ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

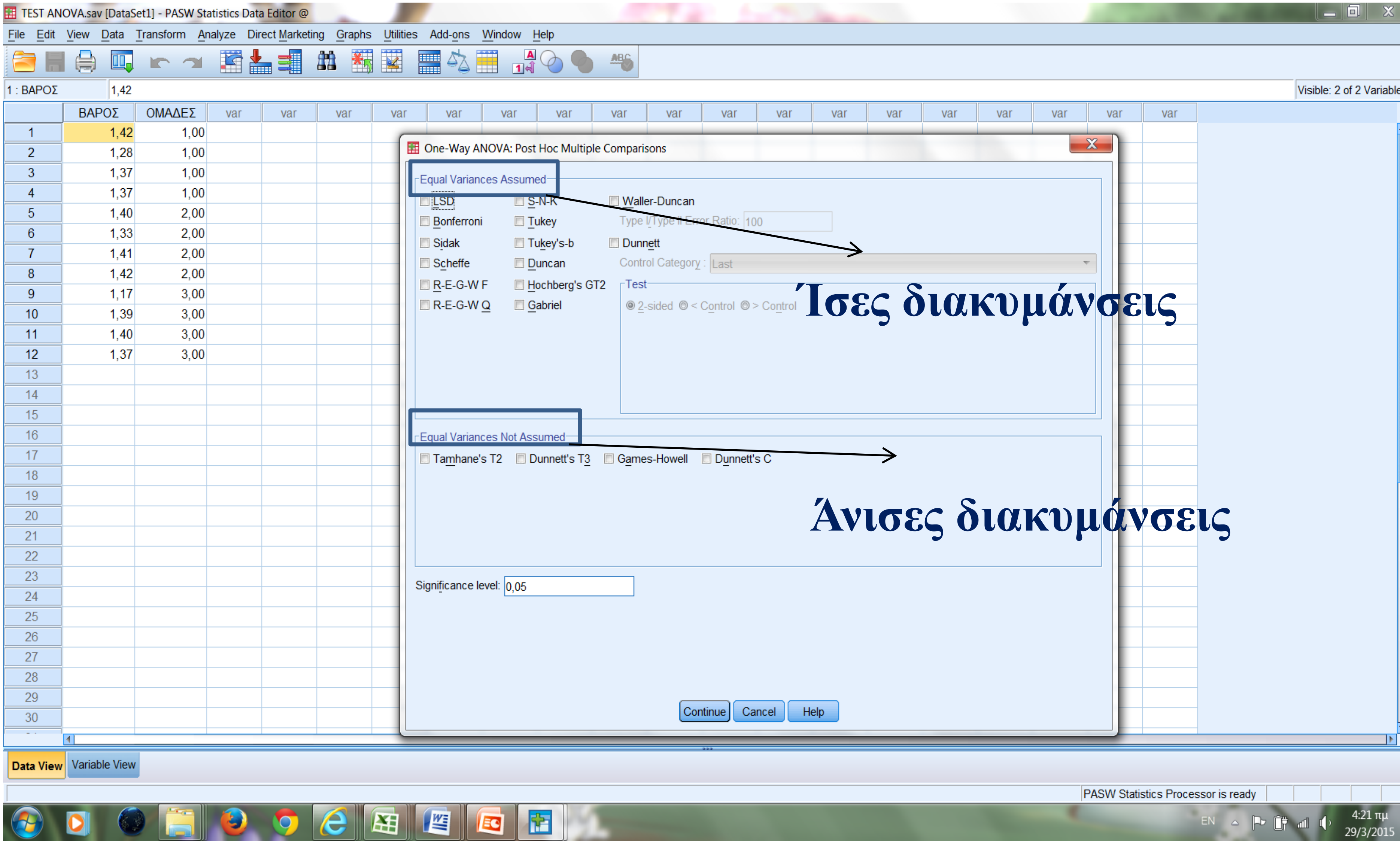
Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2015

Πολλαπλές συγκρίσεις

Στην ανάλυση διακύμανσης ελέγχουμε την ισότητα των μέσων. Αν δεν ισχύει (απορρίπτεται) τότε υπάρχει ένα τουλάχιστον $μ_i$ που διαφέρει από τα υπόλοιπα. Δεν διαθέτουμε όμως καμία πληροφορία σχετικά με ποιο ή ποια από τα $μ_i$ είναι διαφορετικά. Επομένως πρέπει να κάνουμε πολλαπλές συγκρίσεις μεταξύ των ομάδων για να διαπιστώσουμε τις διαφοροποιήσεις καθώς και τα επίπεδα σημαντικότητάς τους.

