classmate
Date
Page

25/6/2024

UNIT-IV

Sampling and statistical Interference

Sampling theory is a study of relationship between populat-ion and samples drawn from the population.

Objectives of sampling

The main objectives of sampling are:

i) Gathering maximum information about the population with minimum effort, cost & time.

ii) To obtain the best possible values of the parameters under specific conditions.

iii) Determine the reliability of these estimates.

Definitions:

① population / universe:
A large collection of individuals or items or qualities or attributes or results which can be numerically specified is called as population / universe.

* A population containing a finite no. of individuals / items / members is a finite population.

* A population containing with an infinite no. of individuals / items / members is called an infinite population.

② parameter:
* A parameter is a statistical measure or constant such as mean, variance, standard deviation etc, obtained from the population.

③ Sample:
A sample is a subset of population or a small selection selected from the population.

④ statistic:
A statistic is a statistical measure or constant such as mean, variance, standard deviation etc obtained from the sample.

⑤ Population size:
The no. of individuals or items in a population is called the population size and it is denoted by 'N'.

⑥ Sample size:
The no. of individuals/items in a sample is called sample size and it is denoted by 'n'.

⑦ Sampling:
The process of selecting a sample from the population is called sampling.

⑧ Random Sampling:
The selection of an individual / or an item from the population, in such a way that, each item has an equal chance of being selected is called as Random sampling.

Suppose, we take a sample size of 'n' from a population size 'N', then we will have

$[{}^{N}C_{n}]$ possible samples without replacement and

$[N^{n}]$ possible samples with replacement

* A sample obtained through the random sampling is called a "Random Sample".

* If $n \geq 30$, then the sample is called "large sample" whereas if $n < 30$, the sample is called a "small sample".

Sampling with Replacement and without Replacement:

* An item of the population may be selected more than once, is called as "Sampling with Replacement",

* On the other hand, if the item of the population can't be chosen more than once is called "Sampling without Replacement".

Sampling Distribution:

It describes how a statistic will vary from one sample to another sample of same size.

Sampling Distribution of Means:

"Sample Mean" is a statistic & we discuss the sampling distribution of this statistic.
We consider all possible random samples of size 'n' and determine the mean of each of these samples.

Here, we discuss the sampling distribution of sample Means for the 2 possible types of Random sampling [with or without replacement] associated with a finite population.

classmate

Date
Page

* A population consists of 4 numbers 3, 7, 11, 15. consider all possible samples of size 2 which can be drawn
   i) with replacement
   ii) without replacement from this population.

In both the cases find
   a) population mean b) the population standard deviation
   c) the mean of sampling distribution of means d) std error of means

NOTE: The standard deviation of a sampling distribution of a statistic is called as standard error

a) population mean $= \frac{\Sigma(3,7,11,15)}{4} = \frac{10+26}{4} = \frac{36}{4} = 9$

b) population standard deviation $= \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{36+4+4+36}{4}} = \sqrt{20}$

$= 2\sqrt{5} = 4.47$

sampling without replacement:

samples: (3, 7) (7, 11) (11, 15)
(3, 11) (7, 15)
(3, 15)

Samples:	Mean
(3, 7)	5
(3, 11)	7
(3, 15)	9
(7, 11)	9
(7, 15)	11
(11, 15)	13

classmate

Date ________
Page ________

sampling distribution of means:

x	5	7	9	11	13
f	1	1	2	1	1

c) $\mu_{\bar{x}} = \frac{\sum f \cdot x}{\sum f} = \frac{5+7+18+11+13}{6} = 9$

d) $\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} = \sqrt{\frac{16+4+0+4+16}{6}}$

