UNIT-1

$P(E) = \frac{n(E)}{n}$

$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots$

$P(E) + P(E^c) = 1 \Rightarrow P(E^c) = 1 - P(E)$

$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cup F) - P(F \cup G) - P(G \cup E) + P(E \cap F \cap G)$

Conditional probability : $P(E|F) = \frac{P(E \cap F)}{P(F)}$

Independent events : $P(E \cap F) = P(E) \cdot P(F)$.

Theorem of total probability : $P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A|E_i)$

Bayes's theorem: $P(E_i|A) = \frac{P(E_i) \cdot P(A|E_i)}{\sum P(E_i) \cdot P(A|E_i)}$

Probability distribution: If a random variable x takes values $x_1, x_2, x_3 \dots x_n$ with respective probabilities $P_1, P_2, P_3 \dots P_n$ then

X	=	x_1	x_2	$\dots$	x_n
$P(x)$	=	P_1	P_2	$\dots$	P_n

is known as probability distribution of X.

$$P(x_1) + P(x_2) + \dots + P(x_n) = 1$$

Probability density function : Let X be a random variable. we say that x is a continuos random variable if there exist a non-negative function f, defined for all real $x \in (-\infty, \infty)$ having a property that for any set B of real numbers.

$$P(x \in B) = \int_B f(x)\,dx.$$

the function 'f' is called the probability density function of RV 'x'

$$P\{x \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx.$$

Let $B = [a, b] \Rightarrow P(x \in B) = P\{x \in [a, b]\}$

$$= P\{a \leq x \leq b\} = \int_a^b f(x)\,dx$$

if $a = b$ then $\int_a^a f(x)\,dx = 0$

$$P(x < a) = P(x \leq a) = \int_{-\infty}^{a} f(x)\,dx.$$

Expected value $(E(x))$

$$E(x) = \sum x\, p(x)$$

Exception of continuous random variable: $E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$.

Variance : $var(x) = E(x^2) - (E(x))^2$

$$Var(ax+b) = a^2 \cdot var(x)$$

Standard variance: $S(x) = \sqrt{var(x)}$

UNIT - 2

$$p(x=i) = nC_i \cdot p^i q^{n-i}$$

$i = 0, 1, 2 \ldots n$; $q = 1-p$ $\Rightarrow nC_i\,(p)^i\,(1-p)^{n-i}$

Mean and variance of binomial distribution:

$$E(x) = np$$

$$Var(x) = npq = np(1-p)$$

Computing binomial distribution function: Suppose 'x' is a binomial with parameter (n,p) then

$$p(x = k+1) = \frac{p}{1-p} \cdot \frac{n-k}{k+1}\, p(x=k)$$

Poisson random variable: $p(X=D) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$; $i = 0, 1, 2 \ldots$

* The expected and variance of poisson random variable are both equal to its parameter λ.

$$E(x) = \lambda$$

$$Var(x) = \lambda\left(1 - \frac{\lambda}{n}\right) \quad \left[\because \frac{\lambda}{n} \approx 0\right]$$

$$= \lambda$$

Moment generating function: The MGT $M(t)$ of the random variable x is defined for all real values of t by

$$M(t) = E[e^{tx}] \Rightarrow M'(t) = E[xe^{tx}]$$

$M(t) = \sum_{x_i} e^{tx_i}\, p(x_i)$ $\Rightarrow$ if x is descrete mass function $p(x)$

$\int_{-\infty}^{\infty} e^{tx}\, p(x)$ $\Rightarrow$ if x is continous with density $f(x)$

if $M(t) = E[e^{tx}]$, $p(x)$ is a mass func.

then $p(x=k) = (e^t p + 1 - p)^n$

Moments (μ_r'): $\mu_r' = [E(x-0)^r] = E(x)^r$

1st moment $\mu_1' = E[(x)] = \bar{x}$

2nd moment $\mu_2' = [E(x)^2]$

3rd moment $\mu_3' = [E(x)^3] \ldots$

Moment about origin

$(\mu_r'') = E[(x-a)^r]$

1st moment $\mu_1'' = E[(x-a)]$

2nd moment $\mu_2'' = E[(x-a)^2]$

3rd moment $\mu_3'' = E[(x-a)^3] \ldots$

moment about point (a)

