

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΜΕΛΕΤΗ ΤΩΝ ΚΥΡΙΩΝ
ΠΡΟΒΛΕΠΤΙΚΩΝ ΔΕΙΚΤΩΝ ΓΙΑ ΤΗΝ
ΕΞΕΛΙΞΗ ΕΝΟΣ ΑΓΩΝΑ ΜΠΑΣΚΕΤ
ΣΤΗΝ ΕΥΡΩΛΙΓΚΑ ΓΥΝΑΙΚΩΝ**

Βασιλική Ν. Παναγιωτοπούλου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Ιούνιος 2026

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΜΕΛΕΤΗ ΤΩΝ ΚΥΡΙΩΝ
ΠΡΟΒΛΕΠΤΙΚΩΝ ΔΕΙΚΤΩΝ ΓΙΑ ΤΗΝ
ΕΞΕΛΙΞΗ ΕΝΟΣ ΑΓΩΝΑ ΜΠΑΣΚΕΤ
ΣΤΗΝ ΕΥΡΩΛΙΓΚΑ ΓΥΝΑΙΚΩΝ**

Βασιλική Ν. Παναγιωτοπούλου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Ιούνιος 2026

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη Συνέλευση του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή της σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Κωνσταντίνος Πολίτης, Καθηγητής (Επιβλέπων)
- Εμμανουήλ Ανδρουλάκης, Επίκουρος Καθηγητής
- Χαράλαμπος Ευαγγελάρας, Αναπληρωτής Καθηγητής

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**A STUDY OF PERFORMANCE
INDICATORS IN EUROLEAGUE
WOMEN'S COMPETITION**

By

Vasiliki N. Panagiotopoulou

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfilment of
the requirements for the degree of Master of Science in
Applied Statistics

Piraeus, Greece
June 2026

*Στους γονείς μου, Νίκο και Στεργιανή,
Στην αδερφή μου, Ραφαέλα,
Στη γιαγιά μου, Αγορίτσα.*

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους με στήριξαν και στάθηκαν δίπλα μου όλα αυτά τα χρόνια της ακαδημαϊκής μου πορείας.

Αρχικά, θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου, κ. Κωνσταντίνο Πολίτη, για την πολύτιμη καθοδήγηση, την επιστημονική υποστήριξη και τις ουσιαστικές συμβουλές του σε όλα τα στάδια εκπόνησης της εργασίας. Η συνεχής διαθεσιμότητα, η υπομονή και η εμπιστοσύνη που μου έδειξε συνέβαλαν σημαντικά στην ολοκλήρωση της παρούσας μελέτης.

Ιδιαίτερες ευχαριστίες οφείλω στην οικογένειά μου για την αμέριστη αγάπη, τη στήριξη και την κατανόηση που μου προσέφερε όλα αυτά τα χρόνια. Η πίστη τους σε εμένα αποτέλεσε διαχρονικά πηγή δύναμης και κινήτρου σε κάθε δυσκολία.

Τέλος, θα ήθελα να ευχαριστήσω θερμά τους φίλους μου για τη συμπαράσταση, την υπομονή και την ψυχολογική υποστήριξη όλα αυτά τα χρόνια. Η παρουσία τους υπήρξε ιδιαίτερα σημαντική και πολύτιμη για εμένα.

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει τους βασικούς προβλεπτικούς δείκτες που σχετίζονται με την αγωνιστική επιτυχία των ομάδων στην EuroLeague Γυναικών. Στόχος της έρευνας είναι η διερεύνηση των στατιστικών μεταβλητών που επηρεάζουν την εξέλιξη και την έκβαση ενός αγώνα, καθώς και η ανάδειξη των παραγόντων που συνδέονται περισσότερο με την πρόκριση και τη συνολική απόδοση των ομάδων στη διοργάνωση.

Για τις ανάγκες της ανάλυσης χρησιμοποιήθηκαν επίσημα αγωνιστικά δεδομένα της FIBA από τις αγωνιστικές περιόδους 2004–2025, τα οποία περιλαμβάνουν στατιστικά στοιχεία 108 διαφορετικών ομάδων της EuroLeague Γυναικών. Οι μεταβλητές που εξετάζονται αφορούν βασικούς αγωνιστικούς δείκτες, όπως οι πόντοι ανά αγώνα, τα ποσοστά ευστοχίας, τα ριμπάουντ, οι ασίστ, τα λάθη, τα κλεψίματα, τα κοψίματα και ο δείκτης αποδοτικότητας (EFFPG).

Αρχικά πραγματοποιείται περιγραφική στατιστική ανάλυση και γραφική διερεύνηση των δεδομένων μέσω χρονοσειρών, scatter plots και Chernoff Faces. Στη συνέχεια εφαρμόζονται στατιστικοί έλεγχοι, αναλύσεις συσχέτισης, μοντέλα λογιστικής και πολυωνυμικής λογιστικής παλινδρόμησης, καθώς και μέθοδοι ανάλυσης κατά συστάδες (nearest neighbor, Ward και k-means), με στόχο την αναγνώριση των σημαντικότερων παραγόντων που σχετίζονται με την αγωνιστική επιτυχία και την πρόκριση στο Final 8.

Τα αποτελέσματα δείχνουν ότι μεταβλητές όπως οι πόντοι ανά αγώνα, το ποσοστό ευστοχίας στα δίποντα, οι ασίστ, τα αμυντικά ριμπάουντ και ο δείκτης αποδοτικότητας σχετίζονται θετικά με την επιτυχία των ομάδων, ενώ τα λάθη και τα προσωπικά φάουλ παρουσιάζουν αρνητική σχέση με την απόδοση. Τέλος, πραγματοποιείται συγκριτική προσέγγιση με αντίστοιχη μελέτη για την EuroLeague Ανδρών, με στόχο την ανάδειξη πιθανών διαφορών και ομοιοτήτων μεταξύ ανδρικού και γυναικείου ευρωπαϊκού μπάσκετ.

Abstract

This master's thesis examines the main predictive indicators associated with team success in the Women's EuroLeague. The aim of the study is to investigate the statistical variables that influence the development and outcome of a basketball game, as well as to identify the factors most strongly related to qualification and overall team performance in the competition.

The analysis is based on official FIBA data from the 2004–2025 seasons, including statistical records from 108 different teams participating in the Women's EuroLeague. The variables examined include key performance indicators such as points per game, shooting percentages, rebounds, assists, turnovers, steals, blocks, and the efficiency rating (EFFPG).

Initially, descriptive statistical analysis and graphical investigation of the data are conducted through time series plots, scatter plots, and Chernoff Faces. Subsequently, statistical tests, correlation analyses, logistic and multinomial logistic regression models, as well as cluster analysis methods (nearest neighbor, Ward, and k-means), are applied in order to identify the most important factors related to team success and qualification to the Final 8 stage.

The results indicate that variables such as points per game, two-point shooting percentage, assists, defensive rebounds, and efficiency rating are positively associated with team success, while turnovers and personal fouls are negatively related to performance. Finally, a comparative approach with a similar study concerning the Men's EuroLeague is conducted in order to highlight possible similarities and differences between men's and women's European basketball regarding the factors affecting team success.

Περιεχόμενα

Περίληψη	j
Abstract	l
Περιεχόμενα	a
Κατάλογος Πινάκων	f
Κατάλογος Σχημάτων	i
Κεφάλαιο 1: Εισαγωγή – Σκοπός και Δομή της Εργασίας	13
1.1 Εισαγωγή	13
1.2 Σκοπός της Εργασίας	13
1.3 Δομή της Εργασίας	14
Κεφάλαιο 2: Ιστορική αναδρομή	16
2.1 Η διοργάνωση της EuroLeague	16
2.1.1 Ιστορική αναδρομή	16
2.1.3 Ιστορική αναδρομή στα διαφορετικά format της διοργάνωσης στο πέρασμα των χρόνων	16
2.2 Η EuroLeague Γυναικών στη σύγχρονη εποχή και ο κοινωνικός-αθλητικός της αντίκτυπος στην Ευρώπη	17
2.2.1 Επιρροή κι αντίκτυπος στην Ευρώπη	17
2.2.2 Κοινωνική διάσταση και σύγκριση με την EuroLeague Ανδρών.	18
2.3 Σημαντικά γεγονότα και milestones της EuroLeague Γυναικών	18
2.4 Ελληνική συμμετοχή	19
2.5 Ξεχωριστές παίκτριες και διεθνής αναγνώριση	19
2.6 Βασικοί Δείκτες Απόδοσης: Βιβλιογραφική Ανασκόπηση	21
Κεφάλαιο 3: Περιγραφική ανάλυση	22
3.1 Παρουσίαση δεδομένων	22
3.2 Πλαίσιο και Σκοπός Ανάλυσης	24
3.2.1 Διερεύνηση για ελλιπή δεδομένα (missing values)	24
3.3 Περιγραφική ανάλυση	24
3.3.1 Περιγραφικά στατιστικά μέτρα όλων των ομάδων για όλες τις σεζόν	24

3.3.2 Περιγραφικά στατιστικά μέτρα όλων των ομάδων βάσει της αγωνιστικής τους πορείας στη διοργάνωση	29
Κεφάλαιο 4 : Στατιστική Ανάλυση και Σύγκριση Δεικτών Απόδοσης	34
4.1 Έλεγχος κανονικότητας.	34
4.1.1 Θεωρητικό υπόβαθρο	34
4.1.2 Δημοφιλείς Έλεγχοι Κανονικότητας	35
4.1.3 Έλεγχος κανονικότητας στα δεδομένα.....	35
4.2 Έλεγχος ισότητας διακυμάνσεων.	36
4.2.1 Θεωρητικό υπόβαθρο	36
4.2.2 Έλεγχος ισότητας διακυμάνσεων στα δεδομένα	37
4.3. Έλεγχοι διαφορών μεταξύ ομάδων.....	38
4.3.1 Θεωρητικό υπόβαθρο	38
4.3.3 Στατιστική ερμηνεία.....	39
4.3.4 T-TEST στα δεδομένα.....	39
4.4 Γραφική απεικόνιση και συντελεστής συσχέτισης.....	46
4.4.1 Θεωρητικό υπόβαθρο	46
4.4.2 Γραφική απεικόνιση	47
4.4.3 Συντελεστής συσχέτισης.....	56
Κεφάλαιο 5 : Ανάλυση παλινδρόμησης.....	58
5.1 Λογιστική παλινδρόμηση	58
5.1.1 Θεωρητικό υπόβαθρο	58
5.1.2 Μοντέλο λογιστικής παλινδρόμησης	60
5.1.3 Πολυσυγγραμμικότητα.....	64
5.1.3.1 Θεωρητικό υπόβαθρο.....	64
5.1.3.2 Έλεγχος πολυσυγγραμμικότητας.....	65
5.1.4 Γραφική απεικόνιση	66
5.1.5 Δείκτης Nagelkerke R^2	67
5.1.5.1 Θεωρητικό υπόβαθρο.....	67
5.1.5.2 Υπολογισμός του δείκτη Nagelkerke R^2	68
5.1.6 Κατάλοιπα απόκλισης	69
5.1.6.1 Θεωρητικό υπόβαθρο.....	69

5.1.6.2 Κατάλοιπα απόκλισης τελικού μοντέλου.....	70
5.1.7 Έλεγχος Hosmer – Lemeshow.....	71
5.1.7.1 Θεωρητικό υπόβαθρο.....	71
5.1.7.2 Έλεγχος Hosmer – Lemeshow στο τελικό μοντέλο	72
5.2 Πολυωνυμική Λογιστική Παλινδρόμηση.....	73
5.2.1 Θεωρητικό υπόβαθρο	73
5.2.2 Πολυωνυμικό λογιστικό μοντέλο	74
5.2.3 Πίνακας ταξινόμησης.....	77
5.2.4 Καμπύλη roc στο πολυωνυμικό μοντέλο	79
5.3 Συμπεράσματα για την ανάλυση παλινδρόμησης.....	82
Κεφάλαιο 6 : Ανάλυση κατά Συστάδες.....	84
6.1 Θεωρητικό υπόβαθρο.....	84
6.2 Μέθοδος Ward	88
6.3 Μέθοδος k-means	89
6.4 Σύγκριση αποτελεσμάτων	92
6.5 Σύγκριση αποτελεσμάτων με τη διεθνή βιβλιογραφία	93
Κεφάλαιο 7: Συμπεράσματα και Συζήτηση Αποτελεσμάτων	95
7.1 Συμπεράσματα της παρούσας έρευνας	95
7.2 Σύγκριση με τη διεθνή βιβλιογραφία και την EuroLeague Ανδρών.....	97
Βιβλιογραφία	99
Ελληνική Βιβλιογραφία	99
Ξενόγλωσση Βιβλιογραφία.....	99
Διαδικτυακές Πηγές.....	100
Παράρτημα	101
Κώδικας.....	101
Ονόματα ομάδων.....	110

Κατάλογος Πινάκων

Πίνακας 2.4. Κορυφαίες σκόρερ όλων των εποχών.....	20
Πίνακας 3.1.1. Οι μεταβλητές.....	23
Πίνακας 3.3.1. Στατιστικά περιγραφικά μέτρα.....	25
Πίνακας 3.3.2 Συντελεστές συσχέτισης Pearson των μεταβλητών με το δείκτη EFFPG	28
Πίνακας 3.3.3. Chernoff faces, effect of variables	29
Πίνακας 3.3.2.1. Στατιστικά περιγραφικά μέτρα για τις ομάδες που δεν προκρίθηκαν στο Final 8	30
Πίνακας 3.3.2.2. Στατιστικά περιγραφικά μέτρα για τις ομάδες που προκρίθηκαν στο Final 8 αλλά όχι στο Final 4	30
Πίνακας 3.3.2.3. Στατιστικά περιγραφικά μέτρα για τις ομάδες που προκρίθηκαν στο Final 4.....	31
Πίνακας 4.1.1. Έλεγχος κανονικότητας	36
Πίνακας 4.2.2.1. Έλεγχος ισότητας διακυμάνσεων	38
Πίνακας 4.3.4.1. Mann-Whitney U test για την μεταβλητή PPG	40
Πίνακας 4.3.4.2. Two Sample t-test για την μεταβλητή 2P%	40
Πίνακας 4.3.4.3. Mann-Whitney U test για την μεταβλητή 3P%	41
Πίνακας 4.3.4.4. Welch Two Sample t-test για την μεταβλητή FT%.....	41
Πίνακας 4.3.4.5. Mann-Whitney U test για την μεταβλητή ORPG	42
Πίνακας 4.3.4.6. Welch Two Sample t-test για την μεταβλητή DRPG	42
Πίνακας 4.3.4.7. Mann-Whitney U test για την μεταβλητή RPG	43
Πίνακας 4.3.4.8. Mann-Whitney U test για την μεταβλητή APG.....	43
Πίνακας 4.3.4.9. Mann-Whitney U test για την μεταβλητή STLPG.....	44
Πίνακας 4.3.4.10. Welch Two Sample t-test για την μεταβλητή TOPG.....	44
Πίνακας 4.3.4.11. Mann-Whitney U test για την μεταβλητή BLKPG.....	45
Πίνακας 4.3.4.12. Mann-Whitney U test για την μεταβλητή PFPG	45
Πίνακας 4.3.4.13. Mann-Whitney U test για την μεταβλητή EFFPG.....	46
Πίνακας 4.3.5. Συνολικά αποτελέσματα	46
Πίνακας 4.4.3.1. Point-biserial correlation	57

Πίνακας 5.1.2.1. Μοντέλο παλινδρόμησης	62
Πίνακας 5.1.2.2. Εκθετικοί συντελεστές (odds ratios) και διαστήματα εμπιστοσύνης	63
Πίνακας 5.1.3.2 Έλεγχος πολυσυγγραμμικότητας.....	65
Πίνακας 5.1.4. Area Under the Curve (AUC).....	67
Πίνακα 5.1.5.2.1 ο δείκτης Nagelkerke R^2 για το τελικό μοντέλο.....	69
Πίνακας 5.1.7.1. Έλεγχος Hosmer–Lemeshow στο τελικό μοντέλο	72
Πίνακας 5.2.2.1. Πολυωνυμικό λογιστικό μοντέλο	76
Πίνακας 5.2.2.2. Odds ratios.....	76
Πίνακας 5.2.2.3. Έλεγχος πολυσυγγραμμικότητας στο πολυωνυμικό μοντέλο	77
Πίνακας 5.2.3.1. Πίνακας ταξινόμησης.....	78
Πίνακας 5.2.3.2. Εκτιμημένες πιθανότητες	79
Πίνακας 5.2.4.1. AUC	82
Πίνακας 6.2. Πίνακας συχνотήτων	89
Πίνακας 6.3.1. Πίνακας συχνотήτων K-means.....	89
Πίνακας 6.3.2. Μέσοι όροι στατιστικών δεικτών ανά συστάδα (k-means).....	90
Πίνακας 6.3.3. Πίνακας σύγχυσης (Confusion Matrix)	91
Πίνακας 6.3.4. Συνολική ακρίβεια της ταξινόμησης	91
Πίνακας 3.1 Όνομα και id των ομάδων.....	110

Κατάλογος Σχημάτων

Σχήμα 3.3.1. Time Series Plots των στατιστικών στοιχείων με τη πάροδο των σεζόν.....	26
Σχήμα 3.3.2. Scatter Plots των στατιστικών στοιχείων με το δείκτη EFFPG	27
Σχήμα 3.3.3. Chernoff faces για κάθε σεζόν.....	29
Σχήμα 3.3.4. Boxplot κάθε μεταβλητής ανάλογα με την κατηγορία πρόκρισης.....	32
Σχήμα 3.3.5. Line plot κάθε μεταβλητής ανάλογα με την κατηγορία πρόκρισης	33
Σχήμα 4.4.2.1. Violin plot για PPG.....	48
Σχήμα 4.4.2.2. Violin plot για 2P%.....	49
Σχήμα 4.4.2.3. Violin plot για 3P%.....	49
Σχήμα 4.4.2.4. Violin plot για FT%	50
Σχήμα 4.4.2.5. Violin plot για ORPG.....	51
Σχήμα 4.4.2.6. Violin plot για DRPG.....	51
Σχήμα 4.4.2.7. Violin plot για RPG	52
Σχήμα 4.4.2.8. Violin plot για APG	53
Σχήμα 4.4.2.9. Violin plot για STLPG	53
Σχήμα 4.4.2.10. Violin plot για TOPG.....	54
Σχήμα 4.4.2.11. Violin plot για BLKPG	55
Σχήμα 4.4.2.12. Violin plot για PFPNG.....	55
Σχήμα 4.4.2.13. Violin plot για EFFPG	56
Σχήμα 5.1.4. Καμπύλη ROC.....	66
Σχήμα 5.1.6. Κατάλοιπα απόκλισης τελικού μοντέλου.....	70
Σχήμα 5.2.4.1. Καμπύλη ROC στο πολυωνυμικό μοντέλο	81
Σχήμα 6.2. Δενδροδιάγραμμα	89
Σχήμα 6.3 Διάγραμμα διασποράς των συστάδων μέσω των δύο πρώτων κύριων συνιστωσών (PCA).....	92

Κεφάλαιο 1: Εισαγωγή – Σκοπός και Δομή της Εργασίας

1.1 Εισαγωγή

Τα τελευταία χρόνια, η ανάλυση δεδομένων έχει αποκτήσει ιδιαίτερη σημασία στον χώρο του αθλητισμού, καθώς η αξιοποίηση στατιστικών πληροφοριών συμβάλλει ουσιαστικά στην κατανόηση της αγωνιστικής απόδοσης ομάδων και αθλητών. Ειδικότερα στο μπάσκετ, η χρήση στατιστικών δεικτών αποτελεί βασικό εργαλείο αξιολόγησης της αποτελεσματικότητας μιας ομάδας, τόσο σε επιθετικό όσο και σε αμυντικό επίπεδο. Μέσω της στατιστικής ανάλυσης καθίσταται δυνατός ο εντοπισμός των παραγόντων που σχετίζονται με την αγωνιστική επιτυχία, καθώς και η κατανόηση των στοιχείων που επηρεάζουν την τελική έκβαση ενός αγώνα.

Η EuroLeague Γυναικών αποτελεί την κορυφαία διασυλλογική διοργάνωση γυναικείου μπάσκετ στην Ευρώπη και συγκεντρώνει ομάδες υψηλού αγωνιστικού επιπέδου από διαφορετικές χώρες. Η συνεχής ανάπτυξη της διοργάνωσης, η αυξανόμενη προβολή του γυναικείου αθλητισμού και η διαθεσιμότητα αναλυτικών αγωνιστικών δεδομένων δημιουργούν ένα ιδιαίτερα ενδιαφέρον πεδίο στατιστικής μελέτης. Παράλληλα, η σύγχρονη τάση αξιοποίησης της αθλητικής ανάλυσης (sports analytics) έχει ενισχύσει τη σημασία της διερεύνησης των βασικών δεικτών απόδοσης, τόσο σε ερευνητικό όσο και σε πρακτικό επίπεδο.

1.2 Σκοπός της Εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η μελέτη των κύριων προβλεπτικών δεικτών που σχετίζονται με την εξέλιξη και την αγωνιστική επιτυχία των ομάδων στην EuroLeague Γυναικών. Μέσα από τη στατιστική επεξεργασία δεδομένων επιχειρείται η διερεύνηση της σχέσης μεταξύ βασικών αγωνιστικών μεταβλητών, όπως οι πόντοι ανά αγώνα, τα ποσοστά ευστοχίας, τα ριμπάουντ, οι ασίστ, τα λάθη και ο δείκτης αποδοτικότητας, με την πορεία των ομάδων στη διοργάνωση.

Η ανάλυση βασίζεται σε επίσημα δεδομένα της FIBA που αφορούν τις αγωνιστικές περιόδους 2004–2025 και περιλαμβάνουν στατιστικά στοιχεία ομάδων της EuroLeague Γυναικών. Μέσα από την εφαρμογή περιγραφικής στατιστικής, στατιστικών ελέγχων, μοντέλων παλινδρόμησης και μεθόδων ανάλυσης κατά συστάδες, επιδιώκεται η αναγνώριση των μεταβλητών που επηρεάζουν περισσότερο την πιθανότητα επιτυχίας και πρόκρισης των ομάδων.

Παράλληλα, στόχος της εργασίας είναι η κατανόηση της αγωνιστικής φυσιολογίας του σύγχρονου ευρωπαϊκού γυναικείου μπάσκετ, καθώς και η ανάδειξη πιθανών προτύπων απόδοσης που χαρακτηρίζουν τις πιο επιτυχημένες ομάδες της διοργάνωσης.

1.3 Δομή της Εργασίας

Η εργασία οργανώνεται σε επτά βασικά κεφάλαια.

Στο δεύτερο κεφάλαιο παρουσιάζεται η ιστορική αναδρομή της EuroLeague Γυναικών, οι σημαντικότερες αλλαγές που πραγματοποιήθηκαν διαχρονικά στη δομή της διοργάνωσης, καθώς και ο κοινωνικός και αθλητικός της αντίκτυπος στην Ευρώπη. Παράλληλα γίνεται αναφορά στις σημαντικότερες αθλήτριες της διοργάνωσης, στις ελληνικές συμμετοχές και στη σχετική βιβλιογραφία που αφορά τους βασικούς δείκτες απόδοσης στο γυναικείο μπάσκετ.

Στο τρίτο κεφάλαιο παρουσιάζεται η περιγραφική στατιστική ανάλυση των δεδομένων. Αρχικά γίνεται παρουσίαση των μεταβλητών και των βασικών περιγραφικών μέτρων, ενώ στη συνέχεια πραγματοποιείται γραφική διερεύνηση των δεδομένων μέσω χρονοσειρών, scatter plots και Chernoff Faces. Επιπλέον εξετάζονται οι σχέσεις μεταξύ των αγωνιστικών μεταβλητών και του δείκτη αποδοτικότητας, καθώς και οι διαφοροποιήσεις των ομάδων ανάλογα με την πορεία τους στη διοργάνωση.

Στο τέταρτο κεφάλαιο πραγματοποιείται η στατιστική ανάλυση και σύγκριση των βασικών αγωνιστικών δεικτών μεταξύ των ομάδων που προκρίθηκαν στο Final 8 και εκείνων που δεν κατάφεραν να προκριθούν. Οι ομάδες διαχωρίζονται σε δύο κατηγορίες επιτυχίας και εξετάζονται οι διαφορές τους ως προς βασικές αγωνιστικές μεταβλητές. Παράλληλα εφαρμόζονται στατιστικοί έλεγχοι και αναλύσεις συσχέτισης με στόχο την ανάδειξη των σημαντικότερων παραγόντων που σχετίζονται με την πρόκριση στη διοργάνωση.

Στο πέμπτο κεφάλαιο εφαρμόζονται μέθοδοι λογιστικής και πολυωνυμικής λογιστικής παλινδρόμησης για τη μοντελοποίηση και πρόβλεψη της αγωνιστικής επιτυχίας των ομάδων. Μέσα από τα μοντέλα εξετάζεται η επίδραση βασικών αγωνιστικών δεικτών στην πιθανότητα πρόκρισης και στη συνολική πορεία των ομάδων στη διοργάνωση. Επιπλέον αξιολογείται η προβλεπτική ικανότητα των μοντέλων και αναδεικνύονται οι σημαντικότεροι δείκτες που επηρεάζουν την επιτυχία των ομάδων.

Στο έκτο κεφάλαιο εφαρμόζονται μέθοδοι ανάλυσης κατά συστάδες (cluster analysis), με στόχο τη διερεύνηση της εσωτερικής δομής των δεδομένων και την αναγνώριση ομάδων με κοινά αγωνιστικά χαρακτηριστικά. Αρχικά εφαρμόζονται ιεραρχικές μέθοδοι ομαδοποίησης, όπως η μέθοδος του κοντινότερου γείτονα (nearest neighbor) και η μέθοδος Ward, με σκοπό τη διερεύνηση της δομής των δεδομένων και τον εντοπισμό πιθανών συστάδων μεταξύ των ομάδων μέσω δενδροδιαγραμμάτων.

Στη συνέχεια εφαρμόζεται η μη ιεραρχική μέθοδος k-means, προκειμένου να πραγματοποιηθεί διαχωρισμός των ομάδων σε ομάδες υψηλότερης και χαμηλότερης αγωνιστικής επιτυχίας. Μέσα από την ανάλυση εξετάζεται κατά πόσο οι ομάδες που παρουσιάζουν παρόμοια στατιστικά χαρακτηριστικά συγκεντρώνονται στις ίδιες συστάδες και πώς αυτές σχετίζονται με την πρόκριση στο Final 8 και τη συνολική αγωνιστική απόδοση των ομάδων.

Τέλος, στο τελευταίο κεφάλαιο παρουσιάζονται τα βασικά συμπεράσματα της έρευνας, συνοψίζονται τα σημαντικότερα ευρήματα της ανάλυσης και εξετάζεται η συμβολή των βασικών αγωνιστικών δεικτών στην αγωνιστική επιτυχία των ομάδων της EuroLeague Γυναικών. Παράλληλα πραγματοποιείται συγκριτική προσέγγιση με αντίστοιχη διπλωματική εργασία που αφορά την EuroLeague Ανδρών, με στόχο την ανάδειξη πιθανών ομοιοτήτων και διαφορών μεταξύ του ανδρικού και του γυναικείου ευρωπαϊκού μπάσκετ ως προς τους παράγοντες που επηρεάζουν την απόδοση και την επιτυχία των ομάδων.

Κεφάλαιο 2: Ιστορική αναδρομή

Στο κεφάλαιο που ακολουθεί παρουσιάζεται μια συνοπτική ιστορική αναδρομή της Ευρωλίγκας Γυναικών στο μπάσκετ, με στόχο την κατανόηση της εξέλιξης και της σημασίας της διοργάνωσης στο ευρωπαϊκό αθλητικό στερέωμα. Αρχικά παρουσιάζεται το πλαίσιο ίδρυσής της και οι βασικοί σταθμοί της πορείας της, στη συνέχεια αναλύονται οι σημαντικότερες αλλαγές που έχουν πραγματοποιηθεί διαχρονικά στην διοργάνωση, ενώ εξετάζεται και ο κοινωνικός της αντίκτυπος στην ανάπτυξη του γυναικείου αθλητισμού. Τέλος, γίνεται αναφορά στις κορυφαίες αθλήτριες που άφησαν το αποτύπωμά τους στη διοργάνωση. Οι πληροφορίες που αξιοποιήθηκαν για το συγκεκριμένο κεφάλαιο αντλήθηκαν κυρίως από την επίσημη ιστοσελίδα της FIBA(<https://www.fiba.basketball>) καθώς και από τη διαδικτυακή εγκυκλοπαίδεια Wikipedia(<https://en.wikipedia.org>).

2.1 Η διοργάνωση της EuroLeague

2.1.1 Ιστορική αναδρομή

Η ευρωλίγκα γυναικών (επίσημα γνωστή ως FIBA EuroLeague Γυναικών) είναι η κορυφαία ευρωπαϊκή διοργάνωση για ομάδες γυναικών και σε αντίθεση με την EuroLeague Ανδρών, η διοργάνωση διοργανώνεται εξ ολοκλήρου από την FIBA Europe. Ξεκίνησε το 1958 ως European Cup for Women's champion clubs. Η αρχική οργάνωση αποτελούνταν από 10 ομάδες και προέκυψε μετά από την επιτυχία που είχε ένα αντίστοιχο τουρνουά για ανδρικούς συλλόγους που διοργανώθηκε στις αρχές τις ίδια χρονιάς.

2.1.2 Αλλαγές που υπέστη η διοργάνωση

Κατά την δεκαετία του 1990 η οργάνωση υπέστη δύο βασικές αλλαγές. Η πρώτη έγινε το 1992 και αφορά την εισαγωγή του final 4 όπου παρέμεινε μέχρι το 2011 στο Εκατερίνμπουργκ. Από την επόμενη σεζόν και μετά η διοργάνωση ακολούθησε μια νέα κατεύθυνση αφού κατήργησε το final 4 και το αντικατέστησε με ένα τουρνουά Final 8. Η δεύτερη αλλαγή έγινε το 1996 κι αφορά το όνομα της διοργάνωσης που άλλαξε από Ευρωπαϊκό Κύπελλο Πρωταθλητριών (European Cup for Women's champion clubs) σε αυτό που είναι γνωστό έως και σήμερα : EuroLeague Γυναικών.

2.1.3 Ιστορική αναδρομή στα διαφορετικά format της διοργάνωσης στο πέρασμα των χρόνων.

Όπως αναφέρθηκε και παραπάνω η οργάνωση ξεκίνησε το 1958 Στο αρχικό τουρνουά, η Σλάβια Σόφιας της Βουλγαρίας στέφθηκε πρωταθλήτρια, νικώντας τη Σοβιετική Ντιναμο Μόσχας με 64–40 εντός έδρας και στη συνέχεια με 44–34 εντός γηπέδου στην Μόσχα.

Η μορφή των δύο εντός και εκτός έδρας αγώνων για τον τελικό παρέμεινε μέχρι το 1976, πριν αλλάξει σε μορφή ενός αγώνα την επόμενη χρονιά. Κατά την διάρκεια όλων αυτών των χρόνων έχουν συμμετέχει στην διοργάνωση ομάδες από διάφορες χώρες.

Αξίζει φυσικά να σημειωθεί ότι μετά την εισβολή της Ρωσίας στην Ουκρανία τον Φεβρουάριο του 2022, οι ρωσικές ομάδες αποβλήθηκαν από το τουρνουά από την FIBA για τα play offs . Η EuroLeague Γυναικών έθεσε σε αποκλεισμό τις ρωσικές ομάδες UMMC Ekaterinburg , Dynamo Kursk και MBA Moscow.

- **Σεζόν 2004-2011**

Στην οργάνωση συμμετέχουν 24 σύλλογοι οι οποίοι χωρίστηκαν σε τέσσερις ομίλους αποτελούμενους από 6 ομάδες ο καθένας. Κάθε όμιλος είχε εντός και εκτός έδρας αγώνες. Οι τέσσερις ομάδες που κατέλαβαν την καλύτερη θέση σε κάθε όμιλο προκρίθηκαν στα play-off. Το Final Eight καθορίστηκε σύμφωνα με την κατάταξη κάθε ομάδας στον προκριματικό γύρο. Αυτός ο γύρος διεξήχθη σε εντός και εκτός έδρας αγώνα. Οι νικητές του προκρίθηκαν στα προημιτελικά. Οι νικητές του προημιτελικού γύρου προκρίθηκαν στο Final Four, το οποίο διοργανώθηκε από μία από τις ομάδες που προκρίθηκαν.

- **Σεζόν 2011–2014**

Οι ομάδες χωρίστηκαν σε τρεις ομίλους, οι οποίοι έπαιζαν μεταξύ τους εντός και εκτός έδρας. Η καλύτερη ομάδα από κάθε όμιλο προκρινόταν απευθείας στους Final Eight, ενώ οι επόμενες 14 καλύτερες ομάδες προκρίνονταν σε έναν γύρο play-off . Το Final Eight διεξήχθη σε δύο ομίλους σε έναν ενιαίο χώρο, με τις καλύτερες ομάδες να προκρίνονται στο Final Four.

- **Σεζόν 2014-σημερα**

Οι ομάδες χωρίζονται σε δύο ομίλους, ο καθένας με εντός και εκτός έδρας αγώνες. Οι τέσσερις πρώτες ομάδες από κάθε όμιλο προκρίνονται στα προημιτελικά, τα οποία παίζονται σε τρία παιχνίδια, και οι νικητές προκρίνονται στο Final Four.

2.2 Η EuroLeague Γυναικών στη σύγχρονη εποχή και ο κοινωνικός-αθλητικός της αντίκτυπος στην Ευρώπη

2.2.1 Επιρροή κι αντίκτυπος στην Ευρώπη

Κατά τη σύγχρονη εποχή, και ιδιαίτερα από τα τέλη της δεκαετίας του 2010 έως σήμερα, η EuroLeague Γυναικών έχει εξελιχθεί σε μία από τις πιο δυναμικά αναπτυσσόμενες αθλητικές διοργανώσεις γυναικείου αθλητισμού στην Ευρώπη. Η διοργάνωση δεν περιορίζεται πλέον σε ένα στενό αγωνιστικό πλαίσιο, αλλά λειτουργεί ως θεσμός με αυξανόμενη κοινωνική, πολιτισμική και αθλητική επιρροή. Σύμφωνα με επίσημα στοιχεία της FIBA, η EuroLeague Γυναικών καταγράφει σταθερά ρεκόρ ψηφιακής ανάπτυξης, με σημαντική αύξηση στην παρακολούθηση αγώνων μέσω διαδικτύου, καθώς και στην αλληλεπίδραση του κοινού στα μέσα κοινωνικής δικτύωσης.

Παράλληλα, έχει αναγνωριστεί ως μία από τις ταχύτερα αναπτυσσόμενες αθλητικές «ιδιότητες» διεθνώς, στοιχείο που αναδεικνύει τη μετατόπιση της ευρωπαϊκής αθλητικής κουλτούρας προς μεγαλύτερη αποδοχή και προβολή του γυναικείου επαγγελματικού αθλητισμού. Σημαντικό ρόλο σε αυτή τη μετάβαση διαδραματίζει η στρατηγική αξιοποίηση των ψηφιακών μέσων και των πλατφόρμων streaming, όπως η

συνεργασία της διοργάνωσης με το DAZN (είναι διεθνές πλατφόρμα αθλητικού streaming σαν «Netflix των σπορ»).

Μεταδίδει ζωντανά και on-demand αθλητικά γεγονότα μέσω ίντερνετ), η οποία ενίσχυσε την προσβασιμότητα των αγώνων και επέκτεινε τον κοινωνικό της αντίκτυπο πέρα από τα εθνικά σύνορα. Συνολικά, η EuroLeague Γυναικών αποτελεί πλέον όχι μόνο κορυφαίο αθλητικό γεγονός, αλλά και φορέα κοινωνικής αλλαγής, συμβάλλοντας ενεργά στη θεσμική αναγνώριση, την ορατότητα και την ανάπτυξη του γυναικείου αθλητισμού στην Ευρώπη.

2.2.2 Κοινωνική διάσταση και σύγκριση με την EuroLeague Ανδρών.

Φυσικά δεν θα μπορούσαμε να αποφύγουμε την σύγκριση με την αντίστοιχη διοργάνωση EuroLeague Ανδρών. Σε κοινωνικό επίπεδο, η EuroLeague Γυναικών παρουσιάζει ουσιώδεις διαφορές σε σχέση με την EuroLeague Ανδρών, οι οποίες αντανakλούν ευρύτερες ανισότητες στον επαγγελματικό αθλητισμό. Παρότι και οι δύο διοργανώσεις εκπροσωπούν το υψηλότερο αγωνιστικό επίπεδο στο ευρωπαϊκό μπάσκετ, η EuroLeague Ανδρών απολαμβάνει ιστορικά μεγαλύτερη θεσμική προβολή, εμπορική αξία και παρουσία στα παραδοσιακά μέσα ενημέρωσης. Αντίθετα, η EuroLeague Γυναικών ιδιαίτερα στη σύγχρονη εποχή, αναπτύσσεται μέσα από διαφορετικά κανάλια, με έμφαση στην ψηφιακή επικοινωνία, στα μέσα κοινωνικής δικτύωσης και στην άμεση σύνδεση με το κοινό. Ωστόσο, αυτή η διαφοροποίηση έχει και κοινωνική σημασία. Η EuroLeague Γυναικών λειτουργεί ως πλατφόρμα ενδυνάμωσης της γυναικείας παρουσίας στον επαγγελματικό αθλητισμό, προβάλλοντας πρότυπα αθλητριών σε ένα πεδίο που παραδοσιακά κυριαρχείται από ανδρικά πρότυπα. Σύμφωνα με τη FIBA, η συνεχής αύξηση της απήχησης της διοργάνωσης συνδέεται άμεσα με την αλλαγή κοινωνικών αντιλήψεων γύρω από το φύλο, τον αθλητισμό και την ισότητα ευκαιριών. Σε αντίθεση με την EuroLeague Ανδρών, η οποία έχει εδραιωθεί εδώ και δεκαετίες ως εμπορικό προϊόν, η EuroLeague Γυναικών βρίσκεται σε φάση διαμόρφωσης ταυτότητας, όπου ο κοινωνικός της ρόλος είναι εξίσου σημαντικός με τον αγωνιστικό. Η σύγχρονη πορεία της αναδεικνύει τον γυναικείο αθλητισμό όχι ως «παράγωγο» του ανδρικού, αλλά ως αυτόνομο και εξελισσόμενο θεσμό με ιδιαίτερη κοινωνική βαρύτητα στην ευρωπαϊκή αθλητική πραγματικότητα. Η εξέλιξη της καταδεικνύει ότι το γυναικείο ευρωπαϊκό μπάσκετ έχει εισέλθει σε μια νέα φάση θεσμικής ωρίμανσης και κοινωνικής αναγνώρισης. Η διοργάνωση δεν αποτελεί πλέον απλώς ένα αγωνιστικό πλαίσιο υψηλού επιπέδου, αλλά έναν πολυδιάστατο θεσμό που επηρεάζει τόσο την αθλητική όσο και την κοινωνική πραγματικότητα στην Ευρώπη.

2.3 Σημαντικά γεγονότα και milestones της EuroLeague Γυναικών

Κατά τη διάρκεια της ιστορίας της, η EuroLeague Γυναικών έχει σημειώσει σειρά από σταθμούς-ορόσημα που διαμόρφωσαν το επίπεδο και την προβολή του γυναικείου μπάσκετ στην Ευρώπη. Η ίδρυση της διοργάνωσης στα τέλη της δεκαετίας του 1950 αποτέλεσε το πρώτο βήμα για τη συγκρότηση μιας πανευρωπαϊκής πλατφόρμας αγώνων γυναικείου μπάσκετ, που ένωσε συλλόγους από διαφορετικές χώρες με κοινό στόχο την ανάδειξη ταλέντων και την ανάπτυξη του αθλήματος. Στις επόμενες δεκαετίες, η διοργάνωση παρουσίασε σημαντική εξέλιξη: η δεκαετία του 1990 σηματοδότησε την πλήρη επαγγελματικοποίηση των ομάδων, με την εισαγωγή συστημάτων αδειοδότησης, την ανάπτυξη οργανωτικής δομής και την πρώτη

τηλεοπτική προβολή των αγώνων. Παράλληλα, αναδείχθηκαν παίκτριες που έγιναν διεθνώς γνωστές και συνέβαλαν στην αύξηση του ενδιαφέροντος του κοινού για το γυναικείο μπάσκετ.

Η περίοδος 2000–2010 χαρακτηρίστηκε από διεθνοποίηση και αύξηση του ανταγωνισμού. Οι ομάδες άρχισαν να εντάσσουν ξένες παίκτριες υψηλού επιπέδου, ενώ η EuroLeague Γυναικών ενίσχυσε τη συνεργασία της με την WNBA (Women's National Basketball Association), δημιουργώντας ένα διεθνές οικοσύστημα για κορυφαίες αθλήτριες. Επιπλέον, σημαντικές καινοτομίες στους κανονισμούς και τη δομή της διοργάνωσης βελτίωσαν την ποιότητα του παιχνιδιού και ενίσχυσαν την προβολή της σε όλη την Ευρώπη.

2.4 Ελληνική συμμετοχή

Στην ιστορία της διοργάνωσης, η ελληνική συμμετοχή είναι σχετικά περιορισμένη σε σχέση με τα μεγαλύτερα ευρωπαϊκά καλαθοσφαιρικά κέντρα, αλλά με ιδιαίτερο συμβολισμό για το γυναικείο μπάσκετ της χώρας. Η Sporting Athens ήταν η πρώτη ομάδα που συμμετείχε στην διοργάνωση κατά την περίοδο της αγωνιστικής της ακμής, εκπροσωπώντας το ελληνικό μπάσκετ στο υψηλότερο ευρωπαϊκό επίπεδο. Κορυφαία στιγμή αποτέλεσε η πρόκρισή της στο Final Four τις σεζόν 1990–1991 και 1991–1992, επίτευγμα που τη συγκαταλέγει στις σημαντικότερες ελληνικές παρουσίες στη διοργάνωση. Έπειτα από αυτό ο Ολυμπιακός αποτελεί τον πιο σταθερό εκπρόσωπο της Ελλάδας στη διοργάνωση, εξασφαλίζοντας συμμετοχή απευθείας στους ομίλους της περιόδου 2025–2026, όπου ανταγωνίζεται ομάδες όπως η Fenerbahçe, η USK Πράγας και η Spar Girona. Η παρουσία έστω και μιας ελληνικής ομάδας στη διοργάνωση έχει ιδιαίτερη σημασία για το ελληνικό γυναικείο μπάσκετ, καθώς αποτελεί τη μοναδική ελληνική εκπροσώπηση στην κορυφαία ευρωπαϊκή διοργάνωση συλλόγων. Μέσω αυτής της συμμετοχής, οι αθλήτριες αποκτούν την ευκαιρία να αγωνιστούν σε υψηλό επίπεδο, να βελτιώσουν τις δεξιότητές τους και να συγκριθούν με τις κορυφαίες παίκτριες της Ευρώπης. Παράλληλα, η διεθνής προβολή της ομάδας λειτουργεί ως πρότυπο και κίνητρο για νεότερες αθλήτριες στην Ελλάδα, ενισχύοντας τη συμμετοχή των κοριτσιών στον αθλητισμό και προωθώντας την ισότητα των φύλων.

Η παρουσία της ελληνικής ομάδας δημιουργεί επίσης προοπτικές για διεθνείς συνεργασίες και μεταγραφές, ενισχύοντας συνολικά την ανταγωνιστικότητα του ελληνικού πρωταθλήματος. Έτσι, αποτελεί και ως σημείο αναφοράς για την ανάπτυξη και την προβολή του γυναικείου μπάσκετ στην Ελλάδα.

2.5 Ξεχωριστές παίκτριες και διεθνής αναγνώριση

Στην ιστορία της EuroLeague Γυναικών έχουν αγωνιστεί και ξεχωρίσει πολλές κορυφαίες παίκτριες που άφησαν ανεξίτηλο αποτύπωμα στην κορυφαία ευρωπαϊκή διοργάνωση. Στην κορυφή των σκόρερ όλων των εποχών, η Ann Wauters κατέχει την πρώτη θέση με περισσότερους από 3.700 πόντους, αποτελώντας μια από τις πιο σταθερές και παραγωγικές αθλήτριες στην ιστορία της διοργάνωσης, ενώ έχει κατακτήσει πολλαπλούς τίτλους με ομάδες όπως η US Valenciennes Olympic και η UMMC Ekaterinburg. Μια άλλη εμβληματική φιγούρα είναι η Diana Taurasi, γνωστή για την ικανότητά της να καθορίζει κρίσιμες στιγμές σε τελικούς και playoffs, και αναγνωρισμένη γενικά ως μία από τις κορυφαίες παίκτριες γενικά στο γυναικείο μπάσκετ. Μεταξύ των μεγάλων ονομάτων είναι επίσης η Sandrine Gruda, μια σταθερή

επιλογή στους μεγάλους ευρωπαϊκούς συλλόγους, και η Maria Stepanova, η οποία ξεχώριζε για τη σωματική της παρουσία και ηγετική συμβολή στο παιχνίδι.

Στους πιο πρόσφατους τίτλους και διακρίσεις MVP της EuroLeague, η Emma Meesseman έχει κυριαρχήσει με τον σύλλογο Fenerbahçe, κατακτώντας τον τίτλο σε πολλαπλές σεζόν, συμπεριλαμβανομένου του 2022–2023 και 2023–2024 όπου αναδείχθηκε MVP και μέλος της All-EuroLeague First Team. Αυτές οι αθλήτριες δεν αποτελούν μόνο στατιστικά υψηλές περιπτώσεις· έχουν διαμορφώσει την εξέλιξη και την εικόνα της EuroLeague, καθιστώντας τη διοργάνωση έναν από τους σημαντικότερους θεσμούς στον διεθνή γυναικείο αθλητισμό. Παρακάτω παρουσιάζεται ο πίνακας 2.4 που συγκεντρώνει τις 20 κορυφαίες σκόρερ όλων των εποχών στην EuroLeague Γυναικών, παρουσιάζοντας τον συνολικό αριθμό πόντων που σημείωσε η κάθε παίκτρια στην καριέρα της στη διοργάνωση. (Πηγή: FIBA EuroLeague Women – All-Time Leaders, επεξεργασία στοιχείων 2026)

Player	Points
Ann Wauters	3771
Sandrine Gruda	3393
Diana Taurasi	3200
Jurgita Streimikyte-Virbickiene	2802
Delisha Milton	2699
Maria Stepanova	2528
Agnieszka Bibrzycka	2503
Eva Viteckova	2490
Zuzana Zirkova	2405
Jelena Brooks	2366
Dewanna Bonner	2346
Gordana Kovacevic	2306
Amaya Valdemoro	2296
Angel McCoughtry	2239
Malgorzata Dydek	2233
Allie Quigley	2183
Endy Miyem	2139
Ivana Matovic	2125
Sonja Vasic	2119
Laura Macchi	2109

Πίνακας 2.4. Κορυφαίες σκόρερ όλων των εποχών

2.6 Βασικοί Δείκτες Απόδοσης: Βιβλιογραφική Ανασκόπηση

Στη σύγχρονη βιβλιογραφία της ανάλυσης απόδοσης στο μπάσκετ γυναικών, ιδιαίτερη έμφαση δίνεται στον εντοπισμό των αγωνιστικών δεικτών (game-related statistics) που σχετίζονται με την επιτυχία των ομάδων, μέσω της επεξεργασίας στατιστικών δεδομένων από αγώνες υψηλού επιπέδου. Οι έρευνες αυτές επιδιώκουν να προσδιορίσουν τα πρότυπα απόδοσης που διαφοροποιούν τις νικήτριες από τις ηττημένες ομάδες, λαμβάνοντας υπόψη τόσο τη συνολική αγωνιστική συμπεριφορά όσο και τις ειδικές συνθήκες του κάθε αγώνα. Οι Gómez, Lorenzo, Sampaio και Ibáñez, στο άρθρο τους “Differences between winning and losing teams in women’s basketball game-related statistics” (2006), ανέλυσαν 178 αγώνες της ισπανικής γυναικείας λίγκας και κατέδειξαν ότι οι νικήτριες ομάδες υπερέχουν κυρίως σε δείκτες όπως η ευστοχία στα δίποντα σουτ, τα αμυντικά ριμπάουντ και οι ασίστ. Ιδιαίτερα τα αμυντικά ριμπάουντ αναδείχθηκαν ως ο πιο σταθερός και καθοριστικός δείκτης, καθώς συμβάλλουν στον έλεγχο της κατοχής και στον περιορισμό των επιθετικών ευκαιριών του αντιπάλου. Επιπλέον, οι συγγραφείς διαχωρίζουν τους αγώνες σε ισορροπημένους (με διαφορά σκορ έως 12 πόντους) και μη ισορροπημένους, δείχνοντας ότι στους πρώτους η έκβαση εξαρτάται κυρίως από την ευστοχία σε ελεύθερες βολές και τρίποντα, καθώς και από την ομαδική συνεργασία, ενώ στους δεύτερους η υπεροχή των νικητών αποτυπώνεται περισσότερο στην αποτελεσματικότητα στα δίποντα και στην αμυντική πίεση (π.χ. κλεψίματα).