Πολλαπλές συγκρίσεις



Κριτήριο ελαχίστης σημαντικής διαφοράς (LSD) -Fisher

Εφαρμόζεται για συγκρίσεις ανά δύο

$$H_0 : \mu_i = \mu_j \Rightarrow \mu_i - \mu_j = 0$$

$$H_1 : \mu_i \neq \mu_j \Rightarrow \mu_i - \mu_j \neq 0$$

Η αρχική υπόθεση H_0 απορρίπτεται ααν:

$$\left|\bar{Y}_i - \bar{Y}_j\right| > t_{a/2} s_v^2 \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

$\bar{Y}_i$: μέση τιμή δείγματος i

n_i : μέγεθος δείγματος i

$t_{n-k,a/2}$

$s_v^2 = \sigma^2$: διασπορά μέσα στα δείγματα

Κριτήριο ελαχίστης σημαντικής διαφοράς (LSD) -Fisher

Το 100(1-α)% δ.ε. για την διαφορά $\mu_i - \mu_j$ είναι

$$\left(\overline{Y}_i - \overline{Y}_j \pm t_{n-k,a/2} \sqrt{s_v^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right)$$

Παράδειγμα

Για την παραγωγή χρωμάτων χρησιμοποιήθηκαν 6 διαφορετικά διαλύματα υδροχλωρικού οξέως (HCL)

Για κάθε διάλυμα έγιναν 5 μετρήσεις και τα αποτελέσματα σε gr παρουσιάζονται στον παρακάτω πίνακα

Διάλυμα	1	2	3	4	5	6
Δειγματικοί μέσοι	505	528	564	498	600	470

Παράδειγμα

Επίσης δίνεται και ο πίνακας ανάλυσης διακύμανσης

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο άθροισμα τετραγώνων	F
Απόκλιση από την H_0 (μεταξύ δειγμάτων)	56360	5	11272	4.6
Κατάλοιπα (εντός δειγμάτων)	58824	24	2451	
Σύνολο	115184	29		

Παρατηρούμε ότι $F=4.6>2.62=F_{5,24,0.05}$ άρα η H_0 απορρίπτεται.

Παράδειγμα

Έλεγχος διαφορών με την LSD μέθοδο

$$H_0 : \mu_i = \mu_j \Rightarrow \mu_i - \mu_j = 0$$

$$H_1 : \mu_i \neq \mu_j \Rightarrow \mu_i - \mu_j \neq 0$$

Υπολογίζω την ποσότητα LSD

$$LSD = t_{a/2} \sqrt{s_v^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

$$t_{n-k, a/2} = t_{24,0.025} = 2.064$$

$$LSD = 2.064 \sqrt{2451 \left(\frac{2}{5} \right)} = 64.63$$

Παράδειγμα

Σχηματίζουμε τις διαφορές διατάσσοντας τους δειγματικούς μέσους από τον μικρότερο σστο μεγαλύτερο

Διάλυμα	6	4	1	2	3	5
Δειγματικοί μέσοι	470	498	505	528	564	600

$$\left|\overline{Y}_6 - \overline{Y}_5\right| = \left|470 - 600\right| = 130 > 64.63$$

$$\left|\overline{Y}_6 - \overline{Y}_3\right| = \left|470 - 564\right| = 94 > 64.63$$

$$\left|\overline{Y}_6 - \overline{Y}_2\right| = \left|470 - 528\right| = 58 < 64.63 \Rightarrow \text{σταματάμε}$$

Παράδειγμα

Διάλυμα	6	4	1	2	3	5
Δειγματικοί μέσοι	470	498	505	528	564	600

$$\left|\overline{Y}_4 - \overline{Y}_5\right| = \left|498 - 600\right| = 102 > 64.63$$

$$\left|\overline{Y}_4 - \overline{Y}_3\right| = \left|498 - 564\right| = 66 > 64.63$$

$$\left|\overline{Y}_4 - \overline{Y}_2\right| = \left|498 - 528\right| = 30 < 64.63 \Rightarrow \text{σταματάμε}$$

$$\left|\overline{Y}_1 - \overline{Y}_5\right| = \left|505 - 600\right| = 95 > 64.63$$

$$\left|\overline{Y}_1 - \overline{Y}_3\right| = \left|505 - 564\right| = 59 < 64.63 \Rightarrow \text{σταματάμε}$$

$$\left|\overline{Y}_2 - \overline{Y}_5\right| = \left|528 - 600\right| = 72 > 64.63$$

$$\left|\overline{Y}_2 - \overline{Y}_3\right| = \left|528 - 564\right| = 36 < 64.63 \Rightarrow \text{σταματάμε}$$

$$\left|\overline{Y}_3 - \overline{Y}_5\right| = \left|564 - 600\right| = 36 < 64.63 \Rightarrow \text{σταματάμε}$$

