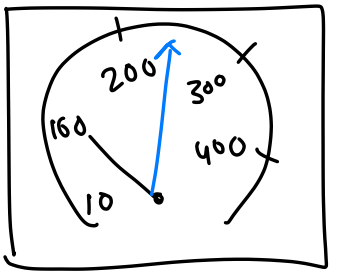
Probabilities

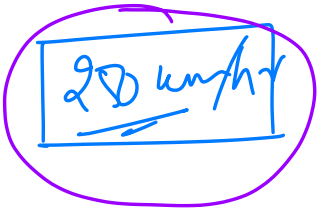
Descriptive Statistics

Statistics

↳ Statistic → A statistic is a number which describes data.

Descriptive Statistics →



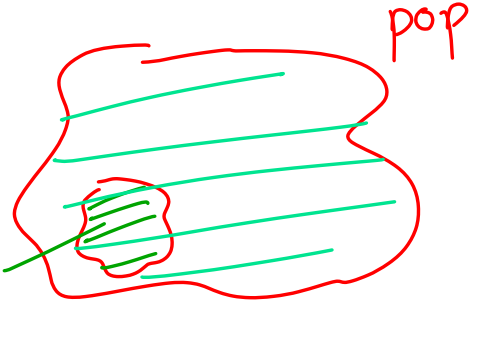


→ Summarising
→ Central tendency, variability

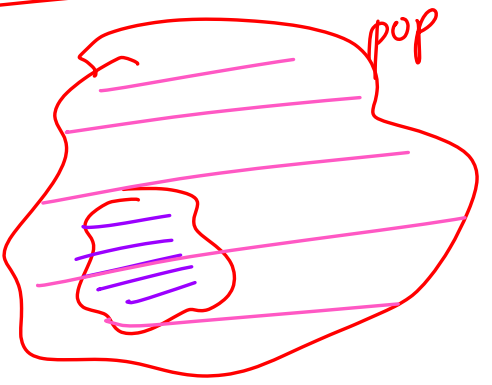
Inferential Statistics → drawing some conclusions

→ 5 feet 7 inches →

Next Module →



EXIT POLL



Glassdoor / Ambitionbox

DS1 Salaries at Google

Sal_lakhs = [30, 30, 35, 40, 40] → Mean, Median

Mean = Sum(Sal_lakhs) / len(Sal_lakhs) = 35

Median = 35

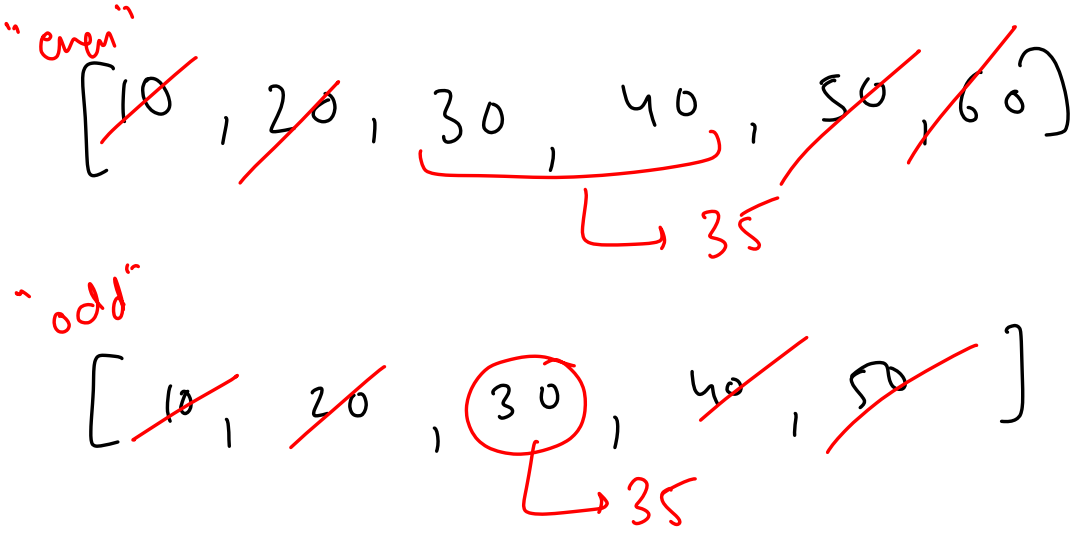
60
80
35
175
5
35

Sal_lakhs = [30, 30, 35, 40, 40, 80]

Mean = 42.5

Median = 37.5

{ medians are robust to outliers }


"even"
[10, 20, 30, 40, 50, 60]
35
"odd"
[10, 20, 30, 40, 50]
35

There are 4 people whose average age is 24.
We know the age of three people: 20, 22, and 28.
What is the median age of these 4 people?

