Unit 2: Statistics

16 April 2023 14:18

① Arithmatic Mean of Ungrouped Data

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$ (no. of obs.)

② Mean of Grouped Data :

$$\boxed{\bar{x} = \frac{\sum f_i \cdot x_i}{N}}$$

$N = \sum f_i$

$$x_i = \frac{\text{lower class limit} + \text{upper class limit}}{2}$$

class limit = $l - 0.5$ to $h + 0.5$

③ Median

Ungrouped

$n$ = odd $\quad M = \left(\frac{n+1}{2}\right)^{th}$ observ.

$n$ = even $\quad M = \frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$ obsv.

Grouped

$$\boxed{M = l + \frac{h}{f}\left(\frac{N}{2} - CF\right)}$$

$l$ = lower limit of median class

$l$ = lower limit of median class
$h$ = width of " "
$N = \Sigma f$
$f$ = freq of " "
$CF$ = cumulative freq. of class preceding median class.

Median class = The class having "CF" just greater or equal to $\frac{N}{2}$.

$CF$ table = Continuous addition of freq.

④ Mode

Grouped / cont. Freq. Distribution

Modal class = class with max freq.

$$\text{Mode} = l + h\left[\frac{f_m - f_1}{2f_m - f_1 - f_2}\right]$$

$l$ = lower limit of modal class
$h$ = width of " "
$f_m$ = freq. of modal class
$f_1$ = freq. of class just "above" modal class
$f_2$ = " " " " "below" " "

## (5) Standard Deviation

$$\boxed{SD = \sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot f}{N}}}$$

$\bar{x} = \frac{\sum x}{N}$

$N \to$ no. of obs.

$N = \sum f_i$ in groped data

$$\boxed{\text{Variance} = \sigma^2}$$

## (6) Coeff. of Variance

$$\boxed{CV = \frac{SD}{\bar{x}} \times 100}$$

$\to$ measure of dispersion.

## (7) Moments

- Measure of force w.r.t its tendency to provide rotation.

$\to$ Moment about mean

$$\boxed{\mu_r = \frac{\sum f(x-\bar{x})^r}{\sum f}}$$

$$\therefore \mu_1 = \frac{\sum f(x-\bar{x})}{\sum f} ; \mu_2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

$$\mu_2 = \sigma^2$$

→ Moment about arbitrary origin

$$\boxed{\mu_r' = \frac{\sum f(x-a)^r}{\sum f}}$$

$$\therefore \mu_0' = \frac{\sum f}{\sum f} = 1$$

$$\mu_1' = \frac{\sum f(x-a)}{\sum f} = \bar{x} - a$$

→ Relation b/w moment about mean & moment about arbitrary value

(i) $\mu_1 = \mu_1' - \mu_1' = 0$

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

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

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

8 Skewness

– Measure of degree of non symmetry of freq. distribution.

- Measure of degree of non symmetry of freq. distribution.

coeff of Skewness = $\gamma$

$$\boxed{\gamma_1 = \sqrt{\beta_1}} \qquad \boxed{\beta_1 = \frac{\mu_3^2}{\mu_2^3}}$$

## 9 Kurtosis

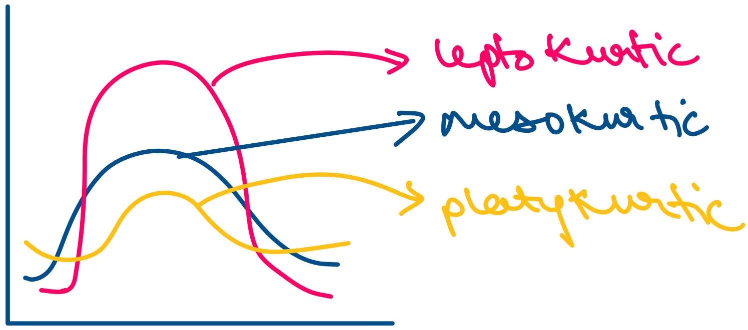
- Measure of thickness of distribution

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$$ → defines peakness of curve

$\beta_2 = 3$ Mesokurtic (Normal)

$\beta_2 > 3$ Leptokurtic (More peak)

$\beta_2 < 3$ Platykurtic (Less peak)




freq. distribution

freq. distribution

## (10) Stack Deviation Method

Suppose $\boxed{d = \frac{x-a}{h}}$ [h = height of class]

$\rightarrow x - a = d.h$

$\because \mu_r' = \frac{\Sigma f (x-a)^r}{\Sigma f}$

$\therefore \boxed{\mu_r' = \frac{\Sigma f d^r}{\Sigma f} . h^r}$

$\mu_1' = h \frac{\Sigma f d}{\Sigma f}$ ; $\mu_2' = h^2 \frac{\Sigma f d^2}{\Sigma f}$

