Στατιστική

Ανάλυση Διασποράς με ένα Παράγοντα

One-Way Anova

Χατζόπουλος Σταύρος

Κεφάλαιο 8ο. Ανάλυση Διασποράς

8.1 Εισαγωγή

- Η ανάλυση διασποράς είναι ένα σύνολο στατιστικών μεθόδων και μοντέλων που ασχολούνται με τις διαφορές των μέσων τιμών μίας μεταβλητής σε ομάδες παρατηρήσεων. Ίσως το σωστότερο όνομα να ήταν ανάλυση μέσων τιμών και όχι ανάλυση διασποράς. Θα διαπιστωθεί παρακάτω ότι όλες οι μέθοδοι που χρησιμοποιούνται βασίζονται σε λόγους διασπορών και για αυτό επιλέχθηκε το όνομα ανάλυση διασποράς ή ANOVA.

- Οι μέθοδοι που σχετίζονται άμεσα με την ανάλυση διασποράς είναι ο έλεγχος μέσων τιμών (t test) και η παλινδρόμηση. Όταν μελετάται η διαφορά των μέσων τιμών δύο ομάδων επιλέγεται ο έλεγχος μέσω του στατιστικού *t*.

- Το t test είναι η απλούστερη από τις διαδικασίες της ανάλυσης διασποράς διότι αφορά τον έλεγχο της διαφοράς των μέσων τιμών δύο ομάδων παρατηρήσεων. Αποδεικνύεται ότι το τετράγωνο του στατιστικού *t* που χρησιμοποιείται για τον έλεγχο της διαφοράς των μέσων τιμών δύο ομάδων ισούται με το στατιστικό *F* που χρησιμοποιείται κατά τη διαδικασία της ανάλυσης διασποράς.

- Συνεπώς η ανάλυση διασποράς, η οποία χρησιμοποιείται για να ελεγχθεί η υπόθεση ότι οι μέσες τιμές τριών η περισσοτέρων ομάδων διαφέρουν ή όχι, αποτελεί επέκταση του t test. Κάποιος θα σκεφτόταν βέβαια να πραγματοποιήσει μια σειρά από t tests ώστε να ελέγξει αν οι μέσες τιμές των ομάδων διαφέρουν, γεγονός που επηρεάζει το σφάλμα τύπου Ι, δηλαδή την απόρριψη της μηδενικής υπόθεσης ενώ είναι αληθής.

- Πιο αναλυτικά, στην ανάλυση διασποράς υπάρχει μια εξαρτημένη μεταβλητή και πλήθος ανεξάρτητων οι οποίες ονομάζονται παράγοντες.

- Η εξαρτημένη μεταβλητή πρέπει να είναι ποσοτική, τουλάχιστον διαστήματος (interval), δηλαδή η απόσταση μεταξύ τιμών της κλίμακας μέτρησης πρέπει να είναι ίδια για όλα τα σημεία της κλίμακας, ενώ οι ανεξάρτητες είναι κατηγορικές μεταβλητές τις κατηγορίες των οποίων τις ονομάζουμε επίπεδα ή στάθμες (levels).

- Αυτός είναι και ο λόγος που καθιστά την ανάλυση διασποράς μια μέθοδο ισοδύναμη με την παλινδρόμηση.

8.2 Προϋποθέσεις για την εφαρμογή της Ανάλυσης Διασποράς

Πριν προχωρήσει κάποιος στην εφαρμογή της μεθόδου πρέπει να ελέγξει αν οι παρακάτω προϋποθέσεις παραβιάζονται ή όχι.

- Τα δείγματα πρέπει να προέρχονται από πληθυσμούς με κανονική κατανομή.

- Οι πληθυσμοί έχουν κοινή διασπορά, δηλαδή πρέπει να γίνεται έλεγχος ομοιογένειας των διασπορών.

- Τα δείγματα πρέπει να είναι ανεξάρτητα μεταξύ τους.

- Σε κάθε δείγμα οι παρατηρήσεις πρέπει να επιλέγονται τυχαία και ανεξάρτητα η μία από την άλλη.

Αξίζει να σημειωθεί ότι μεγαλύτερη προσοχή πρέπει να δίνεται στο γεγονός της ύπαρξης ομοιογένειας διασπορών από ότι στην κανονικότητα των δειγμάτων των επιπέδων.

8.3 Ανάλυση Διασποράς με έναν Παράγοντα

Ας υποτεθεί ότι σε μια μελέτη υπάρχει μια εξαρτημένη μεταβλητή Y με n παρατηρήσεις η οποία επηρεάζεται από έναν παράγοντα. Οι τιμές του παράγοντα ονομάζονται επίπεδα (ομάδες) και έστω ότι είναι k το πλήθος. Το μοντέλο στη συγκεκριμένη περίπτωση δίνεται από τη σχέση:

$$y_{ij} = \mu + a_i + e_{ij}, \text{ όπου}$$

μ είναι ο γενικός μέσος

a_i είναι η κύρια επίδραση (main effect) του παράγοντα στο i επίπεδο, $1 \le i \le k$,

e_{ij} είναι το σφάλμα της j παρατήρησης στο i επίπεδο, $1 \le j \le n_i$. Να σημειωθεί ότι τα σφάλματα ακολουθούν την κανονική κατανομή $N(0, \sigma^2)$

n_i είναι το μέγεθος παρατηρήσεων του επιπέδου i. Ισχύει ότι $\sum_{i=1}^{k} n_i = n$.

Για την κατανόηση των μαθηματικών πράξεων που θα πραγματοποιηθούν παρακάτω απαραίτητοι είναι οι συμβολισμοί:

$\bar{y}_i$ μέση τιμή παρατηρήσεων στο επίπεδο i,

$\bar{y}$ μέση τιμή όλων των παρατηρήσεων.

Εκτιμήσεις των κύριων επιδράσεων των επιπέδων: $\hat{a}_i = \bar{y}_i - \bar{y}$, $\sum_{i=1}^{k} n_i \hat{a}_i = 0$.

Επιπλέον τα υπόλοιπα δίνονται από τη σχέση $e_{ij} = y_{ij} - \bar{y}_i$ και εύκολα διαπιστώνεται ότι

$\sum_{j=1}^{n_i} e_{ij} = 0$ για κάθε $1 \le i \le k$.

Ισχύει ότι $Var(Y) = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2}{n-1}$.

Έστω $SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$.

Αποδεικνύεται ότι

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{Y})^2=\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{y})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 .$$

Αν $SSB=\sum_{i=1}^{k}n_i(\bar{y}_i-\bar{y})^2$ και $SSW=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$ τότε ισχύει

$SST = SSB + SSW$, όπου

SST είναι η συνολική μεταβολή,

SSB είναι η μεταβολή των μέσων τιμών των επιπέδων από τη συνολική μέση τιμή (between groups variation) δηλαδή μεταξύ των ομάδων,

SSW είναι η μεταβολή των παρατηρήσεων από τη μέση τιμή των ομάδων στα οποία αντιστοιχούν (within groups variation), δηλαδή μέσα στις ομάδες.

