Statistical test

① T test (Compare of Mean) (or) (Measure of mean)
→ Sample Standard Deviation (S) is given
→ Sample test

② Z Test (Compare of Mean) (or) (Measure of mean)
→ Population standard Deviation (σ) is given

③ Anova Test (Analysis of Variance)

④ F Test (Comparision of variance)

⑤ Chi-Square (Compare Categorical variable) (or) (Measure or compare between Categorical variable)

① One Sample Z Test

Q) In a population the average IQ is 100 with Standard deviation 15, then the doctor tested the new meditation to find out whether it increases the IQ or decreases the IQ? After 1 month sample of 30 participants were taken and these 30 participants had a mean of 140. Did medication affect intelligence? ($\alpha = 0.05$ given)

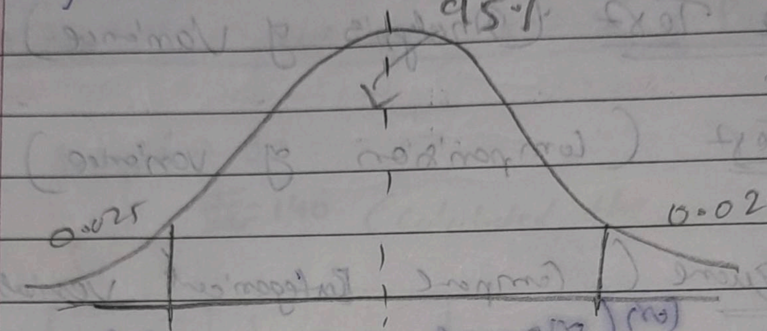
⇒ $\boxed{\mu=100}$ $\boxed{\sigma=15}$ $\alpha = 0.05$

1 month $n = 30$ $\boxed{\bar{x} = 140}$

① Define Null and Hypothesis testing

{Null Hypothesis} = $H_0 \Rightarrow \mu = 100$ ~~not affecte~~d

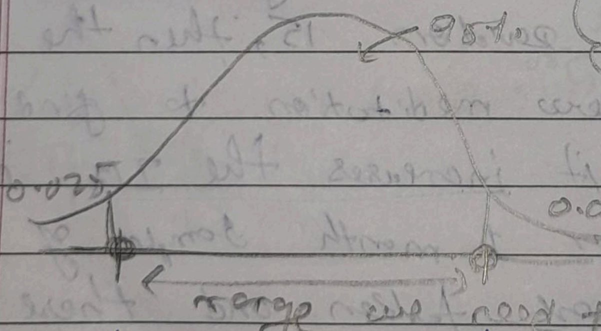
{Alternate Hypothesis} = $H_1 \Rightarrow \mu \neq 100$



② State Alpha $\alpha = 0.05$ (two tailed test)
(Refer graph)

③ State Decision rule

Entire region = 1



-1.96 $+1.96$ (Check Z table)

- Entire region = 1.

1 − 0.025 (Area under the curve)
→ 0.9750

Note • If Z is less than −1.96 or greater than 1.96, we will reject null hypothesis

Z-Score :- It is used to find out how how many Standard Deviation (σ) Away from the Mean (μ)

④ Compute Z- test Statistics

$$\boxed{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}} \Rightarrow \frac{140 - 100}{\frac{15}{\sqrt{30}}} = \frac{40}{2.74} \Rightarrow \underline{14.68}$$

we dont have population data we have divide the σ (SD) by $\sqrt{n}$

⑤ State ~~Res~~ Result

14.60 $\xrightarrow{\text{does not}}$ ~~fall~~ in range -1.96 to $+1.96$

- Since, Z = 14.60 is greater than 1.96.

$\therefore 14.68 \neq -1.96$ to 1.96

Reject the Null Hypothesis

Accept the Alternate Hypothesis

Conclusion :- There is a change in IQ

② One Sample Z test with Proportion

Q) A survey claims that 9 out of 10 to doctor recommend asprin for their patient with Headache. To test this claim a ~~ch~~ random sume of 100 doctor is taken out of those 100 doctors, 82 indicate that they recommend ~~Aspirent~~ Aspirin. Is this claim Accurate?

