Ανάλυση Διασποράς

Η ***Ανάλυση Διασποράς (Analysis of Variance, ANOVA)*** είναι μέθοδος *στατιστικού ελέγχου υποθέσεων* που αναφέρονται σε περισσότερους από δύο πληθυσμούς.

Στην προηγούμενη ενότητα αναφερθήκαμε σε *στατιστικούς ελέγχους υποθέσεων* για την τιμή της *μέσης τιμής,* μ, ή της *διασποράς,* σ^2, ενός πληθυσμού και σε *στατιστικούς ελέγχους υποθέσεων* για τη σύγκριση των *μέσων τιμών,* μ_1 και μ_2, ή των *διασπορών,* σ_1^2 και σ_2^2, δύο πληθυσμών, αντίστοιχα.

Εύλογα γεννάται το ερώτημα: *πώς μπορούμε να συγκρίνουμε, για παράδειγμα, τις μέσες τιμές* $\mu_1, \mu_2, ..., \mu_k$, *k πληθυσμών όταν* $k > 2$*;*

Μια απάντηση θα μπορούσε να είναι η εξής: να γίνει η σύγκριση των k *μέσων τιμών* ανά δύο. Δηλαδή, να κάνουμε όλους τους, $c = \binom{k}{2} = \frac{k \cdot (k-1)}{2}$, ελέγχους:

$$\begin{array}{l} H_0 : \mu_i = \mu_j \\ H_1 : \mu_i \neq \mu_j . \end{array}, \text{ με } i \neq j .$$

Είναι φανερό, ότι μια τέτοια διαδικασία είναι χρονοβόρα ακόμη και όταν το k είναι μικρό. Για παράδειγμα, αν θέλουμε να συγκρίνουμε $k = 4$ *μέσες τιμές,* πρέπει να κάνουμε $c = 6$, συνολικά, ελέγχους. Αν $k = 5$ πρέπει να κάνουμε $c = 10$ ελέγχους, αν $k = 6$ πρέπει να κάνουμε $c = 15$ ελέγχους, ενώ αν $k = 10$, απαιτούνται $c = 45$ έλεγχοι! Όμως, αυτό το πρόβλημα θα μπορούσε να αντιμετωπισθεί (μάλιστα, με χρήση κατάλληλου λογισμικού, πολύ εύκολα). Το πραγματικό πρόβλημα που δημιουργείται είναι άλλο και σοβαρό.

Αν ενδιαφερόμαστε να κάνουμε c **ανεξάρτητους** ελέγχους, σε *επίπεδο σημαντικότητας* α_{Pc} τον καθένα *(πιθανότητα σφάλματος τύπου Ι κατά σύγκριση/per comparison error rate),* τότε η πιθανότητα λανθασμένης απόρριψης της *μηδενικής υπόθεσης* σε έναν τουλάχιστον από αυτούς, δηλαδή, η πιθανότητα, στους c ελέγχους, να συμβεί τουλάχιστον μια φορά *σφάλμα Τύπου Ι (πιθανότητα σφάλματος τύπου Ι κατά πείραμα/per experiment error rate),* είναι: $\alpha_{PE} = 1 - (1 - \alpha_{PC})^c$.

Για παράδειγμα, αν κάνουμε $c = 6, 10, 15, 45$ ανεξάρτητους ελέγχους σε επίπεδο σημαντικότητας $\alpha_{PC} = 0.05$ τον καθένα, τότε, η πιθανότητα, α_{PE}, είναι, αντίστοιχα:

c	6	10	15	45
α_{PE}	0.26	0.40	0.54	0.90

Είναι φανερό ότι το α_{PE} αυξάνεται ραγδαία με τον αριθμό των ελέγχων, c. Ανάλογο πρόβλημα δημιουργείται και στην περίπτωση ***μη ανεξάρτητων*** ελέγχων, όπου, ισχύει, $\alpha_{PE} \leq c \cdot \alpha_{PC}$ *(υποπροσθετική ιδιότητα).* Για την αντιμετώπιση, επομένως, αυτού του προβλήματος, το οποίο στη βιβλιογραφία αναφέρεται ως ***πρόβλημα πολλαπλών συγκρίσεων (multiple comparisons problem)***, **απαιτούνται άλλες προσεγγίσεις**.

Στην ενότητα αυτή θα δούμε πώς απαντάει στο πρόβλημα της σύγκρισης περισσότερων των δύο *μέσων,* η *ανάλυση διασποράς*. Ειδικότερα, θα δούμε πώς και με ποιες προϋποθέσεις μπορούμε να κάνουμε, σε *επίπεδο σημαντικότητας* $\alpha\%$, *έλεγχο ανάλυσης διασποράς* για να ελέγξουμε τη *μηδενική υπόθεση,*

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k \ ,$$

έναντι της *εναλλακτικής,*

$$H_1 : \mu_i \neq \mu_j \text{, για ένα τουλάχιστον ζεύγος (i, j).}$$

Δηλαδή, θα δούμε, πώς μπορούμε να συγκρίνουμε τις *μέσες τιμές* $\mu_1, \mu_2, ..., \mu_k$, $k > 2$ *πληθυσμών,* με έναν μόνο, αντί με πολλαπλούς ελέγχους. Όμως, παρότι φαίνεται ότι με τον έλεγχο *ανάλυσης διασποράς* το πρόβλημα των πολλαπλών ελέγχων παρακάμπτεται, εντούτοις, όπως θα δούμε αργότερα, τελικά δεν το αποφεύγουμε. Και αυτό, γιατί αν με τον *έλεγχο ανάλυσης διασποράς,* η *μηδενική υπόθεση* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$, **απορριφθεί**, πρέπει να μπορούμε να πούμε κάτι για το ποιες από τις διαφορές $\mu_i - \mu_j$ είναι στατιστικά σημαντικές. Όμως, τώρα, ας δούμε τον *έλεγχο ανάλυσης διασποράς*.

Η *ανάλυση διασποράς* προτάθηκε από τον *Sir Ronald A. Fisher* το *1918.* Ευρέως έγινε γνωστή μετά το 1925 όταν εκδόθηκε το κλασικό πλέον βιβλίο του *R. A. Fisher, Statistical Methods for Research Workers,* στο οποίο είχε συμπεριλάβει και την *ανάλυση διασποράς*.

Η *ανάλυση διασποράς* «γεννήθηκε»/προέκυψε κατά την ενασχόληση του *Fisher* με δύσκολα προβλήματα *στατιστικής συμπερασματολογίας* που εμφανίζονται στο ***γεωργικό πειραματισμό*** *(πολλές πηγές μεταβλητότητας και συχνά εμφανιζόμενες ετερογένειες, και μάλιστα, προς διάφορες κατευθύνσεις του πειραματικού αγρού π.χ. ως προς τη γονιμότητα, την κλίση και την υγρασία των εδαφών, τις προηγούμενες καλλιέργειες, κ.τλ.).* Η προσέγγιση της λύσης τέτοιου είδους προβλημάτων που πρότεινε ο *Fisher,* βασίζεται στην ***τυχαιοποίηση*** και στην ***επανάληψη*** και ως μαθηματικό εργαλείο για την υποστήριξη αυτής της προσέγγισης πρότεινε την *ανάλυση διασποράς*. Γι' αυτό, όπως θα δούμε στη συνέχεια, στην *ανάλυση διασποράς* έχει επικρατήσει να χρησιμοποιείται ορολογία (και όχι μόνο) που χρησιμοποιείται στο *γεωργικό πειραματισμό* και γενικότερα στον *πειραματισμό*, παρότι δεν εφαρμόζεται μόνο στην ανάλυση *πειραματικών δεδομένων*. Ας δούμε τέσσερα προβλήματα *στατιστικού ελέγχου υποθέσεων* που αντιμετωπίζονται με *έλεγχο ανάλυσης διασποράς*.

<u>Πρόβλημα-1:</u> Ένας φοιτητής, στο πλαίσιο της πτυχιακής του εργασίας, προκειμένου να συγκρίνει τέσσερα είδη (A1, A2, A3 και A4, αντίστοιχα) πρόσθετης ύλης ζωοτροφών για αύξηση του βάρους νεογέννητων χοίρων, σχεδίασε και εκτέλεσε το εξής πείραμα. Επέλεξε 20 νεογέννητους χοίρους και με μια τυχαία διαδικασία αντιστοίχισε σε 5 από αυτούς την πρόσθετη ύλη A1, σε 5 άλλους την A2, σε 5 άλλους την A3 και σε 5 άλλους την A4 (δες το σχήμα που ακολουθεί). Δημιούργησε έτσι, 4 ομάδες των 5 χοίρων η κάθε μια. Αφού χορήγησε στους χοίρους κάθε ομάδας τροφή με την αντίστοιχη πρόσθετη ύλη για τρεις μήνες, κατέγραψε την αύξηση του βάρους κάθε χοίρου (σε pounds). Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στη μέση αύξηση του βάρους των νεογέννητων χοίρων που να οφείλονται στα τέσσερα είδη πρόσθετης ύλης ζωοτροφών;

A1	*A2*	*A3*	*A4*
81	*78*	*72*	*85*
66	*66*	*70*	*70*
78	*69*	*78*	*83*
76	*64*	*77*	*74*
61	*66*	*69*	*70*

<u>Πρόβλημα-2:</u> Ο επιβλέπων καθηγητής του φοιτητή του Προβλήματος-1, μετά τη στατιστική ανάλυση των πειραματικών δεδομένων, του πρότεινε να εκτελέσει ένα ακόμη πείραμα. Έτσι, ο φοιτητής εκτέλεσε και το εξής πείραμα: Από πέντε <u>διαφορετικές γέννες</u> επέλεξε τυχαία 4 χοίρους από την κάθε μία. Δημιούργησε έτσι, 5 τετράδες χοίρων (Γ1, Γ2, Γ3, Γ4 και Γ5, αντίστοιχα) όπου οι χοίροι της ίδιας τετράδας προέρχονταν από την ίδια γέννα, και στους χοίρους κάθε τετράδας αντιστοίχησε, με μια τυχαία διαδικασία, από μία πρόσθετη ύλη ζωοτροφών (δες το σχήμα που ακολουθεί). Αφού χορήγησε στους χοίρους κάθε ομάδας τροφή με την αντίστοιχη πρόσθετη ύλη στον καθένα για τρεις μήνες, κατέγραψε την αύξηση του βάρους κάθε χοίρου (σε pounds). Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στη μέση αύξηση του βάρους των νεογέννητων χοίρων που να οφείλονται στα τέσσερα είδη πρόσθετης ύλης ζωοτροφών;

Γ1	*Γ2*	*Γ3*	*Γ4*	*Γ5*
Α3 (78)	*Α4 (70)*	*Α2 (70)*	*Α2 (66)*	*Α1 (61)*
Α1 (78)	*Α2 (64)*	*Α4 (83)*	*Α1 (76)*	*Α3 (69)*
Α2 (69)	*Α1 (68)*	*Α3 (73)*	*Α4 (74)*	*Α2 (66)*
Α4 (80)	*Α3 (70)*	*Α1 (76)*	*Α3 (77)*	*Α4 (70)*

<u>Πρόβλημα-3:</u> Ένας φοιτητής του Τμήματος Επιστήμης και Τεχνολογίας Τροφίμων του Γ.Π.Α., στο πλαίσιο της πτυχιακής του εργασίας, συνέκρινε τις ποσότητες χοληστερίνης που περιέχουν τρία διαφορετικά είδη τροφίμων διαίτης, Α1, Α2 και Α3, αντίστοιχα. Από κάθε είδος πήρε ένα τυχαίο δείγμα μεγέθους 8 και μέτρησε την ποσότητα χοληστερίνης (σε milligrams ανά 100gr). Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν οι παρατηρήσεις που πήρε ο φοιτητής ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων ποσοτήτων χοληστερίνης στα τρία είδη τροφίμων;

<u>Πρόβλημα-4:</u> Ένας ερευνητής ενδιαφέρεται να μελετήσει την επίδραση της έντασης του φωτισμού και του επιπέδου της θερμοκρασίας στην ανάπτυξη ενός είδους χλωρόφυτων (φύκη). Μάλιστα, ενδιαφέρεται ιδιαίτερα, να διερευνήσει αν οι δύο αυτοί παράγοντες επιδρούν στην ανάπτυξη του συγκεκριμένου είδους χλωρόφυτου ανεξάρτητα ή η επίδραση του ενός διαφοροποιείται ανάλογα με τη στάθμη του άλλου. Για το σκοπό αυτό, σχεδίασε και εκτέλεσε στο εργαστήριο το εξής πείραμα. Καθόρισε τρία επίπεδα έντασης φωτισμού (Β1, Β2 και Β3) και τρία επίπεδα θερμοκρασίας (10^oC, 20^oC και 30^oC). Για κάθε συνδυασμό, έντασης φωτισμού και επιπέδου θερμοκρασίας, καλλιέργησε 2 χλωρόφυτα του συγκεκριμένου είδους και μετά από μια εβδομάδα τα ζύγισε και κατέγραψε τα βάρη τους (σε gr). Τα δεδομένα που συγκέντρωσε φαίνονται στον πίνακα που ακολουθεί.

		Επίπεδο θερμοκρασίας		
		10^oC	*20^oC*	*30^oC*
Ένταση φωτισμού	*Β1*	*150 160*	*250 230*	*200 200*
	Β2	*170 150*	*260 280*	*190 150*
	Β3	*200 210*	*270 280*	*110 140*

Με βάση αυτά τα πειραματικά δεδομένα και σε επίπεδο σημαντικότητας 5%, τι μπορούμε να πούμε για το αν και πώς το επίπεδο θερμοκρασίας και η ένταση φωτισμού επιδρούν στην ανάπτυξη του συγκεκριμένου είδους χλωρόφυτων;

Είναι προφανές ότι και τα τέσσερα προβλήματα είναι προβλήματα *στατιστικού ελέγχου υποθέσεων* που αφορούν περισσότερες από δύο *μέσες τιμές* (αντίστοιχων πληθυσμών). Στο *Πρόβλημα-3*, τα δεδομένα προκύπτουν από ***δειγματοληψίες***, ενώ στα *Προβλήματα-1, 2* και *4* από ***εκτέλεση πειραμάτων*** και μάλιστα με διαφορετικό *πειραματικό σχέδιο* το καθένα. Παρατηρείστε ότι στα *Προβλήματα-1* και *2*, ενώ το ερευνητικό ερώτημα είναι ακριβώς το ίδιο, ο τρόπος συγκέντρωσης των πειραματικών δεδομένων, δηλαδή το *πειραματικό σχέδιο*, διαφέρει. Παρατηρείστε, επίσης, ότι το ***σχέδιο δειγματοληψίας*** στο *Πρόβλημα-3* και το ***πειραματικό σχέδιο*** στο *Πρόβλημα-1* είναι ίδια.

Για να απαντήσουμε στο *Πρόβλημα-1,* πρέπει να κάνουμε, σε *επίπεδο σημαντικότητας 5%,* στατιστικό έλεγχο της *μηδενικής υπόθεσης,*
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$,
έναντι της *εναλλακτικής,*
$H_1 : \mu_i \neq \mu_j$, *για ένα τουλάχιστον ζεύγος (i, j),*
όπου, $\mu_j,\ j = 1,2,3,4$ η *μέση τιμή* της *τυχαίας μεταβλητής* που εκφράζει την *αύξηση του βάρους νεογέννητων χοίρων μετά τη χορήγηση τροφής που περιέχει πρόσθετη ύλη* $A_j,\ j = 1,2,3,4$*, για τρεις μήνες μετά τη γέννησή τους*. Ουσιαστικά, και με ορολογία *πειραματισμού*, πρέπει να ελέγξουμε, αν επηρεάζονται οι τιμές μιας *ποσοτικής* μεταβλητής, έστω *Y, (αύξηση του βάρους νεογέννητων χοίρων όταν χρησιμοποιείται πρόσθετη ύλη στην τροφή τους για τρεις μήνες μετά τη γέννησή τους),* από τις τιμές (*A1, A2, A3, A4*) μιας *ποιοτικής* μεταβλητής *(το είδος της πρόσθετης ύλης)*.

Στην ορολογία του πειραματισμού, η *ποιοτική* μεταβλητή ονομάζεται ***παράγοντας (factor)*** και οι τιμές της ονομάζονται ***στάθμες*** ή ***επίπεδα (levels)*** του *παράγοντα.* Η εξαρτημένη μεταβλητή *Y* ονομάζεται ***μεταβλητή απόκρισης (response variable)*** γιατί στις τιμές της αντανακλώνται οι αλλαγές στις τιμές του *παράγοντα.*

Σε ένα πείραμα, ο ερευνητής μπορεί να ενδιαφέρεται για συγκεκριμένες τιμές του *παράγοντα* οπότε ως *στάθμες* ορίζει αυτές τις τιμές και τα συμπεράσματα θα αφορούν μόνον αυτές ***(πρότυπο I/πρότυπο καθορισμένων επιδράσεων/fixed effects model)***. Για παράδειγμα, ενδιαφέρεται να μελετήσει την αποτελεσματικότητα τεσσάρων συγκεκριμένων ειδών πρόσθετης ύλης ζωοτροφών. Αν τις τιμές του παράγοντα *(στάθμες)* τις επιλέξει τυχαία από ένα σύνολο δυνατών τιμών του παράγοντα (άπειρο ή πεπερασμένο) τότε τα συμπεράσματά του θα αφορούν όλο τον πληθυσμό των δυνατών τιμών του παράγοντα ***(πρότυπο II/πρότυπο τυχαίων επιδράσεων/random effects model)***. Για παράδειγμα, ενδιαφέρεται να μελετήσει την αποτελεσματικότητα δέκα ειδών ζωοτροφών που υπάρχουν και ως στάθμες ορίζει τέσσερα από αυτά τα είδη που τα επέλεξε τυχαία από τα δέκα που υπάρχουν ή επιλέγει τυχαία πέντε περιοχές της χώρας για να καλλιεργήσει μια ποικιλία καλαμποκιού γιατί ενδιαφέρεται τα συμπεράσματα του να αφορούν όλη τη χώρα.

Στο *Πρόβλημα-2* όπως και στο *Πρόβλημα-1*, ο φοιτητής ενδιαφέρεται να μελετήσει την επίδραση του *παράγοντα «είδος πρόσθετης ύλης»* (για τις ίδιες *στάθμες* όπως στο *Πρόβλημα-1*) στην *«αύξηση του βάρους νεογέννητων χοίρων όταν χρησιμοποιείται πρόσθετη ύλη στην τροφή τους για τρεις μήνες μετά τη γέννησή τους»*. Δηλαδή, τα δύο προβλήματα, διαφοροποιούνται μόνο στο *πειραματικό σχέδιο* που εφαρμόσθηκε.

Στο *Πρόβλημα-3* η *μεταβλητή απόκρισης* εκφράζει τις *«τιμές χοληστερίνης στα τρόφιμα διαίτης»,* και ο *παράγοντας* είναι το *«είδος τροφίμου διαίτης»* με τρεις *στάθμες*.

Στο *Πρόβλημα-4* η *μεταβλητή απόκρισης* εκφράζει την *«ανάπτυξη του συγκεκριμένου είδους χλωρόφυτων»* και ο ερευνητής ενδιαφέρεται να μελετήσει την επίδραση δύο παραγόντων: του *παράγοντα «επίπεδο θερμοκρασίας»* με τρεις *στάθμες* και του *παράγοντα «επίπεδο φωτισμού»* με τρεις επίσης *στάθμες*. Κάθε συνδυασμός από δύο *στάθμες* (μία από κάθε *παράγοντα*) ονομάζεται ***επέμβαση/αγωγή/μεταχείριση (treatment)***. Στο συγκεκριμένο πρόβλημα οι διαφορετικές *επεμβάσεις (οι διαφορετικοί συνδυασμοί επιπέδου θερμοκρασίας-έντασης φωτισμού)* είναι 9 και για κάθε *επέμβαση* ελήφθησαν 2 παρατηρήσεις.

Στην περίπτωση που μελετάμε την επίδραση ενός μόνο *παράγοντα,* οι *επεμβάσεις* είναι οι *στάθμες* του *παράγοντα*. Για παράδειγμα, στο *Πρόβλημα-3*, οι *επεμβάσεις* είναι τα τρία είδη των τροφίμων διαίτης *Α1, Α2* και *Α3.*

Ας δούμε τώρα ποια είναι η βασική, η κεντρική, ιδέα της *ανάλυσης διασποράς*, το νόημά της και πώς εφαρμόζεται. Παρότι, η βασική ιδέα και το νόημα της *ανάλυσης διασποράς* είναι το ίδιο, ανεξάρτητα από το *πειραματικό σχέδιο* ή το *σχέδιο δειγματοληψίας* που χρησιμοποιούμε για τη συγκέντρωση των δεδομένων, εντούτοις, κατά περίπτωση, απαιτούνται κατάλληλες προσαρμογές στην εφαρμογή της μεθόδου (και είναι λογικό). Γι' αυτό, στη συνέχεια θα παρουσιάσουμε τη μέθοδο ανά *πειραματικό σχέδιο*. Ειδικότερα, θα δούμε με ποιες προϋποθέσεις και πώς, μπορούμε να κάνουμε έλεγχο/ελέγχους *ανάλυσης διασποράς* στο ***Εντελώς Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)*** και στο ***Σχέδιο Τυχαιοποιημένων Πλήρων Ομάδων (Randomized Complete Block Design)*** καθώς και στα ***παραγοντικά πειράματα*** και ειδικότερα στο ***a x b Παραγοντικό Πείραμα (a x b Factorial Experiment)***.

Πριν προχωρήσουμε στην παρουσίαση της μεθόδου, αναφέρουμε και υπογραμμίζουμε τις ***υποθέσεις***, ή αλλιώς, τις ***παραδοχές***, που κάνουμε και πρέπει να ελέγχουμε αν ισχύουν/ικανοποιούνται όταν επιλέγουμε να εφαρμόσουμε ελέγχους ***ανάλυσης διασποράς***.

Υποθέσεις για την εφαρμογή ελέγχων ανάλυσης διασποράς

*1. Οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα είναι **κανονικοί** με **κοινή διασπορά**.*

2. Οι υποθέσεις, κατά περίπτωση, που αφορούν τον τρόπο λήψης των παρατηρήσεων, δηλαδή, το πειραματικό σχέδιο.

<u>Εντελώς Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)</u>

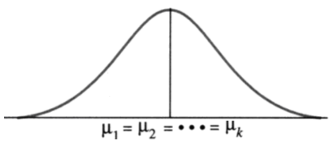
Πρόκειται για το πιο απλό *πειραματικό σχέδιο,* σύμφωνα με το οποίο, εργαζόμαστε με *k* ***ανεξάρτητα τυχαία δείγματα***, ένα από κάθε *πληθυσμό* (ή αλλιώς, ένα από κάθε *στάθμη* του *παράγοντα ή* ένα από κάθε *επέμβαση*) και αποτελεί ευθεία γενίκευση του σχεδίου που γνωρίσαμε στα προηγούμενα όταν μιλήσαμε για τον έλεγχο των *μέσων*, μ_1 και μ_2, δύο *κανονικών πληθυσμών* με *δύο ανεξάρτητα τυχαία δείγματα*. Τέτοια σχέδια είναι αυτά των *Προβλημάτων 1* και *3*.

Προτείνεται για περιπτώσεις όπου θεωρούμε ότι μόνο ένας *παράγοντας* ξεχωρίζει ως αξιοσημείωτη πηγή μεταβλητότητας της *μεταβλητής απόκρισης* και κάθε άλλη μεταβλητότητα στις τιμές της *μεταβλητής απόκρισης*, πέραν αυτής που οφείλεται σε αυτόν τον *παράγοντα* (*επεμβάσεις*), θεωρείται *τυχαία μεταβλητότητα/πειραματικό σφάλμα*. Δηλαδή, θεωρούμε ότι οι *πειραματικές μονάδες* δεν προσθέτουν μεταβλητότητα στις τιμές της *μεταβλητής απόκρισης*. Για παράδειγμα, στο *Πρόβλημα-3*, ο φοιτητής, ως πηγή μεταβλητότητας των τιμών της χοληστερίνης, ξεχώρισε (για να μελετήσει) μόνο τον *παράγοντα είδος τροφίμου διαίτης* και όχι και κάποιον άλλο (π.χ. *παραγωγός του τροφίμου διαίτης*) και στο *Πρόβλημα-1*, ο φοιτητής ξεχώρισε μόνο τον *παράγοντα είδος πρόσθετης ύλης* ως πηγή μεταβλητότητας της αύξησης του βάρους των νεογέννητων χοίρων. Άλλους παράγοντες που πιθανόν επηρεάζουν τη μεταβλητότητα της αύξησης του βάρους των νεογέννητων χοίρων (π.χ. ατομικότητα του ζώου, φυλή, γένος, συνθήκες ενσταυλισμού) δεν τους έλαβε υπόψη του στο σχεδιασμό.

Ας θεωρήσουμε, ότι από καθέναν από k (>2) ***κανονικούς*** πληθυσμούς με ***κοινή διασπορά***, σ^2, και *μέσες τιμές*, αντίστοιχα, $\mu_1, \mu_2, ..., \mu_k$, παίρνουμε ένα ***τυχαίο δείγμα*** μεγέθους, αντίστοιχα, $n_1, n_2, ..., n_k$ για να κάνουμε, με βάση αυτά τα k δείγματα, τον έλεγχο,

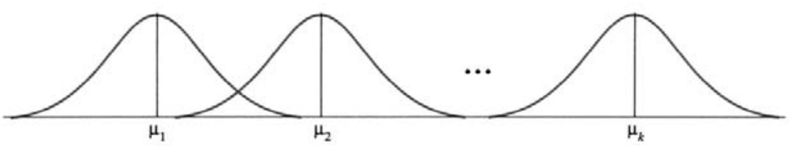
της *μηδενικής υπόθεσης*,

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



έναντι της *εναλλακτικής*,

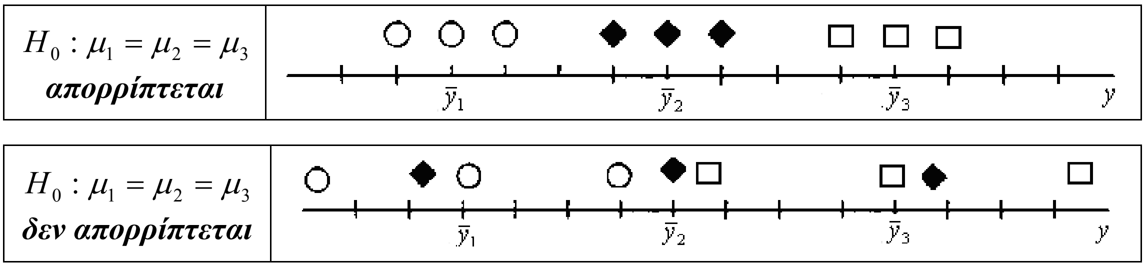
$H_1 : \mu_i \neq \mu_j$ *για ένα τουλάχιστον ζεύγος (i, j)*



Πριν συνεχίσουμε την παρουσίαση της μεθόδου, επισημαίνουμε ότι, επειδή θεωρούμε ότι οι k πληθυσμοί είναι ***κανονικοί*** με ***κοινή διασπορά***, σ^2, τότε, <u>*αν η μηδενική υπόθεση είναι αληθής*</u>, αυτό σημαίνει ότι τα k δείγματα προέρχονται από τον ίδιο πληθυσμό με κατανομή $N(\mu, \sigma^2)$, όπου, $\mu = \mu_1 = \mu_2 = ... = \mu_k$. Σημειώνουμε, επίσης, ότι ο συγκεκριμένος *στατιστικός έλεγχος ανάλυσης διασποράς,* εφαρμόζεται και για $k = 2$ και μάλιστα, <u>είναι ισοδύναμος</u>, δηλαδή οδηγεί σε ταυτόσημα αποτελέσματα, με τον έλεγχο δύο *μέσων* με δύο *ανεξάρτητα τυχαία* δείγματα που γνωρίσαμε στα προηγούμενα.

