

**Μεταπτυχιακό Υπολογιστικής Φυσικής,
2018-2019
Εργασία στο μάθημα 'Ανάλυση Δεδομένων'**

<http://users.auth.gr/dkugiu/Teach/DataAnalysis/index.html>

Δημήτρης Κουγιουμτζής

E-mail: dkugiu@auth.gr

17 Ιανουαρίου 2019

Οδηγίες: Σχετικά με την παράδοση της εργασίας θα πρέπει:

- Το κείμενο της αναφοράς της ανάλυσης που ζητείται να είναι γραμμένο σε κάποιο πρόγραμμα επεξεργασίας κειμένου (π.χ. Word, LaTeX).
- Για κάθε ζήτημα, θα δίνονται τα βασικά αποτελέσματα, όπου ζητούνται, καθώς και τα γραφήματα και οι πίνακες, και τα αντίστοιχα σχόλια με αναφορά στις συναρτήσεις και προγράμματα Matlab, τα οποία θα πρέπει να παραδοθούν ως ξεχωριστά αρχεία μαζί με την εργασία.
- Η εργασία και τα προγράμματα θα πρέπει να αποσταλούν στο διδάσκοντα με το ηλεκτρονικό ταχυδρομείο (e-mail: dkugiu@auth.gr) ως την **Κυριακή 10/2/2019**, καλύτερα σε ένα συμπιεσμένο αρχείο.

Για την απάντηση στα ζητήματα της εργασίας θα χρειαστεί να αναπτύξετε κάποιους αλγόριθμους (σε μορφή συνάρτησης), ενώ άλλοι αλγόριθμοι δίνονται (κάποιοι σε μορφή συνάρτησης Matlab) στην ιστοσελίδα του μαθήματος (στο πλαίσιο των ασκήσεων στο μάθημα). Επίσης θα πρέπει να κάνετε και τα κατάλληλα προγράμματα που καλούν τις συναρτήσεις.

Για την ανάγνωση αρχείου excel (.xls, .xlsx) θα πρέπει να χρησιμοποιήσετε την εντολή `xlsread` του Matlab. Για τη μεταφορά σχήματος από το Matlab στο πρόγραμμα επεξεργασίας κειμένου θα πρέπει να αποθηκεύσετε το σχήμα σε αρχείο χρησιμοποιώντας είτε το παράθυρο εντολών (δες εντολή `print` του

Matlab) ή το μενού στο παράθυρο του σχήματος (File → Export). Για μεταφορά στο Word υπάρχουν διάφορες κατάλληλες μορφές αρχείου εικόνας, η πιο απλή είναι "Enhanced Metafile" (*.emf). Για μεταφορά στο LaTeX η πιο κατάλληλη μορφή είναι "Encapsulated Postscript" (*.eps) ή συμπιεσμένης εικόνας (*.jpg).

Περιγραφή εργασίας και δεδομένα

Η Ευρωπαϊκή Εταιρεία Περιβάλλοντος (European Environment Agency, EEA) συλλέγει στοιχεία για τη μόλυνση του αέρα από διάφορες χώρες. Η ανάλυση της εργασίας θα γίνει σε μετρήσεις του δείκτη αιωρούμενων σωματιδίων P10 που συλλέχθηκαν από την EEA για συγκεκριμένους τομείς δραστηριότητας που συμμετέχουν στη μόλυνση του αέρα. Τα δεδομένα είναι ετήσια για την περίοδο 1990-2007 και προέρχονται από τις χώρες της Ευρωπαϊκής Ένωσης (ΕΕ) των 15 χωρών (πριν την είσοδο των υπολοίπων χωρών). Τα δεδομένα για κάθε τομέα δραστηριότητας δίνονται σε διαφορετικό αρχείο δεδομένων. Το κάθε αρχείο (σε μορφή excel) έχει στην πρώτη στήλη το όνομα του τομέα δραστηριότητας, στη δεύτερη στήλη το έτος και στις υπόλοιπες 15 στήλες τα δεδομένα για κάθε χώρα με την παρακάτω σειρά: Αυστρία, Βέλγιο, Δανία, Φιλανδία, Γαλλία, Γερμανία, Ελλάδα, Ιρλανδία, Ιταλία, Λουξεμβούργο, Ολλανδία, Πορτογαλία, Ισπανία, Σουηδία και Μεγάλη Βρετανία. Η πρώτη γραμμή έχει τα ονόματα των χωρών (στην αντίστοιχη στήλη) και την περιγραφή του P10, π.χ. ένα κελί της πρώτης γραμμής μπορεί να είναι: Emissions - Gg (1000 tonnes) - United Kingdom - Total Particulate Formation PM10.

