"10ο μάθημα"

6/12/21

· Έλεγχος της υπόθεσης $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_I$

- Αν η Ho δεν μπορεί να απορριφθεί, αυτό σημαίνει ότι όλα τα επίπεδα του παράγοντα ισοπιδρούν στην εξαρτ. μεταβλητή Y και η διερεύνηση του μοντέλου τερματίζεται.
- Αν η Ho απορριφθεί, αυτό σημαίνει ότι υπάρχουν επίπεδα του παράγοντα που ασκούν σημαντικότερη επίδραση στην Y και επομένως τίθεται το θέμα του εντοπισμού.

Πως θα κατασκευαστεί τεστ?

- Διευκολύνει η ισοδυναμία $\begin{cases} H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_I (= \alpha) \rightsquigarrow \\ H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_I \rightsquigarrow \end{cases}$

$$\rightsquigarrow Y_{ij} = \overbrace{\mu + \alpha}^{\mu^*} + \varepsilon_{ij} = \mu^* + \varepsilon_{ij}$$

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

Αποδείχτηκε ότι: $E(MSres) = \sigma^2$

$$E(MStr) = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^{I} J_i \cdot \alpha_i^2$$

Αν η Ho δεν μπορεί να απορριφθεί, $MStr \approx MSres$

οπότε $(A \Rightarrow B$, τότε $B \Rightarrow \sim A)$

Αν $MStr \neq MSres \Rightarrow$ η Ho απορ. ή

Αν $MStr \gg MSres \Rightarrow$ η Ho απορ.

Θεώρημα: Υπό τις υποθέσεις για τα σφάλματα και υπό την $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$

(α) $\frac{SSres}{\sigma^2} \sim X^2_{N-I}$, $SSres = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$

(β) $\frac{SStr}{\sigma^2} \sim X^2_{I-1}$, $SStr = \sum_{i=1}^{I} J_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$

Απόδειξη :

(α) Από τις υπόθ. σφ $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$, $i=1,\dots,I$, $j=1,\dots,J_i$
και αποδείξαμε ότι και υπό την Ho $Y_{ij} \sim N(\mu, \sigma^2)$
με $i=1,\dots,I$, $j=1,\dots,J_i$. Αποδείξαμε δηλ. ότι
για $i=1,\dots,I$ $j=1,\dots,J_i$, Y_{iJ_i} τ.δ. από $N(\mu,\sigma^2)$

Ισχύει Αν $W_1,\dots,W_n$ τ.δ. από $N(\mu,\sigma^2)$, τότε

$$\frac{(n-1)S_w^2}{2\sigma^2} \sim X^2_{n-1} \Rightarrow \frac{\sum_{j=1}^{J_i}(Y_{ij} - \overline{Y}_{i\cdot})^2}{\sigma^2} \sim X^2_{J_i-1}$$

$$\Rightarrow \sum_{i=1}^{I} \frac{\sum_{j=1}^{J_i}(Y_{ij} - Y_{i\cdot})^2}{\sigma^2} \sim X^2_{\sum_{i=1}^{I}(J_i-1)}$$

$$\Rightarrow \frac{SSres}{\sigma^2} \sim X^2_{\sum_{i=1}^{I} J_i - \sum_{i=1}^{I} 1} = X^2_{N-I}$$

(β) Για $i=1,\dots,I$, το $Y_{i1}, Y_{i2}, \dots, Y_{iJ_i}$ τ.δ. από $N(\mu,\sigma^2)$

$$\Rightarrow \text{Για } i=1,\dots,I , \quad \overline{Y}_{i\cdot} \sim N\left(\mu, \frac{\sigma^2}{J_i}\right)$$

$$\Rightarrow \frac{\overline{Y}_{i\cdot} - \mu}{\sigma/\sqrt{J_i}} \sim N(0,1) \Rightarrow \frac{J_i(\overline{Y}_{i\cdot} - \mu)^2}{\sigma^2} \sim X^2_1$$

$$\Rightarrow \sum_{i=1}^{I} \frac{J_i(\overline{Y}_{i\cdot} - \mu)^2}{\sigma^2} \sim X^2_{\sum_{i=1}^{I} 1} = X^2_I$$

$$\Rightarrow \frac{\sum_{i=1}^{I} J_i(\bar{Y}_{i\cdot}-\mu)^2}{\sigma^2} \sim X^2_I \Rightarrow \frac{\sum_{i=1}^{I} J_i(\bar{Y}_{i\cdot}-\hat{\mu})^2}{\sigma^2} \sim X^2_{I-1}$$

$$\Rightarrow \frac{\sum_{i=1}^{I} J_i(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2}{\sigma^2} \sim X^2_{I-1} \Rightarrow \frac{SStr}{\sigma^2} \sim X^2_{I-1}$$

