ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

ΔΙΑΛΕΞΗ: ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΙ ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Διδάσκουσα: Ε. Γάκη, Επίκ. Καθηγήτρια

MBA χίος
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
διοικώ σημαίνει αποφασίζω

Περιεχόμενα μαθήματος

- Περιγραφική Στατιστική
- Στοιχεία Θεωρίας Πιθανοτήτων
- Κατανομές
- Εκτιμητική
- Ανάλυση Διακύμανσης
- Συσχέτιση - Παλινδρόμηση
- Παραγοντική Ανάλυση
- Ανάλυση κατά Συστάδες

Σκοπός μαθήματος

Οτιδήποτε συνδέεται με τη συλλογή, επεξεργασία, ανάλυση και ερμηνεία αριθμητικών δεδομένων που προέρχονται από παρατηρήσεις ή μετρήσεις και αναφέρονται σε ιδιότητες φυσικών, οικονομικών ή κοινωνικών φαινομένων ανήκει στην επιστημονική περιοχή της Στατιστικής.

Σκοπός τους μαθήματος είναι

- ✓ η εξοικείωση των φοιτητών με τις βασικές αρχές της Περιγραφικής και Επαγωγικής Στατιστικής
- ✓ η εξοικείωση των φοιτητών με ορισμένες από τις μεθόδους Πολυμεταβλητής Στατιστικής Ανάλυσης
- ✓ η εφαρμογή των παραπάνω στη Διοίκηση Επιχειρήσεων.

Μαθησιακά αποτελέσματα

Με την ολοκλήρωση της μαθησιακής διαδικασίας η φοιτήτρια και ο φοιτητής θα είναι σε θέση :

- ✓ να οργανώνει και να παρουσιάζει ένα πολυπληθές σύνολο δεδομένων, χρησιμοποιώντας τις αρχές της γραφικής σύνοψης αυτών
- ✓ να αξιοποιεί τις τιμές των στατιστικών μέτρων προκειμένου να συνοψίσει ένα πολυπληθές σύνολο δεδομένων
- ✓ να χρησιμοποιεί τις τιμές των στατιστικών μέτρων προκειμένου να εκτιμήσει τις παραμέτρους του υπό μελέτη πληθυσμού
- ✓ να εφαρμόζει γραμμικά μοντέλα σε δεδομένα της παρατήρησης
- ✓ να ελέγχει την καταλληλότητα των μοντέλων,
- ✓ να διερευνά τις υποθέσεις των μοντέλων
- ✓ να αναπτύσσει στατιστική συμπερασματολογία για τις παραμέτρους των μοντέλων
- ✓ να κάνει προβλέψεις με τη χρήση των μοντέλων
- ✓ να εφαρμόζει διάφορες μορφές μη γραμμικών μοντέλων σε δεδομένα και να αξιολογεί την καταλληλότητά τους
- ✓ να εφαρμόζει τη μεθοδολογία της Ανάλυσης Διακύμανσης κατά ένα και κατά δύο παράγοντες

Υλικό μαθήματος

- Διαφάνειες παρουσιάσεων στο https://eclass.aegean.gr/courses/211117/

- Συγγράμματα

- ✓ Αγγελής, Β., Δημάκη Αικ., *Στατιστική Τόμος Α΄,* Χίος, 2010.
- ✓ Bamberg, g., Baur, Fr., Krapp, M. (επιμέλεια Θ. Καλαντζής), *Στατιστική*, Εκδόσεις Προπομπός, 2014
- ✓ Field, A., *Η διερεύνηση της Στατιστικής με τη χρήση του SPSS της IMB, 1η ελληνική έκδοση από την 4η αγγλική,* Εκδόσεις Προπομπός, 2016
- ✓ Ιωαννίδης Δ. Α., *Στατιστικές Μέθοδοι Τόμος Ι,* Εκδόσεις Ζήτη, Θεσσαλονίκης 2001
- ✓ Κουτρουμανίδης Θεόδωρος-Ζαφειρίου Ελένη-Μαλέσιος Χρυσοβαλάντης, Στατιστική ΙΙ, ΕΚΔΟΣΕΙΣ Α. ΤΖΙΟΛΑ & ΥΙΟΙ Α.Ε., 2013
- ✓ Φιλιππάκης Ε. Μιχαήλ, Στατιστικές Μέθοδοι και Ανάλυση Παλινδρόμησης για τις νέες Τεχνολογίες, (Εκδότης): ΦΙΛΙΠΠΑΚΗΣ ΜΙΧΑΗ, 2016
- ✓ Aczel, A. D. and Sounderpandian, J., *Complete Business Statistics*, McGraw – Hill & Irwin, 2002

Περιεχόμενα Διάλεξης

- Στατιστική & Λήψη Αποφάσεων
- Δειγματοληψία & Δειγματοληπτικές Μέθοδοι
- Γραφικές Μέθοδοι Σύνοψης Δεδομένων

Στατιστική ???

Δεδομένα και Διαγράμματα

Παραδείγματα

Στατιστική

Στατιστική

- Βασίζεται σε υπολογισμό αριθμών

- Βασίζεται στο πως επιλέγονται οι αριθμοί και ερμηνεύονται

- Υπάρχουν όμως ορισμένες φορές προβληματικές ερμηνείες

Στατιστική

Παράδειγμα 1 - Παγωτά

Στατιστική

Μια νέα διαφήμιση για τα παγωτά Ben and Jerry's που ανακοινώθηκε πέρυσι το Μάιο είχε ως αποτέλεσμα την αύξηση κατά 30% των πωλήσεων παγωτού τους επόμενους τρεις μήνες.

Επομένως, η διαφήμιση ήταν αποτελεσματική!

……Υπάρχει κάτι προβληματικό σε αυτή τη λογική????

Στατιστική

ΝΑΙ!!!

Η κατανάλωση παγωτού συνήθως αυξάνεται τους καλοκαιρινούς μήνες.

Θεωρήθηκε ότι η διαφήμιση αύξησε τις πωλήσεις. Όμως, μια άλλη μεταβλητή, εδώ η χρονική στιγμή, φαίνεται να επηρεάζει σημαντικά τις πωλήσεις.

Στατιστική

Παράδειγμα 2 – Εκκλησίες και Εγκλήματα

Στατιστική

Όσο περισσότερες εκκλησίες υπάρχουν σε μια πόλη, τόσο περισσότερα εγκλήματα γίνονται.

Επομένως, οι εκκλησίες οδηγούν σε εγκλήματα!

……Υπάρχει κάτι προβληματικό σε αυτή τη λογική????

Στατιστική

ΝΑΙ!!!

Περισσότερες εκκλησίες και περισσότερα εγκλήματα μπορούν να εξηγηθούν από μεγάλους πληθυσμούς.

Στις μεγαλύτερες πόλεις υπάρχουν και τα δύο: περισσότερες εκκλησίες και περισσότερα εγκλήματα.

Εδώ υπάρχει μια τρίτη μεταβλητή που ερμηνεύει τη σχέση των δύο. Λανθασμένα όμως, δεν λαμβάνεται υπόψη και ορισμένοι μπορεί να θεωρήσουν ότι υπάρχει πράγματι σχέση μεταξύ αριθμού εκκλησιών και εγκλημάτων.

Συνοψίζοντας

Η στατιστική ΔΕΝ είναι μόνο δεδομένα και διαγράμματα!

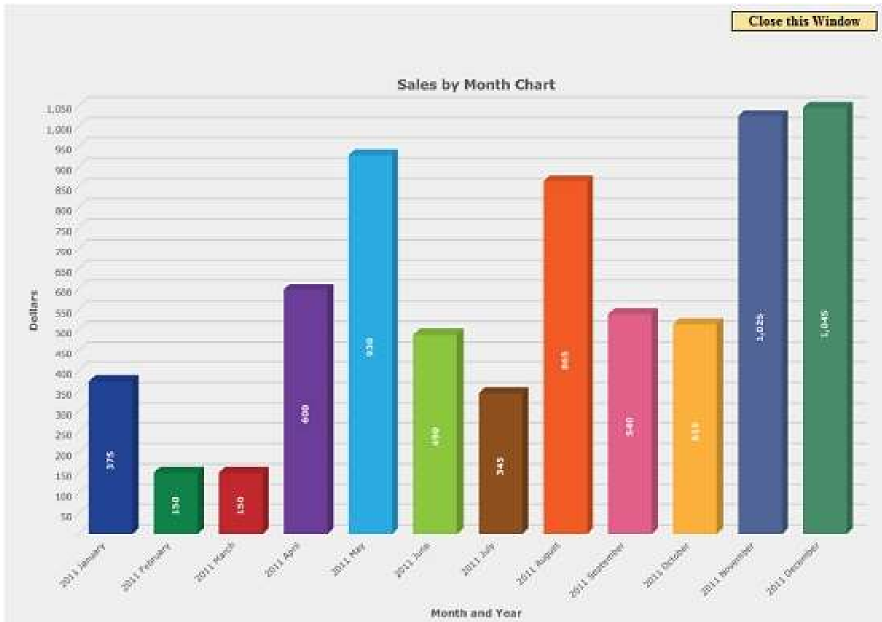
Είναι τεχνικές και διαδικασίες που περιλαμβάνουν

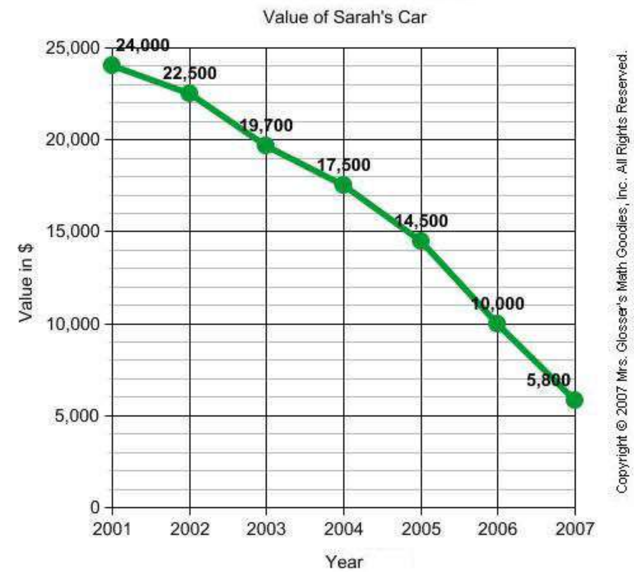
- Ανάλυση
- Παρουσίαση
- Ερμηνεία
- Λήψη αποφάσεων

