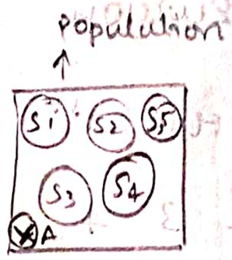
06/12/21

UNIT-IV

SAMPLING TECHNIQUES:-



sampling distributions:-

(X) A → conclude whether A has come from this population.

Random sampling:

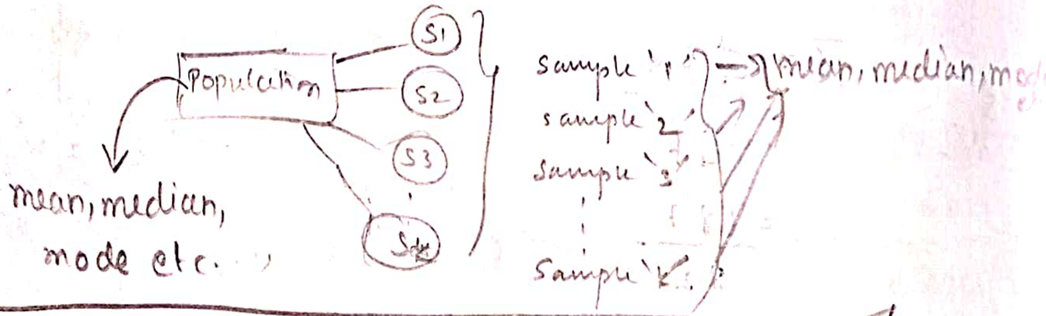
stratified sampling.

Population:

A statistical population is the set of all possible measurements on data corresponding to the entire collection of units for which an inference is to be made.

Sample:

A sample is a part of the statistical population



NOTE:

$\boxed{\text{* Measure of the population is parameter; * " " " samples is statistic.}}$

Conclusion.

- Mean of the Population = Mean of the samplespace

07/12/21

Parameter & statistic:-

Mean, mode, median, etc...., are some characterizations of a statistical distribution. These characteristics are called parameter if they calculated for population and are called statistics if they are calculated for a sample.

sampling distributions.

These statistics vary from sample to sample if repeated random samples of the same size are drawn from a statistical population. The probability distribution of such a statistic is called the sampling distribution.

Eg:

For sampling distribution of mean.

Let us consider the example of a manufacturing process that provides invertors where life times are normally distributed with an arithmetic mean of 2000 hours and a SD of 50 hours. We treat this population as an infinite population.

contd:

No. of samples of size five and compute the mean life time of five invertors in each sample. Each such sample mean is a statistic and these statistics will differ from sample to sample. Thus this statistics is a random variable for which we construct a frequency distribution. The population parameter is 2000 hours and the sample means will be distributed around the parameter value 2000. In some samples it can be < 2000 hrs and in some it may be > 2000 hrs. If we draw a finite no. of samples, the sampling distribution formed by these sample means is referred to as an empirical distribution. If we conceive of drawing all possible samples of a given size, the resulting sampling distribution is a theortical sampling distribution

$\mu = 2000$ (mean of popu.)

$\bar{x} = 1750$ (mean of sample)

$\forall$ is $\mu = \bar{x}$

No !

$\bar{x} \in (2000 - \epsilon , 2000 + \epsilon)$

Standard error: (SE).

SD of the sampling distribution is called standard error. (SE)

$H_0 \rightarrow$ Null hypothesis

$H_1 \rightarrow$ alternative hypothesis.

⇒ H_0 is true and the judge finds that he is innocent (H_0 is accepted)

⇒ H_0 is true and the judge finds that he is guilty (H_0 is rejected)

⇒ H_0 is false and the judge finds that he is innocent (H_0 is accepted)

⇒ H_0 is false and the judge finds that he is guilty (H_0 is rejected).

Procedure:

Step 1 → State the null hypothesis

Step 2 → Decide the alternative hypothesis

Step 3 → choose the significance level

Step 4 → Calculate the test statistic.

$$Z = \frac{(X - E(X))}{SE \text{ of } X.}$$

step 5 → Find the table value of z and decide whether the sample statistic falls within the region of acceptence or rejection

Single → $\geq, \leq$
Double tailed → $=$.
large sample ⇒ ≥ 30.
Small sample ⇒ ≤ 30.

Test for specified mean:

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

Problems

1. A sample of 400 (n) students is found to have a mean height of 171.38 ($\bar{x}$) cms. Can it be reasonably regarded as a sample from a large population with mean height (μ) 171.17 cms and SD 3.30 cms?

single → test specified mean
sample 2nd test → equality of 2 means

soln:

$n = 400$, Population mean $= 171.17$,
Sample mean $= 171.38$, population SD $= 3.30$,

The sample is large

H0: The mean height is 171.17.
H1: The mean height is not 171.17.

$$\boxed{Z = \frac{(X - E(X))}{SE \text{ of } X}} \rightarrow Z = \frac{171.38 - 171.17}{3.30 / \sqrt{400}}$$

$$= 1.27.$$

The table value of z at 5% level is 1.96.

conclusion:

Ho is accepted at 5% level, since the calculated value is less than the table value of z

2. An automatic machine fills in tea in sealed tins with mean weight of tea 1 kg and SD 1 gm. A random sample of 50 tins was examined and it was found that their mean weight was 999.5 gms. Is the machine working properly?

Soln:

n = 50, sample mean = 999.5

Population mean = 1000 gms, population SD = 1 gm

Ho: popu. mean is 1000 gms

H1: " " " not 1000 gms

$$z = \frac{(999.5 - 1000)}{1/\sqrt{50}} = -3.54.$$

Absolute value of z is 3.54.

The table value of z is 1% level is 2.58.

