Inferential Statistics.

Agenda

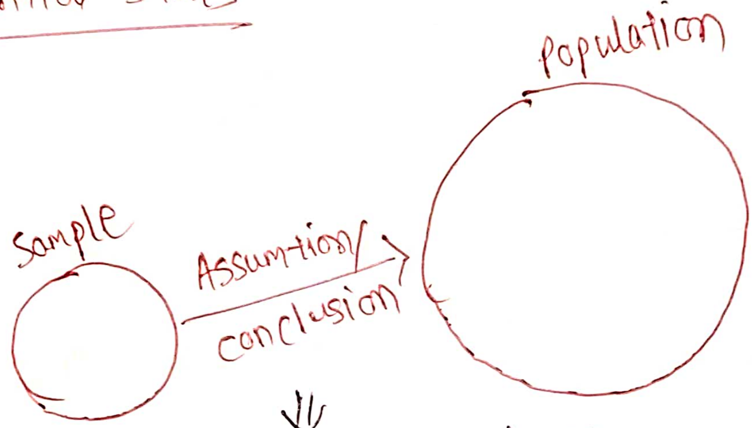
① Hypothesis Testing

② P-Value.

③ Confidence interval

④ significance value.

Inferential stats



⇓

Hypothesis testing.

Steps of Hypothesis testing

① Null hypothesis

Person is not a criminal (default)

coin is fair (default)

② Alternate Hypothesis:

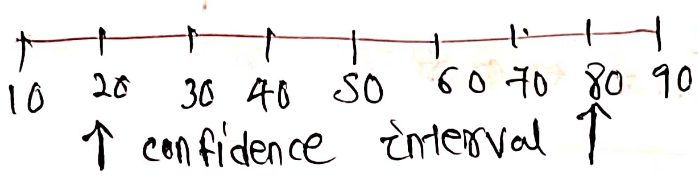
→ Person is a criminal

→ coin is not fair.

→ Perform experiments (tossing a coin 100 times)

→ Why Hypothesis testing ?

In Inferential stats, we make Assumption / conclusion about the population. To validate this we need hyphothesis testing.



50 times head, 80 times, 70 times, 80 times.

↓

Domain expert → confidence interval (20, 80)

10 times (Null hypohtesis rejected)

↳ We fail to reject the null hypothesis.

↳ we reject the null hypothesis.

② Person is criminal or not

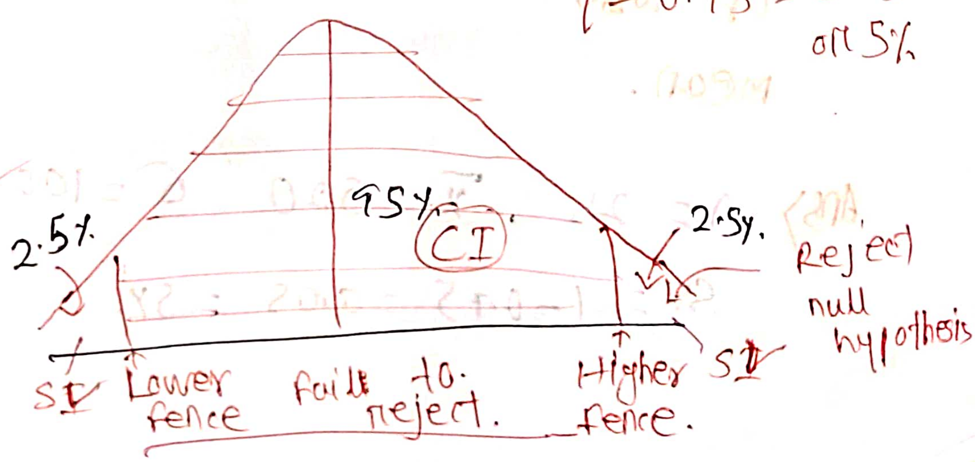
proof : DNA, finger print, eye witness, footage.

