



ΣΤΑΤΙΣΤΙΚΗ (ΜΥ0202)

Μ.Ν. Ντυκέν,
Πανεπιστήμιο Θεσσαλίας
Τ.Μ.Χ.Π.Π.Α.

Βόλος, 2018-2019

ΔΙΑΛΕΞΗ 04

Περιεχόμενο της Διάλεξης

- (1) Περιγραφική Στατιστική: παρουσίαση δεδομένων, γραφικές παραστάσεις,
- (2) Γραφικές παραστάσεις ποιοτικών δεδομένων,
- (3) Γραφικές παραστάσεις ποσοτικών δεδομένων
- (4) Η πυραμίδα ηλικιών

***1. Περιγραφική Ανάλυση
Παρουσίαση Δεδομένων
Γραφικές Παραστάσεις***

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Μετά από την συλλογή των δεδομένων, η πρώτη και πολύ σημαντική επεξεργασία είναι η οργάνωση των δεδομένων με την διαμόρφωση των πινάκων και ειδικότερα, ο υπολογισμός των συχνοτήτων.

Οι γραφικές παραστάσεις συχνοτήτων αποτελούν το επόμενο χρήσιμο εργαλείο διότι επιτρέπει μια πρώτη αναζήτηση της μεταβλητότητας και των ιδιοτήτων μιας μεταβλητής.

Βασικοί κανόνες της γραφικής απεικόνισης των δεδομένων:

- Η επιλογή του τύπου γραφικής παράστασης (Γράφημα) εξαρτάται σε σημαντικό βαθμό από το τύπο μεταβλητής (ποιοτικής / ποσοτικής) και το πλήθος των δεδομένων.
- Το γράφημα πρέπει να είναι κατανοητό δηλαδή να διευκολύνει την κατανόηση του φαινομένου που εξετάζεται.
- Το γράφημα πρέπει να είναι σαφές και ακριβές (να μην παραπλάνια τον αναγνώστη)

1.1. Γραφικές Παραστάσεις Ποιοτικών μεταβλητών

Με Ποιοτικές Μεταβλητές, η μοναδική επιτρεπόμενη επεξεργασία είναι:

ο *υπολογισμός συχνοτήτων*

με βάση τις οποίες μπορούμε να παράγουμε

τις γραφικές παραστάσεις.

Δύο κλασικές μορφές γραφημάτων για ποιοτικές μεταβλητές:

- Κυκλικό διάγραμμα (pie chart)
- Ραβδόγραμμα / Διάγραμμα γραμμών (bar chart)

Οι δύο εναλλακτικές μορφές γραφημάτων απεικονίζουν συχνότητες (n_i) ή σχετικές συχνότητες (f_i) που αντιστοιχούν στις κατηγορίες της μεταβλητής.

Κυκλικό διάγραμμα (pie chart)

- Κατάλληλο για ονομαστική (nominal) μεταβλητή.
- Το σύνολο των κατηγοριών πρέπει να καλύψει το 100% των τιμών.
- Πρόκειται για κυκλικό δίσκο χωρισμένο σε k κυκλικούς τομείς, όπου k = αριθμός κατηγοριών της ονομαστικής μεταβλητής.
- Οι συχνότητες παριστάνονται με κυκλικούς τομείς.

Χρησιμοποιείται όταν:

- a) Έχουμε περιορισμένο αριθμό κατηγοριών (5 έως 7) για να διαβαστεί εύκολα.
- b) Υπάρχουν αποκλείσεις στις συχνότητες (οι συχνότητες των κατηγοριών διαφέρουν αρκετά μεταξύ τους). Αν όλες οι κατηγορίες έχουν περίπου την ίδια συχνότητα, το γράφημα δεν μας προσφέρει σημαντική πληροφορία.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

Ο υπολογισμένος πληθυσμός της Ελλάδας το 2017 ανέρχεται σε 10.768.193 κάτοικους. Η κατανομή τους κατά μεγάλες ομάδες χωρών γέννησης δίνεται στον ακόλουθο πίνακα.

Η μεταβλητή είναι ονομαστική (καμία ιεράρχηση).

Πίνακας 1: Πληθυσμός Ελλάδας κατά μεγάλες ομάδες χωρών γέννησης (2017)

| Χώρες | Αριθμός κατοίκων |
|-------------------------|-------------------|
| 1. Ελλάδα | 9.517.330 |
| 2. Ε.Ε. (έκτος Ελλάδας) | 345.619 |
| 3. Άλλες χώρες | 905.244 |
| Σύνολο | 10.768.193 |

Πηγή: ΕΛΣΤΑΤ, Υπολογισμένος Πληθυσμός

Μεγάλη διαφορά μεταξύ των 3 συχνοτήτων



Υπολογίζουμε τις σχετικές συχνοτήτες οι οποίες διαβάζονται ευκολότερα και παράγουμε το διάγραμμα

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

Πληθυσμός Ελλάδας κατά μεγάλες ομάδες χώρων γέννησης (2017)

| Χώρες | Αριθμός κατοίκων | Σχετικές συχνότητες |
|-------------------------|------------------|---------------------|
| 1. Ελλάδα | 9.517.330 | 88,4 |
| 2. Ε.Ε. (έκτος Ελλάδας) | 345.619 | 3,2 |
| 3. Άλλες χώρες | 905.244 | 8,4 |
| Σύνολο | 10.768.193 | 100 |

$$f_i = \frac{n_i}{n} \acute{\eta}$$

$$f_i = 100 \times \frac{n_i}{n} (\%)$$

Το **κυκλικό διάγραμμα** είναι ένας κύκλος που χωρίζεται σε κυκλικούς τομείς.

Κάθε κυκλικός τομέας αντιπροσωπεύει μια κατηγορία.

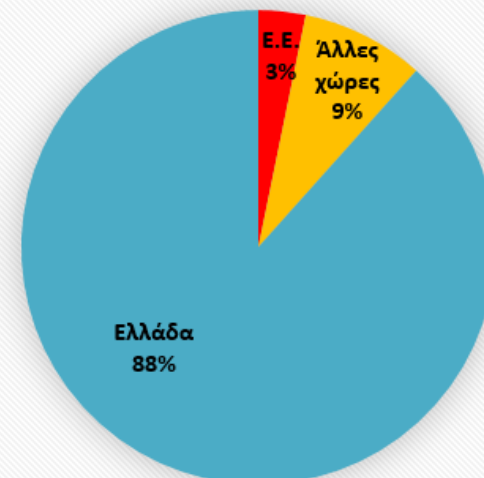
Κυκλικός τομέας $i = 360 \times \frac{n_i}{n} = 360 \times f_i$

Όπου f_i αντιστοιχεί στην απλή αναλογία της κατηγορίας i .