Βασικά ερωτήματα της εργασίας είναι πως κατανέμεται το P10 (1000 tonnes) στις διάφορες χώρες, στα διάφορα έτη, καθώς και στις διάφορες δραστηριότητες. Επίσης θέλουμε να δούμε αν συσχετίζεται το P10 μεταξύ των διαφόρων χωρών αλλά και των δραστηριοτήτων με βάση τα δεδομένα την περίοδο 1990-2007. Θέλουμε ακόμα να εκτιμήσουμε μοντέλο που να προσδιορίζει τη συνολική μέτρηση P10 (NationalTotals) της κάθε χώρας από τις μετρήσεις των επιμέρους δραστηριοτήτων. Τέλος θέλουμε να διερευνήσουμε αν μπορούμε να προβλέψουμε τη μέτρηση του P10 το επόμενο έτος γνωρίζοντας τις μετρήσεις τα προηγούμενα έτη.

Τα ονόματα των αρχείων δεδομένων δίνονται παρακάτω:

A/A	Όνομα
1	EmissionP10EU15Agriculture
2	EmissionP10EU15EnergyIndustries
3	EmissionP10EU15EnergyIndustriesPowerProduction1A1a
4	EmissionP10EU15FugitiveEmissions
5	EmissionP10EU15IndustryEnergy
6	EmissionP10EU15IndustryProcesses
7	EmissionP10EU15OtherEnergy
8	EmissionP10EU15OtherTransport
9	EmissionP10EU15RoadTransport
10	EmissionP10EU15Waste
11	EmissionP10NationalTotalsEU15

Ζητήματα εργασίας

Τα ζητήματα που θα ασχοληθεί ο/η φοιτητής/τρια αντιστοιχούν στους δείκτες από 1 ως 10 που δίνονται για κάθε φοιτητή/τρια στην ιστοσελίδα του μαθήματος.

1. Θέλουμε να ελέγξουμε οπτικά αν το P10 έχει την ίδια κατανομή σε δύο χώρες για την ίδια δραστηριότητα με βάση τα δεδομένα την περίοδο 1990-2007 ή σε δύο δραστηριότητες για την ίδια χώρα. Για αυτό θα πρέπει να φτιάξετε μια **συνάρτηση** που θα παίρνει ως είσοδο δύο διανύσματα (ή έναν πίνακα με δύο στήλες). Τα διανύσματα θα έχουν τα δεδομένα του P10 για δύο χώρες και για κάποια δραστηριότητα ή για δύο δραστηριότητες και για κάποια χώρα. Επίσης στην είσοδο θα δίνονται τα ονόματα των δύο χωρών καθώς και το όνομα της δραστηριότητας (στα αγγλικά) ή τα ονόματα των δύο δραστηριοτήτων και το όνομα της χώρας, αντίστοιχα. Η συνάρτηση θα σχηματίζει τις δύο πολυγωνικές γραμμές των δύο ιστογραμμάτων σε ένα σχήμα. Επειδή τα δεδομένα για τις δύο χώρες ή δραστηριότητες μπορεί να διαφέρουν ως προς τη μέση τιμή και διασπορά, η συνάρτηση θα δίνει και ένα δεύτερο σχήμα του ίδιου τύπου αλλά αφού τα δεδομένα για κάθε χώρα ή δραστηριότητα έχουν κανονικοποιηθεί ώστε να έχουν μέση τιμή μηδέν και τυπική απόκλιση ένα. Στα σχήματα θα πρέπει να φαίνονται οι χώρες ή δραστηριότητες που αντιστοιχούν οι καμπύλες και η δραστηριότητα ή χώρα, αντίστοιχα, στην οποία αναφέρονται.

