DATE : __/__/__
PAGE : ______

Statistics : Study of collecting, describing & draw inferences from data

Types : Descriptive , Inferential Statistics
↳ mean, mode median ; ↳ ~~Stop~~ Sampling → P-values & Hypothesis testing

→ Mean : Avg $\frac{\text{Sum of all obs.}}{\text{Total no. of obs.}} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$ or μ_x

Its univariate analysis, one variable (column) at once
Disadv : outliers affect this majorly.

→ median : middle value i) Sort acending order
ii) pick middle value : odd terms : $\left(\frac{n+1}{2}\right)^{th}$ term
even terms → $\frac{\frac{n}{2}^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$

Adv :- ~~outliers~~ few outliers don't affect ~~cent~~ this.

→ Mode : most common, frequent value
If we have only one mode value is unimodal
If we have 2 mode values its Bimodal value
~~More~~ than 2 mode values : multimodal.

→ Variance (σ^2) : Spread : measure how far a dataset is spread out.
* If all the values are same then no variability (or) zero variance (or) no spread
* Variance is average squared distance

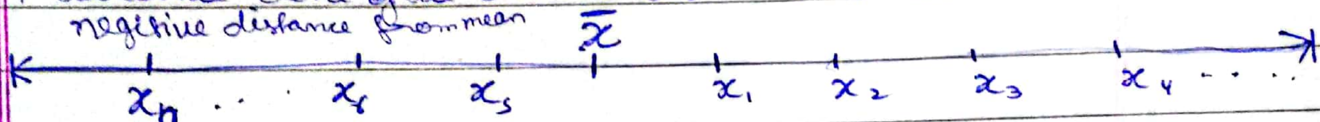
Note : • Complexity of mean → O(n) ; median → O(n log n)
• Median is computationally expensive
• Univariate analysis (1D) is analysis of only one variable/column
• all mean, mean, mode are univariate analysis, even σ^2 & σ

Note : • Mean & median are measure of Central tendency
• Variance, standard deviation, Range, IQR, are measure of Spread

* If spread or variance is small the mean is more accurate compare to large spread

$$\text{variance} = \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

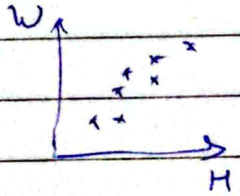
→ variance don't give acctual data as square are taken to ~~cancel~~ omitt negitive distance from mean



→ Standard deiviation = sqrt(variance) $= \sqrt{\sigma^2} = \sigma$

* variance, std dev are impacted by Outlier

* Bivariate: 2D
2 col^n.



H↑ W↑

For Bivariant we use measure of Relationship
→ Co-variance → Co-relation

as, variance $= \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)^2$

covariance $= \frac{1}{n} \sum_{i=1}^{n} \left((x_i - \mu_x)(y_i - \mu_y)\right) = cov(x, y)$

value of cov(x,y) can be positive or negative, so if

$cov(x, y) =$ +(ve) → x↑ ⇒ y↑
&
-(ve) → x↑ ⇒ y↓