



Καθηγητής Ι. Μητρόπουλος

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ

Τηλ.: +030 2610 369213, email: imitro@upatras.gr

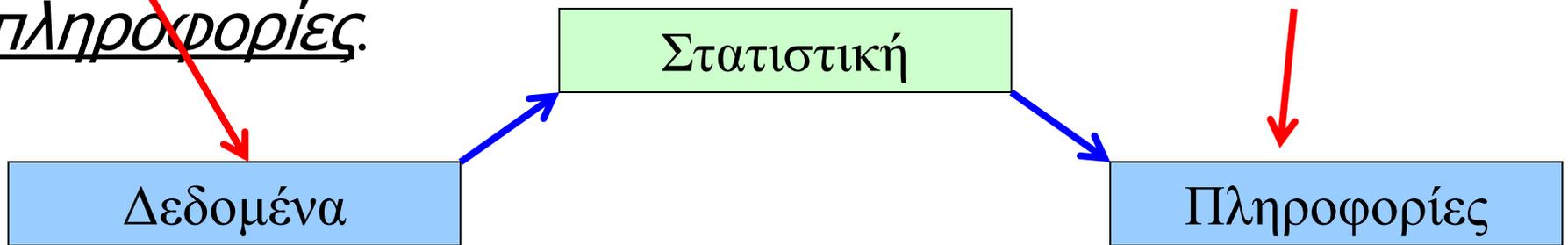
Διεύθυνση: Μεγάλου Αλεξάνδρου 1, 263 34 ΠΑΤΡΑ

Ποσοτικές Μέθοδοι στην Οικονομία και Διοίκηση (I)

ΕΝΟΤΗΤΑ 2^η : Περιγραφικές Τεχνικές & Δειγματοληψία

Εισαγωγή

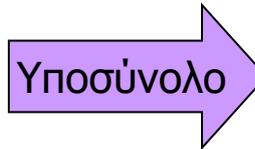
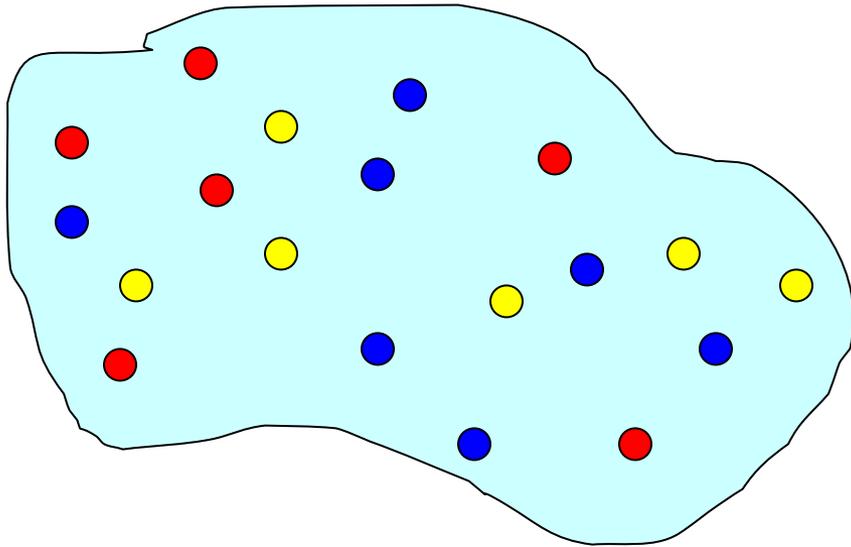
Η **περιγραφική στατιστική** ασχολείται με την αναδιάταξη, τη σύνοψη και την παρουσίαση των δεδομένων με τρόπο που παράγει χρήσιμες πληροφορίες.



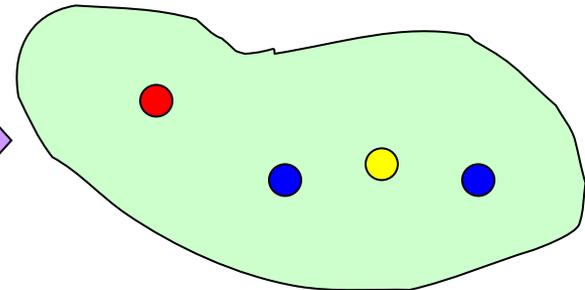
Στις μεθόδους περιλαμβάνονται γραφικές τεχνικές και αριθμητικοί δείκτες (όπως ο αριθμητικός μέσος) για τη σύνοψη και παρουσίαση των δεδομένων.

Πληθυσμοί & Δείγματα

Πληθυσμός



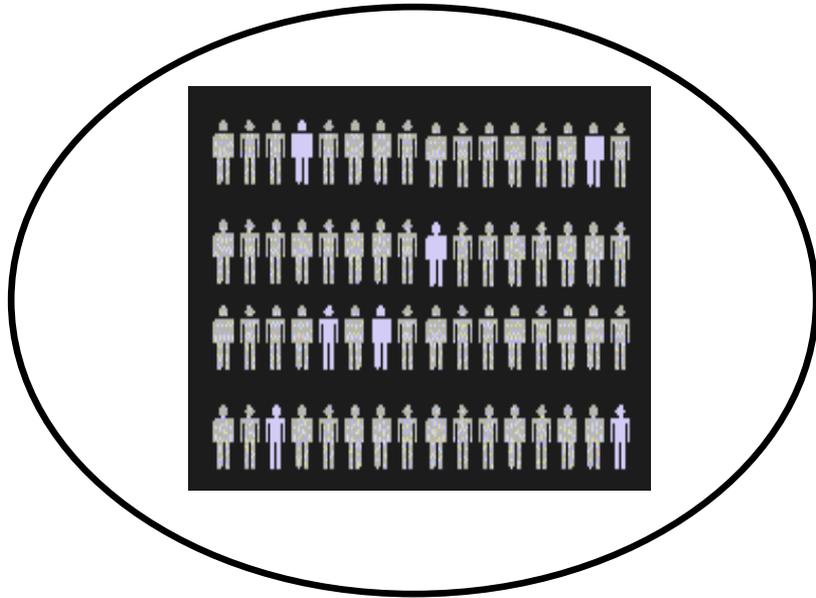
Δείγμα



Οι γραφικές μέθοδοι που θα παρουσιάσουμε εφαρμόζονται σε ολόκληρους πληθυσμούς **και** σε δείγματα που προέρχονται από αυτούς.

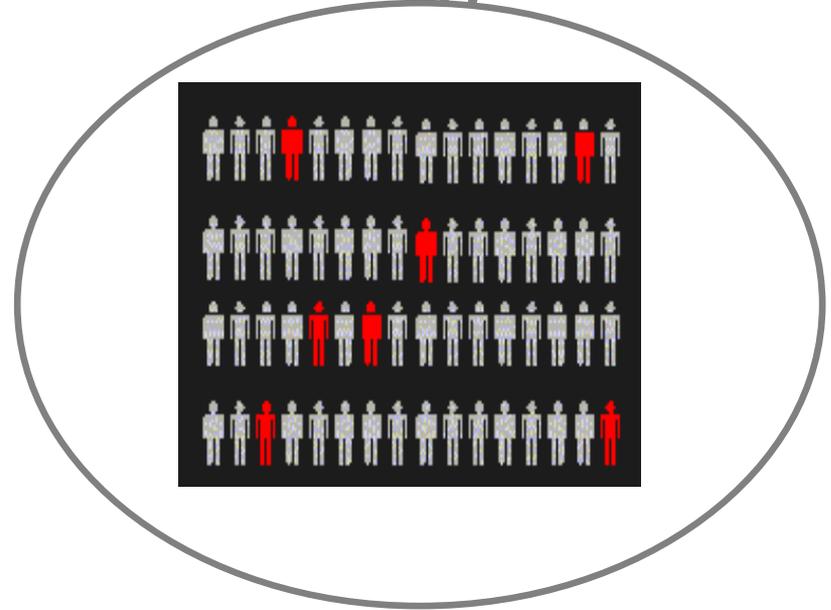
Πληθυσμός έναντι Δείγματος

Πληθυσμός



Όλα τα αντικείμενα ή τα άτομα για τα οποία θέλετε να εξάγετε συμπεράσματα

Δείγμα



Ένα τμήμα του πληθυσμού των αντικειμένων ή των ατόμων

Ορισμοί...

Μεταβλητή είναι ένα χαρακτηριστικό ενός πληθυσμού ή δείγματος.

Π.χ. βαθμοί φοιτητών.

Συνήθως συμβολίζεται με κεφαλαίο γράμμα: X, Y, Z...

Τιμές μιας μεταβλητής είναι το εύρος των δυνατών παρατηρήσεων για μια μεταβλητή.

Π.χ. βαθμοί φοιτητών (0..100)

Δεδομένα είναι οι *τιμές που έχουν παρατηρηθεί* για μια μεταβλητή.

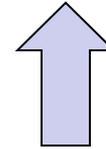
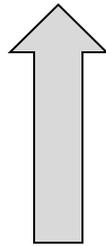
Π.χ. βαθμοί φοιτητών: {67, 74, 71, 83, 93, 55, 48}

Μεταβλητές, Τύποι Δεδομένων & Πληροφοριών

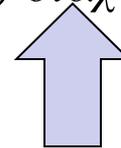
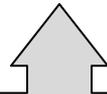
Οι μεταβλητές μπορούν να διαχωριστούν σε δύο κατηγορίες:

(I) Ποσοτικές (*quantitative or numerical*) μεταβλητές

(II) Ποιοτικές (*qualitative*) μεταβλητές



Τα **δεδομένα** (τουλάχιστον για στατιστικούς λόγους) διαχωρίζονται σε τέσσερις κύριες ομάδες:



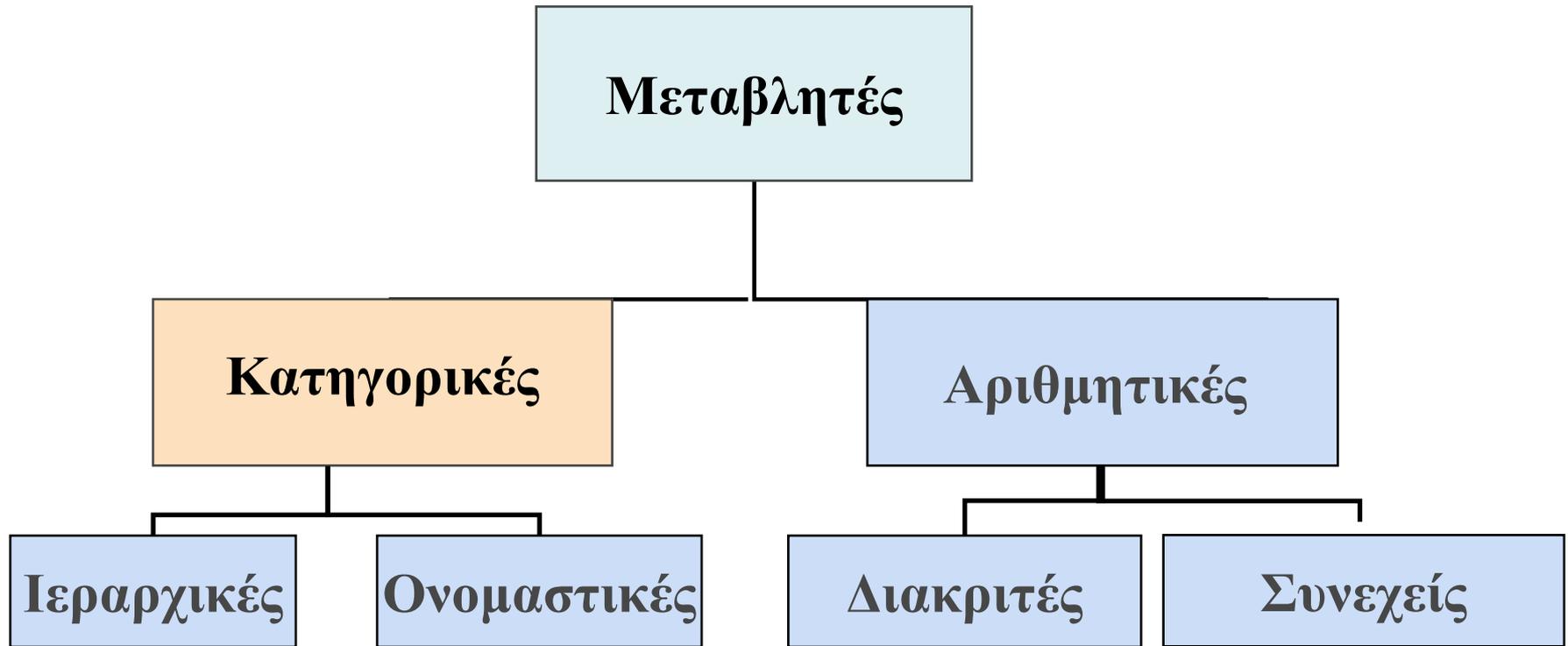
A. Συνεχή Δεδομένα (*Interval Data*)

B. Ασυνεχή ή Διακριτά (*Discrete Data*)

D. Διατακτικά ή Ιεραρχικά Δεδομένα (*Ordinal Data*)

E. Ονομαστικά Δεδομένα (*Nominal Data*)

Τύποι Μεταβλητών



Παραδείγματα:

- Ηλικιακές Κατηγορίες
- Βαθμός Ικανοποίησης
(Κατηγορίες Likert)

Παραδείγματα:

- Οικογενειακή κατάσταση
- Πολιτικό κόμμα
- Χρώμα Ματιών
(ορισμένες κατηγορίες)

Παραδείγματα:

- Αριθμός παιδιών
- Ελαττωματικά ανά ώρα
(Καταμετρημένα αντικείμενα)

Παραδείγματα:

- Βάρος
- Τάση
(Μετρημένα χαρακτηριστικά)

Συνεχή Δεδομένα ...

Συνεχή Δεδομένα

Πραγματικοί αριθμοί, όπως ύψη, βάρη, τιμές, κλπ.
*Ονομάζονται επίσης και **ποσοτικά** ή **αριθμητικά**.*

*Σημείωση: Οι αριθμητικές πράξεις μπορούν να γίνουν μόνο σε Συνεχή Δεδομένα, επομένως έχει νόημα να μιλάμε για $2 * \text{Ύψος}$, ή $\text{Τιμή} + \$1$, κ.ά.*

Διατακτικά Δεδομένα...

Τα **Διατακτικά Δεδομένα** φαίνονται κατηγορικά στη φύση τους, αλλά οι τιμές τους έχουν **διάταξη**, μια ταξινόμηση:

Π.χ. Βαθμολογικό σύστημα εργασιών: ελλιπής = 1, μέτρια = 2, καλή = 3, πολύ καλή = 4, εξαιρετική = 5

Παρόλο που πάλι δεν έχει νόημα να γίνονται πράξεις στα δεδομένα αυτά (π.χ. $2 * \text{μέτρια} = \text{πολύ καλή?!$), μπορούμε να πούμε :

εξαιρετική > ελλιπής ή μέτρια < πολύ καλή

Δηλαδή, η διάταξη διατηρείται ανεξάρτητα από τις αριθμητικές τιμές που αντιστοιχίζονται σε κάθε κατηγορία.

Ονομαστικά Δεδομένα...

Ονομαστικά Δεδομένα

Οι τιμές των **ονομαστικών** δεδομένων είναι **κατηγορίες**.

Π.χ. απαντήσεις σε ερώτηση σχετικά με την οικογενειακή κατάσταση, κωδικοποιούνται ως εξής:

Άγαμος = 1, Έγγαμος = 2, Διαζευγμένος = 3, Χήρος = 4

Αυτά τα δεδομένα είναι **κατηγοριοποιημένα** από τη φύση τους και οι αριθμητικές πράξεις δεν έχουν νόημα (π.χ. ισχύει ότι $\text{Χήρος} \div 2 = \text{Έγγαμος}?!)$

*Τα ονομαστικά δεδομένα καλούνται επίσης **ποιοτικά** ή **κατηγορικά**.*

Υπολογισμοί

Όπως ήδη αναφέραμε,

- Όλοι οι υπολογισμοί επιτρέπονται σε **συνεχή** δεδομένα.
- Μόνο οι υπολογισμοί που αφορούν την ταξινόμηση επιτρέπονται σε **διατακτικά** δεδομένα.
- Κανένας υπολογισμός δεν επιτρέπεται σε **ονομαστικά** δεδομένα, παρά μόνο η καταμέτρηση των παρατηρήσεων κάθε κατηγορίας.

Οδηγούμαστε έτσι στην ακόλουθη “ιεραρχία δεδομένων”...

Ιεραρχία Δεδομένων...

Συνεχή (interval)

Οι τιμές είναι πραγματικοί αριθμοί.

Όλοι οι αριθμητικοί υπολογισμοί επιτρέπονται.

Μπορούν να χρησιμοποιηθούν και ως διατακτικά ή ονομαστικά.

Διατακτικά (ordinal)

Οι τιμές αντιπροσωπεύουν μόνο τη διάταξη των δεδομένων.

Επιτρέπονται μόνο οι υπολογισμοί σχετικά με τη διάταξη.

Μπορούν να χρησιμοποιηθούν και ως ονομαστικά αλλά όχι ως συνεχή.

Ονομαστικά (nominal)

Οι τιμές είναι τυχαίοι αριθμοί που αναπαριστούν κατηγορίες.

Μόνο οι υπολογισμοί συχνοτήτων επιτρέπονται.

Δεν μπορούν να χρησιμοποιηθούν ως ποσοτικά ή διατακτικά.

Γραφήματα & Πίνακες για Ονομαστικά Δεδομένα...

Ο μόνος επιτρεπτός υπολογισμός σε ονομαστικά δεδομένα είναι η καταμέτρηση της συχνότητας κάθε τιμής.

Μπορούμε να συνοψίσουμε τα δεδομένα σε έναν πίνακα που παρουσιάζει τις κατηγορίες και τα αποτελέσματα της μέτρησης, και ονομάζεται ***κατανομή συχνοτήτων***.

Η ***κατανομή σχετικών συχνοτήτων*** παρουσιάζει τις κατηγορίες και το ποσοστό το οποίο κατέχει κάθε μία ως προς το μέγεθος του πληθυσμού.

Παράδειγμα 2.1 Έρευνα προτίμησης μπύρας

Το 2006 η συνολική κατανάλωση ελαφριάς μπύρας στις ΗΠΑ ήταν περίπου 3 εκατομμύρια γαλόνια.

Σε μια τόσο μεγάλη αγορά, οι κατασκευαστές ενδιαφέρονται να γνωρίζουν περισσότερα για τους αγοραστές των προϊόντων τους.

Ο διευθυντής μάρκετινγκ μιας μεγάλης ζυθοποιίας θέλησε να αναλύσει τις πωλήσεις ελαφριάς μπύρας μεταξύ των φοιτητών.