Θα κάνετε **πρόγραμμα** που καλεί τη συνάρτηση ενδεικτικά για διάφορες χώρες και δραστηριότητες.

Φαίνεται να διαφέρουν οι κατανομές του P10 για κάποια δραστηριότητα στις διάφορες χώρες ή για κάποια χώρα σε διάφορες δραστηριότητες ;

2. Για κάθε δραστηριότητα και χώρα θέλουμε να διερευνήσουμε αν η κατανομή του δείκτη P10 είναι κανονική. Θα φτιάξετε ένα **πρόγραμμα** που θα φορτώνει τα αρχεία για τις 10 δραστηριότητες (όχι το ενδέκατο αρχείο) και για κάθε δραστηριότητα και χώρα θα εξετάζει αν η κατανομή του δείκτη P10 με βάση τα δεδομένα την περίοδο 1990-2007 μπορεί να είναι κανονική (έλεγχος καλής προσαρμογής). Θα συλλέγει τις απαντήσεις και θα δίνει το ποσοστό των περιπτώσεων που η κατανομή βρέθηκε να είναι κανονική.

Στη συνέχεια το πρόγραμμα θα αθροίζει για κάποια δραστηριότητα τις τιμές του P10 από όλες τις 15 χώρες (για κάθε έτος ξεχωριστά). Το πρόγραμμα θα επαναλάβει τον έλεγχο κανονικής κατανομής στα δεδομένα των 15 χωρών την περίοδο 1990-2007 για κάθε δραστηριότητα ξεχωριστά. Το πρόγραμμα θα αναφέρει τα αποτελέσματα του ελέγχου για κάθε δραστηριότητα. Τέλος θα κανονικοποιήσει τα 10 δείγματα για τις 10 δραστηριότητες ώστε να έχουν μέση τιμή μηδέν και τυπική απόκλιση ένα και θα δημιουργήσει ένα σχήμα με τις 10 πολυγωνικές γραμμές ιστογράμματος, μια για κάθε δραστηριότητα.

Φαίνεται η κατανομή του δείκτη P10 για κάθε χώρα και δραστηριότητα να είναι κανονική; Φαίνεται να είναι κανονική η κατανομή του δείκτη P10 από τις 15 χώρες σε μια δραστηριότητα ;

3. Θέλουμε να εκτιμήσουμε για κάποια δραστηριότητα τη μέση σχετική τιμή P10 για κάθε χώρα από τα δεδομένα την περίοδο 1990-2007. Η σχετική τιμή P10 για κάποια δραστηριότητα ορίζεται ως ο λόγος της τιμής P10 για αυτήν τη δραστηριότητα ως προς την αντίστοιχη συνολική τιμή (από το αρχείο EmissionP10NationalTotalsEU15).

Για αυτό θα πρέπει να υπολογίσετε διάστημα εμπιστοσύνης σε επίπεδο 95% κάνοντας χρήση της παραμετρικής κατανομής student αλλά και με κατανομή που σχηματίζεται με τη μέθοδο bootstrap (δημιουργώντας μεγάλο πλήθος αντιγράφων του αρχικού δείγματος με τυχαία επαναδειγματοληψία με επανάθεση, δες άσκηση 5.4).

