Επαγωγική Στατιστική
Ανάλυση διακύμανσης

Χαράλαμπος Γναρδέλλης

Πρόγραμμα Μεταπτυχιακών Σπουδών
Βιώσιμη Αλιεία, Υδατοκαλλιέργεια

Ανάλυση διακύμανσης προς έναν παράγοντα

- **Ο έλεγχος των μέσων τιμών δύο ανεξάρτητων πληθυσμών είναι η απλούστερη περίπτωση ενός σύνθετου ζητήματος που συναντάται συχνά στη στατιστική συμπερασματολογία και το οποίο, στη γενική του μορφή, αφορά τη σύγκριση μέσων τιμών περισσοτέρων των δύο.**

- **Από μεθοδολογικής απόψεως, προβλήματα του είδους επιλύονται με τη βοήθεια των τεχνικών της ευρύτερης περιοχής της στατιστικής ανάλυσης που φέρει το όνομα *ανάλυση διακύμανσης* (analysis of variance). Ο λόγος για τον οποίο οι τεχνικές αυτές αναφέρονται με αυτήν την ονομασία, οφείλεται στο γεγονός ότι η λογική τους στοχεύει στο διαμερισμό της συνολικής διασποράς ενός συνόλου δεδομένων σε επιμέρους συνιστώσες.**

Η *ανάλυση διακύμανσης προς έναν παράγοντα (one-way analysis of variance)* είναι η πλέον απλή περίπτωση ελέγχου περισσοτέρων των δύο πληθυσμιακών μέσων τιμών. Το γενικό πλαίσιο στο οποίο τοποθετείται η χρήση της, αφορά *k* ανεξάρτητους πληθυσμούς στους οποίους η κατανομή μιας ποσοτικής μεταβλητής είναι κανονική με μέσες τιμές και τυπικές αποκλίσεις μ_j και σ_j αντίστοιχα, όπου $j = 1,2, . . ., k$. Από κάθε έναν από τους πληθυσμούς αυτούς λαμβάνεται ένα τυχαίο δείγμα μεγέθους n_j με μέση τιμή $\bar{x}_j$ και τυπική απόκλιση s_j

Συμβολική απεικόνιση των πληθυσμιακών και δειγματικών τιμών στην ανάλυση διακύμανσης προς έναν παράγοντα

	1	**2**	**. . .**	***k***
Πληθυσμός				
Μέση τιμή	μ_1	μ_2	. . .	μ_k
Τυπική απόκλιση	σ_1	σ_2	. . .	σ_k
Δείγμα				
Μέση τιμή	$\bar{x}_1$	$\bar{x}_2$	. . .	$\bar{x}_k$
Τυπική απόκλιση	s_1	s_2	. . .	s_k
Μέγεθος δείγματος	n_1	n_2	. . .	n_k

Η μηδενική υπόθεση του ελέγχου των *k* πληθυσμιακών μέσων τιμών μπορεί να διατυπωθεί

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

ενώ η εναλλακτική που προκύπτει λογικά από την απόρριψη της μηδενικής υπόθεσης είναι

H_A **: Δύο τουλάχιστον από τις** $\mu_1, \mu_2, \ldots, \mu_k$ **διαφέρουν μεταξύ τους.**

Μια πρώτη αντιμετώπιση ενός τέτοιου ελέγχου θα μπορούσε να γίνει με μια απλή προέκταση του ελέγχου δύο μέσων τιμών. Να ελεγχθούν δηλαδή ανά δύο όλοι οι δυνατοί συνδυασμοί των *k* μέσων τιμών, χρησιμοποιώντας πολλαπλά *t* tests για ανεξάρτητα δείγματα. Το πρόβλημα που δημιουργείται σε μια τέτοια περίπτωση είναι ότι όταν πολλές μέσες τιμές ελέγχονται ταυτόχρονα ανά δύο, η πιθανότητα να προκύψει εσφαλμένη απόρριψη μιας μηδενικής υπόθεσης αυξάνει ανάλογα με τον αριθμό των δυνατών συγκρίσεων.

