

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ ΔΙΟΙΚΗΣΗ
ΕΠΙΧΕΙΡΗΣΕΩΝ (MBA)**

Διπλωματική Εργασία

**ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΠΕΤΡΕΛΑΙΟΕΙΔΩΝ ΣΤΗΝ ΕΥΡΩΠΑΙΚΗ
ΕΝΩΣΗ**

ΕΥΘΥΜΙΑ ΔΑΝΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ ΜΑΡΑΒΕΛΑΚΗΣ ΠΕΤΡΟΣ

Πειραιάς, 2021

ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

«Δηλώνω υπεύθυνα ότι η διπλωματική εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών, του Πανεπιστημίου Πειραιώς, στη Διοίκηση Επιχειρήσεων : MBA» με τίτλο:

«... *Στρατηγική Ανάλυση Π.Σ.Ρ.Ε.Α.Ι.Ο.Ε.Ι.Δ.Ε.Ν. στην Ευρωπαϊκή Ένωση* ...»

έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας, αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου»

Υπογραφή Μεταπτυχιακού Φοιτητή Ονοματεπώνυμο

Καλλιμα
.....

Αφιερώνεται στους γονείς μου

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΠΕΤΡΕΛΑΙΟΕΙΔΩΝ ΣΤΗΝ ΕΥΡΩΠΑΙΚΗ ΕΝΩΣΗ

Δανία Ευθυμία

Σημαντικοί όροι: πετρέλαιο, κατανάλωση, Ευρωπαϊκή Ένωση

Περίληψη

Στην παρούσα εργασία μελετάται η τελική ενεργειακή κατανάλωση πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης.

Αρχικά, στο πρώτο κεφάλαιο της εργασίας εμπεριέχεται η εισαγωγή. Περιλαμβάνεται ο στόχος και ο αντικειμενικός σκοπός της παρούσας διπλωματικής. Επιπροσθέτως, γίνεται μια σύντομη ιστορική αναδρομή σε ότι αφορά το πετρέλαιο και παρατίθενται κάποια στοιχεία για την κατανάλωση πετρελαιοειδών.

Στη συνέχεια στο κεφάλαιο δύο, περιγράφονται και εξηγούνται συνοπτικά ορισμοί της Στατιστικής και οι μεθοδολογίες που χρησιμοποιήθηκαν στο ερευνητικό μέρος. Επιπλέον, παρουσιάζονται, βασικές γραφικές παραστάσεις, η μέθοδος της γραμμικής παλινδρόμησης και η πρόβλεψη των χρονοσειρών.

Στο τρίτο κεφάλαιο, γίνεται η παρουσίαση των δεδομένων και των μεταβλητών που χρησιμοποιούνται στην παρούσα εργασία. Επιπροσθέτως, αναφέρονται τα ερευνητικά ερωτήματα που θα απαντηθούν στο τέταρτο κεφάλαιο.

Στο τέταρτο κεφάλαιο, το οποίο αποτελεί και το σημαντικότερο μέρος της διπλωματικής εργασίας, παρουσιάζεται η κύρια ανάλυση μέσω της χρήσης των εργαλείων που αναφέρθηκαν στο κεφάλαιο δύο. Συγκεκριμένα, γίνεται μια στατιστική ανάλυση στις μεταβλητές που θα χρησιμοποιηθούν. Στη συνέχεια μελετάται η επίδραση του μετασχηματισμού του αργού πετρελαίου σε καταναλώσιμα προϊόντα, η παραγωγή πετρελαιοειδών, οι εισαγωγές πετρελαίου και φυσικού αερίου, προκειμένου να διερευνηθεί κατά πόσο οι παράγοντες αυτοί επηρεάζουν την τελική κατανάλωση πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης. Τέλος, επιχειρείται η πρόβλεψη της κατανάλωσης πετρελαιοειδών στην Ευρωπαϊκή Ένωση για τα επόμενα δύο χρόνια.

Στο πέμπτο και τελευταίο κεφάλαιο της εργασίας, έχοντας λάβει υπόψη τα αποτελέσματα που προέκυψαν από το ερευνητικό μέρος, αποτυπώνονται τα κύρια συμπεράσματα της παρούσας διπλωματικής, καθώς και προτάσεις για μελλοντική έρευνα.

Ευχαριστίες

Ολοκληρώνοντας τις Μεταπτυχιακές σπουδές , θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μαραβελάκη Πέτρο, για την πολύτιμη βοήθεια και τις συμβουλές που μου έδωσε καθ' όλη την διάρκεια της εκπόνησης της διπλωματικής εργασίας.

Επιπλέον θα ήθελα να εκφράσω την αμέριστη ευγνωμοσύνη μου στην οικογένειά μου που είναι πάντα δίπλα μου και στηρίζει κάθε επιλογή μου.

Τέλος, θέλω να ευχαριστήσω από καρδιάς τις φίλες μου Ελένη και Σπυριδούλα και το αγόρι μου Αποστόλη, για την συμπαράσταση και την υπομονή τους.

Κατάσταση Διαγραμμάτων

Διάγραμμα 1 : Διάγραμμα Διασποράς για τις μεταβλητές ConsOil και TransfOil	38
Διάγραμμα 2 : Διάγραμμα Διασποράς για τις μεταβλητές ConsOil και ImpGas	39
Διάγραμμα 3 : Διάγραμμα Διασποράς για τις μεταβλητές TransfOil και ImpGas	39
Διάγραμμα 4 : Διάγραμμα Διασποράς για τις Μεταβλητές ProdOil και ConsOil	40
Διάγραμμα 5 : Διάγραμμα Διασποράς για τις μεταβλητές ImpOil και ConsOil.....	40
Διάγραμμα 6 : Ιστόγραμμα για την μεταβλητή ConsOil	42
Διάγραμμα 7 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ConsOil.....	43
Διάγραμμα 8 : Θηκόγραμμα για την μεταβλητή ConsOil	43
Διάγραμμα 9 : Ιστόγραμμα για την μεταβλητή TransfOil	44
Διάγραμμα 10 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή TransfOil.....	45
Διάγραμμα 11 : Θηκόγραμμα για την μεταβλητή TransfOil.....	45
Διάγραμμα 12 : Ιστόγραμμα για την μεταβλητή ImpGas	46
Διάγραμμα 13 : Διάγραμμα Ελέγχου Κανονικότητας για την Μεταβλητή ImpGas.....	46
Διάγραμμα 14 : Θηκόγραμμα για την μεταβλητή ImpGas	47
Διάγραμμα 15 : Ιστόγραμμα για την μεταβλητή ConsOil	48
Διάγραμμα 16 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ConsOil.....	49
Διάγραμμα 17 : Θηκόγραμμα για την μεταβλητή ConsOil.....	49
Διάγραμμα 18 : Ιστόγραμμα για την μεταβλητή TransfOil	50
Διάγραμμα 19 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή TransfOil.....	50
Διάγραμμα 20 : Θηκόγραμμα για την μεταβλητή TransfOil.....	51
Διάγραμμα 21 : Ιστόγραμμα για την μεταβλητή ImpGas	52
Διάγραμμα 22 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ImpGas.....	52
Διάγραμμα 23 : Θηκόγραμμα για την μεταβλητή ImpGas	53
Διάγραμμα 24 : Ιστόγραμμα για την μεταβλητή ConsOil	54
Διάγραμμα 25 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ConsOil.....	55
Διάγραμμα 26 : Θηκόγραμμα για την μεταβλητή ConsOil.....	55
Διάγραμμα 27 : Ιστόγραμμα για την μεταβλητή TransfOil	56
Διάγραμμα 28 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή TransfOil.....	56
Διάγραμμα 29 : Θηκόγραμμα για την μεταβλητή TransfOil.....	57
Διάγραμμα 30 : Ιστόγραμμα για την μεταβλητή ImpGas	58
Διάγραμμα 31 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ImpGas.....	58
Διάγραμμα 32 : Θηκόγραμμα για την μεταβλητή ImpGas	59
Διάγραμμα 33 : Διάγραμμα Ελέγχου Κανονικότητας Καταλοίπων	64
Διάγραμμα 34 : Διάγραμμα Καταλοίπων ως προς τις Προβλεπόμενες Τιμές.....	64
Διάγραμμα 35 : Διάγραμμα Ελέγχου Κανονικότητας Καταλοίπων	70
Διάγραμμα 36 : Διάγραμμα καταλοίπων ως προς τις Προβλεπόμενες Τιμές.....	70
Διάγραμμα 37 : Διάγραμμα Χρονοσειρών για την μεταβλητή ConsOil.....	72
Διάγραμμα 38 : Αυτοσυσχέτιση των μεταβλητών ConsOil και ConsOilt-1.	73
Διάγραμμα 39 : Διάγραμμα Χρονοσειρών για την μεταβλητή ConsOil μετά από μία Διαφόριση.....	73
Διάγραμμα 40 : Αυτοσυσχέτιση των μεταβλητών ConsOil και ConsOilt-1 μετά από μία Διαφόριση.....	74
Διάγραμμα 41 : Κατάλοιπα Αυτοσυσχετίσεων.....	77
Διάγραμμα 42 : Κατάλοιπα Μερικών Αυτοσυσχετίσεων.....	77
Διάγραμμα 43 : Προβλέψεις Κατανάλωσης Πετρελαιοειδών για δύο χρόνια	78

Κατάσταση Πινάκων

Πίνακας 1	: Συμβολισμοί Μεταβλητών.....	41
Πίνακας 2	: Έλεγχος Σημαντικότητας Kolmogorov-Smirnov Shapiro-Wilk.....	53
Πίνακας 3	: Έλεγχος Κανονικότητας Kolmogorov-Smirnov και Shapiro-Wilk.....	59
Πίνακας 4	: Πίνακας Συσχετίσεων των μεταβλητών κατά Pearson.....	60
Πίνακας 5	: Πίνακας Συσχετίσεων κατά Spearman	60
Πίνακας 6	: Συντελεστής Προσδιορισμού R.....	61
Πίνακας 7	: Ανάλυση Διακύμανσης ANOVA	62
Πίνακας 8	: Συντελεστές Παλινδρόμησης.....	62
Πίνακας 9	: Τεστ Κανονικότητας Καταλοίπων	63
Πίνακας 10	: Έλεγχος Τυχειότητας για τα Κατάλοιπα.....	65
Πίνακας 11	: Συμβολισμός της Μετασχηματισμένης Εξαρτημένης Μεταβλητής ConsOil	66
Πίνακας 12	: Πίνακας Συσχετίσεων Μεταβλητών κατά Pearson	66
Πίνακας 13	: Πίνακας Συσχετίσεων Μεταβλητών κατά Spearman.....	67
Πίνακας 14	: Συντελεστής Προσδιορισμού R.....	67
Πίνακας 15	: Ανάλυση Διακύμανσης ANOVA.....	68
Πίνακας 16	: Συντελεστές Παλινδρόμησης.....	68
Πίνακας 17	: Τεστ Κανονικότητας Καταλοίπων	69
Πίνακας 18	: Έλεγχος Τυχειότητας για τα Κατάλοιπα.....	71
Πίνακας 19	: Υπολογισμός AIC για Υποψήφια Μοντέλα.....	75
Πίνακας 20	: Έλεγχος Ροών Καταλοίπων	75
Πίνακας 21	: Προβλέψεις Κατανάλωσης Πετρελαιοειδών για δύο χρόνια.....	78

Κατάσταση Σχημάτων

Σχήμα 1	: Κατανάλωση Πετρελαιοειδών στην Ευρωπαϊκή Ένωση	4
Σχήμα 2	: Ενεργειακή Κατανάλωση ανά Καύσιμο	5
Σχήμα 3	: Ιστόγραμμα Κατανομής Συχνοτήτων.....	7
Σχήμα 4	: Κυκλικό Διάγραμμα (pie chart)	8
Σχήμα 5	: Ακιδωτό Διάγραμμα (bar chart)	9
Σχήμα 6	: Σχήματα Κατανομής.....	12
Σχήμα 7	: Διάγραμμα Πλαισίου- Αποκλίσεων	13
Σχήμα 8	: Η Κανονική Κατανομή.....	14
Σχήμα 9	: Διαγράμματα Διασποράς.....	18
Σχήμα 10	: Διαγράμματα Διασποράς, Καμπυλόγραμμη Σχέση	18
Σχήμα 11	: Διάγραμμα Διασποράς	19
Σχήμα 12	: Διάγραμμα Διασποράς, Μετασχηματισμός με χρήση Αλγορίθμου.....	20
Σχήμα 13	: Διαγραμματική Απεικόνιση των Υποθέσεων της Γραμμικής Παλινδρόμησης	22
Σχήμα 14	: Μέτρα Μεταβλητότητας Μοντέλου.....	24
Σχήμα 15	: Παράδειγμα Χρονοσειράς για το Πρόσθετο Μοντέλο (μηνιαία παραγωγή Γάλακτος).....	32
Σχήμα 16	: Παράδειγμα Χρονοσειράς για Πολλαπλασιαστικό Μοντέλο	33

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	1
1.1 Αντικείμενο Εργασίας	1
1.2 Σκοπός Εργασίας.....	1
1.3 Ιστορική Αναδρομή του Πετρελαίου	2
1.4 Κατανάλωση Πετρελαιοειδών στην Ευρωπαϊκή Ένωση	3
ΚΕΦΑΛΑΙΟ 2: ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ	6
2.1 Πίνακες.....	6
2.2 Διαγράμματα	7
2.3 Αριθμητικά Περιγραφικά Μέτρα.....	9
2.3.1. Μέτρηση Κεντρικής Τάσης	10
2.3.2 Μέτρα Διασποράς ή Μεταβλητότητας.....	10
2.3.3. Έλεγχος σχήματος	12
2.4 Διάγραμμα Πλαισίου-Απολήξεων (Boxplot).....	13
2.5 Κανονική κατανομή	14
2.6 Στατιστικός Έλεγχος Υποθέσεων.....	15
2.6.1 Μηδενική και Εναλλακτική Υπόθεση.....	15
2.6.2 Σφάλματα και Επίπεδο σημαντικότητας	15
2.7 Ανάλυση Παλινδρόμησης	16
2.7.1 Συντελεστής Συσχέτισης και Διαγράμματα Διασποράς	16
2.7.2 Απλή Γραμμική Παλινδρόμηση.....	20
2.7.3 Μέθοδος Ελαχίστων Τετραγώνων	22
2.7.4 Η μεταβλητότητα του Μοντέλου.....	23
2.7.5 Έλεγχος Στατιστικής Σημαντικότητας	26
2.8 Πολλαπλή Γραμμική Παλινδρόμηση	27
2.8.1 Εκτίμηση της Εξίσωσης της Πολλαπλής Γραμμικής Παλινδρόμησης.....	28
2.8.2 Συντελεστής Πολλαπλού προσδιορισμού	29
2.8.3 Έλεγχοι Στατιστικής Σημαντικότητας	29
2.9 Χρονοσειρές.....	31
2.9.1 Ορισμός.....	31
2.9.2 Συνιστώσες Χρονοσειρών	31
2.9.3 Το Προσθετικό Μοντέλο.....	31
2.9.4 Το Πολλαπλασιαστικό Μοντέλο	32
2.9.5 Αιτιοκρατία και Στοχαστικότητα.....	33

2.9.6 Στασιμότητα.....	33
2.9.7 Γραμμικά Μοντέλα Πρόβλεψης	33
2.9.8 Moving Average Model (MA).....	34
2.9.9 Auto Regressive Moving Average (ARMA).....	35
2.9.10 Auto Regressive Integrated Moving Average (ARIMA) Model	35
ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ	36
3.1 Δεδομένα	36
3.2 Ερευνητικά ερωτήματα	36
ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ.....	38
4.1 Ανάλυση Μεταβλητών.....	38
4.2 Κατανομές Δεδομένων	41
4.2.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil).....	41
4.2.2 Μεταβλητή Μετασχηματισμός Πετρελαιοειδών (TransfOil)	43
4.2.3 Μεταβλητή Εισαγωγές Φυσικού Αερίου (ImpGas)	45
4.3 Πρώτη Ομάδα Χωρών.....	47
4.3.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil).....	47
4.3.2 Μεταβλητή Μετασχηματισμός Πετρελαιοειδών (TransfOil)	49
4.3.3 Μεταβλητή Εισαγωγή Φυσικού Αερίου (ImpGas).....	51
4.4 Δεύτερη Ομάδα χωρών	53
4.4.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil).....	53
4.4.2 Μεταβλητή Μετασχηματισμός Πετρελαίου (TransfOil).....	55
4.4.3 Μεταβλητή Εισαγωγή Φυσικού Αερίου (ImpGas).....	57
4.5 Ανάλυση Παλινδρόμησης: Πρώτη Ομάδα χωρών.....	59
4.5.1 Έλεγχος Συσχετίσεων Μεταβλητών	60
4.5.2 Αποτελέσματα Γραμμικής Παλινδρόμησης.....	61
4.5.3 Ανάλυση Καταλοίπων	63
4.6 Ανάλυση Παλινδρόμησης: Δεύτερη Ομάδα Χωρών.....	66
4.6.1 Έλεγχος Συσχετίσεων Μεταβλητών	66
4.6.2 Αποτελέσματα Γραμμικής Παλινδρόμησης.....	67
4.6.3 Ανάλυση Καταλοίπων	69
4.7 Προβλέψεις Κατανάλωσης Πετρελαιοειδών.....	71
4.7.1 Έλεγχος Στασιμότητας	72
4.7.2 Πρόβλεψη με τη μέθοδο ARIMA	74
Κεφάλαιο 5: Συμπεράσματα-Προτάσεις για Μελλοντική Έρευνα	79
5.1 Συμπεράσματα.....	79
5.2 Προτάσεις για Μελλοντική Έρευνα	81

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο Εργασίας

Στην παρούσα διπλωματική εργασία γίνεται προσπάθεια να αποτυπωθούν, όσο το δυνατόν με μεγαλύτερη ακρίβεια, οι προσδιοριστικοί παράγοντες που επηρεάζουν την τελική κατανάλωση πετρελαιοειδών στην Ευρωπαϊκή Ένωση. Επιπλέον γίνεται προσπάθεια πρόβλεψης των τιμών της κατανάλωσης για τα έτη, 2019-2020. Η εργασία αποτελείται από τη θεωρητική ανάλυση, και από την στατιστική μελέτη των δεδομένων.

Οι παράγοντες που θα αναλυθούν στη συνέχεια, αποδεικνύεται ότι παίζουν καθοριστικό ρόλο στον τρόπο που καταναλώνονται τα πετρελαιοειδή από τους τελικούς χρήστες. Η κατανάλωση των πετρελαιοειδών αποτελεί αντικείμενο συνεχούς έρευνας και μελέτης, καθώς τα δεδομένα στον ενεργειακό κλάδο αλλάζουν ραγδαία τα τελευταία χρόνια.

Γενικότερα, γίνεται αναφορά στις στατιστικές μεθοδολογίες που χρησιμοποιήθηκαν. Σημαντικό κομμάτι της παρούσας εργασίας αποτελεί το τέταρτο κεφάλαιο, το οποίο περιλαμβάνει την στατιστική ανάλυση των παραγόντων που επηρεάζουν την κατανάλωση, καθώς και την μεθοδολογία που χρησιμοποιείται για την πρόβλεψη των τιμών της.

Η σημασία που έχει το πετρέλαιο, αλλά και τα παράγωγα προϊόντα του, στην παγκόσμια κοινότητα είναι μεγάλη. Αυτό συμβαίνει διότι επηρεάζει αρκετές πτυχές της καθημερινότητας του ανθρώπου. Συγκεκριμένα, η ενεργειακή εξάρτηση της Ευρωπαϊκής Ένωσης από εξωτερικές πηγές ολοένα και αυξάνεται, ιδιαίτερα τα τελευταία χρόνια λόγω της μεγάλης ζήτησης που υπάρχει από τους πολίτες. Το πετρέλαιο συχνά αποκαλείται και μαύρος χρυσός λόγω της τεράστιας αξίας που έχει αποκτήσει, ορίζοντας το μέλλον της οικονομίας με τις αυξομειώσεις της τιμής του.

Στην παρούσα εργασία γίνεται προσπάθεια προσέγγισης των παραπάνω θεματικών ενότητων. Επιπλέον, παρουσιάζονται και αναλύονται τα αποτελέσματα της στατιστικής έρευνας, η οποία βασίστηκε σε δεδομένα, τα οποία αντλήθηκαν από το Ενεργειακό Ισοζύγιο της Ευρωπαϊκής Ένωσης.

1.2 Σκοπός Εργασίας

Η παρούσα εργασία αποτελείται από δύο μέρη, το θεωρητικό και το πρακτικό. Συγκεκριμένα η ανάλυση γίνεται πρώτα σε γενικό επίπεδο και στη συνέχεια σε ειδικό, όπου και αναπτύσσεται το ερευνητικό μέρος.

Σκοπός της παρούσας εργασίας είναι να μελετηθούν και να προσεγγιστούν οι παράγοντες που επηρεάζουν την κατανάλωση πετρελαιοειδών στη Ευρωπαϊκή Ένωση. Εξίσου σημαντικό είναι να προσδιοριστούν μελλοντικές τιμές της κατανάλωσης, όσο το δυνατόν ορθά και σωστά γίνεται, με τα δεδομένα που υπάρχουν διαθέσιμα.

Μέσα από την παρούσα εργασία, οι ενδιαφερόμενοι θα έχουν την δυνατότητα να αποκτήσουν γενικές γνώσεις σε ότι αφορά τη στατιστική ανάλυση δεδομένων και την μεθοδολογία που είναι απαραίτητη για τη δημιουργία ενός μοντέλου παλινδρόμησης, καθώς και στον τρόπο πρόβλεψης χρονοσειρών.

1.3 Ιστορική Αναδρομή του Πετρελαίου

Τα πρώτα κοιτάσματα πετρελαίου ανακαλύφθηκαν περίπου πριν από 160 χρόνια. Από την αρχή βέβαια της δημιουργίας των κοινωνιών ο άνθρωπος στράφηκε στην κατανάλωση ενέργειας για την διασφάλιση της επιβίωσής του. Το πετρέλαιο υπήρξε καθοριστικός παράγοντας για την αρχή της βιομηχανικής επανάστασης, με τις πρώτες μηχανές εσωτερικής καύσης να αποτελούν την δύναμη την οποία ώθησε την παγκόσμια οικονομία και βιομηχανία σε τεχνολογικά άλματα.

Ο Γερμανός ορυκτολόγος Agricola, σύστησε για πρώτη φορά την λέξη πετρέλαιο στην παγκόσμια κοινότητα το 1556.

Ιστορικά, η πρώτη αναφορά του πετρελαίου γίνεται στην Βίβλο. Όπου περιγράφεται πως Νώε κάλυψε το εξωτερικό μέρος της Κιβωτού με ένα υλικό ασφαλτικής σύστασης, για την προστασία του καραβιού από ενδεχόμενη καταιγίδα. Λίγο αργότερα, ο Ηρόδοτος περιγράφει την ύπαρξη ενός πηγαδιού, το οποίο αντί για νερό εμπεριείχε άσφαλο. Επιπλέον, φυλές όπως οι Σουμέριοι και οι Βαβυλώνιοι χρησιμοποίησαν το πετρέλαιο που βρέθηκε, στο Χιτ του ποταμού Ευφράτη, πριν από 5000 χρόνια. Παρόμοιες διαρροές καταγράφηκαν ιστορικά στην περιοχή της Μεσοποταμίας και γενικότερα στην ανατολική Μεσόγειο. Τέλος, η Νεκρά θάλασσα τα χρόνια εκείνα, ονομαζόταν και Ασφαλίτις Λίμνη, διότι στις όχθες της έβγαινε πετρέλαιο από κάποια υποβρύχια διαρροή πιθανώς.

Στις ΗΠΑ, έως τον 19^ο αιώνα, το πετρέλαιο χρησιμοποιούνταν για φως, ακριβώς με τον ίδιο τρόπο που το χρησιμοποιούσαν και οι αρχαίοι Έλληνες και τους Ρωμαίοι. Τον Αύγουστο του 1859 στην δυτική Πενσυλβάνια, έγινε η πρώτη προσπάθεια αναζήτησης πετρελαίου από τον Έντουιν Ντρέικ και με αυτόν τον τρόπο άνοιξε ο δρόμος για τη βιομηχανία του πετρελαίου. Την ίδια χρονολογία ανακαλύφθηκαν κοιτάσματα πετρελαίου στην Ευρώπη και την Ανατολή.

Τον 20^ο αιώνα η Βιομηχανική Επανάσταση ήταν γεγονός. Ο φωτισμός με τη χρήση πετρελαίου σταμάτησε να παίζει πρωταρχικό ρόλο και τελικά το πετρέλαιο αποτέλεσε την πρώτη πηγή ενέργειας παγκοσμίως.

Μέχρι το 1870 το πετρέλαιο που παραγόταν ήταν περίπου 1.000.000 το χρόνο και στη συνέχεια με την πάροδο των χρόνων έφτασε να ξεπερνά τους 3.000.0000 τόνους. Την σημερινή εποχή το πετρέλαιο αποτελεί την κύρια πηγή ενέργειας παγκοσμίως και από αυτό, εξαρτάται το μέλλον της οικονομίας κάθε χώρας.

Η πρώτη πετρελαϊκή εταιρεία των ΗΠΑ ιδρύθηκε από τον δικηγόρο George H. Bissell και από τον επιχειρηματία James Townsend. Την πρωτότυπη για την εποχή ιδέα συνέβαλε άθελά του ο καθηγητής του Γέιλ Dr. Benjamin Silliman, αναλύοντας το πετρέλαιο που συλλέχθηκε από τα νερά της Πενσυλβάνια, υποστήριξε πως χρησιμοποιώντας το θα μπορούσε να δημιουργηθεί ένα πρώτης τάξεως φωτιστικό. Η εταιρία αρχικά ονομάστηκε Pennsylvania Rock Oil Company και αργότερα μετονομάστηκε σε Seneca Oil Company.

Στη συνέχεια ο Bissell, προσπάθησε να εντοπίσει το μέρος από το οποίο χυνόταν το πετρέλαιο στο νερό. Προσλαμβάνοντας τον Edwin L. Drake και αγοράζοντας την έκταση την οποία ενδιαφερόταν να μελετήσει, πέτυχε διάνα, βρίσκοντας όντως αποθέματα πετρελαίου στο συγκεκριμένο σημείο στην πόλη Titusville στην Πενσυλβάνια. Στις 27 Αυγούστου του 1859 εξορύχθηκε λοιπόν, πρώτη φορά από την πηγή του πετρελαίου με την βοήθεια γεωτρήπανου. Η πόλη Titusville καθώς και άλλες πόλεις στην γύρω περιοχή επωφελήθηκαν στο έπακρο την ανακάλυψη αυτή, αφού άνθισαν οικονομικά. Ο John Rockefeller, μεγάλος επιχειρηματίας της εποχής, έσπευσε κατευθείαν να αρπάξει την ευκαιρία, αγοράζοντας έκταση στο Κλιβελαντ. Δημιούργησε ένα μικρό διυλιστήριο, το οποίο τον έκανε ηγετική προσωπικότητα της αμερικανικής πετρελαιοβιομηχανίας. Το 1866 λειτουργούσε ήδη εξαγωγικό γραφείο στη Νέα Υόρκη.

Το 1867, ήταν αρκετά σημαντική χρονιά για την πορεία του πετρελαίου. Ο S.V Harkness και ο Henry Flagler ίδρυσαν την εταιρία Standard Oil Company, ευρέως γνωστή στην πετρελαϊκή βιομηχανία. Το γεγονός πως στην Πενσυλβάνια υπήρχε πηγή πετρελαίου έγινε πολύ γρήγορα γνωστό και αμέσως αρκετοί ήταν αυτοί που θέλησαν να εκμεταλλευτούν την κατάσταση και να διεκδικήσουν μερίδιο από την αυξανόμενη αυτή αγορά.

Ο John Rockefeller, έπειτα από το πλήθος ανθρώπων που εκδήλωσαν ενδιαφέρον για την αγορά του πετρελαίου, αποφάσισε να εξαγοράσει μαζικά τις μικρότερες εταιρείες και να συγχωνεύσει την δική του με τους ανταγωνιστές του. Με αυτό τον τρόπο, έως το 1870 ο Rockefeller κατάφερε να μετατρέψει την Standard Oil σε κυρίαρχο διυλιστήριο σε ολόκληρη την Πενσυλβάνια.

Τον 20^ο αιώνα, η ζήτηση του πετρελαίου άρχισε να ανεβαίνει ραγδαία. Κύριο ρόλο σε αυτό έπαιξαν τα βενζινοκίνητα αυτοκίνητα, καθώς και τα αεροπλάνα, εκτοξεύοντας την ζήτηση σε βενζίνη στα ύψη και φυσικά κάνοντας τη επείγουσα ανάγκη. Σαν αποτέλεσμα η βιομηχανία του πετρελαίου κλήθηκε να καλύψει τις ανάγκες πάρα πολλών ανθρώπων, κάτι που ξεπερνούσε τις δυνατότητες της.

Για τους παραπάνω λόγους, επιτακτική ήταν η ανάγκη για αναζήτηση πετρελαίου σε διαφορετικά μέρη, ενώ αξιοσημείωτο είναι πως την ίδια περίοδο γίνονται οι πρώτες αναφορές για την ενδεχόμενη εξάντληση των αμερικανικών πετρελαϊκών αποθεμάτων. Οι Αμερικάνοι υπουργοί αναγνωρίζοντας την σοβαρότητα της κατάστασης, ξεκινούν τις πιέσεις για να γίνουν έρευνες σε Μέση Ανατολή , Νότια Αμερική και Νοτιοανατολική Ασία. Το εγχείρημα αυτό στέφθηκε με μεγάλη επιτυχία, αφού το 1910, βρέθηκαν αποθέματα πετρελαίου σε Μεξικό, Καναδά, Περσία, Περού, και Βενεζουέλα.

Ορόσημο για το μέλλον του πετρελαίου, στάθηκε ο Β΄ Παγκόσμιος πόλεμος, όπου το πετρέλαιο κέρδισε έδαφος και έγινε το πλέον απαραίτητο καύσιμο της νέας εποχής, μετατρέποντας την βιομηχανία του πετρελαίου σε παγκόσμιο κολοσσό.

Φτάνοντας στην σημερινή εποχή, αυτό που παρατηρήθηκε ήταν πως οι πρώιμες τεχνικές εξόρυξης πετρελαίου, οδήγησαν στην εξάντληση πολλών κοιτασμάτων. Το τραγικό αποτέλεσμα ήταν να απελευθερωθούν μεγάλες ποσότητες φυσικού αερίου. Συγκεκριμένα περισσότερο από το 75% των αποθεμάτων φυσικού αερίου και πετρελαίου κατασπαταλήθηκε στις ΗΠΑ.

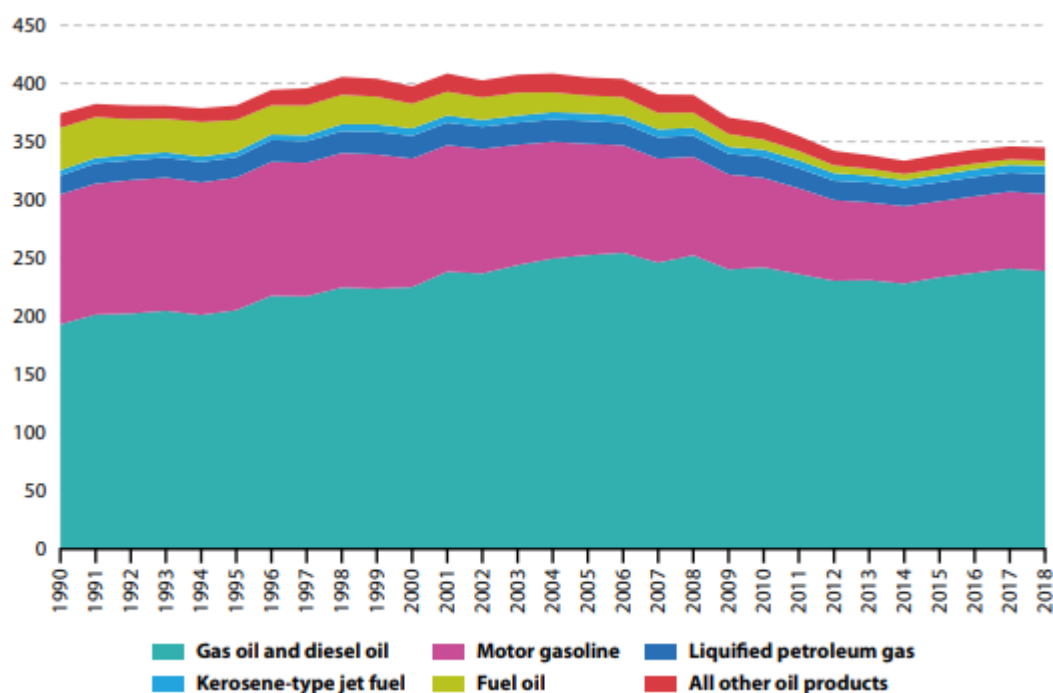
Τέλος, βέβαιο είναι πως ο σύγχρονος κόσμος είναι απαραίτητο να στραφεί σε εναλλακτικές πηγές ενέργειας και να αλλάξει ριζικά τον τρόπο που ζει και καταναλώνει. Σύγχρονες περιβαλλοντικές έρευνες έδειξαν πως τα αποθέματα πετρελαίου έχουν ξεκινήσει να περιορίζονται αρκετά.

1.4 Κατανάλωση Πετρελαιοειδών στην Ευρωπαϊκή Ένωση

Η κατανάλωση πετρελαιοειδών που θα αναλυθεί στα παρακάτω κεφάλαια, περιγράφει το σύνολο όλων των προϊόντων πετρελαίου που καταναλώθηκαν από τους τελικούς χρήστες στην Ευρωπαϊκή Ένωση. Μερικά από τα προϊόντα αυτά είναι τα παρακάτω :

- Υγραέριο
- Καύσιμα
- Κηροζίνη
- Πετρέλαιο Ντίζελ
- Μαζούτ
- Άλλα προϊόντα πετρελαίου

Στο σχήμα που παρατίθεται παρακάτω, φαίνεται η κατανάλωση πετρελαιοειδών από το έτος 1990 έως το 2018:



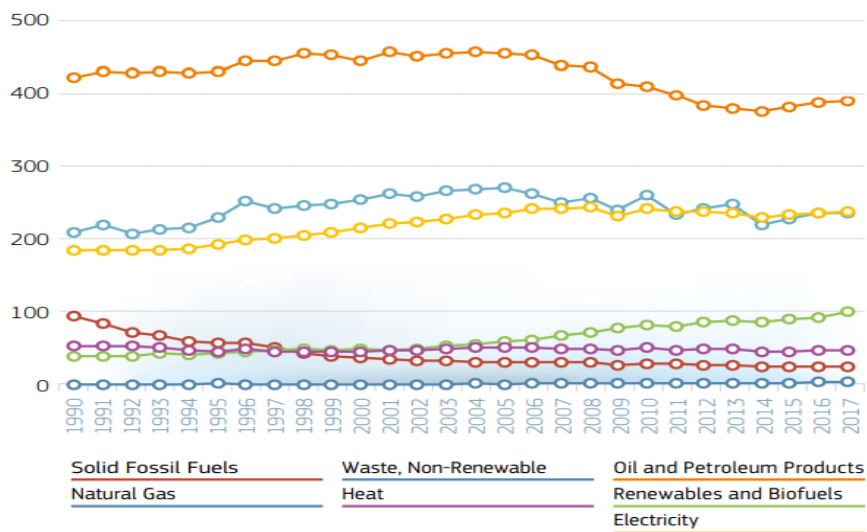
Σχήμα 1: Κατανάλωση Πετρελαιοειδών στην Ευρωπαϊκή Ένωση

Πηγή : *EU ENERGY IN FIGURES*. (2019). Eurostat.

Αυτό που παρατηρείται από το παραπάνω σχήμα είναι μία γενικότερη πτώση στην κατανάλωση πετρελαιοειδών σε σχέση με τις τιμές το 1990, εκτός από την κηροζίνη, η οποία έχει ανοδική πορεία το 2018.

Επιπροσθέτως, στο Σχήμα 2 φαίνεται ξεκάθαρα πως τα πετρελαιοειδή υπερτερούν κατά πολύ σε σύγκριση με άλλα καύσιμα.

**FINAL ENERGY CONSUMPTION – BY FUEL – EU-28 –
1990-2017 (Mtoe)**



Σχήμα 2 : Ενεργειακή Κατανάλωση ανά Καύσιμο

Πηγή : *EU ENERGY IN FIGURES*. (2019). Eurostat.

ΚΕΦΑΛΑΙΟ 2: ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Στο κεφάλαιο αυτό αναλύονται βασικοί ορισμοί και μεθοδολογίες της στατιστικής, στις οποίες βασίζεται όλο το ερευνητικό κομμάτι και πιο συγκεκριμένα η ανάλυση των δεδομένων στο Κεφάλαιο 4 της παρούσας διπλωματικής εργασίας.

Η **Στατιστική** είναι μία μεθοδική μαθηματική επιστήμη που επιχειρεί να εξάγει έγκυρη γνώση χρησιμοποιώντας δεδομένα παρατήρησης ή και πειράματος. Πρόκειται για σημαντική επιστήμη της οποίας οι εφαρμογές εκτείνονται σε πολλούς κλάδους της ανθρώπινης δραστηριότητας όπως στις επιχειρήσεις, καθώς και σε θετικές και συμπεριφορικές ή Κοινωνικές επιστήμες. Η **Περιγραφική Στατιστική** είναι κλάδος της στατιστικής με βασικό αντικείμενο μελέτης την διαδικασία συλλογής δεδομένων, την συνοπτική και αποτελεσματική παρουσίασή τους, καθώς και την ανάλυση και εξαγωγή συμπερασμάτων για τη λήψη ορθών αποφάσεων.

