

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2022

Use case 1: Μελέτη Αγροτικής Οικονομίας των ΗΠΑ

Ονοματεπώνυμο: Βασίλειος Ηλίας Δρούζας

Αριθμός Μητρώου: 3180051

Διδάσκοντες: Ιωάννης Ντζούφρας, Ξανθή Πεντελή

Μάιος 2022

Εργαλεία: R, Rstudio, Latex

Περιεχόμενα

1	Εισαγωγή και περιγραφή της μελέτης περίπτωσης: Η αγροτική οικονομία των ΗΠΑ	2
2	Πρώτη γνωριμία με τα δεδομένα- Έλεγχος κανονικότητας	3
2.1	Πρώτη γνωριμία	3
2.2	Έλεγχος κανονικότητας	4
3	Σχέσεις μεταβλητών	5
3.1	Επιλογή ελέγχου	5
3.2	Σχέσεις μεταβλητών ανά δύο	5
3.3	Ποιοι παράγοντες καθορίζουν τη τιμή του μοσχαρίσιου και χοιρινού κρέατος	6
4	Η επίδραση του χρόνου και άλλων μεταβλητών στις τιμές	8
4.1	Η επίδραση του χρόνου	8
4.2	Η επίδραση άλλων μεταβλητών	9
5	Υπάρχουν σημαντικές διαφορές στους μέσους ή διαμέσους των περιόδων 1925-1930,1931-1935,1936-1941;	10
5.1	Μέσοι	10
5.2	Διάμεσοι	11
6	Συμπεράσματα ανάλυσης	12
7	Παράρτημα	13

1 Εισαγωγή και περιγραφή της μελέτης περίπτωσης: Η αγροτική οικονομία των ΗΠΑ

Στην εποχή μας, η επιστήμη και η τεχνολογία έχουν έρθει για να δώσουν λύσεις και να διευκολύνουν τη ζωή του ανθρώπου. Η επιστήμη της στατιστικής, μαζί με την ανάλυση, μπορούν να δώσουν φως και να εξάγουν συμπεράσματα βάσει των παρατηρούμενων δεδομένων. Η περίοδος 1925-1941 για τις ΗΠΑ βρήκε τον αμερικανικό λαό σε πολλές διαφορετικές καταστάσεις. Η "Great Depression", που προέκυψε το 1929 και κράτησε επί μία δεκαετία ύστερα από μια μεγάλη πτώση στις τιμές των αποθεμάτων προϊόντων στις ΗΠΑ (λόγω του "κραχ" της Wall street) έφερε σημαντική οικονομική ύφεση. Επίσης, οι ΗΠΑ είχαν εμπλοκή στον Β' παγκόσμιο πόλεμο (1939-1945), παράγοντες που επηρέασαν σαφώς και την αγροτική οικονομία της χώρας.

Στην παρούσα ανάλυση θα εξετάσουμε τη κατανάλωση μοσχαρίσιου και χοιρινού κρέατος στις ΗΠΑ τη χρονική περίοδο 1925-1941 και τη σχέση μεταξύ της κατανάλωσης-τιμής καθώς και υπόλοιπων παραμέτρων, οι οποίες αναφέρονται παρακάτω.

Οι μεταβλητές λοιπόν που θα εξεταστούν είναι οι ακόλουθες:

- **PBE** : Η τιμή του μοσχαρίσιου κρέατος (cents/lb) , όπου lb: λίβρα
- **CBE** : Η κατανάλωση μοσχαρίσιου κρέατος ανά άτομο (lbs)
- **PPO** : Η τιμή του χοιρινού κρέατος (cents/lb)
- **CPO** : Η κατανάλωση χοιρινού κρέατος ανά άτομο (lbs)
- **PFO** : Δείκτης τιμών λιανικής πώλησης (1947-1949=100)
- **DINC** : Δείκτης διαθέσιμου εισοδήματος ανά άτομο (Disposable income per capita index, 1947-1949=100)
- **CFO** : Δείκτης κατανάλωσης τροφίμων ανά άτομο (1947-1949=100)
- **RDINC** : Δείκτης πραγματικού διαθέσιμου εισοδήματος ανά άτομο (1947-1949=100)
- **RFP** : Δείκτης τιμών τροφίμων λαμβάνοντας υπόψη και το CPI (1947-1949=100)

Τα δεδομένα που έχουμε δεν είναι πολλά, είναι οι τιμές των μεταβλητών που αναφέραμε παραπάνω για το χρονικό διάστημα 1925-1941. Παίρνοντας τα δεδομένα από το σύνδεσμο στο eclass, τα περνάμε σε ένα csv αρχείο και το εισάγουμε στο Rstudio για να ξεκινήσουμε την ανάλυση.

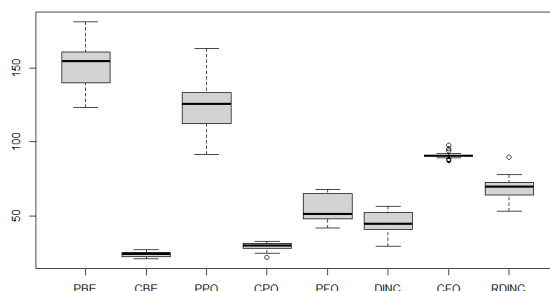
Από το Output 1 του παραρτήματος βλέπουμε ότι όλες οι τιμές λαμβάνουν συνεχείς τιμές, όπως επίσης ότι τα δεδομένα είναι ποσοτικά χαρακτηριστικά (τιμή, κατανάλωση κρέατος και δείκτες τιμών/εισοδήματος/κατανάλωσης).

2 Πρώτη γνωριμία με τα δεδομένα- Έλεγχος κανονικότητας

Πριν ξεκινήσουμε την ανάλυση, μετατρέπουμε τις τιμές σε cents/kg και τη κατανάλωση σε κιλά. Η ισοδυναμία είναι $1 \text{ lb} = 0.4536 \text{ kg}$ και $1 \text{ cent/lb} = 2.2046 \text{ cents/kg}$. Επομένως, οι μεταβλητές PBE και PPO πολ/ζονται με 2.2046 και οι CBE, CPO με 0.4536.

2.1 Πρώτη γνωριμία

Πρώτα καλό είναι να επισκοπήσουμε τα δεδομένα μας, να δούμε αν υπάρχουν ατυπικές τιμές (outliers). Μπορούμε να το πετύχουμε αυτό μέσω των boxplots:

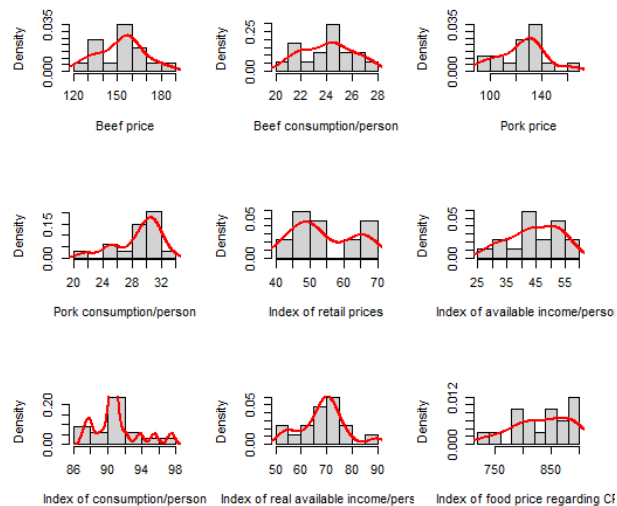


Σχήμα 1: Πλάι-πλάι boxplots των μεταβλητών (πλην της RFP)

Αφαιρέσαμε από το παραπάνω διάγραμμα το boxplot της RFP, διότι οι τιμές της ήταν πολύ αυξημένες σε σχέση με τις υπόλοιπες μεταβλητές και θα μας δυσχέραινε την οπτική. Το boxplot για την μεταβλητή αυτή βρίσκεται στο Output 4. Οι μεταβλητές CFO, RDINC και CPO, όπως βλέπουμε στο σχήμα, παρουσιάζουν μερικές ατυπικές τιμές (η CFO κυρίως). Στους ελέγχους που θα κάνουμε στη συνέχεια δεν προβλέπεται να μας επηρεάσουν κάπου αρνητικά. Στο Output 2 παρουσιάζονται και βασικά περιγραφικά μέτρα για όλες τις μεταβλητές.

