

## 3ο Φροντιστήριο – Στατιστική Ι

**Κεφ. 5 : Παρουσίαση Δεδομένων**

**Ασκήσεις Κατανόησης**

## Άσκηση 1

Ποια είναι η καλύτερη μέθοδος γραφικής παρουσίασης μιας ποιοτικής μεταβλητής: το κυκλικό διάγραμμα ή το ακιδωτό (ραβδόγραμμα) διάγραμμα;

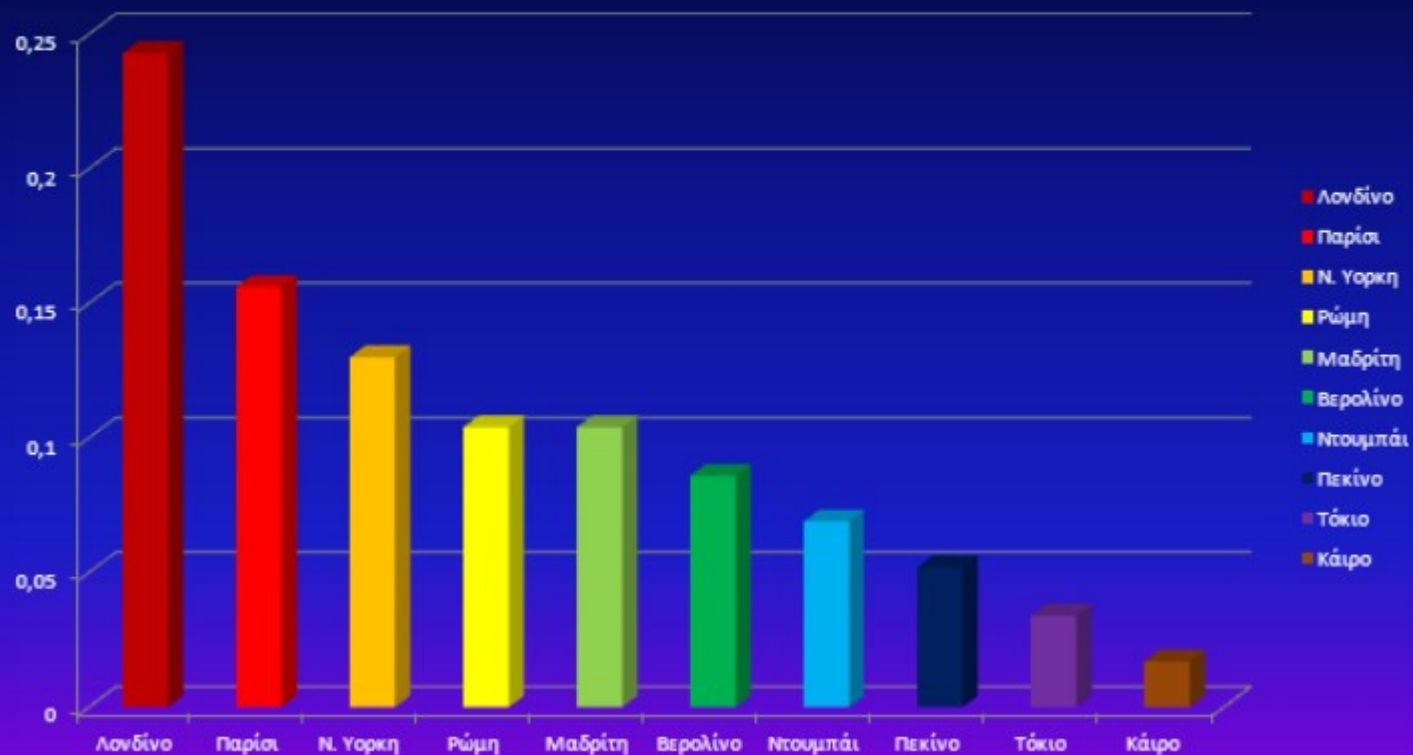
Απάντηση: Ουσιαστικά δεν υπάρχει διαφορά και μπορούν να χρησιμοποιηθούν και οι δύο μέθοδοι. Ωστόσο θα πρέπει να λαμβάνονται υπόψη και οι ιδιαιτερότητες των δεδομένων.

Αν π.χ. η μεταβλητή που θέλουμε να παρουσιάσουμε έχει πολλά επίπεδα, τότε ίσως είναι καλύτερο να χρησιμοποιήσουμε ακιδωτό διότι το κυκλικό ενδεχομένως να είναι δυσανάγνωστο.

