ΥΔΡΟΣΚΟΠΙΟ

ΠΡΟΓΡΑΜΜΑ STRIDE ΕΛΛΑΣ

ΔΗΜΙΟΥΡΓΙΑ ΕΘΝΙΚΗΣ ΤΡΑΠΕΖΑΣ ΥΔΡΟΛΟΓΙΚΗΣ ΚΑΙ ΜΕΤΕΩΡΟΛΟΓΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ

HELLENIC NATIONAL METEOROLOGICAL SERVICE

ΜΑΜΑΣΗΣ

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΟΜΟΙΟΓΕΝΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ, ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΣΦΑΛΜΑΤΩΝ, ΣΥΣΧΕΤΙΣΗΣ ΚΑΙ ΔΙΟΡΘΩΣΗΣ ΑΠΟΚΛΙΣΕΩΝ

DEVELOPMENT OF METHODS FOR HOMOGENIZATION OF DATA, ERROR DETECTION AND CORRECTION

Δ. Κατσιμάρδος και Θ. Χαραντώνης

D. Katsimardos and Th. Charantonis

Αριθμός τεύχους
Report number 5/9

ΑΘΗΝΑ - ΦΕΒΡΟΥΑΡΙΟΣ 1993
ATHENS - FEBRUARY 1993

HYDROSCOPE

STRIDE HELLAS PROGRAMME

DEVELOPMENT OF A NATIONAL DATA BANK FOR HYDROLOGICAL AND METEOROLOGICAL INFORMATION

<u>ΠΕΡΙΕΧΟΜΕΝΑ</u>

ΠΕΡΙΛΗΨΗ

Η εργασία αυτή πραγματοποιήθηκε για το ερευνητικό προγραμμα " Υδροσκόπιο ", στην Εθνική Μετεωρολογική Υπηρεσία, στα πλαίσια του προγράμματος " STRIDE - ΕΛΛΑΣ ".

Με βάση την υπάρχουσα διεθνή και εγχώρια βιβλιογραφία και κυρίως τα Τεχνικά Εγχειρίδια του Παγκοσμίου Μετεωρολογικού Οργανισμού, παρέχονται σ'αυτή την εργασία βασικές στατιστικές μέθοδοι για τον προσδιορισμό λαθών και τον έλεγχο της ομοιογένειας των μετεωρολογικών χρονοσειρών. Επίσης περιγράφονται αρχικά ορισμένες στατιστικές έννοιες που η γνώση τους είναι απαραίτητη για την κατανόηση των μεθόδων που αναφέρονται στα επόμενα κεφάλαια.

Ακολούθως παρατίθενται βασικές μέθοδοι ομοιογενοποίησης των δεδομένων όποτε αυτό είναι δυνατό καθώς και παραδείγματα.

ABSTRACT

The present work has been developed for the research project " Hydroscope ", at the National Meteorological Service, in the framework of the Programme " STRIDE - HELLAS".

On the basis of the existing bibliography (Greek and International) and mainly on the basis of the Technical Notes of the World Meteorological Organization, the basic statistical methods for the determination of errors and the checking of the homogeneity of meteorological time series are given. Also, in the beginning of the work, some statistical definitions and theorems are given, the knowledge of is necessary for the understanding of the methods referred to in the following chapters.

Finally some basic methods for homogeneization of a time series are given whenever this is possible as well as some examples.

1.ΕΙΣΑΓΩΓΗ.

Είναι γνωστό ότι η σωστή εκτίμηση των κλιματολογικών (ή/και στατιστικών) παραμέτρων μιας περιοχής έχει εξαιρετικά μεγάλη σημασία. Αυτές οι παράμετροι υπολογίζονται με βάση τις τιμές των μετεωρολογικών παραμέτρων που περιέχονται στις βάσεις των μετεωρολογικών δεδομένων. Είναι επομένως απόλυτα αναγκαίο, αυτά τα δεδομένα, που στο μεγαλύτερο μέρος τους είναι οι τιμές των μετρήσεων των μετεωρολογικών παραμέτρων, να είναι όσο το δυνατό πλήρη και σωστά, για να είναι τα προκύπτοντα αποτελέσματα της στατιστικής επεξεργασίας αξιόπιστα.

Αν και οι τυπικές διαδικασίες εκτέλεσης των παρατηρήσεων στους σταθμούς και διαβίβασης των μηνυμάτων τους τείνουν να ελαχιστοποιήσουν τα σφάλματα, εν τούτοις δεν τα έχουν εξαλείψει. Γιαυτό, στο χώρο συγκέντρωσης και αποκωδικοποίησης των μηνυμάτων, πρέπει να γίνεται (και γίνεται) ένας ποιοτικός έλεγχος των παρατηρήσεων σύμφωνα με τα διεθνή πρότυπα.

Ο ποιοτικός έλεγχος των παρατηρήσεων συνίσταται κατά βάση:

(α) Στον έλεγχο των τιμών των παραμέτρων ως προς ορισμένα, για την κάθε εξεταζόμενη παράμετρο, όρια, μέσα στα οποία πρέπει να βρίσκονται αυτές.

(β) Στην σύγκριση των τιμών ορισμένων παραμέτρων με τις τιμές άλλων παραμέτρων της ίδιας ώρας, για να διαπιστώνεται η εσωτερική συνέπεια του συνόλου των παρατηρούμενων τιμών.

(γ) Στη σύγκριση των τιμών της κάθε παραμέτρου με τις αντίστοιχες τιμές της ίδιας παραμέτρου κατά την προηγούμενη παρατήρηση, για την ιστορική τους συνέπεια.

(δ) Στη σύγκριση των τιμών των διάφορων παραμέτρων με τις αντίστοιχες τιμές των ίδιων παραμέτρων σε γειτονικούς σταθμούς, για την συνέπειά τους στο χώρο.

Όλοι οι παραπάνω επί μέρους έλεγχοι αποτελούν τα βασικά στάδια του χωροχρονικού ποιοτικού ελέγχου, που πρέπει να γίνουν σύμφωνα με τα διεθνή πρότυπα. Η ανάπτυξη των μεθόδων του ποιοτικού αυτού ελέγχου αποτελεί ιδιαίτερη μελέτη, στα πλαίσια του STRIDE ("ΥΔΡΟΣΚΟΠΙΟ" μελέτη 11/3) και στην παρούσα εργασία θεωρείται ότι στις τιμές των παραμέτρων έχει γίνει ο πρωταρχικός αυτος έλεγχος.

Οι παραπάνω έλεγχοι μπορούν να διαπιστώσουν ένα μέρος των σφαλμάτων που είναι δυνατόν να υπάρξουν. Παραμένουν ομως σφάλματα που προκύπτουν απο την αλλαγή της μεθοδου παρατήρησης, του οργάνου μέτρησης, του παρατηρητή και ακόμα απο μετατοπιση του σταθμού σε γειτονική του θέση, γιατί οι διαφορές των "νέων" τιμών, απο αντίστοιχες "παλιές", που οφείλονται στους παραπάνω λόγους, δεν είναι τόσο σημαντικές ώστε να μπορούν να διαπιστωθούν απο τους προηγούμενους ελέγχους. Αν π.χ. ένας σταθμός αλλάξει το βαρομετρο ή το θερμόμετρο και το νέο οργανο δίνει τιμές μικροτερες ή μεγαλύτερες απο τις τιμές του οργάνου που αντικατέστησε, ο ποιοτικος έλεγχος δεν θα μπορέσει να διαπιστώσει τέτοιο σφάλμα, επειδή οι διαφορές των τιμών των δύο οργάνων δεν μπορεί να διαφέρουν σημαντικά. Ακόμα τα ορια των τιμών μέσα στα οποία πρέπει να ανήκουν οι διάφοροι παράμετροι, είναι αρκετά μεγάλα και λάθη της μορφής π.χ. 21°C αντί 12°C (αναγραμματισμού) δεν μπορούν να διαπιστωθούν στον πρωταρχικό αυτο έλεγχο.

Τέλος ο πρωταρχικός έλεγχος δεν συμπεριλαμβάνει παραμέτρους όπως η ηλιοφάνεια, η ηλιακή ακτινοβολία κλπ. Οι τιμές των παραμέτρων αυτών έχουν απο την φύση τους τόσο ελάχιστη όσο και μέγιστη τιμή.

Για την διαπίστωση τέτοιων σφαλμάτων χρειάζονται άλλοι έλεγχοι, που δεν μπορούν να γίνουν σε REAL TIME όπως οι προηγούμενοι και σκοπος αυτής της μελέτης είναι:

α Η ανάπτυξη μεθοδων διαπίστωσης των σφαλμάτων αυτών και διόρθωσης των τιμών, όπου και όσο αυτό είναι δυνατό και

β Η ανάλυση των στατιστικών παραμέτρων σε συνδυασμό με το ιστορικό των σταθμών για τον έλεγχο της ομοιογένειας των τιμών των μετεωρολογικών παραμέτρων, και η ανάπτυξη των μεθοδων που χρησιμοποιούνται για την ομοιογενοποίηση (όπου αυτό δεν συμβαίνει).

Τα σφάλματα κατά τον Αδαμόπουλο 1960, διακρίνονται σε δύο μεγάλες κατηγορίες: **Τα τυχαία και τα συστηματικά.**

Τα τυχαία σφάλματα οφείλονται σε αίτια που δεν είναι γνωστά και είναι πολύ δύσκολο να διαπιστωθούν, θα αποτελέσουν δε το αντικείμενο του 3ου Κεφαλαίου.

Τα συστηματικά σφάλματα οφείλονται σε αίτια που μπορούν να βρεθούν και είναι δυνατον να διαπιστωθούν με ορισμένα **tests** ομοιογένειας. Τα σφάλματα αυτά επιδρούν κατά το μάλλον ή ήττον ομοιόμορφα σ΄όλες της τιμές της παραμέτρου, θα αποτελέσουν δε το αντικείμενο του 4ου Κεφαλαίου, όπου παρατίθενται και οι βασικές μέθοδοι ελέγχου της ομοιογένειας των χρονοσειρών.

Στο Κεφάλαιο 5 δίνονται οι βασικές μέθοδοι ομοιογενοποίησης των χρονοσειρών όπου και όταν αυτό είναι δυνατό.

Τέλος στο Κεφάλαιο 2 ορίζονται **οι στατιστικές** έννοιες που χρησιμοποιούνται στα επόμενα κεφάλαια και δίνονται τα βασικά θεωρήματα της στατιστικής που η γνώση τους είναι απολύτως αναγκαία για την κατανόηση των προτεινόμενων μεθόδων.

2.ΓΕΝΙΚΕΣ ΕΝΝΟΙΕΣ - ΟΡΙΣΜΟΙ.

2.1 Τυχαία σφάλματα.

Όπως ήδη έχει αναφερθεί, τα τυχαία σφάλματα οφείλονται σε λόγους που είναι άγνωστοι και είναι πολύ δύσκολο να διαπιστωθούν για να μπορέσει κανείς να τα διορθώσει. Για το λόγο αυτό σε συνδυασμό με το ότι η μέση τιμή τους είναι μηδέν, δεν υπάρχουν στην βιβλιογραφία πολλά για τα σφάλματα αυτά όσον αφορά απλές τιμές-μετρήσεις.

Ένα **σφάλμα χαρακτηρίζεται τυχαίο** όταν πληρούνται οι παρακάτω προϋποθέσεις-συνθήκες (Αδαμόπουλος 1960):

(α) Συνθήκη ανεξαρτησίας. Σύμφωνα μ'αυτή, τα αίτια που προκάλεσαν τα σφάλματα είναι ανεξάρτητα μεταξύ τους, δηλαδή αν ένα αίτιο προκάλεσε θετικό ή αρνητικό αποτέλεσμα, το αποτέλεσμα αυτό δεν συνδέεται με άλλο θετικό ή αρνητικό αποτέλεσμα κάποιου άλλου αιτίου.

Παράδειγμα 2.1.α.

Τα σφάλματα αναγραμματισμού είναι ανεξάρτητα μεταξύ τους και είναι τυχαία.

(β) Συνθήκη ομοιογένειας. Σύμφωνα με την συνθήκη αυτή τα αίτια που επιδρούν στις τιμές εμφανίζουν την ίδια ένταση επί όλων των τιμών του πληθυσμού από τον οποίο έχουν ληφθεί οι προς μελέτη τιμές.

Παράδειγμα 2.1.β.

Η ελάχιστη θερμοκρασία το καλοκαίρι είναι μεγαλύτερη από εκείνη του χειμώνα. Εδώ ο παράγοντας "εποχή" δεν επιδρά με τον ίδιο τρόπο στο σύνολο των τιμών της ελάχιστης θερμοκρασίας.

(γ) Συνθήκη συμμετρίας. Οι αιτίες που προκαλούν θετικές αποκλίσεις των τιμών πρέπει να εξουδετερώνουν αυτές που προκαλούν αρνητικές αποκλίσεις, έτσι ώστε το άθροισμα τους να είναι ίσο με μηδέν.

Παράδειγμα 2.1.γ.

Αν πάρουμε μια μόνο τιμή του πληθυσμού η θετική ή η αρνητική της απόκλιση, από την μέση τιμή του πληθυσμού από τον οποίο προέρχεται, δεν εξουδετερώνεται από άλλη, αντίθετή της διαφορά, οπότε δεν πληρούται η συνθήκη της συμμετρίας (μεροληψία).

(δ) **Συνθήκη του μεγάλου αριθμού.** Τα προκαλούντα τα σφάλματα αυτά αίτια πρέπει να είναι πολλά και της ίδιας (κατά το δυνατό) βαρύτητας για να λειτουργούν οι κανόνες της τύχης.

2.2 Τυχαίο δείγμα.

Μια προσφιλής μέθοδος εργασίας για την διαπίστωση άν τα σφάλματα είναι τυχαία, είναι η μέθοδος του δείγματος. Η μέθοδος αυτή συνίσταται στην στατιστική επεξεργασία ενός δείγματος δηλαδή ενός μικρού αριθμού τιμών-στοιχείων της παραμέτρου που μελετάται (μικρού σχετικά με το πλήθος-αριθμό των στοιχείων του συνόλου όλων των διαθέσιμων τιμών της παραμέτρου). Ο αριθμός των τιμών από τις οποίες αποτελείται το δείγμα ονομάζεται μέγεθος του δείγματος και η έκφραση **"δείγμα μεγέθους κ"** θα υποδηλώνει ένα δείγμα με πλήθος στοιχείων-τιμών ίσο με κ. Οι τιμές του δείγματος πρέπει να λαμβάνονται με εντελώς τυχαίο τρόπο από το σύνολο των τιμών της παραμέτρου.

Η μέθοδος του δείγματος εφαρμόζεται όταν:

(α) Η εύρεση των χαρακτηριστικών της παραμέτρου με την στατιστική επεξεργασία ολόκληρου του πληθυσμού, θα μπορούσε να καταστρέψει μέρος ή και ολόκληρο τον πληθυσμό.

Παράδειγμα 2.2.α.

Η αντοχή των αυτοκινήτων που κινούνται με 100Kmh^{-1} σε σύγκρουσή τους με τοίχο, θα κατέστρεφε ολόκληρο τον πλυθυσμό.

(β) Ο χρόνος για τον υπολογισμό των χαρακτηριστικών από το σύνολο των τιμών είναι μεγάλος και η λήψη των αποφάσεων επείγει.

Παράδειγμα 2.2.β.

Τα galop.

(γ) `Όταν δεν απαιτούνται πολύ μεγάλες ακρίβειες.

(δ) `Όταν δεν διατίθεται ολόκληρος ο πληθυσμός αλλά μόνο μικρός αριθμός τιμών.

Για τις μετεωρολογικές παραμέτρους οι παραπάνω προϋποθέσεις δεν υφίστανται, ιδιαίτερα σήμερα, αφού οι ηλεκτρονικοί υπολογιστές έλυσαν και το πρόβλημα του γρήγορου υπολογισμού των στατιστικών παραμέτρων που γίνεται σε ελάχιστο χρονο, ενώ τα αρχεία περιέχουν τις τιμές ολοκληρου του πληθυσμού.

Οι μέθοδοι που προτείνονται παρακάτω αφορούν τον έλεγχο προσδιορισμού σφαλμάτων σε στατιστικά δείγματα (μέσοι οροι, τυπικές αποκλίσεις κλπ), προτείνονται δε και μέθοδοι, ελέγχου και διόρθωσης απλών τιμών, οι οποίες προϋποθέτουν την διάθεση των πρωτογενών στοιχειων (row data).

Πριν δοθούν οι μέθοδοι αυτές καλό είναι να οριστούν ορισμένες έννοιες καθώς και μερικοί συμβολισμοί, που θα συναντά κανεις συχνά παρακάτω, οπως η έννοια του μεγέθους του δείγματος, που δοθηκε ήδη.

2.3 Βασικές στατιστικές παράμετροι και έννοιες.

Ένας συμβολισμός που χρησιμοποιείται αρκετά συχνά είναι αυτός του συνόλου:

$$I_\lambda = \{1,2,3,...,\lambda\}$$

που πολλές φορές απλοποιεί ολόκληρες παραστάσεις. Το I_λ είναι το υποσύνολο του N (σύνολο των φυσικών αριθμών) του οποίου τα στοιχεία είναι οι πρώτοι λ φυσικοί αριθμοί. Παραδειγματα της χρήσης του συμβολισμού είναι τα παρακάτω:

Παράδειγμα 2.3.α.

Οι δώδεκα μήνες του χρόνου (Ιανουάριος μέχρι Δεκέμβριος) αντιστοιχούν στους αύξοντες αριθμούς 1,2,...12. Όταν ζητείται η στατιστική παράμετρος X που προέκυψε απο τιμές ενός οποιουδήποτε αλλά καθορισμένου κάθε φορά μήνα και δεν ενδιαφερόμαστε για το ποιος είναι ο μήνας αυτός, τότε γράφουμε X_i, $i \in I_{12}$ και παίρνει συγκεκριμένη τιμή όταν δοθεί στο i μια απο τις 12 δυνατές τιμές του.

Παράδειγμα 2.3.β.

Αν ένας σταθμός άρχισε να λειτουργεί απο το 1951 στο έτος YEAR αντιστοιχούμε τον αριθμό K=YEAR-1951+1. Το 1993 έχει K=1993-1951+1=43. Αν θελήσουμε να πάρουμε το σύνολο **A** με στοιχεία τις παρατηρήσεις μιας παραμέτρου ενός οποιουδήποτε έτους απο το 1951 μέχρι το 1993 μπορούμε να γράψουμε A_i, $i \in I_{43}$

που γίνεται το σύνολο των στοιχείων ενός συγκεκριμένου χρόνου όταν δοθεί στο ι συγκεκριμένη τιμή.

2.3.1 Δειγματική ή στατιστική παράμετρος και κλιματολογική παράμετρος ή απλά Παράμετρος. Δειγματική παράμετρος θα ονομάζεται κάθε παράμετρος που προκύπτει από την επεξεργασία ενός δείγματος, πχ δειγματικός μέσος όρος, δειγματική τυπική απόκλιση, δειγματική κατανομή, κλπ. Οι ίδιες παράμετροι υπολογισμένες από το σύνολο του πληθυσμού μιας παραμέτρου (π.χ πίεσης) θα αναφέρονται με το όνομά τους.

Η τιμή των δειγματικών παραμέτρων διαφέρει από δείγμα σε δείγμα και εξαρτάται από το μέγεθος του δείγματος **n** και τον τρόπο επιλογής του δείγματος (τυχαίο ή μή), ενώ η παράμετρος που υπολογίζεται από το σύνολο των διατιθέμενων τιμών είναι μοναδική (για το σύνολο των τιμών που διατίθενται).

2.3.2 Ο μέσος όρος. Ο μέσος όρος του πληθυσμού (θεωρείται ότι διατίθενται όλες οι δυνατές τιμές της παραμέτρου που είναι σε πλήθος K αριθμοί) παριστάνεται με μ_x ή απλούστερα μ δίνεται δε από τις σχέσεις:

$$\mu = \mu_x = \frac{\sum_{i=1}^{K} x_i}{K} = \frac{{}_1\Sigma^K x_i}{K} = {}_1\Sigma^K x_i/K \qquad (2.1)$$

και ο **δειγματικός μέσος όρος E(X), ενός δείγματος μεγέθους n** από την ανάλογη σχέση:

$$E(X) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{{}_1\Sigma^n x_i}{n} = {}_1\Sigma^n x_i/n \qquad (2.2)$$

2.3.3. Η τυπική απόκλιση. Για την τυπική απόκλιση χρησιμοποιείται το σ_x ή απλούστερα σ όταν υπολογίζεται από ολόκληρο τον πληθυσμό (πλήθος στοιχείων K όπως στον υπολογισμό του μέσου όρου) και S_x ή/και S όταν πρόκειται για την δειγματική τυπική απόκλιση ενός δείγματος μεγέθους **n**. Τα ίδια αντίστοιχα σύμβολα υψωμένα στο τετράγωνο χρησιμοποιούνται για την διακύμανση.

Η τυπική απόκλιση σ ή σ_x δίνεται από την:

$$\sigma = \sigma_x = \sqrt{\sigma^2} = (\sigma^2)^{0.5} = sqrt(\sigma^2) \qquad (2.3)$$

όπου:

$$\sigma^2 = \sigma_x^2 = \frac{\sum_{i=1}^{K}(x_i-\mu_x)^2}{K} = \frac{{}_i\Sigma^K(x_i-\mu_x)^2}{K} = {}_i\Sigma^K(x_i-\mu_x)^2/K \qquad (2.4)$$

Αν με X^2 παραστήσουμε την τυχαία μεταβλητή-παράμετρο της οποίας τιμές είναι τα τετράγωνα των τιμών της υπό εξέταση παραμέτρου X και με $\mu(X^2)$ την μέση τιμή της, τότε επειδή:

$${}_i\Sigma^K(x_i-\mu_x)^2/K = {}_i\Sigma^K x_i^2/K - 2*\mu_x*\mu_x + \mu_x^2 = \mu(X^2)-\mu_x^2 \qquad (2.5)$$

Η (2.5) μπορεί να πάρει την μορφή:

$$\sigma = \sigma_x = [\mu(X^2)-\mu_x^2]^{0.5} = sqrt[\mu(X^2)-\mu_x^2] \qquad (2.6)$$

Η δειγματική τυπική απόκλιση S ή S_x ενός δείγματος μεγέθους n δίνεται από τις ανάλογες με τις παραπάνω σχέσεις:

$$S = S_x = \sqrt{S^2} = (S^2)^{0.5} = sqrt(S^2) \qquad (2.7)$$

όπου:

$$S^2 = S_x^2 = \frac{\sum_{i=1}^{n}[x_i-E(X)]^2}{K} = \frac{{}_i\Sigma^n[x_i-E(X)]^2}{K} = {}_i\Sigma^n[x_i-E(X)]^2/K \qquad (2.8)$$

Αν με $E(X^2)$ παραστήσουμε την δειγματική μέση τιμή της παραμέτρου X^2 που ορίσαμε παραπάνω, τότε επειδή και πάλι:

$${}_i\Sigma^n[x_i-E(X)]^2/K={}_i\Sigma^n x_i^2/n-2*E(X)*E(X)+E(X)^2=E(X^2)-[E(X)]^2 \qquad (2.9)$$

Η (2.9) μπορεί να πάρει την μορφή:

$$S = S_x = \{E(X^2)-[E(X)^2]\}^{0.5} = sqrt\{E(X^2)-[E(X)^2]\} \qquad (2.10)$$

Για μικρές τιμές του K πολλές φορές για τον υπολογισμό του S ο παρονομαστής K στην (2.8) αντικαθίσταται από το (K-1). Όταν υπολογίζεται μ'αυτό τον τρόπο την παριστάνουμε με S' και η σχέση που συνδέει τις S και S' είναι:

$$S' = S*[K/(K-1)]^{0.5} \qquad (2.11)$$

2.3.4. Εκατοστιαία σημεία. Ο αριθμός α λέγεται Κ εκατοστιαίο σημείο μιας παραμέτρου (ή ενός δείγματος απ'αυτή) όταν επαληθεύει τη σχέση:

$$P(X\leq\alpha)=K/100. \qquad (2.12)$$

Συνήθεις τιμές για το Κ είναι οι 1, 5, 10, 25, 100/3(ή 33.3), 50, 200/3(ή 66.6), 75, 90, 95 και 99.

2.3.5. Άκρες τιμές μιας παραμέτρου. Αν $x_1,x_2,...x_K$ είναι οι Κ τιμές μιας παραμέτρου X, οι αριθμοί:

$$MINX = \min\{x_i/i\in I_K\} \qquad (2.13)$$

$$MAXX = \max\{x_i/i\in I_K\} \qquad (2.14)$$

ονομάζονται **απόλυτα ελάχιστη** και **απόλυτα μέγιστη** τιμή της παραμέτρου, αντίστοιχα. Οι αντίστοιχες τιμές που προκύπτουν από ένα δείγμα μεγέθους **n** θα παριστάνονται με minX και maxX αντίστοιχα και είναι ίσες με:

$$minX=\min\{x_i/i\in I_n\} \qquad (2.15)$$

$$maxX=\max\{x_i/i\in I_n\} \qquad (2.16)$$

Το **πεδίο** τιμών της X το παριστάνουμε με D ή με MMX και το πεδίο τιμών ενός δείγματος με mmX. Τα παραπάνω σύνολα είναι τα διαστήματα:

$$D= MMX = [MINX , MAXX] \text{ και } mmX=[minX , maxX] \qquad (2.17)$$

Είναι προφανές ότι αν x μια τιμή της παραμέτρου X τότε $x\in MMX$ και αν x είναι μια τιμή από ένα δείγμα της παραμέτρου τότε $x\in mmX$ και ακόμη το διάστημα mmX είναι υποσύνολο-υποδιάστημα του MMX.

Οι αριθμοί ΕΥΡΧ=MAXX-MINX και ευρX=maxX-minX ονομάζονται **εύρος των τιμών** της παραμέτρου X και **δειγματικό εύρος** (του συγκεκριμένου δείγματος), αντίστοιχα.

2.3.6. Διαμελισμός συνόλου. Δοθέντος ενός συνόλου Α και μιας οικογένειας $\{A_i\}_{i\in I}$, υποσυνόλων αυτού (όπου I το σύνολο των δεικτών στο οποίο ανήκει το i, το I για τις ανάγκες της συγκεκριμένης μελέτης θεωρείται πεπερασμένο), των οποίων η ένωση είναι ίση με Α, δηλαδή:

$\cup_{i\in I}A_i=A$ λέμε ότι η **οικογένεια $\{A_i\}$ καλύπτει το Α.**

Όταν η οικογένεια $\{A_i\}_{i\in I}$ καλύπτει το σύνολο Α και τα σύνολα της οικογένειας είναι ανά δύο ξένα μεταξύ τους (δηλαδή αν για κάθε $i,j\in I$ με $i\neq j$ ισχύει: $A_i\cap A_j=\emptyset$) τότε λέμε ότι η οικογένεια $\{A_i\}$ **αποτελεί ένα διαμελισμό του συνόλου A.**

Παράδειγμα 2.3.γ.
Εστω A=[0,10] και τα σύνολα: $A_1=[0,1]$, $A_2=(1,2]$, $A_i=(i,i+1]$ με $2<i<11$, τότε τα σύνολα $A_i,i\in I_{10}$ αποτελούν ένα διαμελισμό του συνόλου A γιατί είναι ανά δύο ξένα μεταξύ τους και η ένωσή τους είναι ίση με το A.

Για τον έλεγχο της προσαρμογής της συνάρτησης πυκνότητας πιθανότητας, του συνόλου των τιμών μιας παραμέτρου X ή ενός τυχαίου δείγματός τους, σε κάποια θεωρητική συνάρτηση πυκνότητας πιθανότητας, διαμελίζεται το σύνολο (διάστημα) MMX (ή το mmX) σε κ υποσύνολά του (υποδιαστήματά του). Ο αριθμός κ των υποσυνόλων που προτείνεται σε τέτοιες περιπτώσεις δίνεται από την σχέση:

κ=5*logK

όπου K το πλήθος των διατιθέμενων τιμών και logK ο δεκαδικός λογάριθμος του K.

Ο συμβολισμός: **X≈C(X,α,β,...)** σημαίνει ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή C με παραμέτρους α,β,κλπ. Για παράδειγμα $X\approx N(X,\mu,\sigma^2)$ σημαίνει ότι η τυχαία μεταβλητή X ακολουθεί την κανονική (N) κατανομή με μέση τιμή μ και τυπική απόκλιση σ (μ, σ οι παράμετροι της φυσικής κατανομής).

2.3.7. Τάξη του x_i και τάξη του διατεταγμένου x_i. Εστω ένα σύνολο **n** τιμών μιας παραμέτρου A (δείγμα μεγέθους **n**), οι $x_1,x_2,...,x_n$. Διατάσσουμε τις τιμές αυτές κατά τάξη αύξοντος (ή φθίνοντος) μεγέθους και έστω $y_1,y_2,...,y_n$ οι διατεταγμένες τιμές. **Ο αριθμός i** ονομάζεται **τάξη της x_i. Αν $x_i=y_j$** τότε, **ο αριθμός** j ονομάζεται **τάξη του διατεταγμένου x_i ή διατεταγμένη τάξη του x_i σε αύξουσα (ή φθίνουσα) διάταξη .**

Αν συμβεί: $y_k=y_{k+1}=...=y_{k+m}$ με m>0, δηλαδή αν κάποια από τις **n** τιμές $x_i,i\in I_n$ του συνόλου των τιμών, συμβεί να εμφανίζεται περισσότερες από μια φορές και είναι $y_k=x_{i1}=y_{k+1}=x_{i2}=..,\ =y_{k+m}=x_{im}$, τότε σαν τάξη των διατεταγμένων x_{i1}, x_{i2},..,x_{im} ορίζεται ο **j'**, μέσος όρος των αριθμών κ,κ+1,..,κ+m, ίσος με:

$$j'={}_0\Sigma^m(\kappa+i)/(m+1)=[(\kappa+m)*(\kappa+m+1)-(\kappa-1)*\kappa]/[2*(m+1)]$$

2.4 Χρονοσειρά.

΄Οταν τα δεδομένα-τιμές μιας παραμέτρου έχουν ληφθεί κατά ισαπέχοντα μεταξύ τους χρονικά διαστήματα, τότε λέμε ότι τα δεδομένα αποτελούν μια **χρονοσειρά** της παραμέτρου. Το σταθερό χρονικό διάστημα ΔΤ που χωρίζει δύο διαδοχικές παρατηρήσεις ονομάζεται **περίοδος δείγματος της χρονοσειράς ή απλά περίοδος δείγματος.**

Οι Box και Jenkins 1970 ορίζουν την χρονοσειρά ως εξής:

΄Ενα σύνολο παρατηρήσεων-μετρήσεων που παράγεται σε χρονολογική σειρά ονομάζεται χρονοσειρά. Εαν το σύνολο των τιμών παράγεται από συνεχείς μετρήσεις στο χρόνο τότε η σειρά λέγεται συνεχής ενώ εάν οι παρατηρήσεις λαμβάνονται σε διακριτούς χρόνους η σειρά λέγεται διακριτή. Οι παρατηρήσεις κατά τις χρονικές στιγμές $\tau_1,\tau_2,...\tau_K$, σε μια διακριτή χρονοσειρά μπορούν να παρασταθούν με $x(\tau_1)$, $x(\tau_2)$,..., $x(\tau_K)$. Αν διατίθενται Κ διαδοχικές τιμές μιας χρονοσειράς που έχουν σταθερή διαφορά χρόνου ΔΤ και αν είναι τ_1 ο χρόνος της αρχής των παρατηρήσεων, τότε οι τιμές της χρονοσειράς μπορούν να πάρουν την μορφή $x(\tau_1),x(\tau_1+\Delta T),...x[\tau_1+(K-1)*\Delta T]$. ΄Οταν ο χρόνος τ_1 της έναρξης των παρατηρήσεων δεν ενδιαφέρει και ο ΔΤ εννοείται, τότε η χρονοσειρά μπορεί να πάρει την απλούστερη μορφή x_1, x_2,..., x_K που χρησιμοποιώντας το σύνολο I_K, γίνεται x_ι, $\iota\in I_K$.

Διακριτές χρονοσειρές μπορεί να παράγει κανείς με τους παρακάτω τρόπους.

(α) Εκτελώντας παρατηρήσεις κατά σταθερά χρονικά διαστήματα ΔΤ.

(β) Από μια συνεχή χρονοσειρά με στιγμομέτρηση κατά σταθερά χρονικά διαστήματα ΔΤ.

(γ) Από μια συνεχή αθροιστική χρονοσειρά όπως γίνεται με τις ταινίες των βροχογράφων.

Από τα παραπάνω γίνεται φανερό ότι από μία συνεχή χρονοσειρά μπορεί κανείς να παράγει απειρία διακριτών χρονοσειρών, αρκεί προς τούτο να θεωρεί είτε διαφορετικό χρόνο έναρξης της χρονοσειράς, είτε διαφορετικό ΔΤ πχ 2*ΔΤ.

Τα μετεωρολογικά δεδομένα προέρχονται από παρατηρήσεις (μετρήσεις) των μετεωρολογικών παραμέτρων που λαμβάνονται σε σταθερές ώρες της ημέρας. ΄Ετσι οι τιμές μιας ορισμένης ώρας κάθε μέρας, κάθε μιας απο τις μετρούμενες παραμέτρους δίνει και μία χρονοσειρά περιόδου δείγματος 24h. Αν κάποια παράμετρος μετριέται περισσότερες φορές την ημέρα (πχ 4 για τις παρατηρήσεις των κύριων συνοπτικών ωρών, 8 φορές για τις συνοπτικές παρατηρήσεις κλπ) λαμβάνονται χρονοσειρές μικρότερης περιόδου δείγματος.

Ο δείκτης **i** της τιμής x_i της παραμέτρου X την χρονική στιγμή **i** αποτελεί και την τάξη της x_i.

2.5 Στατιστικές Υποθέσεις. Στοιχεία θεωρίας αποφάσεων.

Στην στατιστική πολλές φορές ζητείται να αποφασίσει κανείς κατά πόσο η διαφορά, (μπορεί να είναι και λόγος) της τιμής μιας στατιστικής παραμέτρου που προέρχεται απο κάποιο δείγμα, απο κάποια υποθετική (ή μη) τιμή, είναι μικρή ώστε να μπορεί να οφείλεται σε τυχαίους παράγοντες. Για την εξαγωγή ανάλογων συμπερασμάτων οι Κούνιας κ.α. 1990 προτείνουν τα παρακάτω βήματα:

1. Ορισμός της Μηδενικής Υπόθεσης H_o (H_o στο εξής).

