Statistics

Date: 2080/11/01
Page: 1

Q1) Suppose a population of 4 computers with their lifetimes 3, 5, 7 and 9 years. Comment on the population distribution. Assuming that you sample with replacement, select all possible samples of $n=2$, and construct sampling distribution of mean and compare the population distribution and sampling distribution of mean. Compare population means and population variance versus variance of sample means and comment on them with the support of theoretical consideration if any.

→ Soln,

Given,

Size of Population $(N) = 4$. $(n=2)$

Size of Sample $(n) = N^n = 4^2 = 16$.

Sample(x_i)	$\bar{x}_i$	$s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2 = \frac{1}{n-1}(\sum x^2 - n\bar{x}^2)$
(3,3)	3	0
(3,5)	4	2
(3,7)	5	8
(3,9)	6	18
(5,3)	4	2
(5,5)	5	0
(5,7)	6	2
(5,9)	7	8
(7,3)	5	8
(7,5)	6	2
(7,7)	7	0
(7,9)	8	2
(9,3)	6	18

Sample distribution of mean variance.

Date:
Page:

(9,5)	7	8
(9,7)	8	2
(9,9)	9	0
	$\Sigma \bar{x}_i = 96$	$\Sigma s^2 = \frac{80}{16} = 5$

$$E(\bar{x}) = \frac{\Sigma \bar{x}_i}{n} = \frac{96}{16} = 6.$$

$$\bar{x} = \mu = \frac{3+5+7+9}{4} = \frac{24}{4} = 6.$$

$$S^2 = \frac{1}{N} \Sigma (x-\bar{x})^2 = \frac{1}{N}\left(\Sigma x^2 - n\bar{\Sigma x}^2\right)$$

$$= \frac{1}{4}\left((3^2+5^2+7^2+9^2) - (4\times 6^2)\right)$$

$$= \frac{1}{4}(164 - 144)$$

$$= 5$$

<u>Therefore;</u>

Population distribution mean (μ) and sampling distribution mean ($\bar{x}$) are equal. i.e.

$$\mu = \bar{x} = 6$$

Population variance $\{V(\sigma)$ or $S^2\}$ and variance of sample mean $\{V(\bar{x})$ or $s^2\}$ are also equal. i.e.

Date:
Page:

$S^2 = s^2 = 5$

var popn — sample var

$E(\bar{x}) = \mu$:→ Sample mean is an unbiased estimate of population mean.

$E(s^2) = S^2$:→ Sample mean square is an unbiased estimate of population mean square

Q2) Describe the concept of sampling distribution of mean with reference to the population data (20, 21, 22 & 23) of size 4. In order to explain this, perform simple random sampling with replacement taking all possible samples with sample size n = 2. While describing the sampling distribution following issues will be covered. Population mean & population variance and its distribution. Sample mean and sample variance and its distribution. Comparision of population mean and sample mean; population variance and sample variance; population distribution and sampling distribution and sampling distribution based on the given data. Standard error of mean, final comment based on ~~the~~ your result.

⇒ Sol^n,

Given,

Population Data : 20, 21, 22 and 23

Population Size : 4 (N)

Sample size (n) = 2

Let, Population mean = μ

Population variance = σ^2

Sample mean = $\bar{x}$

Sample Variance = s^2

Sample (x_i)	$\bar{x}_i$	$s^2 = \frac{1}{n-1}\left(\Sigma x^2 - n\bar{x}^2\right)$
(20, 20)	20	0
(20, 21)	20.5	0.5
(20, 22)	21	2
(20, 23)	21.5	4.5
(21, 20)	20.5	0.5
(21, 21)	21	0
(21, 22)	21.5	0.5
(21, 23)	22	2
(22, 20)	21	2
(22, 21)	21.5	0.5
(22, 22)	22	0
(22, 23)	22.5	0.5
(23, 20)	21.5	4.5
(23, 21)	22	2
(23, 22)	22.5	0.5
(23, 23)	23	0
$n_i = 16$	$\Sigma\bar{x}_i = 344$	$\Sigma s^2 = 20$

Date :

Page:

$$\bar{x} = \frac{\sum x_i}{n_i} = \frac{344}{16} = 21.5$$

$$\mu = \frac{20+21+22+23}{4} = 21.5 \quad (x).$$

$$\sigma^2 = S^2 = \frac{1}{N}\left(\sum x^2 - N\bar{x}^2\right)$$

$$= \frac{1}{4}\left((20^2+21^2+22^2+23^2) - 4 \times 21.5^2\right)$$

$$= 5/4 = 1.25.$$

$$s^2 = \frac{\sum s^2}{n_i} = \frac{20}{16} = 1.25.$$

$\therefore$ Population mean $(\mu) = 21.5$

Population Variance $(\sigma^2) = 1.25$

Sample mean $(\bar{x}) = 21.5$

Sample Variance $(s^2) = 1.25$

Population mean (μ) = Sample mean $(\bar{x}) = 21.5$.