Αν θέλουμε να φανεί ότι τα ερευνητικά δεδομένα είναι τμήμα ενός συνόλου, τότε ίσως είναι προτιμότερο να επιλέξουμε το κυκλικό διάγραμμα αξιοποιώντας τη σχετική συχνότητα (rf) των δεδομένων.

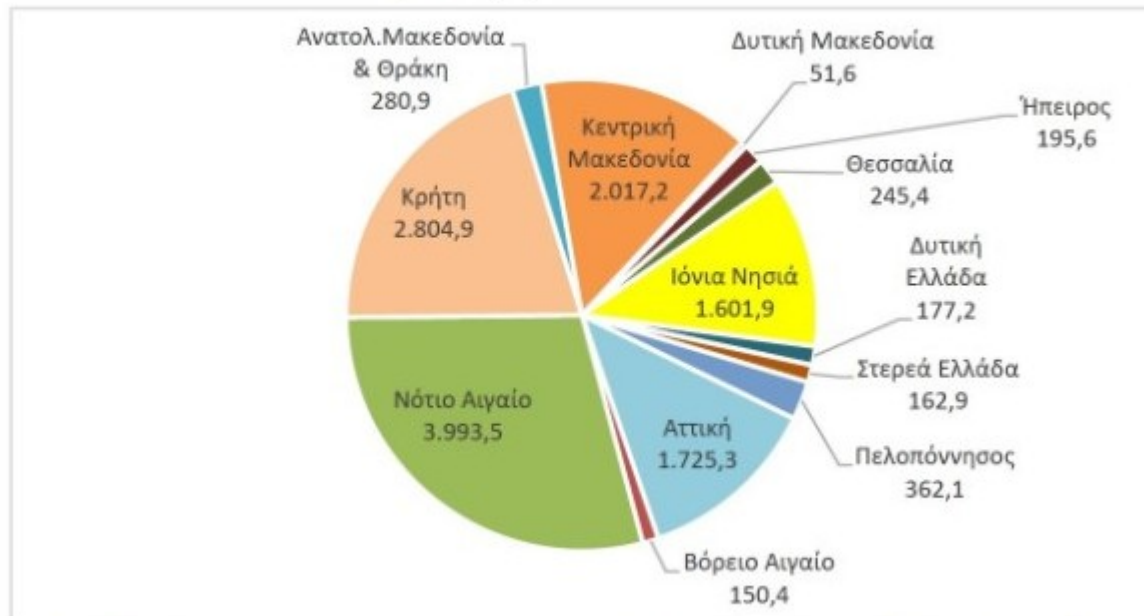
Αν όμως πιστεύουμε ότι θα βοηθούσε για την εξαγωγή πρώτων συμπερασμάτων η παρουσίαση των απόλυτων τιμών των ερευνητικών δεδομένων, τότε επιλέγουμε το ακιδωτό διάγραμμα.

## Ραβδόγραμμα πόλεων-σχετικών συχνοτήτων





**ΚΑΤΑΝΟΜΗ ΤΑΞΙΔΙΩΤΙΚΩΝ ΕΙΣΠΡΑΞΕΩΝ ΣΤΙΣ 13 ΠΕΡΙΦΕΡΕΙΕΣ**  
 Ιανουάριος-Σεπτέμβριος 2018  
 (εκατ. Ευρώ)



ΠΗΓΗ: Τράπεζα της Ελλάδος

Επεξεργασία στοιχείων PDI Infographics για το Euro2day.gr

## Άσκηση 3

Γιατί κάποιος να προτιμήσει σε μία έρευνα ένα φυλλογράφημα αντί για ένα ιστόγραμμα ή ένα πολύγωνο συχνότητας;

Απάντηση: Γιατί με αυτόν τον τρόπο παρουσίασης δεδομένων έχουμε τη δυνατότητα σε πολύ μικρό χώρο να έχουμε μπροστά μας όλες τις μεμονωμένες τιμές που συλλέχθηκαν. Με τις άλλες δύο μεθόδους, δεν μπορούμε να δούμε τις μεμονωμένες τιμές αλλά μόνο τον αριθμό των ατόμων (ή το ποσοστό) που μετρήθηκαν με μία συγκεκριμένη τιμή.

## Παράδειγμα

Έστω ότι έχουμε τα παρακάτω δεδομένα, τα οποία έχουν διαταχθεί κατ' αύξουσα τιμή :

8, 13, 16, 25, 26, 29, 30, 32, 37,  
38, 40, 41, 44, 47, 49, 51, 54, 55,  
58, 61, 63, 67, 75, 78, 82, 86, 95

Εδώ τα διακριτά δεδομένα είναι διψήφιοι αριθμοί.