2.2 Έλεγχος κανονικότητας

Στη συνέχεια θα κάνουμε έλεγχο κανονικότητας του δείγματος. Αρχικά παρουσιάζουμε τα ιστογράμματα όλων των μεταβλητών:



Σχήμα 1: Ιστογράμματα κανονικότητας των μεταβλητών του δείγματος

Οι περισσότερες μεταβλητές φαίνεται να ακολουθούν την κανονικότητα. Για να έχουμε μια πιο ξεκάθαρη εικόνα, μπορούμε να εφαρμόσουμε έλεγχο κανονικότητας του Shapiro στις μεταβλητές, όπου σε επίπεδο σημαντικότητας 5% μόνο οι μεταβλητές CPO,PFO περιέχουν μη κανονικοποιημένα δεδομένα.

```
shapiro-wilk normality test
data: data$PFO
w = 0.87368, p-value = 0.02515

shapiro-wilk normality test
data: data$CPO
w = 0.88831, p-value = 0.04346
```

Πίνακας 1: Έλεγχος Shapiro-wilk για τις μεταβλητές PFO,CPO.

3 Σχέσεις μεταβλητών

Επειδή λοιπόν δεν είναι όλα τα δεδομένα μας κανονικοποιημένα, δε μπορούμε να χρησιμοποιήσουμε τον παραμετρικό έλεγχο μέσω του συντελεστή συσχέτισης του Pearson. Αντί αυτού, θα χρησιμοποιήσουμε τους ελέγχους των Kendall και Spearman (μη παραμετρικοί).

Στον παρακάτω πίνακα του Kendall, μπορούμε να επισκοπήσουμε τις σχέσεις μεταξύ των μεταβλητών. Οι τιμές του δείκτη συσχέτισης κυμαίνονται στο [-1,1]. Τιμές πιο κοντά στο -1 και το 1 δηλώνουν τόσο πιο ισχυρή σχέση (αρνητική και θετική αντίστοιχα) , ενώ τιμές κοντά στο 0 δηλώνουν απουσία συσχέτισης μεταξύ των μεταβλητών. Τα αστεράκια δηλώνουν και αυτά την ισχύ της σχέσης των μεταβλητών (όσο περισσότερα, τόσο πιο ισχυρή η σχέση).



- Η μεταβλητή Year, δεν έχει ιδιαίτερα υψηλή σχέση με άλλη μεταβλητή, όμως η σχέση της με τις PFO,RFP δεν είναι καθόλου αμελητέα.
- Η μεταβλητή RDINC παρουσιάζει υψηλή θετική συσχέτιση με την DINC (αναμενόμενο, αφού εκφράζουν παρόμοια πράγματα) και με την CFO (κατανάλωση τροφίμων ανά άτομο). Δε μας κάνει εντύπωση, αφού αυτό σημαίνει ότι όσο πιο πλούσιο/φτωχό είναι ένα άτομο, τόσο περισσότερο/λιγότερο αντίστοιχα έχει την (οικονομική) δυνατότητα να αγοράσει κρέας.
- Η μεταβλητή CBE παρουσιάζει υψηλή συσχέτιση με την PBE(προφανές, η τιμή του μοσχαρίσιου κρέατος επηρεάζει τη κατανάλωση του). Αντίστοιχα και οι μεταβλητές PPO,CPO για το χοιρινό κρέας.
- Η μεταβλητή CPO παρουσιάζει υψηλή συσχέτιση με τις DINC,PBE.Εύλογα, καθώς όσο μεγαλύτερο το εισόδημα, τόσο μεγαλύτερη και η ευχέρεια για κατανάλωση. Αντίστοιχα , όσο μικρότερη η τιμή, τόσο μεγαλύτερη πάλι η καταναλωτική ευχέρεια.

Προς επιβεβαίωση των παραπάνω, τρέξαμε Chi-squared ελέγχους για τις σχέσεις των παραπάνω μεταβλητών. Τα αποτελέσματα βρίσκονται στο Output 30.

Μια εναλλακτική του Kendall για μη παραμετρικούς ελέγχους, όπως προαναφέραμε, είναι ο δείκτης συσχέτισης του Spearman. Το Output 14 παρουσιάζει τη μορφή του πίνακα αυτού, ο οποίος συμφωνεί σε μεγάλο βαθμό με τον πίνακα του Kendall.

Συνολικά μπορούμε να πούμε ότι ναι μεν υπάρχει συσχέτιση μεταξύ μερισμών μεταβλητών που προαναφέραμε, όμως ίσως μια γενική εικόνα είναι ότι οι περισσότερες μεταβλητές δεν παρουσιάζουν μεγάλη συσχέτιση μεταξύ τους.

3.3 Ποιοι παράγοντες καθορίζουν τη τιμή του μοσχαρίσιου και χοιρινού κρέατος

Παρακάτω εφαρμόζουμε ένα μοντέλο γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή τη τιμή του μοσχαρίσιου και αρχικά με όλες τις υπόλοιπες μεταβλητές. Θα εφαρμόσουμε τη διαδικασία stepwise selection (forward και backward) για την επιλογή των μεταβλητών. Η πρώτη μεταβλητή που αφαιρεί ο έλεγχος είναι η RFP. Στη συνέχεια, ο έλεγχος δεν αφαιρεί κάποια (πρόβλημα: μας αφήνει με σχεδόν όλες τις μεταβλητές, ενώ το AIC score από το επόμενο μοντέλο μεγαλώνει και παρατηρούμε ότι υπάρχει πρόβλημα πολυσυγγραμμικότητας στα δεδομένα με τιμές ακόμα και παραπάνω από 100. Άρα η μέθοδος stepwise selection εδώ δε λειτουργεί καλά, με το πρόβλημα υπερπροσαρμογής (overfitting) στα δεδομένα). Στο Output 48 βλέπουμε ότι οι 3 μεταβλητές που επηρεάζουν σημαντικά είναι οι CBE,PPO,CPO(p-value < 0.05 σε 95% επίπεδο σημαντικότητας).Θα κρατήσουμε αυτές. Το μοντέλο πετυχαίνει adjusted R squared ίσο με 91.7% και είναι το εξής:

$$PBE=476.1912 - 7.5996*CBE - 0.1885*PPO - 4.0498*CPO$$

Στο Output 34 βλέπουμε κάποια βασικά διαγράμματα ελέγχου του παραπάνω μοντέλου. Τώρα ας κάνουμε έναν έλεγχο για τις προϋποθέσεις της γραμμικής παλινδρόμησης:

1. Κανονικότητα καταλοίπων. Για τον έλεγχο του μοντέλου βλέπουμε από το Output 34 και κάνοντας τους ελέγχους shapiro και lillie (output 35) ότι τα δεδομένα είναι κανονικά κατανομημένα.

2. Ανεξαρτησία-τυχειότητα καταλοίπων.

Βλέπουμε από τον έλεγχο του Durbin-Watson στο Output 36 ότι η τιμή του ελέγχου είναι 1.65 (κοντά στο 2) η οποία δεν απορρίπτει την μηδενική υπόθεση της ανεξαρτησίας σε επίπεδο σημαντικότητας $\alpha = 5\%$. Άρα μπορούμε να επιβεβαιώσουμε την ανεξαρτησία και τυχειότητα των καταλοίπων.

3. Ομοσκεδαστικότητα καταλοίπων.