Conclusion:

Ho is rejected at 1% level, ∵ calculated value of z is > the table value

3. A sample of 900 items has mean 3.4 and SD 2.61. Can the sample be regarded as drawn from a population with mean 3.25 at 5% level of significance? (large sample test)

Soln:

sample size = 900, sample mean = 3.4,

sample SD = 2.61, population mean = 3.25.

Ho: The pop. mean is 3.25

H1: " " " " not 3.25

$$Z = \frac{(3.4 - 3.25)}{2.61/\sqrt{900}} = 1.72.$$

1.96

conclusion:

Ho is accepted ∵ the calculated value is less than the table value. ∴ it is likely that the sample belongs to the popu. with mean 3.4.

Test for equality of two means:

Formula → $Z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Problems:

1. Random samples drawn from two places ~~given~~ gave the following data relating to the height of males.

	Place A	Place B
Mean height (inches) $\bar{x}$	68.5 $\bar{x}$	68.58 $\bar{y}$
SD of heights	2.5 σ_1^2 6.1 5.7	3 σ_2^2
sample size	1200 n_1	1500 n_2

Test at 5% level that the mean height is the same for adults in the two places.

Soln:

Ho: Mean height is the same in two places.

H1: " " " different " " "

$$Z = \frac{68.5 - 68.58}{\sqrt{\frac{2.5 \times 2.5}{1200} + \frac{3 \times 3}{1500}}} = -0.76.$$

The absolute value of z is 0.76.

Conclusion:

The Ho is accepted at 5% level, since the calculated value of z is less than the table value of z namely 1.96. Hence the mean height is the same in the two places.

Test for equality of two means

2. The means of two samples of 1000 and 2000 (n_1, n_2) members are respectively 67.5 and 68 (x, y) inches. Can they be regarded as drawn from the same population with SD 2.5 inches? (SD.)

Soln:

Ho: the samples belong to same popu.

H1: " " do not " " " ".

$$Z = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = -5.15.$$

The absolute value of the calculated value of z is > the table value of z (2.58) at 1% level of significance.

∴ we reject the Ho hypothesis.

Test for specified propotion:-

To test the hypothesis that the population propotion P has a specified value P_o.

$$Z = \frac{\overset{\text{sample propotion}}{p} - \overset{\text{Population}}{P_o}}{\sqrt{\frac{P_o Q_o}{n}}}, \quad Q = 1 - P$$

$\frac{2560}{5000}$

Problems:

$Z < 5\%$ level. accept Ho

1. A person threw 10 dice 500 times and obtained 2560 times 4, 5 or 6. Can this be attributed to fluctuations in sampling.

Soln:

sample size $= 10 \times 500 = 5000$

p = propotion of getting 4, 5 or 6 $= 2560/5000$

P_o = popu. propotion for getting 4, 5 or 6 $= \frac{1}{2}$ $\quad \frac{3}{6} = \frac{1}{2}$

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

Ho is $P = \frac{1}{2}$

The alternative hypothesis (H_1) is P not equal to $\frac{1}{2}$.

$$Z = \frac{\frac{2560}{5000} - \frac{1}{2}}{\sqrt{\left(\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{500}\right)}} = 1.697.$$

the table value of z is 1.96 at 5% level.
∴ the calculated value of z is < the table value of z, we accept the Ho at 5% level.

---

2. In a sample of 500 people in kerala 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in the state at 5% level of significance.

Soln:

Ho:- Tea and coffee are equally popular
H1: " " " " not " "

$H_0 \Rightarrow P = 1/2$

$H_1 \Rightarrow P$ is not equal to $1/2$

Sample size $n = 500$.

Proportion of tea drinkers in the population

$p = 280/500$

Proportion of tea drinkers in the population

$P = 1/2$.

$$z = \frac{\frac{280}{500} - \frac{1}{2}}{\sqrt{\left(\frac{1}{2}\frac{1}{2}\frac{1}{500}\right)}} = 2.68.$$

The table value of z at 5% level is 1.96.

Conclusion: The Ho is rejected at 5% level. ∴ the calculated value of z is > ~~than~~ the table value of z.

∴ Coffee and tea are not equally popular in the state

Test for equality of two proportions:-

$H_0 \Rightarrow P_1 = P_2$

$H_1 \Rightarrow P_1$ is not equal to P_2

→ Proportion of P_1

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

→ 2nd

$Q = 1 - P$

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

$\frac{280}{500}$

Problems:

1. In a sample of 600 men from a certain city (n_1) 450 men are found to be smokers. ($450/600 = p_1$) In a sample of 900 from another city 450 are found to be smokers. ($450/900 = P_2$) Do the data indicate that the two cities are significantly different w.r.t prevalence of smoking habit among men.

Soln:

$n_1 = 600$, $n_2 = 900$

Proportion of smokers in the 1 city $= p_1 = 450/600$

" " " " II city $= p_2 = 450/900$

H_0: Proportion of smokers in both cities are equal.

H_1: " " " " " " not equal.

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

$$P = \frac{600 \times \frac{450}{600} + 900 \times \frac{480}{900}}{600 + 900}$$

$P = 0.6$, $Q = 0.4$.

$1 - 0.6$

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{0.75 - 0.5}{\sqrt{0.6 \times 0.4 \left(\frac{1}{600} + \frac{1}{900}\right)}} = 9.62.$$

The table value of z at 1% level is 2.58.
The calculated value of z is > the table value. ∴ the Ho is rejected.

15/12/21 Recall

Procedure:

Step 1 → State the H_0

Step 2 → Decide the H_1

Step 3 → Choose the significance level

Step 4 → Calculate the test static.

$$Z = (X - E(X)) / SE \text{ of } X$$

Step 5 → Find the table value of z and decide whether the sample statistic falls within the region of acceptance or rejection.

Test for a specified mean

t, f, chi squared test

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

$\sqrt{n}$ → size of the sample.

σ → SD of popu.

20/12/21

Test for significance for the difference of SD :-

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

s_i → SD of sample i
σ_i → SD of population i.
n_i → no. of in sample i.