<u>Θεωρώ</u> $F = \frac{MStr}{MSres}$. Το $F = \frac{SStr/(I-1)}{SSres/(N-I)} =$

$$= \frac{(SStr/\sigma^2)(I-1)}{(SSres/\sigma^2)(N-I)} \stackrel{d}{=} \frac{X^2_{I-1}/(I-1)}{X^2_{N-I}/(N-I)} \underset{\text{ανεξ. } X^2}{\overset{\text{ισχύει}}{\sim}} F_{I-1,N-I}$$

υπό την Ho.

<u>Μορφή της κ.π.</u>: Μεγάλες τιμές της F-ποσότητας δηλ. $F \geq c$.

<u>Το κ.σ. c υπολογ.</u> από $a = P(\text{απορ. Ho} \mid \text{Ho αληθ.})$ και $c = F_{I-1,N-I,a}$

Για τον έλεγχο της $H_0: \alpha_1 = \dots = \alpha_I$, η ΣΣΤ είναι $F = \frac{MStr}{MSres}$ με κατανομή υπό την Ho την $F_{I-1,N-I}$

και κπ $F \geq F_{I-1,N-I,a}$

<u>Πολλαπλές Συγκρίσεις</u>

Αν η Ho απορ. ($H_0: \alpha_1 = \alpha_2 = \dots = \alpha_I$), τότε →

→ <u>Πολλαπλές Συγκρίσεις</u> → Ελάχιστης Σημαντικής Διαφοράς (ΕΣΔ)
→ Tukey
→ Bonferroni
→ Γραφικών Αναλύσεων ή Scheffe (*)

Αν η H_0 απορριφθεί, αυτό σημαίνει ότι υπάρχουν επίπεδα του παράγοντα τα οποία είναι πιο σημαντικά τα οποία ασκούν σημαντικότερη επίδραση την Y. Μια μέθοδος πολλαπλών συγκρίσεων σκοπό έχει να εντοπίσει τα επίπεδα που ασκούν σημαντικότερη επίδραση στην Y και ακόμα, να τα κατατάξει.

Μέθοδος Ελάχιστης Σημαντικής Διαφοράς

Έστω α_i και $\alpha_{i'}$ η επίδραση των επιπέδων i και i' στην εξαρτ. μεταβλητή Y για $i,i' = 1,\dots,I$

Υπόθεση: $H_0 : \alpha_i = \alpha_{i'}$ (ισοεπίδραση των επιπέδων i και i' στην Y) με $i,i' = 1,\dots,I$

Ένα τεστ θα στηριχτεί στους εκτιμητές $\hat{\alpha}_i$ και $\hat{\alpha}_{i'}$ των παραμέτρων της H_0.

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

Απορ. της H_0 θα στηριχτεί στη διαφορά $\hat{\alpha}_i - \hat{\alpha}_{i'} =$
$= (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) - (\bar{Y}_{i'\cdot} - \bar{Y}_{\cdot\cdot}) = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$ για $i,i' = 1,\dots,I, \ i \neq i'$

Από τις υποθ. για τα σφάλματα $Y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$
$\Rightarrow Y_{i1},\dots,Y_{iJ_i}$ τδ. από $N(\mu+\alpha_i, \sigma^2)$ $\forall i$
Άρα $\bar{Y}_{i\cdot} \sim N(\mu + \alpha_i, \frac{\sigma^2}{J_i})$, $\forall i = 1,\dots,I$ και
$J_i \rightarrow$ μέγεθος δείγμ.

$\bar{Y}_{i\cdot}$ ανεξ. μεταξύ τους, λόγω της ανεξ. των Y_{ij} $\forall i,j$

Επομένως: $\bar{Y}_{i\cdot} \sim N(\mu + \alpha_i, \frac{\sigma^2}{J_i})$, $\bar{Y}_{i'\cdot} \sim N(\mu + \alpha_{i'}, \frac{\sigma^2}{J_{i'}})$

με $i,i' = 1,\dots,I$ και $i \neq i'$ και $\bar{Y}_{i\cdot}, \bar{Y}_{i'\cdot}$ ανεξ/τα.

