

Ενότητα 7: Ανάλυση Διασποράς με έναν παράγοντα (One way Analysis of Variance)

Σε αυτή την ενότητα θα εξετάσουμε ένα ειδικό πρόβλημα γραμμικής παλινδρόμησης το οποίο εμφανίζεται αρκετά συχνά στις εφαρμογές. Συγκεκριμένα, θέλουμε να διερευνήσουμε πώς επηρεάζεται μία (scaled) μεταβλητή Y από μία κατηγορική μεταβλητή X (ordinal ή nominal) η οποία λαμβάνει k τιμές. Η μεταβλητή X καλείται και *παράγοντας* (factor) με k *στάθμες*. Για το σκοπό αυτό λαμβάνουμε k τυχαία δείγματα της μεταβλητής Y :

$Y_{11}, Y_{12}, \dots, Y_{1,n_1}$ όταν ο παράγοντας X βρίσκεται στην 1^η στάθμη,

$Y_{21}, Y_{22}, \dots, Y_{2,n_2}$ όταν ο παράγοντας X βρίσκεται στην 2^η στάθμη,

...

$Y_{k1}, Y_{k2}, \dots, Y_{k,n_k}$ όταν ο παράγοντας X βρίσκεται στην k -οστή στάθμη,

Θεωρούμε το μοντέλο

$$Y_{1j} = \mu_1 + \varepsilon_{1j}, \quad j=1,2,\dots,n_1, \quad (X \text{ στην } 1^{\text{η}} \text{ στάθμη})$$

$$Y_{2j} = \mu_2 + \varepsilon_{2j}, \quad j=1,2,\dots,n_2 \quad (X \text{ στην } 2^{\text{η}} \text{ στάθμη})$$

...

$$Y_{kj} = \mu_k + \varepsilon_{kj}, \quad j=1,2,\dots,n_k \quad (X \text{ στην } k\text{-οστή στάθμη})$$

όπου τα «σφάλματα» $\varepsilon_{ij}, j=1,2,\dots,n_i, i=1,2,\dots,k$ είναι ανεξάρτητες τ.μ. που ακολουθούν $N(0, \sigma^2)$. Ουσιαστικά θεωρούμε ότι η μεταβλητή $Y \sim N(\mu_i, \sigma^2)$ όταν ο παράγοντας X βρίσκεται στην i -στάθμη. Μας ενδιαφέρει να εκτιμήσουμε τα $\mu_1, \mu_2, \dots, \mu_k$ (και το σ^2) και να ελέγξουμε αν είναι ίσα ή υπάρχει διαφορά μεταξύ τους. Με άλλα λόγια θέλουμε να ελέγξουμε αν η μεταβλητή Y έχει διαφορετική «συμπεριφορά» στις διαφορετικές στάθμες της X . Το συγκεκριμένο πρόβλημα μπορεί να θεωρηθεί και ως μία επέκταση του t -τεστ που χρησιμοποιείται για τον έλεγχο της ισότητας των μέσων μ_1, μ_2 δύο (ανεξάρτητων, κανονικών) πληθυσμών.

Όπως έχει ήδη αναφερθεί, το παραπάνω μοντέλο μπορεί να θεωρηθεί και ως μια ειδική περίπτωση ενός πολλαπλού γραμμικού μοντέλου. Πράγματι, όλες οι παραπάνω σχέσεις γράφονται σε μία σχέση ως εξής:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

όπου

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ \vdots \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & & & & \vdots \\ 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 1 & 0 & \dots & \dots & 0 \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \ddots \\ 0 & \dots & \dots & \dots & 0 & 1 \\ \vdots & & & & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \vdots \\ \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{bmatrix}.$$

όπου $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, $n = n_1 + n_2 + \dots + n_k$, και επομένως έχουμε ένα πολλαπλό γραμμικό μοντέλο με ειδικής μορφής πίνακα σχεδιασμού \mathbf{X} .

7.1. Εκτίμηση των παραμέτρων $\mu_1, \mu_2, \dots, \mu_k$ και σ^2

Για την εκτίμηση των παραμέτρων του μοντέλου μπορούμε να χρησιμοποιήσουμε τους γενικούς τύπους που ισχύουν για το γενικό πολλαπλό γραμμικό μοντέλο (βλ. Ενότητα 6) ή να εργαστούμε εξαρχής αναζητώντας τις ε.μ.π. (ή ισοδύναμα τις εκτιμήτριες ελαχίστων τετραγώνων). Σε κάθε περίπτωση είναι εύκολο να διαπιστώσουμε ότι οι εκτιμήτριες των παραμέτρων έχουν την ακόλουθη απλή μορφή

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \bar{Y}_{i\cdot}, \quad i = 1, 2, \dots, k.$$

Όλες οι ποσότητες που χρησιμοποιήθηκαν στο πολλαπλό γραμμικό μοντέλο μπορούν να χρησιμοποιηθούν και εδώ. Για παράδειγμα, οι προσαρμοσμένες τιμές των Y_{ij} και τα κατάλοιπα θα είναι

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\mu}}, \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}},$$

και συγκεκριμένα, $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$, $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$. Ως εκτιμήτρια της διασποράς σ^2 των σφαλμάτων, μπορούμε να θεωρήσουμε (όπως και στο πολλαπλό γραμμικό μοντέλο) την

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-k} \sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2 = \frac{1}{n-k} \sum_{i,j} (Y_{ij} - \bar{Y}_{i\cdot})^2.$$

7.2. Έλεγχος υποθέσεων και δ.ε. για τις παραμέτρους του μοντέλου.

