

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Οικονομικό Πανεπιστήμιο Αθηνών

Τμήμα Στατιστικής

Εαρινό εξάμηνο

Ονοματεπώνυμο : Εφραίμ Γιαννίκος

Αριθμός μητρώου : 6210015

Μάθημα : Ανάλυση Δεδομένων

Περιεχόμενα

<u>Περιεχόμενα</u>	2
<u>Εισαγωγή</u>	3
<u>Περιγραφική ανάλυση</u>	4
<u>Σχέσεις ανα δύο</u>	7
<u>Ερμηνευτικά μοντέλα</u>	10
<u>Συμπεράσματα</u>	14
<u>Παράρτημα 1 Κώδικας στην R</u>	15

Εισαγωγή

Το μπέιζμπολ αποτελεί ένα από τα πιο δημοφιλή και παραδοσιακά αθλήματα στις Ηνωμένες Πολιτείες, με μακρά ιστορία και έντονη επαγγελματική δραστηριότητα . Το άθλημα αυτό έχει εξελιχθεί σε μία βιομηχανία δισεκατομμυρίων , όπου οι αποδόσεις των παικτών συνδέονται με τις οικονομικές τους απολαβές . Οι επιδόσεις αυτές αξιολογούνται και έπειτα χρησιμοποιούνται τόσο για την στρατηγική των ομάδων όσο και για τις συμβάσεις των παικτών . Οι καλύτεροι παίκτες του μπέιζμπολ λαμβάνουν υψηλούς μισθούς, γεγονός που αντικατοπτρίζει τη σημασία της ατομικής επίδοσης στον επαγγελματικό αθλητισμό . Ο σκοπός της μελέτης είναι να απαντήσει στο ερώτημα «γιατί οι αθλητές του μπέιζμπολ κερδίζουν όσα κερδίζουν».

Πριν προχωρήσουμε στην ανάλυση , ας εξηγήσουμε κάποιους βασικούς κανόνες του μπέιζμπολ :

1. Στόχος του παιχνιδιού

Η ομάδα που σκοράρει τους περισσότερους πόντους κερδίζει. Πόντος σημειώνεται όταν ένας παίκτης της επίθεσης καταφέρει να κάνει τον γύρο των τεσσάρων βάσεων και να επιστρέψει στην αρχική .

2. Οι δύο ομάδες

- Το παιχνίδι παίζεται σε **2 φάσεις**:
 - **Επίθεση**: μία ομάδα προσπαθεί να χτυπήσει την μπάλα και να τρέξει στις βάσεις.
 - **Άμυνα** : η άλλη προσπαθεί να σταματήσει τους επιθετικούς και να κάνει outs.

3. Πως σκοράρει μια ομάδα

- Ένας παίκτης πρέπει να περάσει και τις 4 βάσεις.
- Αν τα καταφέρει, σημειώνεται 1 πόντος

4. Διάρκεια παιχνιδιού

- Το παιχνίδι έχει **9** ημί-περιόδους.
- Σε κάθε ημι-περίοδο, κάθε ομάδα παίζει **μία φορά επίθεση και μία άμυνα**.
- Αν μετά και από τις 9 το σκορ είναι ισόπαλο, πάμε σε **παράταση** .

Περιγραφική ανάλυση

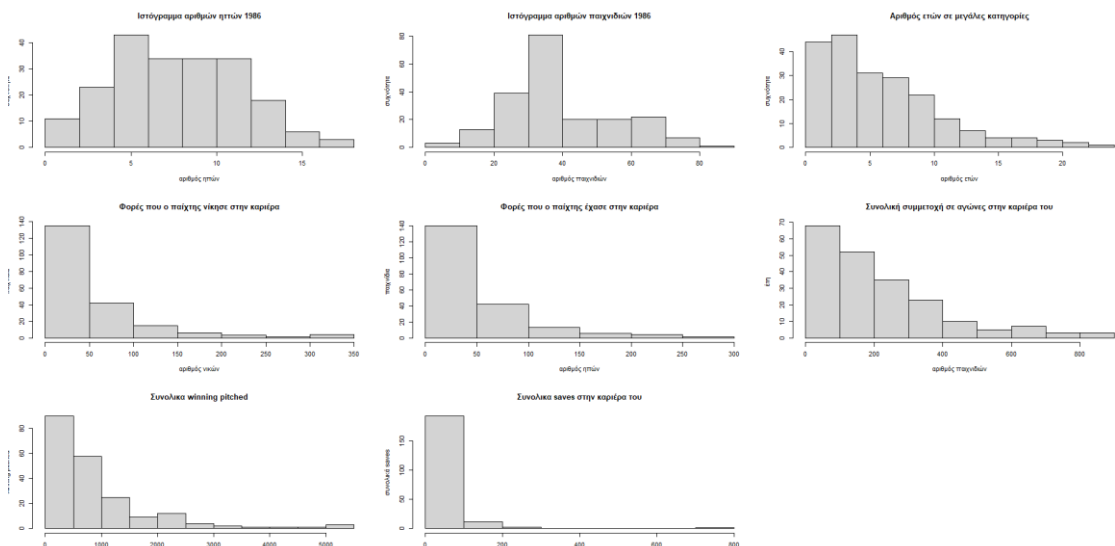
name	Όνομα παίκτη
surname	Επίθετο παίκτη
wins86	Φορές που ο παίκτης νίκησε το 1986
losses86	Φορές που ο παίκτης έχασε το 1986
games86	Αριθμός αγώνων στους οποίους συμμετείχε το 1986
leagues	Αριθμός ετών σε μεγάλες κατηγορίες
wins	Φορές που ο παίκτης νίκησε στην καριέρα του
losses	Φορές που ο παίκτης έχασε στην καριέρα του
games	Αριθμός αγώνων στους οποίους συμμετείχε στην καριέρα του
innings	Φορές innings pitched στην καριέρα του
saves	Φορές saves στην καριέρα του
salary87	Μισθός κατά την έναρξη της αγωνιστικής περιόδου το 1987 (χιλιάδες δολάρια)
league87	Κατηγορία το 1987 (1: Εθνική, 2: Αμερικάνικη)

Τα παραπάνω δεδομένα συλλέχθηκαν από τις κυριότερες κατηγορίες μπέιζμπολ των ΗΠΑ κατά την διάρκεια της αγωνιστικής περιόδου του 1987 . Το δείγμα απαρτίζεται από 206 παίκτες και διαθέτει στο σύνολο 12 μεταβλητές . Αρχικά γίνεται εισαγωγή των δεδομένων στην R με την επέκταση του αρχείου να είναι .sav . Οι μεταβλητές χωρίζονται σε κατηγορικές και ποσοτικές . Παρακάτω δίνονται αναλυτικά όλες οι μεταβλητές . Κατόπιν ελέγχου οι πρώτες 2 μεταβλητές ήταν καταχωρημένες ως “character” (οι οποίες είναι μεταβλητές ιχνηλασιμότητας), ενώ όλες οι υπόλοιπες είχαν καταχωρηθεί ως ποσοτικές (numeric) .Όμως έγινε αλλαγή στην τελευταία (league87) η οποία είναι **κατηγορική** με δύο επίπεδα (1: Εθνική, 2: Αμερικάνικη) .