Για την ομοσκεδαστικότητα των καταλοίπων ελέγχουμε το διάγραμμα Scale-Location (Output 34), όπου μια οριζόντια γραμμή με περίπου ίση ταξινόμηση των αντικειμένων στο χώρο θα ήταν μια καλή ένδειξη (δεν είμαστε ιδιαίτερα σίγουροι στο παράδειγμα μας). Βρίσκουμε όμως με το LeveneTest (Output 46) ότι $p\text{-value}=0.27$, ιοπότε μπορούμε να θεωρήσουμε ότι η ομοσκεδαστικότητα ισχύει.

4. Γραμμικότητα καταλοίπων.

Στο διάγραμμα residuals vs fitted, βλέπουμε ότι η κόκκινη γραμμή δεν είναι ευθεία και πολλά σημεία βρίσκονται μακριά από το 0, κάτι που αποτελεί ένδειξη μη γραμμικότητας.

5. Πολυ-συγγραμμικότητα μοντέλου.

Μιας και πρόκειται για μοντέλο πολλαπλής παλινδρόμησης, οφείλουμε να ελέγξουμε τον δείκτη διόγκωσης διακύμανσης (VIF). Βρίσκουμε ότι ο δείκτης είναι σε όλες τις μεταβλητές < 5 , επομένως δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας του μοντέλου.

Θεωρώντας λοιπόν ότι η υπόθεση της γραμμικότητας καταλοίπων δεν ισχύει, το μοντέλο που προκύπτει δεν μπορεί να θεωρηθεί πλήρως αξιόπιστο οπότε χρειάζεται προσοχή στην ερμηνεία των αποτελεσμάτων.

Για τη τιμή του χοιρινού κρέατος κάνουμε αρχικά τη παλινδρόμηση με ανεξάρτητες τις CPO,PBE,CPE,YEAR. Όμως βρίσκουμε ότι παραβιάζεται ο κανόνας της πολυσυγγραμμικότητας (βρήκαμε τιμές > 10 στις μεταβλητές CBE και PBE). Αφαιρούμε τη PBE (είχε το μεγαλύτερο vif value) και αφού στο μοντέλο αυτό η CBE δε συνεισφέρει πολύ ($p\text{-value}=0.1$), την αφαιρούμε και αυτή. Τελικά παίρνουμε το μοντέλο:

$$PPO=4435.3084 - 5.5657 \cdot CPO - 2.1465 \cdot YEAR,$$

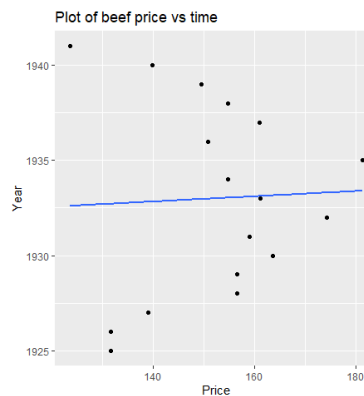
το οποίο έχει 95.4% adjusted R squared , κάτι που το καθιστά πολύ ισχυρό. Έπειτα ακολουθούμε την ίδια διαδικασία για τον έλεγχο των προϋποθέσεων του μοντέλου για τη γραμμική παλινδρόμηση. Αναλυτικά τα αποτελέσματα στα Outputs 37-40 και 42,47.

Τα AIC και BIC scores των μοντέλων παρουσιάζονται στα Output 50,51. Βλέπουμε ότι με μόνο $df=4$ καταφέραμε να κρατήσουμε ένα αξιοπρεπές AIC και BIC score και για χοιρινό και μοσχαρίσιο κρέας, ενώ για ένα καλό μέτρο σύγκρισης το μοντέλο που επιλέξαμε έχει πολύ καλύτερο score από ένα τυχαίο μοντέλο με τον ίδιο αριθμό df . Αυτό είναι ένα ακόμη στοιχείο που αποδεικνύει και αυτό την καλή απόδοση του μοντέλου.

4 Η επίδραση του χρόνου και άλλων μεταβλητών στις τιμές

4.1 Η επίδραση του χρόνου

Θα εξετάσουμε αρχικά τη τιμή του μοσχαρίσιου κρέατος με το χρόνο. Το παρακάτω scatterplot (σχήμα 3), μας βοηθάει για να οπτικοποιήσουμε τυχόν σχέσεις μεταξύ των μεταβλητών.

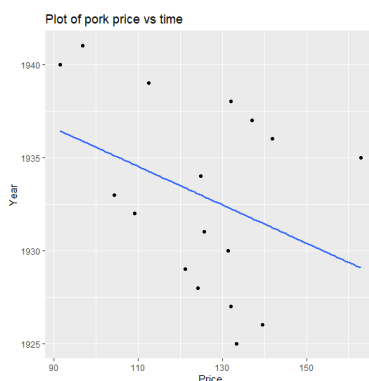


Σχήμα 3: Scatterplot για τη σχέση χρόνου-τιμής μοσχαρίσιου κρέατος.

Το σχήμα 3 δε μας πείθει ότι υπάρχει κάποια ιδιαίτερη σχέση μεταξύ των δύο μεταβλητών που εξετάζουμε. Κάνοντας δοκιμές στο μοντέλο της πολλαπλής παλινδρόμησης, παρατηρήσαμε ότι η πιο σημαντική μεταβλητή είναι η CBE στο

μοντέλο του μοσχαρίσιου κρέατος (δηλ. η κατανάλωσή του), κάτι απόλυτα λογικό. Ο χρόνος δεν είναι ισχυρός predictor, όπως υπέδειξε ο Kendall.

Θα δοκιμάσουμε τώρα τη σχέση μεταξύ της τιμής χοιρινού κρέατος- χρόνου. Βλέπουμε ότι ίσως εδώ η σχέση είναι λίγο πιο δυνατή από ότι προηγουμένως, όχι σίγουρα πολύ ισχυρή αλλά υπάρχει, καθώς και η κλίση της ευθείας είναι μεγαλύτερη.



Σχήμα 3: Scatterplot για τη σχέση χρόνου-τιμής χοιρινού κρέατος.

Επομένως ο χρόνος δεν επηρεάζει τις τιμές του μοσχαρίσιου κρέατος. Για το χοιρινό όμως, θεωρούμε ότι έχει βαρύτητα.

Ακόμη, ο χρόνος επηρεάζει (αρνητικά) το δείκτη τιμών λιανικής πώλησης. Όσο τα χρόνια περνούν, τόσο η τιμή πέφτει, κάτι που επαληθεύεται και από το γεγονός ότι η ευημερία στην οικονομία των ΗΠΑ σταμάτησε λόγω του "κραχ" και του Β' παγκόσμιου πολέμου. Ανάλογα κάνουμε ελέγχους για προϋποθέσεις της γραμμικής παλινδρόμησης.

4.2 Η επίδραση άλλων μεταβλητών

Θα δοκιμάσουμε αρχικά να επιβεβαιώσουμε τις σχέσεις των μεταβλητών που ανακαλύψαμε πριν μέσω του διαγράμματος του Kendall. Η τιμή επηρεάζεται από τη κατανάλωση; Ναι, σύμφωνα και με τα Output 19,20 για μοσχαρίσιο και χοιρινό κρέας αντίστοιχα.

Με βάση τα παραπάνω outputs επιβεβαιώνουμε τη σχέση που είχε προβλέψει ο Kendall. Το CPO (κατανάλωση του μοσχαρίσιου) είναι πολύ καλός εκτιμητής της τιμής του μοσχαρίσιου κρέατος και η κατανάλωση του χοιρινού επηρεάζει επίσης.