Αν τα παραπάνω αθροίσματα τετραγώνων διαιρεθούν με τους βαθμούς ελευθερίας οι οποίοι αντιστοιχούν σε αυτά, δηλαδή το SSB με $k-1$ και το SSW με το $n-k$ προκύπτουν οι ποσότητες:

$MSB=\frac{SSB}{k-1}$: μέτρο μεταβλητότητας των μέσων τιμών των ομάδων. Υποδεικνύει το μέγεθος της συνολικής μεταβλητότητας που οφείλεται στις διαφορές των μέσων τιμών των ομάδων.

$MSW=\frac{SSW}{n-k}$: μέτρο μεταβλητότητας μέσα σε κάθε ομάδα γύρω από τη μέση τιμή του. Υποδεικνύει το μέγεθος της συνολικής μεταβλητότητας που οφείλεται στο τυχαίο σφάλμα.

Ένα μέτρο της σημαντικότητας του παράγοντα είναι ο συντελεστής $\eta^2 = \frac{SSB}{SST}$ ο οποίος μετριέται σε μονάδες επί τοις εκατό. Υποδηλώνει το ποσοστό της διασποράς που εξηγείται από τον παράγοντα. Είναι ένα μέτρο ισοδύναμο με τον συντελεστή προσδιορισμού R^2 που προκύπτει από την παλινδρόμηση.

Το γεγονός ότι η τιμή του βασίζεται στο δείγμα από το οποίο υπολογίζεται δεν επιτρέπει την γενίκευση του για τον πληθυσμό, το οποίο όμως συμβαίνει συχνά διότι είναι το μοναδικό μέτρο που προσφέρει το SPSS με συνέπεια αρκετοί ερευνητές να το χρησιμοποιούν. Υπερεκτιμά το ποσοστό διασποράς και για αυτό το λόγο χρησιμοποιείται συνήθως ένα εναλλακτικό μέτρο που διορθώνει την υπερεκτίμηση που υπολογίζεται από τον συντελεστή η^2.

Το μέτρο αυτό δίνεται από τη σχέση $\omega^2 = \frac{SSB - (k-1)MSW}{SST + MSW}$. Τιμές του ω^2 περίπου 0.01, 0.06 και 0.15 υποδηλώνουν αντίστοιχα μικρό, μεσαίο και μεγάλο μέγεθος επίδρασης του παράγοντα.

Όπως αναφέρθηκε στην εισαγωγή η ανάλυση διασποράς χρησιμοποιείται για να ελέγξει αν υπάρχουν διαφορές στις μέσες τιμές μ_1, μ_2, …, μ_k των k επιπέδων των πληθυσμών από τους οποίους προέρχονται τα δείγματα.

Πριν προχωρήσει κάποιος στην εφαρμογή της ανάλυσης διασποράς αναγκαίο είναι να πραγματοποιήσει έλεγχο ομοιογένειας διασπορών. Ας είναι σ_1^2, σ_2^2, …, σ_k^2 οι διασπορές των k επιπέδων των πληθυσμών από τους οποίους προέρχονται τα δείγματα.

Ο έλεγχος ομοιογένειας των διασπορών με τη βοήθεια του Levene test διατυπώνεται ως εξής:

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$,

H_1: $\sigma_i^2 \neq \sigma_j^2$ για ένα τουλάχιστον ζεύγος διασπορών των επιπέδων i και j.

Έστω $z_{ij} = |e_{ij}| = |y_{ij} - \bar{y}_i|$, για $1 \le i \le k$ και $1 \le j \le n_i$, $\bar{z}_i$ η μέση τιμή των z_{ij} για κάθε επίπεδο i και $\bar{z}$ η μέση τιμή των z_{ij}.

Αποδεικνύεται ότι το στατιστικό $W = \dfrac{(n-k)\sum_{i=1}^{k} n_i(\bar{z}_i - \bar{z})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_i)^2}$ ακολουθεί την κατανομή $F_{k-1,n-k}$.

Η απορριπτική περιοχή της μηδενικής υπόθεσης είναι: $R = \{W > F_{k-1,n-k;a}\}$.

Ο έλεγχος **υπόθεσης της ανάλυσης διασποράς** με ένα παράγοντα διατυπώνεται ως εξής:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ ή $a_1 = a_2 = \ldots = a_k = 0$

H_1: $\mu_i \neq \mu_j$ ή $a_i \neq a_j$ για ένα τουλάχιστον ζεύγος των επιπέδων i και j.

Αποδεικνύεται ότι το στατιστικό $F = \dfrac{MSB}{MSW}$ ακολουθεί την κατανομή $F_{k-1,n-k}$, εφόσον ικανοποιούνται οι προϋποθέσεις για την εφαρμογή της ανάλυσης διασποράς.

Συνεπώς η **απορριπτική περιοχή της μηδενικής υπόθεσης είναι**: $R = \{F > F_{k-1,n-k;a}\}$.

Συνοψίζοντας τους παραπάνω υπολογισμούς σε έναν πίνακα προκύπτει ο πίνακας της ανάλυσης διασποράς για έναν παράγοντα ο οποίος έχει τη μορφή:

Πηγή Μεταβολής	**Αθροίσματα Τετραγώνων (SS)**	**Βαθμοί Ελευθερίας (β.ε.)**	**Μέση Μεταβολή (MS)**	**Στατιστικό F**
Μεταξύ των δειγμάτων (between groups)	$SSB = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$	$k-1$	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
Μέσα στα δείγματα (within groups)	$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$	$n-k$	$MSW = \frac{SSW}{n-k}$	
Συνολική Μεταβολή	$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$	$n-1$		

Παράδειγμα

Ο λογιστής μιας εταιρείας προκειμένου να ελέγξει αν υπάρχει διαφορά στις μέσες τιμές των μισθών των υπαλλήλων σε σχέση με το επίπεδο εκπαίδευσής τους έλαβε δείγματα 5 υπαλλήλων από κάθε επίπεδο εκπαίδευσης και κατέγραψε το μηνιαίο μισθό τους (σε εκατοντάδες €). Τα δεδομένα δίνονται στον παρακάτω πίνακα:

Δημοτικό	**Γυμνάσιο**	**Λύκειο**	**ΑΕΙ / ΤΕΙ**
10	7	13	15
11	8	14	16
9	7	15	10
10	11	14	16
10	12	14	18

Υπάρχει διαφορά του μέσου μισθού ανάλογα με το επίπεδο εκπαίδευσης των υπαλλήλων; Δίνεται a = 0.05.

