

## ΜΑΘΗΜΑ: "Πληροφορική με Εφαρμογές Στατιστικής"

### ΕΡΓΑΣΤΗΡΙΟ 5: Βασικά Στατιστικά Μέτρα & Υπολογισμός με το CALC II

**Άσκηση 1<sup>η</sup>:** Ανοίξτε το CALC και φορτώστε το αρχείο Lab5.ods (υπάρχει στη σελίδα του μαθήματος στο e-class, στα Έγγραφα). Το dataset είναι από το βιβλίο των Berk and Carey (2000) *Data Analysis with Microsoft Excel*, Duxbury (ελαφρώς τροποποιημένο). Κατεβάστε το στον υπολογιστή σας και στη συνέχεια ανοίξτε το με το LibreOffice CALC. Παρατηρήστε ότι έχει ένα μόνο φύλλο εργασίας, στο οποίο δίνονται τιμές των παρακάτω μεταβλητών

**Price:** Τιμή πώλησης του σπιτιού

**Square feet:** Το μέγεθος του σπιτιού σε τετραγωνικά πόδια (μονάδα μέτρησης εμβαδού, 1 square feet = 0.09290304 square meters)

**Age:** Ηλικία του σπιτιού (σε έτη)

**Features:** Πλήθος αντικειμένων οικιακού εξοπλισμού (π.χ. πλυντήριο πιάτων, ψυγείο, φούρνος μικροκυμάτων, το πολύ 11)

**NE Sector:** Βρίσκεται στη νοτιοανατολική πλευρά της πόλης (Yes/No)

**Corner Lot:** Το σπίτι βρίσκεται στη γωνία του δρόμου (Yes/No)

**Offer Pending:** Εκκρεμεί προσφορά για το σπίτι (Yes/No)

**Annual Tax:** Εκτιμώμενο ποσό ετήσιου φόρου ακίνητης περιουσίας

Χρησιμοποιώντας τα δεδομένα της στήλης «**Price**» και τις κατάλληλες συναρτήσεις του CALC, να απαντήσετε στις παρακάτω ερωτήσεις:

- Υπολογίστε τα  $Q_1$ ,  $Q_3$ ,  $IQR = Q_3 - Q_1$ . Στη συνέχεια, υπολογίστε τις τιμές  $KOT = \max\{X_{(1)}, Q_1 - 1.5 \cdot IQR\}$ ,  $AOT = \min\{X_{(n)}, Q_3 + 1.5 \cdot IQR\}$ . Να βρεθεί το ποσοστό των σπιτιών με τιμή πώλησης μικρότερη της KOT, το αντίστοιχο ποσοστό των σπιτιών με τιμή πώλησης μεγαλύτερη της AOT και το συνολικό ποσοστό σπιτιών που ικανοποιούν τη συνθήκη  $\{Τιμή Πώλησης < KOT \text{ ή } Τιμή Πώλησης > AOT\}$ .
- Να υπολογιστεί και να ερμηνευτεί η τιμή του συντελεστή μεταβλητότητας CV (ως προς την ομοιογένεια του δείγματος). Δίνεται ότι  $CV = \frac{s}{\bar{x}}$ .
- Να υπολογιστούν οι τιμές των μέτρων  $\frac{1}{n} \sum |x_i - A|$ , όπου το  $A = \bar{X}$  (άρα θέλετε τη μέση απόλυτη απόκλιση, MAD, και μπορείτε να χρησιμοποιήσετε τη συνάρτηση AVEDEV του CALC) καθώς και η διάμεση απόλυτη απόκλιση (δηλ. η διάμεσος των αποκλίσεων  $|x_i - \delta|$ ,  $i = 1, 2, \dots, n$ , όπου  $\delta$  η δειγματική διάμεσος).
- Να υπολογιστεί ο 10% και ο 20% περικομμένος μέσος (χρησιμοποιήστε τη συνάρτηση =TRIMMEAN() του CALC).
- Να υπολογιστεί ο συντελεστής GINI και η μέση διαφορά GINI για τις τιμές της μεταβλητής Price.

6. Να βρεθεί το μοναδικό πλήθος τιμών της AGE. Η διαδικασία είναι η εξής:

- |          |   |
|----------|---|
| βήμα 1.  | Επιλέξτε το array με τις τιμές της AGE  |
| βήμα 2.  | Από το Menu, επιλέξτε Data / More Filters / Standard Filter.  |
| βήμα 3.  | Στο παράθυρο διαλόγου, στην 1 <sup>η</sup> γραμμή σχετικά Φίλτρα Κριτηρίων (Filter Criteria) δώστε “None” στο Όνομα Πεδίου (Field Name).                              |
| βήμα 4.  | Πατήστε πάνω στο Options (Επιλογές) για να εμφανιστούν περισσότερες επιλογές.   |
| βήμα 5.  | Επιλέξτε το «No duplications» (όχι διπλοεγγραφές, μοναδικές τιμές)  |
| βήμα 6.  | Επιλέξτε το “Copy results to:” και δώστε τον πλήρη προορισμό κελιού στο πεδίο.  |
| βήμα 7.  | Αποεπιλέξτε το “Keep filter criteria”.  |
| βήμα 8.  | Αν το array που επιλέξατε στο 1 δεν περιλαμβάνει τίτλο, αποεπιλέξτε το “Range contains column labels”.  |
| βήμα 9.  | Αν θέλετε να υπάρχει διάκριση μεταξύ Κεφαλαίων ή όχι (εδώ δε θέλετε κάτι τέτοιο γιατί αναζητάτε μοναδικές εγγραφές σε αριθμητικά δεδομένα) επιλέξτε «Case sensitive”. |
| βήμα 10. | Πατάμε OK.  |

**Πριν φύγετε, να αποθηκεύσετε (σε δικό σας φάκελο, είτε στο σκληρό δίσκο του Η/Υ στο εργαστήριο, είτε σε “φλασάκι”) το βιβλίο εργασίας που δημιουργήσαμε στο 5ο εργαστήριο. Να δώσετε το όνομα LAB5.odt. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας πριν φύγετε).**