



Ποσοτική & Ποιοτική Ανάλυση Δεδομένων Βασικές Έννοιες

Παιδαγωγικό Τμήμα Δημοτικής Εκπαίδευσης
Δημοκρίτειο Πανεπιστήμιο Θράκης
Αλεξανδρούπολη
2013 - 2014

Περιγραφική και Επαγωγική Στατιστική

- Η περιγραφική στατιστική περιλαμβάνει μεθόδους για τη σύνοψη, την οργάνωση και την απλοποίηση δεδομένων.
- Η επαγωγική στατιστική περιλαμβάνει μεθόδους που επιτρέπουν την εκτίμηση των παραμέτρων του πληθυσμού, μέσα από τις τιμές των στατιστικών του δείγματος.

Βασικές Έννοιες

Μεταβλητή (ή Χαρακτηριστικό ή Παράμετρος)

Κάθε ιδιότητα ενός αντικειμένου ή μια κατάσταση που παίρνει διαφορετικές τιμές. Οι τιμές αυτές δεν είναι απαραίτητο να είναι αριθμητικές.

Παραδείγματα:

Το βάρος, η νοημοσύνη, η στάση απέναντι στο ρατσισμό, τα πολιτικά κόμματα που μετέχουν σε μια εκλογική αναμέτρηση.

Βασικές Έννοιες

- Οι μεταβλητές διακρίνονται γενικά σε:

Ποσοτικές

Μεταβλητές των οποίων οι τιμές έχουν αριθμητικές ιδιότητες και εκφράζονται με μία μονάδα μέτρησης.

Π.χ.

Εισόδημα,
Βάρος,
Ύψος,
Βαθμολογία

Ποιοτικές

Μεταβλητές οι οποίες αναφέρονται σε κάποιο ποιοτικό χαρακτηριστικό.

Π.χ. Φύλο,

Επίπεδο μόρφωσης,
Περιοχή καταγωγής

Βασικές Έννοιες

- Οι μεταβλητές διακρίνονται επίσης σε:

Ασυνεχείς ή Διακριτές

Όταν μπορούν να λάβουν έναν πεπερασμένο, ακέραιο αριθμό τιμών, δηλαδή υπάρχουν τιμές μεταξύ των οποίων δεν είναι δυνατό να υπάρξει άλλη τιμή.

Π.χ.

Αριθμός μελών Οικογένειας
Αριθμός μαθητών μιας τάξης

Συνεχείς

Όταν μπορούν να λάβουν κάθε πιθανή τιμή μέσα σε ένα διάστημα πραγματικών αριθμών.

Π.χ.

Ηλικία, Ύψος, Βάρος,
Θερμοκρασία

Βασικές Έννοιες

■ Κλίμακες Μέτρησης Μεταβλητών

1. Κατηγορική (nominal)

Κάθε τιμή της μεταβλητής έχει την ίδια αξία, οι τιμές δεν εκφράζουν κάποια διάταξη ή μετρήσιμη ποσότητα (π.χ. στη μεταβλητή «φύλο» η τιμή «άντρας» έχει την ίδια σπουδαιότητα με τη τιμή «γυναίκα»).

2. Ιεραρχική (ordinal)

Οι τιμές εκφράζουν διάταξη αλλά όχι μετρήσιμη ποσότητα (π.χ. η μεταβλητή «σχολική επίδοση» μπορεί να λάβει ακέραιες τιμές σε μια κλίμακα από 1=πολύ κακή έως 5=πολύ καλή).

Βασικές Έννοιες

■ Κλίμακες Μέτρησης Μεταβλητών

3. Ίσων διαστημάτων (interval)

Εκφράζουν μετρήσιμη ποσότητα (επιτρέπονται οι αριθμητικές πράξεις ανάμεσα στις τιμές), η απόσταση ανάμεσα σε δύο τιμές δεν είναι σταθερή και η τιμή μηδέν είναι αυθαίρετη (π.χ. θερμοκρασία, IQ).

4. Αναλογική (ratio)

Εκφράζουν μετρήσιμη ποσότητα (επιτρέπονται οι αριθμητικές πράξεις ανάμεσα στις τιμές), η απόσταση ανάμεσα σε δύο τιμές είναι σταθερή και η τιμή μηδέν σημαίνει απουσία (π.χ. βάρος, ηλικία).

Βασικές Έννοιες

- Να χαρακτηρίσετε τις παρακάτω μεταβλητές ως προς το είδος τους και την κλίμακα μέτρησης:

Φύλο

Ηλικία

Εισόδημα

Επίδοση Μαθητή (κλίμακα 1 ως 10)

Επίδοση Μαθητή (Α, Β, Γ ή Δ)

Επάγγελμα

Αγαπημένο Χρώμα

Θερμοκρασία σε βαθμούς Κελσίου

Αριθμός Μελών Οικογένειας

Ταχυδρομικός Κώδικας

Αριθμός τηλεφώνου

Βασικές Έννοιες

- Στο πλαίσιο μιας έρευνας όπου χρησιμοποιούνται εργαλεία της Επαγωγικής Στατιστικής μια μεταβλητή μπορεί να είναι:

Ανεξάρτητη Μεταβλητή

Είναι μια μεταβλητή που ενδέχεται να επηρεάζει κάποια ή κάποιες άλλες (τις εξαρτημένες)

Εξαρτημένη Μεταβλητή

Είναι μια μεταβλητή που υποθέτουμε ότι δέχεται τις επιδράσεις της άλλης (ανεξάρτητης).

Βασικές Έννοιες

Π.χ. Στα παρακάτω ζευγάρια ποιά είναι η ανεξάρτητη και ποια η εξαρτημένη;

- * Επάγγελμα πατέρα - Επάγγελμα παιδιού
- * Ύψος αγορών ηλεκτρικών ειδών - Αριθμός γάμων
- * Βαθμολογία μαθητή - Ώρες μελέτης
- * Στάση απέναντι στα μαθηματικά - Επίδοση στα μαθηματικά
- * Κατάθλιψη - Μείωση εισοδήματος

Κωδικοποίηση Δεδομένων

■ Απλό Ερωτηματολόγιο

Φύλο : άνδρας ☐ γυναίκα ☐

Ηλικία :

Οικογενειακή Κατάσταση : άγαμος ☐ έγγαμος ☐ χήρος ☐ διαζευγμένος ☐

Αριθμός ατόμων οικογενείας :

Επίπεδο Μόρφωσης : αγράμματος ☐ Α' βάθμια ☐ Β' βάθμια ☐ Γ' Βάθμια ☐

Ποιά είναι η γνώμη σου για τη δημοτικοποίηση των αστικών συγκοινωνιών :

Πρέπει να γίνει ☐

Δεν πρέπει να γίνει ☐

Δεν έχω γνώμη ☐

Κωδικοποίηση Δεδομένων

■ Ποιοτική ή Διακριτή Μεταβλητή

Αντιστοιχούμε έναν ακέραιο αριθμό σε κάθε κατηγορία της μεταβλητής.

Οικογενειακή Κατάσταση	Κωδικός
Άγαμος	1
Έγγαμος	2
Χήρος	3
Διαζευγμένος	4

Κωδικοποίηση Δεδομένων

■ Ποιοτική ή Διακριτή Μεταβλητή

Αντιστοιχούμε έναν ακέραιο αριθμό σε κάθε κατηγορία της μεταβλητής.

Αριθμός παιδιών στην οικογένεια	Κωδικός
0	1
1	2
2	3
3	4
4 και άνω	5

Κωδικοποίηση Δεδομένων

■ Ποσοτική ή Συνεχής Μεταβλητή

Οι τιμές των ποσοτικών μεταβλητών είτε παραμένουν ως έχουν είτε ομαδοποιούνται σε κατηγορίες («κατηγοριοποιούνται»).