Αν π.χ. για τον έλεγχο της ισότητας των μέσων τιμών μ_i και μ_j χρησιμοποιηθεί ως επίπεδο σημαντικότητας η τιμή α=0,05 και για τον έλεγχο των μέσων μ_k και μ_v χρησιμοποιηθεί επίσης η τιμή α=0,05, τότε –σύμφωνα με τον πολλαπλασιαστικό κανόνα των πιθανοτήτων– η συνολική πιθανότητα να απορριφθεί ορθά η μηδενική υπόθεση και για τους δύο ελέγχους είναι

$$(1-0,05)\cdot(1-0,05) = (0,95)^2 = 0,9025 .$$

Επομένως, η πιθανότητα να απορριφθεί εσφαλμένα η μηδενική υπόθεση σε έναν από τους δύο ελέγχους είναι

$$p = 1-0,9025 = 0,0975.$$

Σε έναν έλεγχο που αφορά την ισότητα τριών πληθυσμιακών μέσων, δηλαδή όταν *k*=3, η συνολική πιθανότητα να απορριφθεί εσφαλμένα η μηδενική υπόθεση σε έναν από τους ελέγχους είναι $1-(0,95)^3=0,1426$, ενώ για 5 πληθυσμιακούς μέσους, η συνολική πιθανότητα είναι $1-(0,95)^{10}=0,4012$. Όπως είναι εμφανές, η συνδυασμένη πιθανότητα εσφαλμένης απόρριψης μιας μηδενικής υπόθεσης σε πολλαπλούς ελέγχους, αυξάνει ανάλογα με τον αριθμό των δυνατών συγκρίσεων και μπορεί να γίνει πολύ μεγαλύτερη από το προκαθορισμένο επίπεδο σημαντικότητας ενός ελέγχου, το οποίο συνήθως είναι *α*=0,05. Γενικά, η συνολική πιθανότητα εσφαλμένης απόρριψης μιας μηδενικής υπόθεσης είναι ίση με $1-(0,95)^m$, όπου *m* ο αριθμός των δυνατών συγκρίσεων.

Αριθμός λευκών αιμοσφαιρίων (WBC/mm^3) σε τρεις ομάδων ατόμων

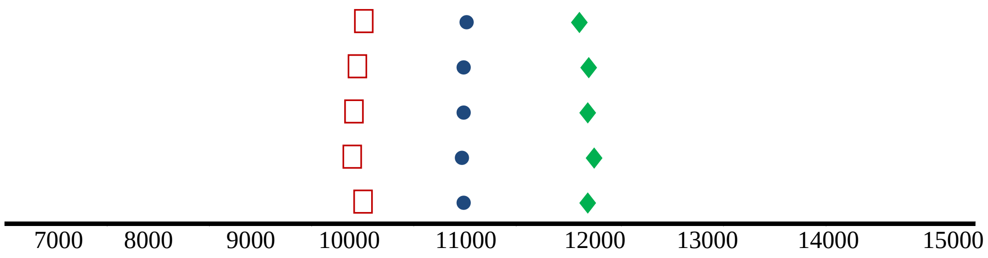
Περίπτωση 1η

Ομάδες		
Α ♦	**Β** ●	**Γ** □
11.800	11.020	10.020
11.840	11.000	10.000
11.820	11.000	9.980
11.780	10.980	9.960
11.760	11.000	10.040
$\bar{x}_1 = 11.800$	$\bar{x}_2 = 11.000$	$\bar{x}_3 = 10.000$

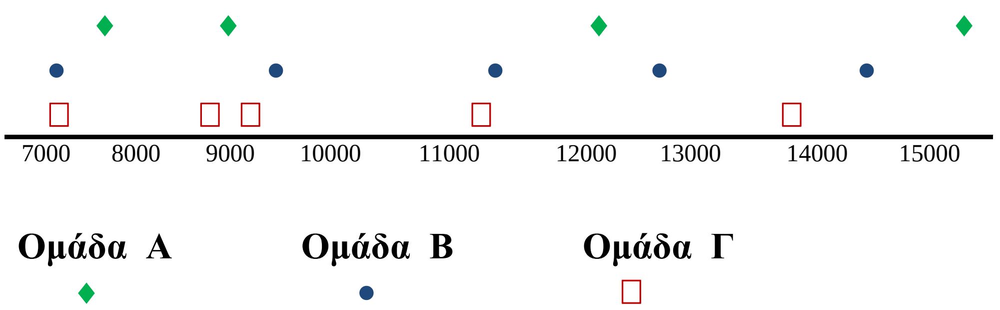
Περίπτωση 2η

Ομάδες		
Α ♦	**Β** ●	**Γ** □
11.800	12.620	9.040
8.840	7.080	13.860
15.020	9.460	8.960
15.780	14.400	11.100
7.560	11.440	7.040
$\bar{x}_1 = 11.800$	$\bar{x}_2 = 11.000$	$\bar{x}_3 = 10.000$

