ΚΕΦΑΛΑΙΟ 3

ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ ΒΑΣΙΣΜΕΝΕΣ ΣΤΙΣ ΤΑΞΕΙΣ ΜΕΓΕΘΟΥΣ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΕΝΟΣ Ή ΔΥΟ ΔΕΙΓΜΑΤΩΝ
(Methods Based on Ranks)

Στο κεφάλαιο αυτό, εξετάζονται ορισμένες τεχικές ανάλυσης δεδομένων, οι οποίες βασίζονται στην τάξη μεγέθους των παρατηρήσεων ενός δείγματος και όχι στις παρατηρήσεις αυτές καθεαυτές.

Οι *βαθμοί (τάξεις μεγέθους, ranks)* των δεδομένων προτιμώνται συχνά από τα δεδομένα αυτά καθεαυτά για διάφορους λόγους. Ο πρώτος λόγος μπορεί να είναι ότι οι τιμές των παρατηρήσεων δεν έχουν νόημα από μόνες τους παρά μόνο όταν θεωρείται η διάταξή τους σε σχέση με τις υπόλοιπες παρατηρήσεις, οπότε οι τιμές δεν εμπεριέχουν περισσότερες πληροφορίες από ό,τι οι τάξεις μεγέθους τους. Αυτή είναι η φύση των δεδομένων που βρίσκονται σε κλίμακα διάταξης (ordinal scale). Ο δεύτερος λόγος μπορεί να είναι ότι οι τιμές των παρατηρήσεων έχουν μεν νόημα από μόνες τους, αλλά η συνάρτηση κατανομής τους δεν είναι κανονική, οπότε η Θεωρία Πιθανοτήτων δεν προσφέρεται για τον προσδιορισμό της κατανομής της στατιστικής συνάρτησης που χρησιμοποιείται για τον έλεγχο υποθέσεων, όταν αυτή είναι συνάρτηση των παρατηρήσεων αυτών καθεαυτών. Αντίθετα, η Θεωρία Πιθανοτήτων, που απαιτείται για τον έλεγχο υποθέσεων με την χρήση στατιστικών συναρτήσεων που βασίζονται στις τάξεις μεγέθους των παρατηρήσεων, είναι σχετικά απλή και δεν εξαρτάται από την κατανομή από την οποία έχουν προέλθει οι παρατηρήσεις σε πολλές περιπτώσεις.

Τα δεδομένα, επομένως, μπορεί να είναι μη αριθμητικά *(καλό, καλύτερο, πολύ καλό)* ή αριθμητικά (9.64, 8.45, κ.λ.π.). Αν τα δεδομένα είναι μη αριθμητικά, αλλά είναι διατεταγμένα, δηλαδή βρίσκονται σε διατεταγμένη κλίμακα (ordinal data), οι μέθοδοι ανάλυσης που θα εξετασθούν είναι συχνά από τις πιο ισχυρές, μεθόδους που είναι διαθέσιμες. Εάν τα δεδομένα είναι αριθμητικά και, επιπλέον, αποτελούν παρατηρήσεις πάνω σε τυχαίες μεταβλητές που έχουν την κανονική κατανομή, έτσι ώστε να πληρούνται όλες οι προϋποθέσεις που απαιτούνται για τους παραμετρικούς ελέγχους, η απώλεια της αποτελεσματικότητας στην οποία οδηγεί η χρήση των μη παραμετρικών μεθόδων αυτού του κεφαλαίου είναι πολύ μικρή. Σε τέτοιες περιπτώσεις, η σχετική αποτελεσματικότητα των ελέγχων οι οποίοι χρησιμοποιούν μόνο την τάξη μεγέθους των παρατηρήσεων είναι συχνά περίπου 0.95, ανάλογα με την περίπτωση.

Οι μέθοδοι αυτού του κεφαλαίου ισχύουν για όλες τις μορφές των πληθυσμών, ανεξάρτητα, δηλαδή, από το εάν ο πληθυσμός είναι συνεχής, διακριτός ή μίγμα των δύο αυτών τύπων. Βέβαια, οι πρώτες μη παραμετρικές μέθοδοι απαιτούσαν την υπόθεση της συνέχειας των τυχαίων μεταβλητών για την ισχύ των ελέγχων που βασίζονται στην τάξη μεγέθους των παρατηρήσεων. Όμως, μεταγενέστερα αποτελέσματα (Conover, 1973) απέδειξαν ότι η υπόθεση της συνέχειας δεν είναι απαραίτητη. Αντίθετα, μπορεί να αντικατασταθεί από την τετριμμένη υπόθεση ότι $P(X = x) < 1$ για κάθε x. Επειδή είναι μάλλον σπάνιο να παίρνει ένας ερευνητής κάποιο δείγμα από έναν πληθυσμό που αποτελείται εξ ολοκλήρου από μία μοναδική τιμή, έστω x_0, η υπόθεση ότι ο πληθυσμός αποτελείται από μία τουλάχιστον τιμή παραλείπεται στα επόμενα.

Μία χρήσιμη μέθοδος για την διάταξη παρατηρήσεων κατά αύξουσα σειρά μεγέθους στηρίζεται, όπως είναι γνωστό, στην κατασκευή του διαγράμματος μίσχου-φύλλου (stem-and-leaf diagram) (Βλέπε, για παράδειγμα, Ι. Πανάρετου και Ε. Ξεκαλάκη "*Εισαγωγή στη Στατιστική σκέψη*", Αθήνα 1993). Ας υποθέσουμε, για παράδειγμα, ότι θέλουμε να διατάξουμε κατά αύξουσα τάξη μεγέθους τις εξής παρατηρήσεις:

74	63	88	69	81	91	75
82	91	87	77	86	86	87
96	84	93	73	74	93	78
70	84	90	97	79	89	93

Χρησιμοποιώντας τα ψηφία των δεκάδων ως *μίσχους* και τα ψηφία των μονάδων ως *φύλλα*, το διάγραμμα μίσχου-φύλλου των παραπάνω δεδομένων έχει την μορφή:

6	3 9
7	4 0 7 3 4 9 5 8
8	2 4 4 8 7 1 6 6 9 7
9	6 1 3 0 7 1 3 3

Από την κατανομή των δεδομένων, όπως αυτή προκύπτει από το διάγραμμα μίσχου-φύλλου, εύκολα μπορούμε να κατατάξουμε τα δεδομένα κατά αύξουσα σειρά μεγέθους των παρατηρήσεων, αν διατάξουμε κατά αύξουσα σειρά μεγέθους τις τιμές που απαρτίζουν τα φύλλα του διαγράμματος μίσχου-φύλλου. Το αποτέλεσμα για τα δεδομένα μας φαίνεται από το διάγραμμα μίσχου-φύλλου που ακολουθεί.

6	3 9
7	0 3 4 4 5 7 8 9
8	1 2 4 4 6 6 7 7 8 9
9	0 1 1 3 3 3 6 7

Από το διάγραμμα αυτό προκύπτει ότι το διατεταγμένο σύνολο δεδομένων έχει την μορφή

63 69 70 73 74 74 75 77 78 79

81 82 84 84 86 86 87 87 88 89

90 91 91 93 93 93 96 97

Εύκολα τώρα μπορούμε να αντιστοιχίσουμε σε κάθε παρατήρηση του διατεταγμένου συνόλου τον *βαθμό* της, δηλαδή τον αριθμό που εκπροσωπεί την θέση που η παρατήρηση κατέχει στο διατεταγμένο δείγμα. Αν δύο ή περισσότερες τιμές ταυτίζονται, τότε ως βαθμός κάθε μιας από αυτές θεωρείται ο μέσος των βαθμών που αυτές θα είχαν αν ήταν διαφορετικές. Η ακολουθία των βαθμών των παραπάνω δεδομένων είναι η εξής:

1 2 3 4 5.5 5.5 7 8 9 10

11 12 13.5 13.5 15.5 15.5 17.5 17.5 19 20

21 22.5 22.5 25 25 25 27 28

Στα επόμενα, εξετάζονται μη παραμετρικές τεχνικές οι οποίες εφαρμόζονται για την εξαγωγή συμπερασμάτων σχετικών με την θέση ενός πληθυσμού ή με την σύγκριση των θέσεων δύο πληθυσμών, όταν ως μέτρο θέσης θεωρείται κυρίως η διάμεσος του πληθυσμού. Η συμπερασματολογία η σχετική με την διάμεσο ενός πληθυσμού είναι σημαντική σε πολλές πρακτικές εφαρμογές για δύο κυρίως λόγους. Ο πρώτος λόγος αναφέρεται στην περίπτωση που ο πληθυσμός παρουσιάζει ασυμμετρία, όπως, για παράδειγμα, στην περίπτωση της κατανομής οικογενειακού εισοδήματος, των πωλήσεων ενός καταστήματος ή του αποθέματος ενός εργοστασίου. Η διάμεσος του πληθυσμού στις περιπτώσεις αυτές βρίσκεται πλησιέστερα στο κέντρο της κατανομής από ό,τι η μέση τιμή του πληθυσμού και, επομένως, ενδέχεται να έχει περισσότερη έννοια ως μέτρο θέσης. Ο δεύτερος λόγος αναφέρεται στην περίπτωση που ο πληθυσμός είναι συμμετρικός. Στην περίπτωση αυτή, η διάμεσος και η μέση τιμή του πληθυσμού ταυτίζονται και επομένως και οι δύο αυτές παράμετροι είναι χρήσιμες ως μέτρα θέσης.

3.1 Ο ΕΛΕΓΧΟΣ WILCOXON ΓΙΑ ΕΝΑ ΔΕΙΓΜΑ ΠΑΡΑΤΗΡΗΣΕΩΝ Ή ΖΕΥΓΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

Ο έλεγχος Wilcoxon χρησιμοποιείται για τον έλεγχο υποθέσεων που αναφέρονται σε παραμέτρους κεντρικής τάσης. Εφαρμόζεται στις περιπτώσεις όπου έχουμε ένα μοναδικό δείγμα παρατηρήσεων, όπως επίσης και στις περιπτώσεις όπου το δείγμα μας αποτελείται από ζεύγη παρατηρήσεων, κατάσταση που οδηγεί σε ένα μοναδικό δείγμα, το δείγμα των διαφορών των μελών των αρχικών ζευγών παρατηρήσεων. (Το ζεύγος παρατηρήσεων (X_i, Y_i) αποτελεί στην πραγματικότητα μία μοναδική παρατήρηση πάνω σε μία διδιάστατη τυχαία μεταβλητή). Όπως και ο προσημικός έλεγχος που εξετάσαμε στο κεφάλαιο 2, έτσι και ο έλεγχος του Wilcoxon έχει ως παραμετρικό ανάλογο τον έλεγχο t (για ένα δείγμα παρατηρήσεων ή για δείγμα ζευγών παρατηρήσεων). Συγκριτικές μελέτες έδειξαν ότι η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου των προσημασμένων τάξεων Wilcoxon σε σχέση με τον αντίστοιχο παραμετρικό έλεγχο T δεν είναι ποτέ χαμηλότερη από την τιμή 0.864. (Στην περίπτωση ζευγών παρατηρήσεων, η μελέτη αυτή έγινε με την προϋπόθεση ότι οι πληθυσμοί από τους οποίους οι παρατηρήσεις ελήφθησαν διαφέρουν μόνο κατά την μέση τιμή τους). Όπως είδαμε στην σχετική ενότητα, ο προσημικός έλεγχος εφαρμόζεται σε δείγματα ζευγών παρατηρήσεων *χαρακτηρίζοντας* κάθε ζεύγος ως "+" ζεύγος, "–" ζεύγος, ή "0" ζεύγος (δηλαδή ζεύγος παρατηρήσεων που οι τιμές τους ταυτίζονται) και εφαρμόζοντας τον διωνυμικό έλεγχο στον προκύπτον απλό δείγμα. Ο έλεγχος της ενότητας αυτής επίσης *μετασχηματίζει* τα ζεύγη παρατηρήσεων (X_i, Y_i), $i = 1, 2, ..., n$ σε απλές παρατηρήσεις θεωρώντας τις διαφορές

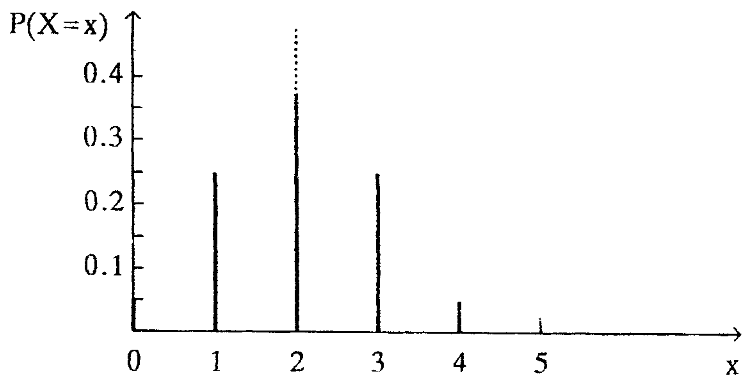
$$D_i = Y_i - X_i\ , i = 1, 2, ..., n.$$

Η ανάλυση τότε εφαρμόζεται πάνω στις μεταβλητές D_1, D_2, ..., D_n, οι οποίες θεωρούνται ως ένα δείγμα απλών παρατηρήσεων. Σε αντίθεση με τον προσημικό έλεγχο, ο οποίος απλώς σημειώνει αν η μεταβλητή D_i $(i = 1, 2, ..., n)$ είναι θετική, αρνητική ή μηδέν, ο έλεγχος της παρούσας ενότητας παίρνει υπόψη του τα μεγέθη των θετικών D_i σε σχέση με τα μεγέθη των αρνητικών D_i. Υπάρχει, δηλαδή, μία διαφορά μεταξύ των δύο αυτών ελέγχων, παρά το γεγονός ότι οι υποθέσεις που ενδιαφερόμαστε να ελέγξουμε στις δύο αυτές περιπτώσεις έχουν πολλές ομοιότητες.

Η σημαντική διαφορά που υπάρχει μεταξύ του προσημικού ελέγχου και του ελέγχου του Wilcoxon είναι ότι στην περίπτωση του τελευταίου ελέγχου απαιτείται μία πρόσθετη υπόθεση για την κατανομή των διαφορών $D_i = Y_i - X_i$, $i = 1, 2, ..., n$. Συγκεκριμένα, απαιτείται η υπόθεση ότι η κατανομή των μεταβλητών $D_1, D_2, ..., D_n$ είναι *συμμετρική*.

Στο σημείο αυτό, θα ήταν χρήσιμο να ορισθεί η έννοια της *συμμετρίας* μιας κατανομής και να εξετασθεί η επίδρασή της πάνω στην κλίμακα μέτρησης των δεδομένων.

Ο ορισμός της συμμετρίας είναι απλός στην περίπτωση διακριτών κατανομών. Μία διακριτή κατανομή είναι *συμμετρική* αν το αριστερό μισό του γραφήματος της συνάρτησης πιθανότητάς της είναι συμμετρικό (*ακριβές αντίγραφο*) του δεξιού μισού του γραφήματος. Για παράδειγμα, η διωνυμική κατανομή είναι συμμετρική αν η συνάρτηση πιθανότητάς της έχει την μορφή του σχήματος 3.1.1:

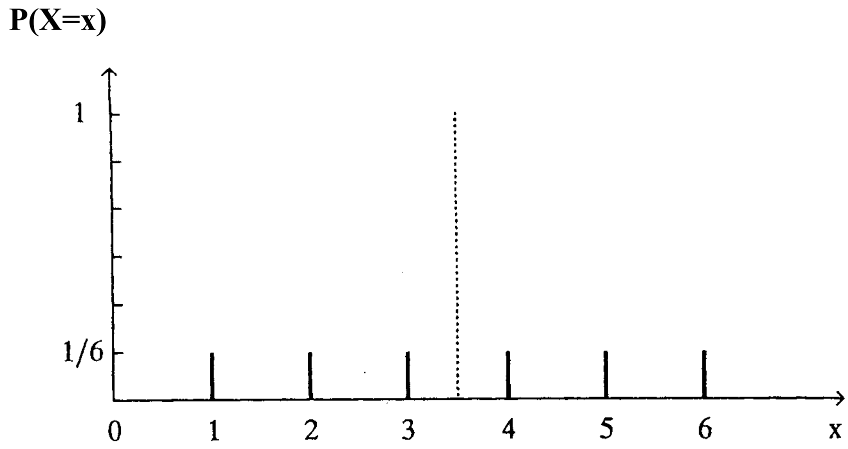


Σχήμα 3.1.1

Συμμετρία στην περίπτωση της διωνυμικής κατανομής

(n=4, p=0.5)

Επίσης, η διακριτή ομοιόμορφη κατανομή είναι πάντοτε συμμετρική (σχήμα 3.1.2). Οι διακεκομμένες γραμμές στα σχήματα αυτά συμβολίζουν τις ευθείες γύρω από τις οποίες οι κατανομές είναι συμμετρικές.



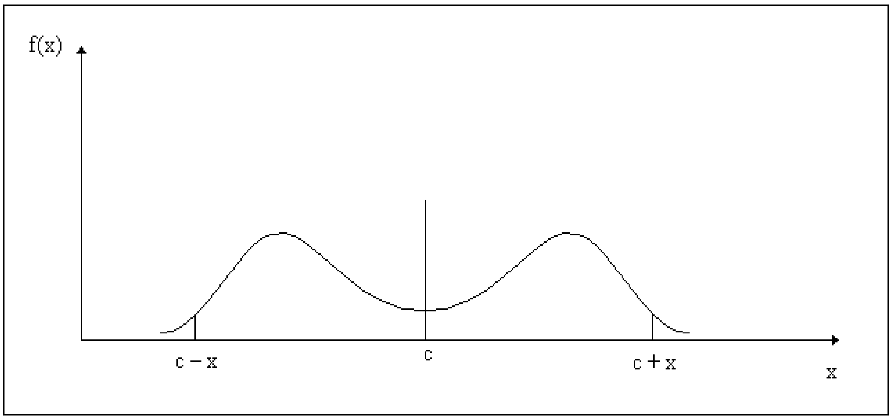
Σχήμα 3.1.2

Συμμετρία στην περίπτωση της διακριτής ομοιόμορφης κατανομής (P(X=x)=1/6, x=1,2,…,6)

Στην περίπτωση των μη διακριτών κατανομών συνηθίζεται περισσότερο να χρησιμοποιείται ο εξής ορισμός της συμμετρίας:

Η κατανομή μιας συνεχούς τυχαίας μεταβλητής X λέγεται *συμμετρική γύρω από την ευθεία* x=c, για κάποια σταθερά c, αν $P(X \le c - x) = P(X \ge c + x)$ για κάθε δυνατή τιμή x.

Η γεωμετρική ερμηνεία της παραπάνω ισότητας είναι ότι οι γραμμοσκιασμένες περιοχές του σχήματος 3.1.3, που ακολουθεί, έχουν ίσα εμβαδά για κάθε x, οπότε το γράφημα της συνάρτησης πυκνότητας πιθανότητας αριστερά της σταθεράς c είναι συμμετρικό του γραφήματος δεξιά της σταθεράς c.



Σχήμα 3.1.3

Συνεχής κατανομή συμμετρική γύρω από την σταθερά c

Είναι προφανές ότι, στο σχήμα 3.1.1, c=2 και ο παραπάνω ορισμός μπορεί εύκολα να επαληθευθεί για κάθε πραγματική τιμή x. Το ίδιο ισχύει και για την κατανομή του σχήματος 3.1.2, όπου c=3.5.

Θα πρέπει να τονισθεί στο σημείο αυτό ότι, παρά το γεγονός ότι συχνά ενδέχεται να μην γνωρίζουμε την πραγματική κατανομή μιας τυχαίας μεταβλητής, μπορούμε να κάνουμε την υπόθεση ότι η κατανομή αυτή είναι συμμετρική. Η υπόθεση αυτή, είναι προφανές ότι δεν είναι το ίδιο ισχυρή όπως η υπόθεση της κανονικότητας, αφού, όπως είναι γνωστό, η κανονική κατανομή είναι συμμετρική, ενώ μία συμμετρική κατανομή δεν είναι κατ' ανάγκη κανονική.

Οπως είναι γνωστό, εάν μία κατανομή είναι συμμετρική, η μέση τιμή της (εάν υπάρχει) ταυτίζεται με την διάμεσο. Μία συνέπεια, επομένως, της προσθήκης της υπόθεσης της συμμετρίας είναι ότι οποιαδήποτε συμπερασματολογία σχετική με την διάμεσο θα ισχύει και για την μέση τιμή της κατανομής.

Μία δεύτερη συνέπεια της προσθήκης της υπόθεσης της συμμετρίας είναι ότι η κλίμακα μέτρησης των δεδομένων απαιτείται να αλλάξει από κλίμακα διάταξης σε κλίμακα διαστήματος. Στην διατεταγμένη κλίμακα, η διάκριση μεταξύ δύο τιμών μιας τυχαίας μεταβλητής γίνεται μόνο στην βάση του ποια είναι μεγαλύτερη και ποια μικρότερη από τις δύο. Δεν είναι απαραίτητο να γνωρίζουμε ποια από τις δύο τιμές είναι πλησιέστερα προς την διάμεσο. Αν η υπόθεση της συμμετρίας έχει έννοια (είναι εύλογη), η απόσταση των παρατηρήσεων από την διάμεσο έχει έννοια και, επομένως, και η απόσταση μεταξύ δύο παρατηρήσεων έχει έννοια. Κατά συνέπεια, η κλίμακα μέτρησης των δεδομένων δεν είναι απλώς διατεταγμένη, αλλά είναι κλίμακα διαστήματος.

Ο έλεγχος, που προτάθηκε από τον Wilcoxon (1945), αναπτύχθηκε για τον έλεγχο της υπόθεσης ότι ένα συγκεκριμένο δείγμα προέρχεται από κάποιον πληθυσμό με μία συγκεκριμένη διάμεσο. Όπως επίσης αναφέρθηκε νωρίτερα, ο έλεγχος αυτός μπορεί να χρησιμοποιηθεί και σε περιπτώσεις όπου το δείγμα αποτελείται από ζεύγη παρατηρήσεων, όπως, για παράδειγμα, στην περίπτωση παρατηρήσεων "*πριν*" (X) και "*μετά*" (Y) που έχουν γίνει πάνω σε κάθε ένα από n άτομα με σκοπό να ελεγχθεί αν η δεύτερη τυχαία μεταβλητή στο ζεύγος (X, Y) έχει την ίδια διάμεσο όπως και η πρώτη.

3.1.1 Ο έλεγχος των Προσημασμένων Τάξεων Μεγέθους του Wilcoxon για την διάμεσο ενός πληθυσμού

(The Wilcoxon Signed Rank Test for a Median)

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα παρατηρήσεων μεγέθους n΄, έστω $X_1, X_2, ..., X_{n'}$, από έναν πληθυσμό ο οποίος περιγράφεται από την τυχαία μεταβλητή X της οποίας η κατανομή είναι συμμετρική. Έστω επίσης ότι $x_{0.5}$ συμβολίζει την διάμεσο της κατανομής της τυχαίας μεταβλητής X (του πληθυσμού). Αν m είναι μια γνωστή σταθερά, οι υποθέσεις που ενδέχεται να ενδιαφερόμαστε να ελέγξουμε σε σχέση με τον δοθέντα πληθυσμό έχουν τη μορφή:

A. (Μονόπλευρος έλεγχος):

H_0: $x_{0.5} \geq m$

H_1: $x_{0.5} < m$

B. (Μονόπλευρος έλεγχος):

H_0: $x_{0.5} \leq m$

H_1: $x_{0.5} > m$

Γ. (Αμφίπλευρος έλεγχος):

H_0: $x_{0.5} = m$

H_1: $x_{0.5} \neq m$.

Παρατήρηση: Είναι προφανές, ότι αν η μέση τιμή του πληθυσμού E(X) υπάρχει, τότε η παράμετρος $x_{0.5}$ μπορεί να αντικατασταθεί από την παράμετρο E(X) στις υποθέσεις των περιπτώσεων A, B και Γ, οποτεδήποτε η μορφή του προβλήματος απαιτεί ελέγχους σχετικά με τη μέση τιμή του πληθυσμού.

Οπως είναι γνωστό, από τον ορισμό της διαμέσου ενός πληθυσμού, ισχύει ότι

$$P(X<x_{0.5}) \leq 0.5 \text{ και } P(X>x_{0.5}) \leq 0.5.$$

Επομένως, κάθε X_i (i = 1, 2, ..., n΄) έχει πιθανότητα το πολύ ίση με 0.5 να υπερβαίνει την διάμεσο $x_{0.5}$ και πιθανότητα το πολύ ίση με 0.5 να είναι μικρότερη από αυτήν. (Στην περίπτωση που ο πληθυσμός είναι συνεχής, οι πιθανότητες αυτές είναι ίσες με 0.5, δηλαδή $P(X<x_{0.5}) = P(X>x_{0.5}) = 0.5$, $i = 1, 2, ..., n'$). Αν, επιπλέον, η διάμεσος του πληθυσμού είναι ίση με το μηδέν ($x_{0.5} = 0$), κάθε X_i ($i = 1, 2, ..., n'$) θα έχει πιθανότητα το πολύ ίση με 0.5 να έχει θετική τιμή (η παρατήρηση να είναι πάνω από την διάμεσο) και πιθανότητα το πολύ ίση με 0.5 να έχει αρνητική τιμή (η παρατήρηση να είναι κάτω από την διάμεσο).

Αν, επομένως, από το αρχικό δείγμα των παρατηρήσεων X_i, $i = 1, 2, ..., n'$ θεωρήσουμε το δείγμα των διαφορών $D_i = m - X_i$, $i = 1, 2, ..., n'$, κάθε παρατήρηση D_i ($i = 1, 2, ..., n'$) του προκύπτοντος δείγματος θα είναι θετική (ή αρνητική) με πιθανότητα το πολύ ίση με 0.5. Επιπλέον, λόγω συμμετρίας της κατανομής του πληθυσμού, τόσο η τιμή της διαφοράς D_i ($i = 1, 2, ..., n'$) όσο και η τάξη μεγέθους της είναι ανεξάρτητες από το αν η διαφορά βρίσκεται πάνω ή κάτω από την διάμεσο $d_{0.5}$ του δείγματος των D_i, $i = 1, 2, ..., n'$, που στην προκειμένη περίπτωση είναι ίση με το μηδέν. (Αν η κατανομή του πληθυσμού δεν ήταν συμμετρική, οι θετικές διαφορές, ενδεχομένως, θα έτειναν να είναι πολύ μεγαλύτερες από τις αρνητικές διαφορές ή αντίστροφα).

Από τα παραπάνω, συνάγεται ότι για τον έλεγχο των υποθέσεων των περιπτώσεων Α, Β και Γ απαιτείται μια στατιστική συνάρτηση, της οποίας η τιμή να είναι ενδεικτική του κατά πόσον οι θετικές διαφορές δεν υπερβαίνουν τις αρνητικές διαφορές τόσο σε μέγεθος όσο και σε συχνότητα εμφάνισης.

Θεωρούμε το δείγμα των διαφορών $D_i = m - X_i$, $i = 1, 2, ..., n'$, από το οποίο εξαιρούμε τις διαφορές που είναι ίσες με το μηδέν. Έστω ότι n, $n \le n'$ συμβολίζει τον αριθμό των διαφορών που απομένουν.

Θεωρούμε στην συνέχεια τις απόλυτες τιμές αυτών των διαφορών, δηλαδή την ακολουθία των τυχαίων μεταβλητών

$$|D_i| = |m - X_i|, \; i = 1, 2, ..., n.$$

Διατάσσουμε κατά αύξουσα τάξη μεγέθους τις τιμές $|D_i|$, $i = 1, 2, ..., n$ και αντιστοιχίζουμε βαθμούς από το 1 μεχρι το n στις παρατηρήσεις του προκύπτοντος διατεταγμένου δείγματος (με το βαθμό 1 να αντιστοιχεί στην μικρότερη απόλυτη διαφορά και το βαθμό n να αντιστοιχεί στην μεγαλύτερη απόλυτη διαφορά). Αν περισσότερες από μία απόλυτες διαφορές είναι ίσες, τότε σε κάθε μία από αυτές αντιστοιχίζουμε τον μέσο των βαθμών που αυτές θα είχαν αν οι τιμές τους διέφεραν. Θεωρούμε την ακολουθία των βαθμών (τάξεων μεγέθους) $R(|D_i|)$, $i = 1, 2, ..., n$ των απολύτων διαφορών $|D_i|$, $i = 1, 2, ..., n$ στο προκύπτον διατεταγμένο δείγμα και ορίζουμε τις μεταβλητές

$$R_i = \begin{cases} +R(|D_i|), & \text{αν } D_i \equiv m - X_i > 0 \\ -R(|D_i|), & \text{αν } D_i \equiv m - X_i < 0 \end{cases} \quad , i = 1, 2, ..., n .$$

Η μεταβλητή R_i ($i = 1, 2, ..., n$) ονομάζεται *προσημασμένος βαθμός ή προσημασμένη τάξη μεγέθους (signed rank).*

Τότε, ως στατιστική συνάρτηση για τον έλεγχο των υποθέσεων των περιπτώσεων Α, Β και Γ ορίζεται η

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}} .$$

Ο αριθμητής $S = \sum_{i=1}^{n} R_i$ της στατιστικής αυτής συνάρτησης είναι το άθροισμα των προσημασμένων τάξεων μεγέθους (signed ranks) των διαφορών $|D_i| = |m - X_i|$, $i = 1, 2, ..., n$. Επειδή οι όροι $R_1, R_2, ..., R_n$ του αθροίσματος αυτού αποτελούν ένα τυχαίο δείγμα από ένα συμμετρικό πληθυσμό με διάμεσο ίση με το μηδέν, η κατανομή της τυχαίας μεταβλητής $S = \sum_{i=1}^{n} R_i$ έχει μέση τιμή $E(S) = 0$ και διασπορά $V(S) = \sum_{i=1}^{n} R_i^2$. Επομένως, η επιλεγείσα στατιστική συνάρτηση T είναι η τυποποιημένη μορφή του αθροίσματος των προσημασμένων τάξεων μεγέθους των διαφορών $|D_i| = |m - X_i|$, $i = 1, 2, ..., n$.

Από τον ορισμό της στατιστικής συνάρτησης T, είναι προφανές ότι αυτή λαβαίνει υπόψη της τα εξής δύο στοιχεία:

(1) Το μέγεθος των διαφορών $D_i, i = 1, 2, ..., n$ (είναι συνάρτηση των τάξεων μεγέθους των διαφορών αυτών) το οποίο είναι ενδεικτικό της απόστασης των $X_i, i = 1, 2, ..., n$ από την τιμή m της διαμέσου τους.

(2) Το πρόσημο των διαφορών $D_i, i = 1,2,...,n$ το οποίο είναι ενδεικτικό του κατά πόσον μία τιμή X_i είναι κάτω από την διάμεσο (+) ή πάνω από την διάμεσο (–).

Σημείωση: Στην περίπτωση που δεν υπάρχουν περιπτώσεις ταύτισης στις τιμές των $R_1, R_2, ..., R_n$, ο παρανομαστής της στατιστικής συνάρτησης T γράφεται με την μορφή

$$V(S) = \sum_{i=1}^{n} R_i^2$$

= άθροισμα των τετραγώνων των n πρώτων φυσικών αριθμών

$= n\,(n+1)(2n+1)/6.$

Επομένως, η στατιστική συνάρτηση T παίρνει στην περίπτωση αυτή την μορφή

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{n(n+1)(2n+1)/6}}.$$

Η κατανομή της στατιστικής συνάρτησης Τ, είτε αυτή έχει την γενική μορφή η οποία επιτρέπει περιπτώσεις ταύτισης τιμών στο δείγμα των $R_1, R_2, ..., R_n$, είτε έχει την ειδική μορφή που αναφέρθηκε παραπάνω, προσεγγίζεται ικανοποιητικά από την τυποποιημένη κανονική κατανομή.

Θα πρέπει να σημειωθεί εδώ ότι στην περίπτωση που οι τιμές των παρατηρήσεων του δείγματος $R_1, R_2, ..., R_n$ είναι διακεκριμμένες (δεν υπάρχουν περιπτώσεις ταύτισης δύο ή περισσοτέρων τιμών) συχνά χρησιμοποιείται η στατιστική συνάρτηση

$$T^+ = \text{άθροισμα των θετικών } R_i$$

για τον έλεγχο των υποθέσεων των περιπτώσεων Α, Β και Γ. Τα ποσοστιαία σημεία w_p της κατανομής της στατιστικής αυτής συνάρτησης περιέχονται στον πίνακα 8 του παραρτήματος.

Ανεξάρτητα από το εάν χρησιμοποιείται η Τ ή η T^+ ως ελεγχοσυνάρτηση, η κρίσιμη περιοχή για τον έλεγχο των υποθέσεων της περίπτωσης Α είναι προφανές ότι αντιστοιχεί στην δεξιά ουρά της κατανομής της ελεγχοσυνάρτησης. Αυτό είναι άμεση συνέπεια του γεγονότος ότι μια μεγάλη τιμή της χρησιμοποιούμενης ελεγχοσυνάρτησης συνεπάγεται ότι οι θετικές διαφορές υπερτερούν των αρνητικών διαφορών σε μέγεθος ή/και αριθμό και, επομένως, το ποσοστό των τιμών του δείγματος οι οποίες είναι κάτω από την διάμεσο είναι μεγαλύτερο από 1/2 και όχι μικρότερο ή ίσο από 1/2, όπως θα συνέβαινε αν η μηδενική υπόθεση ήταν αληθής. Επομένως, ο κανόνας απόφασης για την περίπτωση των υποθέσεων της μορφής Α διατυπώνεται ως εξής:

Η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης υπερβαίνει το (1–α)-ποσοστιαίο σημείο της κατανομής της. (Δηλαδή, στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση Τ, η κρίσιμη περιοχή ορίζεται από την ανισότητα $T>z_{1-\alpha}$. Αντίστοιχα, στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T^{+}, η κρίσιμη περιοχή ορίζεται από την ανισότητα $T^{+}> w_{1-\alpha}$, όπου $w_{1-\alpha}$ προσδιορίζεται από τον πίνακα 8 του παραρτήματος).

Για τον έλεγχο των υποθέσεων της περίπτωσης Β, είναι προφανές ότι οι μικρές τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης θα αποτελούν ένδειξη υπέρ της εναλλακτικής υπόθεσης, αφού θα υπαινίσσονται ότι ο αριθμός των παρατηρήσεων του αρχικού δείγματος που βρίσκονται πάνω από την διάμεσο m υπερτερεί του αριθμού των παρατηρήσεων που βρίσκονται κάτω από την διάμεσο, έτσι ώστε το ποσοστό των τιμών που βρίσκονται πάνω από την διάμεσο να υπερβαίνει το 1/2. Επομένως, ο κανόνας απόφασης για την περίπτωση των υποθέσεων της μορφής Β διατυπώνεται ως εξής:

Η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής της. (Δηλαδή, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α αν $T<z_{\alpha} = -z_{1-\alpha}$ ή, αντίστοιχα $T^{+} <w_{\alpha}$).

Παρόμοια επιχειρηματολογία οδηγεί στο συμπέρασμα ότι, για τον έλεγχο των υποθέσεων της περίπτωσης Γ, η κρίσιμη περιοχή θα αποτελείται από τις τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης, οι οποίες είναι πολύ μεγάλες ή πολύ μικρές. Ο κανόνας απόφασης στην περίπτωση αυτή διατυπώνεται ως εξής:

Η μηδενική υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης υπερβαίνει το (1–α/2)-ποσοστιαίο σημείο, ή είναι μικρότερη από το α/2-ποσοστιαίο σημείο της κατανομής της. Επομένως, στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T, η κρίσιμη περιοχή αποτελείται από την ένωση των διαστημάτων που ορίζονται από τις ανισότητες $T>z_{1-\alpha/2}$ και $T<z_{\alpha/2} = -z_{1-\alpha/2}$. Στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T^+, η κρίσιμη περιοχή ορίζεται από τις ανισότητες $T^+ > w_{1-\alpha/2}$ και $T^+ < w_{\alpha/2}$.

Παράδειγμα 3.1.1: Οι αποπληθωρισμένες ετήσιες αποδόσεις των μετοχικών κεφαλαίων σε ένα χρηματιστήριο κατανέμονται συμμετρικά. Ένα τυχαίο δείγμα τεσσάρων αποπληθωρισμένων ετήσιων αποδόσεων αποτελείται από τις τιμές

+8.4 –4.3 –0.8 +12.5.

Νομίζετε ότι τα δεδομένα παρέχουν ενδείξεις ότι η διάμεσος απόδοση είναι μεγαλύτερη από 3; (α=0.05)

Λύση: Είναι προφανές ότι οι υποθέσεις, των οποίων ο έλεγχος θα μας οδηγήσει σε κάποιο συμπέρασμα, είναι οι εξής:

$$H_0: x_{0.5} \leq 3$$

$$H_1: x_{0.5} > 3.$$

Από τα δεδομένα του προβλήματος, υπολογίζουμε την τιμή της στατιστικής συνάρτησης T που χρησιμοποιείται ως ελεγχοσυνάρτηση για τον έλεγχο των προσημασμένων τάξεων μεγέθους του Wilcoxon. Οι υπολογισμοί συνοψίζονται στον πίνακα που ακολουθεί.

X_i	$D_i = 3 - X_i$	$\lvert D_i \rvert$	$R(\lvert D_i \rvert)$	R_i

+8.4	–5.4	5.4	2	–2
–4.3	+7.3	7.3	3	+3
–0.8	+3.8	3.8	1	+1
+12.5	–9.5	9.5	4	–4

Οπως φαίνεται από τα στοιχεία του πίνακα αυτού, οι τιμές των διαφορών $|D_i|$, i = 1, 2, 3.4 είναι διακεκριμένες (δεν υπάρχουν περιπτώσεις ταύτισης τιμών). Επομένως, ο έλεγχος των παραπάνω υποθέσεων μπορεί να γίνει είτε με την χρήση της στατιστικής συνάρτησης T ως ελεγχοσυνάρτησης είτε με την χρήση της στατιστικής συνάρτησης T^+ .

Χρησιμοποιώντας την στατιστική συνάρτηση T ως ελεγχοσυνάρτηση, έχουμε από τα στοιχεία του παραπάνω πίνακα ότι

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}}$$

$$= \frac{-2+3+1-4}{\sqrt{30}}$$

$$= -0.365.$$

Η τιμή αυτή βρίσκεται εκτός της κρίσιμης περιοχής μεγέθους 0.05, αφού $T = -0.365 > z_{0.05} = -z_{0.95} = -1.645$. Επομένως, η μηδενική υπόθεση δεν απορρίπτεται σε επίπεδο σημαντικότητας 0.05. Δηλαδή, σε επίπεδο σημαντικότητας 5% τα δεδομένα του προβλήματος δεν παρέχουν ενδείξεις υπέρ της υπόθεσης ότι η διάμεσος αποπληθωρισμένη ετήσια απόδοση των μετοχικών κεφαλαίων στο συγκεκριμένο χρηματιστήριο υπερβαίνει τις τρεις μονάδες.

Το κρίσιμο επίπεδο $\hat{\alpha}$ του ελέγχου αυτού μόλις υπερβαίνει το 35.7%. Πράγματι,

$$\begin{aligned}\hat{\alpha} &= \max P(T \le -0.365 \mid H_0) \\ &= \max P(T \le -0.365 \mid x_{0.5} \le 3) \\ &= P(T \le -0.365 \mid x_{0.5} = 3) \\ &= P(T \ge 0.365 \mid x_{0.5} = 3) \\ &= 1 - P(T < 0.365 \mid x_{0.5} = 3) \\ &> 1 - 0.643 \\ &= 0.357.\end{aligned}$$

Επομένως, θα χρειασθεί επίπεδο σημαντικότητας μεγαλύτερο από 35.7% για να απορριφθεί η μηδενική υπόθεση υπέρ της εναλλακτικής με βάση τα παρόντα δεδομένα.

Παρόμοια είναι και τα αποτελέσματα του ελέγχου ο οποίος βασίζεται στην ελεγχοσυνάρτηση T^+. Πράγματι, υπολογίζοντας την τιμή της ελεγχοσυνάρτησης T^+, έχουμε

$$\begin{aligned}T^+ &= \text{άθροισμα θετικών } R_i \\ &= 3+1 \\ &= 4.\end{aligned}$$

Η τιμή αυτή βρίσκεται εκτός της κρίσιμης περιοχής μεγέθους 0.05. Πράγματι, από τον πίνακα 8 του παραρτήματος, το 0.05-ποσοστιαίο σημείο της κατανομής της στατιστικής συνάρτησης T^+ είναι $w_{0.05}=0$.

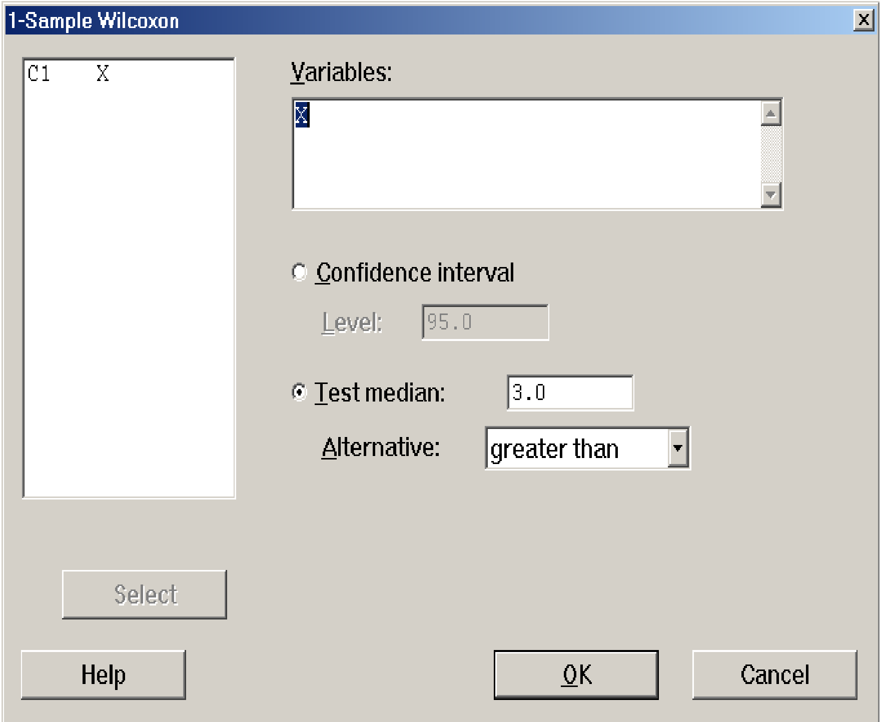
Επομένως, και με αυτήν την ελεγχοσυνάρτηση δεν οδηγούμεθα σε απόρριψη της μηδενικής υπόθεσης με βάση τα δεδομένα μας σε επίπεδο σημαντικότητας 0.05.

Από τον πίνακα 8 του παραρτήματος, παρατηρούμε επίσης ότι το κρίσιμο επίπεδο και αυτού του ελέγχου είναι πολύ υψηλό. Πράγματι,

$$\begin{aligned}\hat{\alpha} &= \max P(T^+ \leq 4 \mid H_0) \\ &= \max P(T^+ \leq 4 \mid x_{0.5} \leq 3) \\ &= P(T^+ \leq 4 \mid x_{0.5} = 3) \\ &= 0.40.\end{aligned}$$

Λύση με το MINITAB: Για την διεξαγωγή του ελέγχου Wilcoxon για την διάμεσο ενός δείγματος, το MINITAB υπολογίζει τις διαφορές των τιμών του δείγματος από την υποτιθέμενη διάμεσο και σε κάθε διαφορά αντιστοιχίζει την τάξη μεγέθους της απόλυτης τιμής της. Κατόπιν βασίζει τον έλεγχο στα αθροίσματα των τάξεων μεγέθους (θετικών και αρνητικών διαφορών χωριστά).

Το δείγμα πρέπει να αποθηκευθεί σε μία μεταβλητή (έστω **X**) και κατόπιν πρέπει να επιλέξουμε **Stat, Nonparametrics, 1-Sample Wilcoxon**. Τότε προκύπτει το ακόλουθο πλαίσιο διαλόγου:



Στο πεδίο **Variables,** δηλώνουμε την μεταβλητή που περιέχει το δείγμα. Στο πεδίο **Test median,** δηλώνουμε την υποτιθέμενη τιμή της διαμέσου και, στο πεδίο **Alternative,** την κατεύθυνση της ανισότητας στην εναλλακτική υπόθεση. Η κατεύθυνση, για το συγκεκριμένο πρόβλημα, πρέπει επομένως να δηλωθεί **greater than**. Τα αποτελέσματα που παίρνουμε είναι τα εξής:

Wilcoxon Signed Rank Test

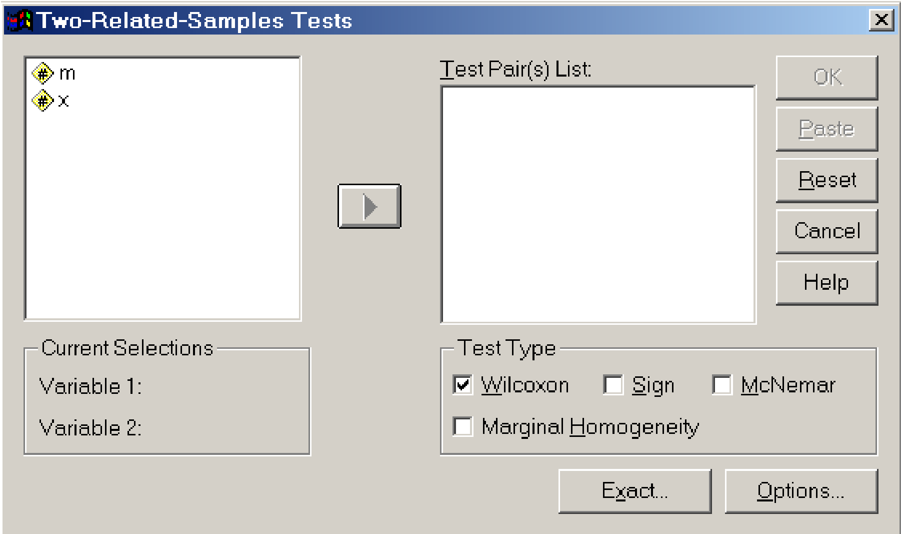
```
Test of median = 3.000 versus median  >  3.000

                N for   Wilcoxon              Estimated
            N    Test  Statistic        P        Median
X           4       4        6.0    0.428         3.950
```

Το MINITAB υπολογίζει τις διαφορές $D_i^{'} = X_i - m$ της διαμέσου από τις παρατηρήσεις (αντί των διαφορών $D_i^{'} = m - X_i$) ως εκ τούτου, μεγάλες θετικές τέτοιες διαφορές θα αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Είναι προφανές ότι μεταξύ των προσημασμένων τάξεων μεγέθους $R_i^{'}$ και R_i των διαφορών $D_i^{'}$ και D_i, αντίστοιχα ισχύει η σχέση $R_i^{'} = -R_i$. Επομένως, η ελεγχοσυνάρτηση που το πακέτο χρησιμοποιεί είναι η T'^{+} = Σ θετικών $R_i^{'}$ = –Σ αρνητικών R_i. Η τιμή της ελεγχοσυνάρτησης αυτής είναι 6 και το κρίσιμο επίπεδο είναι 0.428. Όπως αναμενόταν, το συμπέρασμα στο οποίο οδηγούμεθα σχετικά με το κατά πόσο η μηδενική υπόθεση μπορεί να θεωρηθεί εύλογη δεν διαφέρει από αυτό στο οποίο οδηγηθήκαμε με την αναλυτική λύση του παραδείγματος.