Confidence Interval (CI)

significance value

CI = 95%

1 − 0.95 = 0.05 or 5%



Point estimate: The value of any statistics that estimate the value of parameter is called point estimate.

$$\bar{x} \xrightarrow{\text{make conclusion}} \mu$$

Stats parameter.

$\bar{x} \geq \mu$

$\bar{x} \leq \mu$

Point estimate ($\bar{x}$) $\rightarrow$ Population mean (μ) parameter

$\pm$ margin of error

Lower fence = Point estimate − margin of error

Higher fence = Point estimate + margin of error

$$\text{margin of error} = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

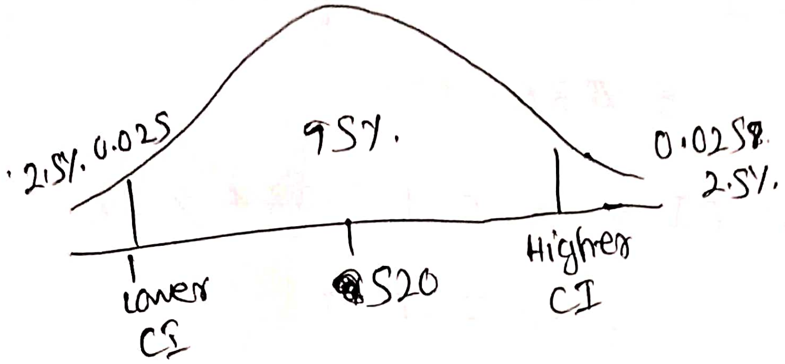
$\boxed{\frac{\sigma}{\sqrt{n}}}$ → standard error.

α = significance value. $\alpha = 1 - CI$

Q. On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a ~~sample~~ population standard deviation of 100. Construct a 95 CI about the mean.

Ans) $n = 25$, $\bar{x} = 520$, $\sigma = 100$, $CI = 95\%$

$SV = 1 - 0.95 = 0.05 = 5\%$



Lower CI = Point estimate − margin of error

$= 520 - Z_{0.05/2} \frac{\sigma}{\sqrt{n}}$

$= 520 - Z_{0.05/2} \frac{100}{5}$

$= 520 - 1.96 \times 20$

$= 480.8$

Higher CI $= 520 + 1.96 \times 20 = 559.2$

Q On the Quant test of cat exam a sample of 25 test takers has a mean of 520 with a sample standard deviation of 80 construct 95% CI about the mean?

Ans. $\bar{x} = 520$ $n = 25$ CI = 95% $S = 80$

S.V $= 1 - 0.95 = 0.05$

→ As sample standard deviation given T test performed

Lower CI $= 520 - t_{0.05/2} \frac{80}{5}$

$\left[\because \bar{x} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}\right]$

Degree of freedom = $n-1 = 25-1 = 24$

Lower CI = $520 - 2.064 \times 16$

$= 486.976$

Higher CI = $520 + 2.064 \times 16$

$= 553.024$

① 1 tail and 2 tail test

Q. College in town A has a placement rate of 85%. A new college recently opened and it was found that a sample of 150 student had a placement rate of 88% with the standard deviation of 4%. Does this college has a different placement rate with 95% CI?

$\neq$ two tail test

if $<$ or $>$ one tail test.

Q. A factory has a machine that fills 80ml of baby medichine in a bottle. An employee belives that the avg amount of baby medichine is not 80ml. Using 40 sample, the avg amount of medichene is displayed by the machine is 78 ml with a standard deviation of 2.5 ml

a) State Null and alternate hypothesis.
b) At 95%. CI is there enough evidence machine is working properly or not.

Ans
a) Null hypothesis

$\mu = 80$ ml.

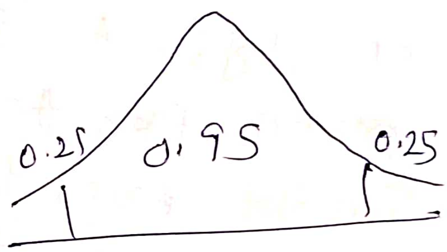
Alternate hypothesis

$\mu \neq 80$ml.