Κάνουμε την ίδια διαδικασία για να βρούμε αν κάποια μεταβλητή καθορίζει τη τιμή του χοιρινού κρέατος. Βρίσκουμε ότι καλός εκτιμητής είναι η κατανάλωση

του χοιρινού (προφανώς) και όπως μας είχε προειδεάσει ο Kendall η τιμή του μοσχαρίσιου κρέατος δεν μπορεί να εκτιμήσει καλά τη τιμή του χοιρινού όπως επίσης και η κατανάλωση του μοσχαρίσιου (το είδαμε στο μοντέλο παλινδρόμησης, που δε χρειαστήκαμε τις 2 αυτές μεταβλητές). Ο μοναδικός δηλαδή σοβαρός εκτιμητής της τιμής του χοιρινού είναι η κατανάλωση του. Η τιμή του χοιρινού κρέατος δεν επηρεάζεται από τον ανταγωνισμό, όπως επίσης και από τους δείκτες (εισοδήματος,κατανάλωσης κλπ.) , πράγμα που το καθιστά πολύ ισχυρό στην αγορά. Αναλυτικά τα αποτελέσματα στα Outputs 23,24.

5 Υπάρχουν σημαντικές διαφορές στους μέσους ή διαμέσους των περιόδων 1925-1930,1931-1935,1936-1941;

Αρχικά θα χωρίσουμε τις μεταβλητές PBE,PPO σε τρία μέρη. Για τη μεταβλητή PBE στα PBE-first,PBE-second και PBE-third που αντιπροσωπεύουν τις τιμές του μοσχαρίσιου κρέατος τα χρονικά διαστήματα 1925-1930,1931-1935 και 1936-1941 αντίστοιχα. Αντίστοιχα χωρίζουμε και την PPO. Οι τρεις νέες μεταβλητές για κάθε τιμή αντιπροσωπεύουν δείγματα ανεξάρτητα μεταξύ τους.

5.1 Μέσοι

Κάνουμε έλεγχο t-test για ανεξάρτητα δείγματα στις μεταβλητές που φτιάξαμε. Ξέρουμε ότι η PPO ακολουθεί τη κανονική κατανομή και έχοντας χωρίσει τη μεταβλητή στα δύο, προφανώς οι διακυμάνσεις είναι ίσες. Ξεκινώντας για την τιμή του μοσχαρίσιου κρέατος έχουμε:

```
welch Two Sample t-test

data: PBE_first and PBE_second
t = -2.5722, df = 8.9912, p-value = 0.0301
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -36.823870 -2.359221
sample estimates:
mean of x mean of y
 146.4589  166.0505

welch Two Sample t-test

data: PBE_first and PBE_third
t = -0.0092978, df = 9.9633, p-value = 0.9928
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17.69281  17.54584
sample estimates:
mean of x mean of y
 146.4589  146.5324
```

```

welch Two Sample t-test
data: PBE_second and PBE_third
t = 2.6507, df = 8.9924, p-value = 0.02647
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.858641 36.177476
sample estimates:
mean of x mean of y
166.0505 146.5324

```

Πίνακας 4: t-tests για να βρούμε αν η διαφορά μεταξύ των μέσων στις 3 περιόδους είναι σημαντική.

- Μεταξύ των περιόδων 1925-1930 και 1931-1935, υπάρχει στατιστικά σημαντική διαφορά καθώς το $p\text{-value}=0.03<0.05$, άρα απορρίπτουμε την H_0 ότι η διαφορά είναι μηδενική.
- Μεταξύ των περιόδων 1925-1930 και 1936-1941 δεν υπάρχει στατιστικά σημαντική διαφορά ($p\text{-value}>0.05$)
- Επίσης μεταξύ των 1931-1935 και 1936-1941 δεν υπάρχει στατιστικά σημαντική διαφορά.

Αντίστοιχα δουλεύουμε και για τη μεταβλητή PPO(τιμή του χοιρινού κρέατος). Προκύπτει ότι:

- Μεταξύ των περιόδων 1925-1930 και 1931-1935, δεν υπάρχει στατιστικά σημαντική διαφορά καθώς το $p\text{-value}>0.05$, άρα δεν απορρίπτουμε την H_0 ότι η διαφορά είναι μηδενική.
- Μεταξύ των περιόδων 1925-1930 και 1936-1941 επίσης δεν υπάρχει στατιστικά σημαντική διαφορά ($p\text{-value}>0.05$)
- Τέλος, μεταξύ των 1931-1935 και 1936-1941 πάλι δεν υπάρχει στατιστικά σημαντική διαφορά.

Τα αποτελέσματα αυτών των tests είναι στο Output 27.

Σημείωση: Αντί για μεμονωμένα t-tests για κάθε περίπτωση, μια εναλλακτική στην περίπτωσή μας θα ήταν one-way ANOVA. Τα πολλά t-tests είναι λιγότερο επιρρεπή σε σφάλματα από την ANOVA, για αυτό και προτιμήθηκε ως τρόπος επίλυσης του προβλήματος.

5.2 Διάμεσοι

Για τον έλεγχο των διαμέσων χρησιμοποιείται ο έλεγχος median-test του Mood και το Wilcoxon test. Στη περίπτωσή μας δε μπορεί να εφαρμοστεί ο πρώτος επειδή έχουμε άνισα μεγέθη των τριών μεταβλητών. Κάνουμε τον έλεγχο του Wilcoxon για κάθε ζεύγος περιόδων για τη τιμή μοσχαρίσιου κρέατος και έχουμε:

(Στον έλεγχο αυτό ως H_0 θεωρείται η περίπτωση που δεν υπάρχει στατιστικά σημαντική διαφορά στις διαμέσους και H_a το αντίθετο.)

- Μεταξύ των περιόδων 1925-1930 και 1931-1935, δεν υπάρχει στατιστικά σημαντική διαφορά καθώς το $p\text{-value} > 0.05$, άρα δεν απορρίπτουμε την H_0 ότι η διαφορά είναι μηδενική.
- Μεταξύ των περιόδων 1925-1930 και 1936-1941 το $p\text{-value} > 0.05$, επομένως πάλι δεν υπάρχει στατιστικά σημαντική διαφορά. .
- Τέλος, μεταξύ των 1931-1935 και 1936-1941 το $p\text{-value} = 0.02 < 0.05$, επομένως μπορούμε να πούμε ότι υπάρχει στατιστικά σημαντική διαφορά.

Κάνουμε τον ίδιο έλεγχο και για τη τιμή του χοιρινού και προκύπτει ότι μεταξύ όλων των περιόδων δεν υπάρχει στατιστικά σημαντική διαφορά (πάντα το $p\text{-value}$ ήταν > 0.05). Τα αποτελέσματα μπορείτε να τα δείτε αναλυτικά στα output 28,29.

6 Συμπεράσματα ανάλυσης

Από την παραπάνω ανάλυση, τα κύρια συμπεράσματά μας είναι τα εξής:

- Ο χρόνος δεν επηρεάζει τις τιμές του μοσχαρίσιου κρέατος. Επηρεάζει όμως τις τιμές του χοιρινού.
- Η τιμή του μοσχαρίσιου κρέατος εύλογα επηρεάζεται από την κατανάλωση του, το δείκτη τιμών των προϊόντων και το δείκτη διαθέσιμου εισοδήματος/άτομο.
- Η τιμή του χοιρινού επηρεάζεται μεν από την κατανάλωση του αλλά όχι από άλλους παράγοντες. Αυτό σημαίνει ότι το χοιρινό κρέας θα προσφέρεται πάντα σταθερά σε συνάρτηση με την κατανάλωσή του χωρίς να έχουν σημασία οι δείκτες εισοδήματος/τιμών τροφίμων.
- Οι μέθοδοι stepwise selection οδηγούν σε overfitting των δεδομένων.
- Υπάρχει σημαντική διαφορά ανάμεσα στους μέσους των τιμών του μοσχαρίσιου κρέατος των περιόδων 1925-1930 και 1931-1935. Πιθανότατα σε αυτό έπαιξε ρόλο και η οικονομική ύφεση που ξεκίνησε το 1929. Η τιμή του χοιρινού ωστόσο δεν επηρεάστηκε, κάτι που το περιμέναμε, αφού όπως είπαμε η τιμή του χοιρινού επηρεάζεται μόνο από την κατανάλωση του.
- Υπάρχει σημαντική διαφορά ανάμεσα στους διάμεσους των τιμών του μοσχαρίσιου κρέατος μεταξύ των περιόδων 1931-1935 και 1936-1941. Για τις διαμέσους στο χοιρινό δεν υπάρχουν στατιστικά σημαντικές διαφορές, όπως αναμέναμε.

