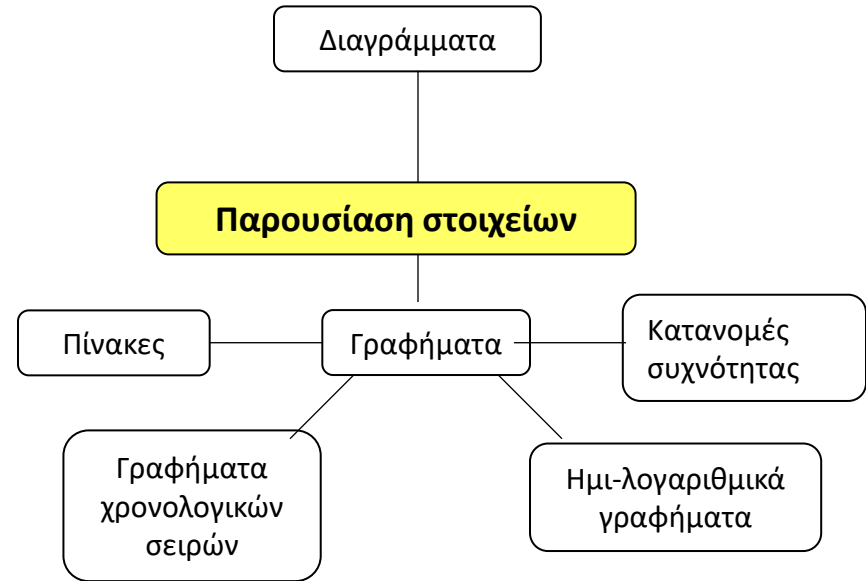


ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

- ✓Έχεις επιλέξει το πεδίο και θέμα της έρευνας
- ✓Έχεις ετοιμάσει ένα ερωτηματολόγιο για την συλλογή των στοιχείων (ή έχεις βρει τις πηγές από όπου θα τα αντλήσεις)
- ✓Έχεις ορίσει ένα πλαίσιο δειγματοληψίας
- ✓Έχεις καταναίμει την ερευνητική εργασία στα μέλη της ερευνητικής ομάδας σου
- ✓Έχεις συγκεντρώσει τα στοιχεία
- Τι πρόκειται να κάνεις με αυτά τα στοιχεία?



- Ο σκοπός της παρουσίασης τω στοιχείων είναι να τα οργανώσεις με ένα λογικό τρόπο αλλά και με ένα τρόπο για να μπορούν να κατανοηθούν και μελετηθούν εύκολα
- Αυτό γίνεται μόνο για περιγραφικούς λόγους και μόνο μια σχετικά περιορισμένη μελέτη μπορεί να γίνει μέσα από τα γραφήματα και τους πίνακες (περιγραφική στατιστική)
- Παρόλα αυτά συχνά η (παραπέρα) στατιστική ανάλυση βασίζεται σε μερικά «αρχικά ευρήματα» που προκύπτουν από τα γραφήματα

ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

ΠΙΝΑΚΕΣ

Ένας Πίνακας αποτελείται από στοιχεία σε γραμμές και στήλες.

- Οι στήλες έχουν ταυτότητα (τίτλο)
- Οι γραμμές έχουν ταυτότητα (τίτλο)

Πρόκειται για ένα δισδιάστατο (**διπλής εισόδου**) πίνακα (2D), που μπορεί να συμπεριλάβει δεδομένα που περιγράφονται από 2 μόνο μεταβλητές

Τίτλος

**Ηλεκτρισμός και παραγωγή γκαζιού στην Ελλάδα
Περίοδος 1955 - 1964**

Τίτλοι
στηλών

Ετος	Παραγωγή Ηλεκτρισμού (millions ΩXB)			Gas Production (millions m3)
	Πετρελαιοκίνηση	Υδροηλεκτρική	Συνολική	
1955	948	310	1.258	17
1956	974	502	1.476	17
1957	1.279	348	1.627	16
1958	1.339	451	1.790	16
1959	1.529	433	1.962	15
1960	1.729	469	2.198	14
1961	1.930	554	2.484	14
1962	2.121	614	2.735	13
1963	2.307	806	3.113	14
1964	2.967	749	3.716	12

Μονάδες
μέτρησης

Σωμα

Σημειώσεις

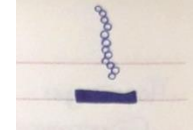
a) Η ενέργεια που αναλώνεται στο εργοστάσιο παραγωγής δεν συμπεριλαμβάνεται

