

**ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΠΡΟΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΜΑΙΟΣ 2024

‘ΣΧΟΛΙΚΟ ΠΡΩΤΑΘΛΗΜΑ’

Γουέρα Ευαγγελία-Ζωή

Εισαγωγή στις Πιθανότητες και Στατιστική με R

Διδάσκοντες: Δ.Καρλής, Β.Χασίωτης

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ.....	3
1.1. Αποτελέσματα	5
2. ΠΕΡΙΠΤΩΣΗ ΑΤΥΧΙΑΣ.....	8
2.1. Αποτελέσματα	10

1. ΕΙΣΑΓΩΓΗ

Στο πλαίσιο αυτής της εργασίας εξετάζουμε το πρωτάθλημα καλαθοσφαίρισης 16 σχολείων από όλη τη χώρα. Οι 16 ομάδες θα παίζουν σε αγώνες νοκ-άουτ μέχρι να αναδειχτεί ο πρωταθλητής. Για λόγους ευκολίας, κατά την διάρκεια του παιχνιδιού η κάθε ομάδα έχει έναν κωδικό A1-A16. Κάθε ομάδα επίσης παίζει με ένα συγκεκριμένο πλάνο αναμετρήσεων και ο αριθμός των πόντων που πετυχαίνουν σε έναν αγώνα κανονικής διάρκειας θα είναι το ακέραιο μέρος μιας τυχαίας μεταβλητής που ακολουθεί κανονική κατανομή με καθορισμένες παραμέτρους. Σε κάθε αγώνα κερδίζει η ομάδα με τους περισσότερους πόντους και σε περίπτωση ισοπαλίας κερδίζει η πιο τυχερή ομάδα.

Θέλοντας να υπολογίσουμε για κάθε ομάδα πόσο πιθανό είναι να φτάσει στον τελικό και να κερδίσει τον τίτλο, ξεκινήσαμε την προσομοίωση του παιχνιδιού θέλοντας να κατασκευάσουμε 2 συναρτήσεις όπου η μία θα επιστρέφει τις ομάδες που έφτασαν στον τελικό και η άλλη, με βάση τα αποτελέσματα της προηγούμενης, την ομάδα που κέρδισε.

Για την πρώτη συνάρτηση:

Η ιδέα είναι η συνάρτηση να χωρίζεται σε τρεις γύρους. Στον πρώτο γύρο θα είναι η πρώτη αναμέτρηση όπου η κάθε ομάδα θα παίζει με την αντίστοιχη ομάδα που έχει δοθεί στο πλάνο. Για αυτόν τον λόγο, φτιάξαμε μια λίστα που θα περιλαμβάνει τα ζεύγη των ομάδων για την πρώτη αναμέτρηση. Επίσης, αποθηκεύσαμε σε ένα διάνυσμα όλες τις μέσες τιμές για τους πόντους των ομάδων.

Ξεκινώντας από τον πρώτο γύρο, θέλουμε να διαλέγουμε το κάθε ένα ζεύγος ομάδων από την λίστα και κάθε φορά οι δύο ομάδες να παίζουν φέρνοντας τυχαία ένα σκορ που θα αντιστοιχεί στην μέση τιμή τους. Για αυτόν τον λόγο, φτιάξαμε μια επαναληπτική διαδικασία μέσα στην οποία διαλέγουμε κάθε φορά την ομάδα 1 και την ομάδα 2 από τα στοιχεία της λίστας, δηλαδή με βάση το πρώτο και το δεύτερο στοιχείο των ζευγών από την λίστα. Για να ορίσουμε τα σκορ θα πρέπει για την κάθε ομάδα να πάρουμε την αντίστοιχη μέση τιμή. Συνεπώς, φτιάξαμε δύο διανύσματα όπου για την ομάδα 1 και για την ομάδα 2 θα μας δίνουν την μέση τιμή από το αρχικό διάνυσμα μέσων τιμών που είχαμε ορίσει. Δηλαδή, αν οι ομάδες που παίζουν είναι οι A1 και A16 τότε από το διάνυσμα μέσων τιμών θέλουμε να πάρουμε την πρώτη και την δέκατη έκτη τιμή για τα σκορ που θα ορίσουμε. Στην συνέχεια, τα σκορ θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή των ομάδων και τυπική απόκλιση κ , όπου κ θα είναι η μεταβλητή εισόδου της συνάρτησης.

Συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε κερδίζει η ομάδα 1 κρατώντας σε ένα κενό διάνυσμα(winners1) την ομάδα 1. Αλλιώς κερδίζει η ομάδα 2 κρατώντας στο κενό διάνυσμα την ομάδα 2.

Συνεχίζουμε στην δεύτερη αναμέτρηση για τις ομάδες που κέρδισαν στον πρώτο γύρο ακολουθώντας τα ίδια βήματα..

Με βάση το πλάνο των αγώνων, αποθηκεύουμε σε μια λίστα τους νικητές του προηγούμενου γύρου ανά ζεύγη. Μέσα από την επαναληπτική διαδικασία, διαλέγουμε την ομάδα 1 και την ομάδα 2 που θα χρειαστεί να αναμετρηθούν από τα στοιχεία της νέας λίστας. Φτιάχνουμε δύο διανύσματα όπου για την ομάδα 1 και για την ομάδα 2 θα δίνουν την μέση τιμή από το αρχικό διάνυσμα μέσω των τιμών που είχαμε ορίσει και δηλώνουμε τα σκορ τα οποία θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή των ομάδων και τυπική απόκλιση κ .

Στην συνέχεια συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε κερδίζει η ομάδα 1 κρατώντας σε ένα κενό διάνυσμα(winners2) την ομάδα 1. Αλλιώς κερδίζει η ομάδα 2 κρατώντας στο κενό διάνυσμα την ομάδα 2.

Η ίδια ακριβώς λογική εφαρμόζεται και στην τρίτη αναμέτρηση για τις ομάδες που κέρδισαν στον δεύτερο γύρο, φτιάχνοντας μια επαναληπτική διαδικασία όπου στο τέλος θα αποθηκεύει κάθε φορά σε ένα κενό διάνυσμα(winners3) ή την ομάδα 1 ή την ομάδα 2. Η συνάρτηση κλείνει ζητώντας να επιστρέφει το διάνυσμα με τις ομάδες που έχουν αποθηκευτεί στον τρίτο γύρο, δηλαδή τις ομάδες που φτάνουν στον τελικό.

