Μαθημα 16ο (1) 4/5/2017

Στατιστικη

Μετρα επαναφορας στην καταλληλοτητα του Μοντελου

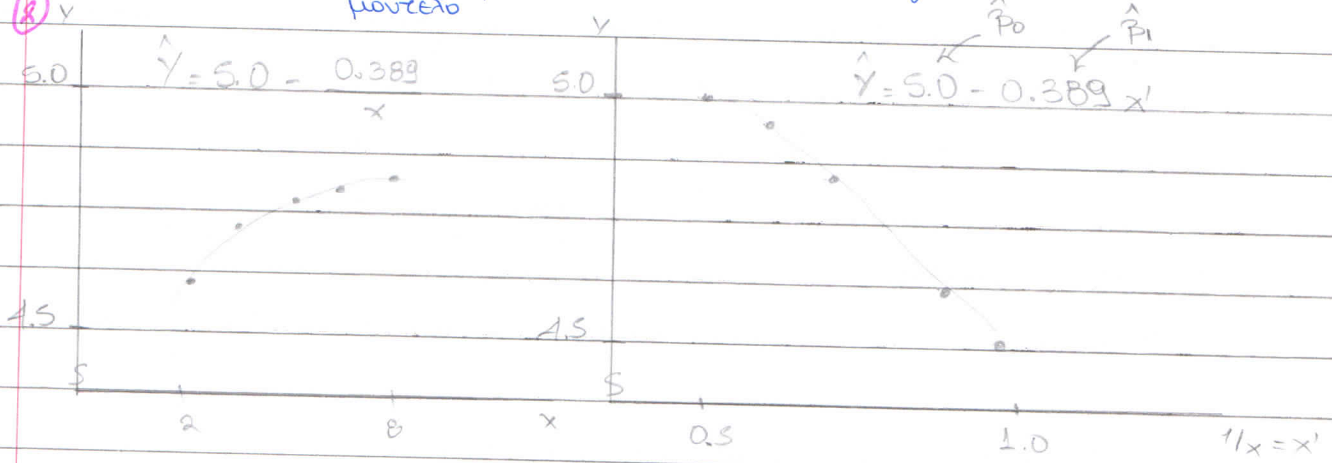
Π.χ. (1) $Y=\gamma_0 \cdot \gamma_1^{x} \cdot \varepsilon$, $E(\varepsilon)=1$, $Var(\varepsilon)=\sigma^2$

$$\underbrace{\log_{10} Y}_{=Y'} = \log_{10}\gamma_0 + \log_{10}\gamma_1^{x} + \log_{10}\varepsilon = \underbrace{\log_{10}\gamma_0}_{=\beta_0} + x\underbrace{(\log_{10}\gamma_1)}_{=\beta_1} + \underbrace{\log_{10}\varepsilon}_{=\varepsilon}$$

Αρα $Y'=\beta_0+\beta_1 x+\varepsilon$

$(x_i, \log_{10} Y_i) \xrightarrow[\text{απλο γραμ. μοντελο}]{} \hat\beta_0, \hat\beta_1 \longrightarrow \hat Y = \hat\gamma_0 \cdot \hat\gamma_1^{x}$

(2)



Αντιστροφος μετασχηματισμος

$Y=\beta_0+\frac{\beta_1}{x}+\varepsilon \xrightarrow{x'=1/x} Y=\beta_0+\beta_1 x'+\varepsilon$

$\hat Y=\hat\beta_0+\frac{\hat\beta_1}{x} \leftarrow \hat Y=\hat\beta_0+\hat\beta_1 x'$

x : χρονια εμπειρια , Y_i : κερδη/h

Πολλαπλη Γραμμικη παλινδρομηση (Y, X ποσοτικες μεταβλητες)

$Y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\dots+\beta_k x_{ki}+\varepsilon_i$ γενικο γραμ. μοντελο παλινδρ.

$Y=\beta_0+\beta_1 x+\beta_2 x^2+\dots+\beta_k x^k+\varepsilon$ πολυωνυμικο μοντελο

→ αν βαλω $x_1=x$, $x_2=x^2$, …, $x_k=x^k$ στην Y_i χρησιμοποιω το γενικο μοντελο π.χ. ποιοτικές

άνδρας, γυναίκα

φτωχός, πλούσιος

A , B

κωδικές τιμές χρησιμοποιώ

δοθω 0 αν είναι άνδρας και 1 αν είναι γυναίκα

ανάλογα με τις κατηγορίες

{ ⊝ Μοντέλο ανάλυσης συνδιακύμανσης όταν έχω ποιοτική και ποσοτική μεταβλητή μαζί }

Συσχέτιση

Για δύο τυχαίες μεταβλητές X και Y ένα μέτρο της συμμεταβλητότητας είναι η συνδιακύμανση (αφορά διδιάστατη κατανομή)

→ συμβ: $(\sigma_{XY} =)\ Cov(X,Y) = E[(X-EX)(Y-EY)] =$

$$= E(XY) - (EX)(EY)$$

$\sigma^2_X \to var(X)$

{ ⊛ μονάδες μέτρ: το γινόμενο των αναμενόμενων τιμών (δηλ. π.χ. $m \cdot m = m^2$) }

Ο "θεωρητικός" συντελεστής συσχέτισης των (X,Y):

$$\rho = \rho(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{varX \cdot varY}} \quad , \ -1 \le \rho \le 1$$

