

παρουσίαση στατιστικών δεδομένων

- Πίνακες
- Διαγράμματα
- Στατιστικές εκθέσεις
- Ασκήσεις εφαρμογές

Τρόποι παρουσίασης δεδομένων

Σημαντικό μέρος της περιγραφικής, στατιστικής αναφέρεται στον τρόπο παρουσίασης των πληροφοριών, ώστε να κινούν το ενδιαφέρον, και να εξαγονται εύκολα τα συμπεράσματα

Οι τρόποι παρουσίασης είναι:

- Στατιστικοί πίνακες
- Στατιστικά διαγράμματα
- Στατιστικές εκθέσεις

πίνακες

- Παρουσιάζουμε τα δεδομένα σε μορφή γραμμών και στηλών, όπου αναφέρουμε τις τιμές της μεταβλητής και τις συχνότητες (ή τα ποσοστά) εμφάνισής τους.
- Οι πιο γνωστοί πίνακες είναι οι πίνακες συχνοτήτων:
 - για μία μεταβλητή (απλοί πίνακες)
 - για δύο μεταβλητές (πίνακες διπλής εισόδου)
 - για περισσότερες μεταβλητές (πολλαπλοί πίνακες)

Πλεονεκτήματα	Μειονεκτήματα
Πληρότητα	Δεν ελκύει το ενδιαφέρον
Ακρίβεια	Δυσκολία συγκρίσεων

Στοιχεία πίνακα

Σε κάθε πίνακα που παρουσιάζει στατιστικά δεδομένα, θα πρέπει να υπάρχουν τα παρακάτω στοιχεία:

- Επικεφαλίδα-τίτλος πίνακα
- Κύριο μέρος με γραμμές στήλες
- Υποσημειώσεις πίνακα
- Πηγή συλλογής δεδομένων

Παράδειγμα πολλαπλού πίνακα

Ι. ΠΛΗΘΥΣΜΟΣ

Ια. Πραγματικός πληθυσμός της Ελλάδας κατά φύλο
και ομάδες ηλικιών

ΑΡΙΘΜΟΣ
ΠΙΝΑΚΑ

ΤΙΤΛΟΣ

ΕΠΙΚΕΦΑΛΙΔΕΣ
ΣΤΗΛΩΝ

ΚΟΡΜΟΣ
ΚΥΡΙΟ
ΣΩΜΑ

Απογραφές

1971(1)

1981(1)

1991(1)

2001(2)

Ελλάδος	8.768.372	9.739.589	10.259.900	10.964.020
Άρρενες	4.286.748	4.779.571	5.055.408	5.431.816
Θήλειες	4.481.624	4.960.018	5.204.492	5.532.204
0 – 14 ετών	2.223.904	2.307.297	1.974.867	1.666.888
15 – 64 ετών	5.587.352	6.192.751	6.880.681	7.423.889
65 ετών και άνω	957.116	1.239.541	1.404.352	1.873.243
Ποσοστιαία κατανομή %				
Άρρενες	48,89	49,07	49,27	49,54
Θήλειες	51,11	50,93	50,73	50,45
0 – 14 ετών	25,36	23,69	19,25	15,20
15 – 64 ετών	63,72	63,58	67,06	67,71
65 ετών και άνω	10,92	12,73	13,69	17,08

ΥΠΟΣΗΜΕΙΩΣΕΙΣ

(1) Ηλεκτρονική επεξεργασία 25% και 10% των ερωτηματολογίων των απογραφών 1971 και 1981, αντίστοιχα, και καθολική επεξεργασία των ερωτηματολογίων της απογραφής 1991.

ΠΗΓΗ

(2) Οριστικά στοιχεία για το σύνολο της Ελλάδος και προσωρινά κατά φύλο και ομάδες ηλικιών.

Πηγή: Ε.Σ.Υ.Ε. «Η Ελλάδα με αριθμούς», 2003, σελ. 3

Παρουσίαση μιας μεταβλητής με πίνακα

Ποιοτική μεταβλητή

Περιοχή καταγωγής

Τιμές περιοχή καταγωγής	Συχνότητα	Σχετική συχνότητα	Αθροιστική συχνότητα	Σχετική αθροιστική συχνότητα
Θράκη	15	$15/80=0,2$ 20%	15	20%
Μακεδονία	24	$24/80=0,3$ 30%	39	50%
Ήπειρος	12	$12/80=0,15$ 15%	51	65%
Θεσσαλία	10	$10/80=0,125$ 12,5%	61	77,5%
Στερεά Ελλάδα	8	$8/80=0,1$ 10%	69	87,5%
Υπόλοιπη Ελλάδα	11	$11/80=0,125$ 12,5%	80	100%
σύνολο	80	1,0 100%		

Πηγή: δεδομένα προσομοίωσης

Παρουσίαση μιας μεταβλητής με πίνακα

Ποσοτική μεταβλητή ασυνεχής

Ημέρες άδειας εργαζομένων

Τιμές πλήθος ημερών άδειας	Συχνότητα	Σχετική συχνότητα	Αθροιστική συχνότητα	Σχετική αθροιστική συχνότητα
3	8	$8/80=0,1$ 10%	8	10%
5	24	$24/80=0,3$ 30%	32	40%
7	20	$18/80=0,25$ 25%	52	65%
10	16	$18/80=0,20$ 20%	68	85%
15	8	$8/80=0,1$ 10%	76	95%
20	4	$4/80=0,05$ 5%	80	100%
σύνολο	80	1,0 100%		

Πηγή: δεδομένα προσομοίωσης

Παρουσίαση μιας μεταβλητής με πίνακα

Ποσοτική μεταβλητή συνεχής

Ετήσιο εισόδημα εργαζομένων

Τιμές ετήσιο εισόδημα	Συχνότητα	Σχετική συχνότητα	Αθροιστική συχνότητα	Σχετική αθροιστική συχνότητα
0-5.000	8	$8/80=0,1$ 10%	8	10%
5.000-15.000	24	$24/80=0,3$ 30%	32	40%
15.000-20.000	20	$18/80=0,25$ 25%	52	65%
20.000-30.000	16	$18/80=0,20$ 20%	68	85%
30.000-40.000	8	$8/80=0,1$ 10%	76	95%
40.000-60.000	4	$4/80=0,05$ 5%	80	100%
σύνολο	80	1,0 100%		

Πηγή: δεδομένα προσομοίωσης

Πίνακας διπλής εισόδου

- Μπορούμε να παρουσιάσουμε κάθε μία από τις δύο μεταβλητές με ένα διαφορετικό πίνακα και να έχουμε δύο πίνακες συχνότητων. Με τον τρόπο αυτό όμως δε παρουσιάζεται η συχνότητα ταυτόχρονης εμφάνισης μιας ιδιότητας της πρώτης μεταβλητής με μια ιδιότητα της δεύτερης μεταβλητής.
- Για να παρουσιάσουμε ακριβώς την συχνότητα που συμπίπτει να εμφανιστεί η ιδιότητα της πρώτης μεταβλητής με μια τιμή ή ιδιότητα της δεύτερης μεταβλητής χρησιμοποιούμε τον πίνακα διπλής εισόδου, που ονομάζεται και πίνακας συμπτώσεων.
- Στον πίνακα αυτό, στην πρώτη στήλη γράφουμε τις τιμές της πρώτης μεταβλητής, ενώ στην πρώτη γραμμή της τιμές της δεύτερης μεταβλητής. Στην διασταύρωση μιας γραμμής της τιμής i (πρώτης μεταβλητής) και μιας στήλης της τιμής j (δεύτερης μεταβλητής) γράφουμε το πλήθος των φορών (συχνότητα) που εμφανίστηκαν συγχρόνως στα δεδομένα μας η τιμή i και η τιμή j . Η διασταύρωση αυτή ονομάζεται κελί ij του πίνακα διπλής εισόδου.

