***Ε Θ Ν Ι Κ Ο  Μ Ε Τ Σ Ο Β Ι Ο  Π Ο Λ Υ Τ Ε Χ Ν Ε Ι Ο***
***ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ & ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ***
***Τομέας Μαθηματικών***
***Πολυτεχνειούπολη – Ζωγράφου ΑΘΗΝΑ - 157 80***
***ΤΗΛ. : 772 1774***
***FAX : 772 1775***

**ΜΑΘΗΜΑ:** *Ανάλυση Δεδομένων με Η/Υ (5$^{ο}$ εξάμηνο)*

**ΔΙΔΑΣΚΩΝ:** *Δημήτρης Φουσκάκης*

ΕΡΓΑΣΙΑ 2$^{η}$

**Θέμα Εργασίας**: Στατιστική Συμπερασματολογία

**Άσκηση 1**

Θεωρήστε ότι οι παρακάτω παρατηρήσεις μας δίνουν τον αριθμό προσπαθειών ενός δείγματος 18 μαθητών της Γ' Γυμνασίου κάποιου σχολείου μέχρι να πετύχουν τρίποντο.

3 3 4 3 2 1 1 2 3 4 2 2 3 5 3 6 1 2

i) Να βρεθεί αναλυτικά η Ε.Μ.Π. του ποσοστού επιτυχίας στα τρίποντα των μαθητών της εν λόγω τάξης.

ii) Να βρεθεί γραφικά η Ε.Μ.Π. του ποσοστού επιτυχίας στα τρίποντα των μαθητών της εν λόγω τάξης με τη βοήθεια της R.

**Άσκηση 2**

Θεωρήστε τα δεδομένα του ακόλουθου συνδέσμου: http://www.math.ntua.gr/~fouskakis/myopsy.txt. Μια οφθαλμολογική κλινική παίδων θέλησε να εξετάσει τους παράγοντες από τους οποίους εξαρτάται η ανάπτυξη της παιδικής μυωπίας. Σε δείγμα 618 παιδιών ηλικίας από 5 έως 9 ετών, συλλέχθηκε πληροφορία για τις ακόλουθες μεταβλητές: **id** = ο αύξων αριθμός του κάθε παιδιού, **age** = η ηλικία του (σε έτη), **gender** = το φύλο (0: αγόρι, 1: κορίτσι), **comphr** = οι ανά εβδομάδα ώρες

που το παιδί έπαιζε βιντεοπαιχνίδια ή καθόταν στον ηλεκτρονικό υπολογιστή, **studyhr** = οι ανά εβδομάδα ώρες διαβάσματος/μελέτης για το σχολείο και **parentmy** = αν τουλάχιστον ένας από τους γονείς είναι μύωπας (0: όχι, 1: ναι). Ακόμη μετρήθηκε ο δείκτης σφαιρικού ισοδύναμου διάθλασης (**SER**) με μονάδα μέτρησης τη διόπτρα (D) ο οποίος αφορά τη δυνατότητα εστίασης του οφθαλμού. Όσο περισσότερο ο εν λόγω δείκτης πέφτει κάτω από την τιμή 0, τόσο πιο υψηλή είναι η μυωπία του ατόμου. Αντίστοιχη περιγραφή των μεταβλητών βρίσκεται στην εκφώνηση της **Εργασίας 1**.

i. 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(·)*). Δώστε ονόματα στις 7 μεταβλητές. Με χρήση της εντολής *sample(·)* της R, επιλέξτε τυχαία 45 παρατηρήσεις χωρίς αγνοούμενες τιμές. Δουλέψτε από εδώ και στο εξής με αυτό το σύνολο δεδομένων.

ii. Με χρήση της εντολής *apply(·)*, κατασκευάστε μια νέα μεταβλητή με όνομα '**totalhr**' η οποία να αντιστοιχεί στο σύνολο ωρών που κάθε παιδί αφιερώνει στην ψυχαγωγία μέσω ηλεκτρονικού υπολογιστή και στο διάβασμα σε εβδομαδιαία βάση. Δώστε μια περιγραφική ανάλυση της τυχαίας αυτής μεταβλητής η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.

