module 5.

Sampling Distributions.

Population: (N)

Large collection of individuals / attributes / numerical data can be regarded as population or universe.

Sample: (n)

Sample is the finite subset of a population.

$n \geq 30$ - large sample.

$n < 30$ - small sample.

Example:

a) Population of India is - (N)
population of a state - (n)

b) Cars produced in India - (N)
A brand of car produced - (n).

- Statistical constants for population. such as mean (μ) and standard deviation (σ) are called parameters.
- Similarly constants for sample drawn from population such as mean (x) and standard deviation (s) are called statistics.

Random sampling.

Selection of an item from a population in such a way that each item has the chance of being selected is called random sampling.

↳ Sampling with replacement - item can be chosen more than once.
↳ sampling without replacement - item can be chosen only once.

If (N) - size of finite population & (n) - size of sample then we have N^n samples. - (with replacement condition).

Sampling where a member cannot be chosen more than once. - sampling without replacement.

total samples = ${}^{N}C_{n}$ samples.

Sampling Distribution.

population — set of samples — mean & SD of set — diff. mean & SD for even sample

frequency distribution of samples — grouped according to frequency.

SAMPLE DISTRIBUTION.

- The Standard Deviation of a sampling distribution is called standard error
- Reciprocal of standard error = precision.

Sampling distribution of means.

Consider a population for which the mean is (μ) and standard deviation is ($\sigma_{\bar{x}}^2$) - suppose we draw a set of samples of size n from the population - and compute the mean ($\mu_{\bar{x}}$) of each of these. The frequency distribution of these means is called sampling distribution of means.

mean = $\mu_{\bar{x}}$

std. deviation = $\sigma_{\bar{x}}^2$

a) $\mu_{\bar{x}} = \mu$

b) $\sigma_{\bar{x}}^2 = \frac{c\sigma^2}{n}$ [c is a constant]. $c = \frac{N-n}{N-1}$.

if n is large: $c \to 1$ as $N \to \infty$.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$ (with replacement)

Q) A population has mean 75 and std. dev 12

a) Random samples of size 121 are taken. Find mean & std. dev of sample.

b) How would the answers to part a) change if size of samples = 400

$\mu = 75$ $n = 121$.

$\sigma = 12$.

1) $\mu_{\bar{x}} = \mu$.

$\Rightarrow \mu_{\bar{x}} = 75$.

$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{12}{\sqrt{121}} = \frac{12}{11} = 1.09$.

2) $\mu_{\bar{x}} = \mu = 75$.

$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{400}} = \frac{12}{20} = 0.6$.

$\therefore$ when (n) increases the value of $\sigma_{\bar{x}}$ reduces

Q) population - $\mu = 5.75$. a) $n = 81$ - find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$
$\sigma_x = 1.02$. b) $n = 25$

Page :
Date :

1) $\mu = \mu_{\bar{x}} = 5.75$.
$\sigma_{\bar{x}} = \frac{1.02}{9} = 0.1133$

2) $\mu = \mu_{\bar{x}} = 5.75$.
$\sigma_{\bar{x}} = \frac{1.02}{5} = 0.204$.

∴ value of $\sigma_{\bar{x}}$ for (b) increases with decrease in (n).

Q) Weights of 1500 ball bearings are normally distributed with a mean of 635g. and std.dev - 1.36 g. If 300 samples of size - 36 are drawn from the population. Determine the expected mean of sampling distribution of means

a) With replacement

b) Without replacement.

a) $\mu = 635$
$\sigma = 1.36$
$n = 36$.
$\mu_{\bar{x}} = \mu = 635$.
$\sigma_{\bar{x}} = \frac{1.36}{6} = 0.2266$.

b) $\sigma_{\bar{x}} = \frac{\sqrt{c}\,\sigma_x}{\sqrt{n}}$.
$= \sqrt{\frac{1500-36}{1499}} \cdot \frac{1.36}{6}$.
$= 0.9882 \times 1.36/6$.
$= 0.224$.