Αρχικά εξετάζουμε τα περιγραφικά μέτρα των ποσοτικών μεταβλητών όπως : μέση τιμή και διάμεσος, μέτρα διασποράς (τυπική Απόκλιση), μέτρα Σχετικής θέσης (Q1, Q3) . Για την αγωνιστική περίοδο 1986 η μέση τιμή νικών ενός παίκτη είναι 8.746 (με τυπική απόκλιση = 4.54) , μέση τιμή ηττών είναι 8.262 (T.A = 3.72) , μέσες συμμετοχές σε αγώνες 39.53 (T.A =15.71) και μέσος μισθός στην έναρξη της περιόδου 494.7 χιλιάδες δολάρια (T.A = 372.79). Έπειτα εξετάζουμε στοιχεία του κάθε παίχτη καθόλη την διάρκεια της καριέρας του : μέση συμμετοχή σε μεγάλες κατηγορίες σε έτη 6.38 (T.A = 4.688) , μέσες νίκες 53.4 (T.A =58.94) , μέσες ήττες 47.64 (T.A =47.77) ,μέση συμμετοχή σε αγώνες 212.63 (T.A = 181.65) , μέσα winning pitched 883.9 (T.A =953.43) και μέση τιμή saves 24.67 (T.A = 68.41

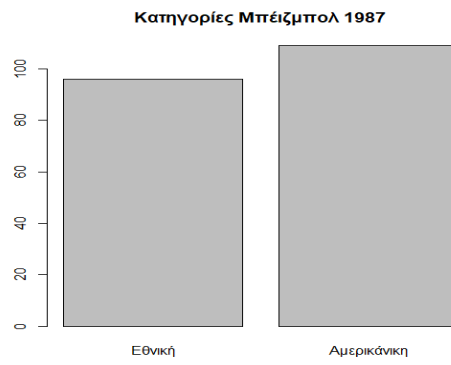
Μεταβλητή	Μέση τιμή	Min	Max	Τυπική απόκλιση	Διάμεσος	Q1	Q3	Κύρτωση	Ασυμμετρία
WINS86	8.746	0	24	4.54	9	5	11	3.09	0.53
LOSSES86	8.26	0	18	3.72	8	5	11	2.39	0.18
GAMES86	39.53	1	83	15.71	35	30	50	2.89	0.55
LEAGUES	6.38	0	23	4.688	6.388	3	9	4.31	1.21
WINS	53.4	1	323	58.94	35	15.25	64.5	9.60	2.36
LOSSES	47.64	1	261	47.77	31.5	14	64.75	7.60	2.00
GAMES	212.63	3.86	853	181.65	154.5	80	298.75	4.70	1.39
WINNINGS	883.9	19	5264	953.43	542.5	260.9	1174.2	9.34	2.32
SAVES	24.67	0	791.1	68.41	3	0	16.75	79.13	7.55
SALARY87	494.7	0	1850	372.79	415	155	750	4.20	1.00

Πίνακας 1 : Περιγραφικά μέτρα



Διάγραμμα 1 : Ιστογράμματα για τις ποσοτικές μεταβλητές

Παρακάτω δίνονται και πληροφορίες σχετικά με την συμμετοχή αθλητών τόσο σε αμερικανικές αγωνιστικές όσο και σε εθνικές .



Διάγραμμα 2 : Ραβδόγραμμα για τις κατηγορίες μπέιζμπολ.

Σχέσεις ανα δύο

Στην συνέχεια εξετάζουμε αν οι μέσες τιμές της κατηγορίας που συμμετάσχει ο κάθε παίχτης (Εθνική ή Αμερικάνικη) με τις υπόλοιπες μεταβλητές παρουσιάζουν στατιστικά σημαντική διαφορά . Για παράδειγμα , αν οι μέσες απολαβές ενός παίχτη της αμερικάνικης κατηγορίας διαφέρει σημαντικά σε σχέση με αυτές ενός παίκτη της εθνικής. Απαραίτητη προϋπόθεση είναι τα δεδομένα κάθε μεταβλητής να κατανέμεται κανονικά σε κάθε κατηγορία .Επομένως πραγματοποιούμε έλεγχο Shapiro-Wilk test . Ο έλεγχος της R μας παρουσίασε p-value = 1.269e-05 για την εθνική και 3.153e-06 για αμερικάνικη , τα οποία <5% άρα δεν ακολουθείται κανονική κατανομή. Αφού δεν ακολουθείται η κανονική κατανομή , θα εξετάσουμε αν μπορεί να προσεγγιστεί . Απαραίτητη προϋπόθεση είναι το δείγμα μας να είναι >50 για κάθε επίπεδο . Το δείγμα μας είναι συνολικά 205, άρα το δείγμα είναι επαρκές. Για να μπορέσουμε να εξετάσουμε αν προσεγγίζεται η κανονική κατανομή προχωρούμε σε έλεγχο συμμετρίας , όπου πρέπει τα δεδομένα να είναι συμμετρικά σε κάθε ομάδα .Ο έλεγχος ονομάζεται Agostino test και τα αποτελέσματα του p-value = 0.0004 (εθνική) και 0.0003 (αμερικάνικη), επομένως δεν υπάρχει συμμετρία και δεν προσεγγίζεται η κανονικότητα και οπότε θα πραγματοποιηθεί μη παραμετρικό έλεγχο Wilcoxon Test . Ο έλεγχος παρουσιάζει p-value = 0.2843 >5% , άρα οι διάμεσοι των μισθών κάθε κατηγορίας δεν απέχουν στατιστικά σημαντικά . Ανάλογη διαδικασία θα ακολουθηθεί και για τις υπόλοιπες ποσοτικές μεταβλητές. Πραγματοποιώντας Shapiro-Wilk Test σε όλες τις ποσοτικές μεταβλητές , παρατηρούμε τα εξής p-value :

Όνομα μεταβλητής	P-value εθνικής	P-value Αμερικάνικης	Outcome
WINS86	0.01216	0.0162	Δεν ακολουθείται κανονική κατανομή για καμία μεταβλητή
LOSSES86	0.02429	0.0429	
GAMES86	9.909e-05	0.0005371	
LEAGUES	1.258e-05	2.263e-08	
WINS	3.019e-09	2.585e-13	
LOSSES	6.88e-09	9.227e-12	
GAMES	3.36e-08	1.102e-08	
INNINGS	3.153e-09	3.59e-13	
SAVES	7.491e-16	2.412e-16	

Πίνακας 2 : Αποτελέσματα Shapiro – Wilk

Επομένως θα πρέπει να ελεγχθεί αν προσεγγίζεται η κανονική κατανομή για κάθε μεταβλητή . Ο έλεγχος Agostino με τα αποτελέσματα της συμμετρίας παρουσιάζονται παρακάτω .