$(\mu_r) = E[(x-\bar{x})^r]$

1st moment $\mu_1 = E[(x-\bar{x})] = 0$

2nd moment $\mu_2 = E[(x-\bar{x})^2]$

$= E(x^2) - (E(x))^2 = Var(x)$

$\Rightarrow \mu_2 = \mu_2' - (\mu_1')^2$

3rd moment $\mu_3 = E[(x-\bar{x})^3]$

$= E(x^3) - 3E(x^2)E(x) + 2(E(x))^3$

$\Rightarrow \mu_3 = \mu_3' - 3\mu_2'\mu_1 + 2(\mu_1')^3 \ldots$

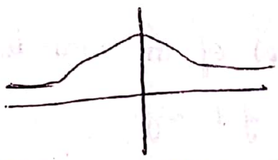
moment about mean $(\bar{x})$

<u>Skewness</u>: $\beta = \frac{(\mu_3)^2}{(\mu_2)^3}$

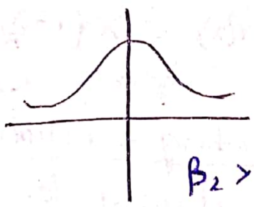
<u>Kurtosis</u>: $\beta_2 = \frac{\mu_4}{\mu_2^2}$

if

$\beta = 0 \Rightarrow$ curve is symmetrical



$\Rightarrow$ mean = median = mode.

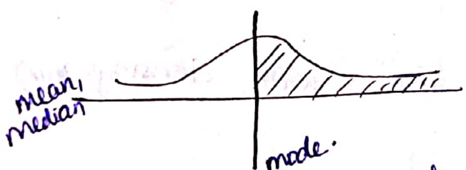


$\beta_2 > 3$: lapto Kurtic

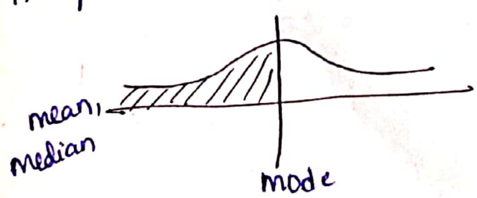
$\beta_2 = 3$: meso Kurtic

$\beta_2 < 3$: platy Kurtic

if $\beta > 0 \Rightarrow$ Curve is +vely skewed



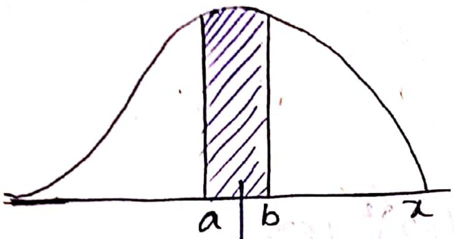
if $\beta < 0 \Rightarrow$ curve is -vely skewed



UNIT-3

Continuous random variable: $P\{x \in B\} = \int_B f(x)\,dx$

Note: $f(x)$ must satisfy $\int_{-\infty}^{\infty} f(x)\,dx = P\{x \in (-\infty, \infty)\} = 1$



$p(a \le x \le b) = \int_a^b f(x)\,dx$

$p\{x = a\} = \int_a^a f(x)\,dx = 0$

$p\{x < a\} = p(x \le a) = F(a) = \int_{-\infty}^{a} f(x)\,dx$

$p(a \le x \le b)$

$F(a) \rightarrow$ cumulative distribution of x

Expectation of cont. RV: $E[x] = \int_{-\infty}^{\infty} x\, f(x)\,dx$

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

$$E[ax+b] = aE(x) + b$$

Variance of cont. RV: $Var(x) = E[x^2] - (E(x))^2$

$$Var(ax+b) = a^2 \cdot Var(x)$$

Uniform Random variable: A RV is said to be uniformly distributed over the interval $(0,1)$ if its probability density func. is given by

