

## Εργασία στο Μάθημα Ανάλυση Δεδομένων

---

Κίμων Ζαγκούρης  
Α.Ε.Μ. 4353  
Email: kzagko@gmail.com

11 Φεβρουαρίου 2019



1

---

<sup>1</sup>Image created by Timashov Sergiy.

## Περιεχόμενα

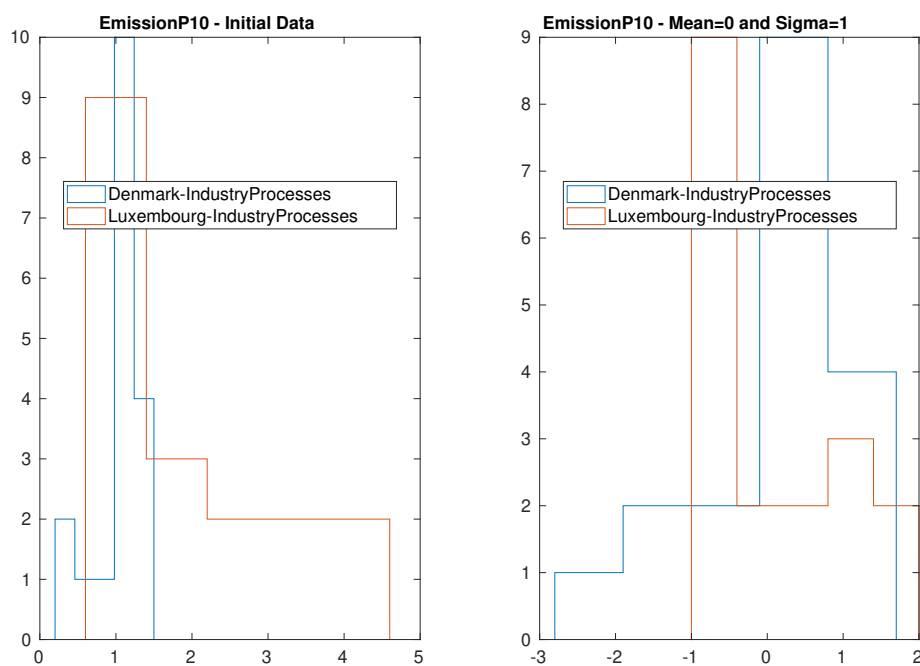
<b>1</b>	<b>Ζητήματα</b>	<b>3</b>
1.1	Ζήτημα 1 . . . . .	3
1.2	Ζήτημα 2 . . . . .	6
1.3	Ζήτημα 3 . . . . .	8
1.4	Ζήτημα 4 . . . . .	12
1.5	Ζήτημα 5 . . . . .	16
1.6	Ζήτημα 6 . . . . .	21
1.7	Ζήτημα 7 . . . . .	25
1.8	Ζήτημα 8 . . . . .	25
1.9	Ζήτημα 9 . . . . .	25
1.10	Ζήτημα 10 . . . . .	25
<b>2</b>	<b>Προγράμματα</b>	<b>25</b>
2.1	Exercise1 . . . . .	25
2.2	DataLoader . . . . .	27
2.3	Chartme . . . . .	28
2.4	Exercise2 . . . . .	29
2.5	Exercise3 . . . . .	32
2.6	CIs . . . . .	33
2.7	MeanP10 . . . . .	34
2.8	Exercise4 . . . . .	34
2.9	stdCIs . . . . .	35
2.10	StdP10 . . . . .	36
2.11	Exercise5 . . . . .	37
2.12	AllCountries . . . . .	38
2.13	corrme . . . . .	39
2.14	Exercise6 . . . . .	39
2.15	AllActivities . . . . .	41
<b>3</b>	<b>Understanding Text</b>	<b>42</b>
3.0.1	Suppose ``chuck" implies throwing. . . . .	42
3.0.2	Suppose ``chuck" implies vomiting. . . . .	42
<b>4</b>	<b>Interpreting Equations</b>	<b>42</b>
4.1	Identify the author of Equation 4.1 below and briefly describe it in English. . . . .	42
4.2	Try to make sense of some more equations. . . . .	42
<b>5</b>	<b>Viewing Lists</b>	<b>43</b>
5.1	Bullet Point List . . . . .	43
5.2	Numbered List . . . . .	43
<b>6</b>	<b>Interpreting a Table</b>	<b>43</b>
6.1	The table above shows the nutritional consistencies of two sausage types. Explain their relative differences given what you know about daily adult nutritional recommendations. . . . .	43
<b>7</b>	<b>Reading a Code Listing</b>	<b>43</b>
7.1	How many luftballons will be output by the Listing 16 above? . . . . .	44
7.2	Identify the regular expression in Listing 16 and explain how it relates to the anti-war sentiments found in the rest of the script. . . . .	44

# 1 Ζητήματα

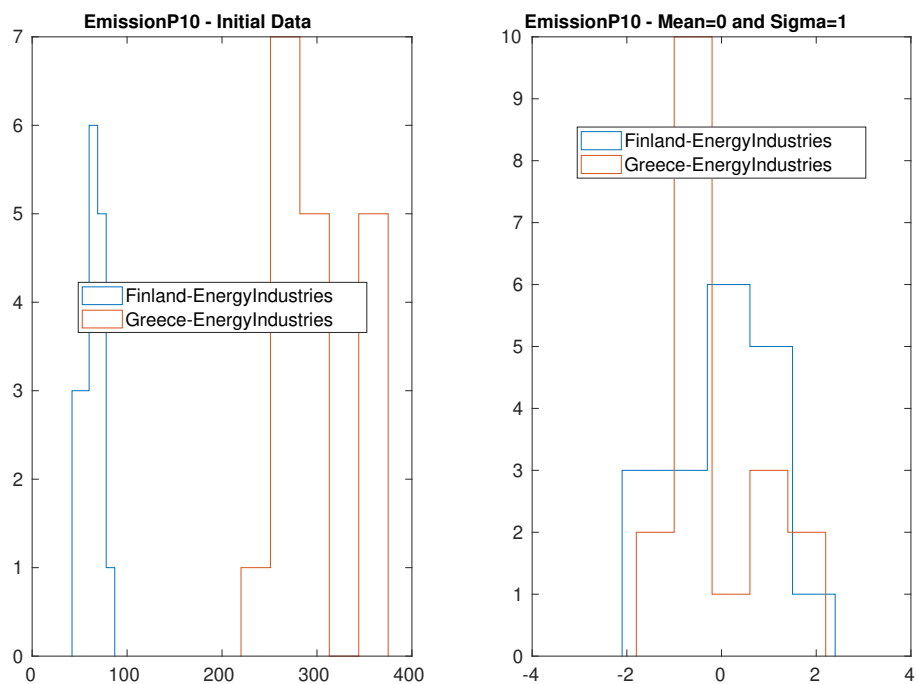
Παρακάτω είναι τα 10 ζητήματα της εργασίας αυτής ανά κεφάλαιο. Στο Κεφάλαιο 2 υπάρχουν τα αντίστοιχα προγράμματα Matlab που χρησιμοποιήθηκαν για την επίλυση των ζητημάτων.

## 1.1 Ζήτημα 1

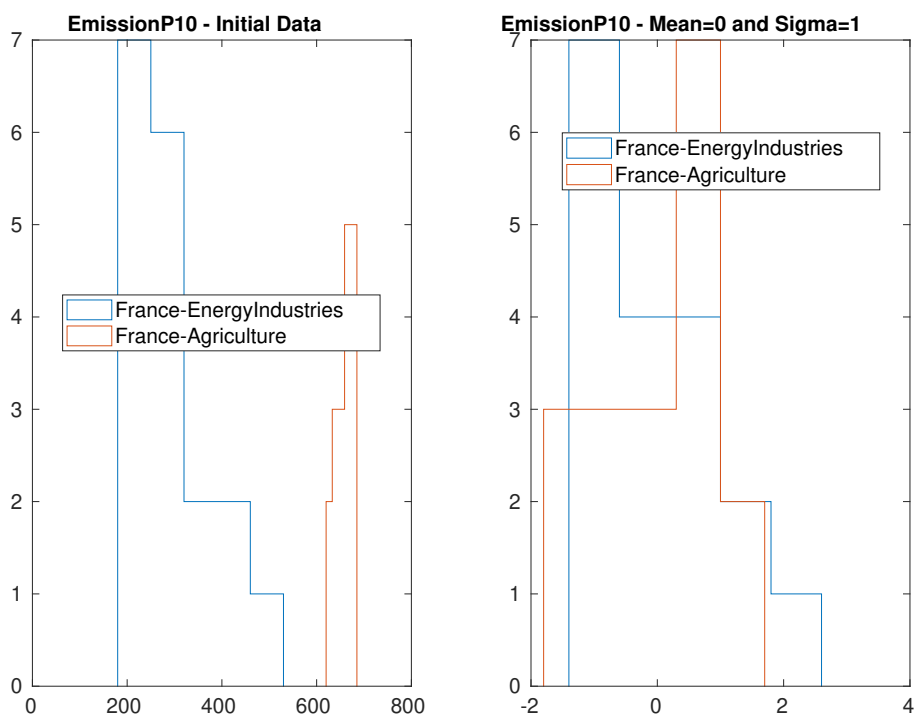
Στο ζήτημα αυτό τρέχουμε τον παρακάτω Κώδικα για 6 περιπτώσεις και παραθέτουμε τα γραφήματα.



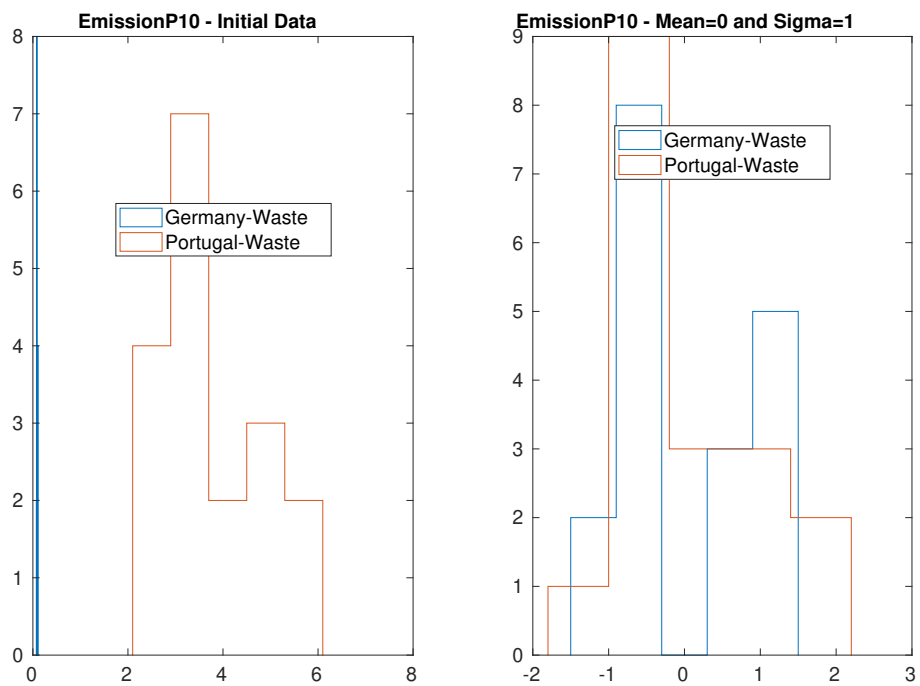
Σχήμα 1.1: PM10 Emissions for Denmark Luxembourg IndustryProcesses.



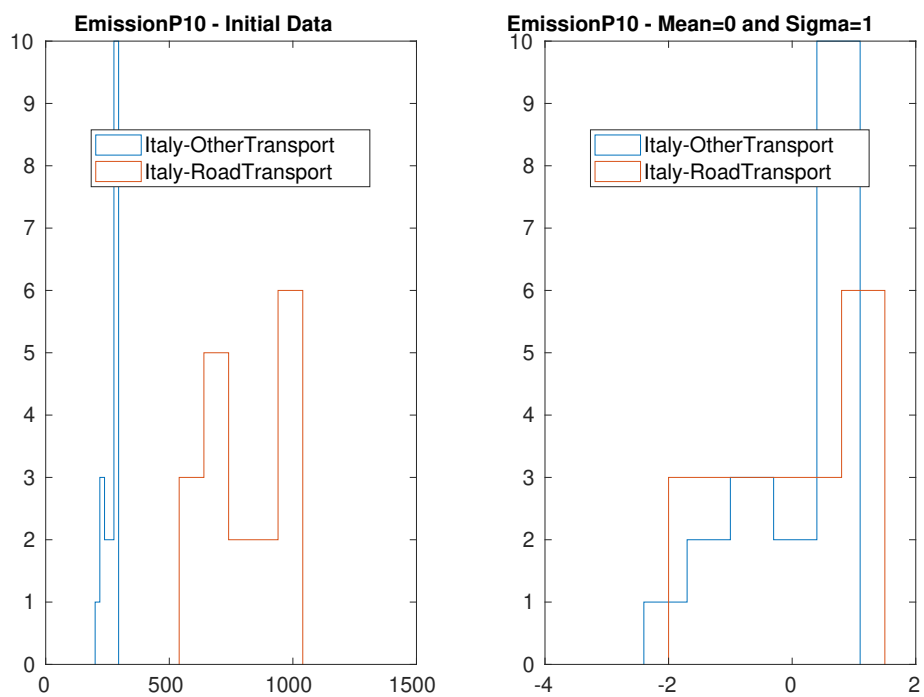
Σχήμα 1.2: PM10 Emissions for Finland Greece EnergyIndustries.