Λύση με το SPSS: Για την διεξαγωγή του ελέγχου Wilcoxon για την διάμεσο ενός πληθυσμού, το δείγμα καταχωρίζεται σε μια μεταβλητή (έστω **x**). Παράλληλα, ορίζεται μια άλλη μεταβλητή (έστω **m**) της ίδιας διάστασης με την **x** και με όλες τις τιμές της ίσες με την υποτιθέμενη τιμή της διαμέσου.

Για τα δεδομένα του παραδείγματος, καταχωρίζουμε στην μεταβλητή **x** το δείγμα και στην μεταβλητή **m** την τιμή 3 τέσσερις φορές. Καλό είναι η στήλη **m** να βρίσκεται δεξιά της **x** γιατί, διαφορετικά, αυτό που εμείς θεωρούμε θετική διαφορά θα θεωρείται από το SPSS αρνητική. (Αυτό δεν αποτελεί πρόβλημα, αλλά μπορεί να προκαλέσει σύγχυση).

Για να διεξαγάγουμε τον έλεγχο Wilcoxon, επιλέγουμε κατά σειρά **Analyze, Nonparametric Tests, 2 Related Samples** και οδηγούμεθα στο ακόλουθο πλαίσιο διαλόγου:



Το πλαίσιο είναι αυτό που χρησιμοποιείται και για τον προσημικό έλεγχο, οπότε ο τρόπος χειρισμού του είναι πλέον γνωστός, μόνο που στο πεδίο **Test Type** τώρα επιλέγουμε **Wilcoxon**. Επειδή έχουμε μικρό δείγμα επιλέγουμε χρησιμοποίηση της ακριβούς κατανομής των στατιστικών συναρτήσεων ελέγχου. Η εκτέλεση του ελέγχου δίνει

Ranks

		N	Mean Rank	Sum of Ranks
M - X	Negative Ranks	2[a]	3.00	6.00
	Positive Ranks	2[b]	2.00	4.00
	Ties	0[c]		
	Total	4		

a M < X
b M > X
c X = M

Test Statistics[b]

	M - X
Z	-.365[a]
Asymp. Sig. (2-tailed)	.715
Exact Sig. (2-tailed)	.875
Exact Sig. (1-tailed)	.438
Point Probability	.125

a Based on positive ranks.
b Wilcoxon Signed Ranks Test

Το SPSS χρησιμοποιεί, ανάλογα με την περίσταση, το άθροισμα των θετικών ή το άθροισμα των αρνητικών τάξεων μεγέθους. Συγκεκριμένα, χρησιμοποιεί το άθροισμα με την μικρότερη τιμή και όταν υπολογίζει το κρίσιμο επίπεδο ενός μονόπλευρου ελέγχου, αναφέρεται σε έλεγχο του οποίου η εναλλακτική υπόθεση υποστηρίζεται από μικρές τιμές του συγκεκριμένου αθροίσματος τάξεων μεγέθους. Στο παράδειγμα μας, είναι μικρότερο το άθροισμα των θετικών τάξεων μεγέθους (γραμμή **Positive Ranks**, στήλη **Sum of Ranks**). Επομένως η τιμή του κρίσιμου επιπέδου (0.438) που δίνει το SPSS αντιστοιχεί στην τιμή του κρίσιμου επιπέδου με εναλλακτική υπόθεση της μορφής

$$H_0 : x_{0.5} > 3,$$

δηλαδή αντιστοιχεί στον έλεγχο που μας ενδιαφέρει.

Σημείωση: Αν είχαμε τον αντίθετης φοράς μονόπλευρο έλεγχο, λόγω του ότι οι τιμές των δύο αθροισμάτων τάξεων μεγέθους έχουν σταθερό άθροισμα που εξαρτάται μόνο από το μέγεθος του δείγματος, η τιμή του κρίσιμου επιπέδου θα ήταν συμπληρωματική αυτής που δίνει το SPSS. Για αμφίπλευρο έλεγχο, το p-value που δίνει το SPSS στο πεδίο Exact Sig. (2-tailed) είναι η τιμή του κρίσιμου επιπέδου που θέλουμε.

Λύση με το SAS: Το SAS διεξάγει τον έλεγχο Signed Rank test υπολογίζοντας την ελεγχοσυνάρτηση S με τον τύπο $\sum R_{i+} - n_t(n_t + 1)/4$, όπου R_{i+} είναι το άθροισμα των θετικών τάξεων μεγέθους και n_t είναι το πλήθος των τιμών, εξαιρώντας τις διαφορές που είναι ίσες με 0. Έτσι, πληκτρολογώντας τις παρακάτω εντολές.

```
data index;
input x @@;
diff=(3-x);
cards;
8.4 -4.3 -0.8 12.5
;
run;
```

```
proc univariate;
var diff;
run;
```

παίρνουμε το αποτέλεσμα που ακολουθεί

```
                              Univariate Procedure   Variable=DIFF

                 Moments                                         Quantiles(Def=5)
 N          4  Sum Wgts     4     100% Max       7.3       99%      7.3
 Mean   -0.95  Sum       -3.8      75% Q3       5.55       95%      7.3
 Std Dev 7.821551  Variance 61.17667 50% Med   -0.8       90%      7.3
 Skewness  -0.06194  Kurtosis  -3.81179  25% Q1   -7.45   10%      -9.5
 USS       187.14  CSS        183.53     0% Min   -9.5     5%      -9.5
 CV       -823.321  Std Mean   3.910776                    1%      -9.5
 T:Mean=0 -0.24292  Pr>|T|     0.8237       Range        16.8
 Num ^= 0       4  Num > 0        2       Q3-Q1        13
 M(Sign)        0  Pr>=|M|     1.0000       Mode        -9.5
 Sgn Rank      -1  Pr>=|S|     0.8750


                                           Extremes

                               Lowest    Obs     Highest    Obs
                                 -9.5(      4)        .(        )
                                 -5.4(      1)     -9.5(       4)
                                  3.8(      3)     -5.4(       1)
                                  7.3(      2)      3.8(       3)
                                    .(       )      7.3(       2)
```

Στο πεδίο **Sgn Rank**, εμφανίζεται η τιμή της ελεγχοσυνάρτησης και κοντά της η τιμή του αμφίπλευρου παρατηρούμενου επιπέδου στατιστικής σημαντικότητας.

Σημειώνεται ότι αν θέλαμε το πακέτο να εκτελέσει τον έλεγχο με την στατιστική συνάρτηση T και να υπολογίσει το μονόπλευρο επίπεδο σημαντικότητας που ζητείται, θα μπορούσαμε να χρησιμοποιήσουμε τις εντολές

```
data d311;
input x @@;
diff=(3-x);
code=1;
if diff>0 then code=2;
absdiff=abs(diff);
if diff=0 then absdiff=.;
cards;
8.4 -4.3 -0.8 12.5
;
proc rank out=ranks;
ranks r;
var absdiff;
run;
proc print data=ranks;
run;
data ranks;
set ranks;
r2=r;
if code=1 then r2=-r;
rr=r2**2;
proc means sum;
```

```
var r2 rr;
output out=sums sum=sumr2 sumrr;
run;
data sums;
set sums;
t=abs(sumr2/(sumrr**(1/2)));
pvalue=(1-cdf('normal',t));
proc print;
run;
```

Η μεταβλητή (**3-X**) με την οποία εργαζόμαστε έχει αποθηκευθεί με το όνομα **diff**, ενώ η μεταβλητή **absdiff** περιέχει τις απόλυτες τιμές των διαφορών (δίδοντας missing value στις τιμές που είναι 0). Η διαδικασία **proc rank** αποδίδει στις τιμές τις τάξεις μεγέθους και τις αποθηκεύει στην μεταβλητή με όνομα **ranks**. Τέλος, η εντολή **pvalue=(1-cdf('normal',t));** δίδει στην μεταβλητή **pvalue** την τιμή της αθροιστικής συνάρτησης της κανονικής κατανομής που αντιστοιχεί στο ζητούμενο επίπεδο στατιστικής σημαντικότητας. Τα αποτελέσματα παρέχονται στον παρακάτω πίνακα.

OBS	X	DIFF	CODE	ABSDIFF	R
1	8.4	-5.4	1	5.4	2
2	-4.3	7.3	2	7.3	3
3	-0.8	3.8	2	3.8	1
4	12.5	-9.5	1	9.5	4

Variable	Sum
R2	-2.0000000
RR	30.0000000

OBS	_TYPE_	_FREQ_	SUMR2	SUMRR	T	PVALUE
1	0	4	-2	30	0.36515	0.35750

Παράδειγμα 3.1.2: Ο υπεύθυνος του τμήματος έρευνας αγοράς μιας μεγάλης εταιρείας κατάρτισε ένα ερωτηματολόγιο με στόχο να μελετήσει τις αντιδράσεις του καταναλωτικού κοινού για ένα προϊόν. Υπέθεσε ότι τα ποσοστά των καταναλωτών στους οποίους άρεσε το προϊόν και αυτών στους οποίους δεν άρεσε το προϊόν είναι ίσα. Για να ελέγξει την υπόθεση αυτή διατύπωσε την εξής ερώτηση: “Πώς αισθάνεσθε για το προϊόν αυτό;” Τα άτομα που επρόκειτο να ερωτηθούν θα έπρεπε να απαντήσουν διαλέγοντας μια απάντηση σε

μια κλίμακα από το 1 ως το 7, στην οποία ο αριθμός 1 υποδήλωνε ότι το προϊόν ήταν της αρεσκείας του ερωτώμενου και ο αριθμός 7 δήλωνε ότι το προϊόν ήταν της απόλυτης αρεσκείας του. Η διάμεση απάντηση, δηλαδή η απάντηση που αντιστοιχούσε στον αριθμό 4, δήλωνε ότι ο ερωτώμενος ήταν αδιάφορος προς το συγκεκριμένο προϊόν. Σε μια ανιχνευτική (πιλοτική) έρευνα που έκανε ενόψει της τελικής διαμόρφωσης του ερωτηματολογίου του, ο ερευνητής πήρε τις εξής απαντήσεις από 12 άτομα:

7 3 5 4 7 1 2 2 5 7 6 5.

Να ελεγχθεί η υπόθεση ότι η διάμεση ανταπόκριση στο συγκεκριμένο προϊόν είναι 4. (α=0.05).

Λύση: Οι προς έλεγχο υποθέσεις διατυπώνονται ως εξής:

$$H_0 : x_{0.5} = 4$$
$$H_1 : x_{0.5} \neq 4.$$

Για τον υπολογισμό της τιμής της ελεγχοσυνάρτησης του Wilcoxon κατασκευάζουμε τον πίνακα που ακολουθεί.

X_i	$D_i = 4 - X_i$	$\lvert D_i \rvert$	$R(\lvert D_i \rvert)$	R_i
7	–3	3	9.5	–9.5
3	+1	1	2.5	+2.5
5	–1	1	2.5	–2.5
4	0	0	–	–

7	–3	3	9.5	–9.5
1	+3	3	9.5	+9.5
2	+2	2	6	+6
2	+2	2	6	+6
5	–1	1	2.5	–2.5
7	–3	3	9.5	–9.5
6	–2	2	6	–6
5	–1	1	2.5	–2.5

Οπως φαίνεται από τα στοιχεία του πίνακα, υπάρχουν αρκετές περιπτώσεις ταύτισης τιμών στο δείγμα των διαφορών. Επομένως, η κατάλληλη στατιστική συνάρτηση για τον έλεγχο των παραπάνω υποθέσεων είναι η

$$T = \sum_{i=1}^{n} R_i \Big/ \sqrt{\sum_{i=1}^{n} R_i^2}\ .$$

Η τιμή της συνάρτησης αυτής, υπολογίζεται από τα δεδομένα ίση με

$$T = -18 \big/ \sqrt{494} = -0.81\ .$$

Η τιμή αυτή βρίσκεται εκτός της κρίσιμης περιοχής μεγέθους 0.05, αφού $|T| < z = 1.96$. Επομένως, σε επίπεδο σημαντικότητας 0.05, δεν απορρίπτεται η μηδενική υπόθεση. Κατά συνέπεια, σε επίπεδο σημαντικότητας 0.05, δεν θα μπορούσαμε να συμπεράνουμε ότι οι αντιδράσεις στο προϊόν έχουν μια διάμεσο διαφορετική από την αντιστοιχούσα στην τιμή 4: “αδιαφορία”.

Το κρίσιμο επίπεδο $\hat{\alpha}$ υπερβαίνει το 20% $\hat{\alpha} > 20\%$). Πράγματι, επειδή η παρατηρηθείσα τιμή της στατιστικής συνάρτησης T είναι αρνητική (μικρότερη από την μέση τιμή της κατανομής της), έχουμε, από τον πίνακα της τυποποιημένης κανονικής κατανομής, ότι

$$\hat{\alpha}/2 = P(T \leq -0.81 \mid x_{0.5} = 4) = 0.209$$

Επομένως, $\hat{\alpha}$ =0.418.

Λύση με το MINITAB: Η χρήση του πακέτου για τον έλεγχο Wilcoxon για την διάμεσο ενός πληθυσμού είναι ήδη γνωστή. Τα αποτελέσματα που μας δίνει το MINITAB είναι:

```
Test of median = 4,000 versus median not = 4.000

                N for   Wilcoxon            Estimated
            N    Test   Statistic        P     Median
x          12      11        42.0    0.450      4.500
```

Η τιμή του κρίσιμου επιπέδου είναι αρκετά μεγάλη ώστε να μπορούμε να θεωρήσουμε την μηδενική υπόθεση ως εύλογη.

Λύση με το SPSS: Η χρήση του πακέτου για τον έλεγχο Wilcoxon για την διάμεσο ενός πληθυσμού είναι ήδη γνωστή. Τα αποτελέσματα που μας δίνει το SPSS είναι:

Ranks

		N	Mean Rank	Sum of Ranks
M - Βαθμολογία	Negative Ranks	7[a]	6,00	42,00
	Positive Ranks	4[b]	6,00	24,00
	Ties	1[c]		
	Total	12		

a M < Βαθμολογία
b M > Βαθμολογία
c Βαθμολογία = M

Test Statistics[b]

M - Βαθμολογία

Z	-,810[a]
Asymp. Sig. (2-tailed)	,418
Exact Sig. (2-tailed)	,481
Exact Sig. (1-tailed)	,241
Point Probability	,041

a Based on positive ranks.
b Wilcoxon Signed Ranks Test

Η τιμή του παρατηρούμενου επιπέδου σημαντικότητας είναι αρκετά μεγάλη ώστενα μπορούμε να θεωρήσουμε την μηδενική υπόθεση ως εύλογη (πεδίο Exact Sig. (2-tailed)).

Λύση με το SAS: Οπως και στο προηγούμενο παράδειγμα, προκειμένου να διεξαγάγουμε τον έλεγχο Wilcoxon θα πρέπει να χρησιμοποιήσουμε τις εντολές:

```
data market;
input x @@;
diff=(4-x);
code=1;
if diff>0 then code=2;
absdiff=abs(diff);
if diff=0 then absdiff=.;
cards;
7 3 5 4 7 1
2 2 5 7 6 5
;
proc rank out=ranks;
ranks r;
var absdiff;
run;
proc print data=ranks;
run;
data ranks;
set ranks;
r2=r;
if code=1 then r2=-r;
rr=r2**2;
proc means sum;
var r2 rr;
output out=sums sum=sumr2 sumrr;
run;
proc print;
run;
data sums;
set sums;
t=abs(sumr2/(sumrr**(1/2)));
pvalue=2*(1-cdf('normal',t));
proc print;
run;
```

Τα αποτελέσματα δίδονται στον πίνακα που ακολουθεί

```
OBS     X     DIFF     CODE     ABSDIFF     R
```

1	7	-3	1	3	9.5
2	3	1	2	1	2.5
3	5	-1	1	1	2.5
4	4	0	1	.	.
5	7	-3	1	3	9.5
6	1	3	2	3	9.5
7	2	2	2	2	6.0
8	2	2	2	2	6.0
9	5	-1	1	1	2.5
10	7	-3	1	3	9.5
11	6	-2	1	2	6.0
12	5	-1	1	1	2.5

Variable	Sum
R2	-18.0000000
RR	494.0000000

OBS	_TYPE_	_FREQ_	SUMR2	SUMRR	T	PVALUE
1	0	12	-18	494	0.80986	0.41802

Θα πρέπει να σημειωθεί ότι, για τον υπολογισμό του παρατηρούμενου επίπεδου στατιστικής σημαντικότητας, έχουμε πολλαπλασιάσει επί 2 σε αντίθεση με το προηγούμενο παράδειγμα, εφόσον σε αυτή την περίπτωση διεξάγουμε αμφίπλευρο έλεγχο.

Παράδειγμα 3.1.3: Ένας παιδοψυχολόγος ενδιαφέρεται να μελετήσει τον βαθμό της κοινωνικότητας παιδιών που πηγαίνουν σε παιδικό σταθμό. Ο παιδοψυχολόγος μετρά την κοινωνικότητα των παιδιών βαθμολογώντας τις απαντήσεις τους σε μία σειρά τυποποιημένων ερωτήσεων σχετικών με διάφορες εικόνες στις οποίες απεικονίζονται διάφορες κοινωνικές καταστάσεις. Ένα τυχαίο δείγμα 30 παιδιών της ίδιας περίπου ηλικίας που ερωτήθηκαν έδωσε τα εξής αποτελέσματα:

23.8	30.7	34.3	36.6	40.6
26.0	31.2	34.9	37.2	41.1
26.9	31.3	35.0	37.3	42.3
27.4	32.8	35.9	37.9	42.8

28.0	33.2	36.1	38.2	44.0
30.3	33.9	36.4	39.6	45.8

Να ελεγχθεί η υπόθεση ότι η μέση βαθμολογία των παιδιών δεν υπερβαίνει την τιμή 30. (α=0.05).

Λύση: Οι προς έλεγχο υποθέσεις έχουν την μορφή

$$H_0 : E(X) \leq 30$$
$$H_1 : E(X) > 30.$$

Ο πίνακας που ακολουθεί συνοψίζει τα βήματα που απαιτούνται για τον υπολογισμό των προσημασμένων τάξεων μεγέθους των παρατηρήσεων του δείγματος.

X_i	$D_i = 30 - X_i$	$\lvert D_i \rvert$	$R(\lvert D_i \rvert)$	R_i
23.8	+6.2	6.2	17	+17
26.0	+4.0	4.0	11	+11
26.9	+3.1	3.1	8	+8
27.4	+2.6	2.6	6	+6
28.0	+2.0	2.0	5	+5
30.3	–0.3	0.3	1	–1
30.7	–0.7	0.7	2	–2
31.2	–1.2	1.2	3	–3
31.3	–1.3	1.3	4	–4
32.8	–2.8	2.8	7	–7
33.2	–3.2	3.2	9	–9
33.9	–3.9	3.9	10	–10
34.3	–4.3	4.3	12	–12
34.9	–4.9	4.9	13	–13
35.0	–5.0	5.0	14	–14
35.9	–5.9	5.9	15	–15
36.1	–6.1	6.1	16	–16
36.4	–6.4	6.4	18	–18
36.6	–6.6	6.6	19	–19
37.2	–7.2	7.2	20	–20
37.3	–7.3	7.3	21	–21
37.9	–7.9	7.9	22	–22
38.2	–8.2	8.2	23	–23
39.6	–9.6	9.6	24	–24
40.6	–10.6	10.6	25	–25
41.1	–11.1	11.1	26	–26
42.3	–12.3	12.3	27	–27
42.8	–12.8	12.8	28	–28
44.0	–14.0	14.0	29	–29
45.8	–15.8	15.8	30	–30

Είναι προφανές, από τα στοιχεία του πίνακα, ότι δεν υπάρχουν περιπτώσεις ταύτισης τιμών στο δείγμα των διαφορών. Επομένως, για τον έλεγχο των παραπάνω υποθέσεων μπορεί να χρησιμοποιηθεί η στατιστική συνάρτηση

$$T^{+} = \text{άθροισμα των θετικών } R_i .$$

Η κρίσιμη περιοχή του ελέγχου μεγέθους 0.05 αντιστοιχεί στην αριστερή ουρά της κατανομής της στατιστικής συνάρτησης T^{+} και ορίζεται από την ανισότητα $T^{+} < w_{0.05}$, όπου $w_{0.05}$ συμβολίζει το 0.05-ποσοστιαίο σημείο της κατανομής της στατιστικής συνάρτησης T^{+}, του οποίου η τιμή προσδιορίζεται από τον πίνακα 8 του παραρτήματος ίση με 152.

Επομένως, η κρίσιμη περιοχή μεγέθους 0.05 αντιστοιχεί σε τιμές της στατιστικής συνάρτησης, οι οποίες είναι μικρότερες από την τιμή 152. Από τα δεδομένα του πίνακα που κατασκευάσθηκε παραπάνω, προκύπτει ότι η τιμή της ελεγχοσυνάρτησης είναι ίση με $\tau^{+} = 47$. Η τιμή αυτή είναι, προφανώς, μέσα στην κρίσιμη περιοχή και, κατά συνέπεια, οδηγεί σε απόρριψη της μηδενικής υπόθεσης σε επίπεδο σημαντικότητας 5%. Άρα, σε επίπεδο σημαντικότητας 0.05 τα δεδομένα παρέχουν ενδείξεις ότι η μέση βαθμολογία των παιδιών υπερβαίνει την τιμή 30.

Το κρίσιμο επίπεδο $\hat{\alpha}$ μπορεί εύκολα να διαπιστωθεί ότι έχει πολύ χαμηλή τιμή. Πράγματι, από τον πίνακα 8 του παραρτήματος, προκύπτει ότι η τιμή του ποσοστιαίου σημείου $w_{\hat{\alpha}} \equiv \tau^{+} = 47$ αντιστοιχεί σε τιμή του $\hat{\alpha}$ μικρότερη του 0.005. Η τιμή του $\hat{\alpha}$ μπορεί να υπολογισθεί κατά προσέγγιση από την σχέση που συνδέει τα ποσοστιαία σημεία w_p της κατανομής της T^{+} με τα αντίστοιχα ποσοστιαία σημεία z_p της τυποποιημένης κανονικής, δηλαδή από την σχέση

$$w_p \cong \frac{n(n+1)}{4} + z_p \sqrt{\frac{n(n+1)(2n+1)}{24}}.$$

Θέτοντας $p = \hat{\alpha}$, έχουμε

$$47 \equiv w_{\hat{\alpha}} = \frac{(30)(31)}{4} + z_{\hat{\alpha}} \sqrt{\frac{(30)(31)(61)}{24}},$$

ή, ισοδύναμα,

$$z_{\hat{\alpha}} = -3.82.$$

Από τον πίνακα της τυποποιημένης κανονικής κατανομής, προκύπτει ότι $\hat{\alpha} = 0.0001$.

Λύση με το MINITAB: Η χρήση του πακέτου για τον έλεγχο Wilcoxon-Mann-Whitney για την διάμεσο ενός πληθυσμού είναι ήδη γνωστή. Τα αποτελέσματα που μας δίνει το MINITAB είναι:

```
Test of median = 30.00 versus median not = 30.00

             N for   Wilcoxon            Estimated
          N   Test  Statistic        P      Median
x        30     30      418.0    0.000       35.15
```

Η τιμή του κρίσιμου επιπέδου είναι πάρα πολύ μικρή και κατά συνέπεια η μηδενική υπόθεση δεν μπορεί να θεωρηθεί εύλογη.

Λύση με το SPSS: Η χρήση του πακέτου για τον έλεγχο Wilcoxon-Mann-Whitney για την διάμεσο ενός πληθυσμού είναι ήδη γνωστή. Τα αποτελέσματα που μας δίνει το SPSS είναι:

Ranks

		N	Mean Rank	Sum of Ranks
M - Βαθμός κοινωνικότητας	Negative Ranks	25[a]	16,72	418,00
	Positive Ranks	5[b]	9,40	47,00
	Ties	0[c]		
	Total	30		

a M < Βαθμός κοινωνικότητας

b M > Βαθμός κοινωνικότητας
c Βαθμός κοινωνικότητας = M

Test Statistics[b]

	M - Βαθμός κοινωνικότητας
Z	-3,815[a]
Asymp. Sig. (2-tailed)	,000
Exact Sig. (2-tailed)	,000
Exact Sig. (1-tailed)	,000
Point Probability	,000

a Based on positive ranks.
b Wilcoxon Signed Ranks Test

Το μονόπλευρο παρατηρούμενο επίπεδο σημαντικότητας (πεδίο Exact Sig. (1-tailed)) είναι μικρότερο του 0.001. Κατά συνέπεια, η μηδενική υπόθεση ότι ο μέσος βαθμός κοινωνικότητας των παιδιών δεν υπερβαίνει το 30 δεν φαίνεται να είναι εύλογη.

Λύση με το SAS: Πληκτρολογούμε και πάλι τις ακόλουθες εντολές

```
data kids;
input x @@;
diff=(30-x);
code=1;
if diff>0 then code=2;
absdiff=abs(diff);
cards;
23.8 30.7 34.3 36.6 40.6
26.0 31.2 34.9 37.2 41.1
26.9 31.3 35.0 37.3 42.3
27.4 32.8 35.9 37.9 42.8
28.0 33.2 36.1 38.2 44.0
30.3 33.9 36.4 39.6 45.8
;
proc rank out=ranks;
ranks r;
var absdiff;
run;
proc print data=ranks;
run;
data ranks;
set ranks;
r2=r;
if code=1 then r2=-r;
rr=r2**2;
proc means sum;
var r2 rr;
output out=sums sum=sumr2 sumrr;
run;
proc print;
run;
data sums;
set sums;
t=sumr2/(sumrr**(1/2));
pvalue=cdf('normal',t);
proc print;
run;
```

Η εκτέλεση των εντολών μας δίδει τα αποτελέσματα που ακολουθούν

OBS	X	DIFF	CODE	ABSDIFF	R
1	23.8	6.2	2	6.2	17
2	30.7	-0.7	1	0.7	2
3	34.3	-4.3	1	4.3	12
4	36.6	-6.6	1	6.6	19
5	40.6	-10.6	1	10.6	25
6	26.0	4.0	2	4.0	11
7	31.2	-1.2	1	1.2	3
8	34.9	-4.9	1	4.9	13
9	37.2	-7.2	1	7.2	20
10	41.1	-11.1	1	11.1	26
11	26.9	3.1	2	3.1	8
12	31.3	-1.3	1	1.3	4
13	35.0	-5.0	1	5.0	14
14	37.3	-7.3	1	7.3	21
15	42.3	-12.3	1	12.3	27
16	27.4	2.6	2	2.6	6
17	32.8	-2.8	1	2.8	7
18	35.9	-5.9	1	5.9	15
19	37.9	-7.9	1	7.9	22
20	42.8	-12.8	1	12.8	28
21	28.0	2.0	2	2.0	5
22	33.2	-3.2	1	3.2	9
23	36.1	-6.1	1	6.1	16
24	38.2	-8.2	1	8.2	23
25	44.0	-14.0	1	14.0	29
26	30.3	-0.3	1	0.3	1
27	33.9	-3.9	1	3.9	10
28	36.4	-6.4	1	6.4	18
29	39.6	-9.6	1	9.6	24
30	45.8	-15.8	1	15.8	30

Variable	Sum
R2	-371.0000000
RR	9455.00

OBS	_TYPE_	_FREQ_	SUMR2	SUMRR	T	PVALUE
1	0	30	-371	9455	-3.81543	.000067974

Το μονόπλευρο παρατηρούμενο επίπεδο στατιστικής σημαντικότητας υπολογίσθηκε μικρότερο από 0.0001 και, κατά συνέπεια, η μηδενική υπόθεση δεν μπορεί να θεωρηθεί εύλογη σε οποιοδήποτε από τα συνήθη επίπεδα σημαντικότητας α.

Παρατήρηση: Ο προσημασμένος έλεγχος των Wilcoxon είναι το μη παραμετρικό ανάλογο του ελέγχου Z για την μέση τιμή ενός ληθυσμού με γνωστή διασπορά στην περίπτωση που η κατανομή του δεν είναι η κανονική.

3.1.2 Ο Ελεγχος των προσημασμένων Τάξεων Μεγέθους του Wilcoxon για Δείγμα Ζευγών Παρατηρήσεων

(The Wilcoxon Signed Rank Test with Paired Data)

Στην ενότητα αυτή θα περιγράψουμε την χρήση του ελέγχου των προσημασμένων τάξεων μεγέθους του Wilcoxon σε προβλήματα σύγκρισης των θέσεων δύο πληθυσμών, όταν η συγκριτική μελέτη βασίζεται σε ένα δείγμα ζευγών παρατηρήσεων. Τα δεδομένα, δηλαδή, αποτελούνται από n΄ ζεύγη τιμών (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) που αποτελούν πραγματοποίηση της ακολουθίας των διδιάστατων τυχαίων μεταβλητών (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), η οποία αποτελεί το τυχαίο δείγμα μεγέθους n΄ από τον διδιάστατο πληθυσμό που περιγράφεται από την μεταβλητή (X, Y).

Θεωρούμε τις διαφορές $D_i = Y_i – X_i$, i = 1, 2, ..., n΄. Οπως και στην περίπτωση του ενός δείγματος, θεωρούμε τις απόλυτες διαφορές $|D_i| = | Y_i – X_i |$, i = 1, 2, ..., n΄, παραλείπουμε (αγνοούμε) τα ζεύγη (X_i, Y_i) για τα οποία $X_i = Y_i$, δηλαδή $D_i = 0$, και διατάσσουμε κατά αύξουσα τάξη μεγέθους τις n ($n \leq n'$) διαφορές $|D_i|$ που απομένουν. Στην συνέχεια, αντιστοιχίζουμε στα n ζεύγη παρατηρήσεων που απομένουν βαθμούς από το 1 μέχρι το n, οι οποίοι δεν είναι τίποτε άλλο από τις τάξεις μεγέθους των απολύτων διαφορών $|D_i| = |Y_i – X_i|$, i = 1, 2, ..., n. Ετσι, ο βαθμός 1 αντιστοιχεί στο ζεύγος (X_i, Y_i) με την μικρότερη απόλυτη διαφορά $|D_i|$, ο βαθμός 2 αντιστοιχεί στο ζεύγος (X_i, Y_i) με την αμέσως μεγαλύτερη απόλυτη διαφορά, ... , ο βαθμός n αντιστοιχεί στο ζεύγος (X_i, Y_i) με την μεγαλύτερη απόλυτη διαφορά. Αν αρκετά ζεύγη έχουν απόλυτες διαφορές οι οποίες ταυτίζονται, τότε σε κάθε ένα από αυτά τα ζεύγη παρατηρήσεων αντιστοιχίζεται ως βαθμός ο μέσος των βαθμών που τα ζεύγη θα είχαν αν οι διαφορές τους δεν ήταν ίσες.

Οπως και στην περίπτωση του ενός δείγματος, υποθέτουμε ότι η κατανομή της μεταβλητής D=Y–X είναι συμμετρική και ότι η κλίμακα μέτρησης των δεδομένων είναι τουλάχιστον κλίμακα διαστήματος.

Συμβολίζουμε με $d_{0.5}$ την διάμεσο του πληθυσμού των διαφορών, ο οποίος περιγράφεται από την μεταβλητή D =Y–X. Προφανώς, η τιμή της διαμέσου $d_{0.5}$ παρέχει πληροφορίες για τις σχετικές θέσεις των δύο πληθυσμών. Οταν οι τιμές της τυχαίας μεταβλητής Y είναι μεγαλύτερες από τις αντίστοιχες τιμές της τυχαίας μεταβλητής X και αυτό συμβαίνει τις περισσότερες φορές στον πληθυσμό των διαφορών D = Y–X, τότε η διάμεσος διαφορά είναι θετική, δηλαδή $d > 0$. Με όμοιο τρόπο, όταν οι τιμές της Y είναι μικρότερες από τις αντίστοιχες τιμές της X και αυτό συμβαίνει τις περισσότερες φορές στον πληθυσμό των διαφορών D=Y–X, τότε η διάμεσος διαφορά είναι αρνητική, δηλαδή $d_{0.5} > 0$. Τέλος, αν οι τιμές της μεταβλητής D=Y–X στον πληθυσμό των διαφορών είναι θετικές και αρνητικές με ίσες συχνότητες, τότε η διάμεσος διαφορά είναι ίση με το μηδέν, δηλαδή $d_{0.5} = 0$. Είναι, δηλαδή, η διάμεσος $d_{0.5}$ μία παράμετρος η οποία μετρά την διαφορά στις θέσεις των δύο πληθυσμών. Αυτό, εξ άλλου, είναι προφανές και από το γεγονός ότι στην περίπτωση που η κατανομή των διαφορών είναι συμμετρική, η διάμεσος του πληθυσμού ταυτίζεται με την μέση τιμή του.

Είναι, επομένως, προφανές από τα προηγούμενα ότι οι υποθέσεις τις οποίες ενδέχεται να ενδιαφερόμαστε να ελέγξουμε διακρίνονται στις εξής κατηγορίες:

Α. (Μονόπλευρος ελέγχος):

H_0: $d_{0.5} \leq 0$

H_1: $d_{0.5} > 0$

Β. (Μονόπλευρος έλεγχος):

H_0: $d_{0.5} \geq 0$

H_1: $d_{0.5} < 0$

Γ. (Μονόπλευρος έλεγχος):

H_0: $d_{0.5} = 0$

H_1: $d_{0.5} \neq 0$.

Η στατιστική συνάρτηση η οποία χρησιμοποιείται για τον έλεγχο των υποθέσεων των παραπάνω περιπτώσεων βασίζεται και πάλι στις προσημασμένες τάξεις μεγέθους R_i, i = 1, 2, ..., n των απολύτων διαφορών $|D_i|$, i = 1, 2, ..., n, οι οποίες ορίζονται όπως και στην περίπτωση του ενός δείγματος. Δηλαδή,

$$R_i = \begin{cases} +R(|D_i|), & \text{αν } D_i \equiv Y_i - X_i > 0 \\ -R(|D_i|), & \text{αν } D_i \equiv Y_i - X_i < 0 \end{cases}, \qquad i = 1, 2, ..., n.$$

Συγκεκριμένα, για τον έλεγχο των υποθέσεων των περιπτώσεων Α, Β και Γ χρησιμοποιείται, όπως ακριβώς και στην περίπτωση του ενός δείγματος, η στατιστική συνάρτηση

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}},$$

ή, όταν όλες οι διαφορές $D_i = Y_i - X_i$ είναι διακεκριμμένες, μία από τις εναλλακτικές στατιστικές συναρτήσεις

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{n(n+1)(2n+1)/6}}$$

ή

$$T^+ = \text{άθροισμα των θετικών } R_i.$$

Οι κρίσιμες περιοχές των ελέγχων των υποθέσεων των μορφών Α, Β και Γ καθορίζονται όπως και στην προηγούμενη ενότητα με την βοήθεια του πίνακα 8 του παραρτήματος, στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T^+ ή του πίνακα 2 της τυποποιημένης κανονικής κατανομής, στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση Τ. Πιο συγκεκριμένα, οι κρίσιμες περιοχές μεγέθους α των ελέγχων των παραπάνω περιπτώσεων ορίζονται από τις ανισότητες:

Α. (Μονόπλευρος έλεγχος): $T > z_{1-\alpha}$ ή $T^+ > w_{1-\alpha}$

Β. (Μονόπλευρος έλεγχος): $T < z_{\alpha} = -z_{1-\alpha}$ ή $T^+ < w_{\alpha}$

Γ. (Αμφίπλευρος έλεγχος):

$$T < z_{\alpha/2} \text{ και } T > z_{1-\alpha/2} \quad \text{ή} \quad T^+ < w_{\alpha/2} \text{ και } T^+ > w_{1-\alpha/2}.$$

Παρατήρηση: Για την διατύπωση του προβλήματος του ελέγχου του Wilcoxon για την περίπτωση δύο μη ανεξαρτήτων δειγμάτων, υποθέσαμε ότι η ακολουθία των διδιάστατων τυχαίων μεταβλητών (X_1, Y_1), (X_2, Y_2), ..., $(X_{n'}, Y_{n'})$ αποτελούσε τυχαίο δείγμα από τον πληθυσμό των τιμών του τυχαίου διανύσματος (X, Y). Στην πραγματικότητα, για την διεξαγωγή του ελέγχου του Wilcoxon απαιτείται μία πιο ήπια υπόθεση. Συγκεκριμένα, αρκούσε να υποθέσουμε ότι η ακολουθία των διδιάστατων μεταβλητών (X_1, Y_1), (X_2, Y_2), ..., $(X_{n'}, Y_{n'})$ ορίζεται έτσι ώστε οι διαφορές D_i =

$Y_i - X_i$, $i = 1, 2, ..., n'$, να είναι ανεξάρτητες και να έχουν την ίδια διάμεσο. Το πλεονέκτημα, όμως, της αυστηρότερης υπόθεσης που θεωρήθηκε για την ακολουθία $(X_1, Y_1), (X_2, Y_2), ..., (X_{n'}, Y_{n'})$ είναι ότι η υπόθεση αυτή επιτρέπει την αντικατάσταση της παραμέτρου $d_{0.5}$ από την παράμετρο $E(X) - E(Y)$ στις υποθέσεις των περιπτώσεων Α, Β και Γ, οποτεδήποτε βέβαια οι μέσες τιμές $E(X)$ και $E(Y)$ υπάρχουν και η χρησιμοποίησή τους έχει έννοια στο πλαίσιο του εξεταζόμενου προβλήματος.

Παράδειγμα 3.1.4: Ενας επιχειρησιακός ερευνητής ενδιαφέρεται να μελετήσει το αποτέλεσμα που η ποσότητα πληροφόρησης έχει πάνω στις πιστωτικές αποφάσεις. Για τον σκοπό αυτό, 40 εθελοντές managers πιστώσεων χωρίσθηκαν σε 20 ζεύγη, με τρόπο ώστε οι managers σε κάθε ζεύγος να έχουν παρόμοιο υπόβαθρο όσο αφορά την εμπειρία τους, τη μόρφωσή τους κ.λ.π.. Ο επιχειρησιακός ερευνητής προετοίμασε δύο υποθετικές περιπτώσεις για μια υποθετική επιχείρηση η οποία ζητά εμπορική πίστωση. Η πρώτη περίπτωση έδινε περιορισμένη πληροφόρηση για την επιχείρηση (για παράδειγμα, τις σειρές προϊόντων της, το χρονικό διάστημα λειτουργίας της, τα περιουσιακά της στοιχεία). Η δεύτερη περίπτωση έδινε λεπτομερή πληροφόρηση, η οποία περιελάμβανε οικονομικά δεδομένα όπως και δεδομένα πάνω στον τρόπο λειτουργίας της επιχείρησης για τα τελευταία χρόνια. Για κάθε ζεύγος managers, χρησιμοποιήθηκε ένας τυχαίος αριθμός για να αποφασισθεί ποιος από τους δύο θα μελετούσε την περίπτωση 1 και ποιος την περίπτωση 2. Αφού οι managers μελέτησαν τις περιπτώσεις που τους δόθηκαν, ζητήθηκε από αυτούς να προσδιορίσουν το μέγιστο ποσό πίστωσης που θα έδιναν στην εταιρεία. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί.

Ζεύγος (i) :	1	2	3	4	5	6	7	8	9	10
Περίπτωση 1 (X_i) :	85	50	85	70	30	75	67	60	60	85
Περίπτωση 2 (Y_i) :	90	60	113	100	45	99	86	85	85	87

Ζεύγος (i) :	11	12	13	14	15	16	17	18	19	20
Περίπτωση 1 (X_i) :	70	85	70	77	70	50	55	75	81	20
Περίπτωση 2 (Y_i) :	50	90	93	115	62	85	72	125	103	60

Με βάση τα στοιχεία του παραπάνω πίνακα, νομίζετε ότι ο επιχειρησιακός ερευνητής θα μπορούσε να συμπεράνει ότι η διάμεσος διαφορά στα μέγιστα ποσά πίστωσης που έδωσαν οι managers της ομάδας με περιορισμένη πληροφόρηση και στα ποσά που έδωσαν οι managers της ομάδας με την λεπτομερή πληροφόρηση είναι μηδέν; (α=0.05)

Λύση: Είναι προφανές ότι ένα κατάλληλο ζευγάρι υποθέσεων, ο έλεγχος των οποίων θα μπορούσε να δώσει απάντηση στο παραπάνω ερώτημα, είναι το εξής:

H_0: διάμεσος διαφορά = 0

H_1: διάμεσος διαφορά $\neq$ 0.

Αν συμβολίσουμε την διάμεσο διαφορά των ποσών που προσδιόρισαν οι managers με $d_{0.5}$, οι παραπάνω υποθέσεις γράφονται με τη μορφή

H_0: $d_{0.5} = 0$

H_1: $d_{0.5} \neq 0$.

Ο πίνακας που ακολουθεί συνοψίζει τα βήματα που απαιτούνται για την εφαρμογή του ελέγχου του Wilcoxon στα παραπάνω δεδομένα.

Ποσό πίστωσης (εκατομμύρια δραχμές)

Ζεύγος (i)	Περίπτωση 1 (X_i)	Περίπτωση 2 (Y_i)	$D_i=Y_i–X_i$	$\lvert D_i \rvert$	$R(\lvert D_i \rvert)$	R_i
1	85	90	+5	5	2.5	+2.5
2	50	60	+10	10	5	+5
3	85	113	+28	28	15	+15
4	70	100	+30	30	16	+16
5	30	45	+15	15	6	+6
6	75	99	+24	24	12	+12
7	67	86	+19	19	8	+8
8	60	85	+25	25	13.5	+13.5
9	60	85	+25	25	13.5	+13.5
10	85	87	+2	2	1	+1
11	70	50	–20	20	9	–9
12	85	90	+5	5	2.5	+2.5
13	70	93	+23	23	11	+11
14	77	115	+38	38	18	+18
15	70	62	–8	8	4	–4
16	50	85	+35	35	17	+17
17	55	72	+17	17	7	+7
18	75	125	+50	50	20	+20
19	81	103	+22	22	10	+10
20	20	60	+40	40	19	+19

Οπως φαίνεται από τα στοιχεία του πίνακα, οι τιμές των προσημασμένων διαφορών R_i δεν είναι διακεκριμένες. Επομένως, η μορφή της στατιστικής συνάρτησης του Wilcoxon, η οποία θα χρησιμοποιηθεί είναι η

$$T = \frac{\sum_{i=1}^{20} R_i}{\sqrt{\sum_{i=1}^{20} R_i^2}}.$$

Η κρίσιμη περιοχή μεγέθους 0.05 του ελέγχου αυτού ορίζεται από τις ανισότητες $T < z_{0.025}$ και $T > z_{0.975}$. Δηλαδή, η μηδενική υπόθεση H_0 θα απορριφθεί σε επίπεδο σημαντικότητας 0.05, αν $T < z_{0.025} = -196$, ή $T > z_{0.975} = 1.96$. Από τα δεδομένα του πίνακα, προκύπτει ότι η τιμή τ της στατιστικής συνάρτησης T είναι ίση με $\tau = \frac{184}{\sqrt{2869}} = 3.435$.

Η παρατηρηθείσα τιμή της στατιστικής συνάρτησης T βρίσκεται, επομένως, μέσα στην κρίσιμη περιοχή μεγέθους 0.05. Άρα, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας 0.05 και, κατά συνέπεια, θα μπορούσε να συμπεράνει κανείς, με πιθανότητα σφάλματος τύπου I ίση με 5%, ότι η διάμεσος των διαφορών στα μέγιστα ποσά πίστωσης που καθόρισαν οι managers των δύο κατηγοριών διαφέρει από το μηδέν.

Το κρίσιμο επίπεδο του ελέγχου αυτού είναι πολύ χαμηλό, όπως εύκολα μπορεί να διαπιστωθεί. Πράγματι, επειδή $\tau = 3.435 > E(T) = 0$, το κρίσιμο επίπεδο του ελέγχου υπολογίζεται με βάση τον πίνακα 2 της τυποποιημένης κανονικής κατανομής ως εξής:

$$\frac{\hat{\alpha}}{2} = P(T \geq 3.435 \mid H_0)$$

$$= 1 - P(T < 3.435 \mid d_{0.5} = 0)$$

$$1 - 0.9997$$

$$= 0.0003.$$

Επομένως, $\hat{\alpha} = 0.0006$.

Λύση με το MINITAB: Για την διεξαγωγή του ελέγχου Wilcoxon για δείγμα ζευγών παρατηρήσεων με το MINITAB, πρέπει να καταχωρίσουμε τα δύο δείγματα σε δύο διαφορετικές μεταβλητές και, στην συνέχεια, να καταχωρίσουμε το δείγμα των διαφορών των ζευγών σε μία τρίτη μεταβλητή. Την μεταβλητή αυτή μεταχειριζόμαστε ως ένα δείγμα από τον πληθυσμό των διαφορών των δύο αρχικών πληθυσμών για την διάμεσο του οποίου ελέγχουμε (με τον έλεγχο Wilcoxon) την υπόθεση ότι ισούται με 0.

Τα αποτελέσματα περιέχονται στον παρακάτω πίνακα.

```
Wilcoxon Signed Rank Test

Test of median = 0.000000 versus median not = 0.000000

                N for   Wilcoxon            Estimated
            N   Test  Statistic        P       Median
D          20     20      197.0    0.001        20.25
```

Στον αμφίπλευρο έλεγχο μπορούμε να χρησιμοποιήσουμε οποιοδήποτε από τα δύο αθροίσματα τάξεων μεγέθους. Εδώ, το MINITAB χρησιμοποιεί αυτό των θετικών διαφορών. Η τιμή αυτού του αθροίσματος που προκύπτει είναι πολύ μεγάλη με πολύ χαμηλό αντίστοιχο κρίσιμο επίπεδο. Κατά συνέπεια, η H_0 δεν μπορεί να θεωρηθεί εύλογη.

Σημείωση: Υπενθυμίζεται ότι όταν η τιμή του ενός αθροίσματος είναι πολύ μεγάλη η τιμή του άλλου είναι πολύ μικρή και ότι η τιμή του κρίσιμου επιπέδου για αμφίπλευρο έλεγχο είναι η ίδια όποιο άθροισμα κι αν χρησιμοποιήσουμε.

Λύση με το SPSS: Για να λύσουμε το παράδειγμα αυτό, πρέπει να καταχωρίσουμε τα δύο δείγματα σε δύο χωριστές μεταβλητές, έστω **x** και **y**. Στην συνέχεια, δηλώνουμε κατά σειρά **Analyze, Nonparametric Tests, 2 Related Samples** και η συνέχεια είναι διαδικαστικά απλή. Τα δείγματα έχουν επαρκές μέγεθος για να χρησιμοποιήσουμε την ασυμπτωτική κανονική προσέγγιση της κατανομής των αθροισμάτων των τάξεων μεγέθους. Τα αποτελέσματα που δίνει το SPSS είναι:

Ranks

		N	Mean Rank	Sum of Ranks
Y - X	Negative Ranks[a]	2	6.50	13.00
	Positive Ranks[b]	18	10.94	197.00
	Ties[c]	0		
	Total	20		

a Y < X
b Y > X
c X = Y

Test Statistics[b]

	Y - X
Z	-3.435[a]
Asymp. Sig. (2-tailed)	.001

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

Παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης T, η οποία στον πίνακα αποτελεσμάτων συμβολίζεται με Z, έχει υπολογισθεί με βάση το άθροισμα των αρνητικών τάξεων μεγέθους μια και αυτό έχει την μικρότερη τιμή από τα δύο. Επειδή, βέβαια, έχουμε αμφίπλευρο έλεγχο, η παραπάνω παρατήρηση δεν έχει ιδιαίτερη σημασία. Η τιμή του κρίσιμου επιπέδου είναι 0.001 και κατά συνέπεια, η μηδενική υπόθεση δεν είναι μία εύλογη υπόθεση.

Σημείωση: Στην περίπτωση μονόπλευρου ελέγχου, θα διαιρούσαμε την τιμή p-value που δίνει το πρόγραμμα διά 2 και ανάλογα με την κατεύθυνση του ελέγχου θα βγάζαμε τα συμπεράσματά μας σχετικά με το εύλογο ή μη της μηδενικής υπόθεσης. (βλέπε παρ. 3.1.1).

