

ΜΑΘΗΜΑ: “Πληροφορική με Εφαρμογές”

ΕΡΓΑΣΤΗΡΙΟ 8: Περιγραφική Στατιστική με το CALC – Ομαδοποιημένα Δεδομένα

Ανοίξτε το CALC και φορτώστε το αρχείο Lab8.ods (φύλλο εργασίας OM2). Στη στήλη **AA** δίνεται ο αύξων αριθμός των δεδομένων στο αρχείο και στη στήλη **X** δίνεται η τιμή της χοληστερίνης 40 ενηλίκων. Χρησιμοποιώντας τα δεδομένα στο φύλλο εργασίας OM2 και τις αντίστοιχες συναρτήσεις του CALC, να κάνετε ομαδοποίηση των δεδομένων σε κλάσεις ίσου πλάτους (να επιλέξετε εσείς τον αριθμό των κλάσεων, σύμφωνα με τα όσα έχουμε πει) και να δώσετε τον πίνακα συχνοτήτων. Δώστε έμφαση στο πως μπορείτε να χρησιμοποιήσετε για το σκοπό αυτό τη συνάρτηση FREQUENCY.

Υπολογισμός του πλήθους k των κλάσεων, για δείγμα n παρατηρήσεων.

Sturges formula: $1 + \log_2 n$ και στρογγυλοποίηση στον μεγαλύτερο ακέραιο. Για το πλάτος της κλάσης, $c = \frac{X_{(n)} - X_{(1)}}{k}$. Το c δε στρογγυλοποιείται κατ' ανάγκη σε ακέραιο αλλά στρογγυλοποιείται προς τα πάνω.

Freedman – Diaconis rule: Πλάτος κλάσης $c = 2 \cdot \frac{IQR}{\sqrt[3]{n}}$. Στη συνέχεια, το $k = \frac{X_{(n)} - X_{(1)}}{c}$ και η στρογγυλοποίηση γίνεται στον μεγαλύτερο ακέραιο.

Ο πίνακας πρέπει να περιέχει τα εξής στοιχεία: Όρια κλάσεων (κάτω και άνω), κεντρικές τιμές y_i για κάθε κλάση, τιμές για απόλυτες και αθροιστικές συχνότητες, σχετικές και σχετικές αθροιστικές συχνότητες (οι δύο τελευταίες, σε μορφή ποσοστού). Η μορφή του να είναι (κάπως) έτσι:

i κλάση	Κάτω Όριο	Άνω Όριο	y_i	f_i	rf_i	$rf_i\%$	$rcf_i\%$
1							
2							
3							
...							
$k-1$							
k							
Σύνολο							

Χρησιμοποιώντας ΜΟΝΟ τα στοιχεία του πίνακα συχνοτήτων, να υπολογίσετε μέση τιμή και τυπική απόκλιση των (ομαδοποιημένων) δεδομένων.

Αφού υπολογίσετε τις τιμές αυτές, στη συνέχεια να υπολογίσετε μέση τιμή και τυπική απόκλιση των μη-ομαδοποιημένων δεδομένων. Παρατηρείτε διαφορές στις τιμές των μέτρων; Που μπορεί να οφείλεται αυτό; Για τον υπολογισμό της μέσης τιμής και της διασποράς στα ομαδοποιημένα δεδομένα, να χρησιμοποιήσετε τους παρακάτω τύπους

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k f_i y_i,$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^k f_i (y_i - \bar{X})^2 = \frac{1}{n-1} (\sum_{i=1}^k f_i y_i^2 - n\bar{X}^2),$$

όπου k είναι το πλήθος των κλάσεων και y_i είναι η κεντρική τιμή της i -οστής κλάσης.

Να επαναλάβετε τα παραπάνω ερωτήματα αν είναι γνωστές οι κλάσεις για την ομαδοποίηση των δεδομένων: [170, 185), [185, 200), [200, 215), [215, 230). Παρατηρείτε διαφορές στις τιμές της μέσης τιμής και της τυπικής απόκλισης έναντι της προηγούμενης ομαδοποίησης; Υπάρχουν διαφορές αν η ομαδοποίηση γίνει στις κλάσεις (170, 185], (185, 200], (200, 215], (215, 230];

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 8ο εργαστήριο. Να δώσετε το όνομα LAB8.ods. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη.