

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Εργασία : Τιμολόγηση Διαμαντιών

Φοιτήτρια :

Ονοματεπώνυμο : Ραδιώτη Νίκη

Αριθμός Μητρώου : 3200165

Περιεχόμενα:

1. Εισαγωγή - περιγραφή μελέτης και προβλήματος
2. Περιγραφική ανάλυση
3. Σχέσεις μεταβλητών ανά δύο
4. Προβλεπτικά ή Ερμηνευτικά Μοντέλα
5. Συμπεράσματα και συζήτηση

ΠΑΡΑΡΤΗΜΑ

Κεφάλαιο 1

Εισαγωγή - Περιγραφή μελέτης και προβλήματος

Η αξιολόγηση της τιμής μιας πέτρας διαμαντιού δεν είναι εύκολη υπόθεση αφού πρέπει να λάβει κάποιος υπ' όψη τα καράτια, το χρώμα, τη διαφάνεια και το κόσψιμο του. Μέσω της στατιστικής προσφέρεται τρόπος να αποφανθεί κάποιος για την τελική τιμή μιας τέτοιας πέτρας. Τα δεδομένα της εργασίας αναφέρονται σε ένα δείγμα 308 παρατηρήσεων τα οποία προέρχονται από μια μελέτη που έγινε στη Σιγκαπούρη σχετικά με τον τρόπο τιμολόγησης των διαμαντιών.

Θα παρουσιάσουμε τα κατάλληλα περιγραφικά μέτρα για τις μεταβλητές: Τελική τιμή, βάρος της πέτρας διαμαντιού σε καράτια, χρώμα, διαφάνεια, οργανισμός πιστοποίησης και θα εξετάσουμε τις μεταξύ τους σχέσεις. Θα εκτιμήσουμε αν υπάρχει διαφορά στην τελική τιμή ανάλογα με το μέγεθος της πέτρας. Θα εκτιμήσουμε ένα στατιστικό μοντέλο που θα εξετάζει την τιμή ενός διαμαντιού σε σχέση με τις άλλες μεταβλητές και θα εντοπίζει ποιά χαρακτηριστικά είναι εκείνα που ανεβάζουν την αξία μιας πέτρας. Ακόμα θα εντοπίζει πόσο η τελική τιμή διαφοροποιείται ανάλογα με τον οργανισμό πιστοποίησης. Θα ελέγξουμε την ύπαρξη ακραίων τιμών και αν είναι απαραίτητη η αφαίρεση τους. Τέλος θα εκτιμήσουμε με βάση το μοντέλο την αναμενόμενη τιμή για ένα τυπικό διαμάντι.

Τα συμπεράσματα που θα προκύψουν είναι χρήσιμα για τις προβλέψεις στο εμπόριο διαμαντιών. Το εμπόριο των διαμαντιών είναι ακριβό και επηρεάζει τις οικονομίες των χωρών - κρατών που εξάγουν ή εισάγουν διαμάντια. Επομένως είναι σημαντικό για την παγκόσμια οικονομία.

Κεφάλαιο 2ο

Περιγραφική ανάλυση

Ξεκινώντας την ανάλυση μας εισάγουμε τα δεδομένα στην R με την οποία θα εργαστούμε¹. Οι μεταβλητές, τιμή και μέγεθος σε καράτια είναι ποσοτικές ενώ οι μεταβλητές διαφάνεια, οργανισμός πιστοποίησης και χρώμα είναι ποιοτικές (κατηγορικές).

Ο έλεγχος της κανονικότητας για τις δύο ποσοτικές μεταβλητές γίνεται χρησιμοποιώντας Q-Q plots και τον έλεγχο Shapiro-Wilk ή Kolmogorov-Smirnov (χρησιμοποιώ έναν από τους δύο ελέγχους γιατί έχω μεγάλο δείγμα $n=308$). Τελικά βλέπουμε ότι καμία από τις δύο ποσοτικές μεταβλητές δεν είναι κανονική (p-values : $3.343e-11$ και $8.404e-11$ αντίστοιχα στον έλεγχο Shapiro-Wilk)².

Στην συνέχεια θα ήταν καλό να πάρουμε κάποια διαγράμματα για να δούμε τη μορφή των δεδομένων σε κάθε ποσοτική μεταβλητή. Τα ενδεδειγμένα διαγράμματα είναι τα ιστογράμματα και τα διαγράμματα πλαισίου και αποπλήξεων (boxplots)³. Επιπλέον για να έχουμε μια πιο πλήρη εικόνα μπορούμε να χρησιμοποιήσουμε τα διαγράμματα "βιολιού"⁴ Τα εν λόγω διαγράμματα μας δίνουνε μία πολύ καλή ιδέα για το πώς κατανέμονται οι μεταβλητές μας όπως φαίνεται παρακάτω:

1: για λεπτομέρειες για την εισαγωγή των δεδομένων βλέπε

ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 1

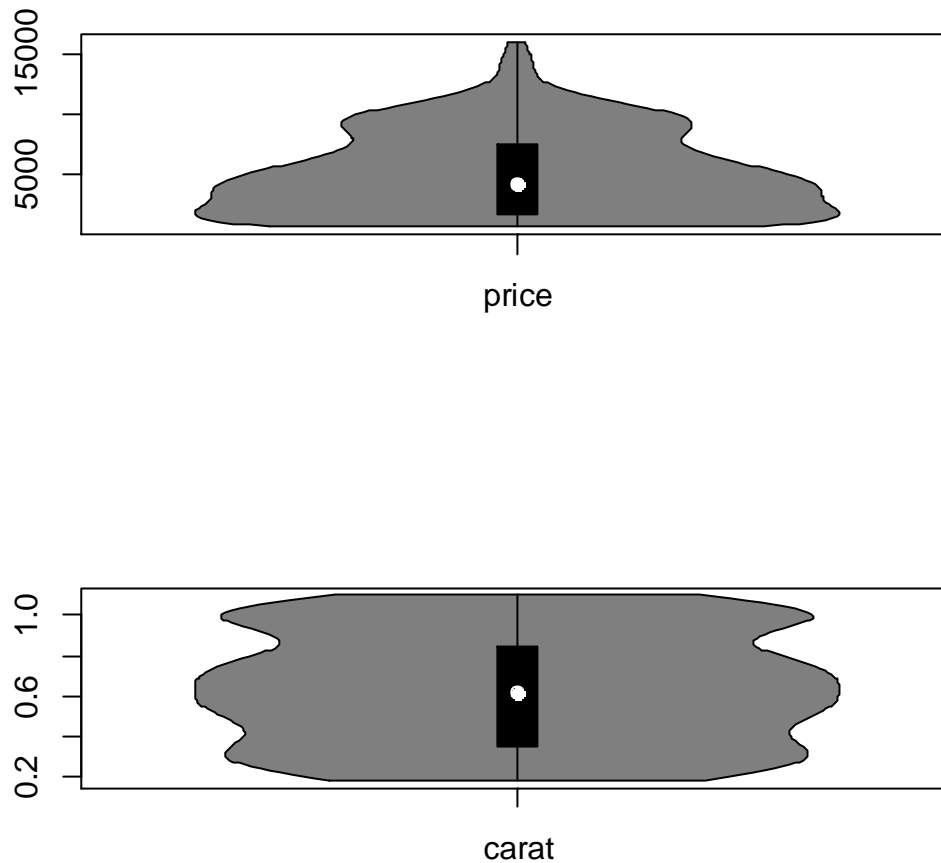
2: βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 2,ΠΙΝΑΚΕΣ 1,2, ΔΙΑΓΡΑΜΜΑΤΑ 1,2

3:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 3,ΔΙΑΓΡΑΜΜΑΤΑ 3,4

4:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 4

ΔΙΑΓΡΑΜΜΑΤΑ "ΒΙΟΛΙΟΥ" για τις ποσοτικές μεταβλητές "ΤΙΜΗ
ΔΙΑΜΑΝΤΙΟΥ" και "ΜΕΓΕΘΟΣ ΔΙΑΜΑΝΤΙΟΥ"

ΔΙΑΓΡΑΜΜΑΤΑ "ΒΙΟΛΙΟΥ" ΓΙΑ ΤΙΣ ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ ΤΙΜΗΣ ΚΑΙ ΜΕΓΕΘΟΥΣ



Τα κατάλληλα μέτρα για τις ποσοτικές θα είναι η μέση τιμή, η διάμεσος, η τυπική απόκλιση, τα τεταρτημόρια και τέλος η κύρτωση και η ασυμμετρία⁵.

Παίρνοντας τα γραφήματα από την R βλέπουμε πως η μεταβλητή τελική τιμή (price) έχει πολλές χαμηλές τιμές και λίγες μόνο πολύ υψηλές τιμές (δεξιά ασυμμετρία όπως άλλωστε περιμέναμε, βλέποντας τους δείκτες ασυμμετρίας). Ενώ η μεταβλητή μέγεθος σε καράτια δεν έχει ούτε δεξιά, ούτε αριστερή ασυμμετρία. Επίσης παρουσιάζει τρεις κορυφές μία για τις μικρές τιμές, μία για τις μεσαίες τιμές και μία για τις μεγάλες τιμές (οι περισσότερες παρατηρήσεις είναι στο μεσαίο μέγεθος).

5. βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 5, ΠΙΝΑΚΕΣ 3,4

Συνεχίζοντας την ανάλυσή μας κατηγοριοποιούμε την μεταβλητή μέγεθος σε καράτια για να την εξετάσουμε ξεχωριστά⁶.

Για τις κατηγορικές μεταβλητές θα χρησιμοποιήσουμε διαγράμματα "πίτας"⁷, ραβδογράμματα⁸, πίνακες συχνοτήτων και σχετικών συχνοτήτων⁹. Παρατηρούμε ότι τα περισσότερα διαμάντια είναι μεσαίου μεγέθους (ποσοστό 46%). Ακόμα για το χρώμα παρατηρούμε ότι 27% είναι F, 21% είναι G και 20% είναι H. Για τη διαφάνεια παρατηρούμε ότι 26% είναι VS1 και 25% είναι VVS2. Για τον οργανισμό πιστοποίησης παρατηρούμε ότι το 50% είναι GIA.

6. βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 6

7. βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 7, ΔΙΑΓΡΑΜΜΑ 5

8. βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 8, ΔΙΑΓΡΑΜΜΑ 6

9. βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 9, ΠΙΝΑΚΕΣ 5

ΠΙΝΑΚΑΣ ΣΧΕΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΚΗ ΜΕΤΑΒΛΗΤΗ
ΤΟΥ ΜΕΓΕΘΟΥΣ

1	2	3
0.31	0.46	0.23

ΠΙΝΑΚΑΣ ΣΧΕΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΚΗ ΜΕΤΑΒΛΗΤΗ
ΤΟΥ ΧΡΩΜΑΤΟΣ

D	E	F	G	H	I
0.05	0.14	0.27	0.21	0.20	0.13

ΠΙΝΑΚΑΣ ΣΧΕΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΚΗ ΜΕΤΑΒΛΗΤΗ
ΤΗΣ ΔΙΑΦΑΝΕΙΑΣ

IF	VS1	VS2	VVS1	VVS2
0.14	0.26	0.17	0.17	0.25

ΠΙΝΑΚΑΣ ΣΧΕΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΚΗ ΜΕΤΑΒΛΗΤΗ
ΤΟΥ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ

GIA	HRD	IGI
0.49	0.26	0.25

Κεφάλαιο 3ο

Σχέσεις μεταβλητών ανά δύο

Έχοντας μελετήσει τις μεταβλητές μας κάθε μία ξεχωριστά σειρά έχει η μελέτη των μεταξύ τους σχέσεων. Οι σχέσεις που έχουν κάποιο νόημα να μελετηθούν είναι οι σχέσεις μεταξύ των μεταβλητών:

τιμή - μέγεθος (δύο ποσοτικές)

τιμή - μέγεθος κατηγοριοποιημένο (μία ποσοτική και μία κατηγορική)

τιμή - διαφάνεια (μία ποσοτική και μία κατηγορική)

τιμή - χρώμα (μία ποσοτική και μία κατηγορική)

τιμή - οργανισμός πιστοποίησης (μία ποσοτική και μία κατηγορική)

χρώμα - διαφάνεια (δύο κατηγορικές)

μέγεθος κατηγοριοποιημένο - διαφάνεια (δύο κατηγορικές)

οργανισμός πιστοποίησης - διαφάνεια (δύο κατηγορικές)

μέγεθος κατηγοριοποιημένο - χρώμα (δύο κατηγορικές)

οργανισμός πιστοποίησης - μέγεθος κατηγοριοποιημένο (δύο κατηγορικές)

διαφάνεια-μέγεθος κατηγοριοποιημένο (δύο κατηγορικές)

Για τη μελέτη των δύο κατηγορικών θα χρησιμοποιήσουμε πίνακες συνάφειας¹⁰ κατά ομάδες και "στοιβαγμένα" ραβδογράμματα¹¹ και θα κάνουμε ελέγχους χ^2 ανεξαρτησίας.

10: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 10, ΠΙΝΑΚΕΣ: 9.1, 9.2, 10.1, 10.2, 11.1, 11.2
12.1, 12.2, 13.1, 13.2, 14.1, 14.2

11: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 11, ΔΙΑΓΡΑΜΜΑΤΑ: 7.1, 7.2, 8.1, 8.2, 9.1,
9.2, 10.1, 10.2, 11.1, 11.2, 12.1, 12.2

Υπάρχει σχέση:

διαφάνειας - οργανισμού πιστοποίησης ($p\text{-value} < 2.2 \times 10^{-16} < 0.05$)

διαφάνειας - μεγέθους σε καράτια ($p\text{-value} = 1.5004 \times 10^{-8} < 0.05$)

μεγέθους σε καράτια - οργανισμού πιστοποίησης ($p\text{-value} < 2.2 \times 10^{-16} < 0.05$)

Ενώ δεν υπάρχει σχέση:

χρώματος - μεγέθους σε καράτια ($p\text{-value} = 0.3446 > 0.05$)

χρώματος - οργανισμού πιστοποίησης ($p\text{-value} = 0.3616 > 0.05$)

χρώματος - διαφάνειας ($p\text{-value} = 0.3966 > 0.05$)¹²

Για τη μελέτη μίας ποσοτικής και μίας κατηγορικής εφαρμόζω ανάλυση διακύμανσης (έλεγχοι για ανεξάρτητα δείγματα). Επειδή τα κατάλοιπα δεν είναι κανονικά και το δείγμα μεγάλο εφαρμόζω μη παραμετρικό έλεγχο Kruskal Wallis test και κάνω έλεγχο για την ισότητα των διαμέσων. Διαγραμματικά κάνω boxplots ανά κατηγορία. Αναλυτικά οι παραπάνω έλεγχοι και τα διαγράμματα βρίσκονται στο παρακάτω κεφάλαιο.

12. βλέπε: ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 12, ΠΙΝΑΚΕΣ 15,16,17,18,19,20

Κεφάλαιο 4ο

Προβλεπτικά ή ερμηνευτικά μοντέλα

Έχοντας παρατηρήσει τις σχέσεις μεταξύ των μεταβλητών μας και έχοντας ξεχωρίσει αυτές που είναι ισχυρές μπορούμε πλέον να προβούμε στην κατασκευή κάποιων ερμηνευτικών μοντέλων σε μια προσπάθεια να κατασκευάσουμε ένα μοντέλο για την τιμολόγηση των διαμαντιών.

Αρχικά θα ξεκινήσουμε πραγματοποιώντας κάποιους έλεγχους υποθέσεων για να έχουμε πληρέστερη εικόνα των δεδομένων μας. Ένας πρώτος έλεγχος που μας ενδιαφέρει είναι να δούμε αν τα δεδομένα που έχουμε αφορούν ένα ακριβό προϊόν. Μια υποκειμενική θεώρηση ότι αν η διάμεση τιμή πώλησης των διαμαντιών είναι κοντά στα 4.000 δολάρια Σιγκαπούρης μπορούμε να πούμε ότι είναι ένα προσιτά οικονομικό προϊόν. Αλλιώς αν έχουμε υψηλότερη τιμή μπορούμε να πούμε ότι είναι ένα ακριβό προϊόν. Παίρνοντας τα αποτελέσματα από την R βλέπουμε πως η μηδενική υπόθεση του οικονομικού προϊόντος (όπως το ορίσαμε εμείς) απορρίπτεται.

(Wilcoxon test p-value = 0.0001444)¹³

Οπότε μπορούμε να ισχυριστούμε ότι πρόκειται για ένα ακριβό προϊόν.

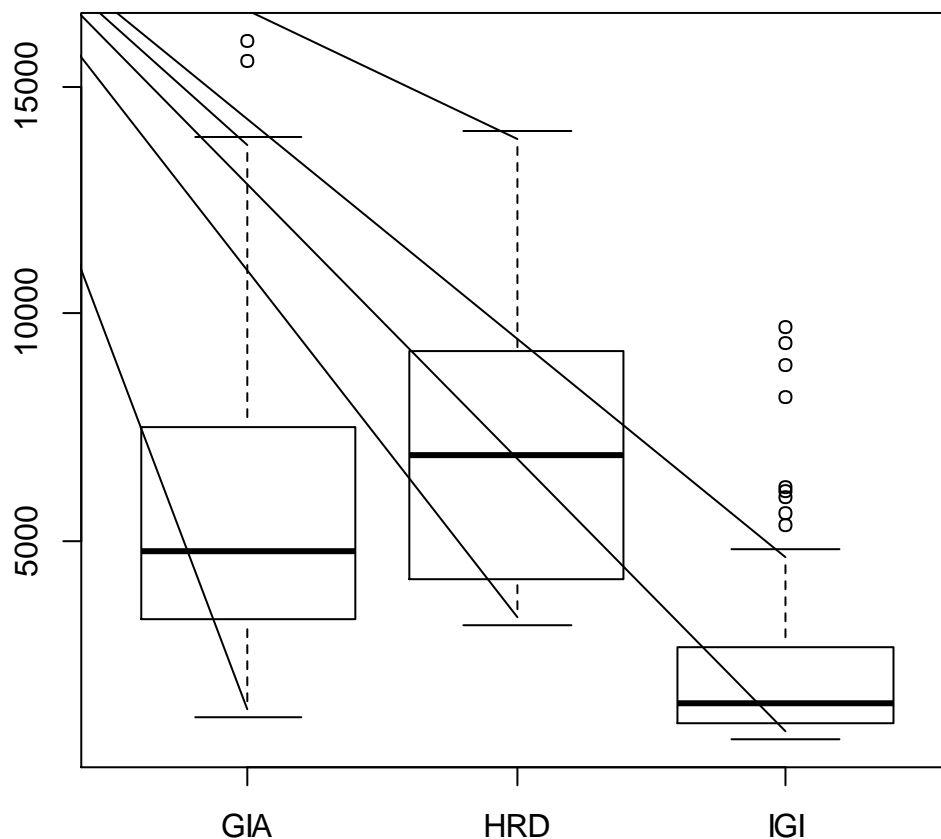
Αυτό που μας ενδιαφέρει στην συνέχεια είναι να δούμε ποιά είναι η τιμή του διαμαντιού για τα διάφορα επίπεδα διαφάνειας. Αυτό που θα πρέπει αρχικά να μελετήσουμε είναι αν υπάρχει κάποια στατιστικά σημαντική διαφορά μεταξύ της τιμής του διαμαντιού για τα διάφορα επίπεδα διαφάνειας. Ένας τρόπος να το κάνουμε αυτό είναι η ανάλυση της διακύμανσης αλλά δυστυχώς οι υποθέσεις της δεν ικανοποιούνται και για αυτό βασιζόμαστε στον αντίστοιχο μη παραμετρικό έλεγχο (Kruskal-Wallis, για την ισότητα των διαμέσων, μεγάλο δείγμα) οποίος μας δείχνει ότι υπάρχει διαφορά της τιμής στα διάφορα επίπεδα διαφάνειας (p-value = 1.405e-07)¹⁴.

13. βλέπε: ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 13, ΠΙΝΑΚΑΣ 21

14. βλέπε: ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 14, ΠΙΝΑΚΑΣ 22

Αύτη τη διαφορά θα τη δούμε καλύτερα με το διάγραμμα boxplot ανά επίπεδο διαφάνειας¹⁵. Ακολουθώντας την ίδια διαδικασία για να βρούμε άμα υπάρχει κάποια στατιστική σημαντική διαφορά μεταξύ της τιμής του διαμαντιού για τους διαφορετικούς οργανισμούς πιστοποίησης, βασιζόμαστε στον αντίστοιχο μη παραμετρικό έλεγχο (Kruskal-Wallis, για την ισότητα των διαμέσων, μεγάλο δείγμα) ο οποίος μας δείχνει ότι υπάρχει διαφορά τιμής για τους διαφορετικούς οργανισμούς πιστοποίησης ($p\text{-value} < 22e-16$) (δεν ικανοποιούνται υποθέσεις της ανάλυσης διακύμανσης). Ακόμα βλέπουμε τη διαφορά, στο αντίστοιχο boxplot ανά οργανισμό πιστοποίησης¹⁶ όπως φαίνεται παρακάτω:

ΣΧΕΣΗ ΤΙΜΗΣ ΜΕ ΟΡΓΑΝΙΣΜΟ ΠΙΣΤΟΠΟΙΗΣΗΣ



Από το παραπάνω διάγραμμα παρατηρούμε ότι για τον οργανισμό πιστοποίησης IGI η τιμή των διαμαντιών είναι αρκετά μικρότερη από ότι στους υπόλοιπους οργανισμούς πιστοποίησης (γιατί τα περισσότερα είναι μεγέθους 1, και είναι διαφάνειας IF). Ενώ για τον οργανισμό πιστοποίησης HRD οι τιμές είναι υψηλότερες.

