Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Σχολή Θετικών Επιστημών, Τμήμα Φυσικής

Εργασία στο Μάθημα Ανάλυση Δεδομένων

Κίμων Ζαγκούρης
A.E.M. 4353
Email: kzagko@gmail.com

12 Φεβρουαρίου 2019

[1]

[1] Image created by Timashov Sergiy.

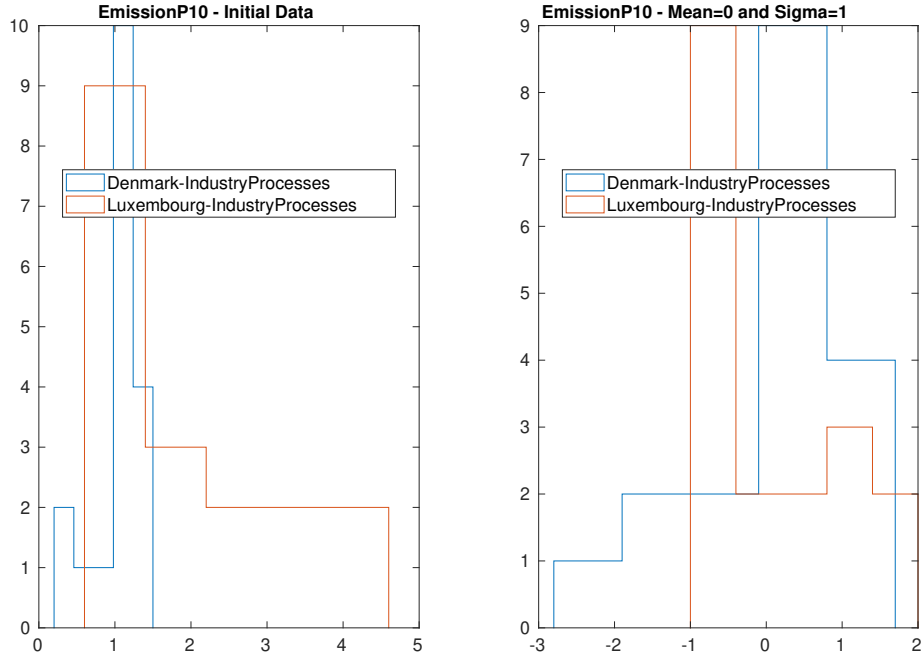
Περιεχόμενα

1 Ζητήματα

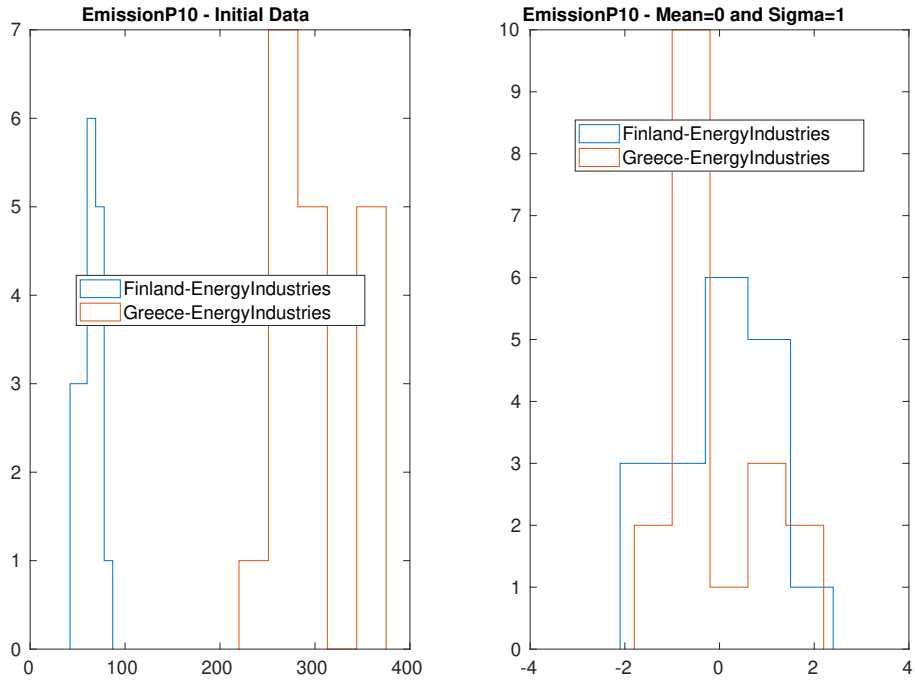
Παρακάτω είναι τα 10 ζητήματα της εργασίας αυτής ανά κεφάλαιο. Στο Κεφάλαιο 2 υπάρχουν τα αντίστοιχα προγράμματα Matlab που χρησιμοποιήθηκαν για την επίλυση των ζητημάτων. Στα προγράμματα υπάρχουν σχόλια για την καλύτερη παρακολούθηση του κώδικα σε κάποια σημεία.

1.1 Ζήτημα 1

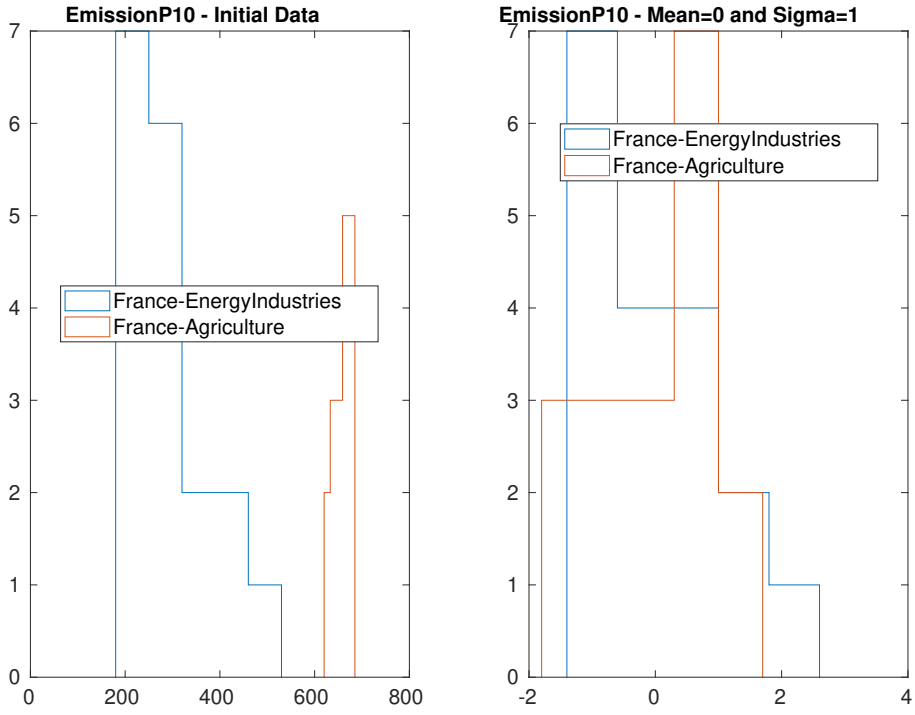
Στο ζήτημα αυτό τρέχουμε τον παρακάτω Κώδικα για 6 περιπτώσεις και παραθέτουμε τα γραφήματα.



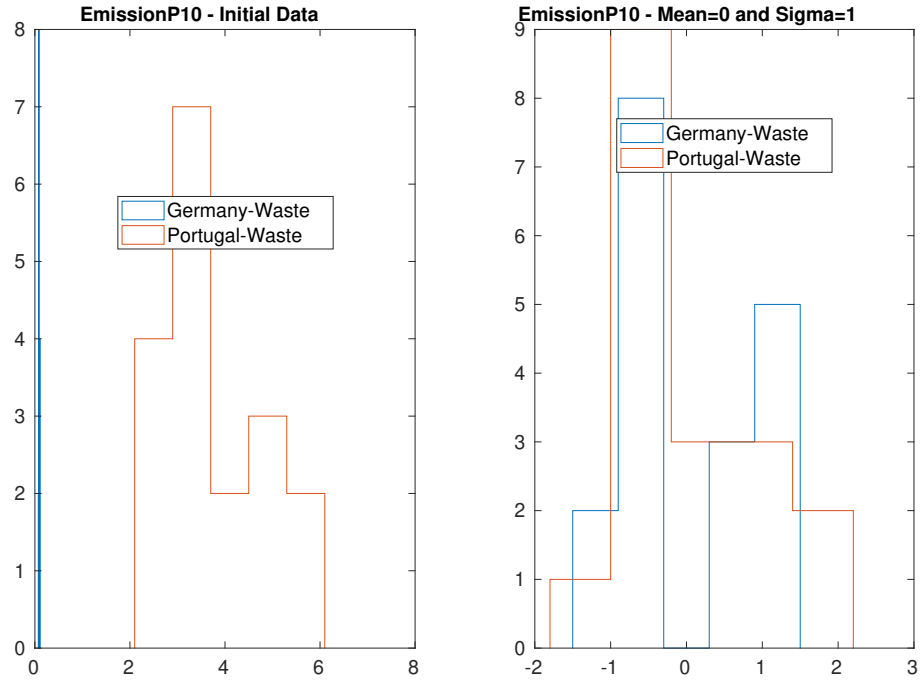
Σχήμα 1.1: PM10 Emissions for Denmark Luxembourg IndustryProcesses.



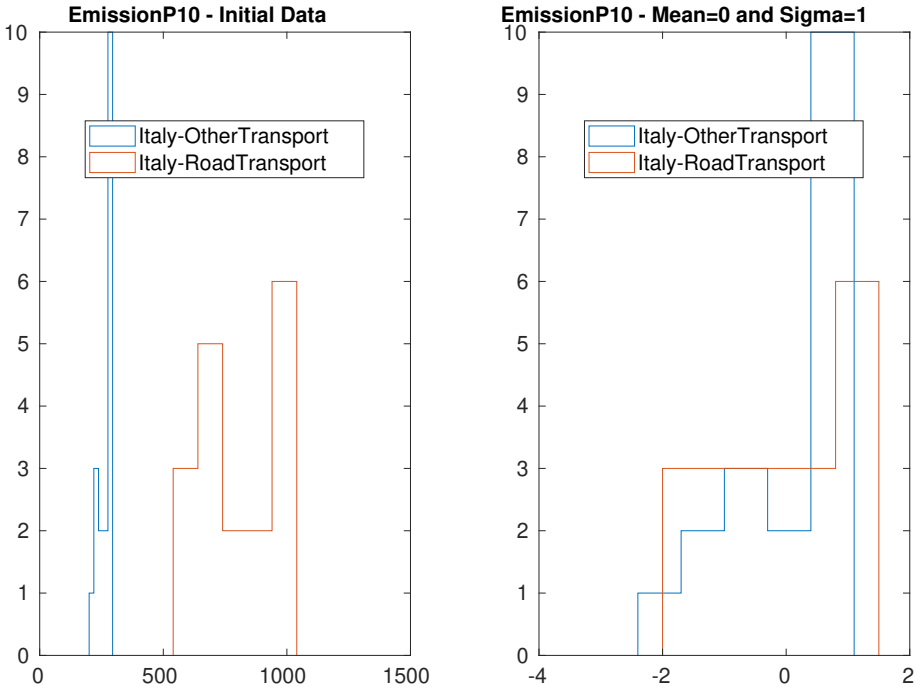
Σχήμα 1.2: PM10 Emissions for Finland Greece EnergyIndustries.



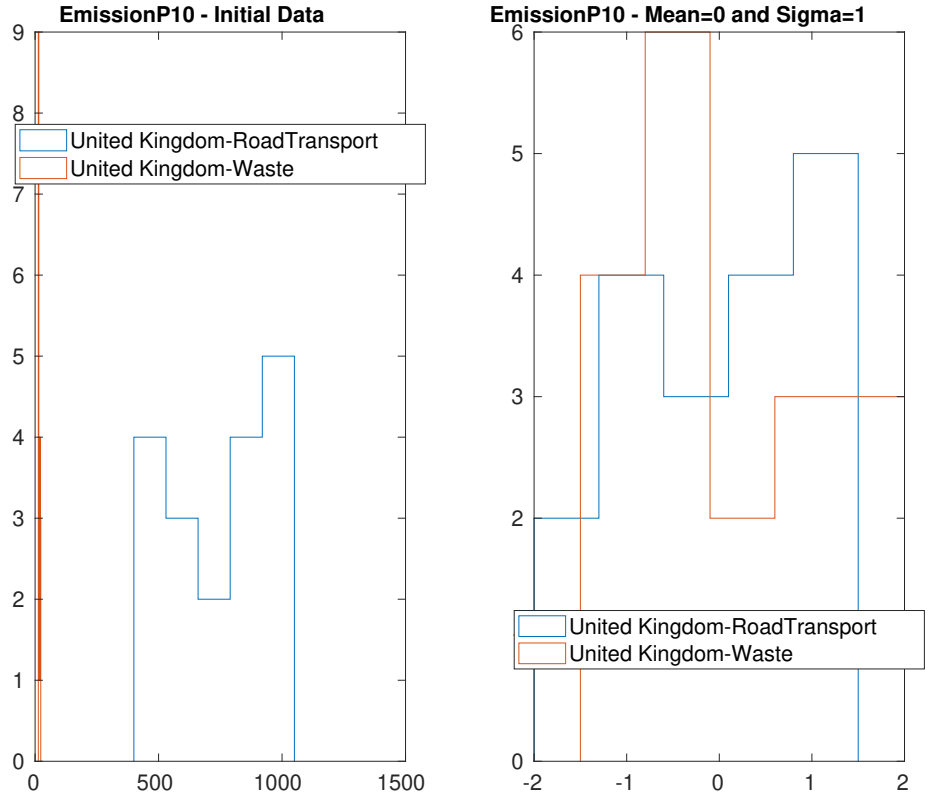
Σχήμα 1.3: PM10 Emissions for France EnergyIndustries Agriculture.



Σχήμα 1.4: PM10 Emissions for Germany Portugal Waste.



Σχήμα 1.5: PM10 Emissions for Italy OtherTransport RoadTransport.

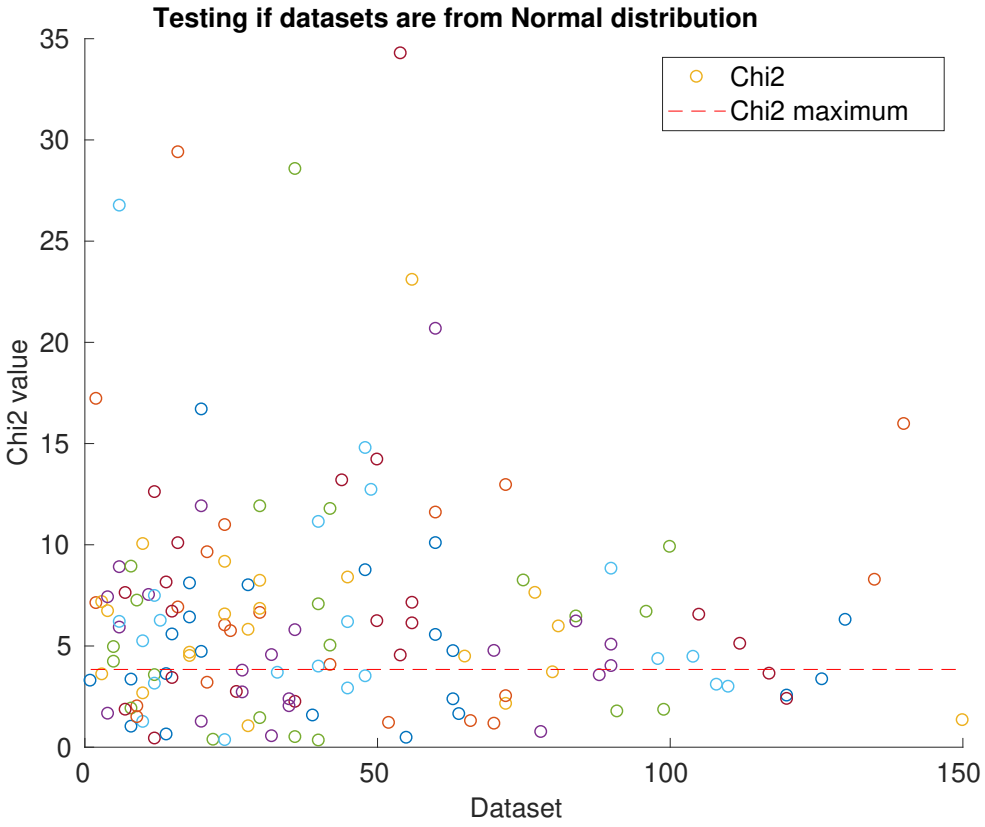


Σχήμα 1.6: PM10 Emissions for United Kingdom RoadTransport Waste.

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 1.3 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 1.5 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 χ^2. Για τον λόγο αυτό στο παραπάνω πρόγραμμα επιλέξαμε να κάνουμε αναλυτική μέτρηση του χ^2. Μετά την εκτέλεση το πρόγραμμα μας επιστρέφει τα παρακάτω σχήματα 1.7 και 1.8.



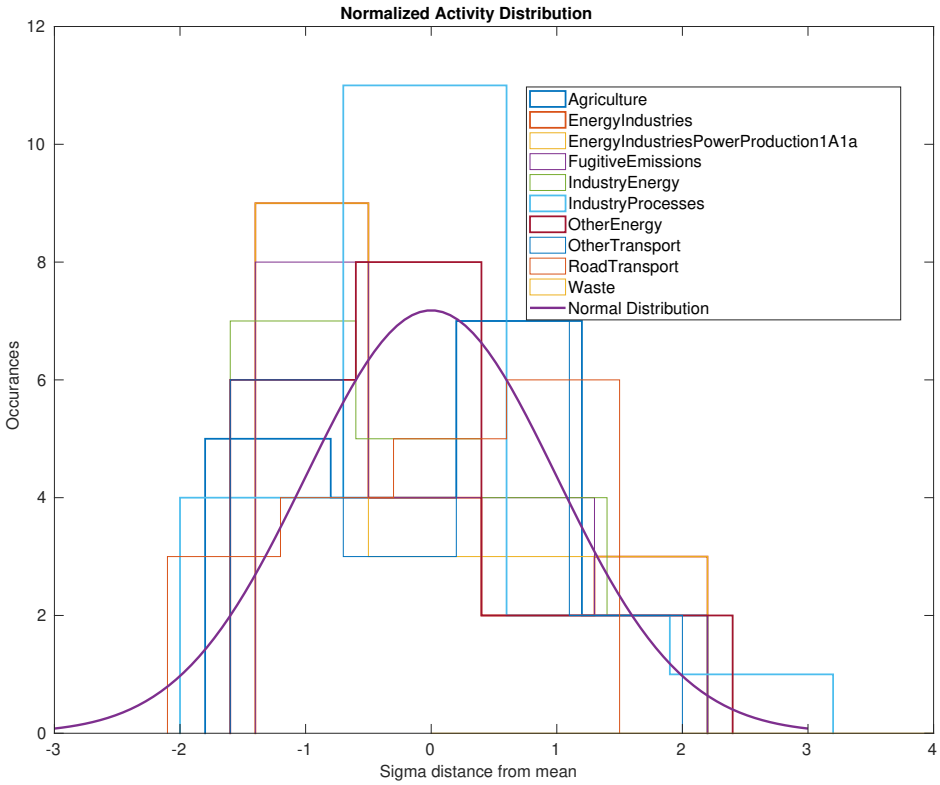
Σχήμα 1.7: Αποτελέσματα χ^2 για κάθε δραστηριότητα και χώρα καθώς και το μέγιστο όριο με διακεκομμένη γραμμή.

Από το σχήμα 1.7 παρατηρούμε ότι αρκετά χ^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 confindence level of  5.00%

Does Agriculture Follow a normal distribution : YES
Does EnergyIndustries Follow a normal distribution : YES
Does EnergyIndustriesPowerProduction1A1a 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 σημαντικότη-

τας βλέπουμε ότι μόνο 4 στις 10 δραστηριότητες θα μπορούσαν να προέρχονται από κανονική κατανομή.

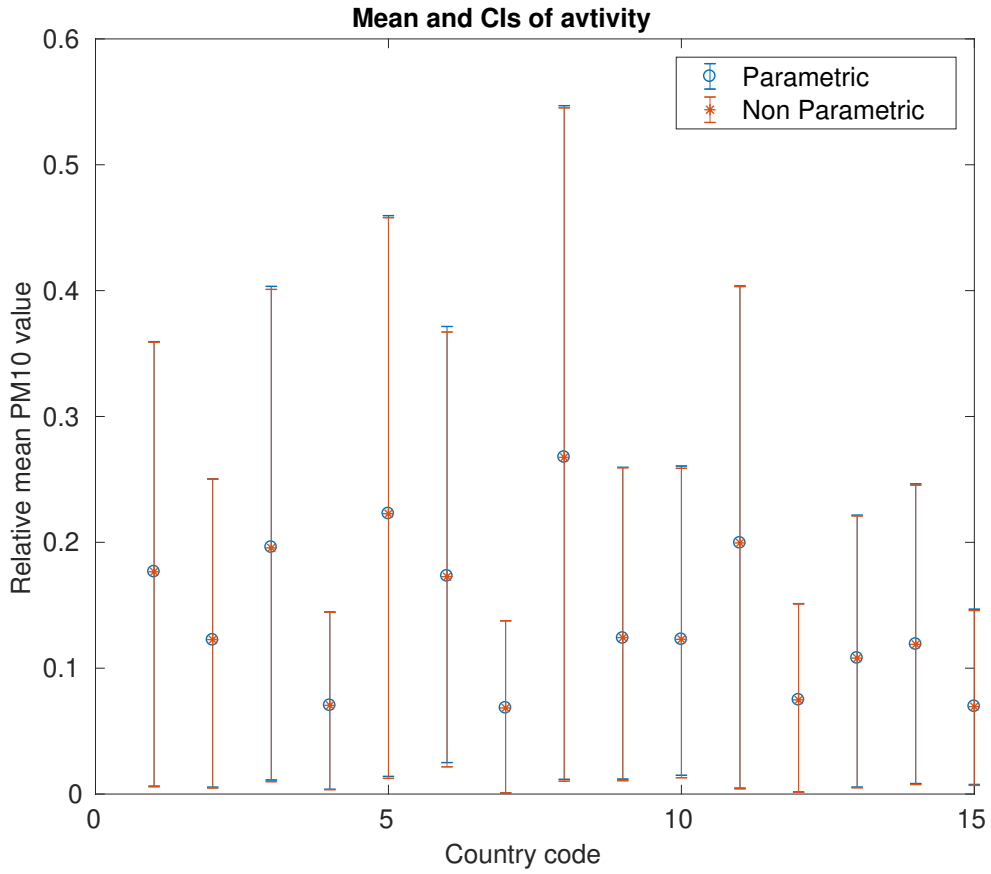


Σχήμα 1.8: Κανονικοποιημένες συνολικές κατανομές για όλες τις δραστηριότητες καθώς και ενδεικτική κανονική κατανομή για σύγκριση. Οι δραστηριότητες με ποιο έντονη γραμμή έχουν περάσει το τεστ κανονικότητας.

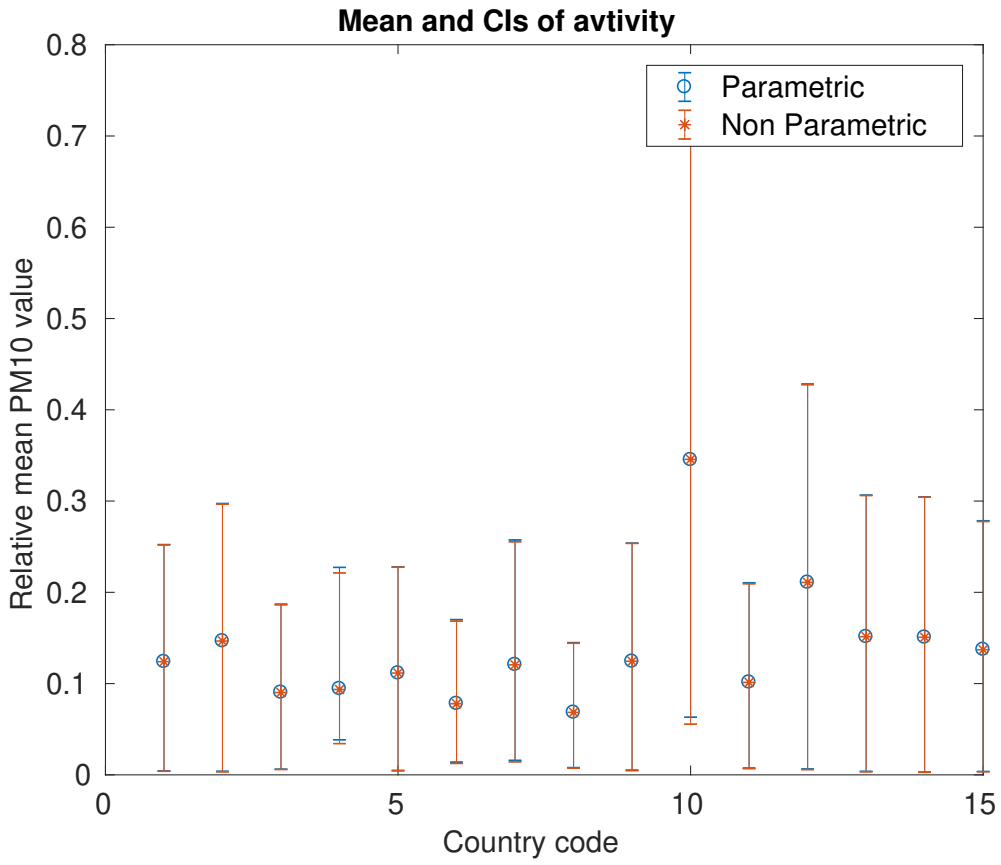
Στο σχήμα 1.8 παρατηρούμε ότι γενικά οι κανονικοποιημένες κατανομές των δραστηριοτήτων διαφέρουν από κανονικές. Πέρα απο τα αποτελέσματα που μας επιστρέφει το πρόγραμμα θα μπορούσε κάποιος οπτικά και μόνο να πει ότι από όλες τις κατανομές αυτές που φαίνεται να ακολουθούν κανονική κατανομή είναι: IndustryProcesses και OtherEnergy.

1.3 Ζήτημα 3

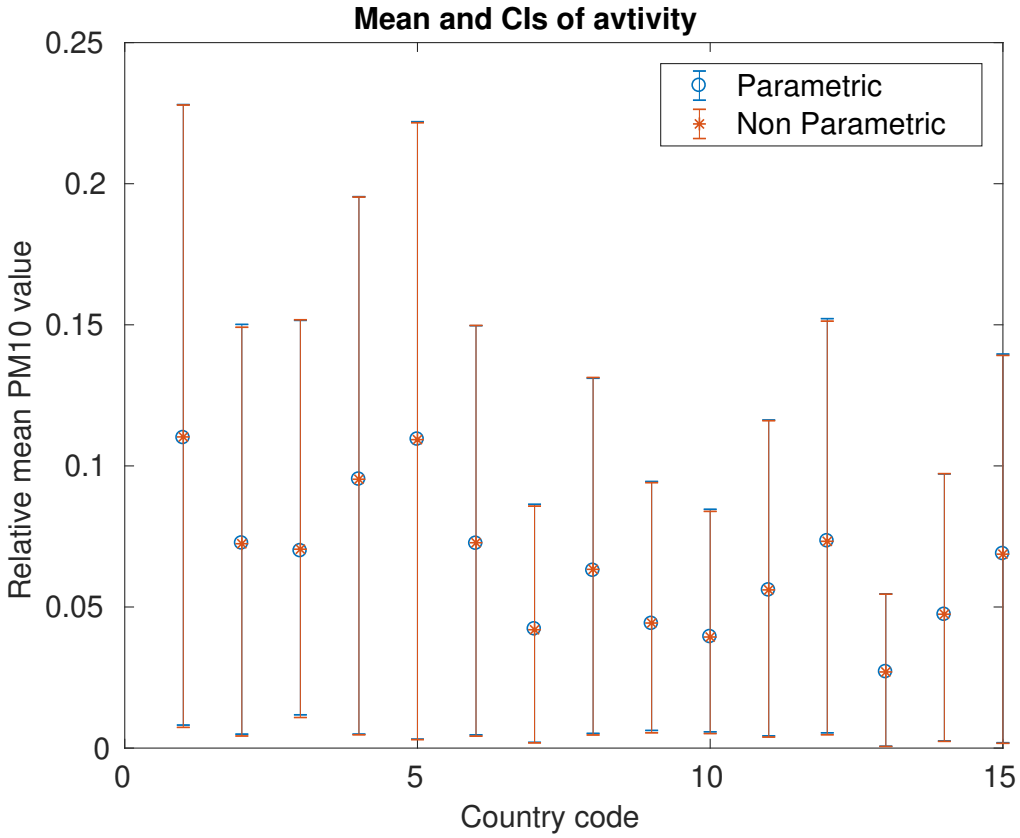
Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 5. Ενδεικτικά εκτελέσαμε τον κώδικα για 4 δραστηριότητες και τα αποτελέσματα φαίνονται στα σχήμα 1.9,σχήμα 1.10,σχήμα 1.11 και σχήμα 1.12.