Παράδειγμα

Επομένως

$$\left|\overline{Y}_i - \overline{Y}_j\right| < LSD$$

σημαίνει ότι δεν εντοπίζεται στατιστικά σημαντική διαφορά

Διάλυμα	6	4	1	2	3	5
Δειγματικοί μέσοι	470	498	505	528	564	600

6 - 5

6 - 3

4 - 5

4 - 3

1 - 5

2 - 5

6,4,1,2 σημαντική διαφορά από 5

6,4 σημαντική διαφορά από 3

Παραδείγματα

Ομάδα χοιριδίων	Αύξηση βάρους			
1	1.42	1.28	1.37	1.37
2	1.4	1.33	1.41	1.42
3	1.17	1.39	1.4	1.37

Multiple Comparisons

ΒΑΡΟΣ
LSD

(I) ΟΜΑΔΕΣ	(J) ΟΜΑΔΕΣ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	-,03000	,05316	,586	-,1503	,0903
	3,00	,02750	,05316	,617	-,0928	,1478
2,00	1,00	,03000	,05316	,586	-,0903	,1503
	3,00	,05750	,05316	,308	-,0628	,1778
3,00	1,00	-,02750	,05316	,617	-,1478	,0928
	2,00	-,05750	,05316	,308	-,1778	,0628

Μέθοδος Τ (Tukey)

- Η πιθανότητα να κάνουμε λάθος σε τουλάχιστον ένα ζευγάρι είναι σημαντική με την χρήση της μεθόδου LSD.
- Η μέθοδος Τ εφαρμόσθηκε αρχικά για ισομεγέθη δείγματα και για διαφορές μέσων ανά δύο αλλά επεκτάθηκε σε ανισομεγέθη δείγματα και γραμμικές αντιπαρεμβολές

Μέθοδος T (Tukey)

Αν X_1, X_2,…, X_r είναι ανεξάρτητες τ.μ. με κατανομή $N(\mu_i,\sigma^2)$, i=1,2,…,r και

1. $W=\max X_i - \min X_i$
2. Έχουμε εκτιμητή s^2 του σ^2 με u-βαθμούς ελευθερίας ανεξάρτητο του W, ώστε

$$\frac{us^2}{\sigma^2} \sim x_u^2$$

τότε ο λόγος $q(r,u)=\dfrac{W}{s}$ λέγεται τυποποιημένο εύρος

Μέθοδος Τ (Tukey)

Ανάλυση διασποράς για ισομεγέθη δείγματα

Έχουμε τ.μ. $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_k$ Ανεξάρτητες με κατανομή $N(\mu_i, \sigma^2/n)$ και s^2 είναι ένας εκτιμητής του σ^2 ανεξάρτητος από τα $\overline{Y}_i$ με

$$\frac{(n-k)s^2}{\sigma^2} \sim x_{n-k}^2$$

και $n = n_1 = n_2 = \ldots = n_k$

Από τον ορισμό του q(r,u) προκύπτει ότι:

Μέθοδος T (Tukey)

Ανάλυση διασποράς για ισομεγέθη δείγματα

Πρόταση

Η πιθανότητα είναι 1-α να ισχύουν συγχρόνως οι k(k-1)/2 σχέσεις:

$$\bar{Y}_i - \bar{Y}_j - Ts \leq \mu_i - \mu_j \leq \bar{Y}_i - \bar{Y}_j + Ts$$

όπου $T = \dfrac{q(k, n-k; 1-a)}{\sqrt{n}}$

Μέθοδος Τ (Tukey)

- Με την μέθοδο Τ για ισομεγέθη δείγματα κατατάσσουμε τους εκτιμώμενους μέσους κατά αύξουσα σειρά και εξετάζουμε αν

$$\left|\overline{Y}_i - \overline{Y}_j\right| < W_k = \frac{q\left(k, n-k, a\right)}{\sqrt{n}}$$

- Όσες διαφορές ικανοποιούν την σχέση αυτή τότε οι αντίστοιχοι μέσοι είναι ίσοι (ανήκουν στην ίδια ομάδα) αλλιώς είναι άνισοι.