→ Για Ελλάδα ($i=1$) $f_1 \times 360^0 = 0,884 \times 360^0 = 318^0$

→ Για τις Άλλες ($i=3$) $f_3 \times 360^0 = 0,084 \times 360^0 = 30^0$

Πληθυσμός Ελλάδας κατά ομάδες Χώρων Γέννησης



Πηγή: ΕΛΣΤΑΤ, Υπολογισμένος Πληθυσμός

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

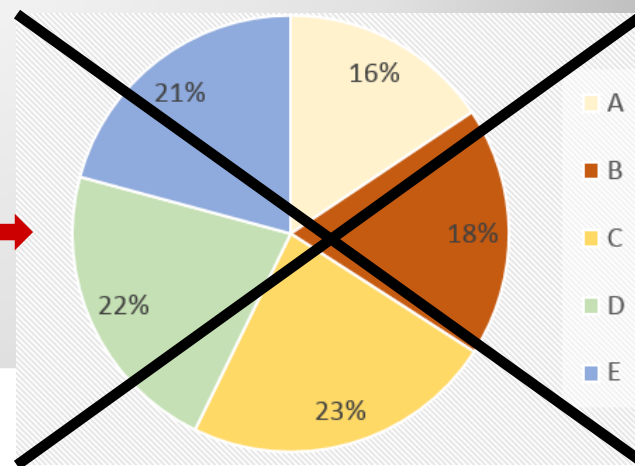
Σε ποιες περιπτώσεις το κυκλικό διάγραμμα **ΔΕΝ** πρέπει να χρησιμοποιηθεί;

- Όταν ο αριθμός των κατηγοριών είναι πολύ μεγάλος.

[Π.χ. Κυκλικό διάγραμμα με 12 κατηγορίες δεν διαβάζεται!]

- Όταν δεν υπάρχουν πραγματικές αποκλίσεις μεταξύ των συχνοτήτων.

Αν όλες οι κατηγορίες έχουν την ίδια ή σχεδόν συχνότητα, η γραφική αναπαράσταση δεν μας προσφέρει σημαντική πληροφορία: οι μικρές διαφορές δεν εμφανίζονται καλά.



Ραβδόγραμμα (bar chart)

- Κατάλληλο για ονομαστική (nominal) και τακτική (ordinal) μεταβλητή.
- Το σύνολο των κατηγοριών πρέπει να καλύψει το 100% των τιμών.
- Οι συχνότητες παριστάνονται με ορθογώνια: το ύψος των ορθογωνίων απεικονίζει τη συχνότητα της κατηγορίας. Τα ορθογώνια έχουν ίδιο μήκος το οποίο επιλέγουμε αυθαίρετα όμως με τρόπο που διευκολύνει την αναγνώριση του διαγράμματος.

Χρησιμοποιείται όταν:

a) Έχουμε σχετικά μεγάλο αριθμό κατηγοριών.

➔ Επιτρέπει τη διάκριση μεταξύ συχνοτήτων που δεν διαφέρουν σημαντικά.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

Με βάση την Έρευνα Εργατικού Δυναμικού (ΕΛΣΤΑΤ), ο πληθυσμός ηλικίας 15 ετών και άνω ανέρχεται σε 9.212,8 χιλιάδες ατόμων. Η κατανομή τους κατά επίπεδο εκπαίδευσης δίνεται στον ακόλουθο πίνακα.

Η μεταβλητή είναι τακτική (ordinal) & αποτελείται από 6 κατηγορίες.

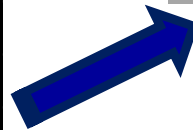
Πίνακας 2: Πληθυσμός ηλικίας 15+ κατά επίπεδο εκπαίδευσης (2016)

| Επίπεδο εκπαίδευσης | Συχνότητα |
|------------------------------|-----------|
| Δεν πήγε καθόλου σχολείο (*) | 375,3 |
| Πρωτοβάθμια εκπαίδευση | 2.095,9 |
| Κατώτερη δευτεροβάθμια | 1.108,5 |
| Ανώτερη δευτεροβάθμια | 3.574,4 |
| Πτυχίο ΑΕΙ / ΤΕΙ | 1.847,7 |
| Μεταπτυχιακός τίτλος | 211,1 |
| Σύνολο πληθυσμού ηλικίας 15+ | 9.212,8 |

(*) Περιλαμβάνει τα άτομα με λίγες τάξεις δημοτικού

Πηγή: ΕΛΣΤΑΤ, Έρευνα Εργατικού Δυναμικού

Μεγάλη διαφορά μεταξύ των συχνοτήτων



Υπολογίζουμε τις σχετικές συχνότητες οι οποίες διαβάζονται ευκολότερα και παράγουμε το διάγραμμα

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβολές

Πληθυσμός ηλικίας 15+ κατά επίπεδο εκπαίδευσης (2016)

| Επίπεδο εκπαίδευσης | Συχνότητα | Σχετική συχνότητα |
|-------------------------------------|----------------|-------------------|
| Δεν πήγε καθόλου σχολείο (*) | 375,3 | 4,1 |
| Πρωτοβάθμια εκπαίδευση | 2.095,9 | 22,7 |
| Κατώτερη δευτεροβάθμια | 1.108,5 | 12,0 |
| Ανώτερη δευτεροβάθμια | 3.574,4 | 38,8 |
| Πτυχίο ΑΕΙ / ΤΕΙ | 1.847,7 | 20,1 |
| Μεταπτυχιακός τίτλος | 211,1 | 2,3 |
| Σύνολο πληθυσμού ηλικίας 15+ | 9.212,8 | 100,0 |

(*) Περιλαμβάνει τα άτομα με λίγες τάξεις δημοτικού

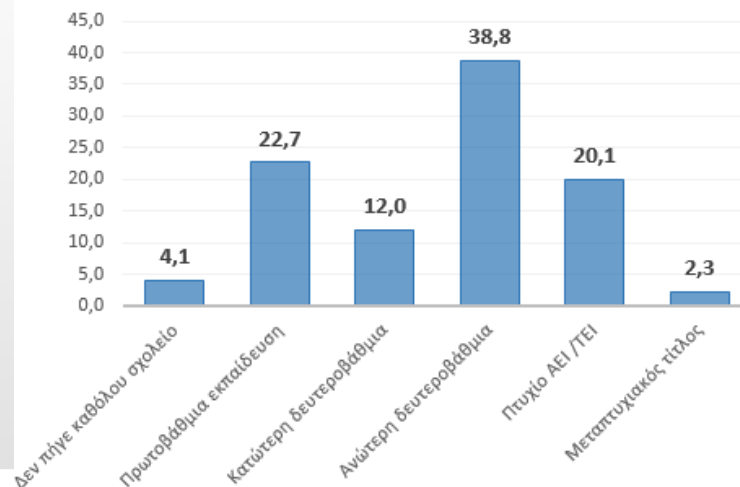
Διακρίνεται εύκολα η διαφορά μεταξύ του 1^{ου} και 6^{ου} επιπέδου εκπαίδευσης

Το **ραβδόγραμμα** αποτελείται από ορθογώνια παραλληλόγραμμα που αντανακλούν τη συχνότητα κάθε κατηγορίας.

Στον οριζόντιο άξονα, η επιλογή της βάσης είναι αυθαίρετη (πλάτος των παραλληλογράμμων).

Ο κάθετος άξονας απεικονίζει τις συχνότητες (τα ύψη των παραλληλογράμμων είναι ανάλογα με τις συχνότητες).

