Φροντιστήριο 7ο

(1) X_1, X_2, X_3 ανεξάρτητες και ισόνομες με μέσο μ και διακύμανση σ^2. Ας είναι οι γραμμικοί συνδυασμοί αυτών $Y_1 = X_1 + X_2$ και $Y_2 = X_2 - X_3$.

N.δ.ο. $Cov(Y_1, Y_2) = \sigma^2$

Γενικά είναι $Cov(Y_1, Y_2) = E\left[(Y_1 - E(Y_1))(Y_2 - E(Y_2))\right]$

$$= E(Y_1 \cdot Y_2) - E(Y_1) \cdot E(Y_2)$$

Όμως:

$$E(Y_1) = E(X_1 + X_2) = E(X_1) + E(X_2) = \mu + \mu = 2\mu$$

και

$$E(Y_2) = E(X_2 - X_3) = E(X_2) - E(X_3) = \mu - \mu = 0$$

$$\Rightarrow Cov(Y_1, Y_2) = E(Y_1 \cdot Y_2) - E(Y_1) \cdot E(Y_2) = E(Y_1 \cdot Y_2) - \mu \cdot 0$$

$$= E(Y_1 \cdot Y_2) =$$

$$= E\left[(X_1 + X_2)(X_2 - X_3)\right] = E\left(X_1X_2 - X_1X_3 + X_2^2 - X_2X_3\right) =$$

$$= E(X_1 \cdot X_2) - E(X_1X_3) + E(X_2^2) - E(X_2X_3) =$$

$$= \cancel{\mu \cdot \mu} - \cancel{\mu \cdot \mu} + E(X_2^2) - \mu \cdot \mu = E(X_2^2) - \mu^2 \quad (*)$$

$$Var(X_2) = E\left[\left(X_2 - E(X_2)\right)^2\right] = E(X_2^2) - \left[E(X_2)\right]^2$$

$$\Rightarrow E(X_2^2) = Var(X_2) + \left[E(X_2)\right]^2$$

$$\Rightarrow E(X_2^2) = \sigma^2 + \mu^2 \quad (**)$$

Επομένως από $(*)$:

$$Cov(Y_1, Y_2) = E(X_2^2) - \mu^2 \overset{(**)}{=\!=} \left(\sigma^2 + \cancel{\mu^2}\right) - \cancel{\mu^2}$$

$$= \sigma^2 \checkmark$$

(2) Ν.δ.ο. ο συντελεστής προσδιορισμού R^2 στην παλινδρόμηση της Y επί της X είναι ίσος με το τετράγωνο της τιμής της συσχέτισης δείγματος ανάμεσα στην X και στην Y, δηλ. ότι $R^2 = r_{XY}^2$

Είναι $$R^2 = \frac{ESS}{TSS} \stackrel{\frac{SS_{Reg}}{SS_{Tot}}}{=} \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad ①$$

Όμως:

$$\stackrel{SS_{Reg}}{\parallel}ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(\overbrace{\hat{b}_0 + \hat{b}_1 X_i}^{\hat{Y}_i} - \bar{Y})^2 = \sum_{i=1}^{n}(\overbrace{\cancel{\bar{Y}} - \hat{b}_1\bar{X}}^{\hat{b}_0} + \hat{b}_1 X_i - \cancel{\bar{Y}})^2 =$$

η εκτίμηση του b_0 με LS

$$= \sum_{i=1}^{n}\left[\hat{b}_1(X_i - \bar{X})\right]^2 = \hat{b}_1^2 \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 =$$

$\hat{b}_1$

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

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

$$① \xrightarrow{②} R^2 = \dots$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{n}(y_i-\bar{y})^2} = \frac{\frac{1}{n^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\frac{1}{n^2}\left[\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{n}(y_i-\bar{y})^2}\right]^2} =$$

$$= \frac{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\left[\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \cdot \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}\right]^2} = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\frac{1}{n}(y_i-\bar{y})^2}}\right]^2 =$$

$$= \left[\frac{S_{xy}}{S_x \cdot S_y}\right]^2 = r_{xy}^2$$

Δειγματική τυπική απόκλιση μεταξύ Χ και Υ.

(Δειγματικός Συντελεστής Συσχέτισης)

Συντελεστής Pearson :

$$r(x,y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \cdot \sum(y_i-\bar{y})^2}}$$

(3) $\hat{Y} = 5.4 + 3.2X$
$\quad\quad(3.1)\quad(1.5)$

$R^2 = 0.26$ και $\hat{\sigma} = 6.2$

$n = 250$

(α) Έλεγχος της:

$H_0: b_1 = 0$

έναντι

$H_1: b_1 \neq 0$ για $\alpha = 5\%$

$$t = \frac{\hat{b}_1 - b_1}{S.e.(\hat{b}_1)} \overset{\text{υπό } H_0}{=} \frac{3.2 - 0}{1.5} = 2.133$$

$$t_c = t_{\alpha/2,\, n-k} = t_{0.025,\, 248} = 1.96$$

Επειδή $|t| > t_c$ $\Rightarrow$ η H_0 απορρίπτεται έναντι της H_1 για $\alpha = 5\%$

γιατί πρόκειται για δίπλευρο έλεγχο

(b) $\hat{b}_1^U = \hat{b}_1 + t_c \cdot S.E(\hat{b}_1) = 3.2 + 1.96 * 1.5 = 6.14$

$\hat{b}_1^L = \hat{b}_1 - t_c \cdot S.E.(\hat{b}_1) = 3.2 - 1.96 * 1.5 = 0.26$

$b_1 \in [0.26,\ 6.14]$

(*) → Παρατηρώ ότι η τιμή $b_1 = 0$ δεν βρίσκεται εντός του διαστήμ. όπως αναμενόταν!

(c) Η (σχέση) υπόθεση $b_1=0$ απορρίφθηκε (για $\alpha=5\%$) που σημαίνει ότι ∃ σχέση (αιτιακή) μεταξύ των X και Y. Επομένως, θα μας εξέπληττε το γεγονός ότι X, Y ανεξάρτητες. (σε περίπτωση που μας έλεγε κάποιος ότι X_i, Y_i ήταν ανεξάρτητες!)

(d) Υποθέστε ότι X_i και Y_i ανεξάρτητες και σχηματίζουν πολλά δείγματα μεγέθους $n=250$, εκτιμώνται παλινδρομήσεις και απαντώνται τα ερωτήματα (a) και (b). Σε ποιό % των δειγμάτων θ' απορριφθεί η H_0 στο (a)?

→ Σε ποιό ποσοστό των δειγμάτων η τιμή $b_1=0$ θα συμπεριλαμβάνεται στο διάστημα εμπιστοσύνης στο (b)?

→ Η $H_0: b_1=0$ απορρ. στο 5% των δειγμάτων με 5% επιπ. σημαντικότητας και το 95% των διαστημ. εμπιστ. θα συμπεριλαμβάνει την $b_1=0$.