Είναι εύκολο να διαπιστωθεί ότι $\hat{\mu}_i = \bar{Y}_{i\cdot} \sim N(\mu_i, \sigma^2/n_i)$ και επομένως,

$$\frac{\bar{Y}_{i\cdot} - \mu_i}{S/\sqrt{n_i}} \sim t_{n-k}$$

από όπου προκύπτει ότι το

$$\left(\bar{Y}_{i\cdot} - \frac{S}{\sqrt{n_i}} t_{n-k} \left(\frac{\alpha}{2} \right), \bar{Y}_{i\cdot} + \frac{S}{\sqrt{n_i}} t_{n-k} \left(\frac{\alpha}{2} \right) \right)$$

είναι ένα δ.ε. για το μ_i συντελεστού $1-\alpha$. Το συγκεκριμένο δ.ε. χρησιμοποιεί ως εκτίμηση του σ^2 την S^2 η οποία βασίζεται στις παρατηρήσεις από όλες τις στάθμες του παράγοντα (pooled estimate). Η εκτιμήτρια αυτή είναι η καλύτερη που μπορούμε να πάρουμε, αρκεί η διασπορά των σφαλμάτων να είναι σταθερή (και ίση με σ^2) σε όλες τις στάθμες του παράγοντα X . Εάν δεν είμαστε σίγουροι ότι κάτι τέτοιο συμβαίνει μπορούμε να εκτιμήσουμε την διασπορά των σφαλμάτων ξεχωριστά σε κάθε στάθμη (internal estimate),

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$$

και να πάρουμε το δ.ε για το μ_i συντελεστού $1-\alpha$,

$$\left(\bar{Y}_{i\cdot} - \frac{S_i}{\sqrt{n_i}} t_{n_i-1} \left(\frac{\alpha}{2} \right), \bar{Y}_{i\cdot} + \frac{S_i}{\sqrt{n_i}} t_{n_i-1} \left(\frac{\alpha}{2} \right) \right).$$

Σε αυτό το μοντέλο μας ενδιαφέρει ο έλεγχος $H_0: \mu_i = \mu_j$ για $i \neq j$. Επειδή,

$$\hat{\mu}_i - \hat{\mu}_j = \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \sim N(\mu_i - \mu_j, \sigma^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right))$$

προκύπτει το δ.ε. συντελεστού $1-\alpha$ για το $\mu_i - \mu_j$

$$\left(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} - S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-k} \left(\frac{\alpha}{2} \right), \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} + S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-k} \left(\frac{\alpha}{2} \right) \right)$$

ενώ θα απορρίπτεται η $H_0: \mu_i = \mu_j$ έναντι της $H_1: \mu_i \neq \mu_j$ όταν

$$|T_{ij}| > t_{n-k}(\frac{\alpha}{2}) \quad \text{όπου} \quad T_{ij} = \frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}}{S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \underset{H_0}{\sim} t_{n-k}$$

με αντίστοιχο p-value (για δείγμα που έδωσε $T_{ij} = t_{ij}$)

$$p\text{-value} = P(|T_{ij}| > |t_{ij}|) = 2(1 - F_{t_{n-k}}(|t_{ij}|)).$$

7.3. Ερμηνεύοντας τη συνολική μεταβλητότητα του μοντέλου

Το μοντέλο που εξετάζουμε αποτελεί υποπερίπτωση του πολλαπλού γραμμικού μοντέλου και επομένως μπορούμε απευθείας και εδώ να πούμε ότι η διασπορά των παρατηρήσεων Y_{ij} χωρίζεται σε δύο αθροίσματα (θα μπορούσε εύκολα να αποδειχθεί και ανεξάρτητα)

$$\sum_{i,j} (Y_{ij} - \bar{Y})^2 = \sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2 + \sum_{i,j} (\hat{Y}_{ij} - \bar{Y})^2 = \sum_{i,j} (Y_{ij} - \bar{Y}_{i\cdot})^2 + \sum_{i,j} (\bar{Y}_{i\cdot} - \bar{Y})^2,$$

τα οποία συμβολίζονται και πάλι με SST , SSE και SSR (ή $SSTr$) αντίστοιχα (το \bar{Y} συμβολίζεται και με $\bar{Y}_{..}$). Αν θεωρήσουμε ότι οι n παρατηρήσεις χωρίζονται σε k ομάδες (μία για κάθε στάθμη του παράγοντα), τότε το SSE μπορεί να θεωρηθεί ως η μεταβλητότητα «εντός των ομάδων» (Sum of Squares Within Groups) ενώ το SSR η μεταβλητότητα «μεταξύ των ομάδων» (Sum of Squares Between Groups). Αποδεικνύεται (εδώ εύκολα) ότι

$$\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2 = \frac{1}{\sigma^2} \sum_{i,j} (Y_{ij} - \bar{Y}_{i\cdot})^2 \sim \chi_{n-k}^2$$

ενώ όταν $\mu_1 = \mu_2 = \dots = \mu_k$ τότε αποδεικνύεται ότι

$$\frac{SSTr}{\sigma^2} \sim \chi_{k-1}^2, \quad \frac{SST}{\sigma^2} \sim \chi_{n-1}^2.$$

(διαφορετικά ακολουθούν κάποιες μη-κεντρικές κατανομές χι-τετράγωνο). Επομένως, αν $\mu_1 = \mu_2 = \dots = \mu_k$, το πηλίκο

$$F = \frac{\frac{SSTr}{\sigma^2} / (k-1)}{\frac{SSE}{\sigma^2} / (n-k)} = \frac{SSTr / (k-1)}{SSE / (n-k)} \sim F_{k-1, n-k}$$

($SSTr$ και SSE είναι ανεξάρτητες). Αντίστοιχα με το πολλαπλό γραμμικό μοντέλο, μπορούμε να κατασκευάσουμε έναν έλεγχο για την υπόθεση¹ $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ (δηλ. ότι η Y συμπεριφέρεται το ίδιο σε όλες της στάθμες του παράγοντα, άρα είναι ανεξάρτητη του παράγοντα). Θα απορρίπτεται η H_0 όταν (ε.σ. α)

$$F = \frac{SSTr / (k-1)}{SSE / (n-k)} > F_{k-1, n-k}(\alpha): \text{άνω } \alpha\text{-σημείο της κατανομής } F \text{ με } k-1 \text{ και } n-k \text{ β.ε.}$$

¹ Στο πολλαπλό γραμμικό μοντέλο ο ίδιος έλεγχος χρησιμοποιείται για την υπόθεση $b_1 = b_2 = \dots = b_{p-1} = 0$ ενώ εδώ για την $\mu_1 = \mu_2 = \dots = \mu_k$ (όχι $= 0$). Η διαφορά εδώ είναι το έχουμε ένα μοντέλο χωρίς σταθερά (ο πίνακας σχεδιασμού X δεν έχει στην πρώτη στήλη μονάδες) διότι

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, k.$$

Εάν όμως θέσουμε $\mu_i = \mu + t_i$ ($\sum n_i t_i = 0$) τότε το παραπάνω γράφεται ισοδύναμα ως μοντέλο με σταθερά, ως εξής,

$$Y_{ij} = \mu + t_i + \varepsilon_{ij}, \quad j = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, k.$$

στο οποίο το F -τεστ ελέγχει αν $t_1 = t_2 = \dots = t_k = 0$ δηλαδή $\mu_1 = \mu_2 = \dots = \mu_k = \mu$.

