: Given $n = 12$

													Total
x	5	2	8	−1	3	0	6	−2	1	5	0	4	$\Sigma x = 31$
x^2	25	4	64	1	9	0	36	4	1	25	0	16	$\Sigma x^2 = 185$

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

Variance $s^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = \frac{185}{12} - (2.58)^2$

$= 8.761$

$\therefore$ SD $s = \sqrt{8.761}$

$= 2.96$

Null Hypothesis H_0: $\bar{x} = \mu$

Alternate hypothesis H_1: $\bar{x} > \mu$

$\alpha = 5\% = 0.05$ LOS

$df = n-1 = 12-1 = 11$

<u>Calculation</u> $t = \dfrac{\bar{x}-\mu}{(s/\sqrt{n-1})} = \dfrac{2.58-0}{(2.96/\sqrt{11})}$

$= 2.89$

Table value of $t = 1.796$ (5% LOS, $Df = 11$)

<u>Conclusion:</u> Cal t > tab t

$\therefore$ Reject H_0

<u>Problem: 4</u>

The mean life time of a sample of 25 bulbs is found as 1550 hours with a SD 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% LOS?

<u>Answer:</u> Given $n = 25$

$\bar{x} = 1550$

$s = 120$ & $\mu = 1600$

Null Hypothesis H_0: $\bar{x} = \mu$

there is no difference b/w the population mean & sample mean.

Alternate Hypothesis H_1: $\bar{x} < \mu$

$\alpha = 5\%$ LOS

$df = n-1 = 25-1 = 24$

Calculation

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{1550 - 1600}{[120/\sqrt{25-1}]}$$

$$= -2.04$$

$|t| = 2.04$

Table value of $t = 1.711$ (5% LOS & Df = 24)

Conclusion:

cal t > tab t

Reject Ho

Problem: 5

A test of a breaking strengths of 6 ropes manufactured by a company showed a mean breaking strength of 3515 kg and a SD of 60 kg where as the manufacturer claimed ~~#~~ a mean breaking strength of 3630 kg. Can we support the manufacturer's claim at the level of significance 0.05.

A/M '2019
N/D '2020

Answer: Given $\bar{x} = 3515$

$\mu = 3630$

$s = 60$

$n = 6$

Null Hypothesis Ho: $\bar{x} = \mu$

there is no difference b/w the sample mean & the population mean

Alternate hypothesis H_1: $\bar{x} < \mu$

$\alpha = 5\%$ (5% LOS)

Calculation:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{3515 - 3630}{(60/\sqrt{6-1})} = -4.286$$

$|t| = 4.286$

Cal. $t = 4.286$

Table value of $t = 2.01$

Conclusion:

Cal t > tab. t

$\therefore$ Reject H_0

Type: 2

t - test for difference of means.

Formula $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $s^2 = \left\{\dfrac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}\right\}$

(or)

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

where s_1, s_2 are SD of given samples.

Problem: 1

Two horses A & 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 (A & B were tested according to the time) the horse A is running faster than B at 5% LOS.

Answer: Given $n_1 = 7, n_2 = 6$

$\Sigma x_1 = 28 + 30 + 32 + 33 + 33 + 29 + 34 = 219$

$\Sigma x_1^2 = 28^2 + 30^2 + 32^2 + 33^2 + 33^2 + 29^2 + 34^2 = 6883$

$\Sigma x_2 = 29 + 30 + 30 + 24 + 27 + 29 = 169$

$\Sigma x_2^2 = 29^2 + 30^2 + 30^2 + 24^2 + 27^2 + 29^2 = 4787$

$\therefore \bar{x}_1 = \frac{\Sigma x_1}{n_1} = 31.29, \quad \bar{x}_2 = \frac{\Sigma x_2}{n_2} = 28.17$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \left(\frac{\Sigma x_1}{n_1}\right)^2 = \frac{6883}{7} - (31.29)^2 = 4.23$

$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \left(\frac{\Sigma x_2}{n_2}\right)^2 = \frac{4787}{6} - (28.17)^2 = 4.28$

