**Introduction to Probability & Statistics**
**Prof Abhay Gopal Bhatt**
**Department of Statistics**
**Indian Statistical Institute Delhi**
**Week - 8**
**Lecture - 29**
**Correlation Coefficient**

agar humein kuch compare karna ho to covariance ka raw number kaafi useful nahi hota, isliye hum ek aur quantity use karte hain jo covariance se judi hai: correlation coefficient (sah-samvandh gunank), jisko hum corr(X,Y) ya symbol ρ(X,Y) se darshate hain, aur iska definition hai ρ(X,Y) = Cov(X,Y) / (σX · σY), jahan σX aur σY standard deviation (maanak vichhilan) of X aur Y hote hain;

$$Cov(ax+b, cy+d) = E\big((ax+b)(cy+d)\big) - E(ax+b)E(cy+d)$$

$$= acCov(x,y)$$

is formula ka effect yeh hota hai ki covariance units par depend karta hai lekin correlation units se free hota hai, kyunki agar hum X ko AX+B se aur Y ko CY+D se scale karein, to Cov(AX+B, CY+D) = AC·Cov(X,Y) hota hai, lekin σAX+B = |A|σX aur σCY+D = |C|σY, isliye numerator me AC aur denominator me |A||C| cancel ho jaate hain, bas unka sign bachta hai—matlab correlation unaffected hota hai; lekin covariance directly scale pe depend karta hai; ab dubara discrete wale example par aate hain jahan X leta hai 100 ya 250, aur Y leta hai 0, 100, 200, aur humne pehle hi E[X], E[Y], E[XY], Cov(X,Y)=1875 nikaal liya tha; ab correlation ke liye hume chahiye variances aur standard deviations; to expected value of $X^2$ = Σ $x^2$ PX(x) = 36,250; expected value of $Y^2$ = Σ $y^2$ PY(y) = 22,500;

variance of X = $E[X^2] - (E[X])^2 = 36{,}250 - 175^2 = 5625 \rightarrow \sigma X = 75$; variance of Y = $E[Y^2] - (E[Y])^2 = 22{,}500 - 125^2 = 6875 \rightarrow \sigma Y = \sqrt{6875}$; to correlation ρ(X,Y) = 1875 / (75 · √6875) jo humein batata hai ki X aur Y ke beech real relationship kitna majboot hai bina units ke effect ke yahi reason hai ki correlation ek behtar aur normalized measure hai relationship ka, jabki covariance ka raw number compare nahi kiya ja sakta across different situations. available hai correlation between x and y is covariance between x and y divided by sigma x sigma y jo ki hai 1875 divided by 75 times 82.92 aur iska maan hai kareeb 0.3015, to covariance to kaafi bada number tha lekin correlation coefficient sirf 0.3015 hai; isi tarah se exactly isi tarah second continuous wale case me covariance approx −0.0267 (−2/75) tha aur usme marginal pdfs same hone ke kaaran expected value of x square aur expected value of y square bhi same aaye 0.2; to variance(x) = $0.2 - (2/5)^2 = 0.04$, aur variance(y) bhi 0.04; to correlation coefficient = covariance / (σx σy) = (−2/75) / (0.2 × 0.2) = −2/3 ≈ −0.6667; isse hum dekh sakte hain ki pehle example me covariance 1875 bahut bada tha par correlation sirf 0.3 tha, aur second example me covariance bahut chhota tha par correlation −0.66 (kaafi strong) tha ye dikhata hai ki covariance number ka magnitude misleading ho sakta hai kyunki woh scale aur units par depend karta hai, jabki correlation normalized hota hai, units se free hota hai, aur behtar measure hota hai sambandh ka; continuous distribution ke example me condition x + y < 1 ke kaaran x aur y directly negatively related hainx bada to y chhota, x chhota to y bada iske kaaran correlation negative aur strong; ab kuch propositions: (A) correlation units par depend nahi karta corr(AX+B, CY+D) = corr(X,Y); (B) correlation hamesha −1 aur +1 ke beech hota hai; (C) agar X aur Y independent hain to

corr(X,Y) = 0 (lekin converse true nahi corr = 0 hone par zaroori nahi X,Y independent ho); (D) correlation = ±1 tab aur sirf tab hota hai jab X aur Y ke beech strict linear relationship ho: Y = aX + b, jahan a ≠ 0, agar a positive hai to corr = +1 aur agar a negative hai to corr = −1; aur last mein ek example batata hai ki corr = 0 hone ka matlab relationship nahi ek PMF me sirf 4 possible points (−4,1), (4,−1), (2,2), (−2,−2) hain yahan X aur Y directly related hain, independence nahi, par expected value of X, expected value of Y, expected value of XY sab 0 ke karan correlation 0; ye batata hai ki correlation sirf linear relationship measure karta hai, general dependency nahi  isliye correlation coefficient ko samajhte waqt ye concept bahut mahatvapurna hai.thank you