$= \sqrt{\frac{20}{3}} = 2.5819$

so, standard error $= 2.5819$

sampling with Replacement

samples:	Mean:
(3,3)	3
(3,7)	5
(3,11)	7
(3,15)	9
(7,3)	5
(7,7)	7
(7,11)	9
(7,15)	11
(11,3)	7
(11,7)	9
(11,11)	11
(11,15)	13
(15,3)	9
(15,7)	11
(15,11)	13
(15,15)	15

sampling distribution of means

x	3	5	7	9	11	13	15
f	1	2	3	4	3	2	1

c) $\mu_{\bar{x}} = \frac{\sum f x}{\sum f} = \frac{144}{16} = 9$

d) $\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} = \frac{\sqrt{1(3-9)^2 + 2(5-9)^2 + \cdots}}{\sqrt{16}}$

$\sigma = 3.16.$

standard error $= 3.16$

NOTE: standard error plays a very important role in the theory of large samples and forms a basis of hypothesis.

⑨ Degrees of Freedom:

The no. of degrees of freedom, usually denoted by 'ν' is the no. of values in a cell which may be assigned arbitrarily. It can be interpreted as the no. of independent values generated by a sample of small size for estimating a population parameter.

⑩ Testing of Hypothesis:

Statistical estimation & testing of statistical hypothesis are 2 important aspects in statistical inference.

Definition:

A positive statement about the parameter of the population which may or may not be true is called statistical hypothesis.

* It is made on the basis of information obtained by experimentation.
* Testing of Hypothesis is a procedure for deciding whether to accept or reject the hypothesis.

* procedures which enables us to decide whether to accept or reject the hypothesis are called "Test of Hypothesis" are also called "Test of significance".

⑪ Null Hypothesis:

The hypothesis formulated for the purpose of rejection under the assumption that it is true. is called the Null hypothesis.

It is denoted by 'H_0'.

* Any hypothesis which is complement to the Null hypothesis is called an "alternate hypothesis" and is denoted by 'H_1'

* Suppose we want to test the null hypothesis that there is no significant difference between population mean and the sample mean.

If 'μ' is the population mean and 'μ_0' is the sample mean then,

$\boxed{H_0 : \mu = \mu_0}$

then the alternate hypothesis will be either

i) $\boxed{H_1 : \mu \neq \mu_0}$ or ii) $\boxed{H_1 : \mu > \mu_0}$ or iii) $\boxed{H_1 : \mu < \mu_0}$

The hypothesis in one (i) is a 2-tailed alternate hypothesis. The hypothesis in ii) is a right tailed alternate hypothesis. The hypothesis in iii) is a left tailed alternate hypothesis.

∴ i) is called 2-tailed test hypothesis.
ii) and iii) are called single tailed test hypothesis

Type 1 & Type 2 errors.:

When a null hypothesis H_0 is tested against an alternate hypothesis H_1, we are likely to commit, one of the two types of errors. as explained,

	Accept	Reject
H0 : True	correct decision	Type-I error
H0: False	Type-II error	correct decision

P(Rejecting Ho when it is true) = P(Type-I-error) = α
P(Accepting Ho when it is false) = P(Type-II-error) = β.

Level of significance

* In type-I error, 'α' is called the level of significance, i.e., the probability level below which, we reject the null hypothesis

* Usually two levels of significance are there i.e, 5% level of significance & 1% level of significance.

When we take 5% as level of significance, then the probability of committing type-I error is 0.05,
similarly if we take
1% level of significance, probability of committing type-I error is 0.01.

A test of statistical hypothesis Ho, involves determination of region in sample space such that the hypothesis will be rejected,

If the sample point falls within the region and it will be accepted if the sample point falls outside the region.
~~it will be~~

This region is called the "critical region" or "region of rejection".
The rest of the sample space is called "region of acceptance"

* The value of a test statistic which seperates the critical region & the region of acceptance is called critical value or significant value.

procedure for Testing of Hypothesis for large samples

We know that, the standard normal variable 'z' is defined by,

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

where $\bar{x}$ is mean of the sample
μ is population mean
σ is standard deviation of the population
'n' is the sample size.

This value of 'z' is called the "Test statistic".