Το μοντέλο της ανάλυσης διασποράς στην συγκεκριμένη περίπτωση δίνεται από τη σχέση:

$y_{ij} = \mu + a_i + e_{ij}$, όπου $1 \le i \le 4$ και $1 \le j \le 5$.

Δημοτικό	Γυμνάσιο	Λύκειο	ΑΕΙ / ΤΕΙ
$y_{11}=10=\mu+a_1+e_{11}$	$y_{21}=7=\mu+a_2+e_{21}$	$y_{31}=13=\mu+a_3+e_{31}$	$y_{41}=15=\mu+a_4+e_{41}$
$y_{12}=11=\mu+a_1+e_{12}$	$y_{22}=8=\mu+a_2+e_{22}$	$y_{32}=14=\mu+a_3+e_{32}$	$y_{42}=16=\mu+a_4+e_{42}$
$y_{13}=9=\mu+a_1+e_{13}$	$y_{23}=7=\mu+a_2+e_{23}$	$y_{33}=15=\mu+a_3+e_{33}$	$y_{43}=10=\mu+a_4+e_{43}$
$y_{14}=10=\mu+a_1+e_{14}$	$y_{24}=11=\mu+a_2+e_{24}$	$y_{34}=14=\mu+a_3+e_{34}$	$y_{44}=16=\mu+a_4+e_{44}$
$y_{15}=10=\mu+a_1+e_{15}$	$y_{25}=12=\mu+a_2+e_{25}$	$y_{35}=14=\mu+a_3+e_{35}$	$y_{45}=18=\mu+a_4+e_{45}$

Διαπιστώνεται ότι $n_1=n_2=n_3=n_4=5$ και $n=20$.

Επιπλέον $\bar{y}_1=10$, $\bar{y}_2=9$, $\bar{y}_3=14$, $\bar{y}_4=15$, $\bar{y}=12$,

$$s_1=\sqrt{\frac{1}{4}\sum_{j=1}^{5}(y_{1j}-10)^2}=0.707, \qquad s_2=\sqrt{\frac{1}{4}\sum_{j=1}^{5}(y_{2j}-9)^2}=2.342,$$

$$s_3=\sqrt{\frac{1}{4}\sum_{j=1}^{5}(y_{3j}-14)^2}=0.707, \qquad s_4=\sqrt{\frac{1}{4}\sum_{j=1}^{5}(y_{4j}-15)^2}=3.000.$$

Τα τυπικά σφάλματα $SE_i=\frac{s_i}{\sqrt{n_i}}$ τα οποία βοηθούν στην κατασκευή των διαστημάτων εμπιστοσύνης των μέσων τιμών είναι:

$$SE_1=\frac{s_1}{\sqrt{n_1}}=0.316, \qquad SE_2=\frac{s_2}{\sqrt{n_2}}=1.049,$$

$$SE_3=\frac{s_3}{\sqrt{n_3}}=0.316, \qquad SE_4=\frac{s_4}{\sqrt{n_4}}=1.342.$$

Τα (1-a)% διαστήματα εμπιστοσύνης για τις μέσες τιμές δίνονται από τη σχέση $\bar{y}_i \pm t_{n_i-1,a/2}\cdot SE_i$, όπου $t_{n_i-1,a/2}=t_{4;0.025}=2.776$.

Επομένως τα 95% δ.ε. για τις μέσες τιμές μ_1, μ_2, μ_3 και μ_4 είναι αντίστοιχα

$(9.12, 10.88)$, $(6.09, 11.91)$,

$(13.12, 14.88)$ και $(11.28, 18.72)$.

- Διαπιστώνεται ότι κατά τον υπολογισμό των δ.ε. για τις μέσες τιμές το τυπικό σφάλμα δίνεται από τη σχέση $SE_i = \frac{s_i}{\sqrt{n_i}}$, όπου s_i^2 είναι ο εκτιμητής της διασποράς σ_i^2 της κάθε ομάδας.

- Στην περίπτωση της ανάλυσης διασποράς ο εκτιμητής της διασποράς είναι η ποσότητα MSW εφόσον οι διασπορές των ομάδων είναι ίσες. Σε αυτή την περίπτωση τα (1-a)% διαστήματα εμπιστοσύνης για τις μέσες τιμές δίνονται από τη σχέση

$$\bar{y}_i \pm t_{n-k;a/2} \cdot \sqrt{\frac{MSW}{n_i}}.$$

Για το συγκεκριμένο παράδειγμα $t_{n-k;a/2} = t_{16;0.025} = 2.12$.

Θα διαπιστωθεί παρακάτω ότι $MSW = 3.875$.

Επομένως τα 95% δ.ε. για τις μέσες τιμές μ_1, μ_2, μ_3 και μ_4 είναι αντίστοιχα:

$(8.134, 11.866)$, $(7.134, 10.866)$,

$(12.134, 15.866)$ και $(13.134, 16.866)$.

Προηγουμένως, χρησιμοποιώντας τα τυπικά σφάλματα των επιπέδων του παράγοντα, βρήκαμε ότι τα 95% δ.ε. είναι

$(9.12, 10.88)$, $(6.09, 11.91)$,

$(13.12, 14.88)$ και $(11.28, 18.72)$.

Οι εκτιμήσεις των επιδράσεων των επιπέδων δίνονται από τη σχέση:

$$\hat{a}_i = \bar{y}_i - \bar{y} .$$

Επομένως $\hat{a}_1 = \bar{y}_1 - \bar{y} = 10 - 12 = -2$, $\hat{a}_2 = -3$,

$\hat{a}_3 = 2$ και $\hat{a}_4 = 3$.

Προφανώς ισχύει ότι $\sum_{i=1}^{4} n_i \hat{a}_i = 5(-2-3+2+3) = 0$.

Τα υπόλοιπα δίνονται από τη σχέση:

$$e_{ij} = y_{ij} - \bar{y}_i$$

Δημοτικό	Γυμνάσιο	Λύκειο	ΑΕΙ / ΤΕΙ
$e_{11} = y_{11} - \bar{y}_1 = 0$	$e_{21} = y_{21} - \bar{y}_2 = -2$	$e_{31} = y_{31} - \bar{y}_3 = -1$	$e_{41} = y_{41} - \bar{y}_4 = 0$
$e_{12} = y_{12} - \bar{y}_1 = 1$	$e_{22} = y_{22} - \bar{y}_2 = -1$	$e_{32} = y_{32} - \bar{y}_3 = 0$	$e_{42} = y_{42} - \bar{y}_4 = 1$
$e_{13} = y_{13} - \bar{y}_1 = -1$	$e_{23} = y_{23} - \bar{y}_2 = -2$	$e_{33} = y_{33} - \bar{y}_3 = 1$	$e_{43} = y_{43} - \bar{y}_4 = -5$
$e_{14} = y_{14} - \bar{y}_1 = 0$	$e_{24} = y_{24} - \bar{y}_2 = 2$	$e_{34} = y_{34} - \bar{y}_3 = 0$	$e_{44} = y_{44} - \bar{y}_4 = 1$
$e_{15} = y_{15} - \bar{y}_1 = 0$	$e_{25} = y_{25} - \bar{y}_2 = 3$	$e_{35} = y_{35} - \bar{y}_3 = 0$	$e_{45} = y_{45} - \bar{y}_4 = 3$