με αντίστοιχο p-value:

$$p\text{-value} = 1 - F_{F_{k-1, n-k}} \left(\frac{SSTr / (k-1)}{SSE / (n-k)} \right)$$

Όλες οι παραπάνω ποσότητες συνοψίζονται στον πίνακα ανάλυσης διασποράς (ANOVA):

<i>Model</i>	Sum of Squares	<i>df</i>	Mean Square	F-Ratio	Sig. (p-value)
Between Groups	$SSTr = \sum_{i,j} (\bar{Y}_{i\cdot} - \bar{Y})^2$	$k-1$	$MSTr = \frac{SSTr}{k-1}$	$\frac{MSTr}{MSE}$	$1 - F_{F_{k-1, n-k}} \left(\frac{MSTr}{MSE} \right)$
Within Groups	$SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_{i\cdot})^2$	$n-k$	$MSE = \frac{SSE}{n-k} = S^2$		
Total	$SST = \sum_{i,j} (Y_{ij} - \bar{Y})^2$	$n-1$			

7.4. Πολλαπλές συγκρίσεις

Μία από τις κύριες επιδιώξεις μας στο μοντέλο αυτό είναι να ελέγξουμε αν η Y παρουσιάζει διαφορετική συμπεριφορά στις διάφορες στάθμες του παράγοντα X . Μπορούμε με βάση το F -τεστ που είδαμε παραπάνω να εξετάσουμε την $H_0: \mu_1 = \mu_2 = \dots = \mu_k$. Εάν απορριφθεί η συγκεκριμένη υπόθεση μας ενδιαφέρει να εξετάσουμε ποια μ_i διαφέρουν από τα υπόλοιπα. Για το σκοπό αυτό μπορούμε να κάνουμε όλες τις ανά δύο συγκρίσεις των μέσων $\mu_i = \mu_j, i \neq j$ (πολλαπλές συγκρίσεις).

Εάν έχουμε k στάθμες τότε θα πρέπει να ελέγξουμε $l = k(k-1)/2$ υποθέσεις ισότητας μέσων (k ανά δύο). Π.χ. αν $k = 3$ μπορούμε να κάνουμε όλους τους ελέγχους

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

(με εναλλακτικές τις αμφίπλευρες $\mu_i \neq \mu_j$). Μπορούμε να ελέγξουμε κάθε μία από τις παραπάνω υποθέσεις ξεχωριστά με t -tests σε ε.σ. a (με τον τρόπο που είδαμε στην Παράγραφο 7.2.) και στη συνέχεια να δούμε ποιοι μέσοι διαφέρουν και να τους κατατάξουμε (ομαδοποιήσουμε).

Π.χ. αν απορρίψουμε τις υποθέσεις $H_0: \mu_1 = \mu_2$ και $H_0: \mu_1 = \mu_3$ (ενώ δεν απορρίψουμε ότι $H_0: \mu_2 = \mu_3$) τότε μπορούμε να πούμε ότι ο μέσος μ_1 διαφέρει από τους άλλους δύο οι οποίοι μπορεί να είναι ίσοι («πιθανές» ομάδες από ίσους μέσους: $\{\mu_1\}, \{\mu_2, \mu_3\}$). Εάν σε άλλη περίπτωση απορρίψουμε μόνο την $H_0: \mu_1 = \mu_3$ (ενώ $\bar{Y}_{1\cdot} \leq \bar{Y}_{2\cdot} \leq \bar{Y}_{3\cdot}$) τότε ο μ_1 διαφέρει από τον μ_3 ενώ ο μ_2 μπορεί να είναι ίσος είτε με τον μ_1 είτε με τον μ_3 («πιθανές» ομάδες από ίσους μέσους: $\{\mu_1, \mu_2\}, \{\mu_2, \mu_3\}$).

Η παραπάνω διαδικασία κατά την οποία συγκρίνουμε όλους τους μέσους ανά δυο (έλεγχος l το πλήθος υποθέσεων μέσω t -test), κάθε φορά σε στάθμη a , καλείται μέθοδος LSD (Least Significant Difference). Η μέθοδος αυτή έχει ένα μειονέκτημα: η πιθανότητα λανθασμένης απόρριψης κάποιας από τις l αυτές υποθέσεις δεν είναι a αλλά αρκετά μεγαλύτερη. Πράγματι, αν A_i είναι το ενδεχόμενο να απορρίψουμε την i -υπόθεση, $i = 1, 2, \dots, l$, τότε η πιθανότητα λανθασμένης απόρριψης κάποιας από τις l υποθέσεις $\mu_i = \mu_j, i \neq j$ (ενώ $\mu_1 = \mu_2 = \dots = \mu_k$) θα είναι

$$P(A_1 \cup A_2 \cup \dots \cup A_l) = 1 - P(A_1^c \cap A_2^c \cap \dots \cap A_l^c) \approx 1 - P(A_1^c)P(A_2^c) \dots P(A_l^c) = 1 - (1-a)^l$$

(θα ίσχυε ισότητα αν οι απορρίψεις είναι ανεξάρτητες, που εδώ δεν είναι), δηλαδή αρκετά μεγαλύτερη του a (ειδικά για μεγάλο l).

Μας ενδιαφέρει να γίνει η πολλαπλή σύγκριση έτσι ώστε με πιθανότητα το πολύ a να κάνουμε τουλάχιστον μια λανθασμένη απόρριψη στους l ελέγχους (δηλ. συνολική $P(\text{σφάλμα τύπου I}) \leq a$). Για το σκοπό αυτό έχουν προταθεί διάφορες μέθοδοι. Στη συνέχεια θα εξετάσουμε μόνο δύο από αυτές: τη μέθοδο Bonferroni και την μέθοδο Tukey

a. Η μέθοδος Bonferroni. Ένας απλός αλλά όχι τόσο αποτελεσματικός τρόπος για να παραμείνει η $P(\text{σφάλμα τύπου I}) \leq a$ στην πολλαπλή σύγκριση είναι να κάνουμε κάθε έναν από l τους ελέγχους

$H_0: \mu_i = \mu_j$ όχι σε ε.σ. a (όπως στην μέθοδο LSD) αλλά σε ε.σ. a/l ώστε η πιθανότητα λανθασμένης απόρριψης τουλάχιστον μιας υπόθεσης να είναι

$$P(A_1 \cup A_2 \cup \dots \cup A_l) \leq P(A_1) + P(A_2) + \dots + P(A_l) = \frac{a}{l} + \dots + \frac{a}{l} = a.$$

