"9ο μάθημα" 29/11/21

$Y_{ij} = \mu + a_i + \varepsilon_{ij}$, με $i = 1, \dots, I$ $\quad j = 1, \dots, J_i$

$\hat{\mu} = \bar{Y}_{..}$, $\hat{a}_i = \bar{Y}_{i.} - \bar{Y}_{..}$ έτσι προέκυψαν αυτά

υπό πλευρικές συνθήκες: $\sum_{i=1}^{I} J_i a_i = 0$ (ΠΣ)

Δεδομένα: $\forall\, i = 1, \dots, I \rightarrow Y_{i1}, Y_{i2}, \dots, Y_{iJ_i} \rightarrow \bar{Y}_{i.}$

Ολική Μεταβλητότητα $= \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2 \pm \bar{Y}_i$

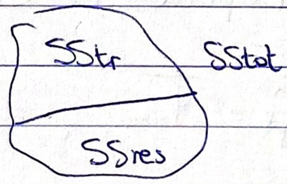
Θεώρημα: Στο μοντέλο ανάλυσης διακύμανσης κατά 1 παράγοντα $\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{I} J_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 +$

$+ \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_{i.})^2 \Rightarrow SStot = SStr + SSres$

Υπόλοιπα: $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{a}_i) =$

$e_{ij} = Y_{ij} - \bar{Y}_{i.}$ $\quad = Y_{ij} - (\cancel{\bar{Y}_{..}} + \bar{Y}_{i.} - \cancel{\bar{Y}_{..}}) \Rightarrow$

$SStot = SStr + SSres$



Απόδειξη: $SStot = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_i)(\bar{Y}_{i.} - \bar{Y}_{..}) = \dots = 0$

ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ μοντέλου της αναλ. διακ. κατά 1 παράγοντα

Πηγή Μεταβλητότητας	SS	Β.Ε.	MS	F-πηλίκο
Δοκιμασίες (επίπεδα του παράγοντα)	SStr	$I-1$	$MStr = \frac{SStr}{I-1}$	$F = \frac{MStr}{MSres}$
Υπόλοιπα	SSres	$(N-1)-(I-1) = N-I$	$MSres = \frac{SSres}{N-I}$	
Ολική	SStot	$N-1$		

Παρατήρηση: $\underset{\sim}{Y} = X \cdot \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \longleftrightarrow Y_{ij} = \mu + \alpha_i + \varepsilon_i$

με $i = 1, \dots, I$ $\quad j = 1, \dots, J_i$

χωρίζουμε τα επίπεδα

$$\underset{\sim}{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1J_1} \\ \hline Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2J_2} \\ \hline \vdots \\ \hline Y_{I1} \\ Y_{I2} \\ \vdots \\ Y_{IJ_I} \end{pmatrix}, \quad X = \left[\begin{array}{ccccc} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 1 & 1 & 0 & \cdots & 0 \\ \hline 1 & 0 & 1 & & 0 \\ 1 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & & 0 \\ \hline & & \vdots & & \\ \hline 1 & 0 & & 0 & 1 \\ 1 & 0 & & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & & 0 & 1 \end{array}\right] \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} J_1 \\ \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} J_2 \\ \\ \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} J_j \end{array}$$

$(J_1 + J_2 + \dots + J_I) \times (I+1) =$
$= N \times (I+1)$

$$\underset{\sim}{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{pmatrix}_{(I+1)\times 1}, \quad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1J_1} \\ \hline \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2J_2} \\ \hline \vdots \\ \hline \varepsilon_{I1} \\ \varepsilon_{I2} \\ \vdots \\ \varepsilon_{IJ_I} \end{pmatrix}_{N \times 1}$$

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

Υποθέσεις για τα σφάλματα

$E(\varepsilon_{ij}) = 0 \quad i = 1, \dots, I \quad j = 1, \dots, J_i \qquad E(Y_{ij}) = \mu + \alpha_i$

$Var(\varepsilon_{ij}) = \sigma^2 \qquad Var(Y_{ij}) = \sigma^2$

$Cov(\varepsilon_{ij}, \varepsilon_{k\ell}) \quad i \neq k, j \neq \ell \quad \xleftrightarrow[Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}]{\text{συνεχείς } Y_{ij}} \quad Cov(Y_{ij}, Y_{k\ell}) = 0$

$\varepsilon_{ij} \sim N(0, \sigma^2) \qquad Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$

→ τα ε_{ij} και Y_{ij} είναι ανεξάρτητα

Ιδιότητες των εκτιμητών (από τις υποθέσεις για τα σφάλματα)

