① Central Limit Theorem

② Infrential Statistics

*) Z test {Z table} [5-6 problems].

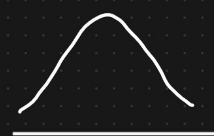
*) t test {t table}

*) Z test propotion population.

*) Chi square (categorical Test)

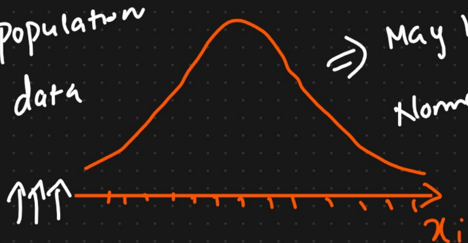
*) ANNOVA (F Test)

Infrential



① Central Limit Theorem

{ population data



⇒ May be Gaussian Normal Distr

$\boxed{n \geq 30}$ ✓ Sample mean distribution

$\longrightarrow$ $[x_1, x_2, x_3, x_4, x_5 \dots x_{30}] \rightarrow \bar{x}_1$

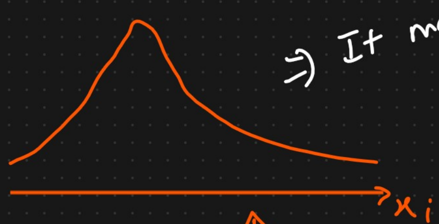
Sample 1

Sample 2 $[x_1, x_2, x_3, x_4, x_5 \dots x_{30}] \rightarrow \bar{x}_2$

$\rightarrow \bar{x}_3$

⋮

Sample m $\rightarrow \bar{x}_m$



⇒ It may not

10,

$\boxed{n \geq 30}$ ✓

Sample mean distribution

Gaussian Distribution

Normal Distribution

② Infuential Statistics {Data Anaysyt, Data Scientist}

① 100k ⇒ T-Shirt ⇒ No ⇒ Sample data ⇒ XL, L, Small

② iNeuron ⟶ Meetup → Hitesh ⇒ 300-400 people → T-shirts ⇓ Ordered L, M, XL, XXL

{ 20% L, 10% XXL, 10% XL, 60% Medium } 500 t-shirts

Next Event ÷ Z test $\bar{x}$ ⇓ $\boxed{\mu, s}$ → C.I [ ]

③ ATM ④ Measure the size of entire sharks

⑤ Amazon delivery { Percentile, Quartiles } ⇒

④ Hypothesis Testing

① A factory has a machine that fills 80ml of baby medicine 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 and Alternate Hypothesis

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

Ans) Step 1 $n = 40$ $\bar{x} = 78$ $s = 2.5$

$H_0: \mu = 80$ {Null Hypothesis}

$H_1 : \mu \neq 80$ {Alternate hypothesis}

Why Z test?

Step 2:

$\alpha = 0.05$ C.I = 95%

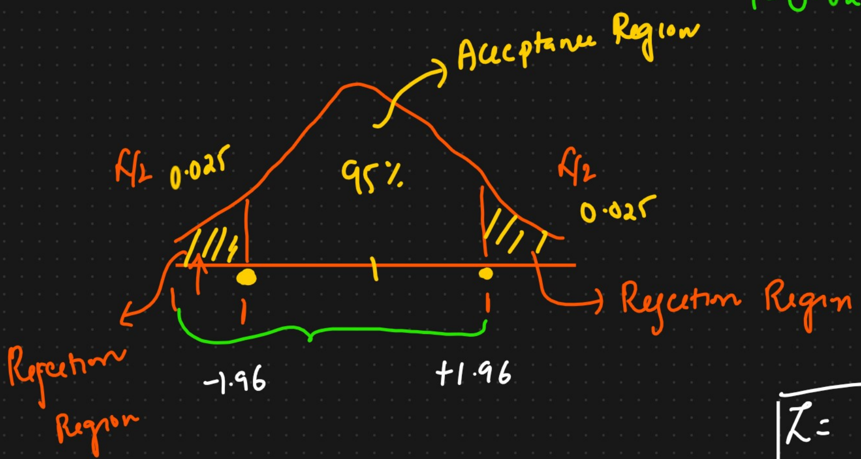
$\boxed{n < 30}$ { ① $n \geq 30$ $n \leq 30$ ② population std or sample std }

