

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Μάθημα: *Ανάλυση Δεδομένων*

Διδάσκων: *Ι.Ντζούφρας*

Εργασία: *Ανάλυση απόδοσης παικτών guard του NBA*

Φοιτητής

Ονοματεπώνυμο: *Περίχαρος Παναγιώτης*

Αριθμός Μητρώου: *6140099*

Έτος Σπουδών: *4ο*

Ημερομηνία Παράδοσης: *13 Σεπτεμβρίου 2018*

Περιεχόμενα

1	Εισαγωγή-Περιγραφή μελέτης και προβλήματος	3
2	Περιγραφική ανάλυση	4
3	Σχέσεις μεταβλητών ανα δύο	6
4	Προβλεπτικά ή ερμηνευτικά μοντέλα	7
5	Συμπέρασμα	9
6	Βιβλιογραφία	10
7	Παράρτημα	11

1 Εισαγωγή-Περιγραφή μελέτης και προβλήματος

Η καλαθοσφαίριση αποτελεί ένα από τα πιο δημοφιλή αθλήματα σε ολόκληρο τον κόσμο. Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι αθλητικές επιδόσεις στο άθλημα από τους επαγγελματίες καλαθοσφαιριστές. Η εργασία αφορά τις επιδόσεις των guard του NBA για την περίοδο 1993-1994.

Σκοπός της εργασίας είναι:

- Να εξεταστεί κατά πόσο η ηλικία επηρεάζει την επίδοση των αθλητών.
- Να ελέγξουμε την επίδοση των αθλητών όσον αφορά τους πόντους που πετυχαίνουν ανά αγώνα.
- Η δημιουργία ενός μοντέλου που θα καθορίζει ποιοι παίκτες έχουν προοπτικές για αυξημένο χρόνο συμμετοχής.

Τα χαρακτηριστικά (για κάθε παίκτη) που θα υποστούν ανάλυση είναι:

1. Ύψος (height)
2. Αριθμός αγώνων στους οποίους συμμετείχε (num.match)
3. Συνολικός χρόνος σε λεπτά που αγωνίστηκε (time.min)
4. Ηλικία (age)
5. Πόντοι ανά αγώνα (points)
6. Assist ανά αγώνα (assist)
7. Rebound ανά αγώνα (rebound)
8. Ποσοστό καλαθιών εντός γραμμής τριπόντου (per.in.area)
9. Ποσοστό ευστοχίας ελευθέρων βολών (per.freethrows)

Με την ανάλυση των προαναφερθέντων χαρακτηριστικών, που θα περιγράψουμε στην συνέχεια της εργασίας, θα αποκτήσουμε μια πιο ενδελεχή εικόνα για τις επιδόσεις των παικτών. Τα συμπεράσματα που θα προκύψουν από την ανάλυση θα παρουσιάσουν ενδιαφέρον για τις ομάδες του NBA σχετικά με το πότε μια μεταγραφή θα θεωρηθεί πετυχημένη.

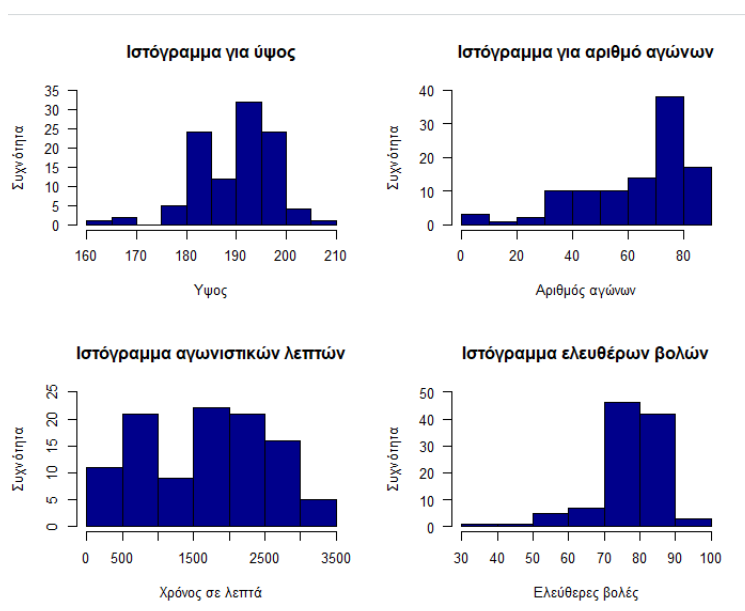
2 Περιγραφική ανάλυση

Στην συνέχεια θα ακολουθήσει μια συνοπτική περιγραφή των δεδομένων για να έχουμε μια αρχική εικόνα της κατανομής τους. Αρχικά με τον παρακάτω πίνακα παρατηρούμε για τις μεταβλητές μας μερικά περιγραφικά στοιχεία πιο συγκεκριμένα τον μέσο όρο, την ελάχιστη και μέγιστη τιμή.

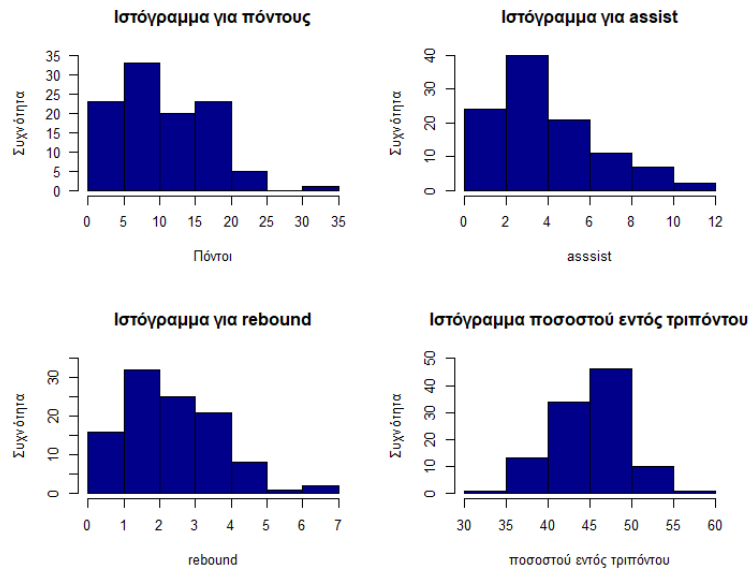
<i>height</i>	<i>num.match</i>	<i>time.min</i>	<i>age</i>	<i>points</i>	<i>assist</i>	<i>rebound</i>	<i>per.in.area</i>	<i>per.free.throws</i>
Min. :160.0	Min. : 8.00	Min. : 55	Min. :22.00	Min. : 1.7	Min. : 0.300	Min. :0.300	Min. :34.80	Min. :37.50
Mean :190.2	Mean :64.15	Mean :1656	Mean :27.53	Mean :10.7	Mean : 3.996	Mean :2.395	Mean :45.13	Mean :77.86
Max. :210.0	Max. :82.00	Max. :3117	Max. :37.00	Max. :32.6	Max. :12.000	Max. :6.700	Max. :59.50	Max. :94.80

Σχήμα 1: Μέσος όρος, μέγιστη και ελάχιστη τιμή.

Έπειτα για την καλύτερη κατανόηση των δεδομένων μας θα ακολουθήσει η γραφική τους αναπαράσταση με ιστογράμματα (για ποσοτικές μεταβλητές) και ραβδόγραμματα (για τις ποιοτικές) για τις παραμέτρους του προβλήματός μας. Ειδικότερα το ραβδόγραμμα απεικονίζει την συχνότητα εμφάνισης των κατηγοριών στις οποίες είναι ταξινομημένα τα δεδομένα μας. Το ιστόγραμμα είναι γραφική απεικόνιση στατιστικών συχνοτήτων περιοχών τιμών ενός μεγέθους και σχηματίζεται από παρακείμενα ορθογώνια. Η επιφάνεια κάθε ορθογωνίου είναι μέτρο της συχνότητας εμφάνισης της συγκεκριμένης περιοχής τιμών ενώ το ύψος του ισούται με το λόγο της συχνότητας προς το εύρος των τιμών που αντιπροσωπεύει το ορθογώνιο.

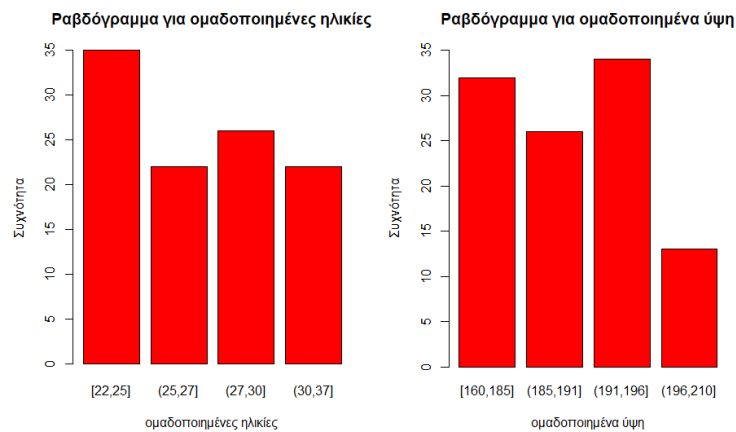


Σχήμα 2: Ιστόγραμμα των μεταβλητών ύψος, αριθμός αγώνων, χρόνος σε λεπτά, ποσοστό ευστοχίας ελεύθερων βολών



Σχήμα 3: Ιστόγραμμα των μεταβλητών πόντοι ανα αγώνα, assist ανα αγώνα, rebound ανα αγώνα, ποσοστό ευστοχίας εντός γραμμής τριπόντου

Μετατρέψαμε τις μεταβλητές ηλικία και Ύψος απο ποσοτικές σε ποιοτικές μεταβλητές, για την καλύτερη διαγραμματική παρουσίαση τους και για πιο εύκολη κατανόηση της δομής τους.



Σχήμα 4: Ραβδογράμματα για τις μετατρεπόμενες σε ποιοτικές μεταβλητές ύψος και ηλικία

3 Σχέσεις μεταβλητών ανα δύο

Σε αυτό το κεφάλαιο θα γίνει μια πιο διεξοδική εξέταση του κατά πόσο η ηλικία επηρεάζει την επίδοση των αθλητών. Πιο συγκεκριμένα θα μελετήσουμε αν οι παράμετροι μας πόντοι, assist, rebound ανα αγώνα για κάθε παίκτη επηρεάζονται από την ηλικία. Καθώς επίσης και την επιρροή της στο ποσοστό καλαθιών εντός γραμμής τριπόντου, το ποσοστό ευστοχίας ελεύθερων βολών, τον συνολικό χρόνο σε λεπτά που αγωνίστηκε ο παίκτης και τον αριθμό αγώνων στους οποίους συμμετείχε.