Παράδειγμα 1ο



Παράδειγμα 2ο



Ανάλυση της συνολικής διασποράς σε επιμέρους συνιστώσες

Χρησιμοποιώντας τους ορισμούς που ετέθησαν προηγουμένως και συμβολίζοντας τις αναλυτικές τιμές των παρατηρήσεων με x_{ij} , $j=1,2,...,k$ (k ο αριθμός των ομάδων) και $i=1,2,...,n_j$ (n_j ο αριθμός των παρατηρήσεων κάθε ομάδας), μέτρο της συνολικής διασποράς των παρατηρήσεων των k ομάδων είναι η ποσότητα

$$TSS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}\right)^2$$

όπου $\bar{x}$ η συνολική μέση τιμή όλων των παρατηρήσεων.

- **Η ποσότητα *SSW*, αποτελεί μέτρο της διασποράς στο εσωτερικό των *k* ομάδων.**

$$SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \overline{x}_j \right)^2$$

- **Ενώ η ποσότητα *SSB*, αποτελεί μέτρο της διασποράς των μέσων τιμών $\overline{x}_j$, *j=1,2,…k* των *k* ομάδων γύρω από τη συνολική μέση τιμή $\overline{x}$.**

$$SSB = \sum_{j=1}^{k} n_j \left(\overline{x}_j - \overline{x} \right)^2$$

Η ποσότητα *TSS* ονομάζεται *συνολικό άθροισμα τετραγώνων* (*Total Sum of Squares*)

η ποσότητα *SSW* ονομάζεται *άθροισμα τετραγώνων στο εσωτερικό των ομάδων* (*Sum of Squares Within-groups*)

και η ποσότητα *SSB άθροισμα τετραγώνων μεταξύ των ομάδων* (*Sum of Squares Between-groups)*

Αποδεικνύεται εύκολα με απλές αλγεβρικές πράξεις ότι μεταξύ των ποσοτήτων *TSS, SSB* και *SSW* ισχύει η εξής σχέση

$$TSS = SSB + SSW .$$

Οι ποσότητες s_T^2 , s_B^2 και s_W^2 οι οποίες προκύπτουν διαιρώντας τις ποσότητες *TSS*, *SSB* και *SSW* με τους βαθμούς ελευθερίας τους, *N*-1, *k-1* και *N-k*, ονομάζονται αντίστοιχα *συνολικό μέσο τετράγωνο*, (total mean square), *μέσο τετράγωνο μεταξύ των ομάδων* (mean square between-groups) και *μέσο τετράγωνο στο εσωτερικό των ομάδων* (mean square within- groups).

$$s_T^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}\right)^2}{N-1} = \frac{TSS}{N-1}$$

Συνολικό μέσο τετράγωνο (total mean square)

$$s_B^2 = \frac{\sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{x}\right)^2}{k-1} = \frac{SSB}{k-1}$$

Μέσο τετράγωνο μεταξύ των ομάδων (mean square between-groups)

$$s_W^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2}{N-k} = \frac{SSW}{N-k}$$