Ένα τυχαίο δείγμα 285 τελειόφοιτων ρωτήθηκε σχετικά με το ποια από τις ακόλουθες μάρκες είναι η αγαπημένη του.

Παράδειγμα 2.1

1. Budweiser Light
2. Busch Light
3. Coors Light
4. Michelob Light
5. Miller Light
6. Natural Light
7. Άλλη

Οι απαντήσεις καταγράφηκαν χρησιμοποιώντας αυτούς του κωδικούς. Να κατασκευάσετε την κατανομή συχνοτήτων και σχετικών συχνοτήτων και να συνοψίσετε γραφικά τα δεδομένα με ένα ραβδόγραμμα και ένα κυκλικό διάγραμμα.

Παράδειγμα 2.1

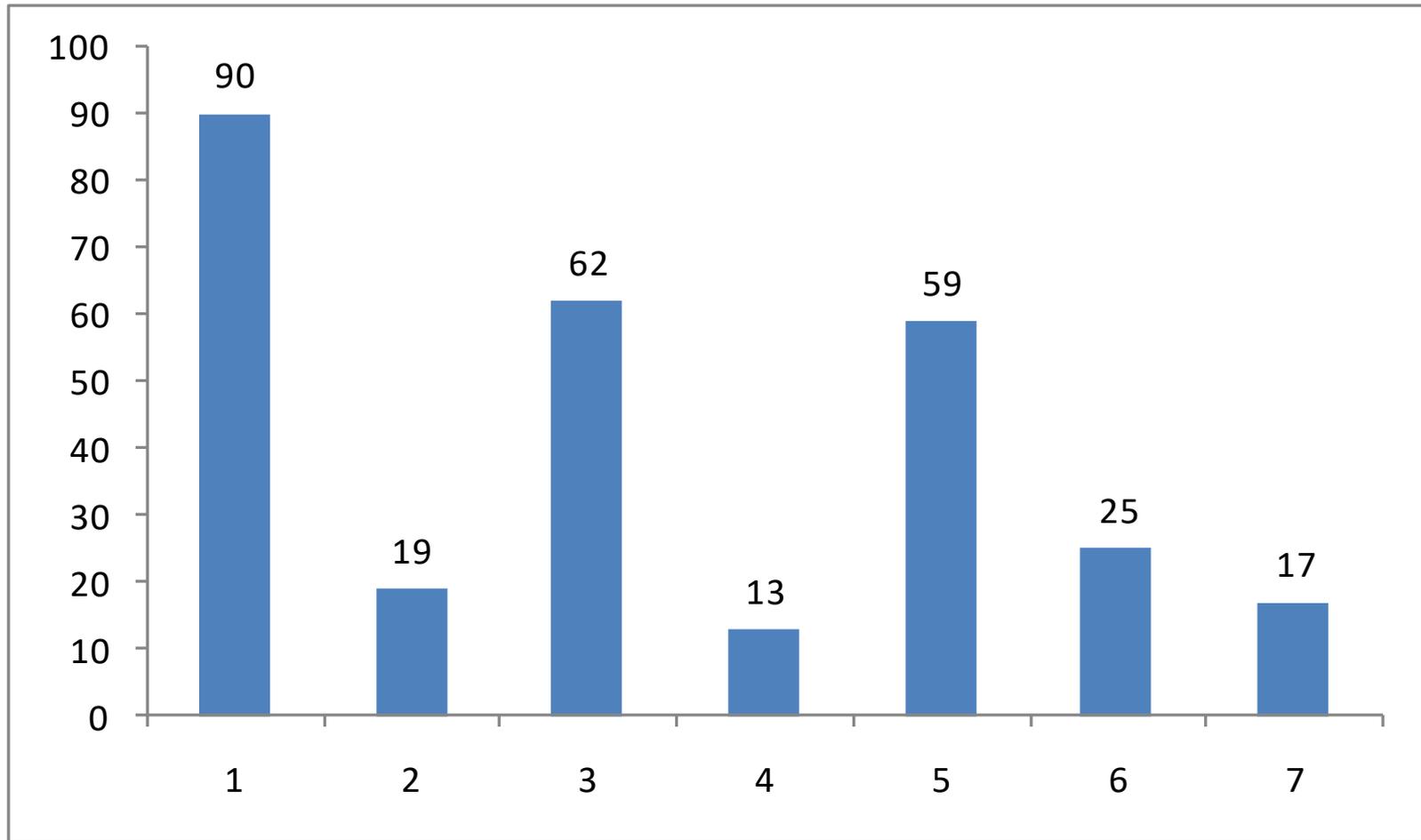
Xm02-01*

1	1	1	1	2	4	3	5	1	3	1	3	7	5	1
1	5	2	1	5	1	3	3	3	1	1	5	3	1	5
5	1	1	3	3	5	5	6	3	5	3	5	5	5	1
1	2	1	1	5	5	3	2	1	6	1	1	4	5	1
3	3	5	4	7	6	6	4	4	6	5	2	1	1	5
3	3	1	3	5	3	3	7	3	7	2	1	5	7	
3	6	2	6	3	6	6	6	5	6	1	1	6	3	
7	1	1	1	5	1	3	1	3	7	7	2	1	1	
2	5	3	1	1	3	1	1	7	5	3	2	1	1	
6	5	7	1	3	2	1	3	1	1	7	5	5	6	
1	4	6	1	3	1	1	5	5	5	5	1	5	5	
6	1	3	3	1	3	7	1	1	1	2	4	1	1	
3	3	7	5	5	1	1	3	5	1	5	4	5	3	
4	1	4	5	3	1	5	3	3	3	1	1	5	3	
5	6	4	3	5	6	4	6	5	5	5	5	3	1	
2	3	2	7	5	1	6	6	2	3	3	3	1	1	
5	1	4	6	3	5	1	1	2	1	5	6	1	1	
5	1	3	5	1	1	1	3	7	3	1	6	3	1	
2	2	5	1	3	5	5	2	3	1	1	3	6	1	
1	1	1	7	3	1	5	3	3	3	5	3	1	7	

Κατανομή Συχνοτήτων και Σχετικών Συχνοτήτων

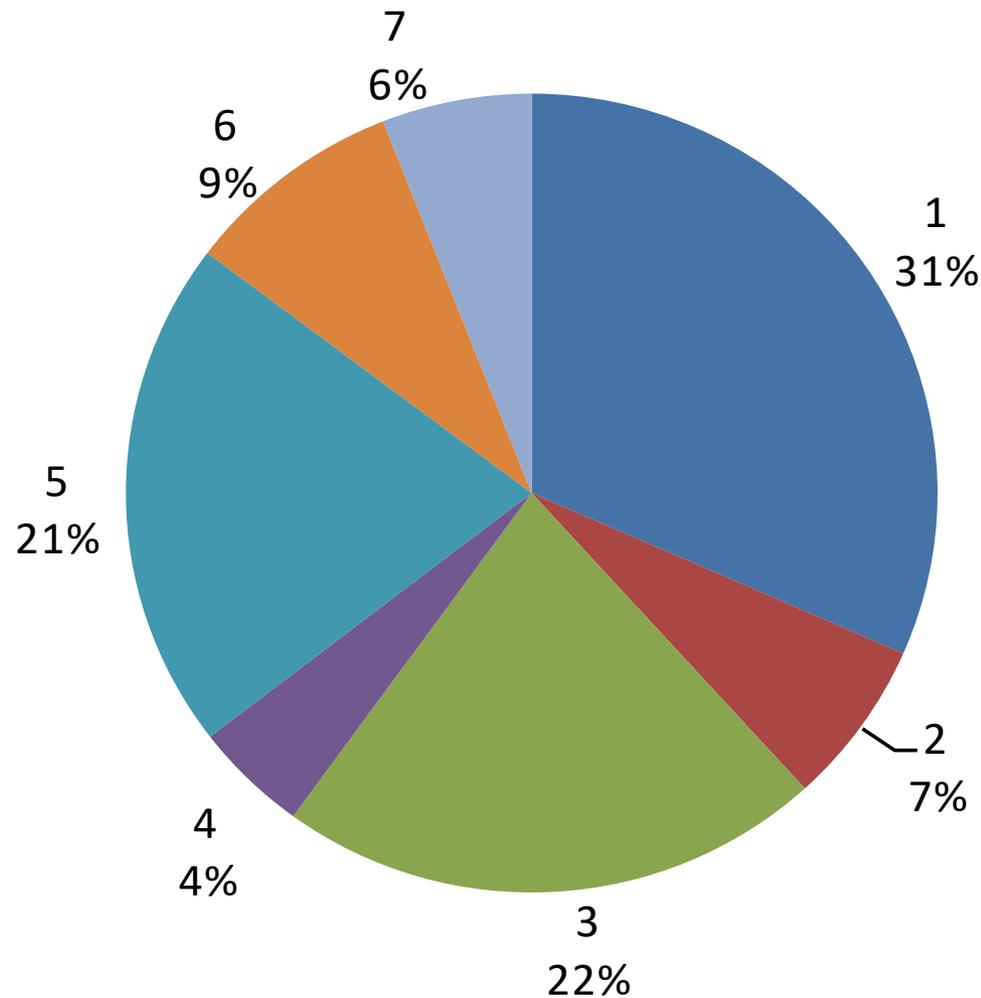
Μάρκα	Συχνότητα (fi)	Σχετική συχνότητα (fi%)
Budweiser Light	90	31.6%
Busch Light	19	6.7
Coors Light	62	21.8
Michelob Light	13	4.6
Miller Lite	59	20.7
Natural Light	25	8.8
Άλλη	17	6.0
Σύνολο	285	100

Ονομαστικά Δεδομένα (Συχνότητα)



Το ραβδόγραμμα χρησιμοποιείται συχνά για *συχνότητες*...

Ονομαστικά Δεδομένα (Σχετική Συχνότητα)

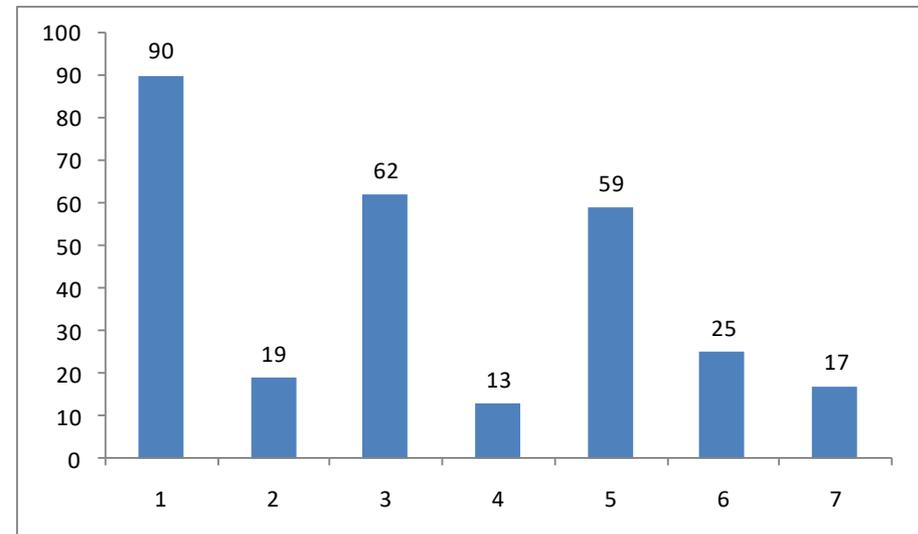
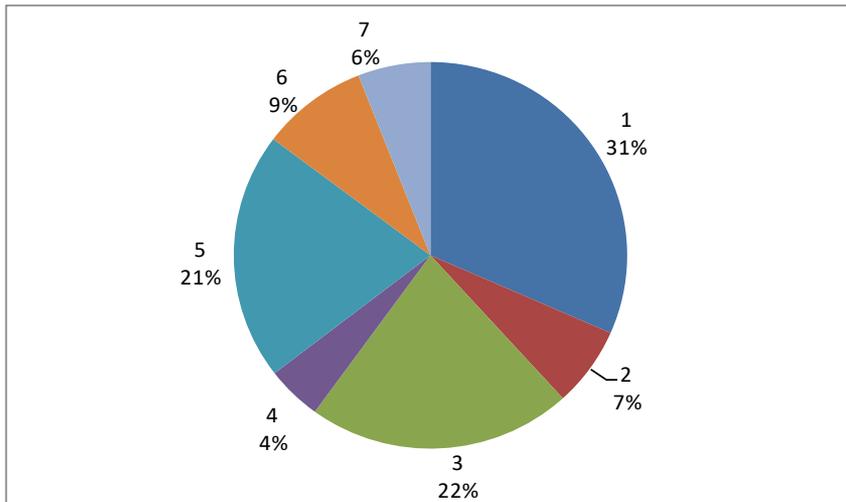


Το κυκλικό διάγραμμα δείχνει **σχετικές συχνότητες...**

Ονομαστικά Δεδομένα

Light Beer Brand	Frequency	Relative Frequency
Budweiser Light	90	31.6%
Busch Light	19	6.7
Coors Light	62	21.8
Michelob Light	13	4.6
Miller Lite	59	20.7
Natural Light	25	8.8
Other brands	17	6.0

Όλα έχουν τις ίδιες **πληροφορίες**,
(με βάση τα ίδια **δεδομένα**).
Διαφέρει απλά η **παρουσίαση**.



Παράδειγμα 2.2

Ο πίνακας 2.3 δείχνει τη συνολική κατανάλωση ενέργειας από όλες τις πηγές στις ΗΠΑ το 2005.

Για διευκόλυνση, όλες οι τιμές έχουν υπολογιστεί σε θερμικά ισοδύναμα μετρικών τόνων (1.000 κιλά) πετρελαίου.

Για παράδειγμα, οι ΗΠΑ κατανάλωσαν άνθρακα και προϊόντα άνθρακα ισοδύναμα με 545.259 μετρικούς τόνους πετρελαίου.

Να χρησιμοποιήσετε την κατάλληλη γραφική μέθοδο για να απεικονίσετε αυτές τις τιμές.

Πίνακας 2.3

Xm02-02*

<u>Μη ανανεώσιμες πηγές</u>	<u>Κατανάλωση</u>
Άνθρακας & προϊόντα άνθρακα	545.258
Πετρέλαιο	903.440
Φυσικό αέριο	517.881
Πυρηνική ενέργεια	209.890
<u>Ανανεώσιμες πηγές</u>	
Υδροηλεκτρική ενέργεια	18.251
Στερεά βιομάζα	52.473
Άλλες πηγές (υγρή βιομάζα, γεω- θερμική, ηλιακή, αιολική, θαλάσσια)	20.533
Σύνολο	2.267.726

Παράδειγμα 2.2



Γραφικές Τεχνικές για Συνεχή Δεδομένα

Υπάρχουν αρκετές γραφικές μέθοδοι που χρησιμοποιούνται όταν τα δεδομένα είναι **συνεχή** (δηλ. αριθμητικά, μη-κατηγορικά).

Η πιο σημαντική γραφική μέθοδος είναι το **ιστόγραμμα**.

Το ιστόγραμμα δεν αποτελεί μόνο μια ισχυρή γραφική τεχνική για τη **σύνοψη** συνεχών δεδομένων, αλλά χρησιμοποιείται και για την **κατανόηση** των πιθανοτήτων.

Παράδειγμα 2.4

Με την απελευθέρωση των τηλεπικοινωνιών, δημιουργήθηκαν αρκετές νέες εταιρίες οι οποίες ανταγωνίζονται στην παροχή τηλεφωνικών υπηρεσιών. Σχεδόν σε όλες τις περιπτώσεις, οι εταιρίες αυτές ανταγωνίζονται στις τιμές αφού οι προσφερόμενες υπηρεσίες είναι παρόμοιες. Ο καθορισμός τιμής ενός προϊόντος σε ένα τόσο σκληρό ανταγωνισμό είναι εξαιρετικά δύσκολο να γίνει. Στους παράγοντες που πρέπει να εξεταστούν είναι η προσφορά, η ζήτηση, η ελαστικότητα τιμής και οι ενέργειες των ανταγωνιστών. Τα μακράς διάρκειας πακέτα μπορεί να περιέχουν χρονοχρέωση, μηνιαίο πάγιο, ή συνδυασμό αυτών.

Ο καθορισμός της κατάλληλης τιμολογιακής πολιτικής διευκολύνεται από τη γνώση της συμπεριφοράς των καταναλωτών και ειδικότερα από το ύψος των μηνιαίων λογαριασμών.

Παράδειγμα 2.4

Στα πλαίσια μιας ευρύτερης έρευνας, μια εταιρία τηλεφωνίας θέλησε να συγκεντρώσει πληροφορίες για το ύψος των μηνιαίων λογαριασμών νέων συνδρομητών κατά τον πρώτο μήνα μετά την εγγραφή τους. Ο διευθυντής μάρκετινγκ διεξήγαγε μια έρευνα σε 200 νέους συνδρομητές στην οποία καταγράφηκε το ύψος του πρώτου μηνιαίου λογαριασμού. Τα στοιχεία αυτά βρίσκονται στο αρχείο [Xm02-04](#). Τα ευρήματα της έρευνας πρέπει να παρουσιαστούν στη διοίκηση της εταιρίας. Ποιες πληροφορίες μπορούν να εξαχθούν;

Παράδειγμα 2.4

Στο Παράδειγμα 2.1 δημιουργήσαμε μια κατανομή συχνοτήτων των 5 κατηγοριών. Στο παράδειγμα αυτό, επίσης θα δημιουργήσουμε μια κατανομή συχνοτήτων μετρώντας τον αριθμό των παρατηρήσεων που ανήκουν σε μια σειρά διαστημάτων, τα οποία ονομάζονται κλάσεις.

Θα εξηγήσουμε αργότερα τον τρόπο επιλογής των κλάσεων.