$\therefore S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = 5.03$

Null Hypothesis: There is no difference b/w

H_0: Population means $(\mu_1 = \mu_2)$

(or)

There is no difference b/w Sample means $(\bar{x}_1 = \bar{x}_2)$

Alternate hypothesis $H_1: \mu_1 \neq \mu_2$

(or)

$\bar{x}_1 \neq \bar{x}_2$

Calculation $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{31.29 - 28.17}{\sqrt{5.03\left(\frac{1}{7} + \frac{1}{6}\right)}} = 2.5$

Conclusion :

Cal t > tab t

Reject H_0

Problem : 2

A group of 10 rats fed on diet A & 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 superiority of diet A over diet B.

N/D '2017

Answer:

Given $n_1 = 10$

$n_2 = 8$

$$\Sigma x_1 = 5+6+8+1+12+4+3+9+6+10 = 64$$

$$\Sigma x_1^2 = 5^2+6^2+8^2+1^2+12^2+4^2+3^2+9^2+6^2+10^2 = 512$$

$$\Sigma x_2 = 2+3+6+8+10+1+2+8 = 40$$

$$\Sigma x_2^2 = 2^2+3^2+6^2+8^2+10^2+1^2+2^2+8^2 = 282$$

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

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

$$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \left(\frac{\Sigma x_1}{n_1}\right)^2 = 10.24$$

$$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \left(\frac{\Sigma x_2}{n_2}\right)^2 = 10.25$$

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

$$= 11.525$$

Null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternate hypothesis $H_1 : \mu_1 > \mu_2$

$\alpha = 5\%$ LOS , $Df = n_1 + n_2 - 2 = 10 + 8 - 2 = 16$

Calculation:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{6.4 - 5}{\sqrt{11.525\left(\frac{1}{10} + \frac{1}{8}\right)}} = 0.869$$

Table value of $t = 1.746$

Conclusion: Cal t < tab t

Accept H_0

Problem: 3

The following random samples, are measurements of the heat producing capacity (in millions of calories per ton) of specimen's of coals from two mines.

Mine 1	8260	8130	8350	8070	8340	-
Mine 2	7950	7890	7900	8140	7920	7840

Use the 0.01 LOS to test whether the difference b/w the means of these two samples is significant

Answer:

$\Sigma x_1 = 8260 + 8130 + 8350 + 8070 + 8340 = 41150$

$\Sigma x_1^2 = 338\,727\,500$

$\Sigma x_2 = 7950 + 7890 + 7900 + 8140 + 7920 + 7840$

$= 47640$

$\Sigma x_2^2 = 37831600$

$\therefore \bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{41150}{5} = 8230$

$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{47640}{6} = 7940$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \left(\frac{\Sigma x_1}{n_1}\right)^2 = 12600$

$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \left(\frac{\Sigma x_2}{n_2}\right)^2 = 9100$

$\therefore S^2 = \left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\right) = \frac{(5)(12600)+(6)(9100)}{5+6-2}$

$= 13066.67$, Table value $t = 3.25$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternate Hypothesis $H_1: \mu_1 \neq \mu_2$

$\alpha = 1\%$ LOS, $Df = n_1+n_2-2$

$= 5+6-2$

$= 9$

Calculation: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{8230-7940}{\sqrt{13066.67\left(\frac{1}{5}+\frac{1}{6}\right)}} = 4.19$

Conclusion: Cal t > Tab t

$\therefore$ Reject H_0

Problem :4 Two independent samples are chosen from two schools A & B, a common test is given in a subject. The scores of the students as follows:

School A	76	68	70	43	94	68	33	-
School B	40	48	92	85	70	76	68	22

Can we conclude that students of school A performed better than students of school B? (Try)

Problem: 5

The independent samples from normal populations with equal variance gave the following

Sample	Size	mean	SD
1	16	23.4	2.5
2	12	24.9	2.8

Is the difference b/w the means significant?

