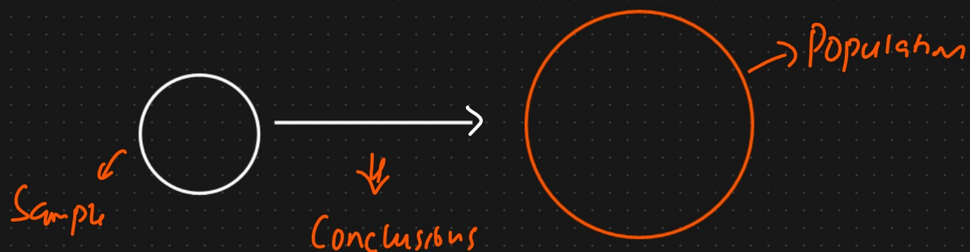
A Hypothesis And Hypothesis Testing Mechanism

Inferential Stats : Conclusion or Inference



Conclusions

⇓

Hypothesis Testing.

Person → Crime → Court

Hypothesis Testing Mechanism

① Null Hypothesis (H_0) - Person is not guilty

- The assumption you are beginning with

② Alternate Hypothesis (H_1) - The person is guilty

- Opposite of Null Hypothesis

③ Experiments → Statistical Analysis {P value, Significance value}

→ Collect Proof (DNA, Finger Test)

④ Accept the Null Hypothesis or Reject the Null Hypothesis

Eg: Colleges at District A states its average passed percentage of Students are 85%. A new college opened in the district and it was found that a Sample of Student 100 have a pass percentage of 90%. with a standard deviation of 4%.

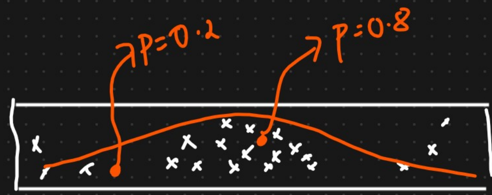
Does this College have a different passed percentage.

Ans) Null hypothesis $(H_0) = \mu = 85\%$

Alternate Hypothesis $(H_1) = \mu \neq 85\%$

Pvalue

The p value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.



Space bar

Out of 100 touches, we tuch around 20 times in this region

Hypothesis Testing

Eg : Coin is Fair or Not {100 times}

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

$P(H)=0.6$ $P(T)=0.4$

$P(H)=0.7$ $P(T)=0.3$

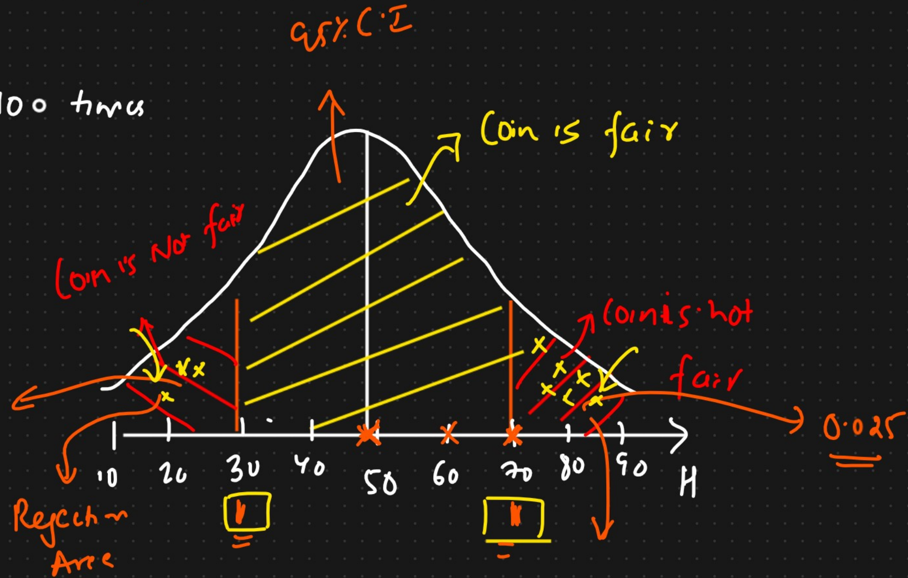
① Null Hypothesis :