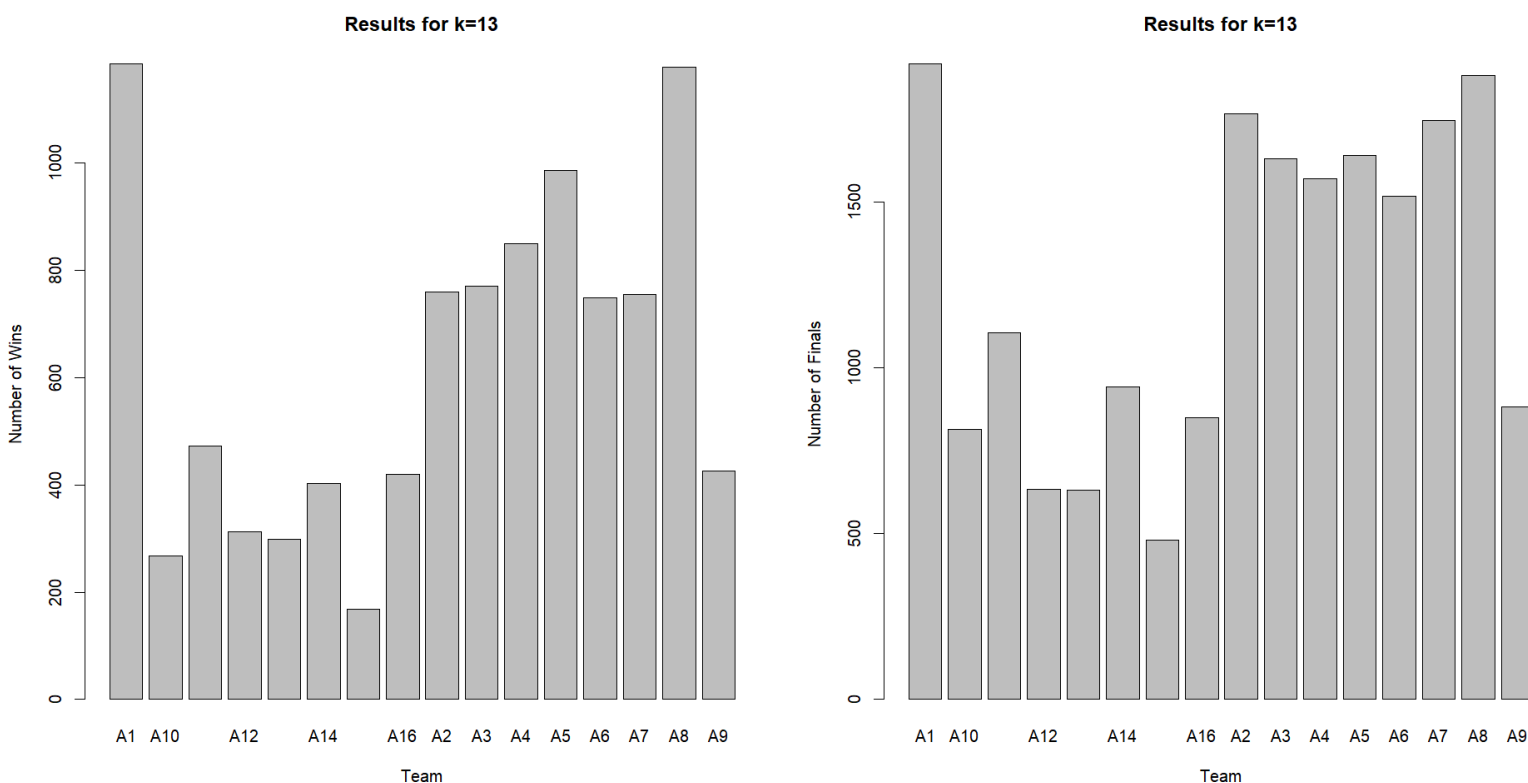
Για την δεύτερη συνάρτηση:

Η συνάρτηση θα έχει επίσης ως μεταβλητή εισόδου την τυπική απόκλιση κ και η διαδικασία θα είναι η ίδια με τους προηγούμενους γύρους. Η διαφορά είναι ότι σε αυτήν την συνάρτηση δεν χρειάζεται να αποθηκεύσουμε τις ομάδες του τελικού σε μια λίστα. Θα χρειαστεί μόνο να ξανά δώσουμε το αρχικό διάνυσμα μέσω των τιμών για όλες τις ομάδες και να δηλώσουμε τις ομάδες του τελικού ως το αποτέλεσμα της πρώτης συνάρτησης. Ως ομάδα 1 ορίζουμε το πρώτο στοιχείο του διανύσματος που επιστρέφει η πρώτη συνάρτηση και ως ομάδα 2 το δεύτερο στοιχείο. Παίρνουμε τις μέσες τιμές των ομάδων και ορίζουμε με αυτές τα αντίστοιχα σκορ τους όπως και στους προηγούμενους γύρους. Συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε η συνάρτηση θα επιστρέφει την ομάδα 1 αλλιώς θα επιστρέφει την ομάδα 2.