Ηλικία
45
20
30
20
25
14
60



Ηλικία	Κωδικός
Μέχρι 17	1
Από 18 έως 24	2
Από 25 έως 30	3
Από 31 έως 40	4
Από 41 έως 60	5
Μεγαλύτερος από 60	6

Κατασκευή Πίνακα Δεδομένων

Φύλο : άνδρας ☒

γυναίκα ☐

Ηλικία :**45**.....

Οικογενειακή Κατάσταση : άγαμος ☐ έγγαμος ☒ χήρος ☐ διαζευγμένος ☐

Αριθμός ατόμων οικογενείας :**2**.....

Επίπεδο Μόρφωσης : αγράμματος ☐ Α' βάρθμια ☐ Β' βάρθμια ☐ Γ' Βάρθμια ☒

Ποιά είναι η γνώμη σου για τη δημοτικοποίηση των αστικών συγκοινωνιών :

Πρέπει να γίνει ☒

Δεν πρέπει να γίνει ☐

Δεν έχω γνώμη ☐

α/α	Φύλο	Ηλικία	Οικ.Κατ.	Μέλη Οικ.	Μόρφ.	Γνώμη
1	1	45	2	2	3	1
2	2	23	1	1	3	3
3	...	...	...	...	...	...

Κατασκευή Πίνακα Δεδομένων

α/α	Φύλο	Ηλικία	Οικ.Κατ.	Μέλη Οικ.	Μόρφ.	Γνώμη
1	1	45	2	2	4	1
2	2	23	1	1	3	3
3	1	30	4	3	2	2
4	1	20	2	4	1	2
5	2	25	1	2	4	1
6	1	52	2	5	4	2
7	1	60	2	6	1	1
8	2	48	3	2	3	2
9	2	14	1	1	2	3
10	1	22	1	1	3	1

10 άτομα - στατιστικές μονάδες (γραμμές) και
6 χαρακτηριστικά - μεταβλητές (στήλες)

Οργάνωση & Παρουσίαση Δεδομένων

- Η οργάνωση των δεδομένων περιλαμβάνει το σύνολο των ενεργειών που γίνονται ώστε τα διαθέσιμα δεδομένα να μπορούν να πάρουν τέτοια μορφή ώστε να αναλύονται και να παρουσιάζονται εποπτικά.
- Έλεγχος Ορθότητας
- Έλεγχος Πληρότητας

Οργάνωση & Παρουσίαση Ποιοτικών Δεδομένων

- Οι πίνακες και τα διαγράμματα που παρουσιάζουν ποιοτικά ή κατηγορικά δεδομένα πρέπει να δείχνουν τη **συχνότητα** (frequency) των περιπτώσεων κάθε κατηγορίας.
- **Συχνότητα:** Ο αριθμός των ατόμων ή των περιπτώσεων που εμπίπτουν σε μια κατηγορία.
- **Απόλυτη Συχνότητα:** Ο αριθμός των ατόμων ή των περιπτώσεων μιας κατηγορίας σε απόλυτο αριθμό.
- **Σχετική Συχνότητα:** Ο αριθμός των ατόμων ή των περιπτώσεων μιας κατηγορίας σε ποσοστό (%) επί του συνόλου.

Πίνακας Συχνοτήτων Ποιοτικών Δεδομένων

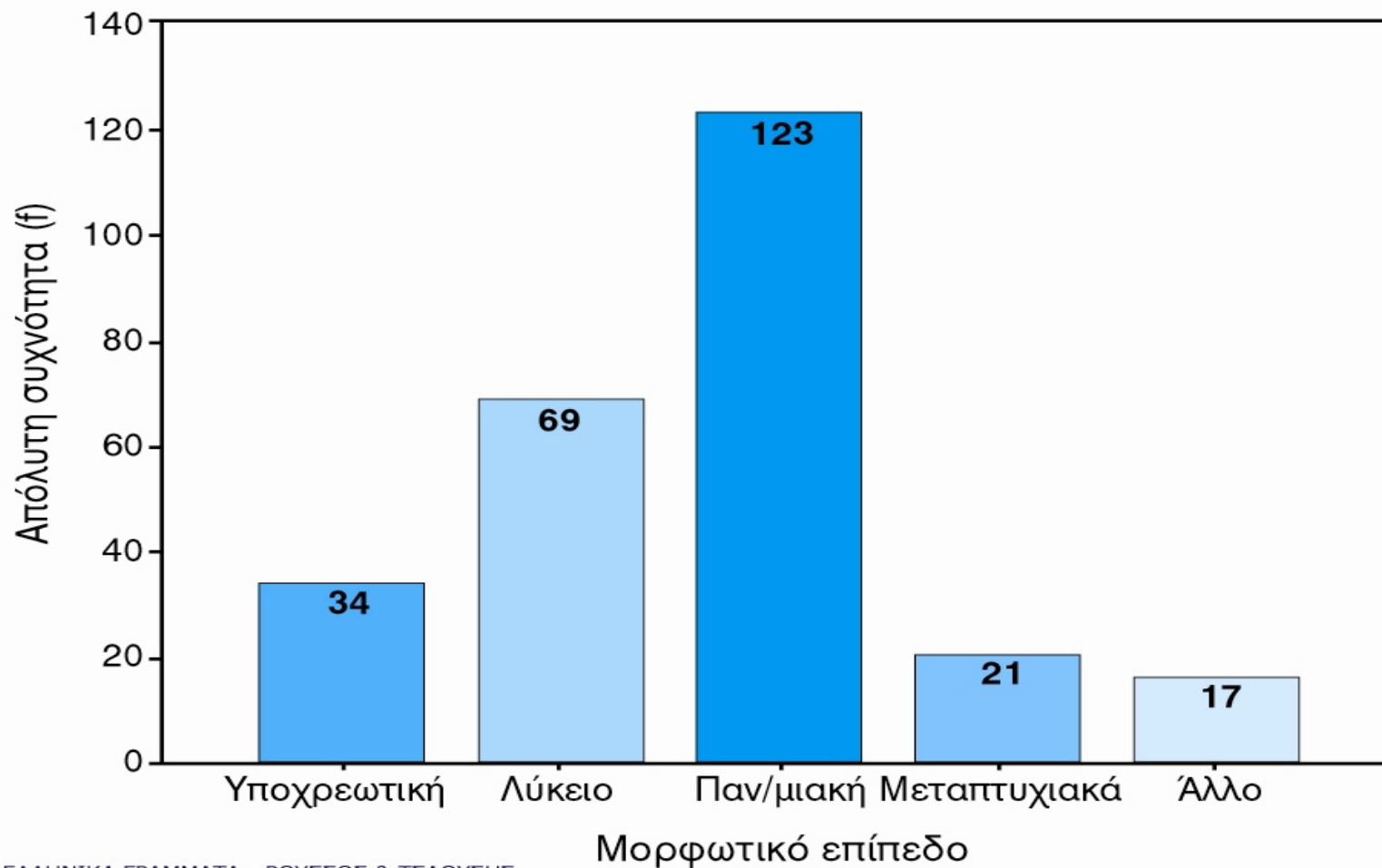
Πίνακας 1. Επίπεδο Εκπαίδευσης Συμμετεχόντων

Εκπαίδευση	Συχνότητα (f)	Σχετική συχνότητα (rf)
Υποχρεωτική	34	12.88
Λύκειο	69	26.14
Πανεπιστημιακή	123	46.59
Μεταπτυχιακά	21	7.95
Άλλο	17	6.44
ΣΥΝΟΛΟ	264	100