Όνομα μεταβλητής	P-value εθνικής	P-value Αμερικάνικης	Συμμετρική
WINS86	0.0258	0.01586	OXI

LOSSES86	0.07523	0.8589	NAI
GAMES86	0.01271	0.02687	OXI
LEAGUES	0.0003379	1.963e-06	OXI
WINS	1.232e-08	2.353e-11	OXI
LOSSES	3.569e-08	8.332e-10	OXI
GAMES	2.916e-07	3.922e-06	OXI
INNINGS	1.68e-08	2.807e-11	OXI
SAVES	8.131e-14	9.474e-15	OXI

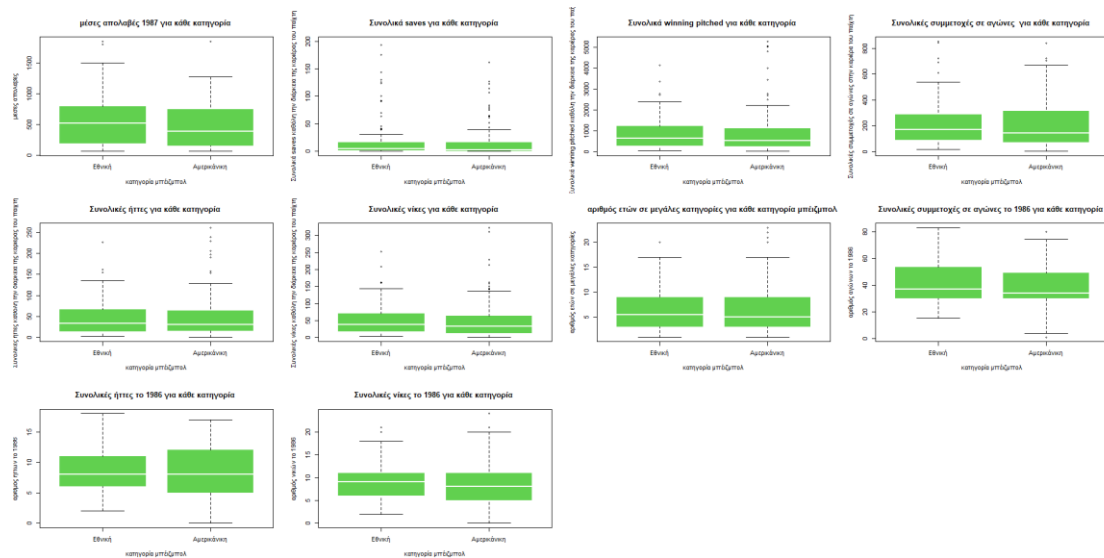
Πίνακας 7 : Αποτελέσματα Agostino Test συμμετρίας

Παρατηρούμε πως όλες οι μεταβλητές δεν είναι συμμετρικές με εξαίρεση την LOSSES86 . Για να μπορεί να προσεγγιστεί η κανονικότητα θα πρέπει η μεταβλητή να είναι και μεσόκυρτη . Με έλεγχο Bonett test $p\text{-value} = 0.018$ (εθνική) και 0.0001 (αμερικάνικη) $<5\%$, άρα δεν είναι μεσόκυρτη .Επομένως δεν προσεγγίζεται η κανονικότητα . Άρα θα κάνουμε έλεγχο Wilcoxon για το σύνολο των μεταβλητών

Όνομα μεταβλητής	P-value	Διαφέρουν στατιστικά σημαντικά οι διάμεσοι
WINS86	0.3544	OXI
LOSSES86	0.8591	OXI
GAMES86	0.3213	OXI
LEAGUES	0.8018	OXI
WINS	0.5434	OXI
LOSSES	0.9633	OXI
GAMES	0.4864	OXI
INNINGS	0.7555	OXI
SAVES	0.2434	OXI

Πίνακας 3 : έλεγχος Wilcoxon

Από τον παραπάνω πίνακα προκύπτει ότι δεσ υπάρχουν σημαντικές διαφορές στις διαμέσους των μεταβλητών μεταξύ των δύο ομάδων. Παρακάτω δίνεται η οπτικοποίηση των δεδομένων μας μέσω boxplot για το σύνολο των μεταβλητών. Έτσι γίνονται εμφανή η ασυμμετρία των δεδομένων, η διασπορά των τιμών, η διάμεσος και το κέντρο, καθώς και οι ακραίες τιμές.



Διάγραμμα 2 : Boxplot κάθε ποσοτικής μεταβλητής για κάθε κατηγορία baseball

Έπειτα εξετάζονται οι σχέσεις των ποσοτικών μεταβλητών μεταξύ τους για φαινόμενα συσχέτισης. Προκειμένου να πραγματοποιηθεί έλεγχος, πρέπει να ελεγχθεί αν οι μεταβλητές κατανέμονται κανονικά. Όπως είδαμε προηγουμένως, αυτό δεν ισχύει, επομένως γίνεται μη-παραμετρικός έλεγχος **Spearman**. Ο πίνακας παρουσιάζει τις συσχετίσεις μεταξύ μεταβλητών που αφορούν την απόδοση παικτών baseball και τους μισθούς τους, με τη χρήση του συντελεστή Spearman. Παρατηρείται ότι οι μεταβλητές που σχετίζονται με τη συμμετοχή και την εμπειρία των παικτών (όπως οι GAMES, INNINGS και LEAGUES) παρουσιάζουν ισχυρές θετικές συσχετίσεις μεταξύ τους. Συγκεκριμένα, ο αριθμός των αγώνων (GAMES) σχετίζεται θετικά τόσο με τα innings ($\rho = 0.974$) όσο και με τη συμμετοχή σε πρωταθλήματα ($\rho = 0.934$), γεγονός που υποδεικνύει ότι οι πιο ενεργοί παίκτες συμμετέχουν ευρύτερα στο πρωτάθλημα και αγωνίζονται περισσότερο. Παράλληλα, οι συνολικές επιδόσεις (WINS και LOSSES) συνδέονται ισχυρά με τα κέρδη (WINNINGS), με συσχετίσεις της τάξης του 0.881 και 0.951 αντίστοιχα. Αυτό δείχνει ότι οι παίκτες που εμπλέκονται σε περισσότερα παιχνίδια και έχουν πιο έντονη παρουσία, ανεξαρτήτως αν πρόκειται για νίκες ή ήττες, συγκεντρώνουν περισσότερα χρηματικά κέρδη. Ο μισθός του 1987 (SALARY87) σχετίζεται επίσης θετικά με τις περισσότερες μεταβλητές απόδοσης, με ισχυρότερη συσχέτιση με τα συνολικά κέρδη ($\rho = 0.800$), τα innings ($\rho = 0.800$), τις συμμετοχές σε πρωταθλήματα ($\rho = 0.739$) και τα games ($\rho = 0.738$). Αυτό υποδηλώνει ότι ο μισθός επηρεάζεται κυρίως από τον όγκο συμμετοχής και την εμπειρία, περισσότερο από συγκεκριμένες επιδόσεις όπως οι νίκες σε μια χρονιά. Ιδιαίτερη σημασία έχει η αρνητική συσχέτιση των saves με τις νίκες ($\rho = -0.441$), τις ήττες και τα games του 1986, κάτι που αντικατοπτρίζει τον διαφορετικό ρόλο των ανακουφιστικών pitchers (closers), οι οποίοι δεν αγωνίζονται συστηματικά αλλά συμμετέχουν σε ειδικές περιστάσεις. Αυτό εξηγεί και την πιο αδύναμη σχέση που παρουσιάζουν οι saves με τον μισθό ($\rho = 0.247$).

