Basic stats.

Questions on hypothesis testing

Que: A factory manufactures cars with warrenty of 5 years or more on engine. An engineer belives that the engine will malfunction in less than 5 years. He tests the sample of 40 cars and finds the average time to be 4.8 years with a standard deviation of 0.5.

1) state null and alternate hypothesis.
2) At 2% significance level, is there enough evidence to support the idea that the warrenty should be revised?

* Before solving this question lets see P-value.

* P-value

→ P-value is used to support or reject null hypothesis.

→ P-value is evidence against null hypothesis.

→ calculate statistic (z, t, etc) at this ~~statistic~~ statistic value calculate area under curve from respective table.

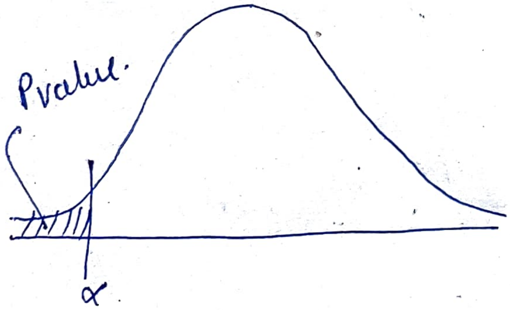
The area under the curve will give P-value.

1) if $P \geq \alpha$ ⟹ Null hypothesis cannot be rejected.

2) if $P < \alpha$ ⟹ Null hypothesis is rejected.

α = Significance value.

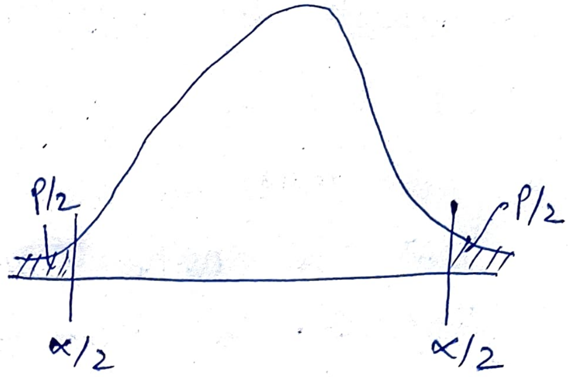
Graphically p-value is the area in the tail of a probablity distribution



one tailed Test

if $P \geq \alpha$ Ho accepted

$P < \alpha$ Ho rejected.



Two tailed test

$P_{net} \Rightarrow P_1 + P_2 = P/2 + P/2$

Now lets solve the previous Question.

1) using Test statistic.

$n = 40, \quad \bar{x} = 4.8 \quad s = 0.5, \quad \alpha = 0.02, \mu = 5$

① Null hypothesis (Ho)

Ho $\Rightarrow$ $\mu \geq 5$

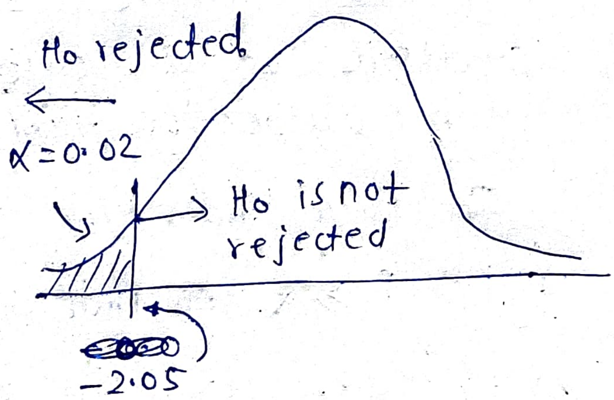
Alternate hypothesis (H_1)

$H_1 \Rightarrow \mu < 5$

↑ one tailed test (left)

② CI = 98% $\alpha = 0.02$

Desision boundary



$\because n \geq 30$

$\therefore$ z test

at $\alpha = 0.02$

from z table $z_{0.02} = -2.05$

③ z test

$$z\text{score} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.8-5}{0.5/\sqrt{40}} = -2.53$$