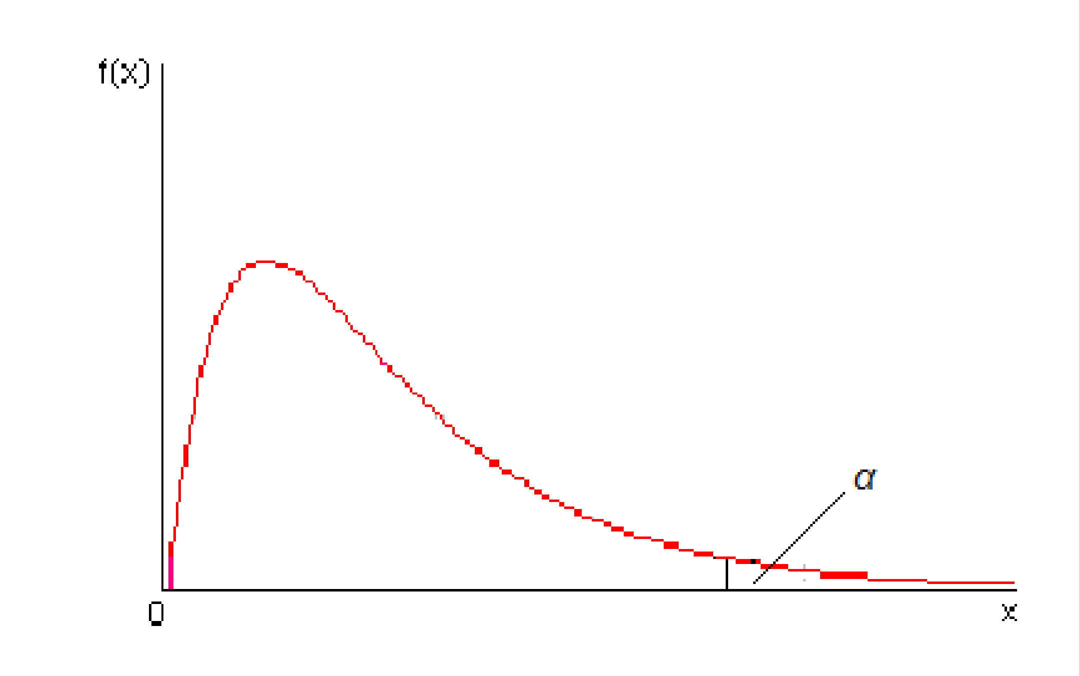
Μέσο τετράγωνο στο εσωτερικό των ομάδων (mean square within- groups).

Υπό την προϋπόθεση, ότι ισχύει η μηδενική υπόθεση,

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

ο λόγος $F = \dfrac{s_B^2}{s_W^2}$

ακολουθεί την κατανομή *F* με *k*-1 και *N*-*k* βαθμούς ελευθερίας.



Αν επομένως, η τιμή του λόγου είναι αρκετά μεγάλη ώστε να ορίζει στη δεξιά ουρά της αντίστοιχης κατανομής *F* μια περιοχή εμβαδού μικρότερου από α=0,05 (και άρα η πιθανότητα *p* να προκύψει μια τόσο μεγάλη τιμή να είναι μικρότερη από 0,05), τότε η υπόθεση της ισότητας των *k* πληθυσμιακών μέσων απορρίπτεται .

Στο SPSS, τα αποτελέσματα της διαδικασίας One-Way ANOVA, οργανώνονται υπό τη μορφή ενός πίνακα, στο εσωτερικό του οποίου εμφανίζονται όλες οι προαναφερθείσες ποσότητες s_B^2 , s_W^2 , *TSS*, *SSB*, *SSW*, *N-k*, *k-1*, μαζί με την τιμή του κριτηρίου *F* και την πιθανότητα *p* (Sig.) του ελέγχου. Ο πίνακας αυτός έχει συγκεκριμένη διάταξη όλων αυτών των ποσοτήτων στο εσωτερικό του και ονομάζεται *πίνακας ανάλυσης διακύμανσης (ANOVA πίνακας – Analysis of Variance table)*.

Πίνακας Ανάλυσης Διακύμανσης (ANOVA Table)

Πηγή διασποράς	Άθροισμα τετραγώνων	Βαθμοί ελευθερίας	Μέσο τετράγωνο	*F*	Sig.
Μεταξύ των ομάδων	*SSB*	*k*-*1*	$s_B^2 = \frac{SSB}{(k-1)}$	$\frac{s_B^2}{s_W^2}$	
Στο εσωτερικό των ομάδων	*SSW*	*N*-*k*	$s_w^2 = \frac{SSW}{(N-k)}$		
Σύνολο	*TSS*	*N*-*1*			

Προϋποθέσεις εφαρμογής της ανάλυσης διακύμανσης

Οι κατανομές της ποσοτικής μεταβλητής στους *k* πληθυσμούς από τους οποίους προέρχονται οι ομάδες θα πρέπει να είναι κανονικές. Ο έλεγχος της κανονικότητας γίνεται διαγραμματικά με τη χρήση θηκογραμμάτων και ιστογραμμάτων. Όταν ο συνολικός αριθμός των παρατηρήσεων είναι μικρός, $N \leq 30$, είναι απαραίτητη η χρήση μη παραμετρικών μεθόδων, όπως αυτή της ανάλυσης διακύμανσης των Kruskal-Wallis.