Παράδειγμα

Δίσκοι από μείγμα χαλκού με άλλο μέταλλο τοποθετήθηκαν σε θειικό οξύ για να μελετηθεί το χάσιμο βάρους κατά την διάβρωση. Το μείγμα είναι χαλκός με άργυρο ή χαλκός με πυρίτιο. Έγιναν 10 μετρήσεις για κάθε μείγμα με αποτελέσματα

	1	2	3	4	5
	Καθαρό	Άργυρος 35%	Πυρίτιο 25%	Άργυρος 87%	Πυρίτιο 50%
$Y_{i.}$	31.8	30.13	30.10	32.58	31.83
S_i^2	1.181	2.085	5.008	1.170	1.64
n_i	10	10	10	10	10

Παράδειγμα

Η κατάταξη των μέσων δίνει

$$\bar{Y}_3 = 30.1, \bar{Y}_2 = 30.13, \bar{Y}_1 = 31.8, \bar{Y}_5 = 31.83, \bar{Y}_4 = 32.58$$

	1	2	3	4	5
	Καθαρό	Άργυρος 35%	Πυρίτιο 25%	Άργυρος 87%	Πυρίτιο 50%
$Y_{i.}$	31.8	30.13	30.10	32.58	31.83
s_i^2	1.181	2.085	5.008	1.170	1.64
n_i	10	10	10	10	10

$$s^2 = \frac{SSE}{45} = \frac{SSE_1 + SSE_2 + \ldots + SSE_5}{45}$$

$$s^2 = \frac{s_1^2 + s_2^2 + \ldots + s_5^2}{5} = \frac{11.084}{5} = 2.216$$

Παράδειγμα

Από τους πίνακες της q παίρνουμε $q(5,45,0.95) = 4.02$

και $W_5 = 4.02\sqrt{\dfrac{2.2168}{10}} = 1.89$

32.58-30.10=2.48>1.89 , όλοι οι μέσοι δεν είναι ίσοι

32.58-30.13=2.45>1.89 , οι τέσσερεις μεγαλύτεροι όχι ίσοι

32.58-31.8= 0.78<1.89, οι τρείς μεγαλύτεροι ανήκουν στην ίδια ομάδα (θεωρούνται ίσοι)

31.83-31.80=0.03<1.89, οι τρείς μικρότεροι στην ίδια ομάδα

Μέθοδος T (Tukey)

- **Ανάλυση διασποράς για ανισομεγέθη δείγματα**

Όταν όλα τα δείγματα δεν έχουν το ίδιο μέγεθος και συγκρίνουμε όλους τους μέσους μεταξύ τους τότε εφαρμόζεται πάλι η μέθοδος T και ονομάζεται μέθοδος Tukey-Kramer.

Διαστήματα εμπιστοσύνης:

$$\widehat{D} \pm Ts\left(\widehat{D}\right)$$

όπου:

$$\widehat{D} = \bar{Y}_i - \bar{Y}_j$$

$$s^2\left(\widehat{D}\right) = s^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$$

$$T = q(k, n-k; 1-\alpha)\frac{1}{\sqrt{2}}$$

Παράδειγμα

Μας ενδιαφέρει να εξετάσουμε 5 δίαιτες που δίνονται σε παιδιά ηλικίας ενός έτους. Το βάρος σε γραμμάρια δινεται από τον παρακάτω πίνακα

Δίαιτα	Βάρος σε γραμμάρια	n_i	$\bar{Y}_{i.}$	$Y_{i.}$
Α	28 24.8 27.9 24.7 28.7 34.8 30.9	7	28.54	199.8
Β	24.7 30.7 30.5 22 26.7 28.6 28.8 22.6 34.8	9	27.71	249.4
Γ	30.5 24.8 30.9 22.5 37.6 28.6 28.7 28.8 33.6 24	10	29	290
Δ	30.4 28.9 27.8 30.7 32.7 30.8 30.5 34.5 32.8 31.7 28.6 38.8	12	31.51	378.2
Ε	28.5 20.8 22.9 17.8 23.6 18.9 24.7 16.3 25.4 20.9 25.1	11	22.26	244.9

$$Y_{i.} = \sum_{j}^{n_i} Y_{ij} \qquad \bar{Y}_{i.} = \frac{1}{n_i}\sum_{j}^{n_i} Y_{ij}$$

Παράδειγμα

Πίνακας Ανάλυσης Διακύμανσης

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο άθροισμα τετραγώνων	F
Απόκλιση από την H_0 (μεταξύ δειγμάτων)	521.262	4	130.315	9.0801 ($p<0.0001$)
Κατάλοιπα (εντός δειγμάτων)	631.508	44	14.352	
Σύνολο	11552.763	48		

Επομένως απορρίπτουμε ότι όλοι οι μέσοι είναι ίσοι.

