part of population = sample.

8/8/23

MODULE - 3

SAMPLING THEORY

Random Sampling:

A large collection of individuals or attributes or numerical data can be understood as a population or universe.

A finite subset of universe is called a Sample.

The number of individuals in a sample is called a Sample size.

The process of selecting a sample from the population is called as sampling.

- The selection of an individual or item from the population in such a way that each has the same chance of being selected is called Random Sampling.
- Sampling where a member of the population may be selected more than once is called as Sampling with replacement, on the other hand if a member cannot be chosen more than once is called as Sampling without replacement.

imp.

Testing of Hypothesis:

In order to arrive at a decision regarding the population through a sample of the population we have to make certain assumptions referred to as hypothesis which may or may not be true.

* → The hypothesis formulated for the purpose of its rejection under the assumption that its true is called the Null hypothesis denoted by H_0.

* → Any hypothesis which is complimentary to the null hypothesis is called Alternative hypothesis, denoted by H_1.

* imp

Errors :

In the context of testing of hypothesis, there are basically 2 types of errors we can make, Type I error and Type II error.

Type I error : means rejection of hypothesis which should have been accepted.

Type II error : means accepting the hypothesis which should have been rejected

	Accepting the hypothesis	Rejecting hypothesis.
Hypothesis true	correct dec. (no error)	Wrong dec (Type I)
Hypothesis false	Wrong dec (Type II)	Correct dec (no error)

A region which amounts to the rejection of null hypothesis is called Critical Region or region of rejection

Significance level :

The probability level, below which leads to the rejection of the hypothesis is known as significance level.

Probability is conventionally fixed at 0.05 or 0.01 i.e, 5% or 1%. These are called significance levels.

Level of significance is the probability of Type I error.

Test of significance and Confidence intervals :

The process which helps us to decide about the acceptance or rejection of the hypothesis is called the test of significance:

Lets us assume that we have a normal population with mean μ and S.D σ. If $\bar{x}$ is the sample mean of a random sampling of size n, the quantity z defined by

$$z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$$ is called Standard Normal Variate (SNV).

$$\Rightarrow -1.96 \leq \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} \leq 1.96$$ (95% of area lies b/w $z = -1.96$ & $z = +1.96$.)

$$-\frac{\sigma}{\sqrt{n}} 1.96 \leq \bar{x} - \mu \leq \frac{\sigma}{\sqrt{n}} 1.96.$$

$$\mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}}(1.96) \quad \& \quad \bar{x} - \frac{\sigma}{\sqrt{n}}(1.96) \leq \mu$$

$$\therefore \bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Similarly, $\bar{x} - 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$

($\because$ 99% of area lies b/w -2.58 and $+2.58$)

<u>One tailed</u> and <u>Two tailed tests</u>:

In our test of acceptance or non acceptance of a hypothesis we concentrate on the value of z on both sides of the mean. This can be categorized that focus of attention lies in the two "tails" of the distribution and hence such a test is called <u>two tailed test</u>.

Sometimes we will be interested in the extreme values to only one side of the mean in which case the region of significance will be a region to one side of the distribution. Obviously the

/ /

area of such a region will be equal to the level of significance itself. Such a test is called a one-tailed test.

Test	critical values of z	
	5 % level	1 % level
one-tailed test	-1.645 or 1.645	-2.33 or 2.33
two-tailed test	-1.96 or 1.96	-2.58 or 2.58

imp

(1) <u>Test of significance for small samples:</u>

<u>Degrees of Freedom:</u>

The number of degrees of freedom (d.f) usually denoted by 'ν'. is the no. of values in a set which may be assigned arbitrarily. It can be interpreted as the number of independent values generated by a sample of small sizes for estimating a population parameter.

<u>Students' t - Distribution</u>:

Let x_i be a random sample of size n drawn from a normal population with mean μ and variance σ^2.