H_0 : Coin is fair

② Alternate Hypothesis :

H_1 : Coin is not fair

③ Experiment : 100 times



④ Significance Value : $\alpha = 0.05$

$C.I = 1 - 0.05 = 0.95$

⑤ Conclusion $P <$ Significance value

Reject the Null Hypothesis

else

Fail to Reject the Null Hypothesis

Hypothesis Testing And Statistical Analysis

① Z Test } ⇒ Average ⇒ Z table → Z score And p value
② t Test } ⇒ Average ⇒ t table
③ CHI SQUARE ⇒ Categorical Data
④ ANNOVA ⇒ Variance

Z test i) Population std ii) $n \geq 30$

1) The average heights of all residents in a city is 168cm with a $\sigma = 3.9$. A doctor believes the mean to be diffrent. He measured the height of 36 individuals and found the average height to be 169.5 cm.

(a) State null and Alternate Hypothesis
(b) At a 95% confidence level, is there enough evidence to reject the null hypothesis.

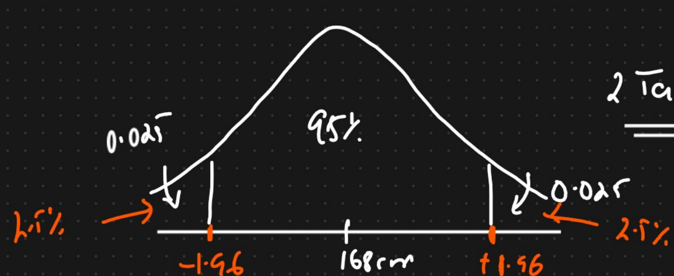
Ans) $\mu = 168cm$ $\sigma = 3.9$ $n = 36$ $\bar{x} = 169.5$

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

① Null Hypothesis $H_0 = \mu = 168cms$

② Alternate Hypothesis $H_1 = \mu \neq 168 cm$

③ Based on C.I We will draw Decision Boundary



2 Tail Test

$1 - 0.025 = 0.9750 \Rightarrow$ Z-score

$\Downarrow$

Area $\Rightarrow +1.96$

if $\underline{\underline{Z}}$ is less than -1.96 or greater than $+1.96$, Reject the Null Hypothesis.

Z-score $= \dfrac{\bar{X} - \mu}{\sigma}$

$\Downarrow$

$\dfrac{\sigma}{\sqrt{n}}$

Z-test

$$Z_d = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

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

$$Z_d = \frac{1.5}{0.65} = \boxed{2.31}$$

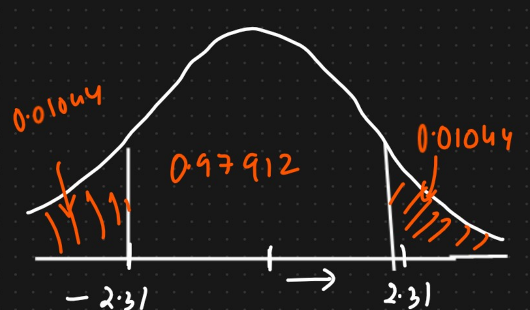
Z-score

Conclusion

$2.31 > 1.96$ Reject the Null Hypothesis

$\underline{\underline{p < 0.05}}$

0.98956



$1 - 0.98956 = 0$

$\Downarrow$

Final Conclusion the Average $\neq$ 168cm

The average height seems to increasing based on sample heights.

(1) p value $= 0.01044 + 0.01044$

$= 0.02088$

$p < 0.05$

$0.02088 < 0.05 \Rightarrow$ Reject the Null Hypothesis

② A factory manufactures bulbs with a average warranty of 5 years with standard deviation of 0.50. A worker believes that the bulb will malfunction in less than 5 years. He tests a sample of 40 bulbs and find the average time to be 4.8 years.