* The critical value at 5% level of significance is 'α'
so, ⇒ z_α

i) If $|z| < z_\alpha$, we accept the hypothesis (null) H_0.
ii) If $|z| > z_\alpha$, then reject the null hypothesis H_0

z_α at 5% level of significance $= 1.96$

* At 1% level of significance, $z_\alpha = 2.58$

Confidence limits for μ:

* 95% confidence limits for μ is

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

i.e., $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

To accept or reject the hypothesis.

* 99% confidence limits of μ is

$$\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$$

$$\text{i.e., } \bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

* The mean life time of sample of 100 fluoroscent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. The company claims that the average lift of the bulbs produced by its 1600 hours. using the level of significance of 0.05, is the claim acceptable?

$\bar{x} = 1570$

$\sigma = 120$ hrs

$\mu = 1600$ hrs

$n = 100$

$H_0 : \mu = \mu_0$

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

$$Z = \frac{1570 - 1600}{120/\sqrt{100}}$$

$$Z = -2.5$$

Z_α at 0.05 significance level = 1.96

$$|Z| > Z_\alpha$$

So, we should reject the hypothesis H_0.

So, his claim is not acceptable.

classmate

Date
Page

* A dice is thrown 9000 times and of these 3240 yielded a 3 or 4. Is this consistent, show that the dice can be unregarded as unbiased one. (use 5% level of significance)

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

$$\mu = np \; \& \; \sigma = \sqrt{npq}$$

Here, H_0 = dice is unbiased

H_1 = The dice is biased

P (getting 3 or 4) = P(getting 3) + P(getting 4)

$$p = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Expected no. of observations (μ) = np

$$= 9000 \times \frac{1}{3} = 3000$$

$\bar{x} = 3240$.

$$\sigma = \sqrt{\overset{1000}{\cancel{9000}} \times \frac{1}{\cancel{3}} \times \frac{2}{\cancel{3}}} = \sqrt{2000}$$

$$Z = \frac{\bar{x} - np}{\sqrt{npq}}$$

$$Z = \frac{3240 - 3000}{\sqrt{2000}/\cancel{\sqrt{1000}}} = \frac{240}{\sqrt{2}} = \frac{240 \times \cancel{3}}{\sqrt{2}}$$

$$Z = 120 \times \cancel{3} \times \sqrt{2}$$

$$Z = \frac{360\sqrt{2}}{3} = 120\sqrt{2}$$

at 5% level of significance

$Z_\alpha = 1.96$

$|Z| > Z_\alpha$

So, we reject the null hypothesis H_0.

<u>NOTE:</u> Let 'x' be the observed number of success in a sample of size n, and $\mu = np$ be the expected no. of success. The associated standard normal variate z is defined as

$$z = \frac{\bar{x} - np}{\sqrt{npq}}$$

classmate

Date ______
Page ______

* A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that coin is unbiased at 5% level of significance.

Ho: The coin is unbiased.

H1: The coin is ~~un~~biased

$n = 400$

$P(\text{getting head}) = \frac{1}{2} = p$

$\mu = np = 400 \times \frac{1}{2} = 200$

$\sigma = \sqrt{npq} = \sqrt{400 \times \frac{1}{4}} = 10$

$\bar{x} = 216$

$Z = \frac{\bar{x} - \mu}{\sigma} = \frac{216 - 200}{10} = \frac{16}{10} = 1.6$

at 5% level of significance,

$Z\alpha = 1.96$

$|Z| < Z\alpha$

NOTE: The probable limits at 5% level of significance is $p \pm 1.96\sqrt{\frac{pq}{n}}$

The probable limits at 1% level of significance is $p \pm 2.58\sqrt{\frac{pq}{n}}$

* A survey was conducted in a slum locality of 2000 families by selecting a sample of size 800. It was revealed that 180 families were illiterates. Find the probable limits of the illiterate families in population of 2000. (5% level of significance)

