

ΜΑΘΗΜΑ: «Πληροφορική με Εφαρμογές Στατιστικής» Περιγραφική Στατιστική με την R (Διαγράμματα)

Για τις παρακάτω ασκήσεις να χρησιμοποιήσετε την R

ΑΣΚΗΣΕΙΣ

Στο προηγούμενο εργαστήριο φτιάξατε ένα data frame στην R, το οποίο και το αποθηκεύσατε ως αρχείο κειμένου (*.txt). Το αρχείο βρίσκεται στη σελίδα του μαθήματος στο eclass με το όνομα dataLab11.txt. **Κατεβάστε το και αποθηκεύστε το στην Επιφάνεια εργασίας του Η/Υ σας¹.**

1. Ανοίξτε την R και φορτώστε το data frame που φτιάξατε, φορτώνοντας με κατάλληλο τρόπο τα δεδομένα από το αρχείο txt. **Υπόδειξη:** Χρησιμοποιήστε την εντολή read.table και εισάγετε τα δεδομένα με τον τίτλο datafile1. Στη συνέχεια με την εντολή attach φορτώστε τα δεδομένα στην R ώστε να είναι έτοιμα για επεξεργασία

2. Να συμπληρώσετε τον παρακάτω πίνακα

Στατιστικό Περιγραφικό Μέτρο	Βάρος	Πίεση
Μέση τιμή		
Τυπική απόκλιση		
1ο τεταρτημόριο		
Διάμεσος		
3ο τεταρτημόριο		
Εύρος		
Ενδοτεταρτημοριακό εύρος		
Περιοκομμένος μέσος 10%		
80% ποσοστιαίο σημείο		
35% ποσοστιαίο σημείο		
Συντελεστής γραμμικής συσχέτισης		

3. Εξηγήστε το αποτέλεσμα που σας δίνει η εντολή
> summary(datafile1)

4. Τι ποσοστό των ατόμων του δείγματος, έχει βάρος μικρότερο από 70kg; Τι ποσοστό έχει ύψος μεγαλύτερο από 175cm; Σε κάθε μια από τις παραπάνω υποομάδες, πόσοι είναι άνδρες, πόσοι γυναίκες;
Υπόδειξη: Να χρησιμοποιήσετε τις εντολές which και length.

5. Να δοθεί και να ερμηνευτεί το διάγραμμα διασποράς για τις μεταβλητές height και weight. Να δώσετε διαφορετικά χρώματα στα σημεία, ώστε να ξεχωρίζει το φύλο.

Υπόδειξη: Χρησιμοποιήστε την εντολή
> plot(height,weight,col=gender,pch=20)

¹ Αν το αποθηκεύσετε σε άλλο φάκελο (π.χ. Λήψεις ή Downloads κλπ), να το αντιγράψετε στην Επιφάνεια Εργασίας.

6. Στο διάγραμμα που κατασκευάσατε στο ερώτημα 3, να προσθέσετε τις παρακάτω γραμμές: $x=160$ και $x=185$ (κάθετες γραμμές), $y=55$ και $y=95$ (οριζόντιες διακεκομμένες γραμμές).

7. Να γίνει το ιστόγραμμα συχνοτήτων για το βάρος και το ύψος των ατόμων καθώς επίσης και το ιστόγραμμα πυκνότητας για το δείκτη μάζας σώματος. Να χρησιμοποιήσετε 5 κλάσεις.

8. Να δοθεί το δενδρόγραμμα των τιμών της χοληστερίνης για τα μέλη του δείγματος. Μπορείτε να εξηγήσετε πως αναπαρίστανται οι μετρήσεις του δείγματος σε αυτό;

9. Να κατασκευάσετε τα θηκογράμματα της συστολικής πίεσης, χωριστά για άνδρες και γυναίκες και να τα απεικονίσετε στο ίδιο γράφημα. Να κάνετε το ίδιο και για την τιμή της χοληστερίνης.

Υπόδειξη: Ένας “γρήγορος” τρόπος είναι χρησιμοποιώντας την εντολή

```
> boxplot(Sys~gender, names=c("Male", "Female"))
```

Μπορείτε όμως να εργαστείτε και με την εντολή `which` και να ορίσετε δύο διανύσματα για τη συστολική πίεση, ένα για άνδρες και ένα για γυναίκες (άσκηση για το σπίτι!).

10. Να δοθεί ο πίνακας συχνοτήτων για κάθε μια από τις κατηγορικές μεταβλητές στα δεδομένα.

11. Να δώσετε τον πίνακα συνάφειας για το ζεύγος κατηγορικών μεταβλητών Φύλο-Καπνιστική συνήθεια. Με ποια εντολή θα πάρετε τις σχετικές συχνότητες σε κάθε έναν από τους παραπάνω πίνακες; Να δώσετε τους πίνακες με τις περιθώριες σχετικές συχνότητες για το ζεύγος Φύλο-Καπνιστική συνήθεια.

12. Να δώσετε το στοιβαγμένο ραβδόγραμμα συχνοτήτων της καπνιστικής συνήθειας ως προς το φύλο.

13. Να δώσετε το ομαδοποιημένο ραβδόγραμμα συχνοτήτων της καπνιστικής συνήθειας ως προς το φύλο.