2. Ορισμός εναλλακτικής (ή των εναλλακτικών) υπόθεσης H_1 (ή H_1, H_2, H_3 κλπ).

3. Ορισμος του test που θα εφαρμοστεί.

4. Ορισμός της απορριπτικής περιοχής της H_o.

5. Εξαγωγή των συμπερασμάτων.

Όταν είναι γνωστή η κατανομή των συχνοτήτων της παραμέτρου απο την οποία προέρχεται το δείγμα είναι πολλές φορές δυνατόν να βρεθεί και η κατανομή των αποκλίσεων απο μια ορισμένη τιμή και, επομένως, για δοθείσα πιθανότητα, να βρεθούν τα όρια εντός των οποίων θα πρέπει να βρίσκεται η τιμή αυτή, ορίζοντας το διάστημα εμπιστοσύνης.

Η Μηδενική Υπόθεση H_o είναι η υπόθεση σύμφωνα με την οποία η διαφορά της τιμής της στατιστικής παραμέτρου που ελέγχεται από μια θεωρητική τιμή μπορεί να οφείλεται σε τυχαία γεγονότα. Εναλλακτική (-ες) υπόθεση (-εις) είναι κάθε συμπληρωματική (-ες) της H_o.

Παράδειγμα 2.5.α.

Τυχαίο δείγμα μεγέθους κ έχει μέση τιμή E(X) ενώ η μέση τιμή του πληθυσμού είναι μ_x. Μπορούμε να κάνουμε τις παρακάτω μηδενικές και εναλλακτικές υποθέσεις.

H_0: 1. $E(X)=\mu_x$ ή $\delta=E(X)-\mu_x=0$

H_0: 2. $E(X)=\mu_x+\alpha$, οπου α σταθερά ή $\delta=E(X)-\mu_x=\alpha$

H_1: α. $\delta\neq0$ ή $\delta\neq\alpha$ (δίπλευρο test)

H_1: β. $\delta>0$ ή $\delta<0$ ή $\delta>\alpha$ ή $\delta<\alpha$ (μονόπλευρο test)

Πάντα δεχόμαστε σαν σωστή την μία από τις H_0 ή H_1 και απορρίπτουμε την άλλη. Το ποια απ'αυτές θα γίνει δεκτή σαν ορθή θα το δώσει το test που θα εφαρμοστεί. Το τελευταίο είναι συνάρτηση της στατιστικής παραμέτρου που εκφράζει η τιμή που ζητείται να πληροί την μηδενική υπόθεση (μέση τιμή, τυπική απόκλιση, ποσοστά κλπ).

Υπάρχει περίπτωση μετά την εφαρμογή του test να προκύψει απόρριψη μιας σωστής υπόθεσης. Το σφάλμα που γίνεται στην περίπτωση αυτή ονομάζεται **σφάλμα πρώτου τύπου** και γράφεται **σφάλμα τύπου I**. Υπάρχει, αντίθετα, περίπτωση να γίνει δεκτή μια λανθασμένη υπόθεση. Το σφάλμα στην τελευταία περίπτωση ονομάζεται **σφάλμα δευτέρου τύπου** και γράφεται **σφάλμα τύπου II**.

Στάθμη σημαντικότητας (σ.σ). Η πιθανότητα α ν'απορριφθεί μια σωστή υπόθεση H_0 ονομάζεται στάθμη σημαντικότητας (σ.σ) του εφαρμοζόμενου test. Είναι δηλαδή:

$\alpha=P(H_0$ απορρίπτεται / H_0 σωστή$)$. (2.18)

Στα επόμενα θα γράφουμε: **(σ.σ): α** αντί της πρότασης: **σε στάθμη σημαντικότητας α.**

Την πιθανότητα α την ορίζει ο μελετητής αλλά στις περισσότερες εφαρμογές χρησιμοποιείται η **α=0.05** και αρκετά συχνά και οι τιμές α=0.01, α=0.025 και α=0.1. Ο αριθμός 1-α κατά τον Ψώινο (1978) ονομάζεται **συντελεστής εμπιστοσύνης.** Πολλές φορές χρησιμοποιείται, έστω και λανθασμένα, ο 1-α αντί του α. Δηλαδή, αντί του ορθού (σ.σ): α=0.05 χρησιμοποιείται το (σ.σ): α=0.95. Στην περίπτωση αυτή ο α εκφράζει την πιθανότητα να δεχτούμε την H_0 όταν αυτή είναι σωστή.

Η πιθανότητα να γίνει σφάλμα τύπου II συμβολίζεται με β και ο αριθμός **γ=1-β** ονομάζεται **ισχύς του test.** Είναι λοιπόν:

$\beta=P(H_0$ δεκτή / H_0 λανθασμένη$)$. (2.19)

Ένα test είναι **συνεπές** όταν αφήνει μικρά περιθώρια για σφάλμα τύπου I. Επειδή το σφάλμα τύπου II είναι δύσκολο να βρεθεί, όταν γίνεται δεκτή η μηδενική υπόθεση H_0 χωρίς να έχει υπολογιστεί το β πρέπει στο συμπέρασμα να αναφέρεται **"γίνεται**

δεκτή η H_o γιατί δεν υπάρχει αρκετή πληροφορία για απόρριψή της" ή **"δεν υπάρχει αρκετή πληροφορία για την απόρριψη της H_o".** 'Οταν το συμπέρασμα του test είναι η απόρριψη της H_o λέμε: "Οι παρατηρηθείσες διαφορές είναι στατιστικά σημαντικές" που σημαίνει ότι τέτοιες διαφορές σαν αυτές που παρατηρήθηκαν δεν μπορεί να οφείλονται μόνο στην τύχη.

Γενικά:

(α) 'Οταν οι παρατηρούμενες διαφορές από την υποθετική τιμή αναμένεται να υπάρχουν σε περισσότερες από 5% ($\alpha>0.05$) φορές, τότε λέγονται **Στατιστικά Μή Σημαντικές (ΣΜΣ).**

(β) 'Οταν αναμένονται μεταξύ 1% και 5% ($0.01<\alpha\leq0.05$), λέγονται **Στατιστικά Σημαντικές (ΣΣ)** και, τέλος,

(γ) όταν αναμένονται το πολύ 1% ($\alpha\leq0.01$) λέγονται **Στατιστικά Πολύ Σημαντικές (ΣΠΣ).**

2.6 Διάστημα Εμπιστοσύνης (Δ.Ε).

Σε πολλές περιπτώσεις δεν είναι δυνατό να διαθέτει κανείς ολόκληρο τον πληθυσμό για να υπολογίσει ακριβώς τις στατιστικές παραμέτρους της τυχαίας μεταβλητής που μελετά. 'Ετσι, ακόμα και σ'ένα πλήρες αρχείο με τις τιμές μιας μετεωρολογικής παραμέτρου ενός σταθμού, διατίθενται πεπερασμένου πλήθους τιμές της παραμέτρου (π.χ. τιμές 50 ετών). Οι στατιστικές, λοιπόν, παράμετροι που υπολογίζονται από τις τιμές αυτές δεν είναι παρά προσεγγιστικές εκτιμήσεις των πραγματικών τιμών των παραμέτρων αυτών, όταν οι διατιθέμενες τιμές είναι πολλές και η προσέγγιση είναι τόσο καλύτερη όσο το πλήθος των διαθέσιμων τιμών μεγαλώνει. Στην παράγραφο 2.3.1 αναφέρθηκε ότι οι στατιστικές παράμετροι που υπολογίζονται με βάση τα διατιθέμενα στοιχεία ενός σταθμού θα αναφέρονται με το ονομά τους, δεχόμενοι ότι οι με τον τροπο αυτό υπολογιζόμενες τιμές τους αποτελούν τις καλύτερες εκτιμήσεις των αντίστοιχων παραμέτρων. 'Οταν όμως δεν διατίθεται αρκετά μεγάλο πλήθος τιμών μιας παραμέτρου ή όταν εργάζεται κανείς μ'ένα σχετικά μικρού μεγέθους δείγμα, τότε οι τιμές των στατιστικών παραμέτρων που υπολογίζει μπορεί να διαφέρουν από τις αντίστοιχες πραγματικές τιμές των παραμέτρων αυτών και μάλιστα οι διαφορές αυτές μπορεί, σε ορισμένες περιπτώσεις, να είναι αρκετά μεγάλες.

Για το λόγο αυτό εισάγεται η έννοια του διαστήματος εμπιστοσύνης.

Ορίζεται σαν διάστημα εμπιστοσύνης (Δ.Ε) μιας παραμέτρου το διάστημα $[\alpha_1,\alpha_2]$ εντός του οποίου οι τιμές της παραμέτρου πρέπει να κυμαίνονται με δοθείσα πιθανότητα. Αν με X παραστήσουμε την παράμετρο τότε το $[\alpha_1,\alpha_2]$ είναι το διάστημα εμπιστοσύνης της X με πιθανότητα 1-α (συντελεστής εμπιστοσύνης) αν ισχύει η σχέση:

$$P(\alpha_1 \leq X \leq \alpha_2)=1-\alpha. \qquad (2.20)$$

που μπορεί να γραφεί και ως εξής:

$$P(X \leq \alpha_1)+P(X \geq \alpha_2)=\alpha \qquad (2.21)$$

που σημαίνει ότι η πιθανότητα ύπαρξης τιμών έξω από το διάστημα D είναι α (σε αντιστοιχία με την στάθμη σημαντικότητας (σ.σ)).

Κατά τους Kendall και Stuart (1973), οι αριθμοί α_2 και α_1 ονομάζονται **ανώτερο και κατώτερο όριο εμπιστοσύνης**, αντίστοιχα. Αν συμβεί να είναι $P(X \leq \alpha_1)=P(X \geq \alpha_2)=\alpha/2$, τότε το διάστημα εμπιστοσύνης ονομάζεται **κεντρικό**, αλλοιώς ονομάζεται **μή κεντρικό**. Εις το εξής, τα διαστήματα εμπιστοσύνης θεωρούνται κεντρικά εκτός αν ρητά μνημονεύονται διαφορετικά.

Παράδειγμα 2.6.α.

Είναι γνωστό ότι αν οι τιμές μιας παραμέτρου X ακολουθούν την κανονική κατανομή $N(X,\mu,\sigma^2)$ τότε στο διάστημα:

$$D=[\mu-1.96*\sigma,\mu+1.96*\sigma]$$

ανήκει το 95% του συνόλου των τιμών της παραμέτρου. Έτσι, για την παράμετρο αυτή το D είναι το διάστημα εμπιστοσύνης για πιθανότητα α=0.05 και το ανώτερο και κατώτερο όριο εμπιστοσύνης είναι οι αριθμοί $\alpha_1=\mu-1.96*\sigma$ και $\alpha_2=\mu+1.96*\sigma$. Αν θέλαμε το διάστημα εμπιστοσύνης για πιθανότητα α=0.318, τότε αυτό θα ήταν το $D_1=[\mu-\sigma,\mu+\sigma]$, όπως προκύπτει από τον πίνακα της φυσικής (κανονικής) κατανομής (Πίνακας ΙΒ του παραρτήματος).

Γενικά, για την κανονική κατανομή και για συντελεστή εμπιστοσύνης 1-α βρίσκεται ο αριθμός $Z_{(1-\alpha)/2}$ από τον πίνακα 1 και το διάστημα εμπιστοσύνης είναι το:

$$D=[\alpha_1,\alpha_2]=[\mu-\sigma*Z_{(1-\alpha)/2}\ ,\ \mu+\sigma*Z_{(1-\alpha)/2}] \qquad (2.22)$$

2.7 Κατανομές δειγματικών παραμέτρων.

Δίνονται παρακάτω τα πιο βασικά θεωρήματα της στατιστικής από τα οποία προκύπτει η κατανομή των στατιστικών παραμέτρων. Στα θεωρήματα αυτά στηρίζεται κατά το μάλλον ή ήττον η διαπίστωση της ομοιογένειας μιας χρονοσειράς ενώ μπορούν να χρησιμοποιηθούν και για τον έλεγχο απλών τιμών.

<u>Θεώρημα 1.</u> Αν οι τιμές μιας παραμέτρου X, από το οποίο προήλθε ένα δείγμα Y μεγέθους η, ακολουθεί την κανονική κατανομή $N(X,\mu,\sigma^2)$, τότε η E(Y) ακολουθεί επίσης την κανονική κατανομή και μάλιστα είναι $E(Y)\approx N(E(Y),\mu,\sigma^2/\eta)$.

Η πρόταση αυτή λέει ουσιαστικά ότι αν κάποιος πάρει πολυάριθμα δείγματα μεγέθους η, από τις τιμές μιας παραμέτρου X και θεωρήσει σαν τυχαία μεταβλητή την τυχαία μεταβλητή της οποίας οι τιμές είναι οι μέσοι όροι E(Y) των δειγμάτων αυτών, τότε οι τιμές της E(Y) ακολουθούν την κανονική κατανομή $N(E(Y),\mu,\sigma^2/\eta)$.

Το γεγονός αυτό βοηθά στην εύρεση του (Δ.Ε) με βάση την εξίσωση (2.24).

Παράδειγμα 2.7.α.

Η ελάχιστη θερμοκρασία σ'ένα σταθμό ακολουθεί την κανονική κατανομή $N(T_{min},\mu,\sigma^2)$. Από τις διατιθέμενες τιμές των τριάντα τελευταίων ετών που θεωρούμε σαν ένα δείγμα μεγέθους $\eta=30*365=10950$ βρέθηκε ότι $E(T_{min})=12°C$ και $\sigma^2=64$. Ζητούμε να βρούμε το διάστημα εμπιστοσύνης του μ για πιθανότητα 0.95.

Είναι: $\sigma=8$, $\eta^{0.5}\approx104$, οπότε το διάστημα εμπιστοσύνης $D=[\alpha_1,\alpha_2]$ θα είναι το $D=[11.85,12.15]$ διότι :

$\alpha_1=E(T_{min})-1.96*(\sigma^2/\eta)^{0.5}=12-1.96*8/104\approx11.85$ και:

$\alpha_2=E(T_{min})+1.96*(\sigma^2/\eta)^{0.5}\approx12.15$.

Παρατηρεί, λοιπόν, κανείς ότι αν θεωρηθούν τυχαία δείγματα μεγέθους $\eta=10950$, τότε η μέση τιμή της παραμέτρου T_{min}, με πιθανότητα $1-\alpha=0.95$, δεν διαφέρει περισσότερο των 0.15°C.

Γενικά, για πιθανότητα $1-\alpha=0.95$ η αναμενόμενη τιμή της μέσης τιμής μ μιας παραμέτρου X με τυπική απόκλιση σ θα βρίσκεται στο διάστημα:

$$D=[E(X)-1.96*\sigma/\sqrt{\eta}\ ,\ E(X)+1.96*\sigma/\sqrt{\eta}] \qquad (2.23)$$

όπου E(X) η μέση τιμή ενός δείγματος μεγέθους **η** (που υποθέτουμε ότι είναι αρκετά μεγάλο) και **σ** η τυπική απόκλιση της παραμέτρου που εδώ υποθέτουμε ότι συμπίπτει μ'αυτή του δείγματος.

'Οταν, λοιπόν, το **η** είναι αρκετά μεγάλο, τότε μπορεί κανείς να υποθέσει ότι μ=E(X), γιατί τότε ο αριθμος $1.96*\sigma/\sqrt{n}$ είναι πολύ μικρος. Στο γεγονός αυτό οφείλεται η παραδοχή που έγινε στην αρχή, σύμφωνα με την οποία, όταν μια στατιστική παράμετρος υπολογίζεται από το σύνολο των διατιθέμενων τιμών ονομάζεται με το όνομά της, παρόλο που δεν είναι η πραγματική τιμή της παραμέτρου αυτής αλλά μια προσεγγιστική εκτίμησή της.

Εάν ήταν γνωστά τα μ και σ της τυχαίας μεταβλητής $X \approx N(X,\mu,\sigma^2)$ και ζητούσαμε διαστήμα εμπιστοσύνης για τους μέσους όρους E(X) δειγμάτων μεγέθους n, αυτό θα είχε την ίδια ακριβώς μορφή με το D που ορίσαμε για το μ, μόνο που στη θέση του E(X) είναι το μ:

$D=[\mu-Z_{(1-\alpha)/2}*\sigma/\sqrt{n}\ ,\ \mu+Z_{(1-\alpha)/2}*\sigma/\sqrt{n}]$ **(2.24)**

Παράδειγμα 2.7.β.

Η πίεση P στην μέση στάθμη θαλάσσης την 06UTC κατά το i ($i \in I_{36}$) δεκαήμερο του έτους ακολουθεί κανονική κατανομή, $P \approx N(P,\mu,\sigma^2)$. Αν το αρχείο μας περιέχει παρατηρήσεις 30 ετών, τότε οι διατιθέμενες τιμές κατά το πρώτο δεκαήμερο θα είναι n=300-κ, όπου κ το πλήθος των ελλειπουσών τιμών, που υποθέτουμε ίσο με μηδέν. Τα όρια του διαστήματος εμπιστοσύνης D του μ, για α=0.05 είναι τώρα το $D=[E(P)-1.96*\sigma/\sqrt{300}\ ,\ E(P)+1.96*\sigma/\sqrt{300}]$, όπου E(P) ο δειγματικός μέσος όρος που προέκυψε από τις 300 τιμές. Αν $\sigma^2=81$, τότε επειδή $1.96*\sigma/\sqrt{300} \approx 1$ το διάστημα εμπιστοσύνης γίνεται:

D=[E(P)-1,E(P)+1]

Το παραπάνω θεώρημα εφαρμόζεται για μεγάλα δείγματα **(n>30)**. 'Όταν το μέγεθος του δείγματος είναι μικρότερο του 30 (μικρό δείγμα), τότε χρησιμοποιείται το θεώρημα 2.

<u>Θεώρημα 2.</u> Αν η κατανομή μιας παραμέτρου X είναι η κανονική με μέση τιμή μ και τυπική απόκλιση σ (δηλαδή αν $X \approx N(X,\mu,\sigma^2)$) και θεωρήσουμε ένα τυχαίο δείγμα Y μεγέθους n με μέση τιμή E(Y) και τυπική απόκλιση S (ή S'), τότε η τυχαία μεταβλητή:

$t = [\mu-E(Y)]/(S'/\sqrt{n}) = [\mu-E(Y)]/[S/(n-1)^{0.5}]$ **(2.25)**

ακολουθεί την t-κατανομή με v=n-1 βαθμούς ελευθερίας.

Στην (2.25) η S' είναι η τυπική απόκλιση που υπολογίστηκε από το δείγμα με βάση τις (2.10) και (2.11). Η τυπική απόκλιση S' που υπολογίζεται με βάση την (2.11) είναι ένας αβίαστος εκτιμητής της τυπικής απόκλισης σ του πληθυσμού.

Αν το n είναι αρκετά μεγάλο (n>30), τότε οι τιμές της t-κατανομής συμπίπτουν με τις τιμές της φυσικής κατανομής και η πρόταση 2 ισοδυναμεί με την πρόταση 1, όπως αναφέρθηκε παραπάνω.

Ο Πίνακας III δίνει τις τιμές του $t_{v,\alpha/2}$ της t_v-κατανομής για τιμές του v από 1 έως 30 και για διάφορες τιμές του α.

Για την εύρεση του διαστήματος εμπιστοσύνης του μ, δοθέντος του α, βρίσκουμε από τον Πίνακα III το t_v που αντιστοιχεί στο α/2 και το D, για την μέση τιμή μ του πληθυσμού, δίνεται από την σχέση:

$$D=[E(Y)-t_{v,\alpha/2}*S'/\sqrt{n}\ ,\ E(Y)+t_{v,\alpha/2}*S'/\sqrt{n}] \qquad (2.26)$$

Εάν είναι γνωστή η μ και ζητείται το διάστημα εμπιστοσύνης για την E(Y), αυτό προκύπτει με αντικατάσταση του E(Y) με μ στον παραπάνω τύπο, δηλαδή:

$$D=[\mu-t_{v,\alpha/2}*S'/\sqrt{n}\ ,\ \mu+t_{v,\alpha/2}*S'/\sqrt{n}] \qquad (2.27)$$

Παράδειγμα 2.7.γ.

Τον Ιανουάριο του 1990, σ'ένα σταθμό, οι 31 τιμές της ελάχιστης θερμοκρασίας έδωσαν: E(X)=9°C, $(S')^2$=12°C.

Αν η ελάχιστη θερμοκρασία του Ιανουάριου στο σταθμό ακολουθεί κανονική κατανομή με μ=8°C και άγνωστη σ, θα βρεθούν:

α. Το διάστημα εμπιστοσύνης της E(X).

β. Το διάστημα εμπιστοσύνης του μ (όταν αυτός είναι άγνωστος και η παράμετρος ακολουθεί κανονική κατανομή).

γ. Αν η διαφορά μ-E(X) είναι σημαντική σε (σ.σ): α=0.05.

Όπως είδαμε το διάστημα εμπιστοσύνης D για την μέση τιμή ενός δείγματος μεγέθους n, όταν είναι γνωστός ο μέσος όρος του πληθυσμού, δίνεται από την (2.26):

$$D=[\mu-t_{v,\alpha/2}*S'/\sqrt{n}\ ,\ \mu+t_{v,\alpha/2}*S'/\sqrt{n}]$$

Για v=n-1=31, α=0.05 από τον πίνακα 2 του t-test βρίσκεται: $t_{30,0.025}$=1.96 και άρα **D=[6.78 , 9.22]**.

Θεωρώντας άγνωστη την μέση τιμή του πληθυσμού μ, το διάστημα εμπιστοσύνης εντός του οποίου αυτή ανήκει σε (σ.σ): α, δίνεται από την (2.25):

$$D=[E(X)-t_{v,\alpha/2}*S'/\sqrt{n}\ ,\ E(X)+t_{v,\alpha/2}*S'/\sqrt{n}]$$

οπότε: **D=[7.78 , 10.22]**.

Αφού E(X)∈D, συμπεραίνεται ότι η διαφορά δ=|8-9|°C δεν είναι σημαντική και διαφορές αυτού του μεγέθους μπορούν να χαρακτηρίζονται τυχαίες (διαφορά ΣΜΣ).

<u>Θεώρημα 3.</u> Αν οι τυχαίες μεταβλητές X_1 και X_2 ακολουθούν την κανονική κατανομή με μέσους όρους μ_1 και μ_2 αντίστοιχα και <u>κοινή τυπική αποκλίση σ</u>, τότε αν Y_1 και Y_2 είναι δύο τυχαία δείγματα μεγέθους n_1 και n_2 και αντίστοιχες μέσες τιμές και τυπικές αποκλίσεις τις $E(Y_1)$, $E(Y_2)$ και $S_{1,2}$ τότε η τυχαία μεταβλητή:

$$t=\{(\mu_1-\mu_2)-[E(Y_1)-E(Y_2)]\}/[S_v*(1/n_1+1/n_2)^{0.5}] \qquad (2.28)$$

όπου:

$$S_v^2 = [(n_1-1)*S_1^2+(n_2-1)*S_2^2]/(n_1+n_2-2) \qquad (2.29)$$

ακολουθεί την t-κατανομή με $v=n_1+n_2-2$ βαθμούς ελευθερίας.

Το διάστημα εμπιστοσύνης της διαφοράς $\mu_1-\mu_2$ δίνεται όπως αυτό του μ στην πρόταση 2 και είναι το:

$$D=[E(Y_1)-E(Y_2)-t_{v,\alpha/2}*SS \;,\; E(Y_1)-E(Y_2)+t_{v,\alpha/2}*SS] \qquad (2.30)$$

όπου:

$$SS = S_v*(1/n_1+1/n_2)^{0.5} \qquad (2.31)$$

Το θεώρημα αυτό χρησιμοποιείται και για τον έλεγχο της διαφοράς των μέσων τιμών δύο δειγμάτων που λήφθηκαν από τον ίδιο πληθυσμό. Στην περίπτωση αυτή $\mu_1=\mu_2$.

Εάν τα δείγματα είναι μεγάλα (δηλαδή τα n_1 και n_2 είναι μεγάλα), τότε η τυχαία μεταβλητή:

$$\Xi=\{(\mu_1-\mu_2)-[E(Y_1)-E(Y_2)]\}/S_v \qquad (2.32)$$

ακολουθεί την κανονική κατανομή $\Xi \approx N(\Xi,(\mu_1-\mu_2),S_v^2)$, όπου η S_v εκτιμάται από την:

$$S_v^2 = S_1^2/n_1 + S_2^2/n_2 \qquad (2.33)$$

Εαν οι πληθυσμοί έχουν διαφορετικές τυπικές αποκλίσεις, τότε η τυχαία μεταβλητή:

$$t_a=\{(\mu_1-\mu_2)-[E(Y_1)-E(Y_2)]\}/S_v \qquad (2.34)$$

ακολουθεί την t-κατανομή, με $S_v^2=S_1^2/n_1+S_2^2/n_2$ και η απόρριψη της μηδενικής υπόθεσης H_o (που μπορεί να είναι $\mu_1=\mu_2$ στην περίπτωση που τα δείγματα προέρχονται από τον ίδιο πληθυσμό ή $\mu_1-\mu_2=\delta$) βασίζεται στο κατά πόσο η ποσότητα $|t_a|$ είναι μεγαλύτερη ή μικρότερη από την ποσότητα:

$$t=(t_1*S_1^2/n_1+t_2*S_2^2/n_2)/S_v^2 \qquad (2.35)$$

όπου t_1 και t_2 οι τιμές των t από τον Πίνακα III για $v_1=n_1-1$ και $v_2=n_2-1$ βαθμούς ελευθερίας και για (σ.σ): α/2. Η H_o

απορρίπτεται όταν $|t|>t_{v,\alpha/2}$

Το διάστημα εμπιστοσύνης της διαφοράς $\delta=\mu_1-\mu_2$ δίνεται από την:

$D=[E(Y_1)-E(Y_2)-t_{v,\alpha/2}*S_v\ ,\ E(Y_1)-E(Y_2)+t_{v,\alpha/2}*S_v$ **(2.36)**

Σε περίπτωση που είναι γνωστές οι μ_1 και μ_2, το διάστημα εμπιστοσύνης της διαφοράς $E(Y_1)-E(Y_2)$ θα δίνεται από την:

$D=[(\mu_1-\mu_2)-t_{v,\alpha/2}*S_v\ ,\ (\mu_1-\mu_2)+t_{v,\alpha/2}*S_v$ **(2.37)**

όπου και πάλι η S_v υπολογίζεται με βάση την (34).

Παράδειγμα 2.7.6.

Η μέση τιμή και η τυπική απόκλιση των 31 ημερήσιων τιμών του Ιανουάριου τα έτη 1959 και 1990 για την ελάχιστη θερμοκρασία ήταν $E(X_{59})=7.5°C$, $E(X_{90})=9°C$ και $S_{59}^2=16$, $S_{90}^2=12$.

Είναι σημαντική η διαφορά των μέσων τιμών $\delta=|7.5-9|=1.5$;

Υποθέτουμε ότι τα δυο δείγματα προέρχονται από τον ίδιο πληθυσμό (ελάχιστες θερμοκρασίες Ιανουαρίου) που ακολουθεί κανονική κατανομή. Είναι λοιπόν $\mu_1=\mu_2$, $v=n_1+n_2-2=31+31-2=60$, $S_v^2=[S_{59}^2*(n_1-1)+S_{90}^2*(n_2-1)]/(n_1+n_2-2)=14$ και $S=S_v*(1/n_1+1/n_2)=14*2/31=0.9$. Για $v=60$ βαθμούς ελευθερίας και $1-\alpha=0.95$ ($\alpha=0.05$ και $\alpha/2=0.025$) από τον Πίνακα III βρίσκεται $t=t_{60,0.025}=1.96$ και άρα $S*t=1.765$. Διαφορές λοιπόν των μέσων τιμών των δειγμάτων μεγέθους $n=30$ (31, ή και 28 ή 29 που είναι οι μέρες του μήνα) μέχρι την τιμή 1.765 θεωρούνται ΣΜΣ και μπορούν να οφείλονται στην τύχη. Η H_0 στην περίπτωση αυτή ήταν: H_0 $E(X_{59})=E(X_{90})$. Το διάστημα εμπιστοσύνης για την διαφορά $\delta=|E(X_{59})-E(X_{90})|$ είναι το $D=[-1.765\ ,\ 1.765]$. Οι δείκτες 59 και 90 που παριστούν τα έτη μπορούν να αντικατασταθούν από άλλα έτη και να βρεθεί κατά πόσο υπάρχει κάποια σημαντική διαφορά μεταξύ των μέσων όρων των τιμών των μεταβλητών. Οι συγκρίσεις των μέσων τιμών με εφαρμογή των προτάσεων 2 και 3, συνδυαζόμενες με την μέθοδο του 1/1000 ή του 1/100 μπορούν σε αρκετές περιπτώσεις να δείξουν πιθανές λανθασμένες τιμές στο αρχείο.

<u>Θεώρημα 4</u>. Αν η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή ($X\approx N(X,\mu=1,\sigma^2=1)$), τότε η τυχαία μεταβλητή $X^2=X_1^2$ ακολουθεί την Χι-τετράγωνο κατανομή με ένα βαθμό ελευθερίας.

<u>Θεώρημα 5</u>. Αν οι τυχαίες μεταβλητές $Z_1,Z_2,..,Z_\kappa$ ακολουθούν την κανονική κατανομή με μέση τιμή $\mu=0$ και διακύμανση $\sigma^2=1$, (δηλαδή αν $Z_i^2 \approx N(Z_i^2,\mu=0,\sigma^2=1), i\in I_\kappa$) <u>και αν είναι στοχαστικά ανεξάρτητες</u> μεταξύ τους, τότε οι τυχαία μεταβλητή X^2 όπου:

$X^2={}_1\Sigma^\kappa Z_i^2=Z_1^2+Z_2^2+...+Z_\kappa^2$

ακολουθεί την Χι-τετράγωνο κατανομή με ν=κ βαθμούς ελευθερίας.

<u>Πόρισμα</u>. Αν οι τυχαίες μεταβλητές $X_1,X_2,..,X_\kappa$ ακολουθούν την κανονική κατανομή με μέση τιμή μ_i και διακύμανση σ_i^2, (δηλαδή αν $X_i^2 \approx N(X_i^2,\mu_i,\sigma_i^2), i\in I_\kappa$) και αν είναι στοχαστικά <u>ανεξάρτητες</u> μεταξύ τους, τότε οι τυχαία μεταβλητή:

$X^2={}_1\Sigma^\kappa Y_i^2=Y_1^2+Y_2^2+...+Y_\kappa^2$

όπου

$Y_i=(X_i-\mu_i)/\sigma_i^2$.

ακολουθεί την Χι-τετράγωνο κατανομή με ν=κ βαθμούς ελευθερίας.

<u>Θεώρημα 6</u>. Αν οι τυχαίες μεταβλητές X_1 και X_2 είναι <u>ανεξάρτητες</u> και η X_1 ακολουθεί την κατανομή Χι-τετράγωνο με ν=κ βαθμούς ελευθερίας και η τυχαία μεταβλητή X_2 την κατανομή Χι-τετράγωνο με ν=λ βαθμούς ελευθερίας, τότε η τυχαία μεταβλητή $X=X_1+X_2$ ακολουθεί την κατανομή Χι-τετράγωνο με ν=κ+λ βαθμούς ελευθερίας.

Όπως αναφέρθηκε στην παράγραφο 2.1 τα σφάλματα για να είναι τυχαία πρέπει να είναι ανεξάρτητα μεταξύ τους. Αν σαν σφάλμα θεωρηθεί η απόκλιση των συχνοτήτων Π_i των τιμών μιας τυχαίας μεταβλητής στο διάστημα A_i, ενός διαμελισμού Q του συνόλου των τιμών της, από τις αντίστοιχες θεωρητικές συχνότητες Θ_i στο ίδιο διάστημα, τότε οι τυχαίες μεταβλητές που ορίζονται από τις σχέσεις:

$$X_i^2=(\Theta_i-\Pi_i)^2/\Theta_i, i\in I_m \qquad (2.38)$$

όπου m το πλήθος των διαστημάτων του διαμελισμού Q, ακολουθούν την Χι-τετράγωνο κατανομή με **ν=1** βαθμό ελευθερίας. Η τυχαία μεταβλητή που προκύπτει από το άθροισμα των m αυτών ανεξάρτητων μεταβλητών σύμφωνα με τα παραπάνω θεωρήματα θα ακολουθεί επίσης την Χι-τετράγωνο κατανομή με **ν=m-1-k k≥0** βαθμούς ελευθερίας. Από το m αφαιρείται η μονάδα διότι το άθροισμα των πραγματικών και των θεωρητικών συχνοτήτων πρέπει να είναι το ίδιο και ίσο με το μέγεθος του δείγματος και το πλήθος των παραμέτρων που απαιτούνται στον ορισμό της θεωρητικής συχνότητας k. Εάν συγκρίνονται συχνότητες μιας παραμέτρου από διαφορετικούς σταθμούς, εκ των οποίων ο ένας θεωρείται μάρτυρας (αξιόπιστος) και επομένως οι τιμές του αντιμετωπίζονται σαν οι θεωρητικές, τότε είναι k=0.

<u>Θεώρημα 7.</u>`Εστω ότι η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με μέση τιμή μ και διασπορά σ^2 ($X \approx N(x,\mu,\sigma^2)$). Αν E(X) και S^2 είναι ο μέσος όρος και η διασπορά ενός τυχαίου δείγματος μεγέθους n της X, τότε η τυχαία μεταβλητή:

$$X^2 = \frac{(n-1)*S^2}{\sigma^2} \qquad (2.39)$$

ακολουθεί την X^2-κατανομή με v=n-1 βαθμούς ελευθερίας.

<u>Θεώρημα 8.</u> Αν X_1, X_2 είναι δυο ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την X^2 κατανομή με v_1, v_2 βαθμούς ελευθερίας αντίστοιχα, τότε η τυχαία μεταβλητή:

$$F = \frac{X_1/v_1}{X_2/v_2} \qquad (2.40)$$

ακολουθεί την F-κατανομή με v_1 και v_2 βαθμούς ελευθερίας δηλαδή:

$X \approx F(v_1, v_2)$

Τα θεωρήματα 7 και 8 χρησιμοποιούνται για τον έλεγχο της ομοιογένειας μιας χρονοσειράς και τον έλεγχο των απλών τιμών με τρόπο που θα συζητηθεί στα ανάλογα εδάφια.