The statistic 't' is defined as follows:

$$t = \frac{\bar{x} - \mu}{(s/\sqrt{n})} = \frac{\bar{x} - \mu}{s}(\sqrt{n}) \quad \text{———} (1)$$

then, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

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

~~ν = n - 1 denotes the~~

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

where $s^2 = \frac{1}{n_1 + n_2 - 2} \left\{ \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2 \right.$

$\nu = n-1$ denotes the no. of degrees of freedom of t,

to test the hypothesis, whether the sample mean ($\bar{x}$) differs significantly from the population mean (μ) or hypothetical value (μ).

we compute students t given by ①

→ If $|t| > t_{0.05}$, then difference b/w $\bar{x}$ and μ is said to be significance at 5% level of significance and hypothesis is rejected.

→ If $|t| > t_{0.01}$, then difference b/w $\bar{x}$ and μ is said to be significance at 1% level of significance and hypothesis is rejected.

→ If $|t|$ is less than the table value at a certain level of significance the data is said to be conformal or consistent with the hypothesis that μ is the mean of population (i.e hypothesis is accepted)

- 95% confidence limits for μ are
  $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05}$

- 99% confidence limits for μ are
  $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.01}$.

(2) Test of significance for large samples:

*> Test of significance of proportions:

Let us suppose that we take N samples, each drawing n members. Let p be the probability of success and q of failure so that $p + q = 1$.

The frequencies of samples with successes 0, 1, 2, ----- n, are the terms of the binomial expansion of $N(q+p)^n$

Mean of Binomial Distribution is np and S.D is $\sqrt{npq}$

i.) Mean proportion of successes $= \frac{np}{n} = p$

ii.) S.D or S.E proportion of success $= \frac{\sqrt{npq}}{n}$

$$= \sqrt{\frac{npq}{n^2}}$$

$$= \sqrt{\frac{pq}{n}}$$

Let x be the observed no. of successes in a sample size of n and $\mu = np$ be the expected number of success.

Let the associated standard normal variable z be defined by.

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

- If $|x| > 2.58$ we conclude that the difference is highly significant and reject the hypothesis. Since p is the probability of successes and

$\sqrt{\frac{pq}{n}}$ is the S.E proportion of successes,

$p \pm 2.58\sqrt{\frac{pq}{n}}$ are the probable limits.

→ <u>Test of significance of difference between means:</u>

Let μ_1 and μ_2 be the mean of 2 populations. Let (x_1, σ_1), (x_2, σ_2) be the mean and S.D of 2 large samples of sizes n_1 and n_2 respectively.

We wish to test the null hypothesis H_0, that there is no difference between the population means that is $H_0 : M_1 = M_2$

The statistic for this test is given by

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

Also confidence limits for the difference of means of the population are

$$(\bar{x}_1 - \bar{x}_2) \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

If the samples are drawn from the same population then, $\sigma_1 = \sigma_2 = \sigma$,

i.e. $$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

<u>Test of significance for Small Sample:</u>
(Sample size ≤ 30)
(imp.)

Students t - distribution :

1.) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ (only 1 sample is given)

$\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

degree of freedom $\gamma = n-1$

2.) $t = \dfrac{\bar{x} - \bar{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ (when 2 samples are given.)

$s^2 = \dfrac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{j=1}^{n}(y_i - \bar{y})^2\right\}$

Note :-

population :	Mean	μ	sample :	Mean	$\bar{x}$
	S.D	σ		S.D	s
	size :	N		size	n

Problems :

1. 10 individuals are chosen at random from a population and their heights in inches are found to be 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis the mean height of the universe is 66 inches.

Ans ($t_{0.05} = 2.265$ for 9 d.f (degree of freedom))

Given sample size, $n = 10$.

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

$\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{63 + 63 + 66 + 67 + 68 + 69 + 70 + 70 + 71 + 71}{10}$

$= \underline{67.8}$

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

x_i	63	63	66	67	68	69	70	70	71	71
$(x_i - \bar{x})$	-4.8	-4.8	-1.8	-0.8	0.2	1.2	2.2	2.2	3.2	3.2
$(x_i - \bar{x})^2$	23.04	23.04	3.24	0.64	0.04	1.44	4.84	4.84	10.24	10.24

$\sum (x_i - \bar{x})^2 = \cancel{84.2} = \underline{81.6}$.