Πηγή: Ρούσσος & Τσαούσης (2006)

$$\frac{34}{264} \times 100$$

Ραβδόγραμμα ή Ραβδωτό Διάγραμμα



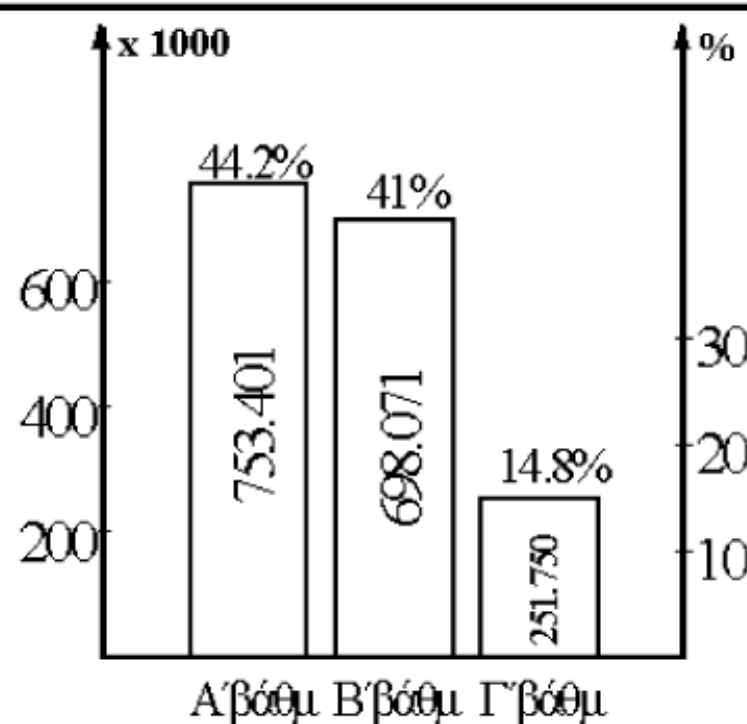
Διάγραμμα 1: Εκπαιδευτικό Επίπεδο Συμμετεχόντων.

Ραβδόγραμμα ή Ραβδωτό Διάγραμμα

Πίνακας 5.2 : Κατανομή των Ελλήνων που φοιτούσαν στις τρεις βαθμίδες της εκπαίδευσης (1991)

βαθμίδα	Αριθμός	%
Αβάθμια	753.401	44.2
Ββάθμια	698.071	41.0
Γβάθμια	251.750	14.8
Σύνολο	1.703.222	100.0

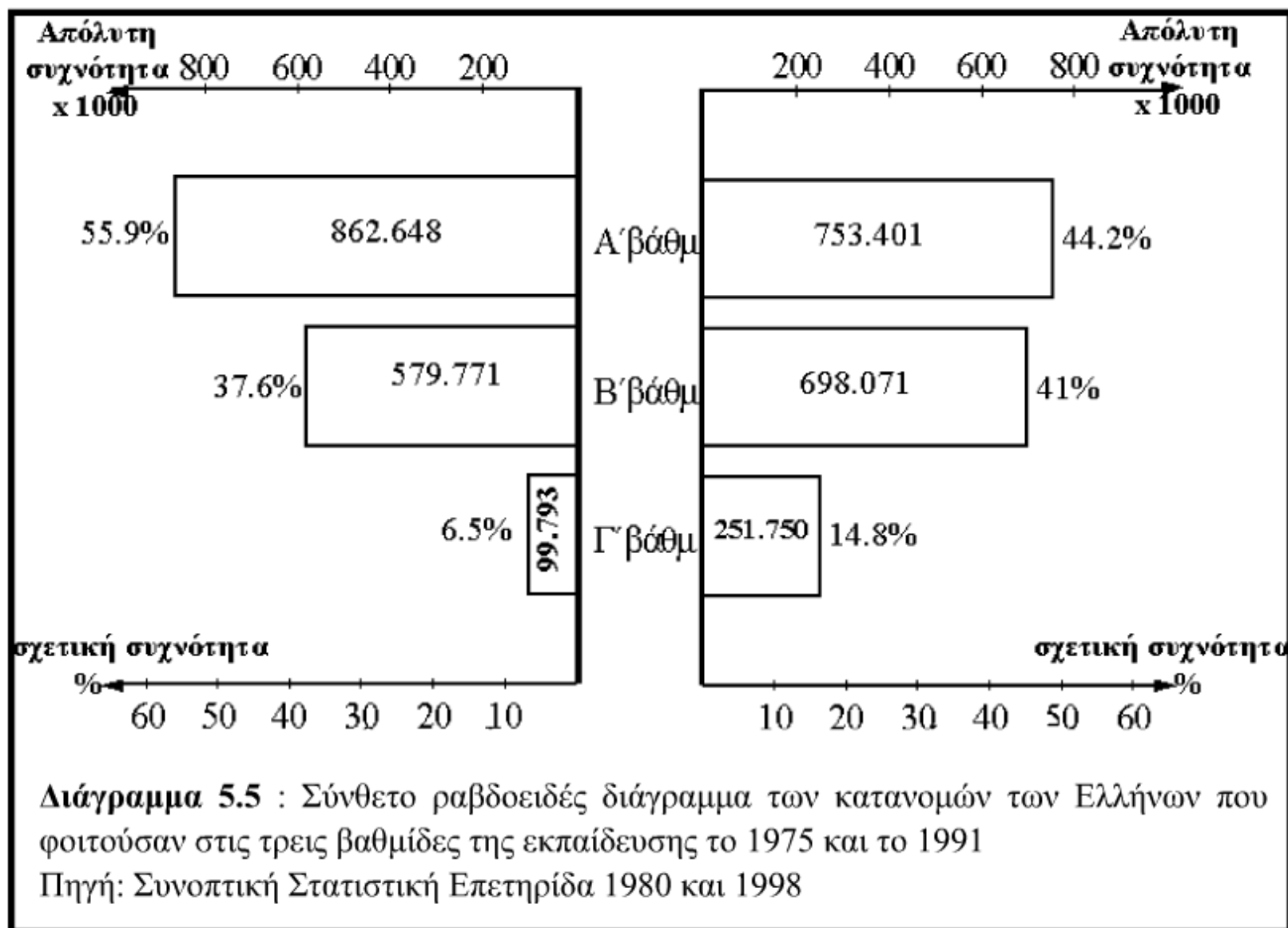
Πηγή :Συνοπτική Στατιστική Επετηρίδα
1998