Επίπεδο Εκπαίδευσης του Πληθυσμού ηλικίας 15+ (2016)



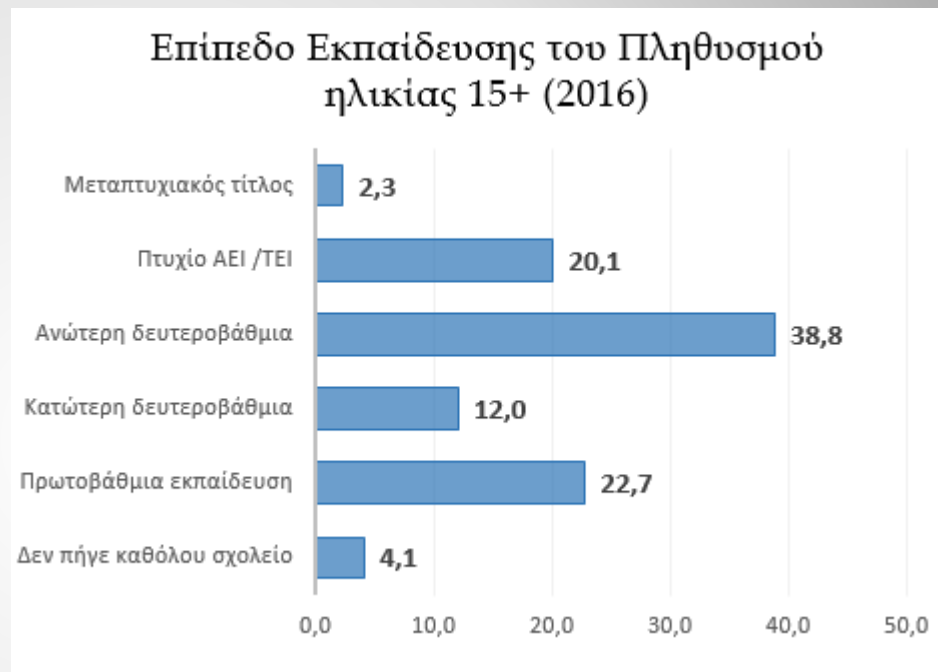
ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

Το **ραβδόγραμμα** μπορεί να σχεδιαστεί διαφορετικά.

Στον κάθετο άξονα, εμφανίζονται οι κατηγορίες της μεταβλητής.

Στον οριζόντιο άξονα, εμφανίζονται οι συχνότητες (το μέγεθος των παραλληλογράμμων είναι ανάλογα με τις συχνότητες).



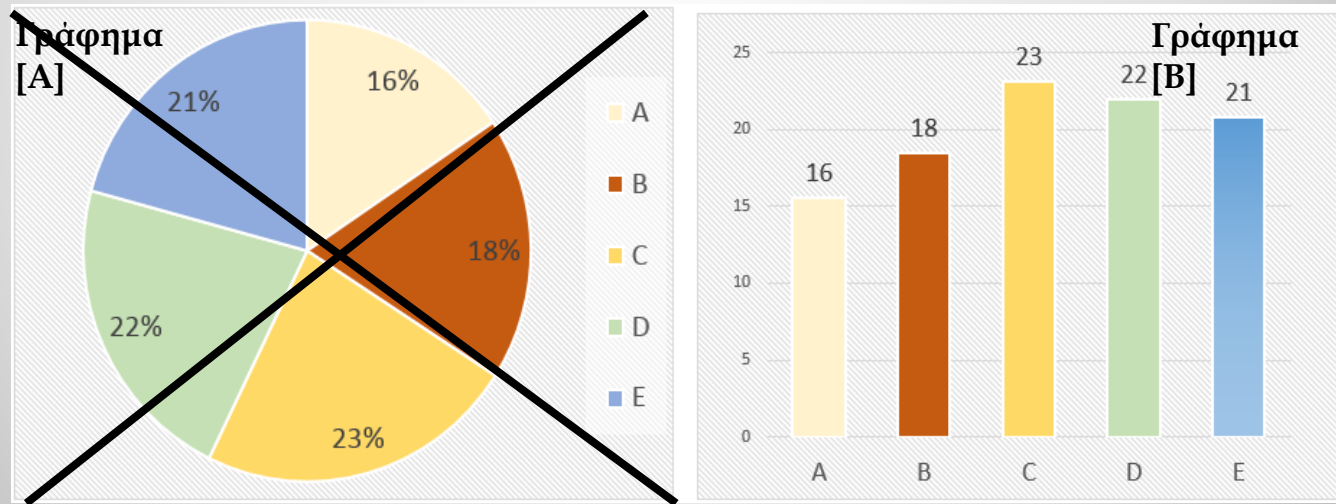
Η επιλογή είναι δική σας!

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβλητές

Σε ποιες περιπτώσεις μπορούμε να χρησιμοποιούμε το ραβδόγραμμα;

- ❑ Ο αριθμός των κατηγοριών μπορεί να είναι αρκετά μεγάλος σε αντίθεση με το κυκλικό διάγραμμα.
- ❑ Με το ραβδόγραμμα, μπορούμε να διακρίνουμε πιο εύκολα τις μικρές διαφορές μεταξύ των συχνοτήτων σε σχέση με το κυκλικό διάγραμμα (βλέπε γράφημα [B]).



- ❑ Με το ραβδόγραμμα, είναι ευκολότερο να κάνουμε συγκρίσεις ανάμεσα σε 2 ή περισσότερες μεταβλητές.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.1 Ποσοτικές Μεταβολές

Σύγκριση επίπεδου εκπαίδευσης του πληθυσμού ηλικίας 15+ μεταξύ του 2001 και του 2016

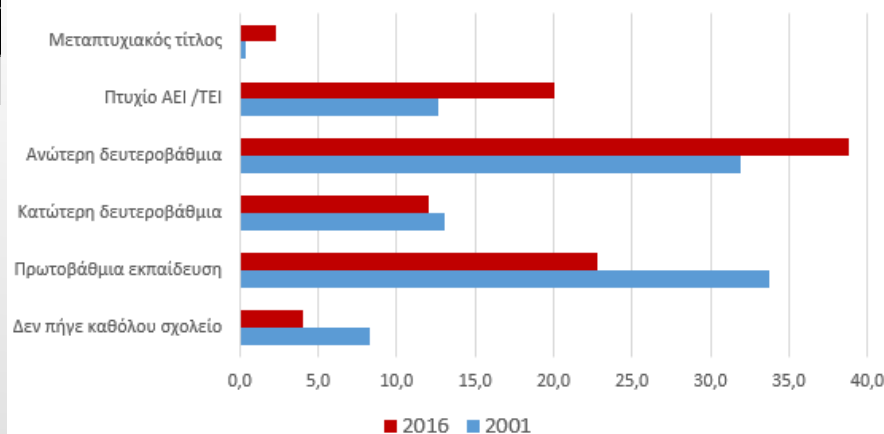
Πληθυσμός ηλικίας 15+ κατά επίπεδο εκπαίδευσης (%)
2001 και 2016

| Επίπεδο εκπαίδευσης | 2001 | 2016 |
|-------------------------------------|--------------|--------------|
| Δεν πήγε καθόλου σχολείο | 8,3 | 4,1 |
| Πρωτοβάθμια εκπαίδευση | 33,8 | 22,7 |
| Κατώτερη δευτεροβάθμια | 13,1 | 12,0 |
| Ανώτερη δευτεροβάθμια | 31,9 | 38,8 |
| Πτυχίο ΑΕΙ /ΤΕΙ | 12,6 | 20,1 |
| Μεταπτυχιακός τίτλος | 0,3 | 2,3 |
| Σύνολο πληθυσμού ηλικίας 15+ | 100,0 | 100,0 |

(*) Περιλαμβάνει τα άτομα με λίγες τάξεις δημοτικού

Για κάθε έτος, υπολογίσαμε τις σχετικές συχνότητες (%)

Επίπεδο εκπαίδευσης πληθυσμού ηλικίας 15+,
το 2001 και το 2016



Με το ραβδόγραμμα, η συγκριτική ανάλυση είναι εύκολη. Σε διάρκεια 15 ετών, εμφανίζεται σημαντική βελτίωση του μορφωτικού επιπέδου των Ελλήνων.