Η βασική ιδέα της μεθόδου είναι η εξής: *όλες οι διαφορές στους δειγματικούς μέσους (στους μέσους των επεμβάσεων),* $\bar{y}_1, \bar{y}_2, ..., \bar{y}_k$, *κρίνονται στατιστικά σημαντικές ή όχι, σε σχέση με τις μεταβλητότητες (διασπορές) εντός των δειγμάτων.*

Έτσι, παρότι στις δύο περιπτώσεις που φαίνονται στο *Σχήμα-1*, οι αντίστοιχοι *δειγματικοί μέσοι* είναι ίσοι, στην πρώτη περίπτωση οι διαφορές που παρατηρούνται στους τρεις *δειγματικούς μέσους* κρίνονται στατιστικά σημαντικές και η *μηδενική υπόθεση* απορρίπτεται γιατί οι μεταβλητότητες ***εντός*** των δειγμάτων είναι μικρές, ενώ στη δεύτερη περίπτωση, ίδιες διαφορές στους *δειγματικούς μέσους,* δεν κρίνονται στατιστικά σημαντικές και η *μηδενική υπόθεση* δεν απορρίπτεται, γιατί οι μεταβλητότητες ***εντός*** των δειγμάτων είναι μεγάλες.



Σχήμα-1

Αυτή η ιδέα, «τεχνικά», υλοποιείται ως εξής: <u>Εκτιμάμε την *κοινή διασπορά,*</u> σ^2, των *k κανονικών πληθυσμών* (είναι άγνωστη συνήθως), <u>με δύο τρόπους</u>:

<u>1ος τρόπος</u>. Με το σταθμισμένο μέσο s_w^2 των *k* *δειγματικών διασπορών*, s_1^2, $s_2^2, ..., s_k^2$, δηλαδή, με το σταθμισμένο μέσο των *διασπορών* ***εντός*** των *δειγμάτων* (όπως κάναμε στον έλεγχο δύο *μέσων* με δύο *ανεξάρτητα τυχαία δείγματα*):

$$s_w^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2 ... + (n_k - 1)\cdot s_k^2}{\nu - k},$$

όπου, $\nu = n_1 + n_2 + ... + n_k$.

Αυτή η εκτίμηση είναι μια «λογική και καλή» εκτίμηση της από *κοινού διασποράς*, σ^2, ανεξάρτητα αν οι *k* μέσοι, $\mu_1, \mu_2, ..., \mu_k$, είναι ίσοι ή όχι, δηλαδή, ανεξάρτητα αν η *μηδενική υπόθεση* είναι ή όχι αληθής.

<u>2ος τρόπος</u>. Με την $s_b^2 = \dfrac{\sum_{j=1}^{k} n_j \cdot (\bar{y}_j - \bar{y})^2}{k-1}$, όπου, $\bar{y}$ ο *γενικός μέσος* (ο μέσος όλων των $\nu = n_1 + n_2 + ... + n_k$ παρατηρήσεων) και $\bar{y}_j$ ο μέσος του *j* δείγματος. Με αυτό τον τρόπο αξιοποιούμε τη σχέση που συνδέει τη *διασπορά* του πληθυσμού, σ^2, με τη *διασπορά* του *δειγματικού μέσου,* $\sigma_{\bar{Y}}^2$, δηλαδή, αξιοποιούμε τη σχέση, $\sigma^2 = n\sigma_{\bar{Y}}^2$, που συνδέει τη *διασπορά* του πληθυσμού με τη *διασπορά* ***μεταξύ*** των *δειγμάτων* που παίρνουμε από αυτόν. Βέβαια, αυτή η εκτίμηση, είναι «λογική» και περιμένουμε να είναι περίπου ίση με την s_w^2, <u>μόνο αν η *μηδενική υπόθεση* είναι αληθής</u>, δηλαδή, μόνο αν $\mu_1 = \mu_2 = ... = \mu_k = \mu$, γιατί στην περίπτωση αυτή τα *k* δείγματα θα προέρχονται από τον ίδιο *κανονικό* πληθυσμό, $N(\mu, \sigma^2)$, και επομένως μπορούμε τότε να αξιοποιήσουμε τη σχέση $\sigma^2 = n\sigma_{\bar{Y}}^2$.

Αν επομένως, η *μηδενική υπόθεση* είναι αληθής, αναμένουμε ο λόγος $\dfrac{s_b^2}{s_w^2}$ να είναι κοντά στο ένα, ενώ αν δεν είναι αληθής, περιμένουμε να αυξάνεται γιατί θα αυξάνεται ο αριθμητής που εμπεριέχει/ενσωματώνει και τη μεταβλητότητα στους *δειγματικούς μέσους*. Έτσι, ως *κριτήριο/στατιστική συνάρτηση ελέγχου* μπορούμε να χρησιμοποιήσουμε το λόγο:

$$\frac{\textit{μεταβλητότητα μεταξύ των δειγμάτων}}{\textit{μεταβλητότητα εντός των δειγμάτων}}$$

Ας συμβολίσουμε αυτόν το λόγο με F_{Tr}. Προφανώς, με βάση τα προηγούμενα, ο λόγος αυτός, μπορεί να εκφρασθεί ως, $F_{tr} = \dfrac{SSTr/(k-1)}{SSE/(\nu-k)} = \dfrac{MSTr}{MSE}$, όπου,

$SSTr = \sum_j n_j \cdot (\bar{y}_j - \bar{y})^2$, ονομάζεται <u>*άθροισμα τετραγώνων των επεμβάσεων (ή αλλιώς, **μεταξύ** των δειγμάτων)*</u> και εκφράζει τη <u>*μεταβλητότητα μεταξύ των μέσων των k επεμβάσεων (ή αλλιώς, μεταξύ των μέσων των δειγμάτων)*</u> και έχει $k-1$ *βαθμούς ελευθερίας,*

$SSE = \sum_{ji} (y_{ij} - \bar{y}_j)^2$, ονομάζεται <u>*άθροισμα τετραγώνων του σφάλματος*</u> και εκφράζει τη <u>*μεταβλητότητα **εντός** των δειγμάτων (ή αλλιώς εντός των επεμβάσεων)*</u> και έχει $\nu - k$ *βαθμούς ελευθερίας,*

$MSTr = \dfrac{SSTr}{k-1}$ και $MSE = \dfrac{SSE}{\nu-k}$ ονομάζονται, αντίστοιχα, <u>*μέσο άθροισμα τετραγώνων των επεμβάσεων*</u> και <u>*μέσο άθροισμα τετραγώνων του σφάλματος*</u> και προφανώς είναι οι δύο εκτιμήτριες της κοινής *διασποράς* σ^2 (που εξηγήσαμε προηγουμένως). Μάλιστα, (εύκολα) αποδεικνύεται ότι,

$$\sum_j n_j (\bar{y}_j - \bar{y})^2 + \sum_{ji} (y_{ij} - \bar{y}_j)^2 = \sum_{ji} (y_{ij} - \bar{y})^2 ,$$

όπου, το διπλό άθροισμα $\sum_{ji} (y_{ij} - \bar{y})^2$, συμβολίζεται με *SSTot,* ονομάζεται <u>*ολικό άθροισμα τετραγώνων*</u>, έχει $\nu - 1$ *βαθμούς ελευθερίας* και προφανώς εκφράζει την *ολική μεταβλητότητα* των παρατηρήσεων, ή αλλιώς, τη μεταβλητότητα όλων των παρατηρήσεων γύρω από το *γενικό δειγματικό μέσο* $\bar{y}$.

Δηλαδή, η <u>*ολική μεταβλητότητα*</u>, $SSTot = \sum_{ji} (y_{ij} - \bar{y})^2$, <u>*αναλύεται*</u> σε δύο συνιστώσες:

- στη μεταβλητότητα ***μεταξύ*** των δειγμάτων: $SSTr = \sum_j n_j \cdot (\bar{y}_j - \bar{y})^2$ και
- στη μεταβλητότητα ***εντός*** των δειγμάτων: $SSE = \sum_{ji} (y_{ij} - \bar{y}_j)^2$.

Δηλαδή, $SSTr + SSE = SSTot$.

Αποδεικνύεται ότι, με την υπόθεση ότι η μηδενική υπόθεση, $H_0 : \mu_1 = \mu_2 = ... = \mu_k$, *είναι αληθής, η τυχαία μεταβλητή* F_{Tr} *ακολουθεί την κατανομή F με* $k-1$ *και* $\nu - k$ *βαθμούς ελευθερίας, δηλαδή,*

$$F_{tr} = \frac{MSTr}{MSE} \sim F_{k-1;\nu-k} .$$

Επομένως, σε επίπεδο σημαντικότητας $\alpha\%$, *η απορριπτική περιοχή του ελέγχου είναι:*

$$F_{tr} = \frac{MSTr}{MSE} > F_{k-1;\nu\text{-k};\alpha} .$$

Για διευκόλυνση στον υπολογισμό των *SSTot, SSTr* και *SSE*, δίνουμε τους παρακάτω τύπους (αποδεικνύονται πολύ εύκολα).

$SSTot = \sum_{ji} y_{ij}^2 - \dfrac{G^2}{\nu}$, $SSTr = \sum_{j=1}^{k} \dfrac{T_j^2}{n_j} - \dfrac{G^2}{\nu}$ και $SSE = SSTot - SSTr$, όπου,

$G = \sum_{ji} y_{ij} = \sum_{j=1}^{k} T_j$, το *γενικό άθροισμα (Grand total)* των παρατηρήσεων, δηλαδή το άθροισμα όλων των $\nu = n_1 + n_2 + ... + n_k$, παρατηρήσεων και T_j, $j = 1, 2, ..., k$, το άθροισμα όλων των παρατηρήσεων του j δείγματος.

Για διευκόλυνσή μας, όλες τις πληροφορίες που είναι απαραίτητες για τον έλεγχο *ανάλυσης διασποράς (αθροίσματα τετραγώνων, βαθμοί ελευθερίας, μέσα αθροίσματα τετραγώνων, τιμή του κριτηρίου)* τις συγκεντρώνουμε σε έναν πίνακα που ονομάζεται ***Πίνακας Ανάλυσης Διασποράς (ANOVA Table)***.

Πίνακας Ανάλυσης Διασποράς *(για το εντελώς τυχαιοποιημένο σχέδιο)*					
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***	***Περιοχή απόρριψης***
Επεμβάσεις (Treatments) ή Παράγοντας (factor) ή μεταξύ των δειγμάτων	$k-1$	$SSTr$	$MSTr = \frac{SSTr}{k-1}$	$F_{Tr} = \frac{MSTr}{MSE}$	$F_{Tr} > F_{k-1;\nu-k;\alpha}$
Σφάλμα (Error) ή εντός των δειγμάτων	$\nu - k$	SSE	$MSE = \frac{SSE}{\nu - k}$		
Ολική	$\nu - 1$	$SSTot$			

<u>Σημείωση:</u> Στην περίπτωση που η κοινή διασπορά, σ^2, των k κανονικών πληθυσμών είναι γνωστή (κάτι το οποίο βέβαια δεν είναι σύνηθες σε πραγματικά προβλήματα), ως στατιστική συνάρτηση ελέγχου χρησιμοποιούμε το λόγο, $\frac{SSTr}{\sigma^2}$. Με την υπόθεση ότι η μηδενική υπόθεση είναι αληθής, αποδεικνύεται ότι $\frac{SSTr}{\sigma^2} \sim \chi^2_{k-1}$ και επομένως, σε επίπεδο σημαντικότητας $\alpha\%$, η απορριπτική περιοχή του ελέγχου είναι, $\frac{SSTr}{\sigma^2} > \chi^2_{k-1;\alpha}$.

Ας δούμε, ως ένα παράδειγμα, πώς εφαρμόζουμε τα προηγούμενα για να απαντήσουμε στο ερώτημα που τίθεται στο *Πρόβλημα-1*.

<u>Παράδειγμα-1(αναφέρεται στο Πρόβλημα-1):</u> Προφανώς, το πείραμα σχεδιάσθηκε με βάση το *εντελώς τυχαιοποιημένο σχέδιο*. Έχουμε στη διάθεσή μας $k = 4$, *ανεξάρτητα τυχαία δείγματα* και κάνοντας τις παραδοχές ότι καθένα από αυτά προέρχεται από *κανονικό* πληθυσμό με *διασπορά*, σ^2, κοινή και για τους τέσσερις πληθυσμούς, θα εφαρμόσουμε τη μέθοδο *ανάλυσης διασποράς* για να ελέγξουμε, σε *επίπεδο σημαντικότητας 5%*, αν οι *μέσες τιμές*, μ_1, μ_2, μ_3, μ_4, αντίστοιχα, των τεσσάρων πληθυσμών είναι ίσες ή όχι. Δηλαδή, θα ελέγξουμε αν τα συγκεκριμένα τέσσερα *ανεξάρτητα τυχαία* δείγματα υποστηρίζουν ή όχι, ότι η *αύξηση του βάρους νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους,* επηρεάζεται από το *είδος πρόσθετης ύλης* (αν διαφοροποιείται από είδος σε είδος πρόσθετης ύλης).

Συγκεκριμένα, σε *επίπεδο σημαντικότητας 5%,* θα κάνουμε τον έλεγχο της *μηδενικής υπόθεσης,*

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 ,$$

έναντι της *εναλλακτικής*,

$$H_1 : \mu_i \neq \mu_j , \text{ για ένα τουλάχιστον ζευγάρι } (i, j), \; i, j = 1,2,3,4 .$$

Τα τέσσερα δείγματα έχουν ίδιο μέγεθος, 5, δηλαδή, $n_1 = n_2 = n_3 = n_4 = 5$ και επομένως συνολικά έχουμε στη διάθεσή μας, $\nu = n_1 + n_2 + n_3 + n_4 = 20$ παρατηρήσεις. Για να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*, κατ' αρχάς, θα υπολογίσουμε τα *αθροίσματα τετραγώνων* $SSTr$, SSE και $SSTot$.

A1	A2	A3	A4
81	78	72	85
66	66	70	70
78	69	78	83
76	64	77	74
61	66	69	70
$T_1 = 362$	$T_2 = 343$	$T_3 = 366$	$T_4 = 382$

Το *γενικό άθροισμα* των $\nu = 20$ παρατηρήσεων είναι, $G = \sum_{ji} y_{ij} = \sum_{j=1}^{4} T_j = 1453$ και επομένως:

$$SSTot = \sum_{ji} y_{ij}^2 - \frac{G^2}{\nu} = (81^2 + 66^2 + ... + 70^2) - \frac{1453^2}{20} = 838.55 , \qquad \text{με} \qquad \nu - 1 = 19$$

βαθμούς ελευθερίας.

$$SSTr = \sum_{j=1}^{4} \frac{T_j^2}{n_j} - \frac{G^2}{\nu} = \frac{362^2 + 343^2 + 366^2 + 382^2}{5} - \frac{1453^2}{20} = 154.15 , \qquad \text{με} \qquad k - 1 = 3$$

βαθμούς ελευθερίας.

$SSE = SSTot - SSTr = 838.55 - 154.15 = 684.40$, με $\nu - k = 16$ *βαθμούς ελευθερίας.*

Μπορούμε πλέον να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*:

Πίνακας Ανάλυσης Διασποράς				
Πηγή μεταβλητότητας	**Β.Ε.**	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***
Παράγοντας ***(ή επεμβάσεις)*** *(είδος πρόσθετης ύλης)*	3	154.15	$MSTr = \frac{154.15}{3} = 51.38$	$F_{Tr} = \frac{51.38}{42.77} = 1.20$
Σφάλμα	16	684.40	$MSE = \frac{684.40}{16} = 42.77$	
Ολική	19	838.55		

Επειδή η τιμή $F_{Tr} = 1.20$ δε βρίσκεται στην *περιοχή απόρριψης* της *μηδενικής υπόθεσης* αφού, $F_{3;16;0.05} = 3.24$, συμπεραίνουμε ότι, σε *επίπεδο σημαντικότητας 5%,* τα πειραματικά δεδομένα **δεν υποστηρίζουν** ότι το *είδος πρόσθετης ύλης*, επηρεάζει την *αύξηση του βάρους νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους.*

Για προβληματισμό: γιατί άραγε, ο επιβλέπων καθηγητής, μετά από αυτό το συμπέρασμα, πρότεινε στο φοιτητή (δες Πρόβλημα-2) να εκτελέσει ένα ακόμη πείραμα, και μάλιστα, με διαφορετικό πειραματικό σχέδιο;

Ας δούμε ένα ακόμη παράδειγμα.

Παράδειγμα-2: *Ένας μεταπτυχιακός φοιτητής του Τμήματος Γεωπονικής Βιοτεχνολογίας του Γ.Π.Α., στο πλαίσιο μιας εργαστηριακής άσκησης, διερεύνησε την περιεκτικότητα σε σωματιδιακές προσμείξεις ενός υγρού φαρμακευτικού σκευάσματος που χορηγείται ενδοφλεβίως σε ταύρους και παράγεται από τρεις διαφορετικές εταιρείες Α1, Α2 και Α3. Συγκεκριμένα, έλεγξε 6 σκευάσματα από κάθε εταιρεία (τυχαία επιλεγμένα) και κατέγραψε τον αριθμό σωματιδίων με διάμετρο μεγαλύτερη των 5 microns ανά λίτρο. Οι μετρήσεις που πήρε δίνονται στον πίνακα που ακολουθεί.*

		Αριθμός σωματιδίων *(ανά λίτρο)*					
Εταιρεία παραγωγής	Α1	255	264	342	331	234	217
	Α2	105	288	98	275	221	240
	Α3	577	515	214	413	401	260

Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των σκευασμάτων των τριών εταιρειών ως προς τη μέση περιεκτικότητά τους σε σωματιδιακές προσμείξεις;

Ο φοιτητής έχει στη διάθεσή του 18 παρατηρήσεις, που συγκέντρωσε παίρνοντας $k = 3$, *ανεξάρτητα τυχαία δείγματα.* Με τις παραδοχές ότι καθένα από αυτά προέρχεται από *κανονικό* πληθυσμό με *διασπορά*, σ^2, κοινή και για τους τρεις πληθυσμούς, εφάρμοσε τη μέθοδο *ανάλυσης διασποράς* για να ελέγξει, σε *επίπεδο σημαντικότητας 5%*, αν οι *μέσες τιμές*, μ_1, μ_2, μ_3, αντίστοιχα, των τριών πληθυσμών είναι ίσες ή όχι. Δηλαδή, για να ελέγξει τη *μηδενική υπόθεση*,

$$H_0 : \mu_1 = \mu_2 = \mu_3 ,$$

έναντι της *εναλλακτικής*,

$$H_1 : \mu_i \neq \mu_j , \text{ για ένα τουλάχιστον ζευγάρι } (i, j),\ i, j = 1,2,3 .$$

Ο φοιτητής χρησιμοποίησε, για τους αναγκαίους υπολογισμούς, το *στατιστικό πακέτο Statgraphics* από το οποίο πήρε τον ακόλουθο *πίνακα ανάλυσης διασποράς*.

```
            ANOVA Table for αριθμός-σωματιδίων by εταιρεία-παραγωγής
                              Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Between groups           113646.0      2      56823.2       5.81       0.0136
Within groups            146754.0     15      9783.58
-----------------------------------------------------------------------------
Total                    260400.0     17
```

Επειδή $P - Value = 0.0136 < 0.05$ (ή επειδή $F_{Tr} = 5.81 > F_{2;15;0.05} = 3.68$), η *μηδενική υπόθεση*, σε *επίπεδο σημαντικότητας 5%* απορρίπτεται. Δηλαδή, τα δεδομένα από τα τρία συγκεκριμένα *ανεξάρτητα* και *τυχαία* δείγματα, δίνουν στατιστικά σημαντικές αποδείξεις ότι ο μέσος αριθμός σωματιδιακών προσμείξεων δεν είναι ίδιος στα τρία φαρμακευτικά σκευάσματα. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 0.05 (μάλιστα, επειδή το *στατιστικό πακέτο* υπολόγισε την *P-Value*, μπορούμε να πούμε ότι αυτή η πιθανότητα είναι το πολύ 0.0136).

Βέβαια, το επόμενο (εύλογο) ερώτημα είναι: *εφόσον η μηδενική υπόθεση απορρίπτεται, τι μπορούμε να πούμε για τις διαφορές των τριών μέσων; Διαφέρουν όλοι μεταξύ τους ή μόνο ένας (και ποιος είναι αυτός);*

Φυσικά, αυτό το ερώτημα, προκύπτει κάθε φορά που σε έναν έλεγχο *ανάλυσης διασποράς*, απορρίπτεται η *μηδενική υπόθεση*. Η απάντηση δεν είναι απλή, και ουσιαστικά, το ερώτημα αυτό, θέτει και πάλι το ***πρόβλημα πολλαπλών συγκρίσεων***.

Στη συνέχεια, θα δούμε κάποιες από τις απαντήσεις που δίνει σε αυτό το πρόβλημα η *στατιστική συμπερασματολογία*.

***(1- α)100% Διαστήματα Εμπιστοσύνης* για <u>*τη μέση τιμή*</u> ενός πληθυσμού (ή αλλιώς, μιας *επέμβασης*) και για <u>*τη διαφορά των μέσων τιμών*</u> δύο πληθυσμών (ή αλλιώς, *δύο επεμβάσεων*) στο <u>*εντελώς τυχαιοποιημένο σχέδιο*</u>**

Ο φοιτητής (του προηγούμενου παραδείγματος), πιστεύει/υποψιάζεται, ότι από τα τρία φαρμακευτικά σκευάσματα, μεγαλύτερο αριθμό σωματιδιακών προσμείξεων περιέχει το σκεύασμα της εταιρείας Α3 και επίσης, ότι τα σκευάσματα των εταιρειών Α1 και Α2 μάλλον δε διαφέρουν (ως προς τον αριθμό σωματιδιακών προσμείξεων).

Σε περιπτώσεις που υπάρχει ειδικό/προσχεδιασμένο ενδιαφέρον για κάποια συγκεκριμένη, από τις $k > 2$, *μέσες τιμές* ή για κάποια, από τις $\binom{k}{2}$, διαφορές $\mu_i - \mu_j$, μια λύση είναι να κατασκευάσουμε ένα $(1-\alpha)\cdot 100\%$ *διάστημα εμπιστοσύνης*, για τη συγκεκριμένη *μέση τιμή* που μας ενδιαφέρει και ένα $(1-\alpha)\cdot 100\%$ *διάστημα εμπιστοσύνης*, για τη συγκεκριμένη *διαφορά* που μας ενδιαφέρει, χρησιμοποιώντας την *κατανομή t*. Μάλιστα, το απαιτούμενο για την κατασκευή των διαστημάτων *s*, μπορούμε να το υπολογίζουμε ως την τετραγωνική ρίζα του *MSE* αφού το *MSE* είναι *αμερόληπτη εκτιμήτρια* της άγνωστης *διασποράς* σ^2 και μάλιστα υπολογίζεται από όλες τις παρατηρήσεις. Βέβαια, στην περίπτωση αυτή, για την *κριτική τιμή* χρησιμοποιούμε $\nu - k$ *βαθμούς ελευθερίας*.

(1- α)100% διάστημα εμπιστοσύνης για ένα μέσο μ_i

$$\bar{y}_j \pm \frac{s}{\sqrt{n_j}} \cdot t_{\nu-k;\alpha/2}$$

(1- α)100% διάστημα εμπιστοσύνης για τη διαφορά $\mu_i - \mu_j$ *δύο μέσων*

$$(\bar{y}_i - \bar{y}_j) \pm s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \cdot t_{\nu-k;\alpha/2}$$

Όπου, $s = \sqrt{MSE}$ και $\nu = n_1 + n_2 + ... + n_k$.

Έτσι, στο παράδειγμά μας, έχουμε $s = \sqrt{MSE} = \sqrt{9783.58} = 98.91$, $\bar{y}_1 = 273.83$, $\bar{y}_2 = 204.5$, $\bar{y}_3 = 396.67$ και επομένως, ένα *95% διάστημα εμπιστοσύνης* για το *μέσο αριθμό*, μ_3, *σωματιδιακών προσμείξεων στο φαρμακευτικό σκεύασμα της εταιρείας Α3*, είναι,

$$\bar{y}_3 \pm \frac{s}{\sqrt{n_3}} \cdot t_{\nu-k;\alpha/2} \text{ ή } 396.67 \pm \frac{98.91}{\sqrt{6}} \cdot t_{15;0.025} \text{ ή } 396.67 \pm \frac{98.91}{\sqrt{6}} \cdot 2.131 \text{ ή } 396.67 \pm 86.03$$

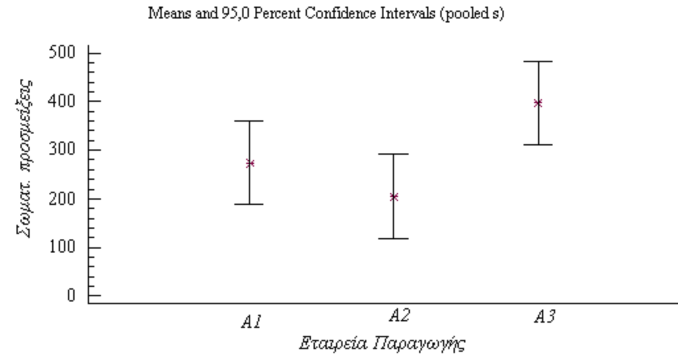
και ένα *95% διάστημα εμπιστοσύνης* για τη *διαφορά* $\mu_2 - \mu_1$ *των μέσων των σωματιδιακών προσμείξεων των σκευασμάτων των εταιρειών Α1 και Α2* είναι,

$$(\bar{y}_2 - \bar{y}_1) \pm s \cdot \sqrt{\frac{1}{n_2} + \frac{1}{n_1}} \cdot t_{\nu-k;\alpha/2} \quad \text{ή} \quad (204.5 - 273.83) \pm 98.91 \cdot \sqrt{\frac{2}{6}} \cdot t_{15;0.025} \quad \text{ή}$$

$$-69.33 \pm 121.69 \text{ ή } [-191.02, \quad 52.36].$$

Παρατηρούμε ότι το *95% διάστημα εμπιστοσύνης*, $[-191.02, \quad 52.36]$, για τη διαφορά $\mu_2 - \mu_1$, περιέχει το 0 και επομένως, σε *επίπεδο σημαντικότητας 5%*, τα δεδομένα από τα δείγματα δεν δείχνουν στατιστικά σημαντική *διαφορά των μέσων* των σωματιδιακών προσμείξεων των σκευασμάτων των εταιρειών Α2 και Α1, δηλαδή, φαίνεται ότι ήταν βάσιμη η υποψία του ερευνητή.

Στο *Σχήμα-2* φαίνονται οι *μέσοι* των τριών δειγμάτων και τα αντίστοιχα *95% διαστήματα εμπιστοσύνης* που κατασκευάσθηκαν όπως περιγράψαμε προηγουμένως. Το διάγραμμα αυτό έγινε με το *στατιστικό πακέτο Statgraphics*.



Σχήμα-2

Παρατηρείστε, ότι υπάρχει επικάλυψη των διαστημάτων που αντιστοιχούν στα σκευάσματα των εταιρειών Α1 και Α2 αλλά και μεταξύ των Α1 και Α3 ενώ δεν υπάρχει μεταξύ των Α2 και Α3. Όμως, αν ο φοιτητής δεν έχει ***προσχεδιάσει***, δεν έχει εκ των προτέρων ***(a priori)*** ειδικό ενδιαφέρον, για κάποια διαφορά (π.χ, για την $\mu_2 - \mu_1$ όπως είδαμε) και ***εκ των υστέρων (a posteriori)*** θέλει να διερευνήσει τη σημαντικότητα όλων των $c = \binom{3}{2}$ διαφορών, δηλαδή, και των διαφορών $\mu_2 - \mu_3$ και $\mu_3 - \mu_1$, τότε απαιτείται άλλη προσέγγιση γιατί, όπως έχουμε εξηγήσει, αυξάνεται ραγδαία η πιθανότητα να αναδείξει τουλάχιστον μια από τις *c* διαφορές στατιστικά σημαντική ενώ δεν είναι *(δηλαδή, η πιθανότητα να συμβεί τουλάχιστον μια φορά σφάλμα τύπου I ή αλλιώς, η πιθανότητα να συμβεί κατά πείραμα σφάλμα τύπου I)*.

