**ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ**

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

****

**ΕΡΓΑΣΙΑ ΜΕ ΘΕΜΑ ΤΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ**

ΜΑΘΗΜΑ:

**«****[12ΥΔ01] ΟΙΚΟΝΟΜΕΤΡΙΑ Ι»**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ:

**Κωνσταντίνος Π. Γιαννακόπουλος**

ΑΕΜ:

**16051**

(Τμήμα Μαθηματικών)

Table of Contents

**1. ΕΙΣΑΓΩΓΗ3**

**2. ΤΟ ΥΠΟΔΕΙΓΜΑ4**

2.1 ΘΕΩΡΗΤΙΚΗ ΘΕΜΕΛΙΩΣΗ4

2.2 ΕΞΕΙΔΙΚΕΥΣΗ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ4

**3. ΕΜΠΕΙΡΙΚΗ ΑΝΑΛΥΣΗ**6

3.1 ΔΕΔΟΜΕΝΑ6

3.2 ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ6

3.3 ΕΚΤΙΜΗΣΗ ΤΟΥ ΥΠΟΔΕΙΓΜΑΤΟΣ9

**4. ΔΙΑΓΝΩΣΤΙΚΟΙ ΕΛΕΓΧΟΙ13**

4.1 ΕΛΕΓΧΟΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ13

4.2 ΕΛΕΓΧΟΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ14

4.3 ΕΛΕΓΧΟΣ ΠΟΛΥΣΥΓΓΡΑΜΙΚΟΤΗΤΑΣ…………………………………………..15

**5. ΕΛΕΓΧΟΙ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ**16

**6. ΕΛΕΓΧΟΣ ΕΥΣΤΑΘΕΙΑΣ17**

**ΣΥΜΠΕΡΑΣΜΑΤΑ18**

**ΒΙΒΛΙΟΓΡΑΦΙΑ…………………………………………………………………………….19**

**ΠΑΡΑΡΤΗΜΑ………………………………………………………………………………20**

**1. ΕΙΣΑΓΩΓΗ**

Η παρούσα εργασία γίνεται στα πλαίσια του μαθήματος της Οικονομετρίας και όπως και το ίδιο το μάθημα, πραγματεύεται το θέμα της γραμμικής παλινδρόμησης. Στόχος της εργασίας είναι η πρακτική εξοικείωση με τις έννοιες και τις πρακτικές που αναφέρθηκαν κατά τη διάρκεια του εξαμήνου. Πιο συγκεκριμένα, λαμβάνοντας κάποια δεδομένα, θα αναλύσω το γραμμικό υπόδειγμα, μέσω οπτικοποίησης αλλά και ελέγχων, προκειμένου να εξετάσουμε κατά πόσο είναι σωστή η γραμμική σχέση, να εξετάσουμε αν υπάρχουν τυχόν παθογένειες και τέλος ναι εξάγουμε τα κατάλληλα συμπεράσματα.

Στο 2ο κεφάλαιο θα αναλύσουμε το dataset και το γραμμικό υπόδειγμα σε ένα γενικό και θεωρητικό πλαίσιο. Στη συνέχεια, στο κεφάλαιο 3 θα αναλύσουμε τα δεδομένα, μέσω γραφικών παραστάσεων για να λάβουμε τις πρώτες ενδείξεις για τις μεταβλητές μας και τέλος να εκτιμήσουμε το υπόδειγμα. Στο 4ο κεφάλαιο, θα πραγματοποιηθούν με τους κατάλληλους ελέγχους, έλεγχοι για αυτοσυσχετιση και ετεροσκεδαστικότητα. Στο 5ο και 6ο κεφάλαιο, θα γίνουν έλεγχοι σημαντικότητας των συντελεστών του υποδείγματος αλλά και ευστάθειας. Τέλος, θα καταλήξουμε σε συμπεράσματα σχετικά με την ανάλυση μας.

Η εργασία πραγματοποιήθηκε σε γλώσσα R (Rstudio).

