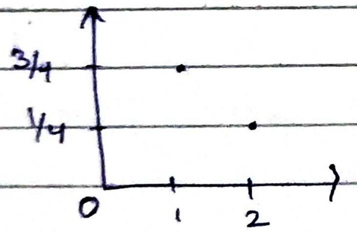
Q1) Represents ~~unc~~ uncertain outcomes from a random event, like rolling a die (e.g. X with value 1-6)

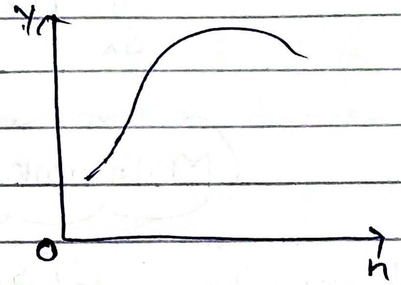
Random variable

Discrete | Continuous

Probability distribution

X (no. of head)	0	1	2
$P(x)$	$1/4$	$2/4$	$1/4$





Weight of Student in college

Probablity Mass function

① $P(x) \geq 0$

② $\sum P(x) = 1$

Mean

$E(x) = \cancel{E(x)\,P(x)}$
$\sum x P(x)$

Probability Density function

① $\int_{-\infty}^{\infty} f(x)\, dx = 1$

② $f(x) \geq 0$

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

Bionomial
Possion Distibution

Exponential
Normal
Uniform

Cumulative Distribution function

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$$

Relationship between distribution function & density function

$$\frac{d}{dx} F(x) = f(x)$$

Mathematical Expectation

denoted by $E(\phi(x))$

Discrete → $\sum_x \phi(x) P(x)$

Continous → $\int_{-\infty}^{\infty} \phi(x) f(x)\, dx$

if $\boxed{\phi(x) = x}$

Mean $= E(x)$

$\bar{x}$

Discrete → $E(x) = \sum_x x\, P(x)$

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

Variance

$$= E(x^2) - [E(x)]^2$$

$$\therefore \int_0^\infty x^{n-1} \cdot e^{-ax} \, dx = \frac{\Gamma n}{a^n}$$

Bivariate random variable

A bivariate random variable involves two random variables Simultaneously describing outcomes (e.g X and Y representing Separate events) often studied for thier joint probabilities and correlations.

<u>Bivariate Discrete</u> : if the possible of (x,y) are finite or countably infinite (x,y) is called a two dimensional discrete random variable

<u>Bivariate Continous</u> : If (X,Y) can be assume all values in specified region $R(x,y)$ is called two dimensional continous random variable

Joint probablity mass function

if (X,Y) is a two dimensional discrete random variable Where $X = x_i$ & $Y = y_j$, i & $J = 1, 2 ---$

Then $P(X=x_i, Y=y_j) = P_{ij}$

① $P_{ij} \geq 0$
② $\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} = 1$

The set of triple (X_i, Y_j, P_{ij}) is called JPD of X, Y

Marginal probability mass function

$$P(X=x_i) = \sum_j P_{ij} = P_{i1} + P_{i2} + \cdots$$

The collection of pair (x_i, P_i) is called marginal PMF of x

Marginal PMF of y

$$P(Y=y_j) = \sum_i P_{ij} = P_{1j} + P_{2j} + P_{3j} + \cdots$$

The pair (y_j, P_j) is called marginal PMF of y

If X and Y are independent

if $P_{xy}(x,y) = P_x(x)\, P_y(y)$

Bionomial Distribution

1. All the trials are independent
2. Number n of trial is finite
3. The probability of sucess is same of each trial

$$\boxed{P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}}$$

$p \to$ Probability of sucess
$q \to$ Probability of failure

$$\boxed{p + q = 1}$$

Probability mass function

$$\boxed{P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}}$$

1. $P(x) \geq 0$
2. $\sum_{x=0}^{n} P(x) = 1$

Moment generating function

$$MGF = M_x(t) = (pe^{t} + q)^{n}$$

Characterstics function

$$\phi_x(t) = E(e^{ixt})$$
$$= (pe^{it} + q)^{n}$$

Probability generating function

$$Z_x(t) = E(z^{x})$$
$$= (zp + q)^{n}$$

Mean

$$E(x) = np$$

Variance

$$\sigma_x^2 = E(x^2) - [E(x)]^2$$
$$= npq$$

Date ..../..../........

Formula Sheet

Frequency distribution

1) Arithmetic mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum f_i x_i}{N}$$

$$\bar{x} = A + \frac{h \sum f d}{N} \qquad \boxed{d = \frac{x - A}{h}}$$

2) Median

$$\text{median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

l = lowest limit
h = diff b/w interval
C = above value of A

3) Mode

$$\text{mode} = l + \frac{h(f_K - f_{K-1})}{2f_K - f_{K-1} - f_{K+1}} \quad \text{(continous)}$$

- Discrete ⇒ Highest frequency of x

4) Geometric mean

$$G = \text{Antilog}\left[\frac{1}{n}\sum_{i=0}^{n} \log x_i\right]$$

Date ..../..../........