Population Variance (σ^2) = sample variance $(s^2) = 1.25$

Date : ....
Page: ....

Standard Error (S.E.) $= \sqrt{V(\bar{x})}$

$= \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}}$

$= \frac{1.118}{\sqrt{2}} = 0.79$

Q3) Explain Sampling distribution of mean with some reference to numerical example. Illustrate the practical implication of central limit theorem in inferential statistics

→ The Sampling distribution of mean is a fundamental concept in statistics. It describe the distribution of sample means obtained from repeated random sampling of population.

The central limit theorem state that, the distribution of the sum of a large number of independent, identical distributed random variables approach a normal distribution, regardless of the original population distribution.

E.g. let's consider a population with the values $\{2, 4, 6, 8, 10\}$

Suppose we take sample size $(n=2)$ and calculate mean from each sample

Date :

Page:

Sample 1 = (2, 4) → Mean = (2 + 4)/2 = 3.
Sample 2 = (6, 8) → Mean = (6+8)/2 = 7
Sample 3 = (4, 10) → Mean = (4 + 10)/2 = 7.

→ The central limit Theorem has practical implications in constructing confidence intervals, hypothesis testing, determining sample sizes, quality control processes and constructing prediction intervals. Essentially, the CLT allows researchers to make reliable inferences about population parameters based on sample data, even when the population distribution is unknown or non-normal.

Q4) A study of 1000 computer engineers conducted by their professional organization reported that 300 stated that their firms greatest concern was to uplift the professional quality of work. In order to conduct a follow up study to estimate the population proportion of computer engineers to fulfill their greatest concern within ± 0.01 with 99% confidence interval how many computer engineers would be required to be surveyed?

Date :
Page:

⇒ Sol^n,
Given,
Sample Population Size $(n) = 1000$
No. of success $(x) = 300$
Probability of success $(P) = \frac{300}{1000} = 0.3$

failure $(q) = 1 - P = 1 - 0.3 = 0.7$
Error margine $(e) = 0.01$

For 99 % C.I. the value of Z is $Z_\alpha = 2.326$
Now,
Sample size$(n) = \frac{Z_\alpha^2 \, P.q}{e^2}$

$= \frac{(2.326)^2 \times 0.3 \times 0.7}{(0.01)^2}$

$= 11361.57 \approx 11362$

∴ 11362 Computer engineer would be required to be surved.

5) What do you understand by estimation? If we want to determine average mechanical aptitude of a large group of workers, how large a random sample is need to be able to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 2.0 points? Assume that population standard deviation is 30.

⇒ Estimation is any procedures used to calculate the value of a population drawn from observations within a sample size drawn from that population.

⇒ Sample size $(n) = ?$
Level of siginificance $(\alpha) = 0.95$
$\alpha = 1 - 0.95 = 0.05$.
Margin Error $(e) = 2$
Standard Deviation $(\sigma) = 30$
We have,

$$n = \frac{Z_{\alpha}^{2}\sigma^{2}}{d^{2}}$$

$$= \frac{1.96^{2} \times 30^{2}}{2^{2}}$$

$$= 864.36 \approx 863$$

∴ Sample size $(n) = 863$ is needed.

6) The average time taken by server to execute an algorithm varies from time to time. From the past experience it is known that the time taken is normally distributed with standard deviation of 6.7 minutes. The IT manager wishes to estimate the average by drawing a random sample such that the probability is 0.95 that the mean of the sample will not deviate by more than 1 minute from the population mean. What should be sample size.

Date:
Page:

⟹ Sol^n // Sample size $(n) = ?$

Standard deviation $(\sigma) = 6.7$ minutes

$1 - \alpha = 0.96$

$\alpha = 0.05$

margin error $(e) = 1$ minutes

we have,

$$n = \frac{Z_{\alpha}^2 \cdot \sigma^2}{e^2}$$

$$= \frac{Z_{0.05}^2 \times 6.7^2}{1^2}$$

$$= \frac{1.96^2 \times 6.7^2}{1^2} = 172.44 \approx 172$$

∴ Sample size should be 172.

7) An effort to estimate the mean amount per Customer for dinner at a major Atlanta restaurant data were collected for a sample of 49 customers and a sample mean is found as \$ 24.8. Assume population standard deviation is \$ 5.

a) Compute standard error of mean

b) Find 95% confidence interval estimate for the population mean.

⟹ Sol^n //

Date: ....
Page: ....

Sample size (n) = 49
Sample mean ($\bar{x}$) = \$24.8
Population Standard Deviation (σ) = \$5
Standard Error (S.E) = ?
Confidence interval (C.I) = ?