με βάση τα δεδομένα

Επομένως

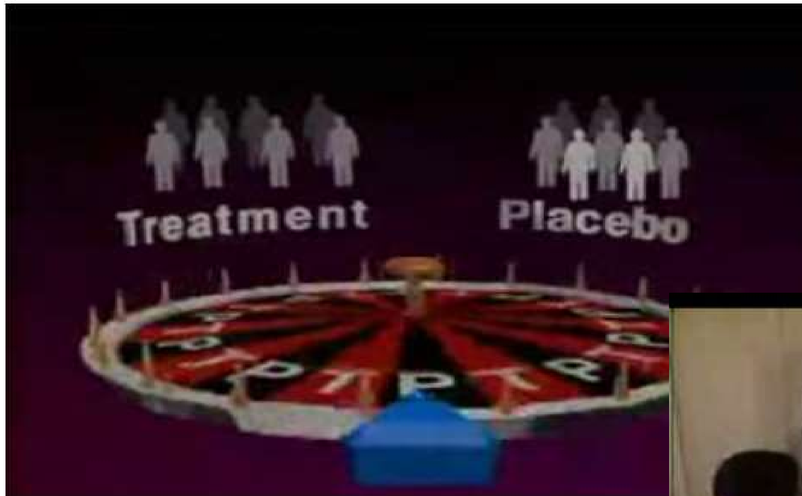
Για να περιγράψουμε δεδομένα χρειαζόμαστε γραφήματα (graphs/charts)





Επομένως

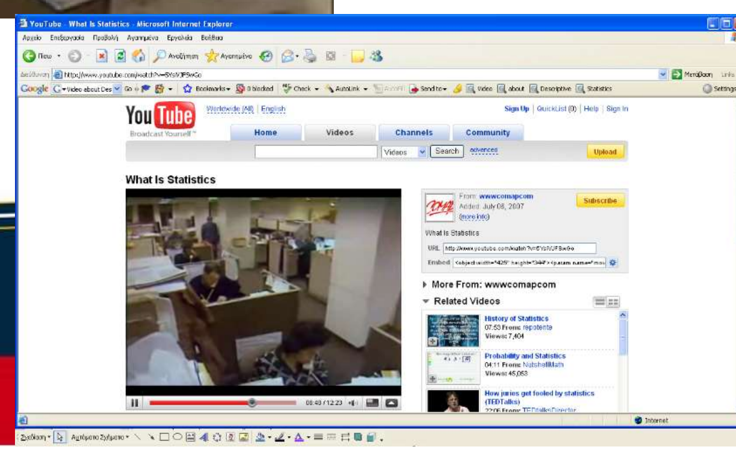
Για να παράγουμε δεδομένα πραγματοποιούμε



πειράματα

Έρευνες πεδίου

Τηλεφωνικές Έρευνες



Επομένως

Για να καταλήξουμε σε αποτελέσματα έχουμε ανάγκη την στατιστική

but……..

*There are three kinds of **lies**;*
*there are lies, damned lies and **statistics**.*

The Royal Statistical Society Schools Lecture 2004: 'Lies and Statistics', Part 1

Why???

Τι προέκυψε από τα παραπάνω???

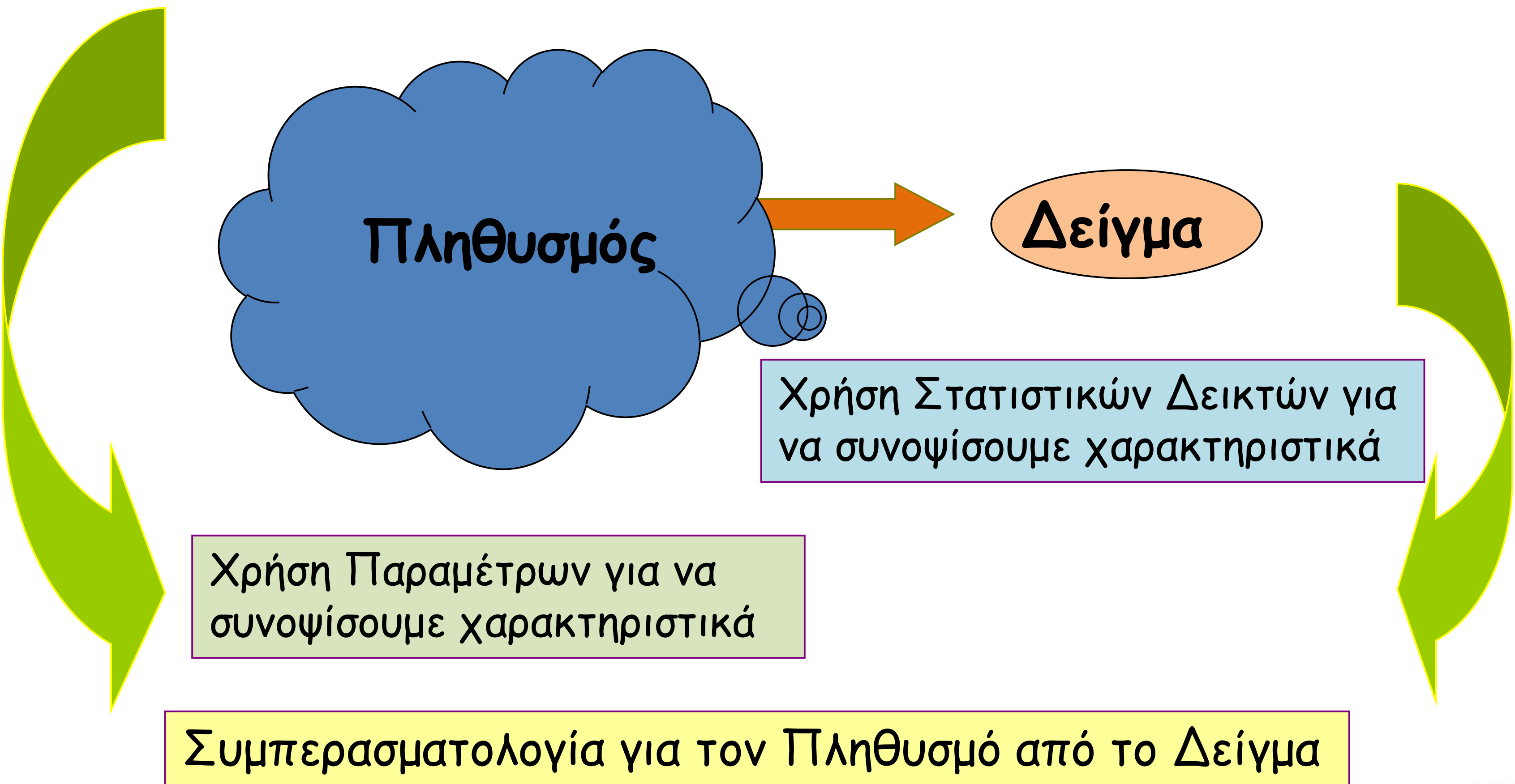
…ότι η Στατιστική σχετίζετε με τρεις ενέργειες:

- Να συλλέξουμε τα στοιχεία
- Να χρησιμοποιήσουμε γραφήματα για να τα περιγράψουμε
- Να χρησιμοποιήσουμε τις αρχές της στατιστικής συμπερασματολογίας για να οδηγηθούμε σε αποτελέσματα

Απαραίτητοι Ορισμοί

- Πληθυσμός είναι μια συλλογή (ή σύνολο) των υπό μελέτη αντικειμένων.
- Δείγμα είναι ένα υπο - σύνολο του πληθυσμού το οποίο επιλέγεται για ανάλυση - επίβλεψη.
- Παράμετρος είναι ένα περιληπτικό μέτρο που περιγράφει ένα χαρακτηριστικό του πληθυσμού.
- Στατιστικός Δείκτης (ή συνάρτηση) είναι ένα περιληπτικό μέτρο που περιγράφει ένα χαρακτηριστικό του δείγματος.

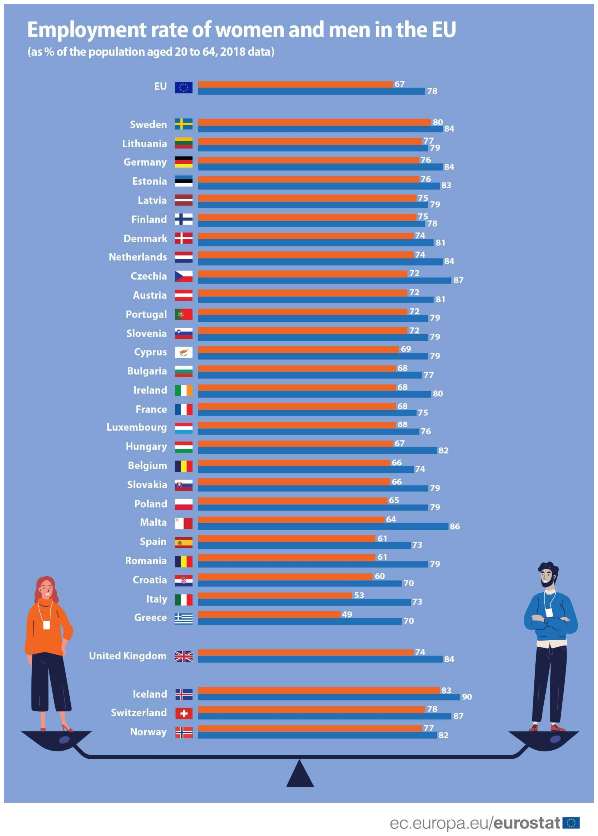
Πληθυσμός & Δείγμα



Στατιστικές Μέθοδοι

- **Περιγραφική Στατιστική**

  Περιγράφει **συστηματικά** και συνοψίζει μερικά από τα **βασικά χαρακτηριστικά των δεδομένων** που προέρχονται από φυσικά ή κοινωνικά φαινόμενα



Στατιστικές Μέθοδοι

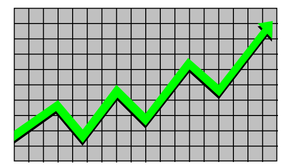
- **Στατιστική Συμπερασματολογία**

Προσπαθεί να **γενικεύσει τα συμπεράσματα** που προκύπτουν από τις περιγραφικές στατιστικές αναλύσεις, να διακρίνει τι υπάρχει «πίσω» από τα δεδομένα, και να **υποβοηθήσει τη διαδικασία λήψης αποφάσεων** .