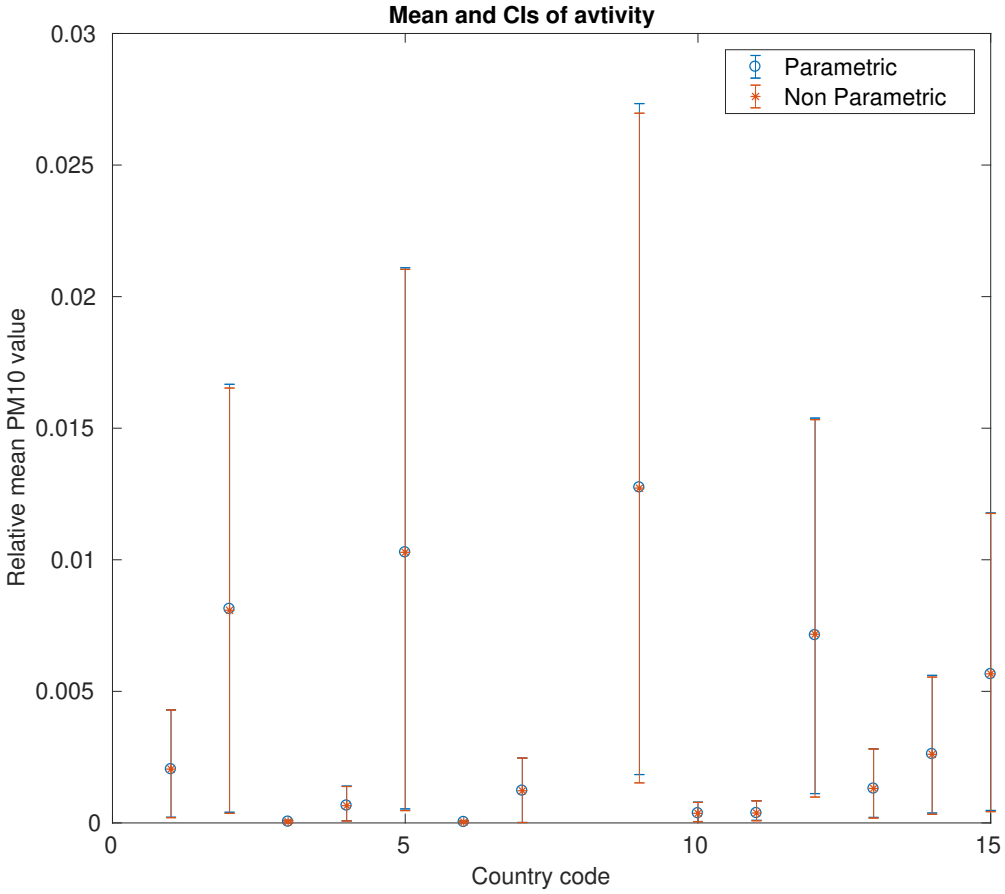
Σχήμα 1.9: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα Agriculture σε κάθε χώρα.



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.

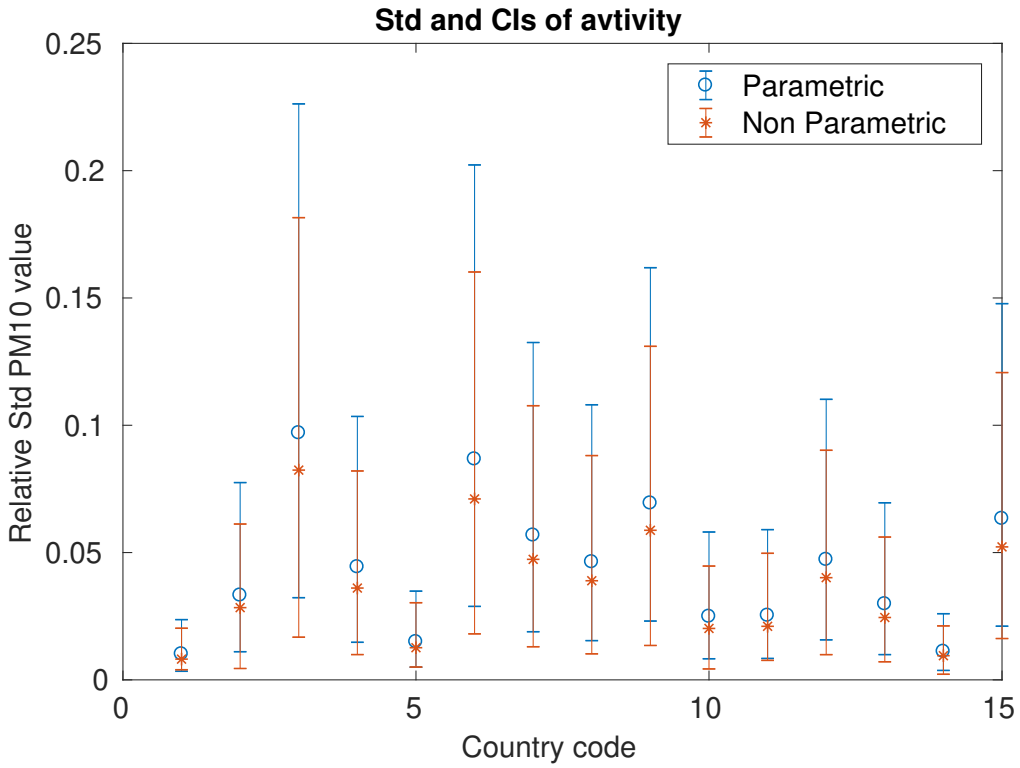


Σχήμα 1.12: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα για την δραστηριότητα Waste σε κάθε χώρα.

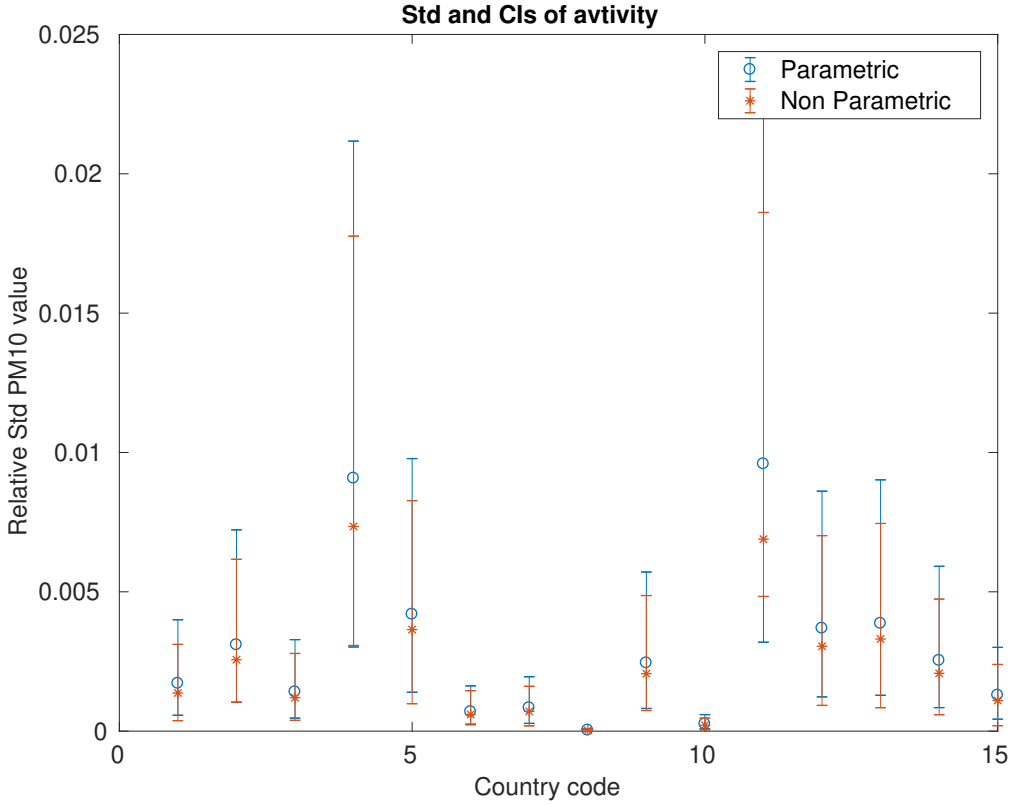
Από τα παραπάνω σχήματα δεν φαίνεται να υπάρχει μεγάλη διαφορά μεταξύ των μέσων τιμών και των διαστημάτων εμπιστοσύνης που προέρχονται από παραμετρική και με παραμετρική εκτίμηση.

1.4 Ζήτημα 4

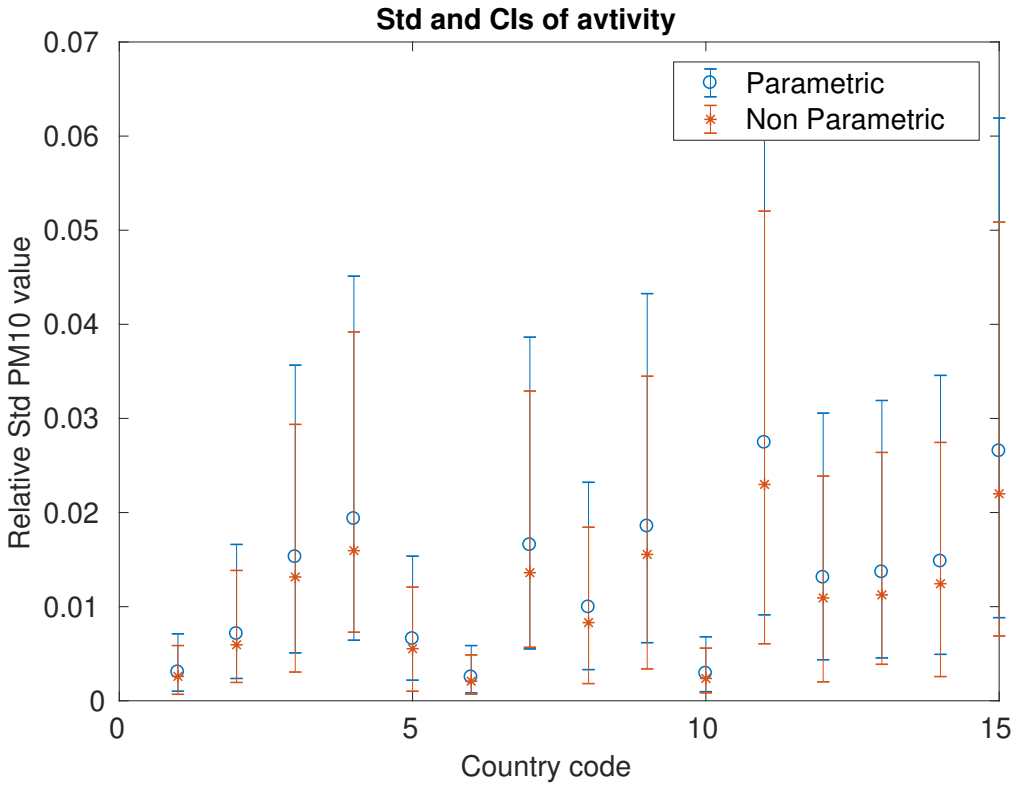
Η ίδια ανάλυση όπως και στο Κεφάλαιο 1.3 αλλά τώρα για την τυπική απόκλιση. Η ανάλυση γίνεται με τον κώδικα 8. Καλέσαμε τον παραπάνω κώδιακ ενδεικτικά για μερικές δραστηριότητες και πήραμε τα παρακάτω σχήμα 1.13,σχήμα 1.14,σχήμα 1.15 και σχήμα 1.16.



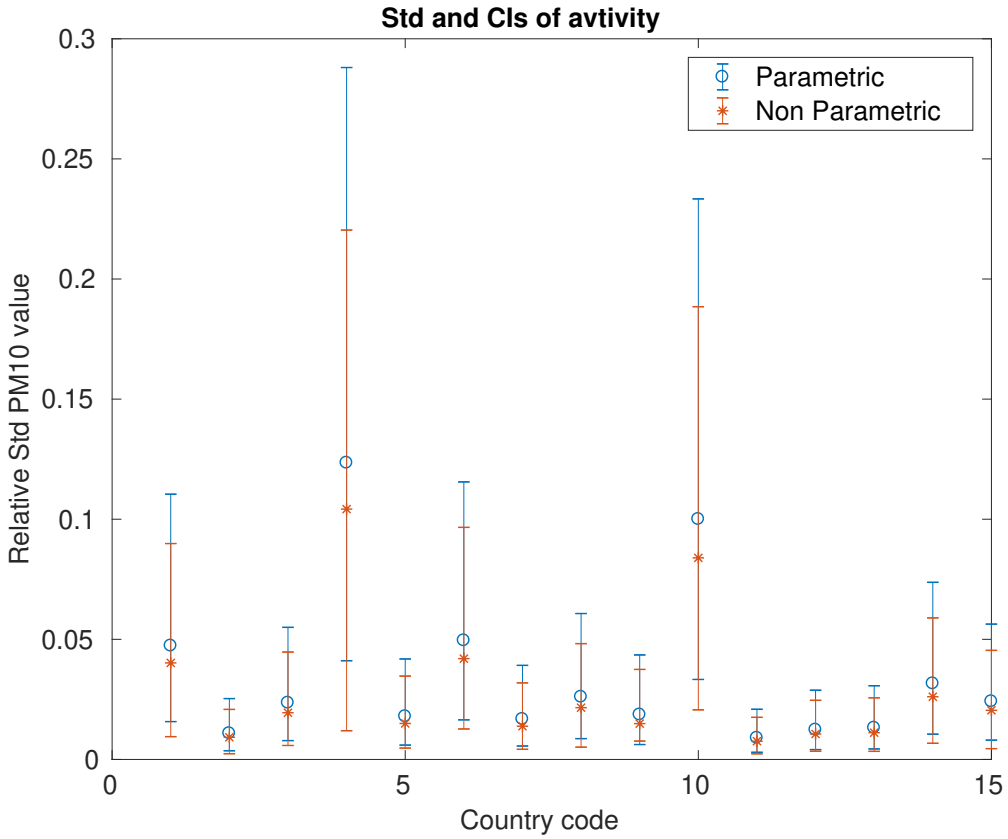
Σχήμα 1.13: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα EnergyIndustries σε κάθε χώρα.



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.

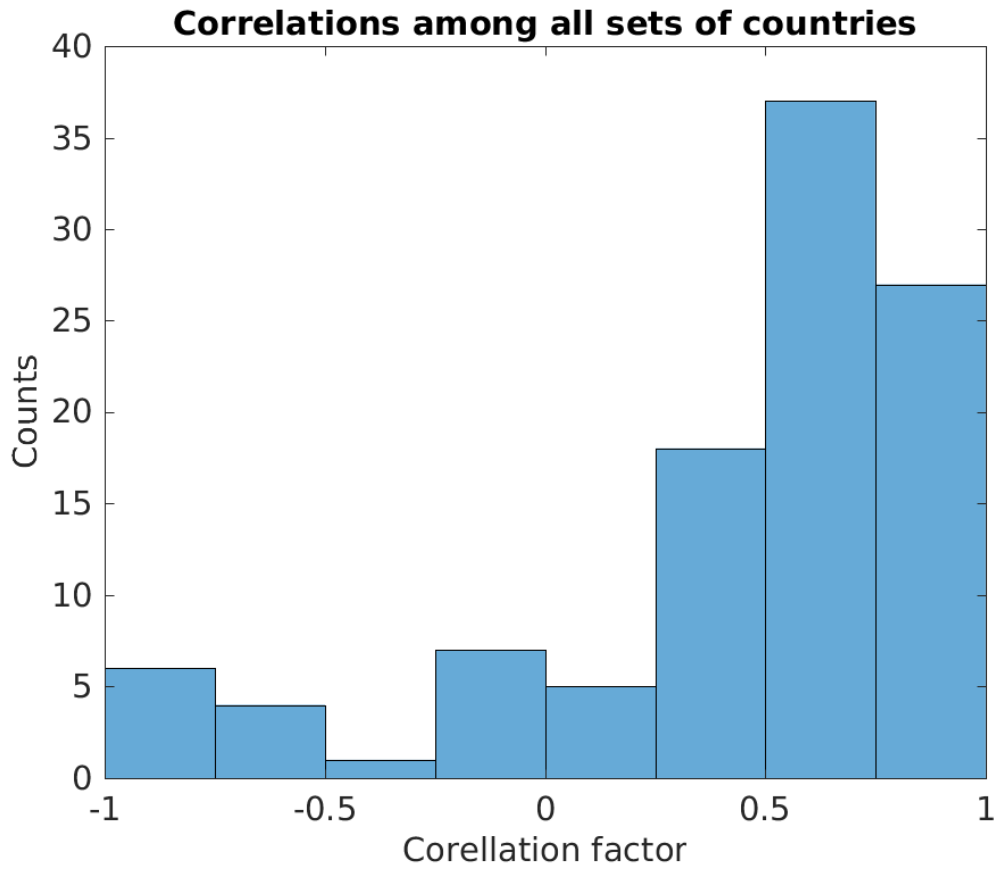


Σχήμα 1.16: Μέσες τιμές και παραμετρικά και μη παραμετρικά διαστήματα της τυπικής απόκλισης για την δραστηριότητα RoadTransport σε κάθε χώρα.

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.

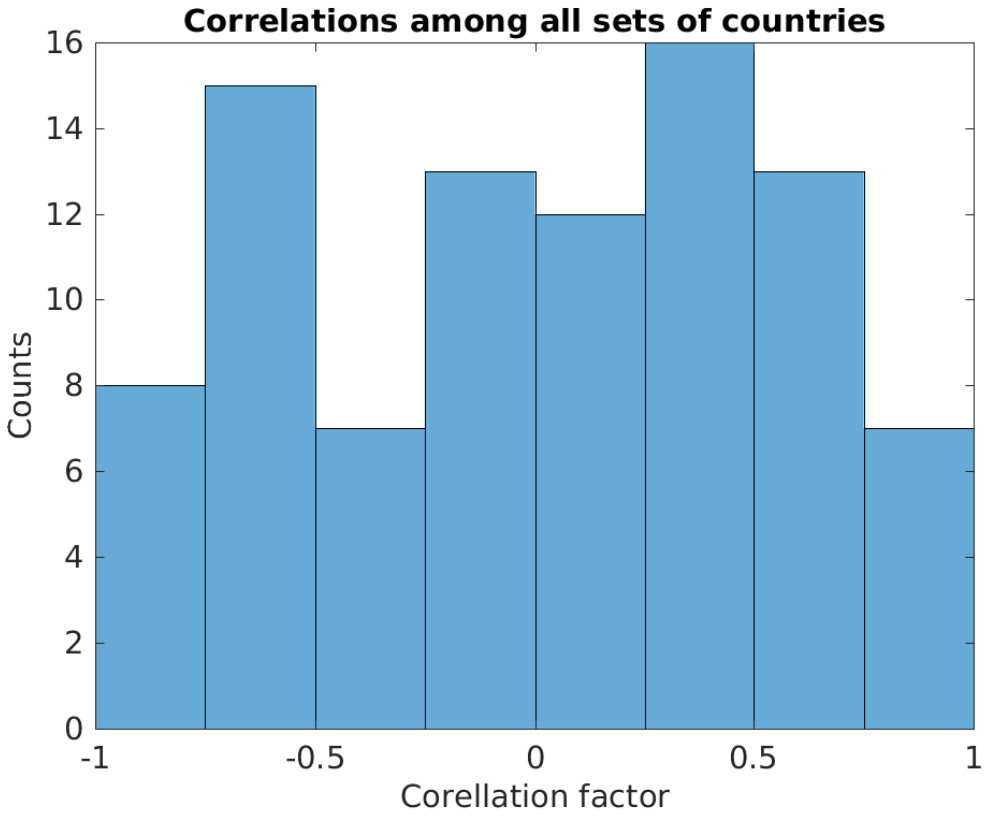
1.5 Ζήτημα 5

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. Στα αποτελέσματα που εκτυπώνονται αναφέρεται η απόλυτη τιμή του συντελεστή συσχέτισης.



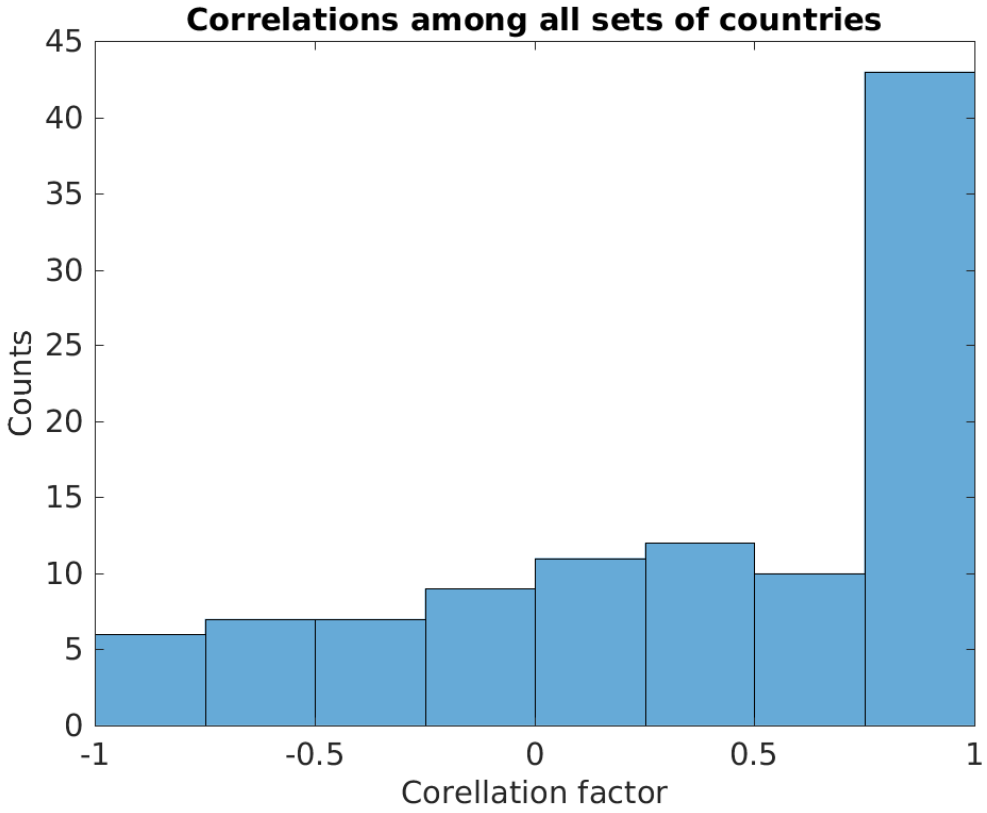
Σχήμα 1.17: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας Agriculture.

```
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%

 24.76% of the sets of countries pass
the NON-Parametric test with a significance level of  5.00%
```



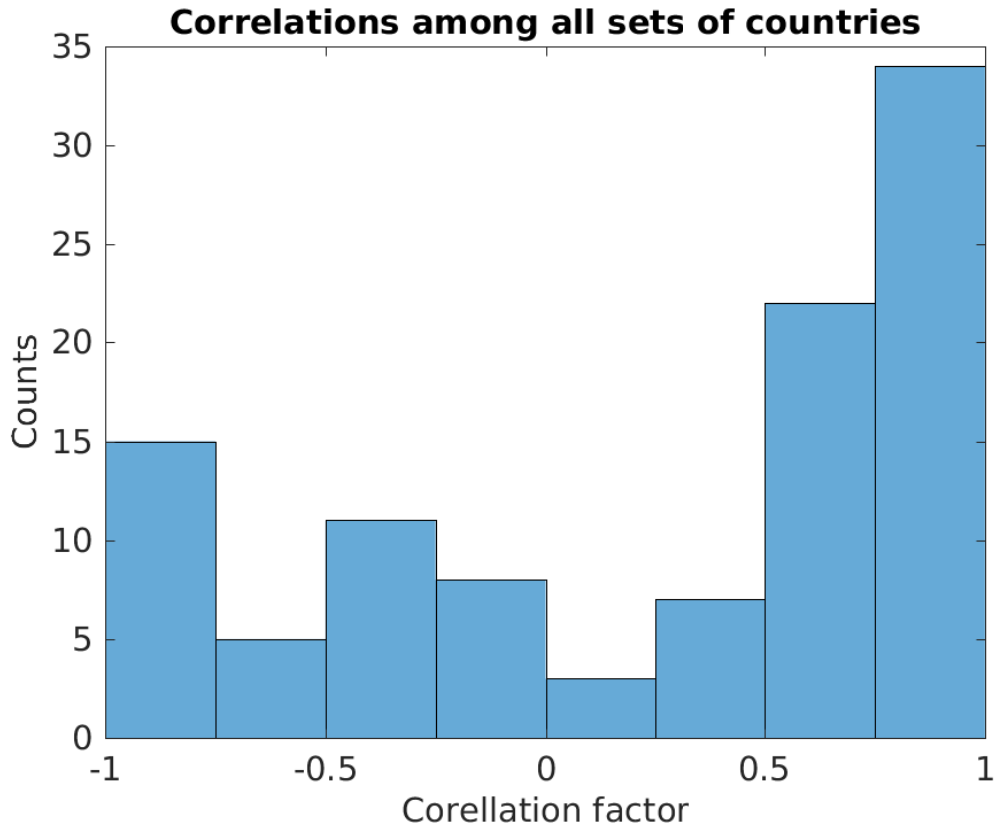
Σχήμα 1.18: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας Waste.

```
The 10 sets of countries with the highest
correlation for Waste are
Correlation         Country1         Country2
 0.88                     Germany         Luxembourg
 0.87                     Austria            Germany
 0.87                      Sweden         United Kingdom
 0.86                    Portugal             Sweden
 0.86                     Austria         Luxembourg
 0.86                    Portugal         United Kingdom
 0.85                  Luxembourg         United Kingdom
 0.83                  Netherlands           Portugal
 0.82                     Germany              Spain
 0.82                  Luxembourg            Portugal

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

 58.10% 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%

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

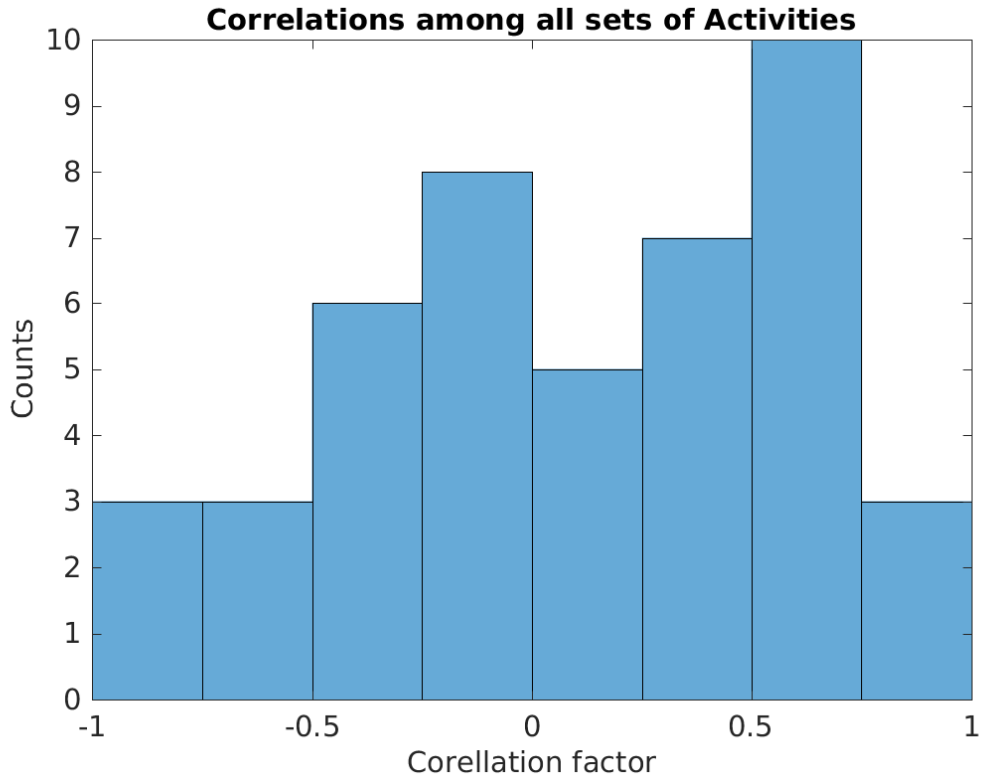


Σχήμα 1.20: Ιστόγραμμα των συντελεστών συσχέτισης της δραστηριότητας RoadTransport.