Για την εξέταση της επιρροής της ηλικίας στις επιδόσεις των αθλητών έγινε χρήση της ανάλυσης διακύμανσης κατά έναν παράγοντα (one-way anova) καθώς χρησιμοποιήσαμε την μετατρεπόμενη μεταβλητή της ηλικίας (την ομαδοποιημένη ηλικία) που είναι ποιοτική με πολλά επίπεδα και μια ποσοτική μεταβλητή (κάποια παράμετρο από τις επιδόσεις). Επιπρόσθετα χρησιμοποιήσαμε τον έλεγχο Kruskal wallis όταν η ανάλυση διακύμανσης κατά έναν παράγοντα δεν ήταν εφαρμόσιμη και ο μέσος όρος δεν ήταν το κατάλληλο μέτρο περιγραφής της κεντρικής θέσης. Άλλες τεχνικές που χρησιμοποιήθηκαν για την εξέταση των σχέσεων των μεταβλητών μας είναι: shapiro-wilk normality test, Levene's Test for Homogeneity of Variance, Pairwise comparisons using Wilcoxon rank sum test, Tukey multiple comparisons of means, Dunn (1964) Kruskal-Wallis multiple comparison p-values adjusted with the Bonferroni method.

Τα συμπεράσματα που εξάγονται από τους παραπάνω ελέγχους είναι τα εξής:

- Οι πόντοι ανα αγώνα φαίνεται να διαφέρουν ανάμεσα στις ηλικιακές ομάδες από 22 έως και 25 και 27 έως και 30.
- Οι assist ανα αγώνα διαφέρουν μεταξύ των ηλικιακών ομάδων από 27 έως και 30 και 30 έως και 37, ενώ παράλληλα παρατηρείται διαφορά και μεταξύ των 22 έως και 25 και 27 έως και 30.
- Τα rebound ανα αγώνα διαφέρουν για τις ηλικιακές ομάδες από 22 έως και 25 και 27 έως και 30, καθώς επίσης μεταξύ των ηλικιών από 27 έως και 30 και 30 έως και 37.
- Ο αριθμός των αγώνων συμμετοχής φαίνεται να διαφέρει για τις ηλικιακές ομάδες από 22 έως και 25 και 27 έως και 30.
- Ως προς το ποσοστό καλαθιών εντός γραμμής τριπόντου δεν παρατηρείται κάποια στατιστικά σημαντική διαφορά λόγω της ηλικίας.
- Ως προς το ποσοστό ευστοχίας ελεύθερων βολών παρατηρείται διαφορά ανάμεσα στις ηλικιακές ομάδες από 22 έως και 25 και 27 έως και 30 ενώ παρατηρείται επίσης διαφορά στις ηλικίες από 25 έως και 27 και 27 έως και 30.
- Τέλος παρατηρούμε διαφορές στον συνολικό χρόνο σε λεπτά που αγωνίστηκε ο κάθε παίκτης ανάμεσα στις ηλικιακές ομάδες από 22 έως και 25 και από 27 έως και 30 και στις ομάδες από 27 έως και 30 και 30 έως και 37.

4 Προβλεπτικά ή ερμηνευτικά μοντέλα

Στο κεφάλαιο αυτό θα ασχοληθούμε με την δημιουργία ενός μοντέλου με σκοπό να καθορίσουμε ποιοι παίκτες έχουν προοπτικές για αυξημένο συνολικό χρόνο συμμετοχής, με βάση τα υπόλοιπα χαρακτηριστικά που διαθέτουμε για τον κάθε παίκτη .

Αρχικά σε περίπτωση προσαρμογής του πλήρους μοντέλου για την πρόβλεψη (δηλαδή μοντέλο της μορφής συνολικός χρόνος σε λεπτά = σταθερά + α *ύψος + β *αριθμός αγώνων + γ *ηλικία + δ *πόντοι ανα αγώνα + ϵ *assist ανα αγώνα + ζ *rebound ανα αγώνα + η *ποσοστό καλαθιών εντός γραμμής τριπόντου + θ *ποσοστό ευστοχίας ελευθέρων βολών), παρατηρούμε πως έχει καλή προσαρμογή στα δεδομένα (υψηλό R^2 adjusted). Ακόμα παρατηρούμε ότι πληρούνται οι προϋποθέσεις του γραμμικού μας μοντέλου (κανονικότητα, ανεξαρτησία και ομοσκεδαστικότητα καταλοίπων) .Ωστόσο υπάρχει ένα πρόβλημα, η ερμηνεία της σταθεράς του μοντέλου δεν έχει νόημα .Συνεπώς λόγω του προβλήματος αυτού οδηγούμαστε στο συμπέρασμα ότι για την καλύτερη συμπερασματολογία οι μεταβλητές που θα χρησιμοποιηθούν θα είναι κεντραρισμένες ως προς τον μέσο τους (για παράδειγμα αυτό σημαίνει ότι η κεντραρισμένη μεταβλητή των πόντων θα ισούται ως οι πόντοι που πετυχαίνει κάθε παίκτης συνολικά πλην τον μέσο όρο των συνολικών πόντων που πετυχαίνουν όλοι οι παίκτες). Συνεπώς το μοντέλο θα είναι της μορφής $Y = b_0. + b_1. * (X_1 - \bar{X}_1) + \dots + b_j. * (X_j - \bar{X}_j)$.Με την κεντροποίηση δεν επηρεάζονται καθόλου οι έλεγχοι των υποθέσεων του μοντέλου.Η μόνη αλλαγή που γίνεται είναι στην ερμηνεία των εκτιμήσεων των συντελεστών του μοντέλου μας.Τα αποτελέσματα που λαμβάνουμε απο το κεντροποιημένο μοντέλο είναι :

```
call:
lm(formula = NBA.data$time.min ~ point.centered + rebound.centered +
  assist.centered + heigth.centered + num.much.centered + age.centered +
  area.centered + frthr.centered)
```

Residuals:

Min	1Q	Median	3Q	Max
-620.46	-154.91	14.36	161.17	574.55

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1656.486	23.404	70.779	< 2e-16 ***
point.centered	60.405	7.287	8.290	6.90e-13 ***
rebound.centered	52.507	33.338	1.575	0.119
assist.centered	75.273	15.269	4.930	3.44e-06 ***
heigth.centered	2.715	4.290	0.633	0.528
num.much.centered	23.759	1.436	16.544	< 2e-16 ***
age.centered	6.219	7.357	0.845	0.400
area.centered	3.807	5.944	0.640	0.523
frthr.centered	-2.282	2.997	-0.761	0.448

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

Residual standard error: 239.8 on 96 degrees of freedom
 Multiple R-squared: 0.9323, Adjusted R-squared: 0.9267
 F-statistic: 165.3 on 8 and 96 DF, p-value: < 2.2e-16

Σχήμα 5: Αρχικό κεντροποιημένο μοντέλο.

Στην συνέχεια παρατηρούμε ότι δεν είναι όλες οι μεταβλητές στατιστικά σημαντικές (το καταλαβαίνουμε μέσω του t test). Αυτό σημαίνει ότι δεν είναι ιδιαίτερα χρήσιμες και ότι μπορούν να αφαιρεθούν από αυτό χωρίς να χάνουμε στατιστικά σημαντική πληροφορία. Προτιμάμε συνεπώς ένα μοντέλο στο οποίο έχουν απαλειφθεί οι στατιστικά μη σημαντικές μεταβλητές και αυτό θα επιτευχθεί μέσω των μεθόδων stepwise procedure, backward procedure. Επίσης θα χρησιμοποιήσουμε την μέθοδο Akaike's Information criterion (AIC) με στόχο το βέλτιστο μοντέλο. Τα αποτελέσματα τα οποία εξάγουμε από τις παραπάνω μεθόδους για το καινούργιο κεντροποιημένο μοντέλο είναι :

```
call:
lm(formula = NBA.data$time.min ~ point.centered + rebound.centered +
    assist.centered + num.much.centered)

Residuals:
    Min       1Q   Median       3Q      Max
-615.20 -161.34   5.14  138.85  583.61

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1656.486    23.169  71.497 < 2e-16 ***
point.centered     61.091     6.373   9.586 8.00e-16 ***
rebound.centered   59.180    30.679   1.929  0.0566 .
assist.centered    70.566    12.403   5.690 1.27e-07 ***
num.much.centered  23.610     1.382  17.084 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 237.4 on 100 degrees of freedom
Multiple R-squared:  0.9309,    Adjusted R-squared:  0.9281
F-statistic: 336.8 on 4 and 100 DF,  p-value: < 2.2e-16
```

Σχήμα 6: Καινούργιο κεντροποιημένο μοντέλο.

Σύμφωνα με τα αποτελέσματα που εξάγουμε από τον παραπάνω πίνακα το βέλτιστο μοντέλο μας θα έχει την μορφή :

- Συνολικός χρόνος σε λεπτά = $1656.5 + 61 \cdot \text{κεντροποιημένοι πόντοι} + 59 \cdot \text{κεντροποιημένα rebound} + 70 \cdot \text{κεντροποιημένες assist} + 23 \cdot \text{κεντροποιημένοι αριθμοί αγώνων συμμετοχής}$

Παράλληλα παρατηρούμε από το outpout της R υψηλή τιμή του συντελεστή προσδιορισμού R^2 adjusted= 0.928 (το οποίο σημαίνει ότι το 92 τοις εκατό της διακύμανσης εκφράζεται από το επιλεγμένο μοντέλο.), ενώ βλέπουμε και τη τιμή του τυπικού σφάλματος των εκτιμητών (Std.Error), τη τιμή του T στατιστικού (t value), το p-value και την τυπική απόκλιση των σφαλμάτων (Residual standard error) ε ακολουθούν $N(0, 237.4^2)$).

Η ερμηνεία των συντελεστών του καινούργιου κεντροποιημένου γραμμικού μοντέλου είναι η εξής :

- Για την σταθερά : Η αναμενόμενη τιμή του συνολικού χρόνου σε λεπτά για κάθε παίκτη είναι 1656.5 λεπτά όταν οι παράμετροι μας πόντοι, rebound, assist και αριθμός αγώνων συμμετοχής είναι ίσοι με τους δειγματικούς μεσους τους ή αλλιώς ένας μέσος παίκτης παίζει συνολικά 1656.5 λεπτά σε μια season .
- Για τον συντελεστή των κεντροποιημένων πόντων : Ο αναμενόμενος συνολικός χρόνος σε λεπτά όταν η παράμετρος των κεντροποιημένων πόντων αυξηθεί κατά μία μονάδα δοθέντος ότι οι υπόλοιπες κεντροποιημένες μεταβλητές παραμένουν σταθερές.
- Για τον συντελεστή των κεντροποιημένων rebound : Ο αναμενόμενος συνολικός χρόνος σε λεπτά όταν η παράμετρος των κεντροποιημένων rebound αυξηθεί κατά μία μονάδα και υπόλοιπες κεντροποιημένες μεταβλητές παραμένουν σταθερές.
- Για τον συντελεστή των κεντροποιημένων assist : Ο αναμενόμενος συνολικός χρόνος σε λεπτά όταν η παράμετρος των κεντροποιημένων assist αυξηθεί κατά μία μονάδα και υπόλοιπες κεντροποιημένες μεταβλητές παραμένουν σταθερές.
- Για τον συντελεστή των κεντροποιημένων αριθμών αγώνων συμμετοχής : Ο αναμενόμενος συνολικός χρόνος σε λεπτά όταν η παράμετρος των κεντροποιημένων αριθμών αγώνων συμμετοχής αυξηθεί κατά μία μονάδα και υπόλοιπες κεντροποιημένες μεταβλητές παραμένουν σταθερές.

