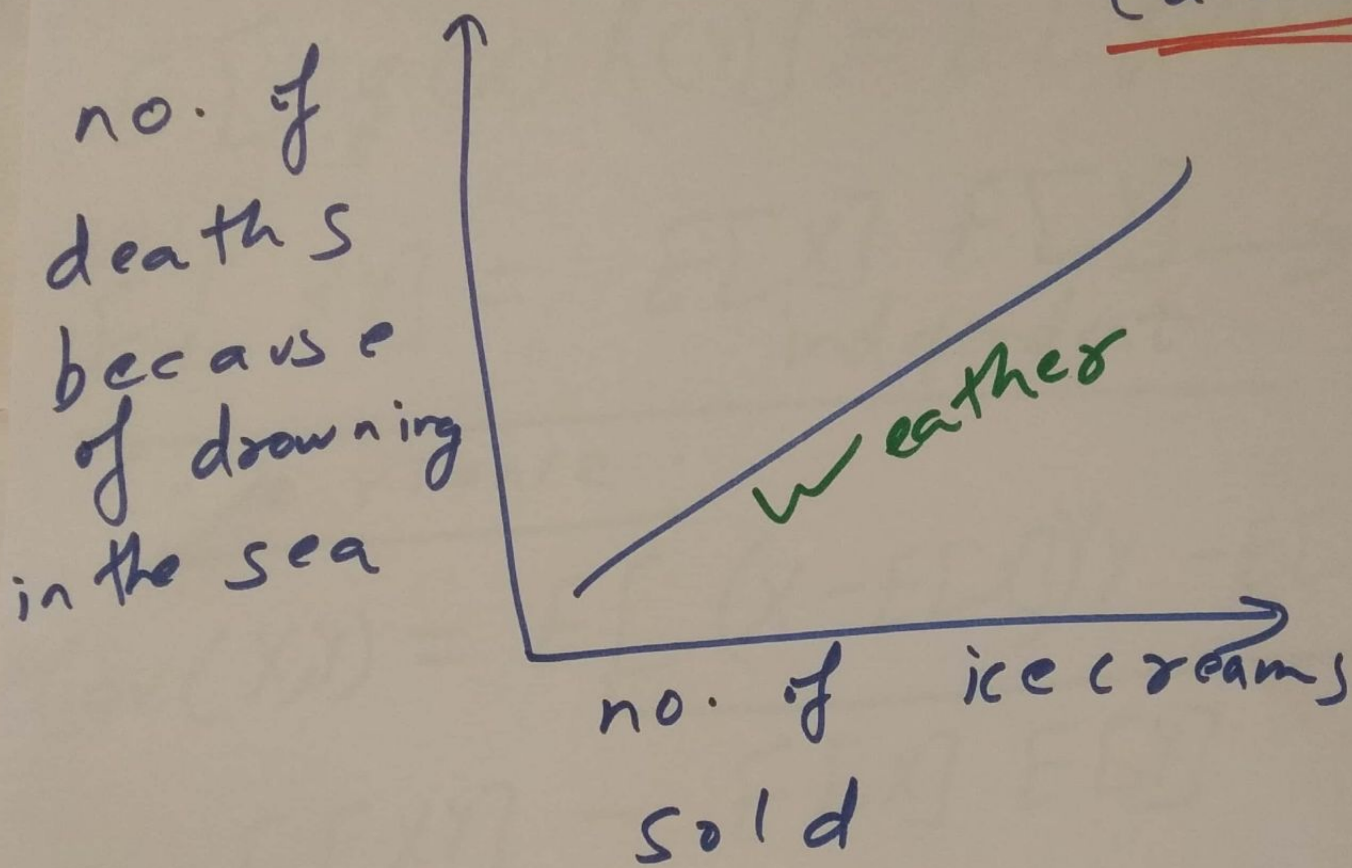
Lecture - 26 P 1

Covariance and Correlation

Causality



No causality here.

Bothe are dependent on the weather.

Theorem:

X, Y are independent.

$$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)]$$

$$E[XY] = E[X]\,E[Y] \rightarrow$$ independent

Covariance:

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

$$= E[XY] - E[X]\,E[Y]$$

if X & Y are independent,

then $\mathrm{Cov}(X,Y) = 0$

X = marks in exams

Y = attendance in lectures

covariance,

correlation

e.g.

$$X = \begin{cases} 0 & 1/3 \\ 1 & 1/3 \\ -1 & 1/3 \end{cases}$$

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{if } X = 0 \end{cases}$$

$$Cov(X,Y) = ? = 0$$

$$\underbrace{E[XY]}_{=\,0} - \underbrace{E[X]}_{=\,0}\ \underbrace{E[Y]}_{=\,1/3} = 0$$

$$P(X=0 \wedge Y=0) = 0 = L.H.S.$$

$$= P(X=0) * P(Y=0)$$

$$= \frac{1}{3} * \frac{2}{3} = \frac{2}{9} = R.H.S.$$

if X & Y are independent then $Cov(X,Y)=0$

if $Cov(X,Y)=0$, then X & Y need not be independent, i.e., they may be dependent

---

Some properties of Covariance.

