

|  |
|--|
| <p style="text-align: center;"><b>ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ</b></p> <p style="text-align: center;"><b>ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ</b></p> <p style="text-align: center;"><b>ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ</b></p> <p style="text-align: center;"><b>ΕΡΓΑΣΙΑ</b></p> |
|--|

Τα δεδομένα (αρχείο *views.sav*) αποτελούν τυχαίο δείγμα 52 ιστοσελίδων και περιέχουν τις ακόλουθες μεταβλητές:

| Όνομα μεταβλητής | Περιγραφή μεταβλητής  |
|------------------|---|
| Country          | Χώρα προέλευσης (1=ελληνική, 0=όχι ελληνική)                                |
| Subject          | Θεματολογία της ιστοσελίδας (1=Αθλητικά, 2=Πολιτικά, 3=Lifestyle)           |
| News             | Ημερήσιος αριθμός νέων αναρτήσεων   |
| Yr               | Παλαιότητα της ιστοσελίδας (1=λειτουργεί λιγότερο από 2 έτη, 0=διαφορετικά) |
| Journalists      | Αριθμός δημοσιογράφων που απασχολούνται στη συγκεκριμένη ιστοσελίδα         |
| Views            | Ετήσιος αριθμός επισκέψεων ( <i>views</i> ) σε συγκεκριμένη ιστοσελίδα      |

Εξαιρώντας την  $i$ -οστή παρατήρηση, όπου  $i$  είναι τα τελευταία δύο ψηφία του αριθμού φοιτητικού μητρώου, να απαντηθούν τα ακόλουθα.

**ΜΕΡΟΣ Α.**

- (i) Να υπολογισθεί η μέση τιμή και η τυπική απόκλιση των ετήσιων αριθμών επισκέψεων των ιστοσελίδων. Να δοθεί η ερμηνεία των αποτελεσμάτων.
- (ii) Να υπολογισθεί η διάμεσος, τα τεταρτημόρια, το 40% ποσοστημόριο και η κορυφή των ετήσιων αριθμών επισκέψεων των ιστοσελίδων. Να δοθεί η ερμηνεία των αποτελεσμάτων.
- (iii) Να ορισθεί κατάλληλα μια νέα μεταβλητή (*Views\_2*), η οποία να ομαδοποιεί τις ιστοσελίδες σε 4 κατηγορίες ανάλογα με τον ετήσιο αριθμό επισκέψεων τους ως εξής:

**1<sup>η</sup> ομάδα:** ιστοσελίδες με ετήσιο αριθμό επισκέψεων μέχρι 22000

**2<sup>η</sup> ομάδα:** ιστοσελίδες με ετήσιο αριθμό επισκέψεων πάνω από 22000 και μέχρι 28000

**3<sup>η</sup> ομάδα:** ιστοσελίδες με ετήσιο αριθμό επισκέψεων πάνω από 28000 και μέχρι 30000

**4<sup>η</sup> ομάδα:** ιστοσελίδες με ετήσιο αριθμό επισκέψεων πάνω από 30000

(α) Να κατασκευασθεί ο πίνακας συχνοτήτων και το κυκλικό διάγραμμα βάσει της νέας μεταβλητής.

(β) Τι ποσοστό των ιστοσελίδων ανήκουν στη 3<sup>η</sup> ομάδα;

- (iv) Να συγκριθούν ως προς τη μεταβλητότητα που παρουσιάζουν οι 4 ομάδες ιστοσελίδων που έχουν δημιουργηθεί βάσει της νέας μεταβλητής (Views \_2). Σχολιάστε τα αποτελέσματα.

## ΜΕΡΟΣ Β.

- (i) Να εξετασθεί σε επίπεδο σημαντικότητας 5% αν οι ελληνικές ιστοσελίδες έχουν την ίδια μέση ετήσια επισκεψιμότητα με τις όχι ελληνικές ιστοσελίδες.
- (ii) Να εξετασθεί σε επίπεδο σημαντικότητας 5% αν οι ιστοσελίδες που λειτουργούν τουλάχιστον 2 έτη, έχουν τον ίδιο μέσο ετήσιο αριθμό επισκέψεων με ιστοσελίδες που λειτουργούν λιγότερο από 2 έτη.
- (iii) Να εξετασθεί σε επίπεδο σημαντικότητας 5% αν ο μέσος ετήσιος αριθμός επισκέψεων των ιστοσελίδων είναι στατιστικά ίσος με 25000 ή όχι.
- (iv) Να εξετασθεί σε επίπεδο σημαντικότητας 5% αν ο παράγοντας Country επηρεάζει τη θεματολογία της ιστοσελίδας.

