Statistics for CS

Assignment 02

Applications of different formulas used in Hypothesis Testing

- Aditi Chaurasia
(17BCS001)
5th sem

★ Hypothesis Testing (z-test)

Ques 1 : Application of one sample test ; Check whether a random sample is drawn from a normal population or not.

use : $z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$, when population variance is known

and $z = \frac{\bar{x} - \mu}{(s/\sqrt{n})}$, when population variance is unknown.

→ A principal at a school claims that the students in his college are above average intelligence. A random sample of 30 students's IQ scores have a mean score of 112.5. Can it be reasonably regarded as a random sample from a large population whose mean population IQ is 100 with std. deviation of 15. Test at $\alpha = 10\%$ LOS.

Soln. $\mu = 100$, $\bar{x} = 112.5$, $n = 30$, $\sigma = 15$

Now, $H_0 : \mu = \mu_0$ i.e. random sample is drawn from the normal population.

H_1 (or H_A) : $\mu \neq \mu_0$ } two-tail test
$\mu > \mu_0$
$\mu < \mu_0$

Since $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ (since σ^2 is known)

$= \frac{112.5 - 100}{15/\sqrt{30}} = \frac{12.5 \times \sqrt{30}}{15} = 4.56$

i.e. $z_{cal} = 4.56$

We also know that z_{tab} (at $\alpha = 5\%$) $= 1.645$

Therefore, $z_{cal} > z_{tab}$ i.e. $4.56 > 1.645$

hence, reject H_0

This implies that given random sample is not drawn from normal population i.e. $N(\underset{\mu}{100}, \underset{\sigma}{15})$

Scanned by CamScanner

→ A manager has claimed that the average age of employees in his company is 30.0. A random sample of 100 employees holding ~~different insurance~~ positions give the following age list.

Age group	16-20	21-25	26-30	31-35	36-40
No. of employees.	12	22	20	30	16

Calculate the arithmetic mean & std. deviation of this distribution & test the claim at the 10% l.o.s.

Soln. $\mu = 30.0$, $n = 100$, $\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{N} = \frac{2880}{100} = 28.8$

C.I.	f_i	x_i	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
16-20	12	18	216	-10.8	116.64	1399.68
21-25	22	23	506	-5.8	33.64	740.08
26-30	20	28	560	-0.8	0.64	12.8
31-35	30	33	990	4.2	17.64	529.2
36-40	16	38	608	9.2	84.64	1354.24
			$\sum f_i x_i = 2880$			$\sum f_i(x_i-\bar{x})^2 = 4036$

H_0: $\mu = \mu_0$ i.e. sample is drawn from normal population $N(30, \sigma)$.

H_1; $\left.\begin{matrix}\mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0\end{matrix}\right\}$ two tail test.

In this case, σ is unknown therefore first calculate 's'.

sample std. deviation, $s = \sqrt{\frac{1}{n-1}\left(\sum f_i (x_i - 28.8)^2\right)}$

$$= \sqrt{\frac{1}{(100-1)} \cdot (4036)} = \sqrt{40.76}$$

$$= 6.384$$

$$\therefore z = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{28.8 - 30.0}{6.384/\sqrt{99}} = \frac{-1.2}{0.641} = -1.872$$

$$\therefore z_{cal} = |-1.872| = 1.872.$$

Hence, $z_{tab} < z_{cal}$ since $1.645 < 1.872$

$\therefore$ reject H_0 i.e. random sample is not drawn from normal population.

Ques 2. Application of two sample test : Check whether there is any significant difference between two populations. We use z-test for two samples to check whether there is significance difference b/w means of populations. If it is so, then both samples are drawn from same population i.e. no significant difference, otherwise both are drawn from two diff. population.

use : • $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ under H_0 when σ_1^2 & σ_2^2 are known

• $z = \dfrac{\bar{x} - \bar{y}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim N(0,1)$ under H_0 when σ_1^2 & σ_2^2 are unknown

• $z = \dfrac{\bar{x} - \mu}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ when pooled variance is known.