Θα πρέπει να φτιάξετε **δύο συναρτήσεις**. Η **πρώτη συνάρτηση** θα παίρνει ως είσοδο δύο διανύσματα μήκους 18, το ένα με την τιμή του P10 για μια δραστηριότητα και χώρα και το άλλο με τη συνολική τιμή P10 για την ίδια χώρα και την περίοδο 1990-2007, καθώς και ένα επίπεδο σημαντικότητας. Η συνάρτηση θα υπολογίζει και θα δίνει στην έξοδο τα δύο διαστήματα εμπιστοσύνης (παραμετρικό και μη-

παραμετρικό) για το δεδομένο επίπεδο σημαντικότητας (ή αντίστοιχα επίπεδο εμπιστοσύνης).

Η **δεύτερη συνάρτηση** θα παίρνει ως είσοδο δύο πίνακες, ο ένας με την τιμή του P10 για μια δραστηριότητα και ο άλλος με τη συνολική τιμή P10 για τις 15 χώρες και την περίοδο 1990-2007, καθώς και ένα επίπεδο σημαντικότητας. Η συνάρτηση θα καλεί για κάθε χώρα την πρώτη συνάρτηση και θα συλλέγει το παραμετρικό και μη-παραμετρικό διάστημα εμπιστοσύνης μέσης τιμής για κάθε χώρα. Στη συνέχεια θα δημιουργεί ένα σχήμα με τη μέση τιμή και τα δύο διαστήματα εμπιστοσύνης για κάθε χώρα (οι χώρες με κωδικό 1 ως 15 στον οριζόντιο άξονα).

Φαίνεται να διαφέρουν οι χώρες ως προς τη μέση σχετική τιμή P10; Φαίνεται να διαφέρουν τα παραμετρικά και μη-παραμετρικά διαστήματα εμπιστοσύνης;

4. Το ίδιο με το ζήτημα 3 αλλά για την τυπική απόκλιση.
5. Θέλουμε να διερευνήσουμε αν υπάρχει συσχέτιση μεταξύ του P10 σε δύο διαφορετικές χώρες (και την ίδια δραστηριότητα). Γι αυτό θα κάνετε κατάλληλο παραμετρικό έλεγχο για μηδενική συσχέτιση χρησιμοποιώντας το στατιστικό της κατανομής Student, καθώς και έλεγχο τυχαιοποίησης χρησιμοποιώντας μεγάλο πλήθος τυχαιοποιημένων δειγμάτων (άσκηση 5.2).

Θα πρέπει να φτιάξετε **δύο συναρτήσεις**. Η **πρώτη συνάρτηση** θα παίρνει ως είσοδο δύο διανύσματα (ίσου μήκους) για την τιμή P10 σε δύο διαφορετικές χώρες (και την ίδια δραστηριότητα) την περίοδο 1990-2007, καθώς και ένα επίπεδο σημαντικότητας. Η συνάρτηση θα δίνει την τιμή του συντελεστή συσχέτισης και δύο p -τιμές, για τον παραμετρικό και τον μη-παραμετρικό έλεγχο (τυχαιοποίησης) σημαντικότητας της συσχέτισης.

Η **δεύτερη συνάρτηση** θα παίρνει ως είσοδο έναν πίνακα 15 στηλών με την τιμή του P10 για μια δραστηριότητα, τις 15 χώρες και την περίοδο 1990-2007, καθώς και ένα επίπεδο σημαντικότητας. Η συνάρτηση θα σαρώνει όλα τα ζεύγη των 15 χωρών και για κάθε ζεύγος θα καλεί την πρώτη συνάρτηση και θα συλλέγει την τιμή του συντελεστή συσχέτισης και τις p τιμές για τον παραμετρικό και μη-παραμετρικό έλεγχο. Στη συνέχεια θα εμφανίζει ένα ιστόγραμμα των συντελεστών συσχέτισης για όλα τα ζεύγη χωρών, θα δίνει μια λίστα με τα 10 ζεύγη χωρών με την υψηλότερη συσχέτιση, και θα δίνει και το ποσοστό των ζευγών

με στατιστικά σημαντική συσχέτιση με βάση τον παραμετρικό και τον μη-παραμετρικό έλεγχο.

