

19-07-2025

Agenda:

Statistics - I ← python
← numpy (computation)

Theory
+
few practicals
+
more code

P.J. Create a class called
calc.

- add → multiply
- subtract → inherit more advancd
- dir func() from other class
called ABCDE.

visualise data,
to to → answer
→ why



Statistics - 2025

P.J. Data →

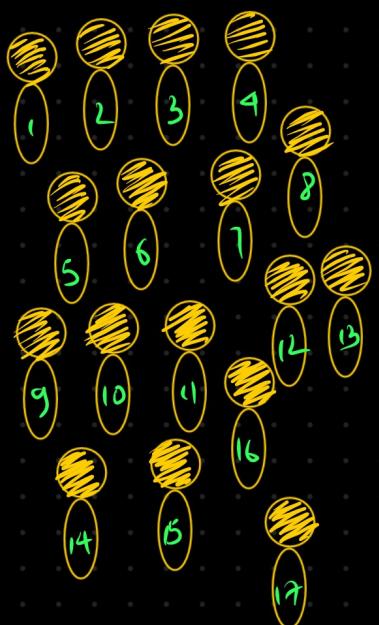
Step →

Definition:

Statistics is a domain of Mathematics / Science where we focus on collecting, organising, analyzing, and interpreting data to make decision.

- Avg height in the class?
- Avg age in the class?
- what is the chance it will rain tomorrow?
- If i do more study, will i get better grades?.

Key Terms



total - student → 17 (population)

Small part of population → 5 (sample)

Parameter - No. Des population (Avg height)

Statistic - No. Des sample (sample Avg height)

Variable - Something we measure (height, weight)

observation - one data point (height)

Data - collection of observation
(height, w, g)

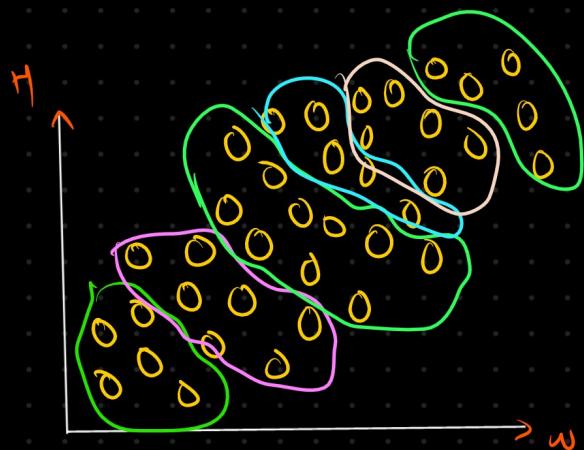
P.J. (clothing) brand



whole earth's
population!

divide → 6

groups / clusters



(1) → → 1000 people

(2) , , →

population is large → parameter is hard to
find

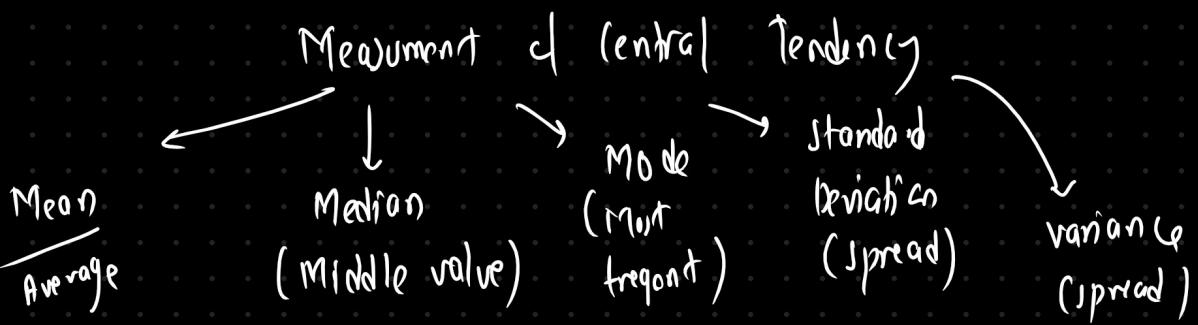
↓
sample & statistic

→ Bigger the sample, better the estimate.

→ Variable (2 types): —

* Qualitative (categorical) → Gender, City, Class, subject, email - ^{spam}

* Quantitative (Numerical) → Age, Height, marks, temp, How much!, bat% per c, salary, distance



→ Mean

$$\begin{array}{c}
 \textcircled{A} \quad \textcircled{B} \quad \textcircled{C} \quad \textcircled{D} \\
 10 \quad 20 \quad 30 \quad 40
 \end{array}
 \rightarrow
 \frac{10 + 20 + 30 + 40}{4} \rightarrow \frac{100}{4} = 25$$

↓
total values

↓
total no.