Οι διακυμάνσεις της ποσοτικής μεταβλητής στους *k* πληθυσμούς θα πρέπει να είναι ίσες. Η προϋπόθεση αυτή είναι απαραίτητο να διασφαλίζεται κατά τη χρήση της μεθόδου. Ο έλεγχος της ισότητας των διακυμάνσεων γίνεται με τη βοήθεια του τεστ του Levene. Η μοναδική περίπτωση να παρακαμφθεί η συγκεκριμένη προϋπόθεση είναι όταν ο αριθμός των παρατηρήσεων ανά ομάδα να είναι περίπου ίδιος. Για τις περιπτώσεις που δεν διασφαλίζεται η ισότητα των διακυμάνσεων, η διαδικασία One-way ANOVA διαθέτει τα τεστ των Brown-Forsythe και Welch.

Τα δείγματα θα πρέπει να έχουν ληφθεί με τυχαία δειγματοληψία από τους αντίστοιχους πληθυσμούς και να είναι ανεξάρτητα το ένα του άλλου. Αυτό σημαίνει ότι η επιλογή του ενός δείγματος δεν επηρεάζεται από την επιλογή των υπολοίπων. Χαρακτηριστική περίπτωση εξαρτημένων δειγμάτων είναι, όταν οι ομάδες αφορούν επαναλαμβανόμενες μετρήσεις της ίδια ποσοτικής μεταβλητής στα ίδια άτομα σε διαφορετικές χρονικές στιγμές. Σε αυτού του τύπου τα σχέδια απαιτείται κατάλληλη μέθοδος ανάλυσης όπως είναι αυτή της *ανάλυσης διακύμανσης επαναλαμβανόμενων μετρήσεων* (*repeated measures analysis of variance*).

Έλεγχοι πολλαπλών συγκρίσεων

Στην ανάλυση διακύμανσης προς έναν παράγοντα, η απόρριψη της μηδενικής υπόθεσης

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

οδηγεί λογικά στην αποδοχή της εναλλακτικής, H_A, σύμφωνα με την οποία δύο τουλάχιστον από τις k πληθυσμιακές μέσες τιμές διαφέρουν μεταξύ τους. Η διατύπωση αυτή της εναλλακτικής υπόθεσης, δεν βοηθά στον ακριβή προσδιορισμό των μέσων τιμών που διαφέρουν και, επομένως, προκειμένου αυτές να εντοπιστούν είναι απαραίτητο να γίνει χρήση συμπληρωματικών ελέγχων. Οι έλεγχοι οι οποίοι χρησιμοποιούνται σε μια τέτοια περίπτωση ονομάζονται *έλεγχοι πολλαπλών συγκρίσεων* (multiple comparisons tests).

Ο έλεγχος ή αλλιώς διόρθωση του Bonferroni βασίζεται σε απλά *t* tests, τα οποία γίνονται μεταξύ όλων των πληθυσμιακών μέσων τιμών αλλά με τροποποιημένο επίπεδο σημαντικότητας.

Έχουμε αναφέρει ότι πραγματοποιώντας πολλαπλά *t* tests μεταξύ *k* πληθυσμιακών μέσων τιμών, αυξάνουμε σημαντικά τη συνολική πιθανότητα για εσφαλμένη απόρριψη της μηδενικής υπόθεσης σε έναν από τους ελέγχους. Με τη διόρθωση του Bonferroni μπορούμε να παρακάμψουμε το συγκεκριμένο πρόβλημα, καθιστώντας τους ελέγχους αυτούς πιο «συντηρητικούς».

Η λογική της διόρθωσης στοχεύει στην ελάττωση του επιπέδου σημαντικότητας των επιμέρους ελέγχων σε τέτοιο βαθμό, ώστε το συνολικό επίπεδο σημαντικότητα να διατηρείται κάτω από μια προκαθορισμένη τιμή.