Η εκτίμηση για την διασπορά των σφαλμάτων είναι:

$$\sigma^2 = 14.352$$

Παράδειγμα

$$s^2 = 14.352, \qquad T = q(5,44;0.95)\frac{1}{\sqrt{2}}$$

$$(\bar{Y}_A - \bar{Y}_B) \pm 2.546\sqrt{14.32\left(\frac{1}{7}+\frac{1}{9}\right)} = 0.83 \pm 4.86$$

$$(\bar{Y}_A - \bar{Y}_\Gamma) \pm 2.546\sqrt{14.32\left(\frac{1}{7}+\frac{1}{10}\right)} = -0.46 \pm 4.75$$

$$(\bar{Y}_A - \bar{Y}_\Delta) \pm 2.546\sqrt{14.32\left(\frac{1}{7}+\frac{1}{12}\right)} = -2.98 \pm 4.58$$

$$(\bar{Y}_A - \bar{Y}_E) \pm 2.546\sqrt{14.32\left(\frac{1}{7}+\frac{1}{11}\right)} = 6.28 \pm 4.66$$

$$(\bar{Y}_B - \bar{Y}_\Gamma) \pm 2.546\sqrt{14.32\left(\frac{1}{9}+\frac{1}{10}\right)} = -1.29 \pm 4.43$$

$$(\bar{Y}_B - \bar{Y}_\Delta) \pm 2.546\sqrt{14.32\left(\frac{1}{9}+\frac{1}{12}\right)} = -3.81 \pm 4.25$$

$$(\bar{Y}_B - \bar{Y}_E) \pm 2.546\sqrt{14.32\left(\frac{1}{9}+\frac{1}{11}\right)} = 5.44 \pm 4.35$$

$$(\bar{Y}_\Gamma - \bar{Y}_\Delta) \pm 2.546\sqrt{14.32\left(\frac{1}{10}+\frac{1}{12}\right)} = -2.52 \pm 4.13$$

$$(\bar{Y}_\Gamma - \bar{Y}_E) \pm 2.546\sqrt{14.32\left(\frac{1}{10}+\frac{1}{11}\right)} = 6.74 \pm 4.21$$

$$(\bar{Y}_\Delta - \bar{Y}_E) \pm 2.546\sqrt{14.32\left(\frac{1}{12}+\frac{1}{11}\right)} = 9.25 \pm 4.11$$

Όταν το διάστημα περιέχει το 0, οι αντίστοιχοι μέσοι είναι ίσοι
Οι μέσοι χωρίζονται σε 2 ομάδες, η δίαιτα Ε έχει το μικρότερο βάρος και οι υπόλοιπες θεωρούνται ισοδύναμες

Μέσοι άνισοι
Ε δίαιτα κοινή

Παραδείγματα

Ομάδα χοιριδίων	Αύξηση βάρους			
1	1.42	1.28	1.37	1.37
2	1.4	1.33	1.41	1.42
3	1.17	1.39	1.4	1.37

Multiple Comparisons

ΒΑΡΟΣ
Tukey HSD

(I) ΟΜΑΔΕΣ	(J) ΟΜΑΔΕΣ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	-,03000	,05316	,842	-,1784	,1184
	3,00	,02750	,05316	,865	-,1209	,1759
2,00	1,00	,03000	,05316	,842	-,1184	,1784
	3,00	,05750	,05316	,548	-,0909	,2059
3,00	1,00	-,02750	,05316	,865	-,1759	,1209
	2,00	-,05750	,05316	,548	-,2059	,0909

