Άσκηση (ΑΒΕΖ) (SOS) ΚΕΦ. 6^{o}

Η εταιρεία ζυμαρικών ΑΒΕΖ ενδιαφέρεται να διαπιστώσει κατά πόσο οι κάτοικοι μιας περιοχής δείχναν ιδιαίτερη προτίμηση ανάμεσα σε 4 τύπους ζυμαρικών που παράγει. Για το σκοπό αυτό γίνεται τυχαία επιλογή 10 καταστημάτων τροφίμων σε καθένα από τα οποία δίνεται για πώληση ένας από τους τέσσερις τύπους ζυμαρικών. Τελικά, δύο από τους τέσσερις τύπους ζυμαρικών δίνονται σε τρία από τα δέκα καταστήματα και οι υπόλοιποι δύο σε δύο από τα καταστήματα ο κάθε τύπος (όλοι οι εξωτερικοί παράγοντες που μπορούν να επιδράσουν για κάθε κατάστημα είναι ίδιοι)

Οι ποσότητες ζυμαρικών που πωλήθηκαν ανά κατάστημα και τύπο, σε διάστημα μιας βδομάδας δίνονται στον ακόλουθο πίνακα.

	Τύπος Ζυμαρικών				
Καταστήματα τροφίμων	1	2	3	4	ΣΥΝΟΛΟ
1	12	14	19	24	
2	18	12	17	30	
3		13	21		
ΣΥΝΟΛΑ	30	39	57	54	180
Δειγματικοί Μέσοι	15	13	19	27	18
Αριθμός Καταστημάτων	2	3	3	2	10

i) Να υπολογιστούν οι εκτιμητές ελαχίστων τετραγώνων για το στατιστικό μοντέλο που αφορά έναν παράγοντα

ii) Να υπολογιστούν τα υπόλοιπα (ή οι αποκλίσεις των παρατηρήσεων από τη δειγματική μέση τιμή τους) και να παρασταθούν στον αντίστοιχο πίνακα.

iii) Να γίνει στατιστική ανάλυση των δεδομένων του παραπάνω πίνακα (μέσω του πίνακα ANOVA)

ΛΥΣΗ

i) Κατά έναν παράγοντα είχαμε δείξει ότι

$Q(\mu_1,\mu_2,\mu_3,\mu_4) = \sum_j \sum_i (Y_{ij}-\mu_j)^2$

Παραγωγίζοντας τη συνάρτηση παίρναμε ότι:

$Q'(\mu_j) = -2 \sum_j \sum_i (Y_{ij}-\hat{\mu}_j) = 0 \Rightarrow \hat{\mu}_j = \overline{Y}_{\cdot j} \quad j=1,2,3,4$

Άρα, $\hat{\mu}_1 = \overline{Y}_1 = 15$, $\hat{\mu}_2 = \overline{Y}_2 = 13$, $\hat{\mu}_3 = \overline{Y}_3 = 19$, $\hat{\mu}_4 = \overline{Y}_4' = 27$

ii) $e_{ij} = Y_{ij} - \hat{\mu}_j = Y_{ij} - \overline{Y}_{\cdot j}$

Άρα, $e_{11} = Y_{11} - \hat{\mu}_1 = 12 - 15 = -3$

$e_{21} = Y_{21} - \hat{\mu}_1 = 18 - 15 = +3$

όμοια αν συνεχίσουμε προκύπτει ο πίνακας

Καταστήματα	1	2	3	4	ΣΥΝΟΛΟ
1	−3	+1	0	−3	
2	+3	−1	−2	+3	
3		0	+2		
ΣΥΝΟΛΟ	0	0	0	0	0

$\leftarrow \sum_{i=1}^{4} e_{ij} = 0,\ j = 1, \ldots, r$

iii) ΠΙΝΑΚΑΣ ΑΝΑΔΙΑ

ΠΡΟΕΛΕΥΣΗ ΜΕΤΑΒΛ.	ΑΘΡ. ΤΕΤΡ. (SS)	Β.Ε	ΜΕΣΑ ΤΕΤΡ. (MS)
ΔΟΚΙΜΕΣ	258	3	86.00
ΥΠΟΛΟΙΠΟ	46	6	7.67
ΟΛΙΚΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑ	304	9	

Στατιστική συνάρτηση:

$$F = \frac{MS_{tr}}{MS_{res}} = 11{,}2$$ και αφού $F = 11{,}2 > 4{,}76 = F_{0.05, 3, 6}$

τότε η H_0 : απορρίπτεται ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$).
Δηλαδή, οι μέσες πωλήσεις των τηλεφωνικών σε κάθε κατάστημα με επίπεδο σημαντικότητας 5% διαφέρουν.