$P(\text{selecting illiterates}) = \frac{180}{800} = 0.225$

$q = 0.775$

The probable limits are

$$0.225 \pm 1.96 \sqrt{\frac{0.225 \times 0.775}{800}}$$

$= (0.196, 0.2539)$

For 2000 families,

$(0.196 \times 2000, 0.2539 \times 2000)$

$= (392.2, 507.8)$

$\approx (392, 508)$ at 5% level of significance.

at 1% level of significance,

$$0.225 \pm 2.58 \sqrt{\frac{0.225 \times 0.775}{800}}$$

$= (0.1869, 0.2630)$

For 2000 families,

$(0.1869 \times 2000, 0.2630 \times 2000)$

$\approx (374, 526)$ at 1% level of significance.

* A sample of 900 members is found to have a mean of 3.4cm. can it be reasonably regarded as a truly random sample from a large population with mean 3.25 cm & SD 1.61cm.

$\bar{x} = 3.4$

$\mu = 3.25$

$\sigma = 1.61$

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

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

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{3.4 - 3.25}{1.61 / \sqrt{900}}$$

$= 2.8$

at 5% $Z_\alpha = 1.96$

$|Z| > Z_\alpha$

We reject H_0, sample is not from a large population

NOTE: * If $\overline{x_1}$ & $\overline{x_2}$ are means of 2 sample sizes n_1, n_2 and 'σ' is the standard deviation of the population then $Z = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

* If $\overline{x_1}$ & $\overline{x_2}$ are means of samples of sizes n_1, n_2 respectively, and σ_1, σ_2 are the standard deviations of the above samples then $Z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

* The means of two large samples of sizes 1000 and 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

By data

$\overline{x_1} = 67.5$

$\overline{x_2} = 68.0$

$\sigma = 2.5$

$n_1 = 1000$

$n_2 = 2000$

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$Z = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{2000} + \frac{1}{1000}}}$$

$$Z = -5.1639$$

$Z\alpha$ (5% level of significance) $= 1.98$

$|Z| > Z\alpha$

Reject H_0

Samples are not drawn from sample population.

* The mean yield of wheat from a district A was 210 pounds with S.D. 10 pounds per year/acre from a sample of 100 plots. In another district the mean yield from 220 pounds with S.D 12 pounds from a sample of 150 plots.
Assuming that the S.D of yield in the entire state was 11

pounds, test whether there is any significant difference between the mean yield of crops in the two districts.

$H_0: \mu_1 = \mu_2$

$\sigma_1 = \sigma_2 = \sigma$

By data, $n_1 = 100$ | $n_2 = 150$

$\bar{x}_1 = 210$ | $\bar{x}_2 = 220$

$\sigma_1 = 10$ | $\sigma_2 = 12$.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{210 - 220}{\sqrt{\frac{100}{100} + \frac{144}{150}}}$$

$$Z = -7.14$$

$Z_\alpha = 1.96$ at 5% level of significance.

$|Z| > Z_\alpha$

Reject the H_0

There is a significant difference b/w the samples

Test for small samples:

consider, Let $x_1, x_2, x_3, \ldots x_n$ be a random sample of size 'n' with mean $\bar{x}$ and standard deviation 's'

The test statistic for 't' is:

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

here 'μ' is the mean of the population

The test statistic 't' is tested at any level of significance with $\nu = n-1$ (degrees of freedom)

at 5% LOS, $t_\alpha = 2.160$

Here, we take,

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

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

* The average breaking strength of steel rods is specified to be 18.5 thousand pounds. To test this a sample of 14 rods was tested. The mean and standard deviation obtained were 17.85 and 1955 respectively. Is the result of the experiment significant with 95% confidence?

$\mu = 18.5$

$\bar{x} = 17.85$

$s = 1.955$

$n = 14$

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

$$t = \frac{17.85 - 18.5}{1.955/\sqrt{14-1}} = -1.198$$

degrees of freedom $\nu = n-1 = 14-1 = 13$

$t_\alpha = 2.160$

$|t| < t_\alpha$ at 5% LOS

Accept H_0

* The heights of 10 males of a given locality are found to be 175, 168, 155, 170, 152, 170, 175, 160, 160 and 165 cms. Based on this sample, find the 95% confidence limits for the height of males in that locality.