Η παραπάνω μέθοδος συνήθως προτιμάται όταν το k είναι μικρό. Ισοδύναμα, μπορούμε να διατηρήσουμε το ίδιο ε.σ. a σε κάθε έναν από τους ελέγχους αλλά να πολλαπλασιάσουμε τα αρχικά p-value (που προκύπτουν από την μέθοδο LSD) με l . Έτσι, αν ένας έλεγχος δίνει p-value = p_0 (με την μέθοδο LSD), θεωρούμε ότι (με την μέθοδο Bonferroni) θα έχει p-value = lp_0 (ή καλύτερα $\min\{lp_0, 1\}$).

b. Η μέθοδος Tukey. Η μέθοδος αυτή βασίζεται στην κατανομή του τυποποιημένου εύρους από ανεξάρτητες κανονικές. Συγκεκριμένα, έστω Z_1, Z_2, \dots, Z_r ανεξάρτητες τ.μ. από την $N(0, d^2)$ και έστω D^2 μία εκτιμήτρια του d^2 τ.μ. ανεξάρτητη των Z_i τέτοια ώστε $nD^2/d^2 \sim \chi_v^2$, για κάποια παράμετρο v . Η κατανομή $F_{R(r,v)}$ του τυποποιημένου εύρους,

$$R_{r,v} = \frac{\max\{Z_1, \dots, Z_r\} - \min\{Z_1, \dots, Z_r\}}{D}.$$

έχει μελετηθεί, και έστω $q(r, v, a)$ τα άνω a -σημεία της.

Η μέθοδος Tukey για ισομεγέθη δείγματα. Αν $n_1 = n_2 = \dots = n_k = m$ (ισομεγέθη δείγματα στις k στάθμες του παράγοντα) τότε οι τ.μ.

$$Z_1 = \bar{Y}_{1\bullet} - \mu_1, Z_2 = \bar{Y}_{2\bullet} - \mu_2, \dots, Z_k = \bar{Y}_{k\bullet} - \mu_k$$

είναι ανεξάρτητες τ.μ. και ακολουθούν $N(0, d^2 = \sigma^2/m)$. Επίσης, αν επιλέξουμε $D = S/\sqrt{m}$ τότε $(n-k)D^2/d^2 \sim \chi_{n-k}^2$ (και D ανεξ. των Z_i) και επομένως, σύμφωνα και με τα παραπάνω, το τυποποιημένο εύρος

$$R_{k,n-k} = \frac{\max\{Z_1, \dots, Z_k\} - \min\{Z_1, \dots, Z_k\}}{D} = \frac{\max_h \{\bar{Y}_{h\bullet} - \mu_h\} - \min_h \{\bar{Y}_{h\bullet} - \mu_h\}}{S/\sqrt{m}}$$

θα έχει γνωστή κατανομή, την $F_{R(k,n-k)}$, και άνω a -σημεία τα $q(k, n-k, a)$. Επομένως

$$P\left(\frac{\max_h \{\bar{Y}_{h\bullet} - \mu_h\} - \min_h \{\bar{Y}_{h\bullet} - \mu_h\}}{S/\sqrt{m}} \leq q(k, n-k, a)\right) = 1 - a.$$

Είναι εύκολο τώρα να επαληθεύσουμε ότι το ενδεχόμενο στην παραπάνω πιθανότητα είναι ίσο με το ενδεχόμενο

$$\frac{|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} - (\mu_i - \mu_j)|}{S/\sqrt{m}} \leq q(k, n-k, a) \text{ για κάθε } i, j$$

και άρα τελικά,

$$P\left(\frac{|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} - (\mu_i - \mu_j)|}{S/\sqrt{m}} \leq q(k, n-k, a) \text{ για κάθε } i, j\right) = 1 - a$$

ή ισοδύναμα

$$P\left((\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) - \frac{S}{\sqrt{m}}q(k, n-k, a) \leq \mu_i - \mu_j \leq (\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) + \frac{S}{\sqrt{m}}q(k, n-k, a) \text{ για κάθε } i, j\right) = 1 - a$$

από όπου προκύπτει ένα **πολλαπλό** δ.ε. (ή ένα σύνολο από ταυτόχρονα δ.ε.) για τις διαφορές $\mu_i - \mu_j$ με **συνολικό** σ.ε. $1 - a$. Επίσης αν απορρίψουμε την υπόθεση $\mu_i = \mu_j$ όταν

$$|T'_{ij}| > q(k, n-k, a) \quad \text{όπου} \quad T'_{ij} = \frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}}{S/\sqrt{m}},$$

για $i, j = 1, 2, \dots, k$, η πιθανότητα να κάνουμε τουλάχιστον μια λάθος απόρριψη (στους $k(k-1)/2$ ελέγχους) ίση με a . Το p-value του ελέγχου θα είναι (για δείγμα που έδωσε $T'_{ij} = t'_{ij}$)

$$p\text{-value} = P(|T'_{ij}| > |t'_{ij}|) = 2(1 - F_{R(k, n-k)}(|t'_{ij}|)).$$

Η μέθοδος Tukey για ανισομεγέθη δείγματα. Η μέθοδος Tukey μπορεί να τροποποιηθεί ώστε να εφαρμόζεται και στην περίπτωση που έχουμε ανισομεγέθη δείγματα n_1, n_2, \dots, n_k στις διάφορες στάθμες του παράγοντα. Παρατηρούμε ότι τώρα, η τ.μ. $\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}$ δεν έχει διασπορά $2/m$ αλλά $1/n_i + 1/n_j$. Εφαρμόζουμε λοιπόν ακριβώς τα ίδια με παραπάνω μόνο που στη θέση του $1/m$ στα δ.ε. για το $\mu_i - \mu_j$ και στον αντίστοιχο έλεγχο $\mu_i = \mu_j$ χρησιμοποιούμε το $(1/n_i + 1/n_j)/2$. Δηλαδή αντί του m χρησιμοποιούμε τον αρμονικό μέσο των n_i, n_j . Το πολλαπλό δ.ε. που προκύπτει θα έχει (προσεγγιστικά) συντελεστή εμπιστοσύνης a .