i) $Y = X$

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

$$Cov(X,X) = E[X^2] - (E[X])^2$$

$$= Var(X)$$

---

ii) $Cov(X,Y) = Cov(Y,X)$

---

iii) $Cov(aX,Y) = a\,Cov(X,Y)$

$$Cov(aX, aY) = a^2\,Cov(X,Y)$$

$$Var(aX) = a^2\,Var(X)$$

iv) Covariance is additive. 5

$$Cov(X+Y, Z) =$$

$$Cov(X, Z) + Cov(Y, Z)$$

$$Cov(\sum X_i, \sum Y_j) = \sum_i \sum_j Cov(X_i, Y_j)$$

$$Cov(X_1 + X_2 + X_3, Y_1 + Y_2) =$$

$$Cov(X_1, Y_1) + Cov(X_1, Y_2) +$$

$$Cov(X_2, Y_1) + Cov(X_2, Y_2) +$$

$$Cov(X_3, Y_1) + Cov(X_3, Y_2)$$

Variance of sum of random variables

$$Var(\sum_{i=1}^{n} X_i)$$

$$= Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j)$$

$Cov(\sum X_i, \sum X_j)$ ⑥

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} \boxed{Cov(X_i, X_i)} + \sum_{i \neq j} \sum Cov(X_i, X_j)$$

$n^2 - n$ terms

$$= \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} \sum Cov(X_i, X_j)$$

if X_i & X_j are independent, then all covariance terms $= 0$

$$= \sum_{i=1}^{n} Var(X_i)$$

when X_i are independent variance of sum = sum of variance e.g. Binomial random variable

(n, p)

$Var(X) = np(1-p)$, already know this

You want to define X_i

s.t. $\boxed{\sum X_i = X}$ and

X_i are independent of

each other.

$Var(X) = Var(\sum X_i)$

$= \sum Var(X_i)$

because X_i are independent.

$$X_i = \begin{cases} 1 & \text{if you get a success on the } i\text{th trial} \\ 0 & \text{if you fail on the } i\text{th trial} \end{cases}$$

$i = 1$ to n

X_i are independent

$$Var(X) = \sum_{i=1}^{n} Var(X_i)$$

$$Var(X_i) = ?$$

$$= E[X_i^2] - \left(E[X_i]\right)^2$$

$$X_i \begin{array}{ll} \nearrow 0 & (1-p) \\ \searrow \underline{1} & p \end{array}$$

$$E[X_i] = p$$

$$E[X_i^2] = 0^2(1-p) + 1^2(p)$$

$$= p$$

$$Var(X_i) = p - p^2 = p(1-p)$$

$$Var(X) = n \cdot Var(X_i) = \underline{\underline{np(1-p)}}$$

Correlation:

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

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

$= 0$ if X & Y are independent

$\rho(X,Y)$ is defined when $\sigma_X \neq 0$ AND $\sigma_Y \neq 0$.

Both need to be non-zero.

When is $Var(X) = 0$?

when X is constant

Correlation is defined when both random variables are non-constants.

$$|\rho(X,Y)| \leq 1$$

$$-1 \leq \rho(X,Y) \leq 1$$

$\rho(X,Y) = 0$ when X & Y are independent

$$Z = \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$$

$$Var(Z) \geq 0$$

$$Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0$$

$$Var\left(\sum X_i\right) = \sum Var(X_i) + \sum_{i \neq j}\sum Cov(X_i, X_j)$$

(11)

$$\mathrm{Var}\left(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}\right)$$

$$= \boxed{\mathrm{Var}\left(\frac{X}{\sigma_X}\right)} + \boxed{\mathrm{Var}\left(\frac{Y}{\sigma_Y}\right)}$$

$$+ 2\,\mathrm{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

$$= \frac{1}{\sigma_X^2}\mathrm{Var}(X) + \frac{1}{\sigma_Y^2}\mathrm{Var}(Y)$$

$$+ \frac{2}{\sigma_X \sigma_Y}\mathrm{Cov}(X,Y)$$

$$= 2 + \frac{2}{\sigma_X \sigma_Y}\mathrm{Cov}(X,Y)$$

$$= 2 + 2\rho(X,Y) \geq 0$$

$$\Rightarrow \boxed{\rho(X,Y) \geq -1}$$

How do you prove that $\rho(X, Y) \leq 1$?

$$W = \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$$

H.W.

When is $\rho = -1$?