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

$$= \frac{1}{10-1} (81.6)$$

$s^2 = 9.0666$

$s = \sqrt{9.0666}$

$= \cancel{3.0099} = 3.011$

$\mu = 66$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{67.8 - 66}{3.0111/\sqrt{10}}$$

$= \cancel{0.18}\ 1.8904 < 2.265$

Thus the hypothesis is expected at 5% level of significance. ($t_{0.05} = 2.265$)

2. A certain stimulous administered to each of the 12 patients resulted in the following change in Blood pressures,

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

Can it be concluded that the stimulus will increase the blood pressure.

Ans (1 sample with 12 people)

$\mu = 0$ (normal.)

$t_{0.05} = 11$ d.f $= 2.201$.

($t_{0.05} = 2.201$ for 11 d.f.)

... hypothesis is rejected at
we conclude with 95 % in confidence, that the stimulus in general is increasing with B.P / /

Given sample size $n = 12$.

$\mu = 0$

$$t = \frac{\bar{x} - \mu}{(s/\sqrt{n})} \quad \text{——— (1)}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5+2+8-1+3+0+6-2+1+5+0+4}{12}$$

$$= 31/12 = \underline{2.583}$$

x_i	5	2	8	-1	3	0	6	-2	1	5	0	4
$(x_i - \bar{x})^2$	5.84	0.72	29.34	12.83	0.17	6.67	11.67	21.00	2.50	5.84	6.67	2.00

$$\sum (x_i - \bar{x})^2 = 105.25$$

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

$$= \frac{1}{12-1} \times 105.25$$

$$s^2 = \underline{9.5681}$$

$$s = \sqrt{9.5681} = \underline{3.093}$$

$$(1) \Rightarrow \quad t = \frac{2.583 - 0}{\frac{3.093}{\sqrt{12}}}$$

$$= 2.892 > 2.201$$

Hence the hypothesis is rejected with 5 % significance we conclude with 95 % in confidence, that the stimuli in general is increasing with blood measure.

10/9/23

Note :

one sample

i) 95 % of confidance limit = $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.05}$. (for μ)

ii) 99 % of confidance limit = $\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.01}$ (for μ)

3. A sample of 11 rats from a central population, had an average blood viscosity of 3.92 with a standard deviation of 0.61. On the basis of this sample, establish 95 % of fiducial limit for μ, the mean blood viscosity of the central population.

Ans: Given, sample size $n = 11$.

$s = 0.61$.

$\bar{x} = \cancel{0.}3.92$

(also given $t_{0.005} = 2.228$ for 10 degree of freedom)

one sample,

95 % of confidance limit for $\mu = \bar{x} \pm \frac{s}{\sqrt{n}} (t_{0.05})$

$$\mu = 3.92 \pm \frac{0.61}{\sqrt{11}} (2.228)$$

$$\Rightarrow \mu = 3.92 + \frac{0.61}{\sqrt{11}} (2.228) = 4.3297 \text{ and.}$$

$$\mu = 3.92 - \frac{0.61}{\sqrt{11}} (2.228) = 3.5102$$

Thus 95 % confidence limits for μ are 3.5102 and 4.3297

4. 2 types of batteries are tested for their length of life and the following results were obtained,

Battery A : $n_1 = 10$, $\bar{x}_1 = 500$ hrs $\sigma_1^2 = 100$

Battery B : $n_2 = 10$ $\bar{x}_2 = 560$ hrs $\sigma_2^2 = 121$

Note: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ($\because$ there are 2 samples)

$$\boxed{s^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma^2}{n_1 + n_2 - 2}}$$

Compute students t - distribution and test whether there is a significant difference in 2 means.

Ans: $s^2 = \dfrac{10 \times 100 + 10 \times 121}{10 + 10 - 2}$

$= \dfrac{1000 + 1210}{18}$

$s = 122.77$

$s = \sqrt{122.77}$

$= 11.0805$

$t = \dfrac{500 - 560}{11.0805\sqrt{\frac{1}{10} + \frac{1}{10}}} = \dfrac{-60}{11.0805\sqrt{\frac{2}{10}}}$

$= \dfrac{-60}{11.0805 \times 0.4472} = \dfrac{-60}{4.9553}$

$t = -12.1081$

$|t| = 12.1081$

This value of t is greater than the table value of 't' with degree of freedom 18 ($n_1 + n_2$ at all levels of significance.