Αντίστοιχα, οι Gasperi, Gómez, Lago-Peñas και Sampaio, στο άρθρο τους “Game Related Statistics Discriminate National and Foreign Players According to Playing Position and Team Ability in the Women’s Basketball EuroLeague” (2020), ανέλυσαν δεδομένα από αγώνες της Ευρωλίγκας γυναικών την χρονιά 2016-2017 και, πέρα από τη σύγκριση παικτριών, κατέληξαν σε σημαντικά συμπεράσματα σχετικά με τη συνολική απόδοση των ομάδων. Συγκεκριμένα, διαπιστώθηκε ότι οι ομάδες υψηλότερης δυναμικότητας παρουσιάζουν μεγαλύτερη αγωνιστική αποτελεσματικότητα, η οποία εκφράζεται μέσω καλύτερων ποσοστών ευστοχίας, περισσότερων ασίστ και αποτελεσματικότερης διαχείρισης των κατοχών, δηλαδή λιγότερα λάθη και καλύτερο έλεγχο του ρυθμού του παιχνιδιού. Παράλληλα, τονίζεται ότι οι δείκτες απόδοσης δεν λειτουργούν μεμονωμένα αλλά συνδυαστικά, καθώς η επιτυχία προκύπτει από την ισορροπία μεταξύ επιθετικής αποτελεσματικότητας και αμυντικής σταθερότητας. Συνολικά, τα ευρήματα των παραπάνω μελετών συγκλίνουν στο ότι η επιτυχία στο μπάσκετ γυναικών υψηλού επιπέδου καθορίζεται από έναν συνδυασμό βασικών δεικτών απόδοσης, όπως τα αμυντικά ριμπάουντ, η ευστοχία στα σουτ και οι ασίστ, οι οποίοι αντικατοπτρίζουν τον έλεγχο της κατοχής, την ποιότητα των επιθετικών επιλογών και την αποτελεσματικότητα της ομαδικής συνεργασίας. Επιπλέον, η σημασία των επιμέρους δεικτών διαφοροποιείται ανάλογα με τις συνθήκες του αγώνα, γεγονός που αναδεικνύει την ανάγκη για πολυπαραγοντική ανάλυση της απόδοσης σε διοργανώσεις υψηλού επιπέδου.

Κεφάλαιο 3: Περιγραφική ανάλυση

Σε αυτό το κεφάλαιο παρουσιάζεται η περιγραφική στατιστική ανάλυση των δεδομένων. Αρχικά παρουσιάζονται συνοπτικά τα δεδομένα και υπολογίζονται τα βασικά περιγραφικά μέτρα. Τα αποτελέσματα αυτά απεικονίζονται γραφικά μέσω διαγραμμάτων time series plots, τα οποία επιτρέπουν την οπτική διερεύνηση της εξέλιξης των δεικτών στις αγωνιστικές περιόδους. Στη συνέχεια της ανάλυσης δημιουργούνται διαγράμματα διασποράς (scatter plots) μεταξύ του δείκτη αποδοτικότητας EFFPG και των υπόλοιπων αγωνιστικών μεταβλητών, με σκοπό τη διερεύνηση της μορφής, της κατεύθυνσης και της έντασης της μεταξύ τους σχέσης. Παράλληλα παρουσιάζονται διαγράμματα Chernoff Faces, τα οποία παρέχουν μια πολυδιάστατη γραφική απεικόνιση των στατιστικών μεταβλητών. Τέλος, οι ομάδες κατηγοριοποιούνται βάσει της αγωνιστικής τους πορείας στη διοργάνωση σε τρία επίπεδα επιτυχίας: ομάδες που δεν προκρίθηκαν στο Final 8 (No), ομάδες που προκρίθηκαν στο Final 8 αλλά δεν έφτασαν στο Final 4 (Final8) και ομάδες που κατέλαβαν θέση στην τελική τετράδα (Final4). Για κάθε κατηγορία παρουσιάζονται τα αντίστοιχα περιγραφικά στατιστικά μέτρα, ενώ οι διαφορές μεταξύ των ομάδων απεικονίζονται γραφικά μέσω boxplots και line plots. Οι πληροφορίες που αξιοποιήθηκαν για το συγκεκριμένο κεφάλαιο αντλήθηκαν κυρίως από την επίσημη ιστοσελίδα της FIBA(<https://www.fiba.basketball>) ,από τη διαδικτυακή εγκυκλοπαίδεια Wikipedia (<https://en.wikipedia.org>) καθώς και από το βιβλίο Εφαρμοσμένη Στατιστική και Στατιστική Μηχανική Μάθηση με χρήση των IBM SPSS Statistics, R, Python (Μπερσίμης κ.α., 2021)

3.1 Παρουσίαση δεδομένων

Παρακάτω στον Πίνακα 3.1.1 παρουσιάζονται αναλυτικά τα δεδομένα που χρησιμοποιήθηκαν για την παρούσα διπλωματική τα οποία αντλήθηκαν από την επίσημη ιστοσελίδα της FIBA(<https://www.fiba.basketball>). Τα δεδομένα αναφέρονται στους μέσους όρους των ομάδων για την κάθε σεζόν. Το σύνολο των δεδομένων περιέχει 22 σεζόν και συγκεκριμένα περιλαμβάνουν τα στατιστικά των ομάδων ξεχωριστά για κάθε σεζόν από το 2004-2025. Κατά την διάρκεια αυτών των χρόνων έχουν σημειωθεί περιπτώσεις αλλαγής ονομασίας ομάδων λόγω χορηγών ή μεταβολών στη διοίκηση. Για λόγους συνέπειας και διευκόλυνσης της ανάλυσης, όλες οι ομάδες καταγράφονται με την πιο πρόσφατη ονομασία τους κατά την τελευταία διαθέσιμη σεζόν. Αυτό επιτρέπει τη σαφή αναφορά και τη σύγκριση των δεδομένων παρά τις αλλαγές ονομασίας στο ιστορικό των συλλόγων. Από το 2004 έως το 2025 έχουν συμμετάσχει έστω και μια φορά 108 διαφορετικές ομάδες. Αναλυτικά τα ονόματα των ομάδων μαζί με τον κωδικό τους παρουσιάζονται στο παράρτημα στον Πίνακα 3.1.

Στην ανάλυση που θα ακολουθήσει χρησιμοποιήθηκαν οι παρακάτω μεταβλητές.

Μεταβλητή	Επεξήγηση
TEAM	Όνομα ομάδας
TEAM ID	Κωδικός ομάδας
SEASON	Κωδικός για τη κάθε σεζόν (1:2004, 2:2005 κ.ο.κ)
QUARTER FINAL	Ένδειξη για το αν η ομάδα προκρίθηκε στα playoffs (1:ναι , 0:όχι)
FINAL 4	Ένδειξη για το αν η ομάδα προκρίθηκε στο Final Four (1:ναι ,0:όχι)
FINAL 4 POSITION	Τελική κατάταξη στο Final Four
PPG	Μέσος όρος πόντων ανά παιχνίδι
2P%	Μέσος όρος δίποντων ανά παιχνίδι
3P%	Μέσος όρος τρίποντων ανά παιχνίδι
FT%	Μέσος όρος ελευθέρων βολών ανά παιχνίδι
ORPG	Μέσος όρος επιθετικών ριμπάουντ ανά παιχνίδι
DRPG	Μέσος όρος αμυντικών ριμπάουντ ανά παιχνίδι
RPG	Μέσος όρος ριμπάουντ ανά παιχνίδι ανά παιχνίδι
APG	Μέσος όρος “Ασίστ” (Τελικές πάσες που οδήγησαν σε καλάθι) ανά παιχνίδι
STLPG	Κλεψίματα ανά παιχνίδι
TOPG	Λάθη ανά παιχνίδι
BLKPG	Κοψίματα ανά παιχνίδι
PFPG	Φάουλ που έκανε η ομάδα ανά παιχνίδι
EFFPG	Δείκτης αξιολόγησης της απόδοσης

Πίνακας 3.1.1 Οι μεταβλητές

Στην EuroLeague Γυναικών όπως και στις περισσότερες διοργανώσεις της FIBA το Efficiency (EFF) ή Performance Index Rating (PIR) για όλη την ομάδα υπολογίζεται αθροίζοντας τις θετικές ενέργειες και αφαιρώντας τις αρνητικές ενέργειες όλων των παικτριών που αγωνίστηκαν στον αγώνα. Ο βασικός τύπος για τον υπολογισμό του EFF της ομάδας είναι:

$$\text{EFF Ομάδας} = (\text{Σύνολο Θετικών Ενεργειών}) - (\text{Σύνολο Αρνητικών Ενεργειών})$$

Αναλυτικότερα ο τύπος περιλαμβάνει τα εξής στοιχεία:

Θετικές ενέργειες (+) : Πόντοι + Συνολικά Ριμπάουντ + Ασίστ + Κλεψίματα + Κοψίματα + Κερδισμένα Φάουλ.

Αρνητικές ενέργειες (-) : Άστοχα Σουτ + Χαμένες Βολές + Λάθη + Φάουλ

3.2 Πλαίσιο και Σκοπός Ανάλυσης

Η παρούσα εργασία αποσκοπεί στη διερεύνηση των βασικών παραγόντων που επηρεάζουν την έκβαση ενός αγώνα στην EuroLeague Γυναικών, αξιοποιώντας δεδομένα από τα τελευταία 22 χρόνια της διοργάνωσης. Μέσα από τη στατιστική ανάλυση των επίσημων αγωνιστικών δεδομένων, επιχειρείται η κατανόηση της αγωνιστικής φυσιογνωμίας του ευρωπαϊκού μπάσκετ γυναικών καθώς και η ανάδειξη των μεταβολών που έχουν σημειωθεί διαχρονικά. Η εργασία στοχεύει στη δημιουργία ενός ολοκληρωμένου αναλυτικού πλαισίου που συνδέει τα αγωνιστικά στατιστικά με την πιθανότητα νίκης. Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά βασικά περιγραφικά στατιστικά μέτρα καθώς γίνεται και η γραφική απεικόνιση των μεταβλητών.

3.2.1 Διερεύνηση για ελλιπή δεδομένα (missing values)

Αρχικά, γίνεται έλεγχος σε κάθε ανάλυση για την ύπαρξη ελλείπων τιμών στα δεδομένα μας. Τα αποτελέσματα δείχνουν ότι υπάρχουν missing values (εμφανίζονται όταν δεν αποθηκεύεται καμία τιμή δεδομένων για τη μεταβλητή σε μια παρατήρηση.). Αυτό ήταν αναμενόμενο αφού η αγωνιστική περίοδος 2019–2020 (season 17) δεν ολοκληρώθηκε κανονικά λόγω της πανδημίας COVID-19, γεγονός που θα μπορούσε να επηρεάσει τα αποτελέσματα. Για αυτό τον λόγο από εδώ και πέρα η season 17 εξαιρέθηκε από το δείγμα της ανάλυσης.

3.3 Περιγραφική ανάλυση

Στην παρούσα παράγραφο παρουσιάζονται τα βασικά περιγραφικά στατιστικά μέτρα που αφορούν τις ποσοτικές μεταβλητές του δείγματος, στο πλαίσιο των αναλύσεων που πραγματοποιήθηκαν. Η έμφαση δίνεται σε μέτρα κεντρικής τάσης και διασποράς, τα οποία επιτρέπουν τη συνοπτική αποτύπωση των αριθμητικών χαρακτηριστικών των δεδομένων. Ως εκ τούτου, από τα αποτελέσματα της συγκεκριμένης ενότητας εξαιρούνται οι κατηγορικές μεταβλητές, καθώς η φύση τους δεν επιτρέπει την εφαρμογή των αντίστοιχων αριθμητικών περιγραφικών μέτρων. Επίσης προηγήθηκε μετατροπή των μεταβλητών σε αριθμητική μορφή, δεδομένου ότι κατά την αρχική εισαγωγή των δεδομένων είχαν αποθηκευτεί ως χαρακτήρες, γεγονός που δεν επέτρεπε την άμεση εφαρμογή αριθμητικών στατιστικών υπολογισμών.

3.3.1 Περιγραφικά στατιστικά μέτρα όλων των ομάδων για όλες τις σεζόν

Σε αυτήν την ενότητα, υπολογίσαμε τα περιγραφικά στατιστικά μέτρα όλων των ομάδων για όλες τις σεζόν. Στον πίνακα 3.3.1 παρουσιάζονται τα περιγραφικά στατιστικά των μεταβλητών της έρευνας. Συγκεκριμένα, ο Mean (Μέσος Όρος) εκφράζει τον αριθμητικό μέσο των τιμών κάθε μεταβλητής, ενώ η SD (Τυπική Απόκλιση) δείχνει τον βαθμό διασποράς των τιμών γύρω από τον μέσο όρο.

Η Median (Διάμεσος) αντιπροσωπεύει τη μεσαία τιμή της κατανομής, τα Q1 και Q3 αντιστοιχούν στο 1ο και 3ο τεταρτημόριο (25ο και 75ο εκατοστημόριο αντίστοιχα), ενώ τα Min και Max δηλώνουν τη μικρότερη και τη μεγαλύτερη παρατηρούμενη τιμή.

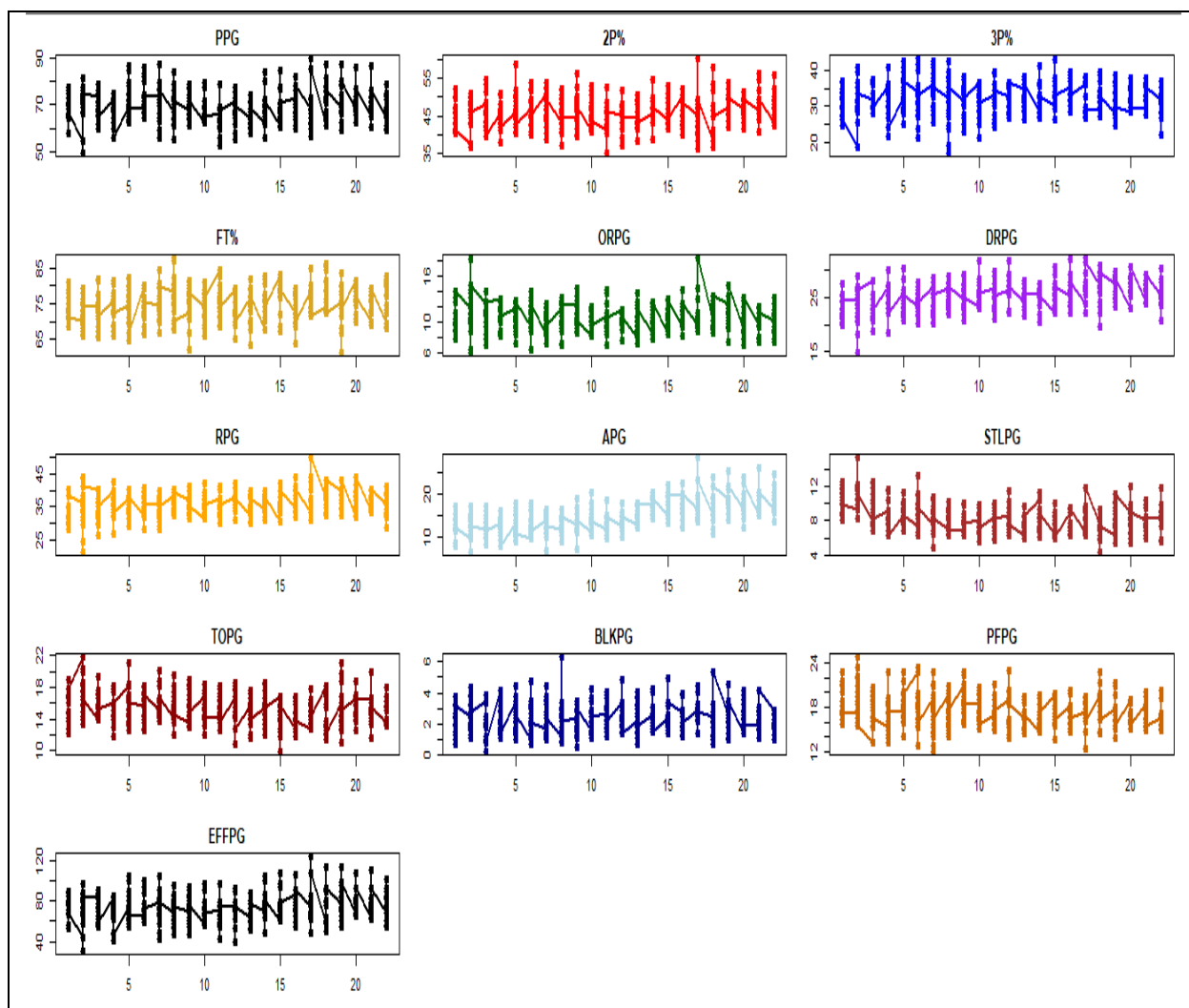
Από τα αποτελέσματα προκύπτει ότι οι περισσότερες μεταβλητές παρουσιάζουν σχετικά μικρή έως μέτρια διασπορά, καθώς οι διάμεσοι είναι κοντά στους μέσους όρους, γεγονός που υποδηλώνει σχετικά συμμετρικές κατανομές. Μεγαλύτερη μεταβλητότητα παρατηρείται στον δείκτη αποδοτικότητας (EFFPG), ενώ μικρότερη διακύμανση εμφανίζεται σε μεταβλητές όπως τα μπλοκ (BLKPG), υποδηλώνοντας μεγαλύτερη ομοιογένεια στη συγκεκριμένη αγωνιστική παράμετρο.

	Mean	SD	Median	Q1	Q3	Min	Max
PPG	69.21	6.27	68.8	65.1	72.7	49.1	87.5
2P%	46.11	3.84	46.1	43.5	48.6	35	58.8
3P%	32.3	4.18	32.5	29.7	35.2	16.9	43.2
FT%	74.89	4.63	75.2	71.8	78	61.3	87.8
ORPG	10.51	1.67	10.5	9.5	11.6	5.9	18.3
DRPG	25.04	2.49	25.2	23.25	26.55	14.8	32
RPG	35.54	3.2	35.5	33.4	37.65	21.2	43.9
APG	14.38	3.21	14	12.1	16.3	6.3	26.1
STLPG	8.43	1.63	8.3	7.3	9.4	4.3	15.4
TOPG	15.39	1.86	15.4	14.05	16.6	9.9	21.8
BLKPG	2.31	0.91	2.2	1.6	2.8	0.2	6.3
PFG	17.54	2.05	17.3	16	18.8	12	24.8
EFFPG	72.62	12.76	72.2	64.7	79.2	30.3	114

Πίνακας 3.3.1 Στατιστικά περιγραφικά μέτρα

Στο Σχήμα 3.3.1 παρουσιάζονται οι χρονοσειρές (time series plots) των βασικών αγωνιστικών δεικτών ανά σεζόν (SEASON). Τα διαγράμματα χρονοσειράς απεικονίζουν τη διαχρονική εξέλιξη των βασικών αγωνιστικών δεικτών ανά σεζόν, επιτρέποντας την παρατήρηση τάσεων, διακυμάνσεων και πιθανών ακραίων τιμών. Γενικά, ένα time series plot χρησιμοποιείται για την ανάλυση της συμπεριφοράς μιας μεταβλητής στον χρόνο και για τον εντοπισμό τυχόν αυξητικών ή μειωτικών τάσεων. Από τα παρακάτω γραφήματα προκύπτει ότι οι περισσότερες μεταβλητές παρουσιάζουν σχετική σταθερότητα, με ήπιες διακυμάνσεις γύρω από τον μέσο όρο και χωρίς έντονες διαρθρωτικές μεταβολές. Οι μεταβλητές 2P%, 3P%, FT% εμφανίζουν μικρό εύρος μεταβολής, γεγονός που υποδηλώνει διαχρονικά σταθερό επίπεδο αποτελεσματικότητας στο σκοράρισμα. Οι δείκτες ριμπάουντ (ORPG, DRPG, RPG) κινούνται επίσης σε σχετικά σταθερά επίπεδα, ενώ οι ασίστ (APG) παρουσιάζουν ελαφρώς μεγαλύτερη μεταβλητότητα, ενδεχομένως υποδηλώνοντας διαφοροποιήσεις μεταξύ των περιόδων. Ο δείκτης αποδοτικότητας (EFFPG) εμφανίζει μεγαλύτερο εύρος διακύμανσης, όπως αναμενόταν, καθώς επηρεάζεται από το σύνολο των επιμέρους αγωνιστικών επιδόσεων.

Συνολικά, τα αποτελέσματα υποδεικνύουν μια ομαλή διαχρονική πορεία των στατιστικών με φυσιολογικές αγωνιστικές διακυμάνσεις και χωρίς έντονες μακροχρόνιες τάσεις.

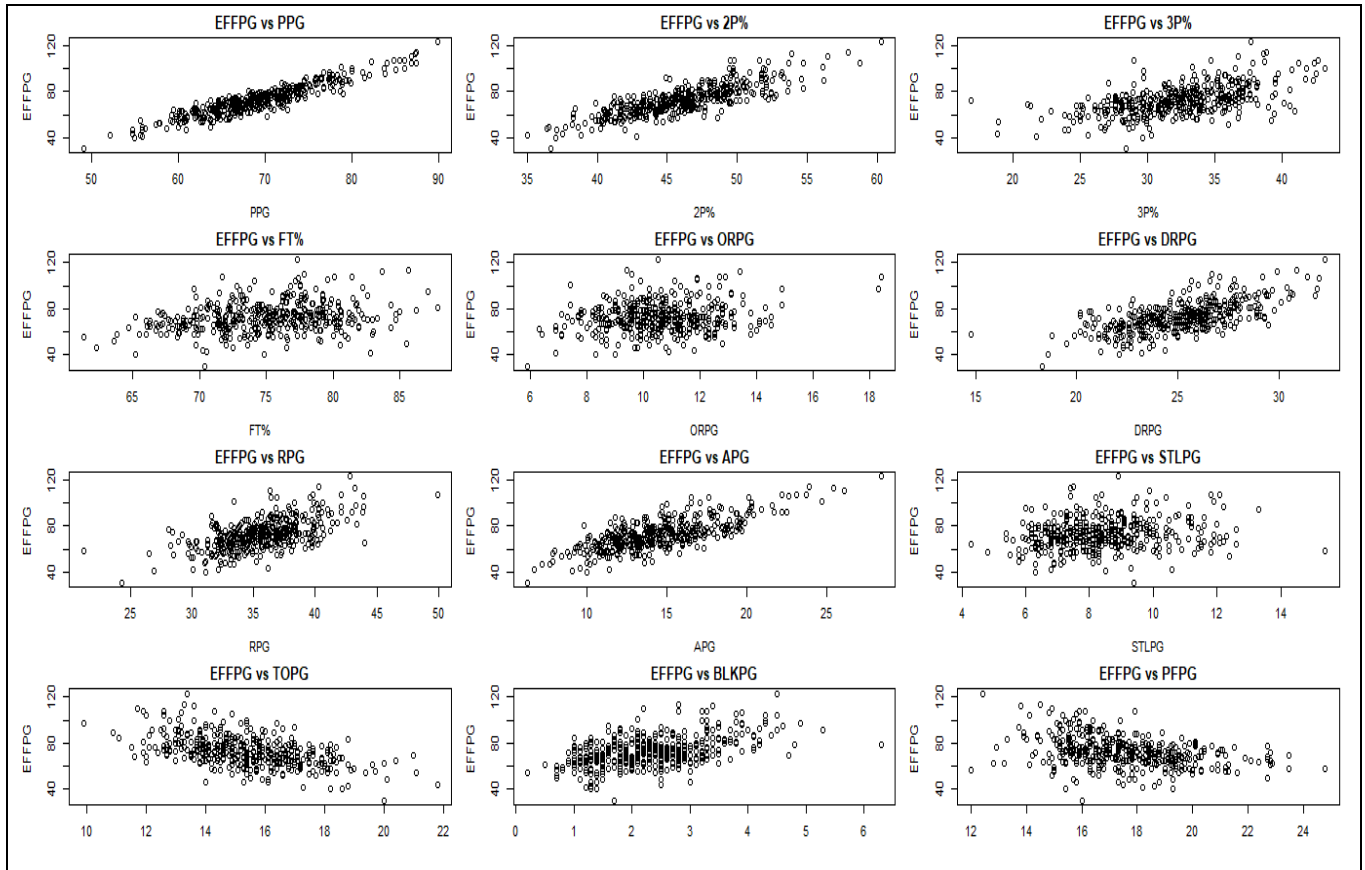


Σχήμα 3.3.1 Time Series Plots των στατιστικών στοιχείων με τη πάροδο των σεζόν

Στη συνέχεια δημιουργήθηκαν διαγράμματα διασποράς (scatter plots) μεταξύ του δείκτη αποδοτικότητας EFFPG και των λοιπών αγωνιστικών μεταβλητών, με στόχο τη διερεύνηση της μορφής, της κατεύθυνσης και της έντασης της μεταξύ τους σχέσης. Μέσω της οπτικής επιθεώρησης των διαγραμμάτων καθίσταται δυνατή η αξιολόγηση της ύπαρξης γραμμικής τάσης και του βαθμού συγκέντρωσης των παρατηρήσεων. Από τα διαγράμματα προκύπτει ιδιαίτερα έντονη θετική γραμμική τάση της EFFPG με τις μεταβλητές PPG, 2P%, APG και DRPG, καθώς τα σημεία διατάσσονται κατά μήκος ανοδικής κατεύθυνσης με σχετικά περιορισμένη διασπορά. Το εύρημα αυτό είναι αναμενόμενο, δεδομένου ότι οι συγκεκριμένοι δείκτες αποτυπώνουν βασικές πτυχές της αγωνιστικής συμβολής ενός παίκτη. Μέτρια θετική τάση παρατηρείται στις μεταβλητές 3P%, RPG, BLKPG, όπου η ανοδική διάταξη των σημείων είναι εμφανής, αν και συνοδεύεται από μεγαλύτερη μεταβλητότητα.

Αντιθέτως, στις μεταβλητές FT%, STLPG και ORPG η τάση εμφανίζεται ασθενέστερη, με μεγαλύτερη διασπορά και λιγότερο σαφή γραμμική διάταξη. Αρνητική τάση παρατηρείται στη μεταβλητή TOPG, όπου οι υψηλότερες τιμές λαθών συνδέονται με χαμηλότερες τιμές αποδοτικότητας.

Παρόμοια αλλά ηπιότερη αρνητική σχέση διαφαίνεται και στα PFPG. Τα διαγράμματα υποδεικνύουν ότι ο δείκτης αποδοτικότητας συνδέεται κυρίως με μεταβλητές που αποτυπώνουν παραγωγικότητα και συνολική συμμετοχή στο παιχνίδι, ενώ επηρεάζεται αρνητικά από δείκτες που σχετίζονται με αγωνιστικά λάθη και παραβάσεις.



Σχήμα 3.3.2 Scatter Plots των στατιστικών στοιχείων με το δείκτη EFFPG

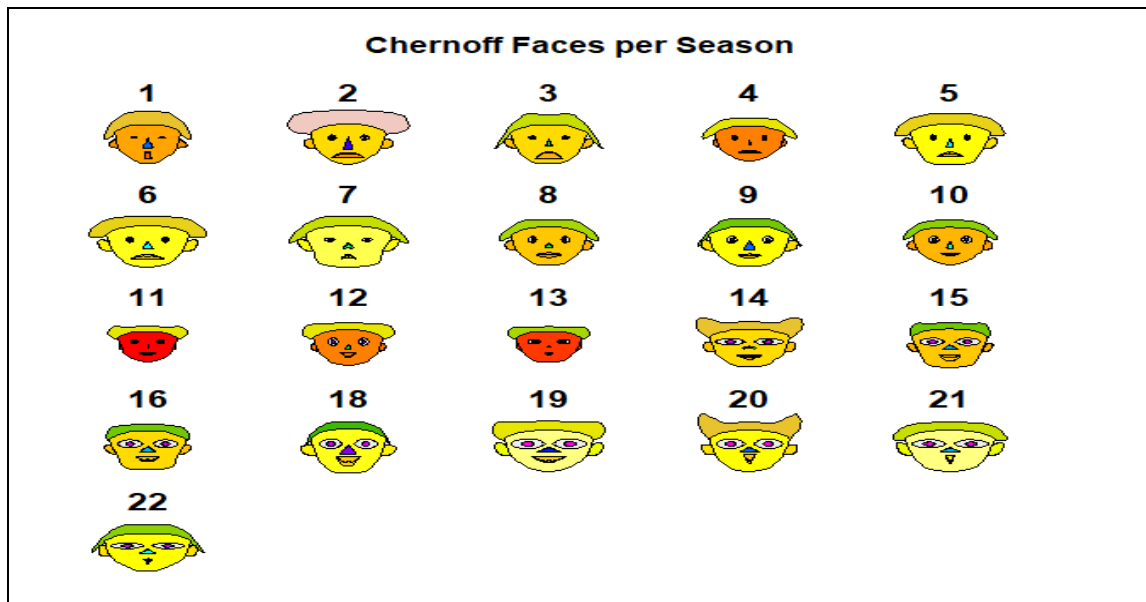
Οι παραπάνω παρατηρήσεις επιβεβαιώνονται και από τον υπολογισμό των συντελεστών συσχέτισης Pearson που φαίνονται στον πίνακα 3.3.2.b. Συγκεκριμένα, η ισχυρότερη θετική συσχέτιση με τον δείκτη αποδοτικότητας παρατηρείται στους πόντους ανά αγώνα (PPG, $r = 0.924$) και στο ποσοστό ευστοχίας στα δίποντα (2P%, $r = 0.812$). Σημαντική θετική συσχέτιση εμφανίζουν επίσης οι ασίστ ανά αγώνα (APG, $r = 0.732$) και τα αμυντικά ριμπάουντ (DRPG, $r = 0.651$), γεγονός που υποδηλώνει ότι οι δείκτες αυτοί συμβάλλουν ουσιαστικά στη συνολική αποδοτικότητα των ομάδων. Μέτρια θετική συσχέτιση παρατηρείται στα συνολικά ριμπάουντ (RPG, $r = 0.597$), στο ποσοστό ευστοχίας στα τρίποντα (3P%, $r = 0.568$) και στα κοψίματα (BLKPG, $r = 0.504$).

Αντίθετα, αρνητική συσχέτιση εμφανίζεται στα λάθη ανά αγώνα (TOPG, $r = -0.517$) και στα προσωπικά φάουλ (PFPG, $r = -0.346$), γεγονός που υποδηλώνει ότι η αύξηση των συγκεκριμένων δεικτών σχετίζεται με χαμηλότερα επίπεδα αποδοτικότητας. Οι υπόλοιπες μεταβλητές παρουσιάζουν ασθενέστερη σχέση με τον δείκτη EFPFG.

PPG	2P%	3P%	FT%	ORPG	DRPG	RPG	APG	STLPG	TOPG	BLKPG
0.924	0.812	0.568	0.273	0.171	0.651	0.597	0.732	0.135	-0.517	0.504
PFPG	EFPFG									
-0.346	1.000									

Πίνακας 3.3.2.b Συντελεστές συσχέτισης Pearson των μεταβλητών με το δείκτη EFPFG

Τέλος παρουσιάζονται στο Σχήμα 3.3.3 τα διαγράμματα Chernoff Faces. Τα Chernoff Faces αποτελούν μια μέθοδο πολυδιάστατης οπτικοποίησης δεδομένων, κατά την οποία κάθε παρατήρηση απεικονίζεται ως ένα ανθρώπινο πρόσωπο και κάθε μεταβλητή αντιστοιχίζεται σε ένα συγκεκριμένο μορφολογικό χαρακτηριστικό του προσώπου (π.χ. ύψος προσώπου, μέγεθος ματιών, χαμόγελο). Η βασική ιδέα της μεθόδου είναι ότι ο ανθρώπινος εγκέφαλος αναγνωρίζει εύκολα διαφορές σε πρόσωπα, επομένως μπορεί να εντοπίσει γρήγορα μοτίβα, ομοιότητες και μεταβολές σε πολυδιάστατα δεδομένα. Στην παρούσα ανάλυση, κάθε πρόσωπο αντιστοιχεί σε μία αγωνιστική περίοδο (season), ενώ τα χαρακτηριστικά του προσώπου αναπαριστούν ποσοτικούς δείκτες απόδοσης (PPG, ποσοστά ευστοχίας, ριμπάουντ, ασίστ, κλεψίματα, λάθη, τάπες, φάουλ και δείκτη αποτελεσματικότητας). Από τη σύγκριση των προσώπων διαπιστώνεται σταδιακή μεταβολή της μορφολογίας τους από τις πρώτες προς τις τελευταίες σεζόν, γεγονός που αντανακλά διαφοροποιήσεις στις αγωνιστικές επιδόσεις. Οι πρώτες σεζόν εμφανίζουν πιο «ουδέτερα» χαρακτηριστικά, υποδηλώνοντας σχετικά μέτρια επίπεδα σκοραρίσματος και συνολικής αποδοτικότητας. Στις μεσαίες σεζόν παρατηρείται εντονότερη διαφοροποίηση στα χαρακτηριστικά που συνδέονται με το σκοράρισμα και τη δημιουργία (ύψος προσώπου, μάτια, χαμόγελο), ένδειξη βελτίωσης επιθετικών και ομαδικών δεικτών. Στις τελευταίες σεζόν τα πρόσωπα παρουσιάζουν μεγαλύτερη συμμετρία και πιο έντονα χαρακτηριστικά που σχετίζονται με την αποτελεσματικότητα (πλάτος και ύψος μύτης – EFPFG), γεγονός που υποδηλώνει σταθεροποίηση και συνολικά υψηλότερο επίπεδο απόδοσης. Συνεπώς, το διάγραμμα Chernoff επιτρέπει την οπτική αποτύπωση της διαχρονικής εξέλιξης των στατιστικών επιδόσεων, καθιστώντας εμφανείς τις περιόδους βελτίωσης και σταθεροποίησης μέσω της συνολικής «έκφρασης» κάθε προσώπου.



Σχήμα 3.3.3 : Chernoff faces για κάθε σεζόν.

Στον παρακάτω πίνακα παρουσιάζεται το πώς το κάθε στατιστικό στοιχείο ανά σεζόν, επηρεάζει τη διαμόρφωση των προσώπων.

effect of variables:	
modified item	Var
"height of face	" PPG"
"width of face	" 2P%
"structure of face"	"3P%
"height of mouth	" FT%
"width of mouth	" ORPG"
"smiling	" DRPG"
"height of eyes	" RPPG"
"width of eyes	" APG"
"height of hair	" STLPG"
"width of hair	" TOPG"
"style of hair	" BLKPG"
"height of nose	" PFP%
"width of nose	" EFFP%
"width of ear	" PPG"
"height of ear	" 2P%

Πίνακας 3.3.3 Chernoff faces, effect of variables

3.3.2 Περιγραφικά στατιστικά μέτρα όλων των ομάδων βάσει της αγωνιστικής τους πορείας στη διοργάνωση

Στη συνέχεια παρουσιάζονται τα περιγραφικά στατιστικά μέτρα των μεταβλητών, με σκοπό την αποτύπωση της κατανομής και της κεντρικής τάσης των αγωνιστικών δεικτών ανά επίπεδο επιτυχίας. Συγκεκριμένα, αναλύονται ο μέσος όρος, η τυπική απόκλιση, η διάμεσος, τα τεταρτημόρια, καθώς και οι ελάχιστες και μέγιστες τιμές για κάθε στατιστική κατηγορία, προκειμένου να διαμορφωθεί μία σαφής εικόνα των διαφοροποιήσεων μεταξύ των ομάδων. Η κατηγοριοποίηση πραγματοποιήθηκε βάσει της αγωνιστικής τους πορείας στη διοργάνωση, διακρίνοντας τρία επίπεδα: ομάδες που δεν προκρίθηκαν στο final 8 (No), ομάδες που έφτασαν στο Final8 αλλά δεν προκρίθηκαν στο Final 4 (Final8) και ομάδες που κατέλαβαν θέση στην τελική τετράδα (Final4).

Η περιγραφική ανάλυση καταδεικνύει σαφή κλιμακωτή διαφοροποίηση των επιδόσεων μεταξύ των τριών κατηγοριών, με τις ομάδες Final4 να εμφανίζουν συστηματικά υψηλότερες τιμές στους περισσότερους θετικούς δείκτες απόδοσης. Παρατηρείται προοδευτική αύξηση της επιθετικής παραγωγικότητας (PPG: 66.56 – 70.33 – 75.13), της συνολικής αποδοτικότητας (EFFPG: 65.96 – 76.39 – 86.54), καθώς και των ποσοστών ευστοχίας στα δίποντα (2P%: 44.76 – 46.66 – 49.14) και τρίποντα (3P%: 30.95 – 33.10 – 35.08) από την κατηγορία No προς την Final4. Παράλληλα, καταγράφεται αύξηση στα αμυντικά και συνολικά ριμπάουντ (RPG: 34.28 – 36.44 – 37.99) και στις ασίστ (APG: 13.28 – 15.05 – 16.63), γεγονός που υποδηλώνει ενισχυμένη ομαδική συνοχή και αποτελεσματικότερο έλεγχο του παιχνιδιού. Αντιθέτως, τα λάθη ανά αγώνα (TOPG) παρουσιάζουν πτωτική τάση (16.09 – 14.65 – 14.26), στοιχείο που συνδέεται με βελτιωμένη διαχείριση στις επιτυχημένες ομάδες.

	Mean	SD	Median	Q1	Q3	Min	Max
PPG	66.56	5.17	66.6	63.75	69.8	49.1	87.3
2P%	44.76	3.55	44.9	42.4	46.8	35	54.7
3P%	30.95	3.98	31.2	28.45	33.55	18.8	40.9
FT%	74.8	4.95	75.3	71.4	78.4	62.3	87.8
ORPG	10.3	1.72	10.3	9.15	11.4	5.9	14.5
DRPG	23.99	2.18	24	22.6	25.5	14.8	29.9
RPG	34.28	2.95	34.1	32.7	35.95	21.2	43.2
APG	13.28	2.88	13.1	11.3	15	6.3	25.4
STLPG	8.27	1.71	8.1	7	9.15	4.3	15.4
TOPG	16.09	1.8	16.1	14.95	17.2	11.1	21.8
BLKPG	2	0.8	1.9	1.4	2.5	0.2	6.3
PFPG	18.06	2.1	18	16.6	19.4	12	24.8
EFFPG	65.96	9.78	66.6	60.45	72.45	30.3	112.6

Πίνακας 3.3.2.1 Στατιστικά περιγραφικά μέτρα για τις ομάδες που δεν προκρίθηκαν στο final 8

	Mean	SD	Median	Q1	Q3	Min	Max
PPG	70.33	5.11	70.15	66.68	72.65	59.1	86.9
2P%	46.66	3.16	46.25	44.6	48.95	41.2	58.8
3P%	33.1	3.84	33.35	30.92	36.05	16.9	42.5
FT%	74.18	4.63	73.95	71.5	77.35	61.3	87.1
ORPG	10.7	1.52	10.6	9.67	11.62	7.1	14.9
DRPG	25.75	1.83	25.6	24.87	26.88	20.6	29.8
RPG	36.44	2.38	36.7	35.1	37.8	30.5	43.4
APG	15.05	2.27	14.8	13.17	16.52	10.9	19.7
STLPG	8.56	1.42	8.4	7.57	9.43	6.1	13.3
TOPG	14.65	1.56	14.4	13.67	15.72	11.5	19.7
BLKPG	2.51	0.87	2.4	1.9	3.02	0.9	4.8
PFPG	17.16	1.82	16.85	15.9	18.02	12.9	23.5
EFFPG	76.39	8.01	76	71.47	80.32	54.9	104.8

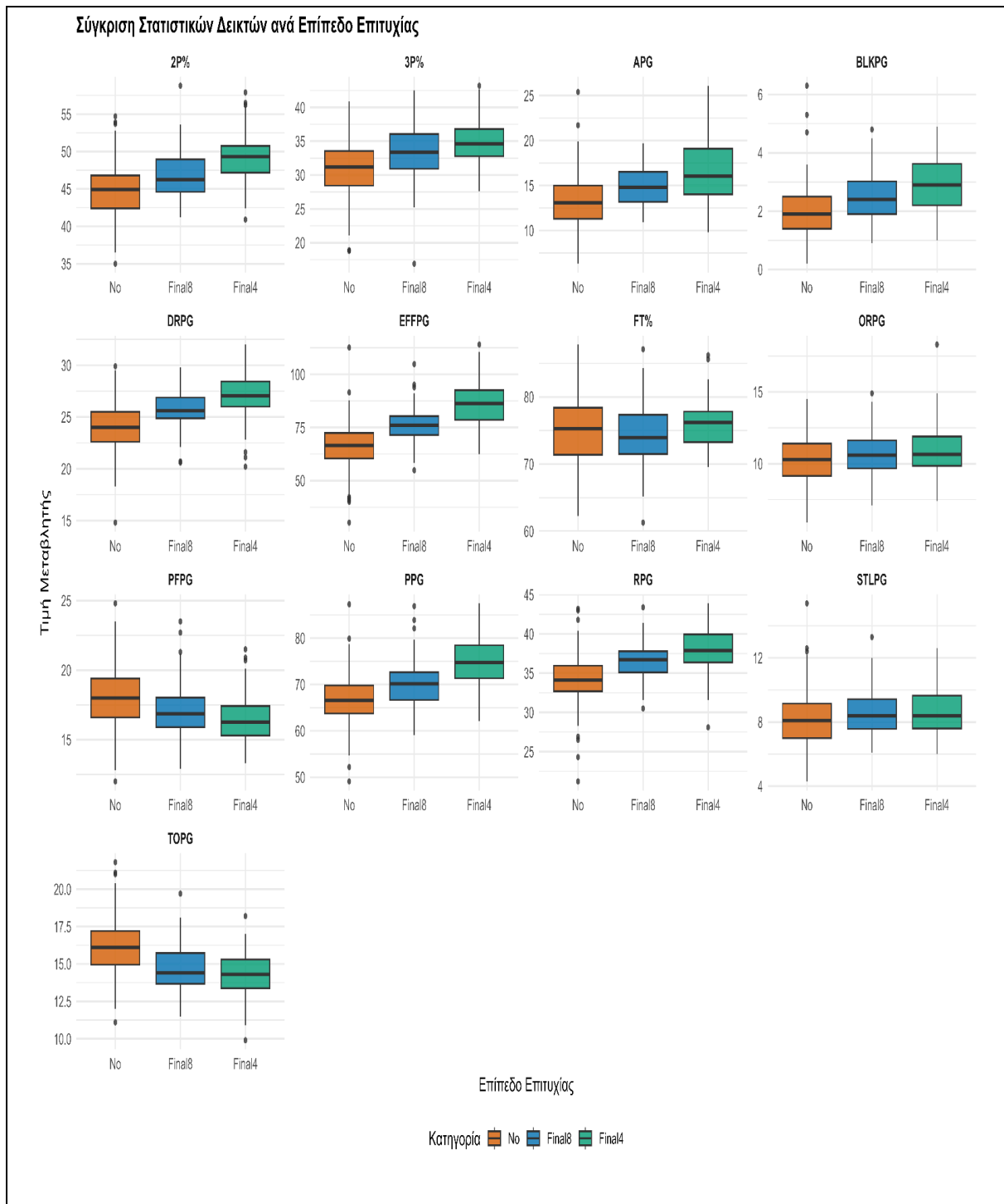
Πίνακας 3.3.2.2 Στατιστικά περιγραφικά μέτρα για τις ομάδες που προκρίθηκαν στο final 8 αλλά όχι στο final 4

	Mean	SD	Median	Q1	Q3	Min	Max
PPG	75.13	5.64	74.7	71.35	78.45	62.1	87.5
2P%	49.14	3.33	49.3	47.17	50.75	40.9	57.9
3P%	35.08	3.43	34.6	32.77	36.8	27.6	43.2
FT%	75.84	3.51	76.2	73.27	77.82	69.6	86.2
ORPG	10.89	1.62	10.65	9.85	11.9	7.4	18.3
DRPG	27.11	2.29	27.05	26	28.42	20.2	32
RPG	37.99	2.85	37.9	36.38	39.95	28.1	43.9
APG	16.63	3.51	16.05	14.02	19.1	9.8	26.1
STLPG	8.71	1.54	8.4	7.6	9.65	6	12.6
TOPG	14.26	1.43	14.3	13.38	15.3	9.9	18.2
BLKPG	2.92	0.87	2.9	2.2	3.62	1	4.9
PFPG	16.51	1.63	16.25	15.3	17.42	13.3	21.5
EFFPG	86.54	10.75	86.2	78.6	92.53	62.5	114

Πίνακας 3.3.2.3 Στατιστικά περιγραφικά μέτρα για τις ομάδες που προκρίθηκαν στο final 4

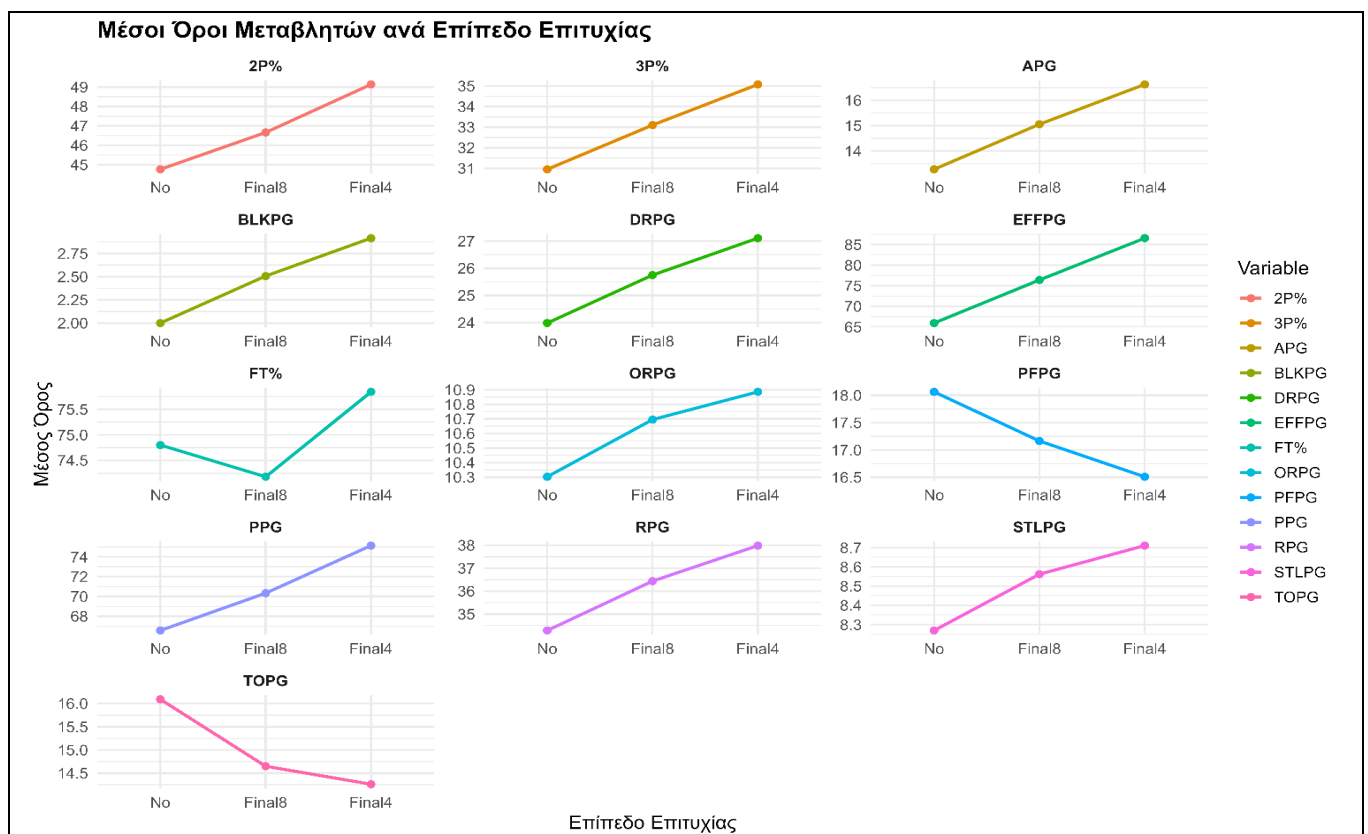
Παρακάτω στο Σχήμα 3.3.2 παρουσιάζονται τα boxplots των βασικών στατιστικών δεικτών ανά επίπεδο επιτυχίας (No, Final8, Final4), επιτρέποντας τη σύγκριση της κατανομής των επιδόσεων μεταξύ των τριών κατηγοριών. Από τη συνολική εικόνα παρατηρείται σαφής διαφοροποίηση κυρίως μεταξύ των ομάδων που δεν προκρίθηκαν (No) και εκείνων που έφτασαν στο Final4. Ειδικότερα, οι μεταβλητές που σχετίζονται με την επιθετική αποτελεσματικότητα, όπως οι πόντοι ανά αγώνα (PPG), το ποσοστό δίποντων (2P%), το ποσοστό τρίποντων (3P%) και ο δείκτης συνολικής αποτελεσματικότητας (EFFPG), εμφανίζουν αυξανόμενη μέση τιμή από την κατηγορία No προς την Final4. Οι ομάδες του Final4 παρουσιάζουν συστηματικά υψηλότερες κεντρικές τιμές και συχνά μικρότερη διασπορά, γεγονός που υποδηλώνει όχι μόνο καλύτερη αλλά και πιο σταθερή απόδοση. Αντίστοιχα, στις μεταβλητές δημιουργίας και συνολικής συμβολής, όπως οι ασίστ (APG) και τα συνολικά ριμπάουντ (RPG), παρατηρείται ανοδική τάση υπέρ των πιο επιτυχημένων ομάδων. Στους αμυντικούς δείκτες (DRPG, STLPG, BLKPG) διακρίνεται επίσης ελαφρά υπεροχή των ομάδων Final4, αν και οι διαφορές είναι λιγότερο έντονες σε σχέση με τους επιθετικούς δείκτες. Αντίθετα, στα λάθη (TOPG) παρατηρείται αντίστροφη τάση, με τις ομάδες Final4 να εμφανίζουν χαμηλότερη μέση τιμή, όπως είδαμε και στους πίνακες παραπάνω, σε σχέση με τις ομάδες No, γεγονός που ενισχύει την υπόθεση ότι η αγωνιστική αποτελεσματικότητα συνδέεται και με τον καλύτερο έλεγχο του παιχνιδιού. Παρατηρείται επίσης ότι σε αρκετές μεταβλητές η κατηγορία Final8 βρίσκεται ενδιάμεσα μεταξύ No και Final4, γεγονός που υποδηλώνει μια σταδιακή κλιμάκωση της απόδοσης ανάλογα με το επίπεδο επιτυχίας.