2.8 Το X^2-test σαν test προσαρμογής και ανεξαρτησίας.

Το X^2 (Χι τετράγωνο) test χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας των διαφορών μεταξύ παρατηρούμενων και θεωρητικών συχνοτήτων. Σε αντίθεση με το t-test και το test για την φυσική κατανομή, το X^2 test μπορεί να εφαρμοστεί και μεταξύ περισσότερων των δυο μεγεθών, αρκεί τα μεγέθη αυτά να εκφράζουν συχνότητες. Χρησιμοποιείται μάλιστα τόσο για ποιοτικό όσο και ποσοτικό έλεγχο των δεδομένων.

Το X^2 υπολογίζεται από την σχέση:

$$X^2={}_1\Sigma^{\kappa}X_i^2 \quad (2.41)$$

όπου:

$$X_i^2=(\Pi_i-\Theta_i)^2/\Theta_i \quad (2.42)$$

και Π_i, Θ_i οι παρατηρηθείσες και οι θεωρητικές συχνότητες σε κάθε μια από τις κ κλάσεις στις οποίες διαμελίζεται το σύνολο των τιμών της υπό εξέταση παραμέτρου.

Οι βαθμοί ελευθερίας (ΒΕ) ν κατά την εφαρμογή του X^2 test είναι:

$$\nu=\kappa-\kappa_1 \quad (2.43)$$

όπου $\kappa_1 \geq 1$ το πλήθος των περιορισμών που προκύπτουν κατά τον υπολογισμό των θεωρητικών συχνοτήτων. Το $\kappa_1=1$ όταν ο μόνος περιορισμός (περιορισμός που υπάρχει πάντα στην εφαρμογή του X^2 test) είναι η χρησιμοποίηση του πλήθους Κ των τιμών του δείγματος (ή του πληθυσμού) και για τις τιμές της θεωρητικής συχνότητας, δηλαδή:

$${}_1\Sigma^{\kappa}\Pi_i={}_1\Sigma^{\kappa}\Theta_i=K. \quad (2.44)$$

Για την προσαρμογή δεδομένων στην φυσική κατανομή είναι $\kappa_1=3$, διότι χρησιμοποιείται το ίδιο πλήθος (απώλεια ενός βαθμού ελευθερίας) και οι παράμετροι μ,σ (ή Ε(Χ), S_x) (απώλεια δύο ακόμη βαθμών ελευθερίας).

Όταν το X^2-test εφαρμόζεται σαν test ανεξαρτησίας των δειγμάτων (ουσιαστικά σαν test ομοιογένειας της χρονοσειράς), τότε οι βαθμοί ελευθερίας (ΒΕ) υπολογίζονται από την:

$$\nu=(\nu_1-1)*(\nu_2-1)*...*(\nu_p-1) \quad (2.45)$$

όπου $\nu_1,\nu_2...\nu_p$ οι καταστάσεις-διαστάσεις του πίνακα συνάφειας (περιγράφεται παρακάτω), που δημιουργείται για την μελέτη της ομοιογένειας αυτής.

Η υπόθεση H_o που γίνεται κατά την εφαρμογή του X^2-test είναι:

$H_{o\pi}$: Προσαρμόζονται οι τιμές της τυχαίας μεταβλητής Χ (παραμέτρου) στην θεωρητική κατανομή των συχνοτήτων F(X); (test προσαρμογής).

$H_{o\alpha}$: Είναι τα χαρακτηριστικά ως προς τα οποία εξετάζεται και κατανέμεται ο πληθυσμός στον πίνακα συνάφειας ανεξάρτητα μεταξύ τους σε (σ.σ): α=0.05 (ή άλλη τιμή του α που ορίζει ο μελετητής); (test ανεξαρτησίας).

Η H_1 (εναλλακτική της H_o) είναι η άρνηση της H_o.

Για να γίνει δεκτή μια από τις H_o ή H_1 υπολογίζεται το X^2 με βάση τις (2.40) και (2.39) και βρίσκεται η τιμή του θεωρητικού $X^2_{v,\alpha}$ από τους Πίνακες IIα,IIβ της κατανομής $X^2_{v,\alpha}$, ($X^2_{v,\alpha/2}$ για δίπλευρο test, που σχεδόν ποτέ δεν εφαρμόζεται στο X^2 test).

Αν συμβεί $X^2 > X^2_{v,\alpha}$ απορρίπτεται η H_o γιατί τιμές του X^2 μεγαλύτερες του $X^2_{v,\alpha}$ δεν μπορεί να οφείλονται μόνο στην τύχη.

2.8.1 Παρατηρήσεις κατά την εφαρμογή του X^2-test.

α. Οι θεωρητικές συχνότητες όταν το D=MMX διαμελίζεται σε δύο μόνο υποσύνολά-κλάσεις πρέπει να είναι μεγαλύτερες ή ίσες του δέκα (10). Για περισσότερες από δύο κλάσεις γίνονται δεκτές θεωρητικές συχνότητες μεγαλύτερες ή ίσες του πέντε (5). Αν βρεθούν κλάσεις με συχνότητα μικρότερη του πέντε αυτές ενοποιούνται με τις γειτονικές τους, ή διαμελίζεται διαφορετικά το D (σε μικρότερο αριθμό κλάσεων).

β. Οι θεωρητικές συχνότητες $\theta_{i,j}$ που προκύπτουν από τους πίνακες συνάφειας πρέπει να είναι μεγαλύτερες ή ίσες του ένα (1) . Ο Βασιλείου (1985) επιτρέπει ακόμα και την μηδενική τιμή στις $\theta_{i,j}$ υπολογίζοντας όμως το X^2 από το άθροισμα των $X^2_{i,j}$ στα οποία $\theta_{i,j} \neq 0$, οπότε κατεβάζει και τους βαθμούς ελευθερίας κατά τον αριθμό των περιπτώσεων στις οποίες $\theta_{i,j}=0$.

γ. Ο έλεγχος με το X^2 test είναι προτιμότερο να γίνεται χρησιμοποιώντας τις απόλυτες και όχι τις επί τα εκατό σχετικές συχνότητες.

δ. Το μέγεθος του χρησιμοποιούμενου δείγματος πρέπει να είναι μεγαλύτερο του πενήντα (50).

Παράδειγμα 2.8.α.

Εφαρμογή του X^2 test σαν test προσαρμογής σε θεωρητική συχνότητα.

Οι συχνότητες που παρατηρήθηκαν στις 10 κλάσεις στις οποίες διαμελίσαμε το σύνολο [0,20] στο οποίο ανήκαν οι 360 τιμές της μέσης τιμής της ημερήσιας θερμοκρασίας κατά το πρώτο δεκαήμερο του έτους, κατά τα έτη από 1954 έως 1990 (υπήρχαν 10 ελλείπουσες τιμές στο αρχείο) δίνονται στον Πίνακα 2.1.

θα ελεγχθεί η προσαρμογή των συχνοτήτων αυτών για :

α. Ομοιόμορφη κατανομή.

β. Τριγωνική κατανομή, για ισοσκελές τρίγωνο βάσης [0,20].

γ. Κανονική κατανομή, με $E(X)=E(T)=10$ και $S^2=S^2(T)=16$.

Στην ομοιόμορφη κατανομή η συχνότητα είναι ανάλογη του μήκους της κλάσης και αφού ορίσαμε κλάσεις ίσου μήκους (2°C) σε κάθε κλάση θα υπάρχουν 360/10=36 τιμές. Για την ισοσκελή τριγωνική κατανομή οι συχνότητες πριν απο το σημείο (20-0)/2=10 και μετά απ'αυτό πρέπει να είναι ίσες μεταξύ τους (ισοσκελές τρίγωνο==>συμμετρικό ως προς την διάμεσο). Ακόμα στο 0 που είναι το άκρο του διαστήματος δεν πρέπει να υπάρχουν τιμές (αλλιώς δεν θα ήταν τριγωνική). 'Αρα η ευθεία που ενώνει την τρίτη κορυφή του τριγώνου με την αρχή των αξόνων θα έχει την μορφή $y=a*x$ και θα πρέπει να ισχύει:

$$180 = 360/2 = 360*F(x\leq 10) = 360*F(0\leq x\leq 10)/2 = \int_0^{10} axdx=100*a/2$$

οπότε a=3.6.

Με βάση την τιμή αυτή του a βρίσκουμε τις συχνότητες που αντιστοιχούν σε κάθε μια απο τις δέκα καταστάσεις στις οποίες διαμελίστηκε το σύνολο [0,20] που είναι οι θεωρητικές συχνότητες για την κατανομή αυτή. Π.χ στο διάστημα [2,4] θα υπάρχουν:

$$\int_2^4 3.6*x*dx=3.6*(16-4)/2=21.6 \text{ τιμές.}$$

Με τον ίδιο τρόπο υπολογίζονται και οι συχνότητες των κλάσεων από 1 μέχρι 5 και κατόπιν λόγω της συμμετρίας (στο συγκεκριμένο παράδειγμα) της κατανομής και οι υπόλοιπες συχνότητες. Τις τιμές αυτές τις βρίσκει κανείς στον πίνακα που ακολουθεί μαζί με τις τιμές της ομοιόμορφης κατανομής. Στον ίδιο πίνακα έχουν τοποθετηθεί και οι τιμές $X^2_i=(\Pi_i-\Theta_i)^2/\Theta_i$.

Για την εύρεση των συχνοτήτων Θ_i στην περίπτωση της κανονικής κατανομής, θεωρούμε την τυχαία μεταβλητή $Z=[X-E(X)]/S$ (τυπικός μετασχηματισμός της τυχαίας μεταβλητής X) και τις τιμές Z_i που αντιστοιχούν στα άκρα των διαστημάτων-κλάσεων X_i, $i\in I_{11}$ (δηλαδή τα σημεία $x_1=0$, $x_2=2$,...,$x_{11}=20$). Η Z ακολουθεί την κανονική κατανομή όταν η X ακολουθεί την κανονική κατανομή και μάλιστα τότε είναι $Z\approx N(Z,0,1)$. Απο τον πίνακα της κανονικής κατανομής βρίσκονται οι πιθανότητες που αντιστοιχούν στα σημεία Z_i και απο αυτές οι συχνότητες που αντιστοιχούν σε κάθε κλάση.

Εξ αιτίας της συμμετρίας της κανονικής κατανομής και επειδή δεχτήκαμε E(X)=10°C που είναι το μέσο του διαστήματος [0,20], βρίσκουμε τις συχνότητες μόνο για X≥0 (ή Z≥0). Δίνεται ο τρόπος υπολογισμού της θεωρητικής συχνότητας στα διαστήματα [10,12) και [12,14). Από τον Πίνακα ΙΒ της φυσικής κατανομής για Z=0.5 αντιστοιχεί πιθανότητα 0.1915 και για Z=1 πιθανότητα 0.3413. Τις τιμές αυτές μετατρέπουμε σε συχνότητες για δείγμα μεγέθους n=360 και παίρνουμε:

$P(0<Z\leq 0.5)*n=0.1915*360=68.9$

$P(0<Z\leq 1)*n=0.3413*360=122.8$

οπότε: $P(0.5<Z\leq 1)*n=[P(0<Z\leq 1)-P(0.5<Z\leq 1)]*n=122.8-68.9=53.9$

Για το διάστημα [18,20] η συχνότητα δεν χρειάζεται να υπολογιστεί (στο συγκεκριμένο παράδειγμα) αφού μπορεί να προκύψει σαν διαφορά της συχνότητας δεξιά του μέσου όρου (εδώ 360/2=180 λόγω συμμετρίας) και της συχνότητας στο διάστημα [10,18) που είναι ίση με $P(0<Z\leq 2)*n=0.4772*360=171.7$, οπότε η ζητούμενη συχνότητα είναι ίση με 180-171.7=8.3.

Επειδή για την ομοιόμορφη κατανομή οι θεωρητικές συχνότητες υπολογίζονται χωρίς να υπολογιστούν κάποιες παράμετροι που συνδέονται με την κατανομή, αυτή αφαιρείται μόνο ένας βαθμός ελευθερίας και έχουμε (ΒΕ)$_o$: v=10-1=9.

Για την τριγωνική κατανομή απαιτήθηκε ο υπολογισμός του a και, επομένως, "χάθηκε" άλλος ένας βαθμός ελευθερίας και έτσι οι βαθμοί ελευθερίας είναι (ΒΕ)$_\tau$: v=10-2=8.

Τέλος, οι βαθμοί ελευθερίας για την κανονική κατανομή είναι v=10-3=7 διότι για τον υπολογισμό των θεωρητικών συχνοτήτων χρησιμοποιήθηκαν οι παράμετροι: n, E(X) και S.

Για (σ.σ): α=0.05 και v=9 ή v=8 ή v=7 είναι:

$X^2_{9,0.05}=16.919$, $X^2_{8,0.05}=15.5$ και $X^2_{7,0.05}=14.067$

Από την τελευταία σειρά του πίνακα βλέπουμε ότι η υπόθεση: "Η κατανομή των συχνοτήτων της μέσης ημερήσιας θερμοκρασίας είναι ομοιόμορφη" δεν ευσταθεί διότι $X^2=115 >> 16.919$.

Αντίθετα, οι υποθέσεις "Η κατανομή είναι τριγωνική (ισοσκελής) ή κανονική" μπορούν να γίνουν δεκτές σε σ.σ α=0.05, αφού $X^2=3.6 < 15.5$ και $X^2=4.79<14.067$.

Πινακας 2.1. Παρατηρηθεισες (Π_i) και θεωρητικές ($\Theta_{i,j}$) συχνοτητες (j=1 για ομοιόμορφη, j=2 για τριγωνική και j=3 για την φυσική κατανομή).

Α/Α	κλάση	Π_i	$\Theta_{i,1}$ Ομ.	$X^2{}_i$ ομ.	$\Theta_{i,2}$ Τριγ.	$X^2{}_i$ Τριγ	Z_i -2.5	$\Theta^2{}_{i,2}$ Μεταξύ.	$X^2{}_i$ Κανον
1	0 - 2	10	36	18.7	7.2	1.1	-2.0	8.3	0.35
2	2 - 4	18	36	9.0	21.6	0.6	-1.0	15.8	0.31
3	4 - 6	40	36	0.4	36.0	0.4	-1.0	33.1	1.44
4	6 - 8	50	36	5.4	50.4	0.0	-0.5	53.9	0.28
5	8 -10	70	36	32.1	64.8	0.4	0.0	68.9	0.02
6	10 -12	60	36	16.0	64.8	0.4	0.5	68.9	1.15
7	12 -14	52	36	7.1	50.4	0.1	1.0	53.9	0.07
8	14 -16	32	36	0.4	36.0	0.4	1.5	33.1	0.04
9	16 -18	20	36	7.1	21.6	0.1	2.0	15.8	1.12
10	18 -20	8	36	21.8	7.2	0.1	2.5	8.3	0.01
	Σύνολο	360	360	115	360	3.6		360	4.79

3. ΕΛΕΓΧΟΣ ΑΠΛΩΝ ΤΙΜΩΝ.

3.1.Γενικά.

Σαν **απλή τιμή** ορίζεται μια τυχαία τιμή που μπορει να πάρει μια μελετούμενη παράμετρος. Όπως προαναφέρθηκε δεν υπάρχουν κανόνες για τον έλεγχο της αξιοπιστίας των απλών τιμών και ο έλεγχος περιορίζεται κυρίως στον ποιοτικό που περιγράφεται στο τεύχος "Ανάπτυξη μεθόδων ποιοτικού ελέγχου χωροχρονικά σύμφωνα με τα διεθνή πρότυπα" του "ΥΔΡΟΣΚΟΠΙΟ". Ο έλεγχος αυτός γίνεται (τουλάχιστον για τα αρχεία της ΕΜΥ) αμέσως μετά την λήψη των τηλεγραφημάτων που περιέχουν τις παρατηρήσεις και δεν συμπεριλαμβάνει τον έλεγχο όλων των παραμέτρων.

Προτείνονται εδώ μερικοί έλεγχοι που μπορούν να γίνουν για ορισμένες μετεωρολογικές παραμέτρους. Οι παράμετροι αυτοί μπορεί να έχουν ή να μην έχουν ελεγχθεί κατά τον πρωτογενή έλεγχο.

Διόρθωση των απλών τιμών δεν είναι δυνατή παρά μόνο σε περίπτωση που διατίθενται τα αρχεία με τα πρωτογενή στοιχεία, με βάση τα οποία διορθώνονται οι τιμές όταν διαπιστωθεί ή υπάρχουν υποψίες ότι αυτές είναι λανθασμένες.

3.2.'Ελεγχος απλών τιμών παραμέτρων με φυσικά όρια.

Πολλές από τις μετεωρολογικές παραμέτρους έχουν από την φύση τους απόλυτα άκρες τιμές. Οι τιμές αυτές λαμβάνονται υπόψη κατά τον πρωτογενή έλεγχο και άν η τιμή μιας παραμέτρου με φυσικά όρια βρίσκεται εκτός των ορίων αυτών απορρίπτεται. Ο πρωτογενής έλεγχος δεν ελέγχει τις τιμές όλων των παραμέτρων που μετρούνται στους μετεωρολογικούς σταθμούς. 'Ετσι το κλάσμα ηλιοφάνειας n/N επαληθεύει την σχέση: $0 \leq n/N \leq 1$, ο λόγος R/Ra, όπου R η ηλιακή ακτινοβολία που μετράται σ'έναν τόπο μια ορισμένη μέρα και Ra η αντίστοιχη θεωρητική-αστρονομική τιμή της, επαληθεύει την ανάλογη σχέση $0 \leq R/Ra \leq 1$ κλπ. Πρέπει, λοιπόν, και για τις παραμέτρους αυτές να γίνει ανάλογος έλεγχος. Τονίζεται ότι ο έλεγχος αυτός για τις δυο παραπάνω παραμέτρους γίνεται στην ΕΜΥ a posteriori και, επομένως απαιτείται να γίνεται για ανάλογες τιμές σταθμών που προέρχονται από άλλους φορείς και δεν έχουν υποστεί ανάλογο έλεγχο.

Ακόμα, αν διατίθεται αρχείο με ωριαίες τιμές της ηλιοφάνειας πρέπει αυτές να είναι μικρότερες από την μονάδα και αν διατίθενται ανάλογες τιμές της ολικής ακτινοβολίας, αυτές πρέπει να είναι μικρότερες από τις αντίστοιχες θεωρητικές τιμές για την ίδια ώρα, τον ίδιο σταθμό και την ίδια ημερομηνία.

Πρέπει, τέλος, ν'αναφερθεί η περίπτωση κατά την οποία αντί της απλής τιμής (απλό ανάγνωσμα του οργάνου) μιας παραμέτρου, το αρχείο περιέχει αθροιστικές τιμές της παραμέτρου, του αθροίσματος υπολογιζόμενου από μια χρονική στιγμή t_o και μετά. Στην περίπτωση αυτή πρέπει οι μετέπειτα τιμές της παραμέτρου να είναι μεγαλύτερες των προηγούμενων τιμών τους. Σαν παράδειγμα αναφέρεται το αθροιστικό ύψος βροχής μιας ημέρας, που όπως είναι γνωστό για μια ημέρα (Day) μετράται από την 18UTC της προηγούμενης μέρας (Day-1 ημέρα) μέχρι την 18UTC της Day μέρας. Στην περίπτωση αυτή αν παραστήσουμε με BP_i $i\in I_{24}$ τις 24 ωριαίες αθροιστικές τιμές της βροχής ενός 24ώρου θα πρέπει να ισχύει: $BP_k < BP_\lambda$ όταν $k \leq \lambda$. Η ίδια ανισότητα πρέπει να ελέγχεται αν ισχύει για όλες τις παραμέτρους των οποίων οι τιμές είναι αθροιστικές, όπως για παράδειγμα οι τιμές του αθροιστικού ανεμομέτρου, τιμές που απαιτούνται για τον υπολογισμό της δυναμικής εξατμισοδιαπνοής με την μέθοδο του Penmann.

Ανάλογοι έλεγχοι μπορούν και πρέπει να γίνονται για τις παραμέτρους:

Σχετική υγρασία	**UU**	**($0 \leq UU \leq 100$)**
Υψος βροχής	**BP**	**($0 \leq BP$)**
Νέφωση	**N**	**($0 \leq N \leq 9$)**
Εξάτμιση	**Ev**	**($0 \leq Ev$)**

3.3. `Ελεγχος του 1/1000 του συνόλου των τιμών και του 1/100 των τιμών του μήνα ή του δεκαημέρου.

Οι τιμές των μετεωρολογικών παραμέτρων είναι γνωστό ότι λαμβάνονται με ορισμένη ακρίβεια η κάθε μια. `Ετσι, η θερμοκρασία και η πίεση στην επιφάνεια έχουν ακρίβεια 0.1°C και 0.1hPa αντίστοιχα, η θερμοκρασία στα ανώτερα στρώματα της ατμόσφαιρας 1°C, ο άνεμος 1Kt κλπ. Στο σύνολο λοιπόν MMX (ή mmX όταν πρόκειται για δείγμα) περιέχονται πεπερασμένου πλήθους τιμές (το πολύ **ΕΥΡΧ/ακρίβεια**). Μπορούμε λοιπόν να βρούμε την συνάρτηση αθροιστικής συχνότητας F(x) της παραμέτρου X και να ελέγχθούν για την ορθότητά τους (εφόσον τα γραπτά αρχεία του σταθμού υπάρχουν ακόμη) οι τιμές x της μελετούμενης παραμέτρου X οι οποίες επαληθεύουν τις σχέσεις:

$F(x \le x_o) \le 0.001$ ή $F(x \ge x_1) \ge 0.999$ (3.1)

δηλαδή να ελεγχθούν οι πολύ μικρές και οι πολύ μεγάλες τιμές των παραμέτρων που δεν έχουν φυσικά όρια (θερμοκρασία, πίεση, θερμοκρασία υγρού θερμομέτρου κλπ). Για τις τιμές που επαληθεύουν την μία από τις δύο παραπάνω σχέσεις αναγράφεται η ημερομηνία παρατήρησης, ώστε να μπορεί κανείς εύκολα να βρει τις αντίστοιχες τιμές που είχαν αναγραφεί στο σταθμό και να διαπιστωθεί, αν δεν έγινε κάποιο λάθος (τηλεπικοινωνιακό ή αναγραμματισμού) μέχρι της τελικής καταχώρησης της τιμής στα αρχεία από τα οποία η τιμή διαβάστηκε τελικά. Οι τιμές δεν απορρίπτονται εαν δεν διαπιστωθεί από τα αρχεία των σταθμών ότι είναι λανθασμένες, οπότε τότε μπορούν να αντικατασταθούν από τις σωστές τιμές.

Αν, από την χρονοσειρά των τιμών της παραμέτρου, δημιουργήσει κανείς τις δώδεκα (ή τριάντα έξη) χρονοσειρές με τιμές, τις τιμές της χρονοσειράς που ανήκουν στον ίδιο μήνα (ή δεκαήμερο), τότε η παραπάνω μέθοδος του 1/1000 μπορεί να εφαρμοστεί για τις τιμές της κάθε μιας από τις δώδεκα (ή τριανταέξη) νέες χρονοσειρές παίρνοντας τώρα αντί το 1/1000, το 1/100 των τιμών που βρίσκονται πλησιέστερα στις τιμές $MINX_ι$ και $MAXX_ι$ $ι \in I_{12}$ ($ι \in I_{36}$), όπου $MINX_ι$ και $MAXX_ι$ οι ελάχιστες και μέγιστες τιμές των νέων χρονοσειρών.

Ο έλεγχος γίνεται τώρα για τις τιμές x της παραμέτρου που κατά τον μήνα (το δεκαήμερο) ι, επαληθεύουν τις σχέσεις:

$F(x \le x_o) \le 0.01$ και $F(x_o \ge x) \ge 0.99$. (3.2)

Οι τιμές της παραμέτρου οι μικρότερες της x_o και οι μεγαλύτερες της x_1 θεωρούνται σαν "ύποπτες" και πρέπει να ελεγθούν. Τις τιμές X_o και x_1 τις ονομάζουμε κρίσιμα όρια (κάτω και πάνω).

Παράδειγμα 3.1.α.

Οι ανηγμένες στην μέση στάθμη θαλάσσης τιμές της πίεσης P (σε hPa) σ'ένα σταθμό ανήκουν στο σύνολο-διάστημα MMX=[982,1052]. Η συνάρτηση F(P) έδωσε: $F(P \le 991) < 0.001$ και $F(P \ge 1047) \ge 0.999$. Οι "ύποπτες" για έλεγχο τιμές είναι λοιπόν οι τιμές οι μικρότερες των 991 (κάτω κρίσιμο όριο) και οι μεγαλύτερες των 1047hPa (άνω κρίσιμο όριο).

Στον ίδιο σταθμό κατά τον Αύγουστο οι τιμές της πίεσης βρέθηκαν ν'ανήκουν στο MMX_{8}=[1002,1041]. Η συνάρτηση αθροιστικής συχνότητας για τις τιμές του Αυγούστου έδωσε: F(P≤1008)=0.0101 και F(P≥1029)=0.990. Οι τιμές της πίεσης που πρέπει να ελεγχθούν για τον Αύγουστο είναι οι μικρότερες του 1008 ή οι μεγαλύτερες του 1029.

Επειδή για την θερμοκρασία και την πίεση γίνονται περισσότερες της μιας μετρήσεις ανά ημέρα και οι τιμές τους παρουσιάζουν μια σαφή περιοδικότητα (ημερήσιος κύκλος), καλό είναι ο παραπάνω έλεγχος να γίνεται για κάθε ώρα που διατίθενται παρατηρήσεις.

3.4. Έλεγχος του 1/1000 των μεγάλων διαδοχικών διαφορών από το σύνολο των τιμών, ή του 1/100 των διαδοχικών διαφορών στον μήνα ή το δεκαήμερο i.

Δοθείσης μιας χρονοσειράς $x_1,x_2,...x_K$ τιμών της παραμέτρου Χ δημιουργείται η χρονοσειρά των διαδοχικών διαφορών $\Delta_1,\Delta_2...\Delta_{K-1}$ με $\Delta_i=x_{i+1}-x_i$. Για την νέα σειρά βρίσκεται η συνάρτηση αθροιστικής κατανομής F(Δ) και εφαρμόζεται ο κανόνας του 1/1000 της παραπάνω παραγράφου. Τον κανόνα αυτό τον εφαρμόζουμε για μεταβλητές που ακολουθούν κατά το μάλλον ή ήττον την κανονική κατανομή (θερμοκρασία, πίεση ύψος ισοβαρικών επιφανειών). .

Με τον ίδιο τρόπο που δημιουργήθηκαν οι σειρές των τιμών του μήνα i, $i \in I_{12}$ (ή του δεκαήμερου i, $i \in I_{36}$) μπορεί να δημιουργήσει κανείς δώδεκα ή τριανταέξη νέες χρονοσειρές με τιμές τις τιμές των διαδοχικών διαφορών μέσα σ'ένα μήνα ή μέσα σ'ένα δεκαήμερο και να εφαρμόσει τον έλεγχο του 1/1000 στις νέες σειρές παίρνοντας όμως το 1/100 των τιμών αντί του 1/1000, όπως παραπάνω.

Παράδειγμα 3.1.β.

Στο παράδειγμα 10 της παραγράφου 2.2 οι διαδοχικές διαφορές των τιμών της πίεσης της 06UTC ανήκουν στο διάστημα $D=MMX_{6P}$=[-21.4 , 25.7]. Δηλαδή η μεγαλύτερη πτώση και η μεγαλύτερη άνοδος της πιεσης που καταγράφηκε στον υπό μελέτη σταθμο είναι 21.4 και 25.7hPa αντίστοιχα. Η συνάρτηση F(δP) έδωσε για κρίσιμα όρια τις τιμές: x_0=-17.4 και x_1=22.2 Αναζητούνται λοιπόν στο αρχείο του σταθμού οι ημερομηνίες κατά τις οποίες σημειώθηκαν μικρότερες ή μεγαλύτερες διαφορές από τις παραπάνω κρίσιμες τιμές.

3.5. Απλοποίηση της διαδικασίας εύρεσης των "ύποπτων" τιμών για την εφαρμογή των ελέγχων 1/1000 (ή του 1/100).

Οι τιμές $x_1, x_2, \ldots x_K$ και $\Delta_1, \Delta_2, \ldots \Delta_{K-1}$, των χρονοσειρών Χ και Δ κατατάσσονται κατά σειρά αύξοντος μεγέθους και έστω $x'_1 \le x'_2 \le \ldots \le x'_K$ και $\Delta'_1 \le \Delta'_2 \le \ldots \le \Delta'_{K-1}$ οι νέες χρονοσειρές των οποίων τα στοιχεία είναι διατεταγμένα. Η τιμή x_0 για την οποία ισχύει η $P(x \le x_0) \le 0.001$ είναι τώρα η τιμή που βρίσκεται στην θέση [0.001*K]+1 και αυτή για την οποία ισχύει $P(x_1 \ge x) \ge 0.999$ η τιμή που έχει τάξη [0.999*K]-1 (όπου η συνάρτηση [x] είναι το ακέραιο μέρος του αριθμού x). Για τις αντίστοιχες τιμές στην χρονοσειρά των διαδοχικών διαφορών, οι αντίστοιχες τιμές είναι αυτές που καταλαμβάνουν τις θέσεις [0.001*(K-1)]+1 και [0.999*(K-1)]-1 αντίστοιχα.

Για τους ελέγχους του 1/100 των τιμών του μήνα j, $j \in I_{12}$ (ή του δεκαήμερου $j \in I_{36}$), κατατάσσονται οι τιμές των χρονοσειρών με τιμές στον μήνα (δεκαήμερο) j, κατά σειρά άυξοντος μεγέθους και ελέγχονται σαν "ύποπτες" οι $[0.01 \ast K_j]+1$ και $[0.99 \ast (K_j-1)]-1$ αντίστοιχα. Οι αριθμοί K_j παριστούν το πλήθος των τιμών της χρονοσειράς του μήνα (του δεκαήμερου).

3.6. Συνδυασμός της μεθόδου 1/100 και των θεωρημάτων 1,2,3. Τα

Τα θεωρήματα 1,2 και 3 μπορούν να εφαρμοστούν για τον έλεγχο απλών τιμών σε συνδυασμό με τις μεθόδους του 1/100 ώς εξής:

Έστω ότι από τους ελέγχους του 1/100 προκύψει κάποια "ύποπτη" τιμή της παραμέτρου Χ κατά το δεκαήμερο i του έτους Κ. Αν η παράμετρος Χ ακολουθεί την κανονική κατανομή, τότε οι δέκα τιμές του δεκαήμερου στο οποίο βρέθηκε η "ύποπτη" αυτή τιμή θεωρείται σαν ένα τυχαίο δείγμα μεγέθους n=10 (μικρό δείγμα). Αν η μέση τιμή, $E(X_i)$, του δεκαήμερου κείται έξω από τα διαστήματα εμπιστοσύνης (Δ.Ε) που ορίζονται από τις (2.24) ή

(2.27), τότε αυξάνεται η "υποψία" της εν λογω τιμής. Αν η τιμή βρίσκεται μέσα στα διαστήματα αυτά, τότε αν Y είναι μια άλλη παράμετρος που ακολουθεί την κανονική κατανομή $Y_i \approx N(Y_i, \mu_{y,i}, \sigma^2 y_{,i})$ και $E(Y_i)$ είναι η μέση τιμή των n τιμών του δεκαήμερου στο οποίο διαπιστώθηκε η "ύποπτη" τιμή, τότε βρίσκεται το διάστημα εμπιστοσύνης της διαφοράς $E(X_i)-E(Y_i)$ από την (2.37), με $\mu_1-\mu_2=\mu_{x,i}-\mu_{y,i}$. Αν η διαφορά $E(X_i)-E(Y_i)$ βρίσκεται έξω από το (Δ.Ε) τότε ενισχύεται ακόμα περισσότερο η άποψη ότι η "ύποπτη" τιμή είναι πράγματι "ύποπτη".

Για την εφαρμογή της μεθόδου αυτής πρέπει να ελεγχθούν οι κατανομές των συχνοτήτων των παραμέτρων X και Y, με τον τρόπο που περιγράφηκε στην παράγραφο 2.8 και στο παράδειγμα 2.8.α.

3.7.Συνδυασμός της μεθόδου 1/100 και του θεωρήματος 7.

Αν $X=N(X,\mu,\sigma^2)$ και $E(X_i)$, S_i $i \in I_k$ οι μέσες τιμές και οι τυπικές αποκλίσεις **κ** τυχαίων δειγμάτων μεγέθους $\mathbf{n_k}$ το καθένα, βρίσκονται οι τυχαίες μεταβλητές $X^2_k, i \in I_k$ που ακολουθούν την X^2-κατανομή με $\mathbf{v_i=n_i-1}$ βαθμούς ελευθερίας.

Αν για την σ.σ.: α, η τιμή $X^2_{v_i,\alpha}$ είναι μεγαλύτερη αυτής που βρέθηκε $X^2_{v_i}$,δηλαδή:

$$X^2_{v_i} > X^2_{v_i,\alpha} \quad (3.3)$$

τότε το δείγμα μεγέθους $v_i=n_i-1$ έχει διασπορά μεγαλύτερη της αναμενόμενης και αυτό πιθανώς να οφείλεται στην ύπαρξη εντός αυτού μιας ή περισσοτέρων λανθασμένων τιμών, γεγονός που πρέπει να ελεγχθεί.

Οι τιμές $X^2_{v,\alpha}$ βρίσκονται στους Πίνακες IIα,IIβ του παραρτήματος.

Στους πίνακες αυτούς δίνονται συναρτήσει των v και α οι τιμές $X^2_{v,\alpha}$ για τις οποίες:

$$P(X^2_v > X^2_{v,\alpha}) = \alpha \quad (3.4)$$

Αν θεωρήσουμε σαν X την παράμετρο "Μέση τιμή του δεκαημέρου (ή του μήνα), τότε η X ακολουθεί την κανονική κατανομή (παραδοχή που είναι αληθής για τις περισσότερες μετεωρολογικές παραμέτρους). Παριστάνουμε με μ_k και σ_k τη μέση τιμή και την τυπική απόκλιση της X και με $E(X_j)$ και S_j η μέση τιμή και η τυπική απόκλιση του δεκαημέρου κ (μήνα) του έτους j.

