19th dec '22

Aerofit Review

let's start @ 9:05

Probability → joint

Probability → marginal

Probability → conditional

⇒ Aerofit

KP281 $1500

KP481 $1750

KP781 $2500

1. Age
2. Income ↑
3. fitness ↑
4. Usage ↑
5. Gender

Data → Basic Exploration (df.info, describe

→ Bivariate Analysis

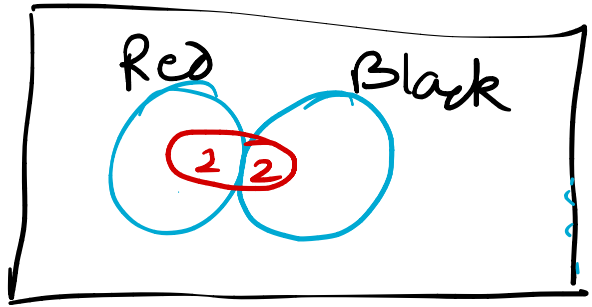
→ Probalities

Target Varible → Product

Marginal	Conditional	Joint
$P(A)$	$P(A \mid B)$	$P(A \cap B)$

$\Rightarrow$ Deck of Cards $\rightarrow$ 52

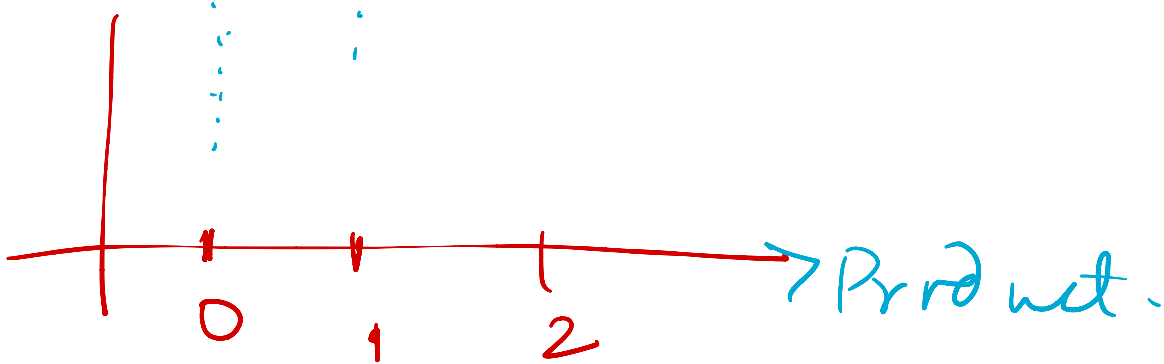
- $P(\text{King}) = 4/52 = 1/13$ ✓ } $\rightarrow$ Marginal Prob.
- $P(\text{King} \mid \text{Red}) = 2/26$ ✓ } Conditional Prob.
- $P(\text{King and Red}) = P(K \cap R) = \dfrac{2}{52}$ } joint Prob.