Η μορφή των boxplots (θέση διαμέσου, εύρος τεταρτημορίων και παρουσία ακραίων τιμών) επιβεβαιώνει ότι οι πιο επιτυχημένες ομάδες δεν υπερέχουν μόνο σε μέσο επίπεδο, αλλά παρουσιάζουν συνολικά ανώτερο στατιστικό προφίλ. Συνολικά, η γραφική απεικόνιση καταδεικνύει ότι η πρόκριση σε υψηλότερο στάδιο της διοργάνωσης συνδέεται με βελτιωμένους επιθετικούς δείκτες, υψηλότερη συνολική αποτελεσματικότητα και καλύτερη διαχείριση λαθών, γεγονός που αναδεικνύει τη σημασία της πολυδιάστατης αγωνιστικής απόδοσης για την επίτευξη επιτυχίας.



Σχήμα 3.3.2 Boxplot κάθε μεταβλητής ανάλογα με την κατηγορία πρόκρισης

Στην συνέχεια παρουσιάζονται και τα διαγράμματα γραμμής (line plot). Το διάγραμμα γραμμής (line plot) αποτελεί γραφική μέθοδο απεικόνισης ποσοτικών δεδομένων που χρησιμοποιείται για την ανάδειξη τάσεων και μεταβολών μεταξύ διακριτών κατηγοριών, συνδέοντας τις τιμές (στην προκειμένη περίπτωση τους μέσους όρους) με ευθύγραμμα τμήματα ώστε να καθίσταται εμφανής η κατεύθυνση της μεταβολής. Σε αντίθεση με το boxplot που παρουσιάζει τη διασπορά και την κατανομή των τιμών, το line plot εστιάζει στην κεντρική τάση και επιτρέπει την εύκολη σύγκριση μεταξύ ομάδων. Από το συγκεκριμένο διάγραμμα προκύπτει σαφής κλιμακωτή τάση βελτίωσης των περισσότερων στατιστικών δεικτών όσο αυξάνεται το επίπεδο επιτυχίας (No → Final8 → Final4). Οι επιθετικοί δείκτες, όπως οι πόντοι ανά αγώνα (PPG), τα ποσοστά ευστοχίας (2P%, 3P%) και ο δείκτης συνολικής αποτελεσματικότητας (EFFPG), παρουσιάζουν σταθερή ανοδική πορεία, γεγονός που υποδηλώνει ότι η αγωνιστική επιτυχία συνδέεται άμεσα με την επιθετική παραγωγικότητα και αποδοτικότητα. Παρόμοια ανοδική τάση εμφανίζουν οι ασίστ (APG) και τα ριμπάουντ (RPG, DRPG, ORPG), στοιχείο που καταδεικνύει καλύτερη ομαδική λειτουργία και έλεγχο του παιχνιδιού από τις πιο επιτυχημένες ομάδες. Αντίθετα, τα λάθη (TOPG) παρουσιάζουν καθοδική τάση, γεγονός που ενισχύει την άποψη ότι η μείωση των λαθών και η αποτελεσματικότερη διαχείριση των κατοχών αποτελούν κρίσιμο παράγοντα επιτυχίας. Επομένως το διάγραμμα επιβεβαιώνει ότι όσο υψηλότερο είναι το επίπεδο πρόκρισης, τόσο βελτιώνεται το συνολικό στατιστικό προφίλ των ομάδων, με ιδιαίτερη έμφαση στην επιθετική αποτελεσματικότητα και στη διαχείριση του παιχνιδιού.



Σχήμα 3.3.3 Line plot κάθε μεταβλητής ανάλογα με την κατηγορία πρόκρισης

Κεφάλαιο 4 : Στατιστική Ανάλυση και Σύγκριση Δεικτών Απόδοσης

Στο παρόν κεφάλαιο πραγματοποιείται η στατιστική ανάλυση των δεδομένων με στόχο τη διερεύνηση διαφορών μεταξύ των ομάδων που προκρίθηκαν στο Final 8 και αυτών που δεν προκρίθηκαν. Για τους σκοπούς της ανάλυσης, οι ομάδες διαχωρίστηκαν σε δύο κατηγορίες: (i) ομάδες που προκρίθηκαν στο Final 8 (Yes) και (ii) ομάδες που δεν προκρίθηκαν (No). Η κατηγοριοποίηση αυτή χρησιμοποιείται ως βασική ανεξάρτητη μεταβλητή σε όλες τις στατιστικές συγκρίσεις που ακολουθούν. Αρχικά, εξετάζεται η κατανομή των δεδομένων μέσω ελέγχων κανονικότητας, προκειμένου να καθοριστεί η καταλληλότητα των στατιστικών μεθόδων που θα εφαρμοστούν. Στη συνέχεια, ελέγχεται η ισότητα διακυμάνσεων μεταξύ των ομάδων, ώστε να επιλεγεί το κατάλληλο παραμετρικό ή μη παραμετρικό τεστ. Με βάση τα αποτελέσματα των ελέγχων αυτών, εφαρμόζονται t-tests ή μη παραμετρικά τεστ (Mann–Whitney U) για τη σύγκριση των ομάδων. Σε όλους τους στατιστικούς ελέγχους που πραγματοποιούνται στο παρόν κεφάλαιο, το επίπεδο στατιστικής σημαντικότητας ορίζεται στο $\alpha = 5\%$. Τέλος, παρουσιάζονται γραφικές απεικονίσεις (violin plots) και υπολογίζονται οι συντελεστές επίδρασης, προκειμένου να αξιολογηθεί όχι μόνο η στατιστική αλλά και η πρακτική σημασία των διαφορών.

4.1 Έλεγχος κανονικότητας.

4.1.1 Θεωρητικό υπόβαθρο

Στη στατιστική ανάλυση, η κατανόηση της κατανομής των δεδομένων αποτελεί θεμελιώδη προϋπόθεση πριν από την εφαρμογή οποιασδήποτε μεθόδου σύγκρισης ή εκτίμησης. Οι μέθοδοι διακρίνονται σε δύο βασικές κατηγορίες: παραμετρικές και μη παραμετρικές. Οι παραμετρικές μέθοδοι, όπως το t-test, η ανάλυση διακύμανσης (ANOVA) ή η Pearson συσχέτιση, βασίζονται στην υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή και η διακύμανση είναι σταθερή μεταξύ των συγκρινόμενων ομάδων. Οι παραμετρικές μέθοδοι χαρακτηρίζονται από υψηλότερη στατιστική ισχύ, δηλαδή μεγαλύτερη ικανότητα να ανιχνεύουν πραγματικές διαφορές όταν αυτές υπάρχουν. Αντιθέτως, οι μη παραμετρικές μέθοδοι, όπως το Mann-Whitney U test ή το Wilcoxon signed-rank test, δεν κάνουν υποθέσεις σχετικά με την κατανομή των δεδομένων και χρησιμοποιούνται όταν τα δεδομένα παραβιάζουν την κανονικότητα ή όταν τα δείγματα είναι μικρά και ευαίσθητα σε ακραίες τιμές. Παρόλο που έχουν λιγότερη στατιστική ισχύ από τις παραμετρικές μεθόδους, παρέχουν ασφαλή και αξιόπιστα αποτελέσματα όταν οι παραμετρικές υποθέσεις δεν ισχύουν. Η κανονικότητα των δεδομένων είναι κρίσιμη επειδή η ακρίβεια των παραμετρικών τεστ εξαρτάται από αυτήν. Συγκεκριμένα, οι υπολογισμοί των t-statistics, των διαστημάτων εμπιστοσύνης και των p-values βασίζονται στην υπόθεση ότι οι παρατηρήσεις προέρχονται από μια συμμετρική, κανονική κατανομή. Όταν αυτή η υπόθεση παραβιάζεται, οι παραμετρικές εκτιμήσεις μπορεί να είναι μεροληπτικές και τα συμπεράσματα να μην είναι έγκυρα.

4.1.2 Δημοφιλείς Έλεγχοι Κανονικότητας

Η αξιολόγηση της κανονικότητας των δεδομένων μπορεί να γίνει με διάφορους στατιστικούς ελέγχους και γραφικές μεθόδους. Ένας από τους πιο ευρέως χρησιμοποιούμενους ελέγχους είναι το Shapiro-Wilk test, το οποίο θεωρείται ιδιαίτερα ισχυρό για μικρά και μεσαία δείγματα. Η μηδενική υπόθεση σε αυτό το τεστ είναι ότι τα δεδομένα προέρχονται από κανονική κατανομή, ενώ η στατιστική W συγκρίνει τις παρατηρούμενες τιμές με τις αντίστοιχες θεωρητικές τιμές μιας κανονικής κατανομής. Εάν το p -value είναι μικρότερο του 0,05, απορρίπτεται η υπόθεση της κανονικότητας, υποδεικνύοντας ότι τα δεδομένα δεν ακολουθούν κανονική κατανομή. Παράλληλα, χρησιμοποιείται και το Kolmogorov-Smirnov test, το οποίο συγκρίνει την κατανομή των παρατηρούμενων δεδομένων με μια θεωρητική κατανομή, όπως η κανονική. Αν και λιγότερο ισχυρό από το Shapiro-Wilk για μικρά δείγματα, το K-S test παρέχει μια γενική ένδειξη της απόκλισης από την κανονικότητα, ιδιαίτερα στις κεντρικές περιοχές της κατανομής. Επιπλέον, το Anderson-Darling test αποτελεί μια παραλλαγή του K-S test που δίνει μεγαλύτερη βαρύτητα στις ουρές της κατανομής, γεγονός σημαντικό όταν τα ακραία δεδομένα μπορούν να επηρεάσουν την ανάλυση. Τέλος, εκτός των ποσοτικών τεστ, συχνά χρησιμοποιούνται και γραφικές μέθοδοι όπως τα Q-Q plots, στα οποία τα δεδομένα αποτυπώνονται σε σχέση με τα θεωρητικά τεταρτημόρια μιας κανονικής κατανομής. Εάν τα σημεία ευθυγραμμίζονται κατά προσέγγιση σε μια ευθεία, η κανονικότητα θεωρείται ικανοποιητική. Η συνδυαστική χρήση αυτών των ελέγχων επιτρέπει την ολοκληρωμένη αξιολόγηση της κατανομής των δεδομένων, παρέχοντας τη βάση για την επιλογή κατάλληλων παραμετρικών ή μη παραμετρικών στατιστικών τεστ. Συνολικά οι έλεγχοι κανονικότητας αποτελούν κρίσιμο βήμα σε οποιαδήποτε στατιστική ανάλυση που χρησιμοποιεί παραμετρικές μεθόδους. Η γνώση της κατανομής των δεδομένων καθορίζει ποια μέθοδος είναι κατάλληλη και εξασφαλίζει αξιόπιστα και έγκυρα συμπεράσματα. Η συνδυαστική χρήση ποσοτικών ελέγχων (Shapiro-Wilk, Anderson-Darling) και γραφικών μεθόδων (Q-Q plots) παρέχει ολοκληρωμένη αξιολόγηση και καθοδηγεί την επιλογή ανάμεσα σε παραμετρικό ή μη παραμετρικό τεστ.

4.1.3 Έλεγχος κανονικότητας στα δεδομένα

Εφαρμόζοντας τον έλεγχο κανονικότητας Shapiro-Wilk ξεχωριστά για τις ομάδες που δεν προκρίθηκαν στο final 8 (Final8_Status = No) και για τις ομάδες που προκρίθηκαν (Final8_Status = Yes), προέκυψαν τα αποτελέσματα που παρουσιάζονται στον Πίνακα 4.1.1. Οι μεταβλητές 2P%, FT%, DRPG και TOPG εμφάνισαν p -value μεγαλύτερο του 0,05 και στις δύο ομάδες, γεγονός που υποδηλώνει ότι δεν απορρίπτεται η υπόθεση της κανονικότητας για τις συγκεκριμένες μεταβλητές. Αντίθετα, οι μεταβλητές PPG, 3P%, ORPG, RPG, APG, STLPG, BLKPG, PFPG και EFFPG παρουσίασαν p -value μικρότερο του 0,05 σε τουλάχιστον μία από τις δύο ομάδες, γεγονός που υποδηλώνει απόκλιση από την κανονική κατανομή. Τα αποτελέσματα αυτά λαμβάνονται υπόψη στην επιλογή των κατάλληλων στατιστικών μεθόδων που χρησιμοποιούνται στη συνέχεια της ανάλυσης.

Για τις μεταβλητές που ακολουθούν κανονική κατανομή, μπορούμε να εφαρμόσουμε παραμετρικά τεστ, όπως το t-test ή το Welch t-test, προκειμένου να συγκρίνουμε τις ομάδες που πέρασαν στο Final 8 με αυτές που δεν πέρασαν. Για τις μεταβλητές που δεν ακολουθούν κανονική κατανομή, η χρήση παραμετρικών τεστ θα μπορούσε να δώσει παραπλανητικά ή μεροληπτικά αποτελέσματα. Σε αυτές τις περιπτώσεις, συνιστάται η εφαρμογή μη παραμετρικών τεστ, όπως το Mann-Whitney U test, που δεν απαιτούν υπόθεση κανονικότητας.

Με βάση τα αποτελέσματα αυτά, το επόμενο βήμα στην ανάλυση θα είναι ο έλεγχος ισότητας διακυμάνσεων (Levene's test) για κάθε μεταβλητή, ώστε να προσδιοριστεί αν η παραμετρική σύγκριση (t-test) μπορεί να εφαρμοστεί με ασφάλεια ή αν πρέπει να χρησιμοποιηθεί η παραλλαγή Welch t-test. Στη συνέχεια, θα εκτελεστούν τα κατάλληλα t-tests ή μη παραμετρικά τεστ για να εντοπιστούν οι μεταβλητές που διαφοροποιούν σημαντικά τις ομάδες σε επίπεδο στατιστικής σημαντικότητας.

Μεταβλητή	p-value (No)	p-value (Yes)	Συμπέρασμα
PPG	0.0319	0.0464	Μη κανονική
2P%	0.5735	0.0814	Κανονική
3P%	0.4133	0.0028	Μη κανονική
FT%	0.1778	0.7449	Κανονική
ORPG	0.7584	0.0003	Μη κανονική
DRPG	0.0627	0.2032	Κανονική
RPG	0.0001	0.1326	Μη κανονική
APG	0.0019	0.0003	Μη κανονική
STLPG	0.0001	0.0001	Μη κανονική
TOPG	0.6819	0.6804	Κανονική
BLKPG	<0.0001	0.0210	Μη κανονική
PFPG	0.2021	<0.0001	Μη κανονική
EFFPG	0.0001	0.0029	Μη κανονική

Πίνακας 4.1.1 Έλεγχος κανονικότητας

4.2 Έλεγχος ισότητας διακυμάνσεων.

4.2.1 Θεωρητικό υπόβαθρο

Στη στατιστική ανάλυση, η υπόθεση της ισότητας διακυμάνσεων (homoscedasticity) αποτελεί μία από τις θεμελιώδεις προϋποθέσεις για πολλά παραμετρικά τεστ, όπως το t-test για δύο ομάδες και η ανάλυση διακύμανσης (ANOVA) για περισσότερες από δύο ομάδες. Η ισότητα διακυμάνσεων σημαίνει ότι οι τιμές των παρατηρήσεων σε κάθε ομάδα παρουσιάζουν συγκρίσιμη διασπορά γύρω από τον μέσο όρο τους. Όταν αυτή η υπόθεση παραβιάζεται, δηλαδή όταν οι διακυμάνσεις διαφέρουν σημαντικά μεταξύ των ομάδων (heteroscedasticity), η χρήση κλασικών παραμετρικών τεστ μπορεί να οδηγήσει σε παραπλανητικά αποτελέσματα, όπως μεροληπτικά p-values, αυξημένο κίνδυνο τύπου I σφάλματος ή λανθασμένες εκτιμήσεις των διαφορών μεταξύ ομάδων. Ο έλεγχος Levene's test αποτελεί τον πιο συνηθισμένο έλεγχο ισότητας διακυμάνσεων.

Η μηδενική υπόθεση (H_0) του τεστ είναι ότι οι διακυμάνσεις είναι ίσες μεταξύ όλων των συγκρινόμενων ομάδων, ενώ η εναλλακτική υπόθεση (H_1) υποστηρίζει ότι τουλάχιστον δύο ομάδες έχουν διαφορετική διακύμανση (Μπερσίμης κ.α., 2021). Η βασική ιδέα του Levene είναι να συγκρίνει τις αποκλίσεις κάθε παρατήρησης από τον μέσο όρο ή τη διάμεσο της ομάδας της, δημιουργώντας ένα νέο δείκτη διασποράς για κάθε παρατήρηση και εφαρμόζοντας ανάλυση διακύμανσης (ANOVA) σε αυτές τις αποκλίσεις. Το τεστ παράγει ένα p-value, με το οποίο ελέγχουμε αν οι διακυμάνσεις είναι στατιστικά ίσες ή διαφορετικές. Υπάρχουν διαφορετικές παραλλαγές του τεστ Levene. Η κλασική εκδοχή χρησιμοποιεί τον μέσο όρο κάθε ομάδας για τον υπολογισμό των αποκλίσεων. Η ερμηνεία των αποτελεσμάτων είναι απλή: αν το p-value του Levene's test είναι μεγαλύτερο από 0,05, η υπόθεση της ισότητας διακυμάνσεων δεν απορρίπτεται, επομένως μπορούμε να χρησιμοποιήσουμε το κλασικό t-test για τη σύγκριση των μέσων όρων μεταξύ ομάδων. Αν το p-value είναι μικρότερο από 0,05, η υπόθεση απορρίπτεται και συνιστάται η χρήση Welch t-test, το οποίο είναι πιο ανθεκτικό στην ανισότητα διακυμάνσεων και παρέχει ασφαλέστερα συμπεράσματα.

4.2.2 Έλεγχος ισότητας διακυμάνσεων στα δεδομένα

Προκειμένου να διασφαλιστεί η εγκυρότητα των στατιστικών συγκρίσεων μεταξύ των ομάδων που πέρασαν στο Final 8 και αυτών που δεν πέρασαν, πραγματοποιήθηκε ο έλεγχος ισότητας διακυμάνσεων (Levene's test) για όλες τις μεταβλητές απόδοσης. Η ισότητα διακυμάνσεων αποτελεί βασική προϋπόθεση για παραμετρικά τεστ. Από τα αποτελέσματα του Levene's test που παρουσιάζονται στον πίνακα 4.2.2.1 παρατηρείται ότι οι μεταβλητές FT%, TOPG, BLKPG και PFPG δεν παρουσιάζουν ισότητα διακυμάνσεων μεταξύ των δύο ομάδων ($p < 0,05$), ενώ οι υπόλοιπες μεταβλητές, όπως PPG, 2P%, 3P%, ORPG, DRPG, RPG, APG, STLPG, EFFPG, εμφανίζουν σταθερές διακυμάνσεις ($p > 0,05$).

Αυτό υποδηλώνει ότι για τις πρώτες μεταβλητές απαιτείται η χρήση τεστ που δεν υποθέτουν ίσες διακυμάνσεις ή η επιλογή του Welch t-test, ενώ για τις υπόλοιπες μπορεί να χρησιμοποιηθεί το κλασικό t-test, υπό την προϋπόθεση ότι πληρούνται και οι υπόλοιπες παραδοχές. Σε συνδυασμό με τα αποτελέσματα του Shapiro-Wilk test, που αξιολόγησαν την κανονικότητα των μεταβλητών, προκύπτει για τη συνέχεια της ανάλυσης:

- Για μεταβλητές που ακολουθούν κανονική κατανομή και έχουν ίση διακύμανση, θα εφαρμοστεί το t-test.
- Για μεταβλητές που ακολουθούν κανονική κατανομή αλλά παρουσιάζουν άνισες διακυμάνσεις, θα εφαρμοστεί το Welch t-test.
- Για μεταβλητές που δεν ακολουθούν κανονική κατανομή, ανεξαρτήτως ισότητας διακυμάνσεων, θα χρησιμοποιηθεί μη παραμετρικό τεστ, όπως το Mann-Whitney U test, το οποίο δεν προϋποθέτει ούτε κανονικότητα ούτε ίσες διακυμάνσεις.

Με αυτή τη διαδικασία εξασφαλίζεται ότι κάθε μεταβλητή αξιολογείται με το κατάλληλο στατιστικό εργαλείο, διασφαλίζοντας ότι οι συγκρίσεις μεταξύ ομάδων βασίζονται σε στατιστικά έγκυρα και αξιόπιστα αποτελέσματα. Τα αποτελέσματα που θα προκύψουν από τα επιλεγμένα τεστ θα επιτρέψουν την ταυτοποίηση των δεικτών απόδοσης που διαφοροποιούν σημαντικά τις ομάδες που προκρίθηκαν στο Final8 από τις υπόλοιπες.

Μεταβλητή	Levene F	p-value	Συμπέρασμα
PPG	3.742	0.054	Ίσες διακυμάνσεις
2P%	0.003	0.954	Ίσες διακυμάνσεις
3P%	0.832	0.362	Ίσες διακυμάνσεις
FT%	7.572	0.006	Μη ίσες διακυμάνσεις
ORPG	2.404	0.122	Ίσες διακυμάνσεις
DRPG	0.216	0.642	Ίσες διακυμάνσεις
RPG	0.315	0.575	Ίσες διακυμάνσεις
APG	0.784	0.376	Ίσες διακυμάνσεις
STLPG	2.535	0.112	Ίσες διακυμάνσεις
TOPG	5.264	0.022	Μη ίσες διακυμάνσεις
BLKPG	5.651	0.018	Μη ίσες διακυμάνσεις
PFFPG	7.621	0.006	Μη ίσες διακυμάνσεις
EFFPG	0.847	0.358	Ίσες διακυμάνσεις

Πίνακας 4.2.2.1 Έλεγχος ισότητας διακυμάνσεων

4.3. Έλεγχοι διαφορών μεταξύ ομάδων

4.3.1 Θεωρητικό υπόβαθρο

Το t-test αποτελεί θεμελιώδες εργαλείο στη στατιστική για τη σύγκριση δύο μέσων όρων και την αξιολόγηση της στατιστικής σημαντικότητας των διαφορών μεταξύ δύο ομάδων. Αναπτύχθηκε αρχικά από τον William Sealy Gosset το 1908, υπό το ψευδώνυμο “Student”, και χρησιμοποιείται ευρέως όταν τα δεδομένα ακολουθούν περίπου την κανονική κατανομή. Το t-test είναι παραμετρικό τεστ, γεγονός που σημαίνει ότι στηρίζεται σε παραδοχές σχετικά με την κατανομή και τη διασπορά των δεδομένων. Ο βασικός σκοπός του t-test είναι να εξετάσει αν οι διαφορές που παρατηρούνται μεταξύ δύο μέσων όρων είναι στατιστικά σημαντικές ή αν μπορούν να αποδοθούν στην τυχαία διακύμανση των παρατηρήσεων. Στην ουσία, το τεστ υπολογίζει μια αναλογία: $t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$ όπου $\bar{X}_1$ και $\bar{X}_2$ είναι οι μέσοι όροι των δύο ομάδων και SE η τυπική σταθερά σφάλματος της διαφοράς. Όσο μεγαλύτερη είναι η διαφορά των μέσων όρων σε σχέση με την ομαδική διασπορά, τόσο μεγαλύτερη είναι η τιμή t, αυξάνοντας την πιθανότητα να απορριφθεί η μηδενική υπόθεση (H_0) που ισχυρίζεται ότι οι μέσοι όροι είναι ίσοι. Για να είναι έγκυρα τα αποτελέσματα ενός t-test, πρέπει να πληρούνται συγκεκριμένες παραδοχές:

1. **Κανονικότητα:** Οι τιμές της μεταβλητής σε κάθε ομάδα πρέπει να ακολουθούν κανονική κατανομή. Ο έλεγχος αυτός γίνεται συχνά με το Shapiro-Wilk test ή άλλα τεστ κανονικότητας.
2. **Ανεξαρτησία παρατηρήσεων:** Οι τιμές κάθε ομάδας πρέπει να προέρχονται από ανεξάρτητα δείγματα, χωρίς αλληλεξάρτηση μεταξύ τους.

3. **Ισότητα διακυμάνσεων (homoscedasticity):** Οι διακυμάνσεις των δύο ομάδων πρέπει να είναι περίπου ίσες. Αυτό ελέγχεται με το Levene's test ή το F-test. Σε περίπτωση παραβίασης αυτής της υπόθεσης χρησιμοποιείται η παραλλαγή Welch t-test, η οποία δεν απαιτεί ίσες διακυμάνσεις.

4.3.2 Εκδοχές του t-test

Υπάρχουν τρεις κύριες εκδοχές του t-test ανάλογα με τη φύση των δεδομένων:

1. **Independent (ανεξάρτητο) t-test:** Χρησιμοποιείται για τη σύγκριση δύο ανεξάρτητων ομάδων, όπως στην περίπτωση μας οι ομάδες που πέρασαν στο Final 8 και αυτές που δεν πέρασαν.
2. **Paired (ζευγαρωτό) t-test:** Χρησιμοποιείται για επαναλαμβανόμενες μετρήσεις στην ίδια ομάδα ή για ζεύγη παρατηρήσεων που σχετίζονται άμεσα.
3. **Welch t-test:** Μια τροποποιημένη εκδοχή του ανεξάρτητου t-test, που είναι πιο ανθεκτική στην παραβίαση της υπόθεσης ισότητας διακυμάνσεων.

4.3.3 Στατιστική ερμηνεία

Το αποτέλεσμα του t-test συνοδεύεται από την τιμή p-value, η οποία δείχνει την πιθανότητα να παρατηρηθεί μια διαφορά ίση ή μεγαλύτερη από αυτή που παρατηρήθηκε, αν η μηδενική υπόθεση ήταν αληθής.

Συχνά χρησιμοποιείται επίπεδο σημαντικότητας ($\alpha = 0.05$) για την απόφαση:

- $p < 0,05 \rightarrow$ Απορρίπτουμε τη H_0 , η διαφορά θεωρείται στατιστικά σημαντική.
- $p \geq 0,05 \rightarrow$ Δεν απορρίπτουμε τη H_0 , η διαφορά δεν θεωρείται σημαντική.

Στο πλαίσιο ανάλυσης των δεικτών απόδοσης στην EuroLeague Γυναικών, το t-test χρησιμοποιείται για να συγκρίνει μεταβλητές μεταξύ των ομάδων που πέρασαν στο Final 8 και αυτών που δεν πέρασαν. Η χρήση του t-test επιτρέπει την ποσοτική αξιολόγηση των διαφορών και τον εντοπισμό δεικτών που διαφοροποιούν σημαντικά τις ομάδες υψηλής απόδοσης από τις υπόλοιπες, συμβάλλοντας στην κατανόηση των παραγόντων που επηρεάζουν την επιτυχία στο γυναικείο μπάσκετ υψηλού επιπέδου.

4.3.4 T-TEST στα δεδομένα

Σε αυτήν την ενότητα, πραγματοποιήθηκαν t-tests για τη σύγκριση των μέσων τιμών. Για κάθε μεταβλητή επιλέχθηκε το κατάλληλο t-test ανάλογα με τα αποτελέσματα που προηγήθηκαν στον έλεγχο κανονικότητας και στον έλεγχο ισότητας διακυμάνσεων.

Για τη μεταβλητή PPG, ο έλεγχος Shapiro–Wilk έδειξε ότι τα δεδομένα δεν ακολουθούν κανονική κατανομή, ενώ το τεστ Levene επιβεβαίωσε την ισότητα διακυμάνσεων μεταξύ των ομάδων. Ως εκ τούτου, εφαρμόστηκε το μη παραμετρικό Mann–Whitney U test.

Τα αποτελέσματα του Πίνακα 4.3.4.1 έδειξαν στατιστικά σημαντική διαφορά ($W = 8038$, $p\text{-value} < 2.2e-16$), υποδηλώνοντας ότι οι ομάδες διαφοροποιούνται ως όχι την επιθετική όχι απόδοση σε πόντους ανά παιχνίδι.

Μεταβλητή: PPG

Κανονικότητα (Shapiro-Wilk): Όχι

Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

W = 8038

p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

Πίνακας 4.3.4.1 Mann-Whitney U test για την μεταβλητή PPG

Για τη μεταβλητή 2P%, διαπιστώθηκε κανονική κατανομή και ισότητα διακυμάνσεων, γεγονός που επέτρεψε την εφαρμογή του ανεξάρτητου t-test. Σύμφωνα με τον Πίνακα 4.3.4.2, προκύπτει στατιστικά σημαντική διαφορά ($t = -8.7366$, $df = 389$, $p\text{-value} < 2.2e-16$), με 95% διάστημα εμπιστοσύνης $[-3.84, -2.43]$. Η ομάδα Yes εμφανίζει υψηλότερη μέση τιμή (47.90) σε σχέση με την ομάδα No (44.76), γεγονός που δείχνει καλύτερη αποτελεσματικότητα στα δίποντα.

Μεταβλητή: 2P%

Κανονικότητα (Shapiro-Wilk): Ναι

Ισότητα διακυμάνσεων (Levene): Ναι

= ΑΠΟΤΕΛΕΣΜΑ T-TEST =

Two Sample t-test

data: data[[var]] by data\$Final8_Status

$t = -8.7366$

$df = 389$

$p\text{-value} < 2.2e-16$

alternative hypothesis: true difference in means between group No and group Yes is not equal to 0

percent confidence interval:

$[-3.841583, -2.430182]$

sample estimates:

mean in group No mean in group Yes

44.76233 47.89821

Πίνακας 4.3.4.2 Two Sample t-test για την μεταβλητή 2P%

Για τη μεταβλητή 3P%, ο έλεγχος κανονικότητας έδειξε απόκλιση από την κανονική κατανομή, ενώ οι διακυμάνσεις κρίθηκαν ίσες. Για τον λόγο αυτό χρησιμοποιήθηκε το Mann-Whitney U test. Τα αποτελέσματα του Πίνακα 4.3.4.3 έδειξαν στατιστικά σημαντική διαφορά ($W = 10276$, $p\text{-value} = 2.102e-14$), υποδηλώνοντας ότι η ομάδα Yes παρουσιάζει υψηλότερη απόδοση στο ποσοστό ευστοχίας τριπόντων.

Μεταβλητή: 3P%

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 10276
p-value = 2.102e-14

alternative hypothesis: true location shift is not equal to 0

Πίνακας 4.3.4.3 Mann-Whitney U test για την μεταβλητή 3P%.

Για τη μεταβλητή FT%, τα δεδομένα ακολουθούν κανονική κατανομή, αλλά οι διακυμάνσεις δεν είναι ίσες, γεγονός που οδήγησε στη χρήση του Welch t-test. Όχι προκύπτει από τον Πίνακα 4.3.4.4, δεν υπάρχει στατιστικά σημαντική διαφορά ($t = -0.45269$, $df = 383.82$, $p\text{-value} = 0.651$), ενώ το 95% διάστημα εμπιστοσύνης $[-1.12, 0.70]$ περιλαμβάνει το μηδέν. Οι μέσες τιμές των ομάδων (74.80 για No και 75.01 για Yes) είναι σχεδόν όχι, επιβεβαιώνοντας την απουσία ουσιαστικής διαφοράς.

Μεταβλητή: FT%

Κανονικότητα (Shapiro-Wilk): Ναι
Ισότητα διακυμάνσεων (Levene): Όχι

=== ΑΠΟΤΕΛΕΣΜΑ T-TEST ===

Welch Two Sample t-test

data: data[[var]] by data\$Final8_Status

$t = -0.45269$
 $df = 383.82$
 $p\text{-value} = 0.651$

alternative hypothesis: true difference in means between group No and group Yes is not equal to 0

95 percent confidence interval:
-1.1187121 0.6997411

sample estimates:
mean in group No mean in group Yes
74.80135 75.01071

Πίνακας 4.3.4.4. Welch Two Sample t-test για την μεταβλητή FT%

Για τη μεταβλητή ORPG, διαπιστώθηκε μη κανονική κατανομή και ισότητα διακυμάνσεων, οπότε εφαρμόστηκε το Mann-Whitney U test. Τα αποτελέσματα του Πίνακα 4.3.4.5 έδειξαν στατιστικά σημαντική διαφορά ($W = 16030$, $p\text{-value} = 0.01456$), υποδηλώνοντας ότι η ομάδα Yes εμφανίζει ελαφρώς υψηλότερες τιμές επιθετικών ριμπάουντ, αν και η διαφορά είναι περιορισμένη σε πρακτικό επίπεδο.

Μεταβλητή: ORPG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 16030
p-value = 0.01456

alternative hypothesis: true location shift is not equal to 0

Πίνακας 4.3.4.5 Mann-Whitney U test για την μεταβλητή ORPG

Για τη μεταβλητή DRPG, πληρούνται οι προϋποθέσεις κανονικότητας και ισότητας διακυμάνσεων, επομένως εφαρμόστηκε το t-test. Σύμφωνα με τον Πίνακα 4.3.4.6, υπάρχει έντονη στατιστικά σημαντική διαφορά ($t = -10.979$, $df = 389$, $p\text{-value} < 2.2e-16$) με 95% διάστημα εμπιστοσύνης $[-2.88, -2.00]$. Η ομάδα Yes παρουσιάζει υψηλότερη μέση τιμή (26.43) έναντι της ομάδας No (23.99), γεγονός που υποδηλώνει καλύτερη αμυντική απόδοση.

Μεταβλητή: DRPG

Κανονικότητα (Shapiro-Wilk): Ναι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ T-TEST ===

Two Sample t-test

data: data[[var]] by data\$Final8_Status

$t = -10.979$
 $df = 389$
 $p\text{-value} < 2.2e-16$

alternative hypothesis: true difference in means between group No and group Yes is not equal to 0

95 percent confidence interval:
 $-2.877924 \ -2.003744$

sample estimates:
mean in group No mean in group Yes
23.98655 26.42738

Πίνακα 4.3.4.6 Two Sample t-test για τη μεταβλητή DRPG.

Για τη μεταβλητή RPG, η μη κανονικότητα των δεδομένων και η ισότητα διακυμάνσεων οδήγησαν στη χρήση του Mann-Whitney U test. Τα αποτελέσματα του Πίνακα 4.3.4.7 έδειξαν στατιστικά σημαντική διαφορά ($W = 8162.5$, $p\text{-value} < 2.2e-16$), υποδηλώνοντας ότι η ομάδα που προκρίθηκε εμφανίζει υψηλότερες τιμές συνολικών ριμπάουντ ανά παιχνίδι.

Μεταβλητή: RPG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 8162.5
p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.7 Mann-Whitney U test για την μεταβλητή RPG

Για τη μεταβλητή APG, διαπιστώθηκε επίσης μη κανονικότητα και ισότητα διακυμάνσεων, οδηγώντας στη χρήση του Mann-Whitney U test. Όπως φαίνεται στον Πίνακα 4.3.4.8, υπάρχει στατιστικά σημαντική διαφορά ($W = 9943$, $p\text{-value} = 1.943e-15$), με την ομάδα Yes να εμφανίζει υψηλότερη απόδοση σε ασίστ ανά παιχνίδι.

Μεταβλητή: APG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 9943
p-value = 1.943e-15

alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.8 Mann-Whitney U test για την μεταβλητή APG

Για τη μεταβλητή STLPG, τα δεδομένα δεν είναι κανονικά και οι διακυμάνσεις είναι ίσες. Επομένως, εφαρμόστηκε το Mann-Whitney U test, το οποίο, σύμφωνα με τον Πίνακα 4.3.4.9, έδειξε στατιστικά σημαντική διαφορά ($W = 15954$, $p\text{-value} = 0.01201$). Η διαφορά αυτή είναι μικρού μεγέθους, με ελαφρώς υψηλότερες τιμές για την ομάδα Yes.

Μεταβλητή: STLPG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 15954
p-value = 0.01201

alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.9 Mann-Whitney U test για την μεταβλητή STLPG

Για τη μεταβλητή TOPG, η κανονικότητα επιβεβαιώθηκε αλλά οι διακυμάνσεις δεν ήταν ίσες, οπότε εφαρμόστηκε το Welch t-test. Τα αποτελέσματα του Πίνακα 4.3.4.10 δείχνουν στατιστικά σημαντική διαφορά ($t = 9.7535$, $df = 384.8$, $p\text{-value} < 2.2e-16$) με 95% διάστημα εμπιστοσύνης [1.30, 1.96]. Η ομάδα No παρουσιάζει υψηλότερη μέση τιμή (16.09) σε σχέση με την ομάδα Yes (14.46), υποδηλώνοντας περισσότερα λάθη ανά παιχνίδι.

Μεταβλητή: TOPG

Κανονικότητα (Shapiro-Wilk): Ναι
Ισότητα διακυμάνσεων (Levene): Όχι

=== ΑΠΟΤΕΛΕΣΜΑ T-TEST ===

Welch Two Sample t-test

data: data[[var]] by data\$Final8_Status

$t = 9.7535$
 $df = 384.8$
 $p\text{-value} < 2.2e-16$

alternative hypothesis: true difference in means between group No and group Yes is not equal to 0

95 percent confidence interval:
1.304283 1.962890

sample estimates:
mean in group No mean in group Yes
16.09013 14.45655

Πίνακας 4.3.4.10 Welch Two Sample t-test για την μεταβλητή TOPG

Για τη μεταβλητή BLKPG, η μη κανονικότητα και η ανισότητα διακυμάνσεων οδήγησαν στη χρήση του Mann-Whitney U test. Σύμφωνα με τον Πίνακα 4.3.4.11, υπάρχει στατιστικά σημαντική διαφορά ($W = 10178$, $p\text{-value} = 1.017e-14$), υποδηλώνοντας καλύτερη αμυντική παρουσία της ομάδας Yes στα blocks.

Μεταβλητή: BLKPG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Όχι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 10178
p-value = 1.017e-14

alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.11 1 Mann-Whitney U test για την μεταβλητή BLKPG

Για τη μεταβλητή PFPG, τα δεδομένα δεν είναι κανονικά και οι διακυμάνσεις δεν είναι ίσες, επομένως εφαρμόστηκε το Mann-Whitney U test. Τα αποτελέσματα του Πίνακα 4.3.4.12 έδειξαν στατιστικά σημαντική διαφορά ($W = 25703$, $p\text{-value} = 2.937e-10$), με την ομάδα που δεν προκρίθηκε να εμφανίζει υψηλότερες τιμές προσωπικών φάουλ.

Μεταβλητή: PFPG

Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Όχι

=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===

Wilcoxon rank sum test with continuity correction

data: data[[var]] by data\$Final8_Status

W = 25703
p-value = 2.937e-10

alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.12 Mann-Whitney U test για την μεταβλητή PFPG

Τέλος, για τη μεταβλητή EFFPG, διαπιστώθηκε μη κανονική κατανομή και ισότητα διακυμάνσεων, οδηγώντας στη χρήση του Mann-Whitney U test. Όπως παρουσιάζεται στον Πίνακα 4.3.4.13, υπάρχει ιδιαίτερα ισχυρή στατιστικά σημαντική διαφορά ($W = 4717$, $p\text{-value} < 2.2e-16$), γεγονός που δείχνει ότι η συνολική αποτελεσματικότητα αποτελεί κρίσιμο παράγοντα διαφοροποίησης των ομάδων.

Μεταβλητή: EFFPG
Κανονικότητα (Shapiro-Wilk): Όχι
Ισότητα διακυμάνσεων (Levene): Ναι
=== ΑΠΟΤΕΛΕΣΜΑ ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΥ (Mann-Whitney U) ===
Wilcoxon rank sum test with continuity correction
data: data[[var]] by data\$Final8_Status
W = 4717
p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

Πίνακα 4.3.4.13 Mann-Whitney U test για την μεταβλητή EFFPG

Στον πίνακα 4.3.5 παρουσιάζονται συνολικά όλα τα αποτελέσματα από τους παραπάνω ελέγχους.

Μεταβλητή	Στατιστικός Έλεγχος	p-value	Στατιστικά σημαντικό
PPG	Mann-Whitney	4.18e-22	Ναι
2P%	t-test (equal var)	7.24e-17	Ναι
3P%	Mann-Whitney	2.10e-14	Ναι
FT%	t-test (Welch)	0.651	Όχι
ORPG	Mann-Whitney	0.0146	Ναι
DRPG	t-test (equal var)	1.30e-24	Ναι
RPG	Mann-Whitney	1.24e-21	Ναι
APG	Mann-Whitney	1.94e-15	Ναι
STLPG	Mann-Whitney	0.0120	Ναι
TOPG	t-test (Welch)	3.14e-20	Ναι
BLKPG	Mann-Whitney	1.02e-14	Ναι
PFP	Mann-Whitney	2.94e-10	Ναι
EFFPG	Mann-Whitney	8.85e-37	Ναι

Πίνακας 4.3.5. Συνολικά αποτελέσματα

4.4 Γραφική απεικόνιση και συντελεστής συσχέτισης

4.4.1 Θεωρητικό υπόβαθρο

Σε αυτήν την ενότητα θα γίνει η γραφική απεικόνιση καθώς και ο υπολογισμός του συντελεστή επίδρασης για κάθε μεταβλητή. Παρακάτω ακολουθεί για κάθε μεταβλητή το αντίστοιχο διάγραμμα violin. Τα διαγράμματα violin αποτελούν έναν συνδυασμό boxplot και εκτίμησης πυκνότητας και χρησιμοποιούνται για την απεικόνιση της κατανομής μιας μεταβλητής. Πιο συγκεκριμένα, το “σώμα” του violin δείχνει την πυκνότητα των δεδομένων: όσο πιο φαρδύ είναι σε ένα σημείο, τόσο περισσότερες παρατηρήσεις συγκεντρώνονται σε αυτή την τιμή.

Παράλληλα, στο εσωτερικό συνήθως περιλαμβάνεται ένα boxplot, το οποίο παρουσιάζει τη διάμεσο, τα τεταρτημόρια και την διασπορά των δεδομένων. Το βασικό πλεονέκτημα των violin plots είναι ότι επιτρέπουν την καλύτερη κατανόηση της μορφής της κατανομής (π.χ. συμμετρία), κάτι που δεν φαίνεται εύκολα μόνο από ένα boxplot. Έτσι, είναι ιδιαίτερα χρήσιμα για τη σύγκριση της κατανομής μιας μεταβλητής μεταξύ διαφορετικών ομάδων. Το point-biserial correlation αποτελεί ειδική περίπτωση του Pearson correlation coefficient και χρησιμοποιείται για τη μέτρηση της σχέσης μεταξύ μιας συνεχούς και μιας διχοτομικής μεταβλητής. Σε αντίθεση με την p-value, η οποία υποδεικνύει αν μια σχέση είναι στατιστικά σημαντική, ο συντελεστής point-biserial εκφράζει το μέγεθος και την κατεύθυνση της σχέσης στην πράξη. Ο συντελεστής αυτός λαμβάνει τιμές στο διάστημα από -1 έως +1, όπου τιμές κοντά στο ± 1 υποδηλώνουν ισχυρή συσχέτιση, ενώ τιμές κοντά στο 0 υποδηλώνουν ασθενή ή μηδενική συσχέτιση.

Θετική τιμή υποδηλώνει ότι υψηλότερες τιμές της συνεχούς μεταβλητής συνδέονται με την κατηγορία που έχει κωδικοποιηθεί ως 1, ενώ αρνητική τιμή υποδηλώνει το αντίθετο.

Από θεωρητική άποψη, ο συντελεστής point-biserial είναι ισοδύναμος με τον συντελεστή Pearson όταν η διχοτομική μεταβλητή κωδικοποιείται ως 0 και 1, γεγονός που τον καθιστά ιδιαίτερα χρήσιμο σε εφαρμογές όπου εξετάζεται η σχέση μεταξύ ποσοτικών χαρακτηριστικών και δυαδικών εκβάσεων (UCLA Statistical Consulting Group, n.d.). Επιπλέον, ο συντελεστής αυτός συνδέεται άμεσα με τον έλεγχο διαφορών μεταξύ δύο ομάδων, καθώς μπορεί να θεωρηθεί ως μετασχηματισμός του t-test, παρέχοντας έτσι μια εναλλακτική προσέγγιση για την αξιολόγηση της σχέσης μεταξύ μιας συνεχούς μεταβλητής και μιας δυαδικής κατηγορίας (Lane, n.d.).

