Descriptive Statistics :-

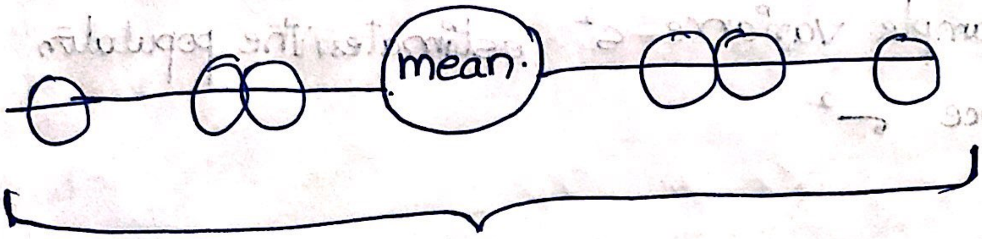
Range = Max - Min

Measures of variability :-

1. Variance
2. Standard deviation
3. Coefficient of variation.

→ Univariate
i.e applies for one variable.

Variance :-



Variance measures the dispersion of a set of data points around their mean

$$\text{population Variance } (\sigma^2) = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

average squared distance from true mean

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

population → population variance is computed with denominator N

Sample → sample variance with denominator (n)

Suppose we draw n independent observations from a population with mean μ and population variance σ^2

μ → unknown

σ^2 usually unknown

→ The sample mean $\bar{x}$ estimates the population mean μ

⇒ The sample variance s^2 estimates the population variance σ^2

since we cant possibly calculate :

$$\frac{\sum (x_i-\mu)^2}{N}$$

→ as true mean is unknown

We could try.

$$\frac{\sum (x_i-\bar{x})^2}{n}$$

→ replacing true mean with best estimate of it as sample mean

⇒ subtracting $\bar{x}$ makes the sum value smaller to compensate that we divide by n-1.

$$s^2 = \frac{\sum (x_i-\bar{x})^2}{n-1}$$

↳ On average, the estimator equals the population variance σ^2

proving that s^2 is an unbiased estimator of σ^2

$$E(s^2) = E\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right] = \sigma^2$$

Let x_1, x_2, x_3, x_4 ----- x_n be n independent observations from a population with mean μ and variance σ^2.

$\boxed{E(x_i) = \mu}$ $\boxed{Var(x_i) = \sigma^2}$

A few relationships that we will use:

$$E(\Sigma x_i) = \Sigma(E(x_i))$$

$$E(cX) = cE(X)$$

$Var(X) = E(X^2) - [E(X)]^2 \rightarrow$ where X is random variable

$$\boxed{E(x_i^2) = \sigma^2 + \mu^2}$$

$$Var(\bar{x}) = E(\bar{x})^2 - E(\bar{x})]^2$$

$$\boxed{E(\bar{x})^2 = \frac{\sigma^2}{n} + \mu^2}$$

Numerator :-

$$E\left[\sum(x_i-\bar{x})^2\right]$$

$$= E\left[\sum\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\right]$$

$$= E\left[\sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2\right]$$

as we are adding up to n times it becomes n times of $\bar{x}^2$

$$= E\left[\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2\right]$$

$\bar{X} \neq \Sigma$ $x_i \neq \Sigma$

$$\boxed{\bar{x} = \frac{\sum x_i}{n}}$$

$$\boxed{\sum x_i = n\bar{x}}$$

$$= E\left[\sum x_i^2 - 2\bar{x}\cdot n\bar{x} + n\bar{x}^2\right]$$

$$= E\left[\sum x_i^2 - n\bar{x}^2\right]$$ $\therefore$ Expectation of sum = sum of the expectations

$$= \sum E(x_i^2) - E(n\bar{x}^2)$$

$$= \sum E(x_i^2) - nE(\bar{x}^2)$$

$$= \sum(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)$$

$$= n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}$$

$$= n\sigma^2 - \sigma^2$$

$$= \sigma^2(n-1)$$

$$E(s^2) = E\left[\frac{\sum(x_i-\bar{x})^2}{n-1}\right]$$

money.

$$= \frac{1}{n-1} E\left[\sum(x_i-\bar{x})^2\right]$$

std $\Rightarrow$ \$

$$= \frac{1}{\cancel{n-1}} \sigma^2 \cancel{(n-1)}$$

Why Variance std ?

coeff.

$$= \sigma^2$$

$\therefore$ Hence proved that s^2 is an unbiased ectimator of σ^2

1/20/2022

Standard Deviation :-

→ It is much more meaningful than variance.

$$\boxed{\text{population SD}(\sigma) = \sqrt{\sigma^2}}$$

$$\boxed{\text{Sample SD}(s) = \sqrt{s^2}}$$

Coefficient of variation :-

⇒ This is also known as relative standard deviation.

$$\boxed{CV = \frac{\text{standard deviation}}{\text{mean}}}$$

population → $C_v = \frac{\sigma}{\mu}$.

sample → $\hat{C}_v = \frac{s}{\bar{x}}$

⇒ standard deviation is the mo[illegible] measure of variability for a single data set

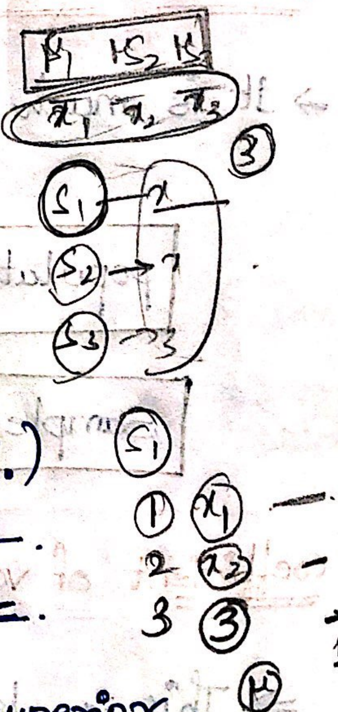
⇒ Coefficient of variation compares two or more data sets

Application of coefficient of variation :—

COV is argubaly a better overall indicator of relative risk when it's used to compare different securities.

Example :- Two different stocks offers different return with each exhibiting a different standard deviation.

	Stock A	Stock B
(mean) return	15%	10%
Standard deviation	10%	5%
C_v	(10%/15%)	(5%/10%)
	0.67	0.5



data suggests that stock B is a superior investment from a risk based perspective.

⇒ The data set which gives the lower coefficient of variation, it offers a better mean relative to SD.

Covariance :- bivariate

→ Measures of relationship between variables.

⇒ The two variables are correlated and the main statistic to measure this correlation is covariance.

⇒ covariance may be >0, =0, <0.

sample formula :-

$$S_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

population formula :-

$$\sigma_{xy} = \frac{\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)}{N}$$

Covariance gives a sense of direction :-

>0, the two variables move together

<0, the two variables moves in opposite direction

=0, the two variables are independent.

Correlation coefficient :-

⇒ correlation adjusts covariance, so that the relationship between the two variables becomes easy and intuitive to interpret.

$$= \frac{cov(x,y)}{stdev(x) * stdev(y)}$$

$\frac{(9-11)^2 + (13-11)^2}{2}$

$\frac{4+4}{2} = \frac{8}{2} = 4$

($\sigma^2 = 4$).

$-1 <$ correlation coefficent < 1.

Perfect positive relation :-

Correlation coeff = 1

- the entire variability of one variable is explained by the other.

correlation of 0 :

- Absolutely independent variables.

Negative correlation :-

Perfect Negative Correlation of -1.

imperfect Negative correlation : (-1,0).