Q) The population consists of 4 numbers 3, 7, 11, 15.

a) find the mean and std.dev of sampling distribution of means by considering sample size 2 with replacement.

b) If N, n denotes respectively the population and sample size. σ, σ_x - std.dev of population & sample. without replacement.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ - (Prove)

$\mu_{\bar{x}} = \mu$ [$\mu_{\bar{x}}$ - mean of sampling distribution] and [μ - population mean

$\mu = 9$
$\Sigma x^2 = 404$
std dev $= \sigma_x = \sqrt{\frac{404}{4} - 81} = \sqrt{20} = 4.4721$.

Camlin

With replacement.

a) $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{4.4721}{\sqrt{2}}$ / but formula to be proved.

samples.

(3,3) (3,7) (3,11) (3,15)

(7,3) (7,7) (7,11) (7,15)

(11,3) (11,7) (11,11) (11,15)

(15,3) (15,7) (15,11) (15,15)

Means of samples.

3, 5, 7, 9, 5, 7, 9, 11, 7, 9, 11, 13, 9, 11, 13, 15.

Frequency distribution of samples.

3	1
5	2
7	3
9	4
11	3
13	2
15	1

$\Rightarrow \mu_{\bar{x}} = 9.$

$\sigma_{\bar{x}} = \sqrt{\frac{1456}{16} - 81} = \sqrt{10}.$

b) without replacement

(3,7) (3,11) (3,15) (7,11) (7,15) (11,15) [no number chosen twice]

mean

5, 7, 9, 9, 11, 13.

5	1
7	1
9	2
11	1
13	1

$\mu_{\bar{x}} = 9$

$\sigma_{\bar{x}} = \sqrt{\frac{526}{6} - 81} = \sqrt{6.66}$

$= 2.58$

Now to prove.

$\sigma_x = 4.4721$

$\frac{\sigma_x}{\sqrt{n}} = 3.16225$

$\frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{4-2}{4-1}} = 1.489.$

Page :
Date :

Central Limit Theorem.

If the variable X has a non normal distribution with mean (μ) and std.dev (σ) then limiting distribution of $z = \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{x-\mu}{\sigma_{\bar{x}}}$

Confidence Limit

The confidence interval for population mean is determined by taking $\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm E$ where $\bar{x}$ is sample mean.
E is marginal error
z_c is critical value.

Q) Random sample of size $N = 100$ is taken from population with std dev $\sigma = 5.1$. given that sample mean $\bar{x} = 21.6$ obtain that 95% confidence interval for population mean (μ)

$z - 95\% - z_c = 1.96$.
$z - 99\% - z_c = 2.58$

$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$

$21.6 \pm 1.96 \frac{(5.1)}{10}$.

21.6 ± 0.9996.

$= 22.5996$ or 20.6004.

We are 95% sure that confidence interval for population mean is between $(20.6004 - 22.5996)$

Q) Certain tubes manufactured by a company has mean life of 800hrs. and std dev of 60 hrs find the probability that random sample of 16 tubes from population will have lifetime

a) b/w 790 hrs — 810 hrs.
b) < 785 hrs.
c) > 820 hrs.
d) b/w 770 - 830

$\mu = 800$
$\sigma = 60$
$n = 16$

$\sigma_{\bar{x}} = \sigma/\sqrt{n} = \frac{60}{4} = 15.$

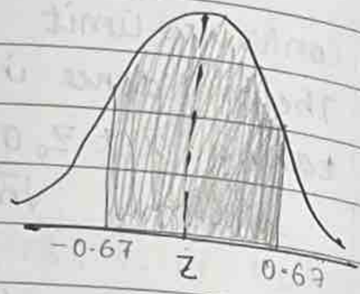
$$z = \frac{\bar{x} - 800}{15} \quad \left[z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \right] \quad (\bar{x}\text{ - question}).$$