Δημιουργήστε πίνακα διπλής εισόδου από
τα παρακάτω δεδομένα δύο μεταβλητών

φύλο	Μέρες άδειας
------	-----------------

Άνδρας	4
--------	---

Γυναίκα	4
---------	---

Άνδρας	2
--------	---

Γυναίκα	4
---------	---

Γυναίκα	2
---------	---

Άνδρας	0
--------	---

Γυναίκα	2
---------	---

Άνδρας	4
--------	---

Άνδρας	3
--------	---

Γυναίκα	3
---------	---

Άνδρας	4
--------	---

Γυναίκα	0
---------	---

Γυναίκα	3
---------	---

Άνδρας	2
--------	---

Γυναίκα	2
---------	---

Άνδρας	0
--------	---

Άνδρας	0
--------	---

Άνδρας	4
--------	---

Γυναίκα	3
---------	---

Άνδρας	3
--------	---

Φύλο και μέρες άδειας
εργαζομένων

	0	2	3	4	συνολο
άνδρας	3	2	2	4	11
γυναίκα	1	3	3	2	9
σύνολο	4	5	5	6	20

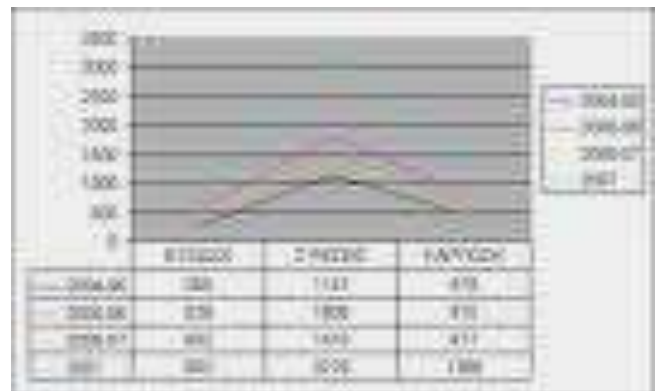
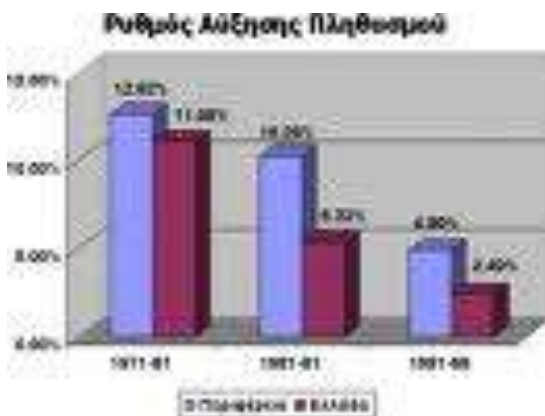
Πηγή: δεδομένα προσομοίωσης

Στον παραπάνω πίνακα,
μπορούμε πολύ ευκολότερα
να εξετάσουμε τη σχέση
φύλου και ημερών άδειας, ή
να κάνουμε συγκρίσεις
μελετώντας ποσοστά

διαγράμματα

- Παρουσιάζουν τα στοιχεία με εικόνες ώστε να ελκύουν το ενδιαφέρον.

Πλεονεκτήματα	Μειονεκτήματα
ελκυστικά	Όχι ακρίβεια στοιχείων
Εύκολα στις συγκρίσεις	Δυσκολία δημιουργίας

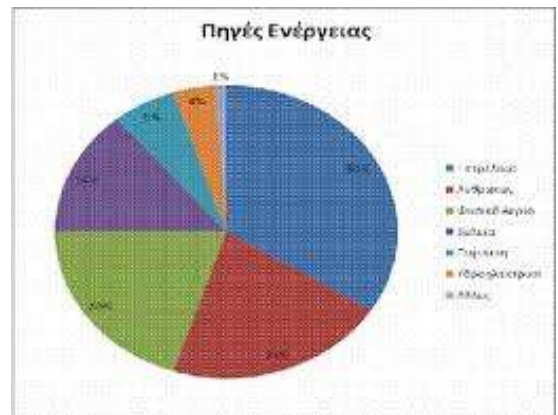
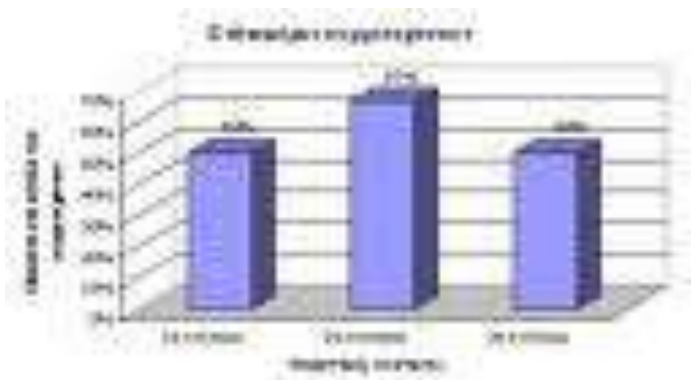
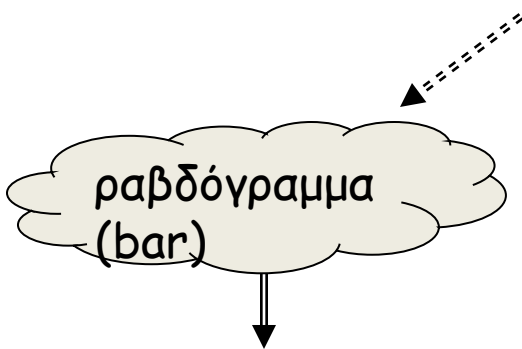


Είδη διαγραμμάτων

- Ραβδόγραμμα
- Κυκλικό διάγραμμα
- Ιστόγραμμα
- Χρονοδιάγραμμα
- Χαρτόγραμμα
- Ειδικό διάγραμμα

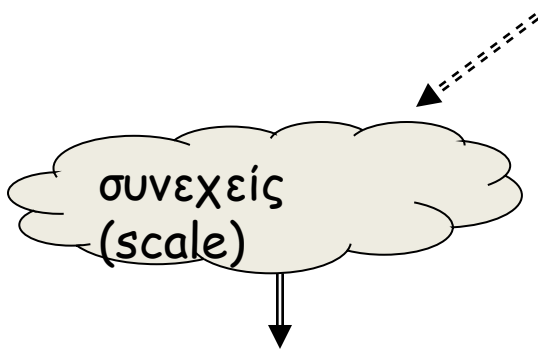
Ποιοτικές μεταβλητές

- Παρουσιάζονται με:

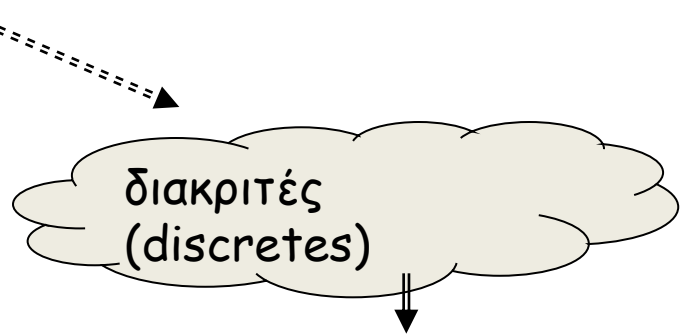
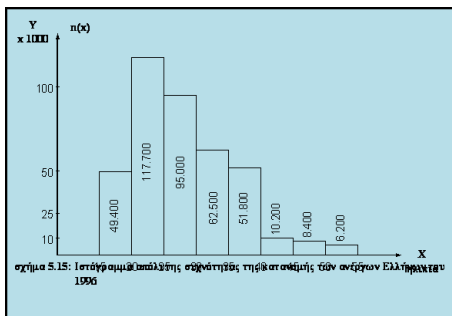


Ποσοτικές μεταβλητές

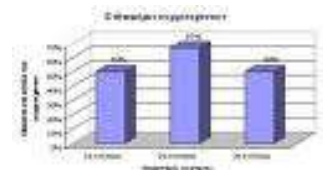
- Παρουσιάζονται με :



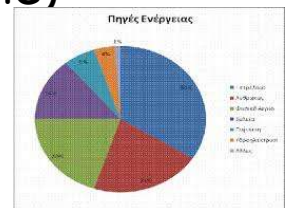
Ιστόγραμμα
(histogram)



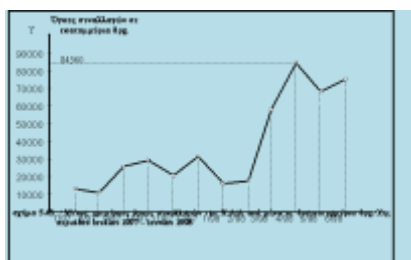
Ραβδόγραμμα (bar)



Κυκλικό (pie)

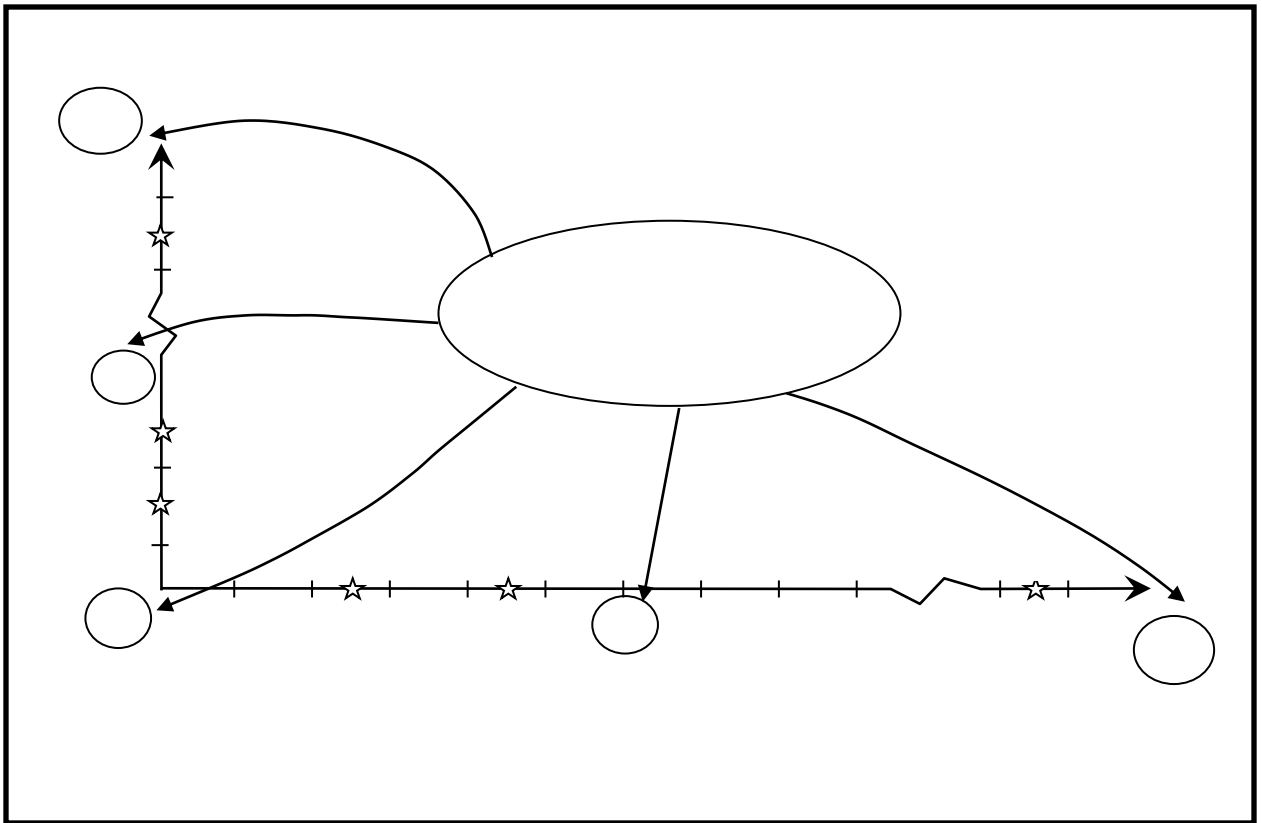


Χρονοδιάγραμμα (όταν
αναφέρονται στο χρόνο)