Λύση με το SAS: Πληκτρολογούμε τις εντολές:

```
data market;
input x y @@;
diff=(y-x);
code=1;
if diff>0 then code=2;
absdiff=abs(diff);
cards;
85 90 50 60 85 113 70 100 30 45
75 99 67 86 60  85 60  85 85 87
70 50 85 90 70  93 77 115 70 62
50 85 55 72 75 125 81 103 20 60
;
proc rank out=ranks;
ranks r;
var absdiff;
run;
proc print data=ranks;
run;
data ranks;
set ranks;
r2=r;
if code=1 then r2=-r;
rr=r2**2;
proc means sum;
var r2 rr;
output out=sums sum=sumr2 sumrr;
run;
proc print;
run;
data sums;
set sums;
t=sumr2/(sumrr**(1/2));
pvalue=2*(1-cdf('normal',t));
proc print;
run;
```

Τα αποτελέσματα που παίρνουμε είναι τα εξής:

OBS	X	Y	DIFF	CODE	ABSDIFF	R
1	85	90	5	2	5	2.5
2	50	60	10	2	10	5.0
3	85	113	28	2	28	15.0
4	70	100	30	2	30	16.0
5	30	45	15	2	15	6.0
6	75	99	24	2	24	12.0
7	67	86	19	2	19	8.0
8	60	85	25	2	25	13.5

9	60	85	25	2	25	13.5
10	85	87	2	2	2	1.0
11	70	50	-20	1	20	9.0
12	85	90	5	2	5	2.5
13	70	93	23	2	23	11.0
14	77	115	38	2	38	18.0
15	70	62	-8	1	8	4.0
16	50	85	35	2	35	17.0
17	55	72	17	2	17	7.0
18	75	125	50	2	50	20.0
19	81	103	22	2	22	10.0
20	20	60	40	2	40	19.0

Variable	Sum
R2	184.0000000
RR	2869.00

OBS	_TYPE_	_FREQ_	SUMR2	SUMRR	T	PVALUE
1	0	20	184	2869	3.43520	.00059211

Το παρατηρούμενο επίπεδο στατιστικής σημαντικότητας έχει τιμή μικρότερη του 0.001. Επομένως, η μηδενική υπόθεση δεν μπορεί να θεωρηθεί εύλογη.

Παράδειγμα 3.1.5: Η γρίπη και οι συνέπειές της θεωρούνται επικίνδυνες για άτομα ηλικίας 60 ετών και άνω και πολλοί γιατροί συνιστούν στους ασθενείς τους που ανήκουν σε αυτή την κατηγορία ηλικίας να εμβολιάζονται με ένα αντιγριπικό εμβόλιο κάθε φθινόπωρο.

Ας υποθέσουμε ότι μία φαρμακευτική εταιρεία κατασκευάζει ένα νέο εμβόλιο για μία συγκεκριμένη μορφή γρίπης. Για να ελεγχθούν τα αποτελέσματα αυτού του εμβολίου, απαιτείται η γνώση της μεταβολής της θερμοκρασίας του σώματος που παρατηρείται ακριβώς πριν από τον εμβολιασμό και μία ώρα μετά τον εμβολιασμό. Εάν το εμβόλιο είναι αποτελεσματικό, ο ασθενής θα πρέπει να παρουσιάσει συμπτώματα ελαφράς γρίπης (ελαφρά καταρροή, ελαφρό πονόλαιμο, ρίγη και άνοδο της θερμοκρασίας του σώματος) μέσα σε μία ώρα από τον εμβολιασμό. Τα αποτελέσματα που ακολουθούν αναφέρονται σε ένα δείγμα n΄=10 αρρένων ασθενών (ηλικίας 60 ετών και άνω), οι οποίοι συμμετείχαν στο πείραμα εθελοντικά.

Θερμοκρασία (?C)	Εθελοντής (i) 1	2	3	4	5	6	7	8	9	10
Πριν (X_i)	37.0	37.0	36.4	36.7	37.0	36.9	37.0	37.0	36.8	37.0
Μετά (Y_i)	38.0	37.2	37.3	38.6	37.8	36.9	36.9	39.6	37.6	37.5

Αποτελούν τα δεδομένα του πίνακα ένδειξη σημαντικής αύξησης στη θερμοκρασία του σώματος αρρένων ασθενών ηλικίας 60 ετών και άνω μία ώρα μετά τον εμβολιασμό; (α=0.05)

Λύση: Οι προς έλεγχο υποθέσεις έχουν την μορφή

$$H_0 : d_{0.5} \geq 0$$

$$H_1 : d_{0.5} > 0$$

όπου $d_{0.5}$ παριστάνει την διάμεσο του πληθυσμού των διαφορών της θερμοκρασίας πριν και μετά τον εμβολιασμό. Για τον έλεγχο των υποθέσεων αυτών με την ελεγχοσυνάρτηση του Wilcoxon, κατασκευάζουμε τον πίνακα που ακολουθεί.

			Θερμοκρασία (° C)			
Εθελοντής i	Πριν X_i	Μετά Y_i	$D_i = Y_i - X_i$	$\lvert D_i \rvert$	$R(\lvert D_i \rvert)$	R_i
1	37.0	38.0	+1.0	1.0	7	+7
2	37.0	37.2	+0.2	0.2	2	+2
3	36.4	37.3	+0.9	0.9	6	+6
4	36.7	38.6	+1.9	1.9	8	+8
5	37.0	37.8	+0.8	0.8	4.5	+4.5
6	36.9	36.9	0.0	0.0	–	–
7	37.0	36.9	–0.1	0.1	1	–1
8	37.0	39.6	+2.6	2.6	9	+9
9	36.8	37.6	+0.8	0.8	4.5	+4.5
10	37.0	37.5	+0.5	0.5	3	+3

Οι τιμές των διαφορών D_i δεν είναι διακεκριμένες. Επομένως, η κατάλληλη ελεγχοσυνάρτηση τύπου Wilcoxon θα έχει την μορφή

$$T = \frac{\sum_{i-1}^{10} R_i}{\sqrt{\sum_{i=1}^{10} R_i^2}}.$$

Από την μορφή της εναλλακτικής υπόθεσης, είναι προφανές ότι η κρίσιμη περιοχή μεγέθους 0.05 του ελέγχου ορίζεται από την ανισότητα $T > z_{0.975} = 1.645$. Από τα δεδομένα του πίνακα, προκύπτει ότι η τιμή τ της στατιστικής συνάρτησης T είναι ίση με

$$\tau = \frac{43}{\sqrt{284.5}} = 2.549.$$

Προφανώς, η παρατηρούμενη τιμή της στατιστικής συνάρτησης T βρίσκεται μέσα στην κρίσιμη περιοχή μεγέθους 0.05 και, κατά συνέπεια, η μηδενική υπόθεση απορρίπτεται. Άρα, σε επίπεδο σημαντικότητας 5% τα δεδομένα παρέχουν ενδείξεις υπέρ της υπόθεσης ότι η θερμοκρασία του σώματος αυξάνει σημαντικά μεταξύ αρρένων ηλικίας 60 ετών και άνω μέσα σε μία ώρα από την στιγμή του εμβολιασμού και, επομένως, το νέο εμβόλιο είναι αποτελεσματικό.

Το κρίσιμο επίπεδο $\hat{\alpha}$ υπολογίζεται από τον πίνακα της τυποποιημένης κανονικής ως εξής:

$$\begin{aligned}\hat{\alpha} &= \max P(T \geq 2.549 \mid H_0) \\ &= P(T \geq 2.549 \mid d_{0.5} = 0) \\ &= 1 - P(T \leq 2.549 \mid d_{0.5} = 0) \\ &= 1 - 0.9946 \\ &= 0.0054.\end{aligned}$$

Λύση με το MINITAB: Καταχωρίζουμε τις διαφορές Y_i-X_i σε μία μεταβλητή (έστω **z**) και εφαρμόζουμε τον έλεγχο Wilcoxon με τον ήδη γνωστό τρόπο. Τα αποτελέσματα του ελέγχου είναι τα εξής:

```
Test of median = 0.000000 versus median  >  0.000000

                N for   Wilcoxon           Estimated
            N   Test  Statistic        P      Median
z          10      9       44.0    0.006      0.8000
```

Το παρατηρούμενο επίπεδο σημαντικότητας είναι 0.006, μικρότερο του α και συνεπώς η μηδενική υπόθεση δεν είναι εύλογη σε επίπεδο σημαντικότητας 5%.

Λύση με το SPSS: Καταχωρίζουμε τις διαφορές Y_i-X_i σε μία μεταβλητή (έστω **d**) και, σε μία άλλη (έστω **m**), την τιμή της διαμέσου κάτω από την μηδενική υπόθεση (δηλαδή, την τιμή 0). Εφαρμόζοντας τον έλεγχο Wilcoxon κατά τα ήδη γνωστά, προκύπτουν τα εξής αποτελέσματα:

Ranks

		N	Mean Rank	Sum of Ranks
M - D	Negative Ranks	8[a]	5.50	44.00
	Positive Ranks	1[b]	1.00	1.00
	Ties	1[c]		
	Total	10		

a M < D
b M > D
c D = M

Test Statistics[b]

	M - D
Z	-2.549[a]
Asymp. Sig. (2-tailed)	.011
Exact Sig. (2-tailed)	.008
Exact Sig. (1-tailed)	.004
Point Probability	.002

a Based on positive ranks.
b Wilcoxon Signed Ranks Test

Οπως έχει ήδη παρατηρηθεί, το SPSS, βασίζει τον έλεγχο σε εκείνο το ζεύγος υποθέσεων του οποίου η εναλλακτική υπόθεση υποστηρίζεται από το μικρότερο παρατηρούμενο άθροισμα τάξεων μεγέθους. Στην περίπτωση του παραδείγματός μας, αυτό είναι το άθροισμα T^{+}. Μικρές τιμές του αθροίσματος αυτού αποτελούν ένδειξη υπέρ της υπόθεσης ότι ο εμβολιασμός προκαλεί αύξηση της θερμοκρασίας του σώματος. Αυτή είναι ακριβώς η εναλλακτική υπόθεση που μας ενδιαφέρει στο συγκεκριμένο πρόβλημα. Το παρατηρούμενο επίπεδο σημαντικότητας είναι 0.0055 (προκύπτει με διαίρεση του ασυμπτωτικού κρίσιμου επιπέδου Asymp. Sig. (2-tailed), που αναλογεί σε αμφίπλευρη εναλλακτική, δια του 2). Επομένως, η μηδενική υπόθεση ότι ο εμβολιασμός δεν προκαλεί σημαντική αύξηση της θερμοκρασίας του σώματος δεν μπορεί να θεωρηθεί εύλογη.

Λύση με το SAS: Χρησιμοποιούμε τις εντολές

```
data flu;
input x y @@;
diff=(y-x);
code=1;
if diff>0 then code=2;
```

```
absdiff=abs(diff);
if diff=0 then absdiff=.;
cards;
37.0 38.0 37.0 37.2 36.4 37.3 36.7 38.6 37.0 37.8
36.9 36.9 37.0 36.9 37.0 39.6 36.8 37.6 37.0 37.5
;
proc rank ties=mean out=ranks;
ranks r;
var absdiff;
run;
proc print data=ranks;
run;
data ranks;
set ranks;
r2=r;
if code=1 then r2=-r;
rr=r2**2;
proc means sum;
var r2 rr;
output out=sums sum=sumr2 sumrr;
run;
proc print;
run;
data sums;
set sums;
t=sumr2/(sumrr**(1/2));
pvalue=(1-cdf('normal',t));
proc print;
run;
```

Η εκτέλεση των παραπάνω εντολών μας δίνει τα εξής αποτελέσματα.

OBS	X	Y	DIFF	CODE	ABSDIFF	R
1	37.0	38.0	1.0	2	1.0	7
2	37.0	37.2	0.2	2	0.2	2
3	36.4	37.3	0.9	2	0.9	6
4	36.7	38.6	1.9	2	1.9	8
5	37.0	37.8	0.8	2	0.8	4
6	36.9	36.9	0.0	1	.	.
7	37.0	36.9	-0.1	1	0.1	1
8	37.0	39.6	2.6	2	2.6	9
9	36.8	37.6	0.8	2	0.8	5
10	37.0	37.5	0.5	2	0.5	3

Variable	Sum
R2	43.0000000
RR	285.0000000

OBS	_TYPE_	_FREQ_	SUMR2	SUMRR	T	PVALUE
1	0	10	43	285	2.54710	.0054311

Το παρατηρούμενο επίπεδο στατιστικής σημαντικότητας είναι μικρότερο του 0.006 και, επομένως, η υπόθεση ότι το εμβόλιο δεν προκαλεί στατιστικά σημαντική αύξηση της θερμοκρασίας του σώματος μία ώρα μετά τον εμβολιασμό δεν μπορεί να θεωρηθεί εύλογη.

3.1.3 Διάστημα Εμπιστοσύνης για την Παράμετρο Θέσης (Μέση τιμή ή Διάμεσο) Ενός Πληθυσμού ή την Διαφορά των Παραμέτρων Θέσης Δύο Πληθυσμών με Βάση Δείγμα Ζευγών Παρατηρήσεων

Διαστήματα εμπιστοσύνης για την θέση ενός πληθυσμού ή και την διαφορά στις θέσεις δύο πληθυσμών με βάση διαφορές ζευγών παρατηρήσεων αποτελούν μία εναλλακτική αντιμετώπιση του ελέγχου υποθέσεων με βάση τον έλεγχο των προσημασμένων τάξεων μεγέθους του Wilcoxon.

Τα δεδομένα αποτελούνται από n παρατηρήσεις της μορφής X_1, X_2, …, X_n πάνω στην μεταβλητή X (στην περίπτωση ενός πληθυσμού) ή της μορφής (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n) πάνω στην μεταβλητή (X, Y), στην περίπτωση δύο πληθυσμών. Θεωρούμε τις μεταβλητές

$$D_i = \begin{cases} X_i, & \text{στην περίπτωση ενός πληθυσμού} \\ Y_i - X_i, & \text{στην περίπτωση δύο πληθυσμών,} \end{cases}$$

τις οποίες διατάσουμε κατ' αύξουσα σειρά μεγέθους, λαμβάνοντας με αυτόν τον τρόπο το εξής διατεταγμένο δείγμα παρατηρήσεων:

$$D^{(1)} \leq D^{(2)} \leq \ldots \leq D^{(n-1)} \leq D^{(n)}.$$

Η κλίμακα μέτρησης των παρατηρήσεων είναι τουλάχιστον κλίμακα διαστήματος.

Με βάση το παραπάνω διατεταγμένο δείγμα παρατηρήσεων, επιθυμούμε να κατασκευάσουμε ένα διάστημα εμπιστοσύνης για την παράμετρο θέσης (μέση τιμή ή διάμεσο) του πληθυσμού

$$D = \begin{cases} X, & \text{στην περίπτωση ενός πληθυσμού} \\ Y - X, & \text{στην περίπτωση δύο πληθυσμών,} \end{cases}$$

ο οποίος είναι συμμετρικός. Η βασική ιδέα είναι να θεωρήσουμε κάθε δυνατό ζεύγος παρατηρήσεων (D_i, D_j) και να χρησιμοποιήσουμε τον μέσο τους $(D_i+D_j)/2$ ως εκτίμηση της παραμέτρου θέσης του πληθυσμού που περιγράφεται από την μεταβλητή D. Επειδή κάθε τέτοια εκτίμηση έχει την ίδια πιθανότητα να βρίσκεται πάνω ή κάτω από την παράμετρο θέσης του πληθυσμού και, επίσης, έχει την ίδια πιθανότητα να συνδέεται με οποιαδήποτε από δύο συμμετρικά διαστήματα ως προς την παράμετρο θέσης του πληθυσμού, οι εκτιμήσεις αυτές αποτελούν εύλογες επιλογές ως εκτιμήσεις της θέσης του πληθυσμού. Περιμένουμε, βέβαια, ότι μερικές από αυτές θα υπερεκτιμούν και άλλες θα υποεκτιμούν την παράμετρο αυτή και, επομένως, είναι διαισθητικά εύλογο να απορρίψουμε μερικές από τις πολύ μικρές τιμές και μερικές από τις πολύ μεγάλες. Ο αριθμός των εκτιμήσεων που θα απορρίψουμε προσδιορίζεται με βάση τον βαθμό εμπιστοσύνης που θέλουμε να έχουμε στο υπό κατασκευή διάστημα τιμών. Πάνω στην λογική αυτή στηρίζεται η κατασκευή του διαστήματος εμπιστοσύνης.

Το διάστημα εμπιστοσύνης για την διάμεσο (αντίστοιχα, την μέση τιμή) του πληθυσμού που περιγράφεται από την μεταβλητή D υπολογίζεται, όπως αναφέρθηκε παραπάνω, χρησιμοποιώντας την ελεγχοσυνάρτηση που Wilcoxon για τον έλεγχο της υπόθεσης H_0:

$d_{.50}$ = m (αντίστοιχα, H_0: μ_D = m), για διάφορες τιμές της σταθεράς m. Η υπόθεση αυτή μπορεί να ελεγχθεί με βάση το δείγμα των D_i αν από κάθε παρατήρηση D_i αφαιρέσουμε την τιμή της σταθεράς m και ελέγξουμε την υπόθεση H_0: $d'_{0.50} = 0$ (αντίστοιχα, H_0: $\mu_{D'} = 0$) για τον πληθυσμό που περιγράφεται από την μεταβλητή D΄=D-m. Αντί, όμως, να αφαιρούμε την τιμή m από κάθε D_i, να διατάσσουμε κατά αύξουσα σειρά μεγέθους τις προκύπτουσες παρατηρήσεις και να υπολογίζουμε την τιμή της στατιστικής συνάρτησης T^+, για κάθε δυνατή τιμή του m, είναι ευκολότερο να εργασθούμε ως εξής:

Θεωρούμε τους n(n+1)/2 δυνατούς μέσους της μορφής $(D_i + D_j)/2$, για όλους τους δείκτες i και j, συμπεριλαμβανομένων των περιπτώσεων i=j. Στην συνέχεια, χρειάζεται να προσδιορίσουμε πόσοι από αυτούς τους μέσους υπερβαίνουν την τιμή m ώστε να προσδιορίσουμε την τιμή της στατιστικής συνάρτησης T^+. Εργαζόμενοι με αντίστροφη σειρά, ξεκινάμε με την κρίσιμη τιμή $w_{\alpha/2}$ της στατιστικής συνάρτησης T^+ και προσδιορίζουμε τους $k=w_{\alpha/2}$ μεγαλύτερους και τους $k=w_{\alpha/2}$ μικρότερους τέτοιους μέσους. (Δεν χρειάζεται να υπολογισθούν όλοι οι N = n(n+1)/2 μέσοι, αλλά μόνο εκείνοι που βρίσκονται πλησιέστερα στον μεγαλύτερο και στον μικρότερο από αυτούς). Το ζητούμενο διάστημα εμπιστοσύνης είναι το διάστημα με άκρα

$$[D^{(k)}, D^{(N-k+1)}].$$

Δηλαδή, το διάστημα με άκρα τον $k=w_{\alpha/2}$ μικρότερο και τον $k=w_{\alpha/2}$ μεγαλύτερο μέσο είναι το ζητούμενο διάστημα εμπιστοσύνης, με συντελεστή εμπιστοσύνης $\geq 1-\alpha$.

Παράδειγμα 3.1.6: Μεταξύ των μέτρων που λαμβάνονται για την πρόληψη αεροπορικών ατυχημάτων, περιλαμβάνεται και η τοποθέτηση προειδοποιητικών μηχανισμών στα αεροπλάνα, οι οποίοι δίνουν ένδειξη στον πιλότο για τον κίνδυνο που διατρέχει όταν η ταχύτητα ενός αεροπλάνου υπερβαίνει κατά περίπου 5 κόμβους την *ταχύτητα απώλειας στήριξης*. 7 τέτοιοι μηχανισμοί επελέγησαν με τυχαίο τρόπο από μία μεγάλη γραμμή παραγωγής και ελέγχθηκαν με την βοήθεια μίας μηχανής που προσομοιώνει συνθήκες χειρισμού αεροπλάνου. Οι ταχύτητες πάνω από την προβλεπόμενη ταχύτητα απώλειας στήριξης στις οποίες οι προειδοποιητικοί μηχανισμοί ενεργοποιήθηκαν (σε κόμβους) κατεγράφησαν με τα εξής αποτελέσματα:

4.8 5.3 6.1 4.2 4.9 5.9 5.1

Να κατασκευασθεί ένα διάστημα τιμών με συντελεστή εμπιστοσύνης 95% για την μέση υπέρβαση του προβλεπόμενου ορίου ταχύτητας στην οποία οι μηχανισμοί της συγκεκριμένης γραμμής παραγωγής ενεργοποιούνται.

Λύση: Από τον πίνακα 8 των ποσοστιαίων σημείων της κατανομής της στατιστικής συνάρτησης T^+, παρατηρούμε ότι, για n=7, $w_{0.025}=3$. Επομένως, χρειάζεται να προσδιορίσουμε τις τρεις μικρότερες και τρεις μεγαλύτερες τιμές των N=7(7+1)/2=28 δυνατών μέσων της μορφής $(D_i+D_j)/2$ που μπορούμε να σχηματίσουμε από τα παραπάνω δεδομένα. Για τον σκοπό αυτό, διατάσσουμε κατ' αύξουσα σειρά μεγέθους τις τιμές του δοθέντος δείγματος:

4.2 4.8 4.9 5.1 5.3 5.9 6.1

Το παραπάνω διατεταγμένο δείγμα, αποτελεί τις τιμές των παρατηρήσεων $D^{(1)}$, $D^{(2)}$, …, $D^{(n)}$. Με την βοήθεια του πίνακα που ακολουθεί, υπολογίζουμε μερικούς από τους μέσους της μορφής $(D_i+D_j)/2$ που είναι πλησιέστερα στην μικρότερη τιμή του δείγματος (η οποία και αποτελεί τον μικρότερο τέτοιο μέσο), όπως, επίσης, μερικούς μέσους οι οποίοι είναι πλησιέστερα στην μεγαλύτερη τιμή του δείγματος (η οποία και αποτελεί τον μεγαλύτερο τέτοιο μέσο).

Πίνακας δυνατών μέσων της μορφής $(D_i+D_j)/2$

	4.2	**4.8**	**4.9**	**5.1**	**5.3**	**5.9**	**6.1**
4.2	4.2	4.5	4.55	4.65			
4.8		4.8	4.85	4.95			
4.9			4.9				
5.1				5.1	5.2	5.5	5.6
5.3					5.3	5.6	5.7
5.9						5.9	6
6.1							6.1

Οπως παρατηρήθηκε προηγουμένως, ο μικρότερος τέτοιος μέσος αντιστοιχεί στην πιο μικρή παρατήρηση, είναι δηλαδή η τιμή 4.2. Ο επόμενος μικρότερος μέσος δίνεται από τον μέσο των τιμών 4.2 και 4.8, δηλαδή $\frac{1}{2}(4.2+4.8)=4.5$. Η επόμενη μικρότερη τιμή είναι $\frac{1}{2}(4.2+4.9)=4.55$. Είναι σαφές ότι οι υπόλοιπες τιμές των μέσων είναι μεγαλύτερες. Επομένως, το κατώτερο όριο του ζητούμενου διαστήματος εμπιστοσύνης είναι η τιμή 4.55. Με παρόμοιο τρόπο, οι τρεις μεγαλύτερες τιμές των μέσων προσδιορίζονται ίσες με 6.1, $\frac{1}{2}(5.9+6.1)=6.0$ και 5.9. Επομένως, το ανώτερο όριο του ζητούμενου διαστήματος εμπιστοσύνης είναι η τιμή 5.9. Η μέση, λοιπόν, υπέρβαση του ορίου ταχύτητας στην οποία οι μηχανισμοί της συγκεκριμένης γραμμής παραγωγής ενεργοποιούνται εκτιμήθηκε με το διάστημα [4.55, 5.9] με συντελεστή εμπιστοσύνης τουλάχιστον ίσο με 0.95.

3.1.4 Γραφική Μέθοδος Κατασκευής Διαστήματος Εμπιστοσύνης

Στην βιβλιογραφία, έχουν αναπτυχθεί αρκετές γραφικές μέθοδοι προσδιορισμού διαστημάτων εμπιστοσύνης τύπου Wilcoxon. Ενδεικτικά, δίνεται στα επόμενα μία συνοπτική περιγραφή μιας τέτοιας μεθόδου στο πλαίσιο ενός παραδείγματος.

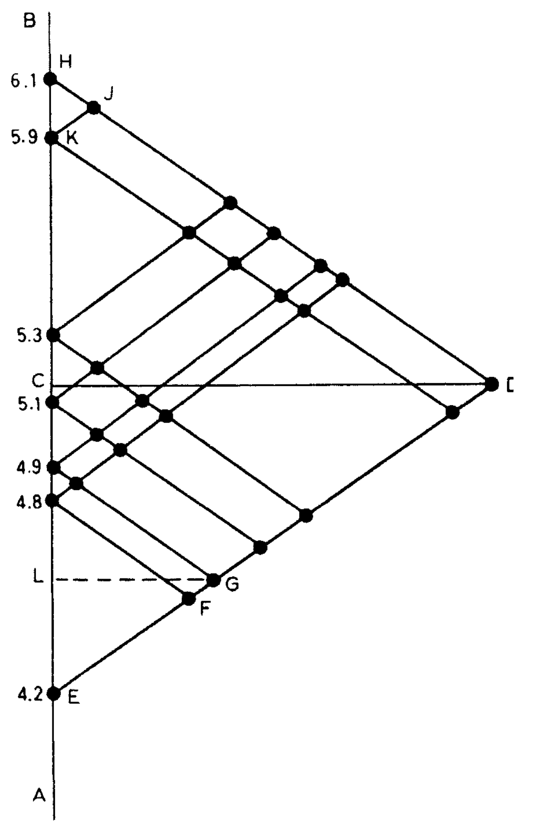
Παράδειγμα 3.1.7: Ας θεωρήσουμε τα δεδομένα του παραδείγματος της προηγούμενης ενότητας διατεταγμένα, για ευκολία, κατ' αύξουσα σειρά μεγέθους:

4.2 4.8 4.9 5.1 5.3 5.9 6.1

Να κατασκευασθεί ένα διάστημα με συντελεστή εμπιστοσύνης 95% για την μέση υπέρβαση του ορίου ταχύτητας από τους προειδοποιητικούς μηχανισμούς της συγκεκριμένης γραμμής παραγωγής.

Λύση: Η περιγραφή της μεθόδου γίνεται ευχερέστερη με την βοήθεια του σχήματος 3.1.4.

Σύρουμε μία κατακόρυφη ευθεία AB, πάνω στην οποία σημειώνουμε, σε κατάλληλη κλίμακα, σημεία που αντιστοιχούν στις τιμές των παρατηρήσεων του δείγματος. Το σημείο C συμβολίζει το μέσο του διαστήματος που ορίζουν η μικρότερη παρατήρηση (4.2) και η μεγαλύτερη παρατήρηση (6.1), οι οποίες συμβολίζονται αντίστοιχα με τα γράμματα E και H στο σχήμα 3.1.4. Στο σημείο C φέρουμε κάθετο στην ευθεία AB και διαλέγουμε ένα τυχαίο σημείο D πάνω σ' αυτήν. Φέρουμε τα ευθύγραμμα τμήματα DE και DH. Από κάθε ένα από τα σημεία που αντιστοιχούν στις παρατηρήσεις του δείγματος πάνω στην ευθεία ΑΒ, φέρουμε παράλληλες προς τα ευθύγραμμα τμήμα DH και DE.



Σχήμα 3.1.4

Γραφική μέθοδος κατασκευής διαστήματος εμπιστοσύνης

Αν από τις τομές αυτών των παραλλήλων φέρουμε καθέτους προς την ευθεία ΑΒ, τα σημεία τομής (βλέπε π.χ. το σημείο L της τομής της καθέτου που φέρεται από το σημείο G προς την ευθεία ΑΒ) μαζί με τα σημεία που αντιστοιχούν στις τιμές του δείγματος, παρέχουν όλους τους δυνατούς μέσους της μορφής $(D_i+D_j)/2$, για κάθε i και j, συμπεριλαμβανομένης της περίπτωσης i=j. Όπως είδαμε στο παράδειγμα 3.1.6, ο αριθμός k των μικροτέρων τέτοιων μέσων και των μεγαλυτέρων τέτοιων μέσων που απορρίπτουμε, είναι $k=w_{0.025}=3$ για το μέγεθος του δείγματός μας (n=7). Επομένως, τα όρια εμπιστοσύνης θα προσδιορισθούν με την βοήθεια καθέτων που θα φέρουμε προς την ευθεία AB από τα σημεία τομής των παραλλήλων που έχουμε γράψει που αντιστοιχούν στα σημεία G (για το κάτω όριο εμπιστοσύνης) και K (για το πάνω όριο εμπιστοσύνης). (Υπενθυμίζεται ότι οι τιμές των δεδομένων είναι και αυτές δυνατές τιμές των μέσων της μορφής $(D_i+D_j)/2$). Το ζητούμενο διάστημα τιμών, επομένως, είναι το διάστημα [4.55, 5.9] με συντελεστή εμπιστοσύνης ≥ 0.95.

3.2 ΠΕΡΙΠΤΩΣΗ ΑΝΕΞΑΡΤΗΤΩΝ ΔΕΙΓΜΑΤΩΝ - - Ο ΕΛΕΓΧΟΣ ΤΩΝ MANN - WHITNEY Ή ΤΟΥ WILCOXON

Ο έλεγχος που παρουσιάζεται στην ενότητα αυτή είναι γνωστός ως έλεγχος των Mann-Whitney και επίσης ως έλεγχος του Wilcoxon. Μελετήθηκε για πρώτη φορά, για ισομεγέθη δείγματα, από τον Wilcoxon το 1945. Η περίπτωση δειγμάτων διαφορετικού μεγέθους μελετήθηκε από τους Mann και Whitney το 1947 και από τον Wilcoxon το 1949. Αυτό εξηγεί το ότι ο έλεγχος αυτός εμφανίζεται στην βιβλιογραφία με διαφορετικά ονόματα. Για τους ίδιους λόγους, ο έλεγχος αυτός συχνά αναφέρεται και ως έλεγχος Wilcoxon-Mann-Whitney. Άλλες ισοδύναμες μορφές του ιδίου ελέγχου έχουν επίσης εμφανισθεί στην βιβλιογραφία κάτω από διάφορα άλλα ονόματα, ενδεχομένως λόγω της διαισθητικής λογικής στην οποία στηρίζεται. Αν και κατά κύριο λόγο είναι έλεγχος που χρησιμοποιείται στην περίπτωση που έχουμε δύο ανεξάρτητα δείγματα, ο έλεγχος των Mann-Whitney μπορεί να εφαρμοσθεί σε πολλές διαφορετικές περιπτώσεις.

Η συνήθης περίπτωση των δύο δειγμάτων είναι αυτή που ο ερευνητής έχει δύο ανεξάρτητα δείγματα τα οποία προέρχονται από δύο πληθυσμούς, ενδεχομένως διαφορετικούς, και επιθυμεί να χρησιμοποιήσει έναν στατιστικό έλεγχο για να εξετάσει αν η μηδενική υπόθεση ότι οι δύο πληθυσμοί ταυτίζονται μπορεί να απορριφθεί. Δηλαδή, ο ερευνητής επιθυμεί να ανιχνεύσει διαφορές μεταξύ των δύο πληθυσμών με βάση τυχαία δείγματα από αυτούς τους πληθυσμούς. Εάν τα δείγματα αποτελούνται από δεδομένα σε διατεταγμένη κλίμακα, η πιο ενδιαφέρουσα διαφορά για έναν ερευνητή θα ήταν μία διαφορά στην θέση των δύο πληθυσμών. Τείνουν οι τιμές του ενός πληθυσμού να είναι μεγαλύτερες από τις τιμές του άλλου πληθυσμού; Εχουν οι δύο πληθυσμοί ίσες διαμέσους; Είναι ίσες οι μέσες τιμές των πληθυσμών;

Μία διαισθητική μέθοδος αντιμετώπισης του προβλήματος των δύο δειγμάτων στηρίζεται στην συνένωση των δύο δειγμάτων σε ένα ενιαίο δείγμα, του οποίου οι τιμές διατάσσονται κατ' αύξουσα σειρά μεγέθους. Στις τιμές του προκύπτοντος δείγματος αντιστοιχίζονται βαθμοί από την μικρότερη στη μεγαλύτερη, ανεξάρτητα από τον πληθυσμό από τον οποίο προέρχεται κάθε μία τιμή. Τότε, η στατιστική συνάρτηση που θα μπορούσε να χρησιμοποιηθεί ως ελεγχοσυνάρτηση, θα μπορούσε να ορισθεί ως το άθροισμα των βαθμών που αντιστοιχούν σε εκείνες τις τιμές που προέρχονται από τον ένα από τους δύο πληθυσμούς. Είναι προφανές, ότι εάν το άθροισμα αυτό είναι πολύ μικρό (ή πολύ μεγάλο), αυτό θα αποτελεί ένδειξη ότι οι τιμές που προέρχονται από αυτόν τον πληθυσμό τείνουν να είναι μικρότερες (ή μεγαλύτερες) από τις τιμές που προέρχονται από τον άλλο πληθυσμό. Επομένως, η μηδενική υπόθεση ότι δεν υπάρχουν διαφορές μεταξύ των πληθυσμών είναι εύλογο να απορριφθεί, όταν οι βαθμοί που σχετίζονται με το ένα δείγμα τείνουν να είναι υψηλότεροι από τους βαθμούς που σχετίζονται με το άλλο δείγμα.

Είναι ενδιαφέρον να σημειωθεί ότι η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου Mann-Whitney δεν είναι ποτέ εξαιρετικά χαμηλή σε σύγκριση με τον έλεγχο t για δύο δείγματα, ο οποίος είναι το σύνηθες παραμετρικό του ανάλογο. (Συγκεκριμένα, η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου Mann-Whitney συγκρινόμενου με τον έλεγχο t έχει μελετηθεί κάτω από την υπόθεση ότι οι δύο πληθυσμοί διαφέρουν μόνο ως προς την μέση τιμή τους. Στην περίπτωση κανονικών πληθυσμών, η ασυμπτωτική σχετική αποτελεσματικότητα είναι 0.955, στην περίπτωση ομοιόμορφων πληθυσμών είναι 1 και στην περίπτωση που οι πληθυσμοί έχουν εκθετικές κατανομές, είναι 1.5. Αν οι δύο πληθυσμοί διαφέρουν μόνο ως προς την παράμετρο θέσης, η ασυμπτωτική σχετική αποτελεσματικότητα δεν είναι ποτέ μικρότερη από 0.864). Επιπλέον, το αντίθετο δεν είναι πάντα αληθές. Η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου t όταν συγκρίνεται με τον έλεγχο Mann-Whitney, ενδέχεται να είναι πολύ χαμηλή και, συχνά, εξαιρετικά χαμηλή (σχεδόν ίση με το μηδέν), ιδιαίτερα στην περίπτωση που το ένα ή και τα δύο δείγματα περιέχουν ακραίες τιμές. Επομένως, ο έλεγχος Mann-Whitney, ο οποίος περιγράφεται στα επόμενα, είναι ένας ασφαλέστερος έλεγχος.

Τα δεδομένα αποτελούνται από δύο αμοιβαία ανεξάρτητα τυχαία δείγματα παρατηρήσεων, των οποίων η κλίμακα μέτρησης είναι τουλάχιστον κλίμακα διάταξης (ordinal scale). Εστω ότι η ακολουθία τυχαίων μεταβλητών $X_1, X_2, ..., X_n$ συμβολίζει το τυχαίο δείγμα μεγέθους n από τον πληθυσμό 1 και έστω ότι η ακολουθία τυχαίων μεταβλητών Y_1, Y_2, ..., Y_m συμβολίζει το τυχαίο δείγμα μεγέθους m από τον πληθυσμό 2. Θεωρούμε το δείγμα μεγέθους n+m που προκύπτει από την συνένωση των δυο δειγμάτων και αντιστοιχίζουμε στις τιμές του τους βαθμούς (τάξεις μεγέθους) 1 έως n+m. Εστω ότι $R(X_i)$ και $R(Y_i)$ οι βαθμοί που αντιστοιχούν στις μεταβλητές X_i και Y_i για κάθε i και j. Για ευκολία στον συμβολισμό, θέτουμε $N = n + m$. (Αν αρκετές από τις τιμές του δείγματος ταυτίζονται, αντιστοιχίζουμε σε κάθε μία από αυτές τον μέσο των βαθμών που θα είχαν αν δεν ταυτίζονταν).

Έστω $F(\cdot)$ και $G(\cdot)$ οι συναρτήσεις κατανομής των τυχαίων μεταβλητών X και Y, αντίστοιχα, οι οποίες περιγράφουν τους πληθυσμούς 1 και 2 αντίστοιχα. Τότε, οι υποθέσεις που θα μπορούσε να ελέγξει κανείς, προκειμένου να ανιχνεύσει διαφορές μεταξύ των δύο πληθυσμών, μπορούν να διατυπωθούν ως εξής:

Α. (Αμφίπλευρος έλεγχος):

H_0: $F(x) = G(x)$, για κάθε x

H_1: $F(x) \neq G(x)$, για κάποιο x.

Β. (Μονόπλευρος έλεγχος):

H_0: $F(x) \leq G(x)$, για κάθε x

H_1: $F(x) > G(x)$, για κάποιο x.

Γ. (Μονόπλευρος έλεγχος):

H_0: $F(x) \geq G(x)$, για κάθε x

H_1: $F(x) < G(x)$, για κάποιο x.

Η μορφή της εναλλακτικής υπόθεσης στην περίπτωση Α συνεπάγεται ότι

$$P(X \leq x) \neq P(Y \leq x), \text{ για τουλάχιστον ένα } x,$$

ή, ισοδύναμα,

$$P(X > x) \neq P(Y > x), \text{ για τουλάχιστον ένα } x.$$

Δηλαδή, η πιθανότητα με την οποία οι τιμές του πληθυσμού 1 υπερβαίνουν την τιμή x δεν είναι ίση με την πιθανότητα με την οποία οι τιμές του πληθυσμού 2 υπερβαίνουν την τιμή x, για ένα τουλάχιστον x. Επομένως, η πιθανότητα με την οποία οι τιμές του πληθυσμού 1 υπερβαίνουν την τιμή x είναι μεγαλύτερη ή μικρότερη από την πιθανότητα με την οποία οι τιμές του πληθυσμού 2 υπερβαίνουν την τιμή x. Με άλλα λόγια, οι τιμές του πληθυσμού 1 τείνουν να είναι μεγαλύτερες από τις τιμές του πληθυσμού 2. Κατά συνέπεια, η πιθανότητα $P(X>Y)$ δεν είναι ίση με την πιθανότητα $P(X<Y)$, όπως υπαινίσσεται η μηδενική υπόθεση. Δηλαδή, $P(X<Y) \neq 1/2$. Για τον λόγο αυτό, συχνά οι παραπάνω υποθέσεις διατυπώνονται με την εξής μορφή:

Α. (Αμφίπλευρος έλεγχος):

H_0: $P(X<Y) = 1/2$

H_1: $P(X<Y) \neq 1/2$.

Με βάση την ίδια επιχειρηματολογία, οι υποθέσεις των περιπτώσεων Β και Γ διατυπώνωνται συχνά με την μορφή:

Β. (Μονόπλευρος έλεγχος):

H_0: $P(X<Y) \leq 1/2$

H_1: $P(X<Y) > 1/2$.

Γ. (Μονόπλευρος έλεγχος):

H_0: $P(X<Y) \geq 1/2$

H_1: $P(X<Y) < 1/2$.

Ο έλεγχος των Mann και Whitney είναι αμερόληπτος και συνεπής.

Ως στατιστική συνάρτηση για τον έλεγχο των παραπάνω υποθέσεων συνήθως χρησιμοποιείται το άθροισμα των βαθμών (των τάξεων μεγέθους) που οι παρατηρήσεις X_1, X_2, ..., X_n έχουν στο συνενωμένο δείγμα των n+m παρατηρήσεων, δηλαδή

$$T = \sum_{i=1}^{n} R(X_i).$$

Η στατιστική αυτή συνάρτηση χρησιμοποιείται όταν δεν υπάρχουν περιπτώσεις ταύτισης τιμών στο δείγμα, ή όταν υπάρχουν μόνο λίγες τέτοιες περιπτώσεις. Αν υπάρχουν πολλές περιπτώσεις ταύτισης τιμών, αντί της στατιστικής συνάρτησης T, χρησιμοποιείται η τυποποιημένη μορφή της:

$$T_1 = \frac{T - E(T)}{\sqrt{V(T)}}$$

ή, ισοδύναμα,

$$T_1 = \frac{T - \dfrac{n(N+1)}{2}}{\sqrt{\dfrac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \dfrac{nm(N+1)^2}{4(N-1)}}} \quad ,$$

όπου το άθροισμα $\sum_{i=1}^{N} R_i^2$ αναφέρεται στο άθροισμα των τετραγώνων των N μέσων βαθμών που στην πραγματικότητα χρησιμοποιούνται στο συνενωμένο δείγμα.

Παρατήρηση: Το γεγονός ότι η στατιστική συνάρτηση T_1 είναι η τυποποιημένη μορφή της στατιστικής συνάρτησης T είναι άμεση συνέπεια του εξής θεωρήματος:

Αν S είναι το άθροισμα n ακεραίων, οι οποίοι έχουν επιλεγεί τυχαία και χωρίς επανάθεση από τους πρώτους N ακεραίους, τότε η μέση τιμή και η διασπορά του S δίνονται από τις σχέσεις

$$E(S)=\frac{n(N+1)}{2} \qquad \text{και} \qquad V(S)=\frac{n(N+1)(N-n)}{12}$$

αντίστοιχα.

Τα ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T δίνονται στον πίνακα 9 του παραρτήματος. Ο πίνακας δίνει μόνο τα κάτω ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T (w_p , p = 0.001, 0.005, 0.01, 0.025, 0.05, 0.10). Τα πάνω ποσοστιαία σημεία της κατανομής μπορούν να υπολογισθούν από την σχέση

$$w_{1-p} = n(N+1) - w_p.$$

Τα ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T_1 προσεγγίζονται από τα αντίστοιχα ποσοστιαία σημεία της τυποποιημένης κανονικής κατανομής (πίνακας 2 του παραρτήματος).

Για τον έλεγχο των υποθέσεων της περίπτωσης Α, είναι προφανές ότι οι πολύ μεγάλες ή οι πολύ μικρές τιμές της στατιστικής συνάρτησης, ανεξάρτητα από την συγκεκριμένη μορφή που αυτή έχει (T ή T_1), θα συνηγορούν υπέρ της απόρριψης της μηδενικής υπόθεσης. Δηλαδή, ο κανόνας απόφασης στην περίπτωση αυτή έχει τη μορφή:

Η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας ίσο με α, αν η τιμή της στατιστικής συνάρτησης είναι μικρότερη από το α/2-ποσοστιαίο σημείο της κατανομής της ή αν είναι μεγαλύτερη από το (1–α/2)-ποσοστιαίο σημείο της κατανομής της.

Αν, δηλαδή, χρησιμοποιείται η στατιστική συνάρτηση T, τότε η μηδενική υπόθεση θα απορρίπτεται αν $T< w_{\alpha/2}$ ή αν $T > w_{1-\alpha/2}$. Με αντίστοιχο τρόπο ορίζεται η κρίσιμη περιοχή στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T_1 .

Στην περίπτωση του ελέγχου των υποθέσεων της μορφής Β, είναι προφανές ότι οι μικρές τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης θα αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α, αν η παρατηρούμενη τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής της. Δηλαδή, η μηδενική υπόθεση απορρίπτεται αν $T< w_{\alpha}$, στην περίπτωση που χρησιμοποιείται η Τ με αντίστοιχη διατύπωση του κανόνα απόφασης στην περίπτωση που η χρησιμοποιούμενη στατιστική συνάρτηση είναι η T_1.

Για τον έλεγχο των υποθέσεων της περίπτωσης Γ, είναι προφανές ότι οι μεγάλες τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης θα αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως ο κανόνας απόφασης στην περίπτωση αυτή διατυπώνεται ως εξής:

Η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α, αν η παρατηρούμενη τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης υπερβαίνει το (1–α)-ποσοστιαίο σημείο της κατανομής της. Επομένως, στην περίπτωση που χρησιμοποιείται η Τ, η μηδενική υπόθεση θα απορρίπτεται σε επίπεδο σημαντικότητας α αν $T > w_{1-\alpha}$. Αντίστοιχη είναι η μορφή της κρίσιμης περιοχής στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T_1.

Παρατήρηση: Συχνά, σε σχέση με τις πρακτικές εφαρμογές, εάν υπάρχει μία διαφορά μεταξύ των συναρτήσεων κατανομής δύο πληθυσμών, αυτή η διαφορά αναφέρεται στην θέση των δύο κατανομών. Δηλαδή, αν η F(x) δεν ταυτίζεται με την G(x), τότε η

F(x) ταυτίζεται με την G(x+c) όπου c είναι μία σταθερά. Στην περίπτωση αυτή, οι υποθέσεις μπορούν να διατυπωθούν ώστε να αναφέρονται στις μέσες τιμές των δύο πληθυσμών, εάν βέβαια αυτές υπάρχουν. Οι περιπτώσεις Α, Β και Γ γράφονται τώρα με την εξής μορφή:

Α. (Αμφίπλευρος έλεγχος):

H_0: $E(X) = E(Y)$

H_1: $E(X) \neq E(Y)$

Β. (Μονόπλευρος έλεγχος):

H_0: $E(X) \geq E(Y)$

H_1: $E(X) < E(Y)$

Γ. (Μονόπλευρος έλεγχος):

H_0: $E(X) \leq E(Y)$

H_1: $E(X) > E(Y)$.

Αντίστοιχες υποθέσεις μπορούν να διατυπωθούν για την περίπτωση που η διαφορά μεταξύ των συναρτήσεων κατανομής των δύο πληθυσμών αναφέρεται στις διαμέσους των δύο πληθυσμών.

Επομένως, ο έλεγχος Mann-Whitney μπορεί να χρησιμοποιηθεί και ως έλεγχος ύπαρξης διαφοράς στην θέση (μέση τιμή ή διάμεσο) δύο πληθυσμών.

Παράδειγμα 3.2.1: Στο γυμνάσιο μιας επαρχιακής πόλης, 12 από τους 48 μαθητές της πρώτης τάξης ζούσαν σε αγροικίες. Ο καθηγητής της Φυσικής Αγωγής αυτού του γυμνασίου, θέλοντας να εξετάσει κατά πόσον τα παιδιά που ζούσαν σε αγροικίες είχαν καλύτερη φυσική κατάσταση από τα παιδιά που ζούσαν στην πόλη, επενόησε ένα τεστ φυσικής κατάστασης στο οποίο και υπέβαλε τους 48 μαθητές. Τα αποτελέσματα του τεστ συνοψίζονται στον πίνακα που ακολουθεί. (Οι υψηλοί βαθμοί αντανακλούν πολύ καλή φυσική κατάσταση):

Αποτελέσματα του τεστ φυσικής κατάστασης

Παιδιά Αγροικιών		Παιδιά Πόλης					
14.8	10.6	12.7	16.9	7.6	2.4	6.2	9.9
7.3	12.5	14.2	7.9	11.3	6.4	6.1	10.6
5.6	12.9	12.6	16.0	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5.0
9.0	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1.0	3.2	5.9	4.0

Λύση: Είναι προφανές ότι καμία από τις δύο ομάδες των παιδιών δεν αποτελεί τυχαίο δείγμα από κάποιο πληθυσμό. Παρ' όλα αυτά, θα μπορούσε να υποθέσει κανείς, ότι οι βαθμοί που τα παιδιά πέτυχαν στο τεστ φυσικής κατάστασης προσομοιάζουν με υποθετικά τυχαία δείγματα από πληθυσμούς παιδιών αγροικιών και παιδιών πόλεων της ίδιας κατηγορίας ηλικίας, τουλάχιστον για παρόμοιες περιοχές. Ο τρόπος με τον οποίο ορίσθηκαν τα δύο δείγματα συνεπάγεται και την ανεξαρτησία τους. Επομένως η προϋπόθεση της ανεξαρτησίας των δύο δειγμάτων φαίνεται να ισχύει.

