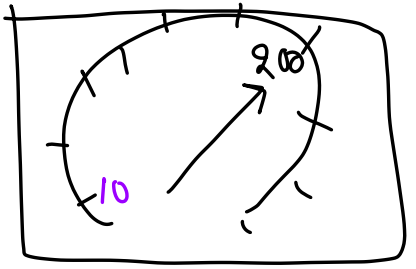
Statistics

↳ Statistic → A number that describes my data.

Descriptive Statistics



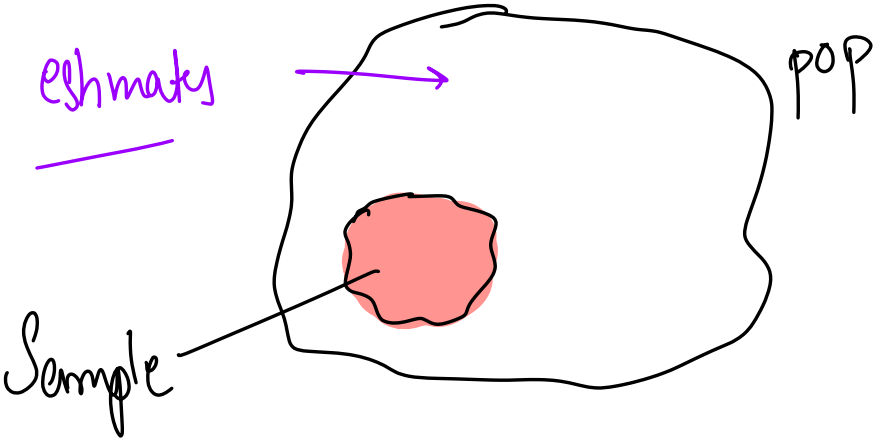
→ Means, Median, Mode.

→ Data → Central Tendency, Vaucance, Std.

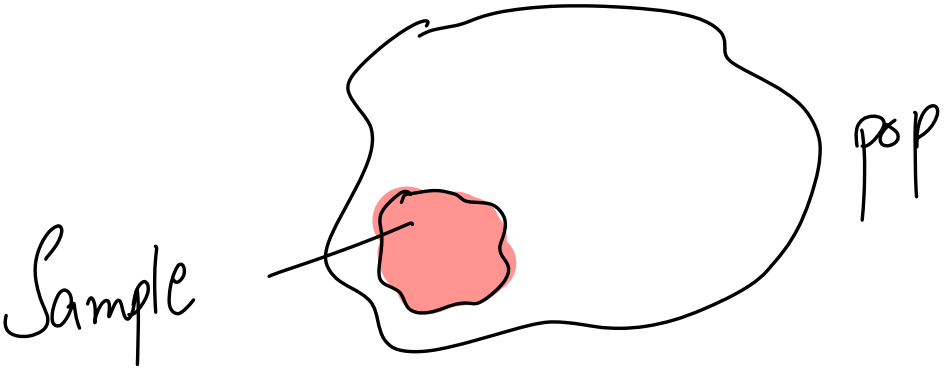
} → Summarizing the data

Inferential Statistics

Avg. Height 5'7".



EXIT POLLS



Glassdoor / Ambition Box

DS1 Salaries at Google

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

→ Mean = 35

→ Median = 35

→ Mode = 30, 40

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

Median = 35

Startup [100]

Sal_lakhs [30, 30, 35, 40, 40, 100]

→ Mean = 45.83

→ Median = 37.5

→ Mode = 30, 40

"Medians are more robust outliers".

even

[10, 20, 30, 40, 50, 60]

→ 30+40 / 2 = 35 → Median

odd

[10, 20, 30, 40, 50]

30 → median

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?

59 users have participated

A	22	8%
B ✓	24	44%
C	25	15%
D	26	32%

$$\frac{A_1+A_2+A_3+A_4}{4} = 24$$

$$\frac{20+22+28+A_4}{4} = 24$$

$$70 + A_4 = 96$$

$$\boxed{A_4 = 26}$$

20, 22, 26, 28

$$\frac{22+26}{2} = \frac{48}{2} = 24$$

Mode: Observation with highest freq.

