

# ΜΑΘΗΜΑ: "Πληροφορική με Εφαρμογές Στατιστικής"

## ΕΡΓΑΣΤΗΡΙΟ 10: Γραφήματα με το CALC

### Ασκήσεις

1. Στον παρακάτω πίνακα δίνονται τα ομαδοποιημένα δεδομένα (Σύνολο A) σχετικά με τα μηνιαία έξοδα φοιτητών & φοιτητριών σε μια επαρχιακή πόλη (δείτε επίσης Lab10.ods, φύλλο EX1):

Έξοδα (σε €)	Ποσοστό φοιτητών (%)
[300-350)	4
[350-400)	6
[400-450)	9
[450-500)	11
[500-550)	18
[550-600)	26
[600-650)	16
[650-700)	4
[700-750)	3
[750-800)	2
[800-850)	1
Σύνολο	100%

(α) Να δοθεί το Ιστόγραμμα σχετικών συχνοτήτων.

(β) Να κατασκευάσετε το γραμμικό διάγραμμα των αθροιστικών σχετικών συχνοτήτων.

(γ) Να κατασκευάσετε το ιστόγραμμα σχετικών συχνοτήτων για τα παραπάνω δεδομένα, μαζί με την πολυγωνική γραμμή. Για να το κάνετε αυτό, αρχικά να θεωρήσετε μια κλάση [250-300) με σχετική συχνότητα 0 καθώς και μια κλάση [850-900), επίσης με σχετική συχνότητα 0. Στη συνέχεια να υπολογίσετε τις κεντρικές τιμές κάθε κλάσης και να μαζί με το ιστόγραμμα τα ζεύγη τιμών ( $y_i$ ,  $rf_i\%$ ) (αρχικά θα βγούνε 2 ράβδοι, η μια δίπλα στην άλλη αλλά αυτό στη συνέχεια θα αλλάξει). Οι κατηγορίες στη 2<sup>η</sup> σειρά δεδομένων θα είναι τα  $y_i$ . Τέλος, στο chart type να επιλέξετε Column and Line.

(δ) Στη διάθεσή μας έχουμε επίσης και τα αντίστοιχα ποσοστά σε καθεμία από τις παραπάνω κλάσεις για τους φοιτητές & φοιτητρίες, οι οποίοι σπουδάζουν σε ένα μεγάλο αστικό κέντρο.

**Σύνολο B:** 7, 9, 12, 13, 18, 14, 10, 7, 5, 3, 2.

Χρησιμοποιώντας τα δεδομένα αυτά, να κατασκευάσετε το γραμμικό διάγραμμα των αθροιστικών σχετικών συχνοτήτων για το Σύνολο B και να το απεικονίσετε στο ίδιο γράφημα με το Σύνολο A (δηλ. για τους φοιτητές που σπουδάζουν στην επαρχιακή πόλη). Να συγκρίνετε τη μορφή και τη συμπεριφορά των δύο κατανομών.

(ε) Να επαναλάβετε το (δ) αλλά τώρα να κατασκευάσετε το area chart.

(στ) Να υπολογιστεί η μέση τιμή και η τυπική απόκλιση των δεδομένων και για τα δύο σύνολα. Να τα συγκρίνετε ως προς την ομοιογένεια χρησιμοποιώντας το συντελεστή μεταβλητότητας CV.

2. Τα παρακάτω δεδομένα αφορούν τις πωλήσεις μπίρας (σε εκατομμύρια βαρέλια), για 10 διαφορετικές ετικέτες, καθώς και την αντίστοιχη διαφημιστική δαπάνη (σε εκατομμύρια \$) γι' αυτές. Τα δεδομένα βρίσκονται στο αρχείο Lab10.ods, φύλλο εργασίας EX2. Χρησιμοποιώντας τα δεδομένα αυτά,

- (i) να υπολογίσετε την τιμή του συντελεστή γραμμικής συσχέτισης  $r$  και να κατασκευάσετε το διάγραμμα διασποράς των δεδομένων. Να ερμηνεύσετε τα αποτελέσματα.
- (ii) να βρεθεί η εξίσωση της προσαρμοσμένης ευθείας παλινδρόμησης  $\hat{y} = \hat{a} + \hat{b}x$ , όπου  $Y$ =πωλήσεις και  $X$ =διαφημιστική δαπάνη.
- (iii) να απεικονίσετε στο διάγραμμα διασποράς στο (i) και την προσαρμοσμένη ευθεία παλινδρόμησης.
- (iv) Υπολογίστε τις διαφορές  $e = y - \hat{y}$  και να κατασκευάσετε το διάγραμμα διασποράς  $(e, \hat{y})$ .
- (v) Χρησιμοποιώντας ως 3<sup>η</sup> μεταβλητή την τιμή του αλκοόλ κάθε μπίρας, να κατασκευάσετε το bubble chart των  $X$ =διαφημιστική δαπάνη και  $Y$ =πωλήσεις.

Ετικέτα	Διαφημιστική δαπάνη (εκατ. \$)	Πωλήσεις (σε εκατ. βαρέλια)	Αλκοόλ (%)
Budweiser	120,0	36,3	5
Bud Light	68,7	20,7	4.2
Miller Lite	100,1	15,9	4.2
Coors Light	76,6	13,2	4.2
Busch	8,7	8,1	5.9
Natural Light	0,1	7,1	4.2
Miller Genuine Draft	21,5	5,6	4.7
Miller High Lite	1,4	4,4	4.2
Busch Lite	5,3	4,3	4.2
Milwaukee's Best	1,7	4,3	4.5

**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 10ο εργαστήριο. Να δώσετε το όνομα LAB10.ods. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας φεύγοντας).**