Η μαθηματική του μορφή δίνεται από τη σχέση:

$$r_{pb} = \frac{M_1 - M_0}{s} \sqrt{pq}$$

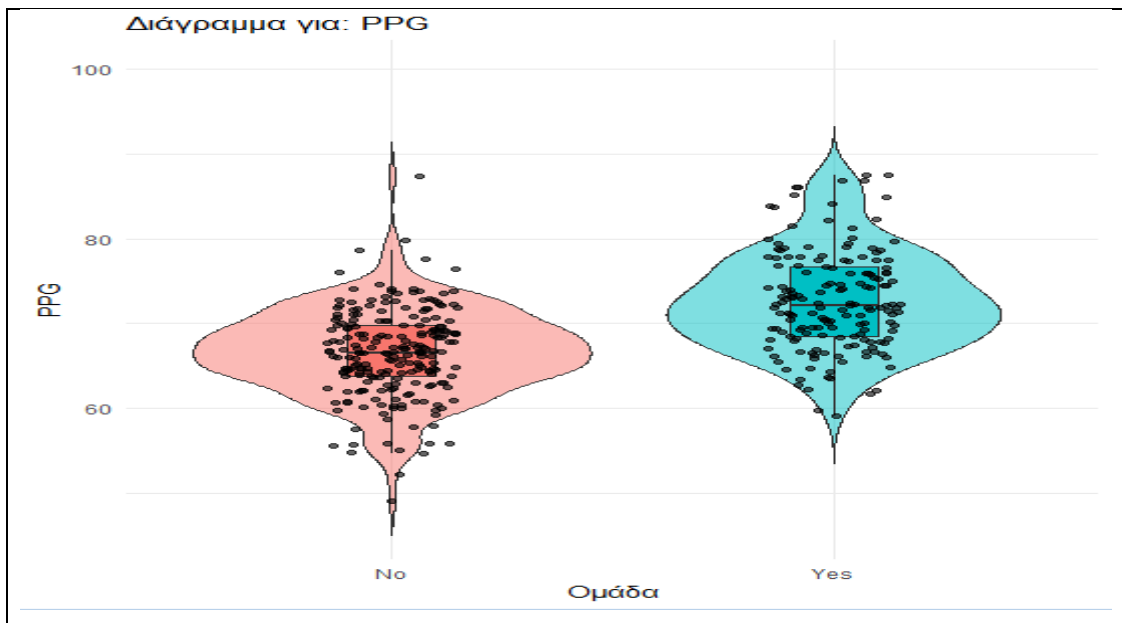
όπου M_1 και M_0 είναι οι μέσοι όροι της συνεχούς μεταβλητής για τις δύο κατηγορίες, s η συνολική τυπική απόκλιση και p, q οι αναλογίες των ομάδων. Η χρήση του συντελεστή point-biserial είναι ιδιαίτερα σημαντική στην ανάλυση δεδομένων, καθώς επιτρέπει την ποσοτικοποίηση της σχέσης μεταξύ μεταβλητών διαφορετικής φύσης και συμβάλλει στην εξαγωγή αξιόπιστων και ερμηνεύσιμων συμπερασμάτων.

4.4.2 Γραφική απεικόνιση

Στο σχήμα 4.4.2.1 φαίνεται το violin plot της μεταβλητής PPG το οποίο παρουσιάζει την κατανομή των πόντων ανά αγώνα για τις ομάδες που προκρίθηκαν (Yes) και αυτές που δεν προκρίθηκαν (No), αναδεικνύοντας σαφή διαφοροποίηση μεταξύ των δύο κατηγοριών, καθώς οι ομάδες που προκρίθηκαν εμφανίζουν γενικά υψηλότερες τιμές. Η κατανομή τους είναι μετατοπισμένη προς μεγαλύτερες τιμές σε σύγκριση με αυτές που δεν προκρίθηκαν γεγονός που υποδηλώνει ανώτερη επιθετική απόδοση.

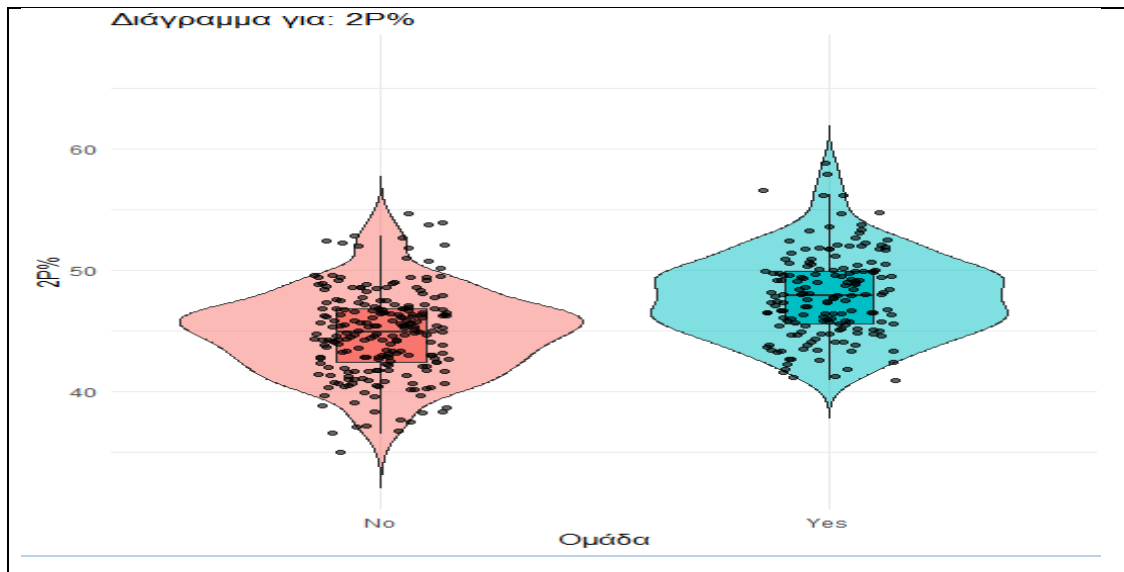
Επίσης παρατηρούνται περιπτώσεις όπου ομάδες με υψηλό PPG δεν καταφέρνουν να προκριθούν, ενώ άλλες με χαμηλότερο PPG προκρίνονται, γεγονός που μπορεί να αποδοθεί στο ότι η επιτυχία μιας ομάδας δεν εξαρτάται αποκλειστικά από την επιθετική της απόδοση αλλά από ένα σύνολο παραγόντων, όπως η αμυντική λειτουργία, τα ριμπάουντ, τα λάθη και η συνολική αποτελεσματικότητα σε κρίσιμες στιγμές. Επιπλέον, ο δείκτης PPG αποτελεί μέση τιμή και δεν αποτυπώνει τη σταθερότητα της απόδοσης ή τη σημασία συγκεκριμένων αγώνων.

Στο διάγραμμα παρατηρούνται επίσης ακραίες τιμές και στις δύο ομάδες, με ιδιαίτερα χαμηλές τιμές στην ομάδα No και ιδιαίτερα υψηλές στην ομάδα Yes, οι οποίες αντανακλούν ειδικές περιπτώσεις ομάδων με ασυνήθιστα χαμηλή ή υψηλή επιθετική παραγωγικότητα· ωστόσο, οι τιμές αυτές δεν φαίνεται να επηρεάζουν τη συνολική τάση των δεδομένων. Συνεπώς, η μεταβλητή PPG αποτελεί έναν σημαντικό αλλά όχι επαρκή από μόνη της δείκτη επιτυχίας.



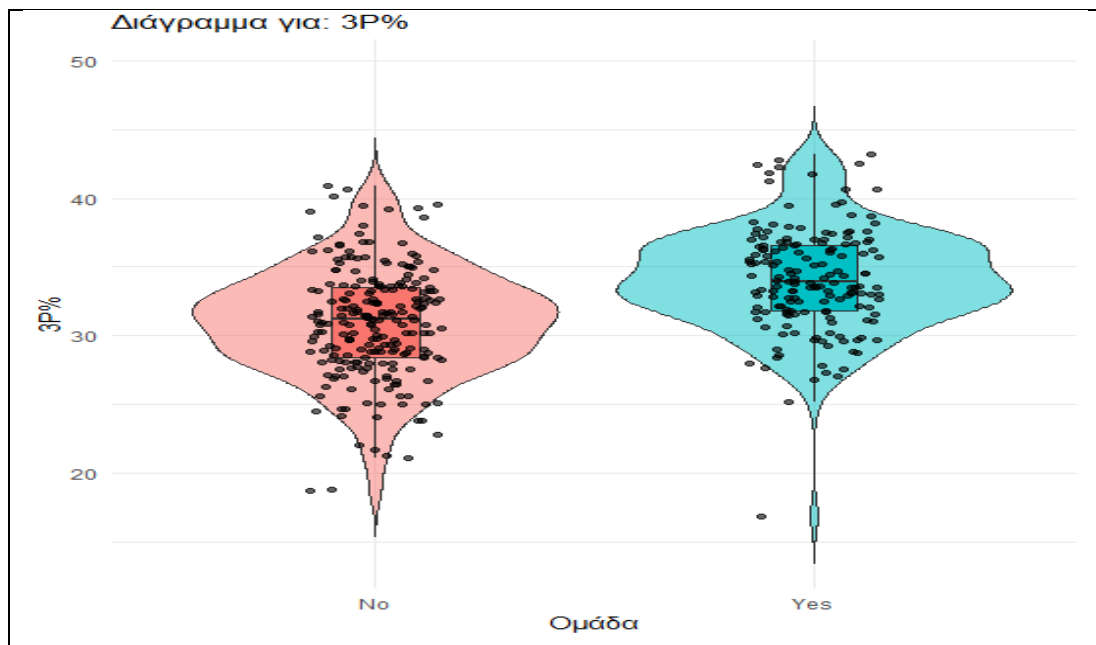
Σχήμα 4.4.2.1 Violin plot για PPG.

Στο σχήμα 4.4.2.2 παρουσιάζεται το διάγραμμα για το ποσοστό ευστοχίας στα δίποντα (2P%) δείχνει ότι οι ομάδες που προκρίθηκαν εμφανίζουν γενικά υψηλότερες τιμές, με την κατανομή τους να είναι μετατοπισμένη προς τα πάνω σε σχέση με αυτές που δεν εξασφάλισαν την πρόκριση. Ωστόσο, παρατηρούνται περιπτώσεις όπου ομάδες με υψηλό ποσοστό ευστοχίας δεν προκρίνονται, γεγονός που υποδηλώνει ότι η αποτελεσματικότητα στα δίποντα, αν και σημαντική, δεν είναι από μόνη της επαρκής για την επιτυχία. Αυτό μπορεί να εξηγηθεί από αδυναμίες σε άλλους τομείς, όπως η άμυνα ή τα λάθη. Παράλληλα, παρατηρούνται ορισμένες ακραίες τιμές, κυρίως υψηλές στην ομάδα Yes και χαμηλότερες στην ομάδα No, οι οποίες πιθανόν αντανακλούν ομάδες με ιδιαίτερα υψηλή ή χαμηλή ευστοχία. Ωστόσο, οι τιμές αυτές δεν μεταβάλλουν τη συνολική εικόνα της θετικής σχέσης.



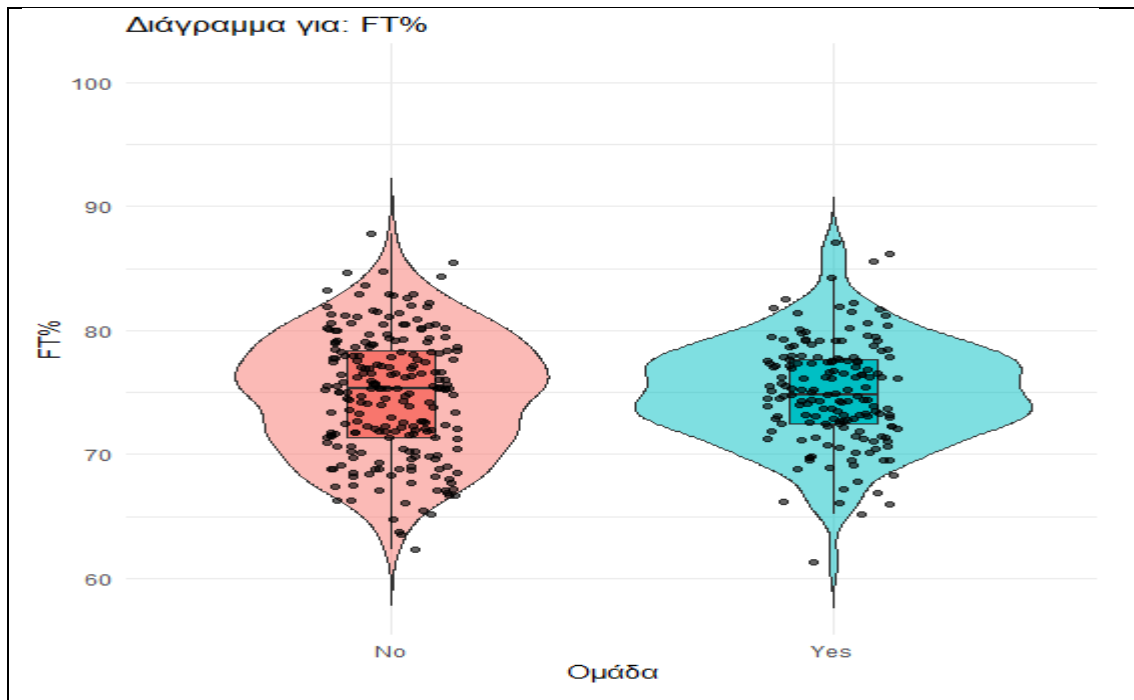
Σχήμα 4.4.2.2 Violin plot για 2P%.

Στο σχήμα 4.4.2.3 για τη μεταβλητή 3P%, οι ομάδες που προκρίθηκαν τείνουν να έχουν ελαφρώς καλύτερα ποσοστά ευστοχίας στα τρίποντα, αν και παρατηρείται σημαντική επικάλυψη μεταξύ των δύο ομάδων. Παρόλα αυτά, ομάδες με υψηλό ποσοστό στα τρίποντα ενδέχεται να μην προκριθούν, κάτι που μπορεί να οφείλεται σε αστάθεια στην απόδοση ή σε ελλείψεις σε άλλες αγωνιστικές παραμέτρους. Επιπλέον, υπάρχουν ακραίες τιμές, τόσο χαμηλές όσο και υψηλές, ιδιαίτερα στην ομάδα Yes, γεγονός που υποδηλώνει μεταβλητότητα στην περιφερειακή απόδοση. Οι ακραίες αυτές παρατηρήσεις δεν αλλοιώνουν τη γενική τάση αλλά αναδεικνύουν τη διαφορετικότητα μεταξύ των ομάδων.



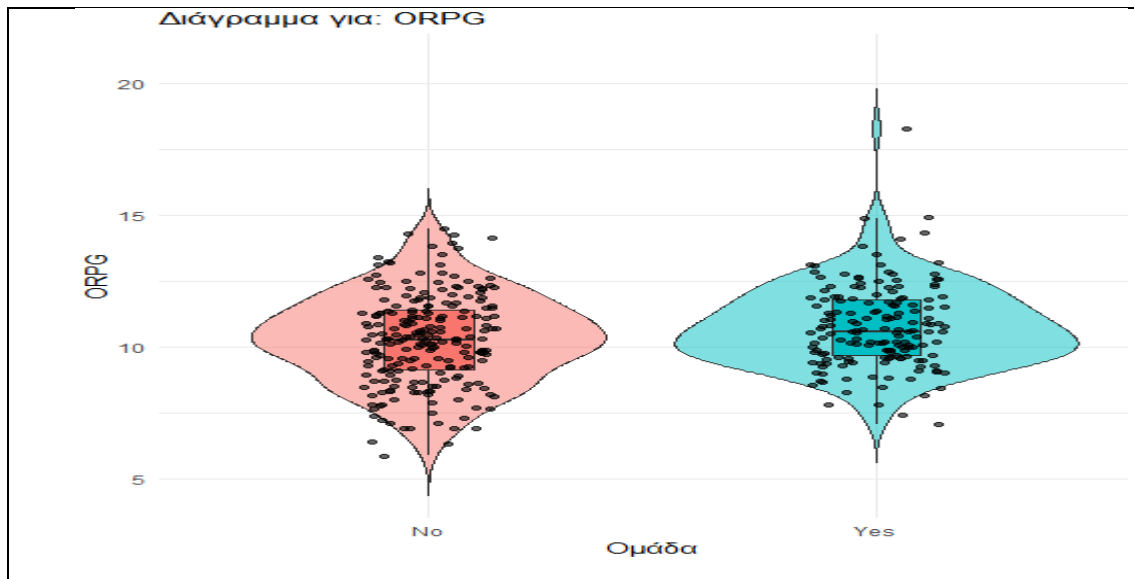
Σχήμα 4.4.2.3 Violin plot για 3P%.

Στο σχήμα 4.4.2.4 το ποσοστό ευστοχίας στις ελεύθερες βολές (FT%) δεν φαίνεται να σχετίζεται με την πρόκριση. Οι δύο ομάδες παρουσιάζουν σχεδόν ταυτόσημες κατανομές, χωρίς σαφή διαφοροποίηση. Παρά την ύπαρξη ορισμένων ακραίων τιμών και στις δύο κατηγορίες, αυτές κατανέμονται συμμετρικά και δεν επηρεάζουν τη συνολική εικόνα, επιβεβαιώνοντας ότι η συγκεκριμένη μεταβλητή δεν αποτελεί καθοριστικό παράγοντα επιτυχίας.



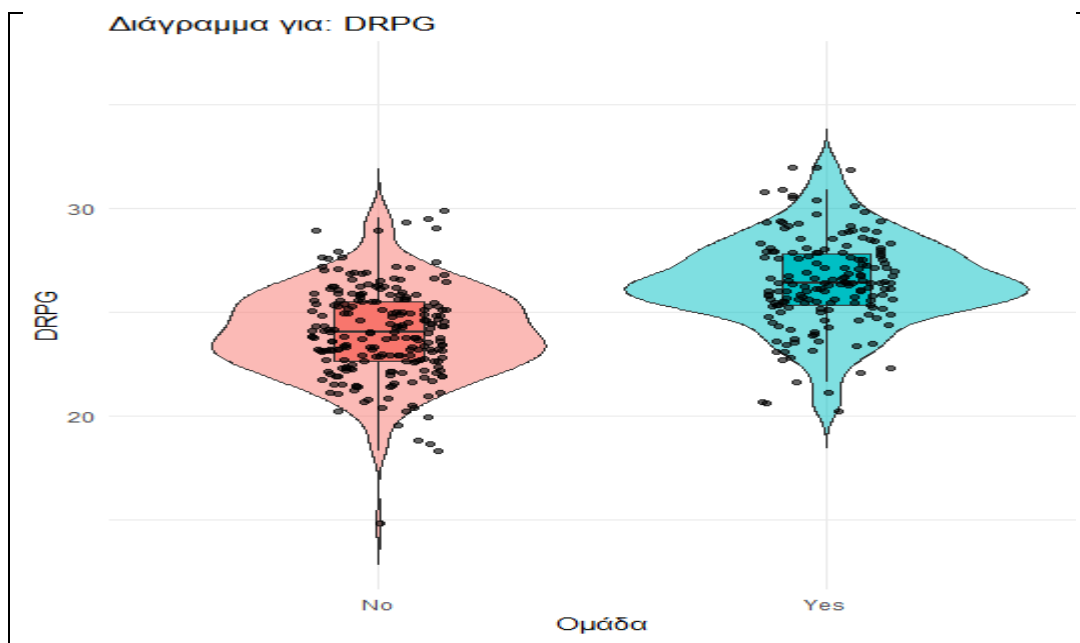
Σχήμα 4.4.2.4 Violin plot για FT %.

Στο σχήμα 4.4.2.5 παρουσιάζονται τα διαγράμματα για την μεταβλητή των επιθετικών ριμπάουντ (ORPG). Οι κατανομές των δύο ομάδων είναι αρκετά παρόμοιες, με μικρές αποκλίσεις. Αν και ορισμένες ομάδες με υψηλές τιμές δεν προκρίνονται, αυτό μπορεί να εξηγηθεί από το γεγονός ότι τα επιθετικά ριμπάουντ δεν επαρκούν από μόνα τους για τη συνολική επιτυχία. Παρατηρούνται ωστόσο ορισμένες υψηλές ακραίες τιμές, κυρίως στην ομάδα Yes, που ενδέχεται να αντιστοιχούν σε ομάδες με έντονη παρουσία στα επιθετικά ριμπάουντ. Παρόλα αυτά, η συνολική επίδραση της μεταβλητής παραμένει χαμηλή.



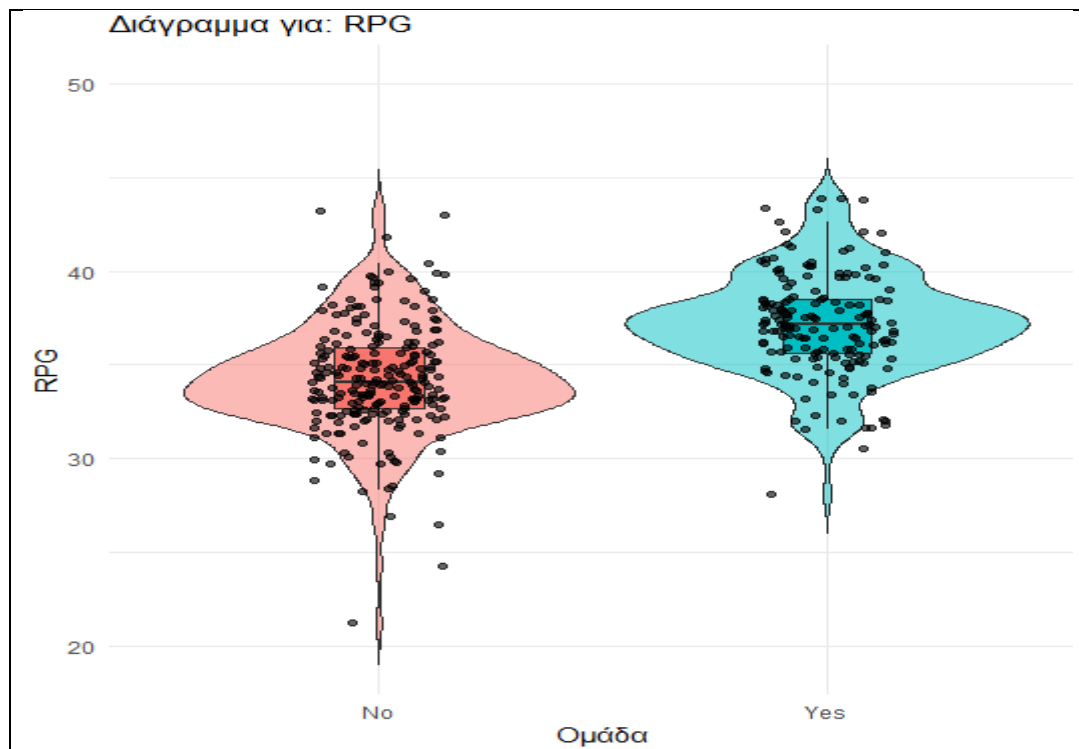
Σχήμα 4.4.2.5 Violin plot για ORPG

Στο σχήμα 4.4.2.6 τα αμυντικά ριμπάουντ (DRPG), οι ομάδες που προκρίθηκαν εμφανίζουν υψηλότερες τιμές και πιο συγκεντρωμένη κατανομή σε σχέση με αυτές που δεν προκρίθηκαν, γεγονός που αναδεικνύει τη σημασία της αμυντικής λειτουργίας. Ωστόσο, ακόμη και ομάδες με υψηλές τιμές μπορεί να μην προκριθούν, εφόσον υστερούν σε άλλες παραμέτρους, όπως η επιθετική αποτελεσματικότητα. Επιπλέον, υπάρχουν ορισμένες ακραίες τιμές, κυρίως χαμηλές στην ομάδα No και υψηλές στην ομάδα Yes, οι οποίες ενισχύουν τη διαφοροποίηση αλλά δεν επηρεάζουν την κύρια τάση.



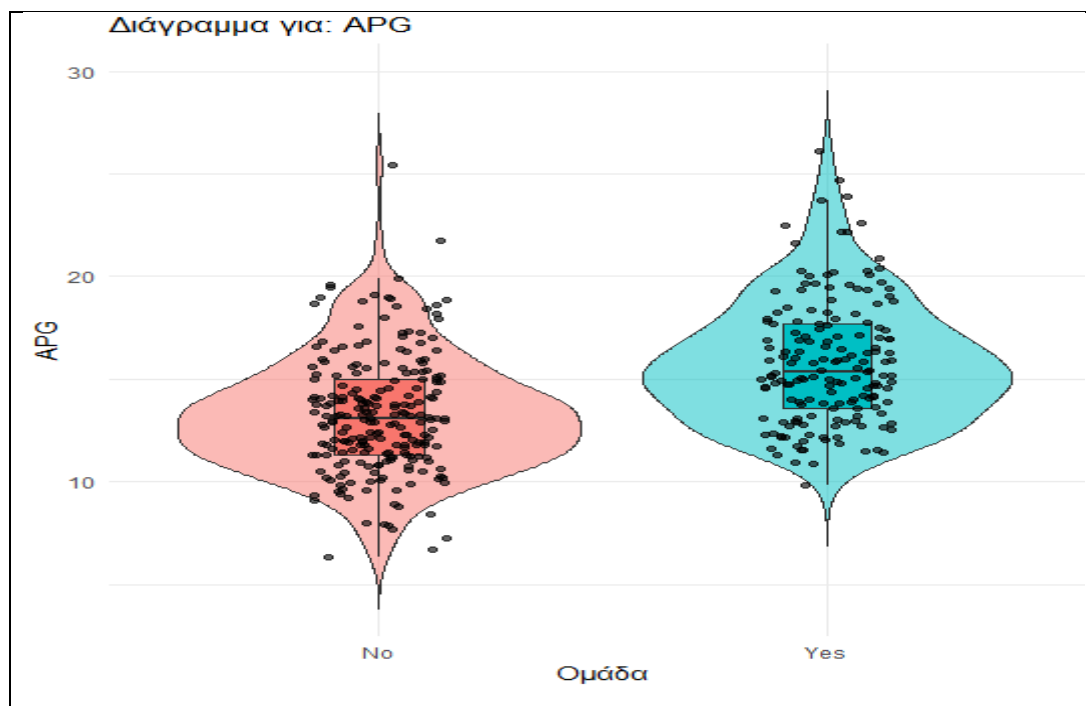
Σχήμα 4.4.2.6 Violin plot για DRPG

Στο σχήμα 4.4.2.7 για τη μεταβλητή RPG, παρατηρείται ότι οι ομάδες που εξασφάλισαν την πρόκριση εμφανίζουν υψηλότερες τιμές, ενώ οι ομάδες που δεν προκρίθηκαν συγκεντρώνονται σε χαμηλότερα επίπεδα. Παρατηρούνται ακραίες τιμές, κυρίως χαμηλές στην ομάδα No, οι οποίες υποδηλώνουν ομάδες με περιορισμένη ικανότητα στα ριμπάουντ. Ωστόσο, αυτές δεν μεταβάλλουν τη γενική εικόνα.



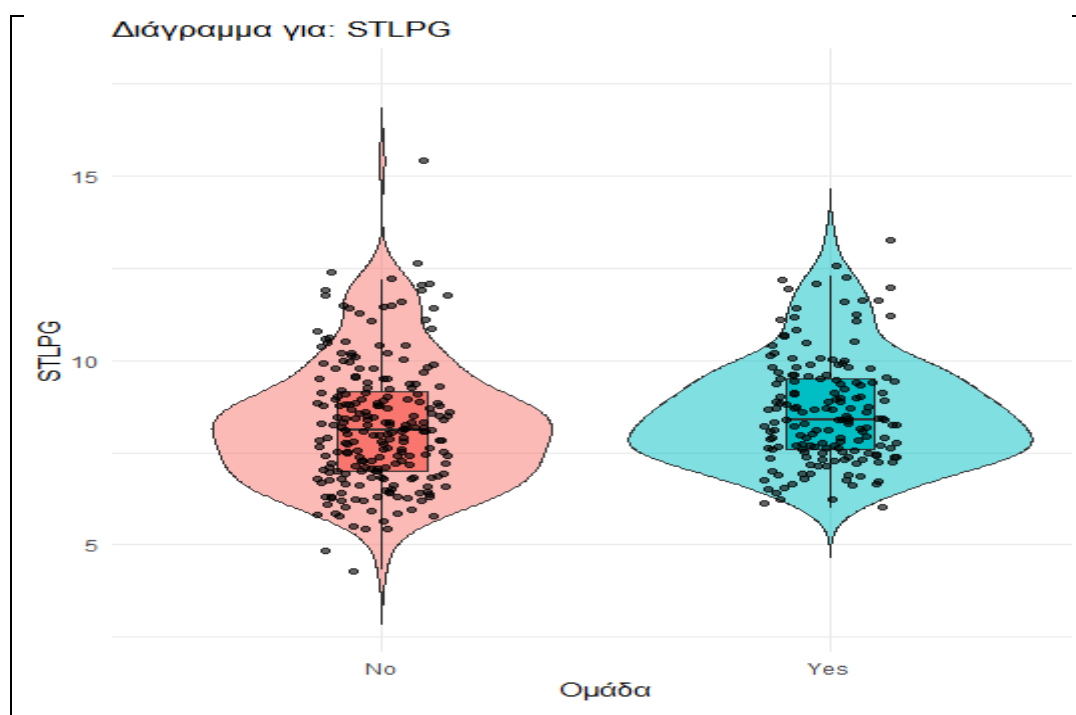
Σχήμα 4.4.2.7 Violin plot για RPG

Το σχήμα 4.4.2.8 αφορά τις ασίστ ανά αγώνα (APG). Οι ομάδες που προκρίθηκαν εμφανίζουν υψηλότερες τιμές, γεγονός που υποδηλώνει καλύτερη ομαδική συνεργασία και δημιουργία παιχνιδιού. Ωστόσο, ομάδες με υψηλό αριθμό ασίστ μπορεί να μην προκριθούν, εάν δεν συνδυάζουν τη δημιουργία με αποτελεσματική εκτέλεση ή αμυντική συνέπεια. Στο διάγραμμα εμφανίζονται ορισμένες ακραίες τιμές, κυρίως υψηλές στην ομάδα Yes, που πιθανόν αντιστοιχούν σε ομάδες με εξαιρετική κυκλοφορία της μπάλας. Οι τιμές αυτές ενισχύουν τη συνολική τάση χωρίς να την αλλοιώνουν.



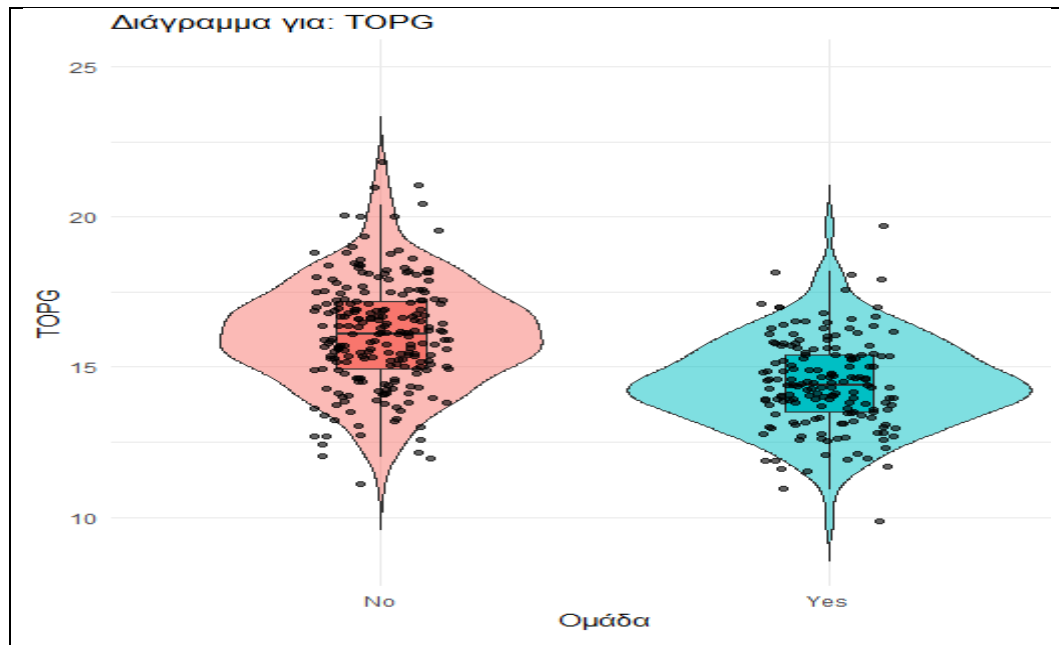
Σχήμα 4.4.2.8 Violin plot για APG

Στο σχήμα 4.4.2.9 για τη μεταβλητή STLPG οι δύο ομάδες εμφανίζουν παρόμοια κατανομή τιμών, γεγονός που υποδηλώνει ότι η συγκεκριμένη μεταβλητή δεν διαφοροποιεί ουσιαστικά την επιτυχία. Ομάδες με υψηλές τιμές ενδέχεται να μην προκριθούν, πιθανώς λόγω συνολικής αγωνιστικής αστάθειας. Υπάρχουν ωστόσο ορισμένες ακραίες τιμές, κυρίως υψηλές στην ομάδα No, οι οποίες πιθανόν αντιπροσωπεύουν ομάδες με έντονη αμυντική δραστηριότητα αλλά όχι απαραίτητα συνολική επιτυχία.



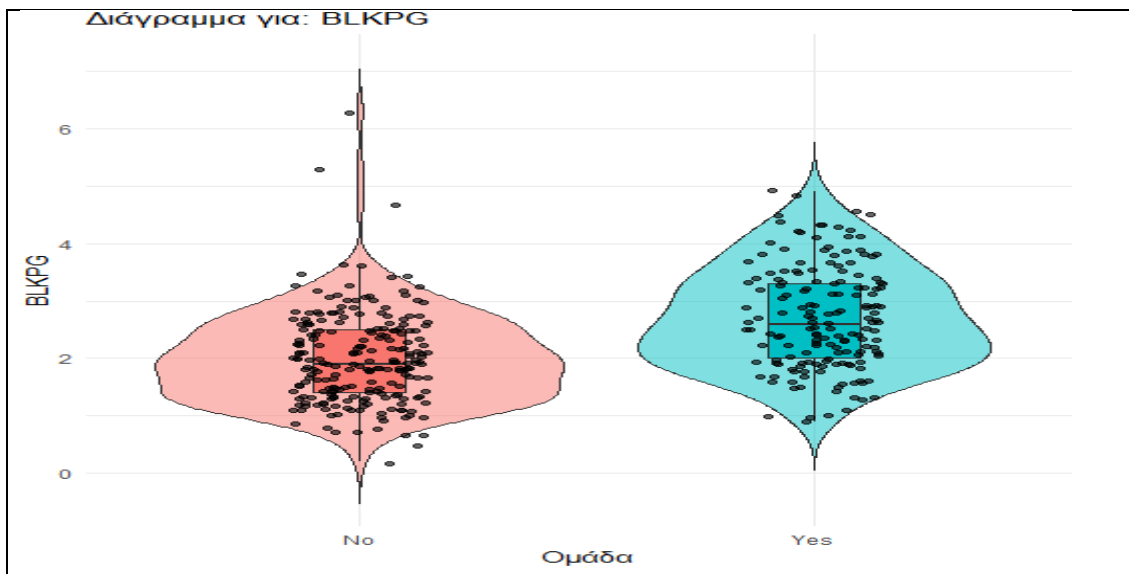
Σχήμα 4.4.2.9 Violin plot για STLPG

Στο σχήμα 4.4.2.10 για τα λάθη ανά αγώνα (TOPG), οι ομάδες που δεν προκρίθηκαν εμφανίζουν υψηλότερες τιμές λαθών, ενώ αυτές που προκρίθηκαν παρουσιάζουν χαμηλότερα επίπεδα. Παρόλα αυτά, ορισμένες ομάδες με σχετικά υψηλά λάθη μπορεί να προκριθούν, εφόσον υπερτερούν σε άλλους τομείς. Παρατηρούνται ακραίες τιμές κυρίως υψηλές στην ομάδα No, γεγονός που ενισχύει τη σημασία της αποφυγής λαθών για την επιτυχία.



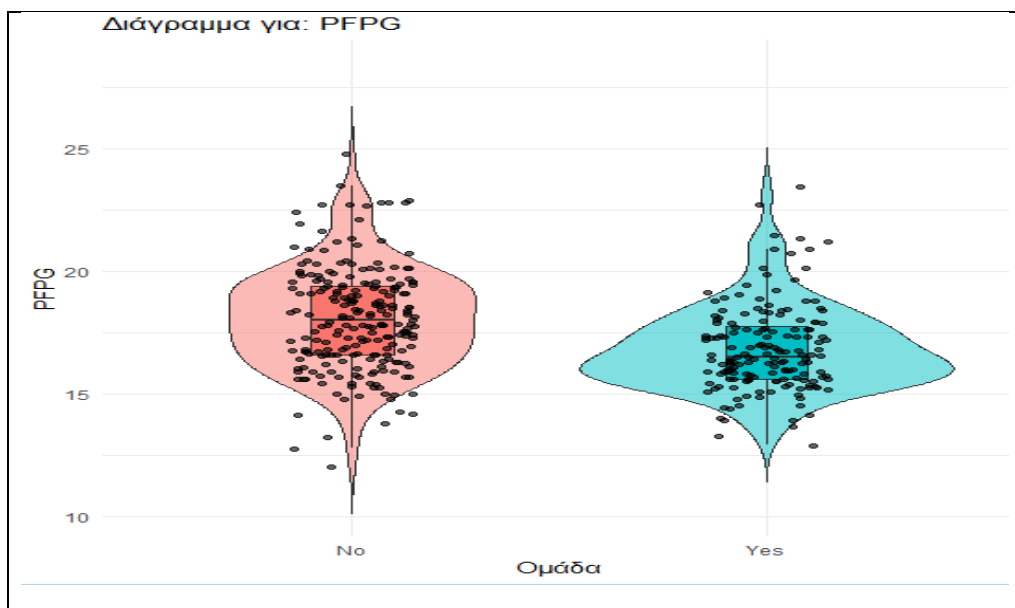
Σχήμα 4.4.2.10 Violin plot για TOPG

Στο σχήμα 4.4.2.11 παρουσιάζονται τα μπλοκ ανά αγώνα (BLKPG). Οι ομάδες που προκρίθηκαν εμφανίζουν υψηλότερες τιμές, γεγονός που υποδηλώνει ισχυρότερη αμυντική παρουσία κοντά στο καλάθι. Ωστόσο, υψηλές τιμές δεν εξασφαλίζουν πρόκριση, καθώς απαιτείται συνολική ισορροπία απόδοσης. Υπάρχουν ορισμένες ακραίες τιμές, κυρίως υψηλές στην ομάδα Yes, που πιθανόν αντιστοιχούν σε ομάδες με ιδιαίτερα έντονη παρουσία στην άμυνα. Οι τιμές αυτές δεν αλλοιώνουν τη συνολική εικόνα.



Σχήμα 4.4.2.11 Violin plot για BLKPG

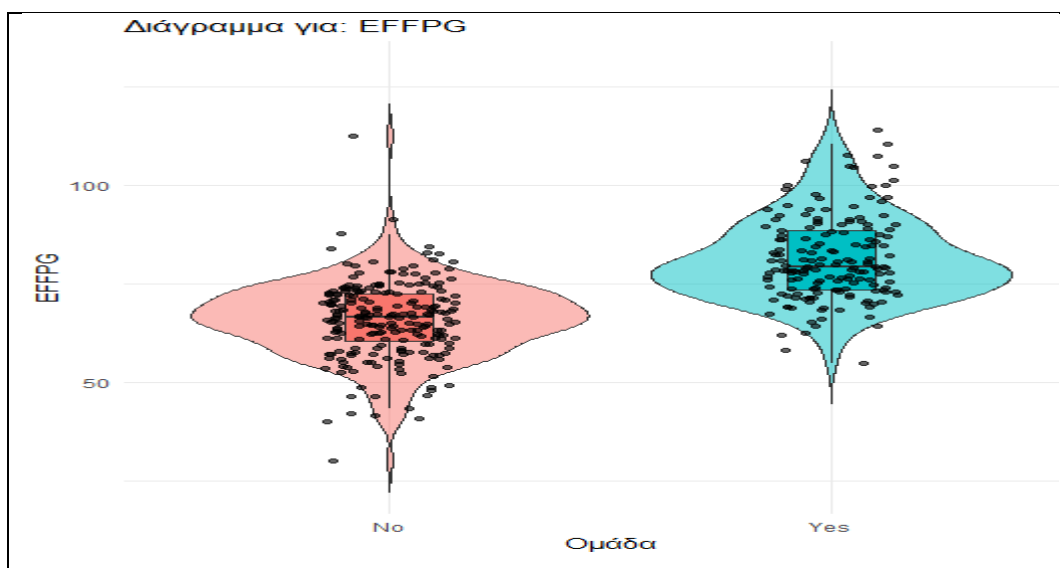
Στο σχήμα 4.4.2.12 είναι το διάγραμμα για τα προσωπικά φάουλ ανά αγώνα (PFPG). Οι ομάδες που δεν προκρίθηκαν παρουσιάζουν ελαφρώς υψηλότερες τιμές, γεγονός που υποδηλώνει ότι η αυξημένη παραβατικότητα ενδέχεται να επηρεάζει αρνητικά την απόδοση. Παρατηρούνται ακραίες τιμές κυρίως υψηλές στην ομάδα No, οι οποίες πιθανόν σχετίζονται με ομάδες που εμφανίζουν πιο επιθετικό ή ασταθές αμυντικό παιχνίδι.



Σχήμα 4.4.2.12 Violin plot για PFPG

Τέλος, στο σχήμα 4.4.2.13 η μεταβλητή EFFPG. Οι ομάδες που προκρίθηκαν εμφανίζουν σαφώς υψηλότερες τιμές, με την κατανομή τους να είναι μετατοπισμένη προς ανώτερα επίπεδα σε σχέση με τις ομάδες που αποκλείστηκαν, γεγονός που αναδεικνύει τη σημασία της συνολικής αγωνιστικής επίδοσης.

Παρόλα αυτά, παρατηρούνται και περιπτώσεις ομάδων με σχετικά υψηλή αποδοτικότητα που δεν προκρίνονται, γεγονός που δείχνει ότι ακόμη και ένας σύνθετος δείκτης όπως το EFFPG δεν επαρκεί από μόνος του για να εξηγήσει πλήρως την πρόκριση, καθώς αυτή εξαρτάται από πολλαπλούς παράγοντες και τη συνολική αγωνιστική συνέπεια. Επιπλέον, στο διάγραμμα εμφανίζονται ακραίες τιμές, κυρίως υψηλές στην ομάδα Yes αλλά και ορισμένες χαμηλές στην ομάδα No, οι οποίες αντανakλούν ιδιαίτερες περιπτώσεις ομάδων με εξαιρετικά υψηλή ή χαμηλή συνολική απόδοση· ωστόσο, οι τιμές αυτές δεν αλλοιώνουν τη γενική εικόνα της ισχυρής θετικής σχέσης μεταξύ της μεταβλητής και της πρόκρισης.



Σχήμα 4.4.2.13 Violin plot για EFFPG

4.4.3 Συντελεστής συσχέτισης

Ο συντελεστής point-biserial συσχέτισης αποτελεί ειδική περίπτωση μέτρου συσχέτισης που χρησιμοποιείται όταν μία μεταβλητή είναι συνεχής και η άλλη διχοτομική. Σύμφωνα με τον Bonett (2019), ο point-biserial αποτελεί ένα ευρέως χρησιμοποιούμενο μέτρο μεγέθους επίδρασης (effect size) σε σχέδια δύο ομάδων, καθώς εκφράζει τη σχέση μεταξύ μιας ποσοτικής μεταβλητής και μιας κατηγορικής μεταβλητής με δύο επίπεδα. Παράλληλα, συνδέεται άμεσα με τη διαφορά μέσων τιμών μεταξύ των δύο ομάδων, καθώς μπορεί να εκφραστεί μέσω της τυποποιημένης διαφοράς μέσων, γεγονός που τον καθιστά εναλλακτικό τρόπο παρουσίασης των αποτελεσμάτων ενός t-test. Επιπλέον, η χρήση του δεν περιορίζεται μόνο στον έλεγχο στατιστικής σημαντικότητας, αλλά επιτρέπει και την εκτίμηση του μεγέθους της σχέσης μεταξύ των μεταβλητών, προσφέροντας πιο ουσιαστική ερμηνεία των αποτελεσμάτων.

Στο πλαίσιο αυτό, ο point-biserial μπορεί να ερμηνευθεί ως δείκτης που αποτυπώνει τον βαθμό στον οποίο οι δύο ομάδες διαφέρουν ως προς τη μέση τιμή της συνεχούς μεταβλητής, εκφρασμένος όμως σε κλίμακα συσχέτισης.

Όπως αναφέρεται στη σχετική βιβλιογραφία (Bonett 2019), η τιμή του εξαρτάται από τη διαφορά των μέσων τιμών, τη διασπορά των δεδομένων και την αναλογία των παρατηρήσεων στις δύο ομάδες, γεγονός που σημαίνει ότι επηρεάζεται τόσο από την ένταση της σχέσης όσο και από τη δομή του δείγματος. Ο συντελεστής λαμβάνει τιμές από -1 έως 1 , όπου οι θετικές τιμές υποδηλώνουν ότι υψηλότερες τιμές της συνεχούς μεταβλητής συνδέονται με τη μία κατηγορία της διχοτομικής μεταβλητής, ενώ οι αρνητικές τιμές υποδηλώνουν την αντίστροφη σχέση.

Από τα αποτελέσματα του συντελεστή point-biserial που παρουσιάζονται στον Πίνακα 4.4.3.1 προκύπτει ότι η ισχυρότερη θετική συσχέτιση παρατηρείται στη μεταβλητή EFFPG ($r_{pb} = 0.602$), υποδηλώνοντας ισχυρή σχέση με τη δυαδική μεταβλητή συμμετοχής στο Final 8. Ισχυρές προς μέτριες θετικές συσχετίσεις εμφανίζουν επίσης οι μεταβλητές PPG ($r_{pb} = 0.488$), DRPG ($r_{pb} = 0.486$), RPG ($r_{pb} = 0.454$) και 2P% ($r_{pb} = 0.405$), γεγονός που υποδηλώνει ότι υψηλότερες τιμές αυτών των δεικτών συνδέονται με μεγαλύτερη πιθανότητα συμμετοχής στο Final 8. Μέτρια συσχέτιση καταγράφεται για τις μεταβλητές APG, BLKPG και 3P%, ενώ ασθενής συσχέτιση εμφανίζουν οι ORPG και STLPG. Αντίθετα, η μεταβλητή FT% παρουσιάζει πρακτικά μηδενική συσχέτιση ($r_{pb} = 0.022$), υποδηλώνοντας ότι δεν σχετίζεται ουσιαστικά με την έκβαση. Τέλος, αρνητική συσχέτιση παρατηρείται για τις μεταβλητές TOPG ($r_{pb} = -0.434$) και PFPG ($r_{pb} = -0.297$), γεγονός που δείχνει ότι υψηλότερες τιμές αυτών των μεταβλητών συνδέονται με μικρότερη πιθανότητα συμμετοχής στο Final 8. Τα αποτελέσματα αυτά είναι συνεπή με τα ευρήματα της γραφικής ανάλυσης, ενισχύοντας την αξιοπιστία των συμπερασμάτων.