7.5. Έλεγχος ομοσκεδαστικότητας των παρατηρήσεων

Για να είναι αξιόπιστα όλα τα παραπάνω θα πρέπει τα σφάλματα (ή ισοδύναμα τα Y_{ij}) να έχουν την ίδια διασπορά σ^2 . Επομένως θα πρέπει στα πλαίσια ελέγχου ορθότητας του μοντέλου να εξετάσουμε αν κάτι τέτοιο πράγματι ισχύει. Μπορούμε και πάλι να χρησιμοποιήσουμε το Levene's test ομοσκεδαστικότητας που έχουμε χρησιμοποιήσει για τον ίδιο σκοπό σε προηγούμενη ενότητα (κατά τον έλεγχο των μέσων δύο ανεξάρτητων πληθυσμών). Το test αυτό βασίζεται στην ανάλυση διασποράς και επομένως τώρα είμαστε σε θέση να το περιγράψουμε. Βασίζεται στις τυχαίες μεταβλητές

$$W_{ij} = |Y_{ij} - \bar{Y}_{i\cdot}|, j = 1, 2, \dots, n_i, i=1, 2, \dots, k.$$

Αν τα Y_{ij} έχουν την ίδια διασπορά σε όλες τις στάθμες, τότε τα W_{ij} θα έχουν την ίδια μέση τιμή σε όλες τις στάθμες του παράγοντα X . Επομένως αρκεί να ελέγξουμε αν οι τ.μ. W_{ij} έχουν την ίδια μέση τιμή σε όλες τις στάθμες του παράγοντα X . Αυτό μπορεί εύκολα να γίνει με την μεθοδολογία που αναπτύχθηκε στην συγκεκριμένη ενότητα (ανάλυση διασποράς).

Το Levene test ουσιαστικά είναι το F -ratio test του πίνακα ANOVA που αντιστοιχεί στο μοντέλο ανάλυσης διασποράς της W ως προς τον παράγοντα X , δηλ. βασίζεται στην στατιστική συνάρτηση

$$F = \frac{SSTr_W / (k-1)}{SSE_W / (n-k)} = \frac{\sum_{i,j} (\bar{W}_{i\cdot} - \bar{W})^2 / (k-1)}{\sum_{i,j} (W_{ij} - \bar{W}_{i\cdot})^2 / (n-k)}.$$

Αν, με βάση αυτό το F -ratio test, απορρίπτεται ότι ο (θεωρητικός) μέσος της W είναι σταθερός σε όλες τις στάθμες της X τότε απορρίπτεται ότι και η (θεωρητική) διασπορά της Y είναι σταθερή σε όλες τις στάθμες της X . Ο έλεγχος αυτός είναι περισσότερο ευσταθής από άλλα παρόμοια test (Bartlett, Cochran, Hartley) στην περίπτωση μη-κανονικότητας των παρατηρήσεων.

Άσκηση 1. Σε μια έρευνα που έγινε στο πανεπιστήμιο Μελβούρνης επελέγη τυχαία ένα δείγμα αντρών και γυναικών διαφόρων ηλικιών το οποίο και υποβλήθηκε σε ένα test αντοχής στον (σωματικό) πόνο. Στον ακόλουθο πίνακα δίνεται για κάθε ένα άτομο του δείγματος ο δείκτης αντοχής στον πόνο ο οποίος εξήχθη με βάση το test αυτό (μεγαλύτερος δείκτης σημαίνει μεγαλύτερη αντοχή). Σε κάθε άτομο του δείγματος καταγράφεται επίσης και το (φυσικό) χρώμα των μαλλιών: ανοιχτό ξανθό (1), σκούρο ξανθό (2), ανοιχτό μελαχρινό (3), σκούρο μελαχρινό (4).

Έχει ενδιαφέρον να εξετάσουμε αν υπάρχουν διαφορές στην μέσο δείκτη αντοχής στον πόνο μεταξύ των ατόμων με διαφορετικό χρώμα μαλλιού.

Χρώμα μαλλιών	Δείκτης αντοχής
1	62
1	60
1	71
1	55
1	48

Χρώμα μαλλιών	Δείκτης αντοχής
2	63
2	57
2	52
2	41
2	43

Χρώμα μαλλιών	Δείκτης αντοχής
3	42
3	50
3	41
3	37

Χρώμα μαλλιών	Δείκτης αντοχής
4	32
4	39
4	51
4	30
4	35

McClave, J. T., and Dietrich II, F. H. (1991). *Statistics*. Dellen Publishing, San Francisco, Exercise 10.20.

1) Για να ληφθεί μία πρώτη εικόνα, κατασκευάστε ένα scatterplot και ένα Boxplot του δείκτη αντοχής για κάθε στάθμη του παράγοντα «χρώμα μαλλιών». Ο δείκτης φαίνεται να επηρεάζεται από το χρώμα των μαλλιών;

2) Αν Y_{ij} είναι δείκτης αντοχής στον πόνο του j -ατόμου με i -χρώμα μαλλιών, εφαρμόστε το μοντέλο

$$Y_{1j} = \mu_1 + \varepsilon_{1j}, \quad j = 1, 2, 3, 4, 5 \quad (n_1 = 5), \quad (\text{παράγοντας «χρώμα» στην 1}^{\text{η}} \text{ στάθμη})$$

$$Y_{2j} = \mu_2 + \varepsilon_{2j}, \quad j = 1, 2, 3, 4, 5 \quad (n_2 = 5) \quad (\text{παράγοντας «χρώμα» στην 2}^{\text{η}} \text{ στάθμη})$$

$$Y_{3j} = \mu_3 + \varepsilon_{3j}, \quad j = 1, 2, 3, 4 \quad (n_3 = 4), \quad (\text{παράγοντας «χρώμα» στην 3}^{\text{η}} \text{ στάθμη})$$

$$Y_{4j} = \mu_4 + \varepsilon_{4j}, \quad j = 1, 2, 3, 4, 5 \quad (n_4 = 5) \quad (\text{παράγοντας «χρώμα» στην 4}^{\text{η}} \text{ στάθμη})$$

(ή συνοπτικά $Y_{ij} = \mu_i + \varepsilon_{ij}$, $i = 1, 2, 3, 4$, $j = 1, 2, \dots, n_i$) όπου μ_i είναι ο μέσος δείκτης αντοχής ατόμου με i χρώμα μαλλιού, και τα «σφάλματα» ε_{ij} είναι ανεξάρτητες τ.μ. που ακολουθούν $N(0, \sigma^2)$. Εκτιμήστε τα $\mu_1, \mu_2, \mu_3, \mu_4$ σημειακά και με δ.ε. (συντελεστού 95% το καθένα). Εκτιμήστε την διασπορά των σφαλμάτων. Ελέγξτε την υπόθεση $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ σε ε.σ. $\alpha = 1\%$. Ο παράγοντας «χρώμα μαλλιών» επιδρά στο δείκτη αντοχής στον πόνο;

3) Να κατασκευάσετε τα γραφήματα των δ.ε. 95% των $\mu_1, \mu_2, \mu_3, \mu_4$.

4) Βρείτε 95% δ.ε. για τη διαφορά $\mu_1 - \mu_4$ και 95% δ.ε. για τη διαφορά $\mu_3 - \mu_4$. Το μ_4 διαφέρει από το μ_1 ; (ε.σ. 5%). Το μ_3 διαφέρει από το μ_4 ; (ε.σ. 5%).

5) Να δοθούν ταυτόχρονα δ.ε. (με συνολικό σ.ε. 95%) για τις διαφορές των μ_i ανά δύο με τις μεθόδους (a) Bonferroni και (b) Tukey. Να γίνει ομαδοποίηση των μέσων $\mu_1, \mu_2, \mu_3, \mu_4$ με τη μέθοδο Tukey (ώστε η συνολική πιθανότητα σφάλματος να είναι 5%).

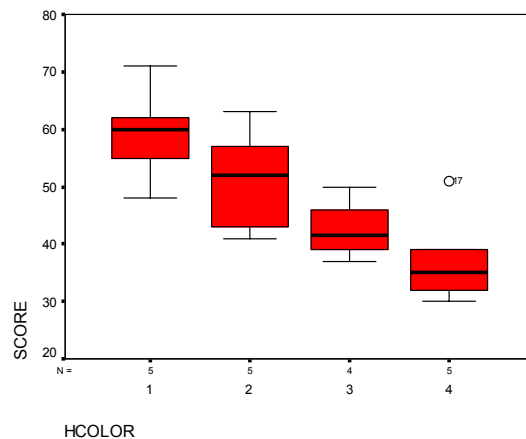
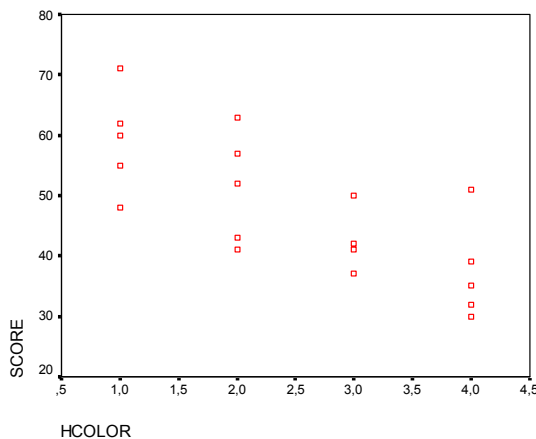
6) Να ελέγξετε αν ο δείκτης αντοχής στον πόνο έχει πράγματι την ίδια διασπορά σ^2 σε όλες τις στάθμες του παράγοντα «χρώμα μαλλιών».

7) Θα μπορούσαμε σε αυτά τα δεδομένα να εφαρμόσουμε ένα απλό γραμμικό μοντέλο;

Λύση. Εισάγουμε τις 19 περιπτώσεις στο SPSS σε δύο μεταβλητές – στήλες: HCOLOR, SCORE

(1) Για το scatterplot επιλέγουμε : Graphs/Scatterplot/simple, Y axis: SCORE, X axis: HCOLOR.

Για Boxplot: Graphs/Boxplot/Simple: Variable: SCORE, Category Axis: HCOLOR.



Και από τα δύο γραφήματα φαίνεται να υπάρχει μία μείωση του μέσου δείκτη όσο πηγαίνουμε από ξανθά σε μελαχρινά άτομα. Δεν μπορούμε όμως άμεσα να πούμε ότι αυτή η μείωση είναι στατιστικά σημαντική (μπορεί π.χ. να οφείλεται στο ότι το τυχαίο δείγμα είναι μικρό).

(2) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: SCORE Factor: HCOLOR, Options: Check Descriptive λαμβάνεται ο πίνακας με τα Descriptives:

Descriptives

SCORE										
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum		
					Lower Bound	Upper Bound				
1	5	59,20	8,53	3,81	48,61	69,79	48	71		
2	5	51,20	9,28	4,15	39,67	62,73	41	63		
3	4	42,50	5,45	2,72	33,83	51,17	37	50		
4	5	37,40	8,32	3,72	27,06	47,74	30	51		
Total	19	47,84	11,46	2,63	42,32	53,36	30	71		

Ο οποίος περιέχει τις εκτιμήσεις των $\mu_1, \mu_2, \mu_3, \mu_4$ (59.2, 51.2, 42.5 37.4) καθώς και τα αντίστοιχα δ.ε. (το καθένα συντελεστού 95% με βάση τα interval estimates του σ^2). Ειδικότερα περιέχει τις ποσότητες

	N	Mean	Std. Dev.	Std. Error	Lower/Upper Bound
i	n_i	$\bar{Y}_{i\cdot}$	$S_i = \sqrt{\frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2}$	$\frac{S_i}{\sqrt{n_i}}$	$\bar{Y}_{i\cdot} \pm \frac{S_i}{\sqrt{n_i}} t_{n_i-1} \left(\frac{\alpha}{2}\right)$

Ενώ η τελευταία γραμμή περιέχει τις ποσότητες

Total	n	\bar{Y}	$S_T = \sqrt{\frac{1}{n-1} \sum_{i,j} (Y_{ij} - \bar{Y})^2}$	$\frac{S_T}{\sqrt{n}}$	$\bar{Y} \pm \frac{S_T}{\sqrt{n}} t_{n-1} \left(\frac{\alpha}{2}\right)$
-------	-----	-----------	--	------------------------	--

(όπου $\alpha = 0.05$)

Επίσης από την ανάλυση αυτή δίνεται και ο πίνακας ANOVA που περιγράψαμε παραπάνω.