Εύκολα, αποδεικνύεται ότι

$$\sum_{j=1}^{5} e_{1j} = \sum_{j=1}^{5} e_{2j} = \sum_{j=1}^{5} e_{3j} = \sum_{j=1}^{5} e_{4j} = 0 \quad \text{και} \quad \sum_{i=1}^{4}\sum_{j=1}^{5} e_{ij} = 0 .$$

Τα τυποποιημένα υπόλοιπα (standardized residuals) θα υπολογιστούν με τη βοήθεια της σχέσης:

$$\tilde{e}_{ij} = \frac{y_{ij} - \bar{y}}{\sqrt{\text{MSW}}}$$

Μετά από πράξεις διαπιστώνεται ότι:

Δημοτικό	Γυμνάσιο	Λύκειο	ΑΕΙ / ΤΕΙ
0.000	-1.016	-0.508	0.000
0.508	-0.508	0.000	0.508
-0.508	-1.016	0.508	-2.540
0.000	1.016	0.000	0.508
0.000	1.524	0.000	1.524

Ισχύει ότι $|\tilde{e}_{43}| > 2.5$ με συνέπεια η παρατήρηση $y_{43} = 10$ να **θεωρείται παράτυπο σημείο**.

Ο έλεγχος ομοιογένειας των διασπορών για το παράδειγμα διατυπώνεται ως εξής:

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

$H_1 : \sigma_i^2 \neq \sigma_j^2$ για ένα τουλάχιστον ζεύγος i και j.

Το στατιστικό $W = \dfrac{16\sum_{i=1}^{k} 5(\bar{z}_i - \bar{z})^2}{3\sum_{i=1}^{k}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_i)^2}$ βασίζεται στην εύρεση των τιμών

$$z_{ij} = |e_{ij}| = |y_{ij} - \bar{y}|, \text{ για } 1 \le i \le 4 \text{ και } 1 \le j \le 5$$

οι οποίες δίνονται στον ακόλουθο πίνακα

Δημοτικό	Γυμνάσιο	Λύκειο	ΑΕΙ / ΤΕΙ
0	2	1	0
1	1	0	1
1	2	1	5
0	2	0	1
0	3	0	3

Εύκολα υπολογίζεται ότι

$\bar{z}_1 = 0.4$, $\bar{z}_2 = 2$, $\bar{z}_3 = 0.4$,

$\bar{z}_4 = 2$ και $\bar{z} = 1.2$ $\sum_{i=1}^{4}(\bar{z}_i - \bar{z})^2 = 4 \cdot 0.64 = 2.56$

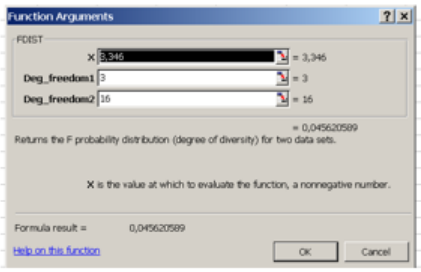
Επιπλέον

$$\sum_{i=1}^{4}\sum_{j=1}^{5}(z_{ij} - \bar{z}_i)^2 = \sum_{j=1}^{5}(z_{1j} - 0.4)^2 + \sum_{j=1}^{5}(z_{2j} - 2)^2 + \sum_{j=1}^{5}(z_{3j} - 0.4)^2 + \sum_{j=1}^{5}(z_{4j} - 2)^2 = 20.4 .$$

Επομένως $W = \dfrac{16 \cdot 5 \cdot 2.56}{3 \cdot 20.4} = 3.346$.

Η τιμή αυτή θα συγκριθεί με την τιμή $F_{3,16;0.05} = 3.24$ και προφανώς $W > F_{3,16;0.05}$ με συνέπεια να απορρίπτεται η μηδενική υπόθεση σε επίπεδο σημαντικότητας 0.05 που αφορά την ισότητα των διασπορών των τεσσάρων ομάδων. Η σημαντικότητα του ελέγχου δίνεται από τη σχέση $\alpha = P(F > W) = P(F > 3.346)$ και θα υπολογιστεί μέσω της κατανομής F, η οποία απαιτεί τη γνώση των $k-1=3$ και $n-k=16$ βαθμών ελευθερίας.

Το Excel προσφέρει τον υπολογισμό της πιθανότητας μέσω της συνάρτησης FDIST, όπως προκύπτει από το ακόλουθο παράθυρο εργασίας:



Επομένως η πιθανότητα απόρριψης της μηδενικής υπόθεσης ενώ είναι αληθής είναι 0.046, πολύ κοντά δηλαδή στην τιμή 0.05 με συνέπεια να μπορεί κάποιος να ισχυριστεί ότι οι διασπορές είναι ίσες σε επίπεδο σημαντικότητας 0.045.

Για τον έλεγχο της υπόθεσης της ανάλυσης διασποράς, δηλαδή της ισότητας των μεσών όρων, δηλαδή:

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

$H_1: \mu_i \neq \mu_j$ για ένα τουλάχιστον ζεύγος i και j

απαραίτητοι είναι οι υπολογισμοί των SST, SSB και SSW. Για το παράδειγμα μετά από πράξεις διαπιστώνεται ότι

$$SSB = \sum_{i=1}^{4} 5(\bar{y}_i - \bar{y})^2 = 5(4+9+4+9) = 130,$$

$$SSW = \sum_{i=1}^{4}\sum_{j=1}^{5}(y_{ij} - \bar{y}_i)^2 = 2+22+2+36 = 62$$

$$SST = \sum_{i=1}^{4}\sum_{j=1}^{5}(y_{ij} - \bar{y})^2 = 22+67+22+81 = 192 = SSB + SSW .$$

$$MSB = \frac{SSB}{3} = 43.333 \quad \text{και} \quad MSW = \frac{SSW}{16} = 3.875,$$

Πηγή Μεταβολής	Αθρ. Τετρ.(SS)	β.ε.	MS	Στατιστικό F
between groups	$SSB = 130$	3	$MSB = 43.333$	$F = \frac{MSB}{MSW} = \frac{43.333}{3.875} = 11.183$
within groups	$SSW = 62$	16	$MSW = 3.875$	
Σύνολο	$SST = 192$	19		

Πηγή Μεταβολής	Αθρ. Τετρ.(SS)	β.ε.	MS	Στατιστικό F
between groups	$SSB = 130$	3	$MSB = 43.333$	$F = \frac{MSB}{MSW} = \frac{43.333}{3.875} = 11.183$
within groups	$SSW = 62$	16	$MSW = 3.875$	
Σύνολο	$SST = 192$	19		

Η απορριπτική περιοχή της μηδενικής υπόθεσης είναι:

$$R = \{F > F_{3,16;0.05}\},$$

όπου $F_{3,16;0.05} = 3.24$

και προφανώς η μηδενική υπόθεση **απορρίπτεται**. Επομένως υπάρχει διαφορά του μέσου μισθού ανάλογα με το επίπεδο εκπαίδευσης των υπαλλήλων.

