ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

ΔΙΑΛΕΞΗ: ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Διδάσκουσα: Ε. Γάκη, Επίκ. Καθηγήτρια

Περιεχόμενα Ενότητας

- Ανάλυση Διακύμανσης
  - Κατά έναν Παράγοντα
  - Κατά δύο Παράγοντες
    - Με μία παρατήρηση ανά κυψελίδα
    - Με m παρατηρήσεις ανά κυψελίδα

Περιεχόμενα Μαθήματος

- Ανάλυση Διακύμανσης
  - Κατά δύο Παράγοντες
    - Με μία παρατήρηση ανά κυψελίδα
    - Με m παρατηρήσεις ανά κυψελίδα

Ανάλυση Διακύμανσης

Η ανάλυση διακύμανσης (analysis of variance) είναι μια τεχνική που χρησιμεύει ως βάση για τη στατιστική ανάλυση μοντέλων που συνδέονται με τις **μεθόδους πειραματικών σχεδιασμών (experimental designs)**.

Ειδικότερα η ανάλυση διακύμανσης ασχολείται με τον προσδιορισμό των πηγών της μεταβλητότητας που παρατηρείται στα δειγματικά δεδομένα.

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες

Η ανάλυση διακύμανσης κατά δύο παράγοντες είναι μια επέκταση της σύγκρισης των μέσων δύο πληθυσμών όταν τα δείγματα που χρησιμοποιούμε έχουν επιλεγεί κατά ζεύγη.

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες

με μία Παρατήρηση ανά Κυψελίδα

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

- Προσφέρεται όταν οι μονάδες που πρόκειται να υποβληθούν σε κάποια αγωγή είναι ανομοιογενείς και θεωρούμε χρήσιμο να τις ομαδοποιήσουμε σε μπλοκ που έχουν παρόμοια χαρακτηριστικά. Τέτοιες ομαδοποιήσεις γίνονται συνήθως σε ιατρικά πειράματα που ομαδοποιούμε ασθενείς με παρόμοιες ασθένειες, ηλικίες, φύλλο κλπ.

- Είναι εφαρμόσιμη όταν είναι δυνατόν να υποθέσουμε ότι δεν υφίσταται αλληλεπίδραση μεταξύ των μπλοκ και των αγωγών που χρησιμοποιούνται στο πείραμα.

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Οι έλεγχοι που κάνουμε είναι:

α) Έλεγχος της υπόθεσης ότι οι αγωγές ασκούν ή όχι επίδραση:

H_o: $\mu_{1.} = \mu_{2.} = \ldots\ldots = \mu_{\kappa.}$
H_1: Δύο, τουλάχιστον, από τους μέσους διαφέρουν

β) Έλεγχος της υπόθεσης ότι τα μπλοκ που χρησιμοποιήθηκαν δεν διαφέρουν στατιστικά μεταξύ τους. Στην περίπτωση αυτή η ισοδύναμη υπόθεση που θα πρέπει να ελέγξουμε είναι η

H_o: $\mu_{.1} = \mu_{.2} = \ldots\ldots = \mu_{.b}$
H_1: Δύο, τουλάχιστον, από τους μέσους διαφέρουν

Υποθέτουμε ότι:

- οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα έχουν κοινή διακύμανση σ^2
- οι αποκλίσεις των παρατηρήσεων είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διακύμανση σ^2 .

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Στην ανάλυση που θα ακολουθήσει **ο ένας παράγοντας (factor)** είναι η **αγωγή (treatment)** και θα αντιστοιχεί στις στήλες του σχεδιασμού, ενώ ο άλλος παράγοντας είναι **τα μπλοκ (blocks)** που θα αντιστοιχούν στις γραμμές του σχεδιασμού

Τότε θα έχουμε συνοπτικά ότι:

ΜΠΛΟΚ	Α Γ Ω Γ Ε Σ							
	1	2	. . .	i	. . .	k	Σύνολα	Μέσοι
1	Y_{11}	Y_{21}	. . .	Y_{i1}	. . .	Y_{k1}	$Y_{.1}$	$\bar{Y}_{.1}$
2	Y_{12}	Y_{22}	. . .	Y_{i2}	. . .	Y_{k2}	$Y_{.2}$	$\bar{Y}_{.2}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
b	Y_{1b}	Y_{2b}	. . .	Y_{ib}	. . .	Y_{kb}	$Y_{.b}$	$\bar{Y}_{.i}$
Σύνολα	$Y_{1.}$	$Y_{2.}$	. . .	$Y_{i.}$	. . .	$Y_{k.}$	$Y_{..}$	
Μέσοι	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	. . .	$\bar{Y}_{i.}$	. . .	$\bar{Y}_{k.}$		$\bar{Y}_{..}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Η αντικατάσταση ενός δείκτη από τελεία υποδηλώνει άθροιση ως προς το δείκτη αυτό. Συγκεκριμένα:

$$Y_{i.} = \sum_{r=1}^{b} Y_{ir} \qquad \overline{Y_{i.}} = Y_{i.} / b$$

$$Y_{.r} = \sum_{i=1}^{k} Y_{ir} \qquad \overline{Y_{.r}} = Y_{.r} / k$$

$$Y_{..} = \sum_{r=1}^{b} Y_{.r} = \sum_{i=1}^{k} Y_{i.} = \sum_{i=1}^{k} \sum_{r=1}^{b} Y_{ir} \qquad \overline{Y_{..}} = Y_{..} / kb$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Ένα μέτρο της συνολικής μεταβλητότητας μπορεί να υπολογισθεί αθροίζοντας τα τετράγωνα των αποκλίσεων κάθε παρατήρησης από τον μέσο $\bar{Y}_{..}$ ο οποίος βασίζεται σε όλες τις παρατηρήσεις. Με άλλα λόγια, η συνολική μεταβλητότητα, ή συνολικό άθροισμα τετραγώνων SST, υπολογίζεται με βάση τον τύπο

$$SST = \sum_{i=1}^{k}\sum_{r=1}^{b}\left(Y_{ir} - \bar{Y}_{..}\right)^2 = \sum_{i=1}^{k}\sum_{r=1}^{b} Y_{ir}^2 - \frac{(Y_{..})^2}{kb}$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Η συνολική μεταβλητότητα μπορεί να αναλυθεί σε τρεις συνιστώσες, εκ των οποίων:

- η πρώτη εκφράζει τη μεταβλητότητα μεταξύ των αγωγών (between treatments variation)
- η δεύτερη εκφράζει τη μεταβλητότητα μεταξύ των μπλοκ (between blocks variation) και
- η τρίτη την ανερμήνευτη μεταβλητότητα (error variation).

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

SST = SSTr + SSBl + SSE

$$SST = \sum_{i=1}^{k}\sum_{r=1}^{b}\left(Y_{ir} - \overline{Y}_{..}\right)^2 = \sum_{i=1}^{k}\sum_{r=1}^{b} Y_{ir}^2 - \frac{(Y_{..})^2}{kb}$$

$$SSTr = \sum_{i=1}^{k}\sum_{r=1}^{b}\left(Y_{ir} - \overline{Y}_{..}\right)^2 = b\sum_{i=1}^{k}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 = \frac{\sum_{i}\left(Y_{i.}\right)^2}{b} - \frac{(Y_{..})^2}{kb}$$

$$SSBl = \sum_{i=1}^{k}\sum_{r=1}^{b}\left(\overline{Y}_{ir} - \overline{Y}_{..}\right)^2 = k\sum_{r=1}^{b}(\overline{Y}_{ir} - \overline{Y}_{..})^2 = \frac{\sum_{r}(Y_{.r})^2}{k} - \frac{(Y_{..})^2}{kb}$$

$$SSE = \sum_{i=1}^{k}\sum_{r=1}^{b} Y_{ir}^2 - \frac{\sum_{i}\left(Y_{i.}\right)^2}{b} - \frac{\sum_{r}\left(Y_{.r}\right)^2}{k} + \frac{\left(Y_{..}\right)^2}{kb}$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Οι εκτιμήσεις για τις αντίστοιχες διακυμάνσεις είναι:

$$MSTr = \frac{SSTr}{k - 1}$$

$$MSBl = \frac{SSBl}{b - 1}$$

$$MSE = \frac{SSE}{(k - 1)(b - 1)}$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

α) Έλεγχος της υπόθεσης ότι οι αγωγές ασκούν ή όχι επίδραση:

H_o: $\mu_{1.} = \mu_{2.} = \ldots\ldots = \mu_{\kappa.}$
H_1: Δύο, τουλάχιστον, από τους μέσους διαφέρουν

Συνάρτηση Ελέγχου: $$F_o = \frac{MSTr}{MSE} \sim F_{k-1,(b-1)(k-1)}$$

Κριτήριο Ελέγχου:
Απορρίπτουμε την H_0 αν: $F_o > F_{k-1,\ (b-1)(k-1),\ 1-a}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

β) Έλεγχος της υπόθεσης ότι τα μπλοκ που χρησιμοποιήθηκαν δεν διαφέρουν στατιστικά μεταξύ τους.

