Some Important Formula

Unit - III

* Correlation Coefficient [Quantitative data]

$$r_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$-1 \leq r_{xy} \leq 1$$

* Rank co-relation coefficient [Qualitative Data]

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$ [Non- tied Rank]

where d = Rank in x - Rank in y

$$r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12} m_1(m_1^2-1) + \frac{1}{12} m_2(m_2^2-1) + \cdots\right]}{n(n^2-1)}$$

where m_1, m_2 & m_3 are the repetition of a number.

* Regression line

→ y on x

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

where $$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

→ x on y

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

where $b_{xy} = \dfrac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$

* Curve Fitting

→ Straight line

$$y = a + bx$$

The normal equations are

$$\sum y = an + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

Note: To make the normal equ^n Multiply the coefficient of constant & take the summation

→ Second Degree Parabola

$$y = a + bx + cx^2$$

The normal equ^n are

$$\sum y = an + b\sum x + c\sum x^2$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

→ $y = ae^{bx}$

$$\log y = \log a + bx$$

$$Y = A + bx$$

The normal equ^n are

$$\Sigma y = An + b\Sigma x$$

$$\Sigma xy = A\Sigma x + b\Sigma x^2$$

* Moments about Mean

$$\mu_r = \frac{\Sigma (x_i - \bar{x})^r}{n}$$ [If data is given in series form]

$$\mu_r = \frac{\Sigma f_i (x_i - \bar{x})^r}{\Sigma f_i}$$ [If data is given in frequency form]

* Skewness

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

+ve skewed if $\beta_1 > 0$

-ve skewed if $\beta_1 < 0$

Symmetric if $\beta_1 = 0$

* Kurtosis

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

Meso-Kurtosis if $\beta_2 = 3$

Lapto-Kurtosis if $\beta_2 > 3$

Platy-kurtosis if $\beta_2 < 3$

* Karl Pearson's coefficient

$\gamma_1 = \sqrt{\beta_1}$ & $\gamma_2 = \beta_2 - 3$

* Multiple Linear Regression

If x depends on y & z

$$x = a + by + cz$$

& the normal eqn are

$$\Sigma x = an + b\Sigma y + c\Sigma z$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2 + c\Sigma yz$$

$$\Sigma xz = a\Sigma z + b\Sigma yz + c\Sigma z^2$$

It is very similar as curve fitting.

Unit . IV

* Bionomail Distribution

Probability of r success

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

where n = no. of trial

p = probability of sucess

q = " " failure

The general form of Bionomail distribution

$(q+p)^n$

→ If trials are repeated N times then the general form will be $N.(q+p)^n$

& Probability of r sucess will becomes Frequency of r sucess & can be calculated as $N.P(r)$.

i.e $N.\ {}^nC_r\, p^r q^{n-r}$

* Note: → $P(0) + P(1) + P(2) + \dots + P(n) = 1$

→ At least 2 sucess i.e [lower limit given]

$P(2) + P(3) + \dots + P(n)$

→ At most 2 sucess i.e [upper limit given]

$P(0) + P(1) + P(2)$

→ Exactly 2 sucess

$P(2)$

→ $p+q=1$

→ Bionomail Distribution is applicable only when

n - finite

p - constant

* Poisson Distribution

Probability of r sucess

$$P(r) = \frac{e^{-\lambda} \lambda^{r}}{r!}$$

where $\lambda = np$ = Mean, it is given either in question or you have to calculate $n \cdot p$

→ Frequency of r sucess $N \cdot P(r)$

$$= N \cdot \frac{e^{-\lambda} \lambda^{r}}{r!}$$

→ Applicable only when p is small & n is large.

* Normal Distribution

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

where μ = Mean

σ = std deviation

→ You have to use table of Normal Distribution

* Addition Law of Probabibity

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

→ If Events A & B are ~~independent~~ mutually exclusive, then

$$P(A \cap B) = 0$$

$$\Rightarrow \quad P(A \cup B) = P(A) + P(B)$$

* Multiplication Law of Probability

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$$

If events are independent, then

$$P(A/B) = P(A) \quad \& \quad P(B/A) = P(B)$$

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

* Baye's theorem

$$P\left(\frac{E_i}{A}\right) = \frac{P(A) \cdot P\left(\frac{A}{E_i}\right)}{\Sigma \, P(A) \cdot P\left(\frac{A}{E_i}\right)}$$

* Random Variable : $S \rightarrow R$

Discrete R. V
→ If Range set is finite

Probability Mass function
→ (i) $\Sigma \, p_i = 1$
(ii) $p_i > 0$

Continous R.V.
→ If Range set is an Interval.

Probability Density function
(i) $f(x) > 0$
(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

* Expectation

$$E(X) = \Sigma x_i p_i$$

* Variance

$$Var(X) = E(X^2) - [E(X)]^2$$

Unit - V

* t- test [If sample size $n < 30$ & compare the Mean]

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad \text{or} \quad \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

where μ = Mean of Population

S = Std deviation of Population

n = Sample size

s = Std deviation of Sample

$\bar{x}$ = Mean of Sample

→ If data is given

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

* Relation b/w s & S

$$(n-1) S^2 = n s^2$$

* Degree of freedom = $n-1$

* ~~F-test [Comp~~

* For two samples

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where } S = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

or

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

$\bar{x}$ = Mean of sample 1

n_1 = Size of sample 1

$\bar{y}$ = Mean of Sample 2

n_2 = Size of Sample 2

D.o.f = $n_1 + n_2 - 2$

* F - Test [Compare the variance]

$$F = \frac{S_1^2}{S_2^2} \quad [S_1^2 > S_2^2]$$

$$S_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} \quad \text{or} \quad \frac{n_1 s_1^2}{(n_1 - 1)}$$

$$S_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 2} \quad \text{or} \quad \frac{n_2 s_2^2}{n_2 - 1}$$

D.o.f is ~~$n_1 + n_2 - 2$~~ $(n_1 - 1), (n_2 - 2)$

* Chi - Square Test [Compare the frequency]

$$\chi^2 = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right)$$

O_i = Observed Frequency

E_i = Expected Frequency

Expected frequency can be calculated as

→ Average — D.o.f. $n-1$

→ B.D — $n-1$

→ P.D. — $n-2$

→ N.D. — $n-3$

Contingency Table

a	b	a + b
c	d	c + d
a + c	b + d	N = a + b + c + d

Expected Frequency

$\frac{(a+b)(a+c)}{N}$	$\frac{(a+b)(b+d)}{N}$
$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(c+d)}{N}$

D.o.f is $(m-1)(n-1)$ where

m = no. of rows

n = no. of columns

* Anova Test

	Sum of Square	D.o.f	Mean of sum of square	F =
within Sample				
b/w Sample				
Total				

* Control charts

→ $\bar{X}$ Chart [Mean]

$CL = \bar{\bar{X}}$

$LCL = \bar{\bar{X}} - A_2 \bar{R}$

$UCL = \bar{\bar{X}} + A_2 \bar{R}$

→ R chart [Range]

$CL = \bar{R}$

$LCL = D_3 \bar{R}$

$UCL = D_4 \bar{R}$

→ C chart

$CL = \bar{c} = \frac{\Sigma c}{n} = \frac{\text{Sum of defect is sample}}{\text{Total no of sample}}$

$LCL = \bar{c} - 3\sqrt{c}$

$UCL = \bar{c} + 3\sqrt{c}$

→ np chart

$$CL = n\bar{p}$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$

where $\bar{p} = \dfrac{\text{No. of defective items}}{\text{Total no. of inspected items}}$

→ p chart

$$CL = \bar{p}$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$