7 Παράρτημα

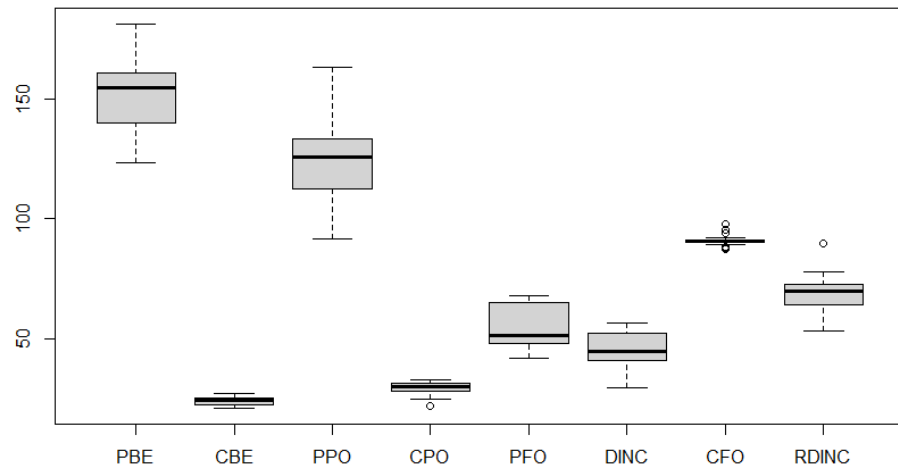
```
> str(data)
'data.frame': 17 obs. of 10 variables:
 $ YEAR : int 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 ...
 $ PBE : num 59.7 59.7 63.71 71.71 74.2 72.1 79.73 1.70.2 ...
 $ CBE : num 58.6 59.4 53.7 48.1 49 48.2 47.9 46 50.8 55.2 ...
 $ PPO : num 60.5 63.3 59.9 56.3 55 59.6 57 49.5 47.3 56.6 ...
 $ CPO : num 65.8 63.3 66.8 69.9 68.7 66.1 67.4 69.7 68.7 62.2 ...
 $ PFO : num 65.8 68 65.5 64.8 65.6 62.4 51.4 42.8 41.6 46.4 ...
 $ DINC : num 51.4 52.6 52.1 52.7 55.1 48.8 41.5 31.4 29.4 33.2 ...
 $ CFO : num 90.9 92.1 90.9 90.9 91.1 90.7 90 87.8 88 89.1 ...
 $ RDINC : num 68.5 69.6 70.2 71.9 75.2 68.3 64 53.9 53.2 58 ...
 $ RFP : int 877 899 883 884 895 874 791 733 752 811 ...
```

Output 1: Ανάλυση μεταβλητών μοντέλου.

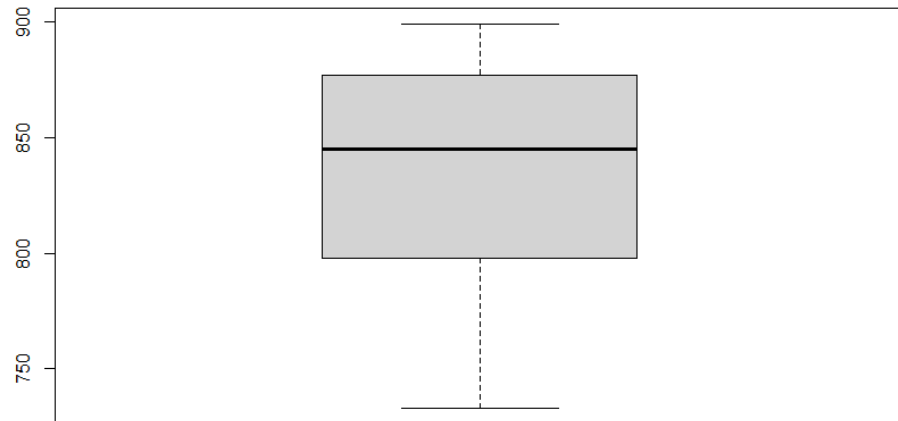
```
summary(data)
  YEAR      PBE      CBE      PPO      CPO      PFO      DINC
Min. :1925 Min. :123.5 Min. :20.87 Min. : 91.49 Min. :21.64 Min. :41.60 Min. :29.40
1st Qu.:1929 1st Qu.:139.8 1st Qu.:22.23 1st Qu.:112.43 1st Qu.:28.21 1st Qu.:47.80 1st Qu.:40.80
Median :1933 Median :154.8 Median :24.36 Median :125.66 Median :29.98 Median :51.40 Median :44.50
Mean :1933 Mean :152.2 Mean :24.08 Mean :124.73 Mean :29.00 Mean :54.22 Mean :44.62
3rd Qu.:1937 3rd Qu.:160.9 3rd Qu.:25.04 3rd Qu.:133.38 3rd Qu.:31.16 3rd Qu.:64.80 3rd Qu.:52.10
Max. :1941 Max. :181.2 Max. :27.22 Max. :162.92 Max. :32.84 Max. :68.00 Max. :56.30

      CFO      RDINC      RFP
Min. :87.30 Min. :53.20 Min. :733.0
1st Qu.:90.00 1st Qu.:64.00 1st Qu.:798.0
Median :90.70 Median :69.60 Median :845.0
Mean :91.01 Mean :68.65 Mean :833.2
3rd Qu.:91.10 3rd Qu.:72.50 3rd Qu.:877.0
Max. :97.50 Max. :89.50 Max. :899.0
```

Output 2: Βασικά περιγραφικά μέτρα για τις μεταβλητές.



Output 3: Πλάι-πλάι boxplots των μεταβλητών (πλην της RFP)



Output 4: Boxplot της RFP.

```
shapiro-wilk normality test
data:  data$PFO
W = 0.87368, p-value = 0.02515
```

```
shapiro-wilk normality test
data:  data$CPO
W = 0.88831, p-value = 0.04346
```

Output 5: Έλεγχος Shapiro-wilk για τις μεταβλητές PFO,CPO αντίστοιχα.

```
shapiro-wilk normality test
data:  data$CBE
W = 0.95064, p-value = 0.4668
```

Output 6: Έλεγχος Shapiro-wilk για τη μεταβλητή CBE.

```
shapiro-wilk normality test
data:  data$CFO
W = 0.89449, p-value = 0.05497
```

Output 7: Έλεγχος Shapiro-wilk για τη μεταβλητή CFO.

```
shapiro-wilk normality test
data:  data$DINC
W = 0.94385, p-value = 0.3667
```


Output 8: Έλεγχος Shapiro-wilk για τη μεταβλητή DINC.

```
> shapiro.test(data$PBE)

      Shapiro-Wilk normality test

data:  data$PBE
W = 0.96482, p-value = 0.723
```

Output 9: Έλεγχος Shapiro-wilk για τη μεταβλητή PBE.

```
      Shapiro-Wilk normality test

data:  data$PPO
W = 0.96691, p-value = 0.7623
```

Output 10: Έλεγχος Shapiro-wilk για τη μεταβλητή PPO.

```
      Shapiro-Wilk normality test

data:  data$RDINC
W = 0.94442, p-value = 0.3744
```

Output 11: Έλεγχος Shapiro-wilk για τη μεταβλητή RDINC.

```
      Shapiro-Wilk normality test

data:  data$RFP
W = 0.94032, p-value = 0.3221
```

Output 12: Έλεγχος Shapiro-wilk για τη μεταβλητή RFP.