1) $P\left[\frac{790-800}{15} < \bar{x} < \frac{810-800}{15}\right]$

$= P\left[-0.667 < z < 0.67\right]$

$= \phi(0.667) + \phi(0.667) \quad 0.24857$

$= 2\phi(0.667) = 2(\cancel{0.2486}) = 0.4972.$



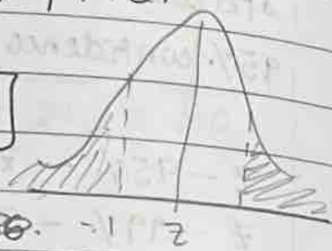
2) $P\left[\bar{x} < \frac{785-800}{15}\right] = P\left[\cancel{z} < -1\right]$ $\quad P(z>1)$

$= 0.5 - \phi(+1)$

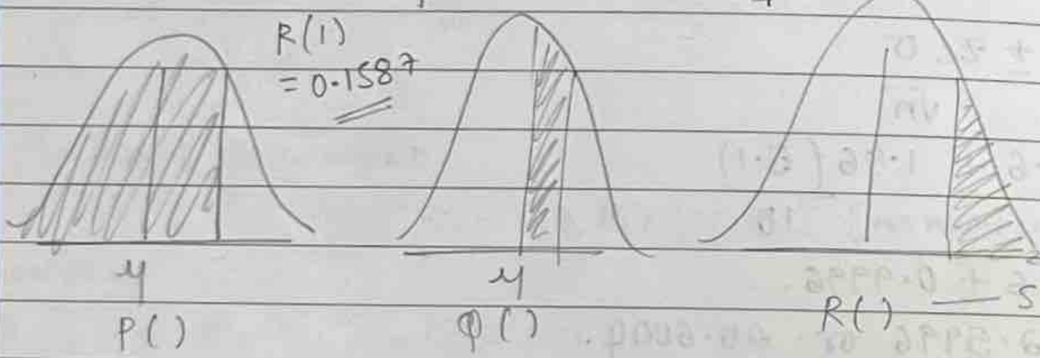
(or) by calci directly

$\cancel{= 0.5 - 0.15866}$

$\cancel{= 0.3413}.$



Imp.

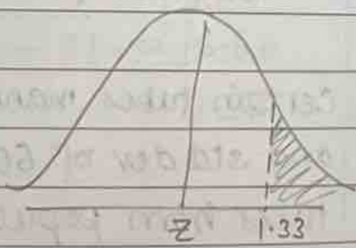


statistics
AC = optn - ▽
P( ) φ( ) — opt 4
R( ).

3) $P\left[\bar{x} > \frac{820-800}{15}\right] = P\left[z > 1.333\right]$

$= 0.09178.$

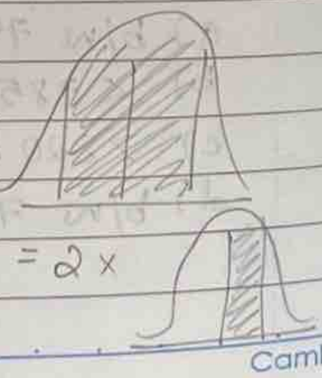
// use calci Directly.



4) $P\left[\frac{-30}{15} < z < \frac{30}{15}\right] = P\left[-2 < z < 2\right]$

$= 2P(0 < z < 2)$

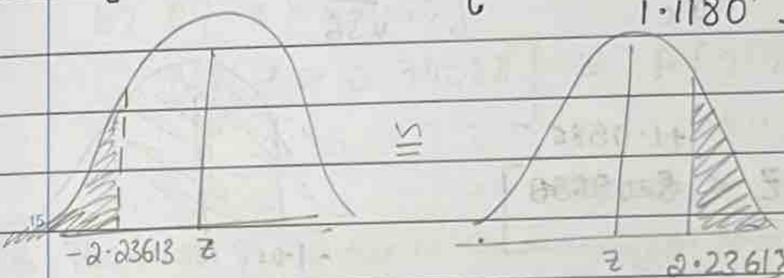
$= 0.9544.$



Q) A prototype automated tyre has a design life of 38500 miles with std.dev - 2500 miles. 5 such tires are manufactured & tested. On the assumption that the actual population std.dev is 2500 miles, find probability that, the sample mean will be < 36000 miles. Assume distribution of lifetime is normal.