H_o: $\mu_{.1} = \mu_{.2} = \ldots.. = \mu_{.b}$
H_1: Δύο, τουλάχιστον, από τους μέσους διαφέρουν

Συνάρτηση Ελέγχου: $$F_o = \frac{MSBl}{MSE} \sim F_{b-1,(b-1)(k-1)}$$

Κριτήριο Ελέγχου:
Απορρίπτουμε την H_0 αν: $F_0 > F_{b-1,\ (b-1)(k-1),\ 1-a}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με μία Παρατήρηση ανά Κυψελίδα

Πίνακας Ανάλυσης Διακύμανσης κατά Δύο Παράγοντες
με Μια Παρατήρηση ανά Κυψελίδα

Πηγή Μεταβλητότητας	SS	DF	MS	F (κάτω από την H_0)
Μεταξύ Αγωγών	SSTr	k-1	SSTr/(k-1)=MSTr	$MSTr/MSE=F_{Tr}$
Μεταξύ Blocks	SSBl	b-1	SSBl/(b-1)=MSBl	$MSBl/MSE=F_{Bl}$
Σφάλμα	SSE	kb-(k+b-1) = = (b-1)(k-1)	SSE/(b-1)(k-1)=MSE	
Σύνολο	SST	kb-1		

Παράδειγμα

Τα δάνεια που χορηγεί μία Τράπεζα κατατάσσονται ανάλογα με τη χρήση για την οποία προορίζονται σε Εμπορικά, Βιομηχανικά και Αγροτικά και ανάλογα με τη διάρκεια τους σε Βραχυπρόθεσμα, Μεσοπρόθεσμα και Μακροπρόθεσμα. Στον Πίνακα εμφανίζεται το ύψος ανά κατηγορία (σε εκ. δρχ.) ενός τυχαίου δείγματος δανείων που χορηγήθηκαν από τη συγκεκριμένη Τράπεζα κατά τη διάρκεια του προηγούμενου οικονομικού έτους.

Είδος / Διάρκεια	Εμπορικά	Βιομηχανικά	Αγροτικά
Βραχυπρόθεσμα	17	9	3
Μεσοπρόθεσμα	4	10	4
Μακροπρόθεσμα	2	10	4

Παράδειγμα

Με βάση τα στοιχεία αυτά:

- Να δημιουργηθεί ο Πίνακας Ανάλυσης Διακύμανσης.
- Να ελεγχθεί σε επίπεδο σημαντικότητας 0,05 η υπόθεση ότι το ύψος του δανείου δεν επηρεάζεται από τη χρήση για την οποία προορίζεται.
- Να ελεγχθεί σε επίπεδο σημαντικότητας 0,05 η υπόθεση ότι το ύψος του δανείου δεν επηρεάζεται από τη διάρκεια του.

Σημείωση: Για τη διευκόλυνση των υπολογισμών δίνονται:

$$\sum_i \left(y_{i.}\right)^2 = 1491 \qquad \sum_{i=1}^{3}\sum_{r=1}^{3} y_{ir}^2 = 631$$

$$\sum_i \left(y_{.r}\right)^2 = 1421 \qquad y_{..} = 63$$

Παράδειγμα

	ΕΙΔΟΣ ΔΑΝΕΙΟΥ				
ΔΙΑΡΚΕΙΑ ΔΑΝΕΙΟΥ	*E_1*	**E_2**	**E_3**	**ΣΥΝΟΛΑ**	**ΜΕΣΟΙ**
Δ_1	17	9	3	$Y_{.1}=29$	$\overline{Y}_{.1} = 9.7$
Δ_2	4	10	4	$Y_{.2}=18$	$\overline{Y}_{.2} = 6.0$
Δ_3	2	10	4	$Y_{.3}=16$	$\overline{Y}_{.3} = 5.3$
ΣΥΝΟΛΑ	$Y_{1.}=23$	$Y_{2.}=29$	$Y_{3.}=11$	$Y_{..}=63$	
ΜΕΣΟΙ	$\overline{Y}_{1.} = 7.7$	$\overline{Y}_{2.} = 9.7$	$\overline{Y}_{3.} = 3.7$		$\overline{Y}_{..} = 7$