Kendall's correlation diagram

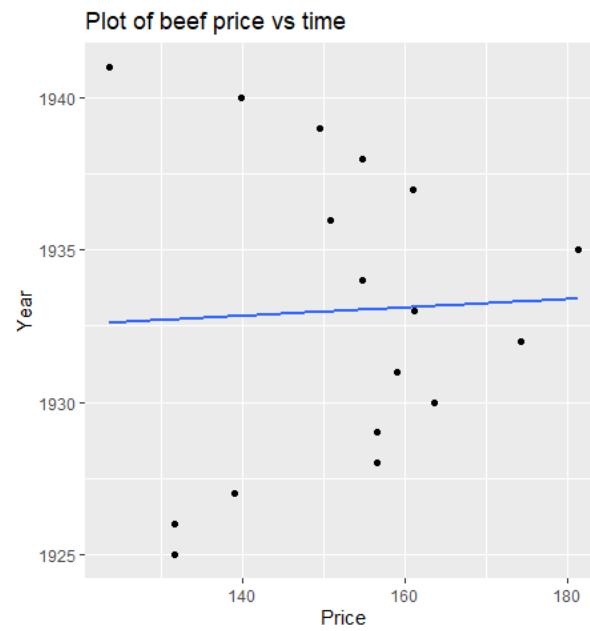
YEAR	-0.44*	-0.38*	-0.29	-0.10	-0.07	-0.02	0.04	0.18	0.24	
RDINC	0.32	0.29	-0.10	0.66***	0.12	-0.38*	0.69***	0.29		
CBE	0.18	0.21	0.17	0.25	-0.28	-0.60***	0.35			
CFO	0.41*	0.36*	-0.22	0.65***	0.22	-0.62***				
PBE	-0.29	-0.22	0.03	-0.45*	-0.04					
CPO	-0.03	-0.12	-0.65***	0.19						
DINC	0.66***	0.60***	-0.02							
PPO	0.29	0.29								
RFP	0.74***									
PFO										
	PFO	RFP	PPO	DINC	CPO	PBE	CFO	CBE	RDINC	YEAR

Output 13: Πίνακας συσχέτισης του Kendall.

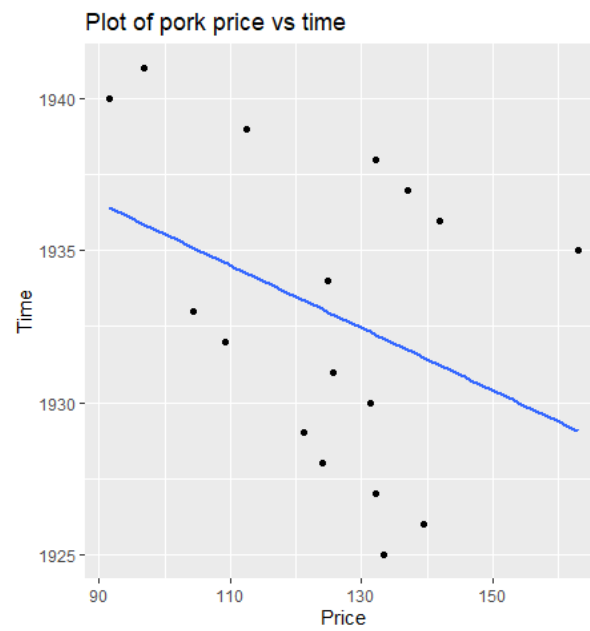
Sparman's correlation diagram

YEAR	-0.62**	-0.58*	-0.32	-0.27	-0.16	-0.01	0.08	0.23	0.29	
RDINC	0.42	0.38	-0.21	0.78***	0.17	-0.56*	0.84***	0.40		
CBE	0.24	0.26	0.23	0.31	-0.44	-0.77***	0.45			
CFO	0.49*	0.38	-0.32	0.81***	0.30	-0.78***				
PBE	-0.41	-0.29	0.06	-0.58*	-0.01					
CPO	-0.02	-0.16	-0.81***	0.25						
DINC	0.84***	0.75***	-0.03							
PPO	0.41	0.49*								
RFP	0.89***									
PFO										
	PFO	RFP	PPO	DINC	CPO	PBE	CFO	CBE	RDINC	YEAR

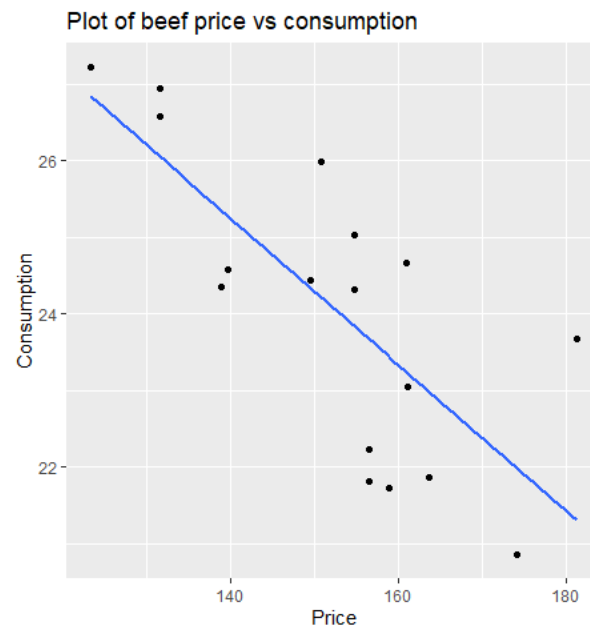
Output 14: Πίνακας συσχέτισης του Spearman.



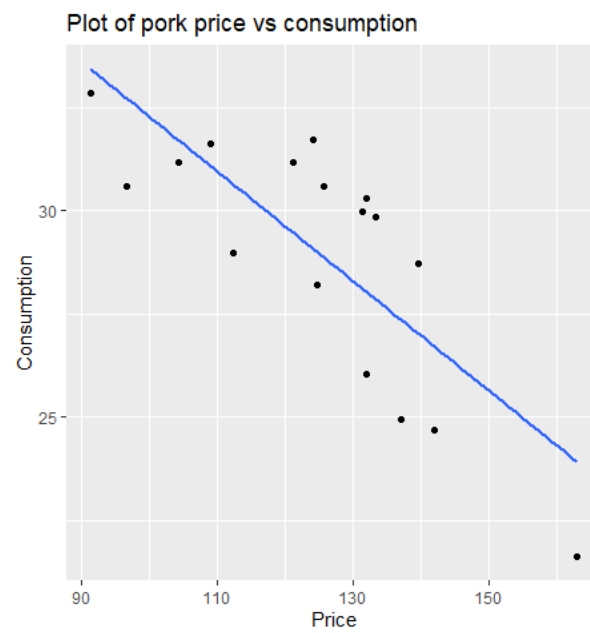
Output 15: Scatterplot για τη σχέση χρόνου-τιμής μοσχαρίσιου κρέατος.



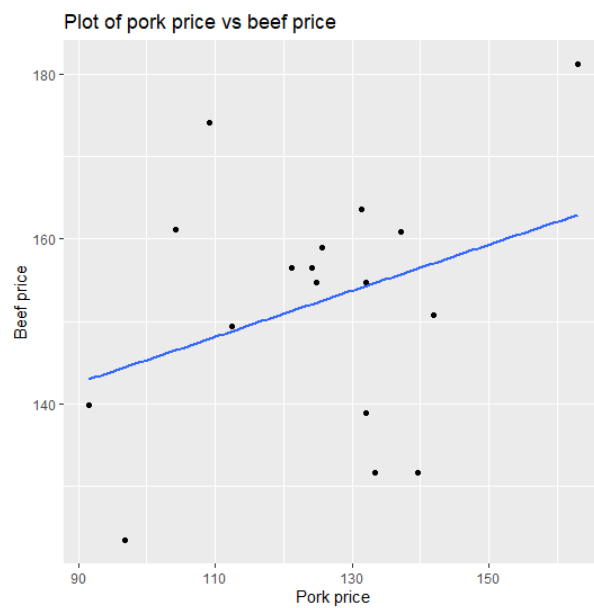
Output 17: Scatterplot για τη σχέση χρόνου-τιμής χοιρινού κρέατος.