→ The amount of a certain trace element in blood is known to vary with a std. deviation of 14.1 ppm for male blood donors & 9.5 ppm for female donors. Random samples of 75 male and 50 female donors yield concentration means of 28 and 33 ppm, respectively. Examine whether there is any significant difference b/w means at 1% LOS.

Soln. $H_0 : \mu_1 = \mu_2$ i.e. no significant difference b/o means.

$H_1 : \mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$ } i.e. there is significance difference between 2 means.

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{28 - 33}{\sqrt{\frac{(14.1)^2}{75} + \frac{(9.5)^2}{50}}} = \frac{-5}{\sqrt{2.65 + 1.81}} = \frac{-5}{\sqrt{4.46}} = -2.37$$

Scanned by CamScanner

$\therefore Z_{cal} = 2.37$ (take +ve value)

and from z-table, we have Z_{tab} at $\alpha = 1\% = 2.580$

$\therefore Z_{cal} < Z_{tab} \Rightarrow$ accept H_0.

Thus, there is no significant difference between the 2 means.

→ Two samples are drawn from two large population give the following result.

	Sample I (X)	Sample II (Y)
mean	1.3	1.6
Std. deviation	0.5	0.3
sample size	22	24

To examine whether there is any significant difference b/w means at 5% l.o.s.

Soln. $H_0 : \mu_I = \mu_{II}$ i.e. no significant difference b/w the means.

$H_a : \left.\begin{array}{l}\mu_I \neq \mu_{II} \\ \mu_I > \mu_{II} \\ \mu_I < \mu_{II}\end{array}\right\}$ i.e. there is a significant difference between the means.

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} = \frac{(1.3 - 1.6)}{\sqrt{\frac{(0.5)^2}{22} + \frac{(0.3)^2}{24}}} = \frac{-0.3}{\sqrt{\frac{0.025}{22} + \frac{0.09}{24}}}$$

$= -2.44$

$\therefore Z_{cal} = 2.44$ (take +ve value)

and from z table, we have Z_{tab} at $\alpha = 5\% = 1.96$

$\therefore Z_{cal} > Z_{tab} \Rightarrow$ reject H_0

Thus, there is a significant difference b/w the means hence they are drawn from different populations.

→ The mean of 2 large sample of 100 and 400 members are 58 inches and 58.5 inches, can the samples be regarded as drawn from the sample population of std. deviation 2.5 inches. Take l.o.s as 1%.

Soln. $H_0 : \mu_1 = \mu_2$ i.e. both are drawn from same population.

$H_1 : \left.\begin{array}{l}\mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2\end{array}\right\}$ i.e. both are drawn from different populations.

$$z = \frac{\bar{x}-\bar{y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{58-58.5}{2.5\sqrt{\frac{1}{100}+\frac{1}{400}}} = \frac{-0.5\times 20\times 10}{2.5\sqrt{500}} = \frac{-100}{2.5\sqrt{5}} = \frac{-64}{\sqrt{5}}$$

$= -1.78$

$\therefore z_{cal} = 1.78$ and $z_{tab} = 2.580$ at $\alpha = 1\%$.

$\therefore z_{cal} < z_{tab} \Rightarrow$ accept H_0 i.e. both are drawn from same population (i.e. rural population).

- Confidence Interval.

→ The average weight of 100 randomly selected adult males is 180 lbs. Assume a population std. deviation of 20 lbs. compute a 95% confidence interval.