Note :- ① if σ_i's are not known, replace σ_i with s_i

② if the samples are taken from same population SD of the population is not given. But SD of the sample is given, use the formula for combined SD & replace σ_i with σ.

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

s_1 & s_2 combined SD formula to find σ

$$z = \frac{s_1 - s_2}{\sigma\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}}$$

1. Frame the hypothesis
2. Decide whether it single tail / double tail test.

3. Decide the level of significance
4. Calculate test statistic Z
5. Compare the values with calculate value.
6. Arrive at a conclusion.

Problem: Random samples drawn from two countries gave the following data:

	country A	country B
Mean height	67.42 inches (x)	67.25 (y)
SD	2.58	2.5
No. of samples	1000	1200

Is the difference b/w SD significant?

Soln:

Null hypothesis H_0:- The difference b/w SD of the samples is not significant ($s_1 = s_2$)

Alternative hypothesis H_1:- The difference b/w SD of the samples is significant ($s_1 \neq s_2$)

$\sigma_1 = ?$, $\sigma_2 = ?$

$n_1 = 1000$, $n_2 = 1200$

$s_1 = 2.58$, $s_2 = 2.5$

{σ_1 & σ_2 are not given
Replace σ_i with s_i & proceed to calculate z}

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$ large samples test

$$z = \frac{2.58 - 2.5}{\sqrt{\frac{2.58^2}{2000} + \frac{2.5^2}{2400}}} = 1.03$$

$|z| = 1.03 < 1.96$ Accept null hypothesis at 5% level of significance.

$s_1 = s_2$.

$H_0: \bar{x} = \bar{y}$ {difference in mean of the samples is not significant}

$H_1:- \bar{x} \neq \bar{y}$ {two tailed test significant difference}

$\bar{x} = 67.42$

$\bar{y} = 67.25$

σ_1 (replace with s_1) $= 2.58$

σ_2 (replace with s_2) $= 2.5$

$n_1 = 1000$, $n_2 = 1200$.

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

unbiased estimator

$$= \frac{67.42 - 67.25}{\sqrt{\frac{2.58^2}{1000} + \frac{2.5^2}{1200}}} = 1.56$$

$|z| = 1.56 < 1.96$. Accept the null hypothesis at 5% level of significance.

Population $\rightarrow \mu$

samples $\rightarrow \bar{x_1}, \bar{x_2}, \bar{x_3}, \bar{x_4}$ & $\bar{x_5}$.

Unbiased estimator :- Expecting (sample mean) = population mean

i.e; $E(\bar{x}) = \mu$

$$\boxed{\bar{x} = \frac{\bar{x_1} + \bar{x_2} + \bar{x_3} + \bar{x_4} + \bar{x_5}}{5}}$$

$E(s_i) = \sigma_i$

21/12/21

Small Sample Test: $(n < 30)$.

Entire (large) sample theory was based on normal distribution (standard normal distribution $z = \frac{X - \mu}{\sigma}$)

If the sample size is small, the distribution is far from normality & hence normal test cannot be applied.

we will discuss

(i) t - test (students t - distribution)

(ii) F - test (F - distribution)

Pdf of student's t - distribution:

$$\left\{ f(t) = \frac{1}{\sqrt{n}\, \beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}} \qquad -\infty < t < \infty \right.$$

with 'n' degree of freedom.

Application of t - distribution. (t - test).

(i) t - Test for single mean

(ii) t - Test for difference of means

(iii) Paired t - test $-\bar{x} \cdot -\bar{y}$.

(iv) Testing the significance of observed sample correlation.

t - test for single mean:

To test;

(i) a random sample x_i $(i = 1, 2, \ldots, n)$ of size n has come out from a normal population with mean μ

(ii) if sample mean differs significantly from the population mean $(\bar{x} = \mu)$.

Null hypothesis H_0:- sample is drawn from the population with μ. $\{\bar{x} = \mu\}$

test - specific sample mean population

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ with $(n-1)$ degrees of freedom $((n-1) - df)$.

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

Average.

Unbiased estimator:-

s_1^2 s_2^2 s_3^2 s_4^2 s_5^2 s_6^2

$$E\left(\frac{n}{n-1} s^2\right) = \sigma^2.$$

---

1. A random sample of 10 boys had the foll. IQ's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a popu. mean IQ of 100?

Soln:-

H_0:- sample has been drawn from the popu $(\bar{x} = \mu = 100)$

H_1: $\mu \neq 100 \longrightarrow$ (two tailed)

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

$$\bar{x} = \frac{70 + 120 + 110 + 101 + 88 + 83 + 95 + 98 + 107 + 100}{10}$$

$$= \frac{972}{10} = 97.2$$

x	$x-\bar{x}$ $\rightarrow (97.2)$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
		1833.60

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

$$= \frac{1833.60}{9} = 203.73.$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{97.2 - 100}{\sqrt{203.73}/\sqrt{10}}$$

$$= -0.62$$

$$|t| = 0.62 < 2.262$$

Table value of t at 5% level of significance with 9 df.

Accept Ho $\quad t <$ Table value

sample has come out from the population.

22/12/21

Small sample test

- t - distribution
  - Test for specified mean
  - Test for diff. of means
  - Paired t - test
  - Observed sample correlation
- F - distribution
  - significant difference in mean
- χ^2 - distribution
  - Goodness of fit independence of attributes.

$n < 30$

degrees of freedom

↓

no. of independent values in a sample

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ with $(n-1)$ df.

Test for diff. in means:

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ with (n_1+n_2-2) df.

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2 + \sum_i (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$ df

1. Given the gain in weight of samples fed on two diets A & B.

A: 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25 (x_i)

B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22. (y_i)

Test, if the two diets differ significantly as regards their effect on Ice in weight.