**2. ΤΟ ΥΠΟΔΕΙΓΜΑ**

2.1 Θεωρητική θεμελίωση

Το dataset που επιλέχθηκε, περιέχει τις παρακάτω μεταβλητές:

**Country**: Το όνομα κάθε χώρας

**ISO\_code**: Τον κωδικό ISO κάθε χώρας

**GPI**: Παγκόσμιος Δείκτης Ειρήνης

**GCI**: Παγκόσμιος Δείκτης Ανταγωνιστικότητας

**Export**: Οι εξαγωγές της χώρας σε USD($)

**Import**: Οι εισαγωγές της χώρας σε USD($)

**Muslim**: Ποσοστό μουσουλμάνων στη χώρα

**Mu2**: Μουσουλμάνοι (σε εκατομμύρια) στη χώρα

**Military**:Ποσοστό στρατιωτικών εξόδων με βάση το ΑΕΠ

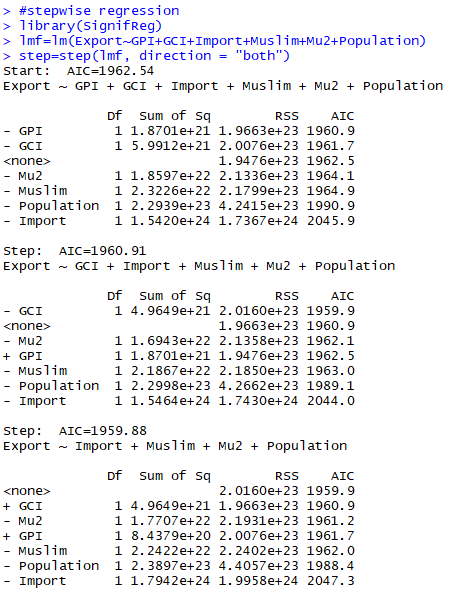
**Population**: Ο πληθυσμός της χώρας

2.2 Εξειδίκευση του υποδείγματος

Στο υπόδειγμα μας, η μεταβλητή στόχος- ανεξάρτητη μεταβλητη, θα είναι το Export. Προκειμένου να εξετάσουμε ποιες θα είναι οι ερμηνευτικές μεταβλητές που θα συμπεριλάβουμε στο υπόδειγμα, θα χρησιμοποιήσουμε το κριτήριο πληροφορίας AIC.

Παρατηρούμε ότι το βέλτιστο μοντέλο είναι το ακόλουθο:

Export ~ Import + Muslim + Mu2 + Population



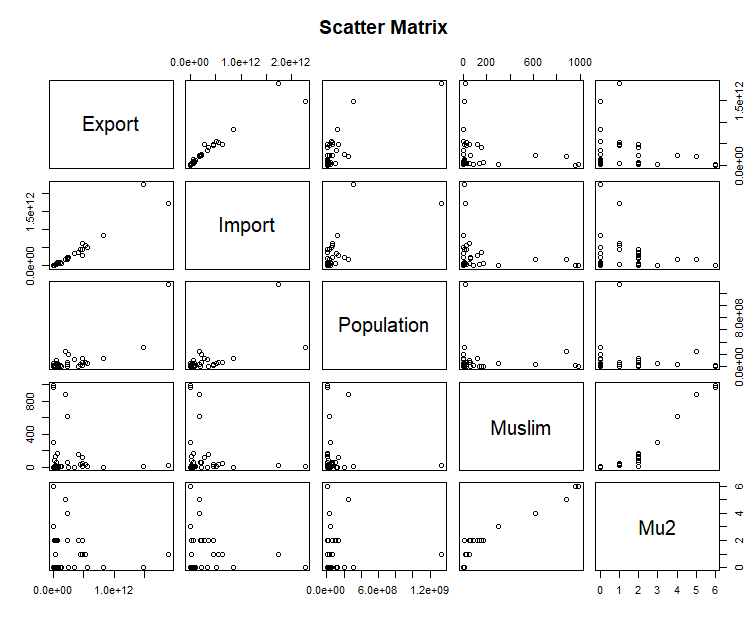
**3. Εμπειρική Ανάλυση**

3.1 Δεδομένα

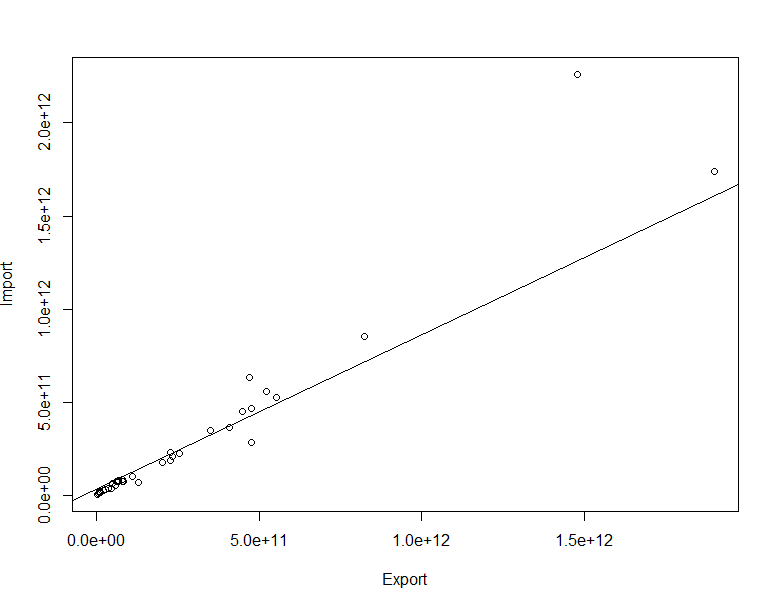
Τα δεδομένα προήλθαν από το http://wiki.stat.ucla.edu. Συγκεκριμένα χρησιμοποιήθηκαν τα "καθαρά" δεδομένα, ώστε να μην υπάρχουν missing values. Επίσης, επιλέχθηκαν τα δεδομένα του 2011 για 39 χώρες και για τις μεταβλητές που αναφέρθηκαν νωρίτερα.

