Maths formula cheatsheet

Unit 1 :-

Measure of central tendencies

- Median (Middle value of the data)
- Mean (The average of the data)
- Mode (Most commonly occurance value)

Mean = $\dfrac{\text{sum of all values in the data}}{\text{no. of values in the data}}$ $(\bar{x}) = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Mean

- Grouped: $\bar{x} = \sum_{i=1}^{n} \dfrac{x_i}{n}$
- Ungrouped: $\bar{x} = \sum_{i=1}^{n} \dfrac{f_i x_i}{n}$

Step dev :- $\underset{\text{assumed mean}}{A} + h\left[\dfrac{\sum u_i f_i}{\sum f_i}\right]$ $u_i = \dfrac{d_i}{h}$ $d_i = x_i - A$

h ~ class diff.

assumed-mean :- $A + \dfrac{\sum d_i f_i}{\sum f_i}$ $d_i = x_i - A$

Median :- For odd terms :- $\left(\dfrac{n+1}{2}\right)^{th}$ term.

For even terms :- $\dfrac{\left(\frac{n}{2}\right)^{th}\text{term} + \left(\frac{n}{2}+1\right)^{th}\text{term}}{2}$

Median $= l + \dfrac{h\left(\frac{N}{2} - C.F\right)}{f}$

l is the lower-limit of median class
N :- no. of observation
f :- frequency of median class
h :- class size
C.f :- C.f of preceeding class

Median class N/2 near to C.F.

Mode :- Most reapeated term.

Mode $= l + h\left(\dfrac{(f_1 - f_0)}{2f_1 - f_0 - f_2}\right)$

f_1 :- frequency of the modal class
f_0 :- " preceding class
f_2 :- succeding class
h :- size of class interval
l :- lower Limit of the modal class.

Modal class :- class with highest frequency.

Quartiles ÷

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ term $\qquad Q_3 = \frac{3}{4}(N+1)^{th}$ term

$$Q_1 = \frac{l + h\left(\frac{N}{4} - c.f\right)}{f}$$

$$Q_3 = l + \frac{h\left(\frac{3N}{4} - c.f\right)}{f}$$

Range = highest value − lowest value.

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

$D_i = \left(\frac{i}{10} * (n+1)\right)^{th}$ term $\qquad D_m = l + \frac{h}{f}\left(\frac{i.n}{10} - c.f\right)$

prev + 0.x (Next − prev)

$P_j = \left(\frac{j(n+1)}{100}\right)^{th}$ term $\qquad P_j = l + \frac{h}{f}\left(\frac{j.n}{100} - c.f\right)$

→ Measure of dispersion ÷

Range ÷ Highest V − lowest value

Coeff of range ÷ $\frac{H-L}{H+L}$

Quartile dev (semi inter quartile range) ÷ $Q_d = \frac{Q_3 - Q_1}{2}$

Coeff of QD ÷ $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

⇒ Mean deviation :

from mean $= \frac{\sum |x - \bar{x}|}{N}$

from median $= \frac{\sum |x - M|}{N}$

For freq distribution

$M.D = \frac{\sum f|x-\bar{x}|}{\sum f}$ $\quad$ M.D (about median) $= \frac{\sum f|x - \text{Median}|}{\sum f}$

M.D (about mode) $= \frac{\sum f|x - \text{Mode}|}{\sum f}$

for a population data mean μ is

$M.D = \frac{\sum f|x - \mu|}{\sum f}$

Co-eff of M.D $= \frac{M.D}{\sum f}$

→ Standard deviation (σ)

$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$ $\quad$ short-cut $\quad \sigma = \sqrt{\left\{ \frac{\sum x^2}{n} - \left(\frac{\sum x}{N}\right)^2 \right\}}$

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

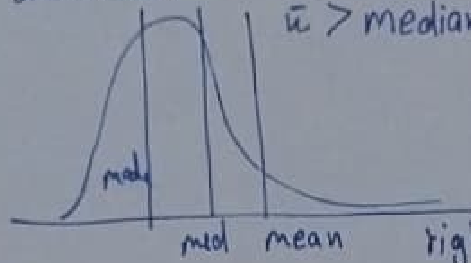
Coeff of S.D (σ) $= \frac{S.D}{\text{mean}} = \frac{\sigma}{\bar{x}}$