$n = 10$

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ ⇒ confidence limits for z-test

for t-test ⇒ $\bar{x} \pm t_{0.05} \sqrt{\frac{s^2}{n}}$

$t_{0.05} = 2.26$ at 95% confidence level.

$\nu = 10-1 = 9$ d.f

x	$(x-\bar{x})$	$(x-\bar{x})^2$
175	10	100
168	3	9
155	-10	100
170	5	25
152	-13	169
170	5	25
175	10	100
160	-5	25
160	-5	25
165	0	0

$\bar{x} = 165$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{10-1} = \frac{578}{9} = 64.22.$$

$$\text{confidence limits} \Rightarrow 165 \pm 2.26 \cdot \frac{(8.01)}{\sqrt{10}}$$

$$= (170.724,\ 159.2754)$$

* A certain stimulus administered to each of the 12 patients resulted in the following change in blood pressure, 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. can it be concluded that the stimulus will increase the blood pressure?

$n = 12$

ν (degrees of freedom) $= n-1 = 12-1 = 11$

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

$x-\bar{x}$	$(x-\bar{x})^2$
5	25
2	4
8	64
-1	1
3	9
0	0
6	36
-2	4
1	1
5	25
0	0
4	16

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

H_0 : stimulus increases the b.p.

$\bar{x} = 2.58$

$$s^2 = \frac{\Sigma (x-\bar{x})^2}{n-1} = \frac{185}{11} = 16.8181$$

$s = 4.1009$

$$t = \frac{2.58}{4.1009/\sqrt{11}} = 2.0865$$

$|t| < t_\alpha$

so, H_0 is accepted.

* A fertilizer mixing machine is set to give 12 kg of nitrate for quintal bag of fertilizer. Ten 100kg bags are examined, the percentage of nitrate per bag as follows : 11, 14, 13, 12, 13, 12, 13, 14, 11, 12. Are there any reasons to believe that the machine is defective? value of t for 9 degree of freedom is 2.262.

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

$n-1 = 9$

$\mu = 12$

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

$\bar{x} = 12.5$

x	$x-\bar{x}$	$(x-\bar{x})^2$
11	-1.5	2.25
14	1.5	2.25
13	0.5	0.25
12	-0.5	0.25
13	0.5	0.25
12	-0.5	0.25
13	0.5	0.25
14	1.5	2.25
11	-1.5	2.25
12	-0.5	0.25

$$\Sigma(x-\bar{x})^2 = 10.5$$

$$s^2 = \frac{1}{9} \times 10.5 = 1.1667$$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} = \frac{12.5-12}{1.0801/\sqrt{9}} = \frac{0.5 \times 3}{1.0801} = \frac{1.5}{1.0801}$$

$$t = 1.3887$$

t calculated < tabulated.

T-test for difference between two sample means

Given, two independent samples $x_1, x_2, \ldots, x_{n_1}$

$y_1, y_2, \ldots, y_{n_2}$

with means $\bar{x}$ & $\bar{y}$ and standard deviations s_x and s_y respectively from a normal population with same variance, we have to test the hypothesis that the population means μ_1 & μ_2 are same.

The test statistic 't' is given by,

$$t = \frac{\bar{x}-\bar{y}}{s \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$ where,

classmate

Date
Page

$$\sigma_S^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)s_x^2 + (n_2-1)s_y^2\right]$$

where the degrees of freedom $\nu = n_1+n_2-2$

we consider, $H_1: \mu_1 \neq \mu_2$

$H_0: \mu_1 = \mu_2$

* Two independent samples of sizes 8 & 7 contained the following values:

sample 1	19	17	15	21	18	18	16	14
sample 2	15	14	15	19	15	18	16	–

Is the difference b/w the sample means significant?