1.2. Γραφικές Παραστάσεις Ποσοτικών μεταβλητών

Με Ποσοτικές Μεταβλητές, τα δεδομένα είναι είτε ατομικά είτε ομαδοποιημένα σε κλάσεις (τάξεις / ομάδες). Υπάρχουν πολλαπλές μορφές περιγραφικής / διερευνητικής ανάλυσης.

Όμως οι γραφικές παραστάσεις παραμένουν χρήσιμο προκαταρκτικό εργαλείο ανάλυσης.

Συνηθεις μορφές γραφημάτων:

- Ιστόγραμμα (Histogram): αν τα δεδομένα είναι ατομικά, πρέπει να γίνει κατασκευή των τάξεων.
- Πολυγωνικές γραμμές.
- Διάγραμμα γραμμών για χρονολογικές σειρές [Π.χ. κκΑΕΠ 2000-2014].

Με Ποσοτικές Μεταβλητές, θα πρέπει να γίνει:

- Κατάταξη των τιμών σε αύξουσα σειρά.
- Υπολογισμός του εύρους των τιμών: μέγιστη και ελάχιστη τιμή.

Όταν έχουμε ατομικά δεδομένα:

- Διαμέριση του εύρους των τιμών σε ένα σχετικό μικρό αριθμό k κλάσεων (ομάδων): $i = 1, \dots, k$
- Προσδιορισμός απόλυτων ή/και σχετικών συχνοτήτων:
 - ποιό το πλήθος ατόμων σε κάθε κλάση i ;
 - Απόλυτη ή σχετική συχνότητα;
 - Οι κλάσεις έχουν ή δεν έχουν το ίδιο πλάτος; (2 διαφορετικές επεξεργασίες)

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Ο υπολογισμένος πληθυσμός της Ελλάδας το 2017 ανέρχεται σε 10.768.193 κάτοικους. Η κατανομή τους κατά τάξεις ηλικιών δίνεται στον ακόλουθο πίνακα.

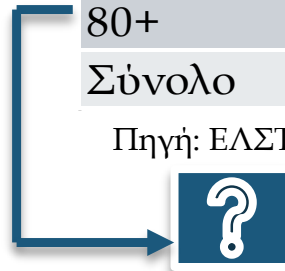
Η μεταβλητή είναι ποσοτική (με ομαδοποίηση σε τάξεις).

Πίνακας 3: Πληθυσμός Ελλάδας κατά τάξεις ηλικιών (2017)

| Ηλικία | Αριθμός ατόμων | f_i |
|---------|----------------|-------|
| 0 - 19 | 2.092.696 | 19,4 |
| 20 - 39 | 2.606.636 | 24,2 |
| 40 - 59 | 3.095.893 | 28,8 |
| 60 - 79 | 2.249.507 | 20,9 |
| 80+ | 723.461 | 6,7 |
| Σύνολο | 10.768.193 | 100,0 |

Τα δεδομένα είναι ήδη ομαδοποιημένα σε 5 τάξεις. Όλες οι τάξεις έχουν το ίδιο εύρος με εξαίρεση την τελευταία (80+).

Πηγή: ΕΛΣΤΑΤ, Υπολογισμένος Πληθυσμός



ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Πίνακας 3: Πληθυσμός Ελλάδας κατά τάξεις ηλικιών (2017)

| Ηλικία | Αριθμός ατόμων | f_i |
|---------|----------------|-------|
| 0 - 19 | 2.092.696 | 19,4 |
| 20 - 39 | 2.606.636 | 24,2 |
| 40 - 59 | 3.095.893 | 28,8 |
| 60 - 79 | 2.249.507 | 20,9 |
| 80+ | 723.461 | 6,7 |
| Σύνολο | 10.768.193 | 100,0 |

Δεδομένου ότι ο πληθυσμός 80 ετών και άνω ανέρχεται σε 6,7% του συνολικού πληθυσμού, θεωρούμε ότι η τελευταία κλάση = 80-99

Ύψος παραλληλογράμμων ανάλογα με τις συχνότητες



Ίδιο πλάτος στον οριζόντιο άξονα = 20 έτη

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Με το ιστόγραμμα, μπορούμε να χρησιμοποιούμε περισσότερες από μια μεταβλητή για να γίνει συγκριτική ανάλυση.

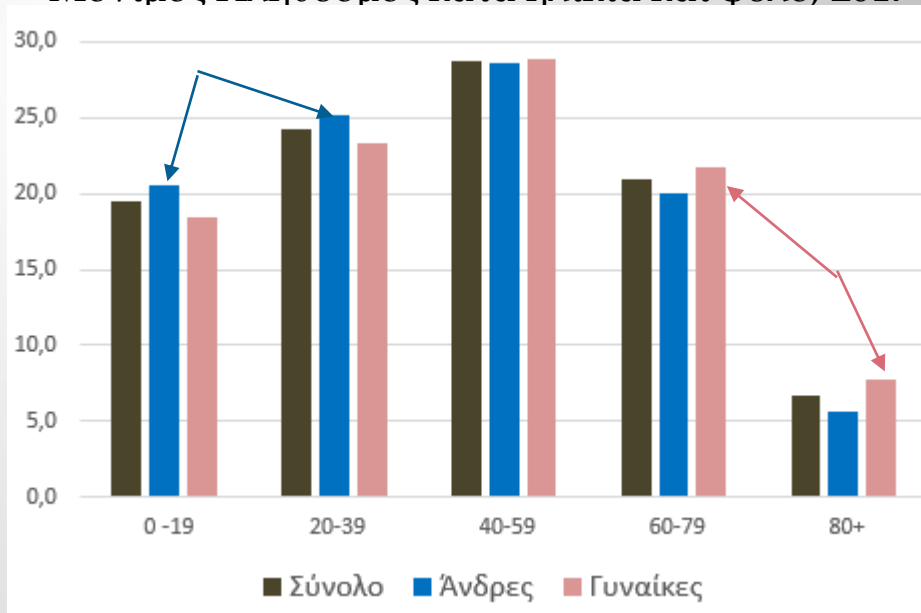
Μπορούμε, για παράδειγμα, να συγκρίνουμε την κατανομή του πληθυσμού κατά ηλικία μεταξύ ανδρών και γυναικών.

Εφόσον οι τάξεις ηλικίας είναι ίσες, μπορεί να γίνει αναπαράσταση της κατανομής του μόνιμου πληθυσμού κατά ηλικία και φύλο στο ίδιο διάγραμμα έτσι ώστε να εμφανιστούν οι ενδεχόμενες διαφορές μεταξύ ανδρών και γυναικών.

| Ηλικία | Σχετικές Συχνότητες | | |
|----------------|---------------------|--------|----------|
| Ηλικία | Σύνολο | Άνδρες | Γυναίκες |
| 0 - 19 | 19,4 | 20,6 | 18,4 |
| 20 - 39 | 24,2 | 25,1 | 23,3 |
| 40 - 59 | 28,8 | 28,6 | 28,8 |
| 60 - 79 | 20,9 | 20,0 | 21,8 |
| 80 - 99 | 6,7 | 5,7 | 7,7 |
| Σύνολο | 100,0 | 100,0 | 100,0 |

Πηγή: ΕΛΣΤΑΤ, Υπολογισμένος Πληθυσμός, 2017

Μόνιμος Πληθυσμός κατά ηλικία και φύλο, 2017



Περισσότεροι άνδρες στις νέες ηλικίες,
περισσότερες γυναίκες στις μεγάλες ηλικίες

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Σύμφωνα με τα αποτελέσματα της Απογραφής Πληθυσμού και Κατοικιών (2011), ο αριθμός κατοικιών στην Ελλάδα ανέρχεται σε 4.122.088 κατοικίες. Η κατανομή των κατοικιών με βάση την πυκνότητά τους δίνεται στον ακόλουθο πίνακα.