Μέθοδος Duncan

- Για να εφαρμόσουμε το τεστ του Duncan σε δείγματα σε ίδιο μέγεθος m, διατάσουμε τους δειγματικούς μέσους σε αύξουσα σειρά και υπολογίζουμε το τυπικό σφάλμα κάθε μέσου

$$s^2_{\overline{Y}_{i.}} = \sqrt{\frac{MSE}{m}} = \sqrt{\frac{s_v^2}{m}}$$

- Για δείγματα με διαφορετικό μέγεθος αντικαθιστούμε το μέγεθος n με τον αρμονικό μέσο m_h των $\{n_i\}$ με

$$m_h = \frac{k}{\sum_{i=1}^{k} \frac{1}{n_i}}$$

Μέθοδος Duncan

- Αν

$$n_1 = n_2 = \ldots = n_k = m$$

τότε $m_h = m$

- Από τους πίνακες του Duncan βρίσκουμε τις τιμές

$$r_a = (p, v), p = 1,2,\ldots,k$$

όπου α επίπεδο σημαντικότητας, και ν οι βαθμοί ελευθερίας για το σφάλμα ν=n-k

- Υπολογίζουμε τις ποσότητες

$$R_p = r_a(p, v) s_{\bar{Y}_{i.}}, p = 2,3,\ldots,k$$

Μέθοδος Duncan

- Ελέγχουμε τις παρατηρούμενες διαφορές μεταξύ των μέσων ξεκινώντας από την μεγαλύτερη ενάντια στην μικρότερη, η οποία ελέγχεται με το ελάχιστο σημαντικό εύρος R_k
- Στην συνέχεια υπολογίζουμε την διαφορά της μεγαλύτερης από την δεύτερη μικρότερη, η οποία συγκρίνεται με το ελάχιστο σημαντικό εύρος R_{k-1}
- Οι συγκρίσεις συνεχίζονται μέχρι να υπολογιστούν όλες οι διαφορές των μέσων με τον μεγαλύτερο μέσο.

Μέθοδος Duncan

- Στην συνέχεια υπολογίζεται η διαφορά του δευτέρου μεγαλύτερου μέσου και του μικρότερου και συγκρίνεται με την τιμή R_{k-1}. Η διαδικασία συνεχίζεται μέχρις ότου εξαντληθούν οι διαφορές των δυνατών ζευγών των μέσων

$$\frac{k(k-1)}{2}$$

- Αν μια παρατηρούμενη διαφορά i είναι μεγαλύτερη από το αντίστοιχο ελάχιστο σημαντικό εύρος R_i τότε το ζεύγος των μέσων διαφέρει σημαντικά

Παράδειγμα

Για την παραγωγή συνθετικών ενδυμάτων χρησιμοποιούνται διάφορα ποσοστά βαμβακιού. Πέντε διαφορετικά ποσοστά χρησιμοποιήθηκαν και για κάθε ποσοστό έγιναν 5 μετρήσεις.

$\overline{Y}_{i.}$	**Βάρος σε γραμμάρια**
Α - 15%	7 7 15 11 9
Β – 20%	12 17 12 18 18
Γ – 25%	14 18 18 19 19
Δ - 35%	19 25 22 19 23
Ε – 35%	7 10 11 15 11

Παράδειγμα

Από τα δεδομένα παρατηρούμε ότι:

$$Y_{..} = 376$$

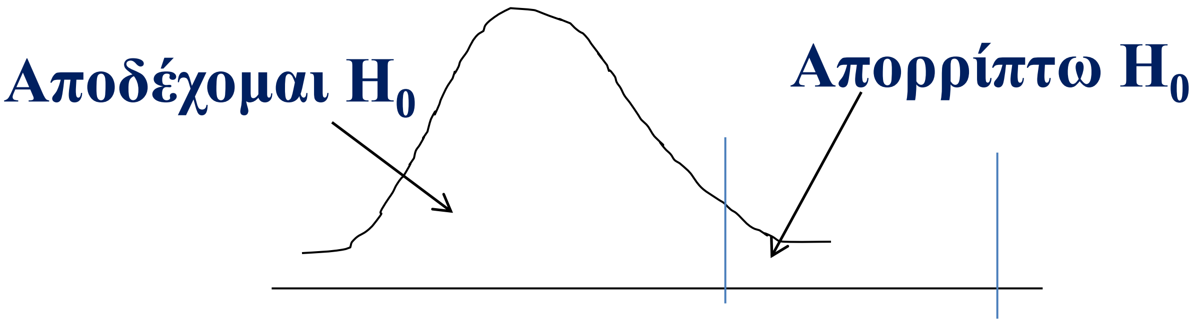
$$\bar{Y}_{..} = 15.04$$

$Y_{i.}$	$\bar{Y}_{i.}$
49	9.8
77	15.4
88	17.6
108	21.6
54	10.8

Πηγή μεταβλητότητας	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο άθροισμα τετραγώνων	F
Απόκλιση από την H_0 (μεταξύ δειγμάτων)	457.76	4	118.94	14.76
Κατάλοιπα (εντός δειγμάτων)	161.20	20	8.06	
Σύνολο	636.96	24		