1.1. Αποτελέσματα

Για τυπική απόκλιση $\kappa=13$:

Προσομοιώνοντας και τις 2 συναρτήσεις για 10000 φορές και παίρνοντας τις συχνότητες των αποτελεσμάτων παρατηρούμε ότι η ομάδα που εμφανίζεται περισσότερες φορές νικήτρια είναι η ομάδα A1 η οποία θα φτάνει και τις περισσότερες φορές στον τελικό. Η αμέσως επόμενη ομάδα με την μεγαλύτερη συχνότητα στον τελικό και στην νίκη είναι η A8. Από τα παρακάτω διαγράμματα συχνοτήτων παρατηρούμε αντίστοιχα ότι η A1 και η A8 έχουν πολύ μικρή διαφορά μεταξύ τους αλλά η A1 θα έχει την μεγαλύτερη συχνότητα εμφάνισης στον τελικό και στην νίκη.



Έπειτα, διαιρώντας τις συχνότητες με το άθροισμα τους, παίρνουμε την πιθανότητα για την κάθε μια ομάδα να φτάσει στον τελικό και να νικήσει. Παρατηρούμε ότι η ομάδα με την μεγαλύτερη πιθανότητα να φτάσει στον τελικό και να νικήσει είναι η A1 με πιθανότητες περίπου 0,19 και 0,11 αντίστοιχα.

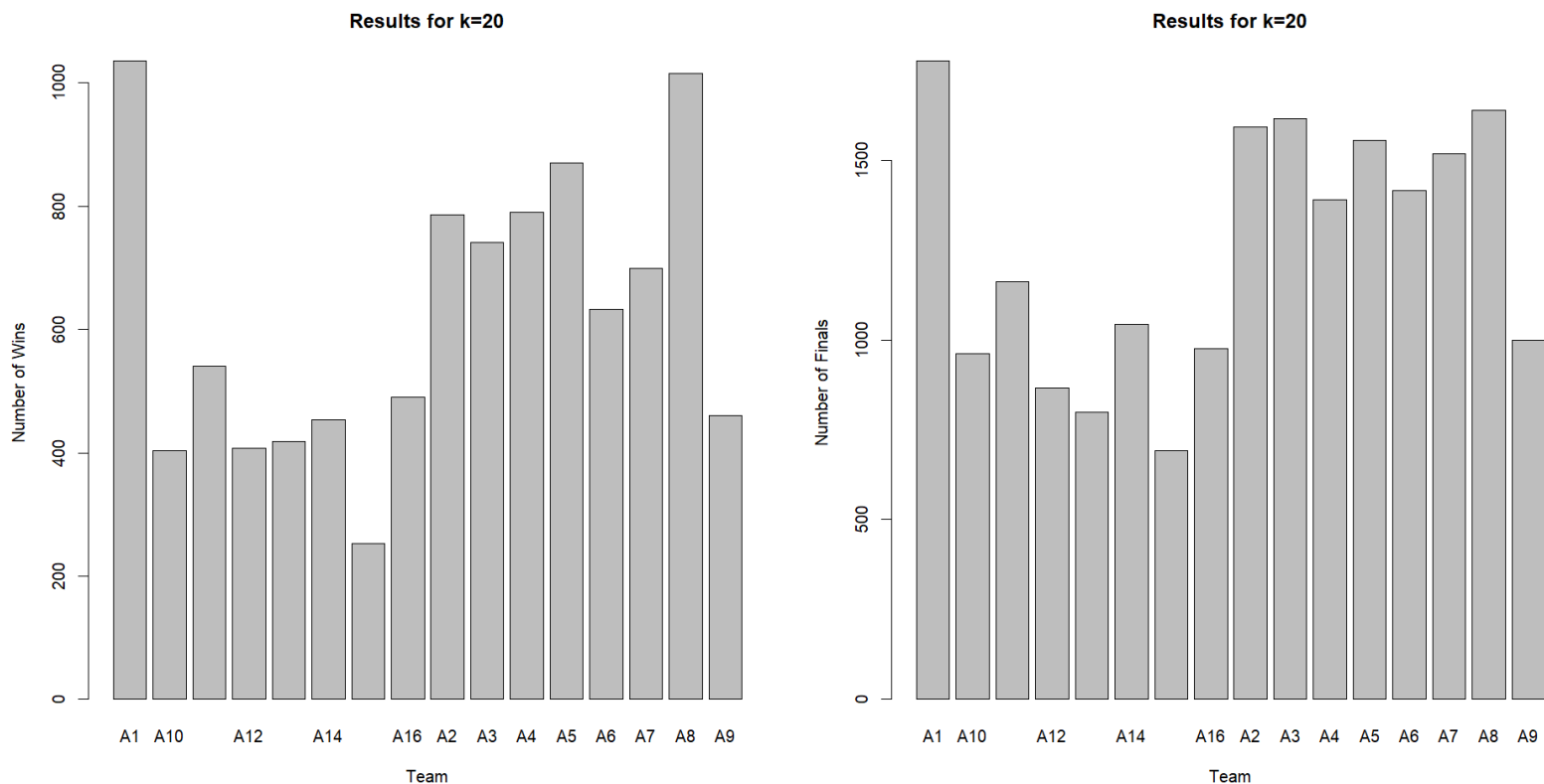
Ομάδες	Τελικός	Νικητής
A1	0,195	0,119
A10	0,079	0,027
A11	0,112	0,046
A12	0,067	0,03
A13	0,065	0,03
A14	0,092	0,038
A15	0,049	0,014
A16	0,084	0,042
A2	0,174	0,082
A3	0,173	0,075
A4	0,149	0,089
A5	0,166	0,099
A6	0,151	0,071
A7	0,166	0,073
A8	0,185	0,112
A9	0,086	0,046
Άθροισμα	2	1