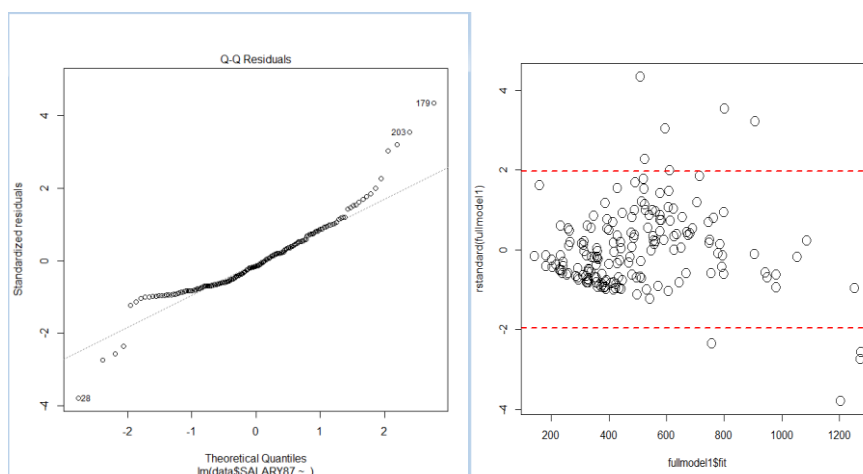
	WINS86	LOSSES86	GAMES86	LEAGUES	WINS	LOSSES	GAMES	INNINGS	SAVES	SALARY87
WINS86		0.439 (.001)	-0.172 (.023)	0.020 (.790)	0.308 (.001)	0.192 (.011)	-0.066 (.382)	0.257 (.001)	-0.441 (.001)	0.212 (.005)
LOSSES86	0.439 (.001)		-0.140 (.064)	0.139 (.066)	0.340 (.001)	0.395 (.001)	0.067 (.377)	0.358 (.001)	-0.355 (.001)	0.204 (.007)
GAMES86	-0.172 (.023)	-0.140 (.064)		0.099 (.193)	-0.046 (.548)	-0.014 (.849)	0.319 (.001)	-0.058 (.444)	0.649 (.001)	0.083 (.272)
LEAGUES	0.020 (.790)	0.139 (.066)	0.099 (.193)		0.881 (.001)	0.907 (.001)	0.934 (.001)	0.898 (.001)	0.409 (.001)	0.739 (.001)
WINS	0.308 (.001)	0.340 (.001)	-0.046 (.548)	0.881 (.001)		0.951 (.001)	0.824 (.001)	0.985 (.001)	0.162 (.032)	0.813 (.001)
LOSSES	0.192 (.011)	0.395 (.001)	-0.014 (.849)	0.907 (.001)	0.951 (.001)		0.844 (.001)	0.974 (.001)	0.196 (.009)	0.764 (.001)
GAMES	-0.066 (.382)	0.067 (.377)	0.319 (.001)	0.934 (.001)	0.824 (.001)	0.844 (.001)		0.825 (.001)	0.609 (.001)	0.738 (.001)
INNINGS	0.257 (.001)	0.358 (.001)	-0.058 (.444)	0.898 (.001)	0.985 (.001)	0.974 (.001)	0.825 (.001)		0.143 (.059)	0.800 (.001)
SAVES	-0.441 (.001)	-0.355 (.001)	0.649 (.001)	0.409 (.001)	0.162 (.032)	0.196 (.009)	0.609 (.001)	0.143 (.059)		0.247 (.001)
SALARY87	0.212 (.005)	0.204 (.007)	0.083 (.272)	0.739 (.001)	0.813 (.001)	0.764 (.001)	0.738 (.001)	0.800 (.001)	0.247 (.001)	

Computed correlation used spearman-method with listwise-deletion.

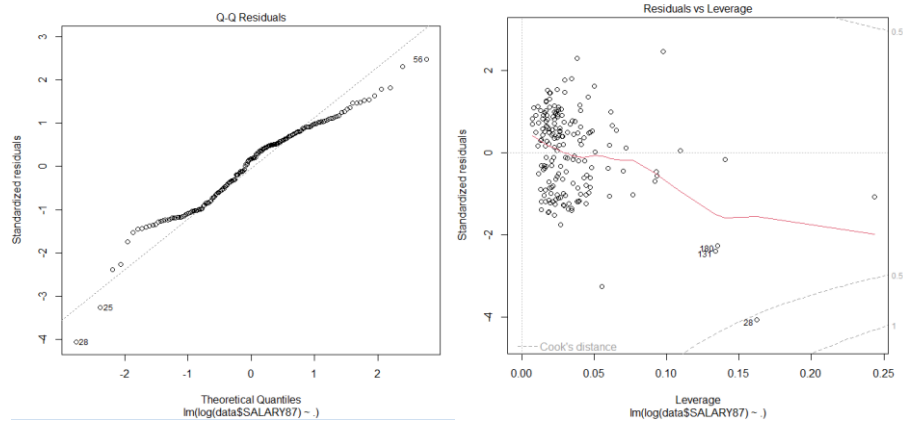
Πίνακας : Σχέσεις συσχέτισης με τα αντίστοιχα p-value

Ερμηνευτικά μοντέλα

Σε αυτό το κεφάλαιο επιχειρείται η ανάπτυξη ενός μοντέλου , το οποίο θα θα εξετάζει την σχέση ανάμεσα στις αποδοχές που λαμβάνει ένας παίκτης με τις υπόλοιπες μεταβλητές . Η αρχική εκτίμηση του μοντέλου γίνεται με ανεξάρτητες μεταβλητές όλες τις ποσοτικές δηλαδή χωρίς το ονοματεπώνυμο του παίκτη και την κατηγορία που συμμετέχει. Το μοντέλο παρουσιάζει πρόβλημα πολυσυγγραμμικότητας, επομένως πραγματοποιείται αφαίρεση κάποιων μεταβλητών. Το νέο μοντέλο διαθέτει τις μεταβλητές wins86, losses86, games86, wins και saves. Λύνεται το πρόβλημα της πολυσυγγραμμικότητας, παραβιάζεται η κανονικότητα και η ομοσκεδαστικότητα . Πραγματοποιείται λογαριθμικός μετασχηματισμός της μεταβλητής salary87 και εκτελείται εκ νέου το μοντέλο. Το πρόβλημα της κανονικότητας και της ομοσκεδαστικότητας λύνεται, όχι όμως της γραμμικότητας