44 users have participated

A	22	18%
B	24	32%
C	25	7%
D	26	43%

End Quiz Now

$A_1 + A_2 + A_3 + A_4 = 24$

$20 + 22 + 28 + x = 24$

4

$x = 26$

$20, 22, 26, 28$

$\frac{22+26}{2} = 24$

$A_1 + A_2 + A_3 + A_4 = Avg$

4

$20, 22, 28, 26$

$20, 22, 26, 28$

mean

$70 + x = 96$

$x = 26$

Mode: Observation with highest frequency.

90, 80, 90, 70, 90, 75, 90 → Mode → 90

90 - 4 70 - 1
80 - 1 75 - 1

90, 90, 70, 70, 80, 75 → Mode → 90, 70

90 - 2 80 - 1
70 - 2 75 - 1

Range: "Max-value - Min Value"

Cricketer Scored Runs

20, 35, 40, 60, 85 = Range = 85 - 20 = 65
(20 = min, 85 = max)

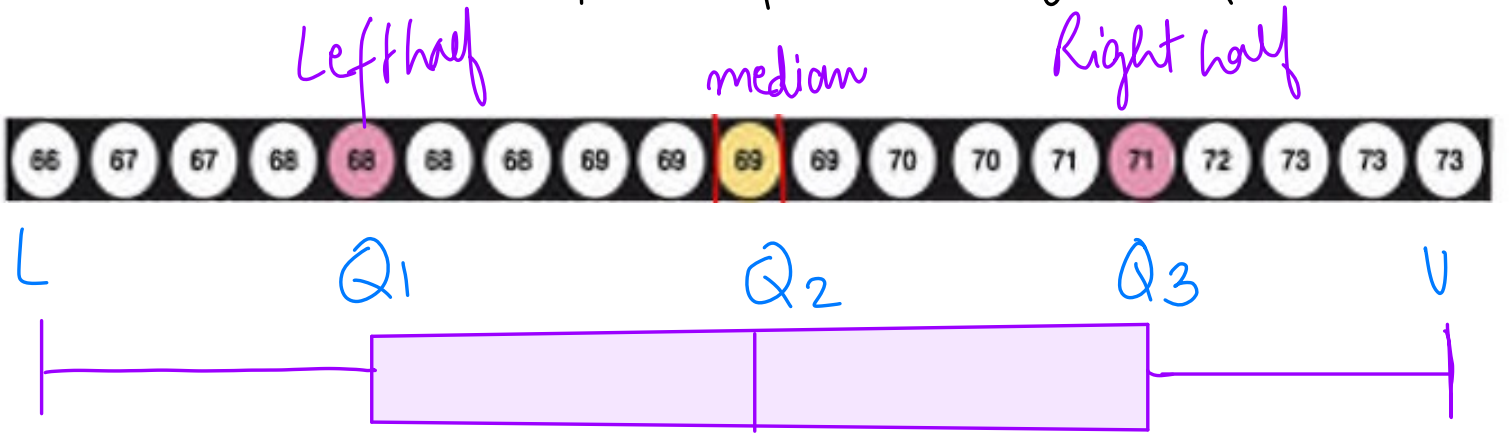
Sehwag 0, 0, 0, 1, 19, 45, 319 ⇒ Range = 319 - 0 = 319

Dravid 30, 40, 50, 60, 32, 49, 52 ⇒ Range = 30

InterQuartile Range: $Q3 - Q1$

$L = Q1 - 1.5(IQR)$

$U = Q3 + 1.5(IQR)$



$IQR = Q_3 - Q_1 = 71 - 68 = 3$

$L = Q_1 - 1.5(IQR) = 68 - 1.5(3) = 63.5$

$U = Q_3 + 1.5(IQR) = 71 + 1.5(3) = 75.5$

Schwag — "Aggressive"

David — "Wall"