N/D'2019

Answer: Given $\overline{x_1} = 23.4,\ \bar{x}_2 = 24.9$

$s_1 = 2.5,\ s_2 = 2.8$

$n_1 = 16,\ n_2 = 12$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{(16)(2.5)^2 + (12)(2.8)^2}{16+12-2}$$

$$= 7.4646$$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternate hypothesis $H_1: \mu_1 \neq \mu_2$

$\alpha = 0.05,\ df = n_1+n_2-2 = 16+12-2 = 26$

Calculation:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{23.4 - 24.9}{\sqrt{7.4646\left(\frac{1}{16}+\frac{1}{12}\right)}}$$

$$= -1.4376$$

$$|t| = 1.4376$$

Table value of $t = 2.056$

Conclusion:

Cal t < tab t

Accept H_0.

F - Distribution (F - Test)

(Snedecor's F - Distribution)

Test for equality of variances.

Formula: $F = \dfrac{S_1^2}{S_2^2}$, $S_1^2 > S_2^2$

where $S_1^2 = \dfrac{n_1 s_1^2}{n_1 - 1}$

$S_2^2 = \dfrac{n_2 s_2^2}{n_2 - 1}$

Problem: 1

A group of 10 rats fed on diet A and another group of 8 rats fed on diet B recorded the following increase of in weight. Find if the variances are significantly different.

Diet A	5	6	8	1	12	4	3	9	6	10	-
Diet B	2	3	6	8	10	1	2	8	-	-	-

Answer:

Given $n_1 = 10$, $n_2 = 8$

											Total
x_1	5	6	8	1	12	4	3	9	6	10	$\Sigma x_1 = 64$
x_1^2	$5^2 = 25$	36	64	1	144	16	9	81	36	100	$\Sigma x_1^2 = 512$
x_2	2	3	6	8	10	1	2	8	–	–	$\Sigma x_2 = 40$
x_2^2	4	9	36	64	100	1	4	64	–	–	$\Sigma x_2^2 = 282$

$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{64}{10} = \frac{32}{5}$

$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \left(\frac{\Sigma x_1}{n_1}\right)^2$

$= \frac{512}{10} - \left(\frac{32}{5}\right)^2$

$= 10.24$

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

$= 11.38$

$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{40}{8} = 5$

$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \left(\frac{\Sigma x_2}{n_2}\right)^2$

$= \frac{282}{8} - 5^2$

$= 10.25$

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

$= 11.71$

Null Hypothesis H_0: There is no different b/w the variances.

Calculation

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

$\therefore F = \frac{S_2^2}{S_1^2} = \frac{11.71}{11.38} = 1.03$

Table value of $F(7, 9) = 3.29$

Conclusion: Cal F < Tab F

$\therefore$ Accept H_0

Problem: 2

Two independent samples of sizes 9 & 7 from a normal population has 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% LOS.

A/M 2017

Answer:

Given $n_1 = 9$, $n_2 = 7$

										Total
Sample I x_1	18	13	12	15	12	14	16	14	15	$\Sigma x_1 = 129$
x_1^2	18^2	13^2	12^2	15^2	12^2	14^2	16^2	14^2	15^2	$\Sigma x_1^2 = 1879$
Sample II x_2	16	19	13	16	18	13	15	-	-	$\Sigma x_2 = 110$
x_2^2	16^2	19^2	13^2	16^2	18^2	13^2	15^2	-	-	$\Sigma x_2^2 = 1760$

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{129}{9}$$

$$s_1^2 = \frac{\Sigma x_1^2}{n_1} - \left(\frac{\Sigma x_1}{n_1}\right)^2$$

$$= \left(\frac{1879}{9}\right) - \left(\frac{129}{9}\right)^2 = \frac{10}{3}$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = 3.75$$

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

$$s_2^2 = \frac{\Sigma x_2^2}{n_2} - \left(\frac{\Sigma x_2}{n_2}\right)^2$$

$$= \frac{1760}{7} - \left(\frac{110}{7}\right)^2 = \frac{220}{49}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = 5.2381$$