5 Συμπέρασμα

Εν κατακλείδι τα αποτελέσματα που μπορούμε να εξάγουμε από την παραπάνω έρευνα ως προς το μοντέλο είναι η πολύ καλή προσαρμογή του (R^2 adjusted= 0.928) για τα δεδομένα που αφορούν τις χρονολογίες 1993 και 1994. Όμως λόγω της παλαιότητας των δεδομένων δεν μπορούμε να κάνουμε καποια ουσιαστική πρόβλεψη για τα σημερινά δεδομένα λόγω μεγάλου σφάλματος (μεγάλο extrapolation) συνεπώς δεν θα ήταν αξιόπιστη η εκτίμηση των επιδόσεων. Ως προς το κεφάλαιο 3 εξάγουμε το συμπέρασμα ότι υπάρχει διαφορά στις επιδόσεις των αθλητών εκτός από τις επιδόσεις που αφορούν το ποσοστό καλαθιών εντός γραμμής τριπόντου. Τέλος σε ότι αφορά το κεφάλαιο ένα παρατηρούμε στις μεταβλητές ύψος, αριθμός αγώνων, rebound και ελεύθερες βολές να υπάρχει αρνητική ασυμμετρία που σημαίνει αύξηση των συχνοτήτων προς της μεγάλες τιμές. Αντίθετα παρατηρούμε θετική ασυμμετρία στις μεταβλητές πόντοι και assist ανά αγώνα καθώς και στο ποσοστό ευστοχίας εντός τριπόντου.

6 Βιβλιογραφία

ΒΙΒΛΙΑ-ΣΗΜΕΙΩΣΕΙΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ:

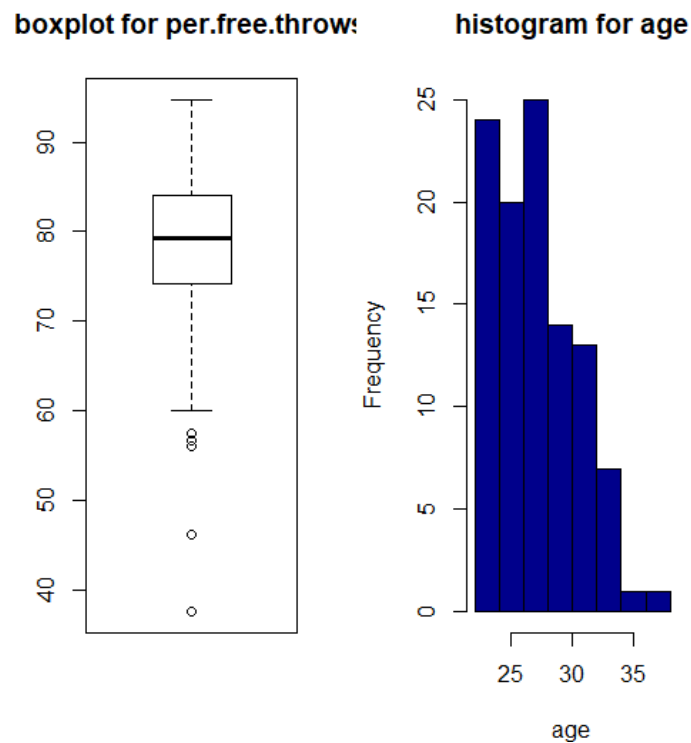
- Ανάλυση Δεδομένων με Χρήση της R (Δ.ΦΟΥΣΚΑΚΗΣ)
- Εισαγωγή στην Στατιστική Ανάλυση με το R (Crawley)
- στατιστική (ΑΓΓΕΛΗΣ Β. ΔΗΜΑΚΗ Κ.)
- ΣΗΜΕΙΩΣΕΙΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ 2016-2017 ΙΩΑΝΝΗΣ ΝΤΖΟΥΦΡΑΣ.
- ΣΗΜΕΙΩΣΕΙΣ ΜΑΘΗΜΑΤΟΣ 'ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ' ΠΑΝΑΓΙΩΤΗΣ ΤΣΙΑΜΥΡΤΖΗΣ.

Τα πακέτα που χρησιμοποιήθηκαν στην R είναι : sjplot, psych, corrplot, FSA, car, PerformanceAnalytics,

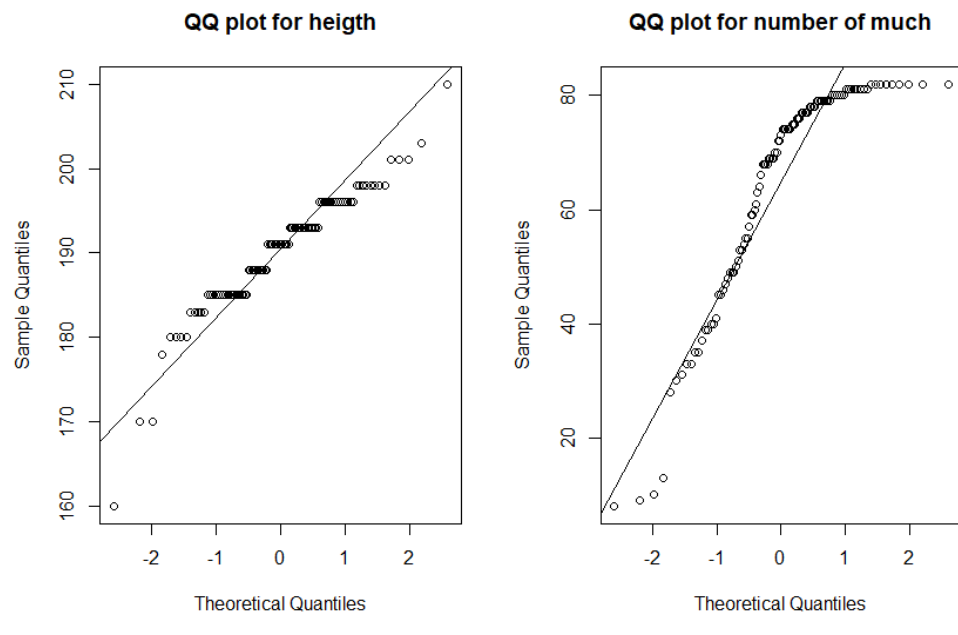
7 Παράρτημα

<i>heighth</i>	<i>num.match</i>	<i>time.min</i>	<i>age</i>	<i>points</i>	<i>assist</i>	<i>rebound</i>	<i>per.in.area</i>	<i>per.free.throws</i>
Min. :160.0	Min. : 8.00	Min. : 55	Min. :22.00	Min. : 1.7	Min. : 0.300	Min. :0.300	Min. :34.80	Min. :37.50
1st Qu.:185.0	1st Qu.:51.00	1st Qu.: 881	1st Qu.:25.00	1st Qu.: 5.4	1st Qu.: 2.200	1st Qu.:1.500	1st Qu.:42.80	1st Qu.:74.20
Median :191.0	Median :73.00	Median :1622	Median :27.00	Median : 9.3	Median : 3.300	Median :2.200	Median :45.30	Median :79.30
Mean :190.2	Mean :64.15	Mean :1656	Mean :27.53	Mean :10.7	Mean : 3.996	Mean :2.395	Mean :45.13	Mean :77.86
3rd Qu.:196.0	3rd Qu.:79.00	3rd Qu.:2430	3rd Qu.:30.00	3rd Qu.:15.1	3rd Qu.: 5.400	3rd Qu.:3.200	3rd Qu.:47.80	3rd Qu.:84.00
Max. :210.0	Max. :82.00	Max. :3117	Max. :37.00	Max. :32.6	Max. :12.000	Max. :6.700	Max. :59.50	Max. :94.80

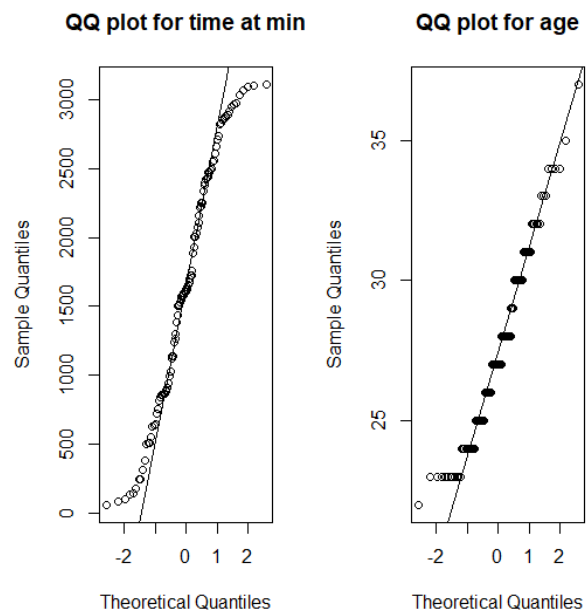
Σχήμα 7: Βασικά περιγραφικά μέτρα για όλες τις παραμέτρους.



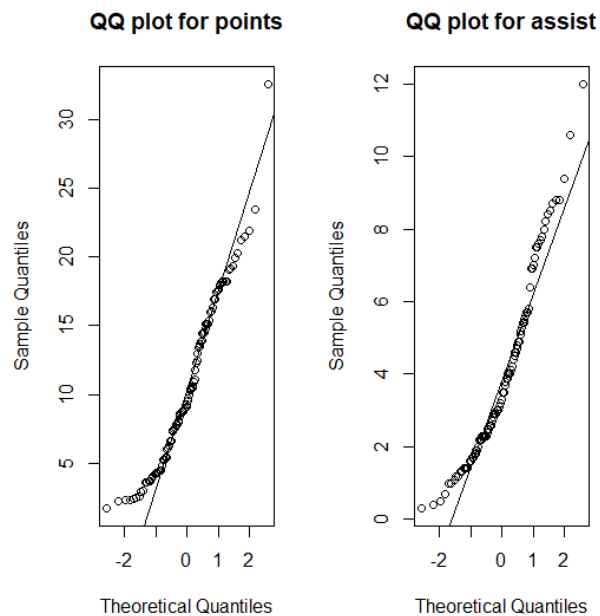
Σχήμα 8: Ιστογράμμο και διάγραμμα πλαισίου απολήξεων για την ποσοτική μεταβλητή ηλικία.