Ως μέτρο σημαντικότητας του παράγοντα επίπεδο εκπαίδευσης υπαλλήλων ορίστηκε ο συντελεστής $\eta^2 = \frac{SSB}{SST}$ του οποίου η τιμή είναι

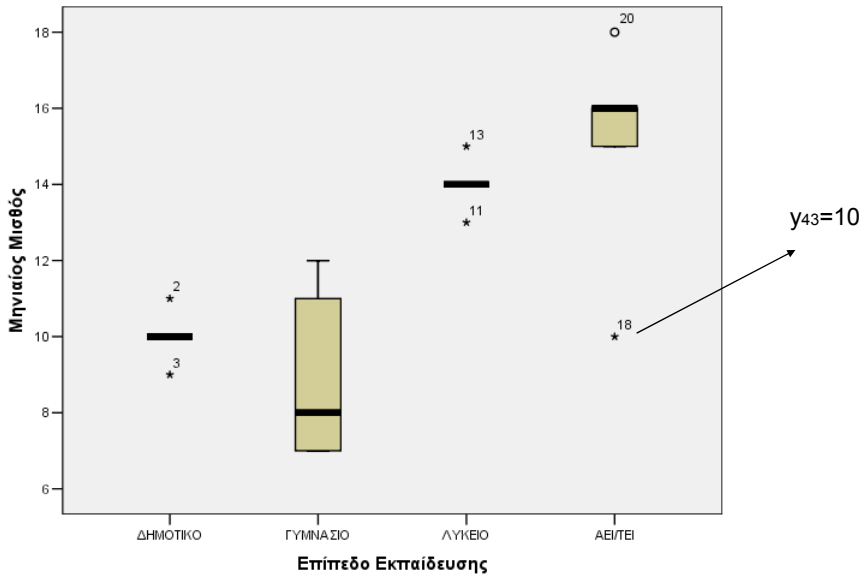
$$\eta^2 = \frac{130}{192} = 0.677$$

και υποδεικνύει ότι το ποσοστό της διασποράς που εξηγείται από τον παράγοντα επίπεδο εκπαίδευσης είναι 67.7%.

Επίσης ορίστηκε και ο συντελεστής $\omega^2 = \frac{SSB - (k-1)MSW}{SST + MSW}$ ως ένα εναλλακτικό μέτρο που διορθώνει την υπερεκτίμηση που υπολογίζεται από τον συντελεστή η^2. Υπολογίζεται ότι $\omega^2 = \frac{130 - 3 \cdot 3.875}{192 + 3.875} = 0.604$.

Προϋποθέσεις εφαρμογής της ανάλυσης διασποράς

Κανονικότητα και Παράτυπα σημεία



Προϋποθέσεις εφαρμογής της ανάλυσης διασποράς

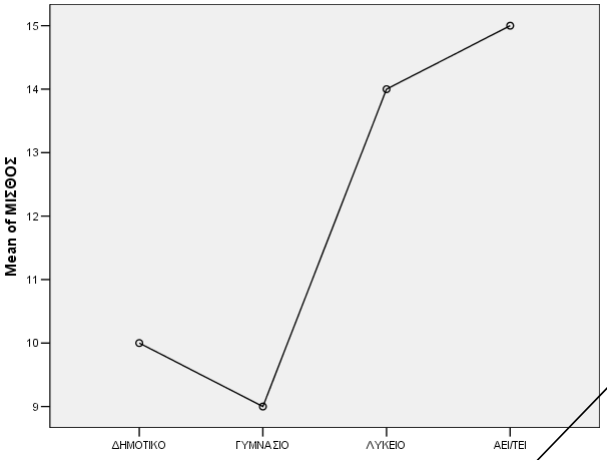
Κανονικότητα με Tests of Normality (διαδικασία Explore)

Από τους ελέγχους κανονικότητας προκύπτει ότι η μηδενική υπόθεση ότι οι παρατηρήσεις κάθε επιπέδου προέρχονται από κανονικά κατανεμημένους πληθυσμούς γίνεται δεκτή (σημαντικότητα >> 0.05).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Επίπεδο Εκπαίδευσης	Statistic	df	Sig.	Statistic	df	Sig.
Μηνιαίος Μισθός	ΔΗΜΟΤΙΚΟ	,300	5	,161	,883	5	,325
	ΓΥΜΝΑΣΙΟ	,265	5	,200*	,836	5	,154
	ΛΥΚΕΙΟ	,300	5	,161	,883	5	,325
	ΑΕΙ/ΤΕΙ	,300	5	,161	,858	5	,222

Γραφική παράσταση μέσων τιμών



95% δ.ε. με χρήση των τυπικών σφαλμάτων και όχι του MSW

Μηνιαίος Μισθός

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
ΔΗΜΟΤΙΚΟ	5	10,00	,707	,316	9,12	10,88	9	11
ΓΥΜΝΑΣΙΟ	5	9,00	2,345	1,049	6,09	11,91	7	12
ΛΥΚΕΙΟ	5	14,00	,707	,316	13,12	14,88	13	15
ΑΕΙ/ΤΕΙ	5	15,00	3,000	1,342	11,28	18,72	10	18
Total	20	12,00	3,179	,711	10,51	13,49	7	18

Ανακεφαλαιώνοντας τις διάφορες μεθόδους που μπορούν να επιλεγούν ώστε να πραγματοποιηθεί έλεγχος ύπαρξης διαφορών στους μέσους όρους των επιπέδων ενός παράγοντα, υπό την προϋπόθεση ότι τα επίπεδα είναι ανεξάρτητα και οι παρατηρήσεις κάθε επιπέδου είναι ανεξάρτητες και τυχαία επιλεγμένες, προκύπτει ότι:

- Η μέθοδος της ανάλυσης διασποράς ενδείκνυται όταν τα δεδομένα προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή, υπάρχει ομοιογένεια διασπορών και δεν υπάρχουν παράτυπα σημεία. Οι μεταβλητές είναι ποσοτικές διαστήματος (μπορεί και κλίμακας).

- Στην περίπτωση που τα δεδομένα δεν προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή, υπάρχει ανομοιογένεια διασπορών και υπάρχουν παράτυπα σημεία απαιτείται μετασχηματισμός των δεδομένων.