(I_1) Οι ΕΕΤ των μ, α_i με $i=1,\dots,I$ είναι αμερόληπτοι

$$E(\hat{\mu}) = \mu, \quad E(\hat{\alpha}_i) = \alpha_i$$

Απόδειξη: $E(\hat{\mu}) = E(\bar{Y}_{..}) = E\left(\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}\right) =$

$$= \frac{1}{N}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i} E(Y_{ij}) = \frac{1}{N}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i}(\mu+\alpha_i) =$$

$$= \frac{1}{N}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mu + \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\alpha_i =$$

$$= \frac{1}{\cancel{N}}\cancel{N}\cdot\mu + \frac{1}{N}\sum_{i=1}^{I}\cancel{J_i\alpha_i}^{\,0} \overset{\Pi\Sigma}{=} 0$$

$$E(\hat{\alpha}_i) = E(\bar{Y}_{i.} - \bar{Y}_{..}) = E(\bar{Y}_{i}) - E(\underset{\substack{\uparrow\\ \hat{\mu}}}{\bar{Y}_{..}}) = E(\bar{Y}_{i.}) - \mu =$$

$$= E\left(\frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij}\right) - \mu = \frac{1}{J_i}\cdot\sum_{j=1}^{J_i} E(Y_{ij}) - \mu =$$

$$= \frac{1}{J_i}\sum_{j=1}^{J_i} E(\mu+\alpha_i+\varepsilon_{ij}) - \mu = \frac{1}{J_i}\cdot\sum_{j=1}^{J_i}\mu + \frac{1}{J_i}\sum_{j=1}\alpha_i - \mu =$$

$$= \frac{1}{\cancel{J_i}}\cdot\cancel{J_i}\cdot\mu + \frac{1}{\cancel{J_i}}\cancel{J_i}\cdot\alpha_i - \mu = \alpha_i$$

(I_2) Οι ΕΕΤ των μ και α_i ταυτίζονται με τους αντίστοιχους Ε.Μ.Π.

Απόδειξη: Οι ΕΕΤ προκύπτουν από την ελαχιστοποίηση ως προς μ και α_i του $S = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\varepsilon_{ij}^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-\alpha_i)^2$ (*)

L = Η από κοινού κατανομή των δεδομένων = Η από κοινού κατανομή των Y_{ij} με $i=1,\dots,I$, $j=1,\dots,J_i$ $\overset{\text{συνεχ}}{Y_{ij}}$

$$=\sum_{i=1}^{I}\sum_{j=1}^{J_i} f_{Y_{ij}}(Y_{ij}) \underset{=\!=\!=}{\overset{Y_{ij}\sim N(\mu+\alpha_i,\sigma^2)}{}} \prod_{i=1}^{I}\prod_{j=1}^{J_i} \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{1}{2\sigma^2}(Y_{ij}-\mu-\alpha_i)^2}$$

$$\Rightarrow L=\frac{1}{\sigma\sqrt{2\pi}}\cdot n.\, e^{\frac{-1}{2\sigma^2}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-\alpha_i)^2}$$

Οι ΕΜΠ των μ, α_i με $i=1,\dots,I$ προκύπταν από τη μεγιστοποίηση ως προς μ, α_i της L ή από μεγιστοποίηση ως προς μ και α_i του:

$$\log L=-N.\log(\sigma\sqrt{2\pi})-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-\alpha_i)^2$$ ή από

μεγιστοποίηση ως προς μ, α_i του $-\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-\alpha_i)^2$

ή ελαχιστοποίηση του $\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\mu-\alpha_i)^2$ (**)

Από (*), (**) οι ΕΕΤ ταυτίζονται με τους ΕΜΠ (προκύπταν από ελαχιστοποίηση της ίδιας ποσότητας.

Άσκηση (Σπίτι): Να βρεθεί ο ΕΜΠ της σ^2

Θεώρημα: Υπό τις υποθέσεις για τα σφάλματα, $E(MS_{res})=\sigma^2$.

Απόδειξη: $MS_{res}=\frac{SS_{res}}{N-I}$, $SS_{res}=\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\bar{Y}_{i.})^2$.

