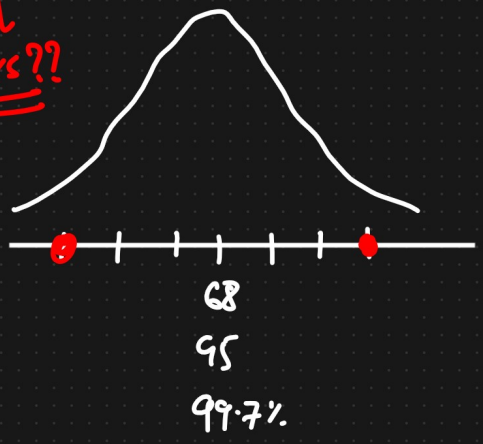
DAY 4 — STATS

1. IQR - python ✓
2. Probability ✓
3. Permutation And Combination ✓
4. Confidence Intervals ✓
5. P value ✓
6. Hypothesis Testing ✓

After 3rd sd outliers??



$$Z_{score} = \frac{x_i - \mu}{\sigma}$$ ✓

4. **Probability** : Probability is a measure of the likelihood of an Event

Eg : Roll a dice {1, 2, 3, 4, 5, 6}

$$Pr(6) = \frac{\#\text{ of way an event can occur}}{\#\text{ of possible outcome}}$$

$$= \frac{1}{6}$$

Toss a coin {H, T}

$$Pr(H) = \frac{1}{2}$$

② Addition Rule (Probability, "or")

Mutual Exclusive Event

Two Events Are mutual exclusive if they cannot occur at the same time

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

$\{1, 2\}$

Tossing a coin $\{H$

$T\}$

Non Mutual Exclusive

Multiple events can occur at the same time

Eg: Deck of cards $\{Q, ♡\}$

① If I Toss a coin, what is the Probability of the coin landing on heads or tails?

Ans) Mutual Exclusive Addition Rule

$$Pr(A \text{ or } B) = P(A) + P(B)$$

$$= \frac{1}{2} + \frac{1}{2}$$

$$\boxed{Pr(A \text{ or } B) = 1}$$

Roll a Dice

$$Pr(1 \text{ or } 3 \text{ or } 6) = Pr(1) + Pr(3) + Pr(6)$$

$$= 1/6 + 1/6 + 1/6$$

$$= 3/6 = 1/2 = 0.5 //.$$

$\frac{52}{52}$ $\frac{2}{}$

Non Mutual Exclusive

You are picking a card randomly from a deck. What is the probability of choosing a card that is queen or a heart? → (52)

Ans) Non mutual Exclusive

$$P(Q) = \frac{4}{52} \qquad P(\heartsuit) = \frac{13}{52} \qquad Pr(Q \text{ and } \heartsuit) = \frac{1}{52}$$

Addition Rule for non mutual exclusive Events

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(Q \text{ or } \heartsuit) = P(Q) + P(\heartsuit) - P(Q \text{ and } \heartsuit)$$

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$= \frac{16}{52} \text{ a} \quad //$$

③ Multiplication Rule

{ Independent Events }

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

1, 1, 2, Each & Every are independent → Red marble ●

Dependent Events → $Pr(Red) = \frac{3}{5}$ → $Pr(Green) = \frac{2}{4}$

Naive Bayes {Conditional Probability}

Independent Events

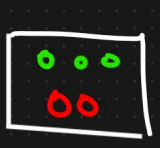
④ What is the probability of rolling a "5" and then a "4" in a dice?

Ans) Independent Event

Multiplication Rule

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

$Pr(5 \text{ and } 4) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$ //.



3/5

$\frac{2}{4}$

⑤ What is the probability of drawing a Queen and then a Aces from a deck of cards?

$P(Q \text{ and } R) = P(Q) * P(R/Q)$ → Event

Ans) Dependent

$P(A \text{ and } B) = P(A) * P(B/A)$ → conditional probability

↓

Bayes theorem

$P(Q \text{ and } A) = P(Q) * P(A/Q)$

$= \frac{4}{52} * 4/51$

⑥ Permutation and Combination

Permutation

School trip {choclate factory} → Dairy, 5 star, Milky bar, Eclairs, Gem, Silk

Student {Assignment}

Student → $\underline{6} \times \underline{5} \times \underline{4} = \underline{120}$