- Η μέθοδος Kruskal-Wallis ενδείκνυται στην περίπτωση που τα δεδομένα δεν προέρχονται από πληθυσμούς που ακολουθούν κανονική κατανομή και οι μεταβλητές είναι μεταβλητές διάταξης.

8.5 Αντιθέσεις στην Ανάλυση Διασποράς με έναν Παράγοντα

Οι αντιθέσεις (contrasts) χρησιμεύουν για να ελεγχθούν υποθέσεις που αφορούν τις μέσες τιμές κάποιων από τα επίπεδα του παράγοντα είτε ένα συνδυασμό τους. Η αντίθεση είναι ένα σταθμισμένο άθροισμα μέσων τιμών. Για παράδειγμα κάποιος επιθυμεί να ελέγξει αν η μέση τιμή του μηνιαίου μισθού των υπαλλήλων που είναι απόφοιτοι Λυκείου είναι ίση με το μέσο όρο των μέσων τιμών των τριών άλλων επιπέδων. Να ελέγξει δηλαδή αν ισχύει ότι $\mu_3 = \frac{\mu_1 + \mu_2 + \mu_4}{3}$.

Χρησιμοποιώντας μαθηματικούς συμβολισμούς ας είναι c_i, $1 \le i \le k$, k το πλήθος βάρη (σταθμίσεις) τέτοια ώστε $\sum_{i=1}^{k} c_i = 0$ και $\boldsymbol{c} = (c_1, c_2, \text{K}, c_k)$ το διάνυσμα βαρών.

Η αντίθεση ορίζεται ως ο γραμμικός συνδυασμός $\psi = \sum_{i=1}^{k} c_i \mu_i$, όπου μ_i είναι οι μέσες τιμές των πληθυσμών από τους οποίους προέρχονται τα δείγματα των επιπέδων. Ως εκτιμητής της αντίθεσης ορίζεται η ποσότητα $\hat{\psi} = \sum_{i=1}^{k} c_i \bar{y}_i$, με $\bar{y}_i$ τις δειγματικές μέσες τιμές.

Να σημειωθεί ότι η επιλογή ορθογώνιων αντιθέσεων διευκολύνει τους υπολογισμούς. Δύο αντιθέσεις $\boldsymbol{c_1}$ και $\boldsymbol{c_2}$ ονομάζονται **ορθογώνιες** αν τα διανύσματα των σταθμίσεων ικανοποιούν τη συνθήκη $\boldsymbol{c_1 c_2'} = 0$, εφόσον τα μεγέθη δειγμάτων των k επιπέδων n_i είναι ίσα. Αναλυτικότερα αν $\boldsymbol{c}_1 = (c_{11}, c_{12}, \dots, c_{1k})$ και $\boldsymbol{c}_2 = (c_{21}, c_{22}, \dots, c_{2k})$ τότε πρέπει να ισχύει ότι

$$\sum_{h=1}^{k} c_{1h} c_{2h} = 0 .$$

Στην περίπτωση που τα μεγέθη δειγμάτων δεν είναι ίσα η συνθήκη που πρέπει να ικανοποιείται ώστε δύο αντιθέσεις να είναι ορθογώνιες διαμορφώνεται ως εξής:

$$\sum_{h=1}^{k} \frac{c_{1h} c_{2h}}{n_h} = 0 .$$

<u>Παρατηρήσεις</u>

1. Ο πίνακας ανάλυσης διασποράς κατασκευάζεται για ένα σύνολο $k-1$ ορθογώνιων αντιθέσεων.
2. Η απόρριψη της μηδενικής υπόθεσης της ανάλυσης διασποράς

   H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ συνεπάγεται την απόρριψη της μηδενικής υπόθεσης για μία τουλάχιστον αντίθεση, δηλαδή αν δεν υπάρχει διαφορά στις μέσες τιμές των επιπέδων τότε μία τουλάχιστον αντίθεση δεν είναι ίση με μηδέν.

3. Η απόρριψη της μηδενικής υπόθεσης για μία αντίθεση δεν συνεπάγεται την απόρριψη της μηδενικής υπόθεσης για τον έλεγχο της ανάλυσης διασποράς.
4. Αν γίνει δεκτή η μηδενική υπόθεση της ανάλυσης διασποράς τότε δεν σημαίνει ότι οι τιμές όλων των αντιθέσεων θα είναι ίσες με μηδέν.

Οι αντιθέσεις που χρησιμοποιούνται στην ανάλυση διασποράς για να πραγματοποιηθούν έλεγχοι υποθέσεων τους οποίους εκ των προτέρων έχει ορίσει ο ερευνητής έχουν τιμή η οποία υπολογίζεται από τη σχέση $\hat{\psi} = \sum_{i=1}^{k} c_i \bar{y}_i$.

Επιπλέον ορίζεται **τυπικό σφάλμα** για κάθε αντίθεση το οποίο εξαρτάται από το MSW, υπολογίζεται διάστημα εμπιστοσύνης, μπορεί να πραγματοποιηθεί έλεγχος της υπόθεσης ότι η τιμή της είναι ίση με μηδέν και υπολογίζεται η συνεισφορά της στο άθροισμα τετραγώνων των αποκλίσεων. Οι παραπάνω υπολογισμοί εξαρτώνται από το αν υπάρχει ή όχι ομοιογένεια διασπορών.

Τυπικό Σφάλμα

Η εύρεση των τυπικών σφαλμάτων εξαρτάται από τον εκτιμητή της διασποράς. **Αν οι διασπορές είναι ίσες** τότε το τυπικό σφάλμα της αντίθεσης δίνεται από τη σχέση

$$SE(\hat{\psi}_i) = \sqrt{MSW \sum_{i=1}^{k} \frac{c_i^2}{n_i}}.$$

Έλεγχοι Υποθέσεων

Ο έλεγχος υπόθεσης για την τιμή μιας αντίθεσης διατυπώνεται ως εξής:

$$H_0: \hat{\psi}_i = 0 \ , \qquad H_1: \hat{\psi}_i \neq 0 \ .$$

Αποδεικνύεται ότι το στατιστικό $t_i = \dfrac{\hat{\psi}_i}{SE(\hat{\psi}_i)}$ ακολουθεί την κατανομή $t_{\beta.\varepsilon.}$ και η απορριπτική περιοχή της μηδενικής υπόθεσης είναι $R = \left\{ |t| > t_{\beta.\varepsilon.,a/2} \right\}$, όπου οι βαθμοί ελευθερίας (β.ε.) στην περίπτωση ίσων διασπορών είναι $n-k$, ενώ στην περίπτωση άνισων διασπορών είναι $\dfrac{\left(\sum_{i=1}^{k} \frac{c_i^2}{n_i} s_i^2 \right)^2}{\sum_{i=1}^{k} \frac{\left(c_i^2 s_i^2 \right)^2}{n_i - 1}}$.

