

1^η Σειρά Ασκήσεων

Ομάδα:

Άγγελος Τσελές (A.M: 3170160)

Ανδρέας Πολυχρονάκης (A.M: 3170140)

1)

A)

STEMPLOTS

• Δεδομένα I

Stem	Plot
30	3
31	01
32	167
33	46
34	25

• Δεδομένα II

Stem	Plot
0	0028
1	24
3	2
4	2
6	4
9	0

• Δεδομένα III

Stem	Plot
0	0168
1	03567788
2	0156
3	059
4	013468
5	24899
6	06
7	
8	16789
9	46

BOXPLOT

• Δεδομένα I

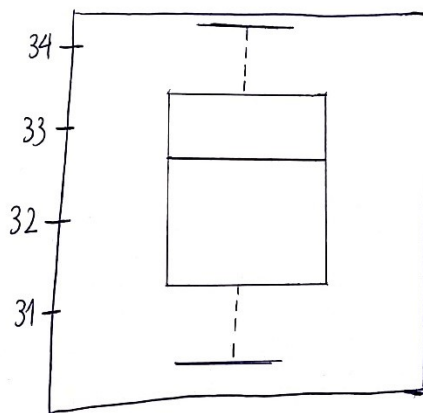
max \rightarrow 34.5

min \rightarrow 30.3

median \rightarrow 32.65

Q1 \rightarrow 31.35

Q2 \rightarrow 33.5



1^η Σειρά Ασκήσεων

BOXPLOT

• Δεδομένα II

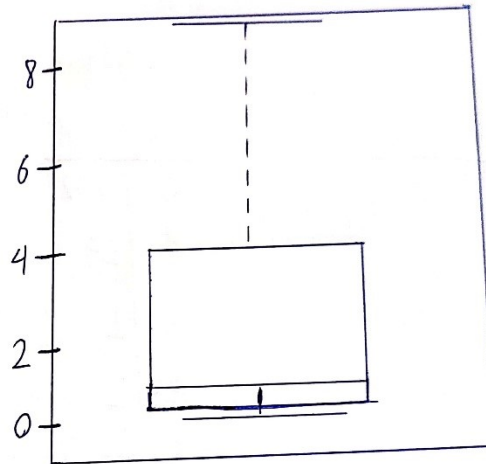
max \rightarrow 9.0

min \rightarrow 0.0

median \rightarrow 1.3

Q1 \rightarrow 0.35