$H_0: \mu_1 = \mu_2$

$\bar{x} = 17.$

$\bar{y} = 16$ $\qquad n_1 = 8,\ n_2 = 7$

$\Sigma(x-\bar{x})^2 = 36$ $\qquad \nu = n_1+n_2-2 = 13$

$\Sigma(y-\bar{y})^2 = 20.$

$$s_x^2 = \frac{1}{8-1}\times 36 = 5.14$$

$$s_y^2 = \frac{1}{7-1}\times 20 = 3.33$$

$$\sigma_S^2 = \frac{1}{13}\left[7(5.14) + 6(3.33)\right]$$

$\Rightarrow \nu = 13$

$\sigma_S^2 = 4.3046$

$\sigma_S = 2.0747$

$$t = \frac{\bar{x}-\bar{y}}{\sigma_S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{17-16}{2.0747\sqrt{\frac{1}{8}+\frac{1}{7}}} = 0.9313$$

$t_{calculated} < t_{tabulated}$.

$t_{tabulated}$ for $\nu = 13 \Rightarrow 2.16$.

classmate

Date

Page

Q.45) Question bank

	size	Mean	SD
Sample 1	8	1234h	36h
Sample 2	7	1036h	40h

By data.

$n_1 = 8$, $n_2 = 7$

$\bar{x} = 1234$ $\bar{y} = 1036$

$s_x = 36$ $s_y = 40$

$$\sigma_s^2 = \frac{1}{13}\left[7(36)^2 + 6(40)^2\right]$$

$$\sigma_s^2 = 1436.30$$

$$\sigma_s = 37.89$$

$$t = \frac{\bar{x} - \bar{y}}{\sigma_s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad ; \quad t = \frac{1234 - 1036}{37.89\sqrt{\frac{1}{8} + \frac{1}{7}}}$$

$$t = 10.09$$

$\nu = 13$, t tabulated $= 2.16$

t calculated > t tabulated $= 2.16$

Ho : Rejected.

so, type 2 bulbs are superior to type 1 bulbs.

Questionbank

50.

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I test	23	20	19	21	18	20	18	17	23	16	19
Marks II test	24	19	22	18	20	22	20	20	23	20	17

Boys	Marks I Test	Marks II Test	$d(T_1 - T_2)$	$(d-\bar{d})^2$
1	23	24	-1	0
2	20	19	+1	4
3	19	22	-3	4
4	21	18	+3	16
5	18	20	-2	1
6	20	22	-2	1
7	18	20	-2	1
8	17	20	-3	4
9	23	23	0	1
10	16	20	-4	9
11	19	17	+2	9
			$\bar{d} = \frac{-11}{11} = -1$	$\Sigma(d-\bar{d})^2 = 50$

$$S^2 = \frac{1}{n-1} \Sigma(d-\bar{d})^2 = \frac{50}{10} = 5$$

$$S = 2.236$$

$$t = \frac{\bar{d}}{2.236/\sqrt{10}} = \frac{-1}{2.236\sqrt{10}} = -1.414$$

$t_{calculated} = -1.414$. (one-tailed)

t_{10} at 5% level of significance = 1.812

$t_{calculated} < t_{tabulated}$

Accept H_0.

Yes, the students have been benifited by extra coaching.

F-test (Fischer's test):

Let $(x_1, x_2, \ldots\ldots, x_{n_1})$ & $(y_1, y_2, \ldots\ldots, y_{n_2})$ be the values of two independent random samples drawn from 2 normal populations having variances σ^2

The test statistic 'F' is denoted by

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

$$S_1^2 > S_2^2$$

where $S_1^2 = \frac{1}{n_1 - 1} \sum (x - \bar{x})^2$

$S_2^2 = \frac{1}{n_2 - 1} \sum (y - \bar{y})^2$

The degrees of freedom are $\nu = (n_1 - 1, n_2 - 1)$

* The two samples of sizes $n_1 = 9$, $n_2 = 8$, the sums of squares of deviations from their respective means equal to 160 & 91 respectively. Can they be regarded as drawn from the same normal population?