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{Otherwise} \end{cases}$$

if $0 < a < b < 1$ then $p\{a \le x \le b\} = \int_a^b f(x)\,dx = \int_a^b 1\,dx = b - a$

$$p\{a \le x \le b\} = b - a$$

* In general, we say that X is a uniform RV on interval (α, β) if its a) probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

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

b) distribution function is given by

$$f(a) = \begin{cases} 0 & a \le \alpha \\ \frac{a - \alpha}{\beta - \alpha} & \alpha < a < \beta \\ 1 & a \ge \beta \end{cases}$$

Normal random variable: We say that X is a NRV with parameters μ and σ^2 if the density of X is given by

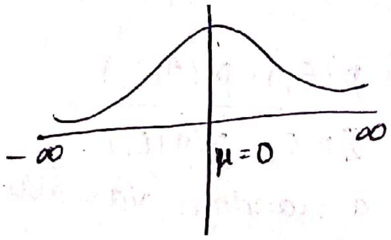
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} , -\infty < x < \infty$$

The density func. is a bell-shaped curve i.e Symmetric about μ.

Mean for NRV = μ

Variance for NRV = σ^2

If $\mu = 0, \sigma = 1 \Rightarrow \underset{\substack{\downarrow \\ \text{Standard NRV}}}{Z} = \frac{x-\mu}{\sigma} \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}, -\infty < z < \infty$



Exponential random variable: A continuous RV whose probability density func is given for some $\lambda > 0$ given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

This is said to be exponential random variable with parameter λ

Mean for ERV = $\frac{1}{\lambda}$

Variance for ERV = $\frac{1}{\lambda^2}$

Moment generating function:

a) uniform distribution: $M(t) = \dfrac{e^{t\beta} - e^{t\alpha}}{t(\beta - \alpha)}$

b) Normal distribution: $M(t) = e^{\left(\mu t + \frac{t^2\sigma^2}{2}\right)}$

$$M(t) = \frac{1}{\beta - \alpha}\left[\beta + \frac{\beta^2 t}{2!} + \frac{\beta^3 t^2}{3!} + \cdots - \alpha - \frac{\alpha^2 t}{2!} - \frac{\alpha^3 t^2}{3!} \cdots\right]$$

Exponential RV: A cont. RV whose probability density func. is given for some $\lambda, \lambda > 0$ is given by $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

$(S, \lambda), \lambda > 0$ is given by $f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{S-1}}{\Gamma(S)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Mean for ERV = $\frac{S}{\lambda}$

Variance for ERV = $\frac{S}{\lambda^2}$

Mean for ERV = $\frac{1}{\lambda}$

Variance for ERV = $\frac{1}{\lambda^2}$

Scanned with CamScanner

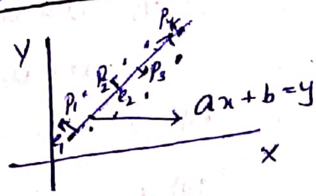
UNIT 4

Principle of least square: It is a procedure to fit a unique curve through given points.

Let $y = f(x)$ be eqn. of curve with data points $P_1(x_1, y_1), P_2(x_2, y_2)$..... $P_n(x_n, y_n)$

Curve fitting by least square method: $e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$
$\Downarrow$
minimum

Fitting of straight line:



Let $f(x) = Ax + B$ then

by least square method,

$$\Sigma x_K y_K = A \Sigma x_K^2 + B \Sigma x_K$$

$$\Rightarrow \boxed{\Sigma y_K = A \Sigma x_K + Bn}$$

*

x	x_1	x_2 ...	x_n
y	y_1	y_2 ...	y_n

then. a) fitting straight line i) $y = ax + bn$

$$\Rightarrow \Sigma y = a \Sigma x + bn \quad \text{—(1)}$$

ii) $yx = ax^2 + bx$

$$\Rightarrow \Sigma yx = a \Sigma x^2 + b \Sigma x \quad \text{—(2)}$$

b) fitting 2° parabola by least square method:

$$y = ax^2 + bx + cn$$

$$\Rightarrow \Sigma y = a \Sigma x^2 + b \Sigma x + \Sigma cn \quad \text{—(1)}$$