Discrete = $\text{Antilog}\left[\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i\right]$

Alternative

Mode = 3 median − 2 mean

5) Harmonic mean

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} \qquad H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}}$$

Moments

1.) About a point

$$\mu_1' = \frac{1}{N}\sum f_i d_i$$

$$\mu_2' = \frac{1}{N}\sum f_i d_i^2$$

$$\mu_3' = \frac{1}{N}\sum f_i d_i^3$$

$$\mu_4' = \frac{1}{N}\sum f_i d_i^4$$

2.) About mean

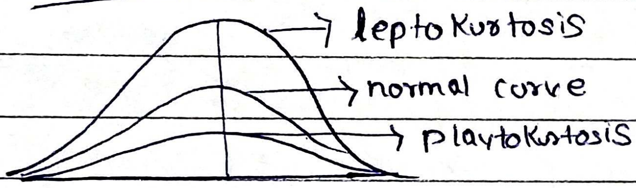
$\mu_1 = \mu_1'$
$\mu_2 = \mu_2' - \mu_1'^2$
$\mu_3 = \mu_3' - 3\mu_3'\mu_1' + 2\mu_1'^3$
$\mu_4 = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^2$

Karl's pearson coefficient

$\beta_1 = \frac{\mu_3^2}{\mu_2^2}$ $\beta_2 = \frac{\mu_4}{\mu_2^2}$

$\gamma_1 = \pm\sqrt{B_1}$ $\gamma_2 = B_2 - 3$

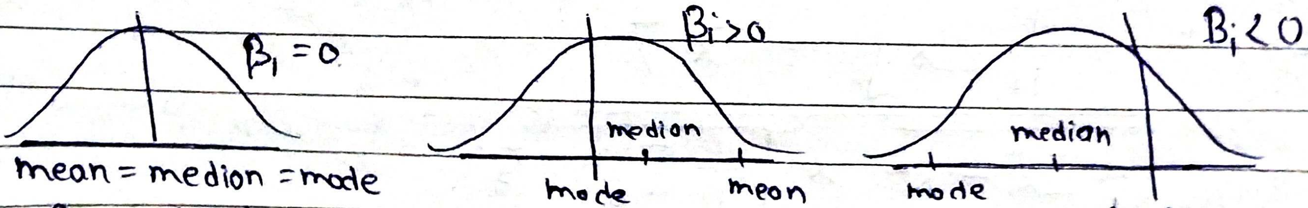
Kurtosis => Flatness



$\beta_2 = \frac{\mu_4}{\mu_2^2}$

1) $\beta_2 = 3$, $\gamma_2 = 0$ → Normal curve (mesokurtosis)
2) $B_2 < 3$, $\gamma_2 < 0$ → Playto Kurtosis
3) $\beta_2 > 3$, $\gamma > 0$ → Laptokurtosis

Skewness → Symmetric



Symmetrical ~~mode~~ ~~Skewed~~

Positive skewed

(−)ve skewed

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

Quartile deviation

$$QD = \frac{Q_3 - Q_1}{2}$$

Cofficient of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

$Q_1 \to$ First quatile
$Q_2 \to$ Second quatile (median)
$Q_3 \to$ third quatile

$$\left.\begin{aligned} Q_1 &\Rightarrow \left(\frac{n+1}{4}\right)^{th} \text{ term} \\ Q_3 &\Rightarrow 3\left(\frac{n+1}{4}\right)^{th} \text{ term} \end{aligned}\right\} \text{discrete}$$

$$\left.\begin{aligned} Q_1 &\Rightarrow L + \left(\frac{N/4 - CF}{f}\right) \times h \\ Q_3 &\Rightarrow L + \left(\frac{3N/4 - CF}{f}\right) \times h \end{aligned}\right\} \text{Continous}$$

Correlation Cofficient

$$r(x,y) = \frac{\frac{1}{n}\sum xy - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2} \times \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}}$$

Rank correlation $\Rightarrow P = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ $\qquad d_i = x_i - y_i$

Date ..../..../........

Rank are repeated

$$P = 1 - 6\left[\frac{\sum D_i^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots}{n(n^2-1)}\right.$$

Curve Fitting :-

1) Straight line

$$y = a + bx$$
$$\sum y = na + b\sum x$$
$$\sum yx = a\sum x + b\sum x^2$$

2) Power curve

$$y = ax^b$$
$$U = A + bV$$
$$\sum U = nA + b\sum V$$
$$\sum UV = A\sum V + b\sum V^2$$

$U = \log y$
$A = \log a$
$V = \log x$

3) Exponential curve

$$y = ab^x$$
$$U = A + xV$$
$$\sum U = nA + xV$$
$$\sum Ux = A\sum x + V\sum x^2$$

$$\boxed{\begin{array}{l} U = \log y \\ A = \log a \\ V = \log b \end{array}}$$

4) Parabola curve

$$y = a + bx + cx^2$$
$$\sum y = na + b\sum x + c\sum x^2$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

Regression :-

↳ avg relation b/w two variable

X on y

$$(x - \bar{x}) = r\frac{\sigma x}{\sigma y}(y - \bar{y})$$

y on x

$$(y - \bar{y}) = r\frac{\sigma x}{\sigma y}(x - \bar{x})$$