Για να βρούμε αν υπάρχει κάποια στατιστική σημαντική διαφορά της τιμής του διαμαντιού για τα διάφορα χρώματα βασιζόμαστε στον αντίστοιχο μη παραμετρικό έλεγχο (Kruskal-Wallis, για την ισότητα των διαμέσων, μεγάλο δείγμα), ο οποίος μας δείχνει ότι δεν υπάρχει διαφορά τιμής για τα διαφορετικά χρώματα διαμαντιών ($p\text{-value} = 0.388$) (δεν ικανοποιούνται οι υποθέσεις της ανάλυσης διακύμανσης). Ακόμα βλέπουμε τη διαφορά, στο αντίστοιχο boxplot ανά χρώμα¹⁷.

Για να βρούμε αν υπάρχει κάποια στατιστικά σημαντική διαφορά της τιμής του διαμαντιού, για τα διαφορετικά επίπεδα της κατηγοροποιημένης μεταβλητής του μεγέθους σε καράτια, βασιζόμαστε στον αντίστοιχο μη παραμετρικό έλεγχο (Kruskal-Wallis, για την ισότητα των διαμέσων, μεγάλο δείγμα), ο οποίος μας δείχνει ότι υπάρχει διαφορά τιμής για τα διαφορετικά μεγέθη διαμαντιών ($p\text{-value} < 2.2e-16$) (δεν ικανοποιούνται οι υποθέσεις της ανάλυσης διακύμανσης). Ακόμα βλέπουμε τη διαφορά, στο αντίστοιχο boxplot ανά μέγεθος¹⁸, όπως φαίνεται στο διάγραμμα της επόμενης σελίδας. Παρατηρούμε ότι το μέγεθος είναι καθοριστικός παράγοντας για τον υπολογισμό της τιμής του διαμαντιού. Όσο μεγαλύτερο είναι το μέγεθος του διαμαντιού, τόσο ακριβότερο είναι.

Θα μελετήσουμε τη σχέση που έχουν οι δύο ποσοτικές μεταβλητές της τιμής του διαμαντιού με το μέγεθός του σε καράτια.

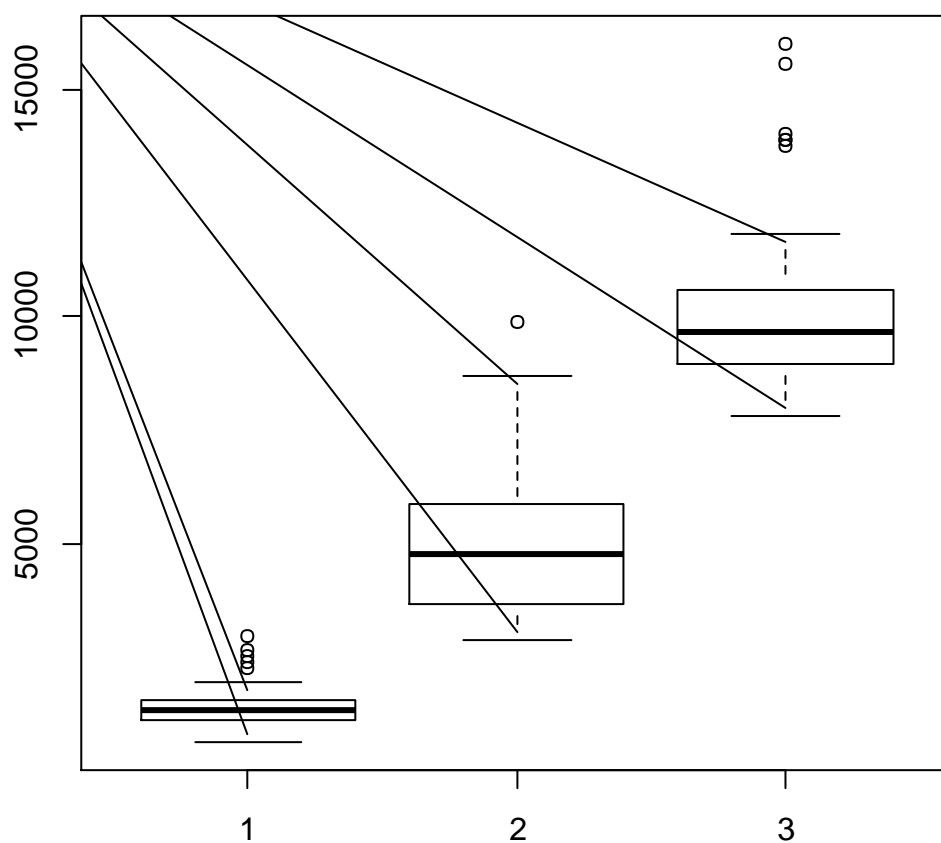
Υπάρχει έντονη γραμμική σχέση ανάμεσα στις δύο αυτές μεταβλητές και αυτό φαίνεται από τον συντελεστή συσχέτισης ($\text{cor}(x, y) = 0.9447$)¹⁹

17: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 17, ΔΙΑΓΡΑΜΜΑ 14, ΠΙΝΑΚΑΣ 24

18: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 18, ΠΙΝΑΚΑΣ 25

19: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 19

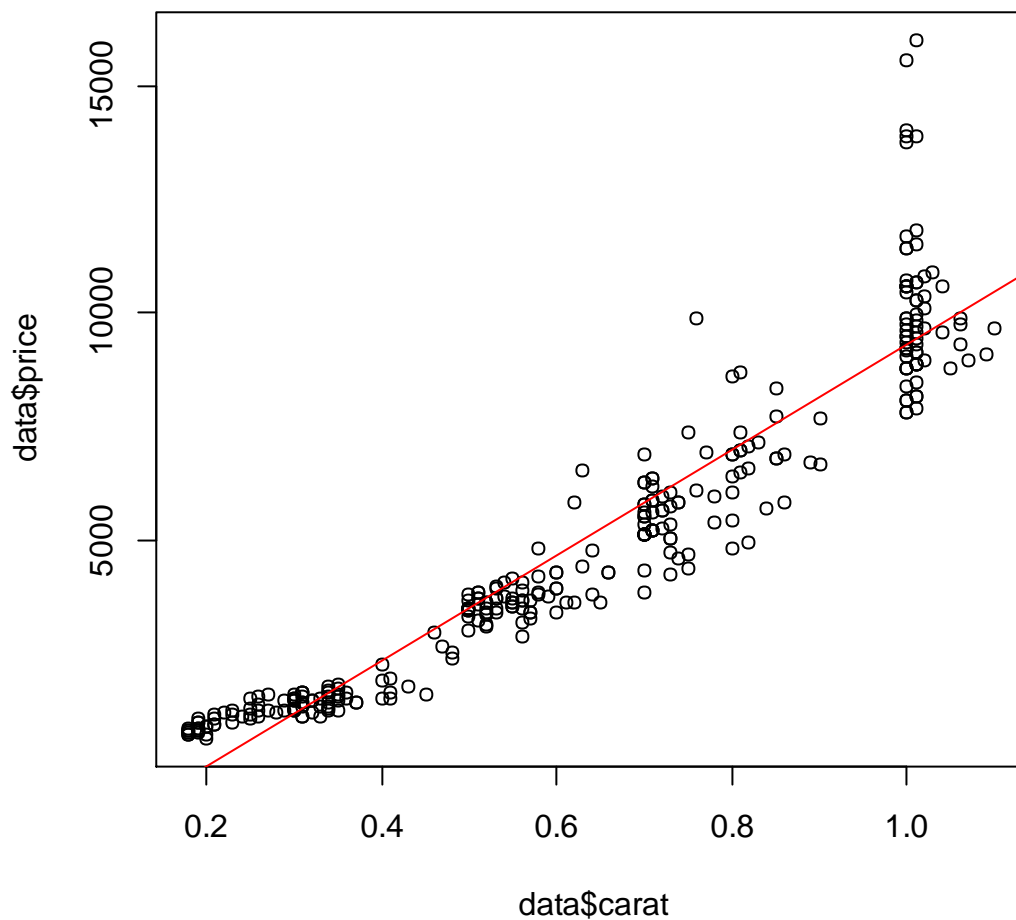
ΣΧΕΣΗ ΤΙΜΗΣ ΜΕΓΕΘΟΥΣ



Διαγραμματικά φαίνεται η συσχέτιση αυτή στο παρακάτω σχήμα²⁰.

Ακόμα στο διάγραμμα φαίνονται κάποιες ακραίες τιμές.

²⁰:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 20



Εφαρμόζοντας το απλό γραμμικό μοντέλο για να βρω τη σχέση της τιμής πώλησης των διαμαντιών με το μέγεθος σε καράτια βρίσκω:

τιμή πώλησης διαμαντιών = $-2298.4 + 11598.9 * \text{μέγεθος σε καράτια}$.

Το μοντέλο προσαρμόζεται πολύ καλά στα δεδομένα (έχει υψηλό $R^2_{\text{adjusted}} = 0.8922$) αλλά δυστυχώς τα αποτελέσματα δεν μπορούν να θεωρηθούν έγκυρα αφού δεν πληρούνται όλες οι προϋποθέσεις του μοντέλου (τα κατάλοιπα δεν είναι κανονικά)²¹.

21:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 21, ΠΙΝΑΚΑΣ 26, ΔΙΑΓΡΑΜΜΑ 15

Δοκιμάζω μετασχηματισμούς²² για να βρω καλύτερο μοντέλο(δοκιμάζω και κεντροποίηση για να έχει ερμηνεία η σταθερά). Από αυτούς προτιμότερος είναι ο ακόλουθος: (οι υπόλοιποι δεν έχουν κανονικά κατάλοιπα)

$$\log(\text{τιμή πώλησης διαμαντιών}) = 6.445 + 2.84 * \text{μέγεθος σε καράτια}$$

$$\text{ή τιμή πώλησης διαμαντιών} = e^{6.445} * e^{2.84 * \text{μέγεθος σε καράτια}}.$$

Το μοντέλο προσαρμόζεται πολύ καλά στα δεδομένα έχει υψηλό $R^2_{\text{adjusted}} = 0.9353$. Από τις προϋποθέσεις ικανοποιούνται οι ακόλουθες:

i) έχει κανονικά κατάλοιπα (p-value=0.05108)

ii)έχει ομοσκεδαστικά κατάλοιπα²³ (p-value=0.05131)

iii)υπάρχει πρόβλημα ανεξαρτησίας καταλοίπων ²⁴(p-value=0)

Ακόμα δοκιμάζουμε μετασχηματισμό box-cox²⁵.

Τέλος εφαρμόζουμε πολυωνυμική παλινδρόμηση²⁶. Δεν είναι πολύ καλύτερη από το γραμμικό μοντέλο και επειδή έχει δύσκολη ερμηνεία δεν την προτιμώ.

22:βλέπε ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 22,ΠΙΝΑΚΑΣ 27

23:βλέπε ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 23,ΠΙΝΑΚΑΣ 28,ΔΙΑΓΡΑΜΜΑΤΑ 16,17

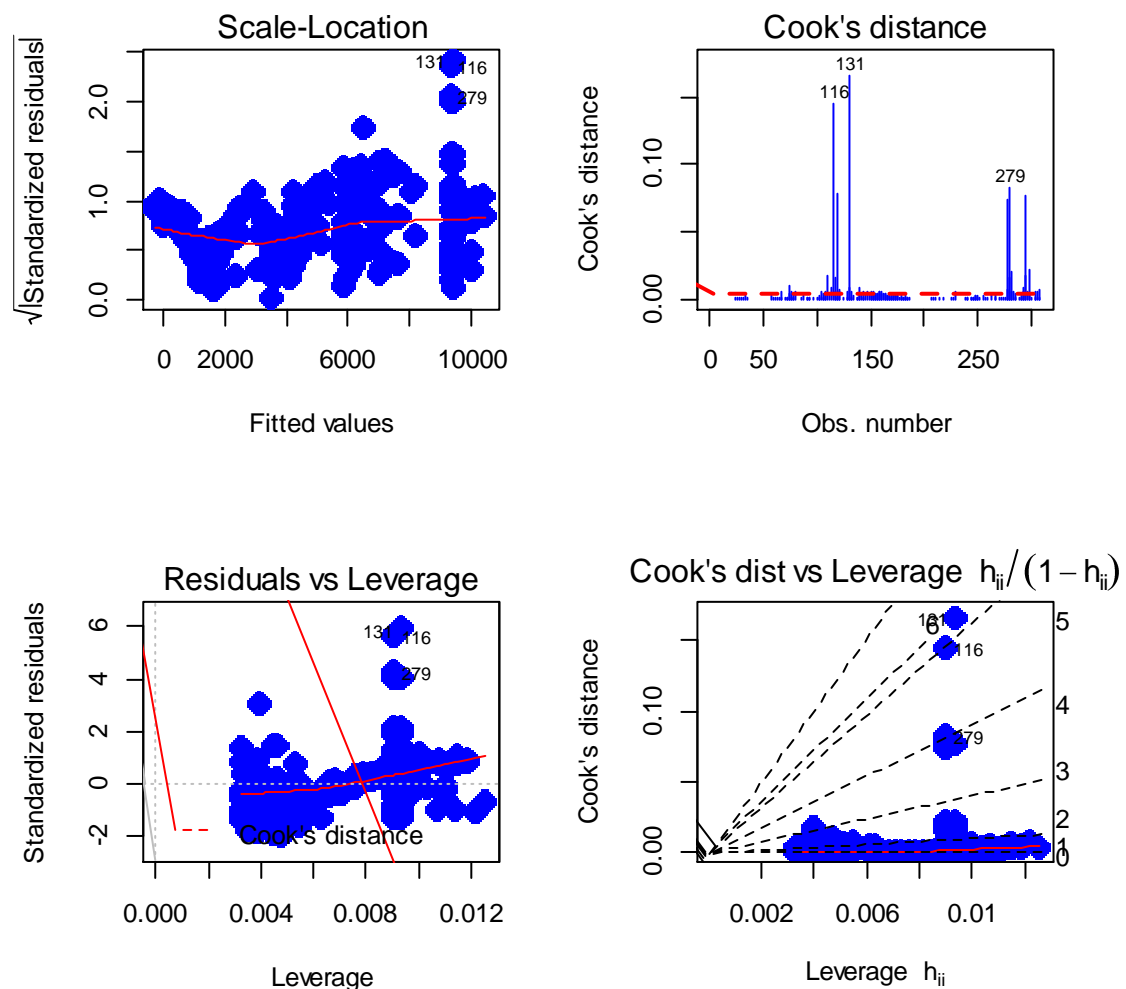
24:βλέπε ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 24,ΠΙΝΑΚΑΣ 29

25:βλέπε ΠΑΡΑΡΤΗΜΑ , ΕΝΤΟΛΗ 25,ΔΙΑΓΡΑΜΜΑ 18

26:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 26,ΠΙΝΑΚΑΣ 30

Τώρα εντοπίζουμε την ύπαρξη ακραίων τιμών²⁷. Εντοπίζουμε ακόμα τα σημεία επιρροής²⁸.

Σαν ακραίες τιμές (ή σημεία επιρροής) θεωρώ τις παρατηρήσεις 131, 116, 279. Πρόκειται για μεγάλα και ακριβά διαμάντια. Αυτά φαίνονται στο παρακάτω διάγραμμα.



27:βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 27

28:βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 28

Δουλεύοντας με δύο ποσοτικές μεταβλητές δεν καταλήξαμε σε ένα μοντέλο που να μας ικανοποιεί πλήρως. Σίγουρα η τιμή του διαμαντιού εξαρτάται πολύ από το μέγεθος.

Έχοντας κάνει την παραπάνω προεργασία μπορούμε να ξεκινήσουμε να δοκιμάζουμε διάφορα μοντέλα για την περιγραφή και στην συνέχεια για την πρόβλεψη της τιμής των διαμαντιών. Θα προσπαθήσουμε να εκφράσουμε την ποσοτική μεταβλητή της τιμής του διαμαντιού συναρτήσει των 4 κατηγορικών μεταβλητών (μέγεθος, διαφάνεια, χρώμα, οργανισμό πιστοποίησης) και των αλληλεπιδράσεών τους.

Δοκιμάζοντας το μοντέλο που έχει τις 4 κατηγορικές χωρίς τις αλληλεπιδράσεις παίρνουμε μοντέλο με υψηλό $R^2_{\text{adjusted}}=0.8895$, αλλά δεν έχει κανονικά κατάλοιπα²⁹(S.W. p.value=0.0009395, K.S. p.value=0.000806).

Μπορούμε να εξετάσουμε αν υπάρχει στατιστικά σημαντική διαφοροποίηση του μοντέλου από το σταθερό. Ο έλεγχος δείχνει ότι υπάρχει διαφοροποίηση, αλλά δυστυχώς δεν ισχύουν οι προϋποθέσεις (δεν είναι κανονικά τα κατάλοιπα)³⁰.

Δοκιμάζοντας και άλλα μοντέλα με 2 κατηγορικές ή με 3 κατηγορικές χωρίς αλληλεπιδράσεις, παίρνουμε μοντέλα που δεν έχουν κανονικά κατάλοιπα³¹.

Δοκιμάζοντας και άλλα μοντέλα με αλληλεπιδράσεις 2 κατηγορικών, παίρνουμε μοντέλα που δεν έχουν κανονικά κατάλοιπα³².

Τελικά συνεχίζοντας το ψάξιμο καταλήγω σε ένα μοντέλο που περιέχει τις 4 κατηγορικές και τις αλληλεπιδράσεις της διαφάνειας με το χρώμα. Αυτό το μοντέλο φαίνεται να ικανοποιεί τις προϋποθέσεις (κανονικότητα και ομοσκεδαστικότητα των καταλοίπων)³³.

29: βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 29, ΠΙΝΑΚΑΣ 31

30:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 30, ΠΙΝΑΚΑΣ 32

31:βλέπε ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 31,ΠΙΝΑΚΑΣ 33

32:βλέπε ΠΑΡΑΡΤΗΜΑ,ΕΝΤΟΛΗ 32,ΠΙΝΑΚΑΣ 34

33:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 33, ΠΙΝΑΚΑΣ 35,ΔΙΑΓΡΑΜΜΑ 19

Οι συντελεστές που χρειάζονται για την ερμηνεία του μοντέλου βρίσκονται στον παρακάτω πίνακα:

	estimate	p-value
(Intercept)	6710.9	7.89e-10 ***
car2	3174.1	< 2e-16 ***
car3	8218.6	< 2e-16 ***
bodyHRD	329.5	0.040504 *
bodyIGI	-669.5	0.001062 **
colourE	-4578.8	8.80e-05 ***
colourF	-3884.8	0.000573 ***
colourG	-4864.3	1.44e-05 ***
colourH	-5045.9	6.21e-06 ***
colourI	-5057.5	8.78e-06 ***
clarityVS1	-4485.3	7.42e-05 ***
clarityVS2	-5901.9	5.71e-06 ***
clarityVVS1	865.5	0.500463
clarityVVS2	-2606.4	0.026648 *
colourE:clarityVS1	3968.9	0.001672 **
colourF:clarityVS1	3434.5	0.004810 **
colourG:clarityVS1	3991.7	0.001128 **

colourH:clarityVS1	4486.0	0.000259 ***
colourI:clarityVS1	3407.0	0.006596 **
colourE:clarityVS2	5823.0	5.57e-05 ***
colourF:clarityVS2	4845.0	0.000459 ***
colourG:clarityVS2	5005.8	0.000298 ***
colourH:clarityVS2	5410.4	7.71e-05 ***
colourI:clarityVS2	4786.0	0.000690 ***
colourE:clarityVVS1	817.3	0.573216
colourF:clarityVVS1	-1690.7	0.215366
colourG:clarityVVS1	-858.7	0.528464
colourH:clarityVVS1	-1529.3	0.264936
colourI:clarityVVS1	-1507.4	0.287129
colourE:clarityVVS2	2682.1	0.041277 *
colourF:clarityVVS2	1465.5	0.244867
colourG:clarityVVS2	2456.7	0.049374 *
colourH:clarityVVS2	2659.8	0.033745 *
colourI:clarityVVS2	1896.3	0.138067

Από τον παραπάνω πίνακα μπορούμε να βγάλουμε κάποια συμπεράσματα:

Η αναμενόμενη τιμή για διαμάντι μικρού μεγέθους (1),οργανισμό πιστοποίησης GIA, χρώματος D και διαφάνειας IF είναι 6710.9 \$.