Αν θεωρηθούν οι **n** μέρες του **κ** δεκαημέρου (μήνα) του έτους j και εαν ένα τυχαίο δείγμα μεγέθους **n** (σταθερό για το ίδιο δεκαήμερο όταν δεν υπάρχουν ελλείπουσες τιμές), τότε βρίσκονται οι:

$$X^2_{v,j} = \frac{(n-1)*S^2_j}{\sigma^2_v}$$

και συγκρίνονται οι τιμές $X^2_{v,j}$ με τις αντίστοιχες τιμές $X^2_{j,\alpha}$ των Πινάκων IIα, IIβ, με v=n-1 και δοθείσα (σ.σ) : α.

Αν $X^2_{v,j} > X^2_{j,\alpha}$, τότε είναι πολύ πιθανό να υπάρχουν λανθασμένες τιμές στο υπό μελέτη δεκαήμερο του έτους j και πρέπει να ελεγχθούν όλες οι τιμές του δακαημέρου. Αν μάλιστα η εφαρμογή του ελέγχου 1/100 ή/και του ελέγχου με την εφαρμογή των θεωρημάτων 1 και 2 δώσουν μια "ύποπτη " τιμή που βρίσκεται στο δεκαήμερο αυτό τότε αυξάνεται ακόμη περισσότερο ο "βαθμός υποψίας" της εν λόγω τιμής.

3.8. Αλλαγή των ορίων των προτεινόμενων για τον πρωταρχικό έλεγχο τιμών.

Από τις προηγούμενες παραγράφους προκύπτει ότι μπορούν να τροποποιηθούν τα πολύ μεγάλα όρια ,εντός των οποίων πρέπει να κυμαίνονται οι τιμές των διαφόρων μετεωρολογικών παραμέτρων στον πρωταρχικό έλεγχο, σε νέα όρια, ειδικά για κάθε σταθμό και για κάθε δεκαήμερο του έτους. Τα νέα αυτά όρια μπορούν να ορισθούν από τα όρια που θα προκύψουν από την εφαρμογή των μεθόδων 1/100 και για τον έλεγχο των απλών τιμών να εφαρμόζεται η παρακάτω διαδικασία:

α. Ελέγχεται η τιμή αν βρίσκεται στα όρια που τίθενται με τον πρωταρχικό έλεγχο, κοινό για όλους τους σταθμούς.

`Οταν η τιμή βρίσκεται εκτός των ορίων αυτών, η τιμή απορρίπτεται.

β. `Οταν η τιμή βρίσκεται εντός των ορίων αυτών, ελέγχεται κατόπιν αν βρίσκεται μεταξύ της ελάχιστης και της μέγιστης τιμής που παρατηρήθηκε στον εν λόγω σταθμό για το συγκεκριμένο δεκαήμερο. Αν βρίσκεται εκτός των ορίων (για μελλοντικές τιμές), τότε γίνεται έλεγχος της από τα πρωτογενή στοιχεία.

γ. Αν βρίσκεται εντός των παραπάνω ορίων, ελέγχεται αν βρίσκεται και εντός ή εκτός των ορίων που προκύπτουν από την εφαρμογή των μεθόδων 1/100 στο συγκεκριμένο δεκαήμερο και για τον συγκεκριμένο σταθμό. Αν βρίσκεται εντός των ορίων γίνεται αποδεκτή σαν σωστή.`Αν όχι, γίνεται έλεγχος της από τα πρωτογενή στοιχεία και αν υπάρχει λάθος διορθώνεται.

4. ΕΛΕΓΧΟΣ ΟΜΟΙΟΓΕΝΕΙΑΣ

4.1. Έννοια ομοιογένειας

Μια χρονοσειρά $x_1, x_2, ..., x_n$, όπου x_i, $i \in I_n$ οι τιμές μιας δειγματικής στατιστικής μετεωρολογικής παραμέτρου (π.χ. μηνιαίες μέσες τιμές, μηνιαία ή ετήσια αθροίσματα κ.λ.π.), λέγεται **ομοιογενής**, άν είναι δείγμα από ένα και μοναδικό πληθυσμό. Εξ ορισμού, **μια κλιματολογική σειρά πρέπει να είναι μία ομοιογενής χρονοσειρά.**

Αν μια χρονοσειρά είναι ομοιογενής, οι μεταβολές της παραμέτρου που αντιπροσωπεύει θα πρέπει να οφείλονται μόνο στις μεταβολές του καιρού και του κλίματος.

Μιά σειρά θα λέγεται **ομοιογενής** όταν είναι στο σύνολό της ομοιογενής και **τμηματικά ή κατά τμήματα ομοιογενής** , όταν δεν είναι ομοιογενής σε όλη την χρονικη περίοδο λειτουργίας του σταθμού αλλά μόνο σε υποπεριόδους αυτής.

Ο έλεγχος της ομοιογένειας της χρονοσειράς μιας μετεωρολογικής παραμέτρου είναι ένα δύσκολο θέμα και απαιτεί εξαιρετική προσοχή, γιατί είναι ποικίλα τα αίτια που μπορούν να δημιουργήσουν την ανομοιογένεια και γιατί είναι δύσκολο να διαχωριστούν οι φυσικές αιτίες από τις τεχνητές παρεμβάσεις που την προκαλούν.

Μπορεί μάλιστα να ειπωθεί ότι μια οποιαδήποτε σειρά δεδομένων μεγαλύτερη των 10 ή 20 χρονων έχει μεγάλη πιθανότητα να έχει κάποιο είδος ανομοιογένειας.

Οι ανομοιογένειες σε μια κλιματολογική σειρά μπορούν να έχουν τη μορφή:

(ι) μιας απότομης αλλαγής στις παραμέτρους της κεντρικής τάσης

(ιι) μιας προοδευτικής αλλαγής (μακραίωνα τάση)

(ιιι) ενός είδους ταλάντωσης

Στις μετεωρολογικές παραμέτρους οι περισσότερες ανομοιογένειες εμπίπτουν στην πρώτη κατηγορία και οι κυριότεροι παράγοντες που οδηγούν σε αυτή είναι:

(α) Η μετακίνηση του σταθμού

(β) Η αλλαγή στην έκθεση των οργάνων

(γ) Η αλλαγή των οργάνων

..

(δ) Η αλλαγή στη μέθοδο λήψης των παρατηρήσεων

(ε) Η αλλαγή του παρατηρητού

(στ) Η αλλαγή στην ώρα της παρατήρησης

Αυτές οι ανομοιογένειες μπορούν να επηρεάσουν τις τιμές των στατιστικών παραμέτρων, όπως τη μέση τιμή, τη διασπορά, κ.λ.π

Στις μακρές σειρές δεδομένων οι ανομοιογένειες οφείλονται σε συνδυασμο κυρίως των παραπάνω παραγόντων αλλά, και άλλων, άγνωστων παραγόντων. Οι μεγάλες χρονοσειρές, οι οποίες έχουν και την μεγαλύτερη αξία σε κλιματικές μελέτες, προέρχονται συνήθως από σταθμούς εγκατεστημένους κοντά σε κατοικημένα κέντρα τα οποία γενικά αναπτύσσονται και επεκτείνονται προοδευτικά. Έτσι, οι τοπικές συνθήκες γύρω από αυτούς τους σταθμούς (μικροκλίμα) αλλάζουν προοδευτικά με αποτέλεσμα και οι κλιματικές-στατιστικές παράμετροι που καθορίζονται από αυτούς τους σταθμούς να μεταβάλλονται κι έτσι να είναι ανόμοιες με εκείνες που είχαν υπολογισθεί προτού το περιβάλλον των σταθμών μεταβληθεί από εξοχικό σε αστικό.

Άλλά, εκτος από την ανάπτυξη και επέκταση των κατοικημένων περιοχών, τα διάφορα προγράμματα τροποποιήσης του περιβάλλοντος, η αποξήρανση των ελών, η κατασκευή τεχνητών λιμνών και δεξαμενών, οι πυρκαγιές των δασών, η αλλαγή της επιφανειακής κάλυψης με διάφορες καλλιέργειες, η μεγάλης κλίμακας άρδευση, τα προγράμματα αναδάσωσης και γενικά τεχνητές τροποποιήσεις του τοπικού περιβάλλοντος, αποτελούν πηγές ανομοιογένειας στα ιστορικά κλιματολογικά δεδομένα.

Όμως οι κύριοι παράγοντες ανομοιογένειας ειναι οι (α) μέχρι (στ) που αναφέρθηκαν παραπάνω και θα πρέπει να τονισθεί οτι ο καλύτερος τροπος για να ελεγχθεί και να διαπιστωθεί η τυχόν ανομοιογένεια, είναι η προσεκτική μελέτη του ιστορικού του σταθμού. Εάν το ιστορικό δείχνει αλλαγές που θα μπορούσαν να προκαλέσουν ανομοιογένειες και που θα μπορούσαν να περιγραφούν ως προς την περίοδο και τον χαρακτήρα τους, τότε μπορούν να εφαρμοσθούν έλεγχοι ανομοιογένειας.

Εάν μία χρονοσειρά θεωρηθεί ανομοιογενής, τότε μπορούν να γίνουν προσαρμογές των δεδομένων έτσι ώστε, η νέα τροποποιημένη χρονοσειρά να είναι ομοιογενής και οι στατιστικοί υπολογισμοί να εκτιμούν με την καλύτερη δυνατή προσέγγιση τις στατιστικές παραμέτρους του πληθυσμού.

Υπάρχουν διάφοροι μέθοδοι ομοιογενοποίησης των δεδομένων, όμως κάθε μια από αυτές μπορεί να χρησιμοποιηθεί μόνο κατά περίπτωση, κάτω από προϋποθέσεις και με μεγάλη προσοχή.

Θα πρέπει, τέλος, να τονισθεί ότι όλες οι μέθοδοι ομοιογενοποίησης που θα αναφερθούν, αφορούν τροποποίηση των δειγματικών στατιστικών παραμέτρων π.χ. των μέσων τιμών (δεκαημέρων, μηνιαίων, ετησίων κλπ), των ολικών ποσών (μηνιαίων, ετησίων κλπ) και **όχι των απλών τιμών (ωριαίων ή ημερήσιων τιμών των παρατηρήσεων).**

Στις επόμενες παραγράφους αρχικά αναπτύσσονται διεξοδικά βασικές μέθοδοι ελέγχου της ομοιογένειας γιατί δεν νοείται να προχωρήσει κανείς σε ομοιογενοποίηση δεδομένων, προτού διαπιστώσει όσο το δυνατόν πιο τεκμηριωμένα την ύπαρξη ανομοιογένειας.

Υπάρχουν διάφορες γραφικές μέθοδοι για τον έλεγχο της ομοιογένειας, (καταχώρηση των διαφορών ή των λόγων των τιμών σταθμών, ανάλυση διπλής μάζας κλπ) , οι οποίες είναι πολύ χρήσιμες για μια πρώτη οπτική ένδειξη ύπαρξης τυχόν ανομοιογένειας, όμως έχουν την αδυναμία ότι η τελική απόφαση στηρίζεται σε υποκειμενική εκτίμηση.

Ακολούθως περιγράφονται τα παραμετρικά και μη-παραμετρικά στατιστικά tests που παρέχουν ποσοτικούς και αντικειμενικούς κανόνες βάσει των οποίων γίνεται αποδεκτή ή απορρίπτεται η υπόθεση της ομοιογένειας με πιθανότητα αξιοπιστίας (στάθμη σημαντικότητας). Σ' αυτά τα tests έχει σημασία το μέγεθος το δείγματος.

Τόσο οι γραφικές μέθοδοι, όσο και τα παραμετρικά ή μη παραμετρικά tests, βασίζονται στη συσχέτιση των τιμών των στατιστικών παραμέτρων ενός σταθμού με τις αντίστοιχες τιμές γειτονικών σταθμών που διαθέτουν χρονοσειρές αποδεδειγμένα ομοιογενείς, ή στη συσχέτιση των στατιστικών παραμέτρων δύο ή περισσοτέρων δειγμάτων της ίδιας μεταβλητής, του ίδιου του σταθμού αλλά σε διαφορετικές χρονικές περιόδους.

Τέλος, παρατίθενται βασικές μέθοδοι ομοιογενοποίησης των δεδομένων.

Θα πρέπει να τονισθεί ότι, αν γίνουν τροποποιήσεις στα δεδομένα, θα πρέπει πάντοτε παράλληλα με τα διορθωμένα (δεδομένα) να διατηρηθούν και τα αρχικά, γιατί τα διορθωμένα εμπεριέχουν κάποια προβλήματα αλλά και γιατί μπορεί να αναπτυχθούν καλύτεροι μέθοδοι ομοιογενοποίησης στο μέλλον οι οποίοι δεν θα μπορούν να εφαρμοσθούν χωρίς την ύπαρξη των αρχικών δεδομένων.

Τέλος, όσα παρατίθενται στα επόμενα κεφάλαια περιγράφουν τις βασικές διαδικασίες ελέγχου ομοιογένειας και ομοιογενοποίησης, οι οποίες μπορούν να εφαρμοσθούν από τους μελετητές των χρονοσειρών. Όμως, το όλο θέμα δεν παύει να αποτελεί αντικείμενο έρευνας και να βρίσκονται νέες καλύτερες μέθοδοι.

4.2. Γραφικοί έλεγχοι της ομοιογένειας

Υπάρχουν διάφορες γραφικές διαδικασίες για τον αρχικό έλεγχο της ομοιογένειας από τις οποίες παραθέτουμε δύο βασικές μεθόδους.

4.2.1. Γραφικός έλεγχος ομοιογένειας με καταχώρηση των διαφορών. Αν, παράλληλα με την υπό έλεγχο χρονοσειρά μιας κλιματολογικής παραμέτρου (π.χ. της μέσης ετήσιας θερμοκρασίας, μέσης μηνιαίας θερμοκρασίας κλπ) $x_1, x_2, ... x_n$, από ένα σταθμό Σ, διατίθενται **για την ίδια χρονική περίοδο**, αντίστοιχες χρονοσειρές τιμών:

$$(a1_1, a1_2, ..., a1_n), (a2_1, a2_2, ... a2_n), ..., (am_1, am_2, ... am_n),$$

από m γειτονικούς σταθμούς $\Sigma_1, \Sigma_2, ..., \Sigma_m$, μια διαδικασία που μπορεί να ακολουθηθεί είναι η εξής:

Υπολογίζονται οι από έτος σε έτος **διαφορές** μεταξύ της υπό μελέτη σειράς και κάθε μιας από τις άλλες σειρές και καταχωρούνται όλες οι προκύπτουσες σειρές διαφορών:

$$\begin{array}{l} (x_1-a1_1, x_2-a1_2, ..., x_n-a1_n), \\ (x_1-a2_1, x_2-a2_2, ..., x_n-a2_n), \qquad (4.1) \\ .... \\ (x_1-am_1, x_2-am_2, ..., x_n-am_n) \end{array}$$

σ'ένα σύστημα ορθογωνίων συντεταγμένων όπου ο οριζόντιος άξονας αναφέρεται στις τιμές του χρόνου από τον οποίο προέκυψαν οι τιμές x_i, $a\zeta_i$, $i \epsilon I_n$, $\zeta \epsilon I_m$ και ο κατακόρυφος άξονας στις τιμές των διαφορών.

Αυτή η καταχώρηση παρέχει μια πρώτη ένδειξη της συγκρισιμότητας μεταξύ των αρχείων της θερμοκρασίας όλων των επιλεγμένων σταθμών. Εάν κάποια από τις τιμές της θερμοκρασίας της υπό μελέτη σειράς περιέχει σημαντικά σφάλματα, αυτή η διαδικασία καταχώρησης τα αποκαλύπτει με την εμφάνιση "αιχμής" στο ίδιο έτος σε όλες (ή τουλάχιστον στις περισσότερες) καταχωρημένες σειρές.

Παράδειγμα 4.2.1.

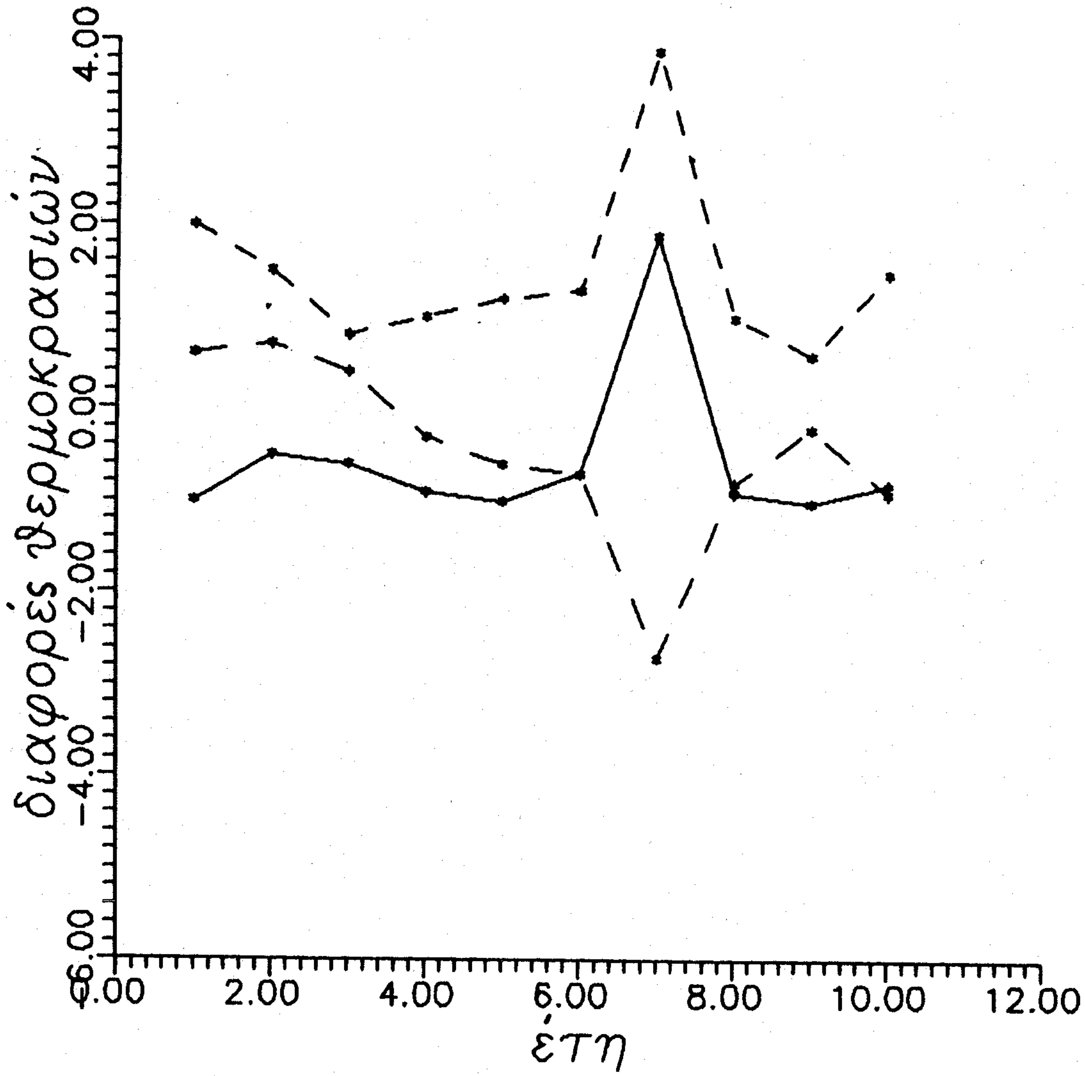
Στον Πίνακα (4.1.) παρατίθενται οι μέσες μηνιαίες τιμές θερμοκρασίας Ιουνίου για 10 χρόνια στο σταθμό $Σ_1$ (2η στήλη) και ακολούθως για τα ίδια χρόνια στους γειτονικούς σταθμούς $Σ_2$, $Σ_3$ και $Σ_4$ που θεωρούνται ομοιογενείς (3η,4η και 5η στήλη).
Οι διαφορές των μέσων μηνιαίων θερμοκρασιών μεταξύ του σταθμού $Σ_1$ και των γειτονικών σταθμών $Σ_2$,$Σ_3$,$Σ_4$ έχουν γραφεί στις επόμενες στήλες (4η, 6η, και 8η αντίστοιχα).

Στο σχήμα (4.1.) έχουν καταχωρηθεί οι σειρές των διαφορών για τα 10 χρόνια και για τους τρεις σταθμούς και διαπιστώνεται εύκολα ότι υπάρχει και στα τρία γραφήματα ένδειξη ανομοιογένειας τον 7ο χρόνο. Γι'αυτό, η τιμή της μέσης μηνιαίας θερμοκρασίας αυτόν τον χρόνο, στο σταθμό $Σ_1$ πρέπει να ερευνηθεί από το ιστορικό του σταθμού (π.χ. αλλαγή οργάνου κλπ.)

Μια άλλη διαδικασία που μπορεί να ακολουθηθεί στην περίπτωση του υετού, είναι να καταχωρηθεί η σειρά των ετήσιων ή εποχικών ολικών ποσών υετού του υπό έλεγχο σταθμού μαζί με συγκρίσιμες σειρές άλλων ομοιογενών σταθμών με την μορφή συσσωρευτικών αθροισμάτων. Αυτή η μέθοδος που λέγεται τεχνική της **"ανάλυσης διπλής μάζας"**, αναλύεται στην επόμενη παράγραφο λεπτομερώς γιατί αποτελεί την καλύτερη γραφική μέθοδο ελέγχου της ομοιογένειας σειράς δεδομένων.

Πίνακας 4.1. Μηνιαίες μέσες τιμές Ιουνίου, 10 ετών, στους σταθμούς Σ_1,Σ_2,Σ_3,Σ_4 καθώς και οι διαφορές των αντίστοιχων τιμών μεταξύ του σταθμού Σ_1 και των τριών γειτονικών του.

Χρονος	Μηνιαία μέση θερμοκρασία Ιουνίου				Διαφορές θερμοκρασιών		
	Σ_1	Σ_2	Σ_3	Σ_4	$\Sigma_1-\Sigma_2$	$\Sigma_1-\Sigma_3$	$\Sigma_1-\Sigma_4$
1	17.4	18.4	15.4	16.8	-1.0	+2.0	+0.6
2	20.9	21.4	19.4	20.2	-0.5	+1.5	+0.7
3	18.7	19.3	17.9	18.3	-0.6	+0.8	+0.4
4	18.6	19.5	17.6	18.9	-0.9	+1.0	-0.3
5	16.9	17.9	15.7	16.3	-1.0	+1.2	-0.6
6	20.8	21.5	19.5	20.1	-0.7	+1.3	-0.7
7	23.9	22.0	20.0	20.2	+1.9	+3.9	-2.7
8	17.9	18.8	16.9	17.1	-0.9	+1.0	-0.8
9	14.2	15.2	13.6	14.0	-1.0	+0.6	-0.2
10	19.2	20.0	17.7	18.3	-0.8	+1.5	-0.9



Σχήμα 4.1. Καταχωρημένες τιμές διαφορών θερμοκρασίας μεταξύ χρονοσειρών σταθμών.

4.2.2.Ανάλυση διπλής μάζας. Στην μετεωρολογική και κυρίως την υδρομετεωρολογική πρακτική χρησιμοποιείται συχνά η "ανάλυση καμπύλης διπλής μάζας" ή απλά "ανάλυση διπλής μάζας" για τον έλεγχο της ομοιογένειας της χρονοσειράς μιας κλιματολογικής παραμέτρου και κυρίως του υετού (Seary και Hardison 1960).

Η μέθοδος αυτή συνίσταται στην γραφική απεικόνιση των αθροιστικών ετήσιων υψών υετού (αλλά και άλλων παραμέτρων) του υπό εξέταση σταθμού και του μέσου όρου από **m** γειτονικούς σταθμούς, των οποίων οι χρονοσειρές των υψών υετού θεωρούνται ομοιογενείς.

Η μέθοδος έχει ως εξής:

Έστω ότι χρειάζεται να ελεγχθεί η ομοιογένεια της σειράς των ετήσιων υψών υετού $x_1, x_2, ... x_n$ του σταθμού Σ_1 για **n** χρόνια και διατίθενται για την ίδια περίοδο των **n** αυτών χρόνων και οι αντίστοιχες χρονοσειρές m γειτονικών σταθμών :
$(a1_1, ... a1_n), (a2_1, ..., a2_n), ..., (am_1, ... am_n)$
που θεωρούνται ομοιογενείς.

Από την σειρά $x_1, x_2, ..., x_n$ δημιουργείται η σειρά των αθροιστικών τιμών ολικού υετού $X_1, X_2, ..., X_n$, όπου:

$$X_k = {}_1\Sigma^k x_i, k \in I_n \quad (4.2)$$

δηλαδή:

$$X_1 = x_1$$
$$X_2 = x_1 + x_2$$
$$X_3 = x_1 + x_2 + x_3$$
$$.................$$
$$X_n = x_1 + x_2 + x_3 + ... + x_n$$

Από τις τιμές των χρονοσειρών των γειτονικών σταθμών υπολογίζεται η χρονοσειρά των μέσων όρων για κάθε έτος $A_1, A_2, ..., A_n$, όπου:

$$A_i = ({}_1\Sigma^m \alpha j_i)/m \quad (\text{άθροιση ως προς } j), i \in I_n \quad (4.3)$$

δηλαδή:

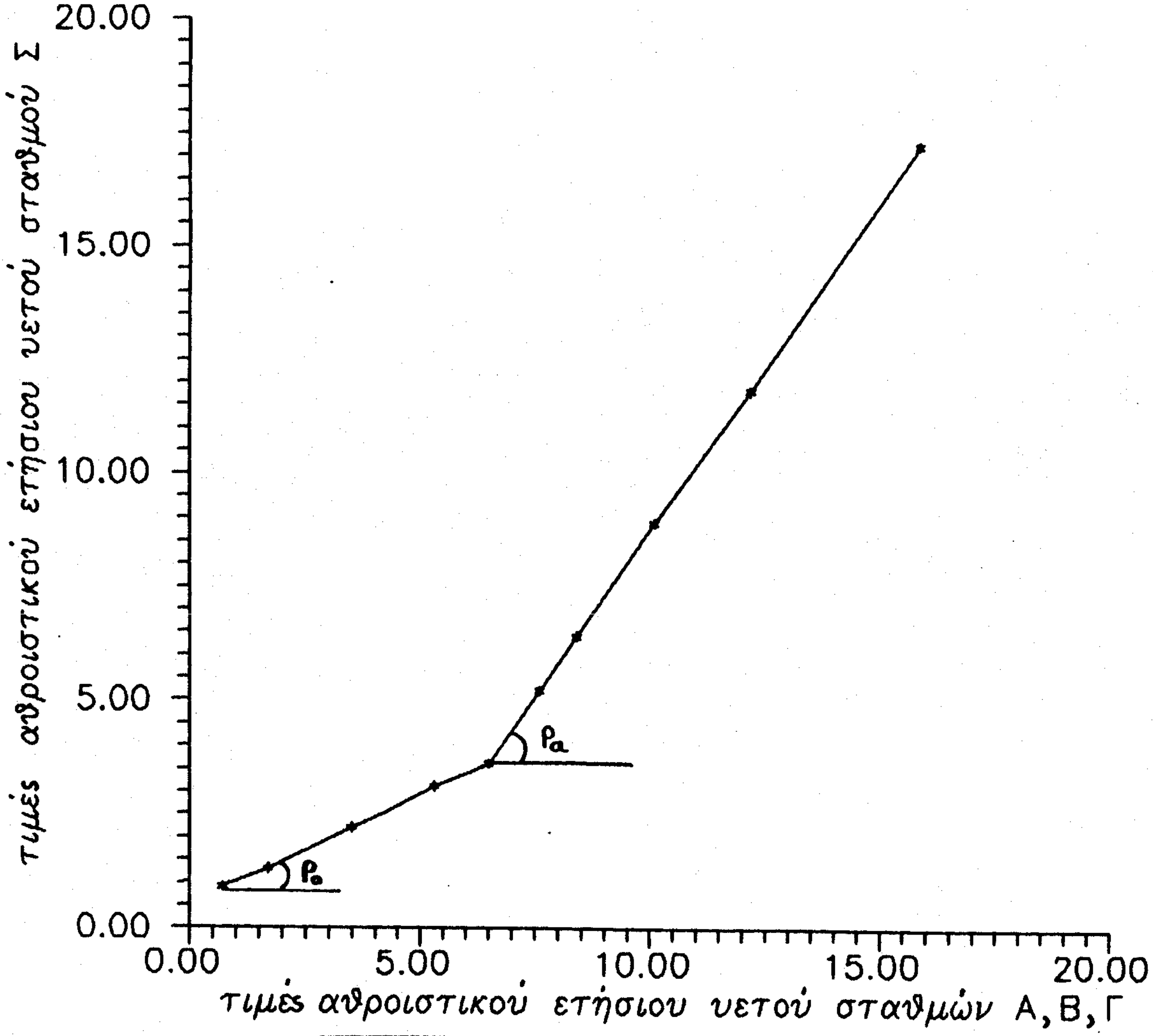
$$A_1 = (a1_1 + a2_1 + ... + am_1)/m$$
$$A_2 = (a1_2 + a2_2 + ... + am_2)/m$$
$$.....................$$
$$A_n = (\alpha 1_n + \alpha 2_n + ... + am_n)/m$$

Ακολούθως υπολογίζεται η σειρά των αθροιστικών υψών υετού $Y_1, Y_2, ..., Y_n$, όπου:

διαπιστώνεται οτι αλλάζει σημαντικά η κλιση του ευθύγραμμου τμήματος απο τον 5ο χρονο και μετά ,πράγμα που αποτελεί μια πρώτη ένδειξη ανομοιογένειας της χρονοσειράς.

Πίνακας 4.2. Ετήσια ύψη υετού και αθροιστικά ύψη υετού για 10 χρόνια στον σταθμό Σ και σε τρεις γειτονικούς του (μέσος όρος) με μονάδα τα 1000 mm .

Χρονος	Ετήσιος υετός στο σταθμό Σ	Αθροιστικος ετήσιος υετός στο σταθμό Σ	Μέσος όρος των ετήσιων υετών στους σταθμούς Α,Β,Γ	Αντίστοιχοι αθροιστικοί ετήσιοι υετοί των σταθμων
1	0.90	0.90	0.70	0.70
2	0.40	1.30	1.00	1.70
3	0.90	2.20	1.80	3.50
4	0.90	3.10	1.80	5.30
5	0.50	3.60	1.20	6.50
6	1.60	5.20	1.10	7.60
7	1.20	6.40	0.80	8.40
8	2.50	8.90	1.70	10.10
9	2.90	11.80	2.10	12.20
10	5.40	17.20	3.70	15.90



Σχήμα 4.2. Καταχωρημένες τιμές αθροιστικού ετήσιου υετού.

4.3. Αντικειμενικές μέθοδοι ελέγχου σχετικής ομοιογένειας

4.3.1. Γενικά. Ενώ οι γραφικές μέθοδοι ελέγχου της ομοιογένειας δεδομένων είναι πολύ χρήσιμες για μια πρώτη ένδειξη της ύπαρξης ή όχι ομοιογένειας, ο αναλυτής δεν μπορεί να βασισθεί μόνο σ'αυτές για μια τελική απόφαση αλλά πρέπει να χρησιμοποιήσει και αντικειμενικές μεθόδους για τον έλεγχο της ομοιογένειάς. Από παλιά, οι αντικειμενικές μέθοδοι ελέγχου της ομοιογένειας έχουν βασισθεί στην εκτίμηση της σχετικής ομοιογένειας **δύο ή περισσοτέρων σειρών**, εξεταζομένων ανά δύο κάθε φορά. Ειδικά στα δεδομένα της θερμοκρασίας και του υετού έχει **εφαρμοσθεί** αυτή η μέθοδος από τους Conrad (1925,1927,1932) και Pollak (1950) και Conrad,Siheier (1927).

Στην περίπτωση της μέσης θερμοκρασίας, αφαιρούνται οι αντίστοιχες τιμές των χρονοσειρών δύο γειτονικών σταθμών και εξετάζεται κατά πόσο οι προκύπτουσες **διαφορές είναι τυχαίες** (μη σημαντικές) με την εφαρμογή ενός κατάλληλου test

Στην περίπτωση του υετού, (μέσων ή ολικών ποσών), οι τιμές της σειράς του ενός σταθμού διαιρούνται (όρο προς όρο) με τις αντίστοιχες τιμές του άλλου σταθμού και εξετάζεται αν οι **λόγοι** που προκύπτουν **είναι τυχαίοι (μη σημαντικοί)** με την ίδια διαδικασία.

Και στις δύο περιπτώσεις, αν η εφαρμογή ενός στατιστικού test οδηγήσει στο συμπέρασμα ότι οι τιμές των διαφορών ή των λόγων είναι στατιστικά μη σημαντικές,τότε οι δύο σειρές λέγονται **σχετικά ομοιογενείς.**

Ο Conrad έχει προτείνει διάφορα tests για τον έλεγχο της σημαντικότητας αυτών των τιμών στην εκτίμηση της σχετικής ομοιογένειας, όπως το "Abbe Criterion" και το "Helmert Criterion" αλλά είναι πολύ αυστηρά με αποτέλεσμα να καθορίζονται μη ομοιογενείς, σειρές που είναι ομοιογενείς.

Σε μελέτες **σχετικής ομοιογένειας,** όταν η σειρά των διαφορών ή των λόγων αποδεικνύεται ότι ακολουθεί προσεγγιστικά την "κανονική" κατανομή (Εργασία του προγράμματος "ΥΔΡΟΣΚΟΠΙΟ" "Σχεδιασμός στατιστικής επεξεργασίας δεδομένων"), τότε για τον έλεγχο της ομοιογένειας μπορεί να χρησιμοποιηθεί ο λόγος **Von Neumann.** Όταν είναι πιθανόν η πηγή της ανομοιογένειας να οφείλεται σε μια **μετατόπιση** του ενός από τους σταθμούς που μελετώνται, τότε μπορεί να εφαρμοσθεί το **"student's test"** στη σύγκριση των διαφορών (ή των λογων) των μέσων πριν και μετά την ημερομηνία μετατόπισης.

Αν η μορφή της ανομοιογένειας είναι άγνωστη και δεν είναι αποδεκτό ότι η κατανομή συχνότητας είναι η κανονική κατανομή, τότε μπορούν να χρησιμοποιηθούν μη-παραμετρικά tests.

Θα πρέπει να τονισθεί ότι η σχετική ομοιογένεια δύο ή περισσοτέρων χρονοσειρών δεν είναι ο τελικός σκοπός του αναλυτή αλλά είναι ένα μέσο γιά τον τελικό σκοπό, που είναι η διαπίστωση της ομοιογένειας μιας και μόνης σειράς, όταν η μία έκ των σειρών είναι η υπό έλεγχο χρονοσειρά ενώ η άλλη είναι μία διαθέσιμη ομοιογενής χρονοσειρά.