Καλέστε τη δεύτερη συνάρτηση για κάποιες δραστηριότητες. Φαίνεται να υπάρχει στατιστικά σημαντική συσχέτιση του P10 μεταξύ κάποιων χωρών ως προς κάποια δραστηριότητα; Φαίνεται να διαφέρουν οι παραμετρικοί και μη-παραμετρικοί έλεγχοι σημαντικότητας για το συντελεστή συσχέτισης;

6. Το ίδιο με το ζήτημα 5 αλλά για τη συσχέτιση μεταξύ του P10 σε δύο διαφορετικές δραστηριότητες και την ίδια χώρα. Θα πρέπει να σχηματιστεί ένας πίνακας για κάθε χώρα που θα έχει τις 10 δραστηριότητες στις στήλες και αυτός ο πίνακας θα πρέπει να δίνεται ως είσοδο στη δεύτερη συνάρτηση.
7. Θέλουμε να διερευνήσουμε αν μπορούμε, και σε ποιες περιπτώσεις (δραστηριότητα και χώρα), να προσδιορίσουμε τη συνολική τιμή του P10 (που σχηματίζεται από όλες τις δραστηριότητες, δίνεται στο αρχείο EmissionP10NationalTotalsEU15) από την αντίστοιχη τιμή σε κάποια δραστηριότητα.

Θα πρέπει να φτιάξετε **δύο συναρτήσεις**. Η **πρώτη συνάρτηση** θα παίρνει ως είσοδο δύο διανύσματα (ίσου μήκους) για την τιμή P10 σε μια δραστηριότητα και χώρα και τη συνολική τιμή P10 για την ίδια χώρα την περίοδο 1990-2007, καθώς και μια ακέραια τιμή για το μέγιστο βαθμό πολυωνύμου. Η συνάρτηση θα βρίσκει το πολυωνυμικό μοντέλο που προσαρμόζεται καλύτερα και δίνει την υψηλότερη τιμή του προσαρμοσμένου συντελεστή προσδιορισμού της συνολικής τιμής από την τιμή σε μια δραστηριότητα. Στην έξοδο θα δίνει αυτήν την τιμή, καθώς και τον αντίστοιχο βαθμό πολυωνύμου. Σημειώνεται ότι στη διερεύνηση θα συμπεριληφθεί και ο βαθμός μηδέν, δηλαδή να μην υπάρχει κάποιο πολυώνυμο που βελτιώνει το σφάλμα (δηλαδή το μοντέλο $y = \beta_0 + \epsilon$).

Η **δεύτερη συνάρτηση** θα παίρνει ως είσοδο δύο πίνακες, ο ένας με την τιμή του P10 για μια δραστηριότητα και ο άλλος με τη συνολική τιμή P10 για τις 15 χώρες και την περίοδο 1990-2007, καθώς και μια ακέραια τιμή για το μέγιστο βαθμού πολυωνύμου. Η συνάρτηση θα καλεί για κάθε χώρα την πρώτη συνάρτηση και θα συλλέγει τις τιμές του προσαρμοσμένου συντελεστή προσδιορισμού καθώς και τον αντίστοιχο βαθμό του πολυωνύμου. Στην έξοδο θα δίνει δύο διανύσματα, ένα με τις 15 τιμές του προσαρμοσμένου συντελεστή προσδιορισμού και ένα με τις αντίστοιχες τιμές του βαθμού πολυωνύμου.