Οι στατιστικοί πίνακες κι οι γραφικές παραστάσεις αποτελούν βασικό κομμάτι της στατιστικής ανάλυσης διότι τα δεδομένα παρουσιάζονται με μεγαλύτερη σαφήνεια και απεικονίζονται με αρκετά κατανοητό τρόπο στον αναγνώστη.

2.1 Πίνακες

Η περιγραφική στατιστική δίνει τη δυνατότητα να απεικονιστεί η συμπεριφορά των δεδομένων με τη χρήση πινάκων. Είναι μία διαδικασία που ευνοεί κυρίως την παρουσίαση δεδομένων σε μεγάλο αριθμό. Επιτρέπει να παρατηρηθεί που συγκεντρώνονται τα δεδομένα σύμφωνα με διάφορα χαρακτηριστικά και μας βοηθά να διακρίνουμε την μεγαλύτερη και τη μικρότερη τιμή τους. Η διαδικασία που χρησιμοποιείται για τους παραπάνω λόγους είναι η **κατανομή συχνοτήτων**.

Η κατανομή συχνοτήτων συνοψίζει τα αριθμητικά δεδομένα σε διαστήματα τιμών. Τα διαστήματα αντιπροσωπεύουν ένα εύρος τιμών και ονομάζονται διάστημα τάξης (classes). Κάθε τιμή μπορεί να περιέχεται μόνο σε μία κατηγορία τάξεων. Αρχικά, για να δημιουργηθεί μία χρήσιμη κατανομή συχνοτήτων είναι απαραίτητο να προσδιοριστεί ο αριθμός των τάξεων (ο οποίος δεν χρειάζεται να είναι ούτε μικρός ούτε πολύ μεγάλος) που είναι κατάλληλος για τα υπάρχοντα δεδομένα. Στη συνέχεια, θα πρέπει να προσδιοριστεί το πλάτος (width) για κάθε διάστημα τάξης, αφαιρώντας την χαμηλότερη από την υψηλότερη τιμή και διαιρώντας με το πλήθος των τάξεων :

$$\text{Interval width} = \frac{\text{highest value} - \text{lowest value}}{\text{number of classes}}$$

Στόχος του πίνακα κατανομής συχνοτήτων είναι να απεικονίζει περιληπτικά και να δώσει επαρκείς πληροφορίες για τα δεδομένα που αναλύονται. Επιπλέον, σκοπός είναι η ενημέρωση του αναγνώστη με ευδιάκριτο τρόπο και όχι η σύγχυση με παρουσία πολύπλοκων διαγραμμάτων. Τα συμπεράσματα που μπορούν να αντληθούν από τον πίνακα κατανομής συχνοτήτων είναι αρκετά χρήσιμα και ποικίλουν.

Αρχικά, γίνεται εύκολα αντιληπτό στον ενδιαφερόμενο το σημείο που συγκεντρώνονται οι περισσότερες τιμές του χαρακτηριστικού που μελετάται και γενικότερα τον τρόπο με το οποίο αυτές διαμοιράζονται στα διάφορα διαστήματα τάξης που έχουν τεθεί. Επιπροσθέτως, αρκετά σημαντικό είναι πως με την βοήθεια του συγκεκριμένου πίνακα μπορούν να υπολογιστούν οι σχετικές συχνότητες των διαστημάτων τάξης ως ποσοστό % του συνόλου των παρατηρήσεων

για όλες τις τάξεις. Τέλος μπορεί να προσδιοριστεί η αθροιστική συχνότητα των τιμών υπολογίζοντας την διαδοχική άθροιση των απλών συχνοτήτων.

2.2 Διαγράμματα

Οργανώνοντας τα δεδομένα τις περισσότερες φορές παρατηρούνται σχέσεις ή ακόμα και μοτίβα μεταξύ τους. Για να εξερευνηθούν και να ανακαλυφθούν, με τον καλύτερο τρόπο τα συγκεκριμένα μοτίβα και τις σχέσεις που ενδέχεται να χαρακτηρίζουν τα δεδομένα, γίνεται προσπάθεια απεικόνισης τους δημιουργώντας διαγράμματα και γραφήματα.

Το **ιστόγραμμα** είναι ο πιο κοινός τρόπος για την παρουσίαση δεδομένων. Απεικονίζει αρκετά κατανοητά και ξεκάθαρα τις κατανομές συχνοτήτων. Για παράδειγμα, μπορεί πολύ εύκολα να καθοριστεί ποιες τιμές είναι κοινές και ποιες είναι πιο σπάνιες. Το ιστόγραμμα είναι ένα γράφημα που αναφέρεται σε ομαδοποιημένα αριθμητικά δεδομένα όπου χρησιμοποιούνται κάθετες στήλες ή ιστοί για την περιγραφή τους. Ο οριζόντιος άξονας Χ απεικονίζει τα διαστήματα τάξης της μεταβλητής για την οποία ενδιαφερόμαστε και ο κάθετος άξονας Υ αντιπροσωπεύει είτε τη συχνότητα, είτε το ποσοστό των παρατηρήσεων που ανήκουν στα διαστήματα τάξης. Το Σχήμα 3 δείχνει ένα ιστόγραμμα συχνοτήτων που περιγράφει το κόστος των γευμάτων σε εστιατόρια κάποιας πόλης. Φαίνεται πως το κόστος των γευμάτων συγκεντρώνεται κυρίως σε τιμές μεταξύ 40 \$ και 50 \$.

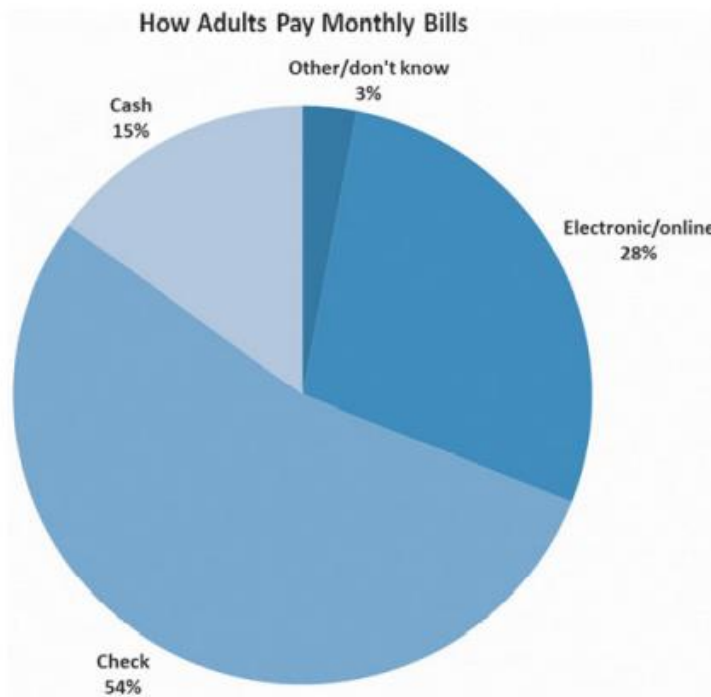


Σχήμα 3 : Ιστόγραμμα Κατανομής Συχνοτήτων

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ. 50

Ένα εξίσου σημαντικό διάγραμμα που περιγράφει με επιτυχία αριθμητικά δεδομένα είναι το **κυκλικό διάγραμμα** (pie chart). Το συγκεκριμένο γράφημα χρησιμοποιείται όταν είναι απαραίτητο να απεικονιστεί η κατανομή των παρατηρήσεων σε διαφορετικές κατηγορίες, με

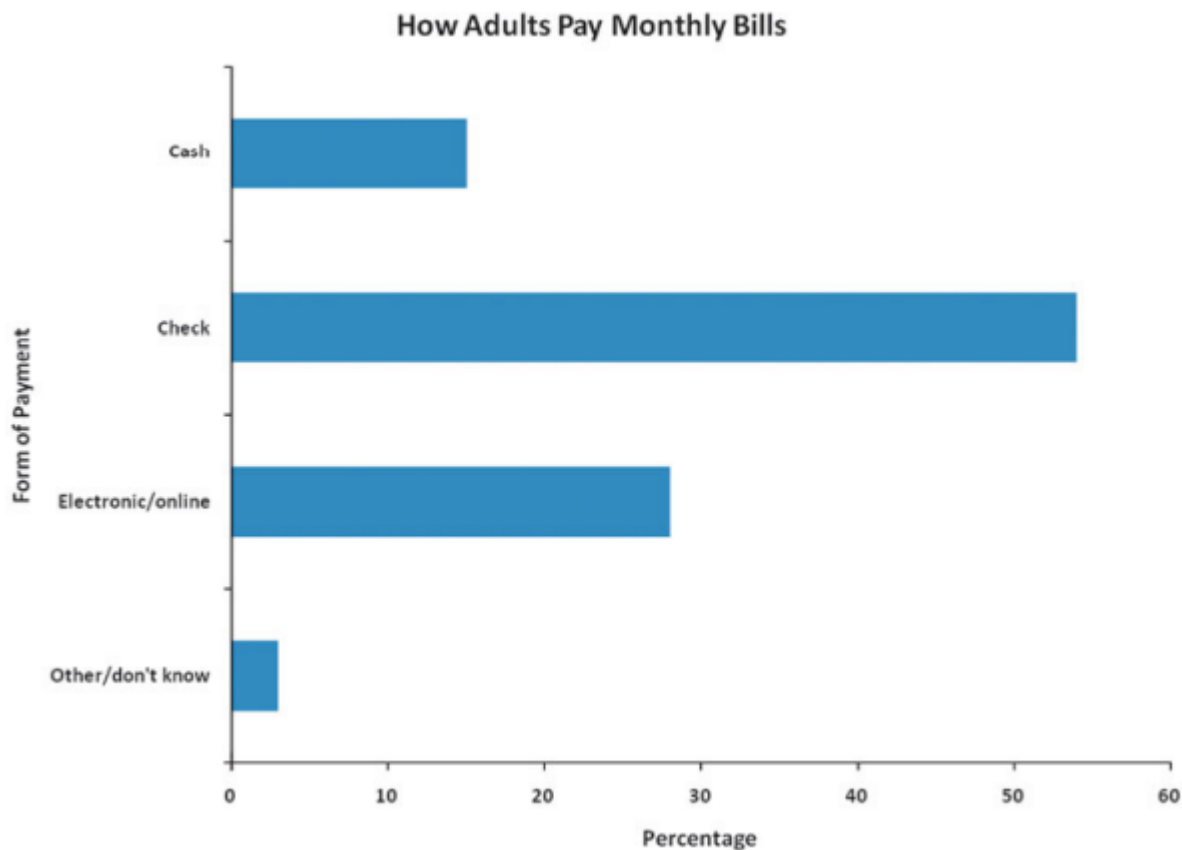
βάση ποιοτικά χαρακτηριστικό. Το μέγεθος κάθε «κομματιού» του κύκλου αντιπροσωπεύει το ποσοστό κάθε κατηγορίας. Για παράδειγμα παρακάτω στο Σχήμα 4 απεικονίζεται σε ένα κυκλικό διάγραμμα, τι ποσοστό των ανθρώπων πληρώνουν ηλεκτρονικά, μετρητά, με επιταγές κλπ. ανά μήνα.



Σχήμα 4 : Κυκλικό Διάγραμμα (pie chart)

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ.43

Τέλος, ένα αρκετά διαδεδομένο διάγραμμα και πολυχρησιμοποιημένο από τις οικονομικές επιστήμες είναι το **ακιδωτό** (bar chart). Είναι ένα γράφημα το οποίο συγκρίνει διαφορετικές κατηγορίες χρησιμοποιώντας μεμονωμένες ράβδους που αντιπροσωπεύουν τις παρατηρήσεις κάθε κατηγορίας. Σε αντίθεση με το ιστόγραμμα, οι ράβδοι στο συγκεκριμένο γράφημα διαχωρίζονται. Παρακάτω φαίνεται ο τρόπος με τον οποίο απεικονίζεται σε ακιδωτό διάγραμμα το προηγούμενο παράδειγμα.



Σχήμα 5 : Ακιδωτό Διάγραμμα (*bar chart*)

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ.42

Γενικότερα, η χρησιμότητα των γραφικών παραστάσεων και των διαγραμμάτων είναι μεγάλη στην ανάλυση δεδομένων. Με μία σύντομη ματιά μπορούν να αποκαλύψουν την συμπεριφορά των δεδομένων και να δώσουν πληροφορίες για το είδος της ανάλυσης που θα πραγματοποιηθεί παρακάτω.

2.3 Αριθμητικά Περιγραφικά Μέτρα

Στην προηγούμενη ενότητα παρουσιάστηκε η διαδικασία με την οποία μπορούν να παρουσιαστούν αριθμητικά δεδομένα έτσι ώστε να γίνουν κατανοητά και ευανάγνωστα. Τα διαγράμματα και οι γραφικές παραστάσεις παίζουν αρκετά σημαντικό ρόλο στην περιγραφική στατιστική, παρόλα αυτά δεν είναι αρκούν για να αποκαλυφθεί όλη η εικόνα των δεδομένων. Μια ολοκληρωμένη ανάλυση εμπεριέχει υπολογισμούς και εκτιμήσεις βασικών χαρακτηριστικών των δεδομένων όπως η μεταβλητότητα, η κεντρική τάση κλπ. Για τους παραπάνω λόγους ορίζονται κάποια ποσοτικά μεγέθη, τα οποία περιγράφουν τα χαρακτηριστικά μίας τυχαίας μεταβλητής x , και ονομάζονται περιγραφικά μέτρα.

Παρακάτω παρουσιάζονται τρεις τύποι περιγραφικών μέτρων :

- Η κεντρική τάση
- Η διασπορά ή μεταβλητότητα

- Το σχήμα

Τα περιγραφικά μέτρα που προκύπτουν από τα δεδομένα ενός πληθυσμού ονομάζονται **παράμετροι**, ενώ τα δεδομένα που αφορούν δείγματα ονομάζονται **εκτιμήσεις**.

2.3.1. Μέτρηση Κεντρικής Τάσης

Τα περισσότερα σύνολα δεδομένων τείνουν να συγκεντρώνονται γύρω από μια κεντρική τιμή. Με αυτό τον τρόπο, σε κάθε σύνολο δεδομένων επιλέγεται μία τιμή η οποία περιγράφει την συμπεριφορά των υπολοίπων. Τρεις είναι οι πιο κοινοί τρόποι μέτρησης της κεντρικής τάσης, ο αριθμητικός μέσος, η διάμεσος και η επικρατούσα τιμή.

Αριθμητικός μέσος (ή δειγματική μέση τιμή) είναι ο πιο κοινός τρόπος μέτρησης της κεντρικής τάσης. Ο αριθμητικός μέσος υπολογίζεται χρησιμοποιώντας όλες τις τιμές που υπάρχουν στο δείγμα, γι' αυτό το λόγο επηρεάζεται και από τις πιο ακραίες. Έστω x_1, x_2, \dots, x_n οι τιμές των παρατηρήσεων του δείγματος για μία τυχαία μεταβλητή X . Ο δειγματικός μέσος όρος συμβολίζεται με \bar{x} και προκύπτει από τον παρακάτω τύπο:

$$\bar{x} = \frac{x_1, x_2, \dots, x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Διάμεσος είναι η μεσαία τιμή ενός συνόλου δεδομένων που έχει ιεραρχηθεί με αύξουσα σειρά. Όταν ο αριθμός των παρατηρήσεων είναι περιττός, η τιμή της διαμέσου βρίσκεται στην θέση $\frac{n+1}{2}$ και όταν ο αριθμός είναι ζυγός τότε η τιμή της διαμέσου βρίσκεται στη θέση $\frac{n}{2}$ και $\frac{n}{2} + 1$. Αρκετά θετικό είναι πως η διάμεσος δεν επηρεάζεται από ακραίες τιμές.

Επικρατούσα τιμή (ή κορυφή) είναι η τιμή σε ένα σύνολο δεδομένων που εμφανίζεται με μεγαλύτερη συχνότητα. Όπως και η διάμεσος, η επικρατούσα τιμή δεν επηρεάζεται από ακραίες τιμές. Σε περίπτωση που υπάρχουν πάνω από μία τιμές με μεγάλη συχνότητα, τότε όλες θεωρούνται επικρατούσες.

2.3.2 Μέτρα Διασποράς ή Μεταβλητότητας

Τα μέτρα κεντρικής τάσης, όπως η μέση τιμή και η διασπορά, περιγράφουν ικανοποιητικά τα δεδομένα όταν οι παρατηρήσεις συγκεντρώνονται γύρω από την κεντρική τιμή. Σε περιπτώσεις όμως που οι παρατηρήσεις είναι αρκετά διασκορπισμένες, τα συγκεκριμένα μέτρα δεν μας καλύπτουν. Τα μέτρα μεταβλητότητας προσδιορίζουν τη διαφορά μεταξύ των παρατηρήσεων, καθώς και τη διασπορά τους στο χώρο. Οι συνηθέστεροι τρόποι μέτρησης της διασποράς είναι πέντε: το εύρος, η τεταρτημοριακή απόκλιση, η διακύμανση, η τυπική απόκλιση και ο συντελεστής μεταβλητότητας.

Το **εύρος** αποτελεί ένα από τα πιο εύκολα μέτρα διασποράς και υπολογίζεται αφαιρώντας την μικρότερη από την μεγαλύτερη τιμή των δεδομένων. Περιγράφεται από την παρακάτω σχέση:

$$Range = X_{largest} - X_{smallest}$$

Τα πλεονεκτήματα του συγκεκριμένου μέτρου διασποράς είναι ο απλός υπολογισμός του και η σαφή και ξεκάθαρη ερμηνεία του. Ο επηρεασμός του βέβαια, από ακραίες τιμές των δεδομένων δεν το καθιστά ιδιαίτερα αξιόπιστο.

Τεταρτημοριακή απόκλιση ή τεταρτημόρια ορίζεται η μέθοδος με την οποία μπορούμε να χωρίσουμε ένα σύνολο παρατηρήσεων σε τέσσερα ίσα μέρη. Το πρώτο τεταρτημόριο συνήθως ονομάζεται Q_1 (lower quartile) και αντιπροσωπεύει την τιμή κάτω από την οποία εμφανίζεται το 25% του συνόλου των παρατηρήσεων. Το δεύτερο τεταρτημόριο συμβολίζεται Q_2 (middle quartile), συμπίπτει με τη διάμεσο και είναι το 50% των παρατηρήσεων. Τέλος, το τρίτο τεταρτημόριο Q_3 (upper quartile) είναι η τιμή κάτω από την οποία εμφανίζεται το 75% των παρατηρήσεων (Lind κ.ά., 2018).

Το ενδοτεταρτημοριακό εύρος δεν αντιμετωπίζει τα μειονεκτήματα του εύρους, διότι δεν επηρεάζεται από μακρινές τιμές και υπολογίζεται από την αφαίρεση του πρώτου (Q_1) και του τρίτου (Q_3) τεταρτημόριου.

Διακύμανση και Τυπική Απόκλιση

Χρησιμοποιώντας σαν μέτρα διασποράς το εύρος και την τεταρτημοριακή απόκλιση, εξάγονται συμπεράσματα για τη συνολική και τη μεσαία διασπορά των παρατηρήσεων, χωρίς να ληφθεί υπόψη η συμπεριφορά των υπολοίπων τιμών του δείγματος. Τα πιο κοινά και σημαντικά μέτρα που περιγράφουν με μεγαλύτερη ακρίβεια την διασπορά όλων των τιμών του δείγματος γύρω από τον μέσο όρο, καθώς και το τρόπο με τον οποίο κατανέμονται στο χώρο, είναι η διακύμανση και η τυπική απόκλιση. Η διακύμανση έχει την ιδιότητα να λαμβάνει υπόψη της όλες τις παρατηρήσεις του δείγματος και ταυτόχρονα να υπολογίζει την συγκέντρωση των τιμών γύρω από τον μέσο όρο. Συμβολίζεται με s^2 και είναι ο μέσος όρος των τετραγωνικών αποκλίσεων των τιμών από τον αριθμητικό μέσο. Η διακύμανση περιγράφεται από τον παρακάτω τύπο:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Υπολογίζοντας την τετραγωνική ρίζα της διακύμανσης προκύπτει η τυπική απόκλιση. Παρακάτω βρίσκεται ο τύπος για την τυπική απόκλιση και συμβολίζεται με s :

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

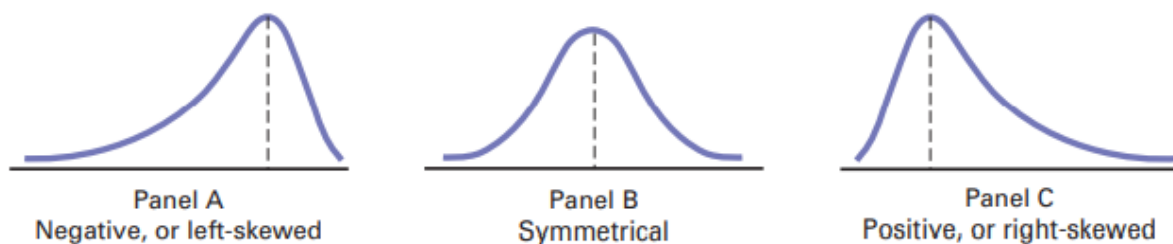
Το άθροισμα τετραγώνων, σύμφωνα με τους κανόνες της άλγεβρας είναι πάντα θετικό. Κατ' επέκταση η διακύμανση και η τυπική απόκλιση δεν μπορούν να πάρουν αρνητικές τιμές. Η μονή ακραία περίπτωση κατά την οποία τα δύο αυτά μέτρα παίρνουν την τιμή μηδέν είναι όταν οι παρατηρήσεις δεν έχουν καμία διασπορά και είναι όλες ίδιες μεταξύ τους. Η μονάδα μέτρησης που εκφράζεται η διασπορά και η τυπική απόκλιση είναι ίδια με αυτή της αρχικής μεταβλητής.

Συντελεστής μεταβλητότητας είναι ένα ιδιαίτερα χρήσιμο μέτρο στην περίπτωση που είναι απαραίτητο να συγκριθεί η διασπορά διαφορετικών ομάδων δεδομένων μεταξύ τους. Μετρά τη διασπορά των τιμών σε σχέση με τον μέσο όρο τους και εκφράζεται ως ποσοστό επί τις εκατό (%). Ορίζεται με τον παρακάτω τύπο:

$$CV = \left(\frac{S}{\bar{x}}\right) 100\%$$

2.3.3. Έλεγχος σχήματος

Το σχήμα είναι ένα περιγραφικό μέτρο το οποίο μας δίνει μια ξεκάθαρη εικόνα της κατανομής που ακολουθούν τα δεδομένα, σε όλο το εύρος των τιμών του δείγματος. Η κατανομή μπορεί να είναι **συμμετρική** ή **ασύμμετρη** (skewed). Σε μία συμμετρική κατανομή, οι τιμές κάτω από τον αριθμητικό μέσο κατανέμονται ακριβώς με τον ίδιο τρόπο με τις τιμές πάνω από τον αριθμητικό μέσο. Σε μία ασύμμετρη κατανομή υπάρχει ανισορροπία μεταξύ των υψηλών και των χαμηλών τιμών του δείγματος. Επιπλέον, το σχήμα της κατανομής μπορεί να επηρεάσει την σχέση μεταξύ του αριθμητικού μέσου με τη διάμεσο. Παρακάτω φαίνονται πιο ξεκάθαρα διάφορες κατανομές:



Σχήμα 6 : Σχήματα Κατανομής

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ. 108

Τα δεδομένα στο πρώτο σχήμα Panel A είναι φανερό πως έχουν αρνητική ασυμμετρία ή ασυμμετρία προς τα αριστερά. Αυτό σημαίνει πως η παραμόρφωση προς τα αριστερά προκαλείται από κάποιες εξαιρετικά μικρές τιμές. Στη συγκεκριμένη περίπτωση η μέση τιμή είναι μικρότερη από την διάμεσο. Στο δεύτερο σχήμα Panel B είναι αρκετά ξεκάθαρο πως υπάρχει συμμετρία ή μηδενική ασυμμετρία, κατά συνέπεια ο αριθμητικός μέσος ισούται με τη διάμεσο. Τέλος, στο τρίτο διάγραμμα παρατηρείται πως υπάρχει θετική ασυμμετρία ή ασυμμετρία προς τα δεξιά. Η ‘ουρά’ της κατανομής, προκαλείται από αρκετά μεγάλες τιμές του δείγματος. Συνεπώς, η τιμή του αριθμητικού μέσου είναι και μεγαλύτερη από την διάμεσο του δείγματος (Berenson κ.ά., 2011). Περιληπτικά :

- Mean < Median: negative, or left-skewed
- Mean = Median: symmetric, or zero skewness
- Mean > Median: positive, or right- skewed

Κάποιες φορές είναι αρκετά πιθανό πως αν δύο κατανομές έχουν την ίδια διακύμανση, τον ίδιο μέσο όρο και την ίδια ασυμμετρία, το σχήμα τους μπορεί να διαφέρει. Σε τέτοιου είδους περιπτώσεις πρέπει να μελετηθεί και ένα άλλο εξίσου σημαντικό μέτρο, η **κύρτωση** (kurtosis) των δεδομένων. Ουσιαστικά η κύρτωση περιγράφει το σχήμα της κορυφής της κατανομής που μελετάται. Μετρά τη σχετική συγκέντρωση των παρατηρήσεων στο κέντρο της κατανομής ενός συνόλου δεδομένων και το συγκρίνει με τις ‘ουρές’ της. Όταν η τιμή της κύρτωσης είναι αρκετά μεγάλη, τότε η κορυφή της καμπάνας της κατανομής θα είναι αρκετά υψηλή. Όταν η κατανομή έχει τη μορφή της καμπάνας, όπως φαίνεται και στα παραπάνω σχήματα, η τιμή της κύρτωσης είναι μηδενική. Αντιθέτως, όταν το σχήμα μίας κατανομής είναι πιο επίπεδο από την κανονική,

τότε η κύρτωση παίρνει αρνητικές τιμές και ονομάζεται πλατύκυρτη. Στην περίπτωση που υπάρχει θετική κύρτωση, σημαίνει πως μία κατανομή έχει πιο ψηλή κορυφή από αυτή της κανονικής και ονομάζεται λεπτόκυρτη.

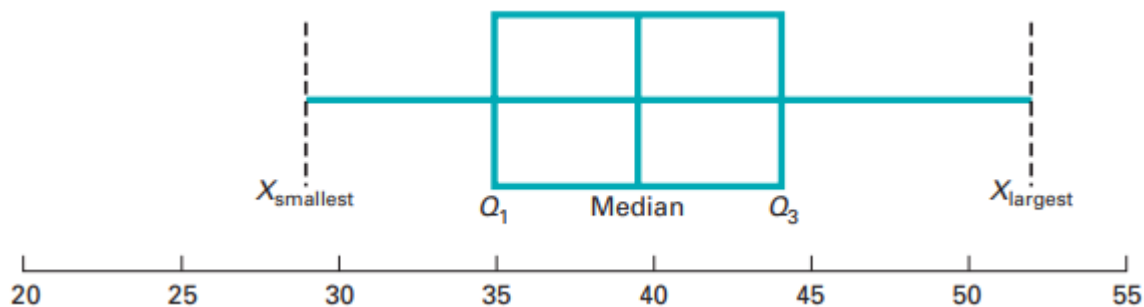
2.4 Διάγραμμα Πλαισίου-Απολήξεων (Boxplot)

Το διάγραμμα πλαισίου- απολήξεων ή θηκόγραμμα παρέχει γρήγορα και εύκολα μία γραφική απεικόνιση ενός συνόλου δεδομένων. Αποτελεί μία πολύ σημαντική στατιστική τεχνική, η οποία επιτρέπει να παρατηρηθούν τα δεδομένα και να εξαχθούν συμπεράσματα για τυχόν συσχετίσεις, τάσεις και ύπαρξη ακραίων τιμών.

Μέσω του διαγράμματος πλαισίου-απολήξεων βγαίνουν συμπεράσματα για πέντε βασικές μετρήσεις στην περιγραφική στατιστική και είναι οι παρακάτω:

- Η διάμεσος
- Το κατώτερο τεταρτημόριο Q_1
- Το ανώτερο τεταρτημόριο Q_3
- Η μικρότερη παρατήρηση
- Η μέγιστη παρατήρηση

Παρακάτω, στο Σχήμα 7 φαίνεται η μορφή που έχει ένα διάγραμμα πλαισίου-αποκλίσεων και αναλύεται ο τρόπος με τον οποίο αντλούνται πληροφορίες για τα δεδομένα.



Σχήμα 7 : Διάγραμμα Πλαισίου- Αποκλίσεων

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ.117

Στο θηκόγραμμα, οι κάθετη γραμμή στη μέση του πλαισίου είναι η διάμεσος. Οι κάθετες πλευρές του πλαισίου αντιπροσωπεύουν το κατώτερο και το ανώτερο τεταρτημόριο, Q_1 και Q_3 αντίστοιχα. Επιπλέον, οι οριζόντιες γραμμές που βρίσκονται δεξιά και αριστερά του πλαισίου ονομάζονται αποκλίσεις και ενώνουν το Q_1 με την μικρότερη παρατήρηση και το Q_3 με την μέγιστη παρατήρηση. Εάν κάποια παρατήρηση έχει απόσταση μεγαλύτερη από 1,5 φορές το τεταρτημοριακό εύρος τότε καλείται πιθανή έκτροπη παρατήρηση, ενώ αν βρίσκεται σε απόσταση 3 φορές μεγαλύτερη από το τεταρτημοριακό εύρος τότε σίγουρα είναι έκτροπη.

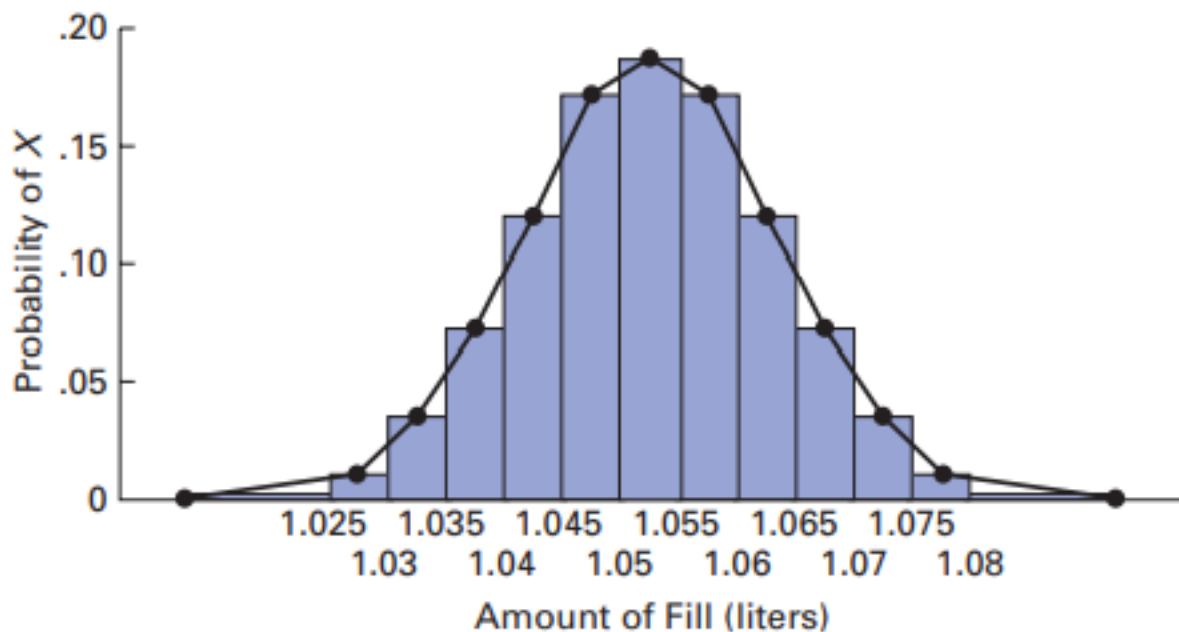
Τα συμπεράσματα που μπορούν να εξαχθούν για τα δεδομένα από ένα τέτοιο διάγραμμα ποικίλουν και είναι μείζονος σημασίας. Πρώτον, αναγνωρίζεται η θέση των δεδομένων μέσω της διαμέσου. Δεύτερον, εντοπίζεται η διασπορά των δεδομένων μέσω του τεταρτημοριακού εύρους και των αποκλίσεων. Τρίτον, εντοπίζεται τυχόν ασυμμετρία ή συμμετρία στα δεδομένα σύμφωνα με τη θέση της διαμέσου και το μήκος των αποκλίσεων. Τέλος, εντοπίζονται πιθανές ή πραγματικές έκτροπες παρατηρήσεις.

2.5 Κανονική κατανομή

Η κανονική κατανομή (γνωστή και ως Γκαουσιανή κατανομή) αναφέρεται σε συνεχή μεταβλητές και χρησιμοποιείται σαν μία πρώτη προσέγγιση για να περιγράψουν τυχαίες μεταβλητές πραγματικών τιμών, οι οποίες τείνουν να συγκεντρώνονται γύρω από τη μέση τιμή. Η κανονική κατανομή αποτελεί την πιο σημαντική κατανομή της στατιστικής μεθοδολογίας για τέσσερις βασικούς λόγους:

- Την κανονική κατανομή ακολουθούν είτε με ακρίβεια είτε με μεγάλη προσέγγιση τα περισσότερα συνεχή φαινόμενα.
- Πολλές ασυνεχείς κατανομές πιθανοτήτων, μπορούν να προσεγγιστούν μέσω της κανονικής κατανομής.
- Η κανονική κατανομή, αποτελεί σύμφωνα με το κεντρικό οριακό θεώρημα τη βάση της επαγωγικής στατιστικής.
- Τυχαία σφάλματα που εμφανίζονται σε διάφορες μετρήσεις έχουν κανονική κατανομή.

Η γραφική παράσταση της κανονικής κατανομής έχει σχήμα ‘καμπάνας’ και για τον λόγο αυτό είναι συμμετρική. Σαν αποτέλεσμα, αριθμητικός μέσος, η επικρατούσα τιμή και η διάμεσος συμπίπτουν. Στο Σχήμα 8 βλέπουμε μια χαρακτηριστική κανονική κατανομή:



Σχήμα 8 : Η Κανονική Κατανομή

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ. 219

Στις συνεχείς μεταβλητές, γίνεται προσπάθεια να προσεγγιστεί η μαθηματική συνάρτηση της κατανομής $f(x)$. Η συνάρτηση μίας συνεχούς μεταβλητής ονομάζεται συνάρτηση πυκνότητας πιθανότητας και εκφράζει την πυκνότητα (συχνότητα) εμφάνισης των παρατηρήσεων στα διάφορα διαστήματα τιμών. Η μαθηματική εξίσωση που την περιγράφει είναι η παρακάτω:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < X < +\infty$$

Από την σχέση που φαίνεται παραπάνω συμπεραίνουμε πως οι πιθανότητες της τυχαίας μεταβλητής X εξαρτώνται μόνο από τις δύο παραμέτρους του πληθυσμού μ και σ .

2.6 Στατιστικός Έλεγχος Υποθέσεων

Υπάρχουν διάφορες μέθοδοι που μας βοηθούν να βγάλουμε συμπεράσματα για τις παραμέτρους ενός πληθυσμού. Ο έλεγχος υποθέσεων είναι μία τέτοια μέθοδος στην στατιστική και ξεκινά με μία θεωρία ή έναν ισχυρισμό, συνήθως για τον αριθμητικό μέσο. Ουσιαστικά, η ερώτηση για την οποία πραγματοποιείται ο έλεγχος, είναι εάν η τιμή της παραμέτρου ισούται με έναν συγκεκριμένο αριθμό.

2.6.1 Μηδενική και Εναλλακτική Υπόθεση

Στον στατιστικό έλεγχο υπάρχει συνήθως μια προκαθορισμένη ιδέα για την τιμή της παραμέτρου. Εκτός εάν αποκτηθούν αρκετές αποδείξεις για το αντίθετο, η αρχική ιδέα ισχύει πάντα, ονομάζεται μηδενική υπόθεση (null hypothesis) και συμβολίζεται με H_0 . Κάθε φορά που καθορίζεται η μηδενική υπόθεση, ταυτόχρονα υπάρχει η εναλλακτική (alternative hypothesis) και συμβολίζεται με H_1 . Αυτό συμβαίνει διότι όταν η H_0 αποδειχθεί λάθος, γίνεται αποδοχή της H_1 . Για παράδειγμα, μπορούμε να ισχυριστούμε ότι ο αριθμητικός μέσος ενός πληθυσμού ισούται με 50 και συμβολίζεται :

$$H_0: \mu = 50$$

Στην περίπτωση που συγκεντρωθούν συμπεράσματα τα οποία αποδείξουν το αντίθετο, μόνο τότε ο ερευνητής αποδέχεται την εναλλακτική υπόθεση και απορρίπτει την μηδενική:

$$H_1: \mu \neq 50$$

Ανάμεσα στις δύο υποθέσεις που αναλύθηκαν παραπάνω μόνο μία μπορεί να είναι αξιόπιστη. Αυτό συμβαίνει διότι ισχυρίζονται ακριβώς δύο αντίθετα πράγματα, ως αποτέλεσμα η αποδοχή για παράδειγμα της μηδενικής υπόθεσης, σημαίνει αυτόματα απόρριψη της εναλλακτικής. Τέλος, στατιστικοί έλεγχοι είναι δυνατό να πραγματοποιηθούν και σε άλλες παραμέτρους του πληθυσμού όπως για παράδειγμα η διασπορά. Επιπλέον, οι δύο αυτές υποθέσεις μπορούν να μελετηθούν και με ανισότητες, τονίζοντας βέβαια σε αυτό το σημείο, πως η ισότητα πρέπει πάντα να εμφανίζεται στη μηδενική υπόθεση.