Ένα διαμάντι μεγέθους μεσαίου (2) αναμένεται να είναι 3174.1\$ ακριβότερο από ένα διαμάντι μικρού μεγέθους (1) όταν έχουν το ίδιο χρώμα, την ίδια διαφάνεια και τον ίδιο οργανισμό πιστοποίησης.

Ένα διαμάντι μεγέθους μεγάλου (3) αναμένεται να είναι 8218.6 \$ ακριβότερο από ένα διαμάντι μικρού μεγέθους (1) όταν έχουν το ίδιο χρώμα, την ίδια διαφάνεια και τον ίδιο οργανισμό πιστοποίησης.

Ένα διαμάντι από τον οργανισμό πιστοποίησης HRD αναμένεται να είναι 329.5\$ ακριβότερο από ένα διαμάντι του οργανισμού πιστοποίησης GIA όταν έχουν το ίδιο μέγεθος, το ίδιο χρώμα και την ίδια διαφάνεια.

Ένα διαμάντι από τον οργανισμό πιστοποίησης IGI αναμένεται να είναι 669.5\$ πιο φθηνό από ένα διαμάντι του οργανισμού πιστοποίησης GIA όταν έχουν το ίδιο μέγεθος, το ίδιο χρώμα και την ίδια διαφάνεια.

Η ερμηνεία των υπόλοιπων εκτιμήσεων είναι πιο πολύπλοκη . Αυτό κάνει το μοντέλο μας λίγο δύσκολο στην ερμηνεία του.

Ακόμη το μοντέλο προσαρμόζεται πολύ καλά στα δεδομένα γιατί έχει υψηλό $R^2_{\text{adjusted}}=0.9068$.

Ακόμη στο ψάξιμο του μοντέλου και για να αποφανθούμε για το ποιες μεταβλητές θα διώξουμε μπορούμε να απευθυνθούμε στις μεθόδους FORWARD SELECTION, BANCWARD ELIMINATION και STEPWISE REGRESSION. Επίσης θα χρησιμοποιήσουμε μεθόδους οι οποίες επιλέγουν το βέλτιστο μοντέλο με βάση το κριτήριο Akaike ή αλλιώς AIC³⁴.

Τέλος θα προσπαθήσουμε να δούμε σε τι συμπεράσματα θα οδηγηθούμε αν αφαιρέσουμε κάποιες ακραίες τιμές. Αποφασίζουμε να αφαιρέσουμε 15 τιμές³⁵.

Θα αναζητήσουμε μοντέλο για το καινούργιο σετ δεδομένων.

34: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 34, ΠΙΝΑΚΑΣ 36

35: βλέπε ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΗ 35, ΠΙΝΑΚΑΣ 37, ΔΙΑΓΡΑΜΜΑ 20

Δοκιμάζοντας το μοντέλο με τις 4 κατηγορικές και χωρίς αλληλεπιδράσεις παίρνουμε μοντέλο που ικανοποιεί τις προϋποθέσεις ,έχει καλή προσαρμογή (υψηλό R^2_{adjusted})³⁶

Οι συντελεστές που χρειάζονται για την ερμηνεία του μοντέλου βρίσκονται στον παρακάτω πίνακα:

	estimate	p-value
(Intercept)	2175.00	6.65e-07 ***
car2	3256.16	< 2e-16 ***
car3	7919.52	< 2e-16 ***
colourE	15.35	0.96637
colourF	-183.88	0.59608
colourG	-439.05	0.21744
colourH	-499.43	0.16766
colourI	-973.90	0.00991 **
bodyHRD	312.71	0.04940 *
bodyIGI	-462.62	0.01733 *
clarityVS1	-337.09	0.16293
clarityVS2	-407.77	0.11125
clarityVVS1	-157.77	0.50802
clarityVVS2	-161.14	0.46494

36:βλέπε ΠΑΡΑΡΤΗΜΑ, ΕΝΤΟΛΗ 36,ΠΙΝΑΚΑΣ 37

Κεφάλαιο 5ο

Συμπεράσματα και συζήτηση

Από τη μελέτη που κάναμε και τα δύο μοντέλα που προτείνουμε έχουμε τα ακόλουθα συμπεράσματα:

Βασικός παράγοντας που επηρεάζει την τιμή ενός διαμαντιού είναι το μέγεθός του. Όμως υπάρχουν συσχετίσεις ανάμεσα στον οργανισμό πιστοποίησης ,στη διαφάνεια, στο χρώμα. Τα περισσότερα διαμάντια μικρού μεγέθους προέρχονται από τον οργανισμό πιστοποίησης IGI, και έχουν διαφάνεια IF. Πολλά διαμάντια μεσαίου μεγέθους προέρχονται από τον οργανισμό πιστοποίησης GIA και HRD. Οι ακραίες τιμές είναι διαμάντια πολύ μεγάλα και πολύ ακριβά. Τα μεγάλα και ακριβά διαμάντια είναι λίγα και αλλάζουν τα αποτελέσματα της ερευνάς μας .Έτσι ψάξαμε και μοντέλο μετά από την αφαίρεσή τους ,που να αφορά την τιμολόγηση των περισσότερων διαμαντιών που είναι μεσαίου και μικρού μεγέθους.

ΠΑΡΑΡΤΗΜΑ

ΕΝΤΟΛΕΣ ΤΗΣ R

ΕΝΤΟΛΗ 1: Διαβάζω τα δεδομένα

```
data<-read.table("E:/02_diamonds.csv",sep=",")  
  
names(data)<-c("index","carat","colour","clarity","body","price")  
  
head(data)  
  
str(data)  
  
data$price<-as.numeric(data$price)  
  
dim(data)
```

ΕΝΤΟΛΗ 2: έλεγχος κανονικότητας για τη ποσοτική μεταβλητή: "τιμή διαμαντιού"

```
qqnorm(data$price,main="normal qq plot for price")  
  
qqline(data$price,col=2)  
  
shapiro.test(data$price)#oxi kanoniko digma#  
  
έλεγχος κανονικότητας για τη ποσοτική μεταβλητή: "μέγεθος διαμαντιού"  
  
qqnorm(data$carat,main="normal qq plot for carat")  
  
qqline(data$carat,col=2)  
  
shapiro.test(data$carat)#oxi kanoniko digma#
```

ΕΝΤΟΛΗ 3: Ιστογράμματα και boxplot

```
par(mfrow=c(2,1))  
  
hist(data$price)  
  
boxplot(data$price)  
  
par(mfrow=c(2,1))  
  
hist(data$carat)
```

```
boxplot(data$carat)
```

ΕΝΤΟΛΗ 4: Διαγράμματα βιολιου

```
install.packages("sm")
```

```
library(sm)
```

```
install.packages("vioplot")
```

```
library(vioplot)
```

```
vioplot(data$price,names="price")
```

```
vioplot(data$carat,names="carat")
```

ΕΝΤΟΛΗ 5: Περιγραφικά μέτρα

```
install.packages("psych")
```

```
library(psych)
```

```
round(t(describe(data$price)),1)
```

```
round(t(describe(data$carat)),1)
```

ΕΝΤΟΛΗ 6: Κατηγοροποίηση ποσοτικής μεταβλητής καράτια

```
data$car<-1*(data$carat<0.5)+2*(data$carat>=0.5&data$carat<1)+3*(data$carat>=1)
```

```
data$car<-as.factor(data$car)
```

ΕΝΤΟΛΗ 7: Πίτες

```
par(mfrow=c(2,2))
```

```
pie(table(data$car))
```

```
title("ΚΑΤΗΓΟΡΙΕΣ ΜΕΓΕΘΟΥΣ")
```

```
pie(table(data$colour))
```



```
title("ΚΑΤΗΓΟΡΙΕΣ ΧΡΩΜΑΤΟΣ")
```

```
pie(table(data$clarity))
```

```
title("ΚΑΤΗΓΟΡΙΕΣ ΔΙΑΦΑΝΕΙΑΣ")
```

```
pie(table(data$body))
```

```
title("ΚΑΤ. ΟΡΓ. ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

ΕΝΤΟΛΗ 8:Ραβδογράμματα

```
par(mfrow=c(2,2))
```

```
barplot(round(prop.table(table(data$car)),2),col=3)
```

```
title("ΚΑΤΗΓΟΡΙΕΣ ΜΕΓΕΘΟΥΣ")
```

```
barplot(round(prop.table(table(data$colour)),2),col=3)
```

```
title("ΚΑΤΗΓΟΡΙΕΣ ΧΡΩΜΑΤΟΣ")
```

```
barplot(round(prop.table(table(data$clarity)),2),col=3)
```

```
title("ΚΑΤΗΓΟΡΙΕΣ ΔΙΑΦΑΝΕΙΑΣ")
```

```
barplot(round(prop.table(table(data$body)),2),col=3)
```

```
title("ΚΑΤ. ΟΡΓ. ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

ΕΝΤΟΛΗ 9:Πίνακες συχνοτήτων

```
table(data$car)
```

```
round(prop.table(table(data$car)),2)
```

```
table(data$colour)
```

```
round(prop.table(table(data$colour)),2)
```

```
table(data$clarity)
```

```
round(prop.table(table(data$clarity)),2)
```

```
table(data$body)
```

```
round(prop.table(table(data$body)),2)
```

ΕΝΤΟΛΗ 10:Πίνακες συνάφειας

```
tab1<-table(data$colour,data$clarity)
```

```
install.packages("gmodels")
```

```
library(gmodels)
```

```
CrossTable(data$colour,data$clarity)
```

```
tab2<-table(data$colour,data$car)
```

```
CrossTable(data$colour,data$car)
```

```
tab3<-table(data$colour,data$body)
```

```
CrossTable(data$colour,data$body)
```

```
tab4<-table(data$clarity,data$body)
```

```
CrossTable(data$clarity,data$body)
```

```
tab5<-table(data$car,data$body)
```

```
CrossTable(data$car,data$body)
```

```
tab6<-table(data$clarity,data$car)
```

```
CrossTable(data$clarity,data$car)
```

ΕΝΤΟΛΗ 11:Ραβδογράμματα στοιβαγμένα και κατά ομάδες

```
mycol1<-1:length(levels(data$colour))
```

```
mycol2<-1:length(levels(data$clarity))
```

```
mycol3<-1:length(levels(data$car))
```

```
barplot(tab1,beside=T,col=mycol1,legend=levels(data$colour),xlab="ΔΙΑΦΑΝΕΙΑ")
```

```
barplot(tab1,beside=F,col=mycol1,legend=levels(data$colour),xlab="ΔΙΑΦΑΝΕΙΑ")
```

```
title("ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΔΙΑΦΑΝΕΙΑΣ")
```

```
barplot(tab2,beside=T,col=mycol1,legend=levels(data$colour),xlab="ΜΕΓΕΘΟΣ  
ΠΕΤΡΑΣ")
```

```
barplot(tab2,beside=F,col=mycol1,legend=levels(data$colour),xlab="ΜΕΓΕΘΟΣ  
ΠΕΤΡΑΣ")
```

```
title("ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΜΕΓΕΘΟΥΣ")
```

```
barplot(tab3,beside=T,col=mycol1,legend=levels(data$colour),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab3,beside=F,col=mycol1,legend=levels(data$colour),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
title("ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab4,beside=T,col=mycol2,legend=levels(data$clarity),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab4,beside=F,col=mycol2,legend=levels(data$clarity),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
title("ΣΧΕΣΗ ΔΙΑΦΑΝΕΙΑΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab5,beside=T,col=mycol3,legend=levels(data$car),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab5,beside=F,col=mycol3,legend=levels(data$car),xlab="ΟΡΓΑΝΙΣΜΟΣ  
ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
title("ΣΧΕΣΗ ΜΕΓΕΘΟΥΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ")
```

```
barplot(tab6,beside=T,col=mycol2,legend=levels(data$clarity),xlab="ΜΕΓΕΘΟΣ")
```

```
barplot(tab6,beside=F,col=mycol2,legend=levels(data$clarity),xlab="ΜΕΓΕΘΟΣ")
```

```
title("ΣΧΕΣΗ ΔΙΑΦΑΝΕΙΑΣ ΜΕΓΕΘΟΥΣ")
```

ΕΝΤΟΛΗ 12: Έλεγχοι χ^2

```
chisq.test(data$colour,data$clarity)
```

```
chisq.test(data$colour,data$car)
```

```
chisq.test(data$colour,data$body)
```

```
chisq.test(data$clarity,data$body)
```

```
chisq.test(data$car,data$body)
```

```
chisq.test(data$clarity,data$car)
```

ΕΝΤΟΛΗ 13:Έλεγχος αν το προϊόν είναι ακριβό

```
wilcox.test(data$price,alternative="greater",mu=4000)
```

ΕΝΤΟΛΗ 14:Σχέση τιμής- διαφάνειας

```
summary(aov(price~clarity,data=data))
```

```
shapiro.test(aov(price~clarity,data=data)$res)#oxi kanonika katalipa#
```

```
leveneTest(price~clarity,data=data,center=mean)#oxi omoskedastika katalipa#
```

```
#megalo digma kano elegxo gia tin isotita ton diameson#
```

```
kruskal.test(price~clarity,data=data)#aporipto Ho ara iparxi diafora#
```

```
boxplot(data$price~data$clarity)
```

ΕΝΤΟΛΗ 15:Σχέση τιμής- διαφάνειας

```
#sxesi timis me diafania#
```

```
summary(aov(price~clarity,data=data))
```

```
shapiro.test(aov(price~clarity,data=data)$res)#oxi kanonika katalipa#
```

```
leveneTest(price~clarity,data=data,center=mean)#oxi omoskedastika katalipa#
```

```
#megalo digma kano elegxo gia tin isotita ton diameson#
```

```
kruskal.test(price~clarity,data=data)#aporipto Ho ara iparxi diafora#
```

```
boxplot(data$price~data$clarity)
```

```
title("ΣΧΕΣΗ ΤΙΜΗΣ ΚΑΙ ΔΙΑΦΑΝΕΙΑΣ")
```

ΕΝΤΟΛΗ 16:Σχέση τιμής -οργανισμός πιστοποίησης

#sxesi timis me organismo pistopiisis#

```
summary(aov(price~body,data=data))
```

```
shapiro.test(aov(price~body,data=data)$res)#oxi kanonika katalipa#
```

```
leveneTest(price~body,data=data,center=mean)#oxi omoskedastika katalipa#
```

#megalo digma kano elegxo mi parametriko gia tin isotita ton diameson#

```
kruskal.test(price~body,data=data)# aporipto Ho ara iparxi diafora#
```

```
boxplot(data$price~data$body)
```

ΕΝΤΟΛΗ 17:Σχέση τιμής-χρώμα

#sxesi timis me xroma#

```
summary(aov(price~colour,data=data))
```

```
shapiro.test(aov(price~colour,data=data)$res)#oxi kanonika katalipa#
```

```
leveneTest(price~colour,data=data,center=mean)#oxi omoskedastika katalipa#
```

#megalo digma kano elegxo gia tin isotita ton diameson#

```
kruskal.test(price~colour,data=data)# den aporipto Ho ara den iparxi diafora#
```

```
boxplot(data$price~data$colour)
```

ΕΝΤΟΛΗ 18

#sxesi timis me katigoriki metabliti karatia#

```
summary(aov(price~car,data=data))
```

```
shapiro.test(aov(price~car,data=data)$res)#oxi kanonika katalipa#
```

```
leveneTest(price~car,data=data,center=mean)#oxi omoskedastika katalipa#
```

#megalo digma kano elegxo mi parametriko gia tin isotita ton diameson#

```
kruskal.test(price~car,data=data)#aporipto Ho ara iparxi diafora#
```

```
boxplot(data$price~data$car)
```

ΕΝΤΟΛΗ 19: Συντελεστές συσχέτισης

```
cor(data$price,data$carat)
```

```
cor(data$price,data$carat,method="kendal")
```

```
cor(data$price,data$carat,method="spearman")
```

ΕΝΤΟΛΗ 20

```
plot(data$carat,data$price)
```

```
abline(lm(data$price~data$carat),col=2)
```

ΕΝΤΟΛΗ 21

```
l1<-lm(data$price~data$carat)
```

```
summary(l1)
```

```
#elegxos kanonikotitas katalipon#
```

```
qqnorm(l1$res)
```

```
qqline(l1$res)
```

```
shapiro.test(l1$res)#oxi kanonika katalipa#
```

```
lillie.test(l1$res)#oxi kanonika katalipa#
```

ΕΝΤΟΛΗ 22: Δοκιμάζω μετασχηματισμούς

```
#dokimazo metasximatismous#
```

```
l2<-lm(log(data$price)~data$carat)
```

```
l3<-lm(log(data$price)~log(data$carat))
```

```
l4<-lm(data$price~log(data$carat))
```

```
l5<-lm(I(1/data$price)~data$carat)
```

```
l6<-lm(data$price~I(1/data$carat))
```

```
l7<-lm(I(1/data$price)~I(1/data$carat))
```

```
l8<-lm(data$price~I(data$carat-mean(data$carat)))
```

```
shapiro.test(l2$res)
```

```
shapiro.test(l3$res)
```

```
shapiro.test(l4$res)
```

```
shapiro.test(l5$res)
```

```
shapiro.test(l6$res)
```

```
shapiro.test(l7$res)
```

```
shapiro.test(l8$res)
```

```
lillie.test(l2$res)
```

```
lillie.test(l3$res)
```

```
ENTOΛH 23
```

```
#elegxo ipolipes proipothesis gia l2 #
```

```
#elegxos omoskedastikotitas katalipon# ###Omoskedastikotha Kataloipwn#####
```

```
##rstandard vs X##,##rstandard vs Yhat##,##st.res.sq vs Yhat##
```

```
#res vs X
```

```
plot( data$carat, rstandard(l2) , ylim=c(-2.2,2.2), pch=16, cex=2, col='blue')
```

```
abline(h=1.96, col='red', lwd=2, lty=2)
```

```
abline(h=-1.96, col='red', lwd=2, lty=2)
```

```
#st.res vs Yhat
```

```
plot( l2$fit, rstandard(l2) , ylim=c(-2.2,2.2), pch=16, cex=2, col='blue')
```

```
abline(h=1.96, col='red', lwd=2, lty=2)
```

```
abline(h=-1.96, col='red', lwd=2, lty=2)
```

```
# Elegxos isothtas Diakumansewn ana tetarthmorio ths X #
```

```
x<-cut(data$carat-mean(data$carat), breaks= quantile(data$carat-  
mean(data$carat)),include.lowest=T)
```

```
table(x)
```

```
var.test(rstandard(l2),x)
```

ENTOAH 24

```
#elegxos anejartisias katalipon#
```

```
dwt(l2)
```

```
durbinWatsonTest(l2)
```

```
durbinWatsonTest(l2,max.lag=7)
```

ENTOAH 25

```
# Matasxhmatismoi Boc Cox#
```

```
library(MASS)
```

```
boxcox(l1, lambda=seq(-5,5,0.1) )
```

```
temp<-boxcox(l1, lambda=seq(-5,5,0.1) )
```

```
temp$x[ temp$y==max(temp$y) ]
```

ENTOAH 26

```
#polionimiki palindromisi#
```

```
hp<-lm(price~poly(carat,6,raw=T),data=data)
```

```
summary(hp)
```

```
AIC(hp)
```

```
BIC(hp)
```

ENTOAH 27

```
###Akraies times#
```



```
#plot Yhat vs st.res#
```

```
par(bty='l')
```

```
plot(data$carat-mean(data$carat), rstandard(l1) , ylim=c(-2.2,2.2), pch=16, cex=2,  
col='blue')
```

```
abline(h=1.96, col='red', lwd=2, lty=2)
```

```
abline(h=-1.96, col='red', lwd=2, lty=2)
```

ΕΝΤΟΛΗ 28

```
#Shmeia ephrrohs#
```

```
plot( l1, pch=16, cex=2, col='blue', add.smooth=T, which=3)
```

```
#cook distance
```

```
plot( l1, pch=16, cex=2, col='blue', which=4)
```

```
abline(h=4/(1078-1-1), col='red', lty=2, lwd=2)###empeiriko shmeio diaxwrismou  
4/(n-p-1)#
```

```
plot( l1, pch=16, cex=2, col='blue', which=5)
```

```
plot( l1, pch=16, cex=2, col='blue', which=6)
```