6.3 Μια ομάδα από 28 υπέρβαρα κορίτσια αποφάσισαν λίγο πριν τις καλοκαιρινές διακοπές να χάσουν βάρος ακολουθώντας τέσσερις διαφορετικές μεθόδους δίαιτας $\Delta_1, \Delta_2, \Delta_3, \Delta_4$. Χωρίστηκαν σε 4 ίσες ομάδες και κάθε ομάδα ακολούθησε μία δίαιτα. Λόγω αποχωρήσεων (αρρώστιες, όψιμη προτίμηση βουνού αντί θάλασσας κλπ.) ο αριθμός των κοριτσιών που τελικά ολοκλήρωσαν τη δίαιτα διέφερε από ομάδα σε ομάδα. Τα δεδομένα πού ακολουθούν αποτελούν τα κιλά που έχασαν τα κορίτσια μετά από δύο μήνες δίαιτα.

Δ_1	7	9	7	8	8	9	
Δ_2	7	7	8	8	7	7	5
Δ_3	6	8	5	6	5	6	
Δ_4	8	9	10	9			

Υπάρχει διαφορά ως προς την αποτελεσματικότητα των τεσσάρων μεθόδων δίαιτας ($\alpha = 0.05$) και αν ναι, ποια μέθοδος είναι η καλύτερη;

ΛΥΣΗ

Τα κορίτσια της Δ1 είναι σε πλήθος $n_1 = 6$
" " " Δ2 " " " $n_2 = 7$
" " " Δ3 " " " $n_3 = 6$
" " " Δ4 " " " $n_4 = 4$
" " Αθροιστικά των Δ1,Δ2,Δ3,Δ4 $n = 23$
Άρα, μπορούμε να κατασκευάσουμε τον εξής Πίνακα:

ΚΟΡΙΤΣΙΑ	Τυπος Διαιτας Δ1	Δ2	Δ3	Δ4	ΣΥΝΟΛΟ
1	7	7	6	8	
2	9	7	8	9	
3	7	8	5	10	
4	8	8	6	9	
5	8	7	5		
6	9	7	6		
7		5			
ΣΥΝΟΛΟ	48	49	36	36	169
Αριθμ. Μέσοι	48/6 = 8	49/7 = 7	36/6 = 6	36/4 = 9	30
Αριθμ. κοριτσ.	6	7	6	4	23

ΑΝΑΔΙΑ

ΠΡΟΕΛΕΥΣΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ	ΑΘΡ. ΤΕΤΡΑΓ. (SS)	Β.Ε	ΜΕΣΑ ΤΕΤΡΑΓ. (MS)
ΔΙΑΙΤΕΣ	25.2174	3	8.4058
ΥΠΟΛΟΙΠΟ	18.0000	19	0.9474
ΟΛΙΚΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑ	43.2174	22	

- $SS_{tr} = \sum_{j=1}^{4} n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 = 25.2174$
- $SS_{tot} = \sum_{j=1}^{4}\sum_{i=1}^{7} (Y_{ij} - \bar{Y}_{..})^2 = 43.2174$
- $SS_{res} = SS_{tot} - SS_{tr} = 18.000$
- $MS_{tr} = \frac{SS_{tr}}{3} = 8.4058$
- $MS_{res} = \frac{SS_{res}}{19} = 0.9474$

Εστω $H_0: \mu_4 - \mu_3 = 0$ ν $H_a: \mu_4 - \mu_3 > 0$

$t_{0,05,19} = 1.729 \Rightarrow$ κρίσιμη περιοχή : $(1.729, +\infty)$

Αλλά, από το στατιστικό

$$t = \frac{\bar{Y}_{.4} - \bar{Y}_{.3}}{\sqrt{MS_{res}\left(\frac{1}{n_3} + \frac{1}{n_4}\right)}} = 4.77 > 1.729.$$ Αρα, $H_0: \mu_4 - \mu_3 = 0$ απορρίπτεται

Αρα, Ισχυει ότι $\mu_4 - \mu_3 > 0$

Δηλαδή υπάρχει διαφορά μεταξύ τως Δ3 και Δ4

(Ανάλογα γίνονται και οι υπόλοιπες συγκρίσεις)

σε επίπεδο σημαντικότητας $a = 0,05$.