ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Στέλιος Ζήμερας
Αναπληρωτής Καθηγητής
Τμήμα Στατιστικής και Αναλογιστικών-Χρηματοοοικονομικών Μαθηματικών

Σαμος

2020

Ανάλυση με δυο παράγοντες

- Στην ανάλυση με έναν παράγοντα ένας παράγοντας(κριτήριο) (factor) είναι η"επίδραση" (treatment) που αντιστοιχεί στις στήλες του σχεδιασμού ενώ ο άλλος παράγοντας(κριτήριο) είναι τα μπλοκ που αντιστοιχούν στις γραμμές του σχεδιασμού.
- Η υπόθεση που θέλουμε να ελέγξουμε είναι η

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

H_1 : Δύο, τουλάχιστον, από τους μέσους διαφέρουν

Ανάλυση με δυο παράγοντες

- Η ανάλυση διακύμανσης κατά δύο κριτήρια που θα αναπτύξουμε είναι κατάλληλη όταν για το πρόβλημα που εξετάζουμε είναι δυνατόν να υποθέσουμε ότι δεν υφίσταται αλληλεπίδραση(interaction) μεταξύ των"μπλοκ" και των "επιδράσεων" που χρησιμοποιούνται στο πείραμα.

Ανάλυση με δυο παράγοντες

- Αν Y_{ij} είναι η τιμή της εξαρτημένης μεταβλητής όταν η i "επίδραση" εφαρμοσθεί στο j "μπλοκ" θα έχουμε

$$Y_{ij} = \mu_i + \beta_j + \varepsilon_{ij}$$

- Υποθέτουμε ότι

$$\mu_i = \mu + \alpha_i$$

όπου μ_i είναι ο μέσος στον πληθυσμό που αντιστοιχεί στην i επίδραση και μ ο συνολικός μέσος θα έχουμε

$$\boxed{Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}}$$

Ανάλυση με δυο παράγοντες

όπου $\mu = \dfrac{\sum \mu_i}{k}$, α_i είναι το αποτέλεσμα της i "επίδρασης" και β_j είναι το αποτέλεσμα της χρησιμοποίησης του j "μπλοκ" (ε_{ij} είναι το τυχαίο λάθος).

Για τα ε_{ij} υποθέτουμε ότι είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή $N(0,\sigma^2)$.

Σύμφωνα με τις υποθέσεις αυτές η αρχική μας υπόθεση είναι ισοδύναμη με την υπόθεση

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$$

(ότι δηλαδή οι επιδράσεις δεν διαφέρουν)

$$H_1 : \text{τουλάχιστον δύο από τα } \alpha_i \text{ διαφέρουν.}$$

Ανάλυση με δυο παράγοντες

Επίσης, σε μερικές περιπτώσεις μας ενδιαφέρει να ελέγξουμε την υπόθεση ότι τα μπλοκ που χρησιμοποιήθηκαν δεν διαφέρουν στατιστικά σημαντικά μεταξύ τους. Στην περίπτωση αυτή η ισοδύναμη υπόθεση που θα πρέπει να ελέγξουμε είναι η

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_b = 0$

H_1 : τουλάχιστον δύο από τα β_i διαφέρουν μεταξύ τους.

Η διαφορά στην ανάλυση που θα ακολουθήσουμε εδώ από εκείνη της ανάλυσης κατά ένα κριτήριο (one-way analysis) είναι ότι ενώ εκεί το SST είχε αναλυθεί σε δύο συνιστώσες (SSE και SSTr) εδώ το SSE χωρίζεται και σε ένα τρίτο παράγοντα ώστε να ληφθεί υπόψη η διαφορά της διακύμανσης που οφείλεται στη χρησιμοποίηση διαφορετικών μπλοκ