iii. Να ελεγχθεί σε ε.σ. 2% η υπόθεση ότι η μέση τιμή της τ.μ. 'totalhr' είναι 5 ώρες με εναλλακτική ότι είναι χαμηλότερη από 5 ώρες.

iv. Περιγράψτε κατάλληλα την τ.μ. 'totalhr' ξεχωριστά για κάθε φύλο (συμπεριλάβετε γραφικές και αριθμητικές μεθόδους). Σχολιάστε τα ευρήματά σας.

v. Να ελεγχθεί σε ε.σ. 2% η υπόθεση ότι η μέση τιμή της τ.μ. 'totalhr' δε διαφέρει ανάμεσα στα δύο φύλα με εναλλακτική ότι διαφέρει. Σχολιάστε το αποτέλεσμα.

vi. Να ελεγχθεί σε ε.σ. 10% η υπόθεση ότι η μέση τιμή του δείκτη SER δε διαφέρει ανάμεσα στα δύο φύλα με εναλλακτική ότι είναι υψηλότερη στα κορίτσια. Σχολιάστε το αποτέλεσμα.

vii. Κατασκευάστε μια νέα κατηγορική μεταβλητή για τα δεδομένα που αφορούν στον δείκτη σφαιρικού ισοδύναμου διάθλασης (*SER*) με τη χρήση 3 κλάσεων, όπου τιμή μικρότερη ή ίση από 0 θεωρείται 'χαμηλή (*low*)', τιμή μεγαλύτερη από 0 και μικρότερη ή ίση από 1 θεωρείται 'καλή (*good*)' και τιμή μεγαλύτερη από 1 θεωρείται 'πολύ καλή (*very good*)'. Δώστε αντίστοιχα ονόματα στις κατηγορίες της νέας αυτής κατηγορικής μεταβλητής. Δώστε

μια περιγραφή της νέας κατηγορικής μεταβλητής για την ποιότητα όρασης σε σχέση με το αν τα παιδιά έχουν τουλάχιστον έναν γονέα με μυωπία ή όχι ('parentmy'), η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους.

viii. Να ελεγχθεί σε ε.σ. 5%, με και χωρίς διόρθωση συνέχειας, η υπόθεση ότι το ποσοστό των παιδιών με 'καλή' όραση είναι 50% με εναλλακτική ότι είναι μικρότερο. Χρησιμοποιήστε την κατηγορική μεταβλητή που κατασκευάσατε στο προηγούμενο ερώτημα.

ix. Να ελεγθεί σε ε.σ. 5% η υπόθεση ότι το αν ένα παιδί έχει 'χαμηλή', 'καλή' ή 'πολύ καλή' όραση εξαρτάται από το αν έχει τουλάχιστον έναν γονέα με μυωπία, με εναλλακτική ότι δεν εξαρτάται.

Οδηγίες

- Η εργασία θα πρέπει να παραδοθεί στο γραφείο μου ή στην θυρίδα μου (δίπλα από την γραμματεία του Τομέα Μαθηματικών) εκτυπωμένη και όχι σε ηλεκτρονική μορφή **μέχρι την Τετάρτη, 7 Ιανουαρίου 2014 στις 13:00μμ.** Καμιά εργασία δεν θα γίνει δεκτή μετά την ώρα αυτή.
- Η εργασία θα πρέπει να περιλαμβάνει τους κώδικες της R με πλήρη επεξήγηση, γραφήματα και πίνακες με κατάλληλους τίτλους και πλήρη επεξήγηση των αποτελεσμάτων. Δε θα πρέπει να υπερβαίνει τις 15 σελίδες σε έκταση, χρησιμοποιώντας μέγεθος γραμματοσειράς 12. Οι σελίδες να είναι αριθμημένες.
- 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**