Περιγραφική Στατιστική

- Συλλογή Δεδομένων
  - π.χ. Δημοσκόπηση

- Παρουσίαση Δεδομένων
  - π.χ. Πίνακες και Διαγράμματα



- Ερμηνεία και Περιγραφή Δεδομένων
  - π.χ. Δειγματικός Μέσος = $\dfrac{\sum X_i}{n}$

Στατιστική Συμπερασματολογία

- Εκτιμητική
  - π.χ. Εκτίμηση του πληθυσμιακού μέσου βάρους μέσω του δειγματικού μέσου βάρους
- Έλεγχος Υποθέσεων
  - π.χ. Έλεγχος του ισχυρισμού ότι το μέσο βάρος του πληθυσμού είναι 80 κιλά

Εξαγωγή Συμπερασμάτων και/ ή αποφάσεων για έναν πληθυσμό με βάση τα δειγματικά αποτελέσματα.

Τύποι Δεδομένων

Δεδομένα

Κατηγορικά (Ποιοτικά)

Τα **ποιοτικά** δεδομένα (qualitative data) είναι παρατηρήσεις που μπορούν να χωρισθούν σε κατηγορίες με βάση κάποιο ποιοτικό χαρακτηριστικό (φύλο, οικογενειακή κατάσταση, γεωγραφική περιφέρεια κ.λ.π).

Αριθμητικά (Ποσοτικά)

Τα **ποσοτικά** δεδομένα (quantitative data) είναι πραγματικοί αριθμοί που αναφέρονται στην τιμή κάποιου χαρακτηριστικού (π.χ. ύψος, βάρος, μισθός κ.λ.π).

Διακριτά

Συνεχή

Για την **μέτρηση** των δεδομένων χρησιμοποιούνται οι παρακάτω τέσσερις κλίμακες μέτρησης:

- Ονομαστική Κλίμακα (Nominal Scale)
- Κλίμακα Διάταξης (Ordinal Scale)
- Κλίμακα Διαστήματος (Interval Scale)
- Κλίμακα Λόγου (ratio Scale)

Τρόποι Συλλογής Δεδομένων

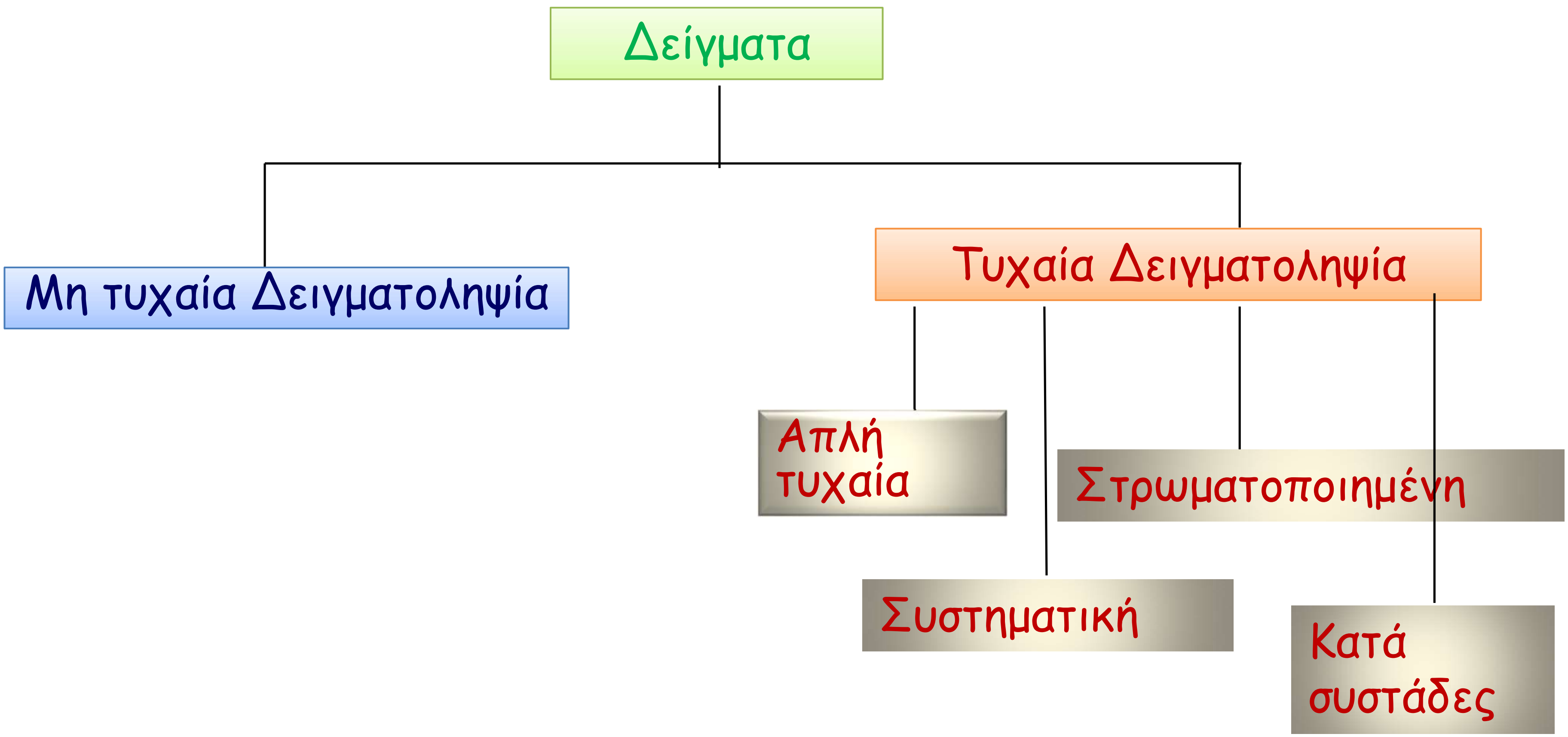
Απογραφική μέθοδος

- Πληροφορίες από κάθε στοιχείο του πληθυσμού.
- Προφανώς μια τέτοια διαδικασία είναι εξαιρετικά χρονοβόρα και πολυδάπανη και για αυτό σπάνια χρησιμοποιείται στην πράξη.

Δειγματοληπτική μέθοδος

- Πληροφορίες από ένα δείγμα το οποίο περιέχει ορισμένα μόνο στοιχεία του πληθυσμού.
- Η δειγματοληπτική μέθοδος είναι σαφώς ταχύτερη και οικονομικότερη της απογραφικής. Το ερώτημα όμως που γεννιέται είναι αν τα συμπεράσματα που βασίζονται σ' αυτή είναι εξίσου αξιόπιστα με τα συμπεράσματα που βασίζονται στην απογραφή

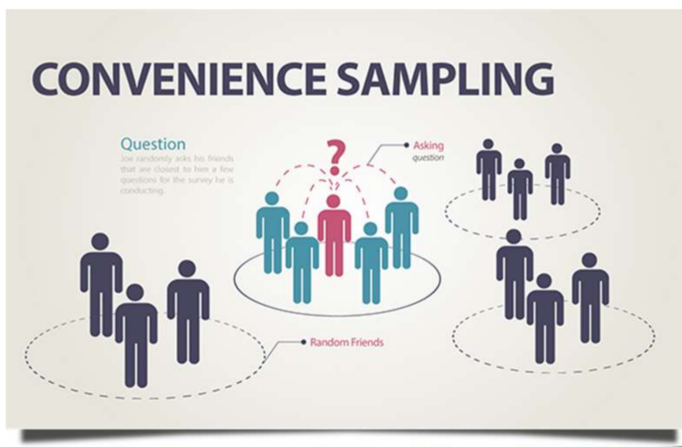
Τύποι δειγματοληπτικών μεθόδων



Μη Τυχαία Δειγματοληψία

Convenience Sampling/ Δειγματοληψία Ευκολίας

- Η επιλογή του δείγματος γίνεται με μοναδικό κριτήριο την ευκολία συλλογής δεδομένων από έναν γνωστό πληθυσμό
- Μεγάλος ο βαθμός ανταπόκρισης των ερωτώμενων
- Τα αποτελέσματα της ανάλυσης δεν είναι μπορούν να γενικευθούν, γιατί το δείγμα είναι άγνωστο για ποιόν πληθυσμό είναι αντιπροσωπευτικό

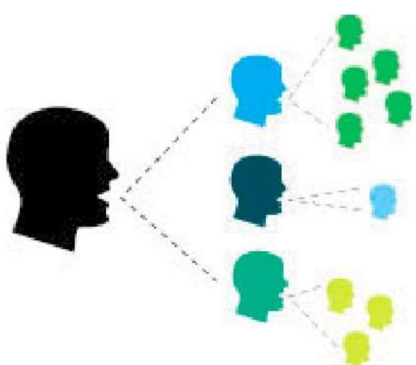


Snowball Sampling

- Χρησιμοποιείται ως μέθοδος δειγματοληψίας σε περιπτώσεις που είναι δύσκολο να έχουμε πρόσβαση σε καποιον πληθυσμό

- Ξεκινάμε την επαφή με μια ομάδα από τον πληθυσμό και εν συνεχεία μέσω αυτης της ομάδας προσεγγίζονται και άλλοι

- Δεν είναι αντιπροσωπευτικό το δείγμα με αποτέλεσμα να μην μπορούν να γενικευτούν τα συμπεράσματα



Quota Sampling/
Δειγματοληψία με ποσοστά

- Το δειγματοληπτικό σχέδιο είναι αποτέλεσμα της κρίσης του ερευνητή και της πρόσβασης στις μονάδες του πληθυσμού
- Το δείγμα που επιλέγεται αντανακλά τις αναλογίες του πληθυσμού σε διαφορετικές κατηγορίες
- Χρησιμοποιείται κυρίως σε έρευνας γνώμης,marketing
- Μπορεί να θεωρηθεί ως αναλογική εκπροσώπηση στο δείγμα επιμέρους κατηγοριών του πληθυσμού

