

2. ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ ΚΑΤΑ ΔΥΟ ΚΡΙΤΗΡΙΑ **(Analysis of Variance for two factor Experiments)** **(Two-Way Analysis of Variance)**

Ο πειραματικός σχεδιασμός για τον οποίο θα μιλήσουμε είναι μια επέκταση της μεθοδολογίας σύγκρισης των μέσων δύο πληθυσμών όταν τα δείγματα που χρησιμοποιούμε έχουν επιλεγεί κατά ζεύγη. Και στις δύο περιπτώσεις οι μέθοδοι που χρησιμοποιούνται αποσκοπούν στο να ελεγχθεί η ανεπιθύμητη διακύμανση των πειραματικών αποτελεσμάτων. Με τον πειραματικό αυτό σχεδιασμό αποβλέπουμε στο να συγκρίνουμε διαφορετικές επιδράσεις χωρίζοντας τις μονάδες στις οποίες εφαρμόζονται οι επιδράσεις σε ομάδες που έχουν κοινά χαρακτηριστικά.

2.1 Σχεδιασμός τυχαιοποιημένων πλήρων σχημάτων **(Randomized Complete block Design) (RCB)**

Σύμφωνα με αυτό τον πειραματικό σχεδιασμό έχουμε παρατηρήσεις για κάθε μία από k επιδράσεις (treatments) σε κάθε ένα από b ομογενή στρώματα (μπλοκ). Η ανάλυση στην περίπτωση αυτή παίρνει υπόψη της την παρουσία των μπλοκ και την συνεισφορά τους στη συνολική διασπορά οπότε η διασπορά που θεωρείται ότι προέρχεται από τις επιδράσεις καθορίζεται καλύτερα.

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

Αν υποτεθεί ότι ο σχεδιασμός που είχε χρησιμοποιηθεί ήταν ένας πλήρως τυχαιοποιημένος σχεδιασμός (*completely randomized design*) και υποθέταμε ότι τα στοιχεία του πίνακα προήλθαν από 18 διαφορετικές δακτυλογράφους (αν αγνοούσαμε δηλαδή το γεγονός ότι είχαν χρησιμοποιηθεί έξι δακτυλογράφοι) και υποθέταμε ότι είχαμε έξι παρατηρήσεις για κάθε επεξεργαστή θα καταλήγαμε στο συμπέρασμα ότι δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των επεξεργαστών. Ο πίνακας ανάλυσης διακύμανσης του προβλήματος αυτού για τον πλήρως τυχαιοποιημένο σχεδιασμό δίνει τα αποτελέσματα που δίνονται στη συνέχεια.

Επειδή $F < F_{2, 15, 0.95} = 3.682$, με τον πλήρως τυχαιοποιημένο σχεδιασμό δεν θα απορρίπταμε την μηδενική υπόθεση ότι υπάρχει ουσιαστική διαφορά μεταξύ των τριών επεξεργαστών.

Η ανάλυση όμως αυτή δεν μπορεί με κανένα τρόπο να θεωρηθεί σωστή αφού γνωρίζουμε ότι οι δακτυλογράφοι που μετείχαν στο τεστ είναι φυσικό να ενεργούν ως ανομοιογενείς μονάδες. Θα πρέπει δηλαδή να λάβουμε υπόψη μας ότι ο ρυθμός εισαγωγής στοιχείων, εκτός από τους τρεις επεξεργαστές, επηρεάζεται και από το συγκεκριμένο άτομο που χρησιμοποιεί τους επεξεργαστές αυτούς. Επομένως αρκετή από την συνολική διακύμανση οφείλεται στην παρουσία των "μπλοκ" στον σχεδιασμό.