$$= \frac{2}{52} = P(K \cap R)$$

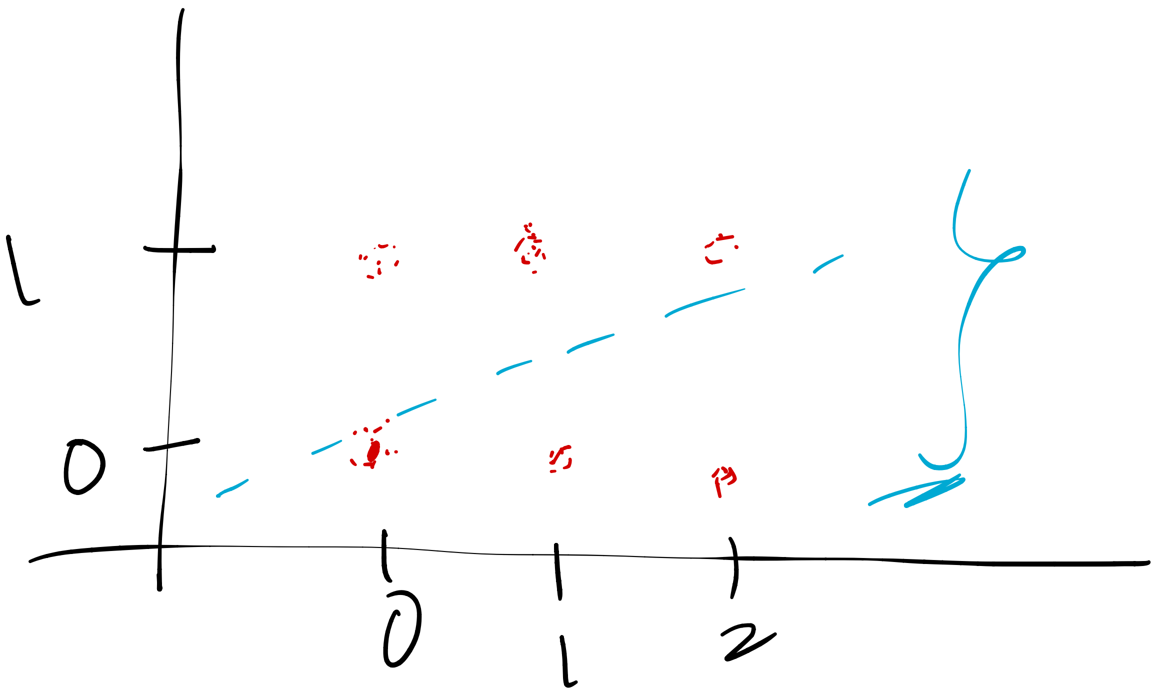
Product [0, 1, 2]

Income [5k, 7k ----]



gender $\rightarrow$ 0, 1

product $\rightarrow$ 0, 1, 2



CLT Central limit theorem

as sample size ↑ ~ Inferring about population with more confidence

→ m ⟶ μ ⟶ true mean

M s (4-5K) Sample & Females (1-2k)

[4.5, 3K - - - - , 2.5]