- Το ψηφίο των δεκάδων είναι ο **μίσχος** (ή κορμός)
- Το ψηφίο των μονάδων είναι το **φύλλο**

0		8
1		3 6
2		5 6 9
3		0 2 7 8
4		0 1 4 7 9
5		1 4 5 8
6		1 3 7
7		5 8
8		2 6
9		5

## Άσκηση 4

Ποια είναι η βασική σχεδιαστική διαφορά ανάμεσα σε ένα ακιδωτό διάγραμμα και σε ένα ιστόγραμμα; Ποια κλίμακα μέτρησης είναι κατάλληλη για το καθένα από αυτά;

Απάντηση: Η βασική διαφορά είναι ότι στο ακιδωτό υπάρχουν κενά ανάμεσα στις ράβδους, ενώ στο ιστόγραμμα οι ράβδοι είναι ενωμένοι, δηλώνοντας τη συνέχεια της μεταβλητής.

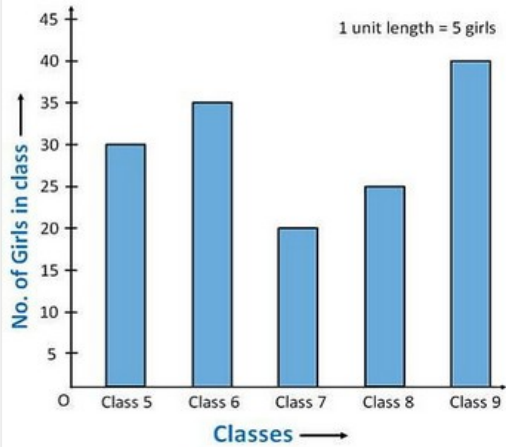
Η πιο κατάλληλη κλίμακα μέτρησης για το ακιδωτό διάγραμμα είναι η κατηγορική, ενώ για το ιστόγραμμα είναι είτε η κλίμακα ίσως διαστημάτων είτε η αναλογική.

# Διαφορά Ακιδωτού διαγράμματος και Ιστογράμματος

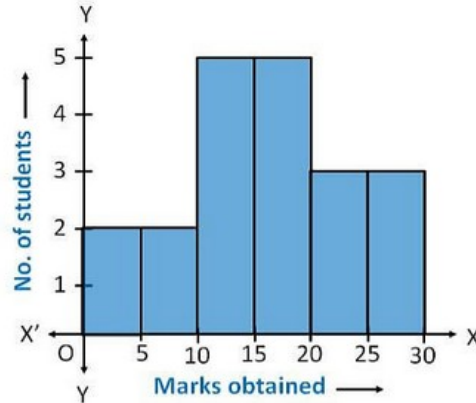
## Difference between Bar Graph & Histogram

teachoo.com

### Bar Graph



### Histogram



In Bar Graph

- Bars have equal space
- On the y-axis, we have numbers & on the x-axis, we have data which can be anything.

In Histogram

- Bars are fixed
- On the y-axis, we have numbers & on the x-axis, we have data which is continuous & will always be number.

## Άσκηση 5

Γιατί είναι σημαντική η απόφαση που θα πάρει μια ερευνήτρια για το εύρος των διαστημάτων που θα χρησιμοποιήσει σε μία ομαδοποιημένη κατανομή; Τι αντίκτυπο μπορεί να έχει μία τέτοια απόφαση στις πληροφορίες που παίρνουμε για τα συγκεκριμένα δεδομένα;

Απάντηση: Είναι σημαντική απόφαση γιατί, ανάλογα με το μέγεθος του εύρους (πλάτους) των διαστημάτων μιας ομαδοποιημένης κατανομής, καθορίζεται η ακρίβεια των χαρακτηριστικών.

Όσο πιο μεγάλο είναι το εύρος τόσο μειώνεται η ακρίβεια των στατιστικών αποτελεσμάτων.