Πίνακας ANOVA

Αιτία Διασποράς	SS	DF	MS	F
Μεταξύ Επιδράσεων	SSTr=27.11	k - 1 = 2	$S_0^2 = \frac{SSTr}{k - 1}$ = 13.56	$\frac{S_0^2}{S_p^2} = 0.08$
Μέσα στις Επιδράσεις (Λάθος)	SSE=2532.67	N - k = 15	$S_p^2 = \frac{SSE}{N - k}$ = 168.84	
Σύνολο	SST=2559.78	N - 1 = 17		

Σημείωση: Μια απλή μέθοδος για να εξαλείψουμε αυτής της μορφής την διακύμανση είναι αυτή που χρησιμοποιεί τα τυποποιημένα αποτελέσματα του πειράματος. (Τα αποτελέσματα αυτά προκύπτουν αν από κάθε παρατήρηση αφαιρέσουμε τον μέσο κάθε "μπλοκ" και διαιρέσουμε με την τυπική απόκλιση στο συγκεκριμένο μπλοκ).
Για το πρόβλημά μας π.χ. έχουμε:

$$Z_1 = (42-44)/1.73 = -1.16, Z_2 = (45-44)/1.73 = 0.58$$

Ο αντίστοιχος πίνακας των τυποποιημένων τιμών θα είναι ο πίνακας που ακολουθεί:

Τυποποιημένες τιμές για τους τρεις επεξεργαστές κειμένου

	1	2	3
1	-1.16	0.58	0.58
2	-0.32	-0.80	1.12
3	-1.09	0.87	0.22
4	-1.11	0.28	0.83
5	0.87	-1.09	0.22
6	-1.12	0.32	0.80

Αν παρατηρήσουμε τον πίνακα των τυποποιημένων τιμών βλέπουμε αμέσως ένδειξη διαφοράς στην απόδοση τους. Είναι προφανές ότι ο επεξεργαστής 3, για παράδειγμα, έχει σταθερά καλύτερη απόδοση από τον επεξεργαστή 1.

Το παράδειγμα αυτό οδηγεί στο συμπέρασμα ότι στην ανάλυση μας θα πρέπει να πάρουμε υπόψη μας και το γεγονός ότι έχουμε χρησιμοποιήσει μπλοκ και να διαμορφώσουμε μια άλλη συμπερασματολογία.

2.2 Ανάλυση Διακύμανσης για το Σχεδιασμό Τυχαιοποιημένων Πλήρων Στρωμάτων (Two-way analysis of variance) (The Analysis of Variance for the Randomized Complete Block Design)

Στην ανάλυση μας αυτή 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:

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

H_1 : Δύο, τουλάχιστον, από τους μέσους διαφέρουν

Γενικά, μπορούμε να παρατηρήσουμε ότι ο σχεδιασμός αυτός του πειράματος χρησιμοποιείται όταν οι μονάδες του πειράματος που πρόκειται να υποβληθούν σε κάποια "επίδραση" είναι ανομοιογενείς και θεωρούμε χρήσιμο να τις ομαδοποιήσουμε σε μπλοκ που έχουν παρόμοια χαρακτηριστικά. Τέτοιες ομαδοποιήσεις γίνονται συνήθως σε ιατρικά πειράματα όπου ομαδοποιούμε ασθενείς με παρόμοιες ασθένειες, ηλικίες, φύλλο, κλπ ώστε να υποστούν την ίδια θεραπεία και να μελετηθούν. Το ίδιο συμβαίνει σε πειράματα όπου ομαδοποιούμε αδέρφια κλπ.

Η ανάλυση διακύμανσης κατά δύο κριτήρια που θα αναπτύξουμε στη συνέχεια 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.

Το μοντέλο

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

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

Υποθέτοντας 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

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

όπου $\mu = \frac{\sum \mu_i}{k}$, 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

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

(ότι δηλαδή οι επιδράσεις δεν διαφέρουν)

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

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

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

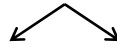
$$H_1 : \text{τουλάχιστον δύο από τα } \beta_i \text{ διαφέρουν μεταξύ τους.}$$

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.