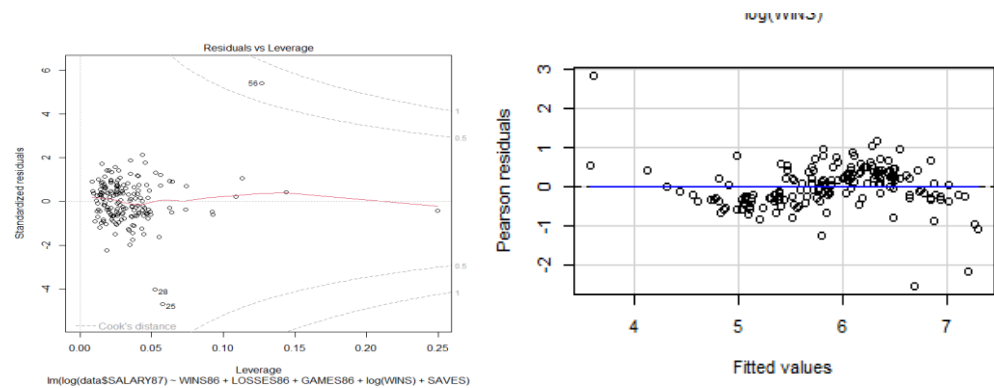


Διάγραμμα 3 : Q-Q plot κανονικότητας και διάγραμμα ομοσκεδαστικότητας



Διάγραμμα 4 : Q-Q plot κανονικότητας για το λογαριθμισμένο μοντέλο και διάγραμμα σημείων επιρροής .

Λογαριθμίζοντας και την μεταβλητή WINS και τρέχοντας εκ νέου το μοντέλο διορθώνεται το πρόβλημα της γραμμικότητας .



Απο το Q-Q plot παρατηρούμε τρεις ακραίες τιμές που αντιστοιχούν στις παρατηρήσεις 25,28,56 όπου η τιμή 56 είναι και σημείο επιρροής . Αρχικά γίνεται αφαίρεση του σημείου μόχλευσης και έπειτα και οι 2 ακραίες τιμές . Σε κάθε περίπτωση το μοντέλο διαφοροποιείται .

Το τελικό μοντέλο είναι της μορφής :

$$\log(\widehat{SALARY87}) = 3.119 - 0.0032 * WINS86 - 0.010 * LOSSES86 + 0.004 * GAMES86 + 0.7702 * \log(WINS) + 0.001 * SAVES + \varepsilon, \varepsilon \sim N(0, 0.4275^2)$$

	Estimate	Std Error	P-value
Intercept	3.1194714	0.1663091	<2e-16 ***
WINS86	-0.0032806	0.0086781	0.7059
LOSSES86	-0.0101798	0.0101381	0.3168
GAMES86	0.0043046	0.0024372	0.0792
log(WINS)	0.7702747	0.0393148	<2e-16 ***
SAVES	0.0018410	0.0009572	0.0562
R² = 77,48%		F = 114.2	

Adjusted R²= 76,8% p-value: < 2.2e-16

***<0,001 **<0,01 *<0,05

Το μοντέλο που προκύπτει είναι στατιστικά σημαντικό ($F=114.2$, $p\text{-value}<2.2e-16$) και μπορεί να ερμηνεύσει το 76,8% της μεταβλητότητας των αποδοχών ενός παίκτη . Στατιστικά σημαντικές μεταβλητές είναι οι WINS86 , LOSSES86 , GAMES86, SAVES . Αυτό σημαίνει πως οι απολαβές που θα λάβει ένας παίκτης εξαρτάται από τι συνολικές συμμετοχές που έλαβε το 1986 , τις νίκες καθώς και ήττες της ίδιας χρονιάς και τέλος τις συνολικές νίκες και saves σε όλη την καριέρα του.

Ο εκτιμώμενος γραμμικός παλινδρομικός τύπος εξηγεί τη λογαριθμική τιμή του μισθού των παικτών το 1987 ($\log(\text{SALARY87})$) βάσει επιδόσεων από το 1986 αλλά και μακροχρόνιων στατιστικών. Ο σταθερός όρος της εξίσωσης είναι 3.119, γεγονός που σημαίνει ότι όταν όλες οι ανεξάρτητες μεταβλητές είναι μηδέν, ο αναμενόμενος λογάριθμος του μισθού είναι 3.119. Άρα το SALARY θα είναι $\exp(3.119)$.

Η μεταβλητή WINS86 έχει συντελεστή -0.0032, κάτι που σημαίνει ότι κάθε επιπλέον νίκη το 1986 σχετίζεται με ελαφρώς χαμηλότερο λογάριθμο μισθού, υπό σταθερότητα των υπολοίπων παραγόντων. Αυτή η αρνητική τιμή είναι μη αναμενόμενη και ενδέχεται να δείχνει ότι οι επιτυχίες μιας μόνο χρονιάς δεν εκτιμήθηκαν θετικά από τις ομάδες όταν ελέγχεται για άλλες παραμέτρους. Παρόμοια, κάθε επιπλέον ήττα το 1986 (LOSSES86) συνδέεται με μείωση του μισθού κατά $\exp(0.010)-1)*100\%=1\%$ κάτι που είναι αναμενόμενο καθώς οι περισσότερες ήττες αποτελούν δείγμα χαμηλότερης απόδοσης. Αντίθετα, η συμμετοχή σε περισσότερα παιχνίδια το 1986 (GAMES86) σχετίζεται θετικά με τον μισθό, με κάθε επιπλέον παιχνίδι να αυξάνει τον μισθό κατά $\exp(0.004)-1)*100\%=0.4\%$. Αυτό υποδηλώνει ότι η γενική παρουσία και διαθεσιμότητα του παίκτη αναγνωρίζεται και αμείβεται. Ιδιαίτερη σημασία έχει η μεταβλητή $\log(\text{WINS})$, η οποία έχει ισχυρά θετική επίδραση (συντελεστής 0.7702). Οι επιπλέον νίκες αυξάνουν τον μισθό κατά 0,7702%.

Τέλος, η μεταβλητή SAVES έχει μικρή αλλά θετική επίδραση (0.001), γεγονός που δείχνει ότι οι παίκτες που εξειδικεύονται στο να "κλείνουν" αγώνες (closers) ανταμείβονται ελαφρώς περισσότερο, χωρίς όμως αυτός ο ρόλος να είναι καθοριστικός παράγοντας στον καθορισμό του μισθού. Συνολικά, το μοντέλο δείχνει ότι ο μισθός των παικτών για το 1987 επηρεάζεται θετικά από τη συμμετοχή, τη συνολική εμπειρία και την ιστορική απόδοση, ενώ οι επιδόσεις του 1986 έχουν μικρότερη ή και αρνητική συνεισφορά όταν ελέγχονται οι υπόλοιποι παράγοντες.

