ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Αναπλ. Καθ. Στέλιος Ζήμερας

Τμήμα Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2020

Μοντελοποίηση

- Έστω για την σύγκριση k δειγμάτων συλλέγουμε n_i παρατηρήσεις Y_{ij} j=1,2,…, p από το δείγμα i με i=1,2,…,k

$$
\begin{array}{l}
1^0 \text{ δειγμα } Y_{11}, Y_{12}, \ldots, Y_{1p} \\
2^0 \text{ δειγμα } Y_{21}, Y_{22}, \ldots, Y_{2p} \\
\vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots \\
k^0 \text{ δειγμα } Y_{k1}, Y_{k1}, \ldots, Y_{kp}
\end{array}
$$

$\Downarrow$ **Πειραματικό σχέδιο**

$$H_0: \mu_1 = \mu_2 = \ldots\ldots = \mu_k$$

Μοντελοποίηση

Δηλ: οι μέσες τιμές των πληθυσμών από τους οποίους λαμβάνονται τα εν λόγω δείγματα είναι ίσες μεταξύ τους

Θεωρούμε ότι $n = n_1 + n_2 + \ldots\ldots + n_k$ όπου n: είναι το μέγεθος του δείγματος.

Θα έχουμε ότι για την παρατήρηση Y_{ij} που γράφεται ως

$$Y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij}, i = 1, 2, \ldots, k; j = 1, 2, \ldots, p$$

όπου οι τ.μ. ε_{ij} είναι ανεξάρτητες και $\varepsilon_{ij} \sim N(0,\sigma^2)$ με $\sigma^2 > 0$

$$\sum_i^k a_i = 0 \qquad \sum_j^p \beta_j = 0$$

Υποθέσεις

$$H_A : a_1 = a_2 = \ldots. = a_k = 0$$

$$H_B : \beta_1 = \beta_2 = \ldots. = \beta_p = 0$$

Μοντελοποίηση

Έλεγχος της υπόθεσης H_A

Παρατηρούμε ότι αφού ισχύει $\sum_{i}^{k} a_i = 0$

τότε η H_A γράφεται ισοδύναμα

$$H_A : a_1 - a_k = 0, = a_2 - a_k = 0, \ldots . a_{k-1} - a_k = 0$$

Αν θέσουμε

$$S = (Y - X\beta)'(Y - X\beta) = \sum_{i}^{k} \sum_{j}^{p} \left(Y_{ij} - \mu - a_i - \beta_j \right)^2 =$$

$$= \sum_{j}^{p} \left(Y_{1j} - \mu - a_1 - \beta_j \right)^2 + \sum_{j}^{p} \left(Y_{2j} - \mu - a_2 - \beta_j \right)^2 + \ldots . +$$

$$+ \sum_{j}^{p} \left(Y_{kj} - \mu - a_k - \beta_j \right)^2$$

Μοντελοποίηση

Θα έχουμε

$$\frac{\partial S}{\partial \mu} = -2\sum_{i}^{k}\sum_{j}^{p}\left(Y_{ij} - \mu - a_i - \beta_j\right) = 0 \Rightarrow$$

$$\Rightarrow \sum_{i}^{k}\sum_{j}^{p} Y_{ij} - kp\mu - \cancelto{0}{\sum_{i}^{k}\sum_{j}^{p} a_i} - \cancelto{0}{\sum_{i}^{k}\sum_{j}^{p} \beta_j} = 0 \Leftrightarrow kp\mu = \sum_{i}^{k}\sum_{j}^{p} Y_{ij}$$

$$\sum_{i}^{k} a_i = 0 \qquad \sum_{j}^{p} \beta_j = 0$$

Μοντελοποίηση

Επομένως η εκτιμήτρια ελαχίστων τετραγώνων της παραμέτρου μ (αν θέσουμε kp=n) έχουμε

$$\widehat{\mu} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{p} Y_{ij} = Y_{..}$$

$$\frac{\partial S}{\partial a_1} = -2\sum_{j}^{p}\left(Y_{1j} - \mu - a_1 - \beta_j\right) = 0$$

$$\frac{\partial S}{\partial a_2} = -2\sum_{j}^{p}\left(Y_{2j} - \mu - a_2 - \beta_j\right) = 0$$

....

$$\frac{\partial S}{\partial a_k} = -2\sum_{j}^{p}\left(Y_{kj} - \mu - a_k - \beta_j\right) = 0$$

