LECTURE 34 : CORRELATION AND COVARIANCE

- TOPICS TO COVER (BASED ON CH 4.8)
  - → CORRELATION AND COVARIANCE

→ CORRELATION AND COVARIANCE

LET X AND Y BE TWO RANDOM VARIABLES WITH FINITE SECOND MOMENTS, I.E.

$EX^2 < \infty$ AND $EY^2 < \infty$.

YOU'LL SEE IT SOON!

DEFINE THREE MEASURES OF A LINEAR? RELATIONSHIP BETWEEN X AND Y:

JOINT EXPECTATION OF X AND Y

CORRELATION BETWEEN X AND Y $= E(XY)$

COVARIANCE BETWEEN X AND Y : $\mathrm{Cov}(X,Y) = E(X-E(X))(Y-E(Y))$

CORRELATION COEFFICIENT BETWEEN X AND Y $= \rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$

OBSERVE THE FOLLOWING :

→ $$E(XY) = \begin{cases} \sum_i \sum_j x_i y_j \, p_{X,Y}(x_i, y_j), & \text{DISCRETE CASE,} \\ \int\int_{\mathbb{R}^2} uv \, f_{X,Y}(u,v)\, dv\, du, & \text{CONT. CASE.} \end{cases}$$

→ $$\mathrm{Cov}(X,X) = E(X-E(X))(X-E(X)) = E(X-E(X))^2 = \mathrm{Var}(X)$$

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

LINEARITY OF EXPECTATION $= E(XY) - E(X\,E(Y)) - E(E(X)\,Y) + E(E(X)\,E(Y))$

E(X) AND E(Y) ARE CONSTANTS $= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$

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

$\rightarrow$ IF $E(X) = 0$ OR $E(Y) = 0$: $\text{Cov}(X,Y) = E(XY)$

DEFINITION : IF $\text{Cov}(X,Y) = 0$: X AND Y ARE CALLED UNCORRELATED.

IF $\text{Cov}(X,Y) > 0$: X AND Y ARE CALLED +VELY UNCORRELATED.

IF $\text{Cov}(X,Y) < 0$: X AND Y ARE CALLED -VELY UNCORRELATED.

IF $\sigma_X^2 = \text{Var}(X) > 0$ AND $\sigma_Y^2 = \text{Var}(Y) > 0$, I.E., $\rho_{X,Y}$ IS WELL-DEFINED,

$\text{Cov}(X,Y) \begin{matrix} = \\ > \\ < \end{matrix} 0 \iff \rho_{X,Y} \begin{matrix} = \\ > \\ < \end{matrix} 0$

INDEPENDENCE AND UNCORRELATEDNESS

THEOREM: X AND Y ARE INDEPENDENT $\Rightarrow$ X AND Y ARE UNCORRELATED.

HOWEVER THE CONVERSE IS NOT TRUE.

PROOF: LET X AND Y ARE INDEPENDENT, I.E.

$$f_{X,Y}(u,v) = f_X(u)\cdot f_Y(v) \quad \forall\ (u,v)\in\mathbb{R}^2 \qquad (1)$$

$$\Rightarrow \iint_{\mathbb{R}^2} uv\, f_{X,Y}(u,v)\, dv\, du = \iint_{\mathbb{R}^2} uv\, f_X(u)\cdot f_Y(v)\, dv\, du \qquad (2)$$

$$\Rightarrow \quad E(XY) = \int_{\mathbb{R}} u f_X(u)\, du \int_{\mathbb{R}} v f_Y(v)\, dv$$

$$\Rightarrow \quad E(XY) = E(X)\,E(Y) \quad \Rightarrow \quad X \text{ AND } Y \text{ ARE UNCORRELATED}$$

THE CONVERSE DOESN'T HOLD TRUE AS $(2) \not\Rightarrow (1)$, I.E., IF TWO INTEGRALS ARE THE SAME $\not\Rightarrow$ INTEGRANDS / FUNCTIONS INTEGRATED ARE THE SAME. E.G.

$$\int_0^1 x\, dx = \int_0^1 (1-x)\, dx = \frac{1}{2} \quad \text{BUT} \quad x = (1-x) \quad x\in[0,1]$$

ANOTHER WAY TO REALIZE THAT THE CONVERSE DOESN'T HOLD IS AS FOLLOWS:

TAKE $X \sim$ UNIFORM $(-1, 1)$ AND DEFINE $Y = X^2$. CLEARLY X AND Y ARE NOT INDEPENDENT.