	Μοντέλο με ακραίες τιμές και μόχλευση						Μοντέλο χωρίς μόχλευση					Μοντέλο με αφαίρεση των 2 ακραίων τιμών					Μοντέλο χωρίς ακραίες τιμές							
Predictors	Estimates	std. Error	CI	Statistic	p	df	Estimates	std. Error	CI	Statistic	p	df	Estimates	std. Error	CI	Statistic	p	df	Estimates	std. Error	CI	Statistic	p	df
(Intercept)	3.60	0.20	3.20–4.00	17.82	<0.001	170.00	3.19	0.20	2.80–3.57	16.13	<0.001	169.00	3.16	0.17	2.83–3.50	18.59	<0.001	167.00	3.12	0.17	2.79–3.45	18.76	<0.001	166.00
WINS86	0.00	0.01	-0.02–0.03	0.40	0.692	170.00	0.01	0.01	-0.01–0.03	0.70	0.486	169.00	-0.00	0.01	-0.02–0.02	-0.23	0.822	167.00	-0.00	0.01	-0.02–0.01	-0.38	0.706	166.00
LOSSES86	-0.02	0.01	-0.05–0.01	-1.42	0.156	170.00	-0.01	0.01	-0.04–0.01	-1.05	0.296	169.00	-0.01	0.01	-0.03–0.01	-0.97	0.332	167.00	-0.01	0.01	-0.03–0.01	-1.00	0.317	166.00
GAMES86	0.00	0.00	-0.00–0.01	0.70	0.485	170.00	0.01	0.00	0.00–0.01	2.11	0.036	169.00	0.00	0.00	-0.00–0.01	1.77	0.079	167.00	0.00	0.00	-0.00–0.01	1.77	0.079	166.00
WINS [log]	0.65	0.05	0.55–0.75	13.21	<0.001	170.00	0.70	0.05	0.61–0.79	15.31	<0.001	169.00	0.75	0.04	0.67–0.83	18.83	<0.001	167.00	0.77	0.04	0.69–0.85	19.59	<0.001	166.00
SAVES	0.00	0.00	-0.00–0.00	1.53	0.128	170.00	0.00	0.00	-0.00–0.00	1.29	0.199	169.00	0.00	0.00	-0.00–0.00	1.95	0.053	167.00	0.00	0.00	-0.00–0.00	1.92	0.056	166.00
Observations	176						175					173					172							
R ² / R ² adjusted	0.614 / 0.602						0.679 / 0.670					0.761 / 0.754					0.775 / 0.768							

Παραπάνω δίνεται και ένας πίνακας σύγκριση όλων των μοντέλων.

Συμπεράσματα

Η μελέτη ανέδειξε ότι οι απολαβές των παικτών μπέιζμπολ εξαρτώνται κυρίως από τη συνολική τους εμπειρία και απόδοση, παρά από τις επιδόσεις μιας μόνο χρονιάς. Μεταβλητές όπως τα $\log(\text{WINS})$, οι συμμετοχές σε αγώνες (GAMES86) και τα saves έχουν θετική επίδραση στους μισθούς, ενώ οι ήττες (LOSSES86) επηρεάζουν αρνητικά. Αντίθετα, οι νίκες του 1986 (WINS86) δεν είναι στατιστικά σημαντικές και ενδέχεται να έχουν αρνητική επίδραση, γεγονός που υποδηλώνει ότι μια καλή σεζόν από μόνη της δεν εξασφαλίζει αυξήσεις. Η λογαριθμική μετατροπή των μεταβλητών βελτίωσε σημαντικά τη συμπεριφορά του μοντέλου, διορθώνοντας προβλήματα γραμμικότητας, κανονικότητας και ομοσκεδαστικότητας. Το τελικό μοντέλο εξηγεί το 76,8% της διακύμανσης στους μισθούς ($R^2 = 0.768$), γεγονός που το καθιστά στατιστικά αξιόπιστο. Συνολικά, επιβεβαιώνεται ότι το ύψος του μισθού αντανακλά όχι μόνο την πρόσφατη απόδοση αλλά και την καθολική καριέρα και τη συνέπεια του παίκτη.

Όπου **Estimates** : Οι συντελεστές παλινδρόμησης , **std.Error** : Το τυπικό σφάλμα , **CI** : το 95% διαστήματος εμπιστοσύνης για τον συντελεστή , **Statistic** : το αποτέλεσμα του στατιστικού ελέγχου , P-value κάθε μεταβλητής , οι συνολικές παρατηρήσεις (Observations) και τα αντίστοιχα R^2 , R^2 ADJUSTED .

Παράρτημα 1 Κώδικας στην R

```
install.packages("foreign")
```

```
library(foreign)
```

```
data<- read.spss(file.choose(), to.data.frame=TRUE ) #ανοιγουμε το αρχειο με τα δεδομενα
```

```
head(data) #μια συντομη παρατηρηση των δεδομενων
```

```
str(data) #εξεταζουμε την δομη των δεδομενων
```

```
data$LEAGUE87<-as.factor(data$LEAGUE87) #απο numeric σε factor την τελευταια μεταβλητη
```

```
levels(data$LEAGUE87)<-c( "Εθνική" , "Αμερικάνικη")
```

```
summary(data)
```

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

```
library(psych)
```

```
describe(data[,3:12])
```

```
#kurtosis and skewness
```

```
install.packages("moments")
```

```
library(moments)
```

```
skewness(data[3:12],na.rm=TRUE)
```

```
kurtosis(data[3:12],na.rm=TRUE)
```

```
hist(data$WINS86 ,
```

```
main = "Ιστόγραμμα αριθμών νικών 1986" ,
```

```
xlab = "αριθμός νικών" ,
```

```
ylab = "συχνότητα")
```

```
par(mfrow=c(3,3))
```

```
hist(data$LOSSES86 ,
```

```
main = "Ιστόγραμμα αριθμών ηττών 1986" ,
```

```
xlab = "αριθμός ηττών" ,
```

```
ylab = "συχνότητα")
```

```
hist(data$GAMES86 ,
```

```
main = "Ιστόγραμμα αριθμών παιχνιδιών 1986" ,
```

```
xlab = "αριθμός παιχνιδιών" ,
```

```
ylab = "συχνότητα")
```

```
hist(data$LEAGUES ,
```

```
main = "Αριθμός ετών σε μεγάλες κατηγορίες" ,
```

```
xlab = "αριθμός ετών" ,
```

```
ylab = "συχνότητα")
```

```
hist(data$WINS ,
```

```
main = "Φορές που ο παίχτης νίκησε στην καριέρα" ,
```

```
xlab = "αριθμός νικών" ,
```

```
ylab = "παιχνίδια")
```

```
hist(data$LOSSES,
```

```
main = "Φορές που ο παίχτης έχασε στην καριέρα" ,
```

```
xlab = "αριθμός ηττών" ,
```

```
ylab = "παιχνίδια")
```

```
hist(data$GAMES,
```



```
main = "Συνολική συμμετοχή σε αγώνες στην καριέρα του",  
xlab = "αριθμός παιχνιδιών",  
ylab = "έτη")
```

```
hist(data$INNINGS,  
main = "Συνολικά winning pitched",  
xlab = "παιχνίδια",  
ylab = "winning pitched")
```

```
hist(data$SAVES,  
main = "Συνολικά saves στην καριέρα του",  
xlab = "παιχνίδια",  
ylab = "συνολικά saves")
```

```
cor(na.omit(data[,3:12]))  
pairs(data[,3:12])  
library(sjPlot)  
tab_corr( na.omit(data[,3:12]), corr.method= "spearman", p.numeric = TRUE)
```