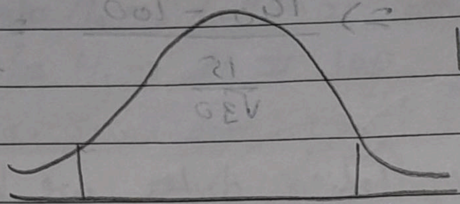
$\alpha = 0.05$

Date :
Page :

(1) Null Hypothesis $H_0 \Rightarrow P = 0.90$

Alternate Hypothesis $H_1 \Rightarrow P \neq 0.90$

(2) Significance value $\alpha = 0.05$
(Two tailed test it is)



-1.96 $+1.96$

$\boxed{1 - 0.05 = 0.9750}$

(3) Decision Rule : (Refer previous sum)

(4) Z Statistics

$$\Rightarrow Z_0 = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$\hat{P} = 0.82$ $P_0 = 0.90$ ~~n = 100~~

n (or) $x = 100$

$$= \frac{0.82 - 0.90}{\sqrt{\frac{0.90(1-10)}{100}}} \Rightarrow -2.667$$

★ (5) State Result

claim is Inaccurate

{ Ans $\Rightarrow -2.667 \notin [-1.96, 1.96]$

~~Reject~~ Accept Null Hypothesis
Reject Null Hypothesis }

Probability

1) It is the measure of the likelihood of an Event.

(1) Eg = Tossing a Coin

Sample Space

$S = \{H, T\}$

Probability = $\dfrac{\text{No. of ways an Event can Occur}}{\text{Total No. of Possible Outcomes}}$

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

(2) Rolling a Dice

$S = \{1, 2, 3, 4, 5, 6\}$

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

→ Two Events are mutually Exclusive if they cannot occur at the same time

(1) Mutual Exclusive Event

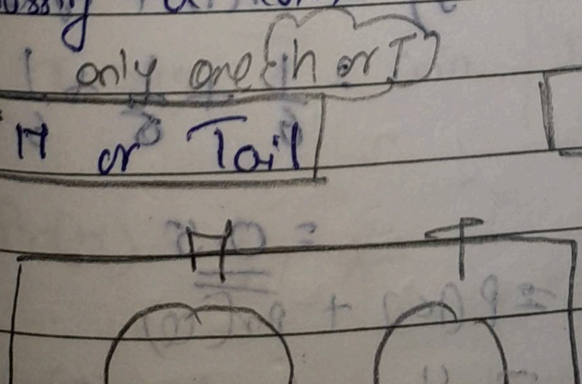
→ Only one event will happen at a time. 2 event cannot happen

• Tossing a Coin

only one (h or T) $\qquad$ not both

$\boxed{\text{H or Tail}}$ ✓ $\qquad$ $\boxed{\text{H \& Tail}}$

Venn Diagram →

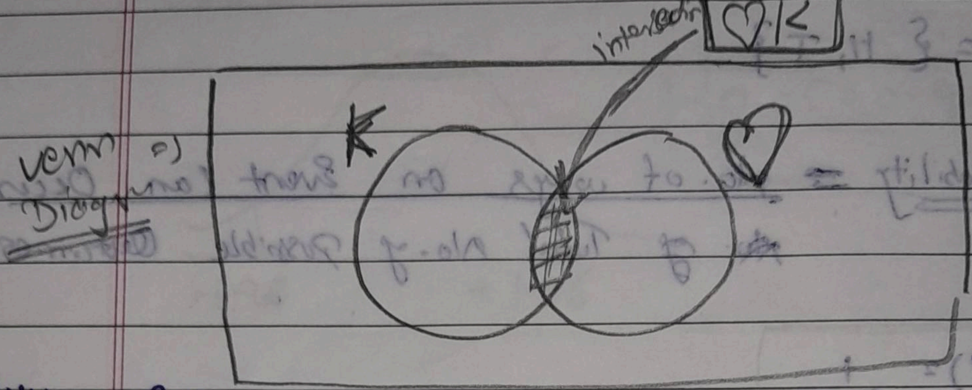


either h or T

(2) Not Mutual Exclusive Event

- Taking out a card from deck

$\boxed{K \ \& \ \heartsuit} \Rightarrow$ Two outcome can come at a time



* Additive Rules for Mutual Exclusive

(1) eg Tossing a Coin

what is Pr of getting,

$\Rightarrow Pr(H \text{ or } Tails) = Pr(H) + Pr(T)$

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

$= 1$

(2) eg Rolling a Dice

$Pr(1 \text{ or } 2 \text{ or } 3) = Pr(1) + Pr(2) + P(3)$

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

$= 0.5$

(3) eg $P(K \text{ or } Queen) = P(K) + Pr(Q)$

$= \frac{4}{52} + \frac{4}{52} = \frac{8}{52} \Rightarrow \frac{3}{3}$

Additive Rule for Non-Mutual Exclusive Event

eg. Taking a card from the deck

Pr(K or ♡)

- Pr(K) = $\frac{4}{52}$
- Pr(♡) = $\frac{13}{52}$
- Pr(K and ♡) = $\frac{1}{52}$