Έτσι έχουμε

$$\text{CR design: } SST = SSTr + SSE_{CR}$$



$$\text{RCB design: } SST = SSTr + SSB + SSE_{RCB}$$

Στην περίπτωση αυτή οι αντίστοιχες στατιστικές συναρτήσεις θα δίνονται από τους εξής τύπους:

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

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

$$SST = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j Y_{ij}^2 - \frac{(Y_{..})^2}{kb}$$

$$SSE = SST - SSTr - SSB$$

Ο αντίστοιχος πίνακας διακύμανσης θα είναι ο εξής.

Πίνακας 2-WAY-ANOVA
Ανάλυση διακύμανσης κατά δύο κριτήρια

Αιτία Διασποράς	SS	BE	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			

Σημείωση 1: Η ανάλυση αυτή στηρίζεται στο γεγονός ότι

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

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

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

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

Σημείωση 2: Πολλές φορές το $\frac{(Y_{..})^2}{kb}$ ονομάζεται *συντελεστής διόρθωσης (correction factor)*

$$CF = \frac{(Y_{..})^2}{kb}$$

Σημείωση 3: Και για την ανάλυση αυτή, προκειμένου να είμαστε σε θέση να χρησιμοποιήσουμε την κατανομή F, είναι απαραίτητο να κάνουμε τις ίδιες υποθέσεις που είχαμε κάνει στον πλήρως τυχαιοποιημένο σχεδιασμό. Οι υποθέσεις αυτές είναι ότι οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα έχουν κοινή διασπορά σ^2 όπως επίσης και ότι οι αποκλίσεις των παρατηρήσεων είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διασπορά σ^2 .

Επανερχόμενοι στο πρόβλημά μας με τους τρεις επεξεργαστές κειμένου μπορούμε να υπολογίσουμε τις τιμές των στατιστικών συναρτήσεων και τα συγκεκριμένα δεδομένα. Έτσι έχουμε

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

$$\bar{y}_1 = 47.33, \quad \bar{y}_2 = 49, \quad \bar{y}_3 = 50.33$$

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

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

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

$$T_{..} \equiv y_{..} = 880$$

$$CF = \frac{(880)^2}{18} = 4302.22$$

$$SST = 45582 - 43022.22 = 2559.78$$

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

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

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

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

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

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

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

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

Επίσης έχουμε:

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

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

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

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

Ο πίνακας ανάλυσης διασποράς είναι ο εξής

Πίνακας 2-WAY-ANOVA
Ανάλυση διακύμανσης κατά δύο κριτήρια

Αιτία Διασποράς	SS	BE	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		

Σημείωση: Είναι ενδιαφέρον να παρατηρήσουμε ότι τα συμπεράσματα που καταλήξαμε με την ανάλυση αυτή είναι τελείως διαφορετικά από εκείνα στα οποία θα καταλήγαμε αν είχαμε χρησιμοποιήσει την ανάλυση διακύμανσης κατά ένα παράγοντα (*one-way analysis of variance*). Μια σύγκριση των δύο αναλύσεων μας δείχνει ότι οι τιμές των SST και SSTr είναι οι ίδιες και στις δύο αναλύσεις. Εδώ όμως η τιμή του SSE έχει μειωθεί από 2532.67 που ήταν στην προηγούμενη ανάλυση σε 31.56 και η αντίστοιχη τιμή του MSE έχει ελαττωθεί από 168.84 σε 3.16. Δεδομένου ότι η τελευταία αυτή τιμή είναι ο διαιρέτης στο λόγο F που χρησιμοποιούμε για τον έλεγχο υποθέσεως και δεδομένου ότι ο αριθμητής είναι σταθερός με τιμή 13.56 είναι φυσικό ότι ο έλεγχος που εφαρμόζουμε τώρα είναι περισσότερο ευαίσθητος σε διαφορές μεταξύ των επιδράσεων. Αυτό φαίνεται και στην τιμή που αντιστοιχεί στη διακύμανση των δεδομένων που οφείλεται στις διαφορές μεταξύ των μπλοκ.