Soln:

H_0: Two diets do not have effect on Ice in weight. $(\bar{A} = \bar{B})$

H_1: Two diets have effect on Ice in weight $(\bar{A} \neq \bar{B})$ (two tailed test)

$$t = \frac{\bar{A} - \bar{B}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ → with (n_1+n_2-2) df.

$$s^2 = \frac{\sum_i (A - \bar{A})^2 + \sum_i (B - \bar{B})^2}{n_1 + n_2 - 2}.$$

	n_1												
A:	25	32	30	34	24	14	32	24	30	31	35	25	336
$A-\bar{A}$:	-3	-4	2	6	-4	-14	4	-4	2	3	7	-3	
$(A-\bar{A})^2$	9	16	4	36	16	196	16	16	4	9	49	9	380

$$A = 336, \quad \bar{A} = \frac{336}{12} = 28, \quad (A-\bar{A})^2 = 380$$

	n_2														
B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22
$B-\bar{B}$	14	4	-8	-20	17	1	10	0	2	5	-12	-9	5	-1	-8
$(B-\bar{B})^2$	196	16	64	400	279	1	100	0	4	25	144	81	25	1	64

$$B = 450, \quad \bar{B} = \frac{450}{15} = 30, \quad (B-\bar{B})^2 = 1410.$$

$$\Sigma (A-\bar{A}) = 380, \quad \Sigma (B-\bar{B}) = 1410.$$

$$S^2 = \frac{\Sigma (A-\bar{A})^2 + \Sigma (B-\bar{B})^2}{n_1+n_2-2} = \frac{380+1410}{12+15-2}$$

$$= \frac{1790}{25} = 71.6.$$

$$t = \frac{\bar{A}-\bar{B}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{28-30}{\sqrt{71.6\left(\frac{1}{12}+\frac{1}{15}\right)}} = -0.609.$$

$|t| = 0.609 < 2.06$ with 25 df at 5% level

$\therefore$ Accept H_0.

2. Samples of two types of electric bulbs were tested for length of life and the foll data were obtained.

	Type I (X)	Type II (Y)
no. of samples	8 n_1	7 n_2
mean	1234 hrs $\bar{x}$	1036 hrs $\bar{y}$
S.D	36	40

Is the difference in the mean sufficient to warrant that type I is superior to type II regarding length of life?

Soln:

H_0: Type I & type II are identical $\{\mu_x = \mu_y\}$

H_1: Type I is superior to type II $\{\mu_x > \mu_y\}$.

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } (n_1 + n_2 - 2) \text{ df}$$

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

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

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

$$S = \frac{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2}{n_1 + n_2 - 2}$$

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

$$= \frac{1}{8 + 7 - 2}\left[8(36)^2 + 7(40)^2\right]$$

$$= 1659.08$$

$$t = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.39.$$

Table value for 5% level & 13 df is 1.77
(single tailed test)..

9.39 > 1.77

Reject H_0 => Accept H_1
i.e; type I is superior to type II.

23/12/21

Small sample test:

(i) $(n < 30)$

(ii) t-distribution

(iii) test for specified mean

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

df => no. of independent values in the data

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

(iv) test for difference in means

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } (n_1 + n_2 - 2)\,df$$

$$s^2 = \frac{\Sigma (x - \bar{x})^2 + \Sigma (y - \bar{y})^2}{n_1 + n_2 - 2}$$

Paired t-test:

{To test the diff. in means}

If the samples are dependent }

test ~~specific~~ statistic $t = \frac{\bar{d}}{s/\sqrt{n}}$ with $(n-1)$ df

$$s = \frac{1}{n-1}\left\{\sum (d-\bar{d})^2\right\} = \frac{1}{n-1}\left\{\sum d^2 - \frac{(\sum d)^2}{n}\right\}$$

$d = x - y$.

---

1. Two laboratories carry out independent estimates of a particular chemicals in a medicine produced by a certain firm. A sample is taken from each batch, halved & the separate halves sent to the two laboratories. The foll-data was obtained.

~~Notes~~

No. of samples $= 10$ n

Mean value of the diff. of estimates $= 0.6 = \bar{d}$

sum of the squares of the diff $= 20$ $\sum (d-\bar{d})^2$

Is the diff. significant?

Given t at 5% level for 9 df is 2.262

Soln:

(H_0: difference in mean is not significant

Sample has come out from the population)

H_1: difference in mean significant.

$t = \frac{\bar{d}}{s/\sqrt{n}}$ with $(n-1)$ df.

$s^2 = \frac{20}{9} = 2.22 \Rightarrow s = \sqrt{2.22}$

$$t = \frac{0.6}{\sqrt{\frac{2.22}{10}}} = 1.274.$$

$t = 1.274 < 2.264$. we accept H_0 at 5%.

level of significance.

---

2. A certain stimulus administrated to each of the 12 patients resulted in foll. ↑se of bp.

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

Can it be concluded that the stimulus will in general be accompanied by an ↑se in bp?

Soln:

H_0: there is no significant diff. in bp before and after taking the medicine.

H_1: there is significant diff. in bp before and after taking the medicine.

d: 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6 (31)

d^2 = 25, 4, 64, 1, 9, 0, 4, 1, 25, 0, 16, 36 (185)

$$\bar{d} = \frac{31}{12} = 2.58.$$

$$s^2 = \frac{1}{n-1}\left\{\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right\}$$

$$= \frac{1}{11}\left\{185 - \frac{31^2}{12}\right\}$$

from the table

$$t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2.58}{\sqrt{\frac{1}{11}\left(185 - \frac{31^2}{12}\right)} / \sqrt{12}} = 2.89.$$

Table value of t at 5% level is 1.80

$t = 2.89 > 1.80$. Reject H_0 at 5% level.

t - test for testing significance of an observed correlation.