Null Hypothesis Ho: $\mu_1 = \mu_2$

there is no diff b/w Population variances

Calculation

Here $S_2^2 > S_1^2$

$$\therefore F = \frac{S_2^2}{S_1^2} = \frac{5.2381}{3.75} = 1.3968$$

Table value of $F(6, 8) = 3.58$

$$\boxed{\begin{aligned} Df &= (n_1 - 1, n_2 - 1) \\ &= (9-1, 7-1) \\ &= (8, 6) \\ &\text{Here } S_2^2 > S_1^2 \\ \therefore Df &= (n_2 - 1, n_1 - 1) \end{aligned}}$$

Conclusion Cal F < Tab F

Accept Ho

Problem: 3

In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations was 102.6. Test whether this difference is significant at 5% LOS, given that the 5 percent of F for $v_1 = 7$ & $v_2 = 9$ degrees of freedom is 3.29

Ans

Given

$n_1 = 8$	$n_2 = 10$
$\Sigma(x_1 - \bar{x}_1)^2 = 84.4$	$\Sigma(x_2 - \bar{x}_2)^2 = 102.6$
$S_1^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1} = 12.06$	$S_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1} = 11.42$

Here $S_1^2 > S_2^2$

Null Hypothesis H_0: There is no significant difference b/w the variances of the samples.

$$D.f = (n_1 - 1, n_2 - 1)$$
$$= (8-1, 10-1) = (7, 9)$$

Calculation: $F = \frac{S_1^2}{S_2^2} = \frac{12.06}{11.42} = 1.06$

Table value of $F(7,9)$ at 5% $= 3.29$

Conclusion: Here cal.F < Tab F

Accept H_0

Problem: 4

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 this difference is significant at 5% LOS.

Ans:

Null Hypothesis H_0: There is no significant difft b/w the variances.

Given $n_1 = 10$ | $n_2 = 12$

$\Sigma(x_1-\bar{x}_1)^2 = 120$ | $\Sigma(x_2-\bar{x}_2)^2 = 314$

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = 13.33 \qquad S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = 28.55$$

$$Df = (n_1-1, n_2-1) = (10-1, 12-1) = (9, 11)$$

Here $S_2^2 > S_1^2$ $\therefore Df = (n_2-1, n_1-1) = (11, 9)$

Calculation

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

Table value of $F_{(11,9)}$ at 5% $= 3.11$

Conclusion: Cal F < Tab F

$\therefore$ Accept H_0.

Problem: 5

Two samples of sizes 9 & 8 gives the sum of squares of deviations from their respective means equal to 160 & 91 respectively. Can they be regarded as drawn from the same normal population?

A/M'2018
N/D'2019.

Answer:

Null Hypothesis H_0:
Samples are drawn from the same normal population.

Given $n_1 = 9$ | $n_2 = 8$

$\Sigma(x_1-\bar{x}_1)^2 = 160$ | $\Sigma(x_2-\bar{x}_2)^2 = 91$

$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = \frac{160}{9-1} = 20$ | $S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = \frac{91}{8-1} = 13$

$Df = (n_1-1, n_2-1) = (9-1, 8-1) = (8, 7)$

Calculation: $F = \frac{S_1^2}{S_2^2}$ Here $S_1^2 > S_2^2$

$= \frac{20}{13} = 1.54$

Table value of $F(8, 7) = 3.73$

Conclusion:

Cal F < Tab F

Accept H_0.

Hence we conclude that the samples might have come from two populations having the same variance.

TEST BASED ON χ^2 - Distribution

Chi-square Test for Goodness of fit

Formula $\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$

where O - observed frequency

E - Expected frequency.

Note: In case of

fitting a Binomial Distribution, d.f. = n-1

fitting a Poisson Distribution, df = n-2

fitting a Normal Distribution, df = n-3

Problem: 1

A company keeps records of accidents. During a recent safety review, a random sample of 60 accidents was selected and classified by the day of the week on which they occured.

Day	Mon	Tue	Wed	Thur	Fri
No. of accidents	8	12	9	14	17

Test whether there is any evidence that accidents are more likely on some days than others

N/D 2018

Answer:

Null Hypothesis H_0: Accidents are equally likely to occur on any day of the week.