2.3 Αλληλεπιδράσεις σε Πειράματα Δύο Παραγόντων (Interaction in Two-Factor Experiment)

Ας υποθέσουμε ότι στο πρόβλημα των τριών επεξεργαστών κειμένου αποφασίζουμε να σχεδιάσουμε το πείραμα με τέτοιο τρόπο ώστε για κάθε μια από τις έξη δακτυλογράφους που χρησιμοποιούμε και για κάθε ένα από τους τρεις υπό εξέταση επεξεργαστές κειμένου να έχουμε τρία ανεξάρτητα σύνολα μετρήσεων για τον αριθμό λέξεων ανά λεπτό που επιτυγχάνουν οι δακτυλογράφοι. Έστω ότι από τον

σχεδιασμό αυτό του πειράματος πήραμε τα αποτελέσματα που δίνονται στη συνέχεια.

Ρυθμός Εισαγωγής Στοιχείων

Δακτυλογράφος	Είδος Επεξεργαστή		
	1	2	3
1	42	45	45
	43	43	44
	39	45	47
2	37	36	40
	36	37	41
	36	34	42
3	53	56	55
	56	55	57
	54	58	55
4	68	73	75
	70	71	78
	65	76	75
5	48	45	47
	44	46	49
	47	45	47
6	36	39	40
	34	39	42
	37	37	38

Τίθεται το ερώτημα αν οι επαναλαμβανόμενες μετρήσεις επηρεάζουν το αποτέλεσμα του πειράματος.

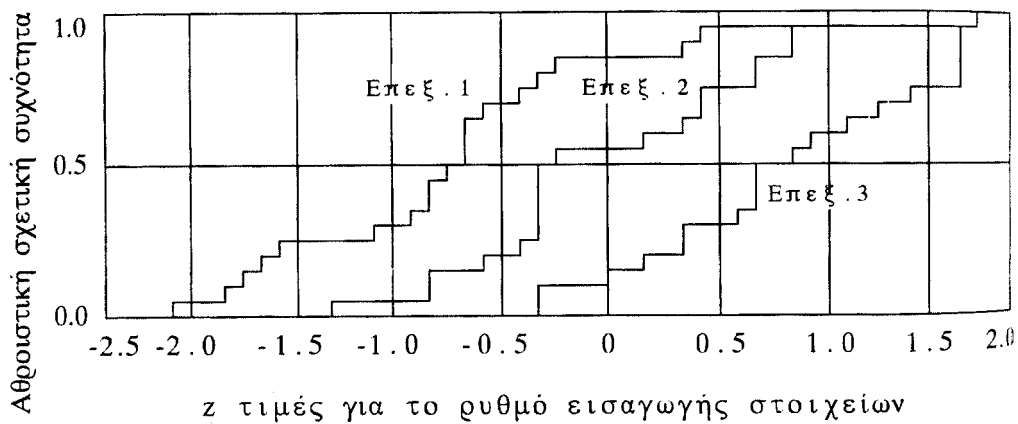
Είναι προφανές ότι οι επαναληπτικές μετρήσεις δημιουργούν μια πρόσθετη διακύμανση στα δεδομένα. Αυτό γιατί οι τρεις διαφορετικές μετρήσεις που παίρνουμε από την κάθε δακτυλογράφο για τον κάθε επεξεργαστή είναι φυσικό να διαφέρουν αφού η κάθε δακτυλογράφος δεν έχει πάντα την ίδια απόδοση.