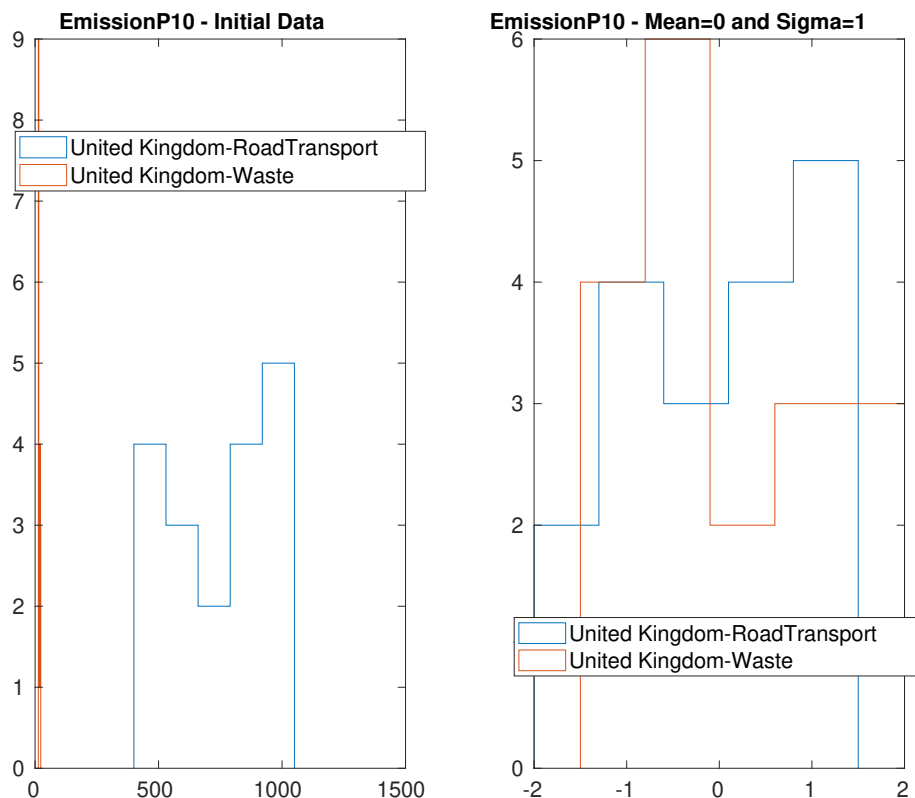
Σχήμα 1.3: PM10 Emissions for France EnergyIndustries Agriculture.



Σχήμα 1.4: PM10 Emissions for Germany Portugal Waste.



Σχήμα 1.5: PM10 Emissions for Italy OtherTransport RoadTransport.

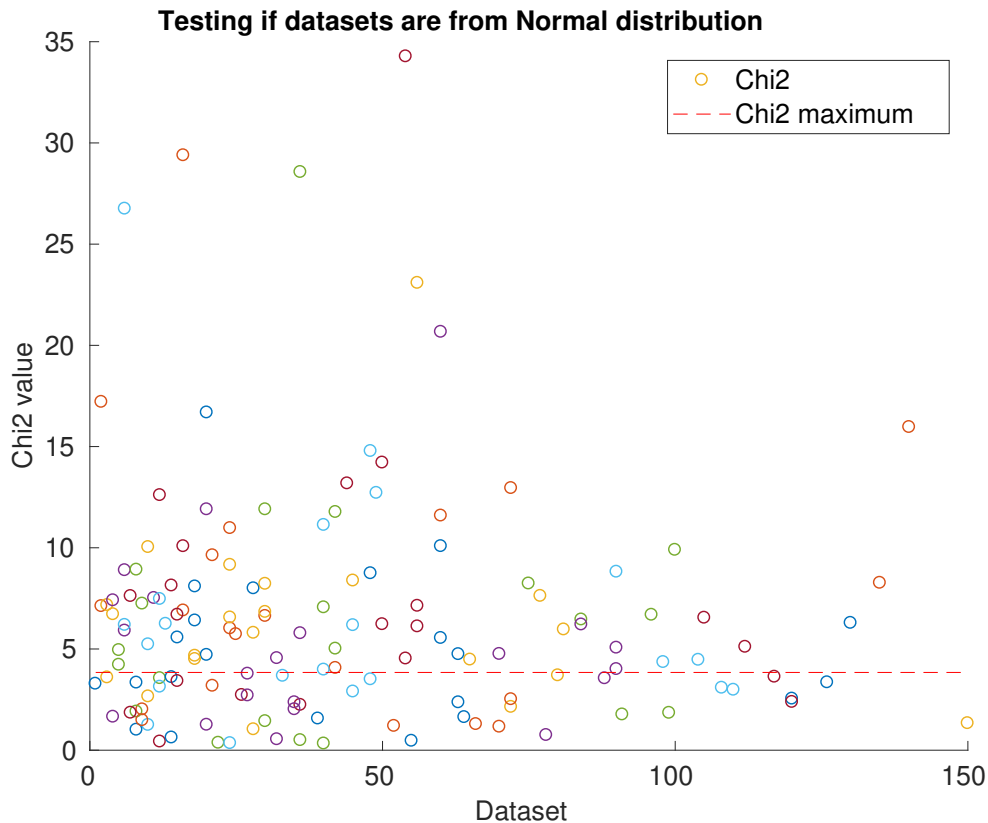


Σχήμα 1.6: PM10 Emissions for United Kingdom RoadTransport Waste.

Τα παραπάνω αποτελούν ένα τυχαίο δείγμα από όλα τα δυνατά αποτελέσματα και συνδυασμούς των χωρών και των δραστηριοτήτων αλλά μπορούμε να σχολιάσουμε μερικά από αυτά. Στο σχήμα 1.1 βλέπουμε ότι οι κατανομές των PM10 πριν την κανονικοποίηση δεν είναι εύκολο να συγκριθούν και να ελεγχθούν αλλά μετά την κανονικοποίηση μπορούμε να πούμε ότι φαίνεται να μην έχουν κοινή κατανομή λόγω του σχήματός τους. Ενώ αντίθετα στο σχήμα 1.1 Για την κατανομή των PM10 στην Ιταλία βλέπουμε ότι όταν κανονικοποιούμε τις κατανομές φαίνεται αυτές να μοιάζουν. Γενικά όμως τα δεδομένα είναι λίγα, στην καλύτερη περίπτωση 18 σημεία, τα οποία όταν κατανέμονται σε ιστόγραμμα δίνουν πολύ μικρό αριθμό στοιχείων, περίπου 4, ανά περιοχή κατανομής (bin). Με αποτέλεσμα να είναι δύσκολο να έχουμε καλή περιγραφή της κάθε κατανομής.

## 1.2 Ζήτημα 2

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 4. Λόγου του μικρού αριθμού δεδομένων η συνάρτηση `chi2gof` του Matlab δεν επιστρέφει αξιόπιστα αποτελέσματα για τον υπολογισμό του  $\chi^2$ . Για τον λόγο αυτό στο παραπάνω πρόγραμμα επιλέξαμε να κάνουμε αναλυτική μέτρηση του  $\chi^2$ . Μετά την εκτέλεση το πρόγραμμα μας επιστρέφει τα παρακάτω σχήματα 1.2 και 1.2.



Σχήμα 1.7: Αποτελέσματα  $\chi^2$  για κάθε δραστηριότητα και χώρα καθώς και το μέγιστο όριο με διακεκομμένη γραμμή.

Από το σχήμα 1.2 παρατηρούμε ότι αρκετά  $\chi^2$  βρίσκονται επάνω από το μέγιστο  $\chi^2_{1-\alpha, n-3}$ . Για τους βαθμούς ελευθερίας χρησιμοποιούμε  $n-3$  γιατί συγκρίνουμε με την συνεχή κανονική κατανομή που έχει δύο ακόμα δεσμευτικές παραμέτρους, την μέση τιμή και την απόκλιση. Το πρόγραμμα επίσης μας επιστρέφει αν βρήκε μια κατανομή να είναι κανονική ή όχι με βάση κάποιο όριο πιθανότητας. Παρακάτω είναι τα αποτελέσματα αυτού του ελέγχου.

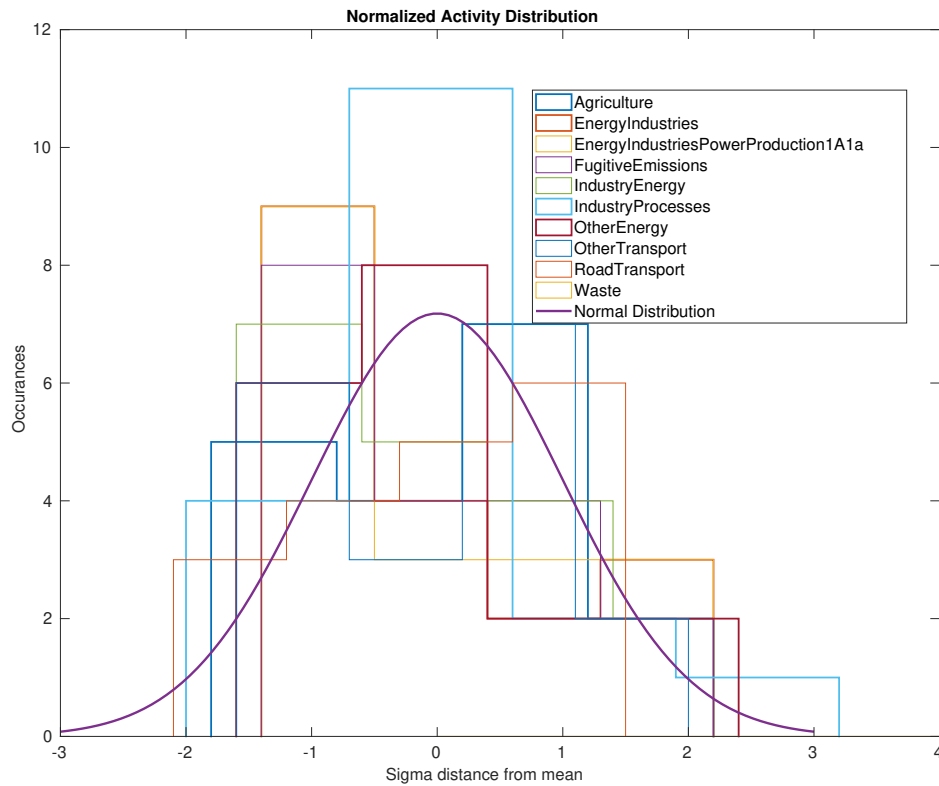
At the significance level of 5.00% the percentage of the datasets that could be from a Normal distribution are 37.33%

At the confidence level of 5.00%

```
Does Agriculture Follow a normal distribution : YES
Does EnergyIndustries Follow a normal distribution : YES
Does EnergyIndustriesPowerProduction1Ala Follow a normal distribution : NO
Does FugitiveEmissions Follow a normal distribution : NO
Does IndustryEnergy Follow a normal distribution : NO
Does IndustryProcesses Follow a normal distribution : YES
Does OtherEnergy Follow a normal distribution : YES
Does OtherTransport Follow a normal distribution : NO
Does RoadTransport Follow a normal distribution : NO
Does Waste Follow a normal distribution : NO
```

Παρατηρούμε ότι στο επίπεδο σημαντικότητας 5% περίπου το 37% από όλα τα σετ θα μπορούσε

να θεωρηθεί ότι προέρχεται από κανονική κατανομή. Επίσης στο ίδιο ποσοστό σημαντικότητας βλέπουμε ότι μόνο 4 στις 10 δραστηριότητες θα μπορούσαν να προέρχονται από κανονική κατανομή.



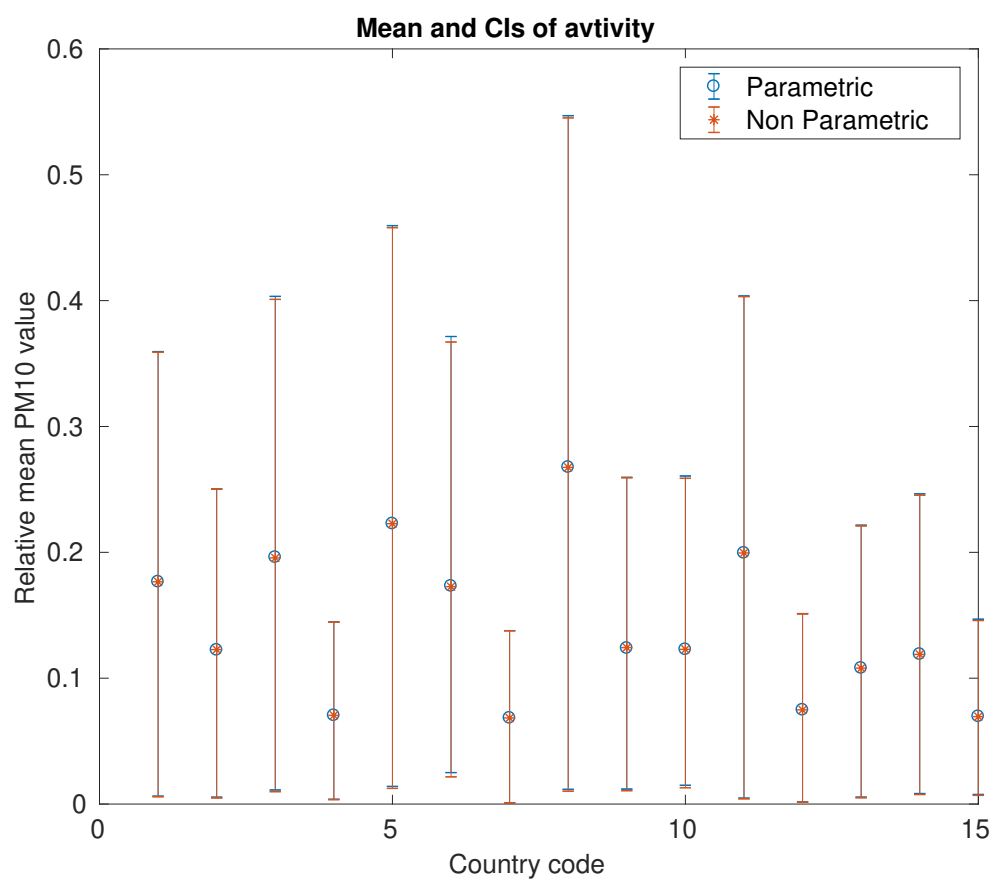
Σχήμα 1.8: Κανονικοποιημένες συνολικές κατανομές για όλες τις δραστηριότητες καθώς και ενδεικτική κανονική κατανομή για σύγκριση. Οι δραστηριότητες με ποιο έντονη γραμμή έχουν περάσει το τεστ κανονικότητας.