Judgment Sampling/ Δειγματοληψία κρίσης

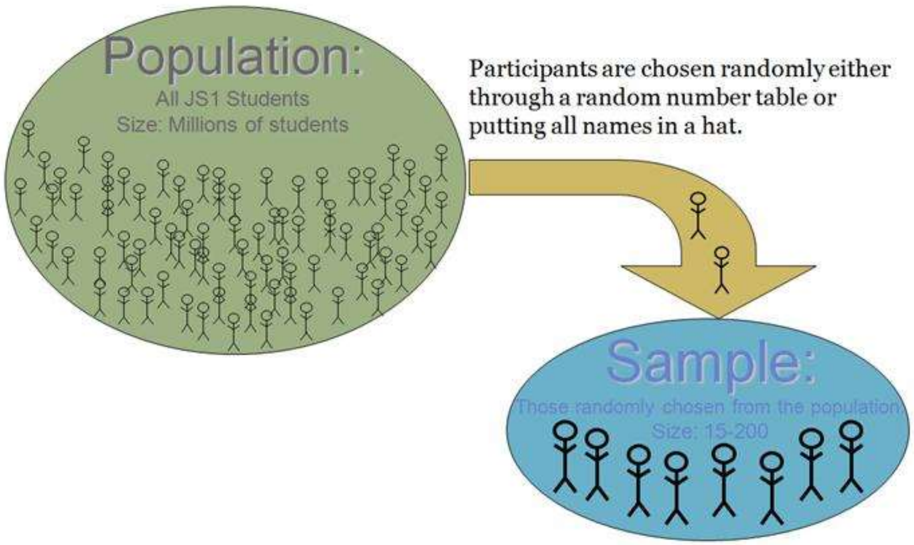
- Το δείγμα που επιλέγεται με βάση την κρίση του ερευνητή
- Τα αποτελέσματα σχετίζονται με τις αντιλήψεις του ερευνητή και δεν μπορούν να γενικευθούν

Τυχαία Δειγματοληψία

(Δειγματοληψία με χρήση Πιθανοτήτων)

Απλή Τυχαία Δειγματοληψία

Είναι δειγματοληπτικό σχέδιο σύμφωνα με το οποίο καθένα από τα ισοπληθή δείγματα που είναι δυνατό να σχηματισθούν από τον πληθυσμό έχει την ίδια με τα άλλα πιθανότητα επιλογής.

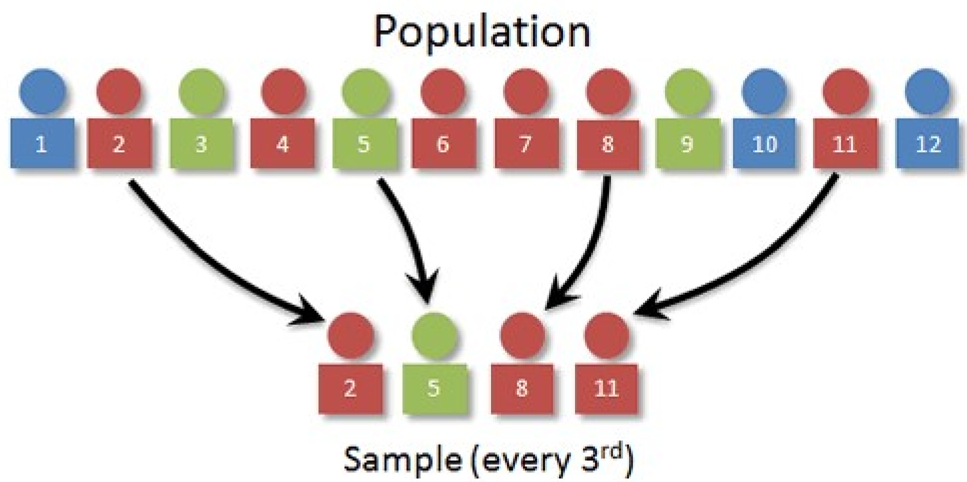


Part of a Table of Random Numbers

61424	20419	86546	00517
90222	27993	04952	66762
50349	71146	97668	86523
85676	10005	08216	25906
02429	19761	15370	43882
90519	61988	40164	15815
20631	88967	19660	89624
89990	78733	16447	27932

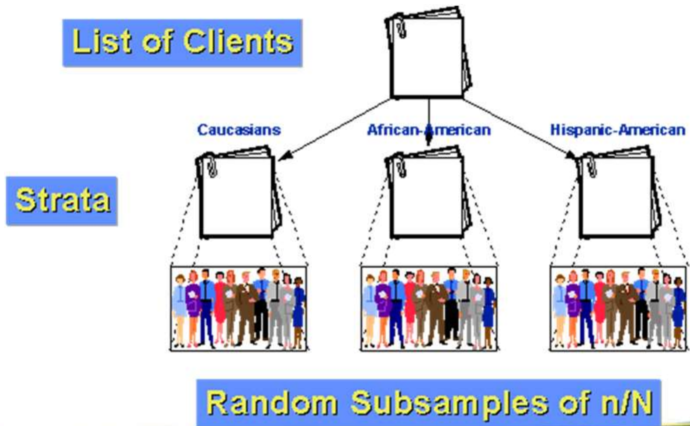
Συστηματική Δειγματοληψία

Είναι το δειγματοληπτικό σχέδιο σύμφωνα με το οποίο τα στοιχεία του δείγματος επιλέγονται από τον πληθυσμό κατά ομοιόμορφα διαστήματα διατεταγμένης τάξης.



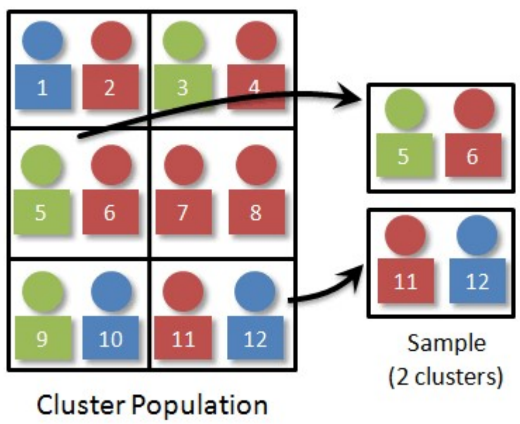
Στρωματοποιημένη Δειγματοληψία

Είναι το δειγματοληπτικό σχέδιο σύμφωνα με το οποίο τα στοιχεία του πληθυσμού διαιρούνται σε ομοιογενείς υποπληθυσμούς ή στρώματα (strata) στη βάση κάποιου σημαντικού χαρακτηριστικού και στη συνέχεια από κάθε ένα από τα στρώματα του πληθυσμού επιλέγεται ένα απλό τυχαίο δείγμα μεγέθους ni, i = 1,2,...,k ανεξάρτητα από τα άλλα. Το ενιαίο δείγμα μεγέθους n = n1 + n2 +...+ nk που προκύπτει από την ένωση των k ανεξαρτήτων απλών τυχαίων δειγμάτων ονομάζεται στρωματοποιημένο τυχαίο δείγμα (stratified sample).



Δειγματοληψία κατά Συστάδες

Είναι η δειγματοληπτική <u>τεχνική</u> η οποία διαιρεί τις στοιχειώδεις μονάδες του πληθυσμού σε ομάδες (clusters) και στη συνέχεια με απλή τυχαία δειγματοληψία επιλέγει ένα δείγμα από τις ομάδες αυτές. Όταν όλα τα στοιχεία των επιλεγμένων ομάδων περιέχονται στο τελικό δείγμα έχουμε δειγματοληψία κατά ομάδες σε ένα στάδιο (single stage cluster sampling). Αντίθετα όταν περιλάβουμε στο ενιαίο δείγμα ένα δείγμα μόνο από τις μονάδες των επιλεγμένων ομάδων έχουμε δειγματοληψία κατά ομάδες σε δύο στάδια. Η δειγματοληψία με το σχέδιο αυτό μπορεί να προχωρήσει σε τρία ή και περισσότερα στάδια (multistage cluster sampling).

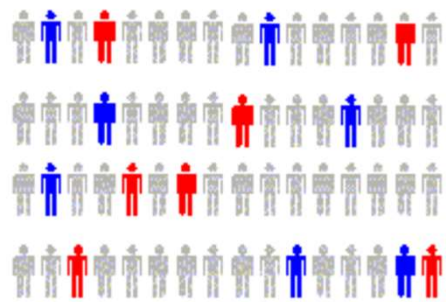


Είδη Σφαλμάτων σε δειγματοληπτικές έρευνες

- Σφάλμα αντιπροσωπ
- Σφάλμα μη απάντησης
- Σφάλμα Δειγματοληψίας
- Σφάλμα μέτρησης

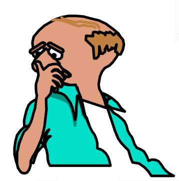
Αποκλείσθηκε από το πλαίσιο Δειγματοληψίας

Παρακολούθηση ατόμων που δεν απάντησαν .



Τυχαίες Διαφορές από δείγμα σε δείγμα.

Κακή ερώτηση!



Γραφικές μέθοδοι σύνοψης δεδομένων

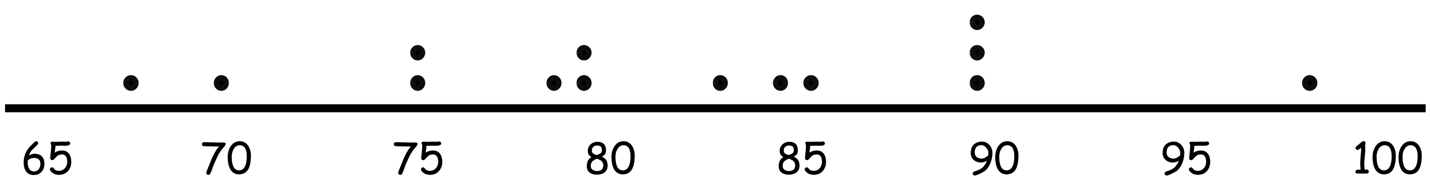
Διάγραμμα Σημείων

- Χρησιμοποιείται όταν έχουμε μικρό αριθμό παρατηρήσεων
- Πρόκειται για την τοποθέτηση όλων των διαθέσιμων τιμών πάνω σε ένα ευθύγραμμο τμήμα.

    Αν υπάρχουν περισσότερες από μια τιμές οι οποίες συμπίπτουν τοποθετούνται η μία πάνω στην άλλη

Παράδειγμα: Οι ετήσιες πωλήσεις (σε εκατ. €) 20 εταιριών που δραστηριοποιούνται στον κλάδο της πληροφορικής δίνεται στη συνέχεια

70 90 82 90 75 85 86 79 84 98 75 80 90 80 68



Διάγραμμα Μίσχου-Φύλλου

- Πρόκειται για περιγραφικό τρόπο παρουσίασης των δεδομένων με τρόπο που να υποδηλώνει την κατανομή τους.
- Κατασκευή διαγράμματος
  - Χωρίζουμε κάθε παρατήρηση σε δύο μέρη, το μίσχο και το φύλλο.
  - Κατατάσσουμε τις τιμές του μίσχου και τοποθετούμε τις τιμές των φύλλων στα δεξιά του αντίστοιχου μίσχου.
- Σημειώνεται ότι ο καθορισμός του φύλλου και μίσχου είναι αυθαίρετος.