$\mu_3' = h^3 \frac{\Sigma f d^3}{\Sigma f}$

Suppose

$u_1 = \frac{\Sigma f d}{\Sigma f}$ ; $u_2 = \frac{\Sigma f d^2}{\Sigma f}$ ; $u_3 = \frac{\Sigma f d^3}{\Sigma f}$

$\therefore \boxed{\mu_1' = h u_1 \ ; \ \mu_2' = h^2 u_2 ; \ \mu_3' = h^3 u_3}$

$$\boxed{\begin{array}{l} (i)\ \mu_1 = 0 \\ (ii)\ \mu_2 = h^2(u_2 - u_1^2) \\ (iii)\ \mu_3 = h^3(u_3 - 3u_2u_1 + 2u_1^3) \\ (iv)\ \mu_4 = h^4(u_4 - 4u_3u_1 + 6u_3u_2 - 3u_1^4) \end{array}}$$

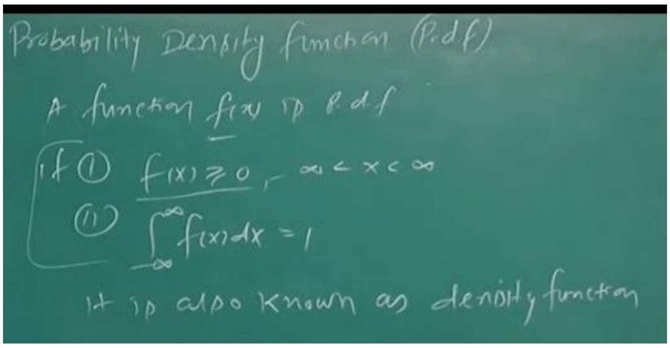
# 11 Mathematical expectations

Random variable = A real value assigned to each sample point of the sample space.
Discrete RV = Random variable takes finite or countable values. Eg coin tossing
Continuous RV = Random variable taking all possible values. Eg weight of all college students

Probability function of x = f(x).
P(X=x) = f(x) is the probability function if it satisfies the following conditions

1. f(x) >= 0 for all values of x
2. Sum of all values of f(x) = 1

If X is discrete RV then its probability function is called = Probability mass function
If X is continuous RV then its probability function is called = Probability density function




Mathematical Expectation:
If X is random variable and f(x) is probability function then mathematical expectation is denoted by E(x).

E(x) = summation xi fi from i=1 to n

E(x) = summation xi.P(xi) for DRV
E(x) =

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

for CRV

E(x^2) = x^2 times P(x) for DRV

if $p(x) = x$

$\bar{x}$ = mean $= E(X)$

- D.R.V: $E(X) = \sum_x x\,p(x)$
- C.R.V: $E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx$

$E(X) = \bar{x}$

Variance $= E(x-\bar{x})^2 = E$

$= E(x^2) - [E(x)]^2$

Correlation coefficient = karls

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

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

$$cov(x,y) = E(xy) - E(x)E(y)$$
$$= \frac{\Sigma xy}{n} - \frac{\Sigma x}{n}\frac{\Sigma y}{n}$$
$$\sigma_x = \sqrt{E(x^2) - (E(x))^2} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$
$$\sigma_y = \sqrt{E(y^2) - (E(y))^2} = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}$$

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

Rank correlation = spearsman

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

$$d^2 = (R_x - R_y)^2$$

$$\text{rank } r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

Regression = statistical technique to estimate or predict a variable from another known variable.
statistical technique that defines relationship between two variables

Regression lines = when a graph is plotted with two variables which are linearly correlated to each other. So the lines have 2 regression equations as well which are

1. Y on x i.e., y depends on x => y = a + bx
2. X on y i.e., x depends on y => x = a + by

→ Regression eq$^n$

1. y on x : $\boxed{y - \bar{y} = b_{yx}(x - \bar{x})}$

$\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}}$ → Regression coeff.

$$\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}} \rightarrow \text{Regression coeff.}$$

(2) $x$ on $y$: $\boxed{x-\bar{x} = b_{xy}(y-\bar{y})}$

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

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

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

→ Properties of Regression Coeff.

(1) Geometric mean $r$ b/w 2 reg. coeff. is given by :

$$r^2 = b_{xy}\cdot b_{yx}$$

$$\boxed{r = \pm\sqrt{b_{xy}\cdot b_{yx}}}$$

(2) Both RC will have same sign.

3 Individual values of both RC can't exceed 1.

* 2 regression lines always intersect @ mid pts. of $x$ & $y$ i.e., $(\bar{x}, \bar{y})$