Soln. By formula

$$\bar{x} - z.\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z.\frac{\sigma}{\sqrt{n}}$$ (use $z_{tab} = 1.96$ at $\alpha = 5\%$ / 0.05)

$\Rightarrow 180 - \frac{1.96\times 20}{\sqrt{100}} < \mu < 180 + \frac{1.96\times 20}{\sqrt{100}}$

$\Rightarrow 180 - 1.96\times 2 < \mu < 180 + 1.96\times 2$

$\Rightarrow 180 - 3.92 < \mu < 180 + 3.92$

$\Rightarrow 176.08 < \mu < 183.92$

$\therefore$ C.I. at 95% = [176.08, 183.92]

→ 100 randomly obese people are assigned group 1 and put on the low fat diet. Another 100 randomly assigned people are put into group 2. & put on a diet of approximately the same amount of food, but not as low in fat. After 4 months, the mean weight loss was 9.31 lps for group 1 ($s_1 = 4.67$) & 7.40 lps ($s = 4.04$) for group 2. compute confidence interval at l.o.s 5%. (or confidence level 95%.)

Soln. $\bar{x} = 9.31$, $s_1 = 4.67$ $\therefore (\bar{x}-\bar{y}) = 9.31 - 7.40 = 1.91$

$\bar{y} = 7.40$, $s_2 = 4.04$

and $$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{(4.67)^2}{100}+\frac{(4.04)^2}{100}} = \sqrt{\frac{21.8089+16.3216}{100}}$$

Scanned by CamScanner

$= \sqrt{\frac{38.1305}{100}} = \frac{6.174}{10} = 0.617$

$\therefore$ C.I. at 95%, $Z_{tab} = 1.96$

$\therefore (\bar{x}-\bar{y}) - Z_\alpha\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}-\bar{y}) + Z_\alpha\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

$\Rightarrow 1.91 - 1.96 \times 0.617 < (\mu_1 - \mu_2) < 1.91 - 1.96 \times 0.617$

$\Rightarrow 1.91 - 1.209 < (\mu_1 - \mu_2) < 1.91 + 1.209$

$\Rightarrow 0.701 < (\mu_1 - \mu_2) < 3.119$

C.I. : $[0.701, 3.119]$ at 95% confidence level.

→ Given mean of two large sample of 100 and 400 members as 65.5 and 66 kgs, compute 95% C.I. if samples are drawn from same population of std. deviation 2.0 kg.

Soln, $Z_{tab} = 1.96$ at $\alpha = 5\%$.

and $\bar{x} - \bar{y} = 65.5 - 66.0 = -0.5$, $\sigma = 2.0$

& $n_1 = 100$, $n_2 = 400$

$\therefore \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 2 \times \sqrt{\frac{1}{100}+\frac{1}{400}} = 2 \times \sqrt{\frac{\cancel{500}}{400 \times \cancel{100}}}$

$= \frac{\cancel{2}}{\cancel{20}} \times \sqrt{5} = \frac{\sqrt{5}}{10} = \frac{2.23}{10} = 0.223$

$\therefore$ C.I. at 95% :

$(\bar{x}-\bar{y}) - Z_\alpha \cdot \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}-\bar{y}) + Z_\alpha \cdot \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$\Rightarrow -0.5 - 1.96 \times 0.223 < \mu_1 - \mu_2 < -0.5 + 1.96 \times 0.223$

$\Rightarrow -0.5 - 0.437 < \mu_1 - \mu_2 < -0.5 + 0.437$

$\Rightarrow -0.937 < \mu_1 - \mu_2 < -0.063$

$\therefore$ CI at 95% $= [-0.937, -0.063]$

Two Standard Deviation Test

Check whether there is significant difference between 2 std. deviations.

Use: $z = \dfrac{|s_1 - s_2|}{\sqrt{\dfrac{s_1^2}{2n_1} + \dfrac{s_2^2}{2n_2}}} \sim N(0,1)$ under H_0

$= 0$, otherwise.

Ques 3. Two random samples drawn from 2 fields gave the following data relating to the yield of crops.

	Field I	Field II
mean	56	54.5
Std. dev.	1.414	1.575
Sample size	300	300

Examine is there any significant difference between std. deviations.