Στις επόμενες παραγράφους παρατίθενται αναλυτικά μέθοδοι ελέγχου της σχετικής ομοιογένειας καθώς και παραδείγματα.

4.3.2. Ο Λόγος Von Neumann ως test ομοιογένειας. Εάν x_i, $i\varepsilon I_n$, είναι οι διαδοχικές δειγματικές παράμετροι μιας μετεωρολογικής παραμέτρου στο σταθμό Σ_1 και $y_i, i\varepsilon I_n$, οι αντίστοιχες τιμές της ίδιας στατιστικής παραμέτρου στο γειτονικό σταθμό Σ_2, ακολουθείται η παρακάτω διαδικασία:

Υπολογίζεται η σειρά των διαφορών $\delta_i=x_i-y_i, i\varepsilon I_n$,(για υετό υπολογίζεται η σειρά των λογων),δηλαδή:

$$\begin{aligned} &\delta_1=x_1-y_1,\\ &\delta_2=x_2-y_2,\\ &\ldots\ldots\ldots\\ &\delta_n=x_n-y_n \end{aligned} \qquad (4.5)$$

και ακολούθως υπολογίζεται η μέση τιμή αυτών $E(\Delta)$ και η διασπορά S_Δ^2 ,σύμφωνα με τους τύπους (2.2) και (2.8).

Ο λόγος Von Neumann ορίζεται ως ο λογος της μέσης τιμής των τετραγώνων των διαφορών των διαδοχικών τιμών της σειράς προς την διακύμανση ή διασπορά S_Δ^2, δηλαδή:

$$V=\frac{\dfrac{{}_1\Sigma^{n-1}(\delta_i-\delta_{i+1})^2}{(n-1)}}{S_\Delta^2} \qquad (4.6)$$

Όταν η σειρά των διαφορών (ή των λόγων) ακολουθεί την **κανονική κατανομή** και ο αριθμός **n** των τιμών είναι αρκετά μεγάλος (**$n \geq 50$**), τότε και ο λόγος Von Neumann ακολουθεί την κανονική κατανομή, με μέση τιμή **$E(V)=2n/(n-1)$** και διασπορά **$S_v=4(n-2)/(n-1)^2$** και, επομένως, για συγκεκριμένη στάθμη σημαντικότητας (σ.σ) α, οι τιμές του V θα επαληθεύουν την ανισότητα:

$$E(V)-Z_{(1-a)/2}*S_v \leq V \leq E(V)+Z_{(1-a)/2}*S_v \qquad (4.7)$$

π.χ. για σ.σ $\alpha=0.05$, είναι $Z_{(1-a)/2}=Z_{0.475}=1.96$ και το διάστημα εμπιστοσύνης D της V θα είναι :

$$D= \left[\frac{2*n-2*1.96*\sqrt{(n-2)}}{n-1}, \frac{2*n+2*1.96.\sqrt{(n-2)}}{n-1} \right] \qquad (4.8)$$

Επομένως, εξετάζεται πρώτα αν η σειρά των διαφορών ή των λόγων ακολουθεί την κανονική κατανομή και εφ'όσον διαπιστωθεί η προσαρμογή σ'αυτή την κατανομή, τότε τίθεται η μηδενική υπόθεση **H_0**, ότι: διαφορές δ_i όπως οι παρατηρηθείσες μπορούν να είναι τυχαίες.

Ακολούθως λαμβάνεται το μονόπλευρο test, και για επιθυμητή σ.σ. α, βρίσκεται η τιμή z_{1-a} της κανονικής κατανομής από τους **Πίνακες Ια, ΙΒ** του παραρτήματος και μετά ελέγχεται αν η υπολογιζόμενη τιμή της V είναι μικρότερη ή όχι από την τιμή

$$V_a=\frac{2*n-2*z_{1-a}*\sqrt{(n-2)}}{n-1} \qquad (4.9)$$

Αν η υπολογιζόμενη τιμή V από το δείγμα είναι μικρότερη από την τιμή V_a, δηλαδή αν $V<V_a$, για την επιθυμητή σ.σ α,τότε απορρίπτεται η H_0 και γίνεται αποδεκτή η εναλλακτική υπόθεση ότι οι τιμές των διαφορών δ_i δεν είναι τυχαίες και επομένως, οι αρχικές σειρές δεν είναι σχετικά ομοιογενείς. Αν όμως, η τιμή της V είναι μεγαλύτερη από την V_a, τότε η μηδενική υπόθεση γίνεται αποδεκτή καθώς θεωρείται ότι οι τιμές της σειράς των διαφορών ή των λόγων είναι τυχαίες και συνεπώς οι σειρές x_i, y_i είναι σχετικά ομοιογενείς.

Παράδειγμα 4.3.2.

Στον Πίνακα (4.3) παρατίθενται οι μηνιαίες μέσες τιμές θερμοκρασίας Ιουνίου για 52 χρόνια, στους σταθμούς Σ_1, και Σ_2 (στήλες:2η, 3η, 6η και 7η) καθώς και οι αντίστοιχες τιμές των διαφορών τους (στήλες: 4η και 8η).

Για τη σειρά δ_i των στηλών 4 και 8 εξετάζεται αν οι τιμές ακολουθούν την κανονική κατανομή εφαρμόζοντας το **X^2-test** σαν test προσαρμογής της παρατηρούμενης προς την θεωρητική συχνότητα, όπως περιγράφεται στην παράγραφο **2.8.1.** Η διαδικασία είναι η εξής :

Διαμελίζεται το διάστημα [-5,7] στο οποίο ανήκουν οι 52 διαφορές δ_i $i\in I_{52}$ σε κ=[5*log52]=[5*1.7]=[8.5]=8 διαστήματα-κλάσεις μήκους [[7-(-5)]/8=1.5,τα οποία είναι:A_1=[-5,-3.5), A_2=[-3.5,-2), A_3=[-2,-0.5), A_4=[-0.5,1), A_5 =[1,2.5), A_6=[2.5,4),A_7=[4,5.5) και A_8=[5.5,7].

Κατόπιν, θεωρείται ο κανονικός μετασχηματισμός **$z=(\Delta-E(\Delta))/S_\Delta$** και υπολογίζονται οι τιμές z_i, $i\in I_8$, με **$z_i=(\alpha_i-E(\Delta))/S_\Delta$**, όπου α_i τα άκρα των διαστημάτων A_i, δηλαδή, το z_i παίρνει τις τιμές:-2.62, -1.96, -1.31, -0.65, 0, 0.65, 1.31, 1.96, 2.62. Στη συνέχεια, από τον Πίνακα ΙΒ της κανονικής κατανομής βρίσκονται οι πιθανότητες **$P(0<z\le z_i)$**,οι οποίες πολλαπλασιαζόμενες με n=52, δίδουν τις αθροιστικές συχνότητες, από τις οποίες βρίσκονται οι θεωρητικές απόλυτες συχνότητες, **θ_i**, για κάθε **διάστημα-κλάση.** Επειδή οι απόλυτες συχνότητες της 1ης, 2ης, 7ης και 8ης κλάσης είναι μικρότερες από 5, ενοποιούνται η 1η με τη 2η στήλη και η 7η με τη 8η, αντίστοιχα, δίνοντας ένα διαμελισμό 6, όχι ισομήκων διαστημάτων, με αντίστοιχες απόλυτες συχνότητες,**θ_i**, όπως παρατίθενται στον Πίνακα 4.4.

Επίσης, με βάση τις τιμές της δοθείσας σειράς, βρίσκονται και οι παρατηρούμενες απόλυτες συχνότητες, **Π_i**.

Πίνακας 4.3. Μηνιαίες μέσες θερμοκρασίες Ιουνίου 52 χρόνων στους σταθμούς $Σ_1$,$Σ_2$ καθώς και οι αντίστοιχες διαφορές τους.

Μηνιαία Μέση Θερμοκρασία				Μηνιαία μέση θερμοκρασία			
Ετος	Στ.$Σ_1$	Στ.$Σ_2$	$δ_i$	Ετος	Στ.$Σ_1$	Στ.$Σ_2$	$δ_i$
1	17.4	20.4	-3.0	27	19.8	17.3	2.5
2	20.9	22.5	-1.6	28	18.3	18.9	-0.6
3	18.7	18.6	0.1	29	19.3	18.4	0.9
4	18.7	17.9	0.8	30	17.5	14.0	3.5
5	16.9	14.9	2.0	31	19.5	21.7	-2.2
6	20.8	18.3	2.5	32	18.6	18.6	0.0
7	20.4	17.2	3.2	33	18.6	16.0	2.6
8	17.9	19.7	-1.8	34	17.9	16.8	1.1
9	18.1	18.5	-0.4	35	17.8	12.3	5.5
10	18.5	18.6	-0.1	36	19.9	19.7	0.2
11	19.5	18.7	0.8	37	20.9	22.6	-1.7
12	18.6	16.3	2.3	38	22.9	20.9	2.0
13	18.6	12.3	6.2	39	18.9	15.8	3.1
14	17.9	15.6	2.3	40	19.2	20.8	-1.6
15	17.8	18.4	-0.6	41	22.0	23.3	-1.3
16	19.9	19.6	0.3	42	18.9	18.3	0.6
17	20.9	21.2	-0.3	43	20.7	18.2	2.5
18	22.9	27.9	-5.0	44	19.7	18.4	1.3
19	18.9	18.0	0.9	45	14.5	14.7	-0.2
20	19.2	16.6	2.6	46	20.3	18.0	2.3
21	22.0	20.8	1.2	47	19.8	18.3	1.5
22	18.9	19.2	-0.3	48	18.3	18.1	0.2
23	20.7	19.4	1.3	49	19.3	12.3	7.0
24	19.7	16.6	3.1	50	17.5	13.7	3.8
25	16.5	20.3	-3.8	51	18.5	15.6	2.9
26	20.3	19.6	0.7	52	16.5	14.3	2.2

Ακολούθως υπολογίζονται οι τιμές X_i^2 σύμφωνα με τον τύπο **(2.40)** και μετά και η τιμή $X^2={}_1\Sigma^6 X_i^2$

Από τους πίνακες της X^2 κατανομής (Πίνακες IIα, IIβ του παραρτήματος) βρίσκουμε για 6-3=3 β.ε. και στάθμη σημαντικότητας α=0.05 την τιμή $X^2_{6-3,0.05}=7.81473$.

Επειδή $X^2<X^2_{3,0.05}$, γίνεται αποδεκτό ότι η σειρά των διαφορών ακολουθεί την κανονική κατανομή και, επομένως, μπορεί να εφαρμοστεί το test Von Neumman.

Όπως έχει ήδη υπολογισθεί είναι:

Η μέση τιμή, $E(z)=.99\approx 1$.

Η διασπορά, $S_\delta^2=5.23$ και συνεπώς $S=2.29$

Επομένως, μπορεί να υπολογισθεί η τιμη V:

$$V=\frac{\dfrac{{}_1\Sigma^{n-1}(\delta_i-\delta_{i+1})^2}{n-1}}{S_\delta^2}=\frac{448.90}{5.23*51}\approx 1.6$$

Η τιμή V_α υπολογίζεται για σ.σ. : α=0.05 και για μονόπλευρο test, από τους Πίνακες I (παράρτημα) της κανονικής κατανομής είναι, $z_{1-\alpha}=1.645$:

$$V_\alpha=\frac{2*52-2*1.645*\sqrt{(52-2)}}{52-1}=1.58$$

Επειδή $V>V_\alpha$, γίνεται αποδεκτό ότι οι δύο αρχικές σειρές είναι σχετικά ομοιογενείς.

Πίνακας 4.4. Προσαρμογή στην κανονική κατανομή.

Διαμελισμός 1ος			Διαμελισμός 2ος				
	Κλάση	θ_i		Κλάση	θ_i	Π_i	X_i^2
1	[-5,-3.5)	1.10					
2	[-3.5,-1)	3.64	1	[-5,-2)	4.9	4	$(4.9-4)^2/4.9$ =0.170
3	[-2,-0.5)	8.48	2	[-2,-0.5)	8.5	7	$(8.5-7)^2/8.5$ =0.260
4	[-0.5, 1)	12.58	3	[-0.5, 1)	12.6	16	$(12.6-16)^2/12.6$=0.920
5	[1 ,2.5)	12.58	4	[1 ,2.5)	12.6	13	$(12.6-13)^2/12.6$=0.013
6	[2.5, 4)	8.48	5	[2.5, 4)	8.5	9	$(8.5-9)^2/8.5$ =0.020
7	[4 ,5.5)	3.64	6	[4 , 7]	4.9	3	$(4.9-3)^2/4.9$ =0.290
8	[5.5, 7]	1.10					
							${}_1\Sigma^6 X_i^2$ =1.673
							$X^2_{3,0.05}$ =7.814

4.3.3.Student's t-test για τον έλεγχο της διαφοράς μεταξύ δύο μέσων. Συχνά χρειάζεται να γίνει σύγκριση των μέσων δύο διαφορετικών περιόδων της ίδιας χρονοσειράς μιας μετεωρολογικής παραμέτρου.

Ας υποτεθεί ότι ένας σταθμός λειτούργησε αρχικά στην θέση Σ_1 για **n** χρόνια και υπάρχει γι'αυτή την περίοδο η σειρά των μηνιαίων τιμών μιας μετεωρολογικής παραμέτρου **X**, $x_1, x_2, \ldots, x_n$ και μετά μεταφέρθηκε στην θέση Σ_2 όπου λειτούργησε για **m** χρόνια και υπάρχουν για την ίδια μετεωρολογική παράμετρο οι μηνιαίες τιμές $y_1, y_2, \ldots, y_m$ στη νέα θέση Σ_2.

Υπολογίζονται οι μέσες τιμές E(x) και E(y) καθώς και οι διασπορές S_x και S_y για τις σειρές x_i, y_j $i \in I_n, j \in I_m$ σύμφωνα με τους τύπους (2.2) και (2.8) .

Σύμφωνα με το θεώρημα **3 (παραγ.2.6)** καθορίζεται ότι αν τα δύο ανεξάρτητα δείγματα αντιστοιχούν σε πληθυσμούς που ακολουθούν την **κανονική κατανομή** και έχουν μέσες τιμές μ_1, μ_2 και ίδια διασπορά σ^2, τότε η τυχαία μεταβλητή t που ορίζεται απο την σχέση :

$$t=\frac{E(x)-E(y)-(\mu_1-\mu_2)}{\left[\frac{n*S_x^2+m*S_y^2}{(n+m-2)}*\left(\frac{1}{n}*\frac{1}{m}\right)\right]^{1/2}} \qquad (4.10)$$

ακολουθεί την κατανομή **student με v=n+m-2** βαθμούς ελευθερίας.

Γι'αυτό, εφ'όσον διαπιστωθεί ότι οι σειρές στα δύο δείγματα ακολουθούν την κανονική κατανομή τίθεται η μηδενική υπόθεση H_o ότι: **οι δύο πληθυσμοί, από τους οποίους προέρχονται τα δείγματα, έχουν ίσους μέσους, $\mu_1=\mu_2=\mu$, και ίδια τυπική απόκλιση, σ, δηλαδή ότι τα δύο δείγματα ανήκουν στον ίδιο πληθυσμό.**

Το test βασίζεται σε μια σύγκριση της απολύτου τιμής $|t_d|$ όπου:

$$t_d=\frac{E(x)-E(y)}{\left[\frac{n*S_x^2+m*S_y^2}{(n+m-2)}*\left(\frac{1}{n}*\frac{1}{m}\right)\right]^{1/2}} \qquad (4.11)$$

με τα σημεία της κατανομής student $t_{v,\alpha/2}$ που αντιστοιχούν σε πιθανότητα 95% (α=0.05) για δίπλευρο test. Αν $t_d>t_{v,\alpha/2}$ απορρίπτεται η H_o και το συμπέρασμα είναι ότι τα δύο δείγματα δεν προέρχονται από τον ίδιο πληθυσμό δηλαδή υπάρχει ανομοιογένεια μεταξύ των θέσεων Σ_1 και Σ_2.

Παράδειγμα 4.3.3.

Στον Πίνακα 4.5 παρέχονται οι μηνιαίες τιμές θερμοκρασίας Αυγούστου για 30 χρόνια στον σταθμό Σ. Τα 12 πρώτα χρόνια ο σταθμός παρέμεινε στην θέση A_1 και μετά μεταφέρθηκε στη θέση A_2. Χρειάζεται να εξετασθεί άν η χρονοσειρά των 30 μηνιαίων θερμοκρασιών είναι ομοιογενής.

Γενικά οι μηνιαίες θερμοκρασίες ακολουθούν την κανονική κατανομή και με την αποδοχή αυτή θα εξετασθεί η ομοιογένεια της σειράς με το Student test.

Η μέση τιμή και η διασπορά των τιμών για τα 12 πρώτα χρόνια, n=12, είναι αντίστοιχα: E(x)=17.87 και S_x^2=1.54, ενώ για τα επόμενα 18 χρόνια, m=18, είναι E(y)=19.28 και S_y^2=3.27

Η τιμή t_d από τον τύπο (4.11) υπολογίζεται: t_d=-1.32

Από τον Πίνακα III (Παράρτημα) για στάθμη σημαντικότητας α=0.05, για β.ε.:v=n+m-2=12+18-2=28 και για δίπλευρο test είναι $t_{v,\alpha/2}=t_{28,0.25}$=2.05

Επομένως, $|t_d|=1.32<t_{28,0.25}=2.05$ και συνεπώς η χρονοσειρά είναι ομοιογενής.

Πίνακας 4.5. Μηνιαίες μέσες θερμοκρασίες Αυγούστου στο σταθμό Σ για 30 χρόνια.

Ετος	Μηνιαία μέση θερμοκρασία Αυγούστου	Ετος	Μηνιαία μέση θερμοκρασία Αυγούστου
1	16.4	16	19.9
2	19.9	17	20.9
3	17.7	18	22.9
4	17.7	19	18.9
5	15.9	20	19.2
6	19.8	21	22.0
7	19.4	22	18.9
8	16.9	23	20.7
9	17.1	24	19.7
10	17.5	25	14.5
11	18.5	26	20.3
12	17.6	27	19.8
13	18.6	28	18.3
14	17.9	29	19.3
15	17.8	30	17.5

4.3.4.`Ελεγχος Cramer για σύγκριση των μέσων υποπεριόδων με το μέσο όλης της περιόδου. Μερικές φορές χρειάζεται να γίνει σύγκριση μεταξύ του μέσου όλης της σειράς και των μέσων τιμών ορισμένων υποπεριόδων του αρχείου. Στις περιπτώσεις αυτές το Cramer test μπορεί να εφαρμοστεί για να καθοριστεί αν οι διαφορές των μέσων δεν είναι μεγαλύτερες από εκείνες που είναι συμβατές με τη μηδενική υπόθεση ότι οι μέσοι των υποπεριόδων ανήκουν στον ίδιο πληθυσμό.

Έστω E(X) και S η μέση τιμή και η τυπική απόκλιση ολόκληρου του αρχείου των τιμών όλης της σειράς και $E(x_k)$ και S_k, η μέση τιμή και η τυπική απόκλιση των n_k τιμών της υποπεριόδου κ.

Αν υπολογισθούν οι μέσες τιμές και οι τυπικές αποκλίσεις σύμφωνα με τους τύπους (2.2) και (2.8), τότε μπορεί να υπολογισθεί και η τιμή τ_k, όπου :

$$\tau_k=\frac{E(x_k)-E(x)}{S} \qquad (4.12)$$

και ακολούθως και η τιμή της t_k, από τον τύπο:

$$t_k=\left[\frac{n_k*(n-2)}{n-n_k(1+\tau_k^2)}\right]^{1/2}*\tau_k \qquad (4.13)$$

Η στατιστική παράμετρος t_k ακολουθεί την **student κατανομή** με **$v=n-2$** βαθμούς ελευθερίας.

Θεωρώντας τη μηδενική υπόθεση H_o ότι οι διαφορές των μέσων δεν είναι σημαντικές, για τυχαίο δείγμα μεγέθους n_k, υπολογίζεται το t_k και η απόλυτη τιμή του $|t_k|$ συγκρίνεται με τα σημεία της student κατανομής $t_{v,\alpha/2}$, που αντιστοιχούν σε πιθανότητα 95% (σ.σ α=0.05) και για δίπλευρο test. Αν για την τιμή t_k ισχύει: $t_k>t_{v,\alpha/2}$, τότε απορρίπτεται η μηδενική υπόθεση H_o και το συμπέρασμα είναι ότι τα δύο δείγματα δεν ανήκουν στον ίδιο πληθυσμό άρα υπάρχει ανομοιογένεια.

Παράδειγμα 4.3.4.

Στον Πίνακα (4.6) παρατίθενται οι μηνιαίες μέσες θερμοκρασίες του Ιουλίου για 30 χρόνια στο σταθμό Σ που μετά από 12 χρόνια συνεχούς λειτουργίας στη θέση A_1 μεταφέρθηκε στη θέση A_2.

Θα εφαρμοσθεί η μέθοδος Cramer για τον έλεγχο της ομοιογένειας.

Υπολογίζεται η μέση τιμή E(X) και η διασπορά S^2 όλης της χρονοσειράς με τους τύπους (2.2) και (2.8) :

$E(X)=19.12$

$S^2=2.26 \quad (S=1.62)$

καθώς και της υποπεριόδου των 12 πρώτων χρόνων (κ=12):

$E_{12}(X)=18.87$

$S_{12}^2=1.54$

Υπολογίζεται η τιμή των παραμέτρων

$\tau_\kappa=(18.87-19.12)/1.62 = -0.15$ και

$$t_\kappa= \left[\frac{12*(30-2)}{30-12*(1+0.15^2)}\right]^{1/2} *(-0.15)=-0.65$$

Από τον Πίνακα III της t-κατανομής (Παράρτημα) για (σ.σ) : α=0.05, για βαθμούς ελευθερίας:ν=30-2=28 και για δίπλευρο test, βρίσκεται $t_{\nu,\alpha/2}\approx 2.05$

Επομένως, $|t_\kappa|=0.65<t_{\nu,\alpha/2}=2.05$, άρα η χρονοσειρά είναι ομοιογενής.

Πίνακας 4.6. Μηνιαίες μέσες θερμοκρασίες Ιουλίου στο σταθμό Σ' για 30 χρόνια.

Ετος	Μηνιαία μέση θερμοκρασία Ιουλίου	Ετος	Μηνιαία μέση θερμοκρασία Ιουλίου
1	17.4	16	19.9
2	20.9	17	20.9
3	18.7	18	22.9
4	18.7	19	18.9
5	16.9	20	19.2
6	20.8	21	22.0
7	20.4	22	18.9
8	17.9	23	20.7
9	18.1	24	19.7
10	18.5	25	14.5
11	19.5	26	20.3
12	18.6	27	19.8
13	18.6	28	18.3
14	17.9	29	19.3
15	17.8	30	17.5

4.3.5. Μέθοδος ελέγχου της ομοιογένειας των υψών βροχής με τις αθροιστικές αποκλίσεις. Ο έλεγχος της ομοιογένειας με αυτή τη μέθοδο προϋποθέτει ότι η υπό μελέτη μεταβλητή ακολουθεί την κανονική κατανομή και βασίζεται στα μερικά αθροίσματα των αθροιστικών αποκλίσεων των τιμών της σειράς από τη μέση τιμή τους.

Θεωρούμε την χρονοσειρά $y_i, i\in I_n$ των **n** μηνιαίων υψών βροχής του ίδιου μήνα για **n** χρόνια και υπολογίζουμε τη μέση τιμή τους E(Y) σύμφωνα με το τύπο (2.8).

Ακολούθως υπολογίζουμε τα μερικά αθροίσματα των αποκλίσεων των **n** τιμών από τη μέση τιμή E(Y):

$$S_k={}_1\Sigma^k(y_i-E(y)),\qquad k\in I_n\,, \tag{4.14}$$

όπου είναι:

$$S_n=0$$

$$S_o=0$$

Για ομοιογενή δείγματα, τα S_k θα πρέπει να κυμαίνονται γύρω από το μηδέν γιατί δεν πρέπει να υπάρχει συστηματική απόκλιση των y_i από τη μέση τιμή $E(y)$.

Μια πιο ευαίσθητη στατιστική πάνω στις αθροιστικές αποκλίσεις επιτυγχάνεται με τα τροποποιημένα αθροίσματα τα οποία ορίζονται από τις σχέσεις:

$$S_k^*=S_k/S_y, \qquad k=0,1,2,\ldots n, \tag{4.15}$$

όπου S_y^2 η διασπορά που υπολογίζεται από τον γνωστό τύπο :

$$S_y^2=\{{}_1\Sigma^n(y_i-E(Y))^2\}/n \tag{4.16}$$

Αν ορισθεί η στατιστική παράμετρος :

$$Q=\max|S^*_k|, \qquad \text{για} \qquad 0\leq k\leq n \tag{4.17}$$

τότε υψηλές τιμές της Q δείχνουν αλλαγή στο επίπεδο της μέσης τιμής. Οι κρίσιμες τιμές για το Q στα διάφορα επίπεδα εμπιστοσύνης δίδονται στον Πίνακα (4.8).

Παράδειγμα 4.3.5.

Στον Πίνακα (4.7) δίνονται τα μηνιαία ύψη υετού του μήνα Αυγούστου για 30 χρόνια στον σταθμό Σ_1 (στήλη 2).

Υπολογίζεται η μέση τιμή της χρονοσειράς $E(Y)=100.0$ και η τυπική απόκλιση $S_y=61.66$

Ακολούθως από τον τύπο (4.14) υπολογίζονται τα μερικά αθροίσματα S_k (στήλη 3) και τα τροποποιημένα αθροίσματα S^*_k (στήλη 4) από τα οποία διαπιστώνεται ότι $Q=\max|S^*_k|=4.96$ και, επομένως, είναι: $Q/\sqrt{n}=Q/\sqrt{30}=4.96/5.48=0.91$.

Από τον Πίνακα (4.8) για στάθμη σημαντικότητας $\alpha=95\%$ και για $n=30$, η οριακή τιμή του $Q/\sqrt{30}=1.24$ και επειδή η υπολογισμένη τιμή είναι 0.91, η χρονοσειρά είναι ομοιογενής.

Πίνακας 4.7. Μηνιαία ύψη υετού, μερικά αθροίσματα και τροποποιημένα αθροίσματα.

Ετος	Μηνιαία ύψη υετού Μαίου	S_k	S^*_k
1	250.	150.0	2.82
2	147.	197.0	3.70
3	83.	180.0	3.38
4	108.	188.0	3.53
5	171.	259.0	4.86
6	62.	221.0	4.15
7	67.	188.0	3.53
8	119.	207.0	3.89
9	157	264.0	4.96
10	23.	187.0	3.51
11	78.	165.0	3.10
12	79.	144.0	2.70
13	85.	129.0	2.42
14	18.	47.0	.88
15	105.	52.0	.98
16	48.	.0	.00
17	41.	-59.0	-1.11
18	44.	-115.0	-2.16
19	133.	-82.0	-1.54
20	158.	-24.0	-0.45
21	54.	-70.0	-1.31
22	72.	-98.0	-1.84
23	49.	-149.0	-2.80
24	110.	-139.0	-2.61
25	100.	-139.0	-2.61
26	125.	-114.0	-2.14
27	57.	-157.0	-2.95
28	206.	-51.0	-0.96
29	107.	-44.0	-0.83
30	144.	0.0	0.00

Πίνακας 4.8. Ποσοστιαία επί τοις % σημεία του $Q/\sqrt{n}$.

n	σ.σ	Q/√n 90%	Q/√n 95%	Q/√n 99%
10		1,05	1,14	1,29
20		1,10	1,22	1,42
30		1,12	1,24	1,46
40		1,13	1,26	1,50
50		1,14	1,27	1,52
100		1,17	1,29	1,55
∞		1,22	1,36	1,63

4.4. Αντικειμενικές μέθοδοι απόλυτης ομοιογένειας

Μια εναλλακτική μορφή ανάλυσης της ομοιογένειας είναι αυτή στην οποία η υπό έλεγχο σειρά δεδομένων συγκρίνεται μόνο με μία άλλη σειρά, η οποία προκύπτει από τους μέσους όρους των αντίστοιχων τιμών των σειρών από σταθμούς που περιβάλλουν τον υπό έλεγχο σταθμό. Στην περίπτωση όμως αυτή θα πρέπει να ληφθούν υπόψη ορισμένα στοιχεία, όπως είναι η συγκεκριμένη παράμετρος που μελετάται, η πυκνότητα των σταθμών στην περιοχή, το μέγεθος του αρχείου που ελέγχεται, η συχνότητα εμφάνισης ελλειπουσών παρατηρήσεων, οι μετατοπίσεις σταθμών καθώς και άλλοι παράγοντες π.χ. η κλιματική ομοιομορφία της περιοχής κ.λ.π.

'Ενα test που μπορεί να χρησιμοποιηθεί ακολούθως για την υπο έλεγχο σειρά και την προκύπτουσα σειρά των μέσων όρων, είναι το test Von Neumann που αναφέρθηκε προηγουμένως.

4.5. Μη παραμετρικές διαδικασίες ελέγχου της ομοιογένειας

4.5.1.Γενικά. Ο έλεγχος ομοιογένειας είναι ένας **στατιστικος έλεγχος υποθέσεων**, ο οποίος απαιτεί μία υπόθεση ομοιογένειας (μηδενική υπόθεση-H_o) και έναν κανόνα για αποδοχή ή απόρριψη αυτής της υπόθεσης.'Ετσι, αν η πιθανότητα αποδοχής της ομοιογένειας είναι μικρή, συμπεραίνεται ότι πιθανώς η σειρά δεν είναι ομοιογενής. Αντίθετα, αν αυτή είναι μεγάλη, τότε λαμβάνεται η απόφαση για αποδοχή της μηδενικής υπόθεσης, δηλαδή, για αποδοχή της ομοιογένειας. Στον κανόνα πρέπει να καθορίζεται η πιθανότητα (στάθμη σημαντικότητας- (σ.σ)) πέρα από την οποία η υπόθεση ομοιογένειας θα πρέπει να απορρίπτεται και να γίνεται αποδεκτή μία εναλλακτική υπόθεση (αναλυτικά παρατίθενται τα παραπάνω στην παράγραφο 1.5).

Αν είναι γνωστή η συνάρτηση κατανομής της μετεωρολογικής παραμέτρου της οποίας ελέγχεται η χρονοσειρά για ομοιογένεια, τότε μπορούν να γίνουν έλεγχοι με βάση την κατανομή. Στις περισσότερες περιπτώσεις όμως, είναι δύσκολο να καθορισθούν οι κατανομές των μεταβλητών που ελέγχονται για ομοιογένεια και γι'αυτό χρησιμοποιούνται **μη-παραμετρικοί έλεγχοι,** όπως λέγονται, στους οποίους δεν είναι απαραίτητη η γνώση της κατανομής της τυχαίας μεταβλητης-παραμέτρου για τον έλεγχο της ομοιογένειας .

Ο πιο γνωστός μη παραμετρικός έλεγχος είναι αυτός του **X^2-test** ενώ πολύ συχνά στη μετεωρολογία χρησιμοποιείται και το "run test",που είναι ευαίσθητο σε όλες αυτές τις εναλλακτικές υποθέσεις. Στις επόμενες παραγράφους περιγράφονται σημαντικοί μη παραμετρικοί έλεγχοι.

4.5.2. Το X^2-test ως test ομοιογένειας χρονοσειράς. Το X^2-test σαν test ομοιογένειας περιγράφεται διεξοδικά στην παράγραφο 5.2 όπου εξετάζεται και η επιλογή του "καλύτερου" γειτονικού σταθμού για την ομοιογενοποίηση ανομοιογενούς σειράς.

4.5.3.`Ελεγχος "run"(Run test). Ο Thom **[Βιβλ.9]** χρησιμοποιεί τον μη-παραμετρικό έλεγχο "run" για να καθορίσει αν μια σειρά δεδομένων είναι ομοιογενής ή, διαφορετικά, αν υπάρχει μια τάση για περιοδική παλινδρόμηση του μέσου.

Σ'αυτόν τον έλεγχο, σε κάθε τιμή της προς έλεγχο χρονοσειράς που είναι μεγαλύτερη από, ή ίση με, τη διάμεσο (median) αντιστοιχεί ένα σύμβολο ή ένας αριθμός, π.χ. Α και σε κάθε τιμή μικρότερη από τη διάμεσο ένα άλλο σύμβολο ή ένας άλλος αριθμός π.χ. Β. Κατόπιν μετράται ο αριθμός των "runs u ", δηλαδή ο αριθμός **των διαδοχικών σειρών από Α ή Β** που ελέγχεται με τη βοήθεια της κατανομής του u.

Για να γίνει πιο κατανοητό το "run test" παρατίθεται το παρακάτω παράδειγμα, όπως ακριβώς αναφέρεται από τον Thom [Βιβλ.9].

Πίνακας 4.9. "Runs" για την παρατηρηθείσα θερμοκρασία της Γενεύης.

1927	17,4	B	1	1937	19.5	A	6	1947	22,0	A	
1928	20,9	A	2	1938	18,6	B		1948	18,9	B	11
1929	18,7	B		1939	18,6	B	7	1949	20,7	A	
1930	18,7	B	3	1940	17,9	B		1950	19,7	A	
1931	16,9	B		1941	17,8	B		1951	14,5	A	12
1932	20,8	A		1942	19,9	A		1952	20,3	A	
1933	20,4	A	4	1943	20,9	A	8	1953	19,8	A	
1934	17,9	B		1944	22,9	A		1954	18,3	B	13
1935	18,1	B	5	1945	18,9	B	9	1955	19,3	A	14
1936	18,5	B		1946	19,2	A	10	1956	17,5	B	15

Από τον Πίνακα (4.9) φαίνεται ότι η διάμεσος βρίσκεται μεταξύ 18,9 και 19,2 και θεωρούμε ότι είναι η τιμή 19,05. Κάθε θερμοκρασία στον Πίνακα (4.9) συγκρίνεται ως προς τήν τιμή αυτή και χαρακτηρίζεται με Α ή Β, ανάλογα με το αν είναι πάνω (μεγαλύτερη) ή κάτω (μικρότερη) από τη διάμεσο, αντίστοιχα. Ακολούθως, τα "runs u" σημειώνονται ως διαδοχικές σειρές των Α και Β. Στο παράδειγμα, ο ολικός αριθμός των "runs" είναι u=15.

