ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ

Διάλεξη 17η

09/05/2018

Πολλαπλή γραμμική παλινδρόμηση:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$ (πολλαπλό γραμμικό μοντέλο)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

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

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3$$

Συσχέτιση: Για δύο τ.μ. ποσοτικές X και Y ένα

μέτρο συμμεταβλητότητας είναι η συνδιακύμανση (συνδιασπορά) και δίνεται από την σχέση:

$$\underset{\text{covariance}}{\text{cov}}(\underbrace{X,Y}_{\text{μπορεί να το δούμε να συμβολίζεται ως: } \sigma_{XY}}) = E(X-EX)(Y-EY) = E(XY) - (EX)(EY)$$

Ο συντελεστής συσχέτισης $\rho = \rho_{X,Y} = \rho(X,Y) = \dfrac{\text{cov}(X,Y)}{\sqrt{Var(X) \times Var(Y)}}$

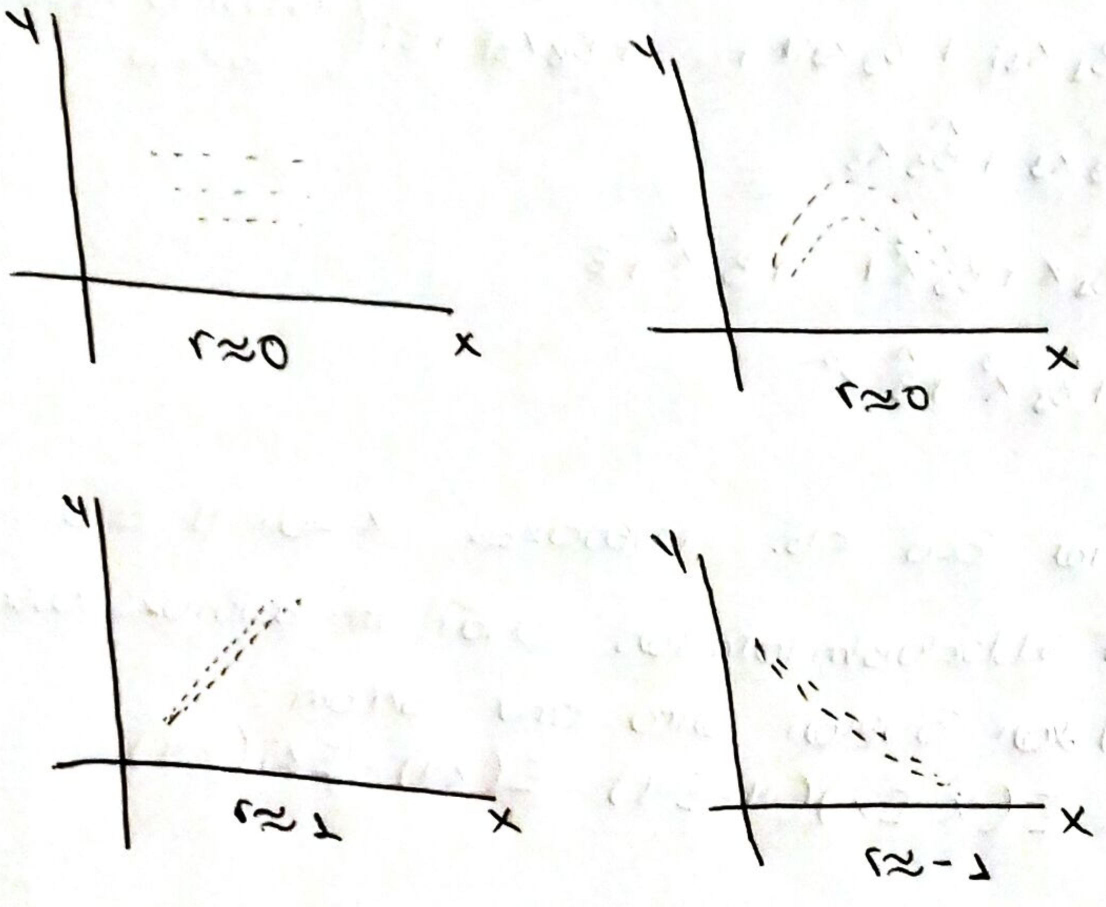
, $-1 \le \rho \le 1$.

για $\rho = 0$: X και Y γραμμικά ασυσχέτιστα

για $\rho = \pm 1$: X και Y γραμμικά εξαρτημένα σε τέλεια σχέση.

Ο συντελεστής συσχέτισης για $n(x_i, y_i)$ δείγμα δίνεται από την σχέση:

$$r = r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}}, \quad -1 \leq r \leq 1$$



$H_0: \rho = 0$ v $H_1: \rho \neq 0$

$t = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}} \overset{H_0}{\sim} t_{n-2}$, κρίσιμη περιοχή $|t| \geq t_{\frac{\alpha}{2}, n-2}$.

(ισοδύναμο με t-test για $B_1 = 0$)

r = δειγματικός συντελεστής συσχέτισης. Αν υψώσουμε το r στο τετράγωνο θα πάρουμε:

$$r^2 = \frac{\left[\sum_{i}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]\left[\sum_{i}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]}{\sum_{i}^{n}(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}$$

$$= \frac{\hat{\beta}_2 \sum_{i}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i}(y_i-\bar{y})^2} = \frac{SS_{reg}}{SS_{tot}} = R^2$$