## Άσκηση 6

Ποιο είναι το βασικό πλεονέκτημα και ποιο το βασικό μειονέκτημα μιας ομαδοποιημένης κατανομής;

Απάντηση: Το βασικό πλεονέκτημα είναι ότι κερδίζουμε χώρο στη δημιουργία πινάκων (λιγότερες κατηγορίες). Αν π.χ. σε μία έρευνα με μεγάλο δείγμα έχει μετρηθεί η ηλικία των συμμετεχόντων, η παρουσίαση των συχνοτήτων της ηλικίας με μία ομαδοποιημένη κατανομή θα μας επιτρέψει να παρουσιάσουμε τα δεδομένα με πιο συνοπτικό τρόπο. Ωστόσο το βασικό μειονέκτημα είναι ότι μπορεί να χάνουμε πληροφορίες, αφού οι μεμονωμένες τιμές μέσα στα διαστήματα δεν παρουσιάζονται.

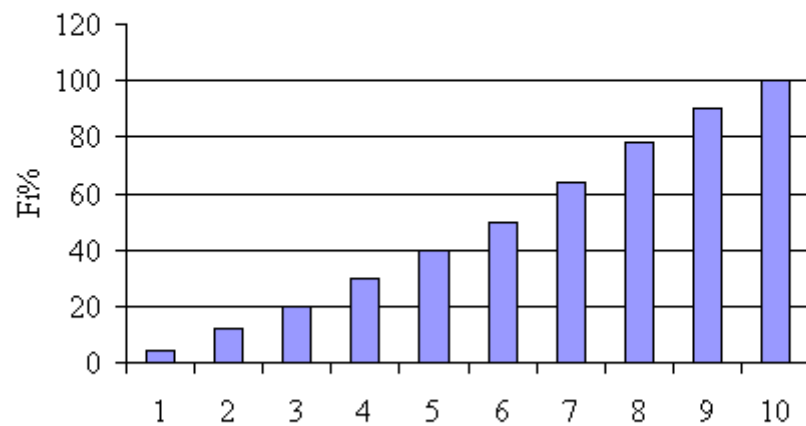
## Άσκηση 7

Πώς μπορούν να μας φανούν χρήσιμες οι πληροφορίες που παίρνουμε από τη σχετική αθροιστική συχνότητα (crf) μιας κατανομής;

Απάντηση: Υπάρχουν περιπτώσεις που μας ενδιαφέρει να γνωρίζουμε το ποσοστό ατόμων που έχουν ίδια ή μικρότερη επίδοση από μία συγκεκριμένη τιμή. Η αθροιστική συχνότητα μας δίνει αυτή την πληροφορία.

®

### διάγραμμα αθροιστικών συχνοτήτων



**επίδοση** 1→βαθμ.0, 2→βαθμ.1, ..., 10→βαθμ.9

## Άσκηση 8

Μία ερευνήτρια ενδιαφέρθηκε να μελετήσει την επίδραση που έχει το φύλο στη σχολική επίδοση. Για να μετρήσει τη σχολική επίδοση χρησιμοποίησε τη βαθμολογία κάθε μαθητή σε όλα τα μαθήματα. Με ποιον τρόπο θα αποδοθούν γραφικά οι δύο συγκεκριμένες μεταβλητές και γιατί;

Απάντηση: Για να παρουσιάσει τα δεδομένα αναφορικά με το φύλο μπορεί να χρησιμοποιήσει είτε ένα ακιδωτό διάγραμμα είτε ένα κυκλικό. Βέβαια σε περιπτώσει που η μεταβλητή λαμβάνει μόνο δύο τιμές, ένα διάγραμμα δεν βοηθάει περισσότερο από την λεκτική ανάλυση.