$\forall\, i=1,\dots,I$ το $Y_{i1}, Y_{i2},\dots,Y_{iJ_i}$ είναι τ.δ. από πληθυσμό με διακύμανση σ^2. Η δειγματική διακύμανση είναι αμερόληπτος εκτιμητής του σ^2 του πληθυσμού, δηλαδή αν $W_1,\dots,W_n$ είναι τ.δ. από πληθ. με διακ/ση σ^2, τότε:

$$E(S_W^2)=E\left(\frac{1}{n-1}\cdot\sum_{i=1}^{n}(W_i-\bar{W})^2\right)=\sigma^2$$

Ισχύει: $E\left(\frac{1}{J_i-1}\sum_{j=1}^{J_i}(Y_{ij}-\bar{Y}_{i.})^2\right)=\sigma^2 \Rightarrow$

$$\Rightarrow E\left(\sum_{j=1}^{J_i}(Y_{ij}-\bar{Y}_{i.})^2\right) = (J_i-1)\cdot\sigma^2, \text{ με } i=1,\dots,I$$

$$E(MSres) = E\left(\frac{1}{N-I}\cdot\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\bar{Y}_{i.})^2\right) = \frac{1}{N-I}\sum_{i=1}^{I}E\left(\sum_{j=1}^{J_i}(Y_{ij}-\bar{Y}_{i.})^2\right) =$$

$$= \frac{1}{N-I}\cdot\sum_{i=1}^{I}(J_i-1)\cdot\sigma^2 = \sigma^2\cdot\frac{1}{N-I}\cdot\sum_{i=1}^{I}(J_i-1) = \sigma^2\cdot\frac{1}{N-I}\left(\sum_{i=1}^{I}J_i - \sum_{i=1}^{I}1\right) =$$

$$= \sigma^2\cdot\frac{1}{\cancel{N-I}}\cancel{(N-I)} = \sigma^2$$

Θεώρημα: Υπό τις υποθέσεις για τα σφάλματα $E(MStr) = \sigma^2 + \frac{1}{I-1}\sum_{i=1}^{I}J_i\alpha_i^2$

Απόδειξη: $MStr = \frac{SStr}{I-1}$ (1), $SStr = \sum_{i=1}^{I}J_i(\bar{Y}_{i.}-\bar{Y}_{..})^2$

Από το μοντέλο $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ και την ΠΣ $\sum_{i=1}^{I}J_i\alpha_i = 0$

Παίρνουμε: $Y_{i.} = J_i\mu + J_i\alpha_i + \varepsilon_{i.}$, $Y_{..} = N\mu + \varepsilon_{..}$, $\varepsilon_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\varepsilon_{ij}$

$\varepsilon_{i.} = \sum_{j=1}^{J_i}\varepsilon_{ij}$

Άρα $E(Y_{i.}) = J_i(\mu+\alpha_i)$, $Var(Y_{i.}) = Var(\varepsilon_{i.}) = J_i\cdot\sigma^2$

$E(Y_{..}) = N\cdot\mu$, $Var(Y_{..}) = Var(\varepsilon_{..}) = N\cdot\sigma^2$

$$E(SStr) = E\left(\sum_{i=1}^{I}J_i(\bar{Y}_{i.}-\bar{Y}_{..})^2\right) = E\left[\sum_{i=1}^{I}J_i\left(\frac{Y_{i.}}{J_i}-\frac{Y_{..}}{N}\right)^2\right] =$$

$$= E\left[\sum_{i=1}^{I}\frac{Y_{i.}^2}{J_i} - \frac{Y_{..}^2}{N}\right] = \sum_{i=1}^{I}\frac{1}{J_i}\cdot E(Y_{i.}^2) - \frac{1}{N}\cdot E(Y_{..}^2) \qquad \underline{\underline{E(w^2) = Var(w) + (Ew)^2}}$$

$$= \sum_{i=1}^{I}\frac{1}{J_i}\left[Var(Y_{i.}) + (EY_{i.})^2\right] - \frac{1}{N}\left[Var(Y_{..}) + (EY_{..})^2\right] =$$

$$= \sum_{i=1}^{I}\frac{1}{J_i}\left[J_i\cdot\sigma^2 + J_i^2(\mu+\alpha_i)^2\right] - \frac{1}{N}\left[N\sigma^2 + N^2\cdot\mu^2\right] = (I-1)\cdot\sigma^2 + \sum_{i=1}^{I}J_i\cdot\alpha_i^2 \quad (2)$$

Από (1), (2): $\underbrace{E(MStr)} = \frac{1}{I-1}\cdot E(SStr) = \frac{1}{I-1}\left(\cdot(I-1)\sigma^2 + \sum_{i=1}^{I}J_i\cdot\alpha_i^2\right) =$

$$\underbrace{= \sigma^2 + \frac{1}{I-1}\cdot\sum_{i=1}^{I}J_i\cdot\alpha_i^2}$$

Έλεγχος της $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_I$: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i=1,\dots,I$, $j=1,\dots,J_i$

Το α_i σχηματοποιεί την επίδραση του i-επιπέδου του παράγοντα στην εξαρτημένη μεταβλητή Y.