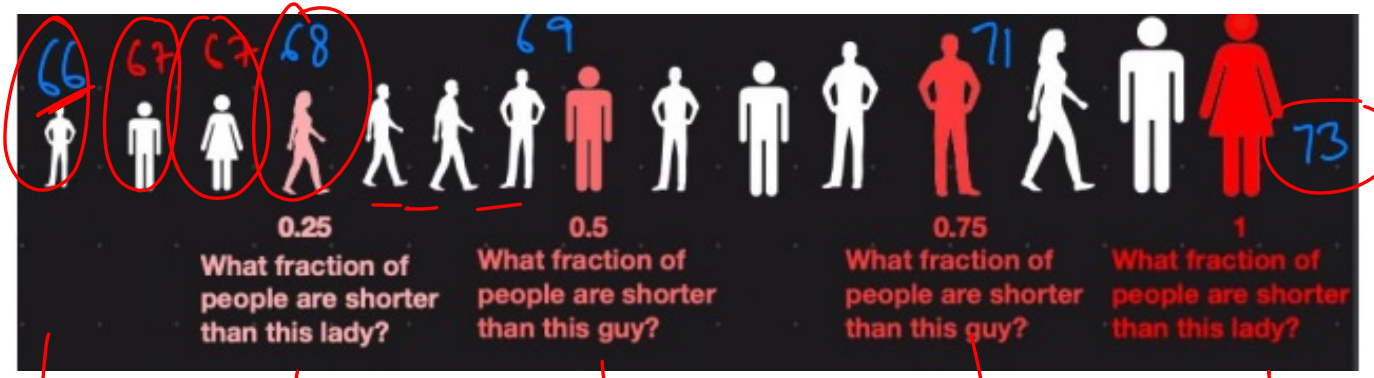
Quantile: A value which tells us that "q" observations are less than this value

Quartile: Quartile when Q is set to Quartile = "4"

$q = 0.25 \times \text{len(data)} \rightarrow 1^{st}$ Quartile

$= 0.5 \times \text{len(data)} \rightarrow 2^{nd}$ Quartile

Percentile: A value which tells us "p%" observation are less than that value.



