DATE 8 Aug
PAGE

★ Statistics

- with replacement : Prob. remain constant : Independent samples.
  without replacement : Prob. will increase : dependent samples

  For eg. If there are 7 person then prob. of choosing ~~one~~ two of them one after the other

$$= \begin{cases} \frac{1}{7} \times \frac{1}{7} & : \text{with replacement} \\ \frac{1}{7} \times \frac{1}{6} & , \text{without replacement.} \end{cases}$$

- Sample : subset of population. Pop$^n$. 'N' → Sample 'n'

- Prob. Sampling methods
  1. Simple Random Sampling (lottery method & random no. method)
  2. Stratified random sampling. (stratum = group)
  3. Systematic Sampling (a/c to some rule eg. all 1's in one group, all 2's in one group.) when identification no. is available (like aadhar no, roll etc.)
  4. Cluster R.S. (base characteristics may / may not be the same)
  5. Probability proportion sampling. (~~prob.~~ prob. of sample will be proportional to the size of the sample)
  6. multi-phase sampling (take a small sample, if it is statistically true, then go for bigger sample & proceed in similar way for every sample).
  7. multi-stage sampling (involves any 2 or more sampling techniques)

- Non-Probability Sampling Techniques
  - Snowball technique. (when objective is very rare)
  - Purposive sampling (biased)
  - volunteer sampling
  - Quota Sampling
  - Regiment/Area Sampling

- large sample Test

z - test → One sample test
z - test → Two sample test.

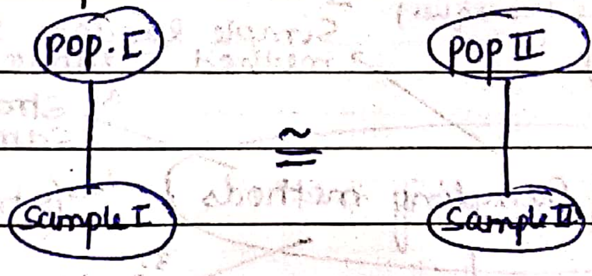
- population : $\mu$ , $\sigma^2$

$\mu$ → pop. mean, $\sigma^2$ → pop. variance

sample : $\bar{x}$ , $s^2$

$\bar{x}$ → sample mean, $s^2$ → sample variance.

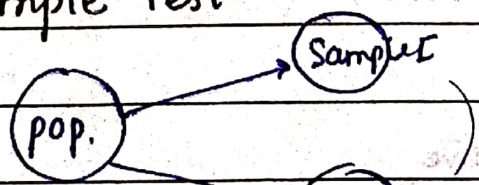
- Two sample test




⇒ If there is no significant difference b/w sample I & sample II then, both pop$^n$. will be same

i.e. Pop(s) can be equal. (conclusion).

One sample Test

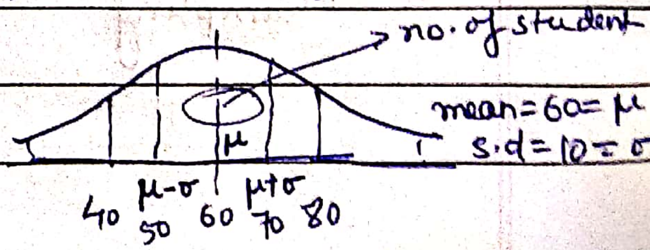



have no significant difference then both samples are drawn from same population. (conclusion).

- large sample ~~size~~ , if sample size $\geq 30$

small sample " " " $< 30$

Range : $(\mu - \sigma, \mu + \sigma)$



★ Hypothesis Testing

1) Null Hypothesis ($H_0$)

assumed hypothesis — $H_0$

$$\left.\begin{aligned} \bar{x} &= \mu_0 \\ E(\bar{x}) &= \mu \end{aligned}\right\} \quad \mu = \mu_0$$

2) Alternate Hypothesis ($H_a$ or $H_1$)

$\mu_0 \neq \mu$ } ⟶ ~~Our~~ Two-sided (Two tail) test

$\mu_0 > \mu_0$, $\mu < \mu_0$ } ⟶ One-sided test (one tail test)

| | Decision | |
|---|---|---|
| Test | $H_0$ Rejected | $H_0$ accepted |
| assumption: $H_0$ is true | (Type I error) wrong | ~~true~~ correct |
| assumption: $H_0$ is false | ~~true~~ correct | wrong (Type II error) |

- $\alpha$ = Type I error = level of significance = Prob. (Reject $H_0$ / $H_0$ is true)

and $\beta$ = Type II error = Prob (Accept $H_0$ / $H_0$ is false)

- level of significance (l.o.s) = 10%, 5%, 1%, 0.1%

- $\mu, \sigma^2$ ⟶ popn — characteristics of popn = parameter

  ↓

  $\bar{x}, s^2$ ⟶ sample — characteristics of sample = statistic

- mean : measure of central tendency
  std. deviation : measure of dispersion.

- Test statistic $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ : decides whether Ho will be accepted or rejected.

Find critical region, take Decision & write conclusion based on decision.

This is called Testing of Hypothesis.

- Power of analysis = $1 - \varepsilon$.

22 Aug

★ Large Sample Test. $[n \geq 30]$

- One Sample Test : let $x_1, x_2, x_3, \ldots x_n$ be random observation of sample 'x' is drawn from a normal population with mean 'μ' and variance is $\sigma^2$ i.e. $N(\mu, \sigma^2)$, respectively.

Null Hypothesis (Ho) : $\mu = \mu_0$ i.e. a random sample drawn from normal population / there is no significant difference b/w sample mean & pop. mean.

Alternate Hypothesis ($H_1$ or $H_a$) : $\mu \neq \mu_0$

$\mu > \mu_0$

Scanned by CamScanner

$\mu < \mu_0$.

The random sample is not drawn from the normal population or there is a significant difference between sample mean & population mean.

L.O.S : $\alpha = 1\%$ or $5\%$ or $10\%$.

Test statistic : let $X \sim N(\mu, \sigma^2)$ (i.e. sample $X$ is drawn from normal popn $N(\mu, \sigma^2)$) then the $z$-statistic or $z$ is defined as

$$z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} \sim N(0,1) \text{ under } H_0 \text{ if}$$

popn variance $\sigma^2$ is known (condn).

and

$$z = \frac{\bar{x} - \mu}{(s/\sqrt{n})} \sim N(0,1) \text{ under } H_0 \text{ if popn.}$$

variance is unknown.

Conclusion : If $z$ calculated value is less than $z$-tabulated value $\underline{z_{cal} < z_{tab}}$ then accept $H_0$ otherwise reject $H_0$ at $\alpha\%$ level of significance (L.O.S.).

- $z_{tab}$ = 1.96 at $\alpha = 5\%$
  1.648 at $\alpha = 10\%$
  2.58 at $\alpha = 1\%$

**Problem 1.)** A sample of 900 members is found to have a mean 3.5 cm. Can it be reasonably regarded as a simple sample from a large population whose mean is 3.38 and std. deviation is 2.4 cm? (Take $\alpha = 5\%$ if not given in the question)

Ans. (It is a two-sided (tail) test since there is no word like increasing or decreasing).