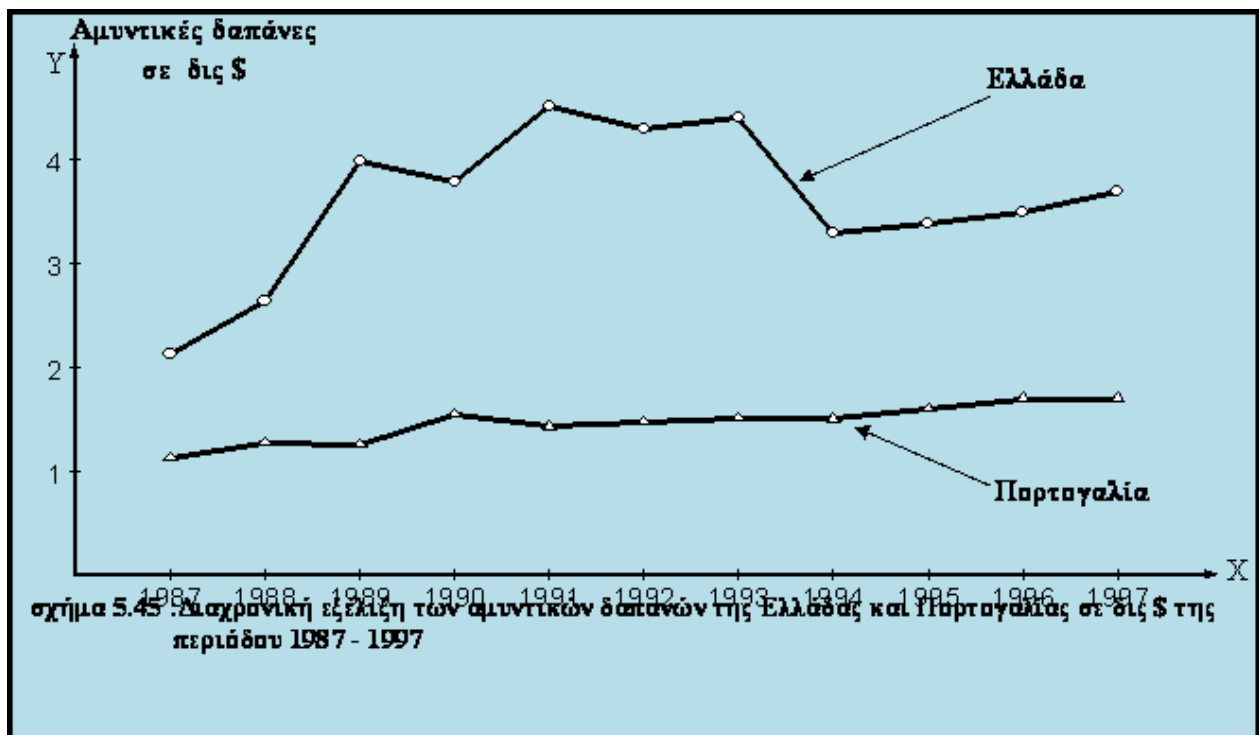
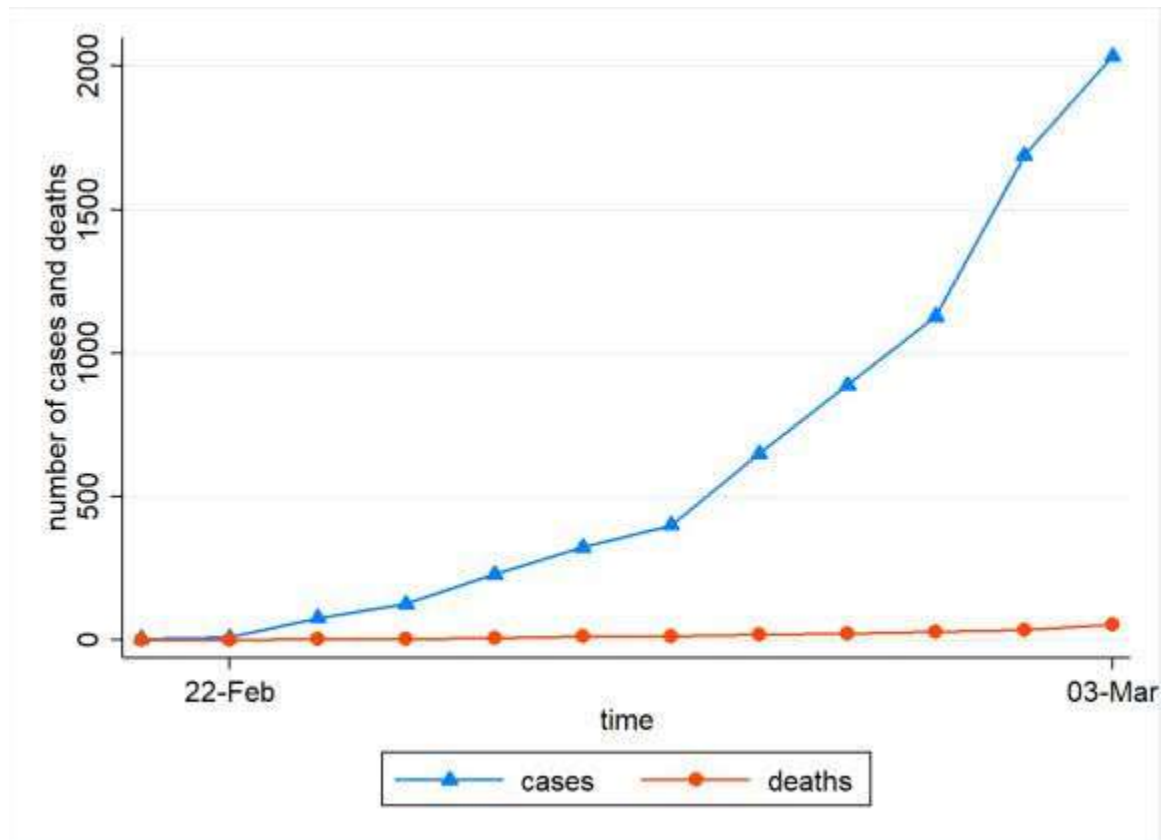


Διαγράμματα σε καρτεσιανό σύστημα

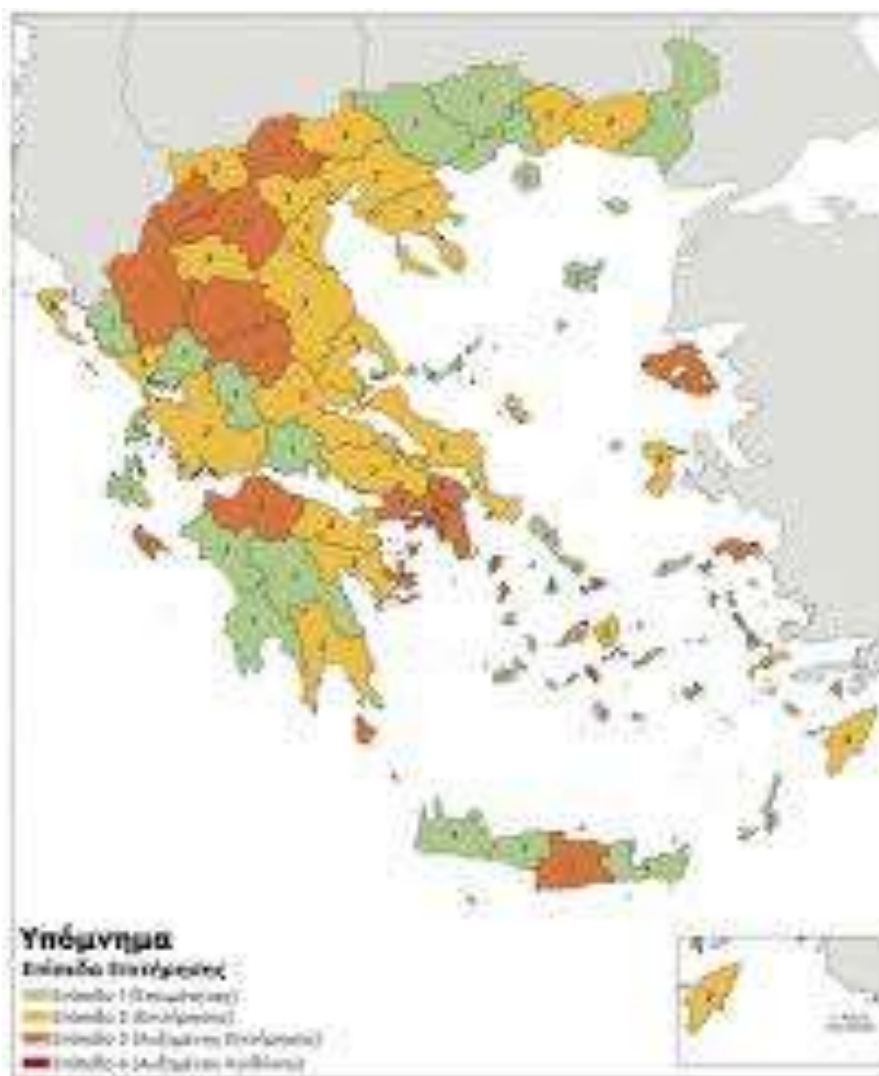
Στην περίπτωση ιστογράμματος ή
χρονοδιαγράμματος, χρησιμοποιούμε το
καρτεσιανό σύστημα και προσέχουμε τα
εξής:



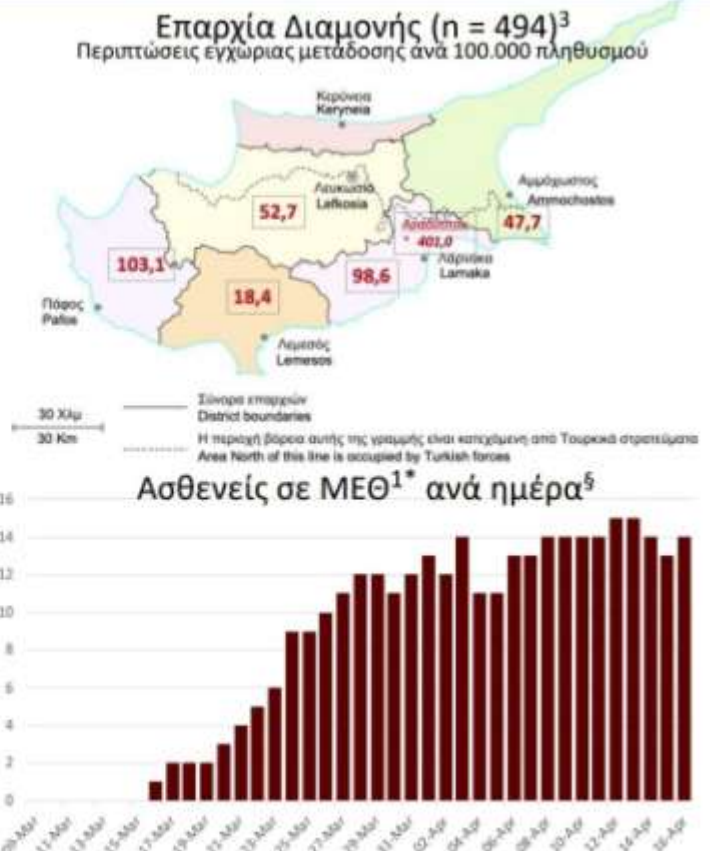
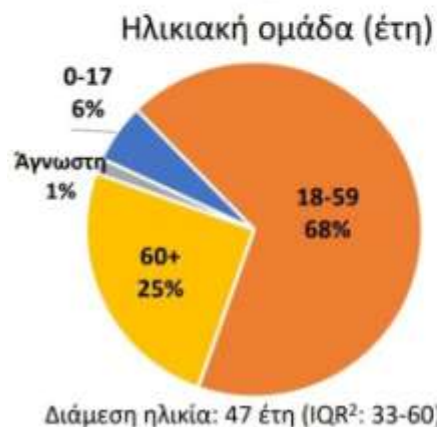
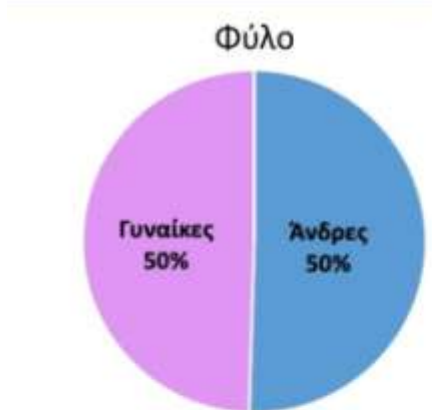
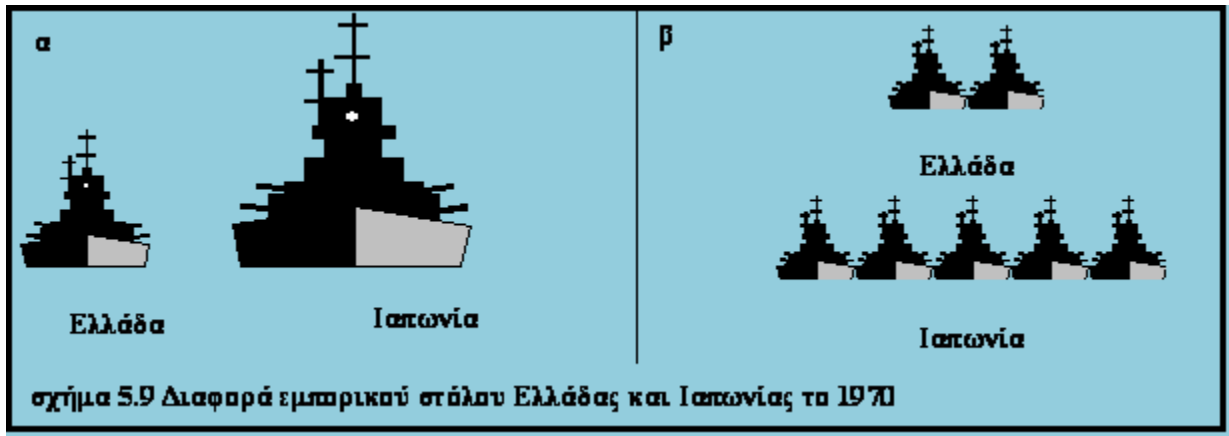
χρονοδιάγραμμα



χαρτογράμματα



Ειδικά διαγράμματα



^{1*}ΜΕΘ: Μονάδα Εντατικής Θεραπείας, δεδομένα υπό επεξεργασία, ²IQR: Ενδοτεταρτημοριακό εύρος, ³Περιπτώσεις εγχώριας μετάδοσης μόνο. Περιπτώσεις με πρόσφατο ιστορικό ταξιδιού εξαιρούνται, ⁴Συμπεριλαμβάνεται η ημέρα εξιτηρίου/θανάτου

