Δευτέρα 14/12/20

Μέθοδος Γραμμικών αντιθέσεων ή μέθοδος scheffe

Ορισμός:

Γραμμική αντίθεση λέγεται κάθε γραμμικός συνδυασμός των κύριων επιδράσεων $a_i, i=1,2,..I$ των επιπέδων που παράγοντα στην εξαρτημένη μεταβλητή της μορφής $L=\sum_{i=1}^{I} c_i a_i$ με $\sum_{i=1}^{I} c_i=0$

Μας ενδιαφέρει ο έλεγχος της $H_0: L=0$, γιατί δίνει την δυνατότητα πλήθους ελέγχων που αφορούν τις επιδράσεις a στην εξαρτ. μεταβλητή

Η γραμμική αντίθεση επιτρέπει τη διατύπωση συγκρίσεων περισσότερων των δύο $a_i, i=1,2,..I$

π.χ. ① $H_0: L=0 \longleftrightarrow H_0: a_i=a_{i'}$ αν $c_i=1$, $c_{i'}=-1$ και $c_i=0$ για τα άλλα i

② Αν $c_i=1$, $c_k=-\frac{1}{2}$, $c_l=-\frac{1}{2}$, $c_j=0$, $j\neq i,k,l$

τότε

$H_0: L=0 \longleftrightarrow H_0: a_i=\frac{1}{2}(a_k+a_l)$

$\hookrightarrow L=a_i-\frac{1}{2}a_k-\frac{1}{2}a_l=0$, $\sum_{i=1}^{I} c_i=0$

Κατασκευή στατιστικού τεστ για τον έλεγχο $H_0: L=0$:

Θεωρούμε ότι οι υποθέσεις για τα σφάλματα ικανοποιούνται τότε $\bar{Y}_{i\cdot} \sim N\left(\mu+a_i, \frac{\sigma^2}{J_i}\right), i=1,..I$. Με την κατά Wald προσεγ.

το τέστ για τον έλεγχο $H_0: L=0$ θα στηρίζεται σε εκτιμήτη της L

$\hat{L}=\sum_{i=1}^{I} c_i\hat{a}_i$, $\hat{a}_i=\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}$

έτσι $L=\sum_{i=1}^{I} c_i(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}) = \sum_{i=1}^{I} c_i\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\cancel{\sum_{i=1}^{I} c_i}^{\,0} = \boxed{\sum_{i=1}^{I} c_i\bar{Y}_{i\cdot}}$

Το $\hat{L}$ είναι γραμμ. συνδυασμός των ανεξ. $\bar{Y}_{i\cdot} \sim N\left(\mu+a_i, \frac{\sigma^2}{J_i}\right)$

$\Rightarrow \hat{L} \sim$ Normal.

$$E\hat{L} = E\left(\sum c_i \bar{Y}_{i\cdot}\right) = \sum c_i E\left(\bar{Y}_{i\cdot}\right) = \sum c_i(\mu + a_i) =$$

$$= \mu \underbrace{\cancel{\sum c_i}}_{0} + \sum c_i a_i = L$$

$$\operatorname{Var}\hat{L} \overset{\bar{Y}_{i\cdot}\ \text{ανεξ.}}{=\!=} \sum c_i^2 \operatorname{Var}\left(\bar{Y}_{i\cdot}\right) = \sum_{i=1}^{I} c_i^2 \frac{\sigma^2}{J_i} = \sigma^2 \sum \frac{c_i^2}{J_i}$$

Άρα υπό την H_0: $L=0 \Rightarrow \boxed{\hat{L} \sim N\left(0, \sigma^2 \sum \frac{c_i^2}{J_i}\right)}$

Άρα $\dfrac{\hat{L}}{\sigma\sqrt{\sum \frac{c_i^2}{J_i}}} \sim N(0,1) \Rightarrow \dfrac{\hat{L}^2}{\sigma^2 \sum \frac{c_i^2}{J_i}} \sim \chi_1^2$ υπό την Ho.