Στο σχήμα 1.2 παρατηρούμε ότι γενικά οι κανονικοποιημένες κατανομές των δραστηριοτήτων διαφέρουν από κανονικές. Πέρα από τα αποτελέσματα που μας επιστρέφει το πρόγραμμα θα μπορούσε κάποιος οπτικά και μόνο να πει ότι από όλες τις κατανομές αυτές που φαίνεται να ακολουθούν κανονική κατανομή είναι: IndustryProcesses και OtherEnergy.

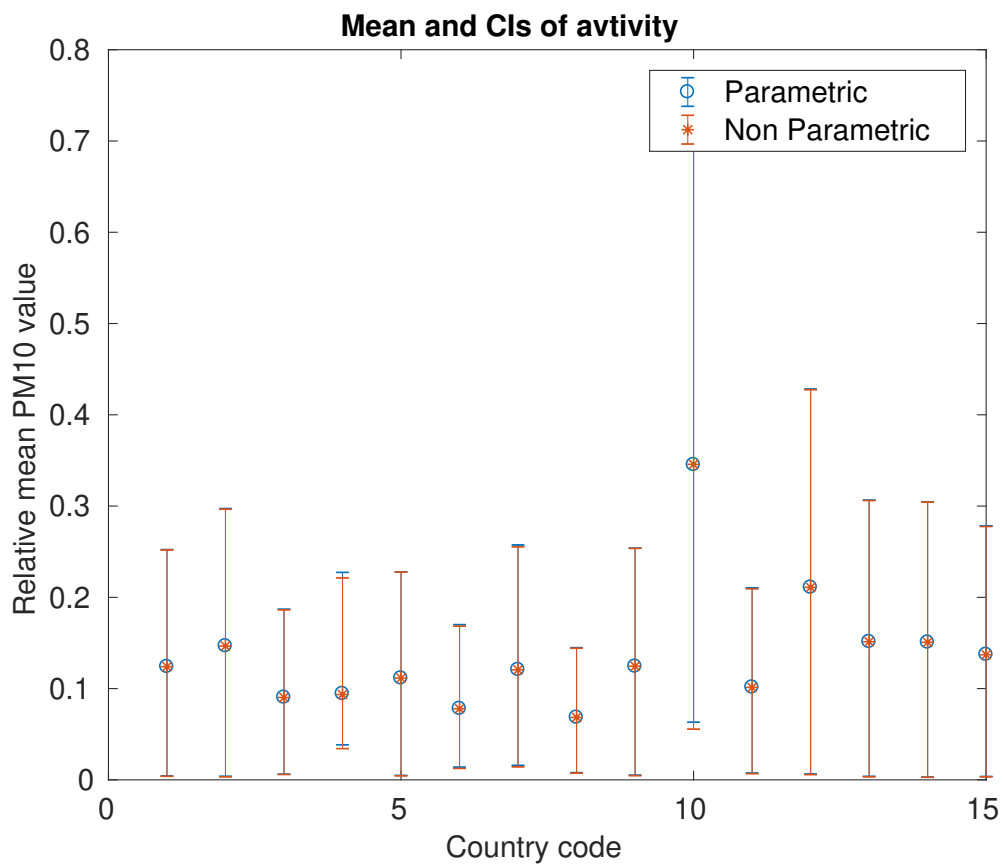
### 1.3 Ζήτημα 3

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 5. Ενδεικτικά εκτελέσαμε τον κώδικα για 4 δραστηριότητες και τα αποτελέσματα φαίνονται στα σχήμα 1.3, σχήμα 1.3, σχήμα 1.3 και σχήμα 1.3.

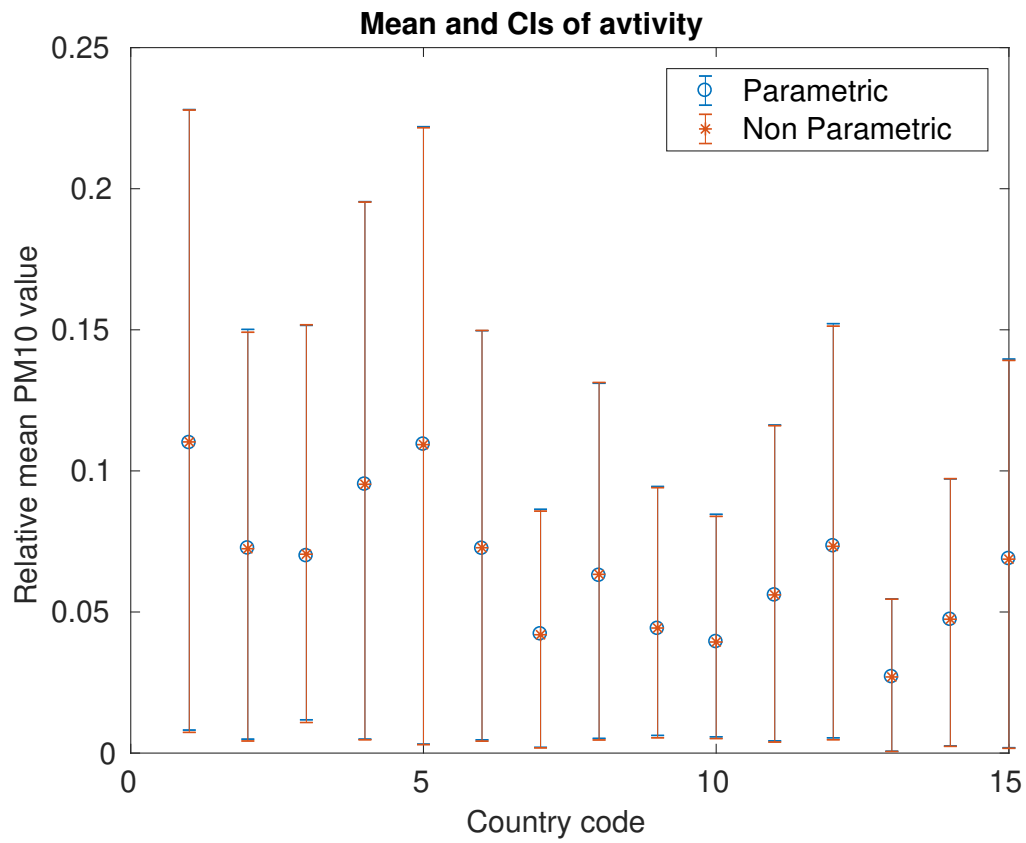




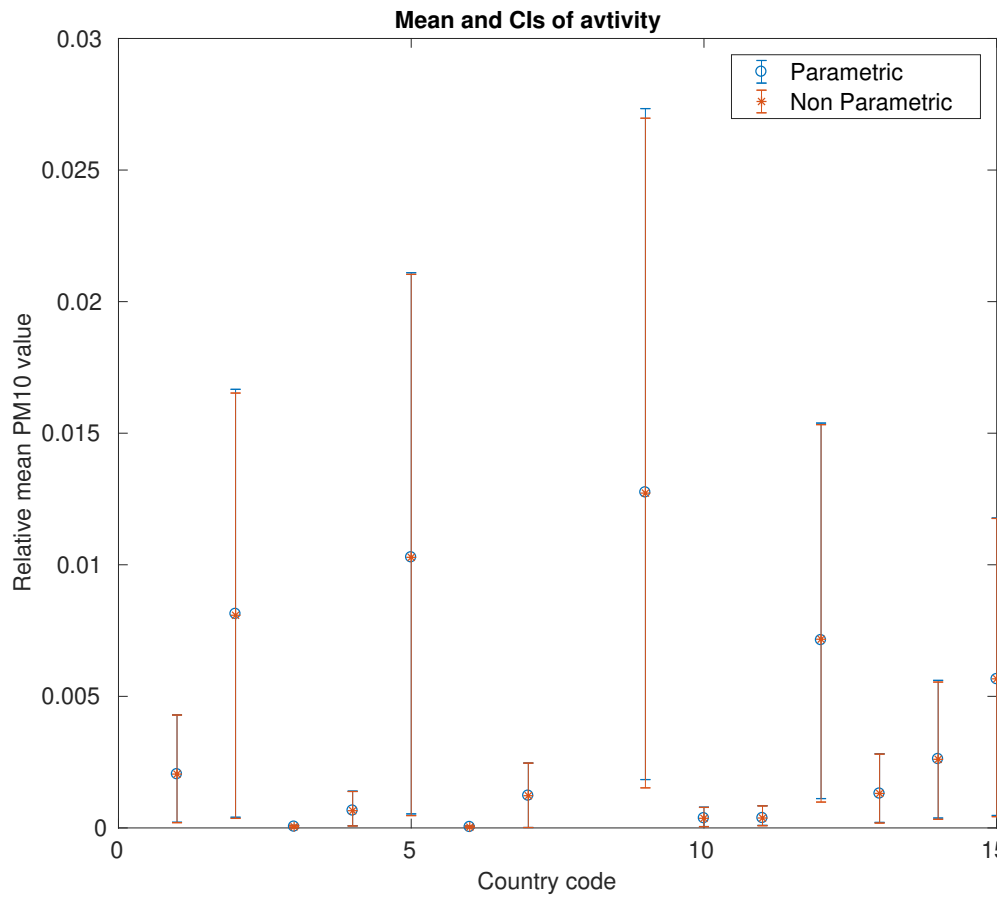
Σχήμα 1.9: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα Agriculture σε κάθε χώρα.



Σχήμα 1.10: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα IndustryEnergy σε κάθε χώρα.



Σχήμα 1.11: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα OtherEnergy σε κάθε χώρα.

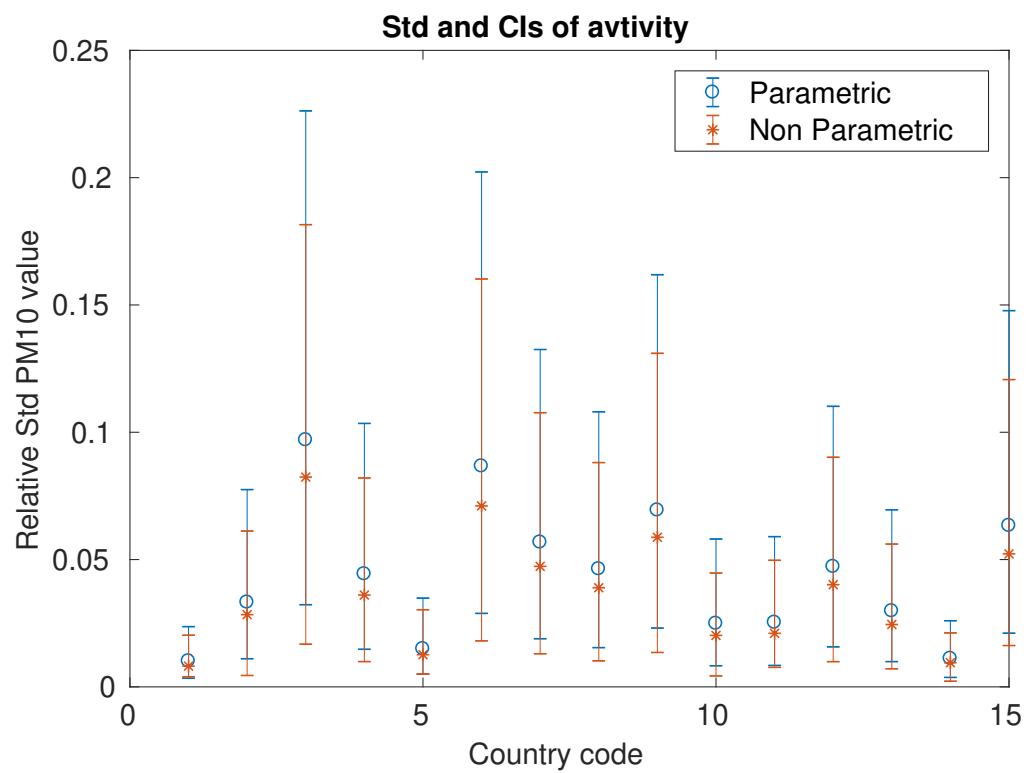


Σχήμα 1.12: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα Waste σε κάθε χώρα.

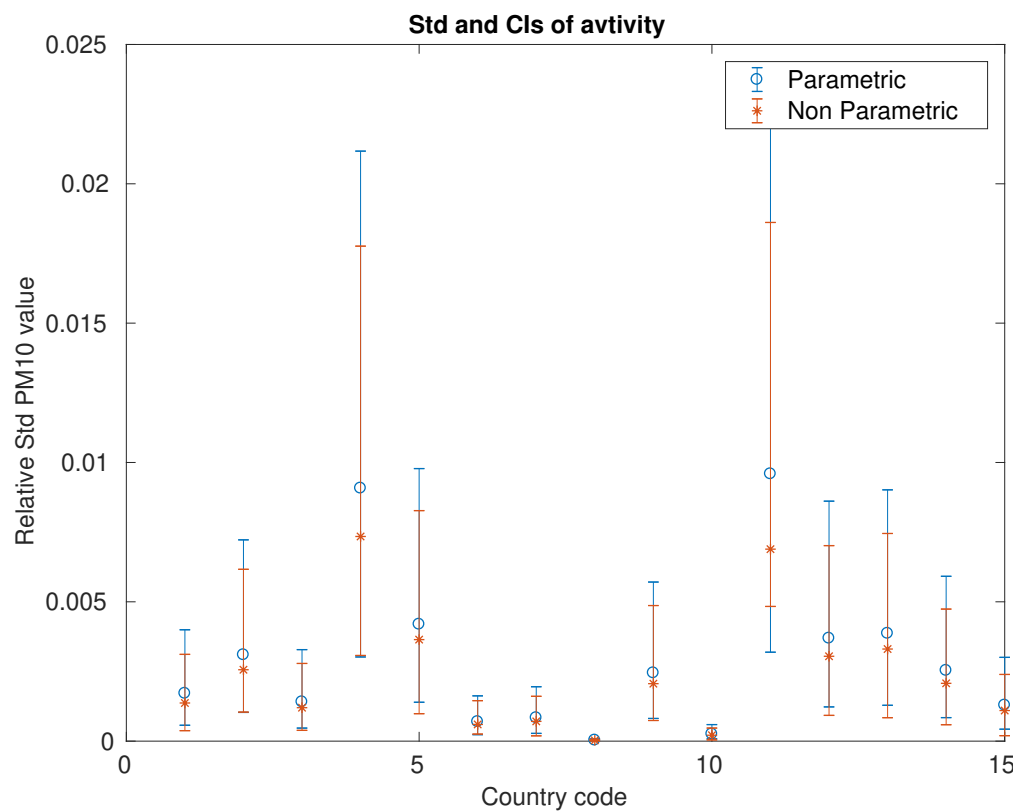
Από τα παραπάνω σχήματα δεν φαίνεται να υπάρχει μεγάλη διαφορά μεταξύ των μέσων τιμών και των διαστημάτων εμπιστοσύνης που προέρχονται από παραμετρική και με παραμετρική εκτίμηση.