Q2 \rightarrow 3.95



BOXPLOT

• Δεδομένα III

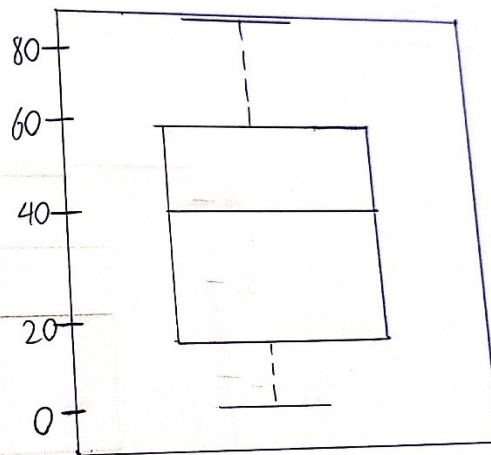
max \rightarrow 96

min \rightarrow 0

median \rightarrow 39.5

Q1 \rightarrow 17.75

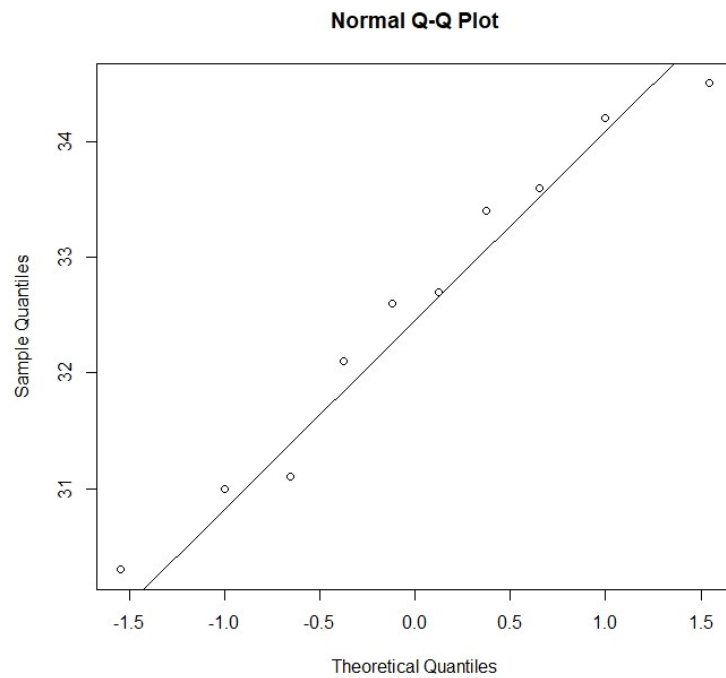
Q2 \rightarrow 59



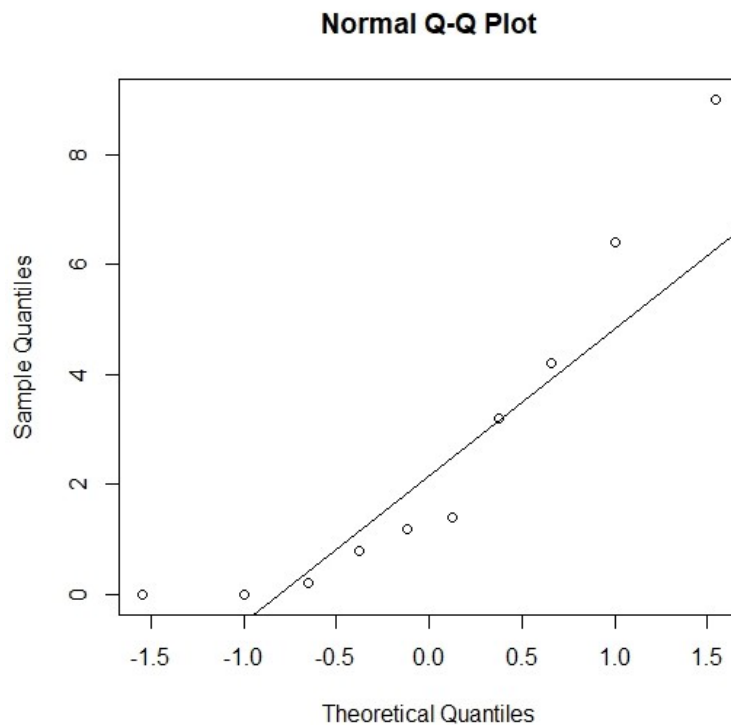
B) Γνωρίζουμε ότι η σύνοψη των 5 αριθμών είναι χρήσιμη στην περίπτωση όπου έχουμε ισχυρά outliers. Παρατηρούμε ότι στα δεδομένα 2 υπάρχουν ισχυρά outliers οπότε η σύνοψη των 5 αριθμών συνοψίζει καλύτερα την κατανομή. Στα δεδομένα 1 και 3 η μέση τιμή και η τυπική απόκλιση μπορούν να συνοψίσουν καλά την κατανομή μιας και βλέπουμε ότι ο κύριος όγκος των περιπτώσεων βρίσκεται στο κέντρο και γύρω από αυτό.