↳ Δεν μπορώ να την χρησιμοποιήσω στο τεστ γιατί έχει τη σ^2 που μας είναι άγνωστη, άρα στόχος να το απαλείψουμε

Επίσης $\dfrac{SSres}{\sigma^2} \sim \chi^2_{N-1}$, Συμβολίζουμε $MS_L = \dfrac{\hat{L}^2}{\sigma^2 \sum \frac{c_i^2}{J_i}}$

Θεωρούμε $F_L = \dfrac{MS_L}{MSres} = \dfrac{\frac{\hat{L}^2}{\cancel{\sigma^2}\sum c_i^2/J_i}}{\frac{SSres}{\cancel{\sigma^2}(N-1)}} = \dfrac{\hat{L}^2 / \sum c_i^2/J_i}{SSres/N-1} \overset{\text{είναι ξ.}}{\underset{H_0}{\sim}} F_{1,N-1}$

$$\boxed{F_L = \frac{(N-1)\hat{L}^2}{SSres \cdot \sum \frac{c_i^2}{J_i}} \sim F_{1,N-1}}$$

Μορφή κρίσιμης περιοχής.

Μεγάλες τιμές του F_L οδηγούν σε μεγάλες τιμές του $\hat{L}$ και επειδή το $\hat{L}$ εκτιμά το L, μεγάλες τιμές του F_L οδηγούν σε απορ. της H_0: $L=0$.

κπ. $\boxed{F_L \geq C}$

$a = P(\text{Απορ. } H_0 \mid H_0 \text{ αληθής}) = P(F_L \geq C \mid F_L \sim F_{1,N-1})$

άρα $\boxed{c = F_{1, N-I, a}}$

Συγκεντρωτικά:

Για τον έλεγχο της μηδενικής υπόθεσης: $H_0: L = 0$ η SSL είναι

$F = \dfrac{MS_L}{MS_{res}}$, όπου $MS_L = \dfrac{L^2}{\sum_{i=1}^{I} \frac{c_i^2}{J_i}}$ με κατανομή

$F_{1, N-I}$ υπο την H_0 Κρ. περιοχή $F_L \geq F_{1, N-I, a}$

Ανάλυση υπολοίπων

Υπόλοιπα: $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{a}_i) = Y_{ij} - (\bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..}) =$
$= Y_{ij} - \bar{Y}_{i.}$, $i = 1, \dots I$, $j = 1 \dots J_i$

Αξιοποιούνται για τον έλεγχο των υποθέσεων για τα σφάλματα όπως και στα μοντέλα πολυνδρόμησης.

Σχέση Ανάλ. διακύμανσης κατά ένα παράγοντα και γενικού Γραμμικού μοντέλου.

$Y_{ij} = \mu + a_i + \varepsilon_{ij}$, $i = 1, \dots I$, $j = 1, \dots J_i$
$\underline{Y} = X \underset{\sim}{\beta} + \varepsilon$

Ορίζουμε $\underset{\sim}{Y} = (Y_{11}, Y_{12}, Y_{1J_1}, Y_{21}, Y_{22}, \dots Y_{2J_2}, \dots Y_{I1}, Y_{I2}, Y_{IJ_I})^T$

$$X = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & \vdots & 1 \\ \vdots & & \vdots \\ 0 & 0 & 1 \end{bmatrix}}_{\sum_{i=1}^{I} J_i \times I} \qquad \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_I \end{bmatrix}$$

$\underset{\sim}{\varepsilon} = (\varepsilon_{11}, \dots \varepsilon_{1J_1}, \varepsilon_{21}, \dots \varepsilon_{2J_2}, \dots \varepsilon_{I1}, \dots \varepsilon_{IJ_I})^T$

Επομένως:

$$\underset{\sim}{y}=X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon} \Leftrightarrow \left\{\begin{array}{ll} Y_{ij}=\mu_i+\varepsilon_{ij}, & \text{αν } \beta_i=\mu_i \\ Y_{ij}=\mu+a_i+\varepsilon_{ij}, & \text{αν } \beta_i=\mu+a_i \end{array}\right\}$$

Παράδειγμα

Στο μοντέλο αναλ. διακύμανσης κατά ένα παράγοντα με $\varepsilon_{ij}\sim N(0,\sigma^2)$, $i=1,\dots,I$, $j=1,\dots,J_i$ και ασυσχέτιστα.

