Infuential Statistics

① Hypothesis Testing
② p - value
③ Confidence Interval
④ Significance Value

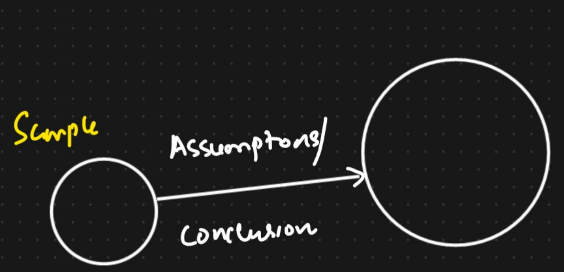
z test
t test
Chi square test
Annova test (F-test)

3 Distributions

① Bernoulli's
② Binomial
③ Power Law

TRANSFORMATION

Infuential Stat



⇓

Hypothesis Testing

Experiment

→ [Coin is fair or not]

$P(H) = 0.5$ $P(T) = 0.5$

Steps of hypothesis Testing

① Null Hypothesis : Coin is fair ⇒ Accepted

② Alternate hypothesis : Coin is not fair

③ Perform Experiments

10 times

75

100 times

⇓

| 50 times Head | ⇒ Fair

| 60 times Head |

60 40

70 30

80 20

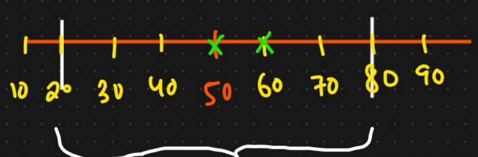
10 → Null hypothesis is Rejected

→ Alternate hypothesis is Accepted

CI = [20-80]

⇓

Coin is fair



C.I ⇒ Confidence Interval

70 times ⇒ Domain Expert

⇓

Confidence Interval

↳ We fail to Reject the Null Hypothesis [within C.I]

↳ We Reject the Null Hypothesis [outside C.I]

⇒ Conclusions

② Person is Criminal or not {Murder Case}

① Null Hypothesis : Person is not Criminal

② Alternate Hypothesis : Person is criminal

③ Experiment / Proof : DNA, finger print, weapons, eye witness, footage

⇓

Judge ⇒

Conclusions

Vaccines ⇒ Medical ⇒ critical

↗ Domain Expert

Confidence Interval : (C.I)

$\boxed{95\%}$

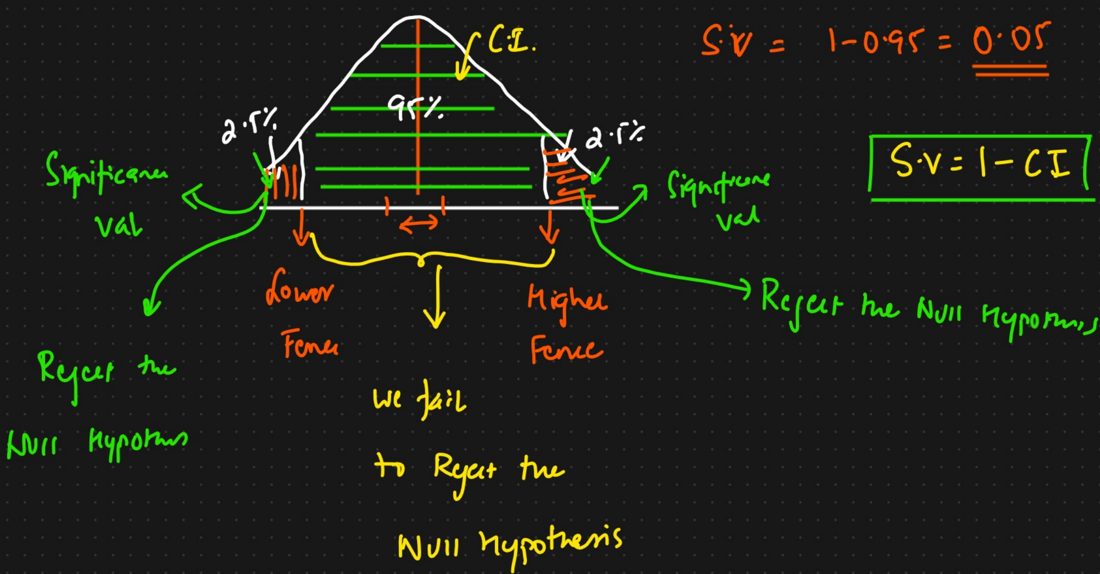
Significance Value

S.V = 1 − C.I

C.I = 95%

S.V = 1 − 0.95 = 0.05

$\boxed{S.V = 1 - CI}$



Point Estimate : The Value of any statistic that estimates the Value of a parameter is called Point Estimate