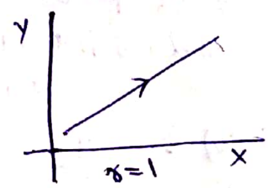
$$yx = ax^3 + bx^2 + cx$$

$$\Rightarrow \Sigma yx = a \Sigma x^3 + b \Sigma x^2 + c \Sigma x \quad \text{—(2)}$$

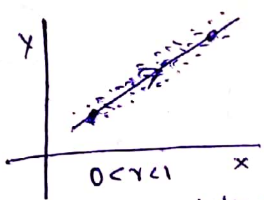
$$yx^2 = ax^4 + bx^3 + cx^2$$

$$\Rightarrow \Sigma yx^2 = a \Sigma x^4 + b \Sigma x^3 + c \Sigma x^2 \quad \text{—(3)}$$

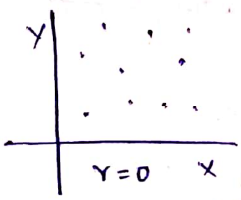
Correlation coefficient (r): Reation b/w 2 variables is correlation



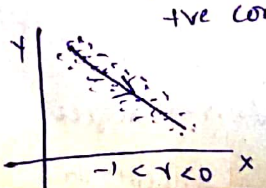
perfect +ve



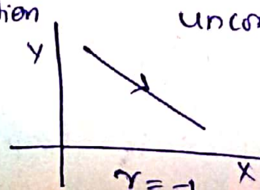
+ve correlation



uncorrelated data



-ve correlation



Perfect -ve

Karl pearson coefficient (or) correlation coefficient (r)

$$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

$$cov(x,y) = E(xy) - E(x)E(y)$$

$$E(xy) = \frac{\sum x_i y_i}{n}$$

$$E(x) = \frac{\sum x_i}{n}, \quad E(y) = \frac{\sum y_i}{n}$$

$$\sigma_x = \sqrt{V(x)} = \sqrt{(E(x))^2 - E(x^2)} = \sqrt{\left(\frac{\sum x_i}{n}\right)^2 - \left(\frac{\sum x_i^2}{n}\right)}$$

$$\sigma_y = \sqrt{V(y)} = \sqrt{(E(y))^2 - E(y^2)} = \sqrt{\left(\frac{\sum y_i}{n}\right)^2 - \left(\frac{\sum y_i^2}{n}\right)}$$

Rank correlation (ρ):

$$\rho_{xy} = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

d = rank of x − rank of y.

Regression: If the curve of regression is a st. line then it is said to be line of regression or linear regression.

Line of regression x on y:

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

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

Line of regression y on x:

$$y - \bar{y} = b_{xy}(x - \bar{x})$$

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

$$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

Angle b/w 2 lines of regression:

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

Critical values of z (z_α):

Critical values (z_α)	Level of significance (α) $\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
2-tailed test	2.58	1.96	1.645
Right tailed test	2.33	1.645	1.28
Left tailed test	−2.33	−1.645	−1.28

Hypothesis testing:

Single proportion

1) $\hat{P}$ - Sample proportion
2) P - population proportion
3) α - Level of significance

2 tailed test	Left tailed test	Right tailed test
$H_0: P = P_a$	$H_0: P = P_a$	$H_0: P = P_a$
$H_1: P \neq P_a$	$H_1: P < P_a$	$H_1: P > P_a$

P_a - assumed proportion.

$$Z_{cal} = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Difference of proportion:

1) P_1 and P_2 (two sample proportion)
2) α = Level of significance

2 tailed test	Left tailed test	Right tailed test
$H_0: P_1 = P_2$	$H_0: P_1 = P_2$	$H_0: P_1 = P_2$
$H_1: P_1 \neq P_2$	$H_1: P_1 < P_2$	$H_1: P_1 > P_2$

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\because P_1 = \frac{x_1}{n_1}$

$P_2 = \frac{x_2}{n_2}$

P - population proportion

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow Z_{cal} = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$Q_1 = 1 - P_1$