2.6.2 Σφάλματα και Επίπεδο σημαντικότητας

Όπως προαναφέρθηκε, στον στατιστικό έλεγχο υποθέσεων μπορούμε να αποδεχθούμε ή να απορρίψουμε είτε την H_0 είτε την H_1 με τις αποδείξεις που υπάρχουν. Είναι όμως οι αποδείξεις αυτές αρκετές για να φτάσει κάποιος στο σωστό συμπέρασμα; Η αλήθεια είναι πως δεν μπορεί κανείς να είναι 100% σίγουρος για το αποτέλεσμα και για αυτόν το λόγο υπάρχει η πιθανότητα να εμφανιστούν δύο τύποι σφαλμάτων.

Το σφάλμα **Τύπου I** λαμβάνει χώρα όταν απορριφθεί η H_0 , στην περίπτωση όμως που είναι αληθής.

Το σφάλμα **Τύπου II** συμβαίνει όταν δεν απορριφθεί την H_0 , στην περίπτωση που είναι λανθασμένη και θα έπρεπε να απορριφθεί.

Για να ελαχιστοποιηθούν τα σφάλματα χρησιμοποιούνται διάφορες μέθοδοι που βοηθούν να φτάσουμε στο συμπέρασμα, να απορρίψουμε ή να δεχθούμε μία υπόθεση. Το **p – value** λοιπόν είναι μία τιμή η οποία θα μας καθοδηγήσει σωστά ακόμα και αν οι αποδείξεις ευνοούν την λάθος κατάληξη.

Το **p – value** ορίζεται ως η πιθανότητα να προκύψουν αποδείξεις από ένα δείγμα μεγέθους n που δεν ευνοούν τον ισχυρισμό της H_0 , ακόμα και αν η μηδενική υπόθεση είναι αληθής. Η παραπάνω πιθανότητα δέχεται τιμές μεταξύ 0 και 1. Στην περίπτωση που η τιμή του **p – value** ισούται με 0 τότε, η μηδενική υπόθεση είναι σίγουρα λανθασμένη, αντίθετα όταν παίρνει την τιμή 1 τότε η H_0 είναι αληθής.

Πρώτου ο ερευνητής προχωρήσει στον έλεγχο υποθέσεων, είναι απαραίτητο να ορισθεί ένα επίπεδο εμπιστοσύνης α , όπου οι σύνηθες τιμές του είναι 5%, 10% και 1%. Προκύπτει λοιπόν πως αν το **p – value** είναι μικρότερο από το επίπεδο σημαντικότητας που έχει οριστεί τότε απορρίπτεται την H_0 . Αντίθετα όταν το **p – value** είναι μεγαλύτερο από το επίπεδο σημαντικότητας α , γίνεται αποδοχή της H_0 .

2.7 Ανάλυση Παλινδρόμησης

Η παλινδρόμηση είναι ένα αρκετά σημαντικό εργαλείο στην στατιστική το οποίο χρησιμοποιείται ευρέως για να εντοπιστεί κάποια ενδεχόμενη συσχέτιση μεταξύ των μεταβλητών. Οι εφαρμογές της ποικίλουν και χρησιμοποιούνται από πολλά επιστημονικά πεδία όπως είναι η φυσική, η χημεία, τα οικονομικά κλπ. Οι μεταβλητές οι οποίες αποτελούν τέτοιου είδους τεχνικές μοντελοποίησης χωρίζονται σε δύο κατηγορίες :

- Την **ανεξάρτητη μεταβλητή (x)**, όπου υπάρχει δυνατότητα ελέγχου των τιμών της και ο καθορισμός αυτών
- Την **εξαρτημένη μεταβλητή (y)**, δηλαδή, η μεταβλητή που γίνεται προσπάθεια πρόβλεψης των τιμών της και στην οποία είναι εμφανή οποιαδήποτε μεταβολή συμβαίνει στις ανεξάρτητες μεταβλητές.

Το πρώτο στάδιο είναι να ανακαλυφθεί η στατιστική σχέση που ενδεχομένως να έχουν οι μεταβλητές που διατίθενται. Όταν η πρωταρχική ανάλυση οδηγήσει στο συμπέρασμα πως υπάρχει συσχέτιση μεταξύ των μεταβλητών, τότε το επόμενο στάδιο είναι η ανάλυση παλινδρόμησης.

2.7.1 Συντελεστής Συσχέτισης και Διαγράμματα Διασποράς

Η συσχέτιση αναφέρεται σε μία ευρεία κατηγορία στατιστικών σχέσεων, αν και από το ευρύ κοινό, αναφέρεται στον βαθμό με τον οποίο δύο μεταβλητές έχουν γραμμική σχέση η μία με την άλλη. Ουσιαστικά, αποτελεί μία ένδειξη για το πόσο καλά κινούνται μαζί οι δύο μεταβλητές σε μία ευθεία γραμμή. Το πιο γνωστό μέτρο της εξάρτησης μεταξύ δύο ποσοτήτων είναι ο συντελεστής συσχέτισης συνδιακύμανσης Pearson, συμβολίζεται με r και περιγράφεται από τον παρακάτω τύπο:

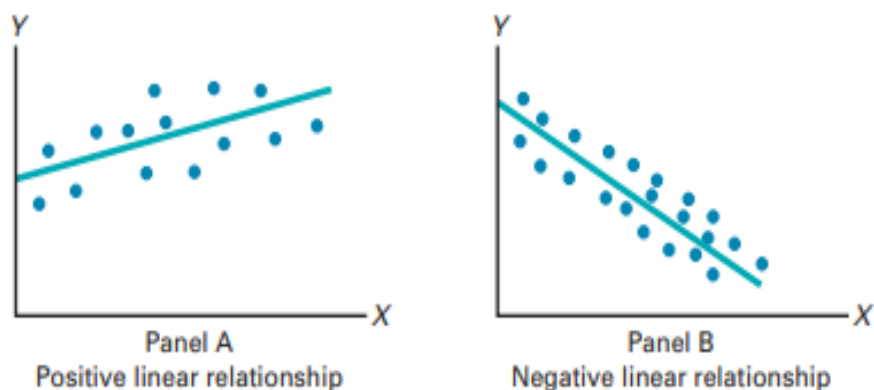
$$r_{x,y} = cor(X, Y) = \frac{cov(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$

Σε αυτό το σημείο, είναι σημαντικό να αναφερθεί, πως όταν χρησιμοποιείται ο συντελεστής συνδιακύμανσης Pearson, για την ένδειξη συσχέτισης μεταξύ δύο μεταβλητών, υπάρχει η υπόθεση πως η ανεξάρτητη και η εξαρτημένη μεταβλητή ακολουθούν κανονική κατανομή με αριθμητικούς μέσους μ_x και μ_y και τυπικές αποκλίσεις σ_x και σ_y . Ο συντελεστής συσχέτισης, λαμβάνει τιμές από -1 έως $+1$. Παρακάτω αναλύεται, πως ερμηνεύεται το r σύμφωνα με τις τιμές που παίρνει κάθε φορά:

- Όταν το $r = 0$ τότε δεν υπάρχει καμία συσχέτιση μεταξύ των δύο μεταβλητών και κατά συνέπεια ούτε γραμμική σχέση.
- Όταν το $r = 1$ συμπεραίνουμε πως υπάρχει η τέλεια συσχέτιση μεταξύ της ανεξάρτητης και εξαρτημένης μεταβλητής. Για παράδειγμα, αν η X αυξηθεί, το ίδιο ακριβώς θα συμβεί στην Y . Τέτοιες περιπτώσεις είναι εξαιρετικά σπάνιο να συμβούν με πραγματικά δεδομένα.
- Όταν το $r = -1$ τότε υπάρχει τέλεια αρνητική γραμμική σχέση. Όπου όταν αυξάνεται μία εκ των δύο μεταβλητών, η άλλη μειώνεται και αντίστοιχα.
- Γενικότερα, όταν οι τιμή του r κυμαίνεται μεταξύ $[-1, 1]$ μπορούμε να βγάλουμε αρκετά συμπεράσματα για την σχέση που έχουν δύο μεταβλητές. Για παράδειγμα αν η τιμή του r είναι ίση με 85, δηλαδή αρκετά κοντά στη μονάδα, φαίνεται να υπάρχει αρκετά ισχυρή γραμμική σχέση μεταξύ των δύο αυτών μεταβλητών.

Ένα άλλο θετικό χαρακτηριστικό που έχει ο συντελεστής συσχέτισης, είναι ότι δεν εκφράζεται σε κάποια μονάδα μέτρησης, κάτι το οποίο επιτρέπει, χωρίς περιορισμούς, να συγκριθούν ζεύγη μεταβλητών μεταξύ τους τα οποία είναι τελείως διαφορετικά.

Ένας εξίσου σημαντικός τρόπος που μας βοηθά να ανακαλύψουμε τις σχέσεις μεταξύ δύο μεταβλητών είναι το διάγραμμα διασποράς (scatter plot). Στο διάγραμμα αυτό, στον οριζόντιο άξονα x' βρίσκονται οι τιμές της ανεξάρτητης μεταβλητής, ενώ στον κάθετο άξονα $y'y$ τοποθετούνται οι τιμές της εξαρτημένης. Στο Σχήμα 9 παρατίθενται κάποια διαγράμματα διασποράς:

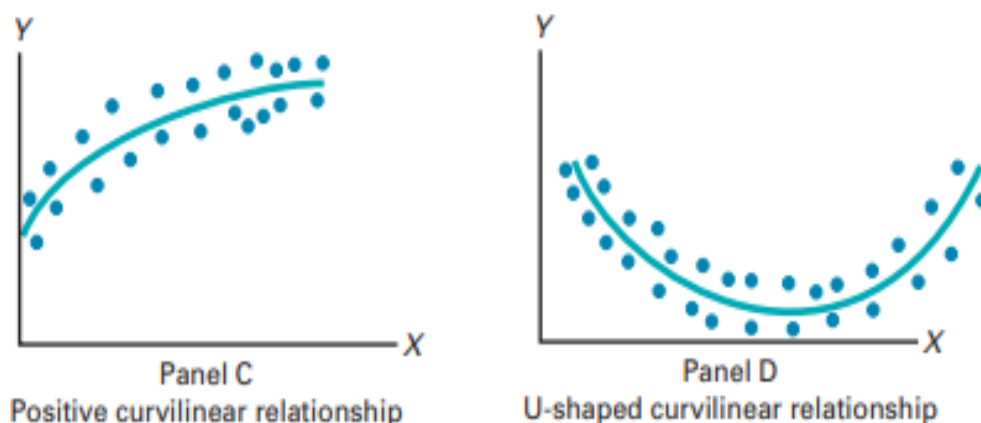


Σχήμα 9 : Διαγράμματα Διασποράς

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ. 523

Όταν τα σημεία συγκεντρώνονται με τέτοιο τρόπο έτσι ώστε να σχηματίζεται μία ευθεία γραμμή, φαίνεται πως υπάρχει αρκετά δυνατή σχέση μεταξύ των μεταβλητών που μελετώνται.

Τέλος, είναι απαραίτητο ο ερευνητής να είναι ιδιαίτερα προσεκτικός με την ερμηνεία του συντελεστή συσχέτισης. Σε διάφορες περιπτώσεις, μπορεί οι μεταβλητές να έχουν μεγάλη συσχέτιση μεταξύ τους, αλλά η σχέση τους να μην είναι γραμμική, αλλά καμπυλόγραμμη όπως φαίνεται στο Σχήμα 10. Για τον παραπάνω λόγο, είναι ξεκάθαρη η χρησιμότητα του διαγράμματος διασποράς, το οποίο εντοπίζει έγκαιρα τέτοιου είδους σχέσεις, απεικονίζοντας απλά τα σημεία σε ένα γράφημα.

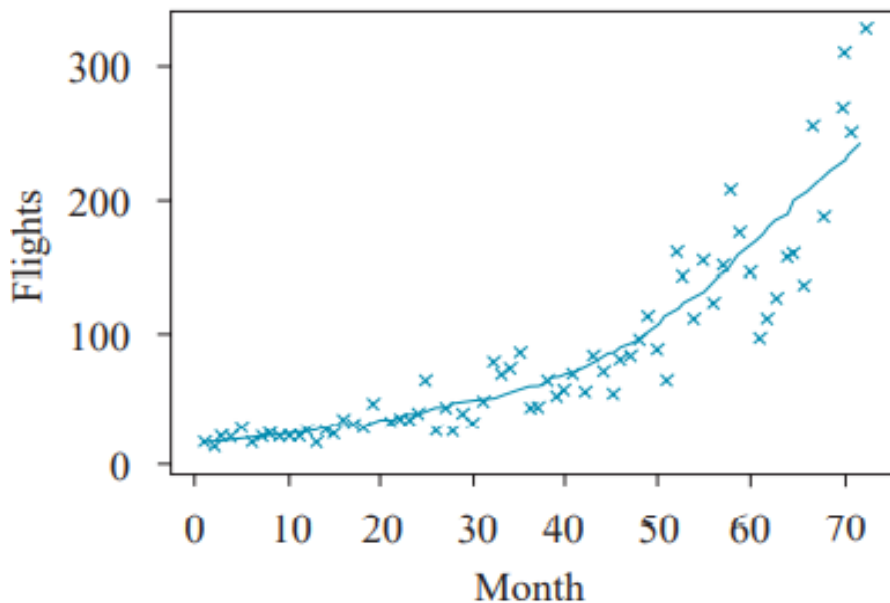


Σχήμα 10 : Διαγράμματα Διασποράς, Καμπυλόγραμμη Σχέση

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ.. 523

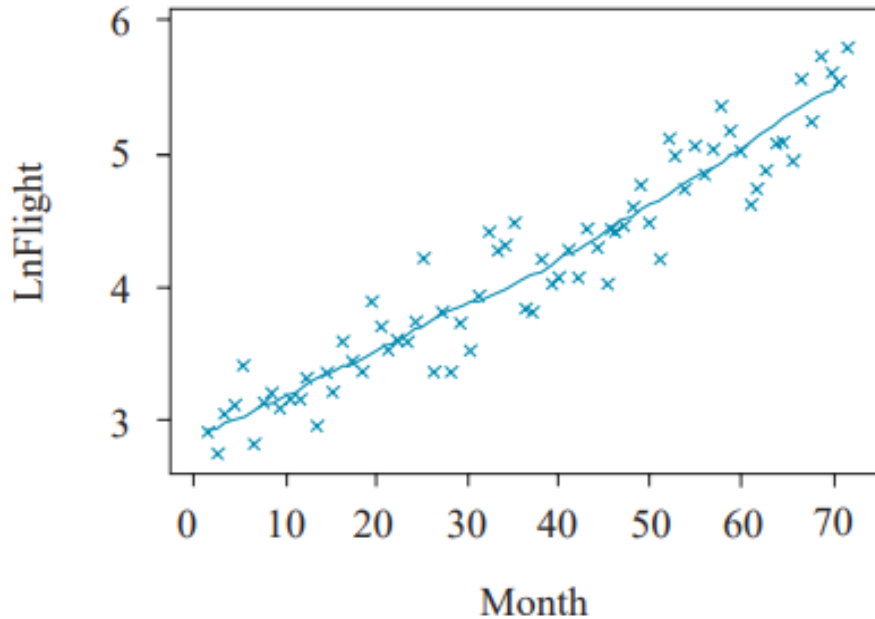
Κάποιες φορές πολλές οικονομικές σχέσεις δεν είναι γραμμικές. Κατά συνέπεια το διάγραμμα διασποράς δεν εμφανίζει την νοητή ευθεία γραμμή γύρω από την οποία συγκεντρώνονται οι

παρατηρήσεις. Στην περίπτωση αυτή, ο ερευνητής μπορεί να επέμβει και να πραγματοποιήσει μετασχηματισμούς, είτε στην ανεξάρτητη μεταβλητή (X), είτε στην εξαρτημένη (Y). Υπάρχει δυνατότητα να χρησιμοποιηθούν αρκετοί μετασχηματισμοί, κάποιοι από τους πιο γνωστούς είναι η τετραγωνική ρίζα (x^2), ο λογάριθμος (\ln) και ο αντίστροφος. Για να βρεθεί ο κατάλληλος μετασχηματισμός, σε κάθε περίπτωση, χρειάζονται αρκετές προσπάθειες. Παρακάτω εμφανίζονται δύο διαγράμματα διασποράς. Στο Σχήμα 11 απεικονίζονται τα δεδομένα πριν τον μετασχηματισμό και στο Σχήμα 12 μετά τον μετασχηματισμό της εξαρτημένης μεταβλητής, με χρήση λογαρίθμου.



Σχήμα 11 : Διάγραμμα Διασποράς

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition



Σχήμα 12 : Διάγραμμα Διασποράς, Μετασχηματισμός με χρήση Αλγορίθμου

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition

Τέλος, είναι απαραίτητο, προτού προβεί κάποιος σε μετασχηματισμό να απαντηθούν δύο ερωτήματα:

1. Υπάρχει μη γραμμική σχέση μεταξύ των δύο μεταβλητών ;
2. Υπάρχει κάποιο μοτίβο που να αποδεικνύει την αύξηση της μεταβλητότητας κατά μήκος του κάθετου άξονα $y'y$;

Εάν σε κάποιο από τα παραπάνω ερωτήματα η απάντηση είναι θετική, τότε η υπόθεση ότι η διακύμανση είναι σταθερή αμφισβητείται και γίνονται οι μετασχηματισμοί που προαναφέρθηκαν.

2.7.2 Απλή Γραμμική Παλινδρόμηση

Η γραμμική παλινδρόμηση είναι μια προσέγγιση μοντελοποίησης της σχέσης μιας εξαρτημένης μεταβλητής Y , με μία ή περισσότερες ανεξάρτητες μεταβλητές X . Οι μεταβλητή X_i (το i εκφράζει τις παρατηρήσεις του δείγματος) δεν θεωρείται τυχαία, ενώ η μεταβλητή Y_i θεωρείται τυχαία μεταβλητή. Στην περίπτωση που έχουμε μόνο μία ανεξάρτητη μεταβλητή X και η σχέση μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής είναι γραμμική, τότε η μοντελοποίηση ονομάζεται απλή γραμμική παλινδρόμηση (simple linear regression). Στην απλή γραμμική παλινδρόμηση υπάρχει ένα σύνολο δειγμάτων τιμών $\{X, Y_i\}$ και στόχος αποτελεί, να βρεθεί το απλούστερο μαθηματικό μοντέλο που θα περιγράψει καλύτερα την σχέση που υπάρχει μεταξύ των δύο μεταβλητών. Το μοντέλο παλινδρόμησης που αναζητάτε από τους ενδιαφερόμενους είναι μία ευθεία γραμμή. Έχοντας το υπόδειγμα αυτό, υπάρχει δυνατότητα πρόβλεψης των τιμών του Y για νέες τιμές του X . Παρακάτω φαίνεται η μορφή αυτής της ευθείας:

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

Όπου: Y_i = η τιμή της εξαρτημένης μεταβλητής (dependent variable)

X_i = η τιμή της ανεξάρτητης μεταβλητής (independent variable)

β_0 = το σημείο τομής του άξονα της Y από ευθεία της παλινδρόμησης
(intercept)

β_1 = η κλίση της ευθείας της παλινδρόμησης (slope)

ε_i = σφάλματα ή κατάλοιπα , δηλαδή η διαφορά μεταξύ της
πραγματικής τιμής της Y και της προβλεπόμενης τιμής που
προκύπτει από το υπόδειγμα (residuals)

Για να είναι αξιόπιστο και αληθές το υπόδειγμα της γραμμικής παλινδρόμησης υπάρχουν κάποιες προϋποθέσεις που είναι απαραίτητο να τηρούνται. Αρχικά, η σχέση μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής πρέπει να είναι γραμμική. Δεύτερον, οι τιμές της ανεξάρτητης μεταβλητής θεωρούνται γνωστές και είναι τυχαίες. Αν υπάρχει κάποιος παράγοντας ο οποίος επηρεάζει την X τότε αυτός προέρχεται από τα σφάλματα ε . Τέλος, η πιο σημαντική προϋπόθεση για την γραμμική παλίνδρομη είναι πως τα σφάλματα πρέπει να κατανέμονται κανονικά με αριθμητικό μέσο ίσο με 0 , $E[\varepsilon_i] = 0$, και σταθερή διακύμανση $Var[\varepsilon_i] = \sigma^2$ για όλες τις παρατηρήσεις i . Επιπλέον, τα σφάλματα δεν πρέπει να συσχετίζονται μεταξύ τους.

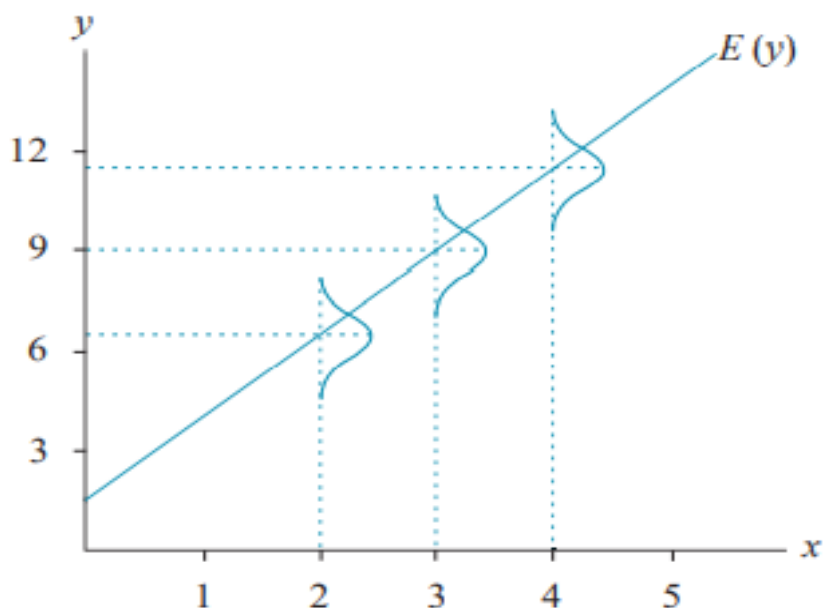
Δύο συνέπειες των παραπάνω υποθέσεων είναι ότι κάθε παρατήρηση Y_i έχει μέση τιμή:

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

Και διακύμανση:

$$Var[X_i] = \sigma^2$$

Στο Σχήμα 13 απεικονίζονται και διαγραμματικά οι παραπάνω υποθέσεις:



Σχήμα 13: Διαγραμματική Απεικόνιση των Υποθέσεων της Γραμμικής Παλινδρόμησης

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition

Όπως φαίνεται και στο σχήμα οι πραγματικές τιμές της εξαρτημένης μεταβλητής κατανέμονται κανονικά, με τις μέσες τιμές να πέφτουν πάνω στην ευθεία της παλινδρόμησης. Επιπλέον, η τυπική απόκλιση είναι η ίδια για όλες τις τιμές της εξαρτημένης μεταβλητής.

Η ευθεία γραμμή της παλινδρόμησης προσδιορίζεται από δύο παραμέτρους, τις β_0 και β_1 και ονομάζονται συντελεστές παλινδρόμησης. Η β_0 δείχνει το σημείο που τέμνει η ευθεία τον άξονα της Y και ονομάζεται τεταγμένη στην αρχή μηδέν, και ο συντελεστής β_1 δείχνει την κλίση της ευθείας και είναι η μέση μεταβολή της εξαρτημένης μεταβλητής Y που αντιστοιχεί σε μεταβολή της ανεξάρτητης X κατά μία μονάδα.

2.7.3 Μέθοδος Ελαχίστων Τετραγώνων

Η μέθοδος αυτή, χρησιμοποιείται για την εκτίμηση των παραμέτρων του μοντέλου παλινδρόμησης, δηλαδή οι συντελεστές β_0 και β_1 . Πρωταρχικός στόχος είναι η ευθεία που θα προκύψει, να περιγράψει με τον καλύτερο δυνατό τρόπο, τη σχέση που υπάρχει μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής. Η γραμμή αυτή είναι σημαντικό να περνάει όσο το δυνατόν πιο κοντά από τα ζεύγη των παρατηρήσεων (x_i, y_i) , έτσι ώστε να ελαχιστοποιούνται τα σφάλματα πρόβλεψης. Η ευθεία συχνά, αναφέρεται και σαν **γραμμή πρόβλεψης** και όταν υπολογιστούν οι εκτιμήσεις αυτές, θα υπάρχει δυνατότητα πρόβλεψης των τιμών της Y με την παρακάτω εξίσωση παλινδρόμησης:

$$\hat{y} = b_0 + b_1 X_1$$

Αυτό που έχει επικρατήσει στην μαθηματική κοινότητα είναι οι τιμές των παραμέτρων του πληθυσμού να συμβολίζονται με μικρά ελληνικά γράμματα και οι εκτιμήσεις των παραμέτρων αυτών με λατινικούς χαρακτήρες.

Τα σφάλματα ή κατάλοιπα συμβολίζονται με e και ισούνται με τη διαφορά των παρατηρούμενων τιμών της Y και των προσαρμοσμένων τιμών της \hat{Y} :

$$e = Y_i - \hat{Y}_i$$

Ουσιαστικά, αυτό που αναζητείται, είναι εκείνες τις τιμές των b_0 και b_1 που θα ελαχιστοποιούν τις αποκλίσεις (σφάλματα ή κατάλοιπα) e . Τα σφάλματα μπορούν να έχουν θετικό και αρνητικό πρόσημο, γι' αυτό το λόγο γίνεται προσπάθεια να ελαχιστοποιηθούν τα τετράγωνά τους και πιο συγκεκριμένα τα αθροίσματα αυτών. Ως εκ τούτου η μέθοδος αυτή ονομάζεται, **μέθοδος ελαχίστων τετραγώνων** και εκφράζεται από την παρακάτω σχέση:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n [Y_i - (b_0 + b_1 X_i)]^2$$

Λύνοντας το παραπάνω σύστημα προκύπτουν οι παρακάτω σχέσεις για τις παραμέτρους:

$$b_1 = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Όπου:

$$\bar{Y} = \frac{\sum_{i=1}^n y_i}{n},$$

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

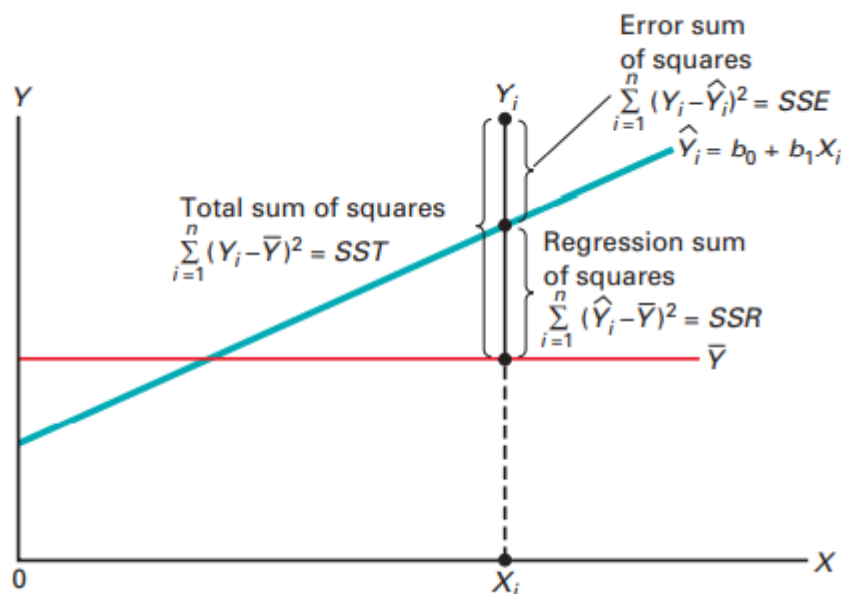
Οι εκτιμητές b_0 και b_1 των ελαχίστων τετραγώνων έχουν τις εξής ιδιότητες :

- Το άθροισμα των καταλοίπων e_i γύρω από τη γραμμή της παλινδρόμησης ισούται με μηδέν.
- Η γραμμή παλινδρόμησης περνάει από το σημείο (\bar{X}, \bar{Y}) που αντιστοιχεί στους μέσους των μεταβλητών X και Y .

2.7.4 Η μεταβλητότητα του Μοντέλου

Το πιο σημαντικά θέματα που είναι απαραίτητο να ληφθούν υπόψη προτού χρησιμοποιηθεί η εξίσωση παλινδρόμησης, αφορούν την προβλεπτική ικανότητα της εξίσωσης και το ποσοστό των μεταβολών της εξαρτημένης Y που οφείλεται στις επιδράσεις της ανεξάρτητης X . Η απάντηση στα παραπάνω ερωτήματα θα καθορίσει κατά πόσο τα αποτελέσματα μίας πρόβλεψης είναι έγκυρα, με βάση την εξίσωση της παλινδρόμησης. Στην ενότητα που ακολουθεί αναλύονται τα ερωτήματα αυτά, καθώς και οι απαντήσεις τους.

Χρησιμοποιώντας τη μέθοδο ελαχίστων τετραγώνων για τον προσδιορισμό των συντελεστών της παλινδρόμησης για ένα σύνολο δεδομένων, είναι απαραίτητο να προσδιοριστούν τα μέτρα μεταβλητότητας. Το πρώτο μέτρο που χρησιμοποιείται είναι το **συνολικό άθροισμα τετραγώνων (Total Sum of Squares, SST)**, το οποίο μετρά τη διασπορά των τιμών Y_i και ορίζεται από το άθροισμα των τετραγώνων των αποκλίσεων των τιμών από τον μέσο όρο τους, \bar{Y} . Η συνολική μεταβλητότητα ή το συνολικό άθροισμα των τετραγώνων της Y χωρίζεται σε δύο συνιστώσες, στην «εξηγημένη» από την εξίσωση της παλινδρόμησης και την «ανεξήγητη». Η εξηγημένη μεταβλητή ή **άθροισμα των τετραγώνων των τιμών της παλινδρόμησης (Sum of Squares Regression, SSR)** αντιπροσωπεύει τη μεταβλητότητα που εξηγείται από τη σχέση μεταξύ της X και της Y και η ανεξήγητη μεταβλητή ή **άθροισμα των τετραγώνων των σφαλμάτων (Sum of Squares Error, SSE)** που οφείλεται στην επίδραση άγνωστων παραγόντων, εκτός της X . Παρακάτω απεικονίζονται γραφικά τα παραπάνω μέτρα διασποράς:



Σχήμα 14 : Μέτρα Μεταβλητότητας Μοντέλου

Πηγή: Berenson M.L., Levine D.M., Krehbiel T.C, 2002 “Basic Business Statistics: Concepts and Business Applications” Twelfth Edition, σελ. 533

ι. Υπολογισμός των αθροισμάτων των τετραγώνων

Το άθροισμα των τετραγώνων της παλινδρόμησης (SSR) βασίζεται στη διαφορά μεταξύ των τιμών \hat{Y}_i και του μέσου όρου \bar{Y} . Το άθροισμα των τετραγώνων των σφαλμάτων αντιπροσωπεύει ένα μέρος της διακύμανσης της Y που δεν εξηγείται από την παλινδρόμηση. Παρακάτω παρουσιάζονται οι σχέσεις που περιγράφουν τα μέτρα διασποράς:

Το συνολικό άθροισμα των τετραγώνων (SST) ισούται με το άθροισμα των τετραγώνων των τιμών της παλινδρόμησης (SSR) συν το άθροισμα των τετραγώνων των σφαλμάτων (SSE),

$$SST = SSR + SSE$$

Επιπλέον, το συνολικό άθροισμα των τετραγώνων (SST) ισούται με το άθροισμα των τετραγώνων των αποκλίσεων των τιμών από το μέσο όρο τους,

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Το άθροισμα των τετραγώνων των τιμών της παλινδρόμησης (SSR) ισούται με το σύνολο των τετραγώνων των αποκλίσεων μεταξύ των προβλεπόμενων τιμών των Y και του μέσου όρου \bar{Y} ,

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

Το άθροισμα των τετραγώνων των σφαλμάτων (SSE) ισούται με το σύνολο των τετραγώνων των αποκλίσεων των παρατηρούμενων τιμών Y και των προβλεπόμενων τιμών \hat{Y} ,

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

ii. Συντελεστής προσδιορισμού

Από τη φύση τους τα παραπάνω μέτρα SSR , SSE και SST παρέχουν λίγες πληροφορίες σχετικά με τη μεταβλητότητα του μοντέλου. Ωστόσο, η αναλογία του αθροίσματος των τετραγώνων των τιμών της παλινδρόμησης (SSR) προς το συνολικό άθροισμα των τετραγώνων (SST), μετρά το ποσοστό της συνολικής μεταβλητότητας της Y που εξηγείται από την ανεξάρτητη μεταβλητή X στο μοντέλο της παλινδρόμησης. Το κλάσμα αυτό ονομάζεται **συντελεστής προσδιορισμού** και συμβολίζεται με R^2 ,

$$R^2 = \frac{SSR}{SST}$$

ή

$$R^2 = 1 - \frac{SSE}{SST}$$

Από τα παραπάνω προκύπτει πως ο συντελεστής προσδιορισμού R^2 μπορεί να λάβει μόνο θετικές τιμές, στο διάστημα $[0,1]$. Τέλος, υπάρχει και εναλλακτικός τρόπος υπολογισμού του συντελεστή προσδιορισμού, R^2 . Στην απλή γραμμική παλινδρόμηση, όπου υπάρχει μόνο μία ανεξάρτητη μεταβλητή X , η συντελεστής προσδιορισμού ισούται με το τετράγωνο του συντελεστή συσχέτισης.

$$R^2 = r^2$$

2.7.5 Έλεγχος Στατιστικής Σημαντικότητας

- **Συντελεστής Προσδιορισμού**

Για να πραγματοποιηθεί έλεγχος στατιστικής σημαντικότητας στον συντελεστή προσδιορισμού, είναι απαραίτητο να βασιστούμε στη σχέση $TSS = SSR + SSE$. Η σχέση αυτή, εμπεριέχει τις δύο συνιστώσες της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής Y . Στην συγκεκριμένη περίπτωση αυτό που είναι βασικό να ελεγχθεί, είναι εάν το ποσοστό μεταβλητότητας της Y , το οποίο επηρεάζεται από την ανεξάρτητη μεταβλητή X , είναι διαφορετικό του μηδενός. Οι υποθέσεις φαίνονται παρακάτω:

H_0 : Η μεταβλητότητα της εξαρτημένης μεταβλητής Y δεν εξηγείται καθόλου από το μοντέλο της παλινδρόμησης.

H_1 : Ένα μέρος της μεταβλητότητας της Y εξηγείται από της εξίσωση της παλινδρόμησης

Συμπεραίνουμε λοιπόν πως είναι απαραίτητο να γίνει σύγκριση των δύο συνιστωσών της TSS . Εάν η εξηγημένη SSR είναι μεγαλύτερη της ανεξήγητης SSE , τότε η επίδραση της εξίσωσης της παλινδρόμησης είναι αρκετά σημαντική. Στην αντίθετη περίπτωση, το ποσοστό που περιγράφεται από την παλινδρόμηση είναι αμελητέο. Ο έλεγχος μεταξύ τους βασίζεται στη κατανομή F και ισούται με:

$$F_{1,(n-2)} = \frac{\sum(\hat{Y} - \bar{Y})^2 / 1}{\sum(\hat{Y} - \bar{Y})^2 / (n - 2)} = \frac{SSR}{MSE}$$

- **Συντελεστής Παλινδρόμησης b_i**

Λαμβάνοντας υπόψη τα παραπάνω, προκύπτει το ασφαλές συμπέρασμα πως ο συντελεστής παλινδρόμησης b_1 είναι διαφορετικός του μηδενός, στην περίπτωση που η ανεξάρτητη μεταβλητή X επιδρά στατιστικά σημαντικά στην εξαρτημένη μεταβλητή Y . Επιπροσθέτως, είναι απαραίτητο να συνυπολογιστεί πως υπάρχει περίπτωση ο συντελεστής παλινδρόμησης b_1 να υπόκειται στα σφάλματα της δειγματοληψίας.

Ουσιαστικά αυτό οδηγεί στην ανάγκη να γίνει γνωστό, όχι μόνο εάν ο β είναι μηδέν, αλλά να βρεθεί και το διάστημα εμπιστοσύνης του.

Στην παρακάτω σχέση φαίνεται ο τρόπος με τον οποίο υπολογίζεται το τυπικό σφάλμα του συντελεστή β :

$$\sigma_{b_1} = \frac{\sigma_\varepsilon}{\sqrt{\sum(X - \bar{X})^2}}$$

Το σ_ε στον αριθμητή του τύπου, είναι το τυπικό σφάλμα της εξίσωσης της παλινδρόμησης, δηλαδή η τετραγωνική ρίζα του σ_ε^2 . Βέβαια το σ_ε , δεν το γνωρίζουμε, σαν αποτέλεσμα θα χρησιμοποιηθεί το s_ε , η εκτίμηση του από τα δεδομένα. Η εκτίμηση του τυπικού σφάλματος συμβολίζεται με s_{b_1} και ισούται :

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum(X - \bar{X})^2}}$$

Οι δύο υποθέσεις φαίνονται παρακάτω :

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Στην περίπτωση αυτή χρησιμοποιείται ο έλεγχος t με $n - 2$ βαθμούς ελευθερίας :

$$t_{n-2} = \frac{b_1}{s_{b_1}}$$

Όταν η τιμή του t με βαθμούς ελευθερίας $|t_{n-2}|$ είναι μεγαλύτερη από την τιμή του $|t_{n-2, \alpha/2}|$, η μηδενική υπόθεση απορρίπτεται. Στην αντίθετη περίπτωση την αποδεχόμαστε.

Στην απλή παλινδρόμηση οι παραπάνω έλεγχοι οδηγούν στο ίδια ακριβώς συμπεράσματα. Για παράδειγμα όταν συντελεστής παλινδρόμησης είναι στατιστικά σημαντικός, τότε σημαντικός θα είναι και ο συντελεστής συσχέτισης και ο συντελεστής προσδιορισμού. Συμπεραίνουμε, πως οποιοσδήποτε έλεγχος από τους παραπάνω τρεις, πραγματοποιηθεί, θα οδηγηθούμε σε εξίσου έγκυρα αποτελέσματα.