Step 3: Decision Boundary

$1 - 0.025 = 0.9750$

Why t test

{ ① Sample std ② $n < 30$ }



$$\boxed{Z = \frac{x_i - \mu}{\sigma / \sqrt{n}}}$$

$n = 1$ }

④ Calculate Test Statistics

$$Z = \frac{\bar{x} - \mu}{\boxed{s/\sqrt{n}}} \Rightarrow \text{Standard Error}$$

Sample Standard deviation

$$= \frac{78 - 80}{2.5/\sqrt{40}} = \frac{-2 \times \sqrt{40}}{2.5} = \frac{-2}{2.5} \times 6.32 = \boxed{-5.05}$$

⑤ State the Results

Decision Rule: If $Z = -5.05$ is less than -1.96 or greater than 1.96, then reject the null hypothesis with 95% C.I

Reject Ho Null hypothesis {There is some fault in the machine}

(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? $\left\{ \begin{array}{c} 95\% \\ \Downarrow \\ C.I \end{array} \right\}$

Ans) $\sigma = 15 \quad n = 30 \quad \bar{x} = 140$

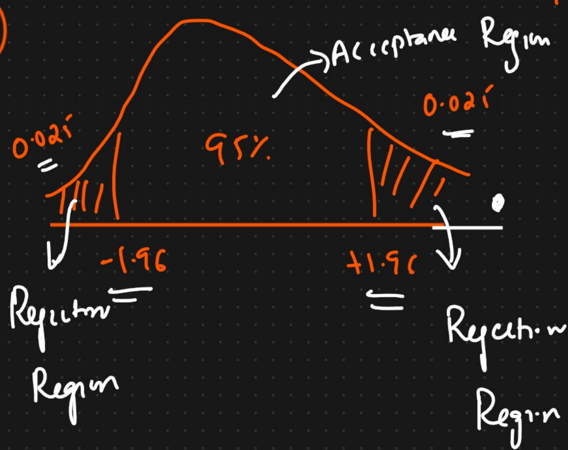
(1) $H_0 : \mu = 100$

$H_1 : \mu \neq 100$

(2) $\alpha = 0.05 \quad C.I = 95\%$

$\alpha = 0.05 \qquad 1 - 0.025 = 0.9750$

(3)



If Z is less than -1.96 or greater than 1.96, reject the Null Hypothesis

(4) $$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{140 - 100}{15/\sqrt{30}} = 14.60$$

$14.60 > 1.96$ Reject the Null Hypothesis

(x) A complain was registred, the boys in the Municipal Primary School are underfed. Average weight of boys of age 10 is 32kgs With S.D = 9kgs. A sample of 25 boys was selected from the municipal school and the average weight was found to be 29.5 kgs? With C.I = 95% Check whether it is True or False?

Ans) $\mu = 32\text{kgs}$ $\sigma = 9\text{kg}$ $n = 25$ $\bar{x} = 29.5$ $\alpha = 0.05$