$\mu = 3.38$ and $\bar{x} = 3.5$, $n = 900$

$\sigma = 2.4$

Now, $H_0 : \mu = \mu_0$

ie $\mu = 3.38$

$H_1 : \mu \neq \mu_0$ and $\alpha = 5\%$ (l.o.s.)

$Z_{cal} = 1.5$ and $Z_{tab} = 1.96$

(since $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.5 - 3.38}{2.4/\sqrt{900}} = \frac{0.12 \times 30}{2.4}$

$= 3/2 = 1.5$ )

Since $Z_{cal} < Z_{tab} \Rightarrow$ accept $H_0$ : conclusion.

ie given sample is drawn from normal population.

Problem2. An insurance agent has claimed that the average age of the policy holder through him for all agent which is 30.5. A random sample of 100 policy holder who had insurance through him give the following age list.

| Age last Birthday | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 |
|---|---|---|---|---|---|
| No. of person | 12 | 22 | 20 | 30 | 16 |

Calculate the arithmetic mean & std. deviation of this distribution and test the claim at the 5% los.

Ans.) $\mu = 30.5$ , $n = 100$

$\bar{x} = \frac{\sum_i f_i x_i}{N} = \frac{2880}{100} = 28.8$

| C.I. | $f_i$ | $x_i$ | $f_i x_i$ |
|---|---|---|---|
| 16-20 | 12 | 18 | 216 |
| 21-25 | 22 | 23 | 506 |
| 26-30 | 20 | 28 | 560 |
| 31-35 | 30 | 33 | 990 |
| 36-40 | 16 | 38 | 608 |

$\sum f_i x_i = 2880$

$H_0 : \mu = \mu_0$

i.e. $\mu = 30.5$

$H_1 : \mu \neq \mu_0$ (two-tailed test)

and s.d., $\sigma = \sqrt{\frac{1}{n-1}\left(\sum f_i (x_i - \bar{x})^2\right)}$ for large sample.

$= \sqrt{\frac{1}{(100-1)} \sum f_i (x_i - 28.8)^2}$

$= \sqrt{\frac{1}{99}(4036)} = \sqrt{40.76} = 6.384$

since

| $(x_i - x)$ | $(x_i - \bar{x})^2$ | $f_i (x_i - \bar{x})^2$ |
|---|---|---|
| $-10.8$ | 116.64 | 1399.68 |
| $-5.8$ | 33.64 | 740.08 |
| $-0.8$ | 0.64 | 12.8 |
| 4.2 | 17.64 | 529.2 |
| 9.2 | 84.64 | 1354.24 |

$\sum f_i (x_i - \bar{x})^2 = 4036$

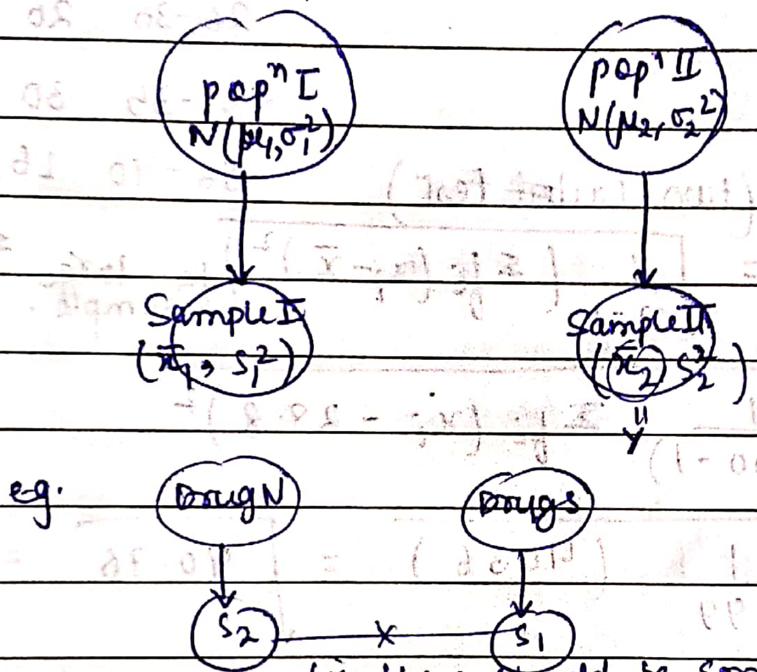
$\therefore z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{28.8 - 30.5}{6.52/10} = -\frac{2.3 \times 10}{6.52}$ 9.94

$\sqrt{n-1}$ 9.94

$= \frac{23}{6.52}$ 22.884 $= 3.52 \rightarrow 3.50$

here, $Z_{cal} > Z_{tab} \Rightarrow$ reject $H_0$.

23 Aug) Two-Sample Test
↓
for comparison
(significant difference between the 2 samples)




(i.e. there should be some significant diff. b/w the effect performed of both medicine i.e. sample, otherwise there is no meaning of two drugs if they have same action hence it is enough i.e. both are same).

Let X be a random sample i.e. $x_1, x_2, \ldots x_n$ drawn from a normal pop with mean $\mu_1$ & variance $\sigma_1^2$ i.e. $N(\mu_1, \sigma_1^2)$ and Y be another random sample drawn from another normal population $N(\mu_2, \sigma_2^2)$.

- Null hypothesis :- $H_0: \mu_1 = \mu_2$ i.e. there is no significant difference b/w means of population(s). [checking whether given populations.

are similar or not]

- $H_1$ or $H_A$ : $\mu_1 \neq \mu_2$

  $\mu_1 > \mu_2$

  $\mu_1 < \mu_2$

i.e. there is a significant difference b/w means of population.

- l.o.s. : $\alpha = 1\%$, $5\%$ or $10\%$

Test statistics :-

case I : $\sigma_1^2$ and $\sigma_2^2$ are known.

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \text{ under } H_0.$$

where $\bar{x}$ and $\bar{y}$ are sample mean & $\sigma_1^2$ & $\sigma_2^2$ are pop. variance and $n_1$ & $n_2$ are sample sizes.

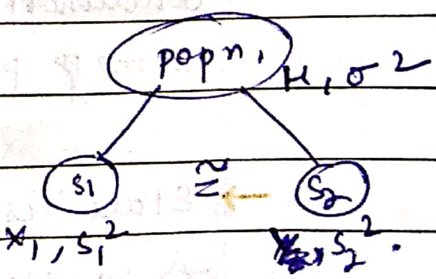
case II : $\sigma_1^2$ & $\sigma_2^2$ are unknown.

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

and $s_1^2$ & $s_2^2$ are sample variance.

case III : Pooled variance is known.

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



Problem 3) The mean of 2 large sample of 1000 and 2000 members are 67.5 inches and 68.0 inches. Can the samples be regarded as drawn from the same population of std. deviation 2.5 inches, Test at 5% L.O.S.

soln.
$$z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{67.5 - 68.0}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{-\cancel{0.5} \times \cancel{10}^2}{\cancel{2.5}_5 \sqrt{3/20}} = -2\sqrt{\frac{20}{3}} = -2 \times 2\sqrt{\frac{5}{3}}$$

$$= -4 \times 1.29 = -5.1639$$

Null Hypothesis : $H_0 : \mu_1 = \mu_2$
both are drawn from same population.