Οι προς έλεγχο υποθέσεις μπορούν να διατυπωθούν με την εξής μορφή:

H_0: Τα παιδιά των αγροικιών δεν τείνουν να έχουν καλύτερη φυσική κατάσταση από τα παιδιά της πόλης.

H_1: Τα παιδιά των αγροικιών τείνουν να έχουν καλύτερη φυσική κατάσταση από τα παιδιά της πόλης.

Εναλλακτικά, οι υποθέσεις αυτές μπορούν να γραφούν με τη μορφή:

$$H_0: P(X<Y) \geq 1/2$$

$$H_1: P(X<Y) < 1/2,$$

όπου X και Y συμβολίζουν την φυσική κατάσταση του πληθυσμού των παιδιών των αγροικιών και του πληθυσμού των παιδιών των πόλεων, αντίστοιχα.

Συνδυάζοντας τα δύο δείγματα σε ένα ενιαίο δείγμα και διατάσσοντας κατά αύξουσα σειρά μεγέθους τις παρατηρήσεις του προκύπτοντος δείγματος, καταλήγουμε στον πίνακα που ακολουθεί.

X	Y	Βαθμός	X	Y	Βαθμός	X	Y	Βαθμός
	1.0	1		6.2	17		11.3	33
	1.8	2	6.3		18	11.4		34
	2.1	3		6.4	19		11.8	35
	2.4	4		6.7	20.5	12.5		36
	2.6	5		6.7	20.5		12.6	37
2.7		6	7.3		22		12.7	38
	3.2	7		7.6	23	12.9		39
	3.6	8		7.9	24		14.2	40
	4.0	9		8.3	25		14.8	41.5
4.2		10	9.0		26	14.8		41.5
	5.0	11		9.1	27		15.3	43

	5.6	13		9.9	28		16.0	44
	5.6	13		10.6	30.5	16.1		45
5.6		13		10.6	30.5		16.9	46
	5.9	15	10.6		30.5		17.7	47
	6.1	16		10.6	30.5		18.6	48

Υπάρχουν τέσσερις περιπτώσεις ταύτισης τιμών, όπως φαίνεται από τις αγκύλες στις στήλες του πίνακα. Επομένως, η κατάλληλη στατιστική συνάρτηση για τον έλεγχο των υποθέσεων του προβλήματος είναι η T_1 . Ο έλεγχος είναι μονόπλευρος και η κρίσιμη περιοχή προφανώς αντιστοιχεί σε μεγάλες τιμές της στατιστικής συνάρτησης T_1. Από τον πίνακα της τυποποιημένης κανονικής κατανομής, παρατηρούμε ότι η κρίσιμη περιοχή μεγέθους α=0.05 αντιστοιχεί σε τιμές της στατιστικής συνάρτησης T_1 μεγαλύτερες από την τιμή $z_{0.95} = 1.645$. Δηλαδή, ο κανόνας απόφασης έχει τη μορφή:

Η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α=0.05, αν $T_1 > 1.645$.

Από τα δεδομένα του προβλήματος, έχουμε ότι n=12, m=36. Επομένως, N = 12+36 = 48. Επίσης,

$$
\begin{aligned}
T &= \sum_{i=1}^{n} R(X_i) \\
&= 6 + 10 + 13 + 18 + 22 + 26 + 30.5 + 34 + 36 + 39 + 41.5 + 45 \\
&= 321,
\end{aligned}
$$

ενώ,

$$\sum_{i=1}^{N} R_i^2 = 38016.$$

Κατά συνέπεια,

$$T_1 = \frac{T - \frac{n(N+1)}{2}}{\sqrt{\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}}}$$

$$= \frac{321 - 13(\frac{49}{2})}{\sqrt{\frac{(12)(36)}{(48)(47)}(38016) - \frac{(12)(36)(49)^2}{4(47)}}}$$

$$= 0.6431.$$

Η παρατηρηθείσα τιμή της στατιστικής συνάρτησης T_1, επομένως, δεν περιέχεται στην κρίσιμη περιοχή, και επομένως η μηδενική υπόθεση δεν απορρίπτεται σε επίπεδο σημαντικότητας 0.05. Κατά συνέπεια, σε επίπεδο σημαντικότητας 0.05, τα δεδομένα δεν παρέχουν ενδείξεις υπέρ της υπόθεσης ότι τα παιδιά των αγροικιών έχουν καλύτερη φυσική κατάσταση από τα παιδιά των πόλεων.

Από τον πίνακα της τυποποιημένης κανονικής κατανομής, παρατηρούμε ότι η τιμή της στατιστικής συνάρτησης T_1 είναι πολύ κοντά στην τιμή του 0.74-ποσοστιαίου σημείου της τυποποιημένης κανονικής κατανομής. Δηλαδή, το κρίσιμο επίπεδο $\hat{\alpha}$ είναι

$$\begin{aligned} \hat{\alpha} &= \max P(T_1 \geq 0.6431 \mid H_0) \\ &= P(Z \geq 0.6431) \\ &= 1 - P(Z < 0.6431) \\ &\cong 1 - 0.74 \\ &= 0.26. \end{aligned}$$

Επομένως, η μηδενική υπόθεση θα απορρίπτεται οποτεδήποτε το επίπεδο σημαντικότητας έχει μία τιμή α η οποία είναι μεγαλύτερη ή ίση από το κρίσιμο επίπεδο $(\alpha \geq \hat{\alpha})$.

Σημείωση: Οπως παρατηρήθηκε προηγουμένως, υπάρχουν τέσσερις περιπτώσεις ταύτισης τιμών. Θα πρέπει να σημειωθεί ότι, επειδή ο αριθμός των περιπτώσεων ταύτισης τιμών δεν είναι μεγάλος, η χρησιμοποίηση της στατιστικής συνάρτησης T στην θέση της T_1 ενδέχεται να είναι αποδεκτή. Αν, λοιπόν, αγνοήσουμε τις περιπτώσεις ταύτισης τιμών και χρησιμοποιήσουμε την στατιστική συνάρτηση T, τότε η κρίσιμη περιοχή του ελέγχου αντιστοιχεί στις τιμές της στατιστικής συνάρτησης T οι οποίες υπερβαίνουν το 0.95-ποσοστιαίο σημείο της, όπως αυτό προκύπτει από τον πίνακα 9 του παραρτήματος. Δηλαδή, η κρίσιμη περιοχή του ελέγχου θα αντιστοιχεί τώρα στις τιμές της στατιστικής συνάρτησης T οι οποίες υπερβαίνουν την τιμή

$$\begin{aligned} w_{0.95} &= n(N+1)/2 + z_{0.95}\sqrt{nm(N+1)/12} \\ &= 294 + (1.645)(42) \\ &= 363.1. \end{aligned}$$

Οπως όμως υπολογίσθηκε προηγουμένως, η παρατηρηθείσα τιμή της στατιστικής συνάρτησης T είναι ίση με 321. Επομένως, όπως και προηγουμένως, η μηδενική υπόθεση δεν απορρίπτεται σε επίπεδο σημαντικότητας 0.05.

Για τον προσδιορισμό του κρισίμου επιπέδου $\hat{\alpha}$, χρησιμοποιούμε και πάλι τον τύπο που δίνεται στο τέλος του πίνακα 9 για την προσέγγιση των ποσοστιαίων σημείων της κατανομής της στατιστικής συνάρτησης T, μόνο που τώρα θα τον χρησιμοποιήσουμε με την αντίστροφη έννοια: θα χρησιμοποιήσουμε την τιμή τ της στατιστικής συνάρτησης T στην θέση της τιμής $w_{0.95}$ και από την προκύπτουσα εξίσωση θα προσδιορίσουμε την τιμή $z_{1-\alpha}$, δηλαδή την τιμή του κατάλληλου ποσοστιαίου σημείου της τυποποιημένης κανονικής κατανομής, από την οποία και θα προκύψει η τιμή του κρισίμου επιπέδου $\hat{\alpha}$.

Συγκεκριμένα, από την εξίσωση

$$\tau = \frac{n(N+1)}{2} + z_{1-\alpha}\sqrt{\frac{nm(N+1)}{12}}$$

προκύπτει ότι

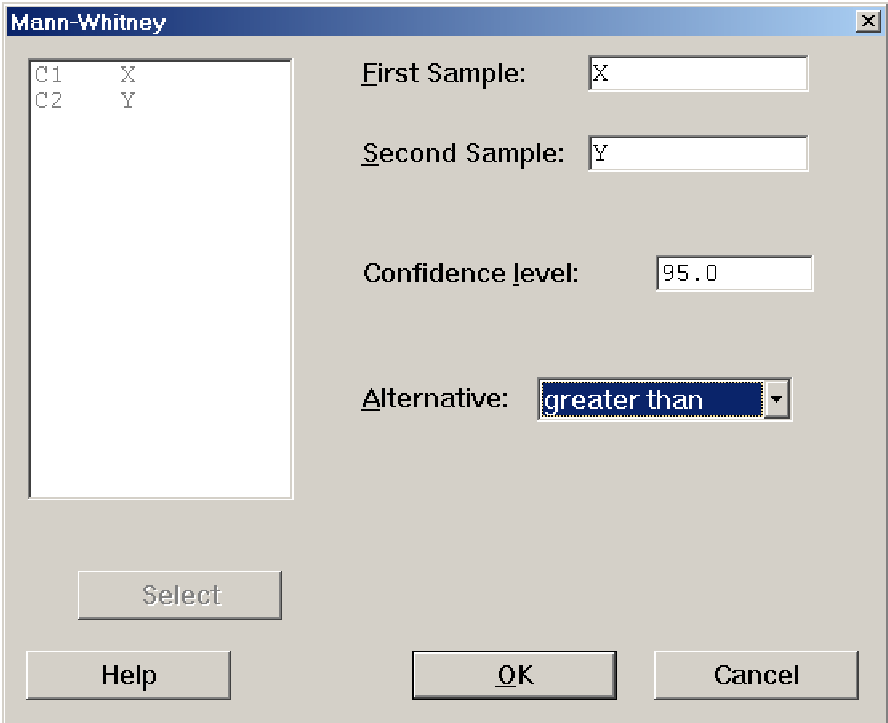
$$z_{1-\alpha} = \frac{\tau - n(N+1)/2}{\sqrt{nm(N+1)/12}}$$

$$= 0.6429.$$

Προφανώς η τιμή αυτή δεν διαφέρει πολύ από την αντίστοιχη τιμή της στατιστικής συνάρτησης T_1, η οποία, όπως παρατηρήσαμε, ταυτίζεται με το 0.74-ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής και, επομένως, το κρίσιμο επίπεδο και στην περίπτωση αυτή είναι 0.26. Δηλαδή, όπως και προηγουμένως, $\hat{\alpha} = 0.26$.

Λύση με το MINITAB: Για την διεξαγωγή του ελέγχου Wilcoxon-Mann-Whitney, το MINITAB χρησιμοποιεί ως στατιστική συνάρτηση ελέγχου το άθροισμα των τάξεων μεγέθους εκείνου του δείγματος που δηλώνουμε πρώτο στο κατάλληλο πλαίσιο διαλόγου. Ανάλογα με την μορφή της εναλλακτικής υπόθεσης, η μηδενική υπόθεση απορρίπτεται για πολύ μικρές ή πολύ μεγάλες τιμές του αθροίσματος αυτού.

Καταχωρίζουμε δύο δείγματα σε δύο μεταβλητές (έστω **X** και **Y**) και, στην συνέχεια, επιλέγουμε **Stat, Nonparametrics, Mann-Whitney**. Αυτό οδηγεί στο ακόλουθο πλαίσιο διαλόγου:



Στα πεδία **First Sample** και **Second Sample**, δηλώνουμε τις μεταβλητές που περιέχουν τα δύο δείγματα. (Η στατιστική συνάρτηση ελέγχου θα είναι το άθροισμα των τάξεων μεγέθους των παρατηρήσεων του δείγματος που δηλώθηκε στο **First Sample**). Στο πεδίο **Confidence level**, δηλώνουμε την τιμή του 1-α. (Αυτή δεν έχει ιδιαίτερη σημασία γιατί το πακέτο υπολογίζει πάντα το κρίσιμο επίπεδο (p-value)). Τέλος, στο πεδίο **Alternative**, επιλέγουμε **greater than**, **equal than** ή **not equal to** ανάλογα με την εναλλακτική υπόθεση παίρνοντας υπ' όψη μας ποια μεταβλητή καταχωρίσθηκε στο πεδίο **First Sample**.

Το MINITAB θεωρεί την μεταβλητή του πεδίου **First Sample** ως εκείνη που επιδεικνύει τιμές μεγαλύτερες (**greater than**), μικρότερες (**less than**) ή ίσες (**equal to**) με τις τιμές της μεταβλητής του πεδίου **Second Sample**. Εμείς εδώ, καταχωρίσαμε την μεταβλητή **X** στο πεδίο

First Sample και, στο συγκεκριμένο παράδειγμα, η εναλλακτική υπόθεση P(X<Y)<1/2 δηλώνει ότι η X τείνει να παίρνει τιμές μεγαλύτερες από την Y. Επομένως, πρέπει να επιλέξουμε **greater than**.

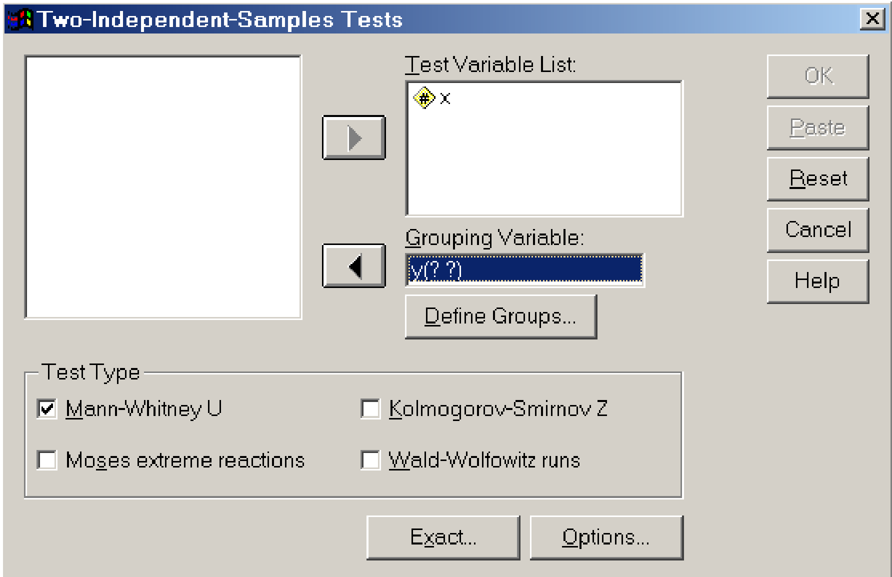
Τα αποτελέσματα του ελέγχου είναι:

Mann-Whitney Confidence Interval and Test

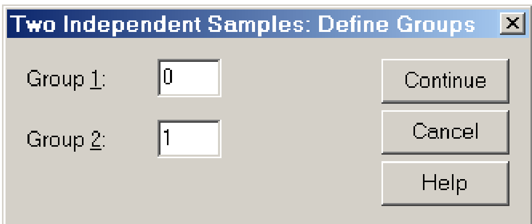
```
X          N =  12     Median =       9.800
Y          N =  36     Median =       7.750
Point estimate for ETA1-ETA2 is       0.800
95.1 Percent CI for ETA1-ETA2 is (-2.300,4.400)
W = 321.0
Test of ETA1 = ETA2  vs  ETA1 > ETA2 is significant at
0.2640
The test is significant at 0.2639 (adjusted for ties)

Cannot reject at alpha = 0.05
```

Η ελεγχοσυνάρτηση συμβολίζεται με **W**, ενώ το παρατηρούμενο επίπεδο σημαντικότητας δίνεται στην γραμμή **The test is significant at**. Για τα δεδομένα του παραδείγματος, το κρίσιμο επίπεδο είναι 0.2639 και, όπως σημειώνεται στην τελευταία γραμμή του παραπάνω πίνακα αποτελεσμάτων, η μηδενική υπόθεση δεν απορρίπτεται σε επίπεδο σημαντικότητας 0.05.

Λύση με το SPSS: Για να κάνουμε τον έλεγχο Wilcoxon-Mann-Whitney, καταχωρίζουμε και τα δύο δείγματα σε μία μεταβλητή, ενώ, σε μία άλλη μεταβλητή, δίνουμε κωδικές τιμές που υποδηλώνουν το δείγμα από το οποίο προέρχεται κάθε παρατήρηση. Έστω ότι καταχωρίζουμε τα δύο δείγματα στην μεταβλητή **x**. Δημιουργούμε την μεταβλητή **y** με τις τιμές 0 (αγροικία – πρώτο δείγμα) και 1 (πόλη – δεύτερο δείγμα). Για να διεξαγάγουμε τον έλεγχο, επιλέγουμε κατά σειρά **Analyze, Nonparametric Tests, 2 Independent Samples** και οδηγούμεθα στο παρακάτω πλαίσιο διαλόγου:



Στο πλαίσιο **Test Variable List**, εισάγουμε την μεταβλητή **x**. Στο πλαίσιο **Grouping Variable**, δηλώνουμε **y**. Κοντά στο σύμβολο **y**, εμφανίζονται δύο ερωτηματικά. Πιέζοντας το πλήκτρο **Define Groups**, εμφανίζεται το εξής πλαίσιο διαλόγου:



Στο πλαίσιο αυτό, πρέπει να δηλώσουμε ποια κωδική τιμή της **y** υποδηλώνει το κάθε δείγμα. Δίνουμε 0 και 1 και πιέζουμε **Continue**. Στο κύριο πλαίσιο διαλόγου, επιλέγουμε **Mann-Whitney U** στο πεδίο **Test Type** και ζητάμε χρησιμοποίηση της ακριβούς συνάρτησης κατανομής της στατιστικής συνάρτησης ελέγχου. Τα αποτελέσματα του ελέγχου είναι:

Ranks

	Y	N	Mean Rank	Sum of Ranks
X	Αγροικία	12	26.75	321.00
	Πόλη	36	23.75	855.00
	Total	48		

Test Statistics[a]

	X
Mann-Whitney U	189.000
Wilcoxon W	855.000
Z	-.643
Asymp. Sig. (2-tailed)	.520
Exact Sig. (2-tailed)	.529
Exact Sig. (1-tailed)	.264
Point Probability	.004

a Grouping Variable: Y

Από τα αθροίσματα των τάξεων μεγέθους των δύο δειγμάτων στο συνενωμένο δείγμα, το SPSS υπολογίζει τις εξής στατιστικές συναρτήσεις

$$U = T - \frac{1}{2} n \cdot (n+1) \text{ και } U' = T' - \frac{1}{2} m \cdot (m+1),$$

όπου T και T΄ είναι τα αθροίσματα των τάξεων μεγέθους των τιμών του πρώτου και του δεύτερου δείγματος, αντίστοιχα. Στην συνέχεια, παίρνει την μικρότερη από αυτές και την εμφανίζει στον δεύτερο πίνακα ως **Mann-Whitney U**. Κατόπιν, υπολογίζει την τιμή του κρίσιμου επιπέδου για τον αμφίπλευρο έλεγχο και την ακριβή τιμή του κρίσιμου επιπέδου εκείνου του μονόπλευρου ελέγχου με τον οποίο η H_0 θα απορριπτόταν για μικρές τιμές του χρησιμοποιούμενου U. Εδώ, μικρότερη τιμή έχει το U΄ (και αυτή θεωρείται ως Mann-Whitney U). Άρα το κρίσιμο επίπεδο που το πακέτο δίνει, αντιστοιχεί σε μονόπλευρο έλεγχο με εναλλακτική υπόθεση

$$H_1 : P(X < Y) < 0.5$$

δηλαδή ότι η X τείνει να παίρνει μεγαλύτερες τιμές από την Y.

Αυτός ακριβώς είναι ο έλεγχος που επιθυμούμε να κάνουμε. Επομένως, το κρίσιμο επίπεδο είναι 0.264.

Σημείωση: Για τον μονόπλευρο έλεγχο με εναλλακτική υπόθεση H_1: $P(X<Y)>0.5$, δηλαδή ότι η X τείνει να παίρνει μικρότερες τιμές από την Y η τιμή του κρίσιμου επιπέδου προκύπτει με αφαίρεση της τιμής που δίνει το πακέτο (εδώ της τιμής 0.264) από την μονάδα, αφού το άθροισμα των τιμών των στατιστικών συναρτήσεων U και U΄ είναι σταθερό και εξαρτάται μόνο από τα μεγέθη των δύο δειγμάτων. (Συγκεκριμένα, $U + U' = mn$).

Λύση με το SAS: Ο έλεγχος Mann-Whitney στο SAS δίδεται ως Wilcoxon rank-sum test. (Ο έλεγχος εμφανίζεται στην βιβλιογραφία και με αυτό το όνομα). Για την διεξαγωγή του ελέγχου αυτού, μπορούμε να χρησιμοποιήσουμε τις εντολές

```
data test;
input score region @@;
cards;
14.8 1 10.6 1  7.3 1 12.5 1  5.6 1 12.9 1
 6.3 1 16.1 1  9.0 1 11.4 1  4.2 1  2.7 1
12.7 2 16.9 2  7.6 2  2.4 2  6.2 2  9.9 2
14.2 2  7.9 2 11.3 2  6.4 2  6.1 2 10.6 2
12.6 2 16.0 2  8.3 2  9.1 2 15.3 2 14.8 2
 2.1 2 10.6 2  6.7 2  6.7 2 10.6 2  5.0 2
17.7 2  5.6 2  3.6 2 18.6 2  1.8 2  2.6 2
11.8 2  5.6 2  1.0 2  3.2 2 5.9  2  4.0 2
;
run;
proc npar1way wilcoxon;
class region;
var score;
run;
```

Τα αποτελέσματα του ελέγχου δίδονται στον πίνακα που ακολουθεί.

```
                                                The SAS System

                                        N P A R 1 W A Y   P R O C E D U R E

                                    Wilcoxon Scores (Rank Sums) for Variable SCORE
                                           Classified by Variable REGION

              Sum of          Expected          Std Dev            Mean
REGION           N              Scores         Under H0         Under H0
Score

1               12             321.0             294.0       41.9817590
26.7500000
```

```
2              36            855.0            882.0       41.9817590
23.7500000
                               Average Scores Were Used for Ties

                     Wilcoxon 2-Sample Test (Normal Approximation)
                     (with Continuity Correction of .5)

                     S =  321.000         Z = 0.631227         Prob >
|Z| = 0.5279

                     T-Test Approx. Significance = 0.5310

                     Kruskal-Wallis Test (Chi-Square Approximation)
                     CHISQ = 0.41362         DF =  1         Prob >
CHISQ = 0.5201
```

Η ελεγχοσυνάρτηση συμβολίζεται με S και η τιμή της ισούται με 321. Η προσεγγιστική λύση (με διόρθωση συνέχειας) υπολογίζεται ίση με 0.631227 και το αντίστοιχο αμφίπλευρο παρατηρούμενο επίπεδο στατιστικής σημαντικότητας είναι ίσο με 0.5279. Το μονόπλευρο παρατηρούμενο επίπεδο που ζητείται στο συγκεκριμένο παράδειγμα προκύπτει με διαίρεση του 0.5279 με το 2 και ισούται με 0.26395.

Θα πρέπει να σημειωθεί ότι με την υποεντολή **wilcoxon;**, το πρόγραμμα παρέχει τον έλεγχο Mann-Whitney, εφόσον οι υπό εξέταση ομάδες είναι δύο (υπάρχουν δύο επίπεδα κατηγοριοποίησης). Αν τα επίπεδα είναι περισσότερα από δύο, το πακέτο αυτόματα παρέχει τον έλεγχο Kruskal-Wallis.

Επίσης, εκτός από τον ασυμπτωτικό έλεγχο, το πακέτο δίνει και το ακριβές παρατηρούμενο επίπεδο στατιστικής σημαντικότητας με την υποεντολή **exact wilcoxon;**. Στην περίπτωση αυτή βέβαια η διαδικασία υπολογισμού είναι ιδιαίτερα χρονοβόρα.

Το επόμενο παράδειγμα αναφέρεται σε μία περίπτωση όπου δεν υπάρχουν σαφώς ορισμένες μεταβλητές.

Παράδειγμα 3.2.2: Προκειμένου να εξετασθεί αν ο πυρόλιθος που προέρχεται από μία περιοχή Α τείνει να έχει τον ίδιο βαθμό σκληρότητας με τον πυρόλιθο που προέρχεται από μια άλλη περιοχή Β, σχεδιάσθηκε το εξής πείραμα. Τέσσερα τεμάχια πυρόλιθου επελέγησαν τυχαία από την περιοχή Α και πέντε τεμάχια από την περιοχή Β. Για να προσδιορισθεί ποιο από οποιαδήποτε δύο τεμάχια πυρόλιθου ήταν σκληρότερο, τα δύο κομμάτια τρίφτηκαν το ένα πάνω στα άλλο για ένα συγκεκριμένο χρονικό διάστημα. Το τεμάχιο που υπέστη λιγότερη ζημιά κρίθηκε ως το σκληρότερο από τα δύο. Με τον τρόπο αυτό, και τα εννέα κομμάτια του πυρόλιθου διατάχθηκαν σύμφωνα με τον βαθμό σκληρότητάς τους. Ο βαθμός 1 αντιστοιχήθηκε στο πιο μαλακό τεμάχιο, ο βαθμός 2 στο αμέσως σκληρότερο κ.ο.κ. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί.

Προέλευση τεμαχίου	Α	Α	Α	Β	Α	Β	Β	Β	Β
Βαθμός	1	2	3	4	5	6	7	8	9

Θα μπορούσε να θεωρηθεί ότι τα δεδομένα αυτά παρέχουν ενδείξεις υπέρ της υπόθεσης ότι τα δύο είδη πυρόλιθου διαφέρουν ως προς τον βαθμό σκληρότητας;

Λύση: Είναι προφανές ότι οι προς έλεγχο υποθέσεις είναι οι εξής:

H_0 : Τα δύο είδη πυρόλιθου δεν διαφέρουν ως προς τον βαθμό σκληρότητας

H_1 : Τα δύο είδη πυρόλιθου διαφέρουν ως προς τον βαθμό σκληρότητας.

Για τον έλεγχο των υποθέσεων αυτών με τον έλεγχο των Mann-Whitney, ορίζουμε την ελεγχοσυνάρτηση

T= άθροισμα των τάξεων μεγέθους των τεμαχίων πυρόλιθου που προέρχονται από την περιοχή Α.

Η τιμή της συνάρτησης αυτής από τα δεδομένα του προβλήματος υπολογίζεται ίση με 11. Πράγματι, τ = 1 + 2 + 3 + 5 = 11.

Η κρίσιμη περιοχή μεγέθους α=0.05 αντιστοιχεί σε τιμές της στατιστικής συνάρτησης Τ που είναι μικρότερες από την τιμή 12 ($T<12$) και σε τιμές της στατιστικής συνάρτησης Τ οι οποίες υπερβαίνουν την τιμή (4)(10)–12=28 ($T>28$). Επειδή η παρατηρηθείσα τιμή της στατιστικής συνάρτησης Τ είναι μικρότερη από την τιμή 12, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας 5% και συμπεραίνεται ότι, σε επίπεδο σημαντικότητας 0.05, τα δεδομένα παρέχουν επαρκείς ενδείξεις υπέρ της υπόθεσης ότι τα δύο είδη πυρόλιθου διαφέρουν ως προς το βαθμό σκληρότητας.

Το κρίσιμο επίπεδο $\hat{\alpha}$ μπορεί να θεωρηθεί ίσο με 0.05 αφού, όπως εύκολα μπορεί να διαπιστωθεί από τον πίνακα 7 του παραρτήματος, από όλες τις τιμές του $\hat{\alpha}$, η τιμή 0.05 είναι η ελάχιστη η οποία οδηγεί σε απόρριψη της μηδενικής υπόθεσης.

Λύση με το MINITAB: Για την διεξαγωγή του ελέγχου Wilcoxon-Mann-Whitney για δύο ανεξάρτητα δείγματα με τη βοήθεια του MINITAB, εργαζόμαστε όπως στο προηγούμενο παράδειγμα. Τα αποτελέσματα του ελέγχου είναι:

```
Mann-Whitney Confidence Interval and Test
a          N =   4     Median =       2.500
b          N =   5     Median =       7.000
Point estimate for ETA1-ETA2 is      -4.000
96.3 Percent CI for ETA1-ETA2 is (-6.999;-1.000)
W = 11.0
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is significant
at 0.0373
```

Το κρίσιμο επίπεδο, όπως δίνεται από το πρόγραμμα στην τελευταία γραμμή, είναι 0.0373.

Λύση με το SPSS: Για την διεξαγωγή του ελέγχου Wilcoxon-Mann-Whitney για δύο ανεξάρτητα δείγματα με την βοήθεια του SPSS, εργαζόμαστε όπως στο προηγούμενο παράδειγμα. Τα αποτελέσματα του ελέγχου είναι:

Ranks

	Είδος πυρόλιθου	N	Mean Rank	Sum of Ranks
Σκληρότητα	A	4	2.75	11.00
	B	5	6.80	34.00
	Total	9		

Test Statistics[b]

	Σκληρότητα
Mann-Whitney U	1.000
Wilcoxon W	11.000
Z	-2.205
Asymp. Sig. (2-tailed)	.027
Exact Sig. [2*(1-tailed Sig.)]	.032[a]
Exact Sig. (2-tailed)	.032
Exact Sig. (1-tailed)	.016
Point Probability	.008

a Not corrected for ties.
b Grouping Variable: Είδος πυρόλιθου

Αν n και m είναι τα μεγέθη του πρώτου και του δεύτερου δείγματος αντίστοιχα (εδώ n=4, m=5) και T και T΄ τα αθροίσματα των τάξεων μεγέθους των παρατηρήσεων τους στο συνενωμένο δείγμα, τότε το SPSS υπολογίζει τις τιμές

$$U = T - \frac{1}{2} n \cdot (n+1) \text{ και } U' = T' - \frac{1}{2} m \cdot (m+1).$$

Μετά, παίρνει την μικρότερη από αυτές και την εμφανίζει στον δεύτερο πίνακα ως **Mann-Whitney U**. Επειδή εδώ έχουμε αμφίπλευρο έλεγχο, δεν ενδιαφερει σε ποιά από τις δύο παραπάνω συναρτήσεις (U ή U΄) αντιστοιχεί η τιμή Mann-Whitney U που δίνει το πακέτο. Το παρατηρούμενο επίπεδο σημαντικότητας (πεδίο **Exact Sig. (2-tailed)**) είναι 0.032.

Λύση με το SAS: Όπως και στο προηγούμενο παράδειγμα, χρησιμοποιούμε τις εντολές

```
data hardness;
input score region $ @@;
cards;
1 A 2 A 3 A 4 B 5 A 6 B 7 B 8 B 9 B
;
run;
proc npar1way wilcoxon;
class region;
var score;
run;
```

Το σύμβολο **$** που ακολουθεί το όνομα της μεταβλητής **region** δηλώνει ότι η μεταβλητή αυτή θα περιέχει χαρακτήρες (αλφαριθμητική μεταβλητή).

Τα αποτελέσματα δίδονται στον πίνακα που ακολουθεί.

```
                  The SAS System
              N P A R 1 W A Y  P R O C E D U R E

         Wilcoxon Scores (Rank Sums) for Variable SCORE
                  Classified by Variable REGION

             Sum of         Expected       Std Dev            Mean
REGION    N  Scores         Under H0      Under H0            Score
A         4   11.0             20.0       4.08248290       2.75000000
B         5   34.0             25.0       4.08248290       6.80000000

              Wilcoxon 2-Sample Test (Normal Approximation)
              (with Continuity Correction of .5)
    S =  11.0000         Z = -2.08207         Prob > |Z| = 0.0373

              T-Test Approx. Significance = 0.0709
              Kruskal-Wallis Test (Chi-Square Approximation)
    CHISQ =  4.8600         DF =  1         Prob > CHISQ = 0.0275
```

Η τιμή της ελεγχοσυνάρτησης υπολογίσθηκε ίση με 11, ενώ το κατά προσέγγιση αμφίπλευρο παρατηρούμενο επίπεδο στατιστικής σημαντικότητας (με διόρθωση συνέχειας) υπολογίσθηκε ίσο με 0.0373.

Παρατήρηση: Ο έλεγχος Mann-Whitney, μπορεί να χρησιμοποιηθεί επίσης για τον έλεγχο της υπόθεσης

$$H_0: E(X) = E(Y) + d,$$

ή, ισοδύναμα, της υπόθεσης

$$H_0: E(X) - E(Y) = d,$$

για οποιαδήποτε συγκεκριμένη τιμή της σταθεράς d. Η διαδικασία διεξαγωγής του ελέγχου αυτού είναι ισοδύναμη με αυτήν που ακολουθούμε για τον έλεγχο της υπόθεσης

$$H_0: E(X) = E(Y'),$$

όπου $Y' = Y + d$. Αυτό που χρειάζεται να κάνουμε είναι να προσθέσουμε την τιμή d σε κάθε μία από τις τιμές του δείγματος πάνω στην μεταβλητή Y και να εφαρμόσουμε τον έλεγχο Mann-Whitney στο αρχικό δείγμα παρατηρήσεων πάνω στην μεταβλητή X και στο προκύπτον δείγμα των παρατηρήσεων πάνω στην μεταβλητή Y΄.

Αντίστοιχος είναι ο τρόπος με τον οποίο μπορούμε να ελέγξουμε υποθέσεις της παραπάνω μορφής για τις διαμέσους $x_{0.5}$ και $y_{0.5}$ των δύο πληθυσμών.

3.2.1. Διάστημα Εμπιστοσύνης για την Διαφορά των Παραμέτρων Θέσης (Μέσων Τιμών ή Διαμέσων) Δύο Πληθυσμών

Οπως αναφέρθηκε προηγουμένως, ο έλεγχος Mann-Whitney μπορεί να χρησιμοποιηθεί για τον έλεγχο της υπόθεσης H_0: $\theta_X - \theta_Y = d$, όπου θ_X, θ_Y συμβολίζουν τις παραμέτρους θέσης (μέσες τιμές ή διάμεσους) των μεταβλητών X και Y αντίστοιχα. Το διάστημα των δυνατών τιμών της σταθεράς d που οδηγούν σε μη απόρριψη αυτής της μηδενικής υπόθεσης, παρέχουν ένα διάστημα εμπιστοσύνης για την διαφορά $\theta_X - \theta_Y$ μεταξύ των παραμέτρων θέσης των δύο πληθυσμών. Αυτό το διάστημα εμπιστοσύνης είναι συχνά περισσότερο χρήσιμο στον ερευνητή από τον έλεγχο της ισότητας των παραμέτρων θέσης των δύο πληθυσμών.

Στην πράξη, η κατασκευή διαστήματος εμπιστοσύνης για την διαφορά στις παραμέτρους θέσης δύο πληθυσμών δεν στηρίζεται στην επανειλημμένη εφαρμογή του ελέγχου Mann-Whitney. Συνίσταται στον προσδιορισμό της μέγιστης και ελάχιστης δυνατής τιμής των διαφορών $X_i - Y_j$ όλων των δυνατών ζευγών (X_i, Y_j) που συνεπάγονται μη απόρριψη της μηδενικής υπόθεσης.

Τα δεδομένα μας αποτελούνται, όπως στην περίπτωση του ελέγχου Mann-Whitney, από δύο αμοιβαία ανεξάρτητα δείγματα παρατηρήσεων $X_1, X_2, \ldots, X_n$ και $Y_1, Y_2, \ldots, Y_m$ πάνω στις μεταβλητές X και Y αντίστοιχα, των οποίων οι κατανομές θεωρούνται ότι διαφέρουν το πολύ ως προς την παράμετρο θέσης τους. (Με άλλα λόγια, υποθέτουμε ότι υπάρχει μία σταθερά d τέτοια ώστε οι μεταβλητές $X + d$ και Y να είναι ισόνομες). Σημειώνεται ότι η συνέχεια των κατανομών από τις οποίες προέρχονται τα δείγματα δεν αποτελεί απαραίτητη προϋπόθεση. Όπως αποδεικνύεται (Noether (1967)), αν ο συντελεστής εμπιστοσύνης είναι 1-α όταν οι πληθυσμοί είναι συνεχείς, για γενικότερης μορφής πληθυσμούς, ο πραγματικός συντελεστής εμπιστοσύνης του ιδίου διαστήματος συμπεριλαμβανομένων των άκρων του (περίπτωση κλειστού διαστήματος) είναι τουλάχιστον ίσος με 1-α. Εάν τα άκρα του διαστήματος δεν περιλαμβάνονται (περίπτωση ανοικτού διαστήματος) τότε ο συντελεστής εμπιστοσύνης είναι το πολύ 1-α. Στα επόμενα, θα θεωρούμε ότι αναφερόμαστε στην πρώτη περίπτωση, δηλαδή την περίπτωση κλειστού διαστήματος τιμών.

Για τον προσδιορισμό όλων των δυνατών διαφορών μεταξύ των παρατηρήσεων των δύο δειγμάτων, είναι χρήσιμο αυτά να διαταχθούν κατ' αύξουσα σειρά μεγέθους. Παρατηρούμε ότι υπάρχουν mn δυνατά ζεύγη. Έστω ότι k συμβολίζει τον αριθμό των ζευγών όπου $X_i > Y_j$, δηλαδή $X_i - Y_j > 0$. Τότε, η τιμή της στατιστικής συνάρτησης $T = \sum_{i=1}^{n} R(X_i)$ που συμβολίζει το άθροισμα των τάξεων μεγέθους των παρατηρήσεων X_i, είναι ίση με $k+n(n+1)/2$. Δηλαδή, η ύπαρξη k ζευγών (X_i , Y_j) με την ιδιότητα $X_i > Y_j$, συνεπάγεται αύξηση της τιμής της στατιστικής συνάρτησης T κατά k μονάδες.

Σύμφωνα με τα προηγούμενα, η τιμή της στατιστικής συνάρτησης T, για την οποία η μηδενική υπόθεση H_0 είναι *οριακά αποδεκτή (δεν απορρίπτεται)*, σε επίπεδο σημαντικότητας α, δίνεται από τον πίνακα 9 του παραρτήματος ως ίση με $w_{\alpha/2}$, όπου $w_{\alpha/2}$ συμβολίζει το α/2-ποσοστιαίο σημείο της κατανομής της T. Επομένως, αφαιρώντας την τιμή $n(n+1)/2$ από το $w_{\alpha/2}$, μπορούμε να προσδιορίσουμε την οριακή τιμή του αριθμού k των ζευγών (X_i , Y_j) με την ιδιότητα $X_i > Y_j$, για την οποία μπορούμε να οδηγηθούμε σε *οριακή αποδοχή* της μηδενικής υπόθεσης H_0. Δηλαδή,

$$k = w_{\alpha/2} - n(n+1)/2.$$

Εχοντας προσδιορίσει την τιμή αυτή του αριθμού k, χρειαζόμαστε να προσδιορίσουμε την τιμή της σταθεράς d, η οποία προστιθέμενη στις τιμές Y_i οδηγεί σ' αυτή την *οριακή* τιμή του αριθμού k, δηλαδή την τιμή της σταθεράς d για την οποία ακριβώς k από τα ζεύγη (X_i, Y_j+d) ικανοποιούν την ανισότητα $X_i > Y_j+d$, ή, ισοδύναμα, την ανισότητα $X_i - Y_j > d$.

Εάν προσθέσουμε την μέγιστη των διαφορών X_i-Y_j σε κάθε μία από τις τιμές Y_i, είναι προφανές ότι καμία από τις προκύπτουσες τιμές δεν θα είναι μικρότερη ή ίση από τις τιμές X_i. Κατά συνέπεια, προσθέτοντας την k τάξης μεγαλύτερη διαφορά X_i-Y_j σε κάθε μία από τις παρατηρήσεις Y_i, επιτυγχάνουμε την εξής *οριακή* περίπτωση: λιγότερα από k ζεύγη ικανοποιούν την ανισότητα $X_i > Y_j+d$ και τουλάχιστον k ζεύγη ικανοποιούν την ανισότητα X_i - Y_j > d. Οι δύο αυτές ανισότητες προσδιορίζουν την μέγιστη τιμή U της σταθεράς d η οποία οδηγεί σε μη απόρριψη της μηδενικής υπόθεσης H_0: $\theta_X = \theta_Y + d$, Εργαζόμενοι με τον ίδιο τρόπο, αλλά με αντίστροφη σειρά, μπορούμε να βρούμε την ελάχιστη τιμή L της σταθεράς d που οδηγεί σε μη απόρριψη της ίδιας μηδενικής υπόθεσης. Το διάστημα των τιμών που αυτές οι δύο τιμές της σταθεράς d ορίζουν αποτελεί το διάστημα εμπιστοσύνης που θέλουμε να κατασκευάσουμε για την διαφορά d = θ_X - θ_Y με συντελεστή εμπιστοσύνης ≥ 1-α. Συμβολικά,

$$P[L \leq \theta_X - \theta_Y \leq U] \geq 1 - \alpha .$$

Οι τιμές L και U συμβολίζουν αντίστοιχα την k κατά σειρά μικρότερη διαφορά X_i-Y_j και την k κατά σειρά μεγαλύτερη διαφορά X_i-Y_j. Δηλαδή, προχωρώντας προς το κέντρο της διατεταγμένης παράθεσης όλων των mn δυνατών διαφορών, οι διαφορές που βρίσκονται k θέσεις από το κάθε άκρο της παράθεσης αυτής αποτελούν τις τιμές L και U.

Παράδειγμα 3.2.3: Δύο αμοιβαία ανεξάρτητα δείγματα 5 ανδρών και 5 γυναικών χρησιμοποιήθηκαν για τις ανάγκες ενός πειράματος σε ένα εργαστήριο με ελεγχόμενες συνθήκες θερμοκρασίας, στο πλαίσιο μίας έρευνας με στόχο να εξετασθεί αν οι άνδρες και οι γυναίκες διαφέρουν ως προς το ποιό θεωρούν ως ανεκτό επίπεδο θερμοκρασίας. Κάθε ένα από τα άτομα των δύο δειγμάτων ρωτήθηκε να προσδιορίσει την θερμοκρασία που θεωρούσαν ως την πιο ανεκτή (σε ^{0}F). Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί.

Ανεκτό επίπεδο θερμοκρασίας σε ^{0}F

Άνδρες (X_i)	**Γυναίκες (Y_i)**
83	84
79	78
82	79
88	77
82	81

Να εκτιμηθεί η μέση διαφορά στα επίπεδα θερμοκρασίας που θεωρούνται ανεκτά από τους άνδρες και τις γυναίκες με ένα διάστημα με συντελεστή εμπιστοσύνης 95%.

Λύση: Διατάσσοντας κατ' αύξουσα σειρά μεγέθους τις τιμές X_i και Y_i προκύπτουν τα εξήςδιατεταγμένα δείγματα τιμών

$X_{(i)}$	$Y_{(i)}$
79	77
82	78
82	79
83	81
88	84

Για n=5, m=5 και α=0.05, ο πίνακας 9 του παραρτήματος δίνει $w_{0.25}$=18. Επομένως k=18-(5)(6)/2=3. Θεωρούμε, στην συνέχεια, τις τρεις μικρότερες και τρεις μεγαλύτερες από τις 5×5=25 δυνατές διαφορές που μπορούμε να σχηματίσουμε.

Μικρότερες Διαφορές	**Μεγαλύτερες Διαφορές**
79 – 84 = –0.5	88 – 77 = 1.1
79 – 81 = –0.2	88 – 78 = 1.0

82 – 84 = –0.2 = L	88 – 79 =0.9 = U

Το διάστημα με άκρα τις τιμές L = –0.2 και U = 0.9 αποτελεί ένα διάστημα τιμών για την ζητούμενη διαφορά στα επίπεδα της θερμοκρασίας που θεωρείται ανεκτή από τους άνδρες και τις γυναίκες με συντελεστή εμπιστοσύνης τουλάχιστον ίσο με 95%. Δηλαδή, (L, U) = (–0.2, 0.9).

3.2.2 Ο Ελεγχος Kruskal-Wallis

Ο έλεγχος Mann-Whitney για δύο ανεξάρτητα δείγματα μπορεί να επεκταθεί για την περίπτωση προβλημάτων που αναφέρονται σε k πληθυσμούς, k>2 (Kruskal-Wallis, (1952)). Η πειραματική κατάσταση αντιστοιχεί στην περίπτωση όπου k ανεξάρτητα τυχαία δείγματα είναι διαθέσιμα, ένα από κάθε ένα από k, ενδεχομένως διαφορετικούς, πληθυσμούς και θέλουμε να ελέγξουμε την μηδενική υπόθεση ότι όλοι οι πληθυσμοί είναι ισόνομοι έναντι της εναλλακτικής ότι κάποιοι από τους πληθυσμούς οδηγούν σε παρατηρούμενες τιμές που είναι μεγαλύτερες από αυτές στις οποίες οδηγούν οι άλλοι πληθυσμοί.

H_0: Οι συναρτήσεις κατανομής των k πληθυσμών είναι ίσες

H_1: Τουλάχιστον ένας από τους πληθυσμούς παρέχει μεγαλύτερες παρατηρήσεις από τουλάχιστον ένα από τους υπολοίπους πληθυσμούς.

Επειδή ο έλεγχος Kruskal-Wallis στοχεύει στον εντοπισμό διαφορών μεταξύ των μέσων των k πληθυσμών, η εναλλακτική υπόθεση συχνά διατυπώνεται ως εξής:

H_1: Τουλάχιστον δύο από τους k πληθυσμούς έχουν διαφορετικές μέσες τιμές.

Ο όρος *μεγαλύτερες τιμές* αναφέρεται σε παρατηρήσεις πάνω σε τυχαίες μεταβλητές. Στην πράξη, βέβαια, οποιεσδήποτε παρατηρήσεις που μπορούν να διαταχθούν κατ' αύξουσα σειρά μεγέθους σύμφωνα με κάποια ιδιότητα (όπως, για παράδειγμα, οι τιμές που είναι ενδεικτικές της ποιότητας) μπορούν να αναλυθούν χρησιμοποιώντας τον έλεγχο Kruskal-Wallis με τρόπο ανάλογο προς αυτόν με τον οποίο αναλύονται αντίστοιχα δεδομένα στην περίπτωση k=2 με τον έλεγχο Mann-Whitney.

Ο πειραματικός σχεδιασμός που περιγράφεται στα επόμενα είναι γνωστός ως *πλήρως τυχαιοποιημένος πειραματικός σχεδιασμός*. Είναι ο ίδιος σχεδιασμός που προϋποθέτει η χρήση του ελέγχου της διαμέσου ως εναλλακτικής μεθόδου ανάλυσης της περίπτωσης k=2 στην θέση του ελέγχου Mann-Whitney. Όπως θα δούμε στα επόμενα, μία γενίκευση του ελέγχου της διαμέσου για c>2 πληθυσμούς, μπορεί σε ορισμένες περιπτώσεις να χρησιμοποιείται στην θέση του ελέγχου Kruskal-Wallis. Η διαφορά τους βρίσκεται στο ότι ο έλεγχος Kruskal-Wallis προϋποθέτει γνώση περισσοτέρων πληροφοριών προερχομένων από τις παρατηρήσεις από ό,τι ο έλεγχος της διαμέσου. Συγκεκριμένα, ο έλεγχος Kruskal-Wallis είναι συνάρτηση των τάξεων μεγέθους των παρατηρήσεων στο συνενωμένο δείγμα (όπως και ο έλεγχος Mann-Whitney στην περίπτωση k=2), ενώ ο έλεγχος της διαμέσου εξαρτάται μόνο από την γνώση του κατά πόσο οι παρατηρήσεις βρίσκονται πιο πάνω ή πιο κάτω από την συνολική διάμεσο. Για τον λόγο αυτό, ο έλεγχος Kruskal-Wallis είναι συνήθως περισσότερο ισχυρός από τον έλεγχο της διαμέσου.