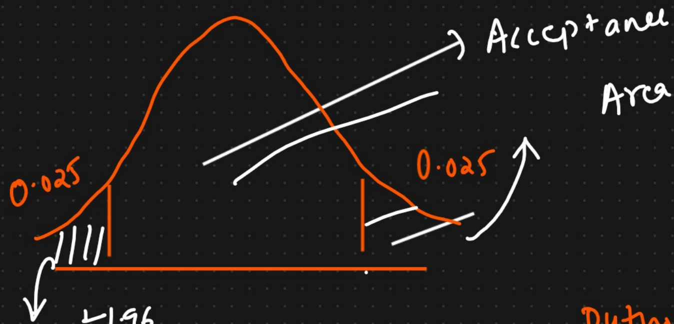
$1 - 0.95 = 0.05$

1) $H_0 = \mu = 32$

$H_1 = \mu < 32$

② $\alpha = 0.05$

③



85%

75%

Python

{z test, p-value}

④ $$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

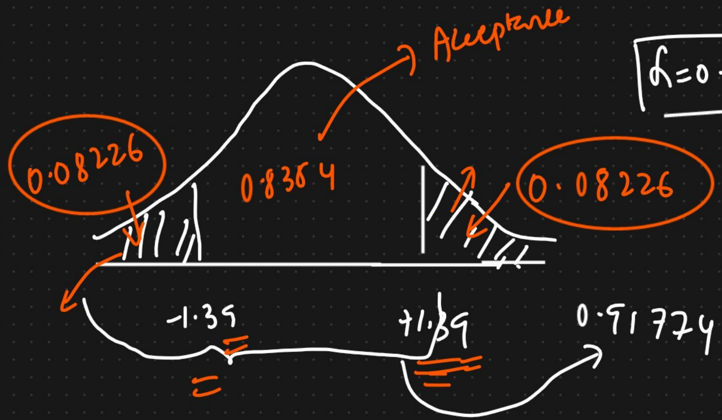
$$= \frac{29.5 - 32}{9/\sqrt{25}} = -1.39$$

Conclusion : $-1.39 > -1.96$ therefore we accept the Null Hypothesis

So, the boys are not underfed.

Significance value ⇒

P value



$\boxed{\alpha = 0.05}$ $\Rightarrow$ p value

$1 - 0.88226 - 0.08226 =$

p value

$1 - 0.91774 = 0.08226$

$\boxed{0.08226 + 0.08226} = \boxed{0.1645}$

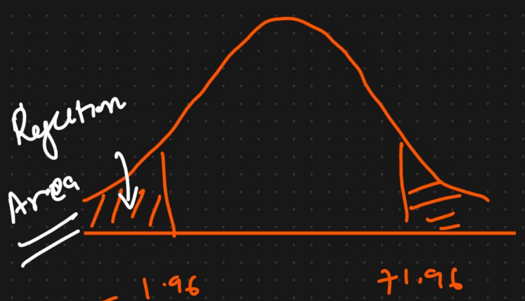
Significance value

$\uparrow$

$0.1645 > 0.05$

$\Downarrow$

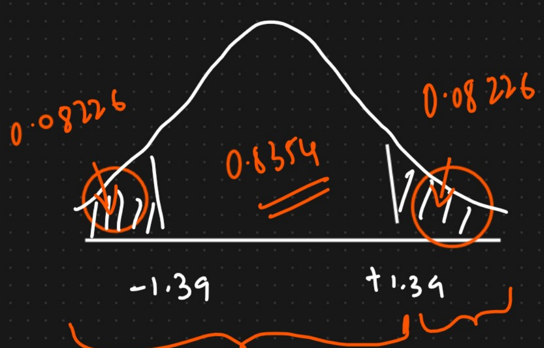
$P \geq 0.05 \Rightarrow$ Accept the Null Hypothesis



Z test

$\Downarrow$

p value



$\Rightarrow$ New p value $= 0.08226 + 0.08226$

$= 0.16$

$1 - 0.08226 - 0.08226$

Domain

$\Downarrow$

$0.1645 >$ Significance value

$\Downarrow$

Accept the Null Hypothesis.