$H_a$ or $H_1$ : $\mu_1 \neq \mu_2$ (Two-sided test)

$$|Z_{cal}| = \left| \frac{\bar{x} - \bar{y}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0,1) \text{ under } H_0.$$

$|Z_{cal}| = 5.16$ and $|Z_{tab}| = 1.96$

since $|Z_{cal}| > |Z_{tab}|$ ∴ reject $H_0$

Conclusion : hence, both samples are not drawn from same population.

→ Statistics depends on dispersion. i.e. Same mean ⇏ samples are drawn from same popn. whereas

same variance (ie s.d.) ie same variation among objects ⇒ (variability among various observation)
samples are drawn from same ~~so~~ population.

**Problem 4)** Two samples are drawn from two large population give the following result.

| | mean | std. | sample size |
|---|---|---|---|
| Sample I : | 250 | 40 | 400 |
| Sample II : | 220 | 55 | 400 |

To examine whether there is any significant difference b/w means 1% l.o.s. ($Z_{tab}$ at 1% is 2.580)

Soln.

$H_0 : \mu_1 = \mu_2$
no significant difference

$H_1 : \mu_1 \neq \mu_2$

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(250 - 220)}{\sqrt{\frac{(40)^2}{400} + \frac{(55)^2}{400}}} = \frac{20 \times 30}{\sqrt{1600 + 3025}}$$

$$= \frac{600}{\sqrt{4625}} = \frac{600}{68} = 8.822$$

$\therefore |Z_{cal}| > |Z_{tab}| \Rightarrow$ reject $H_0$
ie there is significant difference b/w means.

★ Two std. deviation test (when sample means are equal)

let X be a random sample ie $x_1, x_2, \ldots x_n$ drawn from normal population with mean $\mu_1$ and variance $\sigma_1^2$ ie

$N(\mu_1, \sigma_1^2)$ and let $Y$ be another normal pop$^n$ $N(\mu_2, \sigma_2^2)$.

$H_0: \sigma_1 = \sigma_2$ (i.e. there is no significant difference b/w 2 std. deviation).

$H_A$ or $H_1$: $\sigma_1 \neq \sigma_2$, $\sigma_1 > \sigma_2$, $\sigma_1 < \sigma_2$ i.e. there is significant difference b/w 2 std. deviation.

- Test statistics denoted by $Z$, is defined as

$$* \; Z = \frac{|s_1 - s_2|}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \sim N(0,1) \text{ under } H_0.$$

$= 0$ otherwise.

→ If $|Z_{cal}| < |Z_{tab}|$ then accept $H_0$ otherwise reject $H_0$ at $\alpha$ % l.o.s.

- Application.

**Prob.5)** 2 Random samples drawn from 2 countries gave the following data relating to the height of adult males.

| | Country-I | Country-II |
|---|---|---|
| mean | 67.42 | 67.25 |
| S.d. | 2.58 | 2.50 |
| Sample size | 1000 | 1200. |

RANKA
DATE / /
PAGE

Examine: ① is difference b/w mean significant?
② is difference b/w std. deviation significant.
Test at $\alpha = 5\%$ l.o.s.

Sol$^n$ ① $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$z = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{67.42 - 67.25}{\sqrt{\frac{(2.58)^2}{1000} + \frac{(2.5)^2}{1200}}}$$

$$= \frac{0.17 \times 10}{\sqrt{\frac{6.6564}{10} + \frac{6.25}{12}}} = \frac{1.7}{\sqrt{\frac{62.5 + 79.87}{120}}}$$

$$= \frac{1.7}{\sqrt{\frac{142.37}{120}}} = \frac{1.7}{\sqrt{1.186}} = 1.56$$

$$\therefore \underbrace{|z_{cal}|}_{=1.56} < |z_{tab}| = 1.96$$

hence, there is no significant diff.

② $H_0: s_1 = s_2$

$H_1: s_1 \neq s_2$

$$z = \frac{|2.58 - 2.50| \times 10}{\sqrt{(1.186)/2}} = \frac{\sqrt{2} \times 0.08 \times 10}{\sqrt{1.186}}$$

$$= \frac{0.8}{\sqrt{0.593}} = 1.03$$

$$\therefore \underbrace{|z_{cal}|}_{=1.03} < \underbrace{|z_{tab}|}_{=1.96}$$ hence, accept $H_0$.

Scanned by CamScanner

DATE / /
PAGE

Prob 6) Two sets of plot and their variabilities are given by following table.

| | set of 40 plots | set of 60 plots |
|---|---|---|
| mean field | 1258 lb | 1248 lb |
| s.d. | 34 lb | 28 lb |

Examine: is there any significant difference b/w mean of field plot & s.d. of the plots. Test at 5% of l.o.s.

Soln

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1258 - 1248}{\sqrt{\frac{(34)^2}{40} + \frac{(28)^2}{60}}}$$

$$= \frac{10}{\sqrt{\frac{6936 + 3136}{240}}} = \frac{10}{\sqrt{\frac{10072}{240}}}$$

$$= \frac{10}{\sqrt{41.96}} = \frac{10}{6.47} = 1.54$$

since $|z_{cal}| = 1.54 < |z_{tab}| = 1.96$

hence accept $H_0$.

∴ there is no significant diff b/w means of field plot.

and $H_0 : s_1 = s_2$

$H_1 : s_1 \neq s_2$

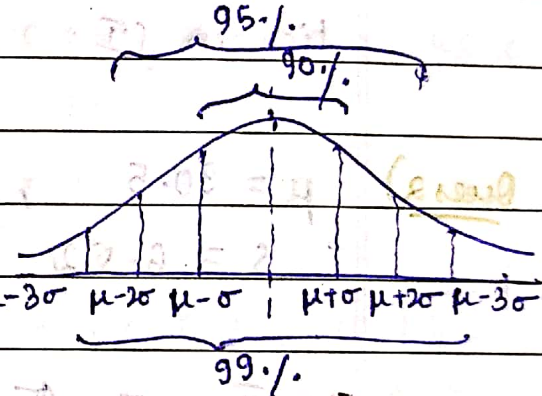
$$z = \frac{34-28}{\sqrt{\frac{41.96}{2}}} = \frac{6\sqrt{2}}{\sqrt{41.96}} = \frac{8.4852}{6.47} = 1.311$$

$\therefore$ $|z_{cal}| < |z_{tab}| \Rightarrow$ accept $H_0$

29 Aug) Confidence Intervals

90%, 95%, 99% : C.I.

mean — one sample / two sample




lower limit / upper limit

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

and two sample:

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

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

from the very beginning Problems.

Ques1) $z = 1.96$ and $n = 900$

$$\bar{x} - z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + z_{\alpha}\left(\sigma/\sqrt{n}\right)$$

$\Rightarrow$ $3.5 - 1.96 \times \frac{2.4}{30} < \mu < 3.5 + 1.96 \times \frac{2.4}{30}$

RANKA DATE / / PAGE

$\Rightarrow 3.5 - 4.704/\sqrt{30} < \mu < 3.5 + (4.704)/\sqrt{30}$

$\Rightarrow 3.5 - 0.1568 < \mu < 3.5 + 0.1568$

$\Rightarrow 3.3432 < \mu < 3.6568$

95% of CI : [3.3432, 3.6568]

Note: $\bar{x}$ value will lie b/w 3.34 and 3.65 i.e. lies in CI.