2.8 Πολλαπλή Γραμμική Παλινδρόμηση

Η προηγούμενη ενότητα επικεντρώθηκε στην απλή γραμμική παλινδρόμηση, όπου χρησιμοποιείται μία ανεξάρτητη μεταβλητή X , για την πρόβλεψη των τιμών της εξαρτημένης

μεταβλητής Y . Στην ενότητα αυτή, θα αναλυθεί το μοντέλο πολλαπλής γραμμικής παλινδρόμησης, όπου χρησιμοποιούνται παραπάνω από μία ανεξάρτητες μεταβλητές για την καλύτερη πρόβλεψη των τιμών της εξαρτημένης Y .

Σκοπός της πολλαπλής παλινδρόμησης είναι να περιγράψει τη σχέση που υπάρχει μεταξύ της εξαρτημένης μεταβλητής Y και των ανεξάρτητων μεταβλητών X_1, X_2, \dots, X_k . Το μοντέλο έχει την εξής μορφή:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Όπου:

Y : η τιμή της εξαρτημένης μεταβλητής

X_1, X_2, \dots, X_k : οι τιμές των ανεξάρτητων μεταβλητών

β_0 : σταθερός όρος

$\beta_1, \beta_2, \dots, \beta_k$: οι συντελεστές παλινδρόμησης που περιγράφουν την επίδραση των ανεξάρτητων μεταβλητών

ε : σφάλμα ή κατάλοιπο, δηλαδή η διαφορά της πραγματικής τιμής της Y και της προβλεπόμενης.

Οι συντελεστές της παλινδρόμησης εκφράζουν, κατά πόσο θα μεταβληθεί η τιμή της εξαρτημένης μεταβλητής Y , όταν αυξηθεί κατά μία μονάδα μία ανεξάρτητη μεταβλητή X_i και οι υπόλοιπες παραμείνουν σταθερές.

Παρακάτω ακολουθούν οι προϋποθέσεις που πρέπει να πληρούνται, για να δημιουργηθεί με ασφάλεια ένα μοντέλο παλινδρόμησης :

- Τα τυπικά σφάλματα που προκύπτουν από την παλινδρόμηση πρέπει να ακολουθούν κανονική κατανομή και να είναι ανεξάρτητα μεταξύ τους.
- Η μέση τιμή των σφαλμάτων να ισούται με το μηδέν $E(\varepsilon_i) = 0$
- Τα σφάλματα είναι απαραίτητο να έχουν σταθερή διακύμανση $Var(\varepsilon_i) = \sigma^2$
- Οι μεταβλητές X_i είναι ανεξάρτητες από τα σφάλματα και θεωρούνται σταθερές ποσότητες

Στην πολλαπλή παλινδρόμηση είναι αρκετά σημαντικό οι μεταβλητές X_i να είναι ανεξάρτητες μεταξύ τους για να μην δημιουργηθεί το πρόβλημα της πολυσυγγραμμικότητας.

2.8.1 Εκτίμηση της Εξίσωσης της Πολλαπλής Γραμμικής Παλινδρόμησης

Η παραπάνω διαδικασία πραγματοποιείται με στόχο να εκτιμηθούν οι παράμετροι του μοντέλου της πολλαπλής παλινδρόμησης, δηλαδή $\beta_1, \beta_2, \dots, \beta_k$. Οι εκτιμήσεις, των συντελεστών της γραμμικής παλινδρόμησης συμβολίζονται με b_1, b_2, \dots, b_k , αντίστοιχα. Η εξίσωση που προκύπτει είναι η παρακάτω:

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Οι συντελεστές b_1, b_2, \dots, b_k δείχνουν τη μερική επίδραση που ασκούν οι ανεξάρτητες μεταβλητές στη Y . Για παράδειγμα, ο συντελεστής b_2 υποδηλώνει την μεταβολή της Y που θα προκύψει εάν η μεταβλητή X_2 μεταβληθεί κατά μία μονάδα και οι άλλες ανεξάρτητες μεταβλητές παραμείνουν σταθερές.

Η εξίσωση που φαίνεται παραπάνω έχει προκύψει από τη μέθοδο ελαχίστων τετραγώνων, όπως και στην απλή παλινδρόμηση. Αναζητούμε, εκείνες τις τιμές των b_1, b_2, \dots, b_k που ελαχιστοποιούν το άθροισμα των τετραγώνων των αποκλίσεων μεταξύ των πραγματικών τιμών της Y και των αναμενόμενων, οι οποίες προκύπτουν από την εξίσωση της παλινδρόμησης.

2.8.2 Συντελεστής Πολλαπλού προσδιορισμού

Ο συντελεστής προσδιορισμού είναι μία πολύ σημαντική παράμετρος στην πολλαπλή παλινδρόμηση, καθώς μετρά το ποσοστό της μεταβλητότητας της Y , που οφείλεται σε όλες τις ανεξάρτητες μεταβλητές μαζί. Ουσιαστικά, δείχνει την συνολική επίδραση που ασκούν οι X_1, X_2, \dots, X_k στην εξαρτημένη Y και περιγράφεται από την σχέση παρακάτω:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{TSS}$$

Στο συγκεκριμένο σημείο είναι απαραίτητο να αναφερθεί πως ο συντελεστής προσδιορισμού υπάρχει περίπτωση να αυξάνεται "τεχνητά", όσο προστίθενται στο μοντέλο παλινδρόμησης περισσότεροι μεταβλητές. Κάτι τέτοιο μπορεί να οδηγήσει εσφαλμένα συμπεράσματα για την εγκυρότητα του μοντέλου, ειδικά στην περίπτωση που οι ανεξάρτητες μεταβλητές είναι περισσότεροι από τον αριθμό του δείγματος. Το συγκεκριμένο πρόβλημα έρχεται να αντιμετωπίσει ο διορθωμένος συντελεστής προσδιορισμού, όπου εκφράζεται από την παρακάτω σχέση:

$$\bar{R}^2 = 1 - \frac{\frac{SSE}{[n - (k + 1)]}}{\frac{SST}{n - 1}} = 1 - (1 - R^2) \frac{n - 1}{n - (k + 1)}$$

Σε περίπτωση του διορθωμένου συντελεστή προσδιορισμού, εάν υπάρξει αύξηση με την πρόσθεση κάποιας μεταβλητής, τότε είναι βέβαιο πως η μεταβλητή αυτή θα πρέπει να συμπεριληφθεί στο μοντέλο παλινδρόμησης.

2.8.3 Έλεγχοι Στατιστικής Σημαντικότητας

Ο πρώτος έλεγχος που είναι απαραίτητο να διενεργηθεί αφορά συνολικά την σημαντικότητα της παλινδρόμησης. Παρακάτω παρατίθενται η μηδενική και η εναλλακτική υπόθεση:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_n = 0$$

$$H_1: \text{τουλάχιστον ένα από τα } \beta_i \neq 0, i = 1, 2, \dots, n$$

Συμπεραίνοντας από τα παραπάνω, στην περίπτωση που απορριφθεί η μηδενική υπόθεση και αποδεχτούμε πως τουλάχιστον ένας από τους παράγοντες β είναι διάφορος του μηδενός, τότε μία τουλάχιστον από τις ανεξάρτητες μεταβλητές έχει στατιστικά σημαντική σχέση με την εξαρτημένη μεταβλητή Y . Στην αντίθετη περίπτωση, απρίπτως την εναλλακτική υπόθεση, αποδεχόμαστε πως οι παράγοντες της εξίσωσης είναι μηδέν, συνεπώς δεν υπάρχει γραμμική σχέση ανάμεσα στην X και Y .

Για να γίνει έλεγχος απόρριψης ή αποδοχής της παραπάνω μηδενικής υπόθεσης, είναι απαραίτητο να χρησιμοποιήσουμε τον F στατιστικό έλεγχο. Με αυτόν τον τρόπο ουσιαστικά εξετάζουμε την αναλογία μεταξύ του μέσου αθροίσματος τετραγώνων της παλινδρόμησης και με του μέσου τετραγωνικού σφάλματος του μοντέλου. Ο F στατιστικός έλεγχος περιγράφεται από την παρακάτω σχέση (Lind κ.ά., 2018) :

$$F_{stat} = \frac{MSR}{MSE}$$

Χρησιμοποιώντας επίπεδο σημαντικότητας α , ισχύει ο παρακάτω κανόνας:

Εάν $F_{k,n-k-1} > F_{(k,n-k-1),\alpha}$, η μηδενική υπόθεση απορρίπτεται.

Εάν $F_{k,n-k-1} < F_{(k,n-k-1),\alpha}$, αποδεχόμαστε την μηδενική υπόθεση.

Στη συνέχεια, σε περίπτωση που απορριφθεί η μηδενική υπόθεση και αποδειχθεί πως εξηγείται μεγάλο μέρος της μεταβλητότητας της Y , προχωράμε στον έλεγχο του διαστήματος εμπιστοσύνης των παραγόντων β_i . Παρακάτω φαίνονται οι υποθέσεις:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Στη συγκεκριμένη περίπτωση χρησιμοποιούμε τον T στατιστικό έλεγχο, με $n - k - 1$ βαθμούς ελευθερίας.

Εάν $|t_{n-k-1}| > \left| t_{n-k-1, \frac{\alpha}{2}} \right|$, η μηδενική υπόθεση απορρίπτεται.

Εάν $|t_{n-k-1}| < \left| t_{n-k-1, \frac{\alpha}{2}} \right|$, αποδεχόμαστε την μηδενική υπόθεση.

Ο παραπάνω έλεγχος αφορά τη συνεισφορά μίας μεταβλητής και ταυτόχρονα περιλαμβάνονται κι άλλες μεταβλητές στην εξίσωση παλινδρόμησης.

2.9 Χρονοσειρές

2.9.1 Ορισμός

Οι χρονοσειρές είναι ένα σύνολο παρατηρήσεων x_i , όπου η κάθε παρατήρηση έχει καταγραφεί σε συγκεκριμένο χρόνο t (R. Maciejewsk, K. S. Candan, "Introduction to Time Series"). Οι τιμές των παρατηρήσεων συμβολίζονται με Y_t και οι συγκεκριμένες χρονικές στιγμές με t_i . Αν λοιπόν ενώσουμε τα δύο παραπάνω σημεία σε ένα καρτεσιανό σύστημα αξόνων θα δημιουργηθεί ένα διάγραμμα, το οποίο θα περιγράφει ουσιαστικά τη συμπεριφορά του φαινομένου που μελετάται, μέσα σε ορισμένο χρόνο.

2.9.2 Συνιστώσες Χρονοσειρών

Οι χρονοσειρές αποτελούνται από τέσσερις διαφορετικές συνιστώσες, οι οποίες διαχωρίζονται από τα δεδομένα που υπάρχουν διαθέσιμα και αναλύονται παρακάτω :

Τάση (trend)

Μια χρονοσειρά έχει τάση όταν υπάρχει σημαντική μεταβολή των τιμών του φαινομένου σε κάποιο χρονικό διάστημα. Η μεταβολή αυτή, μπορεί να είναι θετική ή αρνητική.

Εποχικότητα (seasonality)

Όταν μία χρονοσειρά επηρεάζεται μία συγκεκριμένη χρονική στιγμή με σταθερή συχνότητα μέσα στο χρόνο που μελετάται, έχει εποχικότητα. Για παράδειγμα, μία συγκεκριμένη εποχή(καλοκαίρι, άνοιξη) ή μία συγκεκριμένη μέρα (Εθνική Εορτή κάποια χώρας).

Κυκλικές Κινήσεις (cyclic movements)

Μία χρονοσειρά εμφανίζει κυκλικές κινήσεις όταν οι παρατηρήσεις του φαινομένου εμφανίζουν ακανόνιστα αύξηση ή πτώση στις τιμές τους χωρίς σταθερή ταχύτητα.

Ακανόνιστες Διακυμάνσεις (irregular fluctuations)

Οι ακανόνιστες διακυμάνσεις περιγράφουν γεγονότα σε μία χρονοσειρά που είναι αδύνατο να εξηγηθούν ή να προβλεφθούν από τα υπάρχοντα δεδομένα. Τέτοια γεγονότα είναι για παράδειγμα οι απεργίες ή κάποιο φυσικό φαινόμενο.

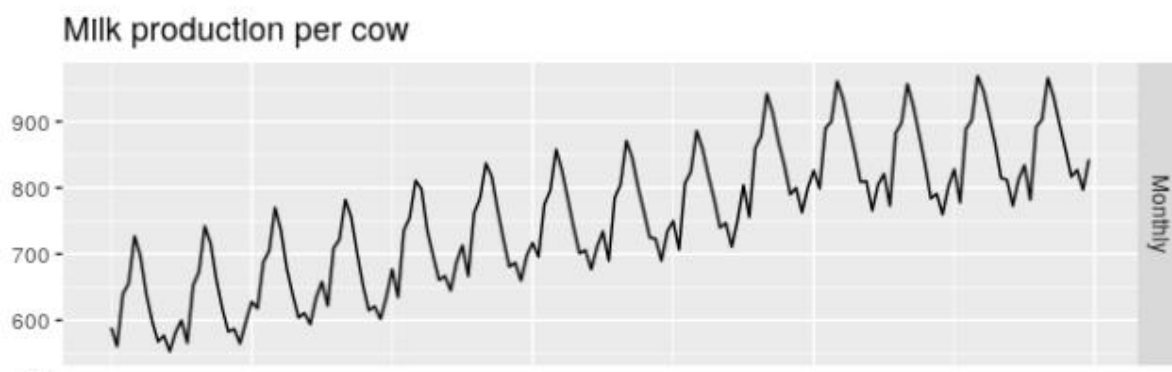
Σε μία χρονοσειρά δεν είναι απαραίτητο να εντοπίζονται όλες μαζί οι παραπάνω συνιστώσες, αλλά κάποιες από αυτές. Η ύπαρξη των συνιστωσών αυτών, αλλά και το κατά πόσο αυτές προσδιορίζουν μία χρονική σειρά εξετάζεται από δύο μαθηματικά μοντέλα. Το προσθετικό μοντέλο και το πολλαπλασιαστικό.

2.9.3 Το Προσθετικό Μοντέλο

Το προσθετικό μοντέλο χρησιμοποιείται όταν το εύρος των μεταβολών της χρονοσειράς παραμένει σταθερό στο χρόνο σε οποιαδήποτε από τις παραπάνω περιπτώσεις. Η σχέση που περιγράφει το προσθετικό μοντέλο εμφανίζεται παρακάτω:

$$Y_t = T_t + C_t + S_t + I_t$$

Το Σχήμα 15 απεικονίζεται μία χρονοσειρά όπου θα μπορούσε να περιγραφεί από το προσθετικό μοντέλο :



Σχήμα 15 : Παράδειγμα Χρονοσειράς για το Πρόσθετο Μοντέλο (μηνιαία παραγωγή Γάλακτος)

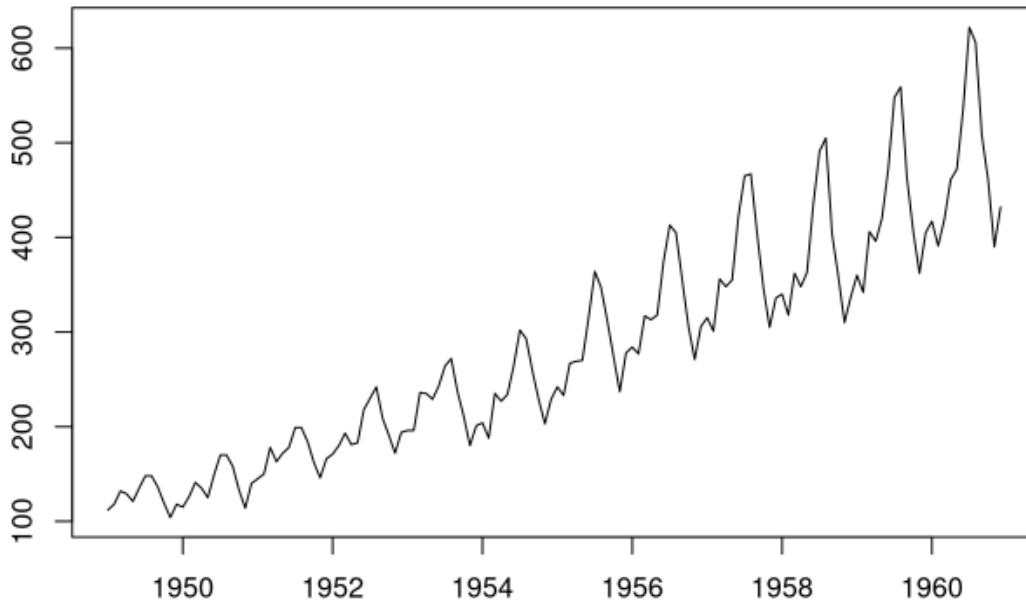
Πηγή: Rob J. Hyndman, George Athanasopoulos,(2018) Forecasting: Principles and Practice

2.9.4 Το Πολλαπλασιαστικό Μοντέλο

Όταν το εύρος των μεταβολών μίας χρονοσειράς αυξάνεται ανάλογα με το μέγεθος της τάσης ή της εποχικότητας, τότε καταλληλότερο να περιγράψει αυτή τη χρονοσειρά κρίνεται το πολλαπλασιαστικό μοντέλο και περιγράφεται από την εξής σχέση :

$$Y_t = T_t * C_t * S_t * I_t$$

Παρακάτω, στο Σχήμα 16, φαίνεται μία χρονοσειρά που περιγράφεται καλύτερα από το πολλαπλασιαστικό μοντέλο:



Σχήμα 16 : Παράδειγμα Χρονοσειράς για Πολλαπλασιαστικό Μοντέλο

Πηγή: www.datascienceblog.net/post/machine-learning/forecasting-an-introduction

2.9.5 Στοχαστικότητα

Οι χρονολογικές σειρές οι οποίες φέρουν πραγματικά μεγέθη, παρουσιάζουν τον λεγόμενο λευκό θόρυβο. Λευκό θόρυβο παρουσιάζουν χρονοσειρές όπου οι μεταβλητές τους κατανέμονται κανονικά, με μέση τιμή μηδέν και με σταθερή διακύμανση, χωρίς να συσχετίζονται μεταξύ τους. Οι χρονοσειρές που παρουσιάζουν λευκό θόρυβο ονομάζονται και στοχαστικές.

2.9.6 Στασιμότητα

Στάσιμη ονομάζεται μία χρονοσειρά της οποίας οι ιδιότητες δεν εξαρτώνται από τον χρόνο κατά τον οποίο παρατηρείται η σειρά. Οι χρονοσειρές με τάσεις ή με εποχικότητα δεν είναι σταθερές. Γενικότερα η τάση και η εποχικότητα επηρεάζουν την αξία των χρονοσειρών (Hyndman & Athanasopoulos, 2018). Η στασιμότητα είναι μία πολύ σημαντική προϋπόθεση πριν πραγματοποιηθεί η ανάλυση των χρονοσειρών. Όταν εμφανίζεται περιοδικότητα ή εποχικότητα μία χρονοσειρά είναι μη στάσιμη και είναι απαραίτητη η εξομάλυνση και ουδετεροποίησή της πριν προχωρήσει ο ερευνητής σε μελλοντικές προβλέψεις.

2.9.7 Γραμμικά Μοντέλα Πρόβλεψης

Τα μοντέλα τα οποία χρησιμοποιούνται στην ανάλυση χρονοσειρών είναι δύο, το Αυτοπαλίνδρομο (autoregressive-AR) και το μοντέλο κινητού μέσου (moving average-MA). Με τον συνδυασμό των δύο παραπάνω σχηματίζονται τα πλέον πιο χρησιμοποιημένα από το

ευρύ κοινό μοντέλα, το Autoregressive Moving Average-ARMA και το Autoregressive Integrated Moving Average-ARIMA. Γενικότερα όλα τα παραπάνω είναι παραλλαγές του μοντέλου ARIMA και ονομάζονται και BOX-Jenkins μοντέλα.

Σε ένα μοντέλο πολλαπλής παλινδρόμησης, προβλέπουμε τη μεταβλητή ενδιαφέροντος χρησιμοποιώντας έναν γραμμικό συνδυασμό προβλέψεων. Σε ένα αυτοπαλίνδρομο μοντέλο, προβλέπουμε τη μεταβλητή ενδιαφέροντος χρησιμοποιώντας έναν γραμμικό συνδυασμό προηγούμενων τιμών της μεταβλητής. Ο όρος autoregression υποδηλώνει μια παλινδρόμηση κάποιας μεταβλητής ενάντια στον εαυτό της (Hyndman & Athanasopoulos, 2018). Παρακάτω φαίνεται ένα Αυτοπαλίνδρομο (AR) μοντέλο :

$$Y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t$$

Όπου:

ε_t : είναι ο λευκός θόρυβος με σταθερή διακύμανση και μέσο όρο μηδενικό

$\alpha_1 \dots \alpha_p$: οι παράμετροι του υποδείγματος

p : αριθμός προηγούμενων παρατηρήσεων

Η διακύμανση των τιμών σε αυτή την περίπτωση μετρά την απόστασή τους από τον μέσο όρο και παρακάτω φαίνεται η σχέση που την περιγράφει:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Όταν μία χρονοσειρά αποτελείται από τυχαίες μεταβλητές, οι οποίες έχουν την ίδια κατανομή, και δεν υπάρχουν συσχετίσεις μεταξύ των διαδοχικών στοιχείων της, καλείται λευκός θόρυβος (white noise). Η κατανομή της συγκεκριμένης χρονοσειράς εκφράζεται ως $WN(0, \sigma_\varepsilon^2)$ με σταθερή διασπορά και μηδενικό μέσο όρο.

2.9.8 Moving Average Model (MA)

Σε αντίθεση με το παραπάνω μοντέλο που χρησιμοποιούσε προηγούμενες παρατηρήσεις για την πρόβλεψη των τιμών, στο μοντέλο κινητού μέσου χρησιμοποιούνται προηγούμενα σφάλματα πρόβλεψης (Hyndman & Athanasopoulos, 2018). Το μοντέλο φαίνεται παρακάτω:

$$Y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_p \varepsilon_{t-p}$$

Το ε_t συμβολίζει ξανά τον λευκό θόρυβο. Αναφερόμαστε σε αυτό το μοντέλο σαν μοντέλο κινητού μέσου τάξεως q .

2.9.9 Auto Regressive Moving Average (ARMA)

Στην συγκεκριμένη περίπτωση θα γίνει αναφορά σε ένα από τα πιο εύχρηστα μοντέλα πρόβλεψης χρονοσειρών, καθώς το μοντέλο ARMA συνδυάζει τα πλεονεκτήματα των δύο παραπάνω. Το μοντέλο ARMA τάξεων (p, q) περιγράφεται από την παρακάτω σχέση:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_p \varepsilon_{t-p}$$

Όπου:

ε_t : ο λευκός θόρυβος

$a_1 \dots a_p$: οι παράμετροι του υποδείγματος για το AR

$\beta_1 \dots \beta_p$: οι παράμετροι του υποδείγματος για το MA

2.9.10 Auto Regressive Integrated Moving Average (ARIMA) Model

Παραπάνω αναφέρθηκαν διάφορα μοντέλα πρόβλεψης χρονοσειρών, τα οποία είναι κατάλληλα όταν οι διαδικασίες είναι στάσιμες. Στον πραγματικό κόσμο, οι περισσότερες παρατηρήσεις παρουσιάζουν τάση ή εποχικότητα, συνεπώς ο μέσος όρος και η διακύμανση εξαρτώνται από τον χρόνο. Για να εφαρμοστεί ένα μοντέλο σε πραγματικά δεδομένα είναι απαραίτητο να εμπεριέχει και τα μη στάσιμα χαρακτηριστικά. Την λύση σε αυτή την περίπτωση καλείται να δώσει το μοντέλο ARIMA. Το συγκεκριμένο μοντέλο, λόγω της ιδιότητας του να προσεγγίζει μη στάσιμες διαδικασίες το καθιστά πολύτιμο εργαλείο σε αναλύσεις χρονοσειρών. Με τη χρήση διαφορών ξεπερνά τους περιορισμούς που είναι τα προηγούμενα υποδείγματα και αυτό επιτυγχάνεται αφαιρώντας την παρατήρηση της τρέχουσας περιόδου από την προηγούμενη. Το μοντέλο ARIMA τάξεως (p, d, q) φαίνεται παρακάτω:

$$y'_t = a_1 y'_{t-1} + a_2 y'_{t-2} + \dots + a_p y'_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_p \varepsilon_{t-p}$$

Όπου:

p : οι παράμετροι της αυτοπαλίνδρομης διαδικασίας

d : ο αριθμός των διαφορών προκειμένου η χρονοσειρά να γίνει στάσιμη

q : οι παράμετροι της διαδικασίας του κινητού μέσου

ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ ΚΑΙ ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ

3.1 Δεδομένα

Τα δεδομένα για την παρούσα εργασία αντλήθηκαν από το Ενεργειακό Ισοζύγιο (Energy Balance) όλων των Ευρωπαϊκών χωρών, το οποίο δημιουργήθηκε από την Ευρωπαϊκή Επιτροπή, με σκοπό την παρακολούθηση των ενεργειακών προϊόντων και τη ροή τους στην οικονομία. Το Ενεργειακό Ισοζύγιο είναι μια συνολική παρουσίαση όλων των δραστηριοτήτων μίας χώρας που σχετίζονται με την ενέργεια. Επιτρέπει στους ενδιαφερόμενους να δουν τη συνολική ποσότητα ενέργειας που εξάγεται από το περιβάλλον, εμπορεύεται, μετασχηματίζεται και χρησιμοποιείται από τους τελικούς χρήστες. Επιπλέον, το Ενεργειακό Ισοζύγιο παρουσιάζει όλα τα στατιστικά σημαντικά ενεργειακά προϊόντα (fuels) μίας χώρας, την παραγωγή τους και την κατανάλωσή τους από διαφορετικούς οικονομικούς παράγοντες (βιομηχανία, μεταφορές κλπ.). Η μονάδα μέτρησης που χρησιμοποιήθηκε είναι η Mtoe (million tonnes of oil equivalent) και είναι κοινή για όλα τα δεδομένα.

Τα δεδομένα περιγράφουν την παραγωγή, τον μετασχηματισμό και τις μορφές κατανάλωσης των πετρελαιοειδών και άλλων καυσίμων σε 28 κράτη-μέλη της Ευρωπαϊκής Ένωσης, για διάρκεια 18 χρόνων. Τα δεδομένα είναι διατεταγμένα σε γραμμές, οι οποίες παρουσιάζουν τις ενεργειακές ροές των καυσίμων και σε στήλες, οι οποίες αντιπροσωπεύουν τις χώρες της Ευρωπαϊκής Ένωσης.

Στην παρούσα μελέτη τα καύσιμα, τα οποία παίζουν κυρίαρχο ρόλο, είναι τα πετρελαιοειδή και το φυσικό αέριο. Παρακάτω φαίνονται αναλυτικότερα οι ενεργειακές ροές των πετρελαιοειδών και κάποιες του φυσικού αερίου :

Ενεργειακές ροές:

- **Πρωτογενής Παραγωγή** (αντιπροσωπεύει κάθε είδους άντληση ενεργειακών προϊόντων από φυσικές πηγές π.χ. η εκχύλιση αργού πετρελαίου)
- **Εισαγωγές** (αντιπροσωπεύει όλες τις εισαγωγές πετρελαιοειδών και φυσικού αερίου σε μία χώρα)
- **Εξαγωγές** (αντιπροσωπεύει όλες τις εξαγωγές πετρελαιοειδών μίας χώρας)
- **Μετασχηματισμός Πετρελαιοειδών** (ο μετασχηματισμός αργού πετρελαίου σε παράγωγα προϊόντα πετρελαίου όπως βενζίνη, κηροζίνη ντίζελ κλπ.)
- **Κατανάλωση Πετρελαιοειδών** (Η συνολική ενέργεια που καταναλώνεται από τους τελικούς χρήστες)

3.2 Ερευνητικά ερωτήματα

Τα τελευταία δέκα περίπου χρόνια η κατανάλωση πετρελαίου μειώθηκε παγκοσμίως και πιο συγκεκριμένα, έφτασε από 33% το 2005 σε 31,6% το 2017. Τα στοιχεία αυτά εγείρουν το ενδιαφέρον για το μέλλον της αγοράς του πετρελαίου αλλά και γενικότερα της ενέργειας. Διάφορες έρευνες προβλέπουν την ραγδαία μείωση της κατανάλωσης πετρελαιοειδών έως το 2050 (Eurostat,2017). Από την άλλη πλευρά, η κατανάλωση του φυσικού αερίου αυξάνεται χρόνο με το χρόνο και φαίνεται να κατακτά έδαφος όχι μόνο στην Ευρώπη αλλά και παγκοσμίως.

Τα γεγονότα αυτά οδηγούν σε εύλογα ερωτήματα σχετικά με το πως επηρεάζεται η κατανάλωση του πετρελαίου αλλά και ποιο θα είναι το μέλλον της στις Ευρωπαϊκές χώρες. Αναλυτικότερα τα ερωτήματα που θα απαντηθούν στην παρούσα εργασία αναφέρονται παρακάτω:

Πως επιδρά η παραγωγή, ο μετασχηματισμός των πετρελαιοειδών και οι εισαγωγές φυσικού αερίου και πετρελαίου στην τελική ενεργειακή κατανάλωση πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης. Αρχικά ερευνάται αν υπάρχει κάποια σχέση μεταξύ της εγχώριας πρωτογενούς παραγωγής πετρελαιοειδών και της κατανάλωσης τους από τους τελικούς χρήστες. Δεύτερον, εξετάζεται κατά πόσο επηρεάζει ο μετασχηματισμός των πετρελαιοειδών σε παράγωγα προϊόντα, την κατανάλωση του πετρελαίου από τους πολίτες κάθε χώρας. Τρίτον, γίνεται προσπάθεια εύρεσης της σχέσης μεταξύ των εισαγωγών πετρελαιοειδών και της κατανάλωσης τους. Τέλος, ερευνάται κατά πόσο υπάρχει σχέση μεταξύ των εισαγωγών φυσικού αερίου στις Ευρωπαϊκές χώρες και στην κατανάλωση πετρελαιοειδών.

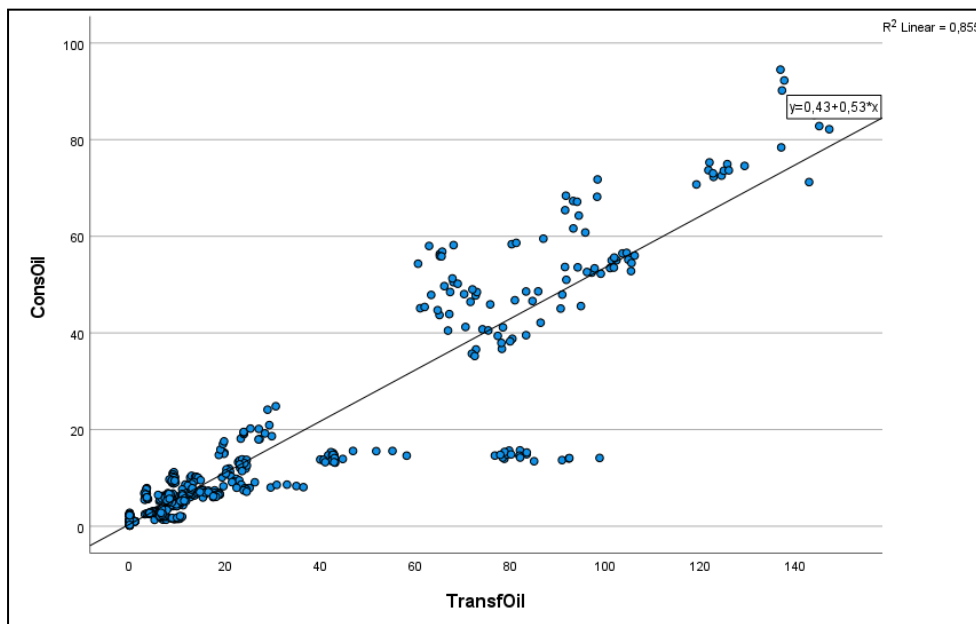
Στόχος είναι να δημιουργηθεί ένα μοντέλο παλινδρόμησης που θα περιγράφει ικανοποιητικά την τελική ενεργειακή κατανάλωση των πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης.

Προβλέψεις για την εξέλιξη της κατανάλωσης των πετρελαιοειδών στις Ευρωπαϊκές χώρες. Σε αυτή την περίπτωση θα γίνει προσπάθεια πρόβλεψης της κατανάλωσης του πετρελαίου με την μέθοδο ARIMA.

ΚΕΦΑΛΑΙΟ 4: ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

4.1 Ανάλυση Μεταβλητών

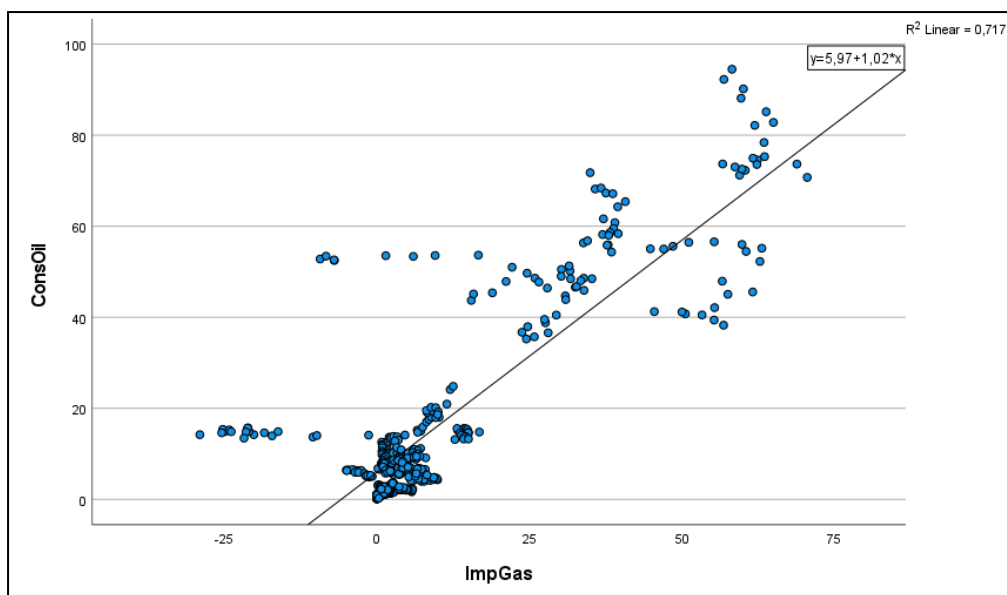
Στην παρακάτω ενότητα απεικονίζονται γραφικά οι μεταβλητές που θα αναλυθούν, σε διαγράμματα διασποράς. Τα διαγράμματα διασποράς, στην ανάλυση παλινδρόμησης, είναι αρκετά αποτελεσματικά εργαλεία καθώς αποκαλύπτουν τη σχέση που έχουν μεταξύ τους οι μεταβλητές.



Διάγραμμα 1 : Διάγραμμα Διασποράς για τις μεταβλητές *ConsOil* και *TransfOil*

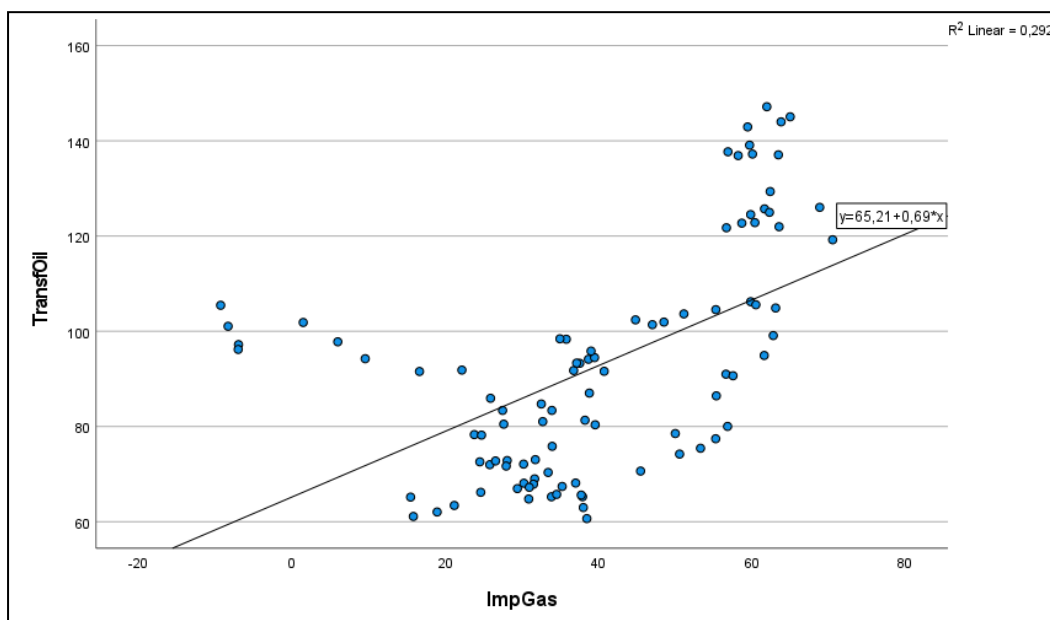
Παραπάνω δίνεται το διάγραμμα διασποράς της κατανάλωσης πετρελαιοειδών σε σχέση με τον μετασχηματισμό πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης. Στο διάγραμμα αυτό έχει σχεδιαστεί η ευθεία ελαχίστων τετραγώνων, η οποία βοηθά να προσδιοριστεί εάν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών. Το συμπέρασμα είναι πως όσο περισσότερα πετρελαιοειδή μετασχηματίζονται σε καταναλώσιμα προϊόντα για τον πληθυσμό κάθε χώρας, τόσο μεγαλύτερη διάθεση θα υπάρχει, άρα και μεγαλύτερη κατανάλωση.