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \quad \text{with } (n-2) \text{ df.}$$

{Assume population correlation co-efficient as 0}

uncorrelated

Example:

A co-efficient of correlation (r) of 0.2 is derived from a random sample of 625 pairs of observations. Is this value of r significant.

Soln::

H_0:: The value of r is not significant

H_1:: The value of r is significant.

{r → sample correlation co-eff. = ρ → pop. correlation coeff.}

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \quad \text{with } (n-2) \text{ df.}$$

$$= \frac{0.2}{\sqrt{1-0.2^2}} \sqrt{625-2} = 5.09.$$

$t = 5.09 > 2.58$ & Reject the null hypothesis at 1% level of significance.

{If the popu. is large, the t-distribution is appr. a normal distribution}.

Hence, value of 2.58 (which is the value of the normaltable for 1% level is considered}.

27/12/21

F-test {F-distribution}.

Test of equality of Two population variance:

$H_0 :- \sigma_1^2 = \sigma_2^2$

$H_1 :- \sigma_1^2 \neq \sigma_2^2$

test statistic $F = \frac{S_1^2}{S_2^2}$ with (n_1-1, n_2-1) df

where, $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ & $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$

Note:

$F \geq 1$ (F is always greater than or equal to 1)

Eg: 1.

Two random samples gave the foll. results.

$n_1 = 10, \ \underbrace{\Sigma (x_i - \bar{x})^2}_{s_1} = 90$

$n_2 = 12, \ \underbrace{\Sigma (y_i - \bar{y})^2}_{s_2} = 108.$

Test whether the samples came from the popu. with same variance.

Soln.

$H_0: \sigma_1^2 = \sigma_2^2$ (samples have come from same population)

$H_1: \sigma^2 \neq \sigma_2^2$ (samples have come from diff. population).

$F = \dfrac{S_1^2}{S_2^2}$ with $(n_1 - 1, n_2 - 1)$ df

$$S_i^2 = \frac{n_i s_i^2}{n_i - 1}$$

$$s_i^2 = \frac{\sum_i (x_i - \bar{x})^2}{n_i} \qquad \boxed{E\left(\frac{ns^2}{n-1}\right) = \sigma^2}$$

$$s_1^2 = \frac{90}{10} = 9, \quad s_2^2 = \frac{108}{12} = 9$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{10(9)}{9} = 10. \qquad E(\bar{x}) = \mu$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{12(9)}{11} = 9.82.$$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.02.$$

df is $(n_1 - 1, n_2 - 1) = (9, 11)$.

Table value of F at (9, 11) for 5% level of significance is 2.90 (Table of F distribution)

calculate value of F < Table value of F.

$\therefore$ we accept H_0.

---

2. values of a variable in two sample are

Sample I	5	6	8	1	12	4	3	9	6	10
Sample II	2	3	6	8	1	10	2	8	-	-

Test the significance of the difference b/w two sample variances.

Soln:

$H_0: \sigma_1^2 = \sigma_2^2$, $H_1 = \sigma_1^2 \neq \sigma_2^2$

$F = \frac{S_1^2}{S_2^2}$ with $(n_1 - 1, n_2 - 1)$ df.

x	x^2
5	25
6	36
8	64
1	1
12	144
4	16
3	9
9	81
6	36
10	100
64	512

$$s_1^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$= \left(\frac{512}{10}\right) - \left(\frac{64}{10}\right)^2$$

$$= 10.24.$$

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

$$= 11.37.$$

y	y^2
2	4
3	9
6	36
8	64
1	1
10	100
2	4
8	64
40	282

$$s_2^2 = \left(\frac{\sum y^2}{n}\right) - \left(\frac{\sum y}{n}\right)^2$$

$$= \frac{282}{8} - \left(\frac{40}{8}\right)^2 = 10.25$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{8(10.25)}{7} = 11.71.$$

$\because S_2^2 > S_1^2$

$$F = \frac{S_2^2}{S_1^2} = \frac{11.71}{11.37} = 1.02.$$

Table value of F with $(n_2 - 1, n_1 - 1)$ df is 3.29 at 5% level.

calculated value is less than table value

Ho is accepted at 5% level.

28/12/21

UNIT - IV (X)

continuation (which is lefted in probability distribution)

(Probability distribution).

Chi - square distribution: (χ^2)

* Goodness of fit
* Independence of attributes

$\Rightarrow X \sim N(\mu, \sigma^2)$ [Normal distribution]

$\Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0,1)$ [standard distribution].

$\Rightarrow \chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$ is a chi-square distribution with 'n' degrees of freedom.

* In case of goodness of fit.

$$\chi^2 = \sum \left(\frac{(O_i - E_i)^2}{E_i}\right) \text{ with } (n-1) \text{ degree freedom}$$

where O_i → observed frequency

E_i → Expected frequency.

5

Note!

→ O_i (or) E_i must be greater than or equal to 5

→ If it is less than 5 group the data so that it is ≥ 5.

→ when it is ≤ 5 there will be error (or) it will not form.

Eg:1

The demand for a particular product in a company is given below.

Days:	M	T	W	Th	F	Sa
Demand:	1124	1125	1110	1120	1126	1115.

Test whether the no. of products doesn't depend upon the day of the week.

Soln:

Ho: sale of the product doesn't depend upon the day.

H1: sale of the product depends upon the day.

$E_i = ?$ (calculate the expected frequency).

$$E_i = \frac{1124 + 1125 + 1110 + 1120 + 1126 + 1115}{6}$$

$E_i = 1120$.

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1124	1120	4	16	0.014
1125	1120	5	25	0.022
1110	1120	-10	100	0.089
1120	1120	0	0	0
1126	1120	6	36	0.032
1115	1120	-5	25	0.022
				0.179

$\chi^2 = 0.179$ with 5 df $(n-1)$ df.

Table value of chi-square (χ^2) for 5 df at 5% level is 11.07.

calculated value < table value.

Accept Ho at 5% level.