$n_1 = 9$, $n_2 = 8$

$\Sigma(x - \bar{x})^2 = 160$, $\Sigma(y - \bar{y})^2 = 91$

$H_0: S_1^2 = S_2^2 = \sigma^2$

$$F = \frac{S_1^2}{S_2^2} = \frac{160/8}{91/7} = \frac{20 \times 7}{91} = \frac{20}{13} = 1.53$$

$$\nu = (n_1 - 1, n_2 - 1) = (8, 7)$$

$F_{8,7} = 3.73$ at 5% LOS

calculated F < Tabulated F

Accept H_0

* The nicotine contents in two random samples of tobacco are given bellow :

x	y	$(x-\bar{x})^2$	$(y-\bar{y})^2$
21	22	12.96	49
24	27	0.36	4
25	28	0.16	1
26	30	1.96	1
27	31	5.76	4
-	36	-	49

$\bar{x} = 24.6$ $\bar{y} = 29$

$\Sigma(x-\bar{x})^2 = 21.2$

$\Sigma(y-\bar{y})^2 = 108$

$s_1^2 = \frac{1}{4}(21.2) = 5.3$

$s_2^2 = \frac{1}{5}(108) = 21.6$

$F = \frac{s_2^2}{s_1^2}$ (as $s_2^2 > s_1^2$)

$F = \frac{21.6}{5.3} = 4.0754$

$\nu(5,4) = 6.26$

F calculated < F tabulated

Accept H_0.

5.

x	y	$(x-\bar{x})^2$
28	29	
30	30	
32	30	
33	24	
33	27	
29	29	
34	-	

continue as previous

$\bar{x} = 31.28$

$\bar{y} = 28.166$

56.

x	y	$(x-\bar{x})^2$	$(y-\bar{y})^2$
20	17	4	49
16	23	36	1
26	32	16	64
27	25	25	1
23	22	1	4
22	24	0	0
18	28	16	16
24	6	4	324
25	31	9	49
19	33	9	81
-	20	-	16
-	27	-	9

$\bar{x} = 22$

$\bar{y} = 24$

$$\Sigma(x-\bar{x})^2 = 120$$

$$\Sigma(y-\bar{y})^2 = 614$$

$$S_1^2 = \frac{1}{9} \times 120 = 13.333$$

$$S_2^2 = \frac{1}{11} \times 614 = 55.818.$$

$$F = \frac{55.818}{13.333} = 4.1874$$

$F_{9,11}$ at 5% LOS = (if no 11 value in table)

so, take 10 or 12.

$F_{9,11} = 2.90$

$F_{calculated} > F_{tabulated}$

so, Reject H_0

χ^2- test (Chi² test)

If $O_1, O_2, O_3, \ldots, O_n$ be a set of observed frequencies and $E_1, E_2, E_3, \ldots, E_n$ are the set of expected frequencies then the test statistic χ^2 is given by,

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \cdots + \frac{(O_n-E_n)^2}{E_n}$$

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

where degrees of freedom $\nu = n-1$

* Five coins are tossed 320 times. The number of heads observed is given below. Examine whether the coin is unbiased

No. of heads	0	1	2	3	4	5	Total
Frequency	15	45	85	95	60	20	320

$p = P(\text{getting Head}) = \frac{1}{2}$, $P(\text{tail}) = \frac{1}{2} = q$

$n = 5$

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

$= {}^5C_x \left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{5-x} = {}^5C_r \times \frac{1}{32}$