⑤

Dairy, Gem, Milky
Milky, Gem, Dairy

$n = 6$
$r = 3$

Permutation

$$^{n}P_{r} = \frac{n!}{(n-r)!} = \frac{6!}{(6-3)!}$$

$$= \frac{6 \times 5 \times 4 \times 3!}{3!}$$

$$= 120$$

⑥ Combination

Dairy	Gem	Eclair
—	—	—

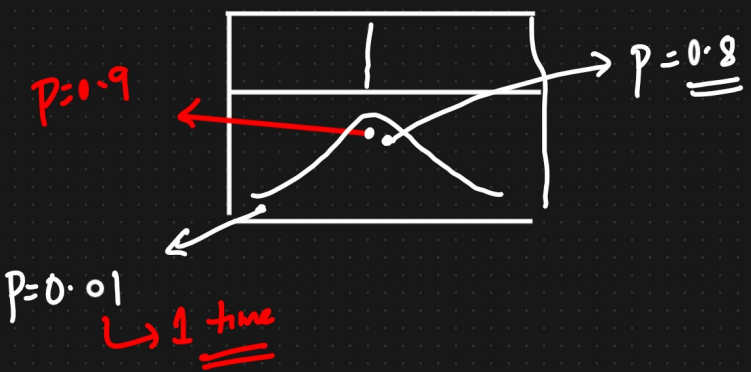
$$^{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{6!}{3!(6-3)!}$$

$$= \frac{6 \times 5 \times 4 \times 3!}{3 \times 2 \times 1 \times 3!}$$

$$= 20$$

① p value { Many people gets confused }

Every 100 time I touch the mouse pad 80 times I touch this specific region



Hypothesis Testing, Confidence Interval, Significance Value, - - - -

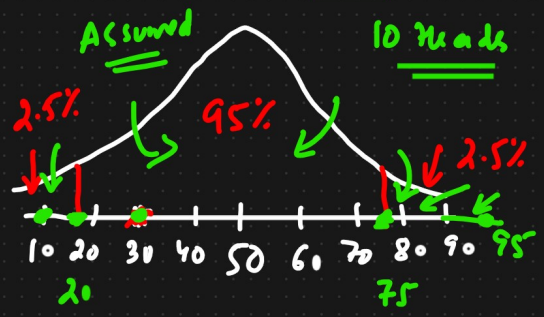
Coin → Test whether this coin is a fair coin or not by performing 100 tosses

Shady coin P(H) = 100%

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

100 → 50 times Head (The coin is fair)

30 Times Head 95 heads



Hypothesis Testing

100% − 5%

⇓

95%

CI

① Null Hypothesis : Coin is fair

✓② Alternate Hypothesis :- Coin is unfair

③ Experiment

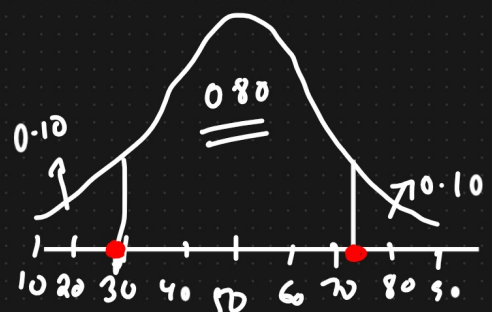
④ Reject or Accept the Null Hypothesis

Significance Value

α = 0.05 { Domain Expert }

α = 0.20 α = 0.3 C.I = 70%

80%



1

25

Today Topics

1. Type 1 and Type 2 Error ✓
2. One Tailed and 2 Tailed Test ✓
3. Confidence Interval ✓
4. Z-test, t-test, Chi-Square Test

1. Type 1 and Type 2 Error

Null Hypothesis (H_0) = Coin is fair

Alternate Hypothesis (H_1) = Coin is not fair

Reality check

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 → Yes

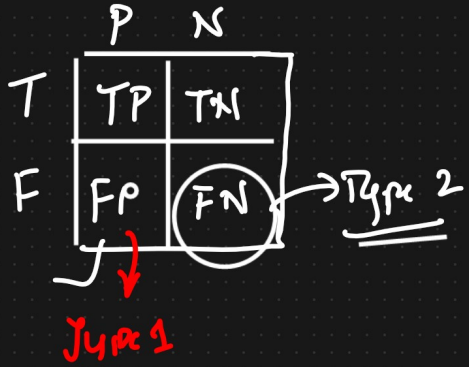
Movie

Person - Death Sentence

Outcome 2 : We reject the Null Hypothesis, When in reality it is true → Type 1 Error

Outcome 3 : We Accept/retain the Null Hypothesis when in reality it is false → Type 2 Error