Ας δούμε πώς μπορούμε να κάνουμε πολλαπλές συγκρίσεις χωρίς να μας «ξεφεύγει» η *πιθανότητα να κάνουμε σφάλμα τύπου I σε τουλάχιστον μια από αυτές,* δηλαδή, η *πιθανότητα σφάλματος τύπου I* <u>*κατά πείραμα*</u> που συμβολίσαμε με α_{PE}.

Έλεγχοι πολλαπλών συγκρίσεων (multiple comparisons tests)

Η πιο απλή (γιατί δεν υπάρχει μόνο μια) εκδοχή του *προβλήματος των πολλαπλών συγκρίσεων*, είναι αυτή στην οποία ήδη έχουμε αναφερθεί, δηλαδή, ο έλεγχος της $H_0 : \mu_i = \mu_j$, έναντι της *εναλλακτικής* $H_1 : \mu_i \neq \mu_j$, <u>για όλα</u> τα *i*, *j* με $i \neq j$.

Από τους πιο γνωστούς *ελέγχους πολλαπλών συγκρίσεων* αυτής της μορφής, είναι η διαδικασία/μέθοδος ***Tukey** (1953)* ***(Tukey's method for paired comparisons)***, γνωστή και ως ***HSD test (Honestly Significant Difference test)***. Στη διαδικασία *Tukey,* όχι απλώς δεν μας «ξεφεύγει» το <u>*κατά πείραμα*</u> *σφάλμα τύπου I*, δηλαδή, *η πιθανότητα να κάνουμε σφάλμα τύπου I σε τουλάχιστον έναν από τους* $c = \binom{k}{2}$ *ελέγχους*, αλλά το θέτουμε υπό τον έλεγχό μας, προκαθορίζοντάς το!! Δηλαδή, η διαδικασία *Tukey*, επιτυγχάνει το εξής: η πιθανότητα, μια τουλάχιστον διαφορά $\mu_i - \mu_j$ να αναδειχθεί, ενώ δεν είναι, στατιστικά σημαντική, να είναι ίση με $\alpha\%$, που εμείς καθορίσαμε και επομένως προσφέρει μεγάλη προστασία από *σφάλματα τύπου Ι* αλλά γι' αυτό θεωρείται (και είναι) συντηρητικός έλεγχος στην ανάδειξη στατιστικά σημαντικών διαφορών.

Η διατήρηση της πιθανότητας *σφάλματος τύπου Ι κατά πείραμα*, σε προκαθορισμένο επίπεδο, επιτυγχάνεται με έναν απλό και έξυπνο τρόπο. Κατασκευάζουμε για όλες τις διαφορές $\mu_i - \mu_j$, $(1-\alpha)\%$ *ταυτόχρονα διαστήματα εμπιστοσύνης* χρησιμοποιώντας την κατανομή *studentized range* $Q_{k;nk-k}$ και οι έλεγχοι της σημαντικότητάς τους γίνονται με αυτά τα διαστήματα. (Η *studentized range* κατανομή $Q_{k;\upsilon}$ είναι η κατανομή της τυχαίας μεταβλητής, $\frac{R}{S}$ όπου, R είναι το *εύρος k κανονικών* τυχαίων μεταβλητών και S μια *εκτιμήτρια* της *τυπικής απόκλισής* τους με υ βαθμούς ελευθερίας.) Αποδεικνύεται (δεν είναι δύσκολο) ότι, αν $\overline{Y}_j$, $j = 1,2,...,k$ είναι οι k *δειγματικοί μέσοι* σε ένα *εντελώς τυχαιοποιημένο σχέδιο* με *k στάθμες,* όπου για κάθε *στάθμη* έχουμε ίδιο αριθμό παρατηρήσεων $n = n_1 = n_2 = ... = n_k$, τότε, τα $\binom{k}{2}$ *ταυτόχρονα* $(1-\alpha)\%$ *διαστήματα εμπιστοσύνης για τις διαφορές* $\mu_i - \mu_j$ είναι,

$$\left(\overline{Y}_i - \overline{Y}_j - Q_{k;nk-k;\alpha}\sqrt{\frac{MSE}{n}},\ \ \overline{Y}_i - \overline{Y}_j + Q_{k;nk-k;\alpha}\sqrt{\frac{MSE}{n}} \right),$$

όπου, $Q_{k;nk-k;\alpha}$ το άνω $\alpha\%$ *ποσοστιαίο σημείο* της $Q_{k;nk-k}$ κατανομής (υπάρχουν σχετικοί πίνακες).

Ας δούμε πώς, κατασκευάζοντας *95% διαστήματα εμπιστοσύνης Tukey,* μπορούμε να συγκρίνουμε, ανά δύο, τους *μέσους,* μ_1, μ_2, μ_3, του *Παραδείγματος-2.*

Παράδειγμα-2 (συνέχεια): Έχουμε, $\bar{y}_1 = 273.83$, $\bar{y}_2 = 204.5$, $\bar{y}_3 = 396.67$, $k = 3$, $n = 6$ και άρα, $Q_{k;nk-k;\alpha}\sqrt{\frac{MSE}{n}} = Q_{3;15;0.05}\sqrt{\frac{9783.58}{6}} = 3.67 \cdot 40.38 = 148.2$.

Έτσι, έχουμε:

$\mu_i - \mu_j$	$\overline{Y}_i - \overline{Y}_j$	$Q_{k;nk-k;\alpha}\sqrt{MSE/n}$ *(όρια +/-)*	*95% Δ.Ε. Tukey*	*Συμπέρασμα για τον έλεγχο:* $H_0 : \mu_i = \mu_j$ $H_1 : \mu_i \neq \mu_j$
$\mu_1 - \mu_2$	69.33	148.2	$(-78.9,\ 217.5)$	*Δεν απορρίπτεται η* H_0
$\mu_1 - \mu_3$	-122.84	148.2	$(-\ 271.0,\ 25.4)$	*Δεν απορρίπτεται η* H_0
$\mu_2 - \mu_3$	-192.17	148.2	$(-340.4,\ -44.0)$	*Απορρίπτεται η* H_0

Επομένως, στατιστικά σημαντική αναδείχθηκε μόνο η διαφορά $\mu_2 - \mu_3$. *Η πιθανότητα να κάνουμε ένα τουλάχιστον σφάλμα τύπου Ι στους τρεις ελέγχους,* είναι το πολύ 5%.

Στον πίνακα που ακολουθεί φαίνονται τα αποτελέσματα που παίρνουμε από το *στατιστικό πακέτο Statgraphics*, για τη *διαδικασία Tukey* που μόλις εφαρμόσαμε.

Παρατηρείστε τη στήλη με τους *δειγματικούς μέσους*. Έχουν γραφεί σε αύξουσα σειρά και στην πρώτη στήλη έχουν γραφεί οι αντίστοιχες *στάθμες* (A2, A1, A3) του παράγοντα *(εταιρεία παραγωγής)*. Παρατηρείστε, επίσης, τη στήλη με τα σύμβολα «X». Αυτά, «διαβάζονται» ως εξής: Στους *μέσους* που αντιστοιχούν σε «X» που βρίσκονται στην ίδια στήλη, δεν διαπιστώθηκε στατιστικά σημαντική διαφορά.

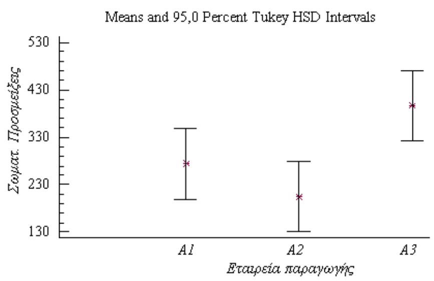
```
          Multiple Range Tests for αριθμός_σωματιδίων by εταιρεία-παραγωγής
--------------------------------------------------------------------------------
Method: 95,0 percent Tukey HSD
Εταιρεία-παραγωγής      Count      Mean        Homogeneous Groups
--------------------------------------------------------------------------------
2                        6       204.5           X
1                        6       273.833         XX
3                        6       396.667          X
--------------------------------------------------------------------------------
Contrast                                  Difference          +/-  Limits
--------------------------------------------------------------------------------
1 - 2                                       69.3333           148.963
1 - 3                                     -122.833            148.963
2 - 3                                    *-192.167            148.963
--------------------------------------------------------------------------------
*denotes a statistically significant difference.
```

Η *διαδικασία Tukey* μπορεί να εφαρμοσθεί και ως εξής: Διατάσσουμε τους τρεις *δειγματικούς μέσους* σε αύξουσα σειρά, $\bar{y}_2 = 204.5$, $\bar{y}_1 = 273.83$, $\bar{y}_3 = 396.67$ και κάνουμε όλους τους ανά δύο ελέγχους με *απορριπτική περιοχή,* $\dfrac{\left|\bar{Y}_i - \bar{Y}_j\right|}{\sqrt{MSE/n}} > Q_{k;nk-k;\alpha}$

ή $\left|\bar{Y}_i - \bar{Y}_j\right| > Q_{k;nk-k;\alpha}\sqrt{MSE/n}$, και αρχίζοντας με τη σύγκριση των *μέσων* που οι αντίστοιχοι *δειγματικοί μέσοι* είναι ο μεγαλύτερος και ο μικρότερος, συνεχίζουμε με την ίδια λογική. Αν η *μηδενική υπόθεση* δεν απορριφθεί για ένα ζευγάρι μέσων, τότε, οι έλεγχοι για τους *μέσους* που οι αντίστοιχοι *δειγματικοί* βρίσκονται εντός του εύρους των *δειγματικών μέσων* των δύο συγκρινόμενων *μέσων*, δε χρειάζεται να γίνουν. Δηλαδή, αν στο παράδειγμά μας, δεν απορριφθεί η $H_0: \mu_3 = \mu_2$, τότε δε χρειάζεται να γίνει και ο έλεγχος της $H_0: \mu_3 = \mu_1$ (γιατί;). Στον πίνακα που ακολουθεί φαίνεται αυτός ο τρόπος εφαρμογής της *διαδικασίας Tukey*.

Σύγκριση μ_i *vs* μ_j	$\left\lvert\bar{Y}_i - \bar{Y}_j\right\rvert$	$\sqrt{MSE/n}$	$q = \dfrac{\left\lvert\bar{Y}_i - \bar{Y}_j\right\rvert}{\sqrt{MSE/n}}$	$Q_{3;15;0.05}$	*Συμπέρασμα για τον έλεγχο:* $H_0: \mu_i = \mu_j$ $H_1: \mu_i \neq \mu_j$
μ_3 *vs* μ_2	192.17	40.38	4.76	3.67	*Απορρίπτεται η* H_0
μ_3 *vs* μ_1	122.84	40.38	3.04	3.67	*Δεν απορρίπτεται η* H_0
μ_1 *vs* μ_2	69.33	40.38	1.72	3.67	*Δεν απορρίπτεται η* H_0

Στο *Σχήμα-3* που ακολουθεί φαίνεται το διάγραμμα *μέσων* με τα αντίστοιχα *95% διαστήματα εμπιστοσύνης Tukey*.



Σχήμα-3

Παρατηρήσεις:

1. *Αρκετά συχνά, όπως συμβαίνει και στο προηγούμενο παράδειγμα, τα αποτελέσματα ενός ελέγχου πολλαπλών συγκρίσεων, μπορεί να φαίνονται διφορούμενα/ασαφή. Παρατηρώντας τη στήλη με τα σύμβολα «Χ», φαίνεται το δείγμα2 και το δείγμα1 να προέρχονται από έναν πληθυσμό και το δείγμα1 και το δείγμα3 από έναν άλλο πληθυσμό. Και, φυσικά, γεννάται η απορία: τελικά από ποιον από τους δύο αυτούς πληθυσμούς προέρχεται το δείγμα1; Η απάντηση είναι εξής: προφανώς η μέθοδος δε μπόρεσε να ταξινομήσει με σαφήνεια το δείγμα1 (συνέβη τουλάχιστον ένα σφάλμα τύπου ΙΙ). Επομένως, στο παράδειγμά μας, το συμπέρασμα από τις τρεις ανά δύο συγκρίσεις με τη διαδικασία Tukey, είναι ότι* $\mu_2 \neq \mu_3$ *και όχι* $\mu_1 = \mu_2 \neq \mu_3 = \mu_1$. *Μια λύση στο πρόβλημα αυτό μπορεί να είναι η επανάληψη της δειγματοληψίας με μεγαλύτερα μεγέθη δειγμάτων, για να αυξήσουμε την ισχύ του ελέγχου ή ισοδύναμα, για να μειώσουμε την πιθανότητα να συμβεί σφάλμα τύπου ΙΙ. Μπορούμε, βέβαια, να κάνουμε κάποιον άλλο, πιο ισχυρό, έλεγχο (αλλά, φυσικά, με το ανάλογο κόστος στην πιθανότητα σφάλματος τύπου Ι).*
2. *Αν τα k δείγματα δεν είναι του ιδίου μεγέθους, η μέθοδος Tukey μπορεί να εφαρμοσθεί με κατάλληλη προσαρμογή στον υπολογισμό των ορίων των διαστημάτων. Αντί* $\sqrt{\frac{MSE}{n}}$ *χρησιμοποιούμε* $\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_2}\right)}$.

Μια άλλη, πολύ γνωστή επίσης, μέθοδος *πολλαπλών συγκρίσεων* είναι η διαδικασία/μέθοδος ***Newman-Keuls*** *(1939; 1952)* γνωστή και ως ***Student-Newman-Keuls/SNK***. Είναι ανάλογη με τη διαδικασία *Tukey* με μία διαφορά. *Η κριτική τιμή* είναι $Q_{p;nk-k;\alpha}$, όπου *p,* ο αριθμός των *μέσων* που βρίσκονται στο εύρος που ορίζουν οι συγκρινόμενοι κάθε φορά *μέσοι*. Δηλαδή η κριτική τιμή, $Q_{p;nk-k;\alpha}$, δεν είναι ίδια για όλους τους ανά δύο ελέγχους.

Ας εφαρμόσουμε τη μέθοδο *SNK* για τα δεδομένα του *Παραδείγματος-2.*

<u>Παράδειγμα-2 (συνέχεια):</u> Διατάσσουμε τους τρεις *δειγματικούς μέσους* σε αύξουσα σειρά: $\bar{y}_2 = 204.5$, $\bar{y}_1 = 273.83$, $\bar{y}_3 = 396.67$

Σύγκριση μ_i *vs* μ_j	$\lvert\bar{Y}_i - \bar{Y}_j\rvert$	$\sqrt{MSE/n}$	$q = \frac{\lvert\bar{Y}_i - \bar{Y}_j\rvert}{\sqrt{MSE/n}}$	p	$Q_{p;15;0.05}$	*Συμπέρασμα για τον έλεγχο:* $H_0: \mu_i = \mu_j$ $H_1: \mu_i \neq \mu_j$
μ_3 *vs* μ_2	192.17	40.38	4.76	3	3.67	*Απορρίπτεται η* H_0
μ_3 *vs* μ_1	122.84	40.38	3.04	2	3.01	*Απορρίπτεται η* H_0
μ_1 *vs* μ_2	69.33	40.38	1.72	2	3.01	*Δεν απορρίπτεται η* H_0

Στατιστικά σημαντικές, με τη *διαδικασία SNK,* βλέπουμε ότι αναδείχθηκαν δύο διαφορές, η $\mu_3 - \mu_2$ και η $\mu_3 - \mu_1$, ενώ με τη *διαδικασία Tukey*, όπως είδαμε, στατιστικά σημαντική αναδείχθηκε μόνο η διαφορά $\mu_3 - \mu_2$. Γενικά, η *διαδικασία SNK* είναι ***πιο ισχυρή*** (πιο ευαίσθητη, ή αλλιώς, λιγότερο συντηρητική, στην ανάδειξη διαφορών) από τη *διαδικασία Tukey,* δηλαδή, τείνει να αναδεικνύει, ως στατιστικά σημαντικές διαφορές, περισσότερες από αυτές που αναδεικνύει η *διαδικασία Tukey*. Όμως, η πιθανότητα α_{PE}, να συμβεί τουλάχιστον *ένα σφάλμα τύπου Ι,* μπορεί να ξεπεράσει το επίπεδο που ορίζουμε (στο παράδειγμά μας, το 5%).

Στον πίνακα που ακολουθεί φαίνονται τα αποτελέσματα που παίρνουμε από το *στατιστικό πακέτο Statgraphics*, για τη *διαδικασία SNK* που μόλις εφαρμόσαμε.

```
          Multiple Range Tests for αριθμός_σωματιδίων by εταιρεία-παραγωγής
--------------------------------------------------------------------------------
Method: 95,0 percent Student-Newman-Keuls
Εταιρεία-παραγωγής     Count       Mean        Homogeneous Groups
--------------------------------------------------------------------------------
2                        6        204.5            X
1                        6        273.833          X
3                        6        396.667           X
--------------------------------------------------------------------------------
Contrast                                    Difference
--------------------------------------------------------------------------------
1 - 2                                         69.3333
1 - 3                                      *-122.833
2 - 3                                      *-192.167
--------------------------------------------------------------------------------
* denotes a statistically significant difference.
```

Παρατηρείστε τη στήλη με τα σύμβολα «Χ». Χρησιμοποιώντας έναν πιο ισχυρό έλεγχο, το πρόβλημα με την ασάφεια στην ταξινόμηση του *δείγματος1* δεν υφίσταται πλέον (μπορούμε δηλαδή να πούμε ότι το συμπέρασμα είναι, $\mu_1 = \mu_2 \neq \mu_3$, αλλά (επειδή όλα στη ζωή έχουν κόστος) μπορεί η πιθανότητα, στις τρεις συγκρίσεις, να κάνουμε *τουλάχιστον ένα σφάλμα τύπου I*, να έχει αυξηθεί.

Στη συνέχεια, θα αναφερθούμε σε βασικά μόνο χαρακτηριστικά και στη λογική πέντε ακόμη *ελέγχων πολλαπλών συγκρίσεων* χωρίς να αναφερθούμε λεπτομερώς σε αυτούς (η εκτέλεσή τους είναι πλέον εύκολη με χρήση *στατιστικών πακέτων*). Συγκεκριμένα, θα αναφερθούμε στους ελέγχους/μεθόδους, ***Scheffe, Dunn, Dunnett, Duncan*** και ***LSD***.

Ο έλεγχος ***Scheffe*** *ή* ***S-Method***, διατηρεί την *πιθανότητα σφάλματος τύπου I <u>κατά πείραμα</u>*, σε προκαθορισμένο επίπεδο, όπως η *διαδικασία Tukey* και επομένως μας προσφέρει μεγάλη προστασία από *σφάλματα τύπου I.* Όμως, έχει σχεδιασθεί όχι για *πολλαπλές συγκρίσεις <u>ανά δύο</u>* αλλά για άλλου τύπου συγκρίσεις ***(γραμμικές αντιθέσεις/linear contrasts)***, στις οποίες δε θα αναφερθούμε. Αναφέρουμε μόνο, ότι σε ένα πείραμα, για παράδειγμα με $k = 5$ *επεμβάσεις*, έλεγχος *γραμμικής αντίθεσης* είναι, για παράδειγμα, ο έλεγχος της *μηδενικής υπόθεσης*, $H_0: \mu_1 - \frac{\mu_2 + \mu_3}{2} = 0$ ή της $H_0: \frac{\mu_1 + \mu_3}{2} - \frac{\mu_4 + \mu_5}{2} = 0$ ή της $H_0: \mu_1 - \mu_3 = 0$. Όπως φαίνεται και από το τελευταίο παράδειγμα, ο έλεγχος *Scheffe* μπορεί να εφαρμοσθεί και για όλες τις *<u>ανά δύο</u> συγκρίσεις,* όμως, στην περίπτωση αυτή, είναι πολύ συντηρητικός στην ανάδειξη στατιστικά σημαντικών διαφορών (περισσότερο και από τη *διαδικασία Tukey*).

Ο έλεγχος ***Dunn*** *(Bonferroni correction)*, χειρίζεται την *πιθανότητα σφάλματος τύπου I κατά πείραμα,* α_{PE}, ως εξής: την προκαθορίζει σε ένα ορισμένο επίπεδο και την κατανέμει (όχι κατ' ανάγκη εξίσου) στις ανά δύο συγκρίσεις που θα γίνουν *(διόρθωση Bonferroni, με βάση τη σχέση* $\alpha_{PE} \leq c \cdot \alpha_{PC}$*)*. Για παράδειγμα, αν αποφασίσουμε να κάνουμε 4 συγκρίσεις και ορίσουμε την *πιθανότητα σφάλματος τύπου I κατά πείραμα,* σε 5%, μπορούμε κάθε σύγκριση να την κάνουμε σε *επίπεδο σημαντικότητας* $\alpha_{PC} = 0.0125$ ή τις τρεις σε 0.01 την κάθε μια, και την τέταρτη σε 0.02, κ.τλ.. Είναι από τους πλέον συντηρητικούς ελέγχους και προσφέρεται για μικρό μόνο αριθμό συγκρίσεων.

Ο έλεγχος ***Dunnett** (1955)*, σχεδιάσθηκε και προσφέρεται για περιπτώσεις που μας ενδιαφέρει η σύγκριση ενός *μάρτυρα* με κάθε μια από ένα σύνολο *επεμβάσεων*. Δηλαδή, όταν δεν μας ενδιαφέρουν όλες οι δυνατές συγκρίσεις αλλά έχουμε **προσχεδιάσει** να κάνουμε μόνο τις συγκεκριμένες συγκρίσεις με το *μάρτυρα*. Και αυτός ο έλεγχος, λαμβάνει υπόψη το *σφάλμα τύπου Ι κατά πείραμα,* α_{PE}.

Ο έλεγχος ***Duncan** (1955)*, όπως και ο *SNK*, χειρίζεται το *σφάλμα τύπου Ι κατά πείραμα,* α_{PE}, λαμβάνοντας υπόψη τον αριθμό των μέσων που βρίσκονται μεταξύ των μέσων που κάθε φορά συγκρίνονται και γι' αυτό και έχει ανάλογη συμπεριφορά με τον έλεγχο *SNK*, δηλαδή, έχει την τάση να αναδεικνύει περισσότερες στατιστικά σημαντικές διαφορές από τη διαδικασία *Tukey*.

Τελευταίο αφήσαμε τον έλεγχο ***ελάχιστης σημαντικής διαφοράς LSD (Least Significant Difference)*** *(Fisher, 1935),* γιατί είναι ο μόνος ο οποίος δε λαμβάνει υπόψη το *σφάλμα τύπου Ι κατά πείραμα,* α_{PE}. Είναι ο πλέον ευαίσθητος στην ανάδειξη στατιστικά σημαντικών διαφορών, όμως παρέχει τη μικρότερη προστασία από *σφάλματα τύπου Ι*. Αν επιλέξουμε να τον εφαρμόσουμε, πρέπει αφενός οι συγκρίσεις να είναι λίγες, και αφετέρου, να έχει προηγηθεί έλεγχος *ανάλυσης διασποράς* της *μηδενικής υπόθεσης* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ και αυτή **να έχει απορριφθεί** *(έτσι μπορεί να δικαιολογηθεί και η χρησιμοποίησή του, παρότι δε λαμβάνει καμία μέριμνα για έλεγχο του σφάλματος τύπου Ι κατά πείραμα, με την έννοια, ότι αφού έχει προηγηθεί απόρριψη της* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$*, περιμένουμε να μην συμβούν πολλά σφάλματα τύπου Ι στις πολλαπλές συγκρίσεις).*

Παρατήρηση: *Παρότι, για την εφαρμογή των ελέγχων πολλαπλών συγκρίσεων δεν απαιτείται να έχει προηγηθεί έλεγχος ανάλυσης διασποράς της μηδενικής υπόθεσης* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ *και απόρριψή της, εντούτοις, συνηθίζεται να εφαρμόζονται μετά από έλεγχο ανάλυση διασποράς και απόρριψη της* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$*. Όμως, επαναλαμβάνουμε, ειδικά ο έλεγχος ελάχιστης σημαντικής διαφοράς (LSD), πρέπει να εφαρμόζεται μόνο* ***μετά από απόρριψη*** *της* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$.

Ολοκληρώνοντας αυτή τη σύντομη αναφορά στους *ελέγχους πολλαπλών συγκρίσεων*, απαιτείται να κάνουμε ένα σχόλιο σχετικά με το πώς, με ποια κριτήρια, αποφασίζουμε, ποιος *έλεγχος πολλαπλών συγκρίσεων* είναι ο καταλληλότερος για την στατιστική ανάλυση των δεδομένων μας.

Για να επιλέξουμε τον κατάλληλο έλεγχο πρέπει πρώτα να αποφασίσουμε αν θέλουμε να κάνουμε όλες τις ανά δύο συγκρίσεις ή μόνο κάποιες συγκεκριμένες. Στις περιπτώσεις που ενδιαφερόμαστε μόνο για κάποιες συγκεκριμένες συγκρίσεις πρέπει να επιλέξουμε μεταξύ των ελέγχων *Dunnett* και *Scheffe* (κατά περίπτωση). Πρέπει επίσης να σταθμίσουμε τη σημασία των *σφαλμάτων τύπου Ι* και *τύπου ΙΙ* στη συγκεκριμένη έρευνα. Δηλαδή, αν πρέπει να «προστατευθούμε» περισσότερο από *σφάλματα τύπου Ι* ή από *σφάλματα τύπου ΙΙ*, αν μας ενδιαφέρει ο έλεγχος της *πιθανότητας σφάλματος τύπου Ι κατά πείραμα,* α_{PE}, ή/και της *πιθανότητας σφάλματος τύπου Ι κατά σύγκριση,* α_{PC}. Φυσικά, η απάντηση σε αυτά τα ερωτήματα εξαρτάται από το συγκεκριμένο κάθε φορά ερευνητικό πρόβλημα αλλά και από την γενικότερη αντίληψη/άποψη του ερευνητή για τη σημασία και τη βαρύτητα των διαφόρων τύπων σφαλμάτων σε μια έρευνα. Αν αποφασίσουμε ότι το *κατά πείραμα σφάλμα τύπου Ι,* α_{PE}, μας ενδιαφέρει, τότε πρέπει να αποκλείσουμε από τις επιλογές μας τον έλεγχο *LSD* και να επιλέξουμε μεταξύ των ελέγχων, *Tukey, Scheffe, Duncan SNK* και *Dunn.*

Το βασικό στοιχείο που πρέπει, στο σημείο αυτό, να λάβουμε υπόψη μας είναι ο τρόπος που ο κάθε έλεγχος χειρίζεται την προστασία από το *σφάλμα τύπου Ι κατά πείραμα*, α_{PE}, γιατί, ουσιαστικά, με βάση αυτόν διαφοροποιούνται. Υπενθυμίζουμε ότι, ο έλεγχος *Scheffe,* όταν χρησιμοποιείται για όλες τις ανά δύο συγκρίσεις είναι πολύ συντηρητικός, δηλαδή, παρέχει τη μικρότερη προστασία από *σφάλματα τύπου ΙΙ* κάτι το οποίο συμβαίνει και με τον έλεγχο *Dunn* όταν ο αριθμός των συγκρίσεων δεν είναι μικρός. Από τους ελέγχους, *Tukey, Duncan* και *SNK* ο πλέον συντηρητικός είναι ο *Tukey* αλλά παρέχει μεγαλύτερη προστασία από *σφάλμα τύπου Ι κατά πείραμα,* α_{PE}, ενώ οι *Duncan* και *SNK* τείνουν να αναδεικνύουν περισσότερες σημαντικές διαφορές γιατί, όπως είδαμε, χειρίζονται το α_{PE}, με άλλον τρόπο («διαβαθμίζουν» την προστασία).