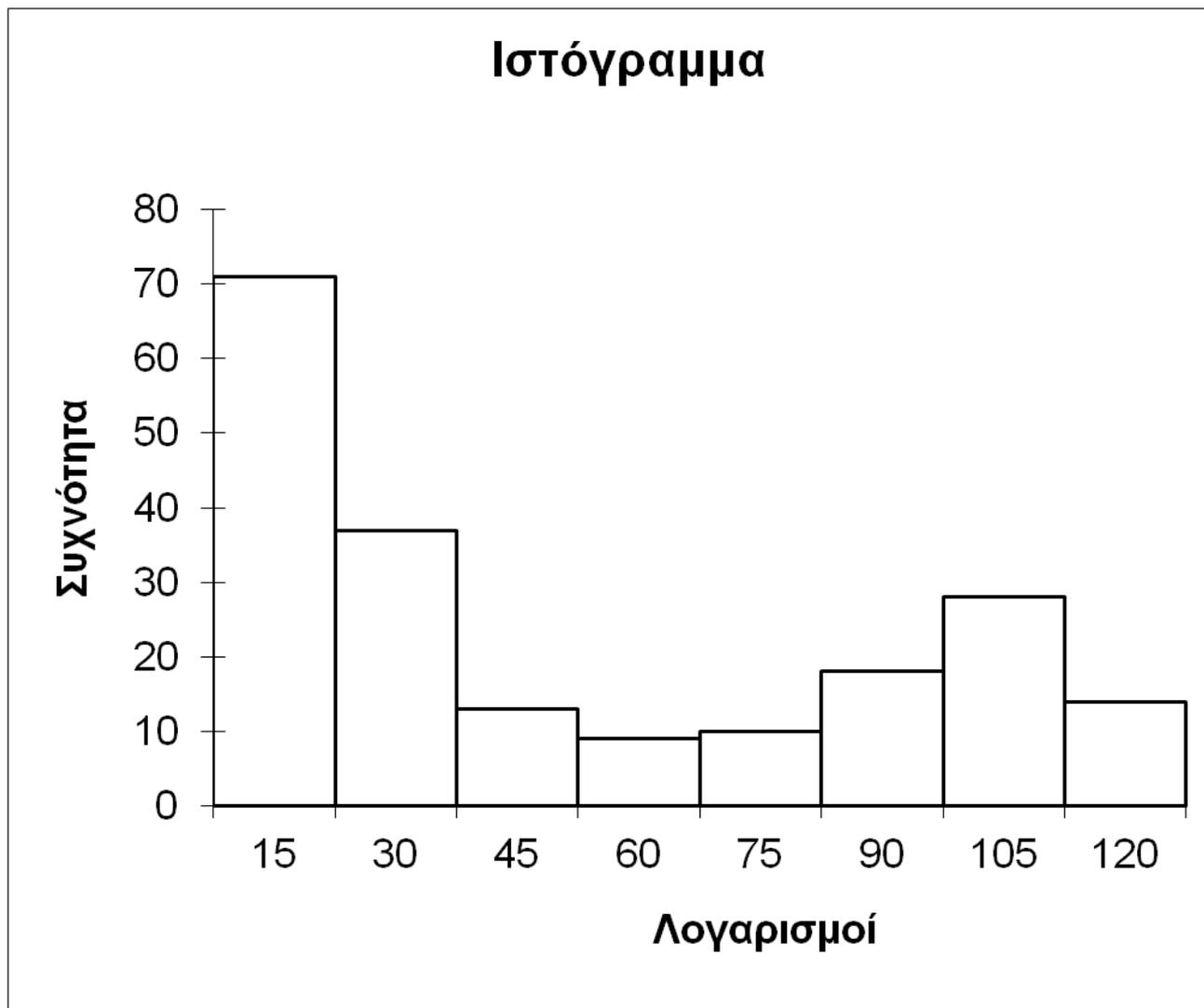
Παράδειγμα 2.4

Επιλέξαμε **οκτώ κλάσεις** με τέτοιο τρόπο ώστε κάθε παρατήρηση να εντάσσεται σε μία και μόνο κλάση. Οι κλάσεις ορίζονται ως εξής:

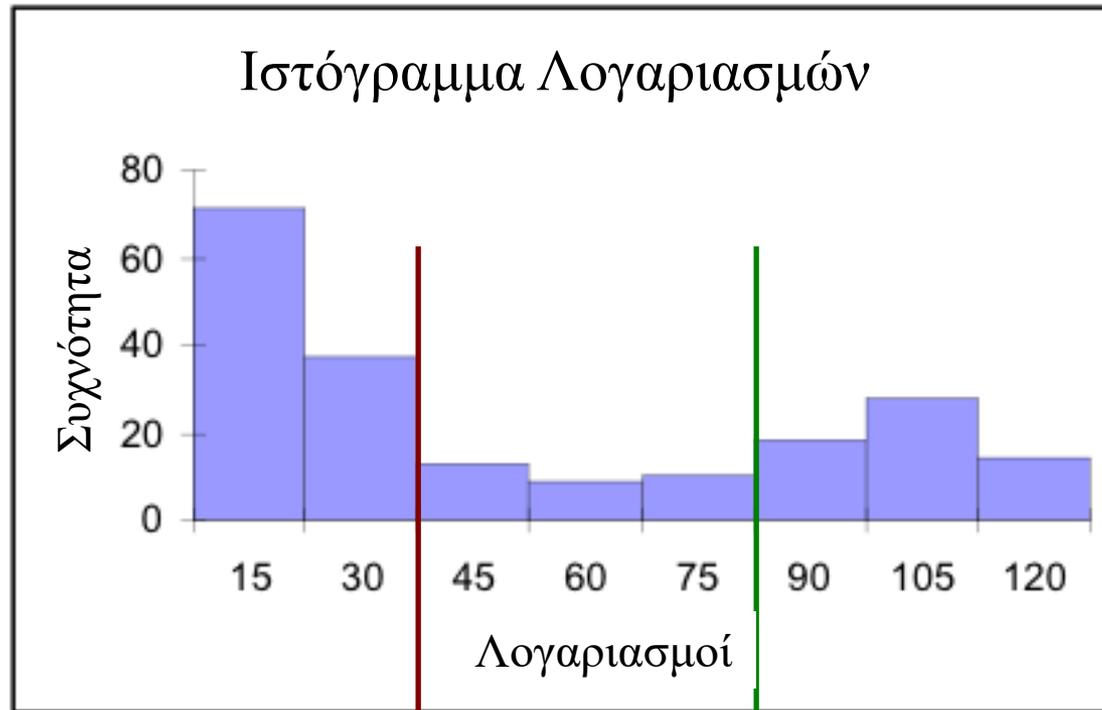
Κλάσεις

1. Ποσά μικρότερα ή ίσα του 15
2. Ποσά μεγαλύτερα του 15 αλλά μικρότερα ή ίσα του 30
3. Ποσά μεγαλύτερα του 30 αλλά μικρότερα ή ίσα του 45
4. Ποσά μεγαλύτερα του 45 αλλά μικρότερα ή ίσα του 60
5. Ποσά μεγαλύτερα του 60 αλλά μικρότερα ή ίσα του 75
6. Ποσά μεγαλύτερα του 75 αλλά μικρότερα ή ίσα του 90
7. Ποσά μεγαλύτερα του 90 αλλά μικρότερα ή ίσα του 105
8. Ποσά μεγαλύτερα του 105 αλλά μικρότερα ή ίσα του 120

Παράδειγμα 2.4



Ερμηνεία...



Περίπου οι μισοί ($71+37=108$) λογαριασμοί είναι "μικροί", δηλ. μικρότεροι από \$30

$(18+28+14=60) \div 200 = 30\%$
δηλ. σχεδόν το ένα τρίτο των λογαριασμών είναι \$90 ή παραπάνω.

Υπάρχουν μόνο λίγοι λογαριασμοί στην ενδιάμεση κατηγορία.

Φτιάχνοντας ένα Ιστόγραμμα...

- 1) Συλλέγουμε τα Δεδομένα
- 2) Δημιουργούμε μια κατανομή συχνοτήτων ...

Πώς;

a) Καθορίζουμε το πλήθος των **κλάσεων** που θα χρησιμοποιήσουμε...

Πώς;

Με 200 παρατηρήσεις, θα πρέπει να έχουμε 7 με 10 κλάσεις...

Table 2.6 Approximate Number of Classes in Frequency Distributions

Number of Observations	Number of Classes
Less than 50	5 - 7
50 - 200	7 - 9
200 - 500	9 - 10
500 - 1,000	10 - 11
1,000 - 5,000	11 - 13
5,000 - 50,000	13 - 17
More than 50,000	17 - 20

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε Τον τύπο του Sturges:

Πλήθος κλάσεων = $1 + 3.3 \log (N)$

N = το πλήθος των στοιχείων - δεδομένων

Φτιάχνοντας ένα Ιστόγραμμα ...

1) Συλλέγουμε τα Δεδομένα

2) Δημιουργούμε μια κατανομή συχνοτήτων...

Πώς;

a) Καθορίζουμε τον αριθμό των **κλάσεων**. [8 για το παράδειγμα]

b) Καθορίζουμε πόσο μεγάλη θα είναι κάθε κλάση...

Πώς;

Βρίσκουμε το **εύρος** των δεδομένων, δηλαδή,

Εύρος = Μεγαλύτερη παρατήρηση – Μικρότερη παρατήρηση

$$\text{Εύρος} = \$119.63 - \$0 = \$119.63$$

Επομένως το πλάτος κάθε κλάσης είναι:

$$\text{Εύρος} \div (\# \text{ κλάσεων}) = 119.63 \div 8 \approx 15$$

Φτιάχνοντας ένα Ιστόγραμμα ...

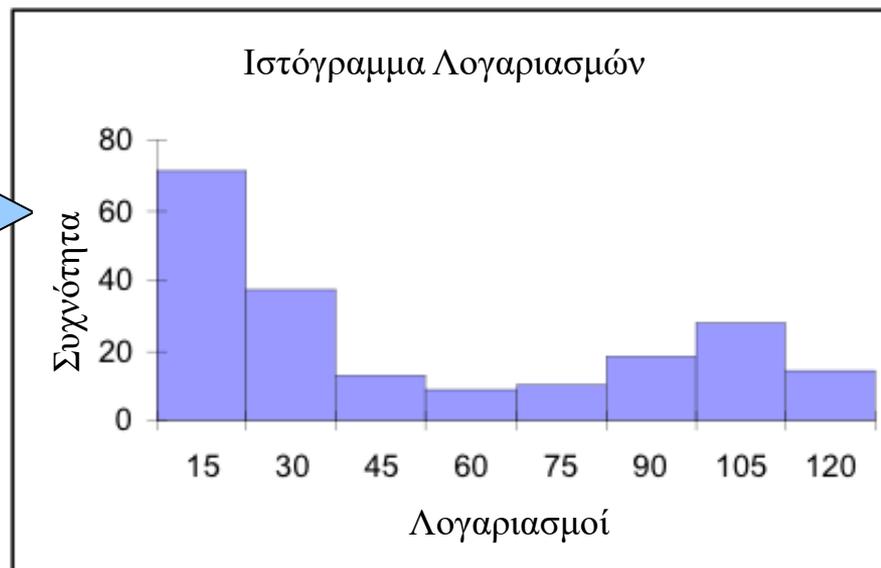
Πίνακας 2.5 Κατανομή Συχνοτήτων
για το Παράδειγμα 2.4

Όρια κλάσεων	Συχνότητα (Frequency:Fi)
0 έως 15	71
15 έως 30	37
30 έως 45	13
45 έως 60	9
60 έως 75	10
75 έως 90	18
90 έως 105	28
105 έως 120	14
Σύνολο	200

Φτιάχνοντας ένα Ιστόγραμμα ...

Πίνακας 2.5 Κατανομή Συχνοτήτων
για το Παράδειγμα 2.4

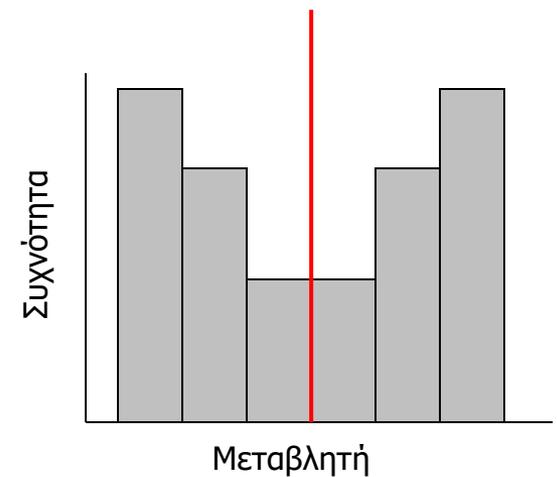
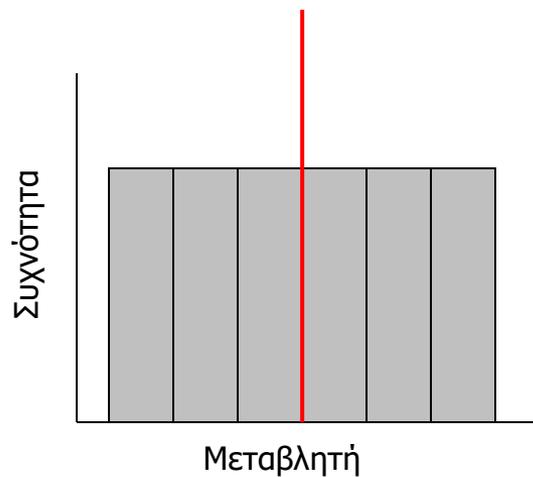
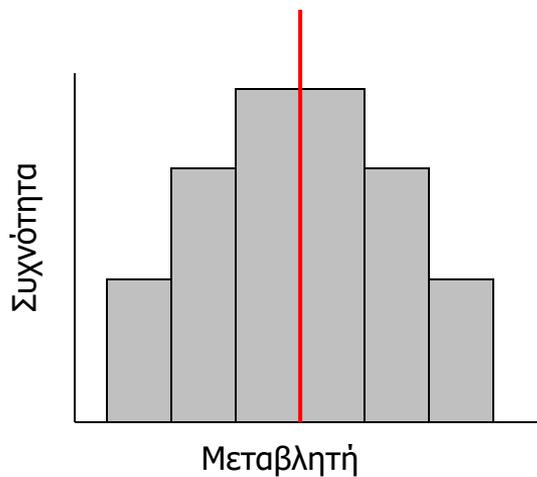
Όρια κλάσεων	Συχνότητα (Frequency:Fi)
0 έως 15	71
15 έως 30	37
30 έως 45	13
45 έως 60	9
60 έως 75	10
75 έως 90	18
90 έως 105	28
105 έως 120	14
Σύνολο	200



Σχήματα Ιστογραμμάτων...

Συμμετρία

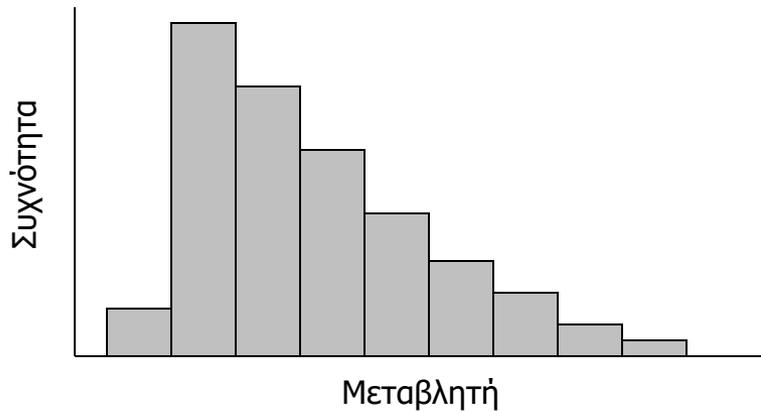
Ένα ιστογράμμα καλείται **συμμετρικό** εάν, σχεδιάζοντας μια **κατακόρυφη ευθεία** στο κέντρο του οι δύο πλευρές είναι ταυτόσημες ως προς το σχήμα και το μέγεθος:



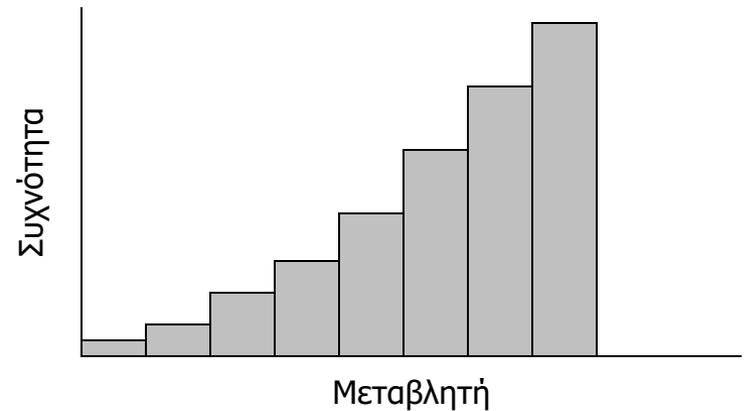
Σχήματα Ιστογραμμάτων...

Ασυμμετρία

Ασύμμετρο ιστογράμμο είναι αυτό που έχει μια «ουρά» στο δεξί ή στο αριστερό άκρο του:



Θετική Ασυμμετρία



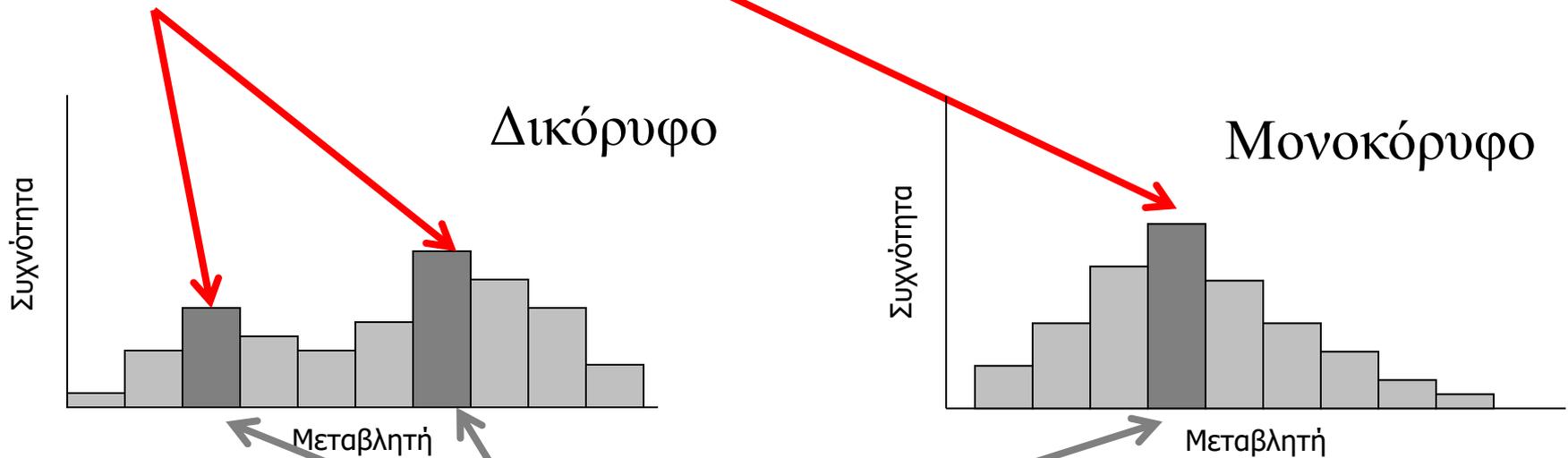
Αρνητική Ασυμμετρία

Σχήματα Ιστογραμμάτων...