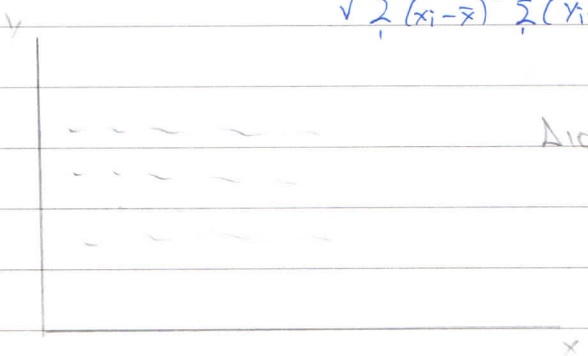
→ καθαρός αριθμός

- Αν $\rho = 0$, οι X και Y ασυσχέτιστες, δε σχετίζονται γραμμικά
- Αν $\rho = \pm 1$, οι X και Y σχετίζονται γραμμικά σε τέλεια σχέση

ρ παράμετρος ⇒ δεν το ξέρω ⇒ χρειάζομαι εκτιμητή

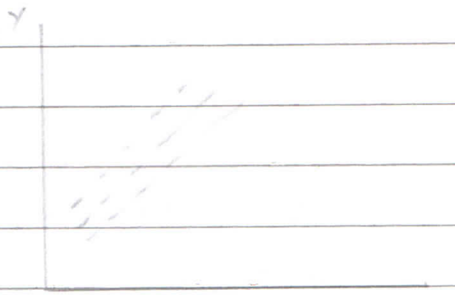
Για τα ζεύγη (x_i, y_i) ο δειγματικός συντελεστής συσχέτισης:

$$r = r(x,y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i-\bar{x})^2 \sum_1^n (y_i - \bar{y})^2}} \quad , \ -1 \le r \le 1$$

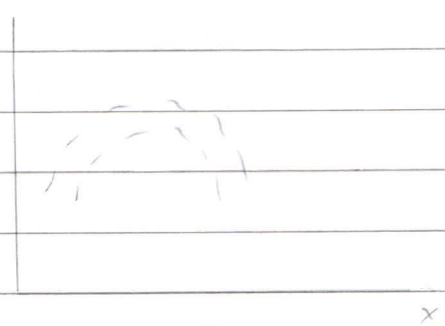


Διάγραμμα Διασποράς

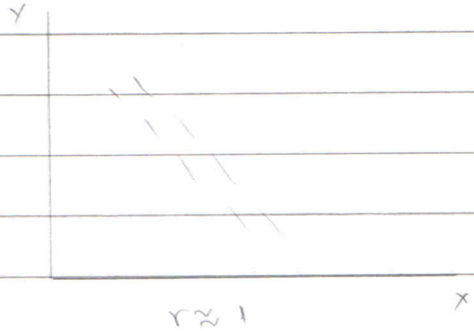
$r \approx 0$



$r \approx 1$



$r \approx 0$



$r \approx 1$

→ Πως η συσχετιση ειναι αμεσα συνδεδεμενη με την απλη γραμ. παλ.

$R^2 = \frac{SS_{reg}}{SS_{tot}}$

$$r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2}} \Rightarrow r^2 = \frac{\overbrace{\Sigma(x_i-\bar{x})(y_i-\bar{y})}^{=\hat{\beta}_1}\;\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2\;\Sigma(y_i-\bar{y})^2} =$$

$$= \frac{\hat{\beta}_1\,\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(y_i-\bar{y})^2} = \frac{SS_{reg}}{SS_{tot}} = R^2$$

→ ισοδυναμος με t-test για $H_0: \beta_1 = 0$

Αν X και Y κανονικες τ.μ τοτε $H_0: \rho = 0 \wedge H_a: \rho \neq 0$

$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$ και $|t| \geq t_{\frac{\alpha}{2}, n-2}$

Για (x_i, y_i) ΣΒΟ : $r = \frac{6800}{\sqrt{3400 * 13660}} = 0.9978$

$t = \frac{0.9978\sqrt{10-2}}{\sqrt{1-0.9978^2}} = 42.6 > t_{0.005,8}\,(=3.355)$ απορ H_0, ∃ γρ. σχεσ.

Παλινδρ : $SS_{reg} = \hat{\beta}_1^2 \sum(x_i-\bar{x})^2$ 1 $MS_{reg} = \frac{SS_{reg}}{1}$

$= \sum(\hat{y}_i - \bar{y})^2 = 226.95$

Υπόλοιπα : $SS_{res} = \sum(y_i - \hat{y}_i)^2$ $n-2$ $= 10$ $MS_{res} = \frac{SS_{res}}{n-2}$

$= SS_{tot} - SS_{reg}$

$= 21.23$

Σύνολο : $SS_{tot} = \sum(y_i - \bar{y})^2$ $n-1$

$= 96.18$ $= 248.18$ 9

$F = \frac{MS_{reg}}{MS_{res}}$, $F \geq F_{a,1,n-2}$

$96.18 > 5.12$ απορ $H_0 : \beta_1 = 0$

$n=11$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 9.27$, $H_0 : \beta_1 = 0$ v $H_a : \beta_1 \neq 0$

(ολη 5.1 οχι)

$\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = 1.44$ $r = \pm\sqrt{R^2}$

$y = 9.27 + 1.44x$ $R^2 = \frac{226.95}{248.18} \cdot 100\% = 91.4\%$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, $t = \frac{\hat{\beta}_1 - \beta_1}{s / \sqrt{\sum(x_i-\bar{x})^2}}$

$|t| \geq t_{\frac{a}{2}, n-2}$ $\to 1.44$

Δ.Ε : $(1-a)\,100\%$ για $\beta_1 : \hat{\beta}_1 \pm t_{\frac{a}{2}, n-2} \cdot \frac{s}{\sqrt{\sum(x_i-\bar{x})^2}}$ $\Rightarrow [1.41, 1.77]$

$t_{0.025,9}$ 2.262