We have,

a) $\Rightarrow$ $S.E = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{49}} = 0.7142$

b) $\Rightarrow$ $1 - \alpha = 0.95$

$\alpha = 0.05$

$Z_{0.05} = 1.96$

$CI = \bar{x} \pm S.E(\bar{x}) Z_\alpha$

$= 24.8 \pm 0.714 \times 1.96$

$= 24.8 \pm 1.399$

Now,

Taking (+ve)	Taking (-ve)
$24.8 + 1.399$	$24.8 - 1.399$
$= 26.19$	$= 23.40$.

$\therefore$ 95% confidence interval estimate for the population mean is 23.40 to 26.19

Date:
Page:

8) A Survey was conducted among 70 students studying B.Sc. CSIT in some Colleges randomly. Among them 50 students secured more than 80% marks in Statistics. Compute 99% and 95% Confidence Interval for population proportion of students who secured more than 80% marks in the subject statistics and comment on the results.

⟹ Soln,

Sample Size $(n) = 70$

Proportion of student secured more than 80% marks in statistics is

$$P = \frac{x}{n} = \frac{50}{70} = 0.71, \quad q = 1-P = 0.29$$

Standard Error $S.E(P) = \sqrt{\frac{Pq}{n}} = \sqrt{\frac{0.71 \times 0.29}{70}}$

$$= 0.054$$

Now,

At 95% Confidence Interval for P,

$= P \pm Z_\alpha S.E(P)$

$= 0.71 \pm 1.645 \times 0.054$

$= 0.71 \pm 0.08$

Taking +ve

$0.71 + 0.08$

$= 0.79$

Taking −ve

$0.71 - 0.08$

$= 0.63$

Hence, 95% Confidence Interval lies between 0.63 and 0.79

Date :

Page:

Again,

At 99 % Confidence Interval for P,

$= P \pm S.E(P)\, Z_\alpha$

$= 0.71 \pm 0.054 \times 2.326$

$= 0.71 \pm 0.125$

Taking +ve

$0.71 + 0.125$

$= 0.83$

Taking -ve

$0.71 - 0.125$

$= 0.58$

Hence, 99 %. Confidence Interval lies betn 0.58 and 0.83

9) A manufacturer of computer paper has a production process that operates continuously throughout an entire production shift. The paper is expected to have an average length of 11 inches and standard deviation is known to be 0.01 inch. Suppose random sample of 100 sheets selected and the average paper length is found to be 10.68 inches. Set up 95% & 99 %. Confidence interval estimate of the Population average Paper length.

⟹ Soln,

Standard deviation of Paper $(\sigma) = 0.01$ inch

Average length of Paper $(\mu) = 11$ inches

Sample of paper sheet size $(n) = 100$

Sample average length of paper $(\bar{x}) = 10.68$

Date :
Page:

- At 95% Confidence Interval of μ (mean)

$$1-\alpha = 0.95$$
$$\alpha = 0.05$$

$$\bar{x} \pm S.E(\bar{x}) Z_\alpha$$
$$= 10.68 \pm \frac{\sigma}{\sqrt{n}} Z_{0.05}$$
$$= 10.68 \pm \frac{0.01}{\sqrt{100}} \times 1.96$$
$$= 10.68 \pm 0.00196$$

Taking +ve
$$10.68 + 0.00196$$
$$= 10.68$$

Taking -ve
$$10.68 - 0.00196$$
$$= 10.67$$

$\therefore$ 95% CI lies between 10.67 and 10.68.

- At 99% CI of μ

$$1-\alpha = 0.99, \quad \alpha = 0.01$$

$$\bar{x} \pm S.E(\bar{x}) Z_\alpha$$
$$= 10.68 \pm \frac{\sigma}{\sqrt{n}} \times Z_{0.01}$$
$$= 10.68 \pm \frac{0.01}{10} \times 2.576$$
$$= 10.68 \pm 0.002576$$

Date:
Page:

Taking +ve

$10.68 + 0.002576$

$= 10.682$

Taking -ve

$10.68 - 0.002576$

$= 10.677$

$\therefore$ 99 CI lies between 10.677 and 10.682.

10) A machine produce metal rods used in an automobile suspension system. A random sample of 6 rods is selected and diameter is measured. The measuring data (in millimeters) are as follows. Assuming that the sample drawn from the normally distributed population.

8.24	8.26	8.20	8.28	8.21	8.23

Find 95% two sided confidence interval on the mean rod diameter, and interpret the result with reference to the given problem.