#### 1.4 Ζήτημα 4

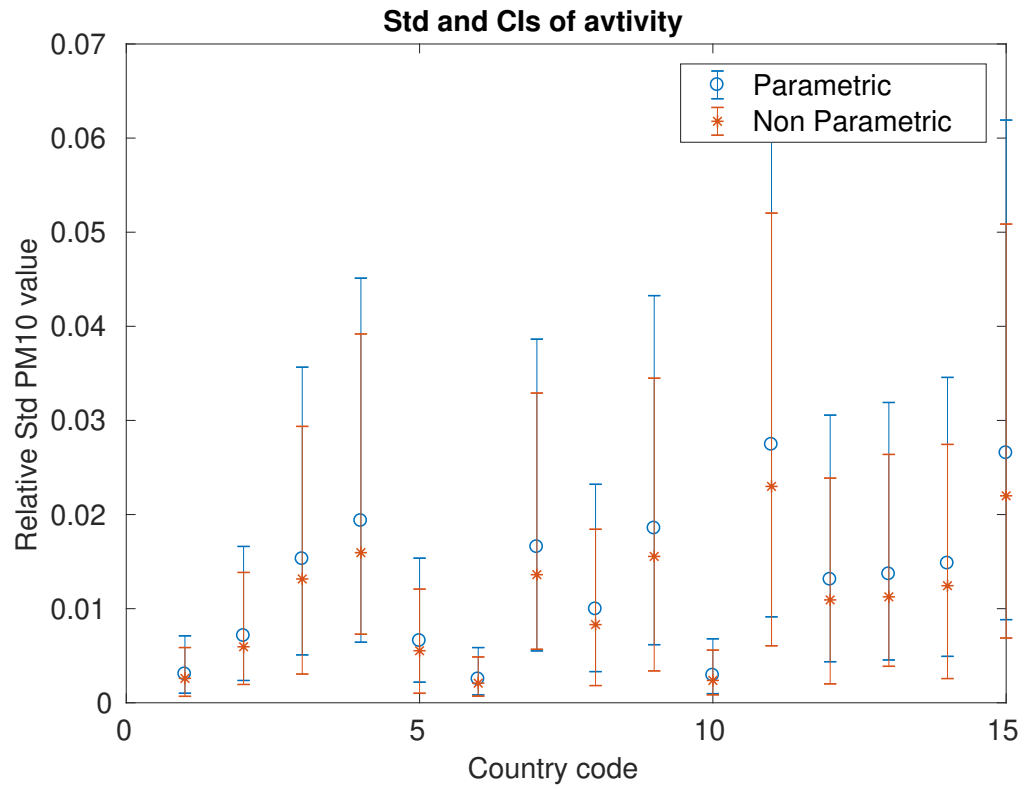
Η ίδια ανάλυση όπως και στο Κεφάλαιο ?? αλλά τώρα για την τυπική απόκλιση. Η ανάλυση γίνεται με τον κώδικα 8. Καλέσαμε τον παραπάνω κώδικα ενδεικτικά για μερικές δραστηριότητες και πήραμε τα παρακάτω σχήμα 1.4, σχήμα 1.4, σχήμα 1.4 και σχήμα 1.4.



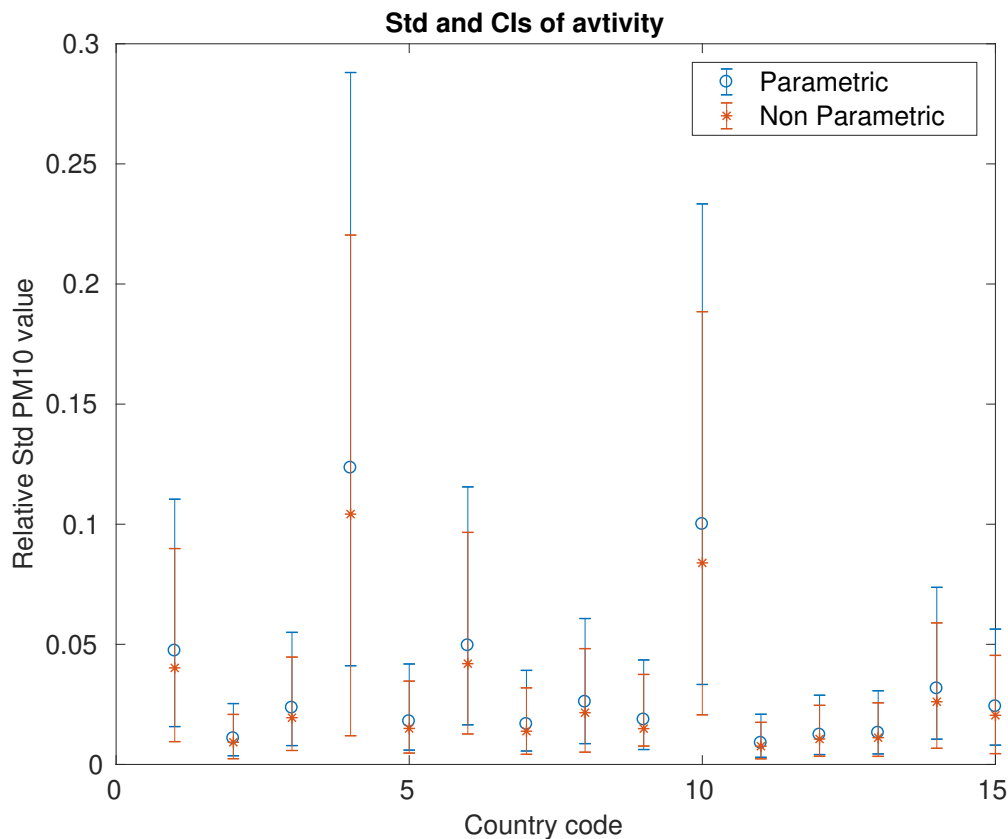
Σχήμα 1.13: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα EnergyIndustries σε κάθε χώρα.



Σχήμα 1.14: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα FugitiveEmissions σε κάθε χώρα.



Σχήμα 1.15: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα Othertransport σε κάθε χώρα.



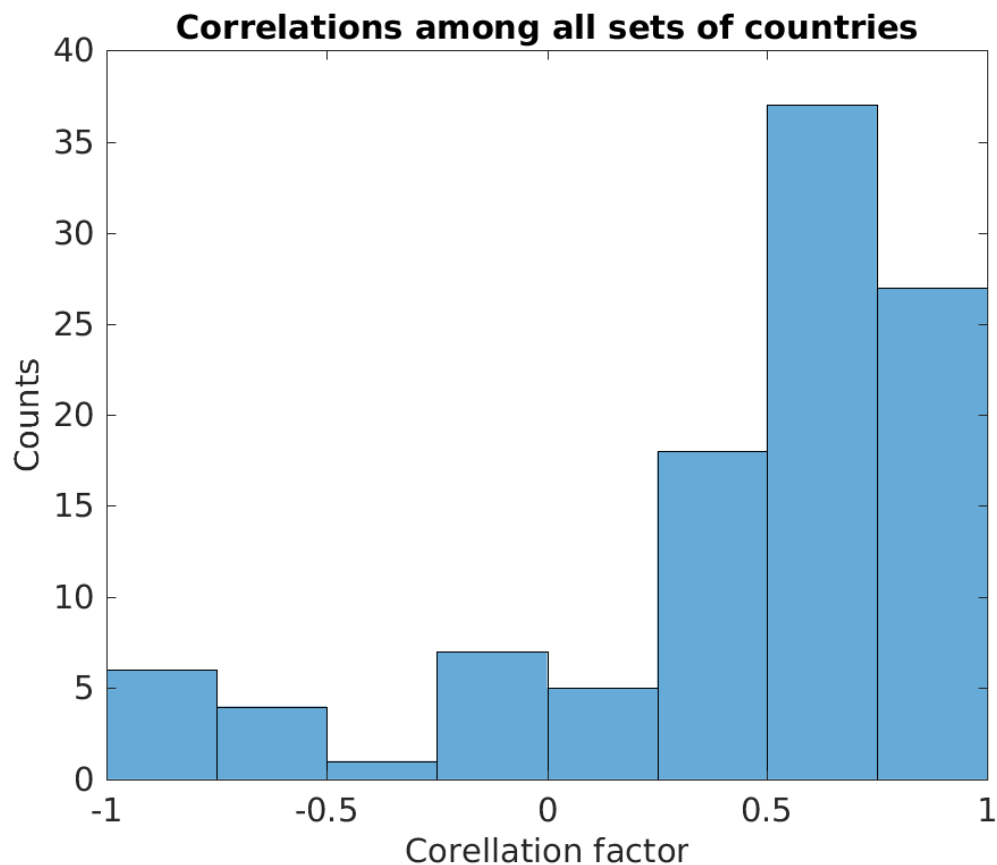
Σχήμα 1.16: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα RoadTransport σε κάθε χώρα.

Από τα παραπάνω σχήματα παρατηρούμε ότι η μέσες τιμές της τυπικής απόκλισης διαφέρουν ανάμεσα στον παραμετρικό και μη τρόπο υπολογισμού τους. Καθώς και ότι σε γενικές γραμμές τα μη παραμετρικά διαστήματα εμπιστοσύνης είναι μικρότερα από τα αντίστοιχα παραμετρικά. Αυτό μπορεί να οφείλεται στον μεγάλο αριθμό (1000) τυχαίων δειγμάτων που δημιουργήσαμε στην μη παραμετρική ανάλυσή μας.

## 1.5 Ζήτημα 5

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 11. Κατά την διάρκεια εκτέλεσης του κώδικα το Matlab επιστρέφει ένα γράφημα και τα αποτελέσματα του τεστ στην οθόνη. Ενδεικτικά καλέσαμε τον παραπάνω κώδικα για 4 δραστηριότητες και τα αποτελέσματα είναι παρακάτω.





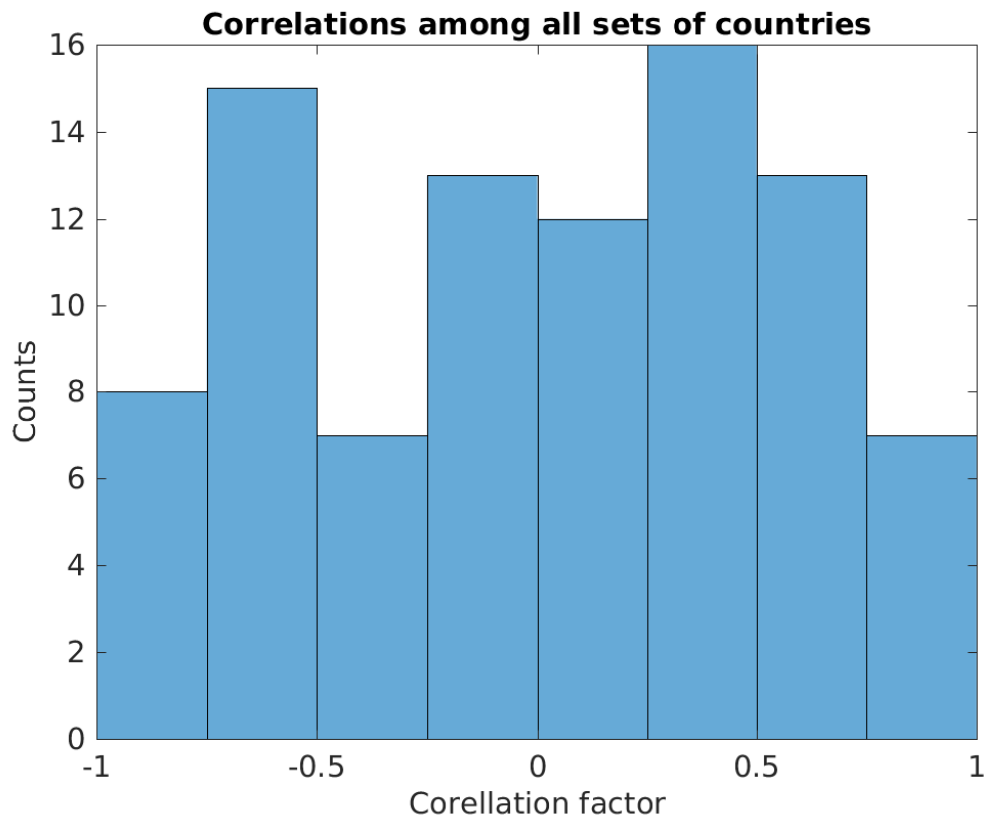
Σχήμα 1.17: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας Agriculture.