Η τεχνική στην οποία στηρίζεται ο έλεγχος Kruskal-Wallis μπορεί να θεωρηθεί ως το μη παραμετρικό ανάλογο της παραμετρικής διαδικασίας που είναι γνωστή ως *ανάλυση διασποράς κατά ένα κριτήριο* ή, μερικές φορές, ως *έλεγχος F κατά ένα κριτήριο*. Για τον λόγο αυτό, η διαδικασία στην οποία στηρίζεται ο έλεγχος αυτός είναι γνωστή στην βιβλιογραφία και ως *μη παραμετρική ανάλυση διασποράς κατά ένα κριτήριο* και ο έλεγχος Kruskal-Wallis είναι περισσότερο γνωστός ως *έλεγχος Kruskal-Wallis για την ανάλυση διασποράς κατά* *ένα κριτήριο με βάση τις τάξεις μεγέθους των παρατηρήσεων (the Kruskal-Wallis one way analysis of variance by ranks).*

Τα δεδομένα αποτελούνται από k ανεξάρτητα τυχαία δείγματα ενδεχομένως διαφορετικού μεγέθους, των οποίων τα στοιχεία μπορούν να ταξινομηθούν σε στήλες

Δείγμα 1	**Δείγμα 2**	**…**	**Δείγμα k**
$X_{1,1}$	$X_{2,1}$	…	$X_{k,1}$
$X_{1,2}$	$X_{2,2}$	…	$X_{k,2}$
…	…	…	…
X_{1,n_1}	X_{2,n_2}	…	X_{k,n_k}

όπου, n_i συμβολίζει το μέγεθος του i δείγματος παρατηρήσεων $X_{i1}, X_{i2}, \ldots, X_{i,n_i}$, $i = 1, 2, \ldots, k$. Η κλίμακα μέτρησης των δεδομένων είναι τουλάχιστον κλίμακα διάταξης.

Θεωρούμε το σύνολο των παρατηρήσεων που προκύπτει από την συνένωση των k δειγμάτων και διατάσσουμε κατ' αύξουσα σειρά μεγέθους τις $N = \sum_{i=1}^{k} n_i$ παρατηρήσεις του. Έστω $R(X_{ij})$ η τάξη μεγέθους της παρατήρησης X_{ij} και R_i το άθροισμα των βαθμών (τάξεων μεγέθους) που αντιστοιχίζονται στις παρατηρήσεις του i δείγματος, δηλαδή

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}), \; i = 1, 2, \ldots, k.$$

Η περίπτωση ισοβαθμουσών παρατηρήσεων αντιμετωπίζεται με τον ίδιο τρόπο που αντιμετωπίσθηκε και στον έλεγχο Mann-Whitney.

Η στατιστική συνάρτηση που χρησιμοποιείται για τον έλεγχο των παραπάνω υποθέσεων δίνεται από την σχέση

$$T=\frac{1}{S^2}\left(\sum_{i=1}^{k}\frac{R_i^2}{n_i}-\frac{N(N+1)^2}{4}\right)$$

όπου,

$$S^2=\frac{1}{N-1}\left(\sum_{i,j}R(X_{ij})^2-N\frac{(N+1)^2}{4}\right).$$

Αν δεν υπάρχουν ισοβαθμούσες παρατηρήσεις, η έκφραση S^2 είναι ίση με N(N+1)/12 και η ελεγχοσυνάρτηση παίρνει την μορφή

$$T=\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3\,(N+1)\,.$$

Τα κρίσιμα σημεία της κατανομής της στατιστικής συνάρτησης Τ περιέχονται στον πίνακα 26 του παραρτήματος για k=3, και $n_i \leq 5$ (i =1, 2, 3) για την περίπτωση κατά την οποία δεν υπάρχουν ισοβαθμούσες παρατηρήσεις. Εκτενέστεροι πίνακες για τα ποσοστιαία σημεία της ακριβούς κατανομής της στατιστικής συνάρτησης Τ παρέχονται από τους Iman, Quade και Alexander (1975). Είναι ενδιαφέρον να παρατηρηθεί ότι η κατανομή της στατιστικής συνάρτησης Τ μπορεί να προσεγγισθεί ικανοποιητικά από την κατανομή χ^2 με k–1 βαθμούς ελευθερίας.

Η κρίσιμη περιοχή του ελέγχου και στις δύο περιπτώσεις, αποτελείται από τιμές της στατιστικής συνάρτησης Τ που υπερβαίνουν το (1–α)-ποσοστιαίο σημείο της χρησιμοποιούμενης κατανομής.

Στην περίπτωση που η μηδενική υπόθεση απορρίπτεται, ο ερευνητής μπορεί να ακολουθήσει μια διαδικασία πολλαπλών συγκρίσεων για να προσδιορίσει ποια ζεύγη πληθυσμών εμφανίζουν διαφορές. Στις περιπτώσεις αυτές, θεωρείται ότι οι πληθυσμοί i και j εμφανίζονται να είναι διαφορετικοί αν ισχύει η εξής ανισότητα

$$\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right| > t_{N\text{-}k,\ 1-\alpha/2}\left(S^2\,\frac{N-1-T}{N-k}\right)^{\frac{1}{2}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{\frac{1}{2}},$$

όπου R_i και R_j είναι τα αθροίσματα των τάξεων μεγέθους των δύο δειγμάτων και $t_{N-k,\ 1-\alpha/2}$ είναι το (1–α/2)-ποσοστιαίο σημείο της κατανομής t με N–k βαθμούς ελευθερίας. Το επίπεδο σημαντικότητας που χρησιμοποιείται εδώ είναι το ίδιο με αυτό του αρχικού ελέγχου Kruskal-Wallis.

Παράδειγμα 3.2.4: Τέσσερις μέθοδοι παραγωγής καλαμποκιού εφαρμόσθηκαν τυχαία σ' ένα μεγάλο αριθμό διαφορετικών εκτάσεων γης. Η παραγωγή ανά στρέμμα, που αντιστοιχεί σε κάθε μέθοδο δίνεται στον ακόλουθο πίνακα:

Μέθοδος	**Παραγωγή**									
1	83	91	94	89	89	96	91	92	90	
2	91	90	81	83	84	83	88	91	89	84
3	101	100	91	93	96	95	94			
4	78	82	81	77	79	81	80	81		

Να ελεγχθεί κατά πόσον οι τέσσερις μέθοδοι διαφέρουν σε αποτελεσματικότητα.

Λύση: Οι προς έλεγχο υποθέσεις έχουν την μορφή

H_0: Οι τέσσερις μέθοδοι είναι ισοδύναμες
(εξίσου αποτελεσματικές)

H_1: Κάποιες από τις μεθόδους τείνουν να οδηγούν σε
μεγαλύτερη παραγωγή από τις υπόλοιπες μεθόδους.

Το δείγμα που προκύπτει αν συνδυάσουμε τα 4 δείγματα σε ένα ενιαίο δείγμα μεγέθους $N = n_1 + n_2 + n_3 + n_4 = 34$ και, στην συνέχεια, διατάξουμε τις παρατηρήσεις του κατά αύξουσα σειρά μεγέθους είναι το εξής:

77	78	79	80	81	81	81	81	82	83	83	83	84	84	88	89	89
89	90	90	91	91	91	91	91	92	93	94	94	95	96	96	100	101

Στις παρατηρήσεις του δείγματος αυτού αντιστοιχίζονται οι βαθμοί 1 (στην μικρότερη, 77) μέχρι 34 (στην μεγαλύτερη, 101). Ισοβαθμούσες παρατηρήσεις αντιστοιχίζονται με τους μέσους των βαθμών που αυτές θα είχαν αν δεν ταυτίζονταν. Ο πίνακας που ακολουθεί παρέχει τα αποτελέσματα.

Μέθοδος

	1		**2**		**3**		**4**	
	Παρατήρηση	Τάξη Μεγέθους	Παρατήρηση	Τάξη Μεγέθους	Παρατήρηση	Τάξη Μεγέθους	Παρατήρηση	Τάξη Μεγέθους
	83	11	91	23	101	34	78	2
	91	23	90	19.5	100	33	82	9
	94	28.5	81	6.5	91	23	81	6.5
	89	17	83	11	93	27	77	1
	89	17	84	13.5	96	31.5	79	3
	96	31.5	83	11	95	30	81	6.5
	91	23	88	15	94	28.5	80	4
	92	26	91	23			81	6.5
	90	19.5	89	17				
			84	13.5				
R_i:		196.5		153.0		207.0		38.5
n_i:		9		10		7		8
N=34								

Η κρίσιμη περιοχή μεγέθους κατά προσέγγιση ίσου με α=0.05, αντιστοιχεί στις τιμές της στατιστικής συνάρτησης T που υπερβαίνουν το 95ο ποσοστιαίο σημείο της κατανομής χ^2 με k–1=3 βαθμούς ελευθερίας. Ορίζεται, δηλαδή, από την ανισότητα

$$T > \chi^2_{3,\,0.95} = 7.815 .$$

Η παρατηρούμενη τιμή της ελεγχοσυνάρτησης T για τα δεδομένα του παραδείγματός μας προκύπτει ίση με τ=25.46 και βρίσκεται μέσα στην περιοχή απόρριψης της μηδενικής υπόθεσης. Η τιμή του κρίσιμου επιπέδου είναι πολύ χαμηλή. Πράγματι,

$$\hat{\alpha} = P(T > 25.46 \mid H_0) = P(\chi^2_3 > 25.46) = 1 - P(\chi^2_3 < 25.46) < 0.001 .$$

Κατά συνέπεια, η μηδενική υπόθεση δεν μπορεί να θεωρηθεί εύλογη και έχει έννοια να χρησιμοποιηθεί η διαδικασία των πολλαπλών συγκρίσεων για τον προσδιορισμό των πληθυσμών οι οποίοι διαφέρουν.

Αγνοώντας τον μικρό αριθμό περιπτώσεων ισοβαθμουσών παρατηρήσεων, χρησιμοποιούμε την απλούστερη μορφή S^2=N(N+1)/12=99.167. Επομένως,

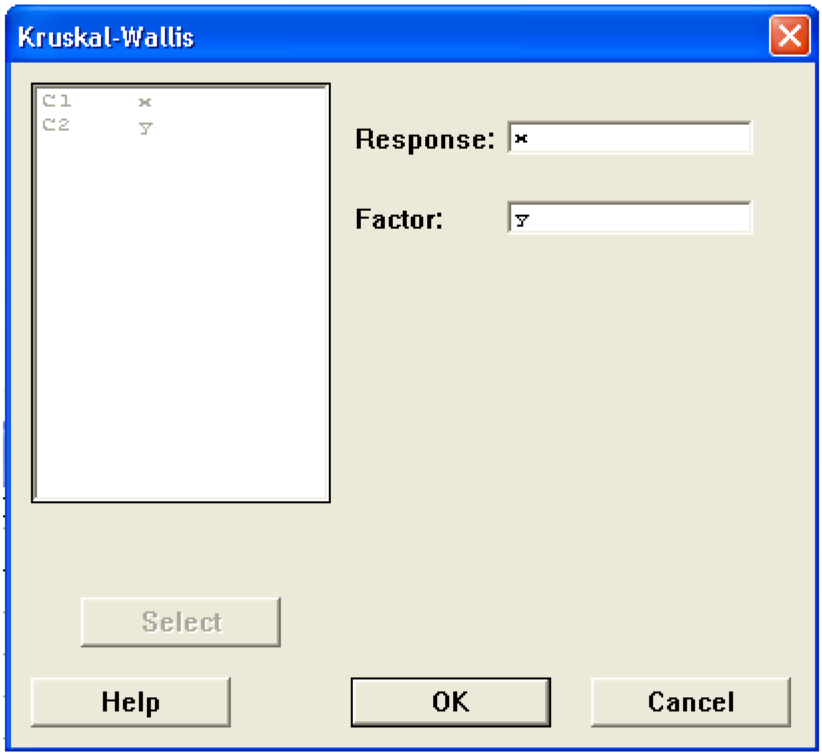
$$\frac{S^2(N-1-T)}{N-k} = \frac{(99.167)\,(33-25.464)}{34-4} = 24.911 .$$

Οι υπόλοιποι υπολογισμοί συνοψίζονται ως εξής:

Πληθυσμοί	$\left\lvert \frac{R_i}{n_i} - \frac{R_j}{n_j} \right\rvert$	$2.041(24.911)^{\frac{1}{2}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{\frac{1}{2}}$
1 και 2	6.533	4.681
1 και 3	7.738	5.134
1 και 4	17.021	4.950
2 και 3	14.271	5.020
2 και 4	10.488	4.832
3 και 4	24.759	5.272

Σε όλες τις περιπτώσεις, οι τιμές στην δεύτερη στήλη υπερβαίνουν αυτές της τρίτης στήλης. Επομένως μπορούμε να συμπεράνουμε ότι η διαδικασία πολλαπλών συγκρίσεων δίνει ενδείξεις ότι όλα τα δυνατά ζεύγη πληθυσμών διαφέρουν.

Λύση με το MINITAB: Για την διεξαγωγή του ελέγχου Kruskal-Wallis με το MINITAB, καταχωρίζουμε σε μία μεταβλητή (έστω **x**) όλες τις μετρήσεις. Σε μία δεύτερη μεταβλητή (έστω **y**), καταχωρίζουμε τιμές-δείκτες που δηλώνουν σε ποιον πληθυσμό (σε ποια μέθοδο στο συγκεκριμένο παράδειγμα) αντιστοιχεί κάθε παρατήρηση (1 για την πρώτη μέθοδο, 2 για τη δεύτερη, κ.ο.κ.). Στην συνέχεια, επιλέγοντας **Stat**, **Nonparametrics**, **Kruskal-Wallis**, οδηγούμεθα στο εξής πλαίσιο διαλόγου:

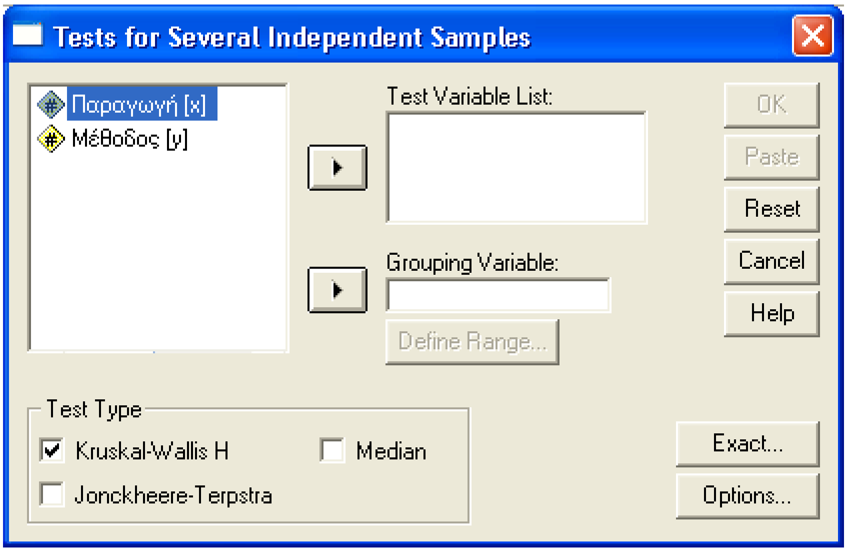


Στο πεδίο **Response**, δηλώνουμε την μεταβλητή **x** που περιέχει τις μετρήσεις και, στο πεδίο **Factor**, δηλώνουμε την μεταβλητή-δείκτη **y** της οποίας οι τιμές δείχνουν σε ποιο πληθυσμό αντιστοιχεί κάθε μέτρηση. Πιέζοντας **OK**, παίρνουμε τα εξής αποτελέσματα:

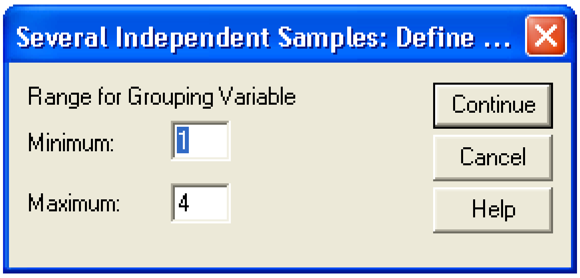
```
Kruskal-Wallis Test on x

y            N    Median    Ave Rank        Z
1            9     91.00        21.8     1.52
2           10     86.00        15.3    -0.83
3            7     95.00        29.6     3.60
4            8     80.50         4.8    -4.12
Overall     34                  17.5

H = 25.46  DF = 3  P = 0.000
H = 25.63  DF = 3  P = 0.000 (adjusted for ties)
```

Η τιμή του κρισίμου επιπέδου είναι εξαιρετικά χαμηλή και, κατά συνέπεια, η μηδενική υπόθεση της ισοδυναμίας των τεσσάρων μέθοδων δεν μπορεί να θεωρηθεί εύλογη.

Λύση με το SPSS: Για την διεξαγωγή του ελέγχου Kruskal-Wallis με το SPSS, καταχωρίζουμε σε μία μεταβλητή (έστω **x**) όλες τις μετρήσεις. Σε μία δεύτερη μεταβλητή (έστω **y**), καταχωρίζουμε τιμές-δείκτες που δηλώνουν σε ποιο πληθυσμό (σε ποια μέθοδο στο συγκεκριμένο παράδειγμα) αντιστοιχεί κάθε παρατήρηση (1 για την πρώτη μέθοδο, 2 για τη δεύτερη, κ.ο.κ.). Στην συνέχεια, επιλέγουμε **Analyze**, **Nonparametric Tests**, **K Independent Samples** και οδηγούμεθα στο ακόλουθο πλαίσιο διαλόγου:



Στο πεδίο **Test Variable List**, δηλώνουμε την μεταβλητή **x** που περιέχει τις μετρήσεις, στο πεδίο **Grouping Variable**, δηλώνουμε την μεταβλητή-δείκτη **y** της οποίας οι τιμές δείχνουν σε ποιο πληθυσμό αντιστοιχεί και, στο **Test Type**, επιλέγουμε **Kruskal-Wallis** H. Πιέζοντας το πλήκτρο **Define Range**, μπορούμε να ορίσουμε όσες από τις τιμές της μεταβλητής-δείκτη **y** αντιστοιχούν στους πληθυσμούς που επιθυμούμε να συγκρίνουμε. Συγκεκριμένα, οδηγούμεθα στο εξής πλαίσιο:



Στο πλαίσιο αυτό, δηλώνουμε την μικρότερη και την μέγιστη τιμή της **y** που θα ληφθεί υπόψη. (Θα μπορούσαμε δηλαδή να εξαιρέσουμε κάποιους πληθυσμούς (κάποιες μεθόδους) από την ανάλυση). Πιέζοντας **Continue** και, στην συνέχεια, **OK** στο επανεμφανιζόμενο αρχικό πλαίσιο, παίρνουμε τα εξής αποτελέσματα:

Ranks

	Μέθοδος	N	Mean Rank
Παραγωγή	1	9	21.83
	2	10	15.30
	3	7	29.57
	4	8	4.81
	Total	34	

Test Statistics

	Παραγωγή
Chi-Square	25.629
df	3
Asymp. Sig.	.000

a Kruskal Wallis Test
b Grouping Variable: Μέθοδος

Οπως μπορούμε να δούμε, η τιμή της ελεγχοσυνάρτησης T είναι τ=25.629 και το παρατηρούμενο επίπεδο σημαντικότητας είναι πάρα πολύ μικρό. Συνεπώς, η μηδενική υπόθεση ότι όλες οι μέθοδοι παραγωγής έχουν την ίδια αποτελεσματικότητα δεν μπορεί να θεωρηθεί εύλογη.

Παρατήρηση: Η διαδικασία των πολλαπλών επιλογών είναι το μη παραμετρικό ανάλογο της συνήθους παραμετρικής διαδικασίας που είναι γνωστή ως *ελάχιστα σημαντική διαφορά του Fisher* με την διαφορά ότι υπολογίζεται με βάση τις τάξεις μεγέθους των παρατηρήσεων και όχι τις τιμές των παρατηρήσεων αυτές καθ' εαυτές.

Παρατήρηση: Οπως αναφέρθηκε προηγουμένως, η μη παραμετρική διαδικασία στην οποία στηρίζεται ο έλεγχος Kruskal-Wallis θεωρείται το μη παραμετρικό ανάλογο της γνωστής *παραμετρικής ανάλυσης διασποράς κατά ένα κριτήριο*. Η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου Kruskal-Wallis σε σχέση με τον έλεγχο F υπερβαίνει πάντα την τιμή 0.864 και τείνει στο άπειρο αν οι κατανομές έχουν ίδια σχήματα αλλά διαφέρουν μόνο κατά την μέση τιμή τους. Αν οι πληθυσμοί είναι κανονικοί, η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου Kruskal-Wallis είναι 0.955, ενώ για ομοιόμορφους πληθυσμούς είναι 1 και για εκθετικούς πληθυσμούς είναι 1.5. Συγκρινόμενος με τον έλεγχο της διαμέσου, ο έλεγχος Kruskal-Wallis έχει ασυμπτωτική σχετική αποτελεσματικότητα ίση με 1.5, 3, και 0.75, αντίστοιχα, για τις τρεις περιπτώσεις κατανομών που αναφέρθηκαν.

3.3 ΕΛΕΓΧΟΙ ΙΣΟΤΗΤΑΣ ΔΙΑΣΠΟΡΩΝ

Οι έλεγχοι που αναπτύχθηκαν στις προηγούμενες ενότητες, πέρα από την χρησιμότητά τους όσο αφορά την σύγκριση μέσων τιμών ή διαμέσων πληθυσμών, αποτελούν, ταυτόχρονα, συνήθεις μεθόδους για την σύγκριση πληθυσμών. Αυτό είναι άμεση συνέπεια του ότι διαφορές στις τιμές των μέσων ή άλλων μέτρων θέσης δύο πληθυσμών συνεπάγονται διαφορά στην μορφή των κατανομών που περιγράφουν τους πληθυσμούς. Συχνά, στην πράξη, ενδιαφέρει να διερευνηθεί η ύπαρξη διαφοράς στην μορφή των πληθυσμών μέσω της σύγκρισης των διασπορών τους. Οι έλεγχοι που περιγράφονται στην συνέχεια παρέχουν αυτή την δυνατότητα, ενώ ταυτόχρονα παρουσιάζουν αυτοί καθ' αυτοί ενδιαφέρον ως έλεγχοι σύγκρισης διασπορών.

3.3.1. Ελεγχος Ισότητας Διασπορών των Siegel-Tukey

Ο έλεγχος αυτός είναι απλός στην διεξαγωγή του, αλλά όχι πολύ ισχυρός. Η βασική ιδέα πίσω από την τεχνική που χρησιμοποιείται είναι ότι, αν δύο δείγματα προέρχονται από πληθυσμούς που διαφέρουν μόνο κατά την διασπορά τους, το δείγμα που προέρχεται από τον πληθυσμό με την μεγαλύτερη διασπορά θα είναι περισσότερο απλωμένο και με μεγαλύτερες ακραίες τιμές. Εάν διατάξουμε κατά αύξουσα σειρά μεγέθους τις παρατηρήσεις του δείγματος που προκύπτει από την συνένωση των δύο επιμέρους δειγμάτων και αντιστοιχίσουμε τον βαθμό 1 στην μικρότερη παρατήρηση, 2 στην μεγαλύτερη, 3 στην επόμενη μεγαλύτερη, 4 και 5 στις επόμενες δύο πιο μικρές, 6 και 7 στις επόμενες πιο μεγάλες κ.ο.κ., το άθροισμα των βαθμών που αντιστοιχούν στις παρατηρήσεις του δείγματος που προέρχεται από τον πληθυσμό με την μεγαλύτερη διασπορά αναμένεται να είναι μικρότερο από ό,τι εάν δεν διαφέρουν οι διασπορές των πληθυσμών. Η εφαρμογή του ελέγχου, βέβαια, δεν θα είναι το ίδιο ικανοποιητική αν, επί πλέον, οι δύο πληθυσμοί διαφέρουν και ως προς την θέση (εάν, δηλαδή, υπάρχουν και διαφορές στις τιμές των μέσων τους ή άλλων μέτρων θέσης). Ένας τρόπος για να υπερπηδηθεί αυτή η δυσκολία, στην περίπτωση που υπάρχουν ενδείξεις ότι οι πληθυσμοί διαφέρουν και ως προς την θέση, είναι να «μετασχηματίσουμε» τις παρατηρήσεις του δείγματος από τον πληθυσμό με την μεγαλύτερη μέση τιμή ή διάμεσο αφαιρώντας από κάθε μία από αυτές μία εκτίμηση της διαφοράς των μέσων ή των διαμέσων των πληθυσμών, αντίστοιχα. Εναλλακτικά, μπορούμε να μετασχηματίσουμε τις παρατηρήσεις του δείγματος που προέρχεται από τον πληθυσμό με την χαμηλότερη μέση τιμή ή διάμεσο προσθέτοντας σε κάθε μία από τις παρατηρήσεις του την εκτίμηση της διαφοράς στις μέσες τιμές ή στις διαμέσους των πληθυσμών, αντίστοιχα. Η διασπορά, όπως είναι γνωστό, δεν επηρεάζεται από μεταβολές στις θέσεις των πληθυσμών. Ο μετασχηματισμός αυτός έχει ως αποτέλεσμα την αύξηση της ισχύος του ελέγχου των Siegel-Tukey.

Οι υποθέσεις που ενδιαφερόμαστε να ελέγξουμε στην περίπτωση αυτή έχουν μία από τις εξής τρεις μορφές:

Α. (Αμφίπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως διαφορετικές μέσες τιμές

H_1: $V(X) \neq V(Y)$

Β. (Μονόπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως
διαφορετικές μέσες τιμές

H_1: $V(X) < V(Y)$

Γ. (Μονόπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως
διαφορετικές μέσες τιμές

H_1: $V(X) > V(Y)$

Οπως αναφέρθηκε παραπάνω, συχνά, σε σχέση με τις πρακτικές εφαρμογές, αν υπάρχει μια διαφορά μεταξύ των κατανομών των δύο μεταβλητών, αυτή αναφέρεται στην διασπορά τους. Στην περίπτωση αυτή, η μηδενική υπόθεση μπορεί να διατυπωθεί ώστε ν' αναφέρεται στις διασπορές των δύο πληθυσμών. Οι περιπτώσεις Α, Β και Γ, τότε, γράφονται με την εξής μορφή:

Α. (Αμφίπλευρος Ελεγχος)

H_0: $V(X) = V(Y)$

H_1: $V(X) \neq V(Y)$

Β. (Μονόπλευρος Ελεγχος)

H_0: $V(X) \geq V(Y)$

H_1: $V(X) < V(Y)$

Γ. (Μονόπλευρος Ελεγχος)

H_0: $V(X) \leq V(Y)$

H_1: $V(X) > V(Y)$

Εστω ότι έχουμε ένα τυχαίο δείγμα παρατηρήσεων $X_1, X_2, ..., X_n$ μεγέθους n πάνω στην τυχαία μεταβλητή X και ένα τυχαίο δείγμα παρατηρήσεων $Y_1, Y_2, ..., Y_m$ μεγέθους m πάνω στην τυχαία μεταβλητή Y. Τα δείγματα αυτά επιλέγονται ανεξάρτητα το ένα του άλλου. Η κλίμακα μέτρησης είναι τουλάχιστον κλίμακα διαστήματος.

Η διαδικασία του ελέγχου όπως περιγράφτηκε παραπάνω, αποτελεί έναν απλό έλεγχο, η εφαρμογή του οποίου είναι ανάλογη της εφαρμογής του ελέγχου των Wilcoxon-Mann-Whitney που αναφέρεται σε διαφορές πληθυσμών στις τιμές των μέσων τους ή άλλων μέτρων θέσης και χρησιμοποιεί τα ποσοστιαία σημεία της στατιστικής συνάρτησης

$$T = \sum_{i=1}^{n} R(X_i)$$

της ενότητας 3.2, τα οποία δίνονται στον πίνακα 9 του παραρτήματος. Ο κανόνας απόφασης διατυπώνεται ως εξής:

Α. (Αμφίδρομος Ελεγχος): Στην περίπτωση του ελέγχου των υποθέσεων της μορφής αυτής, είναι προφανές ότι η μηδενική υπόθεση θα απορρίπτεται σε επίπεδο σημαντικότητας α, αν η στατιστική συνάρτηση T έχει τιμή μικρότερη από το α/2-ποσοστιαίο σημείο της ή μεγαλύτερη από το (1–α/2)-ποσοστιαίο σημείο της. Δηλαδή, η μηδενική υπόθεση απορρίπτεται αν $T < w_{\alpha/2}$ ή $T > w_{1-\alpha/2}$.

Β. (Μονόπλευρος Ελεγχος): Για τον έλεγχο των υποθέσεων της περίπτωσης αυτής, είναι προφανές ότι οι μικρές τιμές της στατιστικής συνάρτησης T αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α, αν η παρατηρούμενη τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής της. Δηλαδή, η H_0 απορρίπτεται αν $T < w_{\alpha}$.

Γ. (Μονόπλευρος Ελεγχος): Για τον έλεγχο των υποθέσεων της περίπτωσης αυτής, είναι προφανές ότι οι μεγάλες τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως, η κρίσιμη περιοχή μεγέθους α για τον έλεγχο αυτό ορίζεται από την ανισότητα $T > w_{1-\alpha}$.

Παράδειγμα 3.3.1: Στα ράφια της βιβλιοθήκης ενός καθηγητή Στατιστικής σε ένα Πανεπιστήμιο της Σκωτίας υπάρχουν 257 βιβλία, από τα οποία 114 είναι βιβλία στατιστικής και 143 βιβλία γενικού ενδιαφέροντος. Ο καθηγητής επιλέγει δύο ανεξάρτητα τυχαία δείγματα βιβλίων, ένα από κάθε κατηγορία, μεγέθους 12 και 16 αντίστοιχα και καταγράφει τους αριθμούς των σελίδων που τα βιβλία αυτά έχουν. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί.

	Αριθμοί σελίδων
Δείγμα 1 (βιβλία στατιστικής, X_i)	126 142 156 228 245 246 370 419 433 454 478 503
Δείγμα 2 (βιβλία γενικού ενδιαφέροντος, Y_j)	29 39 60 78 82 112 125 170 192 224 263 275 276 286 369 756

Ο καθηγητής ενδιαφέρεται να χρησιμοποιήσει τα δύο αυτά δείγματα για να εξετάσει κατά πόσο οι αριθμοί των σελίδων των βιβλίων των δύο κατηγοριών που έχει στην βιβλιοθήκη του παρουσιάζουν διαφορετική μεταβλητότητα. Προκειμένου να έχει μία σημειακή εκτίμηση της διαφοράς των θέσεων των δύο πληθυσμών, σχηματίζει όλες τις δυνατές διαφορές τιμών της μορφής $X_i - Y_j$, i = 1, 2, …, 12, j = 1, 2, …, 16 και χρησιμοποιεί την διάμεσο των 192 διαφορών που προκύπτουν. Έτσι, εκτιμά ότι οι δύο πληθυσμοί παρουσιάζουν διαφορά ως προς την θέση ίση με 133.5. Αυτό, μπορεί εύκολα να διαπιστωθεί, με βάση τον πίνακα που ακολουθεί.

Διαφορές Αριθμών Σελίδων Βιβλίων

$y_i \backslash x_i$	126	142	156	228	245	246	370	419	433	454	478	503
29	97	113	127	199	216	217	341	390	404	425	449	474
39	87	103	117	189	206	207	331	380	394	415	439	464
60	66	82	96	168	185	186	310	359	373	394	418	443
78	48	64	78	150	167	168	292	341	355	376	400	425
82	44	60	74	146	163	164	288	337	351	372	396	421
112	14	30	44	116	133	134	258	307	321	342	366	391
125	1	17	31	103	120	121	245	294	308	329	353	378
170	–44	–28	–14	58	75	76	200	249	263	284	308	333
192	–66	–50	–36	36	53	54	178	227	241	262	286	311
224	–98	–82	–68	4	21	22	146	195	209	230	254	279
263	–137	–121	–107	–35	–18	–17	107	156	170	191	215	240
275	–149	–133	–119	–47	–30	–29	95	144	158	179	203	228
276	–150	–134	–120	–48	–31	–30	94	143	157	178	202	227
286	–160	–144	–130	–58	–41	–40	84	133	147	168	192	217
369	–243	–227	–213	–141	–124	–123	1	50	64	85	109	134
736	–630	–614	–600	–528	–511	–510	–386	–337	–323	–302	–278	–253

Την σημειακή αυτή εκτίμηση της διαφοράς στην θέση των δύο πληθυσμών αριθμών σελίδων προσθέτει σε κάθε μία από τις τιμές για τα βιβλία γενικού ενδιαφέροντος. Στην συνέχεια, συνενώνει τα δύο δείγματα σελίδων σε ένα και διατάσσει κατά αύξουσα σειρά μεγέθους το προκύπτον δείγμα. Σε κάθε μία παρατήρηση του διατεταγμένου δείγματος, αντιστοιχίζει ένα βαθμό με τον τρόπο που περιγράψαμε στα προηγούμενα. Ο πίνακας που ακολουθεί συνοψίζει τα αποτελέσματα. Οι τιμές του πίνακα που εμφανίζονται με υπογράμμιση αντιστοιχούν στις τιμές του ενιαίου δείγματος που συνδέονται με τις τιμές του δείγματος των βιβλίων στατιστικής.

# Σελίδων	$\underline{126}$	$\underline{142}$	$\underline{156}$	162.5	172.5	193.5	211.5	215.5	$\underline{228}$	$\underline{245}$
Βαθμός	$\underline{1}$	$\underline{4}$	$\underline{5}$	8	9	12	13	16	$\underline{17}$	$\underline{20}$
# Σελίδων	245.5	$\underline{246}$	258.5	303.5	325.5	357.5	$\underline{370}$	396.5	408.5	409.5
Βαθμός	21	$\underline{24}$	25	28	27	26	$\underline{23}$	22	19	18
# Σελίδων	$\underline{419}$	419.5	$\underline{433}$	$\underline{454}$	$\underline{478}$	502.5	$\underline{503}$	889.5		
Βαθμός	$\underline{15}$	14	$\underline{11}$	$\underline{10}$	$\underline{7}$	6	$\underline{3}$	2		

Η τιμή της στατιστικής συνάρτησης $T = \sum_{i=1}^{n} R(X_i)$ είναι τ=140. Από τον σχετικό πίνακα του παραρτήματος (πίνακας 7), προκύπτει ότι, σε επίπεδο σημαντικότητας 0.05, η παρατηρηθείσα τιμή της Τ βρίσκεται έξω από την κρίσιμη περιοχή. Πράγματι, για n=12, μ=16, τα 0.025– και 0.975–ποσοστιαία σημεία της Τ είναι $w_{0.025} = 132$ και $w_{0.975} = 12(12+16+1) - w_{0.025} = 348 - 132 = 216$. Επομένως, σε επίπεδο σημαντικότητας 5%, είναι εύλογο να συμπεράνει ο καθηγητής ότι τα βιβλία στατιστικής της βιβλιοθήκης του δεν παρουσιάζουν διαφορετική μεταβλητότητα στους αριθμούς σελίδων τους από τα βιβλία γενικού ενδιαφέροντος.

Λύση με το MINITAB: Το MINITAB δεν διαθέτει τον έλεγχο Siegel-Tukey. Μπορεί μόνο να βοηθήσει στην ταξινόμηση του συνενωμένου δείγματος κατά αύξουσα τάξη μεγέθους.

Λύση με το SPSS: Το SPSS δεν διαθέτει τον έλεγχο Siegel-Tukey. Μπορεί μόνο να βοηθήσει στην ταξινόμηση του συνενωμένου δείγματος κατά αύξουσα τάξη μεγέθους.

3.3.2 Ελεγχος των Τετραγωνικών Τάξεων Μεγέθους για Ισότητα Διασπορών

Ο έλεγχος αυτός, στην εφαρμογή του, είναι ανάλογος με τους ελέγχους που εξετάσθηκαν για την ισότητα μέσων τιμών. Για παράδειγμα, για τον έλεγχο της υπόθεσης H_0: E(X)=E(Y), τα δύο ανεξάρτητα δείγματα που χρησιμοποιούνται, ένα από κάθε πληθυσμό, συνενώνονται σε ένα δείγμα, οι παρατηρήσεις του οποίου διατάσσονται κατά αύξουσα σειρά μεγέθους και το άθροισμα των βαθμών που αντιστοιχούν στις παρατηρήσεις που προέρχονται από το ένα δείγμα, π.χ. το δείγμα των X, χρησιμοποιείται ως ελεγχοσυνάρτηση. Με δεδομένο ότι η διασπορά ενός πληθυσμού ορίζεται ως η αναμενόμενη τιμή του τετραγώνου της διαφοράς της από την μέση τιμή της, για τον έλεγχο της υπόθεσης

$$H_0 : E[(X-\mu_X)^2] = E[(Y-\mu_Y)^2],$$

φαίνεται εύλογο να μεταχειρισθούμε τις τιμές των $(X_i - \mu_X)^2$ και $(Y_j - \mu_Y)^2$ από δύο ανεξάρτητα δείγματα $\{X_1, X_2, \ldots, X_n\}$ και $\{Y_1, Y_2, \ldots, Y_m\}$ όπως μεταχειριζόμαστε τις τιμές των X_i και Y_j για την περίπτωση των μέσων και να χρησιμοποιήσουμε το άθροισμα των τάξεων μεγέθους των τιμών $(X - \mu_X)^2$ ως ελεγχοσυνάρτηση. Στην πράξη, όμως, επιτυγχάνεται μεγαλύτερη ισχύς αν χρησιμοποιηθούν τα τετράγωνα των τάξεων μεγέθους στην θέση των τάξεων μεγέθους αυτών καθεαυτών.

Εστω ότι έχουμε ένα τυχαίο δείγμα παρατηρήσεων $X_1, X_2, \ldots, X_n$ μεγέθους n πάνω στην τυχαία μεταβλητή X και ένα τυχαίο δείγμα παρατηρήσεων $Y_1, Y_2, \ldots, Y_m$ μεγέθους m πάνω στην τυχαία μεταβλητή Y. Τα δείγματα αυτά επιλέγονται ανεξάρτητα το ένα του άλλου. Η κλίμακα μέτρησης είναι τουλάχιστον κλίμακα διαστήματος. Οι υποθέσεις που ενδέχεται να ενδιαφερόμαστε να ελέγξουμε μπορούν να έχουν μία από τις εξής μορφές:

A. (Αμφίπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως διαφορετικές μέσες τιμές

H_1: $V(X) \neq V(Y)$

B. (Μονόπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως διαφορετικές μέ+σες τιμές

H_1: $V(X) < V(Y)$

Γ. (Μονόπλευρος Ελεγχος)

H_0: Οι μεταβλητές X και Y έχουν την ίδια κατανομή με ενδεχομένως διαφορετικές μέσες τιμές

H_1: $V(X) > V(Y)$

Οπως και στην προηγούμενη περίπτωση, η μηδενική υπόθεση μπορεί να διατυπωθεί ώστε ν' αναφέρεται στις διαφορές των δύο πληθυσμών. Οι περιπτώσεις Α, Β και Γ τότε γράφονται ως εξής:

Α. (Αμφίπλευρος Ελεγχος)

H_0: $V(X) = V(Y)$

H_1: $V(X) \neq V(Y)$

Β. (Μονόπλευρος Ελεγχος)

H_0: $V(X) \geq V(Y)$

H_1: $V(X) < V(Y)$

Γ. (Μονόπλευρος Ελεγχος)

H_0: $V(X) \leq V(Y)$

H_1: $V(X) > V(Y)$

Για τον έλεγχο των παραπάνω υποθέσεων θεωρούμε τις μεταβλητές

$$U_i = |X_i - \mu_X|, \qquad i = 1, ..., n$$

και

$$V_j = |Y_j - \mu_Y|, \qquad j = 1, ..., m,$$

όπου

$$\mu_X = E(X) \text{ και } \mu_Y = E(Y).$$

Στην περίπτωση που οι τιμές των μ_X και μ_Y είναι άγνωστες, αντικαθίστανται με τους αντίστοιχους δειγματικούς μέσους $\overline{X}_n$ και $\overline{Y}_m$. Διατάσσουμε κατά αύξουσα σειρά μεγέθους το δείγμα που προκύπτει από την συνένωση των επιμέρους δειγμάτων των U_i, $i = 1, 2, ..., n$ και V_j, $j = 1, 2, ..., m$ και θεωρούμε το δείγμα των βαθμών που αντιστοιχίζονται στις τιμές του προκύπτοντος διατεταγμένου δείγματος. Αν δύο ή περισσότερες τιμές ταυτίζονται, σε κάθε μία από αυτές αντιστοιχίζεται ο μέσος των βαθμών που θα είχαν, αν δεν ταυτίζονταν. Είναι προφανές ότι η διάταξη κατά αύξουσα σειρά μεγέθους των μεταβλητών U_i, $i = 1, 2, ..., n$ και V_j, $j = 1, 2, ..., m$ ισοδυναμεί με την διάταξη κατά αύξουσα σειρά μεγέθους των τιμών $(X_i - \mu_X)^2$, $i = 1,2,...,n$ και $(Y_j - \mu_Y)^2$, $j = 1, 2, ..., m$, αντίστοιχα.

Ας συμβολίσουμε με $R(U_i)$, $i = 1,2,...,n$ και με $R(V_j)$, $j = 1, 2, ..., m$ τους βαθμούς που, με τον παραπάνω τρόπο, αντιστοιχούν στις τιμές που προέρχονται από το δείγμα των X_i, $i = 1, 2, ..., n$ και στις τιμές που προέρχονται από το δείγμα των Y_j, $j = 1, 2, ..., m$, αντίστοιχα. Ως στατιστική συνάρτηση για τον έλεγχο των παραπάνω υποθέσεων, συνήθως χρησιμοποιείται το άθροισμα των τετραγώνων των βαθμών (των τάξεων μεγέθους) που οι παρατηρήσεις $U_1, U_2, ..., U_n$ έχουν στο ενιαίο δείγμα των $n + m$ παρατηρήσεων, δηλαδή

$$T = \sum_{i=1}^{n} [R(U_i)]^2 .$$

Η στατιστική αυτή συνάρτηση χρησιμοποιείται όταν δεν υπάρχουν περιπτώσεις ταύτισης τιμών στο δείγμα. Αν υπάρχουν πολλές περιπτώσεις ταύτισης τιμών, χρησιμοποιείται η τυποποιημένη μορφή της T, η οποία δίνεται από την σχέση

$$T_1 = \frac{T - n\overline{R}^2}{\left[\frac{nm}{N(N-1)}\sum_{k=1}^{N} R_k^4 - \frac{nm}{N-1}(\overline{R}^2)^2\right]^{1/2}},$$

όπου

$N = n + m$, $\overline{R}^2$ συμβολίζει τον μέσο των τετραγώνων των βαθμών των παρατηρήσεων στο συνενωμένο δείγμα, δηλαδή

$$\overline{R}^2 = \frac{1}{N}\left\{\sum_{i=1}^{n}[R(U_i)]^2 + \sum_{j=1}^{m}[R(V_j)]^2\right\}$$

και $\sum R_k^4$ συμβολίζει το άθροισμα της τέταρτης δύναμης των βαθμών των παρατηρήσεων στο συνολικό δείγμα, δηλαδή

$$\sum_{k=1}^{N} R_k^4 = \sum_{i=1}^{n}[R(U_i)]^4 + \sum_{j=1}^{m}[R(V_j)]^4 .$$

Τα ποσοστιαία σημεία της ακριβούς μορφής της κατανομής της στατιστικής συνάρτησης T δίνονται στον πίνακα 9 του παραρτήματος όταν δεν υπάρχουν περιπτώσεις ταύτισης τιμών. Στην περίπτωση μεγάλων δειγμάτων, τα ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T προσεγγίζονται από τα αντίστοιχα ποσοστιαία σημεία της τυποποιημένης κανονικής κατανομής σύμφωνα με την σχέση

$$w_p = \frac{n(N+1)(2N+1)}{6} + z_p\sqrt{\frac{mn(N+1)(2N+1)(8N+1)}{180}},$$

όπου το z_p είναι το p-ποσοστιαίο σημείο της τυποποιημένης κανονικής κατανομής.

Η κατανομή της στατιστικής συνάρτησης T_1, η οποία χρησιμοποιείται οποτεδήποτε υπάρχουν περιπτώσεις ταύτισης τιμών, προσεγγίζεται ικανοποιητικά από την τυποποιημένη κανονική κατανομή.

Στην περίπτωση του ελέγχου των υποθέσεων της μορφής Α, είναι προφανές ότι η μηδενική υπόθεση θα απορρίπτεται σε επίπεδο σημαντικότητας α, αν η στατιστική συνάρτηση T έχει τιμή μικρότερη από το α/2 - ποσοστιαίο σημείο της ή μεγαλύτερη από το (1–α/2) - ποσοστιαίο σημείο της. Δηλαδή, η μηδενική υπόθεση απορρίπτεται αν $T < w_{\alpha/2}$ ή $T > w_{1-\alpha/2}$. Με αντίστοιχο τρόπο ορίζεται η κρίσιμη περιοχή στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T_1.

Για τον έλεγχο των υποθέσεων της περίπτωσης Β είναι προφανές ότι οι μικρές τιμές της στατιστικής συνάρτησης T (αντίστοιχα T_1) θα αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως, η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α, αν η παρατηρούμενη τιμή της χρησιμοποιούμενης στατιστικής συνάρτησης είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής της. Δηλαδή, η H_0 απορρίπτεται αν $T < w_\alpha$, στην περίπτωση που χρησιμοποιείται η T με αντίστοιχη διατύπωση του κανόνα απόφασης στην περίπτωση που η χρησιμοποιούμενη στατιστική συνάρτηση είναι η T_1.

Για τον έλεγχο των υποθέσεων της περίπτωσης Γ, είναι προφανές ότι οι μεγάλες τιμές της χρησιμοποιούμενης στατιστικής συνάρτησης θα αποτελούν ένδειξη υπέρ της απόρριψης της μηδενικής υπόθεσης. Επομένως, η κρίσιμη περιοχή μεγέθους α για τον έλεγχο αυτό ορίζεται από την ανισότητα $T > w_{1-\alpha}$, στην περίπτωση που χρησιμοποιείται η T. Αντίστοιχη είναι η μορφή της κρίσιμης περιοχής στην περίπτωση που χρησιμοποιείται η στατιστική συνάρτηση T_1.

