

**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ****ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ & ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ****Τομέας Μαθηματικών****Πολυτεχνειούπολη – Ζωγράφου ΑΘΗΝΑ - 157 80****ΤΗΛ. : 772 1774****FAX : 772 1775****ΜΑΘΗΜΑ:** *Ανάλυση Δεδομένων με H/Y (6^ο εξάμηνο)***ΔΙΔΑΣΚΩΝ:** *Δημήτρης Φουσκάκης*

ΕΡΓΑΣΙΑ 1^η

Θέμα Εργασίας: Εισαγωγή στην R και Περιγραφική Στατιστική

Άσκηση 1

Διαφημιστική εταιρεία μελετά τους παράγοντες που επηρεάζουν το ύψος των χρημάτων που χρησιμοποιεί για αγορές ένας ιδιώτης κατά τις εορταστικές περιόδους. Για αυτό το λόγο, συλλέχθηκαν πληροφορίες σε δείγμα 53 διαφορετικών ανθρώπων για τις ακόλουθες μεταβλητές: την ηλικία (**age**) του ατόμου σε έτη, την μοναδική εορταστική περίοδο (**holiday**) που πραγματοποιήθηκαν οι αγορές, με κατηγορίες τις "Christmas", "Easter", "Other", (Χριστούγεννα – Πάσχα – Λοιπές), το φύλο (**sex**) του ατόμου που έκανε τις αγορές, με κατηγορίες τις "Man", "Woman", (Αντρας – Γυναίκα), το μέσο χρόνο (**time**) σε λεπτά που θεωρεί πως γίνεται αποδέκτης διαφημιστικών προωθήσεων (από οποιοδήποτε μέσο) στη διάρκεια μιας ημέρας, το μηνιαίο του εισόδημα (**salary**) σε ευρώ και τέλος το ποσό (**spend**) σε ευρώ που ξόδεψε την εορταστική περίοδο. Κάθε άνθρωπος στο δείγμα έχει προβεί αναγκαστικά σε αγορές σε μία μόνο περίοδο από τις τρεις της μεταβλητής **holiday**.

Τα δεδομένα βρίσκονται στο αρχείο:

http://www.math.ntua.gr/~fouskakis/Data_Analysis/Exercises/gifts.txt

Στα δεδομένα μας, με τον χαρακτήρα "*" συμβολίζουμε τις αγνοούμενες τιμές. Εισάγετε τα δεδομένα στην R με χρήση της εντολής `read.table(·)`, αλλάζοντας κατάλληλα το σύμβολο για τις αγνοούμενες τιμές (για αυτόματη αλλαγή, ελέγξτε στο μενού *help* της R τα δυνατά ορίσματα της εντολής `read.table(·)`). Ποια είναι η δομή του αντικειμένου που δημιουργείται από την παραπάνω εντολή; Στη συνέχεια

αφαιρέστε οποιαδήποτε γραμμή περιέχει αγνοούμενη τιμή. Επιπλέον αφαιρέστε, αν υπάρχει, οποιαδήποτε γραμμή περιέχει τιμές από μη ενήλικα άτομα (ηλικία μικρότερη των 18 ετών).

- i. Δώστε μια περιγραφική ανάλυση, για τις τιμές ή κλάσεις (κατηγορίες) κάθε μίας από τις 6 μεταβλητές ξεχωριστά, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- ii. Με τη βοήθεια κατάλληλου διαγράμματος εξετάστε περιγραφικά αν το ποσό που ξοδεύει κάποιος καταναλωτής (**spend**) στο δείγμα διαφοροποιείται ανάλογα με το ποια εορταστική περίοδο (**holiday**) συμβαίνει αυτή η αγορά. Υλοποιήστε παρόμοιες συγκρίσεις, με χρήση διαγραμμάτων, μεταξύ των τιμών της μεταβλητής (**spend**) και των τιμών ή κλάσεων καθεμιάς από τις υπόλοιπες μεταβλητές. Τι συμπεραίνετε;
- iii. Να κατασκευαστεί ο πίνακας συχνοτήτων και σχετικών συχνοτήτων για τα δεδομένα που αφορούν την ηλικία των ατόμων που συμμετέχουν στην έρευνα με τη χρήση 3 κλάσεων: [18-25), [25-38), [38-50), [50 και άνω). Δώστε κατάλληλα ονόματα στις κατηγορίες της νέας αυτής μεταβλητής την οποία ονομάστε την **f_age**. Εν συνεχεία, κατασκευάστε μια ακόμα κατηγορική μεταβλητή, με όνομα **f_time**, με κλάσεις [0,q₁), [q₁,q₂), [q₂,q₃), [q₃ και άνω), όπου q_i (i = 1, 2, 3) είναι το i^ο τεταρτημόριο των τιμών της μεταβλητής **time**. Κατασκευάστε έναν πίνακα συνάφειας συχνοτήτων των μεταβλητών **f_age** και **f_time** στο δείγμα. Δώστε τις σχετικές συχνότητες κελιών και σχολιάστε τα αποτελέσματα. Δημιουργήστε ένα στοιβαγμένο ραβδόγραμμα και σχολιάστε τα αποτελέσματα.