Μεταβλητή	r_{pb}
PPG	0.488
2P%	0.405
3P%	0.372
FT%	0.022
ORPG	0.145
DRPG	0.486
RPG	0.454
APG	0.396
STLPG	0.112
TOPG	-0.434
BLKPG	0.386
PFPG	-0.297
EFFPG	0.602

Πίνακας 4.4.3.1 Point-biserial correlation

Κεφάλαιο 5 : Ανάλυση παλινδρόμησης

Το παρόν κεφάλαιο εξετάζει την αγωνιστική επιτυχία των ομάδων μέσω της εφαρμογής μοντέλων λογιστικής και πολυωνυμικής λογιστικής παλινδρόμησης, με σκοπό τη διερεύνηση των παραγόντων που επηρεάζουν την απόδοσή τους. Η ανάλυση βασίζεται σε ένα σύνολο βασικών αγωνιστικών δεικτών, οι οποίοι χρησιμοποιούνται ως επεξηγηματικές μεταβλητές για την ερμηνεία και πρόβλεψη της επιτυχίας. Αρχικά, εφαρμόζεται δυαδική λογιστική παλινδρόμηση, όπου η εξαρτημένη μεταβλητή εκφράζει την πιθανότητα πρόκρισης των ομάδων στο Final 8. Η προσέγγιση αυτή επιτρέπει τη διάκριση μεταξύ επιτυχίας και μη επιτυχίας, παρέχοντας μια πρώτη εκτίμηση των βασικών παραγόντων που συμβάλλουν στην πρόκριση. Στο πλαίσιο αυτό εξετάζονται η προσαρμογή του μοντέλου, η στατιστική σημαντικότητα των μεταβλητών και η προβλεπτική του ικανότητα μέσω κατάλληλων διαγνωστικών εργαλείων. Στη συνέχεια, εφαρμόζεται διαδικασία επιλογής μεταβλητών, με στόχο τη βελτίωση της ισορροπίας μεταξύ προσαρμογής και πολυπλοκότητας του μοντέλου, καθώς και τη μείωση φαινομένων πολυσυγγραμμικότητας. Μέσω της διαδικασίας αυτής προκύπτει ένα πιο συμπαγές και σταθερό μοντέλο, το οποίο διατηρεί τις σημαντικότερες μεταβλητές και επιτρέπει πιο αξιόπιστη ερμηνεία των αποτελεσμάτων. Ωστόσο, η δυαδική προσέγγιση δεν επαρκεί για την πλήρη αποτύπωση της αγωνιστικής επιτυχίας, καθώς δεν λαμβάνει υπόψη τα διαφορετικά επίπεδα επίδοσης των ομάδων. Για τον λόγο αυτό, η ανάλυση επεκτείνεται σε πολυωνυμική λογιστική παλινδρόμηση, όπου η εξαρτημένη μεταβλητή κατηγοριοποιείται σε τρία επίπεδα επιτυχίας («No», «Final8», «Final4»). Η προσέγγιση αυτή επιτρέπει την ταυτόχρονη σύγκριση πολλαπλών κατηγοριών και την πληρέστερη κατανόηση των διαφορών μεταξύ των επιπέδων απόδοσης. Τέλος, η προβλεπτική ικανότητα των μοντέλων αξιολογείται μέσω κατάλληλων δεικτών και εργαλείων, όπως η καμπύλη ROC και ο πίνακας ταξινόμησης, τα οποία επιτρέπουν την εκτίμηση της ακρίβειας των προβλέψεων τόσο συνολικά όσο και ανά κατηγορία. Συνολικά, η συνδυαστική εφαρμογή των δύο μεθοδολογιών παρέχει μια ολοκληρωμένη και ουσιαστική κατανόηση της επίδρασης των αγωνιστικών δεικτών στην επιτυχία των ομάδων.

5.1 Λογιστική παλινδρόμηση

5.1.1 Θεωρητικό υπόβαθρο

Η ανάλυση παλινδρόμησης αποτελεί ένα από τα σημαντικότερα εργαλεία της στατιστικής, με στόχο τη μελέτη της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και ενός συνόλου ανεξάρτητων μεταβλητών, καθώς και την πρόβλεψη της τιμής της εξαρτημένης μεταβλητής μέσω των ερμηνευτικών μεταβλητών (Κούτρας & Ευαγγελάρας, 2018). Οι μέθοδοι παλινδρόμησης χρησιμοποιούνται ευρέως για την κατανόηση των παραγόντων που επηρεάζουν ένα φαινόμενο και την ανάπτυξη μοντέλων πρόβλεψης.

Στην κλασική μορφή της, η γραμμική παλινδρόμηση εφαρμόζεται όταν η εξαρτημένη μεταβλητή είναι συνεχής και ακολουθεί κανονική κατανομή. Ωστόσο, σε πολλές εφαρμογές η μεταβλητή απόκρισης είναι διχοτομική, λαμβάνοντας τιμές όπως 0 και 1. Στις περιπτώσεις αυτές, η χρήση της γραμμικής παλινδρόμησης δεν είναι κατάλληλη, καθώς παραβιάζονται βασικές υποθέσεις του μοντέλου, όπως η κανονικότητα των καταλοίπων και η ομοσκεδαστικότητα (Μπερσίμης κ.α., 2021). Η λογιστική παλινδρόμηση αποτελεί τη βασική μεθοδολογική προσέγγιση για τη μοντελοποίηση δυαδικών δεδομένων. Στην περίπτωση αυτή, η εξαρτημένη μεταβλητή θεωρείται ότι ακολουθεί την κατανομή Bernoulli και η αναμενόμενη τιμή της ισούται με την πιθανότητα εμφάνισης του γεγονότος, δηλαδή $E(Y)=p$ (Μπερσίμης κ.α., 2021). Η άμεση γραμμική μοντελοποίηση της πιθανότητας δεν είναι κατάλληλη, καθώς μπορεί να οδηγήσει σε τιμές εκτός του διαστήματος $[0,1]$. Για τον λόγο αυτό χρησιμοποιείται η λογιστική συνάρτηση, η οποία μετασχηματίζει τον γραμμικό συνδυασμό των μεταβλητών σε μία μη γραμμική συνάρτηση που λαμβάνει τιμές μόνο μεταξύ 0 και 1. Η συνάρτηση αυτή δίνεται από τη σχέση:

$$p_i = \frac{e^{b_0 + b_1 x_i}}{1 + e^{b_0 + b_1 x_i}}$$

και μέσω μετασχηματισμού οδηγεί στη μορφή logit:

$$\log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1 x_i$$

Ο όρος $\frac{p_i}{1-p_i}$ αποτελεί τον λόγο πιθανοτήτων (odds) και εκφράζει τη σχετική πιθανότητα εμφάνισης του γεγονότος (Μπερσίμης κ.α., 2021). Ο λογαριθμικός μετασχηματισμός καθιστά τη σχέση γραμμική ως προς τις παραμέτρους, επιτρέποντας την εφαρμογή τεχνικών παλινδρόμησης. Η λογιστική παλινδρόμηση εντάσσεται στο πλαίσιο των Γενικευμένων Γραμμικών Μοντέλων, τα οποία αποτελούν γενίκευση της γραμμικής παλινδρόμησης και εφαρμόζονται όταν η εξαρτημένη μεταβλητή δεν ακολουθεί κανονική κατανομή.

Ένα Γενικευμένο Γραμμικό Μοντέλο αποτελείται από την κατανομή της μεταβλητής απόκρισης, τον γραμμικό προβλεπτικό όρο και τη συνάρτηση σύνδεσης (Πολίτης 2023). Στην περίπτωση της λογιστικής παλινδρόμησης, η κατανομή είναι Bernoulli και η συνάρτηση σύνδεσης είναι η logit. Η εκτίμηση των παραμέτρων του μοντέλου πραγματοποιείται με τη μέθοδο της μέγιστης πιθανοφάνειας, η οποία βασίζεται στη μεγιστοποίηση της πιθανότητας παρατήρησης των δεδομένων ως προς τις άγνωστες παραμέτρους. Η συνάρτηση πιθανοφάνειας εκφράζεται ως γινόμενο όρων της $p_i^{y_i} (1 - p_i)^{1-y_i}$, ενώ στην πράξη χρησιμοποιείται η λογαριθμική μορφή της (Μπερσίμης κ.α., 2021). Η εκτίμηση των παραμέτρων πραγματοποιείται μέσω επαναληπτικών διαδικασιών και οι εκτιμητές παρουσιάζουν σημαντικές ιδιότητες, όπως συνέπεια και ασυμπτωτική κανονικότητα (Πολίτης 2023). Για τον έλεγχο της στατιστικής σημαντικότητας των παραμέτρων χρησιμοποιείται ο έλεγχος Wald, καθώς και ο έλεγχος λόγου πιθανοφανειών, ο οποίος συγκρίνει εμφωλευμένα μοντέλα (Μπερσίμης κ.α., 2021; Πολίτης, 2023).

Η προσαρμογή του μοντέλου αξιολογείται μέσω της απόκλισης (deviance), ενώ χρησιμοποιούνται και δείκτες όπως οι συντελεστές προσδιορισμού των Cox και Snell και του Nagelkerke (Μπερσίμης κ.α., 2021). Η ερμηνεία των συντελεστών βασίζεται στους λόγους σχετικών πιθανοτήτων (odds ratios), οι οποίοι προκύπτουν από τον εκθετικό μετασχηματισμό των εκτιμημένων συντελεστών. Για παράδειγμα, μία τιμή odds ratio ίση με 1.2 αυξάνει τον λόγο σχετικών πιθανοτήτων κατά 20% ενώ τιμές μικρότερες της μονάδας υποδηλώνουν μείωση της πιθανότητας (Πολίτης, 2023). Θα πρέπει να σημειωθεί ότι το odds ratio αναφέρεται στον λόγο σχετικών πιθανοτήτων και όχι άμεσα στην πιθανότητα εμφάνισης του γεγονότος. Η εφαρμογή της λογιστικής παλινδρόμησης προϋποθέτει την ικανοποίηση ορισμένων βασικών υποθέσεων, όπως η ανεξαρτησία των παρατηρήσεων, η απουσία έντονης πολυσυγγραμμικότητας μεταξύ των ανεξάρτητων μεταβλητών και η ύπαρξη γραμμικής σχέσης μεταξύ των ανεξάρτητων μεταβλητών και του logit της εξαρτημένης μεταβλητής. Η παραβίαση των υποθέσεων αυτών μπορεί να οδηγήσει σε αναξιόπιστα αποτελέσματα. Η αξιολόγηση της απόδοσης του μοντέλου πραγματοποιείται και μέσω της ανάλυσης της ταξινόμησης των παρατηρήσεων, χρησιμοποιώντας πίνακες σύγχυσης (confusion matrices), οι οποίοι παρουσιάζουν τον αριθμό των σωστών και λανθασμένων προβλέψεων. Μέσω αυτών μπορούν να υπολογιστούν δείκτες όπως η ακρίβεια, η ευαισθησία και η ειδικότητα, οι οποίοι συμβάλλουν στην ολοκληρωμένη αξιολόγηση του μοντέλου. Παράλληλα, ιδιαίτερη σημασία έχει η καμπύλη ROC (Receiver Operating Characteristic), η οποία αποτελεί ένα από τα βασικότερα εργαλεία αξιολόγησης μοντέλων ταξινόμησης. Η καμπύλη ROC (Receiver Operating Characteristic) αποτελεί ένα γραφικό εργαλείο αξιολόγησης της διακριτικής ικανότητας ενός μοντέλου ταξινόμησης, απεικονίζοντας την ευαισθησία ως προς το ποσοστό των ψευδώς θετικών (1-ειδικότητα) για όλες τις δυνατές τιμές ορίου.

Ένα συνολικό μέτρο της απόδοσης του μοντέλου αποτελεί η επιφάνεια κάτω από την καμπύλη ROC (Area Under the Curve – AUC), η οποία συνοψίζει την ικανότητα του μοντέλου να διαχωρίζει σωστά τις δύο κατηγορίες. Τέλος, αξίζει να σημειωθεί ότι η λογιστική παλινδρόμηση αποτελεί μία από τις πλέον διαδεδομένες μεθόδους ανάλυσης δυαδικών δεδομένων, με εφαρμογές σε πολλούς τομείς, όπως η ιατρική, η οικονομία και η αθλητική ανάλυση.

Στην επόμενη ενότητα παρουσιάζεται η εφαρμογή της μεθόδου στα δεδομένα της παρούσας μελέτης, με στόχο τη διερεύνηση των παραγόντων που επηρεάζουν την πρόκριση των ομάδων.

5.1.2 Μοντέλο λογιστικής παλινδρόμησης

Το τελικό μοντέλο λογιστικής παλινδρόμησης εκτιμήθηκε με σκοπό τη διερεύνηση της επίδρασης βασικών αγωνιστικών δεικτών στην πιθανότητα συμμετοχής μιας ομάδας στο Final 8. Τα αποτελέσματα φαίνονται στον πίνακα 5.1.2.1. Ως εξαρτημένη μεταβλητή χρησιμοποιήθηκε η δυαδική μεταβλητή πρόκρισης (Final8 Status), η οποία λαμβάνει την τιμή 1 όταν μια ομάδα προκρίνεται στο Final 8 και την τιμή 0 όταν δεν προκρίνεται. Ως ανεξάρτητες μεταβλητές συμπεριλήφθηκαν οι δείκτες 2P%, 3P%, FT%, APG, STLPG, TOPG και BLKPG.

Η διαδικασία που ακολουθήθηκε για την επιλογή του τελικού υποδείγματος πραγματοποιήθηκε σε διαδοχικά στάδια. Αρχικά εκτιμήθηκε το πλήρες μοντέλο, το οποίο περιλάμβανε το σύνολο των διαθέσιμων επεξηγηματικών μεταβλητών. Στη συνέχεια εφαρμόστηκε η μέθοδος stepwise επιλογής μεταβλητών με βάση το κριτήριο Akaike (AIC), με στόχο τη βελτίωση της ισορροπίας μεταξύ προσαρμογής και πολυπλοκότητας του μοντέλου. Παρότι η stepwise διαδικασία οδήγησε σε μοντέλο με χαμηλότερο AIC, η ερμηνεία των συντελεστών παρουσίασε σημαντικά προβλήματα, καθώς παρατηρήθηκαν μη αναμενόμενα πρόσημα και ενδείξεις πολυσυγγραμμικότητας. Για τον λόγο αυτό επιλέχθηκε μια χειροκίνητη διαδικασία προσαρμογής του μοντέλου, κατά την οποία αφαιρέθηκαν μεταβλητές με υψηλή συσχέτιση και εξετάστηκαν εναλλακτικοί συνδυασμοί, με στόχο τη διαμόρφωση ενός μοντέλου που να είναι ταυτόχρονα στατιστικά επαρκές και ερμηνευτικά σαφές. Το τελικό μοντέλο παρουσιάζει σημαντική βελτίωση σε σχέση με το μηδενικό μοντέλο (Residual deviance: 335.84 έναντι Null deviance: 534.28) και ικανοποιητική τιμή AIC (351.84), γεγονός που υποδηλώνει καλή προσαρμογή. Τα αποτελέσματα δείχνουν ότι τα ποσοστά ευστοχίας στα δίποντα (2P%) και στα τρίποντα (3P%) επηρεάζουν θετικά την πιθανότητα πρόκρισης, όπως και οι ασίστ (APG), τα κλεψίματα (STLPG) και τα κοψίματα (BLKPG), τα οποία συνδέονται με την επιθετική δημιουργία και την αμυντική αποτελεσματικότητα των ομάδων. Αντίθετα, τα λάθη (TOPG) εμφανίζουν αρνητική επίδραση, γεγονός που είναι αναμενόμενο, καθώς μειώνουν τον αριθμό των επιθετικών κατοχών. Το ποσοστό ευστοχίας στις ελεύθερες βολές (FT%) εμφανίζεται στατιστικά σημαντικό αλλά με αρνητικό πρόσημο.

Το αποτέλεσμα αυτό, αν και εκ πρώτης όψεως αντιφατικό, δεν θα πρέπει να ερμηνευθεί ως ένδειξη ότι η βελτίωση στις ελεύθερες βολές μειώνει την πιθανότητα πρόκρισης. Πιθανότερα αντανακλά την επίδραση της ταυτόχρονης παρουσίας άλλων επιθετικών δεικτών στο μοντέλο και πιθανές συσχετίσεις μεταξύ των μεταβλητών. Επιπλέον, το FT% ενδέχεται να λειτουργεί ως δευτερεύων δείκτης απόδοσης σε σχέση με πιο καθοριστικούς παράγοντες, όπως η συνολική επιθετική αποτελεσματικότητα και η δημιουργία φάσεων, με αποτέλεσμα η επίδρασή του να αποτυπώνεται αρνητικά όταν οι υπόλοιπες μεταβλητές διατηρούνται σταθερές. Συνολικά, το τελικό μοντέλο θεωρείται αξιόπιστο και ερμηνεύσιμο, καθώς περιορίζει τα προβλήματα πολυσυγγραμμικότητας και παρέχει αποτελέσματα συμβατά με την αγωνιστική πραγματικότητα.

```

Call:
glm(formula = Final8_Status ~ `2P%` + `3P%` + `FT%` + APG + STLPG +
    TOPG + BLKPG, family = binomial, data = data)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.70943      3.52357  -1.337 0.181369
`2P%`        0.16285      0.04889   3.331 0.000866 ***
`3P%`        0.15172      0.03965   3.827 0.000130 ***
`FT%`       -0.06700      0.03214  -2.085 0.037112 *
APG          0.13777      0.05599   2.461 0.013871 *
STLPG        0.22716      0.09051   2.510 0.012082 *
TOPG        -0.57495      0.09309  -6.176 6.58e-10 ***
BLKPG        0.78283      0.16830   4.652 3.30e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 534.28  on 390  degrees of freedom
Residual deviance: 335.84  on 383  degrees of freedom
AIC: 351.84

Number of Fisher Scoring iterations: 5

```

Πίνακας 5.1.2.1 Μοντέλο παλινδρόμησης

Συμπληρωματικά προς τα παραπάνω, η εκτίμηση των συντελεστών του μοντέλου επιτρέπει τη διατύπωση της εξίσωσης της λογιστικής παλινδρόμησης, η οποία δίνεται από τη σχέση:

$$\text{logit}(p) = -4.709 + 0.163 \cdot 2P\% + 0.152 \cdot 3P\% - 0.067 \cdot FT\% + 0.138 \cdot APG + 0.227 \cdot STLPG - 0.575 \cdot TOPG + 0.783 \cdot BLKPG.$$

Η εξίσωση αυτή αποτυπώνει ποσοτικά την επίδραση κάθε μεταβλητής στην πιθανότητα πρόκρισης, διατηρώντας σταθερές τις υπόλοιπες. Ειδικότερα, οι θετικοί συντελεστές των 2P% και 3P% επιβεβαιώνουν ότι η αυξημένη επιθετική αποτελεσματικότητα ενισχύει την πιθανότητα συμμετοχής στο Final 8, ενώ οι θετικές επιδράσεις των APG και STLPG αναδεικνύουν τη σημασία της δημιουργίας και της αμυντικής δραστηριότητας.

Η μεταβλητή BLKPG παρουσιάζει την ισχυρότερη θετική επίδραση, γεγονός που υποδηλώνει τον καθοριστικό ρόλο της αμυντικής παρουσίας κοντά στο καλάθι. Αντίθετα, ο αρνητικός συντελεστής της TOPG επιβεβαιώνει ότι τα λάθη μειώνουν σημαντικά την πιθανότητα επιτυχίας. Συνολικά, η κατεύθυνση και το μέγεθος των συντελεστών είναι συνεπή με τα προηγούμενα ευρήματα, ενισχύοντας τη συνοχή και την αξιοπιστία της ανάλυσης.

Η ερμηνεία των συντελεστών του τελικού μοντέλου βασίζεται στους εκθετικούς συντελεστές (odds ratios), οι οποίοι εκφράζουν τη σχετική μεταβολή στον λόγο πιθανοτήτων (odds) συμμετοχής στο Final 8 για κάθε μονάδα αύξησης των εξηγηματικών μεταβλητών, διατηρώντας τις υπόλοιπες σταθερές. Παράλληλα, τα 95% διαστήματα εμπιστοσύνης επιτρέπουν την αξιολόγηση της στατιστικής σημαντικότητας: όταν δεν περιλαμβάνουν τη μονάδα (1), η επίδραση θεωρείται στατιστικά σημαντική. Τα αποτελέσματα του Πίνακα 5.1.2.2 δείχνουν ότι το ποσοστό ευστοχίας στα δίποντα (2P%) έχει odds ratio ίσο με 1.18 (95% ΔΕ: [1.07, 1.30]).

Αυτό σημαίνει ότι για κάθε μονάδα αύξησης του δείκτη, ο λόγος πιθανοτήτων συμμετοχής στο Final 8 αυξάνεται κατά περίπου 18%, γεγονός που υποδηλώνει ουσιαστική και στατιστικά σημαντική θετική επίδραση. Αντίστοιχα, το 3P% παρουσιάζει odds ratio 1.16 (95% ΔΕ: [1.08, 1.26]), δηλαδή αύξηση των odds κατά 16%, επιβεβαιώνοντας τη σημασία της περιφερειακής ευστοχίας. Οι ασίστ (APG) εμφανίζουν odds ratio 1.15 (95% ΔΕ: [1.03, 1.28]), γεγονός που αντιστοιχεί σε αύξηση περίπου 15% στον λόγο σχετικών πιθανοτήτων συμμετοχής, υπογραμμίζοντας τη συμβολή της δημιουργίας φάσεων. Τα κλεψίματα (STLPG) παρουσιάζουν odds ratio 1.26 (95% ΔΕ: [1.05, 1.50]), δηλαδή αύξηση της τάξης του 26%, γεγονός που καταδεικνύει τη σημασία της αμυντικής πίεσης. Ακόμη πιο έντονη είναι η επίδραση των κοψιμάτων (BLKPG), με odds ratio 2.19 (95% ΔΕ: [1.59, 3.08]), που σημαίνει ότι ο λόγος σχετικών πιθανοτήτων υπερδιπλασιάζεται (αύξηση περίπου 119%) για κάθε μονάδα αύξησης της μεταβλητής, αναδεικνύοντας την ιδιαίτερα ισχυρή επίδραση της αμυντικής παρουσίας κοντά στο καλάθι. Αντίθετα, τα λάθη (TOPG) παρουσιάζουν odds ratio 0.56 (95% ΔΕ: [0.47, 0.67]), γεγονός που σημαίνει ότι για κάθε επιπλέον λάθος ο λόγος πιθανοτήτων συμμετοχής μειώνεται κατά περίπου 44%. Η τιμή αυτή, σε συνδυασμό με το διάστημα εμπιστοσύνης που βρίσκεται εξ ολοκλήρου κάτω από τη μονάδα, καταδεικνύει ισχυρή και στατιστικά σημαντική αρνητική επίδραση. Τέλος, το ποσοστό ελευθέρων βολών (FT%) εμφανίζει odds ratio 0.94 (95% ΔΕ: [0.88, 1.00]), που αντιστοιχεί σε μείωση περίπου 6% στον λόγο σχετικών πιθανοτήτων συμμετοχής για κάθε μονάδα αύξησης. Ωστόσο, το διάστημα εμπιστοσύνης περιλαμβάνει οριακά τη μονάδα, γεγονός που υποδηλώνει ότι η επίδραση δεν είναι στατιστικά σημαντική και θα πρέπει να ερμηνευθεί με προσοχή. Συνολικά, η ανάλυση τόσο των τιμών των odds ratios όσο και των διαστημάτων εμπιστοσύνης επιβεβαιώνει ότι η επιθετική αποτελεσματικότητα, η αμυντική δραστηριότητα και η αποφυγή λαθών αποτελούν τους βασικούς παράγοντες που επηρεάζουν την πιθανότητα συμμετοχής στο Final 8, με τα αποτελέσματα να παρουσιάζουν υψηλό βαθμό συνοχής με την προηγούμενη ανάλυση του μοντέλου.

Μεταβλητή	Odds Ratio	2.5%	97.5%
Intercept	0.00900903	0.000008272	8.6049182
2P%	1.176862246	1.071260	1.2983774
3P%	1.163839760	1.079112	1.2611313
FT%	0.935196785	0.8772515	0.9954271
APG	1.147711679	1.029561	1.2831339
STLPG	1.255035010	1.052676	1.5025233
TOPG	0.562734880	0.4651452	0.6706962
BLKPG	2.187656754	1.588179	3.0807500

Πίνακας 5.1.2.2 Εκθετικοί συντελεστές (odds ratios) και διαστήματα εμπιστοσύνης

5.1.3 Πολυσυγγραμμικότητα

5.1.3.1 Θεωρητικό υπόβαθρο

Η πολυσυγγραμμικότητα (multicollinearity) αναφέρεται στην ύπαρξη ισχυρής γραμμικής σχέσης μεταξύ δύο ή περισσότερων ανεξάρτητων μεταβλητών σε ένα μοντέλο παλινδρόμησης. Όπως αναφέρεται στη σχετική βιβλιογραφία (Κούτρας & Ευαγγελάρας, 2018), το φαινόμενο αυτό εμφανίζεται συχνά σε εφαρμογές όπου οι ερμηνευτικές μεταβλητές σχετίζονται μεταξύ τους, όπως για παράδειγμα σε οικονομικά, ιατρικά ή κοινωνικά δεδομένα. Η ύπαρξη πολυσυγγραμμικότητας δεν παραβιάζει απαραίτητα τις βασικές υποθέσεις του μοντέλου, ωστόσο δημιουργεί σημαντικά προβλήματα στην εκτίμηση και ερμηνεία των παραμέτρων. Συγκεκριμένα, όταν οι ανεξάρτητες μεταβλητές παρουσιάζουν υψηλή συσχέτιση, καθίσταται δύσκολο να προσδιοριστεί η μεμονωμένη επίδραση κάθε μεταβλητής στην εξαρτημένη. Αυτό έχει ως αποτέλεσμα οι εκτιμήσεις των συντελεστών να εμφανίζουν μεγάλη διακύμανση και να είναι ασταθείς, ενώ μικρές αλλαγές στα δεδομένα μπορεί να οδηγήσουν σε σημαντικές μεταβολές των εκτιμήσεων. Επιπλέον, η πολυσυγγραμμικότητα μπορεί να οδηγήσει σε φαινόμενα όπου μεταβλητές που φαίνονται σημαντικές σε απλά μοντέλα καθίστανται μη στατιστικά σημαντικές σε πολλαπλά μοντέλα, παρά το γεγονός ότι το συνολικό μοντέλο εμφανίζει υψηλή προσαρμογή. Το φαινόμενο αυτό αναδεικνύει τη δυσκολία ερμηνείας των συντελεστών όταν υπάρχει έντονη συσχέτιση μεταξύ των ερμηνευτικών μεταβλητών. Ένα βασικό χαρακτηριστικό της πολυσυγγραμμικότητας είναι ότι, παρότι επηρεάζει σημαντικά τους επιμέρους συντελεστές, δεν μειώνει απαραίτητα την προβλεπτική ικανότητα του μοντέλου. Αντίθετα, μπορεί να παρατηρηθεί υψηλή τιμή του συντελεστή προσδιορισμού R^2 , γεγονός που μπορεί να δημιουργήσει παραπλανητική εικόνα σχετικά με την ποιότητα του μοντέλου.

Για τη διάγνωση της πολυσυγγραμμικότητας χρησιμοποιούνται διάφορα κριτήρια, με σημαντικότερο τον δείκτη διόγκωσης της διακύμανσης (Variance Inflation Factor – VIF). Ο δείκτης αυτός ορίζεται ως:

$$VIF_k = \frac{1}{1 - R_k^2}$$

όπου R_k^2 είναι ο συντελεστής προσδιορισμού που προκύπτει από την παλινδρόμηση της μεταβλητής X_k ως προς τις υπόλοιπες ανεξάρτητες μεταβλητές του μοντέλου. Τιμές του δείκτη VIF κοντά στη μονάδα υποδηλώνουν απουσία πολυσυγγραμμικότητας, ενώ τιμές μεγαλύτερες του 10 αποτελούν ένδειξη σοβαρού προβλήματος. Σε τέτοιες περιπτώσεις, συνιστάται η απομάκρυνση ή ο ανασχηματισμός των μεταβλητών που προκαλούν το φαινόμενο, ώστε να βελτιωθεί η σταθερότητα και η ερμηνευσιμότητα του μοντέλου.

Συνολικά, η πολυσυγγραμμικότητα αποτελεί ένα σημαντικό ζήτημα στη μοντελοποίηση, καθώς μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα ως προς τη σημασία των μεταβλητών, καθιστώντας αναγκαίο τον έλεγχο της πριν την τελική ερμηνεία των αποτελεσμάτων.

5.1.3.2 Έλεγχος πολυσυγγραμμικότητας

Ο έλεγχος πολυσυγγραμμικότητας για το τελικό μοντέλο πραγματοποιήθηκε με τη χρήση του δείκτη διόγκωσης διακύμανσης (Variance Inflation Factor – VIF), προκειμένου να διαπιστωθεί εάν παραμένουν ισχυρές γραμμικές σχέσεις μεταξύ των επεξηγηματικών μεταβλητών που θα μπορούσαν να επηρεάσουν την αξιοπιστία των εκτιμήσεων. Τα αποτελέσματα παρουσιάζονται στον πίνακα 5.1.3.2 και δείχνουν ότι όλες οι μεταβλητές του τελικού μοντέλου παρουσιάζουν πολύ χαμηλές τιμές VIF, συγκεκριμένα: 2P% (≈ 1.21), 3P% (≈ 1.07), FT% (≈ 1.16), APG (≈ 1.20), STLPG (≈ 1.11), TOPG (≈ 1.06) και BLKPG (≈ 1.02). Οι τιμές αυτές βρίσκονται πολύ κοντά στη μονάδα και πολύ χαμηλότερα από τα συνήθη όρια ανησυχίας ($VIF > 5$ ή $VIF > 10$), γεγονός που υποδηλώνει την πρακτική απουσία πολυσυγγραμμικότητας στο τελικό υπόδειγμα. Σε αντίθεση με το αρχικό πλήρες μοντέλο, όπου παρατηρούνταν έντονες αλληλεξαρτήσεις μεταξύ μεταβλητών (ιδιαίτερα μεταξύ των δεικτών ριμπάουντ και των σύνθετων μεταβλητών), το τελικό μοντέλο αποτελείται από μεταβλητές που παρέχουν σε μεγάλο βαθμό ανεξάρτητη πληροφορία. Η σημαντική αυτή μείωση της πολυσυγγραμμικότητας επιβεβαιώνει την αποτελεσματικότητα της διαδικασίας επιλογής μεταβλητών που ακολουθήθηκε, καθώς οδήγησε σε ένα πιο σταθερό και αξιόπιστο μοντέλο. Επιπλέον, η απουσία ισχυρών συσχετίσεων μεταξύ των επεξηγηματικών μεταβλητών ενισχύει την ερμηνευσιμότητα των συντελεστών, επιτρέποντας την απομόνωση της επίδρασης κάθε μεταβλητής στην πιθανότητα πρόκρισης.

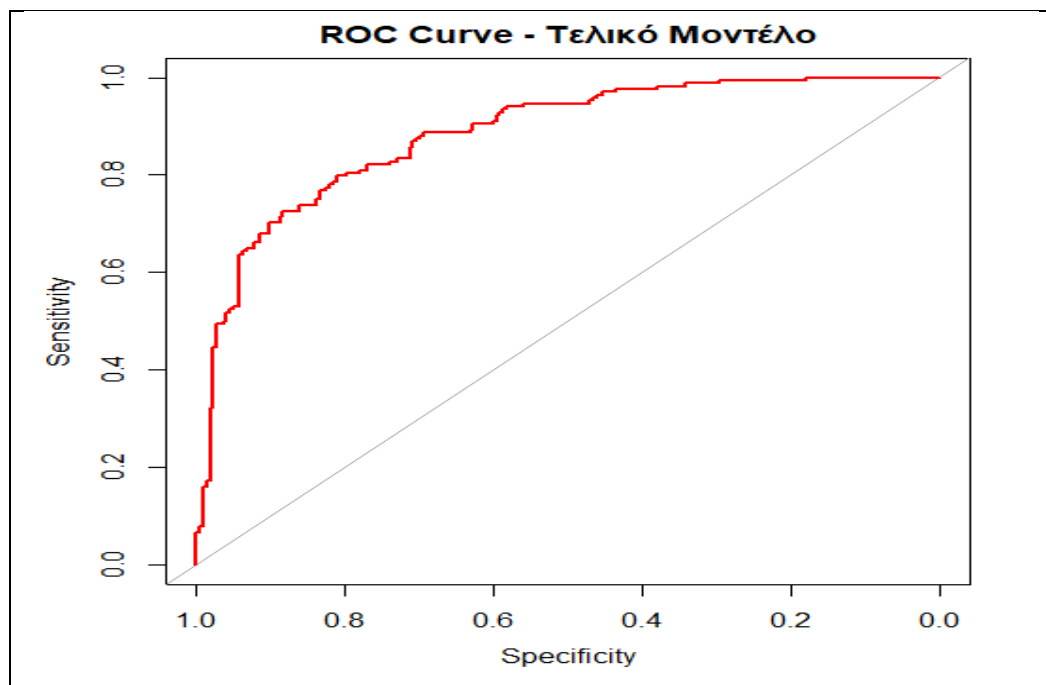
Συνολικά, το τελικό μοντέλο παρουσιάζει άριστη συμπεριφορά ως προς την πολυσυγγραμμικότητα, γεγονός που ενισχύει την εγκυρότητα των εκτιμήσεων και καθιστά τα αποτελέσματα πιο αξιόπιστα και ερμηνεύσιμα σε σύγκριση με το αρχικό υπόδειγμα.

Μεταβλητή	Τιμή VIF
2P%	1.21
3P%	1.07
FT%	1.16
APG	1.20
STLPG	1.11
TOPG	1.06
BLKPG	1.02

Πίνακας 5.1.3.2.α Έλεγχος πολυσυγγραμμικότητας

5.1.4 Γραφική απεικόνιση

Η καμπύλη ROC (Receiver Operating Characteristic) χρησιμοποιείται για την αξιολόγηση της διακριτικής ικανότητας του τελικού μοντέλου ως προς την πρόβλεψη της πρόκρισης στο Final 8. Στο σχήμα 5.1.4 απεικονίζεται η ευαισθησία (Sensitivity) σε συνάρτηση με το 1-ειδικότητα ($1 - \text{Specificity}$) για όλες τις πιθανές τιμές ταξινόμησης. Η ROC καμπύλη παρέχει μια συνολική αποτίμηση της απόδοσης του μοντέλου, ανεξάρτητα από την επιλογή συγκεκριμένου κατωφλίου, καθώς αποτυπώνει τη συμπεριφορά του ταξινομητή σε όλο το εύρος των πιθανών αποφάσεων (Fawcett, 2006). Παρατηρείται ότι η καμπύλη βρίσκεται σημαντικά πάνω από τη διαγώνιο γραμμή αναφοράς (γραμμή τυχαίας ταξινόμησης), γεγονός που υποδηλώνει ότι το μοντέλο διαθέτει ικανοποιητική έως υψηλή διακριτική ικανότητα. Επιπλέον, η καμπύλη προσεγγίζει την άνω αριστερή γωνία του διαγράμματος, κάτι που υποδηλώνει ότι επιτυγχάνει υψηλά επίπεδα ευαισθησίας με ταυτόχρονα χαμηλά επίπεδα ψευδώς θετικών αποτελεσμάτων. Η μορφή αυτή αντανάκλα την ικανότητα του μοντέλου να κατατάσσει σωστά τις παρατηρήσεις, αποδίδοντας υψηλότερες πιθανότητες στις ομάδες που προκρίνονται σε σχέση με εκείνες που δεν προκρίνονται. Συνολικά, η μορφή της καμπύλης δείχνει ότι το τελικό μοντέλο παρουσιάζει καλή ικανότητα διάκρισης μεταξύ ομάδων που προκρίνονται και ομάδων που δεν προκρίνονται στο Final 8, ενισχύοντας την αξιοπιστία του ως εργαλείο πρόβλεψης.



Σχήμα 5.1.4 Καμπύλη ROC

Ένα συνολικό μέτρο της απόδοσης του μοντέλου αποτελεί η επιφάνεια κάτω από την καμπύλη ROC (Area Under the Curve – AUC), η οποία λαμβάνει τιμές μεταξύ 0 και 1. Τιμές κοντά στο 0.5 υποδηλώνουν τυχαία ταξινόμηση, ενώ τιμές κοντά στο 1 υποδηλώνουν υψηλή διακριτική ικανότητα. Τιμές μεγαλύτερες του 0.7 θεωρούνται αποδεκτές, ενώ τιμές άνω του 0.8 υποδηλώνουν πολύ καλή έως εξαιρετική απόδοση του μοντέλου (Πολίτης, 2023; Hosmer & Lemeshow, 2013). Επιπλέον, το AUC μπορεί να ερμηνευτεί ως η πιθανότητα ένα τυχαία επιλεγμένο στοιχείο της θετικής κατηγορίας να λάβει μεγαλύτερη προβλεπόμενη πιθανότητα από ένα στοιχείο της αρνητικής κατηγορίας. Στην παρούσα ανάλυση, η τιμή του AUC για το τελικό μοντέλο ανέρχεται σε 0.8836, γεγονός που υποδηλώνει πολύ καλή ικανότητα διάκρισης μεταξύ των ομάδων που προκρίνονται και εκείνων που δεν προκρίνονται στο Final 8. Η τιμή αυτή υπερβαίνει σαφώς το όριο του 0.8, επιβεβαιώνοντας ότι το μοντέλο παρουσιάζει υψηλή προβλεπτική ικανότητα. Συνεπώς, τα αποτελέσματα της καμπύλης ROC και της τιμής AUC ενισχύουν το συμπέρασμα ότι το τελικό μοντέλο αποτελεί ένα αξιόπιστο εργαλείο πρόβλεψης, ικανό να διαχωρίζει αποτελεσματικά τις δύο κατηγορίες με βάση τα χαρακτηριστικά των ομάδων.

Area under the curve	0.8836
----------------------	--------

Πίνακας 5.1.4 Area Under the Curve (AUC)

5.1.5 Δείκτης Nagelkerke R²

5.1.5.1 Θεωρητικό υπόβαθρο

Στο πλαίσιο της λογιστικής παλινδρόμησης, η αξιολόγηση της προσαρμογής του μοντέλου δεν μπορεί να βασιστεί στον κλασικό συντελεστή προσδιορισμού R² της γραμμικής παλινδρόμησης, καθώς η εξαρτημένη μεταβλητή είναι κατηγορική και η εκτίμηση των παραμέτρων πραγματοποιείται μέσω της μεθόδου μέγιστης πιθανοφάνειας. Για τον λόγο αυτό, η ποιότητα προσαρμογής των μοντέλων αξιολογείται με βάση συναρτήσεις πιθανοφάνειας και μέσω εναλλακτικών δεικτών, γνωστών ως pseudo-R² (Hosmer & Lemeshow, 2013) οι οποίοι βασίζονται στη σύγκριση της πιθανοφάνειας του εκτιμημένου μοντέλου με εκείνη ενός βασικού μοντέλου (null model), το οποίο περιλαμβάνει μόνο τη σταθερά. Με τον τρόπο αυτό εκφράζουν τον βαθμό βελτίωσης που επιτυγχάνεται με την εισαγωγή των ερμηνευτικών μεταβλητών και παρέχουν μία ένδειξη της ερμηνευτικής ικανότητας του μοντέλου. Ένας από τους βασικούς δείκτες είναι ο δείκτης Cox και Snell, ο οποίος ορίζεται ως:

$$R_{\{CS\}}^2 = 1 - \left(\frac{L_0}{L_m} \right)^{\frac{2}{n}}$$

όπου L₀ είναι η πιθανοφάνεια του μοντέλου χωρίς ανεξάρτητες μεταβλητές, L_m η πιθανοφάνεια του εκτιμημένου μοντέλου και n το μέγεθος του δείγματος.

Ωστόσο, ο δείκτης αυτός δεν μπορεί να λάβει μέγιστη τιμή ίση με 1, γεγονός που δυσχεραίνει την ερμηνεία του.

Για τον λόγο αυτό, ο Nagelkerke πρότεινε μία κανονικοποιημένη εκδοχή του δείκτη:

$$R_N^2 = \frac{R_{CS}^2}{\left(1 - L_0^{\frac{2}{n}}\right)}$$

η οποία λαμβάνει τιμές στο διάστημα [0,1], διευκολύνοντας την ερμηνεία.

Μεγαλύτερες τιμές του δείκτη υποδηλώνουν καλύτερη προσαρμογή του μοντέλου στα δεδομένα, χωρίς ωστόσο να ερμηνεύονται με τον ίδιο τρόπο όπως ο κλασικός R^2 της γραμμικής παλινδρόμησης. Συγκεκριμένα, οι τιμές του δεν αντιστοιχούν ακριβώς στο ποσοστό της διακύμανσης που εξηγείται, αλλά χρησιμοποιούνται κυρίως για συγκριτικούς σκοπούς μεταξύ εναλλακτικών μοντέλων. Επιπλέον, χρησιμοποιείται και ο δείκτης McFadden R^2 , ο οποίος αποτελεί έναν από τους πιο διαδεδομένους δείκτες pseudo- R^2 στη λογιστική παλινδρόμηση.

Ο δείκτης αυτός βασίζεται στη σύγκριση της λογαριθμικής πιθανοφάνειας του εκτιμημένου μοντέλου με εκείνη του βασικού μοντέλου (null model) και ορίζεται ως:

$$R_{McF}^2 = 1 - \frac{\ln(L_m)}{\ln(L_0)}$$

όπου $\ln(L_m)$ είναι η λογαριθμική πιθανοφάνεια του εκτιμημένου μοντέλου και $\ln(L_0)$ η αντίστοιχη του μοντέλου χωρίς ανεξάρτητες μεταβλητές. Σε αντίθεση με τον κλασικό συντελεστή R^2 , οι τιμές του δείκτη McFadden είναι συνήθως χαμηλότερες, με τιμές μεταξύ 0.2 και 0.4 να θεωρούνται ενδεικτικές καλής προσαρμογής του μοντέλου. Ο δείκτης αυτός χρησιμοποιείται κυρίως για συγκριτικούς σκοπούς μεταξύ εναλλακτικών μοντέλων και όχι για απόλυτη ερμηνεία της ερμηνευτικής ικανότητας. (Hosmer & Lemeshow, 2013).

5.1.5.2 Υπολογισμός του δείκτη Nagelkerke R^2

Η αξιολόγηση της προσαρμογής του τελικού μοντέλου πραγματοποιήθηκε μέσω δεικτών pseudo- R^2 , και συγκεκριμένα των McFadden, Cox and Snell και Nagelkerke, οι οποίοι βασίζονται στη σύγκριση της πιθανοφάνειας του εκτιμημένου μοντέλου με εκείνη του βασικού μοντέλου (μοντέλο μόνο με σταθερά). Τα αποτελέσματα παρουσιάζονται στον Πίνακα 5.1.5.2.1. Ο δείκτης McFadden R^2 ανέρχεται σε 0.371, τιμή που βρίσκεται εντός του εύρους 0.2–0.4, το οποίο θεωρείται ενδεικτικό καλής προσαρμογής για μοντέλα λογιστικής παλινδρόμησης. Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο παρουσιάζει ικανοποιητική έως υψηλή ποιότητα προσαρμογής στα δεδομένα. Αντίστοιχα, ο δείκτης Cox and Snell λαμβάνει τιμή 0.398, ενώ ο δείκτης Nagelkerke R^2 ανέρχεται σε 0.534. Οι τιμές αυτές υποδηλώνουν ότι το μοντέλο εξηγεί ένα σημαντικό ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής, ενισχύοντας την ερμηνευτική του ικανότητα.

Επιπλέον, ο έλεγχος λόγου πιθανοφανειών (Likelihood Ratio Test) είναι στατιστικά σημαντικός ($\chi^2 = 198.44$, $p < 0.001$), γεγονός που δείχνει ότι το μοντέλο με τις επεξηγηματικές μεταβλητές προσαρμόζεται σημαντικά καλύτερα στα δεδομένα σε σχέση με το βασικό μοντέλο (μόνο με το σταθερό όρο). Το αποτέλεσμα αυτό επιβεβαιώνει τη συνολική χρησιμότητα του μοντέλου. Συνολικά, οι δείκτες pseudo- R^2 και τα αποτελέσματα του ελέγχου πιθανοφάνειας καταδεικνύουν ότι το τελικό μοντέλο παρουσιάζει καλή προσαρμογή και ικανοποιητική ερμηνευτική ικανότητα, καθιστώντας το κατάλληλο για την ανάλυση και πρόβλεψη της πρόκρισης στο Final 8.

Δείκτης	Τιμή
McFadden R^2	0.371
Cox and Snell R^2	0.398
Nagelkerke R^2	0.534
Likelihood Ratio (χ^2)	198.44
p-value	< 0.001

Πίνακα 5.1.5.2.1 ο δείκτης Nagelkerke R^2 για το τελικό μοντέλο.