---

contingency table:

a	b
c	d

In case of 2/2;

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

& $N = a+b+c+d$.

degrees of freedom is $(r-1)(c-1)$.

800/1600

eg:

Out of 8000 graduates in a town 800 are females out of 1600 graduates 120 employees are females. Test whether there is any distinction made in appointment on the basis of gender. ($\chi^2_{0.05} = 3.84$ for 1 df)

gender \ status of employ.	employed	not employed	Total
Male	1480 (a)	5720 (b)	7200
Female	120 (c)	680 (d)	8000
Total	1600	6400.	8000

$N = 8000$. $1600 - 120 = 1480$ (a); $7200 - 1480 = 5720$ (b);

$800 - 120 = 680$ (d); $5720 + 680 = 6400$ (d)

Ho: There is no distinction in the employement w.r.t gender

H1: There is distinction " " " " "

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+d)(c+d)}$$ with $(r-1)(c-1)$ df

no. of columns

no. of rows

$$\chi^2 = \frac{8000[1480 \times 680 - 120 \times 5720]^2}{1600 \times 6400 \times 7200 \times 800}$$ (1 df)

$$\chi^2 = 13.89.$$

Table value of χ^2 for 1 df at 5% level is 3.84.

13.89 > 3.84. Reject the Ho hypothesis.

Note:

- For 1 df at 5% level is 3.54.
- For 5 df at 5% level is 11.07.

29/12/21

Revision for CIA-II.

Problems on expections and variance:-

Q1. Find the expection of the no. on a dice when thrown.

Soln:

X → random variable as a number when a dice is thrown.

x_i; 1, 2, 3, 4, 5, 6.

$P(X=1) = P(X=2) = ,,, = P(X=6) = \frac{1}{6}$.

$E(X) = \sum_i x_i p_i$

$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6}$

$+ 6 \times \frac{1}{6}$

$= \frac{1}{6}[1+2+3+4+5+6]$

$= \frac{6(7)}{2(6)} = \frac{7}{2} = 3.5$ //

---

2. Find the expection of when two dice are thrown.

$\boxed{6 \times 6 = 36}$

Solu: $X \to 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$.

x	2	3	4	5	6	7	8	9	10	11	12
P_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

{(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
⋮
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}

$E(X) = \frac{1}{36}\{2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12\}$

$= \frac{252}{36} = 7.$

S = {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}

$n(S) = 36$

$X \to$ event of getting a sum 10

$$P(X=10) = \frac{3}{36}.$$

---

3. X denote the length of the run of either success or failure starting with first trial in a sequence of Binomial trials. Given $P(X=r) = p^r q + q^r p$. Find $E(X)$ & val (X).

$\hookrightarrow E(X^2) - (E(X))^2$.

<u>soln</u>:

$$E(X) = \sum x\, P(X=x) = \sum_{r=1}^{\infty} r\, P(X=r)$$

$$= \sum_{r=1}^{\infty} r\{p^r q + q^r p\} \qquad |p|<1,\ |q|<1$$

$$= pq \sum_{r=1}^{\infty} r p^{r-1} + pq \sum_{r=1}^{\infty} r q^{r-1}.$$

$$= pq\{1+2p+3p^2+\dots\} + pq\{1+2q+3q^2+\dots\}$$

$$= pq(1-p)^{-2} + pq(1-q)^{-2}.$$

$$= pq(1-p)^{-2} + pq(1-q)^{-2}$$

$$= pq \cdot q^{-2} + pq \cdot p^{-2} = \frac{p}{q} + \frac{q}{p}.$$

$$\boxed{E(X) = \frac{p}{q} + \frac{q}{p}}$$

$$E(X^2) = \sum r^2 P(X=r).$$

$$= \sum [r(r-1) + r]\, P(X=r).$$

$$= \sum r(r-1)\, P(X=r) + \sum r\, P(X=r)$$

consider $\sum_r r(r-1)\, P(X=r)$.

$$= \sum_{r=2}^{\infty} r(r-1)\{p^r q + q^r p\}.$$

$$= \sum_{r=2}^{\infty} r(r-1)\, p^r q + \sum_{r=2}^{\infty} r(r-1) q^r p.$$

$$= p^2 q \sum_{r=2}^{\infty} r(r-1) p^{r-2} + q^2 p \sum_{r=2}^{\infty} r(r-1) q^{r-2}$$

$$= 2p^2 q \underbrace{\boxed{\sum_{r=2}^{\infty} \frac{r(r-1)}{2} p^{r-2}}}_{(1-p)^{-3}} + 2q^2 p \underbrace{\boxed{\sum_{r=2}^{\infty} \frac{r(r-1)}{2} q^{r-2}}}_{(1-q)^{-3}}.$$

$$= 2p^2 q (1-p)^{-3} + 2q^2 p (1-q)^{-3}.$$

$$= \frac{2p^2 q}{q^3} + \frac{2q^2 p}{p^3} = \frac{2p^2}{q^2} + \frac{2q^2}{p^2}.$$

Sub in (A);

$$E(x^2) = \frac{2p^2}{q^2} + \frac{2q^2}{p^2} + \frac{p}{q} + \frac{q}{p}.$$

$$\text{var}(x) = E(x^2) - (E(x))^2 = \frac{2p^2}{q^2} + \frac{2q^2}{p^2} + \frac{p}{q} + \frac{q}{p} - \left(\frac{p}{q} + \frac{q}{p}\right)^2$$

$$= \frac{2p^2}{q^2} + \frac{2q^2}{p^2} + \frac{p}{q} + \frac{q}{p} - \frac{p^2}{q^2} - \frac{q^2}{p^2} - 2$$

p^2/q^2 $\quad q^2/p^2$

$$\text{var}(x) = \frac{p^2}{q^2} + \frac{q^2}{p^2} - 2 + \frac{p}{q} + \frac{q}{p}.$$

$$\text{var}(x) = \left(\frac{p}{q} - \frac{q}{p}\right)^2 + \frac{p}{q} + \frac{p}{q}.$$

