Lecture No.15
Covariance and Correlation

Azeem Iqbal

University of Engineering and Technology, Lahore
(Faisalabad Campus)

Definition **Covariance**

The **covariance** of two random variables X and Y is

$$\mathrm{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

- ✓ Sometimes, the notation $\boldsymbol{\sigma}_{X,Y}$ is used to denote the covariance of X and Y.
- ✓ It is a quantitative measure of the relationship between the two random variables.

- ✓ the magnitude reflects the strength of the relationship, the sign conveys the direction of the relationship.
- ✓ When **$\mathbf{cov(X, Y) = 0}$**, the two random variables are said to be "**uncorrelated**".
- ✓ A **positive correlation** implies, roughly speaking, that they tend to increase or decrease together.
- ✓ A **negative correlation**, on the other hand, implies that when X increases, Y tends to decrease, and vice versa.

Definition **Correlation**

The **Correlation** of two random variables X and Y is

$$r_{X,Y} = E[XY]$$

(a) $\boxed{Cov[X,Y] = r_{X,Y} - \mu_X \mu_Y}$

(b) $Var[X+Y] = Var[X] + Var[Y] + 2\,Cov[X,Y]$.

(c) If $X = Y$, $Cov[X,Y] = Var[X] = Var[Y]$
and $r_{X,Y} = E[X^2] = E[Y^2]$

Example

The probability model for X and Y is given by the following matrix.

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$	$P_X(x)$
$x=0$	0.01	0	0	0.01
$x=1$	0.09	0.09	0	0.18
$x=2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

$Cov = r_{X,Y} - \mu_X \mu_Y$

Find $\mathbf{r}_{X,Y}$ and $Cov[X,Y]$

$E[XY]$ $\quad E[(X-\mu_X)(Y-\mu_Y)]$

Solution

$$r_{X,Y} = E[XY] = \sum_{x=0}^{2} \sum_{y=0}^{2} xy P_{X,Y}(x, y)$$
$$= (1)(1)0.09 + (2)(2)0.81 = 3.33.$$

$$E[X] = (1)(0.18) + (2)(0.81) = 1.80,$$
$$E[Y] = (1)(0.09) + (2)(0.81) = 1.71.$$

$$\mathrm{Cov}[X, Y] = 3.33 - (1.80)(1.71) = 0.252.$$

Download Classroom App from PlayStore:

Code to join the class: **5wphr6r**

Quiz 1 – 24 –December-2020

Syllabus – From start to reliability topic.

Definition

Correlation Coefficient

The correlation coefficient of two random variables X and Y is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

- ✓ Note that the units of the covariance and the correlation are the product of the units of X and Y.

- ✓ Thus, if X has units of kilograms and Y has units of seconds, then $\boldsymbol{Cov}[\boldsymbol{X},\boldsymbol{Y}]$ and $\boldsymbol{r}_{\boldsymbol{X},\boldsymbol{Y}}$ have units of kilogram-seconds. .

- ✓ By contrast, $\rho_{X,Y}$ is a **dimensionless quantity**.
- ✓ An important property of the correlation coefficient is that it is bounded by -1 and 1.

$$-1 \leq \rho_{X,Y} \leq 1$$

- ✓ $\rho_{X,Y}$ describes the information we gain about Y by observing X.
- ✓ For example, a positive correlation coefficient, $\rho_{X,Y} > 0$, suggests that when X is high relative to its expected value, Y also tends to be high, and when X is low, Y is likely to be low.

- ✓ A negative correlation coefficient, $\boldsymbol{\rho_{X,Y} < 0}$, suggests that a high value of X is likely to be accompanied by a low value of Y and that a low value of X is likely to be accompanied by a high value of Y.