Ανάλυση με δυο παράγοντες

$$: SST = SSTr + SSE$$

$$SST = SSTr + \boxed{SSB + SSE}$$

$$SSTr = b\sum_{i}(\bar{Y}_{i.} - \bar{Y}_{..}) = \frac{\sum_{i}(Y_{i.})^{2}}{b} - \frac{(Y_{..})^{2}}{kb}$$

$$SSB = k\sum_{j}(\bar{Y}_{.j} - \bar{Y}_{..}) = \frac{\sum_{j}(Y_{.j})^{2}}{k} - \frac{(Y_{..})^{2}}{kb}$$

$$SST = \sum_{i}\sum_{j}(Y_{ij} - \bar{Y}_{..}) = \sum_{i}\sum_{j}Y_{ij}^{2} - \frac{(Y_{..})^{2}}{kb}$$

$$SSE = SST - SSTr - SSB$$

Ανάλυση με δυο παράγοντες

Πίνακας 2-WAY-ANOVA
Ανάλυση διακύμανσης κατά δύο κριτήρια

Αιτία Διασποράς	SS	ΒΕ	MS	EMS	F (κάτω από την H_0)
Μεταξύ Επιδράσεων (between treatments)	SST	k - 1	$\frac{SSTr}{k-1} = MSTr$	$\sigma^2 + \frac{b\sum \alpha_i^2}{k-1}$	$\frac{MST_r}{MSE} \equiv F_{Tr}$
Μπλοκ (block)	SSB	b - 1	$\frac{SSB}{b-1} = MSB$	$\sigma^2 + \frac{k\sum \beta_i^2}{b-1}$	$\frac{MSB}{MSE} \equiv F_b$
Μέσα στις Επιδράσεις (Λάθος) (within treatments)	SSE	kb-(k+b-1) =(b-1)(k-1)	$\frac{SSE}{(b-1)(k-1)} = MSE$	σ^2	
Σύνολο	SST	kb-1			

Ανάλυση με δυο παράγοντες

- Ισχύει

$$E(MSTr) = E\left(\frac{SSTr}{k-1}\right) = \sigma^2 + \frac{b\sum \alpha_i^2}{k-1}$$

$$E(MSB) = E\left(\frac{SSB}{b-1}\right) = \sigma^2 + \frac{k\sum \beta_i^2}{b-1}$$

$$\text{και } E\,(M\Sigma E) = E\left(\frac{SSE}{(k-1)(b-1)}\right) = \sigma^2$$

Επομένως, κάτω από τη μηδενική υπόθεση, όχι μόνο το MSE είναι αμερόληπτη εκτιμήτρια του σ αλλά και το MST και MSB.

Ανάλυση με δυο παράγοντες

Πρόβλημα: Η Γραμματεία ενός Τμήματος Στατιστικής ενδιαφέρεται να προμηθευτεί ένα νέο επεξεργαστή κειμένου (word processor). Προκειμένου να επιτευχθεί μια επιτυχής επιλογή σχεδιάζεται ένα τεστ με το οποίο μετριέται ο ρυθμός "εισαγωγής" λέξεων στον υπολογιστή (λέξεις ανά λεπτό). Για τον λόγο αυτό επιλέγονται έξι δακτυλογράφοι και η απόδοση τους (λέξεις ανά λεπτό) καταγράφεται για τους τρεις διαφορετικούς επεξεργαστές. Τα αποτελέσματα εμφανίζονται στον πίνακα που ακολουθεί.

Ρυθμός Εισαγωγής Στοιχείων

	Μάρκα Επεξεργαστή		
Δακτ/φος	1	2	3
1	42	45	45
2	37	36	40
3	53	56	55
4	68	73	75
5	48	45	47
6	36	39	40

Ανάλυση με δυο παράγοντες

Ρυθμός Εισαγωγής Στοιχείων

	Μάρκα Επεξεργαστή		
Δακτ/φος	1	2	3
1	42	45	45
2	37	36	40
3	53	56	55
4	68	73	75
5	48	45	47
6	36	39	40

$y_{.1}$

$y_{1.}$

$\boxed{y_{1.} = 284,}$ $\quad y_{2.} = 294, \quad y_{3.} = 302$

$\bar{y}_{1.} = 47.33, \quad \bar{y}_{2.} = 49, \quad \bar{y}_{3.} = 50.33$

$\boxed{y_{.1} = 132,}$ $\quad y_{.2} = 113, \quad y_{.3} = 164$

$y_{.4} = 216, \quad y_{.5} = 140, \quad y_{.6} = 115$

Ανάλυση με δυο παράγοντες

Ρυθμός Εισαγωγής Στοιχείων

	Μάρκα Επεξεργαστή		
Δακτ/φος	1	2	3
1	42	45	45
2	37	36	40
3	53	56	55
4	68	73	75
5	48	45	47
6	36	39	40

