ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Αναπλ. Καθ. Στέλιος Ζήμερας

Τμήμα Στατιστικής και Αναλογιστικών – Χρηματοοικονομικών Μαθηματικών

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

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

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 +……+ n_k όπου n: είναι το μέγεθος του δείγματος.

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

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

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

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

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

Μοντελοποίηση

Έλεγχος της υπόθεσης H_B

Παρατηρούμε ότι αφού ισχύει $\sum_{j}^{p} \beta_i = 0$

τότε η H_B γράφεται ισοδύναμα

$$H_B : \beta_1 - \beta_p = 0, = \beta_2 - \beta_p = 0, \ldots . \beta_{p-1} - \beta_p = 0$$

Αν θέσουμε

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

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

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

Μοντελοποίηση

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

$$\widehat{a}_1 = Y_{1.} - Y_{..}, \widehat{a}_2 = Y_{2.} - Y_{..}, \ldots, \widehat{a}_k = Y_{k.} - Y_{..}$$

$$\widehat{\beta}_1 = Y_{.1} - Y_{..}, \widehat{\beta}_2 = Y_{.2} - Y_{..}, \ldots, \widehat{\beta}_p = Y_{.p} - Y_{..}$$

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 $H_B: \beta_j = 0$, $j=1,2,\ldots,p$ έχουμε να ελαχιστοποιήσουμε το άθροισμα

$$S_B = \sum_{i}^{k}\sum_{j}^{p}\left(Y_{ij} - \mu - a_i\right)^2 = \sum_{j}^{p}\left(Y_{1j} - \mu - a_1\right)^2 +$$

$$+\sum_{j}^{p}\left(Y_{2j} - \mu - a_2\right)^2 + \ldots + \sum_{j}^{p}\left(Y_{kj} - \mu - a_k\right)^2$$

Μοντελοποίηση

Άρα

$$\frac{\partial S_B}{\partial \mu} = -2\sum_i^k \sum_j^p \left(Y_{ij} - \mu - a_i\right) = 0 \Rightarrow$$

$$\Rightarrow \sum_i^k \sum_j^p Y_{ij} - kp\mu = 0$$

Αφού $\sum_i^k a_i = 0$

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

$$\widehat{\mu} = \frac{1}{n}\sum_i^k \sum_j^p Y_{ij} = Y_{..} = \tilde{\tilde{\mu}}$$

Μοντελοποίηση

Θα έχουμε

$$\frac{\partial S_B}{\partial a_1} = -2\sum_i^k \left(Y_{1j} - \mu - a_1\right) = 0$$

$$\frac{\partial S_B}{\partial a_2} = -2\sum_i^k \left(Y_{2j} - \mu - a_2\right) = 0$$

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

$$\frac{\partial S_B}{\partial a_k} = -2\sum_i^k \left(Y_{kp} - \mu - a_k\right) = 0$$

Μοντελοποίηση

- Το σύστημα γράφεται ισοδύναμα αν λάβουμε υπόψη

$$p\left(\mu + a_1\right) = \sum_{j}^{p} Y_{1j}$$

$$p\left(\mu + a_2\right) = \sum_{j}^{p} Y_{2j}$$

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

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

Μοντελοποίηση

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

$$\left(\tilde{\tilde{\mu}}+\tilde{\tilde{a}}_1\right)=\frac{1}{p}\sum_j^p Y_{1j}=Y_{1.}$$

$$\left(\tilde{\tilde{\mu}}+\tilde{\tilde{a}}_2\right)=\frac{1}{p}\sum_j^p Y_{2j}=Y_{2.}$$

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

$$\left(\tilde{\tilde{\mu}}+\tilde{\tilde{a}}_k\right)=\frac{1}{p}\sum_j^p Y_{kj}=Y_{k.}$$

Μοντελοποίηση

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

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

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

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

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

Μοντελοποίηση

Παρατηρούμε ότι ισχύει

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

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

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

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

με βάση

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

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

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

$$\tilde{\tilde{\beta}}_p = 0$$

Μοντελοποίηση

- Η ελάχιστη τιμή του S_B θα είναι

$$\tilde{R}_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - \widehat{\mu} - \widehat{a}_i\right)^2 =$$