④ <u>conclusion</u>

∵ -2.53 ~~doesnt~~ ~~belong to~~ ~~$\notin$~~ is less than -2.05

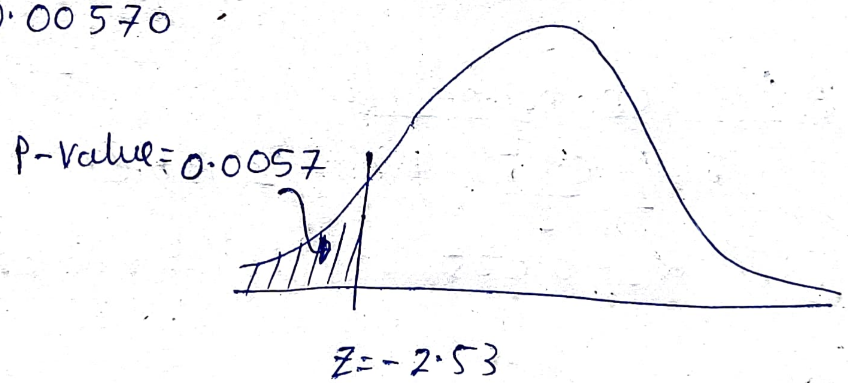
∴ null hypothesis is rejected

Therefore, there is enough evidence to support the idea to revise the warrenty.

② <u>using P- value</u>

at zscore -2.53 from z table area under curve at -2.53 will be p-value

∴ p-value = 0.00570



$\alpha = 0.02$

∵ p-value < α

∴ Null hypothesis is rejected.

ie there is enough evidence to support the idea to reject ~~null~~ null hypothesis and revise the warrenty.

Que: 2 The average weight of residents, in a town XYZ is 168 pounds. A nutritionist believes the mean to be different. She measured the weights of 36 individuals and found the mean to be 169.5 pounds with a standard deviation of 3.9.

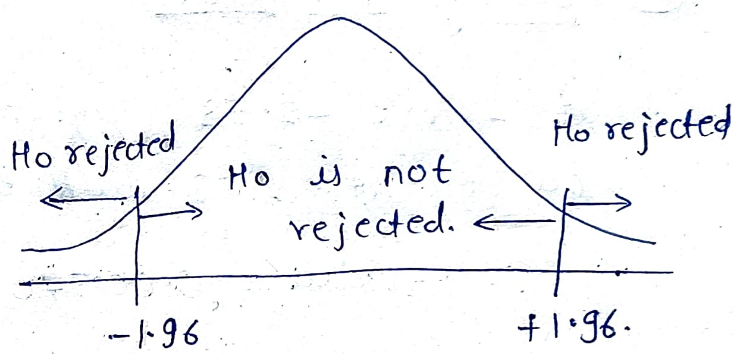
1) State Null and alternate hypothesis.
2) with 95% confidence interval, Is there enough evidence to discard the null hypothesis.

A) using test statistic

① $H_0 \Rightarrow \mu = 168$ pounds.
$H_1 \Rightarrow \mu \neq 168$ pounds
} Two tailed test.

② $CI = 95\%$

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



③ Decision boundary

@ $\alpha/2 = \frac{0.05}{2}$ from Z table

$\alpha/2 = 0.025$

$Z_{\alpha=0.025} = -1.96$

④ Z test $\because$ $n \geq 30$

$$Z_{score} = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{169.5 - 168}{3.9\sqrt{36}} = 2.31$$