Είναι σαφές ότι, πολλά "runs" θα ήταν ένδειξη παλινδρόμησης, ενώ λίγα "runs" θα έδειχναν μια τάση ή μια ολίσθηση της μεσαίας τιμής στη διάρκεια του καταγραμμένου δείγματος. Συνεπώς, αν η πιθανότητα να ξεπερασθεί ένα u είναι μικρή, θα μπορούσε κανείς να υποψιασθεί παλινδρόμηση ενώ αν η πιθανότητα εμφάνισης τιμής μικρότερης του u ενός δείγματος είναι μικρή, τότε θα υποψιαζόταν κανείς μια τάση της διαμέσου. Εάν η πιθανότητα να υπάρξει τιμή μεγαλύτερη ή μικρότερη του u είναι μεγάλη, τότε λέγεται ότι η σειρά των δεδομένων είναι ομοιογενής ή προέρχεται από ένα και μόνο πληθυσμό. Για να γίνει το test χρειάζεται ένας Πίνακας κατανομής του u, ο οποίος παρέχεται παρακάτω (**Πίνακας 4.10**). Επειδή έχει επιλεγεί η διάμεσος τιμή, για τη σύγκριση των τιμών, ο αριθμός των τιμών πάνω από αυτή , N_A, είναι ίσος με τον αριθμό των τιμών κάτω από αυτή, N_B και γι'αυτό ο Πίνακας αφορά την υποθεση $H_0: N_A=N_B$.

Πίνακας 4.10. Κατανομή του αριθμού των "runs u "

				$N_A=N_B$				
	P	P		P	P		P	P
N_A	0.10	0.90	N_A	0.10	0.90	N_A	0.10	0.90
10	8	13	15	12	19	20	16	25
11	9	14	16	13	20	25	22	30
12	9	16	17	14	21	30	26	36
13	10	17	18	15	22	35	31	41
14	11	18	19	16	23	40	35	47
						45	40	52
						50	45	57

Ο Πίνακας (4.10) δίνει τις τιμές του u που αντιστοιχούν στις πιθανότητες που είναι πλησιέστερα στις τιμές 0.10 και 0.90. Εάν ένα δείγμα u, είναι κάτω από το κατώτερο όριο, η ανομοιογένεια οφείλεται σε μία τάση ή ολίσθηση της διαμέσου ενώ αν είναι πάνω από το ανώτερο όριο, οφείλεται σε παλινδρόμηση.

Βλέπουμε ότι στο παράδειγμα, u=15 και $N_A=N_B=15$. Τα όρια για $N_A=15$ στον Πίνακα (4.10) είναι 12 και 19 και συνεπώς, επειδή u=15, δηλαδή το u δεν είναι σημαντικά διαφορετικό από τα u που αναμένονται σε ομοιογενείς σειρές, συμπεραίνουμε ότι η σειρά είναι ομοιογενής.

4.5.4. Κριτήριο kolmogorov-Smirnov για έλεγχο της ομοιογένειας δύο δειγμάτων. Μια μη παραμετρική δοκιμασία για τον έλεγχο της ομοιογένειας, είναι και το **κριτήριο Kolmgorov-Smirnov** για δύο δείγματα **[βιβλ.3]**.

'Εστω ότι ο σταθμός Σ λειτούργησε **n** χρόνια (μήνες ή δεκαήμερα) στη θέση Σ_1 και m χρόνια (μήνες ή δεκαήμερα) στη θέση Σ_2.'Εστω X είναι μία μετεωρολογική παράμετρος και $x_1,x_2,...,x_n$ είναι οι μέσοι όροι ή κάποια άλλη δειγματική παράμετρος (π.χ. αριθμός ημερών παγετού κ.λ.π.) στη θέση Σ_1, πριν απο την μετακίνηση του σταθμού και $y_1, y_2, ..., y_m$ μετά τη μετακίνηση του σταθμού στη θέση $Σ_2$.

Εξετάζουμε τις υποθέσεις **H_o και H_1**, όπου :

H_o: Τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό (είναι ομοιογενή).

H_1: Τα δύο δείγματα προέρχονται από διαφορετικούς πληθυσμούς (διαφορετικές κατανομές).

Ο έλεγχος γίνεται με τη χρήση της παραμέτρου **$D_{m,n}$**, που ορίζεται ως εξής:

$$D_{m,n}=\max_x|F_m(X)-G_n(X)| \tag{4.18}$$

και **$F_m(X)$ και $G_n(Y)$** οι αντίστοιχες αθροιστικές σχετικές συχνότητες των δύο δειγμάτων. Η απορριπτική περιοχή για τον έλεγχο της H_o, είναι αυτή για την οποία ισχύει:

$$D_{m,n}>D_{a,m,n} \tag{4.19}$$

όπου η **$D_{m,n}$** δίνεται από την (4.18) και η τιμή **$D_{a,m,n}$** βρίσκεται από τους σχετικούς Πίνακες **IVα, IVβ** του Παραρτήματος, για τις διάφορες τιμές των m,n και της σ.σ α.

Αν ισχύει **$D_{m,n}<D_{a,m,n}$** σε στάθμη σημαντικότητας α=0.05, δεν μπορεί να απορριφθεί η H_o, πράγμα που σημαίνει ότι τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό.

Παράδειγμα 4.5.4.

'Εστω ότι η παράμετρος Χ είναι ο μηνιαίος αριθμός ημερών υετού και έχουμε δύο δείγματα της Χ στον σταθμό Σ, το ένα από τη θέση $Σ_1$ και το άλλο από τη θέση $Σ_2$, για τους μήνες Ιουνίου και Ιουλίου.

Το πρώτο δείγμα περιέχει τιμές 30 χρόνων ενώ το δεύτερο περιέχει τιμές 15 χρονων, δηλαδή για τους δυο μήνες: n=2*30=60 και m=2*15=30.

Στον Πίνακα (4.11), στην 1η στήλη δίνονται οι δυνατές διακεκριμένες τιμές που βρέθηκαν από τα δείγματα των 60 τιμών x_i, $i \in I_{60}$ και των 30 τιμών y_i, $i \in I_{30}$, αντίστοιχα, για την κλιματολογική παράμετρο Χ "αριθμός ημερών υετού τον Ιούνιο και Ιούλιο", ενώ στην 2η και 3η στήλη είναι οι απόλυτες συχνότητες για το 1ο και το 2ο δείγμα, αντίστοιχα, και οι αθροιστικές σχετικές συχνότητες στην 4η και 5η στήλη. Η 6η στήλη περιέχει τις διαφορές των αθροιστικών συχνοτήτων μεταξύ των δύο δειγμάτων.

Πίνακας 4.11.Συχνότητες μηνιαίου αριθμού ημερών υετού για δύο δείγματα και διαφορές των αντίστοιχων τιμών τους.

Απόλυτες Συχνότητες			Αθροιστικές Συχνότητες-Διαφορές		
X	Σ_1	Σ_2	Σ_1	Σ_2	$\lvert F_m(X)-G_n(X)\rvert$
0	10	7	.167	.233	.067
1	15	8	.417	.500	.083
2	5	2	.500	.567	.067
3	7	3	.617	.667	.050
4	2	2	.650	.733	.083
5	4	2	.717	.800	.083
6	9	2	.867	.867	.000
7	2	1	.900	.900	.000
8	4	2	.967	.967	.000
9	2	1	1.000	1.000	.000
Σύνολο	60	30			

Από την τελευταία στήλη διαπιστώνουμε ότι η μέγιστη τιμή των διαφορών είναι: $D_{60,30}$=0.83.
Απο τους σχετικούς Πίνακες IVα,IVβ, που αναφέρθηκαν παραπάνω για (σ.σ) α=0.05 προκύπτει:

$$D_{60,30,.05}=1.36*\frac{(60+30)^{1/2}}{(60*30)^{1/2}}=0.30$$

Επειδή 0.83=$D_{60,30}>D_{60,30,.05}$=0.30 συμπεραίνουμε ότι δεν μπορούμε να δεχτούμε την H_0, πράγμα που σημαίνει ότι τα δύο δείγματα δεν προέρχονται από τον ίδιο πληθυσμό, στην παραπάνω σ.σ.

4.6. Έλεγχος ανομοιογένειας πολλών δειγμάτων

4.6.1. Γενικά. Πολλές φορές θεωρείται χρήσιμο, αν υπάρχουν αρκετές ανεξάρτητες σειρές παρατηρήσεων της ίδιας μετεωρολογικής παραμέτρου, να εξετάζεται η υπόθεση ότι **αυτές οι σειρές είναι ομοιογενείς**, δηλαδή ότι έχουν όλες την ίδια κατανομή. Σ'αυτή την περίπτωση, όλες οι τιμές μαζί μπορούν να αποτελέσουν ένα ενιαίο δείγμα παρατηρήσεων που το μέγεθος του είναι μεγαλύτερο από το μέγεθος καθενός δείγματος ξεχωριστά και, επομένως, δίνει καλύτερες εκτιμήσεις των στατιστικών παραμέτρων από ό,τι το κάθε δείγμα χωριστά.

Σε μια τέτοια περίπτωση, για κάθε έλεγχο ομοιογένειας της σειράς χρειάζεται να γίνει ο έλεγχος της ομοιογένειας των μέσων των δειγμάτων καθώς και ο έλεγχος της συμβατότητας των μερικών δειγματικών διασπορών τους προς τη διασπορά ολόκληρου του δείγματος, εξετάζοντας το λόγο των διασπορών. Όμως, ένας τέτοιος έλεγχος είναι μόνο προσεγγιστικός και γι'αυτό το λόγο είναι καλύτερα να χρησιμοποιούνται μη παραμετρικά tests, τα οποία εφαρμόζονται συγχρόνως για τον έλεγχο της ομοιογένειας και των μέσων και των διασπορών.

4.6.2. Δοκιμασία ομοιογένειας δύο ανεξαρτήτων δειγμάτων των Mann και Whitney. Το test των Mann-Whitney, γνωστό σαν U-test **[βιβλ.3]**, είναι ένα από τα ισχυρότερα μη-παραμετρικά tests και χρησιμοποιείται για τον έλεγχο της ομοιογένειας δύο δειγμάτων A και B.

Για την εφαρμογή του U-test δεν είναι απαραίτητη η γνώση της κατανομής του πληθυσμού από τον οποίο λήφθηκαν τα τυχαία δείγματα. Τη μόνη προϋπόθεση που πρέπει να πληρούν τα στοιχεία της τυχαίας μεταβλητής X, από την οποία λαμβάνονται τα δείγματα, είναι να είναι αυτά διατάξιμα. Οι μετεωρολογικές παράμετροι στο σύνολο τους θεωρούμενες σαν τυχαίες μεταβλητές πληρούν την προϋπόθεση αυτή και έτσι το U-test είναι κατάλληλο για το σύνολο των μετεωρολογικων παραμέτρων.

Έστω ότι οι τιμές της παραμέτρου X αποτελούν την χρονοσειρά $x_1, x_2, .., x_K$ και ότι οι K αυτές τιμές χωρίζονται σε δύο σύνολα τα Y και W που το καθένα αποτελεί μια χρονοσειρά. **Την Y: $y_1=x_1$, $y_2=x_2, ..., y_n=x_n$ και την W: $w_1=x_{n+1}$,**

$w_2=x_{n+2},\ldots,w_k=x_{n+m}$ **με** $n+m=k$.
Ο δείκτης **n** στη χρονοσειρά Y αντιστοιχεί σε εκείνο τον χρόνο κατά τον οποίο το ιστορικό του σταθμού ή η εφαρμογή μιας απο τις γραφικές μεθόδους που αναφέρθηκαν προηγουμένως μας οδηγούν στην υποψία ότι διακόπτεται η ομοιογένεια της χρονοσειράς, δηλαδή, ότι οι δύο προκύπτουσες χρονοσειρές Y και W προέρχονται από διαφορετικούς πληθυσμούς.

Από τις τιμές της χρονοσειράς Y επιλέγεται τυχαίο δείγμα A μεγέθους n1 και απο τις τιμές της χρονοσειράς W ένα τυχαίο δείγμα B μεγέθους n2 με $\mathbf{n1 \le n2}$.

Το U-test βασίζεται στο ότι, αν η αρχική χρονοσειρά X είναι ομοιογενής τοτε οι κατανομές των τιμών των διαφόρων δειγμάτων απ'αυτή δεν θα πρέπει να διαφέρουν σημαντικά μεταξύ τους. Αυτό όμως σημαίνει ότι αν y είναι μια τυχούσα τιμή από το δείγμα A και w μια τυχούσα τιμή από το B θα ισχύει: $\mathbf{P(y \ge w)=P(y \le w)=0.5}$. Δηλαδή σ'ένα ενιαίο δείγμα, που θα περιέχει τις τιμές του A και του B, ο αριθμός των περιπτώσεων κατά τις οποίες οι τιμές του A προηγούνται (είναι μικρότερες) από τις τιμές του B θα πρέπει να είναι ίσος με τον αριθμό των περιπτώσεων που οι τιμές του A έπονται (είναι μεγαλύτερες) αυτών του B.

Έτσι, αν κατασκευαστεί ένα νέο δείγμα του οποίου τιμές είναι οι $\mathbf{m=n1+n2}$ τιμές των δειγμάτων **A και B** (το τελευταίο δείγμα το συμβολίζουμε με **A+B**) και διαταχθούν αυτές οι τιμές του A+B κατά αύξουσα τάξη μεγέθους (μία απλή διάταξη των τιμών στο A+B), βρίσκουμε τους αριθμούς $\mathbf{U_A}$ **και** $\mathbf{U_B}$ που παριστούν το πλήθος των περιπτώσεων κατά τις οποίες οι τιμές του δείγματος A προηγούνται από τιμές του δείγματος B, (U_A) και, αντίστροφα, οι τιμές του B προηγούνται των τιμών του A, (U_B). Αν η χρονοσειρά X είναι ομοιογενής, για να είναι $\mathbf{P(y \ge w)=P(y \le w)=0.5}$, **πρέπει** $\mathbf{U_A=U_B}$.

Τα στοιχεία $\mathbf{y_i,\ i \in I_m}$ του συνόλου **A+B** καταλαμβάνουν τις θέσεις **i**, $\mathbf{i \in I_m}$. Αν τα στοιχεία του δείγματος A καταλαμβάνουν τις θέσεις $\mathbf{K_1,K_2,\ldots K_{n1}}$ και τα στοιχεία του B τις θέσεις $\mathbf{\Lambda_1,\Lambda_2,\ldots\Lambda_{n2}}$, τότε ορίζονται τα **αθροίσματα των τάξεων** $\mathbf{R_A}$ **και** $\mathbf{R_B}$ **από:**

$$R_A=K_1+K_2+\ldots K_{n1} \quad \text{και} \tag{4.20}$$

$$R_B=\Lambda_1+\Lambda_2+\ldots\Lambda_{n2} \tag{4.21}$$

Για τους αριθμούς U_A, U_B, R_A και R_B ισχύουν:

$U_A=n1*n2-R_A+n1*(n1+1)/2$ (4.22)

$U_B=n1*n2-R_B+n2*(n2+1)/2$ (4.23)

$R_A+R_B=m*(m+1)/2=(n1+n2)*(n1+n2+1)/2$ (4.24)

$U_A+U_B=n1*n2$ (4.25)

Το στατιστικο $U_{n1,n2}$ που απαιτειται για την εφαρμογή του U-test δίνεται απο την σχέση:

$U_{n1,n2}=\min\{U_A,U_B\}$. (4.26)

Για πολύ μικρά δείγματα, $n_1,n_2 \leq 8$, η διαδικασία του U-test ομοιογένειας των Mann και Whitney είναι η εξής:

Αν διατίθενται τα δείγματα Α και Β καθορίζονται πρώτα τα παρακάτω:

α) Η μηδενική υποθεση H_0: τα δείγματα Α και Β προέρχονται απο ομοιογενή πληθυσμό, δηλαδή: $U_A=U_B$.

β) Η εναλλακτική υποθεση H_1: τα δείγματα Α και Β δεν είναι δείγματα απο τον ίδιο πληθυσμό, δηλαδή η χρονοσειρά Χ είναι ανομοιογενής.

Μετά ορίζεται η στάθμη σημαντικοτητας, π.χ. α=0.05, και συγκρίνεται η τιμή του υπολογιζόμενου στατιστικού $U_{n1,n2}$ με εκείνη που βρίσκεται απο την κατανομή του U, εφ'όσον ισχύει η H_0 και που δίνεται απο τους κατάλληλους πίνακες των Mann και Whitney (Πίνακες Vα, Vβ, Vγ του Παραρτήματος). Οι πίνακες αυτοί παρέχουν την πιθανότητα για συγκεκριμένα U, n1, n2 να είναι οι τιμές του U τόσο μεγάλες, όσο το παρατηρούμενο U.

Η υποθεση H_0 απορρίπτεται όταν **η πιθανότητα** να είναι οι τιμές του U μικροτερες ή ίσες απο το υπολογισμένο $U_{n1,n2}$ είναι **μικροτερη** απο την στάθμη σημαντικοτητας α, δηλαδή, $P(U<U_{n1,n2})\leq α=0.05$.

Για δείγματα μεγαλύτερα, όπου $9\leq n1\leq 20$ και $9\leq n2\leq 20$, οι Mann και Whitney παρέχουν άλλους πίνακες (Πίνακες Vδ,Vε,Vστ του Παραρτήματος) στους οποίους δίνονται οι κρίσιμες τιμές του U για διάφορες στάθμες σημαντικότητας α.

Αν η υπολογισμένη τιμή του $U_{n1,n2}$ για τα ορισμένα n1,n2 όρια που αναφέρθηκαν παραπάνω, είναι μικρότερη ή ιση απο την τιμή $U_{n1,n2,\alpha}$ του Πίνακα, τότε η H_0 απορρίπτεται για την **αντίστοιχη στάθμη σημαντικότητας.**

Η κατανομή του U για μεγάλες τιμές των n_1 και n_2, δηλαδή, **$n_1>9$ και $n_2>20$,** είναι κανονική με μέση τιμή $\mu_u=n_1*n_2/2$ και τυπική απόκλιση $\sigma_u^2=n_1*n_2*(n_1+n_2+1)/12$, δηλαδή $U\approx N(U,\mu_u,\sigma_u^2)$.

Η υπόθεση H_o είναι: Μπορεί τόσο μεγάλες ή τόσο μικρές τιμές του U όπως οι παρατηρούμενες να οφείλονται σε τυχαία γεγονότα (οπότε η χρονοσειρά X είναι ομοιογενής), σε σ.σ α (α=0.05);
Η εναλλακτική υπόθεση H_1 είναι: Δεν μπορεί τόσο μεγάλες τιμές του U να οφείλονται στην τύχη.
Και στο U-test ισχύουν όλα όσα ισχύουν και κατά την εφαρμογή των υπόλοιπων tests, δηλαδή σε μονόπλευρο (μόνο τόσο μεγάλες ή μόνο τόσο μικρές τιμές του U) το α λαμβάνεται όπως είναι, ενώ σε δίπλευρο test λαμβάνεται το α/2 αντί του α στον Πίνακα της κανονικής κατανομής (Πίνακας ΙΒ του Παραρτήματος) .

Παράδειγμα 4.6.2.α. (μικρά δείγματα).
῾Εστω ότι το A αποτελείται από τις τιμές: -1,3,2,6 (n_1=4) και το B από τις: -1.2, 2.5, 3.5, 4, 5 (n_2=5). Τα στοιχεία y_i (9 στοιχεία) του διατεταγμένου A+B δίνονται στον Πίνακα 4.12.
Σ΄αυτόν τον Πίνακα, κάτω από κάθε αριθμό y_i, $i\in I_9$, αναγράφεται α ή β ανάλογα με το αν το $y_i\in A$ ή $y_i\in B$ και κάτω από την σειρά αυτή γράφεται η τάξη του y_i μέσα στο διατεταγμένο A+B.

Για να υπολογισθεί το U_A, πρέπει να βρεθεί ο αριθμός των περιπτώσεων που οι τιμές του A προηγούνται από τις τιμές του B. Για το 1εB, δεν υπάρχει καμμιά τιμή του A να προηγείται, για το 4εB, υπάρχουν δύο τιμές, για τις τιμές 6, 7 και 8 του B, προηγούνται 3 τιμές του A και συνεπώς U_A=0+2+3+3+3=11. Επομένως $U_B=n_1*n_2-U_A$=4*5-11=9.
῾Ομως τα U_A και U_B μπορεί να υπολογισθούν και από τις:

$R_A=K_1+K_2+K_3+K_4=2+3+5+9=19$

$R_B=\Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4+\Lambda_5=1+4+6+7+8=26$

$U_A=4*5+4*5/2-R_A=20+10-19=30-19=11$

$U_B=4*5+5*6/2-R_B=20+15-26=35-26=9$

Συνεπώς $U=\min(U_A,U_B)=9$

Σ΄αυτό το παράδειγμα είναι U=9, n_1=4, n_2=5 και από τον Πίνακα Vα, βρίσκεται **P=.452 > α =0.05**, νεπώς δεν μπορεί να απορριφθεί η H_o σε σ.σ. α=0.05 και γίνεται αποδεκτό ότι τα δείγματα προέρχονται από ομοιογενή πληθυσμό.

Πινακας 4.12. Διατεταγμένο Α+Β

-1.2,	-1,	2,	2.5,	3,	3.5,	4,	5,	6
β	α	α	β	α	β	β	β	α
1	2	3	4	5	6	7	8	9
$Λ_1$	K_1	K_2	$Λ_2$	K_3	$Λ_3$	$Λ_4$	$Λ_5$	K_4

Παράδειγμα 4.6.2.β. (μεγάλα δείγματα).
`Εστω ότι n1=20, n2=24 και U=330. Είναι: μ_u=n1*n2/2=240 και σ_u^2=20*24*45/12=1800. Για σ.σ α=0.05 και δίπλευρο test από τον Πίνακα της φυσικής κατανομής (Πίνακας ΙΒ του Παραρτήματος) βρίσκεται z_α=1.96. Εδώ είναι:
$|z|=|(U-\mu_u)/\sigma_u|=(330-240)/\sqrt{1800}=90*\sqrt{2}/(30*2)=1.5*\sqrt{2}\approx 1.5*1.4$ ≈ 2.1, δηλαδή $|z|>z_\alpha$ και άρα δεν μπορεί να γίνει δεκτή η υπόθεση H_o σε σ.σ. α=0.05.

4.6.3.Μέθοδος ελέγχου της ομοιογένειας με το test Kruskal-Wallis. Σ΄αυτή τη μέθοδο ελέγχου της ομοιογένειας, καθεμία παρατήρηση μετατρέπεται σε τάξη, όταν όλες οι παρατηρήσεις διαταχθούν κατά αύξουσα σειρά. **(Kruskal-Wallis [Βιβλ.3,8]).**

Αν υπάρχουν κ δείγματα παρατηρήσεων, **τα $Δ_i, i\in I_\kappa$,** δημιουργείται το δείγμα **$Δ={}_1Σ^\kappa Δ_i$**, διατάσσονται όλες οι τιμές σε αύξουσα σειρά και στη συνέχεια υπολογίζονται τα αθροίσματα των τάξεων **N_i, $i\in I_\kappa$** για κάθε δείγμα $Δ_i$ χωριστά. Το test καθορίζει κατά πόσο τα αθροίσματα αυτά είναι τόσο διαφορετικά που να είναι απίθανο να προέρχονται απο τον ίδιο πληθυσμό.

Αν τα κ δείγματα είναι ομοιογενή, δηλαδή η μηδενική υπόθεση της ομοιογένειας των μέσων ισχύει, τότε η στατιστική παράμετρος Η που ορίζεται από την σχέση:

$$H=\frac{12}{n*(n+1)} * {}_1Σ^\kappa \frac{R_j^2}{n_j} - 3*(n+1) \qquad (4.27)$$

όπου:

κ	:το πλήθος των δειγμάτων
η_j	:το πλήθος των παρατηρήσεων του j δείγματος
$\eta={}_1\Sigma^{\kappa}\eta_j$	:το σύνολο των παρατηρήσεων όλων των δειγμάτων μαζί (δηλαδή του δείγματος Δ)
r_{ij}	:η τάξη στο διατεταγμένο συνολικό δείγμα Δ της παρατήρησης i του δείγματος j
$R_j={}_1\Sigma r_{ij}$	:το άθροισμα των τάξεων στο j δείγμα

ακολουθεί την X^2 κατανομή με $v=\kappa-1$ βαθμούς ελεύθεριας (Πίνακες ΙΙα,ΙΙβ του Παραρτήματος). Η σχέση (4.27) ισχύει εφ'όσον τα μεγέθη των δειγμάτων **δεν** είναι πολύ μικρά, πράγμα που ισχύει για τα μετεωρολογικά δείγματα.

`Οταν όλα τα $\eta^j>5$, και η H_o είναι αληθής, η πιθανότητα να είναι η τιμή του Η μεγαλύτερη από την παρατηρούμενη, δίνεται από τους πίνακες κατανομής της X^2. Αν για δεδομένη στάθμη σημαντικότητας α ισχύει: $H\geq X^2_{v,\alpha}$, τότε δεν μπορεί να γίνει αποδεκτή η H_o ενώ άν $H<X^2_{v,\alpha}$ **αυτή γίνεται αποδεκτή.**

Αν συμβεί κάποιες τιμές να εμφανίζονται στο δείγμα Δ περισσότερες από μία φορές, τότε ο τύπος για το στατιστικό Η διορθώνεται με την ακόλουθη διαδικασία:

Αν η τιμή τ_1 **εμφανίζεται** λ_1 **φορές**

η τιμή τ_2 **εμφανίζεται** λ_2 **φορες**

---, **και** **(4.28)**

η τιμή τ_p **εμφανίζεται** λ_p **φορές**

όπου όλα τα $\lambda_i\geq 2$ **, τότε υπολογίζεται η ποσότητα:**

$$C=1-\frac{{}_1\Sigma^{p}\lambda_j^{3}-{}_1\Sigma^{p}\lambda_j}{\eta^3-\eta}=\frac{{}_1\Sigma^{p}(\lambda_j^{3}-\lambda_j)}{\eta^3-\eta} \qquad (4.29)$$

και ακολούθως υπολογίζεται η Η' που είναι η νέα τιμή της Η από την :

$$H' = \frac{H}{C} = \frac{\dfrac{12}{n*(n+1)} * \sum_{j=1}^{k} \dfrac{R_j^2}{n_j} - 3*(n+1)}{C} \qquad (4.30)$$

Το αποτέλεσμα της διόρθωσης είναι να μεγαλώσει η τιμή του Η και να γίνει πιο σημαντικό το αποτέλεσμα.`Ετσι αν, χωρίς να γίνει η διόρθωση , απορριφθεί η H_o, θα απορριφθεί πολύ περισσότερο μετά την διόρθωση. Αν το ποσοστό των ίσων τιμών είναι μικρότερο από 25%, η πιθανότητα που αντιστοιχεί την τιμή του Η, σπάνια μεταβάλλεται περισσότερο από 10% όταν γίνει η διόρθωση.

Η ομοιογένεια των διασπορών εξετάζεται εφαρμόζοντας το ίδιο test αλλά αντικαθιστώντας αυτή τη φορά κάθε τιμή με την απολυτη τιμή της διαφοράς της από τον γενικό μέσο όρο.

Παράδειγμα 4.6.3.
Στο παράδειγμα που παρατίθεται, χρησιμοποιούνται τα μηνιαία ολικά ύψη υετού του σταθμού των Βρυξελλών στη διάρκεια δέκα συνεχών ετών **[βιβλ.8]**. Για να περιορισθούν οι διαφορές που οφείλονται είτε στην εποχική μεταβλητότητα της μέσης βροχόπτωσης είτε στο άνισο μήκος των διαφόρων μηνών του χρόνου, αυτές οι ποσότητες της βροχόπτωσης έχουν εκφρασθεί σε ποσοστό του μέσου όρου των δέκα ετών. Επιπρόσθετα για να περιορισθεί ο όγκος του δείγματος, εξετάζονται τα ποσά μόνο των μηνών Ιανουαρίου, Μαρτίου, Μαΐου, Ιουλίου, Σεπτεμβρίου και Νοεμβρίου (I,III,V,VII,IX,XI).

Το ολικό διατεταγμένο δείγμα των 60 στοιχείων παρατίθεται στον Πίνακα (4.13), όπου τα ποσά του υετού x_i παρουσιάζονται μαζί με την τάξη τους m, όπως βρίσκεται όταν τεθούν σε αύξουσα σειρά.

Στον Πίνακα (4.14) υπάρχουν οι **ίδιες** τιμές του Πίνακα (4.13) κατά χρονική σειρά στην διάρκεια των δέκα ετών, για κάθε μήνα χωριστά, όπου παρατίθενται οι τάξεις r_j στις αντίστοιχες τιμές x_j, οπως βρισκονται απο την ολική διατεταγμένη σειρά, με εξαιρεση των ισων τιμών οπου το r_j είναι η μέση τιμή των τιμών m που αντιστοιχούν σ'αυτές τις παρατηρήσεις.

Θα ελεγχθεί κατ'αρχήν η ομοιογένεια των μέσων με το Kruskal-Wallis test.

Πίνακας 4.13. Διατεταγμένα μηνιαία ύψη υετού για τους μήνες I,III,V,VI,IX,XI.

Τάξη	τιμή	τάξη	τιμή	τάξη	τιμή
m	x_i	m	x_i	m	x_i
1	15	21	81	41	114
2	22	22	83	42	117
3	24	23	84	43	124
4	32	24	85	44	131
5	33	25	91	45	132
6	37	26	92	46	135
7	41	27	92	47	136
8	42	28	95	48	140
9	45	29	97	49	142
10	51	30	97	50	143
11	55	31	99	51	145
12	55	32	99	52	146
13	56	33	104	53	156
14	58	34	104	54	162
15	62	35	109	55	172
16	62	36	109	56	173
17	67	37	113	57	187
18	69	38	113	58	190
19	75	39	113	59	197
20	75	40	114	60	204

Πίνακας 4.14. Μηνιαία ύψη υετού και αντίστοιχες τάξεις στο συνολικά διατεταγμένο δείγμα των 60 τιμών.

Μήνας	I		II		III		IV		V		VI	
Έτος	x	r	x	r	x	r	x	r	x	r	x	r
1	114	40,5	99	31,5	32	4	15	1	37	6	83	22
2	140	48	131	44	22	2	135	46	146	52	85	24
3	77	20	109	35,5	197	59	97	29,5	92	26,5	58	14
4	62	15,5	104	33,5	132	45	95	28	109	35,5	51	10
5	187	57	99	31,5	113	38	113	38	143	50	136	47
6	92	26,5	97	29,5	113	38	162	54	75	19	56	13
7	104	33,5	142	49	55	11,5	81	21	156	53	67	17
8	114	40,5	91	25	124	43	45	9	33	5	173	56
9	41	7	42	8	24	3	55	11,5	62	15,5	117	42
10	69	18	84	23	190	58	204	60	145	51	172	55
R_{ij}		306,5		310,5		301,5		298		313,5		300

Από τον Πίνακα (4.14) είναι:

$\Sigma R_j = 306,5+310,5+301,5+298+313,5+300=1830=60*61/2$, όπου $n=60$

Έχοντας $n_j=10$ σταθερό για κάθε $j \in I_{10}$ και $n=60$, από την (4.29) προκύπτει ότι $H=0.063$

Εάν εξαιρεθούν οι ομάδες των ίσων τιμών προκύπτει ο Πίνακας (4.15), όπου:

r_m: τάξη των ίσων τιμών,

λ_m: αριθμός ίσων τιμών της ίδιας ομάδας)

Πίνακας 4.15. Πίνακας των ίσων τιμών.

r_m	11,5	15,5	26,5	29,5	31,5	33,5	35,5	38	40,5
λ_m	2	2	2	2	2	2	2	3	2

και υπολογίζεται η ποσότητα:

$\Sigma\lambda=\Sigma(\lambda^3_j-\lambda_j)=(2^3-2)*8+(3^3-3)=72$, οποτε σύμφωνα με την (4.29) θα είναι:

$$C=1-\frac{\Sigma\lambda}{n(n+1)(n-1)}=0.999667$$

Η τιμή C είναι πολύ κοντά στο 1, οποτε με βάση την (4.30) θα είναι $H'=0.062$

Επειδή ο αριθμός των σειρών είναι 6, η κρίσιμη τιμή με την οποία πρέπει να συγκριθεί η τιμή της X, είναι εκείνη της X^2 με $v=6-1=5$ βαθμούς ελευθερίας. Για στάθμη σημαντικότητας $\alpha=0.05$ είναι: $X^2_{5,0.05}=11,07$, και επομένως η υπολογισμένη τιμή είναι πολύ μικρότερη από την κρίσιμη τιμή, πράγμα που επιτρέπει να γίνει αποδεκτή η ομοιογένεια των μέσων.

Για να ελεγχθεί η ομοιογένεια των διασπορών, ο Πίνακας (4.16) έχει οργανωθεί από τις απόλυτες τιμές των διαφορών των τιμών από το μέσο όρο, με τον ίδιο τρόπο που οργανώθηκε ο προηγούμενος πίνακας. Δηλαδή, για κάθε τιμή x_j υπολογίζεται η τιμή $\delta_j=|x_j-100|$ και ακολούθως η τάξη r_j από την συνολική διάταξη των τιμών δ_j.

Κι εδώ πάλι $\Sigma R_j=1830$ και με $n_j=10$, $n=60$ η τιμή της H είναι: $H=10,987$

Ο Πίνακας των ίσων τιμών (4.17) παρατίθεται παρακάτω.

Ἐτσι υπολογίζεται:

$\Sigma\lambda=(2^3-2)*9+(3^3-3)*3=126$ και ακολούθως:

$$C=1-\frac{\Sigma\lambda}{n(n+1)(n-1)}=0.999417$$

και βρίσκεται η διορθωμένη τιμή της H, $H'=H/C=10.993$ η οποία είναι λίγο μικρότερη από την κρίσιμη τιμή 11.07 της X^2 με 5 βαθμούς ελευθερίας και σε σ.σ=0.05 ($X^2_{5,0.05}=11.07$). Ἐτσι η υπόθεση της ομοιογένειας της διασποράς μπορεί να γίνει αποδεκτή.

Πίνακας 4.16. Απόλυτες τιμές των διαφορών των μηνιαίων υψών υετού από τη μέση τιμή των 10 ετών και αντίστοιχη τάξη μέσα στη διατεταγμένη συνολική σειρά των διαφορών.