Πίνακας πιθανοτήτων για τελικό και νίκη με $\kappa=13$

Για τυπική απόκλιση $\kappa=20$:

Προσομοιώνοντας και τις 2 συναρτήσεις για 10000 φορές και παίρνοντας τις συχνότητες των αποτελεσμάτων παρατηρούμε ότι η ομάδα που εμφανίζεται περισσότερες φορές στον τελικό και στο τέλος νικήτρια είναι ξανά η ομάδα A1 αλλά με μικρότερη συχνότητα από αυτή που είχε για $\kappa=13$. Η δεύτερη ομάδα με την μεγαλύτερη εμφάνιση στον τελικό και στην νίκη θα είναι η A8 με αντίστοιχα μικρότερη συχνότητα από αυτή που είχε για $\kappa=13$. Τα ίδια ακριβώς συμπεραίνουμε και από τα παρακάτω διαγράμματα συχνοτήτων.

Έπειτα, διαιρώντας τις συχνότητες με το άθροισμα τους, παίρνουμε την πιθανότητα για την κάθε μια ομάδα να φτάσει στον τελικό και να νικήσει. Παρατηρούμε ότι η ομάδα με την μεγαλύτερη πιθανότητα να φτάσει στον τελικό και να νικήσει είναι η A1 με πιθανότητες περίπου 0,17 και 0,10 αντίστοιχα.



Ομάδες	Τελικός	Νικητής
A1	0,178	0,103
A10	0,092	0,04
A11	0,115	0,055
A12	0,086	0,04
A13	0,08	0,042
A14	0,104	0,045
A15	0,067	0,025
A16	0,096	0,049
A2	0,162	0,078
A3	0,158	0,074
A4	0,139	0,078
A5	0,155	0,086
A6	0,144	0,062
A7	0,154	0,069
A8	0,163	0,1
A9	0,099	0,046
Άθροισμα	2	1

Πίνακας πιθανοτήτων για τελικό και νίκη με κ=20

2. ΠΕΡΙΠΤΩΣΗ ΑΤΥΧΙΑΣ

Στην περίπτωση όπου οι ομάδες A1 και A2 έχουν την ατυχία ο καλύτερος τους παίκτης να τραυματιστεί και να μην παίξει στους πρώτους δυο αγώνες, η ικανότητα των ομάδων θα μειωθεί κατά 5 πόντους κατά μέσο όρο.

Στην περίπτωση αυτή θα χρειαστεί να ξανά φτιάξουμε την πρώτη συνάρτηση που περιλαμβάνει τους τρεις πρώτους γύρους μέχρι τον τελικό προσθέτοντας κάποιες προϋποθέσεις στους πρώτους δυο γύρους για τις ομάδες A1 και A2.