Point Estimate

⇑

$\boxed{\bar{x}} \longrightarrow \mu$

$\bar{x} \geq \mu$

$\bar{x} \leq \mu$

Statistic ↓ $\boxed{\bar{x}}$ ↓ Sample mean

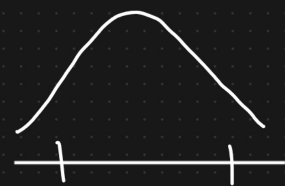
$\boxed{\bar{x}} \longrightarrow \boxed{\mu}$ ← population mean

parameter

$\boxed{\text{Point Estimate}} \pm \boxed{\text{Margin of Error}} = \boxed{\text{Parameter} \Rightarrow \text{population mean}}$

<u>Lower C.I</u> : Point Estimate − Margin of Error

<u>Higher C.I</u> : Point Estimate + Margin of Error



$$\text{Margin of Error} = Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}} \Rightarrow \text{Standard Error}$$

σ → population sd

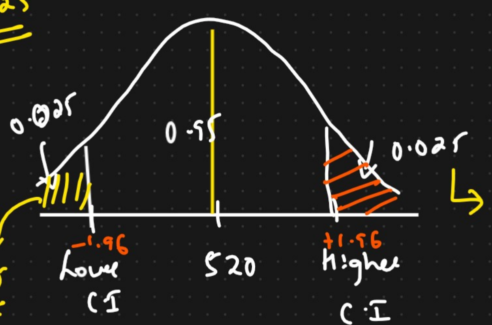
α = Significance Value.

Q) On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a population standard deviation of 100. Construct a 95% C.I about the mean?

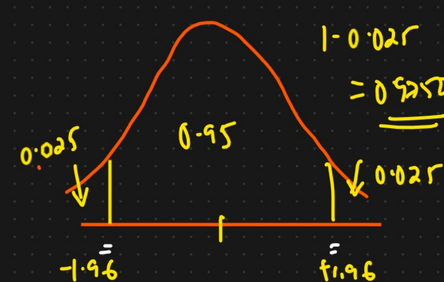
Ans) $n = 25$ $\quad \bar{x} = 520$ $\quad \sigma = 100$ $\quad C.I = 95\%$ $\quad S.V = 1 - C.I = 0.05$

$\frac{5}{2} = 2.5 \Rightarrow 0.025$

$\alpha = 0.05$



$\boxed{1 - 0.025 = 0.975}$



C.I = 0.95

S.V = 1 − 0.95 = 0.05

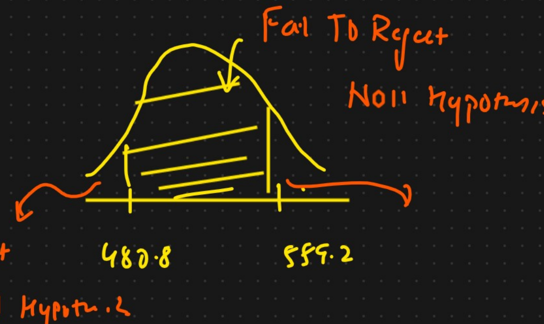
$Z_{0.05/2} \Rightarrow \boxed{Z_{0.025}}$

Lower C.I = Point Estimate − Margin of Error

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

$$= 520 - Z_{0.025} \frac{\cancel{100}\ 20}{\cancel{5}\ 1}$$

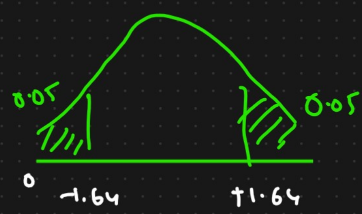
$= 520 - 1.96 \times 20 = \underline{\underline{480.8}}$