Ques 2) $\mu = 30.5$, $n = 100$, $z_{5\%} = 1.96$

$s = 6.52$ and $\bar{x} = 28.8$

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

$\Rightarrow 28.8 - \frac{1.96 \times 6.52}{\sqrt{100}} < \mu < 28.8 + \frac{1.96 \times 6.52}{\sqrt{100}}$

$\Rightarrow 28.8 - \frac{1.96 \times 6.52}{10} < \mu < 28.8 + \frac{1.96 \times 6.52}{10}$

$\Rightarrow 28.8 - 1.277 < \mu < 28.8 + 1.277$

$\Rightarrow 27.522 < \mu < 30.077$

i.e. $\bar{x}$ (= 28.8) will lie into [27.52, 30.07] at 5% l.o.s.

i.e. 95% of CI. [27.52, 30.07]

Ques 3) $\bar{x} - \bar{y} = 67.5 - 68.0 = -0.5$

$\sigma = 2.5$ and $\alpha = 5\%$

$n_1 = 1000$ and $n_2 = 2000$

$$(\bar{x}-\bar{y}) - z_{\alpha/2}\,\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}-\bar{y}) + z_{\alpha/2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,\sigma$$

$$\Rightarrow -0.5 - 1.96\times 2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}} < \mu_1-\mu_2 < -0.5+1.96\times 2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}$$

$$\Rightarrow -0.5 - 1.96\times\frac{2.5}{10}\sqrt{\frac{3}{20}} < \mu_1-\mu_2 < -0.5+1.96\times\frac{2.5}{10}\sqrt{\frac{3}{20}}$$

$$\Rightarrow -0.5 - 0.49\sqrt{\frac{3}{20}} < \mu_1-\mu_2 < -0.5+0.49\sqrt{\frac{3}{20}}$$

$$\Rightarrow -0.5 - 0.189 < \mu_1-\mu_2 < -0.5+0.189$$

$$\Rightarrow -0.689 < \mu_1-\mu_2 < -0.311$$

$$\Rightarrow 0.311 < \mu_2-\mu_1 < 0.689$$

95% of CI : [0.311 , 0.689]

**ques 4)** $z = 2.58$ CI for 99% i.e. $\alpha = 1\%$

$\bar{x}-\bar{y} = 250 - 220 = 30$

$$\& \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{4625}{400}} = \frac{68}{20} = 3.4$$

$$\therefore (\bar{x}-\bar{y}) - z\,(3.4) < \mu_1-\mu_2 < (\bar{x}-\bar{y}) + z\times(3.4)$$

$$\Rightarrow 30 - (2.58\times 3.4) < \mu_1-\mu_2 < 30 + (2.58\times 3.4)$$

$$\Rightarrow 30 - 8.772 < \mu_1-\mu_2 < 30 + 8.772$$

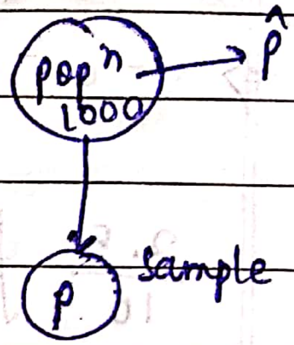
$$21.228 < \mu_1 - \mu_2 < 38.77$$

i.e. 99% of CI : $[21.228, 38.772]$

★ One proportion Test

$$z = \frac{p - \hat{p}}{\sqrt{pq/n}} \sim N(0,1) \text{ under } H_0$$

sample proportion ← $p$; $\hat{p}$ → population proportion; $n$ → sample size




$p = 5\% = 0.05$

$\therefore q = 95\% = 0.95$

Also, $p + q = 1$ (alwe)

prob. of success ($p$), prob. of failure ($q$)

- Let X be a random variable i.e. $x_1, x_2, \dots x_n$ be a random sample drawn from a normal population respectively. Let us consider '$p$' be the proportion of the sample and $\hat{p}$ be the population proportion.

Null Hypothesis ($H_0$) : $p = \hat{p}$ i.e. the sample is drawn from the population / there is no significant difference b/w sample and population proportion.

Alternate Hypotheses ($H_1 / H_A$) : $p \neq \hat{p}$, $p > \hat{p}$, $p < \hat{p}$

i.e. sample is not drawn from normal population / there is a significant difference b/w sample &

normal population proportions.

- level of significance: 5%, 10% or 1%.

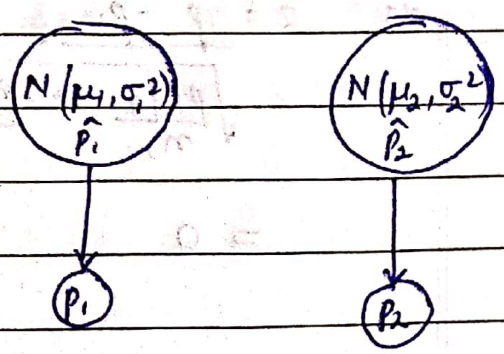
One proportion test statistic is denoted by z & is defined as $z = \dfrac{p-\hat{p}}{\sqrt{pq/n}} \sim N(0,1)$ under $H_0$

$= 0$ otherwise

Conclusion: If $z_{cal} < z_{tab}$ then accept $H_0$ otherwise reject it.

95% of CI: $p - z_{\alpha/2}\sqrt{\dfrac{pq}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\dfrac{pq}{n}}$

30 Aug) Two Sample Proportion Test




let x be random variable i.e. $x_1, x_2, \ldots x_n$ be a random sample drawn from normal population with proportion $\hat{p}_1$.

let us consider y be another random variable i.e. $y_1, y_2, \ldots y_n$ be a random sample drawn from another normal popⁿ, with proportion $\hat{p}_2$.

let us consider $p_1$ and $p_2$ are two sample proportion.

- Null Hypothesis ($H_0$): $\hat{p}_1 = \hat{p}_2$ i.e. the two population proportions are identical or there is no significant diff.

ratio : $N^r$ & $D^r$ are independent .
proportion : $N^r$ & $D^r$ are dependent .

b/w 2 proportions .

- $H_1$ (or $H_a$) : $\hat{p}_1 \neq \hat{p}_2$, $\hat{p}_1 > \hat{p}_2$, $\hat{p}_1 < \hat{p}_2$

i.e the two popn proportions are independent i.e. not identical or there is a significance difference b/w two proportions.

L.o.s : $\alpha$ = 1% , 5% or 10% .

Test statistics : For two sample proportion $p_1$ and $p_2$, the test statistic is denoted by z and is defined as

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1) \text{ under } H_0,$$

$$= 0 \quad \text{otherwise}$$

Critical Region : If $|Z_{cal}| < |Z_{tab}|$ then accept $H_0$, otherwise reject $H_0$.