5.1.6 Κατάλοιπα απόκλισης

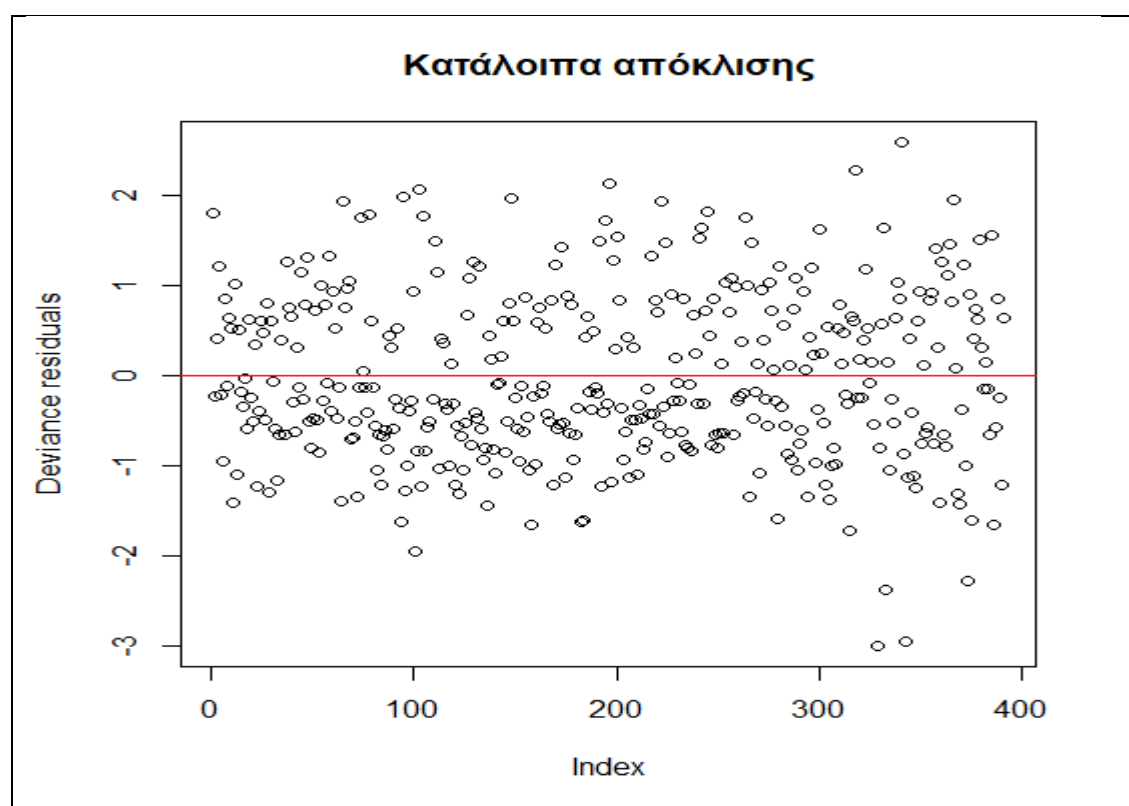
5.1.6.1 Θεωρητικό υπόβαθρο

Στη λογιστική παλινδρόμηση, τα κατάλοιπα δεν ορίζονται με τον ίδιο τρόπο όπως στη γραμμική παλινδρόμηση, λόγω της μη γραμμικής φύσης του μοντέλου και της δυαδικής μορφής της εξαρτημένης μεταβλητής. Για τον λόγο αυτό χρησιμοποιούνται τα κατάλοιπα απόκλισης (deviance residuals), τα οποία βασίζονται στη συνάρτηση πιθανοφάνειας και εκφράζουν τη συνεισφορά κάθε παρατήρησης στη συνολική απόκλιση του μοντέλου. Τα κατάλοιπα απόκλισης αποτελούν μέτρο της διαφοράς μεταξύ των παρατηρούμενων τιμών και των προβλεπόμενων πιθανοτήτων του μοντέλου. Πιο συγκεκριμένα, αποτυπώνουν το κατά πόσο η πρόβλεψη του μοντέλου για μια παρατήρηση αποκλίνει από την πραγματική της κατηγορία, λαμβάνοντας υπόψη τη λογιστική συνάρτηση σύνδεσης (logit link). Με τον τρόπο αυτό, παρέχουν μια πιο κατάλληλη προσέγγιση για την αξιολόγηση της προσαρμογής σε σχέση με τα απλά κατάλοιπα. Η ανάλυση των καταλοίπων απόκλισης πραγματοποιείται συνήθως μέσω γραφημάτων, όπου εξετάζεται η κατανομή τους ως προς τις προβλεπόμενες τιμές ή ως προς τη σειρά των παρατηρήσεων. Σε ένα καλά προσαρμοσμένο μοντέλο, τα κατάλοιπα αναμένεται να κατανέμονται τυχαία γύρω από το μηδέν, χωρίς την ύπαρξη συστηματικών προτύπων, τάσεων ή δομών. Η παρουσία τέτοιων προτύπων μπορεί να υποδηλώνει προβλήματα στο μοντέλο, όπως μη γραμμικότητα, παράλειψη σημαντικών μεταβλητών ή ακατάλληλη μορφή του υποδείγματος. Επιπλέον, τα κατάλοιπα απόκλισης χρησιμοποιούνται για τον εντοπισμό ακραίων ή επιδραστικών παρατηρήσεων. Τιμές με μεγάλο απόλυτο μέγεθος ενδέχεται να υποδηλώνουν παρατηρήσεις που δεν περιγράφονται επαρκώς από το μοντέλο και χρήζουν περαιτέρω διερεύνησης. Ωστόσο, σε δεδομένα πραγματικού κόσμου, η ύπαρξη μεμονωμένων ακραίων τιμών θεωρείται αναμενόμενη και δεν συνεπάγεται απαραίτητα πρόβλημα, εφόσον δεν επηρεάζει τη συνολική συμπεριφορά του μοντέλου.

5.1.6.2 Κατάλοιπα απόκλισης τελικού μοντέλου

Στο σχήμα 5.1.6 παρουσιάζεται το διάγραμμα των καταλοίπων απόκλισης (deviance residuals) ως προς τον δείκτη των παρατηρήσεων. Παρατηρείται ότι τα κατάλοιπα κατανέμονται συμμετρικά γύρω από τη γραμμή του μηδενός, χωρίς να διακρίνεται κάποια εμφανής τάση ή συστηματική δομή. Η απουσία συγκεκριμένων προτύπων, όπως αυξανόμενες ή φθίνουσες τάσεις ή συσσωρεύσεις σημείων σε συγκεκριμένες περιοχές, υποδηλώνει ότι το μοντέλο περιγράφει ικανοποιητικά τη σχέση μεταξύ των μεταβλητών και δεν παρουσιάζει ενδείξεις κακής προσαρμογής.

Η διασπορά των καταλοίπων εμφανίζεται σχετικά ομοιόμορφη σε όλο το εύρος των παρατηρήσεων, γεγονός που δείχνει ότι δεν υπάρχουν σοβαρά προβλήματα μεταβλητότητας που να επηρεάζουν τη σταθερότητα του μοντέλου. Τα περισσότερα κατάλοιπα βρίσκονται εντός του διαστήματος περίπου από -2 έως 2, κάτι που θεωρείται αναμενόμενο για μοντέλα λογιστικής παλινδρόμησης. Παράλληλα, παρατηρούνται λίγες ακραίες τιμές, κυρίως αρνητικές (κάτω από -2.5 έως -3) και σε μικρότερο βαθμό θετικές (πάνω από 2), οι οποίες αντιστοιχούν σε παρατηρήσεις που το μοντέλο δεν προβλέπει με μεγάλη ακρίβεια. Ωστόσο, η παρουσία τους είναι περιορισμένη και δεν υποδηλώνει συστηματική αστοχία του μοντέλου. Συνολικά, το διάγραμμα των καταλοίπων δεν εμφανίζει μοτίβα που να υποδεικνύουν παραβίαση βασικών υποθέσεων ή προβλήματα προσαρμογής, ενισχύοντας την αξιοπιστία και την καταλληλότητα του τελικού μοντέλου για την ανάλυση των δεδομένων.



Σχήμα 5.1.6 Κατάλοιπα απόκλισης τελικού μοντέλου

5.1.7 Έλεγχος Hosmer – Lemeshow

5.1.7.1 Θεωρητικό υπόβαθρο

Η αξιολόγηση της καλής προσαρμογής ενός μοντέλου λογιστικής παλινδρόμησης αποτελεί ένα ιδιαίτερα σημαντικό στάδιο της ανάλυσης, καθώς διασφαλίζει ότι το μοντέλο περιγράφει επαρκώς τη σχέση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών. Σε αντίθεση με τη γραμμική παλινδρόμηση, όπου η αξιολόγηση βασίζεται κυρίως στην ανάλυση της διακύμανσης και στα κατάλοιπα, στη λογιστική παλινδρόμηση η έννοια της απόκλισης (deviance) δεν ακολουθεί γνωστή κατανομή για δίτιμα δεδομένα, γεγονός που καθιστά δύσκολη τη χρήση της για έλεγχο καλής προσαρμογής (Πολίτης, 2023). Για τον λόγο αυτό, έχουν προταθεί εναλλακτικές μέθοδοι αξιολόγησης, με τον πλέον διαδεδομένο να είναι ο έλεγχος Hosmer–Lemeshow (HL). Ο έλεγχος αυτός βασίζεται στη σύγκριση των παρατηρούμενων και των αναμενόμενων συχνοτήτων της δυαδικής απόκρισης, όπως αυτές προκύπτουν από το προσαρμοσμένο μοντέλο. Η βασική ιδέα του ελέγχου είναι η ομαδοποίηση των παρατηρήσεων με βάση τις προβλεπόμενες πιθανότητες επιτυχίας. Συγκεκριμένα, αρχικά υπολογίζονται οι εκτιμημένες πιθανότητες $\hat{p}_i$ για κάθε παρατήρηση και στη συνέχεια οι παρατηρήσεις ταξινομούνται με αύξουσα σειρά ως προς αυτές. Έπειτα, χωρίζονται σε g ομάδες (συνήθως 10 ομάδες ίσου μεγέθους), και για κάθε ομάδα υπολογίζεται ο αριθμός των παρατηρούμενων και των αναμενόμενων επιτυχιών και αποτυχιών.

Με βάση τις παραπάνω ποσότητες, υπολογίζεται μία στατιστική συνάρτηση ελέγχου τύπου Pearson χ^2 , η οποία εκφράζει το μέγεθος της απόκλισης μεταξύ των παρατηρούμενων και των αναμενόμενων τιμών. Υπό τη μηδενική υπόθεση ότι το μοντέλο προσαρμόζεται σωστά στα δεδομένα, η στατιστική αυτή ακολουθεί προσεγγιστικά κατανομή χ^2 με $g - 2$ βαθμούς ελευθερίας.

Η ερμηνεία του ελέγχου είναι η εξής:

- Αν η τιμή p-value είναι μεγάλη (συνήθως μεγαλύτερη από 0.05), τότε δεν απορρίπτεται η μηδενική υπόθεση και συνεπώς δεν υπάρχουν ενδείξεις κακής προσαρμογής του μοντέλου.
- Αντίθετα, μικρή τιμή p-value υποδηλώνει ότι το μοντέλο δεν προσαρμόζεται ικανοποιητικά στα δεδομένα.

Ο έλεγχος Hosmer–Lemeshow χρησιμοποιείται ευρέως στη λογιστική παλινδρόμηση, καθώς λαμβάνει υπόψη την κατανομή των προβλεπόμενων πιθανοτήτων και παρέχει μια πρακτική και εύκολα ερμηνεύσιμη μέθοδο για την αξιολόγηση της συνολικής προσαρμογής του μοντέλου. Ωστόσο, αξίζει να σημειωθεί ότι τα αποτελέσματα του ελέγχου μπορεί να επηρεάζονται από το μέγεθος του δείγματος και τον αριθμό των ομάδων που επιλέγονται, γεγονός που απαιτεί προσεκτική ερμηνεία σε συνδυασμό με άλλους διαγνωστικούς δείκτες.

5.1.7.2 Έλεγχος Hosmer – Lemeshow στο τελικό μοντέλο

Ο έλεγχος Hosmer–Lemeshow εφαρμόστηκε για την αξιολόγηση της καλής προσαρμογής του μοντέλου λογιστικής παλινδρόμησης. Τα αποτελέσματα του ελέγχου παρουσιάζονται στον Πίνακα 5.1.7.1 και έδειξαν ότι η τιμή του στατιστικού ελέγχου είναι $X^2 = 14.093$ με 8 βαθμούς ελευθερίας και $p\text{-value} = 0.0794$. Δεδομένου ότι η τιμή του $p\text{-value}$ είναι μεγαλύτερη από το επίπεδο σημαντικότητας 0.05, δεν απορρίπτεται η μηδενική υπόθεση του ελέγχου, σύμφωνα με την οποία το μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Συνεπώς, δεν προκύπτουν ενδείξεις κακής προσαρμογής και το μοντέλο μπορεί να θεωρηθεί κατάλληλο για την περιγραφή της σχέσης μεταξύ των μεταβλητών. Παρότι το $p\text{-value}$ βρίσκεται σχετικά κοντά στο όριο σημαντικότητας, εξακολουθεί να υποδηλώνει αποδεκτή προσαρμογή, χωρίς να εγείρονται σοβαρές ανησυχίες για την καταλληλότητα του μοντέλου. Το αποτέλεσμα αυτό είναι συνεπές με τα ευρήματα των υπόλοιπων διαγνωστικών ελέγχων, όπως η ανάλυση των καταλοίπων, οι δείκτες pseudo- R^2 και η καμπύλη ROC, τα οποία επίσης υποδεικνύουν καλή προσαρμογή και ικανοποιητική προβλεπτική ικανότητα του μοντέλου.

Hosmer and Lemeshow goodness of fit (GOF) test
--

data: y, fitted(model2)

X-squared = 14.093, df = 8, p-value = 0.07937

Πίνακας 5.1.7.1 Έλεγχος Hosmer – Lemeshow στο τελικό μοντέλο

Συνολικά, τα αποτελέσματα της ανάλυσης καταδεικνύουν ότι το τελικό μοντέλο λογιστικής παλινδρόμησης παρουσιάζει καλή προσαρμογή και υψηλή προβλεπτική ικανότητα. Η ανάλυση των καταλοίπων απόκλισης δεν ανέδειξε την ύπαρξη συστηματικών προτύπων ή αποκλίσεων από τις βασικές υποθέσεις του μοντέλου, γεγονός που υποδηλώνει ότι το μοντέλο περιγράφει ικανοποιητικά τη σχέση μεταξύ των μεταβλητών. Παράλληλα, ο έλεγχος Hosmer–Lemeshow ($p\text{-value} = 0.0794$) επιβεβαιώνει ότι δεν υπάρχουν ενδείξεις κακής προσαρμογής, ενισχύοντας την αξιοπιστία του μοντέλου. Επιπλέον, η τιμή του δείκτη AUC (≈ 0.884) υποδηλώνει πολύ καλή διακριτική ικανότητα, ενώ οι δείκτες pseudo- R^2 καταδεικνύουν ικανοποιητική ερμηνευτική ισχύ. Από τα αποτελέσματα προκύπτει ότι η επιθετική αποτελεσματικότητα (2P%, 3P%), η δημιουργία (APG), η αμυντική δραστηριότητα (STLPG, BLKPG) και η αποφυγή λαθών (TOPG) αποτελούν τους βασικούς παράγοντες που επηρεάζουν την πιθανότητα πρόκρισης. Συνεπώς, το τελικό μοντέλο κρίνεται κατάλληλο για την ερμηνεία και πρόβλεψη της πρόκρισης των ομάδων, καθώς συνδυάζει καλή προσαρμογή, σταθερότητα εκτιμήσεων και υψηλή προβλεπτική ικανότητα.

Στη συνέχεια, η ανάλυση επεκτείνεται στην πολυωνυμική λογιστική παλινδρόμηση, προκειμένου να διερευνηθούν πιο αναλυτικά τα διαφορετικά επίπεδα επιτυχίας των ομάδων.

5.2 Πολυωνυμική Λογιστική Παλινδρόμηση

5.2.1 Θεωρητικό υπόβαθρο

Η πολυωνυμική λογιστική παλινδρόμηση αποτελεί γενίκευση της δυαδικής λογιστικής παλινδρόμησης και εφαρμόζεται όταν η εξαρτημένη μεταβλητή είναι κατηγορική με περισσότερες από δύο δυνατές εκβάσεις (Ηλιόπουλος, 2023). Σε αντίθεση με τη δυαδική περίπτωση, όπου η απόκριση λαμβάνει δύο τιμές, εδώ η μεταβλητή απόκρισης μπορεί να λάβει $J > 2$ κατηγορίες, καθεμία από τις οποίες εμφανίζεται με κάποια πιθανότητα.

Έστω ότι η μεταβλητή απόκρισης Y μπορεί να λάβει J κατηγορίες με πιθανότητες:

$$P(Y = j) = \pi_j, \quad j = 1, 2, \dots, J$$

με τον περιορισμό ότι το άθροισμα των πιθανοτήτων ισούται με τη μονάδα. Στόχος του μοντέλου είναι η έκφραση των πιθανοτήτων αυτών ως συνάρτηση ενός συνόλου επεξηγηματικών μεταβλητών (Ηλιόπουλος, 2023). Δεδομένου ότι οι πιθανότητες δεν είναι ανεξάρτητες μεταξύ τους, χρησιμοποιείται η έννοια των λογαριθμικών λόγων πιθανοτήτων (logits). Για τον σκοπό αυτό επιλέγεται μία κατηγορία ως κατηγορία αναφοράς (baseline category) και οι υπόλοιπες συγκρίνονται με αυτή. Το μοντέλο παίρνει τη μορφή:

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{jp}x_p, \quad j = 2, \dots, J$$

Το παραπάνω σύστημα αποτελείται από $J - 1$ εξισώσεις, μία για κάθε κατηγορία εκτός της κατηγορίας αναφοράς, γεγονός που αυξάνει την πολυπλοκότητα του μοντέλου σε σχέση με τη δυαδική λογιστική παλινδρόμηση (Ηλιόπουλος, 2023). Οι αντίστοιχες πιθανότητες μπορούν να εκφραστούν μέσω εκθετικών συναρτήσεων, διασφαλίζοντας ότι ανήκουν στο διάστημα $[0, 1]$ και αθροίζουν στη μονάδα. Η πολυωνυμική λογιστική παλινδρόμηση βασίζεται στην πολυωνυμική κατανομή, η οποία αποτελεί γενίκευση της διωνυμικής κατανομής για περισσότερες από δύο κατηγορίες (Anderson & Rutkowski, 2007). Στο πλαίσιο αυτό, οι παράμετροι του μοντέλου σχετίζονται με τις πιθανότητες εμφάνισης κάθε κατηγορίας και εκτιμώνται μέσω της μεθόδου της μέγιστης πιθανοφάνειας.

Ένα βασικό χαρακτηριστικό του μοντέλου είναι ότι μπορεί να θεωρηθεί ως ένα σύνολο δυαδικών λογιστικών παλινδρομήσεων που εκτιμώνται ταυτόχρονα, συγκρίνοντας κάθε κατηγορία με την κατηγορία αναφοράς (Anderson & Rutkowski, 2007).

Η ταυτόχρονη εκτίμηση εξασφαλίζει συνέπεια μεταξύ των παραμέτρων και πιο αποδοτική χρήση των δεδομένων. Η ερμηνεία των συντελεστών γίνεται μέσω των λόγων πιθανοτήτων (odds ratios), όπως και στη δυαδική λογιστική παλινδρόμηση, με

τη διαφορά ότι κάθε συντελεστής αναφέρεται στη σύγκριση μιας συγκεκριμένης κατηγορίας με την κατηγορία αναφοράς.

Έτσι, οι συντελεστές δεν περιγράφουν άμεσα πιθανότητες, αλλά τη σχετική μεταβολή των odds μεταξύ κατηγοριών. Επιπλέον, η επιλογή της κατηγορίας αναφοράς δεν επηρεάζει την ποιότητα προσαρμογής του μοντέλου, αλλά επηρεάζει την ερμηνεία των συντελεστών, καθώς όλες οι συγκρίσεις γίνονται σε σχέση με αυτή. Τέλος, αξίζει να σημειωθεί ότι η πολυωνυμική λογιστική παλινδρόμηση μπορεί να εφαρμοστεί τόσο σε ονομαστικές όσο και σε διατακτικές μεταβλητές, αν και στην περίπτωση διατεταγμένων κατηγοριών συχνά προτιμώνται εξειδικευμένα μοντέλα, όπως τα διατακτικά λογιστικά μοντέλα (Ηλιόπουλος, 2023)

5.2.2 Πολυωνυμικό λογιστικό μοντέλο

Στην παρούσα ανάλυση εφαρμόστηκε πολυωνυμικό μοντέλο λογιστικής παλινδρόμησης με σκοπό τη διερεύνηση των παραγόντων που επηρεάζουν το επίπεδο επιτυχίας των ομάδων, το οποίο κατηγοριοποιήθηκε σε τρεις διακριτές κατηγορίες: «No» (μη πρόκριση), «Final8» και «Final4». Ως κατηγορία αναφοράς ορίστηκε η «No», και συνεπώς το μοντέλο εκτιμά ξεχωριστές εξισώσεις για τις συγκρίσεις «Final4 έναντι No» και «Final8 έναντι No». Τα αποτελέσματα παρουσιάζονται στον πίνακα 5.2.2.1. Η επιλογή του τελικού πολυωνυμικού λογιστικού μοντέλου πραγματοποιήθηκε μέσω διαδικασίας οπισθοδρομικής απαλοιφής μεταβλητών (backward elimination), με βάση το κριτήριο Akaike (AIC) και τον έλεγχο λόγου πιθανοφανειών. Η διαδικασία αυτή ξεκινά από το πλήρες μοντέλο, το οποίο περιλαμβάνει όλες τις διαθέσιμες επεξηγηματικές μεταβλητές, και στη συνέχεια αφαιρεί διαδοχικά εκείνες που δεν συμβάλλουν ουσιαστικά στην ερμηνευτική ικανότητα ή την προσαρμογή του μοντέλου. Σε κάθε βήμα, η επιλογή της προς αφαίρεση μεταβλητής βασίστηκε τόσο στη στατιστική σημαντικότητα των συντελεστών (τιμές p), όσο και στη συνολική επίδρασή της στην ποιότητα του μοντέλου, όπως αυτή αποτυπώνεται μέσω του κριτηρίου Akaike (AIC). Η χρήση του AIC επέτρεψε την αντικειμενική σύγκριση μεταξύ διαδοχικών μοντέλων, λαμβάνοντας υπόψη τόσο την καλή προσαρμογή όσο και την πολυπλοκότητα του μοντέλου. Με τον τρόπο αυτό αποφεύχθηκε η υπερπροσαρμογή (overfitting), διασφαλίζοντας ότι το τελικό μοντέλο παραμένει ερμηνεύσιμο.

Παράλληλα, σε κάθε στάδιο πραγματοποιήθηκε έλεγχος λόγου πιθανοφανειών, ώστε να επιβεβαιωθεί ότι η αφαίρεση κάθε μεταβλητής δεν οδηγεί σε στατιστικά σημαντική επιδείνωση της προσαρμογής. Ιδιαίτερη προσοχή δόθηκε σε ζητήματα πολυσυγγραμμικότητας, καθώς μεταβλητές που παρουσίαζαν ισχυρές γραμμικές σχέσεις μεταξύ τους εξετάστηκαν προσεκτικά και, όπου κρίθηκε απαραίτητο, αφαιρέθηκαν ώστε να βελτιωθεί η σταθερότητα των εκτιμήσεων. Επιπλέον, κατά τη διαδικασία επιλογής του τελικού μοντέλου δόθηκε ιδιαίτερη έμφαση στην ερμηνευσιμότητα των αποτελεσμάτων. Συγκεκριμένα, προτιμήθηκαν μοντέλα των οποίων οι συντελεστές μπορούν να ερμηνευτούν με σαφήνεια και συνέπεια ως προς τη θεωρία και τη λογική του αθλήματος. Για τον λόγο αυτό αποφεύχθηκε η ταυτόχρονη

συμπερίληψη μεταβλητών που αποτυπώνουν παρόμοιες διαστάσεις της απόδοσης, καθώς κάτι τέτοιο μπορεί να οδηγήσει σε δυσκολία ερμηνείας και σε φαινόμενα πολυσυγγραμμικότητας.

Η διαδικασία ολοκληρώθηκε όταν δεν παρατηρήθηκε περαιτέρω μείωση του AIC και το μοντέλο παρουσίαζε ταυτόχρονα στατιστική επάρκεια και θεωρητική συνοχή. Η επιλογή των μεταβλητών δεν βασίστηκε αποκλειστικά σε στατιστικά κριτήρια, αλλά και στη δυνατότητα ουσιαστικής ερμηνείας των αποτελεσμάτων. Έτσι, προτιμήθηκε ένα μοντέλο που, πέρα από καλή προσαρμογή στα δεδομένα, επιτρέπει την εξαγωγή σαφών και λογικά συνεπών συμπερασμάτων σχετικά με τους παράγοντες που επηρεάζουν την επιτυχία των ομάδων. Το τελικό μοντέλο περιλαμβάνει τις μεταβλητές PPG, FT%, RPG, APG, TOPG και BLKPG, επιτυγχάνοντας ισορροπία μεταξύ ερμηνευτικής ικανότητας και απλότητας. Οι εκτιμημένοι συντελεστές εκφράζουν μεταβολές στους λογαριθμικούς λόγους πιθανοτήτων (log-odds), ενώ η εκθετική τους μορφή ($\exp(\beta)$) επιτρέπει την ερμηνεία σε όρους λόγων πιθανοτήτων (odds ratios). Τα odds ratios παρουσιάζονται στον πίνακα 5.2.2.2. Δεδομένου ότι η κατηγορία αναφοράς είναι η «No», ένας θετικός συντελεστής υποδηλώνει αύξηση της πιθανότητας συμμετοχής στο Final4 ή στο Final8 σε σχέση με τη μη πρόκριση, ενώ ένας αρνητικός συντελεστής υποδηλώνει μείωση αυτής της πιθανότητας. Η μεταβλητή PPG (πόντοι ανά αγώνα) εμφανίζει θετική και στατιστικά σημαντική επίδραση. Συγκεκριμένα, για τη σύγκριση «Final4 έναντι No», το odds ratio είναι περίπου 1.24, γεγονός που σημαίνει ότι κάθε επιπλέον πόντος αυξάνει κατά 24% τον λόγο σχετικών πιθανοτήτων (odds) συμμετοχής στο Final4 σε σχέση με την κατηγορία αναφοράς (No). Αντίστοιχα, για τη σύγκριση «Final8 έναντι No», το odds ratio είναι περίπου 1.10, υποδηλώνοντας μικρότερη αλλά θετική επίδραση. Η μεταβλητή RPG (ριμπάουντ ανά αγώνα) παρουσιάζει επίσης θετική και στατιστικά σημαντική επίδραση, με odds ratio περίπου 1.33 για το Final4 και 1.20 για το Final8, γεγονός που υποδηλώνει ότι η κυριαρχία στα ριμπάουντ αυξάνει σημαντικά τις πιθανότητες επιτυχίας. Η μεταβλητή APG (assists ανά αγώνα) εμφανίζει θετική και στατιστικά σημαντική επίδραση, με odds ratio περίπου 1.15 για το Final4 και 1.13 για το Final8, αναδεικνύοντας τη σημασία της ομαδικής δημιουργίας παιχνιδιού. Αντίθετα, η μεταβλητή TOPG (λάθη ανά αγώνα) παρουσιάζει αρνητική και ιδιαίτερα στατιστικά σημαντική επίδραση. Το odds ratio είναι περίπου 0.57 για το Final4 και 0.60 για το Final8, γεγονός που σημαίνει ότι η αύξηση των λαθών μειώνει σημαντικά τις πιθανότητες πρόκρισης, επιβεβαιώνοντας τη σημασία του ελέγχου της μπάλας. Η μεταβλητή BLKPG (κοψίματα ανά αγώνα) εμφανίζει θετική επίδραση, με odds ratio περίπου 1.75 για το Final4 και 1.47 για το Final8, γεγονός που υποδηλώνει ότι η αμυντική παρουσία συμβάλλει στην αύξηση των πιθανοτήτων επιτυχίας. Τέλος, η μεταβλητή FT% (ποσοστό ελευθέρων βολών) εμφανίζει αρνητική επίδραση, η οποία είναι στατιστικά σημαντική μόνο στη σύγκριση «Final8 έναντι No», ενώ για το Final4 δεν είναι στατιστικά σημαντική. Το αποτέλεσμα αυτό υποδηλώνει ότι η συγκεκριμένη μεταβλητή δεν αποτελεί ισχυρό και σταθερό διαχωριστικό παράγοντα για την επίτευξη υψηλής επιτυχίας.

Συνολικά, τα αποτελέσματα δείχνουν ότι οι βασικοί παράγοντες που αυξάνουν την πιθανότητα επιτυχίας είναι η επιθετική παραγωγικότητα (PPG), τα ριμπάουντ (RPG),

η δημιουργία παιχνιδιού (APG) και η αμυντική παρουσία (BLKPG), ενώ τα λάθη (TOPG) επηρεάζουν αρνητικά την απόδοση των ομάδων.

Το μοντέλο παρέχει μια συνεκτική, στατιστικά τεκμηριωμένη και ερμηνεύσιμη εικόνα των παραγόντων που διαφοροποιούν τις ομάδες ως προς το επίπεδο επιτυχίας.

```
Call:
mlogit(formula = SUCCESS ~ 1 | PPG + FT. + RPG + APG + TOPG +
      BLKPG, data = data_mlogit, reflevel = "No", method = "nr")

Frequencies of alternatives:choice
      No Final4 Final8
0.57033 0.21483 0.21483

nr method
6 iterations, 0h:0m:0s
g'(-H)^-lg = 9.98E-05
successive function values within tolerance limits

Coefficients :
              Estimate Std. Error z-value      Pr(>|z|)
(Intercept):Final4 -16.047343    4.835700 -3.3185    0.0009050 ***
(Intercept):Final8 -1.508248    3.932757 -0.3835    0.7013424
PPG:Final4         0.216898    0.043064  5.0367 0.0000004737 ***
PPG:Final8         0.093492    0.036761  2.5432    0.0109831 *
FT.:Final4        -0.072139    0.043979 -1.6403    0.1009402
FT.:Final8        -0.098580    0.035332 -2.7901    0.0052687 **
RPG:Final4         0.286205    0.078591  3.6417    0.0002708 ***
RPG:Final8         0.186251    0.066157  2.8153    0.0048735 **
APG:Final4         0.139468    0.067027  2.0808    0.0374552 *
APG:Final8         0.122438    0.059473  2.0587    0.0395230 *
TOPG:Final4        -0.564744    0.124581 -4.5331 0.0000058114 ***
TOPG:Final8        -0.506786    0.103261 -4.9078 0.0000009211 ***
BLKPG:Final4       0.556885    0.222685  2.5008    0.0123923 *
BLKPG:Final8       0.387833    0.199169  1.9473    0.0515047 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -265.67
McFadden R^2:  0.3074
Likelihood ratio test : chisq = 235.83 (p.value = < 0.000000000000000222)
> AIC(model2)
[1] 559.3491
```

Πίνακας 5.2.2.1 Πολυωνυμικό λογιστικό μοντέλο

Μεταβλητή	Final4	Final8
(Intercept)	≈ 0.00	0.22
PPG	1.24	1.10
FT%	0.93	0.91
RPG	1.33	1.20
APG	1.15	1.13
TOPG	0.57	0.60
BLKPG	1.75	1.47

Πίνακας 5.2.2.2 Odds ratios

Παράλληλα, πραγματοποιήθηκε έλεγχος πολυσυγγραμμικότητας μέσω του δείκτη Variance Inflation Factor (VIF), ο οποίος χρησιμοποιείται ευρέως για την ανίχνευση ισχυρών γραμμικών σχέσεων μεταξύ των ανεξάρτητων μεταβλητών ενός μοντέλου παλινδρόμησης. Η ύπαρξη έντονης πολυσυγγραμμικότητας μπορεί να επηρεάσει αρνητικά τη σταθερότητα και την αξιοπιστία των εκτιμώμενων συντελεστών, οδηγώντας σε αυξημένες τυπικές αποκλίσεις, δυσκολία ερμηνείας και ενδεχόμενη αστάθεια των αποτελεσμάτων όπως αναφέραμε στην ενότητα 5.1.3. Για τον λόγο αυτό,

ο έλεγχος πολυσυγγραμμικότητας θεωρείται σημαντικό στάδιο κατά την αξιολόγηση της καταλληλότητας ενός πολυπαραγοντικού μοντέλου.

Στην παρούσα ανάλυση, οι τιμές VIF υπολογίστηκαν για όλες τις ανεξάρτητες μεταβλητές του τελικού πολωνυμικού λογιστικού μοντέλου και βρέθηκαν χαμηλές, καθώς όλες ήταν μικρότερες του 2, με μέγιστη τιμή περίπου 1.87 όπως φαίνεται στον πίνακα 5.2.2.3. Οι τιμές αυτές βρίσκονται σημαντικά κάτω από τα συνήθως αποδεκτά όρια που χρησιμοποιούνται στη βιβλιογραφία ($VIF < 5$ ή $VIF < 10$), γεγονός που υποδηλώνει ότι δεν παρατηρούνται προβλήματα σημαντικής πολυσυγγραμμικότητας μεταξύ των επεξηγηματικών μεταβλητών. Το αποτέλεσμα αυτό επιβεβαιώνει ότι οι μεταβλητές του τελικού μοντέλου συνεισφέρουν στην ερμηνεία της πιθανότητας επιτυχίας των ομάδων. Επιπλέον, η χαμηλή πολυσυγγραμμικότητα ενισχύει την ερμηνευσιμότητα του μοντέλου, καθώς μειώνει το ενδεχόμενο επικάλυψης πληροφορίας μεταξύ μεταβλητών που περιγράφουν παρόμοιες αγωνιστικές διαστάσεις.

Μεταβλητή	Τιμή VIF
PPG	1.87
FT%	1.14
RPG	1.69
APG	1.54
TOPG	1.27
BLKPG	1.46

Πίνακας 5.2.2.3 Έλεγχος πολυσυγγραμμικότητας στο πολωνυμικό μοντέλο

5.2.3 Πίνακας ταξινόμησης

Η αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου πραγματοποιήθηκε μέσω πίνακα ταξινόμησης (classification table ή confusion matrix), ο οποίος χρησιμοποιείται ευρέως στην αξιολόγηση μοντέλων ταξινόμησης (Hosmer et al., 2013). Αν και ο πίνακας αυτός παρουσιάζεται κυρίως στη δυαδική λογιστική παλινδρόμηση, η λογική του επεκτείνεται με φυσικό τρόπο και στην περίπτωση πολωνυμικών μοντέλων, όπου η σύγκριση πραγματοποιείται μεταξύ πολλαπλών κατηγοριών. Στην περίπτωση αυτή, ο πίνακας ταξινόμησης λαμβάνει διαστάσεις $k \times k$, όπου k ο αριθμός των κατηγοριών της εξαρτημένης μεταβλητής, και επιτρέπει την αξιολόγηση της ακρίβειας του μοντέλου σε κάθε κατηγορία ξεχωριστά καθώς και συνολικά. Στην περίπτωση πολωνυμικής λογιστικής παλινδρόμησης, η τελική πρόβλεψη για κάθε παρατήρηση προκύπτει επιλέγοντας την κατηγορία με τη μέγιστη εκτιμημένη πιθανότητα.

Με βάση τα αποτελέσματα του πίνακα 5.2.3.1, το μοντέλο επιτυγχάνει συνολικό ποσοστό σωστής ταξινόμησης ίσο με 72.63%, γεγονός που υποδηλώνει ικανοποιητική έως καλή προβλεπτική ικανότητα.

Εξετάζοντας αναλυτικά τον πίνακα ταξινόμησης, παρατηρείται ότι το μοντέλο παρουσιάζει ιδιαίτερα υψηλή ακρίβεια στην κατηγορία «No», καθώς 207 από τις 223 ομάδες που δεν προκρίθηκαν ταξινομήθηκαν σωστά. Το εύρημα αυτό υποδηλώνει ότι το μοντέλο είναι ιδιαίτερα αποτελεσματικό στον εντοπισμό των ομάδων χαμηλής

απόδοσης. Αντίθετα, η απόδοση του μοντέλου είναι χαμηλότερη στις κατηγορίες «Final8» και «Final4».

Συγκεκριμένα, στην κατηγορία «Final8» ταξινομούνται σωστά 22 από τις 84 ομάδες, ενώ σημαντικός αριθμός αυτών συγχέεται είτε με την κατηγορία «No» είτε με την «Final4». Αντίστοιχα, στην κατηγορία «Final4» ταξινομούνται σωστά 55 από τις 84 ομάδες, ενώ ένα μέρος αυτών κατατάσσεται λανθασμένα κυρίως στην κατηγορία «Final8». Η παρατηρούμενη συμπεριφορά είναι αναμενόμενη, καθώς η διάκριση μεταξύ διαφορετικών επιπέδων επιτυχίας (Final8 και Final4) είναι πιο σύνθετη σε σχέση με τη διάκριση μεταξύ επιτυχίας και μη πρόκρισης. Με άλλα λόγια, το μοντέλο μπορεί ευκολότερα να αναγνωρίσει τις ομάδες που δεν προκρίνονται, ενώ εμφανίζει μεγαλύτερη δυσκολία στην ακριβή κατηγοριοποίηση των ομάδων υψηλής απόδοσης. Συνολικά, τα αποτελέσματα του πίνακα ταξινόμησης επιβεβαιώνουν ότι το πολωνυμικό μοντέλο διαθέτει ικανοποιητική προβλεπτική ικανότητα, ενώ ταυτόχρονα αναδεικνύουν τα όρια της ταξινόμησης όταν πρόκειται για διακριτά επίπεδα επιτυχίας. Το εύρημα αυτό είναι σύμφωνο με τη φύση των δεδομένων και ενισχύει την ανάγκη χρήσης πολυπαραγοντικών μοντέλων για την κατανόηση της απόδοσης των ομάδων.

	No	Final8	Final4
No	207	44	17
Final8	8	22	12
Final4	8	18	55
Mean	0.7263427		

Πίνακας 5.2.3.1 Πίνακας ταξινόμησης

Περαιτέρω, εξετάστηκαν οι εκτιμημένες πιθανότητες που προκύπτουν από το μοντέλο για κάθε κατηγορία. Τα αποτελέσματα παρουσιάζονται στον πίνακα 5.2.3.2. Σημειώνεται ότι στον πίνακα 5.2.3.2 παρουσιάζονται ενδεικτικά οι πρώτες έξι παρατηρήσεις του συνόλου δεδομένων, όπως προκύπτουν από την εντολή preview (head) στο περιβάλλον R. Στην πραγματικότητα, το μοντέλο υπολογίζει εκτιμημένες πιθανότητες για το σύνολο των 391 ομάδων του δείγματος. Η παρουσίαση περιορισμένου αριθμού παρατηρήσεων αποσκοπεί αποκλειστικά στην απεικόνιση της μορφής των αποτελεσμάτων και στη διευκόλυνση της ερμηνείας. Για κάθε παρατήρηση, το μοντέλο υπολογίζει την πιθανότητα συμμετοχής στις κατηγορίες «No», «Final4» και «Final8», και η τελική πρόβλεψη προκύπτει από την κατηγορία με τη μέγιστη πιθανότητα. Ενδεικτικά, παρατηρείται ότι σε ορισμένες περιπτώσεις το μοντέλο αποδίδει πολύ υψηλές πιθανότητες σε μία κατηγορία, όπως για παράδειγμα πιθανότητες άνω του 0.95 για την κατηγορία «No», γεγονός που υποδηλώνει υψηλό βαθμό βεβαιότητας στην πρόβλεψη. Αντίθετα, σε άλλες περιπτώσεις οι πιθανότητες κατανέμονται πιο ισομερώς μεταξύ των κατηγοριών, γεγονός που υποδηλώνει αυξημένη αβεβαιότητα και δυσκολία διάκρισης μεταξύ των επιπέδων επιτυχίας. Η παρατήρηση αυτή είναι συνεπής με τα αποτελέσματα του πίνακα ταξινόμησης, όπου το μοντέλο εμφανίζει υψηλότερη ακρίβεια στην κατηγορία «No» και χαμηλότερη στις κατηγορίες «Final8» και «Final4». Συνεπώς, οι εκτιμημένες πιθανότητες παρέχουν

πρόσθετη πληροφορία σχετικά με τον βαθμό βεβαιότητας των προβλέψεων, επιτρέποντας την καλύτερη κατανόηση της συμπεριφοράς του μοντέλου.

ID	No	Final4	Final8
1	0.5331897	0.18751742	0.27929284
2	0.9261816	0.00843521	0.06538320
3	0.3404618	0.23452952	0.42500868
4	0.7984952	0.04837137	0.15313344
5	0.9710345	0.00396050	0.02500495
6	0.8054350	0.03813244	0.15643252

Πίνακας 5.2.3.2 Εκτιμημένες πιθανότητες

5.2.4 Καμπύλη roc στο πολωνυμικό μοντέλο

Όπως αναφέρθηκε στην ενότητα 5.1.4, η καμπύλη ROC αποτελεί βασικό εργαλείο αξιολόγησης της διακριτικής ικανότητας ενός μοντέλου σε δυαδικό πλαίσιο. Ωστόσο, στην περίπτωση ταξινόμησης σε περισσότερες από δύο κατηγορίες, η εφαρμογή της δεν είναι άμεση και απαιτεί κατάλληλες προσαρμογές. Συγκεκριμένα, η βιβλιογραφία προτείνει τη σύγκριση κάθε κατηγορίας με όλες τις υπόλοιπες (προσέγγιση one-vs-rest), καθώς και τη χρήση γενικευμένων μέτρων AUC που βασίζονται σε ζεύγη κατηγοριών (Hand & Till, 2001).

Η λογική της προσέγγισης αυτής βασίζεται στο γεγονός ότι η ROC ανάλυση αξιολογεί κυρίως την ικανότητα του μοντέλου να διατάσσει (rank) σωστά τις παρατηρήσεις, αποδίδοντας υψηλότερες τιμές πιθανότητας στις σωστές κατηγορίες. Όπως επισημαίνεται από τον Fawcett (2006), η ROC καμπύλη δεν εξαρτάται από συγκεκριμένο σημείο διάκρισης, αλλά αποτυπώνει τη συνολική συμπεριφορά του μοντέλου σε όλο το εύρος των πιθανών αποφάσεων, αξιολογώντας ουσιαστικά την ποιότητα της κατάταξης των παρατηρήσεων. Η ιδιότητα αυτή καθιστά τη μέθοδο ιδιαίτερα κατάλληλη για περιπτώσεις όπου οι εκτιμήσεις του μοντέλου εκφράζονται ως πιθανότητες ή δείκτες και όχι μόνο ως τελικές κατηγορίες.

Στο πλαίσιο αυτό, η ROC ανάλυση επιτρέπει τη μελέτη της διακριτικής ικανότητας του μοντέλου για κάθε κατηγορία ξεχωριστά, αναδεικνύοντας πιθανές διαφοροποιήσεις στην απόδοση μεταξύ των επιπέδων επιτυχίας. Παράλληλα, η γενίκευση που προτείνουν οι Hand και Till (2001) βασίζεται στον υπολογισμό της μέσης διακριτικής ικανότητας μεταξύ όλων των δυνατών ζευγών κατηγοριών, αποφεύγοντας την ανάγκη καθορισμού κόστους σφαλμάτων και επιτρέποντας μια πιο συνολική αξιολόγηση της απόδοσης.

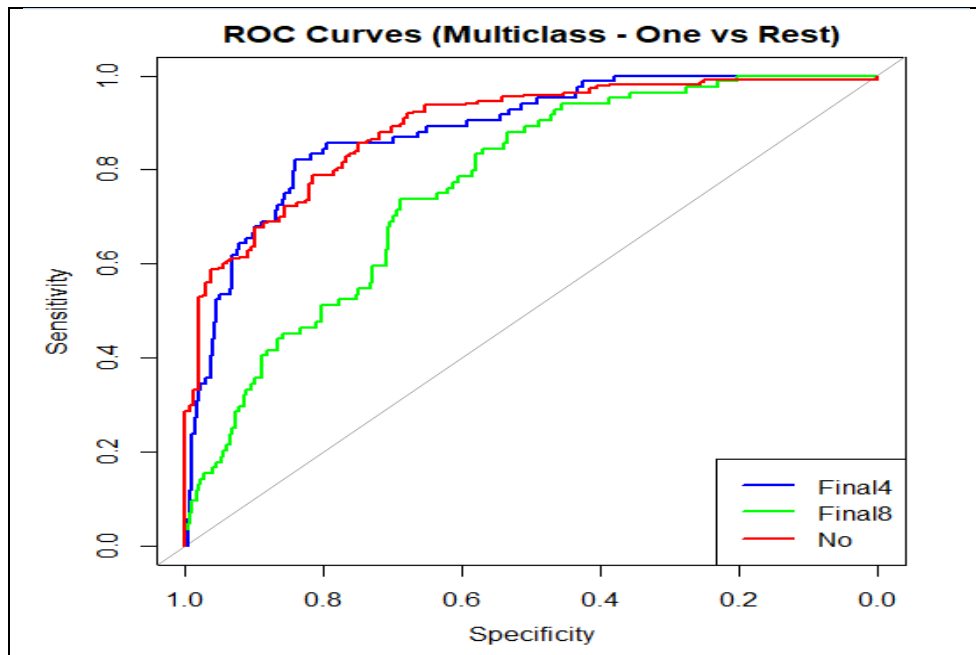
Το Σχήμα 5.2.4 παρουσιάζει τις καμπύλες ROC του πολωνυμικού λογιστικού μοντέλου για τις τρεις κατηγορίες («No», «Final8», «Final4»), παρέχοντας μια συμπληρωματική εικόνα της διακριτικής του ικανότητας σε επίπεδο κατηγορίας. Η ανάλυση αυτή ενισχύει τα ευρήματα που προέκυψαν από τον πίνακα ταξινόμησης και την εξέταση των εκτιμημένων πιθανοτήτων, επιτρέποντας μια πιο ολοκληρωμένη κατανόηση της συμπεριφοράς του μοντέλου.

Όπως παρατηρείται, η καμπύλη της κατηγορίας «No» (κόκκινη γραμμή) εμφανίζει υψηλή διακριτική ικανότητα, γεγονός που επιβεβαιώνει την πολύ καλή απόδοση του μοντέλου στην αναγνώριση των ομάδων που δεν προκρίνονται.

Το εύρημα αυτό είναι πλήρως συνεπές με τον πίνακα ταξινόμησης, όπου καταγράφηκε ιδιαίτερα υψηλό ποσοστό σωστής ταξινόμησης για τη συγκεκριμένη κατηγορία. Επιπλέον, οι εκτιμημένες πιθανότητες δείχνουν ότι σε πολλές περιπτώσεις το μοντέλο αποδίδει πολύ υψηλές τιμές στην κατηγορία «No», γεγονός που υποδηλώνει υψηλό βαθμό βεβαιότητας στις προβλέψεις αυτές. Αντίστοιχα, η κατηγορία «Final4» (μπλε γραμμή) παρουσιάζει επίσης ισχυρή διακριτική ικανότητα, με την καμπύλη της να βρίσκεται κοντά στα ανώτερα επίπεδα του διαγράμματος. Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο μπορεί να διαχωρίσει αποτελεσματικά τις ομάδες υψηλής επιτυχίας, αποδίδοντας υψηλότερες πιθανότητες σε αυτές σε σχέση με τις υπόλοιπες κατηγορίες. Η συμπεριφορά αυτή υποδηλώνει ότι οι ομάδες που φτάνουν στο Final4 χαρακτηρίζονται από πιο έντονα και διακριτά αγωνιστικά πρότυπα. Αντίθετα, η κατηγορία «Final8» (πράσινη γραμμή) εμφανίζει χαμηλότερη διακριτική ικανότητα, καθώς η καμπύλη της βρίσκεται πιο κοντά στη διαγώνιο.

Το εύρημα αυτό συνδέεται άμεσα με τα αποτελέσματα του πίνακα ταξινόμησης, όπου παρατηρήθηκε αυξημένη σύγχυση της κατηγορίας αυτής τόσο με την «No» όσο και με την «Final4». Παράλληλα, οι εκτιμημένες πιθανότητες δείχνουν ότι για αρκετές παρατηρήσεις οι τιμές κατανέμονται πιο ισομερώς μεταξύ των κατηγοριών, γεγονός που αντανakλά μεγαλύτερη αβεβαιότητα και δυσκολία διάκρισης για το συγκεκριμένο επίπεδο επιτυχίας.

Συνολικά, το σχήμα επιβεβαιώνει ότι το μοντέλο παρουσιάζει υψηλότερη διακριτική ικανότητα στις κατηγορίες που αντιστοιχούν σε πιο σαφώς διαφοροποιημένα επίπεδα απόδοσης («No» και «Final4»), ενώ εμφανίζει περιορισμένη απόδοση στην ενδιάμεση κατηγορία («Final8»). Η παρατήρηση αυτή είναι συνεπής με τη γενικότερη συμπεριφορά του μοντέλου και αναδεικνύει ότι η δυσκολία ταξινόμησης αυξάνεται όταν οι κατηγορίες παρουσιάζουν μεγαλύτερη επικάλυψη ως προς τα χαρακτηριστικά τους.



Σχήμα 5.2.4.1 Καμπύλη roc στο πολυωνυμικό μοντέλο

Οι τιμές της περιοχής κάτω από την καμπύλη (AUC) παρουσιάζονται στον πίνακα 5.2.4.1 και επιβεβαιώνουν τη διαφοροποιημένη διακριτική ικανότητα του μοντέλου μεταξύ των κατηγοριών. Συγκεκριμένα, για την κατηγορία «Final4» η τιμή AUC ανέρχεται σε 0.8833, γεγονός που υποδηλώνει υψηλή διακριτική ικανότητα.

Η τιμή αυτή σημαίνει ότι το μοντέλο έχει πολύ μεγάλη πιθανότητα να αποδώσει υψηλότερη εκτιμημένη πιθανότητα σε μια ομάδα που ανήκει στο Final4 σε σχέση με μια ομάδα που δεν ανήκει σε αυτή την κατηγορία. Το αποτέλεσμα αυτό είναι συνεπές με τη μορφή της αντίστοιχης ROC καμπύλης, η οποία βρίσκεται κοντά στην άνω αριστερή περιοχή του διαγράμματος. Αντίστοιχα, για την κατηγορία «No» η τιμή AUC είναι 0.8892, η οποία είναι η υψηλότερη μεταξύ των τριών κατηγοριών. Το εύρημα αυτό επιβεβαιώνει ότι το μοντέλο διακρίνει με ιδιαίτερη ακρίβεια τις ομάδες που δεν προκρίνονται, γεγονός που συνάδει τόσο με τα αποτελέσματα του πίνακα ταξινόμησης όσο και με τις υψηλές εκτιμημένες πιθανότητες που αποδίδονται σε αυτή την κατηγορία. Η υψηλή τιμή AUC υποδηλώνει ότι η κατηγορία αυτή χαρακτηρίζεται από πιο σαφή και διακριτά πρότυπα.