90, 80, 90, 70, 90, 75, 90 } Mode ⇒ 90

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

Unimodal

90, 90, 90
75, 80, 75
75, 70, 42
70, 42, 70

90 - 3 42. 2
75 - 3
70 - 3
80 - 1

90, 90, 70, 80, 70, 75 } — Modes ⇒ 90, 70

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

Bimodal

90, 70, 80, 75 } Nomodes

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

Range : "Max value" − "Min value"

Sehwag → "aggressive"

0, 0, 0, 1, 50, 70, 100, 219

Range : 219 − 0 = 219

Mean : 440 / 8 = 55

Median : 25.5

Mode : 0

Dravid → "The great wall"

10, 28, 40, 45, 50, 52, 68, 70, 100

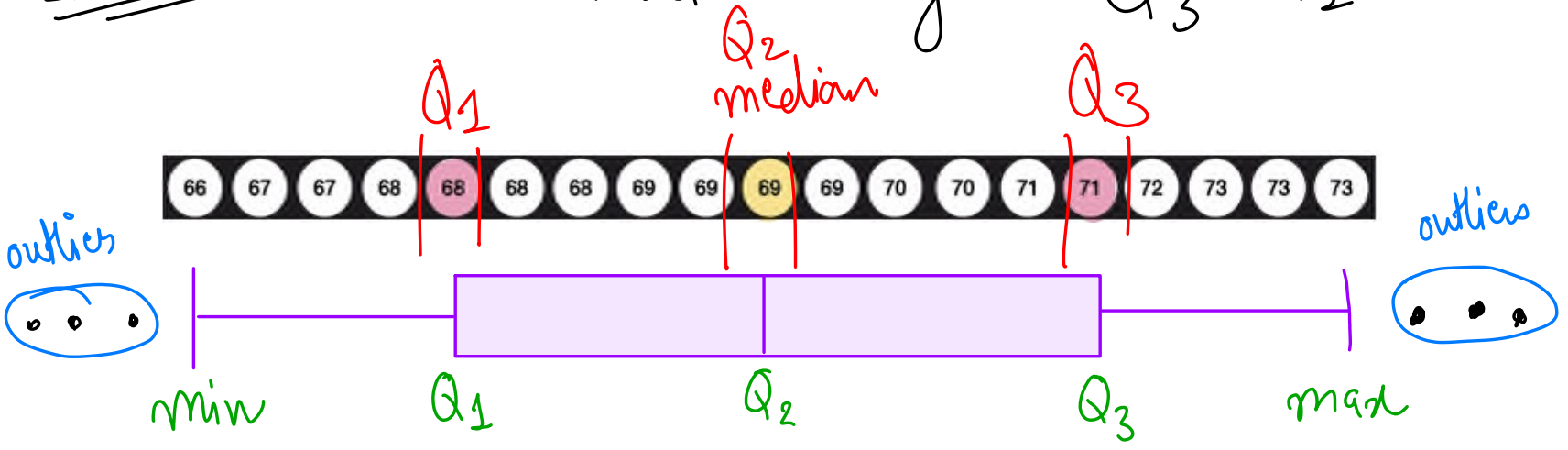
Range = 100 − 10 = 90

Mean = 463 / 9 = 51.44

Median = 50

Mode = No mode.

IQR → Inter Quartile Range = $Q_3 - Q_1$

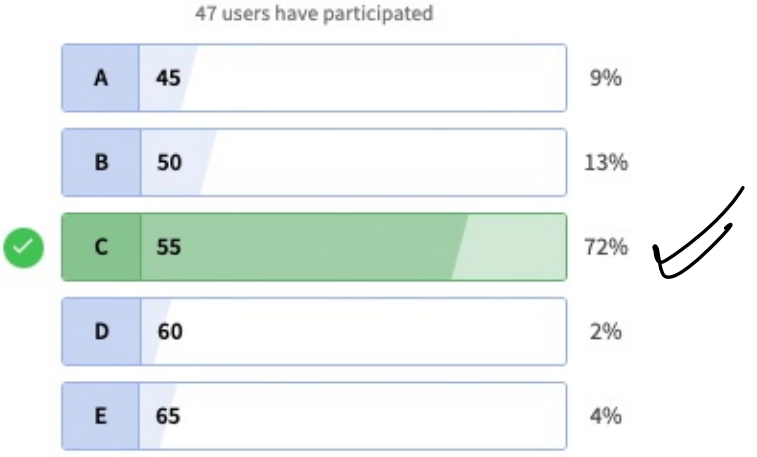


$IQR = Q_3 - Q_1 = 71 - 68 = 3$

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

$U = Q_3 + 1.5(IQR) = 71 + 4.5 = 75.5$

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?