#ελεγχουμε αν salary87 ακολουθει κανονικη κατανομη

```
by(data$SALARY87,data$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθειται κανονικη  
κατανομη
```

table(data\$LEAGUE87) # και τα δυο δειγματα ειναι >50 , προχωρουμε στην εξεταση της συμμετριας

```
install.packages("moments")#κατεβαζουμε το πακετο moments
```

```
library(moments)#το φορτωνουμε στην R
```

```
by(data$SALARY87,data$LEAGUE87,agostino.test) #και οι δυο ομαδες της μεταβλητης salary87  
εχουν ασυμμετρια, επομενως δεν προσεγγιζεται η κανονικη κατανομη
```

#θα κανουμε wilcoxon test για να ελεγχουμε τις διαμεσους

wilcox.test(data\$SALARY87~data\$LEAGUE87) #p-value>5% , αρα δεν διαφερουν οι διαμεσοι των μισθων αναμεσα σε αμερικανικη και εθνικη κατηγορια

#WINS86 ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$WINS86,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθειται κανονικη κατανομη

by(data\$WINS86,data\$LEAGUE87,agostino.test)#και οι δυο ομαδες της μεταβλητης WINS86 εχουν ασυμμετρια, επομενως δεν προσεγγιζεται η κανονικη κατανομη

#θα κανουμε wilcoxon test για να ελεγχουμε τις διαμεσους

wilcox.test(data\$WINS86~data\$LEAGUE87) #p-value>5% , αρα δεν διαφερουν οι διαμεσοι των νικων για το 1986 αναμεσα σε αμερικανικη και εθνικη κατηγορια

#LOSSES86 ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$LOSSES86,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθειται κανονικη κατανομη

by(data\$LOSSES86,data\$LEAGUE87,agostino.test)#και οι δυο ομαδες της μεταβλητης LOSSES86 εχουν συμμετρία

by(data\$LOSSES86 ,data\$LEAGUE87,bonett.test)#p-value < 5% , δεν είναι μεσόκυρτη , άρα δεν προσεγγίζεται η κανονικότητα

#θα κανουμε wilcoxon test για να ελεγχουμε τις διαμεσους

wilcox.test(data\$LOSSES86~data\$LEAGUE87)#p-value>5% , αρα δεν διαφερουν οι διαμεσοι των ηττων για το 1986 αναμεσα σε αμερικανικη και εθνικη κατηγορια

#GAMES86 ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$GAMES86 ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$GAMES86 ,data\$LEAGUE87,agostino.test)#και οι δυο ομάδες της μεταβλητής GAMES86 έχουν ασυμμετρία, επομένως δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$GAMES86 ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί των παιχνιδιών ανάμεσα σε αμερικανική και εθνική κατηγορία

#LEAGUES ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$LEAGUES ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$LEAGUES ,data\$LEAGUE87,agostino.test)#και οι δυο ομάδες της μεταβλητής LEAGUES έχουν ασυμμετρία,δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$LEAGUES ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί ανάμεσα σε αμερικανική και εθνική κατηγορία

#WINS ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$WINS ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$WINS ,data\$LEAGUE87,agostino.test)#και οι δυο ομάδες της μεταβλητής WINS έχουν ασυμμετρία,δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$WINS ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί των νικών ανάμεσα σε αμερικανική και εθνική

#LOSSES ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$LOSSES ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$LOSSES ,data\$LEAGUE87,agostino.test)#p-value <5% , δεν έχω συμμετρία

wilcox.test(data\$LOSSES ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί των ηττών ανάμεσα σε αμερικανική και εθνική

#GAMES ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$GAMES ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$GAMES ,data\$LEAGUE87,agostino.test)#ασυμμετρία,δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$GAMES ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί των παιχνιδιών ανάμεσα σε αμερικανική και εθνική

#INNINGS ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$INNINGS ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$INNINGS ,data\$LEAGUE87,agostino.test)#ασυμμετρία,δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$INNINGS ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί ανάμεσα σε αμερικανική και εθνική

#SAVES ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

by(data\$SAVES ,data\$LEAGUE87,shapiro.test)#p-value < 5% , δεν ακολουθείται κανονική κατανομή

by(data\$SAVES ,data\$LEAGUE87,agostino.test)#ασυμμετρία,δεν προσεγγίζεται η κανονική κατανομή

wilcox.test(data\$SAVES ~data\$LEAGUE87)#p-value>5% ,αρα δεν διαφέρουν οι διαμεσοί ανάμεσα σε αμερικανική των saves και εθνική

par(mfrow=c(3,4))

boxplot(data\$SALARY87~data\$LEAGUE87 , boxcol=0,

main="μέσες απολαβές 1987 για κάθε κατηγορία " ,

ylab="μέσες απολαβές",

```
xlab="κατηγορία μπείζμπολ",
```

```
boxfill=3,
```

```
medlwd=2,
```

```
medcol="white")
```

```
boxplot(data$SAVES~data$LEAGUE87 , boxcol=0,
```

```
main="Συνολικά saves για κάθε κατηγορία " ,
```

```
ylab="Συνολικά saves καθόλη την διάρκεια της καριέρας του παίχτη",
```

```
ylim=c(0,200) ,
```

```
xlab="κατηγορία μπείζμπολ",
```

```
boxfill=3,
```

```
medlwd=2,
```

```
medcol="white") #πολλα outliers - επανελεγχο
```

```
boxplot(data$INNINGS~data$LEAGUE87 , boxcol=0,
```

```
main="Συνολικά winning pitched για κάθε κατηγορία " ,
```

```
ylab="Συνολικά winning pitched καθόλη την διάρκεια της καριέρας του παίχτη " ,
```

```
xlab="κατηγορία μπείζμπολ",
```

```
boxfill=3,
```

```
medlwd=2,
```

```
medcol="white")
```

```
boxplot(data$GAMES~data$LEAGUE87 , boxcol=0,
```

```
main="Συνολικές συμμετοχές σε αγώνες για κάθε κατηγορία " ,
```

```
ylab="Συνολικές συμμετοχές σε αγώνες στην καριέρα του παίχτη " ,
```

```
xlab="κατηγορία μπείζμπολ",
```

```
boxfill=3,
```

```
medlwd=2,  
medcol="white")
```

```
boxplot(data$LOSSES~data$LEAGUE87 , boxcol=0,  
main="Συνολικές ήττες για κάθε κατηγορία " ,  
ylab="Συνολικές ήττες καθόλη την διάρκεια της καριέρας του παίχτη " ,  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
boxplot(data$WINS~data$LEAGUE87 , boxcol=0,  
main="Συνολικές νίκες για κάθε κατηγορία " ,  
ylab="Συνολικές νίκες καθόλη την διάρκεια της καριέρας του παίχτη " ,  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
boxplot(data$LEAGUES~data$LEAGUE87 , boxcol=0,  
main="αριθμός ετών σε μεγάλες κατηγορίες για κάθε κατηγορία μπέιζμπολ " ,  
ylab="αριθμός ετών σε μεγάλες κατηγορίες",  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
boxplot(data$GAMES86~data$LEAGUE87 , boxcol=0,
```

```
main="Συνολικές συμμετοχές σε αγώνες το 1986 για κάθε κατηγορία " ,  
ylab="αριθμός αγώνων το 1986",  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
boxplot(data$LOSSES86~data$LEAGUE87 , boxcol=0,  
main="Συνολικές ήττες το 1986 για κάθε κατηγορία " ,  
ylab="αριθμός ηττών το 1986",  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
boxplot(data$WINS86~data$LEAGUE87 , boxcol=0,  
main="Συνολικές νίκες το 1986 για κάθε κατηγορία " ,  
ylab="αριθμός νικών το 1986",  
xlab="κατηγορία μπέιζμπολ",  
boxfill=3,  
medlwd=2,  
medcol="white")
```

```
#polisigrammikohta
```

```
fullmodel1 <- lm(data$SALARY87~. , data= data[,-c(1,2,12)])
```

```
summary(fullmodel1)
```

```
install.packages("car")
```

```
library(car)
```

```
#polisigrammikitita
```

```
vif(fullmodel1)
```

```
# paraviazetai i kanonikotima, omosk, grammikitita
```

```
fullmodel1 <- lm(data$SALARY87~. , data= data[,-c(1,2,12,13,10,6,9,8)])
```

```
summary(fullmodel1)
```

```
#polisigrammikitita
```

```
vif(fullmodel1)
```

```
#normality
```

```
plot(fullmodel1,which=2)
```

```
#homoskedasticity-linearity
```

```
plot( fullmodel1$fit, rstandard(fullmodel1), cex=2)
```