Η μεταβλητή είναι ποσοτική (με ομαδοποίηση σε τάξεις διαφορετικού εύρους).

Πίνακας 4: Κατοικούμενες κατοικίες κατά πυκνότητα κατοικήσεως (m² ανά κάτοικο)

| i | Πυκνότητα (m ² ανά κάτοικο) | Αριθμός κατοικιών | f _i |
|---|--|-------------------|----------------|
| 1 | < 10 | 26.012 | 0,6 |
| 2 | [10-20) | 465.944 | 11,3 |
| 3 | [20-30) | 988.324 | 24,0 |
| 4 | [30-40) | 801.123 | 19,4 |
| 5 | [40-60) | 962.467 | 23,3 |
| 6 | [60-80) | 438.967 | 10,6 |
| 7 | 80+ | 439.251 | 10,7 |
| | Σύνολο | 4.122.088 | 100,0 |

Τα δεδομένα είναι ήδη ομαδοποιημένα σε 7 τάξεις.

Όμως οι 7 τάξεις δεν έχουν το ίδιο εύρος.

Πηγή: ΕΛΣΤΑΤ, Απογραφή 2011

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Πίνακας 4: Κατοικούμενες κατοικίες κατά πυκνότητα κατοικήσεως (m² ανά κάτοικο)

| i | Πυκνότητα (m ² ανά κάτοικο) | Αριθμός κατοικιών | f _i |
|---|--|-------------------|----------------|
| 1 | < 10 | 26.012 | 0,6 |
| 2 | [10-20) | 465.944 | 11,3 |
| 3 | [20-30) | 988.324 | 24,0 |
| 4 | [30-40) | 801.123 | 19,4 |
| 5 | [40-60) | 962.467 | 23,3 |
| 6 | [60-80) | 438.967 | 10,6 |
| 7 | 80+ | 439.251 | 10,7 |
| | Σύνολο | 4.122.088 | 100,0 |

Οι κλάσεις δεν έχουν όλες το ίδιο πλάτος.

1. Επιλογή εύρους αναφοράς στον οριζόντιο άξονα [π.χ. 10 όπως οι 4 πρώτες κλάσεις]
2. Στις 4 πρώτες τάξεις, αντιστοιχεί **1 μονάδα** για τη βάση του παραλληλόγραμμου.
3. Στην 5^η και 6^η τάξεις, έχουμε **2 μονάδες** (2 x 10) για τη βάση.
4. Στην 7^η τάξη, εξαρτάται από τον προσδιορισμό του άνω ορίου. Θέτοντας π.χ. για την 7^η τάξη: 80-140, έχουμε **6 μονάδες**.

Κατά συνέπεια, υπολογίζουμε τις **διορθωμένες συχνότητες**, λαμβάνοντας υπόψη το πλάτος της κάθε κατηγορίας:

$$f_i^* = f_i / \text{μονάδες} \longrightarrow f_5^* = f_5 / 2 = 23,3 / 2 = 11,6$$

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Κατοικούμενες κατοικίες κατά πυκνότητα κατοικήσεως (m^2 ανά κάτοικο)

| i | Πυκνότητα | n_i | f_i | Μονάδες | f^*_i |
|---|------------------|-----------|-------|---------|---------|
| 1 | < 10 | 26.012 | 0,6 | 1 | 0,6 |
| 2 | [10-20) | 465.944 | 11,3 | 1 | 11,3 |
| 3 | [20-30) | 988.324 | 24,0 | 1 | 24,0 |
| 4 | [30-40) | 801.123 | 19,4 | 1 | 19,4 |
| 5 | [40-60) | 962.467 | 23,3 | 2 | 11,7 |
| 6 | [60-80) | 438.967 | 10,6 | 2 | 5,3 |
| 7 | [80 -140) | 439.251 | 10,7 | 6 | 1,8 |
| | Σύνολο | 4.122.088 | 100,0 | | |

1^η τάξη : 0 έως 10, εύρος = 10 (1 μονάδα)

2^η τάξη : 10 έως 20, εύρος = 10 (1 μονάδα)

...

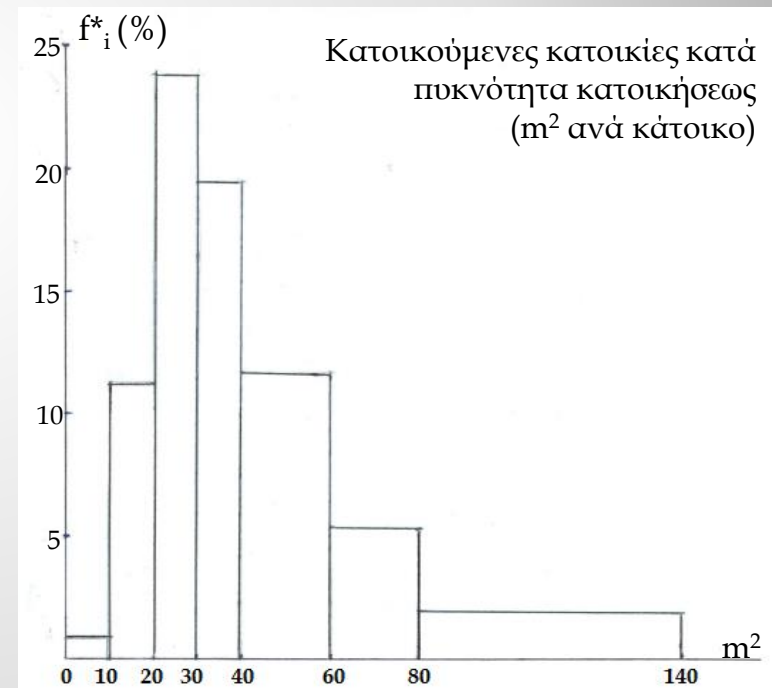
5^η τάξη : 40 έως 60, εύρος = 20 (2 μονάδες)

...

7^η τάξη : 80 έως 140, εύρος = 60 (6 μονάδες)

1.2. Ποσοτικές Μεταβλητές

- Τα ύψη των ορθογωνίων είναι ανάλογα με τις διορθωμένες συχνότητες.
- Οι τάξεις δεν έχουν το ίδιο πλάτος.



Ποσοτική Μεταβλητή με ατομικά δεδομένα και σχετικά μεγάλο πλήθος τιμών.

Βήματα κατασκευής ιστογράμματος με τάξεις ίσου πλάτους

1. Υπολογίζεται το εύρος του δείγματος: $e = X_{\max} - X_{\min}$
2. Πλάτος των τάξεων = εύρος / πλήθος των ομάδων που πρόκειται να δημιουργηθούν
→ Πλάτος = e / k όπου k = πλήθος ομάδων
3. Κριτήρια επιλογής του αριθμού τάξεων:
 - ▶ **Κριτήριο της ομοιογένειας:** πρέπει να φτιάξουμε αρκετές τάξεις έτσι ώστε μέσα στην ίδια τάξη, οι τιμές της μεταβλητής να μη διαφέρουν πολύ, δηλαδή να μη υπάρχει μεγάλη διασπορά τιμών μέσα στην τάξη.
 - ▶ **Κριτήριο της απλότητας:** δεν πρέπει όμως να κατασκευάζουμε πάρα πολλές τάξεις, για να έχουμε ένα σχετικά μικρό σε διαστάσεις πίνακα.