1^η Σειρά Ασκήσεων

Γ)

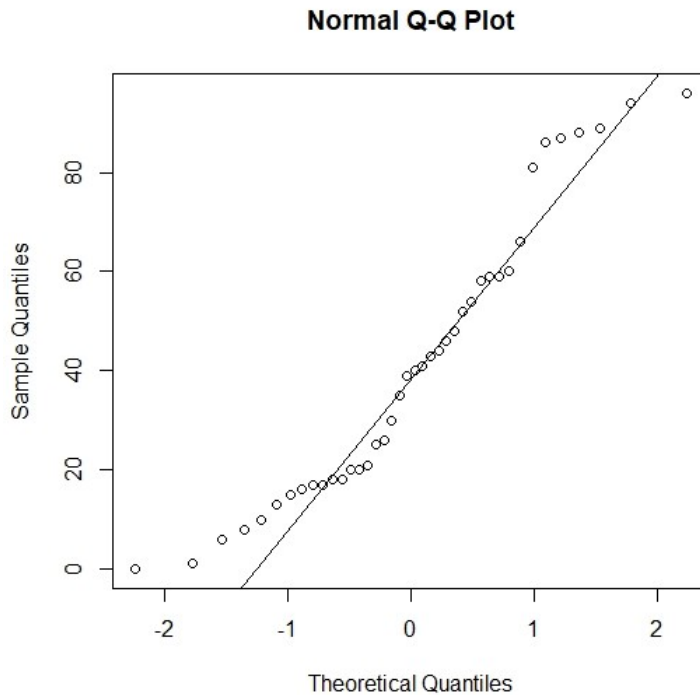


Για τα δεδομένα 1, παρατηρούμε ότι δεν υπάρχουν αποκλίσεις σε όλα τα ποσοστημόρια. Άρα η προσέγγιση της κατανομής των δεδομένων είναι σχετικά ακριβής.



Για τα δεδομένα 2, παρατηρούμε ότι στα δύο ακριανά ποσοστημόρια υπάρχουν αποκλίσεις. Άρα η προσέγγιση της κατανομής των δεδομένων δεν είναι ακριβής στην περίπτωση αυτή.

1^η Σειρά Ασκήσεων



Για τα δεδομένα 2, παρατηρούμε ότι πέρα των ποσοστημόριων -2,-1 δεν υπάρχουν σημαντικές αποκλίσεις. Άρα η προσέγγιση της κατανομής των δεδομένων είναι σχετικά ακριβής στην περίπτωση αυτή.

2)

A) Τα δεδομένα αφορούν όλους τους παίκτες που πέτυχαν γκολ την σεζόν 2019-2020 στο πρωτάθλημα της Super League. Τα δεδομένα προέρχονται από τον επίσημο ιστότοπο της διοργάνωσης(<https://www.slgr.gr/en/statistics/tables/players/18/88/>).

Περιέχονται 148 περιπτώσεις.