Sol^n. H_0 : $\sigma_1 = \sigma_2$ i.e. no significant difference b/w 2 std. deviations

H_a : $\sigma_1 \neq \sigma_2$, $\sigma_1 > \sigma_2$, $\sigma_1 < \sigma_2$ } i.e. there is a significant difference b/w 2 std. deviations.

$$z = \frac{|s_1 - s_2|}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} = \frac{|56 - 54.5|}{\sqrt{\frac{2}{2\times300} + \frac{2.479}{2\times300}}} = \frac{1.5\times10}{\sqrt{\frac{1}{3\times2} + \frac{2.479}{3\times2}}}$$

$$= \frac{15\sqrt{6}}{\sqrt{3.479}} = 15\times1.32 = 19.8$$

$\therefore$ $Z_{cal} > Z_{tab}$ (since $Z_{tab} = 1.96$ at $\alpha = 5\%$ l.o.s.)

$\Rightarrow$ reject H_0 ie there is significant difference b/w the 2 std. deviation.

Test for Proportion

Ques 4. One proportion test : Check whether there is any significant difference b/w sample and population proportion.

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

here, $p + q = 1$

for C.I. : $p - z_{\alpha/2}\sqrt{\frac{pq}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{pq}{n}}$

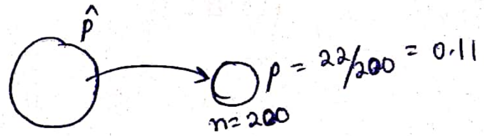
→ The mayor of a town saw an article that claimed the national unemployment rate is 8%. They wondered if this held true in their town, so they took a sample of 200 residents to test $H_0: p = 0.08$ vs $H_a: p \neq 0.08$, where p is the proportion of residents in the town that are unemployed. The sample included 22 residents who were unemployed. Test at $\alpha = 5\%$ l.o.s. Also compute C.I. at 95% confidence level.

Sol^n. $H_0: p = 0.08$ (here, $\hat{p}$ (population proportion) = 0.08 (given))

$H_a: p \neq 0.08$

$\therefore H_0: p = \hat{p}$

and $H_a: p \neq \hat{p}$



now, $z = \dfrac{p - \hat{p}}{\sqrt{pq/n}} = \dfrac{0.11 - 0.08}{\sqrt{\frac{0.11 \times 0.89}{200}}} = \dfrac{0.03 \times 10}{\sqrt{0.048}} = \dfrac{0.3}{0.219}$

$= 1.369$

$\therefore z_{cal} = 1.369 < 1.96$

$\therefore z_{cal} < z_{tab} \Rightarrow$ accept H_0 i.e. sample is drawn from population.

C.I. at 95% is given by

$p - z_{\alpha/2}\sqrt{pq/n} < \hat{p} < p + z_{\alpha/2}\sqrt{pq/n}$

$\Rightarrow 0.11 - 1.96 \times 0.219 < \hat{p} < 0.11 + 1.96 \times 0.219$

$\Rightarrow 0.11 - 0.429 < \hat{p} < 0.11 + 0.429$

$\Rightarrow -0.319 < \hat{p} < 0.539$

$\therefore$ CI at 95% = $[-0.319, 0.539]$

Ques 5. Two proportion test : Check whether two populations proportions are identical. In other words, whether there is any significant difference between two proportions.

Use : $z = \dfrac{p_1 - p_2}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} \sim N(0,1)$ under H_0

$= 0$, otherwise.

C.I. is given by

- $(p_1 - p_2) - z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < \hat{p}_1 - \hat{p}_2 < z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} + (p_1 - p_2)$ (two sided test)
- $(p_1 - p_2) - z_\alpha\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < \hat{p}_1 - \hat{p}_2 < (p_1 - p_2) + z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ (one sided test)

→ The researchers found out that 15 out of 36 small villages had a literacy rate < 75% while 18 out of 25 large villages had a literacy rate < 75%. At $\alpha = 5\%$ l.o.s test the claim that there is no difference in the proportion of small & large villages with a literacy rate < 75%. Also compute 99% of confidence limits.

