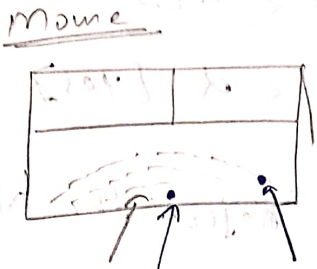
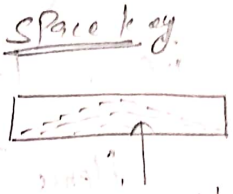
13/04/23

Statistics Day-6

P Value :→

Examples :→



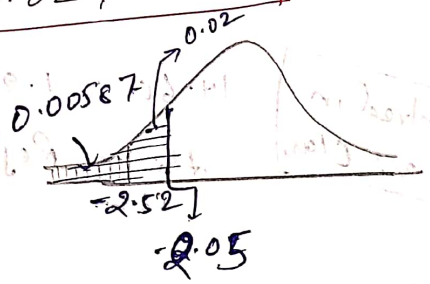


(More P value will be high)

P value ≠ {Significance Value (α)} → Derived from C.I.

Previous Question Solve Using P value (Assignment Q-1)

Q: $\mu = 5,\ n = 40,\ s = 0.50,\ \alpha = 0.02,\ \bar{x} = 4.8$



Sol: $C.I = 1-\alpha$

$= 1-0.02 = 0.98.$

Find Z-score

$$Z\text{-score} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.8-5}{0.50/\sqrt{40}} = \boxed{-2.52}$$

Now check ~~the are~~ in the Z-table

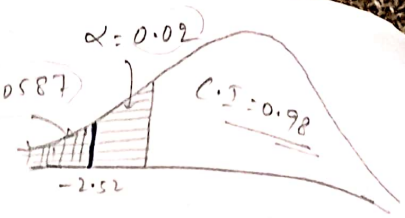
Area under -2.52.

$Z_{-2.52} = 0.00587.$

Calculate

Z_α i.e $Z_{0.02} = -2.05$

i.e area under α is 0.02



Hence

$0.00587 < 0.02$

i.e P-Value $< \alpha$ [Yes]

Hence We Reject the Null Hypothesis.

With the help of Z-score Whatever area we are getting is P-Value.

Assignment Q-2

Q2. $\mu = 100$, $\sigma = 15$, $n = 30$, $\bar{x} = 140$, C.I = 0.95, 2-tail test

Sol^n

$\alpha = 1 - 0.95 = 0.05$

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

Solved in last Class

$14.60 > 1.96$

Hence Reject the Null Hypothesis.

Q. The avg. Weight of all the residents in a town XYZ is 168 pounds. A nutritionist beleives 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.

Two tail test.

(a) State Null & Alterrate Hypothesis.

(b) 95% is there enough evidence to discard the Null Hypothesis?

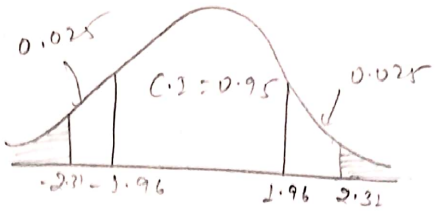
Sol^n

$\mu = 168$, $n = 36$, $\bar{x} = 169.5$, $s = 3.9$, C.I = 0.95

Step1

$H_0 : \mu = 168$

$H_1 : \mu \neq 168$



↓ Lies outside the C.I.

Step2:

C.I = 0.95, $\alpha = 1 - 0.95 = 0.05$

$\alpha/2$ since it is 2 tail test $\rightarrow Z_{\alpha/2}$ i.e $Z_{0.025} = 1.96$.

Step3:

$$\text{Z-score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{169.5 - 168}{3.9/\sqrt{36}}$$

$$= \frac{1.5 \times 6}{3.9}$$

$$= 2.3076 \approx 2.31$$

Step4

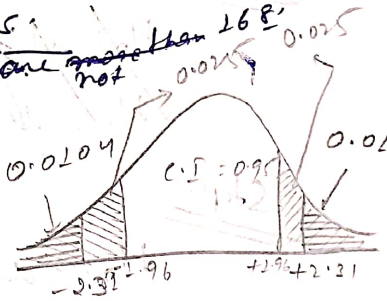
$2.31 > 1.96$

Hence Reject the Null Hypothesis.

Conclusion:- The avg. weight of residents are not 168.

