9ο

10/12/2019

Παλινδρόμηση και Ανάλυση Διακύμανσης

$H_0: a_1 = a_2 = \dots = a_I$

Για τον έλεγχο της H_0, η ΣΣΤ : $F = \frac{MStr}{MSres} \sim F_{I-1, N-I}$ υπο την H_0

και κ.π $F \geq F_{I-1, N-I, a}$.

Αν η H_0 δεν μπορεί να απορριφθεί → τελειώσαμε.

ή Αναλ. Διακυμ. κατά έναν παράγοντα.

Αν η H_0 απορριφθεί, αυτό σημαίνει ότι υπάρχουν επίπεδα a_i, του παράγοντα, που ασκούν σημαντικότερη επίδραση στην Y.

Αν η H_0: απορριφθεί τότε → πολλαπλ.

Συγκρίσεις $\xrightarrow{\text{Κορυφαία}}$ ΕΣΔ.

$$\boxed{ΕΣΔ = t_{N-I, \frac{a}{2}} \sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_{i'}}\right)}}$$

Η $H_0: a_i = a_{i'}$ (το i κ i' επίπεδο ισοεπιδρούν στην Y)

απορρ αν $|\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}| \geq$ ΕΣΔ

Αν η H_0 απορριφθεί καποιο από τα i και i' είναι καλύτερο.

Ποιο είναι καλύτερο?

↳ Ανάλογα με το πρόσημο της διαφοράς $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$

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

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

ΠΑΡΑΔΕΙΓΜΑ

Μια εταιρία κατασκευάζει ενδικτικά όργανα για τα αεροπλάνα με την ιδιότητα να φωσφορίζουν για κάποιο χρονικό διάστημα μετα το σβήσιμο της λάμπας που τα φωτίζει. Η εταιρία για να πετύχει την ιδιότητα αυτή μπορεί να διαλέξει μεταξύ 3 διαφορετικών μεθόδων παραγωγής. Τα παρακάτω δεδομένα δίνουν τον χρόνο σε λεπτά που φωσφορίζουν τα όργανα για κάθε μέθοδο μετα το σβήσιμο της λάμπας. Να αναλυθούν τα δεδομένα και να εξαχθούν συμπεράσματα.

ΜΕΘΟΔΟΣ 1:	52,9	62,1	57,4	50	59,3	61,2	60,8	53,1
ΜΕΘΟΔΟΣ 2:	58,4	55	59,8	62,5	64,7	59,9	54,7	58,4
ΜΕΘΟΔΟΣ 3:	71,3	66,6	63,4	64,7	75,8	65,6	72,9	67,3

ΛΥΣΗ

Χρόνος ← εξαρτάται από την μέθοδο παραγωγής

Μεθ 1 Μεθ 2 Μεθ 3

Έστω Υ ο χρόνος φωσφορισμα, ο οποίος εξαρτάται από έναν παράγοντα (μέθοδος κατασκευής)

Μεθ.1 Μεθ.2 Μεθ.3

$$Y_{ij} = \mu + a_i + \varepsilon_{ij} \quad \boxed{i = 1,2,3} \quad \boxed{j = 1,2,\dots,8}$$

ΑΝΑΔΙΑ

Πηλικ.	SS	β.ε	MS	F-πηλικο
Δοκιμή	584,41	2	292,205	F = 17,044
Υπόλοιπα	360,015	21	17,144	
Ολική	944,425	23		

Η $H_0: \alpha_1 = \alpha_2 = \alpha_3$ απορρ διότι το F πηλίκο

$F = 17{,}044 > F_{2,21,0{,}05} = 5{,}78$ και $F = 17{,}044 > F_{2,21,0{,}01} = 4{,}32$

Έτσι θα περάσουμε σε πολλαπλές συγκρίσεις.

$$E\Sigma\Delta = t_{21,\,0.025}\left(17{,}144\left(\frac{1}{8}+\frac{1}{8}\right) = 4{,}306\right)$$

Αν κάνω πράξεις έχω ότι $\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot} = -2{,}075$

$\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot} = -11{,}35$

$\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot} = -9{,}975$

Επειδή $|\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}| < E\Sigma\Delta \rightarrow$ ΔΕΝ μπορώ να απορρίψω την $H_0: \alpha_1 = \alpha_2$

Άρα θα δεχτώ ότι Μεθ 1 ≡ Μεθ 2.

