

ΜΑΘΗΜΑ: "Εισαγωγή στον Προγραμματισμό"

ΕΡΓΑΣΤΗΡΙΟ 8: Απεικόνιση Δεδομένων

Άσκηση 1 (παραλλαγή Φουσκάκης 2013, 3.1+4.1). Παρακάτω δίνονται οι βαθμοί (με άριστα το 100) 22 φοιτητών της Φιλοσοφικής στο μάθημα της Γλωσσολογίας:

35 55 64 81 37 56 67 61 69 71 33 45 51 32 40 77 92 56 44 20 71 62

- (i). Να βρεθούν η δειγματική μέση τιμή, η δειγματική διάμεσος, η δειγματική τυπική απόκλιση, το 1^ο και το 3^ο τεταρτημόριο, ο συντελεστής μεταβλητότητας καθώς και το ενδοτεταρτημοριακό εύρος.
- (ii). Να κατασκευαστεί το ιστόγραμμα συχνοτήτων και σχετικών συχνοτήτων των βαθμών. Εντοπίστε ομοιότητες και διαφορές μεταξύ τους.
- (iii). Δώστε μια βελτιωμένη εικόνα του ιστογράμματος των βαθμών των φοιτητών δίνοντας χρώμα γκρι στα ορθογώνια, κατάλληλα ονόματα στους άξονες, επεξηγηματικό τίτλο και υπότιτλο. Επίσης να προσθέσετε δύο κάθετες ευθείες διαφορετικού χρώματος για τη μέση τιμή και για τη διάμεσο του δείγματος.

Άσκηση 2 (παραλλαγή Φουσκάκης 2013, 4.3). Θεωρήστε τις παρακάτω παρατηρήσεις μιας τ.μ. X η οποία εκφράζει τη μηνιαία δαπάνη για σούπερ μάρκετ μιας 4μελούς οικογένειας

187.6 125.2 202.1 165.1 143.7 192.3 211.4 208.3 192.8 155.5 173.1 114.6 182.3

- (i). Αναπαραστήστε γραφικά τις τιμές των παρατηρήσεων του δείγματος μέσω της συνάρτησης `plot()` με τα σημεία του γραφήματος να αναπαρίστανται με αστερίσκους και να συνδέονται με γραμμές. Δώστε πράσινο χρώμα στη γραφική παράσταση και επίσης δώστε κατάλληλο τίτλο (χρώματος μπλε) στο γράφημα και κατάλληλους τίτλους στους άξονες.
- (ii). Έστω ότι διαθέτουμε και τις αντίστοιχες μηνιαίες τιμές αλλά το αμέσως προηγούμενο έτος 192.3 184.6 155.4 199.2 203.5 217.4 123.6 139.8 144.5 187.7 188.5 210.6 Προσθέστε τη γραφική παράσταση των νέων παρατηρήσεων (σε μωβ χρώμα) στο γράφημα το προηγούμενου ερωτήματος με χρήση της εντολής `lines()`. Τα νέα σημεία να αναπαρίστανται από κύβους και να ενώνονται με γραμμές μεγαλύτερου πάχους. Δώστε κατάλληλη λεζάντα (`legend`) που να επεξηγεί τα διαφορετικά χρώματα και σχήματα των σημείων του γραφήματος.
- (iii). Στο ήδη υπάρχον γράφημα, προσθέστε δύο οριζόντιες ευθείες (η μια να είναι διακεκομμένη και η άλλη όχι) για τα μέσα μηνιαία έξοδα κάθε έτους. Να χρησιμοποιήσετε τα χρώματα που χρησιμοποιήσατε για κάθε έτος. Δώστε κατάλληλη επεξήγηση με τη βοήθεια της εντολής `text()`.

Άσκηση 3 (παραλλαγή Φουσκάκης 2013, 3.8 + 4.6). Χρησιμοποιήστε το αρχείο DIAMONDSlab.txt (θα το βρείτε στο eclass του μαθήματος) και εισάγετε τα δεδομένα, εκτελώντας τις παρακάτω εντολές:

```
> DFdiam<-read.table("DIAMONDSlab.txt",header=T)
> head(DFdiam)
> attach(DFdiam)
```

Τα δεδομένα προέρχονται από το dataset diamonds που βρίσκεται στη βιβλιοθήκη ggplot2. Αφορούν την τιμή και κάποια άλλα χαρακτηριστικά ενός δείγματος αρκετών χιλιάδων διαμαντιών. Οι μεταβλητές που ενδιαφέρουν είναι οι price (τιμή σε \$), cut (ποιότητα κοπής, Fair, Good, Very Good, Premium, Ideal) και color (από D: χείριστο ως και J: βέλτιστο).