min

Q_1

Q_2

Q_3

max

density

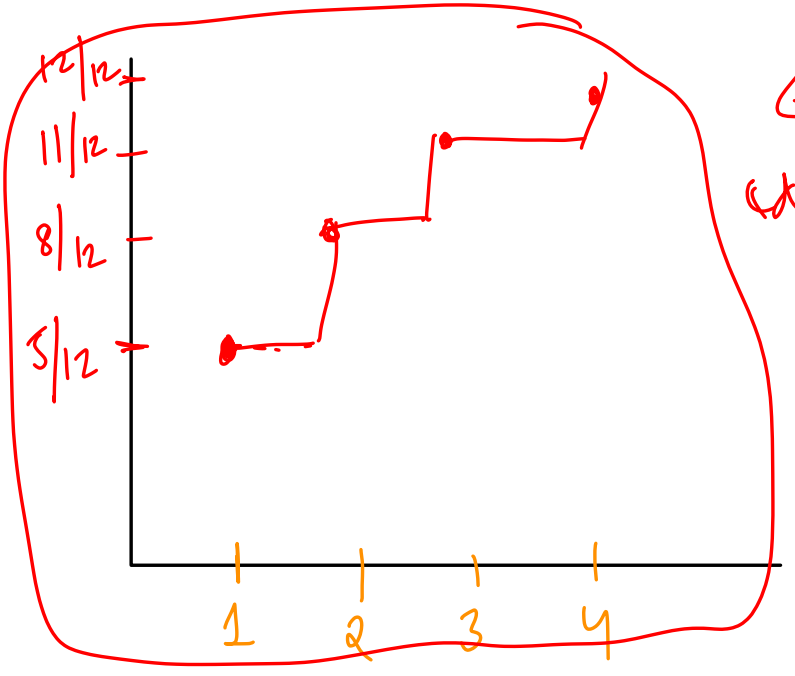
fract^n

pdf → cdf

pmf → cdf

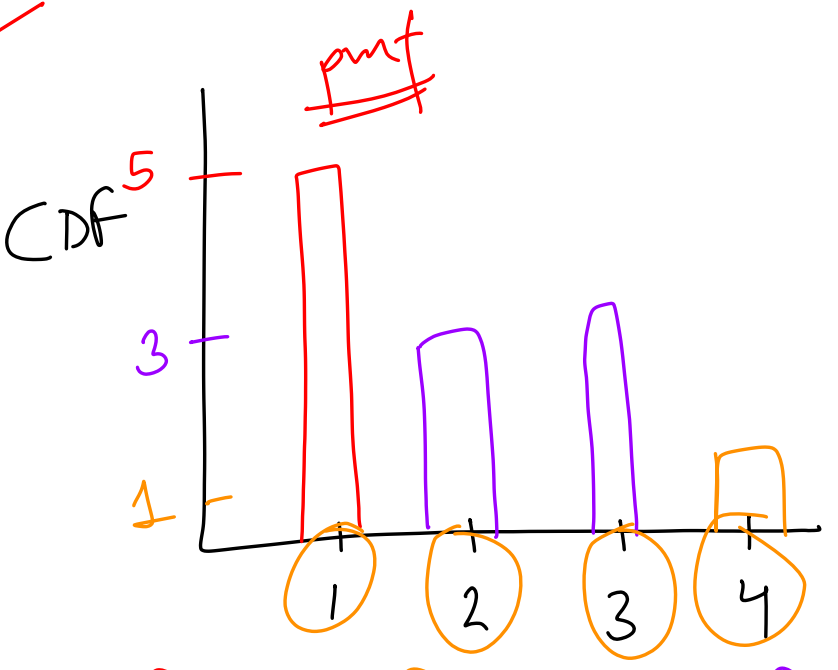
Histograms

distribution plot



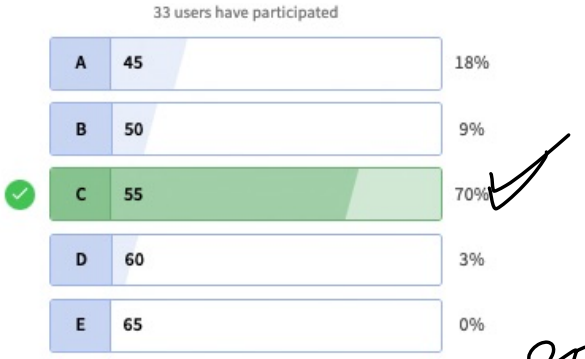
cdf

pmf



1, 2, 3, 1, 2, 3, 1, 1, 3, 4, 1, 5, 2

The mean weight of 2 children in a family is 40 Kgs. If the weight of the mother is included, the mean becomes 45. What is the weight of the mother?



$$\frac{x_1 + x_2}{2} = 40$$

$$x_1 + x_2 = 80$$

80

$$\frac{x_1 + x_2 + M}{3} = 45$$

$$\frac{80 + M}{} = 135$$

$$M = 55$$

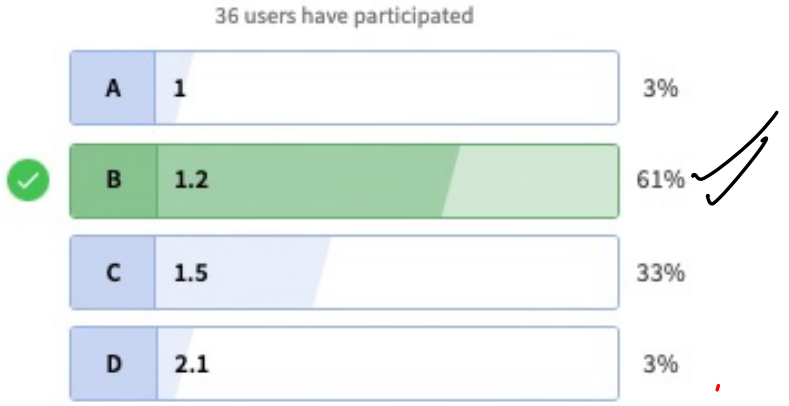
Quiz time!

Quiz Ended!

In a survey about favorite animals, 30 people said cat, 40 people said dog, 20 people said cow. What is the mode of favorite animals in this data?

39 users have participated

A survey of the number of pets in a town saw that 30% people had 0 pets, 40% had 1 pet, 10% had 2 pets, 20% had 3 pets. What is the average number of pets?



odd / km

100

0+0+0+0+0 (30 times) + 1+1+1.. (40 times) + 2+2+2+2.. (10 times) + 3+3+3.... (20 times)

