Μάθημα 8ο 27/11/17

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

Υποθέσεις : $\varepsilon_{ij} \sim N(0, \sigma^2)$, ασυσχέτιστα.

ΘΕΩΡΗΜΑ: Υπό τις υποθέσεις για τα σφάλματα:

α) $E(MS_{res}) = \sigma^2$, $\left(MS_{res} = \frac{1}{N-I} \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i.})^2 \right)$

β) $E(MS_{tr}) = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^{I} J_i a_i^2$, $\left(MS_{tr} = \frac{1}{I-1} \sum_{i=1}^{I} J_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 \right)$

ΑΠΟΔΕΙΞΗ:

α) Αν $W_1, \dots, W_n$ τ.δ. από πληθυσμό με διακύμανση σ^2.
Τότε $E(S_W^2) = \sigma^2$, $S_W^2 = \frac{1}{n-1} \sum_{i=1}^{n} (W_i - \overline{W})^2$ δειγματική διακύμανση.

Επειδή $Y_{i1}, \ldots, Y_{iJ_i}$ είναι τ.δ. από πληθυσμό με διακύμανση σ^2

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

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

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

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

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

$$=E\left[\sum_i J_i\left(\frac{Y_{i.}^2}{J_i^2}+\frac{Y_{..}^2}{N^2}-\frac{2Y_{i.}}{J_i}\frac{Y_{..}}{N}\right)\right]=$$

$$=E\left[\sum_i\frac{Y_{i.}^2}{J_i}+\frac{Y_{..}^2}{N}-\frac{2Y_{..}}{N}\sum_i Y_{i.}\right]=$$

$$=E\left[\sum_i\frac{Y_{i.}}{J_i}+\frac{Y_{..}^2}{N}-\frac{2Y_{..}^2}{N}\right]=$$

$$=E\left[\sum_{i=1}^{I}\frac{Y_{i.}^2}{J_i}-\frac{Y_{..}^2}{N}\right]=$$

$$=\sum_{i=1}^{I}\frac{1}{J_i}E(Y_{i.}^2)-\frac{1}{N}E(Y_{..}^2)$$

$$\xRightarrow{Var(W)=EW^2-(EW)^2}$$

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

Από το μοντέλο $Y_{ij}=\mu+a_i+\varepsilon_{ij}$ και από πλευρική συνθήκη $\sum_{i=1}^{I}J_ia_i=0$

$Y_{i.}=J_i\mu+J_ia_i+\varepsilon_{i.}$, $Y_{..}=N\mu+\varepsilon_{..}$ άρα έχω:

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

$E(Y_{..})=N\mu$, $Var(Y_{..})=Var(\varepsilon_{..})=N\sigma^2$ (3)

Από τις (1), (2), (3) και πράξεις έχω: $E(SStr) = (I-1)\sigma^2 + \sum_{i=1}^{I} J_i a_i^2$

Τελικά, $E(MStr) = \frac{1}{I-1} E(SStr)$

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

Έλεγχος της υπόθεσης $H_0: a_1 = a_2 = \cdots = a_I$

έναντι $H_a: a_\iota = a_i$ για τουλάχιστον ένα ζεύγος $\iota, i = 1, \dots, I$, με $\iota \neq i$

<u>Πρακτική αξία</u>: Αν η H_0 δεν μπορεί να απορριφθεί $\Rightarrow$ όλα τα επίπεδα του παράγοντα ασκούν την ίδια επίδραση στην Y κάτι το οποίο σημαίνει ότι το μοντέλο δεν μπορεί να διευρυνθεί περαιτέρω.
Αν η H_0 απορριφθεί, αυτό σημαίνει ότι κάποιο(-α) από τα επίπεδα του παράγοντα ασκούν σημαντικότερη επίδραση στην Y.
Άρα έχει νόημα να αναζητήσω ποιό(-α) επίπεδο(-α) ασκούν σημαντικότερη επίδραση και ακόμη να τα κατατάξω από το λιγότερο σημαντικό προς το περισσότερο σημαντικό (πολλαπλές συγκρίσεις).