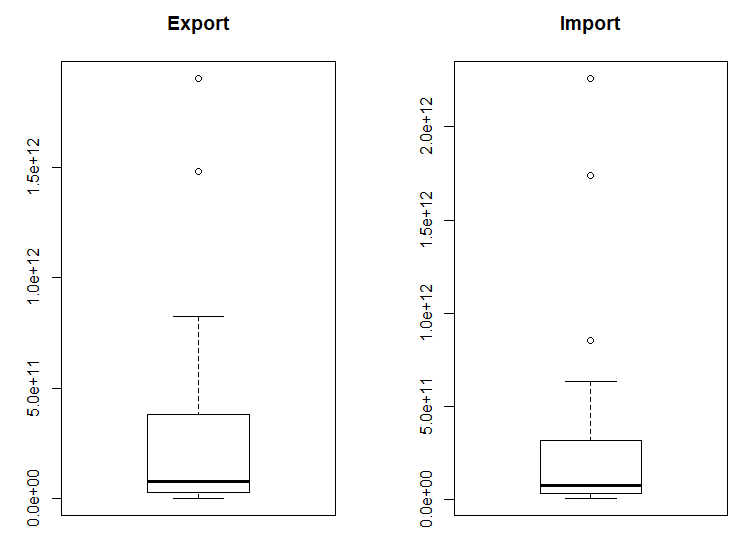
3.2 Γραφικές παραστάσεις

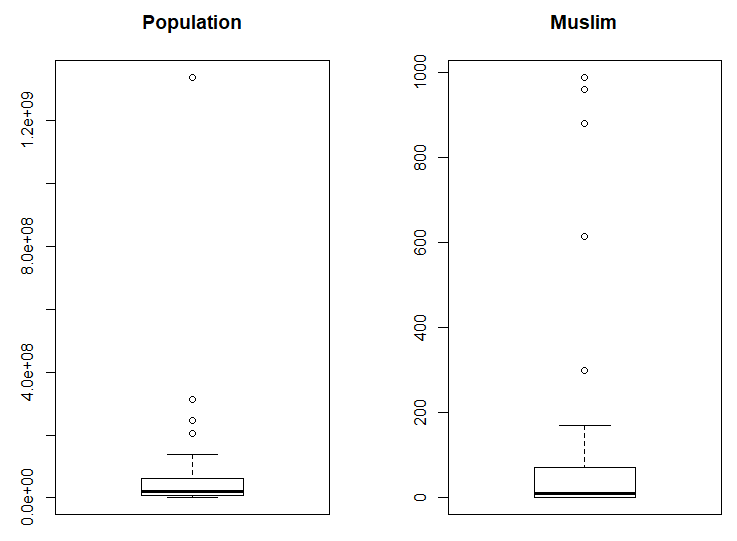
Αρχικά, έχουμε ένα scatter plot όλων των ποσοτικών μεταβλητών ανά δύο:



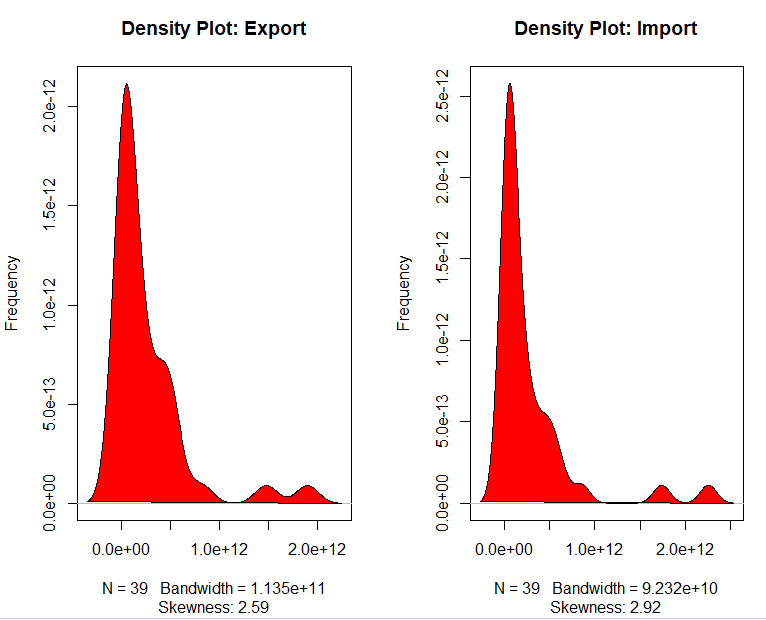
Στη συνέχεια έχουμε ένα scatter plot του Export με το Import, ώστε να παρατηρήσουμε καλύτερα τη γραμμική σχέση.

Στη συνέχεια, έχουμε κάποια boxplot των Export, Import, Population και Muslim.