Παρατηρούνται κάποιες αποκλίσεις στις διασπορές των ζευγών και είναι εμφανής η ύπαρξη ακραίων τιμών. Επιπλέον φαίνεται η συρρίκνωση των ζευγών γύρω από συγκεκριμένες τιμές.



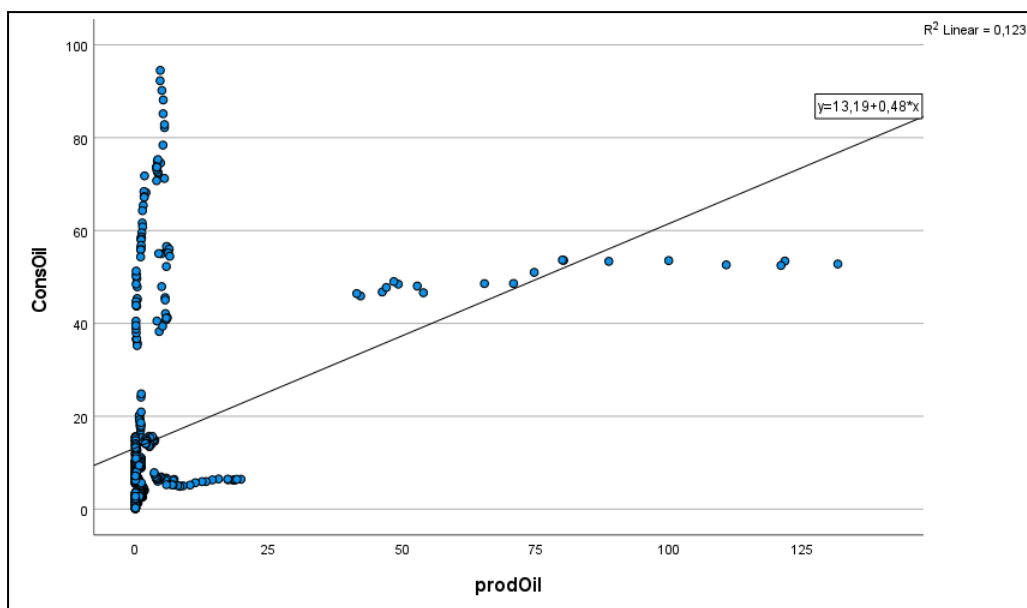
Διάγραμμα 2 : Διάγραμμα Διασποράς για τις μεταβλητές *ConsOil* και *ImpGas*

Στη συγκεκριμένη περίπτωση τα αποτελέσματα είναι παρόμοια. Η κατανάλωση πετρελαιοειδών έχει θετική γραμμική σχέση με την εισαγωγή φυσικού αερίου στη χώρα, κάτι για το οποίο θα περίμενε κανείς το αντίθετο. Στο συγκεκριμένο σημείο είναι απαραίτητο να διευκρινιστεί πως υπάρχουν και αρνητικές τιμές διότι η μεταβλητή *ImpGas* αντιπροσωπεύει τις καθαρές εισαγωγές φυσικού αερίου των Ευρωπαϊκών χωρών.



Διάγραμμα 3 : Διάγραμμα Διασποράς για τις μεταβλητές *TransfOil* και *ImpGas*

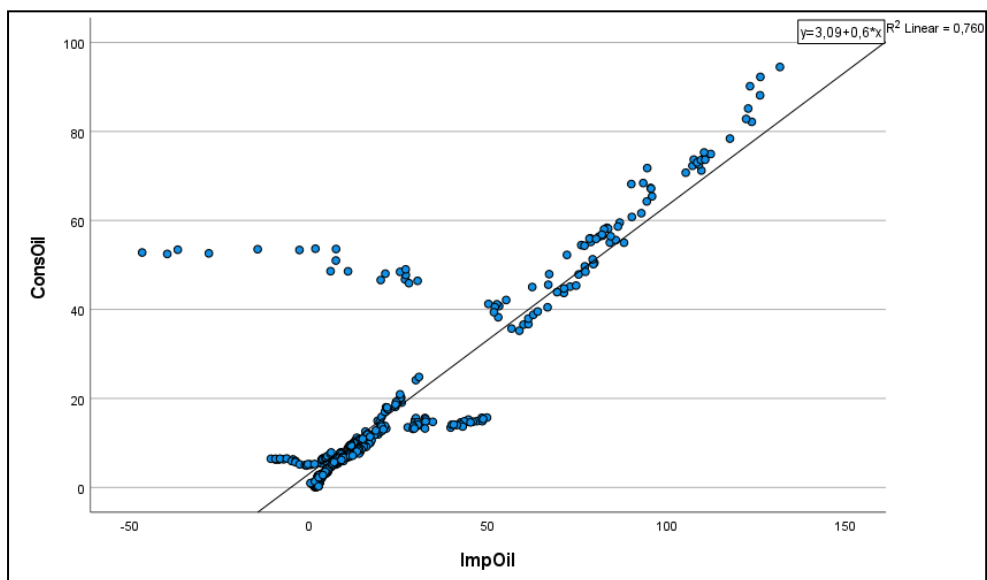
Στο διάγραμμα 3 εξετάζεται η σχέση που υπάρχει μεταξύ των ανεξάρτητων μεταβλητών, όπου δεν φαίνεται κάποια γραμμική σχέση, κάτι το οποίο είναι ιδιαίτερα σημαντικό για τα αποτελέσματα της παλινδρόμησης.



Διάγραμμα 4 : Διάγραμμα Διασποράς για τις Μεταβλητές ProdOil και ConsOil

Παραπάνω απεικονίζεται η σχέση μεταξύ της πρωτογενούς παραγωγής πετρελαιοειδών και στην κατανάλωση τους. Φαίνεται ξεκάθαρα πως δεν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών και αυτό είναι απόλυτα λογικό. Στις Ευρωπαϊκές χώρες η εξόρυξη πετρελαίου είναι μηδαμινή, οι ποσότητες πετρελαιοειδών που χρησιμοποιούνται σε εγχώριες δραστηριότητες προέρχονται κυρίως από εισαγωγές.

Για τους παραπάνω λόγους η συγκεκριμένη μεταβλητή δεν θα συμπεριληφθεί στο μοντέλο παλινδρόμησης.



Διάγραμμα 5 : Διάγραμμα Διασποράς για τις μεταβλητές ImpOil και ConsOil

Σε αυτή την περίπτωση φαίνεται πως υπάρχει γραμμική σχέση μεταξύ των μεταβλητών ImpOil και ConsOil. Ουσιαστικά αυτό σημαίνει πως η κατανάλωση πετρελαιοειδών επηρεάζεται άμεσα από τις εισαγωγές που θα γίνουν στις χώρες της Ευρώπης.

4.2 Κατανομές Δεδομένων

Στο κεφάλαιο αυτό ξεκινάει η επεξεργασία των δεδομένων που χρησιμοποιήθηκαν στην παρούσα μελέτη. Εξετάζονται τα περιγραφικά χαρακτηριστικά κάθε μεταβλητής όπως η μέτρηση κεντρικής τάσης (μέσος όρος), η μέτρηση της διασποράς (διακύμανση, τυπική απόκλιση, εύρος) και το σχήμα (ασυμμετρία, κύρτωση). Επιπλέον με βάση τα παραπάνω μέτρα και διαγράμματα ελέγχεται η ύπαρξη κανονικότητας και η ύπαρξη έκτροπων παρατηρήσεων, οι οποίες και πρέπει εξεταστούν.

Αρχικά, χρησιμοποιήθηκε η μέθοδος της βηματικής παλινδρόμησης (stepwise regression), με σκοπό να επιλεγθούν, μεταξύ της ομάδας των πιθανών ανεξάρτητων μεταβλητών που αναφέρθηκαν στο Κεφάλαιο 3, εκείνες οι μεταβλητές που θα συνεισφέρουν περισσότερο στη μείωση της ανερμήνευτης μεταβλητότητας της εξαρτημένης μεταβλητής Y .

Συμπεραίνοντας από την παραπάνω μέθοδο, η εξαρτημένη μεταβλητή που θα μελετηθεί είναι η κατανάλωση των πετρελαιοειδών στην Ευρωπαϊκή Ένωση και οι ανεξάρτητες είναι η εισαγωγή φυσικού αερίου και ο μετασχηματισμός πετρελαιοειδών. Παρακάτω, στον Πίνακα 1 παρατίθενται όλοι συμβολισμοί των μεταβλητών που χρησιμοποιούνται.

Πίνακας 1 : Συμβολισμοί Μεταβλητών

Συμβολισμός	Επεξήγηση
ConsOil	Κατανάλωση πετρελαιοειδών
TransfOil	Μετασχηματισμός πετρελαιοειδών
ImpGas	Εισαγωγή φυσικού αερίου

4.2.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil)

Από την επεξεργασία με το λογισμικό στατιστικού περιεχομένου SPSS, προκύπτουν τα παρακάτω αποτελέσματα των βασικών στατιστικών χαρακτηριστικών.

Στατιστικά μέτρα για ConsOil

Αριθμός Παρατηρήσεων:532

Μέσος Όρος:15,12

Διάμεσος:6,55

Διασπορά:426,87

Τυπική Απόκλιση:20,6

Ελάχιστη Τιμή:0

Μέγιστη Τιμή:95

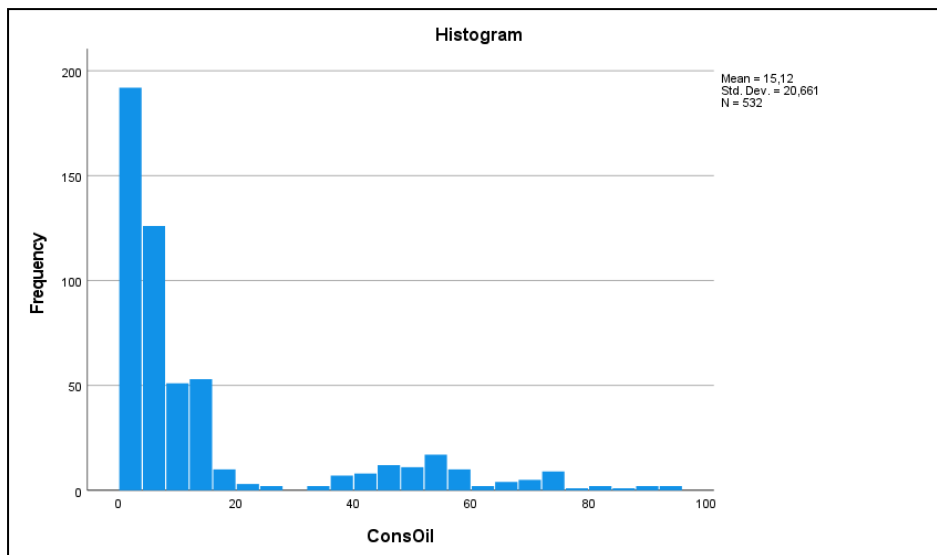
Εύρος:94

Ασυμμετρία:1,8

Κύρτωση:2,38

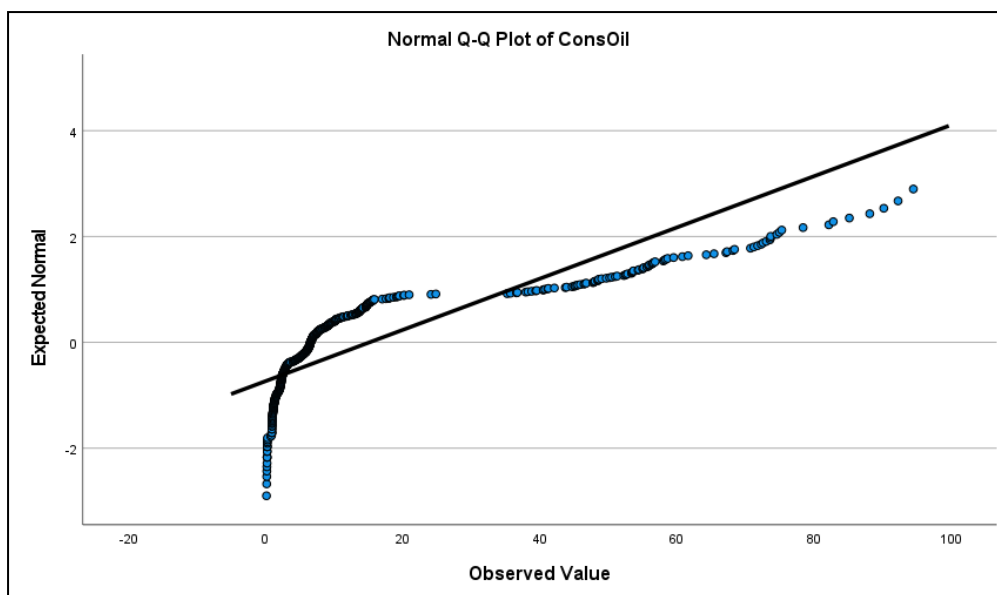
Το σύνολο των παρατηρήσεων όπως φαίνεται παραπάνω είναι 532 για διάρκεια 18 ετών. Ο μέσος όρος (15,12) και η διάμεσος (6,55) αποκλίνουν αρκετά μεταξύ τους, συμπεραίνοντας

πως η κατανομή που ακολουθεί η συγκεκριμένη μεταβλητή δεν είναι συμμετρική. Σημαντικές παράμετροι που πρέπει να ελέγχονται είναι η ασυμμετρία και η κύρτωση. Οι τιμές των παραμέτρων αυτών είναι αντιπροσωπευτικές για τον έλεγχο της κανονικότητας. Όσο πιο κοντά στο μηδέν βρίσκονται οι συγκεκριμένες τιμές, θεωρείται ότι η μεταβλητή ακολουθεί κανονική κατανομή. Οι τιμές της ασυμμετρίας (1,8) και της κύρτωσης (2,38) απέχουν αρκετά από το μηδέν. Περισσότερα συμπεράσματα για την κανονικότητα των δεδομένων αντλούνται από τα παρακάτω διαγράμματα.



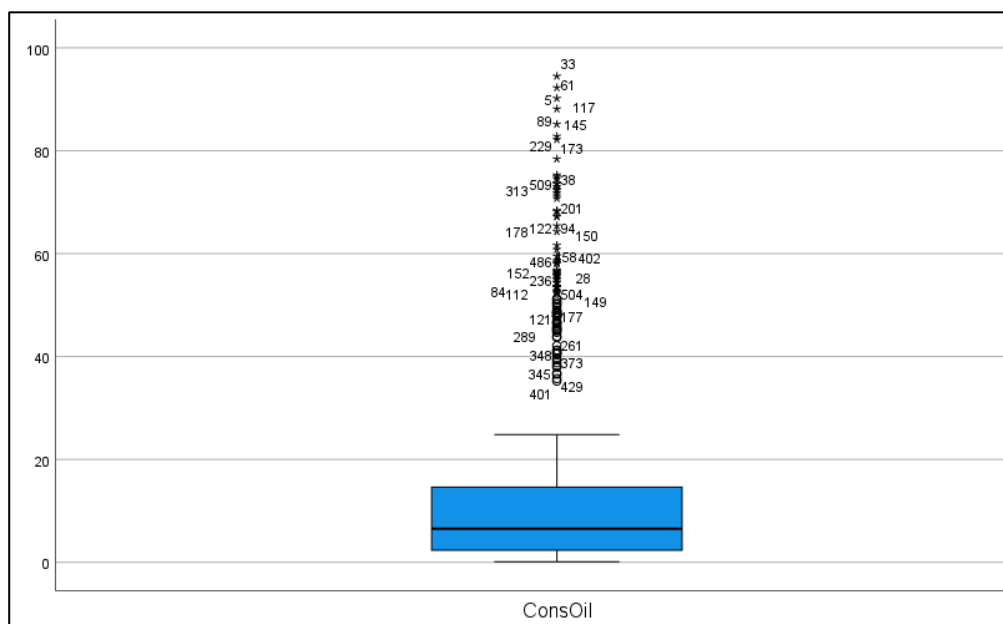
Διάγραμμα 6 : Ιστόγραμμα για την μεταβλητή *ConsOil*

Από το Ιστόγραμμα διακρίνεται μια θετική ασυμμετρία ή ασυμμετρία προς τα δεξιά, με τιμή 1.8. Η ασυμμετρία εξηγεί και ότι ο μέσος όρος είναι μεγαλύτερος από τη διάμεσο. Επιπλέον, η τιμή της κύρτωσης, 2.38, εξηγείται από την συγκέντρωση πολλών τιμών γύρω από ένα συγκεκριμένο σημείο και συνεπώς σχηματίζεται μία λεπτόκυρτη κατανομή. Τέλος, η "ουρά" που παρατηρείται και στα παρακάτω διαγράμματα οφείλεται σε αρκετές ακραίες τιμές που περιέχουν οι παρατηρήσεις.



Διάγραμμα 7 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή *ConsOil*

Στο διάγραμμα ελέγχου της κανονικότητας φαίνεται πως τα σημεία προσεγγίζουν την ευθεία γραμμή, αν και πάλι είναι ξεκάθαρο πως η κατανομή της μεταβλητής επηρεάζεται αρκετά από ακραίες τιμές.



Διάγραμμα 8 : Θηκόγραμμα για την μεταβλητή *ConsOil*

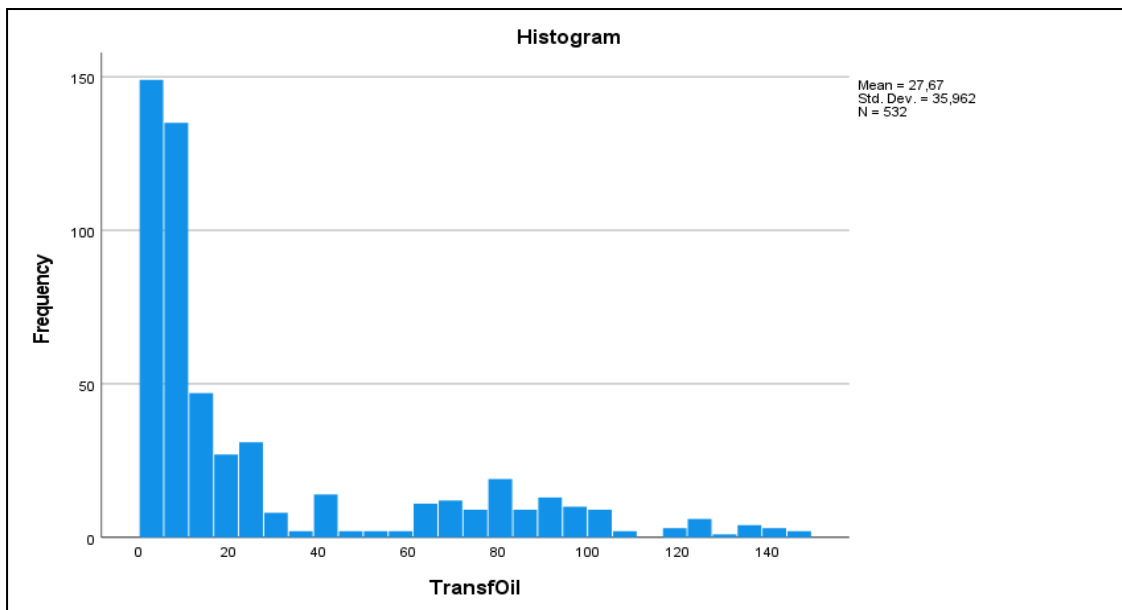
Στο θηκόγραμμα διακρίνονται ξεκάθαρα οι ακραίες τιμές που επηρεάζουν τα δεδομένα.

4.2.2 Μεταβλητή Μετασχηματισμός Πετρελαιοειδών (*TransfOil*)

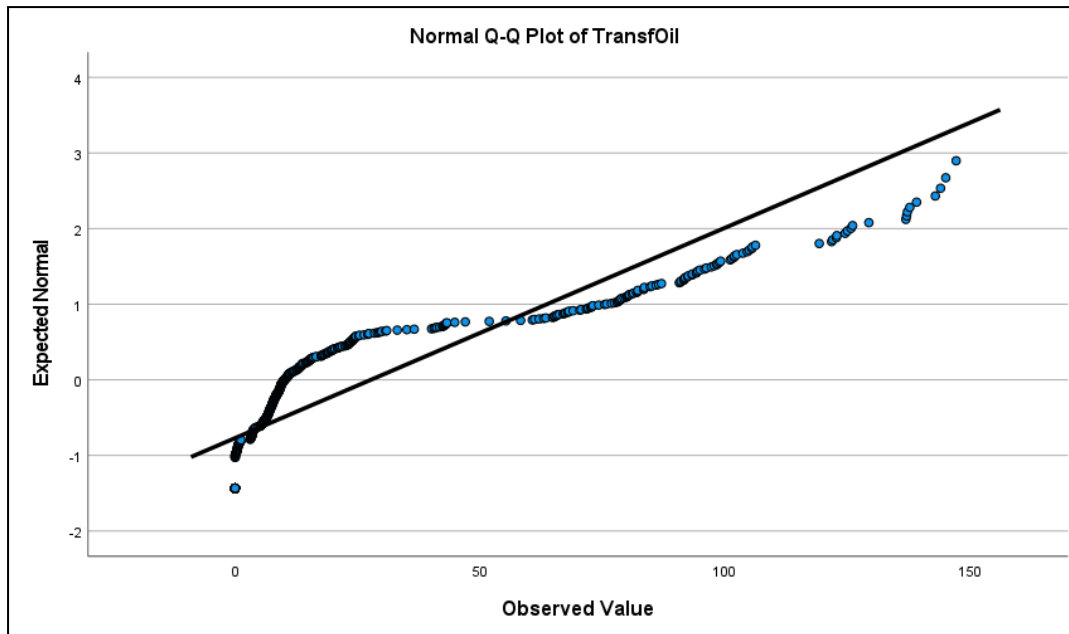
Στατιστικά μέτρα για *TransfOil*

Αριθμός Παρατηρήσεων:532
Μέσος Όρος:27,67
Διάμεσος:10,06
Διασπορά:1293,2
Τυπική Απόκλιση:35,9
Ελάχιστη Τιμή:0
Μέγιστη Τιμή:147
Εύρος:147
Ασυμμετρία:1,48
Κύρτωση:1,09

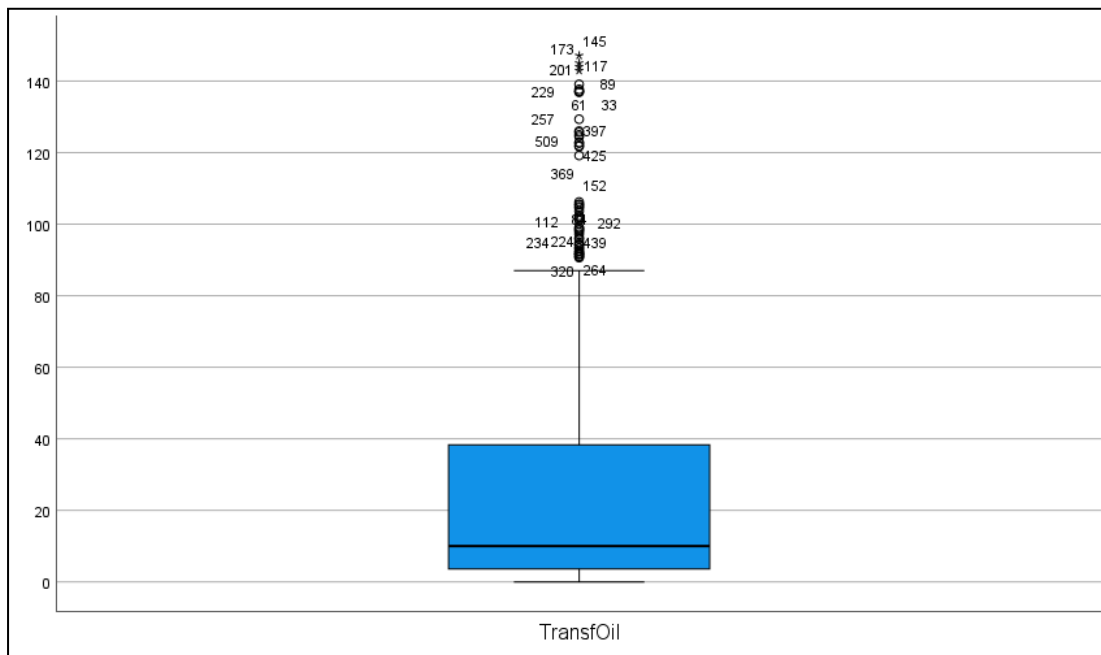
Ο μέσος όρος (27,67) και η διάμεσος (10,06) έχουν μεγάλη διαφορά στις τιμές τους, συμπεραίνοντας πως η κατανομή που ακολουθεί η συγκεκριμένη μεταβλητή δεν είναι συμμετρική. Η ασυμμετρία και η κύρτωση της μεταβλητής φαίνεται να βρίσκονται κοντά στο 0, άρα έχουμε μία ένδειξη κανονικότητας. Παρακάτω στα διαγράμματα, εντοπίζεται ξεκάθαρα μία θετική ασυμμετρία και κύρτωση προ τα δεξιά. Στο θηκόγραμμα είναι εμφανές το πρόβλημα στην κατανομή της μεταβλητής που δημιουργείται ξανά από τις ακραίες τιμές.



Διάγραμμα 9 : Ιστόγραμμα για την μεταβλητή *TransfOil*



Διάγραμμα 10 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή TransfOil



Διάγραμμα 11 : Θηκόγραμμα για την μεταβλητή TransfOil

4.2.3 Μεταβλητή Εισαγωγές Φυσικού Αερίου (ImpGas)

Στατιστικά μέτρα για TransfOil

Αριθμός Παρατηρήσεων:532

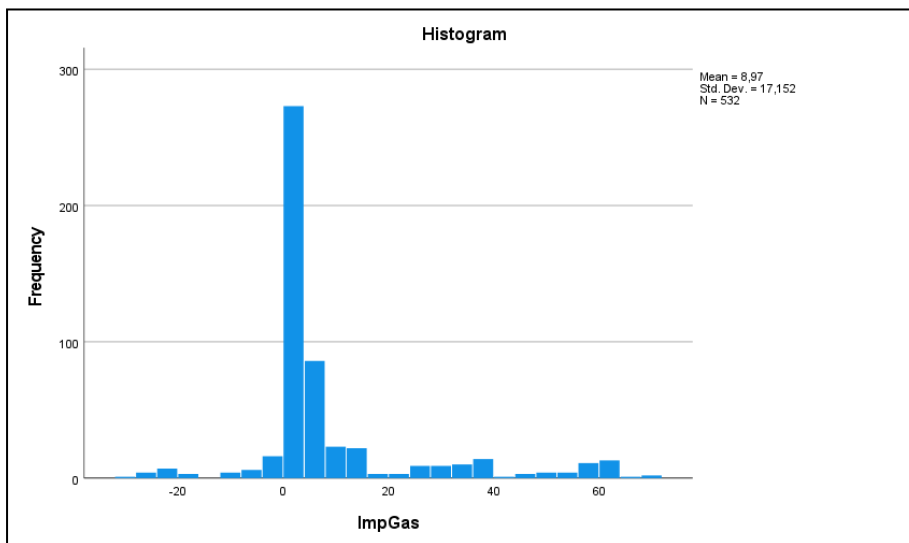
Μέσος Όρος:8.97

Διάμεσος:2.90

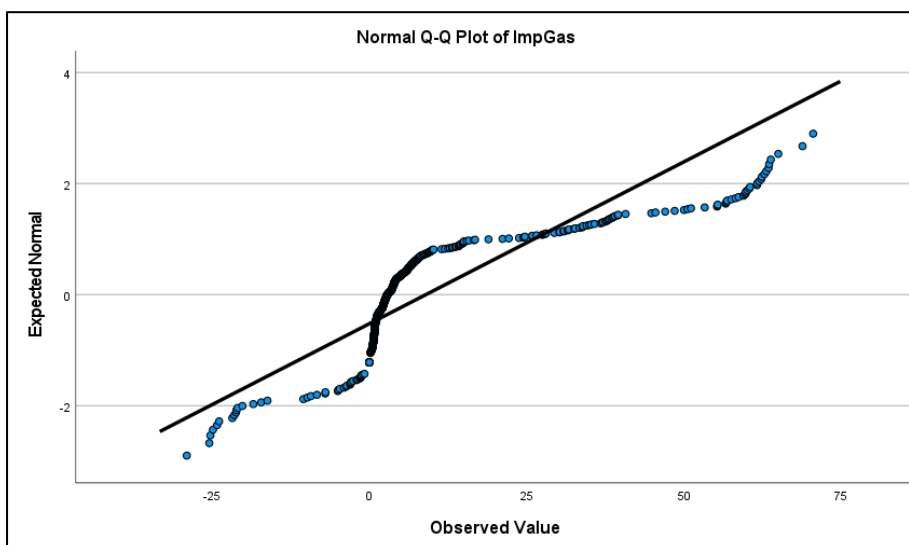
Διασπορά:294.1

Τυπική Απόκλιση:17.1
Ελάχιστη Τιμή:-29
Μέγιστη Τιμή:71
Εύρος:100
Ασυμμετρία:1,77
Κύρτωση:3.04

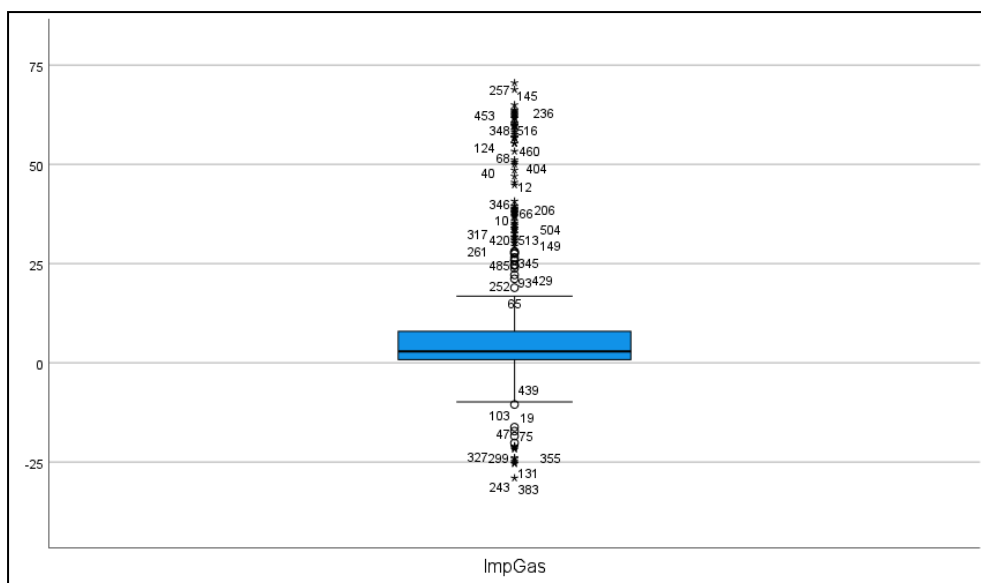
Από τις παραπάνω μετρήσεις ενδιαφέρον παρουσιάζουν οι τιμές της κύρτωσης και της ασυμμετρίας, οι οποίες είναι αρκετά πάνω από το μηδέν με 3,04 και 1,77 αντίστοιχα, κάτι που υποδεικνύει μία λεπτόκυρτη κατανομή με θετική ασυμμετρία. Στα παρακάτω διαγράμματα φαίνεται πιο καθαρά το σχήμα της κατανομής και οι έκτροπες παρατηρήσεις που επηρεάζουν αρκετά την κανονικότητα των δεδομένων.



Διάγραμμα 12 : Ιστόγραμμα για την μεταβλητή ImpGas



Διάγραμμα 13 : Διάγραμμα Ελέγχου Κανονικότητας για την Μεταβλητή ImpGas



Διάγραμμα 14 : Θηκόγραμμα για την μεταβλητή *ImpGas*

Παρατηρώντας τα παραπάνω αποτελέσματα, βγαίνει το συμπέρασμα πως οι μεταβλητές που θα χρησιμοποιηθούν στην ανάλυση δεν ακολουθούν κανονική κατανομή και επηρεάζονται σημαντικά από έκτροπες παρατηρήσεις. Ουσιαστικά οι τιμές των ακραίων αυτών παρατηρήσεων αντιπροσωπεύουν κάποιες συγκεκριμένες ευρωπαϊκές χώρες, κοινές για όλες τις μεταβλητές. Οι χώρες αυτές είναι το Ηνωμένο Βασίλειο, η Γερμανία, η Ισπανία, η Ιταλία και το Βέλγιο, κάτι απόλυτα λογικό διότι συγκρίνοντας τον πληθυσμό των συγκεκριμένων χωρών με αυτόν των υπολοίπων της Ευρωπαϊκής Ένωσης, βγαίνει το συμπέρασμα πως είναι κατά πολύ μεγαλύτερος. Συνεπώς και οι τιμές της κατανάλωσης πετρελαιοειδών, ο μετασχηματισμός τους σε καταναλώσιμα προϊόντα αλλά και οι εισαγωγές φυσικού αερίου θα διαφέρει σημαντικά για τις χώρες αυτές. Σαν αποτέλεσμα οι τιμές που αφορούν τις χώρες αυτές απέχουν αρκετά από το μέσο όρο των υπολοίπων και παρουσιάζονται ως έκτροπες παρατηρήσεις.

Κρίθηκε απαραίτητο λοιπόν, για την ορθότητα των αποτελεσμάτων, να δημιουργηθούν δύο ομάδες χωρών και να μελετηθούν ξεχωριστά. Η πρώτη εμπεριέχει τις παρακάτω χώρες: Ηνωμένο Βασίλειο, Γερμανία, Ισπανία, Ιταλία, Βέλγιο και η δεύτερη τις Βουλγαρία, Τσεχία, Δανία, Εσθονία, Ιρλανδία, Ελλάδα, Κροατία, Κύπρος, Λετονία, Λιθουανία, Ουγγαρία, Μάλτα, Αυστρία, Πορτογαλία, Ρουμανία, Σλοβενία, Σλοβακία, Φιλανδία και Σουηδία.

Πρώτα θα παρουσιαστούν και θα αναλυθούν τα περιγραφικά χαρακτηριστικά των μεταβλητών που ανήκουν στην πρώτη ομάδα χωρών και στη συνέχεια της δεύτερης ομάδας.