Παράδειγμα

SST = SSTr + SSBl + SSE   (1)

$$\sum_{i=1}^{k}\sum_{r=1}^{b} y_{ir}^2 = 17^2 + 9^2 + 3^2 + ... + 10^2 + 4^2 = 631$$ k = 3, b = 3

$$\frac{(y_{..})^2}{kb} = \frac{63^2}{3*3} = 441$$

$$\sum_i (Y_{i.})^2 = 23^2 + 29^2 + 11^2 = 1491$$

$$\sum_r (Y_{.r})^2 = 29^2 + 18^2 + 16^2 + 40^2 = 1421$$

Παράδειγμα

Συνολική μεταβλητότητα : SST = $\sum_{i=1}^{3}\sum_{r=1}^{3} y_{ir}^2 - \frac{(y_{..})^2}{kb} = 631 - 441 \Rightarrow SST = 190$

STr = $\frac{\sum_i (Y_{i.})^2}{b} - \frac{(Y_{..})^2}{kb} = \frac{1491}{3} - 441 \Rightarrow SST_r = 56$

SBl = $\frac{\sum_r (Y_{.r})^2}{k} - \frac{(Y_{..})^2}{kb} = \frac{1421}{3} - 441 \Rightarrow SSBl = 32.7$

(1) $\Rightarrow$ SSE = SST - SSTr - SSBl = 190 - 56 -32.7 $\Rightarrow$ SSE = 101.3

ή SSE = $\sum_{\iota=1}^{\kappa}\sum_{r=1}^{b} y_{ir}^2 - \frac{\sum_i (y_{i.})^2}{b} - \frac{\sum_r (y_{.r})^2}{k} + \frac{(y_{..})^2}{kb} \Rightarrow$

$$SSE = 631 - \frac{1491}{3} - \frac{1421}{3} + 441 \Rightarrow SSE = 101.3$$

Παράδειγμα

Διακυμάνσεις:

$$MSTr = \frac{SSTr}{k-1} = \frac{56}{2} = 28 \Rightarrow MSTr = 28$$

$$MSBl = \frac{SSBl}{b-1} = \frac{32.7}{2} \Rightarrow MSBl = 16.35$$

$$MSE = \frac{SSE}{(k-1)(b-1)} = 1.1, F_{Bl} = \frac{MSBl}{MSE} = 0,65$$

Παράδειγμα

Πίνακας Ανάλυσης Διακύμανσης

ΠΗΓΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ	SS	DF	MS	F
ΜΕΤΑΞΥ ΑΓΩΓΩΝ	56	(κ-1) =3-1=2	28	1.1
ΜΕΤΑΞΥ BLOCKS	32.7	(b-1) =3-1=2	16.35	0.65
ΛΑΘΟΣ	101.3	(b-1)(k-1)=(3-1)(3-1)=4	25.3	
ΣΥΝΟΛΟ	190	Kb-1=3*3-1=8		

Παράδειγμα

Από τον Πίνακα έχουμε ότι

F_{Tr} = 1.1 και F_{Bl} = 0.65 επίσης γνωρίζουμε k = b = 3 , α = 0.05

α) Πρώτη υπόθεση : Το ύψος του δανείου δεν επηρεάζεται από το είδος του

Για να δεχθούμε την H_0 πρέπει $F_{Tr} < F_{k-1, (b-1)(k-1), 1-\alpha}$

$$\left.\begin{array}{l} \mathrm{F}_{\mathrm{Tr}} = 1.1 \\ \mathrm{F}_{\mathrm{k-1,(b-1)(k-1),1-á}} = 2{,}4{,}0.95 = 6.94 \end{array}\right\} \Rightarrow F_{Tr} < F_{2,4,0.95}$$

Δεχόμαστε την H_0. Άρα το ύψος του δανείου δεν επηρεάζεται από το είδος του.