→ Variance (σ^2) $= S.D^2$

Coeff of variance (C.V) $= \frac{S.D}{\text{mean}} \times 100$ $\quad$ C.V is also known as Relative standard dev (RSD)

→ Skewness (assymetry)

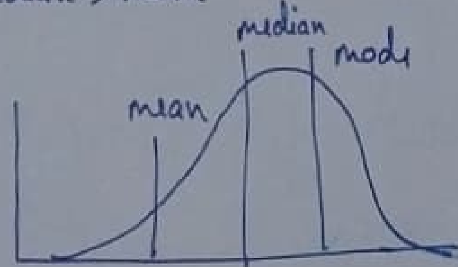
· ~~Negatively~~ Positively skewed distribution $\bar{x} >$ mode
$\bar{x} >$ median $>$ mode



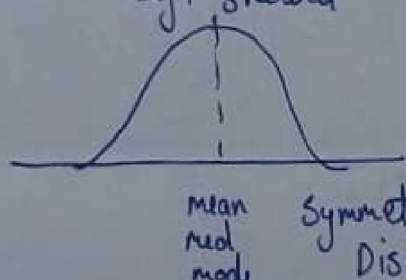
right skewed

· ~~positively~~ Negatively skewed distribution ÷

mode > median > mean



Left skewed



Symmetric Distribution.

→ Measures of skewness

- Karl Pearson's Coeff of skewness

$$SK_p = \frac{\text{Mean} - \text{Mode}}{S.D(\sigma)}$$

$-1 <$ coeff of skewness < 1

$$SK_p = \frac{3(\text{Mean} - \text{Median})}{S.D}$$ (in case of indeterminate mode)

$-3 < SK_p < 3$ for ↑ formula & coeff of skewness $= 0$ for symmetrical data

→ Kelly's Coeff of Skewness

$$SK_K = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$ (Percentile based)

$$SK_K = \frac{D_9 - 2D_5 + D_1}{D_9 - D_1}$$ (Decile based)

→ Bowley's Coeff of skewness

$$SK_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

→ Moments (Wo toh uske saath banti thi) ☺

$$\mu_r = \frac{1}{N}\Sigma(X - \bar{x})^r$$

$$\mu_r = \frac{1}{N}\Sigma f(X - A)^r$$

for grouped data

$$\mu_r = \frac{1}{N}\Sigma f(X - \bar{x})^r$$

$$\mu_r = \frac{1}{N}\Sigma f(X - A)^r$$

$$\mu_r' = \frac{\Sigma f (X-A) \times i^r}{N}$$ – Moment around arbituary point

formulas to convert moment around arbituary point to moment about mean

$$\mu_1 = 0, \quad \mu_2 = \mu_2' - \mu_1'^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$

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

$$\beta_1 = \frac{(\mu_3^2)}{(\mu_2^3)}$$ (Measure of skewness)

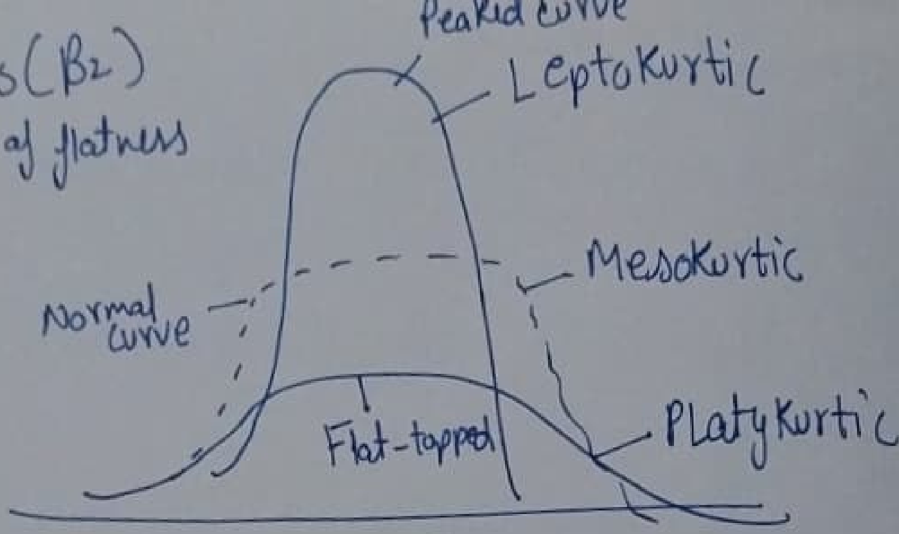
$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}$$

