25ο Μάθημα

22/4/13

Συντελεστής (γραμμικής) συσχέτισης και Συνδιακύμανση

① Παράδειγμα

Πείραμα Τύχης: Επιλογή ατόμου από πληθυσμό

X = Βάρος σε kgr

Y = ύψος σε m

$Cov[X,Y]$

Ένας άλλος μελετητής:

X' = Βάρος σε lb

Y' = Ύψος σε ft

$Cov[X',Y']$

<u>Γενικά</u> $Cov[X,Y] \neq Cov[X',Y']$

διότι $X' = aX$ a: σταθερά μετατροπής kgr → lb

$Y' = cY$ c: σταθερά μετατροπής m → ft

$$Cov[X',Y'] = Cov[aX, cY] = ac\,Cov[X,Y]$$

Ελάττωμα! της Cov: Εξάρτηση από τις μονάδες μετατροπής

② Ορισμός

Συντελεστής (γραμμικής) συσχέτισης των τ.μ. X, Y ορίζεται:

$$\rho(X,Y) = \frac{\overbrace{Cov[X,Y]}^{\sigma_{X,Y}}}{\underbrace{\sqrt{Var[X]}}_{\sigma_X}\,\underbrace{\sqrt{Var[Y]}}_{\sigma_Y}}$$

③ Ιδιότητες

1) $\rho(aX, bY) = \begin{cases} \rho(X,Y) & , \text{αν } ab>0 \\ -\rho(X,Y) & , \text{αν } ab<0 \end{cases}$

2) $\rho(X,Y) \in [-1, 1]$ $\left(\begin{matrix} -1 \\ 1 \end{matrix} \right.$ > ακραίες συσχετίσεις $\left. \right)$

3) X, Y ασυσχέτιστες $\Rightarrow \rho(X,Y) = 0$

4) $\rho(X,Y) = 1 \iff Y = aX + b \quad , a>0$

5) $\rho(X,Y) = -1 \iff Y = aX + b \quad , a<0$

6) **Προσοχή!**

Ενώ $Cov[X+Y, Z] = Cov[X, Z] + Cov[Y, Z]$

$\rho(X+Y, Z) \neq \rho(X,Z) + \rho(Y,Z)$

Αποδείξεις:

1) $\rho(aX, bY) = \dfrac{Cov[aX, bY]}{\sqrt{Var[aX]}\sqrt{Var[bY]}} = \dfrac{ab}{|ab|} \underbrace{\dfrac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}}}_{\rho(X,Y)}$

$= \begin{cases} \rho(X,Y) & , \text{αν } ab>0 \\ -\rho(X,Y) & , \text{αν } ab<0 \end{cases}$

2)

$Var\left(\dfrac{X}{\sigma_X} + \dfrac{Y}{\sigma_Y}\right) \geq 0 \Rightarrow Var\left(\dfrac{X}{\sigma_X}\right) + Var\left(\dfrac{Y}{\sigma_Y}\right) + 2Cov\left[\dfrac{X}{\sigma_X}, \dfrac{Y}{\sigma_Y}\right] \geq 0$

$\Rightarrow \dfrac{1}{\sigma_X^2} Var[X] + \dfrac{1}{\sigma_Y^2} Var[Y] + 2\dfrac{Cov[X,Y]}{\sigma_X \sigma_Y} \geq 0 \Rightarrow$

$\Rightarrow 2 + 2\rho(X,Y) \geq 0 \Rightarrow \rho(X,Y) \geq -1$

και

$$Var\left(\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y}\right)\geq 0 \Rightarrow Var\left(\frac{X}{\sigma_X}\right)+Var\left(\frac{Y}{\sigma_Y}\right)-2Cov\left[\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right]\geq 0$$

$$\Rightarrow \frac{1}{\sigma_X^2}Var[X]+\frac{1}{\sigma_Y^2}Var[Y]-2\frac{Cov[X,Y]}{\sigma_X\sigma_Y}\geq 0 \Rightarrow$$

$$\Rightarrow 2-2\rho(X,Y)\geq 0 \Rightarrow \rho(X,Y)\leq 1$$

$$\left(E[X-E[X]](Y-E[Y])\right)^2 \leq E[(X-E[X])^2]E[(Y-E[Y])^2] \quad C-S$$

3) X, Y ασυσχέτιστα $\Leftrightarrow Cov[X,Y]=0 \Leftrightarrow \rho(X,Y)=0$

5) $\Rightarrow$) $\rho(X,Y)=-1 \Rightarrow Var\left[\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}\right]=0 \Rightarrow \frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}=c$, σταθ