Επειδή $|\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot}| > E\Sigma\Delta \rightarrow$ Απορρίπτω την $H_0: \alpha_1 = \alpha_3$

$\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot} < 0 \Rightarrow$ Μεθ 1 < Μεθ 3

Επειδή $|\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot}| > E\Sigma\Delta \rightarrow$ Απορρίπτω την $H_0: \alpha_2 = \alpha_3$

$\bar{Y}_{2\cdot} - \bar{Y}_{3\cdot} < 0 \Rightarrow$ Μεθ 2 < Μεθ 3

Συγκεντρωτικά $\boxed{\text{Μεθ 1} \equiv \text{Μεθ 2} < \text{Μεθ 3}}$

ΜΕΘΟΔΟΣ ΓΡΑΜΜΙΚΩΝ ΑΝΤΙΘΕΣΕΩΝ ή ΜΕΘΟΔΟΣ ΤΟΥ Scheffe:

Ορισμός: Με τον όρο γραμμική αντίθεση εννοούμε γραμμικό συνδυασμό των επιπέδων α_i, $i = 1, \dots, I$ της εξής μορφής $L = \sum_{i=1}^{I} c_i \alpha_i$ με $\sum_{i=1}^{I} c_i = 0$.

Μας ενδιαφέρει ο έλεγχος της $H_0: L = 0$

ΓΙΑΤΙ?

Γιατί μπορεί να διατυπωθούν πολλές άλλες μηδενικές υποθέσεις για τα α_i, $i = 1, \dots, n$.

$\underline{\pi\chi}$ ▷ $H_0: L=0 \iff H_0: a_i = a_{i'}$

Αν $c_i = 1$, $c_{i'} = -1$ και $c_k = 0$ $k \neq i, i'$

▷ $H_0: L=0 \iff H_0: a_k = \frac{1}{2} a_i + \frac{1}{2} a_{i'}$

Αν $c_k = 1$, $c_i = -\frac{1}{2}$, $c_{i'} = -\frac{1}{2}$ και $c_j = 0$ $j \neq i, i', k$.

Κατασκευή τεστ για έλεγχο της $H_0: L=0$.

Wald: Στηριζόμαστε σε έναν εκτιμητή της ποσότητας που εμφανίζεται στην H_0, δηλαδή σ' έναν εκτιμητή της L.

Θεωρούμε $\hat{L} = \sum_{i=1}^{I} c_i \hat{a}_i$.

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

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

- $E(\hat{L}) = \sum c_i E(\bar{Y}_{i\cdot})$

Επειδή $Y_{ij} \sim N(\mu + a_i, \frac{\sigma^2}{J_i})$ $i = 1, \dots, I$

Έτσι $E(\hat{L}) = \sum_{i=1}^{I} c_i(\mu + a_i) = \mu \cancel{\sum_{i=1}^{I} c_i}^{\,0} + \sum_{i=1}^{I} c_i a_i$.

$$\Rightarrow E(\hat{L}) = \sum_{i=1}^{I} c_i a_i = L$$

- $Var(\hat{L}) = Var\left(\sum_{i=1}^{I} c_i \bar{Y}_{i\cdot}\right) \overset{\text{ανεξ}}{=} \sum_{i=1}^{I} c_i^2 Var(\bar{Y}_{i\cdot})$

$$= \sum_{i=1}^{I} c_i^2 \frac{\sigma^2}{J_i} = \sigma^2 \sum_{i=1}^{I} \frac{c_i}{J_i}$$

Το $\hat{L}$ έχει κανονική ως γραμμικός συνδυασμός ανεξ. κανονικών $(\bar{Y}_{i\cdot})$

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

Υπό την $H_0: L=0$, $\hat{L} \sim N\left(0, \sigma^2 \sum_{i=1}^{I} \frac{c_i}{J_i}\right) \rightarrow \dfrac{\hat{L}^2}{\sigma^2 \sum_{i=1}^{I} \frac{c_i}{J_i}} \sim X^2_1$

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

Έστω $MS_L = \dfrac{\hat{L}^2}{\sum_{i=1}^{I} \frac{c_i}{J_i}}$

Προφανώς $\dfrac{MS_L}{\sigma^2} \sim X_1^2$

Ορίζουμε $F_L = \dfrac{MS_L}{MS_{res}} = \dfrac{\frac{\hat{L}^2}{\sum \frac{c_i}{J_i}}}{\frac{SS_{res}}{(N-I)}} = \dfrac{\frac{\hat{L}^2}{\sigma^2 \sum \frac{c_i}{J_i}}}{\frac{SS_{res}}{\sigma^2(N-I)}} \sim \dfrac{X_1^2}{X^2_{N-I}/(N-I)}$

Άρα $F_L = \dfrac{MS_L}{MS_{res}} \sim F_{1,N-1}$ υπό $H_0: L=0$

Μεγάλες τιμές του F_L, συνηγορούν στην απόρριψη της $H_0: L=0$

Άρα η κ.π έχει τη μορφή $F_L \geq c$

Το κρίσιμο σημείο c υπολογίζεται:

$\alpha = P(\text{Απορρ } H_0 / H_0: \text{αληθής}) = P(F_L \geq c / F_1 \sim F_{1,N-1})$

$\Rightarrow \dots \Rightarrow c = F_{1,N-1,\alpha}$.