Παράδειγμα: Το διάγραμμα μίσχου φύλλου για το προηγούμενο παράδειγμα είναι

Μίσχος	Φύλλο
6	8
7	0 5 9 5
8	2 5 6 4 0 0
9	0 0 8 0

Κατανομή Συχνότητας

(Frequency Distribution)

- Για την ταξινόμηση σε κατανομή συχνοτήτων ακολουθούμε τρία βήματα:
  - Δημιουργία Τάξεων (Classes) για τη μεταβλητή που εξετάζουμε
  - Ένταξη των δεδομένων στις τάξεις
  - Αναγραφή του πλήθους των τιμών της μεταβλητής που εμπίπτουν στην κάθε τάξη

- Για το πρώτο βήμα πρέπει να προσδιορίσουμε κάθε φορά
  - το πλήθος των τάξεων και
  - το πλάτος της κάθε τάξης.

Για να το πετύχουμε αυτό λαμβάνουμε υπόψη μας τα ακόλουθα:

- ✓ Σχεδόν ποτέ οι κατανομές έχουν λιγότερες από 6 ή περισσότερες από 16 τάξεις.
- ✓ Η κάθε τιμή της μεταβλητής εντάσσεται σε μία μόνο τάξη
- ✓ Επιδιώκουμε οι τάξεις να έχουν όλες το ίδιο πλάτος

Κατανομή Συχνότητας

(Παράδειγμα)

Η βαθμολογία (με άριστα το 100) σε ένα τεστ 150 υποψηφίων για θέσεις γραμματέων σε μια μεγάλη εταιρεία φαίνεται στον πίνακα. Να κατασκευαστεί η κατανομή συχνοτήτων των δεδομένων.

27	79	69	40	51	88	55	48	36	61
53	44	94	51	65	42	58	55	69	63
70	48	61	55	60	25	47	78	61	54
57	76	73	62	36	67	40	51	59	68
27	46	62	43	54	83	59	13	72	57
82	45	54	52	71	53	82	69	60	35
41	65	62	75	60	42	55	34	49	45
49	64	40	61	73	44	59	46	71	86
43	69	54	31	56	51	75	44	66	53
80	71	53	56	91	60	41	29	56	57
35	54	43	39	56	27	62	44	85	61
59	89	60	51	71	53	58	26	77	68
62	57	48	69	76	52	49	45	54	41
33	61	80	57	42	45	59	44	68	73
55	70	39	58	69	51	85	46	55	67

Κατανομή Συχνότητας

- Οι βαθμολογίες κυμαίνονται από 13 ως και 94, επιλέγουμε τις εννέα τάξεις 10-20, 20-30,...,90-100
- Εντάσσουμε τις τιμές της μεταβλητής (βαθμολογία υποψηφίων) στις δημιουργηθείσες τάξεις.

Κάθε **τάξη ορίζεται από δύο αριθμούς**.
Ο μικρότερος από αυτούς ονομάζεται
κατώτερο όριο και ο μεγαλύτερος **ανώτερο όριο**

Συχνότητα (frequency) =
το πλήθος των τιμών της μεταβλητής
που εντοπίζονται σε μια δεδομένη τάξη

Ανώτατο όριο=20
Κατώτατο όριο=10

ΤΑΞΕΙΣ	ΠΕΡΙΠΤΩΣΕΙΣ	ΣΥΧΝΟΤΗΤΑ
10-20	I	1
20-30	IIIII I	6
30-40	IIIII IIII	9
40-50	IIIII IIIII IIIII IIIII IIIII IIIII I	31
$\boxed{50\text{-}60}$	IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII II	42
60-70	IIIII IIIII IIIII IIIII IIIII IIIII II	32
70-80	IIIII IIIII IIIII II	17
80-90	IIIII IIIII	10
90-100	II	2
		150

Πλάτος Τάξεως (δ) =
η διαφορά μεταξύ δύο διαδοχικών
κάτω (ή πάνω) ορίων
δ=40-30=10

Οι τάξεις είναι **διαστήματα ανοικτά**
προς τα πάνω.
π.χ. [10,20)

Κεντρικός όρος= ημιάθροισμα
του κατώτερου και ανώτερου ορίου κάθε τάξης $\frac{50+60}{2} = 55$

MBA Χίος
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
διοικώ σημαίνει αποφασίζω

Κατανομή Συχνότητας

- Το ποσοστό όλων των τιμών της μεταβλητής που βρίσκονται σε μία συγκεκριμένη τάξη ονομάζεται **σχετική συχνότητα** (relative frequency).

ΤΑΞΕΙΣ	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ ΩΣ ΠΟΣΟΣΤΟ
10-20	1/150=0,007	0,007*100= 0,7
20-30	6/150=0,040	4,0
30-40	9/150=0,060	6,0
40-50	31/150=0,207	20,7
50-60	42/150=0,280	28,0
60-70	32/150=0,213	21,3
70-80	17/150=0,113	11,3
80-90	10/150=0,067	6,7
90-100	2/150=0,013	1,3
	1,000	**100,0**

Μετά την κατασκευή της κατανομής συχνοτήτων δίνεται και μία γραφική απεικόνιση της.
Η πιο δημοφιλής είναι το ιστόγραμμα (histogram)

Ιστόγραμμα

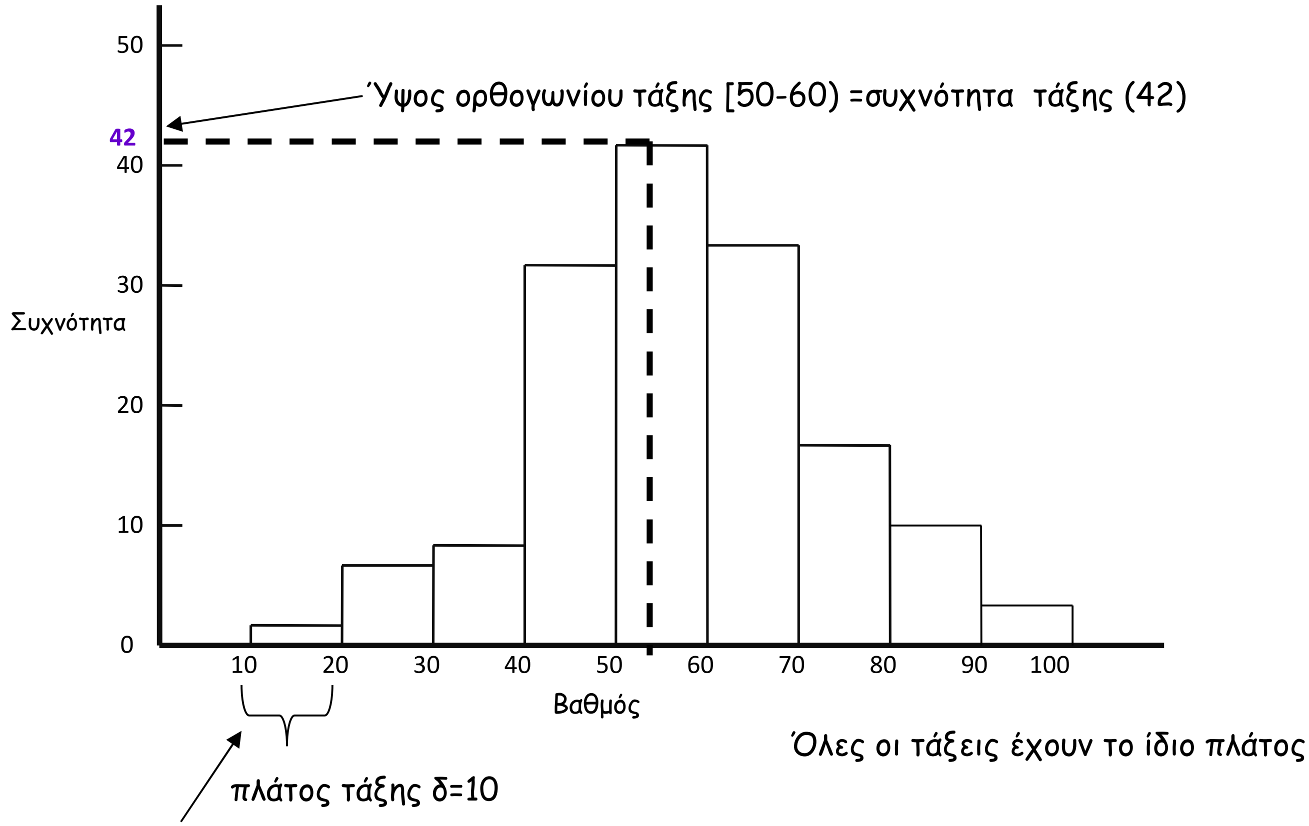
ΚΑΤΑΣΚΕΥΗ ΙΣΤΟΓΡΑΜΜΑΤΟΣ:

1. Στον οριζόντιο άξονα παίρνουμε **ίσα** διαδοχικά ευθύγραμμα τμήματα τα οποία *αντιστοιχούν* στις **τάξεις** της κατανομής.

2. Οι **συχνότητες** της κατανομής τοποθετούνται στον κατακόρυφο άξονα.

3. Στη συνέχεια κατασκευάζουμε ορθογώνια παραλληλόγραμμα με **βάσεις** τα ευθύγραμμα τμήματα που αντιστοιχούν στις **τάξεις** της κατανομής και **ύψη** ανάλογα των **συχνοτήτων** της κάθε τάξης.

Ιστόγραμμα
(Παράδειγμα συν.)