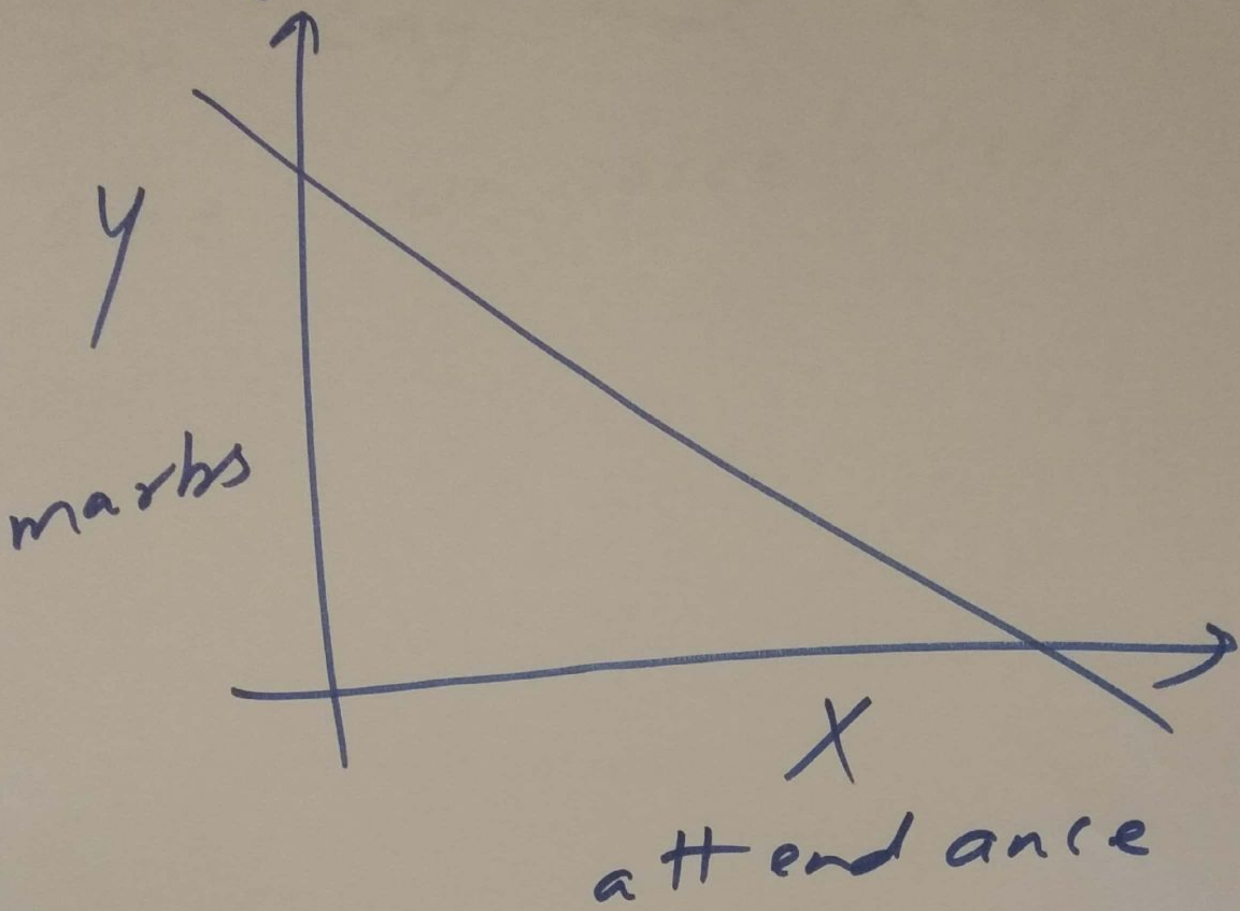
when $Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0$

$$\Rightarrow \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c$$

$$Y = c \cdot \sigma_Y - \frac{\sigma_Y}{\sigma_X} \cdot X$$

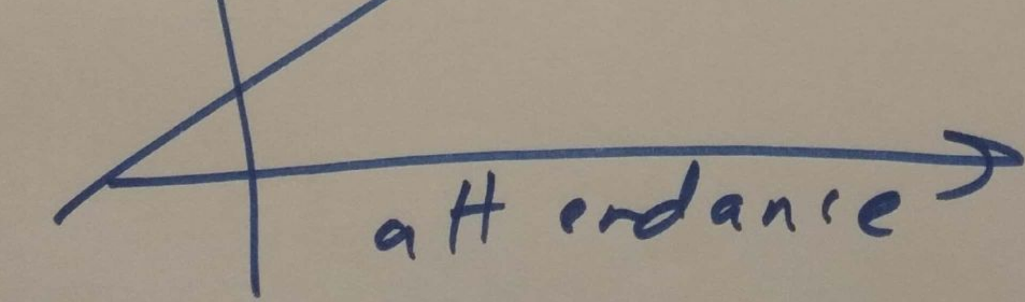
$$Y = a - bX$$

$\rho = -1$



$\rho = 1$

y is linear in x



When $\rho = 0$, then 14

you say that X & Y

are uncorrelated.