⑤ Conclusion

$\because$ 2.31 doesn't lie b/w decision boundary ie −1.96 to +1.96

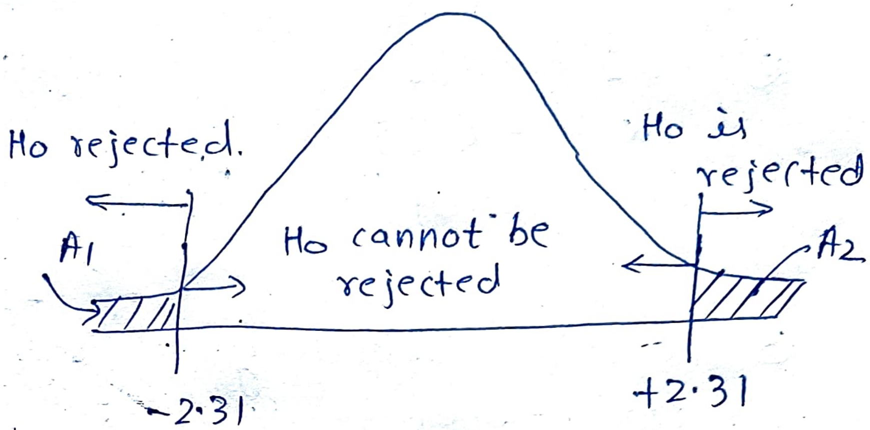
$\therefore$ Null hypothesis is rejected.
and there enough evidence to discard Null hypothesis.

(B) using P-value

at $z = 2.31$, from z table, area under curve at 2.31 will give P-value.

A_1 @ $z = -2.31$

$A_1 = 0.01044$.



A_2 @ $z = 2.31$

$A_2 = 1 - 0.98956$

$A_2 = 0.01044$.

$$P\text{-value} = A_1 + A_2$$
$$= 0.01044 + 0.01044$$
$$= 0.02088$$

or Directly using symmetry
$A_1 = A_2 = 0.01044$.

$\therefore$ P-value $< \alpha$ (0.05)

Null Hypothesis is rejected.

* Question on t-test

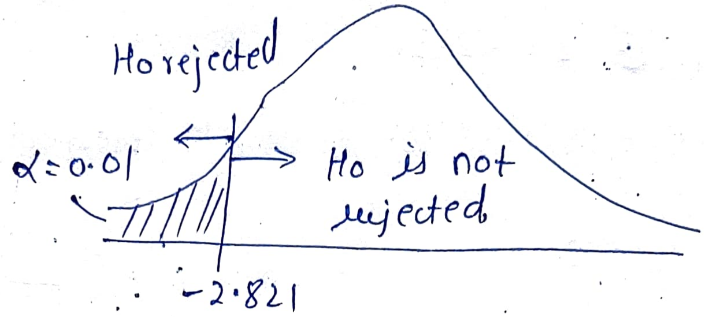
Que: A company manufactures bikes batteries with an average life span of 2 or more years. An Engineer believes that this value is less. Using 10 samples, he measures the average life span to be 1.8 years with standard deviation of 0.15.

1) state Null and alternate hypothesis.
2) at 99% CI, is there enough evidence to discard the Null hypothesis?

① using test statistic

① $H_0 \Rightarrow \mu \geq 2$
$H_1 \Rightarrow \mu < 2$ } one tailed Test Left.

② $CI = 99\%$
$\alpha = 1 - 0.99$
$= 0.01$



③ Decision Boundary
at $\alpha = 0.01$
and
degree of freedom = $n-1$
$dof = 10-1 = 9$

∵ $n < 30$ and s is given
∴ t-test is used.

from t-table

$t_{0.01} = 2.821$ @ $dof = 9$ and ∵ left tailed considering -2.821

④ t-test

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

$\bar{x} = 1.8$, $n = 10$
$\mu = 2$, $s = 0.15$
$\alpha = 0.01$

$t_{score} = -4.22$

⑤ conclusion

∵ $-4.22 < -2.821$
∴ Null hypothesis (H_0) is rejected.

② using p-values
∵ t table only has positive values.

$t_{score} = |-4.22| = 4.22$

at $t = 4.22$ and $dof = 9$

P-value = 0.001 (approximately)

* for actual value use linear Interpolation *

Now ∵ P-value < α (0.01)

∴ Null hypothesis is rejected.

* Z-test with proportions

$$\boxed{Z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}}$$

where

Z = test statistic

n = sample size

P_0 = null hypothesis value.

$\hat{P}$ = observed proportion

Que: A Tech company believes that percentage of residents in town XYZ that owns a cellphone 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 cellphone question.

1) state Null and alternate hypothesis.

2) At 95% CI, is there enough evidence to reject Null hypothesis.