Given
CI = 95, n = 40, $\bar{x} = 78$ S = 25

step-1

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



If $n \geq 30$ or Population s.d (σ) → Z test

if $n < 30$ and Sample s.d (S) → T test

step-2 Z test

$Z_{0.25} = 1.96$

~~Lower CI =~~

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{78 - 80}{2.5/\sqrt{40}} = -5.05$$

Conclusion

Decision rule : If $Z = -5.05$ is less then -1.96 or greater then 1.96 ; reject the null hypothesis with 95% CI.

Reject Null hypothesis { There is fault in the machine.

Q A complen was registered, the boy in a Government schol are underfed, Avg. weight of the boys of age 10 is 32 kgs. with SD 9kgs. A sample of 25 boys were selected from the government school and avg. weight found to be 29.5kgs with CI 95%. check it is true or false?

Ans.

1. we know population SD (given 9kg)

2. Z-test step-1

Null hypothesis

$H_0 : \mu = 32$

Alternate

$H_1 : \mu \neq 32$

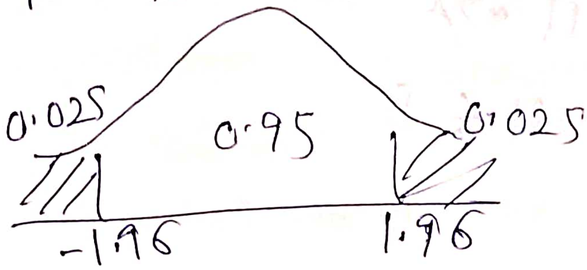
two tail test

step-2

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

step-3

Z test



$$Z \text{ score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{29.5 - 32}{9/\sqrt{25}} = -1.39$$

conclusion

$-1.39 > -1.96$

→ fail to reject the null hypothesis.

The boys are fed well.

Q A factory manufactures car with a warranty of 5 or more year on the engine and transmission. An engineer belives that the engine or transmission will malfunction in less then 5 years. ~~The~~ He tests a sample of 40 cars and find the average time to be 4.8 years with standard deviation of 0.5 years

① State Null and alternate hypothesis.

② At a 2% significance value, is there enough evidence to support the idea that the warenty should be revised?..

~~step-1~~ Null hypothesis

$H_0 : \mu \geq 5$

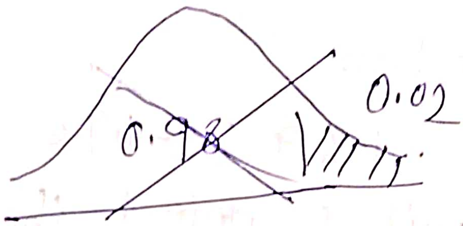
$H_1 : \mu \leq 5$

step-2

$\alpha = 0.02 \quad CI = 0.98$

step-3

$n = 40 \quad S = 0.5$



$n > 30$ ~~and s given~~

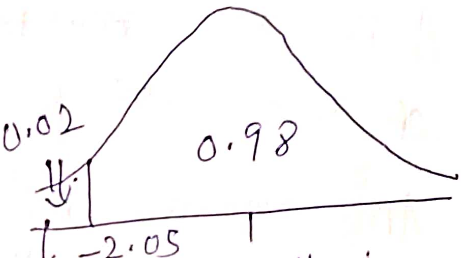
Z test

Left tail
$\because \mu = 5$ and $H_1 \; \mu < 5$

$$Z_{score} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

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

$$= -2.52$$



Reject Null hypothesis.

conclusion

$-2.52 < -2.05$

Reject Null hypothesis.

Warenty needs to be revised.