Method2

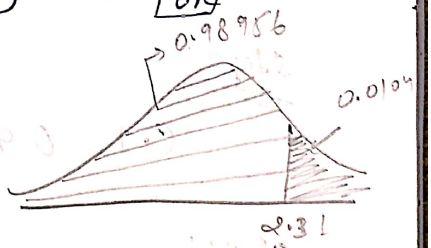
P-Value (Python libraries work on P-values)



from Z-table We can get the value of area under -2.31

i.e 0.0104 $\boxed{Z_{-2.31} = 0.0104}$

2 tail test OR



$\boxed{Z_{2.31} = 0.98956}$

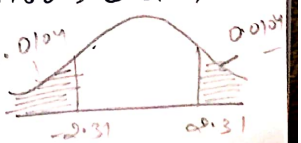
$P\text{-value} = 0.0104 + 0.0104$

$= 0.0208$

$\alpha\text{-value} = 0.05$

Hence $\boxed{0.0208 < 0.05}$

$\boxed{\text{area after } +2.31} = 1 - 0.98956 = 0.0104$



Hence: Reject the Null Hypothesis.

Conclusion: The avg. weight of the residents are not 168.

→ 1-Tail / test

Q: A Company manufactures bikes batteries with an avg. life span of 2 year or more years. An Engg. believes this value to be less. Using 10 samples, he measures the avg. life span to be 1.8 years with a SD of 0.15

(a) state the Null & Alternate Hypothesis.

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

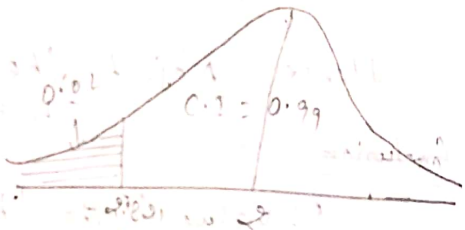
Sol"

$\mu = 2,\ n = 10,\ \bar{x} = 1.8,\ s = 0.15,\ C.I = 0.99$

Step1

~~$H_0: \mu \geq 2$~~ ~~$H_0: \geq 2$~~

~~$H_1: \mu < 2$~~ ~~$H_1:$~~



Step1

$H_0: \mu \geq 2$ {Null Hypothesis}

$H_1: \mu < 2$ {Alternate Hypothesis}

step2:

$C.I = 0.99,\ \alpha = 0.01$ i.e $(1 - 0.99) = 0.01$.

step3:

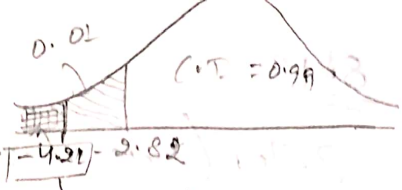
T-test (Since $n < 30$ & 's' is given)

Degree of freedom = $n - 1$

$= 10 - 1 = 9$.

Look into the t table

$T_{0.01} = 2.821$.



Step 4: Calculate t-test statistics

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

$$= \frac{-0.2 \times \sqrt{10}}{0.15}$$

$$= -4.2163$$

→ outside the Confidence Interval

Steps

Conclusion

$-4.2163 < -2.822$. Hence we Reject the Null Hypothesis.

The Avg. life of battery is less than 2 years.

#2. Test with Proportions → getting data in the form of %.

Q) 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. → 2-tail test

(a) State null and Alternate Hypothesis.

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

$n = 200$, $x = 130$

Sol: Step 1

Null Hypothesis: $P_0 = 0.70$ (Proportion)

Alternate Hypothesis: $P_0 \neq 0.70$

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

↓ Proportion of saying yes out of 200.

$q_0 = 1 - P_0 = 1 - 0.70 = 0.30$

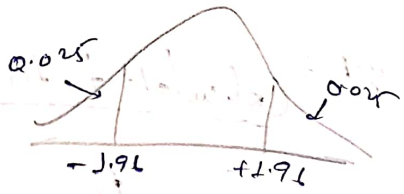
q_0 has been calculated using Alternate Hypothesis.

Step 2

C.I = 0.95

$\alpha = 1 - 0.95 = 0.05$

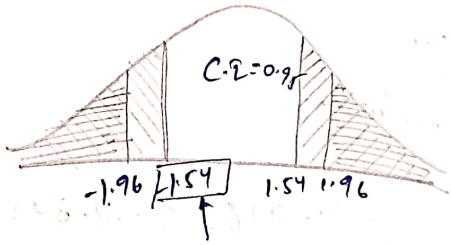
Steps:



Z-Test (Since $n > 30$)

Z-Test with Proportion formula

$$Z\text{-test} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}}$$



P1 comes under the C.I.

$$Z\text{-test} = \frac{0.65 - 0.70}{\sqrt{\frac{0.70 \times 0.30}{200}}}$$

$$= \frac{-0.05 \times \sqrt{200}}{\sqrt{0.21}} \approx -1.54$$

Conclusion

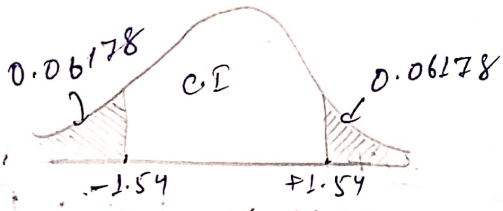
$-1.54 > -1.96$ Hence we fail to Reject the Null Hypothesis.