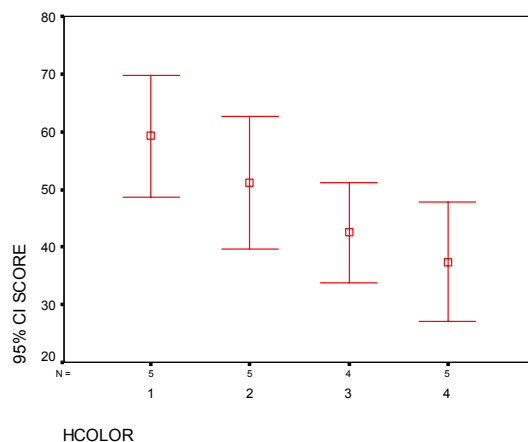
ANOVA

SCORE					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1360,726	3	453,575	6,791	,004
Within Groups	1001,800	15	66,787		
Total	2362,526	18			

Η διασπορά των σφαλμάτων ως γνωστό εκτιμάται από το $SSE/(n-k)$ που δίνεται στον πίνακα ANOVA και είναι 66.787. Αυτό είναι και το pooled estimate της διασποράς σ^2 .

Ο έλεγχος της υπόθεσης $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ έναντι της H_1 : διαφορετικά, γίνεται χρησιμοποιώντας το F - test του πίνακα ANOVA. Εάν ίσχυε η H_0 , η τιμή $F = 6.791$ θα έπρεπε να προέρχεται από την κατανομή $F_{3,15}$ (Snedecor με 3 και 15 β.ε.). Το αντίστοιχο p-value είναι μόλις 0.004 (χοντρικά, μόλις στο 0.004 των περιπτώσεων λαμβάνουμε ένα τόσο «ακραίο» δείγμα υπό την H_0) και επομένως απορρίπτουμε την H_0 (ε.σ. 1%). Οι μέσες τιμές του δείκτη αντοχής στον πόνο δεν είναι ίσες σε όλες τις στάθμες του παράγοντα «χρώμα μαλλιών»

(3) Επιλέγουμε Graphs/error bars/simple, variable: SCORE, Category Axis: HCOLOR (c.i. for mean, level: 95%) από όπου λαμβάνουμε το γράφημα με τα δ.ε. για τα μ_i (το καθένα συντελεστού 95% με βάση τα individual variance estimates).



Από το γράφημα φαίνεται και πάλι μια τάση μείωσης του μέσου δείκτη αντοχής όσο πάμε προς μελαχρινότερα άτομα. Θα μπορούσε όμως π.χ. ο μέσος δείκτης να είναι ίσος μεταξύ των ξανθών (στάθμη 1,2) και ίσος μεταξύ των μελαχρινών ατόμων (στάθμη 3,4) ενώ να υπάρχει διαφορά μεταξύ των ξανθών (1,2) και μελαχρινών (3,4). Για να μπορέσουμε να κάνουμε μια τέτοια «ομαδοποίηση» των μέσων θα πρέπει να προχωρήσουμε σε κάποια μέθοδο πολλαπλών συγκρίσεων.

(4) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: SCORE Factor: HCOLOR, : Post Hoc: LSD απ' όπου λαμβάνουμε τις συγκρίσεις όλων των μέσων ανά δύο (κάθε δ.ε. για τη διαφορά των μέσων είναι συντελεστού 95%):

Multiple Comparisons

Dependent Variable: SCORE

LSD

(I) HCOLOR	(J) HCOLOR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	8,00	5,17	,143	-3,02	19,02
	3	16,70*	5,48	,008	5,02	28,38
	4	21,80*	5,17	,001	10,78	32,82
2	1	-8,00	5,17	,143	-19,02	3,02
	3	8,70	5,48	,133	-2,98	20,38
	4	13,80*	5,17	,017	2,78	24,82
3	1	-16,70*	5,48	,008	-28,38	-5,02
	2	-8,70	5,48	,133	-20,38	2,98
	4	5,10	5,48	,367	-6,58	16,78
4	1	-21,80*	5,17	,001	-32,82	-10,78
	2	-13,80*	5,17	,017	-24,82	-2,78
	3	-5,10	5,48	,367	-16,78	6,58

*. The mean difference is significant at the .05 level.

Ο παραπάνω πίνακας περιέχει σε κάθε γραμμή τις ποσότητες

Mean Difference	s.e.	Sig.	LB, UB
$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$	$2(1 - F_{t_{n-k}}(\frac{ \bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} }{s.e.}))$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm s.e. \cdot t_{n-k}(\frac{\alpha}{2})$

Ένα 95% δ.ε. για τη διαφορά $\mu_1 - \mu_4$ είναι το (10.78, 32.82) ενώ ένα 95% δ.ε. για τη διαφορά $\mu_3 - \mu_4$ είναι το (-6.58, 16.78). Το μ_4 διαφέρει από το μ_1 σε ε.σ. 5% (διότι το αντίστοιχο p-value είναι 0.001) ενώ δεν έχουμε αρκετά στοιχεία για να πούμε ότι το μ_3 διαφέρει από το μ_4 (p-value = 0.367). (οι έλεγχοι μπορούν να γίνουν σε ε.σ. 5% με βάση και τα δ.ε. συντ. 95%).

(5) Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: SCORE Factor: HCOLOR, : Post Hoc: Bonferroni, Tukey απ' όπου λαμβάνουμε τις συγκρίσεις όλων των μέσων ανά δύο με την μέθοδο Bonferroni (ένα σύνολο από ταυτόχρονα δ.ε. για τις διαφορές $\mu_i - \mu_j$ με σν-

ολικό συντελεστή εμπιστοσύνης τουλάχιστον $1 - \alpha$) και τη μέθοδο Tukey (ταυτόχρονα δ.ε. με συνολικό συντελεστή εμπιστοσύνης περίπου $1 - \alpha$ για ανισομεγέθη δείγματα).

Multiple Comparisons

Dependent Variable: SCORE

	(I) HCOLOR	(J) HCOLOR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	8,00	5,17	,436	-6,90	22,90
		3	16,70*	5,48	,037	,90	32,50
		4	21,80*	5,17	,004	6,90	36,70
	2	1	-8,00	5,17	,436	-22,90	6,90
		3	8,70	5,48	,415	-7,10	24,50
		4	13,80	5,17	,074	-1,10	28,70
	3	1	-16,70*	5,48	,037	-32,50	-,90
		2	-8,70	5,48	,415	-24,50	7,10
		4	5,10	5,48	,789	-10,70	20,90
	4	1	-21,80*	5,17	,004	-36,70	-6,90
		2	-13,80	5,17	,074	-28,70	1,10
		3	-5,10	5,48	,789	-20,90	10,70
Bonferroni	1	2	8,00	5,17	,855	-7,69	23,69
		3	16,70*	5,48	,049	5,46E-02	33,35
		4	21,80*	5,17	,004	6,11	37,49
	2	1	-8,00	5,17	,855	-23,69	7,69
		3	8,70	5,48	,800	-7,95	25,35
		4	13,80	5,17	,105	-1,89	29,49
	3	1	-16,70*	5,48	,049	-33,35	-5,46E-02
		2	-8,70	5,48	,800	-25,35	7,95
		4	5,10	5,48	1,000	-11,55	21,75
	4	1	-21,80*	5,17	,004	-37,49	-6,11
		2	-13,80	5,17	,105	-29,49	1,89
		3	-5,10	5,48	1,000	-21,75	11,55

*. The mean difference is significant at the .05 level.