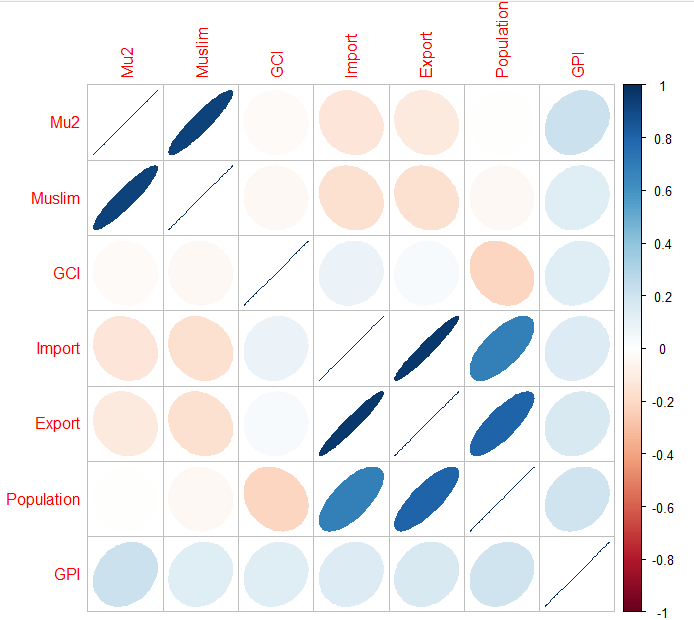


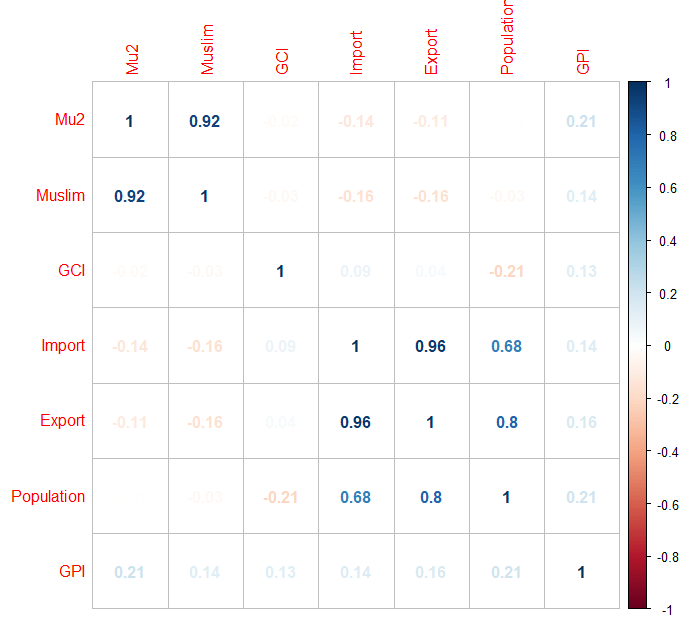
Παρατηρούμε ότι έχουμε κάποια outliers σε όλες τις μεταβλητές.

Τέλος, έχουμε δύο density plot, ώστε να δούμε τις κατανομές των δύο μεταβλητών, που ιδανικά θα θέλαμε να είναι περίπου κανονικές.



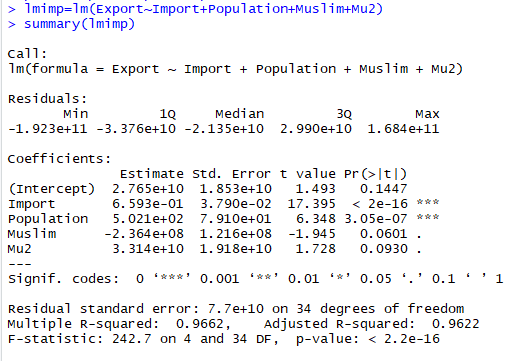
3.3 Εκτίμηση του υποδείγματος

Αρχικά, ας δούμε δύο διαγράμματα σχετικά με τον συντελεστή pearson, ώστε να είμαστε σίγουροι για τη γραμμική σχέση μεταξυ των μεταβλητών.



Στη συνέχεια, προχωρούμε στην εκτίμηση του υποδείγματος

Export ~ Import + Muslim + Mu2 + Population



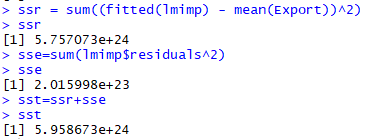
Παρατηρούμε ότι έχουμε το εξής:

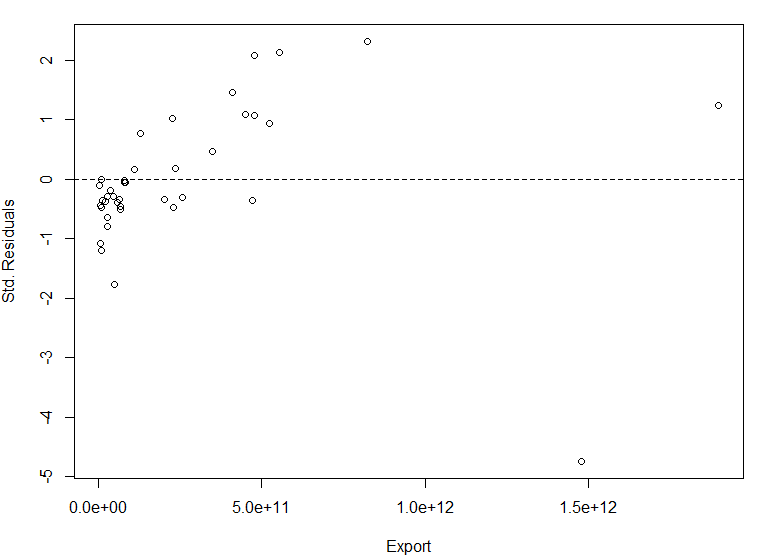
Export= 2.765\*1010+ 0.6593\* import+502.1\*population-2.364\*108\* Muslim +3.31410\*Mu2