Πηγή

Source: National office for statistics

Σημεία κλειδιά:

- καθαρός & συνοπτικός τίτλος
- τίτλοι στηλών & ίσως υπότιτλοι
- συμπεριέλαβε σύνολα στο τέλος δεξιά άκρη
- συχνά χρήσιμη και μια στήλη με ποσοστά
- στρογγυλοποιημένα νούμερα
- Η σειρά των στηλών και των σειρών ανάλογα με την σπουδαιότητά τους



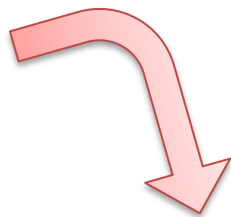
ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΑΣ

Δείγμα – δεδομένα από 10 οικογένειες

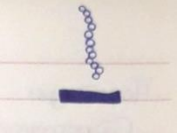
Οικογένεια	Επάγγελμα	Μηνιαίος Μισθός	Αριθμός Παιδιών
1	Εργάτης	700	0
2	Οδηγός	750	1
3	Εργάτης	800	0
4	Δημόσιος Υπάλληλος	700	2
5	Δημόσιος Υπάλληλος	800	2
6	Δημόσιος Υπάλληλος	800	2
7	Δάσκαλος	900	1
8	Ιερέας	1000	1
9	Οδηγός	600	1
10	Εργάτης	600	1

Εφαρμογή



Πίνακας Συχνοτήτων με το επάγγελμα

Αρίθμηση Κατηγοριών v Y_i	<u>Επάγγελμα</u>	Διαλογή	Συχνότητα V_i	Σχετική Συχνότητα f_i	Σχετική Συχνότητα $f_i \%$
1	Εργάτης		3	$3/10 = 0,3$	30%
2	Οδηγός		2	$2/10 = 0,2$	20%
3	Δημόσιος Υπάλληλος		3	$3/10 = 0,3$	30%
4	Δάσκαλος		1	$1/10 = 0,1$	10%
5	Ιερέας		1	$1/10 = 0,1$	10%
Σύνολο >		10	10	$10/10 = 1$	100%



ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΑΣ

Πολύ συχνά τα πρωτογενή στοιχεία είναι μια «μάζα» από ακατάστατα νούμερα. Μερικά επαναλαμβάνονται πολλές φορές – δύσκολο να τα χειριστείς όπως είναι – χρειάζονται οργάνωση

Διαλογή - Κατανομή Συχνοτήτων - Σχετικών Συχνοτήτων

Πήραμε δείγμα 30 πτυχιούχων οδοντιάτρων και ζητήσαμε να μάθουμε τη χρονική διάρκεια που απαιτήθηκε για την απόκτηση πτυχίου.

Οι απαντήσεις έδωσαν τους ακόλουθους χρόνους σε έτη:

Το μέγεθος του δείγματος n ή συχνότερα $n = 30$

Μεταβλητή $X_i = 7, 8, 6, 8, 6, 6, 6, 6, 6, 6, 6, 6, 8, 7, 8, 6, 6, 6, 5, 5, 6, 5, 8, 7, 5, 6, 5, 6, 6, 5$

➤ Σε μία στήλη (την στήλη “Παρατηρήσεις”) τοποθετούμε όλες τις δυνατές τιμές των παρατηρήσεων, από μία φορά την κάθε τιμή. (Κατηγορίες K)

Το πλήθος των κατηγοριών $K = 4$

➤ Δίπλα στη στήλη “Παρατηρήσεις” τοποθετούμε τη στήλη “Διαλογή” και στη στήλη αυτή απέναντι από κάθε παρατήρηση σημειώνουμε μία κατακόρυφη γραμμή για κάθε φορά που η παρατήρηση εμφανίζεται στο δείγμα.

➤ Η συμπλήρωση πέντε εμφανίσεων μιας συγκεκριμένης παρατήρησης σημειώνεται με διαγραφή των τεσσάρων κατακόρυφων γραμμών.

Παρατήρηση	Διαλογή
5	 I
6	 I
7	
8	

Το **πλήθος** των γραμμών που αντιστοιχούν σε κάθε παρατήρηση δείχνει πόσες φορές αυτή εμφανίζεται στο συγκεκριμένο δείγμα και καλείται **συχνότητα της παρατήρησης**,

Διαλογή - Κατανομή Συχνοτήτων - Σχετικών Συχνοτήτων

Όταν το δείγμα είναι πολύ μεγάλο και το εύρος των τιμών είναι επίσης πολύ μεγάλο $R = \max(X) - \min(X)$ ο προηγούμενος πίνακας διαλογής αντικαθίσταται από τον **πίνακα κατανομής συχνοτήτων** :

Πίνακας κατανομής συχνοτήτων Χρονική διάρκεια που απαιτήθηκε για την απόκτηση πτυχίου (έτη).