Ολοκληρώνοντας αυτό το σύντομο σχόλιο, σημειώνουμε ότι στη σχετική βιβλιογραφία, φυσικά, δεν υπάρχουν τυποποιημένα και απόλυτα κριτήρια ή αυστηρά οριοθετημένες περιπτώσεις για το ποιος ή ποιοι έλεγχοι είναι οι καταλληλότεροι. Η απόφαση ανήκει στον ερευνητή και πρέπει να είναι αιτιολογημένη. Απαραίτητη προϋπόθεση γι΄ αυτό, είναι να έχει κατανοήσει τη λογική και τα βασικά, τουλάχιστον, χαρακτηριστικά κάθε ελέγχου.

Υποθέσεις/παραδοχές στο εντελώς τυχαιοποιημένο σχέδιο

Οι υποθέσεις/παραδοχές που κάνουμε στο *εντελώς τυχαιοποιημένο σχέδιο (με καθορισμένες επιδράσεις)*, όπως ήδη έχουμε αναφέρει, είναι:

1. Τα *k* δείγματα είναι *ανεξάρτητα τυχαία* δείγματα (όλες οι παρατηρήσεις y_{ij} είναι μεταξύ τους *ανεξάρτητες*)
2. Κάθε δείγμα προέρχεται από *κανονικό* πληθυσμό
3. Οι *k* πληθυσμοί έχουν *κοινή διασπορά,* σ^2 *(ομοσκεδαστικότητα)*

Αν $\varepsilon_{ij} = Y_{ij} - \mu_j$, $j = 1,2,...,k$, $i = 1,2,...,n_j$ η απόκλιση της Y_{ij} από τη *μέση τιμή* μ_j του πληθυσμού *j*, τότε οι παραπάνω υποθέσεις είναι ισοδύναμες με τις εξής:

1. Τα ε_{ij} είναι μεταξύ τους *ανεξάρτητα*
2. $\varepsilon_{ij} \sim N(0,\sigma^2)$

Τα ε_{ij} ονομάζονται ***υπόλοιπα (residuals)*** και εκφράζουν το ***πειραματικό σφάλμα (experimental error)*** δηλαδή, τη μεταβλητότητα που δεν εξηγείται από τις *επεμβάσεις*.

Όταν τα μεγέθη, $n_1, n_2, ..., n_k$, των *k* δειγμάτων είναι ίσα, η *ανάλυση διασποράς* είναι *«ανθεκτική»* σε μικρές ή μέτριες αποκλίσεις από την υπόθεση τόσο της *κανονικότητας* των πληθυσμών όσο και της *ομοσκεδαστικότητας*. Δηλαδή, η *πιθανότητα σφάλματος τύπου Ι* είναι κοντά στο *επίπεδο σημαντικότητας* που ορίζουμε ακόμη και αν υπάρχουν μικρές ή μέτριες αποκλίσεις από τις δύο αυτές παραδοχές. Αν όμως, τα n_i δεν είναι ίσα, τότε η *πιθανότητα σφάλματος τύπου Ι* μπορεί να διαφέρει αρκετά από το *επίπεδο σημαντικότητας* που ορίζουμε, ανάλογα με το μέγεθος της *ετεροσκεδαστικότητας* (αν μεγάλες *διασπορές* αντιστοιχούν σε μικρά δείγματα, η *πιθανότητα σφάλματος τύπου Ι* μπορεί να είναι αρκετά μεγαλύτερη από το *επίπεδο σημαντικότητας* που ορίζουμε). Επίσης, αν οι κατανομές των πληθυσμών είναι *πλατύκυρτες* ή *λεπτόκυρτες* και τα δείγματα μικρά, η πραγματική *ισχύς* (και επομένως η *πιθανότητα σφάλματος τύπου ΙΙ*) του ελέγχου επηρεάζεται. Βέβαια, αν τα δείγματα δεν είναι πολύ μικρά, μικρές αποκλίσεις από την κανονικότητα, δε δημιουργούν πρόβλημα εγκυρότητας του *ελέγχου ανάλυσης διασποράς*. Σε κάθε περίπτωση, όμως, πρέπει να ελέγχουμε τη βασιμότητα των απαιτούμενων παραδοχών/υποθέσεων.

Για τον έλεγχο της *κανονικότητας* των πληθυσμών από τους οποίους προέρχονται τα δείγματα είτε της *κανονικότητας* των *υπολοίπων*, αλλά και για τον έλεγχο της *ισότητας* των *διασπορών* και της *ανεξαρτησίας* των *υπολοίπων*, η *Στατιστική* μας προσφέρει μια μεγάλη ποικιλία μεθόδων και τεχνικών οι οποίες πλέον, με χρήση κάποιου *στατιστικού πακέτου*, μπορούν πολύ εύκολα να εφαρμοσθούν και να αξιοποιηθούν.

Για τον έλεγχο της *κανονικότητας* μπορούμε να κάνουμε *ελέγχους καλής προσαρμογής (έλεγχος Kolmogorov-Smirnov, έλεγχος Anderson-Darling, έλεγχος* X^2*, κ.ά.)* Επίσης, για τον έλεγχο *ισότητας των διασπορών* μπορούμε να κάνουμε *έλεγχο Bartlett, έλεγχο Cochran, έλεγχο Hartley, κ.ά.*. Δεν θα αναφερθούμε αναλυτικά σε καθέναν από αυτούς τους ελέγχους. Για τον *έλεγχο Bartlett*, επειδή είναι από τους πλέον γνωστούς και χρησιμοποιείται πολύ συχνά, επισημαίνουμε μόνο ότι αν δεν ισχύει η υπόθεση της *κανονικότητας,* επηρεάζεται πολύ και μπορεί να δώσει μη έγκυρα αποτελέσματα. Τέλος, για το έλεγχο της *ανεξαρτησίας των σφαλμάτων* μπορούμε να κάνουμε κάποιο *έλεγχο τυχαιότητας (randomness test)* όπως, *έλεγχο ροών (runs test)*.

Εικόνα για τη βασιμότητα των παραπάνω παραδοχών/υποθέσεων παίρνουμε επίσης κατασκευάζοντας κατάλληλα *ιστογράμματα, θηκογράμματα, κανονικά διαγράμματα πιθανοτήτων (NPP) και διαγράμματα υπολοίπων (residuals plots).*

Αν διαπιστώσουμε ότι η υπόθεση/παραδοχή της *κανονικότητας* ή της *ομοσκεδαστικότητας* (ή και οι δύο) δεν ικανοποιούνται, είναι δυνατόν με κατάλληλο μετασχηματισμό της *μεταβλητής απόκρισης* (των παρατηρήσεων) να «διορθώσουμε» τις αποκλίσεις από αυτές τις παραδοχές. Στο θέμα αυτό δε θα επεκταθούμε περισσότερο, αναφέρουμε μόνο ότι συχνά εφαρμοζόμενοι μετασχηματισμοί είναι, ο λογάριθμος και η τετραγωνική ρίζα.

Αν δεν ικανοποιούνται οι υποθέσεις εφαρμογής *ελέγχου ανάλυσης διασποράς*, η *Στατιστική* μας προσφέρει λύσεις μέσω *μη παραμετρικών* ελέγχων, όπως ο *έλεγχος Kruskal-Wallis*, για τους οποίους οι προϋποθέσεις εφαρμογής είναι πολύ γενικές *(π.χ. ότι η μεταβλητή απόκρισης έχει συνεχή κατανομή).*

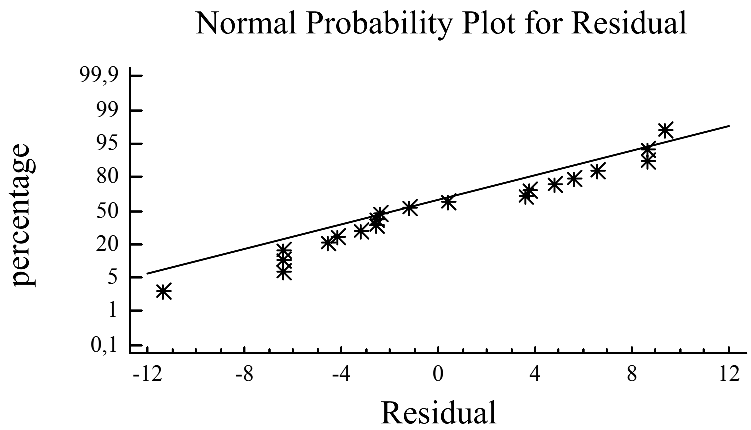
Ας δούμε, με τη βοήθεια του *στατιστικού πακέτου Statgraphics*, αν ικανοποιούνται οι υποθέσεις της *ανάλυσης διασποράς* στο *Παράδειγμα-1*.

Παράδειγμα-1 (συνέχεια): Για να ελέγξουμε αν ικανοποιούνται οι υποθέσεις ότι τα 4 δείγματα προέρχονται από *κανονικούς πληθυσμούς* με κοινή *διασπορά* σ^2, μπορούμε ισοδύναμα να ελέγξουμε αν για τα θεωρητικά υπόλοιπα $\varepsilon_{ij} = Y_{ij} - \mu_j$ ισχύει ότι $\varepsilon_{ij} \sim N(0,\sigma^2)$. Για το σκοπό αυτό, χρησιμοποιούμε τα «παρατηρούμενα» υπόλοιπα $\hat{\varepsilon}_{ij}$, καθένα από τα οποία υπολογίζεται αν από την αντίστοιχη παρατήρηση αφαιρέσουμε τη *μέση τιμή του δείγματος* στο οποίο αυτή ανήκει. Έτσι, επειδή οι τέσσερις *δειγματικοί μέσοι* είναι, αντίστοιχα, $\bar{y}_1 = 72.4, \bar{y}_2 = 68.6, \bar{y}_3 = 73.2$, $\bar{y}_4 = 76.4$, τα παρατηρούμενα *υπόλοιπα* των παρατηρήσεων από την *επέμβαση Α1* είναι, $81 - 72.4 = 8.6$, $66 - 72.4 = -6.4$, $78 - 72.4 = 5.6$, $76 - 72.4 = 3.6$, $61 - 72.4 = -11.4$ (ομοίως υπολογίζονται και για τις υπόλοιπες παρατηρήσεις από τις άλλες *επεμβάσεις*).

Στο *Σχήμα-4* φαίνεται το ***κανονικό διάγραμμα πιθανοτήτων*** για τα *υπόλοιπα* του παραδείγματός μας που κατασκευάσθηκε με το *στατιστικό πακέτο Statgraphics*. Στον οριζόντιο άξονα έχουν σημειωθεί τα $\hat{\varepsilon}_{ij}$ και στον κατακόρυφο οι αντίστοιχες *αναμενόμενες αθροιστικές πιθανότητες* αν η υπόθεση της *κανονικότητας* είναι σωστή. Αν τα σημεία που δημιουργούνται βρίσκονται (κατά προσέγγιση) σε μια ευθεία γραμμή, ή αλλιώς, αν κατανέμονται με τυχαίο τρόπο κοντά σε μια ευθεία γραμμή, τότε έχουμε μια καλή ένδειξη ότι η υπόθεση της *κανονικότητας* των *υπολοίπων* ικανοποιείται.

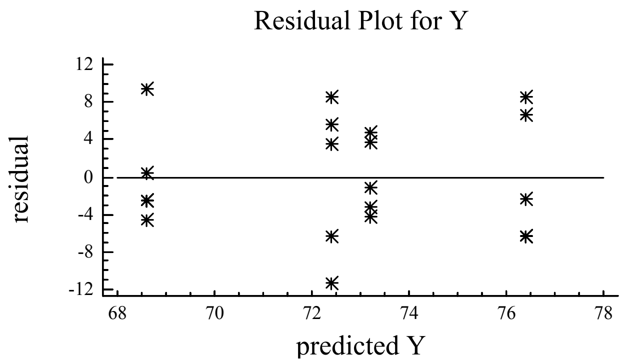
<u>Σημείωση:</u> Στο στατιστικό πακέτο Statgraphics, αλλά και σε πολλά στατιστικά πακέτα, αντί των αναμενόμενων αθροιστικών πιθανοτήτων, $i/(\nu+1)$, *χρησιμοποιούνται οι ελαφρώς τροποποιημένοι λόγοι,* $(i-0.375)/(\nu+0.25)$ *(τα παρατηρούμενα υπόλοιπα τα διατάσσουμε σε αύξουσα σειρά και το i εκφράζει τη θέση του αντίστοιχου παρατηρούμενου υπολοίπου στο διατεταγμένο πλέον δείγμα). Σημειώνουμε, επίσης, ότι το κανονικό διάγραμμα πιθανοτήτων μπορεί να κατασκευασθεί και με άλλους τρόπους, ώστε τα σημεία να αναμένουμε να βρίσκονται στην κύρια διχοτόμο των αξόνων.*

Τα σημεία στο *Σχήμα-4* δε φαίνεται να ακολουθούν κάποιο καμπυλόγραμμο πρότυπο, επομένως έχουμε μια καλή ένδειξη ότι, στο παράδειγμά μας, τα *υπόλοιπα* ακολουθούν *κανονική κατανομή*. Αυτό όμως δεν αρκεί, γι' αυτό πρέπει να κάνουμε και κάποιον (ή κάποιους) *έλεγχο καλής προσαρμογής*. Πράγματι, ο έλεγχος *Kolmogorov-Smirnov* έδωσε $P-Value>0.10$ και ο έλεγχος *Anderson-Darling* έδωσε $P-Value>0.2453$, επομένως δεν έχουμε στατιστικά σημαντικές ενδείξεις ότι μπορεί να απορριφθεί η υπόθεση ότι τα *υπόλοιπα ακολουθούν κανονική κατανομή*.



Σχήμα-4

Μια εικόνα για το αν τα *υπόλοιπα* έχουν *κοινή διασπορά* μπορούμε να πάρουμε από ένα *διάγραμμα υπολοίπων*, όπως αυτό που φαίνεται στο *Σχήμα-5* και αφορά τα *υπόλοιπα* του παραδείγματός μας.



Σχήμα-5

Στον οριζόντιο άξονα έχουν σημειωθεί οι τέσσερις *δειγματικοί* μέσοι ($\bar{y}_1 = 72.4,\ \bar{y}_2 = 68.6,\ \bar{y}_3 = 73.2,\ \bar{y}_4 = 76.4$) και στον κατακόρυφο τα αντίστοιχα *παρατηρούμενα υπόλοιπα* για τις παρατηρήσεις κάθε δείγματος. Αν για όλα τα δείγματα *(επεμβάσεις)* τα αντίστοιχα σημεία είναι περίπου το ίδιο διασκορπισμένα και κατανέμονται ομοιόμορφα (τυχαία) γύρω από την οριζόντια ευθεία που διέρχεται από το 0, έχουμε μια καλή ένδειξη ότι πράγματι τα *υπόλοιπα* έχουν κοινή *διασπορά*. Για το παράδειγμά μας, από το *Σχήμα-5*, δημιουργούνται αμφιβολίες για το αν η *διασπορά των υπολοίπων* είναι ίδια και στις τέσσερις *επεμβάσεις*. Απαιτείται επομένως να κάνουμε κατάλληλο έλεγχο *ισότητας των διασπορών* των τεσσάρων πληθυσμών. Ο έλεγχος *Bartlett* έδωσε $P-Value = 0.5647$ και ο έλεγχος *Cochran* έδωσε $P-Value = 0.5242$, επομένως δεν έχουμε στατιστικά σημαντικές ενδείξεις ότι η υπόθεση της *ομοσκεδαστικότητας* πρέπει να απορριφθεί.

<u>Σχέδιο Τυχαιοποιημένων Πλήρων Ομάδων (Randomized Complete Block Design)</u>

Το σχέδιο των *τυχαιοποιημένων πλήρων ομάδων* αποτελεί ευθεία γενίκευση του σχεδίου που γνωρίσαμε όταν μιλήσαμε για τη ***σύγκριση κατά ζεύγη*** δύο *μέσων*, μ_1 και μ_2. Η διαφορά από τη *σύγκριση κατά ζεύγη* έγκειται μόνο στο ότι αντί να δημιουργούμε έναν αριθμό, έστω *b*, ***<u>ζευγών</u>*** πειραματικών μονάδων επί των οποίων κάνουμε δύο *επεμβάσεις*, δημιουργούμε έναν αριθμό, *b*, ***<u>ομάδων (blocks)</u>*** αποτελούμενων από $k > 2$ πειραματικές μονάδες η κάθε μια (αντί δύο) γιατί κάνουμε περισσότερες από δύο *επεμβάσεις* (ή αλλιώς, γιατί ο *παράγοντας* του οποίου ελέγχουμε την επίδραση στη *μεταβλητή απόκρισης*, έχει περισσότερες από δύο *στάθμες*).

Για τη σκοπιμότητα επιλογής αυτού του σχεδίου και για τον τρόπο/κριτήρια δημιουργίας των *ομάδων*, ισχύουν όσα (πολύ αναλυτικά) αναφέραμε όταν μιλήσαμε για τη *σύγκριση κατά ζεύγη*. Έτσι, επιλέγουμε να εφαρμόσουμε το *σχέδιο των τυχαιοποιημένων πλήρων ομάδων,* όταν θεωρούμε/υποθέτουμε ότι οι πειραματικές μονάδες προσθέτουν μεταβλητότητα στις τιμές της *μεταβλητής απόκρισης*, με την έννοια ότι αυτή επηρεάζεται και από κάποιον άλλο, «εξωγενή/ενοχλητικό», παράγοντα, πέραν του παράγοντα του οποίου ελέγχουμε την επίδραση. Οι *ομάδες* δημιουργούνται ακριβώς όπως τα *ζεύγη* στη *σύγκριση κατά ζεύγη*, δηλαδή, με κριτήριο, να αποτελούνται από πειραματικές μονάδες που έχουν παρόμοια/ομοιογενή συμπεριφορά ως προς αυτόν τον «εξωγενή» παράγοντα. Έτσι, επιτυγχάνουμε, ο έλεγχος της επίδρασης του *παράγοντα* που μας ενδιαφέρει να γίνεται εντός κάθε *ομάδας* σε παρόμοιες συνθήκες και να απομονωθεί η μεταβλητότητα που οφείλεται στον «εξωγενή» παράγοντα.

Αφού δημιουργήσουμε τις *b ομάδες* (με *k* πειραματικές μονάδες την κάθε μια), αντιστοιχίζουμε, με μια <u>τυχαία διαδικασία</u>, μια προς μια, τις *k επεμβάσεις* στις *k* πειραματικές μονάδες κάθε *ομάδας.* Έτσι συνολικά παίρνουμε, $\nu = b \cdot k$ παρατηρήσεις.

Ένα τέτοιο σχέδιο είναι αυτό του *Προβλήματος-2*. Στο πρόβλημα αυτό, ο φοιτητής, πρέπει να απαντήσει στο ίδιο ακριβώς ερώτημα που τίθεται και στο *Πρόβλημα-1,* χρησιμοποιώντας όμως, μετά από υπόδειξη του καθηγητή του, πειραματικά δεδομένα από ένα άλλο πείραμα που σχεδιάσθηκε με το *σχέδιο των τυχαιοποιημένων πλήρων ομάδων*. Ο καθηγητής έκανε αυτή την υπόδειξη στο φοιτητή, γιατί γνωρίζει ότι η αύξηση του βάρους των νεογέννητων επηρεάζεται και από το γένος/φυλή των ζώων κάτι το οποίο δεν ελήφθη υπόψη στον πρώτο σχεδιασμό, με αποτέλεσμα η μεταβλητότητα της αύξησης του βάρους των νεογέννητων χοίρων που οφείλεται σε αυτόν τον παράγοντα να παραμένει ενσωματωμένη στο *SSE,* το οποίο βρίσκεται στον παρανομαστή του *κριτηρίου* του ελέγχου και έτσι, πιθανή σημαντικότητα του παράγοντα *«είδος πρόσθετης ύλης»* να μην αναδεικνύεται αλλά να την «κρύβει» το μεγάλο *SSE*. Δημιουργώντας, επομένως, ομάδες νεογέννητων από διαφορετική γέννα την κάθε μια (στην ίδια ομάδα όμως, τα νεογέννητα να προέρχονται από την ίδια γέννα), θέλησε να υπολογίσει αυτή τη μεταβλητότητα από ομάδα σε ομάδα/από γέννα σε γέννα, δηλαδή, τη μεταβλητότητα που οφείλεται σε γεννετικά/κληρονομικά χαρακτηριστικά και έτσι να αναλύσει περαιτέρω το *SSE* και κατ' επέκταση τη *συνολική μεταβλητότητα (SSTot)*. Δηλαδή, θέλησε να αφαιρέσει/απομονώσει από το *SSE* τη μεταβλητότητα από γέννα σε γέννα και έτσι να το ελαττώσει. Στη συνέχεια, θα δούμε αν «δικαιώθηκε» αυτή η σχεδιαστική επιλογή του καθηγητή (γιατί έχει και κόστος όπως θα δούμε).

Η βασική ιδέα της μεθόδου, όπως ήδη έχουμε αναφέρει, δεν αλλάζει με το *πειραματικό σχέδιο*. Παραμένει η ίδια, και είναι, όπως είδαμε, η ***ανάλυση*** της *συνολικής μεταβλητότητας,* δηλαδή της *συνολικής* ***διασποράς,*** σε συνιστώσες (από εδώ και το όνομά της). Έτσι και σε αυτό το *πειραματικό σχέδιο*, η *συνολική μεταβλητότητα* των τιμών της *μεταβλητής απόκρισης* αναλύεται σε συνιστώσες, όχι όμως σε δύο, αλλά σε τρεις. Η επιπλέον συνιστώσα είναι αυτή που εκφράζει τη *μεταβλητότητα* μεταξύ των *ομάδων* (μεταξύ των *blocks*). Με αυτόν τον τρόπο, επιτυγχάνεται περαιτέρω ελάττωση του *SSE*. Επομένως, στο *σχέδιο τυχαιοποιημένων πλήρων ομάδων* έχουμε,

$$SSTot = SSTr + SSB + SSE\ ,$$

όπου,

$SSTot = \sum_{ij}(y_{ij} - \bar{y})^2 = \sum_{ij} y_{ij}^2 - \dfrac{G^2}{\nu}$, το <u>ολικό άθροισμα τετραγώνων</u> που εκφράζει την <u>ολική μεταβλητότητα</u> των $\nu = b \cdot k$ παρατηρήσεων και έχει $b \cdot k - 1$ *βαθμούς ελευθερίας,*

$SSTr = \sum_{j=1}^{k} \dfrac{T_j^2}{b} - \dfrac{G^2}{\nu}$, το <u>άθροισμα τετραγώνων των επεμβάσεων</u> που εκφράζει τη <u>μεταβλητότητα μεταξύ των μέσων των k επεμβάσεων</u> και έχει $k-1$ *βαθμούς ελευθερίας,*

$SSB = \sum_{i=1}^{b} \dfrac{B_i^2}{k} - \dfrac{G^2}{\nu}$, το <u>άθροισμα τετραγώνων των ομάδων</u> που εκφράζει τη <u>μεταβλητότητα μεταξύ των μέσων των b ομάδων</u> και έχει $b-1$ *βαθμούς ελευθερίας,*

$SSE = SSTot - SSTr - SSB$, το <u>άθροισμα τετραγώνων του σφάλματος</u> και έχει $(b-1)\cdot(k-1)$ *βαθμούς ελευθερίας.*

Με $G = \sum_{ij} y_{ij}$ συμβολίζουμε το *γενικό άθροισμα (Grand total)* των παρατηρήσεων, δηλαδή, το άθροισμα όλων των $\nu = b \cdot k$ παρατηρήσεων, με T_j, $j = 1, 2, ..., k$, συμβολίζουμε το άθροισμα όλων των παρατηρήσεων από την j *επέμβαση* και με B_i, $i = 1, 2, ..., b$, το άθροισμα όλων των παρατηρήσεων στην i *ομάδα*.

Για τη σύγκριση των *μέσων*, $\mu_1, \mu_2, ..., \mu_k$, των *k επεμβάσεων* χρησιμοποιούμε όπως και στο *εντελώς τυχαιοποιημένο σχέδιο* τα *μέσα αθροίσματα τετραγώνων*.

Έτσι, για τον έλεγχο της *μηδενικής υπόθεσης,*

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

έναντι της *εναλλακτικής,*

$$H_1 : \mu_i \neq \mu_j \text{, για ένα τουλάχιστον ζεύγος (i, j),}$$

χρησιμοποιούμε και πάλι το λόγο $F_{tr} = \dfrac{MSTr}{MSE}$ *(τώρα βέβαια το SSE έχει μειωθεί, ελπίζουμε σημαντικά…, αλλά έχει λιγότερους βαθμούς ελευθερίας).*

Αποδεικνύεται ότι, με την υπόθεση ότι η μηδενική υπόθεση είναι αληθής, η τυχαία μεταβλητή F_{Tr} ακολουθεί την κατανομή F με $k-1$ και $(b-1)(k-1)$ βαθμούς ελευθερίας, δηλαδή,

$$F_{tr} = \frac{MSTr}{MSE} \sim F_{k-1;(b-1)(k-1)}.$$

Επομένως, σε επίπεδο σημαντικότητας $\alpha\%$, η απορριπτική περιοχή του ελέγχου είναι:

$$F_{tr} = \frac{MSTr}{MSE} > F_{k-1;(b-1)(\mathrm{k}-1);\alpha}.$$

Παρότι, όπως εξηγήσαμε, επιλέγουμε το *σχέδιο τυχαιοποιημένων πλήρων ομάδων*, για να αναδείξουμε πιο αποτελεσματικά (μειώνοντας το *SSE*) την επίδραση στη *μεταβλητή απόκρισης* του συγκεκριμένου *παράγοντα* που μας ενδιαφέρει (των *επεμβάσεων*), μπορούμε, επιπλέον, να ελέγξουμε αν είναι στατιστικά σημαντική η μεταβλητότητα από *ομάδα* σε *ομάδα*. Δηλαδή, να κάνουμε τον έλεγχο της *μηδενικής υπόθεσης,*

$$H_0 : \text{οι } b \text{ μέσοι των ομάδων είναι ίσοι}$$

έναντι της *εναλλακτικής,*

$$H_1 : \text{τουλάχιστον δύο μέσοι ομάδων διαφέρουν.}$$

Αποδεικνύεται ότι, με την υπόθεση ότι η μηδενική υπόθεση είναι αληθής,

$$F_B = \frac{MSB}{MSE} \sim F_{b-1;(b-1)(k-1)}.$$

Επομένως, σε επίπεδο σημαντικότητας $\alpha\%$, η απορριπτική περιοχή του ελέγχου είναι:

$$F_B = \frac{MSB}{MSE} > F_{b-1;(b-1)(\mathrm{k}-1);\alpha}.$$

Συνοψίζουμε, δίνοντας τον *πίνακα ανάλυσης διασποράς* του σχεδίου.