Αντίθετα, η κατηγορία «Final8» εμφανίζει χαμηλότερη τιμή AUC (0.7646), γεγονός που υποδηλώνει μέτρια διακριτική ικανότητα σε σύγκριση με τις άλλες δύο κατηγορίες. Η διαφορά αυτή αντανάκλα τη μεγαλύτερη δυσκολία του μοντέλου να διαχωρίσει τις ομάδες της συγκεκριμένης κατηγορίας από τις υπόλοιπες. Το εύρημα αυτό είναι συνεπές τόσο με τη μορφή της ROC καμπύλης όσο και με τον πίνακα ταξινόμησης, όπου παρατηρήθηκε αυξημένη σύγχυση της κατηγορίας «Final8» με τις άλλες κατηγορίες.

Συμπερασματικά, οι τιμές AUC επιβεβαιώνουν ότι το μοντέλο παρουσιάζει υψηλή διακριτική ικανότητα στις ακραίες κατηγορίες («No» και «Final4») και χαμηλότερη στην ενδιάμεση κατηγορία («Final8»).

Η διαφοροποίηση αυτή υποδηλώνει ότι η απόδοση του μοντέλου εξαρτάται σε σημαντικό βαθμό από τον βαθμό διαχωρισιμότητας των κατηγοριών, με τις ενδιάμεσες καταστάσεις να αποτελούν πιο σύνθετες περιπτώσεις ταξινόμησης.

Κατηγορία	AUC
Final4	0.8833
Final8	0.7646
No	0.8892

Πίνακας 5.2.4.1 AUC

5.3 Συμπεράσματα για την ανάλυση παλινδρόμησης

Η σύγκριση των τελικών μοντέλων δυαδικής και πολυωνυμικής λογιστικής παλινδρόμησης αναδεικνύει ουσιαστικές διαφορές τόσο ως προς την ερμηνεία των αποτελεσμάτων όσο και ως προς το είδος της πληροφορίας που παρέχουν για την αγωνιστική απόδοση των ομάδων.

Αρχικά, το τελικό δυαδικό μοντέλο επικεντρώνεται αποκλειστικά στη διάκριση μεταξύ πρόκρισης και μη πρόκρισης στο Final 8. Η προσέγγιση αυτή απλοποιεί σημαντικά το πρόβλημα και οδηγεί σε υψηλή προβλεπτική ικανότητα, όπως αποτυπώνεται από την τιμή του δείκτη AUC (0.8836), η οποία υποδηλώνει πολύ καλή διακριτική ικανότητα. Παράλληλα, η ερμηνεία των συντελεστών είναι άμεση, καθώς κάθε μεταβλητή επηρεάζει μία μόνο σύγκριση (πρόκριση έναντι μη πρόκρισης).

Στο τελικό δυαδικό μοντέλο, βασικοί θετικοί παράγοντες πρόκρισης αναδεικνύονται η αποτελεσματικότητα στην επίθεση (2P%, 3P%), η δημιουργία παιχνιδιού (APG) και η αμυντική δραστηριότητα (STLPG, BLKPG), ενώ τα λάθη (TOPG) επηρεάζουν αρνητικά την πιθανότητα επιτυχίας. Το ποσοστό ελευθέρων βολών (FT%) εμφανίζει οριακά αρνητική επίδραση, η οποία δεν ερμηνεύεται ως πραγματική αρνητική συμβολή, αλλά πιθανότατα αντανακλά συσχετίσεις με άλλες μεταβλητές του μοντέλου. Αντίθετα, το πολυωνυμικό μοντέλο προσφέρει μια πιο λεπτομερή και πολυδιάστατη προσέγγιση, καθώς επιτρέπει την ταυτόχρονη ανάλυση τριών επιπέδων επιτυχίας («No», «Final8», «Final4»). Μέσω αυτής της προσέγγισης καθίσταται δυνατή η διάκριση μεταξύ παραγόντων που επηρεάζουν την απλή πρόκριση και εκείνων που διαφοροποιούν την κορυφαία απόδοση (Final4), προσφέροντας βαθύτερη κατανόηση της αγωνιστικής συμπεριφοράς. Τα αποτελέσματα του πολυωνυμικού μοντέλου δείχνουν ότι μεταβλητές όπως οι πόντοι ανά αγώνα (PPG), τα ριμπάουντ (RPG), οι ασίστ (APG) και τα κονσίματα (BLKPG) συμβάλλουν θετικά στην πιθανότητα υψηλότερης επιτυχίας, ενώ τα λάθη (TOPG) επηρεάζουν αρνητικά όλες τις συγκρίσεις.

Το ποσοστό ελευθέρων βολών (FT%) εμφανίζει ασθενή και μη σταθερή επίδραση, γεγονός που υποδηλώνει ότι δεν αποτελεί καθοριστικό παράγοντα διαφοροποίησης μεταξύ των επιπέδων επιτυχίας. Συγκρίνοντας τα δύο μοντέλα ως προς την προβλεπτική τους ικανότητα, το δυαδικό μοντέλο εμφανίζει υψηλή απόδοση, κάτι που είναι αναμενόμενο λόγω της απλούστερης φύσης του προβλήματος.

Ωστόσο, το πολυωνυμικό μοντέλο επιτυγχάνει επίσης ικανοποιητική ακρίβεια ταξινόμησης (περίπου 72.6%), παρά την αυξημένη πολυπλοκότητα, γεγονός που υποδηλώνει ισχυρή ερμηνευτική ικανότητα. Επιπλέον, η αξιολόγηση της διακριτικής ικανότητας του μοντέλου μέσω ROC ανάλυσης και των αντίστοιχων τιμών AUC επιβεβαιώνει τα παραπάνω ευρήματα. Συγκεκριμένα, οι τιμές AUC διαφοροποιούνται μεταξύ των κατηγοριών, με υψηλές τιμές για τις κατηγορίες «No» ($AUC \approx 0.889$) και «Final4» ($AUC \approx 0.883$), ενώ χαμηλότερη τιμή παρατηρείται για την κατηγορία «Final8» ($AUC \approx 0.765$). Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο παρουσιάζει ισχυρότερη διακριτική ικανότητα στα ακραία επίπεδα απόδοσης, ενώ εμφανίζει μεγαλύτερη δυσκολία στη διάκριση της ενδιάμεσης κατηγορίας. Η παρατήρηση αυτή είναι συνεπής με τα αποτελέσματα του πίνακα ταξινόμησης, ενισχύοντας τη συνολική αξιοπιστία των συμπερασμάτων.

Από πλευράς πολυσυγγραμμικότητας, η διαδικασία επιλογής μεταβλητών οδήγησε και στα δύο μοντέλα σε σημαντική μείωση του φαινομένου, με το τελικό δυαδικό μοντέλο να παρουσιάζει ουσιαστικά απουσία πολυσυγγραμμικότητας, γεγονός που ενισχύει τη σταθερότητα και την αξιοπιστία των εκτιμήσεων. Ένα σημαντικό εύρημα αφορά τη διαφοροποίηση της ερμηνείας των μεταβλητών μεταξύ των δύο προσεγγίσεων. Στη δυαδική λογιστική παλινδρόμηση, οι μεταβλητές ερμηνεύονται ως παράγοντες πρόκρισης, ενώ στην πολυωνυμική παλινδρόμηση διαχωρίζονται σε παράγοντες που επηρεάζουν:

- (α) την αποφυγή χαμηλής απόδοσης (No),
- (β) τη μέση επιτυχία (Final8),
- (γ) την κορυφαία επίδοση (Final4).

Η διάκριση αυτή αποτελεί σημαντικό πλεονέκτημα της πολυωνυμικής προσέγγισης, καθώς επιτρέπει πιο λεπτομερή κατανόηση των διαφορών μεταξύ των επιπέδων επιτυχίας. Συνολικά, η εφαρμογή των δύο μεθοδολογιών παρέχει συμπληρωματική πληροφόρηση για την κατανόηση της αγωνιστικής απόδοσης. Το δυαδικό μοντέλο προσφέρει απλή και αξιόπιστη πρόβλεψη της πρόκρισης, ενώ το πολυωνυμικό μοντέλο επιτρέπει βαθύτερη ανάλυση των διαφορών μεταξύ επιπέδων επιτυχίας. Πιο συγκεκριμένα, στο τελικό δυαδικό μοντέλο διατηρήθηκαν οι μεταβλητές 2P%, 3P%, FT%, APG, STLPG, TOPG και BLKPG, ενώ στο τελικό πολυωνυμικό μοντέλο οι μεταβλητές PPG, FT%, RPG, APG, TOPG και BLKPG. Η παρουσία κοινών μεταβλητών, όπως οι APG, TOPG και BLKPG, επιβεβαιώνει τη σημασία τους ως βασικών δεικτών αγωνιστικής απόδοσης, ενώ οι διαφοροποιήσεις μεταξύ των δύο μοντέλων αντανακλούν τις διαφορετικές διαστάσεις που αναδεικνύει κάθε προσέγγιση.

Κεφάλαιο 6 : Ανάλυση κατά Συστάδες

Στο παρόν κεφάλαιο εφαρμόζονται μέθοδοι ανάλυσης κατά συστάδες με στόχο τη διερεύνηση της εσωτερικής δομής των δεδομένων και τον εντοπισμό προτύπων που σχετίζονται με την αγωνιστική απόδοση και την πρόκριση των ομάδων στο Final 8 της διοργάνωσης. Η ανάλυση κατά συστάδες αποτελεί σημαντικό εργαλείο της μη επιβλεπόμενης μάθησης, καθώς επιτρέπει την ομαδοποίηση των παρατηρήσεων με βάση τα κοινά χαρακτηριστικά τους, χωρίς τη χρήση εκ των προτέρων πληροφορίας σχετικά με την έκβαση. Αρχικά εφαρμόζεται η ιεραρχική μέθοδος του κοντινότερου γείτονα (nearest neighbor), με σκοπό τη διερεύνηση της δομής των δεδομένων και τον αρχικό εντοπισμό πιθανών συστάδων μεταξύ των ομάδων. Ωστόσο, λόγω των περιορισμών της συγκεκριμένης μεθόδου και της εμφάνισης του φαινομένου της αλυσιδωτής σύνδεσης (chaining effect), στη συνέχεια εφαρμόζεται η μέθοδος Ward, η οποία βασίζεται στην ελαχιστοποίηση της ενδο-ομαδικής διακύμανσης και οδηγεί σε περισσότερο συμπαγείς και ισορροπημένες συστάδες. Στη συνέχεια εφαρμόζεται η μη ιεραρχική μέθοδος k-means, η οποία στοχεύει στον διαχωρισμό των ομάδων σε προκαθορισμένο αριθμό συστάδων μέσω της ελαχιστοποίησης της ενδο-ομαδικής διακύμανσης. Η μέθοδος αυτή επιτρέπει την ανάδειξη διακριτών προτύπων αγωνιστικής απόδοσης και τη διερεύνηση της σχέσης τους με την πρόκριση των ομάδων στο Final 8. Τέλος, πραγματοποιείται συγκριτική αξιολόγηση των αποτελεσμάτων των μεθόδων Ward και k-means, προκειμένου να διερευνηθεί η αποτελεσματικότητά τους στην ομαδοποίηση των ομάδων και να εντοπιστεί η καταλληλότερη προσέγγιση για την ανάλυση των δεδομένων. Μέσα από τη σύγκριση αυτή επιδιώκεται η εξαγωγή ουσιαστικών συμπερασμάτων σχετικά με τους βασικούς παράγοντες που σχετίζονται με την αγωνιστική επιτυχία των ομάδων.

6.1 Θεωρητικό υπόβαθρο

Η ανάλυση δεδομένων διακρίνεται σε δύο βασικές κατηγορίες μεθόδων: τις επιβλεπόμενες και τις μη επιβλεπόμενες. Στις επιβλεπόμενες μεθόδους, όπως η λογιστική παλινδρόμηση, υπάρχει μία εξαρτημένη μεταβλητή, της οποίας οι τιμές είναι γνωστές και χρησιμοποιούνται για την εκπαίδευση του μοντέλου με στόχο την πρόβλεψη. Αντίθετα, στις μη επιβλεπόμενες μεθόδους δεν υπάρχει μεταβλητή-στόχος και η ανάλυση βασίζεται αποκλειστικά στα χαρακτηριστικά των δεδομένων, με σκοπό την αναγνώριση προτύπων και δομών (Κούτρας, 2025; James et al., 2021). Μία από τις σημαντικότερες μη επιβλεπόμενες τεχνικές είναι η ανάλυση κατά συστάδες (cluster analysis), η οποία στοχεύει στην ομαδοποίηση των παρατηρήσεων σε ομάδες με υψηλή εσωτερική ομοιογένεια και ταυτόχρονα σημαντική διαφοροποίηση μεταξύ των ομάδων. Με άλλα λόγια, οι παρατηρήσεις που ανήκουν στην ίδια συστάδα παρουσιάζουν παρόμοια χαρακτηριστικά, ενώ εκείνες που ανήκουν σε διαφορετικές συστάδες διαφέρουν ουσιαστικά μεταξύ τους (Κούτρας, 2025).

Η έννοια της ομοιότητας μεταξύ των παρατηρήσεων εκφράζεται μέσω μέτρων απόστασης. Η πιο διαδεδομένη μορφή είναι η ευκλείδεια απόσταση, η οποία για δύο παρατηρήσεις x_i και x_j ορίζεται ως:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Ωστόσο, όταν οι μεταβλητές έχουν διαφορετικές μονάδες μέτρησης ή διακυμάνσεις, είναι απαραίτητη η κανονικοποίηση των δεδομένων, ώστε να διασφαλιστεί ότι κάθε μεταβλητή συνεισφέρει ισότιμα στον υπολογισμό της απόστασης. Η κανονικοποίηση πραγματοποιείται συνήθως μέσω της μετατροπής:

$$z = \frac{x - \mu}{\sigma}$$

όπου μ είναι ο μέσος όρος και σ η τυπική απόκλιση της μεταβλητής (Κούτρας, 2025; James et al., 2021). Στην παρούσα εργασία, πριν από την εφαρμογή των μεθόδων ανάλυσης κατά συστάδες, πραγματοποιήθηκε κανονικοποίηση των δεδομένων μέσω z-score standardization, ώστε όλες οι μεταβλητές να συμμετέχουν ισότιμα στον υπολογισμό των αποστάσεων.

Πέρα από την ευκλείδεια απόσταση, χρησιμοποιούνται και άλλα μέτρα, όπως η Manhattan απόσταση:

$$d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

καθώς και η Minkowski απόσταση, η οποία αποτελεί γενίκευση των παραπάνω. Η επιλογή του κατάλληλου μέτρου εξαρτάται από τη φύση των δεδομένων και μπορεί να επηρεάσει σημαντικά το αποτέλεσμα της ομαδοποίησης (Κούτρας, 2025).

Οι μέθοδοι ανάλυσης κατά συστάδες διακρίνονται σε ιεραρχικές και μη ιεραρχικές. Η ιεραρχική ανάλυση βασίζεται στη δημιουργία μιας δομής τύπου δέντρου, όπου οι παρατηρήσεις συγχωνεύονται σταδιακά σε ομάδες. Στη συσσωρευτική προσέγγιση, κάθε παρατήρηση αρχικά αποτελεί ξεχωριστή ομάδα και σε κάθε βήμα συγχωνεύονται οι δύο πιο κοντινές ομάδες μέχρι να σχηματιστεί μία ενιαία ομάδα. Η διαδικασία αυτή απεικονίζεται μέσω του δενδροδιαγράμματος, το οποίο επιτρέπει την οπτική διερεύνηση της δομής των δεδομένων (Κούτρας, 2025). Μία από τις πιο απλές ιεραρχικές μεθόδους είναι η μέθοδος του κοντινότερου γείτονα (nearest neighbor). Σύμφωνα με αυτή, η απόσταση μεταξύ δύο συστάδων A και B ορίζεται ως:

$$d(A, B) = \min_{x \in A, y \in B} d(x, y)$$

Παρά την απλότητά της, η μέθοδος αυτή παρουσιάζει το μειονέκτημα του “chain effect”, κατά το οποίο δημιουργούνται αλυσιδωτές ενώσεις παρατηρήσεων, οδηγώντας σε επιμήκη και μη συμπαγή συστάδες.

Το χαρακτηριστικό αυτό περιορίζει την αποτελεσματικότητά της, ιδιαίτερα σε δεδομένα με θόρυβο (Κούτρας, 2025). Εκτός από τη μέθοδο του κοντινότερου γείτονα, υπάρχουν και άλλες ιεραρχικές προσεγγίσεις, όπως η μέθοδος του μακρινού γείτονα (complete linkage), η μέθοδος της μέσης σύνδεσης (average linkage) και η μέθοδος Ward.

Η μέθοδος Ward διαφοροποιείται από τις υπόλοιπες ιεραρχικές τεχνικές ομαδοποίησης, καθώς δεν βασίζεται άμεσα στην απόσταση μεταξύ μεμονωμένων παρατηρήσεων ή ομάδων, αλλά στην ελαχιστοποίηση της διακύμανσης στο εσωτερικό των συστάδων. Πιο συγκεκριμένα, σε κάθε στάδιο της διαδικασίας συγχώνευσης επιλέγονται οι δύο συστάδες των οποίων η ένωση προκαλεί τη μικρότερη δυνατή αύξηση της συνολικής ενδο-ομαδικής διακύμανσης. Για τον σκοπό αυτό χρησιμοποιείται το άθροισμα των τετραγωνικών αποκλίσεων (Error Sum of Squares – ESS), το οποίο για μία συστάδα C ορίζεται ως:

$$ESS(C) = \sum_{i \in C} d^2(x_i, \bar{x}_C)$$

όπου x_i είναι η i -οστή παρατήρηση της συστάδας, $\bar{x}_C$ το κέντρο (centroid) της συστάδας και $d(x_i, \bar{x}_C)$ η ευκλείδεια απόσταση της παρατήρησης από το κέντρο της ομάδας. Η μέθοδος Ward επιδιώκει σε κάθε βήμα να διατηρεί όσο το δυνατόν μικρότερη την αύξηση του συνολικού ESS, με αποτέλεσμα τη δημιουργία περισσότερο συμπαγών, ομοιογενών και ισορροπημένων συστάδων (Κούτρας, 2025). Σύμφωνα με τον Κούτρα (2025), η μέθοδος Ward θεωρείται μία από τις αποτελεσματικότερες ιεραρχικές τεχνικές ομαδοποίησης, ιδιαίτερα σε ποσοτικά δεδομένα, καθώς παρουσιάζει καλύτερη συνολική συμπεριφορά σε σχέση με τη μέθοδο του κοντινότερου γείτονα. Αντίθετα με τη μέθοδο του κοντινότερου γείτονα (single linkage), η οποία συχνά εμφανίζει το φαινόμενο της αλυσίδας (chaining effect), η μέθοδος Ward τείνει να αποφεύγει τη δημιουργία επιμηκυμένων και ανομοιογενών συστάδων, παράγοντας ομάδες με μεγαλύτερη εσωτερική συνοχή και καλύτερη ερμηνευσιμότητα. Αντίστοιχα, οι James et al. (2021) αναφέρουν ότι οι μέθοδοι του μακρινού γείτονα (complete linkage) και Ward οδηγούν συνήθως σε πιο ισορροπημένες και σταθερές συστάδες σε σύγκριση με τη μέθοδο του κοντινότερου γείτονα, η οποία συχνά δημιουργεί εκτεταμένες συστάδες στις οποίες οι παρατηρήσεις προστίθενται διαδοχικά μία προς μία. Επιπλέον, τονίζεται ότι η επιλογή της μεθόδου σύνδεσης επηρεάζει σημαντικά τη μορφή του δενδρογράμματος και συνεπώς την τελική ομαδοποίηση των δεδομένων. Η εφαρμογή της μεθόδου Ward προϋποθέτει συνήθως τη χρήση ευκλείδειας απόστασης και κανονικοποιημένων δεδομένων, ώστε όλες οι μεταβλητές να συμβάλλουν ισότιμα στον υπολογισμό των αποστάσεων και της ενδο-ομαδικής διακύμανσης. Για τον λόγο αυτό, πριν από την εφαρμογή της ανάλυσης κατά συστάδες πραγματοποιείται συχνά τυποποίηση των δεδομένων μέσω z-score standardization.

Σε αντίθεση με τις ιεραρχικές μεθόδους, οι μη ιεραρχικές μέθοδοι, με κυριότερη την k-means, απαιτούν τον εκ των προτέρων καθορισμό του αριθμού των συστάδων. Η k-means στοχεύει στην ελαχιστοποίηση της ενδο-ομαδικής διακύμανσης, όπως εκφράζεται από τη συνάρτηση:

$$\sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

όπου μ_k είναι το κέντρο κάθε ομάδας.

Το centroid υπολογίζεται ως ο μέσος όρος των παρατηρήσεων της συστάδας:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

Η μέθοδος λειτουργεί επαναληπτικά, αναθέτοντας κάθε παρατήρηση στο πλησιέστερο κέντρο και επαναυπολογίζοντας τα κέντρα μέχρι να επιτευχθεί σύγκλιση (James et al., 2021). Η επιλογή του αριθμού των συστάδων αποτελεί κρίσιμο ζήτημα στη μη ιεραρχική ανάλυση. Μία από τις πιο διαδεδομένες τεχνικές είναι η μέθοδος elbow, η οποία βασίζεται στην παρατήρηση της μείωσης της ενδο-ομαδικής διακύμανσης, καθώς και ο δείκτης silhouette, ο οποίος αξιολογεί την ποιότητα της ομαδοποίησης.

Συνολικά, οι ιεραρχικές και μη ιεραρχικές μέθοδοι παρουσιάζουν διαφορετικά πλεονεκτήματα. Οι ιεραρχικές μέθοδοι προσφέρουν καλύτερη κατανόηση της δομής των δεδομένων, ενώ οι μη ιεραρχικές είναι πιο αποδοτικές και οδηγούν σε πιο σταθερά αποτελέσματα σε μεγάλα σύνολα δεδομένων (Κούτρας, 2025; James et al., 2021). Στην πράξη, οι δύο προσεγγίσεις χρησιμοποιούνται συμπληρωματικά.

Η ανάλυση κατά συστάδες έχει ευρεία εφαρμογή σε πολλούς τομείς, συμπεριλαμβανομένης της ανάλυσης αθλητικών δεδομένων. Στο πλαίσιο της παρούσας εργασίας, χρησιμοποιείται για την ομαδοποίηση ομάδων καλαθοσφαίρισης με βάση στατιστικούς δείκτες απόδοσης, με στόχο την κατανόηση των παραγόντων που σχετίζονται με την πρόκριση σε επόμενες φάσεις της διοργάνωσης.

6.2 Μέθοδος Ward

Αρχικά εφαρμόστηκε η ιεραρχική μέθοδος ανάλυσης κατά συστάδες με χρήση του αλγορίθμου του κοντινότερου γείτονα (nearest neighbor), με σκοπό τη διερεύνηση της δομής των δεδομένων και τον αρχικό εντοπισμό πιθανών ομάδων μεταξύ των παρατηρήσεων. Ωστόσο, τα αποτελέσματα της μεθόδου δεν κρίθηκαν ικανοποιητικά, καθώς παρατηρήθηκε έντονα το φαινόμενο της αλυσίδας (chaining effect), με αποτέλεσμα τη δημιουργία μίας εξαιρετικά μεγάλης συστάδας και μίας μονομελούς συστάδας. Η εικόνα αυτή υποδήλωνε αδυναμία ουσιαστικού διαχωρισμού των ομάδων και περιορισμένη ερμηνευτική αξία της τελικής ομαδοποίησης. Για τον λόγο αυτό, στη συνέχεια εφαρμόστηκε η μέθοδος Ward, η οποία θεωρείται καταλληλότερη για τη δημιουργία περισσότερο συμπαγών και ισορροπημένων συστάδων. Η μέθοδος Ward παρήγαγε σημαντικά πιο σταθερά και ερμηνεύσιμα αποτελέσματα, επιτυγχάνοντας ουσιαστικότερο διαχωρισμό των ομάδων που προκρίθηκαν στο Final 8 από εκείνες που δεν προκρίθηκαν. Συνεπώς, η τελική ερμηνεία της ανάλυσης κατά συστάδες βασίστηκε στα αποτελέσματα της μεθόδου Ward.

Από το δένδροδιάγραμμα (Σχήμα 6.2) παρατηρείται ότι οι παρατηρήσεις οργανώνονται σε δύο βασικές συστάδες, οι οποίες προκύπτουν μέσω σταδιακών συγχωνεύσεων ομάδων με σχετικά παρόμοια χαρακτηριστικά. Η δομή του δένδροδιαγράμματος δείχνει ότι οι συγχωνεύσεις πραγματοποιούνται σε διαφορετικά επίπεδα απόστασης, γεγονός που υποδηλώνει την ύπαρξη διαφοροποίησης μεταξύ των ομάδων ως προς τα αγωνιστικά στατιστικά που χρησιμοποιήθηκαν στην ανάλυση. Παράλληλα, οι συστάδες εμφανίζονται σχετικά συμπαγείς και δεν παρατηρούνται ακραίες ή απομονωμένες παρατηρήσεις.

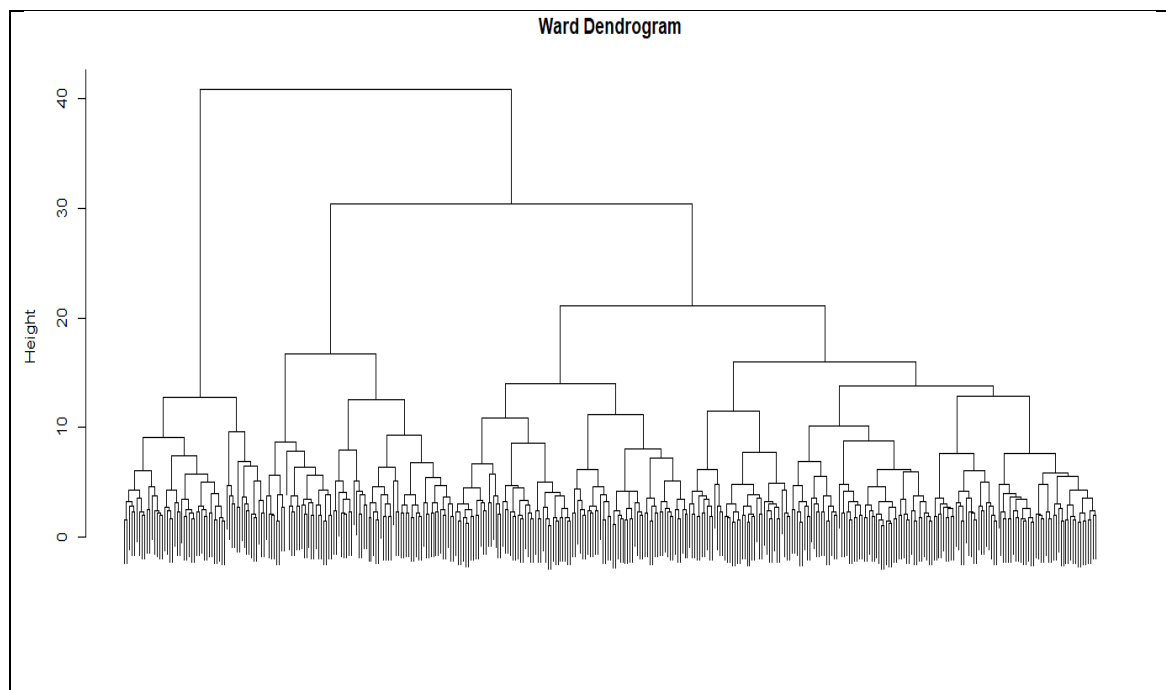
Η εικόνα αυτή επιβεβαιώνεται και από τον πίνακα συχνοτήτων των συστάδων που παρουσιάζεται στον πίνακα 6.2. Συγκεκριμένα, η πρώτη συστάδα περιλαμβάνει 219 ομάδες που δεν προκρίθηκαν στο Final 8 και 115 ομάδες που προκρίθηκαν, ενώ η δεύτερη συστάδα περιλαμβάνει 4 ομάδες που δεν προκρίθηκαν και 53 ομάδες που προκρίθηκαν. Επομένως, η δεύτερη συστάδα χαρακτηρίζεται από ιδιαίτερα υψηλή συγκέντρωση ομάδων που πέτυχαν πρόκριση, γεγονός που υποδηλώνει ότι οι ομάδες αυτές παρουσιάζουν παρόμοιο προφίλ αγωνιστικής απόδοσης και κοινά στατιστικά χαρακτηριστικά.

Η πρώτη συστάδα, σε αντίθεση με τη δεύτερη, περιλαμβάνει κυρίως ομάδες που δεν προκρίθηκαν στο Final 8, αν και σε αυτή εντάσσονται και αρκετές ομάδες που πέτυχαν την πρόκριση. Το αποτέλεσμα αυτό υποδηλώνει ότι, παρότι οι μεταβλητές της ανάλυσης συμβάλλουν ουσιαστικά στη διαφοροποίηση των ομάδων, δεν οδηγούν σε απόλυτο διαχωρισμό μεταξύ επιτυχημένων και μη επιτυχημένων ομάδων. Επομένως, φαίνεται ότι ορισμένες ομάδες που προκρίθηκαν παρουσιάζουν αγωνιστικό προφίλ παρόμοιο με εκείνο ομάδων που δεν προκρίθηκαν, γεγονός που αναδεικνύει την πολυπαραγοντική φύση της αγωνιστικής επιτυχίας.

Συνολικά, τα αποτελέσματα της ανάλυσης δείχνουν ότι η μέθοδος Ward καταφέρνει να εντοπίσει μία συστάδα ομάδων με αυξημένη πιθανότητα επιτυχίας και πρόκρισης στο Final 8, βασιζόμενη στα αγωνιστικά χαρακτηριστικά των ομάδων. Η ομαδοποίηση που προκύπτει θεωρείται ερμηνεύσιμη και παρέχει χρήσιμη πληροφορία σχετικά με τη διαφοροποίηση των ομάδων ως προς την αγωνιστική τους αποτελεσματικότητα.

	No	Yes
1	219	115
2	4	53

Πίνακας 6.2: Πίνακας συχνοτήτων



Σχήμα 6.2 Δενδροδιάγραμμα

6.3 Μέθοδος k-means

Στη συνέχεια εφαρμόστηκε η μη ιεραρχική μέθοδος k-means, με σκοπό τον διαχωρισμό των ομάδων σε δύο συστάδες, που αντιστοιχούν σε ομάδες που προκρίθηκαν και σε ομάδες που δεν προκρίθηκαν στο Final8. Η εφαρμογή της μη ιεραρχικής μεθόδου k-means με αριθμό συστάδων $k = 2$ οδήγησε σε σαφή διαχωρισμό των ομάδων σε δύο διακριτές κατηγορίες. Η σύγκριση των συστάδων με τη μεταβλητή Final8_Status δείχνει ότι η πρώτη αποτελείται κυρίως από ομάδες που δεν προκρίθηκαν στο Final8 (187 έναντι 43 που προκρίθηκαν), ενώ η δεύτερη περιλαμβάνει κυρίως ομάδες που προκρίθηκαν (125 έναντι 36 που δεν προκρίθηκαν) όπως φαίνεται στον πίνακα 6.3.1.

	No	Yes
1	187	43
2	36	125

Πίνακας 6.3.1 Πίνακας Συχνοτήτων K-means

Η διαφοροποίηση αυτή επιβεβαιώνεται και από την ανάλυση των μέσων όρων των μεταβλητών ανά συστάδα όπως φαίνεται στον πίνακα 6.3.2. Συγκεκριμένα, η δεύτερη συστάδα παρουσιάζει σημαντικά υψηλότερες τιμές σε βασικούς δείκτες απόδοσης, όπως πόντους ανά αγώνα (PPG = 74.08 έναντι 65.80), ποσοστά ευστοχίας (2P% = 48.72 έναντι 44.28 και 3P% = 34.42 έναντι 30.82), καθώς και ασίστ (APG = 16.48 έναντι 12.90). Επιπλέον, οι ομάδες της δεύτερης εμφανίζουν υψηλότερο συνολικό δείκτη αποδοτικότητας (EFFPG = 83.73 έναντι 64.84) και λιγότερα λάθη (TOPG = 14.51 έναντι 16.00), γεγονός που υποδηλώνει καλύτερη αγωνιστική οργάνωση. Αντίθετα, η πρώτη συστάδα χαρακτηρίζεται από χαμηλότερες επιδόσεις στους περισσότερους δείκτες, γεγονός που το κατατάσσει σε ομάδες χαμηλότερης αγωνιστικής αποτελεσματικότητας.

Μεταβλητή	Συστάδα 1	Συστάδα 2
PPG	65.80	74.08
2P%	44.28	48.72
3P%	30.82	34.42
FT%	74.21	75.87
ORPG	10.37	10.71
DRPG	23.67	26.99
RPG	34.04	37.69
APG	12.90	16.48
STLPG	8.46	8.39
TOPG	16.00	14.51
BLKPG	1.96	2.80
PFG	18.20	16.59
EFFPG	64.84	83.73

Πίνακας 6.3.2 Μέσοι όροι στατιστικών δεικτών ανά συστάδα (k-means)

Η αξιολόγηση της ομαδοποίησης πραγματοποιήθηκε μέσω πίνακα σύγχυσης (confusion matrix), συγκρίνοντας τις προβλεπόμενες κατηγορίες με τις πραγματικές τιμές της μεταβλητής Final8_Status. Τα αποτελέσματα παρουσιάζονται στον πίνακα 6.3.3 και δείχνουν ότι 187 ομάδες που δεν προκρίθηκαν ταξινομήθηκαν σωστά, καθώς και 125 ομάδες που προκρίθηκαν επιτυχώς. Παράλληλα, καταγράφονται 36 περιπτώσεις ομάδων που δεν προκρίθηκαν αλλά ταξινομήθηκαν ως επιτυχημένες, καθώς και 43 περιπτώσεις ομάδων που προκρίθηκαν χωρίς να εντοπιστούν σωστά από το μοντέλο.

Ο πίνακας σύγχυσης προκύπτει από την αντιστοίχιση των συστάδων με τις κατηγορίες πρόκρισης, βάσει της πλειοψηφίας των παρατηρήσεων σε κάθε συστάδα, γεγονός που εξηγεί τη δομική του αντιστοιχία με τον πίνακα συχνотήτων των συστάδων. Η συνολική ακρίβεια της ταξινόμησης όπως φαίνεται και στον πίνακα 6.3.3 ανέρχεται στο 79.8%, γεγονός που υποδηλώνει υψηλό βαθμό συμφωνίας μεταξύ της ομαδοποίησης και της πραγματικής πρόκρισης. Το αποτέλεσμα αυτό κρίνεται ιδιαίτερα ικανοποιητικό, δεδομένου ότι η μέθοδος k-means είναι μη επιβλεπόμενη και δεν χρησιμοποιεί εκ των προτέρων πληροφορία σχετικά με την πρόκριση των ομάδων.

Πρόβλεψη	No	Yes
No	187	43
Yes	36	125

Πίνακας 6.3.3 Πίνακας σύγχυσης (Confusion Matrix)

Accuracy	0.797954
----------	----------

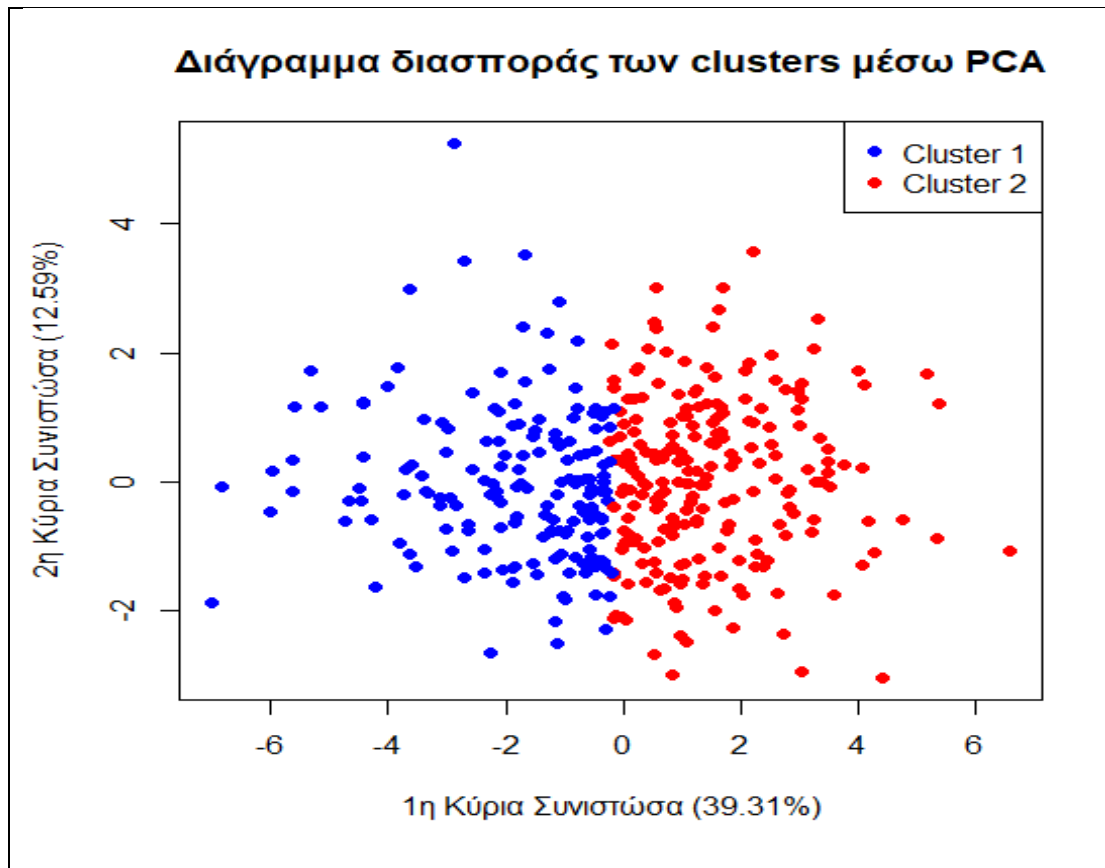
Πίνακας 6.3.4 Συνολική ακρίβεια της ταξινόμησης

Για την οπτική αναπαράσταση των πολυδιάστατων δεδομένων εφαρμόστηκε ανάλυση κύριων συνιστωσών (Principal Component Analysis – PCA), με στόχο τη μείωση της διάστασης των δεδομένων και την απεικόνιση των παρατηρήσεων σε δισδιάστατο χώρο. Η μέθοδος PCA επιτρέπει τον μετασχηματισμό των αρχικών μεταβλητών σε νέες μη συσχετισμένες μεταβλητές, τις κύριες συνιστώσες, οι οποίες διατηρούν το μεγαλύτερο μέρος της συνολικής διακύμανσης των δεδομένων. Στην παρούσα ανάλυση χρησιμοποιήθηκαν οι δύο πρώτες κύριες συνιστώσες, ώστε να καταστεί δυνατή η γραφική απεικόνιση των ομάδων που προέκυψαν από τη μέθοδο k-means.

Η Ανάλυση Κυρίων Συνιστωσών χρησιμοποιείται συχνά συμπληρωματικά στις μεθόδους ανάλυσης κατά συστάδες για την οπτική παρουσίαση πολυδιάστατων δεδομένων σε δισδιάστατο χώρο, διατηρώντας το μεγαλύτερο μέρος της πληροφορίας των αρχικών μεταβλητών (Κούτρας, 2025; James et al., 2021).

Το διάγραμμα διασποράς των συστάδων μέσω PCA παρουσιάζεται στο Σχήμα 6.3. Οι ομάδες που ανήκουν στην πρώτη συστάδα απεικονίζονται με μπλε χρώμα, ενώ οι ομάδες της δεύτερης συστάδας με κόκκινο χρώμα. Από την παρατήρηση του διαγράμματος διαπιστώνεται ότι οι δύο συστάδες παρουσιάζουν σχετικά σαφή διαχωρισμό κυρίως ως προς την πρώτη κύρια συνιστώσα, η οποία εξηγεί το 39,31% της συνολικής διακύμανσης των δεδομένων. Συγκεκριμένα, οι παρατηρήσεις της πρώτης συστάδας συγκεντρώνονται κυρίως σε αρνητικές τιμές της πρώτης κύριας συνιστώσας, ενώ οι παρατηρήσεις της δεύτερης συστάδας εμφανίζονται κυρίως σε θετικές τιμές. Αντίθετα, η δεύτερη κύρια συνιστώσα, η οποία εξηγεί το 12,59% της συνολικής διακύμανσης, δεν φαίνεται να συμβάλλει σημαντικά στον διαχωρισμό των ομάδων, καθώς οι δύο συστάδες παρουσιάζουν παρόμοια κατανομή κατά μήκος του κατακόρυφου άξονα.

Παράλληλα, παρατηρείται περιορισμένη επικάλυψη μεταξύ των δύο συστάδων στο κεντρικό τμήμα του διαγράμματος, ιδιαίτερα κοντά στο μηδέν της πρώτης κύριας συνιστώσας, γεγονός που δείχνει ότι ορισμένες ομάδες εμφανίζουν ενδιάμεσα χαρακτηριστικά απόδοσης και δεν μπορούν να διαχωριστούν πλήρως. Η εικόνα αυτή θεωρείται αναμενόμενη σε δεδομένα αθλητικής απόδοσης, όπου οι διαφορές μεταξύ των ομάδων δεν είναι απόλυτες. Ωστόσο, η συνολική δομή του διαγράμματος επιβεβαιώνει ότι η μέθοδος k-means κατόρθωσε να εντοπίσει δύο σχετικά διακριτά πρότυπα αγωνιστικής απόδοσης, τα οποία σχετίζονται σε σημαντικό βαθμό με την πρόκριση ή μη των ομάδων στο Final 8 της διοργάνωσης.



Σχήμα 6.3 Διάγραμμα διασποράς των συστάδων μέσω των δύο πρώτων κύριων συνιστωσών (PCA)

Συνολικά, τα αποτελέσματα υποδεικνύουν ότι οι βασικοί παράγοντες που σχετίζονται με την πρόκριση στο Final 8 είναι η υψηλή επιθετική παραγωγικότητα, η αποτελεσματικότητα στα σουτ, η δημιουργία παιχνιδιού μέσω ασίστ και η μείωση των λαθών. Η μέθοδος k-means καταφέρνει να αποτυπώσει με σαφήνεια αυτές τις διαφορές και να διαχωρίσει αποτελεσματικά τις ομάδες με βάση την απόδοσή τους.

6.4 Σύγκριση αποτελεσμάτων

Η σύγκριση των αποτελεσμάτων των διαφορετικών μεθόδων ανάλυσης κατά συστάδες αναδεικνύει σημαντικές διαφοροποιήσεις ως προς την ικανότητά τους να αποτυπώνουν τη δομή των δεδομένων και να διαχωρίζουν αποτελεσματικά τις ομάδες. Η ιεραρχική μέθοδος του κοντινότερου γείτονα (nearest neighbor) δεν κατάφερε να εντοπίσει ουσιαστικά διακριτές συστάδες, καθώς παρουσίασε έντονα το φαινόμενο της αλυσιδωτής σύνδεσης (chain effect), οδηγώντας στη δημιουργία μίας ιδιαίτερα μεγάλης συστάδας και μίας εξαιρετικά μικρής. Το αποτέλεσμα αυτό περιόρισε σημαντικά την ερμηνευτική αξία της συγκεκριμένης μεθόδου και δεν επέτρεψε ουσιαστικό διαχωρισμό των ομάδων. Αντίθετα, η εφαρμογή της μεθόδου Ward οδήγησε σε περισσότερο συμπαγείς και ισορροπημένες συστάδες, επιτυγχάνοντας σαφέστερο διαχωρισμό μεταξύ ομάδων που προκρίθηκαν και ομάδων που δεν προκρίθηκαν στο Final 8.

Η μέθοδος αυτή κατόρθωσε να εντοπίσει μία συστάδα με ιδιαίτερα υψηλή συγκέντρωση επιτυχημένων ομάδων, γεγονός που υποδηλώνει ότι συγκεκριμένα αγωνιστικά χαρακτηριστικά σχετίζονται με αυξημένη πιθανότητα πρόκρισης. Παράλληλα, η παρουσία ομάδων που προκρίθηκαν και στις δύο συστάδες δείχνει ότι η αγωνιστική επιτυχία δεν αποτυπώνεται πλήρως από τις μεταβλητές της ανάλυσης, αναδεικνύοντας την πολυπαραγοντική φύση της απόδοσης στον αθλητισμό. Η μη ιεραρχική μέθοδος k-means παρουσίασε ακόμη πιο σαφή και λειτουργικό διαχωρισμό των ομάδων σε δύο διακριτές κατηγορίες. Τα αποτελέσματα της ανάλυσης έδειξαν ότι οι ομάδες της δεύτερης συστάδας εμφάνιζαν σημαντικά υψηλότερες επιδόσεις σε βασικούς αγωνιστικούς δείκτες, όπως πόντους ανά αγώνα, ποσοστά ευστοχίας, ασίστ και συνολικό δείκτη αποδοτικότητας, ενώ παράλληλα παρουσίαζαν λιγότερα λάθη. Η ποιότητα της ομαδοποίησης επιβεβαιώθηκε και από τον πίνακα σύγχυσης, καθώς η συνολική ακρίβεια ταξινόμησης ανήλθε στο 79,8%, ποσοστό ιδιαίτερα υψηλό για μη επιβλεπόμενη μέθοδο. Επιπλέον, η οπτικοποίηση των αποτελεσμάτων μέσω της Ανάλυσης Κυρίων Συνιστωσών (PCA) επιβεβαίωσε τη σαφή διαφοροποίηση των δύο συστάδων κυρίως κατά μήκος της πρώτης κύριας συνιστώσας, η οποία εξηγούσε το μεγαλύτερο μέρος της συνολικής διακύμανσης των δεδομένων. Παρότι παρατηρήθηκε περιορισμένη επικάλυψη μεταξύ των ομάδων στο κεντρικό τμήμα του διαγράμματος, η συνολική εικόνα έδειξε ότι η k-means κατόρθωσε να εντοπίσει δύο σχετικά διακριτά πρότυπα αγωνιστικής απόδοσης που συνδέονται σε σημαντικό βαθμό με την πρόκριση στο Final 8.

Συνολικά, τα αποτελέσματα της παρούσας ανάλυσης δείχνουν ότι η μέθοδος k-means αποτελεί την πιο αποτελεσματική προσέγγιση για τα συγκεκριμένα δεδομένα, καθώς παρέχει σαφή, σταθερό και ερμηνεύσιμο διαχωρισμό των ομάδων.

Παράλληλα, η μέθοδος Ward αποδείχθηκε περισσότερο κατάλληλη από την κλασική ιεραρχική μέθοδο του κοντινότερου γείτονα, προσφέροντας πιο συνεκτικές και ουσιαστικές συστάδες.

Αντίθετα, η μέθοδος του κοντινότερου γείτονα παρουσίασε σημαντικούς περιορισμούς και λειτούργησε κυρίως σε διερευνητικό επίπεδο, χωρίς να συμβάλλει ουσιαστικά στην κατανόηση των παραγόντων που σχετίζονται με την πρόκριση των ομάδων.