$$\Rightarrow Y=-\frac{\sigma_Y}{\sigma_X}X+c\sigma_Y \Rightarrow Y=aX+b,\ a<0$$

$\Leftarrow$)

$$Y=aX+b,\ a<0 \Rightarrow \rho(X,Y)=\frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$

$$=\frac{Cov[X,aX+b]}{\sqrt{Var[X]}\sqrt{Var[aX+b]}}$$

$$=\frac{a\,Cov[X,X]}{|a|\sqrt{Var[X]}\sqrt{Var[X]}}=-1$$

4) ΟΜΟΙΑ με το 5)

$$\rho(X,Y)=1 \Rightarrow Var\left[\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y}\right]=0 \Rightarrow \dots \Rightarrow Y=aX+b,\ a>0$$

Ξέρουμε:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{1\le i\le j\le n} Cov[X_i, X_j]$$

$$Cov\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i, Y_j]$$

④ Εφαρμογή: Εκτίμηση $\mu = E[X]$, $\sigma_x^2 = Var[X]$

$X_1, X_2, \dots, X_n$ n παρατηρήσεις, ανεξ. ισόνομες

$\mu = ;$ $\sigma_x^2 = ;$

Για να εκτιμήσουμε το μ χρησιμοποιούμε τον $\underbrace{\text{μέσο όρο των παρατηρήσεων}}_{\text{Δειγματικός μέσος}}$.

Δειγματικός μέσος ⟵ μέσος του δείγματος

$$\bar{X} = \bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$\mu = E[X_i]$ ⟵ μέσος του πληθυσμού

Είναι σωστό να χρησιμοποιούμε το $\bar{X}$ για να εκτιμήσουμε το μ;

Πρέπει $E[\bar{X}] = \mu$, $Var[\bar{X}] \searrow$ ως προς n (φθίνουσα)
και μάλιστα $\to 0$

Πράγματι,

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$ ✓

$$Var[\bar{X}] = Var\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] \overset{\text{Ανεξ.}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{\cancel{n}\sigma_x^2}{n^{\cancel{2}}} = \frac{\sigma_x^2}{n} \downarrow 0$$

Ερώτημα! Πώς να εκτιμήσουμε το $\sigma_x = E[(x-\mu)^2]$;

Κατά τη λογική του δειγματικού μέσου μία εκτιμήτρια (τυχαία μεταβλητή) για το σ_x^2 μέσω του δείγματος είναι η

$$T = \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\dots+(X_n-\bar{X})^2}{n}$$

Είναι η T καλή εκτιμήτρια του σ_x^2;

Για να είναι η T καλή εκτιμήτρια του σ_x^2 πρέπει:

$E[T] = \sigma_x^2$

$Var[T] \searrow 0$ για $n \to \infty$

$$E[T] = E\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\bar{X})^2] = \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu+\mu-\bar{X})^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(E[(X_i-\mu)^2]+E[2(X_i-\mu)(\mu-\bar{X})]+E[(\mu-\bar{X})^2]\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(E[(X_i-\mu)^2])+2n\frac{\sum_{i=1}^{n}(X_i-\mu)(\mu-\bar{X})}{n}+nE[(\mu-\bar{X})^2]\right)$$

$$= \frac{1}{n}\left(n\sigma_x^2+2nE\left[\frac{\sum_{i=1}^{n}X_i-n\mu}{n}(\mu-\bar{X})\right]+nE[(\mu-\bar{X})^2]\right)$$

$$= \frac{1}{n}\left(n\sigma_x^2-2nVar[\bar{X}]+nVar[\bar{X}]\right)$$

$$= \frac{1}{n}\left(n\sigma_x^2-\cancel{n}\frac{\sigma_x^2}{\cancel{n}}\right) = \frac{(n-1)\sigma_x^2}{n}$$

Η T δεν είναι "καλή" διότι $E[T] = \frac{(n-1)\sigma_x^2}{n} \neq \sigma_x^2$

Αν θέσω $S = \frac{n}{n-1}\cdot T = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ τότε "

$$E[S] = \frac{n}{n-1}E[T] = \sigma_x^2.$$

Αποδεικνύεται ότι $Var[S] \longrightarrow$ καθώς $n \longrightarrow \infty$
Άρα για την εκτίμηση

$\mu = E[X_i]$ χρησιμοποιούμε $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ Δειγμ. μέσος

$\sigma_x^2 = Var[X_i]$ — // — $S = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$ Αμερ. δειγμ. διασπορά