


**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 (5<sup>ο</sup> εξάμηνο)

**ΔΙΔΑΣΚΩΝ:** Μιχάλης Λουλάκης

**ΕΡΓΑΣΙΑ 2<sup>η</sup>**
**Άσκηση 1**

Θεωρήστε ένα συμμετρικό  $n \times n$  πίνακα  $A_n$  του οποίου τα στοιχεία  $A_{ij}$  για  $j \geq i$  είναι ανεξάρτητα δείγματα από την τυπική κανονική κατανομή  $N(0,1)$  (τα υπόλοιπα στοιχεία του  $A$  προσδιορίζονται από την συμμετρία του.) Σύμφωνα με ένα θεώρημα του Wigner, η εμπειρική κατανομή των ιδιοτιμών του πίνακα  $\Lambda_n = (1/\sqrt{n}) A_n$  συγκλίνει καθώς η διάσταση του πίνακα τείνει στο άπειρο σε μια συγκεκριμένη κατανομή. Γραφικά αυτό σημαίνει ότι το ιστόγραμμα των ιδιοτιμών του  $\Lambda_n$  αποκτά ασυμπτωτικά το προφίλ της πυκνότητας πιθανότητας  $f$  αυτής της κατανομής. Θέλουμε να διερευνήσουμε γραφικά το περιεχόμενο του θεωρήματος.

Να γράψετε μια δική σας συνάρτηση στην R η οποία

- θα παίρνει ως όρισμα τη διάσταση  $n$  του συμμετρικού πίνακα,
- θα κατασκευάζει έναν πίνακα  $\Lambda_n$  με τα χαρακτηριστικά που περιγράφηκαν παραπάνω,
- θα υπολογίζει τις ιδιοτιμές του πίνακα με τη βοήθεια της εντολής **eigen()** της R (δείτε τη χρήση της στο Help Menu της R),
- θα επιστρέφει το ιστόγραμμα των ιδιοτιμών του πίνακα  $\Lambda_n$ .

1. Αφού τρέξετε τη συνάρτησή σας για κατάλληλα μεγάλη διάσταση του πίνακα, παρουσιάστε το γραφικό αποτέλεσμα που προβλέπει το θεώρημα του Wigner και σχολιάστε την οριακή πυκνότητα  $f$ .

2. Επαναλάβετε τώρα για την περίπτωση που τα  $A_{ij}$  δεν ακολουθούν τυπική κανονική κατανομή, αλλά ομοιόμορφη κατανομή στο διάστημα  $(-\sqrt{3}, +\sqrt{3})$ , που επίσης έχει μέση τιμή 0 και διασπορά 1. Τι παρατηρείτε για την οριακή πυκνότητα  $f$ ;

**Άσκηση 2**

Οι παρακάτω παρατηρήσεις σε ένα δείγμα 18 γυναικών αφορούν στο πόσες φορές γέννησαν μέχρι να αποκτήσουν το πρώτο τους κορίτσι

3 1 1 1 1 2 1 2 1 2 2 1 1 1 2 4 1 1

- i) Να βρεθεί αναλυτικά η Ε.Μ.Π. της πιθανότητας γέννησης κοριτσιού.
- ii) Να βρεθεί γραφικά η Ε.Μ.Π. της πιθανότητας γέννησης κοριτσιού με τη βοήθεια της R.

**Άσκηση 3**

Το ποσοστό σωματικού λίπους στους άνδρες ηλικίας 20-40 ετών είναι μια τυχαία μεταβλητή  $Y$ . Τιμές άνω του 19% αποτελούν ένδειξη ανδρικής παχυσαρκίας. Σε ένα δείγμα 18 ανδρών της προαναφερθείσας ηλικιακής κατηγορίας, καταγράψαμε το ποσοστό (%) σωματικού λίπους  $Y$ . Τα δεδομένα είναι τα ακόλουθα.