$$\sum_i \sum_j y_{ij}^2 = 45582 \quad k = 3,\ b = 6 \qquad y_{..} = 880$$

$$\frac{(Y_{..})^2}{kb} = \frac{(880)^2}{18} = 4302.22$$

Ανάλυση με δυο παράγοντες

$$SST = 45582 - 43022.22 = 2559.78$$

$$SSTr = \frac{1}{6}(284^2 + 294^2 + 302^2) - 43022.22 = 27.11$$

$$SSB = \frac{1}{3}(132^2+113^2+164^2+216^2+140^2+115^2)-43022.22 = 2501.11$$

$$SSE = 2559.78 - 27.11 - 2501.11 = 31.56$$

$$MSTr = \frac{SST}{3-1} = \frac{27.11}{2} = 13.56$$

$$MSE = \frac{SSE}{(6-1)(3-1)} = \frac{31.56}{10} = 3.16$$

$$F_{tr} = \frac{13.56}{3.16}\left(\frac{MSTr}{MSE}\right) = 4.30$$

Για $\alpha = 0.05$ $F_{2,\,10,\,0.95} = 4.103$

Επομένως, απορρίπτουμε την H_0. Δηλαδή, καταλήγουμε στο συμπέρασμα ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ της αποτελεσματικότητας των τριών επεξεργαστών κειμένου.

Ανάλυση με δυο παράγοντες

$$MSB = \frac{2501.11}{6-1} = 500.22$$

$$MSE = \frac{SSE}{(6-1)(3-1)} = \frac{31.56}{10} = 3.16$$

$$F_b = \frac{500.22}{3.16} = 158.30$$

$$F_{5,\ 10,\ 0.95} = 3.226$$

Βλέπουμε, επομένως, ότι υπάρχει και πολύ σημαντική στατιστικά διαφορά στην απόδοση των 6 δακτυλογράφων.

Ανάλυση με δυο παράγοντες

Πίνακας 2-WAY-ANOVA
Ανάλυση διακύμανσης κατά δύο κριτήρια

Αιτία Διασποράς	SS	ΒΕ	MS	F (κάτω από την H_0)
Μεταξύ Επιδράσεων	27.11	2	13.56	4.30
Μπλοκ (block)	2501.11	5	500.22	158.30
Μέσα στις Επιδράσεις (Λάθος)	31.56	10	3.16	
Σύνολο	2559.78	17		

Ανάλυση με δυο παράγοντες

Έστω ότι το αποτέλεσμα ενός πειράματος εξαρτάται από 2 παράγοντες, Α και Β. Αρχικά υποθέτουμε ότι δεν υπάρχει αλληλεπίδραση (interaction) μεταξύ των παραγόντων. Θα μελετήσουμε τον παράγοντα Α σε m επίπεδα και τον παράγοντα Β σε l επίπεδα. Επομένως υπάρχουν $k = m \cdot l$ συνδυασμοί επιπέδων (treatments) για τους οποίους λαμβάνουμε παρατηρήσεις, έστω μια παρατήρηση σε κάθε treatment. Δηλαδή

$$
\begin{array}{c}
 \qquad B \\
A \quad
\begin{array}{|cccc|c|c|}
\hline
Y_{11} & Y_{12} & \cdots & Y_{1l} & Y_{1\cdot} & \overline{Y}_{1\cdot} \\
Y_{21} & Y_{22} & \cdots & Y_{2l} & Y_{2\cdot} & \overline{Y}_{2\cdot} \\
 & & \vdots & & \vdots & \vdots \\
Y_{m1} & Y_{m2} & \cdots & Y_{ml} & Y_{m\cdot} & \overline{Y}_{m\cdot} \\
\hline
Y_{\cdot 1} & Y_{\cdot 2} & \cdots & Y_{\cdot l} & & \\
\cline{1-4}
\overline{Y}_{\cdot 1} & \overline{Y}_{\cdot 2} & \cdots & \overline{Y}_{\cdot l} & & \\
\hline
\end{array}
\end{array}
$$

