

**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 2^η

Θέμα Εργασίας: Στατιστική Συμπερασματολογία

Άσκηση 1

Μια ερευνήτρια θέλει να εξετάσει τους παράγοντες που επηρεάζουν τον χρόνο που περνούν οι ενήλικες στο τηλέφωνό τους ανά ημέρα. Μέσω ενός ερωτηματολογίου, συλλέχθηκαν πληροφορίες από 70 διαφορετικούς ενήλικες. Οι μεταβλητές που συλλέχθηκαν στο δείγμα για κάθε ενήλικα είναι οι εξής: (α) η ηλικία (**Age**) σε έτη, (β) το φύλο (**Gender**) όπου παρατηρήθηκαν οι κλάσεις “M” (αρσενικό) και “F” (θηλυκό), (γ) το λογισμικό που χρησιμοποιούν (**Operating_System**) στο κινητό τους τηλέφωνο, όπου παρατηρήθηκαν οι κλάσεις “IOS” και “Android”, (δ) ο τύπος εργασίας τους (**Occupation**) χωρισμένος σε κλάσεις “Self_Employed” (αυτοαπασχολούμενος), “Wage Worker” (μισθωτός) και “Other” (άλλο), (ε) η ηλικία του κινητού τους σε ημέρες (**Phone_Age_in_Days**), (στ) ο χρόνος σε ώρες που περάσαν στο κινητό για λόγους ευχαρίστησης (μη εργασιακούς) (**Screen_Time_Leasure**) την προηγούμενη μέρα, (ζ) ο χρόνος σε ώρες που περάσαν στο κινητό για λόγους εργασιακούς (**Screen_Time_Business**) την προηγούμενη μέρα.

Τα δεδομένα βρίσκονται στο αρχείο:

http://www.math.ntua.gr/~fouskakis/Data_Analysis/Exercises/Phone.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. Ένας σύνδεσμος οφθαλμιάτρων αποσκοπεί στην ενημέρωση των πολιτών για την ορθή χρήση κινητών. Ισχυρίζονται ότι ο χρόνος χρήσης κινητού για λόγους ευχαρίστησης για χρονικό διάστημα κατά μέσο όρο μεγαλύτερο των 2.5 ωρών ανά μέρα είναι επιβλαβές για την υγεία των ματιών. Να οριστεί κατάλληλη μεταβλητή και παράμετροι και εν συνεχεία να οριστεί ως έλεγχος υποθέσεων, σε ε.σ. 5%, αν στον πληθυσμό ο μέσος χρόνος χρήσης του κινητού για λόγους ευχαρίστησης υπερβαίνει τις 2.5 ώρες. Να αιτιολογήσετε όποια μέθοδο χρησιμοποιηθεί ως κατάλληλη και να απαντήσετε στο ερευνητικό αυτό ερώτημα.
- ii. Μια διαφημιστική εταιρία σκοπεύει να διαφημίσει ασφαλιστικά προϊόντα μέσω διαφημίσεων στο κινητό. Θεωρούν ότι ασφαλιστικά προϊόντα που αποσκοπούν σε αυτοαπασχολούμενους δεν πρέπει να είναι προτεραιότητα καθώς οι αυτοαπασχολούμενοι δεν χρησιμοποιούν κατά μέσο όρο την ημέρα το κινητό για λόγους ευχαρίστησης το ίδιο συχνά με τις άλλες 2 κατηγορίες. Να κατασκευάσετε μια νέα κατηγορική μεταβλητή (**EmploymentStatus**) με δύο κατηγορίες (0: μη αυτοαπασχολούμενους και 1: αυτοαπασχολούμενους) και περιγράψτε κατάλληλα τις τιμές της τ.μ. **Screen_Time_Leasure** ξεχωριστά για κάθε κατηγορία της νέας αυτής μεταβλητής (συμπεριλάβετε γραφικές και αριθμητικές μεθόδους). Σχολιάστε τα ευρήματά σας.
- iii. Να οριστούν κατάλληλες μεταβλητές και παράμετροι και εν συνεχεία να οριστεί ως έλεγχος υποθέσεων, σε ε.σ. 10%, ο ισχυρισμός διαφημιστικής εταιρίας που αναφέρθηκε στο ερώτημα ii. Να αιτιολογήσετε όποια μέθοδο χρησιμοποιηθεί ως κατάλληλη και να απαντήσετε στο ερευνητικό αυτό ερώτημα.
- iv. Ορμώμενη από στατιστικές μελέτες που καταδεικνύουν διαφοροποίηση στον βαθμό πρεσβυπίας στους άνδρες και τις γυναίκες, η ερευνήτρια θέλει να διερευνήσει τη διαφορά, κατά μέσο όρο, του χρόνου χρήσης κινητού για λόγους ευχαρίστησης ανάμεσα στις δυο αυτές κατηγορίες. Περιγράψτε