28.5	17.8	17.6	27.4	42.0	13.3	11.8	28.3	19.2
23.9	21.4	9.9	25.1	16.2	31.1	25.8	17.4	24.9

- i) Δώστε μια περιγραφική ανάλυση των παραπάνω δεδομένων, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- ii) Να ελεγχθεί σε ε.σ. 5% η υπόθεση ότι η μέση τιμή της τ.μ. Y είναι 19% με εναλλακτική ότι είναι μεγαλύτερη από 19%. Τι συμπέρασμα βγάζετε;

#### Άσκηση 4

Σε ερώτηση σχετικά με τη χρήση φυσικού αερίου για οικιακή θέρμανση, από ένα τυχαίο δείγμα 93 κατοίκων της Αθήνας οι 67 απάντησαν αρνητικά ενώ οι υπόλοιποι απάντησαν θετικά.

- i) Δώστε μια περιγραφική ανάλυση των παραπάνω δεδομένων, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- ii) Να ελεγχθεί σε ε.σ. 10% η υπόθεση ότι το ποσοστό των κατοίκων της Αθήνας που χρησιμοποιούν φυσικό αέριο για οικιακή θέρμανση είναι 40% με εναλλακτική ότι είναι χαμηλότερο. Τι συμπέρασμα βγάζετε;

#### Άσκηση 5

Ένα βασικό χαρακτηριστικό ενός διαμαντιού είναι η ποιότητα κοπής του. Τα παρακάτω δεδομένα αφορούν την τιμή (σε \$) δύο ανεξάρτητων δειγμάτων διαμαντιών. Το δείγμα A περιλαμβάνει διαμάντια με εξαιρετική ποιότητα κοπής και το δείγμα B περιλαμβάνει διαμάντια με μέτρια ποιότητα κοπής.

ΔΕΙΓΜΑ A	2756	552	2703	2699	12648	397	5309	2093	4491	6452	18065	2155	1984
	15750	7238	22586	9751									
ΔΕΙΓΜΑ B	545	2442	2278	1816	529	2394	995	3215	3816	1175	4289	4169	1059
	1127	4821	5178	2161	884	2372	544	2724					

- i) Δώστε μια περιγραφική ανάλυση των παραπάνω δεδομένων, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- iii) Να ελεγχθεί σε ε.σ. 1% η υπόθεση ότι οι δύο ομάδες διαμαντιών κοστολογούνται κατά μέσο όρο το ίδιο, με εναλλακτική ότι οι μέσες τιμές είναι διαφορετικές. Τι συμπέρασμα βγάζετε;

#### Άσκηση 6

Κάνετε μια μελέτη για τη διαφοροποίηση ή όχι ανάλογα με το φύλο, της πρόθεσης φοιτητών να ακολουθήσουν μεταπτυχιακές σπουδές. Πήρατε ένα τυχαίο δείγμα 152 τελειόφοιτων, αποτελούμενο από 80 άνδρες και 72 γυναίκες. 48 από τους άνδρες, και 59 από τις γυναίκες εξέφρασαν την πρόθεσή τους να συνεχίσουν σε μεταπτυχιακές σπουδές.

- i) Δώστε μια περιγραφική ανάλυση των παραπάνω δεδομένων, η οποία να αποτελείται από κατάλληλες αριθμητικές και γραφικές μεθόδους και σχολιάστε τα ευρήματά σας.
- ii) Να ελεγχθεί σε ε.σ. 2%, με και χωρίς διόρθωση συνέχειας, η υπόθεση ισότητας του ποσοστού των γυναικών που προτίθενται να ακολουθήσουν μεταπτυχιακές σπουδές με το αντίστοιχο των ανδρών, με εναλλακτική ότι το ποσοστό προτίμησης μεταπτυχιακών σπουδών είναι υψηλότερο για τις γυναίκες.

#### Οδηγίες

- Η εργασία θα πρέπει να παραδοθεί στο γραφείο μου ή στην θυρίδα μου (δίπλα από την γραμματεία του Τομέα Μαθηματικών) ή στο γραφείο E2.05 της Ε.Χαριτίδου εκτυπωμένη και όχι σε ηλεκτρονική μορφή **μέχρι την Παρασκευή, 3 Φεβρουαρίου 2012 στις 2:00μμ**. Καμιά εργασία δεν θα γίνει δεκτή μετά την ώρα αυτή.
- Η εργασία θα πρέπει να είναι σε μορφή επίσημης αναφοράς και να περιλαμβάνει τους κώδικες της R με πλήρη επεξήγηση, γραφήματα και πίνακες με κατάλληλους τίτλους και πλήρη επεξήγηση των αποτελεσμάτων.
- 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.

**Καλή Επιτυχία**