σχολιάστε

Ι. ΠΛΗΘΥΣΜΟΣ

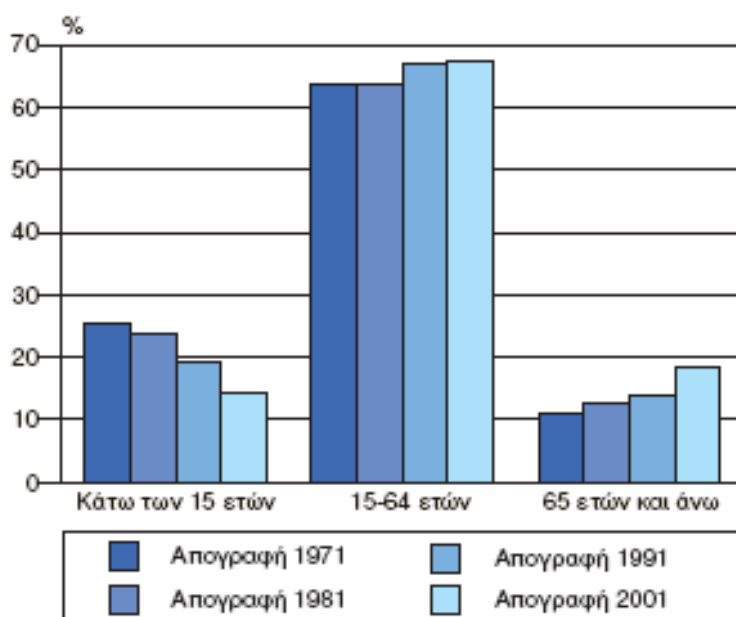
Ια. Πραγματικός πληθυσμός της Ελλάδας κατά φύλο και ομάδες ηλικιών

	Απογραφές			
	1971(1)	1981(1)	1991(1)	2001(2)
Σύνολο Ελλάδος	8.768.372	9.739.589	10.259.900	10.964.020
Άρρενες	4.286.748	4.779.571	5.055.408	5.431.818
Θήλεις	4.481.624	4.960.018	5.204.492	5.532.204
0 – 14 ετών	2.223.904	2.307.297	1.974.867	1.666.888
15 – 64 ετών	5.587.352	6.192.751	6.880.681	7.423.889
65 ετών και άνω	957.116	1.239.541	1.404.352	1.873.243
Ποσοστιαία κατανομή %				
Άρρενες	48,89	49,07	49,27	49,54
Θήλεις	51,11	50,93	50,73	50,45
0 – 14 ετών	25,36	23,69	19,25	15,20
15 – 64 ετών	63,72	63,58	67,06	67,71
65 ετών και άνω	10,92	12,73	13,69	17,08

(1) Δειγματοληπτική επεξεργασία 25% και 10% των ερωτηματολογίων των απογραφών 1971 και 1981, αντίστοιχα, και καθολική επεξεργασία των ερωτηματολογίων της απογραφής 1991.

(2) Οριστικά στοιχεία για το σύνολο της Ελλάδας και προσωρινά κατά φύλο και ομάδες ηλικιών.

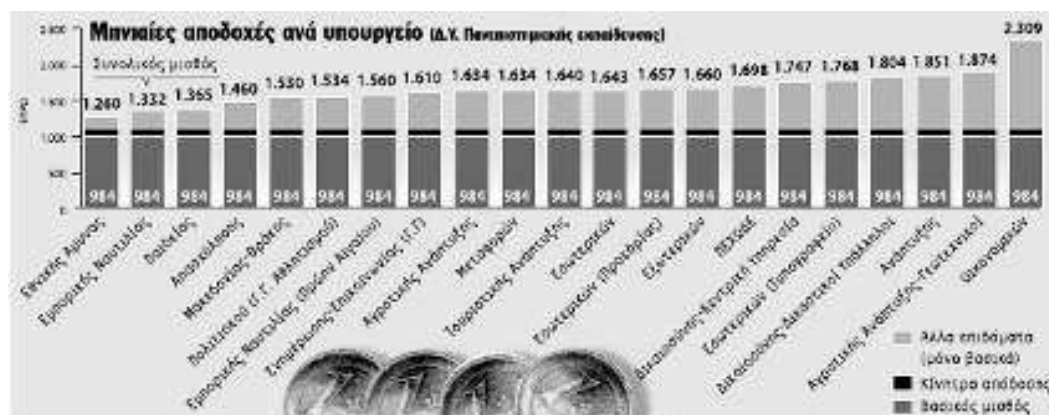
Ποσοστιαία κατανομή του πληθυσμού κατά ομάδες ηλικιών



Συγκρίση μισθολογικού κόστους στον δημόσιο και τον ιδιωτικό τομέα



ΠΗΓΗ: ΚΕΠΕ, Εθνικοί Λογαριασμοί



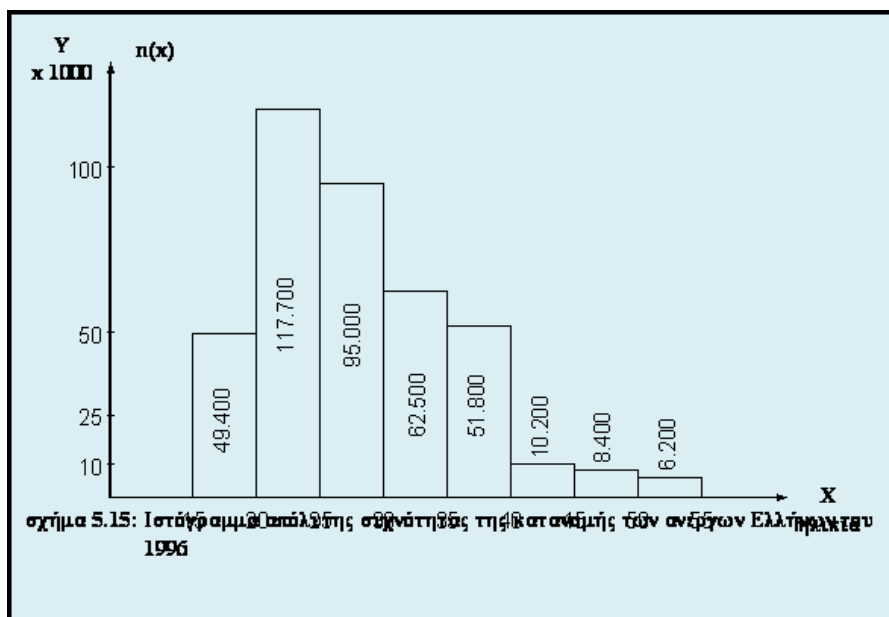
Μισθολογική δαπάνη ανά υπουργείο, 2007



Πηγή: Γενικό Λογιστήριο του Κράτους

παρουσίαση συνεχών ποσοτικών μεταβλητών

- Πίνακες με κλάσεις τιμών
- Ιστόγραμμα
- Παραπλάνηση



Πίνακες κλάσεων (διαστημάτων)

- Παρουσιάζουμε τα δεδομένα σε μορφή γραμμών και στηλών, όπου αναφέρουμε τις τιμές της μεταβλητής και τις συχνότητες (ή τα ποσοστά) εμφάνισής τους.
- Στην περίπτωση συνεχών ποσοτικών μεταβλητών, επειδή οι τιμές είναι πολλές για να γραφεί κάθε μία σε μια γραμμή του πίνακα, δημιουργούμε διαστήματα (κλάσεις) τιμών.
- Το πόσες κλάσεις θα δημιουργήσουμε εξαρτάται από την ακρίβεια που επιθυμούμε να διατηρήσουμε. Περισσότερες κλάσεις-μεγαλύτερη ακρίβεια αλλά πολλές κλάσεις προκαλούν δυσκολίες επεξεργασίας.
- Οι κλάσεις μπορεί να έχουν ίσο πλάτος ή και διαφορετικό πλάτος. Εξαρτάται από το είδος δεδομένων που έχουμε.
- Σε περίπτωση πίνακα με άνισα πλάτη κλάσεων, προσέχουμε πολύ όταν δημιουργούμε το ιστόγραμμα για να μην μας παραπλανήσει. Θα πρέπει το εμβαδόν των ιστών να είναι ανάλογο με τη συχνότητα των κλάσεων. (όχι μόνο το ύψος ανάλογο όπως συμβαίνει με τις κλάσεις ίσου πλάτους).