Παρατηρήσεις [x_j]	5	6	7	8
Συχνότητα [v_j]	6	16	3	5

Το σύνολο των ζευγών της μορφής (x_j, v_j) αποτελεί $\{(5,6) (6,16) (7,3) (8,5)\}$ την **κατανομή συχνοτήτων**.

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος n του δείγματος.
 Δηλαδή: $v_1 + v_2 + v_3 + \dots + v_k = n$

Παράδειγμά μας: $n = v_1 + v_2 + v_3 + v_4 = 6 + 16 + 3 + 5 = 30$

Διαλογή - Κατανομή Συχνοτήτων - Σχετικών Συχνοτήτων

Σε πολλές μελέτες και ειδικότερα όταν το δείγμα είναι πολυπληθές, για να έχουμε μια σαφή συγκριτική εικόνα για το μέρος του συνολικού πλήθους του δείγματος που καταλαμβάνει κάθε τιμή της μεταβλητής που εξετάζουμε, χρησιμοποιούμε τη σχετική συχνότητα (f_j) ή την επί τοις εκατό σχετική συχνότητα ($f_j \%$).

Διάρκεια Σπουδών [x_j]	Συχνότητα [n_j]	Σχετική συχνότητα [f_j]	Η «επί τοις εκατό» σχετική συχνότητα [$f_j \%$]
5	6	0,20	20
6	16	0,53	53
7	3	0,10	10
8	5	0,17	17
Σύνολο:	30	1,00	100

Η Σχετική συχνότητα (f_j) της τιμής x_j ορίζεται ως ο λόγος της συχνότητας n_j ως προς το μέγεθος του δείγματος, δηλαδή $f_j = n_j/n$.

Η «επί τοις εκατό» σχετική συχνότητα ($f_j \%$) της τιμής x_j ορίζεται ως εξής: $f_j \% = (n_j/n) * 100$

ΓΕΝΙΚΑ: Πως κατασκευάζουμε ένα πίνακα συχνοτήτων

Εστω $X_1, X_2, X_3, \dots, X_n$ ένα δείγμα τιμών μιας μεταβλητής X_i Y_1, Y_2, \dots, Y_k οι k διαφορετικές *κατηγορίες* που παρουσιάζει να έχει η μεταβλητή αυτή

Κατασκευάζουμε ένα πίνακα με 6 στήλες

Στη 0^η στήλη η αύξουσα *αρίθμηση* των κατηγοριών

Στη 1^η στήλη η *καταγραφή της κάθε κατηγορίας* σε αύξουσα σειρά

Στη 2^η στήλη η *διαλογή των κατηγοριών*

Στη 3^η στήλη η *συχνότητα εμφάνισης V_i* κάθε κατηγορίας σε *απόλυτο αριθμό*

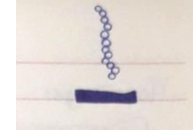
Στη 5^η στήλη η *συχνότητα εμφάνισης κάθε κατηγορίας* σε *σχετικό αριθμό* επι του συνόλου του δείγματος

$$f_i = V_i / n$$

Στη 6^η στήλη η *συχνότητα εμφάνισης κάθε κατηγορίας* σε % *σχετικό αριθμό* $f_i \% = (V_i / n) * 100$

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος n του δείγματος.

Δηλαδή: $v_1 + v_2 + v_3 + \dots + v_k = n$



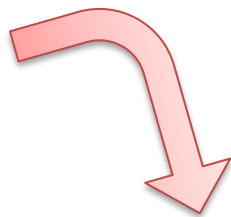
ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΑΣ

Δείγμα – δεδομένα από 10 οικογένειες

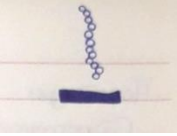
Οικογένεια	Επάγγελμα	Μηνιαίος Μισθός	Αριθμός Παιδιών
1	Εργάτης	700	0
2	Οδηγός	750	1
3	Εργάτης	800	0
4	Δημόσιος Υπάλληλος	700	2
5	Δημόσιος Υπάλληλος	800	2
6	Δημόσιος Υπάλληλος	800	2
7	Δάσκαλος	900	1
8	Ιερέας	1000	1
9	Οδηγός	600	1
10	Εργάτης	600	1