Παράδειγμα 3.3.2: Στο παράδειγμα με τους αριθμούς βιβλίων της προηγούμενης ενότητας, συμβολίζουμε με X_i, i = 1, 2, …,12 τους αριθμούς σελίδων στο δείγμα των 12 βιβλίων στατιστικού περιεχομένου και με Y_j, j = 1, 2, …, 16 τους αριθμούς σελίδων στο δείγμα των 16 βιβλίων γενικού ενδιαφέροντος. Έστω $U_i = \left|X_i - \overline{X}_n\right|$, i = 1, 2, …,12, $V_j = \left|Y_j - \overline{Y}_m\right|$, j = 1, 2, …, 16 τα δείγματα των απολύτων διαφορών των αριθμών των σελίδων των βιβλίων των δύο κατηγοριών από τους αντίστοιχους μέσους τους. Οι παρατηρούμενες τιμές των μέσων είναι $\overline{x} = 316.67$ και $\overline{y} = 208.5$. Στον πίνακα που ακολουθεί περιέχονται οι τιμές αυτών των απολύτων διαφορών συνενωμένες σε ένα ενιαίο δείγμα και διατεταγμένες κατά αύξουσα σειρά μεγέθους (1η γραμμή) με τις τιμές που αντιστοιχούν στα βιβλία στατιστικού περιεχομένου υπογραμμισμένες. Στην 2η και 3η γραμμή του πίνακα περιέχονται οι τάξεις μεγέθους των παρατηρήσεων στο ενιαίο δείγμα και τα τετράγωνά τους αντίστοιχα.

Αποκλίσεις, Βαθμοί (Τάξεις Μεγέθους) και Τετράγωνα Βαθμών

Απόκλιση	15.5	16.5	38.5	$\underline{53.3}$	54.5	66.5	67.5	$\underline{70.7}$	$\underline{71.7}$	77.5
Βαθμός	1	2	3	$\underline{4}$	5	6	7	$\underline{8}$	$\underline{9}$	10
Τετράγωνο Βαθμού	1	4	9	$\underline{16}$	25	36	49	$\underline{64}$	$\underline{81}$	100
Απόκλιση	83.5	$\underline{88.7}$	96.5	$\underline{102.3}$	$\underline{116.3}$	126.5	130.5	$\underline{137.3}$	148.5	
Βαθμός	11	$\underline{12}$	13	$\underline{14}$	$\underline{15}$	16	17	$\underline{18}$	19	
Τετράγωνο Βαθμού	121	$\underline{144}$	169	$\underline{196}$	$\underline{225}$	256	289	$\underline{324}$	361	
Απόκλιση	160.5	$\underline{160.7}$	$\underline{161.3}$	169.5	$\underline{174.7}$	179.5	$\underline{186.3}$	$\underline{190.7}$	547.5	
Βαθμός	20	$\underline{21}$	$\underline{22}$	23	$\underline{24}$	25	$\underline{26}$	$\underline{27}$	28	
Τετράγωνο Βαθμού	400	$\underline{441}$	$\underline{484}$	529	$\underline{576}$	625	$\underline{676}$	$\underline{729}$	784	

Το άθροισμα των υπογραμμισμένων τετραγώνων των βαθμών, τ = 3956, αποτελεί την παρατηρούμενη τιμή της στατιστικής συνάρτησης Τ. Από τον σχετικό πίνακα του παραρτήματος (πίνακας 9) προκύπτει ότι, σε επίπεδο σημαντικότητας 0.05, η παρατηρηθείσα τιμή της στατιστικής συνάρτησης Τ βρίσκεται εκτός της κρίσιμης περιοχής. Πράγματι, για n=12, m=16, το 0.975 – ποσοστιαίο σημείο της κατανομής Τ είναι

$$w_{0.975} = 3306 + z_{0.975} 629.96 = 3306 + (1.96)(629.96) = 4540.53 .$$

Επομένως, σε επίπεδο σημαντικότητας 0.05, δεν απορρίπτεται η μηδενική υπόθεση ισότητας των δύο διασπορών.

Λύση με το MINITAB: Καταχωρίζουμε τα δύο δείγματα στις μεταβλητές **x** (βιβλία στατιστικής) και **y** (βιβλία γενικού ενδιαφέροντος) και υπολογίζουμε τους αντίστοιχους δειγματικούς μέσους. Υπενθυμίζεται ότι ο υπολογισμός των μέσων επιτυγχάνεται με την εντολή **Stat, Basic**

Statistics, Display Descriptive Statistics. Για τα δεδομένα του παραδείγματος προκύπτουν οι εξής δειγματικοί μέσοι:

```
Variable            N       Mean
x                  12      316.7
y                  16      208.5
```

Με την βοήθεια της εντολής **Transform**, **Compute**, δημιουργούμε τις μεταβλητές **u** και **v** των απολύτων αποκλίσεων των **x** και **y**, αντίστοιχα από τους δειγματικούς τους μέσους.

Επειδή η απ’ ευθείας διεξαγωγή του ελέγχου αυτού δεν είναι δυνατή με το MINITAB, μπορεί να ακολουθηθεί η εξής διαδικασία για τον υπολογισμό της τιμής της ελεγχοσυνάρτησης T: Οι μεταβλητές **x** και **y** συνενώνονται σε μία μεταβλητή (έστω **z**) και, με την εντολή **Manip**, **Rank**, υπολογίζονται οι τάξεις μεγέθους των τιμών της και τα τετράγωνά τους. Το άθροισμα των τετραγώνων των τάξεων μεγέθους των τιμών που προέρχονται από την μεταβλητή **x** (δηλαδή η τιμή τ της στατιστικής συνάρτησης T) προκύπτει ίσο με 3956, ίσο δηλαδή με την τιμή που προέκυψε από την αναλυτική λύση του παραδείγματος. Σύγκριση της τιμής αυτής με τα κατάλληλα ποσοστιαία σημεία της κατανομής της ελεγχοσυνάρτησης T οδηγεί στο συμπέρασμα ότι η μηδενική υπόθεση της ισότητας των διασπορών φαίνεται να είναι εύλογη σε επίπεδο σημαντικότητας 5%.

Λύση με το SPSS: Καταχωρίζουμε τα δύο δείγματα στις μεταβλητές **x** (βιβλία στατιστικής) και **y** (βιβλία γενικού ενδιαφέροντος) και υπολογίζουμε τους αντίστοιχους δειγματικούς μέσους. Υπενθυμίζεται ότι ο υπολογισμός των μέσων επιτυγχάνεται με την εντολή **Analyze**,

Descriptive Statistics, **Descriptives**. Για τα δεδομένα του παραδείγματος, προκύπτουν οι εξής δειγματικοί μέσοι:

Descriptive Statistics

	N	Mean
Βιβλία στατιστικής	12	316.67
Βιβλία γενικού ενδιαφέροντος	16	208.50
Valid N (listwise)	12	

Με την βοήθεια της εντολής **Transform**, **Compute**, δημιουργούμε τις μεταβλητές **u** και **v** των απολύτων αποκλίσεων των **x** και **y**, αντίστοιχα από τους δειγματικούς τους μέσους. (Αυτό επιτυγχάνεται με τον ορισμό δύο ψευδομεταβλητών, μιας, της οποίας όλες οι τιμές είναι ίσες με τον δειγματικό μέσο της μεταβλητής **x** και μιας, της οποίας όλες οι τιμές είναι ίσες με τον δειγματικό μέσο της **y**.

Επειδή η απ' ευθείας διεξαγωγή του ελέγχου αυτού δεν είναι δυνατή με το SPSS, μπορεί να ακολουθηθεί η εξής διαδικασία για τον υπολογισμό τςη τιμής της ελεγχοσυνάρτησης T: Οι μεταβλητές **x** και **y** συνενώνονται σε μία μεταβλητή (έστω **z**) και, με την εντολή **Transform**, **Rank Cases**, υπολογίζονται οι τάξεις μεγέθους των τιμών της και τα τετράγωνά τους. Το άθροισμα των τετραγώνων των τάξεων μεγέθους των τιμών που προέρχονται από την μεταβλητή **x** (δηλαδή η τιμή τ της στατιστικής συνάρτησης T) προκύπτει ίσο με 3956, ίσο δηλαδή με την τιμή που προέκυψε από την αναλυτική λύση του παραδείγματος. Σύγκριση της τιμής αυτής με τα κατάλληλα ποσοστιαία σημεία της κατανομής της ελεγχοσυνάρτησης T οδηγεί στο συμπέρασμα ότι η μηδενική υπόθεση της ισότητας των διασπορών φαίνεται να είναι εύλογη σε επίπεδο σημαντικότητας 5%.

Λύση με το SAS: Ο έλεγχος δεν παρέχεται στην συγκεκριμένη έκδοση του πακέτου (έχει προστεθεί σε μεταγενέστερη έκδοση). Για τον υπολογισμό της ελεγχοσυνάρτησης, μπορούμε να ακολουθήσουμε την έμμεση διαδικασία που περιγράψαμε στην επίλυση με τα πακέτα MINITAB και SPSS. Έτσι, ο χειρισμός των μεταβλητών και οι πράξεις που απαιτούνται, γίνονται με τις παρακάτω εντολές.

```
data pages;
input x code @@;
class='statistical';
if code=2 then class='general';
cards;
126 1  142 1  156 1  228 1   245 1  246 1
370 1  419 1  433 1  454 1   478 1  503 1
 29 2   39 2   60 2   78 2    82 2  112 2
125 2  170 2  192 2  224 2   263 2  275 2
276 2  286 2  369 2  756 2
;
run;
proc sort; by class;
proc means mean;
var x;
by class;
run;
data pages;
set pages;
xcenter=0;
if code=1 then xcenter=abs(x-316.66667);
if code=2 then xcenter=abs(x-208.5);
run;
proc rank out=ranks;
var xcenter;
ranks r;
run;
proc print;
run;
data all; merge pages(in=aa)ranks(in=bb);
if aa and bb;
r2=r*r;
run;
proc print;
run;
proc univariate noprint;
var r2;
by class;
output out=results sum= sum1 sum2;
run;
proc print;
run;
```

Τα αποτελέσματα των εντολών, φαίνονται στους πίνακες που ακολουθούν.

```
                     The SAS System

                 Analysis Variable : X
------------------------------- CLASS=general --------------------
                          Mean
                      ------------
                       208.5000000
                      ------------

----------------------------- CLASS=statistical -----------------

                          Mean
                      ------------
                       316.6666667
                      ------------

                      The SAS System
                 OBS    CLASS          SUM1

                  1     general        3758
                  2     statistical    3956
```

Στον πρώτο πίνακα, δίνονται οι μέσες τιμές των αριθμών σελίδων ανά κατηγορία βιβλίων (έτσι ώστε να κατασκευάσουμε τις απόλυτες διαφορές των παρατηρήσεων από τις αντίστοιχες μέσες τιμές). Ο δεύτερος πίνακας δίνει τα αθροίσματα των τετραγώνων των τάξεων μεγέθους ανά κατηγορία βιβλίων. Η τιμή της ελεγχοσυνάρτησης που ζητάμε είναι η τιμή που αντιστοιχεί στα βιβλία στατιστικού περιεχομένου, εδώ ίση με 3956. Σύγκριση της τιμής αυτής με τα κατάλληλα ποσοστιαία σημεία της κατανομής της ελεγχοσυνάρτησης οδηγεί στο συμπέρασμα της μη απόρριψης της μηδενικής υπόθεσης σε επίπεδο σημαντικότητας 5%.

3.4 ΜΕΤΡΑ ΣΥΣΧΕΤΙΣΗΣ ΤΑΞΗΣ ΜΕΓΕΘΟΥΣ
(Measures of Rank Correlation)

Μέτρο Συσχέτισης είναι μία τυχαία μεταβλητή η οποία χρησιμοποιείται σε περιπτώσεις όπου τα δεδομένα αποτελούνται από ζεύγη τιμών, δηλαδή, σε περιπτώσεις που έχουμε ένα διμεταβλητό τυχαίο δείγμα μεγέθους n, έστω (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n). Συχνά, θα χρησιμοποιούμε τον συμβολισμό (X, Y) όταν αναφερόμαστε στο ζεύγος (X_i, Y_i), i = 1, 2, ..., n, εν γένει. Δηλαδή, θα υποθέτουμε ότι οι παρατηρήσεις του δείγματος (X_i, Y_i), i = 1, 2, ..., n είναι ισόνομες και ότι η κοινή διμεταβλητή κατανομή τους είναι η ίδια με την διμεταβλητή κατανομή του τυχαίου διανύσματος (X, Y).

Παραδείγματα διμεταβλητών τυχαίων μεταβλητών περιλαμβάνουν την περίπτωση όπου X_i παριστάνει το ύψος του i ατόμου ενός δείγματος και Y_i παριστάνει το ύψος του πατέρα του, ή την περίπτωση όπου X_i παριστάνει το αποτέλεσμα ενός τεστ για το i άτομο και Y_i παριστάνει το βαθμό κατάρτισης του ατόμου αυτού στο συγκεκριμένο θέμα. Οι τυχαίες μεταβλητές X και Y μπορεί ακόμη να είναι και ανεξάρτητες, όπως θα μπορούσαν να είναι στην περίπτωση που X_i παριστάνει την μέση βαθμολογία ενός παίκτη του basketball και Y_i παριστάνει την βαθμολογία της φίλης του σε ένα συγκεκριμένο μάθημα.

Ενα μέτρο συσχέτισης μεταξύ των μεταβλητών X και Y πρέπει να ικανοποιεί τις εξής προϋποθέσεις για να είναι αποδεκτό.

1. Η τιμή του μέτρου συσχέτισης θα πρέπει να είναι πάντα μεταξύ –1 και +1.

2. Αν οι μεγαλύτερες τιμές της μεταβλητής X τείνουν να αντιστοιχούν στις μεγαλύτερες τιμές της μεταβλητής Y και,

επομένως, οι μικρότερες τιμές της μεταβλητής X τείνουν να αντιστοιχούν στις μικρότερες τιμές της μεταβλητής Y, τότε το μέτρο συσχέτισης θα πρέπει να είναι θετικό και να πλησιάζει την τιμή +1, αν η τάση αυτή είναι ισχυρή. Στην περίπτωση αυτή, θα μιλάμε για *θετική συσχέτιση* μεταξύ των μεταβλητών X και Y.

3. Αν οι μεγαλύτερες τιμές της μεταβλητής X τείνουν να αντιστοιχούν στις μικρότερες τιμές της μεταβλητής Y και αντίστροφα, τότε, το μέτρο συσχέτισης θα πρέπει να έχει μία τιμή αρνητική, η οποία να είναι κοντά στην τιμή +1, αν η τάση είναι ισχυρή. Στην περίπτωση αυτή, θα λέμε ότι οι μεταβλητές X και Y είναι *αρνητικά συσχετισμένες*.

4. Αν οι τιμές της τυχαίας μεταβλητής X φαίνονται να αντιστοιχούν με τυχαίο τρόπο σε τιμές της τυχαίας μεταβλητής Y, το μέτρο συσχέτισης θα πρέπει να έχει μία τιμή αρκετά κοντά στο 0. Αυτή θα ήταν η περίπτωση όπου οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες και, ενδεχομένως, κάποιες άλλες περιπτώσεις όπου οι μεταβλητές X και Y δεν είναι ανεξάρτητες. Θα λέμε στις περιπτώσεις αυτές ότι οι τυχαίες μεταβλητές X και Y είναι *ασυσχέτιστες*, ή ότι *δεν σχετίζονται*, ή ότι *έχουν συσχέτιση* 0.

Το πιο συχνά χρησιμοποιούμενο μέτρο συσχέτισης είναι *ο συντελεστής συσχέτισης του Pearson*, ο οποίος συμβολίζεται με r και ορίζεται ως εξής:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]^{1/2}} ,$$

όπου $\overline{X}$ είναι ο μέσος των τιμών $X_1, X_2, ..., X_n$ και $\overline{Y}$ είναι ο μέσος των τιμών $Y_1, Y_2, ..., Y_n$.

Μία αλλη μορφή του συντελεστή συσχέτισης r του Pearson, με την οποία είναι ευρύτατα γνωστός και η οποία προσφέρεται πολύ περισσότερο για ταχύτερους υπολογισμούς, είναι η εξής:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}\right]^{1/2}\left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]^{1/2}}.$$

Αν ο αριθμητής και ο παρονομαστής στο δεξί μέλος της πρώτης σχέσης διαιρεθούν με n, ο συντελεστής συσχέτισης r μπορεί να γραφεί με τη μορφή

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_I - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]^{1/2}},$$

η οποία μπορεί να ερμηνευθεί ως ο λόγος της συνδιασποράς του δείγματος προς το γινόμενο των τυπικών αποκλίσεων των δύο περιθωρίων δειγμάτων. Είναι προφανές ότι ο συντελεστής αυτός ικανοποιεί τις προϋποθέσεις 1–4. Όμως, η κατανομή του συντελεστή αυτού εξαρτάται από την διμεταβλητή κατανομή του διανύσματος (X, Y). Επομένως, ο συντελεστής r δεν προσφέρεται ως ελεγχοσυνάρτηση για μη παραμετρικούς ελέγχους ή ως στατιστική συνάρτηση για την κατασκευή διαστημάτων εμπιστοσύνης εκτός, βέβαια, αν η κατανομή του διανύσματος (X, Y) είναι γνωστή.

Πολλά άλλα μέτρα συσχέτισης έχουν προταθεί, τα οποία ικανοποιούν τις προαναφερθείσες προϋποθέσεις για να είναι αποδεκτά μέτρα συσχέτισης. Μερικά από αυτά τα μέτρα έχουν κατανομές οι οποίες δεν εξαρτώνται από την κατανομή του διανύσματος (X, Y) αν οι μεταβλητές X και Y είναι ανεξάρτητες και, επομένως, μπορούν να χρησιμοποιηθούν ως ελεγχοσυναρτήσεις σε μη παραμετρικούς ελέγχους ανεξαρτησίας. Τα μέτρα συσχέτισης, τα οποία εξετάζονται στην συνέχεια, εξαρτώνται μόνο απο τις τάξεις μεγέθους των παρατηρήσεων των δειγμάτων και έχουν κατανομές, οι οποίες είναι ανεξάρτητες από την κατανομή του διανύσματος (X, Y), αν οι μεταβλητές X και Y είναι ανεξάρτητες και συνεχείς.

3.4.1 Ο Συντελεστής ρ του Spearman

Εστω (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) ένα δείγμα n παρατηρήσεων πάνω στο τυχαίο διάνυσμα (X, Y). Έστω $R(X_i)$ ο βαθμός ή η τάξη μεγέθους της μεταβλητής X όταν αυτή συγκρίνεται με τις άλλες X τιμές, για i = 1, 2, ..., n. Δηλαδή, $R(X_i) = 1$, αν X_i είναι η μικρότερη από τις τιμές $X_1, X_2, ..., X_n$, $R(X_i) = 2$, αν η μεταβλητή X_i είναι η επόμενη μικρότερη τιμή, κ.ο.κ, με τον βαθμό n να αντιστοιχεί στην μεγαλύτερη τιμή από τις $X_1, X_2, ..., X_n$. Με όμοιο τρόπο, έστω ότι $R(Y_i)$ έχει την τιμή 1, 2, ..., n ανάλογα με το σχετικό μέγεθος της μεταβλητής Y_i, όταν αυτή συγκρίνεται με τις υπόλοιπες Y τιμές.

Τα δεδομένα μπορούν να αποτελούνται και από μη αριθμητικές παρατηρήσεις, οι οποίες εμφανίζονται σε ζεύγη, αν οι παρατηρήσεις είναι τέτοιες που να μπορούν να διαταχθούν κατά αύξουσα σειρά μεγέθους με τον τρόπο που μόλις περιγράψαμε. Στην περίπτωση αυτή, η διάταξη μπορεί να βασίζεται στην ποιότητα των παρατηρήσεων (από την *χειρότερη* παρατήρηση στην *καλύτερη* παρατήρηση) ή στον βαθμό προτίμησης που μπορεί να αντιστοιχηθεί στις παρατηρήσεις κ.ο.κ.

Στις περιπτώσεις όπου δύο ή περισσότερες από τις τιμές ταυτίζονται (tie), αντιστοιχίζουμε σε κάθε μία από τις ίσες αυτές τιμές τον μέσο των βαθμών που θα είχαν αν δεν ταυτίζονταν.

Το μέτρο συσχέτισης που προτάθηκε από τον Spearman το 1904 δεν είναι άλλο από τον συντελεστή r του Pearson υπολογιζόμενο, όμως, με βάση τις τάξεις μεγέθους των παρατηρήσεων και όχι αυτές καθεαυτές τις παρατηρήσεις. Δηλαδή,

$$\rho = \frac{\sum_{i=1}^{n}\left[R(X_i)-\overline{R(X)}\right]\left[R(Y_i)-\overline{R(Y)}\right]}{\left(\sum_{i=1}^{n}\left[R(X_i)-\overline{R(X)}\right]^2\right)^{1/2}\left(\sum_{i=1}^{n}\left[R(Y_i)-\overline{R(Y)}\right]^2\right)^{1/2}},$$

όπου $\overline{R(X)} = \sum_{i=1}^{n} R(X_i)/n$ και $\overline{R(Y)} = \sum_{i=1}^{n} R(Y_i)/n$.

Είναι προφανές, ότι εάν δεν υπάρχουν περιπτώσεις ίσων X τιμών (αντίστοιχα Y τιμών), τότε

$$\overline{R(X)} = \frac{1}{n}\sum_{i=1}^{n} R(X_i) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2},$$

με αντίστοιχη έκφραση για τον μέσο βαθμό των Y τιμών. Επιπλέον,

$$\sum_{i=1}^{n}\left[R(X_i)-\overline{R(X)}\right]^2 = \sum_{i=1}^{n}\left[i-\frac{n+1}{2}\right]^2$$

$$= \sum_{i=1}^{n}\left[i^2+\left[\frac{n+1}{2}\right]^2-2\,\frac{i\,(n+1)}{2}\right]$$

$$= \sum_{i=1}^{n} i^2 + \frac{n\,(n+1)^2}{4} - (n+1)\sum_{i=1}^{n} i$$

$$= \frac{n\,(n+1)\,(2n+1)}{6} + \frac{n\,(n+1)^2}{4} - \frac{n\,(n+1)^2}{2}$$

$$= \frac{n\,(n+1)\,(2n+1)}{6} - \frac{n\,(n+1)^2}{4}$$

$$= \frac{n\,(n+1)}{2}\left[\frac{2n+1}{3}-\frac{n+1}{2}\right]$$

$$= \frac{n\,(n+1)}{12}(n-1)$$

$$= \frac{n\,(n^2-1)}{12},$$

με αντίστοιχη έκφραση για τις Y τιμές. Επομένως, αν όλες οι παρατηρήσεις είναι διακεκριμένες, ο συντελεστής ρ του Spearman μπορεί να γραφεί με την ισοδύναμη μορφή

$$\rho = \frac{\sum_{i=1}^{n}\left(R(X_i)-\frac{n+1}{2}\right)\left[R(Y_i)-\frac{n+1}{2}\right]}{n\left(n^2-1\right)/12}.$$

Στην περίπτωση αυτή, συχνά, χρησιμοποιείται μία ισοδύναμη μορφή του συντελεστή ρ, η οποία προσφέρεται περισσότερο για ταχύτερους υπολογισμούς:

$$\rho = 1-\frac{6T}{n\left(n^2-1\right)},$$

όπου

$$T = \sum_{i=1}^{n} \left[R(X_i) - R(Y_i) \right]^2 .$$

Αν οι X τιμές (αντίστοιχα οι Y τιμές) δεν είναι όλες διακεκριμένες, δηλαδή υπάρχουν περιπτώσεις ίσων τιμών, τότε χρησιμοποιείται η εξής μορφή του συντελεστή ρ:

$$\rho = \frac{\sum_{i=1}^{n} \left[R(X_i) - \frac{n+1}{2} \right] \left[R(Y_i) - \frac{n+1}{2} \right]}{\sqrt{\sum_{i=1}^{n} \left[R(X_i) - \frac{n+1}{2} \right]^2 \sum_{i=1}^{n} \left[R(Y_i) - \frac{n+1}{2} \right]^2}}$$

$$= \frac{\sum_{i=1}^{n} R(X_i) R(X_i) - n \left[\frac{n+1}{2} \right]^2}{\sqrt{\left[\sum_{i=1}^{n} R(X_i)^2 - n\left(\frac{n+1}{2} \right)^2 \right] \left[\sum_{i=}^{n} R(Y_i)^2 - n\left(\frac{n+1}{2} \right)^2 \right]}} .$$

Παράδειγμα 3.4.1: Δώδεκα ζεύγη διδύμων υποβλήθηκαν σε ένα ψυχολογικό τεστ για να μετρηθεί η επιθετικότητά τους. Η έμφαση ήταν στην εξέταση του βαθμού ομοιότητας μεταξύ των διδύμων του ιδίου ζεύγους. Τα δεδομένα παριστάνουν μετρήσεις της επιθετικότητας και συνοψίζονται στον πίνακα που ακολουθεί.

Ζεύγος διδύμων i	1	2	3	4	5	6	7	8	9	10	11	12
Πρωτότοκος X_i	86	71	77	68	91	72	77	91	70	71	88	87
Δευτερότοκος Y_i	88	77	76	64	96	72	65	90	65	80	81	72

Οι πρωτότοκοι όλων των ζευγαριών διδύμων διατάχθηκαν ως προς την επιθετικότητά τους κατά αύξουσα τάξη μεγέθους, όπως και οι δευτερότοκοι των ζευγαριών αυτών, με τα εξής αποτελέσματα:

Ζεύγος διδύμων i	1	2	3	4	5	6	7	8	9	10	11	12
$R(X_i)$	8	3.5	6.5	1	11.5	5	6.5	11.5	2	3.5	10	9
$R(Y_i)$	10	7	6	1	12	4.5	2.5	11	2.5	8	9	4.5
$[R(X_i) - R(Y_i)]$	4	12.25	0.25	0	0.25	0.25	16	0.25	0.25	20.25	1	20.25

Από τα δεδομένα του πίνακα αυτού, προκύπτει ότι η τιμή της στατιστικής συνάρτησης Τ είναι

$$\tau = \sum_{i=1}^{12} \left[R(X_i) - R(Y_i) \right]^2 = 75.$$

Επομένως, ο συντελεστής συσχέτισης ρ του Spearman είναι

$$\rho = 1 - \frac{6\tau}{n\,(n^2 - 1)}$$

$$= 1 - \frac{6\,(75)}{12\,(144 - 1)}$$

$$= 0.7378.$$

Ο συντελεστής συσχέτισης του Spearman χρησιμοποιείται συχνά ως ελεγχοσυνάρτηση για τον έλεγχο της ανεξαρτησίας μεταξύ δύο τυχαίων μεταβλητών. Στην πραγματικότητα, ο συντελεστής συσχέτισης ρ του Spearman δεν είναι ευαίσθητος σε ορισμένες μορφές εξάρτησης. Για τον λόγο αυτό, είναι προτιμότερο να είναι κανείς συγκεκριμένος όσο αφορά την μορφή της εξάρτησης που επιθυμεί να ελέγξει. Επομένως, οι δυνατές υποθέσεις που ενδέχεται να ενδιαφερόμαστε να ελέγξουμε παίρνουν την εξής μορφή:

Α. (Αμφίπλευρος έλεγχος)

H_0: Οι μεταβλητές X και Y είναι αμοιβαία ανεξάρτητες.

H_1: Είτε υπάρχει τάση οι μεγαλύτερες τιμές της μεταβλητής X να αντιστοιχούν στις μεγαλύτερες τιμές της μεταβλητής Y, είτε υπάρχει τάση στις μικρότερες τιμές της μεταβλητής X να αντιστοιχούν στις μεγαλύτερες τιμές της μεταβλητής Y.

B. (Μονόπλευρος έλεγχος για θετική συσχέτιση)

H_0: Οι μεταβλητές X και Y είναι αμοιβαία ανεξάρτητες.

H_1: Υπάρχει τάση οι μεγαλύτερες τιμές της μεταβλητής X να αντιστοιχούν στις μεγαλύτερες τιμές της μεταβλητής Y και αντίστροφα.

Γ. (Μονόπλευρος έλεγχος για αρνητική συσχέτιση)

H_0: Οι μεταβλητές X και Y είναι αμοιβαία ανεξάρτητες.

H_1: Υπάρχει τάση οι μικρότερες τιμές της μεταβλητής X να αντιστοιχούν στις μεγαλύτερες τιμές της μεταβλητής Y και αντίστροφα.

Οι εναλλακτικές υποθέσεις που θεωρήθηκαν παραπάνω διατυπώνουν την ύπαρξη συσχέτισης μεταξύ X και Y. Τότε, μία μηδενική υπόθεση "*μη ύπαρξης συσχέτισης μεταξύ* X *και* Y" θα ήταν περισσότερο ακριβής από την υπόθεση της " *ύπαρξης ανεξαρτησίας μεταξύ* X *και* Y", όπως θεωρήθηκε παραπάνω. Όμως, η μηδενική υπόθεση δόθηκε και στις τρεις παραπάνω περιπτώσεις με την δεύτερη μορφή της, γιατί αυτή χρησιμοποιείται περισσότερο και είναι ευκολότερο να ερμηνευθεί.

Ο συντελεστής συσχέτισης ρ του Spearman μπορεί να χρησιμοποιηθεί ως ελεγχοσυνάρτηση για τις παραπάνω υποθέσεις. Ο πίνακας 10 του παραρτήματος δίνει τα ποσοστιαία σημεία της κατανομής του συντελεστή ρ κάτω από την μηδενική υπόθεση της ανεξαρτησίας των μεταβλητών X και Y. Τότε, η μηδενική υπόθεση H_0 της περίπτωσης B απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή του συντελεστή ρ είναι πολύ μεγάλη, δηλαδή, αν η τιμή του ρ υπερβαίνει το (1-α)-ποσοστιαίο σημείο της κατανομής του ρ. Αντίστοιχα, η μηδενική υπόθεση H_0 της περίπτωσης Γ απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή του συντελεστή ρ είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής του. Τέλος, η μηδενική υπόθεση H_0 της περίπτωσης A απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή του συντελεστή ρ υπερβαίνει το (1–α/2)-ποσοστιαίο σημείο της κατανομής του ρ ή αν είναι μικρότερη από το α/2-ποσοστιαίο σημείο της κατανομής του ρ.

Αν, επομένως, επιθυμούμε να ελέγξουμε τις υποθέσεις της περίπτωσης A σε επίπεδο σημαντικότητας 0.05, τότε θα πρέπει να συγκρίνουμε την παρατηρηθείσα τιμή του συντελεστή ρ με τις τιμές των 0.025 και 0.975 ποσοστιαίων σημείων του σχετικού πίνακα του παραρτήματος. Από τον πίνακα αυτόν προκύπτει ότι

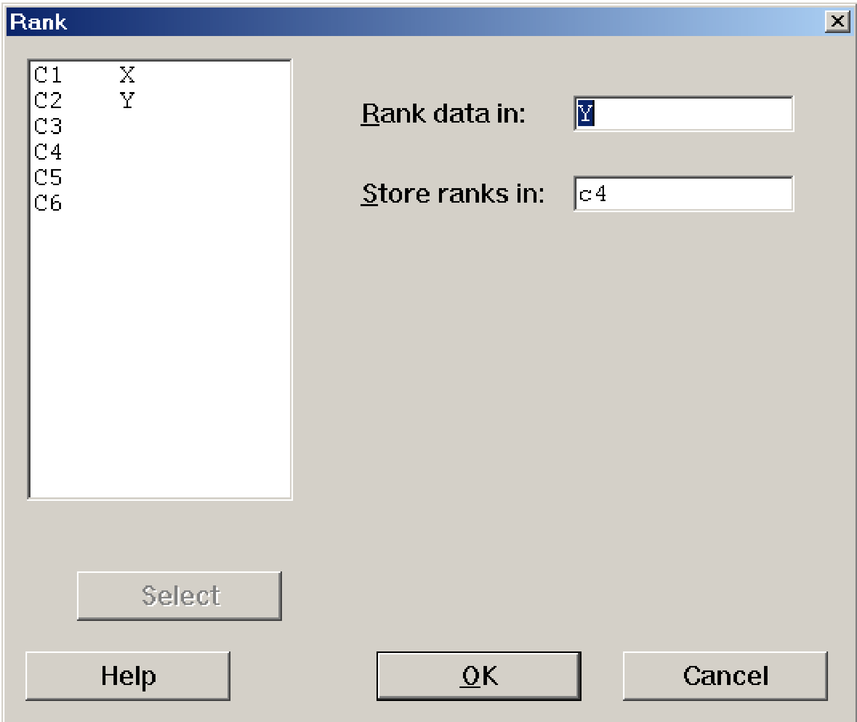
$$w_{0.975} = 0.5804 \text{ και } w_{0.025} = -w_{0.975} = -0.5804.$$

Επομένως, η παρατηρηθείσα τιμή 0.7378 του ρ υπερβαίνει την κρίσιμη τιμή του 0.975 ποσοστιαίου σημείου και, κατά συνέπεια, η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας 5%.

Το κρίσιμο επίπεδο του ελέγχου αυτού είναι περίπου 0.01, όπως προκύπτει από τον σχετικό πίνακα του παραρτήματος.

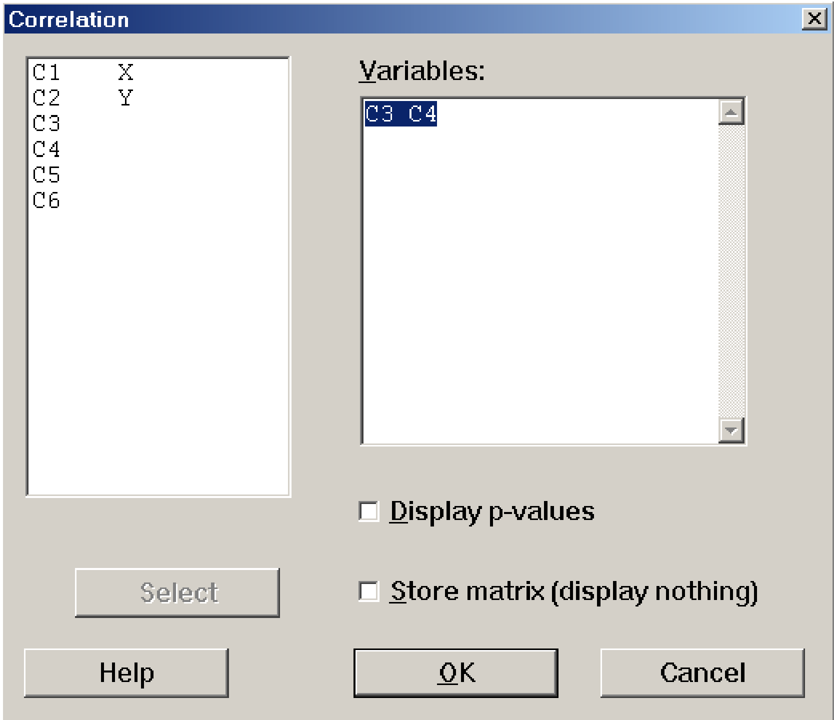
Λύση με το MINITAB: Το MINITAB δεν δίνει απ’ ευθείας το συντελεστή συσχέτισης του Spearman. Μπορεί, όμως να υπολογισθεί με την εξής διαδικασία:

Καταχωρίζουμε τα δύο δείγματα στις στήλες **C1** και **C2** (έστω με ονόματα **X** και **Y** αντίστοιχα) και κατόπιν επιλέγουμε **Manip, Rank** το οποίο οδηγεί στο πλαίσιο διαλόγου:



Στο πεδίο **Rank data in**, δηλώνουμε το όνομα της στήλης για τις παρατηρήσεις της οποίας ζητάμε τις τάξεις μεγέθους. Στο πεδίο **Store ranks in**, δηλώνουμε για κάθε μία από τις μεταβλητές **X** και **Y** το όνομα μίας στήλης στην οποία θα καταχωρισθούν οι τάξεις μεγέθους της, π.χ. **C3** (για την μεταβλητή **X**) και **C4** (για την μεταβλητή **Y**). Ο συντελεστής συσχέτισης του Spearman για τις μεταβλητές **X** και **Y** που θέλουμε να υπολογίσουμε είναι ο συνήθης συντελεστής συσχέτισης των **C3** και **C4** κατά Pearson.

Ο υπολογισμός του συντελεστή αυτού επιτυγχάνεται ως εξής: Επιλέγοντας **Stat, Basic Statistics, Correlation**, προκύπτει το εξής πλαίσιο διαλόγου:

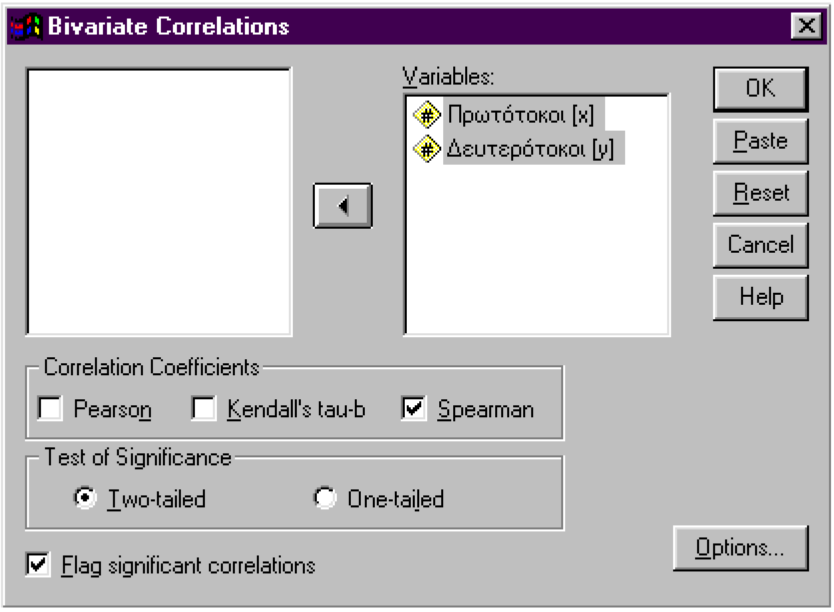


Στο πεδίο **Variables**, δηλώνουμε τις δύο στήλες που περιέχουν τις τάξεις μεγέθους (**C3**, **C4**). Δεν έχει νόημα να ζητηθεί ο υπολογισμός της τιμής του κρίσιμου επιπέδου (πεδίο **p-values**), αφού αυτός θα γίνει με βάση την κατανομή του συντελεστή συσχέτισης κατά Pearson και δεν έχει έννοια για το παράδειγμά μας. Το αποτέλεσμα που παίρνουμε είναι:

```
Correlation of C3 and C4 = 0.735
```

Σύγκριση της τιμής του συντελεστή με τα ποσοστιαία σημεία που δίνει ο πίνακας 10 για α=0.05 και n=12 μας οδηγεί στο συμπέρασμα ότι η μηδενική υπόθεση ότι δεν υπάρχει συσχέτιση μεταξύ **X** και **Y** δεν μπορεί να θεωρηθεί εύλογη σε επίπεδο σημαντικότητας 0.05.

Λύση με το SPSS: Για να υπολογίσουμε το συντελεστή συσχέτισης του Spearman, καταχωρίζουμε τα δύο δείγματα σε δύο μεταβλητές (έστω **x** και **y**) και επιλέγουμε **Analyze, Correlate, Bivariate**. Αυτό οδηγεί στο παρακάτω πλαίσιο διαλόγου:



Στο πεδίο **Variables**, δηλώνουμε όλες τις μεταβλητές των οποίων θέλουμε να υπολογίσουμε τους συντελεστές συσχέτισης. (Αν δηλώσουμε περισσότερες από δύο μεταβλητές, θα μας δοθεί ο πίνακας των συντελεστών συσχέτισής τους). Στο πεδίο **Correlation Coefficients**, δηλώνουμε το είδος των συντελεστών συσχέτισης που θέλουμε. Για τις ανάγκες του παραδείγματός μας, επιλέγουμε Spearman. Στο πεδίο **Test of Significance**, δηλώνουμε αν επιθυμούμε μονόπλευρο ή αμφίπλευρο έλεγχο. Για τις ανάγκες του παραδείγματος δηλώνουμε **Two-tailed**. Τέλος, επιλέγοντας **Flag significant correlations**, ζητάμε από το πρόγραμμα μόνο του να μας δώσει ένδειξη για το κατά πόσο ο συντελεστής συσχέτισης που θα υπολογισθεί είναι στατιστικά σημαντικός. Πιέζοντας **ΟΚ**, οδηγούμεθα στον εξής πίνακα αποτελεσμάτων:

Correlations

			Πρωτότοκοι	Δευτερότοκοι
Spearman's rho	Πρωτότοκοι	Correlation Coefficient	1.000	.735**
		Sig. (2-tailed)	.	.006
		N	12	12
	Δευτερότοκοι	Correlation Coefficient	.735**	1.000
		Sig. (2-tailed)	.006	.
		N	12	12

** Correlation is significant at the .01 level (2-tailed).

Λύση με το SAS: Στο παράθυρο εντολών πληκτρολογούμε

```
data d341;
input x y @@;
cards;
86 88 71 77 77 76 68 64 91 96 72 72 77 65 91 90 70 65 71 80 88 81 87 72
;
run;
proc print;
run;
proc corr spearman;
var x y;
run;
```

Το αποτέλεσμα περιέχεται στον εξής πίνακα:

```
Spearman Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 12

                                X                  Y

                  X          1.00000            0.73545
                              0.0                0.0064

                  Y          0.73545            1.00000
                              0.0064             0.0
```

Τα στοιχεία στην πρώτη γραμμή των κελλιών του πίνακα αντιστοιχούν στις τιμές του συντελεστή συσχέτισης, ενώ η δεύτερη στις τιμές των κρίσιμων επιπέδων.

Παρατήρηση: Συχνά, για τον έλεγχο των υποθέσεων των περιπτώσεων Α, Β και Γ, αντί να χρησιμοποιηθεί ο συντελεστής ρ του Spearman, χρησιμοποιείται η στατιστική συνάρτηση

$$T = \sum_{i=1}^{n}\left[R(X_i) - R(Y_i)\right]^2 .$$

Θα πρέπει να σημειωθεί, βέβαια, ότι, οποτεδήποτε υπάρχουν αρκετές περιπτώσεις ταύτισης τιμών, θα πρέπει να χρησιμοποιείται ο συντελεστής ρ.

Ο έλεγχος, ο οποίος στηρίζεται στην Τ, είναι γνωστός ως *έλεγχος των Hotelling και Pabst*. Τα ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης Τ δίνονται στον πίνακα 11 του παραρτήματος. Ας σημειωθεί, όμως, ότι η τιμή της στατιστικής συνάρτησης Τ είναι μεγάλη όταν η τιμή της στατιστικής συνάρτησης ρ είναι μικρή και αντιστρόφως.

Επομένως, η μηδενική υπόθεση H_0 της περίπτωσης Β απορρίπτεται σε επίπεδο σημαντικότητας α, αν η τιμή της στατιστικής συνάρτησης Τ είναι μικρότερη από το α-ποσοστιαίο σημείο της κατανομής της. Επίσης, η μηδενική υπόθεση H_0 της περίπτωσης Γ απορρίπτεται σε επίπεδο σημαντικότητας α αν η τιμή της στατιστικής συνάρτησης Τ υπερβαίνει το (1–α)-ποσοστιαίο σημείο της κατανομής της.

Παράδειγμα 3.4.2: Ας θεωρήσουμε τα δεδομένα του προηγούμενου παραδείγματος και ας υποθέσουμε ότι θέλουμε να ελέγξουμε τις υποθέσεις:

H_0: Τα μέτρα επιθετικότητας των δύο διδύμων είναι αμοιβαία ανεξάρτητα.

H_1: Υπάρχει είτε θετική συσχέτιση είτε αρνητική συσχέτιση μεταξύ των δύο μέτρων επιθετικότητας.

Λύση: Ας υποθέσουμε ότι επιθυμούμε ο έλεγχος των υποθέσεων αυτών να γίνει σε επίπεδο σημαντικότητας 0.05. Από τον σχετικό πίνακα του παραρτήματος, τα 0.025 και 0.975-ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T (για n=12) είναι

$$w_{0.025} = 120 \quad \text{και} \quad w_{0.975} = \frac{1}{3}n(n^2-1) - w_{0.025} = 452.$$

Οπως, όμως, είδαμε στο προηγούμενο παράδειγμα, η τιμή της στατιστικής συνάρτησης T είναι τ=75. Η τιμή αυτή βρίσκεται, επομένως, μέσα στην κρίσιμη περιοχή μεγέθους 0.05, αφού είναι μικρότερη από την κρίσιμη τιμή 120 του 0.025 ποσοστιαίου σημείου της κατανομής της T. Επομένως, η μηδενική υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας 0.05. Το κρίσιμο επίπεδο και αυτού του ελέγχου εκτιμάται με βάση τον πίνακα 11 του παραρτήματος ίσο περίπου με 0.01.

Λύση με το MINITAB: Στο MINITAB μπορούμε να υπολογίσουμε την τιμή της ελεγχοσυνάρτησης T αν, σε μία στήλη (έστω **C5**), καταχωρίσουμε τα τετράγωνα των διαφορών των τιμών των στηλών **C3** και **C4** και, στην συνέχεια, υπολογίσουμε το άθροισμα των προκυπτουσών τιμών της. Και τα δύο αυτά βήματα γίνονται με τη βοήθεια της επιλογής **Calc**, **Calculator** κατά τα ήδη γνωστά από προηγηθέντα παραδείγματα.

Λύση με το SPSS: Το SPSS δεν υπολογίζει απ' ευθείας την τιμή της συνάρτησης T. Ο υπολογισμός της τιμής της μπορεί να γίνει έμμεσα. Αρκεί να καταχωρίσουμε τις τάξεις μεγέθους των παρατηρήσεων σε δύο νέες μεταβλητές και να δημιουργήσουμε μία μεταβλητή που θα περιέχει τα τετράγωνα των διαφορών τους. Το άθροισμα των τιμών αυτής της στήλης είναι η τιμή της T.

Λύση με το SAS: Στο παράθυρο εντολών πληκτρολογούμε

```
data d341;
input x y @@;
cards;
86 88 71 77 77 76 68 64 91 96 72 72 77 65 91 90 70 65 71 80 88 81 87 72
;
run;
proc print;
run;
proc corr spearman;
var x y;
run;
```

Το αποτέλεσμα περιέχεται στον εξής πίνακα αποτελεσμάτων:

```
Spearman Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 12

                                              X                  Y

                            X            1.00000            0.73545
                                          0.0                0.0064

                            Y            0.73545            1.00000
                                          0.0064             0.0
```

Τα στοιχεία στην πρώτη γραμμή των κελλιών του πίνακα αντιστοιχούν στις τιμές του συντελεστή συσχέτισης, ενώ η δεύτερη στις τιμές των κρίσιμων επιπέδων.

3.4.2 Ο Συντελεστής Συσχέτισης τ του Kendall

Ο συντελεστής συχέτισης τ του Kendall, γνωστός και ως συντελεστής *εναρμόνισης* του Kendall, μοιάζει με τον συντελεστή ρ του Spearman ως προς το ότι υπολογίζεται με βάση την τάξη μεγέθους των παρατηρήσεων και όχι με βάση τις παρατηρήσεις αυτές καθεαυτές και, επιπλέον, η κατανομή του δεν εξαρτάται από την κατανομή των μεταβλητών X και Y, όταν αυτές είναι ανεξάρτητες και συνεχείς. Το κύριο πλεονέκτημα του μέτρου αυτού σε σχέση με το μέτρο ρ του Spearman είναι ότι τείνει στην κανονική κατανομή σχετικά γρήγορα. Αποτέλεσμα αυτού είναι ότι η προσέγγιση της κατανομής του συντελεστή τ από την κανονική κατανομή είναι καλύτερη από την αντίστοιχη προσέγγιση της κατανομής του συντελεστή ρ του Spearman, όταν αληθεύει η μηδενική υπόθεση της ανεξαρτησίας μεταξύ των μεταβλητών X και Y. Ενα άλλο πλεονέκτημα του συντελεστή τ του Kendall βρίσκεται στο γεγονός ότι μπορεί άμεσα και απλά να ερμηνευθεί μέσω των πιθανοτήτων με τις οποίες παρατηρούμε *εναρμονισμένα ή συσχετισμένα (concordant)* ζεύγη τιμών και *μη εναρμονισμένα ή μη συσχετισμένα (discordant)* ζεύγη τιμών, όπως αυτά ορίζονται στην συνέχεια.