4.3 Πρώτη Ομάδα Χωρών

4.3.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil)

Στατιστικά μέτρα για ConsOil

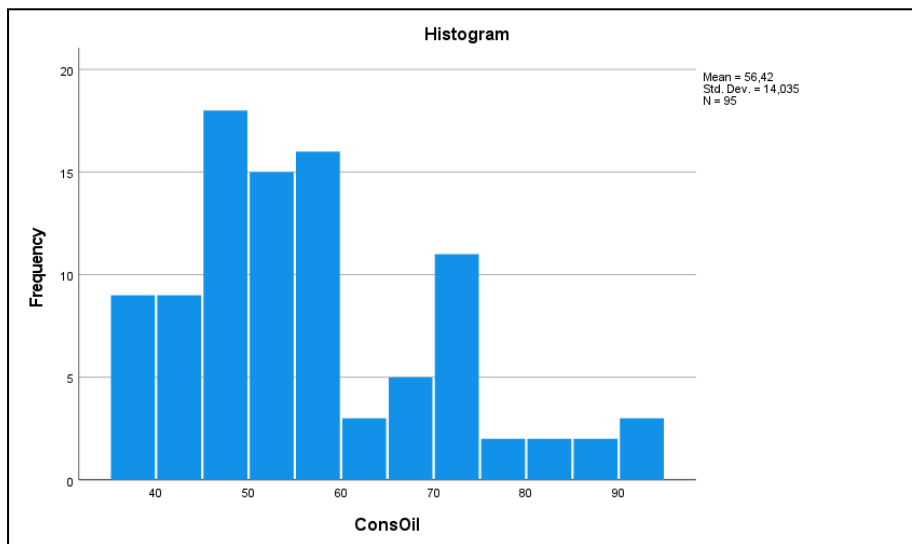
Αριθμός Παρατηρήσεων:95

Μέσος Όρος:56,42

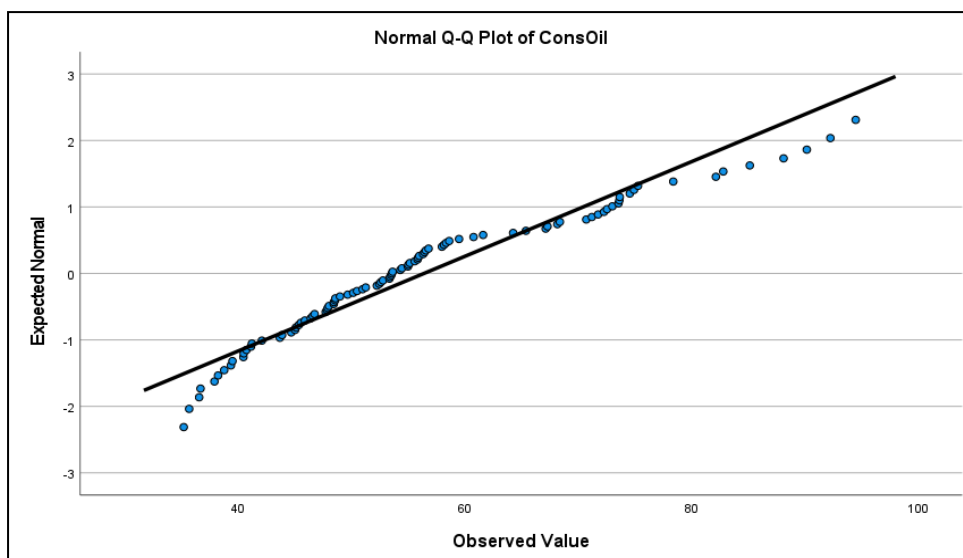
Διάμεσος:53,59

Διασπορά:196,97
Τυπική Απόκλιση:14,03
Ελάχιστη Τιμή:35
Μέγιστη Τιμή:95
Εύρος:59
Ασυμμετρία:0,7
Κύρτωση:0,02

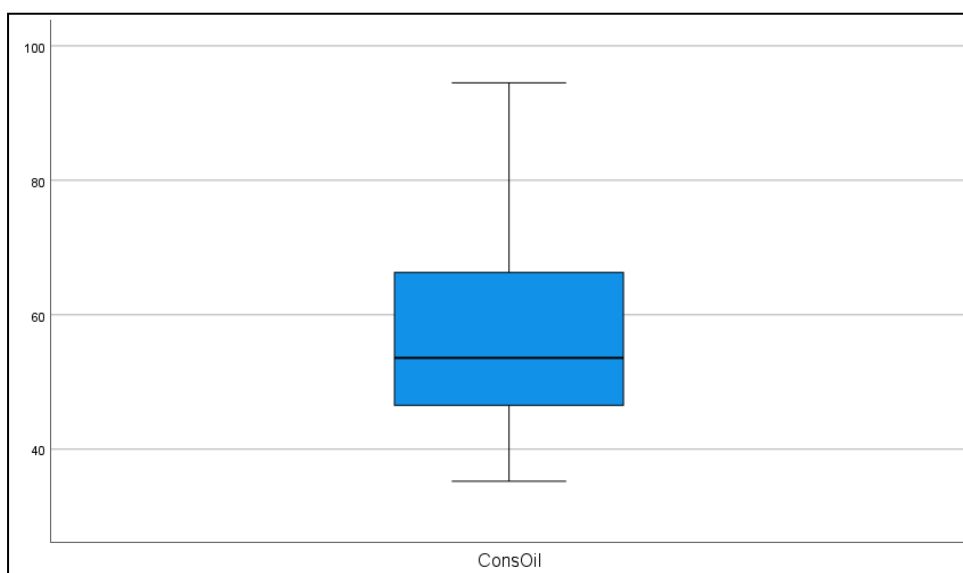
Το σύνολο των παρατηρήσεων μετά τον διαχωρισμό των χωρών είναι 95. Φαίνεται ότι τα αποτελέσματα είναι σχετικά καλύτερα σε σχέση με την πρώτη ανάλυση. Η διάμεσος (53,59) και ο μέσος (56,42) έχουν αρκετά κοντινές τιμές, κάτι που υποδηλώνει μία συμμετρική κατανομή. Επιπλέον σημαντικό είναι πως οι τιμές της κύρτωσης και ασυμμετρίας βρίσκονται κοντά στο μηδέν, με 0,02 και 0,7 αντίστοιχα, κάτι που υποδηλώνει μία ένδειξη κανονικότητας στα δεδομένα. Παρακάτω, στα διαγράμματα φαίνεται πως οι μεταβλητές προσεγγίζουν την κανονική κατανομή, αν και υπάρχουν κάποιες αποκλίσεις, όπως στο διάγραμμα ελέγχου κανονικότητας τα δεδομένα δεν ευθυγραμμίζονται τελείως πάνω στην ευθεία γραμμή. Επιπλέον, στο θηκόγραμμα δεν υπάρχουν έκτροπες παρατηρήσεις, παρόλα αυτά παρατηρείται ασυμμετρία προς τα δεξιά, αυτό φαίνεται και από το ιστόγραμμα συχνοτήτων.



Διάγραμμα 15 : Ιστόγραμμα για την μεταβλητή *ConsOil*



Διάγραμμα 16 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή *ConsOil*



Διάγραμμα 17 : Θηκόγραμμα για την μεταβλητή *ConsOil*

4.3.2 Μεταβλητή Μετασχηματισμός Πετρελαιοειδών (*TransfOil*)

Στατιστικά μέτρα για *TransfOil*

Αριθμός Παρατηρήσεων:95

Μέσος Όρος:92.47

Διάμεσος:91

Διασπορά:566.4

Τυπική Απόκλιση:23.7

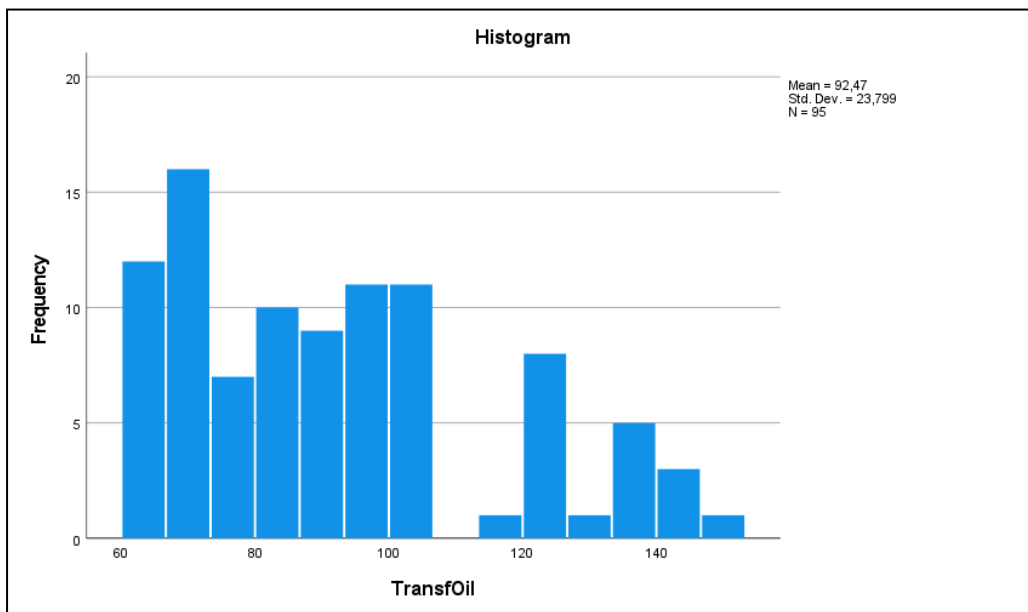
Ελάχιστη Τιμή:61

Μέγιστη Τιμή:147

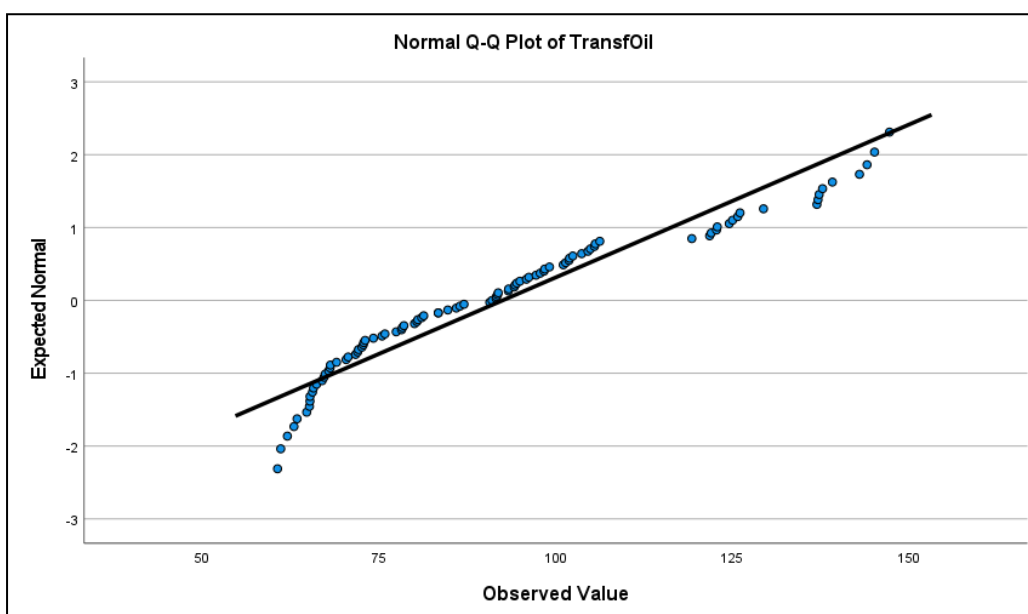
Εύρος:86

Ασυμμετρία:0,7
Κύρτωση:-0.4

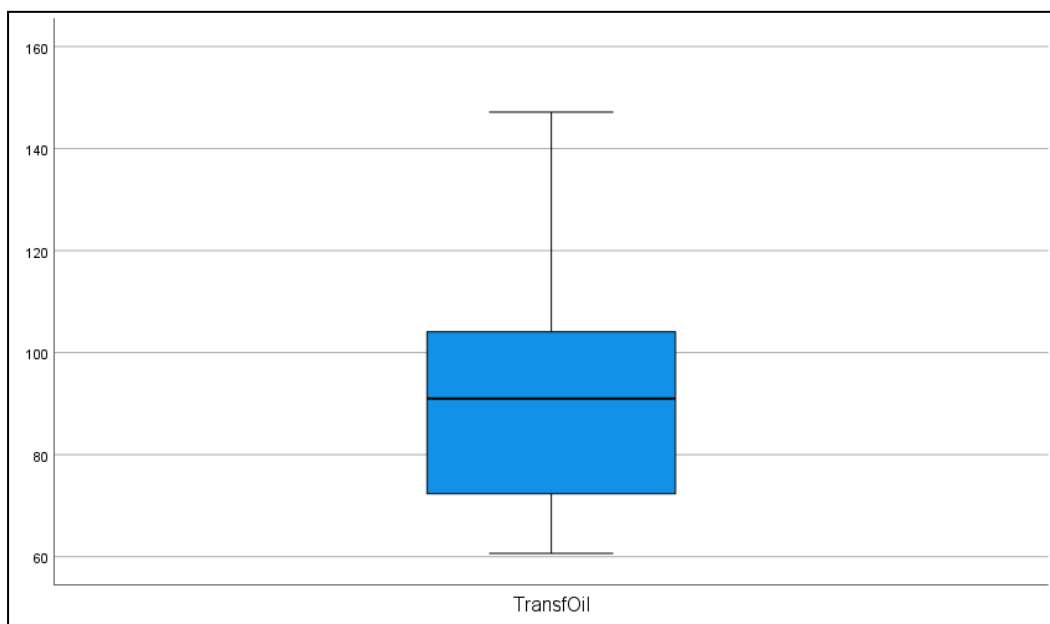
Για την μεταβλητή TransfOil, τα αποτελέσματα επίσης εμφανίζουν μια ένδειξη κανονικότητας με τον μέσο όρο και διάμεσο να έχουν κοντινές τιμές. Το ίδιο μπορούμε να πούμε και για την κύρτωση και την ασυμμετρία, όπου παρουσιάζουν τιμές κοντά στο μηδέν. Παρακάτω φαίνονται και τα διαγράμματα που εξετάζονται για την ύπαρξη της κανονικότητας, στα οποία παρατηρούμε παρόμοια προβλήματα στην ασυμμετρία με αυτά της προηγούμενης μεταβλητής (ConsOil).



Διάγραμμα 18 : Ιστόγραμμα για την μεταβλητή TransfOil



Διάγραμμα 19 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή TransfOil



Διάγραμμα 20 : Θηκόγραμμα για την μεταβλητή *TransfOil*

4.3.3 Μεταβλητή Εισαγωγή Φυσικού Αερίου (ImpGas)

Στατιστικά μέτρα για ImpGas

Αριθμός Παρατηρήσεων:95

Μέσος Όρος:39.51

Διάμεσος:37.90

Διασπορά:347.3

Τυπική Απόκλιση:18.6

Ελάχιστη Τιμή:-9

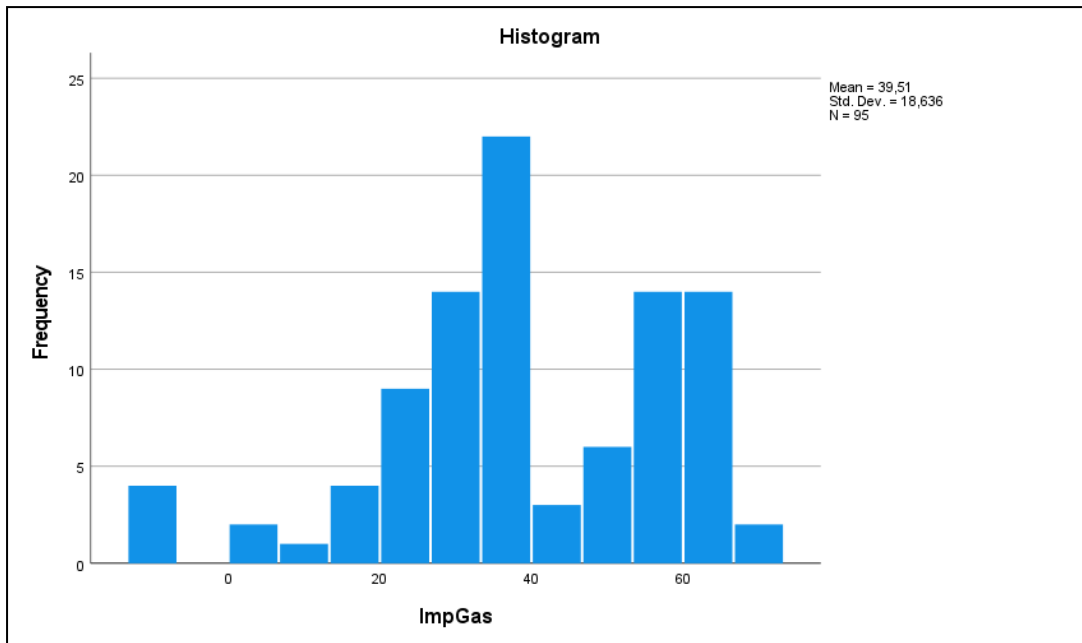
Μέγιστη Τιμή:71

Εύρος:80

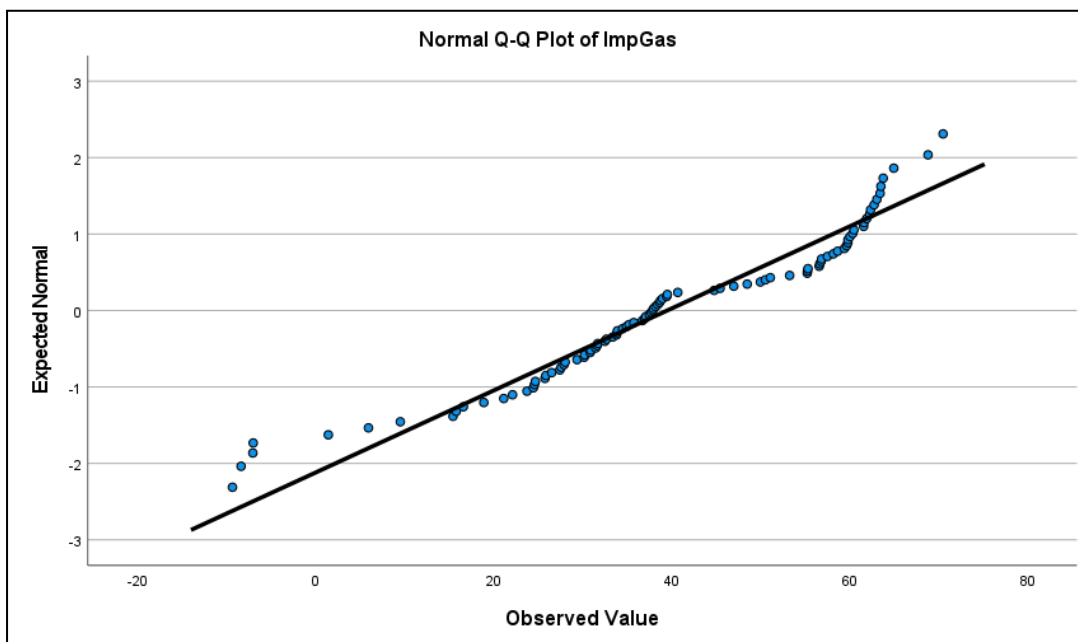
Ασυμμετρία:-0,5

Κύρτωση:0,04

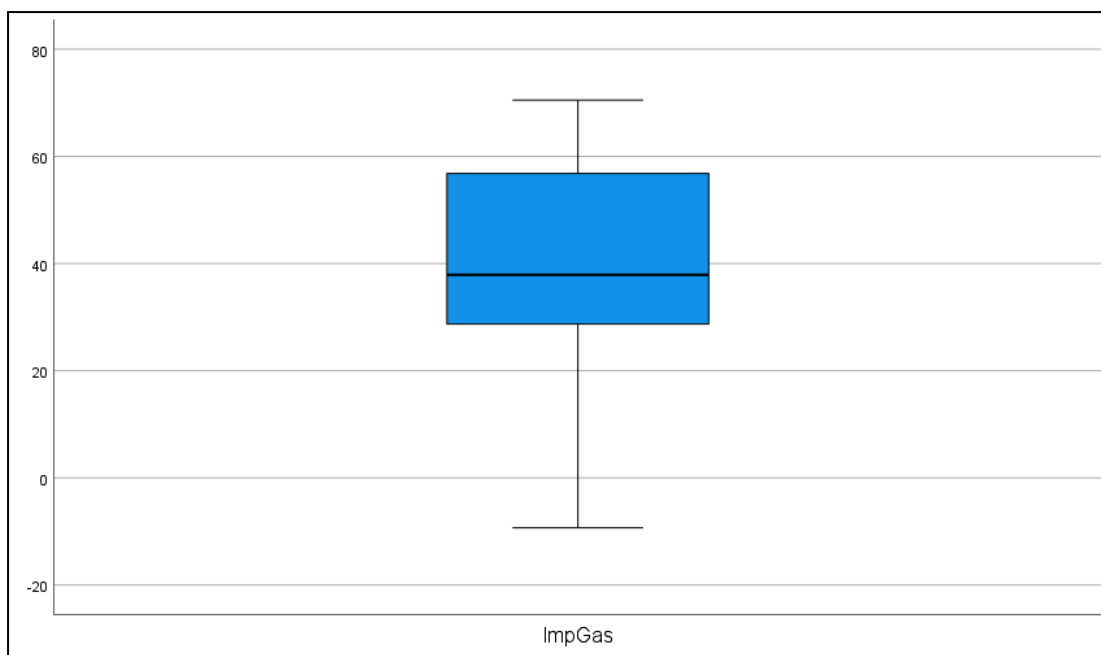
Τα περιγραφικά χαρακτηριστικά της συγκεκριμένης μεταβλητής, και σε αυτή την περίπτωση φαίνεται να έχουν βελτιωθεί αρκετά μετά τον διαχωρισμό των Ευρωπαϊκών χωρών. Οι τιμές της διαμέσου και του μέσου όρου είναι και πάλι κοντινές. Επιπλέον, οι τιμές της κύρτωσης και της ασυμμετρίας βρίσκονται κοντά στο μηδέν, πληρώντας ένα κριτήριο της κανονικότητάς. Στη συγκεκριμένη περίπτωση υπάρχουν ξανά αποκλίσεις από την κανονικότητα. Παρατηρώντας τα διαγράμματα είναι εμφανής η ασυμμετρία προς τα αριστερά.



Διάγραμμα 21 : Ιστόγραμμα για την μεταβλητή ImpGas



Διάγραμμα 22 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ImpGas



Διάγραμμα 23 : Θηκόγραμμα για την μεταβλητή ImpGas

Πραγματοποιώντας τον έλεγχο κανονικότητας Kolmogorov-Smirnov και Shaphiro-Wilk παίρνουμε τα παρακάτω αποτελέσματα για τις μεταβλητές:

Πίνακας 2 : Έλεγχος Σημαντικότητας Kolmogorov-Smirnov Shapiro-Wilk

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TransfOil	,101	95	,018	,920	95	,000
ImpGas	,117	95	,003	,943	95	,000
ConsOil	,141	95	,000	,937	95	,000

a. Lilliefors Significance Correction

Οι τιμές των p-value για όλες τις μεταβλητές είναι μικρότερες από το επίπεδο σημαντικότητας, $\alpha = 0,05$, σαν αποτέλεσμα δεν επαληθεύεται η μηδενική υπόθεση, ότι δηλαδή τα δεδομένα ακολουθούν κανονική κατανομή.

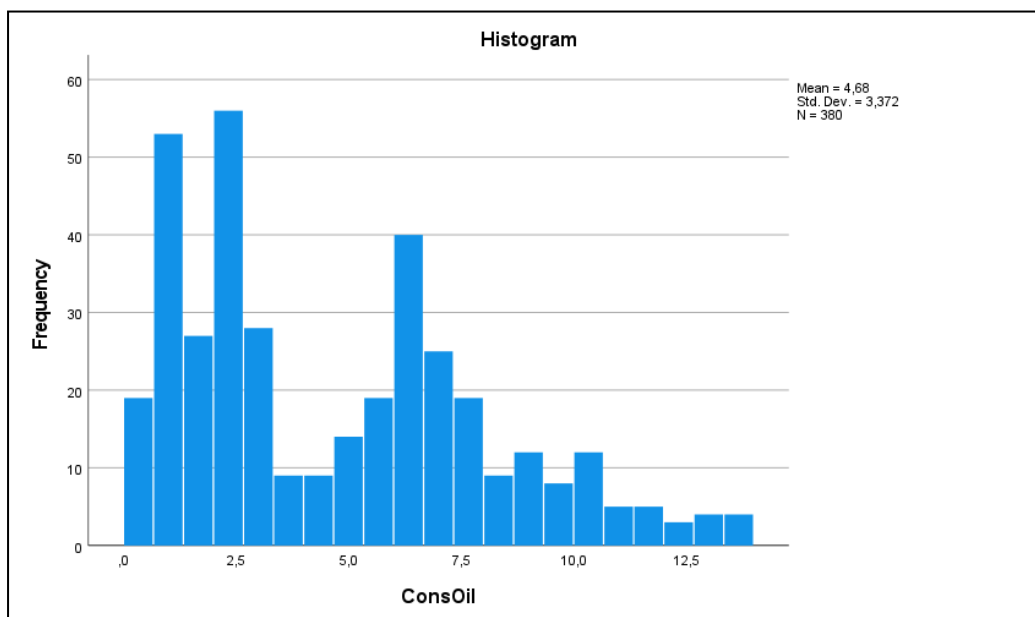
4.4 Δεύτερη Ομάδα χωρών

4.4.1 Μεταβλητή Κατανάλωση Πετρελαιοειδών (ConsOil)

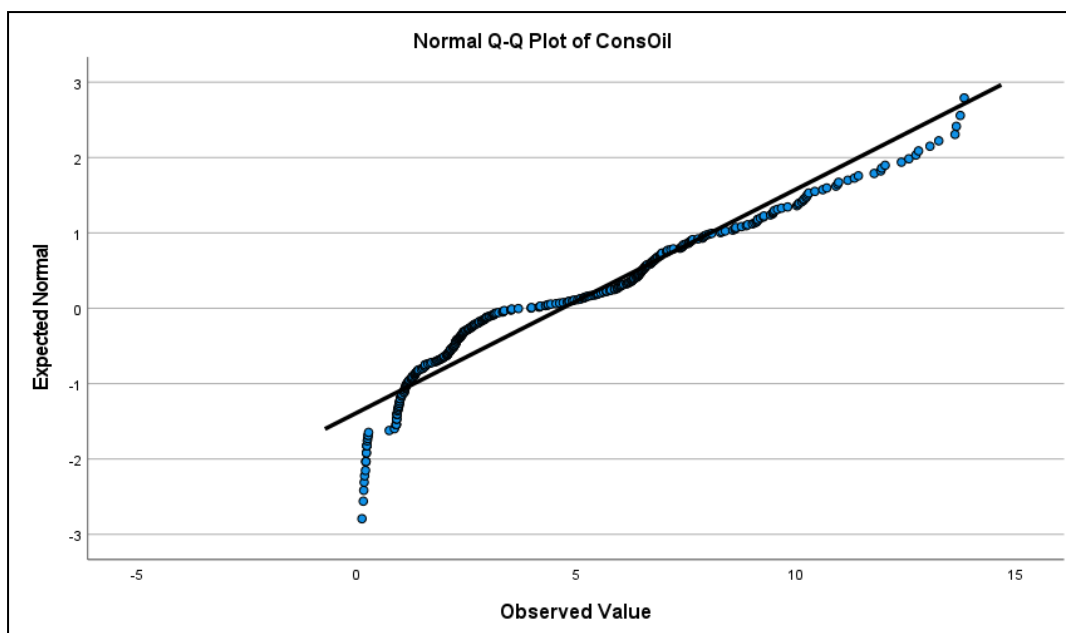
Στατιστικά μέτρα για ConsOil
Αριθμός Παρατηρήσεων:380

Μέσος Όρος:4,68
Διάμεσος:3,84
Διασπορά:11,37
Τυπική Απόκλιση:3,37
Ελάχιστη Τιμή:0
Μέγιστη Τιμή:14
Εύρος:14
Ασυμμετρία:0,6
Κύρτωση:-,04

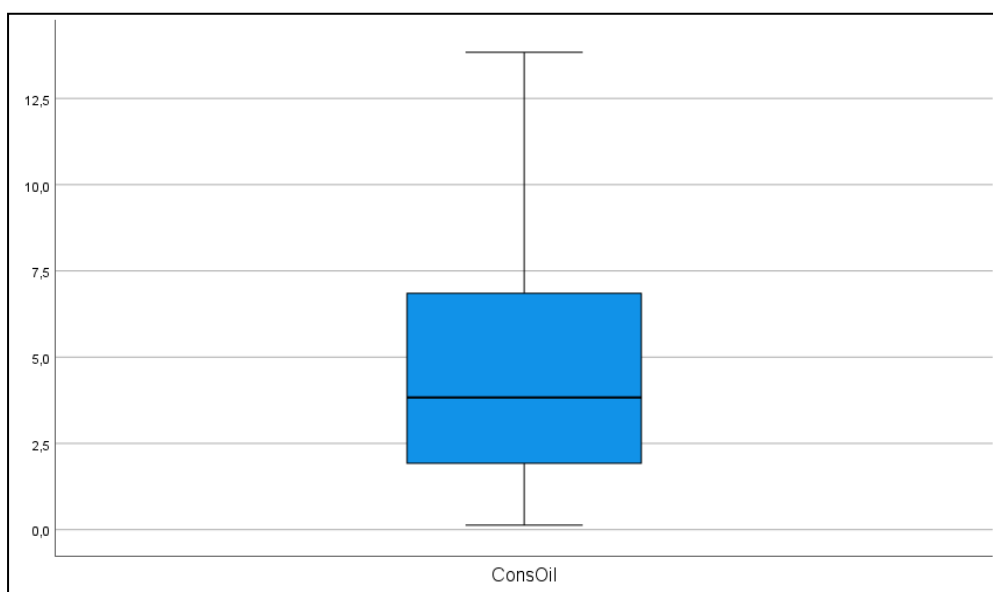
Σε αυτή την περίπτωση το σύνολο των παρατηρήσεων είναι 380. Τα περιγραφικά χαρακτηριστικά της μεταβλητής ConsOil, που αφορά τις υπόλοιπες είκοσι χώρες της Ευρωπαϊκής Ένωσης, φαίνεται να είναι αρκετά ελπιδοφόρα όσον αφορά την κανονικότητα των δεδομένων. Η πρώτη ένδειξη κανονικότητας είναι οι τιμές του μέσου όρου (4,68) και της διάμεσου (3,84), όπου είναι κοντινές. Δεύτερη ένδειξη αποτελούν οι τιμές της κύρτωσης και της ασυμμετρίας των δεδομένων, οι οποίες βρίσκονται κοντά στο μηδέν. Παρότι, οι παραπάνω ενδείξεις ήταν αρκετά ικανοποιητικές για την ύπαρξη κανονικότητας, προχωρώντας στα διαγράμματα, είναι φανερό πως υπάρχουν κάποιες αποκλίσεις από την κανονική κατανομή. Παρατηρείται από το ιστόγραμμα και το θηκόγραμμα η ύπαρξη θετικής ασυμμετρίας και όσον αφορά το διάγραμμα ελέγχου κανονικότητας οι παρατηρήσεις δεν ευθυγραμμίζονται τελείως πάνω στην ευθεία γραμμή.



Διάγραμμα 24 : Ιστόγραμμα για την μεταβλητή ConsOil



Διάγραμμα 25 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή ConsOil



Διάγραμμα 26 : Θηκόγραμμα για την μεταβλητή ConsOil

4.4.2 Μεταβλητή Μετασχηματισμός Πετρελαίου (TransfOil)

Στατιστικά μέτρα για TransfOil

Αριθμός Παρατηρήσεων:380

Μέσος Όρος:8

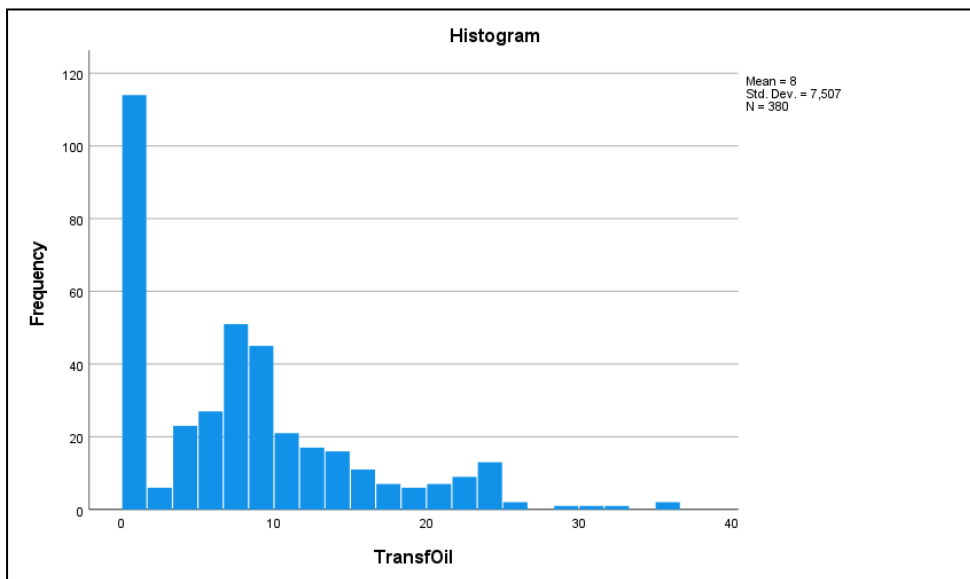
Διάμεσος:7,29

Διασπορά:56,3

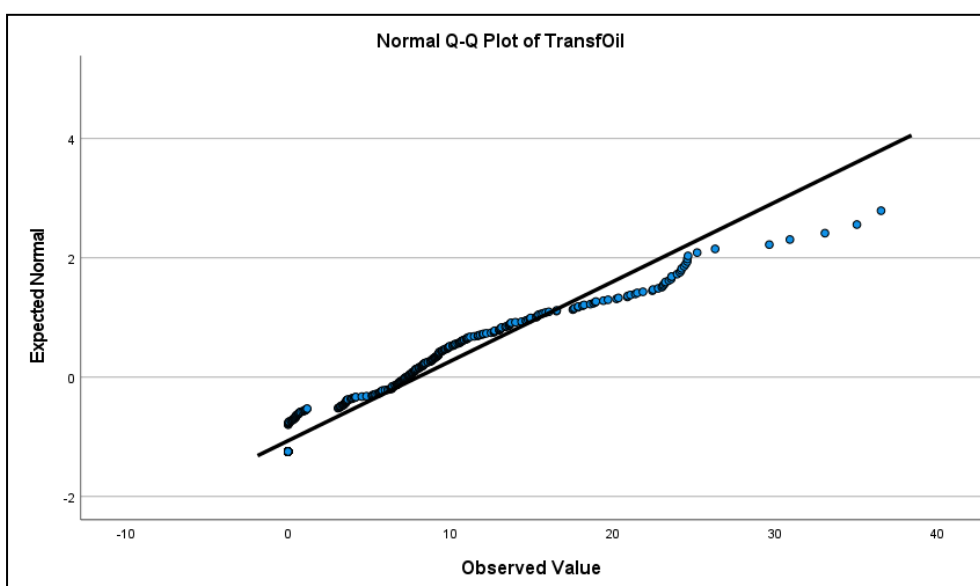
Τυπική Απόκλιση:7,5

Ελάχιστη Τιμή:0
Μέγιστη Τιμή:37
Εύρος:37
Ασυμμετρία:1,005
Κύρτωση:0,7

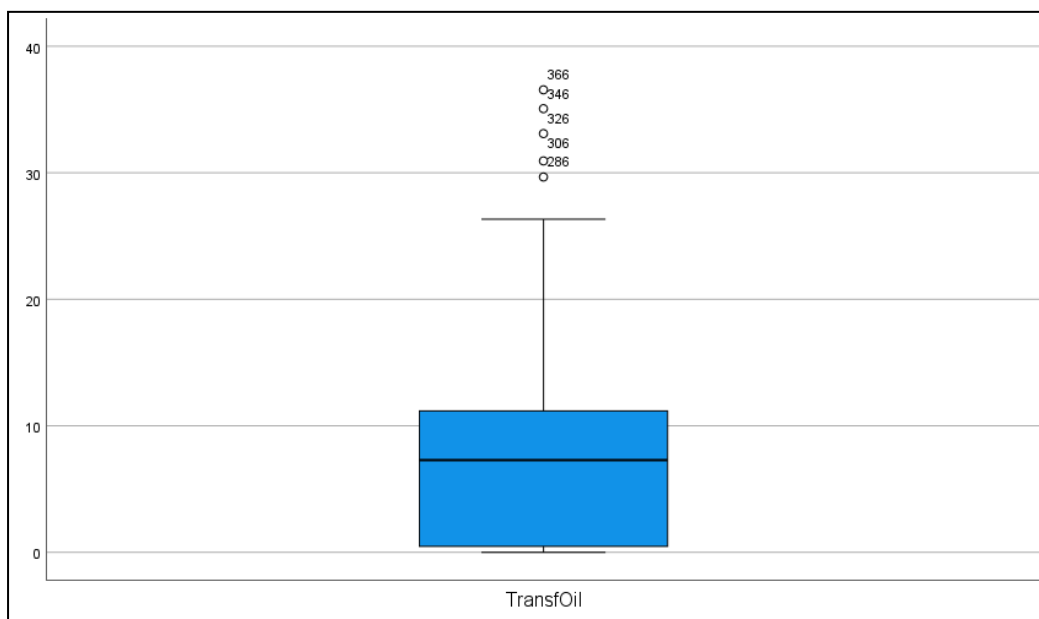
Από τις τιμές της ασυμμετρίας και της κύρτωσης παρατηρείται ότι υπάρχει ένδειξη κανονικότητας. Στο ιστόγραμμα το σχήμα καμπάνας δεν είναι εμφανές και φαίνεται να υπάρχει πρόβλημα με την κύρτωση (λεπτόκυρτη κατανομή) και την ασυμμετρία. Επιπλέον στο διάγραμμα ελέγχου κανονικότητας ενώ τα περισσότερα σημεία ευθυγραμμίζονται με την ευθεία, στο τέλος υπάρχουν κάποιες παρατηρήσεις που ξεφεύγουν. Αυτό εξηγείται από το θηκόγραμμα όπου και φαίνονται ξεκάθαρα κάποιες έκτροπες παρατηρήσεις.