Διαστήματα Εμπιστοσύνης

Τα (1-a)% δ.ε. για τις τιμές των αντιθέσεων δίνονται από τη σχέση $\hat{\psi}_i \pm t_{\beta.\varepsilon.,a/2} \cdot SE(\hat{\psi}_i)$, όπου οι βαθμοί ελευθερίας δόθηκαν παραπάνω.

Αθροίσματα Τετραγώνων

Δίνονται από τη σχέση $SS(\hat{\psi}_i) = \dfrac{\hat{\psi}_i^2}{\sum_{i=1}^{k} \frac{c_i^2}{n_i}}$. Επιπλέον αν οριστούν $k-1$ ορθογώνιες αντιθέσεις τότε ισχύει ότι οι $SSB = \sum_{i=1}^{k-1} SS(\hat{\psi}_i)$.

Ο έλεγχος υποθέσεων για την τιμή μιας αντίθεσης μπορεί να πραγματοποιηθεί χρησιμοποιώντας το στατιστικό $F = \dfrac{SS(\hat{\psi}_i)}{MSW}$ το οποίο ακολουθεί την κατανομή $F_{1,\beta.\varepsilon.}$.

Μέγεθος Επίδρασης

Χρησιμοποιείται ο συντελεστής $\omega^2 = \frac{SS(\hat{\psi}_i) - MSW}{SST + MSW}$. Τιμές του ω^2 περίπου 0.01, 0.06 και 0.15 υποδηλώνουν αντίστοιχα μικρό, μεσαίο και μεγάλο μέγεθος επίδρασης της αντίθεσης. Ο συντελεστής ω^2 υποδεικνύει το ποσοστό της διασποράς της εξαρτημένης μεταβλητής που εξηγείται από την αντίθεση.

Η κατανόηση όλων των παραπάνω θα πραγματοποιηθεί χρησιμοποιώντας τα δεδομένα του παραδείγματος της παραγράφου 8.3 ορίζοντας εκ των προτέρων ποιες υποθέσεις θα ελεγχθούν.

Παράδειγμα

Στον παρακάτω πίνακα δίνεται ο μηνιαίος μισθός των υπαλλήλων σε σχέση με το επίπεδο εκπαίδευσής τους. Να ελεγχθούν οι υποθέσεις σε επίπεδο σημαντικότητας a = 0.05.

i) Ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Δημοτικό είναι ίσος με τον μέσο όρο των μισθών των τριών άλλων επιπέδων.

ii) Ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Λύκειο είναι ίσος με τον μέσο όρο των μισθών των υπαλλήλων με επίπεδο εκπαίδευσης Γυμνάσιο και ΑΕΙ/ΤΕΙ.

iii) Ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Γυμνάσιο είναι ίσος με τον μέσο μισθό των υπαλλήλων με επίπεδο εκπαίδευσης ΑΕΙ/ΤΕΙ.

Δημοτικό	Γυμνάσιο	Λύκειο	ΑΕΙ / ΤΕΙ
10	7	13	15
11	8	14	16
9	7	15	10
10	11	14	16
10	12	14	18

Η ανάλυση διασποράς ελέγχει την υπόθεση

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 και j.

Επιπλέον ζητείται ο έλεγχος των υποθέσεων:

H_0: $\mu_1 = \frac{\mu_2 + \mu_3 + \mu_4}{3}$ ή $3\mu_1 - 1\mu_2 - 1\mu_3 - 1\mu_4 = 0$, H_1: $3\mu_1 - 1\mu_2 - 1\mu_3 - 1\mu_4 \neq 0$,

H_0: $\mu_3 = \frac{\mu_2 + \mu_4}{2}$ ή $0\mu_1 - 1\mu_2 + 2\mu_3 - 1\mu_4 = 0$, H_1: $0\mu_1 - 1\mu_2 + 2\mu_3 - 1\mu_4 \neq 0$ και

H_0: $\mu_2 = \mu_4$ ή $0\mu_1 + 1\mu_2 + 0\mu_3 - 1\mu_4 = 0$, H_1: $0\mu_1 + 1\mu_2 + 0\mu_3 - 1\mu_4 \neq 0$.

Οι αντιθέσεις **ορίζονται** ως εξής:

$\psi_1 = 3\mu_1 - 1\mu_2 - 1\mu_3 - 1\mu_4$,

$\psi_2 = 0\mu_1 - 1\mu_2 + 2\mu_3 - 1\mu_4$ και

$\psi_3 = 0\mu_1 + 1\mu_2 + 0\mu_3 - 1\mu_4$

με συνέπεια τα **διανύσματα βαρών** να είναι $\boldsymbol{c}_1 = (3,-1,-1,-1)$,

$\boldsymbol{c}_2 = (0,-1,2,-1)$ και

$\boldsymbol{c}_3 = (0,1,0,-1)$.

Προφανώς ισχύει ότι $\sum_{h=1}^{4} c_{ih} = 0$ για $1 \le i \le 3$.

Παρατηρώντας ότι τα μεγέθη δειγμάτων των επιπέδων είναι ίσα αποδεικνύεται ότι οι αντιθέσεις είναι **ορθογώνιες** διότι:

$\sum_{h=1}^{4} c_{1h}c_{2h} = 3\cdot 0 + (-1)\cdot(-1) + (-1)\cdot 2 + (-1)\cdot(-1) = 0$,

$\sum_{h=1}^{4} c_{1h}c_{3h} = 3\cdot 0 + (-1)\cdot 1 + (-1)\cdot 0 + (-1)\cdot(-1) = 0$ και

$\sum_{h=1}^{4} c_{2h}c_{3h} = 0\cdot 0 + (-1)\cdot 1 + 2\cdot 0 + (-1)\cdot(-1) = 0$.

Στην παράγραφο 8.3 είχαν υπολογιστεί

$\bar{y}_1 = 10$, $\bar{y}_2 = 9$, $\bar{y}_3 = 14$, $\bar{y}_4 = 15$, $\bar{y} = 12$,

$s_1^2 = 0.5$, $s_2^2 = 5.5$, $s_3^2 = 0.5$, $s_4^2 = 9$ και $MSW = 3.875$.

Οι τιμές των **εκτιμητών των αντιθέσεων** είναι

$\hat{\psi}_1 = 3\bar{y}_1 - \bar{y}_2 - \bar{y}_3 - \bar{y}_4 = -8$,

$\hat{\psi}_2 = -\bar{y}_2 + 2\bar{y}_3 - \bar{y}_4 = 4$ και

$\hat{\psi}_3 = \bar{y}_2 - \bar{y}_4 = -6$.