Ο παραπάνω πίνακας περιέχει σε κάθε γραμμή που αφορά τη μέθοδο Tukey τις ποσότητες

Mean Difference	s.e.	Sig.	LB, UB
$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$	$2(1 - F_{R(k, n-k)}(\sqrt{2} \frac{ \bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} }{s.e.}))$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm s.e. \cdot \frac{q(k, n-k, \alpha)}{\sqrt{2}}$

ενώ σε κάθε γραμμή που αφορά τη μέθοδο Bonferroni περιέχει τις ποσότητες ($l = k(k-1)/2$)

Mean Difference	s.e.	Sig.	LB, UB
$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$	$S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$	$\min\{l \cdot 2(1 - F_{t_{n-k}}(\frac{ \bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} }{s.e.}))\}$	$\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet} \pm s.e. \cdot t_{n-k}(\frac{\alpha}{2l})$

Παρατηρούμε ότι, σύμφωνα με την μέθοδο Tukey, υπάρχει στατιστικά σημαντική διαφορά (ε.σ. 5%) μεταξύ των μέσων μ_1 και μ_3 , και μεταξύ των μέσων μ_1 και μ_4 . Για όλες τις υπόλοιπες συγκρίσεις δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε την ισότητα.

Επομένως ο μέσος μ_1 διαφέρει από τους μ_3, μ_4 (που μπορεί να είναι ίσοι) ενώ ο μ_2 μπορεί να είναι ίσος είτε με τον μ_1 είτε με τους μ_3, μ_4 ($\bar{Y}_{1\bullet} \geq \bar{Y}_{2\bullet} \geq \bar{Y}_{3\bullet} \geq \bar{Y}_{4\bullet}$). Άρα τα άτομα με ανοιχτό ξανθό χρώμα μαλλιών (στάθμη 1) έχουν μεγαλύτερη ανοχή στον πόνο από ότι άτομα με μελαχρινό (ανοιχτό ή σκούρο) χρώμα μαλλιών (στάθμες 3,4). Για τις υπόλοιπες συγκρίσεις δεν μπορούμε να αποφανθούμε.

Με βάση τα παραπάνω προκύπτουν δύο «πιθανές» ομογενείς ομάδες μέσων: η ομάδα $\{\mu_1, \mu_2\}$ και η ομάδα $\{\mu_2, \mu_3, \mu_4\}$ οι οποίες δίνονται και από το SPSS:

Homogeneous Subsets

SCORE

HCOLOR	N	Subset for alpha = .05	
		1	2
Tukey HSD ^{a,b} 4	5	37,40	
3	4	42,50	
2	5	51,20	51,20
1	5		59,20
Sig.		,086	,461

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 4,706.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Το πρώτο p-value που δίνεται στον πίνακα (0.086) αφορά την ομογένεια της πρώτης ομάδας των μέσων μ_2, μ_3, μ_4 . Αντιστοιχεί στον έλεγχο της υπόθεσης $H_0: \mu_2 = \mu_4$ (και επειδή το μ_3 εκτιμάται μεταξύ των μ_1, μ_2 η υπόθεση αυτή είναι ισοδύναμη με την $\mu_2 = \mu_3 = \mu_4$) μέσω της μεθόδου Tukey θεωρώντας ως πλήθος παρατηρήσεων m σε κάθε στάθμη ίσο με τον αρμονικό μέσο του πλήθους των παρατηρήσεων σε όλες τις κλάσεις (δηλ $m = ((1/5+1/4+1/5+1/5)/4)^{-1} = 4.706$) δηλ.

$$p\text{-value} = 2(1 - F_{R(k,n-k)}(\frac{|\bar{Y}_{2\bullet} - \bar{Y}_{4\bullet}|}{S/\sqrt{m}})) = 0.086$$

Το δεύτερο p-value (0.461) αφορά αντίστοιχα την ομογένεια της δεύτερης ομάδας των μέσων μ_1, μ_2 . Αντιστοιχεί στον έλεγχο της υπόθεσης $H_0: \mu_1 = \mu_2$ μέσω της μεθόδου Tukey και πάλι με βάση τον αρμονικό μέσο m .

(6) Για να ελέγξουμε αν οι διασπορές των παρατηρήσεων είναι ίδιες σε όλες τις στάθμες του παράγοντα χρησιμοποιούμε το τεστ Levene. Επιλέγουμε Analyze/compare means/one-way ANOVA, Dependent list: SCORE Factor: HCOLOR, Options: check Homogeneity of variance από όπου παίρνουμε (βλ. παρ. 7.5.)

Test of Homogeneity of Variances

SCORE

Levene Statistic	df1	df2	Sig.
,493	3	15	,692

Δεν μπορούμε να απορρίψουμε ότι οι διασπορές των παρατηρήσεων είναι ίσες σε όλες τις στάθμες (p-value=0.692)

(7) Για να εφαρμόσουμε ένα γραμμικό μοντέλο της μορφής

$$SCORE = b_0 + b_1 \cdot HCOLOR + \varepsilon$$

θα πρέπει η HCOLOR να είναι scaled. Σε αυτή την περίπτωση η HCOLOR είναι απλώς ordinal (μπορούμε να διατάξουμε τις 4 στάθμες της). Θα μπορούσαμε να την θεωρήσουμε scaled με τιμές 1,2,3,4 αλλά προφανώς οι τιμές αυτές είναι εντελώς αυθαίρετες και η σχέση που έχουν μεταξύ τους δεν συμφωνεί απαραίτητα με την σχέση που έχουν οι 4 κλάσεις του χρώματος μαλλιών με συνέπεια να οδηγηθούμε σε αμφίβολα αποτελέσματα.