Ανάλυση της Διακύμανσης Κατά Δύο Παράγοντες

Γενικά έχω δύο παράγοντες Α και Β που ενδέχεται να επηρεάζουν μια εξαρτημένη μεταβλητή Y.

π.χ. Y=βαθμός μαθήματος

Α=διδασκαλία: μέθοδος #1, μέθοδος #2, μέθοδος #3 } 3 επίπεδα

Β=τύπος σχολείου: δημόσια, ιδιωτικά } 2 επίπεδα

Γενικά έχω I επίπεδα για τον Α και J για τον Β.

Συνήθως θέλω να ελέγξω τα παρακάτω:

H_A : Ο παράγοντας Α δεν επηρεάζει.

H_B : Ο παράγοντας Β δεν επηρεάζει.

H_{AB} : Υπάρχει αλληλεπίδραση παραγόντων.

Γενικά:

$$\begin{array}{cc|cccccc} & & & & \text{B} & & & \\ & & 1 & 2 & \dots & j & \dots & J \\ \hline & 1 & & & & & & \\ & 2 & & & & & & \\ & \vdots & & & & & & \\ \text{A} & i & & & & \begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijn_{ij}} \end{pmatrix} \leftarrow (i\,,j) \text{ κυψελίδα (cell)} & & \\ & \vdots & & & & & & \\ & I & & & & & & \end{array}$$

Μοντέλα:

$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$, $i = 1,...,I$, $j = 1,\dots,J$, $k = 1,\dots,n_{ij}$.

Το παραπάνω μοντέλο ονομάζεται προσθετικό.

Υπάρχει και το μοντέλο με αλληλεπίδραση:

$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, $i = 1,...,I$, $j = 1,\dots,J$, $k = 1,\dots,n_{ij}$.

Θα εξετάσουμε πρώτα το προσθετικό μοντέλο με μία παραρτήρηση ανά κυψελίδα, $n_{ij} = 1, \forall i, j$.

Συνεπώς δε χρειάζομαι το δείκτη k και το μοντέλο είναι:

$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, $i = 1,...,I$, $j = 1,\dots,J$.

Προσθετικό Μοντέλο(1 παρατήρηση ανά κυψελίδα)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},\ i = 1,...,I,\ j = 1,\dots,J,\ \varepsilon_{ij} \sim N(0,\sigma^2).$$

Έχω ότι μ_{ij} =E(Y_{ij})= $\mu + \alpha_i + \beta_j$.

Για να είναι ορισμένο καλά το μοντέλο εισάγω τους περιορισμούς

$\sum_{i=1}^{I} \alpha_i = 0$, $\sum_{j=1}^{J} \beta_j = 0$.

Οι παραπάνω περιορισμοί είναι μέρος του μοντέλου.

$\sum_{j=1}^{J} \mu_{ij} = \mu_{i.} = J\ \mu + J\ \alpha_i \Rightarrow \sum_{i=1}^{I} \mu_{i.} = \mu_{..} = I\ J\ \mu \Rightarrow \mu = \overline{\mu}_{..}$.

Άρα, η παράμετρος μ είναι ο μέσος όρος των μ_{ij} .

Από τα παραπάνω έχω ότι: $\alpha_i = \overline{\mu}_{i.} - \mu = \overline{\mu}_{i.} - \overline{\mu}_{..}$.

$\beta_j = \overline{\mu}_{.j} - \mu = \overline{\mu}_{.j} - \overline{\mu}_{..}$.

<u>Έλεγχοι Υποθέσεων:</u>

H_A: O παράγων A δεν επηρεάζει.

Η υπόθεση αυτή σημαίνει ότι η μέση επίδραση του i επιπέδου του παράγοντα Α ως προς τα επίπεδα του Β παραμένει σταθερή για όλα τα i . Δηλαδή $\overline{\mu}_{i.}$ =σταθερά για όλα τα i άρα

H_A: α_i =0 $\forall i$.

Κατά αντιστοιχία H_B : ο παράγων Β δεν επηρεάζει $\Leftrightarrow H_B$: β_j =0 $\forall j$.

Θα χρησιμοποιήσω τη μέθοδο του αθροίσματος των επιπλέον τετραγώνων.