Output 19: Scatterplot για τη σχέση τιμής-κατανάλωσης μοσχαρίσιου κρέατος.



Output 23: Scatterplot για τη σχέση τιμής-κατανάλωσης χοιρινού κρέατος.



Output 24: Scatterplot για τη σχέση τιμής μοσχαρίσιου-χοιρινού κρέατος.

```
welch Two Sample t-test
data: PBE_first and PBE_second
t = -2.5722, df = 8.9912, p-value = 0.0301
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -36.823870 -2.359221
sample estimates:
mean of x mean of y
146.4589 166.0505

welch Two Sample t-test
data: PBE_first and PBE_third
t = -0.0092978, df = 9.9633, p-value = 0.9928
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17.69281 17.54584
sample estimates:
mean of x mean of y
146.4589 146.5324
```

```

welch Two Sample t-test
data: PBE_second and PBE_third
t = 2.6507, df = 8.9924, p-value = 0.02647
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.858641 36.177476
sample estimates:
mean of x mean of y
166.0505 146.5324

```

Output 26: t-tests για να βρούμε αν η διαφορά μεταξύ των μέσων στις 3 περιόδους είναι σημαντική(μοσχάρισιο).

```

welch Two Sample t-test
data: PPO_first and PPO_second
t = 0.464, df = 4.5539, p-value = 0.664
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-23.24612 33.12272
sample estimates:
mean of x mean of y
130.2919 125.3536

```

```

welch Two Sample t-test
data: PPO_first and PPO_third
t = 1.2661, df = 5.9374, p-value = 0.2529
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-10.92089 34.21616
sample estimates:
mean of x mean of y
130.2919 118.6442

```

```

welch Two Sample t-test
data: PPO_second and PPO_third
t = 0.49559, df = 8.3926, p-value = 0.6329
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-24.25701 37.67568
sample estimates:
mean of x mean of y
125.3536 118.6442

```

Output 27: t-tests για να βρούμε αν η διαφορά μεταξύ των μέσων στις 3 περιόδους είναι σημαντική(χοιρινό).

```

wilcoxon rank sum test with continuity correction
data: PBE_first and PBE_second
W = 5, p-value = 0.08144
alternative hypothesis: true location shift is not equal to 0

```

```

wilcoxon rank sum test with continuity correction
data: PBE_first and PBE_third
W = 19, p-value = 0.936
alternative hypothesis: true location shift is not equal to 0

```

```

wilcoxon rank sum test with continuity correction
data: PBE_second and PBE_third
W = 27.5, p-value = 0.0281
alternative hypothesis: true location shift is not equal to 0

```

Output 28: Wilcox tests για να βρούμε αν η διαφορά μεταξύ των διαμέσων στις 3 περιόδους είναι σημαντική(μοσχάρισιο).

```

wilcoxon rank sum exact test
data: PPO_first and PPO_second
W = 20, p-value = 0.4286
alternative hypothesis: true location shift is not equal to 0

```

```

wilcoxon rank sum test with continuity correction
data: PPO_first and PPO_third
W = 21.5, p-value = 0.6304
alternative hypothesis: true location shift is not equal to 0

```

```

wilcoxon rank sum exact test
data: PPO_second and PPO_third
W = 16, p-value = 0.9307
alternative hypothesis: true location shift is not equal to 0

```

Output 29: Wilcox tests για να βρούμε αν η διαφορά μεταξύ των διαμέσων στις 3 περιόδους είναι σημαντική(χοιρινό).

```

Pearson's Chi-squared test
data: data$PBE and data$CBE
X-squared = 221, df = 208, p-value = 0.2556

```

```

Pearson's Chi-squared test
data: data$PBE and data$CPO
X-squared = 191.25, df = 182, p-value = 0.3044

```

```

Pearson's Chi-squared test
data: data$PBE and data$PPO
X-squared = 208.25, df = 195, p-value = 0.2451

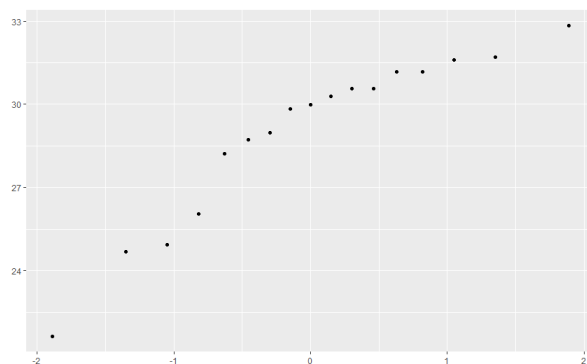
```

```

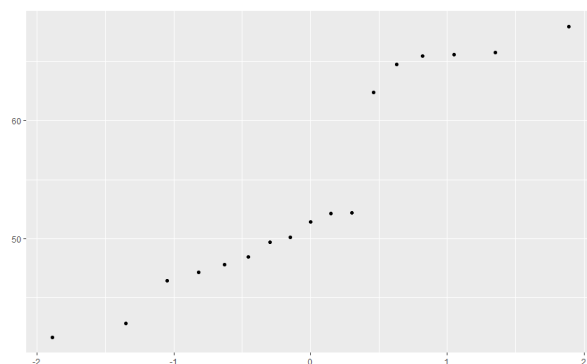
Pearson's Chi-squared test
data: data$CPO and data$DINC
X-squared = 238, df = 224, p-value = 0.2484

```


Output 30: Chi-squared tests για τη διερεύνηση σχέσεων μεταξύ των μεταβλητών.



Output 31: QQPlot για τη μεταβλητή CPO.



Output 32: QQPlot για τη μεταβλητή PFO.

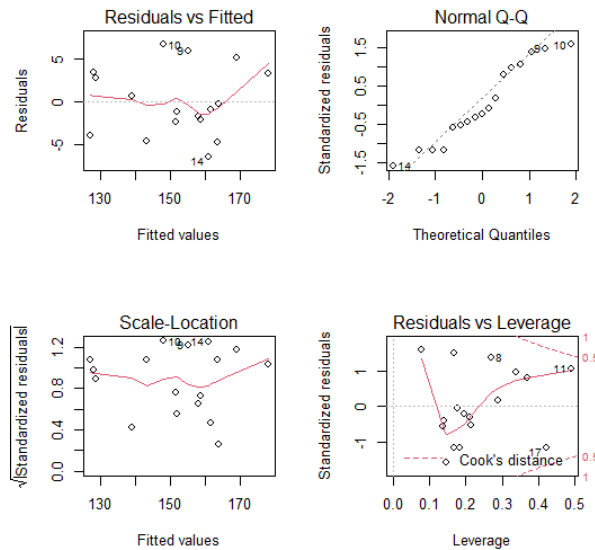
```
Call:
lm(formula = data$PBE ~ data$CBE + data$PPO + data$CPO)

Residuals:
    Min       1Q   Median       3Q      Max
-6.3336 -2.3479 -0.8562  3.2936  6.6282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  476.1912    36.8975   12.906 8.69e-09 ***
data$CBE      -7.5996     0.6071  -12.518 1.26e-08 ***
data$PPO     -0.1885     0.1020   -1.848  0.0875 .
data$CPO     -4.0498     0.6273   -6.456 2.15e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.373 on 13 degrees of freedom
Multiple R-squared:  0.9332,    Adjusted R-squared:  0.9178
F-statistic: 60.52 on 3 and 13 DF,  p-value: 6.786e-08
```

Output 33: Μοντέλο πολλαπλής γραμμικής παλινδρόμησης για τη τιμή του μοσχαρίσιου κρέατος.



