ST. LAWRENCE HIGH SCHOOL

A JESUIT CHRISTIAN MINORITY INSTITUTION

STUDY MATERIAL-6

SUBJECT – STATISTICS

Pre-test

Chapter: BIVARIATE ANALYSIS **Class: XII**

Topic:CORRELATION FOR GROUPED DATA **Date: 08.06.2020**

CORRELATION

PART 4

Suppose we are studying two variables x & y simultaneously. A statement of possible pairs of values assumed by x & y, together with the corresponding probabilities, is called bivatriate probability distribution or joint probability distribution.

Y / X	y_1	y_2	y_3	y_n	marginal prob Of x_i
X_1	p_{11}	p_{12}	p_{13}	p_{1n}	p_{10}
X_2	p_{21}	p_{22}	p_{23}	p_{2n}	p_{20}
X_3	p_{31}	p_{32}	p_{33}	p_{3n}	p_{30}
X_m	p_{m1}	p_{m2}	p_{m3}	p_{mn}	p_{m0}
Marginal prob of y_j	P_{01}	p_{02}	p_{03}	p_{0n}	1

Where p_{ij} = joint probability of x_i and y_j

= $P\left(X = x_i \ \cap Y = \ y_j \right)$

The marginal probabilities are given by

Marginal probability of $X = x_{i\,,}$ p_{io} = $\sum_j p_{ij}$

Marginal probability of $Y = y_{i\,,}$ p_{0j} = $\sum_i p_{ij}$

Some important formulae :

$$E\left(\varphi(X,Y)\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \varphi(x,y)p_{ij}$$

So from the above table,

$$E(X) = \sum_{i=1}^{m} x_i p_{i0} \ and \ E(Y) = \sum_{j=1}^{n} y_j p_{oj}$$

$$E(X^2) = \sum_{i=1}^{m} x_i^2 p_{i0} \ and \ E(Y^2) = \sum_{i=1}^{n} y_j^2 p_{0j}$$

$$E(XY) = \sum_{i=1}^{m}\sum_{i=1}^{n} x_i y_j p_{ij}$$

$$\sigma_X^2 = E(X^2) - \ (E(X))^2$$

$$\sigma_Y^2 = E(Y^2) - \ (E(Y))^2$$

$$cov(X,Y) = \ E(XY) - \ E(X)E(Y)$$

Correlation coefficient $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

Regression equation y on x is $Y - E(Y) = \beta_{YX} (X - E(X))$

Regression coefficient y on x, $\beta_{YX} = \frac{cov(X,Y)}{\sigma_X^2} = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$

Some impotant theorems and results :

- $E(X+Y) = E(X) + E(Y)$

Pf: $E(X+Y) = \sum_{i=1}^{m} \sum_{i=1}^{n} (x_i + y_{j)}\, p_{ij}$

$= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i p_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} y_j p_{ij}$

$= \sum_{i=1}^{m} (x_i \sum_{j=1}^{n} p_{ij}) + \sum_{j=1}^{n} (y_j \sum_{i=1}^{m} p_{ij})$

$= \sum_{i=1}^{m} x_i p_{i0} + \sum_{j=1}^{n} y_j p_{0j}$

$= E(X) + E(Y)$

- If X and Y are two independent variables, then $E(XY) = E(X).E(Y)$

Pf: If X and Y are independent, then

$P\big(X = x_i \cap Y = y_j \big) = P(X = x_i).P(Y = y_j)$

$\Rightarrow p_{ij} = p_{i0}.p_{0j}$

So $E(XY) = \sum_{i=1}^{m} \sum_{i}^{n} x_i y_j p_{ij}$

$= \sum_{i=1}^{m} \sum_{i}^{n} x_i y_j p_{i0} p_{0j}$

$= \sum_{i=1}^{m} (x_i p_{i0} \sum_{j=1}^{n} y_j p_{oj})$

$= \sum_{i=1}^{m} x_i p_{i0} E(Y)$

$= E(Y).E(X)$

- $V(X_1 + X_2 + \ \dots\dots\dots + X_n)$
  $= E((X_1 + X_2 + \ \dots\dots\dots + X_n) - E(X_1 + X_2 + \ \dots\dots\dots + X_n))^2$
  $= E(E(X_1 - E(X_1))^2 + \ E(X_2 - E(X_2)^2 + \ \dots\dots..+E(X_n - E(X_n)^2)$
  $$= \sum_{i=1}^{n} V(X_i) + \ 2 \sum_{i=1}^{n}\sum_{j=1}^{n} cov(X_{i,}X_j)$$
  $i<j$
  $$= \sum_{i=1}^{n} V(X_i) + \ \sum_{i=1}^{n}\sum_{j=1}^{n} cov\left(X_{i,}X_j\right)$$
  $i \ \neq j$

  It comes from the square formula. As it was told already that we get the formula for variance from

  $$\left(\sum_{i=1}^{n} x_i\right)^2 \ = \ \sum_{i=1}^{n} x_i^2 + 2 \ \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j$$
  $i < j$
  $$= \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j$$
  $i \neq j$

Prepared by
Sanjay Bhattacharya