

Joint Variables (II)

Goal: Study expected value and variation when dealing with more than one r.v.

Suppose X, Y are two random variables

either both discrete or both continuous

Let $f(x,y) \rightarrow$ joint pmf of X, Y if both discrete
 (freq) \rightarrow joint pdf of X, Y if both continuous.

If $h(x,y)$ is a function of X and Y

can define the expected value of $h(x,y)$

$$E(h(x,y)) = \begin{cases} \sum_{(x,y)} h(x,y) p(x,y) & \text{if } X, Y \text{ are discrete} \\ \iint_{-\infty}^{\infty} h(x,y) f(x,y) dx dy & \text{if } X, Y \text{ are both continuous.} \end{cases}$$

Covariance and Correlation

Want: A way to measure relationship between X and Y .

Let

$\mu_X = E(X) \rightarrow$ expected value of X

$\mu_Y = E(Y) \rightarrow$ expected value of Y .

Let

$$\begin{aligned} h(x,y) &= (x - \mu_X)(y - \mu_Y) \\ &= \text{"product of the deviation of } X \text{ and } Y \text{ from their respective expected values"} \end{aligned}$$

Note:

- ① If large values of X are associated with large values of $Y \rightarrow$ i.e. $(x - \mu_X) > 0 \Rightarrow (y - \mu_Y) > 0$
 and
 small values of X are associated with small values of $Y \rightarrow$ i.e. $(x - \mu_X) < 0 \Rightarrow (y - \mu_Y) < 0$

We have

$$h(x,y) = (x - \mu_X)(y - \mu_Y) \geq 0$$

as terms on right are both positive or both negative

- ② If large X values are associated with small Y values
 \rightarrow i.e. $(x - \mu_X) > 0 \Rightarrow (y - \mu_Y) < 0$
 and
 large Y values are associated with small X values
 \rightarrow i.e. $(y - \mu_Y) > 0 \Rightarrow (x - \mu_X) < 0$

We will have:

$$h(x,y) = (x - \mu_X)(y - \mu_Y) \leq 0$$

Since terms on the right are of opposite sign.

- \therefore The expected value of $h(x,y) = (x - \mu_X)(y - \mu_Y)$ \rightarrow a good candidate to measure relationship between X and Y .

Define the covariance of X and Y :

$$\begin{aligned} \text{Cov}(X,Y) &= E((x - \mu_X)(y - \mu_Y)) = E(XY) - E(X)E(Y). \\ &= \text{"the expected value of the product of deviation of } X \text{ from } \mu_X \text{ and } Y \text{ from } \mu_Y"} \end{aligned}$$

$$\begin{aligned} \text{Cov}(X,Y) &= \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x,y) && \text{if } X, Y \text{ are discrete} \\ &= \iint_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x,y) dx dy && \text{if } X, Y \text{ are continuous} \end{aligned}$$

Note:

$$\text{① If } X=Y, \quad \text{Cov}(X,Y) = \text{Cov}(X,X) = V(X)$$

- ② If X and Y have "strong positive relationship" \rightarrow large values of X associated with large values of Y and small values of X are associated with small values of Y .

$\text{Cov}(X,Y)$ will be positive.

- ③ If X, Y have a "strong negative relationship" \rightarrow large values of X associated with small values of Y and large values of Y associated with small values of X
 \rightarrow $\text{Cov}(X,Y)$ will be negative.

- ④ If $\text{Cov}(X,Y)$ is positive \rightarrow can say X and Y are positively associated
- $\text{Cov}(X,Y)$ is negative \rightarrow X and Y are negatively associated.

Not a good measure of association!

Depends on the units in which X and Y are measured
 \rightarrow can be made arbitrarily large/small by changing units.

The Correlation Coefficient of X and Y

Let $\sigma_X = \sqrt{V(X)}$ and $\sigma_Y = \sqrt{V(Y)}$,

To make $\text{Cov}(X,Y)$ \rightarrow divide by $\sigma_X \cdot \sigma_Y$ (units).

\therefore The correlation of X and Y is

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Note:

$$\text{① } S_{X,Y} \text{ does not depend on the unit of } X \text{ and } Y$$

$$\text{② } S_{X,Y} \text{ satisfies } -1 \leq S_{X,Y} \leq 1.$$

$$\text{③ If } a, b \text{ are both positive or both negative,} \quad \text{Cov}(aX+b, cY+d) = \text{Cov}(X,Y)$$

In particular, if $y = ax + b$

$$\begin{aligned} \text{Corr}(X,Y) &= \text{Corr}(X, ax+b) = \text{Sign}(a) \text{Corr}(X,X) \\ &= \text{Sign}(a) = \pm 1 \end{aligned}$$

$\text{Corr}(X,Y)$ measures the degree of linear relationship between X and Y .

