DATE : / /
PAGE NO.:

PAS UNIT - 1

Mean:

1. Direct Method: $\frac{\Sigma fx}{\Sigma f}$

2. Assumed mean method: $A + \frac{\Sigma fd}{\Sigma f}$ — A - Assumed mean, $d = x - A$

3. Step-deviation method: $A + \frac{\Sigma fd}{\Sigma f} \times h$ — h - height of class, $d = x - A/h$

4. Combined mean $= \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$

Median: (Middle most value)

- Discrete Series: $\frac{N}{2}$ → nearest value to that.

- Continous series: $l + \frac{h}{f}\left(\frac{N}{2} - c\right)$ (median class - class with highest freq)

  l - lower limit of Median class | h - height of class
  f - frequency of median class | c - c.f of class preceding median class

Mode: (most repeating value)

- Discreate series: mode is ~~freq~~ most repeating value.

- Continous Series: $l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$ — modal class - class with highest freq.

  l - lower limit of modal class | f_0 - freq of pre modal class
  h - height of interval. | f_1 - freq of modal class
  f_2 - freq of post modal class

DATE : / /
PAGE NO.:

Mode when 2 freq have highest no repeating.
(Grouping method)

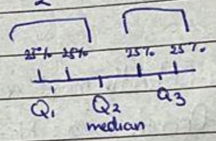
Relation between mean, median & mode

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Quartile Deviation :

$$QD = \frac{Q_3 - Q_1}{2}, \quad \text{coeff of } QD = \frac{Q_3 - Q_1}{Q_3 Q_1} \times 100$$

Q_1 - First Quartile
Q_2 - 2nd Quartile
Q_3 - 3rd Quartile



- For Individual and discrete series: (for individual, arrange in acending)

$$Q_1 = \left(\frac{n+1}{4}\right)^{th}, \quad Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ term}$$

if a^{th} term is decimal = The term + 0.-- × [+1th term + th term]

- Continuous series :

$$QD = \frac{Q_3 - Q_1}{2}, \quad \text{Coeff of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Q_1 = L + \left(\frac{\frac{N}{4} - CF}{f} \times h\right)$$
, l - lower limit of $\frac{N}{4}^{th}$ class, CF - above of the class from $\frac{N}{4}$
f - freq. of $\frac{N}{4}^{th}$ class, h = height of class

$$Q_3 = L + \left(\frac{3N/4 - CF}{f} \times h\right)$$
l - lower limit of $3N/4^{th}$ class, CF - above of the class from $\frac{3N}{4}$
f - freq. of $3N/4^{th}$ class, h = height of class.

DATE: / /
PAGE NO.:

- Deciles:
  - for individual and discrete series: (for individual, arrange in acending)

$$D_1 = \left(\frac{n+1}{10}\right), D_2 = 2\left(\frac{n+1}{10}\right), D_n = n\left(\frac{n+1}{10}\right)$$

  - Continous Series:

$$D_1 = L + \left(\frac{\frac{N}{10} - cf}{f} \times h\right), D_2 = L + \left(\frac{\frac{2N}{10} - cf}{f} \times h\right)$$

$$D_n = L + \left(\frac{N^2/10 - cf}{f} \times h\right)$$

- Percentiles:
  - for individual and discrete series: (for individual, arrange in acending)

$$P_1 = \left(\frac{n+1}{100}\right), P_2 = 2\left(\frac{n+1}{100}\right), P_n = n\left(\frac{n+1}{100}\right)$$

  - Continous Series:

$$P_1 = L + \left(\frac{\frac{N}{100} - cf}{f} \times h\right), P_2 = L + \left(\frac{\frac{2N}{100} - cf}{f} \times h\right)$$

$$P_n = L + \left(\frac{\frac{N^2}{100} - cf}{f} \times h\right)$$

- Mean Deviation:
  - Individual series:

MD about mean : $\Sigma|x-\bar{x}|/n$

MD about median : $\Sigma|x-\text{median}|/n$

MD about mode : $\Sigma|x-\text{mode}|/n$

  - Discrete & continous series:

MD about mean = $\Sigma f|x-\bar{x}|/\Sigma f$

MD about median : $\Sigma f|x-\text{median}|/\Sigma f$

MD about mode : $\Sigma f|x-\text{mode}|/\Sigma f$

coeff of MD = $(MD/\bar{x}) \times 100$.