$\Rightarrow$ Soln,

Sample size (n) = 6

x	x^2
8.24	67.89
8.26	68.22
8.20	67.24
8.28	68.55
8.21	67.40
8.23	67.73
$\Sigma x = 49.42$	$\Sigma x^2 = 407.03$

Date :
Page:

$$\bar{x} = \frac{\sum x}{n} = \frac{49.42}{6} = 8.23$$

$$S^2 = \frac{1}{n-1}\left[\sum x^2 - n\bar{x}^2\right]$$

or, $S^2 = \frac{1}{6-1}\left[407.03 - 6\times(8.23)^2\right]$

or, $S^2 = 0.12$

$\therefore S = \sqrt{0.12} = 0.34$

For Confidence Interval

$\alpha = 0.05$

$\bar{x} \pm S.E(\bar{x})\, t_\alpha (n-1)$

$= \bar{x} \pm \frac{s}{\sqrt{n}}\, t_{0.05}(n-1)$

$= 8.23 \pm \frac{0.34}{\sqrt{6}}\, t_{0.05}(5)$

$= 8.23 \pm 0.138 \times 2.571$

$= 8.23 \pm 0.35$

Taking +ve

$8.23 + 0.35$

$= 8.58$

Taking -ve

$8.23 - 0.35$

$= 7.88$

Hence, CI of 95% lies between 7.88 and 8.58

Date:

Page:

11) A random sample of 10 bulbs has the following life in moths; 24, 26, 32, 28, 20, 20, 23, 34, 30 and 43. Obtain the 95% confidence limit for the population mean life of bulbs.

→ Sol^n//

Sample size $(n) = 10$

x	x^2
24	576
26	676
32	1024
28	764
20	400
20	400
23	529
34	1156
30	900
43	1849
$\Sigma x = 280$	$\Sigma x^2 = 8294$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{280}{10} = 28$$

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

$$= \sqrt{\frac{1}{10-1}\left(8294 - 10 \times 28^2\right)}$$

$$= 7.10$$

Date :
Page:

For Confidence limit

$\alpha = 0.05$

$\bar{x} \pm S.E(\bar{x})\, t_\alpha(n-1)$

$= 28 \pm \frac{s}{\sqrt{n}}\, t_{0.05}(9)$

$= 28 \pm \frac{7.10}{\sqrt{10}} \times 2.262$

$= 28 \pm 5.078$

Taking +ve

$28 + 5.078$

$= 33.078$

Taking -ve

$28 - 5.078$

$= 22.922$

Hence, 95% confidence limit for the population mean life of bulbs lies between 22.922 and 33.078.

12) In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected ~~selected~~ times is 37.7, with a sample standard deviation of 9.2. At the 1% level of significance, do these data provide considerable evidence that the mean number of concurrent users

Date : ....
Page: ....

is greater than 35 ? Draw your conclusion based on your result.

⇒ Sol^n //

Sample size $(n) = 100$
average Sample $(\bar{x}) = 37.7$
Sample S.D $(S) = 9.2$
$\alpha = 1\% = 0.01$

Problem to test

$H_0 : \mu = 35$
$H_1 : \mu_1 > 35$

Test statistics

$$Z = \bar{x} - \mu / (S/\sqrt{n})$$
$$= \frac{37.7 - 35}{9.2/\sqrt{100}} = \frac{2.7}{0.92}$$
$$= 2.93$$

Critical Value

At $\alpha = 0.01$, Critical value, $Z_\alpha = 2.326$

Decision

$Z = 2.93 > \overset{Z_\alpha =}{2.326}$
Reject H_0 at $\alpha = 0.01$.

Conclusion

The mean no. of concurrent users is greater than 35.

Date : ....
Page: ....

13) A dealer of a DELL Company located at New Road Claimed that the average life time of a multimedia projector produced by Dell Company is greater than 60,000 hours with standard deviation of 6000 hrs. In order to test his claim, sample of 100 DELL projectors are taken and the average life time was monitored and it was found to be 55,000 hours. Test the Claim of the dealer at 5% level of significance.

→ Sol^n,

Sample size $(n) = 100$
Average life time of Mul. projector $(\mu) = 60{,}000$
Sample Average " " $(\bar{x}) = 55{,}000$
Standard deviation $(\sigma) = 6000$
level of significance $(\alpha) = 0.05$

Problem to test
$H_0: \mu = 60{,}000$
$H_1: \mu > 60{,}000$ (one tail)

Test statistic:

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

$$= \frac{55{,}000 - 60{,}000}{6000/10}$$

$$= -8.33$$

Date :
Page:

Critical value $(\alpha = 0.05)$

Critical value is $Z_\alpha = 1.645$

Decision

$|Z| = 8.33 > Z_\alpha = 1.645$

$\therefore$ Reject H_0 at $\alpha = 0.05$.

Conclusion

The average life time of multimedia projector produced by Dell Company is greater than 60,000 hours.

14) It is Claimed that samsung and Redmi mobiles are equally popular in Kathmandu. A random Sample of 500 people from Kathmandu showed 300 have Samsung mobile. Test the claim at 5% level of significance.

$\Rightarrow$ Soln,

Sample size $(n) = 500$

Samsung mobile proportion $(p) = \frac{x}{n} = \frac{300}{500} = 0.6$

$q = 1 - p$

$= 1 - 0.6 = 0.4$.