Ans: Given: $P_0 = 70\%$ $\quad \bar{x} = 130$ $\quad CI = 95\%$

$\therefore P_0 = 0.7$ $\quad n = 200$

$$\hat{P} = \frac{\bar{x}}{n} = \frac{130}{200} = 0.65$$

① $\left.\begin{array}{l} H_0 \Rightarrow P_0 = 0.70 \\ H_1 \Rightarrow P_0 \neq 0.70 \end{array}\right\}$ ∴ Two Tailed Test.

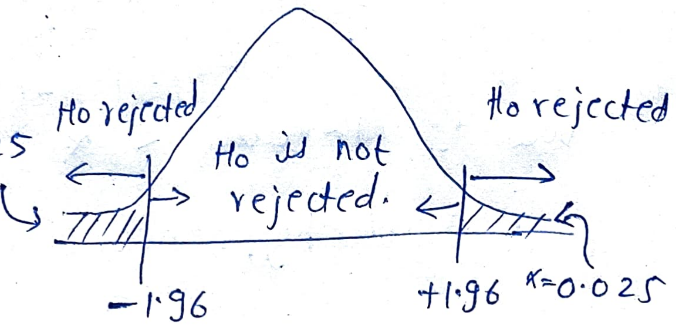
$q_0 = 1 - P_0 = 0.3$

② $CI = 0.95 \quad \alpha = 0.05$

$\because$ 2 tailed Test.

$Z_{\alpha/2} = Z_{0.025} = \pm 1.96$

③ Decision boundary



④ Z-test with proportions

$$Z\,score = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.65 - 0.7}{\sqrt{\frac{0.7(1-0.7)}{200}}} = -1.543$$

⑤ conclusion

$\because -1.543$ lies b/w -1.96 and $+1.96$ ie decision boundary

$\therefore$ we failed to reject Null hypothesis

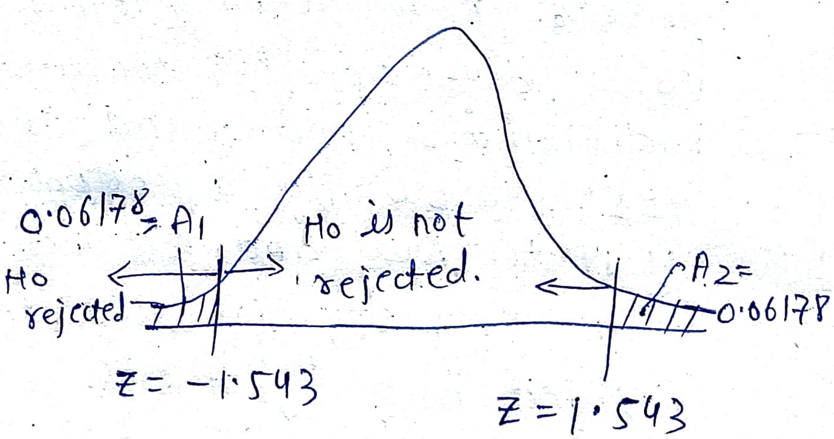
$\Rightarrow$ using p-value

using symmetry

$A_1 = A_2$

at $z = -1.543$

$A_1 = A_2 = 0.06178$



$\therefore$ Pvalue $= A_1 + A_2$

$= 0.06178 + 0.06178$

Pvalue $= 0.12356$

$\because$ p-value $> \alpha$ (0.05)

$\therefore$ we fail to reject Null hypothesis.

Que: A car company believes that the percentage of residents in city ABC that owns a car (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 Question.

1) State Null and alternate hypothesis.
2) At 10% significance level, is there enough evidence, to support the idea that vehicle ownership in city ABC is 60% or less.