95% of C.I : $(p_1 - p_2) - Z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} <$

$$\hat{p}_1 - \hat{p}_2 < (p_1 - p_2) + Z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

"two sided test"


RANKA DATE / / PAGE

$$(p_1-p_2) - z_\alpha\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}} < (\hat{p}_1-\hat{p}_2) < (p_1-p_2) + z_\alpha\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

→ one-sided test

**Ques]** In a nursing home, study mention in the chapter opening statistic today, the researchers found that 12 out of 34 small nursing home has a resident vaccin-ation rate less than 80%. while 17 out of 24 large nursing homes had a vaccination rate < 80%. At $\alpha = 5\%$ l.o.s. test the claim that there is no difference in the proportion of small & large nursing home with a resident vaccination rate < 80%. Calculate 90% confidence limit.

- $z_{10\%} = 1.64$

$p_1 = 12/32 = 0.35$ and $p_2 = 17/24 = 0.70$

and $q_1 = 0.65$ & $q_2 = 0.30$

$H_0 : \hat{p}_1 = \hat{p}_2$ & $H_a : \hat{p}_1 \neq \hat{p}_2$

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} = \frac{0.35}{\sqrt{\frac{0.35\times0.65}{34}+\frac{0.3\times0.7}{24}}}$$

$$= \frac{0.35}{\sqrt{\frac{0.2275}{34}+\frac{0.21}{24}}} = \frac{0.35}{\sqrt{0.00669+0.00875}}$$

$$= \frac{0.35}{\sqrt{0.01544}} = \frac{0.35}{0.1245} = 2.811$$

$\therefore \; |z_{cal}| = 2.811 > |z_{tab}| = 1.96$

Hence reject $H_0$.

- CI at 90%

$$(p_1 - p_2) - z_\alpha \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < \hat{p}_1 - \hat{p}_2 < (p_1 - p_2) + z_\alpha \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$\Rightarrow 0.35 - 1.64\sqrt{0.01544} < \hat{p}_1 - \hat{p}_2 < 0.35 + 1.64\sqrt{0.01544}$$

$$\Rightarrow 0.35 - 1.64 \times 0.1245 < \hat{p}_1 - \hat{p}_2 < 0.35 + 1.64 \times 0.1245$$

$$\Rightarrow 0.35 - 0.204 < \hat{p}_1 - \hat{p}_2 < 0.35 + 0.204$$

$$\Rightarrow 0.145 < \hat{p}_1 - \hat{p}_2 < 0.554$$

**Ques)** In a sample of 200 workers, 45% said that they missed work bcz of personal illness. 10 years ago in a sample of 200 workers, 35% said that they missed work bcz of personal illness. At $\alpha = 1\%$ of l.o.s., is there a difference b/w above proportions? Calculate 95% of CI.

**Ans)** $p_1 = 45\% = 0.45 \Rightarrow q_1 = 0.55$

& $p_2 = 35\% = 0.35 \Rightarrow q_2 = 1 - 0.35 = 0.65$

and $n_1 = 200$ & $n_2 = 200$

$H_0 : \hat{p}_1 = \hat{p}_2$

& $H_1 : \hat{p}_1 \neq \hat{p}_2$

$$z = \frac{0.45 - 0.35}{\sqrt{\frac{0.45 \times 0.55}{200} + \frac{0.35 \times 0.65}{200}}} = \frac{0.1 \times 10}{\sqrt{\frac{0.2475 + 0.2275}{2}}}$$

Scanned by CamScanner

$$= \frac{1}{\sqrt{\frac{0.475}{2}}} = \frac{1}{\sqrt{0.2375}} = \frac{1}{0.487} = 2.054$$

$\therefore Z_{cal} = 2.054 < Z_{tab} = 2.58$

CI at 95%

and $(p_1 - p_2) - Z_\alpha \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < \hat{p}_1 - \hat{p}_2 < (p_1 - p_2) + Z_\alpha \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

$\Rightarrow 0.1 - 1.96 \sqrt{\frac{0.2375}{100}} < \hat{p}_1 - \hat{p}_2 < 0.1 + 1.96 \sqrt{\frac{0.2375}{100}}$

$\Rightarrow 0.1 - \frac{1.96 \times 0.487}{10} < \hat{p}_1 - \hat{p}_2 < 0.1 + \frac{1.96 \times 0.487}{10}$

$\Rightarrow 0.1 - 0.0954 < \hat{p}_1 - \hat{p}_2 < 0.1 + 0.095$

$\Rightarrow 0.0045 < \hat{p}_1 - \hat{p}_2 < 0.195$

5 Sept) One - sample test ($n \le 29$)

- One - sample student 't' test

let 'X' be a random variable i.e. $x_1, x_2, \ldots, x_n$ be a random sample drawn from a normal population with mean '$\mu$' and variance '$\sigma^2$' i.e. $N(\mu, \sigma^2)$. let us consider $(\bar{x}, s^2)$ be the sample mean and sample variance.

- $H_0: \mu = \mu_0$ i.e. the sample is drawn from normal population. or the means are identical.
- $H_A: \mu_0 \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$ } i.e. the samples are not drawn from the normal population & means are not identical.

- L.O.S. $\alpha = 1\%$, $5\%$ or $10\%$
- Test Statistics : $$t = \frac{\bar{x} - \mu_0}{\sqrt{S/_{n-1}}} \sim N(0,1) \text{ under } H_0$$

degree of freedom : out of n samples, popn depends on only one sample on which $\bar{x}$ chosen, hence (n-1) remaining samples are free i.e. (n-1)

where n : sample size , $\mu_0$ : $pop^n$ mean
$\bar{x}$ : sample mean , s : std. deviation of sample.

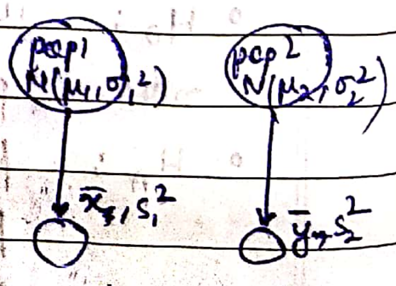
- Decision Rule :- If $t_{cal} < t_{tab}$ with (n-1) d.f. (degree of freedom) at $\alpha\%$ L.O.S. then accept $H_0$, otherwise reject $H_0$.

- Two sample student 't' test

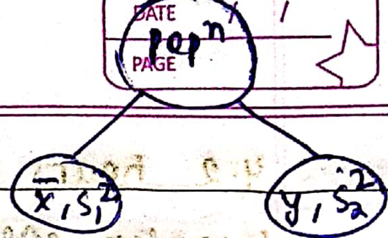
Let 'X' be a random variable i.e. $x_1, x_2, \ldots x_n$ is random sample drawn from normal population. $N(\mu_1, \sigma_1^2)$. Let us consider 'Y' be a random variable i.e. $y_1, y_2, \ldots y_n$ be another random sample drawn from another normal population $N(\mu_2, \sigma_2^2)$.

Let us consider $\bar{x}$ and $\bar{y}$ are sample mean and $s_1^2$ & $s_2^2$ be sample variances drawn from the normal population.

$H_0 : \mu_1 = \mu_2$ i.e. the two samples are drawn from normal population or populations are identical.