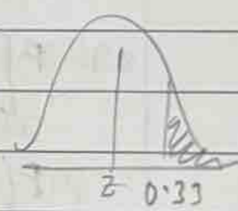
$\mu = 38500 = 38.5$ K miles.
$\sigma = 2500 = 2.5$ K miles.
$\bar{x} = < 36$ K miles.
$n = 5$.
$\sigma_{\bar{x}} = \frac{2.5}{\sqrt{5}} = 1.118067$.

$P[\bar{x} < 36000] = P\left[z < \frac{36-38.5}{1.1180}\right] = P[z < -2.23613]$.
$= 0.0126$.



Q) An automobile battery manufacturer claims that it midgrade battery has mean life 50 months with a std.dev of 6 months. Suppose the distribution of battery lives of this particular brand is normal.
a) On assumption that - claims are true find probability that randomly selected battery will last < 48 months.
b) P(sample size 36 - < 48 months).

$\mu = 50$
$\sigma = 6$
$z = \bar{x} - 50/6$. [case a)]
$\sigma_{\bar{x}} = 6/\sqrt{36} = 1$.



a) $P[z < -1/3] = P[z < -0.33] = P[z > 0.33]$
$= 0.36957$.
≈ 0.3707

b) $P(\bar{x} < 48) = P(Z < -2/1)$

$= P(Z < -2) \simeq P(Z > 2)$

$= 0.0228.$

Q> The weight of 1500 ball bearings are normally distributed with mean of 635 g and std. dev of 1.36 g. If 300 samples of size 36 are drawn from the population. In the case of replacement - find how many random samples would have mean b/w.

a) 634.76 g and 635.24 g.

b) > 635.6 g.

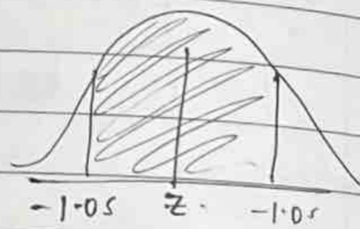
c) < 634.5 g or more than 635.24 g.

$n = 36$ $\qquad Z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$ $\qquad \sigma_{\bar{x}} = \dfrac{1.36}{\sqrt{36}} = 0.22667.$

$\sigma = 635.$

$\sigma = 1.36 g.$

a) $P\left[-1.0586 < Z < \cancel{} +1.0586\right].$



$= 2P(1.0586)$

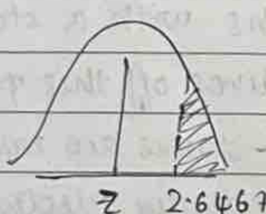
$= 0.71022.$ → for 300 samples → 300×0.71022

$= 213.066.$

$= 213$ bearings.

b) $P(\bar{x} > 635.6)$

$P(Z > 2.6467).$



$= 4.0641 \times 10^{-3}$

for 300 samples → $4.0641 \times 10^{-3} \times 300 = 1.219$

$= 1$ bearing.

c) $P[\bar{x} < 634.5] + P[\bar{x} > 635.24].$

$P(Z < -2.2055) + P(Z > 1.0586).$

$0.013709 + 0.14489$

$= 0.158599$

$\simeq \underline{\underline{0.1586}} \times 300 = 47.5797.$

$= 48$ bearings.

Q) 500 ball bearings have a mean weight of 142.30g & std.dev of 8.5g find the probability that the random sample of 100 ball bearings chosen from this group will have combined weight b/w 1) 140.61g and 141.75g.