Τα **τυπικά σφάλματα υποθέτοντας ότι οι διασπορές είναι ίσες** βρίσκονται από τη σχέση $SE(\hat{\psi}_i) = \sqrt{MSW \sum_{h=1}^{k} \frac{c_{ih}^2}{n_h}}$. Επομένως

$$SE(\hat{\psi}_1) = \sqrt{3.875\left(\frac{3^2}{5} + \frac{(-1)^2}{5} + \frac{(-1)^2}{5} + \frac{(-1)^2}{5}\right)} = 3.050,$$

$$SE(\hat{\psi}_2) = \sqrt{3.875\left(\frac{0^2}{5} + \frac{(-1)^2}{5} + \frac{2^2}{5} + \frac{(-1)^2}{5}\right)} = 2.156 \text{ και}$$

$$SE(\hat{\psi}_3) = \sqrt{3.875\left(\frac{0^2}{5} + \frac{1^2}{5} + \frac{0^2}{5} + \frac{(-1)^2}{5}\right)} = 1.245.$$

Ο έλεγχος μηδενικής υπόθεσης για την ισότητα της τιμής μιας αντίθεση με το μηδέν βασίζεται στον υπολογισμό των στατιστικών $t_i = \frac{\hat{\psi}_i}{SE(\hat{\psi}_i)}$.

Στην περίπτωση **ίσων διασπορών** ισχύει ότι

$$t_1 = \frac{\hat{\psi}_1}{SE(\hat{\psi}_1)} = \frac{-8}{3.050} = -2.623,$$

$$t_2 = \frac{4}{2.156} = 1.855 \text{ και}$$

$$t_3 = \frac{-6}{1.245} = -4.819.$$

Οι απόλυτες τιμές των παραπάνω στατιστικών συγκρίνονται με την τιμή $t_{n-k;\alpha/2} = t_{16;0.025} = 2.12$.

Διαπιστώνεται επομένως, σε επίπεδο σημαντικότητας 0.05, ότι η τιμή της 2ης αντίθεσης μπορεί να είναι ίση με το μηδέν με συνέπεια ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Λύκειο να είναι ίσος με τον μέσο όρο των μισθών των υπαλλήλων με επίπεδο εκπαίδευσης Γυμνάσιο και ΑΕΙ/ΤΕΙ.

Αντίθετα ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Δημοτικό δεν είναι ίσος με τον μέσο όρο των μισθών των άλλων τριών επιπέδων και ο μέσος μισθός των υπαλλήλων με επίπεδο εκπαίδευσης Γυμνάσιο δεν είναι ίσος με τον μέσο μισθό των υπαλλήλων με επίπεδο εκπαίδευσης ΑΕΙ/ΤΕΙ.

Τα αθροίσματα τετραγώνων των αποκλίσεων που εξηγούνται από τις αντιθέσεις υπολογίζονται από τη σχέση $SS(\hat{\psi}_i) = \dfrac{\hat{\psi}_i^2}{\sum_{h=1}^{k} \frac{c_{ij}^2}{n_h}}$. Προφανώς

$$SS(\hat{\psi}_1) = \frac{(-8)^2}{\left(\frac{3^2}{5} + \frac{(-1)^2}{5} + \frac{(-1)^2}{5} + \frac{(-1)^2}{5}\right)} = 26.667,$$

$$SS(\hat{\psi}_2) = \frac{4^2}{\left(\frac{0^2}{5} + \frac{(-1)^2}{5} + \frac{2^2}{5} + \frac{(-1)^2}{5}\right)} = 13.333 \text{ και}$$

$$SS(\hat{\psi}_3) = \frac{(-6)^2}{\left(\frac{0^2}{5} + \frac{1^2}{5} + \frac{5^2}{5} + \frac{(-1)^2}{5}\right)} = 90.$$

Εφόσον οι αντιθέσεις είναι ορθογώνιες ισχύει ότι:

$SS(\hat{\psi}_1) + SS(\hat{\psi}_2) + SS(\hat{\psi}_3) = SSB = 130$.

Κατά τη λύση του παραδείγματος στην παράγραφο 8.3 ο έλεγχος υπόθεσης της ανάλυσης διασποράς της ισότητας των μηνιαίων μισθών των υπαλλήλων είχε απορριφθεί υποθέτοντας είτε ότι οι διασπορές είναι ίσες ή όχι. Σύμφωνα με τη 2η παρατήρηση η απόρριψη της μηδενικής υπόθεσης συνεπάγεται την απόρριψη της μηδενικής υπόθεσης για μία τουλάχιστον αντίθεση, γεγονός που στο συγκεκριμένο παράδειγμα συμβαίνει για δύο αντιθέσεις, την 1η και την 3η.

Συνοψίζοντας και ανακεφαλαιώνοντας τους παραπάνω υπολογισμούς ο πίνακας της ανάλυσης διασποράς με αντιθέσεις έχει τη μορφή:

Πηγή Μεταβολής	**Αθρ. Τετρ.(SS)**	**β.ε.**	**MS**	**Στατιστικό F**
between groups	$SSB = 130$	3	$MSB = 43.333$	$F = \frac{MSB}{MSW} = 11.183$
$\hat{\psi}_1$	$SS(\hat{\psi}_1) = 26.667$	1	$MS(\hat{\psi}_1) = \frac{SS(\hat{\psi}_1)}{1} = 26.667$	$F_1 = \frac{MS(\hat{\psi}_1)}{MSW} = 6.882$
$\hat{\psi}_2$	$SS(\hat{\psi}_2) = 13.333$	1	$MS(\hat{\psi}_2) = \frac{SS(\hat{\psi}_2)}{1} = 13.333$	$F_2 = \frac{MS(\hat{\psi}_2)}{MSW} = 3.441$
$\hat{\psi}_3$	$SS(\hat{\psi}_3) = 90$	1	$MS(\hat{\psi}_3) = \frac{SS(\hat{\psi}_3)}{1} = 90$	$F_3 = \frac{MS(\hat{\psi}_3)}{MSW} = 23.226$
within groups	$SSW = 62$	16	$MSW = 3.875$	
Σύνολο	$SST = 192$	19		

Ο έλεγχος της μηδενικής υπόθεσης ισότητας τη τιμής της αντίθεσης με το μηδέν μπορεί να πραγματοποιηθεί, όπως ήδη αναφέρθηκε, χρησιμοποιώντας τα στατιστικά F_1, F_2 και F_3 και συγκρίνοντας την τιμή τους με την τιμή $F_{1,n-k;a}$ η οποία στη συγκεκριμένη περίπτωση είναι $F_{1,16;0.05} = 4.49$.

Προφανώς

$F_1 > F_{1,16;0.05}$,

$F_3 > F_{1,16;0.05}$, ενώ

$F_2 < F_{1,16;0.05}$

με συνέπεια να απορρίπτεται η μηδενική υπόθεση για τον 1ο και 3ο έλεγχο, ενώ γίνεται δεκτή για τον 2ο έλεγχο που πραγματοποιήθηκε.