Higher C.I $= 520 + 1.96 \times 20 = \underline{\underline{559.2}}$

① $\bar{x} = 480$ $\quad \sigma = 85$ $\quad n = 25$ $\quad$ C.I = 90% $\quad$ Significance

$= 1 - 0.90 = \boxed{0.10}$



Lower C.I $= 480 - Z_{0.10/2}\left[\dfrac{85}{5}\right]$

$= 480 - Z_{0.05}\left[\dfrac{85}{5}\right]$

$= 480 - 1.64\,[17]$

$= 480 - 27.8 = \underline{\underline{452.12}}$

Higher C.I $= 480 + 1.64\,[17]$

$= 480 + 27.8$

$= \underline{\underline{507.8}}$

$[452.12 \rightleftarrows 507.8]$.

② 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% C.I about the mean?

Ans) $\bar{x} = 520$ $\quad S = 80$ $\quad$ C.I = 95% $\quad$ S.V = $1 - 0.95 = \underline{\underline{0.05}}$

$n = 25$

$\bar{x} \pm t_{\alpha/2}\left(\dfrac{S}{\sqrt{n}}\right)$.

t test

Degree of freedom $= \boxed{n-1} = 25 - 1 = \boxed{24}$

$\Rightarrow 0.025$

$\underline{\underline{0.05}}$

Lower $C.I = 520 - t_{0.05/2}\left(\frac{\overset{16}{80}}{\sqrt{81}}\right)$. $\quad \underline{\underline{t_{0.025}}}$

$= 520 - 2.064 * 16$

Lower C.I $= 486.976$

Higher C.I $= \underline{\underline{553.024}}$

① 1 Tail and 2 Tail Test

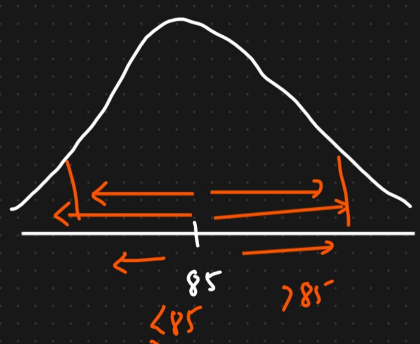
Q) Colleges in Town A has 85% placement rate. A new college was recently opened and it was found that a sample of 150 students had a placement rate of $\boxed{88\%}$ with a standard deviation of $\boxed{4\%}$. Does this college has a different placement rate with 95% C.I?

↓

greater than 85%

lesser than 85%

Two Tail.



1 tail Right side

① Z-test
② t-test

Hypothesis Testing

④ A factory has a machine that fills 80ml of Baby medicines in a bottle. An employee believes the average amount of baby medicine is not 80ml. Using 40 samples, he measures the average amount dispersed by the machine to be 78ml with a standard deviation of 2.5.

(a) State Null & Alternate Hypothesis

(b) At 95% C.I, is there enough evidence to support Machine is working properly or not

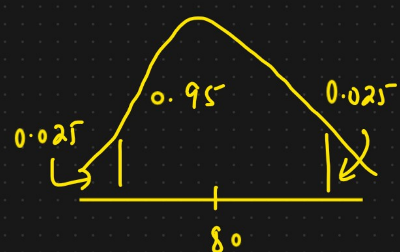
$\boxed{\mu = 80ml \quad n = 40 \quad \bar{x} = 78 \quad s = 2.5}$

Step 1

Ans) Null Hypothesis $\mu = 80$ →

Alternate Hypothesis $\mu \neq 80$ →

Step 2 : $C.I = 0.95$ $\quad S.V(\alpha) = 1 - 0.95 = \boxed{0.05}$



Step 3 :

① $\boxed{n \geq 30 \text{ or population sd}}$ } → Z test

② $\boxed{n < 30 \text{ and sample std}}$ } → t test

$n = 40$

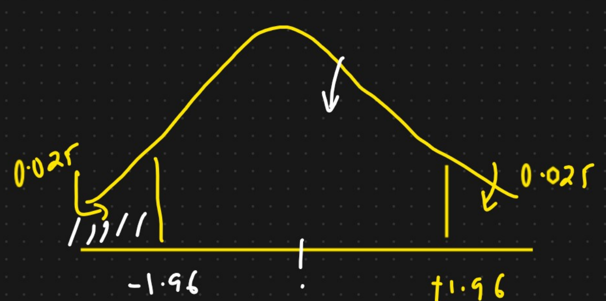
$s = 2.5$

Z test

Let perform the Experiment

$1 - 0.025 = 0.975$

Decision Boundary



* Calculate test statistics (z-test)

$$z = \frac{\bar{x} - \mu}{\boxed{s/\sqrt{n}}} \rightarrow \text{Standard Error} \quad = \frac{78 - 80}{2.5/\sqrt{40}} = -5.05$$