40+20+60 / 100 = 120/100 = 1.2

100

S_1 40 } 40
S_2 40 }
S_3 (30) 20
S_4 45 } 40
S_5 45 }

$$\frac{40+40+30+45+45}{5} = 40$$

40 + 30 + 45
40 20 40

$S_1 \rightarrow$ 40 x 2
30 x 1
45 x 2

40 --- (2) + 30 x 1 (1) + 45 ++ (2)

$$\frac{80+30+90}{5} = 40$$

#pets	weightage	
0	30	0.3
1	40	0.4
2	10	0.1
3	20	0.2
Values	weightages	

weighted avg

$$\Rightarrow \frac{(0\times 30)+(40\times 1)+(10\times 2)+(20\times 3)}{100}$$

$$\frac{0+40+20+60}{100} = \frac{120}{100}$$

$$= 1.2$$

2. Original Salaries : 30, 32, 35, 35, 38 LPA

Which metrics will remain unchanged under the effect of addition of 5 LPA bonus to each of the above salary ?

Subtraction

Addition

37 users have participated

A	Mean, Median	19%
B	Median, Mode	35%
C	Range, IQR	32%
D	Mode, Range	14%

35, 37, 40, 40, 43

⇒ 39 Mean

⇒ 40 Media

30, 32, 35, 35, 38

Q_1 Q_2 Q_3

Mean - 34

Median - 35

Range - 8

IQR - 3

Mode = 35

⇒ Range = 8

X

IQR = 3

Mode = 40

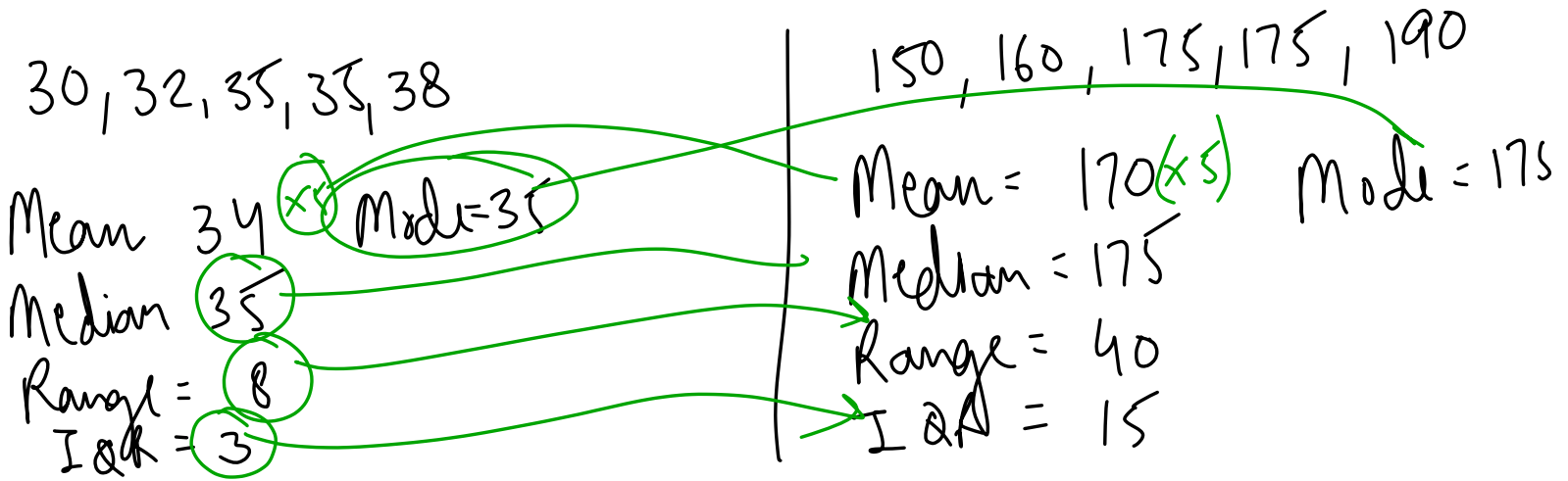
3. **Original Salaries : 30, 32, 35, 35, 38 LPA**