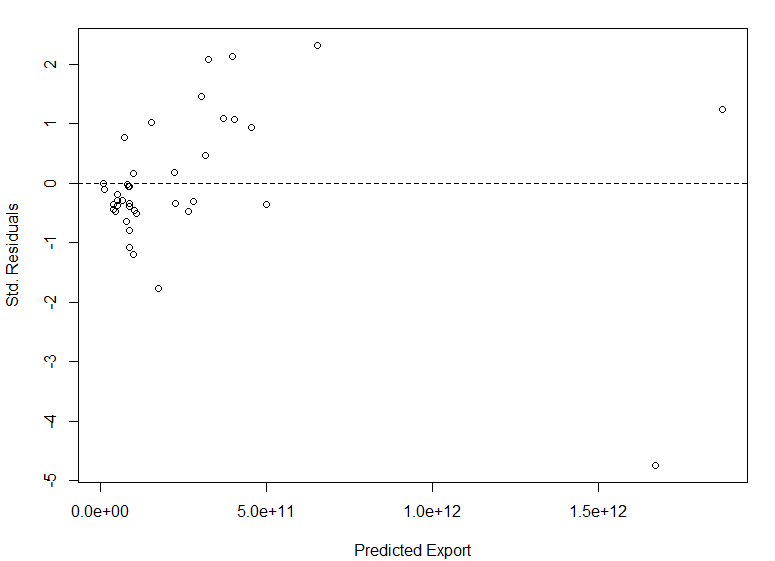
To R2=96.62% και το adj-R2= 96.22%

Στη στήλη Std. Error βλέπουμε τα τυπικά σφάλματα των μεταβλητών αλλά και του σταθερού όρου.

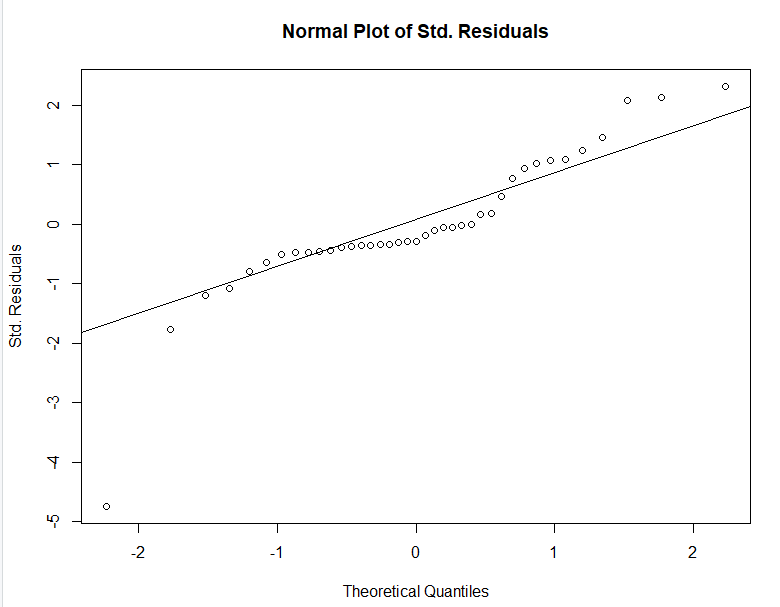
Παρακάτω βλέπουμε τα SSR, SSE και SST:



Στη συνέχεια έχουμε ένα διάγραμμα καταλοίπων- πραγματικών τιμών:

Και ένα διάγραμμα καταλοίπων - εκτιμημένων τιμών:

Τέλος, έχουμε ένα διάγραμμα των καταλοίπων:



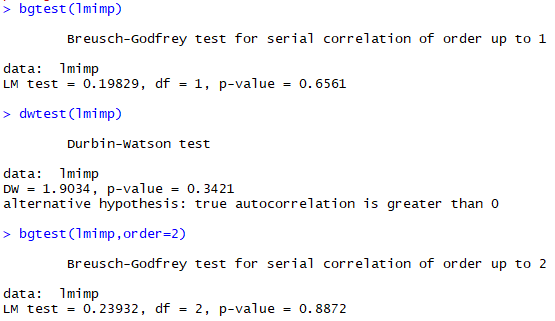
**4. Διαγνωστικοί έλεγχοι**

4.1 Έλεγχος αυτοσυσχέτισης

Σχετικά με την αυτοσυσχέτιση, πραγματοποιούμε ελέγχους Breusch-Godfrey αλλά και Durbin- Watson.

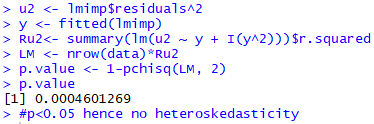
Παρατηρούμε ότι η p-value είναι μεγάλη και στα δύο αυτά tests, επομένως θα πρέπει να δεχτούμε τη μηδενική υπόθεση, που μας δίνει ότι δεν έχουμε αυτοσυσχέτιση πρώτης τάξης στα κατάλοιπα μας.

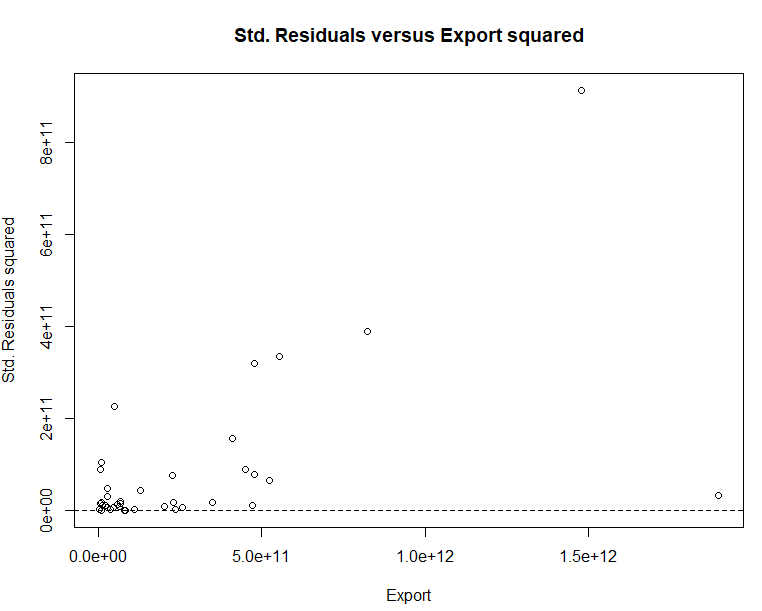
Πραγματοποιούμε έλεγχο Breusch-Godfrey και για δεύτερης τάξης, αλλά τα αποτελέσματα είναι τα ίδια.



4.2 Έλεγχος ετεροσκεδαστικότητας

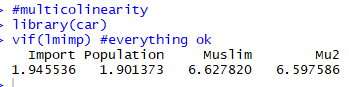
Προκειμένου να ελέγξουμε την ετεροσκεδαστικότητα των καταλοίπων θα πραγματοποιήσουμε έλεγχο White:



Παρατηρούμε ότι η p-value= 0.0004<0.10 ή 0.05, ανάλογα με τη στάθμη σημαντικότητας. Επομένως, μπορούμε να απορρίψουμε την H0 και να δεχτούμε την εναλλακτική υπόθεση, ότι δηλαδη υπάρχει ετεροσκεδαστικότητα.