σ

* Conclusions

Decision Rule: If $z = -5.05$ is less than -1.96 or greater $+1.96$, Reject the Null Hypothesis with 95% C.I

Reject the Null Hypothesis { There is some fault in the machine.

* A complain was registered, the boys in a Government School are underfed. Average weight of the boys of age 10 is $\boxed{32}$ kgs with $\boxed{S.D = 9kgs.}$ A sample of 25 boys were selected from the Government School and the average weight was found to be 29.5 kgs? With CI: 95%. Check it is True or False.

Ans) Conditions for z-test $n=25$ $\mu=32$ $\sigma=9$ $\bar{x}=29.5$

① We know the population sd. $\boxed{OR}$

② We do not know the population sd but our sample is large $n \geq 30$

Conditions For T test

① We donot know the population std

② Our sample size is small $n < 30$

③ Sample std is given.

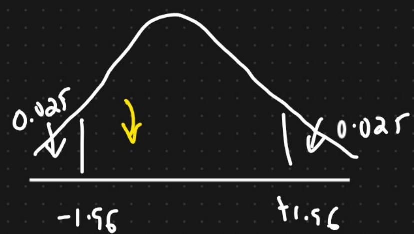
Step 1

① $H_0 : \mu = 32$

$H_1 : \mu \neq 32$

② $C.I = 0.95$ $\qquad \alpha = 1 - 0.95 = 0.05$

③ Z test



s

σ

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

Conclusion $= -1.39 > -1.96$ Accept the Null Hypothesis 95% C.I

We fail to Reject the Null Hypothesis

The Boys are fed well.

① A factory manufactures cars with a warranty of 5 years or more on the engine and transmission. An engineer believes that the engine or transmission will malfunction in less than 5 years. He tests a sample of 40 cars and finds the average time to be 4.8 years with a standard deviation of 0.50. ① State the null & alternate hypothesis

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

Z-score = -2.5298

Step 1 : H_0 : $\mu \geq 5$

H_1 : $\mu < 5$

Step 2 : $\alpha = 0.02$ C.I = 0.98

② In the population the average IQ is 100 with a standard deviation of 15. A team of scientists wants to test a new medication to see if it has a +ve or -ve effect, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect intelligence? { 95% }

Z-score = 14.96

④ A factory manufactures cars with a warranty of 5 years or more on the engine and transmission. An engineer believes that the engine or transmission will malfunction in less than 5 years. He tests a sample of 40 cars and finds the average time to be 4.8 years with a standard deviation of 0.50. ① State the null & alternate hypothesis

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

Ans) $n = 40$ $\bar{x} = 4.8$ years $s = 0.50$

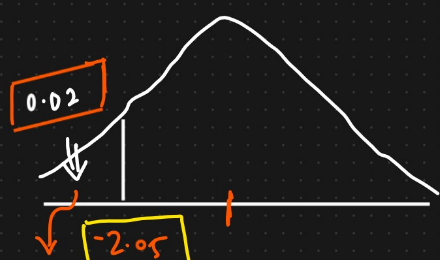
Step 1

$H_0: \mu \geq 5$ {Null Hypothesis}

$H_1: \mu < 5$ {Alternate Hypothesis}

Step 2: $\alpha = 0.02$ $C.I = 1 - 0.02 = 0.98 = 98\%$

Step 3:



Reject the Null Hypothesis

Step 4:

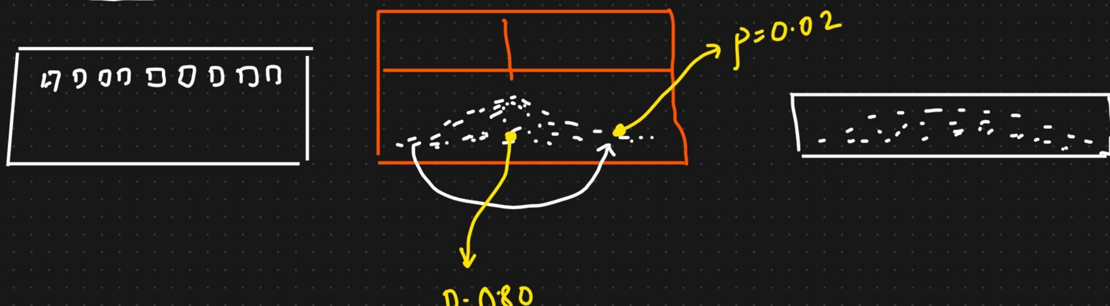
$$Z\text{-score} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{4.8 - 5}{0.50/\sqrt{40}} = -2.5298$$

Conclusion: $-2.52 < -2.05$ Reject the Null Hypothesis

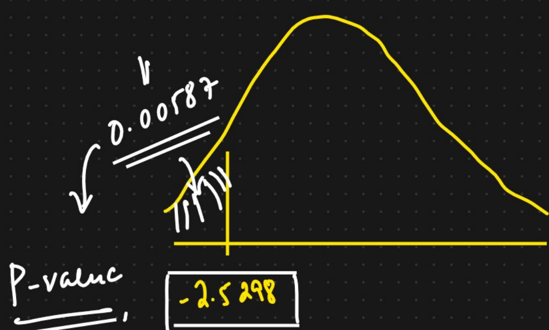
Warranty needs to be revised.

P_value ≠ {Significance value}. → C.I

Out of all 100 random touches



$\alpha = 0.02$



P-value < α [Yes]

Reject the Null Hypothesis

1) The average weight of all residents in a town XYZ is 168 pounds. A nutritionist believes the true mean to be different. She measured the weight of 36 individuals and found the mean to be 169.5 pounds with a standard deviation of 3.9

(a) Null & Alternate Hypothesis}

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

Ans) $\bar{x} = 169.5$ $S = 3.9$ $n = 36$ $\mu = 168$ $C.I = 0.95$

Step 1

H_0 $\mu = 168$

H_1 $\mu \neq 168$.

Step 2 : $C.I = 0.95$ $\alpha = 0.05$

Step 3

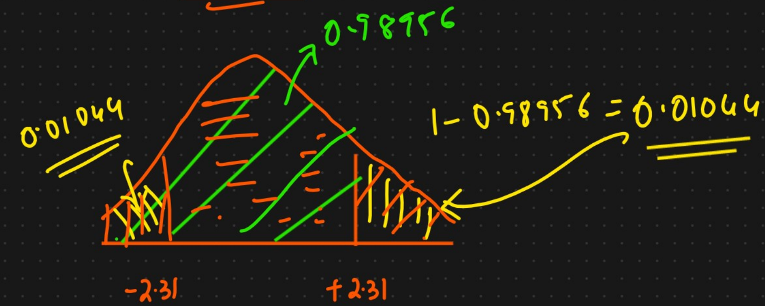
Step 4 : $z\text{-score} = \dfrac{169.5 - 168}{3.9/\sqrt{36}} \approx \boxed{2.31} \Rightarrow 2.3076$

$2.31 > 1.96$ { Reject the Null Hypothesis }

2 tail test

A-value

2 Tail Test



p.value = 0.01044 + 0.01044 = 0.02088

$0.02088 < 0.05$ ($\downarrow \alpha$) { Reject the Null Hypothesis }.

(4) A company manufactures bikes batteries with an average life span of 2 year or more years. An Engineer believes this value to be less. Using 10 Samples, he measures the average 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% C.I, is there enough evidence to discard the H_0?

Ans) (1) $H_0 : \mu \geq 2$

$H_1 : \mu < 2$.

(2) Step 2

$\boxed{n < 30}$ Sample standard deviation

C.I = 0.99 $\alpha = 0.01$

(3) Step 3

Degree of freedom $= n-1$

$= 10-1 = 9$

④ Calculate t test statistics

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.8 - 2}{0.15/\sqrt{10}} = -4.216$$

⑤ $-4.216 < -2.82$ { Reject the Null Hypothesis }.

The average life of the battery is less than 2 years.