Q. The average weight of all recident in a town XYZ is 168 pounds. A nutritionist belives the true mean to be different. she measuried the weight of 36 individuals and faund the mean to be 169.5 pounds with the standard deviation of 3.9

a) Null & alternate hypothesis

b) 95% CI. Is there enough evidence to discard the null hypothesis.

Ans

step 1

$H_0 : \mu = 168$

$H_1 : \mu \neq 168$

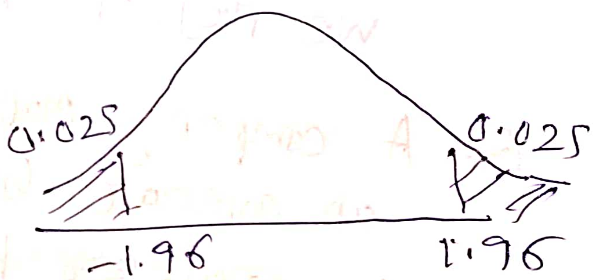
step-2

$CI = 95\%, \alpha = 0.05$

step 3

$n = 36 \quad s = 3.9$

z test



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

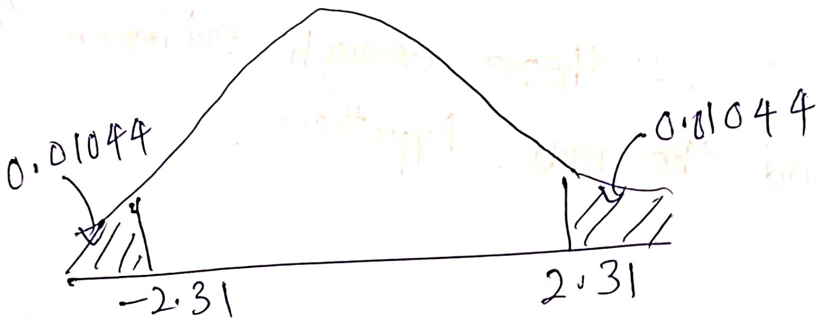
$$= \frac{169.5 - 168}{3.9/\sqrt{36}}$$

$$= \frac{1.5 \times \cancel{6}^{2}}{\cancel{3.9}_{1.3}} = \frac{3}{1.3} = 2.31$$

Conclusion

As $2.31 > 1.96$
we reject the null hypothesis.

By using P value



$$P \text{ value} = 0.01044 + 0.01044$$
$= 0.02088$ [total area under curve of significance region]

As $P = 0.02088 < 0.05$

we reject Null hypothesis.

Q.. A company manufactures bikes batteries with an average life span of 2 or more years. An engineer belives this value to be less using 10 sample, he measures the avg. life span to be 1.8 years with a standard deviation of 0.15

a) State the Null and Alternate hypothesis

b) At a 99% CI, is there enough evidence to discard the H_0.

Step-1

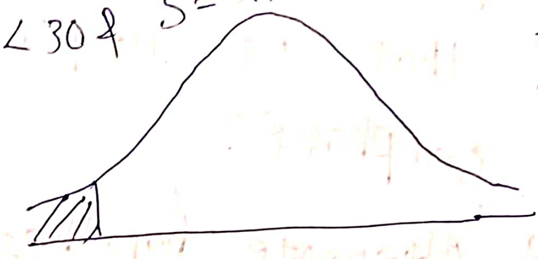
$H_0 : \mu \geq 2$

$H_1 : \mu < 2$

Step 2

$CI = 99\%$ $\alpha = 0.01$ Degree of freedom $= n-1$ $= 10-1 = 9$

As $n = 10 < 30$ & $s = 0.15$



-2.821

$$T \text{ score} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{1.8 - 2}{0.15/\sqrt{10}}$$

$$= -4.21$$

As $-4.21 < -2.821$

Conclusion

we reject the Null hypothesis.

The average life of the battery is less then 2 year.

Z test with proportion

Q. A tech company belives that the percentage of recidents in town XYZ that owns a cellphone is 70%. A marketing manager belives that this value to be different. He conduct of survey of 200 individuals and found that 130 responding yes owning a cell phone?

a) Null and Alternate hypothesis (H_0 & H_1)

b) At a 95%, is there enough evidence to reject the H_0.