ιστόγραμμα

Τι είναι ένα ιστόγραμμα;

- Μια αναπαράσταση των δεδομένων που δείχνει πόσες από τις παρατηρήσεις μας βρίσκονται μέσα σε κάποια διαστήματα τιμών.

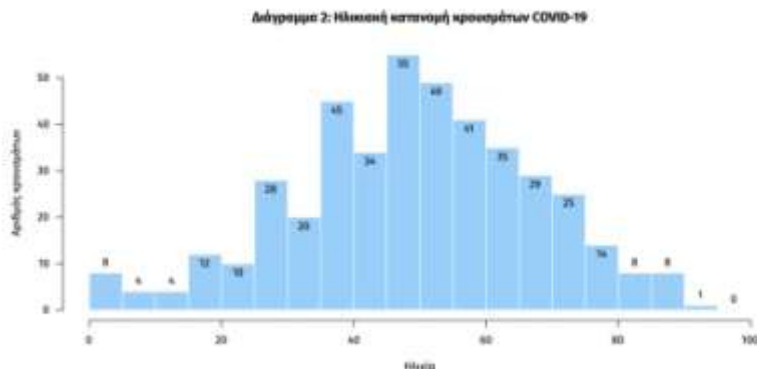
Πως φτιάχνουμε ένα ιστόγραμμα;

- 1. Διαιρούμε το εύρος των δεδομένων σε κλάσεις τιμών κατά προτίμηση ίσου πλάτους.
- 2. Μετρούμε το πλήθος δεδομένων που ανήκει σε κάθε κλάση.
- 3. Σχεδιάζουμε το ιστόγραμμα χρησιμοποιώντας δύο κάθετους άξονες. Τοποθετούμε τις κλάσεις συνήθως στον οριζόντιο άξονα, και τις συχνότητες τους στον κατακόρυφο άξονα.

Για ποιο λόγο χρησιμοποιούμε ένα ιστόγραμμα;

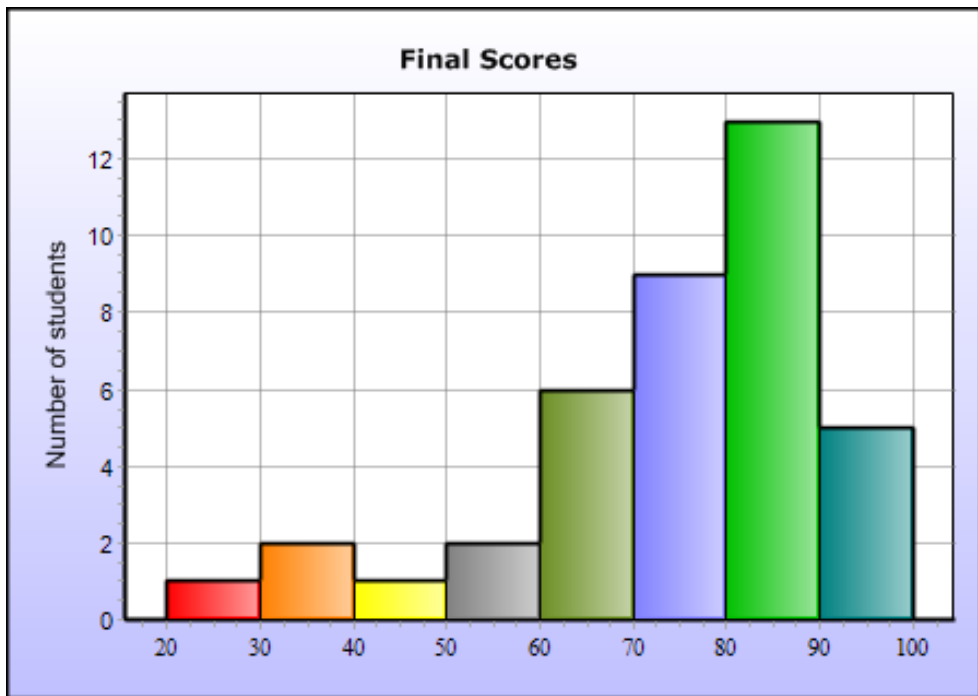
- Το ιστόγραμμα αναπαριστά ένα μεγάλο σύνολο δεδομένων, με τρόπο που να κατανοούμε εύκολα τα βασικά τους χαρακτηριστικά

<http://www.connectseward.org/shs/da1/review07/Histogram2.htm>



Παράδειγμα- ιστόγραμμα

Τιμές (κλάσεις)	συχνότητα
20-30	1
30-40	2
40-50	1
50-60	2
60-70	6
70-80	9
80-90	13
90-100	5
σύνολο	39



http://www.cyber-wit.com/gallery_histogram.html

Παραπλανητικό ιστόγραμμα

- Διαστρέβλωση, πολλές φορές, της πραγματικότητας, γίνεται από κακή χρήση της στατιστικής.
- Δυστυχώς από άγνοια ή σκοπιμότητα, ολοένα και πιο συχνά, γίνεται εσφαλμένη χρήση των στατιστικών μεθόδων παρουσίασης, ανάλυσης και ερμηνείας δεδομένων με αποτέλεσμα να δημιουργείται ψευδής και στρεβλή εικόνα για την πραγματικότητα.
- Οι εντυπώσεις όμως μένουν! Φυσικά δεν ευθύνεται η στατιστική!