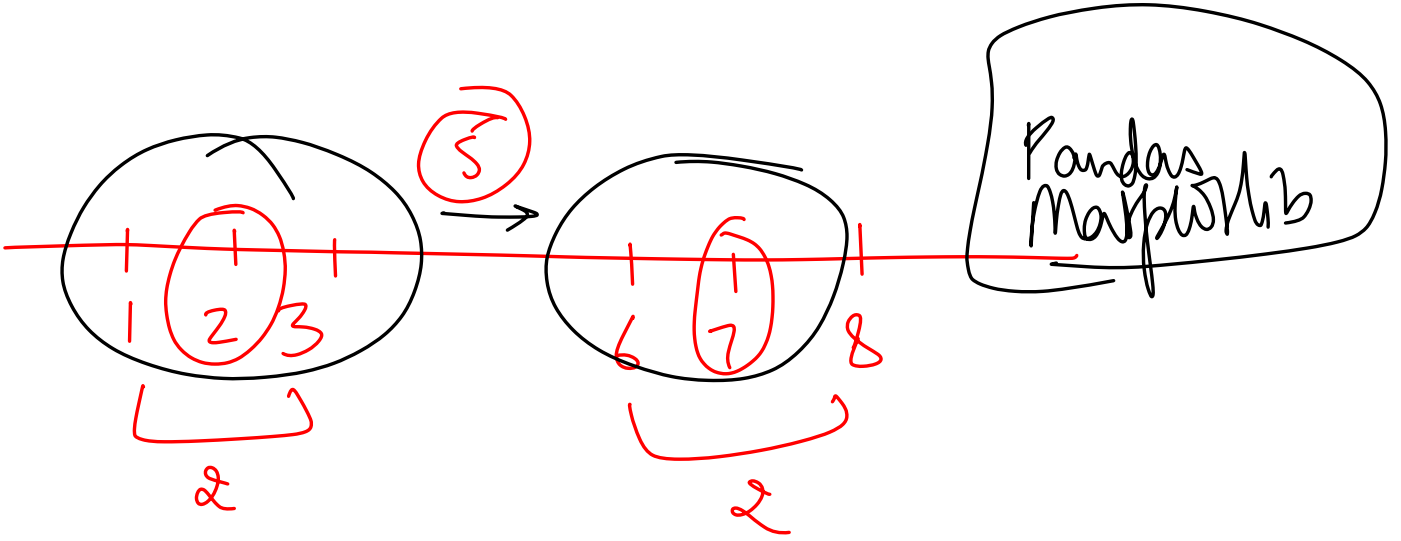
Which metrics will remain unchanged under the effect of multiplication by 5 to each of the above salary ?

division

36 users have participated

A	Mean, Median	3%
B	Median, Mode	11%
C	Range, IQR	33%
✓ D	None	47%
E	IQR, Median	6%

