

30th November, 2022.

DSML: CC Maths

Probability 5 - Descriptive Statistics

Recap: (a) Probability theory .
(b) Bayes' theorem .
(c) Combinatorics .

Today : (a) Statistics : Working with data .

- (b) Mean, median, mode .
- (c) Percentiles, Quartiles, IQR
- (d) CDF
- (e) Variance .

Class starts
@
9:05 p.m.

Statistics

Descriptive: summarize data.

Central tendency, variability.

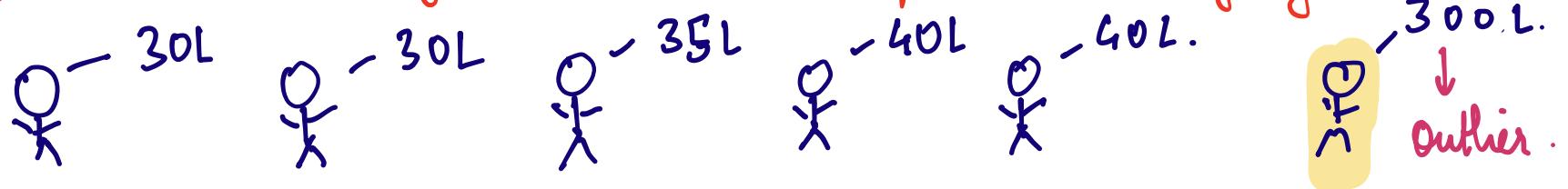
- 1] You are driving @ 65 km/h.
- 2] A \rightarrow 70% vote share.

Inferential Statistics: Draw conclusions.
Confidence intervals, Hypothesis testing
regression.

- 1] You will reach home in \approx 30 mins.
- 2] Exit poll:
A \rightarrow 70% vote share.



Glassdoor / lewls.tji → Salaries for DS @ Google.



Mean: $\frac{(30 + 30 + 35 + 40 + 40)}{5} = 35L$.

Median: 35 L.

Mean: $\frac{(30 + 30 + 35 + 40 + 40 + 300)}{6} = 79L$.

Median: $\frac{35 + 40}{2} = 37.5L$.

Takeaway: Median is more robust to outliers.

Median

10, 20, 30, 40, 50, 60, 70.
0 1 2 3 4 5 6. → odd.

Median.

$$\begin{array}{ccccccccc} 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80. \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \rightarrow \text{Even.}$$



$$\frac{40+50}{2} = \underline{\underline{45}}$$

Quiz.

Quiz 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?

$$\frac{20 + 22 + 28 + \underline{x}}{4} = 24.$$

$$x = 24 \times 4 - (20 + 22 + 28) \\ = 96 - 70 = \underline{26}.$$

$$20 \quad \underline{22} \quad \underline{26} \quad 28. \\ \frac{22 + 26}{2} = \underline{\underline{24}}$$

Mode

90, 90, 90, 80, 90, 70, 95, 90.

90 - 5

80 - 1

70 - 1

95 - 1

Mode : 90.

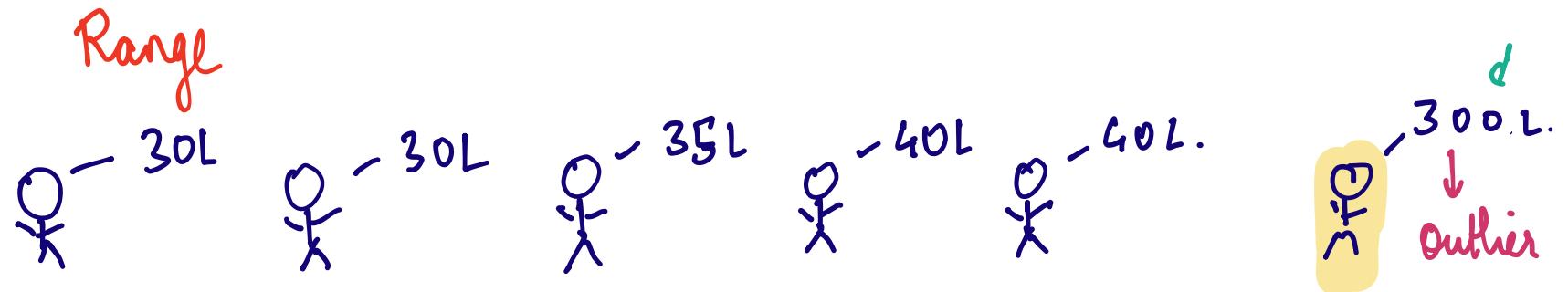
2, 2, 3, 3, 5

2 : 2

3 : 2

5 : 1

Bimodal dataset.
two modes.