Αν καλέσω $\underline{Y}$ =(Y_{ij}), $\underline{\varepsilon}$ =(ε_{ij}) και $\underline{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_I \\ \beta_1 \\ \vdots \\ \beta_J \end{pmatrix}$,

τότε μπορώ να γράψω τις εξισώσεις με τη μορφή πινάκων:

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

Θα χρησιμοποιήσω τη στατιστική συνάρτηση:

$F_A = \dfrac{\{SS_{res}(H_A M) - SS_{res}(\Pi M)\} / \beta.\varepsilon.}{SS_{res}(\Pi M) / \beta.\varepsilon.}$.

Πρέπει να υπολογίσω τις εκτιμήτριες ελαχίστων τετραγώνων για το πλήρες μοντέλο.

Ελαχιστοποιούμε $\phi(\mu, \alpha_1, ..., \alpha_I, \beta_1, ..., \beta_J) = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \mu - \alpha_i - \beta_j)^2$,

με $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$.

Αλλά $\dfrac{\partial\phi}{\partial\mu} = -2\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \mu - \alpha_i - \beta_j) = 0 \Rightarrow \hat{\mu} = \overline{Y}_{..}$.

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

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

Οπότε $\hat{Y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \overline{Y}_{i.} + \overline{Y}_{.j} - \overline{Y}_{..}$.

$$SS_{res}(\Pi M) = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2 .$$

Όταν ισχύει η H_A έχω $Y_{ij} = \mu + \beta_j + \varepsilon_{ij}$.

Με ανάλογο τρόπο βρίσκω $\hat{\hat{\mu}} = \overline{Y}_{..}$, $\hat{\hat{\beta}}_j = \overline{Y}_{.j} - \overline{Y}_{..}$, οπότε

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

Όμως $SS_{res}(H_A M) = \sum_{i=1}^{I}\sum_{j=1}^{J}\{(Y_{ij} - \overline{Y}_{.j} - \overline{Y}_{i.} + \overline{Y}_{..}) - (\overline{Y}_{i.} - \overline{Y}_{..})\}^2$

$$= SS_{res}(\Pi M) + J\sum_{i=1}^{I}(\overline{Y}_{i.} - \overline{Y}_{..})^2 .$$

Άρα, $F_A = \dfrac{J\sum_{i=1}^{I}(\overline{Y}_{i.} - \overline{Y}_{..})^2 / \beta.\varepsilon.}{\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2 / \beta.\varepsilon.}$.

Θα βρω τους β.ε.

Για τον αριθμητή είναι (I -1) αφού έχω $\sum_{i=1}^{I}\alpha_i = 0$.

Ισχύει το εξής:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{..})^2 = \overbrace{I\sum_{j=1}^{J}(\overline{Y}_{.j} - \overline{Y}_{..})^2}^{SS_B} + \overbrace{J\sum_{i=1}^{I}(\overline{Y}_{i.} - \overline{Y}_{..})^2}^{SS_A} + \overbrace{\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} - \overline{Y}_{..})^2}^{SS_{res}}$$

SS_{tot}	=	SS_A	+	SS_B	+	SS_{res}
$I\ J$ -1		(I -1)		(J -1)		(I -1)(J -1)

Άρα, $F_A = \dfrac{MS_A}{MS_{res}}$ και απορρίπτω την H_0 αν $F_A > F_{a;(I-1);(I-1)(J-1)}$.

Ανάλογα κατασκευάζω και F-ελεγχοσυνάρτηση για H_B: β_j =0.

$F_B = \dfrac{MS_B}{MS_{res}}$ και απορρίπτω αν $F_B > F_{a;(J-1);(I-1)(J-1)}$.

<u>ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ</u>

Παράγων Α	SS_A	$(I-1)$	$MS_A = \frac{SS_A}{I-1}$	$F_A = \frac{MS_A}{MS_{res}}$
Παράγων Β	SS_B	$(J-1)$	$MS_B = \frac{SS_B}{J-1}$	$F_B = \frac{MS_B}{MS_{res}}$
Υπόλοιπα	SS_{res}	$(I-1)(J-1)$	$MS_{res} = \frac{SS_{res}}{(I-1)(J-1)}$	
Ολικό	SS_{tot}	$I\ J-1$		