β) Δεύτερη υπόθεση : Το ύψος του δανείου δεν επηρεάζεται από τη διάρκεια του

Για να απορρίψουμε την H_0 πρέπει $F_{B1} > F_{(b-1), (k-1)(b-1), 1-\alpha}$

$$\left.\begin{array}{l} \mathrm{F}_{\mathrm{B1}} = 0.65 \\ \mathrm{F}_{\mathrm{(b-1),(k-1)(b-1),1-á}} = 2{,}4{,}0.95 = 6.94 \end{array}\right\} \Rightarrow F_{B1} < F_{2,4,0.95}$$

Δεχόμαστε την H_0. Άρα το ύψος του δανείου δεν επηρεάζεται από τη διάρκεια του.

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες

με m Παρατηρήσεις ανά Κυψελίδα

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

- Είναι εφαρμόσιμη όταν είναι δυνατόν να υποθέσουμε ότι υφίσταται αλληλεπίδραση μεταξύ των μπλοκ και των αγωγών που χρησιμοποιούνται στο πείραμα.

- Ο ένας παράγοντας (factor) είναι η αγωγή (treatment) και θα αντιστοιχεί στις στήλες του σχεδιασμού, ενώ ο άλλος παράγοντας είναι τα μπλοκ (blocks) που θα αντιστοιχούν στις γραμμές του σχεδιασμού.

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

	Επίδραση				
Μπλοκ	**1**	**2**	**i.**	**k**	
1	Y_{111}	Y_{211}	Y_{i11}	Y_{k11}	$Y_{.11}$
	Y_{112}	Y_{212}	Y_{i12}	Y_{k12}	$Y_{.12}$
j	:	:	:	:	:
m	Y_{11m}	Y_{21m}	Y_{i1m}	Y_{k1m}	$Y_{.1m}$
	$Y_{11.}$	$Y_{21.}$	$Y_{i1.}$	$Y_{k1.}$	$Y_{.1.}$
2	Y_{121}	Y_{221}		Y_{k21}	$Y_{.21}$
	Y_{122}	Y_{222}		Y_{k22}	$Y_{.22}$
j	:	:		:	:
	Y_{12m}	Y_{22m}		Y_{k2m}	
	$Y_{12.}$	$Y_{22.}$	$Y_{i2.}$	$Y_{k2.}$	$Y_{.2.}$
r	Y_{1r1}	Y_{2r1}	Y_{ir1}	Y_{kr1}	$Y_{.r1}$
	Y_{1r2}	Y_{2r2}	Y_{ir2}	Y_{kr2}	$Y_{.r2}$
j	:	:	:	:	:
	Y_{1rm}	Y_{2rm}	Y_{irm}	Y_{krm}	$Y_{.rm}$
	$Y_{1r.}$	$Y_{2r.}$	$Y_{ir.}$	$Y_{kr.}$	$Y_{.r.}$
b	Y_{1b1}	Y_{2b1}		Y_{kb1}	
	Y_{1b2}	Y_{2b2}		Y_{kb2}	
j	:	:		:	
m	Y_{1bm}	Y_{2bm}		Y_{kbm}	
	$Y_{1b.}$	$Y_{2b.}$	$Y_{ib.}$	$Y_{kb.}$	$Y_{.b.}$
	$Y_{1..}$	$Y_{2..}$	$Y_{i..}$	$Y_{k..}$	$Y_{...}$
	$\bar{Y}_{1..}$	$\bar{Y}_{2..}$	$\bar{Y}_{i..}$	$\bar{Y}_{k..}$	$\bar{Y}_{...}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Αποδεικνύεται ότι η μεταβλητότητα μπορεί να αναλυθεί σε τέσσερις συνιστώσες:

- το άθροισμα τετραγώνων που οφείλεται στις αγωγές (sum of squares due to treatments),
- το άθροισμα τετραγώνων που οφείλεται στα μπλοκ (sum of squares due to blocks),
- το άθροισμα των τετραγώνων που οφείλεται στην αλληλεπίδραση αγωγής και μπλοκ (sum of squares due to interaction) και
- το άθροισμα τετραγώνων των λαθών (error sum of squares).