$1 - 0.91774 = 0.08226$

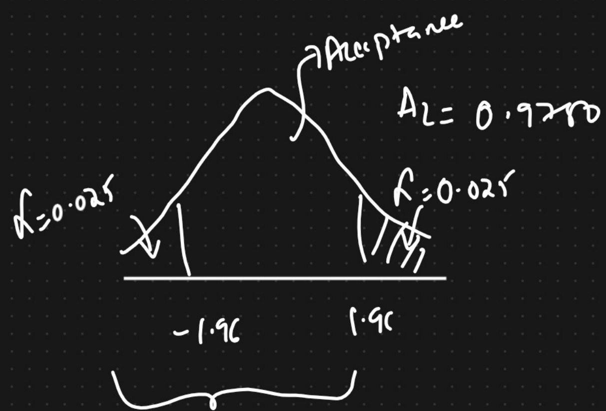
(1) The average weight of all residents in town XYZ is 168 lbs. A nutritionist believes the true mean to be different. She measured

(1) Assignment the weight of 36 individuals and found the mean to be 169.5 lbs with a standard deviation of 3.9.

(a) At 95% C.I is there enough evidence to discard the Null Hypothesis ??

Ans) $H_0 : \mu = 168$ $\qquad n = 36$ $\qquad \bar{x} = 169.5$ $\qquad s = 3.9$

$H_1 : \mu \neq 168$ $\qquad C = 0.95$ $\qquad \alpha = 1 - C.I = 0.05$



Z-test , t-test

$\Downarrow$

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = 2.31$$

$2.31 > 1.96$ Reject the Null Hypothesis

(2) A Company manufactures bike batteries with an average life span of 2 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

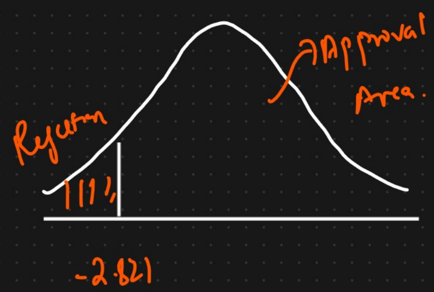
b) At a 99% C.I, is there enough evidence to discard the H_0?

Ans) $H_0 : \mu \geq 2$ $n = 10$ $\bar{x} = 1.8$ $s = 0.15$ {Sample std is given}

$H_1 : \mu < 2$ < 30 t-test??

② $\alpha = 0.01$ $\alpha = 1 - C.I = 1 - 0.99 = 0.01$

③



Degree of freedom $= n-1$
$= 9$

④ Calculate t-test Statistics:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.8 - 2}{0.15/\sqrt{10}} = \frac{-0.2}{0.15/3.1622} = -4.216$$

⑤ Conclusion

$-4.216 < -2.821$ Reject the Null Hypothesis.}

$\Downarrow$.

Z test with proportions

(1) 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 to

Owning a cell phone

(a) State the Null and Alternate Hypothesis?

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

$n = 200$ $\quad x = 130$

An) $H_0 \quad p_0 = 0.70.$

$H_1 \quad p_0 \neq 0.70$

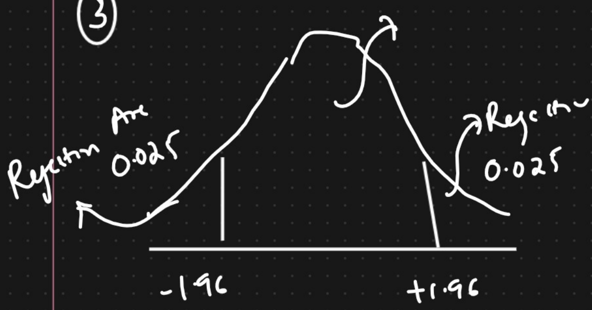
$$\hat{p} = \frac{x}{n} = \frac{130}{200} = \frac{13}{20} = 0.65$$

$q_0 = 1 - p_0$

② $\alpha = 0.05 \qquad C.I = 95\%$

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

③



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

At 95% C.I there is

$-1.54 > -1.96$, So we accept the Null Hypothesis

pvalue

$2 \times 0.06168 > 0.05$

Accept Null Hypothesis.

$1 - 0.93832 = 0.06168$

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?

p value = 0.014