Εφόσον τα δυο κριτήρια αλληλοσυγκρούονται, τότε πρέπει να βρούμε μια ενδιάμεση λύση, ούτε λίγες τάξεις - ομάδες, ούτε πολλές. Σύμφωνα με τα δεδομένα και το εύρος των τιμών, ένας αριθμός από 5 έως 15 θεωρείται συνήθως αρκετός.

ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ – ΠΟΣΟΤΙΚΕΣ ΜΕΤΑΒΛΗΤΕΣ [05]

4. Υπάρχει **εμπειρική διαδικασία** για να επιλέγουμε τον αριθμό ομάδων, παίρνοντας ως αριθμό ομάδων k :

$$k = 1 + 3,3 \log_{10}(n) \text{ όπου } n = \text{πλήθος των παρατηρήσεων}$$
$$\log_{10}(n) = \text{λογάριθμος (δεκαδικός λογάριθμος)}^{(*)}$$

Ο αριθμός k αποτελεί ενδεικτικό αριθμός τάξεων.

5. Κατασκευάζεται ο πίνακας συχνοτήτων: Υπολογίζονται η συχνότητα και η σχετική συχνότητα κάθε τάξης (αριθμός παρατηρήσεων και ποσοστό)
6. Διαμόρφωση του ιστογράμματος:
- ο στον οριζόντιο άξονα έχουμε τα άκρα των τάξεων και
 - ο στον κατακόρυφο, οι συχνότητες (απόλυτες ή σχετικές)

(*) Βλέπε αποτελέσματα της συνάρτησης $\log_{10}(n)$ με τη χρήση του Excel

ΠΙΝΑΚΕΣ ΜΕ ΤΙΜΕΣ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ : $\log_{10}(n)$

| Αριθμός παρατηρήσεων | $\log_{10}(n)$ |
|----------------------|----------------|
| n | |
| 10 | 1,000 |
| 11 | 1,041 |
| 12 | 1,079 |
| 13 | 1,114 |
| 14 | 1,146 |
| 15 | 1,176 |
| 16 | 1,204 |
| 17 | 1,230 |
| 18 | 1,255 |
| 19 | 1,279 |
| 20 | 1,301 |
| 21 | 1,322 |
| 22 | 1,342 |
| 23 | 1,362 |
| 24 | 1,380 |
| 25 | 1,398 |

| Αριθμός παρατηρήσεων | $\log_{10}(n)$ |
|----------------------|----------------|
| n | |
| 26 | 1,415 |
| 27 | 1,431 |
| 28 | 1,447 |
| 29 | 1,462 |
| 30 | 1,477 |
| 31 | 1,491 |
| 32 | 1,505 |
| 33 | 1,519 |
| 34 | 1,531 |
| 35 | 1,544 |
| 36 | 1,556 |
| 37 | 1,568 |
| 38 | 1,580 |
| 39 | 1,591 |
| 40 | 1,602 |
| 41 | 1,613 |

| Αριθμός παρατηρήσεων | $\log_{10}(n)$ |
|----------------------|----------------|
| n | |
| 42 | 1,623 |
| 43 | 1,633 |
| 44 | 1,643 |
| 45 | 1,653 |
| 46 | 1,663 |
| 47 | 1,672 |
| 48 | 1,681 |
| 49 | 1,690 |
| 50 | 1,699 |
| 51 | 1,708 |
| 52 | 1,716 |
| 53 | 1,724 |
| 54 | 1,732 |
| 55 | 1,740 |
| 60 | 1,778 |
| 70 | 1,845 |

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Εξετάζουμε το μέγεθος (αριθμός κατοίκων) σε 20 χωριά. Η καταγραφή μας έδωσε τα ακόλουθα αποτελέσματα:

| Χωριά | Αριθμός κατοικιών σε κάθε χωριό |
|-------|------------------------------------|
| 1 | 170 |
| 2 | 120 |
| 3 | 110 |
| 4 | 135 |
| 5 | 150 |
| 6 | 50 |
| 7 | 105 |
| 8 | 130 |
| 9 | 160 |
| 10 | 125 |
| 11 | 85 |
| 12 | 90 |
| 13 | 160 |
| 14 | 145 |
| 15 | 120 |
| 16 | 130 |
| 17 | 70 |
| 18 | 155 |
| 19 | 190 |
| 20 | 100 |

Με βάση τον πίνακα, έχουμε:

- Πλήθος = $n = 20$ (Χωριά)
- Εύρος = $e = 190 - 50 = 140$ (max - min)
- Αριθμός τάξεων
 $k = 1 + 3,3 \cdot \log_{10}(n) = 1 + 3,3 \cdot \log_{10}(20) = 5,3 \rightarrow k = 5$
- Πλάτος τάξης = $e / k = 140 / 5 = 28 \rightarrow$ κατά προσέγγιση, **πλάτος = 30**
- Η πρώτη τάξη = $[50 - 80[$, η 2^η τάξη = $[80 - 110[$ κ.ά.

$\log_{10}(20) = 1,301 \rightarrow$ στο Excel χρησιμοποιούμε τη συνάρτηση : =LOG10(20)

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Ομαδοποίηση των 20 χωριών σε 5 τάξεις

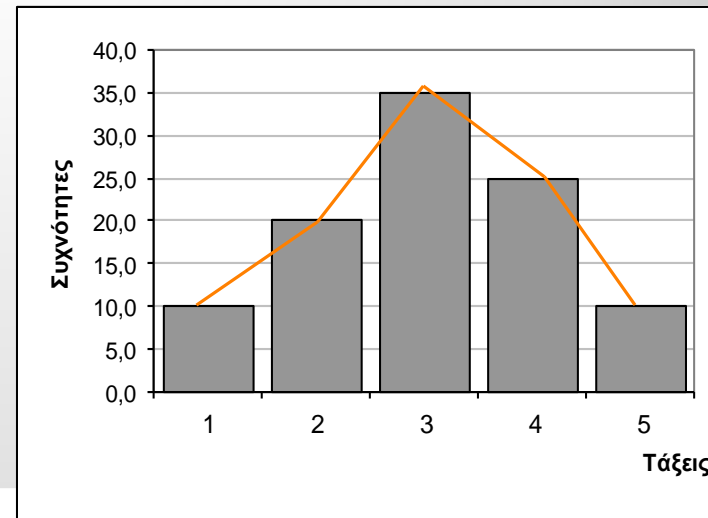
| I | τάξεις | Κέντρο της τάξης x_i | Αριθμός χωριών n_i | % f_i | % F_i |
|---|-----------|---------------------------|-------------------------|------------|------------|
| 1 | [50 - 80) | 65 | 2 | 10,0 | 10,0 |
| 2 | [80 -110) | 95 | 4 | 20,0 | 30,0 |
| 3 | [110-140) | 125 | 7 | 35,0 | 65,0 |
| 4 | [140-170) | 155 | 5 | 25,0 | 90,0 |
| 5 | [170-200) | 185 | 2 | 10,0 | 100,0 |
| | | | 20 | 100,0 | |

Οι 5 τάξεις έχουν το ίδιο πλάτος, η οπτική παρουσίαση είναι απλή.