Μοντελοποίηση

Το σύστημα γράφεται ισοδύναμα αν λάβουμε υπόψη

$$\sum_{j}^{p} \beta_j = 0$$

$$\rho\left(\mu + \alpha_1\right) = \sum_{j}^{p} Y_{1j}$$

$$\rho\left(\mu + \alpha_2\right) = \sum_{j}^{p} Y_{2j}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\rho\left(\mu + \alpha_k\right) = \sum_{j}^{p} Y_{kj}$$

Μοντελοποίηση

Με βάση το παραπάνω σύστημα εξισώσεων αν οι μεταβλητές είναι εκτιμήτριες ελαχίστων τετραγώνων των μ, α_1 , α_2 ,…., α_k , τότε έχουμε

$$\left(\widehat{\mu}+\widehat{\alpha}_1\right)=\frac{1}{\rho}\sum_{j}^{p}Y_{1j}=Y_{1.}$$

$$\left(\widehat{\mu}+\widehat{\alpha}_2\right)=\frac{1}{\rho}\sum_{j}^{p}Y_{2j}=Y_{2.}$$

……………………………………………

$$\left(\widehat{\mu}+\widehat{\alpha}_k\right)=\frac{1}{\rho}\sum_{j}^{p}Y_{kj}==Y_{k.}$$

Μοντελοποίηση

$$\widehat{\mu} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{p} Y_{ij} = Y_{..}$$

$$\widehat{a}_1 = Y_{1.} - Y_{..}$$

$$\widehat{a}_2 = Y_{2.} - Y_{..}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\widehat{a}_k = Y_{k.} - Y_{..}$$

Μοντελοποίηση

$$S=\sum_{i}^{k}\left(Y_{i1}-\mu-a_i-\beta_1\right)^2+\sum_{i}^{k}\left(Y_{i2}-\mu-a_i-\beta_2\right)^2+\ldots.+$$

$$+\sum_{i}^{k}\left(Y_{ip}-\mu-a_i-\beta_p\right)^2$$

Θα έχουμε

$$\frac{\partial S}{\partial \beta_1}=-2\sum_{i}^{k}\left(Y_{i1}-\mu-a_i-\beta_1\right)=0$$

$$\frac{\partial S}{\partial \beta_2}=-2\sum_{i}^{k}\left(Y_{i2}-\mu-a_i-\beta_2\right)=0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial S}{\partial \beta_p}=-2\sum_{i}^{k}\left(Y_{ip}-\mu-a_i-\beta_p\right)=0$$

Μοντελοποίηση

- Το σύστημα γράφεται ισοδύναμα αν λάβουμε υπόψη

$$\sum_{i}^{k} a_i = 0$$

$$k\left(\mu + \beta_1\right) = \sum_{i}^{k} Y_{i1}$$

$$k\left(\mu + \beta_2\right) = \sum_{i}^{k} Y_{i2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$k\left(\mu + \beta_p\right) = \sum_{i}^{k} Y_{ip}$$

Μοντελοποίηση

Με βάση το παραπάνω σύστημα εξισώσεων αν οι μεταβλητές είναι εκτιμήτριες ελαχίστων τετραγώνων των μ, β_1 , β_2 ,…., β_p , τότε έχουμε

$$\left(\widehat{\mu}+\widehat{\beta}_1\right)=\frac{1}{k}\sum_i^k Y_{i1}=Y_{.1}$$

$$\left(\widehat{\mu}+\widehat{\beta}_2\right)=\frac{1}{k}\sum_i^k Y_{i2}=Y_{.2}$$

……………………………………………

$$\left(\widehat{\mu}+\widehat{\beta}_p\right)=\frac{1}{k}\sum_i^k Y_{ip}=Y_{.p}$$

Μοντελοποίηση

$$\widehat{\mu} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{p} Y_{ij} = Y_{..}$$

$$\widehat{\beta}_1 = Y_{.1} - Y_{..}$$

$$\widehat{\beta}_2 = Y_{.2} - Y_{..}$$

....

$$\widehat{\beta}_p = Y_{.p} - Y_{..}$$

Μοντελοποίηση

Το άθροισμα τετραγώνων των αποκλίσεων

$$R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - \widehat{\mu} - \widehat{\alpha}_i - \beta_j\right)^2 =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{..} - \left(Y_{i.} - Y_{..}\right) - \left(Y_{.j} - Y_{..}\right)\right]^2 =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{i.} - Y_{.j} + Y_{..}\right]^2$$

Μοντελοποίηση

- Κάτω από την υπόθεση $H_A : \alpha_i = 0$, i=1,2,…,k έχουμε να ελαχιστοποιήσουμε το άθροισμα

$$S_A = \sum_i^k \sum_j^p \left(Y_{ij} - \mu - \beta_j\right)^2 = \sum_i^k \left(Y_{i1} - \mu - \beta_1\right)^2 +$$

$$+\sum_i^k \left(Y_{i2} - \mu - \beta_2\right)^2 + \dots + \sum_i^k \left(Y_{ip} - \mu - \beta_p\right)^2$$