B) Κατηγορικές Μεταβλητές: Player, Team

Ποσοτικές Μεταβλητές: Appearances , Goals

Μεταβλητή Player: Το ονοματεπώνυμο κάθε παίκτη

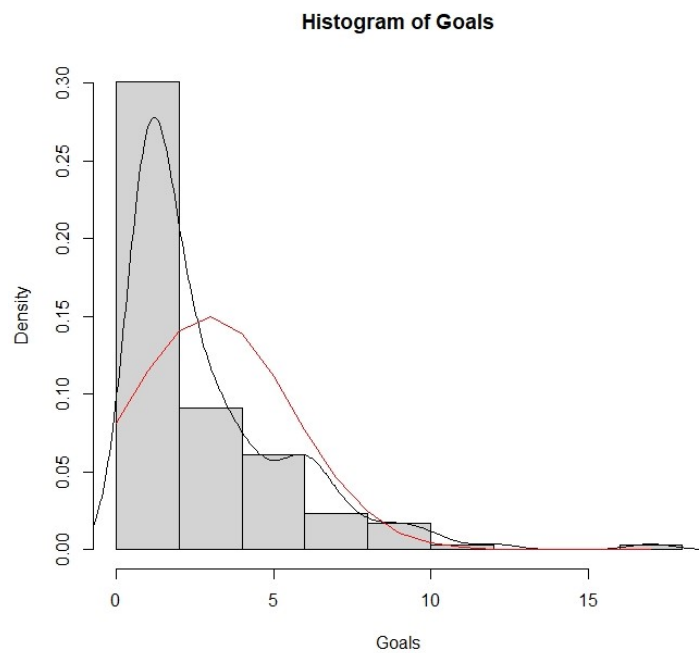
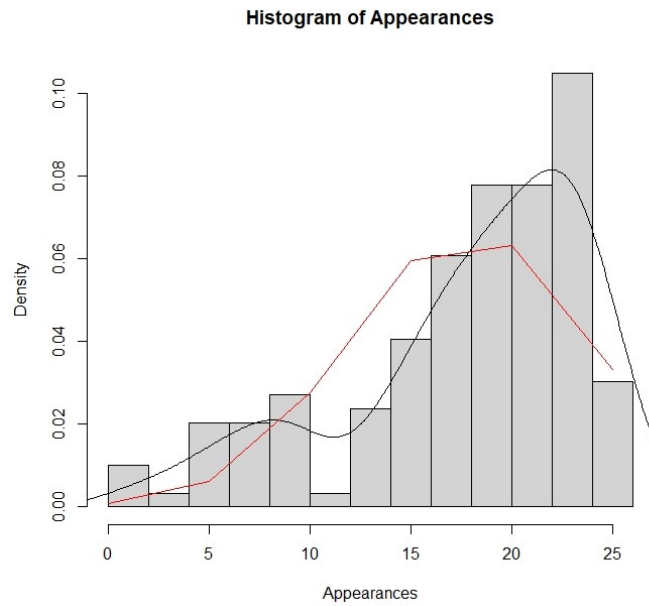
Μεταβλητή Team: Η ομάδα στην οποία αγωνίζεται ο παίκτης

Μεταβλητή Appearances: Οι συμμετοχές του παίκτη στην διάρκεια της σεζόν

Μεταβλητή Goals: Τα τέρματα που πέτυχε ο παίκτης στην διάρκεια της σεζόν

1^η Σειρά Ασκήσεων

Γ)



Μαύρη γραμμή = Προσέγγιση με συνάρτηση πυκνότητας

Κόκκινη Γραμμή = Προσέγγιση με Κανονική Κατανομή

Στο ιστόγραμμα των Appearances ,παρατηρούμε συγκέντρωση των περιπτώσεων στις τιμές 15-25.Αυτό είναι λογικό μιας και οι περισσότεροι παίκτες ,που σκόραραν, έκαναν πολλές εμφανίσεις μέσα στην σεζόν.

Στο ιστόγραμμα των Goals,παρατηρούμε συγκέντρωση των περιπτώσεων στις χαμηλότερες τιμές(0 έως 5).Αυτό είναι λογικό καθώς πολλοί και διαφορετικοί παίκτες κάθε ομάδας σκοράρουν κατά την διάρκεια της σεζόν.

1^η Σειρά Ασκήσεων

Δ)

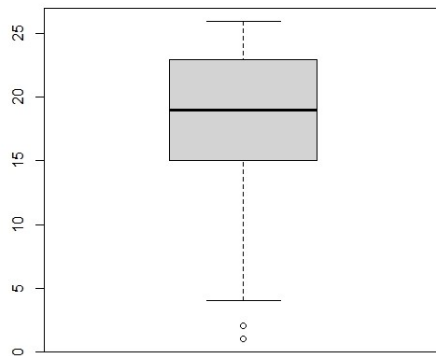
Για την μεταβλητή Appearances:

Μέση Τιμή = 17.93243

Τυπική Απόκλιση = 5.946087m

Σύνοψη των πέντε αριθμών : min=1 , Q1= 15 , median=19 , Q3= 23 , max= 26

Επειδή υπάρχουν ισχυρά outliers (το οποίο είναι φανερό από το boxplot) η σύνοψη των 5 αριθμών είναι καταλληλότερος.