Τι σημαίνει [50 - 80) ;

Η τιμή 50 περιλαμβάνεται στην 1^η τάξη ενώ η τιμή 80 δεν περιλαμβάνεται.

Η γραμμή που ενώνει τις κορυφές του ιστογράμματος είναι το **πολύγωνο συχνοτήτων**.



ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.2. Ποσοτικές Μεταβλητές

Ομαδοποίηση 31 Χώρων της Ευρώπης με βάση το κίνδυνο φτώχειας και κοινωνικού αποκλεισμού

| | Χώρες | % Πληθυσμού σε κίνδυνο φτώχειας και αποκλεισμού |
|----|--------------------|---|
| 1 | Βέλγιο | 20,3 |
| 2 | Βουλγαρία | 38,9 |
| 3 | Τσεχική Δημοκρατία | 12,2 |
| 4 | Δανία | 17,2 |
| 5 | Γερμανία | 19,0 |
| 6 | Εσθονία | 23,4 |
| 7 | Ιρλανδία | 24,2 |
| 8 | Ελλάδα | 34,8 |
| 9 | Ισπανία | 26,6 |
| 10 | Γαλλία | 17,1 |
| 11 | Κροατία | 27,9 |
| 12 | Ιταλία | 28,9 |
| 13 | Κύπρος | 25,2 |
| 14 | Λετονία | 28,2 |
| 15 | Λιθουανία | 29,6 |
| 16 | Λουξεμβούργο | 21,5 |
| 17 | Ουγγαρία | 25,6 |
| 18 | Μάλτα | 19,2 |
| 19 | Κάτω Χώρες | 17,0 |
| 20 | Αυστρία | 18,1 |
| 21 | Πολωνία | 19,5 |
| 22 | Πορτογαλία | 23,3 |
| 23 | Ρουμανία | 35,7 |
| 24 | Σλοβενία | 17,1 |
| 25 | Σλοβακία | 16,3 |
| 26 | Φινλανδία | 15,7 |
| 27 | Σουηδία | 17,7 |
| 28 | Ηνωμένο Βασίλειο | 22,2 |
| 29 | Ισλανδία | 12,2 |
| 30 | Νορβηγία | 16,1 |
| 31 | Ελβετία | 17,8 |

Με βάση τον πίνακα, έχουμε:

- Πλήθος = $n = \dots$ (Χώρες)
- Εύρος = $e = \dots = \dots$ (max - min)
- Αριθμός τάξεων:

$$k = 1 + 3,3 \cdot \log_{10}(n) = \dots = \dots$$

→ $k = \dots$

Πλάτος τάξης = $e / k = \dots / \dots = \dots$

→ κατά προσέγγιση, **πλάτος = ...**

○ Η 1^η τάξη = [... - ...],

η 2^η τάξη = [... - ...] κ.ά.

Πηγή: Eurostat

1.3. Ειδική Γραφική Παράσταση: η Πυραμίδα ηλικιών

Η Πυραμίδα των Ηλικιών αποτελεί ειδική γραφική παράσταση η οποία απεικονίζει την «κατανομή του συνόλου των ατόμων ενός πληθυσμού κατά φύλο και ηλικία» (Κοτζαμάνης, 2009:49).

Πρόκειται για ιδιότυπο ιστόγραμμα το οποίο βασίζεται είτε στις απόλυτες συχνότητες (αριθμός ατόμων) είτε στις σχετικές συχνότητες (%) των ηλικιακών ομάδων ανά φύλο.

Συνήθως:

- Χρησιμοποιούνται οι σχετικές συχνότητες (ειδικά όταν πρόκειται για μεγάλο πληθυσμό) για να υπάρχει δυνατότητα σύγκρισης μεταξύ διαφορετικών χωρικών ενοτήτων ή ακόμα μεταξύ διαφορετικών περιόδων για την ίδια χωρική ενότητα.
- Οι ομάδες ηλικιών αφορούν 5ετείς ομάδες ηλικιών δηλαδή 0-4, 5-9, 10-14 κ.ά.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.3. Πυραμίδα Ηλικιών

Για την κατασκευή της Πυραμίδας, πρέπει να διατηρηθούν ορισμένοι κανόνες.

Η προετοιμασία των δεδομένων υπό την μορφή πίνακας είναι πολύ σημαντική (βλέπε Πίνακα 5):

- οι γραμμές αντιστοιχούν στις **5ετείς ομάδες ηλικιών**,
- οι στήλες δίνουν τον αριθμό ατόμων για το σύνολο του πληθυσμού, τους άνδρες και τις γυναίκες.

Βήμα 1: Υπολογίζουμε τις σχετικές συχνότητες για κάθε ομάδα ηλικιών τόσο για τους άνδρες όσο και για τις γυναίκες. Οι σχετικές συχνότητες υπολογίζονται στο **σύνολο του πληθυσμού** που εξετάζεται → το εμβαδόν της πυραμίδας αντιστοιχεί στο συνολικό πληθυσμό ανεξαρτήτως φύλου.

Βήμα 2: Παράγουμε **2** ιστογράμματα σε ένα **κοινό διάγραμμα** όπου η κατανομή των ανδρών κατά ηλικία εμφανίζεται στον αριστερό τμήμα του διαγράμματος και η αντίστοιχη κατανομή των γυναικών στο δεξί τμήμα. Στον **κάθετο άξονα**, εμφανίζονται οι 5ετείς ομάδες ηλικίας. Στον **οριζόντιο άξονα**, εμφανίζονται οι σχετικές συχνότητες (τα ύψη των παραλληλογράμμων είναι ανάλογα με τις σχετικές συχνότητες).

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.3. Πυραμίδα Ηλικιών

Η κατανομή του πληθυσμού της Ελλάδας το 2010 κατά φύλο και ηλικία δίνεται στον ακόλουθο πίνακα:

| Ηλικίες | ΣΥΝΟΛΟ | ΑΝΔΡΕΣ | ΓΥΝΑΙΚΕΣ | (%) ΑΝΔΡΕΣ | (%) ΓΥΝΑΙΚΕΣ |
|---------|------------|-----------|-----------|------------|--------------|
| 0-4 | 553.769 | 283.010 | 270.759 | 2,55 | 2,44 |
| 5-9 | 533.513 | 272.813 | 260.700 | 2,45 | 2,34 |
| 10-14 | 540.253 | 275.494 | 264.759 | 2,48 | 2,38 |
| 15-19 | 586.760 | 300.251 | 286.509 | 2,70 | 2,58 |
| 20-24 | 655.860 | 337.699 | 318.161 | 3,04 | 2,86 |
| 25-29 | 799.334 | 415.240 | 384.094 | 3,73 | 3,45 |
| 30-34 | 849.638 | 435.265 | 414.373 | 3,91 | 3,73 |
| 35-39 | 848.304 | 428.046 | 420.258 | 3,85 | 3,78 |
| 40-44 | 844.986 | 419.814 | 425.172 | 3,78 | 3,82 |
| 45-49 | 766.777 | 374.917 | 391.860 | 3,37 | 3,52 |
| 50-54 | 733.516 | 357.639 | 375.877 | 3,22 | 3,38 |
| 55-59 | 659.982 | 322.370 | 337.612 | 2,90 | 3,04 |
| 60-64 | 636.501 | 306.149 | 330.352 | 2,75 | 2,97 |
| 65-69 | 521.608 | 248.188 | 273.420 | 2,23 | 2,46 |
| 70-74 | 558.427 | 253.168 | 305.259 | 2,28 | 2,75 |
| 75-79 | 480.906 | 212.385 | 268.521 | 1,91 | 2,41 |
| 80-84 | 333.033 | 137.812 | 195.221 | 1,24 | 1,76 |
| 85-89 | 147.536 | 56.905 | 90.631 | 0,51 | 0,82 |
| 90-94 | 50.643 | 17.875 | 32.768 | 0,16 | 0,29 |
| 95+ | 17.943 | 5.825 | 12.118 | 0,05 | 0,11 |
| Συνολο | 11.119.289 | 5.460.865 | 5.658.424 | 49,1 | 50,9 |