Soln. $z_{1\%} = 2.58$ and $z_{5\%} = 1.96$

$p_1 = 15/36 = 0.41$ and $q_1 = 1 - p_1 = 1 - 0.41 = 0.59$

$p_2 = 18/25 = 0.72$ and $q_2 = 1 - p_2 = 1 - 0.72 = 0.28$

$H_0 : \hat{p}_1 = \hat{p}_2$ i.e. proportions are identical

$H_a : \hat{p}_1 \neq \hat{p}_2$ i.e. proportions are not identical.

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} = \frac{0.41 - 0.72}{\sqrt{\frac{0.41 \times 0.59}{36} + \frac{0.72 \times 0.28}{25}}} = \frac{-0.31}{\sqrt{\frac{0.2419}{36} + \frac{0.2016}{25}}}$$

$$= \frac{-0.31}{\sqrt{0.0067 + 0.008}} = \frac{-0.31}{\sqrt{0.0147}} = \frac{-0.31}{0.121} = -2.56$$

$\therefore z_{cal} = 2.56 > 1.96$ i.e. $z_{cal} > z_{tab} \Rightarrow$ reject H_0

C.I. at 99%,

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

$\Rightarrow -0.31 - 2.58 \times 0.121 < \hat{p}_1 - \hat{p}_2 < -0.31 + 2.58 \times 0.121$

$\Rightarrow -0.31 - 0.312 < \hat{p}_1 - \hat{p}_2 < -0.31 + 0.312$

$\Rightarrow -0.622 < \hat{p}_1 - \hat{p}_2 < 0.002$

$\therefore$ CI at 99% = $[-0.002, 0.622]$

One - sample and two - sample test $(n \le 29)$

Ques 6. One sample student 't' test : Check whether the sample is drawn from normal population or whether means are identical or not

Use : $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}} \sim N(0,1)$ under H_0 with d.f. $(n-1)$

→ A random sample of 9 values from a normal population showed a mean of 41.5 and the sum of squares of deviations from the mean equal to 72. Test whether the assumption of mean 44.5 in the population is reasonable. Also calculate 95% CI.

Soln. $n = 9$, $\mu = 44.5$, $\bar{x} = 41.5$ and $\sum_{i=1}^{9} (x_i - \bar{x})^2 = 72$

$H_0 : \mu = \mu_0$ ie. means are identical.

$H_a : \left.\begin{array}{l} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{array}\right\}$ two tail test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{41.5 - 44.5}{s/\sqrt{9}} \quad — (1)$$

here, $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{8}(72)$ (given)

$s^2 = 9 \Rightarrow s = 3$

$$\therefore t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{-3}{\sqrt{9}/3} \quad \text{(from (1))}$$

$\Rightarrow t = -3$.

we know that t_{tab} at $\alpha = 5\%$ with d.f. $= 9 - 1 = 8$ is 2.306

$\therefore |t_{cal}| > |t_{tab}|$ ie. $3 > 2.306$ hence reject H_0.

Therefore, means are not identical.

C.I. at 95%,

$$\bar{x} - t_{\alpha/2}\sqrt{\frac{s^2}{n}} < \mu < \bar{x} + t_{\alpha/2} \times s/\sqrt{n}$$

$\Rightarrow 41.5 - 2.306 \times 3/3 < \mu < 41.5 + 2.306 \times 3/3$

$\Rightarrow 41.5 - 2.306 < \mu < 41.5 + 2.306$

$\Rightarrow 39.194 < \mu < 43.806$.

CI. at 95% : [39.194, 43.806]

(Independent ...) Ques 7. Two sample ... drawn from ... Use : ...

(Independent)

Ques 7. Two sample student 't' test : Check whether two samples are drawn from normal population or populations are identical.

Use : • $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}}} \sim N(0,1)$ under H_0.