Range : $\frac{\text{Max value}}{300L} - \frac{\text{Min value}}{30L}$

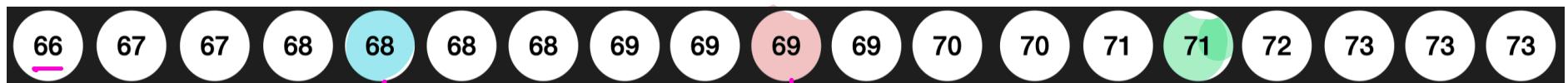
 $= \underline{270L}$

20, 25, 60, 100.

$$100 - 20 = 80.$$

Simply calculating
the range is not
always informative.

Percentiles, Quartiles, Inter Quartile Range (IQR)



Q_1 : First quartile

It is that value such that 25% of the data is less than or equal to it.

$P_{25} \rightarrow$ Percentile-25.

Median.

Q_2 : Second Quartile.

It is that value such that 50% of the data is less than or equal to it.

$P_{50} -$ Percentile 50

$Q_3 - Q_1$

P_{75}
↓
Percentile 75.

$$71 - 68$$

=

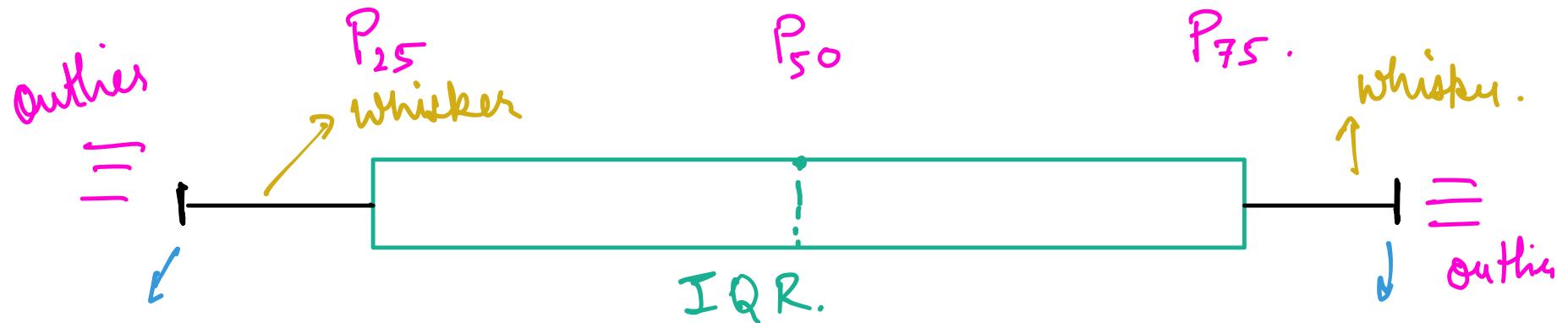
3.

* Percentile : P_x : It is that value such that $x\%$ of the data is less than it



Box Plot :

$P_0 \rightarrow 0\%$ of the data is less than this.



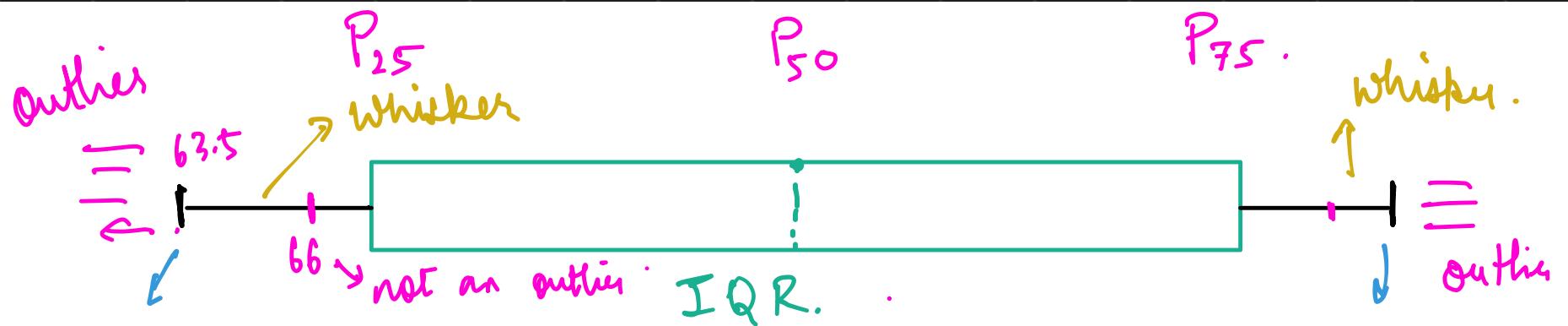
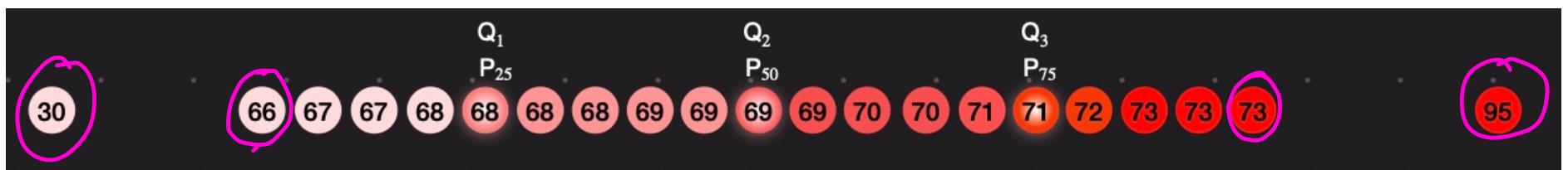
$$Q_1 - 1.5 * \text{IQR}$$

