27ο Μάθημα Πιθανοτήτων 10/5/2010

Συνδιακύμανση - Συντελεστής Συσχέτισης

① "Ελάττωμα" Συνδιακύμανσης

Τυχαίο Πείραμα: Μέτρηση ύψους-βάρους

X = ύψος σε cm $\longrightarrow Cov(X,Y)$

Y = βάρος σε kgr

X' = ύψος σε ft $= c_1 X \longrightarrow Cov(X',Y') = Cov(c_1X, c_2Y) = c_1c_2\,Cov(X,Y)$

Y' = ύψος σε lb $= c_2 Y$

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 (γραμμικής) 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 συντελ. συσχετ. είναι ανεξ. 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) $\rho(X,Y) = 1 \Leftrightarrow Y = aX + b,\ a>0$.

$\rho(X,Y)$

ΑΠΟΔΕΙΞΕΙΣ:

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

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

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

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

3) Προφανές

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

$\Rightarrow Y = -\dfrac{\sigma_Y}{\sigma_X}X + \sigma_Y\cdot c = aX+b, \quad a<0.$

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

5) Όμοια

④ Π̲Ρ̲Ο̲Σ̲Ο̲Χ̲Η̲!!!

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

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

⑤ Π̲α̲ρ̲ά̲δ̲ε̲ι̲γ̲μ̲α̲ (Το πρόβλημα συναντήσεων)

n άτομα και n τα καπέλα τους

Τα πετάνε και ο καθένας παίρνει ένα

N = # ατόμων που πήραν το καπέλο τους

(έχουμε ότι $P(N=0) = \sum_{k=0}^{n} \dfrac{(-1)^k}{k!}$ με αρχή εγκλ.-αποκλ.)

$E[N] = ;\quad Var[N] = ;$

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

$E[N] = E[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} E[I_i]$, $Var[N] = Var[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} Var[I_i] +$

$+ 2 \sum_{1 \le i < j \le n} Cov[I_i, I_j]$

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

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

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

$i \neq j$ $Cov[I_i, I_j] = E[I_i, I_j] - E[I_i]E[I_j] = P\underbrace{\begin{pmatrix}\text{οι } i, j \text{ να} \\ \text{πάρουν το καπ.}\end{pmatrix}}_{\frac{1}{n} \cdot \frac{1}{n-1}} - \frac{1}{n^2}$

Άρα: $E[N] = n \cdot \frac{1}{n} = 1$.

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		

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

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

6) Παράδειγμα

$X_1, X_2, \dots, X_n$ ανεξ. και ισόνομες τ.μ. με $E(X_i) = \mu$ (ίδια κατανομή)

και $Var[X_i] = \sigma^2$

Στη Στατιστική παίρναμε το $\sum_{i=1}^{n} X_i / n = \bar{X}$ (δειγματικός μέσος) για να εκτιμήσουμε το μ.

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

Πράγματι:

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

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

Για να εκτιμήσω τη διασπορά χρησιμοποιώ $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$, $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$

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

$a - b$

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

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

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

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

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

Γι' αυτό στη Στατιστική χρησιμοποιούμε: $S^2 = \frac{\sum_{i=1}^{n}(X-\bar{X})^2}{n-1}$ που

αμερόληπτη
δειγματική διασπορά

έχει την ιδιότητα: $E[S^2] = \sigma^2$

ΓΕΝΙΚΑ: $\mu \leftarrow \frac{\sum_{i=1}^{n} X_i}{n}$ $\qquad \sigma^2 \leftarrow \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$