$W_1 + W_2 = 40 \times 2$

$W_1 + W_2 + M_1 = 45 \times 3$

$M_1 + 80 = 135$

$M_1 = 55$

Quiz time!

TIME LEFT: 0 Secs

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?

48 users have participated

A	Cat	4%
B	Dog	96%
C	Cow	0%

Quantile: A value which tells us "q" observation are less than q value.

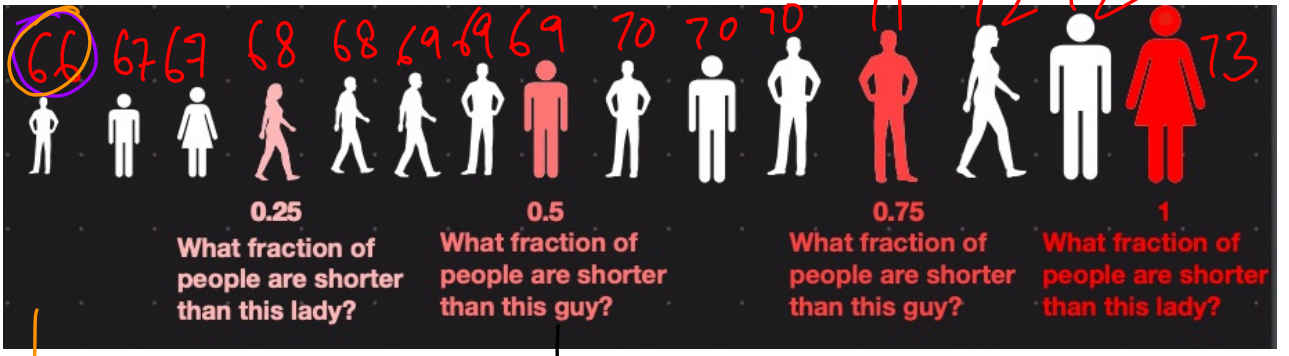
Quartile: Quantile when Q is Quarter → Q = 4

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

$0.50 \times \text{len(data)} \rightarrow 2^{nd}$ quartile

Percentiles: A value which tells us "p%" observations are less than this value. (10th → 10%)

