17ο online μάθημα

21/5/2020

ΕΝΑΛΛΑΚΤΙΚΗ ΙΣΟΔΥΝΑΜΗ ΜΟΡΦΗ

$$y_{ij} = \mu + a_i + \varepsilon_{ij} \quad , \quad i = 1 \dots I, \quad j = 1 \dots n_i$$

Τι αλλάζει ;

μ: κοινή επίδραση στην Y όλων των επιπέδων
a_i: ατομική επίδραση στην Y του i-οστού επιπέδου

Προφανώς $\mu, a_1, \ldots, a_I$ άγνωστες παράμετροι. Πώς θα βρούμε τους εκτιμητές τους;
Με τη μέθοδο ελαχίστων τετραγώνων που έγκειται στην ελαχιστοποίηση ως προς $\mu, a_1, \ldots, a_I$ της

$$S(\mu, a_1, \ldots, a_I) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \mu - a_i)$$

$$S(\mu, a_1, \ldots, a_I) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - (\mu + a_i))^2 =$$

$$= \sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij}^2 - 2 \sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij} (\mu + a_i) + \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\mu + a_i)^2 =$$

$$= \sum_i \sum_j Y_{ij}^2 - 2\mu \sum_i \sum_j Y_{ij} - 2 \sum_i \left[a_i \sum_j Y_{ij} \right] +$$

$$+ \sum_i \sum_j \mu^2 + 2 \sum_i \sum_j \mu a_i + \sum_i \sum_j a_i^2$$

Άρα

$$S(\mu, a_1, \ldots, a_I) = \sum_i \sum_j Y_{ij}^2 - 2\mu Y_{..} - 2 \sum_i a_i Y_{i.} + N\mu^2 + 2\mu \sum_i n_i a_i +$$

$$+ \sum_i n_i a_i^2$$

ή

$$S(\mu, a_1, \ldots, a_I) = \sum_i \sum_j Y_{ij}^2 - 2\mu Y_{..} - 2(a_1 Y_{1.} + \ldots + a_I Y_{I.}) + N\mu^2$$

$$+ 2\mu (n_1 a_1 + \ldots + n_I a_I) + (n_1 a_1^2 + \ldots + n_I a_I^2)$$

Επομένως:

$$\frac{\partial S(\mu, a_1, \ldots, a_I)}{\partial \mu} = -2Y_{..} + 2\mu N + 2 \sum_{i=1}^{I} n_i a_i$$

και $\frac{\partial S}{\partial a_i}(\mu, a_1, \dots, a_I) = -2Y_{i.} + 2\mu n_i + 2n_i a_i$

Επομένως προσδιορίζονται από το σύστημα των εξισώσεων:

$$Y_{..} = N\mu + \sum_{i=1}^{I} a_i n_i \quad \text{και}$$

$$Y_{i.} = n_i(\mu + a_i) \quad i = 1, \dots I$$

ή

$$Y_{i.} = n_i\mu + n_i a_i \quad i = 1 \dots I$$

Παρατηρώ το εξής

$$\begin{cases} Y_{1.} = n_1\mu + n_1 a_1 \\ Y_{2.} = n_2\mu + n_2 a_2 \\ \vdots \\ Y_{I} = n_I\mu + n_I a_I \end{cases}$$

Αθροίζω: $Y_{..} = \mu \sum_{i=1}^{I} n_i + \sum_{i=1}^{I} n_i a_i$

ή $Y_{..} = \mu N + \sum_{i=1}^{I} n_i a_i$

Επομένως το σύστημα των κανονικών εξισώσεων δεν έχει μοναδική λύση!

Εύρεση μοναδικής λύσης; Πώς;

Υιοθέτηση υποθέσεων γνωστών ως πλευρικών συνθηκών

π.χ. $\mu = 0 \longrightarrow$ πάω στο μοντέλο $Y_{ij} = \mu_i + \varepsilon_{ij}$

ή $\sum_{i=1}^{I} a_i n_i = 0$

Υπό την 2η υπόθεση ότι $\sum_{i=1}^{I} a_i n_i = 0$

έχουμε ότι:

$$Y_{..} = N\hat{\mu} \Rightarrow \hat{\mu} = \bar{Y}_{..}$$

και τότε

$$Y_{i\cdot} = n_i\hat{\mu} + n_i\hat{a}_i \Rightarrow \frac{Y_{i\cdot}}{n_i} - \frac{n_i\hat{\mu}}{n_i} = \hat{a}_i \rightarrow$$

$$\hat{a}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} \quad , i = 1 \dots I$$

Προσοχή: Πώς προέκυψε η πλευρική συνθήκη; Θυμηθείτε

$$Y_{\cdot\cdot} = N\mu + \sum_{i=1}^{I} a_i n_i$$

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

Ολική μεταβλητότητα

$$SS_{tot} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

άθροισμα τετραγώνων των διαφορών κάθε παρατήρησης από το γενικό μέσο.