Ans

Step 1

Null hypothesis (H_0) $P_0 = 0.70$ (Given proportion)

Alternate Hypothesis $P_0 \neq 7.0$

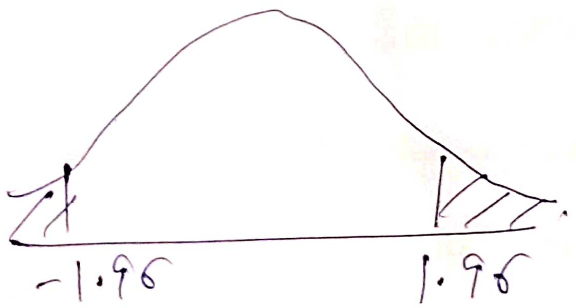
$q_0 = 1 - P_0 = 0.30$ (Remaining proportion)

$n = 200$

$\hat{p} = \frac{130}{200} = 0.65$

Step-2

$CI = 0.95 \quad \alpha = 0.05$



As P_0 & q_0 given

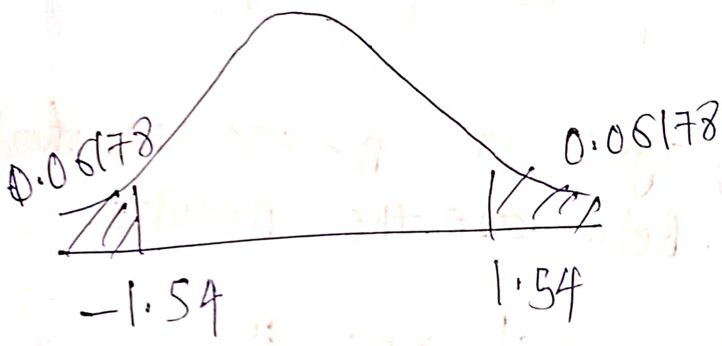
$$Z_{test} = \frac{\bar{p}' - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.65 - 0.7}{\sqrt{\frac{0.7 \times 0.3}{200}}}$$

$$= -1.54$$

conclusion

As $-1.54 > -1.96$
so we fail to reject Null hypothesis.

Using p value



$P = 0.06178 + 0.06178 = 0.12356$

As p value $= 0.12356 > 0.05$.
we fail to reject Null hypothesis.

Chi square test

- Chi square test cliams about population proportions.
- It is a non parametric test that is performed on categorical data.

Q. In the 2000 US sensu; the age of individuals in a small town found to be the following.

<18	$18-35$	>35
20%	30%	50%

In 2010, ages of $n=500$ individuals were sampled. Below are the results.

<18	$18-35$	>35
121	288	91

using $\alpha = 0.05$, would you conclude that the population distribution of ages has changed in the last 10 years?

Ans

<18	18 - 35	>35
20%	30%	50%

In 500

	<18	18 - 35	>35
observed	121	288	91
expected	100	150	250

Step-1 :

Null hypothesis

H_0 : Data meets expected distribution

H_1 : Data doesn't meet expected distribution

Step-2 : $\alpha = 0.05$ $CI = 0.95$

step-3

Degree of freedom = categories - 1

$df = c - 1 = 3 - 1 = 2$

step 4 Decision boundry

χ^2 5.991

step-5

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-250)^2}{250}$$

$$= 232.494$$

Conclusion

$\chi^2 > 5.99$ Reject ~~not~~ H_0

ANOVA test

This statistical test used to determine if there is a statistically significant difference between two or more categorical groups by testing for difference of means using a variance.

example

~~f1~~ ~~output~~ Scores

Teaching Method A : 75, 80, 85, 70, 90.

Teaching Method B : 65, 70, 75, 80, 85

Teaching Method C : 95, 100, 105, 110, 115