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

Μοντελοποίηση

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

$$\tilde{R}_1^2 - R_0^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - Y_{i.}\right)^2 - \sum_{i=1}^{k}\sum_{j=1}^{p}\left[Y_{ij} - Y_{i.} - Y_{.j} + Y_{..}\right]^2 \Rightarrow$$

$$\Rightarrow \tilde{R}_1^2 - R_0^2 = k\sum_{j}^{p}\widehat{\beta}_j^2 = k\sum_{j=1}^{p}\left[Y_{.j} - Y_{..}\right]^2$$

Μοντελοποίηση

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

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

Μοντελοποίηση

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο άθροισμα τετραγώνων	F
Απόκλιση από την H_A (μεταξύ δειγμάτων)	$R_1^2 - R_0^2$	k-1	$\rho\sum_{i=1}^{k}\left[Y_{i.} - Y_{..}\right]^2 \Big/ (k-1)$	$F_A = \dfrac{\rho\sum_{i=1}^{k}\left[Y_{i.} - Y_{..}\right]^2 \Big/ (k-1)}{R_0^2 \Big/ (k-1)(\rho-1)}$
Απόκλιση από την H_B (μεταξύ δειγμάτων)	$\tilde{R}_1^2 - R_0^2$	p-1	$k\sum_{j=1}^{p}\left[Y_{.j} - Y_{..}\right]^2 \Big/ (p-1)$	$F_B = \dfrac{k\sum_{j=1}^{p}\left[Y_{.j} - Y_{..}\right]^2 \Big/ (p-1)}{R_0^2 \Big/ (k-1)(\rho-1)}$
Κατάλοιπα (εντός δειγμάτων)	R_0^2	(k-1)(p-1)	$R_0^2 \Big/ (k-1)(p-1)$	
Σύνολο	$\sum_{i=1}^{k}\sum_{j=1}^{p}\left(Y_{ij} - Y_{..}\right)^2$	n-1		

Απλοποιήσεις

- Το ολικό άθροισμα τετραγώνων υπολογίζεται από

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

- Στην περίπτωση H_A έχουμε

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

$$=\rho\sum_{i}^{k}Y_{i.}^{\;2}-\rho\frac{1}{k}k^2Y_{..}^{\;2}=\rho\sum_{i}^{k}Y_{i.}^{\;2}-\rho kY_{..}^{\;2}=\rho\sum_{i}^{k}Y_{i.}^{\;2}-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2$$

$$\boxed{R_1^2-R_0^2=\rho\sum_{i}^{k}Y_{i.}^{\;2}-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2}$$

Απλοποιήσεις

- Το ολικό άθροισμα τετραγώνων υπολογίζεται από

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

- Στην περίπτωση H_B έχουμε

$$\tilde{R}_1^2-R_0^2=k\sum_{j}^{p}\widehat{\beta}_j^2=k\sum_{j}^{p}\left[Y_{.j}-Y_{..}\right]^2=k\left[\sum_{j}^{p}Y_{.j}^2-\frac{1}{p}\left(\sum_{j}^{p}Y_{.j}\right)^2\right]=$$

$$=k\sum_{j}^{p}Y_{.j}^2-k\frac{1}{p}p^2Y_{..}^{\;2}=k\sum_{j}^{p}Y_{.j}^2-kpY_{..}^{\;2}=k\sum_{j}^{p}Y_{.j}^2-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2$$

$$\boxed{\tilde{R}_1^2-R_0^2=k\sum_{j}^{p}Y_{.j}^2-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2}$$