→ quick summary of the data

- avg mark
- avg age
- avg income

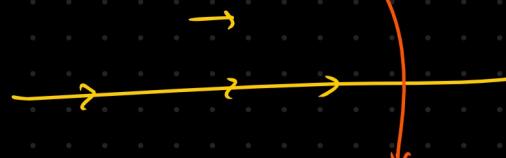
→ Drawback

→ sensitive to outlier

$$\text{group A} \rightarrow 10,000, 40,000, 30000, 20000 \rightarrow \underline{\underline{25,000}}$$

$$\text{group B} \rightarrow 10,000, 40,000, 30000, 20000, \underline{\underline{80,000}}$$

$$\frac{180,000}{5} = \underline{\underline{36000}}$$



Outlier

1, 2, 3, 4, 5, 6, 7, 8, 9, 50,

56, 57, 58

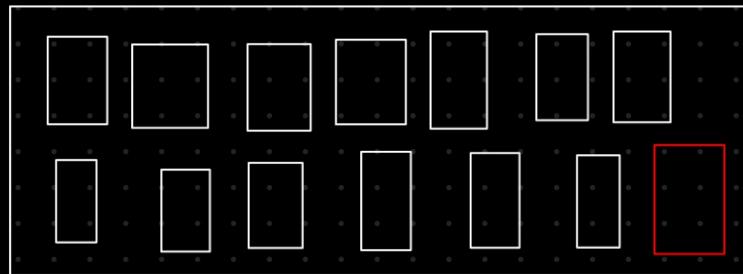
(100)

Summ - $42^{\circ}, 40^{\circ}, 37^{\circ}, 29^{\circ}$

temp - $42^{\circ}, 40^{\circ}, 37^{\circ}, 29^{\circ}, 35^{\circ}, 38^{\circ}, 2^{\circ}, 1^{\circ}, 3^{\circ}$
icon in cols

Square - car

circle - biker

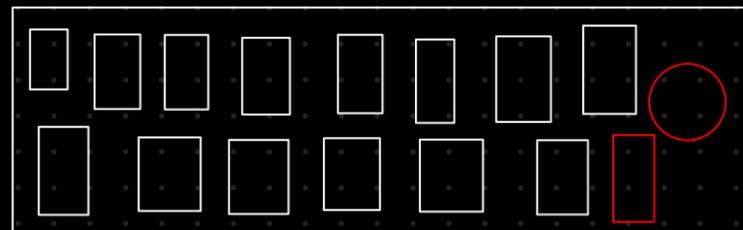


P.J.

find outlier
in parking lot

A.

find no
outlier B



(2) Median (Middle value)

len \rightarrow even \rightarrow avg of
2 centre
value
odd \rightarrow

A \rightarrow 60, 50, 40, 30, 10, 50, 40, 60, 55

A \rightarrow 10, 30, 40, 40, 50, 50, 55, 60, 60 \rightarrow 50

9 \downarrow 9

$$A \rightarrow 60, 50, 40, 30, 10, 50, 40, 90, 55$$

$$A \rightarrow 10, 30, 40, 40, 50, 50, 55, 60, 90 \rightarrow 50$$

$\overbrace{\quad}^4 \qquad \downarrow \qquad \underbrace{\quad}_{4}$

y odd

$$A \rightarrow 60, 50, 40, 30, 10, 50, 40, 90, 55, 95$$

$$A \rightarrow 10, 30, 40, 40, 50, 50, 55, 60, 90, 95$$

$\overbrace{\quad}^4 \qquad \overbrace{\quad}^5 \qquad \overbrace{\quad}^4$

y median

$\frac{50+50}{2} \rightarrow \frac{100}{2} \rightarrow [50]$

$(10, 10, 10, 10, 10, 10, 10, 90, 50, 60, 70)$
 score \uparrow class \rightarrow


(3) Mode

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
B	R	B	G	B	R	G	W	Y	Y

→ Most frequent color → (B)

mean & median?

↓
No

$$\left\{ 30, 40, 60, 70, \underline{\underline{70}}, \underline{\underline{70}}, 80 \right\} \rightarrow \underline{\underline{70}}$$

$$\overbrace{\quad}^4 - \overbrace{160}^4 \rightarrow \overbrace{\quad}^4$$

\rightarrow Standard deviation

How far values are spread from the mean.