ΕΝΤΟΛΗ 29:Αναζήτηση μοντέλου

```
anova1<-lm(price~car+colour+clarity+body,data=data)
```

```
leveneTest(anova1$res~data$car*data$colour*data$body*data$clarity)
```

```
shapiro.test(anova1$res)
```

```
lillie.test(anova1$res)
```

ΕΝΤΟΛΗ 30:Σύγκριση σταθερού με πλήρες μοντέλο

```
null<-aov(price~1,data=data)
```

```
main.effects<-aov(price~car+colour+body+clarity,data=data)
```

```
anova(null,main.effects)
```

```
#Kanonikothta#
```

```
shapiro.test(main.effects$res)
```

```
lillie.test(main.effects$res)
```

ENTOΛΗ 31:Δοκιμάζω μοντέλα

```
#dokimazo me 2 h 3 katigorikes#
```

```
a2<-lm( price~car+body, data=data)
```

```
a3<-lm( price~colour+car, data=data)
```

```
a4<-lm( price~clarity+car, data=data)
```

```
a5<-lm( price~body+clarity, data=data)
```

```
a6<-lm( price~body+colour, data=data)
```

```
a7<-lm( price~clarity+colour, data=data)
```

```
a8<-lm( price~car+body+colour, data=data)
```

```
a9<-lm( price~car+body+clarity, data=data)
```

```
a10<-lm( price~colour+body+clarity, data=data)
```

```
a11<-lm( price~colour+car+clarity, data=data)
```

```
shapiro.test(a2$res)
```

```
shapiro.test(a3$res)
```

```
shapiro.test(a4$res)
```

```
shapiro.test(a5$res)
```

```
shapiro.test(a6$res)
```

```
shapiro.test(a7$res)
```

```
shapiro.test(a8$res)
```

```
shapiro.test(a9$res)
```

```
shapiro.test(a10$res)
```

```
shapiro.test(a11$res)
```

ΕΝΤΟΛΗ 32 :Δοκιμάζω μοντέλα με αλληλεπιδράσεις

```
#dokimazo me allilepidraseis#
```

```
a12<-lm(price~car*colour,data=data)
```

```
shapiro.test(a12$res)
```

```
a13<-lm(price~car*clarity,data=data)
```

```
shapiro.test(a13$res)
```

```
a14<-lm(price~car*body,data=data)
```

```
shapiro.test(a14$res)
```

```
a15<-lm(price~clarity*body,data=data)
```

```
shapiro.test(a15$res)
```

```
a16<-lm(price~clarity*colour,data=data)
```

```
shapiro.test(a16$res)
```

```
a17<-
```

```
lm(price~car*colour+car*clarity+car*body+clarity*body+clarity*colour+body*colour,data=data)
```

```
shapiro.test(a17$res)
```

ΕΝΤΟΛΗ 33":Βρήκα μοντέλο

```
a18<-lm(price~car+body+colour+clarity+colour*clarity,data=data)
```

```
shapiro.test(a18$res)
```

```
lillie.test(a18$res)
```

```
qqnorm(a18$res)
```

```
qqline(a18$res)
```

```
leveneTest(a18$res~data$car*data$colour*data$clarity*data$body)
```

ΕΝΤΟΛΗ 34:Επιλογή μοντέλου με AIC

```
mnull<-lm(price~1,data=data)
```

```
summary(step(mnull, scope=list(lower=mnull,upper=a18), direction='both'))
```

```
summary(step(mnull, scope=list(lower=mnull,upper=a18), direction='forward'))
```

ΕΝΤΟΛΗ 35:Αφαίρεση ακραίων τιμών

```
data1<-data[-c(131,116,279),]
```

```
l<-lm(data1$price~data1$carat)
```

```
par(mfrow=c(2,2))
```

```
plot( l, pch=16, cex=2, col='blue', which=4)
```

```
data2<-data1[-c(291,276,119),]
```

```
n<-lm(data2$price~data2$carat)
```

```
plot( n, pch=16, cex=2, col='blue', which=4)
```

```
data3<-data2[-c(289,277,291),]
```

```
k<-lm(data3$price~data3$carat)
```

```
plot( k, pch=16, cex=2, col='blue', which=4)
```

```
data4<-data3[-c(110,116,275),]
```

```
v<-lm(data4$price~data4$carat)
```

```
plot( v, pch=16, cex=2, col='blue', which=4)
```

```
data5<-data4[-c(118,139,293),]
```

ΕΝΤΟΛΗ 36:Μοντέλο με καινούργιο data set και έλεγχοι προϋποθέσεων

```
anova1<-lm(price~car+colour+body+clarity,data=data5)
```

```
leveneTest(anova1$res~data5$car*data5$colour*data5$body*data5$clarity)
```

```
shapiro.test(anova1$res)
```

```
lillie.test(anova1$res)
```

```
summary(anova1)
```

ΠΙΝΑΚΕΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΠΙΝΑΚΑΣ 1: Έλεγχος κανονικότητας ποσοτικής μεταβλητής τιμή

Shapiro-Wilk normality test

data: data\$price

W = 0.92629, p-value = 3.343e-11

ΠΙΝΑΚΑΣ 2 : Έλεγχος κανονικότητας ποσοτικής μεταβλητής μεγέθους

Shapiro-Wilk normality test

data: data\$carat

W = 0.93053, p-value = 8.404e-11

ΠΙΝΑΚΑΣ 3: Περιγραφικά μέτρα για την ποσοτική μεταβλητή "TIMH ΔΙΑΜΑΝΤΙΟΥ"

X1

vars 1.0

n 308.0

mean 5019.5

sd 3403.1

median 4215.0

trimmed 4739.6

mad 3977.8

min 638.0

max 16008.0

range 15370.0

skew 0.7

kurtosis -0.4

se 193.9

ΠΙΝΑΚΑΣ 4: Περιγραφικά μέτρα για την ποσοτική μεταβλητή "ΜΕΓΕΘΟΣ ΔΙΑΜΑΝΤΙΟΥ"

X1

vars 1.0

n 308.0

mean 0.6

sd 0.3

median 0.6

trimmed 0.6

mad 0.4

min 0.2

max 1.1

range 0.9

skew 0.0

kurtosis -1.3

se 0

ΠΙΝΑΚΑΣ 5: Πίνακας συχνοτήτων για το μέγεθος.

1 2 3

94 143 71

ΠΙΝΑΚΑΣ 6: Πίνακας συχνотήτων για το χρώμα.

D E F G H I

16 44 82 65 61 40

ΠΙΝΑΚΑΣ 7: Πίνακας συχνотήτων για τη διαφάνεια.

IF VS1 VS2 VVS1 VVS2

44 81 53 52 78

ΠΙΝΑΚΑΣ 8: Πίνακας συχνотήτων για τον Οργανισμό Πιστοποίησης.

GIA HRD IGI

151 79 78

ΠΙΝΑΚΑΣ 9.1: Πίνακας συνάφειας για τις μεταβλητές χρώμα-διαφάνεια

IF VS1 VS2 VVS1 VVS2

D 1 7 2 2 4

E 5 18 7 5 9

F 8 25 14 18 17

G 11 12 9 12 21

H 11 11 14 10 15

I 8 8 7 5 12

ΠΙΝΑΚΑΣ 9.2: Πίνακας συνάφειας για τις μεταβλητές χρώμα-διαφάνεια

| data\$clarity

data\$colour | IF | VS1 | VS2 | VVS1 | VVS2 | Row Total |

-----|-----|-----|-----|-----|-----|-----|

D | 1 | 7 | 2 | 2 | 4 | 16 |

H	11	11	14	10	15	61	
	0.600	1.585	1.169	0.009	0.013		
	0.180	0.180	0.230	0.164	0.246	0.198	
	0.250	0.136	0.264	0.192	0.192		
	0.036	0.036	0.045	0.032	0.049		
----- ----- ----- ----- ----- ----- -----							
I	8	8	7	5	12	40	
	0.914	0.603	0.002	0.455	0.345		
	0.200	0.200	0.175	0.125	0.300	0.130	
	0.182	0.099	0.132	0.096	0.154		
	0.026	0.026	0.023	0.016	0.039		
----- ----- ----- ----- ----- ----- -----							
Column Total	44	81	53	52	78	308	
	0.143	0.263	0.172	0.169	0.253		
----- ----- ----- ----- ----- ----- -----							

ΠΙΝΑΚΑΣ 10.1: Πίνακας συνάφειας για τις μεταβλητές χρώμα-μέγεθος

1 2 3

D 5 6 5

E 12 22 10

F 27 41 14

G 23 30 12

H 16 31 14

I 11 13 16

ΠΙΝΑΚΑΣ 10.2: Πίνακας συνάφειας για τις μεταβλητές χρώμα-μέγεθος

data\$car				
data\$colour	1	2	3	Row Total
----- ----- ----- ----- -----				
D	5	6	5	16
	0.003	0.275	0.466	
	0.312	0.375	0.312	0.052
	0.053	0.042	0.070	
	0.016	0.019	0.016	
----- ----- ----- ----- -----				
E	12	22	10	44
	0.152	0.121	0.002	
	0.273	0.500	0.227	0.143
	0.128	0.154	0.141	
	0.039	0.071	0.032	
----- ----- ----- ----- -----				
F	27	41	14	82
	0.156	0.225	1.272	
	0.329	0.500	0.171	0.266
	0.287	0.287	0.197	

		0.088		0.133		0.045			
----- ----- ----- ----- -----									
G		23		30		12		65	
		0.504		0.001		0.594			
		0.354		0.462		0.185		0.211	
		0.245		0.210		0.169			
		0.075		0.097		0.039			
----- ----- ----- ----- -----									
H		16		31		14		61	
		0.368		0.253		0.000			
		0.262		0.508		0.230		0.198	
		0.170		0.217		0.197			
		0.052		0.101		0.045			
----- ----- ----- ----- -----									
I		11		13		16		40	
		0.119		1.671		4.984			
		0.275		0.325		0.400		0.130	
		0.117		0.091		0.225			
		0.036		0.042		0.052			
----- ----- ----- ----- -----									
Column Total		94		143		71		308	
		0.305		0.464		0.231			
----- ----- ----- ----- -----									

ΠΙΝΑΚΑΣ 11.1: Πίνακας συνάφειας για τις μεταβλητές χρώμα-οργανισμός πιστοποίησης

GIA HRD IGI			
D	10	2	4
E	25	10	9
F	40	19	23
G	26	17	22
H	29	22	10
I	21	9	10

ΠΙΝΑΚΑΣ 11.2: Πίνακας συνάφειας για τις μεταβλητές χρώμα-οργανισμός πιστοποίησης

data\$body				
data\$colour	GIA	HRD	IGI	Row Total
----- ----- ----- ----- -----				
D	10	2	4	16
	0.593	1.079	0.001	
	0.625	0.125	0.250	0.052
	0.066	0.025	0.051	
	0.032	0.006	0.013	
----- ----- ----- ----- -----				
E	25	10	9	44
	0.545	0.146	0.412	

		0.568		0.227		0.205		0.143	
		0.166		0.127		0.115			
		0.081		0.032		0.029			
	-----	-----	-----	-----	-----	-----	-----		
F		40		19		23		82	
		0.001		0.196		0.240			
		0.488		0.232		0.280		0.266	
		0.265		0.241		0.295			
		0.130		0.062		0.075			
	-----	-----	-----	-----	-----	-----	-----		
G		26		17		22		65	
		1.080		0.006		1.864			
		0.400		0.262		0.338		0.211	
		0.172		0.215		0.282			
		0.084		0.055		0.071			
	-----	-----	-----	-----	-----	-----	-----		
H		29		22		10		61	
		0.027		2.580		1.921			
		0.475		0.361		0.164		0.198	
		0.192		0.278		0.128			
		0.094		0.071		0.032			
	-----	-----	-----	-----	-----	-----	-----		
I		21		9		10		40	

	0.098	0.155	0.002	
	0.525	0.225	0.250	0.130
	0.139	0.114	0.128	
	0.068	0.029	0.032	
----- ----- ----- ----- -----				
Column Total	151	79	78	308
	0.490	0.256	0.253	
----- ----- ----- ----- -----				

ΠΙΝΑΚΑΣ 12.1: Πίνακας συνάφειας για τις μεταβλητές διαφάνεια-οργανισμός πιστοποίησης

GIA HRD IGI				
IF	6	4	34	
VS1	61	13	7	
VS2	36	15	2	
VVS1	15	23	14	
VVS2	33	24	21	

ΠΙΝΑΚΑΣ 12.2: Πίνακας συνάφειας για τις μεταβλητές διαφάνεια-οργανισμός πιστοποίησης

data\$body				
data\$clarity	GIA	HRD	IGI	Row Total
----- ----- ----- ----- -----				
IF	6	4	34	44
	11.240	4.703	46.886	

		0.136		0.091		0.773		0.143	
		0.040		0.051		0.436			
		0.019		0.013		0.110			
	-----	-----	-----	-----	-----	-----	-----		
VS1		61		13		7		81	
		11.413		2.910		8.902			
		0.753		0.160		0.086		0.263	
		0.404		0.165		0.090			
		0.198		0.042		0.023			
	-----	-----	-----	-----	-----	-----	-----		
VS2		36		15		2		53	
		3.861		0.145		9.720			
		0.679		0.283		0.038		0.172	
		0.238		0.190		0.026			
		0.117		0.049		0.006			
	-----	-----	-----	-----	-----	-----	-----		
VVS1		15		23		14		52	
		4.319		7.000		0.052			
		0.288		0.442		0.269		0.169	
		0.099		0.291		0.179			
		0.049		0.075		0.045			
	-----	-----	-----	-----	-----	-----	-----		
VVS2		33		24		21		78	

	0.718	0.797	0.079	
	0.423	0.308	0.269	0.253
	0.219	0.304	0.269	
	0.107	0.078	0.068	
----- ----- ----- ----- -----				
Column Total	151	79	78	308
	0.490	0.256	0.253	
----- ----- ----- ----- -----				

ΠΙΝΑΚΑΣ 13.1: Πίνακας συνάφειας για τις μεταβλητές μέγεθος-οργανισμός πιστοποίησης

GIA HRD IGI

1 35 0 59

2 80 48 15

3 36 31 4

ΠΙΝΑΚΑΣ 13.2: Πίνακας συνάφειας για τις μεταβλητές μέγεθος-οργανισμός πιστοποίησης

data\$body				
data\$car	GIA	HRD	IGI	Row Total
----- ----- ----- ----- -----				
1	35	0	59	94
	2.666	24.110	52.034	
	0.372	0.000	0.628	0.305
	0.232	0.000	0.756	

	0.114	0.000	0.192	
2	80	48	15	143
	1.396	3.495	12.427	
	0.559	0.336	0.105	0.464
	0.530	0.608	0.192	
	0.260	0.156	0.049	
3	36	31	4	71
	0.041	8.981	10.870	
	0.507	0.437	0.056	0.231
	0.238	0.392	0.051	
	0.117	0.101	0.013	
Column Total	151	79	78	308
	0.490	0.256	0.253	

ΠΙΝΑΚΑΣ 14.1: Πίνακας συνάφειας για τις μεταβλητές μέγεθος-διαφάνειας

1 2 3

IF 32 10 2

VS1 24 37 20

VS2 10 23 20

VVS1 10 32 10

VVS2 18 41 19

ΠΙΝΑΚΑΣ 14.2: Πίνακας συνάφειας για τις μεταβλητές μέγεθος-διαφάνειας

data\$car				
data\$clarity	1	2	3	Row Total
----- ----- ----- ----- -----				
IF	32	10	2	44
	25.684	5.324	6.537	
	0.727	0.227	0.045	0.143
	0.340	0.070	0.028	
	0.104	0.032	0.006	
----- ----- ----- ----- -----				
VS1	24	37	20	81
	0.021	0.010	0.094	
	0.296	0.457	0.247	0.263
	0.255	0.259	0.282	
	0.078	0.120	0.065	
----- ----- ----- ----- -----				
VS2	10	23	20	53
	2.358	0.105	4.957	
	0.189	0.434	0.377	0.172
	0.106	0.161	0.282	
	0.032	0.075	0.065	
----- ----- ----- ----- -----				
VVS1	10	32	10	52

	2.171 2.557 0.329	
	0.192 0.615 0.192 0.169	
	0.106 0.224 0.141	
	0.032 0.104 0.032	
----- ----- ----- ----- -----		
VVS2	18 41 19 78	
	1.416 0.632 0.058	
	0.231 0.526 0.244 0.253	
	0.191 0.287 0.268	
	0.058 0.133 0.062	
----- ----- ----- ----- -----		
Column Total	94 143 71 308	
	0.305 0.464 0.231	
----- ----- ----- ----- -----		

ΠΙΝΑΚΑΣ 15: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$sclarity and data\$body

X-squared = 112.75, df = 8, p-value < 2.2e-16

ΠΙΝΑΚΑΣ 16: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$sclarity and data\$car

X-squared = 52.254, df = 8, p-value = 1.504e-08

ΠΙΝΑΚΑΣ 17: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$car and data\$body

X-squared = 116.02, df = 4, p-value < 2.2e-16

ΠΙΝΑΚΑΣ 18: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$colour and data\$car

X-squared = 11.167, df = 10, p-value = 0.3446

Warning message:

In chisq.test(data\$colour, data\$car) :

Chi-squared approximation may be incorrect

ΠΙΝΑΚΑΣ 19: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$colour and data\$body

X-squared = 10.947, df = 10, p-value = 0.3616

Warning message:

In chisq.test(data\$colour, data\$body) :

Chi-squared approximation may be incorrect

ΠΙΝΑΚΑΣ 20: Αποτελέσματα ελέγχου χ^2

Pearson's Chi-squared test

data: data\$colour and data\$clarity

X-squared = 21.009, df = 20, p-value = 0.3966

Warning message:

In chisq.test(data\$colour, data\$clarity) :

Chi-squared approximation may be incorrect

ΠΙΝΑΚΑΣ 21: Έλεγχος αν είναι το διαμάντι ακριβό προϊόν

Wilcoxon signed rank test with continuity correction

data: data\$price

V = 29464, p-value = 0.0001444

alternative hypothesis: true location is greater than 4000

ΠΙΝΑΚΑΣ 22: Έλεγχος σχέσης τιμής διαφάνειας

```
> summary(aov(price~clarity,data=data))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
clarity	4	2.997e+08	74917212	6.972	2.22e-05 ***
Residuals	303	3.256e+09	10745078		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> shapiro.test(aov(price~clarity,data=data)$res)#oxi kanonika katalipa#
```

Shapiro-Wilk normality test

data: aov(price ~ clarity, data = data)\$res

W = 0.94416, p-value = 2.115e-09

```
> leveneTest(price~clarity,data=data,center=mean)#oxi omoskedastika katalipa#
```

Levene's Test for Homogeneity of Variance (center = mean)

	Df	F value	Pr(>F)
group	4	0.8191	0.5138
	303		

Kruskal-Wallis rank sum test

data: price by clarity

Kruskal-Wallis chi-squared = 37.524, df = 4, p-value = 1.405e-07

ΠΙΝΑΚΑΣ 23: Έλεγχος σχέσης τιμής οργανισμού πιστοποίησης

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
body	2	9.727e+08	486373280	57.44	<2e-16 ***
Residuals	305	2.583e+09	8467806		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data: aov(price ~ body, data = data)\$res

W = 0.93975, p-value = 7.092e-10

Levene's Test for Homogeneity of Variance (center = mean)

	Df	F value	Pr(>F)
--	----	---------	--------

group	2	10.047	5.949e-05 ***
-------	---	--------	---------------

	305		
--	-----	--	--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kruskal-Wallis rank sum test

data: price by body

Kruskal-Wallis chi-squared = 101.85, df = 2, p-value < 2.2e-16

ΠΙΝΑΚΑΣ 24: Έλεγχος σχέσης τιμής χρώματος

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
--	----	--------	---------	---------	--------

colour	5	1.082e+08	21645956	1.896	0.0948 .
--------	---	-----------	----------	-------	----------

Residuals	302	3.447e+09	11414561		
-----------	-----	-----------	----------	--	--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data: aov(price ~ colour, data = data)\$res

W = 0.94906, p-value = 7.612e-09

Levene's Test for Homogeneity of Variance (center = mean)

	Df	F value	Pr(>F)
--	----	---------	--------

group	5	4.8504	0.0002778 ***
-------	---	--------	---------------

	302		
--	-----	--	--

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kruskal-Wallis rank sum test

data: price by colour

Kruskal-Wallis chi-squared = 5.2339, df = 5, p-value = 0.388

ΠΙΝΑΚΑΣ 25: Έλεγχος σχέσης τιμής μεγέθους (κατηγορική)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
--	----	--------	---------	---------	--------

car 2 3.047e+09 1.523e+09 913.2 <2e-16 ***

Residuals 305 5.088e+08 1.668e+06

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data: aov(price ~ car, data = data)\$res

W = 0.91602, p-value = 4.132e-12

Levene's Test for Homogeneity of Variance (center = mean)