6.5 Σύγκριση αποτελεσμάτων με τη διεθνή βιβλιογραφία

Τα ευρήματα της παρούσας εργασίας ενισχύονται από πρόσφατες ερευνητικές μελέτες που αξιοποιούν μεθόδους ανάλυσης κατά συστάδες στον χώρο του γυναικείου μπάσκετ. Συγκεκριμένα, η μελέτη των Réndola-Reinecke et al. (2025) εφάρμοσε συνδυασμό ανάλυσης κύριων συνιστωσών (PCA) και της μεθόδου k-means σε δεδομένα της EuroLeague γυναικών, με στόχο την αναγνώριση προτύπων αγωνιστικής απόδοσης των παικτών. Μέσω της ανάλυσης αυτής, προέκυψαν διακριτές κατηγορίες αγωνιστικών ρόλων, όπως οι οργανωτές παιχνιδιού, οι παίκτριες με πολύπλευρη αγωνιστική συμβολή και οι εξειδικευμένοι παίκτες.

Τα αποτελέσματα της μελέτης ανέδειξαν ότι οι σύγχρονες αγωνιστικές απαιτήσεις δεν περιορίζονται στις παραδοσιακές θέσεις των παικτριών, αλλά χαρακτηρίζονται από μεγαλύτερη ευελιξία και πολυλειτουργικότητα, γεγονός που αντανακλάται στους στατιστικούς δείκτες απόδοσης. Παράλληλα, η μελέτη των Courel-Ibáñez et al. (2024), βασισμένη σε δεδομένα δέκα αγωνιστικών περιόδων του ισπανικού πρωταθλήματος γυναικών, αξιοποίησε επίσης τεχνικές ανάλυσης κατά συστάδες για την κατηγοριοποίηση των παικτών σε διακριτά προφίλ απόδοσης. Η ανάλυση αυτή οδήγησε στον εντοπισμό διαφορετικών επιπέδων αγωνιστικής αποτελεσματικότητας, με τις παίκτριες υψηλής απόδοσης να παρουσιάζουν σημαντικά καλύτερες επιδόσεις σε βασικούς στατιστικούς δείκτες, όπως οι πόντοι, οι ασίστ και τα ριμπάουντ. Επιπλέον, διαπιστώθηκε ότι οι παίκτριες αυτές εμφανίζονται συχνότερα σε ομάδες υψηλής κατάταξης, γεγονός που υποδηλώνει ισχυρή σχέση μεταξύ ατομικής απόδοσης και συλλογικής επιτυχίας. Τα ευρήματα των παραπάνω μελετών παρουσιάζουν σημαντική σύγκλιση με τα αποτελέσματα της παρούσας εργασίας. Ειδικότερα, η εφαρμογή της μεθόδου k-means οδήγησε σε σαφή διαχωρισμό των ομάδων σε συστάδες που αντανακλούν διαφορετικά επίπεδα αγωνιστικής απόδοσης. Οι ομάδες που εντάσσονται στη συστάδα υψηλής απόδοσης χαρακτηρίζονται από αυξημένη επιθετική παραγωγικότητα, υψηλότερα ποσοστά ευστοχίας, μεγαλύτερο αριθμό ασίστ και καλύτερο έλεγχο της μπάλας, όπως αποτυπώνεται από τον μικρότερο αριθμό λαθών. Αντίθετα, οι ομάδες χαμηλότερης απόδοσης εμφανίζουν υστέρηση στους παραπάνω δείκτες.

Η συμφωνία των αποτελεσμάτων με τη διεθνή βιβλιογραφία ενισχύει την εγκυρότητα της μεθοδολογικής προσέγγισης που ακολουθήθηκε και επιβεβαιώνει ότι η ανάλυση κατά συστάδες αποτελεί ένα αξιόπιστο και αποτελεσματικό εργαλείο για την κατανόηση της αγωνιστικής απόδοσης και τον εντοπισμό των βασικών παραγόντων που σχετίζονται με την επιτυχία στο γυναικείο μπάσκετ.

Κεφάλαιο 7: Συμπεράσματα και Συζήτηση Αποτελεσμάτων

Στο παρόν κεφάλαιο παρουσιάζονται συνοπτικά τα βασικά ευρήματα της παρούσας διπλωματικής εργασίας και πραγματοποιείται ερμηνεία των αποτελεσμάτων που προέκυψαν από τη στατιστική ανάλυση των δεδομένων της EuroLeague Γυναικών. Αρχικά, συνοψίζονται τα σημαντικότερα συμπεράσματα σχετικά με τους αγωνιστικούς δείκτες που συνδέονται με την επιτυχία και την πρόκριση των ομάδων στις τελικές φάσεις της διοργάνωσης. Στη συνέχεια, τα αποτελέσματα συγκρίνονται με αντίστοιχα ευρήματα της διεθνούς βιβλιογραφίας και με σχετικές μελέτες που αφορούν την EuroLeague Ανδρών, με στόχο την ανάδειξη πιθανών ομοιοτήτων και διαφορών μεταξύ του ανδρικού και του γυναικείου ευρωπαϊκού μπάσκετ. Τέλος, παρουσιάζονται γενικά συμπεράσματα σχετικά με τη συμβολή της στατιστικής ανάλυσης στην κατανόηση της αγωνιστικής απόδοσης στο σύγχρονο μπάσκετ.

7.1 Συμπεράσματα της παρούσας έρευνας

Η παρούσα διπλωματική εργασία επικεντρώθηκε στη μελέτη των βασικών προβλεπτικών δεικτών που σχετίζονται με την αγωνιστική επιτυχία των ομάδων στην EuroLeague Γυναικών. Μέσα από την αξιοποίηση επίσημων αγωνιστικών δεδομένων της FIBA για την περίοδο 2004–2025, επιχειρήθηκε η διερεύνηση των στατιστικών μεταβλητών που επηρεάζουν περισσότερο την εξέλιξη και την τελική έκβαση ενός αγώνα, καθώς και η συνολική πορεία των ομάδων στη διοργάνωση.

Αρχικά, μέσω της περιγραφικής στατιστικής ανάλυσης παρατηρήθηκαν σαφείς διαφοροποιήσεις μεταξύ των ομάδων ανάλογα με την αγωνιστική τους πορεία στη διοργάνωση. Οι ομάδες που κατάφεραν να προκριθούν στις τελικές φάσεις της EuroLeague παρουσίαζαν υψηλότερες τιμές σε μεταβλητές που συνδέονται με την επιθετική αποτελεσματικότητα και τη συνολική ποιότητα του παιχνιδιού τους. Ειδικότερα, οι ομάδες που έφταναν στο Final 4 εμφάνιζαν μεγαλύτερους μέσους όρους σε πόντους ανά αγώνα (PPG), ποσοστά ευστοχίας στα δίποντα (2P%), ασίστ (APG), αμυντικά ριμπάουντ (DRPG) και στον δείκτη αποδοτικότητας (EFFPG), γεγονός που δείχνει ότι η αγωνιστική επιτυχία σχετίζεται άμεσα με την αποτελεσματική επιθετική λειτουργία, τη συνεργασία μεταξύ των παικτριών και τον έλεγχο του ρυθμού του παιχνιδιού. Αντίθετα, οι ομάδες που δεν κατάφεραν να προκριθούν στις επόμενες φάσεις της διοργάνωσης παρουσίαζαν μεγαλύτερο αριθμό λαθών (TOPG), χαμηλότερη αποτελεσματικότητα στην εκτέλεση των επιθέσεων και μικρότερη συνολική αποδοτικότητα. Το εύρημα αυτό ενισχύει την άποψη ότι η σωστή διαχείριση των κατοχών και η αποφυγή λαθών αποτελούν κρίσιμους παράγοντες για την επιτυχία σε υψηλό επίπεδο ανταγωνισμού. Η γραφική διερεύνηση των δεδομένων μέσω scatter plots, violin plots, line plots και Chernoff Faces ανέδειξε επιπλέον τη δομή και τη συμπεριφορά των μεταβλητών ανάμεσα στις διαφορετικές κατηγορίες ομάδων. Μέσα από τα διαγράμματα παρατηρήθηκε ότι οι επιτυχημένες ομάδες παρουσιάζουν μεγαλύτερη σταθερότητα στους βασικούς αγωνιστικούς δείκτες και μικρότερη διακύμανση στην απόδοσή τους κατά τη διάρκεια των σεζόν.

Παράλληλα, η θετική σχέση του δείκτη αποδοτικότητας (EFFPG) με μεταβλητές όπως οι πόντοι, οι ασίστ και τα ριμπάουντ φάνηκε ιδιαίτερα έντονη, γεγονός που υποδηλώνει ότι ο συγκεκριμένος δείκτης μπορεί να χρησιμοποιηθεί ως συνολικός δείκτης αγωνιστικής ποιότητας μιας ομάδας. Στη συνέχεια, μέσω των στατιστικών ελέγχων διαφορών μεταξύ ομάδων διαπιστώθηκε ότι αρκετές από τις μεταβλητές διαφοροποιούν σημαντικά τις ομάδες υψηλής επιτυχίας από τις υπόλοιπες. Η ευστοχία στα δίποντα, τα αμυντικά ριμπάουντ, οι ασίστ και ο δείκτης αποδοτικότητας παρουσίασαν στατιστικά σημαντικές διαφορές μεταξύ των ομάδων που προκρίθηκαν και εκείνων που αποκλείστηκαν νωρίτερα από τη διοργάνωση. Αντίθετα, μεταβλητές όπως τα λάθη και τα προσωπικά φάουλ συνδέθηκαν αρνητικά με την αγωνιστική επιτυχία.

Ιδιαίτερα σημαντικά ήταν και τα αποτελέσματα της ανάλυσης παλινδρόμησης. Τα μοντέλα λογιστικής και πολυωνυμικής λογιστικής παλινδρόμησης που εφαρμόστηκαν στην παρούσα εργασία επέτρεψαν την εκτίμηση της πιθανότητας πρόκρισης των ομάδων στις τελικές φάσεις της διοργάνωσης. Από τα αποτελέσματα προέκυψε ότι ο δείκτης αποδοτικότητας (EFFPG), οι ασίστ, οι πόντοι ανά αγώνα και τα αμυντικά ριμπάουντ αποτελούν από τους σημαντικότερους προβλεπτικούς παράγοντες της επιτυχίας. Παράλληλα, τα μοντέλα παρουσίασαν ικανοποιητική προβλεπτική ικανότητα, όπως φάνηκε μέσα από τις τιμές AUC και τους ελέγχους προσαρμογής, γεγονός που υποδεικνύει ότι οι στατιστικές μέθοδοι μπορούν να αξιοποιηθούν αποτελεσματικά στην πρόβλεψη της αγωνιστικής πορείας των ομάδων.

Επιπλέον, η εφαρμογή μεθόδων ανάλυσης κατά συστάδες ανέδειξε την ύπαρξη ομάδων με παρόμοια αγωνιστικά χαρακτηριστικά. Οι ομάδες υψηλής επιτυχίας συγκεντρώνονταν σε συστάδες που χαρακτηρίζονταν από μεγαλύτερη επιθετική αποτελεσματικότητα, υψηλότερη συνολική αποδοτικότητα και καλύτερη διαχείριση των κατοχών. Αντίθετα, οι λιγότερο επιτυχημένες ομάδες εμφάνιζαν μεγαλύτερη αστάθεια στην απόδοσή τους και χαμηλότερη συνολική αποτελεσματικότητα. Τα αποτελέσματα αυτά επιβεβαιώνουν ότι η επιτυχία στο σύγχρονο γυναικείο ευρωπαϊκό μπάσκετ δεν εξαρτάται από έναν μόνο παράγοντα, αλλά αποτελεί συνδυασμό επιθετικής ποιότητας, αμυντικής σταθερότητας και ομαδικής λειτουργίας.

Συνολικά, η παρούσα έρευνα ανέδειξε ότι οι βασικοί παράγοντες που σχετίζονται με την αγωνιστική επιτυχία στην EuroLeague Γυναικών είναι η αποτελεσματικότητα στην επίθεση, η σωστή κυκλοφορία της μπάλας μέσω των ασίστ, ο έλεγχος των αμυντικών ριμπάουντ και η περιορισμένη πραγματοποίηση λαθών. Η στατιστική ανάλυση των δεδομένων προσφέρει σημαντικές πληροφορίες σχετικά με τη δομή και τα χαρακτηριστικά των επιτυχημένων ομάδων και μπορεί να αποτελέσει χρήσιμο εργαλείο τόσο για την ερευνητική μελέτη του αθλήματος όσο και για την πρακτική αξιοποίηση από προπονητές, αναλυτές απόδοσης και αθλητικούς οργανισμούς.

7.2 Σύγκριση με τη διεθνή βιβλιογραφία και την EuroLeague Ανδρών

Τα αποτελέσματα της παρούσας έρευνας παρουσιάζουν σημαντικές ομοιότητες με τα ευρήματα προηγούμενων μελετών που αφορούν το γυναικείο και το ευρωπαϊκό μπάσκετ υψηλού επιπέδου. Ιδιαίτερα έντονη είναι η συμφωνία με τις μελέτες των Gómez, Lorenzo, Sampaio και Ibáñez (2006), των Gasperi, Gómez, Lago-Peñas και Sampaio (2020), καθώς και με τη διπλωματική εργασία του Καμίτση (2023) για την EuroLeague Ανδρών.

Αρχικά, οι Gómez et al. (2006), μελετώντας αγώνες γυναικείου μπάσκετ υψηλού επιπέδου, κατέληξαν στο συμπέρασμα ότι οι νικήτριες ομάδες υπερτερούν κυρίως σε μεταβλητές όπως η ευστοχία στα δίποντα, τα αμυντικά ριμπάουντ και οι ασίστ. Τα αποτελέσματα της παρούσας εργασίας επιβεβαιώνουν σε μεγάλο βαθμό τα συγκεκριμένα ευρήματα, καθώς οι ομάδες που προκρίνονταν στις τελικές φάσεις της EuroLeague Γυναικών παρουσίαζαν υψηλότερα ποσοστά ευστοχίας στα δίποντα, μεγαλύτερο αριθμό ασίστ και καλύτερη επίδοση στα αμυντικά ριμπάουντ συγκριτικά με τις ομάδες που αποκλείονταν νωρίτερα από τη διοργάνωση. Η συμφωνία αυτή ενισχύει την άποψη ότι η αγωνιστική επιτυχία στο γυναικείο μπάσκετ σχετίζεται σε μεγάλο βαθμό με την ποιότητα της επιθετικής ανάπτυξης και την αποτελεσματική διαχείριση των κατοχών.

Παράλληλα, οι Gómez et al. υποστήριζαν ότι στα ισορροπημένα παιχνίδια η ομαδική συνεργασία και η αποτελεσματικότητα στις επιθέσεις παίζουν καθοριστικό ρόλο στην τελική έκβαση του αγώνα. Το συμπέρασμα αυτό φαίνεται να επιβεβαιώνεται και στην παρούσα εργασία μέσω της σημαντικής επίδρασης των ασίστ και του δείκτη αποδοτικότητας (EFFPG) στα μοντέλα παλινδρόμησης που εφαρμόστηκαν. Οι ομάδες με υψηλότερη κυκλοφορία της μπάλας και καλύτερη συνολική επιθετική λειτουργία εμφάνιζαν μεγαλύτερη πιθανότητα πρόκρισης στις επόμενες φάσεις της διοργάνωσης.

Αντίστοιχα, ιδιαίτερα σημαντική είναι και η συμφωνία με τη μελέτη των Gasperi et al. (2020), οι οποίοι ανέλυσαν δεδομένα της EuroLeague Γυναικών και διαπίστωσαν ότι οι ομάδες υψηλής δυναμικότητας παρουσιάζουν καλύτερη επιθετική αποτελεσματικότητα, μεγαλύτερο αριθμό ασίστ και αποτελεσματικότερη διαχείριση του ρυθμού του παιχνιδιού. Στην παρούσα εργασία, οι ίδιες μεταβλητές αναδείχθηκαν ως από τους σημαντικότερους παράγοντες που σχετίζονται με την αγωνιστική επιτυχία. Συγκεκριμένα, οι ομάδες που έφταναν στο Final 4 παρουσίαζαν υψηλότερο δείκτη αποδοτικότητας, περισσότερες ασίστ και μικρότερο αριθμό λαθών, γεγονός που υποδηλώνει μεγαλύτερη σταθερότητα και καλύτερο έλεγχο των κατοχών. Επιπλέον, οι Gasperi et al. τόνισαν ότι οι δείκτες απόδοσης δεν λειτουργούν μεμονωμένα αλλά συνδυαστικά, καθώς η επιτυχία εξαρτάται από την ισορροπία μεταξύ επιθετικής αποτελεσματικότητας και αμυντικής σταθερότητας.

Το συγκεκριμένο συμπέρασμα συμφωνεί πλήρως με τα αποτελέσματα της παρούσας εργασίας, όπου οι πιο επιτυχημένες ομάδες δεν παρουσίαζαν υπεροχή μόνο σε μία μεταβλητή, αλλά συνολικά καλύτερη εικόνα σε πολλαπλούς αγωνιστικούς δείκτες, όπως οι πόντοι ανά αγώνα, τα ποσοστά ευστοχίας, τα αμυντικά ριμπάουντ, οι ασίστ και ο δείκτης EFFPG.

Ιδιαίτερο ενδιαφέρον παρουσιάζει και η σύγκριση με τη διπλωματική εργασία του Καμίτση (2023), η οποία επικεντρώθηκε στη EuroLeague Ανδρών και στην αναζήτηση των βασικών προβλεπτικών παραγόντων που επηρεάζουν την πρόκριση των ομάδων στα Playoffs και στο Final Four. Στην εργασία εκείνη εφαρμόστηκαν αντίστοιχες στατιστικές τεχνικές, όπως περιγραφική ανάλυση, λογιστική παλινδρόμηση, clustering και classification models, προκειμένου να αναγνωριστούν τα χαρακτηριστικά των επιτυχημένων ομάδων. Τα αποτελέσματα των δύο εργασιών εμφανίζουν αρκετές ομοιότητες ως προς τους βασικούς δείκτες επιτυχίας. Στη διπλωματική του Καμίτση, οι πόντοι, ο δείκτης PIR και η συνολική επιθετική αποτελεσματικότητα αναδείχθηκαν ως βασικοί παράγοντες πρόκρισης των ομάδων της EuroLeague Ανδρών. Αντίστοιχα, στην παρούσα εργασία οι πόντοι ανά αγώνα, ο δείκτης αποδοτικότητας EFFPG και οι ασίστ παρουσίασαν ιδιαίτερα ισχυρή σχέση με την αγωνιστική επιτυχία των ομάδων της EuroLeague Γυναικών. Επιπλέον, και στις δύο έρευνες τα μοντέλα λογιστικής παλινδρόμησης παρουσίασαν ικανοποιητική προβλεπτική ικανότητα, γεγονός που δείχνει ότι οι συγκεκριμένοι δείκτες μπορούν να αξιοποιηθούν αποτελεσματικά στην πρόβλεψη της πορείας των ομάδων. Παρόλα αυτά, παρατηρούνται και ορισμένες διαφοροποιήσεις μεταξύ του ανδρικού και του γυναικείου ευρωπαϊκού μπάσκετ. Στην εργασία του Καμίτση δίνεται μεγαλύτερη έμφαση στη συνεισφορά των κορυφαίων παικτών μέσω του δείκτη PIR, καθώς σημαντικό μέρος της ανάλυσης επικεντρώνεται στην επίδραση των παικτών με την υψηλότερη ατομική απόδοση στην πορεία των ομάδων τους. Αντίθετα, στην παρούσα εργασία η έμφαση δίνεται περισσότερο στη συνολική ομαδική λειτουργία και σε μεταβλητές που σχετίζονται με τη συνεργασία μεταξύ των παικτριών, όπως οι ασίστ και η συνολική κυκλοφορία της μπάλας. Επιπλέον, στην EuroLeague Γυναικών τα αμυντικά ριμπάουντ και η σταθερότητα στην επιθετική ανάπτυξη φαίνεται να έχουν ιδιαίτερα αυξημένη σημασία για την επιτυχία των ομάδων. Το στοιχείο αυτό πιθανόν συνδέεται με τη διαφορετική αγωνιστική φυσιολογία του γυναικείου μπάσκετ, όπου η οργανωμένη επιθετική λειτουργία και η αποτελεσματική διαχείριση των κατοχών διαδραματίζουν πιο καθοριστικό ρόλο στη διαμόρφωση του τελικού αποτελέσματος. Συνολικά, τα αποτελέσματα της παρούσας εργασίας επιβεβαιώνουν σε μεγάλο βαθμό τα συμπεράσματα τόσο της διεθνούς βιβλιογραφίας όσο και της αντίστοιχης έρευνας για την EuroLeague Ανδρών. Οι μεταβλητές που σχετίζονται με την επιθετική αποτελεσματικότητα, τον έλεγχο των κατοχών και τη συνολική αποδοτικότητα των ομάδων αποτελούν κοινούς και σταθερούς παράγοντες επιτυχίας και στις δύο διοργανώσεις. Παράλληλα, οι επιμέρους διαφοροποιήσεις που παρατηρούνται αναδεικνύουν τη διαφορετική αγωνιστική ταυτότητα του ανδρικού και του γυναικείου ευρωπαϊκού μπάσκετ και επιβεβαιώνουν την ανάγκη ξεχωριστής στατιστικής προσέγγισης για κάθε αγωνιστικό περιβάλλον.

Βιβλιογραφία

Ελληνική Βιβλιογραφία

Ηλιόπουλος, Γ. (2023). Σημειώσεις μαθήματος Γενικευμένων Γραμμικών Μοντέλων. Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς.

Καμίτσης, Α. (2023). Δείκτες για την αξιολόγηση της απόδοσης παικτών και ομάδων σε αγώνες μπάσκετ και παράγοντες που τους επηρεάζουν (Διπλωματική εργασία, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Π.Μ.Σ. Εφαρμοσμένη Στατιστική)

Κούτρας, Μ. (2021–2023). Σημειώσεις Εφαρμοσμένης Πολυμεταβλητής Ανάλυσης. Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς.

Κούτρας, Μ., & Ευαγγελάρας, Χ. (2018). Ανάλυση Παλινδρόμησης: Θεωρία και Εφαρμογές. Θεσσαλονίκη: Εκδόσεις Τζιόλα.

Μπερσίμης, Σ., Μπαρτζής, Γ., Παπαδάκης, Γ., & Σάχλας, Α. (2021). Εφαρμοσμένη Στατιστική και Στατιστική Μηχανική Μάθηση με χρήση των IBM SPSS Statistics, R, Python. Θεσσαλονίκη: Εκδόσεις Τζιόλα.

Πολίτης, Κ. (2023–2024). Σημειώσεις και διαφάνειες μαθήματος Γενικευμένα Γραμμικά Μοντέλα. Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική, Πανεπιστήμιο Πειραιώς.

Ξενόγλωσση Βιβλιογραφία

Bonett, D. G. (2019). Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical and Statistical Psychology*.

Courel-Ibáñez, J., Piñar-López, M. I., Contreras-García, J. M., & Ibáñez, S. J. (2025). Profiling high performers in elite women's basketball: Functional roles and normative benchmarks from 10 seasons of Spanish First Division data.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*,

Gasperi, L., Gómez, M. A., Lago-Peñas, C., & Sampaio, J. (2020). Game related statistics discriminate national and foreign players according to playing position and team ability in the Women's Basketball EuroLeague.

Gómez, M. A., Lorenzo, A., Sampaio, J., & Ibáñez, S. J. (2006). Differences between winning and losing teams in women's basketball game-related statistics.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*.

Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression (2nd ed.). New York: Wiley.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. New York: Springer.

Péndola-Reinecke, M. I., Jiménez-Sáiz, S., Mochales Cuesta, I., & Bustamante-Sánchez, Á. (2025). Exploring new dimensions in the classification of positions in women's basketball: A statistical approach. Applied Sciences, 15, 6159.

Διαδικτυακές Πηγές

FIBA Women's EuroLeague. Διαθέσιμο στο:
<https://www.fiba.basketball/euroleaguewomen/>

Wikipedia contributors. EuroLeague Women. Διαθέσιμο στο:
https://en.wikipedia.org/wiki/EuroLeague_Women

DAZN. Women's EuroLeague Basketball Streaming and Coverage. Διαθέσιμο στο:
<https://www.dazn.com/>

Παράρτημα

Κώδικας

Παρακάτω παρουσιάζεται ο κώδικας που χρησιμοποιήθηκε για τη στατιστική επεξεργασία και ανάλυση των δεδομένων. Ο κώδικας υλοποιήθηκε στο στατιστικό λογισμικό R και περιλαμβάνει τις διαδικασίες εισαγωγής και επεξεργασίας των δεδομένων, την παραγωγή γραφημάτων και περιγραφικών μέτρων, τους στατιστικούς ελέγχους, τα μοντέλα λογιστικής και πολυωνυμικής λογιστικής παλινδρόμησης, καθώς και τις μεθόδους ανάλυσης κατά συστάδες που εφαρμόστηκαν στην παρούσα έρευνα. Η παράθεση του κώδικα πραγματοποιείται με στόχο τη διαφάνεια της μεθοδολογίας και τη δυνατότητα αναπαραγωγής των αποτελεσμάτων της ανάλυσης.

```
# ΚΕΦΑΛΑΙΟ 1 – ΕΙΣΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ#
install.packages("readxl")
install.packages("dplyr")
library(readxl)
library(dplyr)
data<-read_xlsx("C:/Users/User/Desktop/data.xlsx")
attach(data)
class(data)
sum(is.na(data))
# ΑΦΑΙΡΕΣΗ ΤΗΣ SEASON 17#
data <- data %>%
  filter(SEASON != 17)
# ΚΕΦΑΛΑΙΟ 3 – ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ#
# ΠΕΡΙΓΡΑΦΙΚΑ ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ#
cols <- c("PPG", "2P%", "3P%", "FT%",
          "ORPG", "DRPG", "RPG",
          "APG", "STLPG", "TOPG",
          "BLKPG", "PFPG", "EFFPG")
data[, cols] <- lapply(data[, cols], function(x) {
  x <- gsub(",", ".", x)
  x <- gsub("%", "", x)
  as.numeric(x)
})
names(data)
vars <- data[, cols]
descriptive_table <- data.frame(
  Mean   = sapply(vars, mean, na.rm = TRUE),
  SD     = sapply(vars, sd, na.rm = TRUE),
  Median = sapply(vars, median, na.rm = TRUE),
  Q1     = sapply(vars, quantile, probs = 0.25, na.rm = TRUE),
  Q3     = sapply(vars, quantile, probs = 0.75, na.rm = TRUE),
  Min    = sapply(vars, min, na.rm = TRUE),
  Max    = sapply(vars, max, na.rm = TRUE)
)
descriptive_table <- round(descriptive_table, 2)
descriptive_table
```

```

# TIME SERIES PLOTS#
install.packages("reshape2")
library(ggplot2)
library(reshape2)
par(mfrow=c(5,3), mar=c(3,3,2,1))
plot(SEASON, PPG, type="o", col="black", lwd=2,
      xlab="SEASON", ylab="PPG", main="PPG")
# SCATTER PLOTS#
par(mfrow=c(4,3), mar=c(4,4,2,1))
plot(PPG, EFFPG, main="EFFPG vs PPG")
# ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ#
cor_eff <- cor(data[, vars], use = "complete.obs")["EFFPG"]
round(cor_eff, 3)
# CHERNOFF FACES#
install.packages("aplpack")
library(aplpack)
season_data <- data %>%
  group_by(SEASON) %>%
  summarise(across(all_of(cols), mean, na.rm = TRUE))
data_scaled <- scale(season_data[, -1])
par(mfrow = c(5,5), mar = c(1,1,2,1))
faces(data_scaled,
      labels = season_data$SEASON,
      main = "Chernoff Faces per Season")
# ΧΩΡΙΣΜΟΣ ΟΜΑΔΩΝ ΑΝΑ ΠΡΟΚΡΙΣΗ#
data$SUCCESS <- ifelse(data$`FINAL 4` == 1, "Final4",
  ifelse(data$`QUARTER FINAL` == 1, "Final8", "No"))
data$SUCCESS <- factor(data$SUCCESS,
  levels = c("No", "Final8", "Final4"))
#ΠΕΡΙΓΡΑΦΙΚΑ ΜΕΤΡΑ #
data[vars] <- lapply(data[vars], function(x) as.numeric(as.character(x)))
create_table <- function(df) {
  data.frame(
    Variable = vars,
    Mean = round(sapply(df[vars], mean, na.rm=TRUE), 2),
    SD = round(sapply(df[vars], sd, na.rm=TRUE), 2),
    Median = round(sapply(df[vars], median, na.rm=TRUE), 2),
    Q1 = round(sapply(df[vars], quantile, probs=0.25, na.rm=TRUE), 2),
    Q3 = round(sapply(df[vars], quantile, probs=0.75, na.rm=TRUE), 2),
    Min = round(sapply(df[vars], min, na.rm=TRUE), 2),
    Max = round(sapply(df[vars], max, na.rm=TRUE), 2)
  )
}
table_No <- create_table(filter(data, SUCCESS == "No"))
table_Final8 <- create_table(filter(data, SUCCESS == "Final8"))
table_Final4 <- create_table(filter(data, SUCCESS == "Final4"))
# BOXPLOTS & LINE PLOTS#
library(ggplot2)
library(tidyr)
data_long <- data %>%

```

```

select(all_of(vars), SUCCESS) %>%
pivot_longer(cols = all_of(vars),
              names_to = "Variable",
              values_to = "Value")
ggplot(data_long, aes(x = SUCCESS, y = Value, fill = SUCCESS)) +
  geom_boxplot(alpha = 0.8, outlier.shape = 16, outlier.size = 1.5) +
  facet_wrap(~Variable, scales = "free", ncol = 3) +
  scale_fill_manual(values = c(
    "No" = "#D55E00",      # κόκκινο-πορτοκαλί
    "Final8" = "#0072B2",  # μπλε
    "Final4" = "#009E73"   # πράσινο
  )) +
  theme_minimal(base_size = 12) +
  theme(
    legend.position = "bottom",
    strip.text = element_text(face = "bold"),
    plot.title = element_text(face = "bold", size = 14)
  ) +
  labs(title = "Σύγκριση Στατιστικών Δεικτών ανά Επίπεδο Επιτυχίας",
       x = "Επίπεδο Επιτυχίας",
       y = "Τιμή Μεταβλητής",
       fill = "Κατηγορία")
means_by_success <- data %>%
  group_by(SUCCESS) %>%
  summarise(across(all_of(vars), mean, na.rm = TRUE))
means_long <- means_by_success %>%
  pivot_longer(-SUCCESS,
               names_to = "Variable",
               values_to = "Mean")
ggplot(means_long, aes(x = SUCCESS, y = Mean,
                      group = Variable, color = Variable)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  facet_wrap(~Variable, scales = "free", ncol = 3) +
  theme_minimal(base_size = 13) +
  theme(
    strip.text = element_text(face = "bold"),
    plot.title = element_text(face = "bold")
  ) +
  labs(title = "Μέσοι Όροι Μεταβλητών ανά Επίπεδο Επιτυχίας",
       x = "Επίπεδο Επιτυχίας",
       y = "Μέσος Όρος")
# ΚΕΦΑΛΑΙΟ 4 – ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ#
# ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ#
data$Final8_Status <- ifelse(data$`QUARTER FINAL` == 1, "Yes", "No")
data$Final8_Status <- factor(data$Final8_Status, levels = c("No", "Yes"))
normality_table <- data.frame(
  Variable = vars,
  Pvalue_No = NA,
  Pvalue_Yes = NA,

```

```

Conclusion = NA
)
for(i in seq_along(vars)){var <- vars[i]
x_no <- na.omit(data[data$Final8_Status=="No", var][[1]])
x_yes <- na.omit(data[data$Final8_Status=="Yes", var][[1]])
p_no <- shapiro.test(x_no)$p.value
p_yes <- shapiro.test(x_yes)$p.value
normality_table$Pvalue_No[i] <- round(p_no,4)
normality_table$Pvalue_Yes[i] <- round(p_yes,4)
if(p_no > 0.05 & p_yes > 0.05){
  normality_table$Conclusion[i] <- "Κανονική"
} else {
  normality_table$Conclusion[i] <- "Μη κανονική"
}}
# SHAPIRO-WILK TEST#
shapiro_result <- shapiro.test(var_data)
W_val <- shapiro_result$statistic
p_val <- shapiro_result$p.value
cat("Μεταβλητή:", var, "\n")
cat("Shapiro-Wilk W =", round(W_val, 4), ", p-value =", round(p_val, 4), "\n")
if (p_val < 0.05) {
  cat("-> Δεν ακολουθεί κανονική κατανομή\n\n")
  normality_results$Normality[i] <- "Όχι"
} else {
  cat("-> Ακολουθεί κανονική κατανομή\n\n")
  normality_results$Normality[i] <- "Ναι"
}
normality_results$Shapiro_W[i] <- W_val
normality_results$Shapiro_pvalue[i] <- p_val
}
print(normality_results)
# ΕΛΕΓΧΟΣ ΔΙΑΚΥΜΑΝΣΕΩΝ#
install.packages("car")
library(car)
data$Final8_Status <- ifelse(data$`QUARTER FINAL` == 1, "Yes", "No")
data$Final8_Status <- factor(data$Final8_Status, levels = c("No", "Yes"))
variance_results <- data.frame(
  Variable = vars,
  Levene_F = NA_real_,
  Levene_pvalue = NA_real_,
  Equal_Variance = NA_character_,
  stringsAsFactors = FALSE
)
# ΕΛΕΓΧΟΣ ΙΣΟΤΗΤΑΣ ΔΙΑΚΥΜΑΝΣΕΩΝ (LEVENE) ΓΙΑ ΚΑΘΕ
ΟΜΑΔΑ#
levene_res <- leveneTest(data[[var]] ~ data$Final8_Status)
F_val <- levene_res$`F value`[1]
p_val <- levene_res$`Pr(>F)`[1]
variance_results$Levene_F[i] <- F_val
variance_results$Levene_pvalue[i] <- p_val

```

```

variance_results$Equal_Variance[i] <- ifelse(p_val < 0.05, "Οχι", "Ναι")
cat("Μεταβλητή:", var, "\n")
cat("Levene F =", round(F_val,4), ", p-value =", round(p_val,4), "\n")
cat("-> Σταθερή διακύμανση μεταξύ ομάδων:",
variance_results$Equal_Variance[i], "\n\n")
}
cat("===== ΣΥΝΟΛΙΚΟΣ ΠΙΝΑΚΑΣ LEVENE =====\n")
print(variance_results)

# T-TEST / MANN WHITNEY#
test_results <- data.frame(
  Variable = vars,
  Test = NA_character_,
  p_value = NA_real_,
  Significant = NA_character_,
  stringsAsFactors = FALSE
)
for(i in seq_along(vars)){
  var <- vars[i]
  normal <- normality_results$Normality[i]
  equal_var <- variance_results$Equal_Variance[i]
  cat("Μεταβλητή:", var, "\n")
  if(normal == "Ναι"){
    if(equal_var == "Ναι"){
      t_res <- t.test(data[[var]] ~ data$Final8_Status, var.equal = TRUE)
      test_type <- "t-test (equal var)"
    } else {
      t_res <- t.test(data[[var]] ~ data$Final8_Status, var.equal = FALSE)
      test_type <- "t-test (Welch)"
    }
    p_val <- t_res$p.value
    print(t_res)
  } else {
    mw_res <- wilcox.test(data[[var]] ~ data$Final8_Status)
    p_val <- mw_res$p.value
    test_type <- "Mann-Whitney"
    print(mw_res)
  }
  test_results$Test[i] <- test_type
  test_results$p_value[i] <- p_val
  test_results$Significant[i] <- ifelse(p_val < 0.05, "Ναι", "Οχι")
  cat("-----\n\n")
}

# VIOLIN PLOTS & EFFECT SIZE#
install.packages(c("ggplot2", "effsize"))
library(ggplot2)
library(effsize)
get_effect <- function(var, data){
  df <- data[!is.na(data[[var]]) & !is.na(data$Final8_Status), ]
  group_numeric <- ifelse(df$Final8_Status == "Yes", 1, 0)

```

```

# POINT-BISERIAL (PEARSON ME ΔΥΑΔΙΚΗ ΜΕΤΑΒΛΗΤΗ)
r_pb <- cor(df[[var]], group_numeric, method = "pearson")
return(paste0("r_pb = ", round(r_pb, 3), " (Point-Biserial)"))
}
effects <- sapply(vars, function(v) get_effect(v, data))
results <- data.frame(
  Variable = vars,
  Point_Biserial_r = effects
)

# PPG
effect_PPG <- get_effect("PPG", data)
p_PPG <- ggplot(data, aes(x = Final8_Status, y = .data[["PPG"]], fill =
Final8_Status)) +
  geom_violin(trim = FALSE, alpha = 0.5) +
  geom_boxplot(width = 0.2, outlier.shape = NA) +
  geom_jitter(width = 0.15, alpha = 0.6) +
  labs(title = "Διάγραμμα για: PPG", x = "Ομάδα", y = "PPG") +
  theme_minimal() + theme(legend.position = "none") +
  expand_limits(y = max(data$PPG, na.rm = TRUE) * 1.15)
print(p_PPG)

# ΚΕΦΑΛΑΙΟ 5 – ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ#
# ΔΥΑΔΙΚΗ ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ#
model <- glm(Final8_Status ~ PPG + `2P%` + `3P%` + `FT%` +
  ORPG + DRPG + RPG + APG + STLPG +
  TOPG + BLKPG + PFPG + EFFPG,
  data = data,
  family = binomial(link=logit))
summary(model)
vif(model)
library(MASS)
model_full <- glm(Final8_Status ~ PPG + `2P%` + `3P%` + `FT%` +
  ORPG + DRPG + RPG + APG + STLPG +
  TOPG + BLKPG + PFPG + EFFPG,
  data = data,
  family = binomial)
model_step <- stepAIC(model_full, direction = "both")
summary(model_step)
vif(model_step)
model2 <- glm(Final8_Status ~ `2P%` + `3P%` + `FT%`
  + APG + STLPG +
  TOPG + BLKPG,
  data = data,
  family = binomial)
summary(model2)
# ΛΟΓΟΙ ΠΙΘΑΝΟΤΗΤΩΝ (Odds Ratios)#
exp(coef(model2))
# ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ#
exp(confint(model2))

```

```

vif(model2)
# ROC CURVE ANALYSIS#
prob_final <- predict(model2, type = "response")
roc_final <- roc(data$Final8_Status, prob_final, levels = c("No", "Yes"))
plot(roc_final,
     col = "red",
     lwd = 2,
     main = "ROC Curve - Τελικό Μοντέλο")
# AUC#
auc_final <- auc(roc_final)
# NAGELKERKE#
install.packages("rcompanion")
library(rcompanion)
nagelkerke(model2)
# ΚΑΤΑΛΟΙΠΑ ΑΠΟΚΛΙΣΗΣ#
res_dev <- residuals(model2, type = "deviance")
plot(res_dev,
     ylab = "Deviance residuals",
     main = "Κατάλοιπα απόκλισης")

abline(h = 0, col = "red")
# HOSMER – LEMESHOW#
install.packages("ResourceSelection", dependencies = TRUE) # μόνο 1 φορά
library(ResourceSelection)
y <- as.numeric(data$Final8_Status) - 1
hoslem.test(y, fitted(model2), g = 10)

# ΠΟΛΥΩΝΥΜΙΚΟ ΜΟΝΤΕΛΟ#
#ΧΩΡΙΣΜΟΣ ΟΜΑΔΩΝ ΑΝΑ ΠΡΟΚΡΙΣΗ#
data$SUCCESS <- ifelse(data$`FINAL 4` == 1, "Final4",
                      ifelse(data$`QUARTER FINAL` == 1, "Final8", "No"))
data$SUCCESS <- factor(data$SUCCESS,
                      levels = c("No", "Final8", "Final4"))

install.packages("mlogit")
library(mlogit)

data$id <- 1:nrow(data)
data$SUCCESS <- relevel(data$SUCCESS, ref = "No")
data_mlogit <- mlogit.data(data,
                          choice = "SUCCESS",
                          shape = "wide")
model2 <- mlogit(
  SUCCESS ~ 1 | PPG + FT. + RPG
  + APG +
  TOPG+ BLKPG,
  data = data_mlogit, reflevel = "No"
)
summary(model2)
AIC(model2)

```

```

exp(coef(model2))
prob <- predict(model2, newdata = data_mlogit, type = "probabilities")
head(prob)
nrow(prob)
pred <- colnames(prob)[apply(prob, 1, which.max)]
actual <- data$SUCCESS
length(pred)
length(actual)
pred <- factor(pred, levels = c("No", "Final8", "Final4"))
actual <- factor(actual, levels = c("No", "Final8", "Final4"))
table(Predicted = pred, Actual = actual)
mean(pred == actual)
# ROC ΓΙΑ ΚΑΘΕ ΚΑΤΗΓΟΡΙΑ#
roc_final4 <- roc(actual == "Final4", prob[, "Final4"])
roc_final8 <- roc(actual == "Final8", prob[, "Final8"])
roc_no <- roc(actual == "No", prob[, "No"])
# ΔΙΑΓΡΑΜΜΑΤΑ#
plot(roc_final4, col = "blue", lwd = 2,
     main = "ROC Curves (Multiclass - One vs Rest)")
plot(roc_final8, col = "green", lwd = 2, add = TRUE)
plot(roc_no, col = "red", lwd = 2, add = TRUE)
legend("bottomright",
     legend = c("Final4", "Final8", "No"),
     col = c("blue", "green", "red"),
     lwd = 2)
roc_final4 <- roc(actual == "Final4", prob[, "Final4"])
auc(roc_final4)
roc_final8 <- roc(actual == "Final8", prob[, "Final8"])
auc(roc_final8)
roc_no <- roc(actual == "No", prob[, "No"])
auc(roc_no)

#ΚΕΦΑΛΑΙΟ 6-ΑΝΑΛΥΣΗ ΚΑΤΑ ΣΥΣΤΑΔΕΣ #
# SCALING#
data_scaled <- scale(data_sub)
dist_matrix <- dist(data_scaled)

#ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ#
dist_matrix <- dist(data_scaled, method = "euclidean")
hc_single <- hclust(dist_matrix, method = "single")

# WARD#
hc_ward <- hclust(dist_matrix,
     method = "ward.D2")

#ΔΕΝΔΡΟΔΙΑΓΡΑΜΜΑ#
plot(hc_ward,
     labels = FALSE,
     main = "Ward Dendrogram")
clusters_ward <- cutree(hc_ward,

```

```

        k = 2)
# ΠΙΝΑΚΑΣ ΣΥΧΝΟΤΗΤΩΝ#
table(clusters_ward,
      data$Final8_Status)
#ΜΗ ΙΕΡΑΡΧΙΚΗ#
# K-MEANS#
kmeans_res <- kmeans(data_scaled, centers = 2, nstart = 25)
data$Cluster_kmeans <- kmeans_res$cluster
table(data$Cluster_kmeans, data$Final8_Status)
aggregate(data[, vars],
          by = list(Cluster = data$Cluster_kmeans),
          mean)
tab <- table(data$Cluster_kmeans, data$Final8_Status)
best_cluster <- which.max(tab[, "Yes"])
data$Predicted <- ifelse(data$Cluster_kmeans == best_cluster, "Yes", "No")
# Confusion matrix#
table(Predicted = data$Predicted,
      Actual = data$Final8_Status)
# ΑΚΡΙΒΕΙΑ#
mean(data$Predicted == data$Final8_Status)
#PCA ΓΙΑ ΟΠΤΙΚΟΠΟΙΗΣΗ#
pca_res <- prcomp(data_scaled,
                  center = TRUE,
                  scale. = TRUE)
# ΠΟΣΟΣΤΟ ΕΞΗΓΟΥΜΕΝΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ#
var_explained <- summary(pca_res)$importance[2, ] * 100
pca_df <- data.frame(
  PC1 = pca_res$x[,1],
  PC2 = pca_res$x[,2],
  Cluster = as.factor(data$Cluster_kmeans)
)
# Scatter Plot#
plot(pca_df$PC1,
     pca_df$PC2,
     col = ifelse(pca_df$Cluster == 1,
                  "blue",
                  "red"),
     pch = 19,
     xlab = paste0("1η Κύρια Συνιστώσα (",
                   round(var_explained[1], 2),
                   "%)"),
     ylab = paste0("2η Κύρια Συνιστώσα (",
                   round(var_explained[2], 2),
                   "%)"),
     main = "Διάγραμμα διασποράς των clusters μέσω PCA")
legend("topright",
      legend = c("Cluster 1",
                  "Cluster 2"), col = c("blue",
                  "red"), pch = 19)

```

Ονόματα ομάδων

Από το 2004 έως το 2025 έχουν συμμετάσχει έστω και μια φορά 108 διαφορετικές ομάδες στην διοργάνωση . Παρακάτω στον πίνακα 3.1 παρουσιάζονται αναλυτικά τα ονόματα των ομάδων μαζί με τον κωδικό τους.

TEAM	TEAM ID
ACS Sepsi-SIC	1
AGÜ Spor	3
Arka Gdynia	5
Arras	6
Atomeromu KSC Szekszard	7
Baltiyskaya Zvezda	8
Basket Landes	9
BC Polkowice	12
BC Volgaburmash	11
Beretta Famila Schio	13
Besiktas Cola Turka	17
BK IMOS Brno	19
BK Loko	20
BLMA	21
Bourges Basket	23
Carolo Basket	24
Casademont Zaragoza	25
Castors Braine	26
CB Rivas	27
CB Taranto	28
CCC Polkowice	30
CIMSA CBK Mersin	29
CMB Cargo Uni Györ	31
Cras Basket Taranto	32
CSKA Moscow	33
Cukurova Basketbol Mersin	35
CUS Chieti	36
Delta ICP	37
Dexia Namur	38
DVTK HUNTHERM	40
Dynamo Kursk	42
Energa Torun	44
ESBVA-LM	45
Famila Schio	14
Fenerbahce Opet	50
Frisco Brno	54
Galatasaray	56

Gelecek Koleji Cukurova	59
Germano Zama Faenza	60
Good Angels Kosice	61
Gorzow	62
Gospic CO	63
Halcon Avenida	65
Hatay BB	66
Izmit Belediyespor	67
Jolly JBS	68
K Cero ICP	69
Kangoeroes Mechelen	70
Kayseri Kaski spor	71
KGHM BC Polkowice	72
Kibirkstis-Vici	73
Kosit 2013	75
Kozachka - Zalk	76
KSC Szekszard	77
Lattes Montpellier	78
Lavezzini Parma	79
LDLC ASVEL Feminin	80
Lietuvos Telekomas	81
London Lions	82
MAV Coop SE	86
Maxima Broker Kosice	87
MBA Moscow	95
MBK	89
MKB Euroleasing	47
Nadezhda	96
Olympiacos	97
Pays D'Aix Basket 13	99
Pécs 2010	91
Perfumerias Avenida	101
Phard Napoli	103
Reyer Venezia	105
Rivas Ecopolis	106
Ros Casares	108
Serco UNI Gyor	109
Sopron Basket	110
Spar Girona	111
Sparta&K Moscow Region	113
Sporting Athens	115
Szeviép	116
Tango Bourges Basket	117
Tarbes GB	119

Tarsus	120
TEO Vilnius	121
TTT Riga	123
UE Sopron	124
Umana Reyer Venezia	125
UMMC Ekaterinburg	127
Uni Gyor	128
Union Hainaut Basket	130
Uniqs Sopron	131
Universitat FC Barcelona	132
USK Pre Praha	134
USO Mondeville	136
US Valenciennes Olympic	133
Valencia Basket Club	138
VBM-SGAU	140
VBW Arka Gdynia	83
VICI Aistes	142
Villeneuve d'Ascq LM	143
Virtus Segafredo Bologna	144
Vojvodina Nis-Gas	145
WBC Dynamo	146
WBC Novi Zagreb	147
Wisla CANPACK	122
Yakin Dogu Universitesi	151
Zabiny Brno	57
ZVVZ USK Praha	152

Πίνακας 3.1 όνομα και id των ομάδων.