★ Arithmatic mean ($\bar{x}$)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\Sigma x_i}{N}$$

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{n} = \frac{\Sigma x_i f_i}{N}$$

★ Assumed mean method

$d_i = x_i - A$

$\Rightarrow \bar{x} = A + \frac{1}{N} \Sigma f_i d_i$

for continuous frequency distribution

$d_i = \frac{x_i - A}{h}$

$\Rightarrow \bar{x} = A + \frac{h}{N} \Sigma f_i d_i$

★ Median

i) find N, where $N = \Sigma f_i$

ii) find the observation where c.f is just greater than $N/2$

iii) corresponding 'x' is a median. (x)

for continuous frequency distribution

iii) corresponding class is median class ($l - h$)

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

gap ($h-l$) — frequency of Median class — Cummulative frequency of class preceding median class.

★ Mode

↳ (x_i) corresponding to maximum frequency.

for continuous

corresponding class is modal class ($l - h$) with frequency (f_1)

$$\text{Mode} = l + \frac{h(f_1 - f_0)}{(2 f_1 - f_0 - f_2)}$$

★ Mode = 3 Median − 2 ~~Mode~~ Mean

★ Geometric Mean (G)

$$G = (x_1 \times x_2 \times x_3 \times \dots \times x_n)^{1/n}$$

★ Harmonic Mean (H)

$$\frac{1}{H} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}{n}$$

★ $(A.M) \times (H.M) = (G.M)^2$

★ $a \& b \Rightarrow A.M = \frac{a+b}{2}, \ G.M = \sqrt{ab}, \ H.M = \frac{2ab}{a+b}$

★ ~~Mean~~ Standard deviation (σ^2)

$$\sigma^2 = \frac{1}{N} \sum f_i (x_i - \overline{x})^2 = \frac{1}{N} \sum f_i x_i^2 - [\overline{x}]^2$$

if $d_i = x_i - A$

$$\Rightarrow \sigma_d^2 = \frac{1}{N} \sum f_i d_i^2 - \left[\frac{1}{N} \sum f_i d_i\right]^2$$

if $d_i = (x_i - A)/h$

$$\Rightarrow \sigma_x^2 = \frac{h^2}{N} \sum f_i d_i^2 - \left[\frac{h}{N} \sum f_i d_i\right]^2 = h^2 \sigma_d^2$$

★ Coefficient of Variance (COV)

$$\Rightarrow \frac{\sigma}{\overline{x}} \times 100$$

★ $\sigma_1, \overline{x}_1$ $\sigma_2, \overline{x}_2$ $\sigma_3, \overline{x}_3$

f_1 f_2 f_3

$$\Rightarrow \overline{x} = \frac{\overline{x}_1 \times f_1 + \overline{x}_2 \times f_2 + \overline{x}_3 \times f_3}{f_1 + f_2 + f_3}$$

$$\Rightarrow \sigma^2 = \frac{f_1(\sigma_1^2 + d_1^2) + f_2(\sigma_2^2 + d_2^2) + f_3(\sigma_3^2 + d_3^2)}{f_1 + f_2 + f_3}$$

$$\Rightarrow d_1 = \overline{x} - \overline{x}_1, \ d_2 = \overline{x} - \overline{x}_2, \ d_3 = \overline{x} - \overline{x}_3.$$

Example 4.17, 4.25

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

* Probability

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

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \& \quad P(B/A) = \frac{P(A \cap B)}{P(A)}$$

probability of happening of event 'A' given that event 'B' has alredy completed

if A & B are independent events $\Rightarrow P(A \cap B) = P(A) \times P(B)$

→ Difference between mutually exclusive events and independent events is that in mutually exclusive case, we have 2 events that cannot occur at same time whereas incase of independent event one event remains unaffected by the occurence of other event.

(*) $S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

X(No. of Heads)	0	1	2	3
$P(X)$	$1/8$	$3/8$	$3/8$	$1/8$

Table. probability distribution

Random Variable (discrete)

P.m.f (Probability mass function) $\Rightarrow P_i \geq 0 \;\; \& \;\; \Sigma P_i = 1$

$F(x_i) = \Sigma P_i$ where $P(X \leq x_i) \Rightarrow F(X) = \begin{cases} 1/8 & x \leq 0 \\ 4/8 & x \leq 1 \\ 7/8 & x \leq 2 \\ 1 & x \leq 3 \end{cases}$

Distribution function / (c.d.f) Cummulative distribution function.

