Unit 2

Short Notes of Statistics

Mean

- Simple mean $\bar{X} = \frac{\Sigma X}{N}$

- Simple mean when class interval is given with frequency distribution:-

e.g.

Class	freq
10-20	5
30-40	10

In this case, $\bar{X} = \frac{\Sigma fx}{(\Sigma f = N)}$ ← also called Direct method

- mean by shortcut method

$$\bar{X} = \underset{\text{Assumed mean}}{A} \pm \frac{\Sigma fd}{N \to \Sigma f}$$

d = difference = X − A

f = frequency

- weighted Arithmatic mean

$$\bar{X} = \frac{N_1X_1 + N_2X_2}{N_1 + N_2}$$

- Step-Deviation method

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times C$$

C ← common diff. b/w the class (or) Step deviation

$d = X - A$, $d' = \frac{X-A}{C}$

Median

(i) In case of odd Items

$$\text{Median} = \text{Size of} \left[\left(\frac{N+1}{2}\right)\right]^{th} \text{item}$$

(ii) In case of even items

$$\text{median} = \text{Size of} \left[\frac{\left(\frac{n}{2}\right)^{th} \text{item} + \left(\frac{n}{2}+1\right)^{th} \text{item}}{2}\right]$$

- Median of Discrete Series:-

$$\text{Med} = \text{Size of} \left(\frac{N+1}{2}\right)^{th} \text{item} \to \text{odd}$$

$$\text{Med} = \text{Size of} \left(\frac{N}{2}\right)^{th} \text{item} \to \text{even}$$

e.g.

Size of pen	f	CF
2	10	10
3	20	30
4	25	55

odd

$M = \text{Size of} \left(\frac{N+1}{2}\right)^{th}$

$= \text{Size of} \left(\frac{55+1}{2}\right)$

$= 28^{th}$ item

$= 20$ Ans

- Median for continuous series.

$$\text{Med.} = L_1 + \frac{\frac{N}{2} - \text{C.f}}{f} \times (i)$$

L_1 → Lower Limit of median class

i → class interval of median class

e.g.

class
10-20 → i = 10
30-40

Mode

$$\text{Mode} = L_1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i\right)$$

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

Dispersion

Std. Deviation :- (σ)

(a) Std. Deviation from actual mean

↳ used when the mean is a whole No.

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

(b) when the mean is fractional value

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} \quad d = X - A$$

For Discrete Series

(a) Actual mean method:-

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}} \quad d = x - \bar{x}$$

(b) Assumed mean method:-

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \quad d = x - A$$

Coefficient of variation

Population

$$CV = \frac{\sigma}{\mu} \times 100\%$$

σ ← popular std Deviation; μ ← popular mean

Sample

$$CV = \frac{s}{\bar{x}} \times 100\%$$

s ← sample std; x̄ ← sample mean.

For continuous series → e.g. 10-20, 30-40

shortcut method

$$\text{Mean } (\bar{x}) = A + \frac{\Sigma fd'}{N} \times c$$

shortcut metho

$$\sigma \text{ (std. Deviation)} = \sqrt{\frac{\Sigma fd'^2}{N \to \Sigma f} - \left(\frac{\Sigma fd'}{N}\right)^2} \times c$$

Correlation and Regression

Methods of finding correlation (r) coeff. (correlation coefficient)

1. Karl pearson's product moment method
2. Spearman's Rank correlation coeff

etc.

(a) when direct value are given use this

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

$$cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

(b) when large variable are given then use,

x	y
–	–
–	–
–	–

Here n=3

Direct method (n: no. of pair of observation)

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

Shortcut method (or Assumed mean)

$$r = \frac{n\Sigma dx dy - \Sigma dx \Sigma dy}{\sqrt{n\Sigma dx^2 - (\Sigma dx)^2}\sqrt{n\Sigma dy^2 - (\Sigma dy)^2}}$$

Coefficient of correlation is the product of two regression coefficients

$$r = \pm\sqrt{b_{xy} \times b_{yx}}$$

Rank correlation

$$\text{ranking} = 1 - \left(\frac{6 \times \Sigma d^2}{n(n^2-1)}\right)$$

6 ← sif

$d = R(x) - R(y)$; R(x) ← Rank of x

Regression Analysis

Regression line of y on x

① $y - \bar{y} = b_{yx}(x - \bar{x})$

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

Regression line of x on y

① $x - \bar{x} = b_{xy}(y - \bar{y})$

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

To apply various test Check,

Testing of Hypothesis

Short notes

① If $n < 30$ apply t-test.

② If $n > 30$ apply z-test

③ If variance is given apply → f-test

④ If single attribute is given - 1 way ANOVA (for one factor)

⑤ If Double attribute is given - 2 way ANOVA (for two factor)

⑥ For categorical sample → Chi-square

Estimation

$$\boxed{\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}}$$

Student's t-test

① To test the significance of the mean of a random sample

$$t = \frac{(\bar{x} - \mu)}{S} \times \sqrt{n}, \quad S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{\underbrace{(n-1)}_{\nu}}}$$

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{n}}$$

Confidence interval estimate (for α LOS)

② one tailed test $\boxed{\bar{x} \pm t_\alpha \frac{S}{\sqrt{n}}}$ $\alpha = 0.05$

② Two test the difference b/w mean of two samples (Independent samples)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S} \times \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}}$$

$$S = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{\underbrace{(n_1 + n_2 - 2)}_{\nu}}}$$

③ To test the significance of an observed correlation coefficient.

