(43) 6/5

Πολλαπλή Γραμμική Παλινδρόμηση

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k$$

$$Y = \beta_0 + \beta_1 X^1 + \dots + \beta_k X^k + \varepsilon$$

$\boxed{X_1 = X}$ ... $\boxed{X_k = X^2}$

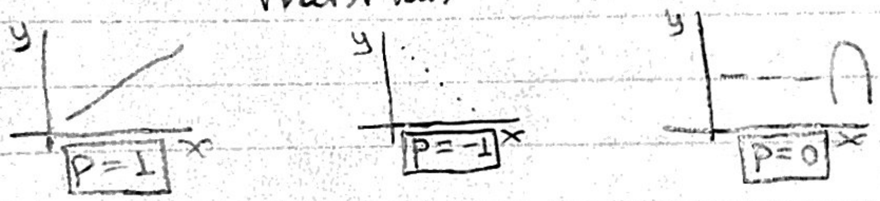
Συσχέτιση

Έστω ότι X και Y είναι ποσοτικές τ.μ.. Ένα μέτρο της συμμεταβλητότητας των X κ Y είναι η συνδιακύμανση (ή συνδιασπορά) και δίνεται από τον τύπο:

$$Cov(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - (EX)(EY)$$

→ Ο συντελεστής των τ.μ. X και Y δίνεται απ' τη σχέση:

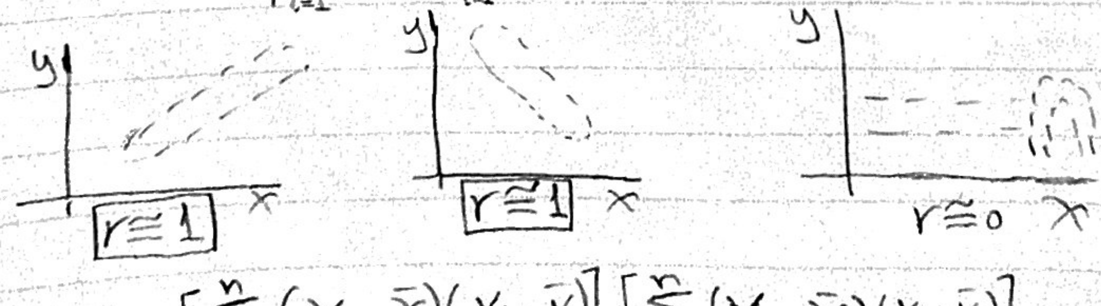
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{VarX \cdot VarY}} \quad \text{με } -1 \le \rho \le 1$$



ασυσχέτιστες γραμμικά

→ Ο δειγματικός συντελεστής συσχέτισης (του Pearson) για τα ζεύγη (X_i, Y_i) με $i=1,\dots,n$ δίνεται απ' τον τύπο:

$$r(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad \text{με } -1 \le r \le 1$$



$$r^2 = \frac{\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right] \cdot \left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]}{\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \cdot \left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]} =$$

$$= \hat{\beta}_1 \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \frac{SS_{reg}}{SS_{tot}} = R^2$$

συντελεστής προσδιορισμού

$H_0: \rho = 0$ έναντι $H_a: \rho \neq 0$

(Αν X, Y κανονικές τ.μ.) Χρησιμοποιούμε το στατιστικό:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2} \text{ όταν } \rho = 0$$

Και για επίπεδο σημαντικότητας α, η κρίσιμη περιοχή (δίπλευρη) είναι $|t| \geq t_{\alpha/2, n-2}$ (ισοδύναμο με τον t-τεστ εδώ για $\beta_1 = 0$)