• $t = \dfrac{\bar{x} - \bar{y}}{s_p\sqrt{\dfrac{1}{n_1-1} + \dfrac{1}{n_2-1}}}$ (pool std. deviation : s_p)

where $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

here, d.f. $= (n_1+n_2-2)$

C.I. at $(100-\alpha)\%$: • $(\bar{x}-\bar{y}) - t_{\alpha/2}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}} < (\mu_1 - \mu_2) < (\bar{x}-\bar{y}) + t_{\alpha/2}\sqrt{\dfrac{s_1^2}{(n_1-1)} + \dfrac{s_2^2}{(n_2-1)}}$ (for two-tail test)

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

→ The prices of ghee were compared in two cities, for this purpose, ten shops were selected at random in each city. The following table gives per kg prices of ghee in two cities.

City A	361	363	356	364	359	360	362	361	358	357
City B	368	369	370	366	367	365	371	372	366	367

Test whether the average price of ghee is of same order in 2 cities. Also compute CI at 95% confidence interval.

soln: $H_0 : \mu_A = \mu_B$ i.e. average price of ghee is of same order in cities A and B.

$H_1 : \mu_A \neq \mu_B$ (two tail test) average price of ghee is of different order in cities A and B.
$\mu_A > \mu_B$
$\mu_A < \mu_B$

use $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

Scanned by CamScanner

	City A			City B	
x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	y_j	$y_j-\bar{y}$	$(y_j-\bar{y})^2$
361	0.9	0.81	368	-0.1	0.01
363	2.9	8.41	369	0.9	0.81
356	-4.1	16.81	370	1.9	3.61
364	3.9	15.21	366	-2.1	4.41
359	-1.1	1.21	367	-1.1	1.21
360	-0.1	0.01	365	-3.1	9.61
362	1.9	3.61	371	+2.9	8.41
361	0.9	0.81	372	3.9	15.21
358	-2.1	4.41	366	-2.1	4.41
357	-3.1	9.61	367	-1.1	1.21
3601		60.9	3681		48.9

$\bar{x} = \frac{3601}{10} = 360.1$ and $\bar{y} = \frac{3681}{10} = 368.1$

$$S_p^2 = \frac{1}{(10+10-2)}\left((n_1-1)s_1^2 + (n_2-1)s_2^2\right) = \frac{1}{18}(60.9 + 48.9)$$

$$= \frac{109.8}{18} = 6.1$$

$$\therefore t = \frac{\bar{x}-\bar{y}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{360.1 - 368.1}{\sqrt{6.1} * \sqrt{\frac{1}{10}+\frac{1}{10}}} = \frac{-8}{\sqrt{\frac{6.1}{5}}} = \frac{-8}{1.1045}$$

$$= -7.24$$

We know that $t_{tab} = 2.10$ at $\alpha = 5\%$ with d.f. 18 (n_1+n_2-2) & for two tail test.

$\therefore$ $|t_{cal}| \not< |t_{tab}|$ $\therefore$ reject H_0 i.e. average prices of ghee in both the cities are different.

C.I. at 95% :

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

$$\Rightarrow -8 - 2.1 \times 1.1045 < (\mu_1-\mu_2) < -8 + 2.1 \times 1.1045$$

$\Rightarrow \quad -8 - 2.319 \quad < (\mu_1 - \mu_2) < -8 + 2.319$

$\Rightarrow \quad -10.31 \quad < (\mu_1 - \mu_2) < -5.69$

, Given is the following data for male and female ~~working~~ working efficiency. At the 1% l.o.s., is there sufficient evidence to conclude that a difference exist b/w the mean no. of working hour per week for male & female employee. compute 99% confidence interval.

~~Soln.~~

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

Soln. $H_0: \mu_1 = \mu_2$ i.e. no difference between means.

$H_a: \mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$ } two tail test.

$$t = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{S_1^2}{(n_1-1)} + \frac{S_2^2}{(n_2-1)}}} = \frac{2.5 - 3.8}{\sqrt{\frac{2.2}{17} + \frac{3.5}{19}}} = \frac{-1.3}{\sqrt{0.129 + 0.184}}$$

$$t = \frac{1.3}{\sqrt{0.313}} = \frac{1.3}{0.559} = 2.32$$

at $\alpha = 1\%$ and d.f. $= n_1 + n_2 - 2 = 18 + 20 - 2 = 36$ (for two-tail test)

$t_{tab} = 2.719$

$\therefore \quad t_{cal} < t_{tab} \Rightarrow$ accept H_0 i.e. means are identical.