Ανάλυση με δυο παράγοντες

$$
\begin{array}{c|cccc|c|c|}
 & & & B & & & \\
\cline{2-7}
 & Y_{11} & Y_{12} & \cdots & Y_{1l} & Y_{1\cdot} & \overline{Y}_{1\cdot} \\
 & Y_{21} & Y_{22} & \cdots & Y_{2l} & Y_{2\cdot} & \overline{Y}_{2\cdot} \\
A & & & \vdots & & \vdots & \vdots \\
 & Y_{m1} & Y_{m2} & \cdots & Y_{ml} & Y_{m\cdot} & \overline{Y}_{m\cdot} \\
\cline{2-7}
 & Y_{\cdot 1} & Y_{\cdot 2} & \cdots & Y_{\cdot l} & & \\
\cline{2-5}
 & \overline{Y}_{\cdot 1} & \overline{Y}_{\cdot 2} & \cdots & \overline{Y}_{\cdot l} & & \\
\cline{2-7}
\end{array}
$$

$$
\begin{aligned}
Y_{i\cdot} &= \sum_{j=1}^{l} Y_{ij} \\
\overline{Y}_{i\cdot} &= \frac{1}{l} Y_{i\cdot}
\end{aligned}
$$

$$
\begin{aligned}
Y_{\cdot j} &= \sum_{i=1}^{m} Y_{ij} \\
\overline{Y}_{\cdot j} &= \frac{1}{m} Y_{\cdot j}
\end{aligned}
$$

$$
\overline{Y}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{l} Y_{ij} = \frac{1}{m} \sum_{i=1}^{m} \overline{Y}_{i\cdot} = \frac{1}{l} \sum_{i=1}^{l} \overline{Y}_{\cdot j} \,.
$$

Ανάλυση με δυο παράγοντες

$$
\begin{aligned}
Y_{ij} &= \mu_i + \epsilon_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, l \\
\epsilon_{ij} &: \text{τυχαία σφάλματα, } \epsilon_{ij} \overset{\text{ανεξ.}}{\sim} N(0, \sigma^2) \\
\mu_{ij} &= \mu_{..} + \alpha_i + \beta_j \\
\alpha_i &: \text{επίδραση του } i \text{ επιπέδου του παράγοντα A} \\
\beta_j &: \text{επίδραση του } j \text{ επιπέδου του παράγοντα B}
\end{aligned}
$$

$$\text{έτσι ώστε } \sum_{i=1}^{m} \alpha_i = 0 \text{ και } \sum_{j=1}^{l} \beta_j = 0$$

$$, Y_{ij} \sim N(\mu_{ij}, \sigma^2).$$

Ανάλυση με δυο παράγοντες

$$\sum_{i=1}^{m} \alpha_i = 0 \qquad \sum_{j=1}^{l} \beta_j = 0.$$

$$\mu_{ij} = \mu'_{..} + \alpha'_i + \beta'_j$$

$$\overline{\alpha}' = \frac{1}{m} \sum_{i=1}^{m} \alpha'_i = 0 \qquad \overline{\beta}' = \frac{1}{l} \sum_{j=1}^{l} \beta'_j = 0.$$

$$\mu_{ij} = \underbrace{\mu'_{..} + \overline{\alpha}' + \overline{\beta}'}_{\mu_{..}} + \underbrace{\alpha'_i - \overline{\alpha}'}_{\alpha_i} + \underbrace{\beta'_i - \overline{\beta}'}_{\beta_j}$$

$$\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} (\alpha'_i - \overline{\alpha}') = 0 \qquad \sum_{i=j}^{l} \beta_j = \sum_{j=1}^{l} (\beta'_j - \overline{\beta}') = 0.$$

Ανάλυση με δυο παράγοντες

Για να ελέγξουμε αν υπάρχει επίδραση του παράγοντα Α κάνουμε τον έλεγχο

$$\begin{aligned} H_0 &: \mu_{11} = \mu_{21} = \cdots = \mu_{m1} \\ H_1 &: 2 \text{ τουλάχιστον διαφέρουν} \end{aligned}$$

Ισοδύναμα

$$\begin{aligned} H_0 &: \alpha_1 = \alpha_2 = \cdots = \alpha_m \\ H_1 &: \alpha_i \neq \alpha_j \text{ για τουλάχιστον ένα ζεύγος } i, j \end{aligned}$$

Αντίστοιχα για να ελέγξουμε αν υπάρχει επίδραση του παράγοντα Β

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \cdots = \beta_m \\ H_1 &: \beta_i \neq \beta_j \text{ για τουλάχιστον ένα ζεύγος } i, j \end{aligned}$$