Μοντελοποίηση

Άρα

$$\frac{\partial S_A}{\partial \mu} = -2\sum_{i}^{k}\sum_{j}^{p}\left(Y_{ij} - \mu - \beta_j\right) = 0 \Rightarrow$$

$$\Rightarrow \sum_{i}^{k}\sum_{j}^{p} Y_{ij} - kp\mu = 0$$

Αφού $\sum_{j}^{p} \beta_j = 0$

Επομένως η εκτιμήτρια ελαχίστων τετραγώνων της παραμέτρου μ (αν θέσουμε kp=n) έχουμε

$$\widehat{\mu} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{p} Y_{ij} = Y_{..} = \tilde{\mu}$$

Μοντελοποίηση

Θα έχουμε

$$\frac{\partial S_A}{\partial \beta_1} = -2\sum_{i}^{k}\left(Y_{i1} - \mu - \beta_1\right) = 0$$

$$\frac{\partial S_A}{\partial \beta_2} = -2\sum_{i}^{k}\left(Y_{i2} - \mu - \beta_2\right) = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial S_A}{\partial \beta_p} = -2\sum_{i}^{k}\left(Y_{ip} - \mu - \beta_p\right) = 0$$

Μοντελοποίηση

- Το σύστημα γράφεται ισοδύναμα αν λάβουμε υπόψη

$$k\left(\mu+\beta_1\right)=\sum_i^k Y_{i1}$$

$$k\left(\mu+\beta_2\right)=\sum_i^k Y_{i2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$k\left(\mu+\beta_p\right)=\sum_i^k Y_{ip}$$

Μοντελοποίηση

Με βάση το παραπάνω σύστημα εξισώσεων αν οι μεταβλητές είναι εκτιμήτριες ελαχίστων τετραγώνων των μ, β_1 , β_2 ,…., β_p , τότε έχουμε

$$\left(\tilde{\mu} + \tilde{\beta}_1\right) = \frac{1}{k}\sum_{i}^{k} Y_{i1} = Y_{.1}$$

$$\left(\tilde{\mu} + \tilde{\beta}_2\right) = \frac{1}{k}\sum_{i}^{k} Y_{i2} = Y_{.2}$$

………………………………………………

$$\left(\tilde{\mu} + \tilde{\beta}_p\right) = \frac{1}{k}\sum_{i}^{k} Y_{ip} = Y_{.p}$$

Μοντελοποίηση

$$\widehat{\mu} = \frac{1}{n}\sum_{i}^{k}\sum_{j}^{p} Y_{ij} = Y_{..} = \tilde{\mu}$$

$$\tilde{\beta}_1 = Y_{.1} - Y_{..}$$

$$\tilde{\beta}_2 = Y_{.2} - Y_{..}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\tilde{\beta}_p = Y_{.p} - Y_{..}$$

Μοντελοποίηση

Παρατηρούμε ότι ισχύει

$$\tilde{\beta}_1 = \widehat{\beta}_1$$

$$\tilde{\beta}_2 = \widehat{\beta}_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\tilde{\beta}_p = \widehat{\beta}_p$$

με βάση

$$\tilde{a}_1 = 0$$

$$\tilde{a}_2 = 0$$

$$\cdots\cdots\cdots$$

$$\tilde{a}_k = 0$$

Μοντελοποίηση

- Η ελάχιστη τιμή του S_A θα είναι

$$R_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - \widehat{\mu} - \widehat{\beta}_j\right)^2 =$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{..} - Y_{.j} + Y_{..}\right]^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{.j}\right]^2$$

Μοντελοποίηση

- Το άθροισμα τετραγώνων που οφείλεται στην απόκλιση από την H_A θα είναι

$$R_1^2 - R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - Y_{.j}\right)^2 - \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{i.} - Y_{.j} + Y_{..}\right]^2 \Rightarrow$$

$$\Rightarrow R_1^2 - R_0^2 = \rho\sum_{i}\widehat{a}_i^2 = \rho\sum_{i=1}^{k}\left[Y_{i.} - Y_{..}\right]^2$$

Μοντελοποίηση

- Η στατιστική συνάρτηση F για τον έλεγχο της H_A θα είναι

$$F_A = \frac{\rho \sum_{i=1}^{k} \left[Y_{i.} - Y_{..} \right]^2 \Big/ (k-1)}{R_0^2 \Big/ (k-1)(\rho - 1)}$$