Διάγραμμα 27 : Ιστόγραμμα για την μεταβλητή *TransfOil*



Διάγραμμα 28: Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή *TransfOil*



Διάγραμμα 29 : Θηκόγραμμα για την μεταβλητή *TransfOil*

4.4.3 Μεταβλητή Εισαγωγή Φυσικού Αερίου (ImpGas)

Στατιστικά μέτρα για ImpGas

Αριθμός Παρατηρήσεων:380

Μέσος Όρος:2.43

Διάμεσος:2.02

Διασπορά:6.8

Τυπική Απόκλιση:2.6

Ελάχιστη Τιμή:-5

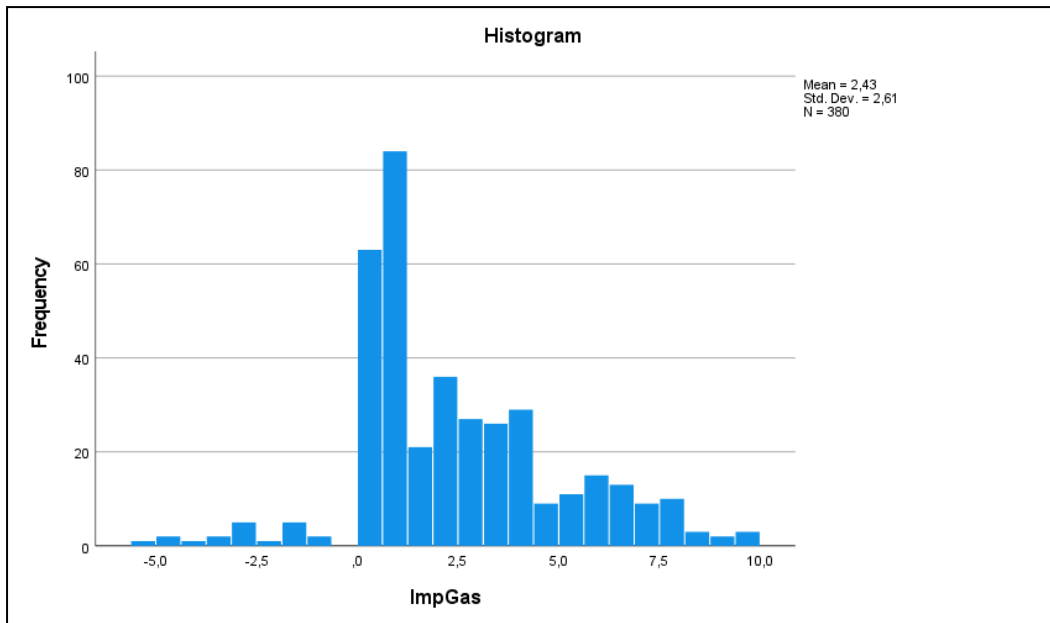
Μέγιστη Τιμή:10

Εύρος:15

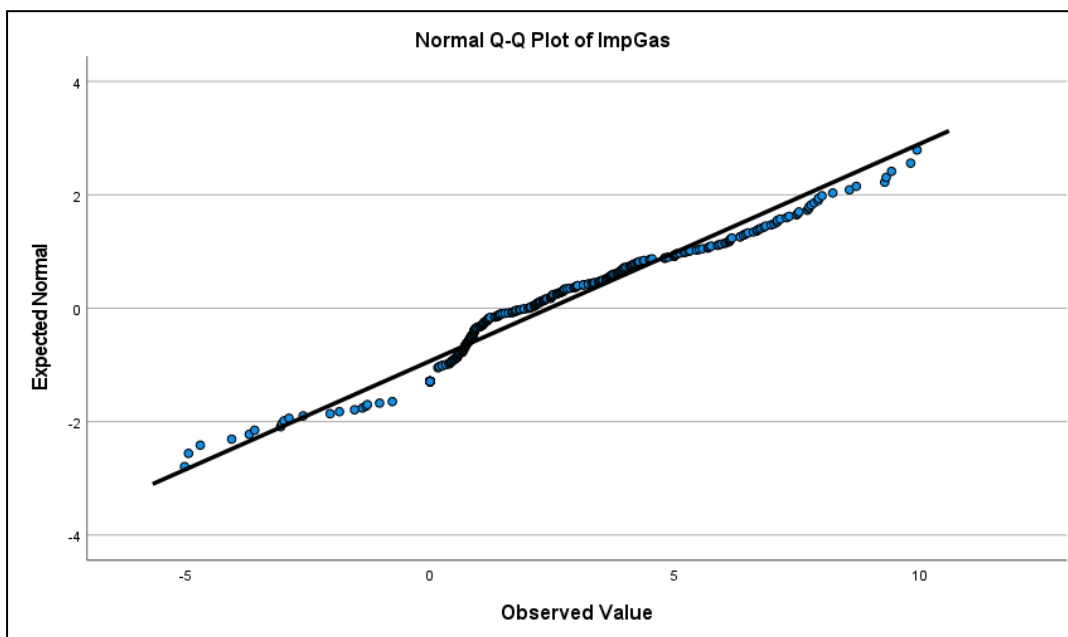
Ασυμμετρία:0.43

Κύρτωση:0.34

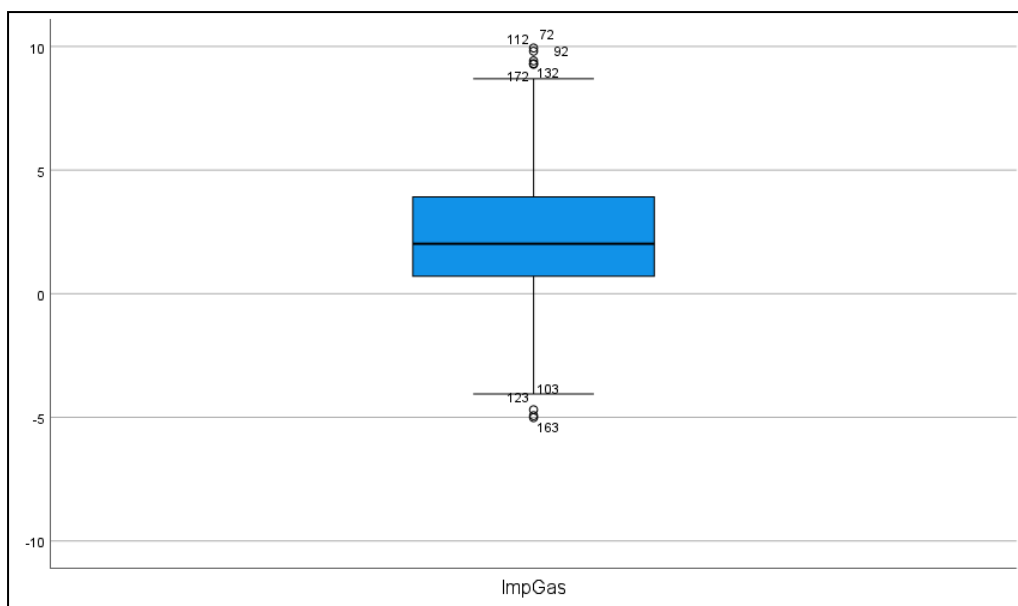
Όπως φαίνεται από τις τιμές της ασυμμετρίας και της κύρτωσης η μεταβλητή ImpGas ακολουθεί κανονική κατανομή, αν και από το ιστόγραμμα παρακάτω φαίνεται να υπάρχει θετική ασυμμετρία. Στο διάγραμμα ελέγχου κανονικότητας τα περισσότερα σημεία ακουμπούν πάνω στην ευθεία γραμμή. Τέλος στο ιστόγραμμα υπάρχουν κάποιες έκτροπες παρατηρήσεις.



Διάγραμμα 30 : Ιστόγραμμα για την μεταβλητή *ImpGas*



Διάγραμμα 31 : Διάγραμμα Ελέγχου Κανονικότητας για την μεταβλητή *ImpGas*



Διάγραμμα 32 : Θηκόγραμμα για την μεταβλητή *ImpGas*

Παρακάτω διεξάγεται έλεγχος κανονικότητας για τις συγκεκριμένες μεταβλητές:

Πίνακας 3 : Έλεγχος Κανονικότητας *Kolmogorov-Smirnov* και *Shapiro-Wilk*

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TransfOil	,143	380	,000	,891	380	,000
ImpGas	,126	380	,000	,955	380	,000
ConsOil	,149	380	,000	,927	380	,000

a. Lilliefors Significance Correction

Όπως φαίνεται από τους παραπάνω ελέγχους το p-value είναι μικρότερο από το επίπεδο σημαντικότητας, $\alpha = 0,05$ (95% διάστημα εμπιστοσύνης), για όλες τις μεταβλητές. Ως αποτέλεσμα δεν μπορεί να επιβεβαιωθεί η μηδενική υπόθεση για την ύπαρξη της κανονικότητας παρόλο που τα περιγραφικά χαρακτηριστικά και κάποια από τα διαγράμματα προσέγγισαν αρκετά την κανονική κατανομή.

4.5 Ανάλυση Παλινδρόμησης: Πρώτη Ομάδα χωρών

Σε αυτή την ενότητα πραγματοποιείται η μέθοδος ανάλυσης παλινδρόμησης για τις μεταβλητές που αφορούν την πρώτη ομάδα χωρών. Παρακάτω περιγράφονται συνοπτικά τα βήματα της ανάλυσης που θα ακολουθήσει:

1. Θα πραγματοποιηθεί έλεγχος συσχετίσεων όλων των εξαρτημένων και ανεξάρτητων μεταβλητών. Οι έλεγχοι θα γίνουν με τους πίνακες συσχετίσεων κατά Pearson και κατά Spearman Rank, ο οποίος είναι λιγότερο ευαίσθητος στις μάκρυνες παρατηρήσεις.

Από τους πίνακες συσχετίσεων ελέγχονται δύο πράγματα. Πρώτον, με ποιες μεταβλητές συσχετίζεται η εξαρτημένη μεταβλητή και δεύτερον εάν υπάρχουν συσχετίσεις μεταξύ των ανεξάρτητων μεταβλητών.

2. Δημιουργούμε τα δύο μοντέλα πολλαπλής παλινδρόμησης. Το πρώτο αφορά τις πέντε χώρες, Ηνωμένο Βασίλειο, Γερμανία, Ισπανία, Ιταλία, Βέλγιο και το δεύτερο τις Βουλγαρία, Τσεχία, Δανία, Εσθονία, Ιρλανδία, Ελλάδα, Κροατία, Κύπρος, Λετονία, Λιθουανία, Ουγγαρία, Μάλτα, Αυστρία, Πολωνία, Πορτογαλία, Ρουμανία, Σλοβενία, Σλοβακία, Φιλανδία και Σουηδία.
3. Στο τελευταίο βήμα της ανάλυσης αλλά και στο πιο σημαντικό γίνεται έλεγχος για την ισχύ των προϋποθέσεων της παλινδρόμησης. Γίνεται έλεγχος για την ύπαρξη κανονικότητας στα κατάλοιπα. Επιπλέον, μέσα από τα διαγράμματα των καταλοίπων, ελέγχεται αν το μέσο σφάλμα είναι μηδέν με σταθερή διασπορά. Τέλος, μέσω διαφόρων ελέγχων, βγαίνει το συμπέρασμα για την τυχαιότητα και την ύπαρξη αυτοσυσχέτισης στα κατάλοιπα.

4.5.1 Έλεγχος Συσχετίσεων Μεταβλητών

Στο σημείο αυτό ελέγχεται ο πίνακας συσχετίσεων κατά Pearson και κατά Spearman Rank, διότι η δεύτερη μέθοδος είναι κατάλληλη σε περιπτώσεις που τα οι μεταβλητές δεν ακολουθούν κανονική κατανομή, όπως φάνηκε στο προηγούμενο κεφάλαιο:

Πίνακας 4 : Πίνακας Συσχετίσεων των μεταβλητών κατά Pearson

		TransfOil	ImpGas	ConsOil
TransfOil	Pearson Correlation	1	,540**	,823**
	Sig. (2-tailed)		,000	,000
	N	95	95	95
ImpGas	Pearson Correlation	,540**	1	,501**
	Sig. (2-tailed)	,000		,000
	N	95	95	95
ConsOil	Pearson Correlation	,823**	,501**	1
	Sig. (2-tailed)	,000	,000	
	N	95	95	95

Πίνακας 5 : Πίνακας Συσχετίσεων κατά Spearman

		TransfOil	ImpGas	ConsOil	
Spearman's rho	TransfOil	Correlation Coefficient	1,000	,602**	,720**
		Sig. (2-tailed)	.	,000	,000
		N	95	95	95
	ImpGas	Correlation Coefficient	,602**	1,000	,567**

	Sig. (2-tailed)	,000	.	,000
	N	95	95	95
ConsOil	Correlation Coefficient	,720**	,567**	1,000
	Sig. (2-tailed)	,000	,000	.
	N	95	95	95

Οι παραπάνω πίνακες δείχνουν τις συσχετίσεις μεταξύ όλων των μεταβλητών. Στην πρώτη γραμμή αναγράφονται οι συντελεστές συσχέτισης. Οι συντελεστές αυτοί μπορούν να πάρουν τιμές από -1 έως $+1$ και μετράνε τον βαθμό συσχέτισης μεταξύ των μεταβλητών. Στη δεύτερη σειρά γράφεται ο αριθμός των ζευγών των τιμών των δεδομένων που χρησιμοποιούνται για τον υπολογισμό των συντελεστών. Ο τρίτος αριθμός που αναγράφεται είναι το $p - value$ και εξετάζει τη στατιστική σημαντικότητα των εκτιμώμενων συντελεστών συσχέτισης. Οι τιμές με $p - value$ μικρότερο από $0,05$, φανερώνουν στατιστικά σημαντικές μη μηδενικές συσχετίσεις σε 95% επίπεδο σημαντικότητας.

Συμπεραίνοντας από τους παραπάνω πίνακες, υπάρχουν ισχυρές συσχετίσεις μεταξύ της εξαρτημένης μεταβλητής ConsOil και των ανεξάρτητων TransfOil και ImpGas. Οι ανεξάρτητες μεταβλητές εμφανίζουν και αυτές μία μικρή συσχέτιση μεταξύ τους. Παρακάτω θα ελεγχθεί το θέμα της πολυσυγγραμμικότητας που αφορά τις ανεξάρτητες μεταβλητές.

4.5.2 Αποτελέσματα Γραμμικής Παλινδρόμησης

Στην ανάλυση παλινδρόμησης μεταξύ της εξαρτημένης μεταβλητής ConsOil και των ανεξάρτητων TransfOil και ImpGas, χρησιμοποιούνται 95 παρατηρήσεις, οι οποίες αντιστοιχούν στα έτη 2000 έως και 2018. Από την επεξεργασία των δεδομένων, προκύπτει το παρακάτω μοντέλο παλινδρόμησης :

$$ConsOil = 11.501 + 0.46TransfOil + 0.055ImpGas$$

Στον Πίνακα 6 παρουσιάζονται τα αποτελέσματα για τον συντελεστή προσδιορισμού του μοντέλου παλινδρόμησης :

Πίνακας 6 : Συντελεστής Προσδιορισμού R

Model Summary					
Model	R	R Square	Adjusted Square	R	Durbin-Watson
1	.826 ^a	.682	.675	.000	2.686

a. Predictors: (Constant), ImpGas, TransfOil
b. Dependent Variable: ConsOil

Παρατηρώντας τα αποτελέσματα του πίνακα φαίνεται πως ο συντελεστής προσδιορισμού R ισούται με 0,68. Αυτό πρακτικά σημαίνει πως το 68% της μεταβλητότητας της εξαρτημένης μεταβλητής ConsOil οφείλεται στις επιδράσεις όλων μαζί των ανεξάρτητων μεταβλητών, TransfOil και ImpGas. Ουσιαστικά ο συντελεστής προσδιορισμού μετράει τη συνολική επίδραση που δέχεται η εξαρτημένη μεταβλητή από τις ανεξάρτητες. Ως αποτέλεσμα των παραπάνω, θα μπορούσε κάποιος να πει πως αν γνωρίζουμε το ποσοστό του μετασχηματισμού πετρελαιοειδών και των εισαγωγών φυσικού αερίου ,στις χώρες της Ευρωπαϊκής Ένωσης, μπορούμε κατά 68% να προσδιορίσουμε την κατανάλωση πετρελαιοειδών.

Επιπροσθέτως, μία ακόμα πληροφορία που δίνει ο Πίνακας 6 είναι ο στατιστικός έλεγχος , Durbin-Watson και αφορά την ύπαρξη αυτοσυσχέτισης στα κατάλοιπα. Ο συγκεκριμένος έλεγχος εξετάζει τη μηδενική υπόθεση πως τα κατάλοιπα δεν έχουν αυτοσυσχέτιση, ενάντια στην εναλλακτική, πως τα κατάλοιπα έχουν υψηλή αυτοσυσχέτιση. Όταν το στατιστικό d ισούται με 2 τότε αποδεικνύεται πως δεν υπάρχει αυτοσυσχέτιση μεταξύ των καταλοίπων. Η τιμή του d κυμαίνεται πάντα μεταξύ 0 και 4. Εάν τα αποτελέσματα του στατιστικού ελέγχου Durbin – Watson είναι μικρότερα από 2, υπάρχουν ενδείξεις θετικής συσχέτισης. Κατά γενικό κανόνα, εάν το Durbin – Watson είναι μικρότερο από 1,0, μπορεί να υπάρχει λόγος ανησυχίας. Εάν το d είναι κατά πολύ μεγαλύτερο του 2 και πλησιάζει το 4, οι διαδοχικοί όροι σφάλματος συσχετίζονται αρνητικά. Στην συγκεκριμένη περίπτωση φαίνεται πως το στατιστικό d ισούται με 2,6. Συμπεραίνουμε λοιπόν πως επειδή ο αριθμός είναι αρκετά κοντά στο 2 επαληθεύεται η μηδενική υπόθεση ότι τα κατάλοιπα δεν έχουν αυτοσυσχέτιση.

Πίνακας 7 : Ανάλυση Διακύμανσης ANOVA

ANOVA^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12622.585	2	6311.293	98.524	.000 ^b
	Residual	5893.376	92	64.058		
	Total	18515.961	94			

a. Dependent Variable: ConsOil
b. Predictors: (Constant), ImpGas, TransfOil

Στον Πίνακα 7 με τίτλο <<ANOVA>> παρουσιάζονται τα αποτελέσματα που προέκυψαν από τον έλεγχο σημαντικότητας μεταξύ των μεταβλητών που έχουν οριστεί. Από την ανάλυση διακύμανσης προκύπτει ότι το $p - value$ είναι μικρότερο από 0,05, επομένως υπάρχει μια στατιστικά σημαντική σχέση ανάμεσα στις ανεξάρτητες μεταβλητές και την εξαρτημένη σε 95% επίπεδο σημαντικότητας.

Πίνακας 8 : Συντελεστές Παλινδρόμησης

Coefficients							
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics
		B	Std. Error	Beta			VIF
1	(Constant)	11.501	3.311		3.474	.000	
	TransfOil	.462	.041	.784	11.211	.000	1.412
	ImpGas	.055	.053	.073	1.051	.296	1.412

Όπως φαίνεται και στον Πίνακα 8 το $p - value$ της μεταβλητής TransfOil είναι μικρότερο από 0,05, άρα η παράμετρος αυτή είναι σημαντικά διαφορετική από το 0, επομένως υπάρχει μία στατιστικά σημαντική πρόβλεψη της εξαρτημένης μεταβλητής με την ανεξάρτητη.

Το αποτέλεσμα αυτό οδηγεί στο συμπέρασμα ότι αν αυξηθεί ο μετασχηματισμός πετρελαιοειδών κατά μία μονάδα και οι υπόλοιπες ερμηνευτικές μεταβλητές παραμείνουν σταθερές, τότε η κατανάλωση πετρελαιοειδών θα αυξηθεί κατά 46% ή 0,46 μονάδες.

Από την άλλη πλευρά το $p - value$ της μεταβλητής ImpGas είναι αρκετά μεγαλύτερο από 0,05, όπου πρακτικά αυτό σημαίνει πως η παράμετρος αυτή δεν είναι σημαντικά διαφορετική από το 0. Ως αποτέλεσμα δεν υπάρχει στατιστικά σημαντική πρόβλεψη της εξαρτημένης μεταβλητής Cons Oil από την ανεξάρτητη.

Επιπλέον, ο Πίνακας 8 δίνει και τον πληθωριστικό παράγοντα διασποράς (VIF). Για τον παράγοντα αυτόν ισχύουν κατά κανόνα τα εξής:

- Αν $VIF_k \cong 1$, η αντίστοιχη μεταβλητή δεν έχει πρόβλημα πολυσυγγραμμικότητας
- Αν $VIF_k > 10$, η αντίστοιχη μεταβλητή εμφανίζει πρόβλημα πολυσυγγραμμικότητας σε σχέση με τις υπόλοιπες ανεξάρτητες μεταβλητές.

Άρα, εφόσον οι τιμές του παράγοντα για την κάθε ανεξάρτητη μεταβλητή είναι μικρότερες του 10 και μάλιστα αρκετά κοντά στη μονάδα, τότε βγαίνει το συμπέρασμα ότι δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας. Ο λόγος που υπάρχουν συσχετίσεις μεταξύ των μεταβλητών, όπως φάνηκε και στον πίνακα συσχετίσεων κατά Spearman, είναι διότι τα δεδομένα συσχετίζονται μεταξύ τους από τη φύση τους.

4.5.3 Ανάλυση Καταλοίπων

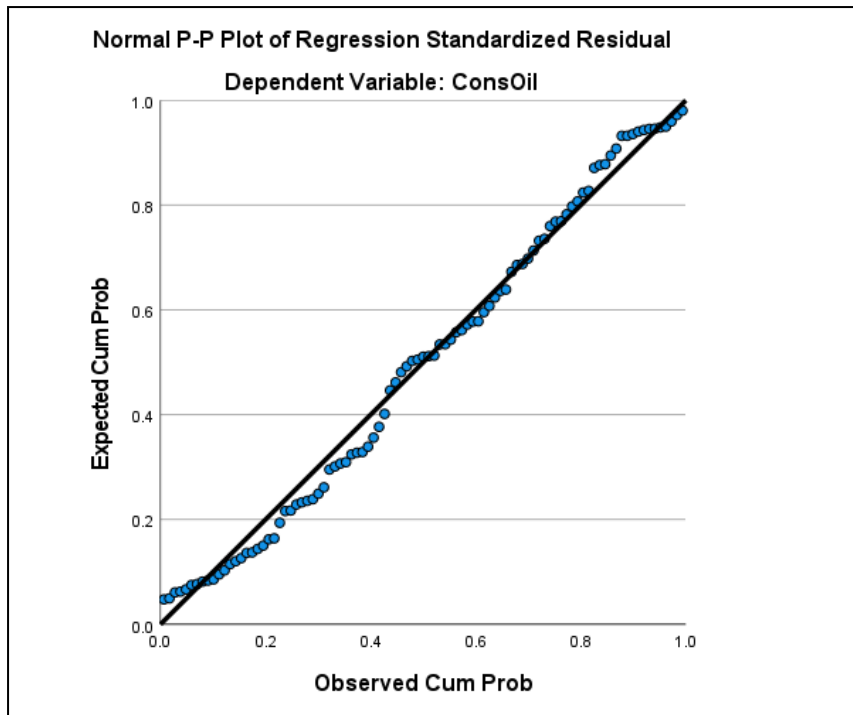
Στην τελευταία αλλά και πιο σημαντική ενότητα για την συγκεκριμένη ανάλυση θα εξεταστεί αν ισχύουν οι προϋποθέσεις της παλινδρόμησης.

Χρησιμοποιώντας το μη παραμετρικό τεστ του Kolmogorov-Smirnov, ελέγχεται η κανονικότητα των καταλοίπων. Όπως φαίνεται παρακάτω η τιμή του $p - value$ είναι 0,2 μεγαλύτερη του 0,05. Προφανώς τα κατάλοιπα ακολουθούν κανονική κατανομή.

Πίνακας 9 : Τεστ Κανονικότητας Καταλοίπων

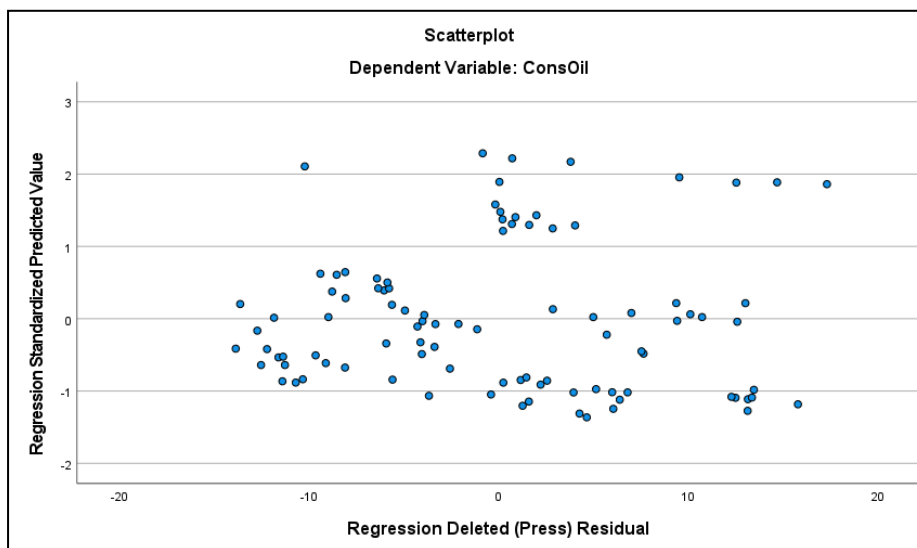
Tests of Normality			
	Kolmogorov-Smirnov ^a		
Standardized Residual	.063	95	.200*
*. This is a lower bound of the true significance.			
a. Lilliefors Significance Correction			

Ακόμα μία ένδειξη κανονικότητας αντικατοπτρίζεται, στο παρακάτω διάγραμμα. Φαίνεται ξεκάθαρα πως όλα τα σημεία "πέφτουν" πάνω στην ευθεία γραμμή, χωρίς κάποια απόκλιση.



Διάγραμμα 33 : Διάγραμμα Ελέγχου Κανονικότητας Καταλοίπων

Παρακάτω απεικονίζεται το διάγραμμα των καταλοίπων (άξονας x) ως προς τις προβλεπόμενες τιμές (άξονας y). Παρατηρείται πως η διασπορά έχει παντού το ίδιο εύρος, συνεπώς δεν υπάρχει το πρόβλημα της ετεροσκεδαστικότητας.



Διάγραμμα 34 : Διάγραμμα Καταλοίπων ως προς τις Προβλεπόμενες Τιμές

Τέλος, ένας επιπλέον μη παραμετρικός έλεγχος που διεξήχθη για να βγει το ασφαλές συμπέρασμα της τυχαιότητας και ανεξαρτησίας των καταλοίπων είναι το Wald–Wolfowitz runs test. Όπως φαίνεται παρακάτω και στις τρεις περιπτώσεις, μέσος, διάμεσος και εύρος το p – $value$ είναι μεγαλύτερο από το 0,05, επομένως επαληθεύεται η μηδενική υπόθεση, ότι δηλαδή τα κατάλοιπα είναι τυχαία και ανεξάρτητα.

Πίνακας 10 : Έλεγχος Τυχαιότητας για τα Κατάλοιπα

Runs Test	
	Standardized Residual
Test Value ^a	,02673
Cases < Test Value	47
Cases >= Test Value	48
Total Cases	95
Number of Runs	57
Z	1,755
Asymp. Sig. (2-tailed)	,079
a. Median	

Runs Test 2	
	Standardized Residual
Test Value ^a	,0000000
Cases < Test Value	45
Cases >= Test Value	50
Total Cases	95
Number of Runs	57
Z	1,786
Asymp. Sig. (2-tailed)	,074
a. Mean	

Runs Test 3	
	Standardized Residual
Test Value ^a	2,06259 ^b
Cases < Test Value	94
Cases >= Test Value	1
Total Cases	95
Number of Runs	3
Z	,147
Asymp. Sig. (2-tailed)	,883
a. Mode	

4.6 Ανάλυση Παλινδρόμησης: Δεύτερη Ομάδα Χωρών

Στην παρακάτω μελέτη, συμμετέχουν όλες οι μεταβλητές που αναλύθηκαν περιγραφικά σε προηγούμενη ενότητα. Οι χώρες οι οποίες αποτελούν τις παρατηρήσεις για τις συγκεκριμένες μεταβλητές είναι οι εξής: Βουλγαρία, Τσεχία, Δανία, Εσθονία, Ιρλανδία, Ελλάδα, Κροατία, Κύπρος, Λετονία, Λιθουανία, Ουγγαρία, Μάλτα, Αυστρία, Πολωνία, Πορτογαλία, Ρουμανία, Σλοβενία, Σλοβακία, Φιλανδία και Σουηδία.

Για την ορθότητα των αποτελεσμάτων και την ισχύ των προϋποθέσεων της παλινδρόμησης κρίθηκε απαραίτητο να γίνουν κάποιοι μετασχηματισμοί στην εξαρτημένη μεταβλητή ConsOil. Ο καταλληλότερος για το συγκεκριμένο μοντέλο παλινδρόμησης είναι ο $\log(\text{ConsOil} + 1)$. Στον Πίνακα 11 παρατίθεται ο συμβολισμός της συγκεκριμένης μεταβλητής.

Πίνακας 11 : Συμβολισμός της Μετασχηματισμένης Εξαρτημένης Μεταβλητής ConsOil

Συμβολισμός	Επεξήγηση
X_1cons	Κατανάλωση Πετρελαιοειδών

4.6.1 Έλεγχος Συσχετίσεων Μεταβλητών

Στο σημείο αυτό ελέγχεται ο πίνακας συσχετίσεων κατά Pearson και κατά Spearman Rank, όπως συνέβη και στην προηγούμενη ανάλυση.

Πίνακας 12 : Πίνακας Συσχετίσεων Μεταβλητών κατά Pearson

Correlations		TransfOil	ImpGas	x_1cons
TransfOil	Pearson Correlation	1	,250**	,755**
	Sig. (2-tailed)		,000	,000
	N	380	380	380
ImpGas	Pearson Correlation	,250**	1	,385**
	Sig. (2-tailed)	,000		,000
	N	380	380	380
x_1cons	Pearson Correlation	,755**	,385**	1
	Sig. (2-tailed)	,000	,000	
	N	380	380	380

** . Correlation is significant at the 0.01 level (2-tailed).

Πίνακας 13 : Πίνακας Συσχετίσεων Μεταβλητών κατά Spearman

			TransfOil	ImpGas	x_1cons
Spearman's rho	TransfOil	Correlation Coefficient	1,000	,457**	,785**
		Sig. (2-tailed)	.	,000	,000
		N	380	380	380
	ImpGas	Correlation Coefficient	,457**	1,000	,591**
		Sig. (2-tailed)	,000	.	,000
		N	380	380	380
	x_1cons	Correlation Coefficient	,785**	,591**	1,000
		Sig. (2-tailed)	,000	,000	.
		N	380	380	380

** . Correlation is significant at the 0.01 level (2-tailed).

Όπως φαίνεται από τους παραπάνω πίνακες η εξαρτημένη μεταβλητή x_1cons είναι ισχυρά συσχετισμένη με τις ανεξάρτητες TransfOil και ImpGas.

4.6.2 Αποτελέσματα Γραμμικής Παλινδρόμησης

Στην ανάλυση παλινδρόμησης μεταξύ της εξαρτημένης μεταβλητής x_1cons και των ανεξάρτητων TransfOil και ImpGas, χρησιμοποιούνται 380 παρατηρήσεις, οι οποίες αντιστοιχούν στα έτη 2000 έως και 2018. Προκύπτει το παρακάτω μοντέλο παλινδρόμησης :

$$x_{1cons} = 0.925 + 0.061TransfOil + 0.053ImpGas$$

Στον Πίνακα 14 παρουσιάζονται τα αποτελέσματα για τον συντελεστή προσδιορισμού του μοντέλου παλινδρόμησης :

Πίνακας 14 : Συντελεστής Προσδιορισμού R

Model Summary ^b								
Model	R	R Square	Adjusted Square	R	Std. Error of the Estimate	Change Statistics		Durbin-Watson
						R Change	Sig. F Change	
1	.782 ^a	.611	.609		.40918	.611	.000	2.374

a. Predictors: (Constant), ImpGas, TransfOil
b. Dependent Variable: x_1cons

Παρατηρώντας τα αποτελέσματα του πίνακα φαίνεται πως ο συντελεστής προσδιορισμού R ισούται με 0,61. Αυτό πρακτικά σημαίνει πως το 61% της μεταβλητότητας της εξαρτημένης μεταβλητής ConsOil οφείλεται στις επιδράσεις όλων μαζί των ανεξάρτητων μεταβλητών , TransfOil και ImpGas. Ως αποτέλεσμα των παραπάνω, θα μπορούσε κάποιος να πει πως αν γνωρίζουμε το ποσοστό του μετασχηματισμού πετρελαιοειδών και των εισαγωγών φυσικού αερίου ,στις χώρες της Ευρωπαϊκής Ένωσης, μπορούμε κατά 61% να προσδιορίσουμε την κατανάλωση πετρελαιοειδών.

Επιπλέον στον στατιστικός έλεγχο Durbin-Watson, φαίνεται πως το στατιστικό d ισούται με 2,3. Συμπεραίνουμε λοιπόν πως επειδή ο αριθμός είναι αρκετά κοντά στο 2 επαληθεύεται η μηδενική υπόθεση ότι τα κατάλοιπα δεν έχουν αυτοσυσχέτιση.

Πίνακας 15 : Ανάλυση Διακύμανσης ANOVA

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	99.105	2	49.553	295.960	.000 ^b
	Residual	63.121	377	.167		
	Total	162.226	379			

a. Dependent Variable: x_1cons
b. Predictors: (Constant), ImpGas, TransfOil

Στον Πίνακα 15 με τίτλο <<ANOVA>> παρουσιάζονται τα αποτελέσματα που προέκυψαν από τον έλεγχο σημαντικότητας μεταξύ των μεταβλητών που έχουν οριστεί. Από την ανάλυση διακύμανσης προκύπτει ότι το $p - value$ είναι μικρότερο από 0,05, επομένως υπάρχει μια στατιστικά σημαντική σχέση ανάμεσα στις μεταβλητές σε 95% επίπεδο σημαντικότητας.

Πίνακας 16 : Συντελεστές Παλινδρόμησης

Coefficients ^a							
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics
		B	Std. Error	Beta			VIF
1	(Constant)	.925	.034		27.272	.000	
	TransfOil	.061	.003	.702	21.170	.000	1.067
	ImpGas	.053	.008	.210	6.322	.000	1.067

a. Dependent Variable: x_1cons

Όπως φαίνεται και στον Πίνακα 16 το $p - value$ της μεταβλητής TransfOil είναι μικρότερο από 0,05, άρα η παράμετρος αυτή είναι σημαντικά διαφορετική από το 0, επομένως υπάρχει μία στατιστικά σημαντική πρόβλεψη της εξαρτημένης μεταβλητής με την ανεξάρτητη.

Το αποτέλεσμα αυτό οδηγεί στο συμπέρασμα ότι αν αυξηθεί ο μετασχηματισμός πετρελαιοειδών κατά μία μονάδα και οι υπόλοιπες ερμηνευτικές μεταβλητές παραμείνουν σταθερές, τότε η κατανάλωση πετρελαιοειδών θα αυξηθεί κατά 6% ή 0,6 μονάδες.

Επιπροσθέτως, το $p - value$ της μεταβλητής ImpGas είναι σε αυτή την περίπτωση, μικρότερο 0,05, όπου πρακτικά αυτό σημαίνει πως η παράμετρος είναι σημαντικά διαφορετική από το 0, άρα υπάρχει στατιστικά σημαντική πρόβλεψη της εξαρτημένης μεταβλητής ConsOil από την ανεξάρτητη.

Λόγω των παραπάνω αποτελεσμάτων, αν αυξηθεί η εισαγωγή φυσικού αερίου στις χώρες της Ευρωπαϊκής Ένωσης κατά μία μονάδα και οι υπόλοιπες ερμηνευτικές μεταβλητές παραμείνουν σταθερές, τότε η κατανάλωση πετρελαιοειδών θα αυξηθεί κατά 5% ή 0,5 μονάδες.

Επιπλέον, ο Πίνακας 16 δίνει και τον πληθωριστικό παράγοντα διασποράς (VIF) με τιμή 1,067 και για τις δύο ανεξάρτητες μεταβλητές.

Άρα, εφόσον οι τιμές του παράγοντα για την κάθε ανεξάρτητη μεταβλητή είναι μικρότερες του 10 και μάλιστα αρκετά κοντά στη μονάδα, τότε βγαίνει το συμπέρασμα ότι δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας.

4.6.3 Ανάλυση Καταλοίπων

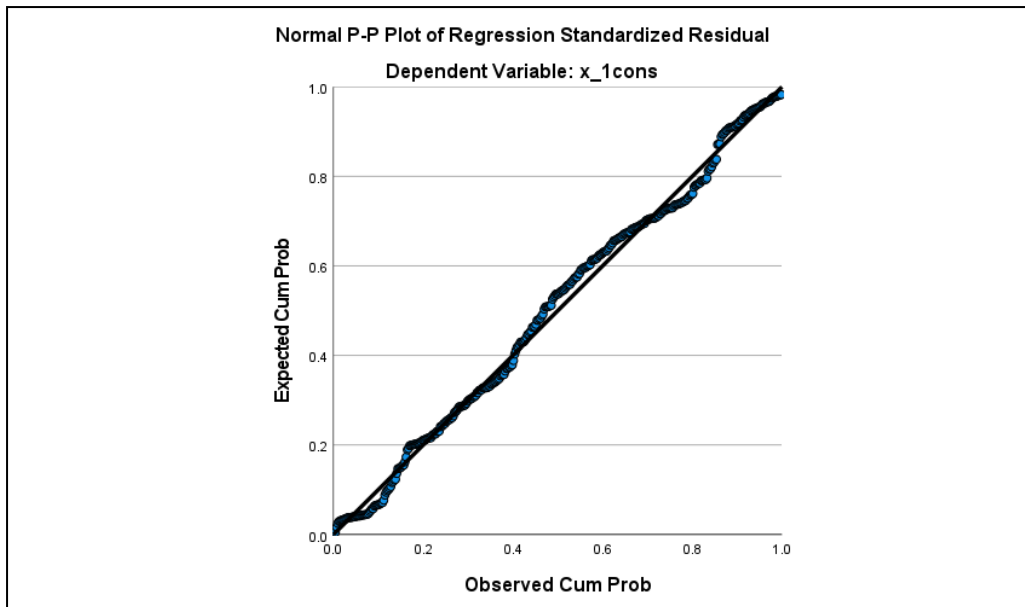
Στην συγκεκριμένη περίπτωση εξετάζονται ξανά τα περιγραφικά χαρακτηριστικά των καταλοίπων του μοντέλου παλινδρόμησης.

Χρησιμοποιώντας το μη παραμετρικό τεστ του Kolmogorov-Smirnov, ελέγχεται η κανονικότητα των καταλοίπων. Όπως φαίνεται παρακάτω η τιμή του $p - value$ είναι 0,09, μεγαλύτερη του 0,05. Προφανώς τα κατάλοιπα ακολουθούν κανονική κατανομή.