(*) continuous random variable

Probability density function (p.d.f) $[f(x)]$

i) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ ii) $f(x) \geq 0, \; -\infty < x < \infty$.

a) Arithmatic mean $= \int_a^b x f(x)\,dx$

b) Geometric mean (G) $\Rightarrow \log G = \int_a^b \log x \, f(x)\,dx$

c) harmonic mean (H) $\Rightarrow \frac{1}{H} = \int_a^b \frac{f(x)}{x}\,dx$

d) μ_r' (about origin) $= \int_a^b x^r f(x)\,dx$; μ_2 (about mean) $= \int_a^b (x-\bar{x})^2 f(x)\,dx$

Mean

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

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

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

$\therefore$ Mean $= \mu_1'$

Variance $= \mu_2$ (raw)

Page No.:
Date:

$$\int_a^{Q_1} f(x)\,dx = \frac{1}{4} \; ; \; \int_a^{Q_3} f(x)\,dx = \frac{3}{4}$$

e) Median (M) → Divides distribution into 2 parts.

$$\int_a^{M} f(x)\,dx = \int_M^{b} f(x)\,dx = \frac{1}{2}$$

f) Mode ⇒ f(x) is maximum

$$\Rightarrow f'(x) = 0 \;\; \& \;\; f''(x) < 0$$

g) Probability $P(E) = \int_E f(x)\,dx$

i) continuous/cummulative distribution function $[F(x)]$

$$\Rightarrow F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx \;, \; -\infty < x < \infty$$

$$\Rightarrow \frac{dF(x)}{dx} = f(x)$$

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

$$= P(X \le b) - P(X \le a) = \underline{\underline{F(b) - F(a)}}$$

Covariance

$$Cov(X,Y) = E[\{X-E(X)\}\{Y-E(Y)\}]$$
$$= E(XY) - E(X)\cdot E(Y)$$

if X & Y are independent $\Rightarrow E(XY) = E(X)\cdot E(Y)$
$\Rightarrow Cov(X,Y) = 0$ if X & Y are independent.

1. $Cov(X+a, Y+b) = Cov(X,Y)$
2. $Cov(aX, bY) = ab\, Cov(X,Y)$
3. $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$

Correlation Coefficient $(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \times \sigma_Y}$

* if $U = a_1X_1 + a_2X_2 + \dots + a_nX_n$

$$E(U) = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n)$$
$$U - E(U) = a_1[X_1 - E(X_1)] + a_2[X_2 - E(X_2)] + \dots + a_n[X_n - E(X_n)]$$
$$E[U-E(U)]^2 = a_1^2E[X_1-E(X_1)]^2 + a_2^2E[X_2-E(X_2)]^2 + \dots + a_n^2E[X_n-E(X_n)]^2$$
$$+ 2\sum_i^n\sum_j^n a_ia_j E[\{X_i - E(X_i)\}\{X_j - E(X_j)\}]$$

$$\Rightarrow V(U) = a_1^2V(X_1) + a_2V(X_2) + \dots + a_nV(X_n) + 2\sum_{i=1}^n\sum_{j=1}^n a_ia_j cov(X_i, X_j)$$

$$\Rightarrow V(X_1+X_2+\dots+X_n) = V(X_1) + V(X_2) + \dots + V(X_n) + 2\sum_{i=1}^n\sum_{j=1}^n cov(X_i, X_j)$$

if $X_1, X_2, \dots, X_n$ are independent $\Rightarrow cov(X_i, X_j) = 0$

$$\Rightarrow V(X_1+X_2+\dots+X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

* $V(X_1 \pm X_2) = V(X_1) + V(X_2) \pm 2Cov(X_1, X_2)$

if X_1 & X_2 are independent
$$\Rightarrow V(X_1 \pm X_2) = V(X_1) + V(X_2)$$

Moment generating function ($M_x(t)$)

$$\Rightarrow M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \text{ , continuous}$$
$$= \sum^{\infty} e^{tx} f(x) \text{ , discrete}$$

* $M_{cx}(t) = M_x(ct)$
* $M_{x_1+x_2+\dots+x_n}(t) = M_{x_1}(t)\times M_{x_2}(t)\times\dots\times M_{x_n}(t)$ $\rightarrow$ Additive property of distribution

Fitting a Binomial
↳ take mean & equate with np
the $f_e = N \times P(x)$.

Date:

* Uniqueness theorem of Moment generating function
↳ The moment generating functions of a distribution uniquely determines the distribution.

$$\mu_1' = E(x) = \left.\frac{d(M_x(t))}{dt}\right|_{at\, x=0}$$

$$\mu_2' = E(x^2) = \left.\frac{d^2(M_x(t))}{dt}\right|_{at\, x=0}$$

* Mean $= E(x) = \mu_1'$, * Variance $= E(x^2) - [E(x)]^2 = \mu_2' - (\mu_1')^2$

$$\Rightarrow \mu_r' = E(x^r) = \left.\frac{d^r(M_x(t))}{dt}\right|_{at\, x=0}$$

* Bernoulli Distribution $\xrightarrow[q \to 0]{p \to 1}$, $p+q=1$

$\mu_2' = E(x^2) = 0^2 \times q + 1^2 \times p = \underline{p}$.