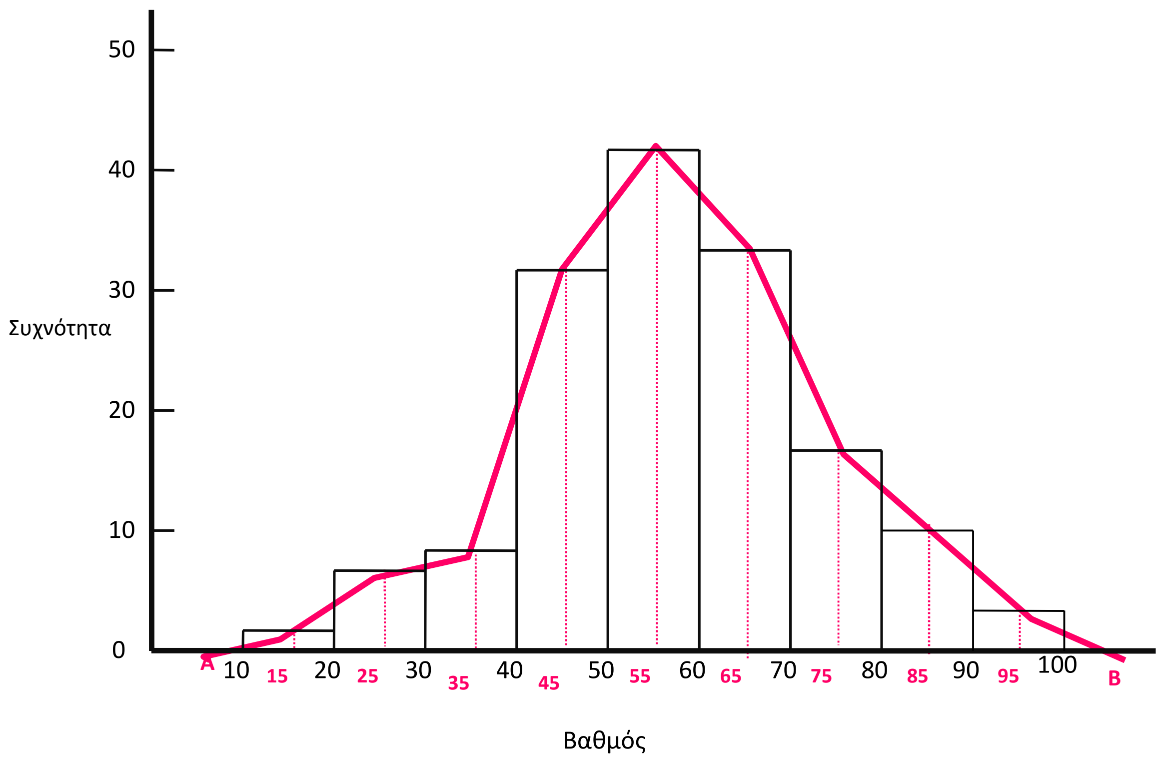
Πολύγωνο Συχνοτήτων

ΚΑΤΑΣΚΕΥΗ ΠΟΛΥΓΩΝΙΚΗΣ ΓΡΑΜΜΗΣ

- Τοποθετούμε στον οριζόντιο άξονα τους κεντρικούς όρους των τάξεων της κατανομής
- Υψώνουμε στα σημεία αυτά κάθετα ευθύγραμμα τμήματα ίσα προς τις αντίστοιχες συχνότητες των τάξεων.
- Ενώνουμε με ευθύγραμμα τμήματα τα διαδοχικά σημεία που έχουν προκύψει.
- Η πολυγωνική γραμμή είναι μία κλειστή τεθλασμένη

Για να το πετύχουμε αυτό θεωρούμε <u>δύο υποθετικές τάξεις</u> με <u>συχνότητα μηδέν</u> στα δύο άκρα της κατανομής και συνδέουμε τα άκρα Α και Β της τεθλασμένης που έχει ήδη προκύψει με τους κεντρικούς όρους των δύο αυτών υποθετικών τάξεων.

Πολυγωνική Γραμμή



Πίνακας Διπλής Εισόδου

Υπάρχουν περιπτώσεις που οι συχνότητες αναφέρονται σε *ζευγάρια μεταβλητών*. Ο πίνακας με τον οποίο επιτυχαίνομε τη σύνοψη των δεδομένων μας ονομάζεται **Πίνακας Διπλής Εισόδου**.

Παράδειγμα: Έστω ότι διαθέτουμε 149 ζευγάρια τιμών που αφορούν τις υπερωρίες και τις αντίστοιχες ωριαίες αποζημιώσεις των εργαζομένων μιας πολυεθνικής

Πίνακας Διπλής Εισόδου

Έστω ότι διαθέτουμε 149 ζευγάρια τιμών που αφορούν τις υπερωρίες και τις αντίστοιχες ωριαίες αποζημιώσεις των εργαζομένων μιας πολυεθνικής

Υπερωρίες	Αποζημίωση	Υπερωρίες	Αποζημίωση	Υπερωρίες	Αποζημίωση	Υπερωρίες	Αποζημίωση	Υπερωρίες	Αποζημίωση
174	50	162	60	169	57	162	66	166	55
161	47	165	58	165	62	156	58	164	60
157	58	168	65	169	51	153	59	170	53
173	60	162	54	162	55	158	55	159	53
169	63	170	64	164	56	156	61	158	53
171	63	172	60	163	49	165	61	151	49
174	61	160	50	164	61	173	57	166	67
161	69	159	52	170	64	165	54	171	68
164	55	168	54	160	51	168	73	166	57
166	57	164	63	164	56	169	55	168	65
157	71	159	56	159	53	161	69	167	59
153	43	152	49	158	57	164	55	162	49
158	52	164	61	161	47	168	70	158	55
170	65	170	53	158	57	169	70	169	62
166	59	160	52	161	51	168	64	169	54
165	62	175	69	154	53	160	52	155	45
156	60	172	69	168	71	168	59	178	56
169	59	169	65	165	53	171	64	167	72
162	52	163	62	167	62	157	54	151	44
163	62	167	57	166	54	163	68	164	53
160	50	171	58	166	69	166	55	152	56
154	60	167	65	172	65	159	56	159	49
159	51	160	62	157	46	164	60	161	59
176	72	160	61	163	70	163	65	159	58
159	50	158	58	165	53	165	50	159	55
165	71	173	63	160	51	160	58	160	50
164	56	170	64	172	66	166	63	157	57
161	50	167	61	165	56	159	51	163	59
164	63	167	65	160	49	155	56	159	70
169	63	159	52	167	60	164	70		

Πίνακας Διπλής Εισόδου

Υπερωρίες Αποζημιώσεις	150-154	154-158	158-162	162-166	166-170	170-174	174-178	ΣΥΝΟΛΟ
39-47	2	2	-	-	-	-	-	4
47-55	2	6	24	8	4	3	-	47
55-63	2	8	9	23	12	5	1	60
63-71	-	-	3	8	13	8	1	33
71-79	-	1	-	1	2	-	1	5
ΣΥΝΟΛΟ	6	17	36	40	31	16	3	149

Από τα σύνολα των συχνοτήτων της τελευταίας **γραμμής** προκύπτουν οι αντίστοιχες κατανομές συχνοτήτων για τις υπερωρίες των 149 εργαζομένων.

Από τα σύνολα των συχνοτήτων της τελευταίας **στήλης** προκύπτουν οι αντίστοιχες κατανομές συχνοτήτων για τις ωριαίες αποζημιώσεις των 149 εργαζομένων.

ΤΑΞΕΙΣ	ΣΥΧΝΟΤΗΤΑ
150-154	6
154-158	17
158-162	36
162-166	40
166-170	31
170-174	16
174-178	3
	149

ΤΑΞΕΙΣ	ΣΥΧΝΟΤΗΤΑ
39-47	4
47-55	47
55-63	60
63-71	33
71-79	5
	149

Αθροιστική Κατανομή Συχνοτήτων

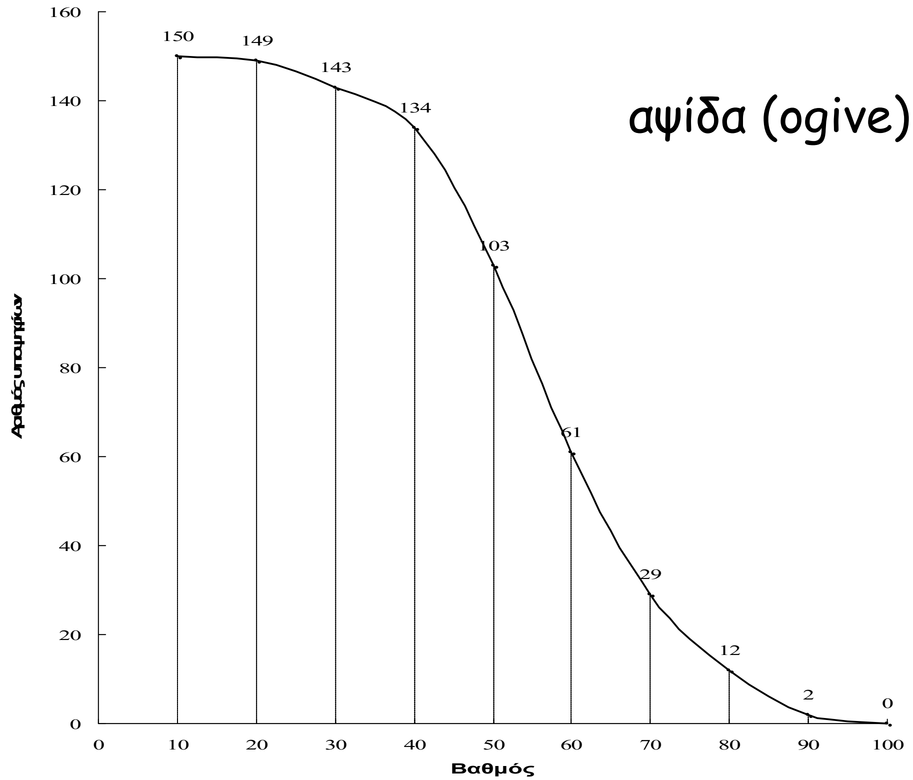
Τα δεδομένα πολλές φορές παρουσιάζονται **με αθροιστική κατανομή συχνοτήτων (cumulative frequency distribution)**

ΤΑΞΕΙΣ	ΣΥΧΝΟΤΗΤΑ	ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ	ΣΧΕΤΙΚΗ ΑΘΡΟΙΣΤΙΚΗ ΣΥΧΝΟΤΗΤΑ ΩΣ %
10-20	1	*1*	*1/150=0,007*	*0,7*
20-30	6	1+6=7	7/150=0,047	4,7
30-40	9	7+9=*16*	16/150=0,106	10,6
40-50	31	16+31=*47*	47/150=0,313	31,3
50-60	42	47+42=*89*	89/150=0,593	59,3
60-70	32	89+32=*121*	121/150=0,806	80,6
70-80	17	121+17=*138*	138/150=0,92	92
80-90	10	138+10=*148*	148/150=0,986	98,6
90-100	2	148+2=*150*	150/150=1	100
	150			