Ξεκινάμε την συνάρτηση δίνοντας την λίστα με τα ζεύγη των ομάδων για τον πρώτο γύρο αγώνων και το διάνυσμα με όλες τις μέσες τιμές των ομάδων. Φτιάχνουμε μία επαναληπτική διαδικασία μέσα στην οποία διαλέγουμε κάθε φορά την ομάδα 1 και την ομάδα 2 από τα στοιχεία της λίστας. Για να ορίσουμε τα σκορ θα πρέπει για την κάθε ομάδα να πάρουμε την αντίστοιχη μέση τιμή. Για αυτόν τον λόγο φτιάχνουμε δύο διανύσματα όπου για την ομάδα 1 και για την ομάδα 2 θα μας δίνουν την μέση τιμή από το αρχικό διάνυσμα μέσων τιμών. Με βάση το πως έχουμε ορίσει τα ζευγάρια που παίζουν μεταξύ τους στον πρώτο γύρο μέσα στην λίστα, βασιζόμενοι στο πλάνο αναμετρήσεων, παρατηρούμε ότι οι ομάδες A1 και A2 θα μπαίνουν στην ομάδα 1 καθώς αποτελούν το πρώτο στοιχείο των ζευγών από την λίστα.

Έτσι βάζουμε ως προϋπόθεση ότι αν στην ομάδα 1 ανήκουν οι ομάδες A1 και A2 τότε το σκορ της ομάδας 1 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή των ομάδων μείον 5 βαθμούς και τυπική απόκλιση 13 αλλιώς το σκορ της ομάδας 1 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή που αντιστοιχεί στην ομάδα και τυπική απόκλιση 13. Εξίσου το σκορ για την ομάδα 2 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή που αντιστοιχεί στην ομάδα και τυπική απόκλιση 13.

Στην συνέχεια, συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε κερδίζει η ομάδα 1 κρατώντας σε ένα κενό διάνυσμα(winners1) την ομάδα 1. Αλλιώς κερδίζει η ομάδα 2 κρατώντας στο κενό διάνυσμα την ομάδα 2.

Το ίδιο ακριβώς εφαρμόζουμε και στο δεύτερο μέρος της συνάρτησης για τις ομάδες που κέρδισαν στον πρώτο γύρο. Αποθηκεύουμε σε μια λίστα τους νικητές του προηγούμενου γύρου ανά ζεύγη. Μέσα από την επαναληπτική διαδικασία, διαλέγουμε την ομάδα 1 και 2 που θα χρειαστεί να αναμετρηθούν από τα στοιχεία της νέας λίστας και φτιάχνουμε ξανά δύο διανύσματα όπου για την ομάδα 1 και για την ομάδα 2 θα μας δίνουν την μέση τιμή από το αρχικό διάνυσμα μέσων τιμών.

Αν στην ομάδα 1 ανήκουν οι ομάδες A1 και A2 τότε το σκορ της ομάδας 1 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή των ομάδων μείον 5 βαθμούς και τυπική απόκλιση 13 αλλιώς το σκορ της ομάδας 1 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή που αντιστοιχεί στην ομάδα και

τυπική απόκλιση 13. Εξίσου το σκορ για την ομάδα 2 θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή που αντιστοιχεί στην ομάδα και τυπική απόκλιση 13.