$P = 1/2,\ Q = 1/2$.

Problem to test

$H_0 : P = 1/2$ i.e. equally Popular

$H_1 : P \neq 1/2$ i.e. not " "

Date :
Page:

Test statistic

$$Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.6 - 0.5}{\sqrt{(0.5 \times 0.5)/500}}$$

$$= \frac{0.1}{0.022} = 4.47$$

Critical value

At $\alpha = 0.05$, critical value is $Z_{\alpha} = 1.96$

Decision

$Z = 4.47 > Z_{\alpha} = 1.96$

Reject H_0 at $\alpha = 0.05$

Conclusion

Samsung and Redmi mobiles are not equally popular in Kathmandu.

15) A sample of 250 items from lot A contains 10 defective items and a sample of 300 items from lot B is found to contain 18 defective items. At significance level $\alpha = 0.05$, is there a significant difference between the quality of the two lots?

⟹ Soln,

let, sample size $= n$

Deffective Items $= x$.

$n_1 = 250$ $\quad n_2 = 300$

$x_1 = 10$ $\quad x_2 = 18$

$\alpha = 0.05$.

Problem Test

$H_0 :$ $P_1 = P_2$

$H_1 :$ $P_1 \neq P_2$

$$P_1 = \frac{x_1}{n_1} = \frac{10}{250} = 0.04$$

$$P_2 = \frac{x_2}{n_2} = \frac{18}{300} = 0.06$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{10+18}{250+300} = 0.05$$

$$Q = 1 - P = 1 - 0.05$$
$$= 0.95$$

Date : ....
Page: ....

Test Statistics

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

$$= \frac{0.04 - 0.06}{\sqrt{0.05 \times 0.95\left(\frac{1}{250}+\frac{1}{300}\right)}}$$

$$= -1.071$$

At $\alpha = 0.05$

Critical value: $Z_\alpha = 1.96$

Decision

$|Z| = 1.071 < Z_\alpha = 1.96$

Accept H_0 at $\alpha = 0.05$

Conclusion

No, there is not significant difference between the quality of the two lots.

16) In location I there are 250 corona positive cases out of 460 persons and in location 2, 250 positive cases reported out of 650 persons. Can it be concluded that proportion of corona positive cases is higher in location 1 compared to location 2? Test at 10 % level of significance.

⇒ Sol^n,

Sample size $= n$

Corona Positive Cases $= x$

$n_1 = 460$, $n_2 = 650$

$x_1 = 250$ $x_2 = 250.$

$\alpha = \cancel{0.01}\ 0.1$

Problem to Test

$H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$

$$P_1 = \frac{x_1}{n_1} = \frac{250}{450} = 0.55$$

$$P_2 = \frac{x_2}{n_2} = \frac{250}{650} = 0.38$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{250+250}{460+650} = 0.45$$

$$Q = 1 - P = 0.55$$

Test Statistics

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

$$= \frac{0.55 - 0.38}{\sqrt{0.45 \times 0.55\left(\frac{1}{460} + \frac{1}{650}\right)}}$$

$$= 5.60$$

Date :
Page:

Critical value

At $\alpha = \cancel{0.01}\ 0.1$, critical value is $Z_\alpha = 1.282$

Decision

$Z = 5.60 > Z_\alpha = 1.282$

Reject H_0 at $\alpha = 0.10$

Conclusion

Corona positive cases is higher in location 1 compared to location 2.

17) What do you mean by hypothesis? Describe null and alternative hypothesis. A company claims that its light bulbs are superior to those of the competitor on the basis of study which showed that a sample of 40 of its bulbs had an average life time 628 hours of continuous use with a standard deviation of 27 hours. While sample of 30 bulbs made by the competitor had an average life time 619 hours of continous use with a standard deviation of 25 hours. Test at 5% level of significance whether this claim is justified?

Date :
Page:

⇒ Sol^n//

Hypothesis is certain assumption about parameter of the population. There are two types: they are Null hypothesis (H_0) and Alternative hypothesis (H_1).

a) Null hypothesis

↳ Hypothesis of no difference and also called hypothesis of insignificant.

b) Alternative hypothesis

↳ Hypothesis of difference and It is also called hypothesis of significant.

⇒

Let, A company's light bulbs = 1

The Competitor's Company's light bulbs = 2.

sample standard deviation = S

$n_1 = 40$ $\qquad$ $n_2 = 30$

$\bar{x}_1 = 628$ $\qquad$ $\bar{x}_2 = 619$

$S_1 = 27$ $\qquad$ $S_2 = 25$

$\alpha = 0.05$

Test of hypothesis

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

CamScanner

Date.: ....
Page: ....

