Descriptive Statistics

Numerical Measure

- Measure of location "Central of Tendency"
  - Mean, Median, Mode, Percentile, Quartile
- Measure of variation "Dispersion"
  - Range, Interquartile Range IQR
  - Variance, standard deviation
  - coefficient of variation (CV)

* Measure of variation:

These numbers measure the spread "dispersion" of the Data

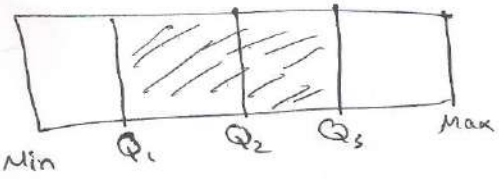
[1] Range: Max - Min

* Note: Range is highly influenced by outliers "Extreme Value"

Exp:- Find the Range : 50, 90, 3, 100, 107, 4, 2

Range = 107 - 2 = 105

[2] Interquartile Range (IQR): is the range of the middle 50% of the data.



$IQR = Q_3 - Q_1$

Exp: Given data :- 20, 40, 10, 25, 10 Find IQR

Sol: Sort data :- 10, 10, 20, 25, 40

$Q_1 = P_{25} \longrightarrow i = \frac{25}{100} \times 5 = 1.25 \rightsquigarrow i = 2$

$\longrightarrow \boxed{X_2 = 10}$ (Q_1)

$Q_3 = P_{75} \longrightarrow i = \frac{75}{100} \times 5 = 3.75 \rightsquigarrow i = 4$

$\longrightarrow \boxed{X_4 = 25}$ (Q_3)

$IQR = 25 - 10 = 15$

[3] & [4] Standard deviation & Variance

$\text{Variance} = (\text{Standard deviation})^2$

$\text{Standard deviation} = \sqrt{\text{Variance}}$

* Standard deviation and variance measure whether the data value cluster around the Mean

→ for sample Sample standard deviation $= S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$

Sample variance $= S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

for population $\underbrace{\sigma}_{\text{Pop. st. dev.}} = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N}}$

$$\underbrace{\sigma^2}_{\text{Pop. variance}} = \frac{\Sigma(x_i-\mu)^2}{N}$$

We will use the SD Mode to find std. dev. and variance

* Some observation about the measure of variation

(1) The standard deviation is the most ~~important~~ widely reported measure of variation

(2) The Range influenced by extreme value

(3) The Range, IQR, Std. dev., variance are always positive

(4) The relatively Small standard deviation indicate →
→ The data value cluster close to the mean

The relatively Large Standard deviation indicate →
→ The data value widely scattered from the mean

Casio - FX-82 MS

Exp. Given the sample

22, 18, 14, 40, 30, 14.

Find the sample mean, sample s.d, sample variance

Solution: $\bar{X} = 23$

$S = 10.26$

$S^2 = 105.2$

Exp: Given the data 200, 150, 400, 550, 310
Find the pop. Std. dev. and pop. variance

Sol.

Mode 2 200 M+ 150 M+ ------ 310 M+

Shift 2 2 = → $\sigma = 143.30$

X^2 = → $\sigma^2 = 20536$

5 Coefficient of variation C.V.

$$C.V = \frac{\text{Std. dev.}}{\text{Mean}} * 100\%$$

* Usually C.V used to compare variability

Exp: The mean and standard deviation for the grades of two sections in Stat 2361 were

Section 1: Mean = 64.87, Std. dev. = 14.8

Section 2: Mean = 70.45, Std. dev. = 12.71

Find C.V for each section, then determine which section has more variable grades :-

Sol. section ①: $C.V_1 = \left(\frac{14.8}{64.87}\right) \times 100\% = 22.81\%$

Section ②: $CV\text{-}2 = \left(\frac{12.71}{70.45}\right) \times 100\% = 8.04\%$

Section ① has more variability in grades