50M Population

50M

n=100, → Sample size

Total samples = 1000

$S_1 \longrightarrow \mu_1$

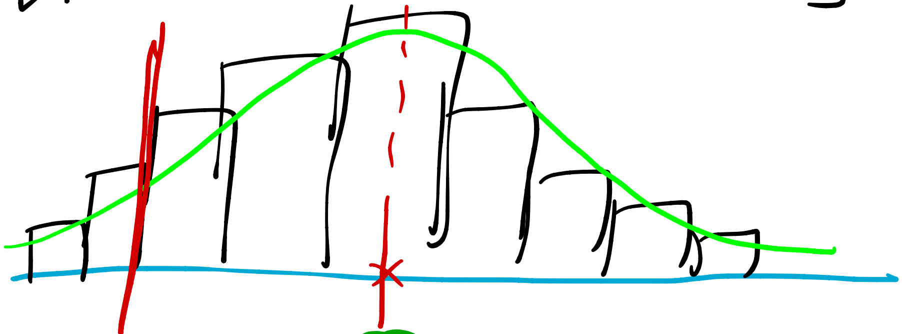
$S_2 \longrightarrow \mu_2$

⋮

$S_{1000} \longrightarrow \mu_{1000}$

means → BOOTSTRAPING

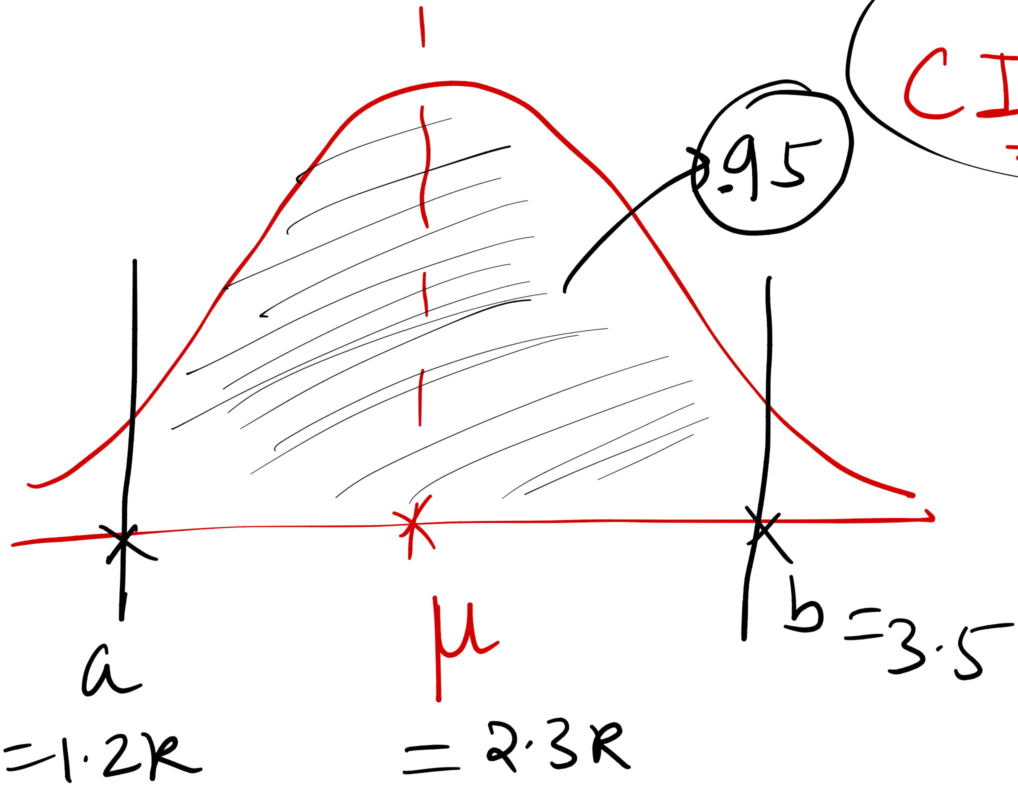
$[\mu_1, \mu_2 - - - - , \mu_{1000}]$



μ

(~ very close to population mean)

50000 → records } → 5M, 50M, 50B



CI ~ 95%

[a, b]

↓

[1.2k, 3.5k]

↓

CI ~ 95%

Males $\xrightarrow[\text{CI}]{95\%}$ [1.2K , 3.5K]

Females $\xrightarrow[\text{CI}]{95\%}$ [0.8K , 2.0K]

mean Amount spent

heights : $[65, 69, 47, 52, \ldots, 75]_{500}$

$^{100}C_r \rightarrow$ Population $\longrightarrow$ CLT

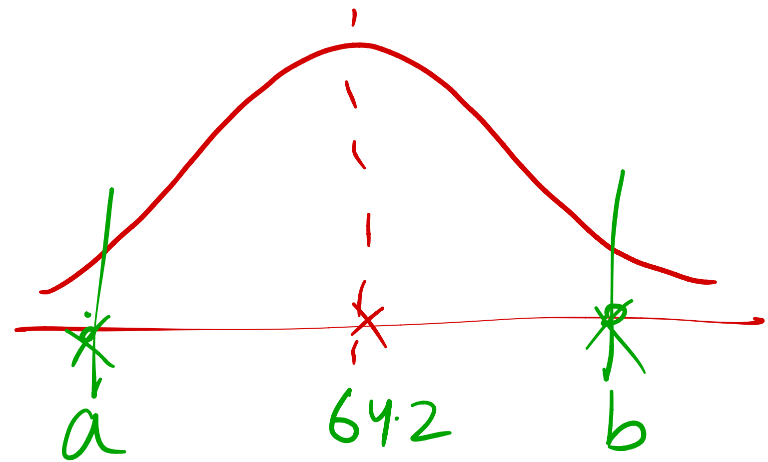
$n = 100,$

$\mu_1 = 63.5$

$\mu_2 = 62.7$

$\vdots$

$\mu_{100} = 65.2$



is Pop mean

with 95% CI

$[60, 66]$