The null hypothesis that there is no significant difference is rejected at all significant levels.

5. A group of 10 boys fed on a diet A and another group of 8 boys fed on a different diet B for a period of 6 months, recorded the

/ /

following increasing in weight (lbs)

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

Diet B : 2 3 6 8 10 1 2 8

Test whether Diet A and B differ significantly regarding their effect on increase in weight

table of value : $t_{0.005}$ = 16 d.f => 2.12.

Ans Given : $n_1 = 10$ $n_2 = 8$.

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

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left\{ \Sigma_i (x_i - \bar{x})^2 + \Sigma_i (y_i - y)^2 \right\}$$

$$\bar{x} = \frac{5+6+8+1+12+4+3+9+6+10}{10} = \cancel{14.5}\ \underline{6.4}$$

$$\bar{y} = \frac{2+3+6+8+10+1+2+8}{8} = \underline{5}$$

x	$(x_i - \bar{x})^2$		y	$(y - \bar{y})^2$
5	1.96		2	9
6	0.16		3	4
8	2.56		6	1
1	29.16		8	9
12	31.36		10	25
4	5.76		1	16
3	11.56		2	9
9	6.76		8	9
6	0.16		Σ_i =	<u>82</u>
10	12.96			

$\Sigma_i = \underline{102.4}$

$$s^2 = \frac{1}{10+8-2}\left\{ 102.4 + 82 \right\}$$

$$s^2 = \frac{1}{16}\left\{ 184.4 \right\} = \underline{11.525}$$ $s = \underline{3.3948}$

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{6.4 - 5}{3.3948\sqrt{\frac{1}{10} + \frac{1}{8}}}$$

$$= \frac{1.4}{3.3948 \times 0.4743} =$$

$$= \underline{0.8694} < 2.12$$
$(t_{0.05})$

11/8/23

Note:

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

(Question).
Null ~~h~~ hypothesis is diet A and diet B do not differ significantly.
we get $t = 0.8694 < 2.12$
Thus the hypothesis is accepted at ~~0.~~ 5% level of significance.
∴ Diet A and Diet B do not differ significantly

6. A group of boys and girls were given an intelligence test. The mean score and standard deviation score and numbers in each groups are as follows

	Boys	Girls
Mean	74	70
S.D	8	10
No.	12	10

~~Is~~ Is the ~~g~~ difference b/w the mean of the 2 groups significant at 5% level of significance.
$t_{0.05} = 2.086$ for 20 d.f.

Ans $n_1 = 12$, $n_2 = 10$.

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

chi square

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

$$s^2 = \frac{12(8)^2 + 10(10)^2}{12 + 10 - 2} = \frac{768 + 1000}{20}$$

$= 88.4$

$s^2 = 88.4$

$s = \sqrt{88.4}$

$= 9.4021$

$$t = \frac{74 - 70}{9.4021\sqrt{\frac{1}{12} + \frac{1}{10}}} = \frac{4}{9.4021 \times 0.42281}$$

$= 0.9986 < 2.086$ ($t_{0.05}$)

Thus hypothesis is accepted at 5% level of significance.

7. Chi-Square distribution :

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

$O_i \rightarrow$ Observed outcome

$E_i \rightarrow$ Expected outcome.

7. A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured 3rd class, 90 had secured 2nd class and 20 had secured 1st class. To these figures, support the general examination result which is in the ratio 4:3:2:1 for the respective categories.

$(\chi^2_{0.05} = 7.81$ for 3 df$)$.

Ans: $4:3:2:1$.

Let us take the hypothesis that these figures support to a general result in ratio $4:3:2:1$

The expected frequencies in respective category are

$n = 500$,

$$\frac{4}{10} \times 500, \quad \frac{3}{10} \times 500, \quad \frac{2}{10} \times 500, \quad \frac{1}{10} \times 500.$$

$$= 200, \quad 150, \quad 100, \quad 50$$

(these are expected values.)

Oi	220	170	90	20
Ei	200	150	100	50.