Αριθμός Κορυφών

Μονοκόρυφο ιστογράμμα είναι αυτό που έχει μόνο μία κορυφή, ενώ

Δικόρυφο ιστογράμμα είναι αυτό που έχει δύο κορυφές



Επικρατούσα κλάση είναι η κλάση με το μεγαλύτερο πλήθος παρατηρήσεων

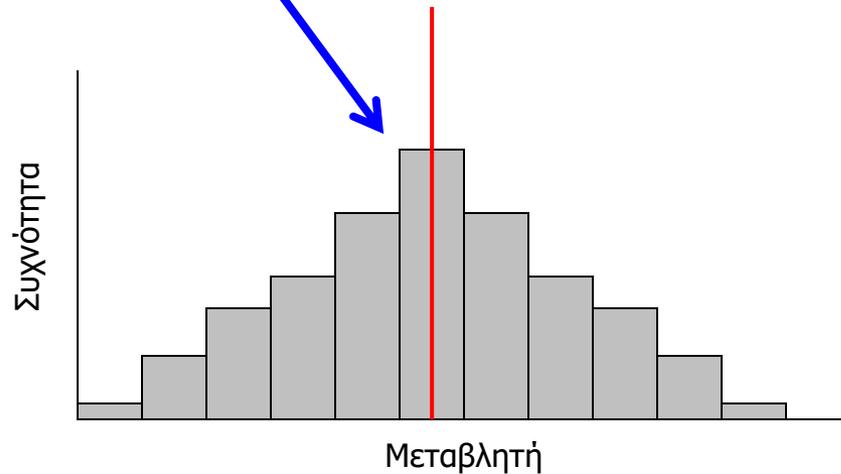
Σχήματα Ιστογραμμάτων...

Καμπάνα

Ένας ειδικός τύπος **μονοκόρυφου συμμετρικού** ιστογραμματος-τος είναι η **καμπάνα**:

Πολλές στατιστικές τεχνικές απαιτούν ο πληθυσμός να έχει σχήμα καμπάνας.

Ο σχεδιασμός του ιστογράμματος βοηθά στην επαλήθευση του σχήματος του πληθυσμού.

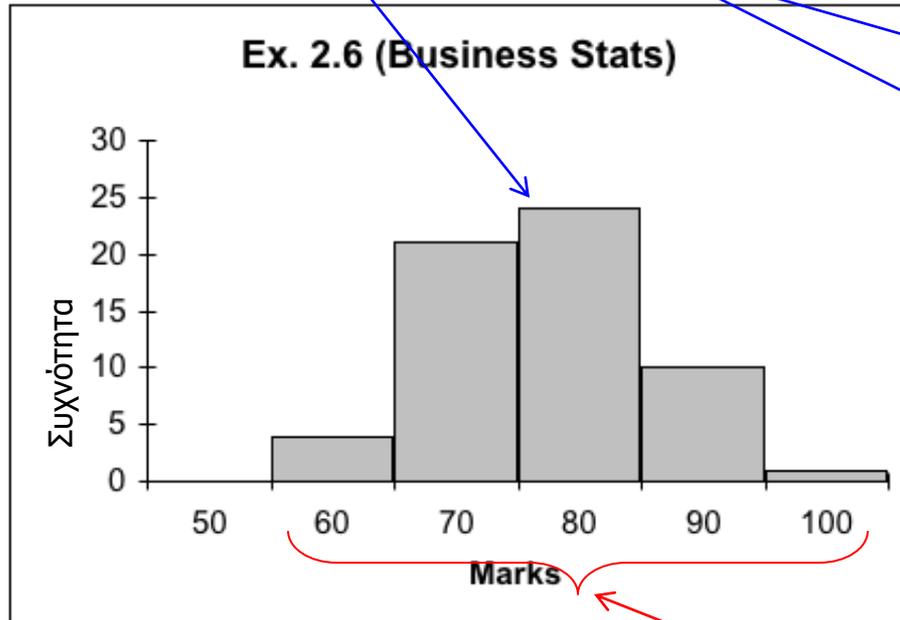


Καμπάνα

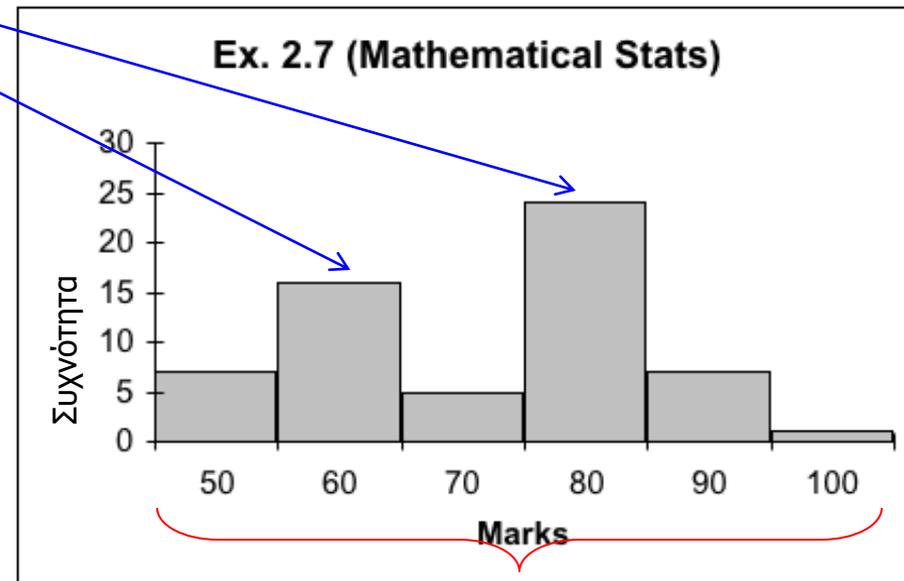
Σύγκριση Ιστογραμμάτων...

Να συγκρίνετε τα ακόλουθα ιστογράμματα που προκύπτουν από τα [Π.χ. 2.6](#) και [Π.χ. 2.7](#):

Μονοκόρυφο vs. Δικόρυφο



Τα δύο μαθήματα, Στατιστική Επιχειρήσεων και Μαθηματική Στατιστική έχουν πολύ διαφορετικά ιστογράμματα...



διασπορά βαθμών (στενότερη | πλατύτερη)

Φυλλογράφημα...

- Διατηρεί πληροφορίες σχετικά με ατομικές παρατηρήσεις οι οποίες θα χάνονταν στη δημιουργία του ιστογράμματος.
- Διασπά κάθε παρατήρηση σε δύο μέρη, τον **κορμό** και το **φύλλο**:
- π.χ. Παρατηρούμενη τιμή: **42,19**
- Υπάρχουν πολλοί τρόποι διάσπασης...
- στη δεκαδική υποδιαστολή:
- ή στη θέση της "δεκάδας"
(στρογγυλοποιώντας στον πλησιέστερο ακέραιο)

Stem	Leaf
42	19
4	2

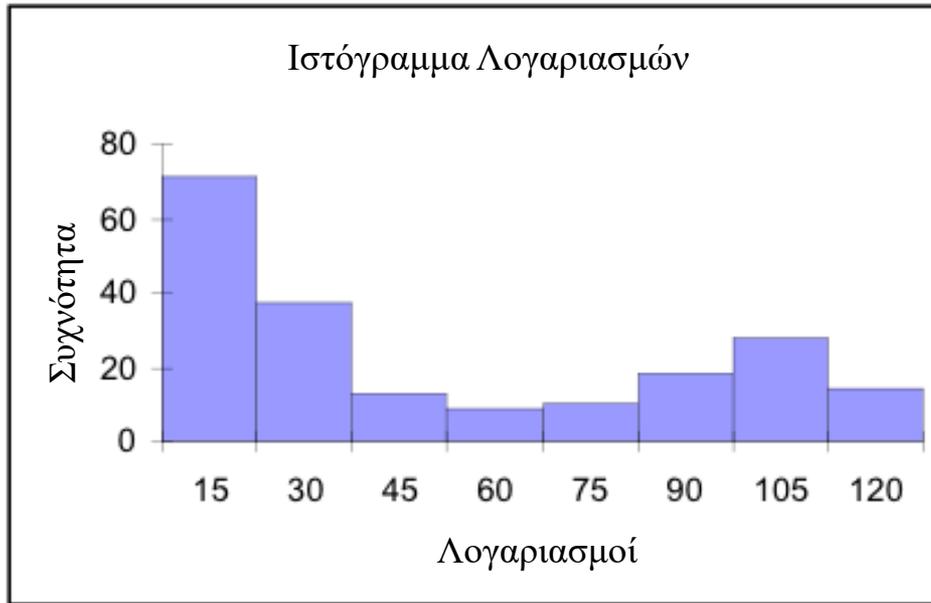
Φυλλογράφημα...

- Συνεχίζουμε για όλες τις παρατηρήσεις. Μετά, χρησιμοποιούμε τους “κορμούς” για τις κλάσεις και κάθε φύλλο γίνεται μέρος του ιστογράμματος (δεδομένα από το παράδειγμα 2.4) ως εξής...

Κορμός	Φύλλο
0	0000000001111122222233333455555666666778888999999
1	000001111233333334455555667889999
2	0000111112344666778999
3	001335589
4	124445589
5	33566
6	3458
7	022224556789
8	334457889999
9	00112222233344555999
10	001344446699
11	124557889

Έτσι, έχουμε ακόμα πρόσβαση στην τιμή των αρχικών δεδομένων

Ιστόγραμμα και Φυλλογράφημα ...



Κορμός Φύλλο

0	000000000111112222223333334455555566666677888888999999
1	0000011112333333334455555566678889999
2	0000111112344666778999
3	001335589
4	124445589
5	33566
6	3458
7	022224556789
8	334457889999
9	00112222233344555999
10	001344446699
11	124557889

Συγκρίνετε τα *σχήματα* των γραφημάτων...

Αθροιστικό πολυγωνικό διάγραμμα (Ogive)...

- Είναι το γράφημα της **κατανομής αθροιστικών συχνοτήτων**.
- Δημιουργούμε μια «αψίδα» σε τρία βήματα...
- Αρχικά, από την κατανομή συχνοτήτων που φτιάξαμε προηγουμένως, υπολογίζουμε **σχετικές συχνότητες**.
- Σχετική συχνότητα =
$$\frac{\# \text{ παρατηρήσεων μιας κλάσης}}{\text{συνολικό } \# \text{ παρατηρήσεων}}$$

Σχετική Συχνότητα...

Για παράδειγμα, είχαμε 71 παρατηρήσεις στην πρώτη κλάση (λογαριασμοί από \$0.00 έως \$15.00). Επομένως, η σχετική συχνότητα για την κλάση αυτή είναι $71 \div 200$ (το συνολικό # λογαριασμών) = 0.355 (ή 35.5%)

Πίνακας 2.7 Κατανομή Σχετικής Συχνότητας για το Παράδειγμα 2.4

Όρια κλάσεων	Σχετική Συχνότητα
0 to 15	$71/200 = .355$
15 to 30	$37/200 = .185$
30 to 45	$13/200 = .065$
45 to 60	$9/200 = .045$
60 to 75	$10/200 = .050$
75 to 90	$18/200 = .090$
90 to 105	$28/200 = .140$
105 to 120	$14/200 = .070$
Σύνολο	$200/200 = 1.0$

Αθροιστικό πολυγωνικό διάγραμμα ...

- Είναι το γράφημα της **κατανομής αθροιστικών συχνοτήτων**.
- Δημιουργούμε μια αψίδα σε τρία βήματα...
- 1) Υπολογίζουμε σχετικές συχνότητες. ✓
- 2) Υπολογίζουμε **αθροιστικές σχετικές συχνότητες** προσθέτοντας τη σχετική συχνότητα της τρέχουσας κλάσης στην αθροιστική σχετική συχνότητα της προηγούμενης κλάσης.
(Για την πρώτη κλάση, η αθροιστική σχετική συχνότητα είναι απλά η σχετική συχνότητα)

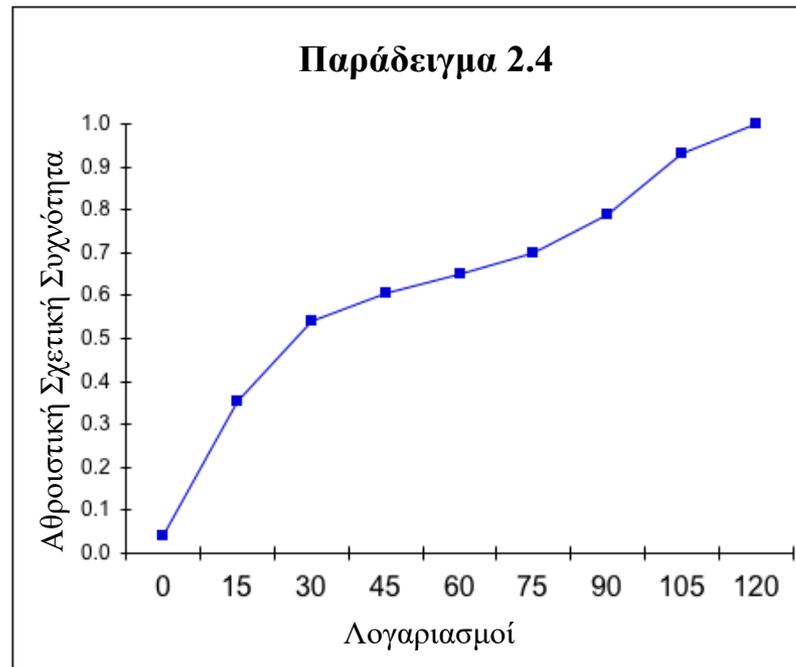
Αθροιστική Σχετική Συχνότητα...

Πίνακας 2.8 Κατανομή Αθροιστικής Σχετικής Συχνότητας Παράδειγμα 2.4

Ορια κλάσεων	Σχετική Συχνότητα	Αθροιστική Σχετική Συχνότητα	
0 to 15	$71/200 = .355$	$71/200 = .355$	πρώτη κλάση...
15 to 30	$37/200 = .185$	$108/200 = .540$	επόμενη : $.355 + .185 = .540$
30 to 45	$13/200 = .065$	$121/200 = .605$	
45 to 60	$9/200 = .045$	$130/200 = .650$:
60 to 75	$10/200 = .05$	$140/200 = .700$:
75 to 90	$18/200 = .09$	$158/200 = .790$	
90 to 105	$28/200 = .14$	$186/200 = .930$	
105 to 120	$14/200 = .07$	$200/200 = 1.00$	τελευταία: $.930 + .070 = 1.00$

Αθροιστικό πολυγωνικό διάγραμμα ...

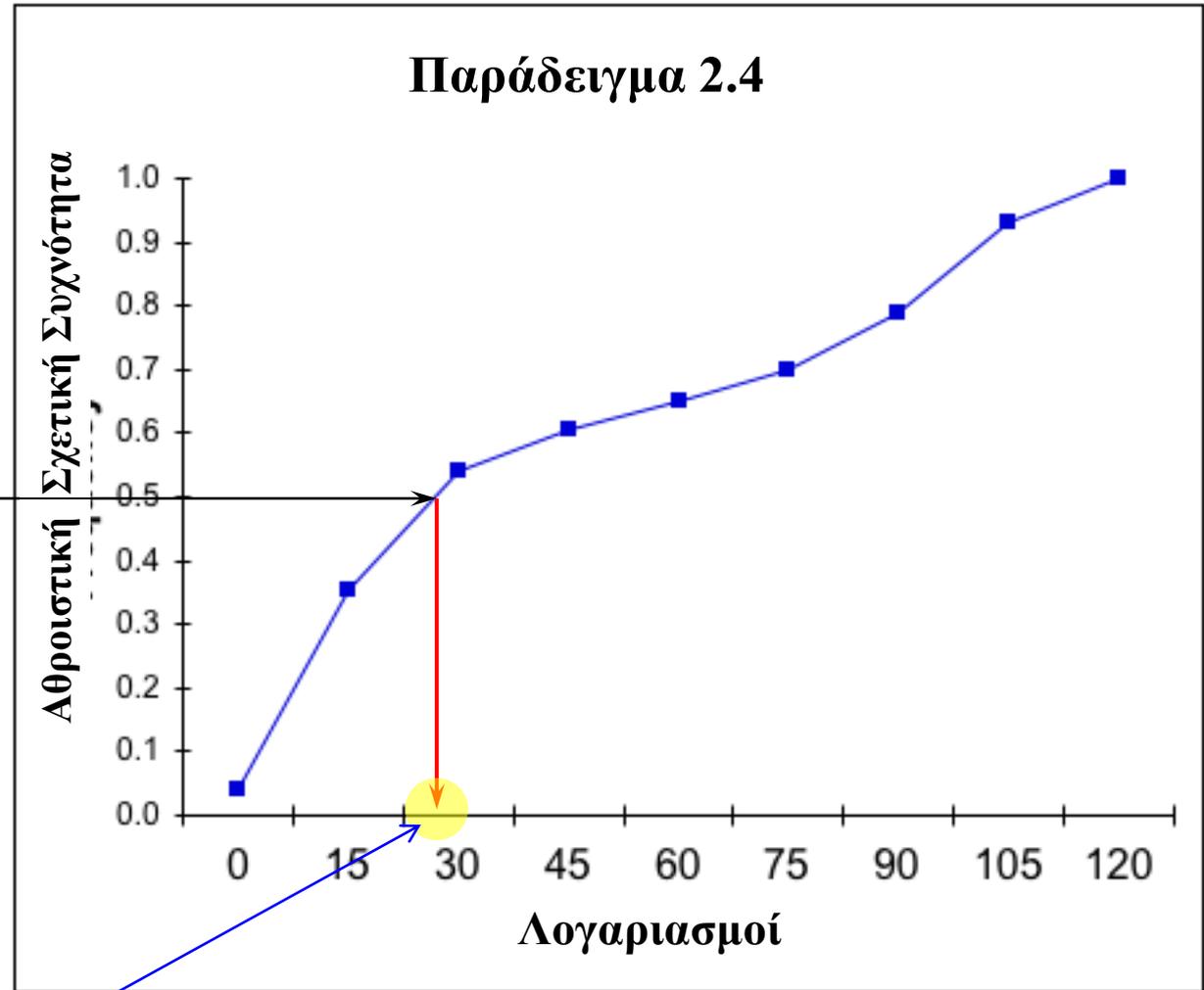
- Είναι το γράφημα της **κατανομής αθροιστικών συχνοτήτων**.
- 1) Υπολογίζουμε σχετικές συχνότητες. ✓
- 2) Υπολογίζουμε αθροιστικές σχετικές συχνότητες. ✓
- 3) Σχεδιάζουμε τις αθροιστικές σχετικές συχνότητες ...