```
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%

 26.67% of the sets of countries pass
the NON-Parametric test with a significance level of  5.00%
```

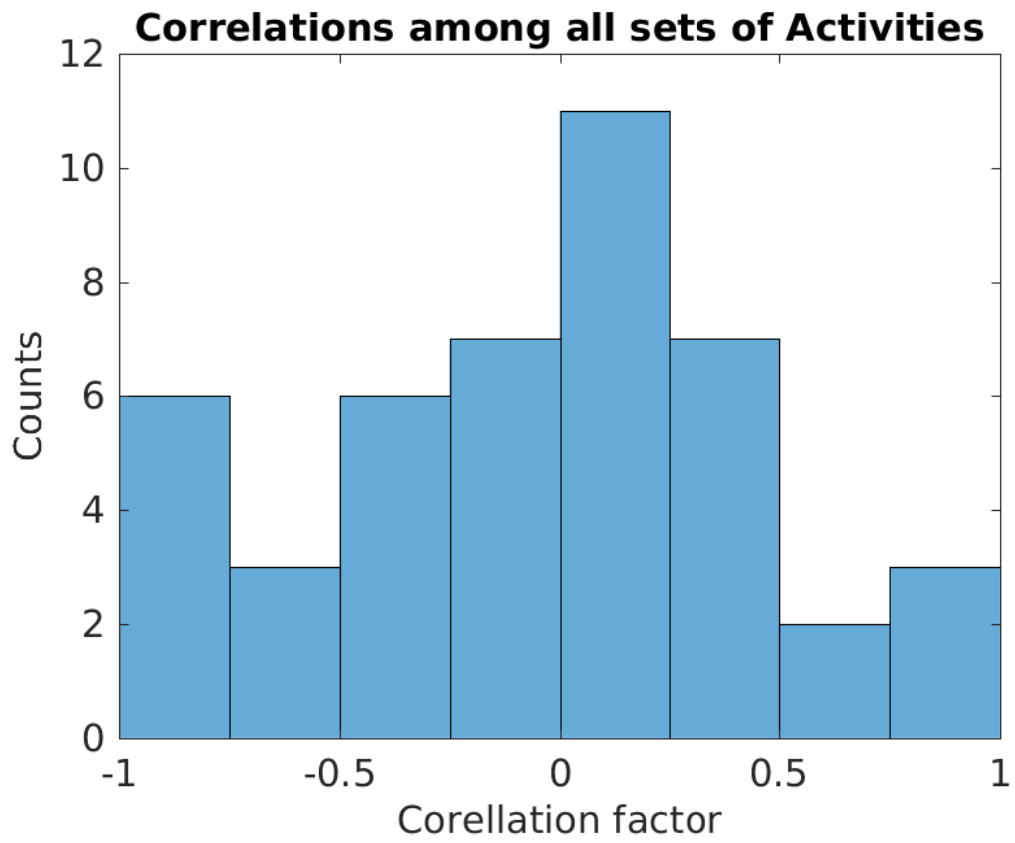
1.6 Ζήτημα 6

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



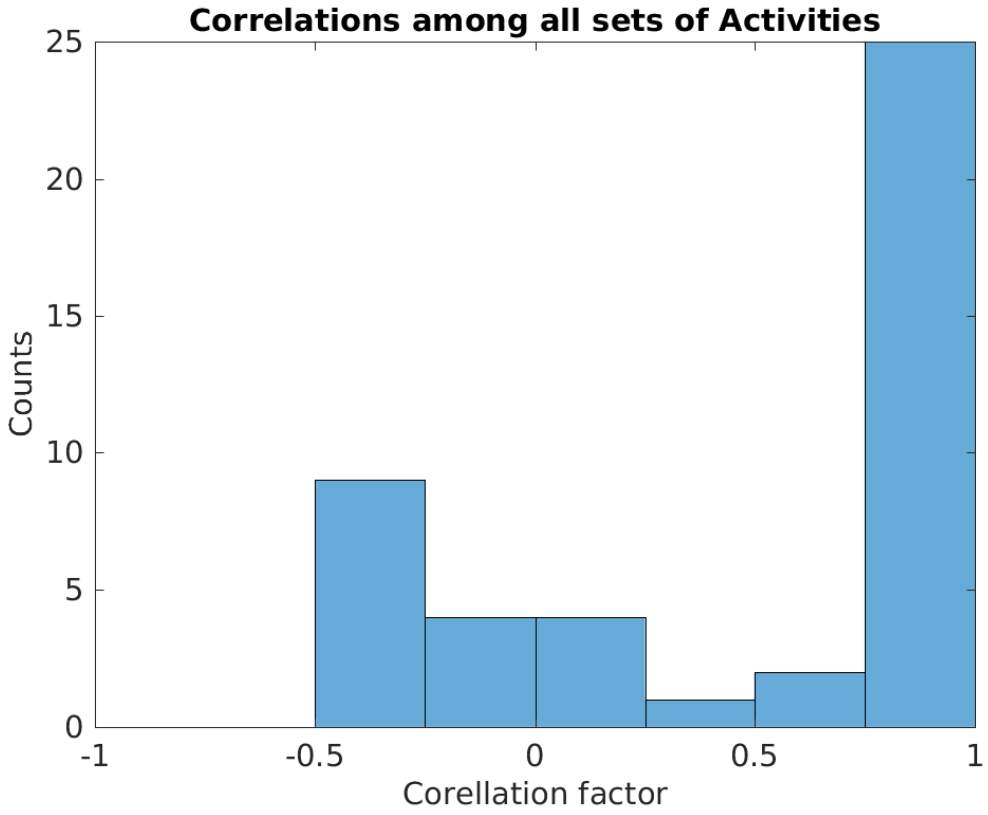
Σχήμα 1.21: Ιστόγραμμα των συντελεστών συσχέτισης για την Austria.

```
The 10 sets of Activities with the highest correlation for Austria are
Correlation         Activity1                  Activity2
 1.00             EnergyIndustries                 EnergyIndustriesPowerProduction1A
 0.91                  Agriculture                        RoadTransport
 0.82            FugitiveEmissions                          OtherEnergy
 0.82                  Agriculture                              Waste
 0.79                  OtherEnergy                              Waste
 0.76                 RoadTransport                             Waste
 0.75         EnergyIndustriesPowerProduction1A1a                      IndustryEner
 0.74         EnergyIndustriesPowerProduction1A1a                    IndustryProcess
 0.74             EnergyIndustries                    IndustryProcesses
 0.72             EnergyIndustries                       IndustryEnergy

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

 51.11% of the sets of Activities pass
the NON-Parametric test with a significance level of  5.00%
```



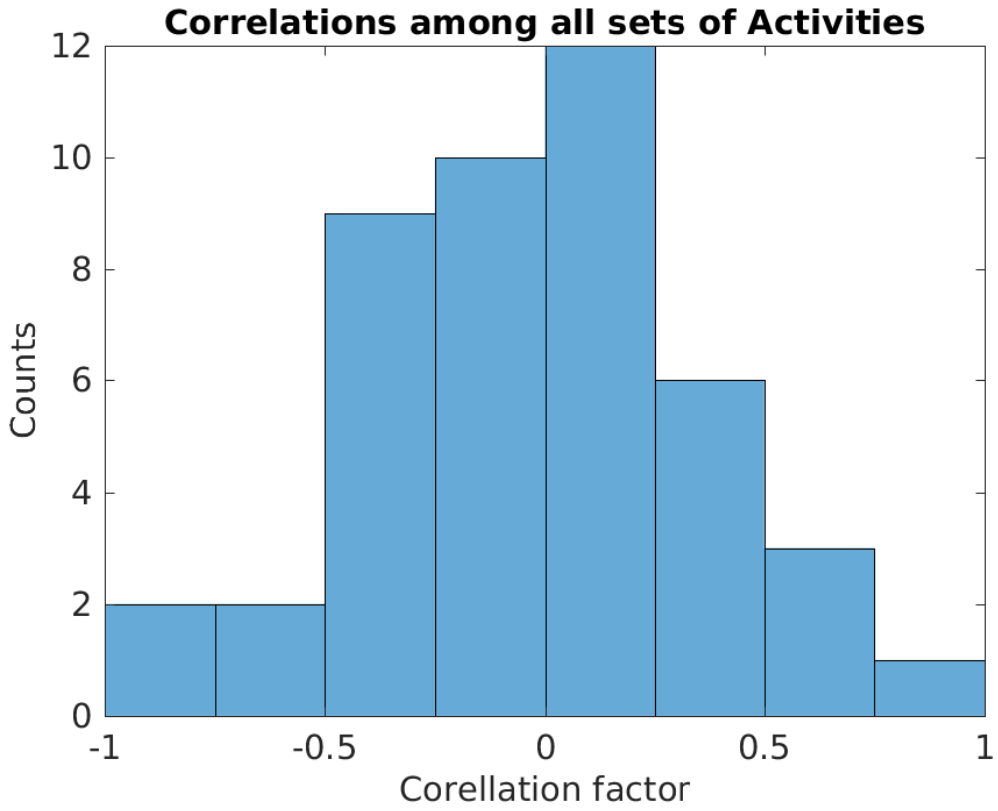
Σχήμα 1.22: Ιστόγραμμα των συντελεστών συσχέτισης για την Spain.

```
The 10 sets of Activities with the highest correlation for Spain are
Correlation         Activity1                  Activity2
 0.99             EnergyIndustries                EnergyIndustriesPowerProduction1A
 0.92               OtherTransport                            Waste
 0.88                RoadTransport                            Waste
 0.86           FugitiveEmissions                     OtherTransport
 0.83              OtherTransport                      RoadTransport
 0.81           FugitiveEmissions                             Waste
 0.79                 Agriculture                   EnergyIndustries
 0.79                 Agriculture                  FugitiveEmissions
 0.77           FugitiveEmissions                      RoadTransport
 0.74                 Agriculture                EnergyIndustriesPowerProduction1A

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

 66.67% 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                 EnergyIndustriesPowerProduction1A
 0.97             EnergyIndustries                        RoadTransport
 0.96         EnergyIndustriesPowerProduction1A1a                          RoadTranspo
 0.96             EnergyIndustries                     IndustryEnergy
 0.96         EnergyIndustriesPowerProduction1A1a                        IndustryEner
 0.95                 Agriculture                   EnergyIndustries
 0.95                 Agriculture                 EnergyIndustriesPowerProduction1A
 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%
```

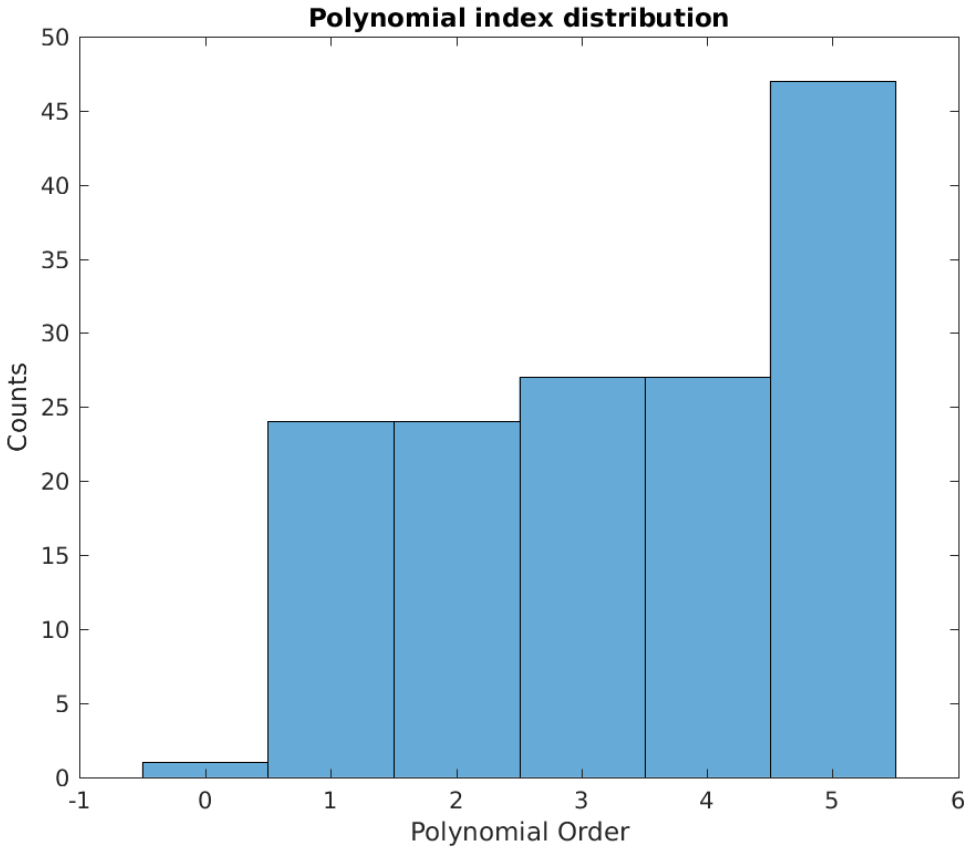


Σχήμα 1.24: Ιστόγραμμα των συντελεστών συσχέτισης για την Greece.

```
The 10 sets of Activities with the highest correlation for Greece are
Correlation         Activity1                  Activity2
 1.00            EnergyIndustries                EnergyIndustriesPowerProduction1A
 0.92         EnergyIndustriesPowerProduction1A1a                        IndustryEner
 0.92            EnergyIndustries                   IndustryEnergy
 0.66            EnergyIndustries                    RoadTransport
 0.65              IndustryEnergy                  IndustryProcesses
 0.62         EnergyIndustriesPowerProduction1A1a                         RoadTranspo
 0.59                Agriculture                   IndustryEnergy
 0.54                Agriculture                  IndustryProcesses
 0.49              IndustryEnergy                    RoadTransport
 0.47                Agriculture                EnergyIndustriesPowerProduction1A

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

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

1.7 Ζήτημα 7

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 16.



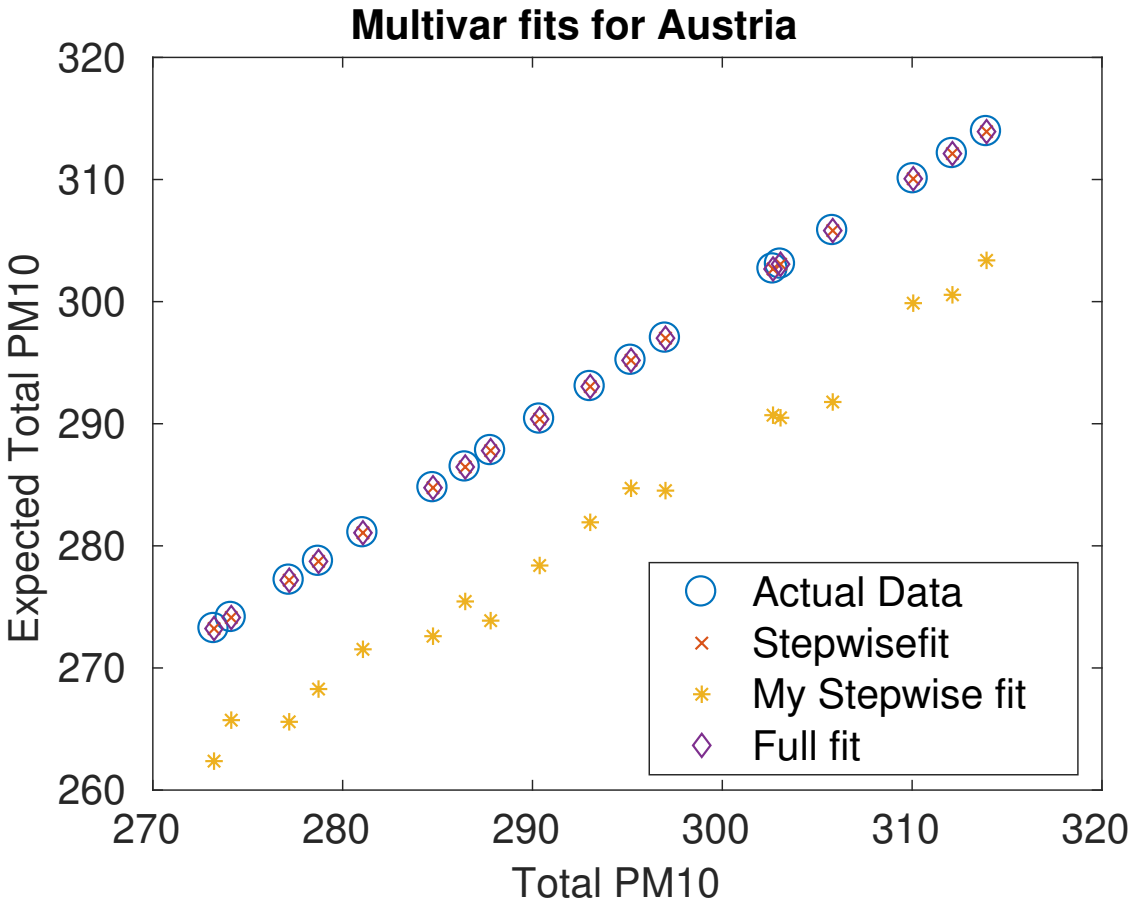
Σχήμα 1.25: Ιστόγραμμα των πολυωνυμικών όρων με το μεγαλύτερο $AdjR^2$.

	AdjR2	Order	Country
Agriculture	0.96671	5	'Netherlands'
EnergyIndustries	0.99606	5	'Italy'
EnergyIndustriesPowerProduction1A1a	0.99382	5	'Italy'
FugitiveEmissions	0.98047	4	'France'
IndustryEnergy	0.99629	4	'Luxembourg'
IndustryProcesses	0.9754	4	'Sweden'
OtherEnergy	0.97918	4	'Germany'
OtherTransport	0.98636	5	'Germany'
RoadTransport	0.99763	3	'United Kingdom'
Waste	0.93096	2	'Luxembourg'

1.8 Ζήτημα 8

Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 19. Παρακάτω δίνονται τα σχήματα και τα αποτελέσματα των βηματικών προσαρμόσεων για κάθε χώρα ξεχωριστά. Στα σχήματα υπάρχει οπτική ανάλυση του πόσο καλά προσαρμόζονται οι εκτιμήσεις στα πραγματικά δεδομένα. Ενώ στο επιστρεφόμενο κείμενο από των κώδικα υπάρχει η πληροφορία για το $AdjR^2$ καθώς και τους συντελεστές της κάθε προσαρμογής. Οι κατηγορίες Stepwizefit, MyStepwise και Full αντιστοιχούν στις προσαρμογές που έγιναν με την stepwisefit function του Matlab, την δική μασ συνάρτηση και το πλήρες μοντέλο αντίστοιχα.

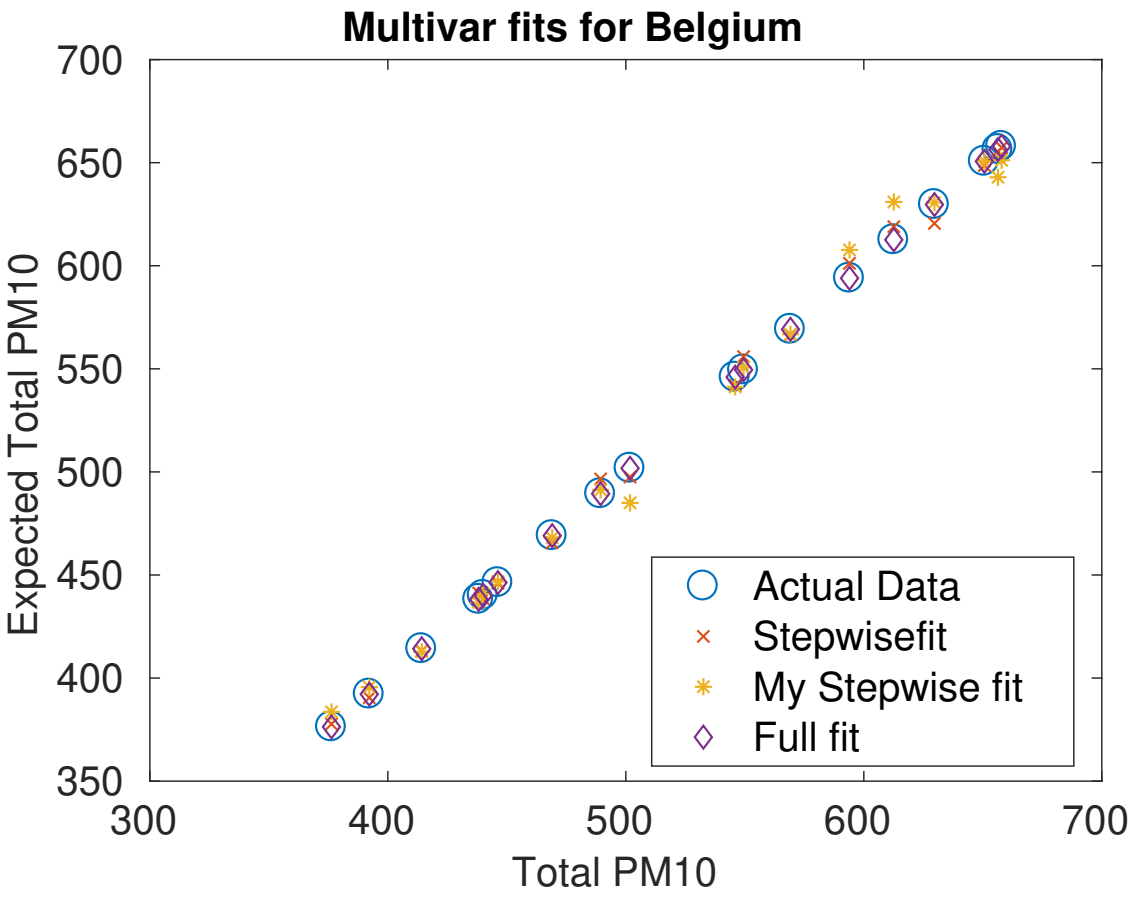
Σε γενικές γραμμές οι προσαρμογές αναπαριστούν τα αρχικά δεδομένα με καλή ακρίβεια. Υπάρχουν μερικά προβλήματα όπως στην περίπτωση της Greece και Ireland όπου υπάρχουν κενά στα δεδομένα με αποτέλεσμα να επηρεάζουν την διαδικασία προσαρμογής.



Σχήμα 1.26: Πραγματικές και εκτιμούμενες τιμές για την Austria.

Multivariable fitting for Activities in Country
Austria

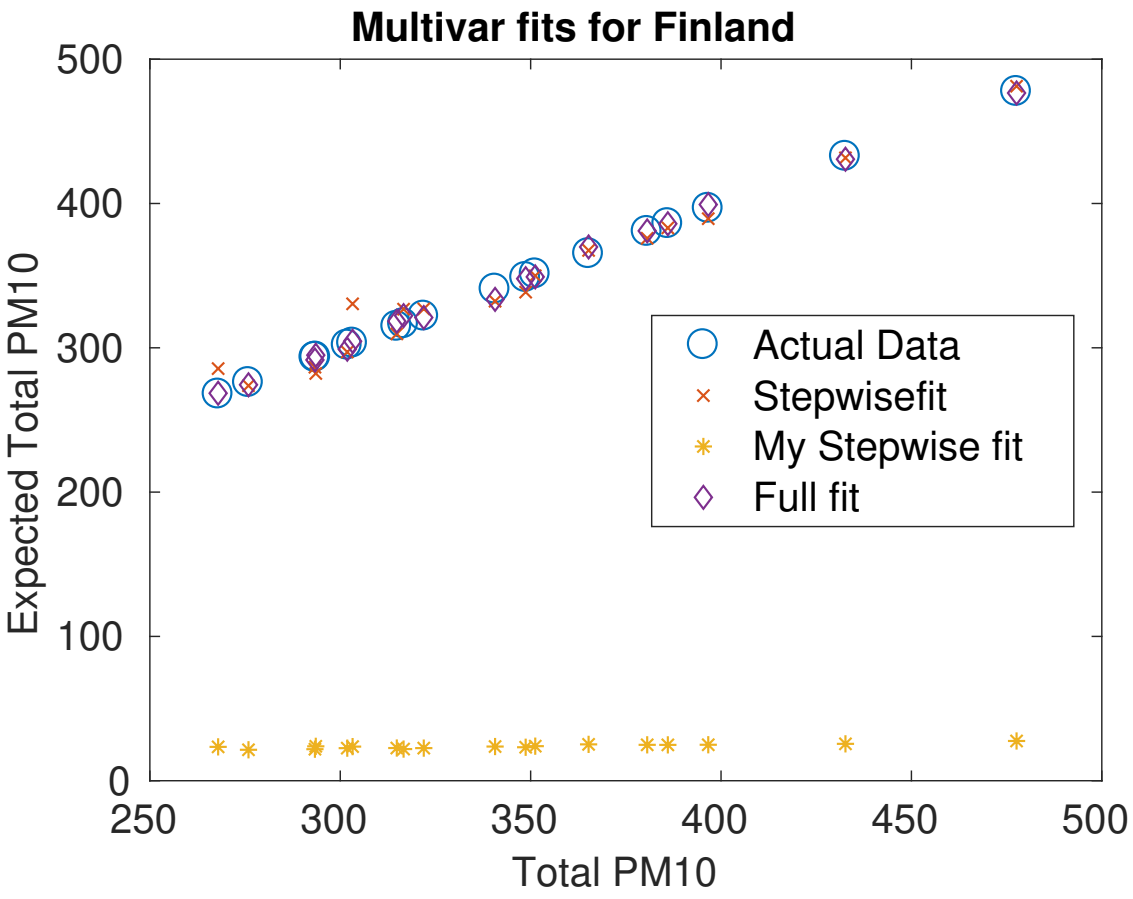
	Stepwisefit	MyStepwise	Full
AdjR2	1	0.98418	1
Intercept	0.57636	125.45	0.57587
Agriculture	0.99855	0	0.99856
EnergyIndustries	1.0001	0	1.0001
EnergyIndustriesPowerProduction1A1a	0	1.8351	-2.3227e-05
FugitiveEmissions	0.99155	0	0.99156
IndustryEnergy	0.99937	0	0.99937
IndustryProcesses	0.99973	0	0.99973
OtherEnergy	1	1.025	1
OtherTransport	0.99631	0	0.99632
RoadTransport	0.99999	0.85459	0.99999
Waste	1.0331	0	1.0331



Σχήμα 1.27: Πραγματικές και εκτιμούμενες τιμές για την Belgium.

```
Multivariable fitting for Activities in Country
                    Belgium

                                                Stepwisefit    MyStepwise      Full
                                                ___________    __________    _________

    AdjR2                                       0.9973         0.99169             1
    Intercept                                     23.9          72.856        0.043911
    Agriculture                                      0               0         0.99905
    EnergyIndustries                            1.2977          1.3598         0.99873
    EnergyIndustriesPowerProduction1A1a              0               0       0.0006567
    FugitiveEmissions                                0               0          1.0159
    IndustryEnergy                                   0               0          1.0068
    IndustryProcesses                           1.9938               0         0.99274
    OtherEnergy                                      0               0          1.0051
    OtherTransport                                   0               0         0.98095
    RoadTransport                               1.7411          2.0741          1.0033
    Waste                                            0               0         0.97523
\end{[fontsize=\small]erbatim}


\begin{figure}[H]

        \centering
```

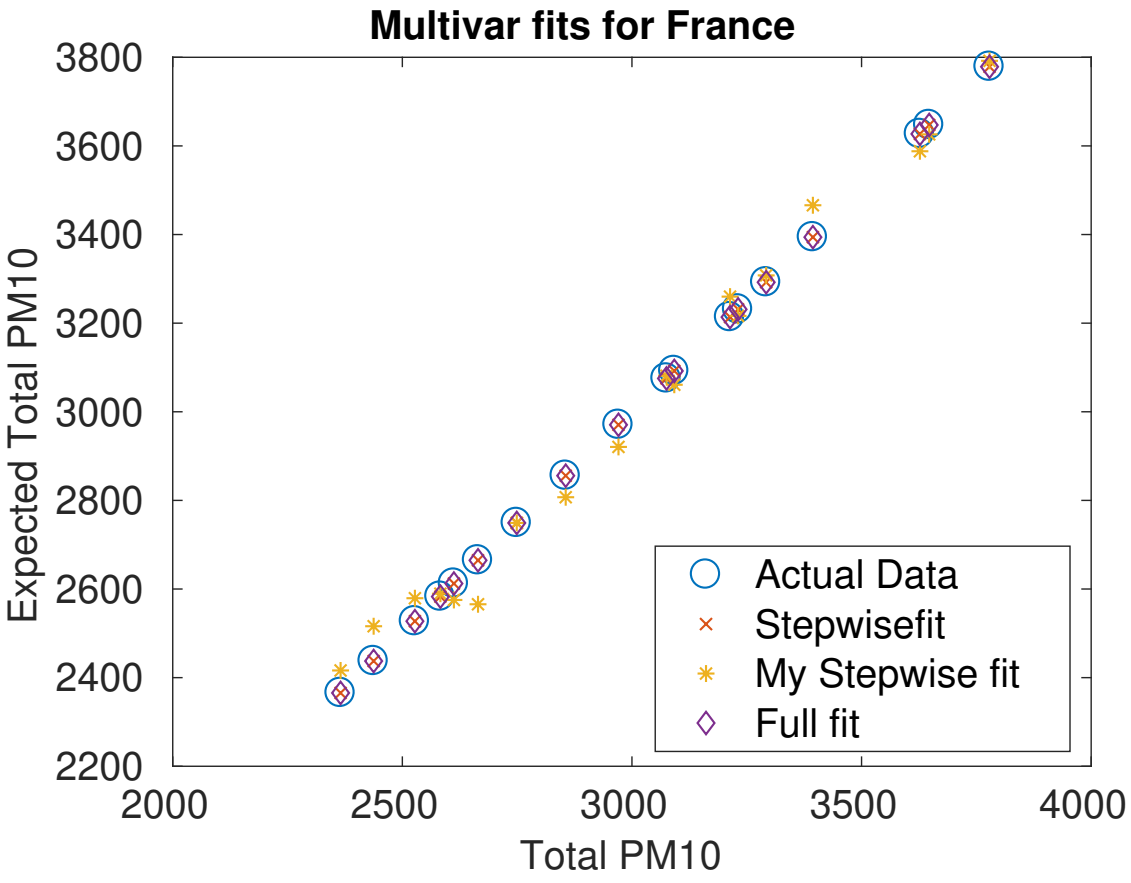
```
        \includegraphics[width=1\columnwidth]{Ex8/Denmark_MultivarFits.eps}
\caption{Πραγματικές και εκτιμούμενες τιμές για την Denmark.}
\label{fig:z83}
\end{figure}


\begin{Verbatim}[fontsize=\small]
Multivariable fitting for Activities in Country
                    Denmark

                                          Stepwisefit    MyStepwise      Full
                                          ___________    __________    _______

    AdjR2                                 0.99992        0.99872       0.99999
    Intercept                              99.167         77.733        25.889
    Agriculture                           0.97628              0       0.97618
    EnergyIndustries                            0              0       0.60969
    EnergyIndustriesPowerProduction1A1a     1.015         1.0301       0.39125
    FugitiveEmissions                           0              0       0.15607
    IndustryEnergy                              0              0       0.74619
    IndustryProcesses                     -6.6942              0       0.12353
    OtherEnergy                                 0              0       0.90714
    OtherTransport                              0              0       0.72493
    RoadTransport                          1.5213         2.1126        1.1612
    Waste                                 -1943.7              0       -579.13
```



Σχήμα 1.28: Πραγματικές και εκτιμούμενες τιμές για την Finland.