Δηλαδή:

$$SST = SSTr + SSBl + SSI + SSE$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

$$SST = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m}\left(Y_{irj} - \overline{Y}_{..}\right)^2 = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m} Y_{irj}^2 - \frac{\left(Y_{...}^2\right)}{kbm}$$

$$SSTr = mb\sum_{i=1}^{k}\left(\overline{Y}_{i..} - Y_{...}\right)^2 = \frac{\sum_{i}\left(Y_{i..}\right)^2}{mb} - \frac{\left(Y_{...}\right)^2}{kmb}$$

$$SSBl = mk\sum_{r=1}^{b}\left(\overline{Y}_{.r.} - \overline{Y}_{...}\right)^2 = \frac{\sum_{r}\left(Y_{.r.}\right)^2}{mk} - \frac{\left(Y_{...}\right)^2}{kbm}$$

$$SSI = m\sum_{i=1}^{k}\sum_{j=1}^{b}\left(\overline{Y}_{ir.} - \overline{Y}_{i..} - \overline{Y}_{...}\right)^2 = \frac{1}{m}\sum_{i=1}^{k}\sum_{r=1}^{b} Y_{ir.}^2 - \frac{1}{mk}\sum_{r=1}^{b} Y_{.r.}^2 - \frac{1}{mk}\sum_{i=1}^{k} Y_{i..}^2 + \frac{\left(Y_{...}\right)^2}{kbm}$$

$$SSE = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m} Y_{irj}^2 - \sum_{i=1}^{k}\sum_{j=1}^{b}\frac{\left(Y_{ir.}\right)^2}{m} = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m}\left(Y_{irj} - \overline{Y}_{ir.}\right)^2 = Y_{irj}^2 - \frac{\left(Y_{ir.}\right)^2}{m}$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Αν τα SSTr, SSBl, SSI και SSE διαιρεθούν με τους αντίστοιχους βαθμούς ελευθερίας θα προκύψουν εκτιμήσεις για τις αντίστοιχες διακυμάνσεις. Συγκεκριμένα,

$$MSTr = \frac{SSTr}{k-1} \qquad MSBl = \frac{SSBl}{b-1}$$

$$MSE = \frac{SSE}{kb(m-1)} \qquad MSI = \frac{SSI}{(b-1)(k-1)}$$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

<u>Έλεγχος της υπόθεσης ότι όλες οι αγωγές ασκούν την ίδια επίδραση:</u>

H_0: Όλοι οι μέσοι είναι ίσοι ($\mu_{1..}=\mu_{2..}= \ldots =\mu_{\kappa..}$)
H_1: Δύο τουλάχιστον από τους μέσους διαφέρουν ($\mu_{i..} \neq \mu_{j..}$, για $i \neq j$, $i,j=1,2,\ldots,k$)

Συνάρτηση Ελέγχου: $F_{Tr} = \dfrac{MSTr}{MSE}$

Κριτήριο Ελέγχου: Απορρίπτουμε την H_0 αν: $F_{Tr} > F_{k-1,\, kb(m-1),\, 1-\alpha}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Έλεγχος της υπόθεσης ότι όλα τα blocks ασκούν την ίδια επίδραση:

H_0: Όλοι οι μέσοι είναι ίσοι ($\mu_{.1.} = \mu_{.2.} = \ldots = \mu_{.\kappa.}$)

H_1: Δύο τουλάχιστον από τους μέσους διαφέρουν ($\mu_{.i.} \neq \mu_{.j.}$, για $i \neq j$, $i,j=1,2,\ldots,k$)

Συνάρτηση Ελέγχου: $F_{Bl} = \dfrac{MSBl}{MSE}$

Κριτήριο Ελέγχου: Απορρίπτουμε την H_0 αν: $F_{Bl} > F_{b-1,\, kb(m-1),\, 1-\alpha}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Έλεγχος της υπόθεσης ότι δεν υπάρχει αλληλεπίδραση αγωγών και blocks:

H_0: Δεν υπάρχει αλληλεπίδραση αγωγών και blocks
H_1: Υπάρχει αλληλεπίδραση αγωγών και blocks

Συνάρτηση Ελέγχου: $F_I = \dfrac{MSI}{MSE}$

Κριτήριο Ελέγχου: Απορρίπτουμε την H_0 αν: $F_I > F_{(k-1)(b-1),\, kb(m-1),\, 1-\alpha}$

Ανάλυση Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Πίνακας Ανάλυσης Διακύμανσης κατά Δύο Παράγοντες με m Παρατηρήσεις ανά Κυψελίδα