→ If μ_3 is +ive the skewness is +ive & vice versa

$$\beta_2 = \frac{\mu_4}{(\mu_2^2)}$$ (Kurtosis)

$$\beta_2 = \gamma_2 - 3$$

→ Kurtosis (β_2)
→ Degree of flatness



Leptokurtic : $\beta_2 > 3$
Mesokurtic : $\beta_2 = 3$
Platykurtic : $\beta_2 < 3$

UNIT- 2

$$P = \frac{\text{no. of favourable events}}{\text{Total events}}$$

$P(\bar{A}) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$

$P(A \cap B) = P(A) \times P(B)$

→ Bernoulli distribution (Discrete)
Bernoulli trial – random exp with only 2 outcomes
It is a binomial distribution with single trial ($n=1$).

→ BINOMIAL DISTRIBUTION : (Discrete)

$$\boxed{P(r) = {}^nC_r \, p^r q^{n-r}}$$

n = no. of trials
r = no. of successes
p = P(success)
$q = 1-p$

Mean of BND : $\boxed{\mu = np}$

Variance (σ^2) = $\boxed{\sigma^2 = npq}$

S.D (σ) = $\boxed{\sigma = \sqrt{npq}}$

→ Poisson Distribution ÷ (Discrete)

If $n > 20$ $p \to 0$ $p < 0.05$ we use Poisson instead of BND

$$P(X=x) = \frac{m^x e^{-m}}{x!}$$

$m = np$ variance $= np = m$

→ Normal distribution ÷ (Continuous)

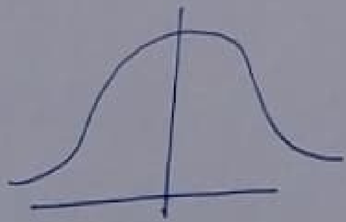
- mean = median = mode

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ $$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

μ = mean of x

σ = S.D of x

$$Z = \frac{x-\mu}{\sigma}$$



$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

---x---x---

$$\mu = \frac{\sum p_i x_i}{\sum p_i}$$

$$\sigma^2 = \sum p_i (x_i - \mu)^2$$

$$\sigma^2 = \sum p_i x_i^2 - \mu^2$$ if μ is not a whole no.

$$S.D = +\sqrt{\sigma^2}$$

---x---x---

Expectation

$E(x)$ = mean $\quad E(x) = \sum_{i=1}^{n} p_i x_i$

$= \sum x \cdot p(x)$

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$E(x^2) = \sum x^2 f(x)$$

$\mu_1 = E(x)$

var $= E(x^2) - (E(x))^2$ $\quad \mu_2 = E(x^2) - [E(x)]^2$ (variance)

→ Correlation Analysis :-

- Karl Pearson's coefficient or ~~rank~~ co-variance method.

$$r \text{ or } r(xy) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

Cov - covariance.

$$Cov(x,y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{n} = \frac{\sum xy}{n}$$

$$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}, \quad \sigma_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n}}$$

$$\boxed{r(x,y) = \frac{\sum xy}{\sqrt{\sum x^2}\sqrt{\sum y^2}} \qquad x = X-\bar{X}, \quad y = Y-\bar{Y}}$$

- Spearman's Rank Correlation:

$$r_s \text{ or } \rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

D = diff in ranks.

when ranks aren't given assign rank 1 to small val & go on.

when ranks repeat.

$$\rho = 1 - \frac{6\left[\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \_\_\right]}{n(n^2-1)}$$

- Regression Analysis:

regression

x on Y $= X = a + by$

→ Least square method :-   – Curve fitting.

① $y = a + bx$ — ①

② $\sum y = an + b\sum x$ — ②

③ Multiple ① both sides by x

$xy = ax + bx^2$

$\sum xy = a\sum x + b\sum x^2$ — ③

Solve for a & b using ② & ③

→ Regression analysis

$x = a + by$ $y - \bar{y} = b_{yx}(x - \bar{x})$

simple mean method

$$\boxed{b_{yx} = \frac{\sum XY}{\sum X^2}}$$

$$b_{xy} = \frac{\sum XY}{\sum Y^2}$$

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

bxy is regression coeff of X on Y.

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

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

r = coef or correlation.

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