$Q_2 = 1 - P_2$

Difference of standard deviation:

1) σ_1, σ_2
2) α

2 tailed test	Left ~~tan~~tailed test	Right tailed test
$H_0: \sigma_1 = \sigma_2$	$H_0: \sigma_1 = \sigma_2$	$H_0: \sigma_1 = \sigma_2$
$H_1: \sigma_1 \neq \sigma_2$	$H_1: \sigma_1 < \sigma_2$	$H_1: \sigma_1 > \sigma_2$

$$Z_{cal} = \frac{S_1 - S_2}{\sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}}$$

UNIT- 5

Test for single mean (small samples): if size of sample < 30 then it is called small samples or exact sample.

T-test:

a) **T-test for single mean:** i) $H_0 : \mu = \mu_0$

ii) $H_1 : \mu \neq \mu_0$

$\mu < \mu_0$ (or) $\mu > \mu_0$

iii) $\alpha = 1\%$ (or) 5% (or) 10%

iv) $t_{cal} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ (raw data)

$= \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$ (discrete data)

t_{tab} : $V = n-1$ degree of freedom @ α - level - 1 tail

$\alpha/2$ level - 2 tail

if $t_{cal} < t_{tab}$,
accept H_0

b) **t-test for difference of mean:** i) $H_0 = \mu_1 = \mu_2$

ii) $H_1 : \mu_1 \neq \mu_2$

$\mu_1 < \mu_2$ (or) $\mu_1 > \mu_2$

iii) $\alpha = 5\%$ or 1% or 2% or 10%

iv) $t = \dfrac{\bar{x}_0 - \bar{y}_0}{\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \times S}$

$$S = \sqrt{\frac{1}{n_1 + n_2 - 2}\left(\Sigma (x_i - \bar{x})^2 + \Sigma (y_i - \bar{y})^2\right)}$$

t_{tab} : $V = n_1 + n_2 - 2$ degree of freedom @ α level - 1 tail

$\alpha/2$ level - 2 tail

if $t_{cal} < t_{tal}$

$\Rightarrow$ accept H_0

c) **Paired sample t-test (To test dependent data):** i) $H_0 = \mu = 0$

ii) $H_1 = \mu > 0$

iii) $\alpha = 5\%$ or 1% or 10% or 2%

iv) $t = \dfrac{\bar{d}}{\frac{S}{\sqrt{n}}}$

$d = x - y$

n = ~~no of~~ sample size

S = SD

t_{tab} : $V = n-1$ degrees

$t_{cal} < t_{tab} \Rightarrow$ Accept H_0

Testing significance of observation:

Correlation coefficient:

$$t = \frac{r}{\sqrt{1 - r^2}} \times \sqrt{n-2}$$

r: correlation coefficient
n: sample size
$V = n-2$ degrees

H_0 : correlated
H_1 : uncorrelated

$t_{cal} < t_{tab} \Rightarrow$ accept H_0

F-test:

Equality of variance:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$\sigma_1^2 > \sigma_2^2$ (or) $\sigma_1^2 < \sigma_2^2$

$\alpha = 5\%$ or 1%

$$F = \frac{S_1^2}{S_2^2} \quad (S_1^2 > S_2^2) \quad \left[\frac{\text{Greater variance}}{\text{Smaller variance}}\right]$$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} \; ; \; S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1}$$

$V_1 = n_1 - 1$, $V_2 = n_2 - 1$ degrees

if $F_{cal} < F_{tab} \Rightarrow$ accept H_0

Chi-square test (χ^2):

goodness of fit:

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i^2}$$

$E_i = NP_i$

O_i : observed frequency in ith class interval
E_i : Expected frequency in that ith " "
P_i : theoretical hypothesized probability associated with ith CI.

degree of freedom: $K - S - 1$
S: No. of parameters

Sample Size	No. of class intervals
20	Do not use chi-square test
50	5 to 10
100	10 to 20
>100	$\sqrt{n}$ to $n/5$

Recomendations for no. of CI for cont. data ↑