Από Μαθηματική άποψη μας εξυπηρετεί να θεωρήσουμε $H_0^*: a_1 = \cdots = a_I = 0$.

Αν H_0 αληθής και $a_1 = a_2 = \cdots = a_I = a \rightsquigarrow Y_{ij} = \mu + a + \varepsilon_{ij} =$

Αν H_0^* αληθής και $a_1 = a_2 = \cdots = a_I = 0 \rightsquigarrow Y_{ij} = \mu + \varepsilon_{ij}$

ΕΡΩΤΗΜΑ: Που θα βασιστώ για να κατασκευάσω τεστ για έλεγχο της H_0 ή της H_0^*;

Είδαμε: $E(MSres) = \sigma^2$ και $E(MStr) = \sigma^2 + \frac{1}{I-1}\sum_{i=1}^{I} J_i a_i^2$

Αν H_0 ή H_0^* αληθής τότε $E(MSres) \equiv E(MStr)$ ή $MSres \approx MStr$
Ξέρω αν $A \Rightarrow B$ τότε $\sim B \Rightarrow \sim A$. $\Bigg\} \Rightarrow$

$\Rightarrow$ Αν το MSres πολύ διαφορετικό από MStr $\Rightarrow$ η H_0 ή H_0^* πρέπει να απορριφθεί

Ένα τεστ πρέπει να βασίζεται στη σύγκριση του $MStr$ με το $MSres$ ή στο F-πηλίκο που ορίζεται $F = \frac{MStr}{MSres}$

ΘΕΩΡΗΜΑ Έστω ότι οι υποθέσεις για τα σφάλματα ισχύουν, δηλ. $\varepsilon_{ij} \sim N(0, \sigma^2)$ και είναι ασυσχέτιστα. Τότε:

α) $\frac{SSres}{\sigma^2} \sim \chi^2_{N-I}$ β) Υπό την $H_0^*: a_1 = \dots = a_I = 0$, $\frac{SStr}{\sigma^2} \sim \chi^2_{I-1}$

ΑΠΟΔΕΙΞΗ: $SSres = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i.})^2$.

α)

Αν $W_1, \dots, W_n$ τ.δ. από $N(\mu, \sigma^2)$ τότε $\frac{(n-1)S_W^2}{\sigma^2} \sim \chi^2_{n-1}$

Επειδή $Y_{i1}, \dots, Y_{iJ_i}$ τ.δ. από $N(\mu + a_i, \sigma^2)$ το $\frac{\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i.})^2}{\sigma^2} \sim \chi^2_{J_i - 1}$, $i = 1, I$

Επιπλέον, επειδή Y_{ij} ασυσχέτιστα και κανονικές οι Y_{ij} είναι και ανεξάρτητες.

Άρα $\sum_{i=1}^{I} \frac{\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i.})^2}{\sigma^2} \sim \chi_{\sum_{i=1}^{I}(J_i - 1)}$ δηλαδή $\frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_{i.})^2}{\sigma^2} \sim \chi^2_{\sum_{i=1}^{I} J_i - \sum_{i=1}^{I} 1}$

$\Rightarrow \frac{SSres}{\sigma^2} \sim \chi^2_{N-I}$

β) $SStr = \sum_{i=1}^{I} J_i (\overline{Y}_{i.} - \overline{Y}_{..})^2$

Αν $W_1, \dots, W_n$ τ.δ. από $N(\mu, \sigma^2)$ τότε $\overline{W} \sim N(\mu, \sigma^2/n)$

Επειδή $Y_{i1}, \dots, Y_{iJ_i}$ τ.δ. από $N(\mu + a_i, \sigma^2)$. το $\overline{Y}_{i.} \sim N\left(\mu + a_i, \frac{\sigma^2}{J_i}\right)$ $i = 1, \dots, I$.