Τυποποιημένες τιμές των Ρυθμών Εισαγωγής Στοιχείων

Δακτυλογράφος	Είδος Επεξεργαστή		
	1	2	3
1	-0.73	0.58	0.58
	-0.29	-0.29	0.15
	-2.04	0.58	1.45
2	-0.25	-0.62	0.85
	-0.62	-0.25	1.24
	-0.62	-1.36	1.61
3	-1.62	0.37	-0.29
	0.37	-0.29	1.03
	-0.96	1.69	-0.29
4	-1.04	0.16	0.64
	-0.56	-0.32	1.35
	-1.75	0.88	0.64
5	0.98	-0.91	0.35
	-1.54	-0.28	1.61
	0.35	-0.91	0.35
6	-0.85	0.43	0.85
	-1.71	0.43	1.71
	0.43	-0.43	0.00

Μια πρώτη εκτίμηση της επίδρασης που έχει η εισαγωγή των επαναλαμβανομένων μετρήσεων επιτυγχάνεται με την γραφική παράσταση της αθροιστικής εμπειρικής συνάρτησης κατανομής του δείγματος. Για να κατασκευάσουμε την αθροιστική εμπειρική συνάρτηση κατανομής του δείγματος τυποποιούμε τις παρατηρήσεις σε κάθε μπλοκ. Τα αποτελέσματα δίνονται στον παρακάτω πίνακα.

Η γραφική παράσταση των αθροιστικών συναρτήσεων καταλοίπων για καθένα από τους επεξεργαστές δίνεται στο σχήμα που ακολουθεί.



Γραφική παράσταση των αθροιστικών συναρτήσεων καταλοίπων για καθένα από τους επεξεργαστές

Η γραφική παράσταση δίνει μια ένδειξη προφανούς διαφοράς των τριών εμπειρικών συναρτήσεων κατανομών. Παρέχει επίσης μιά ένδειξη ότι η διάταξη των μέσων των τριών πληθυσμών είναι $\mu_1 < \mu_2 < \mu_3$. Στη συνέχεια θα προσπαθήσουμε να αντιμετωπίσουμε το πρόβλημα με την μέθοδο της ανάλυσης διασποράς.

2.4 Σχεδιασμός Τυχαιοποιημένων Πλήρων Σχημάτων με Αλληλεπιδράσεις (Randomized Complete Block Design with Interaction)

Όπως παρατηρήσαμε ήδη οι επαναληπτικές μετρήσεις δημιουργούν πρόσθετη διακύμανση στα δεδομένα. Ο σχεδιασμός όμως με τον τρόπο αυτό είναι καλύτερος. Αυτό γιατί δεν χρειάζεται πια να υποθέσουμε ότι η επιρροή των μπλοκ (*block effect*) είναι η ίδια από επίδραση σε επίδραση (υπόθεση που χρησιμοποιήσαμε στην προηγούμενη ενότητα). Συγκεκριμένα, στην προηγούμενη ενότητα υποθέταμε ότι ο μέσος του Y_{ij} άλλαζε κατά μια ποσότητα α_i γιατί το Y_{ij} προερχόταν από το μπλοκ j ανεξάρτητα από το ποιά "επίδραση" χρησιμοποιούνταν. Η μη προσθετικότητα των επιρροών από την χρησιμοποίηση των μπλοκ (*nonadditivity of block effects*) εκφράζεται από τον όρο γ_{ij} που αναφέρεται στην αλληλεπίδραση (*interaction*) (ή,

στην μη προσθετικότητα (*non additivity*) μεταξύ της "επίδρασης" i και του μπλοκ j), δηλαδή

$$Y_{ijm} = \mu_i + \beta_j + \gamma_{ij} + \varepsilon_{ijm} \text{ ή, αν } \mu = \frac{\sum \mu_i}{k}$$

$$Y_{ijm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijm} .$$

Υποθέτουμε ότι,

$$\alpha) \varepsilon_{ijm} \text{ είναι ανεξάρτητα } N(0, \sigma^2)$$

$$\beta) \sum \alpha_i = 0 \quad \sum \beta_j = 0 \quad \sum_i \gamma_{ij} = 0 \quad \sum_j \gamma_{ij} = 0$$

Με τις επαναληπτικές μετρήσεις είναι δυνατόν να υπολογίσουμε ένα μέσο για κάθε κυψέλη (δηλαδή για κάθε συνδυασμό δακτυλογράφου και επεξεργαστή).