$\Rightarrow$ Mean $= E(x) = \underline{\underline{p}}$ & Variance $= E(x^2) - [E(x)]^2 = p - p^2 = p(1-p) = \underline{\underline{pq}}$

* Binomial Distribution

$p(x) = {}^nC_x p^x q^{n-x}$; $x = 0, 1, \ldots, n$; $p+q=1$

$= 0$; otherwise

$\rightarrow \sum P(x) = \sum {}^nC_x p^x q^{n-x} = (p+q)^n = \underline{\underline{1}}$

Conditions

i) Only 2 possible outcomes (p&q) i.e success & faliure

ii) 'n' is finite

iii) each trial is independent of each other

iv) p & q are constant througout experiment.

$\rightarrow \mu_1 = np$, $\mu_2 = n(n-1)p^2 + np$, $\mu_3 = n(n-1)(n-2)p^3 + n(n-1)p^2 + np$ ...

Mean $\underset{E(x)}{} = \mu_1 = \underline{\underline{np}}$, Variance $\underset{V(x)}{} = \mu_2 - \mu_1^2 = \underline{\underline{npq}}$, mode $= [(n+1)p]$,
if integer than
∴ 2 modes $(n+1)p, (n+1)p-1$

$\rightarrow$ Moment generating function

$M_x(t) = E[e^{tx}] = {}^nC_x e^{tx} p^x q^{n-x} = {}^nC_x (e^t p)^x q^{n-x} = (\underline{\underline{pe^t + q}})^n$

$\rightarrow$ reccurence relation $\rightarrow \dfrac{P(x)}{P(x-1)} = \dfrac{n-x+1}{x} \times \left(\dfrac{p}{q}\right)$

Poission Distribution

↳ when $n \to \infty$ & $p \to 0$ but λ (rate) → finite constant $(n \times p)$

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$
$$= 0, \quad \text{otherwise}$$

→ $\sum P(x) = \sum \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right) = e^{-\lambda} \cdot e^{\lambda} = \underline{1}$

→ Mean = $\underline{\lambda}$, Variance = $\underline{\lambda}$

→ Mode = $[\lambda]$, when λ is not integer
$= \underline{\lambda} \;\&\; \underline{\lambda - 1}$

→ Reccurence relation ⇒ $\frac{P(x)}{P(x-1)} = \frac{\lambda}{x}$

→ MGF $(M_x(t))$ ⇒ $M_x(t) = E(e^{tx}) = \sum e^{tx} \times \frac{e^{-\lambda}\lambda^x}{x!}$

$$\Rightarrow M_x(t) = e^{-\lambda} \sum \frac{(\lambda e^t)^x}{x!} = e^{-\lambda}\left(e^{\lambda e^t}\right)$$

$$\Rightarrow M_x(t) = e^{-\lambda} \cdot e^{\lambda e^t} \Rightarrow e^{\lambda(e^t - 1)}$$

Negative Binomial Distribution

→ $f(x; r, p)$ → probability of 'x' failures preceeding r^{th} success in $(x+r)$ trails.

$\underbrace{\qquad\qquad}_{\substack{x \text{ failures} \\ r-1 \text{ success}}} \uparrow$

$$\Rightarrow f(x) = {}^{x+r-1}C_{r-1}\, p^{r-1} q^x, \quad x = 0, 1, \ldots$$
$$= 0, \quad \text{otherwise}$$

let ⇒ $P = \frac{1}{Q}$ & $q = \frac{P}{Q}$

⇒ M.G.F $= (Q - Pe^t)^{r}$, Mean $(E(x)) = rP$, Variance $(V(x)) = \underline{rPQ}$

Page No.:
Date:
YOUVA

Geometric Distribution

→ probability that there are 'x' failures preceeding first sucess

$P(x) = q^x p \quad , x = 0, 1, 2, \ldots$
$= 0 \quad$ otherwise

$M.G.F (M_x(t)) = E[e^{xt}] = \sum q^x p \times e^{xt}$
$= \sum (qe^t)^x p = p \times \frac{1}{1-qe^t} = \frac{p}{1-qe^t}$

* Mean $= \frac{q}{p}$, Variance $= \frac{q}{p^2}$

Lack of Memory

$\Rightarrow P(Y=t \mid X \geq K) = pq^t$, where $Y = X - K$

Normal Distribution

$f(x;\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ → P.d.f of Normal Distribution $\sim N(\mu, \sigma^2)$