Test statistics

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{628 - 619}{\sqrt{\frac{27^2}{40} + \frac{25^2}{30}}}$$

$$= 1.44$$

Critical value

At $\alpha = 0.05$, critical value is $Z_\alpha = 1.645$

Decision

$Z = 1.44 < Z_\alpha = 1.645$

Accept H_0 at $\alpha = 0.05$

Conclusion

It's true, that a Company claims its light bulbs are superior to those of the competitor.

Date :
Page:

18) Based on the following information, performed the following:

i) Test whether two mean are significance different ($\alpha = 5\%$) using independent t-test.

ii) Compute 95% confidence interval estimation for the difference of mean.

iii) Show the linkage between testing of hypothesis and confidence interval estimation in this problem

	Group A	Group B
Sample mean	10	15
Sample Standard Deviation	3	5
Sample Size	49	64

→ Soln,

$n_A = 49$ $\qquad$ $n_B = 64$

$\bar{x}_A = 10$ $\qquad$ $\bar{x}_B = 15$

$S_A = 3$ $\qquad$ $S_B = 5$

Problem ~~to~~ Test

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$S^2 = \frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}$$

$$= \frac{49 \times 3^2 + 64 \times 5^2}{49 + 64 - 2}$$

$$= 18.38$$

Test Statistics

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{S^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

$$= \frac{10 - 15}{\sqrt{18.38\left(\frac{1}{49} + \frac{1}{64}\right)}} = -6.14$$

Date :
Page:

Critical value

At $\alpha = 0.05$, critical value is $t_\alpha(n_A+n_B-2)$

$= t_{0.05}(111) = 1.980$

Decision

~~At $\alpha = 0$~~ $|t| = 6.14 > t_\alpha(n_A+n_B-2) = 1.980$

Reject H_0 at $\alpha = 0.05$

Conclusion

There is no significant difference.

ii) Confidence interval estimation

$$\bar{x}_A - \bar{x}_B \pm \sqrt{s^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}\; t_\alpha(n_1+n_2-2)$$

$$10-15 \pm \sqrt{18.38\left(\frac{1}{49}+\frac{1}{64}\right)}\; t_\alpha(111)$$

$$-5 \pm 0.813 \times 1.980$$

$$-5 \pm 1.609$$

Taking +ve

$-5+1.609$

$= -3.391$

Taking -ve

$-5-1.609$

$= -6.609$

$\therefore$ Confidence interval lies between -6.609 to -3.391.

Date : ....

Page: ....

Q19) Two random sample of Nepalese people taken from rural and urban region gave the following data of income;

Sample from	Size	Average daily income	sd
Rural region	15	800	50
Urban region	10	1250	30

Test whether the average daily income of rural people is significantly less than that of urban people.

⇒ Soln, Let Rural Region = 1

Urban Region = 2

$n_1 = 15$, $\bar{x}_1 = 800$, $s_1 = 50$

$n_2 = 10$, $\bar{x}_2 = 1250$, $s_2 = 30$.

Problem to test:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

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

$$= \frac{15 \times 50^2 + 10 \times 30^2}{15 + 10 - 2}$$

$$= 2021.739$$

Test statistics

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{800 - 1250}{\sqrt{2021.739\left(\frac{1}{15} + \frac{1}{10}\right)}}$$

$$= \frac{-450}{18.356} = -24.515$$

CamScanner

Critical Value

At $\alpha = 0.05$, critical value is
$t_\alpha(n_1+n_2-2) = t_{0.05}(23) = 2.069$

Decision

$|t| = 24.515 > t_\alpha(n_1+n_2-2) = 2.069$
Reject H_0 at $\alpha = 5\%$.

Conclusion

The avg. Income of rural people is Significantly less than of Urban people.

Q20) Previous literature has reported that the average age of B.Sc.CSIT enrolling students in TU is 22 years. A researcher has doubts on this inform and he feels that the average age to be less than 22 years. In order to examine this followin Sample data were collected randomly from the enrolling Students of CSIT

Age in years: 20 19 22 23 19 20 21 22 19 20

Set up null & alternative hypothesis and test whether the researcher doubt will be justed. Use 5% level of significance. Assume that the Parent population from which Samples are drawn is normally distributed.

Date : ....

Page: ....

⇒ Sol^n

Age (x)	x^2
20	400
19	361
22	484
23	529
19	361
20	400
21	441
20	400
19	361
20	400
$\Sigma x = 203$	$\Sigma x^2 = 4137$

$$\bar{x} = \frac{\Sigma n}{n} = \frac{203}{10} = 20.3$$

$$S = \sqrt{\frac{1}{n-1}\left(\Sigma x^2 - n\bar{x}\right)^2}$$

$$= \sqrt{\frac{1}{10-1}\left(4137 - 10 \times (20.3)^2\right)}$$

$$= 1.33$$

Problem to test:

$H_0 : \mu = 22$

$H_1 : \mu < 22$

Date: ....
Page: ....