Αθροιστικό πολυγωνικό διάγραμμα ...

Μπορούμε να απαντήσουμε ερωτήσεις όπως:

Ποιος λογαριασμός είναι στο 50% ;



“σχεδόν \$35”

Περιγραφή Χρονολογικών Σειρών

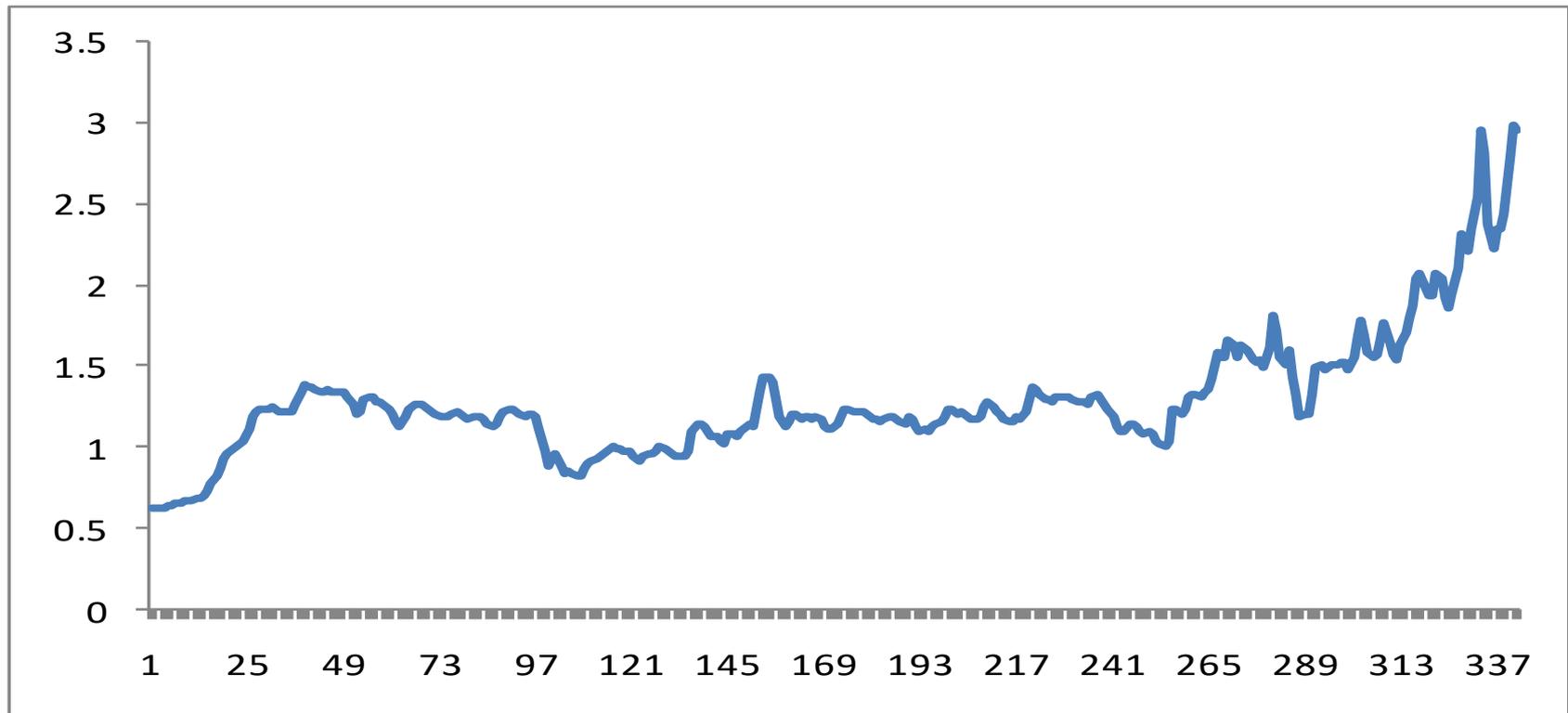
Παρατηρήσεις που συλλέγονται ταυτόχρονα καλούνται **διαστρωματικά δεδομένα**.

Παρατηρήσεις που καταγράφονται σε μια ακολουθία χρονικών σημείων καλούνται **χρονολογικές σειρές**.

Οι χρονολογικές σειρές απεικονίζονται σε **γραμμικό διάγραμμα**, το οποίο παριστάνει την τιμή της μεταβλητής στον κάθετο άξονα και τη χρονική περίοδο στον οριζόντιο άξονα.

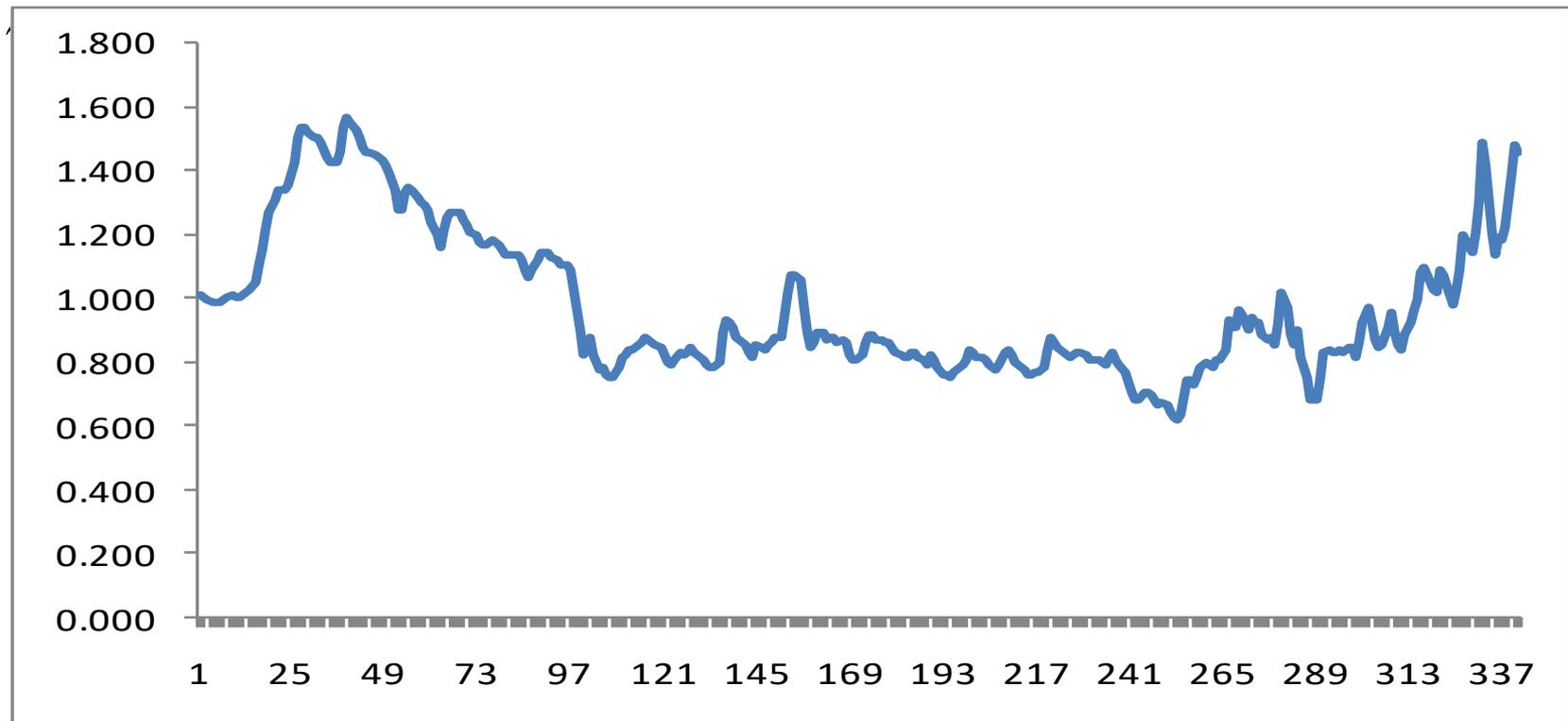
Παράδειγμα 2.8

Καταγράψαμε τις μέσες λιανικές τιμές της βενζίνης από το 1978. [Xm02-08](#). Να σχεδιάσετε ένα γραμμικό διάγραμμα για να περιγράψετε τα δεδομένα και να σχολιάσετε το αποτέλεσμα.



Παράδειγμα 2.9 Τιμή της βενζίνης σε δολάρια 1982-84

Xm02-09 Να αφαιρέσετε τις συνέπειες του πληθωρισμού από τις τιμές του Παραδείγματος 2.8. Να δείτε εάν οι τιμές της βενζίνης είναι υψηλότερες από ό,τι ήταν στο παρελθόν μετά από την αφαίρεση



Σχέση μεταξύ δύο ονομαστικών μεταβλητών...

Μέχρι τώρα είδαμε γραφικές τεχνικές για μία μεταβλητή (είτε για ονομαστικά είτε για συνεχή δεδομένα).

Ο **πίνακας διπλής εισόδου (ή πίνακας διπλής εισόδου ή πίνακας συνάφειας, CROSS-TABULATION)** χρησιμοποιείται για να περιγραφεί η σχέση μεταξύ **δύο** ονομαστικών μεταβλητών.

Ο πίνακας διπλής εισόδου καταγράφει τη **συχνότητα** **κάθε συνδυασμού** των τιμών των δύο μεταβλητών...

ΠΑΡΑΔΕΙΓΜΑ ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ / ΔΙΠΛΗΣ ΕΙΣΟΔΟΥ / ΔΙΑΣΤΑΥΡΩΣΕΩΝ (1)

Χ: ΜΕΤΑΒΛΗΤΗ ΚΑΠΝΙΣΤΕΣ

Υ: ΜΕΤΑΒΛΗΤΗ ΠΑΣΧΟΝΤΕΣ

ΚΑΠΝΙΣΤΕΣ

ΠΑΣΧΟΝΤΕΣ

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ΝΑΙ	80	80,0	80,0	80,0
ΟΧΙ	20	20,0	20,0	100,0
Total	100	100,0	100,0	

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ΝΑΙ	32	32,0	32,0	32,0
ΟΧΙ	68	68,0	68,0	100,0
Total	100	100,0	100,0	

ΠΑΣΧΟΝΤΕΣ * ΚΑΠΝΙΣΤΕΣ Crosstabulation

Count

		ΚΑΠΝΙΣΤΕΣ		Total
		ΝΑΙ	ΟΧΙ	
ΠΑΣΧΟΝΤΕΣ	ΝΑΙ	30	2	32
	ΟΧΙ	50	18	68
Total		80	20	100

Περιθώρια στήλη

Περιθώρια γραμμή

ΠΑΡΑΔΕΙΓΜΑ ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ /ΔΙΠΛΗΣ ΕΙΣΟΔΟΥ /ΔΙΑΣΤΑΥΡΩΣΕΩΝ (2)

ΠΑΣΧΟΝΤΕΣ * ΚΑΠΝΙΣΤΕΣ Crosstabulation

			ΚΑΠΝΙΣΤΕΣ		Total
			ΝΑΙ	ΟΧΙ	
ΠΑΣΧΟΝΤΕΣ	ΝΑΙ	Count	30	2	32
		% within ΠΑΣΧΟΝΤΕΣ	93,8%	6,3%	100,0%
	ΟΧΙ	Count	50	18	68
		% within ΠΑΣΧΟΝΤΕΣ	73,5%	26,5%	100,0%
Total		Count	80	20	100
		% within ΠΑΣΧΟΝΤΕΣ	80,0%	20,0%	100,0%

% γραμμής

ΠΑΣΧΟΝΤΕΣ * ΚΑΠΝΙΣΤΕΣ Crosstabulation

			ΚΑΠΝΙΣΤΕΣ		Total
			ΝΑΙ	ΟΧΙ	
ΠΑΣΧΟΝΤΕΣ	ΝΑΙ	Count	30	2	32
		% within ΚΑΠΝΙΣΤΕΣ	37,5%	10,0%	32,0%
	ΟΧΙ	Count	50	18	68
		% within ΚΑΠΝΙΣΤΕΣ	62,5%	90,0%	68,0%
Total		Count	80	20	100
		% within ΚΑΠΝΙΣΤΕΣ	100,0%	100,0%	100,0%

% στήλης

% συνόλου

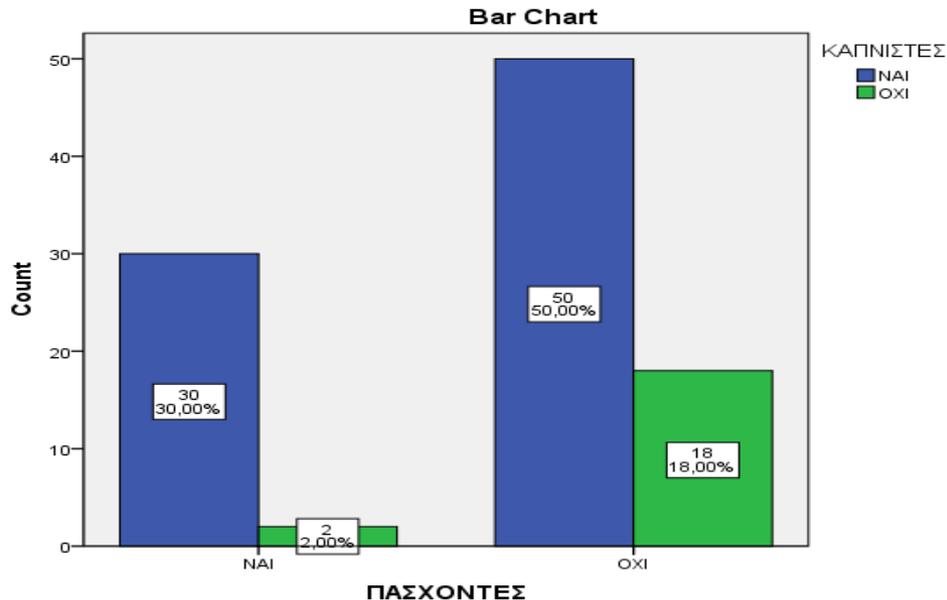
ΠΑΣΧΟΝΤΕΣ * ΚΑΠΝΙΣΤΕΣ Crosstabulation

			ΚΑΠΝΙΣΤΕΣ		Total
			ΝΑΙ	ΟΧΙ	
ΠΑΣΧΟΝΤΕΣ	ΝΑΙ	Count	30	2	32
		% of Total	30,0%	2,0%	32,0%
	ΟΧΙ	Count	50	18	68
		% of Total	50,0%	18,0%	68,0%
Total		Count	80	20	100
		% of Total	80,0%	20,0%	100,0%

ΠΑΡΑΔΕΙΓΜΑ ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ / ΔΙΠΛΗΣ ΕΙΣΟΔΟΥ / ΔΙΑΣΤΑΥΡΩΣΕΩΝ (3)

ΠΑΣΧΟΝΤΕΣ * ΚΑΠΝΙΣΤΕΣ Crosstabulation

			ΚΑΠΝΙΣΤΕΣ		Total
			ΝΑΙ	ΟΧΙ	
ΠΑΣΧΟΝΤΕΣ	ΝΑΙ	Count	30	2	32
		% within ΠΑΣΧΟΝΤΕΣ	93,8%	6,3%	100,0%
		% within ΚΑΠΝΙΣΤΕΣ	37,5%	10,0%	32,0%
		% of Total	30,0%	2,0%	32,0%
	ΟΧΙ	Count	50	18	68
		% within ΠΑΣΧΟΝΤΕΣ	73,5%	26,5%	100,0%
		% within ΚΑΠΝΙΣΤΕΣ	62,5%	90,0%	68,0%
		% of Total	50,0%	18,0%	68,0%
Total	Count	80	20	100	
	% within ΠΑΣΧΟΝΤΕΣ	80,0%	20,0%	100,0%	
	% within ΚΑΠΝΙΣΤΕΣ	100,0%	100,0%	100,0%	
	% of Total	80,0%	20,0%	100,0%	



Παράδειγμα 2.10

Ένα δείγμα αναγνωστών ρωτήθηκε για το ποια εφημερίδα διαβάζει: Globe and Mail (1) Post (2), Star (3), Sun (4), και για το εάν ασκούν χειρονακτική εργασία (1), γραφική εργασία (2), ή ελεύθερο επάγγελμα (3).

Οι απαντήσεις βρίσκονται στο αρχείο [Xm02-10](#).

Παράδειγμα 2.10

Σε μια μεγάλη Αμερικανική πόλη υπάρχουν τέσσερις εφημερίδες που ανταγωνίζονται μεταξύ τους, οι: Post, Globe and Mail, Sun, και Star.

Για το σχεδιασμό της διαφημιστικής πολιτικής, τα τμήματα διαφήμισης των εφημερίδων θέλουν να γνωρίζουν ποια τμήματα της αγοράς διαβάζουν κάθε εφημερίδα.

Πραγματοποιήθηκε μια έρευνα για να εξεταστεί η σχέση μεταξύ επαγγέλματος και ανάγνωσης εφημερίδων.

Παράδειγμα 2.10

Μετρώντας πόσες φορές εμφανίζεται κάθε ένας από τους 12 συνδυασμούς, δημιουργούμε τον Πίνακα 2.9.

Εφημερίδα	Απασχόληση			Σύνολο
	Χειρονακτική εργασία	Γραφική εργασία	Ελεύθερο επάγγελμα	
G&M	27	29	33	89
Post	18	43	51	112
Star	38	21	22	81
Sun	37	15	20	72
Σύνολο	120	108	126	354

Παράδειγμα 2.10

Εάν η απασχόληση και η εφημερίδα συνδέονται, τότε θα υπάρχουν διαφορές μεταξύ των επαγγελμάτων. Ένας εύκολος τρόπος είναι να μετατρέψουμε τις συχνότητες κάθε στήλης σε σχετικές συχνότητες. Δηλαδή, να υπολογίσουμε το άθροισμα της στήλης και να διαιρέσουμε κάθε συχνότητα με το άθροισμα αυτό.

Απασχόληση

	Χειρονακτική εργασία	Γραφική εργασία	Ελεύθερο επάγγελμα
G&M	$27/120 = .23$	$29/108 = .27$	$33/126 = .26$
Post	$18/120 = .15$	$43/108 = .40$	$51/126 = .40$
Star	$38/120 = .32$	$21/108 = .19$	$22/126 = .17$
Sun	$37/120 = .31$	$15/108 = .14$	$20/126 = .16$

Παράδειγμα 2.10

Ερμηνεία: Οι σχετικές συχνότητες στις στήλες 2 & 3 είναι παρόμοιες, αλλά υπάρχουν μεγάλες διαφορές μεταξύ των στηλών 1 και 2, και μεταξύ των στηλών 1 και 3.