Multivariable fitting for Activities in Country
Finland

	Stepwisefit	MyStepwise	Full
AdjR2	0.95853	1	0.99306
Intercept	-94.508	1.146e-12	61.973
Agriculture	12.453	1	12.9
EnergyIndustries	0	-7.0018e-14	-36.27
EnergyIndustriesPowerProduction1A1a	1.0375	3.7329e-14	37.026
FugitiveEmissions	-8.9045	-3.6758e-14	-12.193
IndustryEnergy	0	1.4721e-14	-2.7344
IndustryProcesses	0	-5.0031e-15	1.8925
OtherEnergy	3.1798	-6.584e-15	1.8574
OtherTransport	0	0	-1.8328
RoadTransport	0	5.6423e-15	3.6639
Waste	0	7.2905e-13	-139.65

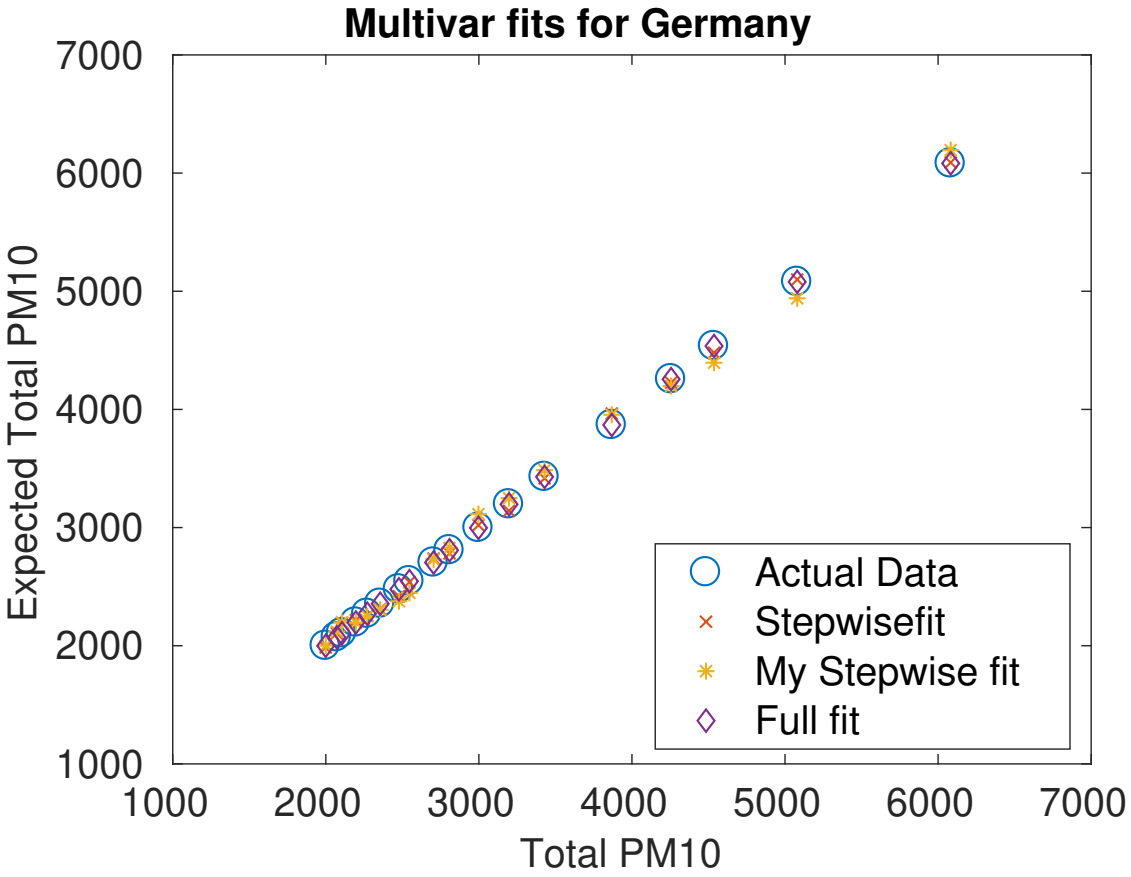


Σχήμα 1.29: Πραγματικές και εκτιμούμενες τιμές για την France.

Multivariable fitting for Activities in Country
France

	Stepwisefit	MyStepwise	Full

AdjR2	1	0.98652	1
Intercept	1.6751	910.65	1.2303
Agriculture	1.0024	0	1.0034
EnergyIndustries	1.0005	0	0.99795
EnergyIndustriesPowerProduction1A1a	0	0	0.0025802
FugitiveEmissions	0.99873	0	0.99905
IndustryEnergy	0.9998	3.2067	1.0002
IndustryProcesses	1.0048	0	1.0015
OtherEnergy	0.99996	3.0522	0.99977
OtherTransport	0.99356	0	0.9952
RoadTransport	1.0003	0	1.0004
Waste	0.98949	0	0.98796

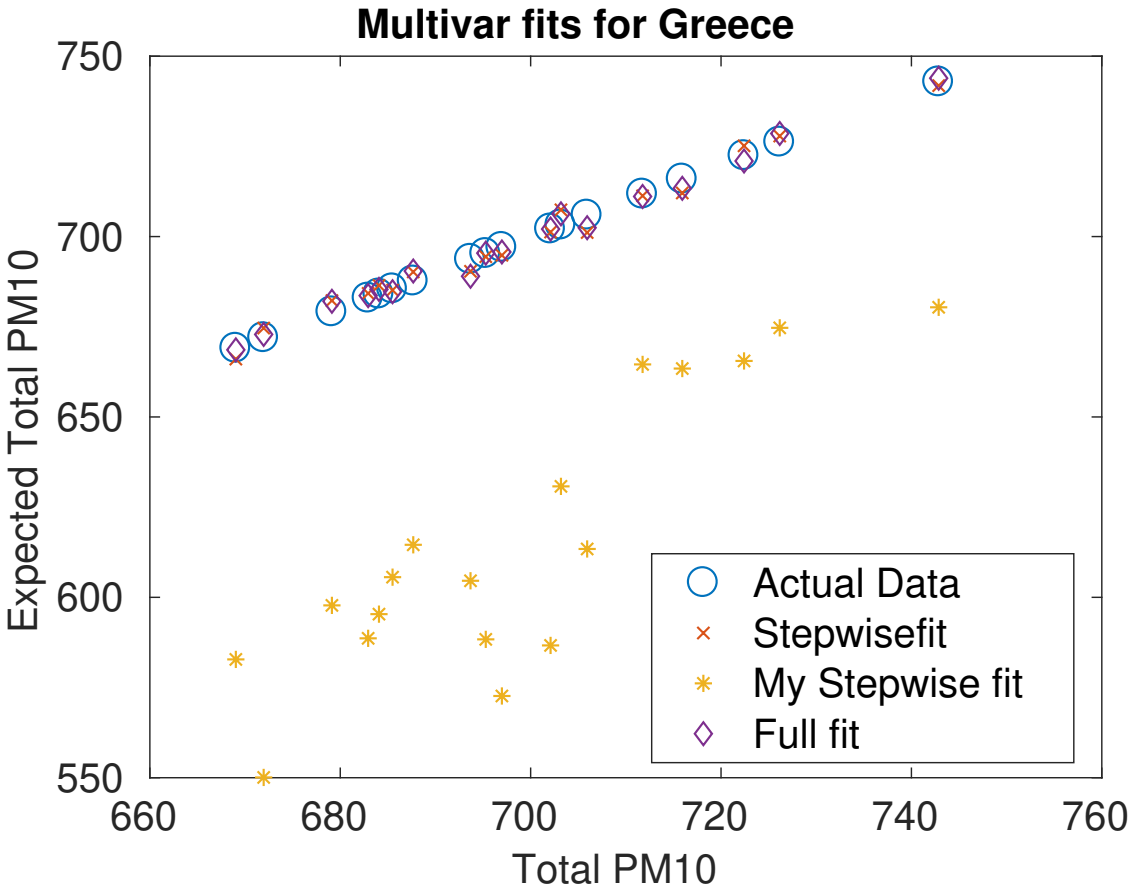


Σχήμα 1.30: Πραγματικές και εκτιμούμενες τιμές για την Germany.

Multivariable fitting for Activities in Country
Germany

	Stepwisefit	MyStepwise	Full
AdjR2	0.99805	0.99425	1
Intercept	307.72	795.46	11.214
Agriculture	0	0	1.0008
EnergyIndustries	0	0	1.0005
EnergyIndustriesPowerProduction1A1a	0	0	-0.0010719
FugitiveEmissions	-31.288	0	1.0111

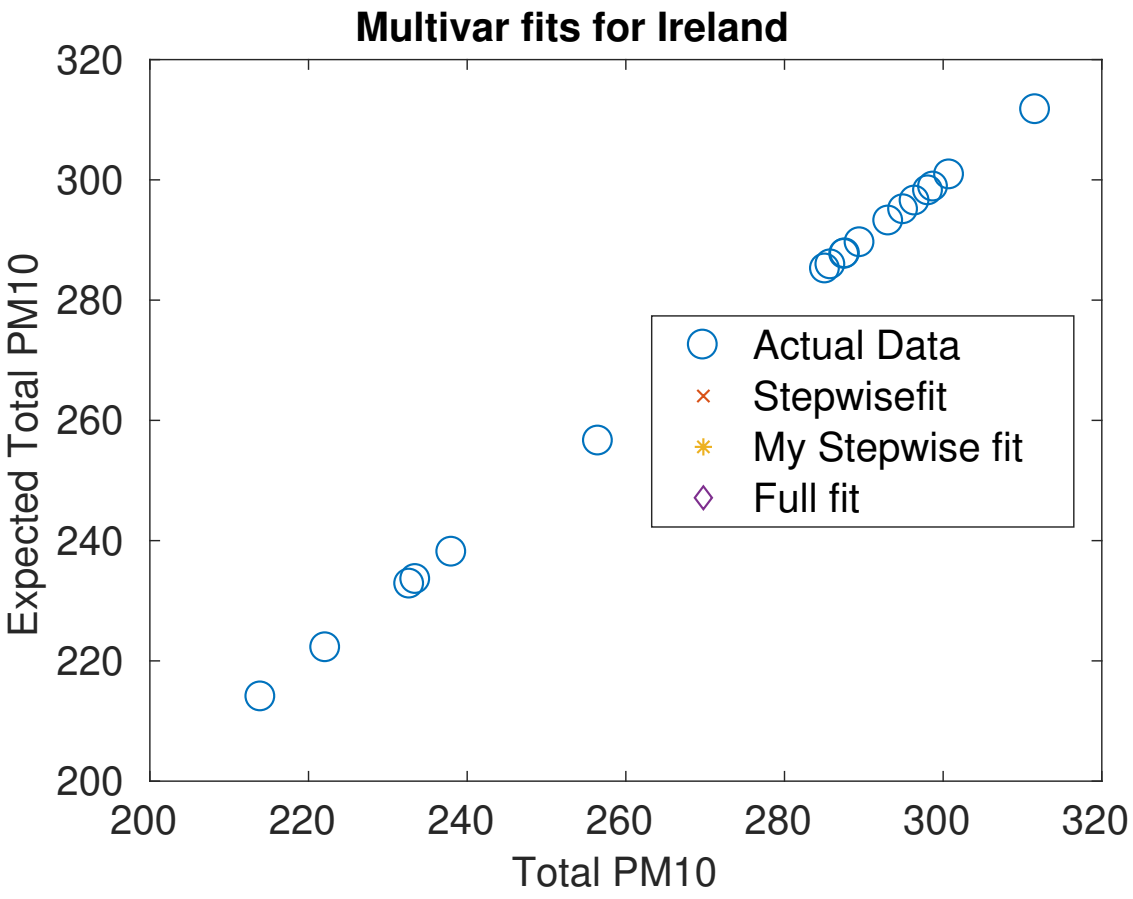
IndustryEnergy	3.8509	3.2964	0.99969
IndustryProcesses	2.5244	0	0.99878
OtherEnergy	0	0	1
OtherTransport	13.477	11.207	1.0002
RoadTransport	0	0	1.0001
Waste	0	0	2.9314



Σχήμα 1.31: Πραγματικές και εκτιμούμενες τιμές για την Greece.

Multivariable fitting for Activities in Country
Greece

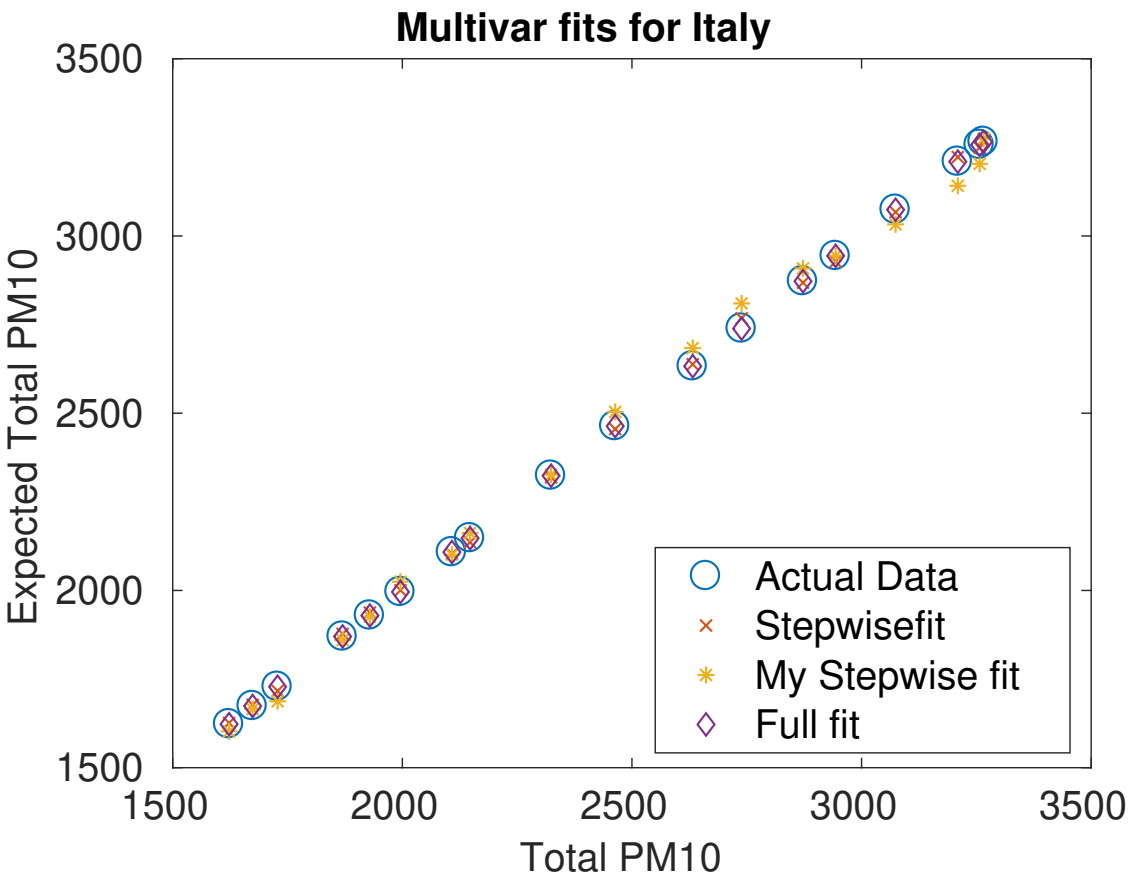
	Stepwisefit	MyStepwise	Full
AdjR2	0.97082	0.89938	0.97
Intercept	48.079	192.78	-67.767
Agriculture	0	0	1.6209
EnergyIndustries	0	0	0.81668
EnergyIndustriesPowerProduction1A1a	1.1346	1.0904	0.28362
FugitiveEmissions	0	0	1.733
IndustryEnergy	0.89349	0	0.92504
IndustryProcesses	4.5881	1.1541	3.9211
OtherEnergy	0.94408	0	0.78521
OtherTransport	0.90138	0.98916	0.8848
RoadTransport	0.68783	0	0.97042
Waste	0	0	0



Σχήμα 1.32: Πραγματικές και εκτιμούμενες τιμές για την Ireland.

Multivariable fitting for Activities in Country
Ireland

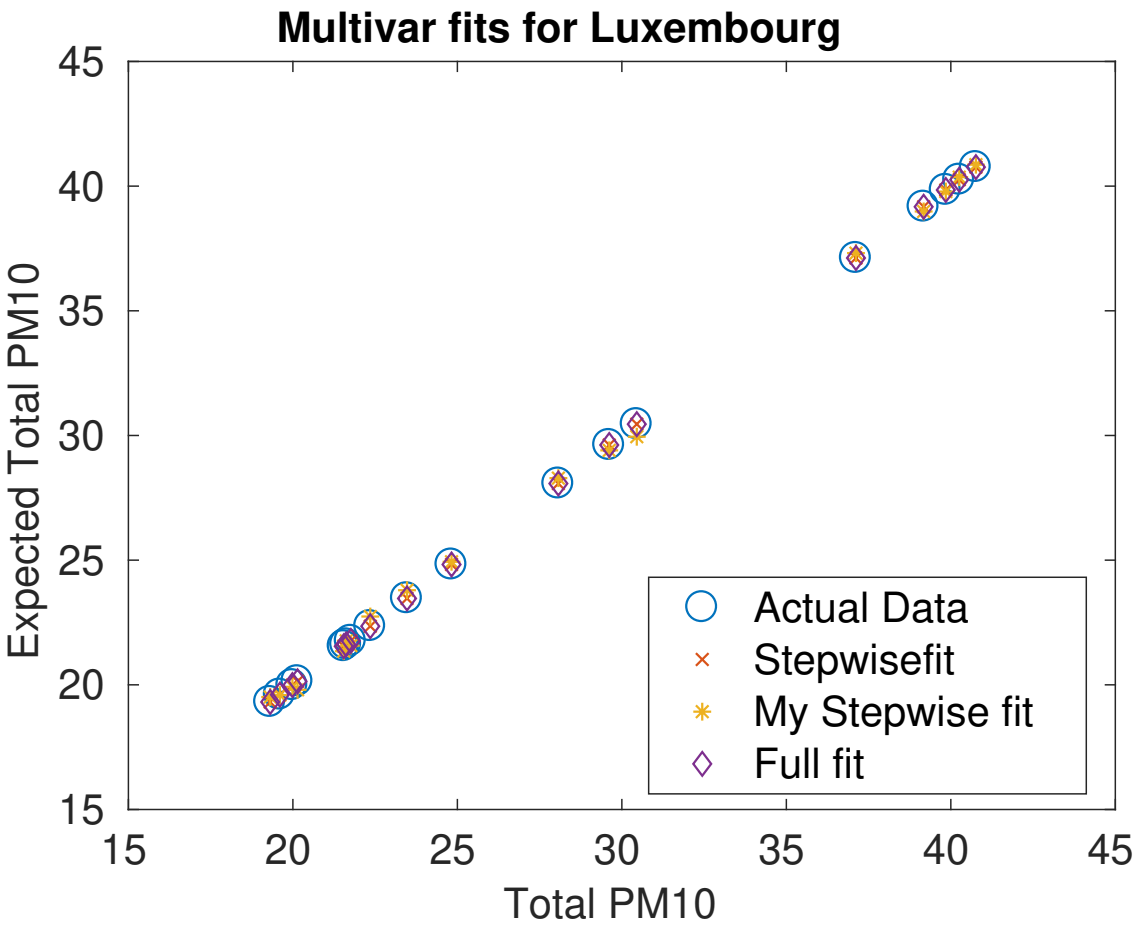
	Stepwisefit	MyStepwise	Full
AdjR2	NaN	NaN	NaN
Intercept	0	0	0
Agriculture	NaN	0	0
EnergyIndustries	NaN	0	0
EnergyIndustriesPowerProduction1A1a	NaN	0	0
FugitiveEmissions	NaN	0	0
IndustryEnergy	NaN	0	0
IndustryProcesses	NaN	0	0
OtherEnergy	NaN	0	0
OtherTransport	NaN	0	0
RoadTransport	NaN	0	0
Waste	NaN	0	0



Σχήμα 1.33: Πραγματικές και εκτιμούμενες τιμές για την Italy.

Multivariable fitting for Activities in Country
Italy

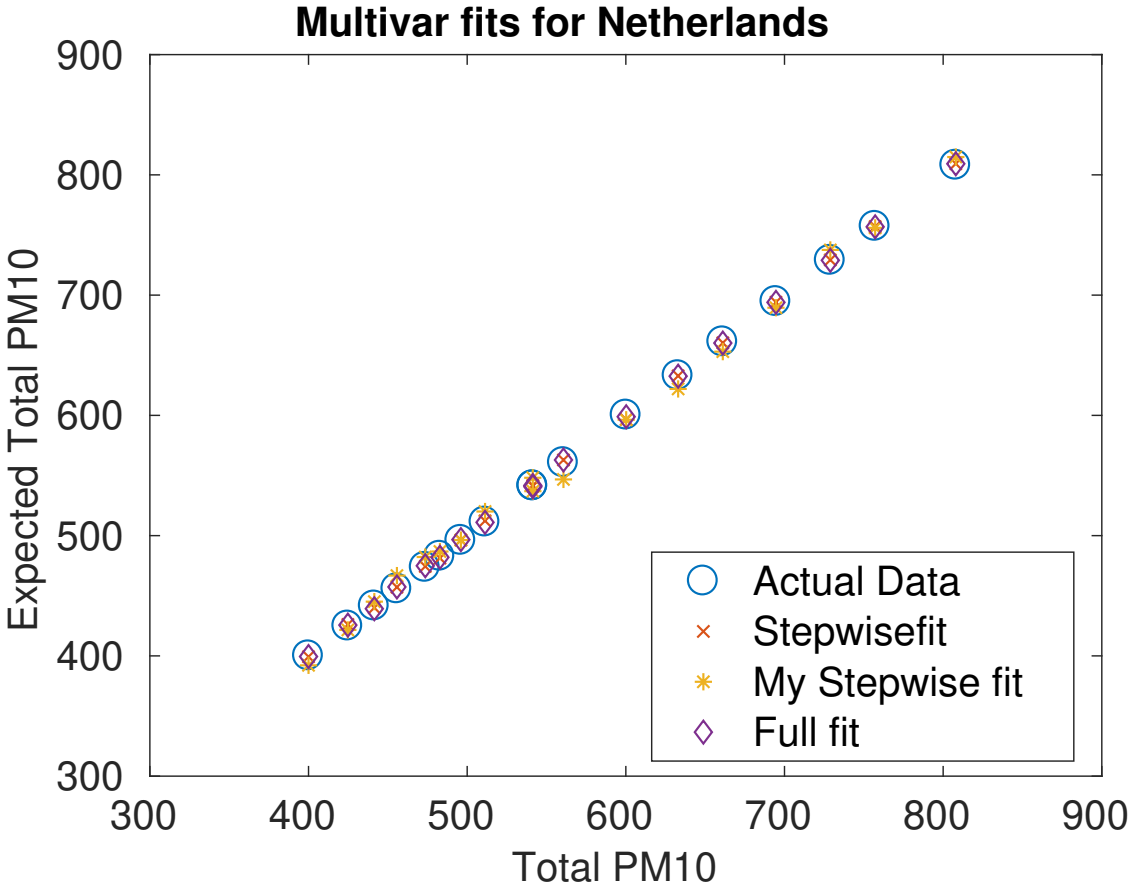
	Stepwisefit	MyStepwise	Full
AdjR2	0.99955	0.99535	1
Intercept	236.84	795.77	-8.4128
Agriculture	0	0	1.0478
EnergyIndustries	1.1083	1.601	1.015
EnergyIndustriesPowerProduction1A1a	0	0	-0.020811
FugitiveEmissions	0	0	0.96018
IndustryEnergy	0.73009	0	0.99561
IndustryProcesses	0	0	0.99623
OtherEnergy	2.4875	0	1.0249
OtherTransport	0	0	1.0093
RoadTransport	1.3897	0.95283	0.99966
Waste	0	0	0.94412



Σχήμα 1.34: Πραγματικές και εκτιμούμενες τιμές για την Luxembourg.

```
Multivariable fitting for Activities in Country
                    Luxembourg

T =

  12×3 table
```

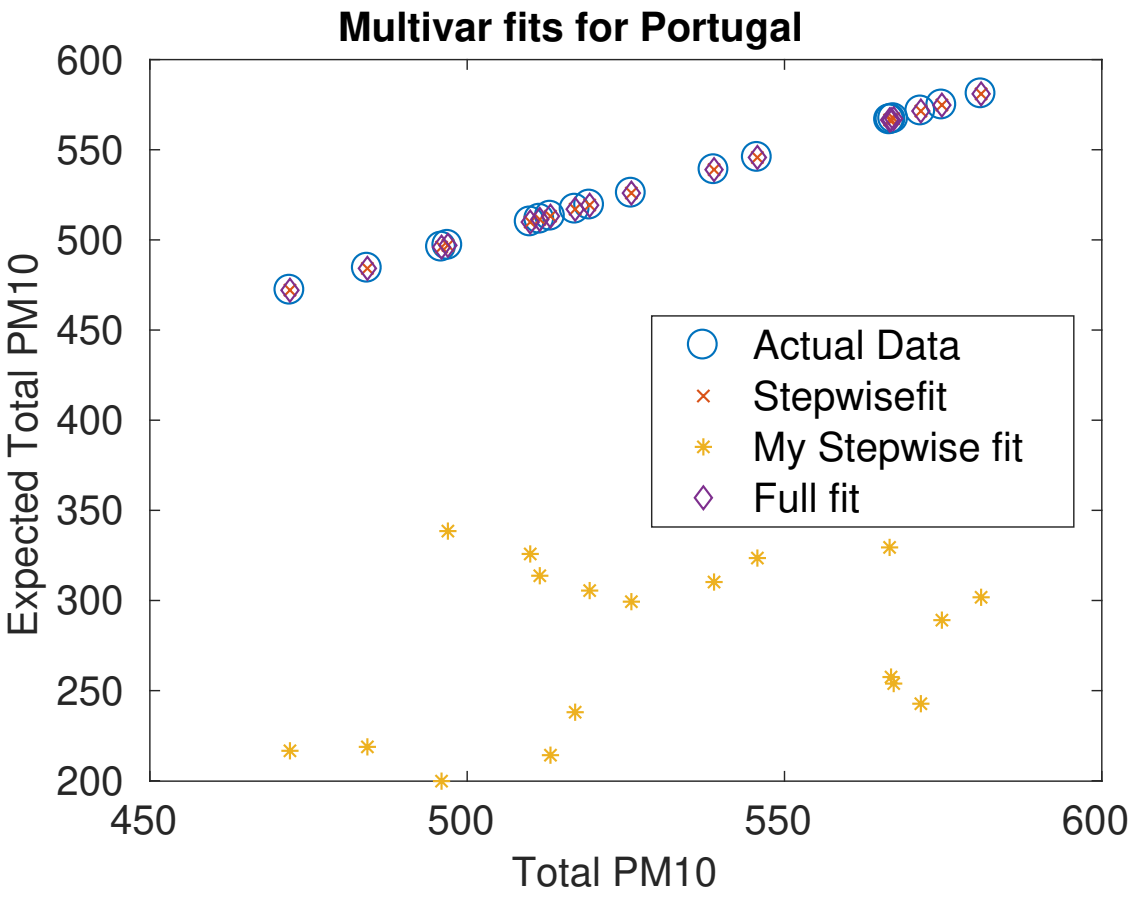
	Stepwisefit	MyStepwise	Full
AdjR2	1	0.99902	1
Intercept	0.44334	8.2746	0.43625
Agriculture	1.0431	0	1.0431
EnergyIndustries	0.98625	0	0.98621
EnergyIndustriesPowerProduction1A1a	0	0	0
FugitiveEmissions	-9.0633	0	-9.0182
IndustryEnergy	1.0004	1.1554	1.0004
IndustryProcesses	1.001	0	1.0011
OtherEnergy	0.99722	0	0.99727
OtherTransport	0.9718	0	0.97168
RoadTransport	1.0058	0.72766	1.0058
Waste	0	0	0.57505



Σχήμα 1.35: Πραγματικές και εκτιμούμενες τιμές για την Netherlands.

