(45) 7 | 5

(α) Δεδομένα ΑΒΕΖ

Καταστήματα γραφείων	1	2	3	4	Σύνολο
1	12	14	19	24	
2	18	12	17	30	
3		13	21		
Σύνολο	30	39	57	54	180
Δειγμ. Μέσοι	15	13	19	27	18
Αριθμ. Καταστ.	2	3	3	2	10

(β) Συμβολισμός

	Δοκιμασίες (j)				
Δειγματική	1	2	3	4	Σύνολο
1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	
2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	
3		Y_{32}	Y_{33}		
Σύνολο	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	$Y_{..}$
Δειγμ. Μέσοι	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{.3}$	$\bar{Y}_{.4}$	$\bar{Y}_{..}$
Μέγεθος δείγματος	n_1	n_2	n_3	n_4	n

όπου Y_{ij}: i-οστή παρατ. μονάδες για τη j δοκιμασία

$n=\sum_j \sum_i n_{ij}$, $Y_{.j}=\sum_{i=1}^{n_j} Y_{ij}$, $\bar{Y}_{.j}=\frac{Y_{.j}}{n_j}$, $Y_{..}=\sum_j \sum_i Y_{ij}$, $\bar{Y}_{..}=\frac{Y_{..}}{n}$

Ανάλυση της διακύμανσης (κατά ένα παράγοντα)

Το στατιστικό μοντέλο

$Y_{ij}=\mu_j+\varepsilon_{ij}$ με $i=1,\dots,n_j$ και $j=1,\dots,r$ (r-δοκιμασίες) $n=\sum_{j=1}^{r} n_j$

όπου τα ε_{ij} θα τα ονομάζουμε τυχαία σφάλματα με $\varepsilon_{ij}\sim N(0,\sigma^2)$

άρα $Y_{ij}\sim N(\mu_j,\sigma^2)$, $\varepsilon_{ij}=Y_{ij}-\mu_j$ (άγνωστη)

Έχουμε $H_0:\mu_1=\dots=\mu_r$ έναντι H_a: όχι όλα τα μ_j ίσα μεταξύ τους

(Γιατί όχι αλλεπάλληλα t-τεστ)

Έστω $r=4$ δοκιμασίες δηλ. $\binom{4}{2}=6$ συγκρίσεις

$H_{0_1}:\mu_1=\mu_2,\dots,H_{0_6}:\mu_3=\mu_4$ Έστω ανεξ. $\alpha=0.05$

τότε το $\alpha'=P$(απορρ. H_0 | H_0 αληθεύει) =

= P(τουλάχ. μία από τις H_{0_i} απορρ. | H_0 αληθεύει) =

= P(τουλάχ. επιτυχ. σε 6 δοκιμές με $p=\alpha$) =

= 1 − P(καμία επιτυχ. σε $n=6$ δοκιμές με $p=\alpha$) =

$=1-\binom{6}{0}\alpha^0(1-\alpha)^6=0{,}265>\alpha$

μέσω διωνυμικής κατανομής $B(\alpha,6)$

Εκτιμητές Ελαχίστων Τετραγώνων

$$Q(\mu_1,\dots,\mu_r)=\sum_{j=1}^{r}\sum_{i=1}^{n_j}\varepsilon_{ij}^2=\sum_{j=1}^{r}\sum_{i=1}^{n_j}(y_{ij}-\mu_j)^2 \text{ (ελαχιστοποιούμε το } \mu_j)$$

$$\Rightarrow Q(\mu_1,\dots,\mu_r)=\sum_{i=1}^{n_1}(y_{i1}-\mu_1)^2+\dots+\sum_{i=1}^{n_r}(\mu_{ir}-\mu_r)^2$$

Έστω $Q_j(\mu_j)=\sum_{i=1}^{n}(y_{ij}-\mu_j)^2$, $j=1,\dots,r$

$$\frac{\partial Q_j}{\partial \mu_j}=-2\sum_{i=1}^{n_j}(y_{ij}-\mu_j)=0 \Rightarrow \sum_{i=1}^{n}y_{ij}=\sum_{i=1}^{n_j}\hat{\mu}_i$$

με $\bar{Y}_{.j}$ έχω τον μέσο όρο που αντιστοιχεί στην j-δοκιμασία

$\bar{Y}_{..}\rightarrow$ συνολικό μέσο Άθροισμα

$\bar{Y}_{.1}\rightarrow$ έχω αθροίσει ως ε_j της παρατήρησης που αντιστοιχεί στην 1 δοκιμασία