Θα κάνετε ένα **πρόγραμμα** που θα καλεί τη δεύτερη συνάρτηση για

κάθε μια από τις 10 δραστηριότητες. Το πρόγραμμα θα δίνει για κάθε δραστηριότητα την υψηλότερη τιμή του προσαρμοσμένου συντελεστή προσδιορισμού, το βαθμό και τη χώρα για την οποία βρέθηκε η υψηλότερη τιμή (καλύτερα σε μορφή πίνακα). Επίσης θα δίνει ένα ραβδόγραμμα με τις συχνότητες εμφάνισης του κάθε βαθμού πολυωνύμου στο σύνολο των δραστηριοτήτων και χωρών.

Φαίνεται να υπάρχει κάποια χώρα όπου η συνολική τιμή P10 προσδιορίζεται με μεγαλύτερη ακρίβεια από την τιμή P10 σε κάποια δραστηριότητα; Φαίνεται κάποιος βαθμός πολυωνύμου να δίνει καλύτερη προσαρμογή από άλλους στο σύνολο των δραστηριοτήτων και χωρών;

8. Θέλουμε να διερευνήσουμε αν η συνολική τιμή του P10 (που σχηματίζεται από όλες τις δραστηριότητες (οι τιμές αυτές δίνονται στο αρχείο EmissionP10NationalTotalsEU15) μπορεί να προσδιορισθεί ικανοποιητικά από την τιμή P10 σε κάποιες δραστηριότητες. Θα δοκιμάσουμε μοντέλα με δύο βηματικές μεθόδους. Το πρώτο είναι το κλασσικό μοντέλο βηματικής παλινδρόμησης που έχει το Matlab (stepwisefit). Το δεύτερο θα το κατασκευάσετε εσείς όπως περιγράφεται παρακάτω.

Θα πρέπει να φτιάξετε **δύο συναρτήσεις**. Η **πρώτη συνάρτηση** θα παίρνει ως είσοδο έναν πίνακα με 11 στήλες, όπου η πρώτη στήλη θα έχει τη συνολική τιμή P10 και οι άλλες 10 στήλες την τιμή P10 στις επιμέρους δραστηριότητες για κάποια χώρα την περίοδο 1990-2007. Η συνάρτηση θα διαμορφώνει το μοντέλο παλινδρόμησης με βηματική τεχνική, όπου το κριτήριο για την επιλογή του νέου όρου στο μοντέλο μεταξύ όλων των υποψηφίων είναι ο προσαρμοσμένος συντελεστής προσδιορισμού. Στην έξοδο θα δίνει την τιμή του προσαρμοσμένου συντελεστή προσδιορισμού για το τελικό μοντέλο και τους συντελεστές του μοντέλου (αν το τελικό μοντέλο δεν είναι πλήρες κάποιοι συντελεστές θα είναι μηδέν).

Η **δεύτερη συνάρτηση** θα έχει την ίδια είσοδο με την πρώτη συνάρτηση. Θα καλεί την πρώτη συνάρτηση και επίσης θα καλεί τη συνάρτηση stepwisefit. Θα υπολογίζει επίσης και το πλήρες μοντέλο με όλες τις 10 μεταβλητές. Θα συγκρίνει τα τρία μοντέλα ως προς τον προσαρμοσμένο συντελεστή προσδιορισμού καθώς και ως προς τους συντελεστές του μοντέλου και θα τυπώνει αν τα τρία μοντέλα είναι ίδια, τους συντελεστές του κάθε μοντέλου και τον προσαρμοσμένο συντελεστή προσδιορισμού του κάθε μοντέλου. Επίσης θα δίνει ένα σχήμα όπου στον οριζόντιο άξονα θα είναι η πραγματική συνολική τιμή του P10 και στον κάθετο άξονα η εκτιμώμενη συνολική τιμή του P10, και θα φαίνονται και οι τρεις εκτιμήσεις (π.χ. με διαφορετικό χρώμα ή/και σύμβολο).

Θα κάνετε ένα **πρόγραμμα** που θα φορτώνει τα δεδομένα και για κάθε χώρα και θα καλεί τη δεύτερη συνάρτηση.