The 10 sets of countries with the highest correlation for Agriculture are

Correlation	Country1	Country2
0.98	Denmark	Italy
0.97	Netherlands	United Kingdom
0.95	Denmark	Netherlands
0.95	Italy	United Kingdom
0.94	Italy	Netherlands
0.94	Austria	Belgium
0.94	Denmark	United Kingdom
0.91	Belgium	United Kingdom
0.91	Austria	Luxembourg
0.91	Austria	United Kingdom

29.52% of the sets of countries pass the Parametric test with a significance level of 5.00%

25.71% of the sets of countries pass the NON-Parametric test with a significance level of 5.00%



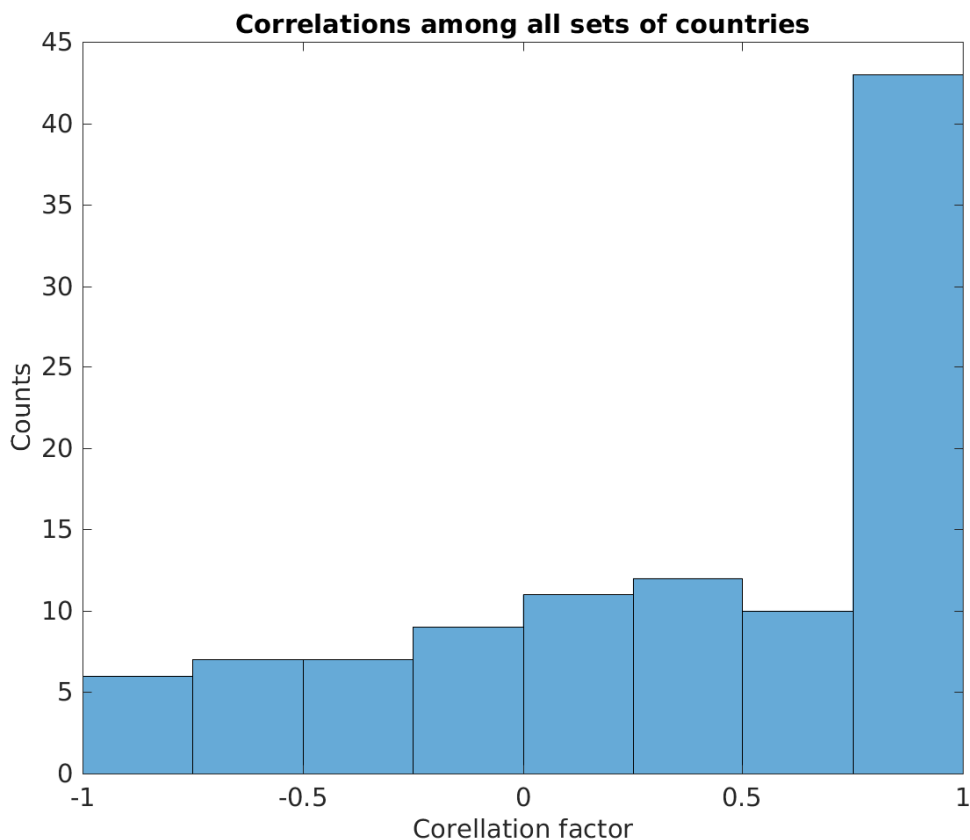
Σχήμα 1.18: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας Waste.

The 10 sets of countries with the highest correlation for Waste are

Correlation	Country1	Country2
NaN	Austria	Ireland
NaN	Belgium	Ireland
NaN	Denmark	Ireland
NaN	Finland	Ireland
NaN	France	Ireland
NaN	Germany	Ireland
NaN	Greece	Ireland
NaN	Ireland	Italy
NaN	Ireland	Luxembourg
NaN	Ireland	Netherlands

44.76% of the sets of countries pass the Parametric test with a significance level of 5.00%

57.14% of the sets of countries pass the NON-Parametric test with a significance level of 5.00%



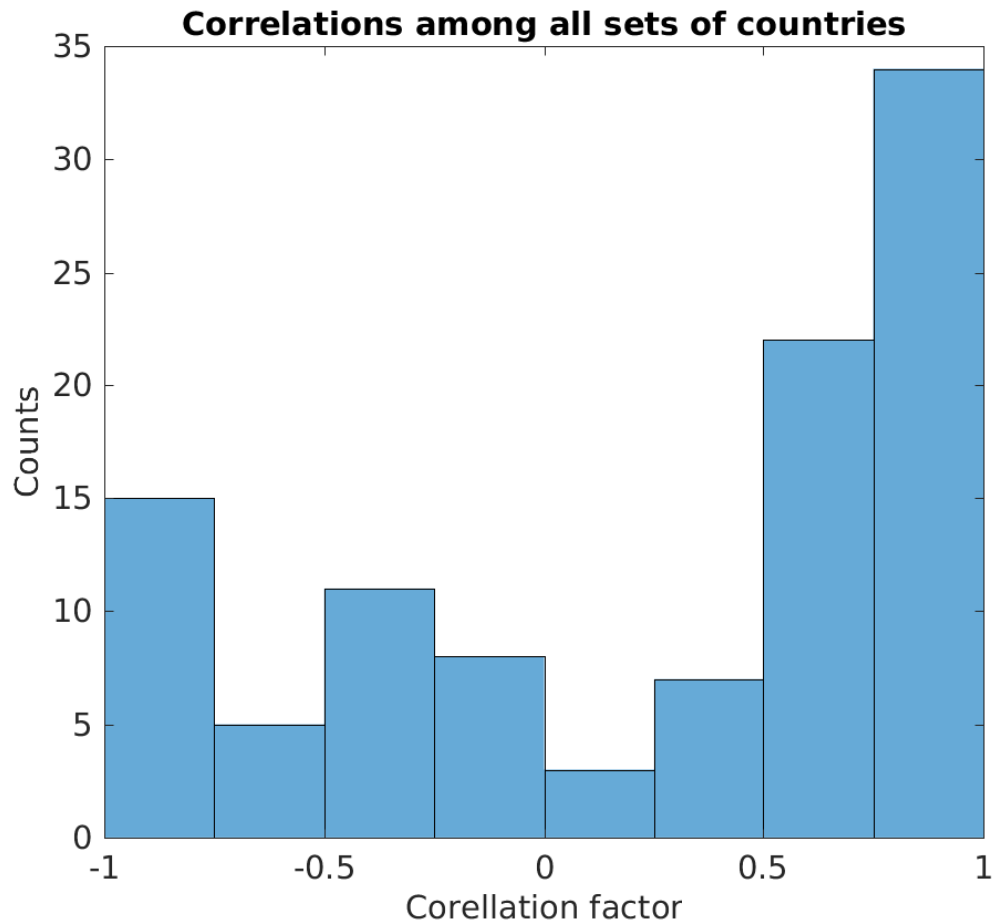
Σχήμα 1.19: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας IndustryEnergy.

The 10 sets of countries with the highest correlation for IndustryEnergy are

Correlation	Country1	Country2
0.99	France	Italy
0.98	Luxembourg	United Kingdom
0.97	Italy	Luxembourg
0.97	France	United Kingdom
0.97	Belgium	United Kingdom
0.97	Italy	United Kingdom
0.97	France	Luxembourg
0.97	Sweden	United Kingdom
0.96	Belgium	Italy
0.96	France	Sweden

35.24% of the sets of countries pass the Parametric test with a significance level of 5.00%

33.33% of the sets of countries pass the NON-Parametric test with a significance level of 5.00%



Σχήμα 1.20: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας RoadTransport.

The 10 sets of countries with the highest correlation for RoadTransport are

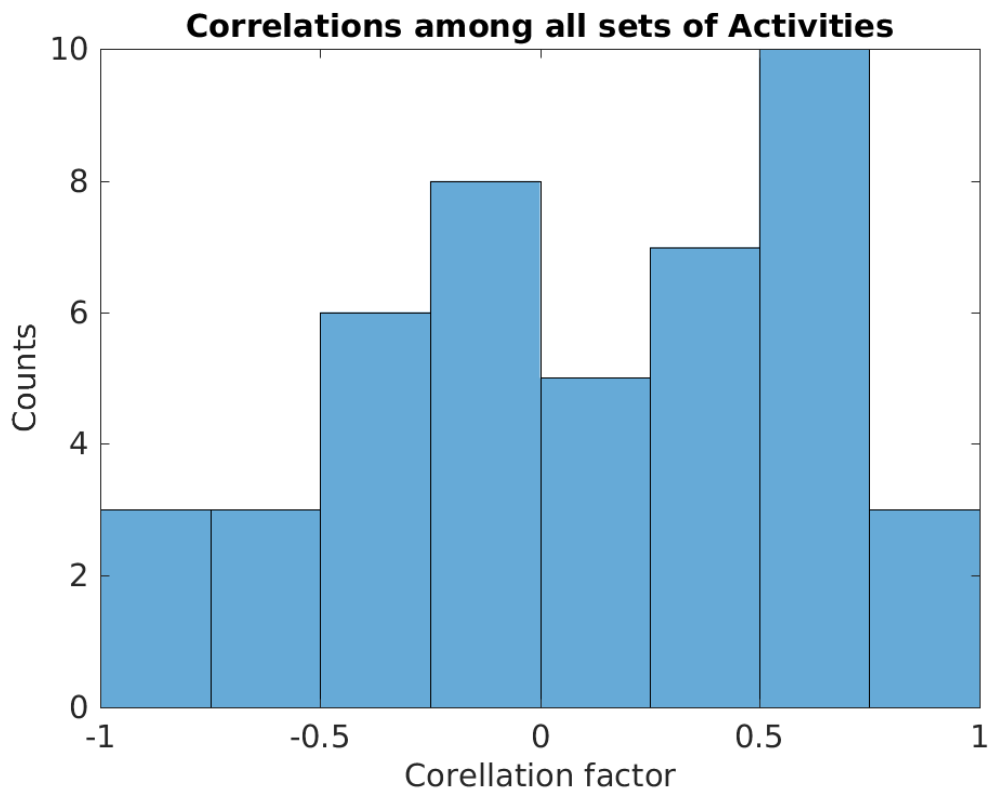
Correlation	Country1	Country2
1.00	Denmark	United Kingdom
0.99	Denmark	France
0.99	Belgium	France
0.99	Denmark	Sweden
0.99	Germany	United Kingdom
0.99	France	United Kingdom
0.99	France	Italy
0.99	Sweden	United Kingdom
0.99	France	Sweden
0.99	Denmark	Italy

25.71% of the sets of countries pass the Parametric test with a significance level of 5.00%

25.71% of the sets of countries pass the NON-Parametric test with a significance level of 5.00%

## 1.6 Ζήτημα 6

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 14.



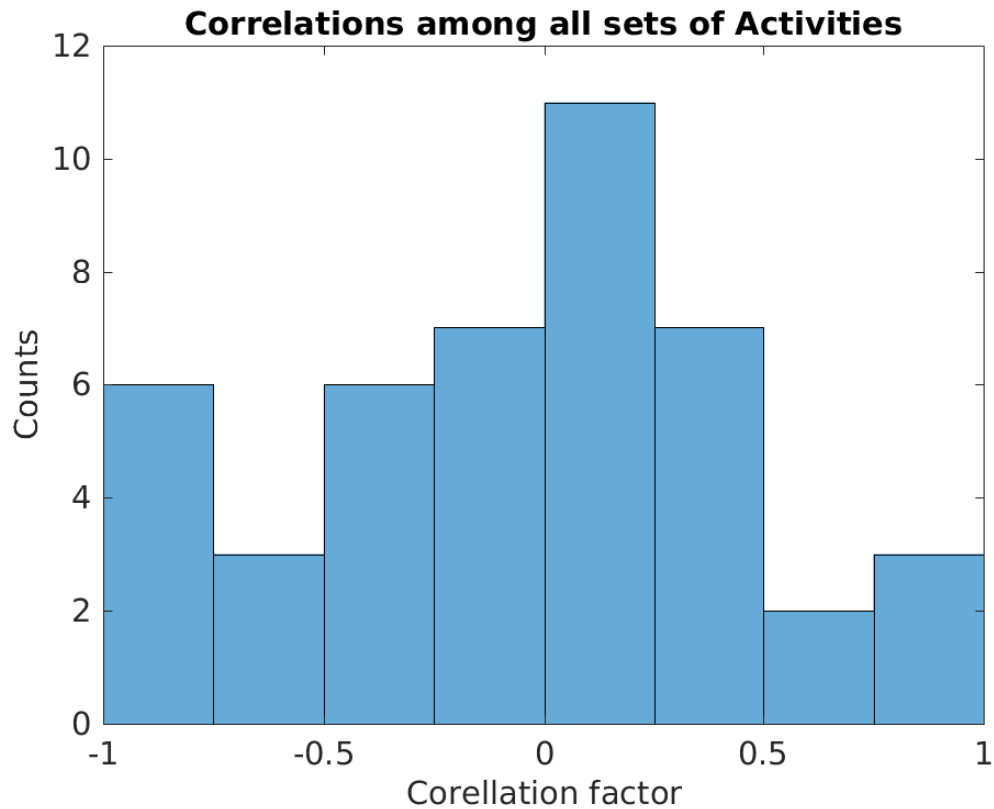
Σχήμα 1.21: Ιστόγραμμα των συντελεστών συσχέτισης για την Austria.