Outcome 4 : We Accept the Null Hypothesis when in reality it is true → Good

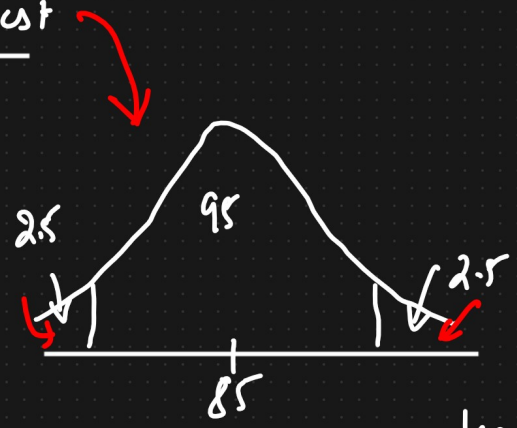


② 1 Tail and 2 tail Test

Eg: Colleges in Karnataka have an 85% placement rate. A new college was recently opened and it was found that a sample of 150 students had a placement rate of 88%. With a standard deviation 4%. Does this college has a different placement rate ? $\alpha = 0.05$

↳ 85%.

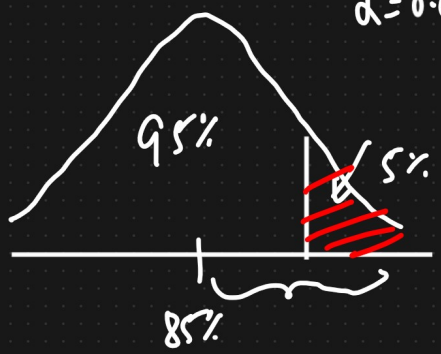
2 tailed Test



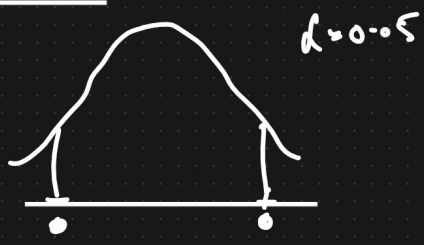
lenes

Does this college have a placement rate greater than 85% ?

$\alpha = 0.05$



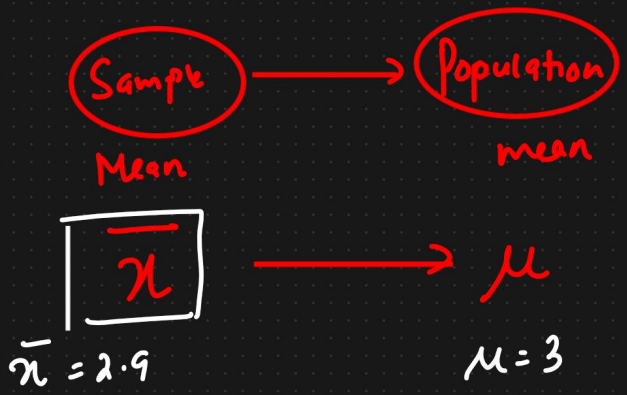
③ Confidence Intervals



Point Estimate

The value of any statistic that estimates the value of a parameter

Influential Stats



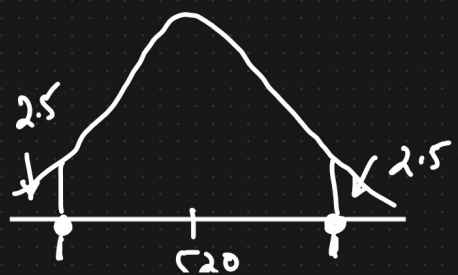
Confidence Intervals

Point Estimate ± Margin of Error

Q) On the quant test of CAT Exam, the standard deviation is known to be 100. A sample of 25 test takers has a mean of 520 score. Construct a 95% CI about the mean?

Ans) $\sigma = 100$ $n = 25$ $\alpha = 0.05$ $\bar{x} = 520$

$\alpha = 1 - 0.95 = 0.05$



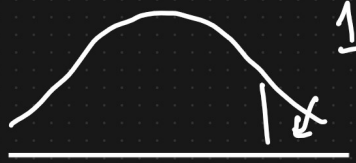
{ ① Population std is given } → Z test
{ ② $n \geq 30$ }

Point Estimate ± Margin of Error

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$ → Standard Error

$$Z_{\frac{0.05}{2}} = Z_{0.025}$$

$$1 - 0.025 = 0.975$$



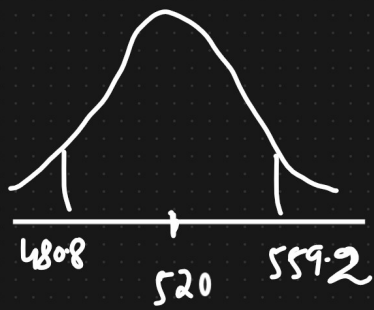
1.96

Upper bound $= \bar{x} + Z_{\frac{0.05}{2}} \frac{100}{\sqrt{25}}$

Lower bound $= \bar{x} - Z_{\frac{0.05}{2}} \frac{100}{\sqrt{25}}$

Upper $= 520 + 1.96\,(20) = 559.2$

Lower $= 520 - 1.96\,(20) = 480.8$



Stats

Find the average size of the sharks throughout the world?