Ανάλυση με δυο παράγοντες

$$\begin{aligned} SST &= \sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{..})^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{l}\left((\overline{Y}_{i.}-\overline{Y}_{..})+(\overline{Y}_{.j}-\overline{Y}_{..})+(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})\right)^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{i.}-\overline{Y}_{..})^2+\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})^2+\sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})^2\,, \end{aligned}$$

μπορεί να αποδειχθεί ότι όλα τα αθροίσματα διπλασίων γινομένων που σχηματίζονται από $(\overline{Y}_{i.}-\overline{Y}_{..})$, $(\overline{Y}_{.j}-\overline{Y}_{..})$ και $(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})$ είναι 0. Επομένως

$$SST = \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{i.}-\overline{Y}_{..})^2}_{SSA} + \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(\overline{Y}_{.j}-\overline{Y}_{..})^2}_{SSB} + \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{l}(Y_{ij}-\overline{Y}_{i.}-\overline{Y}_{.j}+\overline{Y}_{..})^2}_{SSE}$$

SSA : άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα A

SSB : άθροισμα τετραγώνων μεταξύ επιπέδων του παράγοντα B

SSE : άθροισμα τετραγώνων των τυχαίων σφαλμάτων.

Ανάλυση με δυο παράγοντες

Κάτω από την H_0 **:** $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ έχουμε $\frac{SSA}{\sigma^2} \sim X^2_{m-1}$. Άρα απορρίπτουμε την H_0 αν $F_A = \frac{SSA/m-1}{SSE/(m-1)(l-1)} > F_{1-\alpha,m-1,(m-1)(l-1)}$.

Κάτω από την H_0 **:** $\beta_1 = \beta_2 = \cdots = \beta_m = 0$ έχουμε $\frac{SSB}{\sigma^2} \sim X^2_{l-1}$. Άρα απορρίπτουμε την H_0 αν $F_B = \frac{SSB/l-1}{SSE/(m-1)(l-1)} > F_{1-\alpha,l-1,(m-1)(l-1)}$.

Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	af	MS	$F = \frac{MSF}{MSE}$	p–value
Παράγοντας A	SSA	$m-1$	$MSA = \frac{SSA}{m-1}$	$F_A = \frac{MSA}{MSE}$	$\star$
Παράγοντας B	SSB	$l-1$	$MSB = \frac{SSB}{l-1}$	$F_B = \frac{MSB}{MSE}$	
Σφάλμα(Error)	SSE	$(m-1)(l-1)$	$MSE = \frac{SSE}{(m-1)(l-1)}$		
Σύνολο(Total)	SST	$ml-1$			

Ανάλυση με δυο παράγοντες

Παράδειγμα 3.3 *Μας ενδιαφέρει να μελετήσουμε την κατανάλωση βενζίνης από αυτοκίνητα μιας συγκεκριμένης κατηγορίας. Έχουμε 3 μάρκες αυτοκινήτων και 4 διαφορετικούς τύπους βενζίνης. Ο αριθμός των χιλιομέτρων (km) ανά γαλόνι βενζίνης για καθέναν από τους συνδυασμούς είναι*

μάρκα αυτοκινήτου	τύπος βενζίνης				$\overline{Y}_{i\cdot}$
	1	2	3	4	
1	16	18	21	21	19
2	14	15	18	17	16
3	15	15	18	16	16
$\overline{Y}_{\cdot j}$	15	16	19	18	$\overline{Y}_{\cdot\cdot}$=17

Μοντέλο $Y_{ij} = \mu + \alpha_i + \beta_j$, $i = 1,2,3$, $j = 1,2,3,4$. *Μας ενδιαφέρει να ελέγξουμε κατά πόσο η μέση κατανάλωση βενζίνης διαφέρει για τους διάφορους τύπους βενζίνης σε επίπεδο στατιστικής σημαντικότητας* $\alpha = 1\%$. *Δηλαδή, θέλουμε να ελέγξουμε*

$$H_0 \quad : \quad \beta_1 = \beta_2 = \cdots = \beta_m$$

$$vs$$

$$H_1 \quad : \quad \beta_i \neq \beta_j \text{ για τουλάχιστον ένα ζεύγος } i, j$$