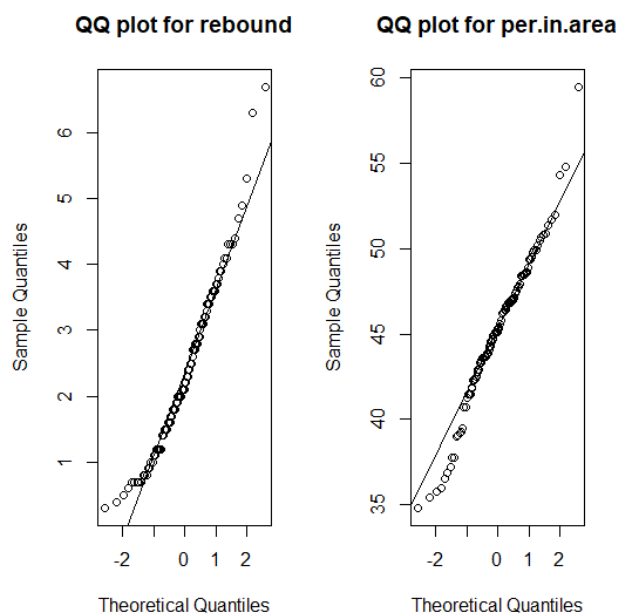
Σχήμα 9: Διάγραμμα για έλεγχο κανονικότητας των μεταβλητών ύψος και αριθμός αγώνων.



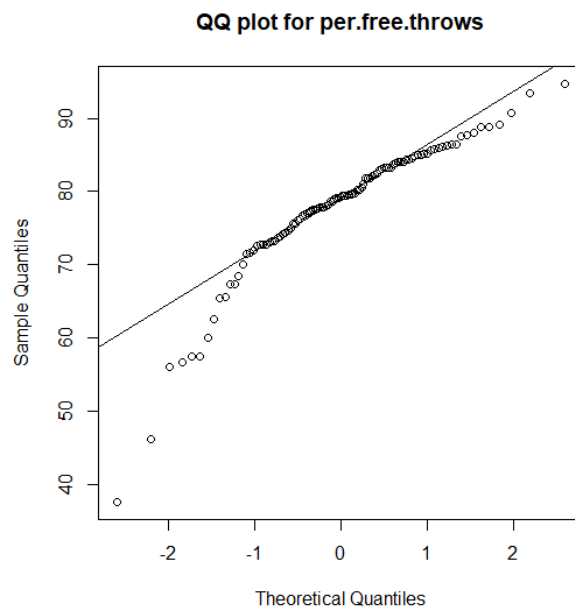
Σχήμα 10: Διάγραμμα για έλεγχο κανονικότητας των μεταβλητών χρόνος σε λεπτά και ηλικία.



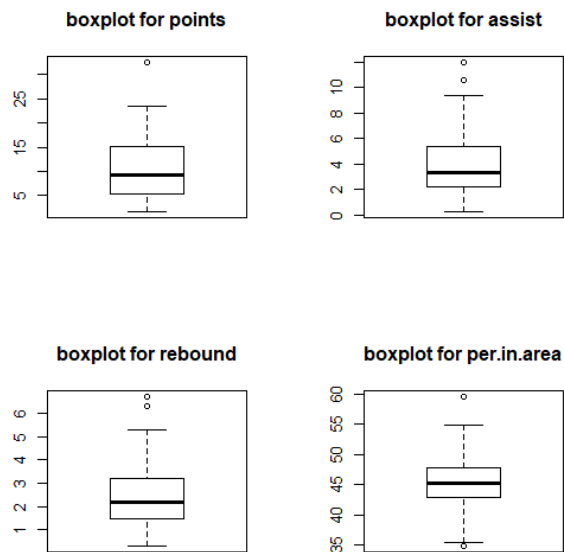
Σχήμα 11: Διάγραμμα για έλεγχο κανονικότητας των μεταβλητών πόντοι ανα αγώνα και assist ανα αγώνα.



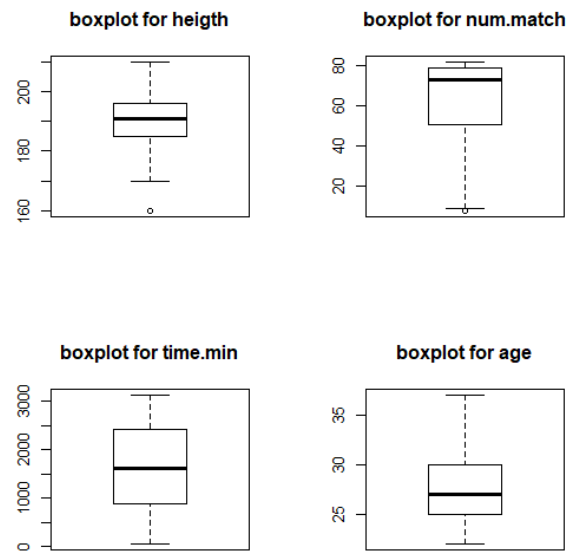
Σχήμα 12: Διάγραμμα για έλεγχο κανονικότητας των μεταβλητών rebound ανα αγώνα, ποσοστό καλαθιών δυο πόντων.



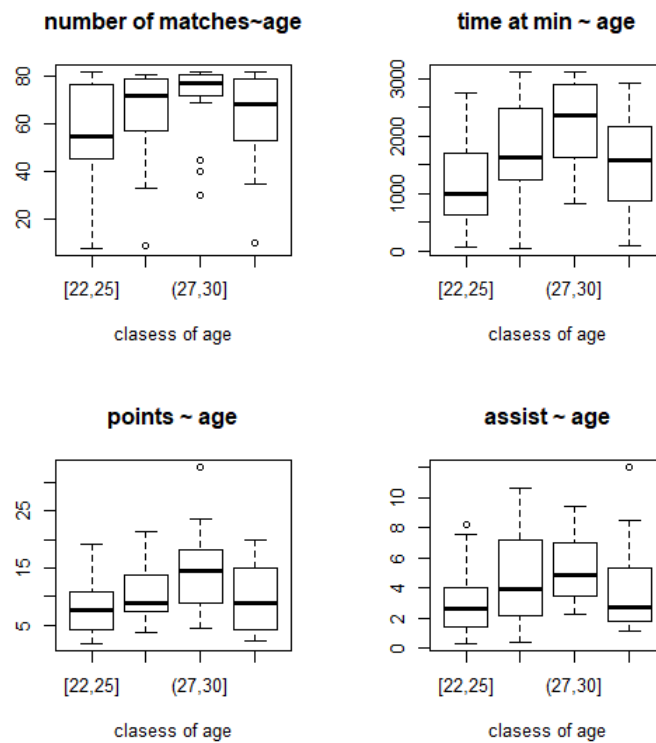
Σχήμα 13: Διάγραμμα για έλεγχο κανονικότητας της μεταβλητής ποσοστό ευστοχίας ελεύθερων βολών.



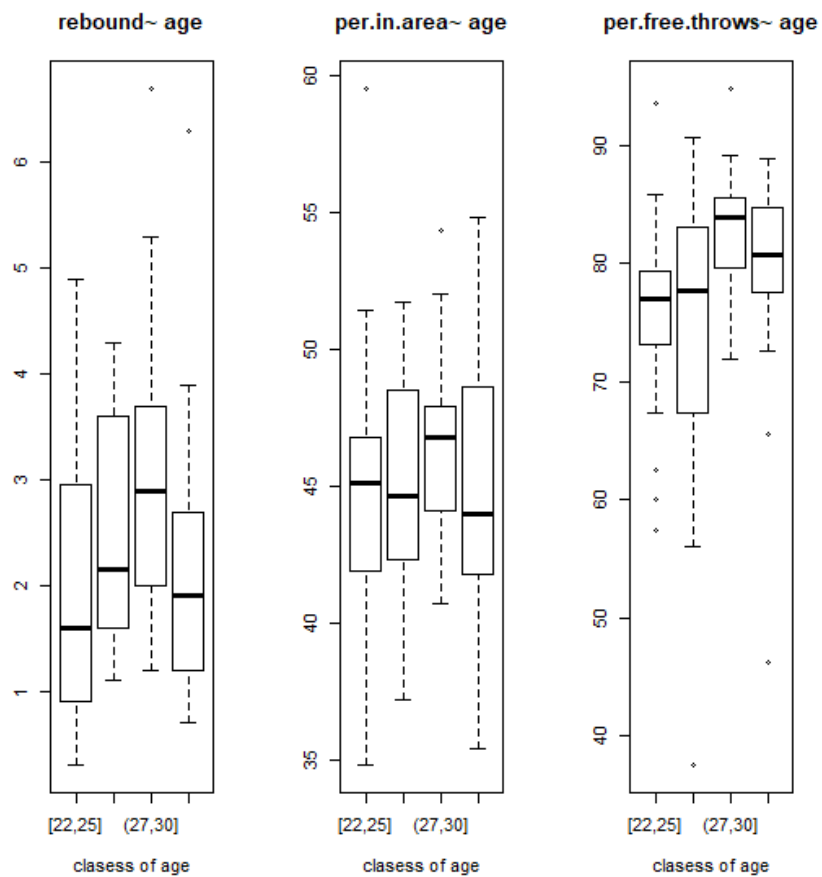
Σχήμα 14: Διάγραμμα πλαισίου απολήξεων για τις μεταβλητές πόντοι,assist,rebound ανα αγώνα,ποσοστό καλαθιών δυο πόντων.



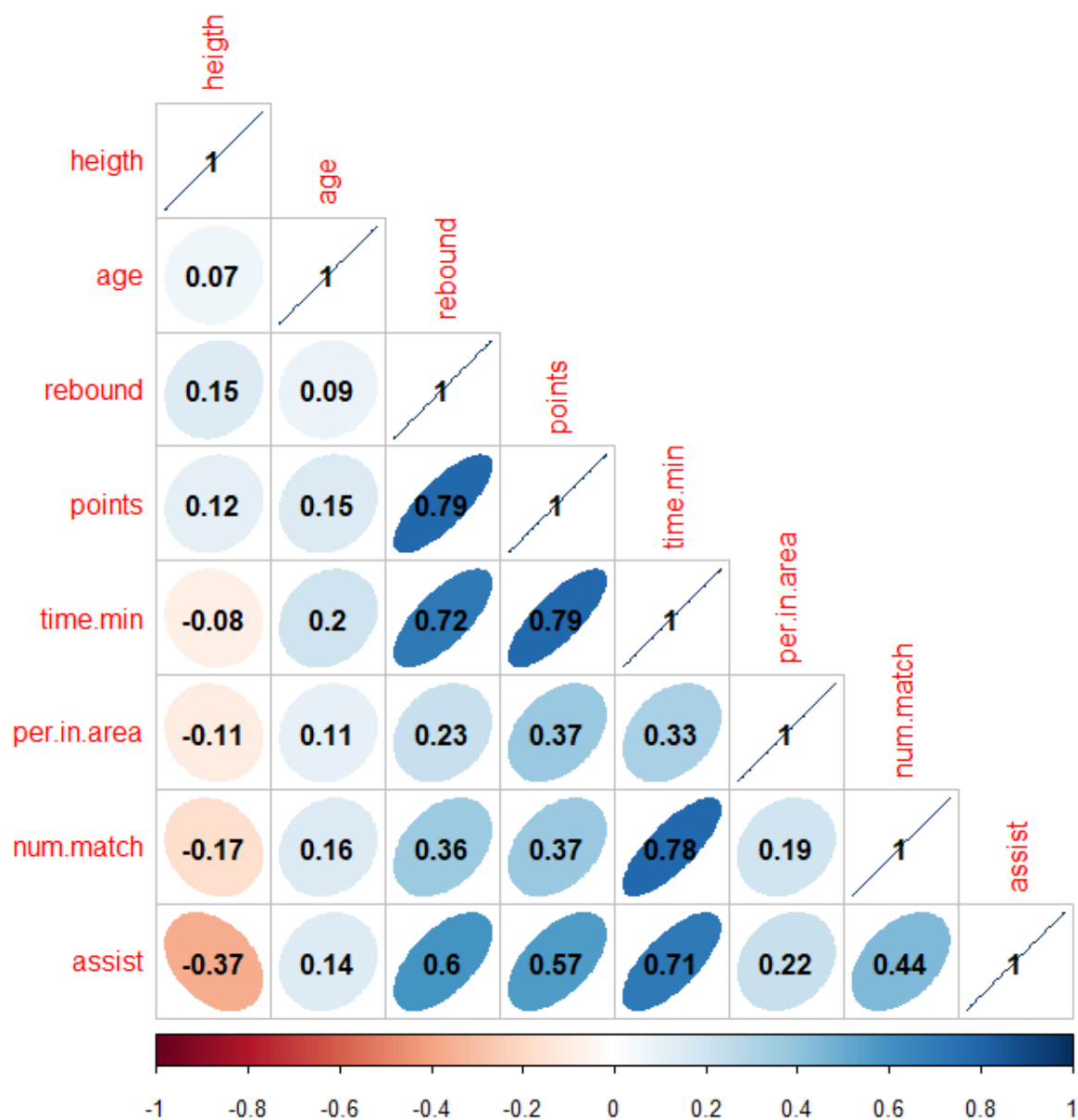
Σχήμα 15: Διάγραμμα πλαισίου απολήξεων για τις μεταβλητές ύψος, αριθμός αγώνων, χρόνος σε λεπτά, ηλικία.