Επομένως: $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \sim N\left(\mu + \alpha_i - \mu - \alpha_{i'}, \frac{\sigma^2}{J_i} + \frac{\sigma^2}{J_{i'}}\right)$

ή $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \sim N\left(\alpha_i - \alpha_{i'}, \sigma^2\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)\right)$

Υπό την Ho $\alpha_i = \alpha_{i'}$

$\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \sim N\left(0, \sigma^2\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)\right)$ $i, i' = 1, \ldots, I$, $i \neq i'$

ή $\dfrac{\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}}{\sigma^2\sqrt{\frac{1}{J_i} + \frac{1}{J_{i'}}}} \sim N(0,1)$ υπό την Ho: $\alpha_i = \alpha_{i'}$

Γνωρίζω: $\dfrac{SSres}{\sigma^2} \sim X^2_{N-I}$

$\Rightarrow \dfrac{\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}}{\sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}} \sim t_{N-I}$ υπό την Ho: $\alpha_i = \alpha_{i'}$

Μεγάλες τιμές του $\dfrac{\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}}{\sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}}$ οδηγούν σε

απόρριψη της Ho: $\alpha_i = \alpha_{i'}$

Άρα για τον έλεγχο της Ho: $\alpha_i = \alpha_{i'}$, $i, i' = 1, \ldots, I$

χρησιμοποιείται η ΣΣΤ: $\dfrac{\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}}{\sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}}$ με κατανομή

t_{N-I} υπό την Ho και Κ.Π $\dfrac{|\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}|}{\sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}} \sim t_{N-I, \frac{a}{2}}$

Η $H_0: \alpha_i = \alpha_{i'}$, $i,i' = 1,\dots,I$, $i \neq i'$ απορρίπτεται αν

$$\underbrace{|\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}| \geq t_{N-I,\frac{\alpha}{2}} \sqrt{MS_{res}\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}}_{\text{Ε.Σ.Δ.}}$$

Αν η $H_0: \alpha_i = \alpha_{i'}$ απορριφθεί, τότε κάποιο από τα i και i' είναι σημαντικότερο (από τα 2 επίπεδα)

Αν η $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \geq 0$, τότε το i ασκεί σημαντικότερη επίδραση στην Y από το i'

Αν $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \leq 0$, τότε το i' σημαντικότερο από το i

Μέθοδος Γραμμικών Αντιθέσεων ή Scheffe

Επιτρέπει τη διατύπωση υποθέσεων για περισσότερα από δύο επίπεδα του παράγοντα

Ορισμός: Γραμμική Αντίθεση ονομάζεται κάθε γρ. συνδυασμό κυρίων αντιδράσεων α_i, $i=1,\dots,I$ της μορφής
$L = \sum_{i=1}^{I} c_i \alpha_i$, με $\sum_{i=1}^{I} c_i = 0$

Παρατήρηση: ⓐ Αν $c_i = 1$, $c_{i'} = -1$ και c_k, $k \neq i,i'$

τότε $L = \alpha_i - \alpha_{i'}$ και προφανώς η $L=0$ είναι ισοδύναμη με $\alpha_i = \alpha_{i'}$.

ⓑ Αν $c_i = 1$, $c_k = -\frac{1}{2} = c_l$, $c_s = 0$ $\forall s \neq i,k,l$, τότε

$L = \alpha_i - \frac{1}{2}\alpha_k - \frac{1}{2}\alpha_l$ και η $L = 0 \Leftrightarrow \alpha_i = \frac{1}{2}\alpha_k + \frac{1}{2}\alpha_l$

Δηλ. η κύρια επίδραση του i-επιπέδου είναι ταυτόσημη με το ημιάθροισμα των κυρίων επιδράσεων των k και l επιπέδων

Ενδιαφέρει ο έλεγχος $H_0: L=0$. Το τέστ θα βασιστεί στην εκτιμήτρια της L, την $\hat{L}=\sum_{i=1}^{I} c_i \cdot \hat{\alpha}_i =$

$$=\sum_{i=1}^{I} c_i(\bar{Y}_{i.}-\bar{Y}_{..})=\sum_{i=1}^{I} c_i\bar{Y}_{i.}-\bar{Y}_{..}\cancelto{0}{\sum_{i=1}^{I} c_i}$$ και έτσι

$$\hat{L}=\sum_{i=1}^{I} c_i\bar{Y}_{i.}$$

Αλλά $Y_{i1},Y_{i2},\dots,Y_{iJ_i}$ από $N(\mu+\alpha_i,\sigma^2) \Rightarrow$

$\Rightarrow \bar{Y}_{i.}\sim N\left(\mu+\alpha_i,\frac{\sigma^2}{J_i}\right)$ και ανεξάρτητα.