(a) State null and alternate hypothesis

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

Ans) $\mu = 5$ $\sigma = 0.50$ $n = 40$ $\bar{x} = 4.8$

a) Null hypothesis $H_0 : \mu = 5$

Alternate Hypothesis $H_1 : \mu < 5$ {1 Tail Test}

b) Decision Boundary

c) Z-test

$$Z_d = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{4.8 - 5}{0.50/\sqrt{40}}$$

$$= -2.53$$

Area under curve with z score $-2.53 = 0.0570$.

p-value = 0.0570

$\alpha = 0.02$

Compare p-value with α

$0.0570 < 0.02 \Rightarrow$ False

We accept the Null Hypothesis

We Fail to Reject the Null Hypothesis.

Student t distribution

In z stats when we perform any analysis using z-score
we require σ (population standard deviation) $\rightarrow$ is already known

How do we perform any analysis when we dont know the population standard deviation?

$\Downarrow$

Student's t distribution

t stats

S = Sample standard deviation

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

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

Z table

t table $\Rightarrow$ t test

Degree of freedom

$$dof = n-1 = 3-1 = 2$$

3 people

T-stats ÷ T test → One Sample t- test.

① In the population the average IQ is 100. A team of researchers want to test a new medication to see if it has either a positive or negative effect on intelligence, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140 with a standard deviation of 20. Did the medication affect intelligence? $C.I = 95\%$ $\alpha = 0.05$

Ans) $\mu = 100$ $n: 30$ $\bar{x} = 140$ $S = 20$ $C.I = 95\%$ $\alpha = 0.05$

① Null Hypothesis $H_0 ÷ \mu = 100$

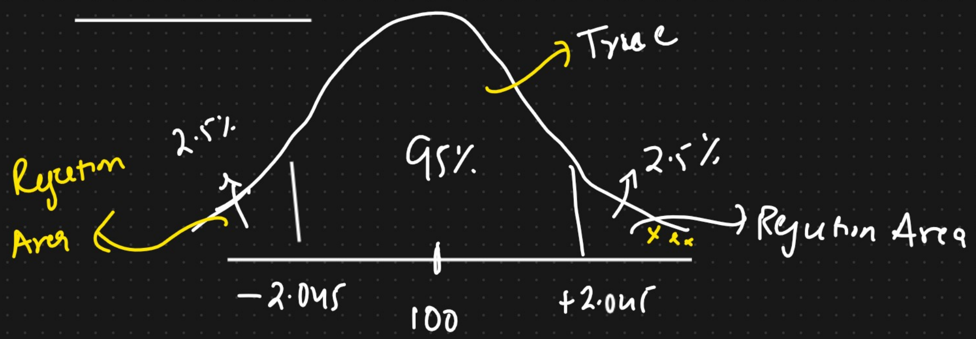
Alternate Hypothesis $H_1 ÷ \mu \neq 100$ {2 Tail Test}

② $\alpha = 0.05$

③ Degree of freedom

$$dof = n-1 = 30-1 = 29.$$

④ Decision Rule



if t test is less than −2.045 or greater than 2.045, reject the null hypothesis

(5) Calculate Test statistics

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{140 - 100}{20/\sqrt{30}} = \frac{40}{3.65} = 10.96$$