Εφαρμογή



Πίνακας Συχνοτήτων με το επάγγελμα

Αρίθμηση Κατηγοριών v Y_i	<u>Επάγγελμα</u>	Διαλογή	Συχνότητα V_i	Σχετική Συχνότητα f_i	Σχετική Συχνότητα $f_i \%$
1	Εργάτης		3	$3/10 = 0,3$	30%
2	Οδηγός		2	$2/10 = 0,2$	20%
3	Δημόσιος Υπάλληλος		3	$3/10 = 0,3$	30%
4	Δάσκαλος		1	$1/10 = 0,1$	10%
5	Ιερέας		1	$1/10 = 0,1$	10%
Σύνολο >		10	10	$10/10 = 1$	100%

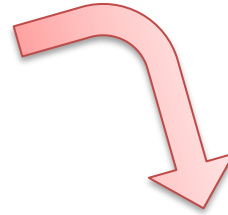


ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ

ΚΑΤΑΝΟΜΕΣ ΣΥΧΝΟΤΗΤΑΣ

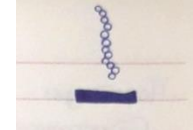
Δείγμα – δεδομένα από 10 οικογένειες

Οικογένεια	Επάγγελμα	Μηνιαίος Μισθός	Αριθμός Παιδιών
1	Εργάτης	700	0
2	Οδηγός	750	1
3	Εργάτης	800	0
4	Δημόσιος Υπάλληλος	700	2
5	Δημόσιος Υπάλληλος	800	2
6	Δημόσιος Υπάλληλος	800	2
7	Δάσκαλος	900	3
8	Ιερέας	1000	2
9	Οδηγός	600	4
10	Εργάτης	600	1



Άσκηση:

Κατασκευάστε πίνακα κατανομής συχνοτήτων με βάση το αριθμό των παιδιών



Διαλογή - Κατανομή Συχνοτήτων - Αθροιστικές Συχνότητες

Πολλές φορές μας ενδιαφέρει να μάθουμε το πλήθος ή το ποσοστό των παρατηρήσεων, που οι τιμές τους είναι *μικρότερες ή ίσες ορισμένης* τιμής x_j της *ποσοτικής μεταβλητής X*. π.χ αν θέλαμε να απαντήσουμε στο ερώτημα «πόσοι οδοντίατροι του προηγούμενου παραδείγματος πήραν πτυχίο το πολύ σε 7 χρόνια», θα υπολογίζαμε σύμφωνα με τον πίνακα τη λεγόμενη

αθροιστική συχνότητα: $N_3 = v_1 + v_2 + v_3 = 6 + 16 + 3 = 25$.

ή σε όρους σχετικής αθροιστικής συχνότητας

$F_3 \% = f_1 \% + f_2 \% + f_3 \% = 20 + 53 + 10 = 83\%$ των οδοντιάτρων.

Διάρκεια Σπουδών [x_j]	Συχνότητα [v_j]	Σχετική συχνότητα [f_j]	Η «επί τοις εκατό» σχετική συχνότητα [$f_j \%$]
5	6	0,20	20
6	16	0,53	53
7	3	0,10	10
8	5	0,17	17
Σύνολο:	30	1,00	100