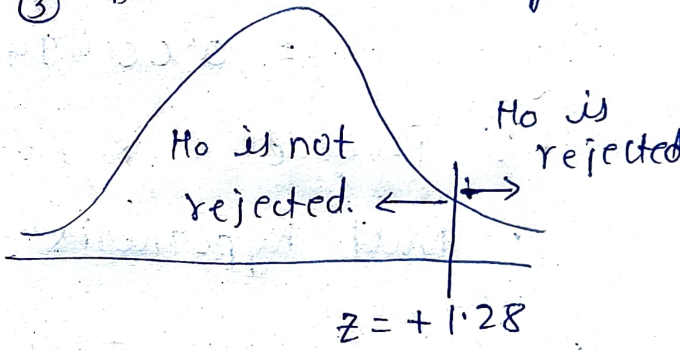
① <u>using ~~test~~ statistic</u>

① $H_0 \Rightarrow \quad P_0 \leq 0.60$

$H_1 \Rightarrow \quad P_0 > 0.60$

single tailed Test (Right)

③ Decision Boundary



② $\alpha = 0.1$, $CI = 0.99$.

$\therefore$ Z value at $\alpha = 0.1$

$Z = 1.28$

④ <u>Z-test with proportions</u>

$P_0 = 0.6$ $\quad x = 170$

$n = 250$

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

$$Z\text{score} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 - (1-P_0)}{n}}}$$

$$= \frac{0.68 - 0.6}{\sqrt{\frac{0.6(1-0.6)}{250}}}$$

$Z\text{score} = 2.582$

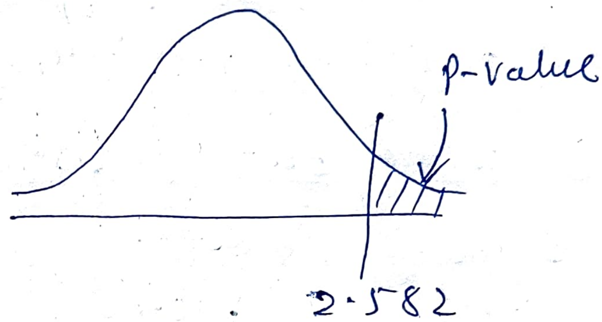
⑤ conclusions

∵ z value ie 2.582 is greater than decision boundary value of 1.28

∴ Null hypothesis is rejected

∴ we can conclude that more than 60% of residents own a vehicle in city ABC

② using p-value



at $Z = 2.582$

P value $= 1 - 0.99506$

$= 0.00494$

∵ P-value < α (0.1)

∴ Null hypothesis is rejected.

* Chi-Square Test

It claims about population proportions.

It is a non-parametric test that is performed on categorical data (ordinal and nominal)
eg: (rank) eg: (weekdays)

* chi-square ~~stat~~ statistic (X^2)

$$\boxed{X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}}$$

where X^2 = chi-square statistic
f_o = observed value
f_e = expected values.

degree of freedom = Number of categories minus one.

$\therefore \boxed{dof = \begin{pmatrix} \text{No. of} \\ \text{categories} \end{pmatrix} - 1}$

Note:

1) if $X^2 \leq$ Decision Boundary $\Rightarrow$ Ho cannot be rejected.
2) if $X^2 >$ Decision Boundary $\Rightarrow$ Ho is rejected

Question

In Year 2000, The US census of the age of individuals in a small town found to be following

Age	< 18 years	18 – 35	> 35 years
value	20 %	30 %	50 %

In year 2010, Ages of 500 individuals were sampled in same town and results are as below:

Age	< 18 years	18 – 35	> 35 years
Value	121	288	91

using $\alpha = 0.5$, would you conclude that the population distribution of ages has changed in last 10 years?

soln: No. of categories, $= 3$

$\alpha = 0.05$

$n = 500$ (No. of samples)

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

20% of 500 — 30% of 500 — 50% of 500

① $H_0 \Rightarrow$ population distribution is not changed.
$H_1 \Rightarrow$ population distribution is changed.

② $\alpha = 0.05$

dof = No. of categories − 1
= 3 − 1
= 2.

③ Decision Boundary
∴ from chi-square table
$\alpha = 0.05$, dof = 2

Area ⇒ 5.991
∴ Decision boundary ⇒ 5.991.

④ chi-square statistic (X^2)

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

$X^2 = 232.494$

⑤ conclusion

∵ X^2 > Decision Boundary

∴ Null hypothesis is rejected.

and we can conclude that population distribution has changed.