Df F value Pr(>F)

group 2 42.474 < 2.2e-16 ***

305

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kruskal-Wallis rank sum test

data: price by car

Kruskal-Wallis chi-squared = 262, df = 2, p-value < 2.2e-16

ΠΙΝΑΚΑΣ 26:Γραμμικό μοντέλο τιμής με ποσοτική μεταβλητή μέγεθος

Call:

lm(formula = data\$price ~ data\$carat)

Residuals:

Min 1Q Median 3Q Max

-2264.7 -604.3 -116.1 435.1 6591.5

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2298.4 158.5 -14.50 <2e-16 ***

data\$carat 11598.9 230.1 50.41 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1118 on 306 degrees of freedom

Multiple R-squared: 0.8925, Adjusted R-squared: 0.8922

F-statistic: 2541 on 1 and 306 DF, p-value: < 2.2e-16

Shapiro-Wilk normality test

data: 11\$res

W = 0.84234, p-value < 2.2e-16

Lilliefors (Kolmogorov-Smirnov) normality test

data: 11\$res

D = 0.11143, p-value = 6.967e-10

ΠΙΝΑΚΑΣ 27: Έλεγχοι κανονικότητας καταλοίπων μετασχηματισμών

Shapiro-Wilk normality test

data: 12\$res

W = 0.98999, p-value = 0.03324

Shapiro-Wilk normality test

data: 13\$res

W = 0.98838, p-value = 0.01446

Shapiro-Wilk normality test

data: 14\$res

W = 0.89696, p-value = 1.291e-13

Shapiro-Wilk normality test

data: 15\$res

W = 0.92281, p-value = 1.614e-11

Shapiro-Wilk normality test

data: 16\$res

W = 0.89613, p-value = 1.12e-13

Shapiro-Wilk normality test

data: 17\$res

W = 0.87762, p-value = 5.86e-15

Lilliefors (Kolmogorov-Smirnov) normality test

data: 12\$res

D = 0.051054, p-value = 0.05108

Lilliefors (Kolmogorov-Smirnov) normality test

data: 13\$res

D = 0.056215, p-value = 0.02016

ΠΙΝΑΚΑΣ 28: Έλεγχος ομοσκεδαστικότητας καταλοίπων για 112

F test to compare two variances

data: rstandard(12) and x

F = 0.80022, num df = 307, denom df = 307, p-value = 0.05131

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.6395428 1.0012752

sample estimates:

ratio of variances

0.8002239

ΠΙΝΑΚΑΣ 29 :Έλεγχος ανεξαρτησίας καταλοίπων

lag Autocorrelation D-W Statistic p-value

1 0.6087599 0.7674872 0

Alternative hypothesis: rho != 0

lag Autocorrelation D-W Statistic p-value

1 0.6087599 0.7674872 0

Alternative hypothesis: rho != 0

ΠΙΝΑΚΑΣ 30:Πολυωνυμικός μετασχηματισμός

Call:

lm(formula = price ~ poly(carat, 6, raw = T), data = data)

Residuals:

Min	1Q	Median	3Q	Max
-2181.5	-373.4	-40.0	269.6	5931.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-16549	9385	-1.763	0.07885 .
poly(carat, 6, raw = T)1	246053	122915	2.002	0.04620 *
poly(carat, 6, raw = T)2	-1348872	616711	-2.187	0.02950 *
poly(carat, 6, raw = T)3	3687762	1529235	2.412	0.01649 *

poly(carat, 6, raw = T)4 -5227358 1996542 -2.618 0.00929 **

poly(carat, 6, raw = T)5 3704103 1313734 2.820 0.00513 **

poly(carat, 6, raw = T)6 -1035165 343270 -3.016 0.00278 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 984.6 on 301 degrees of freedom

Multiple R-squared: 0.9179, Adjusted R-squared: 0.9163

F-statistic: 561.1 on 6 and 301 DF, p-value: < 2.2e

ΠΙΝΑΚΑΣ 31: Έλεγχος ο κανονικότητας και ομοσκεδαστικότητας καταλοιπων μοντέλου με 4 κατηγορικές

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 126 1.1277 0.2286

181

Shapiro-Wilk normality test

data: anova1\$res

W = 0.98279, p-value = 0.0009395

Lilliefors (Kolmogorov-Smirnov) normality test

data: anova1\$res

D = 0.070671, p-value = 0.000806

ΠΙΝΑΚΑΣ 32 : Σύγκριση σταθερού μοντέλου με πλήρες(χωρίς αλληλεπιδράσεις)

Analysis of Variance Table

Model 1: price ~ 1

Model 2: price ~ car + colour + body + clarity

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
--	--------	-----	----	-----------	---	--------

1	307	3555427347				
---	-----	------------	--	--	--	--

2	294	376140958	13	3179286389	191.15	< 2.2e-16 ***
---	-----	-----------	----	------------	--------	---------------

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data: main.effects\$res

W = 0.98279, p-value = 0.0009395

Lilliefors (Kolmogorov-Smirnov) normality test

data: main.effects\$res

D = 0.070671, p-value = 0.000806

ΠΙΝΑΚΑΣ 33: Έλεγχοι κανονικότητας καταλοίπων για διάφορα μοντέλα

Shapiro-Wilk normality test

data: a2\$res

W = 0.92178, p-value = 1.306e-11

data: a3\$res

W = 0.9659, p-value = 1.204e-06

data: a4\$res

W = 0.93048, p-value = 8.301e-11

data: a5\$res

W = 0.94636, p-value = 3.732e-09

data: a6\$res

W = 0.96025, p-value = 1.926e-07

data: a7\$res

W = 0.96177, p-value = 3.113e-07

data: a8\$res

W = 0.97322, p-value = 1.678e-05

data: a9\$res

W = 0.94141, p-value = 1.063e-09

data: a10\$res

W = 0.9654, p-value = 1.017e-06

data: a11\$res

W = 0.97274, p-value = 1.4e-05

ΠΙΝΑΚΑΣ 34: Έλεγχοι κανονικότητας καταλοίπων για διάφορα μοντέλα

Shapiro-Wilk normality test

data: a12\$res

W = 0.95663, p-value = 6.438e-08

data: a13\$res

W = 0.96452, p-value = 7.586e-07

data: a14\$res

W = 0.92512, p-value = 2.611e-11

data: a15\$res

W = 0.96749, p-value = 2.076e-06

data: a16\$res

W = 0.96908, p-value = 3.639e-06

data: a17\$res

W = 0.97328, p-value = 1.717e-05

ΠΙΝΑΚΑΣ 35: Έλεγχοι προϋποθέσεων για μοντέλο α18

Shapiro-Wilk normality test

data: a18\$res

W = 0.99034, p-value = 0.03989

Lilliefors (Kolmogorov-Smirnov) normality test

data: a18\$res

D = 0.050841, p-value = 0.05294

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 126 1.1277 0.2286

181

ΠΙΝΑΚΑΣ 36: Αναζήτηση μοντέλου με κριτήριο AIC

Start: AIC=5010.59

price ~ 1

	Df	Sum of Sq	RSS	AIC
+ car	2	3046642739	508784608	4415.8
+ body	2	972746561	2582680786	4916.1
+ clarity	4	299668847	3255758499	4991.5
<none>			3555427347	5010.6
+ colour	5	108229779	3447197568	5011.1

Step: AIC=4415.77

price ~ car

	Df	Sum of Sq	RSS	AIC
+ colour	5	85882489	422902118	4368.8
+ body	2	12877401	495907207	4411.9
+ clarity	4	13199220	495585387	4415.7
<none>			508784608	4415.8
- car	2	3046642739	3555427347	5010.6

Step: AIC=4368.83

price ~ car + colour

	Df	Sum of Sq	RSS	AIC
+ body	2	15514700	407387418	4361.3
+ clarity	4	19687855	403214263	4362.1
<none>			422902118	4368.8
- colour	5	85882489	508784608	4415.8
- car	2	3024295450	3447197568	5011.1

Step: AIC=4361.31

price ~ car + colour + body

	Df	Sum of Sq	RSS	AIC
+ clarity	4	31246460	376140958	4344.7
<none>			407387418	4361.3
- body	2	15514700	422902118	4368.8
- colour	5	88519789	495907207	4411.9

- car 2 2054216509 2461603926 4911.3

Step: AIC=4344.74

price ~ car + colour + body + clarity

	Df	Sum of Sq	RSS	AIC
+ colour:clarity	20	80432525	295708433	4310.6
<none>			376140958	4344.7
- clarity	4	31246460	407387418	4361.3
- body	2	27073306	403214263	4362.1
- colour	5	99146771	475287729	4406.8
- car	2	2050505567	2426646525	4914.9

Step: AIC=4310.63

price ~ car + colour + body + clarity + colour:clarity

	Df	Sum of Sq	RSS	AIC
<none>			295708433	4310.6
- body	2	21001979	316710412	4327.8
- colour:clarity	20	80432525	376140958	4344.7
- car	2	1848082176	2143790609	4916.8

ΠΙΝΑΚΑΣ 37: Έλεγχοι προϋποθέσεων για μοντέλο μετά την αφαίρεση των ακραίων τιμών

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	118	1.1255	0.2379

Shapiro-Wilk normality test

data: anova1\$res

$W = 0.98029$, p-value = 0.000458

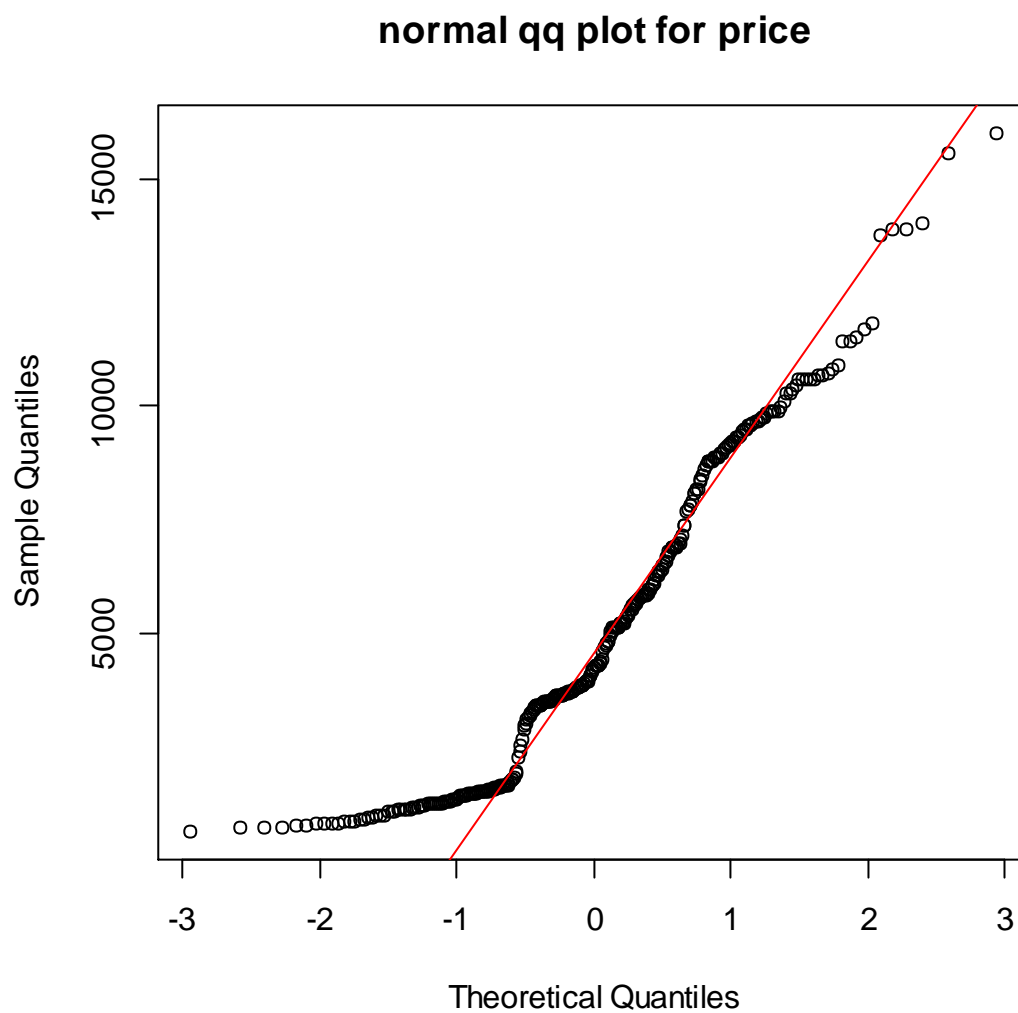
Lilliefors (Kolmogorov-Smirnov) normality test

data: anova1\$res

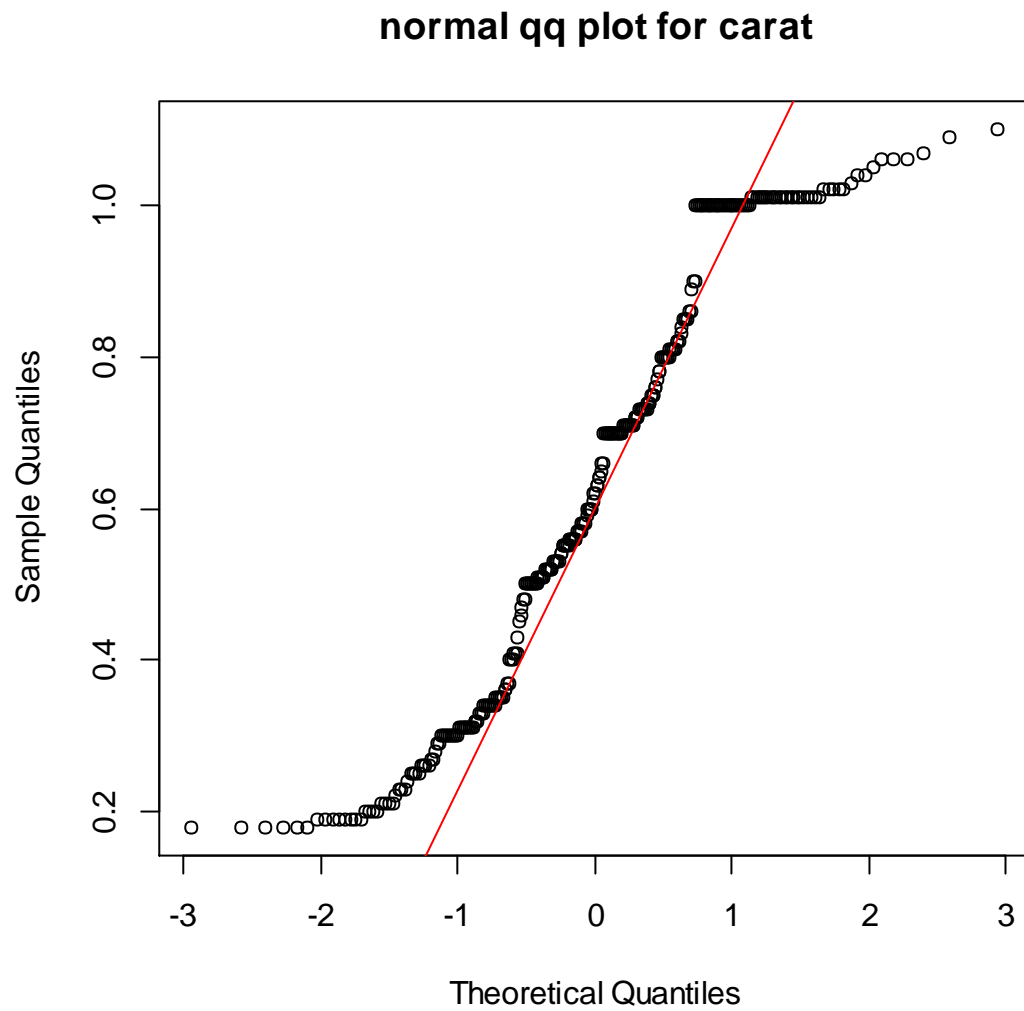
$D = 0.052162$, p-value = 0.0524

ΔΙΑΓΡΑΜΜΑΤΑ

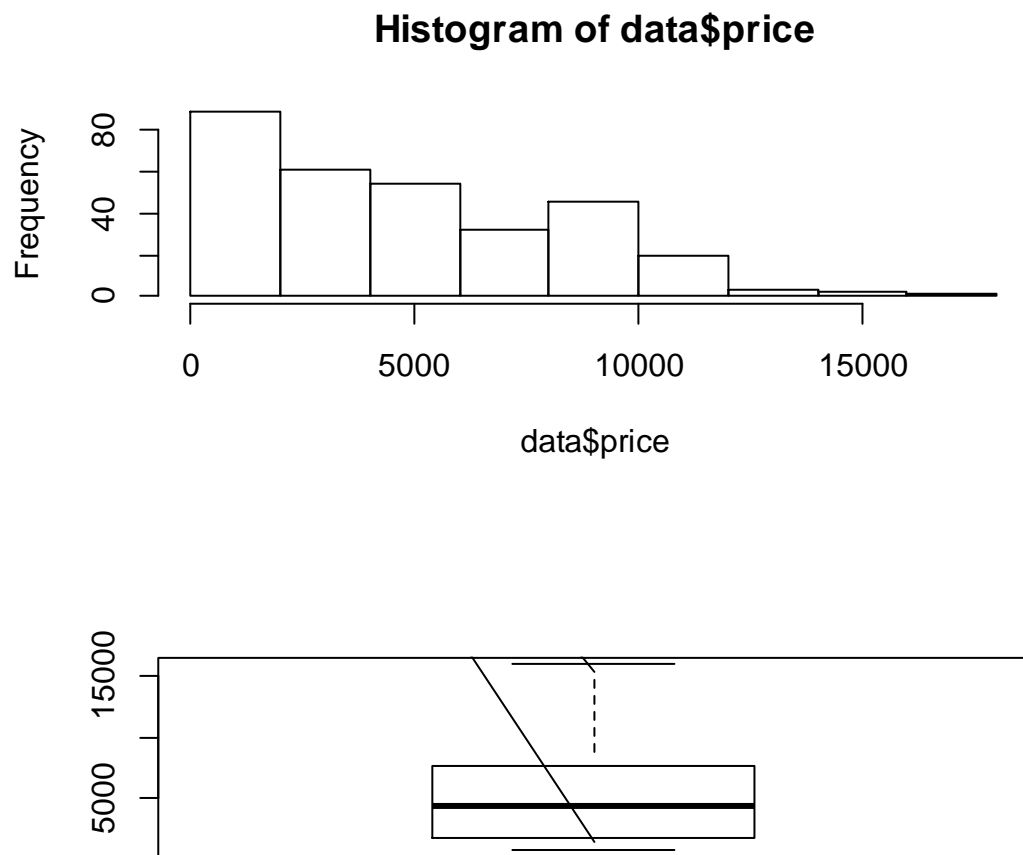
ΔΙΑΓΡΑΜΜΑ 1: Έλεγχος κανονικότητας για την ποσοτική μεταβλητή "ΤΙΜΗ ΔΙΑΜΑΝΤΙΟΥ"



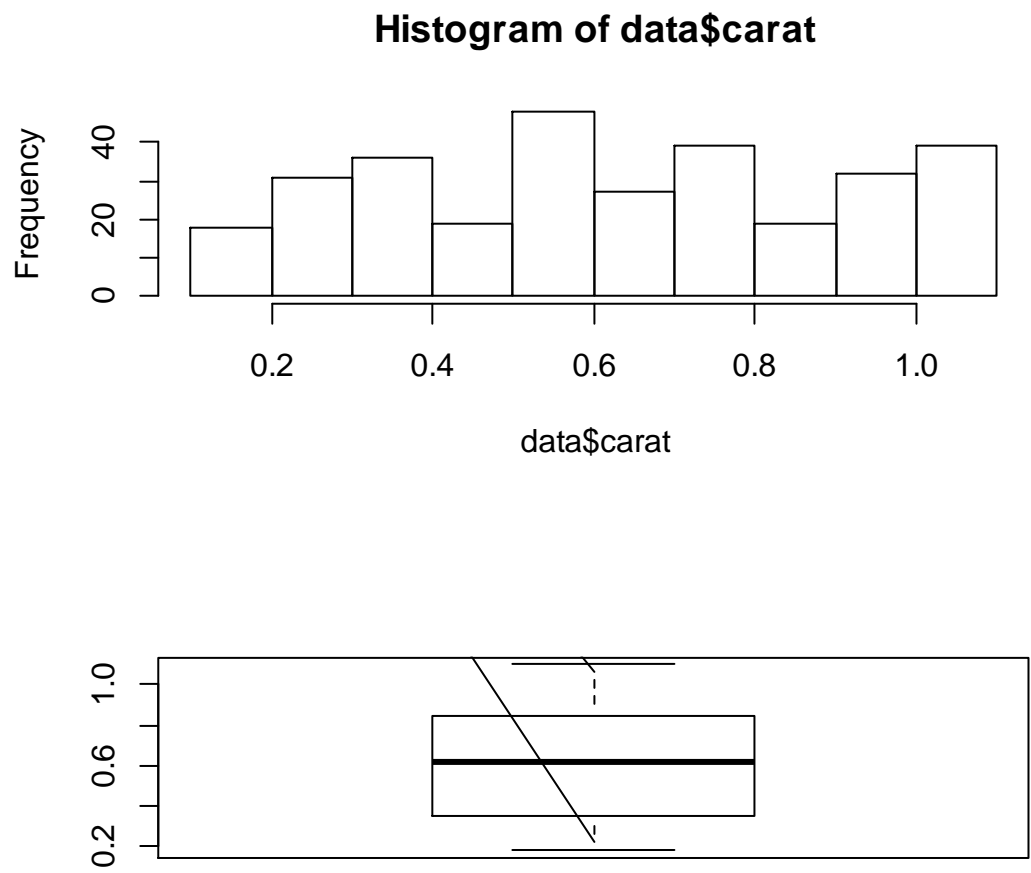
ΔΙΑΓΡΑΜΜΑ2: Έλεγχος κανονικότητας για την ποσοτική μεταβλητή "ΜΕΓΕΘΟΣ
ΔΙΑΜΑΝΤΙΟΥ"