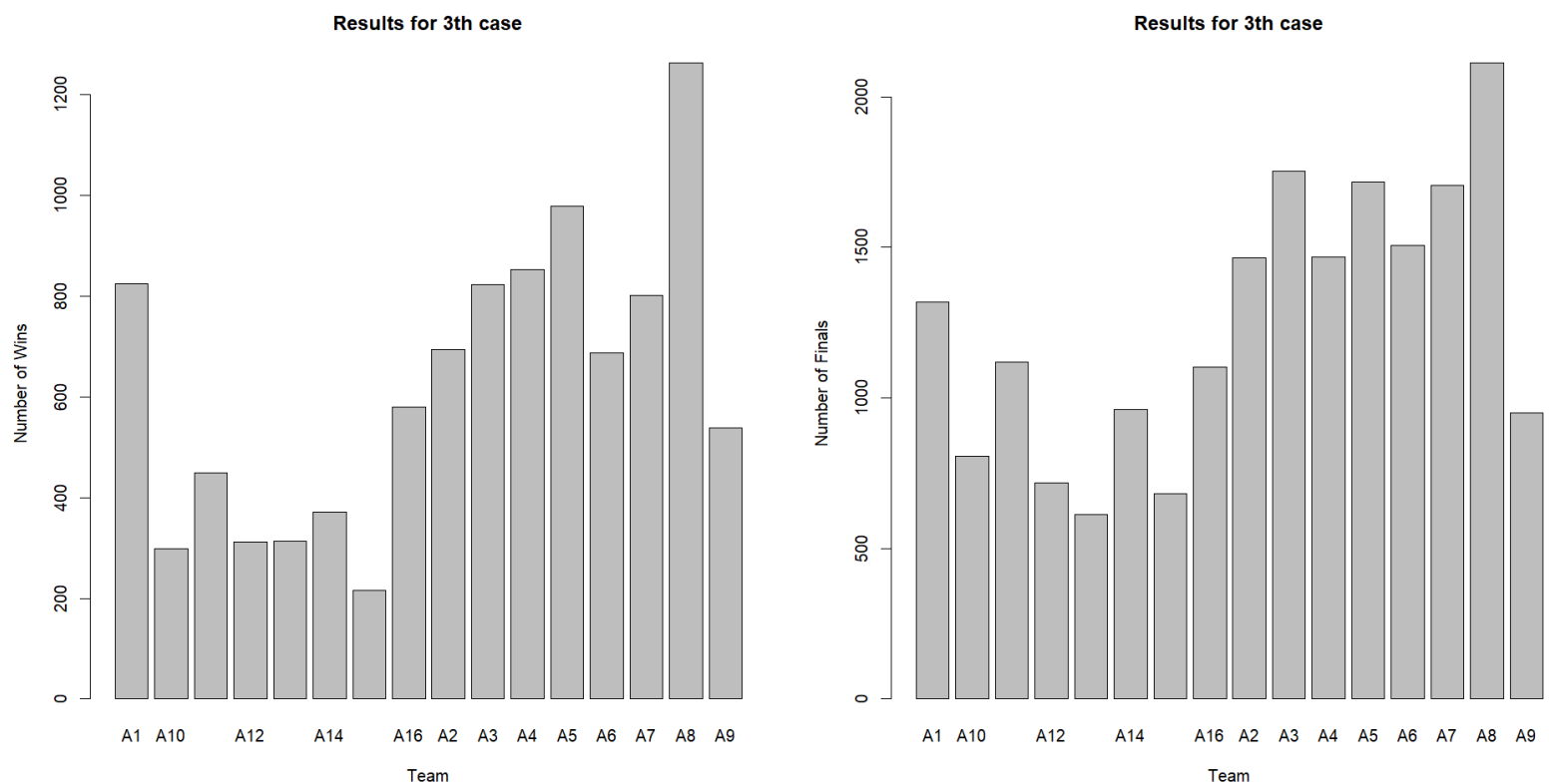
Στην συνέχεια συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε κερδίζει η ομάδα 1 κρατώντας σε ένα κενό διάνυσμα(winners2) την ομάδα 1. Αλλιώς κερδίζει η ομάδα 2 κρατώντας στο κενό διάνυσμα την ομάδα 2.