Multivariable fitting for Activities in Country
Netherlands

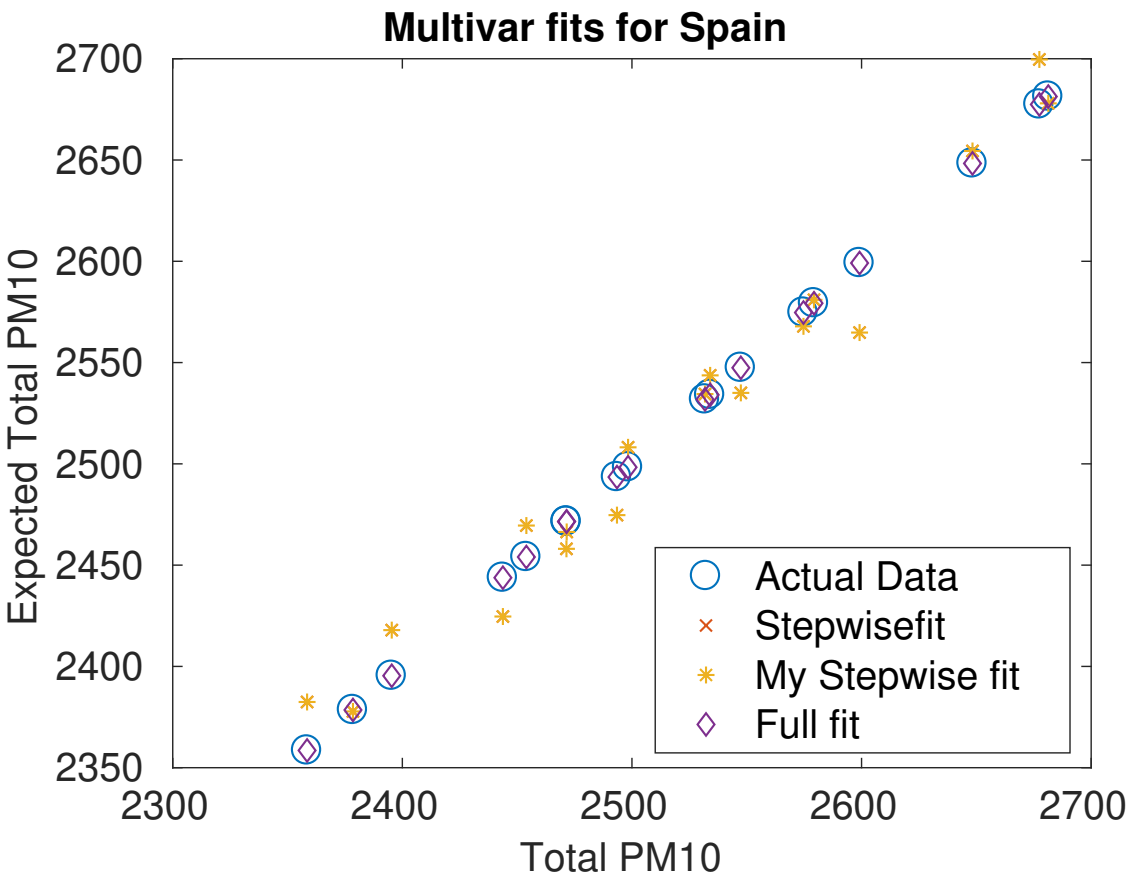
	Stepwisefit	MyStepwise	Full
AdjR2	0.99979	0.99557	0.99979
Intercept	-27.708	2.0338	-24.21
Agriculture	0.72279	0	0.82979
EnergyIndustries	0.96859	0	0.97796
EnergyIndustriesPowerProduction1A1a	0	0	-0.05186
FugitiveEmissions	1.4303	0	1.3791
IndustryEnergy	1.0973	1.2597	1.0776
IndustryProcesses	0.87159	0	0.95956
OtherEnergy	0	0	0.63861
OtherTransport	1.3765	0	1.153
RoadTransport	1.2703	2.944	1.1673
Waste	0	0	1.3544



Σχήμα 1.36: Πραγματικές και εκτιμούμενες τιμές για την Portugal.

Multivariable fitting for Activities in Country
Portugal

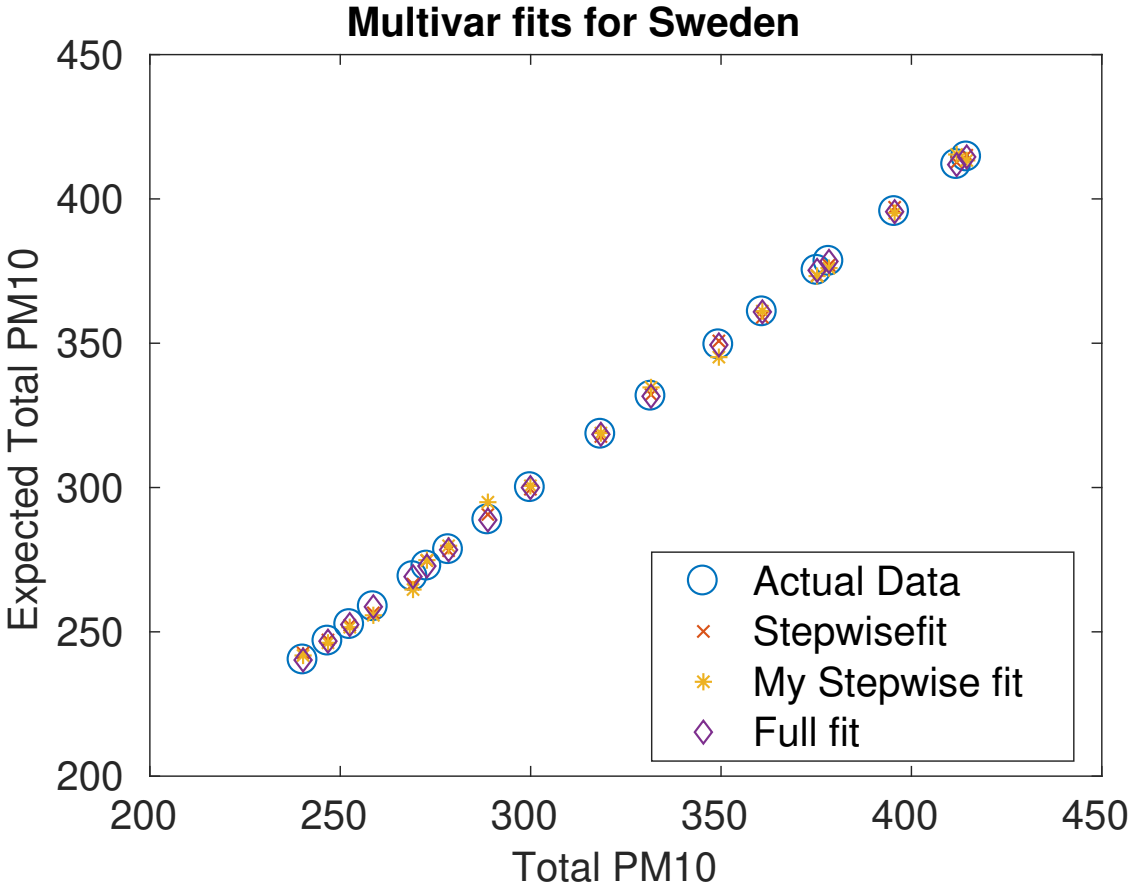
	Stepwisefit	MyStepwise	Full
AdjR2	0.99999	0.86944	1
Intercept	3.4454	117.28	2.8042
Agriculture	0.89938	9.5993	0.93988
EnergyIndustries	0.99466	0.94915	0.98177
EnergyIndustriesPowerProduction1A1a	0	0	0.01454
FugitiveEmissions	0.91417	0	0.97735
IndustryEnergy	1.0218	0	1.0156
IndustryProcesses	0.99696	0	0.99753
OtherEnergy	1.027	0	1.0136
OtherTransport	0.9982	0	0.98878
RoadTransport	1.0002	-3.9099	1.0026
Waste	0.95391	0	0.97481



Σχήμα 1.37: Πραγματικές και εκτιμούμενες τιμές για την Spain.

Multivariable fitting for Activities in Country
Spain

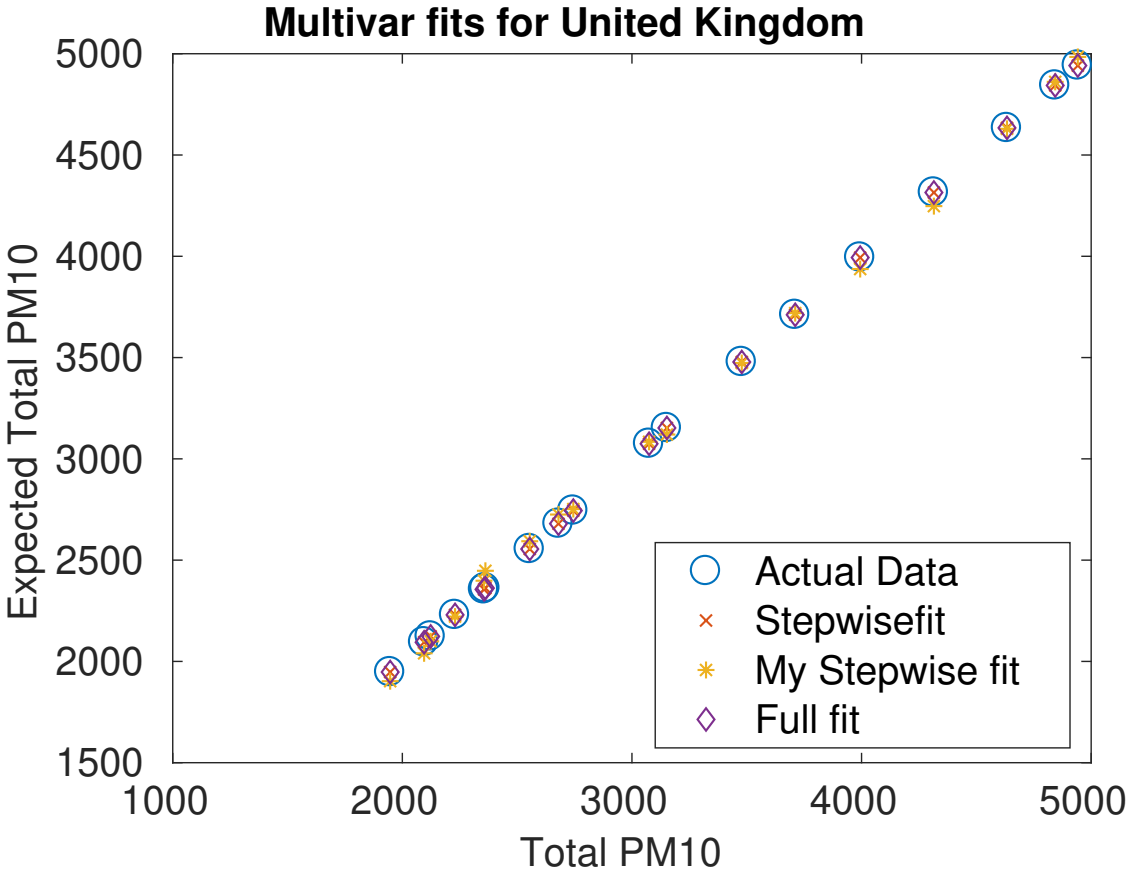
	Stepwisefit	MyStepwise	Full
AdjR2	0.96882	0.96882	1
Intercept	1275.1	1275.1	1.7906e-05
Agriculture	0	0	1
EnergyIndustries	0.91608	0.91608	1
EnergyIndustriesPowerProduction1A1a	0	0	4.7977e-08
FugitiveEmissions	0	0	1
IndustryEnergy	0	0	1
IndustryProcesses	0	0	1
OtherEnergy	5.5264	5.5264	1
OtherTransport	0	0	1
RoadTransport	0	0	1
Waste	0	0	1



Σχήμα 1.38: Πραγματικές και εκτιμούμενες τιμές για την Sweden.

Multivariable fitting for Activities in Country Sweden

	Stepwisefit	MyStepwise	Full
AdjR2	0.9992	0.9976	1
Intercept	56.276	59.476	1.1726
Agriculture	0	0	1.0118
EnergyIndustries	0	0	1.0408
EnergyIndustriesPowerProduction1A1a	0.81219	0	-0.036913
FugitiveEmissions	0	0	0.96369
IndustryEnergy	0	0	0.99073
IndustryProcesses	1.8515	2.2998	0.98606
OtherEnergy	0	0	1.0346
OtherTransport	0	0	0.99516
RoadTransport	1.4246	1.4017	0.99781
Waste	0	0	0.59146



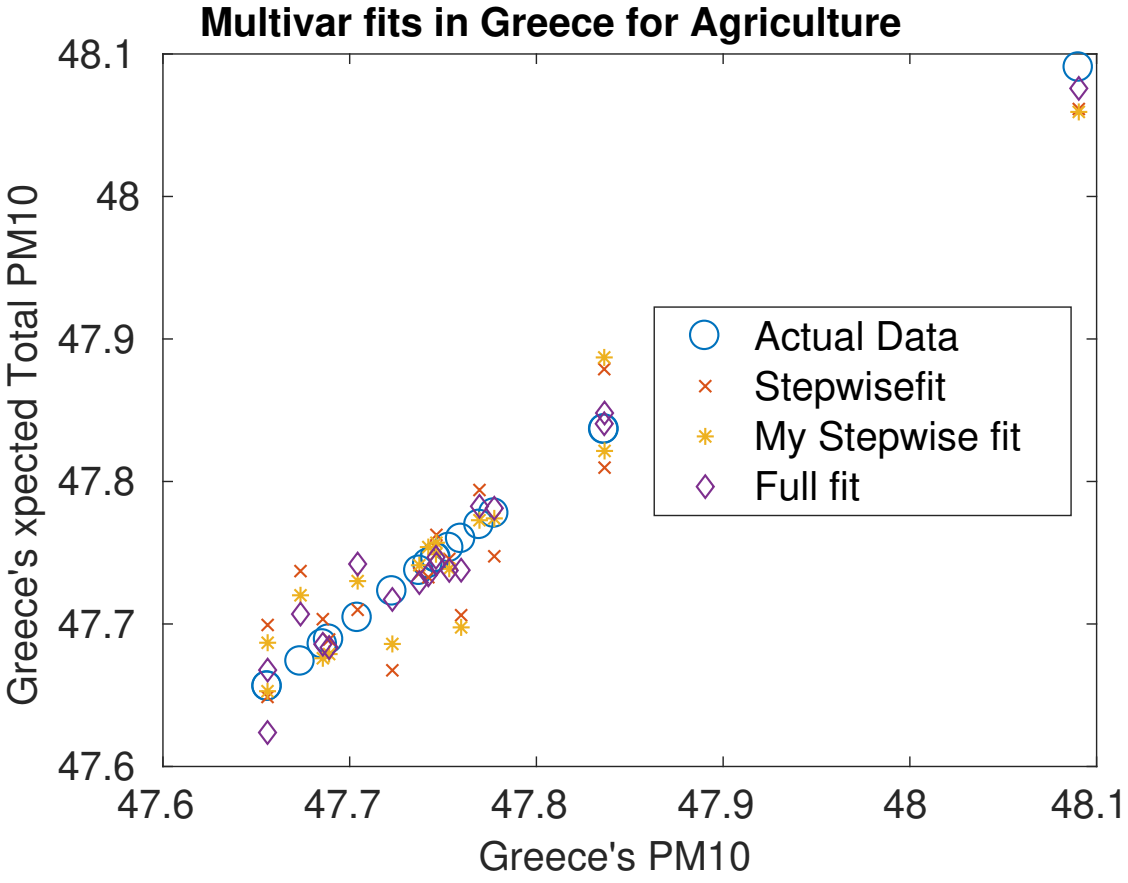
Σχήμα 1.39: Πραγματικές και εκτιμούμενες τιμές για την United Kingdom.

Multivariable fitting for Activities in Country
United Kingdom

	Stepwisefit	MyStepwise	Full
AdjR2	1	0.99806	1
Intercept	4.7657	530.96	11.531
Agriculture	1.0324	0	1.0613
EnergyIndustries	1.0006	1.1866	0.92625
EnergyIndustriesPowerProduction1A1a	0	0	0.074051
FugitiveEmissions	0.97923	0	1.0038
IndustryEnergy	1.0092	0	1.008
IndustryProcesses	1.0837	0	1.1626
OtherEnergy	0.99949	0	1.0115
OtherTransport	0.98561	0	0.96598
RoadTransport	0.98992	1.5858	0.98476
Waste	0.94384	0	0.87251

1.9 Ζήτημα 9

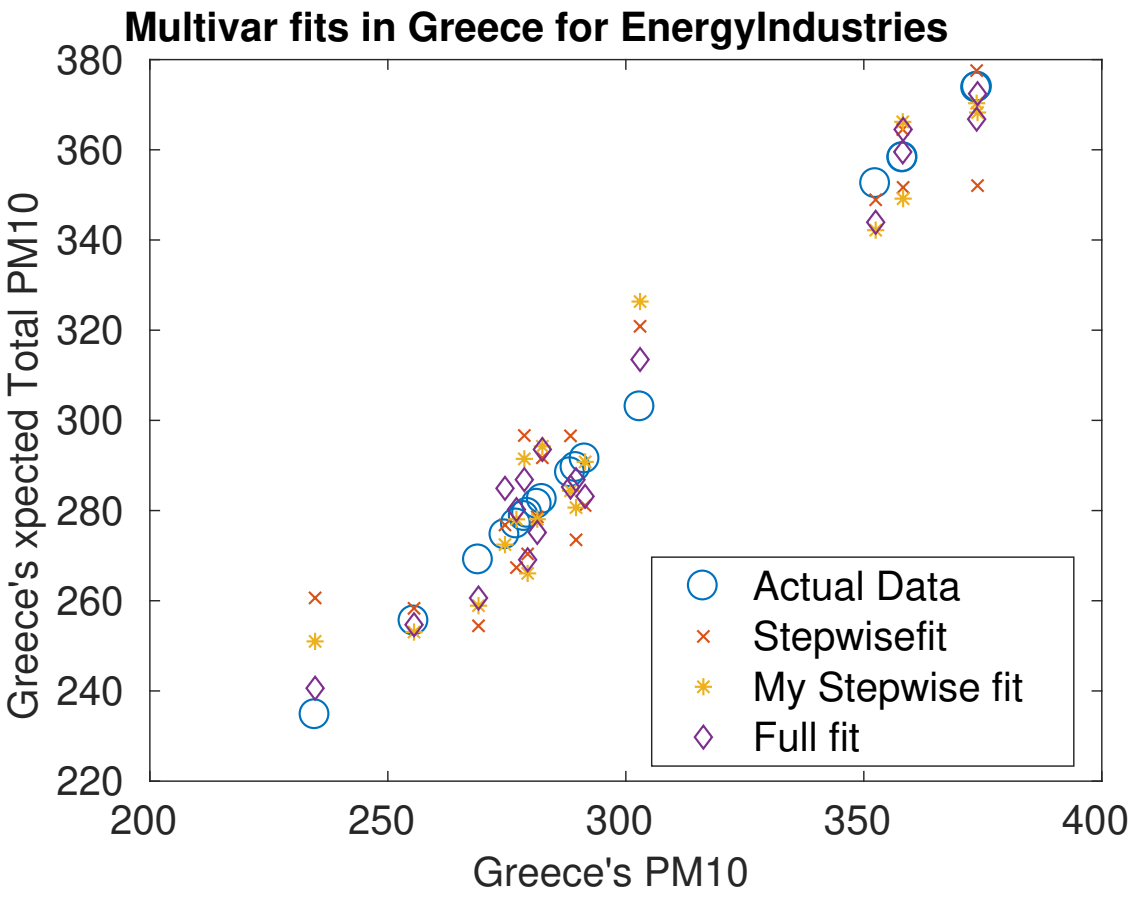
Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 22. Εκτελώντας τον παραπάνω κώδικα για κάθε δραστηριότητα για την Greece, παίρνουμε τα παρακάτω αποτελέσματα.



Σχήμα 1.40: Πραγματικές και εκτιμούμενες τιμές για την Agriculture.

Multivariable fitting in Greece for activity
Agriculture

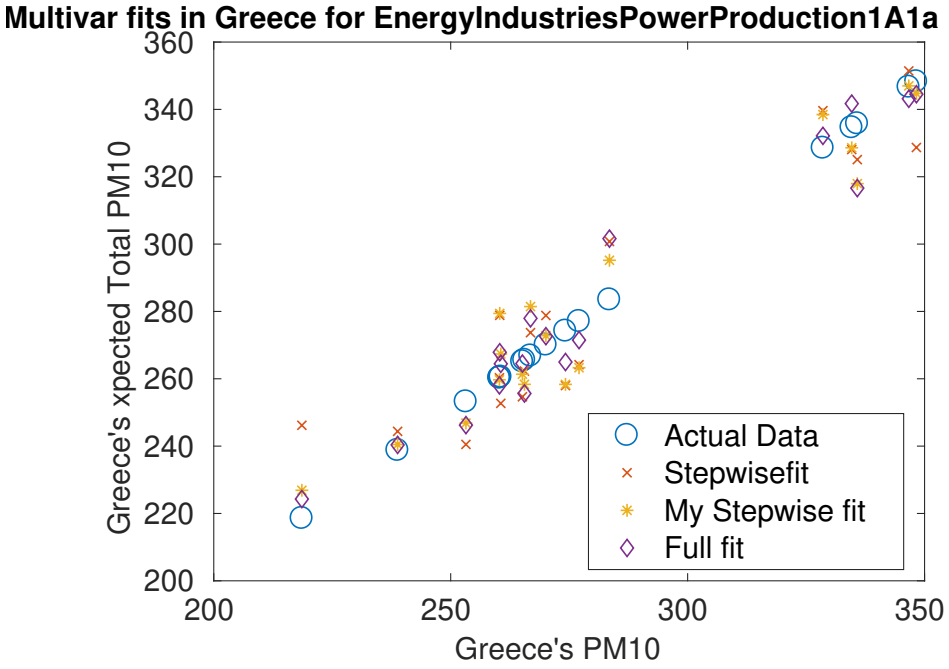
	Stepwisefit	MyStepwise	Full
AdjR2	0.86178	0.88512	0.82286
Intercept	45.596	45.702	47.088
Austria	0	0	0.096335
Belgium	-0.015028	-0.015874	-0.017953
Denmark	0	0.00443	-0.015499
Finland	0	0	0.027114
France	0	0	-0.0072636
Germany	-0.0039902	-0.0041459	-4.3904e-05
Ireland	0	0	0.031979
Italy	0	0	0.0064466
Luxembourg	0.81344	0.85913	-0.79274
Netherlands	0	0	0.0072244
Portugal	0	0	0.07024
Spain	0	0	-0.0046717
Sweden	0	0	-0.048242
United Kingdom	0.01214	0.010109	-0.00068957



Σχήμα 1.41: Πραγματικές και εκτιμούμενες τιμές για την EnergyIndustries.

Multivariable fitting in Greece for activity
EnergyIndustries

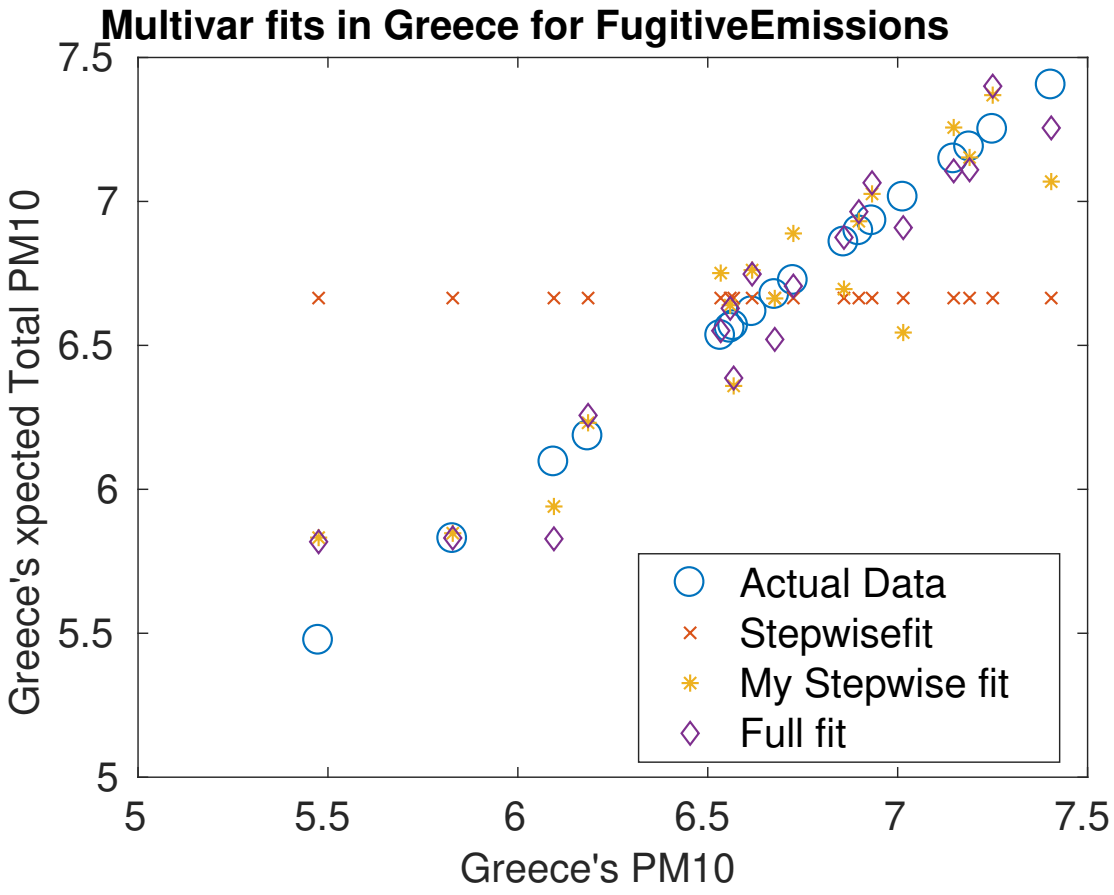
	Stepwisefit	MyStepwise	Full
AdjR2	0.90198	0.92286	0.83018
Intercept	474.51	467.32	296.64
Austria	0	0	1.949
Belgium	0	0	1.9156
Denmark	0	0	0.052545
Finland	0	-0.61138	0.69142
France	0	0	0.25209
Germany	0	0	0.11984
Ireland	-2.0901	-0.9729	-3.9346
Italy	0	0	-0.20664
Luxembourg	0	47.594	-49.602
Netherlands	0	0	0.83065
Portugal	0	0	1.3159
Spain	0	-0.077521	0.02111
Sweden	0	0	0.88024
United Kingdom	0	0	-0.245



Σχήμα 1.42: Πραγματικές και εκτιμούμενες τιμές για την EnergyIndustriesPowerProduction1A1a.