Πηγή	DF	SS	MS	F
Αγωγή (παράγων A)	k-1	SSTr	MSTr=SSTr/k-1	F_1=MSTr/MSE
Μπλοκ (Παράγων B)	b-1	SSBl	MSBI=SSBI/b-1	F_2=MSBI/MSE
Αλληλοεπίδραση (Interaction)	(k-1)(b-1)	SSI	MSI=SSI/(k-1)(b-1)	F_3=MSI/MSE
Λάθος (Error)	Kb(m-1)	SSE	MSE=SSE/kb(m-1)	
Σύνολο	kbm - 1	SST		

Παράδειγμα

Ένας δείκτης υγείας μιας οικονομίας είναι η ικανότητά της να δημιουργεί θέσεις εργασίας, και μια όψη αυτού του δείκτη είναι ο αριθμός των διαφορετικών επαγγελμάτων που κάνει κάθε εργαζόμενος στη διάρκεια της ζωής του. Για τις ανάγκες μιας σχετικής έρευνας (πηγή: Statistical Abstract of the United States) επελέγη τυχαίο δείγμα 80 Αμερικανών 37-45 ετών και καταγράφηκε το φύλο, το επίπεδο εκπαίδευσης και ο συνολικός αριθμός των διαφορετικών επαγγελμάτων που έχουν εξασκήσει μέχρι στιγμής. Το επίπεδο εκπαίδευσης κωδικοποιήθηκε ως εξής:

Κωδικός	Περιγραφή
E1	Δεν ολοκλήρωσε τη δευτεροβάθμια εκπαίδευση
E2	Ολοκλήρωσε τη δευτεροβάθμια εκπαίδευση
E3	Κάποιες σπουδές σε πανεπιστήμιο - κολέγιο χωρίς πτυχίο
E4	Πτυχιούχος πανεπιστημίου – κολεγίου

Παράδειγμα

Επιπλέον είναι γνωστό ότι το δείγμα ήταν μοιρασμένο ισομερώς σε άνδρες και γυναίκες, ενώ από κάθε επίπεδο εκπαίδευσης υπήρχε ίδιος αριθμός περιπτώσεων.

Μια πρώτη επεξεργασία των δεδομένων συνοψίζεται στον παρακάτω πίνακα.

		Επίπεδο Εκπαίδευσης			
		E1	E2	E3	E4
ΑΝΔΡΕΣ	Άθροισμα	126	110	106	90
ΓΥΝΑΙΚΕΣ	Μέση τιμή	11,5	11,2	9,4	8,1

Παράδειγμα

Με βάση τα στοιχεία αυτά:

α. Να δημιουργηθεί ο Πίνακας Ανάλυσης Διακύμανσης.

β. Να ελεγχθεί εάν υπάρχει αλληλεπίδραση μεταξύ φύλου επιπέδων εκπαίδευσης;

Σημείωση 1: Για διευκόλυνση των υπολογισμών δίνονται: $\sum_{i=1}^{4}\sum_{r=1}^{2}\sum_{j=1}^{10} Y_{irj}^2 = 9574$ $\sum_{r=1}^{2}\left(\mathrm{Y}_{.r.}\right)^2 = 348228$

$$\sum_{i=1}^{4}\sum_{r=1}^{2}\left(\mathrm{Y}_{ir.}\right)^2 = 88478$$

Σημείωση 2: Αγωγές θεωρούνται τα επίπεδα εκπαίδευσης και Blocks το φύλο. Επίπεδο σημαντικότητας α = 0,05.

Σημείωση 3: Για τη χρήση των πινάκων μπορείτε να στρογγυλοποιήσετε στην πλησιέστερη τιμή

Παράδειγμα

		Επίπεδο Εκπαίδευσης					
ΦΥΛΟ		E1	E2	E3	E4		
	ΑΝΔΡΕΣ					y.r.	mesos y.r.
	yir.	126	110	106	90	432	10,8
	mesos yir.	12,60	11,00	10,60	9,00		
	ΓΥΝΑΙΚΕΣ					y.r.	mesos y.r.
	yir.	115,00	112,00	94,00	81,00	402,00	10,05
	mesos yir.	11,50	11,20	9,40	8,10		
	yi..	241	222	200	171	y…	834
	mesos yi..	12,05	11,10	10,00	8,55		