CONSIDER $$E(X) = \int_{-1}^{1} u\cdot\frac{1}{2}\, du = \frac{u^2}{4}\Big|_{-1}^{1} = \frac{1^2-(-1)^2}{4} = 0$$

$$E(Y) = E(X^2) = \int_{-1}^{1} u^2\cdot\frac{1}{2}\, du = \frac{u^3}{6}\Big|_{-1}^{1} = \frac{1^3-(-1)^3}{6} = \frac{2}{6}$$

$$E(XY) = E(X\,X^2) = E(X^3)$$

$$= \int_{-1}^{1} u^3 \cdot \frac{1}{2}\, du = \frac{u^4}{8}\Big|_{-1}^{1} = \frac{1^4 - (-1)^4}{6} = 0$$

$$\Rightarrow \quad Cov(X,Y) = E(XY) - E(X)\cdot E(Y)$$

$$= 0 - 0 \cdot \frac{1}{3}$$

$$= 0$$

$\Rightarrow$ X AND Y ARE UNCORRELATED.

BUT $\underbrace{Y = X^2}$ ARE CLEARLY DEPENDENT.

NON-LINEAR RELATIONSHIP!

CORRELATION CAN ONLY DETECT LINEAR RELATIONSHIPS! YOU SEE IT NOW!

EXERCISE. PROVE THE ABOVE THEOREM FOR THE DISCRETE CASE.

UNCORRELATEDNESS OF A SET (POSSIBLY OF SIZE > 2) OF RANDOM VARIABLES.

DEFINITION: A SET (POSSIBLY OF SIZE > 2) OF RANDOM VARIABLES IS CALLED UNCORRELATED IF THEY ARE PAIRWISE UNCORRELATED, I.E. LET

$S = \{X_1, \dots, X_n\}$ BE A SET OF n RVS, S IS CALLED UNCORRELATED IF $\mathrm{Cov}(X_i, X_j) = 0 \quad \forall\ i \neq j = 1, \dots, n$.

REMARK: RECALL THAT UNLIKE UNCORRELATEDNESS, PAIRWISE INDEPENDENCE IS NOT SUFFICIENT FOR MUTUAL INDEPENDENCE.

OBSERVE THE FOLLOWING:

→ $\mathrm{Cov}(X+Y, U+V) = \mathrm{Cov}(X,U) + \mathrm{Cov}(X,V) + \mathrm{Cov}(Y,U) + \mathrm{Cov}(Y,V)$

PROOF: CONSIDER:

$$\begin{aligned}
\mathrm{Cov}(X+Y, U+V) &= E\left((X+Y)(U+V)\right) - E(X+Y)E(U+V) \\
&= E(XU) + E(XV) + E(YU) + E(YV) \\
&\quad - \left(E(X)E(U) + E(X)E(V) + E(Y)E(U) + E(Y)E(V)\right) \\
&= E(XU) - E(X)E(U) + E(XV) - E(X)E(V) \\
&\quad + E(YU) - E(Y)E(U) + E(YV) - E(Y)E(V)) \\
&= \mathrm{Cov}(X,U) + \mathrm{Cov}(X,V) + \mathrm{Cov}(Y,U) + \mathrm{Cov}(Y,V)
\end{aligned}$$

→ $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X,Y)$ $\quad a, b, c, d$ ARE CONSTANTS

PROOF: EXERCISE. HINT: VERY SIMILAR TO THE PROOF OF THE VARIANCE OF A RANDOM VARIABLE IS AFFECTED BY THE CHANGE OF SCALE BUT NOT OF THE ORIGIN.