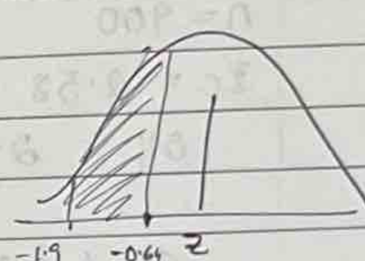
b) $> 144.60g$.

$n = 100$ $\quad \sigma_{\bar{x}} = \frac{8.5}{10} = 0.85$.

$\sigma = 8.5$

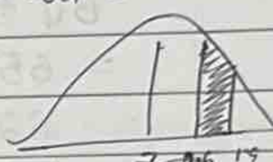
$\mu = 142.30$

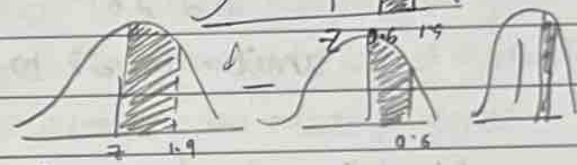
a) $P[-1.9882 < z < -0.64705]$.



$= P(1.9822) - P(0.64705)$ [use $\phi()$]

$= 0.47627 - 0.2412 = 0.23507$.





b) $P[\bar{x} > 144.60]$.

$P[z > 2.70588] = R(2.70588) = 0.0034062$.

Q) The mean & std.dev of the maximum loads supported by 60 cables are 11.09 tonnes and 0.73 tonnes. find

a) 95% confidence limit

b) 99% confidence limit.

for mean of max.load of all cables

$\mu = 11.09$ $\quad n = 60$.

$\sigma = 0.73$

Confidence limit $= \bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$

95% $- z_c = 1.96$.

$11.09 \pm \frac{1.96(0.73)}{\sqrt{60}}$.

11.09 ± 0.18471.

1) 11.2747

2) 10.905

limit = 10.905 – 11.2747

99% $z_c = 2.58$.

$11.09 \pm \frac{2.58(0.73)}{\sqrt{60}}$.

11.09 ± 0.24314.

11.33 and. 10.846.

Q) A sample 900 men are found to have a mean height of 64 inch if this sample has been drawn from a normal population with std-dev 20 in. find 99% confidence limits for mean height of men.

$\mu = 64 = \bar{x}$

$\sigma = 20$

$n = 900$

$z_c = 2.58$

$$64 \pm \frac{2.58(20)}{\sqrt{900}}$$

$$64 \pm 1.72$$

$$= 65.72$$

$$62.28$$

limit - 62.28 to 65.72.

Q) Sample of 5000 students have average height was 62.5 kg with standard deviation of 22kg. find the 95% confidence limit of the average weight of students.

$n = 5000$

$\bar{x} = 62.5$

$\sigma = 22$

$95\% \rightarrow z_c = 1.96$

$$62.5 \pm \frac{1.96(22)}{\sqrt{5000}}$$

$$62.5 \pm 0.6098 \longrightarrow 63.109 - 61.8902.$$

Q) Systolic BP of 566 males was taken. Mean BP was found to be 128.8mm and stdev of 13.05mm. 95% confidence. within which population mean would lie

$$128.8 \pm \frac{13.05(1.96)}{\sqrt{566}}$$

$$= 128.8 \pm 1.0751$$

$$129.87 - 127.72.$$

Q) Standard deviation of blood sugar level in population - 6 mg
If population mean is not known, within what limits is it likely to lie if random sample of 100 has a mean of 80mg %. - 95 %.

Page :
Date :

$\mu = \bar{x} = 80$

$\sigma = 6.$

$\Rightarrow \quad 80 \pm \frac{1.96(6)}{10}.$

$80 \pm 1.176.$

$81.176 - 78.824$

Q) To know the mean weights of all 10 year old boys in Delhi - a sample of 225 was taken. The mean weight of the sample was found to be 67 pounds with std. dev of 12 pounds. What can we infer about the mean weight of the population. [95 and 99%.]