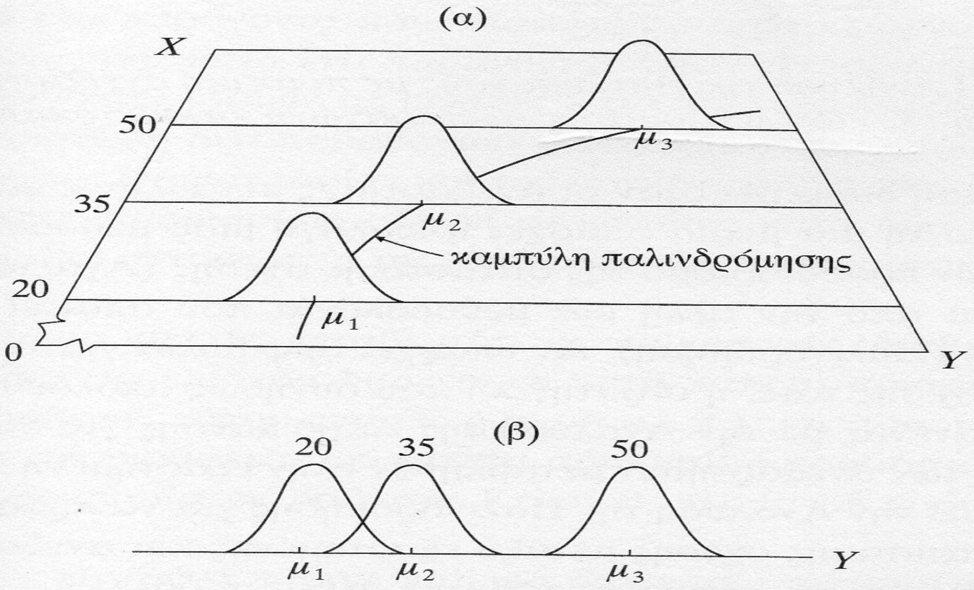
Πχ. (ΕΒΟ)

$$r = \frac{6800}{\sqrt{3400 \cdot 13.600}} = 0,9978$$

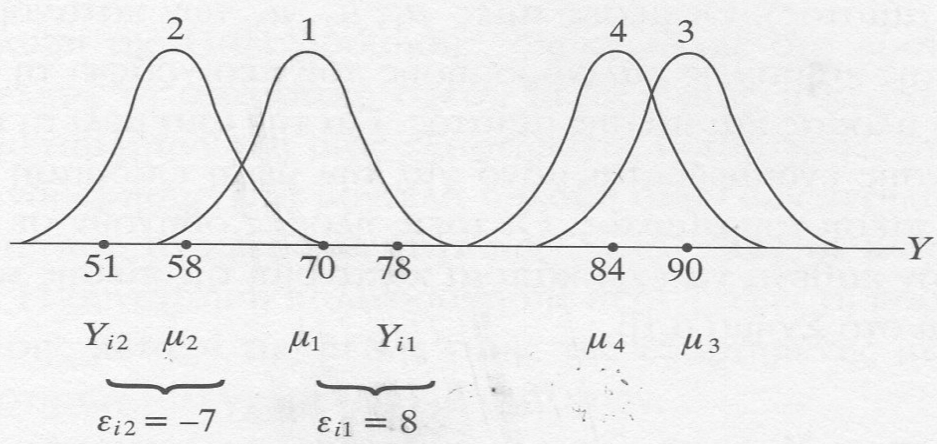
$$t = \frac{0,9978\sqrt{10-2}}{\sqrt{1-0,9956}} = 42,6 > 33,355 = t_{0,005,8}$$ Άρα απορ. την H_0

ίδιο με αυτό για $\beta_1 = 0$ έναντι $\beta_1 \neq 0$

§Κεφάλαιο 6 (Ανάλυση της διακύμανσης)



Σχέση μεταξύ Παλινδρόμησης και Ανάλυσης της Διακύμανσης (X: ηλικία, Y: πίεση αίματος)



Σχήμα 6.2 *Η Ανάλυση της Διακύμανσης με ποιοτική ανεξάρτητη μεταβλητή (X: είδη λιπάσματος, Y: παραγωγή καλαμποκιού)*