$$t = 10.96$$

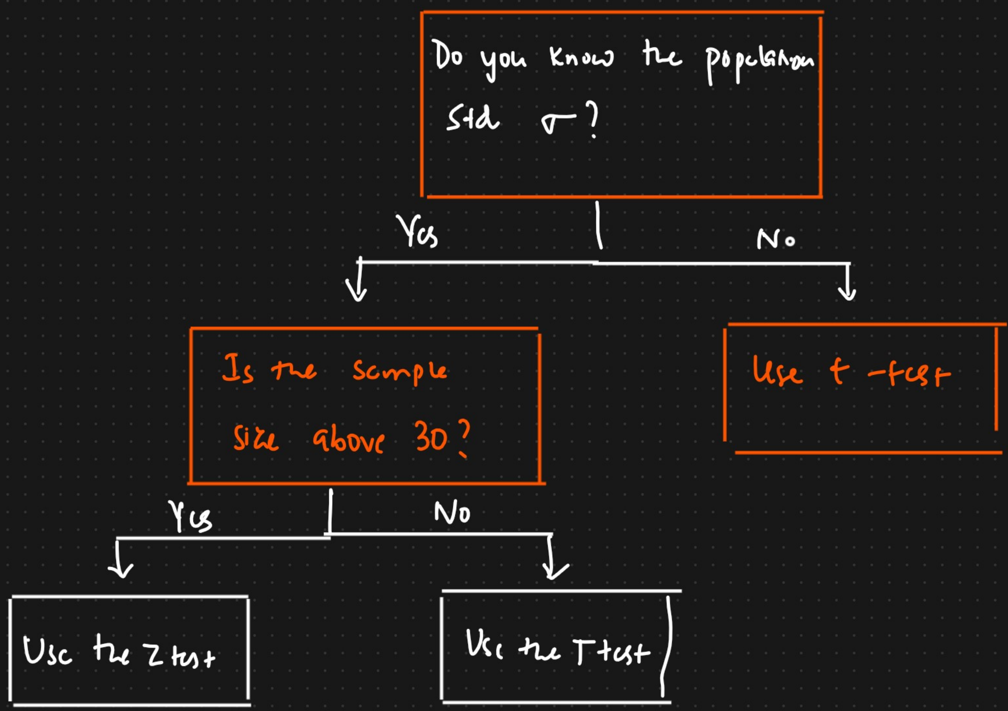
Since

$t = 10.96 > 2.045$ {Reject the Null Hypothesis}.

Conclusion : Medication used has affected the Intelligence

Medication has increased the Intelligence

When To Use T-tests Vs Z-test



Type 1 and Type 2 Errors

Reality : Null hypothesis is True or Null Hypothesis is False

Decision : Null Hypothesis is True or Null Hypothesis is False

Outcome 1 : We reject the Null Hypothesis when in reality it is false → Good

Outcome 2 : We reject the Null Hypothesis when in reality it is True → Type 1 Error

Outcome 3 : We retain the Null Hypothesis, when in reality it is False → Type 2 Error

Outcome 4 : We retain the Null Hypothesis when in reality it is True → Good

Bayes Statistics (Bayes Theorem)

Bayesian statistics is an approach to data analysis and parameter estimation based on Bayes' theorem.

Baye's Theorem

Probability → Independent Events
→ Dependent Events

① Independent Events

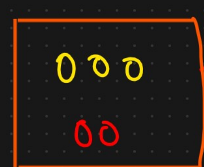
Eg: Rolling a dice
$\{1, 2, 3, 4, 5, 6\}$

$Pr(1) = \frac{1}{6}$ $\quad Pr(2) = \frac{1}{6}$ - - -

Tossing a coin

$Pr(H) = 0.5$ $\quad Pr(T) = 0.5$

② Dependent Event



Red → $Pr(R) = \frac{2}{5}$ $\xrightarrow{\text{Yellow}}$ $Pr(Y) = \frac{3}{4}$

⇒ Conditional probability

$$\boxed{Pr(R \text{ and } Y) = P(R) * \boxed{Pr(Y/R)}}$$

$$= 2/5 * 3/4 = \frac{6}{20}$$

$$Pr(A \text{ and } B) = Pr(B \text{ and } A)$$

$$Pr(A) * Pr(B/A) = Pr(B) * Pr(A/B)$$

$$\boxed{Pr(B/A) = \frac{Pr(B) * Pr(A/B)}{Pr(A)}}$$ ⇒ Baye's theorem

⇓

$$\boxed{Pr(A/B) = \frac{Pr(A) * Pr(B/A)}{Pr(B)}}$$

A, B = events

Pr(A/B) = Probability of A given B is true

Pr (B/A) = " " B " A is true

Pr(A), Pr(B) = Independent probabilities of A and B

DATASET

Independent

O/p / dependent

Size of House	No. of Rooms	Location	Price
x_1	x_2	x_3	y

$$\boxed{Pr(y/x_1, x_2\ x_3) = \frac{Pr(y) * Pr(x_1\ x_2\ x_3/y)}{Pr(x_1, x_2, x_3)}}$$

⇓

Baye's theorem