$n = 225.$

$\bar{x} = 67.$

$\sigma = 12.$

$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}.$

$95\% = 67 \pm \frac{1.96\ (12)}{\sqrt{225}} = 67 \pm 1.568.$

$= 65.432 - 68.568$ pounds.

$99\% = 67 \pm \frac{2.58\ (12)}{\sqrt{225}.} = 67 \pm 2.064$

$64.936 - 69.064$ pounds.

Infer that population mean weight lies b/w ___ ___.

Q1 The mean and standard deviation of diameters of a sample of 250 rivet heads manufactured by a company are 7.2642mm and 0.0058 mm. Find.

a) 99% confidence limits

b) 95% confidence limits

$n = 250$

$\bar{x} = 7.2642$

$\sigma = 0.0058$.

$$99\% — 7.2642 - \frac{0.0058}{\sqrt{250}}(2.58).$$

$$7.2632 - 7.2651.$$

$$95\% - 7.2642 - \frac{0.0058}{\sqrt{250}}(1.96)$$

$$7.26348 - 7.2649.$$

Q) A sample of 80 student is surveyed.

$\bar{x} = 593.84\$$

$\sigma = 369.34\$$

95%.

$$\therefore \quad 593.84 \pm \frac{369.34}{\sqrt{80}}(1.96)$$

$$593.84 \pm 80.935.$$

$$512.904 - 674.775.$$

Q) 400 items are sampled from normally distributed population with

$\bar{x} = 22.1$

$\sigma = 12.8$.

95% confidence interval.

Q) $n = 100$ [students-marks]

$\bar{x} = 67.45$

$\sigma = 2.92$.

find a) 95% confidence interval b) 99% confidence intervals.

Statistical Decision

For reaching statistical decisions - we start with some assumptions or guesses about the populations involved. Such assumptions or guesses which may or may not be true - are called as statistical hypothesis.

HYPOTHESIS

a) NULL hypothesis.

a statistical hypothesis with no difference or with null attitude is called as null hypothesis. In other words - a null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true. It is denoted by H_0

Ex :

→ The average height of the competetors in a game is 160 cm.

→ Average life time of an electrical bulb manufactured by a company is 1800 hrs.

Ex 1) $H_0 : \mu = 160$ cm

2) $H_0 : \mu = 1800$ hrs.

b) Alternative hypothesis.

A statistical hypothesis which is complementary to the null hypothesis is called an alternative hypothesis which is denoted by H_1

It is clear that null hypothesis is meaningful when we formulate an alternative hypothesis.

Ex.

- If the null hypothesis "the average height of competetors in a game is 160 cm" ie $H_0 : \mu = 160$cm then alternative hypothesis will be

a) The average height of competetors in a game is not equal to 160c

$H_1 : \mu \neq 160$cm. - two tail hypothesis

b) $H_1 : \mu > 160$cm

c) $H_1 : \mu < 160$cm

one tail hypothesis.

Ex. Null hypothesis $H_0 : \mu = 1800H$.

alternate
- $H_1 : \mu \neq 1800$
- $H_1 : \mu > 1800$
- $H_1 : \mu < 1800$.

Two tailed Test

$H_0 : \mu = 160$.

$\Rightarrow \mu \geq 160$ or $\mu \leq 160$.

one tailed

(or)

$H_0 : \mu = 160$	$\mu \geq 160$	$\mu \leq 160$
$H_0 : \mu \neq 160$	$\mu > 160$	$\mu < 160$.
Two tailed	Left tailed	Right tailed.

Tests of Hypothesis — and significance.

Procedures which enable us to decide whether to accept or reject a hypothesis or to determine whether observed samples differ significantly from expected results — are called test of hypothesis, test of significance or Rules of Decision

Errors in test of hypothesis

a) Type I

In a hypothesis test — a type I error occurs when the null hypothesis is rejected when it is in fact true i.e. H_0 is wrongly rejected.