\therefore If $S_{X,Y}$ close to 1 \rightarrow strong positive linear association between X and Y .
 Close to -1 \rightarrow strong negative linear association.

$S_{X,Y}=0 \rightarrow X, Y$ are uncorrelated \rightarrow no linear relationship.
 \therefore does not mean that there is no relationship.

④ If X, Y are independent $\Rightarrow \text{Cov}(X,Y)=0$
 $\Rightarrow \text{Corr}(X,Y)=0$.

$\text{Cov}(X,Y)=0$ does not necessarily imply that X and Y are independent!

⑤ Good old adage:

Covariance does not imply Correlation!

only implies association between variables

possible that there is an underlying variable that is "causing" the positive association.

⑥ Note:

$$S_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \Rightarrow \text{Cov}(X,Y) = S_{X,Y} \cdot \sigma_X \cdot \sigma_Y.$$

Linear Combination of Random Variables

Suppose $X_1, X_2, X_3, \dots, X_m$ are n -random variables such that:

$$\begin{aligned} \text{(i)} \quad E(X_i) &= 1, \quad i=1, 2, 3, \dots, n \\ \text{(ii)} \quad V(X_i) &= \sigma_i^2 \quad i=1, 2, \dots, n. \end{aligned}$$

Let $a_1, a_2, \dots, a_n \in \mathbb{R}$, can define the "linear combination" as

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n$$

What is $E(Y)$? $V(Y)$?

Then:

$$\begin{aligned} \text{(i)} \quad E(Y) &= E(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) \\ &= a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad V(Y) &= V(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j) \end{aligned}$$

In particular, if $X_1, X_2, X_3, \dots, X_m$ are independent

$$\begin{aligned} V(Y) &= a_1^2 V(X_1) + a_2^2 V(X_2) + \dots + a_n^2 V(X_n) \\ &= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2 \end{aligned}$$

For Two Random Variables

$Z = aX + bY$, then

$$\text{(i)} \quad E(Z) = a\mu_X + b\mu_Y$$

$$\text{(ii)} \quad V(Z) = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \text{Cov}(X, Y)$$

If $a=1, b=-1$, $Z = X-Y$

$$\begin{aligned} E(X-Y) &= \mu_X - \mu_Y, \\ V(X-Y) &= \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y). \end{aligned}$$

Given $p(x,y) \rightarrow$ joint pmf of X and Y .

$$h: X \times Y \rightarrow \mathbb{R}$$

$$(x,y) \mapsto h(x,y)$$

$$Z = h(X,Y)$$

$$E(Z) = ?$$

$$E(h(x,y)) = \sum_{(x,y) \in X \times Y} h(x,y) p(x,y)$$

$$= \sum_{x \in X} \sum_{y \in Y} h(x,y) p(x,y)$$

if

$$\textcircled{1} \quad h(x,y) = X$$

$$E(h(x,y)) = E(X) = \mu_X$$

$$\textcircled{2} \quad h(x,y) = Y \rightarrow E(h(x,y)) = E(Y) = \mu_Y$$

$$\text{Let } \mu_X = E(X), \mu_Y = E(Y)$$

$$\text{Define: } h(x,y) = (X - \mu_X)(Y - \mu_Y)$$

= product of the deviations of X and Y from their respective means.

large values of X are associated with large values of Y
 $(X - \mu_X) > 0 \Rightarrow (Y - \mu_Y) > 0$

small values X are associated with small values of Y
 $(X - \mu_X) < 0 \Rightarrow (Y - \mu_Y) < 0$

$$h(x,y) = (X - \mu_X)(Y - \mu_Y) > 0$$

$$\underline{E(h(x,y)) > 0}$$

$$\text{Cov}(X,Y) = E((X - \mu_X)(Y - \mu_Y))$$

= Expected value of the product of the deviations.

{

If $\text{Cov}(X,Y)$ is positive \rightarrow positive association
~~between X and Y~~

$\text{Cov}(X,Y)$ is negative \rightarrow negative association
 between X and Y .

Note: $\text{Cov}(X,Y)$ depends on the units in which X and Y are measured.

units of $\text{Cov}(X,Y)$ = product of the units of X and Y .

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \rho_{X,Y}$$

$\rho_{X,Y}$ \rightarrow measures the extent of "linear relationship" between X and Y .

$$-1 \leq \rho_{X,Y} \leq 1 \rightarrow \rho_{X,Y} = \pm 1 \Leftrightarrow Y = aX + b$$

{ perfect linear relationship between X and Y .

Scatter Plots $\rightarrow (x,y)$ in \mathbb{R}^2