$$68 - 1.5 \times 3 \\ = 63.5$$



$$Q_3 + 1.5 * \text{IQR}$$

$$71 + 1.5 \times 3 \\ = 75.5$$



$$Q_1 - 1.5 * \text{IQR}.$$

$$68 - 1.5 \times 3 \\ = 63.5$$

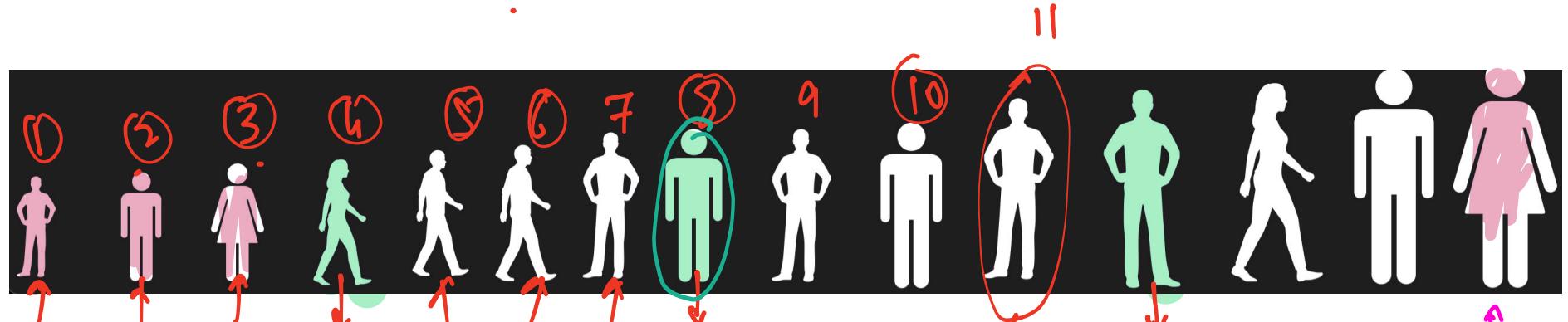
$$66 > 63.5$$

$$Q_3 + 1.5 * \text{IQR}$$

$$71 + 1.5 \times 3 \\ = 75.5$$

$30 \rightarrow$ Outlier

Cumulative Distribution Function (CDF).



↑
5 feet.

P_{25}

Median.

P_{75} .

↑
7 feet

CDF.