Το επίπεδο σημαντικότητας που πρέπει να διασφαλιστεί σε κάθε επιμέρους έλεγχο, ώστε να «προστατεύεται» το συνολικό επίπεδο σημαντικότητας όλων των ελέγχων, εξαρτάται από τον αριθμό των δυνατών συγκρίσεων που μπορούν να γίνουν. Για να θέσουμε τη συνολική πιθανότητα εσφαλμένης απόρριψης της μηδενικής υπόθεσης π.χ. στο 0,05, πρέπει να χρησιμοποιήσουμε ως επίπεδο σημαντικότητας του κάθε επιμέρους ελέγχου τουλάχιστον το

$$\alpha^* = \frac{0{,}05}{m}$$

όπου *m* είναι ο αριθμός των δυνατών συγκρίσεων μεταξύ των *k* μέσων τιμών .

Στην περίπτωση όπου έχουμε π.χ. k=3 ομάδες, απαιτούνται να γίνουν συνολικά 3 επιμέρους έλεγχοι. Επομένως αν θέλουμε να διασφαλίσουμε ότι το συνολικό επίπεδο σημαντικότητας δεν θα υπερβεί το 0,05, πρέπει να χρησιμοποιήσουμε ως επίπεδο σημαντικότητας κάθε επιμέρους ελέγχου το

$$\alpha^* = \frac{0{,}05}{3} = 0{,}016$$

Το πρόβλημα με τη διόρθωση του Bonferroni είναι ότι, όταν ο αριθμός των συγκρίσεων αυξάνει, ελαττώνεται σημαντικά η ισχύς του συνολικού ελέγχου, με αποτέλεσμα να μην μπορεί να τεκμηριώσει σημαντικές διαφορές και σε περιπτώσεις όπου αυτές πράγματι υπάρχουν. Και αυτό διότι, σε μια τέτοια περίπτωση, το συνολικό επίπεδο σημαντικότητας με βάση το οποίο θα γίνουν όλοι οι έλεγχοι, γίνεται πλέον πολύ μικρό και επομένως δύσκολα μπορεί να απορριφθούν οι μηδενικές υποθέσεις των επιμέρους ελέγχων. Το τεστ του Tukey είναι μια πιο αποτελεσματική εναλλακτική λύση σε αυτές τις περιπτώσεις .

Από τα υπόλοιπα multiple comparisons tests που διαθέτει το SPSS, αξίζει να αναφερθεί (για ιστορικούς λόγους) ο έλεγχος της ελάχιστης σημαντικής διαφοράς (least significant difference, LSD) του R.A. Fisher. Ο συγκεκριμένος έλεγχος είναι αρκετά ελαστικός (ουσιαστικά είναι μια διευρυμένη εκδοχή του *t* test), ο οποίος σε αρκετές περιπτώσεις δεν διασφαλίζει τη συνολική πιθανότητα όλων των ελέγχων σε επίπεδα κάτω του 0,05. Το ίδιο έχει αποδειχθεί και για τους ελέγχους S-N-K και Duncan.

Στην αντίθετη ακριβώς πλευρά, ο έλεγχος του Scheffé θεωρείται πολύ «συντηρητικός», διότι το επίπεδο σημαντικότητας που διασφαλίζει, αφορά όχι μόνο την περίπτωση των συγκρίσεων των μέσων τιμών ανά δύο, αλλά και οποιοδήποτε άλλο γραμμικό συνδυασμό των μέσων τιμών μπορεί να οριστεί.

Τέλος το τεστ του Dunnett, έχει ενδιαφέρον σε περιπτώσεις όπου συγκρίνεται ένα σύνολο μέσων τιμών διαφορετικών ομάδων (μεταχειρίσεων) σε σχέση με τη μέση τιμή μιας ομάδας ελέγχου (control group).

Έλεγχοι που χρησιμοποιούνται σε περιπτώσεις άνισων διακυμάνσεων

Το SPSS διαθέτει τέσσερις ελέγχους, σε περίπτωση που οι πληθυσμιακές διακυμάνσεις των *k* ομάδων αποδειχθεί ότι διαφοροποιούνται μεταξύ τους. Είναι οι Tamhane's T2, Dunnett's T3, Games-Howell και Dunnett's C. Από τους ελέγχους αυτούς, ο Tamhane's T2 θεωρείται αρκετά «συντηρητικός» όσον αφορά την ισχύ του ενώ αντίθετα ο έλεγχος των Games-Howell θεωρείται αρκετά ελαστικός. Προτείνεται η χρήση των δύο τεστ του Dunnett (Dunnett's T3 και Dunnett's C) .