Advance Statistics - 03

Topics:

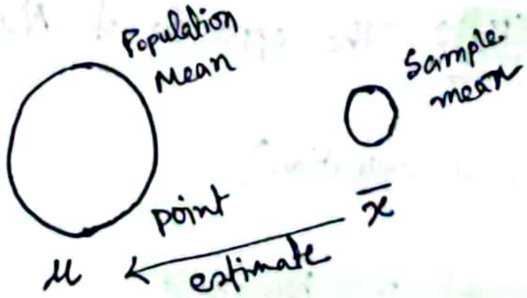
1. Estimate
2. Hypothesis testing and mechanism
3. P Value
4. Z test with examples
5. Student T Distribution
6. T stats and T test

Estimate: It is an observed numerical value used to estimate an unkown population parameter

Types → 1) **Point estimate:** Single numerical value used to estimate unknown population parameter.

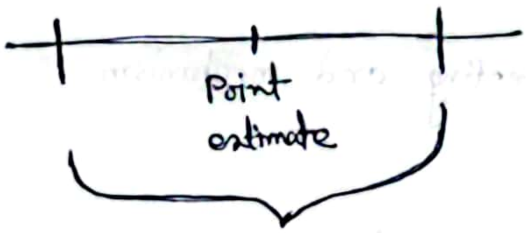
Example:

Sample mean is a point estimate of a population mean.



② **Interval estimate:** It is a range of values that is used to estimate the unkown population parameters

Interval estimates of population parameters are called confidence interval.



The range is **Confidence Interval.**

Hypothesis testing and Mechanism:

Using inferential statistics we can come to conclusions and inferences.
Using sample data from the population data, we derive a conclusion
To derive these conclusions, we use hypothesis testing.

Hypothesis testing mechanism:

① **Null Hypothesis (H_0):** The assumption you are begining with.

② **Alternate Hypothesis (H_1):** The opposite of Null hypothesis.

③ **Experiments** → (Proof collection)

④ **Accept or ~~reget reg~~ reject Null hypothesis**

Example A criminal is taken to the court

Null hypothesis: He is innocent until found guilty (by default)

Alternate hypothesis: He is a criminal

Experiments: Proof collections (DNA matching, fingerprint, etc.)

Accept of reject: According to the proof Null hypothesis will be accepted or ~~is~~ rejected.

P Value:

The P value is a number, calculated from a statistical test, that describes how likely you are to have a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.

Example: A coin is fair or not fair.

$Pr(Head) = 0.5$, $Pr(tail) = 0.5$ (Fair)

$Pr(Head) = 0.6$, $Pr(tail) = 0.4$ (Fair)

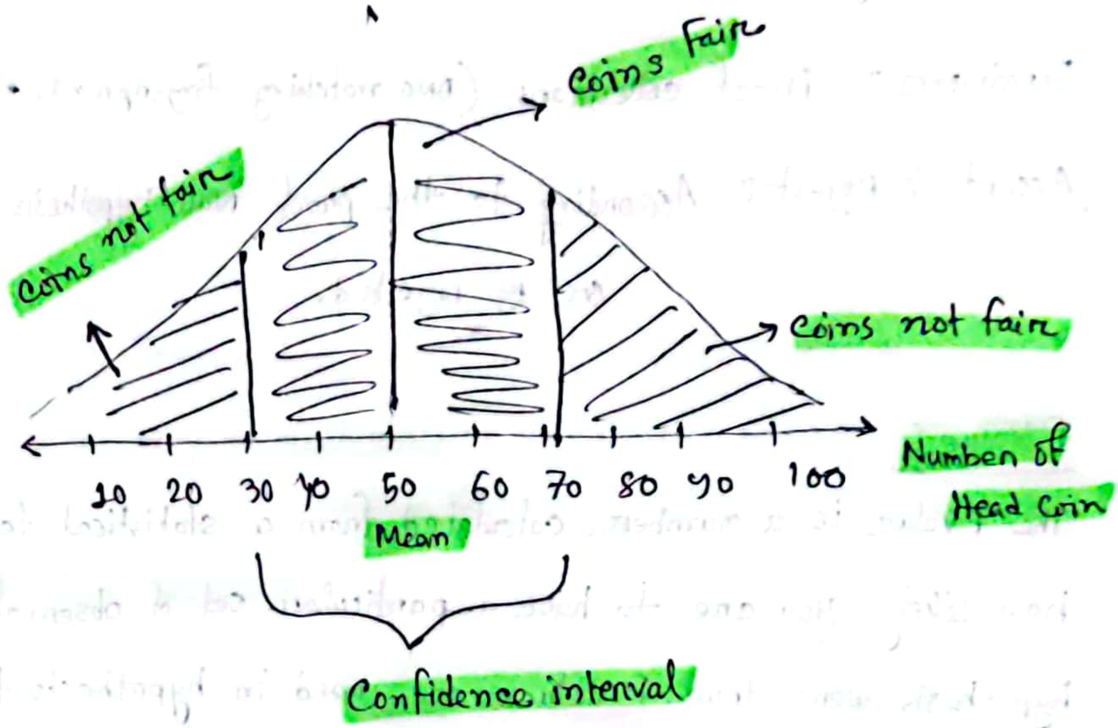
$Pr(Head) = 0.7$, $Pr(tail) = 0.3$ (not fair)

P.T.O

Null hypothesis (H_0): The coin is fair

Alternate hypothesis (H_1): The coin is not fair.