Multivariable fitting in Greece for activity
EnergyIndustriesPowerProduction1A1a

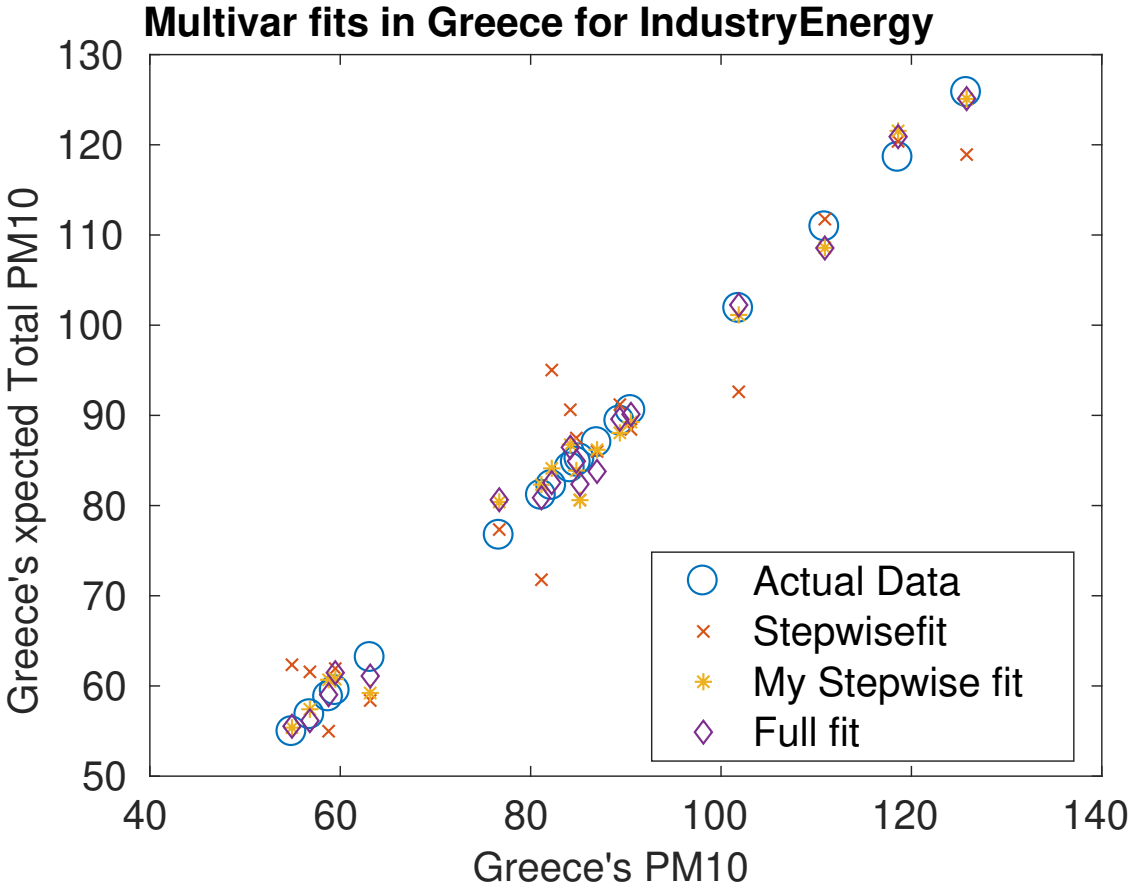
	Stepwisefit	MyStepwise	Full
AdjR2	0.87445	0.89708	0.71327
Intercept	434.55	387.12	259.41
Austria	0	0	-1.088
Belgium	0	0.73171	3.2609
Denmark	0	0	0.6597
Finland	0	-0.39525	1.8287
France	0	0	0.16972
Germany	0	0	0.20092
Ireland	-1.8594	-1.3344	-4.7954
Italy	0	-0.14575	-0.37673
Luxembourg	0	0	-91.87
Netherlands	0	0	0.79198
Portugal	0	0	2.0357
Spain	0	0	0.13377
Sweden	0	1.8285	-1.9564
United Kingdom	0	0	-0.28303



Σχήμα 1.43: Πραγματικές και εκτιμούμενες τιμές για την FugitiveEmissions.

Multivariable fitting in Greece for activity FugitiveEmissions

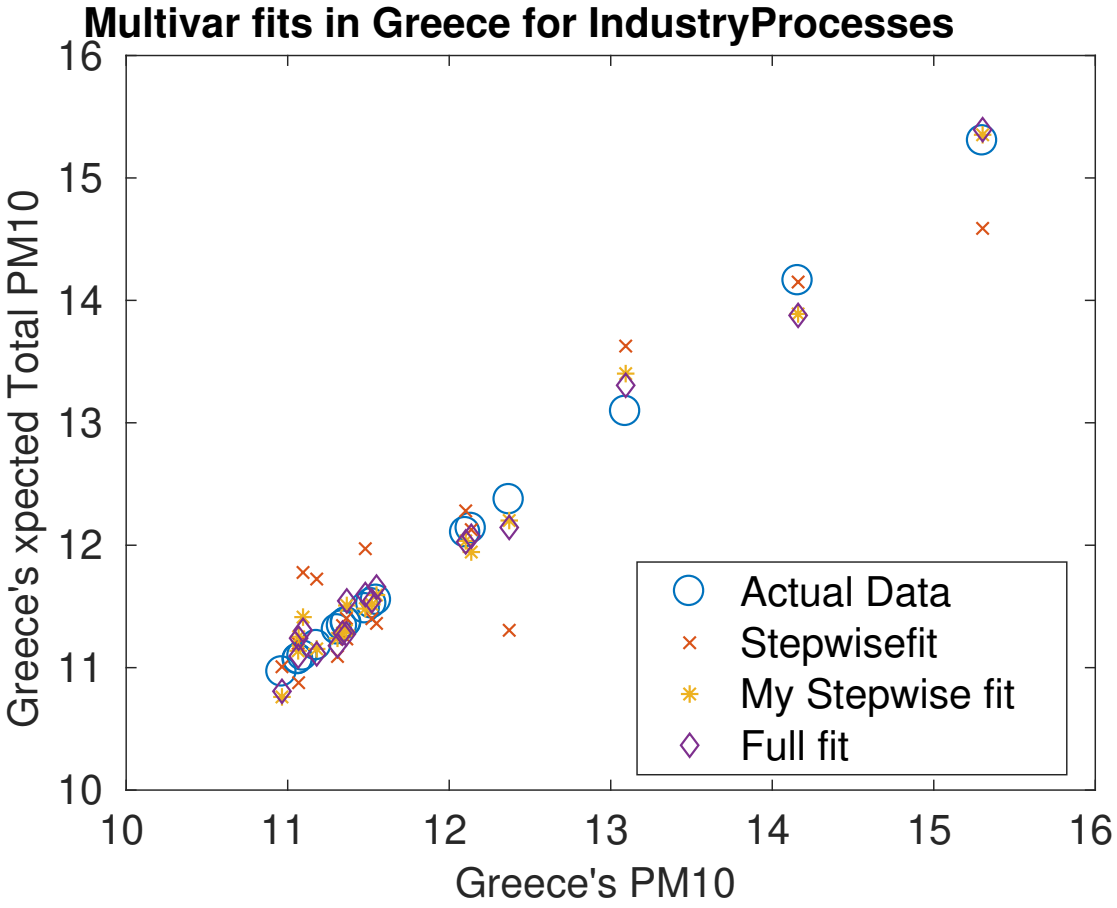
	Stepwisefit	MyStepwise	Full
AdjR2	0	0.77567	0.53648
Intercept	6.6644	9.592	10.814
Austria	0	0.99219	1.6424
Belgium	0	0	-0.092541
Denmark	0	0	0.76422
Finland	0	0.076773	0.085452
France	0	0	0.022749
Germany	0	0	0.016665
Ireland	0	0	-32.693
Italy	0	0	0.032679
Luxembourg	0	0	-6.5289
Netherlands	0	0	-0.01197
Portugal	0	-0.58438	-0.62403
Spain	0	0.041141	0.0066155
Sweden	0	0	-0.36538
United Kingdom	0	-0.18922	-0.3308



Σχήμα 1.44: Πραγματικές και εκτιμούμενες τιμές για την IndustryEnergy.

Multivariable fitting in Greece for activity
IndustryEnergy

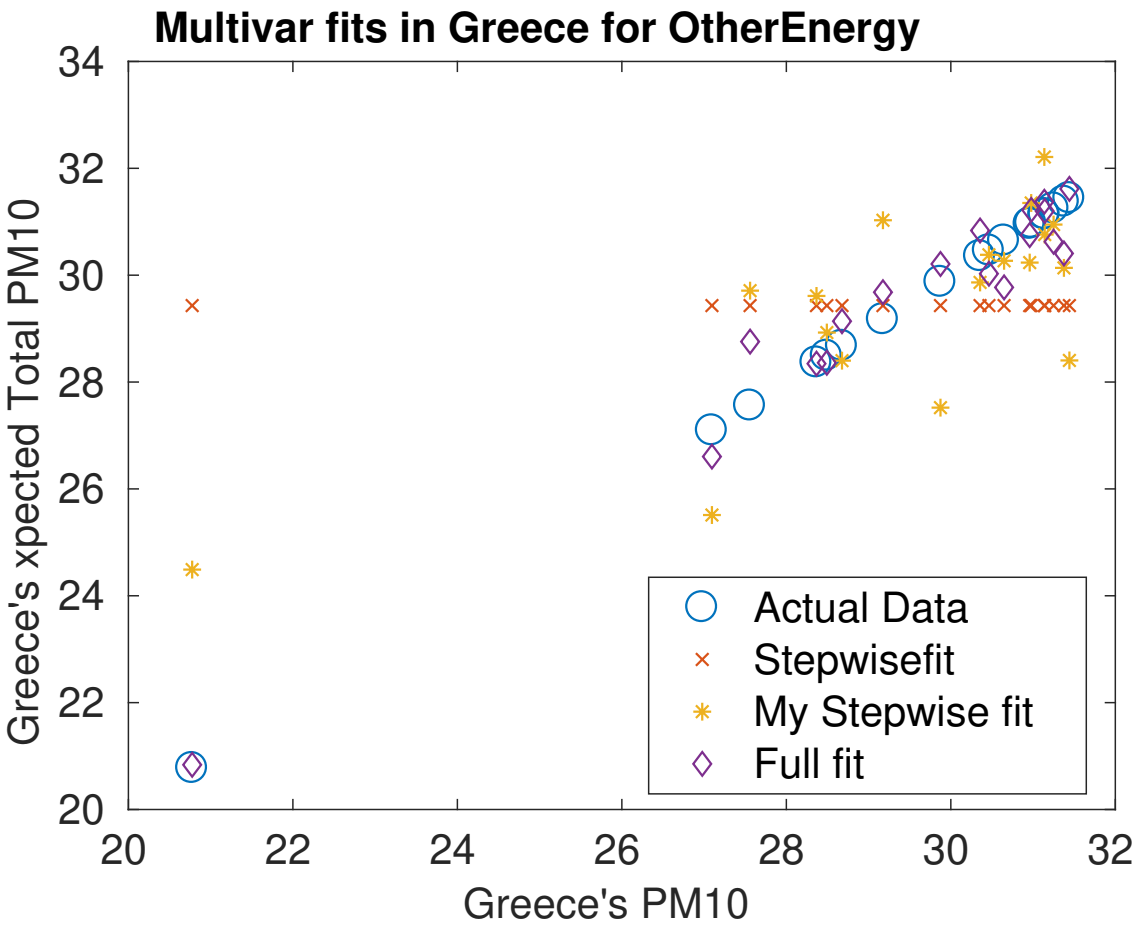
	Stepwisefit	MyStepwise	Full
AdjR2	0.90343	0.97125	0.95425
Intercept	-119.16	-35.916	119.11
Austria	0	0	2.1411
Belgium	0	2.0229	1.8359
Denmark	0	0	-5.0327
Finland	0.37344	-0.015284	-0.1505
France	0.35764	0.017361	0.12145
Germany	0	0.03927	0.12252
Ireland	0	-0.19345	-1.042
Italy	0	0	-0.093878
Luxembourg	0	2.9127	-3.5578
Netherlands	0	0	-1.2819
Portugal	0.63502	0.73275	0.54247
Spain	0	-0.16366	-0.52836
Sweden	0	1.4636	1.8461
United Kingdom	0	-0.39097	0.10122



Σχήμα 1.45: Πραγματικές και εκτιμούμενες τιμές για την IndustryProcesses.

Multivariable fitting in Greece for activity IndustryProcesses

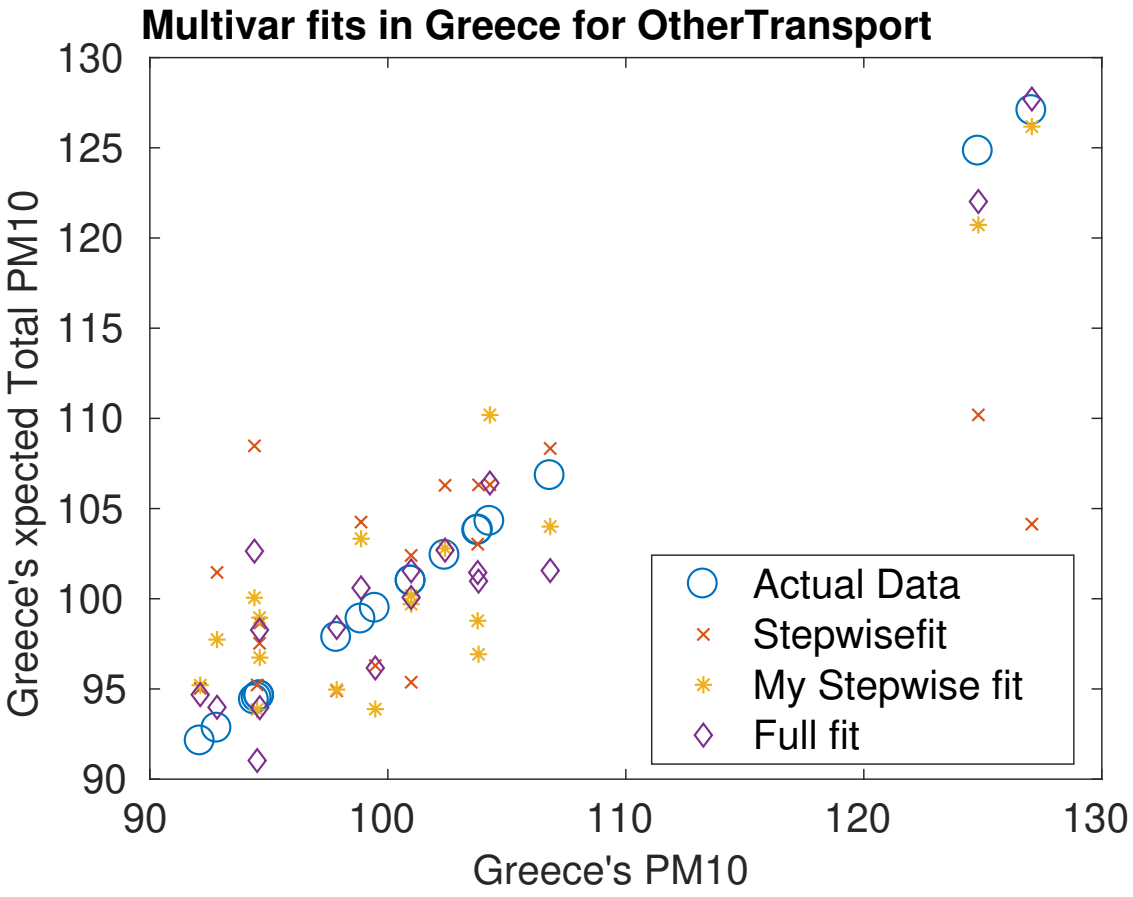
	Stepwisefit	MyStepwise	Full
AdjR2	0.8488	0.94484	0.9072
Intercept	3.5538	5.9051	4.7513
Austria	0.33126	0.29805	0.45469
Belgium	0	0.13114	0.14683
Denmark	0	0	0.40205
Finland	0.14777	-0.046504	-0.0028813
France	0	-0.14279	-0.15543
Germany	0	0.022111	0.021385
Ireland	0	0	-0.65945
Italy	0	-0.026603	-0.031864
Luxembourg	0	3.6695	3.6702
Netherlands	0	-0.053754	-0.03875
Portugal	0	0.019561	0.024573
Spain	0	0.34451	0.30929
Sweden	0	0	-0.0072315
United Kingdom	0	-0.092307	-0.096864



Σχήμα 1.46: Πραγματικές και εκτιμούμενες τιμές για την OtherEnergy.

Multivariable fitting in Greece for activity
OtherEnergy

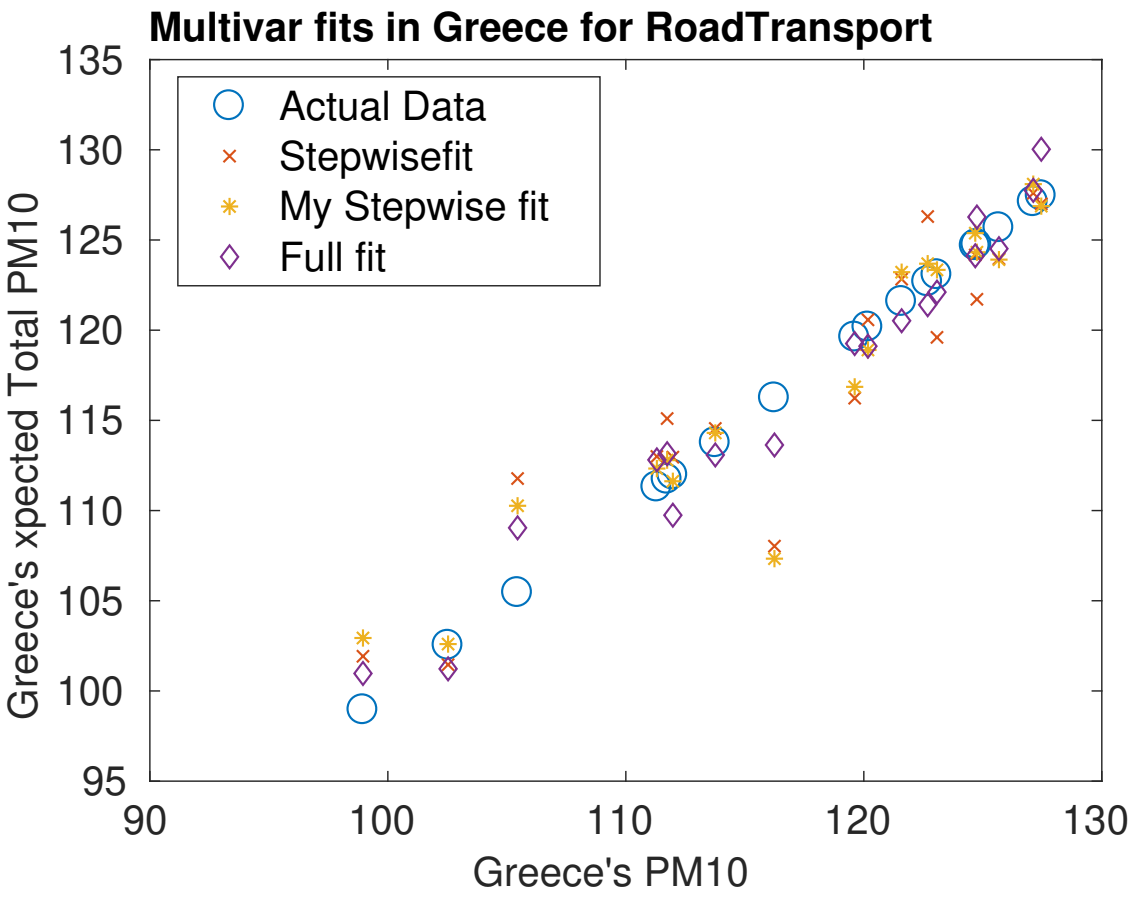
	Stepwisefit	MyStepwise	Full
AdjR2	0	0.43077	0.74738
Intercept	29.431	63.618	56.321
Austria	0	-0.73635	-1.4445
Belgium	0	0.95897	1.903
Denmark	0	-0.58404	-1.8441
Finland	0	0	0.48462
France	0	0	-0.19306
Germany	0	0.0010798	-0.076534
Ireland	0	0	3.0946
Italy	0	0	0.4919
Luxembourg	0	-32.7	-39.493
Netherlands	0	0	0.53744
Portugal	0	0	-0.69296
Spain	0	0	0.047386
Sweden	0	0	-0.64824
United Kingdom	0	0	0.050059



Σχήμα 1.47: Πραγματικές και εκτιμούμενες τιμές για την OtherTransport.

Multivariable fitting in Greece for activity
OtherTransport

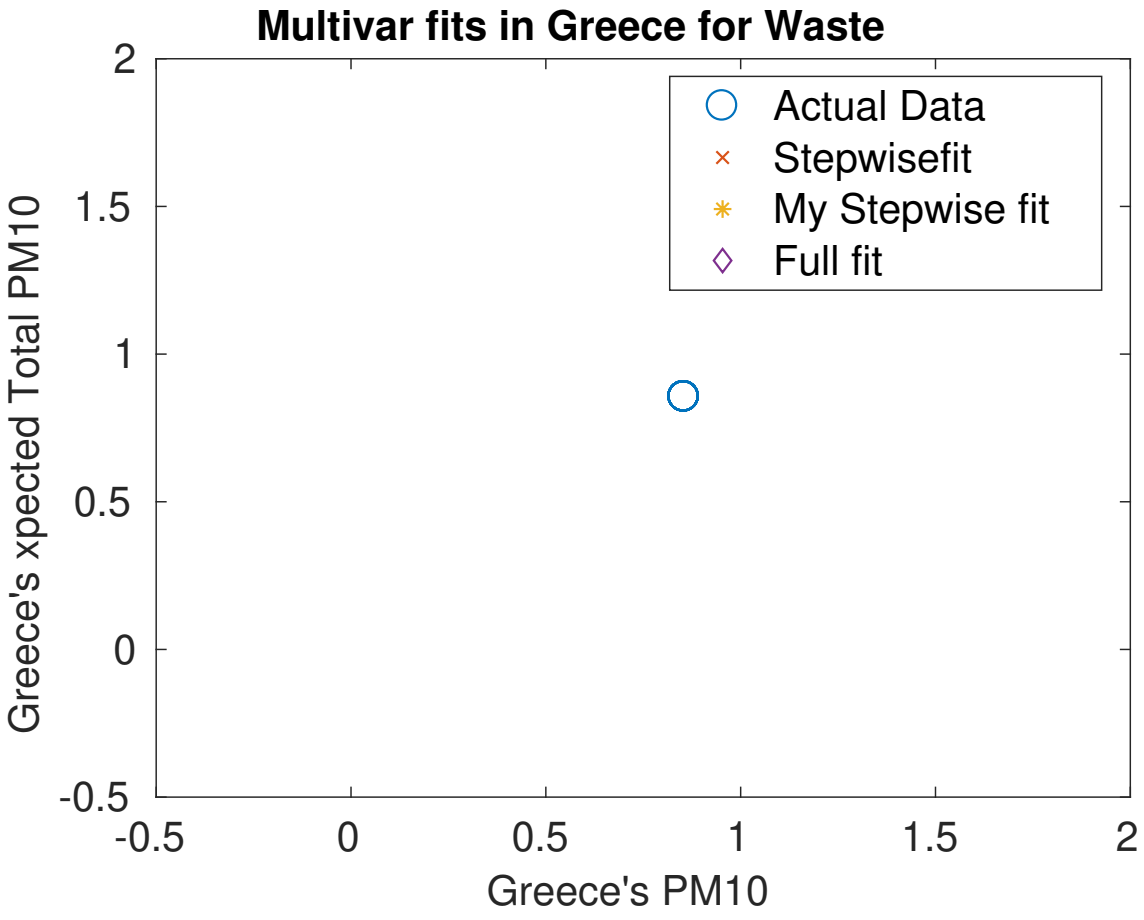
	Stepwisefit	MyStepwise	Full
AdjR2	0.244	0.70262	0.40267
Intercept	85.82	72.405	591.2
Austria	0	0	-6.0897
Belgium	0	1.2661	2.106
Denmark	0	0	-4.3533
Finland	0	0.91441	1.3149
France	0	0.83505	0.91368
Germany	0	0	0.071008
Ireland	0	0	6.0115
Italy	0	0	-0.4334
Luxembourg	82.645	37.353	14.79
Netherlands	0	0	-1.1693
Portugal	0	-2.6332	-1.0819
Spain	0	0	-0.92951
Sweden	0	-4.6855	-6.1015
United Kingdom	0	-0.10652	-0.081946



Σχήμα 1.48: Πραγματικές και εκτιμούμενες τιμές για την RoadTransport.

Multivariable fitting in Greece for activity RoadTransport

	Stepwisefit	MyStepwise	Full
AdjR2	0.83639	0.85769	0.77331
Intercept	10.724	38.682	42.24
Austria	0	0	0.33988
Belgium	0	0	0.63679
Denmark	0	0	-1.3146
Finland	0	0	0.13043
France	0	0	0.089271
Germany	0	0	0.0024107
Ireland	0	0	-0.025922
Italy	0	0	-0.14636
Luxembourg	0	0	-0.43498
Netherlands	0	-0.20742	-0.48155
Portugal	-0.29601	-0.163	-1.0821
Spain	0.24652	0.1776	0.28207
Sweden	0	0	-0.46734
United Kingdom	0	0.042191	0.21825



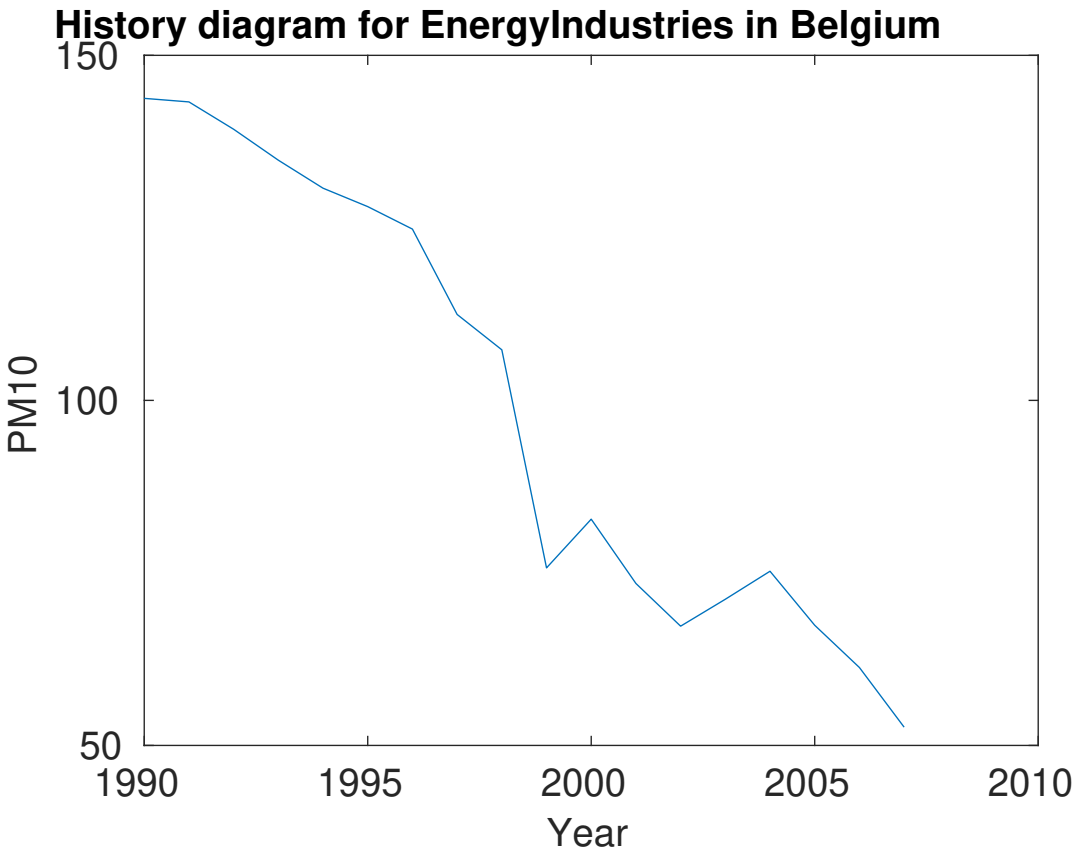
Σχήμα 1.49: Πραγματικές και εκτιμούμενες τιμές για την Waste.