και επειδή υποθέτω την H_0^* το $\overline{Y}_{i.} \sim N(\mu, \sigma^2/J_i)$

$\Rightarrow \frac{\overline{Y}_{i.} - \mu}{\sigma/\sqrt{J_i}} \sim N(0,1)$ ή $\frac{\sqrt{J_i}(\overline{Y}_{i.} - \mu)}{\sigma} \sim N(0,1)$

$\Rightarrow \frac{J_i (\overline{Y}_{i.} - \mu)^2}{\sigma^2} \sim N(0,1)^2 \equiv \chi_1^2$

$$\overset{\text{ανεξ.}}{\underset{Y_{ij}}{\Longrightarrow}} \sum_{i=1}^{I} \frac{J_i(\bar{Y}_{i.}-\mu)^2}{\sigma^2} \sim \chi^2_{\sum_{i=1}^{I} 1} \equiv \chi^2_I$$

$$\Downarrow$$

$$\frac{\sum_{i=1}^{I} J_i(\bar{Y}_{i.}-\mu)^2}{\sigma^2} \sim \chi^2_I \Rightarrow \frac{\sum^{I} J_i(Y_{i.}-\bar{Y}_{..})^2}{\sigma^2} \sim \chi^2_{I-1}$$

$$\Rightarrow \frac{SStr}{\sigma^2} \sim \chi^2_{I-1}$$

Επιστρέφω στον έλεγχο $H_0: a_1=a_2=\cdots=a_I$ ή $H_0^*: a_1=\cdots=a_I=0$.

Η H_0 ή H_0^* απορρίπτεται αν το $F=\frac{MStr}{MSres}$ παίρνει μεγάλες τιμές. Επομένως

η κ.π. θα έχει τη μορφή $F \geq C$

Για να υπολογίσω την C χρειάζομαι κατανομή F υπό την H_0 ή H_0^*.

$$F=\frac{MStr}{MSres}=\frac{SStr/(I-1)}{SSres/(N-I)}=\frac{SStr/\sigma^2(I-1)}{SSres/\sigma^2(N-I)}$$

$$\Rightarrow F \equiv \frac{\chi^2_{I-1}/(I-1)}{\chi^2_{N-I}/(N-I)} \underset{\text{SStr ανεξ SSres}}{\overset{\text{ισχύει}}{\sim}} F_{I-1,N-I} \text{ υπό την } H_0^*$$

Προσδιορισμός της C:

$\alpha=P(\text{Απορ. } H_0 \mid H_0 \text{ αληθ.})=P(F\geq C \mid F\sim F_{I-1,N-I}) \Rightarrow C=F_{I-1,N-I,a}$

Συμπερασματικά: Για τον έλεγχο της $H_0: a_1=\cdots=a_I$ χρησιμοποιείται η ΣΣΤ $F=\frac{MStr}{MSres}$ με κατανομή $F_{I-1,N-I}$ υπό H_0 και κ.π. $F\geq F_{I-1,N-I,a}$

Στην πράξη αναμένουμε η $H_0: a_1=\cdots=a_I$ να απορριφθεί. Τότε έχει αξία το μοντέλο γιατί τότε σημαίνει ότι $\exists$ επίπεδα του παράγοντα που ασκούν σημαντικότερη επίδραση στην εξαρτημένη μεταβλητή Y.

Ποιό(-α) επίπεδο(-α) ? $\xrightarrow{\text{Απάντηση}}$

Πολλαπλές συγκρίσεις

- Ελάχιστης Σημαντικής Διαφοράς (ΕΣΔ) Fisher ←
- Μέθοδος Tukey
- Bonferron
- Μέθοδος Γραμμικών Αντιθέσεων Scheffe ←

Μέθοδος Ελάχιστης Σημαντικής Διαφοράς (ΕΣΔ)

Εφαρμόζεται αν απορριφθεί η $H_0: a_1=\cdots=a_I$ οπότε υιοθετείται η $H_a: a_\iota\neq a_i$ για $\iota, i=1,\dots,I$ με $i\neq\iota$

Άρα έχει νόημα να συγκρίνω ανά δύο τις επιδράσεις a_i, a_ι των επιπέδων του παράγοντα.