Test Statistic

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

$$= \frac{20.3 - 22}{1.33/\sqrt{10}} = -4.04$$

Critical value

At $\alpha = 0.05$, the critical value is $t_\alpha(n-1)$

$= t_{0.05}(10-1) = 1.833$

Decision

$|t| = 4.04 > t_\alpha(9) = 1.833$

Reject H_0 at $\alpha = 0.05$

Conclusion

The average age is less than 22.

Date :
Page:

Q.21) Two kinds of manure were applied to sixteen one hector plot, other condition remaining the same. The yield in quintals are given below.

Manure I	51	18	20	36	50	49	36	34	49
Manure II	29	28	26	35	30	44	46		

Is there any significant difference between the mean yields? Use 5% level of significance.

⇒ Soln

Manure I (x_1)	Manure II (x_2)	x_1^2	x_2^2
51	29	2601	841
18	28	324	784
20	26	400	676
36	35	1296	1225
50	30	2500	900
49	44	2401	1936
36	46	1296	2116
34		1156	$\Sigma x_2^2 = 8478$
49	$\Sigma x_2 = 238$	2401	
$\Sigma x_1 = 343$		$\Sigma x_1^2 = 14375$	

$$\bar{x}_1 = \frac{\Sigma x_1}{n_1} = \frac{343}{9} = 38.11$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{238}{7} = 34$$

$$s^2 = \frac{\sum x_1^2 - n_1\bar{x}_1^2 + \sum x_2^2 - n_2\bar{x}_2^2}{n_1+n_2-2}$$

$$= \frac{14375 - 9\times(38.11)^2 + 8478 - 7\times 34^2}{9+7-2}$$

$$= 120.69$$

Problem to test

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Test statistics

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{38.11 - 34}{\sqrt{120.69\left(\frac{1}{9}+\frac{1}{7}\right)}}$$

$$= 0.742$$

Critical value

At $\alpha = 0.05$, the critical value is $t_\alpha(n_1+n_2-2)$

$= t_{0.05}(14) = 2.145$

Decision

so,

$t = 0.742 < t_\alpha(14) = 2.145$ ∴ Accept H_0 at $\alpha = 0.05$

Conclusion :

There is not significant difference.

CamScanner

Date:
Page:

22) Define type I and type II error in testing of hypothesis. A psychologist wishes to verify that a certain drug increases the reaction time to given stimulus. The following reaction times (in tenth of seconds) were recorded before and after injection of the drug for each of four subjects

	Subject	1	2	3	4
x^n time	Before	7	2	12	12
	After	13	3	18	13

Test at 5% level of significance to determine whether the drug significantly increases reaction time.

⇒ Sol^n//

	H_0	H_1
H_0 (true)	Correct Decision	Wrong Decision (Type I error)
H_1 (true)	Wrong Decision (Type II error)	Correct Decision

In testing of hypothesis if we accept H_0 when H_0 is true is correct decision. If we reject H_0 when H_0 is true is wrong decision. It is called type I error.

If we reject H_0 when H_0 is false is correct decision. If we accept H_0 when H_0 is false is wrong decision. It is called type II error.

Date : ....
Page: ....

Reaction time

Subject	Before (x)	After (y)	$d = y - x$	d^2
1	7	13	6	36
2	2	3	1	1
3	12	18	6	36
4	12	13	1	1
			$\Sigma d = 14$	$\Sigma d^2 = 74$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{14}{4} = 3.5$$

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

$$= \sqrt{\frac{1}{3}\left(74 - 4 \times 3.5^2\right)}$$

$$= 2.88$$

Problem to test

$H_0 : \mu_x = \mu_y$

$H_1 : \mu_x < \mu_y$

Test Statistics

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{3.5}{2.88/\sqrt{4}} = 2.43$$

Critical Value: At $\alpha = 0.05$ the critical value is

$$t_\alpha(n-1) = t_{0.05}(3) = \cancel{3.182}\ 2.353$$

Decision:

$$t = 2.43 > t_\alpha(n-1) = \cancel{3.182}\ 2.353$$

Date :
Page:

~~Accept~~ Reject H_0 at $\alpha = 0.05$

Conclusion :

· The drug reaction time is significantly ~~same.~~ increases.

23) Following are the scores obtained by ~~10~~ 9 university staffs on the computer proficiency skills before training and after training. It was assumed that the proficiency of computer skills is expected to be increased after training.

	Staff	1	2	3	4	5	6	7	8	9
~~Score~~	Before training	50	30	15	22	34	45	40	10	26
	After Training	55	40	30	30	36	45	41	30	40

Test at 5% level of significance whether the training is effective to improve the computer proficiency skills applying appropriate statistical test. Assume that the given score follows normal distribution.

⇒

Date : ....

Page: ....

23

$\longrightarrow$ Sol^n//

Let, Before training $= x_1$

After training $= x_2$.