Τα δεδομένα αποτελούνται από ένα διμεταβλητό τυχαίο δείγμα μεγέθους n παρατηρήσεων (X_i, Y_i), $i = 1, 2, ..., n$, πάνω στο τυχαίο διάνυσμα (X, Y).

Ορισμός: Δύο παρατηρήσεις, έστω (X_j, Y_j) και (X_k, Y_k), ονομάζονται *εναρμονισμένες ή συσχετισμένες (concordant),* αν και τα δύο μέλη της μίας παρατήρησης είναι μεγαλύτερα (ή μικρότερα) από τα αντίστοιχα μέλη της άλλης παρατήρησης. Δηλαδή, αν $X_j > X_k$ (αντίστοιχα, $X_j < X_k$), τότε $Y_j > Y_k$ (αντίστοιχα, $Y_j < Y_k$).

Οι παρατηρήσεις (X_j, Y_j) και (X_k, Y_k) θα ονομάζονται *μη εναρμονισμένες ή μη συσχετισμένες (discordant)*, αν η διάταξη των πρώτων μελών τους είναι αντίθετη από την διάταξη των δεύτερων μελών τους, δηλαδή, αν $X_j > X_k$ (αντίστοιχα, $X_j < X_k$), τότε $Y_j < Y_k$ (αντίστοιχα, $Y_j > Y_k$).

Ισοδύναμα, δύο ζεύγη παρατηρήσεων (X_j, Y_j) και (X_k, Y_k) θα ονομάζονται *εναρμονισμένα* αν οι διαφορές $X_j - X_k$ και $Y_j - Y_k$ έχουν το ίδο πρόσημο (αν $(X_j - X_k)(Y_j - Y_k) > 0$). Τα ζεύγη (X_j, Y_j) και (X_k, Y_k) θα ονομάζονται *μη εναρμονισμένα* αν οι διαφορές $X_j - X_k$ και $Y_j - Y_k$ έχουν αντίθετο πρόσημο (αν $(X_j - X_k)(Y_j - Y_k) < 0$).

Έστω N_c και N_d οι αριθμοί των εναρμονισμένων και μη εναρμονισμένων ζευγών παρατηρήσεων, αντίστοιχα. Τα ζεύγη των παρατηρήσεων (X_j, Y_j) και (X_k, Y_k), για τα οποία ισχύει ότι $X_j = X_k$ ή/και $Y_j = Y_k$, δεν είναι ούτε εναρμονισμένα ούτε μη εναρμονισμένα. Τα ζεύγη αυτά ονομάζονται *ισοβαθμούντα (tied)*.

Εστω N_0 ο αριθμός των ισοβαθμούντων ζευγών παρατηρήσεων. Επειδή οι n παρατηρήσεις μπορούν να συνδυασθούν ανά δύο με $\binom{n}{2} = n(n-1)/2$ διαφορετικούς τρόπους, έπεται ότι $N_c + N_d + N_0 = \binom{n}{2}$.

Τα δεδομένα μπορούν, επίσης, να αποτελούνται από μη αριθμητικές παρατηρήσεις, οι οποίες εμφανίζονται κατά n ζεύγη, με την προϋπόθεση ότι οι παρατηρήσεις αυτές είναι τέτοιες, ώστε μπορούν να ορισθούν εναρμονισμένα και μη εναρμονισμένα ζεύγη παρατηρήσεων και να είναι δυνατός ο υπολογισμός των αριθμών N_c και N_d.

Το μέτρο συσχέτισης που προτάθηκε από τον Kendall το 1938 ορίζεται ως εξής:

$$\tau = \frac{N_c - N_d}{\binom{n}{2}} = \frac{N_c - N_d}{n(n-1)/2}.$$

Ο συντελεστής τ, δηλαδή, παριστάνει την διαφορά μεταξύ των ποσοστών των εναρμονισμένων και μη εναρμονισμένων ζευγών παρατηρήσεων.

Αν όλα τα ζεύγη παρατηρήσεων είναι εναρμονισμένα, τότε ο συντελεστής τ είναι ίσος με 1. Αν όλα τα ζεύγη είναι μη εναρμονισμένα, τότε η τιμή του συντελεστή τ είναι –1. Είναι, δηλαδή, οι τιμές του συντελεστή τ μεταξύ –1 και 1. Επιπλέον, ο συντελεστής τ ικανοποιεί όλες τις προϋποθέσεις που προαναφέρθηκαν.

Ο υπολογισμός του συντελεστή τ γίνεται απλούστερος, αν οι παρατηρήσεις (X_i, Y_i), i = 1, 2, ..., n διαταχθούν σε μία στήλη κατά αύξουσα τάξη μεγέθους των τιμών των παρατηρήσεων πάνω στην τυχαία μεταβλητή X. Τότε, κάθε Y τιμή χρειάζεται να συγκριθεί μόνο με τις Y τιμές που είναι *"κάτω"* από αυτήν. Έτσι, κάθε ζεύγος παρατηρήσεων εξετάζεται μόνο μία φορά και ο αριθμός των συσχετισμένων και μη συσχετισμένων ζευγών προσδιορίζεται γρηγορότερα.

Παράδειγμα 3.4.3: Ας θεωρήσουμε τα δεδομένα πάνω στην επιθετικότητα των διδύμων. Διατάσσοντας τις παρατηρήσεις (X_i, Y_i), i = 1, 2, ..., n κατά αύξουσα τάξη μεγέθους των τιμών των παρατηρήσεων X_i, i = 1, 2, ..., n, καταλήγουμε στον εξής πίνακα:

	$(X^{(i)},\ Y_i^*)$	Εναρμονισμένα ζεύγη κάτω από το $(X^{(i)},\ Y_i^*)$	Μη εναρμονισμένα ζεύγη κάτω από το $(X^{(i)},\ Y_i^*)$
	(68, 64)	11	0
	(70, 65)	9	0
ισοβαθμία	(71, 77)	4	4
	(71, 80)	4	4
	(72, 72)	5	1
ισοβαθμία	(77, 65)	5	0
	(77, 76)	4	1
	(86, 88)	2	2
	(87, 72)	3	0
	(88, 81)	2	0
ισοβαθμία	(91, 90)	0	0
	(91, 96)	0	0
	Σύνολο	$N_c = 49$	$N_d = 12$

Εδώ $X^{(i)}$, i = 1,…, n είναι η διατεταγμένη ακολουθία των παρατηρήσεων X_i, και Y_i^*, i = 1, …, n η προκύπτουσα αναδιάταξη των αντιστοιχουσών σ' αυτές τιμών των Y_i. Η δεύτερη στήλη του πίνακα δίνει, τον αριθμό των ζευγών $(X^{(i+1)},\ Y_{i+1}^*)$ για τα οποία $Y_i^* < Y_{i+1}^*$ όταν $X^{(i)}<X^{(i+1)}$, (i = 1, …, n-1). Η τρίτη στήλη δίνει τον αριθμό των ζευγών $(X^{(i+1)},\ Y_{i+1}^*)$ για τα οποία $Y_i^* > Y_{i+1}^*$ όταν $X^{(i)}< X^{(i+1)}$ (i = 1, …, n-1).

Με βάση τα στοιχεία του πίνακα, προκύπτει ότι η τιμή του συντελεστή συσχέτισης τ του Kendall είναι

$$\tau = \frac{N_c - N_d}{n(n-1)/2} = \frac{49-12}{(12)(11)/2} = 0.5606.$$

Υπάρχει, επομένως, θετική συσχέτιση τάξης μεγέθους μεταξύ των μετρήσεων της επιθετικότητας των διδύμων, όπως προκύπτει από την μέτρηση του συντελεστή συσχέτισης τ του Kendall.

Ο συντελεστής τ μπορεί, επίσης, να χρησιμοποιηθεί ως ελεγχοσυνάρτηση για τον έλεγχο της μηδενικής υπόθεσης της ανεξαρτησίας μεταξύ των τυχαίων μεταβλητών X και Y, με αμφίπλευρες ή μονόπλευρες εναλλακτικές, όπως εξάλλου και στην περίπτωση του συντελεστή συσχέτισης ρ του Spearman. Περισσότερο συχνή, όμως, είναι η χρήση της διαφοράς $N_c - N_d$ ως ελεγχοσυνάρτησης για τον έλεγχο των υποθέσεων αυτών. Χρησιμοποιούμε, δηλαδή, ως ελεγχοσυνάρτηση την στατιστική συνάρτηση

$$T = N_c - N_d,$$

την οποία ονομάζουμε *ελεγχοσυνάρτηση του Kendall*.

Τα ποσοστιαία σημεία της κατανομής της στατιστικής συνάρτησης T δίνονται από τον πίνακα 12 του παραρτήματος. Είναι προφανές ότι μεγάλες τιμές της στατιστικής συνάρτησης T αποτελούν ένδειξη εναντίον της υπόθεσης H_0 και υπέρ της μονόπλευρης εναλλακτικής υπόθεσης θετικής συσχέτισης μεταξύ των μεταβλητών X και Y. Αντίστοιχα, μικρές τιμές της στατιστικής συνάρτησης T αποτελούν ένδειξη εναντίον της υπόθεσης H_0 και υπέρ της μονόπλευρης εναλλακτικής υπόθεσης αρνητικής συσχέτισης μεταξύ των μεταβλητών X και Y. Επομένως, θεωρώντας την ταξινόμηση των ζευγών υποθέσεων που θα μπορούσαν να μας ενδιαφέρουν στο πλαίσιο του παρόντος προβλήματος ή παρόμοιων προβλημάτων, στις κατηγορίες

Α. (Αμφίπλευρος έλεγχος συσχέτισης)

Β. (Μονόπλευρος έλεγχος θετικής συσχέτισης)

Γ. (Μονόπλευρος έλεγχος αρνητικής συσχέτισης),

ο κανόνας απόφασης διαμορφώνεται ως εξής:

Α. Η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν

$T > w_{1-\alpha/2}$ ή αν $T < w_{\alpha/2} = -w_{1-\alpha/2}$.

Β. Η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν

$T > w_{1-\alpha}$.

Γ. Η υπόθεση H_0 απορρίπτεται σε επίπεδο σημαντικότητας α, αν

$T < w_{\alpha} = -w_{1-\alpha}$.

Επιστρέφοντας στο παράδειγμά μας, ας υποθέσουμε ότι ενδιαφερόμαστε να ελέγξουμε την υπόθεση H_0: οι μεταβλητές X και Y είναι ασυσχέτιστες, εναντίον της εναλλακτικής H_1: οι μεταβλητές X και Y είναι συσχετισμένες (αμφίπλευρος έλεγχος).

Από τον πίνακα που κατασκευάσθηκε παραπάνω, έχουμε ότι η παρατηρούμενη τιμή της στατιστικής συνάρτησης T είναι

$$N_c - N_d = 49-12 = 37.$$

Από τον σχετικό πίνακα του παραρτήματος, προκύπτει ότι τα ποσοστιαία σημεία για έναν αμφίπλευρο έλεγχο μεγέθους α=0.05, για n=12, είναι $w_{0.975} = 28$ και $w_{0.025} = -w_{0.975} = -28$. Επομένως, η τιμή της στατιστικής συνάρτησης T υπερβαίνει το 0.975-ποσοστιαίο σημείο της κατανομής της και, κατά συνέπεια, η μηδενική υπόθεση ότι δεν υπάρχει συσχέτιση μεταξύ των X και Y απορρίπτεται σε επίπεδο σημαντικότητας 0.05. Το κρίσιμο επίπεδο του ελέγχου αυτού είναι, όπως φαίνεται από τον σχετικό πίνακα του παραρτήματος, περίπου ίσο με

$$\hat{\alpha} \cong 2(0.005) = 0.01.$$

Λύση με το MINITAB: Το MINITAB δεν δίνει την δυνατότητα ελέγχου της υπόθεσης ύπαρξης συσχέτισης με βάση τον συντελεστή συσχέτισης τ του Kendall και δεν προσφέρεται για την διεξαγωγή του με έμμεσο υπολογισμό της τιμής του τ με εύκολο τρόπο.

Λύση με το SPSS: Ο συντελεστής συσχέτισης του Kendall είναι διαθέσιμος από το ίδιο πλαίσιο που χρησιμοποιήσαμε για τον συντελεστή συσχέτισης του Spearman. Ο υπολογισμός του από το SPSS δίνει:

Correlations

			Πρωτότοκοι	Δευτερότοκοι
Kendall's tau_b	Πρωτότοκοι	Correlation Coefficient	1.000	.583**
		Sig. (2-tailed)	.	.010
		N	12	12
	Δευτερότοκοι	Correlation Coefficient	.583**	1.000
		Sig. (2-tailed)	.010	.
		N	12	12

* Correlation is significant at the .05 level (2-tailed).

Ο συντελεστής συσχέτισης είναι 0.583 και η τιμή του κρίσιμου επιπέδου για αμφίπλευρο έλεγχο είναι 0.01. Όπως δηλώνει το πρόγραμμα, η τιμή που βρήκαμε είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας α=0.05 και, επομένως, η υπόθεση της έλλειψης συσχέτισης μεταξύ X και Y δεν είναι εύλογη σε επίπεδο σημαντικότητας 0.05.

Σημειώνεται ότι και στην περίπτωση τουν συντελεστή τ, το SPSS δεν παρέχει την τιμή της ελεγχοσυνάρτησης T.

Λύση με το SAS: Οι εντολές που εισάγουμε στο παράθυρο εντολών είναι:

```
data d351;
input x y @@;
cards;
```

```
86 88 71 77 77 76 68 64 91 96 72 72 77 65 91 90 70 65 71 80 88 81 87 72
;
run;
proc print;
run;
proc corr kendall;
var x y;
run;
```

Το αποτέλεσμα που παρέχει το πακέτο δίνεται στον παρακάτω πίνακα:

```
Kendall Tau b Correlation Coefficients / Prob > |R| under Ho: Rho=0/N=12
                                               X                  Y

                            X             1.00000            0.58270
                                           0.0                0.0103

                            Y             0.58270            1.00000
                                           0.0103             0.0
```

Ο συντελεστής συσχέτισης υπολογίσθηκε ίσος με 0.5827, με παρατηρούμενο επίπεδο στατιστικής σημαντικότητας ίσο με 0.0103. Η μηδενική υπόθεση που ελέγχεται δηλώνεται από το πακέτο στον τίτλο του πίνακα αποτελεσμάτων (**Prob > |R| under Ho: Rho=0/N=12**).

Παρατήρηση: Οπως ήδη είδαμε, ο συντελεστής συσχέτισης ρ του Spearman, για τα ίδια δεδομένα, είχε μία τιμή μεγαλύτερη από αυτή του συντελεστή τ του Kendall ($\rho = 0.7378 > 0.5606 = \tau$). Βέβαια, οι δύο έλεγχοι των υποθέσεων οι οποίοι έγιναν χρησιμοποιώντας τους συντελεστές αυτούς (ή ισοδύναμες με αυτούς στατιστικές συναρτήσεις) ως ελεγχοσυναρτήσεις, οδήγησαν στα ίδια σχεδόν αποτελέσματα. Εν γένει, ο συντελεστής ρ του Spearman τείνει να είναι μεγαλύτερος κατά απόλυτη τιμή από τον συντελεστή τ του Kendall. Παρ' όλα αυτά, όσο αφορά τους ελέγχους υποθέσεων, δεν υπάρχουν επαρκείς λόγοι για τους οποίους πρέπει ο ένας έλεγχος να προτιμάται έναντι του άλλου.

Μία άλλη χρήση του συντελεστή συσχέτισης ρ του Spearman προτάθηκε από τον Daniels το 1950. Συγκεκριμένα, ο Daniels πρότεινε την χρήση του συντελεστή ρ για τον έλεγχο ύπαρξης τάσης σε ένα σύνολο δεδομένων $X_1, X_2, ..., X_n$, θεωρώντας τα ζεύγη (X_i, Y_i), $i = 1, 2, ..., n$, όπου Y_i είναι ο χρόνος κατά τον οποίο έγινε η μέτρηση X_i (ή η χρονική σειρά με την οποία η μέτρηση αυτή έγινε). Οι μεταβλητές X_i , $i = 1, 2, ..., n$ είναι ανεξάρτητες. Η μηδενική υπόθεση, στην περίπτωση αυτή, είναι ότι οι μεταβλητές X_i, $i = 1, 2, ..., n$ είναι ισόνομες. Η εναλλακτική υπόθεση είναι ότι η κατανομή των μεταβλητών X_i, $i = 1, 2, ..., n$ σχετίζεται με τον χρόνο, με την έννοια ότι, με την πάροδο του χρόνου, οι μετρήσεις X_i, $i = 1, 2, ..., n$ τείνουν να γίνονται μεγαλύτερες (αντίστοιχα μικρότερες). Ο συντελεστής τ του Kendall μπορεί επίσης να χρησιμοποιηθεί για τον έλεγχο τέτοιων υποθέσεων.

Ελεγχοι ύπαρξης τάσης βασιζόμενοι στον συντελεστή συσχέτισης ρ ή τ θεωρούνται εν γένει περισσότερο ισχυροί από τον έλεγχο των Cox και Stuart. (Οπως ο Stuart (1956) παρατηρεί, η ασυμπτωτική σχετική αποτελεσματικότητα του ελέγχου των Cox και Stuart, όταν αυτός εφαρμόζεται σε τυχαίες μεταβλητές που είναι γνωστό ότι κατανέμονται κανονικά, είναι περίπου 0.78, ενώ οι αντίστοιχες τιμές της ασυμπτωτικής σχετικής αποτελεσματικότητας για τους συντελεστές ρ και τ είναι περίπου 0.98 κάτω από τις ίδιες συνθήκες). Παρά το γεγονός αυτό, οι έλεγχοι αυτοί δεν χρησιμοποιούνται το ίδιο ευρέως όπως ο έλεγχος των Cox και Stuart.

Στην συνέχεια, δίνεται ένα παράδειγμα όπου ο συντελεστής ρ του Spearman και ο συντελεστής τ του Kendall μπορούν να χρησιμοποιηθούν για τον έλεγχο της υπόθεσης ύπαρξης τάσης στους όρους μιας ακολουθίας παρατηρήσεων πάνω σε μια τυχαία μεταβλητή X.

Παράδειγμα 3.4.4: Ας υποθέσουμε ότι η συνολική ετήσια βροχόπτωση (σε εκατοστόμετρα) για 19 χρόνια, από το 1950 έως το 1968, μιας συγκεκριμένης περιοχής ήταν η εξής:

45.25, 45.83, 41.77, 36.26, 45.37, 52.25, 35.37,
57.16, 35.37, 58.32, 41.05, 33.72, 45.73, 37.90,
41.72, 36.07, 49.83, 36.24, 39.90.

Με βάση τα στοιχεία αυτά, να ελεγχθεί η υπόθεση ότι η ποσότητα βροχής τείνει να ελαττώνει ή να αυξάνει με την πάροδο του χρόνου.

Λύση: Θεωρούμε τα ζεύγη (X_i, Y_i), $i = 1, 2, ..., 19$, όπου Y_i είναι η i τιμή του διανύσματος των ετών από το 1950 έως το 1968. Παρατηρούμε ότι τα ζεύγη των παρατηρήσεων είναι ήδη διατεταγμένα κατά αύξουσα σειρά μεγέθους των τιμών Y_i, $i = 1, 2, ..., 19$. Επομένως, για τον υπολογισμό των στατιστικών συναρτήσεων ρ και τ ή των ισοδυνάμων με αυτές στατιστικώνσυναρτήσεων $T = \sum_{i=1}^{19}\left[R(X_i) - R(Y_i)\right]^2$ και $N_c - N_d$, αντίστοιχα, κατασκευάζουμε τον πίνακα που ακολουθεί.

X_i	Y_i	$R(X_i)$	$R(Y_i)$	$\left[R(X_i) - R(Y_i)\right]^2$	$N_c^{(i)}$	$N_d^{(i)}$
45.25	1950	12	1	121	7	11
45.83	1951	15	2	169	4	13
41.77	1952	11	3	64	6	10
36.26	1953	6	4	4	10	5
45.27	1954	13	5	64	5	9
52.25	1955	17	6	121	2	11

35.37	1956	2.5	7	20.25	10	11
57.16	1957	18	8	100	1	10
35.37	1958	2.5	9	42.25	9	1
58.32	1959	19	10	81	0	9
41.05	1960	9	11	4	3	5
33.72	1961	1	12	121	7	0
45.73	1962	14	13	1	1	5
37.90	1963	7	14	49	3	2
41.72	1964	10	15	25	1	3
36.07	1965	4	16	144	3	0
49.83	1966	16	17	1	0	2
36.24	1967	5	18	169	1	0
39.90	1968	8	19	121	0	0
				Σύνολο T=1421.5	N_c=73	N_d=97

($N_c^{(i)}$ και $N_d^{(i)}$ παριστάνουν τους αριθμούς των εναρμονισμένων και μη εναρμονισμένων ζευγών, αντίστοιχα, κάτω από το ζεύγος (X_i, Y_i)).

Με βάση τα στοιχεία του πίνακα αυτού, προκύπτει ότι ο κατά Spearman αμφίδρομος έλεγχος οδηγεί στην μη απόρριψη της μηδενικής υπόθεσης ότι δεν υπάρχει τάση αύξησης της βροχόπτωσης με την πάροδο του χρόνου σε επίπεδο σημαντικότητας 0.05. Πράγματι, η παρατηρούμενη τιμή της στατιστικής συνάρτησης

$$T = \sum_{i=1}^{19}\left[R(X_i) - R(Y_i)\right]^2$$

είναι ίση με 1421.5. Αυτή δεν είναι μικρότερη από την τιμή $w_{0.025}$ = 618 του 0.025-ποσοστιαίου σημείου της κατανομής της T, όπως και δεν υπερβαίνει την τιμή $w_{0.975} = \frac{1}{3}\,19(361-1) - 618 = 1662$ του 0.975-ποσοστιαίου σημείου της κατανομής της, όπως αυτά προσδιορίζονται από τον πίνακα 11 του παραρτήματος για n=19. Στο ίδιο αποτέλεσμα, δηλαδή, στην μη απόρριψη της μηδενικής υπόθεσης σε επίπεδο σημαντικότητας 5%, οδηγεί και ο κατά Kendall αμφίδρομος έλεγχος, αφού η παρατηρηθείσα τιμή $\tau = -24$ της στατιστικής συνάρτησης $T = N_c - N_d$ δεν είναι μικρότερη από την τιμή $w_{0.025} = -w_{0.975} = -55$ του 0.025-ποσοστιαίου σημείου της κατανομής της, όπως αυτό προσδιορίζεται από τον πίνακα 12 του παραρτήματος, για n=19.

Λύση με το MINITAB: Το MINITAB δεν δίνει την δυνατότητα διεξαγωγής του ελέγχου ύπαρξης τάσης με βάση τους συντελεστές Kendall και Spearman ή των ισοδυνάμων με αυτούς στατιστικών συναρτήσεων που αναφέρθησαν παραπάνω. Μπορούμε μόνο να χρησιμοποιήσουμε το πακέτο για τον υπολογισμό της στατιστικής συνάρτησης $T = \sum_i [R(X_i) - R(Y_i)]^2$ στην περίπτωση του συντελεστή Spearman. (Οπως προαναφέρθηκε, το MINITAB δεν προσφέρεται γαι τον υπολογισμό των τιμών N_c και N_d). Για τον υπολογισμό της τιμής της Τ, καταχωρίζουμε στην στήλη **C1** το δείγμα και στην στήλη **C2** τις τιμές 1 εώς 19 κατ' αύξουσα σειρά. (Δεν είναι απαραίτητο να καταχωρίσουμε τις χρονολογίες αφού και αυτές θα αντικατασταθούν με τις τάξεις μεγέθους τους. Στην στήλη **C3**, καταχωρίζουμε τις τάξεις μεγέθους των παρατηρήσεων του δείγματος, και, στην στήλη **C4**, καταχωρίζουμε τα τετράγωνα των διαφορών των στοιχείων των στηλών **C2** και **C3**. Το άθροισμα των τιμών της στήλης **C4** παρέχει την τιμή της Τ. Σύγκριση της τιμής αυτής με τα ποσοστιαία σημεία που δίνει ο πίνακας 11 του παραρτήματος μας οδηγεί στο συμπέρασμα ότι η μηδενική υπόθεση μή ύπαρξης τάσης μπορεί να θεωρηθεί εύλογη σε επίπεδο σημαντικότητας 0.05.

Λύση με το SPSS: Για τον έλεγχο ύπαρξης τάσης, καταχωρίζουμε σε μία μεταβλητή (έστω **X**) το δείγμα και σε μία άλλη (έστω **Y**) τις τιμές από 1 έως 19 σε αύξουσα σειρά. (Δεν είναι απαραίτητο να καταχωρίσουμε τις ακριβείς χρονολογίες αφού μόνο η τάξη μεγέθους τους χρησιμεύει για τον έλεγχο). Ζητώντας τον υπολογισμό του συντελεστή συσχέτισης τ του Kendall και ρ του Spearman, παίρνουμε τα εξής αποτελέσματα:

Correlations

			X	Y
Kendall's tau_b	X	Correlation Coefficient	1.000	-.141
		Sig. (2-tailed)	.	.401
		N	19	19
	Y	Correlation Coefficient	-.141	1.000
		Sig. (2-tailed)	.401	.
		N	19	19
Spearman's rho	X	Correlation Coefficient	1.000	-.247
		Sig. (2-tailed)	.	.307
		N	19	19
	Y	Correlation Coefficient	-.247	1.000
		Sig. (2-tailed)	.307	.
		N	19	19

Οι τιμές των συντελεστών συσχέτισης ρ του Spearman και τ του Kendall είναι –0.247 και –0.141 αντίστοιχα. (Οι τιμές αυτές δεν είναι στατιστικά σημαντικές όπως προκύπτει από τις αντίστοιχες τιμές των κρίσιμων επιπέδων. Κατά συνέπεια, η υπόθεση ότι δεν υπάρχει τάση στις τιμές της X μπορεί να θεωρηθεί εύλογη.

Λύση με το SAS: Προκειμένου να υπολογισθούν οι συντελεστές συσχέτισης κατά Spearman και κατά Kendall και να διεξαχθούν οι σχετικοί έλεγχοι, πληκτρολογούμε τις παρακάτω εντολές

```
data d352;
input x y @@;
cards;
45.25 1950 45.83 1951 41.77 1952
36.26 1953 45.27 1954 52.25 1955
35.37 1956 57.16 1957 35.37 1958
58.32 1959 41.05 1960 33.72 1961
45.73 1962 37.90 1963 41.72 1964
36.07 1965 49.83 1966 36.24 1967
39.90 1968
;
run;
proc print;
run;
proc corr spearman kendall;
var x y;
run;
```

Το αποτέλεσμα περιέχεται στον πίνακα που ακολουθεί

```
Spearman Correlation Coefficients / Prob > |R| under Ho: Rho=0
/ N = 19

                                  X                  Y

             X              1.00000           -0.24748
                             0.0               0.3070

             Y             -0.24748            1.00000
                             0.3070             0.0

Kendall Tau b Correlation Coefficients / Prob > |R| under Ho: Rho=0
/ N = 19

                                  X                  Y

             X              1.00000           -0.14076
                             0.0               0.4008

             Y             -0.14076            1.00000
                             0.4008             0.0
```

Παρατήρηση: Η ακριβής κατανομή των στατιστικών συναρτήσεων ρ και τ είναι εύκολο να προσδιορισθεί, αν και, στην πράξη, η διαδικασία είναι πολύ χρονοβόρα ακόμα και για μέτριο μέγεθος δείγματος n. Και οι δύο κατανομές μπορούν να προσδιορισθούν κάτω από την υπόθεση ότι οι μεταβλητές X_i και Y_i είναι ανεξάρτητες και ισόνομες. Τότε, οι n! διατάξεις των βαθμών (τάξεων μεγέθους) των μεταβλητών X_i, θεωρουμένων κατά ζεύγη με τους αντίστοιχους βαθμούς των μεταβλητών Y_i, είναι ισοπίθανες. Οι συναρτήσεις κατανομής προκύπτουν με απλή απαρίθμηση των διατάξεων, οι οποίες οδηγούν σε μία συγκεκριμένη τιμή του ρ ή του τ και με διαίρεση του αριθμού αυτών των διατάξεων με n! για να προσδιορισθεί η πιθανότητα της συγκεκριμένης τιμής του ρ ή του τ.

Επειδή, τόσο η στατιστική συνάρτηση ρ όσο και η τ αποτελούν αθροίσματα τυχαίων μεταβλητών, μπορεί κανείς να χρησιμοποιήσει μια μορφή του κεντρικού οριακού θεωρήματος για να προσεγγίσει τις κατανομές τους στην περίπτωση μεγάλων δειγμάτων. Και οι δύο συντελεστές έχουν συμμετρικές κατανομές γύρω από το μηδέν και, επομένως, έχουν και οι δύο μέση τιμή ίση με το μηδέν. Οι διασπορές των στατιστικών αυτών συναρτήσεων είναι δυσκολότερο να προσδιορισθούν. Διαίρεση, επομένως, των στατιστικών συναρτήσεων ρ και τ με τις αντίστοιχες διασπορές τους οδηγεί σε τυχαίες μεταβλητές, οι οποίες κατά προσέγγιση έχουν την κανονική κατανομή για μεγάλες τιμές του n. Όπως αναφέρθηκε και προηγουμένως, η προσέγγιση αυτή είναι καλύτερη στην περίπτωση της στατιστικής συνάρτησης τ για $n \geq 8$. Δεν είναι, όμως, εξ ίσου ικανοποιητική όταν χρησιμοποιείται για τον κατά προσέγγιση προσδιορισμό των ποσοστιαίων σημείων της κατανομής της στατιστικής συνάρτησης ρ. Στην περίπτωση που τα ζεύγη (X_i, Y_i), $i = 1, 2, ..., n$ είναι ανεξάρτητες και ισόνομες διδιάστατες κανονικές μεταβλητές, και οι δύο συντελεστές έχουν ασυμπτωτική σχετική αποτελεσματικότητα ίση με $9/\pi^2 = 0.912$, σε σχέση με τον παραμετρικό έλεγχο που χρησιμοποιεί τον συντελεστή r του Pearson ως στατιστική συνάρτηση (Stuart, 1954).

3.4.3 Συντελεστής Μερικής Συσχέτισης του Kendall *(Kendall' s Partial Correlation Coefficient)*

Συχνά, η ύπαρξη συσχέτισης μεταξύ δύο μεταβλητών οφείλεται σε συσχέτιση κάθε μιας από τις δύο μεταβλητές με μια τρίτη μεταβλητή. Για παράδειγμα, μεταξύ μαθητών σχολείου διαφορετικών ηλικιών, μπορεί κανείς να βρει μια υψηλή τιμή συντελεστή συσχέτισης μεταξύ του ύψους των μαθητών και του μεγέθους του λεξιλογίου τους. Η συσχέτιση αυτή μπορεί να μην αντανακλά κάποια γνήσια ή άμεση σχέση μεταξύ των δύο αυτών μεταβλητών, αλλά ενδέχεται να προκύπτει από το γεγονός ότι και οι δύο μεταβλητές (το μέγεθος του λεξιλογίου και το ύψος) σχετίζονται με μία τρίτη μεταβλητή, την ηλικία.

Στατιστικά, το πρόβλημα αυτό μπορεί να αντιμετωπισθεί με μεθόδους *μερικής συσχέτισης*. Η μερική συσχέτιση, είναι ένας τρόπος ανάλυσης, ο οποίος απομακρύνει την επίδραση της μεταβλητότητας που προκαλείται από μία τρίτη μεταβλητή πάνω στην σχέση που υφίσταται μεταξύ δύο μεταβλητών Χ και Υ. Με απλά λόγια, η συσχέτιση μεταξύ των μεταβλητών Χ και Υ προσδιορίζεται κρατώντας την τρίτη μεταβλητή σταθερή. Για παράδειγμα, ενδέχεται να ενδιαφερόμαστε να μελετήσουμε την σχέση μεταξύ της ικανότητας απομνημόνευσης και της ικανότητας επίλυσης διαφόρων ειδών προβλημάτων. Και οι δύο αυτές ικανότητες μπορεί να σχετίζονται με την νοημοσύνη. Επομένως για να προσδιορίσουμε την άμεση σχέση ικανότητας απομνημόνευσης και ικανότητας επίλυσης προβλημάτων, θα πρέπει να ελεγχθεί η επίδραση των διαφορών που υπάρχουν στην νοημοσύνη μεταξύ των διαφόρων ατόμων. Ένας τρόπος για να γίνει αυτό είναι να διαλέξουμε άτομα τα οποία έχουν περίπου την ίδια νοημοσύνη. Αν, όμως, αυτό δεν είναι δυνατόν, τότε μπορούμε να χρησιμοποιήσουμε στατιστικές μεθόδους ελέγχου. Τέτοιες μέθοδοι βασίζονται σε τεχνικές μερικής συσχέτισης, με τις οποίες μπορούμε να σταθεροποιήσουμε την επίδραση της νοημοσύνης πάνω στην σχέση μεταξύ της ικανότητας απομνημόνευσης και την ικανότητας επίλυσης προβλημάτων και να προσδιορίσουμε, έτσι, την έκταση της άμεσης σχέσης μεταξύ των δύο αυτών ικανοτήτων.

Στην συνέχεια, εξετάζεται μία μέθοδος *στατιστικού ελέγχου* της επίδρασης που μπορεί να έχει πάνω στην σχέση δύο μεταβλητών μία τρίτη μεταβλητή. Η μέθοδος αυτή βασίζεται στην ανάπτυξη μιας στατιστικής συνάρτησης, η οποία επιτρέπει την μελέτη της συσχέτισης μεταξύ μεταβλητών, όταν άλλες μεταβλητές κρατούνται σταθερές. Η στατιστική αυτή συνάρτηση είναι γνωστή ως *συντελεστής μερικής συσχέτισης του Kendall*.

Αν διαπιστωθεί ότι, κρατώντας μία τρίτη μεταβλητή σταθερή, μειώνεται η συσχέτιση μεταξύ δύο μεταβλητών, τότε μπορούμε να συμπεράνουμε ότι η εξάρτησή τους προκύπτει, εν μέρει, μέσω της συσχέτισής τους με αυτή την τρίτη μεταβλητή. Αν η μερική συσχέτιση είναι μηδέν, μπορούμε να συμπεράνουμε ότι η εξάρτηση των δύο μεταβλητών οφείλεται εξ ολοκλήρου στην συσχέτισή τους με αυτή την τρίτη μεταβλητή. Αντίστροφα, αν η μερική συσχέτιση είναι μεγαλύτερη από την συσχέτιση μεταξύ των δύο μεταβλητών, μπορούμε να συμπεράνουμε ότι η τρίτη μεταβλητή επισκιάζει την πραγματική συσχέτιση μεταξύ των μεταβλητών. Πρέπει, όμως, να θυμόμαστε πάντα ότι δεν μπορούμε να υποθέσουμε οποιασδήποτε μορφής σχέση αιτίου-αιτιατού. Οποιοσδήποτε ισχυρισμός της μορφής αυτής θα πρέπει να στηρίζεται σε επιπλέον πληροφορίες. Ας υποθέσουμε ότι, στον πίνακα που ακολουθεί, έχουμε τις τάξεις μεγέθους τριων μεταβλητών X, Y και Z που αναφέρονται σε τρία χαρακτηριστικά τεσσάρων ατόμων.

Άτομο	α	β	γ	δ

R(Z)	1	2	3	4
R(X)	3	1	2	4
R(Y)	2	1	3	4

Επιθυμούμε να προσδιορίσουμε την συσχέτιση που υφίσταται μεταξύ των μεταβλητών X και Y, όταν η μεταβλητή Z κρατείται σταθερή.

Ο αριθμός των δυνατών ζευγών των τάξεων μεγέθους οποιασδήποτε από τις τρεις μεταβλητές είναι $\binom{4}{2} = 6$. Έχοντας ήδη τις τάξεις μεγέθους των τιμών της μεταβλητής Z διατεταγμένες κατά αύξουσα σειρά μεγέθους, ας θεωρήσουμε όλα τα δυνατά ζεύγη τάξεων μεγέθους των τιμών της μεταβλητής X (αντίστοιχα της μεταβλητής Y) και της μεταβλητής Z. Ας συμβολίσουμε με "+" εκείνα τα ζεύγη, στα οποία η μικρότερη τάξη μεγέθους προηγείται της μεγαλύτερης τάξης μεγέθους και με "–" εκείνα τα ζεύγη, στα οποία η μεγαλύτερη τάξη μεγέθους προηγείται της μικρότερης τάξης μεγέθους. Τότε, προκύπτει ο εξής πίνακας:

Ζεύγος	(α,β)	(α, γ)	(α, δ)	(β, γ)	(β, δ)	(γ, δ)
R(Z)	+	+	+	+	+	+
R(X)	–	–	+	+	+	+
R(Y)	–	+	+	+	+	+

Σύμφωνα με τον πίνακα αυτό, για την μεταβλητή X, το ζεύγος (α, β) είναι ένα "–" ζεύγος, γιατί οι τάξεις μεγέθους στα άτομα α και β (3 και 1, αντίστοιχα) εμφανίζονται με αντίθετη σειρά, δηλαδή η υψηλότερη τάξη μεγέθους προηγείται της χαμηλότερης.

Οι πληροφορίες που περιέχονται στον παραπάνω πίνακα μπορούν να συνοψισθούν σε έναν 2×2 πίνακα συναφείας:

Ας θεωρήσουμε πρώτα τα τρία πρόσημα που αντιστοιχούν στο ζεύγος (α, β) του παραπάνω πίνακα. Για το ζεύγος αυτό, στις μεταβλητές X και Y έχει αντιστοιχισθεί ένα –, ενώ στην μεταβλητή Z, έχει αντιστοιχισθεί ένα +. Επομένως, μπορούμε να πούμε ότι οι μεταβλητές X και Y δεν *«συμφωνούν»* με την Z. Συνεχίζοντας με τον ίδιο τρόπο, μπορούμε να ταξινομήσουμε τα ζεύγη (α,β), (α,γ), (α,δ) κ.λ.π. στα κελλιά του πίνακα που ακολουθεί.

	Ζεύγη με ομόσημες τις τάξεις R(Y) και R(Z)	Ζεύγη με ετερόσημες τις τάξεις R(Y) και R(Z)	Σύνολο
Ζεύγη με ομόσημες τις τάξεις R(X) και R(Z)	A=4	B=0	A+B=4
Ζεύγη με ετερόσημες τις τάξεις R(X) και R(Z)	C=1	D=1	C+D=2
Σύνολο	A+C=5	B+C=1	$\binom{n}{2} = \binom{4}{2} = 6$

Ο *συντελεστής μερικής συσχέτισης του* Kendall μεταξύ των μεταβλητών X και Y, όταν απομακρύνεται η έμμεση συσχέτισή τους που οφείλεται στην μεταβλητή Z, συμβολίζεται με $\tau_{xy.z}$ και ορίζεται από την σχέση

$$\tau_{xy.z} = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}.$$

Στην περίπτωση των δεδομένων του παραδείγματός μας, ο συντελεστής μερικής συσχέτισης των μεταβλητών X και Y όταν η επίδραση της μεταβλητής Z κρατείται σταθερή, είναι

$$\tau_{xy.z} = \frac{(4)\,(1) - (1)\,(0)}{\sqrt{(4)\,(2)\,(5)\,(1)}} = 0.632$$

Ας υπολογίσουμε, τώρα, τον συντελεστή συσχέτισης μεταξύ των μεταβλητών X και Y, χωρίς να λάβουμε υπόψη μας την επίδραση της μεταβλητής Z. Διατάσσοντας τα ζεύγη των τιμών $(R(X_i), R(Y_i))$, $i = 1, 2, 3, 4$ κατά αύξουσα σειρά μεγέθους των τιμών της $R(X_i)$, $i = 1, 2, 3, 4$, έχουμε τον εξής πίνακα:

$(R(X_i), R(Y_i))$	$N_c^{(i)}$	$N_d^{(i)}$
(1, 1)	3	0
(2, 3)	1	1
(3, 2)	1	0
(4,4)	0	0
	N_c=5	N_d=1

Από τον πίνακα αυτό, προκύπτει ότι η τιμή του συντελεστή συσχέτισης του Kendall μεταξύ των μεταβλητών X και Y είναι

$$\tau_{xy} = \frac{N_c - N_d}{\binom{n}{2}} = \frac{5-1}{\binom{4}{2}} = 0.667.$$

Συγκρίνοντας τις τιμές των συντελεστών τ_{xy} και $\tau_{xy.z}$, μπορούμε να συμπεράνουμε ότι οι συσχετίσεις μεταξύ των μεταβλητών X και Z και μεταξύ των μεταβλητών Y και Z μόνο ελάχιστα επηρεάζουν την παρατηρούμενη συσχέτιση μεταξύ των μεταβλητών X και Y. Η μορφή, βέβαια, αυτής της συμπερασματολογίας θα πρέπει να γίνεται με πολλή προσοχή. Ο λόγος είναι ότι ο συντελεστής μερικής συσχέτισης δεν μπορεί να χρησιμοποιηθεί ως ελεγχοσυνάρτηση για τον έλεγχο υποθέσεων ύπαρξης μερικής συσχέτισης μεταξύ δύο μεταβλητών, γιατί η δειγματική κατανομή του δεν έχει προσδιορισθεί.

Αν και η μέθοδος που ακολουθήσαμε για τον υπολογισμό του συντελεστή $\tau_{xy.z}$ είναι χρήσιμη γιατί αποκαλύπτει την φύση της μερικής συσχέτισης. Καθώς το n αυξάνει, η μέθοδος αυτή γίνεται πολύπλοκη και χρονοβόρα, κυρίως λόγω της ταχείας αύξησης της τιμής του $\binom{n}{2}$. Το 1949, ο Kendall απέδειξε ότι η τιμή του συντελεστή $\tau_{xy.z}$ είναι ίση με

$$\tau_{xy.z} = \frac{\tau_{xy} - \tau_{xz}\tau_{yz}}{\sqrt{(1-\tau_{xz}^2)\,(1-\tau_{yz}^2)}}.$$

Ο τύπος αυτός προσφέρεται περισσότερο για ταχύτερους υπολογισμούς. Έτσι, εύκολα μπορεί να διαπιστωθεί ότι ο συντελεστής $\tau_{xy.z}$ για τα στοιχεία του παραδείγματός μας προκύπτει από την παραπάνω σχέση, όπου $\tau_{xy} = 0.667$, $\tau_{yz} = 0.667$, $\tau_{xz} = 0.33$.

Η μερική συσχέτιση μπορεί να εκτιμηθεί και από τον *συντελεστή μερικής συσχέτισης του Pearson*, ο οποίος ορίζεται με αντίστοιχο τρόπο:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)\,(1-r_{yz}^2)}},$$

όπου οι συντελεστές r_{xy}, r_{xz} και r_{yz} είναι οι συνήθεις συντελεστές συσχέτισης του Pearson μεταξύ των μεταβλητών X και Y, X και Z και Y και Z, αντίστοιχα.

Ο συντελεστής ρ του Spearman, επίσης, μπορεί να επεκταθεί για να μετρά μερική συσχέτιση με τον ίδιο ακριβώς τρόπο που περιεγράφη για τον συντελεστή τ του Kendall. Ορίζεται, δηλαδή, ως συντελεστής μερικής συσχέτισης του Spearman ο συντελεστής

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{(1-\rho_{xz}^2)(1-\rho_{yz}^2)}}.$$

Οι κατανομές των συντελεστών $\rho_{xy.z}$, $\tau_{xy.z}$ και $r_{xy.z}$ εξαρτώνται από την από κοινού κατανομή των μεταβλητών X, Y και Z και, επομένως, δεν μπορούν να χρησιμοποιηθούν ως ελεγχοσυναρτήσεις μη παραμετρικών ελέγχων. Εξαίρεση αποτελούν οι κατανομές των δύο πρώτων συντελεστών στην περίπτωση που οι τρεις μεταβλητές X, Y και Z είναι αμοιβαία ανεξάρτητες. Ο ενδιαφερόμενος αναγνώστης μπορεί να αναζητήσει περισσότερες πληροφορίες στα άρθρα των Simon (1977a,b), Agresti (1977), και Wolfe (1977).

Εκτιμήσεις της μερικής συσχέτισης μεταξύ δύο μεταβλητών X_1 και X_2, όταν απομακρύνεται η έμμεση συσχέτιση που οφείλεται στις μεταβλητές X_3, X_4, …, X_n, λαμβάνονται από τους ανάλογα οριζόμενους συντελεστές $\tau_{12.34\ldots n}$ (όταν χρησιμοποιείται η επέκταση του συντελεστή τ του Kendall), $r_{12.34\ldots n}$ (όταν χρησιμοποιείται η επέκταση του συντελεστή r του Pearson) ή $\rho_{12.34\ldots n}$ (όταν χρησιμοποιείται η επέκταση του συντελεστή ρ του Spearman). Ενδεικτικά, δίνουμε τον ορισμό του συντελεστή μερικής συσχέτισης του Pearson μεταξύ των μεταβλητών X_1 και X_2, όταν απομακρύνεται η επίδραση της συσχέτισης που οφείλεται στις μεταβλητές X_3, $X_4, ..., X_n$:

Ας συμβολίσουμε με r τον συντελεστή του Pearson για την εκτίμηση του συντελεστή συσχέτισης μεταξύ δύο μεταβλητών X_i, X_j, $i, j = 1, 2, ..., n$. Δηλαδή,

$$r_{ij} = \begin{cases} 1, & \text{αν } i = j \\ \dfrac{\sum_{k=1}^{m}\left(X_{ik} - \overline{X}_i\right)\left(X_{jk} - \overline{X}_j\right)}{\left[\sum_{k=1}^{m}\left(X_{ik} - \overline{X}_i\right)^2 \sum_{k=1}^{m}\left(X_{jk} - \overline{X}_j\right)^2\right]^{1/2}}, & \text{αν } i \neq j \end{cases}$$

όπου X_{ik} είναι η k παρατήρηση πάνω στην μεταβλητή X_i και $\overline{X}_i = \sum_{k=1}^{m} X_{ik}/n$, $i = 1, 2, ..., n$, $k = 1, 2, ..., m$.

Εστω

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1n} \\ . & 1 & r_{23} & \cdots & r_{2n} \\ . & . & 1 & & \\ . & . & . & \cdots & \\ r_{n1} & r_{n2} & r_{n3} & \cdots & 1 \end{bmatrix}_{nxn},$$

ο πίνακας των συντελεστών συσχέτισης του τυχαίου διανύσματος $(X_1, X_2, ..., X_n)$. Τότε, ο συντελεστής μερικής συσχέτισης του Pearson για τις μεταβλητές X_1 και X_2, όταν απομακρύνεται η επίδραση της συσχέτισης που οφείλεται στις μεταβλητές $X_3, X_4, ..., X_n$, ορίζεται από την σχέση

$$r_{12.34...n} = \frac{r_{12} - \mathbf{r}_{13}^{'}\mathbf{R}_{33}^{-1}\mathbf{r}_{23}}{\sqrt{\left(1-\mathbf{r}_{13}^{'}\mathbf{R}_{33}^{-1}\mathbf{r}_{13}\right)\left(1-\mathbf{r}_{23}^{'}\mathbf{R}_{33}^{-1}\mathbf{r}_{23}\right)}},$$

όπου $\mathbf{r}_{13}$, $\mathbf{r}_{23}$ και $\mathbf{R}_{33}$ είναι οι υποπίνακες του εξής διαμερισμού του πίνακα $\mathbf{R}$ των συντελεστών συσχετίσεως:

$$\mathbf{R} = \begin{bmatrix} \underset{2\times 2}{\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}} & \underset{2\times(n-2)}{\begin{bmatrix} r_{13}^{t} \\ r_{23}^{t} \end{bmatrix}} \\ \underset{(n-2)\times 2}{\left[\mathbf{r}_{13} \quad \mathbf{r}_{23}\right]} & \underset{(n-2)\times(n-2)}{\mathbf{R}_{33}} \end{bmatrix}$$

Δηλαδή,

$$\mathbf{r}_{13} = \begin{bmatrix} r_{13} \\ r_{14} \\ . \\ . \\ . \\ r_{1n} \end{bmatrix}, \quad \mathbf{r}_{23} = \begin{bmatrix} r_{23} \\ r_{24} \\ . \\ . \\ . \\ r_{2n} \end{bmatrix}$$

και

$$\mathbf{R}_{33} = \begin{bmatrix} 1 & r_{34} & r_{35} & .... & r_{3n} \\ & 1 & r_{45} & .... & r_{4n} \\ & & 1 & .... & r_{5n} \\ & & & . & . \\ & & & . & . \\ & & & . & . \\ & & & & 1 \end{bmatrix}.$$

Παράδειγμα 3.4.5: Σε μία διεπιστημονική μελέτη του φαινομένου της γήρανσης των ανθρώπων, η οποία έγινε το 1963, μία ομάδα υγιών ανδρών ηλικίας 65 και άνω υποβλήθηκε σε μία σειρά ψυχολογικών τεστ. Οι βαθμολογίες X_1, X_2, X_3, X_4, X_5 46 ατόμων σε πέντε από τα τεστ περιέχονται στον πίνακα 3.4.1 που ακολουθεί:

Πίνακας 3.4.1
Βαθμολογίες 46 ατόμων σε 5 ψυχολογικά τεστ.