$\mu_1 = \mu_2 \Rightarrow$ drawn from * identical $pop^n$




- $H_a$ or $H_1$ : $\mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$

if $\mu_1 = \mu_2$
⇒ drawn from * same population

i.e. the two samples are not drawn from normal populations or populations are not identical.

- l.o.s. $\alpha = 1\%, 5\%$ or $10\%$
- Test statistic : $t = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}}}$ $\sim N(0,1)$ under $H_0$

$$* \quad t = \frac{\bar{x} - \bar{Y}}{s_p \sqrt{\dfrac{1}{n_1 - 1} + \dfrac{1}{n_2 - 1}}}$$

Pooled std. deviation, $s_p = \sqrt{\dfrac{(n_1-1)\sigma_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

* samples are drawn from same population

where $\bar{X}$ & $\bar{Y}$ are sample means.
$n_1$ & $n_2$ " " size
$s_1$ & $s_2$ " " std. deviation.

- Decision rule : If $t_{cal} < t_{tab}$ with $[n_1 + n_2 - 2]$ d.f. at $\alpha\%$ l.o.s. then accept $H_0$ otherwise reject $H_0$.

Ques) General adult population volunteer an average.

4.2 hours per week. A random sample of 20 females college student and 18 male college student indicated these results concerning the amount of time spent in volunteer ~~the~~ increases per week. At the 1% l.o.s., is there sufficient evidence to conclude that a difference exist b/w the mean no. of volunteer hours per week for male & female college student? Compute the 99% confidence limit :

two tailed test

| | male | female |
|---|---|---|
| Sample mean | 2.5 | 3.8 |
| Sample variance | 2.2 | 3.5 |
| Sample size | 18 | 20 |

Soln. $H_0$ : $\mu_1 = \mu_2$

i.e. there is no significant difference b/w the mean no. of volunteer.

$H_A$ : $\mu_1 \neq \mu_2$ — two-tailed test

$\mu_1 < \mu_2$

$\mu_1 > \mu_2$

$$t = (\bar{x} - \bar{y}) \Big/ \sqrt{\frac{s_1^2}{(n_1-1)} + \frac{s_2^2}{(n_2-1)}}$$

$$= \frac{2.5 - 3.8}{\sqrt{\frac{2.2}{17} + \frac{3.5}{19}}} = \frac{-1.3}{\sqrt{0.129 + 0.184}}$$

RANKA DATE / / PAGE

$= \dfrac{1.3}{\sqrt{0.313}} = \dfrac{1.3}{0.559} = 2.32$

and $\alpha = 1\% = 0.01$

since $d.f = n_1 + n_2 - 2$
$= 18 + 20 - 2$
$= 36$

$t_{cal} = 2.32 < t_{tab} = 2.719$

$\Rightarrow$ accept $H_0$ i.e. $\mu_1 = \mu_2$ : no significant difference b/w mean calculated by male & female students.

99% C.I. : $(\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}} < (\mu_1 - \mu_2) <$

$(\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}}$ for two-tailed test

99% CI : $(\bar{x} - \bar{y}) - t_{\alpha}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}} < (\mu_1 - \mu_2) <$

$(\bar{x} - \bar{y}) + t_{\alpha}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}}$ for one-sided test.

CI for 99% :

$\underbrace{1.3}_{(\bar{x}-\bar{y})} - t_{\alpha/2}\underbrace{(0.559)}_{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} < \mu_1 - \mu_2 < 1.3 + t_{\alpha/2}(0.559)$

$\Rightarrow 1.3 - (2.719 \times 0.559) < \mu_1 - \mu_2 < 1.3 + (2.719 \times 0.559)$

$\Rightarrow 1.3 - 1.519 < \mu_1 - \mu_2 < 1.3 + 1.519$

$\Rightarrow -0.219 < \mu_1 - \mu_2 < 2.819$

here $(\bar{x} - \bar{y})$ i.e 1.3 lies b/w this range / interval

i.e. C.I = $[-0.219, 2.819]$

DATE / /

PAGE

Ques2) The no. of points held ~~by~~ a sample of Hockey tournam[ent] highest score for both the Eastern and Western is shown below. At $\alpha = 0.05$, can it be concluded that there is a difference in mean based on these data? compute 95% ~~CI~~.

| Eastern Team, $X$ | Western Team, $Y$ |
|---|---|
| 83, 60, 75, 58, 78, 59, 70, 58, 62, 61, 59 | 77, 59, 72, 58, 37, 57, 66, 55, 61 |

Sol$^n$.

$$\bar{X} = \frac{83+60+75+58+78+59+70+58+62+61+59}{11}$$

$$= \frac{723}{11} = 65.72$$

$$\text{and } \bar{Y} = \frac{77+59+72+58+37+57+66+55+61}{9}$$

$$= \frac{542}{9} = 60.22$$

To calculate $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

$$s_x^2 = \frac{1}{11}\left[(83-65.72)^2 + (60-65.72)^2 + \cdots + (59-65.72)^2\right]$$

$$= \frac{1}{11}\left[(17.28)^2 + (-5.72)^2 + (9.28)^2 + (-7.72)^2 + (12.28)^2 + (-6.72)^2 + (4.28)^2 + (-7.72)^2 + (-3.72)^2 + (-4.72)^2 + (-6.72)^2\right]$$

RANKA
DATE / /
PAGE

$$= \frac{1}{11}\Big[298.59 + 32.72 + 86.12 + 59.50 + 150.80 + 45.16 + 18.32 + 59.60 + 13.84 + 22.27 + 45.16\Big]$$

$$= \frac{877.34}{11} = 79.75$$

$$S_y^2 = \frac{1}{9}\Big[(77-60.22)^2 + (59-60.22)^2 + \dots + (61-60.22)^2\Big]$$

$$= \frac{1}{9}\Big[(16.78)^2 + (-1.22)^2 + (11.78)^2 + (-2.22)^2 + (-23.22)^2 + (-3.22)^2 + (5.78)^2 + (-5.22)^2 + (0.78)^2\Big]$$

$$= \frac{1}{9}\Big[281.56 + 1.48 + 138.76 + 4.92 + 539.16 + 10.36 + 33.40 + 27.24 + 0.60\Big]$$

$$= \frac{1037.48}{9} = 115.27$$

$\therefore$ for l.o.s. $\alpha = 0.05 = 5\%$ — one-tailed test

and d.f. $= n_1 + n_2 - 2 = 11 + 9 - 2 = \cancel{20}\ 18$

$\therefore t_{tab} = 1.734$

$H_0 : \mu_1 = \mu_2$

and $H_1 : \left.\begin{array}{l} \mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{array}\right\}$ *

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} = \frac{65.72 - 60.22}{\sqrt{\frac{(78.75)}{10} + \frac{(115.27)}{8}}}$$

$$= \frac{5.5}{\sqrt{7.875 + 14.408}} \qquad \cancel{= \frac{5.5}{\sqrt{\frac{6201.56}{10} + \frac{13287.17}{8}}}}$$

$$= \frac{5.5}{\sqrt{22.28}} \qquad \cancel{= \frac{5.5}{\sqrt{620.15 + 1660.89}}} \qquad \cancel{= \frac{5.5}{\sqrt{2281.04}}}$$

$$= \frac{5.5}{4.72} \qquad \cancel{= \frac{5.5}{47.76} = 0.115}$$

$$= 1.165$$

$t_{cal} = \overset{1.162}{\cancel{0.115}} < t_{tab} = 1.734 \Rightarrow$ accept $H_0$.