Το τρίτο μέρος της συνάρτησης παραμένει ίδιο με την αρχική συνάρτηση καθώς η προϋπόθεση για τις ομάδες A1 και A2 θα ισχύει μόνο στους 2 πρώτους αγώνες. Αποθηκεύουμε σε μια λίστα τους νικητές του δεύτερου γύρου ανά ζεύγη. Μέσα από την επαναληπτική διαδικασία, διαλέγουμε την ομάδα 1 και 2 από τα στοιχεία της νέας λίστας, φτιάχνουμε δύο διανύσματα όπου για την ομάδα 1 και για την ομάδα 2 θα μας δίνουν την μέση τιμή από το αρχικό διάνυσμα μέσω των τιμών και ορίζουμε τα σκορ τα οποία θα είναι το ακέραιο μέρος της κανονικής κατανομής με μέση τιμή αυτή των ομάδων και τυπική απόκλιση 13. Συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε κερδίζει η ομάδα 1 κρατώντας σε ένα κενό διάνυσμα(winners3) την ομάδα 1. Αλλιώς κερδίζει η ομάδα 2 κρατώντας στο κενό διάνυσμα την ομάδα 2. Η συνάρτηση κλείνει ζητώντας να μας επιστρέφει τις ομάδες που έχει αποθηκεύσει στον τρίτο γύρο, δηλαδή τις ομάδες που φτάνουν στον τελικό.

Η δεύτερη συνάρτηση που θα καθορίσει τον νικητή παραμένει ακριβώς η ίδια. Ξανά δίνουμε το αρχικό διάνυσμα μέσω των τιμών για όλες τις ομάδες και δηλώνουμε τις ομάδες του τελικού ως το αποτέλεσμα της παραπάνω συνάρτησης. Ως ομάδα 1 ορίζουμε το πρώτο στοιχείο του διανύσματος που επιστρέφει η προηγούμενη συνάρτηση και ως ομάδα 2 το δεύτερο στοιχείο. Παίρνουμε τις μέσες τιμές των ομάδων και ορίζουμε με αυτές τα αντίστοιχα σκορ τους. Συγκρίνουμε τα δύο σκορ λέγοντας ότι αν το σκορ της ομάδας 1 είναι μεγαλύτερο από το σκορ της ομάδας 2 ή αν σε περίπτωση ισοπαλίας παίρνουμε έναν τυχαίο αριθμό από μια ομοιόμορφη κατανομή ο οποίος είναι μεγαλύτερος από 0.5 τότε η συνάρτηση θα επιστρέφει την ομάδα 1 αλλιώς θα επιστρέφει την ομάδα 2.

2.1. Αποτελέσματα

Προσομοιώνοντας και τις 2 συναρτήσεις για 10000 φορές και παίρνοντας τις συχνότητες των αποτελεσμάτων παρατηρούμε ότι η ομάδα που εμφανίζεται περισσότερες φορές στον τελικό και στο τέλος νικήτρια είναι η A8. Με βάση και τα παρακάτω διαγράμματα συχνοτήτων, συμπεραίνουμε ότι αυτή η ατυχία της ομάδας A1 μείωσε τόσο πολύ την εμφάνιση της στον τελικό και στην νίκη όπου η A8 την ξεπέρασε.



Έπειτα, διαιρώντας τις συχνότητες με το άθροισμα τους, παίρνουμε την πιθανότητα για την κάθε μια ομάδα να φτάσει στον τελικό και να νικήσει. Παρατηρούμε ότι η ομάδα με την μεγαλύτερη πιθανότητα να φτάσει στον τελικό και να νικήσει είναι πλέον η A8 με πιθανότητες περίπου 0,21 και 0,12 αντίστοιχα.

Ομάδες	Τελικός	Νικητής
A1	0,131	0,082
A10	0,08	0,029
A11	0,111	0,044
A12	0,071	0,031
A13	0,061	0,031
A14	0,096	0,037
A15	0,068	0,021
A16	0,11	0,057
A2	0,146	0,069
A3	0,175	0,082
A4	0,146	0,085
A5	0,171	0,097
A6	0,15	0,068
A7	0,17	0,08
A8	0,211	0,126
A9	0,095	0,053
Άθροισμα	2	1

Πίνακας πιθανοτήτων για τελικό και νίκη για την 3^η περίπτωση