Πίνακας Ανάλυσης Διασποράς
(για το σχέδιο τυχαιοποιημένων πλήρων ομάδων)

Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***	***Περιοχή απόρριψης***
Επεμβάσεις ή Παράγοντας ή μεταξύ των δειγμάτων	$k-1$	$SSTr$	$MSTr = \frac{SSTr}{k-1}$	$F_{Tr} = \frac{MSTr}{MSE}$	$F_{Tr} > F_{k-1;(b-1)(k-1);\alpha}$
Ομάδες	$b-1$	SSB	$MSB = \frac{SSB}{b-1}$	$F_B = \frac{MSB}{MSE}$	$F_B > F_{b-1;(b-1)(k-1);\alpha}$
Σφάλμα ή εντός των δειγμάτων	$(b-1)(k-1)$	SSE	$MSE = \frac{SSE}{(b-1)(k-1)}$		
Ολική	$k \cdot b - 1$	$SSTot$			

Έλεγχοι πολλαπλών συγκρίσεων *(στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων)*

Όπως και στην περίπτωση του *εντελώς τυχαιοποιημένου σχεδίου*, όταν η *μηδενική υπόθεση* $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ απορριφθεί, δηλαδή, όταν τα πειραματικά δεδομένα υποστηρίζουν ότι οι *μέσοι*, $\mu_1, \mu_2, ..., \mu_k$, των *k επεμβάσεων* διαφέρουν στατιστικά σημαντικά, μπορούμε να κάνουμε *ελέγχους πολλαπλών συγκρίσεων (π.χ. έλεγχο Tukey)* για να διερευνήσουμε ποιοι από τους *k μέσους* διαφέρουν στατιστικά σημαντικά.

Παρατηρήσεις:

1. *<u>Ερμηνεία του ελέγχου για την επίδραση των ομάδων:</u> Αν ο έλεγχος αυτός οδηγήσει σε στατιστικά σημαντικό αποτέλεσμα, δηλαδή, αν οι διαφορές στη μεταβλητή απόκρισης αποδειχθούν στατιστικά σημαντικές από ομάδα σε ομάδα, αυτό σημαίνει ότι σωστά επιλέξαμε το σχέδιο των τυχαιοποιημένων πλήρων ομάδων και έτσι αυτή η σχεδιαστική επιλογή μας «δικαιώθηκε». Αν οι διαφορές αυτές δεν αποδειχθούν στατιστικά σημαντικές, αυτό σημαίνει ότι η μεταβλητότητα που απομονώσαμε (δημιουργώντας τις ομάδες) δεν είναι σημαντική και επομένως δεν επωφεληθήκαμε σε κάτι από τα πλεονεκτήματα αυτού του σχεδίου, ενώ αντίθετα, μειώθηκαν οι βαθμοί ελευθερίας του SSE, χωρίς «αντισταθμιστικό» όφελος (αξιόλογη μείωση του SSE) γεγονός που δημιουργεί πρόβλημα στην ισχύ του ελέγχου <u>για την επίδραση του παράγοντα που μας ενδιαφέρει</u> (αφού το MSE βρίσκεται στον παρανομαστή κριτηρίου για τον έλεγχο της σημαντικότητας του παράγοντα). Βέβαια, υπάρχει και η περίπτωση, να μην ήταν η ιδέα επιλογής του συγκεκριμένου σχεδίου κακή, αλλά ο τρόπος εφαρμογής του (π.χ. αστοχία στη συγκρότηση των ομάδων). Σημειώνουμε, τέλος, ότι μεταξύ των επιστημόνων υπάρχουν διαφορετικές απόψεις για την ερμηνεία αλλά και την αξία του ελέγχου για τη σημαντικότητα των ομάδων. Δε θα επεκταθούμε όμως περισσότερο σε θέματα πειραματικού σχεδιασμού.*
2. *Όπως στο εντελώς τυχαιοποιημένο σχέδιο, όπου για $k = 2$ ο έλεγχος ανάλυσης διασποράς είναι ισοδύναμος με τον έλεγχο δύο μέσων με δύο ανεξάρτητα τυχαία δείγματα, έτσι και στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων, ο έλεγχος ανάλυσης διασποράς για $k = 2$, δηλαδή για δύο μόνο επεμβάσεις σε κάθε ομάδα, είναι ισοδύναμος με τον έλεγχο ζευγαρωτών παρατηρήσεων (δες Πρόβλημα-10).*

Ας δούμε ένα παράδειγμα.

<u>Παράδειγμα-3:</u> *Μια ομάδα ερευνητών προκειμένου να συγκρίνει την επίδραση τριών μεθόδων προετοιμασίας του εδάφους στην ανάπτυξη ενός είδους πεύκης κατά το πρώτο έτος από τη φύτευση, σχεδίασε το εξής πείραμα. Επέλεξε τέσσερις εκτάσεις (ίδιου εμβαδού) σε τέσσερις διαφορετικές δασικές περιοχές και κάθε μια τη διαίρεσε σε τρία πειραματικά τεμάχια ίδιου εμβαδού. Σε καθένα από τα τρία πειραματικά τεμάχια κάθε έκτασης αντιστοίχησε, με μια τυχαία διαδικασία, μια από τις τρεις μεθόδους προετοιμασίας του εδάφους. Στη συνέχεια, σε κάθε πειραματικό τεμάχιο, αφού προετοίμασε το έδαφος με την αντίστοιχη μέθοδο, φύτεψε ένα δενδρύλλιο πεύκης. Στον πίνακα που ακολουθεί, δίνονται (σε cm) οι μετρήσεις που πήραν οι ερευνητές ένα έτος μετά τη φύτευση .*

		Περιοχή φύτευσης			
		E1	*E2*	*E3*	*E4*
Μέθοδος	*M1*	*11*	*13*	*16*	*10*
	M2	*15*	*17*	*20*	*12*
	M3	*10*	*15*	*13*	*10*

Οι ερευνητές θέλησαν να απομονώσουν τη μεταβλητότητα στις τιμές της *μεταβλητής απόκρισης* (*ανάπτυξη δενδρυλλίων πεύκης κατά το πρώτο έτος από τη φύτευση*) που οφείλεται στην *περιοχή φύτευσης* και γι' αυτό δημιούργησαν 4 ομάδες (καθόρισαν 4 διαφορετικές *περιοχές*) και σε κάθε μια από αυτές έκαναν και τις 3 *επεμβάσεις* (*μέθοδοι προετοιμασίας του εδάφους*) γιατί θεώρησαν ότι η ανάπτυξη των δενδρυλλίων πεύκης επηρεάζεται, εκτός από τον παράγοντα *μέθοδος προετοιμασίας του εδάφους* (που τους ενδιαφέρει να μελετήσουν την επίδρασή του), και από την *περιοχή φύτευσης.*

Ας δούμε αν η επίδραση του παράγοντα *μέθοδος προετοιμασίας του εδάφους* είναι, σε *επίπεδο σημαντικότητας 5%* , στατιστικά σημαντική.

Θα κάνουμε, για $\alpha = 5\%$, τον έλεγχο της *μηδενικής υπόθεσης,* $H_0 : \mu_1 = \mu_2 = \mu_3$, έναντι της *εναλλακτικής,* $H_1 : \mu_1 \neq \mu_2$ *ή* $\mu_1 \neq \mu_3$ *ή* $\mu_2 \neq \mu_3$, όπου, μ_1, μ_2, μ_3, οι *μέσοι* των τριών *επεμβάσεων M1, M2, M3,* αντίστοιχα (δηλαδή, οι *μέσες τιμές* των αυξήσεων των δενδρυλλίων όταν για την προετοιμασία του εδάφους εφαρμόζεται η μέθοδος *M1, M2, M3,* αντίστοιχα).

Για κάθε *επέμβαση* έχουμε 4 παρατηρήσεις επομένως συνολικά έχουμε στη διάθεσή μας $\nu = 12$ παρατηρήσεις. Για να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*, κατ' αρχάς, θα υπολογίσουμε τα *αθροίσματα τετραγώνων* *SSTr*, *SSB*, *SSE* και *SSTot.*

		Ομάδα (περιοχή φύτευσης)				
		E1	*E2*	*E3*	*E4*	T_j
Παράγοντας/Επεμβάσεις (μέθοδος)	*M1*	*11*	*13*	*16*	*10*	*50*
	M2	*15*	*17*	*20*	*12*	*64*
	M3	*10*	*15*	*13*	*10*	*48*
	B_i	*36*	*45*	*49*	*32*	$G = 162$

Το *γενικό άθροισμα* των $\nu = 12$ παρατηρήσεων είναι, $G = \sum_{ij} y_{ij} = 162$ και επομένως:

$SSTot = \sum_{ij} y_{ij}^2 - \frac{G^2}{\nu} = (11^2 + 15^2 + ... + 10^2) - \frac{162^2}{12} = 2298 - 2187 = 111$, με $\nu - 1 = 11$ *βαθμούς ελευθερίας.*

$SSTr = \sum_{j=1}^{3} \frac{T_j^2}{b} - \frac{G^2}{\nu} = \frac{50^2 + 64^2 + 48^2}{4} - \frac{162^2}{12} = 2225 - 2187 = 38$, με $k - 1 = 2$ *βαθμούς ελευθερίας.*

$SSB = \sum_{i=1}^{4} \frac{B_i^2}{k} - \frac{G^2}{\nu} = \frac{36^2 + 45^2 + 49^2 + 32^2}{3} - \frac{162^2}{12} = 2248.67 - 2187 = 61.67$ με $b - 1 = 3$ *βαθμούς ελευθερίας.*

$SSE = SSTot - SSTr - SSE = 111 - 38 - 61.67 = 11.33$, με $(b-1)(k-1) = 6$ *βαθμούς ελευθερίας.*

Μπορούμε πλέον να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*:

Πίνακας Ανάλυσης Διασποράς				
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***
Επεμβάσεις *(μέθοδοι)*	*2*	*38*	$MSTr = \frac{38}{2} = 19$	$F_{Tr} = \frac{19}{1.89} = 10.06$
Ομάδες *(περιοχές)*	*3*	*61.67*	$MSB = \frac{61.67}{3} = 20.56$	$F_B = \frac{20.56}{1.89} = 10.88$
Σφάλμα	6	*11.33*	$MSE = \frac{11.33}{6} = 1.89$	
Ολική	*11*	*111*		

Επειδή η τιμή του κριτηρίου, $F_{Tr} = 10.06$, βρίσκεται στην *περιοχή απόρριψης* της *μηδενικής υπόθεσης* αφού, $10.06 > F_{2;6;0.05} = 5.14$, συμπεραίνουμε ότι τα πειραματικά δεδομένα υποστηρίζουν ότι η *μέθοδος προετοιμασίας του εδάφους* επηρεάζει την *ανάπτυξη των δενδρυλλίων πεύκης κατά το πρώτο έτος από τη φύτευση*. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 5%.

Ας δούμε τώρα, αν η μεταβλητότητα από *ομάδα* σε *ομάδα,* δηλαδή, από *περιοχή* σε *περιοχή,* είναι στατιστικά σημαντική. Θα κάνουμε, σε *επίπεδο σημαντικότητας 5%* , τον έλεγχο της *μηδενικής υπόθεσης,* H_0 : *οι τέσσερις μέσοι των ομάδων είναι ίσοι* έναντι της *εναλλακτικής,* H_1 : *τουλάχιστον δύο μέσοι ομάδων διαφέρουν.*

Έχουμε ήδη υπολογίσει την τιμή του κριτηρίου, F_B, και επειδή, $F_B = 10.88 > F_{3;6;0.05} = 4.76$, συμπεραίνουμε ότι τα πειραματικά δεδομένα, σε *επίπεδο σημαντικότητας 5%* , υποστηρίζουν ότι η μεταβλητότητα μεταξύ των περιοχών, είναι στατιστικά σημαντική. Αυτό σημαίνει ότι «δικαιώθηκε» η επιλογή των ερευνητών να χρησιμοποιήσουν *πειραματικό σχέδιο* που απομόνωσε αυτή τη μεταβλητότητα από το *SSE*.

<u>Παράδειγμα-4 (αναφέρεται στο Πρόβλημα-2):</u> Όπως εξηγήσαμε, ο καθηγητής, υπέδειξε ως *πειραματικό σχέδιο* το *σχέδιο τυχαιοποιημένων πλήρων ομάδων*. Στον πίνακα που ακολουθεί φαίνονται τα πειραματικά δεδομένα που συγκέντρωσε ο φοιτητής ταξινομημένα ανά *ομάδα* και ανά *επέμβαση*.

		Παράγοντας/Επεμβάσεις (πρόσθετη ύλη)			
		A1	*A2*	*A3*	*A4*
Ομάδα (γέννα)	*Γ1*	*78*	*69*	*78*	*80*
	Γ2	*68*	*64*	*70*	*70*
	Γ3	*76*	*70*	*73*	*83*
	Γ4	*76*	*66*	*77*	*74*
	Γ5	*61*	*66*	*69*	*70*

Ας δούμε αν η επίδραση του *παράγοντα πρόσθετη ύλη ζωοτροφών* είναι, σε *επίπεδο σημαντικότητας 5%*, στατιστικά σημαντική, δηλαδή αν η *αύξηση του βάρους των νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους,* επηρεάζεται από *το είδος πρόσθετης ύλης ζωοτροφών* που περιέχεται στην τροφή τους.

Θα κάνουμε, σε *επίπεδο σημαντικότητας 5%,* τον έλεγχο της *μηδενικής υπόθεσης,*

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 ,$$

έναντι της *εναλλακτικής*,

$$H_1 : \mu_i \neq \mu_j, \textit{ για ένα τουλάχιστον ζευγάρι } (i, j),\ i, j = 1,2,3,4,$$

όπου, μ_j, $j = 1,2,3,4$, η *μέση αύξηση του βάρους των νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους, όταν στην τροφή τους περιέχεται η πρόσθετη ύλη* A_j.

Ο φοιτητής, για τους αναγκαίους υπολογισμούς, χρησιμοποίησε το *στατιστικό πακέτο Statgraphics* από το οποίο πήρε τον ακόλουθο *πίνακα ανάλυσης διασποράς.*

Analysis of Variance for example4

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Factor	192.6	3	64.2	6.37	0.0079
Blocks	312.3	4	78.075	7.75	0.0025
RESIDUAL	120.9	12	10.075		
TOTAL	625.8	19			

Επειδή $P-Value = 0.0079 < 0.05$ (ή επειδή $F_{Tr} = 6.37 > F_{3;12;0.05} = 3.49$), η *μηδενική υπόθεση*, σε *επίπεδο σημαντικότητας 5%* απορρίπτεται. Δηλαδή, τα πειραματικά δεδομένα από τις τρεις επεμβάσεις, δίνουν στατιστικά σημαντικές αποδείξεις ότι η *μέση αύξηση του βάρους των νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους* **επηρεάζεται** από το είδος της *πρόσθετης ύλης που περιέχεται στην τροφή τους.* Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 0.05 (μάλιστα, επειδή το *στατιστικό πακέτο* υπολόγισε την *P-Value*, μπορούμε να πούμε ότι αυτή η πιθανότητα είναι το πολύ 0.0079). Βλέπουμε, δηλαδή, ότι απομονώνοντας ο φοιτητής από το *SSE,* εκτός από τη μεταβλητότητα που οφείλεται στον παράγοντα που μελετάει *(είδος πρόσθετης ύλης)*, και τη μεταβλητότητα που οφείλεται σε μια άλλη πηγή (τη γέννα) η ***ισχύς*** του ελέγχου αυξήθηκε και η επίδραση του *είδους της πρόσθετης ύλης* αναδείχθηκε στατιστικά σημαντική. Κάτι το οποίο δε συνέβη όταν χρησιμοποίησε το *εντελώς τυχαιοποιημένο σχέδιο* όπου κάθε είδους μεταβλητότητα πέραν αυτής που οφείλεται στον παράγοντα που μελετάει *(είδος πρόσθετης ύλης)* εμπεριέχεται/παραμένει ενσωματωμένη στο *SSE (δες Παράδειγμα-1)*. Φαίνεται επομένως, να δικαιώνεται η υπόδειξη του καθηγητή, κάτι το οποίο επιβεβαιώνεται και από τον έλεγχο σημαντικότητας της μεταβλητότητας μεταξύ των *ομάδων*, δηλαδή από τον έλεγχο της *μηδενικής υπόθεσης*,

$$H_0 : \textit{οι πέντε μέσοι των ομάδων είναι ίσοι}$$

έναντι της *εναλλακτικής,*

$$H_1 : \textit{τουλάχιστον δύο μέσοι ομάδων διαφέρουν.}$$

όπου, η *μηδενική υπόθεση,* σε *επίπεδο σημαντικότητας 5%* απορρίπτεται ($F_B = 7.75 > F_{4;12;0.05} = 3.26$).

Υποθέσεις/παραδοχές στο σχέδιο των τυχαιοποιημένων πλήρων ομάδων

Οι *b* παρατηρήσεις, σε κάθε μία από τις *k επεμβάσεις,* συγκεντρώνονται από τις *ομάδες* (μία παρατήρηση από κάθε ομάδα). Επομένως, οι *k* παρατηρήσεις κάθε *ομάδας* είναι προφανώς ***εξαρτημένες*** (γιατί σε κάθε μια αντανακλώνται οι συνθήκες που ενυπάρχουν στη συγκεκριμένη ομάδα). Δηλαδή, η ανεξαρτησία των παρατηρήσεων, στο *σχέδιο των τυχαιοποιημένων πλήρων ομάδων*, δεν υφίσταται για όλες τις παρατηρήσεις. Γι' αυτό, στο σχέδιο αυτό, παίζει πολύ σημαντικό ρόλο η τυχαία αντιστοίχηση των *επεμβάσεων* στις *πειραματικές μονάδες* κάθε ομάδας.

Στο *σχέδιο των τυχαιοποιημένων πλήρων ομάδων*, πέραν των δύο υποθέσεων ότι,

1. καθένα από τα $k \cdot b$ δείγματα προέρχεται από *κανονικό* πληθυσμό,
2. οι $k \cdot b$ πληθυσμοί έχουν *κοινή διασπορά,* σ^2,

επιπλέον, κάνουμε την υπόθεση ότι,

3. οι *επεμβάσεις* και οι *ομάδες* επιδρούν στη *μεταβλητή απόκρισης* προσθετικά/ανεξάρτητα.

Οι υποθέσεις, (1) και (2) είναι ισοδύναμες με τις υποθέσεις ότι τα *υπόλοιπα,* ε_{ij}, ακολουθούν *κανονικές* κατανομές με *μέση τιμή* 0 και κοινή *διασπορά* σ^2, δηλαδή, ότι $\varepsilon_{ij} \sim N(0,\sigma^2)$. Οι υποθέσεις αυτές ελέγχονται όπως στα προηγούμενα.

α x b Παραγοντικό Πείραμα (α x b Factorial Experiment)

Ας δούμε ένα παράδειγμα *παραγοντικού πειράματος*.

Παράδειγμα-5: *Ένας υποψήφιος διδάκτορας, στο πλαίσιο της διδακτορικής του διατριβής που είχε ως αντικείμενο τη μελέτη του κύκλου ζωής της λευκής μύγας (whitefly), η οποία δημιουργεί πολύ σοβαρό πρόβλημα στις γεωργικές καλλιέργειες, σχεδίασε και εκτέλεσε στο εργαστήριο ένα πείραμα που σκοπό είχε τη διερεύνηση της επίδρασης στον κύκλο ζωής της λευκής μύγας δύο συγκεκριμένων παραγόντων, του επιπέδου της θερμοκρασίας και του είδους φυτών που προσβάλλει. Συγκεκριμένα, σχεδίασε και εκτέλεσε το εξής πείραμα. Καθόρισε τρία επίπεδα θερμοκρασίας (70^oF, 77^oF και 82^oF) και δύο είδη φυτών (βαμβάκι και αγγουριά). Στη συνέχεια, τοποθέτησε λευκές μύγες και στα δύο είδη φυτών υπό τις τρεις διαφορετικές θερμοκρασίες για όλους τους διαφορετικούς συνδυασμούς, είδος φυτού-επίπεδο θερμοκρασίας, και μέτρησε τον αριθμό των αυγών που γέννησαν. Για κάθε συνδυασμό είδος φυτού-επίπεδο θερμοκρασίας πήρε 5 παρατηρήσεις. Τα δεδομένα που συγκέντρωσε φαίνονται στον πίνακα που ακολουθεί.*

		Επίπεδο θερμοκρασίας		
		70^oF	*77^oF*	*82^oF*
Είδος φυτού	*Βαμβάκι*	*37 21 36 43 31*	*34 54 40 42 16*	*46 32 41 36 38*
	Αγγουριά	*50 53 25 37 48*	*59 53 31 69 51*	*62 71 49 43 59*

Ο μεταπτυχιακός φοιτητής σχεδίασε και εκτέλεσε ένα *2 x 3 παραγοντικό πείραμα* με 5 παρατηρήσεις για κάθε μια από τις 6 *επεμβάσεις* (για κάθε συνδυασμό, *είδος φυτού-επίπεδο θερμοκρασίας*), δηλαδή, συνολικά συγκέντρωσε 30 παρατηρήσεις.

Ένα άλλο παράδειγμα *παραγοντικού πειράματος* είναι αυτό του *Προβλήματος-4*, όπου ο ερευνητής σχεδίασε και εκτέλεσε ένα *3 x 3 παραγοντικό πείραμα* με δύο παρατηρήσεις για κάθε μια από τις 9 *επεμβάσεις* (για κάθε συνδυασμό, *ένταση φωτισμού-επίπεδο θερμοκρασίας*) και επομένως, συνολικά, συγκέντρωσε 18 παρατηρήσεις.

Γενικότερα, όταν θέλουμε να μελετήσουμε την επίδραση (στη *μεταβλητή απόκρισης)* δύο παραγόντων, έστω *A* και *B*, με α *στάθμες* ο *A* και *b στάθμες* ο *B,* μπορούμε να σχεδιάσουμε και να εκτελούμε ένα ***α x b παραγοντικό πείραμα***. Σε ένα τέτοιο πείραμα, <u>κάθε *στάθμη* του ενός παράγοντα συνδυάζεται με κάθε *στάθμη* του άλλου παράγοντα</u>, δηλαδή, συνολικά έχουμε $\alpha \cdot b$ διαφορετικές *επεμβάσεις* (διαφορετικούς *συνδυασμούς από δύο στάθμες*, μια από κάθε παράγοντα) που ονομάζονται και *κελιά (cells)*. Οι παράγοντες μπορεί να είναι, όπως στο *Παράδειγμα-5* και στο *Πρόβλημα-4*, *καθορισμένων επιδράσεων* και οι δύο *(πρότυπο I)* ή *τυχαίων επιδράσεων* και οι δύο *(πρότυπο II)* ή ο ένας *καθορισμένων* και ο άλλος *τυχαίων επιδράσεων (μεικτό πρότυπο ή πρότυπο III)*. Σημειώνουμε, τέλος, χωρίς να επεκταθούμε περισσότερο σε θέματα *πειραματικού σχεδιασμού*, ότι ένα *παραγοντικό πείραμα*, αυτό καθαυτό, δε θεωρείται είδος/τύπος *πειραματικού σχεδίου*. Σε ένα παραγοντικό πείραμα, για τη συγκέντρωση των πειραματικών δεδομένων, μπορεί να χρησιμοποιηθεί, είτε το *εντελώς τυχαιοποιημένο σχέδιο*, είτε το *σχέδιο τυχαιοποιημένων πλήρων ομάδων* ή κάποιο άλλο σχέδιο (στα προηγούμενα παραδείγματα οι παρατηρήσεις συγκεντρώθηκαν με βάση το *εντελώς τυχαιοποιημένο σχέδιο*). Στη συνέχεια θα ασχοληθούμε με προβλήματα *παραγοντικών πειραμάτων* στα οποία και οι δύο παράγοντες είναι *καθορισμένων επιδράσεων (πρότυπο I)* και οι παρατηρήσεις έχουν συγκεντρωθεί με το *εντελώς τυχαιοποιημένο σχέδιο*.

Αν δε μπορούμε να υποθέσουμε (να κάνουμε την παραδοχή), ότι οι δύο παράγοντες επιδρούν στη *μεταβλητή απόκρισης* ανεξάρτητα, πρέπει για κάθε *επέμβαση* να πάρουμε περισσότερες από μια παρατηρήσεις. Διευκρινίζουμε ότι, στην περίπτωση αυτή, δεν είναι απαραίτητο να πάρουμε ίδιο αριθμό παρατηρήσεων για κάθε *επέμβαση*. Αυτό που είναι απαραίτητο, είναι οι παρατηρήσεις για κάθε *επέμβαση* να είναι περισσότερες από μια. Στη συνέχεια, θεωρούμε ότι για κάθε *επέμβαση* παίρνουμε ίδιο αριθμό παρατηρήσεων έστω, *r,* και επομένως, ότι συνολικά έχουμε $\alpha \cdot b \cdot r$ παρατηρήσεις.

Αν $r > 1$, πέραν της επίδρασης του *παράγοντα Α* και της επίδρασης του *παράγοντα Β*, μπορούμε να ελέγξουμε και αν, στη *μεταβλητή απόκρισης,* υπάρχει στατιστικά σημαντική επίδραση που να οφείλεται στην ***αλληλεπίδραση*** των *παραγόντων Α* και *Β*. Αυτό σημαίνει ότι μπορούμε να ελέγξουμε αν η επίδραση του ενός παράγοντα εξαρτάται από τη *στάθμη* του άλλου παράγοντα ή αλλιώς, αν η επίδραση του ενός παράγοντα διαφοροποιείται ανάλογα με τη *στάθμη* του άλλου παράγοντα (περισσότερο από όσο μπορεί να αποδοθεί στην τύχη). Δηλαδή, η ***αλληλεπίδραση*** εκφράζει την επίδραση του ενός παράγοντα στη συμπεριφορά του άλλου. Αν δεν υπάρχει *αλληλεπίδραση*, οι δύο παράγοντες συμπεριφέρονται ανεξάρτητα και επιδρούν στο χαρακτηριστικό που μελετάμε προσθετικά/ανεξάρτητα.

Αν $r = 1$, δηλαδή αν για κάθε *επέμβαση* παίρνουμε μόνο μια παρατήρηση (συνολικά $\alpha \cdot b$ παρατηρήσεις), δε μπορούμε να ελέγξουμε αν υπάρχει ή όχι στατιστικά σημαντική επίδραση που να οφείλεται στην *αλληλεπίδραση* των παραγόντων. Γι' αυτό, σχεδιάζουμε και εκτελούμε ένα *α x b παραγοντικό πείραμα* με μια παρατήρηση για κάθε *επέμβαση*, μόνο όταν μπορούμε να υποθέσουμε ότι οι παράγοντες επιδρούν στη *μεταβλητή απόκρισης* ανεξάρτητα.