Experiment:



Confidence Interval Region = Accepted Region (30 – 70)

Rejected Region (00 - 30) and (71 – 100)

if p value found in Accepted Region, then we accept the Null Hypothesis, else we reject the Null Hypothesis if the p value found in rejected region.

Z test with Examples:

Use Z test when population std and population size(n) is given

Question: Context

The average heights of all residents in a city is 168 cm with a $\sigma = 3.9$

A doctor believes the mean to be different. He measured the height of 36 individuals and found the average height to be 169.5 cm

Questions → a) State null and alternate hypothesis

b) At a 95% confidence level, is there enough evidence to reject the null hypothesis?

a) $\mu = 168$ cm, $\sigma = 3.9$, $n = 36$, $\bar{x} = 169.5$ cm

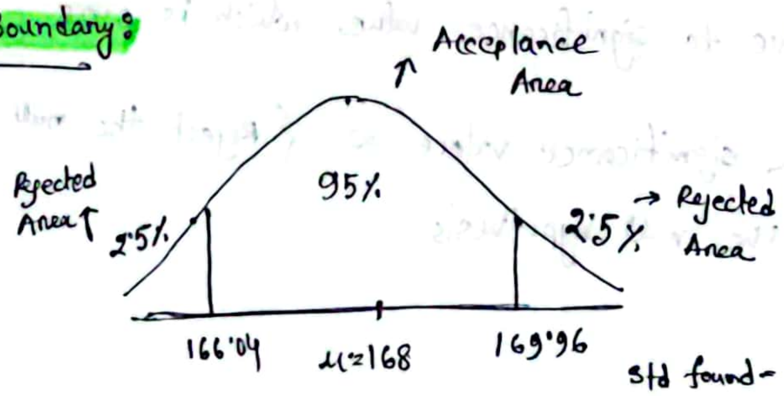
Null Hypothesis (H_0) → Average height, $\mu = 168$ cm

Alternate Hypothesis (H_1) → $\mu \neq 168$ cm {2 tail test}

↳ Ans can be greater or smaller

b) C.I = 0.95 significance value $\alpha = 1 - 0.95 = 0.05$

c) **Decision Boundary:**



std found - 1.96 from Zscore And z table

Z-test Technique:

Statistical Analysis:

$$Z.test = \frac{\bar{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

For population data,

$$Ztest = \frac{\bar{x}_i - \mu}{\sigma}$$

For sample data,

$$Ztest = \frac{\bar{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{169.5 - 168}{\frac{3.9}{\sqrt{36}}}$$

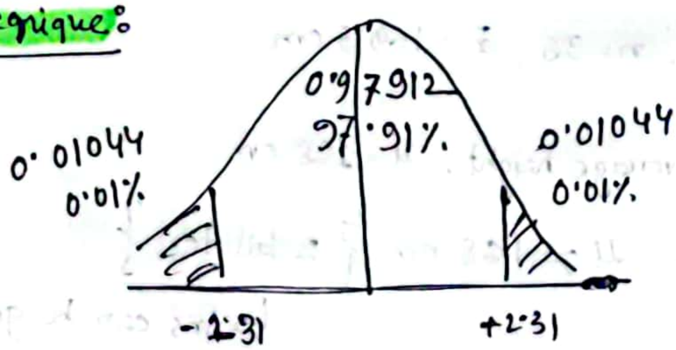
$$= 2.31 \text{ std}$$

(this value of std we found by Z-score)

if Z-test value < -1.96 or > 1.96, we have to ~~rg~~ reject the null hypothesis or else we accept the null hypothesis.

As, Z test $= 2.31 > 1.96$, So, we reject the null hypothesis.

P-value tecqnique:



$$p \text{ value} = 0.01044 + 0.01044 = 0.02088$$

Compare p value to significance value which is 0.05

if p value $<$ significance value α { Reject the null hypothesis

else accept the null hypothesis.

Question 2: Context:

A Factory manufacture bulbs with a average warranty of 5 years with std of 0·50. A worker believes that the bulb will malfunction in less than 5 years. He tests a sample of 40 ~~bulb~~ bulbs and find the average time to be 4·8 years.