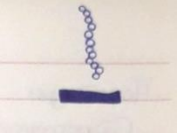
Έως εδώ



25

83%

Αθροιστικά έως εδώ

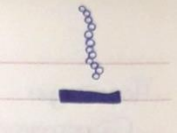


Διαλογή - Κατανομή Συχνοτήτων - Αθροιστικές Συχνότητες

Πίνακας κατανομής Αθροιστικών συχνοτήτων

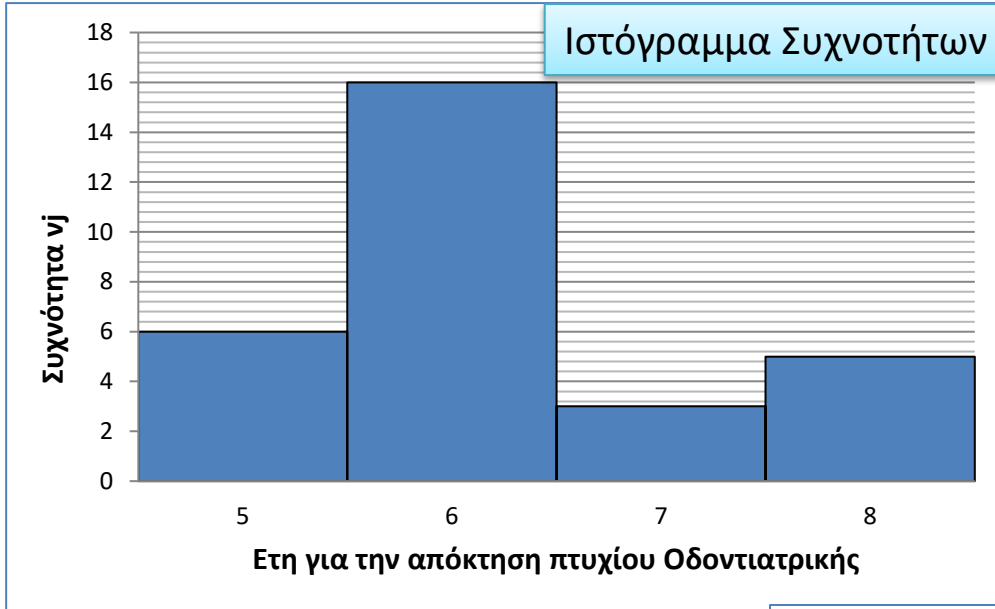
Χρονική διάρκεια που απαιτήθηκε για την απόκτηση πτυχίου Οδοντιάτρων(έτη).

Διάρκεια Σπουδών x_j	Συχνότητα n_j	Σχετική Συχνότητα f_j	Η % σχετική συχνότητα $f_j \%$	Σχετική αθροιστική συχνότητα έως n_j	Η % σχετική συχνότητα έως $f_j \%$	Σχετική αθροιστική συχνότητα πάνω από n_j	Η % σχετική συχνότητα πάνω από $f_j \%$
5	6	0,20	20	0,20	20	0,80	80
6	16	0,53	53	0,73	73	0,27	27
7	3	0,10	10	0,83	83	0,17	17
8	5	0,17	17	1,00	100	0	0
Σύνολο:	30	1,00	100				



ΠΑΡΟΥΣΙΑΣΗ ΣΤΟΙΧΕΙΩΝ “Οπτικοποίηση” – Απεικόνιση με **ΔΙΑΓΡΑΜΜΑΤΑ ΣΥΧΝΟΤΗΤΩΝ**

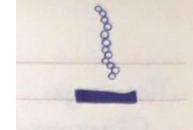
Τις περισσότερες φορές είναι πολύ χρήσιμο να έχουμε ένα είδος «εικόνας» των δεδομένων. Η πληροφορία που παίρνουμε από το σχήμα της κατανομής τους είναι σημαντική για την κατανόηση του χαρακτηριστικού που μελετάμε.



Κατανομή Απλών Συχνοτήτων για την διάρκεια σπουδών στην Οδοντιατρική	
Διάρκεια Σπουδών	Συχνότητα
x_j	n_j
5	6
6	16
7	3
8	5

Στο **ιστόγραμμα** οι **ιστοί εφάπτονται**, το **εμβαδόν** δε κάθε ιστού είναι **ανάλογο** της **συχνότητας** των **δεδομένων** με τιμές που βρίσκονται μεταξύ των δύο άκρων του διαστήματος της βάσης του ιστού.





Κατανομή Συχνοτήτων - Ομαδοποίηση των παρατηρήσεων

Στην περίπτωση διακριτής μεταβλητής, όταν το πλήθος των τιμών της είναι μεγάλο, αλλά κυρίως σε συνεχή μεταβλητή X που μπορεί να πάρει οποιαδήποτε τιμή στο διάστημα ορισμού της, ταξινομούμε τα δεδομένα σε μικρό πλήθος ομάδων που ονομάζονται **κλάσεις ή τάξεις έτσι, ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση. Ομαδοποιούμε τις παρατηρήσεις χωρίζοντας το διάστημα ορισμού (α_0, α_k) της μεταβλητής X σε κλάσεις, δηλαδή σε υποδιαστήματα της μορφής $[\alpha_{j-1}, \alpha_j)$.**

Τα άκρα των κλάσεων ονομάζονται **όρια των κλάσεων και η διαφορά: $c = \alpha_{j-1} - \alpha_j = \text{πλάτος της κλάσης } j$**

Στην πράξη συνήθως χρησιμοποιούμε κλάσεις ίσου πλάτους

Για να προσδιορίσουμε τον αριθμό των κλάσεων k χρησιμοποιούμε τον εμπειρικό τύπο, $k = 1 + 3,32 \cdot \log v$, γνωστό ως κανόνα του Sturges όπου v είναι το πλήθος των παρατηρήσεων.