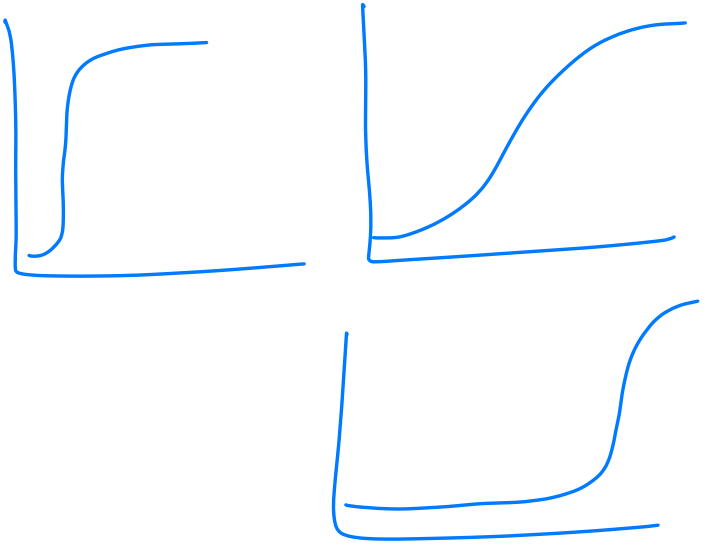


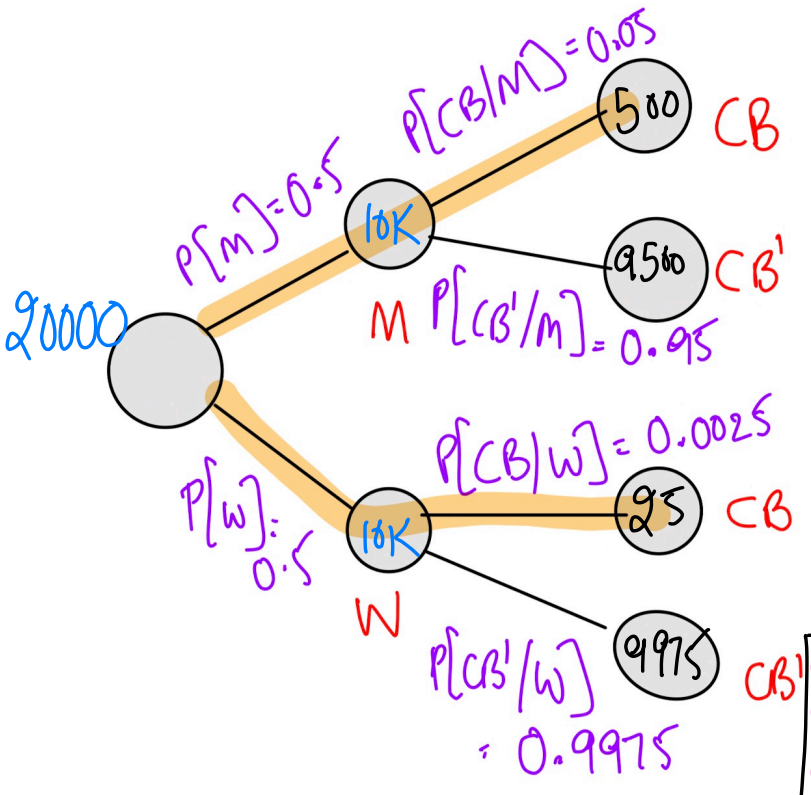


Oct → 12 → 2 (18) = 20

Sept → 4 pm → 16 (13) = 29

4 pm →





$$P[CB/M] = \frac{500}{10000} = 0.05$$

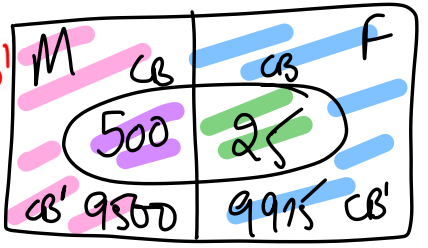
$$P[M/\underline{\underline{CB}}] = \frac{500}{500+25}$$

$$P[CB \cap M] = \frac{500}{20100} = 0.025$$

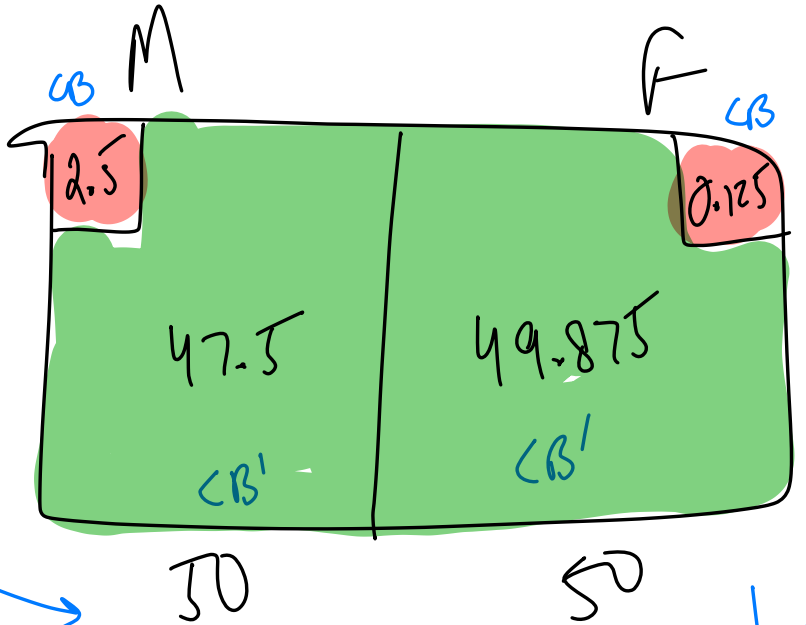
$$= P[CB/M] \cdot P[M]$$

$$= \frac{5}{100} \times \frac{50}{100} =$$

$$= 0.025$$



Suppose 5 % Men and 0.25% Women are color blind. A randomly colour blind person is chosen. What is the prob that this person is a male? Assume same number of males and females.



$2.5 \times 0.05 = 0.125$

$P[CB/M] = \frac{2.5}{50}$

$P[M/CB] = \frac{2.5}{2.5+0.125}$

$P[CB/F] = \frac{0.125}{50}$

$P[F/CB] = \frac{0.125}{2.5+0.125}$

$P[CB \cap M] = 2.5\% = P[CB/M] \cdot P[M] = \frac{2.5}{50} \times \frac{50}{100}$