- (i). Δώστε τον πίνακα σχετικών συχνοτήτων της ποιότητας κοπής των διαμαντιών.
- (ii). Κατασκευάστε το ομαδοποιημένο ραβδόγραμμα του χρώματος των διαμαντιών, δοθείσης της ποιότητας κοπής.
- (iii). Κατασκευάστε τα θηκογράμματα της τιμής των διαμαντιών ως προς την ποιότητα κοπής τους.
- (iv). Κατασκευάστε το στοιβαγμένο ραβδόγραμμα του πίνακα διπλής εισόδου σχετικών συχνοτήτων της ποιότητας κοπής (γραμμές) και του χρώματος (στήλες) των διαμαντιών στο δείγμα. Χρησιμοποιήστε ως τιμή του ορίσματος col το terrain.colors(6) και δώστε κατάλληλη λεζάντα (legend) για επεξήγηση των στοιβών. Δώστε τίτλο και υπότιτλο διαφορετικής γραμματοσειράς και μεγέθους στο γράφημα.

Άσκηση 4 (Homework, Φουσκάκης 2013, 4.11). Έστω ότι θέλουμε να κατασκευάσουμε την προοπτική απεικόνιση της συνάρτησης $z = \sin(\sqrt{x^2 + y^3})$ στο χωρίο $[0.01,5] \times [0.01,5]$. Δημιουργήστε μια ακολουθία από 50 ζεύγη τιμών στο εν λόγω χωρίο (π.χ. με τη seq()) και αναπαράγετε το ζητούμενο γράφημα με τη βοήθεια της ενολής persp(). Αρχικά, μη δώσετε τιμές στα ορίσματα theta, phi (αζιμουθιακή και πολική γωνία). Στη συνέχεια δώστε αντίστοιχα τις τιμές 15 και 20. Επίσης, χρωματίστε το γράφημα δίνοντας το όρισμα col=topo.colors(100). Πειραματιστείτε με διάφορες τιμές του ορίσματος expand.

Άσκηση 5 (Homework, Φουσκάκης 2013, 4.12). Έστω η συνάρτηση με τύπο

$$f(x) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}, x > 0, a > 0, b > 0.$$

- (i). Να κατασκευαστεί το διάγραμμα ισοϋψών καμπυλών της $f(x)$ αν $a \in [0.01,10]$, η $X \in [1,5]$ ενώ το $b = 2$.
- (ii). Για τις παραπάνω τιμές, κατασκευάστε την αντίστοιχη επιφάνεια με τη βοήθεια της

εντολής `image()` χρησιμοποιώντας τη χρωματική παλέτα `rainbow(30)`.

Άσκηση 6 (Φουσκάκης 2013, 3.5 + 4.9). Φορτώστε στην R το dataset `trees`, δίνοντας τις εντολές

```
> attach(trees)
```

```
> head(trees) # για epibebaiosi
```

Τα δεδομένα αφορούν 31 αγριοκερασιές και οι μεταβλητές που συμπεριλαμβάνονται είναι `Girth` (διάμετρος, inches), `Height` (ύψος, ft), `Volume` (όγκος, ft³).

- (i). Υπολογίστε κατάλληλα βασικά περιγραφικά μέτρα (μαζί με μέτρα ασυμμετρίας/κύρτωσης).
- (ii). Να γίνουν σε ένα πλαίσιο 2×2 (χρησιμοποιήστε την `par(mfrow=c(2,2))` για το σκοπό αυτό) τα ιστογράμματα συχνοτήτων των 3 μεταβλητών στο dataset.
- (iii). Με χρήση της εντολής `plot()` απεικονίστε τις τιμές της μεταβλητής `Volume`. Προσαρμόστε κατάλληλα τα όρια τιμών των αξόνων. Με χρήση της εντολής `identify()` προσδιορίστε την πιο ακραία από τις τιμές. Στη συνέχεια, προσθέστε ένα βέλος η αιχμή του οποίου να δείχνει το συγκεκριμένο ακραίο σημείο, ενώ στην αρχή του βέλους προσθέστε ένα κομμάτι κειμένου που να αναγράφει την τιμή διαμέτρου που αντιστοιχεί σε αυτό το σημείο.
- (iv). Να κατασκευαστεί το ιστόγραμμα των παρατηρήσεων της διαμέτρου των αγριοκερασιών, διαμορφώνοντας κατάλληλα την εμφάνιση του γραφήματος. Δώστε χρώμα στα ορθογώνια των κλάσεων του ιστογράμματος χρησιμοποιώντας στο όρισμα `col` την τιμή `heat.colors(10)` (για περισσότερες λεπτομέρειες, `?heat.colors`).
- (v). Με χρήση της εντολής `pairs()` κατασκευάστε το διάγραμμα διασποράς των σημείων κάθε ζεύγους μεταβλητών του dataset. Τα σημεία κάθε διαγράμματος στον «πίνακα» διαγραμμάτων να αναπαρίστανται από κύβους, να είναι χρώματος «light blue» και να έχουν μέγεθος κατά 50% μεγαλύτερο από το προκαθορισμένο.