Μήνας	I		II		III		IV		V		VI	
`Ετος	d	r	d	r	d	r	d	r	d	r	d	r
1	14	16.5	1	1.5	68	51	85	56	63	49	17	20.5
2	40	34	31	26.5	78	55	35	30	46	42	15	18
3	23	23	9	11	97	59	3	3.5	8	8.5	42	35.5
4	38	32.5	4	5.5	32	28	5	7	9	11	49	43
5	87	57	1	1.5	13	14	13	14	43	37	36	31
6	8	8.5	3	3.5	13	14	62	48	25	25	36	38
7	4	5.5	42	35.5	45	40	19	22	56	45	44	29
8	14	16.5	9	11	24	24	55	44	67	50	73	53
9	59	47	58	46	76	54	45	40	38	32.5	17	20.5
10	31	26.5	16	19	90	58	104	60	45	40	72	52
R_j	267		161		397		324.5		340		340.5	

Πίνακας 4.17. Πίνακας των ίσων τιμών.

r_e	1.5	3.5	5.5	8.5	11	14	16.5	20.5	26.5	32.5	35.5	40
λ_e	2	2	2	2	3	3	2	2	2	2	3	

4.6.4. Εφαρμογή του θεωρήματος 8 για τον έλεγχο της ομοιογένειας. Θεωρούμε ότι ο σταθμός Σ λειτούργησε n_1 χρόνια με την **"παλιά μέθοδο"** (στη θέση $Σ_1$ ή με τη μέθοδο παρατήρησης $Σ_1$ ή /και με το όργανο $Σ_1$) και n_2 χρόνια με τη **"νέα μέθοδο"** (στη θέση $Σ_2$ ή/και με τη μέθοδο παρατήρησης $Σ_2$ ή/και με το όργανο $Σ_2$) και έστω $S^2{}_1$ και $S^2{}_2$ οι τυπικές αποκλίσεις με την "παλιά" και "νέα" μέθοδο αντίστοιχα. Αν το σύνολο των τιμών του σταθμού (τιμές "παλιάς" και "νέας" μεθόδου) προέρχονται από **τον ίδιο πληθυσμό**, τότε οι S^2_1 και S^2_2 δεν θα πρέπει να διαφέρουν σημαντικά μεταξύ τους.

Το **θεώρημα 8** δίνει ένα τρόπο ελέγχου της ομοιογένειας των δύο δειγμάτων ως εξής:

Εισάγουμε τις τυχαίες μεταβλητές:

$$X_1 = \frac{(n_1-1)*S^2_1}{\sigma^2_1} \qquad (4.31)$$

και

$$X_2 = \frac{(n_2-1)*S^2_2}{\sigma^2_2} \qquad (4.32)$$

Οι X_1 και X_2 σύμφωνα με το **θεώρημα 7** ακολουθούν την X^2-κατανομή με $v_1=n_1-1$ **και** $v_2=n_2-1$ βαθμούς ελευθερίας αντίστοιχα. Επομένως η τυχαία μεταβλητή:

$$F = \frac{X_1/v_1}{X_2/v_2} = \frac{S^2_1*\sigma^2_2}{S^2_2*\sigma^2_1} \qquad (4.33)$$

στην περίπτωση που οι X_1 και X_2 είναι ανεξάρτητες, σύμφωνα με το **θεώρημα 8**, θα ακολουθεί την **F-κατανομή με v_1,v_2 βαθμούς ελευθερίας.**

Στους Πίνακες VIα,β,γ,δ του Παραρτήματος δίνονται, συναρτήσει των v_1, v_2 και της σ.σ.: α, οι τιμές $F_{v_1,v_2,\alpha}$ για τις οποίες ισχύει : $P(F>F_{v_1,v_2,\alpha})=\alpha$ (4.34)

Πρέπει να σημειωθεί ότι όλες οι τιμές $F_{v_1,v_2,\alpha}$ των πινάκων είναι μεγαλύτερες της μονάδας και γι'αυτό στο κλάσμα με το οποίο υπολογίζεται η τιμή F πρέπει να είναι αριθμητής η μεγαλύτερη από τις δύο τιμές S_1 και S_2.

Εφ'όσον ληφθεί υπόψη η παρατήρηση αυτή, **για μονόπλευρο**

test, τίθεται η μηδενική υποθεση H_o: $\sigma_1=\sigma_2=\sigma$ (συνεπώς, επειδή $\sigma^2{}_2/\sigma^2{}_1=1$, είναι $F=S^2{}_1/S^2{}_2$) και η εναλλακτική υποθεση: H_1= $\sigma_1>\sigma_2$, οποτε $\sigma^2{}_1/\sigma^2{}_2>1$.

Η μηδενική υποθεση H_o, όταν δίνονται τα v_1,v_2, απορρίπτεται για σ.σ. α, όταν για S_1,S_2 με $S_1>S_2$ η τιμή $F= S^2{}_1/S^2{}_2$ είναι μεγαλύτερη του $F_{v1,v2,\alpha}$ ($F>F_{v1,v2,\alpha}$).

Για **διπλευρο test**, θα πρέπει να αναφερθούν τα παρακάτω:

Για την F-κατανομή ισχύει η σχέση:

$$F_{v2,v1,1-\alpha/2} = \frac{1}{F_{v1,v2,\alpha}} \tag{4.35}$$

Τίθεται η μηδενική υποθεση H_o: Τα δύο δείγματα προέρχονται από τον ίδιο πληθυσμό, δηλαδή, $\sigma^2{}_1=\sigma^2{}_2=\sigma^2$, οποτε $\sigma^2{}_1/\sigma^2{}_2=1$ και η εναλλακτική υποθεση H_1: $\sigma^2{}_1\neq\sigma^2{}_2$, **οποτε** $\sigma^2{}_1/\sigma^2{}_2\neq1$.

Επειδή, οπως αναφέρθηκε, η τυχαία μεταβλητη ακολουθεί την F-κατανομή, θα πρέπει σε σ.σ α και v_1, v_2, βαθμούς ελευθερίας να ισχύει :

$$F_{v1,v2,1-\alpha/2} < F < F_{v1,v2,\alpha/2} \tag{4.36}$$

Ή

$$F_{v1,v2,1-\alpha/2} \leq \frac{S^2{}_1*\sigma^2{}_2}{S^2{}_2*\sigma^2{}_1} \leq F_{v1,v2,\alpha/2} \tag{4.37}$$

Επειδή για την F-κατανομή ισχύει:

$$F_{v1,v2,1-\alpha/2}=\frac{1}{F_{v2,v1,\alpha/2}} \tag{4.38}$$

Η (4.37) γίνεται :

$$\frac{1}{F_{v2,v1,\alpha/2}} \leq \frac{S^2{}_1*\sigma^2{}_2}{S^2{}_2*\sigma^2{}_1} \leq F_{v1,v2,\alpha/2} \tag{4.39}$$

Επομένως, σε διπλευρο test, αν ισχύει η H_o και είναι

$\sigma^2_1/\sigma^2_2=1$ θα πρέπει:

$$\frac{1}{F_{v2,v1,\alpha/2}} \leq \frac{S^2_1}{S^2_2} \leq F_{v1,v2,\alpha/2} \qquad (4.40)$$

Αν, συνεπώς, ο λογος S^2_1/S^2_2 επαληθεύει τη σχέση (4.40), τότε γίνεται αποδεκτή η υπόθεση H_o διαφορετικά, απορρίπτεται η H_o και γίνεται αποδεκτή η εναλλακτική της, H_1, όπου $\sigma_1 \neq \sigma_2$.

Παράδειγμα 4.6.4.

Οι μηνιαίες μέσες τιμές της μέγιστης θερμοκρασίας ενος σταθμού Σ που λειτούργησε για $n_1=16$ χρόνια στην θέση Σ_1 και για $n_2=11$ χρόνια στην θέση Σ_2 όπου μεταφέρθηκε από την θέση Σ_1 και εξακολουθεί να λειτουργεί, έδωσαν $S^2_1=2.9$ και $S^2_2=1$.

Θα εξετασθεί αν η χρονοσειρά των μέσων αυτών τιμών μπορεί να θεωρηθεί ομοιογενής σε σ.σ.: $\alpha=0.1$.

Σε δίπλευρο test η H_o και η εναλλακτική της υπόθεση H_1 είναι:

H_o: $\sigma_1=\sigma_2=\sigma$, οπότε $\sigma^2_1/\sigma^2_2=1$

H_1: $\sigma_1 \neq \sigma_2$

Είναι: $\alpha/2=0.05$, $v_1=16-1=15$, $v_2=11-1=10$, οπότε:

$F_{15,10,0.05}=2.85$, $F_{10,15,0.05}=2.54$ και $F=S^2_1/S^2_2=2.9$

Διαπιστώνεται ότι η τιμή F δεν περιλαμβάνεται στο διάστημα D=[2.54 , 2.85] και, επομένως, συμπεραίνεται ότι η σειρά δεν είναι ομοιογενής.

Σε μονόπλευρο test, αν ληφθεί σαν H_o η : $\sigma_1=\sigma_2$, η εναλλακτική της, H_1 θα είναι: $\sigma_1>\sigma_2$, (σ'αυτό μας οδηγεί η μεγαλύτερη της μονάδας τιμή του F, F=2.9) και σ.σ: $\alpha=0.05$, τότε έχουμε και πάλι $F_{15,10,0.05}=2.85$ οπότε $S^2_1/S^2_2>F_{15,10,0.05}$.

Άρα και με το μονόπλευρο test καταλήγουμε στο ίδιο συμπέρασμα (όπως άλλωστε αναμενόταν).

5.ΟΜΟΙΟΓΕΝΟΠΟΙΗΣΗ

5.1.Γενικά.

Η ανομοιογένεια σε σειρές κλιματολογικών δεδομένων οφείλεται συνήθως σε κάποιο διαταρακτικό παράγοντα, όπως έχουμε αναφέρει λεπτομερώς στα προηγούμενα. Αν και έχει επιχειρηθεί πολλές φορές στο παρελθόν να ομοιογενοποιηθούν σειρές δεδομένων με τέτοιες ανομοιογένειες, εν τούτοις πρέπει να γίνει σαφές ότι **δεν είναι δυνατόν** να ομοιογενοποιηθεί μια σειρά διακεκριμένων τιμών και να προκύψει χρονοσειρά με τις ίδιες ιδιότητες. Με άλλα λόγια, εάν τα δεδομένα μιας συγκεκριμένης περιόδου από ένα σταθμό είναι ακατάλληλα, είναι αδύνατο να αναπαραχθούν οι απλές τιμές της σειράς αυτής της περιόδου. Αυτό οφείλεται στο ότι **κάθε προσαρμογή** μιας χρονοσειράς καταστρέφει τη μεταβλητότητά της και συνεπώς αλλάζει την κλίμακα ή την διασπορά της κατανομής συχνότητας. Αντίθετα, είναι δυνατόν να προσαρμοστούν ή να ρυθμιστούν (adjust) συγκεκριμένες στατιστικές παράμετροι της σειράς, έτσι ώστε οι "νέες" προσαρμοσμένες τιμές να δίνουν το ίδιο αποτέλεσμα που θα έδιναν οι "παλιές" τιμές, δηλαδή, οι υπολογισμένες από δείγματα της αρχικής χρονοσειράς.

5.2. Ομοιογενοποίηση των δεδομένων με γραφική μέθοδο.

Η γραφική μέθοδος μπορεί να εφαρμοσθεί όταν έχει χρησιμοποιηθεί η **"ανάλυση διπλής μάζας"** **(παρ.4.2.2)** για τον έλεγχο της ομοιογένειας μιας χρονοσειράς.

Η διαδικασία έχει ως εξής:

Όταν η ανάλυση διπλής μάζας αποκαλύπτει αλλαγή στην κλίση του γραφήματος από κάποιο σημείο και μετά, και ύστερα από προσεκτική μελέτη του ιστορικού του σταθμού, προκύπτει ότι υπάρχει λόγος ανομοιογένειας, μπορεί να ομοιογενοποηθεί η σειρά των δεδομένων και να γίνουν συμβατά, τα προ της αλλαγής της κλίσης δεδομένα με εκείνα τα μετά την αλλαγή, χρησιμοποιώντας τον λόγο των δύο κλίσεων των καμπύλων διπλής μάζας, με βάση τον τύπο:

$$M_\alpha=\frac{\rho_\alpha}{\rho_o}*M_o \qquad (5.1)$$

οπου:

M_o είναι η αρχική τιμή της παραμέτρου (πριν απο την ομοιογενοποίηση), **M_α** είναι η αντίστοιχη τροποποιημένη τιμή, **ρ_o** είναι η κλίση της καμπύλης διπλής μάζας που αντιστοιχεί σε εκείνες τις τιμές της παραμέτρου που ανήκουν στην περίοδο που θα ομοιογενοποιηθεί και **ρ_α** η κλίση της καμπύλης διπλής μάζας που αντιστοιχεί στις τιμές προς τις οποίες θα προσαρμοσθούν οι υπόλοιπες τιμές (οι προς ομοιογενοποίηση).

Οι M_o και M_α είναι στατιστικές παράμετροι και όχι απλές τιμές.

Παράδειγμα 5.2.

'Εστω ότι στο **παράδειγμα (4.2.2)** που αναφέρεται στη μέθοδο **ανάλυσης διπλής μάζας (παραγ.4.2.2)** χρειάζεται να προσαρμοσθούν οι τιμές των αθροιστικών ετήσιων υψών υετού των τεσσάρων (4) πρώτων χρόνων στις τιμές των τελευταίων ετών.

'Οπως διαπιστώνεται:

$\rho_o=0.45$

$\rho_\alpha=1.45$

οπου ρ_o, η κλίση της αρχικής ευθείας και ρ_α η κλιση της δεύτερης ευθείας **(σχήμα 4.2.2)**.

Επομένως : $\rho_o/\rho_\alpha=3.22$

Τα τροποποιημένα αθροιστικά ετήσια ύψη υετού των τεσσάρων πρώτων χρόνων υπολογίζονται με βάση τον **Πίνακα (4.2) των υψών**, από όπου λαμβάνονται τα M_o και την εξίσωση (5.1) ως εξής:

1ος χρονος: $M_1=3.22*0.90=2.90$

2ος χρονος: $M_2=3.22*0.40=1.29$

3ος χρονος: $M_3=3.22*0.90=2.90$

4ος χρονος: $M_4=3.22*0.90=2.90$

5.3. Ομογενοποίηση στατιστικών παραμέτρων με τις μεθόδους της διαφοράς ή του αναλογίας.

Η πλέον κοινή εφαρμογή τέτοιων προσαρμογών είναι αυτή που χρησιμοποιείται για τους μέσους όρους μιας σειράς μετεωρολογικών δεδομένων με σκοπό την εύρεση των κανονικών τιμών της (normals) της μετεωρολογικής παραμέτρου. Τέτοιες προσαρμογές συνίσταται να γίνονται, εφόσον είναι δυνατόν, μόνο όταν είναι διαπιστωμένη η ανομοιογένεια.

Μπορεί να δειχθεί με θεωρητική ανάλυση ότι οι κλασσικές μέθοδοι **της διαφοράς και της αναλογίας** είναι πολύ καλές για την προσαρμογή των μέσων τιμών των θερμοκρασιών και του υετού, καθώς και των άλλων μετεωρολογικών παραμέτρων. Τέτοιες προσαρμογές γίνονται συχνά για να συμπληρωθούν ελλείπουσες τιμές-παρατηρήσεις ή για να αρθούν τυχόν ανομοιογένειες. Η μέθοδος της διαφοράς εφαρμόζεται συνήθως σε σειρές μέσων τιμών μετεωρολογικών παραμέτρων (θερμοκρασίας κ.λ.π.) ενώ η μέθοδος της αναλογίας σε σειρές κυρίως αθροιστικών τιμών των μετεωρολογικών παραμέτρων αλλά και σε μέσες τιμές (μέσα ετήσια ύψη υετού κ.λ.π.).

Οι μέθοδοι απαιτούν την χρήση ενός τουλάχιστον σταθμού με ομοιογενή σειρά τιμών σε όλη την αντίστοιχη χρονική περίοδο στην οποία μελετάται ο υπό εξέταση σταθμός. Αυτός ο βοηθητικός σταθμός πρέπει να είναι όσο το δυνατόν πλησιέστερα στον υπό μελέτη σταθμό και να διαθέτει αποδεδειγμένα ομοιογενή χρονοσειρά παρατηρήσεων, καθ'όσον η αποτελεσματικότητα της ομοιογενοποίησης εξαρτάται από τη συσχέτιση των τιμών των δύο σταθμών. Συνήθως ένας σταθμός σε απόσταση **μικρότερη από τα 50 μίλια (100 Km)** που βρίσκεται στην ίδια κλιματική περιοχή και έχει παρόμοια χαρακτηριστικά ως προς την έκθεση στα καιρικά συστήματα με τον υπό προσαρμογή σταθμό, θα εξυπηρετήσει γι'αυτόν το σκοπό.

Αν υπάρχουν περισσότεροι από ένας βοηθητικοί σταθμοί, μπορεί να χρησιμοποιηθεί μια βοηθητική σειρά δεδομένων που θα αποτελείται από τους μέσους όρους των αντίστοιχων τιμών των σταθμών. Συνήθως, όμως, αυτή η διαδικασία δεν βελτιώνει σημαντικά τις ομοιογενοποιημένες τιμές.

Αν ο πλησιέστερος βοηθητικός σταθμός για την ομοιογενοποίηση δεν διαθέτει μια πλήρη χρονοσειρά, αυτός χρησιμοποιείται μόνο για την περίοδο που καλύπτει το αρχείο του και για την υπόλοιπη περίοδο χρησιμοποιείται ο "επόμενος" πλησιέστερος σταθμός, κ.ο.κ. μέχρι να καλυφθεί ολόκληρη η περίοδος των παρατηρήσεων του υπό ομοιογενοποίηση σταθμού.

5.3.1. Ομοιογενοποίηση με τη μέθοδο της διαφοράς. Στη μέθοδο αυτή ακολουθείται η παρακάτω διαδικασία:

Ας θεωρήσουμε σταθμό $Σ_1$ που διαθέτει αρχείο **κ** ετών μιας παραμέτρου X, ανομοιογενές στο σύνολό του αλλά ομοιογενές κατά τμήματα και έστω $\kappa_1, \kappa_2, ..., \kappa_\lambda$ με $\kappa_1+\kappa_2+...+\kappa_\lambda=\kappa$ **οι λ ομοιογενείς υποπερίοδοι της σειράς, διαρκείας κ_i χρόνων η κάθε μία, $i \in I\lambda$.**

Για την ομοιογενοποίηση της σειράς, θεωρείται κατάλληλος σταθμός Σ_2 που διαθέτει ομοιογενή σειρά σε όλη την περίοδο των κ αντίστοιχων ετών.

Με τη μέθοδο της διαφοράς βρίσκεται η τιμή του διορθωτικού όρου **α** από την εξίσωση:

$$\alpha=E(Y_i)-E(W_i) \qquad (5.2)$$

όπου **$E(Y_i)$ και $E(W_i)$** οι μέσες τιμές της ίδιας παραμέτρου X στους σταθμούς Σ_1 και Σ_2 αντίστοιχα, σε μία κοινή υποπερίοδο (ομοιογενή) **i, $i \in I_\lambda$** (κυρίως στην τελευταία, δηλαδή στην λ) ως προς την οποία είναι επιθυμητό να προσαρμοσθεί ή να ομοιογενοποιηθεί όλη η χρονοσειρά.

Ακολούθως, η μέση τιμή της παραμέτρου σε κάθε μία από τις υποπεριόδους του σταθμού Σ_1 προσαρμόζεται με βάση τον διορθωτικό όρο **α**, από τον τύπο:

$$E'(Y_i)=E(W_i)+\alpha \qquad (5.3)$$

όπου **$E'(Y_i)$** είναι η διορθωμένη μέση τιμή της ομοιογενούς υποπεριόδου $i \neq \lambda$ και $E(W_i)$ η μέση τιμή της παραμέτρου για την ίδια ακριβώς περίοδο i στον βοηθητικό σταθμό **Σ_2** που διαθέτει ομοιογενή σειρά στο σύνολό της.

Αφού η διαδικασία αυτή γίνει για όλες τις υποπεριόδους και υπολογισθούν όλες οι διορθωμένες μέσες τιμές $E'(Y_i)$, ακολούθως η μέση τιμή της παραμέτρου βασισμένη σε όλη την περίοδο των **κ** ετών υπολογίζεται με βάση τον τύπο:

$$E'(Y)=\frac{{}_1\Sigma^\lambda \kappa_i * E'(Y_i)}{{}_1\Sigma^\lambda \kappa_i} = \frac{{}_1\Sigma^\lambda \kappa_i * E'(Y_i)}{\kappa} \qquad (5.4)$$

Παράδειγμα 5.3.1.

Ας θεωρήσουμε ένα δείγμα ετήσιων μέσων θερμοκρασιών του σταθμού της Γενεύης το οποίο είναι σκόπιμα ανομοιογενές

Thom [Βιβλ.9] γιατί απο τις τιμές των 12 πρώτων χρονων έχει αφαιρεθεί 1°C, απο τις τιμές των 8 επομενων χρονων έχουν αφαιρεθεί 0.5°C ενώ οι τιμές των τελευταίων 10 χρονων έχουν μείνει αμετάβλητες. Μια τέτοια σειρά θα μπορούσε να προκύψει σ'ένα σταθμό Σ που θα εργαζόταν για 12 χρόνια στην "ψυχρή" θέση $Σ_1$, για 8 χρόνια στην θέση $Σ_2$ και για τα υπόλοιπα χρόνια στην τελική του θέση Σ.

Χρειάζεται να διορθωθεί **η μέση τιμή της θερμοκρασίας των 30 ετών του σταθμού** με βάση τις τιμές των τελευταίων 10 χρόνων.

Αυτο είναι ένα τυπικο πρόβλημα ομοιογενοποίησης και και θα ακολουθηθεί η παρακάτω διαδικασία:

Η χρονοσειρά των ετήσιων μέσων τιμών της θερμοκρασίας είναι συνολικά ανομοιογενής αλλά κατά τμήματα ομοιογενής στις τρεις χρονικές περιόδους διαρκείας,**$κ_1$=12, $κ_2$=8 και $κ_3$=10.**

Χρησιμοποιείται σαν κατάλληλος βοηθητικος σταθμός, ο σταθμός της Λωζάνης που διαθέτει ομοιογενή σειρά για όλη την περίοδο 1927-1956.

Αρχικά υπολογίζονται οι μέσοι οροι των ετήσιων μέσων θερμοκρασιών, $E(Y_i)$ και $E(W_i)$ για τις τρεις υποπεριόδους **$κ_i$, i∈I3** και στους δύο σταθμούς και οι τιμές τους φαίνονται στο παρακάτω **Πίνακα (5.1).**

Πίνακας 5.1. Ομοιογενοποίηση μέσης θερμοκρασίας για τον σταθμό της Γενεύης.

Υπο-περίοδοι	**k_i**	**Λωζάνη Μέσοι οροι $E(w_i)$**	**Γενεύη Μη διορθωμένοι μέσοι οροι $E(Y_i)$**	**Γενεύη Διορθωμένοι μέσοι οροι $E'(Y_i)$**
1927-1938	(12)	17.9	17.9	19.3*
1939-1946	(8)	18.4	19.0	19.8*
1947-1956	(10)	18.2	19.6	19.6

Προσαρμοσμένος μέσος των 30 ετών 19.5

Εφαρμόζοντας τον τύπο (5.1), για την τρίτη περίοδο των 10 ετών υπολογίζεται ο όρος **α**:

$\alpha=E(Y_3)-E(W_3)=19.6-18.2=1.4$

Εισάγοντας αυτή την τιμή στην (5.2) και αντικαθιστώντας διαδοχικά τις τιμές 17.9 και 18.4 του βοηθητικού σταθμού για τις αντίστοιχες υποπεριόδους, υπολογίζονται:

$E'(Y_1)=17.9+1.4=19.3$ και $E'(Y_2)=18.4+1.4=19.8$

Ακολούθως, χρησιμοποιώντας τον τύπο (5.3) και αντικαθιστώντας τις αντίστοιχες τιμές κ_i και $E'(Y_i), i\varepsilon I_3$ θα είναι :

$$E(Y)=\frac{12*19.3+8*19.8+10*19.6}{12+8+10}=19.5$$

Αυτή η τιμή είναι υπολογισμένη στην ομοιογενοποιημένη σειρά των 30 χρόνων. Η όλη διαδικασία αποτελεί την καλύτερη προσέγγιση του υποθετικού μέσου για την ανομοιογενή περίοδο 1927-1956, βασισμένη στην ομοιογενή υποπερίοδο 1947-1956 του σταθμού της Γενεύης και στον βοηθητικό σταθμό της Λωζάνης.

5.3.2. Ομοιογενοποίηση με την μέθοδο της αναλογίας. Στη μέθοδο αυτή που είναι ανάλογη με τη μέθοδο της διαφοράς αλλά που χρησιμοποιείται ο λόγος, ακολουθείται η παρακάτω διαδικασία:

Ας θεωρήσουμε σταθμό Σ_1 που διαθέτει αρχείο **κ** ετών μιας παραμέτρου Χ, ανομοιογενές στο σύνολο του αλλά ομοιογενές κατά τμήματα και έστω $\kappa_1,\kappa_2,...,\kappa_\lambda$ **με** $\kappa_1+\kappa_2+...+\kappa_\lambda=\kappa$ οι λ ομοιογενείς υποπερίοδοι της σειράς, διαρκείας κ_i χρόνων η κάθε μία, $i\varepsilon I\lambda$.

Για την ομοιογενοποίηση της σειράς, θεωρείται κατάλληλος σταθμός Σ_2 που διαθέτει ομοιογενή σειρά σε όλη την περίοδο των κ αντίστοιχων ετών.

Στη μέθοδο της αναλογίας βρίσκεται η τιμή του διορθωτικού όρου **β** από την εξίσωση:

$$\beta=(\Sigma Y)_i/(\Sigma W)_i \qquad (5.5)$$

όπου $(\Sigma Y)_i$ **και** $(\Sigma W)_i$ τα αθροίσματα των τιμών της παραμέτρου Χ στους σταθμούς Σ_1 και Σ_2 αντίστοιχα, σε μία υποπερίοδο (ομοιογενή) **i**, $i\varepsilon I_\lambda$, (κυρίως στην τελευταία, δηλαδή στην λ), ως προς την οποία είναι επιθυμητό να προσαρμοσθεί η να ομοιογενοποιηθεί όλη η χρονοσειρά.

Ακολούθως, το άθροισμα των τιμών της παραμέτρου σε κάθε μία από τις υποπεριόδους του σταθμού Σ_1 προσαρμόζεται με βάση τον διορθωτικό όρο **β** από τον τύπο:

$$(\Sigma Y)'_i=\beta*(\Sigma W)_i \tag{5.6}$$

όπου $(\Sigma Y)'_i$ είναι η διορθωμένη τιμή του αθροίσματος των τιμών της ομοιογενούς υποπεριόδου $i\neq\lambda$ **και** $(\Sigma W)_i$ η τιμή του αθροίσματος των τιμών της παραμέτρου για την ίδια ακριβώς περίοδο i στον βοηθητικό σταθμό Σ_2 που διαθέτει ομοιογενή σειρά στο σύνολό της.

Αφού η διαδικασία αυτή γίνει για όλες τις υποπεριόδους και υπολογισθούν όλα τα διορθωμένα αθροίσματα $(\Sigma Y)'_i$,ακολούθως, η μέση τιμή της παραμέτρου βασισμένη σε όλη την περίοδο των **κ** ετών υπολογίζεται με βάση τον τύπο:

$$E'(Y)=\frac{{}_{i}\Sigma^{\lambda}\kappa_i*(\Sigma Y)'_i}{{}_{i}\Sigma^{\lambda}\kappa_i}=\frac{{}_{i}\Sigma^{\lambda}\kappa_i*(\Sigma Y)'_i}{\kappa} \tag{5.7}$$

Παράδειγμα 5.3.2.

Ας θεωρήσουμε ένα δείγμα μηνιαίων υψών υετού του σταθμού της Γενεύης για τον μήνα Αύγουστο, το οποίο είναι σκόπιμα ανομοιογενές, **Thom [Βιβλ.9]**, γιατί οι τιμές των 12 πρώτων χρόνων έχει πολλαπλασιασθεί επί 1.20, οι τιμές των 8 επόμενων χρόνων έχουν πολλαπλασιασθεί επί 0.90 ενώ οι τιμές των τελευταίων 10 χρόνων έχουν μείνει αμετάβλητες, **Πίνακας 5.2.**

Μια τέτοια σειρά θα μπορούσε να προκύψει σ'ένα σταθμό Σ που θα εργαζόταν για 12 χρόνια στην θέση Σ_1, για 8 χρόνια στην θέση Σ_2 και για τα υπόλοιπα χρόνια στην τελική του θέση Σ.

Χρειάζεται να υπολογισθεί το μέσο μηνιαίο ύψος υετού Αυγούστου των 30 ετών του σταθμού με βάση τις τιμές των τελευταίων 10 χρόνων.

Αυτό είναι ένα τυπικό πρόβλημα ομοιογενοποίησης και και θα ακολουθηθεί η παρακάτω διαδικασία:

Η χρονοσειρά των μηνιαίων υψών υετού είναι συνολικά ανομοιογενής αλλά κατά τμήματα ομοιογενής στις τρεις χρονικές περιοδους διαρκειας **$κ_1$=12, $κ_2$=8 και $κ_3$=10.**

Χρησιμοποιειται σαν κατάλληλος βοηθητικος σταθμος, ο σταθμός της Λωζάνης που διαθέτει ομοιογενή σειρά για όλη την περιοδο 1927-1956.

Αρχικά υπολογίζονται τα αθροίσματα των μηνιαίων υψών υετού, $(ΣY)_i$ και $(ΣW)_i$ για τις τρεις υποπεριόδους **$κ_i$, iεI3** και στους δύο σταθμούς που φαίνονται στον παρακάτω **Πίνακα 5.3.**

Εφαρμόζοντας τον τύπο (5.5) για την τρίτη περίοδο των 10 ετών υπολογίζεται ο όρος **β:**

$β=(ΣY)_3/(ΣW)_3=1024/1267=0.8082$

Εισάγοντας αυτή την τιμή στην σχέση (5.6) και αντικαθιστώντας διαδοχικά τις τιμές 1602 και 753 του βοηθητικού σταθμού για τις αντίστοιχες υποπεριόδους, υπολογίζονται

$(ΣY)'_1=1602*.8082=1295$ και $(ΣY)'_2=753*.8082=609$.

Ακολούθως, χρησιμοποιώντας τον τύπο (5.7) και αντικαθιστώντας τις αντίστοιχες τιμές **$κ_i$ και $(ΣY)'_i$,iεI$_3$** θα είναι :

$$E(Y)=\frac{12*1295+8*609+10*102419.6}{12+8+10} = 97.6$$

Αυτή η τιμή είναι υπολογισμένη στην ομογενοποιημένη σειρά των 30 χρόνων. Η όλη διαδικασία αποτελεί την καλύτερη προσέγγιση του υποθετικού μέσου μηναίου ύψους υετού για την ανομοιογενή περίοδο 1927-1956, βασισμένη στην ομοιογενή υποπερίοδο 1947-1956 του σταθμού της Γενεύης και στον βοηθητικό σταθμό της Λωζάνης.

Πίνακας 5.2. Ανομοιογενής σειρά ολικών υψών υετού Γενεύης τον Αύγουστο

1927	300	1937	94	1947	54
1928	176	1938	95	1948	72
1929	100	1939	77	1949	49
1930	130	1940	16	1950	110
1931	205	1941	95	1951	100
1932	74	1942	43	1952	125
1933	89	1943	37	1953	57
1934	143	1944	40	1954	206
1935	188	1945	120	1955	107
1936	28	1946	142	1956	144

Πίνακας 5.3. Ομοιογενοποίηση μηνιαίων υψών υετού Αυγούστου στον σταθμό της Γενεύης.

Υπο-περίοδοι	k_i	Λωζάνη Αθροισ. $(\Sigma W)_i$	Γενεύη Μη διορθωμένα αθροισ. $(\Sigma Y)_i$	Γενεύη Διορθωμένα αθροισ. $(\Sigma Y)'_i$
1927-1938	(12)	1602	1613	1295
1939-1946	(8)	753	570	609
1947-1956	(10)	1267	1024	1024

Προσαρμοσμένος μέσος των 30 ετών 97.6

5.4.Εύρεση του κατάλληλου γειτονικού σταθμού με την μέθοδο του συντελεστή συσχέτισης του Spearman.

Για την ομοιογενοποίηση μιας ανομοιογενούς στο σύνολό της, αλλά κατά τμήματα ομοιογενούς χρονοσειράς παρατηρήσεων μιας παραμέτρου X, απαιτείται μια ομοιογενής χρονοσειρά ενός κατάλληλου γειτονικού σταθμού. Η εύρεση του καταλληλότερου σταθμού βασίζεται στη συσχέτιση των τιμών της παραμέτρου X με τις αντίστοιχες τιμές της ίδιας παραμέτρου σε γειτονικούς σταθμούς για κάθε ομοιογενή υποπερίοδο του σταθμού από τον οποίο προέρχεται η προς ομοιογενοποίηση χρονοσειρά. Σαν καταλληλότερος σταθμός εκλέγεται εκείνος για τον οποίο ο συντελεστής συσχέτισης ρ στη περίοδο ως προς την οποία είναι επιθυμητό να γίνει η ομοιογενοποίηση, είναι ο μεγαλύτερος.

Ο "κλασσικός" συντελεστής συσχέτισης δύο μεταβλητών X,Y δίνεται από την σχέση:

$$\rho=\rho_{xy}=\frac{COV(X,Y)}{[\sigma^2{}_x*\sigma^2{}_y]^{0.5}}=\frac{COV(X,Y)}{\sigma_x*\sigma_y} \qquad (5.8)$$

ή την:

$$r=r_{xy}=\frac{S_{xy}}{S'_x*S'_y} \qquad (5.9)$$

όπου:

$$COV(X,Y)=E\{[X-E(X)]*[Y-E(Y)]\} \qquad (5.10)$$

η συνδιακύμανση (Covariance) των X και Y.