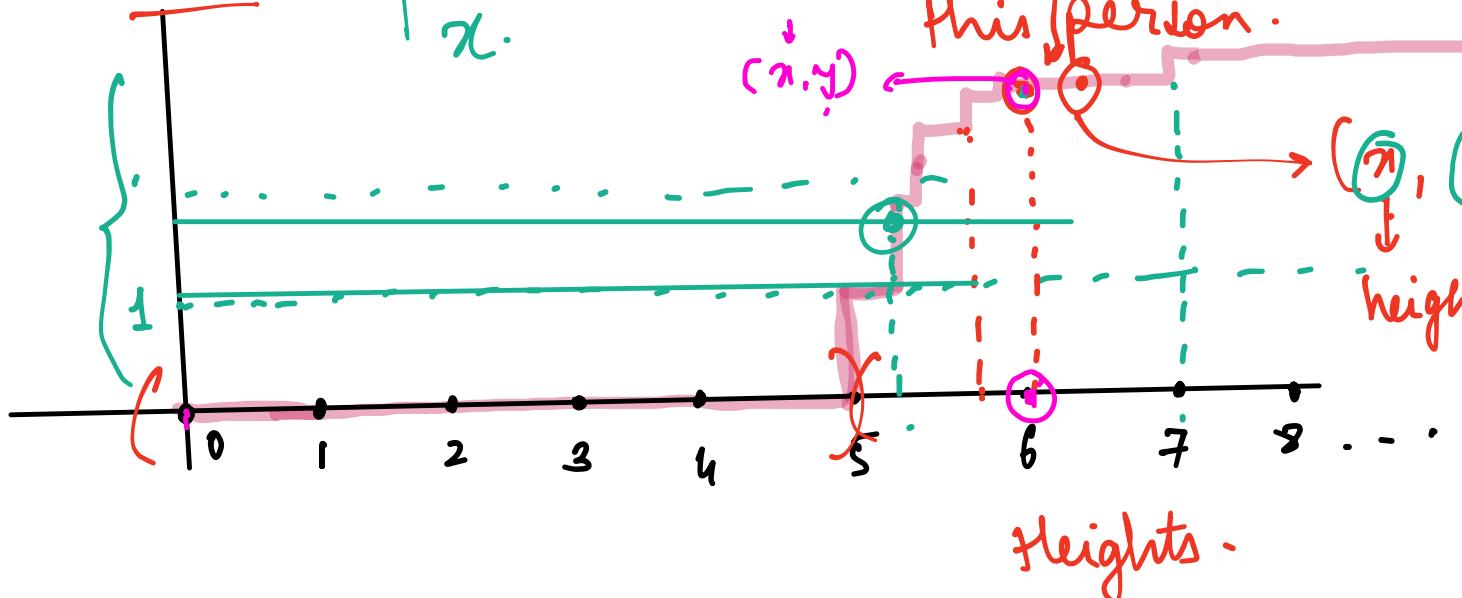
P_x .

P_{50} : 50% of the people are less than this person.

(x, y)

(x, y)
height

of people whose height is $\leq x$.



Simple Arithmetic

* Mean.	34.	→ 39	→ 59.8.
* Median	35.	→ 40	→ 61.6.
* Mode	35.	→ 40	→ 61.6.
* Q ₁	31	→ 36..	→ 54.56
* Q ₂	35	→ 40.	→ 61.6.
* Q ₃	36.5.	→ 41.5;	→ 64.24. × 1.76.
{ * Range	8.	→ 8 (not affected).	→ 14.08.
{ * IQR.	5.5	→ 5.5 (not affected).	→ <u>9.68.</u> × 1.76.

$$(30, 32, 35, 35, 38) \xrightarrow{+5L} (\underline{35, 37, 40, 40, 43}).$$

→ × 1.76 → (52.8, 56.32, 61.6, 61.6, 66.88).

H.W. (Show): ① What method is used in np. percentile, and how does it relate to the method discussed in class?

② Boxplot method for IQR.

Q] Given 5 children, 

(a) $P(\text{exactly 3 boys}) = \frac{10}{32}$

(b) $P(\underbrace{\text{at least 2 girls}}_{\text{}})$.

{ } } \rightarrow 32 possibilities.

Count possibilities s.t. 3 Boys.
Permutations of these. ↓
 $\rightarrow B B B G G$. \rightarrow $\frac{5!}{2^2} = 10$. ${}^5C_3 = \frac{5!}{3! \cdot 2!}$

⑥ BBB GG.

⑦ BB G BG

⑧ B B GG B.

⑨ B G B GB.

⑩ B GG BB.

⑪ G BG BB.

⑫ G GB BB.

⑬

⑭

⑮

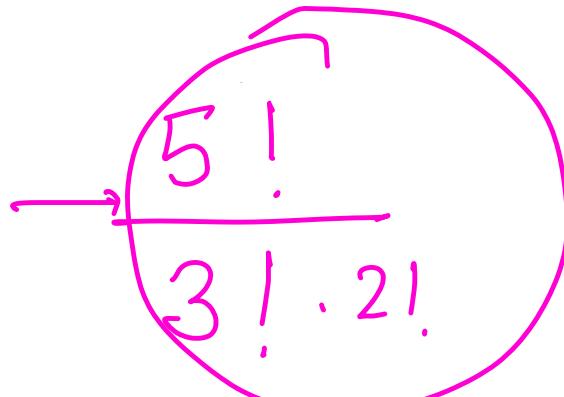
$B_1 \ B_2 \ B_3 \rightarrow 3!$

$G_1, G_2 \rightarrow 2!$

$B, B, B, G, G.$

$\underline{B, B, B, G, G}.$

5C_3



$B \ B \ B \ - G \ G.$

5C_3

1 → 0 girls.

? → 1 girl?

5

G B B . B . B

B G B B B

.

!

$$1 - \frac{6}{32} = \frac{32 - 6}{32} = \frac{26}{32},$$