Multivariable fitting in Greece for activity
Waste

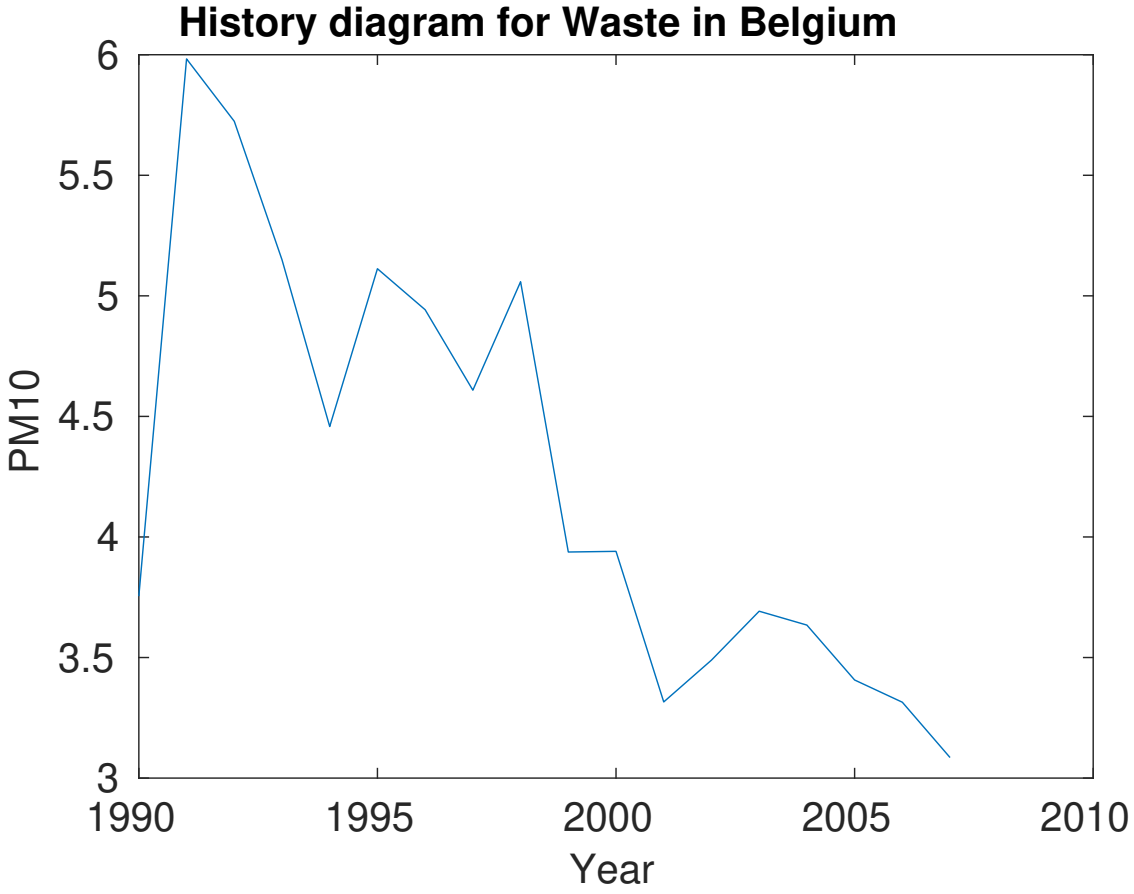
	Stepwisefit	MyStepwise	Full
AdjR2	NaN	NaN	NaN
Intercept	0	0	0
Austria	NaN	0	0
Belgium	NaN	0	0
Denmark	NaN	0	0
Finland	NaN	0	0
France	NaN	0	0
Germany	NaN	0	0
Ireland	NaN	0	0
Italy	NaN	0	0
Luxembourg	NaN	0	0
Netherlands	NaN	0	0
Portugal	NaN	0	0
Spain	NaN	0	0
Sweden	NaN	0	0
United Kingdom	NaN	0	0

1.10 Ζήτημα 10

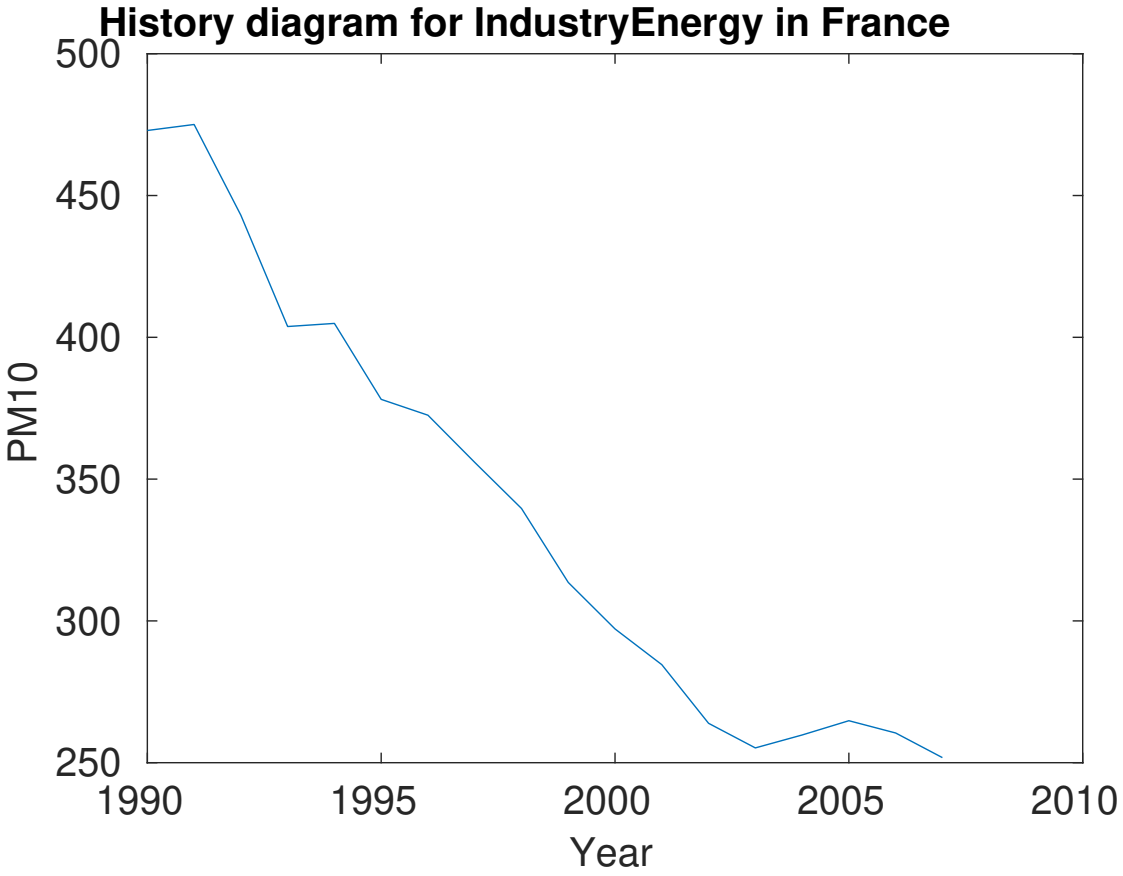
Για την επίλυση αυτού του ζητήματος κάνουμε χρήση του κώδικα 25. Εκτελώντας τον παραπάνω κώδικα για 10 τυχαία ζεύγη δραστηριοτήτων και χωρών και επιλέγοντας σε κάθε ένα πολυώνυμο βαθμού 1 παίρνουμε τα παρακάτω αποτελέσματα.



Σχήμα 1.50: Διάγραμμα ιστορίας για την χώρα Belgium και δραστηριότητα EnergyIndustries.

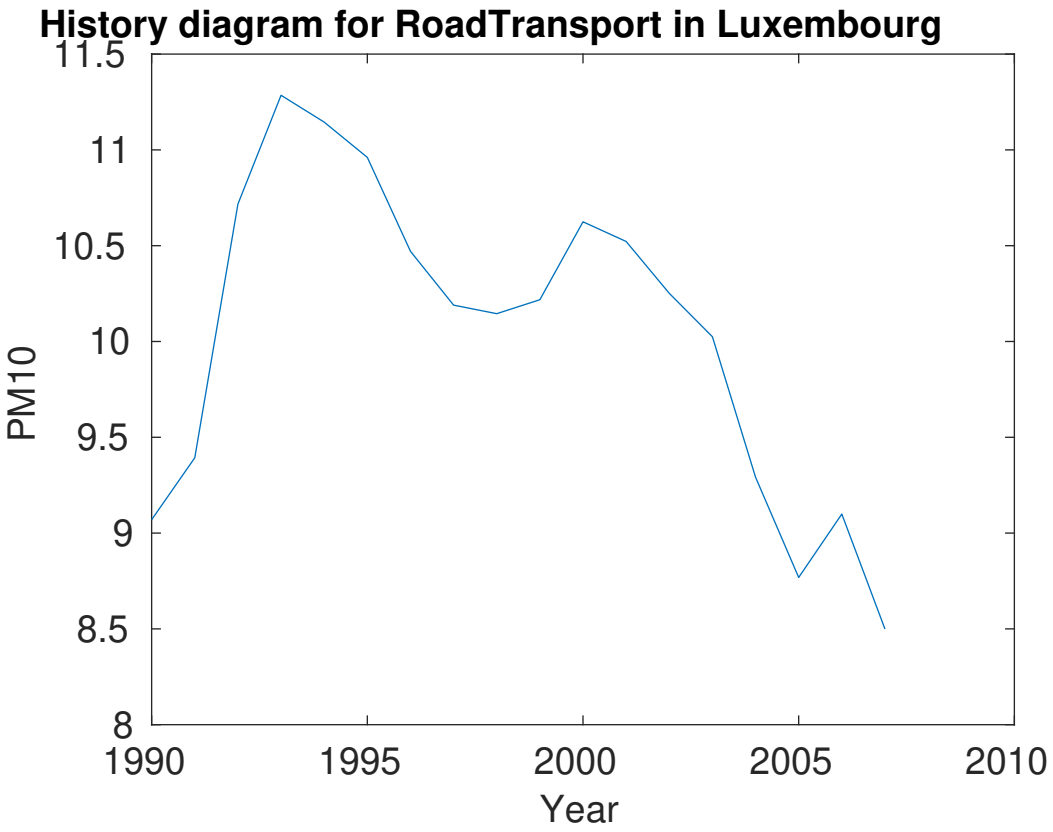


Σχήμα 1.51: Διάγραμμα ιστορίας για την χώρα Belgium και δραστηριότητα Waste.

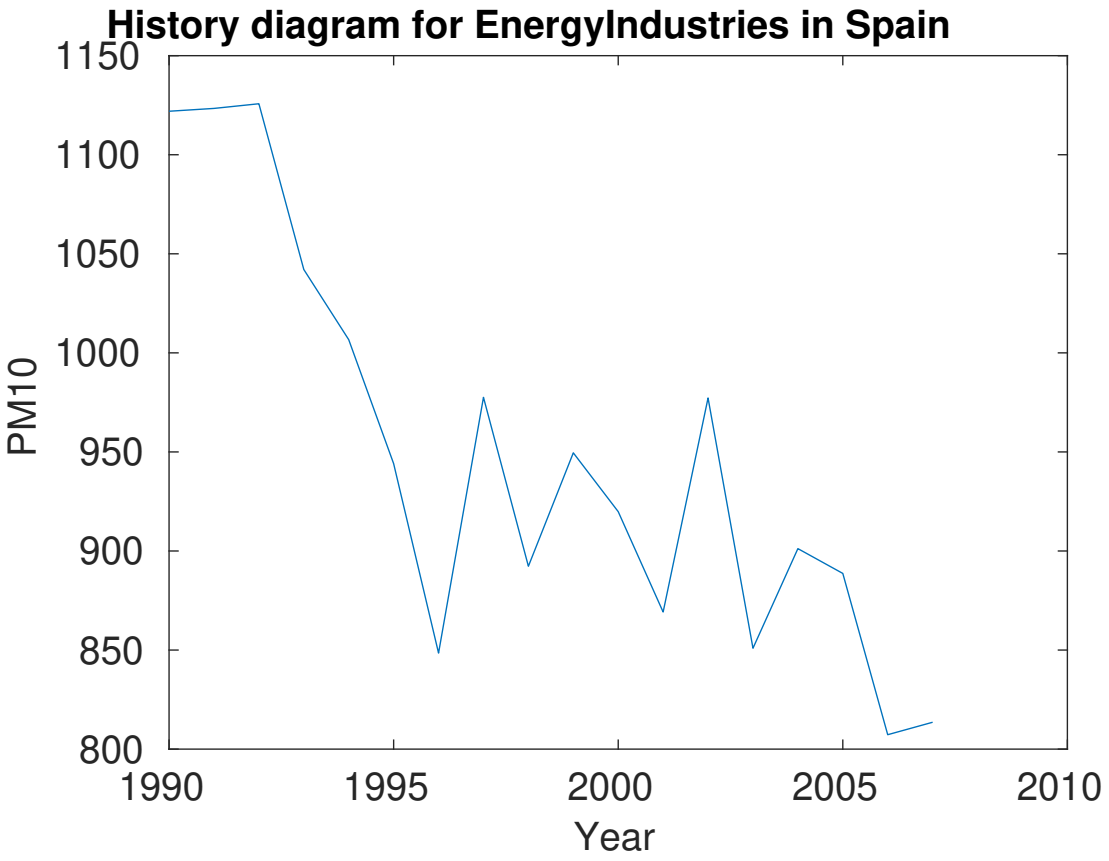


Σχήμα 1.52: Διάγραμμα ιστορίας για την χώρα France και δραστηριότητα IndustryEnergy.

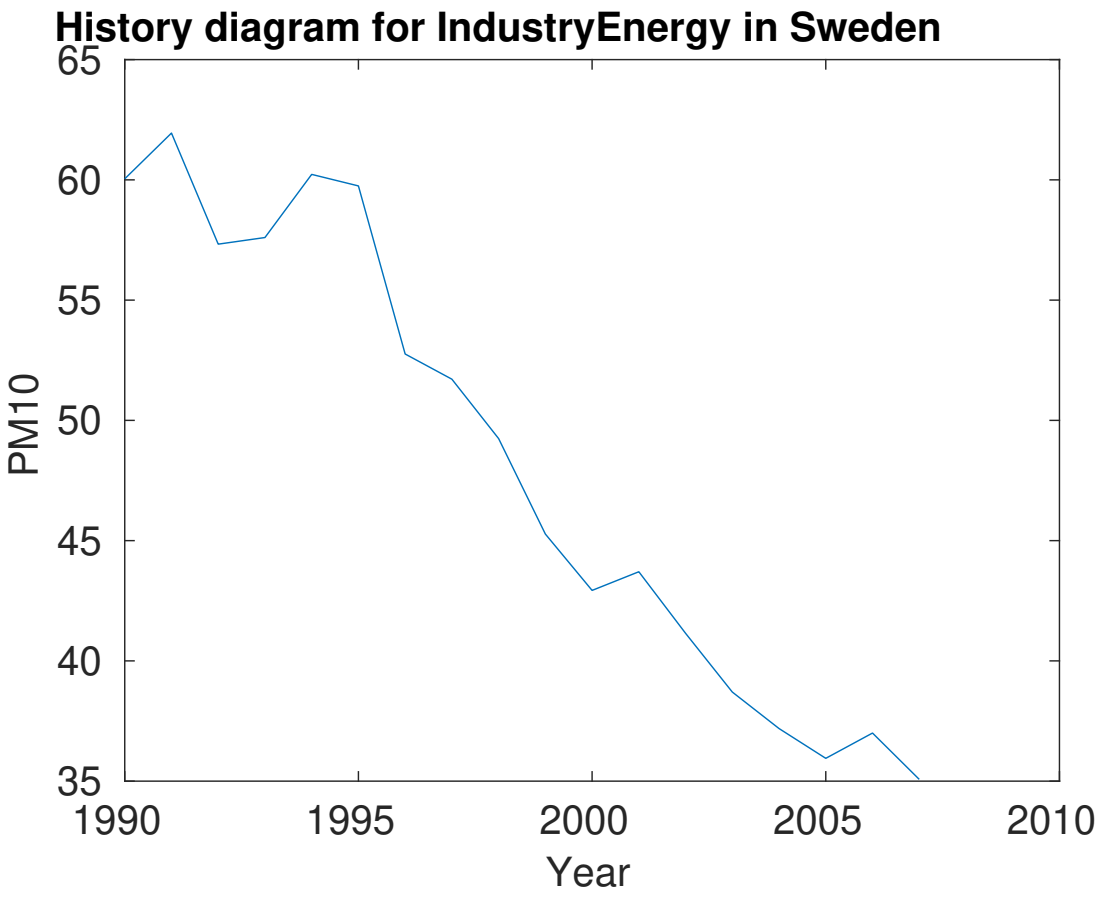
Σχήμα 1.53: Διάγραμμα ιστορίας για την χώρα Italy και δραστηριότητα Waste.



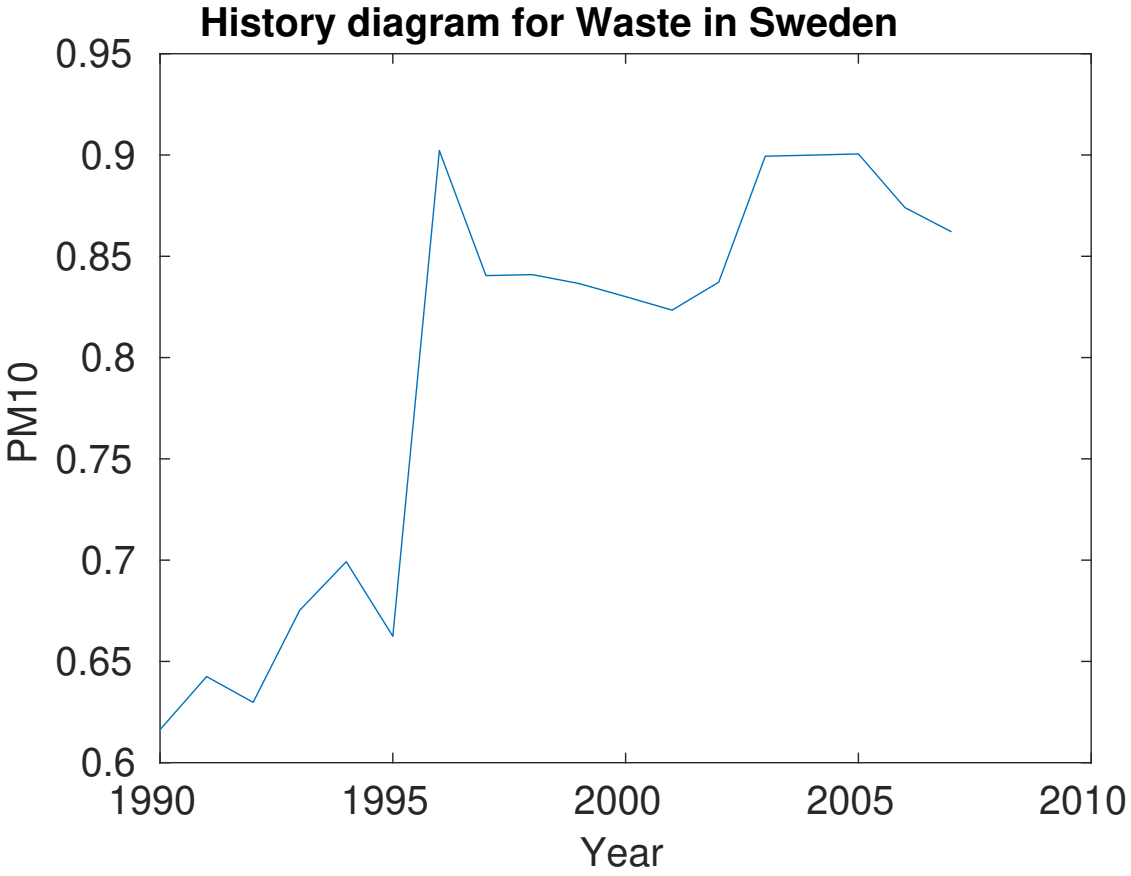
Σχήμα 1.54: Διάγραμμα ιστορίας για την χώρα Luxembourg και δραστηριότητα RoadTransport.



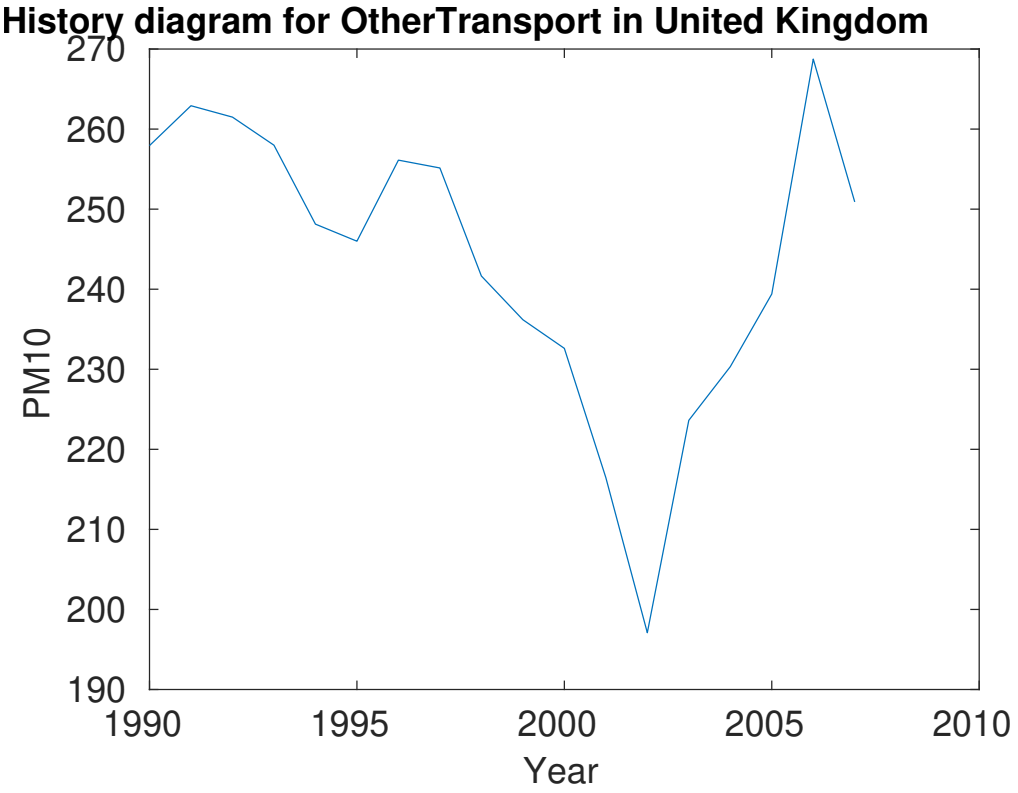
Σχήμα 1.55: Διάγραμμα ιστορίας για την χώρα Spain και δραστηριότητα EnergyIndustries.



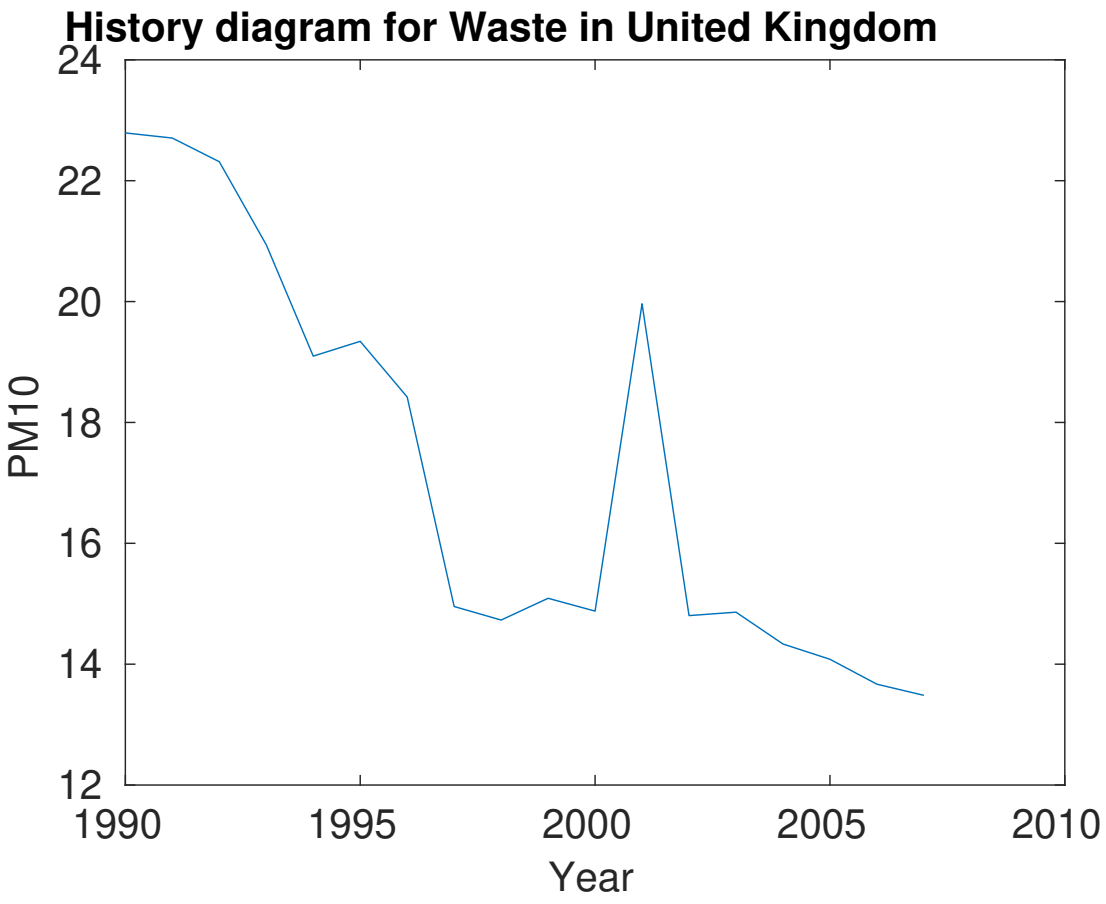
Σχήμα 1.56: Διάγραμμα ιστορίας για την χώρα Sweden και δραστηριότητα IndustryEnergy.



Σχήμα 1.57: Διάγραμμα ιστορίας για την χώρα Sweden και δραστηριότητα Waste.



Σχήμα 1.58: Διάγραμμα ιστορίας για την χώρα United Kingdom και δραστηριότητα OtherTransport.



Σχήμα 1.59: Διάγραμμα ιστορίας για την χώρα United Kingdom και δραστηριότητα Waste.

Παρακάτω δίνονται σε μορφή πίνακα τα αποτελέσματα των παραπάνω ζευγών. Στην τρίτη στήλη (Trend) με 1 συμβολίζεται αν τα αρχικά δεδομένα είχαν τάση και με 0 αν δεν είχαν. Στην τέταρτη στήλη (AutoCorr) με 1 συμβολίζεται αν τα υπόλοιπα μετά την αφαίρεση του βαθμού του πολυωνύμου έχουν σημαντική αυτοσυσχέτιση ή 0 αν δεν έχουν.

Countries	Activities	Trend	AutoCorr
'Spain'	'EnergyIndustries'	1	0
'Sweden'	'Waste'	1	0
'Belgium'	'Waste'	1	0
'Sweden'	'IndustryEnergy'	1	0
'Luxembourg'	'RoadTransport'	1	1
'Belgium'	'EnergyIndustries'	1	1
'France'	'IndustryEnergy'	1	1
'Italy'	'Waste'	0	0
'United Kingdom'	'OtherTransport'	1	1
'United Kingdom'	'Waste'	1	0

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

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
   store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);

%%MAIN program
%Prompt user to choose from countries and activities
prompt = ['Choose type of plot\n One(1) for plotting DIFFERENT
   quantities'...
    ' for the same country\n Two(2) for plotting the SAME
       quantity for'...
    ' different countries\n Please insert 1 or 2\n'];

while 1
    chk1 = int8(input(prompt));
    if chk1 == 1 | chk1 ==2
        break;
    end
end
fprintf('\n The available countries are\n');
for i=1:CountLength
    fprintf('%d \t %s\n',i,Countries{i});
end

if chk1 ==1
```

```
39      prompt2 = ['Choose one country by the number on the left: '
            ];
40      while 1
41          Count1 = int8(input(prompt2));
42          if isinteger(Count1) & 1 <= Count1 & Count1 <=
                CountLength
43              break;
44          end
45      end
46
47      fprintf('\n The available activities are \n');
48
49      for i=1:ActLength
50      fprintf('%d \t %s\n',i,Activities{i});
51      end
52
53
54      prompt3 = ['Choose the first activity by the number on the
            left: '];
55      while 1
56          Count2 = int8(input(prompt3));
57          if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
58              break;
59          end
60      end
61
62
63      prompt4 = ['Choose the second activity by the number on the
            left: '];
64      while 1
65          Count3 = int8(input(prompt4));
66          if isinteger(Count3) & 1 <= Count3 & Count3 <= ActLength
67              break;
68          end
69      end
70      %load the data for the country
71      D1 = DataLoader(filelist,Count2,Count1);
72      D2 = DataLoader(filelist,Count3,Count1);
73
74      Chartme(Countries(Count1),Activities(Count2),Activities(
            Count3),D1,D2,...
75          years,chk1)
76
77  else
78      prompt2 = ['Choose the first country by the number on the
            left: '];
79      while 1
80          Count1 = int8(input(prompt2));
81          if isinteger(Count1) & 1 <= Count1 & Count1 <=
                CountLength
```

```
82              break;
83          end
84      end
85      prompt3 = ['Choose the second country by the number on the
            left: '];
86      while 1
87          Count2 = int8(input(prompt3));
88          if isinteger(Count2) & 1 <= Count2 & Count2 <=
                CountLength
89              break;
90          end
91      end
92
93      fprintf('\n The available activities are \n');
94
95      for i=1:ActLength
96      fprintf('%d \t %s\n',i,Activities{i});
97      end
98
99      prompt4 = ['Choose one activity by the number on the left: '
            ];
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(
            Count3),D1,D2,...
111         years,chk1)
112 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 throught the filenames and returns
          the data for
3     %a specific country and activity
4     datafile = filenames(ActCount).name;
5
6     [data,txt,raw] = xlsread(datafile);
7     if nargin<4 | chk ==0
8         Dar = data(:,Count);
```