Άσκηση 2

α) Να γράψετε μια συνάρτηση στην R (με όνομα **imitono**) η οποία να υπολογίζει προσεγγιστικά και να επιστρέφει την τιμή του $\sin(x)$, όπου x είναι γωνία σε ακτίνια

(rad), με χρήση της σχέσης
$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} \dots + \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$
 Η

συνάρτηση θα δέχεται ως παραμέτρους εισόδου τον φυσικό αριθμό n και τον πραγματικό αριθμό x . Η συνάρτηση θα ελέγχει αν ο αριθμός n ανήκει στους φυσικούς αριθμούς, διαφορετικά θα επιστρέφει κατάλληλο μήνυμα λάθους και θα τερματίζει.

β) Να γράψετε τη συνάρτηση **imitono2**, η οποία θα λειτουργεί όπως η **imitono**, με την διαφορά ότι δεν μπορείτε να χρησιμοποιήσετε βρόχους for, while ή repeat.

γ) Να γράψετε τη συνάρτηση **imitono3**, η οποία θα λειτουργεί όπως η **imitono**, με τη διαφορά ότι δεν θα οριστεί τιμή n, αλλά η διαδικασία θα σταματάει όταν το σφάλμα προσέγγισης είναι μικρότερο του χιλιοστού. Δηλαδή η διαδικασία δεν ολοκληρώνεται όσο $|x_{n+1} - x_n| \geq 0.001$, όπου x_n είναι η τιμή που έχει πάρει η προσέγγιση του ημιτόνου κατά την n-οστή επανάληψη. Η συνάρτηση επιπλέον θα εκτυπώνει στην οθόνη, το πλήθος των επαναλήψεων που χρειάστηκαν μέχρι επιτευχθεί η συγκεκριμένη ακρίβεια.

δ) Στο ίδιο διάγραμμα να σχηματιστεί με διαφορετικό χρώμα το γράφημα του ημιτόνου (για $x \in [0, 2\pi]$) με χρήση της έτοιμης συνάρτησης *sin()* και με χρήση μιας οποιασδήποτε εκ των τριών που κατασκευάσατε στα προηγούμενα ερωτήματα.

Οδηγίες

- Η εργασία θα πρέπει να **παραδοθεί ηλεκτρονικά** στον Γιώργο Τζουμέρκα στο email του, tzoum_giorgos@hotmail.gr.
- Η εργασία που θα παραδώσετε πρέπει να είναι **σε pdf μορφή**. Παρακαλώ χρησιμοποιήστε τον **ακόλουθο τίτλο στο pdf αρχείο σας**: Surname-Name-Ex1.pdf, όπου Surname είναι το επώνυμό σας (με λατινικούς χαρακτήρες) και Name το όνομα σας (με λατινικούς χαρακτήρες). Π.χ. αν παρέδιδα εγώ εργασία θα την ονόμαζα ως εξής: Fouskakis-Dimitris-Ex1.pdf.
- Παρακαλώ χρησιμοποιήστε **ένα εξώφυλλο στο pdf αρχείο σας**, στο οποίο να αναγράφεται ο τίτλος της εργασίας (Εισαγωγή στην R και Περιγραφική Στατιστική), **το ονοματεπώνυμό σας, το email σας, η Σχολή σας και ο αριθμός μητρώου σας**.
- Θα πρέπει να **αποστείλετε ένα μόνο αρχείο**. Η εργασία θα πρέπει να περιλαμβάνει τους κώδικες της R, όχι σε παράρτημα αλλά στην απάντηση του κάθε ερωτήματος.
- Η εργασία θα πρέπει να αποσταλεί **μέχρι τη Δευτέρα 24/4/2023 στις 13:00**. **Καμιά εργασία δεν θα γίνει δεκτή μετά την ώρα αυτή**.
- Η εργασία θα πρέπει να είναι σε μορφή **αναφοράς** και να περιλαμβάνει τους κώδικες της R με πλήρη επεξήγηση, γραφήματα και πίνακες με

κατάλληλους τίτλους και πλήρη επεξήγηση των αποτελεσμάτων. Επίσης, δε θα πρέπει να υπερβαίνει τις **15 σελίδες** με μέγεθος γραμματοσειράς **12**.

- Θα δοθεί ιδιαίτερη σημασία **στην παρουσίαση της εργασίας**. Η εργασία πρέπει να είναι κατανοητή και να περιγράφει οτιδήποτε χρησιμοποιήσατε πειστικά για κάποιον που δεν γνωρίζει πολλά για το αντικείμενο.

Εύχομαι Επιτυχία