Πίνακας 2. 10 Σχετικές Συχνότητες Στηλών για το Παράδειγμα 2.8

Εφημερίδα	Απασχόληση		
	Χειρονακτική εργασία	Γραφική εργασία	Ελεύθερο επάγγελμα
G&M	27/120 = .23	29/108 = .27	33/126 = .26
Post	18/120 = .15	43/108 = .40	51/126 = .40
Star	38/120 = .32	21/108 = .19	22/126 = .17
Sun	37/120 = .31	15/108 = .14	20/126 = .16

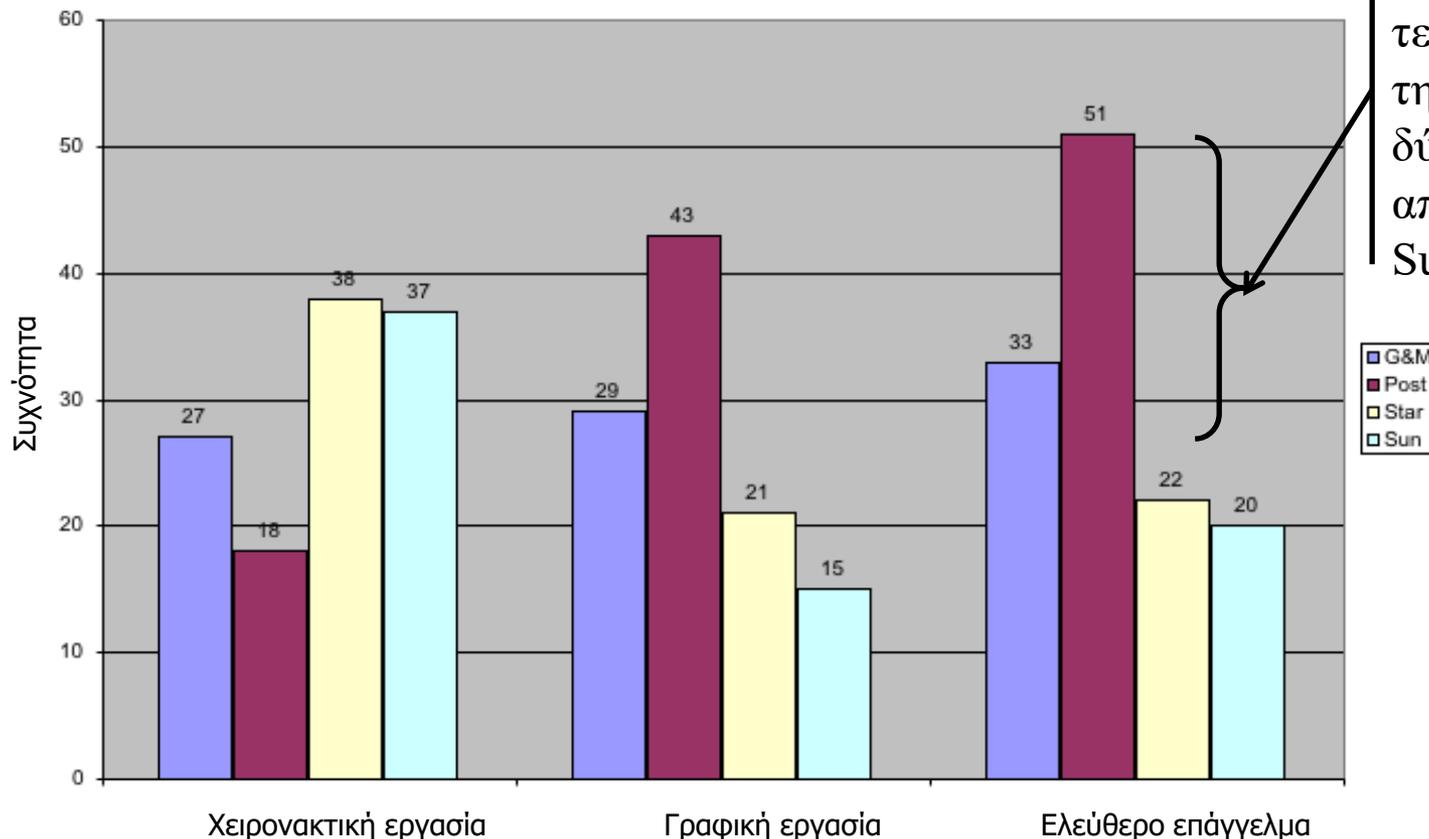
ανόμοιες

Άρα οι εργαζόμενοι χειρονακτικά τείνουν να διαβάζουν διαφορετικές εφημερίδες από τους εργαζόμενους γραφικά και τους ελεύθερους επαγγελματίες, ενώ οι εργαζόμενοι γραφικά και οι ελεύθεροι επαγγελματίες κάνουν παρόμοιες επιλογές σχετικά με την εφημερίδα που διαβάζουν.

Απεικόνιση της Σχέσης Δύο Ονομαστικών Μεταβλητών...

Χρησιμοποιούμε τα δεδομένα από τον πίνακα διπλής εισόδου για να κατασκευάσουμε ραβδογράμματα.

Ραβδογράμματα για το Παράδειγμα 2.8

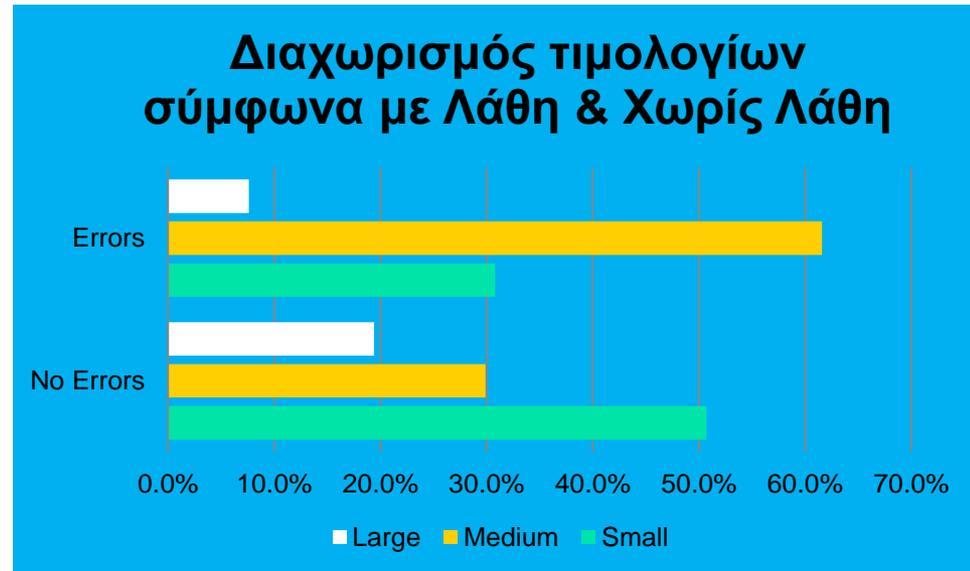


Οι επαγγελματίες τείνουν να διαβάζουν την Globe & Mail δύο φορές συχνότερα από ότι την Star ή την Sun...

Γραφική Απεικόνιση Κατηγορικών Δεδομένων: Πολλαπλά Ραβδογράμματα

- Το **πολλαπλό ραβδόγραμμα** αναπαριστά τα δεδομένα από ένα πίνακα συνάφειας.

	Χωρίς Λάθη	Με Λάθη	Σύνολο
Μικρό Ποσό	50.75%	30.77%	47.50%
Μεσαίο Ποσό	29.85%	61.54%	35.00%
Μεγάλο Ποσό	19.40%	7.69%	17.50%
Σύνολο	100.0%	100.0%	100.0%



Τιμολόγια με λάθη είναι περισσότερο πιθανό να είναι μεσαίου μεγέθους (61.54% έναντι 30.77% και 7.69%)

Γραφική Απεικόνιση Κατηγορικών Δεδομένων: Το Διάγραμμα Pareto

- Χρησιμοποιείται για να αναπαραστήσει κατηγορικά δεδομένα
- Ένα κάθετο ραβδόγραμμα, όπου οι κατηγορίες παρουσιάζονται σε φθίνουσα σειρά σύμφωνα με τη συχνότητα
- Ένα αθροιστικό πολύγωνο φαίνεται στο ίδιο γράφημα
- Χρησιμοποιείται για να διαχωρίσει τις “σημαντικές λίγες” από τις “ασήμαντα πολλές”

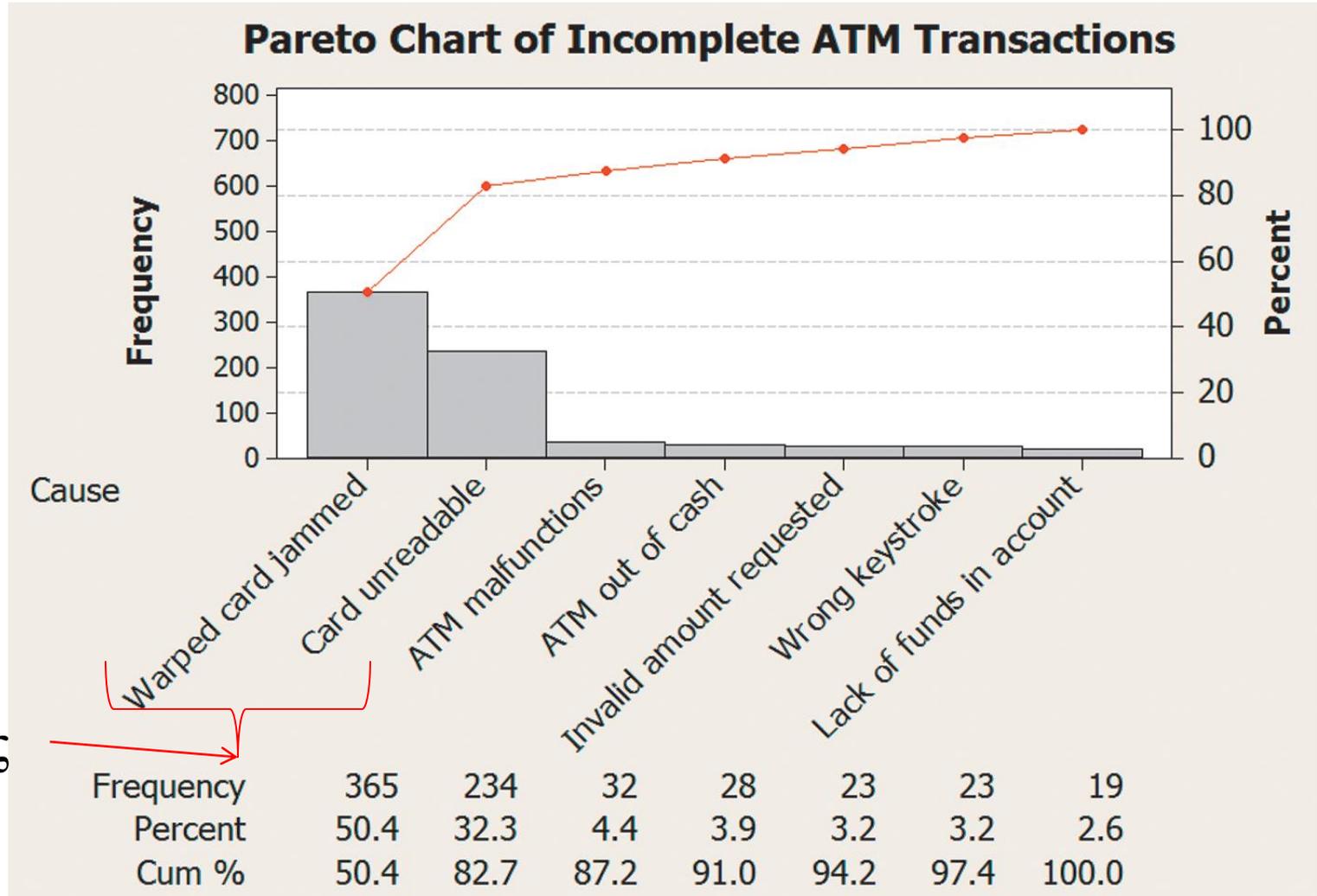
Γραφική Απεικόνιση Κατηγορικών Δεδομένων: Το Διάγραμμα Pareto (συνέχεια)

Πίνακας Συχνοτήτων για τις αιτίες μη ολοκληρωμένων συναλλαγών στα ATMs

Αιτία	Συχνότητα	Ποσοστό	Αθροιστικό
			Ποσοστό
Μπλοκάρισμα κάρτας	365	50.41%	50.41%
Αδυναμία αναγνώρισης κάρτας	234	32.32%	82.73%
Δυσλειτουργίες ATM	32	4.42%	87.15%
Έλλειψη ρευστότητας στα ATM	28	3.87%	91.02%
Λανθασμένο ποσό	23	3.18%	94.20%
Λάθος πληκτρολόγηση	23	3.18%	97.38%
Έλλειψη κεφαλαίου στο λογαριασμό	19	2.62%	100.00%
Σύνολο	724	100.00%	

Πηγή: A. Bhalla, "Don't Misuse the Pareto Principle," *Six Sigma Forum Magazine*, May 2009, pp. 15–18.

Γραφική Απεικόνιση Κατηγορικών Δεδομένων: Το Διάγραμμα Pareto (συνέχεια)



Απεικόνιση της Σχέσης Δύο **Συνεχών** Μεταβλητών...

Περνώντας από τα ονομαστικά στα συνεχή δεδομένα, μας ενδιαφέρει τώρα το πώς συνδέονται δύο συνεχείς μεταβλητές.

Για να εξετάσουμε αυτή τη σχέση χρησιμοποιούμε το **διάγραμμα διασποράς**, στο οποίο σχεδιάζεται η μία μεταβλητή ως προς την άλλη.

Η **ανεξάρτητη** μεταβλητή καλείται X και συνήθως τοποθετείται στον οριζόντιο άξονα, ενώ η άλλη, η **εξαρτημένη** μεταβλητή, καλείται Y , και απεικονίζεται στον κατακόρυφο άξονα.

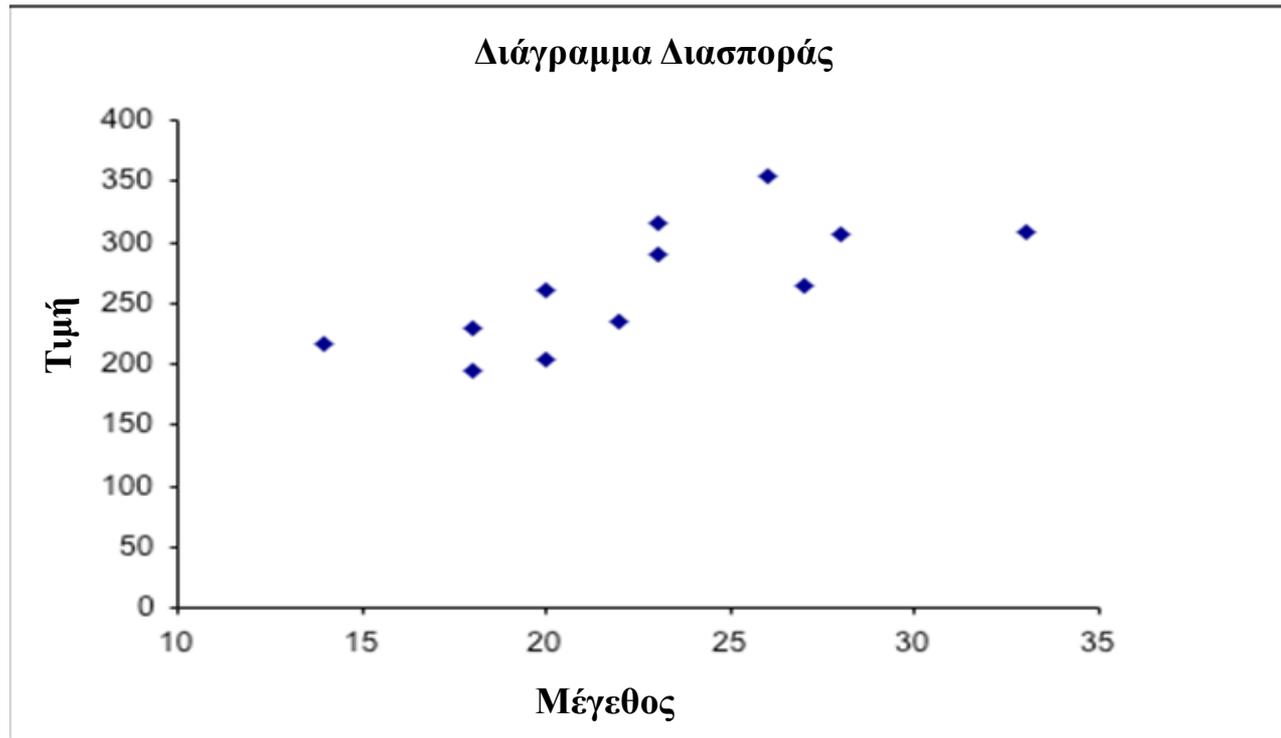
Παράδειγμα 2.12

Ένα κτηματομεσιτικό γραφείο θέλει να μάθει σε ποιο βαθμό η τιμή πώλησης ενός σπιτιού σχετίζεται με το μέγεθός του. Για να αποκτήσει την πληροφορία αυτή, χρησιμοποίησε δείγμα 12 σπιτιών τα οποία πουλήθηκαν πρόσφατα, καταγράφοντας την τιμή σε χιλιάδες δολάρια και το μέγεθος σε εκατοντάδες τετραγωνικά πόδια. Τα δεδομένα παρουσιάζονται στον παρακάτω πίνακα. Να χρησιμοποιήσετε ένα γράφημα για να περιγράψετε τη σχέση μεταξύ τιμής και μεγέθους. [Xm02-12](#)

Μέγεθος	23	18	26	20	22	14	33	28	23	20	27	18
Τιμή	315	229	355	261	234	216	308	306	289	204	265	195

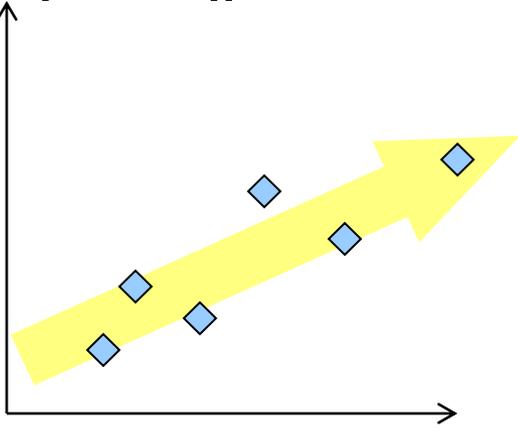
Παράδειγμα 2.12

Φαίνεται ότι πράγματι υπάρχει κάποια σχέση, δηλαδή, όσο μεγαλύτερο είναι το σπίτι τόσο μεγαλύτερη είναι η τιμή ΤΟΥ...