DATE : / /
PAGE NO.:

- Range: $L-S$

  L- largest observation, S- smallest observation

  coeff of range $= \frac{L-S}{L+S} \times 100$

- Standard Deviation (σ), coeff of SD $= \frac{\sigma}{\bar{x}}$

  - Individual Series:

    n- no of observation * mean method: $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

    * Assumed mean: $\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$, $d = x - A$

  - Discrete & Continous Series:

    * Assumed mean

    N- sum of observation $\sigma = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$, $d = X - A$

    $\bar{X} = A + \frac{\Sigma f d}{\Sigma f}$

    * Step deviation method

    $\sigma = \left(\sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}\right) \times h$, $d = \frac{X-A}{h}$

    $\bar{X} = A + \frac{\Sigma f d}{\Sigma f} \times h$

    * normal mean method

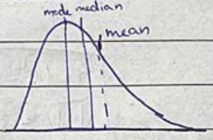
    $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

Variance (σ^2) : square of standard deviation

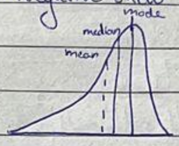
coeff of variance = $\frac{\sigma}{\bar{x}} \times 100$

· **Skewness :** (Lack of Symmetry) (Normally -1 to 1)

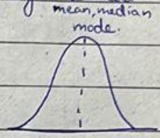
★ Positive skew



★ Negative skew



★ Symmetrical



measure of skewness = β_1

· Karl Pearson Coefficient of Skewness.

$$Sk = \frac{\text{mean} - \text{mode}}{SD}$$

SD - standard deviation

$$Sk = \frac{3(\text{mean} - \text{median})}{SD}$$

mode = 3median − 2mean

· Bowley's Coefficient of Skewness.

$$Sk = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

· Moments :

about mean :

· Individual : $\mu_r = \frac{1}{N}\Sigma(x-\bar{x})^r$, $r = 1, 2, 3, 4$

· Discrete & Continous : $\mu_r = \frac{1}{N}\Sigma f(x-\bar{x})^r$, $r = 1, 2, 3, 4$

DATE : / /
PAGE NO.:

about Arbitrary point 'A':

- Individual series: $\mu_r' = \frac{1}{N}\sum(X-A)^r$, $r = 1,2,3,4$
- Discrete & Continous: $\mu_r' = \frac{1}{N}\sum f(X-A)^r$, $r = 1,2,3,4$

about Origin:

$$\mu_r' = \frac{1}{N}\sum f x^r \text{ , } r = 1,2,3,4$$

Relation b/w μ_r & μ_r' (After getting μ_r' using Arbitrary, convert to μ_r)

$$\mu_1 = 0$$
$$\mu_2 = \mu_2' - \mu_1'^2$$
$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_2'^3$$
$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$\boxed{\begin{array}{l} a + \mu_1' = \text{mean} \\ \sqrt{\mu_2} = SD \end{array}}$$

Kurtosis

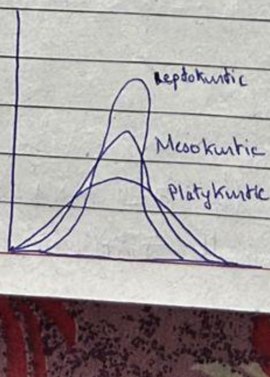
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

(measure of Skewness)

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$ (Measure of ~~Skewness~~ Kurtosis)

$$\gamma_1 = \sqrt{\beta_1}$$

$$\gamma_1 = \beta_2 - 3$$



$\beta_2 > 3$, leptokurtic or $\gamma_1 > 0$

$\beta_2 < 3$, Platykurtic or $\gamma_1 < 0$

$\beta_2 = 3$, mesokurtic or $\gamma_1 = 0$

DATE : / /
PAGE NO.:

PAS UNIT-2

Correlation: relationship between two or more variables.
(coeff of correlation)

A 10 20 30 40 or 40 30 20 10
B 100 200 300 400 or 400 300 200 100

* Positive Correlation: variables change in same direction.

A 10 20 30 40 50
B 100 90 80 70 60

* Negative Correlation: variables change in opposite direction.