We can Conclude that,

Hence 70.1 People definately will have Cell Phone.

Method 2.

Using P-value :-



P-value = 0.06178 + 0.06178

= 0.12356.

Hence P-value > significance value (α)

$0.12356 > 0.05$ Hence we Reject the Null Hypothesis.

and we can Conclude that

70.1 of the people are having cell Phones.

1 tailed test.

Q. A Car Company believes that the percentage of residents in ABC that owns a vehicle is 60% or less. A Sales manager disagree with this. He conducts a Hypothesis testing Surveying 250 residents and found that 170 responded Yes to owning a vehicle.

(a) State the Null and Alternate Hypothesis.

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

Soln

Step1 Null Hypothesis: $P_0 \leq 0.6$

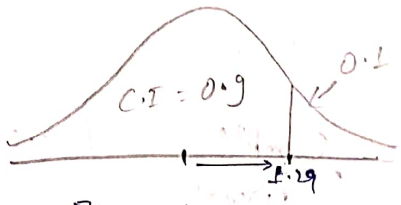
Alternate Hypothesis: $P_0 > 0.6$

$n = 250, \quad x = 170$

$\hat{P} = \frac{170}{250} = 0.68.$

Step2 $\alpha = 0.1 \quad C.I = 1 - 0.1 = 0.9$

$q_0 = 1 - 0.6 = 0.4$



$Z_{0.9} = 1.29$

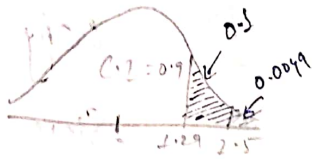
Step3 Z-Test (Since $n > 30$)

1-Tail test

~~$Z_{0.1}$~~ =

$$\text{Z-test} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{n}}} = \frac{0.68 - 0.60}{\sqrt{\frac{0.68 \times 0.60}{250}}}$$

$$= \frac{0.08 \times 5 \times 10}{\sqrt{0.68 \times 0.60}}$$



Outside the C.I

Step-4 $= \frac{0.08}{0.0309} = 2.588.$

Conclusion

$2.588 > 1.29$ Hence we Reject the Null Hypothesis

$P(2.58) = 1 - 0.9951 = 0.0049$: P-value

P-value = $0.0049 < 0.1$ P-value < α we Reject the Null Hypothesis

(α ↑ Significance value)

Chi Square test

(+) Chi square test claims about population proportions.

(+) It is a Non parametric test ie performed in Categorical data.

- Ordinal Data (Order Matters) eg- Rank.
- Nominal Data

Q. In the 2000 Us Census the age of individuals in a small town found to be the following.

Population Data:

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

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

Sample Data:

<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: Why is it chi-square test?

Since here we have three categories which is <18, 18-35 and >35.

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

Observed Data

$n = 500$

	<18	18-35	>35
Observed	121	288	91
Expected	20% of 500 = 100	30% of 500 = 150	50% of 500 = 250

Step1

Null Hypothesis: H_0: The data meets the expected distribution.

H_1: The data doesn't meet the expected distr.

Step2:

$\alpha = 0.05$ C.I = 1 - 0.05 = 0.95 i.e 95%.

Step3: Degree of freedom :→ In case of Chi-Square test we calculate the degree of freedom based on no. of Categories.

$df = c - 1$

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

Here we have total 3 Categories i.e <18, 18-35 & >35.

Step4. Decision Boundary

decision Boundary

$\alpha \rightarrow \boxed{\chi^2_{0.05} = 5.991}$ → Look into the Chi-Square table & find the value of χ_α.

χ^2_α

Step5. Chi-Square test statistics

observed, expected

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

fe → expected

$$= \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-500)^2}{250}$$

Chi-Square Value → $\chi^2 = 232.494$

Conclusion :-

$\chi^2 > 5.99$ Hence Reject the Null Hypothesis

Hence the data doesn't meet the expected distribution.