Παράδειγμα

Έτσι η υπόθεση H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ έναντι της εναλλακτικής H_1: τουλάχιστον ένα από τα μ_i διαφέρει απορρίπτεται αφού $F=14.76>F_{4,20,0.05}=2.87$



$$F_{4,20,0.05} = 2.87 \quad F = 14.76$$

Θα εφαρμόσουμε την μέθοδο Duncan για να κάνουμε τον έλεγχο

$$H_0: \mu_i - \mu_j = 0$$

$$H_0: \mu_i - \mu_j \neq 0$$

Παράδειγμα

Επομένως έχουμε:

$$MSE = s_v^2 = 8.06$$

$$n = 25$$

$$m = 5$$

$$k = 5$$

$$v = n - k = 20$$

Διατάσουμε τους δειγματικούς μέσους σε αύξουσα σειρά:

$\overline{Y}_{1.}$	$\overline{Y}_{5.}$	$\overline{Y}_{2.}$	$\overline{Y}_{3.}$	$\overline{Y}_{4.}$
9.8	10.8	15.4	17.6	21.6

Τυπική απόκλιση μέσου:

$$s_{\overline{Y}_{i.}} = \sqrt{\frac{8.06}{5}} = 1.27$$

Παράδειγμα

Από τον πίνακα του Duncan για 20 β.ε. και α=0.05 έχουμε:

$$r_{0.05}(2,20) = 2.95$$

$$r_{0.05}(3,20) = 3.10$$

$$r_{0.05}(4,20) = 3.18$$

$$r_{0.05}(5,20) = 3.25$$

Υπολογίζουμε τις τιμές:

$$R_2 = r_{0.05}(2,20), s_{\overline{Y}_{i.}} = 2.95 * 1.27 = 3.75$$

$$R_3 = r_{0.05}(3,20), s_{\overline{Y}_{i.}} = 3.10 * 1.27 = 3.94$$

$$R_4 = r_{0.05}(4,20), s_{\overline{Y}_{i.}} = 3.18 * 1.27 = 4.04$$

$$R_5 = r_{0.05}(5,20), s_{\overline{Y}_{i.}} = 3.25 * 1.27 = 4.13$$

Παράδειγμα

Οι συγκρίσεις δίνουν:

$$4 \text{ vs } 1: 21.6 - 9.8 = 11.8 > 4.13$$

$$4 \text{ vs } 5: 21.6 - 10.8 = 10.8 > 4.04$$

$$4 \text{ vs } 2: 21.6 - 15.4 = 6.2 > 3.94$$

$$4 \text{ vs } 3: 21.6 - 17.6 = 4 > 3.75$$

$$3 \text{ vs } 1: 17.6 - 9.8 = 7.8 > 4.04$$

$$3 \text{ vs } 5: 17.6 - 10.8 = 6.8 > 3.95$$

$$3 \text{ vs } 2: 17.6 - 15.4 = 2.2 < 3.75 *$$

$$2 \text{ vs } 1: 15.4 - 9.8 = 5.6 > 3.94$$

$$2 \text{ vs } 5: 15.4 - 10.8 = 4.6 > 3.75$$

$$5 \text{ vs } 1: 10.8 - 9.8 = 1 < 3.75 *$$

Μη στατιστικά σημαντική διαφορά

Μέθοδος Scheffe

- Η μέθοδος αυτή εφαρμόζεται και στις περιπτώσεις όπου το πλήθος των παρατηρήσεων δεν είναι ίδιο σε όλες τις ομάδες. Όταν έχουμε ίδιο μέγεθος παρατηρήσεων σε όλες τις ομάδες τότε είναι προτιμότερη η μέθοδος Τα, διότι δίνει μικρότερα διαστήματα εμπιστοσύνης

Μέθοδος Scheffe

Πρόταση

- Η πιθανότητα είναι 1-α να ισχύουν συγχρόνως οι σχέσεις:

$$\widehat{L} - Ss\left(\widehat{L}\right) \leq L \leq \widehat{L} + Ss\left(\widehat{L}\right)$$