Alternate Hypothesis H_1: Accidents are not equally likely to occur on the days of the week

Expected frequency $E = \frac{60}{5} = 12$

Formula $\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$

where O - Observed frequency

E - Expected frequency.

						Total
O	8	12	9	14	17	-
E	12	12	12	12	12	-
$\frac{(O-E)^2}{E}$	1.333	0	0.75	0.333	2.083	4.499

Calculation $\psi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 4.499$

Table value of $\psi^2 = 9.488$ (5% LOS)

$Df = n-1$
$= 5-1$
$= 4$

Conclusion Here Cal ψ^2 < Tab ψ^2

$\therefore$ Accept H_0

Problem: 2

The following data gives the number of aircraft accidents that occured during the various days of the week. Find whether the accidents are uniformly distributed over the week.

Days	Sun	Mon	Tue	Wed	Thurs	Fri	Sat
No. of accidents	14	16	8	12	11	9	14

Answer:

Null Hypothesis H_0: The accidents are uniformly distributed over the week

Alternate Hypothesis H_1: The accidents are not uniformly distributed.

Expected frequency $E = \frac{84}{7} = 12$

								Total
O	14	16	8	12	11	9	14	84
E	12	12	12	12	12	12	12	—
$\frac{(O-E)^2}{E}$	0.333	1.333	1.333	0	0.083	0.75	0.333	4.165

Degrees of freedom $= n-1$
$= 7-1$
$= 6$

Calculation:

$$\psi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$$

$$= 4.165$$

Table value of $\psi^2 = 12.592$ (5% LOS)
$Df = 6$

Conclusion: cal. ψ^2 < Tab ψ^2

Accept H_0

Problem: 3

In 120 throws of a single die, the following distribution 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.

Answer:

Null Hypothesis H_0: The die is unbiased

Alternate hypothesis H_1: The die is biased.

Expected frequency $E = \frac{120}{6} = 20$

							Total
O	30	25	18	10	22	15	–
E	20	20	20	20	20	20	–
$\frac{(O-E)^2}{E}$	5	1.25	0.2	5	0.2	1.25	12.9

Degrees of freedom $= n-1$

$= 6-1$

$= 5$

Calculation $\psi^2 = \Sigma \frac{(O-E)^2}{E}$

$= 12.9$

Table value of $\psi^2 = 11.07$ (5% LOS)

$Df = 5$

Conclusion: Cal ψ^2 > Tab ψ^2

Reject H_0

Problem: 4 A sample analysis of examination results of 1000 students were made and it was found that 260 failed, 110 first class, 420 second class and rest obtained third class. Do these data support the general examination result in the ratio 2:1:4:3 N/D'19

Answer:

Null Hypothesis :

Given data support the general examination result.

ie. The results in the four categories are in the ratio 2:1:4:3

Alternate Hypothesis:

Given data not support the expected result.

To find E

Rough

$\frac{2}{2+1+4+3} \times$ Total stu.

$\frac{2}{10} \times 1000 = 200$

$\frac{1}{10} \times 1000 = 100$

$\frac{4}{10} \times 1000 = 400$, $\frac{3}{10} \times 1000 = 300$

					Total
O	260	110	420	210	-
E	200	100	400	300	-
$\frac{(O-E)^2}{E}$	18	1	1	27	47

Degrees of from $= n-1 = 4-1 = 3$

Calculation

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 47$$

Table value of $\chi^2 = 7.815$

Conclusion: Here cal χ^2 > Tab χ^2

Reject H_0

Problem:5

A survey of 320 families with 5 children each revealed the following distribution.

No. of boys	5	4	3	2	1	0	–
No. of girls	0	1	2	3	4	5	–
No. of families	14	56	110	88	40	12	–

Is this result consistent with the hypothesis that male and female births are equally probable?

Answer:

Null Hypothesis H_0: Male and female births are equally probable.

Alternate Hypothesis: Male and female births are not equally probable.