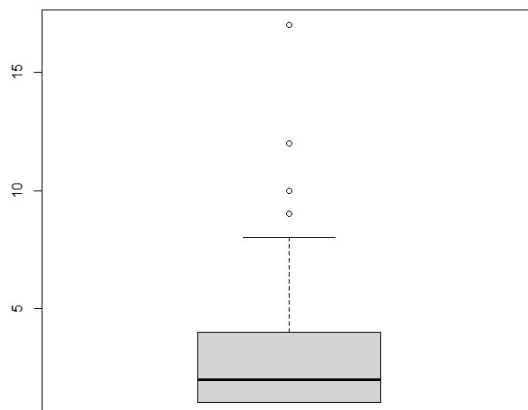
Για την μεταβλητή Goals:

Μέση Τιμή = 2.939

Τυπική Απόκλιση = 2.655315

Σύνοψη των πέντε αριθμών : min=1 , Q1= 1 , median=2 , Q3= 4 , max= 17

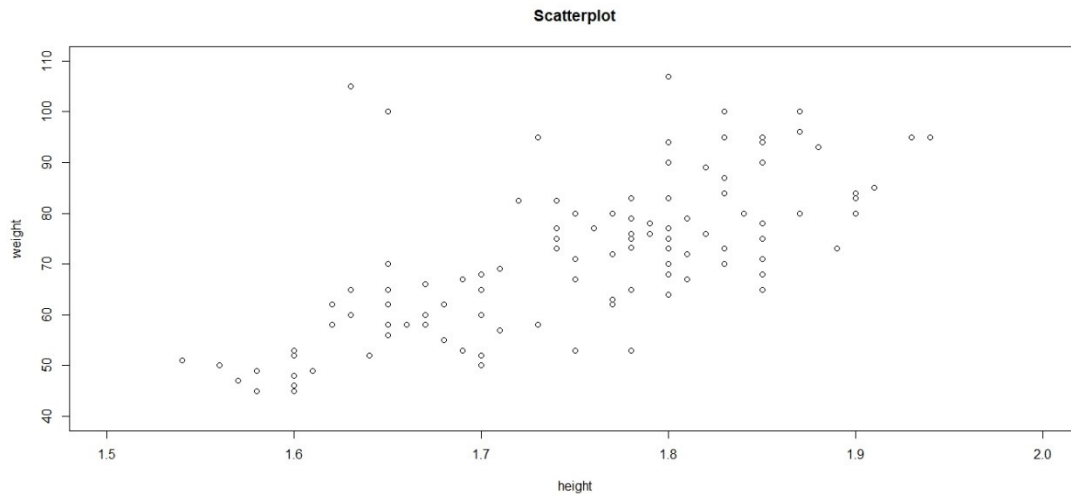
Επειδή υπάρχουν ισχυρά outliers (το οποίο είναι φανερό από το boxplot παρακάτω) η σύνοψη των 5 αριθμών είναι καταλληλότερος.



1^η Σειρά Ασκήσεων

Ε) Επιλέγουμε τις μεταβλητές appearances και goals. Είναι ποσοτικές μεταβλητές και υπάρχει σχέση μεταξύ τους. Η σχέση τους είναι συσχετισμένη καθώς οι τιμές της μίας δίνουν πληροφορία για τις τιμές της άλλης. Όμως δεν είναι αιτιατή καθώς αλλαγές στην τιμή της μιας δεν προκαλούν αλλαγές στην τιμή της άλλης. Πιο συγκεκριμένα, αν αλλάξει η τιμή της μεταβλητής appearances δεν θα αλλάξει η τιμή της μεταβλητής goals και το αντίστροφο.

3)



Διερευνούμε τις μεταβλητές weight και height.

A) Μορφή : Γραμμική με ελάχιστα ατυπικά σημεία
Κατεύθυνση : Αύξουσα
Σχέση : Μέτρια Ισχυρή

B) Συντελεστής Συσχέτισης = 0.6393622

(Υπολογίστηκε με την χρήση της μεθόδου cor)

```
<
> model <- lm(weight[height<2]~height[height<2])
> model

Call:
lm(formula = weight[height < 2] ~ height[height < 2])

Coefficients:
 (Intercept)  height[height < 2]
      -123.6             111.2

> |
```

1^η Σειρά Ασκήσεων