<u>Συγκεντρωτικά</u>: Για τον έλεγχο $H_0: L=0$ η ΣΣΤ είναι $F_L = \dfrac{MS_L}{MS_{res}}$

με κατανομή $F_{L,N-1}$ υπό H_0 και κ.π $F_L \geq F_{1,N-1,\alpha}$.

<u>Ανάλυση Υπολοίπων</u>: Εννοούμε αξιοποίηση των υπολοίπων για τον έλεγχο της ισχύος των υποθέσεων για τα σφάλματα.

$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{\alpha}_i) = Y_{ij} - (\bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..})$

$\Rightarrow e_{ij} = Y_{ij} - \bar{Y}_{i.} \qquad i = 1, \dots, I$
$\qquad\qquad\qquad\qquad\qquad j = 1, \dots, J_i$

Ισοδυναμία Ανάλυσης Διακύμανσης κατά 1 παράγοντα και γραμμ. παλινδρόμησης (όχι τόσο σημαντικό)

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

Αν θεωρήσω: $\underset{\sim}{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1J_1} \\ Y_{21} \\ \vdots \\ Y_{IJ_I} \end{pmatrix}_{N\times 1}$ ορίζω $X = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{N\times I}$

$\underset{\sim}{\beta} = \begin{pmatrix} \mu+\alpha_1 \\ \mu+\alpha_2 \\ \vdots \\ \mu+\alpha_I \end{pmatrix}_{N\times 1}$ $\qquad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{1J_1} \\ \vdots \\ \varepsilon_{I1} \\ \vdots \\ \varepsilon_{IJ_I} \end{pmatrix}_{N\times 1}$

Μοντέλο Ανάλυσης διακύμανσης κατά δύο παράγοντες

Y ← Εξαρτάται από δύο παράγοντες

↑ Ποσοτική

A: $\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_I$

B: $\beta_1 \quad \beta_2 \quad \cdots \quad \beta_J$

πχ Y < Επίπεδο μόρφωσης πατέρα
Επίπεδο μόρφωσης μητέρας

Π: Υποχρ (α_1), Λύκειο (α_2), Πανεπ (α_3), Μεταπτ (α_4)

Μ: (β_1), (β_2), (β_3), (β_4)

Μορφή Δεδομένων

Παράγοντας A	Παράγοντας B β_1	β_2	$\cdots$	β_J		
α_1	Y_{11}	Y_{12}	$\cdots$	Y_{1J}	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$
α_2	Y_{21}	Y_{22}	$\cdots$	Y_{2J}	$Y_{2\cdot}$	$\bar{Y}_{2\cdot}$
$\vdots$	$\vdots$					
α_I	Y_{I1}	Y_{I2}	$\cdots$	Y_{IJ}	$Y_{I\cdot}$	$\bar{Y}_{I\cdot}$
	$Y_{\cdot 1}$	$Y_{\cdot 2}$	$\cdots$	$Y_{\cdot J}$	$Y_{\cdot\cdot}$	
	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$	$\cdots$	$\bar{Y}_{\cdot J}$		$\bar{Y}_{\cdot\cdot}$

Y_{ij} : παρατήρηση που έχει συλλεγεί στο i-επίπεδο του A και στο j-επίπεδο του B.

$$Y_{i\cdot} = \sum_{j=1}^{J} Y_{ij}$$

$$\bar{Y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij} = \frac{1}{J} Y_{i\cdot}$$

$$Y_{\cdot j} = \sum_{i=1}^{I} Y_{ij}$$

$$\bar{Y}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} Y_{ij} = \frac{1}{I} Y_{\cdot j}$$

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

$$\bar{Y}_{\cdot\cdot} = \frac{1}{IJ} Y_{\cdot\cdot} = \frac{1}{IJ}\sum^{I}\sum Y_{ij} = \frac{1}{I}\sum_{i=1}^{I} Y_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$$

Έτσι καταλήγω στο μοντέλο αναλ. διακ. κατά δυο παράγοντες

Κοινή Επίδραση των ΑΚΒσταηνΥ

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

σφάλματα

η ij παρατήρηση της Y.