* Simple Correlation: b/w only 2 variables.
* Multiple Correlation: b/w 2 or more variables.
* Partial Correlation: one variable constant, relationship b/w others variables.

* linear Correlation: plot the variables, straight line - positive. - negative.
* Non Linear Correlation: plot variables, not straight line.

- Karl Pearson's coeff of correlation

$$r(x,y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2 \cdot \Sigma(Y-\bar{Y})^2}}$$

- Spearman's Rank Correlation coefficient

$$R = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

- Rank given
- Rank not given (give lowest as rank 1 and given in increasing order for x and y.)

DATE : / /
PAGE NO.:

- When Ranks are repeating:

$$R = 1 - \frac{6\left\{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots\right\}}{N(N^2-1)}$$

m => no of times the observation is repeating

N => total no of observation.

Regression

Least Square Method

X on Y

$X = a + bY$

$\Sigma X = Na + b\Sigma Y$

$\Sigma XY = a\Sigma Y + b\Sigma Y^2$

Y on X

$Y = a + bX$

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X^2$

Regression Coeff method

X on Y

$X - \bar{X} = b_{xy}(Y - \bar{Y})$

$$b_{xy} = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Y on X

$Y - \bar{Y} = b_{yx}(X - \bar{X})$

$$b_{yx} = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$r = \sqrt{b_{xy}\, b_{yx}}$$

r => correlation coeff, σ_x = SD of x, σ_y = SD of y.

DATE: / /
PAGE NO.:

Random Variable

$X: S \rightarrow R \rightarrow$ Real Numbers.

X ↓ Random Variable, S ↓ Sample space

① Discrete RV: $x = 0, 1, 2, 3 \dots$
② Continous RV: $a \le x \le b$

Probability Density function / Probability Function

$(x_i, f(x_i)) \rightarrow$ P.d.f

- $f(x_i) \ge 0$ & $\sum f(x_i) = 1$ x is discrete
- $f(x_i) \ge 0$ & $\int_{-\infty}^{\infty} f(x_i)dx = 1$ x is Continous.

$E[X] = \begin{cases} \sum x f(x), & X \text{ is discrete RV} \\ \int_{-\infty}^{\infty} x f(x) dx, & X \text{ is continous RV.} \end{cases}$

$E(X)$

★ mean $= \mu = E(X) = E[X]$

★ Variance $(X) = E(X^2) - [E(X)]^2$

$E(X^2) = \sum x^2 f(x)$ - descrete.

$\int_{-\infty}^{\infty} x^2 f(x) dx$ - continous.

P.T.O $\rightarrow$

$P \rightarrow$ Success $\rightarrow 1$

$q = 1 - P \{ \rightarrow$ Failure $\rightarrow 0$

DATE : / /
PAGE NO.:

Bernoulli's Distribution (n=1) (only if yes or no values)

$f(x,p) = p^{x}(1-p)^{1-x}$ where $x=0,1$

* mean $= \mu = E[X] = p$
* Variance $= pq$

· trails are finite

Binomial Distribution (n= ......) (only if multiple times)

$f(x,np) = {}^{n}C_{x}\, p^{x} q^{n-x}$ where $x = 0,1,2 \ldots n$

* mean $= \mu = E[X] = np$
* Variance $= npq$

· trails are finite.

Poisson Distribution (n is average) (only if average is being used)

$f(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, $x = 0,1,2 \ldots$ where λ is average no of success.

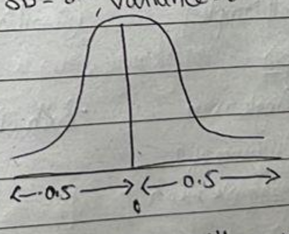
* mean $= \mu = E[X] = \lambda$
* Variance $= \lambda$.

Normal Distribution

$$f(x) = f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0$$

$Z = \frac{X-\mu}{\sigma}$, $\mu \to$ mean

SD $= \sigma$, Variance $= \sigma^2$



to get ~~ans~~ multiply with [100 for percentage
[sample size ~~1000~~ for no of values.

DATE : / /
PAGE NO.:

PAS Unit-3

Hypothesis : Quantative statement about population.

