10.05.10 27ο μάθημα

Συνδιακύμανση - Συντελεστής Συσχέτισης

① "Ελάττωμα" συνδιακύμανσης

Τυχαίο Πείραμα: Μέτρηση ύψους-βάρους

X = ύψος σε cm

Y = βάρος σε kgr $\rightarrow Cov(X,Y)$

X' = ύψος σε ft $= c_1 X$

Y' = βάρος σε lb $= c_2 Y$ $\rightarrow Cov(X',Y') = Cov(c_1X, c_2Y) = c_1c_2\,Cov(X,Y)$

Η συνδιακύμανση εξαρτάται από τις μονάδες μέτρησης. Γι' αυτό εισάγεται ο συντελεστής συσχέτισης.

② Ορισμός

X, Y τ.μ. με $E[X] = \mu_X$, $E[Y] = \mu_Y$, $Var[X] = \sigma_X^2$, $Var[Y] = \sigma_Y^2$ και $Cov[X,Y] = \sigma_{X,Y}$, τότε

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

↑

συντελεστής συσχέτισης

③ Ιδιότητες

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

   Ο συντελεστής συσχέτισης είναι ανεξάρτητος των μονάδων μέτρησης

2. $\rho(X,Y) \in [-1, 1]$
3. $\rho(X,Y) = 0 \Leftrightarrow X, Y$ ασυσχέτιστες

4. $\rho(X,Y) = -1 \Leftrightarrow Y = aX + b \ , \ a<0$

5. $\rho(X,Y) = 1 \Leftrightarrow Y = aX + b \ , \ a>0$

Αποδείξεις:

1. $$\rho(aX, bY) = \frac{Cov[aX, bY]}{\sqrt{Var[aX]Var[bY]}} = \frac{ab\,Cov[X,Y]}{|a|\cdot|b|\sqrt{Var[X]Var[Y]}} = \begin{cases} \rho(X,Y), & ab>0 \\ -\rho(X,Y), & ab<0 \end{cases}$$

$\rho(X,Y)$

2. $$Var\left[\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right] \geq 0 \Leftrightarrow Var\left[\frac{X}{\sigma_X}\right] + Var\left[\frac{Y}{\sigma_Y}\right] + 2Cov\left[\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right] \geq 0$$

$$\Leftrightarrow \underbrace{\frac{Var[X]}{\sigma_X^2}}_{=1} + \underbrace{\frac{Var[Y]}{\sigma_Y^2}}_{=1} + 2\underbrace{\frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}}_{=\rho(X,Y)} \geq 0$$

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

Όμοια:

$$Var\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right] \geq 0 \Leftrightarrow \rho(X,Y) \leq 1.$$

3. Προφανές

4. $(\Rightarrow)$ $$\rho(X,Y) = -1 \Leftrightarrow Var\left[\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right] = 0 \Leftrightarrow \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c, \text{ σταθερά}$$

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

$(\Leftarrow)$ $$Y = aX + b \ , \ a<0 \Rightarrow \rho(X,Y) = \rho(X, aX+b) = -\rho(X,X) = -\frac{\sigma_{X,X}}{\sigma_X\sigma_X} = -1$$

Προσοχή! Ο $\rho(X,Y)$ δεν είναι γραμμικός τελεστής

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

5. Όμοια

⑤ Παράδειγμα (Το πρόβλημα των συναντήσεων)

n άτομα και τα n καπέλα τους

Τα πετάνε κι ο καθένας παίρνει ένα στην τύχη

$N = \#$ ατόμων που πήραν το καπέλο τους

(Έχουμε δει $P(N=0) = \sum_{i=0}^{n} \frac{(-1)^i}{i!}$ με αρχή εγκλεισμού-αποκλεισμού)

$E[N] =;$

$Var[N] =;$

$N = \sum_{i=1}^{n} I_i$ με $I_i = \begin{cases} 1, & \text{αν ο } i \text{ πήρε το καπέλο του} \\ 0, & \text{διαφορετικά} \end{cases}$

$E[N] = E[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} E[I]$

$Var[N] = Var[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} Var[I_i] + 2\sum_{1\leq i<j\leq n} Cov[I_i, I_j]$

$E[I_i] = P(I_i = 1) = P(\text{ο } i \text{ να πάρει το καπέλο του}) = \frac{1}{n}$

$E[I_i^2] = \frac{1}{n}$

$Var[I_i] = E[I_i^2] - E^2[I_i] = \frac{1}{n} - (\frac{1}{n})^2 = \frac{n-1}{n^2}$

$i \neq j$ $Cov[I_i, I_j] = E[I_i I_j] - E[I_i]E[I_j] = P\left(\begin{matrix}\text{οι } i,j \text{ να} \\ \text{πάρουν το} \\ \text{καπέλο τους}\end{matrix}\right) - \frac{1}{n^2}$

$\quad = \frac{1}{n} \cdot \frac{1}{n-1}$

άρα $E[N] = n \cdot \frac{1}{n} = 1$

π.χ. για $n=3$:

1	2	3		
1	2	3	$\frac{1}{6}$	$N=3$
1	3	2	0	$N=2$
2	1	3	$\frac{1}{2}$	$N=1$
2	3	1	$\frac{1}{3}$	$N=0$
3	1	2		
3	2	1	$E[N] = \frac{1}{6} \cdot 3 + \frac{1}{2} \cdot 1 = 1$	

$Var[N] = n\frac{n-1}{n^2} + 2\binom{n}{2}\left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) = 1$

$\binom{n}{2} = \frac{n(n-1)}{2}$

⑥ Παράδειγμα

$X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τ.μ. με $E[X_i]=\mu$, $Var[X_i]=\sigma^2$
ίδια κατανομή

Στη στατιστική παίρνουμε το $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ για να εκτιμήσουμε το μ
δειγματικός μέσος

Για να είναι η εκτίμηση δικαιολογημένη περιμένω $E[\bar{X}]=\mu$ και $Var[\bar{X}] \searrow$ ως προς n.

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

$$Var[\bar{X}] = Var\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \searrow$$

Για να εκτιμήσω τη διασπορά χρησιμοποιώ ...

Μια καλή ιδέα για να εκτιμήσουμε τη διασπορά σ^2 φαίνεται να είναι η $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$ ισχύει όμως $E\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}\right] = \sigma^2$;

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

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

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

$= n\sigma^2 - 2n E[(\bar{X}-\mu)^2] + n E[(\bar{X}-\mu)^2] = n\sigma^2 - n E[(\bar{X}-\mu)^2]$

Όμως $E[\bar{X}]=\mu$ οπότε $E[(\bar{X}-\mu)^2] = Var[\bar{X}] = \frac{\sigma^2}{n}$

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

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

Γι' αυτό στη Στατιστική χρησιμοποιούμε

$S^2 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ · που έχει την ιδιότητα $E[S^2] = \sigma^2$

↳ Αμερόληπτη Δειγματική Διασπορά