Ex.

In a clinical trial of a new drug — the null hypothesis might be that — the new drug is no better on average that the current drug

H_0 : There is no difference b/w the two drugs on average.

A type I error would occur if we conclude that the two drugs produced are of different effects.

The probability of a type I error can be precisely computed as

$$P(\text{type I error}) = \text{significance level} = \alpha.$$

A type I error can also be referred to as an error of the first kind.

b) Type II

In a hypothesis test - a type II error occurs when a null hypothesis H_0 is not rejected when it is in fact false.

Ex.

H_0: There is no difference b/w the two drugs on average.

↳ false notion.

but the false H_0 is accepted - then type II error.

$P(\text{type II error}) = \beta$.

Type II error can also be reffered as error of the second kind.

Type I - rejecting when H_0 is true.

Type II - not rejecting when H_0 is false.

$P(\text{Type I}) = \alpha$.

$P(\text{Type II}) = \beta$.

$P(\text{rejecting } H_0 \text{ when false}) = 1 - \beta$.

Decision	H_0	
	True	False.
Reject H_0.	Type I	✓ [no error]
Not rejecting / Accepting H_0	✓ [no error]	Type II error.

Level of significance. [confidence limit]

suppose that under a given hypothesis H, the sampling distribution of statistics (S) There is a normal distribution with mean (μ_S) and std. dev (σ_S) then $z = \dfrac{S - \mu_S}{\sigma_S}$ is the std. normal variate associated with S. So that for the distribution of S $\mu_S = 0$ $\sigma_S = 1$.

Accordingly for the distribution of z the $z\%$ confidence interval is $(-z_c, z_c)$ This means that we can be $z\%$ confident that if the hypothesis H is true then value of z will be b/w $-z_c$ to z_c.

If we reject the hypothesis (H) on the grounds that the value of H lies outside interval $(-z_c, z_c)$ - we would be making a type I error and probability of making error is $100 - z\%$.

The hypothesis (H) is rejected at a 100 − z% level of confidence. Thus a level of significance is the probability level below which we reject a hypothesis.

Page :
Date :

CIE − II

1) Module 5 − Sampling distribution [optimisation − not included].
2) Module 1 − vector [complete].

T series distribution :

T series followed for a sample size $n < 30$ [⇒ small samples].
degrees of freedom for $n = (n-1)$ degrees.
if t value < given value ⇒ acceptable hypothesis.
t value > given value ($t_{0.05}$ or $t_{0.01}$) ⇒ hypothesis − rejected

$$t = \left(\frac{\bar{x} - \mu}{s}\right)\sqrt{n-1} \quad \text{or} \quad t = \frac{\bar{x} - \mu}{s\sqrt{n}}. \qquad s^2 = \frac{1}{n-1}\Sigma(x - \bar{x})^2.$$

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

s = sd of sample.

Q) $n = 25$. $\bar{x} = 47.5$. $s = 8.4$.
hypothesis : mean of population = 42.1
given $t_{0.05} = 2.064$ for 9 df

$$t = \left(\frac{47.5 - 42.1}{8.4}\right)\sqrt{25-1} = 3.1493 \simeq 3.21.$$

as $3.21 > t_{0.05}$.
⇒ then hypothesis is rejected.

Q) $n = 10$.
70, 120, 110, 101, 88, 83, 45, 98, 107, 100.
H_0 : mean of population IQ of 100 at 5% level of significance.
$t_{0.05} = 2.262$ for 9 d.f.
sample mean = 97.2 ($\bar{x}$).
Sample sd = $\sqrt{\frac{\Sigma x^2 - (\Sigma x)^2}{(n-1)}} = 14.2735$.

$$t = \frac{97.2 - 100}{14.27\sqrt{10}} = 0.6203$$

$t < t_{0.05}$
↳ acceptable.

Camlin