Output 34: Βασικά διαγράμματα ελέγχου μοντέλου γραμμικής παλινδρόμησης (μοσχάρσιο κρέας).

```
> shapiro.test(rstandard(recommended_model)) #Shapiro test
      Shapiro-Wilk normality test
data:  rstandard(recommended_model)
W = 0.93121, p-value = 0.228

> lillie.test(rstandard(recommended_model)) #Lillie test
      Lilliefors (Kolmogorov-Smirnov) normality test
data:  rstandard(recommended_model)
D = 0.13022, p-value = 0.6213
```

Output 35: Έλεγχοι Shapiro και Lillie (μοσχάρσιο κρέας).

```
lag Autocorrelation D-W Statistic p-value
1      0.140275      1.646788    0.106
Alternative hypothesis: rho != 0
```

Output 36: Έλεγχος Durbin-Watson (μοσχάρσιο κρέας).

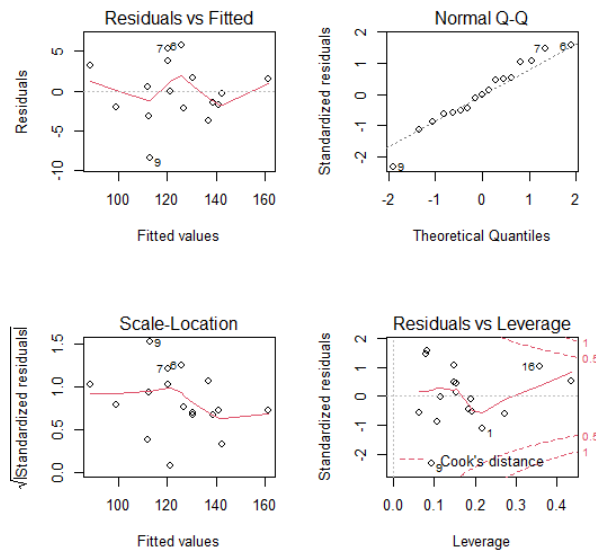
```
Call:
lm(formula = data$PPO ~ data$CPO + data$YEAR)

Residuals:
    Min       1Q   Median       3Q      Max
-8.4073 -2.0128 -0.0178  1.6950  5.7060

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4435.3084   380.4533   11.66 1.35e-08 ***
data$CPO      -5.5657     0.3274  -17.00 9.64e-11 ***
data$YEAR     -2.1465     0.1955  -10.98 2.90e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.812 on 14 degrees of freedom
Multiple R-squared:  0.9599,    Adjusted R-squared:  0.9541
F-statistic: 167.4 on 2 and 14 DF, p-value: 1.677e-10
```

Output 37: Μοντέλο πολλαπλής γραμμικής παλινδρόμησης για τη τιμή του χοιρινού κρέατος.



Output 38: Βασικά διαγράμματα ελέγχου μοντέλου γραμμικής παλινδρόμησης (χοιρινό κρέας).

```
> shapiro.test(rstandard(recommended_model2_3)) #shapiro test
      Shapiro-Wilk normality test
data:  rstandard(recommended_model2_3)
W = 0.96759, p-value = 0.775

> lillie.test(rstandard(recommended_model2_3)) #Lillie test
      Lilliefors (Kolmogorov-Smirnov) normality test
data:  rstandard(recommended_model2_3)
D = 0.090057, p-value = 0.9717
```

Output 39: Έλεγχοι Shapiro και Lillie (χοιρινό κρέας).

```
> dwt(recommended_model2_3)
lag Autocorrelation D-W Statistic p-value
1      0.2893569      1.330378      0.04
Alternative hypothesis: rho != 0
```

Output 40: Έλεγχος Durbin-Watson (χοιρινό κρέας).

```
> vif(recommended_model)
data$CBE data$PPO data$CPO
1.162126 2.756463 2.991482
```

Output 41: Έλεγχος VIF (μοσχαρίσιο κρέας).

```
> vif(recommended_model2_3)
data$CPO data$YEAR
1.073157 1.073157
```

Output 42: Έλεγχος VIF (χοιρινό κρέας).

```
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 3 1.4399 0.2764
13
```

Output 46: Έλεγχος ομοσκεδαστικότητας για το πολυμεταβλητό μοντέλο (μοσχαρίσιο κρέας).

```
> leveneTest(rstandard(recommended_model2_3),qfits)
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group 3 0.7976 0.5169
13
```

Output 47: Έλεγχος ομοσκεδαστικότητας για το πολυμεταβλητό μοντέλο (χοιρινό κρέας).

```

Call:
lm(formula = data$PBE ~ ., data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-2.6424 -1.2473 -0.2206  1.5579  2.9213

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 5937.01753 2949.64964   2.013 0.084020 .
YEAR         -2.83718    1.53663  -1.846 0.107337
CBE          -8.98768    1.37738  -6.525 0.000326 ***
PPO          -0.99901    0.40506  -2.466 0.043064 *
CPO          -8.41057    2.76308  -3.044 0.018743 *
PFO           2.80889    4.33101    0.649 0.537314
DINC         -5.50692    5.73590   -0.960 0.368995
CFO           2.30349    2.35403    0.979 0.360411
RDINC        2.91347    3.77360    0.772 0.465337
RFP          -0.03814    0.24928   -0.153 0.882717
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.728 on 7 degrees of freedom
Multiple R-squared:  0.986,    Adjusted R-squared:  0.968
F-statistic: 54.77 on 9 and 7 DF,  p-value: 1.165e-05

```

Output 48: Μοντέλο πολλαπλής παλινδρόμησης για τη τιμή του μοσχαρίσιου κρέατος με όλες τις μεταβλητές.

```

Call:
lm(formula = data$PPO ~ ., data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-2.3900 -0.8946  0.3874  1.0135  1.7948

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 5575.9648 1399.0955   3.985 0.00529 **
YEAR         -2.7393    0.7509  -3.648 0.00820 **
PBE          -0.4654    0.1887  -2.466 0.04306 *
CBE          -5.4954    1.3949  -3.940 0.00561 **
CPO          -7.3037    0.8024  -9.103 3.96e-05 ***
PFO           3.5768    2.7269    1.312 0.23102
DINC         -5.7460    3.5537  -1.617 0.14993
CFO           2.2561    1.4858    1.518 0.17269
RDINC        3.2395    2.3874    1.357 0.21694
RFP          -0.1278    0.1634   -0.782 0.45989
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.862 on 7 degrees of freedom
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9891
F-statistic: 161.6 on 9 and 7 DF,  p-value: 2.797e-07

```

Output 49: Μοντέλο πολλαπλής παλινδρόμησης για τη τιμή του χοιρινού κρέατος με όλες τις μεταβλητές.

```

> AIC(full,model2,recommended_model1,random_model1_beef) #Get AIC score for the models
      df      AIC
full      11 89.28572
model2     10 87.34247
recommended_model1 5 103.85200
random_model1_beef 5 135.95384
> AIC(full_pork,model2_pork,recommended_model2_3,random_model1_pork)
      df      AIC
full_pork    11 76.30018
model2_pork   10 75.72370
recommended_model2_3 4 98.43639
random_model1_pork 4 150.45991

```

Output 50: AIC scores για τα μοντέλα πολλαπλής παλινδρόμησης που εφαρμόσαμε (πρώτα για μοσχαρίσιο και μετά για χοιρινό κρέας)

```

> BIC(full,model2,recommended_model1,random_model1_beef) #Get BIC score for the models
      df      BIC
full      11 98.45107
model2     10 95.67461
recommended_model1 5 108.01807
random_model1_beef 5 140.11990
> BIC(full_pork,model2_pork,recommended_model2_3,random_model1_pork)
      df      BIC
full_pork    11 85.46552
model2_pork   10 84.05584
recommended_model2_3 4 101.76924
random_model1_pork 4 153.79276

```

Output 51: BIC scores για τα μοντέλα πολλαπλής παλινδρόμησης που εφαρμόσαμε (πρώτα για μοσχαρίσιο και μετά για χοιρινό κρέας)