άρα $\hat{\mu}_{.j}=\bar{Y}_{.j}$, $j=1,\dots,r$

Η ανάλυση της διακύμανσης

$$y_{ij}-\bar{y}_{..}=\underbrace{\bar{y}_{.j}-\bar{y}_{..}}_{\text{γενική μέση τιμή}}+\underbrace{y_{ij}-y_{.j}}_{\text{Υπόλοιπα}}$$

έχω $\varepsilon_{ij}=y_{ij}-\mu_j$, $e_{ij}=y_{ij}-\hat{\mu}_j$:υπόλοιπα

$$\Rightarrow \underbrace{\sum_{j=1}^{r}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_{..})^2}_{SS_{tot}}=\underbrace{\sum_j\sum_i(\bar{y}_{.j}-\bar{y}_{..})^2}_{SS_{tr}}+\underbrace{\sum_j\sum_i(y_{ij}-\bar{y}_{.j})^2}_{SS_{res}}+\cancel{2\sum_j\sum_i(\bar{y}_{.j}-\bar{y}_{..})(\bar{y}_{ij}-\bar{y}_{.j})}\ 0$$

γιατί $\sum_j\sum_i(\bar{y}_{.j}-y_{..})(y_{ij}-y_{.j})=\sum_j(\bar{y}_{.j}-\bar{y}_{..})\sum_i(\bar{y}_{ij}-\bar{y}_{.j})$

$$=\sum_j(\bar{y}_{ij}-\bar{y}_{..})\underbrace{(y_{.j}-n_j\bar{y}_{.j})}_{y_{ij}-y_{.j}}=0 \quad \text{όπου}$$

SS_{tot} : Συνολικό Άθροισμα Τετραγώνων

SS_{tr} : Άθροισμα Τετρ. των Δοκιμασιών

SS_{res} : Άθροισμα Υπολοίπων άρα

$$SS_{tot}=SS_{tr}+SS_{res}$$

$$\text{β.ε.: } n-1=(r-1)+(n-r)$$

$\sum_j\sum_i(y_{ij}-\bar{y}_{..})=0$ $\quad\sum_j n_j(y_{.j}-y_{..})=0$ $\quad\sum_i(y_{ij}-y_{.j})=0$, $j=1,\dots,r$

$$SS_{tot}=\sum_j\sum_i y_{ij}^2-\frac{y_{..}^2}{n}$$

$$SS_{tr}=\sum_j\frac{y_{.j}^2}{n_j}-\frac{y_{..}^2}{n}$$

$$SS_{res}=\sum_j\sum_i y_{ij}^2-\sum_j\frac{y_{.j}}{n_j}=SS_{tot}-SS_{tr}$$

ΑΝΑΔΙΑ (Πίνακας Ανάλυσης της Διακύμανσης)

Προέλευση Μεταβλητότητας	Αθροίσματα Τετραγώνων	Βαθμ. Ελευθερίας	Μέσα Τετράγωνα
Δοκιμασίες ή Μεταξύ των Δοκιμασιών	$SS_{tr} = \sum_{j=1}^{r} n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$	$r-1$	$MS_{tr} = \frac{SS_{tr}}{r-1}$
Υπόλοιπα Δοκιμασιών	$SS_{res} = \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2$	$n-r$	$MS_{res} = \frac{SS_{res}}{n-r}$
Ολική Μεταβλητότητα	$SS_{tot} = \sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2$	$n-1$	

$\Rightarrow E(MS_{res}) = \sigma^2$ (πάντα) ενώ

$$E(MS_{tr}) = \sigma^2 + \frac{1}{r-1} \sum_j n_j (\mu_j - \mu)^2$$

Αν $F = \frac{MS_{tr}}{MS_{res}}$ τότε: $\frac{SS_{res}}{\sigma^2} \sim X^2_{n-r}$ και $\frac{SS_{tr}}{\sigma^2} \sim X^2_{r-1}$

Άρα $F = \frac{MS_{tr}}{MS_{res}} \overset{H_0}{\sim} F_{r-1,n-r}$ με

Κρίσιμη περιοχή: $F \geq F_{\alpha, r-1, n-r}$