If $var(x)=1$, then $var(2x+3)$.

soln:

$$var(ax+b) = E(ax+b)^2 - \{E(ax+b)\}^2$$

$$= E[a^2x^2 + 2abx + b^2] - \{aE(x)+b\}^2$$

$$= a^2E(x^2) + \cancel{2ab\,E(x)} + \cancel{b^2} - a^2(E(x))^2$$

$$= a^2\{E(x^2) - (E(x))^2\} - \cancel{2ab\,E(x)} - \cancel{b^2}$$

$$= a^2 var(x).$$

$$var(2x+3) = 2^2\, var(x) = 4(1) = 4 //$$

---

4. Two r.v's x & y have the following joint PDF.

$$f(x,y) = \begin{cases} K(4-x-y) & 0 \le x \le 2;\ 0 \le y \le 2. \\ 0 & \text{otherwise} \end{cases}$$

find (i) const K (ii) Marginal density for of x & y. (iii) conditional density for (iv) $var(x)$, $var(y)$, $cov(x,y)$.

$E(x^2) - (E(x))^2$ $\quad E(y^2) - (E(y))^2$

$E(xy) - E(x)E(y)$

soln:

wkt, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1.$

$$\int_0^2\int_0^2 K(4-x-y)\,dx\,dy = 1.$$

$$\boxed{r = \frac{cov(x,y)}{\sigma_x \sigma_y}}$$

$$\frac{1}{K} = \int_0^2\int_0^2 \{4-x-y\}\,dx\,dy.$$

$$= \int_0^2 \left\{4x - \frac{x^2}{2} - y\cdot x\right\}_0^2 dy.$$

$$= \int_0^2 \{8-2-2y\}\,dy = \left\{8y - 2y - \frac{\cancel{2}y^2}{\cancel{2}}\right\}_0^2$$

$$= 12 - 4 = 8$$

$\therefore k = \frac{1}{8}$.

$$f(x,y) = \begin{cases} \frac{1}{8}(4-x-y) & 0 \le x, y \le 2. \\ 0 & \text{elsewhere}. \end{cases}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$= \int_0^2 \frac{1}{8}(4-x-y)\, dy = \frac{1}{8}\left\{4y - xy - \frac{y^2}{2}\right\}_0^2$$

$$= \frac{1}{8}\{8 - 2x - 2\}.$$

$$f_X(x) = \begin{cases} \frac{6-2x}{8} & 0 \le x \le 2 \\ 0 & \text{otherwise}. \end{cases}$$

$|||^{ly}$ $$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

$$= \begin{cases} \frac{6-2y}{8} & 0 \le y \le 2. \\ 0 & \text{elsewhere}. \end{cases}$$

$$f_{X/Y}(x,y) = \frac{f(x,y)}{f_Y(y)} = \frac{\cancel{\frac{1}{8}}(4-x-y)}{\cancel{\frac{1}{8}}(6-2y)} \quad 0 \le (x,y) \le 2.$$

$$f_{Y/X}(x,y) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{8}(4-x-y)}{\frac{1}{8}(6-2x)} \quad 0 \le (x,y) \le 2$$

$$\operatorname{var}(X) = E(X^2) - (E(X))^2$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^2 x \frac{1}{8}(6-2x)\, dx$$

$$= \frac{1}{8}\int_0^2 \{6x - 2x^2\}\, dx$$

$$= \frac{1}{8}\left\{\frac{6x^2}{2} - \frac{2x^3}{3}\right\}_0^2$$

$$= \frac{1}{8}\left\{12 - \frac{16}{3}\right\} = \frac{1}{24}(20) = \frac{5}{6}.$$

$$E(x^2) = \int_0^2 x^2 \cdot \frac{1}{8}(6-2x)\,dx$$

$$= \frac{1}{8}\int_0^2 \{6x^2 - 2x^3\}\,dx$$

$$= \frac{1}{8}\left\{\frac{\cancel{6}^{2} x^3}{\cancel{3}} - \frac{\cancel{2}x^4}{\cancel{4}_2}\right\}_0^2 = \frac{1}{8}\{16-8\} = 1.$$

$$= \frac{1}{8}\left\{2x^3 - \frac{x^4}{2}\right\}_0^2 = \frac{1}{8}\{16-8\} = 1$$

$$\text{var}(x) = E(x^2) - (E(x))^2.$$

$$= 1 - \left(\frac{5}{6}\right)^2 = 1 - \frac{25}{36} = \frac{11}{36}.$$

$$E(y) = \frac{5}{6} \quad \& \quad \text{var}(y) = \frac{11}{36}.$$

$$E(xy) = \int_0^2\int_0^2 xy\, f(x,y)\,dx\,dy.$$

$$= \int_0^2\int_0^2 xy\left\{\frac{4-x-y}{8}\right\}dx\,dy.$$

$$= \int_0^2\int_0^2 \frac{y}{8}\{4x - x^2 - xy\}\,dx\,dy$$

$$= \int_0^2 \frac{y}{8}\left\{\frac{4x^2}{2} - \frac{x^3}{3} - \frac{x^2 y}{2}\right\}_0^2 dy.$$

$$= \int_0^2 \frac{y}{8}\left\{8 - \frac{8}{3} - 2y\right\}dy$$

$$= \int_0^2 \frac{y}{8}\left\{\frac{16}{3} - 2y\right\}dy = \frac{1}{24}\int_0^2 \{16y - 6y^2\}\,dy$$

$$= \frac{1}{24}\left\{ \frac{\overset{8}{\cancel{16}}y^2}{\cancel{2}} - \frac{\overset{2}{\cancel{6}}y^3}{\cancel{3}} \right\}_0^2$$

$$= \frac{1}{24}\left\{32 - 16\right\} = \frac{16}{24} = \frac{2}{3}.$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$= \frac{2}{3} - \left(\frac{5}{6}\right)\cdot\left(\frac{5}{6}\right) = \frac{2}{3} - \frac{25}{36}.$$

$$= \frac{24}{36} - \frac{25}{36} = \frac{-1}{36}.$$

30/12/21 **Revision on sampling theory.**

Sampling theory:

To test whether a sample has come out from the population.

compare measures of population & sample.