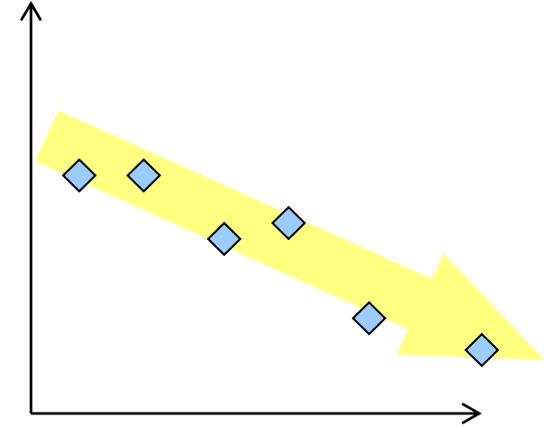


Σχήματα Διαγραμμάτων Διασποράς...

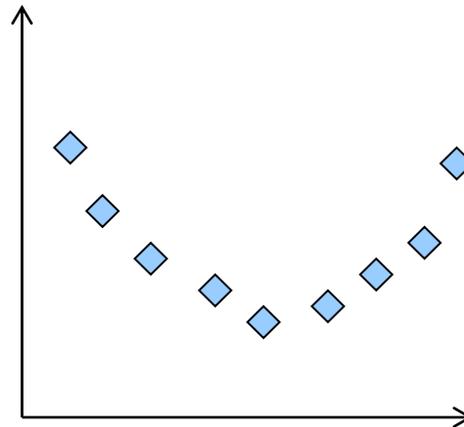
Η **Γραμμικότητα** και η **Κατεύθυνση** είναι δύο χαρακτηριστικά που μας ενδιαφέρουν



Θετική Γραμμική Σχέση



Αρνητική Γραμμική Σχέση



Ασθενής ή Μη-γραμμική Σχέση

Παράδειγμα

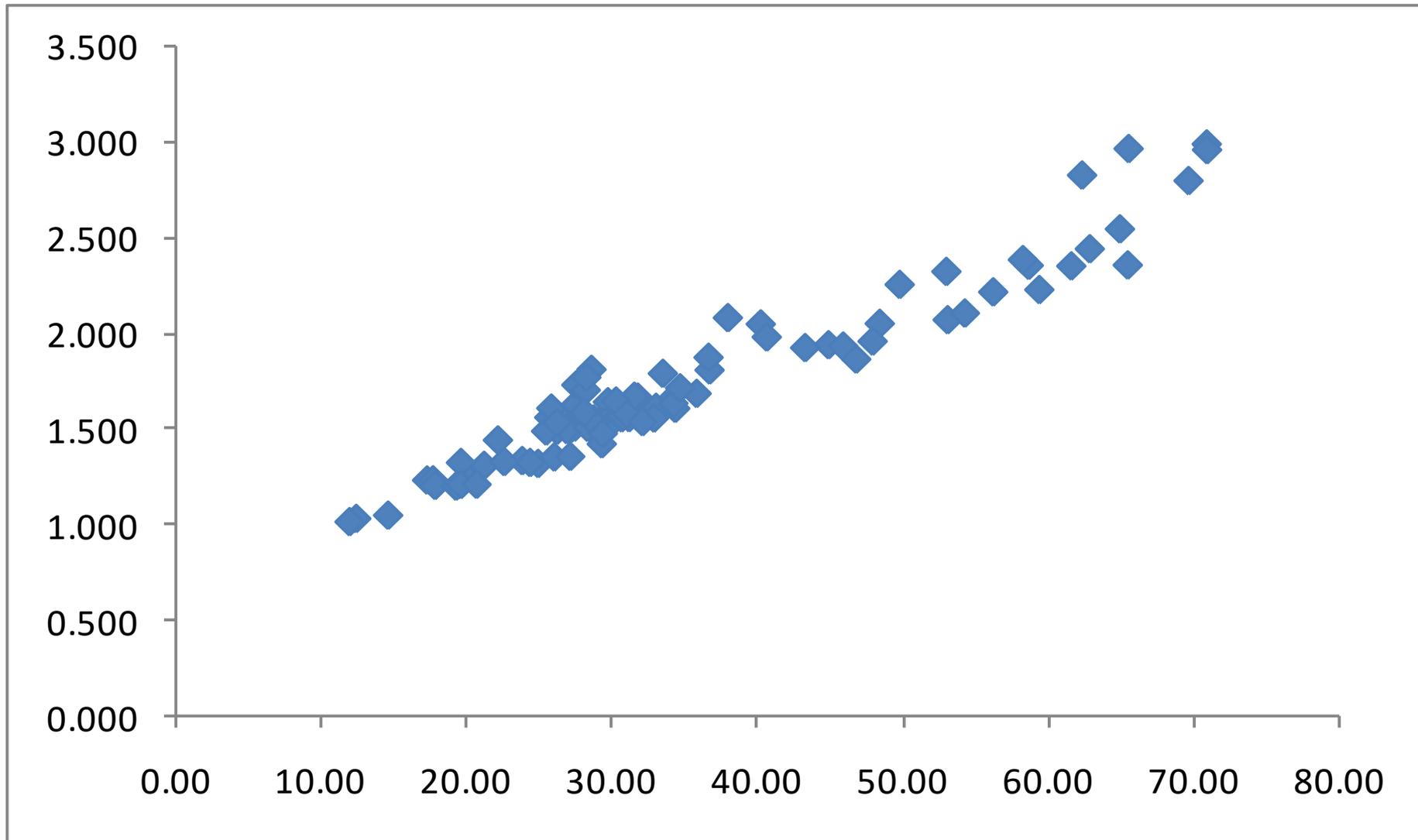
ΤΙΜΕΣ ΑΡΓΟΥ ΠΕΤΡΕΛΑΙΟΥ ΚΑΙ ΒΕΝΖΙΝΗΣ 1999-2006

Τον Δεκέμβριο του 1998 η μέση λιανική τιμή της βενζίνης ήταν \$0.913 το γαλόνι, και η τιμή του πετρελαίου (West Texas αργό πετρέλαιο) ήταν \$11.28 το βαρέλι. Στα 8 χρόνια που ακολούθησαν και οι δύο τιμές αυξήθηκαν σημαντικά. Πολλοί οδηγοί παραπονούνται ότι οι πετρελαϊκές εταιρίες έφταιγαν για την εκτόξευση των τιμών. Πίστευαν, πώς όταν η τιμή του πετρελαίου αυξανόταν το ίδιο συνέβαινε και με την τιμή της βενζίνης, αλλά όταν η τιμή του πετρελαίου μειωνόταν η μείωση στην τιμή της βενζίνης καθυστερούσε.

Για να καθορίσουν εάν η εντύπωση αυτή είναι σωστή, κατέγραψαν τις μηνιαίες τιμές για τα δύο είδη. [Xm02-00](#)

Να απεικονίσετε γραφικά τα δεδομένα αυτά και να περιγράψετε τα ευρήματα.

Παράδειγμα



Ανακεφαλαίωση Ι...

Παράγοντες που Καθορίζουν τη Χρήση Κατανομών Συχνοτήτων και Σχετικών Συχνοτήτων, Ραβδογραμμάτων και Κυκλικών Διαγραμμάτων.

1. Στόχος : Περιγραφή ενός μόνο συνόλου δεδομένων.
2. Τύπος δεδομένων : Ονομαστικά

Παράγοντες που Καθορίζουν τη Χρήση Ιστογραμμάτων, Αθροιστικών Πολυγωνικών Διαγραμμάτων, ή Φυλλογραφημάτων

1. Στόχος : Περιγραφή ενός μόνο συνόλου δεδομένων.
2. Τύπος δεδομένων : Συνεχή

Παράγοντες που Καθορίζουν τη Χρήση Πίνακα Διπλής Εισόδου

1. Στόχος : Περιγραφή της σχέσης μεταξύ δύο μεταβλητών.
2. Τύπος δεδομένων : Ονομαστικά

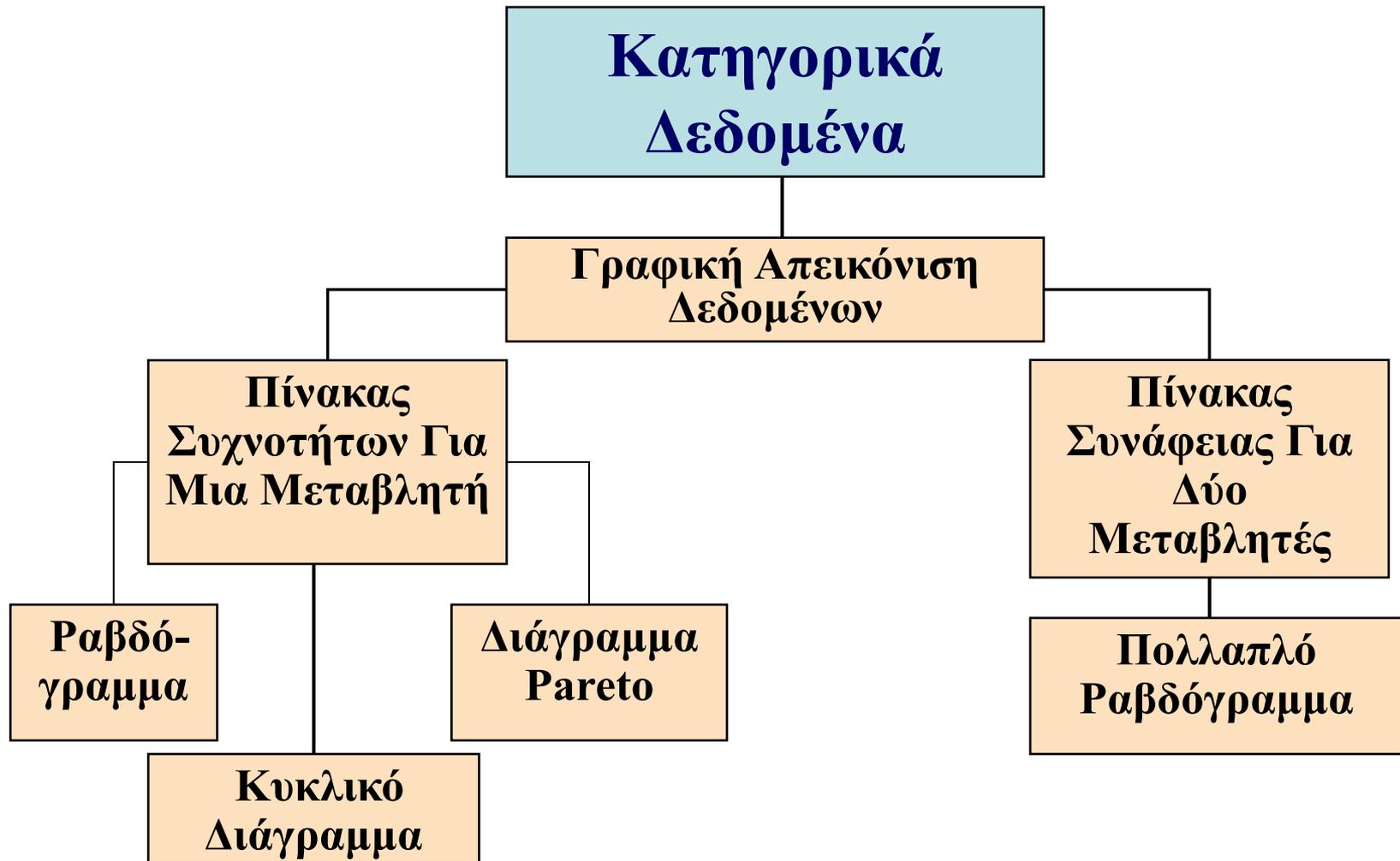
Παράγοντες που Καθορίζουν τη Χρήση Διαγραμμάτων Διασποράς

1. Στόχος : Περιγραφή της σχέσης μεταξύ δύο μεταβλητών.
2. Τύπος δεδομένων : Συνεχή

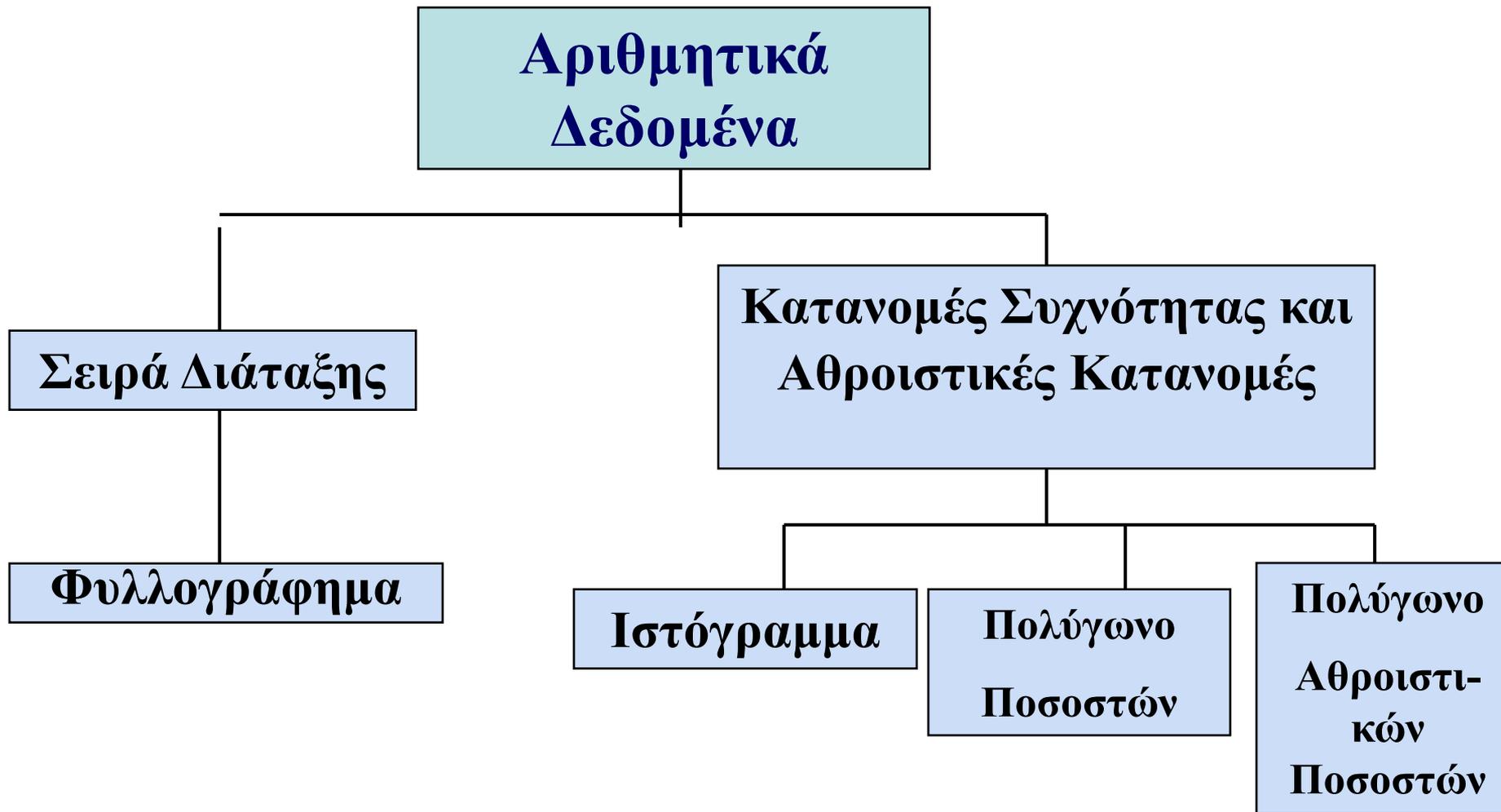
Ανακεφαλαίωση ΙΙ...