$\Rightarrow$ Pr(K or ♡) = Pr(K) + Pr(♡) − P(K ∩ ♡)

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

$= \frac{16}{52} \Rightarrow \frac{4}{13}$ //

Additive Rule

ME $\Rightarrow$ Pr(A or B) = Pr(A) + Pr(B)

NME $\Rightarrow$ Pr(A or B) = Pr(A) + Pr(B) − Pr(A ∩ B)

Multiplicative Rule

Independent Event : Two Event are Said to be independent if they do not affect one another

(1) Independ Event

eg Tossing a coin | Tossing a coin for 3 times {1, 2, 3} H T H

Pr(H) = $\frac{1}{2}$ * No. of Total outcomes will not reduce

Pr(T) = 1

(2) Dependent Event

(i) eg :- Taking a card from a deck

1st Experiment (K) $Pr(K) = \frac{1}{\boxed{52}}$

2nd Second (Q) $Pr(Q) = \frac{1}{\boxed{51}}$.

3rd Experiment (J) $\frac{1}{52}$

$Pr(J) = \frac{1}{\boxed{50}}$

Note

Here we are able to see that w.r.t Dependent Event the outcomes will be changing, there probability will be changing, because the no. of cards will be reducing

Independent Event

eg :- Rolling a dice =) $\boxed{Pr(1 \text{ and } 3)}$

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

$Pr(1 \text{ and } 3) = Pr(1) * Pr(3)$

$= \frac{1}{6} * \frac{1}{6} = \boxed{\frac{1}{36}}$

Live class

eg =) what is the pr of rolling a "5" and then "3"

=) $Pr(5 \text{ and } 3) = Pr(5) * Pr(3)$

$= \frac{1}{6} \times \frac{1}{6} = \boxed{\frac{1}{36}}$

without replacing putting king back?

Dependent Event

eg:- Q) [Pr(K and Q)] ... Q removing cards from deck ← what is Pr of

Q) What is the Pr of removing K and Q?

Bayes theorem

$$Pr(A \text{ and } B) = \boxed{Pr(A) * Pr(B|A)} \leftarrow \text{given } A$$

Naive Bayes — Conditional Events

$$Pr(K \text{ and } Q) = Pr(K) * Pr(Q|K)$$

$$= \frac{4}{52} * \frac{4}{51} \quad Q \text{ given } K$$

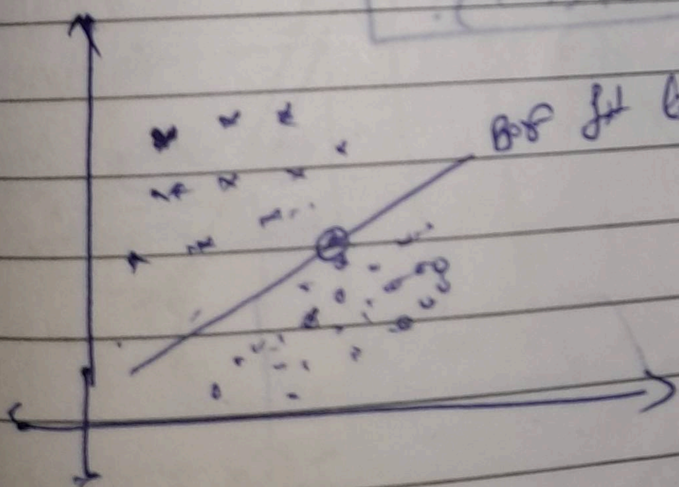
Multiplicative

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

Dependent ⇒ $Pr(A \text{ and } B) = Pr(A) * Pr(B|A)$ ← Conditional Event

Application ML

eg:- Classification problem



ML values will usually in pr like, 0.3/1

30% (Belongs A)

70% (Belongs B)

All this types of scenario are there in ROC, AUC curve, in performance matrix

* Permutation And Combination

• **Permutation:** (order Matters)

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

n = Total no. of

r = Number of records i have

\# All the possible options and over those the order is ever order is basically treated as a separate count.