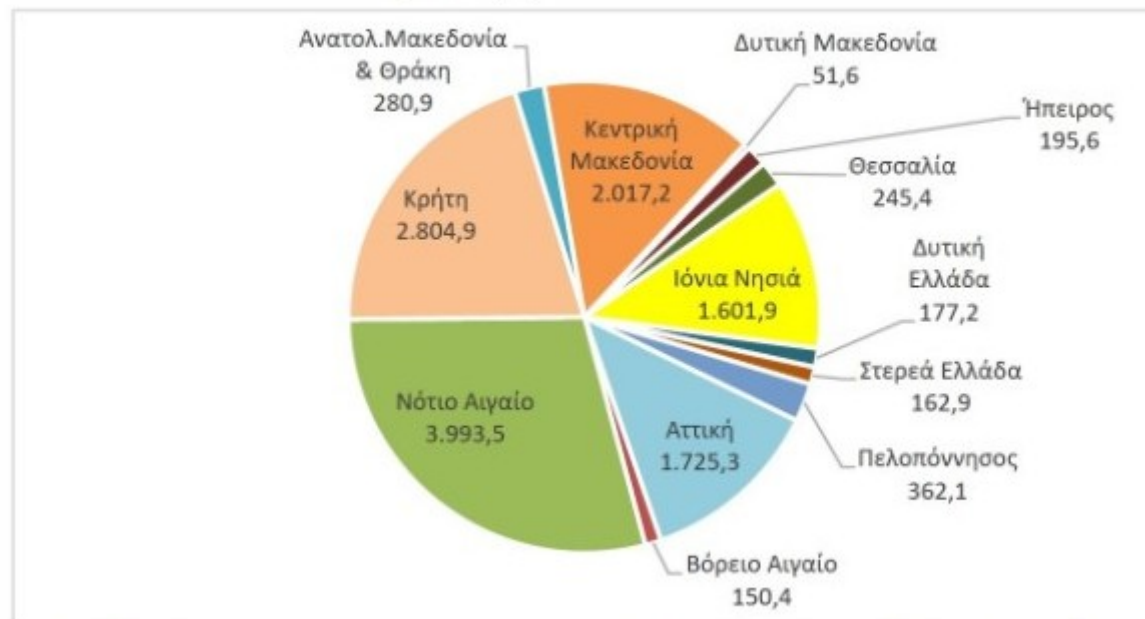
Για να παρουσιάσει τη σχολική επίδοση (μετρημένη με κλίμακα ίσως διαστημάτων) θα μπορούσε να χρησιμοποιήσει ένα ιστόγραμμα είτε ένα πολύγωνο συχνότητας.

## Άσκηση 9

Ένας ερευνητής ενδιαφέρεται να μελετήσει την επίδραση του επαγγέλματος στην ικανοποίηση από την εργασία. Μετρώντας τη μεταβλητή “επάγγελμα” διέκρινε 10 διαφορετικά επαγγέλματα. Για να αποδώσει γραφικά τα δεδομένα που συγκέντρωσε από 516 άτομα που συμμετείχαν στην έρευνα, αποφάσισε να χρησιμοποιήσει το κυκλικό διάγραμμα. Πιστεύετε ότι αυτός είναι ο πιο ενδεικτικός τρόπος παρουσίασης των δεδομένων αυτών;

Απάντηση: Το κυκλικό διάγραμμα δεν είναι ο καλύτερος τρόπος παρουσίασης, όταν υπάρχουν 10 τιμές, διότι είναι δυσανάγνωστος. Πολύ καλύτερα θα μπορούσαν να παρουσιαστούν τα αποτελέσματα με ένα ακιδωτό διάγραμμα (με 10 ράβδους) είτε οριζόντια είτε κάθετα.