Ανάλυση με δυο παράγοντες

Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	df	MS	$F = \frac{MSF}{MSE}$
Παράγοντας A	$SSA = 24$	$m-1=2$	$MSA = \frac{24}{2} = 12$	$F_A = \frac{12}{2/3} = 18$
Παράγοντας B	$SSB = 30$	$l-1=3$	$MSB = \frac{30}{3} = 10$	$F_B = \frac{10}{2/3} = 15$
Σφάλμα (Error)	$SSE = 4$	$(m-1)(l-1) = 6$	$MSE = \frac{4}{6} = \frac{2}{3}$	
Σύνολο (Total)	$SST = 58$	$ml-1=11$		

Σε $\alpha = 1\%$ *απορίπτουμε την* H_0 *εαν* $F_B > F_{0.99,3,6} \Rightarrow F_B = 15 > F_{0.99,3,6} = 9.78$, *άρα απορρίπτουμε την* H_0. *Δηλαδή η μέση κατανάλωση βενζίνης διαφέρει για τους διάφορους τύπους βενζίνης, στις 3 μάρκες αυτοκινήτων που εξετάστηκαν.*

Ανάλυση με δυο παράγοντες

Παράδειγμα 4. Στις Ηνωμένες Πολιτείας της Αμερικής η Εθνική Ομοσπονδί Καλαθοσφαίρισης (NBA) κάνει μια μελέτη για να συγκρίνει την απόδοση των ομάδω που συγκροτούν την Ανατολική, Κεντρική και Δυτική περιφέρεια του πρωταθλήματος Για τον σκοπό αυτό συγκρίνουν την μέση επίθεση που είχαν 5 ομάδες κάθ περιφέρειας που βρίσκονταν στην ίδια θέση του βαθμολογικού πίνακα. Τ αποτελέσματα ήταν τα ακόλουθα:

Περιφέρεια	$1^η$ ομάδα	$2^η$ ομάδα	$3^η$ ομάδα	$4^η$ ομάδα	$5^η$ ομάδα
Ανατολική	98	84	78	74	63
Κεντρική	101	86	76	72	64
Δυτική	99	87	76	67	63

Ανάλυση με δυο παράγοντες

	Blocks					Αθροίσματα
Επιδράσεις	1	2	3	4	5	
1	98	84	78	74	63	397
2	101	86	76	72	64	399
3	99	87	76	67	63	392
Αθροίσματα	**298**	**257**	**230**	**213**	**190**	**1188**

Ο τύπος της συνολικής διακύμανσης είναι: $SS_T = \left(\sum_{i=1}^{t} \sum_{j=1}^{b} y_{ij}^2 \right) - \frac{y_{..}^2}{N}$

$$\left(\sum_{i=1}^{3} \sum_{j=1}^{5} y_{ij}^2 \right) = 98^2 + 84^2 + \dots 67^2 + 63^2 = 96446$$

Και $SS_T = \left(\sum_{i=1}^{3} \sum_{j=1}^{5} y_{ij}^2 \right) - \frac{y_{..}^2}{N} = 96446 - \frac{1188^2}{15} = 2356$

Ανάλυση με δυο παράγοντες

Η διακύμανση μεταξύ των επιδράσεων είναι: $SS_{TR} = \frac{1}{b}\left(\sum_{i=1}^{t} y_{i.}^2\right) - \frac{y_{..}^2}{N}$

$$\frac{1}{5}\left(\sum_{i=1}^{3} y_{i.}^2\right) = \frac{1}{5}(397^2 + 399^2 + 392^2) = 156824$$

$$SS_{TR} = \frac{1}{3}\left(\sum_{i=1}^{6} y_{i.}^2\right) - \frac{y_{..}^2}{N} = 156824 - \frac{1188^2}{15} = 5{,}2$$

Η διακύμανση μέσα στο μπλοκ είναι: $SS_{bl} = \frac{1}{t}\left(\sum_{i=1}^{b} y_{.j}^2\right) - \frac{y_{..}^2}{N}$

$$\frac{1}{3}\left(\sum_{i=1}^{5} y_{.j}^2\right) = \frac{1}{3}(298^2 + 257^2 + 230^2 + 213^2 + 190^2) = 96407$$

Ανάλυση με δυο παράγοντες

$$SS_{bl} = \frac{1}{3}\left(\sum_{i=1}^{5} y_{.j}^{2}\right) - \frac{y_{..}^{2}}{N} = 96407 - \frac{1188^{2}}{15} = 2317$$