Αριστερόστροφη Αθροιστική Κατανομή Συχνοτήτων

Αν ενδιαφερόμαστε
για το πόσες τιμές της
μεταβλητής είναι
μεγαλύτερες
από κάποια τιμή
Χρησιμοποιούμε
**Αριστερόστροφη
αθροιστική κατανομή**

ΑΡΙΣΤΕΡΟΣΤΡΟΦΗ ΑΘΡΟΙΣΤΙΚΗ ΚΑΤΑΝΟΜΗ		
ΒΑΘΜΟΛΟΓΙΑ	ΥΠΟΨΗΦΙΟΙ	
	ΑΡΙΘΜΟΣ	%
10 ΚΑΙ ΠΑΝΩ	150	100,00
20 ΚΑΙ ΠΑΝΩ	149	99,53
30 ΚΑΙ ΠΑΝΩ	143	95,33
40 ΚΑΙ ΠΑΝΩ	134	89,33
50 ΚΑΙ ΠΑΝΩ	103	68,66
60 ΚΑΙ ΠΑΝΩ	61	40,66
70 ΚΑΙ ΠΑΝΩ	29	19,33
80 ΚΑΙ ΠΑΝΩ	12	8,00
90 ΚΑΙ ΠΑΝΩ	2	1,33

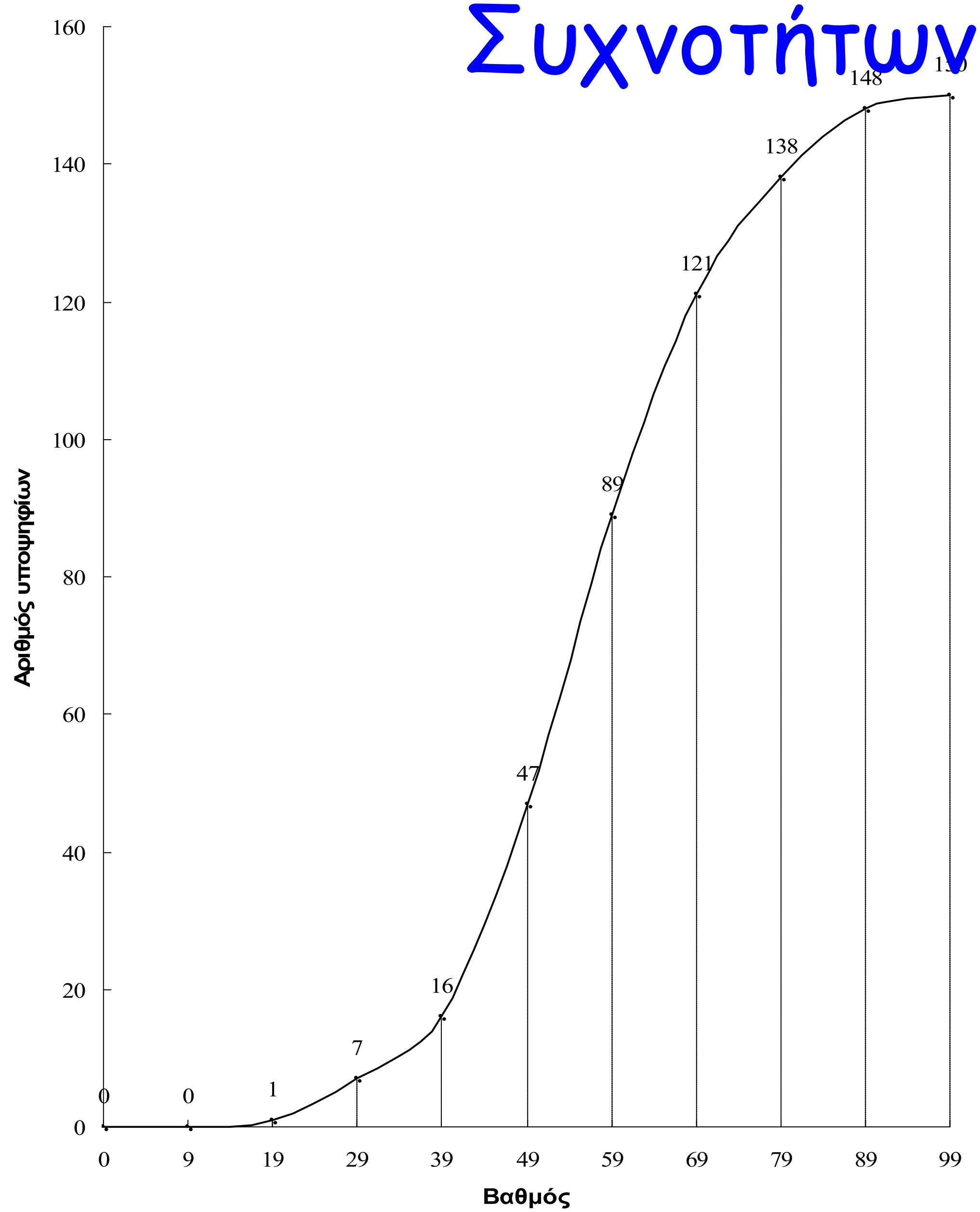


αψίδα (ogive)

Δεξιόστροφη Αθροιστική Κατανομή Συχνοτήτων

Αν ενδιαφερόμαστε
για το πόσες τιμές της
μεταβλητής είναι
μικρότερες
από κάποια τιμή
χρησιμοποιούμε
δεξιόστροφη
αθροιστική κατανομή

ΔΕΞΙΟΣΤΡΟΦΗ ΑΘΡΟΙΣΤΙΚΗ ΚΑΤΑΝΟΜΗ		
ΒΑΘΜΟΛΟΓΙΑ	ΥΠΟΨΗΦΙΟΙ	
	ΑΡΙΘΜΟΣ	%
ΚΑΤΩ ΤΟΥ 20	1	0,67
ΚΑΤΩ ΤΟΥ 30	7	4,67
ΚΑΤΩ ΤΟΥ 40	16	10,67
ΚΑΤΩ ΤΟΥ 50	47	31,33
ΚΑΤΩ ΤΟΥ 60	89	59,33
ΚΑΤΩ ΤΟΥ 70	121	80,67
ΚΑΤΩ ΤΟΥ 80	138	92,00
ΚΑΤΩ ΤΟΥ 90	148	98,67
ΚΑΤΩ ΤΟΥ 100	150	100,00



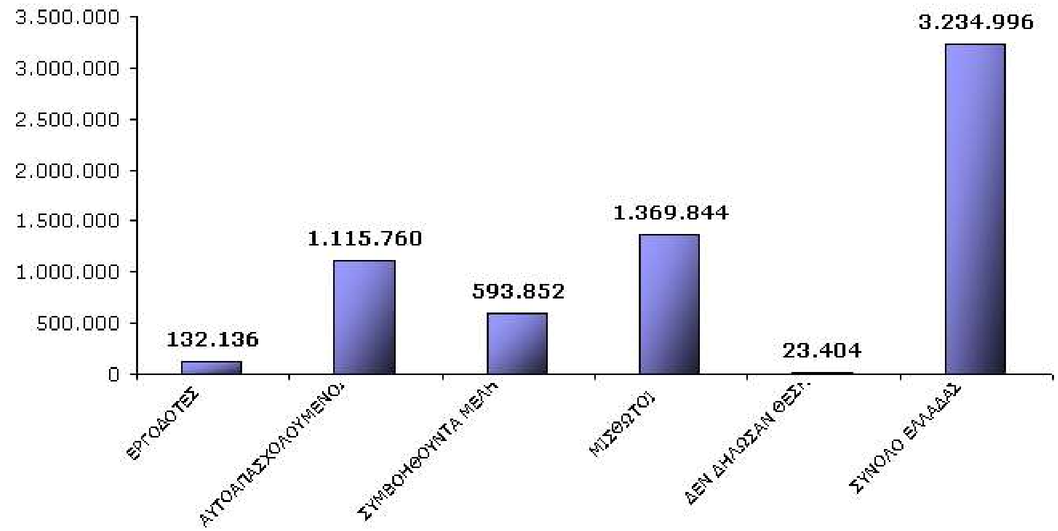
Ραβδόγραμμα-Bar Chart

Το ραβδόγραμμα απεικονίζει τη συχνότητα των κατηγοριών στις οποίες είναι ταξινομημένα τα δεδομένα ως ένα ορθογώνιο κάθετο στον οριζόντιο ή τον κατακόρυφο άξονα.

Όλα τα ορθογώνια έχουν το ίδιο (αυθαίρετο πλάτος) ενώ το μήκος κάθε ορθογωνίου είναι ανάλογο της συχνότητας της αντίστοιχης κατηγορίας δεδομένων.

Παράδειγμα

ΘΕΣΗ ΣΤΟ ΕΠΑΓΓΕΛΜΑ	ΣΥΝΟΛΟ
ΕΡΓΟΔΟΤΕΣ	132.136
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΙ	1.115.760
ΣΥΜΒΟΗΘΟΥΝΤΑ ΜΕΛΗ	593.852
ΜΙΣΘΩΤΟΙ	1.369.844
ΔΕΝ ΔΗΛΩΣΑΝ ΘΕΣΗ	23.404
ΣΥΝΟΛΟ ΕΛΛΑΔΑΣ	3.234.996

