MAT132

Unit - III →

($\bar{x}$) Arithmetic mean :-

1. Ind obs → $\bar{x} = \frac{\sum_1 x_i}{n}$

2. Dis obs → $\bar{x} = \frac{\sum_i f x}{\sum_i f}$ → frequency

For unequal distribution data $A + \frac{\sum_i fd}{\sum_i f}$ (i) will not come

3. cont. obs → $\bar{x} = A + \frac{\sum_i fd}{\sum_i f} \times i$ → length of the CI

A → Assumed mean (middle value)

$d = \frac{m - A}{i}$

m → middle value of CI

i → length of CI.

4. combined mean → $\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \cdots + n_k \bar{x}_k}{n_1 + n_2 + \cdots + n_k}$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

Median :-

1. Ind obs → size of $\left(\frac{N+1}{2}\right)^{th}$ item

N → No. of observations

2. Dis obs → size of $\left(\frac{N+1}{2}\right)^{th}$ item.

N → sum of all frequencies

i.e; $N = \sum_i f$

3. cont. obs → $\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i$

N → $\sum_i f$

cf → cummulative frequency

L → lower limit of median class

f → frequency of median class.

i → length of CI

4. $Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times i$ (I^{st} quartile)

5. $Q_3 = L + \frac{\frac{3N}{4} - cf}{f} \times i$ (3rd quartile)

6. Median means middle value of the data.

7. cf how to calculate means 1st frequency v shld write and then v shld add then v shld c the frequency and the last cf. v shld calculate $\frac{N}{2}$ from the getting value v shld c the nearest no. in the cf then that line only v shld take f and lower limit. and the above no. is cf.

Mode:-

1. It occurs which most highest frequency [ie; which is the maximum frequency that shld b considered as the mode].

2. Relationship b/w mean, median, mode is

Mode = 3 median - 2 mean.

3. cont. obs → $\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$ → length of CI

where $\Delta_1 = f_1 - f_0$ → b4
f_1 → highest frequency

$\Delta_2 = f_1 - f_2$ → after.

$\Delta_1 = f_1 - f_0$, $\Delta_2 = f_1 - f_2$

Geometric mean:-

$\sqrt{(-a)(-b)} = \sqrt{ab}$

$\sqrt{(-1)^2} = 1 \Rightarrow i = -\sqrt{1}$

Geometeric mean of combined group:-

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2 + \cdots + n_k \log G_k}{n_1 + n_2 + \cdots + n_k}$$

Harmonic mean:-

1. Ind obs $\rightarrow$ $HM = \dfrac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$

2. $HM = \dfrac{1}{\frac{1}{N}\sum_{i=1}\left(\frac{f_i}{x_i}\right)}$ $\rightarrow$ where $N = \sum_i f_i$.

Methods of dispersion:-

1. Range $\rightarrow$ $L - S$.

Co-eff of range $\rightarrow$ $\dfrac{L-S}{L+S}$

2. Quartile deviation $(QD) = \dfrac{Q_3 - Q_1}{\cancel{Q_3 + Q_1}\ 2}$

Co-eff of QD $\Rightarrow$ $\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

3. Mean deviation $\rightarrow$ $\dfrac{1}{N}\sum_i f_i |x_i - \alpha|$ $MD = \dfrac{\sum f |m - \cancel{a}|}{\sum f}$

Coeff of MD $\rightarrow$ $\dfrac{MD}{mean}$

Coeff of median $\Rightarrow$ $\dfrac{MD}{median}$

Coeff of mode $\rightarrow$ $\dfrac{MD}{mode}$.

4. Standard deviation $\rightarrow$

$$\sigma = \sqrt{\frac{1}{N}\sum_i f_i (x_i - \bar{x})^2}$$

$N = \sum_i f_i$.

$\sigma^2 \rightarrow$ variance.

5. Equivalent formula for SD $\rightarrow$

$$\sigma = \sqrt{\frac{\sum_i f d^2}{\sum_i f} - \left(\frac{\sum_i f d}{\sum_i f}\right)^2} \times i$$

$$d = \frac{m - A}{i} \qquad d = x_i - \bar{x}.$$

6. co-eff of variation $\rightarrow$ $cov(x) = \frac{\sigma}{\bar{x}} \times 100$.