Για να βρούμε την διακύμανση μέσα στις επιδράσεις (SS_E) χρησιμοποιούμε τον τύπο:

Δηλαδή:

$$SS_T = SS_{TR} + SS_E + SS_{bl} \Leftrightarrow SS_E = SS_T - SS_{TR} - SS_{bl} = 2356 - 5{,}2 - 2317 = 33{,}47$$

Οι βαθμοί ελευθερίας είναι: Για το SS_{TR}: t-1 = 3 -1 = 2

Για το SS_E: (t-1) (b-1) = (3 - 1) (5 – 1) = 8

Για το SS_{bl} : b-1 = 5 – 1 = 4

Ανάλυση με δυο παράγοντες

$$MS_{TR} = \frac{SS_{TR}}{t-1} = \frac{5{,}2}{2} = 2{,}6$$

$$MS_{\mathrm{E}} = \frac{SS_{\mathrm{E}}}{(t-1)(b-1)} = \frac{33{,}47}{8} = 4{,}18$$

$$MS_{bl} = \frac{SS_{bl}}{b-1} = \frac{2317}{4} = 579$$

Η συνάρτηση ελέγχου F_o για τον έλεγχο της ισότητας μεταξύ των μέσων τιμών των επιδράσεων είναι:

H_o: Οι μέσες τιμές μεταξύ των επιδράσεων είναι ίσες
H_1: Οι μέσες τιμές μεταξύ των επιδράσεων δεν είναι ίσες

$$F_o = \frac{MS_{TR}}{MS_E} = \frac{5{,}2}{4{,}18} = 1{,}24$$

Ανάλυση με δυο παράγοντες

Η συνθήκη απόρριψης της μηδενικής υπόθεσης είναι η: $\mathbf{F_o > F_{\alpha, t-1, (t-1)(b-1)}}$. Το $F_{\alpha, t-1, (t-1)(b-1)}$ είναι $F_{0,05, 3-1, (3-1)(5-1)}$ ή $F_{0,05, 2, 8}$. Από τον πίνακα εκατοστημορίων της κατανομής F βρίσκουμε ότι η $F_{0,05, 2, 8}$ ισούνται με 4,46.

Αφού 1,24 < 4,46 η μηδενική υπόθεση δεν απορρίπτεται. Η παρατηρούμενη ανισότητα μεταξύ των μέσων τιμών είναι τυχαία ή πιο πρακτικά οι μέσες τιμές των επιδράσεων είναι ίσες.

Η συνάρτηση ελέγχου F_o για τον έλεγχο της ισότητας μεταξύ των μέσων τιμών των μπλοκς είναι:

H_o: Οι μέσες τιμές μεταξύ των μπλοκς είναι ίσες
H_1: Οι μέσες τιμές μεταξύ των μπλοκς δεν είναι ίσες

$$F_o = \frac{MS_{bl}}{MS_E} = \frac{2317}{33,47} = 69,2$$

Ανάλυση με δυο παράγοντες

Η συνθήκη απόρριψης της μηδενικής υπόθεσης είναι η: $\mathbf{F_o > F_{\alpha, b-1, (t-1)(b-1)}}$. Το $F_{\alpha, b-1, (t-1)(b-1)}$ είναι $F_{0,05, 4, 8}$. Από τον πίνακα εκατοστημορίων της κατανομής F βρίσκουμε ότι η $F_{0,05, 4, 8}$ ισούνται με 3,84.

Αφού 69 > 3,84 η μηδενική υπόθεση απορρίπτεται (έστω και οριακά) και κατά συνέπεια οι μέσες τιμές των μπλοκς δεν είναι ίσες.

Αιτία διασποράς	**Βαθμοί ελευθερίας**	**Αναλύσεις διακύμανσης**	**Μέσες διακυμάνσεις**	**F**	**P**
Μεταξύ επιδράσεων (TR)	2	$SS_{TR} = 5,2$	$MS_{TR} = 2,6$	1,24	0,56
Μέσα το μπλοκ (bl)	4	$SS_{bl} = 2317,7$	$MS_{bl} = 579,4$	69,2	0,00
Μέσα στην επίδραση (E)	8	$SS_E = 33,47$	$MS_E = 4,8$		
Σύνολο (T)	14	$SS_T = 2356,4$			