

# ΜΑΘΗΜΑ: “Πληροφορική με Εφαρμογές Στατιστικής”

## ΕΡΓΑΣΤΗΡΙΟ 10: Γραφήματα με το CALC

### Ασκήσεις

1. Τα παρακάτω δεδομένα (δείτε επίσης Lab8.ods, φύλλο EX1) αφορούν τις απαντήσεις που έδωσαν 60 φοιτητές στα πλαίσια της αξιολόγησης ενός μαθήματος το οποίο είχε παρακολουθήσει. Η ερώτηση αφορούσε το πόσο ικανοποιημένοι είναι από το μάθημα. Οι (δυνατές) απαντήσεις κωδικοποιημένες είναι οι εξής:

“1”=Καθόλου, “2”=Λίγο, “3”=Ούτε πολύ, ούτε λίγο, “4”=Πολύ, “5”=Πάρα πολύ.

3	4	4	5	1	5	3	4	5	2	4	5	3	4	4
4	5	5	4	1	4	5	4	2	5	4	2	4	4	4
5	5	3	4	5	5	2	4	3	4	5	4	3	5	4
4	3	5	4	5	4	3	5	3	4	4	3	5	3	3

(α) Εξηγήστε γιατί τα παραπάνω δεδομένα είναι ποιοτικά (κατηγορικά) δεδομένα.

(β) Να δοθεί ο πίνακας συχνοτήτων με τις απόλυτες και τις σχετικές συχνότητες των τιμών του δείγματος. Ποια είναι η κορυφή (επικρατούσα τιμή) των δεδομένων. Να χρησιμοποιήσετε την εντολή FREQUENCY του CALC.

(γ) Να κατασκευαστεί το ραβδόγραμμα και το κυκλικό διάγραμμα σχετικών συχνοτήτων. Να φαίνονται οι σχετικές συχνότητες πάνω στο κυκλικό διάγραμμα.

(δ) Τι συμπέρασμα συνάγετε από την ανάλυση που κάνατε σχετικά με το βαθμό ικανοποίησης των φοιτητών;

(ε) Έστω ότι έχετε καταγράψει και το φύλο κάθε φοιτητή/τριας (στήλη Gender). Να φτιάξετε τον πίνακα διπλής εισόδου για τις μεταβλητές X και φύλο. Να χρησιμοποιήσετε τη διαδικασία Pivot Tables (Συγκεντρωτικός Πίνακας).

Η διαδικασία περιγράφεται παρακάτω:

Επιλέξτε αρχικά όλα τα δεδομένα (στήλες A, B, C) και από το Δεδομένα / Συγκεντρωτικός Πίνακας/Επιλογή: Τρέχουσα επιλογή και πατάμε Εντάξει. Στο **Πεδία στήλης**, πατάμε πάνω στο Δεδομένα και με το πλήκτρο Delete το σβήνουμε. Από το **Διαθέσιμα πεδία**, επιλέγουμε το X και με drag and drop το «σέρνουμε» στο **Πεδία γραμμής**. Κάνουμε το ίδιο και για το Gender (από το **Διαθέσιμα πεδία**) και το «σέρνουμε» στο **Πεδία στήλης**. Τέλος, στο Πεδία δεδομένων βάζουμε όποια από τις μεταβλητές X ή Gender θέλουμε (εδώ θα βάλουμε τη X, η οποία εμφανίζεται ως Άθροισμα - X). Πατάμε διπλό αριστερό κλικ πάνω στο Άθροισμα - X και επιλέγουμε στο παράθυρο Πεδίο δεδομένων / Συνάρτηση το Πλήθος (μόνο αριθμοί). Πατάμε Εντάξει και ξανά Εντάξει και θα προκύψει ο ζητούμενος πίνακας διπλής εισόδου (σε νέο φύλλο εργασίας. Να το μετονομάσετε σε 2-way table).

(στ-**homework**) Να επαναλάβετε το ερώτημα (ε), χρησιμοποιώντας τη συνάρτηση =COUNTIFS(...). Θα χρειαστεί λίγη περισσότερη δουλειά για να φτιάξετε τον πίνακα αλλά οι συχνότητες μπορούν να υπολογιστούν με την COUNTIFS.