Απλοποιήσεις

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ ANOVA ΜΕ 2 ΠΑΡΑΓΟΝΤΕΣ

B A	B_1	B_2	....	B_ρ	Άθροισμα γραμμών	Μέσοι γραμμών	Τετρ. γραμμών	Άθροισμα Τετραγ. γραμμών
A_1	Y_{11}	Y_{12}	....	$Y_{1\rho}$	$\sum_j^p Y_{1j}$	$Y_{1.}$	$Y_{1.}^2$	$\sum_j^p Y_{1j}^2$
A_2	Y_{21}	Y_{22}	....	$Y_{2\rho}$	$\sum_j^p Y_{2j}$	$Y_{2.}$	$Y_{2.}^2$	$\sum_j^p Y_{2j}^2$
⋮	⋮	⋮	....	⋮	⋮	⋮	⋮	⋮
A_κ	$Y_{\kappa 1}$	$Y_{\kappa 2}$	....	$Y_{\kappa\rho}$	$\sum_j^p Y_{kj}$	$Y_{k.}$	$Y_{k.}^2$	$\sum_j^p Y_{kj}^2$
Άθροισμα Στηλών	$\sum_i^k Y_{i1}$	$\sum_i^k Y_{i2}$	....	$\sum_i^k Y_{ip}$	$\sum_i^k \sum_j^p Y_{ij}$	$\sum_i^k Y_{i.}$	$\sum_i^k Y_{i.}^2$	
Μέσοι Στηλών	$Y_{.1}$	$Y_{.2}$	....	$Y_{.p}$	$\sum_j^p Y_{.j}$	$Y_{..}$		
Τετρ. Στηλών	$Y_{.1}^2$	$Y_{.2}^2$	....	$Y_{.p}^2$	$\rightarrow \sum_j^p Y_{.j}^2$			
Άθροισμα Τετραγ. Στηλών	$\sum_i^k Y_{i1}^2$	$\sum_i^k Y_{i2}^2$	....	$\sum_i^k Y_{ip}^2$				$\sum_i^k \sum_j^p Y_{ij}^2$

Παράδειγμα

ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ ANOVA ΜΕ 2 ΠΑΡΑΓΟΝΤΕΣ

Ειδος τροφης ομάδα	B_1	B_2	B_3	Άθροισμα γραμμών	Μέσοι γραμμών	Τετρ. γραμμών	Άθροισμα Τετραγ. γραμμών
A_1	7	14	8.5	29.5	9.83	96.62	317.25
A_2	16	15	16.5	48	16	256	768.5
A_3	10.5	15	9.5	35	11.67	136.18	425.5
A_4	13.5	21	13.5	48	16	256	805.5
Άθροισμα Στηλών	47	65.5	48	160.5	53.5	744.81	
Μέσοι Στηλών	11.75	16.36	12	40.11	13.38		
Τετρ. Στηλών	138.06	267.64	144	549.71			
Άθροισμα Τετραγ. Στηλών	597.5	1102.25	617				2316.75

Παράδειγμα

$$\sum_{i=1}^{4}\sum_{j=1}^{3}\left(Y_{ij}-Y_{..}\right)^2=\sum_{i=1}^{4}\sum_{j=1}^{3}Y_{ij}^{\ 2}-\frac{1}{n}\left(\sum_{i=1}^{4}\sum_{j=1}^{3}Y_{ij}\right)^2=2316.75-\frac{1}{12}\left(160.5\right)^2=170.06$$

$$R_1^2-R_0^2=\rho\sum_{i}^{k}Y_{i.}^2-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2=3\sum_{i}^{4}Y_{i.}^2-\frac{1}{12}\left(\sum_{i}^{4}\sum_{j}^{3}Y_{ij}\right)^2=....=87.76$$

$$\tilde{R}_1^2-R_0^2=k\sum_{j}^{p}Y_{.j}^2-\frac{1}{n}\left(\sum_{i}^{k}\sum_{j}^{p}Y_{ij}\right)^2=4\sum_{j}^{3}Y_{.j}^2-\frac{1}{12}\left(\sum_{i}^{4}\sum_{j}^{3}Y_{ij}\right)^2=...=30.13$$

Παράδειγμα

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο άθροισμα τετραγώνων	F
Απόκλιση από την H_A (μεταξύ δειγμάτων)	87.76	3	29.25	F_A=5.82
Απόκλιση από την H_B (μεταξύ δειγμάτων)	52.16	2	26.08	F_B=5.19
Κατάλοιπα (εντός δειγμάτων)	30.13	6	5.022	
Σύνολο	170.06	11		

F_A=5.82>4.76 άρα απορρίπτουμε την H_A

F_B=5.19>5.14 άρα απορρίπτουμε την H_B