Σχήμα 16: Διάγραμμα πλαισίου απολήξεων ανα ομάδες της κατηγορικής μεταβλητής ηλικία.



Σχήμα 17: Διάγραμμα πλαισίου απολήξεων ανα ομάδες της κατηγορικής μεταβλητής ηλικία.



Σχήμα 18: Πίνακας συσχετίσεων των μεταβλητών.

```
> shapiro.test(NBA.data$height) ##P-value<0.05 aporiptw Ho ara
      shapiro-wilk normality test
data:  NBA.data$height
W = 0.93095, p-value = 3.709e-05

>
> shapiro.test(NBA.data$num.match) ## P value<0.05 aporiptw Ho
      shapiro-wilk normality test
data:  NBA.data$num.match
W = 0.83746, p-value = 2.223e-09

>
> shapiro.test(NBA.data$time.min) ## P value<0.05 aporiptw Ho
      shapiro-wilk normality test
data:  NBA.data$time.min
W = 0.95746, p-value = 0.001958

>
> shapiro.test(NBA.data$age) ## P value<0.05 aporiptw Ho
      shapiro-wilk normality test
data:  NBA.data$age
W = 0.95395, p-value = 0.001102
```

Σχήμα 19: Ελέγχοι κανονικότητας για τις παραμέτρους height, num,much, time.min, age.

```
> shapiro.test(NBA.data$points) ## P value<0.05 aporiptw Ho
      shapiro-wilk normality test
data:  NBA.data$points
W = 0.94825, p-value = 0.0004486
>
> shapiro.test(NBA.data$assist) ## P value<0.05 apor Ho
      shapiro-wilk normality test
data:  NBA.data$assist
W = 0.92988, p-value = 3.214e-05
>
> shapiro.test(NBA.data$rebound) ## P value<0.05 apor Ho
      shapiro-wilk normality test
data:  NBA.data$rebound
W = 0.95859, p-value = 0.002365
>
> shapiro.test(NBA.data$per.in.area) ## P value>0.05 DEN APORIPTW Ho
      shapiro-wilk normality test
data:  NBA.data$per.in.area
W = 0.98308, p-value = 0.2019
>
> shapiro.test(NBA.data$per.freethrows) ## P value<0.05 aporiptw Ho
      shapiro-wilk normality test
data:  NBA.data$per.freethrows
W = 0.88654, p-value = 2.02e-07
```

Σχήμα 20: Ελέγχοι κανονικότητας για τις παραμέτρους points, assist, rebound, per.in.area, per.freethrows.

```
> anova1<-aov(NBA.data$points~classage,data = NBA.data)
> summary(anova1)
              Df Sum Sq Mean Sq F value    Pr(>F)
classage       3     613   204.47    6.446 0.000488 ***
Residuals    101    3204    31.72
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> ### gia kanonikotita kataloipwn
> shapiro.test(anova1$res)#aporiptw ho p-value<0.05
```

shapiro-wilk normality test

```
data:  anova1$res
W = 0.95732, p-value = 0.001911
```

```
> #ara aporiptw ipothesi kanonikotitas kataloipwn
> ## ARA KRUSKAL TEST
> kruskal.test(NBA.data$points~classage,data = NBA.data)
```

kruskal-wallis rank sum test

```
data:  NBA.data$points by classage
kruskal-wallis chi-squared = 15.164, df = 3, p-value = 0.001682
```

Σχήμα 21: Ελεγχος ισότητας διαμέσων για την μεταβλητή points ανα επίπεδο της κατηγορικής classage .

```
> dunnTest(NBA.data$points~classage,data = NBA.data,method = "bonferroni")
Dunn (1964) kruskal-wallis multiple comparison
p-values adjusted with the Bonferroni method.

      Comparison      Z      P.unadj      P.adj
1 (25,27] - (27,30] -2.0187114 0.0435172255 0.2611033529
2 (25,27] - (30,37]  0.5445723 0.5860477149 1.0000000000
3 (27,30] - (30,37]  2.5855202 0.0097232174 0.0583393046
4 (25,27] - [22,25]  1.4888859 0.1365174043 0.8191044259
5 (27,30] - [22,25]  3.8232906 0.0001316824 0.0007900947
6 (30,37] - [22,25]  0.8854000 0.3759409108 1.0000000000
> ## ara stis ilikiakes omades (27,30] kai [22,25] parathritai diafora stis diammesous
> ##allos tropos
> pairwise.wilcox.test(NBA.data$points,classage,p.adjust.method = "bonferroni")
```

Pairwise comparisons using wilcoxon rank sum test

```
data:  NBA.data$points and classage
```

```
      [22,25] [25,27] [27,30]
(25,27] 0.69326 -      -
(27,30] 0.00098 0.14454 -
(30,37] 1.00000 1.00000 0.11284
```

```
P value adjustment method: bonferroni
```

Σχήμα 22: Συγκρίσεις ανα δύο με διόρθωση bonferroni για τις μεταβλητές classage, points .

```
> anova2<-aov(NBA.data$assist~classage,data=NBA.data)
> summary(anova2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
classage	3	85.9	28.627	5.173	0.00229 **
Residuals	101	558.9	5.534		

```
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #gia kanonikotita kataloipwn
> shapiro.test(anova2$res)### aporiptw ho p-value<0.05

      Shapiro-Wilk normality test

data:  anova2$res
W = 0.93103, p-value = 3.747e-05

> ##aporiptw ipothsi kanonikotitas katalipwn
> kruskal.test(NBA.data$assist~classage,data = NBA.data)##aporiptw ho

      Kruskal-Wallis rank sum test

data:  NBA.data$assist by classage
kruskal-wallis chi-squared = 16.906, df = 3, p-value = 0.0007391
```

Σχήμα 23: Ελεγχος ισότητας διαμέσων για την μεταβλητή assist ανα επίπεδο της κατηγορικής classage .

```
> dunnTest(NBA.data$assist~classage,data = NBA.data,method = "bonferroni")
Dunn (1964) kruskal-wallis multiple comparison
p-values adjusted with the Bonferroni method.
```

	Comparison	Z	P.unadj	P.adj
1	(25,27] - (27,30]	-1.3657029	1.720322e-01	1.0000000000
2	(25,27] - (30,37]	1.2551836	2.094121e-01	1.0000000000
3	(27,30] - (30,37]	2.6721394	7.536934e-03	0.0452216045
4	(25,27] - [22,25]	2.2717274	2.310298e-02	0.1386178948
5	(27,30] - [22,25]	3.9153234	9.028312e-05	0.0005416987
6	(30,37] - [22,25]	0.8807539	3.784510e-01	1.0000000000

```
> #diafora stois (27,30]-(30,37] kai (27,30]-[22,25]
> ##allos tropos
> pairwise.wilcox.test(NBA.data$assist,classage,p.adjust.method = "bonferroni")

      Pairwise comparisons using wilcoxon rank sum test

data:  NBA.data$assist and classage
```

	[22,25]	(25,27]	(27,30]
(25,27]	0.17512 -	-	
(27,30]	0.00057 1.00000 -		
(30,37]	1.00000 1.00000 0.04270		

```
P value adjustment method: bonferroni
```

Σχήμα 24: Συγκρίσεις ανα δύο με διόρθωση bonferroni για τις μεταβλητές classage, assist .

```

> anova3<-aov(NBA.data$rebound~classage,data=NBA.data)
> summary(anova3)
              Df Sum Sq Mean Sq F value Pr(>F)
classage       3  17.93   5.976   3.908  0.011 *
Residuals    101 154.46   1.529
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #για κανονικότητα καταλοίπων
> shapiro.test(anova3$res)### aporiptw ho p-value<0.05

      Shapiro-wilk normality test

data:  anova3$res
W = 0.94167, p-value = 0.000167

> ##aporiptw ipothsi κανονικότητας καταλοίπων
> kruskal.test(NBA.data$rebound~classage,data = NBA.data)##aporiptw ho

      Kruskal-wallis rank sum test

data:  NBA.data$rebound by classage
kruskal-wallis chi-squared = 12.021, df = 3, p-value = 0.007311

```

Σχήμα 25: Ελεγχος ισότητας διαμέσων για την μεταβλητή rebound ανα επίπεδο της κατηγορικής classage .

```

> dunnTest(NBA.data$rebound~classage,data = NBA.data,method = "bonferroni")
Dunn (1964) kruskal-wallis multiple comparison
p-values adjusted with the Bonferroni method.

      Comparison      Z      P.unadj      P.adj
1 (25,27] - (27,30] -1.1748671 0.240047877 1.000000000
2 (25,27] - (30,37]  1.3445880 0.178758316 1.000000000
3 (27,30] - (30,37]  2.5743587 0.010042609 0.060255657
4 (25,27] - [22,25]  1.7708113 0.076592091 0.459552545
5 (27,30] - [22,25]  3.1754073 0.001496263 0.008977578
6 (30,37] - [22,25]  0.2807613 0.778893473 1.000000000
> #διαφορά stis ilikiakes omades (27,30]-[22,25]
> #allos tropos
> pairwise.wilcox.test(NBA.data$rebound,classage,p.adjust.method = "bonferroni")

      Pairwise comparisons using wilcoxon rank sum test

data:  NBA.data$rebound and classage

      [22,25] (25,27] (27,30]
(25,27] 0.362    -        -
(27,30] 0.024    1.000    -
(30,37] 1.000    1.000    0.032

P value adjustment method: bonferroni

```

Σχήμα 26: Συγκρίσεις ανα δύο με διόρθωση bonferroni για τις μεταβλητές classage, rebound .