Y_{ijm} είναι η m παρατήρηση της i "επίδρασης" στο j "μπλοκ". (Η m παρατήρηση ($m=1,2,3$) όταν η j δακτυλογράφος ($j=1,2,\dots,6$) χρησιμοποιεί τον i ($i=1,2,3$ επεξεργαστή)). Έτσι, γενικά, θα έχουμε

Μπλοκ	Επίδραση (Treatment)			
	1	2	...	K
1	Y_{111}	Y_{211}		Y_{k11}
	Y_{112}	Y_{212}		Y_{k12}
	.	.		.
	.	.		.
	Y_{11n}	Y_{21n}		Y_{k1n}
2	Y_{121}	Y_{221}		Y_{k21}
	Y_{122}	Y_{222}		Y_{k22}
	.	.		.
	.	.		.
	Y_{12n}	Y_{22n}		Y_{k2n}
.	.	.		.
.	.	.		.
.	.	.		.
b	Y_{1b1}	Y_{2b1}		Y_{kb1}
	Y_{1b2}	Y_{2b2}		Y_{kb2}
	.	.		.
	.	.		.
	Y_{1bn}	Y_{2bn}		Y_{kbn}

Οι σημειακές εκτιμήτριες των παραμέτρων μ , α_i , β_j , γ_{ij} είναι, αντίστοιχα

$$\hat{\mu} = \bar{Y}_{...}$$

$$\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$$

$$\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$$

$$\hat{\gamma}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$$

Οι υποθέσεις που πρέπει να ελεγχθούν είναι οι

H_0 : Όλες οι επιδράσεις είναι ίδιες

H_1 : τουλάχιστον δύο επιδράσεις διαφέρουν.

Με δεδομένες τις υποθέσεις

$$\sum \alpha_i = 0 \quad \sum \beta_j = 0 \quad \sum_i \gamma_{ij} = 0 \quad \sum_j \gamma_{ij} = 0$$

η μηδενική υπόθεση είναι ισοδύναμη με το σύνολο των υποθέσεων

$$H_{0(1)} : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

$$H_{0(2)} : \beta_1 = \beta_2 = \dots = \beta_b = 0$$

$$H_{0(3)} : \gamma_{11} = \gamma_{12} = \gamma_{21} = \dots = \gamma_{kb} = 0$$

Τα αντίστοιχα αθροίσματα των τετραγώνων θα είναι:

$$SST = \sum_{ijm} (Y_{ijm} - \bar{Y}_{...})^2 = \sum_{ijm} Y_{ijm}^2 - \frac{(Y_{...})^2}{kbn}$$

$$SSTr = nb \sum_i (\bar{Y}_{i.} - \bar{Y}_{...})^2 = \frac{\sum_i (Y_{i.})^2}{nb} - \frac{(Y_{...})^2}{kbn}$$

$$SSB = nk \sum_j (\bar{Y}_{.j} - \bar{Y}_{...})^2 = \frac{\sum_j (Y_{.j})^2}{nk} - \frac{(Y_{...})^2}{kbn}$$

$$SSI = n \sum_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{...})^2 = \frac{1}{n} \sum_{ij} Y_{ij.}^2 - \frac{1}{nk} \sum_j Y_{.j}^2 - \frac{1}{nb} \sum_i Y_{i.}^2 + \frac{(Y_{...})^2}{kbn}$$

$$SSE = \sum_{ijm} (Y_{ijm} - \bar{Y}_{ij.})^2 = \sum_{ijm} Y_{ijm}^2 - \frac{\sum_{ij} (Y_{ij.})^2}{n}$$

$$Eδώ \quad CF = \frac{(Y_{...})^2}{kbn}$$

Ο αντίστοιχος πίνακας διασποράς δίνεται παρακάτω.