On the assumption H_0, the expected frequencies are given by the terms of $N(q+p)^n$

$$= 320\left[\frac{1}{2}+\frac{1}{2}\right]^5$$

$$= \frac{320}{32}\left\{5c_0 + 5c_1 + 5c_2 + 5c_3 + 5c_4 + 5c_5\right\}$$

$$= 10\{1+5+10+10+5+1\}$$

$\therefore$ The expected frequencies are

(10, 50, 100, 100, 50, 10

O	14	56	110	88	40	12	–
E	10	50	100	100	50	10	–
$\frac{(O-E)^2}{E}$	1.6	0.72	1	1.44	2	0.4	7.16

Degrees of freedom $= n-1 = 6-1$
$= 5$

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

$= 7.16$

Table value of $\psi^2 = 11.07$ (5% LOS)
Df = 5

Conclusion: Cal ψ^2 < Tab ψ^2

Accept Ho

ψ^2 Test (Independence of attributes)

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

(or)

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

$Df = (r-1)(c-1) = (2-1)(2-1) = 1$

Problem :1

Find if there is any association b/w extravagance in fathers and extravagance in sons from the following data.

	Extravagant Father	Miserly Father
Extravagant son	327	741
Miserly son	545	234

M/J '2013

Determine the co-efficient of association also.

Answer:

H_0: There is no significant difference b/w the extravagance in sons & fathers.

H_1: " significant "

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

Given $\alpha = 0.05$

$df = (r-1)(c-1) = 1$

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

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

$= 279.77$

Table value for $\chi^2 = 3.841$

Conclusion: 1) Cal χ^2 > Tab χ^2

Reject H_0

2) Coefficient of attributes

$$= \frac{ad-bc}{ad+bc}$$

$= -0.6814$

Problem: 2

From the following information state whether the condition of the child is associated with the condition of the house.

Condition of House

Condition of child	clean	dirty	Total
Clean	69	51	120
Fairly clean	81	20	101
Dirty	35	44	79
Total.	185	115	300

Answer:

Null Hypothesis H_0:

The given attributes are independent

Alternate Hypothesis H_1:

The given attributes are not independent.

Degrees of freedom $= (r-1)(s-1)$

$= (3-1)(2-1)$

$= 2$

Observed values:

			Total
	$a = 69$	$b = 51$	120
	$c = 81$	$d = 20$	101
	$e = 35$	$f = 44$	79
Total	185	115	300

To find the expected values.

		Total
$E(a) = \frac{\text{Row total} \times \text{column total}}{\text{Grand Total}} = 74$	$\frac{115 \times 120}{300} = 46$	120
$\frac{185 \times 101}{300} = 62$	$\frac{115 \times 101}{300} = 39$	101
$\frac{185 \times 79}{300} = 49$	$\frac{115 \times 79}{300} = 30$	79
Total 185	115	300 Grand total

To find χ^2 value

O	E	$\frac{(O-E)^2}{E}$
69	74	0.34
51	46	0.54
81	62	5.82
20	39	9.26
35	49	4
44	30	6.53
		26.49

Cal $\chi^2 = 26.49$

Table value of $\chi^2 = 5.991$

Conclusion: Cal $\chi^2 >$ Tab χ^2

$\therefore$ Reject H_0.

Problem : 3

1000 students at college level were graded according to their IQ and their economic conditions what conclusions can you draw from the following data.

IQ level

Economic conditions	High	Low
Rich	460	140
Poor	240	160

Answer:

Null Hypothesis H_0:

The given attributes are independent

Alternate Hypothesis H_1:

The given attributes are not independent

Degrees of freedom $= (r-1)(c-1)$

$= (2-1)(2-1)$

$= 1$

Formula $\psi^2 = \dfrac{(ad-bc)^2 (a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$

$$= \frac{(460 \times 160 - 140 \times 240)^2 \{460 + 140 + 240 + 160\}}{(460+140)(240+160)(460+240)(140+160)}$$

$= 31.75$

Table value of $\psi^2 = 3.841$

Conclusion

Here cal χ^2 > Tab χ^2

Reject H_0.

---

4) What are the applications / uses of χ^2 distribution

Ans:

(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 co-efficient

— x — x —