④ Z test with proportions

④ A tech company believes that the percentage of residents in town XYZ that owns a cell phone is 70%. A marketing manager believes that this value to be different. He conducts a survey of 200 individuals and found that 130 responded Yes. owning a cell phone?

① State Null And Alternate Hypothesis?

② At a 95% C.I, is there enough evidence to reject the Null Hypothesis?

Ans) Step 1

Null Hypothesis : $P_0 = 0.70$ ✓

Alternate Hypothesis : $P_0 \neq 0.70$ ✓

$q_0 = 1 - P_0 = \boxed{0.30}$

$\boxed{n = 200}$ $\boxed{x = 130}$

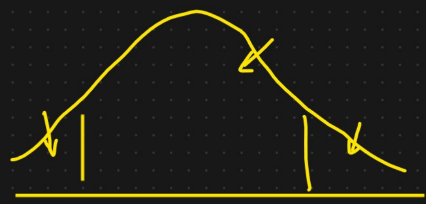
$$\hat{p} = \frac{130}{200} = \boxed{0.65}$$

Step 2 : $C.I = 0.95$ $\alpha = 0.05$

Step 3 :

Step 4 : Z test with proportion

$$Z_{test} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0}}$$

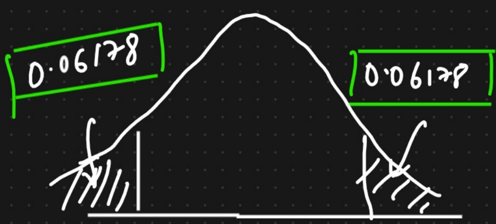


$$= \frac{0.65 - 0.70}{\sqrt{\frac{0.7 \times 0.3}{200}}} \approx -1.54$$

Conclusion

$-1.54 > -1.96$ Fail to Reject the Null hypothesis

Ratio



P-value = 0.06178 + 0.06178

= 0.12356

Pvalue > Significance value Fail To Reject Null Hypothesis.

④ A car company believes that the percentage of residents in City ABC that owns a vehicle is 60% or less. A sales manager disagrees with this. He conducts a hypothesis testing surveying 250 residents and found that 170 responded yes to owning a vehicle.

(a) State the Null & Alternate Hypothesis

(b) At 10% significance level, is there enough evidence to support the idea that vehicle ownership in City ABC is 60% or less?

Ans)

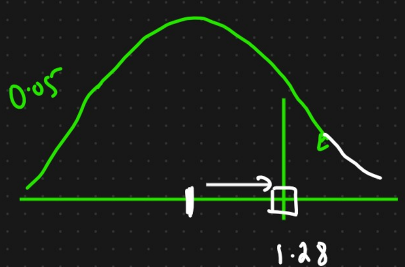
$H_0 : p_0 \leq 0.60$

$H_1 : p_0 > 0.60$

$\hat{p} = \frac{170}{250} = 0.68$

$q_0 = 0.40$

$Z\text{-score} = \dfrac{0.68 - 0.60}{\sqrt{\dfrac{0.6 \times 0.4}{250}}}$



$= \dfrac{0.08}{0.0309} = 2.588$

Reject the Null Hypothesis

① Chi Square test

① Chi Square test claims about population proportions.

ORDINAL DATA

NOMINAL DATA

↑

It is a non parametric test that is performed on categorical data.

⇑

① In the 2000 U.S census 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 the population distribution of ages has changed in the last 10 years?

Ans)

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

n:500

	<18	18-35	>35
Observed	121	288	91
Expected	100	150	250

Step 1 : Null hypothesis H_0 : The data meets the expected distribution

H_1 : The data does not meet the " "

Step 2 : $\alpha = 0.05$ C.I = 95%

Step 3 : Degree of freedom {categories}.

$df = c - 1 = 3 - 1 = \boxed{2}$

↳ No. of categories.

$\alpha = 0.05$

Step 4 : Decision Boundary = $\boxed{5.991}$ {chi square table}

Step 5 : Chi square Test Statistics

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

$$\boxed{\chi^2 = 232.494}$$

Conclusion

$\chi^2 > 5.99$ Reject H_0.

$\Downarrow$ $\Downarrow$ $\Downarrow$

* ANOVA TEST [F Test] → Assignment

3 Distribution

1. Binomial
2. Bernoulli
3. Power Law of Pareto,
4. Poisson.