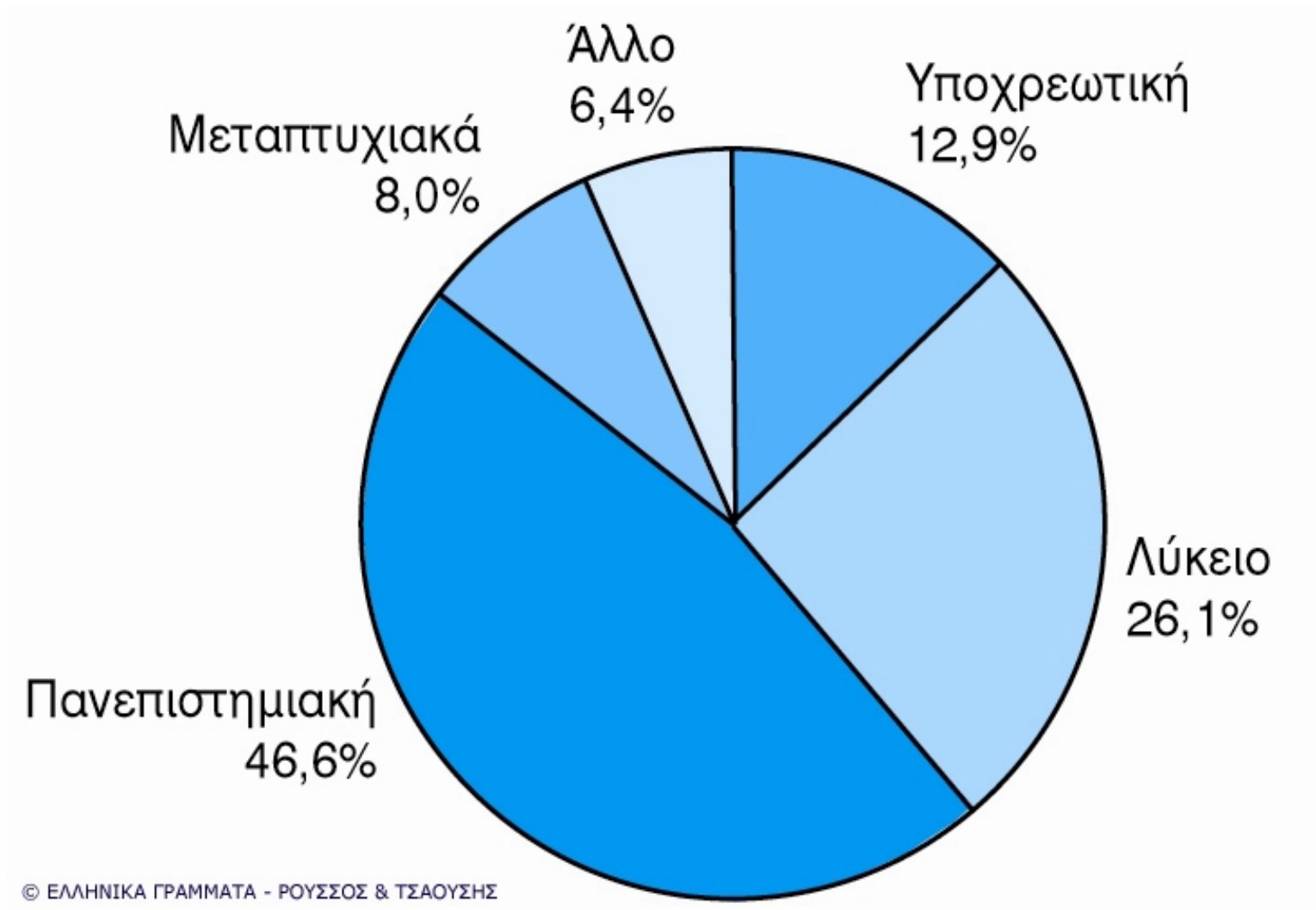
Διάγραμμα 5.4 : Κατανομή των Ελλήνων που φοιτούν στις τρεις βαθμίδες της εκπαίδευσης (1991)

Πηγή :Συνοπτική Στατιστική Επετηρίδα
1998

Ραβδόγραμμα ή Ραβδωτό Διάγραμμα



Κυκλικό Διάγραμμα



Διάγραμμα 2: Εκπαιδευτικό Επίπεδο Συμμετεχόντων.

Οργάνωση & Παρουσίαση Ποσοτικών Δεδομένων

- Ο πιο συνηθισμένος τρόπος οργάνωσης μιας ομάδας ποσοτικών δεδομένων είναι η **κατανομή συχνότητας** (frequency distribution). Η κατανομή συχνότητας δείχνει πόσο συχνά συναντάται κάθε τιμή της μεταβλητής.
- **Ομαδοποιημένη Κατανομή Συχνότητας:** Πίνακας στον οποίο παρουσιάζεται ο αριθμός των ατόμων ή των περιπτώσεων που εμπίπτουν σε κάθε διάστημα τιμών

Πίνακας Συχνοτήτων Ποσοτικής Μεταβλητής

**Ερώτημα: Δηλώστε το βαθμό συμφωνίας ή
διαφωνίας σας με την πρόταση
‘Η Στατιστική με φοβίζει’ (N=80)**

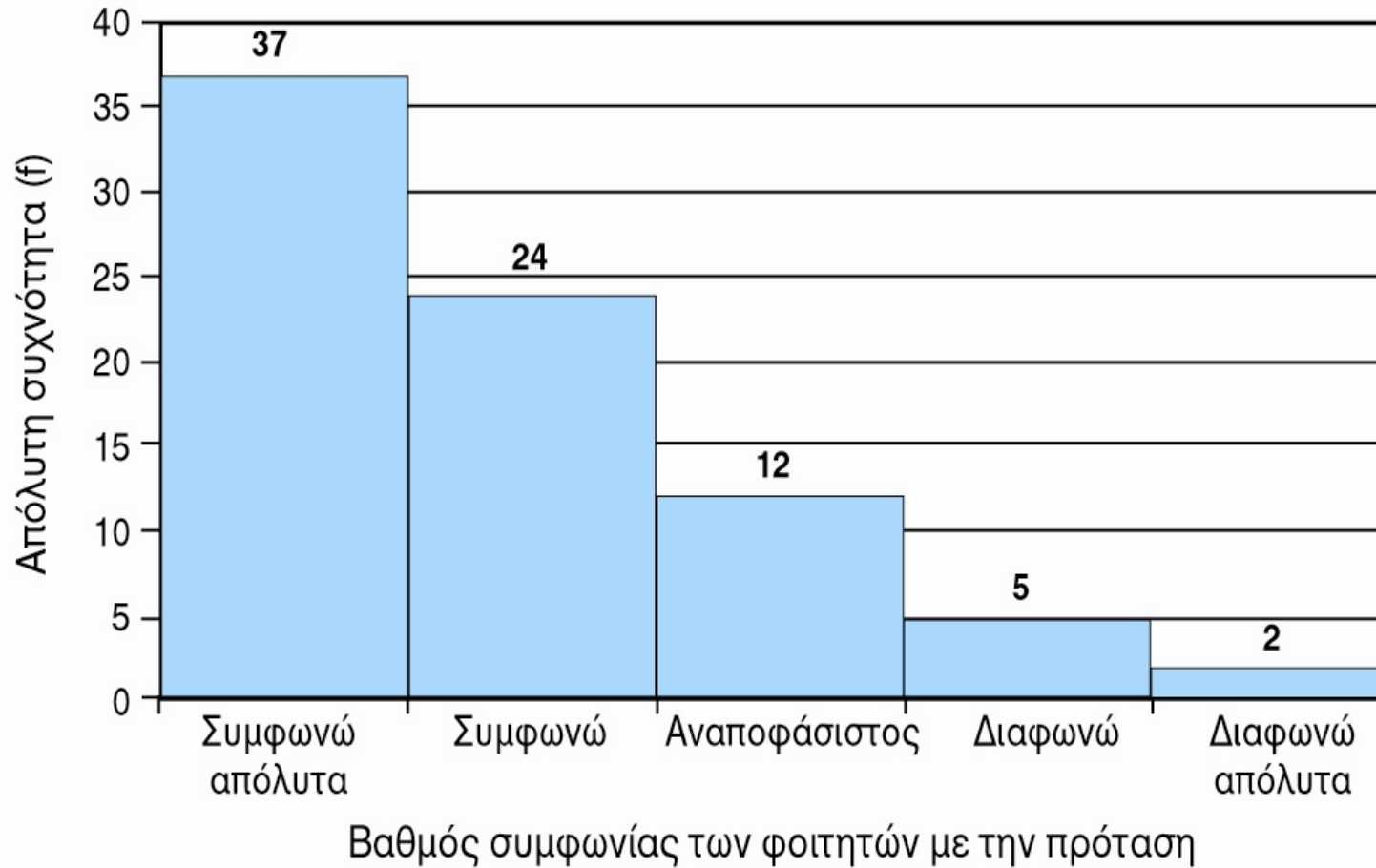
Πίνακας 1. Κατανομή Συχνότητας Εμφάνισης Ηλικιών

Γνώμη		Συχνότητα	Σχετική Συχνότητα (%)
1	Συμφωνώ απόλυτα	37	46,25
2	Συμφωνώ	24	30,00
3	Αναποφάσιστος	12	15,00
4	Διαφωνώ	5	6,25
5	Διαφωνώ απόλυτα	2	2,50
Σύνολο		80	100,0

Ιστόγραμμα

Ερώτημα: 'Η Στατιστική με φοβίζει' (N=80)

© ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ - ΡΟΥΣΣΟΣ & ΤΣΑΟΥΣΗΣ



Διάγραμμα 1: Στάση των Φοιτητών Απέναντι στο Μάθημα της Στατιστικής

Πίνακας Συχνοτήτων Ποσοτικής Μεταβλητής

Πίνακας 2. Ομαδοποιημένη Κατανομή Συχνότητας Εμφάνισης Ηλικιών

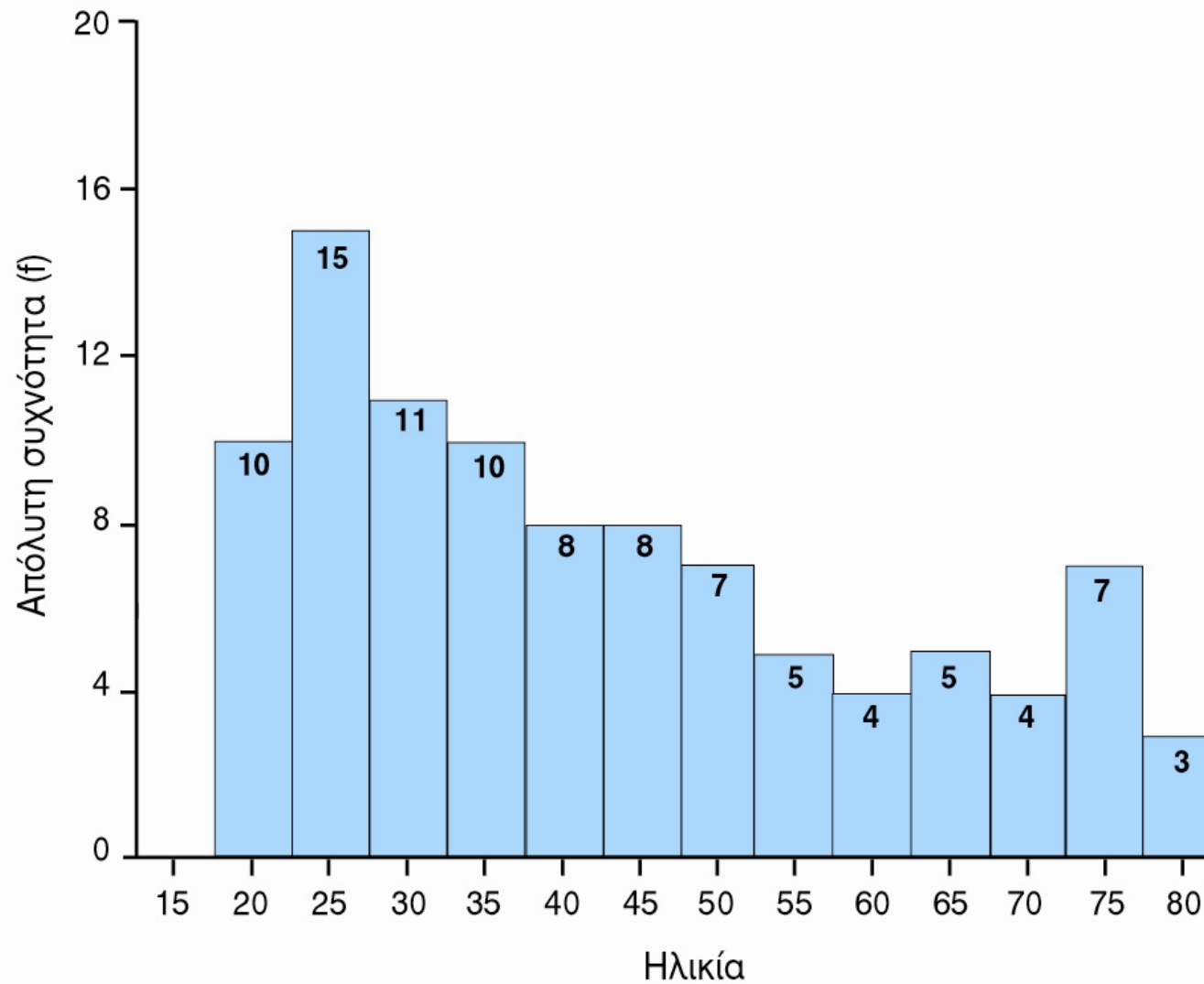
Διαστήματα τιμών	Μέσο διαστήματος	Συχνότητα	Σχετική Συχνότητα (%)
18-22	20	10	10,31
23-27	25	15	15,46
28-32	30	11	11,34
33-37	35	10	10,31
38-42	40	8	8,25
43-47	45	8	8,25
48-52	50	7	7,22
53-57	55	5	5,15
58-62	60	4	4,12
63-67	65	5	5,15
68-72	70	4	4,12
73-77	75	7	7,22
78-82	80	3	3,09
Σύνολο		97	100

Πίνακας Συχνοτήτων Ποσοτικής Μεταβλητής

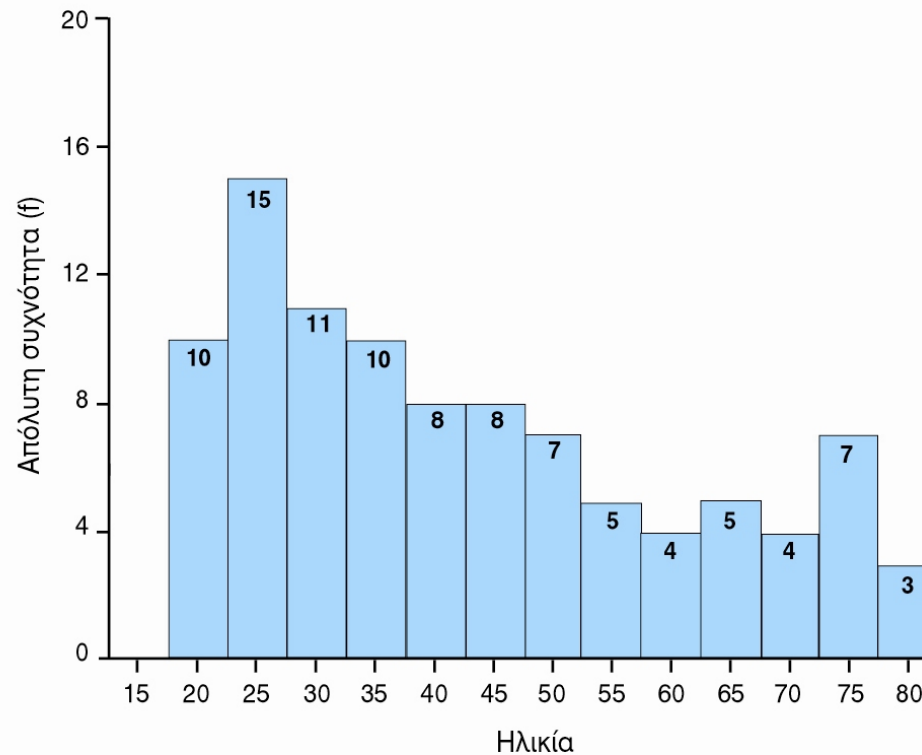
Πίνακας 3. Ομαδοποιημένη Κατανομή Συχνότητας Εμφάνισης Ηλικιών

Διαστήματα τιμών	Συχνότητα		Αθροιστική Συχνότητα	
	Απόλυτη	Σχετική (%)	Απόλυτη	Σχετική (%)
18-22	10	10,31	10	10,31
23-27	15	15,46	25	25,77
28-32	11	11,34	36	37,11
33-37	10	10,31	46	47,42
38-42	8	8,25	54	55,67
43-47	8	8,25	62	63,92
48-52	7	7,22	69	71,14
53-57	5	5,15	74	76,29
58-62	4	4,12	78	80,41
63-67	5	5,15	83	85,56
68-72	4	4,12	87	89,68
73-77	7	7,22	94	96,90
78-82	3	3,09	97	100,00
Σύνολο	97	100		