Δηλ. έχει νόημα να ελέγξω $H_0^A: a_i=a_\iota$, $\iota, i=1,\dots,I$, $\iota\neq i$

Οπότε πρέπει να κατασκευάσω ένα τεστ για τον έλεγχο της $H_0: a_i=a_\iota$ όπου και στηρίζομαι σε έναν εκτιμητή που εμφανίζεται στην H_0, (κατά Wald θεώρηση)

Ξέρω ότι: $\left.\begin{array}{l}\hat{a}_i=\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot}\\ \text{ενώ } \hat{a}_\iota=\bar{Y}_{\iota\cdot}-\bar{Y}_{\cdot\cdot}\end{array}\right\}\Rightarrow \hat{a}_i-\hat{a}_\iota=\bar{Y}_{i\cdot}-\bar{Y}_{\iota\cdot}$

Θα στηριχτώ στη διαφορά $\bar{Y}_{i\cdot}-\bar{Y}_{\iota\cdot}$ και αρκεί να βρω την κατανομή της υπό την μηδενική υπόθεση.

Κατανομή

Αφού $Y_{i1},\dots,Y_{iJ_i}$ τ.δ. από κατανομή $\sim N(\mu+a_i,\sigma^2)$, το $\bar{Y}_{i\cdot}\sim N\left(\mu+a_i,\frac{\sigma^2}{J_i}\right)$

Επίσης $Y_{\iota1},\dots,Y_{\iota J_\iota}$ τ.δ. από κατανομή $\sim N(\mu+a_\iota,\sigma^2)$, το $\bar{Y}_{\iota\cdot}\sim N\left(\mu+a_\iota,\frac{\sigma^2}{J_\iota}\right)$ ανεξάρτητα

$$\Rightarrow \bar{Y}_{i\cdot}-\bar{Y}_{\iota\cdot}\sim N\left(a_i-a_\iota,\frac{\sigma^2}{J_i}+\frac{\sigma^2}{J_\iota}\right) \text{ από μηδενική υπόθεση } \equiv N\left(0,\frac{\sigma^2}{J_i}+\frac{\sigma^2}{J_\iota}\right)$$

$$\Rightarrow \frac{\bar{Y}_{i\cdot}-\bar{Y}_{\iota\cdot}}{\sigma\sqrt{\frac{1}{J_i}+\frac{1}{J_\iota}}}\sim N(0,1) \text{ υπό } H_0: a_i=a_\iota$$

Θυμάμαι ότι $\frac{SSres}{\sigma^2}\sim X^2_{N-I}$ οπότε διαιρώντας :

$$\frac{\dfrac{\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}}{\sigma\sqrt{1/J_i + 1/J_l}}}{\sqrt{\left(\frac{SSres}{\sigma^2}\right)/(N-I)}} = \frac{\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}}{\sqrt{MSres}\sqrt{1/J_i + 1/J_l}} \sim t_{N-I} \text{ υπό την } H_0$$

Για τον έλεγχο της $H_0: a_i = a_l$ η στατιστική συνάρτηση του τεστ είναι η:

$$\frac{\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}}{\sqrt{MSres}\sqrt{1/J_i + \frac{1}{J_l}}} \sim t_{N-I}$$ υπό την H_0 και κρίσιμη περιοχή μεγάλες τιμές του πηλίκου $|t| > C$ δηλ. $\left|\frac{\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}}{\sqrt{MSres}\sqrt{1/J_i + 1/J_l}}\right| \geq C = \cdots = \cdots t_{N-I, a/2}$

οπότε $|\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}| \geq t_{N-I, \frac{a}{2}} \sqrt{MSres}\sqrt{\frac{1}{J_i} + \frac{1}{J_l}}$

Πολλαπλές συγκρίσεις

$$E.\Sigma.\Delta = t_{N-I, \frac{a}{2}} \sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_l}\right)}$$

Η μέθοδος εφαρμόζεται ως εξής:

Αν $|\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}| > E\Sigma\Delta$ τότε H_0 απορρίπτεται.

Αν η $H_0: a_i = a_l$ απορριφθεί αυτό σημαίνει ότι κάποιο από τα επίπεδα i και l ασκεί σημαντικότερη επίδραση στη Y.