```
> anova5<-aov(NBA.data$per.freethrows~classage,data=NBA.data)
> summary(anova4)
              Df Sum Sq Mean Sq F value Pr(>F)
classage      3   71.1    23.71   1.237    0.3
Residuals    101 1935.2    19.16
> #για κανονικότητα καταλοίπων
> shapiro.test(anova5$res)##aporiptw ho p-value>0.05

      shapiro-wilk normality test

data:  anova5$res
W = 0.90324, p-value = 1.222e-06

> ##aporiptw ipothsi κανονικότητας καταλοίπων
> kruskal.test(NBA.data$per.freethrows~classage,data = NBA.data)

      kruskal-wallis rank sum test

data:  NBA.data$per.freethrows by classage
kruskal-wallis chi-squared = 18.4, df = 3, p-value = 0.0003637
```

Σχήμα 27: Ελεγχος ισότητας διαμέσων για την μεταβλητή per.freethrows ανα επίπεδο της κατηγορικής classage.

```
> dunnTest(NBA.data$per.freethrows~classage,data = NBA.data,method = "bonferroni")
Dunn (1964) kruskal-wallis multiple comparison
p-values adjusted with the Bonferroni method.

      Comparison      Z      P.unadj      P.adj
1 (25,27] - (27,30] -3.144420 1.664161e-03 0.0099849679
2 (25,27] - (30,37] -1.561843 1.183249e-01 0.7099494791
3 (27,30] - (30,37]  1.518802 1.288122e-01 0.7728733641
4 (25,27] - [22,25]  0.431507 6.660998e-01 1.0000000000
5 (27,30] - [22,25]  3.971644 7.137827e-05 0.0004282696
6 (30,37] - [22,25]  2.162316 3.059386e-02 0.1835631450
> pairwise.wilcox.test(NBA.data$per.freethrows,classage,p.adjust.method = "bonferroni")

      Pairwise comparisons using wilcoxon rank sum test

data:  NBA.data$per.freethrows and classage

      [22,25] (25,27] (27,30]
(25,27] 1.0000 - -
(27,30] 0.0002 0.0355 -
(30,37] 0.1248 0.9751 0.6804

P value adjustment method: bonferroni
```

Σχήμα 28: Συγκρίσεις ανα δύο με διόρθωση bonferroni για τις μεταβλητές classage, per.freethrows.


```

> anova6<-aov(NBA.data$time.min~classage,data=NBA.data)
> summary(anova6)
              Df    Sum Sq Mean Sq F value    Pr(>F)
classage      3 15498488 5166163   7.897 8.75e-05 ***
Residuals    101 66069588  654154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #gia kanonikotita kataloipwn
> shapiro.test(anova6$res)###DEN aporiptw ho p-value>0.05 (ORIAKA P-VALUE=0.0542)

      Shapiro-Wilk normality test

data:  anova6$res
W = 0.97607, p-value = 0.0542

> ##DEN aporiptw ipothsi kanonikotitas kataloipwn
> ###gia diakimanseis
> leveneTest(anova6)##DEN aporiptw ho
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.4184 0.7402
      101
> ###ara den aporiptw ipothesi iswn diakimansewn

```

Σχήμα 29: Ανάλυση διακύμανσης κατα έναν παράγοντα μεταξύ των μεταβλητών time.min , classage .

```

> TukeyHSD(anova6)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = NBA.data$time.min ~ classage, data = NBA.data)

$classage
              diff          lwr          upr      p adj
(25,27]-[22,25]  528.5260    -46.32996 1103.38191 0.0832139
(27,30]-[22,25] 1007.8407    460.81062 1554.87070 0.0000307
(30,37]-[22,25]  352.0260   -222.82996  926.88191 0.3834252
(27,30]-(25,27]  479.3147   -132.73906 1091.36843 0.1783314
(30,37]-(25,27] -176.5000   -813.54574  460.54574 0.8873930
(30,37]-(27,30] -655.8147  -1267.86843  -43.76094 0.0307477

> #diafora gia (27,30]-[22,25] kai (30,37]-(27,30]

```

Σχήμα 30: Πολλαπλές συγκρίσεις Tukey για να βρώ ποια επίπεδα της κατηγορικής μεταβλητής classage έχουν στατιστικά σημαντική διαφορά σε σχέση με την ποσοτική μου μεταβλητή time.min .


```

> anova4<-aov(NBA.data$per.in.area~classage,data=NBA.data)
> summary(anova4)
              Df Sum Sq Mean Sq F value Pr(>F)
classage      3   71.1    23.71   1.237    0.3
Residuals   101 1935.2    19.16
> #για κανονικότητα καταλοίπων
> shapiro.test(anova4$res)###DEN aporiptw ho p-value>0.05 (pvalue=0.3031)

      Shapiro-Wilk normality test

data:  anova4$res
W = 0.98534, p-value = 0.3031

> ##DEN aporiptw ipothsi kanonikotitas kataloipwn
> ###για διακιμανσεις
> leveneTest(anova4)##DEN aporiptw ho
Levene's Test for Homogeneity of Variance (center = median)
              Df F value Pr(>F)
group        3   1.4459 0.2339
            101
> ####ara den aporiptw ipothesi iswn diakimansewn
> #apo anova den aporiptw ho ara den uparxei diafora sthn mesos times twn ilikiakwn
> #omadwn gia to pososto kaliathiwn entos tripontou
.

```

Σχήμα 31: Ανάλυση διακύμανσης κατα έναν παράγοντα μεταξύ των μεταβλητών per.in.area. , classage .

	<i>heighth</i>	<i>num.match</i>	<i>time.min</i>	<i>age</i>	<i>points</i>	<i>assist</i>	<i>rebound</i>	<i>per.in.area</i>	<i>per.freethrows</i>
<i>heighth</i>									
<i>num.match</i>	-0.17								
<i>time.min</i>	-0.08	0.78***							
<i>age</i>	0.07	0.16	0.20*						
<i>points</i>	0.12	0.37***	0.79***	0.15					
<i>assist</i>	-0.37***	0.44***	0.71***	0.14	0.57***				
<i>rebound</i>	0.15	0.36***	0.72***	0.09	0.79***	0.60***			
<i>per.in.area</i>	-0.11	0.19*	0.33***	0.11	0.37***	0.22*	0.23*		
<i>per.freethrows</i>	-0.06	0.31**	0.38***	0.25*	0.41***	0.24*	0.22*	0.28**	

Computed correlation used pearson-method with listwise-deletion.

Σχήμα 32: Πίνακας γραμμικών συσχετίσεων κατα Pearson .

	<i>heighth</i>	<i>num.match</i>	<i>time.min</i>	<i>age</i>	<i>points</i>	<i>assist</i>	<i>rebound</i>	<i>per.in.area</i>	<i>per.freethrows</i>
<i>heighth</i>									
<i>num.match</i>	-0.18								
<i>time.min</i>	-0.05	0.78***							
<i>age</i>	0.07	0.21*	0.27**						
<i>points</i>	0.14	0.44***	0.82***	0.20*					
<i>assist</i>	-0.31**	0.50***	0.75***	0.21*	0.65***				
<i>rebound</i>	0.17	0.43***	0.78***	0.13	0.81***	0.69***			
<i>per.in.area</i>	-0.01	0.34***	0.35***	0.09	0.40***	0.23*	0.26**		
<i>per.freethrows</i>	-0.04	0.26**	0.35***	0.36***	0.37***	0.21*	0.16	0.27**	

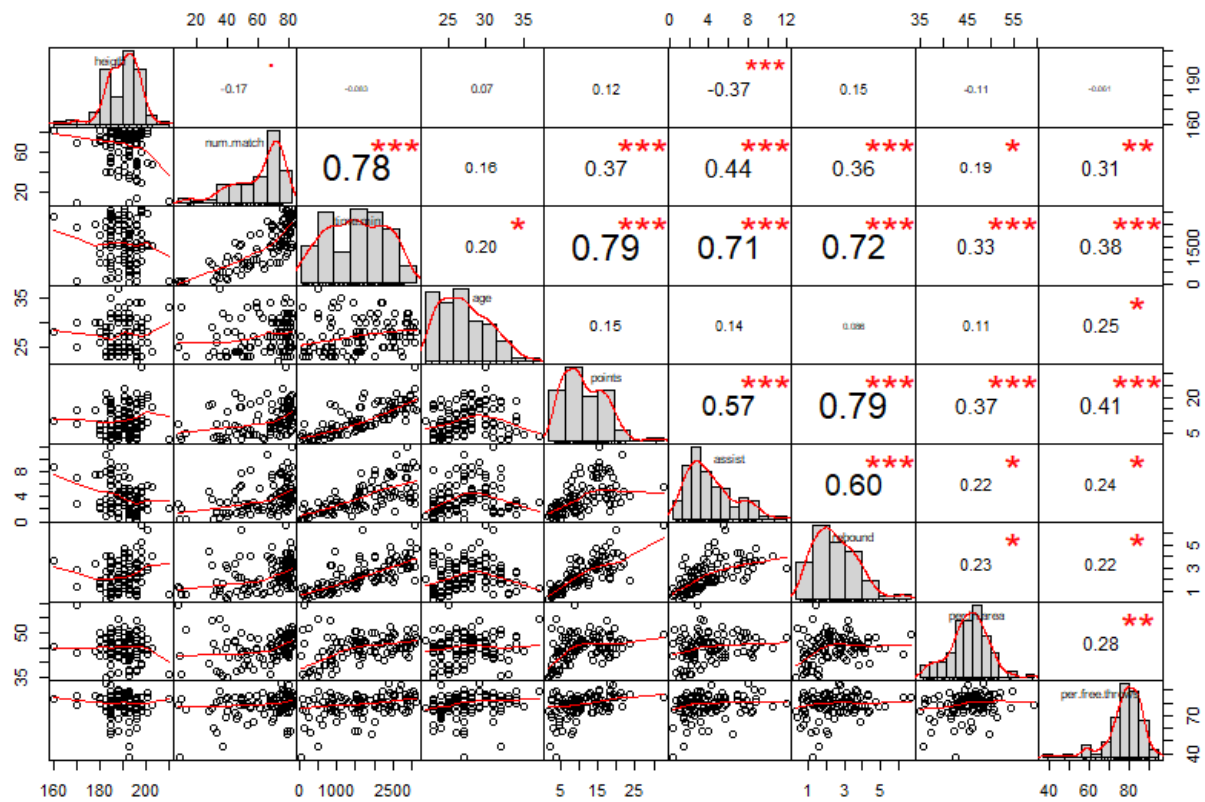
Computed correlation used spearman-method with listwise-deletion.

Σχήμα 33: Μη παραμετρικός πίνακας δεικτών συσχετίσεων κατά Spearman .

	<i>heighth</i>	<i>num.match</i>	<i>time.min</i>	<i>age</i>	<i>points</i>	<i>assist</i>	<i>rebound</i>	<i>per.in.area</i>	<i>per.freethrows</i>
<i>heighth</i>									
<i>num.match</i>	-0.13								
<i>time.min</i>	-0.03	0.60***							
<i>age</i>	0.05	0.15*	0.19**						
<i>points</i>	0.09	0.31***	0.64***	0.16*					
<i>assist</i>	-0.22**	0.35***	0.57***	0.16*	0.46***				
<i>rebound</i>	0.13	0.31***	0.60***	0.10	0.62***	0.51***			
<i>per.in.area</i>	-0.01	0.26***	0.26***	0.07	0.29***	0.16*	0.17*		
<i>per.freethrows</i>	-0.03	0.18**	0.23***	0.24***	0.26***	0.14*	0.11	0.19**	

Computed correlation used kendall-method with listwise-deletion.

Σχήμα 34: Μη παραμετρικός πίνακας δεικτών συσχετίσεων κατά Kendall .



Σχήμα 35: Διαγράμματα σημείων και πίνακας συσχετίσεων μεταξύ των μεταβλητών μας .