C.I for 95%.

$$(\bar{x} - \bar{y}) - t_\alpha \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}} < \mu_1 - \mu_2 < (\bar{x} - \bar{y}) + t_\alpha \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

$$\Rightarrow 5.5 - 1.734 \times 4.72 < \mu_1 - \mu_2 < 5.5 + 1.734 \times 4.72$$

$$\Rightarrow 5.5 - 8.18 < \mu_1 - \mu_2 < 5.5 + 8.18$$

$$\Rightarrow \underset{\text{lower limit}}{-2.68} < \mu_1 - \mu_2 < \underset{\text{upper limit}}{13.68}$$

9 Sept) Paired Student 't' test

(Note: $\mu_D = 0$ means no significance diff b/w after and before test). 
mean difference

let $x$ be a random variable & $x_1, x_2, \ldots x_n$ be random sample drawn from normal population with mean $\mu$ & std. deviation $\sigma$ i.e. $N(\mu, \sigma^2)$.

RANKA DATE / / PAGE

• Null hypothesis ($H_0$) : $\mu_D = 0$ i.e. there is no significan difference b/w before & after test procedure.

• Alternate hypothesis ($H_1$) : $\mu_D \neq 0$, $\mu_D > 0$, $\mu_D < 0$

i.e. there is a significant difference b/w before & after test procedure.

• level of significance, $\alpha = 1\%$, $5\%$, or $10\%$.

• Test statistic of paired student test is denoted by $t_D$, & is defined as

$$t_D = \frac{\bar{D} - \mu_D}{S/\sqrt{n}} \sim N(0,1) \text{ under } H_0$$

where $D_i = X_1 - X_2$

$$\bar{D} = \frac{\Sigma D_i}{n}$$

$$^*S_D = \sqrt{\frac{n(\Sigma D_i^2) - (\Sigma D_i)^2}{n(n-1)}}$$

$n$ = sample size

and • 95% of CI : $\bar{D} - t_{\alpha/2}\left[\frac{S_D}{\sqrt{n}}\right] < \mu_D < \bar{D} + t_{\alpha/2}\left[\frac{S_D}{\sqrt{n}}\right]$

• Decision-rule : If $t_{cal} < t_{tab}$ with $(n-1)$ d.f. at $\alpha\%$ los then accept $H_0$, otherwise reject $H_0$.

• Conclusion : Summarise the result based on the decision.

Scanned by CamScanner

Ques) A physical education director claim by taking a special vitamin, a weightlifter can increase its strength after 2 weeks of regular training. Compute the test at 5% l.o.s and 95% confidence limit.

| Athletes : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Before ($X_1$): | 210 | 230 | 182 | 205 | 262 | 253 | 219 | 216 |
| After ($X_2$): | 219 | 236 | 179 | 204 | 270 | 250 | 222 | 216 |

Soln. $D_i = (X_1 - X_2)$ : -9 -6 3 1 -8 3 -3 0

since last $D_i$ value is 0 ⇒ no effect ⇒ ~~do~~ reduce value of n by 1 i.e. $n = 8-1 = 7$

$$\therefore \bar{D} = \frac{\sum D_i}{n} = \frac{-9-6+3+1-8+3-3}{7}$$

$$= \frac{-19}{7} = -2.714$$

and $D_i^2$ : 81 36 9 1 64 9 9

$$\therefore \sum D_i^2 = 81+36+9+1+64+9+9$$

$$= 209$$

$$\therefore S_p = \sqrt{\frac{n \times (\sum D_i^2) - (\sum D_i)^2}{n(n-1)}}$$

$$= \sqrt{\frac{7 \times 209 - (\cancel{-2.7}\ -19)^2}{7 \times 6}}$$

$$= \sqrt{\frac{1463 - 361}{42}} = \sqrt{\frac{1102}{42}} = \sqrt{26.23}$$

$$= 5.1215$$

RANKA DATE / / PAGE

two tailed test

$$\therefore t_D = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \quad (\text{since } \mu_D = 0)$$

$$= \frac{-2.714 - 0}{5.1215/\sqrt{7}} = \frac{-2.714}{1.93} = -1.406$$

and $t_{5\%} = 2.\cancel{447}$ with d.f. $= 6\ (= n-1)$

since $t_{cal} < t_{D\,tab} \Rightarrow$ accept $H_0$.

95% of CI :

$t_{tab}$ at 95%

$$-2.714 - \cancel{2.365}\,2.447\,(\cancel{2.109}\,1.93) < \mu_D < -2.714 + \cancel{2.365}\,2.447\,(\cancel{2.109}\,1.93)$$

$$-2.714 - 4.72271 < \mu_D < -2.714 + 4.722$$

$$-7.436 < \mu_D < 2.00871$$

lower limit    upper limit

Ques) Students in a statistics class were asked to report the no. of hours they spent on week night & ~~weekends~~ let ~~evide~~ $\alpha = 5\%$ l.o.s.%, is there any sufficient evidence that there is difference in the mean no. of hours?

~~Sol~~

| | Student | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Days | M to F | 8.0 | 5.5 | 7.5 | 8.0 | 7.0 | 6.0 | 6.5 | 80 |
| | Sat & sun | 4 | 7.0 | 10.5 | 12.0 | 11.0 | 9.0 | 6.0 | 9.5 |

compute the 95% of ~~solution~~ CI.

Soln. Scop- ~~written~~ : 9/6 −1.5 −4 −4 −3 −3 + 10.5 10 +

wrong

$$t_D = \frac{\cancel{8-7-1}\,-1.56 - 0}{n(n-1)}$$

$$= \frac{7 \times 70.75 - (12.5)^2}{7.8}$$

$D_i$ : 4  −0.5  −3.0  −4  −4  −3  0.5  −1.5

$$\therefore \bar{D} = \frac{\Sigma D_i}{N} = \frac{4-0.5-3-4-4-3+0.5-1.5}{8}$$

$$= \frac{-12.5}{8} = -1.56$$

$D_i^2$ ; 16  2.25  9  16  16  9  0.25  2.25

$$\Sigma D_i^2 = 16+2.25+9+16+16+9+0.25+2.25$$

$$= 70.75$$

$$S_D = \sqrt{\frac{n \times (\Sigma D_i^2) - (\Sigma D_i)^2}{n(n-1)}}$$

$$= \sqrt{\frac{8 \times 70.75 - (-12.5)^2}{7 \times 8}}$$

$$= \sqrt{\frac{566 - 156.25}{56}} = \sqrt{\frac{409.75}{56}}$$

$$= 2.704$$

$$\therefore t_D = \frac{-1.56 - \mu_0}{S_D/\sqrt{n}} = \frac{-1.56 - 0}{2.704/\sqrt{8}} = \frac{-1.56}{0.95}$$