Null hypothesis : claim or statement about population, Parameter which is supposed to be true. (H_0)

Alternate hypothesis: compliment of Null hypothesis (H_1)

Test of Hypothesis : Deciding whether to accept or reject hypothesis.

Type I error: Reject H_0, when H_0 is true

Type II error: Not Reject H_0, when H_0 is false.

level of signifance: probability of error in accepting or rejecting H_0.

Tests

(SD given)	(SD not given)	(SD given)	(nothing)
Z-test	t-test	F-test	Chi Square test
(large sample)	(Small sample)	(Small sample)	(Small sample)
* types	types		
• Single mean	• Single mean	• Both SD given	• No parametric test.
• two sample mean!	• Difference of 2 Sample mean		
• difference of SD	• Dependent Sample		
• Single Proportion			
• difference of propotion			

*

DATE: / /
PAGE NO.:

Z-test

★ Single mean [Large Sample] $n \geq 30$

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$$ $\bar{X}$ = mean, μ = popl'n mean, $\sigma_{\bar{x}}$ = SD of mean.

- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (when popl^n SD is given)
- $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$ (when Sample SD is given)

★ Two Sample mean [Large Sample] $n \geq 30$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{x}_1\bar{x}_2}}$$

$$\sigma_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ (when popl^n SD is given)

$$\sigma_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$ (when Sample SD is given)

$$\sigma_{\bar{x}_1\bar{x}_2} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$ (when SD is common or same)

P.T.O →

DATE : / /
PAGE NO.:

★ Difference of SD [Large Sample] $n \geq 30$.

$$Z = \frac{S_1 - S_2}{\sigma_{S_1 S_2}}$$

S_1 = SD of sample 1, S_2 = SD of sample 2

$$\sigma_{S_1 S_2} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$ (Popl^n SD is given)

$$\sigma_{S_1 S_2} = \sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}$$ (Sample SD is given)

★ Single Proportion [Large Sample] $n \geq 30$

$$Z = \frac{p - P}{SE(p)}$$

P = popl^n proportion, p = Sample proportion.

- $SE(p) = \sqrt{\frac{PQ}{n}}$ (Popl^n proportion is given)

- $SE(p) = \sqrt{\frac{pq}{n}}$ (Sample proportion is given)

★ Difference of Proportion [Large Sample] $n \geq 30$.

- $Z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, $q = 1 - p$

- $Z = \dfrac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$

DATE : / /
PAGE NO.:

T-Test

* mean of Random Sample [Small sample]

$$t = \frac{\bar{X} - \mu}{S} \times \sqrt{n}$$

• degree of freedom = n−1

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

* difference of two Sample mean [Small sample]

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

• degree of freedom = $n_1 + n_2 - 2$

$$S = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}} \Rightarrow \text{Combined SD.}$$

* Dependent Sample [Small sample] $n_1 = n_2$

$$t = \frac{\bar{d}}{S} \times \sqrt{n}$$

• $\bar{d} = \frac{\Sigma d}{n}$

$$S = \sqrt{\frac{\Sigma(d_1-\bar{d})^2}{n-1}}$$

• degree of freedom = n−1

P.T.O →

DATE : / /
PAGE NO.:

F-Test

* ~~SD is given~~ Two Sample [Small sample] SD is given

$f_{v_1 v_2} = \frac{S_1^2}{S_2^2}$, $(S_1^2 > S_2^2)$ · degree of freedom: $v_1 = n_1 - 1$
$v_2 = n_2 - 1$

$f_{v_1 v_2} = \frac{S_2^2}{S_1^2}$, $(S_2^2 > S_1^2)$

$S_1^2 = \frac{\Sigma (X_1 - \bar{X}_1)^2}{n_1 - 1}$, $S_2^2 = \frac{\Sigma (X_2 - \bar{X}_2)^2}{n_2 - 1}$

* Chi-Square Test / χ^2-Test [Small Sample] (Non-Parametric Test)

O_i $(i = 1, 2 \dots) \Rightarrow$ set of observed freq

E_i $(i = 1, 2 \dots) \Rightarrow$ set of experimental/ hypothetical freq

* $\Sigma O = \Sigma E$
* $E > 5$ (always) (combine to get >5)

* $E = n \times p$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$