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

$$\chi^2 = \frac{(220-200)^2}{200} + \frac{(170-150)^2}{150} + \frac{(90-100)^2}{100} + \frac{(20-50)^2}{50}$$

$$= \frac{400}{200} + \frac{40\cancel{0}}{15\cancel{0}} + \frac{100}{100} + \frac{90\cancel{0}}{5\cancel{0}}$$

$$= 2 + \frac{8}{3} + 1 + 18$$

$$= \underline{23.667} > 7.8.$$

Thus hypothesis is rejected.

The expected value is not matching with observed value.

8. Fit a poisson distribution for the following data and test goodness of fit.

Given that $\chi^2_{0.05} = 7.815$ for 3 df.

x	0	1	2	3	4
f	122	60	15	2	1

Ans

Given : $n = 5$, $N = 200$ (Σf)

$$P(x) = \frac{m^x e^{-x}}{x!}$$

$$\text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{(122 \times 0) + 60 + 30 + 6 + 4}{200}$$

$$= \frac{1}{2}$$

(when freq not given $\bar{x} = \frac{\Sigma x}{n}$.

mean $= 0.5$. (m).

$$f(x) = \Sigma f \times P(x) = 200 \times \frac{(0.5)^x e^{-0.5}}{x!}$$

We have poisson distribution, $P(x) = \frac{m^x e^{-x}}{x!}$.

thus, $m = \frac{\Sigma fx}{\Sigma f} = \frac{100}{200} = \frac{1}{2} = 0.5$.

then, $f(x) = \Sigma f \times P(x)$

$$= N \times P(x) = 200 \times \frac{(0.5)^x e^{-m}}{x!}$$

$\Rightarrow f(0) = 200 \times \frac{e^{-0.5}}{0!} = \cancel{200}\ 121.30 = 121.$

$f(1) = 200 \times \frac{(0.5)^1 e^{-0.5}}{1!} = 60.65 = 61.$

$f(2) = 200 \times \frac{(0.5)^2 e^{-0.5}}{2!} = 15.163 = 15$

$f(3) = 200 \times \frac{(0.5)^3 e^{-0.5}}{3 \times 2} = 2.527 = 2$

$f(4) = 200 \times \frac{(0.5)^4 e^{-0.5}}{4 \times 3 \times 2} = 0.3159 = 0$

x	0	1	2	3	4
$f = O_i$	122	60	15	2	1
E_i	121	61	15	3	0

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

$$= \frac{(122-121)^2}{121} + \frac{(60-61)^2}{61} + \frac{(15-15)^2}{15} + \frac{(2-3)^2}{3} + 0$$

$$= \qquad \frac{(1-0)^2}{0} = 1$$

$$= \underline{\underline{0.0246 < 7.815}}$$

Thus the hypothesis that the fitness is good can be accepted

9. 5 dice were thrown 96 times and the numbers 1, 2, ~~or~~ 3 appearing on the face of the die follows the frequency distribution as below.

No. of dice showing;

1,2 or 3	5	4	3	2	1	0
frequency	7	19	35	24	8	3

Test the hypothesis that the data follows binomial distribution,

$\chi^2_{0.05} = 11.07$ for 5 d.f.

Ans

$n = 5$ (no. of dices)

$N = 96$

$p = \frac{3}{6}$ (3 no.s out of 6)

$= \frac{1}{2} = \underline{0.5}$

$q = 1 - \underline{0.5} = \underline{0.5}$

x	0	1	2	3	4
f = Oi	122	60	15	2	1
Ei	121	61	15	3	0

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

$$= \frac{(122-121)^2}{121} + \frac{(60-61)^2}{61} + \frac{(15-15)^2}{15} + \frac{(2-3)^2}{3} + \frac{(1-0)^2}{0} = 1$$

$$= 0.0246 < 7.815$$

Thus the hypothesis that the fitness is good can be accepted

9. 5 dice were thrown 96 times and the numbers 1, 2, or 3 appearing on the face of the die follows the frequency distribution as below.

No. of dice showing;

1,2 or 3	5	4	3	2	1	0
frequency	7	19	35	24	8	3

Test the hypothesis that the data follows binomial distribution,

$\chi^2_{0.05} = 11.07$ for 5 d.f.

