Ανάλυση της Διακύμανσης κατά ένα Παράγοντα

Δεδομένα:

$$
\begin{array}{lcccl}
 & Y_{11} & \ldots & Y_{1n_1} & \sim N(\mu_1, \sigma^2) \\
 & Y_{21} & \ldots & Y_{2n_2} & \sim N(\mu_2, \sigma^2) \\
 & \vdots & & & \\
i \text{ επίπεδο} & Y_{i1} & \ldots & Y_{in_i} & \sim N(\mu_i, \sigma^2) \\
 & \vdots & & & \\
 & Y_{I1} & \ldots & Y_{In_I} & \sim N(\mu_I, \sigma^2)
\end{array}
$$

$H_0:\ \mu_1 = \mu_2 = \ldots = \mu_I$ προς $H_1: \mu_i \neq \mu_j$ για τουλάχιστον ένα ζεύγος (i, j).

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

Πηγή Μετα-βλητότητας	Αθροίσματα Τετραγώνων	β.ε	Μέσα Τετράγωνα	F ελεγχο-συνάρτηση
Παράγοντας	$SS_{tr} = SS_{bet} = \sum_{i=1}^{I} n_i (\overline{Y}_{i.} - \overline{Y}_{..})^2$	$I-1$	$MS_{bet} = \frac{SS_{bet}}{I-1}$	
Υπόλοιπα	$SS_{res} = SS_{with} = \sum_{i=1}^{I}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2$	$n-I$	$MS_{with} = \frac{SS_{with}}{n-I}$	$F = \frac{MS_{bet}}{MS_{with}}$
Ολικό	$SS_{tot} = \sum_{i=1}^{I}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{..})^2$			

Παρατήρηση:

Αφού $Y_{ij} \sim N(\mu, \sigma^2)$ έχουμε ότι $Y_{ij} = \mu_i + \varepsilon_{ij}$ όπου τα ε_{ij} τυχαίο δέιγμα από $N(0, \sigma^2)$. Άρα, $\hat{\mu}_i = \overline{Y}_{i.}$ (εκτιμήτριες).

Ανάλογα αποτελέσματα έχω με τη μέθοδο ελαχίστων τετραγώνων:
Αναζητώ εκείνα τα μ_i τα οποία ελαχιστοποιούν

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i} \varepsilon_{ij}^{\,2}$$

Παραγωγίζοντας έχω ότι $\hat{\mu}_i = \overline{Y}_{i.}$.

Οι εκτιμώμενες τιμές είναι $\hat{Y}_{ij} = \hat{\mu}_i = \overline{Y}_{i.}$.

Άρα τα υπόλοιπα είναι $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i.}$

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

Θα χρησιμοποιήσουμε τη μέθοδο του επιπλέον αθροίσματος τετραγώνων για να κατασκευάσουμε την F ελεγχοσυνάρτηση. Η βασική ιδέα είναι να μετασχηματίσουμε το μοντέλο ΑΝΑΔΙΑ σε μοντέλο παλινδρόμησης.

Γράφουμε $Y_{ij} = \mu_i + \varepsilon_{ij} \equiv \mu + \alpha_i + \varepsilon_{ij}$ (1).

Θέτουμε περιορισμούς στην (1) έτσι ώστε τα μ και α_i να είναι μονοσήμαντα ορισμένα.

Συνήθως χρησιμοποιούμε $\sum n_i \alpha_i = 0$ αλλά γενικότερα μπορούμε να χρησιμοποιήσουμε $\sum w_i \alpha_i = 0$ όπου $\{w_i\}_{i=1}^{I}$ είναι συντελεστές της επιλογής μας.

Αφού $\mu_i = \mu + \alpha_i \Rightarrow n_i \mu_i = n_i \mu + n_i \alpha_i \Rightarrow \sum_{i=1}^{I} n_i \mu_i = \sum_{i=1}^{I} n_i \mu \Rightarrow \mu = \dfrac{\sum_{i=1}^{I} n_i \mu_i}{n}$

και συνεπώς $\alpha_i = \mu_i - \mu$.

Άρα οι παράμετροι μ και α_i έχουν ορισθεί μονοσήμαντα.

(1)

$$
\begin{array}{ccccccc}
 & & & \mathrm{X} & \underline{\beta} & & \underline{\varepsilon} \\
\underline{Y} & = & \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline \vdots \\ \hline Y_{I1} \\ \vdots \\ Y_{In_I} \end{pmatrix} & = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ 1 & 1 & 0 & \dots & 0 \\ \hline & & & & \\ \hline 1 & 0 & 0 & \dots & 1 \\ \vdots & & & & \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} & \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{pmatrix} & + & \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \hline \vdots \\ \hline \varepsilon_{I1} \\ \vdots \\ \varepsilon_{In_I} \end{pmatrix} \\
 & & n \times 1 & n \times (I+1) & (I+1) \times 1 & & n \times 1
\end{array}
$$

Δηλαδή έχω $\underline{Y} = \mathrm{X}\underline{\beta} + \underline{\varepsilon}$

και μπορώ να εφαρμόσω τη μέθοδο του επιπλέον αθροίσματος τετραγώνων.

$$F = \frac{\{SS_{res}(H_0 M) - SS_{res}(\Pi M)\} / \beta.\varepsilon.}{SS_{res}(\Pi M) / \beta.\varepsilon.}.$$

Άρα $H_0 : \alpha_i = 0, \forall i$.

Υπολογισμός του $SS_{res}(\Pi M)$:

Έχω ότι $SS_{res}(\Pi M) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2 = SS_{with}$ με $n - I$ β.ε.

Υπολογισμός του $SS_{res}(H_0 M)$:

Έχω ότι $SS_{res}(H_0 M) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\hat{Y}}_{ij})$, όπου $\hat{\hat{Y}}_{ij}$ οι εκτιμήσεις όταν ισχύει η H_0.

Αλλά $\hat{\hat{Y}}_{ij} = \hat{\hat{\mu}} = \overline{Y}_{..}$.

Άρα $SS_{res}(H_0 M) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{..})^2 = SS_{tot}$.

Άρα $SS_{res}(H_0 M) - SS_{res}(\Pi M) = SS_{tot} - SS_{with} = SS_{bet}$ με $(I-1)$ β.ε.

Συνεπώς $F = \dfrac{MS_{bet}}{MS_{with}}$.

Εκτιμήτριες ελαχίστων τετραγώνων:

Παραγωγίζοντας $\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \mu - \alpha_i)^2$ έχω ότι:

$$\frac{\partial}{\partial\mu} = -2\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \mu - \alpha_i) = 0 \Rightarrow \hat{\mu} = \overline{Y}_{..} .$$

$$\frac{\partial}{\partial\alpha_i} = -2\sum_{j=1}^{n_i}(Y_{ij} - \mu - \alpha_i) = 0 \Rightarrow \overline{Y}_{i.} - n_i\mu - n_i\alpha_i = 0 \Rightarrow \hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_{..} .$$

<u>Παρατηρήσεις</u>

1. Η συνθήκη $\sum \alpha_i n_i = 0$ διευκολύνει τους υπολογισμούς. Αν χρησιμοποιήσω $\sum \alpha_i w_i = 0$ τότε δεν έχω τις ίδιες εκτιμήτριες αλλά το F-test παραμένει το ίδιο.
2. $Y_{ij} \quad = \quad \mu \quad + \quad \alpha_i \quad + \quad \varepsilon_{ij}$.

♬	♬	♬
απόδοση	συνολική μέση τιμή	επίπεδο ή επίδραση
(response)	(grand mean)	(effect or level)