The 10 sets of Activities with the highest correlation for Austria are

Correlation	Activity1	Activity2
1.00	EnergyIndustries	EnergyIndustriesPowerProduction1A1a
0.82	FugitiveEmissions	OtherEnergy
0.76	RoadTransport	Waste
0.75	EnergyIndustriesPowerProduction1A1a	IndustryEnergy
0.74	EnergyIndustriesPowerProduction1A1a	IndustryProcesses
0.74	EnergyIndustries	IndustryProcesses
0.72	EnergyIndustries	IndustryEnergy
0.70	IndustryProcesses	OtherEnergy
0.70	Agriculture	OtherEnergy
0.64	EnergyIndustriesPowerProduction1A1a	OtherEnergy

53.33% of the sets of Activities pass  
the Parametric test with a significance level of 5.00%

53.33% of the sets of Activities pass  
the NON-Parametric test with a significance level of 5.00%



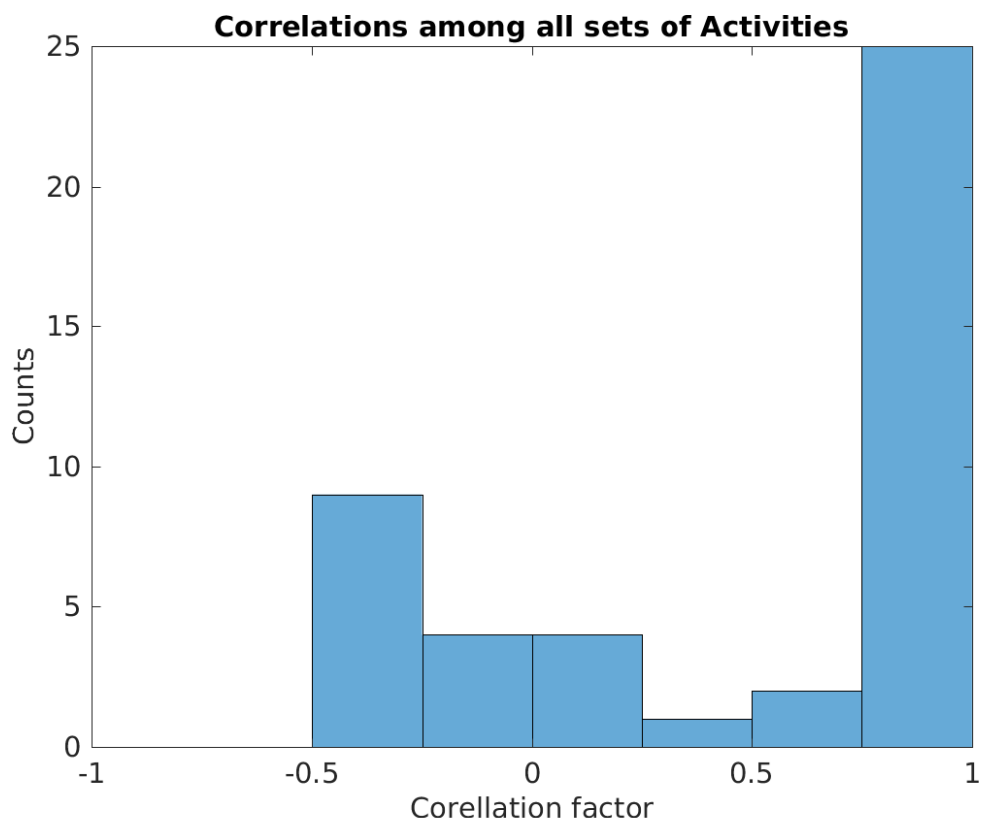
Σχήμα 1.22: Ιστόγραμμα των συντελεστών συσχέτισης για την Spain.

The 10 sets of Activities with the highest correlation for Spain are

Correlation	Activity1	Activity2
0.99	EnergyIndustries	EnergyIndustriesPowerProduction1A1a
0.92	OtherTransport	Waste
0.77	FugitiveEmissions	RoadTransport
0.66	Agriculture	OtherTransport
0.55	EnergyIndustries	FugitiveEmissions
0.48	Agriculture	Waste
0.46	EnergyIndustriesPowerProduction1A1a	FugitiveEmissions
0.44	EnergyIndustries	RoadTransport
0.35	EnergyIndustriesPowerProduction1A1a	RoadTransport
0.32	OtherEnergy	RoadTransport

64.44% of the sets of Activities pass  
the Parametric test with a significance level of 5.00%

62.22% of the sets of Activities pass  
the NON-Parametric test with a significance level of 5.00%



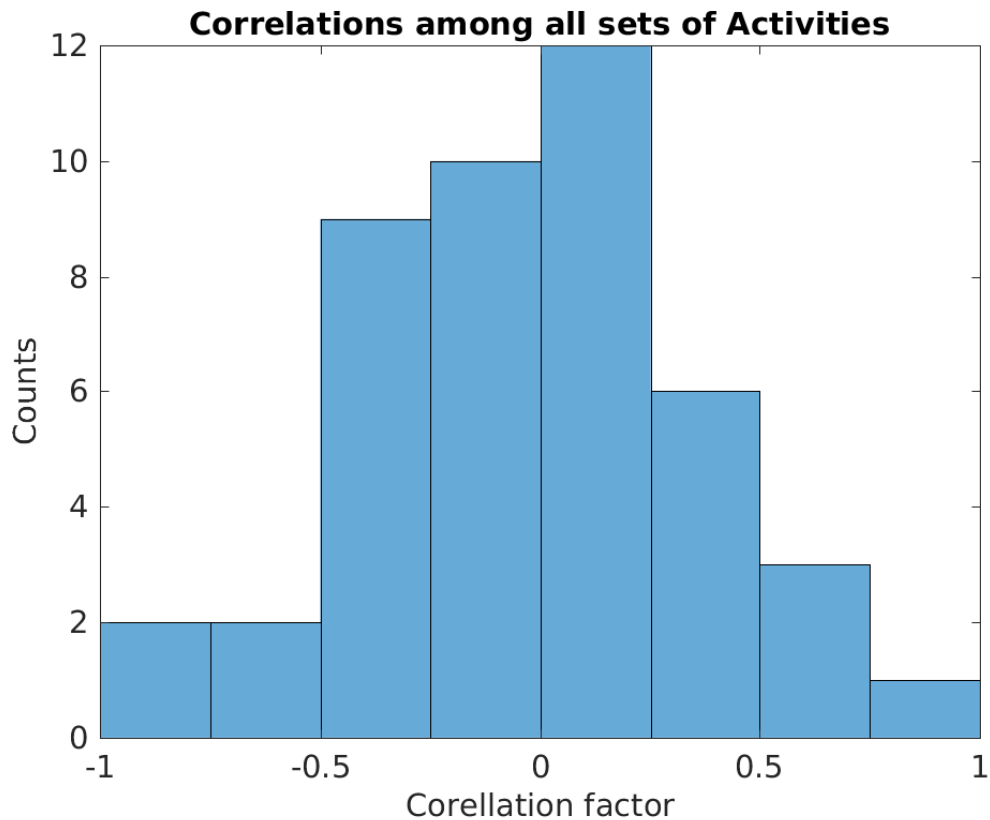
Σχήμα 1.23: Ιστόγραμμα των συντελεστών συσχέτισης για την Italy.

The 10 sets of Activities with the highest correlation for Italy are

Correlation	Activity1	Activity2
1.00	EnergyIndustries	EnergyIndustriesPowerProduction1A1a
0.97	EnergyIndustries	RoadTransport
0.96	EnergyIndustriesPowerProduction1A1a	RoadTransport
0.96	EnergyIndustries	IndustryEnergy
0.96	EnergyIndustriesPowerProduction1A1a	IndustryEnergy
0.95	Agriculture	EnergyIndustries
0.95	Agriculture	EnergyIndustriesPowerProduction1A1a
0.93	Agriculture	RoadTransport
0.92	Agriculture	IndustryEnergy
0.91	IndustryEnergy	RoadTransport

40.00% of the sets of Activities pass  
the Parametric test with a significance level of 5.00%

37.78% of the sets of Activities pass  
the NON-Parametric test with a significance level of 5.00%



Σχήμα 1.24: Ιστόγραμμα των συντελεστών συσχέτισης για την Greece.

The 10 sets of Activities with the highest correlation for Greece are

Correlation	Activity1	Activity2
1.00	EnergyIndustries	EnergyIndustriesPowerProduction1A1a
0.65	IndustryEnergy	IndustryProcesses
0.59	Agriculture	IndustryEnergy
0.54	Agriculture	IndustryProcesses
0.49	IndustryEnergy	RoadTransport
0.45	OtherEnergy	RoadTransport
0.37	Agriculture	RoadTransport
0.32	IndustryEnergy	OtherEnergy
0.32	OtherTransport	RoadTransport
0.29	OtherEnergy	OtherTransport

77.78% of the sets of Activities pass  
the Parametric test with a significance level of 5.00%

75.56% of the sets of Activities pass  
the NON-Parametric test with a significance level of 5.00%



### 1.7 Ζήτημα 7

### 1.8 Ζήτημα 8

### 1.9 Ζήτημα 9

### 1.10 Ζήτημα 10

## 2 Προγράμματα

Στο κεφάλαιο αυτό παραθέτονται οι κώδικες σε γλώσσα Matlab που χρησιμοποιήθηκαν για την επίλυση των παραπάνω ζητημάτων. Σε περίπτωση που κώδικας καλεί επί μέρους συναρτήσεις αυτές αναφέρονται και παραθέτονται.

### 2.1 Exercise1

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.1. Το οποίο με την σειρά του καλεί τα προγράμματα DataLoader και Chartme.

Listing 1: Project2019Ex1.m

```
1 clc ;
2 clear all ;
3
4 %%Preload
5 %find all data and activities
6 filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
7 names = {filelist(:).name};
8 Activities = extractBetween(names(:), "EmissionP10", "EU15");
9 ActLength = length(Activities);
10 %Load a file to get the names of countries and years
11 [data, txt, raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
12 dataname = txt(2,1);
13 years = str2double(txt(2:end,2));
14 Countries = [];
15 for i=1:length(data(1,:)) %get the name of the countries and
    store them
16     Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")"];
17 end
18 Countries = Countries';
19 CountLength = length(Countries);
20
21 %%MAIN program
22 %Prompt user to choose from countries and activities
23 prompt = ['Choose type of plot\n One(1) for plotting DIFFERENT
    quantities'...
24     ' for the same country\n Two(2) for plotting the SAME
    quantity for'...
25     ' different countries\n Please insert 1 or 2\n'];
26
27 while 1
```

```

28     chk1 = int8(input(prompt));
29     if chk1 == 1 | chk1 ==2
30         break;
31     end
32 end
33 fprintf('\n The available countries are\n');
34 for i=1:CountLength
35     fprintf('%d \t %s\n',i,Countries{i});
36 end
37
38 if chk1 ==1
39     prompt2 = ['Choose one country by the number on the left: '
40               ];
41     while 1
42         Count1 = int8(input(prompt2));
43         if isinteger(Count1) & 1 <= Count1 & Count1 <=
44             CountLength
45             break;
46         end
47     end
48
49     fprintf('\n The available activities are \n');
50
51     for i=1:ActLength
52         fprintf('%d \t %s\n',i,Activities{i});
53     end
54
55     prompt3 = ['Choose the first activity by the number on the
56               left: '];
57     while 1
58         Count2 = int8(input(prompt3));
59         if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
60             break;
61         end
62     end
63
64     prompt4 = ['Choose the second activity by the number on the
65               left: '];
66     while 1
67         Count3 = int8(input(prompt4));
68         if isinteger(Count3) & 1 <= Count3 & Count3 <= ActLength
69             break;
70         end
71     end
72
73     %load the data for the country
74     D1 = DataLoader(filelist ,Count2 ,Count1);
75     D2 = DataLoader(filelist ,Count3 ,Count1);

```

```

74     Chartme(Countries(Count1),Activities(Count2),Activities(
75         Count3),D1,D2,...
76         years,chk1)
77 else
78     prompt2 = ['Choose the first country by the number on the
79         left: '];
80     while 1
81         Count1 = int8(input(prompt2));
82         if isinteger(Count1) & 1 <= Count1 & Count1 <=
83             CountLength
84             break;
85     end
86     prompt3 = ['Choose the second country by the number on the
87         left: '];
88     while 1
89         Count2 = int8(input(prompt3));
90         if isinteger(Count2) & 1 <= Count2 & Count2 <=
91             CountLength
92             break;
93     end
94     fprintf('\n The available activities are \n');
95     for i=1:ActLength
96         fprintf('%d \t %s\n',i,Activities{i});
97     end
98     prompt4 = ['Choose one activity by the number on the left: '
99         ];
100    while 1
101        Count3 = int8(input(prompt4));
102        if isinteger(Count3) & 1 <= Count3 & Count3 <= ActLength
103            break;
104        end
105    end
106    %load the data for the activity
107    D1 = DataLoader(filelist,Count3,Count1);
108    D2 = DataLoader(filelist,Count3,Count2);
109
110    Chartme(Countries(Count1),Countries(Count2),Activities(
111        Count3),D1,D2,...
112        years,chk1)
113 end

```

## 2.2 DataLoader

Πρόγραμμα για την επιλογή και εισαγωγή δεδομένων προς ανάλυση στο Matlab

Listing 2: Dataloader.m

```

1 function Dar = DataLoader( filenames , ActCount , Count , chk)
2     %Function that searches through the filenames and returns
3     the data for
4     %a specific country and activity
5     datafile = filenames(ActCount).name;
6
7     [data ,txt ,raw] = xlsread( datafile );
8     if nargin<4 | chk ==0
9         Dar = data (: , Count);
10    else
11        Dar = data ;
12    end

```

## 2.3 Chartme

Πρόγραμμα για την δημιουργία Ιστογραμμάτων για το Ζητούμενο 1.1.

Listing 3: Chartme.m

```

1 function Chartme( StrA , StrB , StrC , D1 ,D2 , years , OpChk , nbins1 )
2
3     %Default bin number follows Sturge 's Law
4
5     if nargin<8
6         nbins1 = round(1+3.322*log10( length( D1) ));
7     end
8     %Plot initial data
9     subplot(1,2,1);
10    histogram(D1,nbins1,'DisplayStyle','stairs')
11    hold on;
12    histogram(D2,nbins1,'DisplayStyle','stairs')
13    if OpChk == 2 %Two countries same activity
14        leg1 = strcat(StrA,"-",StrC);
15        leg2 = strcat(StrB,"-",StrC);
16        legend(leg1,leg2,'Location','Best')
17    else %Same country different activities
18        leg1 = strcat(StrA,"-",StrB);
19        leg2 = strcat(StrA,"-",StrC);
20        legend(leg1,leg2,'Location','Best')
21    end
22    title('EmissionP10 - Initial Data');
23
24    %find and remove mean value and normalize sigma
25    D1 = D1-mean(D1);
26    D2 = D2-mean(D2);
27    D1 = D1./std(D1);
28    D2 = D2./std(D2);
29    subplot(1,2,2);
30    histogram(D1,nbins1,'DisplayStyle','stairs')

```

```

31 hold on;
32 histogram(D2,nbins1,'DisplayStyle','stairs')
33 legend(leg1,leg2,'Location','Best')
34 title('EmissionP10 - Mean=0 and Sigma=1');
35 set(findall(gcf,'-property','FontSize'),'FontSize',14);
36
37 end

```

## 2.4 Exercise2

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.2. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader.