**ΚΑΤΑΝΟΜΗ ΤΑΞΙΔΙΩΤΙΚΩΝ ΕΙΣΠΡΑΞΕΩΝ ΣΤΙΣ 13 ΠΕΡΙΦΕΡΕΙΕΣ**  
**Ιανουάριος-Σεπτέμβριος 2018**  
**(εκατ. Ευρώ)**



ΠΗΓΗ: Τράπεζα της Ελλάδος

Επεξεργασία στοιχείων PDI Infographics για το Euro2day.gr

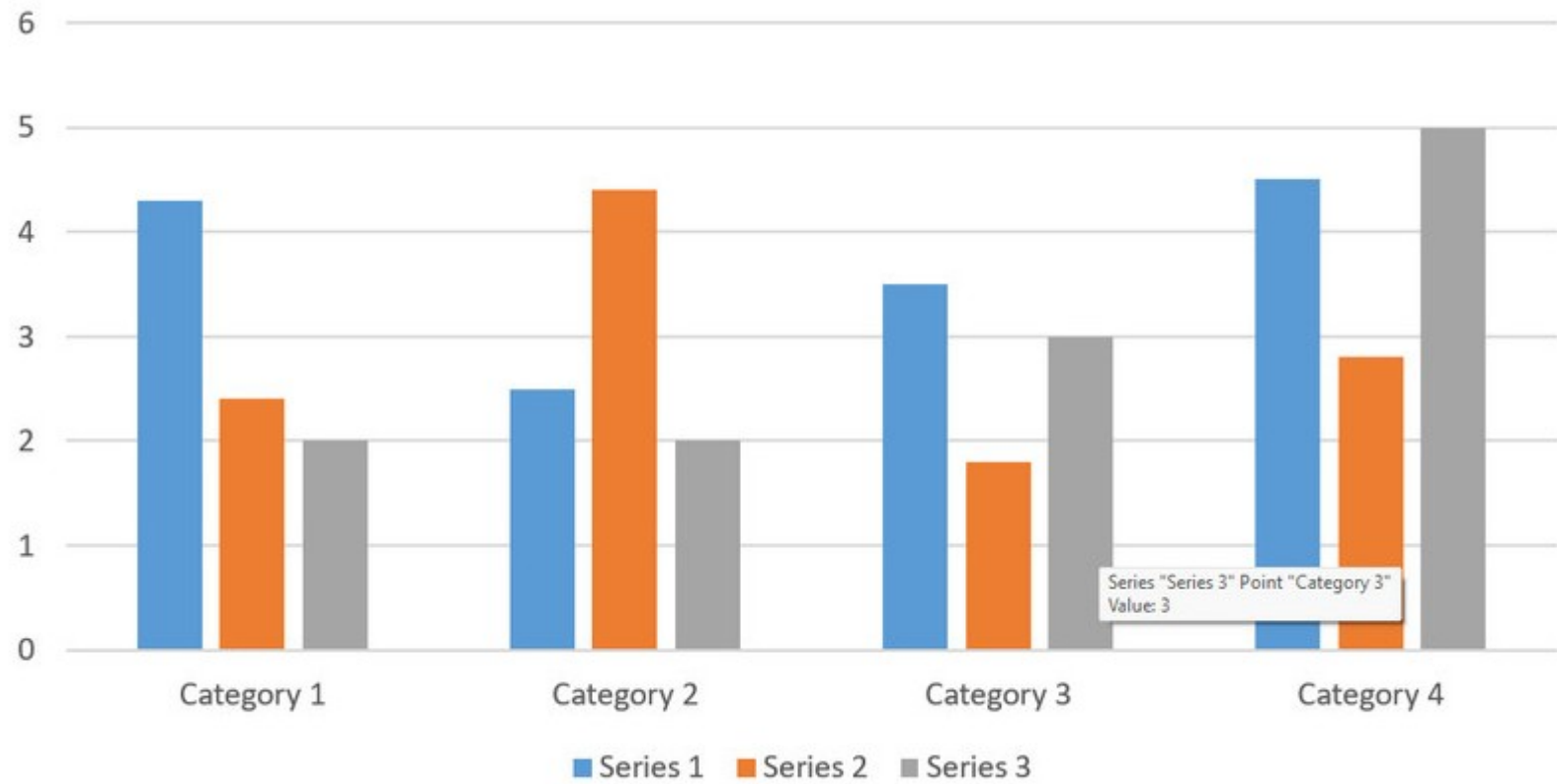
## Άσκηση 10

Μία ερευνήτρια θέλει να παρουσιάσει σε ένα επιστημονικό συνέδριο τη σύγκριση μεταξύ τριών μεταβλητών, χρησιμοποιώντας μία μόνο διαφάνεια λόγω έλλειψης χρόνου. Γι' αυτό το λόγο σχεδιάζει μία γραφική παράσταση στην οποία παρουσιάζει τρία πολύγωνα συχνότητας, ένα για κάθε μεταβλητή, χρησιμοποιώντας διαφορετικά χρώματα.

## Άσκηση 10

Στον οριζόντιο άξονα τοποθέτησε τις τιμές της κλίμακας με την οποία μέτρησε τις μεταβλητές (και οι τρεις με την ίδια κλίμακα, αλλά με διαφορετικά δείγματα π.χ. παιδιά, έφηβοι, ενήλικες) και στον κάθετο άξονα τοποθέτησε τις απόλυτες συχνότητες ( $f$ ). Θα σχολιάζατε κάτι για τον τρόπο με τον οποίο η ερευνήτρια σχεδίασε τη γραφική παράσταση;

Απάντηση: Όταν χρησιμοποιούμε ένα πολύγωνο συχνότητας (αλλά και ένα ιστόγραμμα συχνότητας) προκειμένου να παρουσιάσουμε συγκρητικά τα διαφορετικά επίπεδα μιας μεταβλητής (π.χ. διαφορετικές ηλικιακές ομάδες) ως προς μια άλλη μεταβλητή (την ίδια και για τα τρία επίπεδα) είναι προτιμότερο να χρησιμοποιούμε τις σχετικές συχνότητες αντί για τις απόλυτες συχνότητες ( $f$ ) ώστε να είναι δυνατή η σύγκρισή τους στην περίπτωση που οι ομάδες έχουν διαφορετικό αριθμός ατόμων.





ΤΕΛΟΣ