7. combined SD $\rightarrow$

$$\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + \dots + n_k(\sigma_k^2 + d_k^2)}{n_1 + n_2 + \dots + n_k}.$$

(or)

$$\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2 + \dots}$$

where, $d_1 = \bar{x}_1 - \bar{x}$

$d_2 = \bar{x}_2 - \bar{x}$

$d_3 = \bar{x}_3 - \bar{x}$.

Moments:-

1. Relationship b/w μ_r and μ_r' $\rightarrow$ $\bar{x} - A = \mu_1'$.

 * $\mu_1 = \mu_1' - \mu_1' = 0$.

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

3. Kurtosis $\rightarrow$ $\beta_2 = \frac{\mu_4}{\mu_2^2}$

3. $\mu_1 = \mu_1' - \mu_1' = 0$

4. $\mu_2 = \mu_2' - (\mu_1')^2$

large → cv (x); individual $n_1\bar{x}_1$, $n_2\bar{x}_2$ quata ^ ave - combined, ^ SD.

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

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

correlation :-

1. $$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2}\sqrt{N\sum y^2 - (\sum y)^2}}$$

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

(or) $$r = \frac{N\sum dx\,dy - \sum dx\,dy}{\sqrt{N\sum dx^2 - (\sum dx)^2}\sqrt{N\sum dy^2 - (\sum dy)^2}}$$

$\mu_1 = 0$

$\mu_2 = \frac{\sum fd^2}{\sum f}$

$\mu_3 = \frac{\sum fd^3}{\sum f}$

$\mu_4 = \frac{\sum fd^4}{\sum f}$

2. $$r = \frac{N\sum f\,dx\,dy - \sum f\,dx \sum f\,dy}{\sqrt{N\sum f dx^2 - (\sum f dx)^2}\sqrt{N\sum f dy^2 - (\sum f dy)^2}}$$

where; $n = \sum f$

$dx = \frac{y - A}{h}$

$dy = \frac{y - B}{k}$

3. For finding corrected value. the 1st box is taken 1st and in that the value of x sh'd be taken as (−) and correct value ie; 2nd box x we shld take (+). vice versa [for wrong box x is (−) ve & correct box x is (+) ve.

For correct value y is :-

wrong value

x	y
2	6
4	8

corrected value

x	y
9	8
7	6

−2 − 4 + 9 + 7

~~−6~~ − ~~8~~ + 8 + 6

~~−14~~ ~~−7+~~

4. Rank correlation →

$$\rho = 1 - \frac{6\{\sum_i D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \dots\}}{N(N^2 - 1)}$$

where; $D = R_x - R_y$.

m = No. of times a particular rank gets repeated

5. If three Ranks given means

$D_1 = R_x - R_y$, $D_2 = R_x - R_z$, $D_3 = R_y - R_z$

ρ_{xy}, ρ_{yz}, ρ_{xz} ⇒

ρ_{xy}, ρ_{xz}, ρ_{yz}.

Regression lines:

1. x depends on y.

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

where $b_{xy} = \frac{N\sum_i xy - \sum_i x \sum y}{N\sum_i y^2 - (\sum_i y)^2}$

$r \cdot \frac{\sigma_x}{\sigma_y}$

2. y depends on x.

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

where; $b_{yx} = \frac{N\sum_i xy - \sum x \sum y}{N\sum_i x^2 - (\sum_i x)^2}$

$r \cdot \frac{\sigma_y}{\sigma_x}$

3. Co-eff of regression → $r^2 = b_{xy} b_{yx}$.

4. If both b_{xy} and b_{yx} are (-)ve the $r = -ve$.

5. For finding regression lines $\bar{x} = \frac{\Sigma x}{N}$, $\bar{y} = \frac{\Sigma y}{N}$

6. Given x, find y means use y on x

7. Given y, find x means use x on y.

8. $b_{yx} = \frac{N \Sigma dx dy - \Sigma dx \Sigma dy}{N \Sigma dx^2 - (\Sigma dx)^2}$

UNIT-IV.

* Large sample test →

1. Test for specified mean

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

means they will give one mean, one SD, one 'n'. and one population.

2. Test for ~~spe~~ equality of two means :→

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

means they will give 2 sample means, 2 'n''s, 2 population, 2 SD.