ΔΙΑΓΡΑΜΜΑ 3: Ιστόγραμμα και boxplots για ποσοτική μεταβλητή "ΤΙΜΗ ΔΙΑΜΑΝΤΙΟΥ"

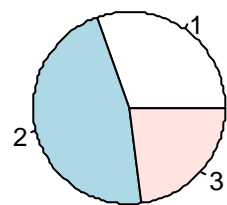


ΔΙΑΓΡΑΜΜΑ 4: Ιστόγραμμα και boxplots για ποσοτική μεταβλητή "ΜΕΓΕΘΟΣ ΔΙΑΜΑΝΤΙΟΥ"

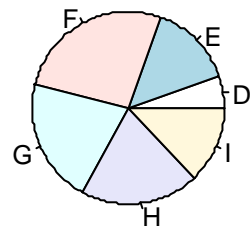


ΔΙΑΓΡΑΜΜΑ 5: "ΠΙΤΕΣ" Για τις κατηγορικές μεταβλητές

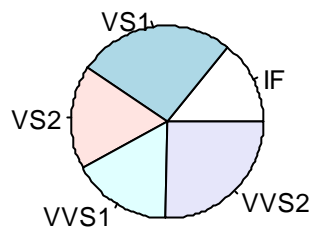
ΚΑΤΗΓΟΡΙΕΣ ΜΕΓΕΘΟΥΣ



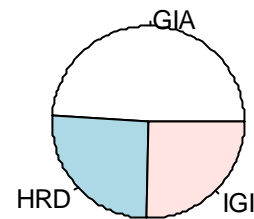
ΚΑΤΗΓΟΡΙΕΣ ΧΡΩΜΑΤΟΣ



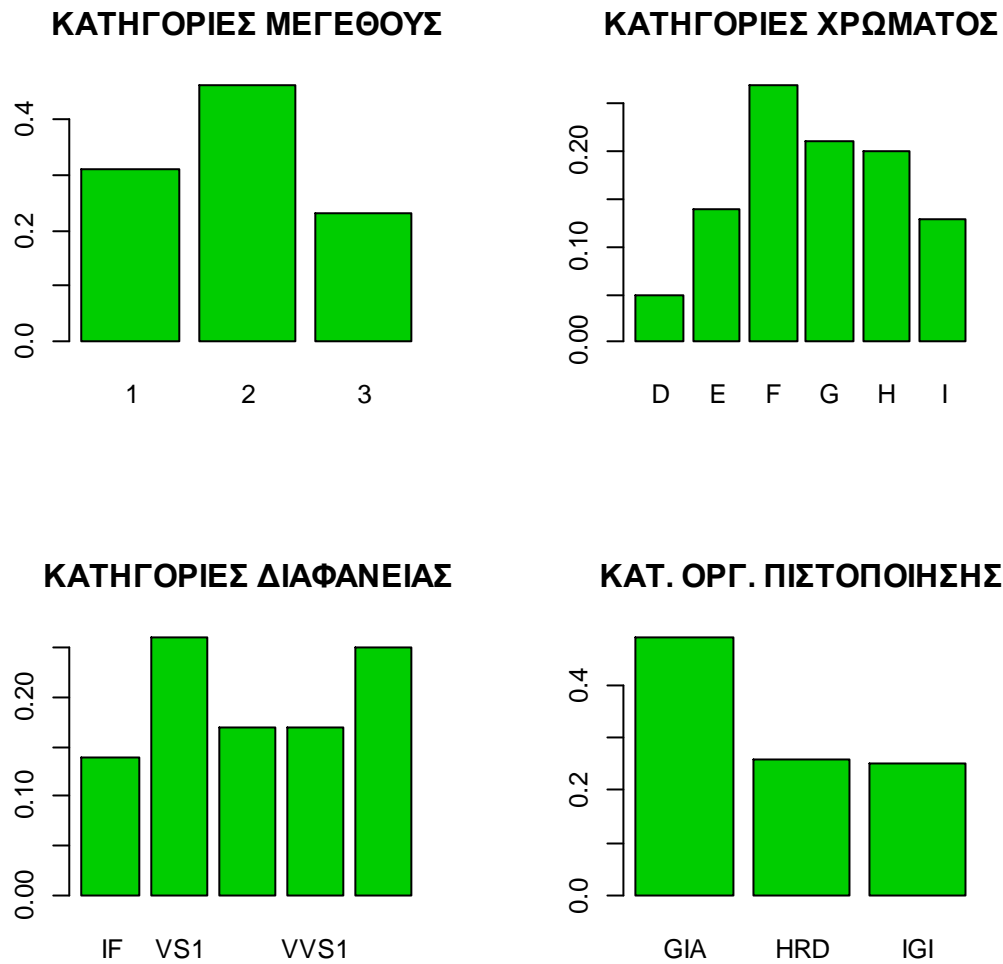
ΚΑΤΗΓΟΡΙΕΣ ΔΙΑΦΑΝΕΙΑΣ



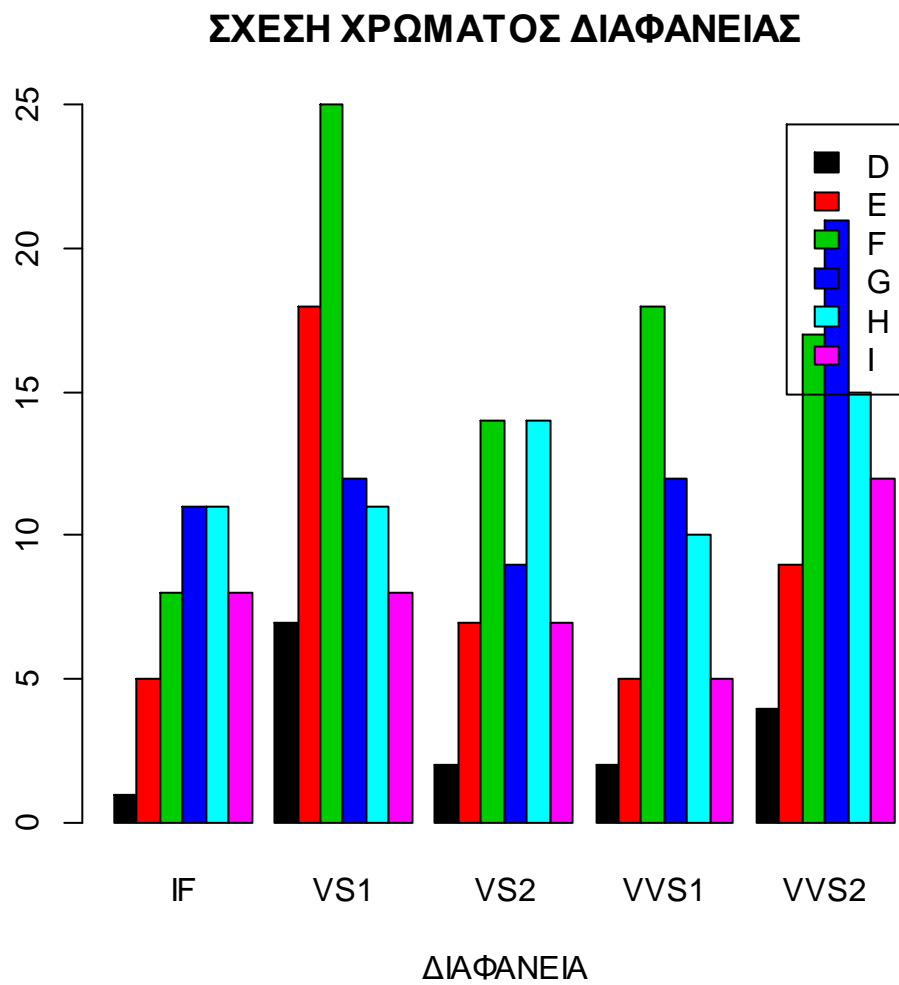
ΚΑΤ. ΟΡΓ. ΠΙΣΤΟΠΟΙΗΣΗΣ



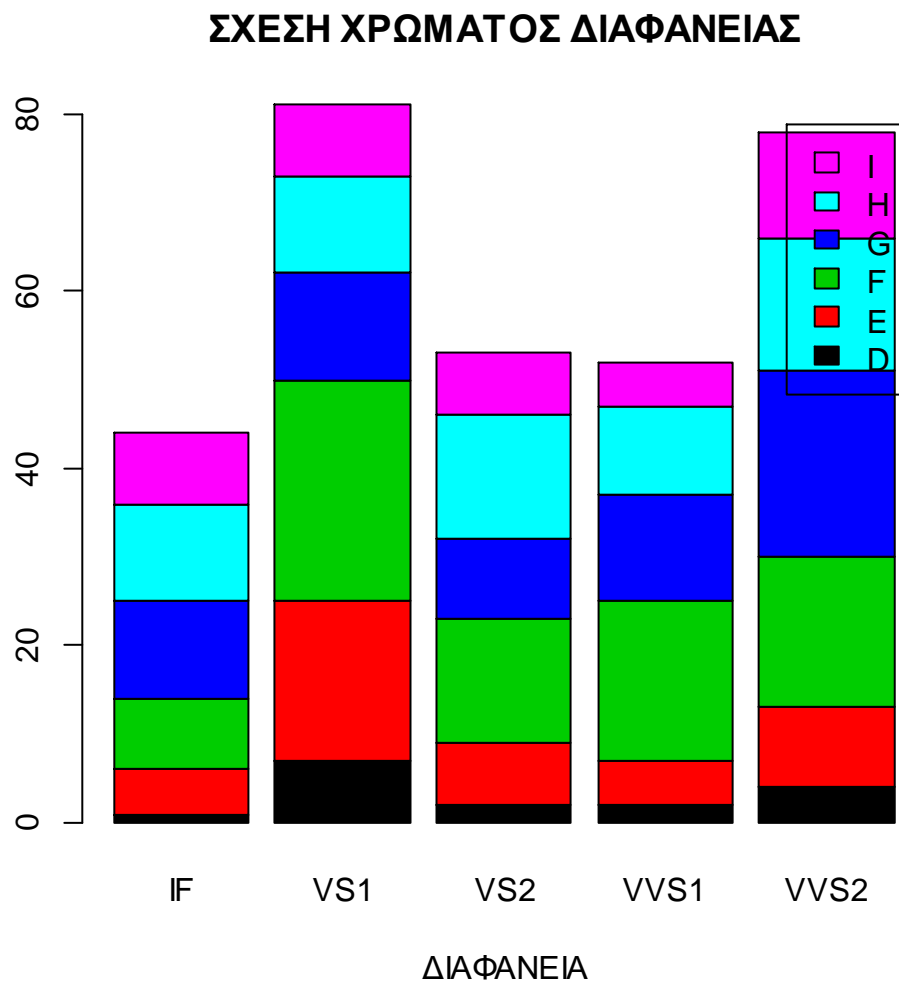
ΔΙΑΓΡΑΜΜΑ 6:"ΡΑΒΔΟΓΡΑΜΜΑΤΑ" Για τις κατηγορικές μεταβλητές



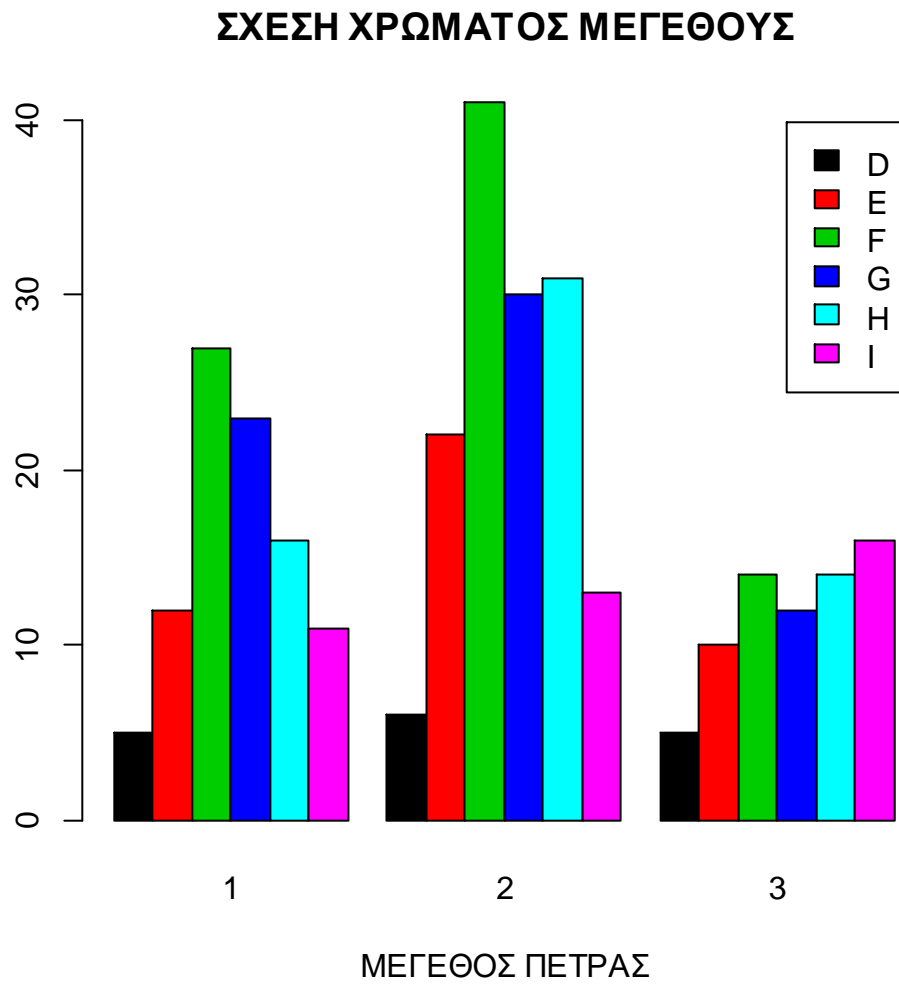
ΔΙΑΓΡΑΜΜΑ 7.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΔΙΑΦΑΝΕΙΑΣ



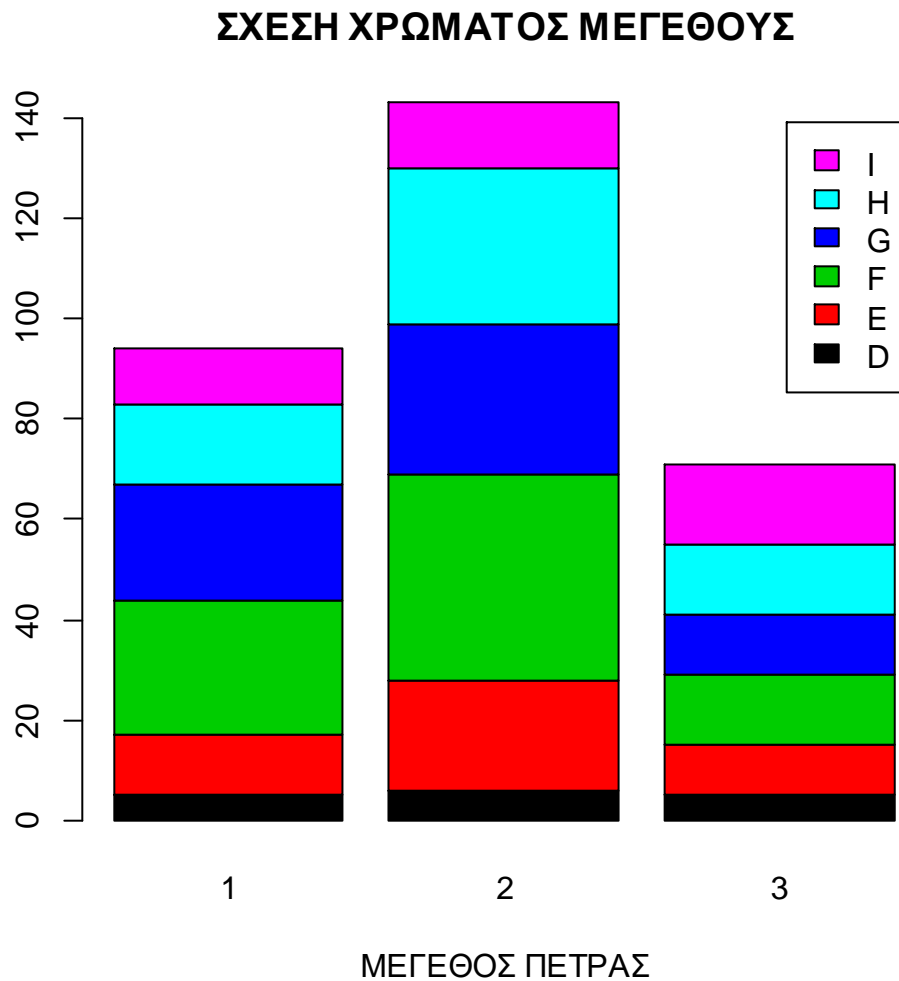
ΔΙΑΓΡΑΜΜΑ 7.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΔΙΑΦΑΝΕΙΑΣ



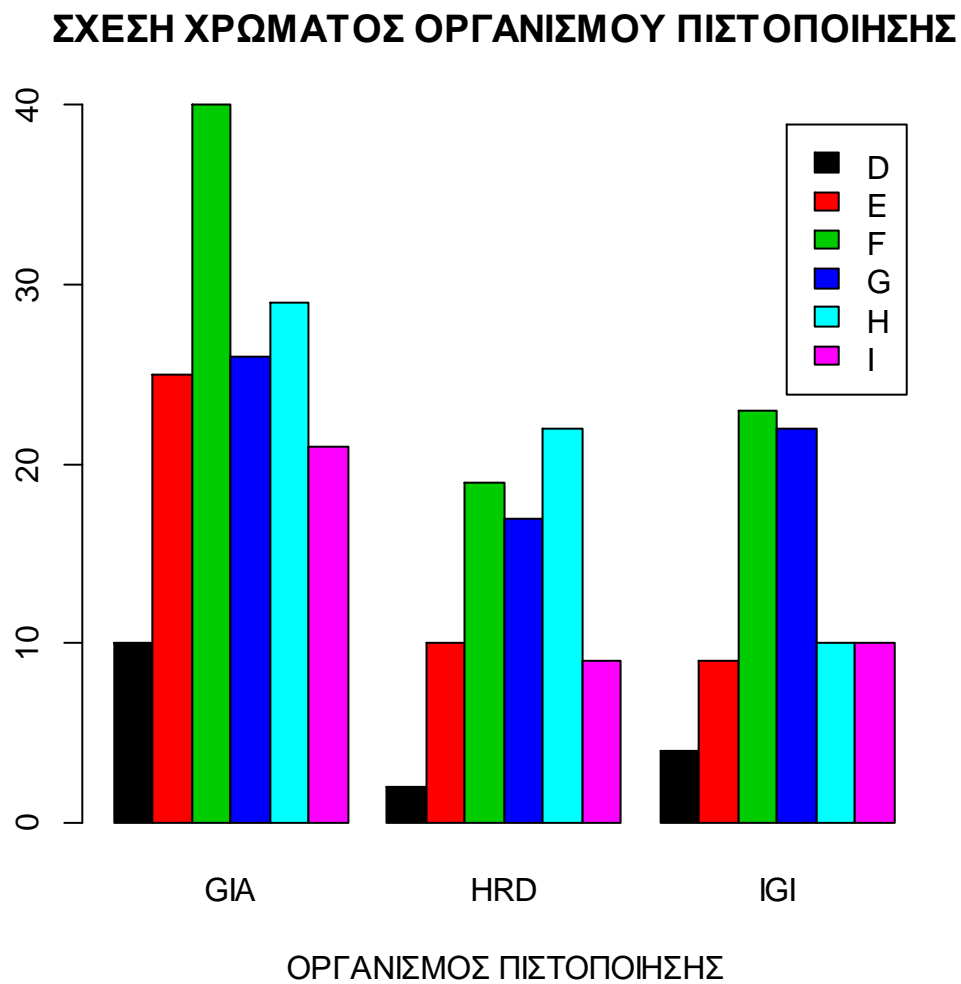
ΔΙΑΓΡΑΜΜΑ 8.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΜΕΓΕΘΟΥΣ



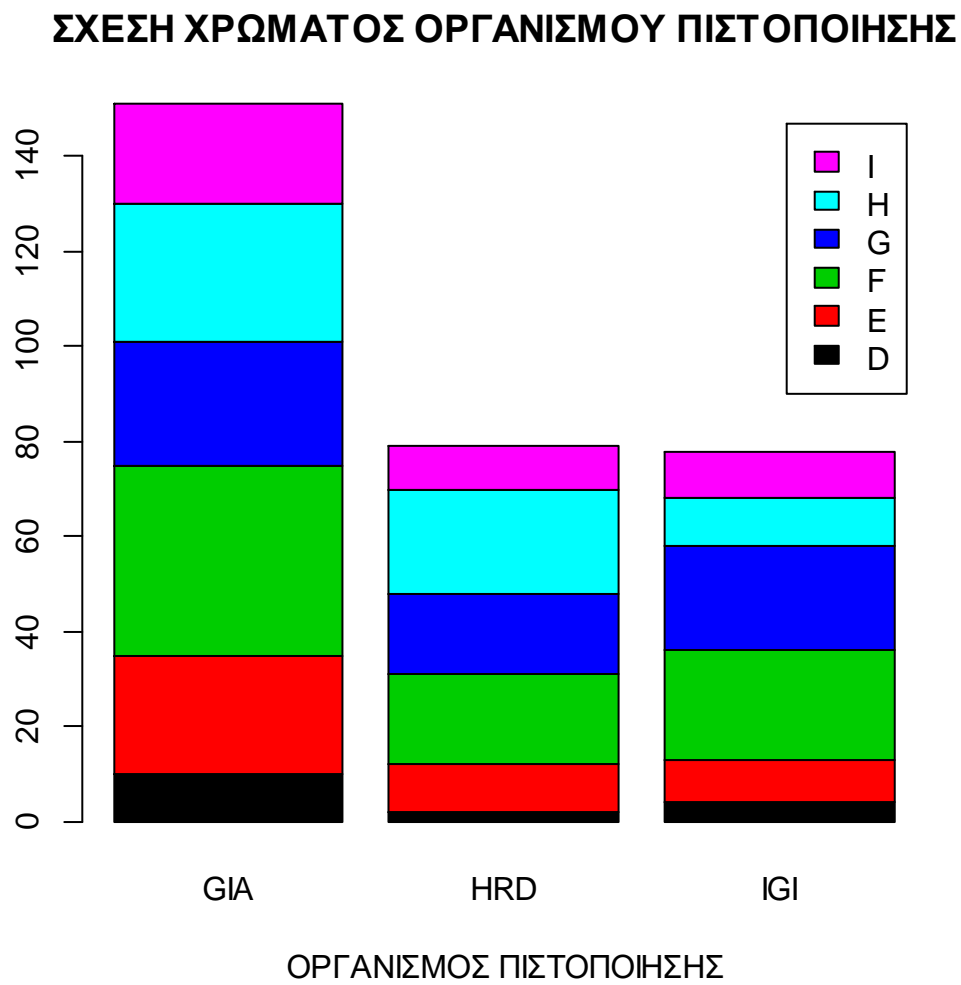
ΔΙΑΓΡΑΜΜΑ 8.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΜΕΓΕΘΟΥΣ



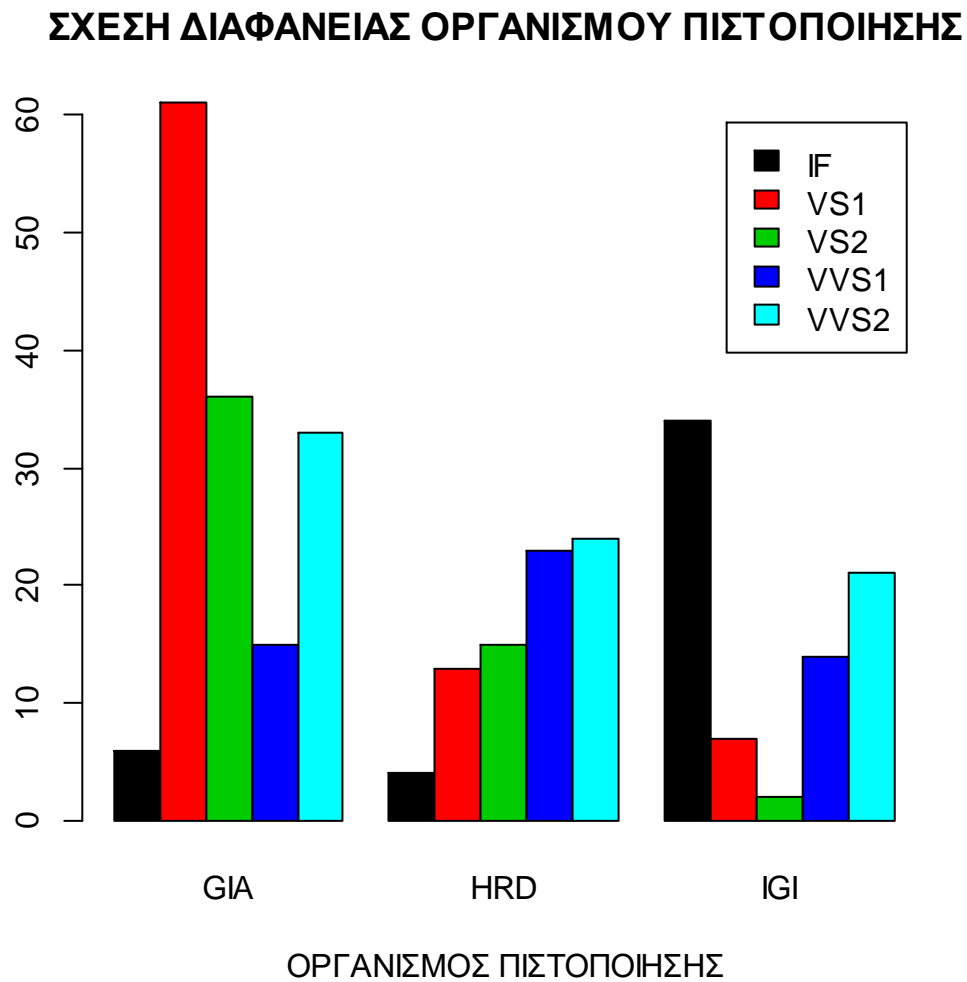
ΔΙΑΓΡΑΜΜΑ 9.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



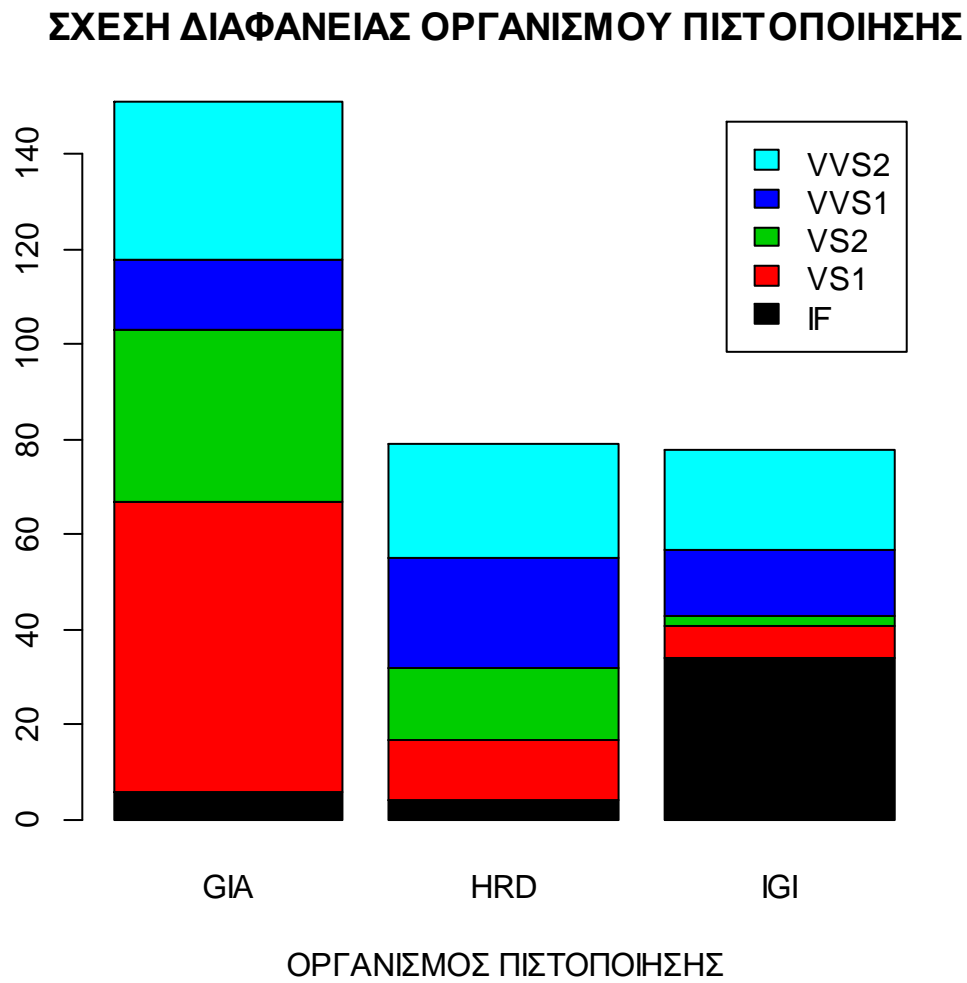
ΔΙΑΓΡΑΜΜΑ 9.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΧΡΩΜΑΤΟΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



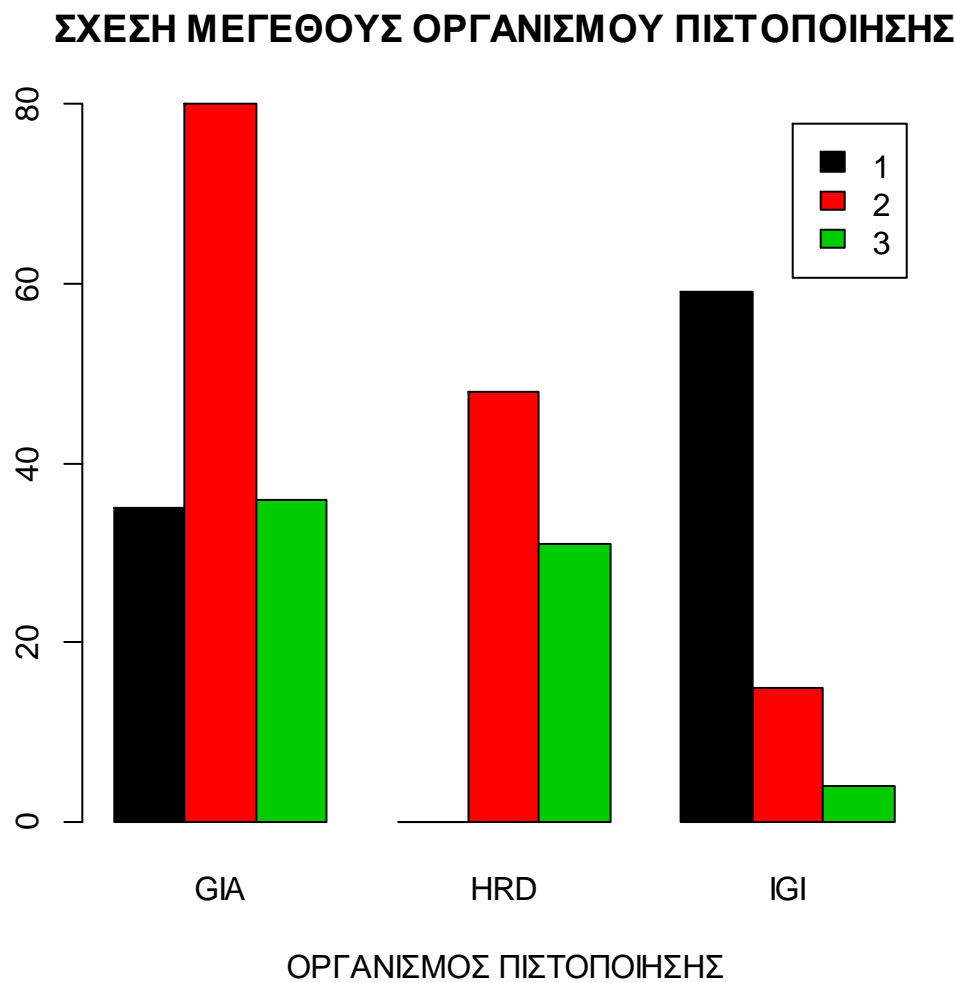
ΔΙΑΓΡΑΜΜΑ 10.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΔΙΑΦΑΝΕΙΑΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



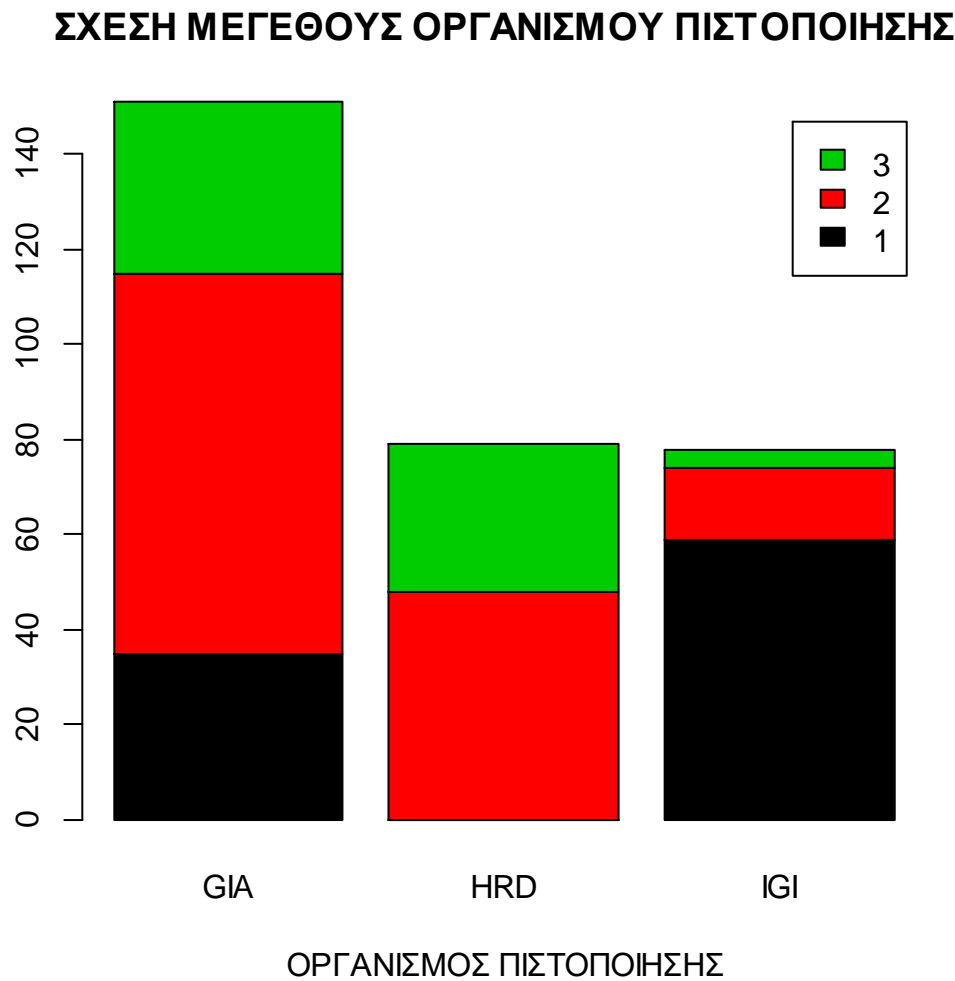
ΔΙΑΓΡΑΜΜΑ 10.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΔΙΑΦΑΝΕΙΑΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



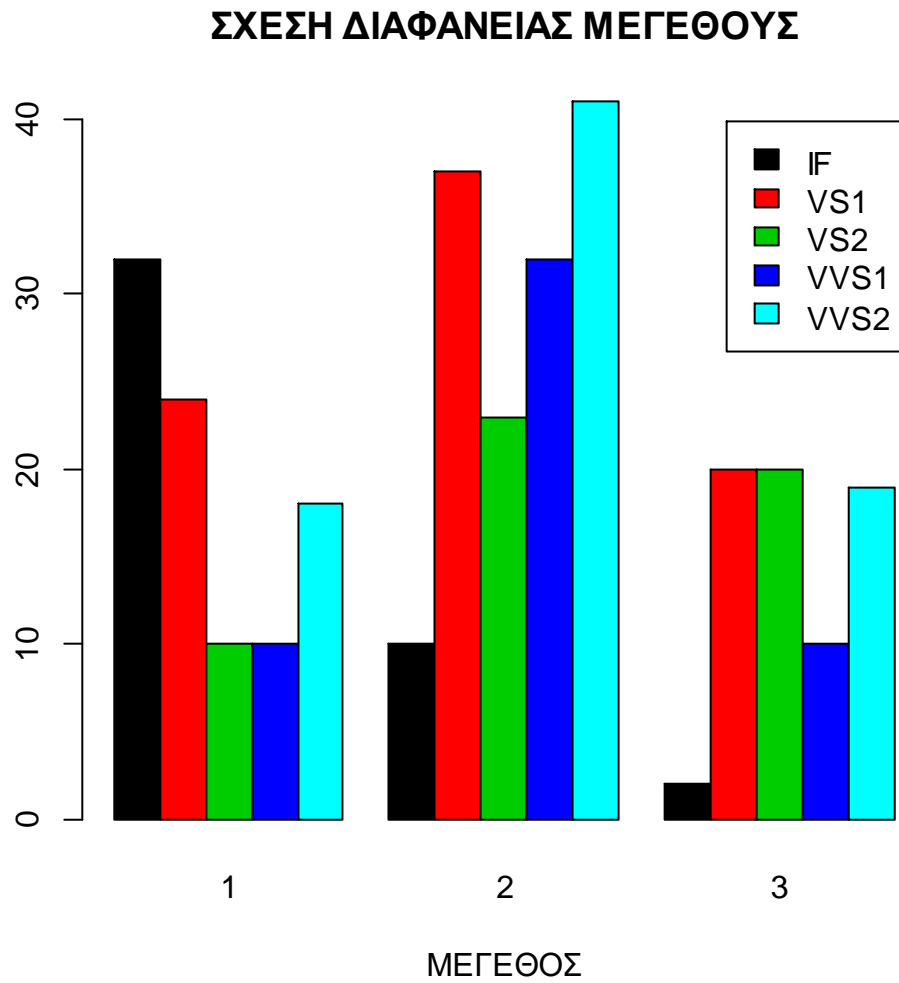
ΔΙΑΓΡΑΜΜΑ 11.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΜΕΓΕΘΟΥΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