Επίσης είδαμε ότι:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad i = 1 \dots I$$
$$j = 1 \dots n_i$$

με $\hat{\mu}_i = \bar{Y}_i$

Ξεκινώντας από τη σχέση:

$$Y_{ij} - \bar{Y}_{\cdot\cdot} = Y_{ij} - \bar{Y}_{i\cdot} + \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$$

Θα δείξουμε ότι:

$$\sum_i\sum_j(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = \sum_i\sum_j(Y_{ij} - \bar{Y}_{i\cdot})^2 + \sum_i\sum_j(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

Αρκεί ν.δ.ο.

$$\sum_i\sum_j(Y_{ij} - \bar{Y}_{i\cdot})(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = 0$$

Είναι

$$\sum_i\sum_j(Y_{ij} - \bar{Y}_{i\cdot})(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = \sum_i\left[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})\right]$$

Όμως $\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.}) = \sum_{j=1}^{n_i} Y_{ij} - n_i \bar{Y}_{i.} = \sum_{j=1}^{n_i} Y_{ij} - n_i \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = 0$

που αποδεικνύει το ζητούμενο.

Άρα το SS_{tot} "διασπάται" σε δύο μέρη:

$\sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 =$ άθροισμα τετραγώνων των υπολοίπων

αφού $\hat{Y}_{ij} = \hat{\mu}_i = \bar{Y}_{i.}$

Δηλαδή $\sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 = SS_{res}$

Το 2ο μέρος είναι:

$$\sum_{i=1}^{I} \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 \underset{j}{\overset{\text{δεν υπάρχει}}{=\!=\!=}} \sum_{i=1}^{I} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 =$$ άθροισμα τετραγώνων αποκλίσεων δειγματικών μέσων από το γενικό δειγματικό μέσο

$$\| \quad SS_{tr}$$

Άρα $SS_{tot} = SS_{tr} + SS_{res}$

Προφανώς πιο εύκολα στον υπολογισμό τα SS_{tot} και SS_{tr}

Άρα $SS_{res} = SS_{tot} - SS_{tr}$

ΒΑΘΜΟΙ ΕΛΕΥΘΕΡΙΑΣ

$$SS_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$$

οι β.ε. είναι $N-1$

διότι
- έχει εκτιμηθεί μια παράμετρος και έχω χρησιμοποιήσει N τιμές Y_{ij} $i=1,\dots I$, $j=1\dots n_i$
- ή ισχύει ότι $\sum_i \sum_j (Y_{ij} - \bar{Y}_{..}) = 0$ άρα χρειάζονται $n_1 + \dots + n_i - 1$ σχέσεις

$$SS_{res} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})$$

οι β.ε. είναι $N-1$

Πάλι δύο τρόποι σκέψης ; Ποιοι ;

Ποιοι είναι οι β.ε. του SS_{tr} ;

$$\underbrace{\text{β.ε. } SS_{tot}}_{N-1} = \underbrace{\text{β.ε. } SS_{res}}_{N-I} + \underbrace{\text{β.ε. } SS_{tr}}_{x}$$

άρα $x = N-1-N+I = I-1$

- Ορίζουμε $MS_{tr} = \dfrac{SS_{tr}}{\text{β.ε. } SS_{tr}} = \dfrac{SS_{tr}}{I-1}$

$$MS_{res} = \frac{SS_{res}}{\text{β.ε. } SS_{res}} = \frac{SS_{res}}{N-I}$$

Υπό τις υποθέσεις για τα σφάλματα αποδεικνύεται ότι:

$$E(MS_{res}) = \sigma^2$$

Επιπλέον υπό την θεώρηση του μοντέλου

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}$$

και της πλευρικής συνθήκης

$$\sum_{i=1}^{I} a_i n_i = 0$$

έχουμε ότι

$$E(MS_{tr}) = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^{I} n_i a_i^2$$

Έχουμε το μοντέλο $Y_{ij} = \mu + a_i + \varepsilon_{ij}$ $\quad i = 1 \dots I$

$j = 1 \dots n_i$

ΠΛΕΥΡΙΚΗ ΣΥΝΘΗΚΗ $\sum_{i=1}^{I} a_i n_i = 0$

Έλεγχος της

$$H_0: a_1 = a_2 = \dots = a_I = 0$$

είναι εννοιολογικά ίδιας με τον

H_0' : $\alpha_1 = \dots = \alpha_I$

γιατί αν

$\alpha_1 = \dots = \alpha_I = \alpha$

τότε $Y_{ij} = \mu + \alpha + \varepsilon_{ij}$ ή $Y_{ij} = \mu^* + \varepsilon_{ij}$

που είναι ισοδύναμο με

$\alpha_1 = \dots = \alpha_I = 0$

Έχουμε δει ότι

$E(MS_{res}) = \sigma^2$

$E(MS_{tr}) = \sigma^2 + \dfrac{1}{I-1} \sum_{i=1}^{I} n_i \alpha_i^2$

Επομένως αν $\alpha_i = 0$ $i = 1 \dots I$ ένα τεστ θα βασίζεται στη σύγκριση των MS_{res}, MS_{tr}

Υπό την H_0:

$$\left.\begin{aligned} \frac{SS_{res}}{\sigma^2} &\sim \chi^2_{N-I} \\ \frac{SS_{tr}}{\sigma^2} &\sim \chi^2_{I-1} \end{aligned}\right\} \text{ανεξάρτητες}$$

Άρα $F = \dfrac{MS_{res}}{MS_{tr}} \overset{H_0}{\sim} F_{N-I, I-1}$

Περιοχή απόρριψης $F \geq F_{\alpha, N-I, I-1}$