Ans $n = 5$ (no. of dices)

$N = 96$

$p = \frac{3}{6}$ (3 no.s out of 6)

$= \frac{1}{2} = \underline{0.5}$

$q = 1 - \underline{0.5} = \underline{0.5}$

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

$P(x) = {}^5C_x\, (0.5)^x (0.5)^{5-x}$

$f(x) = N \times P(x)$

$= 96 \times {}^5C_x (0.5)^{5-x+x} = 96 \times {}^5C_x (0.5)^5$

$f(0) = 96 \times {}^5C_0 \times (0.5)^5$

$= \underline{3}$

$f(1) = 96 \times {}^5C_1 \times (0.5)^5$

$= \underline{15}$

$f(2) = 96 \times {}^5C_2 \times (0.5)^5$

$= \underline{30}$

$f(3) = 96 \times {}^5C_3 \times (0.5)^5$

$= \underline{30}$

$f(4) = 96 \times {}^5C_4 \times (0.5)^5$

$= \underline{15}$

$f(5) = 96 \times {}^5C_5 \times (0.5)^5$

$= \underline{3}$

1, 2, 3 shown:	5	4	3	2	1	0
~~f(x)~~ = O_i	7	19	35	24	8	3
$f(x_i) = E_i$	3	15	30	30	15	3

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

$$= \frac{(7-3)^2}{3} + \frac{(19-15)^2}{15} + \frac{(35-30)^2}{30} + \frac{(24-30)^2}{30} + \frac{(8-15)^2}{15} + \frac{(3-3)^2}{3}$$

$$= \frac{16}{3}$$

$$= 11.7 > 11.07$$

Thus hypothesis is rejected at 5% significance.
∴ Data does not follow binomial distribution.

Test for Significance for large Sample:

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

where $x \rightarrow$ no. of desired success
$\mu \rightarrow np$ be the expected no. of success.

* Probable limits / Confidence limit
at 99% level is : $p \pm 2.58\sqrt{\frac{pq}{n}}$

at 95% level is : $p \pm 1.96\sqrt{\frac{pq}{n}}$

* $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The confidence limit for the difference of mean is;

$$\bar{x}_1 - \bar{x}_2 \pm z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- When 2 samples are drawn from same population then:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Test of Significance of Proportions:

10. A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin is unbiased.

Ans: Null hypothesis : Let the coin is unbiased.

$$z = \frac{x - \mu}{\sigma} = x - np$$

observed outcome $x = \underline{540}$.

$\because$ coin is unbiased, $p = \frac{1}{2}$ (1 out of head or tail in 1 coin)

$$= \underline{0.5}$$

$$q = 1 - 0.5 = \underline{0.5}.$$

$$n = \underline{1000}.$$

Expected outcome $= np = 1000 \times 1/2$

$$= \underline{500}.$$

$$z = \frac{540 - 500}{\sqrt{\underset{250}{500} \times \frac{1}{2}}} = \frac{40}{\sqrt{250}} = \underline{2.5298}$$

$\therefore$ hypothesis is accepted at 1% level of significance, and rejected at 5% level of significance.

($2.5298 > 1.96$
< 2.58.)

11. Result extracts revealed that in a certain school over a period of 5 years, 725 students had passed and 615 students failed. Test the hypothesis that success and failure are in equal proportion.

Ans: Null hypothesis : Let success and failure be in equal proportion.

given, $p = \frac{1}{2}$, $q = \frac{1}{2}$

$n = 725 + 615 = 1340$. $(p\ q)$

Observed outcome $= x = 725$. (success)

Expected outcome $= np = 1340 \times \frac{1}{2}$

$= 670$

$\therefore z = \frac{x - np}{\sqrt{pqn}} = \frac{725 - 670}{\sqrt{670 \times \frac{1}{2}}}$

$= \frac{55}{\sqrt{335}}$

$= 3.0049$.

$3.0049 > 2.58$

Thus hypothesis is rejected at 5% level of significance.

Hence the hypothesis that success and failure are in equal proportion is rejected.

12. A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Is it reasonable to think that a die is unbiased one?

Ans: Null hypothesis: Let the die is unbiased.

probability of getting 3 or 4,

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

$q = \frac{2}{3}$ $n = 9000$.