ν.δ.ο. α) οι ΕΕΤ. των μ και a_i είναι αμερόληπτοι

β) οι ΕΕΤ. » » ταυτίζονται με τους ΕΜΠ

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

$$=\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i}E(Y_{ij})=\frac{1}{N}\sum_i\sum_j E(\mu+a_i+\varepsilon_{ij})=$$

$$=\frac{1}{N}\sum_i\sum_j(\mu+a_i+E(\varepsilon_{ij}))=\frac{1}{N}\sum\sum\mu+\sum_i\sum_{j=1}^{J_i}a_i$$

$$=\frac{1}{N}(N\mu+\underbrace{\cancel{\sum J_i a_i}}_{\to 0}) \text{ πλευρική συνθήκη}$$

$$=\mu.$$

$\hat{a}_i=\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}$

$$E(\hat{a}_i)=E(\bar{Y}_{i\cdot})-E(\bar{Y}_{\cdot\cdot})=E\left(\frac{1}{J_i}\sum_{j=1}^{J_i}Y_{ij}\right)-E\left(\frac{1}{N}\sum\sum Y_{ij}\right)$$

$$=\frac{1}{J_i}\sum E(Y_{ij})-\frac{1}{N}\sum\sum E(Y_{ij})=$$

$$=\frac{1}{J_i}\sum_{j=1}E(\mu+a_i+\varepsilon_{ij})-\frac{1}{N}\sum\sum E(\mu+a_i+\varepsilon_{ij})=$$

$$=\frac{1}{J_i}\sum(\mu+a_i)-\frac{1}{N}\sum_{i=1}^{I}J_i(\mu+a_i)=$$

$$= \frac{1}{J_i}(J_i\mu + J_i a_i) - \frac{1}{N}(N\mu + \sum J_i a_i) = \mu + a_i - \mu = a_i$$

β) Οι ΕΕΤ προκύπτουν από την ελαχιστοποίηση του

$$Q = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \varepsilon_{ij}^2 = \sum\sum (y_{ij} - \mu - a_i)^2 \; (1)$$ ως προς μ και a_i

Οι ΕΜΠ προκύπτουν πάντα από τη μεγιστοποίηση της πιθανοφάνειας δηλ. της από κοινού κατανομής των δεδομένων y_{ij}.

$$L = \frac{1}{(\sigma\sqrt{2\pi})^N} e^{-\frac{1}{2\sigma^2}\sum\sum (y_{ij} - \mu - a_i)^2}$$

$\varepsilon_{ij} \sim N(0, \sigma^2)$

$y_{ij} \sim N(\mu + a_i, \sigma^2)$

Οι ΕΜΠ προκύπτουν από τη μεγιστοποίηση της L ή του $\log L$ ως προς N και a_i

$$-N\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum\sum (y_{ij} - \mu - a_i)^2$$ ή της ποσότητας

$$-\sum\sum (y_{ij} - \mu - a_i)^2 \; (2)$$

Από τις (1) και (2) οι ΕΕΤ$(\mu, a_i) \equiv$ ΕΜΠ(μ, a_i)

Παράδειγμα:

Μια εταιρεία κατασκευάζει ενδεικτικά όργανα για αεροπλάνα με την ιδιότητα τα όργανα αυτά να φωσφορίζουν για κάποιο χρονικό διάστημα μετά το σβήσιμο της λάμπας που τα φωτίζει. Η εταιρεία για να πετύχει την ιδιότητα αυτή μπορεί να επιλέξει μεταξύ τριών μεθόδων (διαφορετικών) παραγωγής. Τα ακόλουθα δεδομένα δίνουν το χρόνο σε sec που φωσφορίζουν τα εν λόγω όργανα για κάθε μια από τις τρεις μεθόδους παραγωγής. Να αναλυθούν τα δεδομ

μεθ. 1 59.2, 62.1, 57.4, 50, 59.3, 61.2, 60.8, 53.1

μεθ. 2 58.4, 55, 59.8, 62.5, 64.7, 59.9, 54.7, 58.4

μεθ. 3 71.3, 60.6, 63.4, 64.7, 675.8, 65.6, 72.9, 67.3

Έστω Y ο χρόνος σε sec που φωσφορίζουν και τρεις μεθόδους παραγωγής που πιθανώς επηρεάζουν το χρόνο Y.