↳ value → D

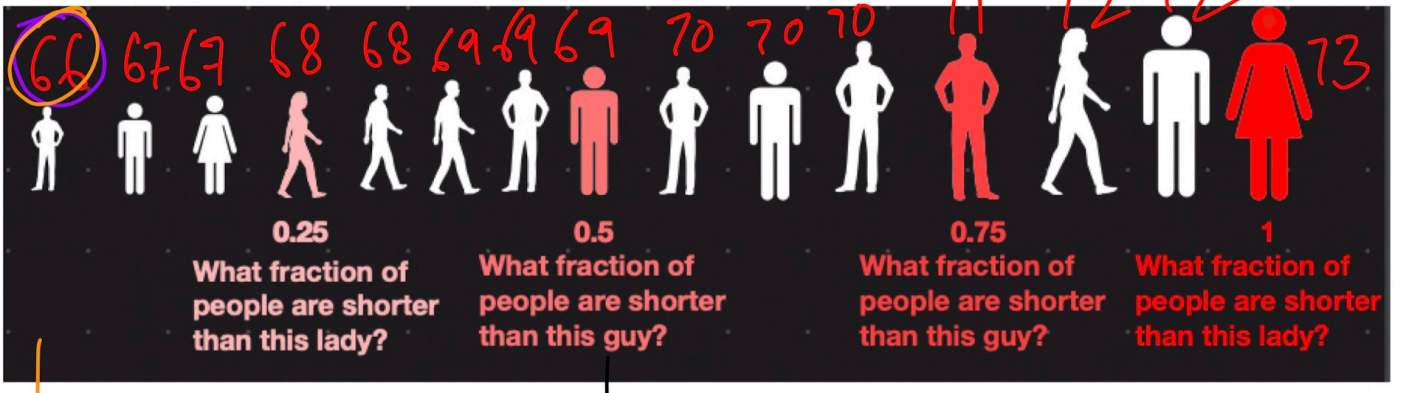


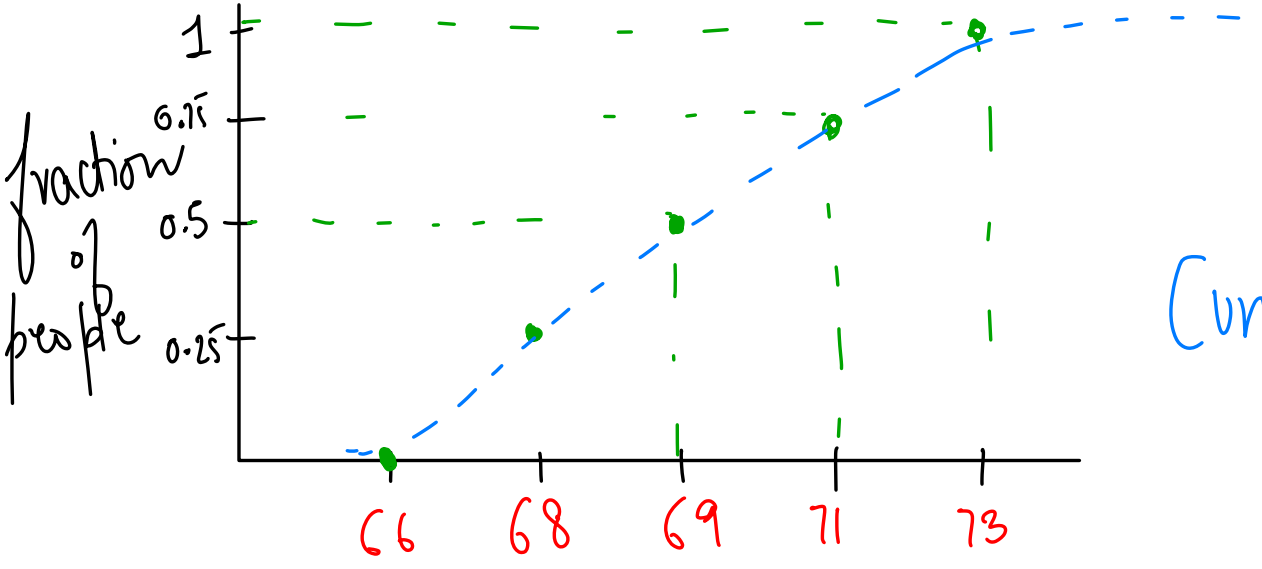
0th percentile

50th percentile → 69

50% of data is less than 69

100th percentile → Max value





CDF

Cumulative density function

Variable

DISCRETE

PMF → Prob Mass function

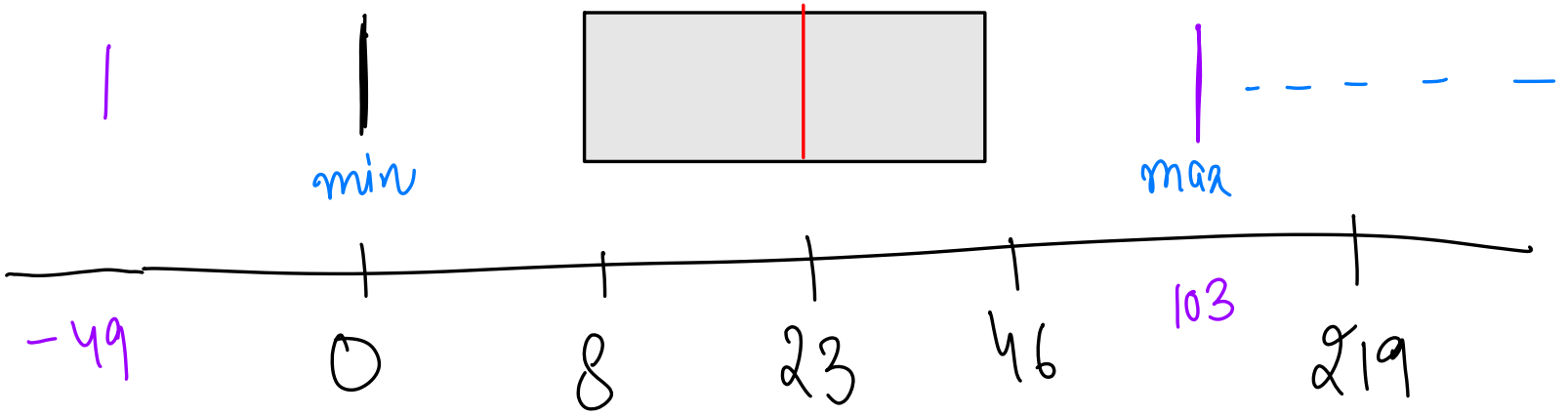
CONTINOUS

PDF → Prob. density function

KDE

CDF →

CDF →



$IQR = 38$

$L = -49$

$U = 103$

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?

35 users have participated

A	1	6%
B ✓	1.2	60%
C	1.5	34%
D	2.1	0%

✓

100 → 30% 40% 10% 20%

pets		
0	0.3	30
1	0.4	40
2	0.1	10
3	0.2	20

weighted Avg:

$$\frac{0\times0.3+1\times0.4+2\times0.1+3\times0.2}{0.3+0.4+0.1+0.2}$$

$$= \frac{0+0.4+0.2+0.6}{1}$$

$$= 1.2$$

$$\text{Avg} = \frac{\underbrace{0+0+0+0+0}_{30\text{ times}} + \underbrace{1+1+1+\cdots}_{40\text{times}} + \underbrace{2+2+2\times 2+2}_{10\text{times}} + \underbrace{3+3+3+3}_{20\text{ times}}}{100}$$