ANOVA (RCB με αλληλεπιδράσεις)

Αίτια	DF	SS	MS	EMS	F
Επίδραση (Παράγων Α)	k-1	SSTr	$MSTr = \frac{SSTr}{k-1}$	$\sigma^2 + \frac{nb \sum a_i^2}{k-1}$	$\frac{MSTr}{MSE}$
Μπλοκ (Παράγων Β)	b-1	SSB	$MSB = \frac{SSB}{b-1}$	$\sigma^2 + \frac{nk \sum b_j^2}{b-1}$	$\frac{MSB}{MSE}$
Αλληλ/ση (Interaction)	(k-1)(b-1)	—	$MSI = \frac{SSI}{(k-1)(b-1)}$	$\sigma^2 + \frac{n \sum_{ij} (\gamma_{ij})^2}{(k-1)(b-1)}$	$\frac{MSI}{MSE}$
Λάθος (Error)	kb(n-1)	SSE	$MSE = \frac{SSE}{kb(n-1)}$	σ^2	
Σύνολο	kbn-1	SST			

Χρησιμοποιώντας τον πίνακα αυτό θα απορρίψουμε των $H_{0(1)}$ αν

$$F_1 = \frac{MSTr}{MSE} > F_{k-1, kb(n-1), 1-\alpha}$$

Επίσης, θα απορρίψουμε την $H_{0(2)}$ αν

$$F_2 = \frac{MSB}{MSE} > F_{b-1, kb(n-1), 1-\alpha}$$

και την $H_{0(3)}$ αν

$$F_3 = \frac{MSI}{MSE} > F_{(k-1)(b-1), kb(n-1), 1-\alpha}$$

Στο παράδειγμά μας

$$Y_{...} = 2642 \quad \sum_{ijm} Y_{ijm}^2 = 137418, \quad k=3, \quad b=6, \quad n=3$$

$$CF = \frac{(2642)^2}{6(3)(3)} = 129262.296$$

$$Y_{.1.} = Y_{11.} + Y_{21.} + Y_{31.} = 124 + 133 + 136 = 393$$

$$Y_{.2.} = 339, Y_{.3.} = 499, Y_{.4.} = 651, Y_{.5.} = 418, Y_{.6.} = 342$$

$$Y_{1..} = 845, Y_{2..} = 880, Y_{3..} = 917$$

$$Y_{11.} = 124, Y_{21.} = 133, Y_{31.} = 136$$

$$Y_{12.} = 109, Y_{22.} = 107, Y_{32.} = 123$$

$$Y_{13.} = 163, Y_{23.} = 169, Y_{33.} = 167$$

$$Y_{14.} = 203, Y_{24.} = 220, Y_{34.} = 228$$

$$Y_{15.} = 139, Y_{25.} = 136, Y_{35.} = 143$$

$$Y_{16.} = 107, Y_{26.} = 115, Y_{36.} = 120$$

$$SST = 137418 - 129262.30 = 8155.704$$

$$SSTr = \frac{1}{3(6)} \{845^2 + 880^2 + 917^2\} - CF$$

$$= 129406.333 - 129262.296$$

$$= 144.037$$

$$SSB = \frac{1}{3(3)} \{393^2 + 339^2 + 499^2 + 651^2 + 418^2 + 342^2\} - CF$$

$$= 137095.556 - 129262.296$$

$$= 7833.259$$

$$SSI = \frac{1}{3} \{124^2 + 133^2 + \dots + 120^2\} - 129406.333 - 137095.556 +$$

$$CF$$

$$= 137324 - 129406.333 - 137095.556 + 129262.296$$

$$= 84,407$$

$$SSE = 8155.704 - 144.037 - 7833.259 - 84,407 = 94,000$$

$$MSTr = 144.037 / 2 = 72,019$$

$$MSE = 94.000 / (6)(3)(3-1) = 2.611$$

$$F = \frac{MSTr}{MSE} = \frac{73.019}{2.611} = 27.58$$

Από τους πίνακες έχουμε

$$F_{2, 36, 0.95} = 3.316$$

Επομένως η $H_{0(1)}$ απορρίπτεται στο $\alpha = 0.05$ επίπεδο σημαντικότητας.