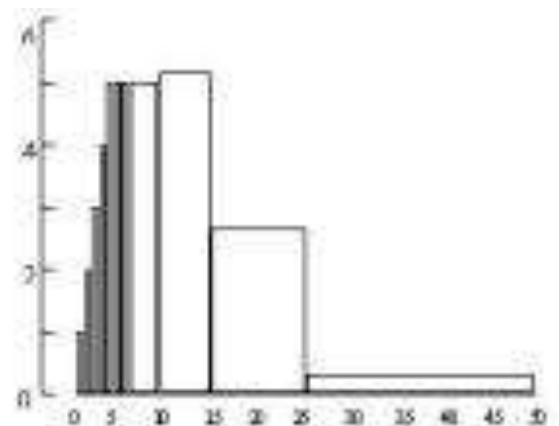
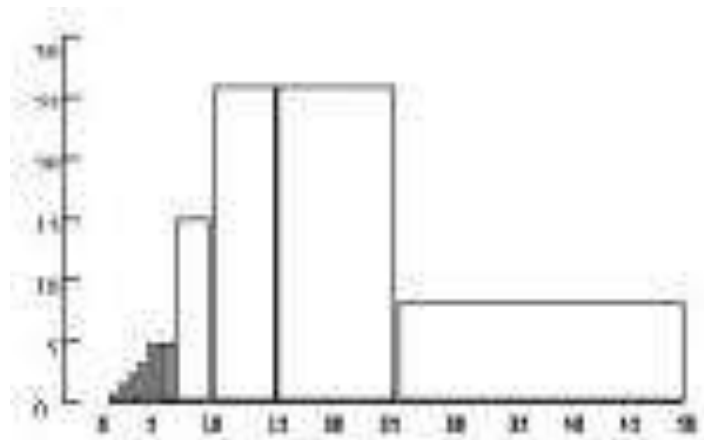
Είναι γνωστή άλλωστε η φράση:

figures do not lie, but liars figure

(Τα σχήματα δεν ψεύδονται αλλά οι ψεύτες σχεδιάζουν)

Παράδειγμα μη σωστού ιστογράμματος

- Τα παρακάτω ιστογράμματα κατασκευάστηκαν για να περιγράψουν την κατανομή των οικογενειών στις ΗΠΑ ως προς το ετήσιο εισόδημά τους το 1973.
- Το πρώτο ιστογράμμο είναι παραπλανητικό καθώς δημιουργεί την εντύπωση ότι η οικονομική κατάσταση των οικογενειών στις ΗΠΑ είναι καλύτερη από την πραγματική. Φαίνεται ότι οι οικογένειες που έχουν εισόδημα μεγαλύτερο από 25.000 δολάρια είναι περισσότερες από αυτές που έχουν εισόδημα μικρότερο από 7.000 δολάρια.
- Βέβαια τα πραγματικά δεδομένα λένε το ακριβώς αντίθετο (8% και 25% αντίστοιχα).
- Ο δημιουργός του ιστογράμματος είτε δεν γνώριζε είτε αγνόησε σκοπίμως το στοιχειώδες ότι στην κατασκευή ενός ιστογράμματος τα εμβαδά των ορθογωνίων και όχι τα ύψη αντιστοιχούν στις συχνότητες (απόλυτες ή σχετικές αναλόγως). Το σωστό ιστογράμμο είναι το δεύτερο που αποδίδει και τη σωστή εικόνα της κατανομής!



Γ.Παπαδόπουλος, Εργαστήριο Μαθηματικών & Στατιστικής, ΓΠΑ

<http://antirisis.wordpress.com/2010/01/13/%CF%85%CF%80%CE%AC%CF%81%CF%87%CE%BF%CF%85%CE%BD-%CF%88%CE%AD%CE%BC%CE%BC%CE%B1%CF%84%CE%B1-%CE%BC%CE%B5%CE%B3%CE%AC%CE%BB%CE%B1-%CF%88%CE%AD%CE%BC%CE%B1%CF%84%CE%B1-%CE%BA%CE%B1%CE%B9-%CE%B7-%CF%83/>
Προσπελάστηκε στις 11-3-2011

Στατιστικές εκθέσεις

- Είναι οι αναφορές για στατιστικά δεδομένα, οι οποίες παρουσιάζουν πίνακες και διαγράμματα με ταυτόχρονο σχολιασμό τους και παροχή επιπλέον σχολίων.
- Στις στατιστικές εκθέσεις μπορεί να αναφέρεται και η κρίση του συγγραφέα, ή συγκριτικά στοιχεία από άλλες στατιστικές εκθέσεις.
- Κύριο χαρακτηριστικό τους είναι ο συνδυασμός διαγραμμάτων, πινάκων και σχολίων – συμπερασμάτων, ώστε να κατανοηθεί καλύτερα ένα φαινόμενο.

Παρουσιάστε με διάγραμμα

κέρδη (χιλ. €)	Συχνότητα
0-10	10
10-20	20
20-40	30
40-60	15
60-100	20
100-200	30
Σύνολο	125

κέρδη (χιλ. €)	Συχνότητα	Υψος ιστογράμματος
0-10	10	
10-20	20	
20-40	30	
40-60	15	
60-100	20	
100-200	30	
Σύνολο	125	

Παρουσιάστε με διάγραμμα

Πίνακας συχνοτήτων

τιμές	συχνότητες
κίτρινο	3
κόκκινο	2
μαύρο	2
πράσινο	3
Σύνολο	10

Κοινόχρηστα (€)	Συχνότητα
80-100	4
100-120	7
120-140	9
140-160	13
160-180	9
180-200	5
200-220	3
Σύνολο	50

Άσκηση

Εξηγείστε με ποιο διάγραμμα θα παρουσιάσετε τα παρακάτω στατιστικά δεδομένα:

- Τα δεδομένα πληθυσμού των διαφόρων πόλεων της Ελλάδας
- Τα δεδομένα προτίμησης οδοντόκρεμας του πληθυσμού της Ελλάδας
- Τα δεδομένα για το μέγεθος σεισμών της Ελλάδας
- Τα δεδομένα για τον χρόνο εξυπηρέτησης ενός πελάτη σε μια τράπεζα
- Τα δεδομένα για τη διάρκεια ζωής ηλεκτρικών λαμπτήρων
- Τα δεδομένα για την πρόβλεψη των εκλογικών αποτελεσμάτων
- Τα δεδομένα για την βαθμολογία ενός μαθήματος στην εξεταστική
- Τα δεδομένα για το ποσοστό επιτυχίας όλων των μαθημάτων σε μια εξεταστική

Σχολιάστε με μια στατιστική έκθεση

Ι. ΠΛΗΘΥΣΜΟΣ

Ια. Πραγματικός πληθυσμός της Ελλάδας κατά φύλο και ομάδες ηλικιών

	Απογραφές			
	1971(1)	1981(1)	1991(1)	2001(2)
Σύνολο Ελλάδος	8.768.372	9.739.589	10.259.900	10.964.020
Άρρενες	4.286.748	4.779.571	5.055.408	5.431.816
Θήλειες	4.481.624	4.960.018	5.204.492	5.532.204
0 – 14 ετών	2.223.904	2.307.297	1.974.867	1.666.888
15 – 64 ετών	5.587.352	6.192.751	6.680.681	7.423.889
65 ετών και άνω	957.116	1.239.541	1.404.352	1.873.243
Ποσοστιαία κατανομή %				
Άρρενες	48,89	49,07	49,27	49,54
Θήλειες	51,11	50,93	50,73	50,45
0 – 14 ετών	25,36	23,69	19,25	15,20
15 – 64 ετών	63,72	63,58	67,06	67,71
65 ετών και άνω	10,92	12,73	13,69	17,08

(1) Δειγματοληπτική επεξεργασία 25% και 10% των ερωτηματολογίων των απογραφών 1971 και 1981, αντίστοιχα, και καθολική επεξεργασία των ερωτηματολογίων της απογραφής 1991.

(2) Οριστικά στοιχεία για το σύνολο της Ελλάδας και προσωρινά κατά φύλο και ομάδες ηλικιών.

Ποσοστιαία κατανομή του πληθυσμού κατά ομάδες ηλικιών