- Αν $(\bar{Y}_{i\cdot} - \bar{Y}_{l\cdot}) > 0$ το i-επίπεδο ασκεί σημαντικότερη επίδραση από το l, αλλιώς συμβαίνει το αντίθετο (κρίνω με βάση το πρόσημο)

ΠΑΡΑΔΕΙΓΜΑ: 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 διαλέξει μεταξύ 3-μεθόδων παραγωγής. Τα παρακάτω δεδομένα δίνουν τον χρόνο σε δευτερο που φωσφορίζουν τα εν λόγω όργανα για κάθε μια από τις 3-μεθόδος.

Να αναλυθούν τα δεδομένα αυτά.

ΜΕΘΟΔΟΣ 1	59.2	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: Έστω Y ο χρόνος που φωσφορίζουν και 3-μέθοδοι παραγωγής που πιθανώς επηρεάζουν το χρόνο Y.

$Y \longrightarrow$ Μέθοδοι παραγωγής παράγοντας $<$ Μέθοδος 1, Μέθοδος 2, Μέθοδος 3.

Μοντέλο: $Y_{ij} = \mu + a_i + \varepsilon_{ij}$ $\quad i=1,2,3$, $J=1,\dots,J_i$ (οι παρατηρήσεις που έχω σε κάθε επίπεδο)
$J_1 = J_2 = J_3 = 8$.

Υποθέτω λοιπόν ότι ισχύουν οι υποθέσεις για τα σφάλματα.
a_i παριστά την κύρια επίδραση της i-μεθόδου στο χρόνο Y.

ΠΙΝΑΚΑΣ ΑΝΑΝΙΑ.

μεταβλητότητα	Α.Τ	Β.Ε	Μ.Τ	F-πηλίκο
μοντέλο	$SS_{tr} = 584.41$	$I-1=2$	$MStr = 292.205$	$F = 17.044$
υπόλοιπα	$SSres = 360.015$	$N-I=21$	$MSres = 17.144$	
ολικά	$SStot = 944.425$	$N-1=23$		

Υπόθεση $H_0: a_1 = a_2 = a_3$ (υπόθεση ισοεπίδρασης επιπέδων, δηλαδή υπόθεση ισοδυναμίας των τριών μεθόδων παραγωγής)

Επειδή $F = 17.044 \geqslant F_{I-1,N-I,a} = F_{2,21} \begin{cases} \xrightarrow{a=0.01} = 4.32 \\ \xrightarrow{a=0.05} = 5.78 \end{cases}$

και επομένως απορρίπτω τη μηδενική υπόθεση!!
Άρα, τώρα εφαρμόζω ΕΣΔ: $\overline{Y}_{1.} = 57.1$, $\overline{Y}_{2.} = 59.175$, $\overline{Y}_{3.} = 68.45$ οπότε
$ΕΣΔ = t_{N-I,\frac{a}{2}} \sqrt{MSres\left(\frac{1}{J_i} + \frac{1}{J_i}\right)} \overset{a=0.05}{=\!=} 2.08 \cdot \sqrt{17.144\left(2 \cdot \frac{1}{8}\right)} = 4.306.$

Οπότε παίρνω τις διαφορές των μέσων:
1) $(\overline{Y}_{1.} - \overline{Y}_{2.}) = 2.075$
2) $(\overline{Y}_{1.} - \overline{Y}_{3.}) = -11.35$
3) $(\overline{Y}_{2.} - \overline{Y}_{3.}) = -9.275$

Επειδή λοιπόν $|\overline{Y}_{1.} - \overline{Y}_{2.}| < 4.306$ άρα $2=1$
$|\overline{Y}_{2.} - \overline{Y}_{3.}| = 9.275 > 4.306$ και άρα $1 \neq 3$ (απορ. H_0 και καλύτερη είναι η 3 από πρόσημο)
$|\overline{Y}_{1.} - \overline{Y}_{3.}| = 11.35 > 4.306$ και άρα $2 \neq 3$ (καλύτερη η 3 από πρόσημο)

Τελικά $1=2<3$.