Listing 4: Project2019Ex2.m

```

1 clc;
2 clear all;
3
4 %%Preload
5 %find all data and activities
6 filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
   data files
7 filelist(7)=[];%remove the Nationatotals filename
8 names = {filelist(:).name};
9 Activities = extractBetween(names(:),'EmissionP10','EU15');
10 ActLength = length(Activities);
11 %Load a file to get the names of countries and years
12 [data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
13 dataname = txt(2,1);
14 years = str2double(txt(2:end,2));
15 Countries = [];
16 for i=1:length(data(1,:)) %get the name of the countries and
   store them
17     Countries = [Countries, extractBetween(txt(1,i+2),') - ',' -
        " "];
18 end
19 Countries = Countries';
20 CountLength = length(Countries);
21
22 alpha = 0.05;% Set the significance level Default 5%
23 numbin = 4;
24 SumH0 = 0;
25 degfree = numbin-3;%because continuous distributions with mean
   and sigma and sum npoints
26 chi2limit = chi2inv(1-alpha,degfree);%the upperlimit in chi2
   above which H0 is rejected
27 for i=1:ActLength
28
29     expval=0;
30     for j=1:CountLength
31         dataV = DataLoader(filelist,i,j);

```

```

32     npoints = length(dataV);
33     %dataV = dataV-mean(dataV);
34     %dataV = dataV./std(dataV);
35     %Calculate basic statistics
36     meanV = mean(dataV);
37     stdV = std(dataV);
38     [ncounts, Edges] = histcounts(dataV,numbin);
39
40     for k=1:numbin
41         expval(k) = npoints*(normcdf(Edges(k+1),meanV,stdV)-
42             normcdf(Edges(k),meanV,stdV));
43         %expval(k) = npoints*(normcdf(Edges(k+1))-normcdf(
44             Edges(k)));
45     end
46     xi2 = sum((ncounts-expval).^2./expval);
47     %pval = chi2cdf(xi2,degfree,'upper');
48     if xi2 < chi2limit
49         h = 0;
50     else
51         h = 1;
52     end
53     p1 = scatter(i*j,xi2);
54     hold on;
55     % chi2 = sum((dataV-meanV).^2./dataV);
56     %if i ~=10 & j ~= 7%remove problematic dataset
57     %pd = fitdist(dataV,'Normal');
58     %[h,p,stats] = chi2gof(dataV,'Alpha',alpha,'NBins
59         ',7,'CDF',pd);
60
61     %else
62     % h = 0;
63     %end
64
65     SumH0 = SumH0 + h;
66     %fprintf('%d %d %d\n',i,j,SumH0);
67 end
68
69 p2 = line([1 ActLength*CountLength],[chi2limit chi2limit],'Color
70     ','red','LineStyle','--');
71 title('Testing if datasets are from Normal distribution');
72 ylabel('Chi2 value');
73 xlabel('Dataset');
74 legend([p1,p2],{'Chi2','Chi2 maximum'});
75 set(findall(gcf,'-property','FontSize'),'FontSize',14)
76 hold off;
77 fprintf(['\nAt the significance level of %5.2f%% the percentage
78     of the datasets\n' ...
79     'that could be from a Normal distribution are %5.2f%% \n

```

```

        '],100*alpha , ...
77         100*(1-SumH0/(ActLength*CountLength)));
78
79 fprintf('\nAt the confidence level of %5.2f%%\n',100*alpha);
80 %%Part two
81 figure();
82 DataM = zeros(length(years),ActLength);
83
84 for i=1:ActLength
85
86     for j=1:CountLength
87         dataV = DataLoader(filelist,i,j);
88         DataM(:,i) = DataM(:,i) + dataV;
89     end
90     npoints = length(years);
91     %dataV = dataV-mean(dataV);
92     %dataV = dataV./std(dataV);
93     %Calculate basic statistics
94     meanV = mean(DataM(:,i));
95     stdV = std(DataM(:,i));
96     [ncounts, Edges] = histcounts(DataM(:,i),numbin);
97
98     for k=1:numbin
99         expval(k) = npoints*(normcdf(Edges(k+1),meanV,stdV)-
100             normcdf(Edges(k), ...
101                 meanV,stdV));
102         %expval(k) = npoints*(normcdf(Edges(k+1))-normcdf(
103             Edges(k)));
104     end
105     xi2 = sum((ncounts-expval).^2./expval);
106     %pval = chi2cdf(xi2,degfree,'upper');
107     fprintf('\nDoes %s Follow a normal distribution : ', ...
108         Activities{i});
109     if xi2 < chi2limit
110         h = 0;
111         fprintf('YES');
112         lwid = 1.5;
113     else
114         h = 1;
115         fprintf('NO');
116         lwid = 0.5;
117     end
118     NormAct = DataM(:,i)-meanV;
119     NormAct = NormAct./std(NormAct);
120     h1 = histogram(NormAct,numbin,'DisplayStyle','stairs','
121         LineWidth',lwid);
122     hold on;
123 end
124 fprintf('\n');

```

```

123 x1 = linspace(-3,3);
124 plot(x1,18*normpdf(x1),'LineWidth',2);
125 legend({Activities{:}},'Normal Distribution','Location','Best');
126 title('Normalized Activity Distribution');
127 ylabel('Occurances');
128 xlabel('Sigma distance from mean');
129 set(findall(gcf,'-property','FontSize'),'FontSize',14)

```

## 2.5 Exercise3

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.3. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, CIs και MeanP10.

Listing 5: Project2019Ex3.m

```

1  clc;
2  clear all;
3
4  %%Preload
5  %find all data and activities
6  filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
   data files
7  % set the filename for National Totals
8  TotFilename = 'EmissionP10NationalTotalsEU15.xls';
9  filelist(7) = [];%remove the Nationaltotals filename
10 names = {filelist(:).name};
11 Activities = extractBetween(names(:),'EmissionP10','EU15');
12 ActLength = length(Activities);
13 %Load a file to get the names of countries and years
14 [data,txt,row] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
15 dataname = txt(2,1);
16 years = str2double(txt(2:end,2));
17 Countries = [];
18 for i=1:length(data(1,:)) %get the name of the countries and
   store them
19     Countries = [Countries, extractBetween(txt(1,i+2),') - ',' -
   ')'];
20 end
21 Countries = Countries';
22 CountLength = length(Countries);
23 alpha = 0.05;%Set the significance level
24
25 %%Main program
26 fprintf('\n The available activities are \n');
27
28 for i=1:ActLength
29     fprintf('%d \t %s\n',i,Activities{i});
30 end
31
32 prompt3 = ['Choose an activity by the number on the left: '];
33 while 1

```



```

34     Count2 = int8(input(prompt3));
35     if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
36         break;
37     end
38 end
39
40 A = DataLoader(filelist, Count2, 1, 1); %get the values for a
    specific activity
41 Ball = xlsread(TotFilename); %get the totals for all countries
42
43 MeanP10(A, Ball, alpha);

```

## 2.6 CIs

Πρόγραμμα για τον υπολογισμό διαστημάτων εμπιστοσύνης για το Ζητούμενο 1.3.

Listing 6: CIs.m

```

1 function [pCIs, npCIs] = CIs(A, B, alpha)
2 %Calculate the Confidence Intervals (CIs) for the relative mean
    value of
3 %A compared to B for a significance level of alpha. Returns
    bothe
4 %the parametric and non parametric CIS.
5
6 D3 = A./B; %relative value
7 meanD3 = mean(D3); %mean of relative
8 stdD3 = std(D3); %std of relative
9 nsize = length(D3);
10 %parametric analysis
11 ta = tinv(1-alpha/2, nsize-1);
12 pCIs = [meanD3-ta*stdD3/sqrt(nsize), meanD3+ta*stdD3/sqrt(nsize)
    ]; %parametric intervals
13
14 %non parametric analysis
15 m = 1000; %number of randomly selecting sets of pairs
16 %create random pairs fo A and B by randomly selecting the
    indices of array
17 %B
18 D4 = zeros(m, 1);
19
20 for i = 1:m
21     indx = unidrnd(nsize, nsize, 1);
22     D4(i) = mean(A(indx)./B(indx));
23 end
24
25 D4 = sort(D4);
26 minCI = round(m*alpha/2);
27 maxCI = round(m*(1-alpha/2));
28
29 npCIs = [D4(minCI), D4(maxCI)];

```

```

30
31
32 end

```

## 2.7 MeanP10

Πρόγραμμα για τον υπολογισμό των μέσων τιμών για το Ζητούμενο 1.3.

Listing 7: MeanP10.m

```

1 function MeanP10(AM,BM, alpha)
2     nCount = length(AM(1,:));%get the number of countries
3     pCIs = zeros(nCount,2);
4     npCIs = zeros(nCount,2);
5     mpCIs = zeros(nCount,1);
6     mnpCIs = zeros(nCount,1);
7
8     for i=1:nCount
9         [pCIs(i,:), npCIs(i,:)] = CIs(AM(:,i),BM(:,i),alpha);
10    end
11
12    mpCIs = mean(pCIs,2);
13    mnpCIs = mean(npCIs,2);
14    figure();
15    errorbar(1:nCount,mpCIs,pCIs(:,1),pCIs(:,2),'o');
16    hold on;
17    errorbar(1:nCount,mnpCIs,npCIs(:,1),npCIs(:,2),'*');
18    title('Mean and CIs of avtivity');
19    ylabel('Relative mean PM10 value');
20    xlabel('Country code');
21    legend('Parametric','Non Parametric');
22    set(findall(gcf,'-property','FontSize'),'FontSize',14);
23 end

```

## 2.8 Exercise4

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.4. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, stdCIs και StdP10.

Listing 8: Project2019Ex4.m

```

1 clc;
2 clear all;
3
4 %%Preload
5 %find all data and activities
6 filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
7 % set tha filename for National Totals
8 TotFilename = 'EmissionP10NationalTotalsEU15.xls';
9 filelist(7)=[];%remove the Nationatotals filename
10 names = {filelist(:).name};

```

```

11 Activities = extractBetween(names(:) ,"EmissionP10 ","EU15");
12 ActLength = length(Activities);
13 %Load a file to get the names of countries and years
14 [data ,txt ,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
15 dataname = txt(2,1);
16 years = str2double(txt(2:end,2));
17 Countries = [];
18 for i=1:length(data(1,:)) %get the name of the countries and
    store them
19     Countries = [Countries , extractBetween(txt(1,i+2) ," - " , " -
        ") ]];
20 end
21 Countries = Countries';
22 CountLength = length(Countries);
23 alpha = 0.05;%Set the significance level
24
25 %%Main program
26 fprintf('\n The available activities are \n');
27
28 for i=1:ActLength
29     fprintf('%d \t %s\n',i ,Activities{i});
30 end
31
32 prompt3 = ['Choose an activity by the number on the left: '];
33 while 1
34     Count2 = int8(input(prompt3));
35     if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
36         break;
37     end
38 end
39
40 A = DataLoader(filelist ,Count2 ,1 ,1);%get the values for a
    specific activity
41 Ball = xlsread(TotFilename);%get the totals for all countries
42
43 StdP10(A,Ball , alpha);

```

## 2.9 stdCIs

Πρόγραμμα για τον υπολογισμό διαστημάτων εμπιστοσύνης για το Ζητούμενο 1.4.

Listing 9: stdCIs.m

```

1 function [pCIs ,npCIs] = stdCIs(A,B,alpha)
2 %Calculate the Confidence Intervals(CIs) for the relative std
    value of
3 %A compared to B for a significance level of alpha. Returns
    bothe
4 %the parametric and non parametric CIS.
5
6 D3 = A./B; %relative value

```

```

7 meanD3 = mean(D3); %mean of relative
8 stdD3 = std(D3); %std of relative
9 nsize = length(D3);
10 %parametric analysis
11 ta = chi2inv(1-alpha/2, nsize-1);
12 tb = chi2inv(alpha/2, nsize-1);
13 pCIs = [sqrt((nsize-1)*stdD3^2/ta), sqrt((nsize-1)*stdD3^2/tb)]; %
    parametric intervals
14
15 %non parametric analysis
16 m = 1000; %number of randomly selecting sets of pairs
17 %create random pairs fo A and B by randomly selecting the
    indices of array
18 %B
19 D4 = zeros(m,1);
20
21 for i=1:m
22     indx = unidrnd(nsize, nsize, 1);
23     D4(i) = std(A(indx)./B(indx));
24 end
25
26 D4 = sort(D4);
27 minCI = round(m*alpha/2);
28 maxCI = round(m*(1-alpha/2));
29
30 npCIs = [D4(minCI), D4(maxCI)];
31
32
33 end

```

## 2.10 StdP10

Πρόγραμμα για τον υπολογισμό των μέσων τιμών για το Ζητούμενο 1.4.

Listing 10: StdP10.m

```

1 function StdP10(AM, BM, alpha)
2     nCount = length(AM(1,:)); %get the number of countries
3     pCIs = zeros(nCount, 2);
4     npCIs = zeros(nCount, 2);
5     mpCIs = zeros(nCount, 1);
6     mnpCIs = zeros(nCount, 1);
7
8     for i=1:nCount
9         [pCIs(i,:), npCIs(i,:)] = stdCIs(AM(:, i), BM(:, i), alpha);
10    end
11
12    mpCIs = mean(pCIs, 2);
13    mnpCIs = mean(npCIs, 2);
14    figure();
15    errorbar(1:nCount, mpCIs, pCIs(:, 1), pCIs(:, 2), 'o');

```

```

16     hold on;
17     errorbar(1:nCount, mnpCIs, npCIs(:,1), npCIs(:,2), '* ');
18     title('Std and CIs of avtivity');
19     ylabel('Relative Std PM10 value');
20     xlabel('Country code');
21     legend('Parametric', 'Non Parametric');
22     set(findall(gcf, '-property', 'FontSize'), 'FontSize', 14);
23 end

```

## 2.11 Exercise5

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.5. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, AllCountries και corrme.