$N = 320$

Expected frequencies are $E_r = 320 \times \frac{1}{32} \times {}^5C_r$

$E_x = 10 \times {}^5C_r$

at $x = 0, 1, 2, 3, 4, 5$

$E_0 = 10$ $\quad$ $E_4 = 50$

$E_1 = 50$ $\quad$ $E_5 = 10$

$E_2 = 100$

$E_3 = 100$

O	E
15	10
45	50
85	100
95	100
60	50
20	10

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{(5)^2}{10} + \frac{(5)^2}{50} + \frac{(15)^2}{100} + \frac{(5)^2}{100} + \frac{(10)^2}{50} + \frac{(10)^2}{10}$$

$$= 2.5 + 0.5 + 2.25 + 0.25 + 2 + 10$$

$$= 5.5 + 12$$

$$\chi^2 = 17.5 \quad \text{(calculated)}$$

$$\nu = n-1 \Rightarrow 5-1 = 4$$

χ^2 (tabulated) $= 9.485$ (5% LOS)

χ^2 calculated $>$ χ^2 tabulated

Ho is Rejected.

so, coin is biased.

* Fit a poisson Distribution for following data test the goodness of fit

x:	0	1	2	3	4	5	6	Total
f:	273	70	30	17	7	2	1	390

$$\text{Mean } (\mu) = \frac{\sum x f}{\sum f} = \frac{70+60+21+28+10+6}{390}$$

$$\Rightarrow \frac{195}{390} = 0.5$$

$\mu = 0.5$

classmate

Date
Page

$$P(X=r) = \frac{e^{-\mu}.\mu^r}{r!} = \frac{e^{-0.5}.(0.5)^r}{r!}$$

Expected frequencies:

$$E_r = 390 \times \frac{e^{-0.5}.(0.5)^r}{r!}$$

$E_0 = 236.5469$

$E_1 = 118.2734$

$E_2 = 29.56$

$E_3 = 4.9280$

$E_4 = 0.6160$

$E_5 = 0.0616$

$E_6 = 0.0051$

since some of the observed frequencies are less than 10, so we group those frequencies together.

O	E	$(O_i - E_i)^2/E_i$
273	236.5469	5.61761
70	118.2734	19.7028
30	29.56	6.54×10^{-3}
(7+7+2+1) 17	5.6	23.2071

$\chi^2 = 48.525$ Calculated

NOTE: For poisson distribution, to test the goodness of fit, we consider

$\nu = n-2$

$\nu = 4-2 = 2$

χ^2tabulated $= 5.99$ (0.05% LOS)

χ^2 (calculated) > χ^2 tabulated

Ho is Reject.

The fit is not good.

* Among 64 offspring of a certain cross between Guinea pigs 34 were red, 10 were black & 20 were white. According to the genetic model these numbers should be in the ratio 9:3:4. Are the data consistent with the model at 5% level.

$16x = 64$

$x = 4$

O	E	$(O_i - E_i)^2/E_i$
34	36	0.111
10	12	0.333
20	16	1

χ^2 calculated $= 1.44$

Expected frequencies are:

$E_1 = 64 \times \frac{9}{16} = 36$

$E_3 = 64 \times \frac{4}{16} = 16$

$E_2 = 64 \times \frac{3}{16} = 12$

$n = 3$

$\nu = n - 1 = 2$

χ^2 at $\nu = 2$ (0.05 LOS)

$= 5.99$

χ^2 calculated $< \chi^2$ tabulated

Ho : accept.

NOTE: For testing of attributes,

	Y	N	
Y	a	b	a+b
N	c	d	c+d
	a+c	b+d	

$N = (a+b) + (c+d)$

(or)

$(a+c) + (b+d)$

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

$\nu = n - 1$ (n = no. of attributes)

problem 62 Q.B.

	Got disease	Did not get disease	Sum
vaccinated	9	42	51
Not vaccinated	17	28	45
	26	70	

$\boxed{N = 96}$

classmate

Date

Page

$$\chi^2 = \frac{96(252 - 714)^2}{(51)(45)(26)(70)} = 4.9057$$

$\chi^2_{0.05}$ at $\nu = n-1 = (2-1) = 1 = 3.84$

$$\begin{array}{r} 134 \\ 68 \\ \hline 714 \end{array}$$

χ^2 calculated $>$ χ^2 tabulate

H_0 : rejected

so, not independent.