Τα ποσοστά (σχετικές συχνότητες υπολογίζονται με βάση το συνολικό πληθυσμό = 11.119.289

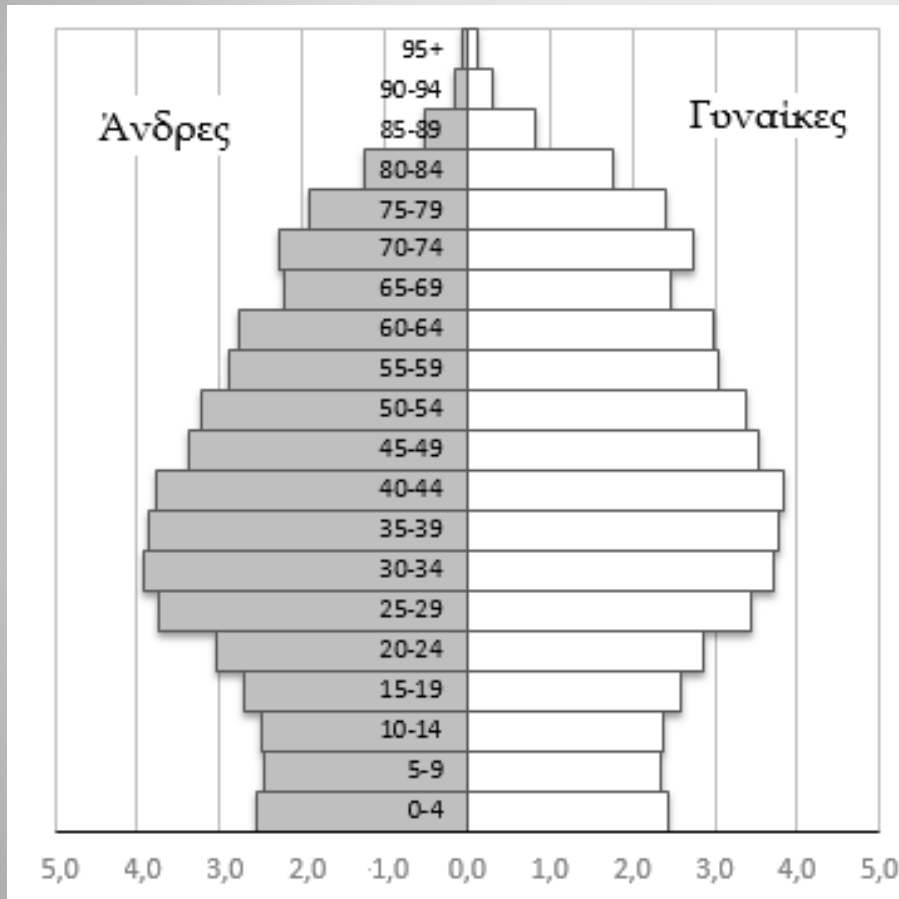
Οι άνδρες 20-24 ετών ανέρχονται σε 337.699 δηλαδή: 3,04% του συνολικού πληθυσμού της Ελλάδας το 2010:

$$f_5 = 100 \times \frac{337,699}{11,119,289} = 3,04$$

Το άθροισμά όλων των συχνοτήτων = 100%

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.3. Πυραμίδα Ηλικιών

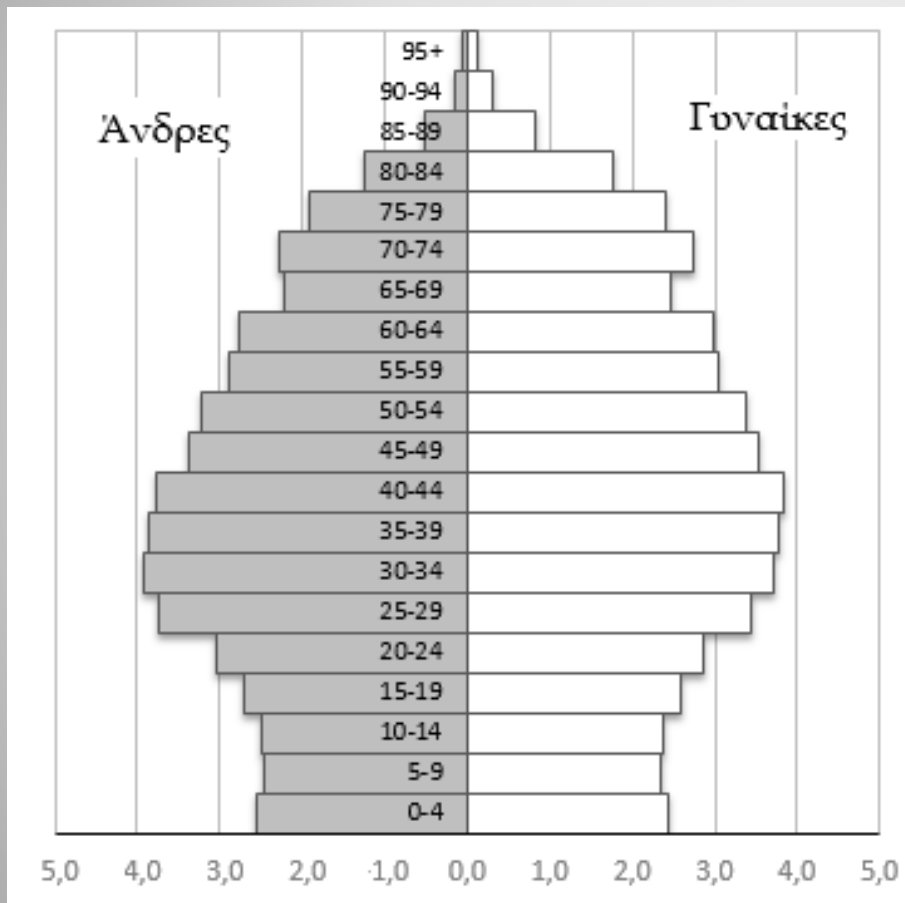


Πυραμίδα
Ηλικιών για τον
Πληθυσμό της
Ελλάδας, 2010

Μικρή βάση που αναδεικνύει το πρόβλημα γεννητικότητας της Ελλάδας με αποτέλεσμα να υπάρχει ζήτημα αναπαραγωγής του πληθυσμού και μεγάλης γήρανσης.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.3. Πυραμίδα Ηλικιών



Πυραμίδα
Ηλικιών για τον
Πληθυσμό της
Ελλάδας, 2010

Μικρή βάση που αναδεικνύει το πρόβλημα γεννητικότητας της Ελλάδας με αποτέλεσμα να υπάρχει ζήτημα αναπαραγωγής του πληθυσμού και μεγάλης γήρανσης.

ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

1.3. Πυραμίδα Ηλικιών

Η σύγκριση μεταξύ δύο ετών αναδεικνύει τις δομικές εξελίξεις του πληθυσμού