Two players (A & B):

$$A \rightarrow 50, 51, 49, 50, 52 \rightarrow 50.4 \text{ (mean)}$$

$$B \rightarrow 0, 100, 10, 90, 50 \rightarrow 50 \text{ (mean)}$$

why of SD:

- * Mean alone does not tell the full story
- * They all can have different spread

$$\text{Score} = [10, 20, 30]$$

$$\star \text{step 1: Mean} = (10 + 20 + 30) / 3 = 20$$

\star step 2: find difference from mean:

$$(10 - 20)^2 = 100$$

$$(20 - 20)^2 = 0$$

$$(30 - 20)^2 = 100$$

$$(1 - 2)^2 = (-1)^2 \rightarrow 1$$

$$(2 - 1)^2 = (1)^2 \rightarrow 1$$

\star step 3 → average those differences

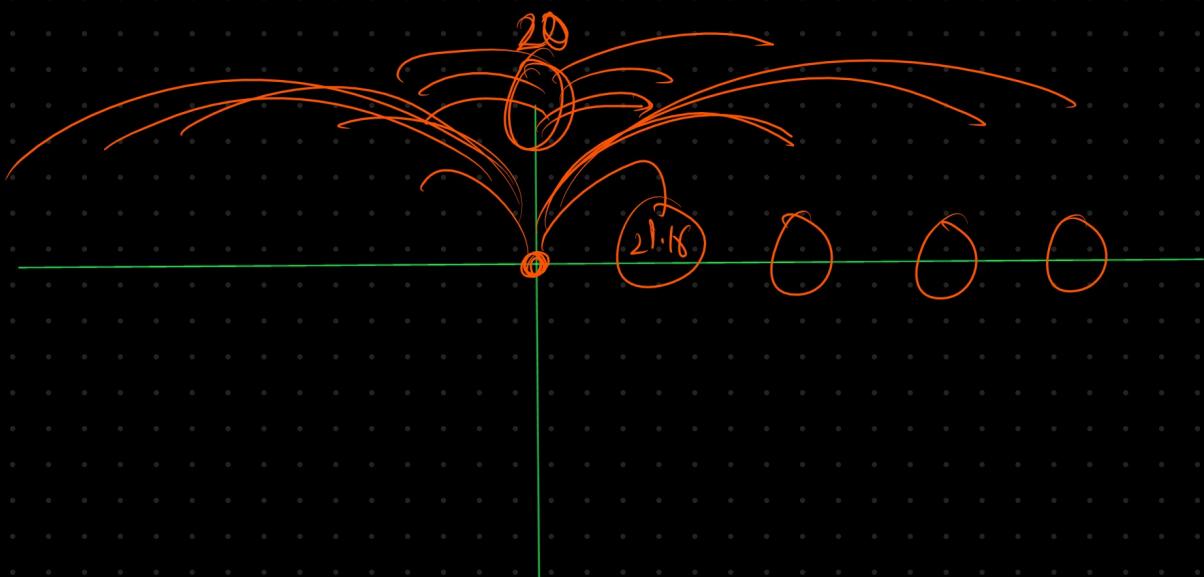
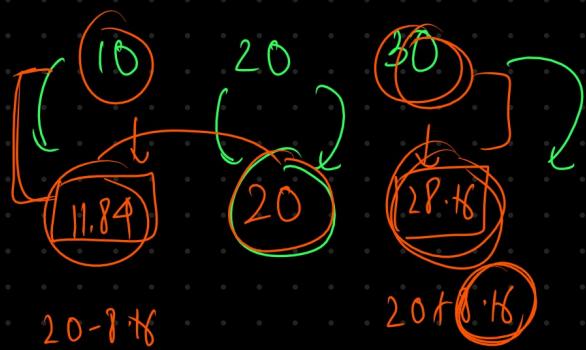
{ → always positive
→ mentally will be higher if difference is higher.

$$(100 + 0 + 100) / 3 \rightarrow 66.67 \rightarrow \underline{\text{Variance}}$$

* Take square root \rightarrow Standard Deviation $\rightarrow \sqrt{\text{variance}} \rightarrow 2$

$$\rightarrow \sqrt{88.67} \approx 8.16$$

On average, numbers deviate from mean by 8.16



$$[10, 20, 30] \rightarrow \underline{\text{marks}} \rightarrow 88.67 \rightarrow \begin{matrix} \text{mark}^2 \\ \text{marks} \end{matrix}$$

Diagram illustrating variance and standard deviation:

$[10, 20, 30] \rightarrow \underline{\text{marks}}$ $\rightarrow 88.67 \rightarrow \begin{matrix} \text{mark}^2 \\ \text{marks} \end{matrix}$

8.16 marks

→ Byte → 10129001100 → MB → 129