Y ← μέθοδος παραγωγής — 1, 2, 3

↑ παράγοντας

Μοντέλο

$Y_{ij} = \mu + a_i + \varepsilon_{ij}$, $i = 1,2,3$, $j = 1,2 \ldots J_i$, $J_1 = J_2 = J_3 = 8$

Υποθέτω ότι οι υποθέσεις για τα σφάλματα ισχύουν.

Έτσι $a_i =$: παριστά την κύρια επίδραση της i-μεθόδου πάνω στο χρόνο Y.

Πίνακας ΑΝΑΔΙΑ

μεταβλ.	SS	β.Ε	MS	f-πηλ.
~~Δοκιμ~~				
Δοκιμασίες	584,410	2	292,205	$\boxed{17,044}$
Υπόλοιπα	360,015	21	17,144	
Ολική μεταβλ.	944,425	23	///	

Έλεγχος υπόθεσης $H_0: a_1 = a_2 = a_3$ ισοεπίδραση μεθόδων

$F = 17,044 \geq F_{I-1, N-I, a} = f_{2,21,0.05} = 3,78$

Άρα απορ. η $H_0: a_1 = a_2 = a_3$

Επομένως πάω σε Πολλαπλές συγκρίσεις. (ΕΣΔ).

$$ΕΣΔ = t_{N-I, \frac{a}{2}} \sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)} = 2,08 \cdot \sqrt{17,144 \cdot \frac{2}{8}} = 4,306$$

$\bar{Y}_{1\cdot} = 57,100$ $\bar{Y}_{2\cdot} = 59,175$ $\bar{Y}_{3\cdot} = 68,45$

$\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot} = -2,075$ $\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot} = -11,350$
$\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot} = -9,275$

$|\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}| = 2.075 < 4.306$ Αποδ. $H_0: a_1 = a_2$
$|\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot}| = 11,350 > 4,306$ Απορ. $H_0: a_1 = a_3$.
$|\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot}| = 9,275 > 4,306 \rightarrow$ Απορ $H_0: a_2 = a_3$.

Επειδή $\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot} < 0$, $\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot} < 0$
Η μέθοδος 1 χειρότερη της μεθ. 3 και η μεθ 2 χειρότερη από την 3.

Τελικά η 1η είναι ισοδύναμη με τη 2η και οι δύο τους είναι χειρότερες από την 3η.

Μοντέλο ανάλυσης διακύμανσης κατά 2 παράγοντες

Διερεύνηση του τρόπου που δύο παράγοντες επηρεάζουν την Y

Y ποσοτική

A ← I επίπεδα

B ← J επίπεδα.

(π.χ.) Y = επίδοση μαθητών, A επίπεδο μόρφωσης πατέρα B επίπεδο μητέρας

Δεδομένα για ανάλυση του μοντέλου.
Έστω Y_{ij} παρατήρηση που έχει ληφθεί για την Y. πάνω στο i-επίπεδο του παράγοντα A και το j-επίπεδο του παράγοντα B, $i = 1, .. I$, $j = 1, .. J$.

Παραγ. A	Παράγοντας B 1	2	.	J	Σύνολα	Μέσοι όροι
1	y_{11}	y_{12}		y_{1J}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}		y_{2J}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮						
I	y_{I1}	y_{I2}		y_{Ij}	$y_{I\cdot}$	$\bar{y}_{I\cdot}$
Σύνολα	$y_{\cdot 1}$	$y_{\cdot 2}$		$y_{\cdot j}$		
Μέσοι όροι	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$		$\bar{y}_{\cdot j}$		

$$\boxed{y_{i\cdot} = \sum_{j=1}^{J} y_{ij}} \qquad \boxed{\bar{y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} y_{ij}}$$

$$\boxed{y_{\cdot j} = \sum_{i=1}^{I} y_{ij}} \qquad \boxed{\bar{y}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} y_{ij}}$$

$$\boxed{y_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}} = \sum_{i=1}^{I} y_{i\cdot} = \sum_{j=1}^{J} y_{\cdot j}$$

$$\boxed{\bar{y}_{\cdot\cdot} = \frac{1}{I\cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{\cdot\cdot}} = \frac{1}{I}\sum_{i=1}^{I}\bar{y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J}\bar{y}_{\cdot j} = \frac{1}{IJ} y_{\cdot\cdot}$$

Μοντέλο:

ατομική επίδραση στην Y του i επιπέδου A

$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ (σφάλματα), $\forall\, i = 1, \dots I$, $j = 1, \dots J$.

y_{ij}: παρατήρηση της εξαρ. μετ.