Ιστόγραμμα ομαδοποιημένης κατανομής



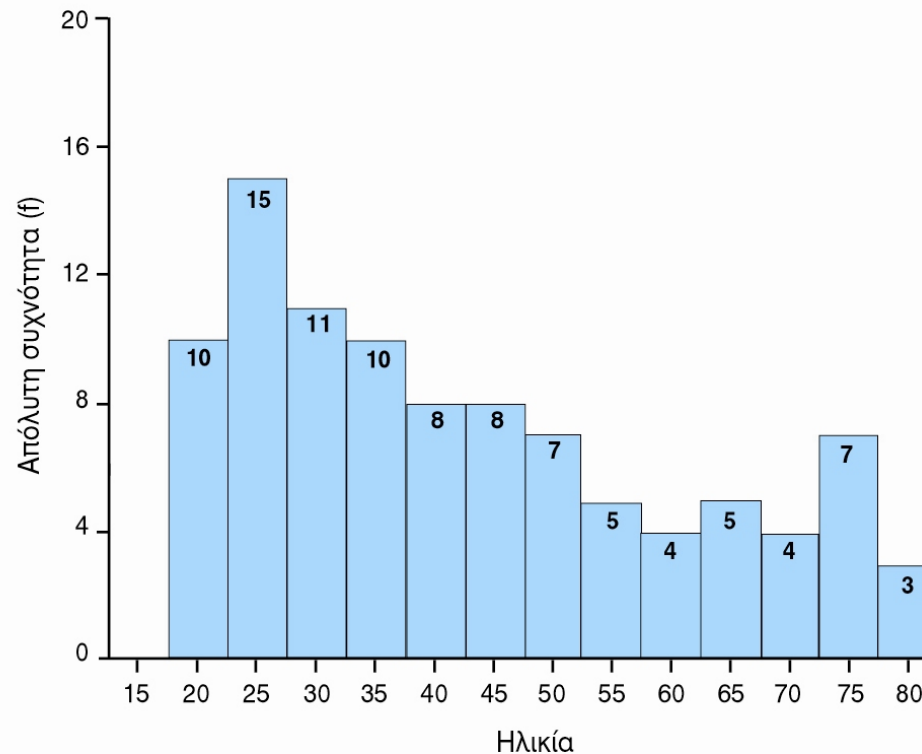
Ιστόγραμμα ομαδοποιημένης κατανομής



© ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ - ΡΟΥΣΣΟΣ & ΤΣΑΟΥΣΗΣ

➤ Τα διαστήματα αναπαρίστανται από την μέση τιμή τους στο κέντρο κάθε ράβδου. Όλες οι ράβδοι έχουν το ίδιο πλάτος. Η επιφάνεια της κάθε ράβδου είναι ανάλογη του αριθμού (ή του ποσοστού) των περιπτώσεων που αναπαριστά.