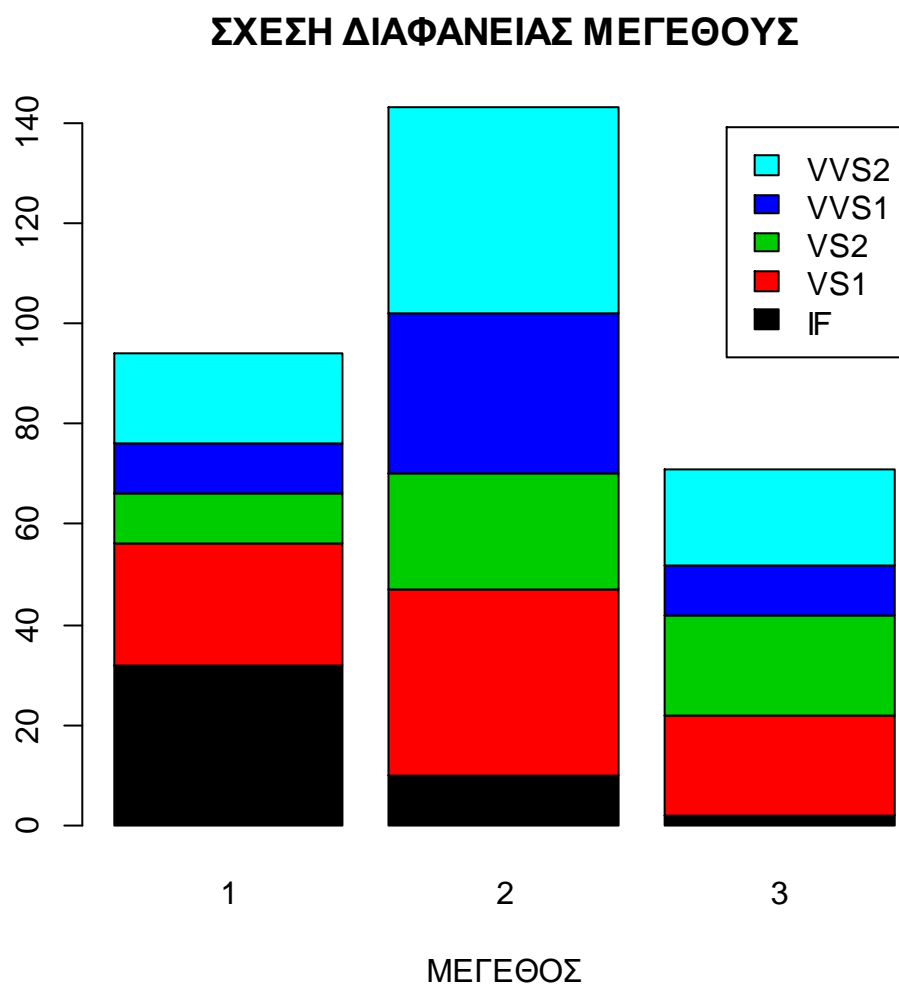
ΔΙΑΓΡΑΜΜΑ 11.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΜΕΓΕΘΟΥΣ ΟΡΓΑΝΙΣΜΟΥ ΠΙΣΤΟΠΟΙΗΣΗΣ



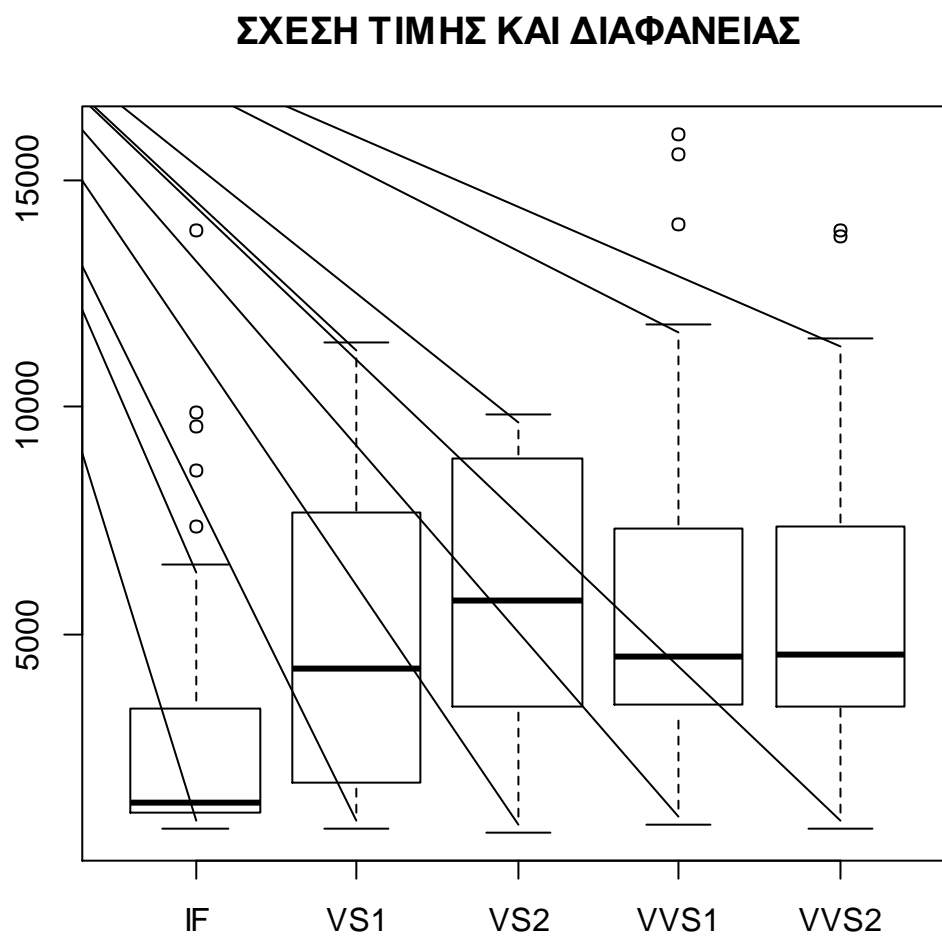
ΔΙΑΓΡΑΜΜΑ 12.1: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΜΕΓΕΘΟΥΣ ΔΙΑΦΑΝΕΙΑΣ



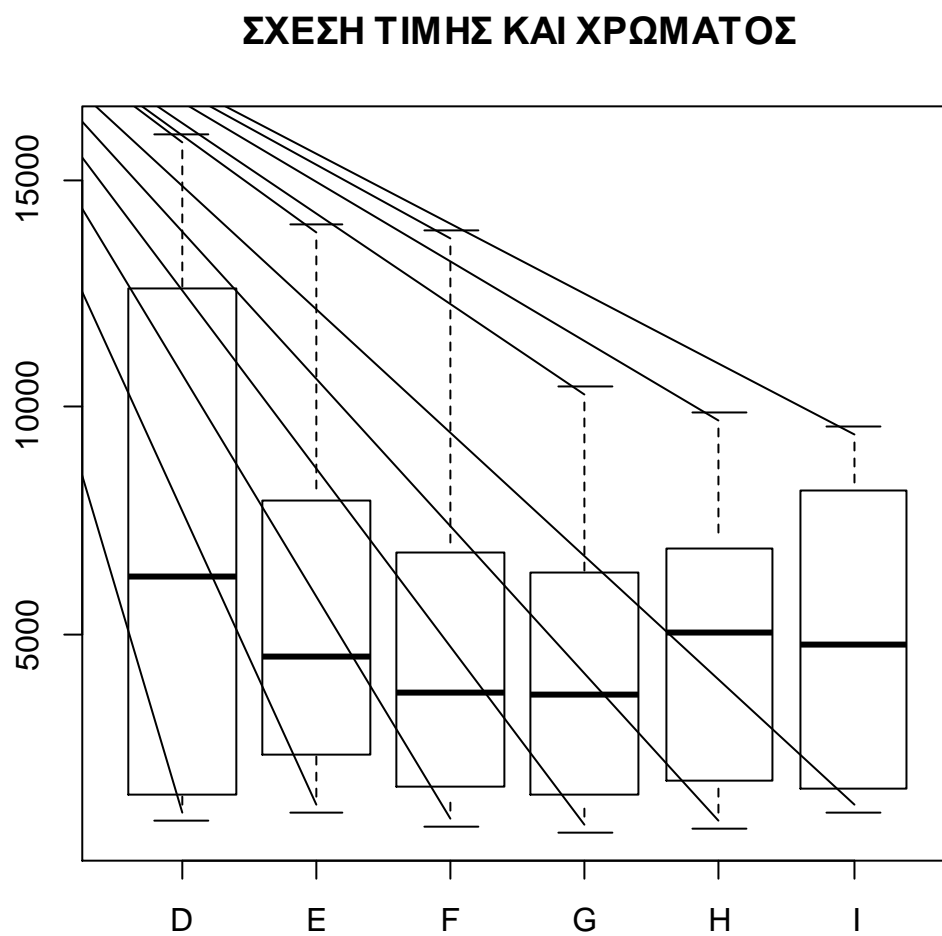
ΔΙΑΓΡΑΜΜΑ 12.2: ΡΑΒΔΟΓΡΑΜΜΑ ΣΧΕΣΗ ΜΕΓΕΘΟΥΣ ΔΙΑΦΑΝΕΙΑΣ



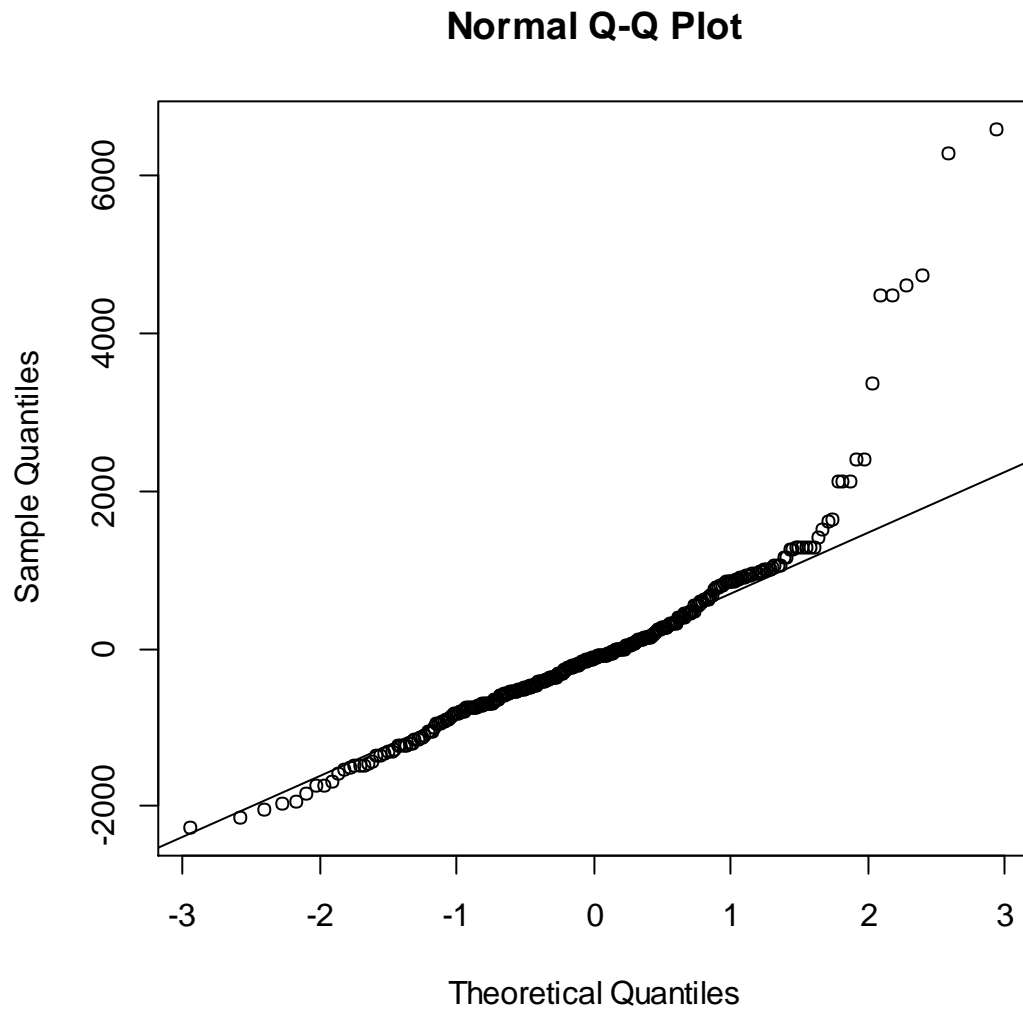
ΔΙΑΓΡΑΜΜΑ 13 :boxplot σχέσης τιμής και διαφάνειας



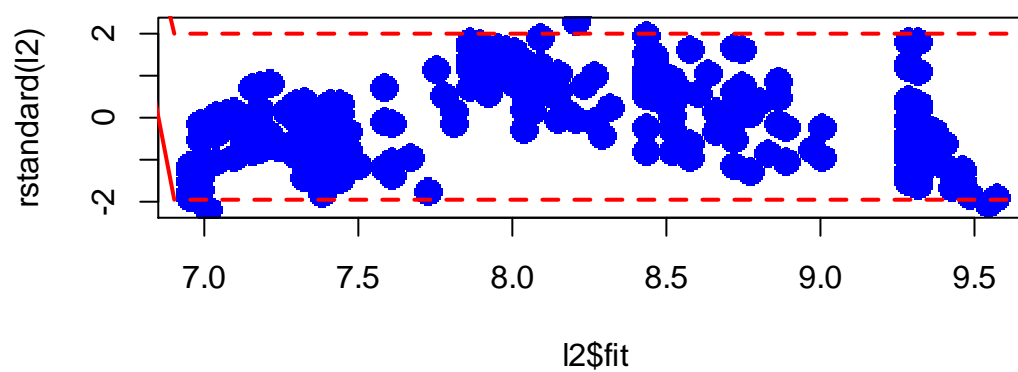
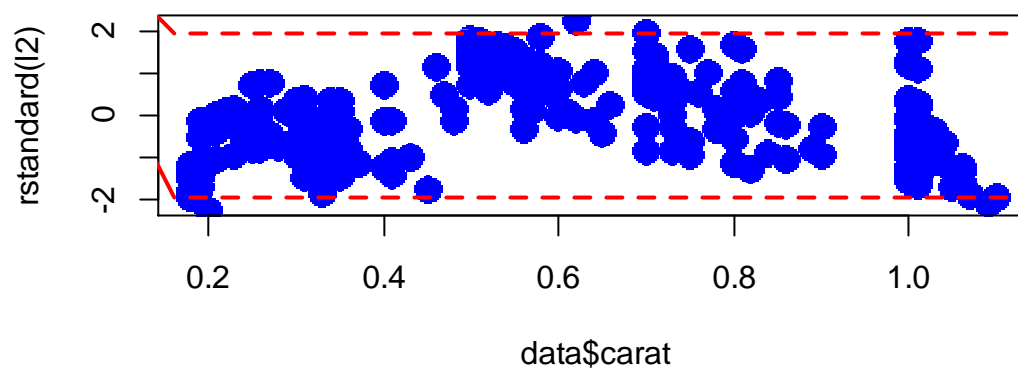
ΔΙΑΓΡΑΜΜΑ 14: boxplot σχέσης τιμής και χρώματος



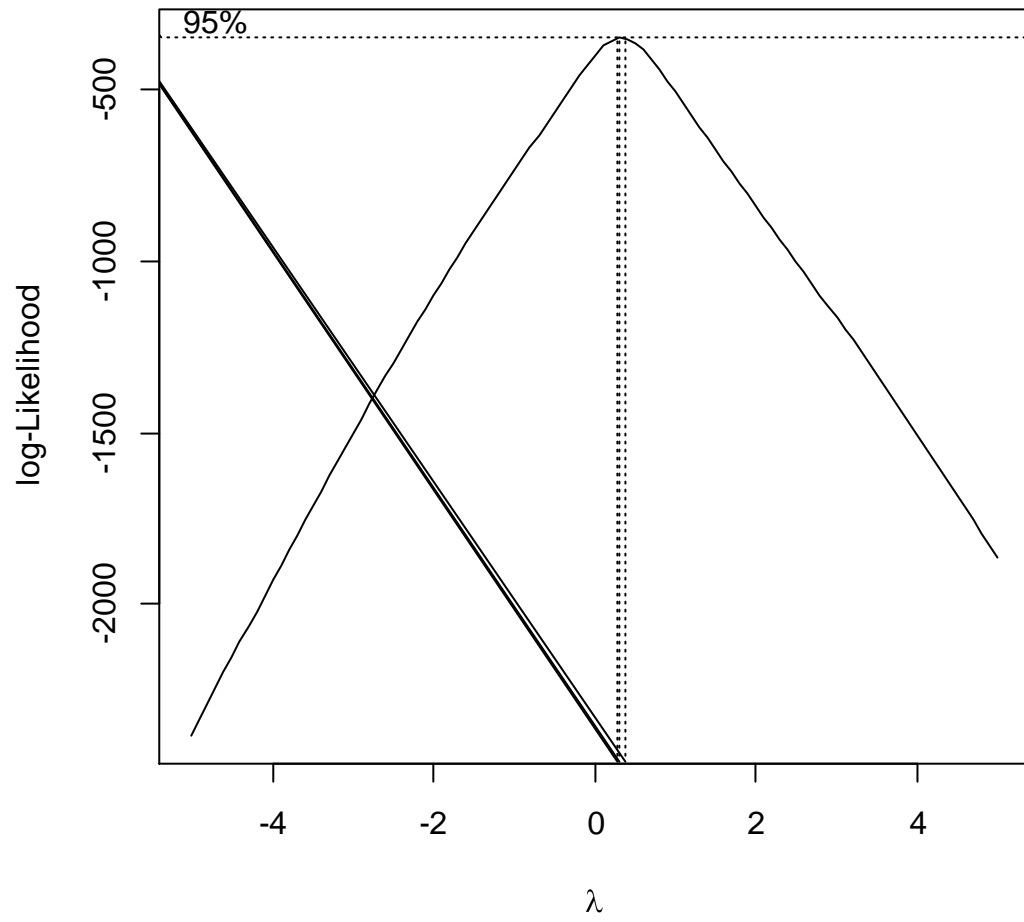
ΔΙΑΓΡΑΜΜΑ 15 ΚΑΝΟΝΙΚΟΤΗΤΑ ΚΑΤΑΛΟΓΙΩΝ



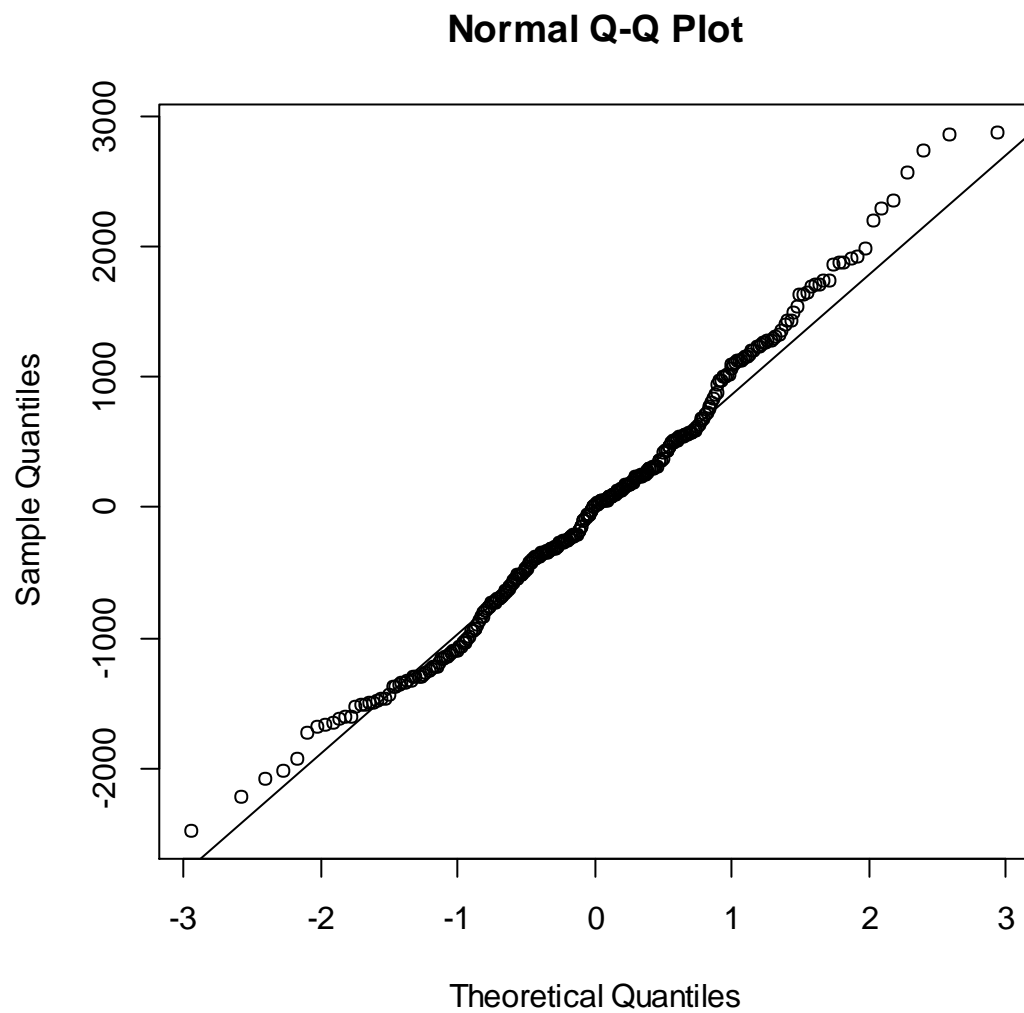
ΔΙΑΓΡΑΜΜΑΤΑ 16,17 ΟΜΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑ ΚΑΤΑΛΟΙΠΩΝ



ΔΙΑΓΡΑΜΜΑ 18 :ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ BOX-COX



ΔΙΑΓΡΑΜΜΑ 19 : ΔΙΑΓΡΑΜΜΑΤΙΚΟΣ ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ
ΚΑΤΑΛΟΙΠΩΝ



ΔΙΑΓΡΑΜΜΑ 20: ΑΦΑΙΡΕΣΗ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