Parameter → measures of population

Statistic → measures of samples.

$\boxed{\text{Statistic} \simeq \text{parameter.}}$

Sampling theory.

- * Parametric test
  - * {Parameters are known}
  - * {Equivalent to population → measures are known}
  - Large sample test — $n \geq 30$
    1. Test for specified mean
    2. Test for difference in mean
  - Small sample test — Otherwise.
    - (a) t-test
      1. Test for specified mean
- * ~~Stat~~ Non-parametric test
  - * {Parameters are unknown}
  - * {Population → measure are not known}

③ Test for specified proportion | ② Test for diff. mean
④ Test for difference in proportion | ③ Test for correl-ation
④ Test for SD/variance

* F-test → test the difference in means
* χ^2-test → " " independence of attributes

$$Z = \frac{t - E(t)}{SE(t)} \simeq N(0,1).$$

$SE(t)$ ↳ standard error ↓ also called. standard deviation of the sample

Steps:

1. Frame the Hypothesis. (Null, Alternative) hypothesis H_0 H_1.

H_0 and H_1 are complement to each other

2. Use the appropriate test static.
3. Compare the table value with calculated value (obtained from step 2)
4. Decide the acceptance (or) rejection of H_0

Level of significance; Confidence interval.

Acceptance range (95%) (99%)

Eg: $Z = \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| \leq 2.58$. $\bar{x} = \mu$.

$\bar{x} \in (\; , \;)$.

It shld lie b/w some values

Test static:

1. $Z = \dfrac{p - P}{\sqrt{\frac{PQ}{n}}}$

$p \to$ proportion of sample
$P \to$ " " " " population
$Q \Rightarrow 1 - P$.

this is the test for specified proportion.

2. Test for specified mean:

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

$\bar{x} \to$ sample mean
$\mu \to$ popu. mean
$\sigma \to$ SD of the popu.
$n \to$ order / no. of samples

3. Test for diff. in means:-

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\underbrace{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}_{\text{SE (standard error)}}}$$

$\bar{x}_i \to$ sample mean of S_i
$\sigma \to$ SD of popu.
$n_i \to$ no. in sample S_i.

~~A. Test for~~

(i) If two popu. are given

SE would became,

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$\sigma_i \to$ SD of popu. P_i.

(ii) Samples SD are given (one popu).

$$\sigma = \frac{(n_1 - 1)(s_1)^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

4. Test for difference in proportions:-

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

$p_i \to$ sample proportion

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

and $Q = 1 - P$.

Chi-square

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad \text{with } (n-1) \text{ degrees of freedom}$$

$$= \frac{N(ad-bc)}{(a+c)(b+d)(a+b)(c+d)} \quad \text{with } (r-1)(c-1)\, df$$

a	b
c	d

If it is like this

F-test: $(F \geq 1)$.

$$F = \frac{s_1^2}{s_2^2}; \qquad s_i^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x}_i)^2$$

Small-sample tests: (Student's t-distribution).

1. Test for specified mean
2. Test for diff. in means
3. Paired t-test (X)
4. Test for correlation co-eff.

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

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } (n_1 + n_2 - 2)\, df$$

(X) $$t = \frac{\bar{d}}{s/\sqrt{n}} \quad \text{with } (n-1)\, df$$

$\bar{d} \to x_i - y_i$

$i = 1, 2, 3 \ldots$

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \quad \text{with } (n-2)\, df$$

1. Random samples of 400 men and 600 women were asked whether they would like to have a fly over near their residance. 200 men & 325 women were in favour of the proporsal. Test the hypothesis that proportion of men & women in favour of the proporsal are same against that they are not. (Use test for diff. in proportion)

Soln.

This is the type of test for diff. in proportion.

$H_0: P_1 = P_2$ $\qquad p_1 = \frac{200}{400}$, $p_2 = \frac{325}{600}$

$H_1: P_1 \neq P_2$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$Q = 1 - P.$

$P = \frac{525}{1000}$, $Q = 1 - \frac{525}{1000} = \frac{475}{1000}$

---

2. The means of 2 single large samples of 1000 and 2000 members are 67.5 and 68 respectively. Can the sampes be regarded as drawn from the same population of SD 2.5? (Use test of diff. in mean).

Soln.

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

$H_1: \bar{x}_1 \neq \bar{x}_2$ (Two tailed test)

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

3. Two researches adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence level. The results are:-

No. of Students

Researcher	Below avg	Avg	above Avg	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

would u say that sampling techniques adopted by the two researchers are significantly different. (use the chi-square)

Soln:

Ho: Sampling techniques adopted by researchers are different.

H1: " " " " " " are ~~same~~ not different.

$$E(86) = \frac{200 \times 126}{300}$$

$$E(33) = \frac{100 \times 93}{300}$$

$$\vdots$$

$$E(2) = \frac{100 \times 12}{300} = 4$$

$$E = \frac{\text{Row total} \times \text{column total}}{N}$$

O_i - observed frequency

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1 86				
2 40				
3 60				
4 33				
5 44				
6 25				
7 $\left\{\begin{matrix}10\\2\end{matrix}\right.$	$\left.\begin{matrix}8\\4\end{matrix}\right\}12$			

$$E(10) = \frac{200 \times \cancel{12}^{4}}{\cancel{300}} = 8$$

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

No. of rows -2

" " coloumn -4

df $(2-1)(4-1)$

$= 1(3) = 3$