* Single tailed ⇒ $>, <$
* Double tailed ⇒ $=$ (constant)
* large sample test (≥ 30)
* Small sample test (≤ 30)

3. Test for specified proportion →

$$z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}}$$

p → sample proportion

P_0 → popu. propo.

$Q = 1 - P$.

4. Test for equality of two proportions →

$H_0 \Rightarrow P_1 = P_2$ → proportion of Ist

$H_1 \Rightarrow P_1 \neq P_2$ → " " " 2nd

$$Z = \frac{P_1 - P_2}{\sqrt{PQ \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where;

$Q = 1 - P$.

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

5. Steps to write LST →

- Step 1 → State the H_0
- Step 2 → Decide H_1
- Step 3 → choose the significance level
- Step 4 → calculate the test static
- Step 5 → Find the table value of z and decide wheather the sample static falls within the region of acceptance or rejection.

6. 1% level → 2.58.

7. 5% level → 1.96

Z, t value < than the table value means accept H_0 or else reject H_0.

Small sample test

1. test specific

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$ with $(n-1)$ d.f.

$\bar{x}$ → sample mean

μ → population.

$$\bar{x} = \frac{x_i}{n} \qquad s = \frac{\sum_1 (x - \bar{x})^2}{n-1}$$

$$s^2 = \frac{1}{n-1} \underset{\substack{\downarrow \\ \text{Average.}}}{(x_i - \bar{x})^2} \Rightarrow s^2 = \frac{(x-\bar{x})^2}{n-1}$$

2. Test for diff. in means →

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ → 2 means, 2 'n'. with $(n_1 + n_2 - 2)$ df.

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2 + \sum_i (y_j - \bar{y})^2}{n_1 + n_2 - 2 \text{ df}}$$

3. Paired t-test. → (differences)

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$ with $(n-1)$ df.

$$s = \frac{1}{n-1}\left\{ \sum (d-\bar{d})^2 \right\} = \boxed{\frac{1}{n-1}\left\{ \sum_i d^2 - \frac{(\sum_i d)^2}{n} \right\}}$$

$$d = x - y.$$

$$\bar{d} = \frac{d \text{ total}}{\text{Total no. of obs.}}$$

4. t-test for testing significance of an observed correlation → (correlation nu given na use this formula)

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$ with $(n-2)$ df

co-eff of correlation → r.

⇒ ~~For correlation v shld take n as $\sqrt{n}$~~

F-test → variance nu irukum apdina athu with (n_1-1, n_2-1) df vanthu f-test

1. $F = \frac{S_1^2}{S_2^2}$

$S_1^2 = \frac{n_1 s_1^2}{n_1-1}$, $S_2^2 = \frac{n_2 s_2^2}{n_2-1}$

always $F \geq 1$

→ Test of equality of two population variance.

$$S_1^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$S_2^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2$$

If they didn't give s_1 and s_2

chi-square test → Avungale kuduthuruppang

1. $$\chi^2 = \sum \left(\frac{(O_i - E_i)^2}{E_i}\right)$$ with $(n-1)$ df.

O_i - observed ie; given

$E_i = \frac{\text{Add all the } O_i}{\text{Total no. of observation}}$

E_i - same value.

$O_i - E_i \rightarrow (O_i - E_i)^2 \rightarrow \frac{(O_i - E_i)^2}{E_i}$

2. contingency table

a	b
c	d

In case of 2|2.

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)} \quad \text{with } (r-1)(c-1)$$

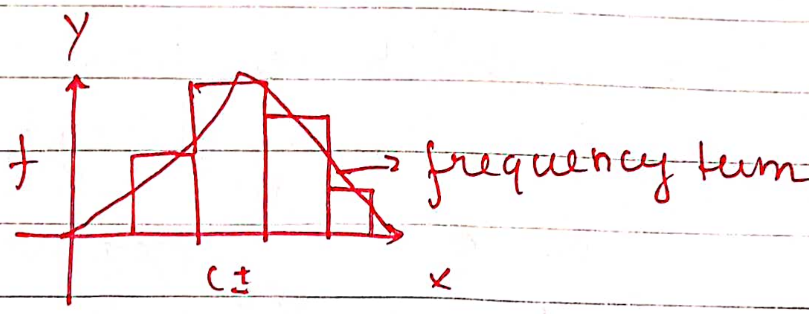
row, column

$N = a+b+c+d.$

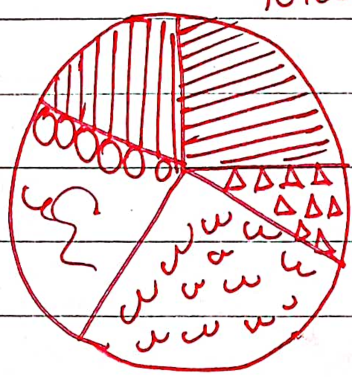
3. For 1 df at 5% level is 3.54.
4. For 5 df at 5% level is 11.07.