Φαίνεται να υπάρχει διαφορά μεταξύ των δύο τεχνικών επιλογής μεταβλητών στο μοντέλο παλινδρόμησης; Καταλήγουν οι μέθοδοι επιλογής μεταβλητών στο πλήρες μοντέλο; Συμφωνούν τα αποτελέσματα για όλες τις χώρες;

9. Το ίδιο με το ζήτημα 8 αλλά για το πρόβλημα παλινδρόμησης, όπου η εξαρτημένη μεταβλητή είναι η τιμή του P10 για κάποια δραστηριότητα για την Ελλάδα και ανεξάρτητες μεταβλητές είναι οι τιμές του P10 για την ίδια δραστηριότητα για τις άλλες 14 χώρες. Το πρόγραμμα θα καλεί τη (δεύτερη) συνάρτηση για κάθε δραστηριότητα.

Φαίνεται να υπάρχει διαφορά μεταξύ των δύο τεχνικών επιλογής μεταβλητών στο μοντέλο παλινδρόμησης; Καταλήγουν οι μέθοδοι επιλογής μεταβλητών στο πλήρες μοντέλο; Συμφωνούν τα αποτελέσματα για όλες τις δραστηριότητες;

10. Θέλουμε να διερευνήσουμε αν η τιμή του P10 για κάποια χώρα και δραστηριότητα σε κάποιο έτος συσχετίζεται με την αντίστοιχη τιμή σε προηγούμενο έτος ή προηγούμενα έτη.

Για αυτό θα κάνετε μια **συνάρτηση** που θα παίρνει ως είσοδο ένα διάγραμμα της τιμής P10 για κάποια χώρα και δραστηριότητα την περίοδο 1990-2007, δηλαδή μια χρονοσειρά. Η συνάρτηση θα δίνει πρώτα το διάγραμμα ιστορίας της χρονοσειράς. Μετά θα ρωτά το χρήστη αν θέλει να απαλείψει την τυχόν τάση δίνοντας το βαθμό του πολυωνυμικού μοντέλου του χρόνου (βαθμός μηδέν σημαίνει ότι δε γίνεται απαλοιφή τάσης με κάποιο πολυωνυμικό μοντέλο). Αυτό γίνεται με χρήση της συνάρτησης `input` του Matlab. Θα γίνεται απαλοιφή της τάσης με το πολυωνυμικό μοντέλο του χρόνου του δεδομένου βαθμού (ή καμιά απαλοιφή αν ο βαθμός είναι μηδέν). Στη συνέχεια η συνάρτηση θα παίρνει τη χρονοσειρά των υπολοίπων (μετά την απαλοιφή της τάσης) ή την αρχική χρονοσειρά (αν ο χρήστης δεν επέλεξε απαλοιφή της τάσης) και θα υπολογίζει την αυτοσυσχέτιση για ένα σύνολο υστερήσεων π.χ. ως την υστέρηση 5. Με βάση τα όρια σημαντικότητας της αυτοσυσχέτισης ($\pm 2/\sqrt{n}$) θα αποφασίζει αν υπάρχουν αυτοσυσχετίσεις ή θα θεωρείται η χρονοσειρά λευκός θόρυβος, δηλαδή χωρίς συσχετίσεις. Η συνάρτηση θα δίνει στην έξοδο αν η χρονοσειρά έχει τάση και αν η χρονοσειρά (ενδεχομένως μετά την απαλοιφή της τάσης) έχει αυτοσυσχέτιση.

Θα κάνετε ένα πρόγραμμα που θα καλεί τη συνάρτηση για 10 τυχαία επιλεγμένες χρονοσειρές (διαφορετική δραστηριότητα και χώρα) και θα

αναφέρει σε ποιες από τις 10 χρονοσειρές βρέθηκε να υπάρχει τάση και στατιστικά σημαντική αυτοσυσχέτιση.