```
9       else
10          Dar = data;
11      end
12  end
```

2.3 Chartme

Πρόγραμμα για την δημιουργία Ιστογραμμάτων για το Ζητούμενου 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

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
    store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);

alpha = 0.05;% Set the significance level Default 5%
numbin = 4;
SumH0 = 0;
degfree = numbin-3;%because continuous distributions with mean
    and sigma and sum npoints
chi2limit = chi2inv(1-alpha,degfree);%the upperlimit in chi2
    above which H0 is rejected
for i=1:ActLength

    expval=0;
    for j=1:CountLength
        dataV = DataLoader(filelist,i,j);
        npoints = length(dataV);
        %dataV = dataV-mean(dataV);
        %dataV = dataV./std(dataV);
        %Calculate basic statistics
        meanV = mean(dataV);
        stdV =  std(dataV);
        [ncounts, Edges] = histcounts(dataV,numbin);

        for k=1:numbin
```

```
41              expval(k) = npoints*(normcdf(Edges(k+1),meanV,stdV)-
                    normcdf(Edges(k),meanV,stdV));
42              %expval(k) = npoints*(normcdf(Edges(k+1))-normcdf(
                    Edges(k)));
43          end
44          xi2 = sum((ncounts-expval).^2./expval);
45          %pval = chi2cdf(xi2,degfree,'upper');
46          if xi2 < chi2limit
47              h = 0;
48          else
49              h = 1;
50          end
51          p1 = scatter(i*j,xi2);
52          hold on;
53  %         chi2 = sum((dataV-meanV).^2./dataV);
54          %if i ~=10 & j ~= 7%remove problematic dataset
55              %pd = fitdist(dataV,'Normal');
56              %[h,p,stats] = chi2gof(dataV,'Alpha',alpha,'NBins
                    ',7,'CDF',pd);
57          %else
58          %    h = 0;
59          %end
60
61          SumH0 = SumH0 + h;
62          %fprintf('%d %d %d\n',i,j,SumH0);
63      end
64
65
66  end
67
68  p2 = line([1 ActLength*CountLength],[chi2limit chi2limit],'Color
        ','red','LineStyle','--');
69  title('Testing if datasets are from Normal distribution');
70  ylabel('Chi2 value');
71  xlabel('Dataset');
72  legend([p1,p2],{'Chi2', 'Chi2 maximum'});
73  set(findall(gcf,'-property','FontSize'),'FontSize',14)
74  hold off;
75  fprintf(['\nAt the significance level of %5.2f%% the percentage
        of the datasets\n' ...
76           'that could be from a Normal distribution are %5.2f%% \n
                '],100*alpha, ...
77           100*(1-SumH0/(ActLength*CountLength)));
78
79  fprintf('\nAt the confindence 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)-
                    normcdf(Edges(k), ...
100                  meanV,stdV));
101             %expval(k) = npoints*(normcdf(Edges(k+1))-normcdf(
                    Edges(k)));
102         end
103         xi2 = sum((ncounts-expval).^2./expval);
104         %pval = chi2cdf(xi2,degfree,'upper');
105         fprintf('\nDoes %s Follow a normal distribution : ', ...
106             Activities{i});
107         if xi2 < chi2limit
108             h = 0;
109             fprintf('YES');
110             lwid = 1.5;
111         else
112             h = 1;
113             fprintf('NO');
114             lwid = 0.5;
115         end
116     NormAct = DataM(:,i)-meanV;
117     NormAct = NormAct./std(NormAct);
118     h1 = histogram(NormAct,numbin,'DisplayStyle','stairs','
            LineWidth',lwid);
119     hold on;
120
121 end
122 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

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
% set tha filename for National Totals
TotFilename ='EmissionP10NationalTotalsEU15.xls';
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
   store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);
alpha = 0.05;%Set the significance level

%%Main program
fprintf('\n The available activities are \n');

for i=1:ActLength
    fprintf('%d \t %s\n',i,Activities{i});
end

prompt3 = ['Choose an activity by the number on the left: '];
while 1
    Count2 = int8(input(prompt3));
    if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
        break;
    end
end

A = DataLoader(filelist,Count2,1,1);%get the values for a
   specific activity
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

```
function MeanP10(AM,BM, alpha)
    nCount = length(AM(1,:));%get the number of countries
    pCIs = zeros(nCount,2);
    npCIs = zeros(nCount,2);
    mpCIs = zeros(nCount,1);
    mnpCIs = zeros(nCount,1);

    for i=1:nCount
        [pCIs(i,:), npCIs(i,:)] = CIs(AM(:,i),BM(:,i),alpha);
    end

    mpCIs = mean(pCIs,2);
    mnpCIs = mean(npCIs,2);
    figure();
    errorbar(1:nCount,mpCIs,pCIs(:,1),pCIs(:,2),'o');
    hold on;
    errorbar(1:nCount,mnpCIs,npCIs(:,1),npCIs(:,2),'*');
    title('Mean and CIs of avtivity');
    ylabel('Relative mean PM10 value');
    xlabel('Country code');
    legend('Parametric','Non Parametric');
    set(findall(gcf,'-property','FontSize'),'FontSize',14);
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

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
% set tha filename for National Totals
TotFilename ='EmissionP10NationalTotalsEU15.xls';
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
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
```

```
%non parametric analysis
m = 1000; %number of randomly selecting sets of pairs
%create random pairs fo A and B by randomly selecting the
    indices of array
%B
D4 = zeros(m,1);

for i=1:m
    indx = unidrnd(nsize,nsize,1);
    D4(i) = std(A(indx)./B(indx));
end

D4 = sort(D4);
minCI = round(m*alpha/2);
maxCI = round(m*(1-alpha/2));

npCIs = [D4(minCI),D4(maxCI)];


end
```

2.10 StdP10

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

Listing 10: StdP10.m

```
function StdP10(AM,BM,alpha)
    nCount = length(AM(1,:));%get the number of countries
    pCIs = zeros(nCount,2);
    npCIs = zeros(nCount,2);
    mpCIs = zeros(nCount,1);
    mnpCIs = zeros(nCount,1);

    for i=1:nCount
       [pCIs(i,:), npCIs(i,:)] = stdCIs(AM(:,i),BM(:,i),alpha);
    end

    mpCIs = mean(pCIs,2);
    mnpCIs = mean(npCIs,2);
    figure();
    errorbar(1:nCount,mpCIs,pCIs(:,1),pCIs(:,2),'o');
    hold on;
    errorbar(1:nCount,mnpCIs,npCIs(:,1),npCIs(:,2),'*');
    title('Std and CIs of avtivity');
    ylabel('Relative Std PM10 value');
    xlabel('Country code');
    legend('Parametric','Non Parametric');
    set(findall(gcf,'-property','FontSize'),'FontSize',14);
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

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
% set tha filename for National Totals
TotFilename ='EmissionP10NationalTotalsEU15.xls';
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
   store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);
alpha = 0.05;%Set the significance level

%%Main program
fprintf('\n The available activities are \n');

for i=1:ActLength
    fprintf('%d \t %s\n',i,Activities{i});
end

prompt3 = ['Choose an activity by the number on the left: '];
while 1
    Count2 = int8(input(prompt3));
    if isinteger(Count2) & 1 <= Count2 & Count2 <= ActLength
        break;
    end
end

%A = DataLoader(filelist,Count2,1);%get the values for a
   specific activity
B = DataLoader(filelist,Count2,2,1);
```

```
[Rmax CountriesMax Per1 Per2] = AllCountries(B,alpha);

fprintf('\nThe 10 sets of countries with the highest \
   ncorrelation for %s are\n',Activities{Count2});
fprintf('Correlation\t Country1\t Country2\n');
for i=1:10
    fprintf('%5.2f\t \t %10s\t %10s \n',Rmax(i),Countries{
       CountriesMax(i,1)},Countries{CountriesMax(i,2)});
end

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);
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

```
function [Rmax combmax count1 count2] = AllCountries(AM,alpha)
    nCount = length(AM(1,:));
    comb = combnk(1:nCount,2);%get all possible combinations of
       countries
    ncomb = length(comb(:,1));
    RPP = zeros(ncomb,3);

    for i=1:ncomb
        [RPP(i,1),RPP(i,2),RPP(i,3)] = corrme(AM(:,comb(i,1)),AM
           (:,comb(i,2)),alpha);
    end


    figure();
    h = histogram(RPP(:,1),-1:0.25:1);
    title('Correlations among all sets of countries');
    ylabel('Counts');
    xlabel('Corellation factor');
    set(findall(gcf,'-property','FontSize'),'FontSize',14);
    %sort the R values and pick the 10 highest
    RPP(:,1)=abs(RPP(:,1));
    RPP(isnan(RPP(:,1)),1)=-inf;
    [Rsorted I] = sort(RPP(:,1),'descend');
    RPP(isinf(RPP(:,1)),1)=NaN;
    Rmax = Rsorted(1:10);
    Imax = I(1:10);
    combmax = comb(Imax,:);
    count1 = sum(RPP(:,2)>=alpha)/ncomb;
```

```
27      count2 = sum(RPP(:,3)>=alpha)/ncomb;
28  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*
           alpha/2)), ...
23     %     ' and ',num2str(round((1-alpha/2)*m)), ' and is ',
           num2str(round(sum(ta<ta0))),'\n']);
24     maxind = round(sum(ta<ta0));
25     p2 = (1-maxind/(1.0*m));
26
27
28 end
```

2.14 Exercise6

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

Listing 14: Project2019Ex6.m

```
1  clc;
2  clear all;
3
4  %%Preload
5  %find all data and activities
```

```
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
% set tha filename for National Totals
TotFilename ='EmissionP10NationalTotalsEU15.xls';
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
   store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);
alpha = 0.05;%Set the significance level

%%Main program
fprintf('\n The available Countries are \n');

for i=1:CountLength
    fprintf('%d \t %s\n',i,Countries{i});
end

prompt3 = ['Choose a Country by the number on the left: '];
while 1
    Count2 = int8(input(prompt3));
    if isinteger(Count2) & 1 <= Count2 & Count2 <= CountLength
        break;
    end
end
CM = zeros(length(years),ActLength);
%load all the activities for a specific country
for i=1:ActLength
    CM(:,i) = DataLoader(filelist,i,Count2);
end

[Rmax ActivitiesMax Per1 Per2] = AllActivities(CM,alpha);

fprintf('\nThe 10 sets of Activities with the highest
   correlation for %s are\n',Countries{Count2});
fprintf('Correlation\t Activity1\t \t Activity2\n');
for i=1:10
    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 \nthe
     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 \nthe 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     RPP(:,1)=abs(RPP(:,1));
20     RPP(isnan(RPP(:,1)),1)=-inf;
21     [Rsorted I] = sort(RPP(:,1),'descend');
22     RPP(isinf(RPP(:,1)),1)=NaN;
23     Rmax = Rsorted(1:10);
24     Imax = I(1:10);
25     combmax = comb(Imax,:);
26     count1 = sum(RPP(:,2)>=alpha)/ncomb;
27     count2 = sum(RPP(:,3)>=alpha)/ncomb;
28 end
```

2.16 Exercise7

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.7. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, AdjRCountries και polyme.

Listing 16: Project2019Ex7.m

```
clc;
clear all;

%%Preload
%find all data and activities
filelist = dir('EmissionP10*EU15.xls');%Get the filenames of the
    data files
% set tha filename for National Totals
TotFilename ='EmissionP10NationalTotalsEU15.xls';
filelist(7)=[];%remove the Nationatotals filename
names = {filelist(:).name};
Activities = extractBetween(names(:),"EmissionP10","EU15");
ActLength = length(Activities);
%Load a file to get the names of countries and years
[data,txt,raw] = xlsread('EmissionP10EnergyIndustriesEU15.xls');
dataname = txt(2,1);
years = str2double(txt(2:end,2));
Countries = [];
for i=1:length(data(1,:)) %get the name of the countries and
    store them
    Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
        ")];
end
Countries = Countries';
CountLength = length(Countries);
alpha = 0.05;%Set the significance level
norder = 5; %maximum polynomial order to fit for

%%Main program
AM = xlsread(TotFilename);
R2M = zeros(CountLength,ActLength);
O1M = zeros(CountLength,ActLength);

for i=1:ActLength
    BM = DataLoader(filelist,i,1,2);
    [R2M(:,i) O1M(:,i)] = AdjRCountries(AM,BM,norder);
end
R2M = abs(R2M);
R2M(isnan(R2M)) = -inf;
[R2Ms Ind] = sort(R2M,'descend');
R2M(isinf(R2M)) = NaN;
R2Max = R2Ms(1,:);
idx = sub2ind(size(O1M), Ind(1,:), [1:ActLength]);
OrdMax = O1M(idx);
CountriesMax = {Countries{Ind(1,:)}};
%ActivitiesMax = [Activities{:}];
T = table(R2Max',OrdMax',CountriesMax','RowNames',Activities,'
    VariableNames',{'AdjR2','Order', 'Country'})

```

```
figure();
histogram(O1M);
title('Polynomial index distribution');
ylabel('Counts');
xlabel('Polynomial Order');
```

2.17 AdjRCountries

Πρόγραμμα για τον υπολογισμό του συντελεστή συσχέτισης για το Ζητούμενο 1.7.

Listing 17: AdjRCountries.m

```
function [R2 O1] = AdjRCountries(AM,BM,n)
    nCount = length(AM(1,:));

    R2 = zeros(nCount,1);
    O1 = zeros(nCount,1);
    for i =1:nCount

        [R2(i) O1(i)] = polyme(AM(:,i),BM(:,i),n);

    end


end
```

2.18 polyme

Πρόγραμμα για τον υπολογισμό πολυωνύμου για το Ζητούμενο 1.7.

Listing 18: polyme.m

```
function [AdjR nm] = polyme(A,B,n)
%A and B are arrays of the same activity and n is the maximum
    polynomial
%order to fit for. This function find the best polynomial order
    fit
%depending on the AdjR. A are the totals and B the independent
    values.
    npoints = length(A);
    adjr2 = zeros(n+1,1);
    if sum(isnan(B))==npoints
        AdjR = NaN;
        nm = 0;
        fprintf('Nan Array found');
    else

        for i=1:(n+1)
            [p S mu] = polyfit(B,A,i-1);
            r2 = 1 - sum((A-polyval(p,B,S,mu)).^2)/sum((A-mean(A
                )).^2);
            adjr2(i) = 1-((npoints-1)/(npoints-i))*(1-r2);
```

```
17          end
18          adjr2 = abs(adjr2);
19          adjr2(isnan(adjr2)) = -Inf;
20          [AdjR nm] = sort(adjr2,'descend');
21          adjr2(isinf(adjr2)) = NaN;
22          AdjR = AdjR(1);
23          nm = nm(1)-1; %to get the actual order of the polynomial
                .
24      end
25
26      if AdjR >1
27          AdjR=NaN;
28      end
29
30  end
```

2.19 Exercise8

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.8. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, mystepwise και StepfitCountries.

Listing 19: Project2019Ex8.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 Nyears = length(years);
18 Countries = [];
19 for i=1:length(data(1,:)) %get the name of the countries and
       store them
20      Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
            ")];
21 end
22 Countries = Countries';
23 CountLength = length(Countries);
24 alpha = 0.05;%Set the significance level
25 norder = 5; %maximum polynomial order to fit for
```

```
26
27 %%Main program
28 AM = xlsread(TotFilename);%load the totals
29 BM = zeros(Nyears,ActLength);
30
31
32 for j=1:CountLength
33     for i = 1:ActLength
34         BM(:,i) = DataLoader(filelist,i,j);
35     end
36     CM = [AM(:,j) BM];
37     StepfitCountries(CM,Countries{j},Activities);
38 end
```

2.20 mystepwise

Πρόγραμμα για τον υπολογισμό βηματικού μοντέλου για το Ζητούμενο 1.8.

Listing 20: mystepwise.m

```
1  function [AdjR20 bfinal] = mystepwise(CM)
2  %function to calculate the best stepwise model based on adjR. CM
       is a
3  %matrix with first column the totals which in this case are the
       dependent
4  %variable and the rest columns are the Activities which are the
       independent
5  % variables
6
7  y = CM(:,1); % get the total data
8
9  %%There are many ways to do a stepwise fit here we will attemtp
       one of the
10 %%simplest. We first fit y with each variable one at a time and
       sort the
11 %%variables by their AdjR. Then we try to fit following that
       order and
12 %%check if the AdjR gets better or not. This stops when the AdjR
       falls in
13 %%value.
14
15 n = length(y);
16 VarSize = length(CM(1,2:end));% the amount of variables
17 R1 = zeros(VarSize, 1);
18 for i =1:VarSize
19     X = [ones(size(y)) CM(:,i+1)];%load variable one at a time
20     [b,bint,r,rint,stats] = regress(y,X);
21     R1(i) = stats(1);
22 %      [b,Sigm,E,CovB,logL] = mvregress(X,y);
23 %
24 %      R1(i) = 1- sum(E.^2)/sum((y-mean(y)).^2);
```

```
25
26 end
27
28 [R1s Ind] = sort(R1,'descend');
29
30 k = 1;
31 AdjR20 = 1-((n-1)/(n-(k+1)))*(1-R1s(1)); %get the AdjR2 of the
    best single variable fit
32 X = [ones(size(y)) CM(:,Ind(1)+1)];%load the best fitting
    variable
33 InitVar = 1:VarSize;%array with all the variable indices
34 SelVar = Ind(1);%Array with the selected variable indices
35 InitVar(Ind(1)) = []; %remove the variable that was accepted in
    the fit
36 remcount = length(InitVar);%remaining variables to test
37 while 1
38
39     AdjR2a = zeros(remcount,1);
40     b1 = zeros(1+k+1,remcount);
41     for j=1:remcount
42         Xa = [X CM(:,InitVar(j)+1)]; %append the next best
               variable
43         [b1(:,j),bint,r,rint,stats] = regress(y,Xa);
44         R2i = stats(1);
45         AdjR2a(j) = 1-((n-1)/(n-(k+1+1)))*(1-R2i);
46     end
47     [AdjR2aS IndS] = sort(AdjR2a,'descend');
48
49     if AdjR2aS(1) < AdjR20
50         break;
51
52     else
53         AdjR20 = AdjR2aS(1);
54         b = b1(:,IndS(1));
55         SelVar = [SelVar InitVar(IndS(1))];
56         InitVar(IndS(1)) = [];
57         k = length(SelVar);
58         X = [X CM(:,SelVar(end))];
59         remcount = length(InitVar);%remaining variables to test
60     end
61     if remcount ==0
62         break;
63     end
64 end
65
66 bfinal = zeros(VarSize+1,1);
67 bfinal(1) = b(1);
68 bfinal(SelVar+1) = b(2:end);
69
70
```

```
71 end
```

2.21 StepfitCountries

Πρόγραμμα για τον υπολογισμό βηματικού πολυωνύμου για το Ζητούμενο 1.8.

Listing 21: StepfitCountries.m

```
1  function StepfitCountries(CM, Country, Actlist)
2
3      y = CM(:,1);
4      Nsize = length(y);
5      %%Calculate the stepwisefit
6      [b,se,pval,inmodel,stats,nextstep,history]= stepwisefit(CM
           (:,2:end),y);
7      bstep = b.*inmodel';
8      intercept = stats.intercept;
9      bstep = [intercept bstep'];
10     R2step = 1-stats.SSresid/stats.SStotal;
11     k= sum(inmodel);
12     AdjR2step = 1-((Nsize-1)/(Nsize-k-1))*stats.SSresid/stats.
           SStotal;
13
14     %%Calculate my stepwise method
15     [AdjR2me bme] = mystepwise(CM);
16
17     %%Calculate the full model
18     NM = [ones(size(y)) CM(:,2:end)];
19     [btot,bint,r,rint,stats] = regress(y,NM);
20     R1 = stats(1);
21     k = length(CM(1,2:end));
22     AdjR2tot = 1-((Nsize-1)/(Nsize-(k+1)))*(1-R1);
23
24
25     %%return results
26
27     fprintf('\nMultivariable fitting for Activities in Country\n
           \t \t %s\n',Country);
28     %fprintf('Activities\t Coefficients');
29     Varlist = ['AdjR2' 'Intercept' {Actlist{:}}];
30     V1 = [AdjR2step bstep'];
31     V2 = [AdjR2me bme'];
32     V3 = [AdjR2tot btot'];
33     T=table(V1,V2,V3,'RowNames',Varlist','VariableNames',{'
           Stepwisefit','MyStepwise','Full'})
34
35     figure();
36     plot(y,y,'o','MarkerSize',10);
37     hold on;
38     plot(y,bstep(1)+CM(:,2:end)*bstep(2:end),'x');
39     plot(y,bme(1)+CM(:,2:end)*bme(2:end),'*');
```

```
40      plot(y,btot(1)+CM(:,2:end)*btot(2:end),'d');
41      titl1 = strcat({'Multivar fits for '}, Country);
42      title(titl1);
43      xlabel('Total PM10');
44      ylabel('Expected Total PM10');
45      legend('Actual Data', 'Stepwisefit', 'My Stepwise fit', '
            Full fit','Location', 'best');
46      figname = strcat(Country,'_MultivarFits');
47      set(findall(gcf,'-property','FontSize'),'FontSize',14);
48      saveas(gcf,figname,'epsc');
49
50  end
```

2.22 Exercise9

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.9. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, mystepwiseAct και StepfitActivoties.

Listing 22: Project2019Ex9.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 Nyears = length(years);
18 Countries = [];
19 for i=1:length(data(1,:)) %get the name of the countries and
       store them
20      Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
            ")];
21 end
22 Countries = Countries';
23 CountLength = length(Countries);
24 alpha = 0.05;%Set the significance level
25 norder = 5; %maximum polynomial order to fit for
26
27 %%Main program
28 %AM = xlsread(TotFilename);%load the totals
```

```
%Set the mark to 7 in order to check for Greece
mark = 7;
for j=1:ActLength
    BM = DataLoader(filelist,j,1,2);

    StepfitActivities(BM, Activities{j}, Countries, mark);
end
```

2.23 mystepwiseAct

Πρόγραμμα για τον υπολογισμό βηματικού μοντέλου για το Ζητούμενο 1.9.

Listing 23: mystepwiseAct.m

```
function [AdjR20 bfinal] = mystepwiseAct(CM,y)
%function to calculate the best stepwise model based on adjR. CM is a
%matrix with the independent variables


%%There are many ways to do a stepwise fit here we will attemtp one of the
%%simplest. We first fit y with each variable one at a time and sort the
%%variables by their AdjR. Then we try to fit following that order and
%%check if the AdjR gets better or not. This stops when the AdjR falls in
%%value.

n = length(y);
VarSize = length(CM(1,:));% the amount of variables
R1 = zeros(VarSize, 1);
for i =1:VarSize
    X = [ones(size(y)) CM(:,i)];%load variable one at a time
    [b,bint,r,rint,stats] = regress(y,X);
    R1(i) = stats(1);

end

[R1s Ind] = sort(R1,'descend');

k = 1;
AdjR20 = 1-((n-1)/(n-(k+1)))*(1-R1s(1)); %get the AdjR2 of the best single variable fit
X = [ones(size(y)) CM(:,Ind(1))];%load the best fitting variable
InitVar = 1:VarSize;%array with all the variable indices
SelVar = Ind(1);%Array with the selected variable indices
InitVar(Ind(1)) = []; %remove the variable that was accepted in the fit
remcount = length(InitVar);%remaining variables to test
```

```
while 1

    AdjR2a = zeros(remcount,1);
    b1 = zeros(1+k+1,remcount);
    for j=1:remcount
        Xa = [X CM(:,InitVar(j))]; %append the next best
            variable
        [b1(:,j),bint,r,rint,stats] = regress(y,Xa);
        R2i = stats(1);
        AdjR2a(j) = 1-((n-1)/(n-(k+1+1)))*(1-R2i);%there is an
            extra +1 here
        %beacause k is not refreshed yet to the new value
    end
    [AdjR2aS IndS] = sort(AdjR2a,'descend');

    if AdjR2aS(1) < AdjR20
        break;

    else
        AdjR20 = AdjR2aS(1);
        b = b1(:,IndS(1));
        SelVar = [SelVar InitVar(IndS(1))];
        InitVar(IndS(1)) = [];
        k = length(SelVar);
        X = [X CM(:,SelVar(end))];
        remcount = length(InitVar);%remaining variables to test
    end
    if remcount ==0
        break;
    end
end

bfinal = zeros(VarSize+1,1);
bfinal(1) = b(1);
bfinal(SelVar+1) = b(2:end);


end
```

2.24 StepfitActivities

Πρόγραμμα για τον υπολογισμό βηματικού πολυωνύμου για το Ζητούμενο 1.9.

Listing 24: StepfitActivities.m

```
function StepfitActivities(CM, Activity, Countlist, mark)

    y = CM(:,mark);
    %remove specific country for datasets
    Countlist(mark)=[];
    CM(:,mark) = [];
```

```
7       Nsize = length(y);
8       %%Calculate the stepwisefit
9       [b,se,pval,inmodel,stats,nextstep,history]= stepwisefit(CM,y
            );
10      bstep = b.*inmodel';
11      intercept = stats.intercept;
12      bstep = [intercept bstep'];
13      R2step = 1-stats.SSresid/stats.SStotal;
14      k= sum(inmodel);
15      AdjR2step = 1-((Nsize-1)/(Nsize-k-1))*stats.SSresid/stats.
            SStotal;
16
17      %%Calculate my stepwise method
18      [AdjR2me bme] = mystepwiseAct(CM,y);
19
20      %%Calculate the full model
21      NM = [ones(size(y)) CM];
22      [btot,bint,r,rint,stats] = regress(y,NM);
23      R1 = stats(1);
24      k = length(CM(1,:));
25      AdjR2tot = 1-((Nsize-1)/(Nsize-(k+1)))*(1-R1);
26
27
28      %%return results
29
30      fprintf('\nMultivariable fitting in Greece for activity\n \t
            \t %s\n',Activity);
31      %fprintf('Activities\t Coefficients');
32      Varlist = ['AdjR2' 'Intercept' {Countlist{:}}];
33      V1 = [AdjR2step bstep'];
34      V2 = [AdjR2me bme'];
35      V3 = [AdjR2tot btot'];
36      T=table(V1,V2,V3,'RowNames',Varlist','VariableNames',{'
            Stepwisefit','MyStepwise','Full'})
37
38      figure();
39      plot(y,y,'o','MarkerSize',10);
40      hold on;
41      plot(y,bstep(1)+CM*bstep(2:end),'x');
42      plot(y,bme(1)+CM*bme(2:end),'*');
43      plot(y,btot(1)+CM*btot(2:end),'d');
44      titl1 = strcat({'Multivar fits in Greece for '}, Activity);
45      title(titl1);
46      xlabel('Greece''s PM10');
47      ylabel('Greece''s xpected Total PM10');
48      legend('Actual Data', 'Stepwisefit', 'My Stepwise fit', '
            Full fit','Location', 'best');
49      figname = strcat(Activity,'_Greece_MultivarFits');
50      set(findall(gcf,'-property','FontSize'),'FontSize',14);
51      saveas(gcf,figname,'epsc');
```

```
52
53 end
```

2.25 Exercise10

Πρόγραμμα Matlab για την επίλυση του Ζητούμενου 1.10. Το οποίο με την σειρά του καλεί το προγράμμα DataLoader, mytisan.

Listing 25: Project2019Ex10.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 Nyears = length(years);
18 Countries = [];
19 for i=1:length(data(1,:)) %get the name of the countries and
      store them
20     Countries = [Countries, extractBetween(txt(1,i+2),") - "," -
           ")];
21 end
22 Countries = Countries';
23 CountLength = length(Countries);
24 alpha = 0.05;%Set the significance level
25
26
27 %%Main program
28 %% pick nSet random countries and activities
29 nSet = 10;
30 rng(0,'twister');%initialise seed for repetability
31 CountRnd = randi(CountLength,nSet,1);
32 ActRnd = randi(ActLength,nSet,1);
33 Trend = zeros(nSet,1);
34 Corr =  zeros(nSet,1);
35 for i=1:nSet
36     Country = Countries{CountRnd(i)};
37     Activity = Activities{ActRnd(i)};
38     BM = DataLoader(filelist,ActRnd(i),CountRnd(i));
```

```
39
40      [Trend(i) Corr(i)] = mytisan(BM, years, Country, Activity);
41  end
42
43  T=table({Countries{CountRnd}}', {Activities{ActRnd}}', Trend, Corr
        ,...
44      'VariableNames',{'Countries','Activities','Trend', 'AutoCorr
            '})
```

2.26 mytisan

Πρόγραμμα για τον υπολογισμό τάσης για το Ζητούμενο 1.10.

Listing 26: mytisan.m

```
1  function [Trend1 Corr1] = mytisan(A, years, Country, Activity)
2
3  %plot the history diagram of the requested quantity
4      figure();
5      plot(years,A);
6      titl1 = strcat('History diagram for ',{' '},Activity,' in '
           ,{' '},Country);
7      title(titl1);
8      ylabel('PM10');
9      xlabel('Year');
10     set(findall(gcf,'-property','FontSize'),'FontSize',14);
11
12     %Ask the user to define the order of the polynomial trend to
           remove
13     nsize = length(A);
14     prompt1 = ['\nChoose the order of the polynomial trend to
           remove from\n the data (0 removes nothing!): '];
15 while 1
16     Count1 = int8(input(prompt1));
17     if isinteger(Count1) & 0 <= Count1 & Count1 <= nsize
18         break;
19     else
20         fprintf('Enter an integer between %d and %d',0,nsize);
21     end
22 end
23 x = (1:length(years))';
24 if Count1 > 0
25     [p1,S1,mu1] = polyfit(x,A,Count1);
26     yres = A-polyval(p1,x,S1,mu1);
27 else
28     yres = A;
29 end
30
31 [ACF,lags,bounds] = autocorr(yres,5);
32 [ACF0,lags0,bounds0] = autocorr(A,5);
33
```

```
34
35 %check if there was original trend in the data by the autocorr
36 if sum(abs(ACF0(2:end))>bounds0(1)) > 0
37     Trend1 = 1;
38 else
39     Trend1 = 0;
40 end
41
42 if sum(abs(ACF(2:end))>bounds(1)) > 0
43     Corr1 = 1;
44 else
45     Corr1 = 0;
46 end
47
48 figname = strcat(Country,'_',Activity,'_Trend_and_Corr');
49 saveas(gcf,figname,'epsc');
50
51
52
53 end
```