→ standard Normal Variate $(z) = \frac{x-\mu}{\sigma}$

$\Rightarrow E(Z) = 0$ & $Var(Z) = 1 \Rightarrow Z \sim N(0,1)$

P.d.f of $z \Rightarrow \phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$ $\quad \phi(-z) = P(Z \leq -z) = P(Z \geq z) = 1 - P(Z \leq z) = 1 - \phi(z)$

distribution function of $z \Rightarrow F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$

$F(-z) = 1 - F(z)$

→ $[P(x)]_{max}$ at $x = \mu$

→ $\mu_{2r+1} = 0$ & $\mu_{2r} = 1 \times 3 \times 5 \times \ldots (2r-1)\sigma^{2r}$

→ point of inflection $\Rightarrow x = \mu \pm \sigma$

→ Mean, Mode & Median coincide & are equal to μ.

→ $MGF(M_x(t)) = e^{\mu t + t^2\sigma^2/2}$ of $f(x)$

$MGF(M_x(t)) = e^{t^2/2}$ of standard normal variate $(z = \frac{x-\mu}{\sigma})$

→ Mean at point A $\Rightarrow \mu_1 = \mu_1' + A$

* Central limit theorem

if $x_1, x_2, \ldots, x_n$ are independent random variables with $\underset{(x_i)}{\text{Mean}} = \mu_i$ & $\underset{(x_i)}{\text{Variance}} = \sigma_i^2$

then under general conditions

$S_n = \sum^{N} x_i$ follows normal distribution $n \to \infty$

with $\mu = \sum \mu_i$ & $\sigma^2 = \sum \sigma_i^2$

→ $n \geq 30$ is considered large.

* Continuous approximation of x.

$P(X \leq x) = P(X < x + 0.5)$

$P(x \leq X) = P(x - 0.5 < X)$

$P(x = X) = P(x - 0.5 < X < x + 0.5)$

* Estimator & Sampling Distribution

→ Standard deviation of $\bar{x}$ = $\dfrac{\text{Population S.D}}{\sqrt{\text{Sample size}}} = \dfrac{\sigma}{\sqrt{n}}$

→ S.N.V $(t) = \dfrac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ → mean; $\left(\frac{\sigma}{\sqrt{n}}\right)$ → standard deviation of $\bar{x}$

→ Population proportion (P) & sample proportion (p)

Mean of $p = P$

Standard deviation of $p = \sqrt{\dfrac{PQ}{n}}$

→ S.N.V $(t) = \dfrac{p - P}{\sqrt{\frac{PQ}{n}}}$

→ Standard Error = $\sqrt{\dfrac{PQ}{n}}$ & estimate of S.E = $\sqrt{\dfrac{pq}{n}}$

→ Unbiased estimator of population variance (σ^2) = $\left(\dfrac{n}{n-1}\right) s^2$

where $s^2 = \dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2$

Page No.:
Date:

* Confidence level

i) 95% $\Rightarrow \left(p - 1.96\sqrt{\frac{pq}{n}},\ p + 1.96\sqrt{\frac{pq}{n}}\right)$

99% $\Rightarrow$ 2.58

for almost certainity $\Rightarrow$ 3

* Maximum liklyhood

$L = \prod_{i=1}^{N} f(x_i, \alpha)$

i) take log on both sides $\ln L = \sum_{i=1}^{N} \ln(f(x_i, \alpha))$

ii) $\left.\frac{\partial \ln L}{\partial \alpha}\right|_{max} = 0$

iii) we have value of parameter α.

* Unbaisness $(E(T(x)))$
* Efficiency → i) Unbaisedness
  ii) min. variance

* M.V.U.E → PATA NHI KYA HAI YE ?

M	T	W	T	F	S	S

Page No.:
Date:
YOUVA

Chi-square test (χ^2) $\xrightarrow{*}$ No parent parameters

(Non-parametric tests & Distribution-free test).

* goodness of fit

conditions $\Rightarrow$ $N > \underline{50}$ & $f_i \nless \underline{5}$

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O)_i - (E)_i}{(E)_i} \right] \quad \& \quad \sum_{i=1}^{N} O_i = \sum_{i=1}^{n} E_i$$

* for a 2x2 table

a	b	a+b
c	d	c+d
a+c	b+d	

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(a+c)(c+d)(b+d)} \quad , \quad \chi^2 = \frac{N\left[\,|ad-bc| - \frac{N}{2}\right]^2}{(a+b)(a+c)(b+d)(c+d)}$$