Observed outcome : $x = 3240$

Expected outcome : $np = 9000 \times \frac{1}{3} = 3000$

$z = \frac{x - np}{\sqrt{npq}} = \frac{3240 - 3000}{\sqrt{3000 \times \frac{2}{3}}} = \frac{240}{\sqrt{2000}} = 5.3665$

$5.3665 > 2.58$

Hence the hypothesis is rejected.

$\therefore$ The ~~coin~~ die is biased.

13. In 324 throws of a 6 ~~phase~~ faced die, an odd no. turns up 181 times. Is it reasonable to think the die is an unbiased one?

Ans: odd no.s = 1, 3, 5; = 3.

$\therefore$ Total throws, $n = 324$.

From sum, $p = \frac{3}{6} = \frac{1}{2}$

$q = \frac{1}{2}$

Observed outcome $= x = 181$ (success $\to 1, 3, 5$)

Expected outcome $= np = 324 \times \frac{1}{2}$

$= 162$

$z = \frac{x - np}{\sqrt{npq}} = \frac{181 - 162}{\sqrt{162 \times 1/2}} = \frac{19}{\sqrt{81}} = \frac{19}{9}$

$= 2.111$

$2.11 < 2.58$ ($\because$ its z)

Hence hypothesis is accepted at 5% level of significance.

$\therefore$ Die is unbiased.

$5.3685 > 2.58$

Hence the hypothesis is rejected.

$\therefore$ The ~~coin~~ die is biased.

13. In 324 throws of a 6 ~~phase~~ faced die, an odd no. turns up 181 times. Is it reasonable to think the die is an unbiased one?

Ans Odd no.s = 1, 3, 5; = 3.

$\therefore$ Total throws, $n = 324$.

From ~~ans~~, $p = \frac{3}{6} = \frac{1}{2}$

$q = \frac{1}{2}$

Observed outcome = $x = 181$ (success $\to$ 1,3,5)

Expected outcome = $np = 324 \times \frac{1}{2}$

$= 162$

$$z = \frac{x - np}{\sqrt{npq}} = \frac{181 - 162}{\sqrt{162 \times 1/2}} = \frac{19}{\sqrt{81}} = \frac{19}{9}$$

$= 2.111$

$2.11 < 2.58$ ($\because$ its z)

Hence hypothesis is accepted at 5% level of significance.

$\therefore$ Die is unbiased.

16/8/23

14. A sample of 100 days is taken from weather report records of a certain district and 10 of them are found to be foggy. What are the probable limit of the % of the foggy days in the district?

Ans to find probable limit of 99%,

$$P \pm 2.58\sqrt{\frac{pq}{n}}.$$

Probability of foggy days in a sample of 100 days, is given by

$$p = \frac{10}{100} = \underline{0.1}$$

$$q = 1 - 0.1 = 0.9$$

$$n = 100$$

$$\therefore \quad 0.1 \pm 2.58\sqrt{\frac{0.1 \times 0.9}{100}} =$$

$$= \underline{0.1774} \quad \text{and} \quad \underline{0.0226}$$

$$\Rightarrow \quad \underline{2.26\%} \quad \text{and} \quad \underline{17.74\%}$$

Thus, the % of foggy days lies between $\underline{2.26\%}$ and $\underline{17.74\%}$.

15. To know the mean weight of all 10 years old boys in Delhi, a sample 225 was taken. The mean weight of a sample was found to be 67 with standard deviation of 12. What can be inferred about the mean weight of the population?

Ans Given $\bar{x} = 67$, $\sigma = 12$

$n = 225$

Confidence limits for the mean of the population corresponding to the given sample:

95%	99%
$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$
$= 67 \pm 1.96\left(\frac{12}{\sqrt{225}}\right)$	$= 67 \pm 2.58\left(\frac{12}{\sqrt{225}}\right)$
$= \underline{65.432}$ and $\underline{68.568}$	$= \underline{64.936}$ and $\underline{69.064}$

we can say with 95 % of confidence that the mean weight of the population 65.43 and 68.56 and also, with 99 % of confidence we can say that the mean weight lies b/w 64.93 to 69.064 = 69.1.