	Συνεχή Δεδομένα	Ονομαστικά Δεδομένα
Ένα μόνο σύνολο δεδομένων	Ιστόγραμμα	Κατανομή Συχνοτήτων και Σχετικών Συχνοτήτων, Ραβδόγραμμα & Κυκλικό Διάγραμμα
Σχέση μεταξύ δύο μεταβλητών	Διάγραμμα Διασποράς	Πίνακας Διπλής Εισόδου, Ραβδόγραμμα

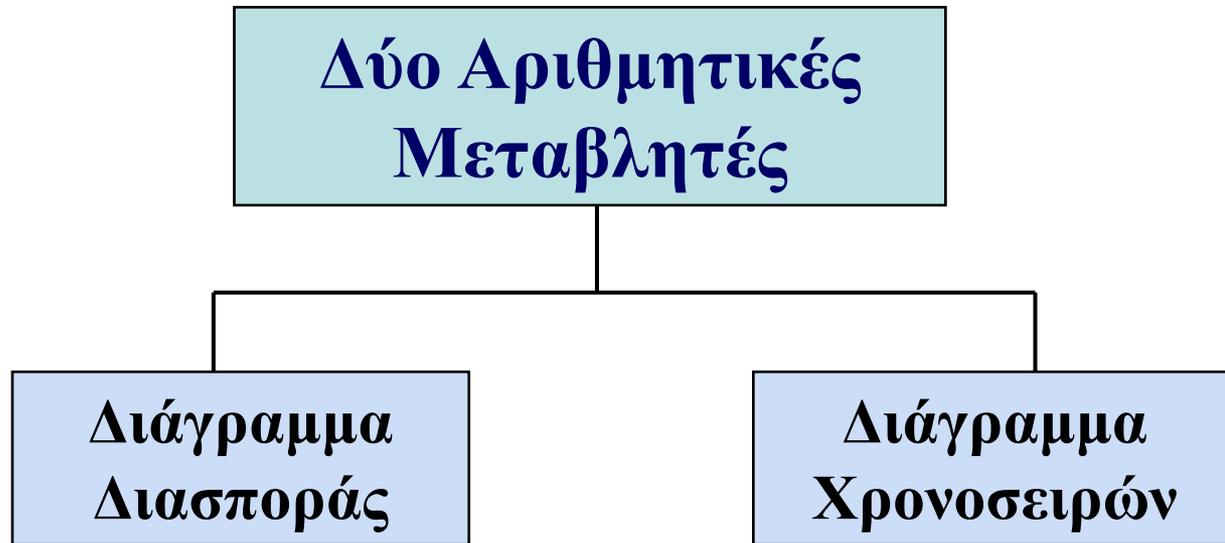
Ανακεφαλαίωση ΙΙ... Απεικόνιση Κατηγορικών Δεδομένων



Απεικόνιση Αριθμητικών Δεδομένων με την Χρήση Γραφικών Απεικονίσεων



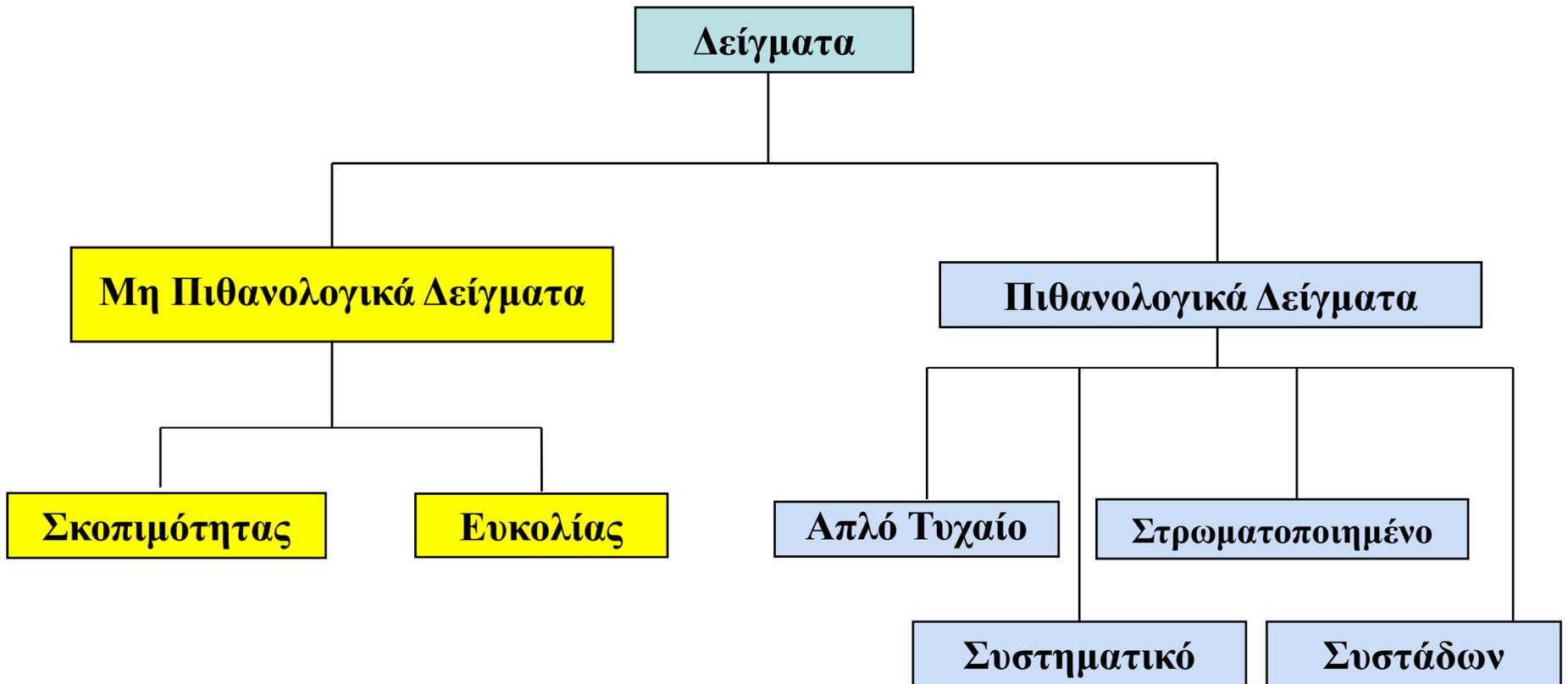
Απεικόνιση Δύο Αριθμητικών Μεταβλητών με τη Χρήση Γραφικών Απεικονίσεων



Δειγματοληψία ...

- Η επαγωγική στατιστική μας επιτρέπει να εξάγουμε συμπε-ράσματα για έναν πληθυσμό εξετάζοντας ένα δείγμα του.
- Δειγματοληψία (δηλ. επιλογή ενός υποσυνόλου του πληθυσμού) γίνεται συχνά για λόγους **κόστους** (είναι οικονομικότερο να εξετάσεις 1.000 τηλεθεατές από ότι 100 εκατομμύρια τηλεθεατές) και **πρακτικότητας** (είναι αδύνατο να γίνει δοκιμή πρόσκρουσης σε κάθε παραγόμενο αυτοκίνητο).
- Σε κάθε περίπτωση, ο **πληθυσμός του δείγματος** και ο **πληθυσμός-στόχος** θα πρέπει να **είναι όμοιοι**.

Είδη Δειγμάτων

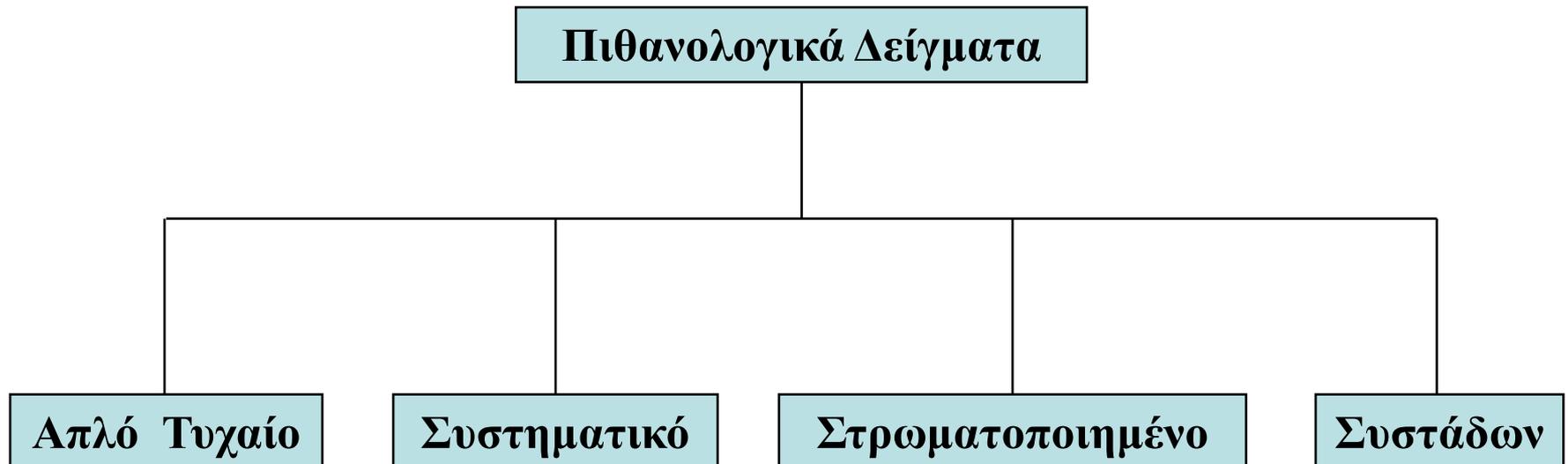


Είδη Δειγμάτων: Μη Πιθανολογικό Δείγμα

- Σε ένα μη πιθανολογικό δείγμα, τα στοιχεία που περιλαμβάνονται επιλέγονται ανεξάρτητα από την πιθανότητα εμφάνισής τους.
 - Σε ένα **δείγμα ευκολίας**, τα στοιχεία επιλέγονται με βάση μόνο το γεγονός ότι είναι εύκολο, οικονομικό ή βολικό για δειγματοληψία.
 - Σε ένα **δείγμα σκοπιμότητας**, μπορείτε να λάβετε τις απόψεις των προεπιλεγμένων ειδικών σχετικά με το θέμα.

Είδη Δειγμάτων: Πιθανολογικό Δείγμα

- Σε ένα *πιθανολογικό δείγμα*, τα στοιχεία του δείγματος επιλέγονται βάσει γνωστών πιθανοτήτων.



Πιθανολογικό Δείγμα: Απλό Τυχαίο Δείγμα

- Κάθε άτομο ή στοιχείο από το πλαίσιο έχει ίσες πιθανότητες να επιλεγεί.
- Η επιλογή μπορεί να γίνει με επανατοποθέτηση (το επιλεγμένο άτομο ή στοιχείο επιστρέφει στο πλαίσιο για πιθανή επανεπιλογή) ή χωρίς επανατοποθέτηση (το επιλεγμένο άτομο ή στοιχείο δεν επιστρέφει στο πλαίσιο).
- Τα δείγματα λαμβάνονται από πίνακα τυχαίων αριθμών ή από γεννήτριες τυχαίων αριθμών του υπολογιστή.

Επιλογή ενός Απλού Τυχαίου Δείγματος χρησιμοποιώντας έναν Πίνακα Τυχαίων Αριθμών

Δειγματοληπτικό Πλαίσιο
Για Πληθυσμό Με 850
Στοιχεία

Όνομα Στοιχείου Στοιχείο #

Bev R.	001
Ulan X.	002
.	.
.	.
.	.
.	.
Joann P.	849
Paul F.	850

Τμήμα ενός Πίνακα Τυχαίων Αριθμών

49280 88924 35779 00283 81163 07275
11100 02340 12860 74697 96644 89439
09893 23997 20048 49420 88872 08401

Τα πρώτα 5 στοιχεία σε ένα απλό τυχαίο δείγμα

Στοιχείο # 492

Στοιχείο # 808

Στοιχείο # 892 -- δεν υπάρχει οπότε αγνοήστε το

Στοιχείο # 435

Στοιχείο # 779

Στοιχείο # 002

Πιθανολογικό Δείγμα: Συστηματικό Δείγμα

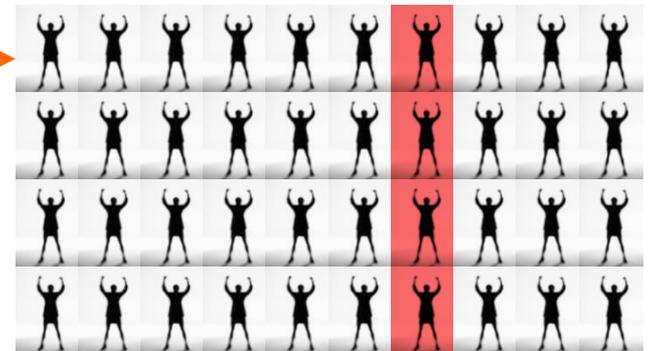
- Επιλέξτε το μέγεθος του δείγματος: n
- Χωρίστε το πλαίσιο των N ατόμων σε ομάδες των k ατόμων:
 $k=N/n$
- Τυχαία επιλέξτε ένα άτομο από την 1η ομάδα
- Επιλέξτε στη συνέχεια κάθε k ο άτομο

$$N = 40$$

$$n = 4$$

$$k = 10$$

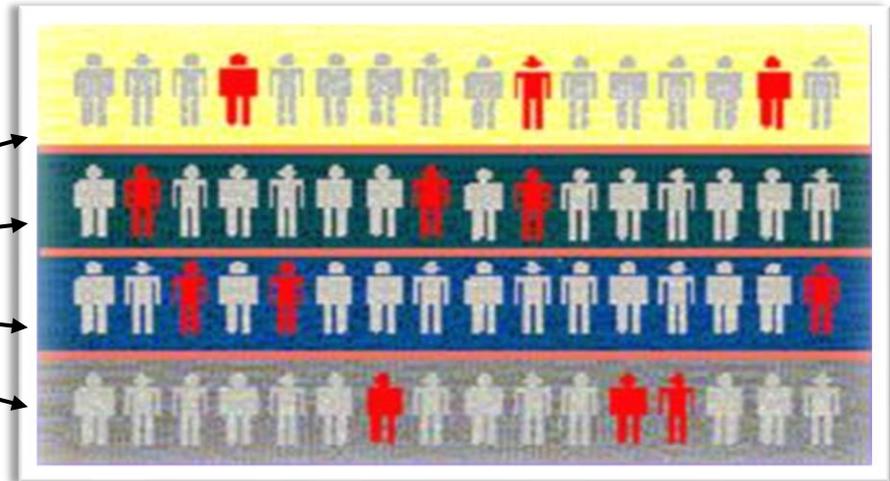
Πρώτη Ομάδα



Πιθανολογικό Δείγμα: Στρωματοποιημένο Δείγμα

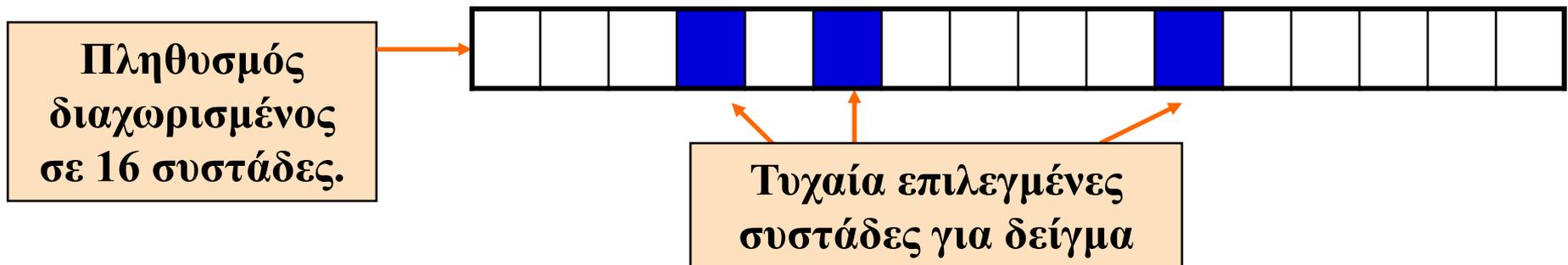
- Διαχωρίστε τον πληθυσμό σε δύο ή περισσότερες υποομάδες (που ονομάζονται στρώματα) σύμφωνα με ορισμένα κοινά χαρακτηριστικά
- Ένα απλό τυχαίο δείγμα επιλέγεται από κάθε υποομάδα, με μεγέθη δείγματος ανάλογα με τα μεγέθη των στρωμάτων
- Δείγματα από τις υποομάδες συνδυάζονται σε ένα
- Αυτή είναι μια κοινή τεχνική κατά τη δειγματοληψία του πληθυσμού των ψηφοφόρων, με στρωματοποίηση στις φυλετικές ή κοινωνικοοικονομικές γραμμές.

**Πληθυσμός
χωρισμένος
σε 4
στρώματα**



Πιθανολογικό Δείγμα: Δείγμα Συστάδων

- Ο πληθυσμός χωρίζεται σε διάφορες “συστάδες”, καθεμιά αντιπροσωπευτική του πληθυσμού
- Επιλέγεται ένα απλό τυχαίο δείγμα συστάδων
- Μπορούν να χρησιμοποιηθούν όλα τα στοιχεία στις επιλεγμένες συστάδες ή τα στοιχεία μπορούν να επιλεγούν από μια συστάδα χρησιμοποιώντας μια άλλη τεχνική δειγματοληψίας πιθανότητας
- Μια κοινή εφαρμογή δειγματοληψίας συστάδων περιλαμβάνει δημοσκοπήσεις έξω από τα εκλογικά κέντρα, όπου επιλέγονται ορισμένες εκλογικές περιφέρειες και λαμβάνονται δείγματα.



Πιθανολογικό Δείγμα: Σύγκριση Δειγματοληπτικών Μεθόδων

- **Απλό τυχαίο δείγμα και Συστηματικό δείγμα**
 - Απλό στην χρήση
 - Ίσως να μην είναι μια καλή αντιπροσώπευση των υποκείμενων χαρακτηριστικών του πληθυσμού
- **Στρωματοποιημένο δείγμα**
 - Εξασφαλίζει την αντιπροσώπευση ατόμων σε ολόκληρο τον πληθυσμό
- **Δείγμα συστάδων**
 - Περισσότερο οικονομικά αποδοτικό
 - Λιγότερο αποτελεσματικό (χρειάζεται μεγαλύτερο δείγμα για να αποκτήσετε το ίδιο επίπεδο ακρίβειας)

Αξιολόγηση της Αξιοπιστίας της Έρευνας

- Ποιος είναι ο σκοπός της έρευνας;
 - Είναι η έρευνα βασισμένη σε ένα πιθανολογικό δείγμα;
 - Σφάλμα κάλυψης – κατάλληλο πλαίσιο;
 - Σφάλμα μη απάντησης – επανεξέταση
 - Σφάλμα μέτρησης - καλές ερωτήσεις προκαλούν καλές απαντήσεις
 - Δειγματοληπτικό σφάλμα - υπάρχει πάντα
-

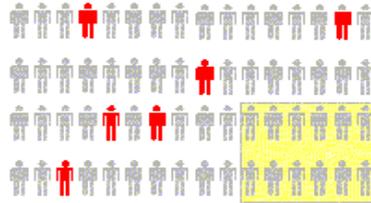
Τύποι Σφαλμάτων Έρευνας

- Σφάλμα κάλυψης ή μεροληπτική επιλογή
 - Υπάρχει αν ορισμένες ομάδες αποκλείονται από το πλαίσιο και δεν έχουν καμία πιθανότητα επιλογής
- Σφάλμα μη απάντησης ή μεροληπτικότητα
 - Τα άτομα που δεν ανταποκρίνονται μπορεί να διαφέρουν από εκείνα που ανταποκρίνονται
- Δειγματοληπτικό σφάλμα
 - Η μεταβλητότητα από δείγμα σε δείγμα θα υπάρχει πάντα
- Σφάλμα μέτρησης
 - Λόγω των αδυναμιών στο σχεδιασμό και / ή το σφάλμα του ερωτηθέντος

Τύποι Σφαλμάτων Έρευνας

DCOVA
(συνέχεια)

- Σφάλμα κάλυψης



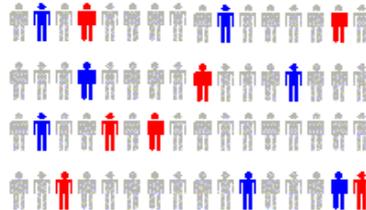
Αποκλεισμός από το
πλαίσιο

- Σφάλμα μη απάντησης



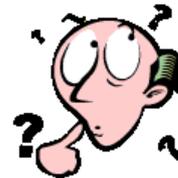
Επανεξέταση των
μη απαντήσεων

- Δειγματοληπτικό σφάλμα



Τυχαίες διαφορές από
δείγμα σε δείγμα

- Σφάλμα μέτρησης



Κακή ή βασική
ερώτηση

Ανακεφαλαίωση ΙΙΙ...

Σε αυτή την ενότητα αναλύσαμε:

- Τους τύπους μεταβλητών που χρησιμοποιούνται στην στατιστική
- Πώς να συλλέγετε δεδομένα
- Τους διαφορετικούς τρόπους να συλλέγετε ένα δείγμα
- Τους τύπους σφαλμάτων έρευνας