κατάλληλα τις τιμές της τ.μ. **Screen_Time_Leasure** ξεχωριστά για κάθε φύλο (συμπεριλάβετε γραφικές και αριθμητικές μεθόδους). Σχολιάστε τα ευρήματά σας. Εν συνεχεία να οριστούν κατάλληλες μεταβλητές και παράμετροι και να οριστεί ως έλεγχος υποθέσεων, σε ε.σ. 1%, ο εν λόγω ισχυρισμός. Να αιτιολογήσετε όποια μέθοδο χρησιμοποιηθεί ως κατάλληλη και να απαντήσετε στο ερευνητικό αυτό ερώτημα.

- v. Να κατασκευάσετε μια νέα κατηγορική μεταβλητή (**Screen_Time_Leasure_f**) με δύο κατηγορίες (0: χαμηλή χρήση και 1: υψηλή χρήση) η οποία να είναι 0 όταν **Screen_Time_Leasure** < 3.5 ώρες και 1 διαφορετικά. Η ερευνήτρια θεωρεί ότι η πιθανότητα οι γυναίκες να κάνουν υψηλή χρήση του κινητού για λόγους ευχαρίστησης, είναι υψηλότερη αυτής των ανδρών. Με χρήση ενός ομαδοποιημένου ραβδογράμματος δείτε αν τα δεδομένα σας συμφωνούν με αυτόν τον ισχυρισμό και σχολιάστε. Εν συνεχεία να οριστούν κατάλληλες μεταβλητές και παράμετροι και να οριστεί ως έλεγχος υποθέσεων, σε ε.σ. 5%, ο εν λόγω ισχυρισμός. Να αιτιολογήσετε όποια μέθοδο χρησιμοποιηθεί ως κατάλληλη και να απαντήσετε στο ερευνητικό αυτό ερώτημα.
- vi. Να ελεγχθεί σε ε.σ. 5% η υπόθεση ότι ο τύπος λογισμικού του κινητού (**Operating_System**) εξαρτάται από το γένος (**Gender**) του ατόμου που κατέχει το κινητό. Να αιτιολογήσετε όποια μέθοδο χρησιμοποιηθεί ως κατάλληλη και να απαντήσετε στο ερευνητικό αυτό ερώτημα. Σε περίπτωση απόρριψης της μηδενικής υπόθεσης, με χρήση κατάλληλων περιγραφικών εργαλείων, να αποφανθείτε που έγκειται η εν λόγω εξάρτηση.

Άσκηση 2

- 1) Ένας τηλεοπτικός σταθμός της Ελληνικής τηλεόρασης έχει βρει σημαντική επιτυχία με ένα τηλεοπτικό παιχνίδι κατά το οποίο ένας διαγωνιζόμενος από το κοινό οδηγείται μπροστά σε τρεις κλειστές πόρτες. Πίσω από μία από αυτές, τυχαία, με πιθανότητα $1/3$, βρίσκεται ένα αυτοκίνητο (το έπαθλο), ενώ πίσω από τις άλλες δύο υπάρχουν κατσίκες. Ο παίκτης καλείται να επιλέξει μία από τις τρεις πόρτες. Η επιλογή του γίνεται τυχαία με πιθανότητα $1/3$. Στη συνέχεια, ο παρουσιαστής (ο οποίος πάντα γνωρίζει πού βρίσκεται το έπαθλο) ανοίγει μία από τις δύο πόρτες που δεν επέλεξε ο παίκτης. Αν πίσω από τις δυο αυτές πόρτες βρίσκεται το έπαθλο, τότε πάντα ο παρουσιαστής ανοίγει

την άλλη πόρτα. Αν το έπαθλο δεν βρίσκεται σε καμία από τις πόρτες (που δεν επέλεξε ο παίκτης), τότε ο παρουσιαστής ανοίγει μία από αυτές τις πόρτες τυχαία (με πιθανότητα $1/2$). Ο παρουσιαστής στη συνέχεια δίνει στον παίκτη τη δυνατότητα είτε να παραμείνει με την αρχική του επιλογή είτε να αλλάξει και να διαλέξει την άλλη κλειστή πόρτα.