Test of Significance of Difference b/w Means :

16. In an elementary school examination, the mean grade of 32 boys was 72 with a standard deviation of 8. while the mean grade of 36 girls was 75 with a standard deviation of 6. Test the hypotheses that the performance of girls is better than boys.

Ans We have, $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_1^2}{n_2}}}$

~~20.~~ Given, $x_1 = 72$, $\sigma_1 = 8$, $n_1 = 32$

$x_2 = 75$, $\sigma_2 = 6$, $n_2 = 36$.

$$\Rightarrow \quad z = \frac{72 - 75}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}} = \frac{-3}{\sqrt{\frac{64}{32} + \frac{36}{36}}} = \frac{-3}{\sqrt{2+1}}$$

$$= \frac{-3}{\sqrt{3}} = -\sqrt{3}$$

$$z = \underline{-1.732}$$

$|z| = 1.732 > 1.645$ × (1 - tail test.)

$1.732 < 2.33$ ✓

~~The difference in~~ the performance of girls and boys is significant at 5 % level but not at 1 %.

17 A sample of 100 ~~gin~~ bulbs produced by a company A, showed a mean life of 1190 hours and standard deviation 90 hours. Also a sample of 75 bulbs produced by a company 'B', showed a mean life of 1230 hrs and a standard deviation of 120 hrs. If there is a difference between the mean lifetimes of the bulbs produced by 2 company. at.

(a) 5% level of significance

(b) 1% level of significience.

Ans Null hypothesis: There is no difference b/w the mean lifetime of bulbs produced by 2 companies.

Given, $\bar{x}_1 = 1190$, $\sigma_1 = 90$, $n_1 = 100$

$\bar{x}_2 = 1230$, $\sigma_2 = 120$, $n_2 = 75$.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}}} = \frac{1190 - 1230}{\sqrt{\frac{90^2}{100^2} + \frac{120^2}{75}}} =$$

$$= \frac{-40}{16.5227}$$

$$z = -2.4209$$

$$|z| = 2.4209$$

$2.4209 > 1.96$ ×

$2.4209 < 2.58$

The null hypothesis is rejected at 5% level, but not at 1% level of significence.

18. A mean of 2 large samples of 1000 and 2000 numbers are 168.75 cm. and 170 cm respectively. Can the sample be regarded as drawn from the same population of S.D 6.25 cm?

Ans Let the sample be regarded as drawn from the same population of 6.25 cm.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

$\bar{x}_1 = 168.75$, $n_1 = 1000$ $\sigma = 6.25.$

$\bar{x}_2 = 170$, $n_2 = 2000$.

$$\Rightarrow z = \frac{168.75 - 170}{6.25\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.1639$$

$|z| = 5.1639 > 2.58$

Thus, so hypothesis is rejected at 1% level of significance; we conclude that the samples can't be regarded

19. A random sample for ~~thousand~~ 1000 workers in company has mean wage of ₹50 per day and standard deviation ₹15. Another sample of 1500 workers from another company has mean wage of ₹45 per day and standard deviation of ₹20. Does the mean rate of wages varies b/w 2 companies? Find 95% confidence limits for the difference of the mean wages of population of 2 companies.

Ans We have $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$

Null hypothesis: The mean rate of wages doesn't vary b/w 2 companies.

Given, $\bar{x}_1 = 50$, $\sigma_1 = 15$, $n_1 = 1000$

$\bar{x}_2 = 45$, $\sigma_2 = 20$, $n_2 = 1500$.

We have $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{50 - 45}{\sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}} = \underline{\underline{7.1307}}$$

$7.1307 > \cancel{2.58}\ 1.96$

$\therefore$ The null hypothesis is rejected at 5% level. Thus mean difference in wages varies.

Confidence limits for the mean of the diff. in wages: at 95% level:

(1 tailed) $z_e = 1.96$ (95%)

x ~~$\bar{x}_1 \pm 1.96$~~

$$(\bar{x}_1 - \bar{x}_2) \pm z_e \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(50 - 45) \pm 1.96 \sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}$$

$\Rightarrow 5 \pm \underline{\underline{1.3743}}$

$= \underline{\underline{3.6257}}$ and $\underline{\underline{6.3743}}$

Thus 95% confidence that the difference between mean wages of population lies between 3.6257 and 6.3743