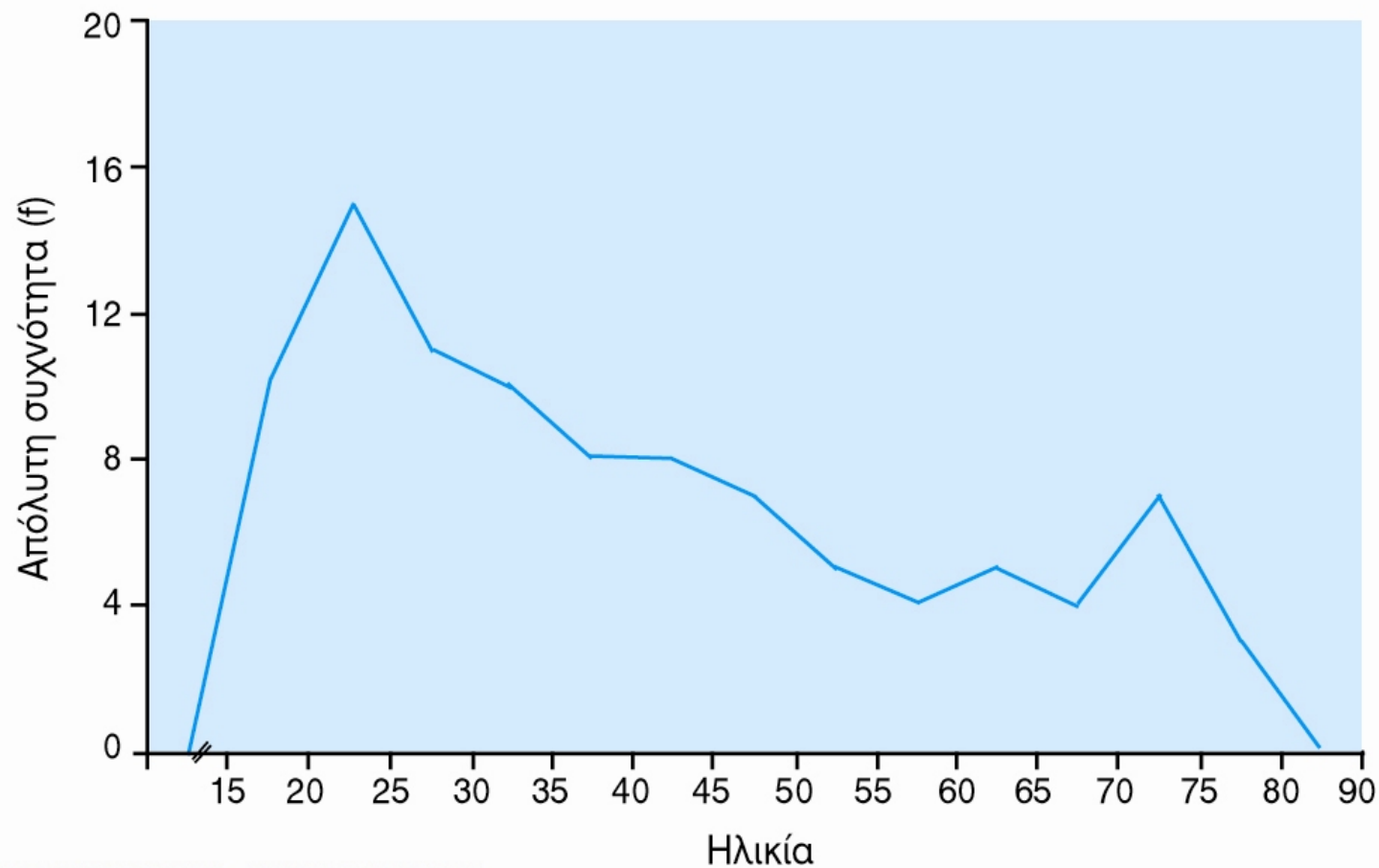
Ιστόγραμμα ομαδοποιημένης κατανομής



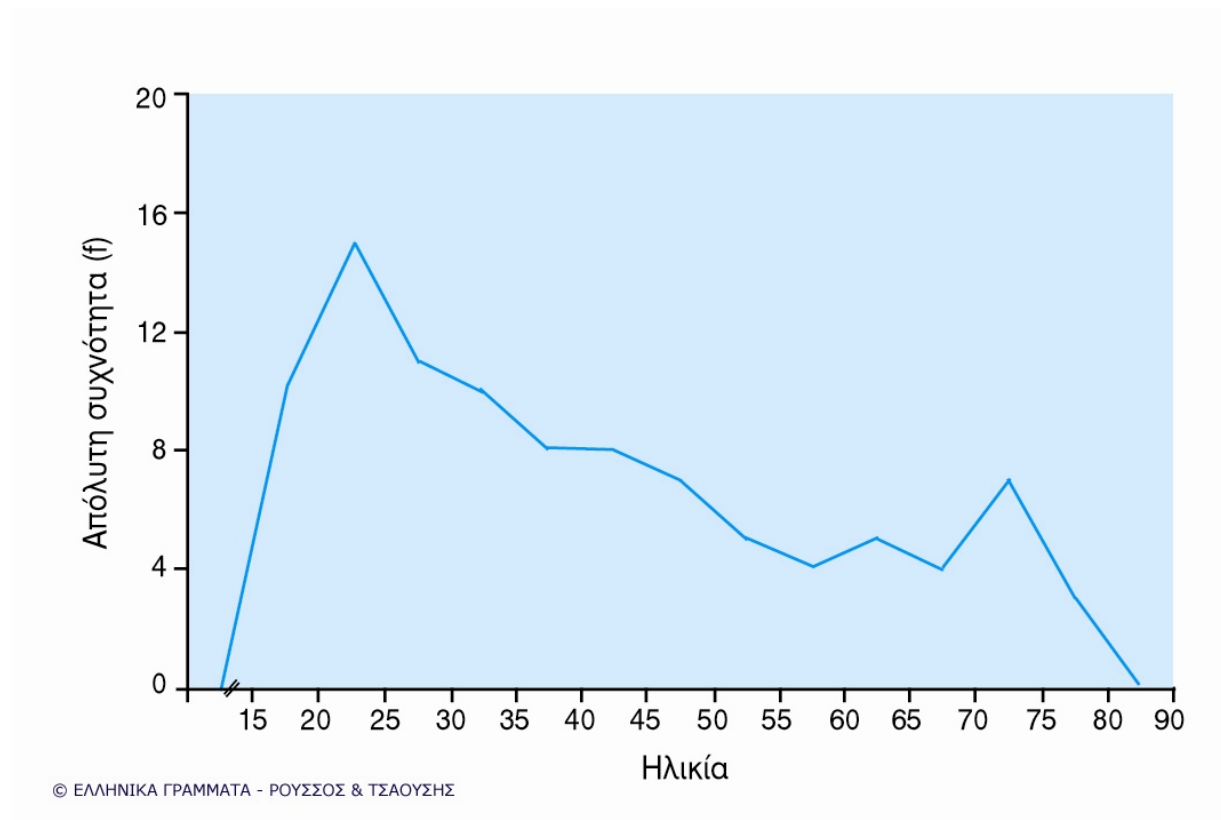
© ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ - ΡΟΥΣΣΟΣ & ΤΣΑΟΥΣΗΣ

➤ Τα ιστογράμματα μας δίνουν πληροφορίες για τα χαρακτηριστικά της κατανομής των δεδομένων. Πχ ποιά είναι η τιμή με την υψηλότερη συχνότητα, αν υπάρχουν ακραίες τιμές στην κατανομή, και (το σημαντικότερο) πώς διασπείρονται (κατανέμονται) οι διάφορες τιμές.

Πολύγωνο συχνότητας

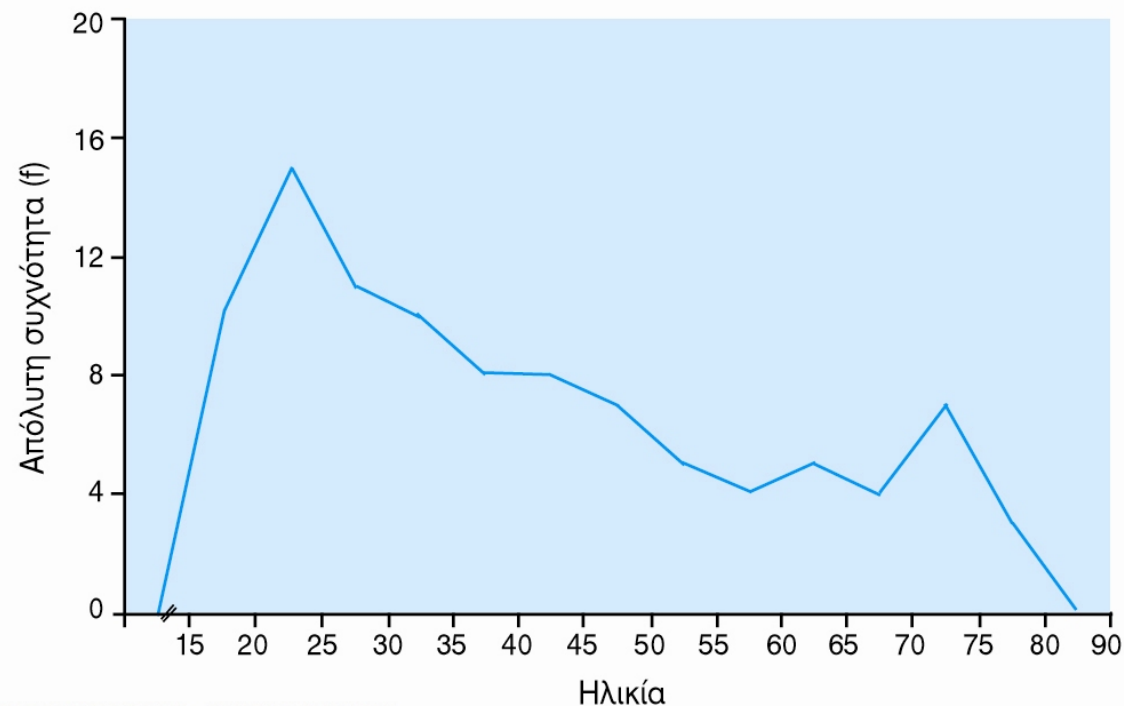


Κατασκευή πολυγώνου συχνότητας



➤ Βάζουμε μία κουκίδα στο μέσο του ανώτερου σημείου της κάθε ράβδου, και ενώνουμε τις κουκίδες.