Listing 11: Project2019Ex5.m

```

1  clc;
2  clear all;
3
4  %%Preload
5  %find all data and activities
6  filelist = dir('EmissionP10*EU15.xls'); %Get the filenames of the
    data files
7  % set tha filename for National Totals
8  TotFilename = 'EmissionP10NationalTotalsEU15.xls';
9  filelist(7) = []; %remove the Nationatotals filename
10 names = {filelist(:).name};
11 Activities = extractBetween(names(:), 'EmissionP10', 'EU15');
12 ActLength = length(Activities);
13 %Load a file to get the names of countries and years
14 [data, txt, raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
15 dataname = txt(2,1);
16 years = str2double(txt(2:end,2));
17 Countries = [];
18 for i=1:length(data(1,:)) %get the name of the countries and
    store them
19     Countries = [Countries, extractBetween(txt(1,i+2), '-' , '-'
        ")]];
20 end
21 Countries = Countries';
22 CountLength = length(Countries);
23 alpha = 0.05; %Set the significance level
24
25 %%Main program
26 fprintf('\n The available activities are \n');
27
28 for i=1:ActLength
29     fprintf('%d \t %s\n', i, Activities{i});
30 end
31
32 prompt3 = ['Choose an activity by the number on the left: '];

```

```

33 while 1
34     Count2 = int8(input(prompt3));
35     if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
36         break;
37     end
38 end
39
40 %A = DataLoader(filelist ,Count2,1);%get the values for a
    specific activity
41 B = DataLoader(filelist ,Count2,2,1);
42 [Rmax CountriesMax Per1 Per2] = AllCountries(B,alpha);
43
44 fprintf('\nThe 10 sets of countries with the highest \
    ncorrelation for %s are\n',Activities{Count2});
45 fprintf('Correlation\t Country1\t Country2\n');
46 for i=1:10
47     fprintf(' %5.2f\t \t %10s\t %10s \n',Rmax(i),Countries{
        CountriesMax(i,1)},Countries{CountriesMax(i,2)});
48 end
49
50 fprintf('\n %5.2f%% of the sets of countries pass \nthe
    Parametric test with a significance level of %5.2f%%\n',100*
    Per1,100*alpha);
51 fprintf('\n %5.2f%% of the sets of countries pass \nthe NON-
    Parametric test with a significance level of %5.2f%%\n',100*
    Per2,100*alpha);

```

## 2.12 AllCountries

Πρόγραμμα για τον υπολογισμό διαστημάτων εμπιστοσύνης για το Ζητούμενο 1.5.

Listing 12: AllCountries.m

```

1 function [Rmax combmax count1 count2] = AllCountries(AM,alpha)
2     nCount = length(AM(1,:));
3     comb = combnk(1:nCount,2);%get all possible combinations of
        countries
4     ncomb = length(comb(:,1));
5     RPP = zeros(ncomb,3);
6
7     for i=1:ncomb
8         [RPP(i,1),RPP(i,2),RPP(i,3)] = corrme(AM(:,comb(i,1)),AM
            (:,comb(i,2)),alpha);
9     end
10
11
12     figure();
13     h = histogram(RPP(:,1),-1:0.25:1);
14     title('Correlations among all sets of countries');
15     ylabel('Counts');
16     xlabel('Corellation factor');

```

```

17     set(findall(gcf, '-property', 'FontSize'), 'FontSize', 14);
18     %sort the R values and pick the 10 highest
19     [Rsorted I] = sort(RPP(:,1), 'descend');
20     Rmax = Rsorted(1:10);
21     Imax = I(1:10);
22     combmax = comb(Imax, :);
23     count1 = sum(RPP(:,2) >= alpha) / ncomb;
24     count2 = sum(RPP(:,3) >= alpha) / ncomb;
25 end

```

### 2.13 corrme

Πρόγραμμα για τον υπολογισμό των μέσων τιμών για το Ζητούμενο 1.5.

Listing 13: corrme.m

```

1 function [r1,p1,p2] = corrme(A,B,alpha)
2     nsize = length(A);
3
4
5     %parametric correlation
6     [r1,p1] = corrcoef(A,B, 'alpha', alpha);
7     r1 = r1(1,2);
8     p1 = p1(1,2);
9     ta0 = abs(r1*sqrt((nsize-2)/(1-r1^2)));
10
11     %non parametric correlation
12     m = 1000;
13     ta = zeros(m,1);
14     ra = zeros(m,1);
15     for i=1:m
16         indx = randperm(nsize);
17         r0 = corrcoef(A,B(indx), 'alpha', alpha);
18         ra(i) = r0(1,2);
19         ta(i) = abs(ra(i)*sqrt((nsize-2)/(1-ra(i)^2)));
20     end
21     ta = sort(ta);
22     %fprintf(['t10 should be between indices ', num2str(round(m*
23         alpha/2)), ...
24         ' and ', num2str(round((1-alpha/2)*m)), ' and is ',
25         num2str(round(sum(ta < ta0))), '\n']);
26     maxind = round(sum(ta < ta0));
27     p2 = (1-maxind/(1.0*m));
28 end

```

### 2.14 Exercise6

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.6. Το οποίο με την σειρά του καλεί το πρόγραμμα DataLoader, AllActivities και corrme.

Listing 14: Project2019Ex6.m

```

1 clc;
2 clear all;
3
4 %%Preload
5 %find all data and activities
6 filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
7 % set the filename for National Totals
8 TotFilename = 'EmissionP10NationalTotalsEU15.xls';
9 filelist(7) = [];%remove the Nationatotals filename
10 names = {filelist(:).name};
11 Activities = extractBetween(names(:), 'EmissionP10', 'EU15');
12 ActLength = length(Activities);
13 %Load a file to get the names of countries and years
14 [data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
15 dataname = txt(2,1);
16 years = str2double(txt(2:end,2));
17 Countries = [];
18 for i=1:length(data(1,:)) %get the name of the countries and
    store them
19     Countries = [Countries, extractBetween(txt(1,i+2), '-' , '-'
        " ")];
20 end
21 Countries = Countries';
22 CountLength = length(Countries);
23 alpha = 0.05;%Set the significance level
24
25 %%Main program
26 fprintf('\n The available Countries are \n');
27
28 for i=1:CountLength
29     fprintf('%d \t %s\n',i,Countries{i});
30 end
31
32 prompt3 = ['Choose a Country by the number on the left: '];
33 while 1
34     Count2 = int8(input(prompt3));
35     if isinteger(Count2) & 1 <= Count2 & Count2 <= CountLength
36         break;
37     end
38 end
39 CM = zeros(length(years),ActLength);
40 %load all the activities for a specific country
41 for i=1:ActLength
42     CM(:,i) = DataLoader(filelist,i,Count2);
43 end
44
45 [Rmax ActivitiesMax Per1 Per2] = AllActivities(CM,alpha);
46

```



```

47 fprintf('\n The 10 sets of Activities with the highest
    correlation for %s are\n', Countries{Count2});
48 fprintf('Correlation\t Activity1\t \t Activity2\n');
49 for i=1:10
50     fprintf('%5.2f\t %20s\t \t %20s \n', Rmax(i), Activities{
        ActivitiesMax(i,1)}, Activities{ActivitiesMax(i,2)});
51 end
52
53 fprintf('\n %5.2f%% of the sets of Activities pass \n the
    Parametric test with a significance level of %5.2f%%\n', 100*
    Per1, 100*alpha);
54 fprintf('\n %5.2f%% of the sets of Activities pass \n the NON-
    Parametric test with a significance level of %5.2f%%\n', 100*
    Per2, 100*alpha);

```

## 2.15 AllActivities

Πρόγραμμα για τον υπολογισμό διαστημάτων εμπιστοσύνης για το Ζητούμενο 1.6.

Listing 15: AllActivities.m

```

1 function [Rmax combmax count1 count2] = AllActivities(AM, alpha)
2     nCount = length(AM(1,:));
3     comb = combnk(1:nCount,2); %get all possible combinations of
        Activities
4     ncomb = length(comb(:,1));
5     RPP = zeros(ncomb,3);
6
7     for i=1:ncomb
8         [RPP(i,1), RPP(i,2), RPP(i,3)] = corrme(AM(:, comb(i,1)), AM
            (:, comb(i,2)), alpha);
9     end
10
11
12     figure();
13     h = histogram(RPP(:,1), -1:0.25:1);
14     title('Correlations among all sets of Activities');
15     ylabel('Counts');
16     xlabel('Corellation factor');
17     set(findall(gcf, '-property', 'FontSize'), 'FontSize', 14);
18     %sort the R values and pick the 10 highest
19     [Rsorted I] = sort(RPP(:,1), 'descend');
20     Rmax = Rsorted(1:10);
21     Imax = I(1:10);
22     combmax = comb(Imax,:);
23     count1 = sum(RPP(:,2) >= alpha) / ncomb;
24     count2 = sum(RPP(:,3) >= alpha) / ncomb;
25 end

```

### 3 UNDERSTANDING TEXT

#### 3.0.1 Suppose ``chuck" implies throwing.

According to the Associated Press (1988), a New York Fish and Wildlife technician named Richard Thomas calculated the volume of dirt in a typical 25--30 foot (7.6--9.1 m) long woodchuck burrow and had determined that if the woodchuck had moved an equivalent volume of wood, it could move ``about 700 pounds (320 kg) on a good day, with the wind at his back".

#### 3.0.2 Suppose ``chuck" implies vomiting.

A woodchuck can ingest  $361.92 \text{ cm}^3$  (22.09 cu in) of wood per day. Assuming immediate expulsion on ingestion with a 5% retainment rate, a woodchuck could chuck  $343.82 \text{ cm}^3$  of wood per day.

**BONUS: SUPPOSE THERE IS NO WOODCHUCK.** Fusce varius orci ac magna dapibus porttitor. In tempor leo a neque bibendum sollicitudin. Nulla pretium fermentum nisi, eget sodales magna facilisis eu. Praesent aliquet nulla ut bibendum lacinia. Donec vel mauris vulputate, commodo ligula ut, egestas orci. Suspendisse commodo odio sed hendrerit lobortis. Donec finibus eros erat, vel ornare enim mattis et.

### 4 INTERPRETING EQUATIONS

#### 4.1 Identify the author of Equation 4.1 below and briefly describe it in English.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4.1)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porttitor arcu luctus, imperdiet urna iaculis, mattis eros. Pellentesque iaculis odio vel nisl ullamcorper, nec faucibus ipsum molestie. Sed dictum nisl non aliquet porttitor. Etiam vulputate arcu dignissim, finibus sem et, viverra nisl. Aenean luctus congue massa, ut laoreet metus ornare in. Nunc fermentum nisi imperdiet lectus tincidunt vestibulum at ac elit. Nulla mattis nisl eu malesuada suscipit.

#### 4.2 Try to make sense of some more equations.

$$\begin{aligned} (x+y)^3 &= (x+y)^2(x+y) \\ &= (x^2 + 2xy + y^2)(x+y) \\ &= (x^3 + 2x^2y + xy^2) + (x^2y + 2xy^2 + y^3) \\ &= x^3 + 3x^2y + 3xy^2 + y^3 \end{aligned} \quad (4.2)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

$$A = \begin{bmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{bmatrix} \quad (4.3)$$

Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

## 5 VIEWING LISTS

### 5.1 Bullet Point List

- First item in a list
  - First item in a list
    - \* First item in a list
    - \* Second item in a list
  - Second item in a list
- Second item in a list

### 5.2 Numbered List

1. First item in a list
2. Second item in a list
3. Third item in a list

## 6 INTERPRETING A TABLE

<i>Per 50g</i>	<b>Pork</b>	<b>Soy</b>
Energy	760kJ	538kJ
Protein	7.0g	9.3g
Carbohydrate	0.0g	4.9g
Fat	16.8g	9.1g
Sodium	0.4g	0.4g
Fibre	0.0g	1.4g

Πίνακας 6.1: Sausage nutrition.

**6.1 The table above shows the nutritional consistencies of two sausage types. Explain their relative differences given what you know about daily adult nutritional recommendations.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porttitor arcu luctus, imperdiet urna iaculis, mattis eros. Pellentesque iaculis odio vel nisl ullamcorper, nec faucibus ipsum molestie. Sed dictum nisl non aliquet porttitor. Etiam vulputate arcu dignissim, finibus sem et, viverra nisl. Aenean luctus congue massa, ut laoreet metus ornare in. Nunc fermentum nisi imperdiet lectus tincidunt vestibulum at ac elit. Nulla mattis nisl eu malesuada suscipit.

## 7 READING A CODE LISTING

Listing 16: Luftballons Perl Script.

```
1 #!/usr/bin/perl
2
3 use strict;
4 use warnings;
```

```

5
6 for (1..99) { print $_." Luftballons\n"; }
7
8 # This is a commented line
9
10 my $string = "Hello World!";
11
12 print $string."\n\n";
13
14 $string =~ s/Hello/Goodbye Cruel/;
15
16 print $string."\n\n";
17
18 finale();
19
20 exit;
21
22 sub finale { print "Fin.\n"; }

```

### 7.1 How many luftballons will be output by the Listing 16 above?

Aliquam arcu turpis, ultrices sed luctus ac, vehicula id metus. Morbi eu feugiat velit, et tempus augue. Proin ac mattis tortor. Donec tincidunt, ante rhoncus luctus semper, arcu lorem lobortis justo, nec convallis ante quam quis lectus. Aenean tincidunt sodales massa, et hendrerit tellus mattis ac. Sed non pretium nibh. Donec cursus maximus luctus. Vivamus lobortis eros et massa porta porttitor.

### 7.2 Identify the regular expression in Listing 16 and explain how it relates to the anti-war sentiments found in the rest of the script.

Fusce varius orci ac magna dapibus porttitor. In tempor leo a neque bibendum sollicitudin. Nulla pretium fermentum nisi, eget sodales magna facilisis eu. Praesent aliquet nulla ut bibendum lacinia. Donec vel mauris vulputate, commodo ligula ut, egestas orci. Suspendisse commodo odio sed hendrerit lobortis. Donec finibus eros erat, vel ornare enim mattis et.