Staff	x_1	x_2	$d = x_2 - x_1$	d^2
1	50	55	5	25
2	30	40	10	100
3	15	30	15	225
4	22	30	8	64
5	34	36	2	4
6	45	45	0	0
7	40	41	1	1
8	10	30	20	400
9	26	40	14	196
			$\Sigma d = 75$	$\Sigma d^2 = 1015$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{75}{9} = 8.33$$

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

$$= \sqrt{\frac{1}{8}\left(1015 - 9 \times 8.33^2\right)}$$

$$= 6.98$$

problem to test

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$.

CamScanner

Date :
Page:

Test statistic

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

$$= \frac{8.33}{6.98/\sqrt{9}}$$

$$= 3.580$$

Critical value

At $\alpha = 0.05$ the critical value is $t_\alpha(n-1)$

$$= t_{0.05}(8) = 1.860$$

Decision

$t = 3.580 > t_\alpha(n-1) = 1.860$

Reject H_0 at $\alpha = 0.05$.

Conclusion.

Yes, The training is ~~not~~ effective to improve the Computer proficiency skills.

Date : ....
Page: ....

24) A drug was given to 10 patients. The Change in blood pressure of patients were recorded as 8, 10, −2, 0, 5, −1, 9, 12, 6 & 5. Is it reasonable to believe that drug has increase blood pressure? Use 5% level of significance.

⇒ sol^n //

Change in bp (d)	d^2
8	64
10	100
−2	4
0	0
5	25
−1	1
9	81
12	144
6	36
5	25
$\Sigma d = 52$	$\Sigma d^2 = 480$

$$\bar{d} = \frac{\Sigma d}{10} = \frac{52}{10} = 5.2$$

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

$$= 4.82$$

Date :
Page:

Problem to test

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

Test statistics

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

$$= \frac{5.2}{4.82/\sqrt{10}}$$

$$= 3.411$$

Critical value:

At $\alpha = 0.05$, the critical value is $t_\alpha(n-1)$

$= t_{0.05}(9) = 1.833$

Decision:

$t = 3.411 > t_\alpha(n-1) = 1.833$

Reject H_0 at $\alpha = 0.05$

Conclusion:

Yes, Drug has increase blood pressure.

Date :

Page:

25) Define Central limit theorem. The life of a certain brand of an electric bulb may be considered a random variable with mean 1350 hours and standard deviation 550 hours. Using central limit theorem. Find the probability that the average life time of 100 bulbs exceeds 1440 hrs.

⟹ Soln,

For a large sample size whatever be the distribution of population the sample follows normal distribution is called central limit theorem.

→ Mean $(\mu) = 1350$

standard deviation $(\sigma) = 550$

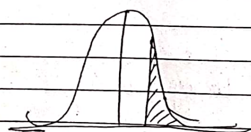
sample size $(n) = 100$

Sample mean $(\bar{x}) = 1440$

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

$$= \frac{1440 - 1350}{550/\sqrt{100}}$$

$$= 1.6364$$



$z=0$ $z=1.6364$

Date :

Page:

$$P(z > 1.6364) = 0.5 - P(0 < z < 1.6364)$$

$$= 0.5 - 0.4484$$

$$= 0.0516$$

∴ The probability that the average life time of 100 bulbs exceeds 1440 hrs is 0.0516.

CamScanner

Date:
Page:

26) Write short notes on
i) Properties of good estimator
⇒ A good estimator in statistics passes several desirable properties that contribute its effectiveness and reliability in estimating population parameter. Here are some key properties of a good estimator.

a) Unbiasedness:
An estimator is unbiased if an average it produces estimates that are equal to the true population parameter. In another words, the expected value of the estimator is equal to the true parameter.

b) Efficiency:
An efficient estimator has a several varience, meaning that it's estimates are relatively close to each other. Lower variance indicates greater ~~pressure~~ precision and reliability in estimation.

c) Consistency:
A consistent estimator converges to the true parameter as the sample size increases. In other words, the probability that the estimate deviate from the true parameter decreases as the sample size grows.

Date:
Page:

d) Sufficiency:
A sufficient statistic contains all the information necessary to make an unbiased estimation of the population parameter. It helps to reduce the dimensionality of the data while retaining relevant information.

ii) Estimation of minimum sample size for given proportion
→ To estimate the minimum sample size for a given proportion with a desired level of confidence and margin of error, you can use following formula:

$$n = \frac{Z^2 pq}{e^2} \qquad (\because q = p-1)$$

Where, n is the minimum sample size.
Z is the Z-score corresponding to the desired level of confidence. For example, for a 95% confidence level, Z would be approximately 1.96.
p is the estimated proportion of population
e is the margin of error.

The actual population size doesn't play a significant role unless it's very small (less than 30). If the population is large, a sample size above 30 is often sufficient.