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

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

```
quantcut<-function(x){ cut(x,breaks=quantile(x),include.lowest=T)}
```

```
qfits<-quantcut(fullmodel1$fit)
```

```
leveneTest(rstandard(fullmodel1),qfits)#eteroskedastikitita
```

```
residualPlots(fullmodel1)
```

```
# paraviazetai i kanonikotima, omosk, grammikitita
```

```
fullmodel1 <- lm(data$SALARY87~. , data= data[,-c(1,2,12,13,10,6,9,8)])
```

```
summary(fullmodel1)
```

```
#polisigrammikitita
```

```
vif(fullmodel1)
```

```
#normality
```

```
plot(fullmodel1,which=2)
```



```
#homoskedasticity-linearity

plot( fullmodel1$fit, rstandard(fullmodel1), cex=2)

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

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

quantcut<-function(x){cut(x,breaks=quantile(x),include.lowest=T)}

qfits<-quantcut(fullmodel1$fit)

leveneTest(rstandard(fullmodel1),qfits)#eteroskedastikotita

residualPlots(fullmodel1)
```

```
# paraviazetai i grammikotita

fullmodel1 <- lm(log(data$SALARY87)~. , data= data[,-c(1,2,12,13,10,6,9,8)])

summary(fullmodel1)

#normality

plot(fullmodel1,which=2)

#homoskedasticity-linearity

plot( fullmodel1$fit, rstandard(fullmodel1), cex=2)

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

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

quantcut<-function(x){cut(x,breaks=quantile(x),include.lowest=T)}

qfits<-quantcut(fullmodel1$fit)

leveneTest(rstandard(fullmodel1),qfits)#p>5%-omoskedastikotita

residualPlots(fullmodel1)
```

```
# vrika akraies times kai ena moxl.

fullmodel1 <- lm(log(data$SALARY87)~WINS86+LOSSES86+GAMES86+log(WINS)+SAVES ,
data= data[,-c(1,2,12,13,10,6,9,8)])

summary(fullmodel1)
```

```

#normality

plot(fullmodel1,which=2)

#homoskedasticity-linearity

plot( fullmodel1$fit, rstandard(fullmodel1), cex=2)

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

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

quantcut<-function(x){ cut(x,breaks=quantile(x),include.lowest=T)}

qfits<-quantcut(fullmodel1$fit)

leveneTest(rstandard(fullmodel1),qfits)# homoskedastikotita

residualPlots(fullmodel1)

#outlier

plot(fullmodel1,which=4)

outlierTest(fullmodel1)

abline(h=4/(nrow(data)-2-1),col='blue',lty=2)

plot(fullmodel1,which=5)

```

#ΑΦΑΙΡΕΣΑ ΤΩ ΣΗΜΕΙΟ ΜΟΧΛΕΥΣΗΣ

```

data2=data[-56,]

fullmodel2 <- lm(log(data2$SALARY87)~WINS86+LOSSES86+GAMES86+log(WINS)+SAVES ,
data= data2[,-c(1,2,12,13,10,6,9,8)])

summary(fullmodel1)

#normality

plot(fullmodel2,which=2)

shapiro.test(fullmodel2$res) # kanonikotita

#homoskedasticity-linearity

plot( fullmodel2$fit, rstandard(fullmodel2), cex=2)

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

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

```

```

quantcut<-function(x){cut(x,breaks=quantile(x),include.lowest=T)}

qfits<-quantcut(fullmodel2$fit)

leveneTest(rstandard(fullmodel2),qfits)#omoskedastikotita

residualPlots(fullmodel2)

#outlier

plot(fullmodel2,which=4)

outlierTest(fullmodel2)

abline(h=4/(nrow(data)-2-1),col='blue',lty=2)

plot(fullmodel2,which=5)

#ΔΙΑΦΟΡΟΠΟΙΕΙΤΑΙ ΤΟ ΜΟΝΤΕΛΟ

#ΑΦΑΙΡΟΥΜΕ ΚΑΙ 2 ΑΚΡΑΙΕΣ ΤΙΜΕΣ

data3=data[-c(25,28,56),]

fullmodel3 <- lm(log(data3$SALARY87)~WINS86+LOSSES86+GAMES86+log(WINS)+SAVES ,
data= data3[,-c(1,2,12,13,10,6,9,8)])

summary(fullmodel3)

#ΤΟ ΜΟΝΤΕΛΟ ΔΙΑΦΟΡΟΠΟΙΕΙΤΑΙ

#ΔΟΚΙΜΑΣΑ ΝΑ ΑΦΑΙΡΕΣΩ ΚΑΙ ΤΗΝ ΝΕΑ ΑΚΡΑΙΑ ΤΙΜΗ ΑΛΛΑ ΔΕΝ ΘΑ ΑΛΛΑΞΕΙ ΤΟ
ΜΟΝΤΕΛΟ.

data4=data[-c(131,25,28,56),]

fullmodel4 <- lm(log(data4$SALARY87)~WINS86+LOSSES86+GAMES86+log(WINS)+SAVES ,
data= data4[,-c(1,2,12,10,6,9,8)])

summary(fullmodel4)

tab_model( fullmodel1 ,fullmodel2,fullmodel3 ,fullmodel4, show.se=TRUE, show.stat=TRUE ,
show.df=TRUE

, show.r2=TRUE , emph.p=TRUE ,dv.labels = c("Μοντέλο με ακραίες τιμές και μόχλευση " ,
" Μοντέλο χωρίς μόχλευση " , "Μοντέλο με αφαίρεση των 2 ακραίων τιμών " ,

```

"Μοντέλο χωρίς ακραίες τιμές "))