Άρα $\hat{L}=\sum_{i=1}^{I} c_i\bar{Y}_{i.}\sim$ Normal γρ. συνδυασμός Normal ανεξ. normal

$$E(\hat{L})=\sum_{i=1}^{I} c_i E(\bar{Y}_{i.})=\sum_{i=1}^{I} c_i(\mu+\alpha_i)=\mu\cdot\cancelto{0}{\sum_{i=1}^{I} c_i}+\sum_{i=1}^{I} c_i\alpha_i=L$$

$$Var(\hat{L})=\sum_{i=1}^{I} c_i^2\cdot Var(\bar{Y}_{i.})=\sum_{i=1}^{I} c_i^2\frac{\sigma^2}{J_i}=\sigma^2\cdot\sum_{i=1}^{I}\frac{c_i^2}{J_i}$$

Άρα $\hat{L}\sim N\left(L,\sigma^2\cdot\sum_{i=1}^{I}\frac{c_i^2}{J_i}\right)$ υπό την $H_0: L=0$

$\hat{L}\sim N\left(0,\sigma^2\cdot\sum_{i=1}^{I}\frac{c_i^2}{J_i}\right)$. Άρα υπό την $H_0: L=0$

τότε $\dfrac{\hat{L}}{\sigma^2\sqrt{\sum_{i=1}^{I}\frac{c_i^2}{J_i}}}\sim N(0,1)\Rightarrow \dfrac{\hat{L}^2}{\sigma^2\cdot\sum_{i=1}^{I}\frac{c_i^2}{J_i}}\sim X_1^2$

$\dfrac{SSres}{\sigma^2}\sim X^2_{N-I}$

Έστω $F_L=\dfrac{MS_L}{MSres}=\dfrac{\hat{L}^2/\sum_{i=1}^{I}\frac{c_i^2}{J_i}}{SSres/(N-I)}=\dfrac{\hat{L}^2/\sigma^2\cdot\sum_{i=1}^{I}\frac{c_i^2}{J_i}}{SSres/\sigma^2(N-I)}\overset{d}{=}$

$$\overset{d}{=}\frac{X_1^2/1}{X^2_{N-I}/(N-I)}\overset{\text{ανεξ}}{\sim}F_{1,N-I}$$

Άρα $F_1 = \frac{MS_L}{MS_{res}}$, $MS_L = \frac{\hat{L}^2}{\sum_{i=1}^{I} \frac{c_i^2}{J_i}}$, $F_L \sim F_{1,N-I}$ υπό $H_0: L=0$

Μορφής της κ.π.: Μεγάλες τιμές του F_L, $F_L \geq c$.

γιατί μεγάλες τιμές του F_L σημαίνει μεγάλες τιμές του αριθμητή $\hat{L}^2$, δηλ. μεγάλες τιμές του $\hat{L}$, δηλ. μεγάλες τιμές του L που συνεπάγεται σε απόρ. της $H_0: L=0$.

Υπολογισμός κσ. c: $a = P(F_L \geq c \mid F_L \sim F_{1,N-I}) \Rightarrow c = F_{1,N-I,a}$

Συγκεντρωτικά: Για τον έλεγχο $H_0: L=0$, η ΣΣΤ είναι $F_L = \frac{MS_L}{MS_{res}} \sim F_{1,N-I}$ υπό την H_0, $MS_L = \frac{\hat{L}^2}{\sum_{i=1}^{I} \frac{c_i^2}{J_i}}$

και κ.π. $F_L \geq F_{1,N-I,a}$

Ανάλυση Υπολοίπων:

$e_{ij} = Y_{ij} - \overline{Y}_{i\cdot}$, $i=1,\dots,I$ $j=1,\dots,J_i$

αξιοποιείται για έλεγχο υποθέσεων για σφάλματα

Άσκηση: Μια εταιρία κατασκευάζει όργανα για αεροπλάνα τα οποία φωσφορίζουν για κάποιο χρον. διάστημα μετά το σβήσιμό τους. Η εταιρία για να πετύχει το σκοπό της μπορεί να διαλέξει μεταξύ τριών μεθόδων παραγωγής. Τα παρακάτω δεδομένα δίνουν το χρόνο σε δευτερόλεπτα που φωσφορίζουν τα εν λόγω όργανα για

καθε μια απο τις μεθόδους Να αναλύσω τα δεδο-
μένα αυτά

Μέθοδος 1 : 52.9, 62.1, 57.4, 50, 59.3, 61.2, 60.8, 53.1

Μέθοδος 2 : 58.4, 55, 59.8, 62.5, 64.7, 59.9, 54.7, 58.4

Μέθοδος 3 : 71.3, 66.6, 63.4, 64.7, 75.8, 65.6, 72.9, 67.3