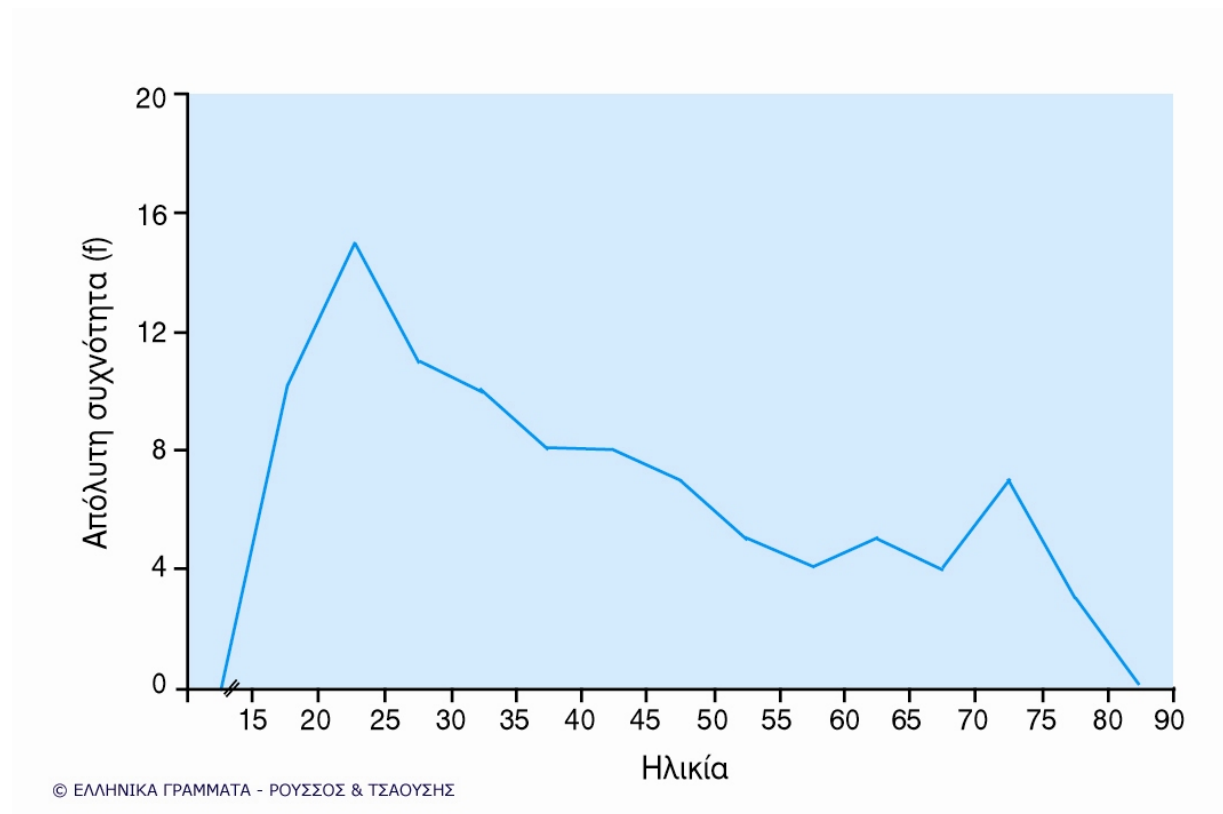
Κατασκευή πολυγώνου συχνότητας



© ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ - ΡΟΥΣΣΟΣ & ΤΣΑΟΥΣΗΣ

✧Κάτω από την κατώτερη και πάνω από την ανώτερη τιμή συμπεριλαμβάνουμε μηδενικές τιμές (συχνότητες) στον οριζόντιο άξονα, κλείνοντας το πολύγωνο, γιατί υπάρχουν κι αυτές οι τιμές στην κλίμακα, οι οποίες είχαν συχνότητα μηδέν. Η γραμμή και στα δύο άκρα της κατανομής ξεκινάει από το μηδέν.

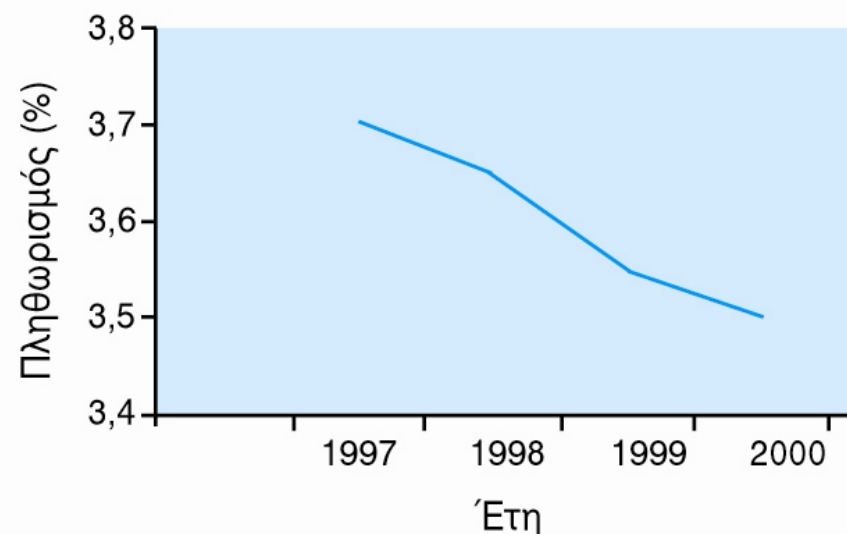
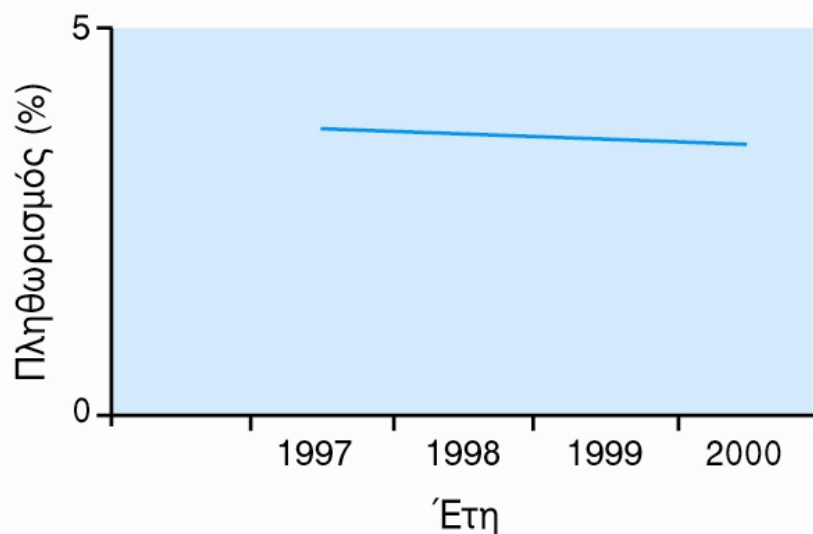
Κατασκευή πολυγώνου συχνότητας



➤ Προσέχουμε να συμπεριλαμβάνουμε την τιμή μηδέν, όταν η συχνότητα ενός διαστήματος είναι μηδέν. Πχ αν δεν υπάρχει κανείς συμμετέχων που να έχει ηλικία από 13 έως 17, δεν παραλείπουμε να σημειώσουμε την τιμή μηδέν.

Κατασκευή πολυγώνου συχνότητας

Η επίδραση του χειρισμού του κάθετου άξονα στη μορφή του πολυγώνου...



© ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ - ΡΟΥΣΣΟΣ & ΤΣΑΟΥΣΗΣ

Πληθωρισμός 1997: 3.7%
Πληθωρισμός 2000: 3.5%

Πίνακες Συνάφειας ή Διπλής Εισόδου

■ Δύο μεταβλητών (απολύτων συχνοτήτων)

Πίνακας 4.16: Πίνακας συμπτώσεων των μεταβλητών ηλικία και επίπεδο μόρφωσης

	Επίπεδο μόρφωσης			
Ηλικία	A' βάθμια	B' βάθμια	Γ' βάθμια	Σύνολο
18-25	10	110	30	150
25-35	20	320	320	660
35-50	50	350	280	680
50-65	120	220	170	510
Σύνολο	200	1000	800	2000

- Στο κάθε ένα από τα 12 κελιά του Πίνακα εμφανίζεται ο αριθμός των ατόμων που έχουν συγχρόνως τις δύο ιδιότητες που αντιστοιχούν στη γραμμή και στήλη που το ορίζουν.

Πίνακες Συνάφειας ή Διπλής Εισόδου

■ Δύο μεταβλητών (σχετικών συχνοτήτων)

Πίνακας 4.17: Πίνακας συμπτώσεων σχετικών συχνοτήτων με ερμηνεία κατά στήλες

Ηλικία	Επίπεδο μόρφωσης			
	A' βάθμια	B' βάθμια	Γ' βάθμια	Σύνολο
18-25	0,050	0,110	0,038	0,075
25-35	0,100	0,320	0,400	0,330
35-50	0,250	0,350	0,350	0,340
50-65	0,600	0,220	0,212	0,255
Σύνολο	1,000	1,000	1,000	1,000

- Προκύπτει όταν τα στοιχεία κάθε στήλης ενός πίνακα συμπτώσεων (απολύτων συχνοτήτων) διαιρεθούν δια του αθροίσματός τους (το τελευταίο στοιχείο της στήλης).

Πίνακες Συνάφειας ή Διπλής Εισόδου

■ Δύο μεταβλητών (σχετικών συχνοτήτων)

Πίνακας 4.18: Πίνακας συμπτώσεων σχετικών συχνοτήτων με ερμηνεία κατά γραμμές

Ηλικία	Επίπεδο μόρφωσης			
	A' βάθμια	B' βάθμια	Γ' βάθμια	Σύνολο
18-25	0,067	0,733	0,200	1,000
25-35	0,030	0,485	0,485	1,000
35-50	0,074	0,515	0,411	1,000
50-65	0,235	0,432	0,333	1,000
Σύνολο	0,100	0,500	0,400	1,000

- Προκύπτει όταν τα στοιχεία κάθε γραμμής ενός πίνακα συμπτώσεων (απολύτων συχνοτήτων) διαιρεθούν δια του αθροίσματός τους (το τελευταίο στοιχείο της γραμμής).



#end