Σε ένα *α x b παραγοντικό πείραμα με* $r > 1$ *παρατηρήσεις ανά επέμβαση*, μπορούμε να κάνουμε τους εξής ελέγχους:

- *Για την **αλληλεπίδραση** των δύο παραγόντων Α και Β:*

Κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

$H_{0\gamma}$: *οι παράγοντες Α και Β **δεν** αλληλεπιδρούν*

έναντι της εναλλακτικής,

$H_{1\gamma}$: *οι παράγοντες Α και Β αλληλεπιδρούν.*

- *Για την **κύρια επίδραση** του παράγοντα Α:*

Κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

$H_{0\alpha}$: *ο παράγοντας Α **δεν** επιδρά, ή αλλιώς, δεν υπάρχουν διαφορές μεταξύ των* α *μέσων του παράγοντα Α*

έναντι της εναλλακτικής,

$H_{1\alpha}$: *ο παράγοντας Α επιδρά, ή αλλιώς, τουλάχιστον δύο από τους* α *μέσους του παράγοντα Α διαφέρουν.*

- *Για την **κύρια επίδραση** του παράγοντα Β:*

Κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

H_{0b} : *ο παράγοντας Β **δεν** επιδρά, ή αλλιώς, δεν υπάρχουν διαφορές μεταξύ των b μέσων του παράγοντα Β*

έναντι της εναλλακτικής,

H_{1b} : *ο παράγοντας Β επιδρά, ή αλλιώς, τουλάχιστον δύο από τους b μέσους του παράγοντα Β διαφέρουν.*

Για την κατασκευή των *κριτηρίων* αυτών των ελέγχων, ακολουθούμε το γνωστό μας πλέον πρότυπο της *ανάλυσης της ολικής μεταβλητότητας των τιμών της μεταβλητής απόκρισης, σε συνιστώσες.* Έτσι, το *ολικό άθροισμα τετραγώνων*, $SSTot = \sum_{i,j,h} (y_{ijh} - \bar{y})^2$, των $v = \alpha \cdot b \cdot r$ πειραματικών δεδομένων, αναλύεται σε τέσσερις συνιστώσες, $SSTot = SSA + SSB + SS(AB) + SSE$, όπου,

$SSA = \sum_{i=1}^{\alpha} \frac{A_i^2}{b \cdot r} - \frac{G^2}{v}$, το *άθροισμα τετραγώνων του παράγοντα A* που εκφράζει τη *μεταβλητότητα μεταξύ των* α *μέσων του παράγοντα A* και έχει $a-1$ *βαθμούς ελευθερίας,*

$SSB = \sum_{j=1}^{b} \frac{B_j^2}{\alpha \cdot r} - \frac{G^2}{v}$, το *άθροισμα τετραγώνων του παράγοντα B* που εκφράζει τη *μεταβλητότητα μεταξύ των* b *μέσων του παράγοντα B* και έχει $b-1$ *βαθμούς ελευθερίας,*

$SS(AB) = \sum_{i,j} \frac{(AB)_{ij}^2}{r} - \frac{G^2}{v} - SSA - SSB$, το *άθροισμα τετραγώνων της αλληλεπίδρασης των παραγόντων A και B* που εκφράζει *τη μεταβλητότητα που απομένει αν από τη μεταβλητότητα μεταξύ των* $\alpha \cdot b$ *κελιών αφαιρέσουμε τη μεταβλητότητα που οφείλεται στον παράγοντα A και τη μεταβλητότητα που οφείλεται στον παράγοντα B*, και έχει $(\alpha-1)\cdot(b-1)$ *βαθμούς ελευθερίας,*

$SSE = SSTot - SSA - SSB - SS(AB)$, το *άθροισμα τετραγώνων του σφάλματος και εκφράζει τη μεταβλητότητα των παρατηρήσεων εντός των κελιών,* και έχει $\alpha \cdot b \cdot (r-1)$ *βαθμούς ελευθερίας.*

Με $G = \sum_{i,j,h} y_{ijh}$ συμβολίζουμε το *γενικό άθροισμα (Grand total)* των $v = \alpha \cdot b \cdot r$ παρατηρήσεων, με A_i, $i = 1,2,...,\alpha$, το άθροισμα όλων των παρατηρήσεων στην *i στάθμη* του παράγοντα *A,* με B_j, $j = 1,2,...,b$, το άθροισμα όλων των παρατηρήσεων στην j *στάθμη* του παράγοντα *B* και με $(AB)_{ij}$ το άθροισμα των *r* παρατηρήσεων που πήραμε από την *i στάθμη* του παράγοντα *A* και τη *j στάθμη* του παράγοντα *B*.

Τέλος, αποδεικνύεται ότι, $SSTot = \sum_{ijh} y_{ijh}^2 - \frac{G^2}{v}$, με $\alpha \cdot b \cdot r - 1$ *βαθμούς ελευθερίας.*

Σημείωση: Το άθροισμα $\sum_{i,j} \frac{(AB)_{ij}^2}{r} - \frac{G^2}{v}$ *εκφράζει τη μεταβλητότητα* ***μεταξύ των κελιών*** *(SSCells) και αναλύεται, όπως φαίνεται από τις παραπάνω σχέσεις, σε τρεις συνιστώσες,* $SSCells = \sum_{i,j} \frac{(AB)_{ij}^2}{r} - \frac{G^2}{v} = SSA + SSB + SS(AB)$. *Έτσι,, το SSTot, ουσιαστικά, αναλύεται στη συνιστώσα που εκφράζει τη μεταβλητότητα* ***μεταξύ των κελιών*** *(SSCells) και στη συνιστώσα που εκφράζει τη μεταβλητότητα* ***εντός των κελιών*** *(SSE), δηλαδή,* $SSTot = SSCells + SSE = SSA + SSB + SS(AB) + SSE$.

Είναι λογικό, οι έλεγχοι να γίνονται με το αντίστοιχο, κατά περίπτωση, *F* κριτήριο.

Στον *πίνακα ανάλυσης διασποράς* που ακολουθεί, φαίνονται το *F* κριτήριο και οι *απορριπτικές περιοχές* των αντίστοιχων ελέγχων.

Πίνακας Ανάλυσης Διασποράς *(για το a x b παραγοντικό πείραμα, με $r>1$ παρατηρήσεις ανά επέμβαση)*					
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***	***Περιοχή απόρριψης***
Παράγοντας Α	$\alpha-1$	*SSA*	$MSA=\dfrac{SSA}{\alpha-1}$	$F_A=\dfrac{MSA}{MSE}$	$F_A>F_{\alpha-1;ab(r-1);\alpha}$
Παράγοντας Β	$b-1$	*SSB*	$MSB=\dfrac{SSB}{b-1}$	$F_B=\dfrac{MSB}{MSE}$	$F_B>F_{b-1;ab(r-1);\alpha}$
Αλληλεπίδραση ΑΒ	$(\alpha-1)(b-1)$	*SS(AB)*	$MS(AB)=\dfrac{SS(AB)}{(\alpha-1)(b-1)}$	$F_{AB}=\dfrac{MS(AB)}{MSE}$	$F_{AB}>F_{(\alpha-1)(b-1);ab(r-1);\alpha}$
Σφάλμα	$\alpha\cdot b\cdot(r-1)$	*SSE*	$MSE=\dfrac{SSE}{\alpha\cdot b\cdot(r-1)}$		
Ολική	$\alpha\cdot b\cdot r-1$	*SSTot*			

Έλεγχοι πολλαπλών συγκρίσεων *(σε ένα a x b παραγοντικό πείραμα)*
Αν τα πειραματικά δεδομένα δεν υποστηρίζουν στατιστικά σημαντική *αλληλεπίδραση* μπορούμε να κάνουμε *ελέγχους πολλαπλών συγκρίσεων (π.χ, Tukey)* για να διερευνήσουμε ανά (στατιστικά σημαντικό) παράγοντα ποιοι *μέσοι* διαφέρουν. Αν υπάρχει στατιστικά σημαντική *αλληλεπίδραση,* νόημα έχει η διερεύνηση των διαφορών <u>μεταξύ των *μέσων των επεμβάσεων (κελιών)*</u> παρά μεταξύ των *μέσων* ανά παράγοντα (που είναι στατιστικά σημαντικός).

Ας δούμε τώρα, τι απαντήσεις προκύπτουν από τα πειραματικά δεδομένα που συγκέντρωσε ο μεταπτυχιακός φοιτητής του *Παραδείγματος-5*.

<u>***Παράδειγμα-5 (συνέχεια):***</u> Ας συμβολίσουμε τον παράγοντα *είδος φυτού* με *A* (και με A_1, A_2 τις δύο στάθμες του, *βαμβάκι* και *αγγουριά,* αντίστοιχα) και τον παράγοντα *επίπεδο θερμοκρασίας* με *B* (και με B_1, B_2 και B_3 τις τρεις στάθμες του, *70°F, 77°F* και *82°F,* αντίστοιχα*)*.

Θα ελέγξουμε, σε *επίπεδο σημαντικότητας 5%,* αν υπάρχει στατιστικά σημαντική επίδραση που να οφείλεται στον παράγοντα *A* (*είδος φυτού)*, στον παράγοντα *B* (*επίπεδο θερμοκρασίας*) αλλά και στην *αλληλεπίδραση* των δύο παραγόντων *A* και *B*. Δηλαδή, θα κάνουμε, σε *επίπεδο σημαντικότητας 5%,* τους εξής ελέγχους:
$H_{0\gamma}$: *οι παράγοντες A και B* ***δεν*** *αλληλεπιδρούν*
έναντι της,
$H_{1\gamma}$: *οι παράγοντες A και B αλληλεπιδρούν.*

$H_{0\alpha}$: *ο παράγοντας A* ***δεν*** *επιδρά ή αλλιώς,* ***δεν*** *υπάρχει διαφορά μεταξύ του μέσου στη στάθμη A_1 και του μέσου στη στάθμη A_2*
έναντι της,
$H_{1\alpha}$: *ο παράγοντας A επιδρά ή αλλιώς, ο μέσος στη στάθμη A_1 και ο μέσος στη στάθμη A_2 διαφέρουν.*

H_{0b} : *ο παράγοντας B* ***δεν*** *επιδρά ή αλλιώς,* ***δεν*** *υπάρχει διαφορά μεταξύ των μέσων στις στάθμες B_1, B_2* και *B_3*

έναντι της,

H_{1b} : *ο παράγοντας Β επιδρά ή αλλιώς, υπάρχει διαφορά μεταξύ των μέσων στις στάθμες* B_1, B_2 και B_3.

<u>*Σημείωση:*</u> *Στη συνέχεια, θα δούμε γιατί από τους τρεις ελέγχους, γράψαμε (και θα κάνουμε) πρώτο τον έλεγχο για ύπαρξη στατιστικά σημαντικής επίδρασης που να οφείλεται στην αλληλεπίδραση των δύο παραγόντων.*

Ο πίνακας που ακολουθεί, είναι ο *πίνακας ανάλυσης διασποράς* που παίρνουμε, για το συγκεκριμένο πρόβλημα, από το *στατιστικό πακέτο Statgraphics.* Στο επόμενο παράδειγμα θα υπολογίσουμε τα *αθροίσματα τετραγώνων* αναλυτικά για να δείξουμε πώς εφαρμόζουμε τους αντίστοιχους τύπους. Κατ' αρχάς, θέλουμε να εστιάσουμε στην ερμηνεία των αποτελεσμάτων των ελέγχων.

Analysis of Variance for eggs

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A: Plant	1512.3	1	1512.3	12.29	0.0018
B: Temperature	487.467	2	243.733	1.98	0.1598
INTERACTIONS					
AB	111.2	2	55.6	0.45	0.6417
RESIDUAL	2952.4	24	123.017		
TOTAL	5063.37	29			

All F-ratios are based on the residual mean square error.

Επειδή, $F_{AB} = 0.45$ και $F_{2;24;0.05} = 3.40$ (ή επειδή $P-Value = 0.6417$), η *μηδενική υπόθεση*, $H_{0\gamma}$: *οι παράγοντες Α και Β* ***δεν*** *αλληλεπιδρούν,* σε *επίπεδο σημαντικότητας 5%, δεν* απορρίπτεται, και επομένως, τα πειραματικά δεδομένα δεν υποστηρίζουν ότι υπάρχει επίδραση στο *μέσο αριθμό αυγών που γεννά η λευκή μύγα* που οφείλεται στην *αλληλεπίδραση* του *είδους φυτού* και του *επιπέδου θερμοκρασίας.*

Εφόσον δε διαπιστώθηκε στατιστικά σημαντική επίδραση που να οφείλεται στην *αλληλεπίδραση* των δύο παραγόντων, ας δούμε αν τα πειραματικά δεδομένα υποστηρίζουν ότι υπάρχει στατιστικά σημαντική επίδραση που να οφείλεται στον κάθε παράγοντα ξεχωριστά.

Επειδή, $F_A = 12.29 > F_{1;24;0.05} = 4.26$ (ή επειδή $P-Value = 0.0018$), η *μηδενική υπόθεση,* $H_{0\alpha}$: *ο παράγοντας Α* ***δεν*** *επιδρά,* σε *επίπεδο σημαντικότητας 5%,* απορρίπτεται. Επομένως, τα πειραματικά δεδομένα υποστηρίζουν ότι υπάρχει στατιστικά σημαντική επίδραση στο *μέσο αριθμό αυγών που γεννά η λευκή μύγα* που οφείλεται στο *είδος φυτού.* Δηλαδή, υπάρχει στατιστικά σημαντική διαφορά στο *μέσο* αριθμό *αυγών που γεννά η λευκή μύγα* στα δύο είδη φυτών *(βαμβάκι και αγγουριά).* Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 5%.

Στο *διάγραμμα μέσων* στο *Σχήμα-6*, φαίνεται και παραστατικά ότι ο *μέσος* αριθμός *αυγών που γεννά η λευκή μύγα* στα δύο είδη φυτών, δεν είναι ίδιος. Τα 95% *διαστήματα εμπιστοσύνης* που επίσης φαίνονται στο διάγραμμα, δίνονται από τον τύπο $\bar{y}_i \pm \frac{\sqrt{MSE}}{\sqrt{15}} \cdot t_{24;0.025}$, $i = 1, 2$ *ή* $\bar{y}_i \pm \frac{\sqrt{123.017}}{\sqrt{15}} \cdot t_{24;0.025}$, όπου, $\bar{y}_1 = 36.47$ ο *δειγματικός μέσος* των 15 παρατηρήσεων από το βαμβάκι και $\bar{y}_2 = 50.67$ ο *δειγματικός μέσος* των 15 παρατηρήσεων από την αγγουριά.

Επειδή, $F_B = 1.98$ και $F_{2;24;0.05} = 3.40$ (αλλά και επειδή $P - Value = 0.1598$), η *μηδενική υπόθεση*, H_{0b} : *ο παράγοντας Β **δεν** επιδρά,* σε *επίπεδο σημαντικότητας 5%,* δεν απορρίπτεται. Επομένως, τα πειραματικά δεδομένα δεν υποστηρίζουν ότι υπάρχει στατιστικά σημαντική επίδραση στο *μέσο αριθμό αυγών που γεννά η λευκή μύγα* που να οφείλεται στο *επίπεδο θερμοκρασίας*. Δηλαδή, δεν υπάρχει στατιστικά σημαντική διαφορά στο *μέσο* αριθμό *αυγών που γεννά η λευκή μύγα* υπό τα συγκεκριμένα τρία *επίπεδα θερμοκρασίας (70^{o}F, 77^{o}F και 82^{o}F).*

Στο *διάγραμμα μέσων* στο *Σχήμα-7*, φαίνεται και παραστατικά ότι ο *μέσος* αριθμός *αυγών που γεννά η λευκή μύγα* δε διαφέρει (στατιστικά σημαντικά) από *επίπεδο* σε *επίπεδο θερμοκρασίας*. Οι παρατηρούμενες διαφορές στους τρεις *δειγματικούς μέσους* αποδίδονται στην τύχη. Τα *95% διαστήματα εμπιστοσύνης* που επίσης φαίνονται στο διάγραμμα, δίνονται από τον τύπο $\bar{y}_j \pm \frac{\sqrt{MSE}}{\sqrt{10}} \cdot t_{24;0.025}$, $j = 1, 2, 3$ *ή* $\bar{y}_j \pm \frac{\sqrt{123.017}}{\sqrt{10}} \cdot t_{24;0.025}$, όπου, $\bar{y}_1 = 38.1$ ο *δειγματικός μέσος* των 10 παρατηρήσεων από το *επίπεδο θερμοκρασίας 70^{o}F,* $\bar{y}_2 = 44.9$ ο *δειγματικός μέσος* των 10 παρατηρήσεων από το *επίπεδο θερμοκρασίας 77^{o}F* και $\bar{y}_3 = 47.7$ ο *δειγματικός μέσος* των 10 παρατηρήσεων από το *επίπεδο θερμοκρασίας 82^{o}F.*

Στα *Σχήματα 8* και *9*, φαίνονται, τα ***διάγραμματα αλληλεπίδρασης* (interaction plot)** των δύο παραγόντων *Α* και *Β*. Ας δούμε πώς κατασκευάζεται και πώς «διαβάζεται» ένα τέτοιο διάγραμμα.

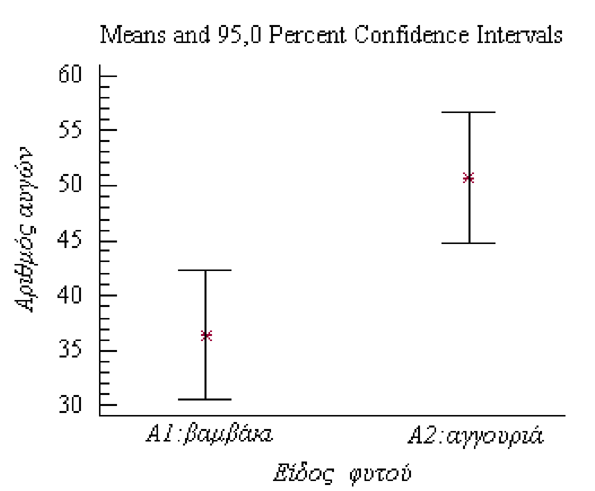
Με ένα *διάγραμμα αλληλεπίδρασης* αναπαριστούμε γραφικά τους *μέσους* των παρατηρήσεων που βρίσκονται στα $\alpha \cdot b$ *κελιά*, δηλαδή, τους *δειγματικούς μέσους* των $\alpha \cdot b$ *επεμβάσεων* ή αλλιώς τους *μέσους* των $\alpha \cdot b$ δειγμάτων, ένα για κάθε *επέμβαση*. Στο παράδειγμά μας, οι *επεμβάσεις (κελιά)* είναι 6 και οι αντίστοιχοι *δειγματικοί μέσοι* που αναπαρίστανται με 6 σημεία στα *διαγράμματα αλληλεπίδρασης* φαίνονται στον πίνακα που ακολουθεί. Στον πίνακα αυτό, επιπλέον, φαίνονται και οι *δειγματικοί μέσοι* ανά *στάθμη* του κάθε παράγοντα (που χρησιμοποιήσαμε προηγουμένως για να κατασκευάσουμε τα *διαστήματα εμπιστοσύνης*).

	B_1 *(70^{o}F)*	B_2 *(77^{o}F)*	B_3 *(82^{o}F)*	
Α1 (βαμβάκι)	*33.6*	*37.2*	*38.6*	*36.47*
Α2 (αγγουριά)	*42.6*	*52.6*	*56.8*	*50.67*
	38.1	*44.9*	*47.7*	

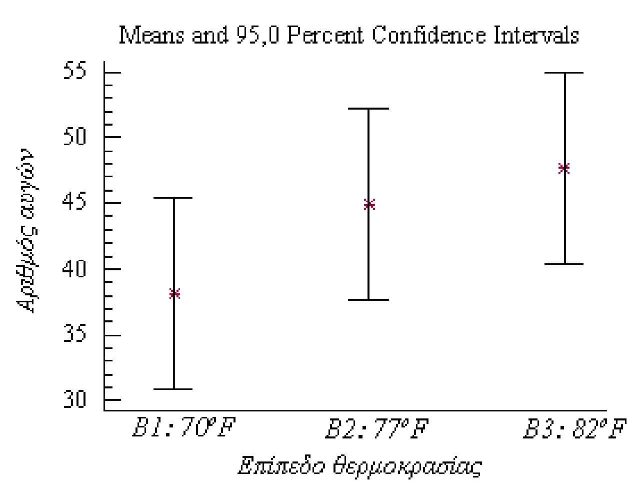
Η γραφική αναπαράσταση αυτών των *δειγματικών μέσων* των *επεμβάσεων (κελιών)* μπορεί να γίνει με δύο τρόπους. Στο *Σχήμα-8* φαίνεται ο ένας τρόπος και στο *Σχήμα-9* ο άλλος. Στο *Σχήμα-8*, στον οριζόντιο άξονα σημειώνονται οι τρεις στάθμες του παράγοντα *Β* ενώ στο *Σχήμα-9* στον οριζόντιο άξονα σημειώνονται οι δύο στάθμες του παράγοντα *Α*. Και στα δύο διαγράμματα στον κατακόρυφο άξονα αναπαρίσταται η *μεταβλητή απόκρισης (αριθμός αυγών που γεννά η λευκή μύγα)*. Ας δούμε πώς «διαβάζεται» το *διάγραμμα αλληλεπίδρασης* του *Σχήματος-8* (το αντίστοιχο διάγραμμα στο *Σχήμα-9* «διαβάζεται» ανάλογα).

Στο *Σχήμα-8*, τα δύο σημεία που αντιστοιχούν στη *στάθμη* B_1 *(70°F)* του παράγοντα *B,* αναπαριστούν τους *δειγματικούς μέσους* 33.6 και 42.6 που αντιστοιχούν στις *στάθμες* A_1 και A_2 του παράγοντα *A* και η κατακόρυφη απόστασή τους μας δείχνει πώς επιδρά ο παράγοντας *A* όταν συνδυάζεται με τη στάθμη B_1 *(70°F)* του *B*. Ομοίως, η κατακόρυφη απόσταση των δύο σημείων που αντιστοιχούν στη *στάθμη* B_2 *(77°F)* μας δείχνει πώς επιδρά ο παράγοντας *A* όταν συνδυάζεται με τη *στάθμη* B_2 και η κατακόρυφη απόσταση των δύο σημείων που αντιστοιχούν στη *στάθμη* B_3 *(82°F)* μας δείχνει πώς επιδρά ο παράγοντας *A* όταν συνδυάζεται με τη *στάθμη* B_3. Αν αυτές οι αποστάσεις είναι ίσες ή περίπου ίσες, αυτό σημαίνει ότι η επίδραση του παράγοντα *A* δεν διαφοροποιείται στις τρεις στάθμες του *B* και επομένως δεν υπάρχει επίδραση που να οφείλεται σε *αλληλεπίδραση* των δύο παραγόντων. Για να είναι πιο παραστατική η γραφική απεικόνιση της ύπαρξης ή όχι *αλληλεπίδρασης*, ενώνουμε τα σημεία που αντιστοιχούν στην ίδια *στάθμη* του παράγοντα *A* και έτσι δημιουργούμε τεθλασμένες γραμμές (όσες τα επίπεδα του παράγοντα *A*). Αν δεν υπάρχει στατιστικά σημαντική επίδραση που να οφείλεται στην *αλληλεπίδραση* των δύο παραγόντων, όλα τα αντίστοιχα ευθύγραμμα τμήματα των τεθλασμένων γραμμών πρέπει να είναι περίπου παράλληλα. Αυτό συμβαίνει, όπως βλέπουμε, στο *Σχήμα-8,* αφού στο παράδειγμά μας δε διαπιστώθηκε στατιστικά σημαντική *αλληλεπίδραση*. Η μικρή απόκλιση από την παραλληλία αποδίδεται στην τύχη. Επομένως, από το *βαμβάκι* στην *αγγουριά* φαίνεται να υπάρχει αύξηση στο *μέσο αριθμό αυγών που γεννά η λευκή πεταλούδα* που βέβαια δεν είναι ακριβώς ίδια στις τρεις *στάθμες θερμοκρασίας* αλλά αυτή η *αλληλεπίδραση* δεν κρίθηκε όπως είδαμε στατιστικά σημαντική (αποδίδεται στην τύχη).

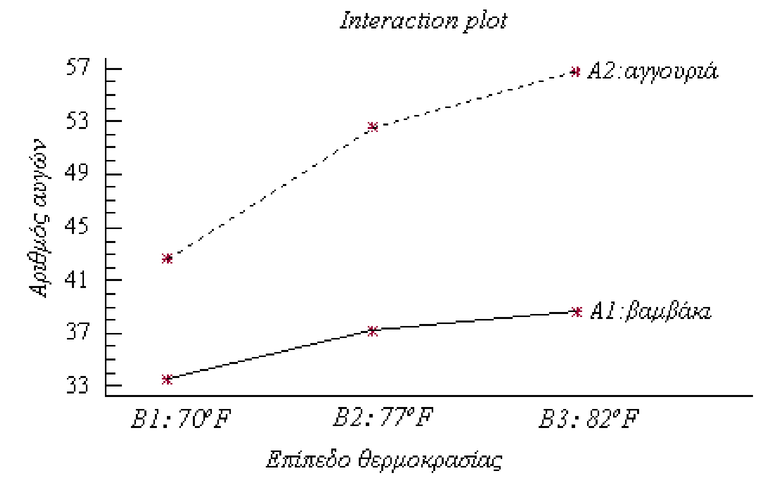
Αν διαπιστωθεί στατιστικά σημαντική επίδραση που να οφείλεται στην *αλληλεπίδραση* των δύο παραγόντων, τότε στα *διαγράμματα αλληλεπίδρασης,* οι τεθλασμένες δε θα αποτελούνται από παράλληλα ή περίπου παράλληλα ευθύγραμμα τμήματα. Κάποια από αυτά ή και όλα θα απέχουν αρκετά από το να είναι παράλληλα, μπορεί και να τέμνονται (αλλαγή κατεύθυνσης της επίδρασης). Μια τέτοια περίπτωση θα δούμε στο *Παράδειγμα-6.*



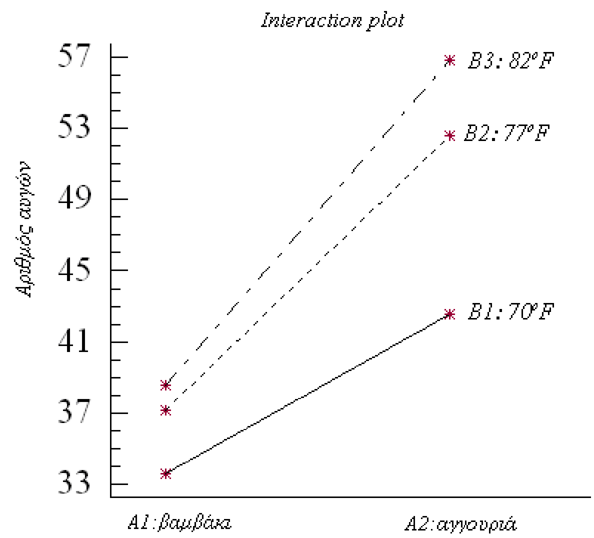
Σχήμα-6



Σχήμα-7



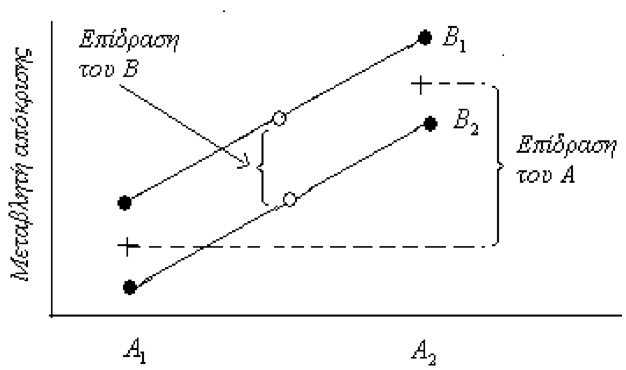
Σχήμα-8



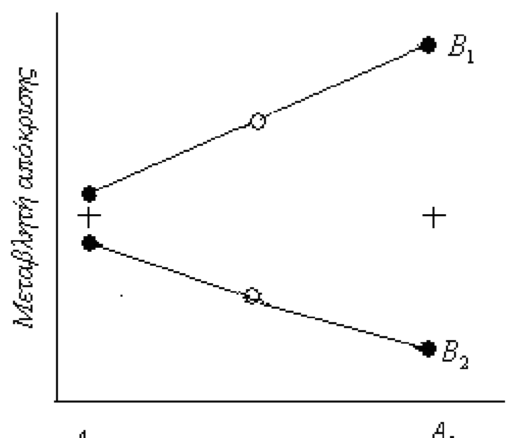
Σχήμα-9

<u>Σημείωση:</u> Η τεθλασμένη γραμμή που αντιστοιχεί στο επίπεδο A_1 (βαμβάκι) του παράγοντα A αναπαριστά γραφικά πώς μεταβάλλεται ο μέσος αριθμός αυγών που γεννά η λευκή πεταλούδα <u>στο βαμβάκι</u> στις τρεις στάθμες της θερμοκρασίας, δηλαδή, πώς επιδρά η θερμοκρασία στον αριθμό των αυγών που γεννά η λευκή μύγα όταν αυτή βρίσκεται στο βαμβάκι. Αντίστοιχα, η τεθλασμένη γραμμή που αντιστοιχεί στο επίπεδο A_2 (αγγουριά) του παράγοντα A αναπαριστά πώς επιδρά η θερμοκρασία στον αριθμό των αυγών που γεννά η λευκή μύγα όταν αυτή βρίσκεται στην αγγουριά. Δηλαδή, οι δύο τεθλασμένες γραμμές αναπαριστούν ό,τι αναπαριστά το διάγραμμα μέσων στο Σχήμα-7 (την επίδραση του παράγοντα B, δηλαδή, του επιπέδου θερμοκρασίας) αλλά ξεχωριστά για τις δύο στάθμες του παράγοντα A (δηλαδή, του είδους φυτού). Επομένως, στο διάγραμμα αλληλεπίδρασης αποτυπώνεται και το συμπέρασμα του ελέγχου για την επίδραση του παράγοντα B. Τέλος, παρατηρούμε, ότι οι δύο τεθλασμένες είναι αρκετά απομακρυσμένες μεταξύ τους. Αυτό αποτυπώνει γραφικά το συμπέρασμα του ελέγχου ότι ο παράγοντας A είναι στατιστικά σημαντικός. Δηλαδή, στο διάγραμμα αλληλεπίδρασης, αναπαρίστανται γραφικά τα αποτελέσματα και των τριών ελέγχων! Στο Σχήμα-10 δείχνουμε, για την περίπτωση δύο παραγόντων A, B με δύο στάθμες ο

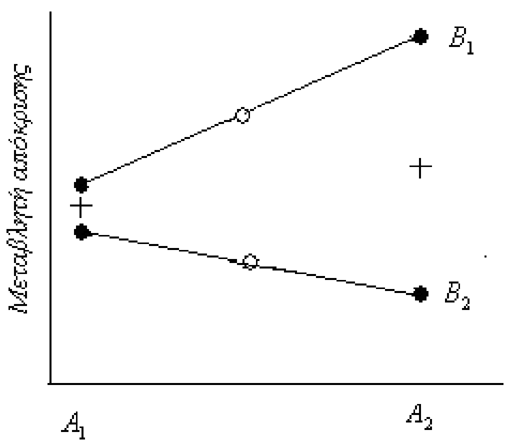
καθένας (A_1, A_2 και B_1, B_2, αντίστοιχα) αυτό ακριβώς. Δηλαδή, πώς στο διάγραμμα αλληλεπίδρασης, εκτός από την ύπαρξη αλληλεπίδρασης (μη σημαντική ή σημαντική) που φαίνεται με την παραλληλία ή όχι των ευθυγράμμων τμημάτων, μπορούμε να διακρίνουμε και την επίδραση (σημαντική ή όχι) καθενός παράγοντα. Με «+» έχουμε συμβολίσει τη μέση τιμή των παρατηρήσεων σε κάθε στάθμη του παράγοντα Α και με «ο» τη μέση τιμή των παρατηρήσεων σε κάθε στάθμη του παράγοντα Β. Τέλος, στα Σχήματα 11, 12 και 13 φαίνονται τρεις περιπτώσεις διαγραμμάτων αλληλεπίδρασης (δύο παραγόντων με δύο στάθμες ο καθένας, επίσης). Στο Σχήμα-11, φαίνεται να μην υπάρχει επίδραση του παράγοντα Α, να υπάρχει μεγάλη επίδραση του Β και μεγάλη επίδραση που οφείλεται στη αλληλεπίδραση των δύο παραγόντων. Στο Σχήμα-12, φαίνεται να υπάρχει μικρή επίδραση του παράγοντα Α (πιθανόν όχι στατιστικά σημαντική) μεγάλη επίδραση του Β και μεγάλη επίδραση που οφείλεται στη αλληλεπίδραση των δύο παραγόντων. Στο Σχήμα-13, φαίνεται να υπάρχει μεγάλη επίδραση του παράγοντα Α, να μην υπάρχει επίδραση του Β και να υπάρχει επίδραση που οφείλεται στη αλληλεπίδραση των δύο παραγόντων (με αλλαγή κατεύθυνσης της επίδρασης).