Unit - II

1. Histogram →



2. Pie-chart → $\frac{x}{\text{Total}} \times 360$



3. Mode graphically

$$L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$\Delta_1 = f_1 - f_0$

$\Delta_2 = f_1 - f_2$

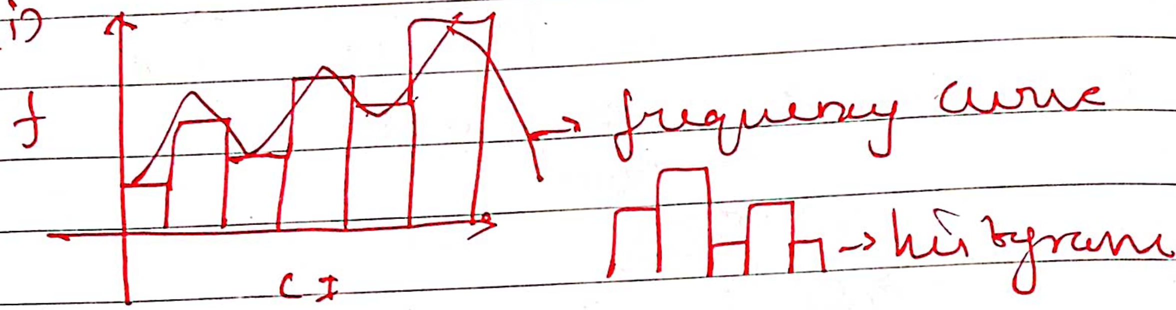
4. Median graphically (OGIVE CURVES).

cf (<) , cf (>)
↓
How to find means normally how we will do like that only. For cf (>) we shld take 59 - minus (f).

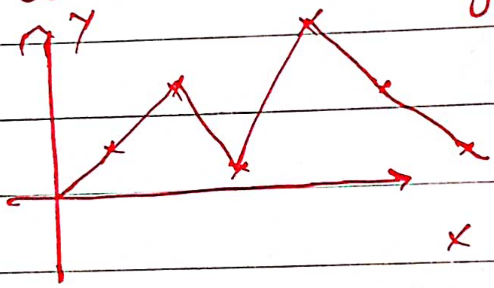
* lower limit of CI > cf
  lower limit and > cf.
* upper limit of CI < cf
  upper limit and cf <.

5. construct the histogram & frequency curve (polygon) ->

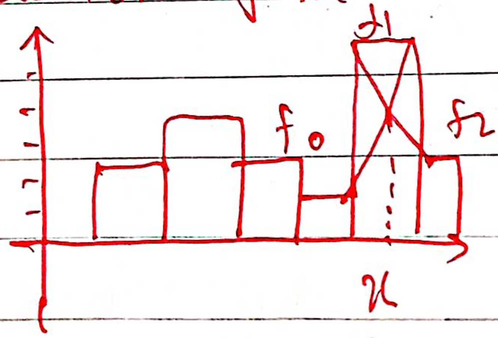
(i)



(ii) without histogram ->



6. calculation of mode



↓
mode

7. Lorenz curve:→

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

8. How will u construct an histogram?

* On the vertical axis, place frequencies.
* On the horizontal axis, place the lower value of each interval.
* Draw a bar extending from the lower value of each interval to the lower value of the next interval.

9. What is mean by OGIVE CURVES and advantage of OGIVE CURVES, how will u draw the OGIVE CURVES?

An OGIVE CHART is a curve of the cf distribution. For drawing such a curve, the frequencies must be expressed as a % of the total frequency.

10. construct a ~~histogram~~ piechart?

* convert the data to %. The 1st step is to convert the data to %.
* calculate the angle for each pie segment A complete circle has 360°.
* Draw the piechart. ~~For~~
* Add labels. it requires a title and lables.

11. construct a frequency table either

univariate , bivariate set of data.