99% CI : $(\bar{X} - \bar{Y}) - t_{\alpha/2}\sqrt{\frac{S_1^2}{(n_1-1)} + \frac{S_2^2}{(n_2-1)}} < (\mu_1 - \mu_2) < (\bar{X} - \bar{Y}) + t_{\alpha/2}\sqrt{\frac{S_1^2}{(n_1-1)} + \frac{S_2^2}{(n_2-1)}}$

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

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

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

i.e. CI. $= [-0.219, 2.819]$

• Paired Student 't' test: used when sample sizes are equal i.e. $n_1 = n_2 = n$, and when samples are not independent, instead they are paired together. It tests if the sample means differ significantly or not.

use: $t_D = \frac{\bar{D} - \mu_D}{S/\sqrt{n}} \sim N(0,1)$ under H_0 and d.f = n-1,

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

C.I: $\bar{D} - t_{\alpha/2}\left[\frac{S_D}{\sqrt{n}}\right] < \mu_D < \bar{D} + t_{\alpha/2}\left[\frac{S_D}{\sqrt{n}}\right]$

→ A certain stimulus administered to each of 12 calves resulted in the following changes in blood sugar levels 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6. Can it be concluded that stimulus will in general be accomplished by increase in blood sugar level? Test at 5% l.o.s.

Soln. H_0: $\mu_x = \mu_y$ or $\mu_d = 0$ i.e. no difference in blood sugar levels of the calves before and after the administering drug.

H_1: $\mu_x < \mu_y$ or $\mu_d < 0$ i.e. stimulus result in increase in blood sugar level of calves.

D_i:	5	2	8	-1	3	0	-2	1	5	0	4	6	$\sum_{i=1}^{12} d_i = 31$
D_i^2:	25	4	64	1	9	0	4	1	25	0	16	36	$\sum_{i=1}^{12} d_i^2 = 185$

$$\bar{D} = \frac{\sum_{i=1}^{12} D_i}{N} = \frac{31}{12} = 2.583$$

$$S^2 = \frac{1}{(n-1)}\left[\left(\sum_{i=1}^{12} D_i^2\right) - \frac{\left(\sum_{i=1}^{12} D_i\right)^2}{n}\right]$$

$$= \frac{1}{11}\left[185 - \frac{(31)^2}{12}\right] = \frac{1}{11}(185 - 80.08)$$

$$= 104.92/11 = 9.538$$

$s = \sqrt{9.538} = 3.088$

$\therefore \quad t = \dfrac{\bar{D} - \mu_D}{s/\sqrt{n}} = \dfrac{2.583 - 0}{3.088/\sqrt{12}}$ (since $\mu_D = 0$ under H_0)

$= \dfrac{2.583}{0.8915} = 2.897$

Since it is one tail test ~~as~~ there may increase in blood sugar level

$\therefore t_{tab} = 1.80$ at d.f. $= (n-1) = 12-1 = 11$ and $\alpha = 5\%$

$\therefore t_{cal} = 2.897 > t_{tab}$ $\therefore$ reject H_0.

hence, stimulus is effective in increasing sugar levels.

C.I. at 95%,

$\bar{D} - t_{\alpha/2}\left(\dfrac{S_D}{\sqrt{n}}\right) < \mu_D < \bar{D} + t_{\alpha/2}\left(\dfrac{S_D}{\sqrt{n}}\right)$

$\Rightarrow 2.583 - 1.8 \times 0.89 < \mu_D < 2.583 + 1.8 \times 0.89$

$\Rightarrow 2.583 - 1.602 < \mu_D < 2.583 + 1.602$

$\Rightarrow 0.981 < \mu_D < 4.185$

CI : $[0.981, 4.185]$

Scanned by CamScanner