όπου:

$$S = \sqrt{(k-1)F_{k-1,n-k,a}}$$

$$s\left(\widehat{L}\right) = s\sqrt{\sum_{i=1}^{k}\frac{c_i^2}{n_i}}$$

$$L = c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k$$

$$\widehat{L} = c_1\overline{Y}_1 + c_2\overline{Y}_2 + \ldots + c_k\overline{Y}_k$$

$$c_1 + c_2 + \ldots + c_k = 0$$

Μέθοδος Scheffe

Η διαδικασία που χρησιμοποιούμε για να κατατάξουμε σε ομάδες τους μέσους με την S-μέθοδο είναι:

1. Τοποθετούμε τα $\overline{Y}_{i.}$ κατά αύξουσα (ή φθίνουσα) σειρά μεγέθους $\overline{Y}_{\{1.\}}, \overline{Y}_{\{2.\}}, \ldots, \overline{Y}_{\{k.\}}$
2. Υπολογίζουμε την ποσότητα

$$D(i, j) = s\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)(k-1)F_{k-1,n-k,a}}$$

3. Αν $\left|\overline{Y}_{i.} - \overline{Y}_{j.}\right| < D(i, j)$ δεχόμαστε ότι οι μέσοι είναι ίσοι (ανήκουν στην ίδια ομάδα) διαφορετικά είναι άνισοι.

Παράδειγμα

Έχουμε : s^2=14.352, k=5, n=49

Διατάσουμε τους δειγματικούς μέσους της δίαιτας σε αύξουσα σειρά:

$\bar{Y}_{E.}$	$\bar{Y}_{B.}$	$\bar{Y}_{A.}$	$\bar{Y}_{\Gamma.}$	$\bar{Y}_{\Delta.}$
22.264 11	27.711 9	28.543 7	29 10	31.517 12

Παράδειγμα

$$D(E,\Delta)=\sqrt{14.352\left(\frac{1}{11}+\frac{1}{12}\right)}4(2.6)=5.099 \Rightarrow 31.517-22.264=9.253>5.099$$

$$D(\mathrm{B},\Delta)=\sqrt{14.352\left(\frac{1}{9}+\frac{1}{12}\right)}4(2.6)=5.387 \Rightarrow 31.517-27.711=3.806<5.099$$

$$D(E,\Gamma)=\sqrt{14.352\left(\frac{1}{11}+\frac{1}{10}\right)}4(2.6)=5.338 \Rightarrow 29-22.264=7.736>5.338$$

$$D(E,\mathrm{A})=\sqrt{14.352\left(\frac{1}{11}+\frac{1}{7}\right)}4(2.6)=5.907 \Rightarrow 28.543-22.264=6.279>5.907$$

$$D(E,\mathrm{B})=\sqrt{14.352\left(\frac{1}{11}+\frac{1}{9}\right)}4(2.6)=5.491 \Rightarrow 27.711-22.264=5.447<5.491$$

E vs Δ

E vs Γ

E vs A

Η δίαιτα Ε διαφέρει σημαντικά

Παράδειγμα

$$D(E,\Delta)=\sqrt{14.352\left(\frac{1}{11}+\frac{1}{12}\right)4(2.6)}=5.099 \Rightarrow 31.517-22.264=9.253>5.099$$

$$\text{διάστημα εμπιστοσύνης } (9.253 \pm 5.099)=(4.154, 14.352) \Rightarrow \text{δεν περιέχεται το } 0$$

$$D(\mathrm{B},\Delta)=\sqrt{14.352\left(\frac{1}{9}+\frac{1}{12}\right)4(2.6)}=5.387 \Rightarrow 31.517-27.711=3.806<5.099$$

$$\text{διάστημα εμπιστοσύνης } (3.806 \pm 5.387)=(-1.581, 9.193) \Rightarrow \text{περιέχεται το } 0$$

Παραδείγματα

Ομάδα χοιριδίων	Αύξηση βάρους			
1	1.42	1.28	1.37	1.37
2	1.4	1.33	1.41	1.42
3	1.17	1.39	1.4	1.37

Multiple Comparisons

ΒΑΡΟΣ
Scheffe

(I) ΟΜΑΔΕΣ	(J) ΟΜΑΔΕΣ	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00	2,00	-,03000	,05316	,855	-,1851	,1251
	3,00	,02750	,05316	,876	-,1276	,1826
2,00	1,00	,03000	,05316	,855	-,1251	,1851
	3,00	,05750	,05316	,577	-,0976	,2126
3,00	1,00	-,02750	,05316	,876	-,1826	,1276
	2,00	-,05750	,05316	,577	-,2126	,0976