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

r → correlatn coeff

n → size of sample

Chi-square test

① (for two categorical sample) formula $\chi^2 = \frac{\Sigma(O - E)^2}{E}$

O → observed freq, E → Expected freq.

$$E = \frac{RT \times CT}{N}$$

RT → Row total

CT → column total

N → Total no. of observat

$\nu = (r-1)(c-1)$ → DOF

② χ^2 test for populatn variance.

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2} \times (n-1), \quad \nu = n-1$$

σ_s^2 → sample variance

σ_p^2 → populatn variance

F-test

$$F = \frac{\sigma_1^2}{\sigma_2^2} \text{ OR } F = \frac{S_1^2}{S_2^2}$$

Larger estimate variance

Smaller estimate variance

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{n}, \quad S^2 = \frac{\Sigma(x - \bar{x})^2}{n-1}$$

DOF

V_1 (numerator) $= n_1 - 1$ (larger variance)

V_2 (Denom.) $= n_2 - 1$ (smaller variance)

one way ANOVA

Source of variation	Sum of Square	Degree of freedom	mean square	F
B/w the Sample	SSC =	$V_1 = c-1$	$MSC = \frac{SSC}{V_1}$	$\frac{MSC}{MSE} = F$
within the sample	SSE =	$V_2 = n-c$	$MSE = \frac{SSE}{V_2}$	

$$SSC = \Sigma(\bar{x}_A - \bar{\bar{x}})^2 + (\bar{x}_B - \bar{\bar{x}})^2 + \text{----}$$

$$SSE = \Sigma(A - \bar{X}_A)^2 + \Sigma(B - \bar{X}_B)^2 + \text{----}$$

Two way ANOVA

Source of variation	Sum of Square	DoF	mean sum of square	Ratio of F
B/w the column	SSC =	$V_1 = c-1$	$MSC = \frac{SSC}{V_1}$	$\frac{MSC}{MSE}$
B/w the Row	SSR =	$V_2 = r-1$	$MSR = \frac{SSR}{V_2}$	$\frac{MSR}{MSE}$
Residual (or) error	SSE =	$V = (c-1) \times (r-1)$	$MSE = \frac{SSE}{V}$	
	SST =			

$$SSC \Rightarrow \frac{(\Sigma A)^2}{n_A} + \frac{(\Sigma B)^2}{n_B} + \text{----} - \text{Correction Factor}\left(\frac{T^2}{N}\right)$$

$$SSR \Rightarrow \frac{(\Sigma P)^2}{n_P} + \frac{(\Sigma Q)^2}{n_Q} + \text{----} - \left(\frac{T^2}{N}\right)$$

SST ⇒ Square of all elements eg. $(4)^2 + (5)^2 + (6)^2$ ---- $-\left(\frac{T^2}{N}\right)$

$$SSE \Rightarrow SST - (SSC + SSR)$$

Statistical quality control

$\bar{X}$-chart → (mean, Average)

R-chart → (Range)

P-chart → (fraction/proportion)

np-chart → (number of diffective)

C-chart → (Defects)

NOTE :-

1. P-chart and np chart only tells that whether any product is diffective or not
2. C-chart not only tells the defect of product but also shows the no. of defects

To find

P-chart

Distinct - Distinct

Sample No. (m)	Item inspected (n)	Defective Items (d)	$P = \frac{d}{n}$

np-chart

Sample No. (m)	Sample of each (n)	No. of defective Items (np)

C-chart (Tells No. of defects also)

Sample No. (m)	No. of defects (c)

$\bar{X}$-chart / R-chart

values like [$n=4$, $A_2=0.5$, $D3=4$, $D4=2$] will be given

Formulas

	C.L (central line)		upper central lim (UCL)	Lower central lim (LCL)
P-chart	$\bar{P} = \frac{\Sigma P}{m}$	$P = \frac{d}{n}$	UCL = $\bar{P} + 3 \times \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$	LCL = $\bar{P} - 3 \times \sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$
np-chart	$n\bar{p} = \frac{\Sigma np}{m}$	$\bar{P} = \frac{n\bar{p}}{n}$	$n\bar{p} + 3 \times \sqrt{n\bar{p}(1-\bar{p})}$	$n\bar{p} - 3 \times \sqrt{n\bar{p}(1-\bar{p})}$
C-chart	$\bar{c} = \frac{\Sigma c}{m}$		$\bar{c} + 3\sqrt{\bar{c}}$	$\bar{c} - 3\sqrt{\bar{c}}$
$\bar{X}$-chart	$C.L = \bar{\bar{X}}$	$\bar{\bar{X}} = \frac{\Sigma \bar{X}}{m}$	$\bar{\bar{X}} + A_2\bar{R}$	$\bar{\bar{X}} - A_2\bar{R}$
R-chart	$C.L = \bar{R} = \frac{\Sigma R}{m}$	$\Sigma R = X_H - X_L$	$D_4\bar{R}$	$D_3\bar{R}$

X_H → Higher value