Πίνακας 17 : Τεστ Κανονικότητας Καταλοίπων

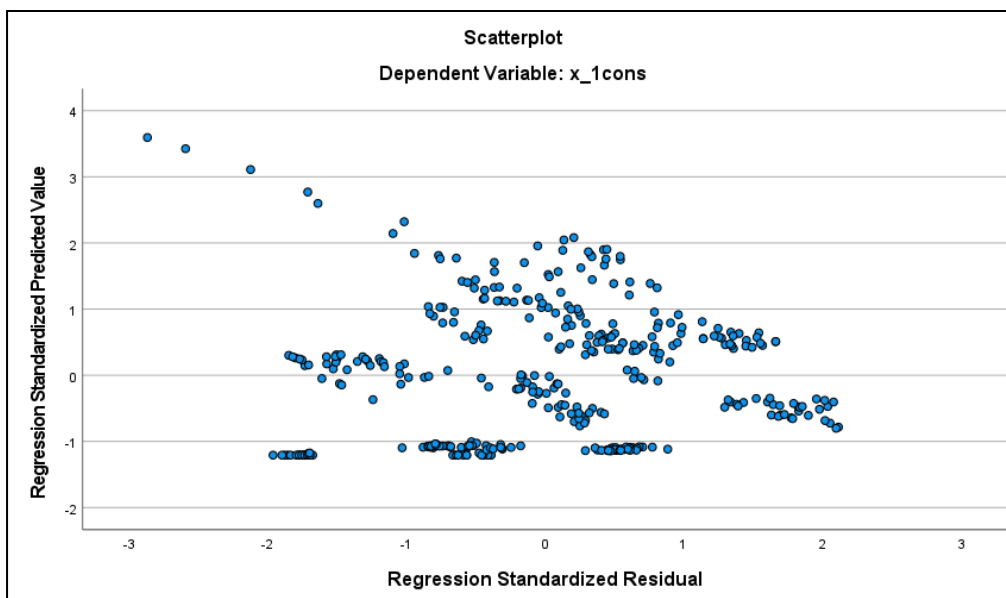
Tests of Normality			
	Kolmogorov-Smirnov ^a		
Standardized Residual	.043	380	.095
a. Lilliefors Significance Correction			

Επιπλέον, παρακάτω φαίνεται το διάγραμμα ελέγχου κανονικότητας των καταλοίπων, όπου είναι ξεκάθαρο πως όλες οι παρατηρήσεις εφάπτονται ακριβώς πάνω στην ευθεία γραμμή.



Διάγραμμα 35 : Διάγραμμα Ελέγχου Κανονικότητας Καταλοίπων

Στο παρακάτω διάγραμμα παρατηρείται πως τα σημεία δημιουργούν το λεγόμενο “νέφος”, παρόλα αυτά υπάρχουν κάποιες έκτροπες παρατηρήσεις και σε κάποια σημεία φαίνεται να σχηματίζεται μοτίβο. Για αυτό τον λόγο προχωράμε και σε περαιτέρω έλεγχο της τυχαιότητας και ανεξαρτησίας των καταλοίπων.



Διάγραμμα 36 : Διάγραμμα καταλοίπων ως προς τις Προβλεπόμενες Τιμές

Τέλος, ο έλεγχος που διεξήχθη για να βγει το ασφαλές συμπέρασμα της τυχαιότητας και ανεξαρτησίας των καταλοίπων είναι το Wald–Wolfowitz runs test. Όπως φαίνεται παρακάτω και στις τρεις περιπτώσεις, μέσος, διάμεσος και εύρος το $p - value$ είναι μεγαλύτερο από το 0,05, επομένως επαληθεύεται η μηδενική υπόθεση, ότι δηλαδή τα κατάλοιπα είναι τυχαία και ανεξάρτητα.

Πίνακας 18 : Έλεγχος Τυχαιότητας για τα Κατάλοιπα

Runs Test	
	x_1cons
Test Value ^a	1,58
Cases < Test Value	190
Cases >= Test Value	190
Total Cases	380
Number of Runs	190
Z	-,103
Asymp. Sig. (2-tailed)	,918
a. Median	

Runs Test 2	
	x_1cons
Test Value ^a	1,5417
Cases < Test Value	189
Cases >= Test Value	191
Total Cases	380
Number of Runs	188
Z	-,308
Asymp. Sig. (2-tailed)	,758
a. Mean	

Runs Test 3	
	x_1cons
Test Value ^a	1,18
Cases < Test Value	119
Cases >= Test Value	261
Total Cases	380
Number of Runs	141
Z	-2,804
Asymp. Sig. (2-tailed)	,051
a. Mode	

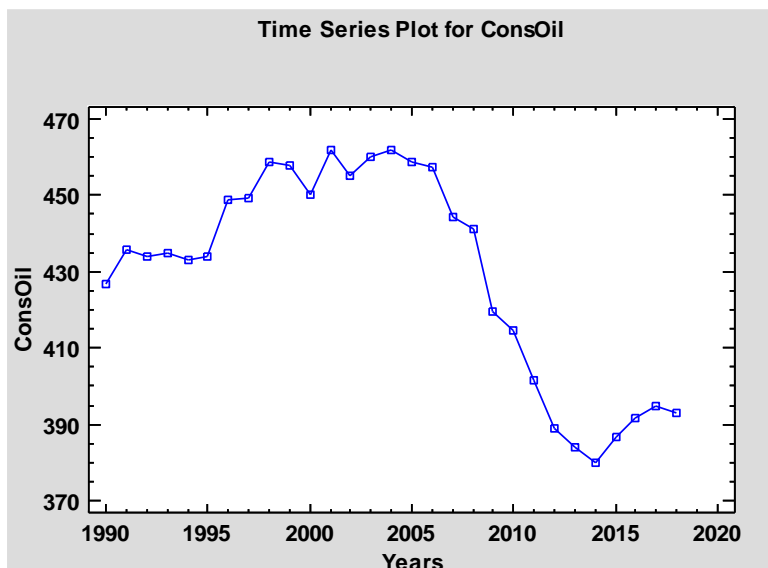
4.7 Προβλέψεις Κατανάλωσης Πετρελαιοειδών

Στη συγκεκριμένη ενότητα, γίνεται προσπάθεια πρόβλεψης της κατανάλωσης των πετρελαιοειδών στην Ευρωπαϊκή Ένωση. Οι παρατηρήσεις οι οποίες συλλέχτηκαν γι' αυτό τον σκοπό αφορούν τα έτη 1990 έως 2018. Το στατιστικό πακέτο το οποίο τα χρησιμοποιηθεί για την επεξεργασία των δεδομένων ονομάζεται Statgraphics.

Η ανάλυση και η επεξεργασία των χρονοσειρών θα γίνει με τη χρήση Box and Jenkins ARIMA modeling μεθοδολογίας. Η τεχνική ARIMA δίνει τη μορφή γενικευμένου υποδείγματος σε μία διακριτή χρονική σειρά.

4.7.1 Έλεγχος Στασιμότητας

Όπως αναφέρθηκε και στο θεωρητικό μέρος, προτού προχωρήσει κανείς στην συγκεκριμένη ανάλυση είναι απαραίτητο να εξεταστεί η χρονοσειρά για την ύπαρξη στασιμότητάς. Το διάγραμμα παρακάτω περιγράφει την κατανάλωση των πετρελαιοειδών σε συνάρτηση με το χρόνο.

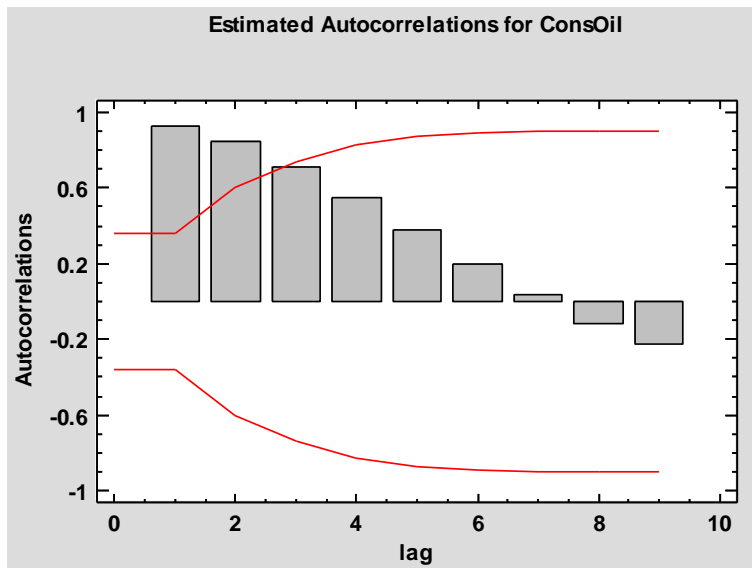


Διάγραμμα 37 : Διάγραμμα Χρονοσειρών για την μεταβλητή *ConsOil*

Το συγκεκριμένο διάγραμμα δίνει αρκετές πληροφορίες για την συμπεριφορά της κατανάλωσης των πετρελαιοειδών στην Ευρωπαϊκή Ένωση. Αρχικά αυτό που παρατηρείται είναι μια ανοδική πορεία με κάποιες αυξομειώσεις από το 1990 μέχρι το 2005. Στη συνέχεια, φαίνεται μία ραγδαία μείωση στη κατανάλωση με αποκορύφωμα το 2014 όπου έχει και την χαμηλότερη τιμή της, μέσα σε αυτά τα 28 χρόνια που μελετώνται. Η συγκεκριμένη μείωση πιθανόν οφείλεται στην οικονομική κρίση που ξεκίνησε το 2008. Τέλος το 2015, η κατανάλωση αρχίζει να παίρνει ανοδική πορεία, κρατώντας βέβαια αρκετά χαμηλά νούμερα σε σχέση με αυτά του 1990.

Από την όψη του διαγράμματος παρατηρείται πως δεν υπάρχει στασιμότητα. Είναι φανερό πως η μέση τιμή είναι διαφορετική από χρονιά σε χρονιά και η διακύμανση δεν είναι σταθερή.

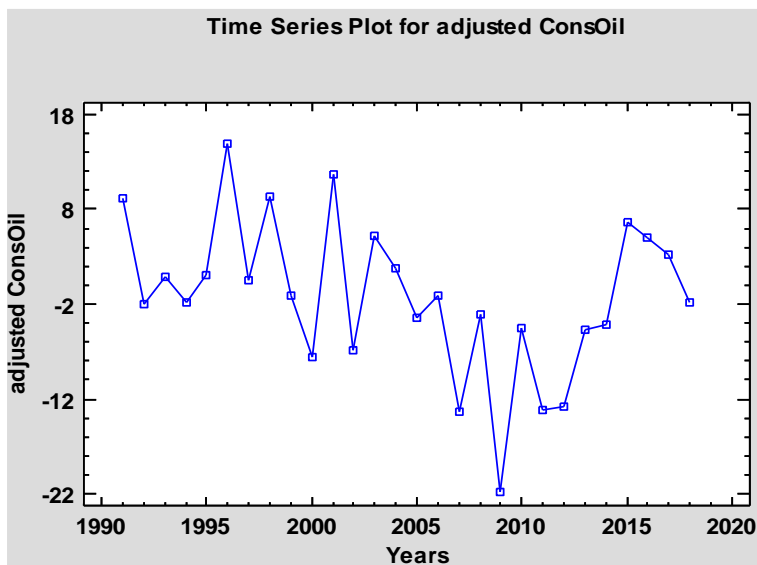
Παρακάτω δίνεται το διάγραμμα αυτοσυσχέτισης, το οποίο περιγράφει τη σχέση που υπάρχει ανάμεσα στην μεταβλητή *ConsOil* και την μεταβλητή $ConsOil_{t-1}$. Ουσιαστικά, το ιδανικό αποτέλεσμα θα ήταν οι χρονουστερήσεις να βρίσκονται εντός των ορίων, κάτι που δεν υποδεικνύει καμία συσχέτιση μεταξύ των δύο παραπάνω μεταβλητών.



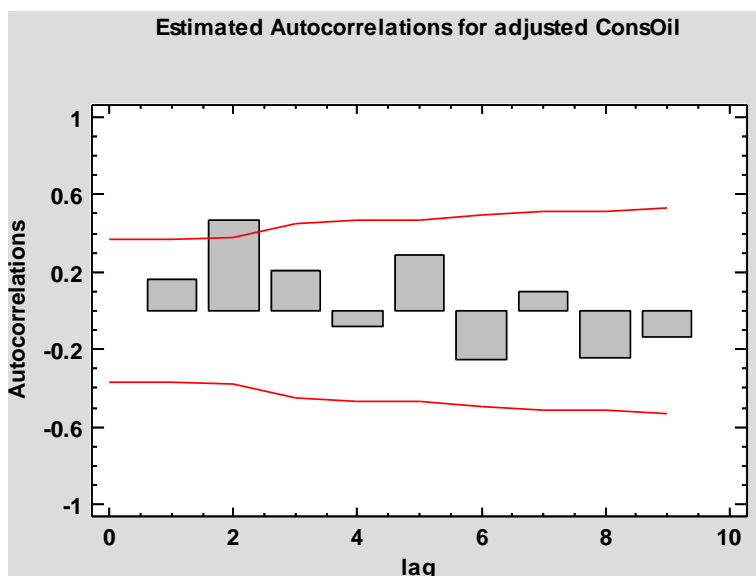
Διάγραμμα 38 : Αυτοσυσχέτιση των μεταβλητών $ConsOil$ και $ConsOil_{t-1}$.

Όπως φαίνεται οι χρονουστερήσεις βρίσκονται αρκετά πάνω από τα όρια κάτι που δείχνει ισχυρή συσχέτιση.

Για να λυθεί το πρόβλημα της μη στασιμότητας, γίνεται μία φορά διαφόριση και αφαιρείται από τη χρονοσειρά η τάση όπου φαίνεται να υπάρχει στο Διάγραμμα 38. Παρακάτω φαίνονται τα αντίστοιχα διαγράμματα:



Διάγραμμα 39 : Διάγραμμα Χρονοσειρών για την μεταβλητή $ConsOil$ μετά από μία Διαφόριση



Διάγραμμα 40 : Αυτοσυσχέτιση των μεταβλητών *ConsOil* και $ConsOil_{t-1}$ μετά από μία Διαφόριση

Συμπεραίνοντας, αφαιρώντας την τάση από τη χρονοσειρά τα διαγράμματα έχουν την επιθυμητή μορφή. Στο Διάγραμμα 39 φαίνεται ότι υπάρχει στασιμότητα, αν και η μείωση της κατανάλωσης το 2010 είναι εμφανή και εδώ. Στο Διάγραμμα 40 μόνο ένα σημείο ξεπερνά τα όρια, επομένως δεν υπάρχουν σοβαρές συσχετίσεις μεταξύ της μεταβλητής και των παρελθοντικών της τιμών.

4.7.2 Πρόβλεψη με τη μέθοδο ARIMA

Σύμφωνα με το λογισμικό πακέτο Statgraphics παρακάτω φαίνονται τα πιθανά μοντέλα πρόβλεψης:

- (A) Random walk
- (B) Random walk with drift = -1.19357
- (C) Constant mean = 429.619
- (D) Linear trend = 461.864 + -2.14967 t
- (E) Simple moving average of 2 terms
- (F) Simple exponential smoothing with alpha = 0.9946
- (G) Brown's linear exp. smoothing with alpha = 0.6206
- (H) Holt's linear exp. smoothing with alpha = 0.9999 and beta = 0.0242
- (I) ARIMA (0,1,2)
- (J) ARIMA (1,1,2)
- (K) ARIMA (2,1,2)

(L) ARIMA (2,1,0)

(M) ARIMA (2,2,2)

Για να προκριθεί το κατάλληλο μοντέλο είναι απαραίτητο να πληρούνται κάποια κριτήρια. Αρχικά, για κάθε ένα από τα υποψήφια μοντέλα υπολογίζεται το αντίστοιχο AIC και επιλέγεται αυτό με τη μικρότερη αλγεβρικά τιμή. Αυτό συμβαίνει διότι ο συγκεκριμένος δείκτης παρουσιάζει πόση πληροφορία έχει χαθεί σε κάθε πιθανό μοντέλο. Συνεπώς επιλέγεται αυτό με την μικρότερη δυνατή απώλεια.

Πίνακας 19 : Υπολογισμός AIC για Υποψήφια Μοντέλα

Model	RMSE	MAE	MAPE	ME	MPE	AIC	HQC
(A)	8.09649	6.17929	1.4501	-1.19357	-0.309305	4.18286	4.18286
(B)	8.15497	6.02929	1.41182	2.03012E-15	-0.030343	4.26622	4.28099
(C)	27.8394	23.7133	5.63945	2.50895E-13	-0.421561	6.72187	6.73663
(D)	21.3611	18.3094	4.25095	2.39134E-13	-0.234037	6.26107	6.29061
(E)	9.9162	7.52593	1.78537	-2.16296	-0.554895	4.65731	4.67207
(F)	8.10499	5.97797	1.40309	-1.16005	-0.300621	4.25393	4.26869
(G)	7.78627	6.0521	1.4161	0.405723	0.122058	4.17369	4.18846
(H)	8.22422	5.89135	1.37882	-0.271782	-0.0849394	4.3521	4.38163
(I)	5.86962	4.58327	1.06896	-0.844862	-0.204157	3.67751	3.70704
(J)	5.91706	4.48229	1.04302	-0.823442	-0.197801	3.76258	3.80688
(K)	6.01757	4.58924	1.06644	-0.772772	-0.182309	3.86523	3.92429
(L)	7.17619	5.16846	1.20514	-0.54427	-0.124721	4.07947	4.109
(M)	6.83243	5.21496	1.22289	-0.369332	-0.0622959	4.11922	4.17829

Παρατηρείται παραπάνω πως το μοντέλο με τη μικρότερη αλγεβρικά τιμή είναι το (I), ARIMA (0,1,2). Στη διαδικασία επιλογής του καταλληλότερου μοντέλου, λαμβάνονται υπόψη και άλλα κριτήρια. Παρακάτω φαίνονται τα αποτελέσματα πέντε ελέγχων ροών των καταλοίπων που παρέχει το στατιστικό πακέτο με τις επωνυμίες RUNS, RUNM, AUTO, MEAN, VAR.

Πίνακας 20 : Έλεγχος Ροών Καταλοίπων

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	8.09649	OK	OK	OK	*	OK
(B)	8.15497	OK	OK	OK	*	OK
(C)	27.8394	**	***	***	**	**
(D)	21.3611	***	***	***	OK	OK

(E)	9.9162	OK	OK	***	*	OK
(F)	8.10499	OK	OK	OK	*	OK
(G)	7.78627	OK	OK	OK	OK	OK
(H)	8.22422	OK	OK	OK	*	OK
(I)	5.86962	OK	OK	OK	OK	OK
(J)	5.91706	OK	OK	OK	OK	OK
(K)	6.01757	OK	OK	OK	OK	OK
(L)	7.17619	OK	OK	OK	OK	OK
(M)	6.83243	OK	OK	OK	OK	OK

Είναι απαραίτητο για την πρόκριση του μοντέλου, να υπάρχει η ένδειξη "OK" σε όλους τους ελέγχους που έχουν τεθεί. Αυτή η ένδειξη συνεπάγεται πως το μοντέλο περνάει επιτυχώς τον αντίστοιχο έλεγχο.

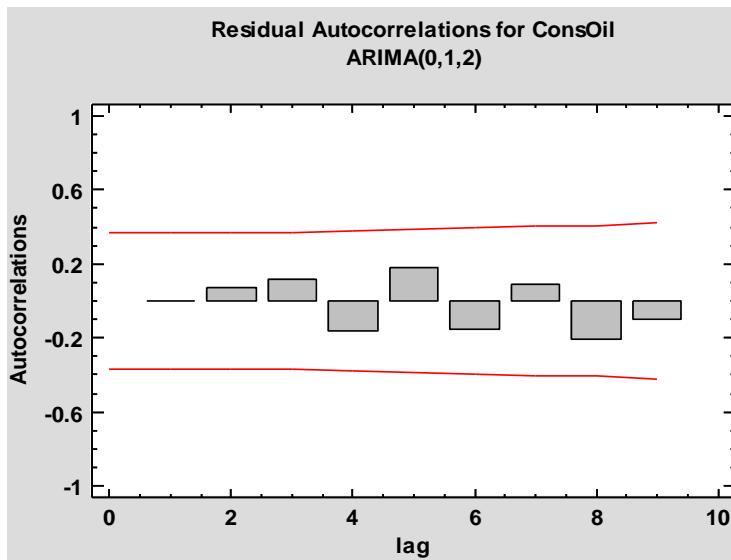
Αν υπάρχει η ένδειξη "***" αυτό σημαίνει ότι ο έλεγχος αποτυγχάνει στο 95% επίπεδο εμπιστοσύνης.

Αν υπάρχει η ένδειξη "****" αυτό σημαίνει ότι ο έλεγχος αποτυγχάνει στο 99% επίπεδο εμπιστοσύνης.

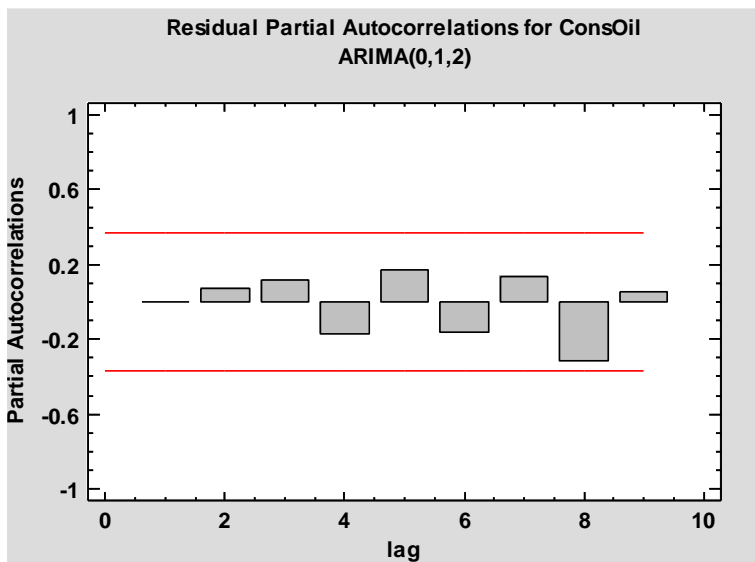
Αν υπάρχει η ένδειξη "*****" αυτό σημαίνει ότι ο έλεγχος αποτυγχάνει στο 99.9% επίπεδο εμπιστοσύνης.

Όπως φαίνεται στον Πίνακα 20 υποψήφιο μοντέλο έχει την ένδειξη "OK" και στους πέντε ελέγχους. Αυτό σημαίνει πως το μοντέλο προσαρμόζεται επαρκώς στα δεδομένα και εφόσον έχει την μικρότερη τιμή AIC, όπως προαναφέρθηκε, προκρίνεται και είναι αυτό που θα κάνει την πρόβλεψη της αντίστοιχης χρονοσειράς.

Τέλος, το τελευταίο κριτήριο που θα πρέπει να ληφθεί υπόψιν είναι τα διαγράμματα αυτοσυσχετίσεων (Residuals Autocorrelation) και μερικών αυτοσυσχετίσεων (Residuals Partial Autocorrelation). Τα κατάλοιπα στα διαγράμματα αυτά πρέπει να βρίσκονται εντός ορίων. Παρακάτω φαίνονται τα συγκεκριμένα διαγράμματα.



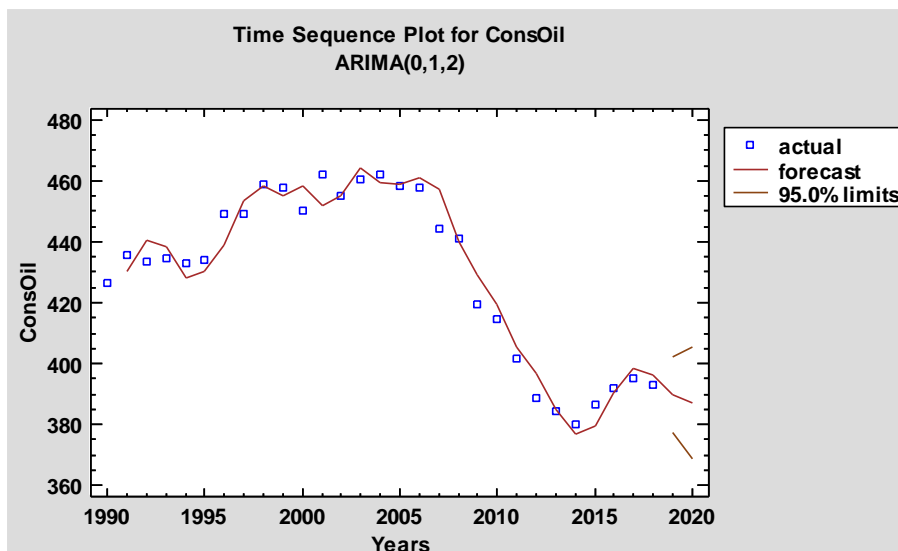
Διάγραμμα 41 : Κατάλοιπα Αυτοσυσχετίσεων



Διάγραμμα 42 : Κατάλοιπα Μερικών Αυτοσυσχετίσεων

Στα παραπάνω διαγράμματα φαίνεται ξεκάθαρα πως τα κατάλοιπα είναι εντός ορίων και για τα δύο διαγράμματα. Το μοντέλο λοιπόν πληροί όλες τις προδιαγραφές και κρίνεται αποδεκτό για την πρόβλεψη της χρονοσειράς.

Οι προβλέψεις που εξάγονται από το μοντέλο ARIMA (0,1,2), καθώς και τα διαστήματα εμπιστοσύνης για τα επόμενα δύο χρόνια από την τελευταία παρατήρηση παρουσιάζονται παρακάτω:



Διάγραμμα 43 : Προβλέψεις Κατανάλωσης Πετρελαιοειδών για δύο χρόνια

Πίνακας 21 : Προβλέψεις Κατανάλωσης Πετρελαιοειδών για δύο χρόνια

		<i>Lower 95%</i>	<i>Upper 95%</i>
<i>Period</i>	<i>Forecast</i>	<i>Limit</i>	<i>Limit</i>
2019	389.797	377.495	402.099
2020	387.063	368.91	405.216

Παρατηρείται από τα παραπάνω αποτελέσματα πως η κατανάλωση πετρελαιοειδών συνεχίζει να μειώνεται και να φτάνει το 2019 την τιμή 389,797 *Mtoe* (εκατομμύρια τόνους πετρελαίου) και το 2020 την ακόμα χαμηλότερη τιμή 387,063 *Mtoe*. Αξίζει να σημειωθεί πως το πετρέλαιο είναι ένα αγαθό το οποίο καθορίζει γενικότερα την οικονομία και οι επιχειρήσεις που επιβιώνουν από την διανομή και την επεξεργασία του είναι πάρα πολλές. Το μέλλον για τους συγκεκριμένους κλάδους, αν και εφόσον η κατανάλωση των πετρελαιοειδών συνεχίσει να είναι φθίνουσα προμηνύεται αβέβαιο και φυσικά ανησυχητικό.

Κεφάλαιο 5: Συμπεράσματα-Προτάσεις για Μελλοντική Έρευνα

5.1 Συμπεράσματα

Στο παρόν κεφάλαιο παρουσιάζονται τα πιο σημαντικά συμπεράσματα της παρούσας εργασίας. Αρχικά, τα δεδομένα που χρησιμοποιήθηκαν για την ανάλυση και των δύο ερευνητικών ερωτημάτων, προήλθαν από το Ενεργειακό Ισοζύγιο όλων των Ευρωπαϊκών χωρών. Το Ενεργειακό Ισοζύγιο, δημιουργήθηκε από την Ευρωπαϊκή Επιτροπή με σκοπό την παρακολούθηση των ενεργειακών προϊόντων και τη ροή τους στην οικονομία. Τα δεδομένα περιγράφουν την παραγωγή, τον μετασχηματισμό και τις μορφές κατανάλωσης των πετρελαιοειδών και άλλων καυσίμων σε 28 κράτη-μέλη της Ευρωπαϊκής Ένωσης.

Το πρώτο ερώτημα που ερευνήθηκε και έγινε προσπάθεια απάντησης του, ήταν *πως επιδρά η παραγωγή, ο μετασχηματισμός και οι εισαγωγές φυσικού αερίου και πετρελαίου στην τελική ενεργειακή κατανάλωση πετρελαιοειδών στις χώρες της Ευρωπαϊκής Ένωσης*. Το στατιστικό πακέτο το οποίο έχει χρησιμοποιηθεί για την επεξεργασία των δεδομένων ονομάζεται SPSS και τα δεδομένα αφορούν τα έτη 2000 έως 2018.

Αναλύοντας τις μεταβλητές και παραθέτοντας τα περιγραφικά χαρακτηριστικά τους, παρατηρήθηκε πως ήταν απαραίτητο να δημιουργηθούν δύο ομάδες χωρών και να μελετηθούν ξεχωριστά. Η ενέργεια αυτή πραγματοποιήθηκε διότι κάποιες χώρες παρουσίαζαν παρόμοια συμπεριφορά στην κατανάλωση πετρελαιοειδών συγκριτικά με άλλες. Η μία ομάδα αφορούσε τις εξής χώρες: Βέλγιο, Ισπανία, Γερμανία, Ιταλία, Γερμανία και η δεύτερη τις Βουλγαρία, Τσεχία, Δανία, Εσθονία, Ιρλανδία, Ελλάδα, Κροατία, Κύπρος, Λετονία, Λιθουανία, Ουγγαρία, Μάλτα, Αυστρία, Πολωνία, Πορτογαλία, Ρουμανία, Σλοβενία, Σλοβακία, Φιλανδία και Σουηδία.

Στη συνέχεια, αποδείχθηκε πως η παραγωγή πετρελαιοειδών κάθε χώρας δεν έχει καμία σχέση με την κατανάλωση τους. Αυτό συμβαίνει διότι οι χώρες της Ευρωπαϊκής Ένωσης κυρίως εισάγουν το αργό πετρέλαιο με τον σημαντικότερο προμηθευτή να είναι η Ρωσία. Έπειτα, από διαγράμματα συσχέτισης που δημιουργήθηκαν, παρατηρήθηκε πως η μεταβλητή TransfOil (μετασχηματισμός του αργού πετρελαίου σε καταναλώσιμα αγαθά, όπως η βενζίνη), επηρεάζει αρκετά την κατανάλωση πετρελαιοειδών. Αξίζει να σημειωθεί πως η συγκεκριμένη μεταβλητή είχε θετική γραμμική συσχέτιση με την ConsOil. Το ίδιο συνέβη και με την μεταβλητή ImpGas (εισαγωγή φυσικού αερίου), η οποία αποδείχθηκε πως έχει θετική γραμμική συσχέτιση με την κατανάλωση πετρελαιοειδών. Επιπλέον οι εισαγωγές πετρελαίου στην Ευρωπαϊκή Ένωση, είχαν αρκετά μεγάλη και θετική συσχέτιση με την κατανάλωση πετρελαιοειδών. Τα συμπεράσματα αυτά αφορούν και τις δύο ομάδες χωρών.

Για την ορθότητα των αποτελεσμάτων, χρησιμοποιήθηκε η μέθοδος της βηματικής παλινδρόμησης (stepwise regression).

Από την παραπάνω μέθοδο, προέκυψε το συμπέρασμα πως οι μεταβλητές οι οποίες συνεισφέρουν περισσότερο στη μείωση της ανερμήνευτης μεταβλητότητας της εξαρτημένης μεταβλητής *ConsOil*, ήταν οι *TransfOil* και *ImpGas*.

Παρακάτω φαίνονται τα μοντέλα παλινδρόμησης και για τις δύο περιπτώσεις:

Πρώτη Ομάδα Χωρών

$$ConsOil = 11.501 + 0.46TransfOil + 0.055ImpGas$$

Δεύτερη Ομάδα Χωρών

$$x_{1cons} = 0.925 + 0.061TransfOil + 0.053ImpGas$$

Εξίσου σημαντικό είναι να αναφερθεί, είναι πως ο συντελεστής προσδιορισμού για την πρώτη ομάδα και για την δεύτερη ομάδα χωρών είναι 0,68 και 0,61 αντίστοιχα. Συμπερασματικά, εξηγώντας πάνω από το 50% της μεταβλητότητας της εξαρτημένης μεταβλητής ConsOil, σε πραγματικά δεδομένα, είναι αρκετά επιτυχές.

Το δεύτερο ερευνητικό ερώτημα, το οποίο προέκυψε διερευνώντας το πρώτο, ήταν η πρόβλεψη της κατανάλωσης πετρελαιοειδών στην Ευρωπαϊκή Ένωση. Οι παρατηρήσεις, αφορούν τα έτη 1990 έως 2018. Το στατιστικό πακέτο το οποίο έχει χρησιμοποιηθεί για την επεξεργασία των δεδομένων ονομάζεται Statgraphics.

Αρχικά, δημιουργήθηκε ένα διάγραμμα ιστορικότητας της μεταβλητής ConsOil. Τα συμπεράσματα που προέκυψαν, ήταν πως η κατανάλωση πετρελαίου στην Ευρωπαϊκή Ένωση ξεκίνησε με σταθερή αύξουσα πορεία και στη συνέχεια, άρχισε να φθίνει. Συγκεκριμένα, αυτό που παρατηρήθηκε, είναι μια ανοδική πορεία με κάποιες αυξομειώσεις από το 1990 μέχρι το 2005. Στη συνέχεια, φαίνεται μία ραγδαία μείωση στη κατανάλωση με αποκορύφωμα το 2014 όπου έχει και την χαμηλότερη τιμή της, μέσα σε αυτά τα 28 χρόνια που μελετώνται. Η συγκεκριμένη μείωση πιθανόν οφείλεται στην οικονομική κρίση που ξεκίνησε το 2008. Τέλος το 2015, η κατανάλωση αρχίζει να παίρνει ανοδική πορεία, κρατώντας βέβαια αρκετά χαμηλά νούμερα σε σχέση με αυτά του 1990.

Χρησιμοποιώντας, την Box and Jenkins ARIMA modeling μεθοδολογία, έγινε προσπάθεια πρόβλεψης για δύο επόμενα έτη, 2019 και 2020. Τα αποτελέσματα από την συγκεκριμένη πρόβλεψη, έδειξαν πως η κατανάλωση των πετρελαιοειδών θα συνεχίσει να μειώνεται. Παρακάτω στον Πίνακα 21 φαίνονται οι τιμές που προβλέπεται πως θα φτάσει.

Πίνακας 21: Προβλέψεις Κατανάλωσης Πετρελαιοειδών για δύο χρόνια

		<i>Lower 95%</i>	<i>Upper 95%</i>
<i>Period</i>	<i>Forecast</i>	<i>Limit</i>	<i>Limit</i>
2019	389.797	377.495	402.099
2020	387.063	368.91	405.216

Η ανάλυση της παρούσας εργασίας μπορεί να δώσει αρκετές πληροφορίες σχετικά με τους τρόπους που επηρεάζεται η τελική κατανάλωση πετρελαιοειδών στην Ευρωπαϊκή Ένωση. Συγκεκριμένα, συμπεραίνουμε πως οι εισαγωγές φυσικού αερίου, καθώς και ο μετασχηματισμός αργού πετρελαίου σε καταναλώσιμα προϊόντα, παίζουν καθοριστικό ρόλο

στην κατανάλωση πετρελαιοειδών από τους τελικούς χρήστες. Κάτι που είναι απόλυτα λογικό διότι, αν δεν υπάρχουν προϊόντα διαθέσιμα στην αγορά, τα οποία δημιουργούνται από πετρέλαιο, προφανώς δεν θα μπορούν να καταναλωθούν.

5.2 Προτάσεις για Μελλοντική Έρευνα

Στην περίπτωση της ανάλυσης παλινδρόμησης, θα μπορούσαν να συμπεριληφθούν επιπλέον μεταβλητές που επηρεάζουν την κατανάλωση πετρελαίου. Αρχικά θα μπορούσε να μελετηθεί ο τρόπος που επηρεάζεται η κατανάλωση του πετρελαίου από την ύπαρξη των βιοκαυσίμων (biofuels), τα οποία τα τελευταία χρόνια χρησιμοποιούνται από το ευρύ κοινό σαν εναλλακτικές μορφές ενέργειας, φιλικές προς το περιβάλλον. Επιπροσθέτως, σίγουρα θα ήταν σημαντικό να εξεταστεί η παραγωγή του πετρελαίου από τον Οργανισμό Εξαγωγών Πετρελαιοπαραγωγών Χωρών, και αν είναι τελικά αρκετή για να καλύψει την παγκόσμια ζήτηση που συνεχώς αυξάνεται.

Όσον αφορά την περίπτωση της οικονομετρικής διερεύνησης μέσω της μεθόδου Box-Jenkins, οι προτάσεις για επέκταση της μελέτης είναι να εξεταστούν με τον ίδιο τρόπο, η κατανάλωση του φυσικού αερίου και άλλων εναλλακτικών καυσίμων, τα οποία όπως φαίνεται κατακτούν όλο και περισσότερο έδαφος.

BIBΛΙΟΓΡΑΦΙΑ

Βιβλία

- Anderson, D, Sweeney, D & Williams, (2010), *Statistics for Business and Economics*, (11th edition), Ohio: Southwestern College Publication.
- Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2011). *Basic Business Statistics—Concepts and Applications (12nd Edition)*. Prentice Hall.
- Davis, R. A., & Brockwell, P. J. (2016). *Introduction to Time Series and Forecasting* (3rd edition). Springer International Publishing.
- Denis Cook, R., & Weisberg, S. (28/07/1994). *An Introduction to Regression Graphics*. John Wiley and Sons.
- George, D & Malley, (2005), *SPSS for windows step by step*, Boston: Allyn and Bacon.
- Hamilton, J. D. (1994). *Time Series Analysis* (1st edition). Princeton University Press.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd edition). Otexts.
- Lind, D., Marchal, W., & Wathen, S. (2018). *Basic Statistics for Business and Economics* (9th edition). McGraw-Hill Education.
- Longnecker, M. T., & Lyman Ott, R. (2000). *An Introduction to Statistical Methods and Data Analysis* (5th edition). Duxbury Press.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability and Statistics)* (1st edition). Wiley.
- Montgomery, D. C., Vining, G. G., & Peck, E. A. (2012). *Introduction to Linear Regression Analysis* (5th edition). Wiley.
- Pallant, J (2010), *SPSS survival manual*, (4th edition), Buckingham: Open University Press.

Άρθρα

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research* 3, 58–67.

Definition of the Revised APEC Energy Data Reporting Format. Eurostat.

Energy balance guide, Methodology guide for the construction of energy balances & Operational guide for the energy balance builder tool. (Ιανουάριος, 2019). Eurostat.

Energy data. (2020). Eurostat.

European Union, ENERGY IN FIGURES. (2019). Eurostat.