- Γράψτε μια συνάρτηση στην R που θα προσομοιώνει το παραπάνω παιχνίδι και θα έχει ως όρισμα τον αριθμό επαναλήψεων του και την στρατηγική του παίκτη.
- Η συνάρτηση θα επιτρέπει 2 στρατηγικές:
 - Ο παίκτης παραμένει πάντα με την αρχική του επιλογή.
 - Ο παίκτης αλλάζει πάντα την επιλογή του μετά την προτροπή του παρουσιαστή.
- Εκτιμήστε και συγκρίνετε τα ποσοστά επιτυχίας για κάθε στρατηγική βασιζόμενοι στις αντίστοιχες σχετικές συχνότητες που προέρχονται από ένα μεγάλο αριθμό επαναλήψεων του παιχνιδιού.
- Εξηγήστε αν είναι συμφέρουσα ή όχι η αλλαγή πόρτας, με βάση τα αποτελέσματα της προσομοίωσης.

2) Το ανταγωνιστικό κανάλι, ανήσυχο από την επιτυχία του παιχνιδιού, αποφασίζει να λανσάρει τη δική του εκδοχή του παιχνιδιού. Μην γνωρίζοντας αυστηρά την φόρμουλα του άλλου παιχνιδιού, ο παρουσιαστής σε αυτή την εκδοχή ανοίγει μία εκ των δύο πορτών που δεν επέλεξε ο παίκτης πάντα τυχαία (με πιθανότητα $1/2$), αφήνοντας έτσι το ενδεχόμενο ο διαγωνιζόμενος να χάσει αυτόματα αν ανοιχτεί η πόρτα με το έπαθλο. Οι διαθέσιμες στρατηγικές στον/στην διαγωνιζόμενο/η είναι οι ίδιες με το προγενέστερο τηλεοπτικό παιχνίδι.

- Γράψτε μια συνάρτηση στην R που θα προσομοιώνει το παραπάνω παιχνίδι και θα έχει ως όρισμα τον αριθμό επαναλήψεων του και την στρατηγική του παίκτη.
- Η συνάρτηση θα επιτρέπει 2 στρατηγικές:
 - Ο παίκτης παραμένει πάντα με την αρχική του επιλογή.
 - Ο παίκτης αλλάζει πάντα την επιλογή του μετά την προτροπή του παρουσιαστή, εκτός αν έχει ήδη ανοιχθεί η πόρτα με το έπαθλο.

- Εκτιμήστε και συγκρίνετε τα ποσοστά επιτυχίας για κάθε στρατηγική βασιζόμενοι στις αντίστοιχες σχετικές συχνότητες που προέρχονται από ένα μεγάλο αριθμό επαναλήψεων του παιχνιδιού.
- Εξηγήστε αν είναι συμφέρουσα ή όχι η αλλαγή πόρτας, με βάση τα αποτελέσματα της προσομοίωσης.

Παρατηρήστε και περιγράψτε τις διαφορές ανάμεσα στα παιχνίδια της εν λόγω Άσκησης.

Οδηγίες

- Η εργασία θα πρέπει να **παραδοθεί ηλεκτρονικά** στον **Σωτήρη Ζαμπέλη** στο email του, szampelis.emp@gmail.com.
- Η εργασία που θα παραδώσετε πρέπει να είναι **σε pdf μορφή**. Παρακαλώ χρησιμοποιήστε τον **ακόλουθο τίτλο στο pdf αρχείο σας**: Surname-Name-Ex2.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: Fouskakakis-Dimitris-Ex2.pdf.
- Παρακαλώ χρησιμοποιήστε **ένα εξώφυλλο στο pdf αρχείο σας**, στο οποίο να αναγράφεται ο τίτλος της εργασίας (Στατιστική Συμπερασματολογία), **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 την 28/4/2025 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