```

call:
lm(formula = NBA.data$time.min ~ NBA.data$height + NBA.data$num.match +
    NBA.data$age + NBA.data$points + NBA.data$assist + NBA.data$rebound +
    NBA.data$per.in.area + NBA.data$per.free.throws, data = NBA.data)

Residuals:
    Min       1Q   Median       3Q      Max
-620.46 -154.91   14.36  161.17  574.55

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1622.277    911.598   -1.780   0.0783 .
NBA.data$height     2.715     4.290    0.633   0.5284
NBA.data$num.match   23.759     1.436   16.544 < 2e-16 ***
NBA.data$age         6.219     7.357    0.845   0.4000
NBA.data$points      60.405     7.287    8.290 6.90e-13 ***
NBA.data$assist      75.273    15.269    4.930 3.44e-06 ***
NBA.data$rebound     52.507    33.338    1.575   0.1186
NBA.data$per.in.area  3.807     5.944    0.640   0.5234
NBA.data$per.free.throws -2.282     2.997   -0.761   0.4483
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 239.8 on 96 degrees of freedom
Multiple R-squared:  0.9323,    Adjusted R-squared:  0.9267
F-statistic: 165.3 on 8 and 96 DF,  p-value: < 2.2e-16

```

Σχήμα 36: Γραμμικό μοντέλο μη κεντροποιημένων μεταβλητών.

```

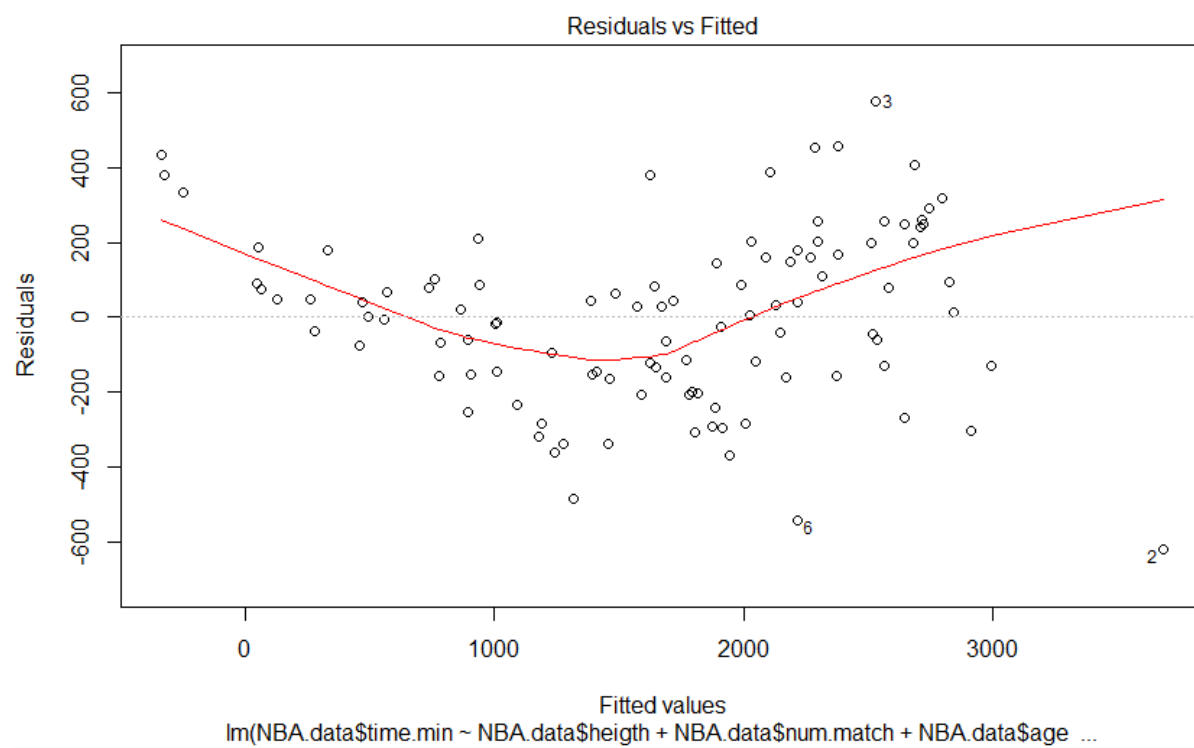
> shapiro.test(rstandard(lm1))

      shapiro-wilk normality test

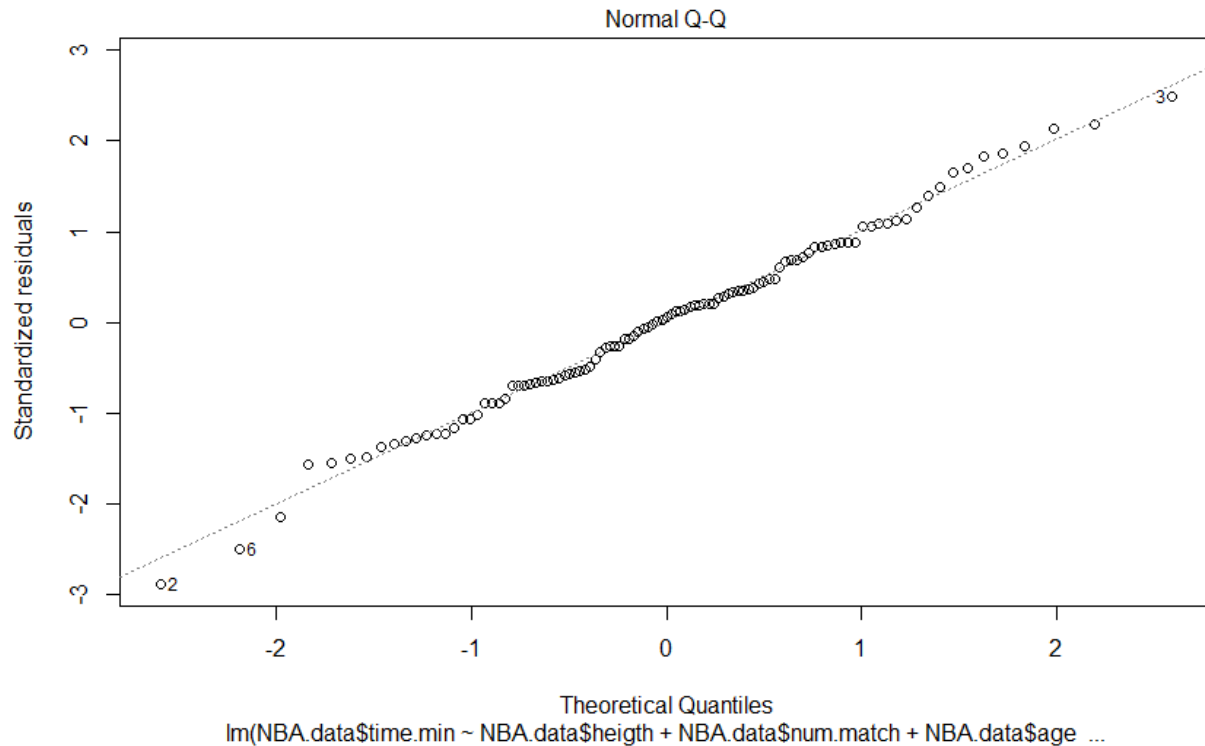
data:  rstandard(lm1)
W = 0.99368, p-value = 0.9125

```

Σχήμα 37: Ελεγχος κανονικότητας καταλοίπων μη κεντραρισμένου μοντέλου



Σχήμα 38: Διαγραμμα για έλεγχο γραμμικότητας.



Σχήμα 39: Διαγραμμα για έλεγχο κανονικότητας καταλοίπων του μη κεντραρισμένου μοντέλου.

```
Step: AIC=1153.53
NBA.data$time.min ~ NBA.data$num.match + NBA.data$points + NBA.data$assist +
  NBA.data$rebound
```

	Df	Sum of Sq	RSS	AIC
<none>			5636204	1153.5
- NBA.data\$rebound	1	209725	5845929	1155.4
- NBA.data\$assist	1	1824533	7460737	1181.0
- NBA.data\$points	1	5179376	10815580	1220.0
- NBA.data\$num.match	1	16449594	22085798	1294.9

Σχήμα 40: Επιλογή μοντέλου μη κεντραρισμένου με την διαδικασία stepwise procedution .

```
Step: AIC=1153.53
NBA.data$time.min ~ NBA.data$num.match + NBA.data$points + NBA.data$assist +
NBA.data$rebound
```

	Df	Sum of Sq	RSS	AIC
<none>			5636204	1153.5
- NBA.data\$rebound	1	209725	5845929	1155.4
- NBA.data\$assist	1	1824533	7460737	1181.0
- NBA.data\$points	1	5179376	10815580	1220.0
- NBA.data\$num.match	1	16449594	22085798	1294.9

```
Call:
lm(formula = NBA.data$time.min ~ NBA.data$num.match + NBA.data$points +
NBA.data$assist + NBA.data$rebound, data = NBA.data)
```

```
Coefficients:
      (Intercept) NBA.data$num.match NBA.data$points
        -935.60         23.61         61.09
NBA.data$assist NBA.data$rebound
         70.57         59.18
```

Σχήμα 41: Επιλογή μοντέλου μη κεντραρισμένου με την διαδικασία backward procedution .

```
Call:
lm(formula = NBA.data$time.min ~ NBA.data$num.match + NBA.data$points +
NBA.data$assist + NBA.data$rebound, data = NBA.data)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-615.20 -161.34   5.14  138.85  583.61
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -935.597     83.498  -11.205 < 2e-16 ***
NBA.data$num.match    23.610      1.382   17.084 < 2e-16 ***
NBA.data$points     61.091      6.373    9.586 8.00e-16 ***
NBA.data$assist     70.566     12.403    5.690 1.27e-07 ***
NBA.data$rebound    59.180     30.679    1.929  0.0566 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 237.4 on 100 degrees of freedom
Multiple R-squared:  0.9309,    Adjusted R-squared:  0.9281
F-statistic: 336.8 on 4 and 100 DF,  p-value: < 2.2e-16
```

Σχήμα 42: summary μοντέλου μη κεντραρισμένου με την διαδικασία stepwise procedution .