Η (5.8) χρησιμοποιείται όταν είναι γνωστά όλα τα δυνατά ζεύγη των τιμών των παραμέτρων X και Y και η (5.9) όταν ο υπολογισμός του συντελεστή γραμμικής συσχέτισης υπολογίζεται από δείγμα μεγέθους **n** ζευγών **(x_i,y_i) $i\in I_n$** της διδιάστατης μεταβλητής (X,Y). Για να χρησιμοποιηθούν οι ρ και r, που δίνονται από τους (5.8) και (5.9), για τον έλεγχο της ανεξαρτησίας των παραμέτρων X και Y απαιτείται όπως αυτές ακολουθούν την κανονική κατανομή. Επειδή η υπόθεση αυτή δεν ευσταθεί πάντα ή επειδή δεν είναι επιθυμητό να γίνεται εκ των προτέρων μια τέτοια υπόθεση, εισάγουμε τον **συντελεστή συσχέτισης Spearman**, με βάση τον οποίο δεν συσχετίζονται οι τιμές x_i και y_i αλλά οι τάξεις τους μέσα στο πληθυσμό ή το δείγμα.

Έστω ένα σύνολο η τιμών μιας παραμέτρου Α (δείγμα μεγέθους η), οι $x_1, x_2, ..., x_n$. Διατάσσουμε τις τιμές αυτές κατά τάξη αύξοντος ή φθίνοντος μεγέθους και έστω $y_1, y_2, ..., y_n$ οι διατεταγμένες τιμές. Ο αριθμός i ονομάζεται **τάξη** της x_i. **Αν $x_i=y_j$, τότε ονομάζεται διατεταγμένη τάξη της τιμής x_i ο αριθμός j.**

Αν συμβεί: $y_k=y_{k+1}=...=y_{k+m}$ με $m>0$, δηλαδή αν κάποια από τις τιμές $x_i, i\in I_n$ του συνόλου συμβεί να εμφανίζεται περισσότερο από μια φορές και είναι $y_k=x_{i1}$, $y_{k+1}=x_{i2}, ..., y_{k+m}=x_{im}$, τότε σαν διατεταγμένη τάξη των x_{i1}, $x_{i2}, ..., x_{im}$ ορίζεται ο j', μέσος όρος των αριθμών $k, k+1, ..., k+m$, ίσος με:

$j'={}_0\Sigma^m(k+i)/(m+1)=[(k+m)*(k+m+1)-(k-1)*k]/[2*(m+1)]$ (5.11)

Μετά απ'αυτά **ο συντελεστής συσχέτισης Spearman** ορίζεται ως εξής:

Έστω (x_i, y_i), $i\in I_n$ ένα δείγμα μεγέθους **η** αντίστοιχων τιμών των παραμέτρων X και Y. Διατάσσονται οι x_i και οι y_i κατά τάξη αύξοντος (ή φθίνοντος) μεγέθους και έστω j_i και k_i οι **διατεταγμένες τάξεις των x_i και y_i.**

Στο ζεύγος (x_i, y_i) αντιστοιχούμε το τετράγωνο της διαφοράς των διαταταγμένων τάξεών τους:

$d_i^2=(j_i-k_i)^2$ (5.12)

και βρίσκουμε το άθροισμα:

$D={}_1\Sigma^n d^2_i$ (5.13)

Ο **συντελεστής του Spearman** των X και Y δίνεται από την:

$$r_s=r_{sxy}=1-\frac{6*D}{n^3-n} \qquad (5.14)$$

-Αν **$r_s=0$** οι X και Y θεωρούνται ανεξάρτητες-ασυσχέτιστες.

-Αν **$r_s=1$** τότε $d_i=0$ για κάθε $i\in I_n$ και οι X και Y έχουν **"τέλεια" θετική συσχέτιση.**

-Αν **$r_s=-1$** τότε για κάθε $i\in I_n$ ισχύει η σχέση:

$j_i+k_i=n+1$ (5.15)

δηλαδή στην μεγαλύτερη τιμή του j_i αντιστοιχεί η μικρότερη του k_i, στην δεύτερη κατά σειρά μεγαλύτερη τιμή του j_i αντιστοιχεί η δεύτερη μικρότερη τιμή του k_i κ.ο.κ και τότε υπάρχει τέλεια **αρνητική συσχέτιση μεταξύ των X και Y.** Τέλος, αν $-1<r_d<1$ τότε η συσχέτιση των X και Y εξαρτάται από την τιμή του **η** και του r_d και είναι **τόσο καλύτερη όσο η $|r_d|$ μεγαλώνει.**

Οι υποθέσεις H_o και H_1 ως προς τις οποίες ελέγχεται η εξάρτηση ή μη των X και Y είναι:

H_o: $r_d=0$ (πλήρης ανεξαρτησία)

H_1: $r_d\neq0$ (εξάρτηση, εξαρτώμενη από τα **r_d και η).**

Σαν H_1 μπορούμε να πάρουμε και την:

H'_1: $r_d>0$ (θετική εξάρτηση σε μονόπλευρο test).

Για **η<6** δεν υπάρχει test από το οποίο να συμπεραίνεται η εξάρτηση ή μη των X και Y.

Για **5<η<11** υπάρχουν πίνακες που δίνουν τις κρίσιμες τιμές $S_{n,\alpha}$ σε συνάρτηση με το η και την (σ.σ): α. Εδώ, ο σχετικός πίνακας **(Πίνακας 5.5) περιέχει τις τιμές του $S_{n,\alpha}$ για τις τιμές 0.1, 0.05 και 0.01 του α.**

Στην περίπτωση αυτή, η H_o γίνεται αποδεκτή σε στάθμη σημαντικότητας (σ.σ): α, όταν $|r_d|<S_{n,\alpha}$ (μονόπλευρο test) ή όταν $|r_d|<S_{n,\alpha/2}$ (δίπλευρο test).

Για **η>10** η τυχαία μεταβλητή t_d όπου:

$$t_d = \frac{r_d*(\eta-2)^{0.5}}{(1-r_d)^{0.5}} \qquad (5.16)$$

ακολουθεί την t-κατανομή με v=η-2 βαθμούς ελευθερίας, οπότε:

αν $|t_d|<t_{n,\alpha/2}$ σε περίπτωση δίπλευρου test ή

αν $|t_d|<t_{n,\alpha}$ στην περίπτωση μονόπλευρου test

η υπόθεση H_o γίνεται δεκτή ενώ αν επαληθεύονται οι αντίστροφες προς τις προηγούμενες ανισότητες τότε γίνεται δεκτή η **H_1 ή H'_1.**

Παράδειγμα 5.4

Οι ετήσιες μέσες θερμοκρασίες των εννέα (η=9) τελευταίων χρόνων του αρχείου του σταθμού Σ_1, του πλησιέστερου σ'αυτόν σταθμού Σ_2 και του αμέσως επόμενου γειτονικού στον Σ_1 σταθμού, Σ_3, δίνονται στον **Πίνακα 5.4.**

Από το ιστορικό του σταθμού ή από την εφαρμογή μιας από της μεθόδους ελέγχου της ομοιογένειας, διαπιστώθηκε ότι ο σταθμός Σ_1 είναι ανομοιογενής στο σύνολο του αλλά κατά τμήματα ομοιογενής και μάλιστα το τελευταίο ομοιογενές τμήμα της περιόδου είναι αυτό των 9 τελευταίων χρόνων. Ανάλογος έλεγχος στα αρχεία των Σ_2 και Σ_3 έδειξε ότι αυτοί είναι ομοιογενείς.

Σκοπεύουμε να ελέγξουμε αν ο $Σ_2$ είναι ο καταλληλότερος σταθμός για την ομοιογενοποίηση των μέσων τιμών της θερμοκρασίας του $Σ_1$ για τα προηγούμενα ομοιογενή τμήματα της χρονοσειράς των τιμών του σταθμού αυτού, ώστε η προκύπτουσα χρονοσειρά των μέσων ετησίων που θα προκύψει να είναι ομοιογενής ή αν πρέπει να εκλεγεί ο $Σ_3$, έστω και αν είναι λίγο περισσότερο απομακρυσμένος.

Από τον Πίνακα 5.4 και την σχέση (5.4) προκύπτει ότι:

$r_{d12}=1-6*34/(9^3-9)=0.717$, και

$r_{d13}=1-6*8.5/(9^3-9)=0.929$

Για n=9 και για (σ.σ): α=0.05 ο **Πίνακας 5.5** δίνει $S_{9,0.05}=0.683$ δηλαδή είναι $r_{d12}>S_{n,\alpha}$ και $r_{d13}>S_{9,0.05}$. Αλλά είναι $r_{d13}>r_{d12}$ και αποτέλεσμα του γεγονότος αυτού είναι να προτιμηθεί ο $Σ_3$ και όχι ο $Σ_2$.

Εδώ, όπως και σ'όλες τις περιπτώσεις εφαρμογής του συντελεστή συσχέτισης του Spearman για τις μετεωρολογικές παραμέτρους, ενδιαφερόμαστε για τη θετική συσχέτιση των παραμέτρων του προς ομοιογενοποίηση των τιμών του σταθμού και του σταθμού που τελικά θα επιλεγεί σαν μάρτυρας και, επομένως, εφαρμόζεται το μονόπλευρο test, τόσο για μικρές τιμές του **n** (όπως στο παράδειγμα που δίνεται αμέσως παρακάτω) όπως και για μεγάλες τιμές του n όπου εφαρμόζεται το t-test.

Πίνακας 5.4.

Τάξη της παρατήρησης (στήλη 1)

`Ετος (στήλη 2)

Ετήσια μέση θερμοκρασία του $Σ_1$ (στήλη 3)

Ετήσια μέση θερμοκρασία του $Σ_2$ (στήλη 4)

Ετήσια μέση θερμοκρασία του $Σ_3$ (στήλη 5)

Διατεταγμένη τάξη των τιμών

j_i του $Σ_1$ (στήλη 6)

k_i του $Σ_2$ (στήλη 7)

l_i του $Σ_3$ (στήλη 8)

Διαφορές $d_{i12}=j_i-k_i$ (στήλη 9)

Τετράγωνα των d_{i12} δηλ. d_{i12}^2 (στήλη 10)

Διαφορές $d_{i13}=j_i-l_i$ (στήλη 11)

Τετράγωνα των d_{i13} δηλ. d_{i13}^2 (στήλη 12)

αριθμός στήλης											
1	2	3	4	5	6	7	8	9	10	11	12
1	1982	14.9	14.5	13.9	9	7	8	2	4	1	1
2	1983	13.0	14.8	13.1	4	5	4	-1	1	0	0
3	1984	13.7	15.3	13.5	6	9	6.5	-3	9	0.5	.25
4	1985	14.0	13.4	13.5	7	3	6.5	4	16	0.5	.25
5	1986	13.4	14.0	13.2	5	6	5	-1	1	0	0
6	1987	12.1	12.8	13.0	1	2	3	-1	1	2	4
7	1988	12.7	13.6	12.8	3	4	2	-1	1	1	1
8	1989	14.5	14.9	14.2	8	8	9	0	0	1	1
9	1990	12.5	12.5	12.5	2	1	1	1	1	1	1
					Σύνολα D των d_i^2				34		8.5

Πίνακας 5.5. Τιμές του $S_{n,\alpha}$ για διάφορες τιμές του μεγέθους n του δείγματος των ζευγών και για διά φορες τιμές της (σ.σ): α, που απαιτούνται για τον έλεγχο της H_o: $r_s=0$ όπου r_s ο συντελεστής συσχέτισης του Spearman

n	α=0.01	α=0.05	α=0.1
6	0.829	0.886	1.
7	0.714	0.786	0.929
8	0.643	0.738	0.885
9	0.600	0.683	0.845
10	0.564	0.648	0.808
11	0.520	0.620	0.773
12	0.496	0.591	0.741
13	0.475	0.566	0.716
14	0.456	0.544	0.695

5.5. Μέθοδοι ελέγχου ομοιογένειας και ομοιογενοποίηση γενικά των παραμέτρων της βάσης δεδομένων.

Η ομοιογένεια των χρονοσειρών των περισσοτέρων παραμέτρων που αναφέρονται στο **τεύχος 10/2 του " Υδροσκόπιου ", στο Κεφάλαιο** 3, μπορεί να ελεγχθεί με τις μεθόδους ελέγχου ομοιογένειας που έχουν αναφερθεί στο Κεφάλαιο 4.

`Οσον αφορά δε την ομοιογενοποίηση, οι προτεινόμενες μέθοδοι ομοιογενοποίησης μπορούν να εφαρμοσθούν στις μετεωρολογικές παραμέτρους που παρατίθενται στον παρακάτω Πίνακα 5.6.

Τα κωδικά γράμματα της 3ης στήλης του Πίνακα έχουν τη σημασία:

α. Μέση ή αθροιστική ετήσια τιμή.

β. Μέση ή αθροιστική μηνιαία τιμή.

γ. Μέση ή αθροιστική τιμή δεκαημέρου.

Οι κωδικοί αριθμοί της 4ης στήλης του Πίνακα έχουν τη σημασία:

1. Καθε διαθέσιμης ώρας παρατήρηση και κυρίως των συνοπτικών ωρών.
2. Μέση ημερήσια τιμή ή αθροιστική τιμή.
3. Μέγιστη ή ελάχιστη ημερήσια τιμή από όργανα άμεσης ανάγνωσης.
4. Συχνότητα ειδικών γεγονότων (π.χ αριθμός ημερών μερικού παγετού ανά έτος).
5. Συνάρτηση της παραμέτρου (π.χ Βαθμοημέρες).
6. Μέγιστη ή ελάχιστη ημερήσια τιμή από αυτογραφικά όργανα ή υπολογισμό.
7. Διαφορά των τιμών της παραμέτρου μεταξύ δύο διαδοχικών παρατηρήσεων.

Πινακας 5.6 Μετεωρολογικές παράμετροι, στις οποίες μπορούν να εφαρμοστούν οι μέθοδοι ομοιογενοποίησης των τιμών.

α/α	Μετεωρολογική Παράμετρος	Τιμές στις οποίες εφαρμόζονται οι μέθοδοι	Κωδικός αριθμός
1.1	Θερμοκρασία (T) αέρα	α,β,γ	1,2,3,4,5,6,7
1.2	T υγρού θερμομέτρου	α,β,γ	1,2,7
1.3	Σημείο δρόσου (DP)	α,β,γ	1,2,7
1.4	T χλοερού εδάφους (T_ε)	α,β,γ	1,2,6
1.5	T_ε σε διάφορα βάθη dd(cm)	α,β,γ	1,2,6
1.6	T επιφάνειας θαλάσσης(T_θ)	α,β	1,2
1.7	T >> λιμνών (T_λ)	α,β	1,2
1.8	T αέρα σε ισοβαρικές επιφάνειες ddd (hPa)	α,β	1,2,7
1.9	T υγρού θερμομέτρου σε ισοβαρικές επιφάνειες	α,β	1,2,7
1.10	Αλλες T (Δυναμική κλπ.)	α,β	1,2,6
2.1	Ατμοσφ. πίεση σταθμού (P)	α,β,γ	1,2,4,6,7
2.2	P στην μέση στάθμη θάλασσ.	α,β,γ	1,2,4,6,7
2.3	Υψος ισοβ. επιφ. ddd(hPa)	α,β,γ	1,2,7
3.1	Σχετική υγρασία	α,β	1,2,4,6,7
3.2	Απόλυτη υγρασία	α,β	1,2,6
4.1	Εξάτμιση Pich ή Pan	α,β	2
4.2	Δυναμική εξάτμιση	α,β	2
5.	Νέφωση (ολική)	α,β	1,2,4
6.1	Ηλιοφάνεια	α,β,γ	2,4
6.2	Κλάσμα ηλιοφάνειας	α,β,γ	2,4
7.1	Ολική Ηλιακή ακτινοβολία	α,β,γ	2,4
7.2	Σχετική ολική GL	α,β,γ	2,4

α/α	Μετεωρολογική Παράμετρος	Τιμές στις οποίες εφαρμόζονται οι μέθοδοι	Κωδικός αριθμός
8.	Ορατότητα (οριζόντια)	α,β,γ	1,3,4
9.1	Υψος βροχής	α,β,γ	1,2,4
9.2	Διάρκεια βροχής	α,β,γ	2
10	Υψος χιονιού	α,β,γ	2,4

ΑΝΑΦΟΡΕΣ

1. **Αδαμόπουλος,Λ.,Α:** Στατιστική. Πανεπιστημιακαί παραδόσεις. Τρεις τομοι. Θεσσαλονικη 1963.
2. **Βασιλείου, Χ.,Π. :** Στοχαστικές μέθοδοι στις επιχειρησιακές έρευνες. Εκδοσεις Χριστοδουλίδη. Θεσσαλονικη 1985.
3. **Κούνια,Σ.- Κολυβά-Μαχαίρα,Φ.- Μπαγιάτη,Κ.- Μπορα-Σέντα,Ε. :** Εισαγωγή στην Στατιστική. Εκδοσεις Χριστοδουλίδη. Θεσσαλονικη 1990.
4. **Μιμίκου Μ.:**Τεχνολογία Υδατικών Πορων.Αθήνα 1990.
5. **Χατζηδάκη-Θέου:**Στατιστικά χαρακτηριστικά θερμοκρασίας αέρος στο Αστεροσκοπείο Αθηνών (1931-1990).Μάρτιος 1991.

---

6. **Box,P.,E.,Geor. and Jenkins,M.,G.:** Time series analysis. Forecasting and control. HOLDEN-DAY Series in time series analysis. Gwilym M. Jenkins and Emanuel Parzen, editors. 1970.
7. **Buishand.T.A.:**Some methods for testing the homogeneity of rainfall records.Journal of Hydrology,58 (1982).
8. **R.Sneyers:** On the statistical analysis of series of observations.WMO N.415.T.N.143. 1990.
9. **Thom H.C.S.:**Some methods of Climatological analysis.WMO N.199.T.N.81. 1966.
10. **WMO. Climatic change.**WMO N.195.T.N.79. 1966.
11. **WMO. Guide to Climatological Practices.**WMO N.100.
12. **WMO. International Workshop on precipitation measurements.** WMO. Instruments and observing methods.N.48 (WMO/TD-328). 1989.
13. **WMO. CCL Working Group on Climate Data.**Geneva ,11-15 November 1991.WCDMP.N.18
14. **WMO. Guidelines for computerized data processing in operational Hydrology and land and water management.** WMO.N.634. 1985.

ΜΑΘΗΜΑΤΙΚΟΙ ΣΥΜΒΟΛΙΣΜΟΙ

$$E(X)=\frac{{}_1\Sigma^k x_i}{k}=\frac{x_1+x_2+...+x_k}{k}$$:η μέση τιμή του πληθυσμού της παραμέτρου X.

$I_k=\{1,2,...,k\}$:το σύνολο των φυσικών αριθμών από τον 1 μέχρι τον κ.

$i\varepsilon I_k$:η τιμή i είναι ένας από τους φυσικούς αριθμούς από το 1 μέχρι τον κ

$$\mu_x=\mu=\frac{{}_1\Sigma^k x_i}{k}=\frac{x_1+x_2+...+x_k}{k}$$:η μέση τιμή του δείγματος μεγέθους κ στοιχείων της παραμέτρου X.

N :το σύνολο των φυσικών αριθμών.

${}_1\Sigma^k x_i=x_1+x_2+...+x_k$:εκφράζει το άθροισμα των κ διαδοχικών τιμών:x_i,$i\varepsilon I_k$.

$$\sigma_x{}^2=\sigma^2=\frac{{}_1\Sigma^k(x_i-\mu_k)^2}{k}$$:η διασπορά του πληθυσμού της παραμέτρου X.

$\sigma_x=\sqrt{\sigma_x{}^2}$:η τυπική απόκλιση του πληθυσμού της παραμέτρου X.

$$S_x{}^2=\frac{{}_1\Sigma^n(x_i-E(X))^2}{n}$$:η διασπορά ή διακύμανση του δείγματος μεγέθους n της παραμέτρου X.

$S_x=\sqrt{S_x{}^2}$:η τυπική απόκλιση του δείγματος της παραμέτρου X.

$\sqrt{}$:η τετραγωνική ρίζα.

/ :η διαίρεση.

* :ο πολλαπλασιασμός.

. :η δεκαδική τελεία

∞ :το άπειρο σύνολο

$\neq$:η διαφορετική τιμή

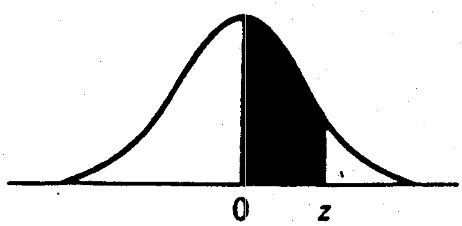
ΠΑΡΑΡΤΗΜΑ

ΣΤΑΤΙΣΤΙΚΟΙ ΠΙΝΑΚΕΣ

Πινακας Ια.[10]

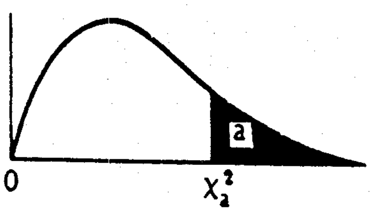
Βαθμοι ελευθεριας	πιθανοτητα επι τοις %			
	95*	97.5†	99	99.5
	κανονικη κατανομη Gauss			
(Infinite)	1.645	1.960	2.326	2.576
	Student's t-κατανομη			
100	1.660	1.984	2.365	2.626
60	1.671	2.000	2.390	2.660
40	1.684	2.021	2.423	2.704
30	1.697	2.042	2.457	2.750
25	1.708	2.060	2.485	2.787
20	1.725	2.086	2.528	2.845
15	1.753	2.131	2.602	2.947
12	1.782	2.179	2.681	3.055
10	1.812	2.228	2.764	3.169
8	1.860	2.306	2.896	3.355
6	1.943	2.447	3.143	3.707
4	2.132	2.776	3.747	4.604

Πινακας ΙΒ. Της πιθανοτητας $P(0<Z<z)$ για την κανονική κατανομή $N(0,1)$. [3]



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Πινακας IIα. Των τιμών $X^2_{\nu,\alpha}$ της X^2 κατανομής για τις οποίες ισχύει $P(X^2>X^2_{\nu,\alpha})=\alpha$. [3]

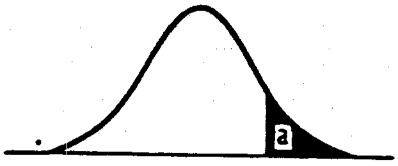


Β.ε	a = .995	a = .990	a = .975	a = .950	a = .900
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

Πινακας ΙΙβ. Των τιμών $X^2_{\nu,\alpha}$ της X^2 κατανομής για τις οποίες ισχύει $P(X^2 > X^2_{\nu,\alpha}) = \alpha$. [3]

a = .10	a = .05	a = .025	a = .010	a = .005	B.E.
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8454	31.9999	34.2672	16
24.7690	27.5871	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
51.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1539	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100

Πινακας III. Των τιμών της t- κατανομής ώστε P $(t_v > t_{v,\alpha}) = \alpha$.[3]



Β.ε	a = .10	a = .05	a = .025	a = .010	a = .005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576

Πινακας IVα. Των τιμών του $D_{a,n}$ στο κριτήριο Kolmogorov-Smornov. [3]

Μέγεθος δείγματος n	Στάθμη σημαντικότητας a .20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.829
4	.494	.525	.564	.624	.734
5	.446	.474	.510	.563	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.409	.486
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.391
17	.250	.266	.286	.318	.380
18	.244	.259	.278	.309	.370
19	.237	.252	.272	.301	.361
20	.231	.246	.264	.294	.352
25	.21	.22	.24	.264	.32
30	.19	.20	.22	.242	.29
35	.18	.19	.21	.23	.27
40				.21	.25
50				.19	.23
60				.17	.21
70				.16	.19
80				.15	.18
90				.14	
100				.14	
Προσεγγιστικοί τύποι	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Πινακας IVβ. Των τιμών του $D_{m,n}$ στο κριτήριο Kolmogorov-Smornov. [3]

n \ m	1	2	3	4	5	6	7	8	9	10	12	15
1	* *	* *	* *	* *	* *	* *	* *	* *	* *	* *		
2		* *	* *	* *	* *	* *	* *	7/8 *	16/18 *	9/10 *		
3			* *	* *	12/15 *	5/6 *	18/21 *	18/24 *	7/9 8/9		9/12 11/12	
4				3/4 *	16/20 *	9/12 10/12	21/28 24/28	6/8 7/8	27/36 32/36	14/20 16/20	8/12 10/12	
5					4/5 4/5	20/30 25/30	25/35 30/35	27/40 32/40	31/45 36/45	7/10 8/10		10/15 11/15
6						4/6 5/6	29/42 35/42	16/24 18/24	12/18 14/18	19/30 22/30	7/12 9/12	
7							5/7 5/7	35/56 42/56	40/63 47/63	43/70 53/70		
8								5/8 6/8	45/72 54/72	23/40 28/40	14/24 16/24	
9									5/9 6/9	52/90 62/90	20/36 24/36	
10										6/10 7/10		15/30 19/30
12											6/12 7/12	30/60 35/60
15												7/15 8/15

Σημ. 1. Οι πάνω και κάτω τιμές των τετραγώνων αναφέρονται σε επίπεδα σημαντικότητας a = .05 και a = .01 αντίστοιχα.

Σημ. 2. Ο *(αστερίσκος) σημαίνει αποδοχή της αρχικής υπόθεσης, δηλαδή ότι τα δείγματα προέρχονται από τον ίδιο πληθυσμό.

Σημ. 3. Για μεγάλα m και n τα δύο a-σημεία δίνονται κατά προσέγγιση από τους τύπους.

$$a = 0.05{:}1.36\sqrt{\frac{m+n}{mn}} \cdot a = 0.01{:}1.63\sqrt{\frac{m+n}{mn}}$$

Πινακας Vα. Των πιθανοτήτων που αντιστοιχούν σε τιμές του U μικροτερες απο την παρατηρούμενη στο test Mann-Whitney. [3].

$n_2 = 3$

U \ n_1	1	2	3
0	.250	.100	.050
1	.500	.200	.100
2	.750	.400	.200
3		.600	.350
4			.500
5			.650

$n_2 = 4$

U \ n_1	1	2	3	4
0	.200	.067	.028	.014
1	.400	.133	.057	.029
2	.600	.267	.114	.057
3		.400	.200	.100
4		.600	.314	.171
5			.429	.243
6			.571	.343
7				.443
8				.557

$n_2 = 5$

U \ n_1	1	2	3	4	5
0	.167	.047	.018	.008	.004
1	.333	.095	.036	.016	.008
2	.500	.190	.071	.032	.016
3	.667	.286	.125	.056	.028
4		.429	.196	.095	.048
5		.571	.286	.143	.075
6			.393	.206	.111
7			.500	.278	.155
8			.607	.365	.210
9				.452	.274
10				.548	.345
11					.421
12					.500
13					.579

$n_2 = 6$

U \ n_1	1	2	3	4	5	6
0	.143	.036	.012	.005	.002	.001
1	.286	.071	.024	.010	.004	.002
2	.428	.143	.048	.019	.009	.004
3	.571	.214	.083	.033	.015	.008
4		.321	.131	.057	.026	.013
5		.429	.190	.086	.041	.021
6		.571	.274	.129	.063	.032
7			.357	.176	.089	.047
8			.452	.238	.123	.066
9			.548	.305	.165	.090
10				.381	.214	.120
11				.457	.268	.155
12				.545	.331	.197
13					.396	.242
14					.465	.294
15					.535	.350
16						.409
17						.469
18						.531

Πίνακας V8. Των πιθανοτήτων που αντιστοιχούν σε τιμές του U μικροτερες από την παρατηρούμενη στο test Mann-Whitney. [3].

$n_2 = 7$

U \ n_1	1	2	3	4	5	6	7
0	.125	.028	.008	.003	.001	.001	.000
1	.250	.056	.017	.006	.003	.001	.001
2	.375	.111	.033	.012	.005	.002	.001
3	.500	.167	.058	.021	.009	.004	.002
4	.625	.250	.092	.036	.015	.007	.003
5		.333	.133	.055	.024	.011	.006
6		.444	.192	.082	.037	.017	.009
7		.556	.258	.115	.053	.026	.013
8			.333	.158	.074	.037	.019
9			.417	.206	.101	.051	.027
10			.500	.264	.134	.069	.036
11			.583	.324	.172	.090	.049
12				.394	.216	.117	.064
13				.464	.265	.147	.082
14				.538	.319	.183	.104
15					.378	.223	.130
16					.438	.267	.159
17					.500	.314	.191
18					.562	.365	.228
19						.418	.267
20						.473	.310
21						.527	.355
22							.402
23							.451
24							.500
25							.549

Πινακας Vγ. Των πιθανοτήτων που αντιστοιχούν σε τιμές του U μικροτερες απο την παρατηρούμενη στο test Mann-Whitney. [3].

$n_2 = 8$

U \ n_1	1	2	3	4	5	6	7	8	t	Κανονική
0	.111	.022	.006	.002	.001	.000	.000	.000	3.308	.001
1	.222	.044	.012	.004	.002	.001	.000	.000	3.203	.001
2	.333	.089	.024	.008	.003	.001	.001	.000	3.098	.001
3	.444	.133	.042	.014	.005	.002	.001	.001	2.993	.001
4	.556	.200	.067	.024	.009	.004	.002	.001	2.888	.002
5		.267	.097	.036	.015	.006	.003	.001	2.783	.003
6		.356	.139	.055	.023	.010	.005	.002	2.678	.004
7		.444	.188	.077	.033	.015	.007	.003	2.573	.005
8		.556	.248	.107	.047	.021	.010	.005	2.468	.007
9			.315	.141	.064	.030	.014	.007	2.363	.009
10			.387	.184	.085	.041	.020	.010	2.258	.012
11			.461	.230	.111	.054	.027	.014	2.153	.016
12			.539	.285	.142	.071	.036	.019	2.048	.020
13				.341	.177	.091	.047	.025	1.943	.026
14				.404	.217	.114	.060	.032	1.838	.033
15				.467	.262	.141	.076	.041	1.733	.041
16				.533	.311	.172	.095	.052	1.628	.052
17					.362	.207	.116	.065	1.523	.064
18					.416	.245	.140	.080	1.418	.078
19					.472	.286	.168	.097	1.313	.094
20					.528	.331	.198	.117	1.208	.113
21						.377	.232	.139	1.102	.135
22						.426	.268	.164	.998	.159
23						.475	.306	.191	.893	.185
24						.525	.347	.221	.788	.215
25							.389	.253	.683	.247
26							.433	.287	.578	.282
27							.478	.323	.473	.318
28							.522	.360	.368	.356
29								.399	.263	.396
30								.439	.158	.437
31								.480	.052	.481
32								.520		

Πινακας Vδ. Κρισιμες τιμές του U στο test Mann-Whitney.[3]

Για μονόπλευρο test a = .001
Για δίπλευρο test a = .002

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2												
3									0	0	0	0
4		0	0	0	1	1	1	2	2	3	3	3
5	1	1	2	2	3	3	4	5	5	6	7	7
6	2	3	4	4	5	6	7	8	9	10	11	12
7	3	5	6	7	8	9	10	11	13	14	15	16
8	5	6	8	9	11	12	14	15	17	18	20	21
9	7	8	10	12	14	15	17	19	21	23	25	26
10	8	10	12	14	17	19	21	23	25	27	29	32
11	10	12	15	17	20	22	24	27	29	32	34	37
12	12	14	17	20	23	25	28	31	34	37	40	42
13	14	17	20	23	26	29	32	35	38	42	45	48
14	15	19	22	25	29	32	36	39	43	46	50	54
15	17	21	24	28	32	36	40	43	47	51	55	59
16	19	23	27	31	35	39	43	48	52	56	60	65
17	21	25	29	34	38	43	47	52	57	61	66	70
18	23	27	32	37	42	46	51	56	61	66	71	76
19	25	29	34	40	45	50	55	60	66	71	77	82
20	26	32	37	42	48	54	59	65	70	76	82	88

Πινακας Vε. Κρισιμες τιμές του U στο test Mann-Whitney.[3]

Για μονόπλευρο test a = .025
Για δίπλευρο test a = .05

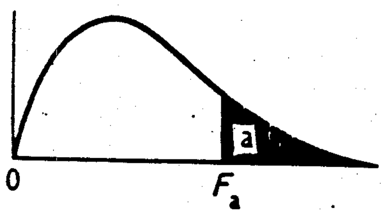
n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Πινακας Vστ. Κρισιμες τιμές του U στο test Mann-Whitney.[3]

Για μονόπλευρο test a = .05
Για δίπλευρο test a = .10

n_2 \ n_1	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

Πινακας VIα. Των τιμών $F_{\nu 1,\nu 2,\alpha}$ της F-κατανομής για τις οποίες ισχύει $P(F>F_{\nu 1,\nu 2,\alpha})=\alpha$. [3]



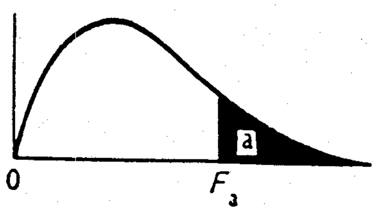
Βαθμοί ελευθερίας (a = .01)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Πίνακας VIβ. Των τιμών $F_{\nu_1,\nu_2,\alpha}$ της F-κατανομής για τις οποίες ισχύει $P(F>F_{\nu_1,\nu_2,\alpha})=\alpha$. [3]

10	12	15	20	24	30	40	60	120	∞	ν_1 / ν_2
6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	1
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	2
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	3
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	4
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	5
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	6
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	7
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	8
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	9
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	10
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	11
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	12
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	13
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	14
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	15
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	16
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	17
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	18
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	19
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	20
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	21
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	22
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	23
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	24
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	25
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	26
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	27
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	28
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	29
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	30
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	40
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	60
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	120
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	∞

Πινακας VIγ. Των τιμών $F_{\nu_1,\nu_2,\alpha}$ της F-κατανομής για τις οποίες ισχύει $P(F>F_{\nu_1,\nu_2,\alpha})=\alpha$. [3]



Βαθμοί Ελευθερίας (a = .05)

ν_1 / ν_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Πίνακας VIβ. Των τιμών $F_{v1,v2,\alpha}$ της F-κατανομής για τις οποίες ισχύει $P(F>F_{v1,v2,\alpha})=\alpha$. [3]

10	12	15	20	24	30	40	60	120	∞	ν_1 / ν_2
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	1
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	2
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	3
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	4
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	5
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	6
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	7
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	8
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	9
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	10
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	11
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	12
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	13
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	14
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	15
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	16
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	17
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	18
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	19
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	20
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	21
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	22
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	23
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	24
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	25
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	26
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	27
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	28
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	29
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	30
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	40
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	60
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	120
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	∞