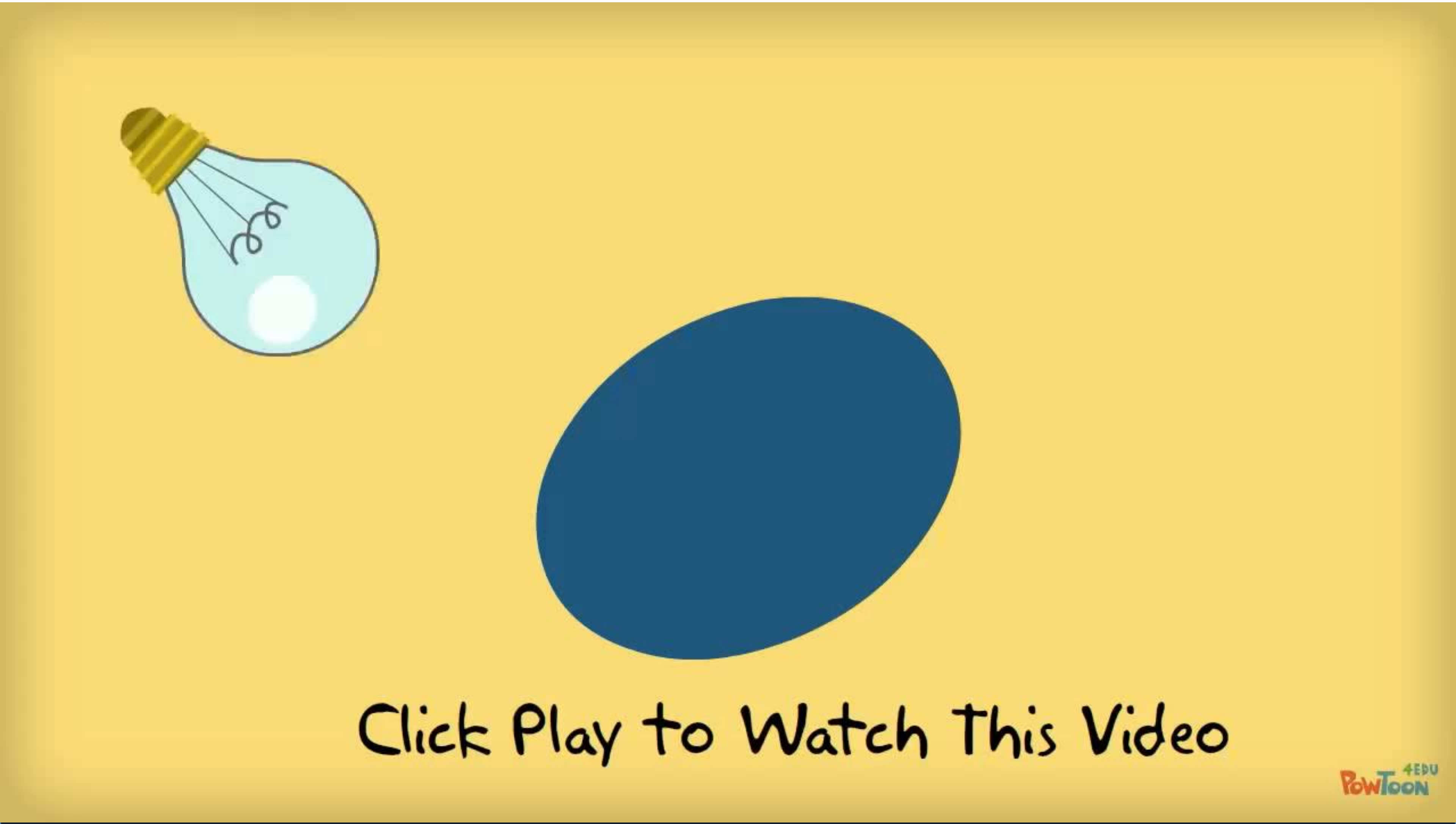
- ✓ A linear relationship between X and Y produces the extreme values, $\boldsymbol{\rho_{X,Y} = \pm 1}$.

Some examples of positive, negative, and zero correlation coefficients include:

- X is the height of a student. Y is the weight of the same student. $\mathbf{0} < \boldsymbol{\rho}_{X,Y} < \mathbf{1}$.
- X is the distance of a cellular phone from the nearest base station. Y is the power of the received signal at the cellular phone. $\mathbf{-1} < \boldsymbol{\rho}_{X,Y} < \mathbf{0}$.
- X is the temperature of a resistor measured in degrees Celsius. Y is the temperature of the same resistor measured in degrees Kelvin. $\boldsymbol{\rho}_{X,Y} = 1$.
- X is the gain of an electrical circuit measured in decibels. Y is the attenuation, measured in decibels, of the same circuit. $\boldsymbol{\rho}_{X,Y} = -1$.

- X is the telephone number of a cellular phone. Y is the CNIC number of the phone's owner. $\boldsymbol{\rho_{X,Y} = 0}$.

Lighting up Statistics - Correlation/Covariance
https://www.youtube.com/watch?v=eRlzmCrdTWw



Example

The joint probability density function of random variables X and Y is

$$f_{X,Y}(x, y) = \begin{cases} xy & 0 \leq x \leq 1, 0 \leq y \leq 2, \\ 0 & \textit{otherwise}. \end{cases}$$

Find the following quantities.

(1) $E[X]$ *and* $\mathrm{Var}[X]$

(2) $E[Y]$ *and* $\mathrm{Var}[Y]$

(3) The correlation $r_{X,Y} = E[XY]$

(4) The covariance $\mathrm{Cov}[X, Y]$

(5) The correlation coefficient $\rho_{X,Y}$

Solution

The first and second moments of X are

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 2x^2\, dx = \frac{2}{3}$$

$$E\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx = \int_0^1 2x^3\, dx = \frac{1}{2}$$

The variance of X is $\mathrm{Var}[X] = E[X^2] - (E[X])^2 = 1/18$.

Solution

The first and second moments of Y are

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^2 \frac{1}{2} y^2\, dy = \frac{4}{3}$$

$$E\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy = \int_0^2 \frac{1}{2} y^3\, dy = 2$$

The variance of Y is $\text{Var}[Y] = E[Y^2] - (E[Y])^2 = 2 - 16/9 = 2/9$.

Solution

The correlation of X and Y is

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\, dx, dy$$

$$= \int_{0}^{1} \int_{0}^{2} x^2 y^2\, dx, dy = \frac{x^3}{3}\bigg|_0^1 \frac{y^3}{3}\bigg|_0^2 = \frac{8}{9}$$

Solution

The covariance of X and Y is

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = \frac{8}{9} - \left(\frac{2}{3}\right)\left(\frac{4}{3}\right) = 0.$$

Since $\text{Cov}[X, Y] = 0$, the correlation coefficient is $\rho_{X,Y} = 0$.