## ΜΕΡΟΣ Γ.

- (i) Να εξετασθεί αν η μεταβλητή views ( $Y$ ) εξαρτάται γραμμικά από τις μεταβλητές country, subject, news, journalists, yr. Να βρεθεί το βέλτιστο γραμμικό μοντέλο (σε επίπεδο σημαντικότητας 1%) και να δοθεί η γραμμική εξίσωση που αντιστοιχεί σε αυτό.
- (ii) Χρησιμοποιώντας το πλήρες μοντέλο, να εκτιμηθούν οι συντελεστές της γραμμικής του εξίσωσης. Να δοθεί η ερμηνεία των αποτελεσμάτων.
- (iii) Χρησιμοποιώντας το πλήρες μοντέλο, να εκτιμηθεί σημειακά και με διάστημα εμπιστοσύνης 99% ο αναμενόμενος επιπρόσθετος ετήσιος

αριθμός επισκέψεων, που θα παρουσιάσει μία ελληνική ιστοσελίδα, έναντι μίας όχι ελληνικής με τα ίδια χαρακτηριστικά.

### ΜΕΡΟΣ Δ.

Δημιουργούμε μία νέα μεταβλητή (journalists\_2) που ομαδοποιεί τις ιστοσελίδες ανάλογα με τον αριθμό δημοσιογράφων που απασχολούν ως ακολούθως:

**1<sup>η</sup> ομάδα:** ιστοσελίδες με αριθμό δημοσιογράφων μέχρι 8

**2<sup>η</sup> ομάδα:** ιστοσελίδες με αριθμό δημοσιογράφων πάνω από 8 μέχρι και 20

**3<sup>η</sup> ομάδα:** ιστοσελίδες με αριθμό δημοσιογράφων πάνω από 20

- (i) Εφαρμόζοντας κατάλληλο στατιστικό μοντέλο, να εξετασθεί σε επίπεδο σημαντικότητας 5% αν ο ετήσιος αριθμός επισκέψεων μίας ιστοσελίδας εξαρτάται από το αν η ιστοσελίδα ανήκει στην 1<sup>η</sup>, 2<sup>η</sup> ή 3<sup>η</sup> ομάδα βάσει του παράγοντα journalists\_2 και από τη χώρα προέλευσης. Δώστε την τελική μορφή του μοντέλου στην οποία καταλήξατε και σχολιάστε τα αποτελέσματα.
- (ii) Χρησιμοποιώντας την τελική μορφή του μοντέλου που καταλήξατε στο ερώτημα (i), να δοθούν οι σημειακές εκτιμήσεις και τα διαστήματα εμπιστοσύνης 95% για τους μέσους ετήσιους αριθμούς επισκέψεων των ιστοσελίδων για κάθε μία από τις ομάδες που έχουν σχηματισθεί βάσει της μεταβλητής journalists\_2. Σχολιάστε τα αποτελέσματα.

### ΜΕΡΟΣ Ε.

- (i) Να εφαρμοσθεί κατάλληλη στατιστική μέθοδος ώστε να διευκρινιστεί το αν οι μεταβλητές journalists, country, subject, news που αντιστοιχούν σε μία ιστοσελίδα είναι επαρκείς πληροφορίες ώστε να μπορούμε να προβλέψουμε την παλαιότητα της συγκεκριμένης ιστοσελίδας. Να βρεθεί το βέλτιστο μοντέλο πρόβλεψης σε επίπεδο σημαντικότητας 10% και να δοθεί η εξίσωση που αντιστοιχεί σε αυτό.
- (ii) Χρησιμοποιώντας το βέλτιστο μοντέλο, να προβλεφθεί η παλαιότητα μίας ελληνικής ιστοσελίδας για την οποία γνωρίζουμε ότι απασχολεί 12 δημοσιογράφους, πραγματεύεται θέματα αθλητικής επικαιρότητας και στην οποία αναρτώνται ημερησίως 15 νέα άρθρα.

**ΚΑΤΑΛΗΚΤΙΚΗ ΗΜΕΡΟΜΗΝΙΑ ΥΠΟΒΟΛΗΣ: 20 ΣΕΠΤΕΜΒΡΙΟΥ 2014**