σ, $\bar{x}$, n, $\alpha = 0.05$

Q) On the quant test of CAT exam, a sample of 25 test takers has a mean of 520 with a standard deviation of 80. Construct 95% confidence interval about the mean?

Ans) Condition $\quad n = 25 \quad \bar{x} = 520 \quad S = 80 \quad \alpha = 0.05$

Here population sd is not given → t-test

Point Estimate ± Margin of Error

$$\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \rightarrow \text{Standard Error}$$

$$t_{\frac{0.05}{2}} = 2.064$$

$$\text{Upper bound} = \bar{x} + t_{0.05/2}\left(\frac{s}{\sqrt{n}}\right)$$

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

$$= 520 + 2.064\left(\frac{80}{5}\right)$$

$$= 553.024$$

$$\text{Lower bound} = \bar{x} - t_{0.05/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$= 520 - 2.064\left(\frac{80}{5}\right)$$

$$= 486.97$$

$$[486.97 \longleftrightarrow 553.024]$$

① One Sample Z-Test

① Population sd is given

② Sample size $n \geq 30$

(*) In the population, the average IQ is 100 with a sd of 15. Researchers wants to test a new medication to see if there is 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. Did the medication affect the intelligence?

$\alpha = 0.05$ (CI = 95%)

→ 110 ✓

Ans) 1) Define Null Hypothesis

$H_0 = \mu = 100$

② Alternate Hypothesis H_1 $\mu \neq 100$

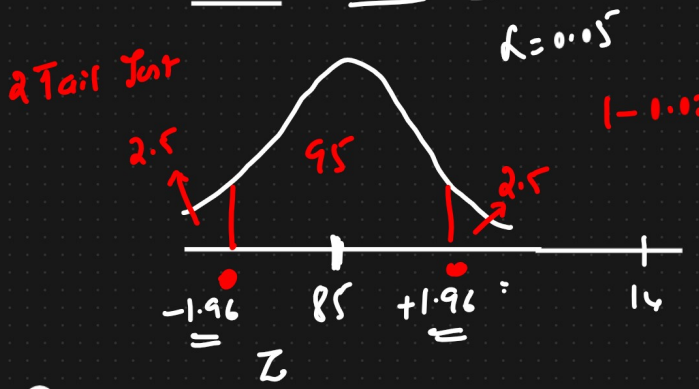
③ State Alpha

$\alpha = 0.05$

④ State Decision Rule

Z table

$1 - 0.025 = 0.975$



$P \leq 0.05$

⑤ Calculate Test Statistics

Sample data

$$Z = \frac{\bar{x} - \mu}{\left\{\frac{\sigma}{\sqrt{n}}\right\}} \rightarrow \text{Standard Error}$$

$$= \frac{140 - 100}{\frac{15}{\sqrt{30}}} = \frac{40}{15} \times \sqrt{30} = 14.60$$

State our Decision

$\{Z = -0.2\}$

$Z = 14.60$

$14.60 > 1.96$

If Z is less than -1.96 or greater than 1.96, reject the null hypothesis

Medication improve the intelligence or decrease ?!

Improve the intelligence

② One Sample t-test

Z-test ⇒ population std

t-test ⇒ unknown population std

EQ) Population the average IQ = 100

$n = 30 \quad \bar{x} = 140 \quad s = 20$

Did the medication affect intelligence?

$\alpha = 0.05$

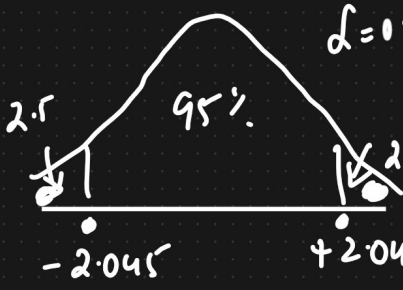
Ans) ① $H_0 = \mu = 100$

② $H_1 \quad \mu \neq 100$

③ Calculate the degree of freedom

$n - 1 = 30 - 1 = 29$

④ State Decision Rule



$\alpha = 0.05 \quad t = 10.96 > 2.045$

Reject Null Hypothesis

$P \leq$ significance value

⇓

{ Increased the intelligence. }

⑤ T Test

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

$\bar{x} = 140$

$\mu = 100$

$s = 20$

$n = 30$

$= 10.96$

Reject the Null Hypothesis

Real World Problem