(ζ) Να γίνει το ραβδόγραμμα (column chart) για τις απαντήσεις των φοιτητών (τιμές της X) ανά φύλο. Να το κάνετε ως ομαδοποιημένο (grouped) αλλά και ως στοιβαγμένο (stacked). Αρχικά να το κάνετε με χρήση των συχνοτήτων (δηλ. να είναι ραβδόγραμμα συχνοτήτων). Στη συνέχεια να κάνετε τα αντίστοιχα ραβδογράμματα σχετικών συχνοτήτων.

Επίσης, να γίνει το ραβδόγραμμα (συχνοτήτων και σχετικών συχνοτήτων, ομαδοποιημένο και στοιβαγμένο) για το φύλο ως προς τις τιμές της X.

Ένας απλός τρόπος είναι ο εξής (για το ομαδοποιημένο ραβδόγραμμα συχνοτήτων των τιμών της X ανά φύλο): Επιλέξτε τις συχνότητες που έχετε στον συγκεντρωτικό πίνακα. Στη συνέχεια, από το Εισαγωγή / Διάγραμμα, εμφανίζεται άμεσα το διάγραμμα. Η 1<sup>η</sup> από τις 3 μορφές του διαγράμματος είναι αυτή του ομαδοποιημένου ραβδογράμματος (grouped bar chart). Η 2<sup>η</sup> είναι το στοιβαγμένο (stacked bar chart) ενώ η 3<sup>η</sup> είναι το στοιβαγμένο ραβδόγραμμα σχετικών συχνοτήτων (με τις σχετικές συχνότητες να δίνονται ως ποσοστό). Πατήστε Τέλος και θα εμφανιστεί το ζητούμενο διάγραμμα.

Δοκιμάστε να κάνετε τα παραπάνω χωρίς να χρησιμοποιήσετε το αποτέλεσμα του ομαδοποιημένου πίνακα.

(η) Στη στήλη Expenses έχουμε καταγράψει τα μηνιαία έξοδα για διατροφή & διασκέδαση κάθε φοιτητή/τριας που συμμετείχε στην έρευνα. Χρησιμοποιήστε τις συναρτήσεις =AVERAGEIF(...), =AVERAGEIFS(...) για να υπολογίσετε

(i) το μέσο όρο των μηνιαίων εξόδων στα μέλη του δείγματος που αξιολόγησαν το μάθημα με 1. Αφού καταλάβετε πως γίνεται, να το κάνετε και για τις υπόλοιπες κατηγορίες της μεταβλητής X και να φτιάξετε τον σχετικό πίνακα.

(ii) το μέσο όρο των μηνιαίων εξόδων για τους φοιτητές και για τις φοιτήτριες, ξεχωριστά.

(iii) το μέσο όρο των μηνιαίων εξόδων για τους φοιτητές που αξιολόγησαν το μάθημα με 2 (προσοχή, θα πρέπει να ικανοποιούνται ταυτόχρονα 2 κριτήρια).

(iv) το μέσο όρο των μηνιαίων εξόδων για τις φοιτήτριες που αξιολόγησαν το μάθημα με 4 ή με 5.

Στα (iii), (iv), χωρίς να είναι υποχρεωτικό, μπορείτε να χρησιμοποιήσετε και τις συναρτήσεις AND(...), OR(...)

2. Στον παρακάτω πίνακα δίνονται τα ομαδοποιημένα δεδομένα (Σύνολο A) σχετικά με τα μηνιαία έξοδα φοιτητών & φοιτητριών σε μια επαρχιακή πόλη (δείτε επίσης Lab8.ods, φύλλο EX2):

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 & φοιτήτριες, 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.

3. 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 Lab8.ods, φύλλο εργασίας EX3. 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 8ο εργαστήριο. Να δώσετε το όνομα LAB8.xlsx. Ενδέχεται να το χρησιμοποιήσουμε και σε επόμενη διάλεξη (οπότε φροντίστε να το πάρετε μαζί σας φεύγοντας).**