β_j: ατομική επίδραση στην Y του j επιπέδου του B

μ: μια κοινή επίδραση στην Y όλων των επιπέδων, 2 παραγόντων (δηλ. όλων των δοκιμασιών).

Εκτιμητές ελαχίστων τετραγώνων των παραμέτρων: μ, a_i, b_j

Ελαχιστοποίηση της $S^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \varepsilon_{ij}^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} (Y_{ij} - \mu - a_i - b_j)^2$

$$\frac{\partial S^2}{\partial \mu} = 0 \quad \text{και} \quad \frac{\partial S^2}{\partial a_i} = 0 \ \forall i \quad \text{και} \quad \frac{\partial S^2}{\partial b_j} = 0 \ \forall j$$

κανον. εξ.

$$\Rightarrow \begin{cases} IJ\mu + J\sum a_i + I\sum b_j = Y_{..} \\ J\mu + Ja_i + \sum b_j = Y_{i.} \\ I\mu + \sum a_i + Jb_j = Y_{.j} \end{cases}$$

Το σύστημα δεν έχει μοναδική λύση, επομένως για να πετύχω μοναδική λύση καταφεύγω στην υιοθέτηση πλασματικών συνθηκών. Διαισθητικά ένας εκτιμητής του μ πρέπει να είναι $\bar{Y}_{..}$. Έτσι πρέπει $\sum a_i = \sum b_j = 0$

Άρα
$$\begin{cases} \hat{\mu} = \dfrac{Y_{..}}{IJ} = \bar{Y}_{..} \\ \hat{a}_i = \bar{Y}_{i.} - \bar{Y}_{..} \quad \forall i = 1, \dots I \\ \hat{b}_j = \bar{Y}_{.j} - \bar{Y}_{..} \quad \forall j = 1, \dots J \end{cases}$$

Υποθέσεις για τα σφάλματα:

Συνέπειες σφαλμάτων στο μοντέλο

1) $E(\varepsilon_{ij}) = 0 \rightsquigarrow E(Y_{ij}) = \mu + a_i + b_j$
2) $Var(\varepsilon_{ij}) = \sigma^2 \rightsquigarrow Var(Y_{ij}) = \sigma^2$
3) $Cov(\varepsilon_{ij}, \varepsilon_{kl}) = 0 \rightsquigarrow Cov(Y_{ij}, Y_{kl}) = 0$
4) $\varepsilon_{ij} \sim N(0, \sigma^2) \rightsquigarrow Y_{ij} \sim N(\mu + a_i + b_j, \sigma^2)$

Ιδιότητες των ΕΕΤ

1) Αν ε_{ij} ικανοποιούν τις υποθέσεις για τα σφάλματα τότε
$E(\hat{\mu}) = \mu$, $E(\hat{a}_i) = a_i$, $E(\hat{b}_j) = b_j$, $i = 1, 2, \dots I$.

2) Ισχύει ότι $EET(\mu, a_i, b_j) = EM\Pi(\mu, a_i, b_j)$

Η πιθανοφάνεια είναι $L = \prod_{i=1}^{I} \prod_{j=1}^{J} f_{Y_{ij}}(y_{ij}) = \prod \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(Y_{ij} - \mu - a_i - b_j)^2}$

$= \frac{1}{\sigma^{IJ}(2\pi)^{IJ/2}} e^{-\frac{1}{2\sigma^2}\sum\sum(Y_{ij} - \mu - a_i - b_j)^2}$

Έπειτα παίρνουμε λογάριθμο της L και εκτελούμε τα ίδια που κάναμε σε προηγούμενα μαθήματα