Question → a) State null and alternate hypothesis

b) At a 2% significance level, is the enough evidence to support the idea that the warranty should be revised?

a) $\mu = 5$, $\sigma = 0{\cdot}50$, $n = 40$, $\bar{x} = 4{\cdot}8$

Null hypothesis: Average warranty $\mu = 5$ years

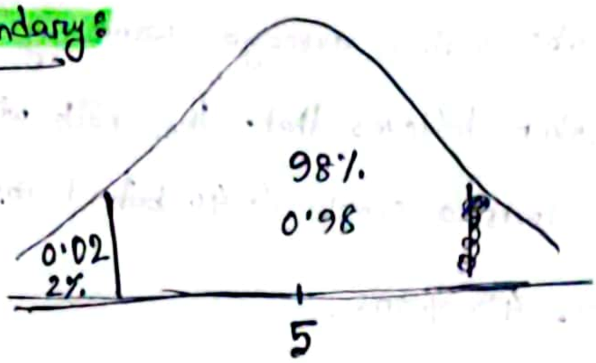
Alternate hypothesis: Average warranty ~~≠ 5 years~~ ~~{2 tail test}~~
$\mu < 5$ years

b) We know, confidence interval = 1 − significance interval

$\Rightarrow$ C.I = 1 − 0·02
= 0·98
= 98% confidence interval.

~~Desicion~~

Decision Boundary:

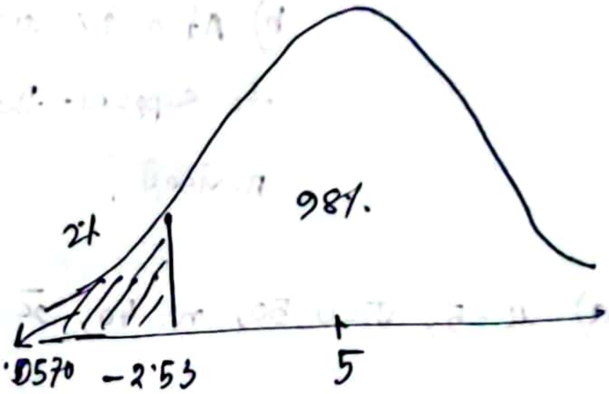


P Value:

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

$$= \frac{4.8 - 5}{\frac{0.5}{\sqrt{40}}}$$

$$= -2.53 \text{ std}$$



↳ Go to Z table and find value for -2.53

(-2.5 in left, 0.03 in right)

value found → 0.0570 which is the area under the curve

P-Value = 0.0570

As P-Value >= 0.02 (significance value)

We can't reject the null hypothesis

We accept the null hypothesis.

~~So, Warranty should not be revised~~

(I have contradictory answer with the lecture)

(confused)

Student T Distribution:

For statistical analysis using Z score we need population std

But if we don't have population std, in that scenario we use Student T Distribution.

Instead of taking population std, we take sample std

From → $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, we use $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ s = sample std

For t distribution,

Degree of freedom:

$dof = n - 1$ [n: sample size]

[We also have t table like the Z table]

Problem on T-stats:

In the population the average IQ is 100, A team of researchers want to test a new medication to see if it has either a positive or negative effect on intelligence, or no effect at all. A sample of 30 participants who have taken the medications has a mean of 140 with a standard deviation of 20. Did the medication affect intelligence? C.I = 95%

Ans) $\mu = 100$, $n = 30$, $\bar{x} = 140$, $S = 20$, $C.I = 0.95$, $\alpha = 0.05$

a) **Null hypothesis:** $\mu = 100$

Alternate hypothesis: $\mu \neq 100$ {2 ~~test~~ tail test}

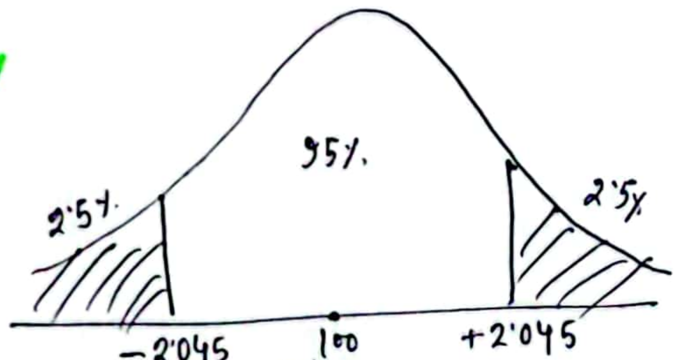
↳ ans canbe greater or lesser

b) $\alpha = 1 - 95\% = 0.05$

c) **Degree of freedom:**

$dof = 30 - 1 = 29$

d) **Decision Boundary:**



Go to t table → Find value in 2 tail ~~row~~ for 0.05 (α) column

Find value in 29 (dof)

2 tail	0.05
29	2.045

if t-test < -2.045 and > ~~or~~ 2.045, we reject the the null hypothesis.

e) Calculate t-test stat:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{140 - 100}{20/\sqrt{30}} = 10.96$$

f) conclusion:

$t = 10.96 > 2.045$, So we are rejecting the null hypothesis.

As, t test value comes positive, the medication positively affect intelligence (increases the intelligence)