	Βαθμολογία				
Άτομο	X_1	X_2	X_3	X_4	X_5
1	12	61	6	17	46
2	14	67	6	17	47
3	12	81	5	27	33
4	26	67	1	36	16
5	7	49	6	19	57
6	24	80	2	26	33
7	17	81	7	21	35
8	22	70	7	15	51
9	12	24	8	10	74
10	43	58	6	24	31
11	7	34	8	9	76
12	19	83	4	24	28
13	12	84	7	18	36
14	26	81	5	21	34
15	8	83	6	17	43
16	42	90	4	28	28
17	19	70	2	24	25
18	16	82	3	29	35
19	14	71	7	21	52
20	35	92	3	35	18
21	22	45	5	24	49
22	15	55	8	19	47
23	23	77	2	26	24
24	12	63	7	23	46
25	11	37	3	19	51
26	14	54	8	12	68
27	8	64	6	18	53
28	13	36	7	20	59
29	8	82	3	23	47
30	11	58	1	27	37
31	11	38	8	11	66
32	27	77	5	30	25
33	26	84	7	25	34
34	22	59	5	19	36
35	45	83	1	38	13
36	20	77	1	19	25
37	17	61	1	26	35
38	27	48	2	23	34
39	29	83	2	35	32
40	26	70	1	27	29
41	18	61	1	41	24
42	35	79	1	43	14
43	29	67	2	28	32
44	10	60	3	24	39
45	14	63	2	33	28
46	26	59	4	31	27
Μέσος	19.70	66.26	4.33	23.96	38.52
Τυπική Απόκλιση	9.7207	16.3536	2.4590	7.7458	15.0787

Ας υποθέσουμε ότι ενδιαφερόμαστε να εκτιμήσουμε τον συντελεστή μερικής συσχέτισης μεταξύ των αποδόσεων X_1 και X_2, απομακρύνοντας την επίδραση της συσχέτισης που οφείλεται στις μεταβλητές X_3, X_4, X_5.

Από τα δεδομένα, εύκολα μπορεί να διαπιστωθεί, ότι ο πίνακας των συνήθων συντελεστών συσχέτισης του Pearson για το τυχαίο διάνυσμα $(X_1, X_2, X_3, X_4, X_5)$ είναι

$$\mathbf{R} = \begin{bmatrix} 1 & -0.6659 & 0.4135 & -0.3927 & 0.5922 \\ & 1 & -0.6910 & 0.7379 & -0.8381 \\ & & 1 & -0.3299 & 0.4885 \\ & & & 1 & -0.7378 \\ & & & & 1 \end{bmatrix}.$$

Ο πίνακας αυτός, εύκολα, μπορεί να διαμερισθεί ως εξής:

$$\mathbf{R} = \begin{bmatrix} \underset{2\times 2}{\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}} & \underset{2\times 3}{\begin{bmatrix} r_{13}^t \\ r_{23}^t \end{bmatrix}} \\ \underset{3\times 2}{[\mathbf{r}_{13} \quad \mathbf{r}_{23}]} & \underset{3\times 3}{\mathbf{R}_{33}} \end{bmatrix},$$

όπου

$$r_{12} = -0.6659,$$

$$\mathbf{r}_{13} = \begin{bmatrix} 0.4135 \\ -0.3927 \\ 0.5922 \end{bmatrix},$$

$$\mathbf{r}_{23} = \begin{bmatrix} -0.6910 \\ 0.7379 \\ -0.8381 \end{bmatrix}$$

$$\mathbf{R}_{33} = \begin{bmatrix} 1 & -0.3299 & 0.4885 \\ & 1 & -0.7378 \\ & & 1 \end{bmatrix},$$

Επομένως,

$$\mathbf{r}_{13}'\mathbf{R}_{33}^{-1}\mathbf{r}_{23} = -0.5310, \quad \mathbf{r}_{13}'\mathbf{R}_{33}^{-1}\mathbf{r}_{13} = 0.3744, \quad \mathbf{r}_{23}'\mathbf{R}_{33}^{-1}\mathbf{r}_{23} = 0.8442.$$

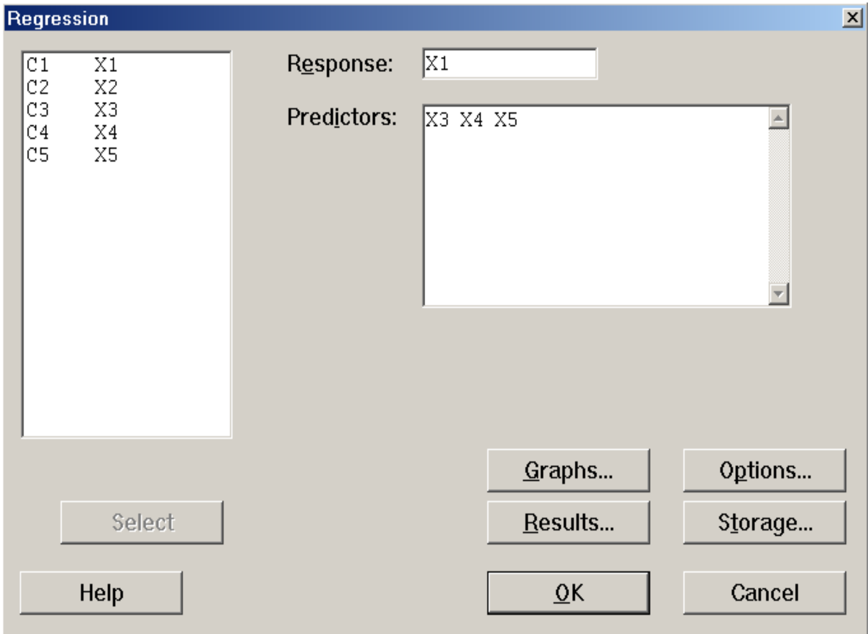
Αντικατάσταση των παραπάνω τιμών στον σχετικό τύπο, οδηγεί στην εξής εκτίμηση του συντελεστή μερικής συσχέτικης του Pearson για τις αποδόσεις X_1 και X_2, όταν οι επιδράσεις της συσχέτισης που οφείλεται στις μεταβλητές X_3, X_4 και X_5 διατηρούνται σταθερές:

$$r_{12.345} = \frac{-0.6659-(-0.5310)}{\sqrt{(1-0.3744)(1-0.8442)}} = -0.4321.$$

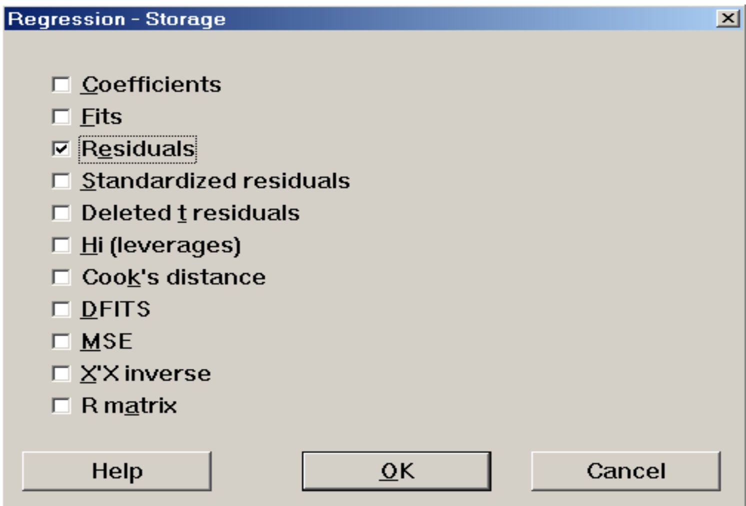
Παρατηρούμε ότι $r_{12.345} < r_{12}$, δηλαδή, η απομάκρυνση της επίδρασης της συσχέστισης που οφείλεται στις μεταβλητές X_3, X_4 και X_5 οδήγησε σε μείωση της συσχέτισης μεταξύ των μεταβλητών X_1 και X_2. Μπορούμε, επομένως, να συμπεράνουμε ότι η αλληλοεξάρτηση των μεταβλητών X_1 και X_2 οφείλεται, εν μέρει, στην επίδραση της συσχέτισης που οφείλεται στις μεταβλητές X_3, X_4 και X_5.

Λύση με το MINITAB: Ο συντελεστής συσχέτισης μεταξύ των μεταβλητών **X1** και **X2** υπολογίζεται ως συντελεστής συσχέτισης μεταξύ των καταλοίπων των παλινδρομήσεων της **X1** και της **X2** πάνω στις μεταβλητές **X3**, **X4** και **X5** ως εξής:

Καταχωρίζουμε τις μεταβλητές **X1**, …, **X5** σε πέντε στήλες (έστω **C1**, …, **C5**). Επιλέγοντας **Stat, Regression, Regression** προκύπτει το εξής πλαίσιο διαλόγου:



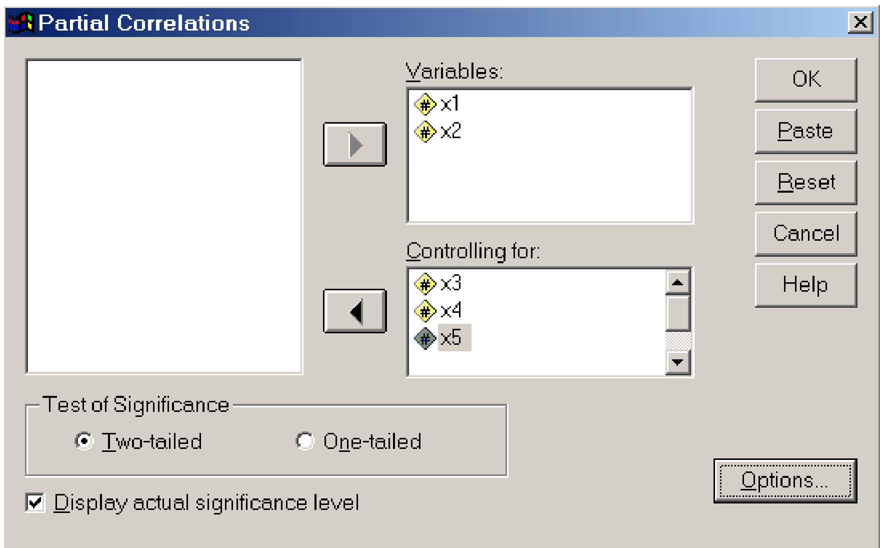
Στο πεδίο **Response**, δηλώνουμε την εξαρτημένη μεταβλητή (π.χ. **X1**) και, στο πεδίο **Predictors**, δηλώνουμε τις μεταβλητές πάνω στις οποίες θα γίνει η παλινδρόμηση (δηλαδή, **X3**, **X4**, **X5**). Για να αποθηκεύσουμε τα κατάλοιπα, πιέζουμε **Storage** και οδηγούμεθα στο παρακάτω πλαίσιο διαλόγου:



Επιλέγοντας **Residuals**, τα κατάλοιπα της παλινδρόμησης θα καταχωρισθούν στην πρώτη διαθέσιμη κενή στήλη (**C6**) με το όνομα **RESI1**. Επαναλαμβάνοντας την ίδια διαδικασία για την μεταβλητή **X2**, τα κατάλοιπα της παλινδρόμησής της πάνω στις **X3**, **X4**, **X5** καταχωρίζονται στην επόμενη κενή στήλη (**C7**) με το όνομα **RESI2**. Η τιμή του συντελεστή συσχέτισης των μεταβλητών **RESI1** και **RESI2** που προκύπτει (0.157) είναι η ζητούμενη τιμή του συντελεστή μερικής συσχέτισης μεταξύ των **X1** και **X2**.

Λύση με το SPSS: Το SPSS δεν παρέχει την δυνατότητα υπολογισμού μη παραμετρικών μερικών συντελεστών συσχέτισης. Επιτρέπει μόνο τον υπολογισμό του μερικού συντελεστή συσχέτισης του Pearson.

Για το συγκεκριμένο παράδειγμα, καταχωρίζουμε τις τιμές των πέντε τυχαίων μεταβλητών σε πέντε μεταβλητές του προγράμματος με τα ονόματα **x1**, **x2**, **x3**, **x4** και **x5**. Στην συνέχεια, επιλέγουμε **Analyze**, **Correlate**, **Partial** και οδηγούμεθα στο εξής πλαίσιο διαλόγου:



Στο πεδίο **Variables**, δηλώνουμε τις μεταβλητές των οποίων θέλουμε να υπολογίσουμε το μερικό συντελεστή συσχέτισης, ενώ, στο πεδίο **Controlling for**, δηλώνουμε τις μεταβλητές την επίδραση των οποίων θέλουμε να απομακρύνουμε. Στο πεδίο **Test of Significance**, δηλώνουμε το είδος του ελέγχου που θέλουμε (μονόπλευρο ή αμφίπλευρο έλεγχο) και έχουμε τη δυνατότητα να ζητήσουμε από το πρόγραμμα τον υπολογισμό του κρίσιμου επιπέδου του ελέγχου (πεδίο **Display actual significance level**). Τα αποτελέσματα που παίρνουμε έχουν την μορφή:

```
- - -  P A R T I A L   C O R R E L A T I O N   C O E F F
I C I E N T S  - - -

Controlling for..     X3         X4         X5

                X1          X2

X1             1.0000      -.1574
             (     0)    (    41)
             P= .        P= .314

X2             -.1574      1.0000
             (    41)    (     0)
             P= .314     P= .

(Coefficient / (D.F.) / 2-tailed Significance)
```

Η τιμή του μερικού συντελεστή συσχέτισης προκύπτει ίση με –0.1574, ενώ η τιμή του κρίσιμου επιπέδου είναι ίση με 0.314. Συνεπώς, η μηδενική υπόθεση μη ύπαρξης συσχέτισης μεταξύ **x1** και **x2** όταν η επίδραση των υπολοίπων μεταβλητών έχει αφαιρεθεί μπορεί να θεωρηθεί εύλογη.

Λύση με το SAS: Προκειμένου να υπολογισθούν οι συντελεστές συσχέτισης και μερικής συσχέτισης που επιθυμούμε (π.χ. Pearson, Spearman, Kendall), χρησιμοποιούμε τις εξής εντολές:

```
data d361;
```

```
input x1 x2 x3 x4 x5;
cards;
12      61      6      17      46
14      67      6      17      47
...
;
run;

proc corr pearson spearman kendall;
var x1 x2;
partial x3 x4 x5;
run;
proc corr pearson spearman kendall;
var x1 x2;
run;
```

Τα αποτελέσματα του πακέτου δίνονται με την εξής μορφή:

```
                             The SAS System
                                     Correlation Analysis

                        3 'PARTIAL' Variables:  X3       X4       X5
                        2 'VAR'     Variables:  X1       X2


                                      Simple Statistics


Partial   Partial
   Variable   N      Mean   Std Dev    Median   Minimum   Maximum
Variance   Std Dev

   X3         46    4.3261    2.4590    4.5000    1.0000    8.0000
.          .
   X4         46   23.9565    7.7458   24.0000    9.0000   43.0000
.          .
   X5         46   38.5217   15.0787   35.0000   13.0000   76.0000
.          .
   X1         46   19.6957    9.7202   17.5000    7.0000   45.0000
52.8239    7.2680
   X2         46   66.2609   16.3536   67.0000   24.0000   92.0000
127.3   11.2843


Pearson Partial Correlation Coefficients / Prob > |R| under Ho: Partial
Rho=0 / N = 46

                                    X1                X2

                        X1          1.00000          -0.15735
                                     0.0               0.3136

                        X2         -0.15735           1.00000
                                     0.3136            0.0



Spearman Partial Correlation Coefficients / Prob > |R| under Ho: Partial
Rho=0 / N = 46


                                    X1                X2

                        X1          1.00000          -0.10321
                                     0.0               0.5101

                        X2         -0.10321           1.00000
                                     0.5101            0.0
```

Kendall Partial Tau b Correlation Coefficients / N = 46

	X1	X2
X1	1.00000	0.07199
X2	0.07199	1.00000

The SAS System

Correlation Analysis

2 'VAR' Variables: X1 X2

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
X1	46	19.695652	9.720241	17.500000	7.000000	45.000000
X2	46	66.260870	16.353640	67.000000	24.000000	92.000000

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 46

	X1	X2
X1	1.00000 0.0	0.41347 0.0043
X2	0.41347 0.0043	1.00000 0.0

Spearman Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 46

	X1	X2
X1	1.00000 0.0	0.38027 0.0091
X2	0.38027 0.0091	1.00000 0.0

Kendall Tau b Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 46

	X1	X2
X1	1.00000 0.0	0.27959 0.0078
X2	0.27959 0.0078	1.00000 0.0

ΑΣΚΗΣΕΙΣ

1. Σε ένα καινούργιο εργάτη ανατίθεται ο χειρισμός μιας μηχανής που κατασκευάζει βίδες. Καθημερινά, ένα τυχαίο δείγμα από τις βίδες που παράγει η μηχανή εξετάζεται και καταγράφεται το ποσοστό των ελαττωματικών. Αποτελούν τα δεδομένα που ακολουθούν ένδειξη σημαντικής βελτίωσης του εργάτη αυτού στον χειρισμό της μηχανής με την πάροδο του χρόνου;

Ημέρα	Ποσοστό ελαττωματικών (%)	Ημέρα	Ποσοστό ελαττωματικών (%)
1	6.1	7	5.3
2	7.5	8	4.5
3	7.7	9	4.9
4	5.9	10	4.6
5	5.2	11	3.0
6	6.1	12	4.0
		13	3.7

α) Να χρησιμοποιηθεί ο συντελεστής συσχέτισης ρ του Spearman.

β) Να χρησιμοποιηθεί ο συντελεστής συσχέτισης τ του Kendall.

2. Ας θεωρήσουμε τα στοιχεία της άσκησης 15 του προηγούμενου κεφαλαίου με τις επιδόσεις ενός ζευγαριού στο bowling:

Γύρος	1	2	3	4	5	6	7	8	9	10
Άνδρας	147	158	131	142	183	151	196	129	155	158
Γυναίκα	122	128	125	123	115	120	108	143	124	123

α) Να υπολογισθεί ο συντελεστής συσχέτισης ρ του Spearman.

β) Να υπολογισθεί ο συντελεστής συσχέτισης τ του Kendall.

γ) Να ελεγχθεί η υπόθεση της ανεξαρτησίας των επιδόσεων των συζύγων χρησιμοποιώντας έναν αμφίδρομο έλεγχο βασισμένο στον συντελεστή ρ.

δ) Να ελεγχθεί η παραπάνω υπόθεση χρησιμοποιώντας τον συντελεστή τ ως ελεγχοσυνάρτηση.

3. Υπάρχει σημαντική συσχέτιση μεταξύ της ηλικίας, στην οποία ένας πρόεδρος των Η.Π.Α. αναλαμβάνει τα καθήκοντά του για πρώτη φορά και της ηλικίας στην οποία πεθαίνει;

Όνομα	Ηλικία Ανάληψης καθηκόντων	Ηλικία Θανάτου	Όνομα	Ηλικία Ανάληψης καθηκόντων	Ηλικία Θανάτου
Washington	57	67	Buchanan	65	77
J. Adams	61	90	Lincoln	52	56
Jefferson	57	83	A. Johnson	56	66
Madison	57	85	Grant	46	63
Monroe	58	73	Hayes	54	70
J. O. Adams	57	80	Garfield	49	49
Jackson	61	78	Arthur	50	56
Van Buren	54	79	Clevelant	47	71
Harrison	68	68	Harrison	55	67
Tyler	51	71	McKinley	54	58
Polk	49	53	T. Roosevelt	42	60
Taylor	64	65	Taft	51	72
Filmore	50	74	Wilson	56	67
Pierce	48	64	Harding	55	57
Coolidge	51	60	Eisenhower	62	78
Hoover	54	90	Kenndey	43	46
F. Roosevelt	51	63	L. Johnson	55	64
Truman	60	88			

α) Να χρησιμοποιηθεί ο συντελεστής συσχέτισης ρ του Spearman.

β) Να χρησιμοποιηθεί ο συντελεστής συσχέτισης τ του Kendall.

(Ας σημειωθεί ότι τα δεδομένα του παραπάνω πίνακα δεν αντιπροσωπεύουν ένα τυχαίο δείγμα. Θα μπορούσε, όμως, να υποτεθεί ότι αυτά συμπεριφέρονται ως ένα τυχαίο δείγμα όλων των προέδρων των Η.Π.Α. του παρελθόντος, του παρόντος και του μέλλοντος).

4. Ποια η διαφορά μεταξύ του προσημικού και του ελέγχου του Wilcoxon;

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 2000)

5. Εστω ότι οι μηνιαίες χιλιομετρικές αποστάσεις που διανύονται από τα αυτοκίνητα της εταιρείας Χ και της εταιρείας Υ δίνονται στον πίνακα που ακολουθεί.

Μήνας	Χιλιομετρικές αποστάσεις της εταιρείας Χ	Χιλιομετρικές αποστάσεις της εταιρείας Υ
Ιανουάριος	100	350
Φεβρουάριος	750	100
Μάρτιος	440	110
Απρίλιος	220	200
Μάιος	630	280
Ιούνιος	250	200
Ιούλιος	600	100
Αύγουστος	525	320
Σεπτέμβριος	800	850
Οκτώβριος	600	350
Νοέμβριος	550	100
Δεκέμβριος	620	110

Υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μέσων μηνιαίων χιλιομετρικών αποστάσεων που διανύουν τα αυτοκίνητα των δύο εταιρειών;

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 2000)

6. Ερευνητής ισχυρίζεται ότι, σε οικογένειες με 2 αγόρια, το μεγαλύτερο είναι πιο βαρύ από το μικρότερο. Για τον λόγο αυτό επελέγησαν με τυχαίο τρόπο 12 οικογένειες με δύο αγόρια ηλικίας τουλάχιστον 20 ετών (έτσι ώστε να έχει συμπληρωθεί η ανάπτυξη και των δύο παιδιών και να μην υπάρχουν διαφορές στα βάρη οφειλόμενες στην διαφορά ηλικίας). Τα βάρη που παρατηρήθηκαν περιέχονται στον παρακάτω πίνακα:

	1	2	3	4	5	6	7	8	9	10	11	12
1^o αγόρι	86	71	77	68	91	72	77	91	70	71	88	87
2^o αγόρι	88	77	76	64	96	72	65	90	65	80	81	72

Πιστεύετε ότι τα στοιχεία παρέχουν ενδείξεις υπέρ του ισχυρισμού του ερευνητή;

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 2000)

7. Ο έλεγχος Wilcoxon για δείγμα ζευγών παρατηρήσεων (πάνω στις τυχαίες μεταβλητές X και Y) ελέγχει υποθέσεις για την διάμεσο $d_{0.5}$ της τυχαίας μεταβλητής Y-X. Έστω ότι, εφαρμόζοντας τον έλεγχο αυτό σε ένα δείγμα, βρίσκουμε ότι η ελεγχοσυνάρτηση

$$T = \frac{\sum R_i}{\sqrt{\sum R_i^2}}$$

έχει τιμή τ=1.8713. Αν η μηδενική υπόθεση είναι H_0: $d_{0.5} \leq 0$, ποιά από τις ενέργειες που ακολουθούν θα ήταν εύλογη;

α) Αποδοχή της H_0 σε επίπεδο σημαντικότητας α=0.1

β) Απόρριψη της H_0 σε επίπεδο σημαντικότητας α=0.05

γ) Αποδοχή της H_0 για κάθε επίπεδο σημαντικότητας $\alpha \geq 0.05$

δ) Απόρριψη της H_0 σε επίπεδο σημαντικότητας α=0.01

(Οικ. Παν/μιο Αθηνών – Εξετ. Δεκ. 2000)

8. Τα δείγματα 16, 21 15, 18, 19, 17, 24, 21, 14, 13, 20, 22 και 12, 18, 17, 14, 18, 10, 9, 11, 19, 8, 7, 16 προέρχονται από τους πληθυσμούς τιμών των τυχαίων μεταβλητών X και Y αντίστοιχα. Να ελεγχθεί η μηδενική υπόθεση ότι τα δείγματα προέρχονται από τον ίδιο πληθυσμό ως προς την εναλλακτική υπόθεση ότι $E(X) < E(Y)$, σε επίπεδο σημαντικότητας 5%. Συγκρίνατε τον έλεγχο που χρησιμοποιείτε με τον αντίστοιχο παραμετρικό έλεγχο.

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 1998)

9. Οι παρακάτω παρατηρήσεις αναφέρονται στις αποστάσεις στις οποίες εφθάρησαν ελαστικά αυτοκινήτων τύπου A και B.

	Απόσταση (χιλ. μίλια)
Ελαστικά τύπου A:	14.8, 7.3, 5.6, 6.3, 9.0, 4.2, 10.6, 12.5, 12.9, 16.1, 11.4, 2.7
Ελαστικά τύπου B:	12.7, 14.2, 12.6, 2.1, 17.7, 11.8, 16.9, 7.9, 16.0, 10.6, 5.6, 5.6,
	7.6, 11.3, 8.3, 6.7, 3.6, 1.0, 2.4, 6.4, 9.1, 6.7, 18.6, 3.2,
	6.2, 6.1, 15.3, 10.6, 1.8, 5.9, 9.9, 10.6, 14.8, 5.0, 2.6, 4.0

α) Αποτελούν τα δεδομένα ένδειξη ότι τα ελαστικά του τύπου A τείνουν να έχουν μεγαλύτερη αντοχή από τα ελαστικά του τύπου B;

β) Ποιο είναι το κρίσιμο επίπεδο απόρριψης της μηδενικής υπόθεσης;

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 1998)

10. Ενα πείραμα έγινε για να συγκριθεί η αντοχή δύο τύπων ανακυκλώσιμου χαρτιού περιτυλίγματος με το ίδιο βάρος. Η επεξεργασία του πρώτου τύπου χαρτιού έγινε χωρίς τη χρήση χημικής ουσίας, ενώ του δεύτερου τύπου έγινε με τη χρήση κάποιας χημικής ουσίας. Από την παραγωγή, επελέγησαν τυχαία δέκα τεμάχια από κάθε τύπο χαρτιού, των οποίων οι μετρήσεις αντοχής εμφανίζονται στον παρακάτω πίνακα.

Μέτρηση αντοχής χαρτιού πρώτου τύπου	Μέτρηση αντοχής χαρτιού δευτέρου τύπου
1.21	1.49
1.43	1.37
1.35	1.67
1.51	1.50
1.39	1.31
1.17	1.29
1.48	1.52
1.42	1.37
1.29	1.44
1.40	1.53

Μπορείτε να ισχυρισθείτε ότι ο χημικά επεξεργασμένος τύπος χαρτιού εμφανίζει μεγαλύτερη αντοχή;

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 1999)

11. Εστω ότι έχουμε επιλέξει δύο ανεξάρτητα τυχαία δείγματα 15 παρατηρήσεων το καθένα πάνω στις τυχαίες μεταβλητές X και Y με σκοπό να ελέγξουμε την υπόθεση

H_0: η X παίρνει συνήθως μικρότερες τιμές από την Y.

Κατατάσσουμε κατά αύξουσα σειρά μεγέθους όλες τις παρατηρήσεις και έστω ότι το άθροισμα των τάξεων μεγέθους των παρατηρήσεων πάνω στην X είναι τ=282. Ποιά από τις ενέργειες που ακολουθούν θα ήταν εύλογη;

α) Απόρριψη της H_0 σε επίπεδο σημαντικότητας α=0.01

β) Απόρριψη της H_0 σε επίπεδο σημαντικότητας α=0.05

γ) Αποδοχή της H_0 σε επίπεδο σημαντικότητας α=0.05

δ) Αποδοχή της H_0 σε επίπεδο σημαντικότητας α=0.1

(Οικ. Παν/μιο Αθηνών – Εξετ. Δεκ. 2000)

12. Ο αναλογιστής μιας συγκεκριμένης ασφαλιστικής εταιρείας θέλει να εξετάσει τα αρχεία δηλώσεων κλοπών που κατατέθηκαν από άτομα που είχαν ασφαλίσει τα σπίτια τους για μικροκλοπή. Στο παρελθόν, η διάμεσος αιτούμενη αποζημίωση ήταν 85000 δρχ. Τα ύψη των αποζημιώσεων, οι οποίες ζητήθηκαν από ένα τυχαίο δείγμα ασφαλισμένων, που είχαν πέσει θύματα μικροκλοπών είναι τα εξής (σε χιλιάδες δραχμές):

140, 92, 35, 202, 80, 87, 80, 100, 47, 25, 160, 68, 50, 65, 310, 90, 75, 120

Υπάρχουν ενδείξεις ότι το ύψος της διαμέσου αιτούμενης αποζημίωσης έχει αυξηθεί σημαντικά; (α=5%).

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 1998)

13. Να σχολιάσετε αναλυτικά τον στατιστικό έλεγχο των Mann-Whitney σε σύγκριση με τον αντίστοιχό του παραμετρικό έλεγχο και να αναφέρετε πότε είναι προτιμότερο να χρησιμοποιείται ο κάθε ένας από αυτούς.

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 2000)

14. Οπως είναι γνωστό, η ελεγχοσυνάρτηση του ελέγχου Mann-Whitney για τον έλεγχο δύο ανεξάρτητων πληθυσμών, με βάση ανεξάρτητα δείγματα τιμών $X_1,X_2,\ldots,X_n$ (από τον ένα πληθυσμό) και $Y_1,Y_2,\ldots,Y_m$ (από τον άλλο), είναι η $T=\sum_{i=1}^{n} R(X_i)$. Να βρεθεί η ακριβής κατανομή της T κάτω από την μηδενική υπόθεση για την περίπτωση n=3 και m=2.

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 2000)

15. Ο έλεγχος Mann-Whitney μπορεί να εφαρμοσθεί σε δεδομένα των οποίων η κλίμακα μέτρησης είναι ονομαστική.

α. Σωστό β. Λάθος

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 2000)

16. Ο μέσος βαθμός ενός μαθήματος ιστορίας σε ένα πανεπιστημιακό τμήμα κατά την τελευταία δεκαπενταετία είναι 6.5 μονάδες. Την τελευταία χρονιά άλλαξε το σύστημα εισαγωγής των φοιτητών στο τμήμα αυτό και μετά την εξέταση του μαθήματος, ο υπεύθυνος καθηγητής επιθυμεί να ελέγξει αν διαφοροποιήθηκαν σημαντικά οι βαθμολογίες των φοιτητών σε σχέση με εκείνες των φοιτητών των προηγούμενων δεκαπέντε ετών. Για τον λόγο αυτό συνέλεξε ένα τυχαίο δείγμα 15 φοιτητών των οποίων οι βαθμοί είναι οι παρακάτω:

4, 4.5, 1, 9.5, 2.5, 6, 7, 9, 9.5, 8, 6.5, 10, 9, 6, 7.

Πιστεύετε ότι τα παραπάνω δεδομένα παρέχουν σημαντικές ενδείξεις ότι η βαθμολογία των νέων φοιτητών διαφέρει σημαντικά από την βαθμολογία των φοιτητών των προηγουμένων ετών;

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 1999)

17. Ενας καθηγητής ψυχολογίας δίδαξε σε 2 τάξεις. Η πρωινή του τάξη αποτελείτο από 9 φοιτητές και η απογευματινή του από 12. Στις τελικές εξετάσεις εξετάσθηκαν όλοι οι φοιτητές του μαζί. Η βαθμολογία όλων των φοιτητών δίνεται στον παρακάτω πίνακα. Μπορείτε να υποστηρίξετε, σε επίπεδο στατιστικής σημαντικότητας 5%, ότι οι φοιτητές της πρωινής τάξης έγραψαν χειρότερα απ’ αυτούς της απογευματινής;

Πρωινή Τάξη: 73 87 79 75 82 66 95 75 70

Απογευματινή Τάξη: 86 81 84 88 90 85 84 92 83 91 53 84

18. Μία βιομηχανία καλλυντικών ενδιαφέρεται να μελετήσει τις αντιδράσεις της αγοράς σε επτά νέες σειρές καλλυντικών. Οι κυριότερες αγορές είναι η Βρεττανική και Αμερικάνικη. Για να έχει μία εικόνα των προτιμήσεων στις δύο αυτές αγορές έκανε μία δειγματοληπτική έρευνα βασισμένη σε ένα δείγμα 200 γυναικών, 100 Βρεττανών και 100 Αμερικανίδων, από τις οποίες ζητήθηκε να αξιολογήσουν τις σειρές των καλλυντικών από το 1 (για την πιο αρεστή) έως το 7 (για την λιγότερο αρεστή). Για κάθε μία χώρα οι 100 βαθμοί προτίμησης για κάθε σειρά καλλυντικών αθροίστικαν και στην σειρά με το μικρότερο άθροισμα αντιστοιχήθηκε ο βαθμός 1, στην σειρά με το αμέσως μεγαλύτερο άθροισμα ο βαθμός 2 κ.ο.κ. Η αξιολόγηση που προέκυψε με τον τρόπο αυτό γαι κάθε μία από τςι δύο χώρες είναι η εξής:

Σειρά καλλυντικών

	Α	**Β**	**Γ**	**Δ**	**Ε**	**ΣΤ**	**Ζ**
Βρεττανές	5	3	2	4	1	6	7
Αμερικανίδες	3	1	4	2	7	6	5

Να υπολογισθούν οι συντελεστές ρ του Spearman και τ του Kendall και να χρησιμοποιηθούν ως ελεγχοσυναρτήσεις για τον έλεγχο της υπόθεσης ότι οι προτιμήσεις των γυναικών των δύο χωρών είναι συσχετισμένες.

19. Οι παρακάτω παρατηρήσεις αποτελούν τα σκορ, τα οποία επέτυχαν παντρεμένες και ανύπαντρες γυναίκες που υποβλήθηκαν σε τεστ προσωπικότητας.

Ανύπαντρες	85	65	74	79	60	77	75	68	69
Παντρεμένες	72	76	66	73	73	63	70	70	71

Μπορεί να υποστηρισθεί ο ισχυρισμός ότι οι ανύπαντρες γυναίκες έχουν ισχυρότερη προσωπικότητα από τις παντρεμένες; (α =10%)

20. Να εξετασθεί αν οι παρακάτω παρατηρήσεις στο τυχαίο διάνυσμα (X, Y) δείχνουν γραμμική εξάρτηση μεταξύ των X και Y.

(3.6, 13) (4.7, 19) (1.4, 9) (5.5, 15) (4.8, 27) (4.3, 14) (3.0, 6)
(4.2, 11) (6.0, 24) (6.8, 26) (4.1, 19) (3.2, 9) (4.0, 8) (1.9, 6)
(0.4, 7) (4.9, 14) (5.6, 18) (5.6, 20)

21. Ενας ερευνητής υπολόγισε το βάρος των μελών των νεογνών τα οποία γεννήθηκαν σε 8 διαφορετικές γέννες χοίρων, με σκοπό να προσδιορίσει εάν το βάρος επηρεάζεται από τον αριθμό των νεογνών που προέρχονται από μια μόνο γέννα. Με τα στοιχεία που συνέλεξε ο ερευνητής αυτός να ελέγξετε την υπόθεση ότι δεν υπάρχει διαφορά στα μέσα βάρη των νεογνών των χοίρων που προέρχονται από γέννες διαφορετικού αριθμού νεογνών έναντι της διαζευκτικής ότι τα μέσα βάρη των νεογέννητων που προέρχονται από γέννες διαφορετικού αριθμού νεογέννητων δεν είναι όλα ίσα (α=0.05).

Γέννες

1η: 2.0 2.8 3.3 3.2 4.4 3.6 1.9 3.3 2.8 1.1

2η: 3.5 2.8 3.2 3.5 2.3 2.4 2.0 1.6

3η: 3.3 3.6 2.6 3.1 3.2 3.3 2.9 3.4 3.2 3.2

4η: 3.2 3.3 3.2 2.9 3.3 2.5 2.6 2.8

5η: 2.6 2.6 2.9 2.0 2.0 2.1

6η: 3.1 2.9 3.1 2.5

7η: 2.6 2.2 2.2 2.5 1.2 1.2

8η: 2.5 2.4 3.0 1.5

(Οικ. Παν/μιο Αθηνών – Εξετ. Φεβρ. 1992)

22. Τρεις κριτές βαθμολόγησαν 8 υποψηφίους. Ελέγξατε αν οι υποψήφιοι μπορούν να θεωρηθούν ισοδύναμοι.

Κριτές	1	2	3	4	5	6	7	8
Α	2	1	3	5	4	8	7	6
Β	1	2	4	5	7	6	8	3
Γ	3	2	1	7	5	8	6	4

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 1993)

23. Προκειμένου να εξετασθεί αν οι αξιολογήσεις 10 γραπτών από 85 εξεταστές διαφέρουν, επελέγη ένα τυχαίο δείγμα 6 εξεταστών. Αέ βαθμολογίες τους για τα 10 γραπτά δίνονται στον πίνακα που ακολουθεί.

		Γραπτό									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	1	22	30	27	30	28	28	28	28	36	29
Εξεταστής	**2**	20	28	25	29	28	25	29	34	40	30
	3	22	28	29	28	25	29	33	29	33	27
	4	24	29	30	28	29	27	30	30	34	30
	5	30	41	37	41	34	32	35	29	42	34
	6	27	27	32	33	33	23	36	22	42	29

Να ελεγχθεί αν τα δεδομένα παρέχουν ενδείξεις ότι η βαθμολογία κάποιων εξεταστών τείνει να είναι υψηλότερη από την βαθμολογία άλλων.

(Οικ. Παν/μιο Αθηνών – Εξετ. Ιαν. 1994)

24. Μία αλυσίδα καταστημάτων τροφίμων διεξήγαγε έρευνα σχετικά με την προτίμηση των καταναλωτών ως προς την ύπαρξη ή μη μουσικής υπόκρουσης μέσα σ' αυτά. Για τον σκοπό αυτό έγινε έρευνα, η οποία στηρίχθηκε σε τυχαίο δείγμα δεκαπέντε καταστημάτων. Κατά την διάρκεια μίας εβδομάδας, σε πέντε από αυτά ακουγόταν μουσική σε αργό ρυθμό και χαμηλή ένταση (είδος Ι), σε άλλα πέντε από αυτά ακουγόταν μουσική με αργό ρυθμό και μέτρια ένταση (είδος ΙΙ) και στα υπόλοιπα πέντε ακουγόταν μουσική σε γρήγορο ρυθμό και μέτρια ένταση (είδος ΙΙΙ). Μετά από μία εβδομάδα ύπαρξης μουσικής υπόκρουσης, εκατό από τους πελάτες κάθε καταστήματος που επελέγησαν τυχαία, ρωτήθηκαν εάν αυτό τους άρεσε. Τα ποσοστά των ατόμων που τους άρεσε η ύπαρξη μουσικής υπόκρουσης και επιθυμούσαν την συνέχισή της δίνονται στον παρακάτω πίνακα, κατά είδος μουσικής.

Είδος Μουσικής Ι (%)	Είδος Μουσικής ΙΙ (%)	Είδος Μουσικής ΙΙΙ (%)
94	84	81
97	89	76
90	82	73
86	90	79
91	78	84

Να ελέγξετε αν υπάρχουν ικανοποιητικές ενδείξεις που να αποδεικνύουν διαφορές στα επίπεδα αποδοχής των τριών ειδών μουσικής; (α=10%)

(Οικ. Παν/μιο Αθηνών – Εξετ. Σεπτ. 1997)

25. Μπορεί να υποστηριχθεί ο ισχυρισμός ότι τα παρακάτω ανεξάρτητα τυχαία δείγματα προέρχονται από πληθυσμούς με την ίδια διάμεσο;

α)	3	21	24	19	19	26	21	22	20	
β)	21	20	1	13	14	13	18	21	19	14
γ)	31	30	21	23	26	25	24			
δ)	8	12	11	7	9	11	10	11		

26. Ενα τυχαίο δείγμα 6 οδηγών, που συνελήφθησαν να οδηγούν μεθυσμένοι, υποβλήθηκαν σε δύο διαφορετικά «τεστ» για τον καθορισμό του επιπέδου του οινοπνεύματος στο αίμα τους με τα εξής αποτελέσματα:

		Οδηγός					
		1	**2**	**3**	**4**	**5**	**6**
Τεστ	**Α**	0.170	0.190	0.2	0.183	0.187	0.178
	Β	0.197	0.178	0.15	0.176	0.205	0.153

α) Είναι εύλογο να υποθέσει κανείς ότι οι ενδείξεις των δύο τεστ είναι αρνητικά συσχετισμένες;

β) Να υπολογισθεί το κρίσιμο επίπεδο.

27. Προκειμένου να αξιολογηθεί η αποτελεσματικότητα δύο αντικαταθλιπτικών φαρμάκων, 12 κλινικοί ασθενείς που έπασχαν από κατάθλιψη στον ίδιο, περίπου, βαθμό χωρίσθηκαν με τυχαίο τρόπο σε 2 ομάδες των 6 ασθενών. Στους ασθενείς της ομάδας 1, χορηγήθηκε το φάρμακο Α για μία περίοδο 6 μηνών, ενώ, στους ασθενείς της ομάδας 2, χορηγήθηκε το φάρμακο Β για την ίδια εξάμηνη χρονική περίοδο. Έξι εβδομάδες αργότερα, και οι 12 ασθενείς «βαθμολογούνται» ως προς το επίπεδο της κατάθλιψης, σε μία κλίμακα από το 0 μέχρι το 10, από έναν ψυχίατρο, ο οποίος δεν γνωρίζει ποιοι ασθενείς ακολουθούν την αγωγή Α και ποιοι την αγωγή Β. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί. (Οι μεγαλύτεροι βαθμοί αντιστοιχούν σε ασθενείς με υψηλότερα επίπεδα κατάθλιψης).

Ομάδα 1 (φάρμακο Α):	10	10	9	1	0	0
Ομάδα 2 (φάρμακο Β):	6	6	5	5	4	4

Παρέχουν τα δεδομένα ενδείξεις ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των διασπορών των δύο ομάδων;

28. Για να ελέγξει κατά πόσο δύο διαφορετικές μέθοδοι διδασκαλίας οδηγούν σε απόδοση των μαθητών με διαφορετικού βαθμού διακυμάνσεις, ένας καθηγητής Μαθηματικών εφαρμόζει αυτές τις μεθόδους σε 2 ομάδες μαθητών της ίδιας περίπου ικανότητας στα Μαθηματικά. Η ομάδα 1, αποτελούμενη από 5 μαθητές, διδάσκεται το μάθημα μέσω των παραδόσεων και ενός συμβατικού βιβλίου (μέθοδος Α). Η ομάδα 2, αποτελούμενη από 6 μαθητές, διδάσκεται το μάθημα μέσω της χρήσης ενός υπολογιστικού πακέτου (μέθοδος Β). Οι εξετάσεις, στο τέλος της χρονιάς, έδωσαν τα εξής αποτελέσματα:

Ομάδα 1 (μέθοδος Α):	14	12	10	10	8

Ομάδα 2 (μέθοδος Β):	18	16	14	14	10	8

Θα μπορούσε να συμπεράνει ο καθηγητής ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των διακυμάνσεων των αποδόσεων των 2 ομάδων μαθητών;

29. Ενας ψυχολόγος διεξάγει έρευνα για να προσδιορίσει αν ο θόρυβος επηρεάζει αρνητικά την μάθηση. 15 άτομα που συμφώνησαν να μετέχουν στο πείραμα, χωρίζονται με τυχαίο τρόπο σε τρεις ομάδες των 5 ατόμων. Σε κάθε ένα από τα μέλη των ομάδων, δίνεται χρόνος 20 λεπτών για να απομνημονεύσει έναν κατάλογο 10 ακατάλληλων συλλαβών, προκειμένου να εξετασθεί σ' αυτές την επόμενη μέρα. Τα 5 άτομα της ομάδας 1 (συνθήκες πλήρους ησυχίας) μελετούν τον κατάλογο των συλλαβών σε μία τελείως ήσυχη αίθουσα. Τα 5 άτομα της ομάδας 2 (συνθήκες μέτριου θορύβου) μελετούν τον κατάλογο των συλλαβών ενώ ακούν κλασική μουσική. Τα 5 άτομα της ομάδας 3 (συνθήκες ακραίου θορύβου) μελετούν τον κατάλογο των συλλαβών ακούγοντας μουσική ροκ. Οι αριθμοί των ακατάλληλων συλλαβών που θυμόντουσαν σωστά τα 15 άτομα την ημέρα της εξέτασης ήταν οι εξής:

Ομάδα 1 (πλήρης ησυχία):	8	10	9	10	9
Ομάδα 2 (κλασική μουσική):	7	8	5	8	5
Ομάδα 3 (μουσική ροκ):	4	8	7	5	7

Αποτελούν τα παραπάνω δεδομένα ένδειξη ότι οι συνθήκες θορύβου στον χώρο μελέτης επηρεάζουν την μάθηση, όπως αυτή αξιολογείται με την απόδοση στα πειράματα αυτής της μορφής;

30. Ο καθηγητής ενός πειραματικού σχολείου έχει 4 τάξεις μαθητών στο μάθημα των Μαθηματικών. Υπάρχουν 6 μαθητές στην τάξη 1, 7 μαθητές στην τάξη 2, 8 μαθητές στην τάξη 3 και 6 μαθητές στην τάξη 4. Ο καθηγητής χρησιμοποιεί, στην τάξη 1, ένα βιβλίο Μαθηματικών παράλληλα με ένα υπολογιστικό πακέτο, στην τάξη 2, ένα συμβατικό σχολικό βιβλίο, στην τάξη 3, διανέμει σημειώσεις των παραδόσεών του και στην τάξη 4 δεν δίνει καθόλου έντυπο υλικό. Στο τέλος του πρώτου εξαμήνου, ζητά από έναν άλλο καθηγητή να ταξινομήσει τους μαθητές του ως προς την ικανότητά τους στα Μαθηματικά με βάση την απόδοσή τους σε ένα διαγνωστικό τεστ. Οι σχετικές τους θέσεις στην ταξινόμηση αυτή είναι οι εξής:

Τάξη 1:	1	2	4	6	8	9		
Τάξη 2:	10	14	18	20	21	25	26	
Τάξη 3:	3	5	7	11	12	16	17	22
Τάξη 4:	13	15	19	23	24	27		

(Οι χαμηλές σχετικές θέσεις (βαθμοί) αντιστοιχούν σε καλύτερους μαθητές). Παρέχουν τα αποτελέσματα της συγκριτικής αξιολόγησης ενδείξεις ότι οι 4 μέθοδοι διδασκαλίας που εφαρμόζει ο καθηγητής διαφέρουν ως προς την αποτελεσματικότητά τους;