$$\text{Avg} = \frac{0+40+20+60}{100} = 1.2$$

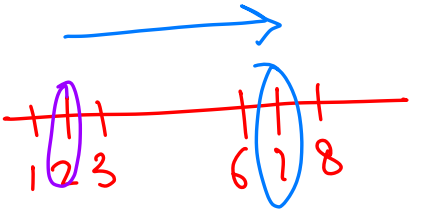
2. **Original Salaries : 30, 32, 35, 35, 38 LPA**

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

39 users have participated

A	Mean, Median	21%
B	Median, Mode	21%
✓ C	Range, IQR	38%
D	Mode, Range	21%

2 Lakh



[30, 32, 35, 35, 38]

+5 +5 +5 +5 +5

35, 37, 40, 40, 43

Mean = 34 (+5)

Median = 35 (+5)

Mode = 35 (+5)

Range = 8 X

IQR = 3 X

Mean = 39

Median = 40

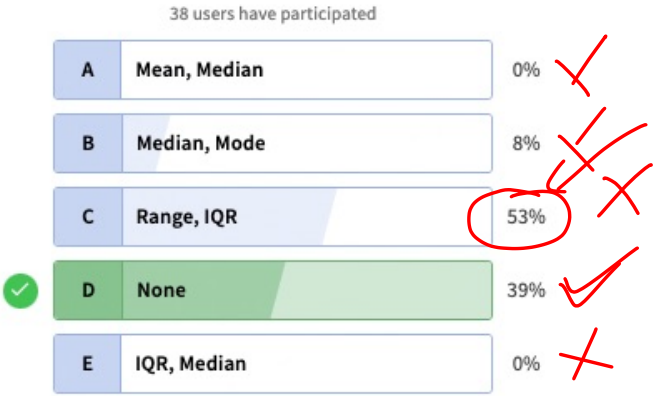
Mode = 40

Range = 8

IQR = 3

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 ?



30, 32, 35, 35, 38

×5 ×5 ×5 ×5 ×5

150 160 175 175 190

Mean = 34 (×5) Mean = 170

Median = 35 (×5) Median = 175

Mode = 35 (×5) Mode = 175

Range = 8 (×5) Range = 40

IQR = 3 (×5) IQR = 15