```
> vif(stp1)
NBA.data$num.match    NBA.data$points    NBA.data$assist    NBA.data$rebound
      1.280223          2.750587          1.759844          2.878740
```

Σχήμα 43: έλεγχος για πολυσυγγραμμικότητα μοντέλου μη κεντραρισμένου, καμία τιμή >10 άρα δεν υπάρχει πολυσυγγραμμικότητα.

```
##kentropoihmenes metavlites
point.centered<-NBA.data$points-mean(NBA.data$points)

rebound.centered<-NBA.data$rebound-mean(NBA.data$rebound)

assist.centered<-NBA.data$assist-mean(NBA.data$assist)

heigth.centered<-NBA.data$heigth-mean(NBA.data$heigth)

num.much.centered<-NBA.data$num.match-mean(NBA.data$num.match)

age.centered<-NBA.data$age-mean(NBA.data$age)

area.centered<-NBA.data$per.in.area-mean(NBA.data$per.in.area)

frthr.centered<-NBA.data$per.freethrows-mean(NBA.data$per.freethrows)
```

Σχήμα 44: εντολές για μετατροπή των μεταβλητών σε κεντραρισμένες.

```
Step:  AIC=1153.53
NBA.data$time.min ~ point.centered + rebound.centered + assist.centered +
num.much.centered
```

	Df	Sum of Sq	RSS	AIC
<none>			5636204	1153.5
+ age.centered	1	41984	5594219	1154.7
+ heigth.centered	1	30386	5605817	1155.0
+ frthr.centered	1	19956	5616248	1155.2
+ area.centered	1	14178	5622025	1155.3
- rebound.centered	1	209725	5845929	1155.4
- assist.centered	1	1824533	7460737	1181.0
- point.centered	1	5179376	10815580	1220.0
- num.much.centered	1	16449594	22085798	1294.9

Σχήμα 45: Επιλογή μοντέλου κεντραρισμένου με την διαδικασία stepwise procedution .


```
Step: AIC=1153.53
NBA.data$time.min ~ point.centered + rebound.centered + assist.centered +
num.much.centered
```

	Df	Sum of Sq	RSS	AIC
<none>			5636204	1153.5
- rebound.centered	1	209725	5845929	1155.4
- assist.centered	1	1824533	7460737	1181.0
- point.centered	1	5179376	10815580	1220.0
- num.much.centered	1	16449594	22085798	1294.9

```
Call:
lm(formula = NBA.data$time.min ~ point.centered + rebound.centered +
assist.centered + num.much.centered)
```

```
Coefficients:
(Intercept)      point.centered rebound.centered assist.centered
      1656.49              61.09              59.18              70.57
num.much.centered
      23.61
```

Σχήμα 46: Επιλογή μοντέλου κεντραρισμένου με την διαδικασία backward procedution .

```
Call:
lm(formula = NBA.data$time.min ~ point.centered + rebound.centered +
assist.centered + num.much.centered)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-615.20 -161.34    5.14  138.85  583.61
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1656.486    23.169   71.497 < 2e-16 ***
point.centered    61.091     6.373    9.586 8.00e-16 ***
rebound.centered  59.180    30.679    1.929  0.0566 .
assist.centered  70.566    12.403    5.690 1.27e-07 ***
num.much.centered 23.610     1.382   17.084 < 2e-16 ***
---

```

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

```
Residual standard error: 237.4 on 100 degrees of freedom
Multiple R-squared:  0.9309,    Adjusted R-squared:  0.9281
F-statistic: 336.8 on 4 and 100 DF,  p-value: < 2.2e-16
```

Σχήμα 47: summary μοντέλου κεντραρισμένου με την διαδικασία stepwise procedution .

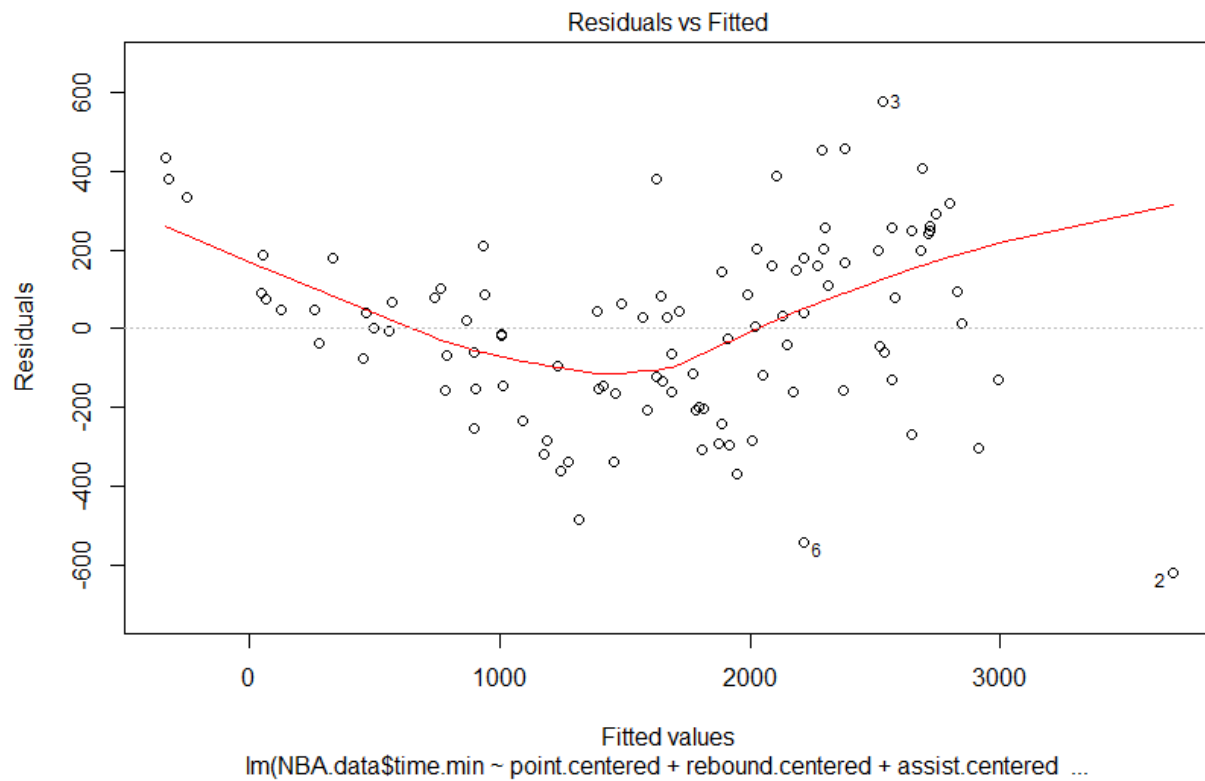
shapiro-wilk normality test

```
data: rstandard(stp2)
w = 0.99525, p-value = 0.9768
```

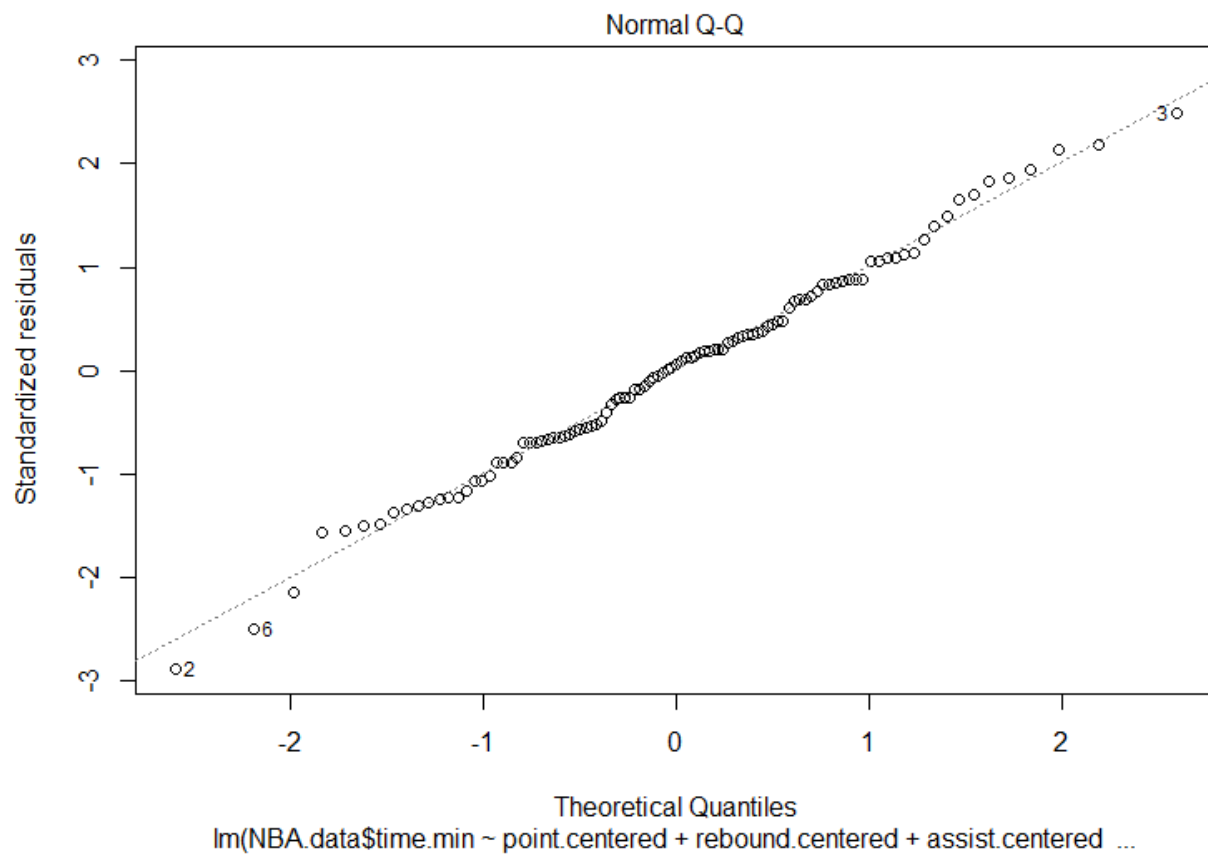
Σχήμα 48: Διαγράμμα για έλεγχο κανονικότητας καταλοίπων του κεντραρισμένου μοντέλου.

```
> vif(stp2)
      point.centered rebound.centered assist.centered num.much.centered
              2.750587              2.878740              1.759844              1.280223
```

Σχήμα 49: έλεγχος για πολυσυγγραμμικότητα μοντέλου κεντραρισμένου, καμία τιμή >10 άρα δεν υπάρχει πολυσυγγραμμικότητα.



Σχήμα 50: διάγραμμα Κατάλοιπων και fitted values για το κεντραρισμένο μοντέλο.



Σχήμα 51: διάγραμμα κανονικότητας Κατάλοιπων για το κεντραρισμένο μοντέλο.