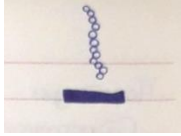
Παράδειγμα: 60 χωράφια σε στρέμματα - πρωτογενή στοιχεία

66,71	72,34	89,31	77,90	92,29	53,42
72,48	105,87	83,08	79,74	83,34	68,62
56,68	83,75	66,45	87,17	64,07	94,48
57,55	57,13	55,85	119,76	59,25	72,95
109,04	58,39	62,38	49,37	61,75	76,80
98,82	42,07	86,85	50,08	78,15	78,44
70,08	68,14	89,80	70,26	68,79	85,44
73,39	73,20	78,34	86,38	56,34	39,20
62,04	94,75	104,35	77,44	77,82	61,63
44,12	96,37	61,12	78,21	83,42	79,13



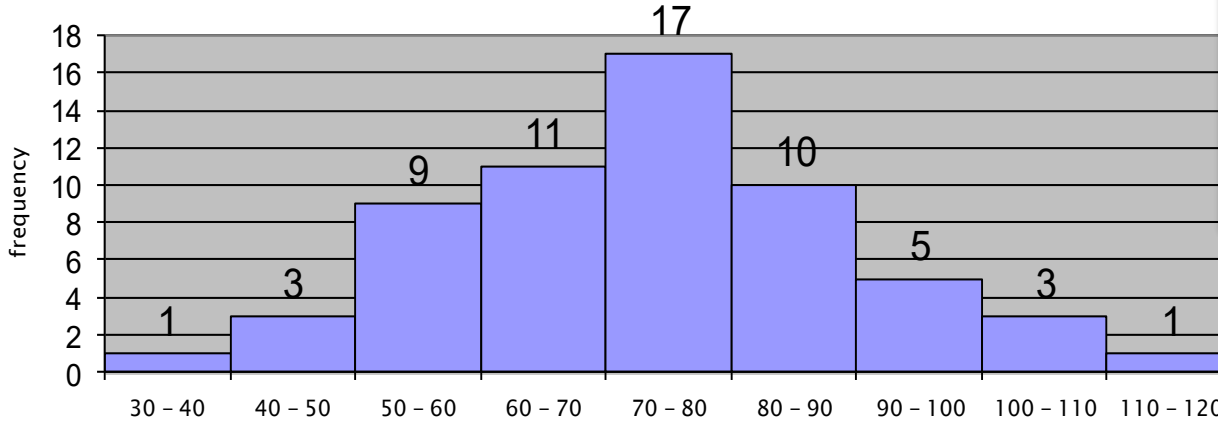
60 χωράφια οργανωμένα σε ομαδοποιημένη κατανομή συχνότητας

Τάξεις (διαστήματα) σε στρέμματα	Μέγεθος (ευρος) κάθε διαστήματος	Πλήθος περιουσιών σε κάθε τάξη (διάστημα)	Πλήθος χωραφιών σε κάθε διάστημα / τάξη / Συχνότητα
30 -< 40	10	/	1
40 -< 50	10	///	3
50 -< 60	10	//// // /	9
60 -< 70	10	//// // //	11
70 -< 80	10	///// // // // //	17
80 -< 90	10	///// // //	10
90 -< 100	10	/////	5
100 -< 110	10	///	3
110 -< 120	10	/	1
Total			60



Ορισμός: *Ένα Ιστόγραμμα είναι μια γραφική απεικόνιση μιας κατανομής συχνότητας*

ΙΣΤΟΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΑΣ



Γενικά - Ένα διάγραμμα:

- Δίνει μια οπτική απεικόνιση των στοιχείων
- Είναι ίσως λιγότερο αναλυτικό από ένα πίνακα αλλά αναδεικνύει ένα η δυο «σημεία κλειδιά» των στοιχείων

Παράδειγμα: 60 χωράφια σε ομαδοποιημένη κατανομή συχνότητας

Μια πολυγωνική γραμμή σχεδιάζεται από ένα ιστόγραμμα ενώνοντας τις κορυφές κάθε ράβδου με μια ευθεία γραμμή