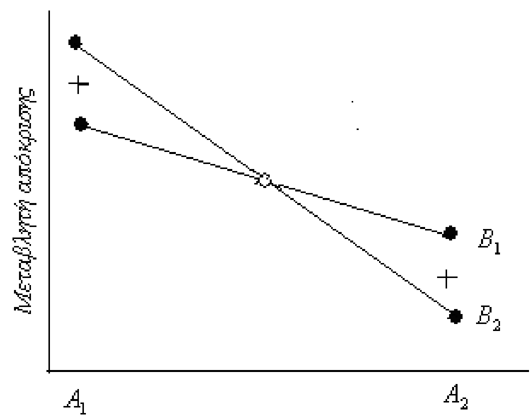
Σχήμα-10



Σχήμα-11



Σχήμα-12



Σχήμα-13

<u>Παράδειγμα-6 (αναφέρεται στο Πρόβλημα-4):</u> Ο ερευνητής ενδιαφέρεται να διερευνήσει αν και πώς οι δύο *παράγοντες*, *επίπεδο θερμοκρασίας*, έστω *Α* (με τρεις *στάθμες* A_1, A_2 και A_3, αντίστοιχα, *10°C, 20°C* και *30°C*) και *ένταση φωτισμού*, έστω

B (με τρεις *στάθμες* B_1, B_2 και B_3) επιδρούν στην ανάπτυξη του συγκεκριμένου είδους χλωρόφυτων.

Θα κάνουμε, σε *επίπεδο σημαντικότητας 5%*, τους παρακάτω ελέγχους.
$H_{0\gamma}$: *οι παράγοντες Α και Β* ***δεν*** *αλληλεπιδρούν*
έναντι της,
$H_{1\gamma}$: *οι παράγοντες Α και Β αλληλεπιδρούν.*

$H_{0\alpha}$: *ο παράγοντας Α* ***δεν*** *επιδρά*
έναντι της,
$H_{1\alpha}$: *ο παράγοντας Α επιδρά*

H_{0b} : *ο παράγοντας Β* ***δεν*** *επιδρά*
έναντι της,
H_{1b} : *ο παράγοντας Β επιδρά*

Θα υπολογίσουμε τα *αθροίσματα τετραγώνων SSA*, *SSB*, *SS(AB)*, *SSE* και *SSTot*, και θα κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς.*

		Επίπεδο Θερμοκρασίας (Παράγοντας Α)			
		A_1 *:10ºC*	A_2 *:20ºC*	A_3 *:30ºC*	B_j
Ένταση φωτισμού (Παράγοντας Β)	B_1	310	480	400	*1190*
	B_2	320	540	340	*1200*
	B_3	410	550	250	*1210*
	A_i	*1040*	*1570*	*990*	$G = 3600$

$$SSTot = \sum_{ijh} y_{ijh}^2 - \frac{G^2}{\nu} = 150^2 + 160^2 + \ldots + 110^2 + 140^2 - \frac{3600^2}{18} = 46600$$

$$SSA = \sum_{i=1}^{\alpha} \frac{A_i^2}{b \cdot r} - \frac{G^2}{\nu} = \frac{1040^2 + 1570^2 + 990^2}{6} - \frac{3600^2}{18} = 34433.3$$

$$SSB = \sum_{j=1}^{b} \frac{B_j^2}{\alpha \cdot r} - \frac{G^2}{\nu} = \frac{1190^2 + 1200^2 + 1210^2}{6} - \frac{3600^2}{18} = 33.3333$$

$$SS(AB) = \sum_{i,j} \frac{(AB)_{ij}^2}{r} - \frac{G^2}{\nu} - SSA - SSB =$$

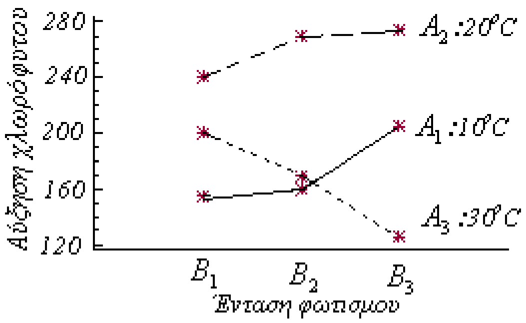
$$= \frac{310^2 + \ldots + 250^2}{2} - \frac{3600^2}{18} - 34433.3 - 33.3333 = 10133.3$$

$$SSE = SSTot - SSA - SSB - SS(AB) = 2000$$

Μπορούμε πλέον να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*:

Πίνακας Ανάλυσης Διασποράς				
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***
Παράγοντας Α *(Επίπεδο θερμοκρασίας)*	2	*34433.3*	$MSA = \frac{34433.3}{2} = 17216.7$	$F_A = \frac{17216.7}{222.222} = 77.48$
Παράγοντας Β *(Ένταση φωτισμού)*	*2*	*33.3*	$MSB = \frac{33.3}{2} = 16.6$	$F_B = \frac{16.6}{222.222} = 0.07$
Αλληλεπίδραση ΑΒ	4	*10133.3*	$MS(AB) = \frac{10133.3}{4} = 2533.3$	$F_{AB} = \frac{2533.3}{222.222} = 11.40$
Σφάλμα	9	*2000*	$MSE = \frac{2000}{9} = 222.222$	
Ολική	17	*46600*		

Επειδή, $F_{AB} = 11.40 > F_{4;9;0.05} = 3.63$, η *μηδενική υπόθεση,* $H_{0\gamma}$: *οι παράγοντες Α και Β **δεν** αλληλεπιδρούν,* σε *επίπεδο σημαντικότητας 5%,* απορρίπτεται . Επομένως, τα πειραματικά δεδομένα υποστηρίζουν ότι, σε *επίπεδο σημαντικότητας 5%,* υπάρχει στατιστικά σημαντική επίδραση στην *ανάπτυξη του συγκεκριμένου είδους χλωρόφυτων* που οφείλεται στην *αλληλεπίδραση επιπέδου θερμοκρασίας* και *έντασης φωτισμού.* Έτσι, για τον ερευνητή, νόημα πλέον έχει η διερεύνηση της συμπεριφοράς της *αύξησης του χλωρόφυτου* στους διάφορους συνδυασμούς *επίπεδο θερμοκρασίας-ένταση φωτισμού* και όχι του κάθε παράγοντα ξεχωριστά. Παρόλα αυτά, παρατηρείστε ότι η επίδραση του παράγοντα *επίπεδο θερμοκρασίας* ($F_A = 77.48 > F_{2;9;0.05} = 4.26$) είναι στατιστικά σημαντική ενώ του παράγοντα *ένταση φωτισμού*, δεν είναι στατιστικά σημαντική ($F_B = 0.07$ ενώ $F_{2;9;0.05} = 4.26$). Δείτε στο *διάγραμμα αλληλεπίδρασης* στο *Σχήμα-14,* πώς διαφοροποιείται η επίδραση του *επιπέδου θερμοκρασίας* στη *μέση αύξηση του χλωρόφυτου* στις τρεις στάθμες *έντασης φωτισμού.*



Σχήμα-14

Υποθέσεις/παραδοχές στο α x b παραγοντικό πείραμα με $r > 1$ ***παρατηρήσεις ανά επέμβαση***

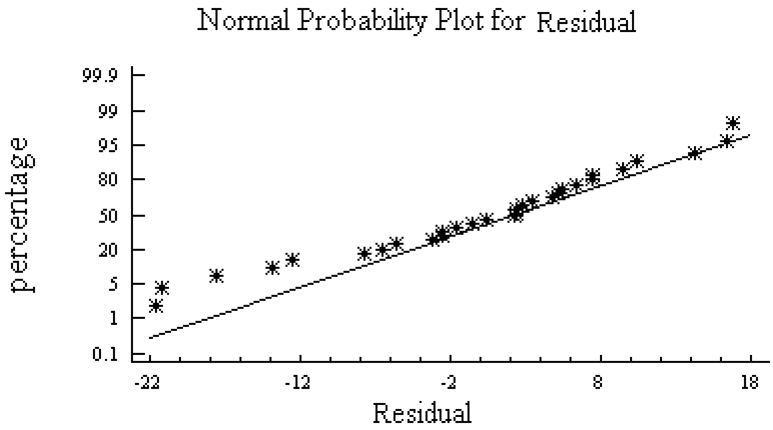
Στο *α x b παραγοντικό πείραμα με* $r > 1$ *παρατηρήσεις ανά επέμβαση*, που μελετήσαμε, οι παρατηρήσεις συγκεντρώνονται σύμφωνα με το ***εντελώς τυχαιοποιημένο σχέδιο***. Οι υποθέσεις/παραδοχές που κάνουμε σε ένα τέτοιο πείραμα, είναι οι εξής:

1. Τα $\alpha \cdot b$ δείγματα είναι *ανεξάρτητα τυχαία* δείγματα (όλες οι παρατηρήσεις y_{ijh} είναι μεταξύ τους *ανεξάρτητες*)
2. Κάθε δείγμα προέρχεται από *κανονικό* πληθυσμό
3. Οι $\alpha \cdot b$ πληθυσμοί έχουν *κοινή διασπορά,* σ^2 *(ομοσκεδαστικότητα)*

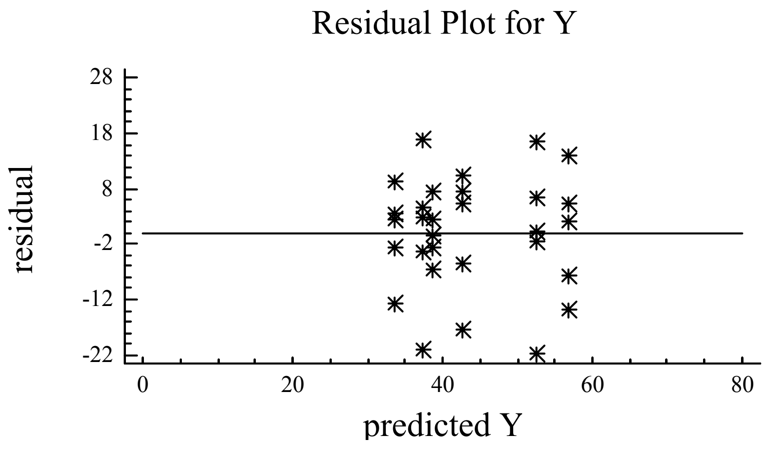
Αν ε_{ijh} η απόκλιση της Y_{ijh} από τη *μέση τιμή* μ_{ij}, του πληθυσμού *ij*, τότε οι παραπάνω υποθέσεις είναι ισοδύναμες με τις εξής:

1. Τα *υπόλοιπα* ε_{ijh} είναι μεταξύ τους *ανεξάρτητα*
2. $\varepsilon_{ijh} \sim N(0, \sigma^2)$

Για τον έλεγχο αυτών των υποθέσεων/παραδοχών, εφαρμόζονται οι μέθοδοι/τεχνικές που έχουμε αναφέρει και εφαρμόσει στα προηγούμενα. Για παράδειγμα, από το *κανονικό διάγραμμα πιθανοτήτων* των *υπολοίπων* και από το *διάγραμμα υπολοίπων* του *Παραδείγματος-5* (*Σχήμα-15* και *Σχήμα-16*, αντίστοιχα) φαίνεται να ικανοποιούνται οι υποθέσεις τις *κανονικότητας* και της *ομοσκεδαστικότητας* κάτι το οποίο βέβαια πρέπει να ελεγχθεί και με κατάλληλους στατιστικούς ελέγχους.



Σχήμα-15



Σχήμα-16

α x b παραγοντικό πείραμα, με μια παρατήρηση ανά επέμβαση ($r=1$ ***)***

Επιλέγουμε να σχεδιάσουμε και να εκτελέσουμε ένα *α x b παραγοντικό πείραμα με μια παρατήρηση ανά επέμβαση (* $r=1$ *)*, όταν μπορούμε να υποθέσουμε/να κάνουμε την παραδοχή ότι οι δύο παράγοντες επιδρούν στη *μεταβλητή απόκρισης* ανεξάρτητα/προσθετικά, δηλαδή, ότι δεν υπάρχει επίδραση στη *μεταβλητή απόκρισης* που να οφείλεται στην *αλληλεπίδραση* των δύο παραγόντων. Έτσι, κάνουμε τους ελέγχους:

- *Για την **επίδραση** του παράγοντα Α:*

Κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

$H_{0\alpha}$ *: ο παράγοντας Α **δεν** επιδρά, ή αλλιώς, δεν υπάρχουν διαφορές μεταξύ των* α *μέσων του παράγοντα Α*

έναντι της εναλλακτικής,

$H_{1\alpha}$ *: ο παράγοντας Α επιδρά, ή αλλιώς, τουλάχιστον δύο από τους* α *μέσους του παράγοντα Α διαφέρουν.*

- *Για την **επίδραση** του παράγοντα Β:*

Κάνουμε τον έλεγχο της μηδενικής υπόθεσης,

H_{0b} *: ο παράγοντας Β **δεν** επιδρά, ή αλλιώς, δεν υπάρχουν διαφορές μεταξύ των b μέσων του παράγοντα Β*

έναντι της εναλλακτικής,

H_{1b} *: ο παράγοντας Β επιδρά, ή αλλιώς, τουλάχιστον δύο από τους b μέσους του παράγοντα Β διαφέρουν.*

Για την κατασκευή των *κριτηρίων* αυτών των ελέγχων, αναλύουμε την *ολική μεταβλητότητα* των τιμών της *μεταβλητής απόκρισης,* σε τρεις συνιστώσες. Έτσι έχουμε, $SSTot = SSA + SSB + SSE$, όπου,

$SSTot = \sum_{ij}(y_{ij}-\bar{y})^2 = \sum_{ij} y_{ij}^2 - \frac{G^2}{\nu}$, το <u>*ολικό άθροισμα τετραγώνων*</u> που εκφράζει την *ολική μεταβλητότητα* των $\nu = \alpha \cdot b$ παρατηρήσεων και έχει $\alpha \cdot b - 1$ *βαθμούς ελευθερίας,*

$SSA = \sum_{i=1}^{\alpha}\frac{A_i^2}{b} - \frac{G^2}{\nu}$, το <u>*άθροισμα τετραγώνων του παράγοντα Α*</u> που εκφράζει τη *μεταβλητότητα μεταξύ των μέσων των παρατηρήσεων στις* α *στάθμες του παράγοντα Α* και έχει $\alpha - 1$ *βαθμούς ελευθερίας,*

$SSB = \sum_{j=1}^{b}\frac{B_j^2}{\alpha} - \frac{G^2}{\nu}$, το <u>*άθροισμα τετραγώνων του παράγοντα Β*</u> που εκφράζει τη *μεταβλητότητα μεταξύ των μέσων των παρατηρήσεων στις b στάθμες του παράγοντα Β* και έχει $b-1$ *βαθμούς ελευθερίας,*

$SSE = SSTot - SSA - SSB$, το <u>*άθροισμα τετραγώνων του σφάλματος*</u> και έχει $(\alpha-1)\cdot(b-1)$ *βαθμούς ελευθερίας.*

Με $G = \sum_{ij} y_{ij}$ συμβολίζουμε το *γενικό άθροισμα (Grand total)* των παρατηρήσεων, δηλαδή, το άθροισμα όλων των $\nu = \alpha \cdot b$ παρατηρήσεων, με A_i, $i = 1,2,...,\alpha$, συμβολίζουμε το άθροισμα όλων των παρατηρήσεων της *i στάθμης* του παράγοντα *Α* και με B_j, $j = 1,2,...,b$, το άθροισμα όλων των παρατηρήσεων της j *στάθμης* του παράγοντα *Β*.

Είναι λογικό, οι έλεγχοι να γίνονται, και στην περίπτωση αυτή, με το αντίστοιχο F κριτήριο. Στον *πίνακα ανάλυσης διασποράς* που ακολουθεί, φαίνονται το F κριτήριο και οι *απορριπτικές περιοχές* των αντίστοιχων ελέγχων.

Πίνακας Ανάλυσης Διασποράς *(για το α x b παραγοντικό πείραμα, με μια παρατήρηση ανά επέμβαση, $r=1$)*					
Πηγή μεταβλητότητας	**Β.Ε.**	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***	***Περιοχή απόρριψης***
Παράγοντας Α	$\alpha-1$	SSA	$MSA=\dfrac{SSA}{\alpha-1}$	$F_A=\dfrac{MSA}{MSE}$	$F_A>F_{\alpha-1;(\alpha-1)(b-1);\alpha}$
Παράγοντας Β	$b-1$	SSB	$MSB=\dfrac{SSB}{b-1}$	$F_B=\dfrac{MSB}{MSE}$	$F_B>F_{b-1;(\alpha-1)(b-1);\alpha}$
Σφάλμα	$(\alpha-1)\cdot(b-1)$	SSE	$MSE=\dfrac{SSE}{(\alpha-1)(b-1)}$		
Ολική	$\alpha\cdot b-1$	$SSTot$			

Ας δούμε ένα παράδειγμα τέτοιου *παραγοντικού πειράματος*.

<u>Παράδειγμα-7:</u> *Ένας ερευνητής για να διερευνήσει αν επηρεάζεται το αποτέλεσμα μιας χημικής αντίδρασης από το είδος του καταλύτη και από τη θερμοκρασία, καθόρισε τέσσερις στάθμες καταλύτη (A_1, A_2, A_3, A_4) και τρεις στάθμες θερμοκρασίας (B_1, B_2, B_3)* και για κάθε συνδυασμό *καταλύτη-θερμοκρασίας,* εκτέλεσε το πείραμα μία φορά. Τα αποτελέσματα των πειραμάτων δίνονται στον πίνακα που ακολουθεί *(σε ppm)*.

		Καταλύτης (Παράγοντας Α)			
		A_1	A_2	A_3	A_4
Θερμοκρασία (Παράγοντας Β)	B_1	53	59	58	50
	B_2	57	65	62	60
	B_3	52	62	54	52

Ο ερευνητής έκανε την παραδοχή ότι οι δύο παράγοντες δεν αλληλεπιδρούν. Έτσι, θα κάνουμε, σε *επίπεδο σημαντικότητας 5%,* τον έλεγχο της *μηδενικής υπόθεσης,*

$H_{0\alpha}$: *ο παράγοντας Α (καταλύτης)* ***δεν*** *επιδρά,*

έναντι της εναλλακτικής,

$H_{1\alpha}$: *ο παράγοντας Α επιδρά.*

και τον έλεγχο της *μηδενικής υπόθεσης,*

H_{0b} : *ο παράγοντας Β (θερμοκρασία)* ***δεν*** *επιδρά*

έναντι της *εναλλακτικής,*

H_{1b} : *ο παράγοντας Β επιδρά.*

Θα υπολογίσουμε τα *αθροίσματα τετραγώνων* SSA, SSB, SSE και $SSTot$, και θα κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς.*

		Καταλύτης (Παράγοντας A)				
		A_1	A_2	A_3	A_4	B_j
Θερμοκρασία (Παράγοντας B)	B_1	53	59	58	50	220
	B_2	57	65	62	60	244
	B_3	52	62	54	52	220
	A_i	162	186	174	162	$G = 684$

$$SSTot = \sum_{ij} y_{ij}^2 - \frac{G^2}{\nu} = 53^2 + 57^2 + ... + 60^2 + 52^2 - \frac{684^2}{12} = 252$$

$$SSA = \sum_{i=1}^{4} \frac{A_i^2}{3} - \frac{G^2}{\nu} = \frac{162^2 + 186^2 + 174^2 + 162^2}{3} - \frac{684^2}{12} = 132$$

$$SSB = \sum_{j=1}^{3} \frac{B_j^2}{4} - \frac{G^2}{\nu} = \frac{220^2 + 244^2 + 220^2}{4} - \frac{684^2}{12} = 96$$

$$SSE = SSTot - SSA - SSB = 252 - 132 - 96 = 24$$

Μπορούμε πλέον να κατασκευάσουμε τον *πίνακα ανάλυσης διασποράς*:

Πίνακας Ανάλυσης Διασποράς				
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***
Παράγοντας A *(Καταλύτης)*	3	*132*	$MSA = \frac{132}{3} = 44$	$F_A = \frac{44}{4} = 11$
Παράγοντας B *(Θερμοκρασία)*	*2*	*96*	$MSB = \frac{96}{2} = 48$	$F_B = \frac{48}{4} = 12$
Σφάλμα	6	*24*	$MSE = \frac{24}{6} = 4$	
Ολική	11	*252*		

Επειδή, $F_A = 11 > F_{3;6;0.05} = 4.76$, η *μηδενική υπόθεση,* $H_{0\alpha}$ *: ο παράγοντας A* ***δεν*** *επιδρά, σε επίπεδο σημαντικότητας 5%,* απορρίπτεται. Επομένως, τα πειραματικά δεδομένα δίνουν στατιστικά σημαντικές αποδείξεις ότι το *είδος καταλύτη* επηρεάζει το αποτέλεσμα του πειράματος. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 5%. Επίσης, επειδή, $F_B = 12 > F_{2;6;0.05} = 5.14$, η *μηδενική υπόθεση,* H_{0b} *: ο παράγοντας B* ***δεν*** *επιδρά, σε επίπεδο σημαντικότητας 5%,* απορρίπτεται. Επομένως, τα πειραματικά δεδομένα δίνουν στατιστικά σημαντικές αποδείξεις ότι το *επίπεδο θερμοκρασίας* επηρεάζει το αποτέλεσμα του πειράματος. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος, είναι το πολύ 5%.

Υποθέσεις/παραδοχές στο a x b παραγοντικό πείραμα με μια παρατήρηση ανά επέμβαση

Στο *a x b παραγοντικό πείραμα με μία παρατήρηση ανά επέμβαση*, που μελετήσαμε, οι παρατηρήσεις συγκεντρώνονται σύμφωνα με το ***εντελώς τυχαιοποιημένο σχέδιο***. Οι υποθέσεις/παραδοχές που κάνουμε σε ένα τέτοιο πείραμα, είναι οι εξής:

1. Δεν υπάρχει *αλληλεπίδραση* των δύο παραγόντων, δηλαδή, οι δύο παράγοντες επιδρούν προσθετικά/ανεξάρτητα.

2. Τα $\alpha \cdot b$ δείγματα είναι *ανεξάρτητα τυχαία* δείγματα (όλες οι παρατηρήσεις y_{ij}, είναι μεταξύ τους *ανεξάρτητες*)
3. Κάθε δείγμα προέρχεται από *κανονικό* πληθυσμό
4. Οι $\alpha \cdot b$ πληθυσμοί έχουν *κοινή διασπορά*, σ^2 *(ομοσκεδαστικότητα)*

Οι υποθέσεις, (2)-(4) είναι ισοδύναμες με τις εξής:

1. Τα *υπόλοιπα* ε_{ij} είναι μεταξύ τους *ανεξάρτητα*
2. $\varepsilon_{ij} \sim N(0,\sigma^2)$

Για τον έλεγχο της υπόθεσης της *προσθετικότητας*, υπάρχει σχετικός στατιστικός έλεγχος και διάφορες εμπειρικές τεχνικές διάγνωσης, όμως, δε θα επεκταθούμε περισσότερο. Οι υπόλοιπες υποθέσεις ελέγχονται όπως στα προηγούμενα.

Ερώτηση: *Πώς σχολιάζετε, συγκριτικά, τον έλεγχο ανάλυσης διασποράς στο α x b παραγοντικό πείραμα με μία παρατήρηση ανά επέμβαση (και εντελώς τυχαιοποιημένο) και στο σχέδιο τυχαιοποιημένων πλήρων ομάδων;*

Προβλήματα και Ασκήσεις

Για κάθε πρόβλημα που ακολουθεί, εκτός των ερωτημάτων που διατυπώνονται, να γίνουν (με τη βοήθεια κάποιου στατιστικού πακέτου) και τα εξής:
α) Έλεγχος των αναγκαίων, για την εφαρμογή ελέγχου ανάλυσης διασποράς, υποθέσεων/παραδοχών.
β) Έλεγχοι πολλαπλών συγκρίσεων (στις περιπτώσεις που διαπιστώνονται στατιστικά σημαντικές επιδράσεις) και σχολιασμός των συμπερασμάτων που προκύπτουν από αυτούς.
γ) Κατάλληλα διαγράμματα για τη γραφική αναπαράσταση των συμπερασμάτων.
δ) Κατάλληλος μη παραμετρικός έλεγχος, όταν οι αναγκαίες για τον έλεγχο ανάλυσης διασποράς υποθέσεις/παραδοχές δεν ικανοποιούνται.