4.3 Έλεγχος πολυσυγγραμμικότητας

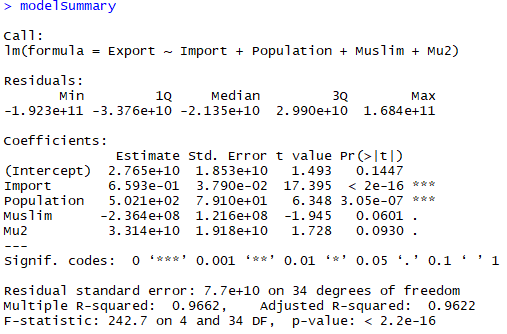
Εδώ θα χρησιμοποιήσουμε ως κριτήριο τον παράγοντα διόγκωσης διακύμανσης.



Παρατηρούμε ότι το VIF<10, επομένως δεν έχουμε πρόβλημα πολυσυγγραμμικότητας στο μοντέλο μας και καμιά μεταβλητή δεν πρέπει να εξαιρεθεί.

**5. ΕΛΕΓΧΟΙ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ**

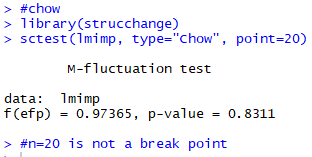
Επανερχόμαστε στο γραμμικό μοντέλο μας, ώστε να εξετάσουμε τη σημαντικότητα των ερμηνευτικών μεταβλητών.



Παρατηρούμε ότι σε στάθμη σημαντικότητας 10%, οι μεταβλητές είναι όλες σημαντικές, ενώ σε στάθμη σημαντικότητας 5% είναι μόνο οι Import και Population.

**6. Έλεγχος ευστάθειας- Έλεγχος Chow**

Χωρίσαμε το δείγμα των παρατηρήσεων σε 20 και 19, δηλαδή εξετάζουμε αν το 20 είναι break point για το μοντέλο μας.



Παρατηρούμε ότι η p-value=0.8311>0.05, επομένως δεχόμαστε την Η0, δηλαδή έχουμε ευσταθείς συντελεστές.

**7. Συμπεράσματα**

Έπειτα από πολλούς ελέγχους που πραγματοποιήσαμε, η μόνη παθογένεια προκύπτει από τον έλεγχο του White, δηλαδή έχουμε ετεροσκεδαστικά κατάλοιπα. Επομένως, θα πρέπει πριν προχωρήσουμε να αντιμετωπίσουμε αυτό το πρόβλημα, ενδεχομένως χρησιμοποιώντας robust τυπικά σφάλματα- πίνακα συνδιακύμανσης συνεπή ως προς την ετεροσκεδαστικότητα, προκειμένου να είναι σωστοί όλοι οι ελεγχοι και τα συμπεράσματα μας για το υπόδειγμα αυτό.

Σε γενικές γραμμές είδαμε ότι το export της χώρας εξαρτάται κυρίως από το import αλλά και το population αυτής.

Στη συνέχεια θα μπορούσαμε ίσως να προχωρήσουμε σε προβλέψεις, είτε να χρησιμοποιήσουμε και παλαιότερα δεδομένα ώστε να φτιάξουμε ένα νεό υπόδειγμα.

**Βιβλιογραφία**

[1] Φ. Κολυβά- Μαχαίρα Ε. Μπόρα- Σέντα, Στατιστική, Εκδόσεις Ζήτη, 2014

[2] Torsten Hothorn, Brian S. Everitt, A Handbook of Statistical Analyses Using R, second edition, Chapman and Hall, 2009

[3] http://r-statistics.co/Linear-Regression.html

[4] http://stats.stackexchange.com

[5] http://wiki.stat.ucla.edu/socr/index.php/SOCR\_Data\_GlobalPeaceIndex\_2001\_2011#Partial\_.28Clean.29\_Annual\_Global\_Peace.2C\_Trade\_and\_Competitiveness\_data\_.282001-2011.29

**Παράρτημα**