$\bar{D}$

$$= 1.63$$

and $t_{D\,tab} = 2.365$ at $(n-1) = 7$ d.f.

i.e. $t_{cal} < t_{D\,tab}$ ⇒ accept $H_0$.

and 95% CI.

$$\underbrace{-1.56}_{\bar{D}} - 2.365 \times (0.956) < \mu_D < -1.56 + 2.365 \times$$

$S/\sqrt{n}$  $t_D$

$\Rightarrow -1.56 - 2.260 < \mu_0 < -1.56 + 2.260$

$\Rightarrow -3.820 < \mu_0 < 0.70$

→ All the previous techniques are parametric (quantitative)

- Chi-Square ($\chi^2$) test (Qualitative Techniques)

1) Test for goodness of fit: (if taken from population)
2) Test for independence (if data already exists)
3) Yate's correction $\chi^2$ - test: (only for 2x2 contigency table) (i.e. not more than that)

→ Test for goodness of fit

observed values: $O_i$ } $\forall i = 1, 2, \ldots, n$

expected values: $E_i$

$H_0$: There is no association b/w observed & expected value i.e. $O_i = E_i$

$H_A$ or $H_1$: There is an association b/w observed & expected value i.e. $O_i \neq E_i$

LOS: $\alpha = 1\%$, $5\%$ or $10\%$.

Test statistic:

Case I: let $x_1, x_2, \ldots x_n$ be a random sample drawn from any population then ~~$\chi^2$~~ chi-square test ~~square~~ statistic is denoted by $\chi^2$ & is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{[(R-1)(C-1)]}$$

↑ d.f.

Case 2 : If the data 2x2 contigency table then

$$\chi^2 = \frac{[|ad - bc|]^2 * n}{[(a+b)(c+d)(a+c)(b+d)]}$$

$\sim \chi^2(1)$ d.f. (here, d.f. = 1)

since $R-1 = 2-1 = 1$ & $C-1 = 2-1 = 1$

**Decision Rule :**

If $\chi^2_{cal} < \chi^2_{tab}$ with $[(R-1)(C-1)]$ d.f. (#rows, #coln)

then accept $H_0$ otherwise reject $H_0$.

**Conclusion :** Summarise the result as per the decision rule.

Ques) A reader read that fireman related deaths for people 1 to 18 years distributed as follows.

74% works accidental

16% works home suicides

10% works other reason. In this district, there were 68 accidental, 27 home suicide, and 5 were other reason during the last year. At $\alpha = 10\%$ L.o.s. claim that %ages are equal.

Soln: $O_1 = 68$ $O_2 = 27$ & $O_3 = 5$

Expected values :

$E_1 = 74$ $E_2 = 16$ $E_3 = 10$

and df : $(R-1)(C-1) = (2-1)(3-1) = 2$

$\therefore$ use $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(2)$ d.f.

| $O_i$ | $E_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $(O_i - E_i)^2/E_i$ |
|---|---|---|---|---|
| 68 | 74 | −6 | 36 | $36/74 = 0.486$ |
| 27 | 16 | 11 | 121 | $121/16 = 7.562$ |
| 5 | 10 | −5 | 25 | $25/10 = 2.5$ |

$$\chi^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \underline{0.486 + 7.562 + 2.5}$$

$$= 10.548$$

Since d.f. = 2 $\therefore$ $\chi^2(2)$ for 10% l.o.s. = 4.604 (using table)

$\because \chi^2_{cal} > \chi^2_{tab} \Rightarrow$ reject $H_0$.
$(10.548) \quad (4.604)$

Ques) A research to determine whether there is a relationship b/w gender of an individual and the amount of alcohol consumed. A sample of 68 people is selected & a following data is obtained. At $\alpha = 5\%$ l.o.s., can the researcher conclude that the alcohol consumption is related to gender.

Sol

| Gender | Alcohol low | moderate | high | Total |
|---|---|---|---|---|
| Male | 10 $O_{11}$ | 9 $O_{12}$ | 8 $O_{13}$ | 27 |
| Female | 13 $O_{21}$ | 16 $O_{22}$ | 12 $O_{23}$ | 41 |
| Total | 23 | 24 | 20 | 68 = N |

$$d.f. = (R-1)(C-1) = (2-1)(3-1) = 2.$$

$$E_{11} = \frac{R_{11} C_{11}}{N} = \frac{27 * 23}{68} = 9.13$$

$$E_{12} = \frac{R_{11} R_{12}}{N} = \frac{27 * 25}{68} = 9.92$$

| $\therefore O_{ij}$ | $E_{ij}$ | $(O_{ij} - E_{ij})$ | $(O_{ij} - E_{ij})^2$ | $\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ |
|---|---|---|---|---|
| 10 | $\frac{27 \times 23}{68} = 9.13$ | 0.87 | 0.7569 | 0.082 |
| 9 | $\frac{27 \times 25}{68} = 9.92$ | −0.92 | 0.8464 | 0.085 |
| 8 | $\frac{27 \times 20}{68} = 7.94$ | 0.06 | 0.0036 | 0.0004 |
| 13 | $\frac{41 \times 23}{68} = 13.86$ | 0.86 | 0.7396 | 0.0533 |
| 16 | $\frac{41 \times 25}{68} = 15.07$ | 0.93 | 0.8649 | 0.0573 |
| 12 | $\frac{41 \times 20}{68} = 12.05$ | −0.05 | 0.0025 | 0.0002 |

$$\therefore \chi^2 = \sum_{i=1}^{6} \left( \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right)$$

$$= 0.2782$$

Since $d.f. = 2$ and $\alpha = 5\% = 0.05$

$\therefore \chi^2_{tab} = 5.99$

$\therefore \underset{(0.27)}{\chi^2_{cal}} < 5.99$ i.e. $\chi^2_{cal} < \chi^2_{tab}$ ∴ accept $H_0$.

Ques) Data on smoking habits and presence of lung cancer is given below.

Smoking

| Cancer | Yes | No | Total | |
|---|---|---|---|---|
| Yes | a 20 | b 10 | 30 | $R_{11} = a+b$ |
| No | c 30 | d 140 | 170 | $R_{12} = c+d$ |
| Total | $C_{11}$ 50 | $C_{12}$ 150 | $200 = N (= R_1 + R_2 = C_1 + C_2)$ | |

Since d.f. = 1, use formula $\chi^2 = \dfrac{[|ad-bc|] * N}{[(a+b)(c+d)(a+c)(b+d)]} \sim \chi^2(1)$ d.f.

$$\chi^2 = \frac{(|ad-bc|) * N}{R_1 * R_2 * C_1 * C_2}$$

$$= \frac{(20 \times 140 - 10 \times 30) \times 200}{38,250,000}$$

$$= \frac{(2800 - 300) \times \cancel{200}}{382500\cancel{00}}$$

$$= \frac{25\cancel{00} \times 2}{3825\cancel{00}} = \frac{50}{3825}$$

$$= 0.013$$

Since $\chi^2_{tab}$ at $\alpha = 5\%$ (0.05 default value) & d.f. = 1

$\chi^2_{tab} = 3.84$

$\chi^2_{cal} < \chi^2_{tab} \Rightarrow$ accept $H_0$