ατομική Επίδραση των Επιπέδων του A

ατομική Επίδραση των Επιπέδων του B

ΥΠΟΘΕΣΕΙΣ ΓΙΑ ΤΑ ΣΦΑΛΜΑΤΑ ΤΟΥ ΜΟΝΤΕΛΟΥ

1. $E(\varepsilon_{ij}) = 0$
2. $Var(\varepsilon_{ij}) = \sigma^2$
3. ε_{ij} : ασυσχέτιστα
4. $\varepsilon_{ij} \sim N(0,\sigma^2)$ $\quad i = 1,\dots I \quad j = 1,\dots,J$

Συνοπτικά

$\varepsilon_{ij} \sim N(0,\sigma^2)$ και είναι ασυσχ (ανεξάρτητα)

$\xrightarrow[\text{για τα } Y]{\text{Συνέπειες}}$

1. $E(Y_{ij}) = \mu + \alpha_i + \beta_j$
2. $Var(Y_{ij}) = \sigma^2$
3. Y_{ij} ασυσχ.
4. $Y_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ $\quad i = 1,\dots,I \quad j = 1,\dots,J$

$$\underbrace{\qquad\qquad}_{Y_{ij} \sim N(\mu+\alpha_i+\beta_j,\ \sigma^2)\ \text{ασυσχέτιστα}}$$

ΕΕΤ των παραμέτρων α_i, β_j, μ:

Ελαχιστοποιώντας το $S = \sum_i \sum_j \varepsilon_{ij}^2$

$$= \sum_i \sum_j (Y_{ij} - \mu - \alpha_i - \beta_j)^2$$

$$\frac{\partial S}{\partial \mu} = 0 \quad \underbrace{\frac{\partial S}{\partial \alpha_i} = 0 \quad \frac{\partial S}{\partial \beta_j} = 0}$$

Προκύπταν οι κανονικές εξισώσεις

:

$$\boxed{IJ\mu + J\sum_{i=1}^{I} \alpha_i + I\sum_{j=1}^{J} \beta_j = Y_{..}}$$

$$\boxed{J\mu + J\alpha_i + \sum_{j=1}^{J} \beta_j = Y_{i.} \quad i = 1, .., I}$$

$$\boxed{I\mu + \sum_{i=1}^{I} \alpha_i + I\beta_i = Y_{.j} \quad j = 1, .., J}$$

Υποθέτω ότι οι πλευρικές συνθήκες που οδηγούν σε εκτιμητή για την μ το $\bar{Y}_{..}$ τις εξής:

$$\sum_{i=1}^{I} \alpha_i = 0 \quad \text{και} \quad \sum_{j=1}^{J} \beta_j = 0$$

Υπό τις 2 παραπάνω πλευρικές συνθήκες, οι ΕΕΤ

είναι $\hat{\mu} = \bar{Y}_{..}$, $\hat{\alpha_i} = \bar{Y}_{i.} - \bar{Y}_{..}$ $\quad i = 1, .., I$

και $\hat{\beta_j} = \bar{Y}_{.j} - \bar{Y}_{..}$ $\quad j = 1, ..., J$

ΙΔΙΟΤΗΤΕΣ

1 Αν οι υποθέσεις για τα σφάλματα ικανοποιούνται, τότε

$E(\hat{\mu}) = \mu$, $E(\hat{\alpha_i}) = \alpha_i$, $E(\hat{\beta_j}) = \beta_j$ $\quad i = 1, .., I$, $j = 1, ..., J$

Αποδ

Θδο $E(\hat{\alpha_i}) = \alpha_i$ $\quad i = 1, .., I$ (ομοίως για τα άλλα)

$$E(\hat{\alpha_i}) = E(\bar{Y}_{i.} - \bar{Y}_{..}) = E(\bar{Y}_{i.}) - E(\bar{Y}_{..})$$

$$E(\bar{Y}_{i.}) = \frac{1}{J} E\left(\sum_{j=1}^{J} Y_{ij}\right) = \frac{1}{J} \sum_{j=1}^{J} E(Y_{ij}) = \frac{1}{J} \sum_{j=1}^{J} E(\mu + \alpha_i + \beta_j + \varepsilon_{ij}) \overset{E(\varepsilon_{ij})=0}{=} \frac{1}{J} \Sigma(\mu + \alpha_i + \beta_j)$$

$$= \frac{1}{J}\left(J\mu + J\alpha_i + \sum_{j=1}^{J} \beta_j\right) = \mu + \alpha_i \quad (1)$$

$$E(\bar{y}_{..}) = E\left(\frac{1}{IJ}\sum\sum y_{ij}\right)$$

$$= \frac{1}{JI}\sum\sum E(\mu+\alpha_i+\beta_j+\varepsilon_{ij}) = \frac{1}{JI}\sum\sum(\mu+\alpha_i+\beta_j)$$

$$= \frac{1}{JI}\left(\sum\sum\mu + \sum_j \cancel{\left(\sum\alpha_i\right)}^{0} + \sum_i \cancel{\left(\sum\beta_j\right)}^{0}\right)$$

$$= \frac{1}{JI}\sum\sum\mu = \frac{1}{JI} IJ\mu \Longrightarrow E(\bar{y}_{..}) = \mu \quad (2)$$

Από (1) κ (2)

$$E(\hat{\alpha}_i) = E(\bar{y}_{i.}) - E(\bar{y}_{..})$$

$$= \mu+\alpha_i-\mu = \alpha_i \quad \forall i=1,..,I.$$