1. Ένας ερευνητής προκειμένου να συγκρίνει τρία σιτηρέσια εκτροφής κοτόπουλων (Σ1, Σ2 και Σ3, αντίστοιχα), σχεδίασε και εκτέλεσε το εξής πείραμα. Επέλεξε 15 νεογέννητα κοτόπουλα και με μια τυχαία διαδικασία αντιστοίχησε σε 5 από αυτά το σιτηρέσιο Σ1, σε 5 άλλα το σιτηρέσιο Σ2 και σε 5 άλλα το σιτηρέσιο Σ3. Δημιούργησε έτσι τρεις ομάδες των πέντε κοτόπουλων η κάθε μία. Αφού χορήγησε στα κοτόπουλα κάθε ομάδας το αντίστοιχο σιτηρέσιο (για συγκεκριμένο χρονικό διάστημα από τη γέννησή τους), μέτρησε το (μικτό) βάρος τους *(σε kgr)*. Οι μετρήσεις που πήρε δίνονται στον πίνακα που ακολουθεί.

		Μικτό Βάρος *(σε kgr)*				
Σιτηρέσιο	Σ1	2.65	2.31	2.61	2.09	2.39
	Σ2	1.83	2.46	2.54	1.72	1.98
	Σ3	2.50	2.24	2.72	2.31	2.65

   α) Τι τύπου *πειραματικό σχέδιο* επέλεξε να εφαρμόσει ο ερευνητής; β) Με βάση αυτά τα πειραματικά δεδομένα και σε *επίπεδο σημαντικότητας 5%*, υπάρχουν στατιστικά σημαντικές διαφορές στη μέση αύξηση του βάρους των κοτόπουλων που να οφείλονται στα τρία σιτηρέσια; γ) Ποιες υποθέσεις χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (β);

2. Ένας φοιτητής του Τμήματος Επιστήμης και Τεχνολογίας Τροφίμων του Γ.Π.Α., στο πλαίσιο της πτυχιακής του εργασίας, συνέκρινε τις ποσότητες χοληστερίνης που περιέχουν τέσσερα διαφορετικά είδη τροφίμων διαίτης, Α1, Α2, Α3 και Α4, αντίστοιχα. Από κάθε είδος πήρε ένα τυχαίο δείγμα μεγέθους 3 και μέτρησε την ποσότητα χοληστερίνης *(σε milligrams ανά 170gr)*. Οι μετρήσεις που πήρε δίνονται στον πίνακα που ακολουθεί.

		Ποσότητα Χοληστερίνης *(σε milligrams/170gr)*		
Είδος τροφίμου διαίτης	Α1	3.6	4.1	4.0
	Α2	3.1	3.2	3.9
	Α3	3.2	3.5	3.5
	Α4	3.5	3.8	3.8

   α) Σε *επίπεδο σημαντικότητας 5%,* υποστηρίζουν αυτά τα δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων ποσοτήτων χοληστερίνης στα τέσσερα είδη τροφίμων; β) Εξηγείστε τι έλεγχο, γιατί και με ποιες παραδοχές κάνατε για να απαντήσετε στο ερώτημα (α).

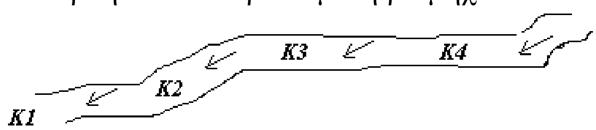
3. Σε ένα πείραμα για τη διερεύνηση της επίδρασης της πρωινής διατροφής στη συγκέντρωση της προσοχής των μαθητών Α΄ Δημοτικού κατά τα πρώτα είκοσι

λεπτά της πρώτης πρωινής ώρας μαθημάτων, επελέγησαν 15 μαθητές της Α΄ Δημοτικού και με μια τυχαία διαδικασία, σε 5 μαθητές δε δόθηκε πρωινό, σε άλλους 5 δόθηκε ελαφρύ πρωινό και σε άλλους, επίσης, 5 δόθηκε πλήρες πρωινό. Στον πίνακα που ακολουθεί, φαίνεται ο χρόνος/διάρκεια συγκέντρωσης της προσοχής (*σε min*), καθενός από τους 15 μαθητές (κατά τα πρώτα είκοσι λεπτά της πρώτης πρωινής ώρας μαθημάτων).

Χρόνοι συγκέντρωσης *(σε min)* ανά κατηγορία πρωινής διατροφής		
Όχι πρωινό	Ελαφρύ πρωινό	Πλήρες πρωινό
8	14	10
7	16	12
9	12	16
13	17	15
10	11	12

α) Τι τύπου *πειραματικό σχέδιο* επελέγη; β) Με βάση αυτά τα πειραματικά δεδομένα και σε *επίπεδο σημαντικότητας 5%*, υπάρχει στατιστικά σημαντική επίδραση στη διάρκεια συγκέντρωσης της προσοχής των μαθητών Α΄ Δημοτικού κατά τα πρώτα είκοσι λεπτά της πρώτης πρωινής ώρας μαθημάτων, που να οφείλεται στην πρωινή διατροφή; γ) Ποιες υποθέσεις χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (β);

4. Μια ομάδα ερευνητών μελέτησε τη μόλυνση των νερών ενός ποταμού από βιομηχανικά απόβλητα. Ως μέρος αυτής της μελέτης, συνέκρινε τη συγκέντρωση διαλυμένου οξυγόνου στα νερά του ποταμού σε 4 περιοχές της κοίτης του, Κ1, Κ2, Κ3 και Κ4. Η περιοχή Κ1 βρίσκεται στις εκβολές του ποταμού ενώ κοντά στην Κ3 ρίχνονται απόβλητα από παρακείμενη βιομηχανία.



Από κάθε περιοχή οι ερευνητές πήραν, με βάση ένα σχέδιο τυχαίας δειγματοληψίας, έξι δείγματα νερού. Στον πίνακα που ακολουθεί δίνονται οι συγκεντρώσεις διαλυμένου οξυγόνου *(σε ppm)* στις τέσσερις περιοχές του ποταμού (πέντε από τα δείγματα νερού που ελήφθησαν, χάθηκαν).

		Συγκέντρωση διαλυμένου οξυγόνου *(σε ppm)*					
Περιοχή	Κ1	6.1	6.3	6.1	6.0		
	Κ2	6.0	6.2	6.1	5.8		
	Κ3	4.8	4.3	5.0	4.7	5.1	
	Κ4	6.3	6.6	6.4	6.4	6.5	6.3

α) Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν αυτά τα δεδομένα ότι μεταξύ των τεσσάρων περιοχών υπάρχουν στατιστικά σημαντικές διαφορές στη μέση συγκέντρωση διαλυμένου οξυγόνου; β) Εξηγείστε τι έλεγχο, γιατί και με ποιες παραδοχές κάνατε για να απαντήσετε στο ερώτημα (α).

5. Στο πλαίσιο μιας περιβαλλοντικής μελέτης, μια ερευνητική ομάδα του Τμήματος Γεωπονικής Βιοτεχνολογίας του Γ.Π.Α., μελέτησε την ανάπτυξη της βλάστησης σε τέσσερις ελώδεις περιοχές της χώρας στις οποίες δεν είχε γίνει ανθρώπινη παρέμβαση (καλλιέργειες, προσχώσεις, κ.τλ.). Ειδικότερα, και ως μέρος της μελέτης, μια προκαθορισμένη ημέρα του Μαΐου η ερευνητική ομάδα επέλεξε τυχαία 6 φυτά ενός συγκεκριμένου είδους από κάθε περιοχή και μέτρησε το μήκος των φύλλων κάθε φυτού *(σε cm)*. Στον πίνακα που ακολουθεί δίνεται το

μέσο μήκος είκοσι τυχαία επιλεγμένων φύλλων κάθε φυτού που επελέγη για να χρησιμοποιηθεί στην έρευνα.

		Μέσο μήκος 20 φύλλων κάθε φυτού *(σε cm)*					
Περιοχή	1	5.7	6.3	6.1	6.0	5.8	6.2
	2	6.2	5.3	5.7	6.0	5.2	5.5
	3	5.4	5.0	6.0	5.6	4.9	5.2
	4	3.7	3.2	3.9	4.0	3.5	3.6

α) Σε *επίπεδο σημαντικότητας 5%*, δίνουν αυτά τα δεδομένα αποδείξεις ότι υπάρχουν στατιστικά σημαντικές διαφορές στο μέσο μήκος των φύλλων των φυτών του συγκεκριμένου είδους μεταξύ των τεσσάρων περιοχών; β) Εξηγείστε τι έλεγχο, γιατί και με ποιες παραδοχές κάνατε για να απαντήσετε στο ερώτημα (α); γ) Ποια παραδοχή από αυτές που χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (α) είναι εύλογο να θεωρήσετε ότι πράγματι ισχύει (κατά προσέγγιση).

6. Να συμπληρώσετε τον πίνακα ANOVA που ακολουθεί.

Πηγή μεταβλητότητας	Β.Ε.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Επεμβάσεις	4			6.40
Σφάλμα			10.60	
Ολική		377.36		

Ο παράγοντας του οποίου η επίδραση μελετάται, είναι σε *επίπεδο σημαντικότητας 5%* στατιστικά σημαντικός;

7. Ο Οργανισμός Ελληνικών Γεωργικών Ασφαλίσεων (ΕΛ.Γ.Α.) θέλει να ελέγξει αν οι εκτιμήσεις των ζημιών στις γεωργικές καλλιέργειες που κάνουν οι τρεις γεωπόνοι (Γ1, Γ2 και Γ3, αντίστοιχα) που χρησιμοποιεί ως εκτιμητές στο νομό Λάρισας, διαφέρουν μεταξύ τους. Για το σκοπό αυτό, επέλεξε από το νομό Λάρισας τέσσερις διαφορετικές καλλιέργειες (Κ1, Κ2, Κ3 και Κ4, αντίστοιχα) που υπέστησαν ζημιά από δυσμενή καιρικά φαινόμενα και ζήτησε από τους τρεις εκτιμητές να εκτιμήσουν, ο καθένας ανεξάρτητα από τους άλλους, τη ζημιά σε κάθε μία από τις τέσσερις καλλιέργειες. Οι εκτιμήσεις που έκαναν οι τρεις γεωπόνοι/εκτιμητές *(σε €/στρέμμα)* δίνονται στον πίνακα που ακολουθεί.

		Καλλιέργεια			
		Κ1	Κ2	Κ3	Κ4
Εκτιμητής	Γ1	169	182	185	173
	Γ2	184	193	191	189
	Γ3	173	187	180	190

α) Εξηγείστε τι τύπου *πειραματικό σχέδιο* επέλεξε να εφαρμόσει ο ΕΛ.Γ.Α. β) Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στις μέσες εκτιμήσεις των ζημιών μεταξύ των τριών γεωπόνων/εκτιμητών; γ) Από το αποτέλεσμα της ανάλυσης που κάνατε, κρίνετε ότι επιβεβαιώθηκε η σχεδιαστική επιλογή του ΕΛ.Γ.Α. ή μήπως όχι;

8. Ο πίνακας ANOVA που ακολουθεί προέκυψε από τις παρατηρήσεις που πήραμε από ένα πείραμα που έγινε με το *σχέδιο τυχαιοποιημένων πλήρων ομάδων.*

Πηγή μεταβλητότητας	Β.Ε.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Επεμβάσεις	4	14.2		
Ομάδες		18.9		
Σφάλμα	24			
Ολική	34	41.9		

α) Πόσες ομάδες δημιουργήθηκαν; β) Πόσες παρατηρήσεις ελήφθησαν για κάθε στάθμη των επεμβάσεων; γ) Πόσες παρατηρήσεις ελήφθησαν από κάθε ομάδα; δ) Να *συμπληρώσετε* τον πίνακα. ε) Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν αυτά τα δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων των επεμβάσεων; Μεταξύ των μέσων των ομάδων;]

9. Ένας ερευνητής προκειμένου να συγκρίνει τέσσερις χημικές ουσίες (Α1, Α2, Α3 και Α4, αντίστοιχα) που χρησιμοποιούνται για τη μείωση της συγκράτησης/απορρόφησης νερού από νήματα, σχεδίασε το εξής πείραμα. Από καθένα από τρία διαφορετικά είδη νήματος, έκοψε τυχαία ένα κομμάτι ορισμένου μήκους, και στη συνέχεια, κάθε κομμάτι το έκοψε σε τέσσερα ίσα κομμάτια. Δημιούργησε έτσι, 3 ομάδες των τεσσάρων κομματιών νήματος η κάθε μια (Κ1, Κ2 και Κ3, αντίστοιχα). Τέλος, στα κομμάτια κάθε ομάδας αντιστοίχησε, με μια τυχαία διαδικασία, από μία χημική ουσία και εκτέλεσε το πείραμα. Στον πίνακα που ακολουθεί φαίνονται το πειραματικό σχέδιο και οι μετρήσεις που πήρε ο ερευνητής (μικρότερη τιμή σημαίνει συγκράτηση λιγότερης υγρασίας)

Κ1	Κ2	Κ3
Α3 9.9	Α4 13.4	Α2 12.7
Α1 10.1	Α2 12.9	Α4 12.9
Α2 11.4	Α1 12.2	Α3 11.4
Α4 12.1	Α3 12.3	Α1 11.9

α) Εξηγείστε τι τύπου *πειραματικό σχέδιο* επέλεξε να εφαρμόσει ο ερευνητής. β) Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στη μέση συγκράτηση υγρασίας μεταξύ των τεσσάρων επεμβάσεων; γ) Από το αποτέλεσμα της ανάλυσης που κάνατε, κρίνετε ότι «δικαιώθηκε» η σχεδιαστική επιλογή του ερευνητή;

10. Για να συγκρίνουμε δυο αντιδιαβρωτικά επιστρώματα σωλήνων, έστω Α και Β, κάναμε το εξής πείραμα. Σε κάθε μία από 10 τυχαία επιλεγμένες περιοχές τοποθετήσαμε μέσα στο έδαφος δύο σωλήνες, τον ένα δίπλα στον άλλο, στο ίδιο βάθος και για ίδιο χρονικό διάστημα. Ο ένας σωλήνας από τους δύο που τοποθετήθηκαν σε κάθε περιοχή, είχε επιστρωθεί με το αντιδιαβρωτικό Α και ο άλλος με το αντιδιαβρωτικό Β. Στον πίνακα που ακολουθεί φαίνεται ο βαθμός διάβρωσης κάθε σωλήνα (σε $10^{-3}\,in$) στις δέκα περιοχές.

	Περιοχή									
	1	2	3	4	5	6	7	8	9	10
Βαθμός διάβρωσης με επίστρωμα Α (σε $10^{-3}\,in$)	47	52	55	76	57	58	39	54	46	71
Βαθμός διάβρωσης με επίστρωμα Β (σε $10^{-3}\,in$)	45	42	46	69	58	54	40	46	42	53

Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν τα πειραματικά δεδομένα ότι τα δύο αντιδιαβρωτικά δεν έχουν την ίδια αποτελεσματικότητα;
Στο ερώτημα αυτό απαντήσαμε σε προηγούμενη ενότητα κάνοντας κατάλληλο *t-test*. Μπορείτε να απαντήσετε στο ίδιο ερώτημα εφαρμόζοντας άλλη μέθοδο στατιστικής ανάλυσης; Πώς συγκρίνετε τις δύο μεθόδους;

11. Ένας ερευνητής, για να μελετήσει πώς επηρεάζεται η βλάστηση των φασολιών από το είδος μυκητοκτόνων που χρησιμοποιούνται και την ποικιλία των φασολιών, σχεδίασε ένα *3x3 παραγοντικό πείραμα (εντελώς τυχαιοποιημένο)* με τέσσερις επαναλήψεις για κάθε συνδυασμό μυκητοκτόνου -ποικιλίας. Στον πίνακα που ακολουθεί δίνονται οι μετρήσεις που πήρε ο ερευνητής μετά την εκτέλεση του πειράματος (ως ποσοστό βλάστησης, %).

		Μυκητοκτόνο		
		M1	M2	M3
Ποικιλία	Α	78 62 72 68	82 78 70 75	92 85 87 90
	Β	65 70 75 69	72 68 73 76	85 79 84 80
	Γ	81 78 75 85	87 83 82 85	94 90 89 95

α) Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχει στατιστικά σημαντική επίδραση στη βλάστηση των φασολιών που να οφείλεται i) στην αλληλεπίδραση ποικιλίας και μυκητοκτόνου ii) στην ποικιλία των φασολιών iii) στο είδος μυκητοκτόνου β) Ποιες υποθέσεις/παραδοχές χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (α);

12. Στο πλαίσιο μιας έρευνας για τη διερεύνηση του βαθμού ικανοποίησης των ασθενών από τις προσφερόμενες υπηρεσίες περίθαλψης σε νοσοκομεία του Ε.Σ.Υ., επελέγησαν τυχαία 5 ασθενείς από κάθε μία από 3 κλινικές (Καρδιολογική, Παθολογική και Ορθοπεδική, αντίστοιχα) δύο νοσοκομείων του Ε.Σ.Υ. (έστω Α και Β). Στον πίνακα που ακολουθεί δίνεται ο βαθμός ικανοποίησης κάθε ασθενούς σε μια κλίμακα από 0 (ελάχιστη ικανοποίηση) έως 10 (μέγιστη ικανοποίηση).

		Νοσοκομείο	
		Α	Β
Κλινική	Καρδ.	7 6 5 7 4	8 9 7 9 8
	Παθ.	8 7 5 5 6	6 5 6 7 5
	Ορθ.	4 3 5 6 5	7 6 5 5 7

α) Τι τύπου *πειραματικό σχέδιο* εφαρμόσθηκε; β) Να κάνετε κατάλληλη στατιστική ανάλυση των δεδομένων που προέκυψαν από την έρευνα. γ) Να διατυπώσετε τα συμπεράσματά σας σε μια μικρή παράγραφο.

13. Το τμήμα ποιοτικού ελέγχου ενός εργοστασίου θεωρεί/υποψιάζεται ότι ο αριθμός ελαττωματικών προϊόντων που παράγονται, εξαρτάται από τον τύπο της μηχανής παραγωγής (M1, M2, M3) και από τον χειριστή της μηχανής (A1, A2, A3). Για να ελέγξει αν πράγματι έτσι συμβαίνει, το τμήμα ποιοτικού ελέγχου, για κάθε συνδυασμό, *τύπος μηχανής-χειριστής*, επέλεξε τυχαία τέσσερις παρτίδες προϊόντων (των 100 προϊόντων) και για κάθε μια κατέγραψε τον αριθμό ελαττωματικών προϊόντων. Οι σχετικές παρατηρήσεις φαίνονται στον πίνακα που ακολουθεί.

		Τύπος μηχανής		
		M1	M2	M3
Χειριστής μηχανής	A1	2 5 4 7	8 3 11 5	9 6 7 12
	A2	6 7 3 1	2 5 7 9	5 6 8 11
	A3	4 7 5 3	7 9 8 5	6 10 12 11

α) Σε *επίπεδο σημαντικότητας 1%*, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχει στατιστικά σημαντική επίδραση στον αριθμό ελαττωματικών προϊόντων που να οφείλεται i) στην αλληλεπίδραση μεταξύ, χειριστή της μηχανής και του τύπου της μηχανής ii) στον τύπο μηχανής iii) στο χειριστή της μηχανής β) Ποιες υποθέσεις/παραδοχές χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (α);

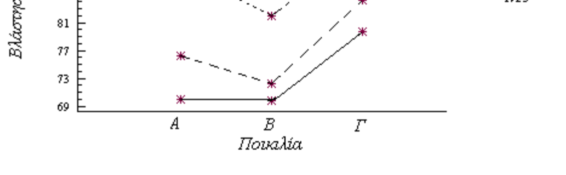
14. Ένας ερευνητής, για να μελετήσει την επίδραση στην παραγωγή γαρίδας, της θερμοκρασίας και της αλμυρότητας του νερού, καθόρισε τρία επίπεδα θερμοκρασίας και τρία επίπεδα αλμυρότητας και εκτέλεσε ένα *3x3 παραγοντικό πείραμα* με μια παρατήρηση σε κάθε επέμβαση. Στον πίνακα που ακολουθεί δίνονται οι μετρήσεις που πήρε ο ερευνητής μετά την εκτέλεση του πειράματος (σε kgr).

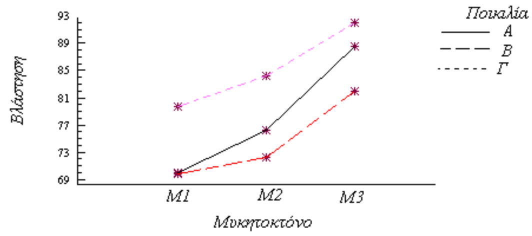
		Αλμυρότητα (σε ppm)		
		700	1400	2100
Θερμοκρασία	60^oF	3	5	4
	70^oF	11	10	12
	80^oF	16	21	17

α) Ποια παραδοχή έκανε ο ερευνητής και σχεδίασε το πείραμα με μία παρατήρηση για κάθε επέμβαση; β) Να κάνετε στατιστική ανάλυση των πειραματικών δεδομένων, σε *επίπεδο σημαντικότητας 5%.* γ) Ποιες υποθέσεις χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (β);

ΑΠΑΝΤΗΣΕΙΣ

1. *α) Εντελώς τυχαιοποιημένο σχέδιο β) Όχι, αφού* $F = 2.57$ *ενώ* $F_{2;12;0.05} = 3.89$ *γ) Τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.*
2. *α) Όχι, αφού* $F = 2.25$ *ενώ* $F_{3;8;0.05} = 4.07$ *β) Πρόκειται για εντελώς τυχαιοποιημένο σχέδιο με τέσσερις επεμβάσεις επομένως ελέγχουμε τους μέσους τεσσάρων πληθυσμών με τέσσερα ανεξάρτητα δείγματα. Με την υπόθεση ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές κάνουμε ANOVA με έναν παράγοντα.*
3. *α) Πρόκειται για εντελώς τυχαιοποιημένο σχέδιο (ελέγχουμε τους μέσους τριών πληθυσμών με τρία ανεξάρτητα δείγματα) β) Ναι, αφού* $F = 4.99 > F_{2;12;0.05} = 3.89$ *γ) Τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.*
4. *α) Ναι, αφού* $F = 68.28 > F_{3;15;0.05} = 3.29$ *β) βλ. ερώτημα (β) της άσκησης 2.*
5. *α) Ναι, αφού* $F = 57.38 > F_{3;20;0.05} = 3.10$ *β) βλ. ερώτημα (β) της άσκησης 2 γ) Η υπόθεση της κανονικότητας των πληθυσμών γιατί από κάθε φυτό, ως παρατήρηση, παίρνουμε ένα δειγματικό μέσο (το μέσο μήκος 20 φύλλων).*
6. *Ναι, αφού* $F = 6.40 > F_{4;10;0.05} = 3.48$.
7. *α) Τυχαιοποιημένων πλήρων ομάδων με 4 blocks (καλλιέργειες) και 3 επεμβάσεις (εκτιμητές) σε κάθε block β) Ναι, αφού* $F = 6.46 > F_{2;6;0.05} = 5.14$ *γ) Όχι, γιατί δε βρέθηκαν στατιστικά σημαντικές διαφορές στις μέσες εκτιμήσεις που να οφείλονται στις καλλιέργειες (μεταξύ των blocks) (*$F = 3.75$ *ενώ* $F_{3;6;0.05} = 4.76$*).*
8. *α) 7 β) 7 γ) 5 ε) Ναι,* $F = 9.68 > F_{4;24;0.05} = 2.78$. *Ναι,* $F = 8.59 > F_{6;24;0.05} = 2.51$.
9. *α) Τυχαιοποιημένων πλήρων ομάδων με 3 blocks (κομμάτια νήματος) και 4 επεμβάσεις (χημικές ουσίες) σε κάθε block β) Ναι, αφού* $F = 19.44 > F_{3;6;0.05} = 4.76$ *γ)Ναι, γιατί βρέθηκαν στατιστικά σημαντικές διαφορές στη μέση συγκέντρωση υγρασίας που οφείλονται στα κομμάτια νήματος (δηλ. μεταξύ των blocks) (*$F = 40.21 > F_{2;6;0.05} = 5.14$*).*
10. *Με ANOVA (για τυχαιοποιημένες πλήρεις ομάδες με 10 blocks και 2 επεμβάσεις σε κάθε block) προκύπτει το ίδιο συμπέρασμα αφού* $F = 10.95 > F_{1;9;0.05} = 5.12$. *Οι δυο έλεγχοι είναι ισοδύναμοι. Παρατηρείστε ότι* $F = 10.95 = t^2 = 3.3^2$ *και* $F_{1;9;0.05} = 5.12 = (t_{9;0.025})^2 = 2.262^2$.
11. *α) i. Όχι, αφού* $F = 0.8713$ *ενώ* $F_{4;27;0.05} = 2.73$

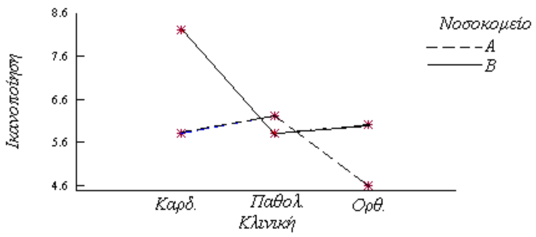


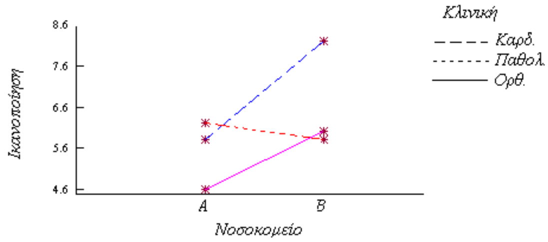


ii. Ναι, αφού $F = 21.37 > F_{2;27;0.05} = 3.35$ *iii. Ναι, αφού* $F = 39.08 > F_{2;27;0.05} = 3.35$.

β) Για κάθε συνδυασμό ποικιλίας-μυκητοκτόνου, τα αντίστοιχα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.

12. *α) 3x2 παραγοντικό πείραμα εντελώς τυχαιοποιημένο με πέντε επαναλήψεις για κάθε συνδυασμό κλινικής-νοσοκομείου. β) Υπάρχει στατιστικά σημαντική αλληλεπίδραση των δύο παραγόντων στη μέση ικανοποίηση των ασθενών (αφού* $F = 4.25 > F_{2;24;0.05} = 3.40$*).*





13. *α) i. Όχι, αφού* $F = 0.12$ *ενώ* $F_{4;27;0.01} = 4.11$. *ii. Ναι, αφού* $F = 7.53 > F_{2;27;0.01} = 5.49$ *iii. Όχι, αφού* $F = 0.91$ *ενώ* $F_{2;27;0.01} = 5.49$.

β) Για κάθε συνδυασμό τύπος μηχανής-χειριστής, τα αντίστοιχα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.

14. *α) Ότι οι δύο παράγοντες επιδρούν ανεξάρτητα/προσθετικά. β) Η επίδραση της αλμυρότητας δεν είναι στατιστικά σημαντική (*$F = 1$ *και* $F_{2;4;0.05} = 6.94$*), ενώ της θερμοκρασίας είναι (*$F = 49 > F_{2;4;0.04} = 6.94$*). γ) Για κάθε συνδυασμό θερμοκρασίας-αλμυρότητας, τα αντίστοιχα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.*