• **Combination:** (order Does Not matter)

eg → Rock Jim Bred } 1 combination

→ Jim Rock Bred X

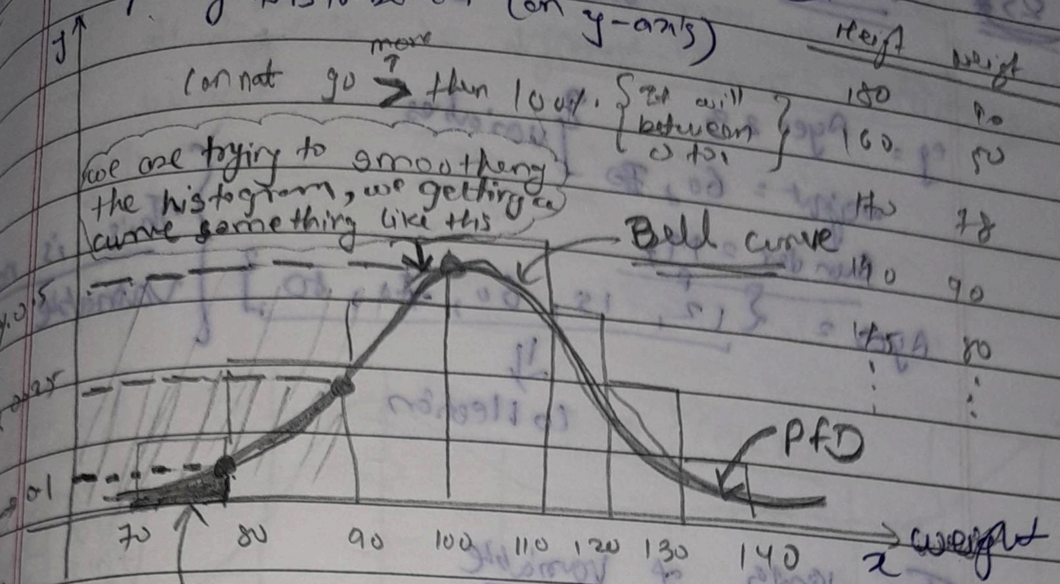
$$\boxed{{}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}}$$

* PDF (Probability Density Function)

% of Distribution (on y-axis)

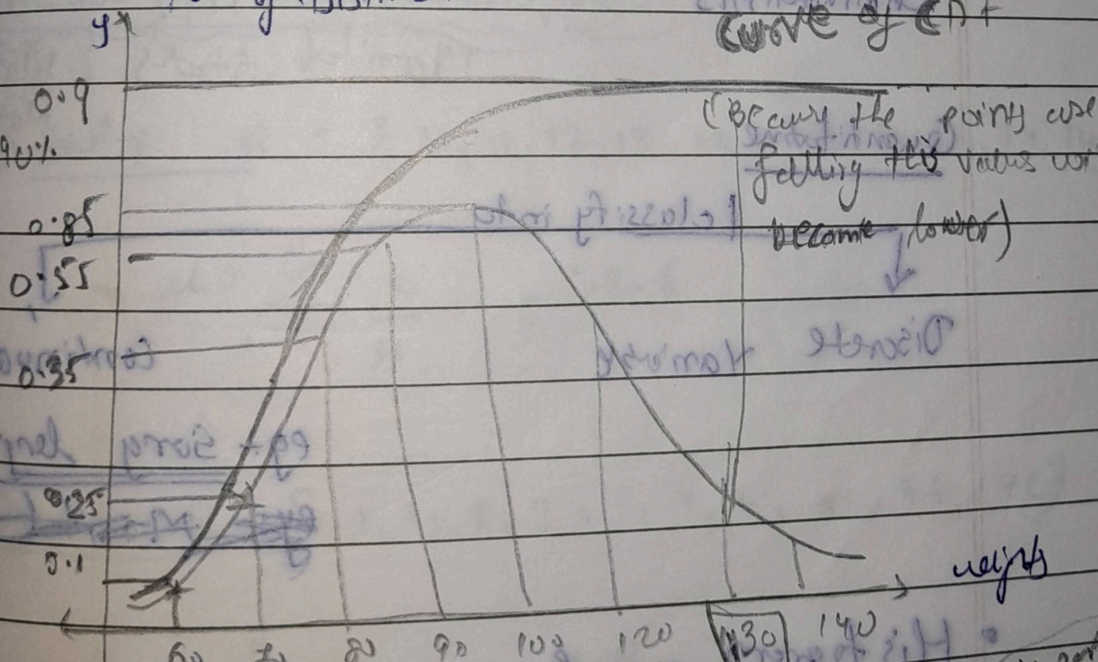
cannot go more than 100%. {It will between 0 to 1}

Height	Weight	O/P
180	90	obese
160	50	slim
170	78	fit
140	90	fit
145	80	slim
⋮	⋮	



% of the Distribution is present in this function {what is the % of Distribution is present in this area, region, % of points that are located in this particular region}

* CDF (Cumulative ~~Distribution~~ Density function)



130kg

(in the particular dataset)

$0.1 + 0.25 = 0.35$

{90% of Distribution is less than 130kg
101 weight are then 130kg.