→ $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

PROOF: CONSIDER:

$$Var(X+Y) = Cov(X+Y,\, X+Y)$$
$$= Cov(X,X) + Cov(X,Y) + Cov(Y,X) + Cov(Y,Y)$$
$$= Var(X) + Cov(X,Y) + Cov(X,Y) + Var(Y)$$
$$= Var(X) + Var(Y) + 2\,Cov(X,Y)$$

→ IF X AND Y ARE UNCORRELATED THEN: $Var(X+Y) = Var(X) + Var(Y)$

PROOF: EXERCISE. HINT: FOLLOWS DIRECTLY FROM THE RESULT ABOVE.

→ $\rho_{aX+b,\, cY+d} = \rho_{X,Y}$ FOR $a, c > 0$

PROOF: CONSIDER:

$$\rho_{\underbrace{aX+b}_{\text{FUNCTION OF } X},\, cY+d} = \frac{Cov(aX+b,\, cY+d)}{\sqrt{Var(aX+b)\,Var(cY+d)}} = \frac{ac\,Cov(X,Y)}{\sqrt{a^2\,Var(X)\;c^2\,Var(Y)}}$$

SCALE

ORIGIN

REPRESENTING THE CHANGE OF ORIGIN AND SCALE

$$= \frac{ac\,Cov(X,Y)}{ac\sqrt{Var(X)\,Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

$$= \rho_{X,Y}$$

STANDARDIZED VERSION OF X

STANDARDIZED VERSION OF Y

→ $$Cov\left(\underbrace{\frac{X - E(X)}{\sigma_X}}_{\text{MEAN}=0,\ \text{VAR}=1},\ \frac{Y - E(Y)}{\sigma_Y}\right) = \rho_{X,Y}$$

CAN BE REGARDED AS THE STANDARDIZED VERSION OF $Cov(X,Y)$

PROOF: CONSIDER:

$$\text{cov}\left(\frac{X - E(X)}{\sigma_X}, \frac{Y - E(Y)}{\sigma_Y}\right) \overset{\text{CHANGE OF ORIGIN}}{=} \text{cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

$$\overset{\text{CHANGE OF SCALE}}{=} \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \rho_{X,Y}$$

RANGE OF THE CORRELATION COEFFICIENT

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

REFER TO THE TEXTBOOK!

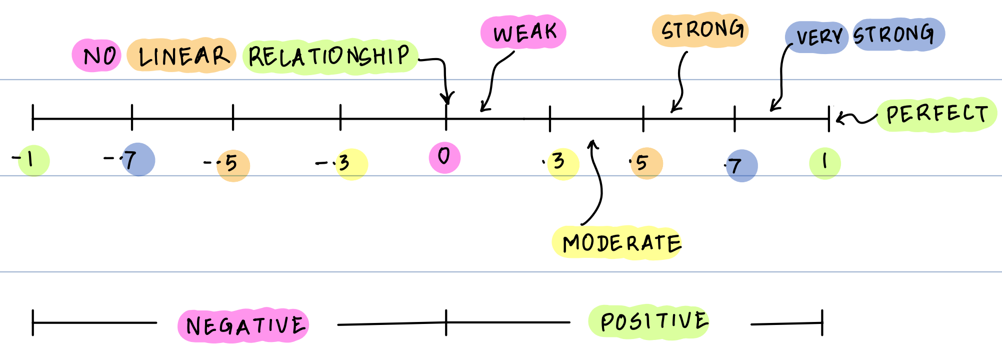
PROOF: THE PROOF FOLLOWS FROM THE CAUCHY-SCHWARZ INEQUALITY.

RELATED TO THE TRIANGLE INEQUALITY.

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

PERFECT NEGATIVE LINEAR RELATIONSHIP
$Y = aX + b$ FOR SOME $a < 0$

PERFECT POSITIVE LINEAR RELATIONSHIP
$Y = aX + b$ FOR SOME $a > 0$



EXERCISE: CALCULATE THE CORRELATION COEFFICIENT ρ_{XY} BETWEEN THE HW SCORE (X) AND EXAM SCORE (Y) FOR THE FOLLOWING DATA:

- HW SCORE (X): 9 10 7 5 8 5
- EXAM SCORE (Y): 8 9 6 4 7 3

CLASSIFY ρ_{XY} INTO ONE OF THE ABOVE RANGES.