Παράδειγμα

$$\sum_{i=1}^{4}\sum_{r=1}^{2}\sum_{j=1}^{10}Y_{irj}^2 = 9574$$

$$\frac{(Y_{...})^2}{kbm} = \frac{834^2}{4*2*10} = \frac{695556}{80} = 8694,45$$

$$\frac{\sum_i (Y_{i..})^2}{mb} = \frac{241^2 + 222^2 + 200^2 + 171^2}{10*2} = \frac{58081 + 49284 + 40000 + 29241}{20} = \frac{176606}{20} = 8830.3$$

$$\frac{\sum_r (Y_{.r.})^2}{mk} = \frac{432^2 + 402^2}{10*4} = \frac{186624 + 161604}{40} = \frac{348228}{40} = 8705.7$$

$$\sum_{i=1}^{k}\sum_{j=1}^{b}\frac{(Y_{ir.})^2}{m} = \frac{126^2 + 110^2 + 106^2 + 90^2 + 115^2 + 112^2 + 94^2 + 81^2}{10} = \frac{15876+12100+11236+8100+13225+12544+8836+6561}{10}$$

$$= \frac{88478}{10} = 8847.8$$

Παράδειγμα

$$SST = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m} Y_{irj}^2 - \frac{\left(Y_{...}^2\right)}{kbm} = 9574 - 8694.45 = 879.55$$

$$SSTr = \frac{\sum_i \left(Y_{i..}\right)^2}{mb} - \frac{\left(Y_{...}\right)^2}{kmb} = 8830.3 - 8694.45 = 135.85$$

$$SSBl = \frac{\sum_r \left(Y_{.r.}\right)^2}{mk} - \frac{\left(Y_{...}\right)^2}{kbm} = 8705.7 - 8694.45 = 11.25$$

$$SSE = \sum_{i=1}^{k}\sum_{r=1}^{b}\sum_{j=1}^{m} Y_{irj}^2 - \sum_{i=1}^{k}\sum_{j=1}^{b} \frac{\left(Y_{ir.}\right)^2}{m} = 9574 - 8847.8 = 726.2$$

$$SSI = SST - SSTr - SSBl - SSE = 879.55 - 135.85 - 11.25 - 726.20 = 6.25$$

Παράδειγμα

$$MSTr = \frac{SSTr}{k-1} = \frac{135.85}{4-1} = 45.28$$

$$MSBl = \frac{SSBl}{b-1} = \frac{11.25}{2\text{-}1} = 11.25$$

$$MSI = \frac{SSI}{(b-1)(k-1)} = \frac{6.25}{(2-1)(4-1)} = 2.08$$

$$MSE = \frac{SSE}{kb(m-1)} = \frac{726.20}{4*2(10-1)} = 10.09$$

$$\mathrm{F_{Tr}} = \frac{\mathrm{MSTr}}{\mathrm{MSE}} = 2.7318$$

$$\mathrm{F_{Bl}} = \frac{\mathrm{MSBl}}{\mathrm{MSE}} = 3.9739$$

$$\mathrm{F_{I}} = \frac{\mathrm{MSI}}{\mathrm{MSE}} = 2.7318$$

Παράδειγμα

Πίνακας Ανάλυσης Διακύμανσης

Πηγή	DF	SS	MS	F
Μεταξύ Αγωγών	3	135,85	45,28	F_{Tr}=2,7318
Μεταξύ Blocks	1	11,25	11,25	F_{Bl}=3,9739
Αλληλοεπίδραση	3	6,25	2,08	F_I=2,7318
Σφάλμα	72	726,20	10,09	
Σύνολο	79	879,55		

<u>Έλεγχος της υπόθεσης ότι δεν υπάρχει αλληλεπίδραση:</u>

H_0: Δεν υπάρχει αλληλεπίδραση

H_1: Υπάρχει αλληλεπίδραση

Συνάρτηση Ελέγχου: $F_I = \frac{MSI}{MSE} = 2.7318$

Κριτήριο Ελέγχου: Απορρίπτουμε την H_0 αν: $F_I > F_{(k-1)(b-1),\ kb(m-1),\ 1-a}$

$F_{(k-1)(b-1),\ kb(m-1),\ 1-a} = F_{3,72,0,95} = F_{3,60,0,95} = 8,572$

F_I = 2,7318<8,572 άρα δεν απορρίπτουμε την H_0

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

διοικώ σημαίνει αποφασίζω