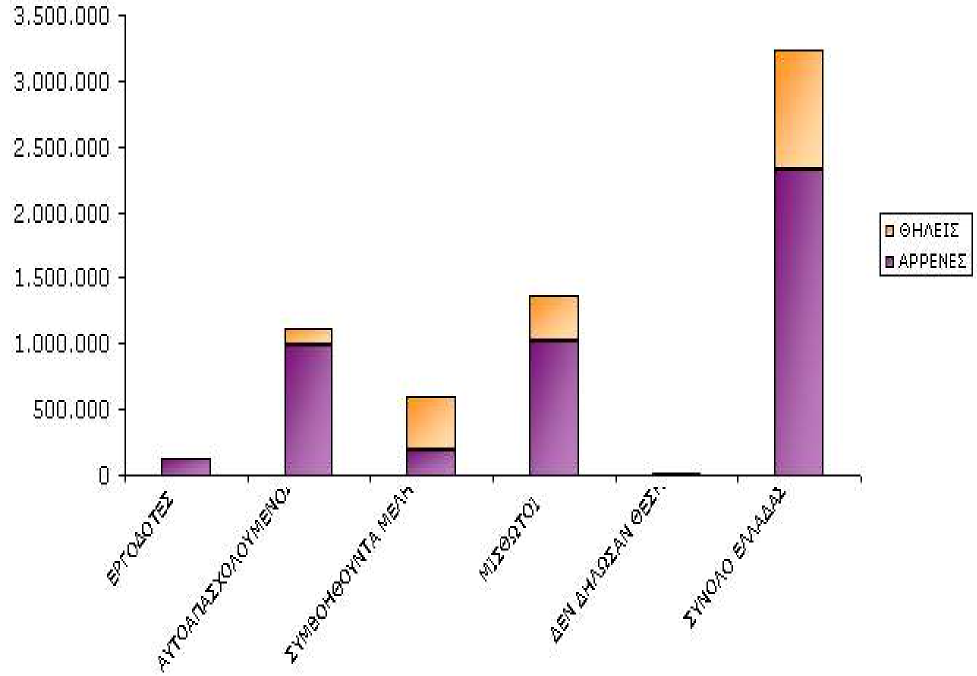


Ραβδόγραμμα Συνιστωσών

Στο ραβδόγραμμα συνιστωσών **κάθε** ορθογώνιο χωρίζεται σε τμήματα (συνιστώσες) που το καθένα αντιστοιχεί σε μία υποδιαίρεση της κατηγορίας που αντιπροσωπεύει το ορθογώνιο

Το **μήκος** κάθε συνιστώσας είναι ανάλογο της συχνότητας (ή σχετικής συχνότητας) της αντίστοιχης υποδιαίρεσης.

ΘΕΣΗ ΣΤΟ ΕΠΑΓΓΕΛΜΑ	ΣΥΝΟΛΟ	ΑΡΡΕΝΕΣ	ΘΗΛΕΙΣ
ΕΡΓΟΔΟΤΕΣ	132.136	120.832	11.304
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΙ	1.115.760	982.280	133.480
ΣΥΜΒΟΗΘΟΥΝΤΑ ΜΕΛΗ	593.852	183.696	410.156
ΜΙΣΘΩΤΟΙ	1.369.844	1.026.236	343.608
ΔΕΝ ΔΗΛΩΣΑΝ ΘΕΣΗ	23.404	16.544	6.860
ΣΥΝΟΛΟ ΕΛΛΑΔΑΣ	3.234.996	2.329.588	905.408



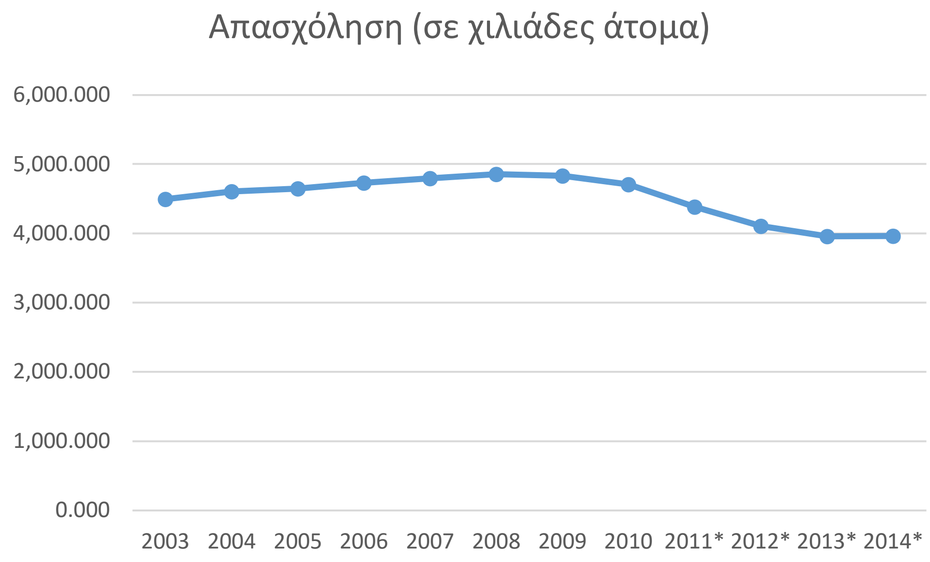
Γραμμογράφημα-Line Chart

- Τα γραμμογραφήματα χρησιμοποιούνται κυρίως όταν τα δεδομένα μας είναι χρονολογικές σειρές
- Ο οριζόντιος άξονας εκφράζει τις χρονικές στιγμές στις οποίες έχουμε κατατάξει τα δεδομένα και ο κατακόρυφος τις αντίστοιχες συχνότητες
- Για να κατασκευάσουμε ένα γραμμογράφημα προσδιορίζουμε τις συχνότητες για κάθε κατηγορία (χρονική στιγμή) και τις ενώνουμε με ευθύγραμμα τμήματα.

Γραμμογράφημα
(Παράδειγμα)

Παράδειγμα: Η εξέλιξη της απασχόλησης για τα έτη 2003-2014

Έτος	Απασχόληση (σε χιλιάδες άτομα)
2003	4.495,681
2004	4.604,092
2005	4.646,871
2006	4.731,335
2007	4.795,069
2008	4.856,363
2009	4.829,003
2010	4.705,483
2011*	4.381,819
2012*	4.105,223
2013*	3.957,225
2014*	3.962,604



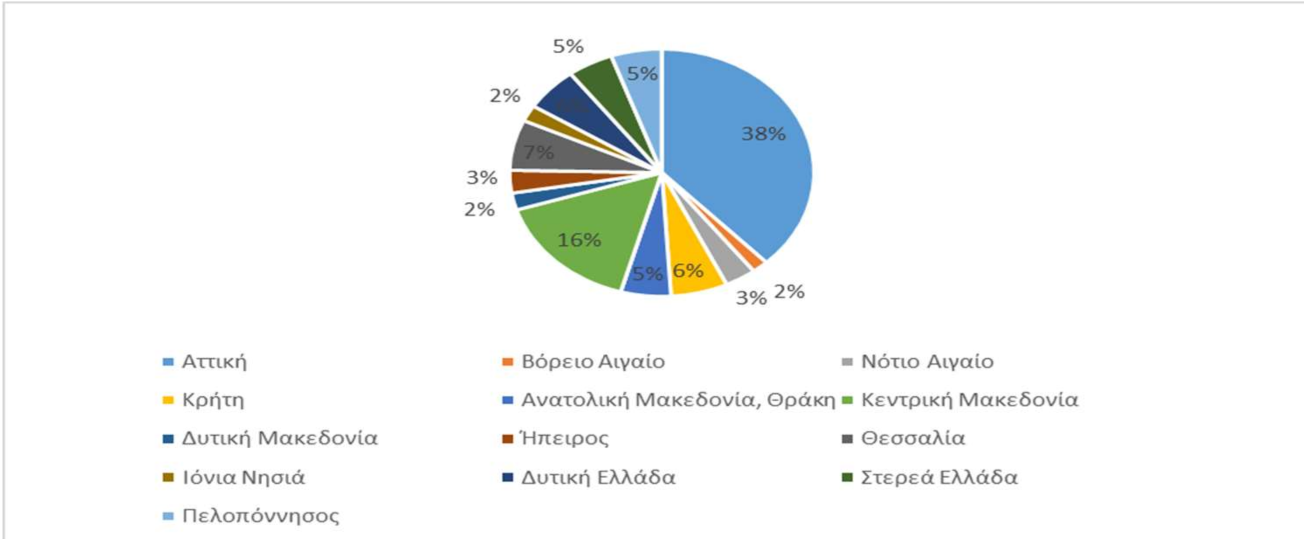
Κυκλικό Διάγραμμα

- Τα κυκλικά διαγράμματα χρησιμοποιούνται για τη γραφική απεικόνιση δεδομένων που αποτελούν υποδιαιρέσεις ενός συνόλου.
  - Αν οι υποδιαιρέσεις εκφράζονται σε ποσοστά το διάγραμμα αναφέρεται **ως κυκλικό διάγραμμα ποσοστών (Percentage pie chart)**.

- Για να κατασκευάσουμε ένα κυκλικό διάγραμμα
  - χαράσσουμε καταρχήν ένα κύκλο με αυθαίρετη ακτίνα
  - Τον χωρίζουμε σε κυκλικούς τομείς οι οποίοι αντιστοιχούν στις διάφορες κατηγορίες των παρατηρήσεων και το εμβαδόν τους είναι ανάλογο με τη συχνότητα της κάθε κατηγορίας.

Κυκλικό Διάγραμμα

Παράδειγμα: Η απασχόληση στις περιφέρειες της χώρας για το 2013



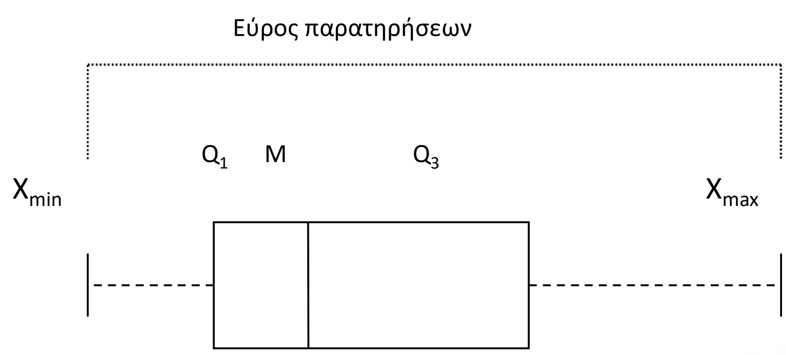
Αττική	1.501,815
Βόρειο Αιγαίο	68,514
Νότιο Αιγαίο	131,336
Κρήτη	238,637
Ανατολική Μακεδονία, Θράκη	210,662
Κεντρική Μακεδονία	624,234
Δυτική Μακεδονία	88,172
Ήπειρος	116,972
Θεσσαλία	261,806
Ιόνια Νησιά	83,660
Δυτική Ελλάδα	226,638
Στερεά Ελλάδα	191,402
Πελοπόννησος	213,377
ΕΛΛΑΔΑ	**3.957,225**

Διαγράμματα Πλαισίου Απολήξεων

- Απεικονίζει τις τιμές της διαμέσου και των τεταρτημορίων (μέτρα που δεν επηρεάζονται από ακραίες τιμές των δεδομένων).

- Συνδυάζει αυτά τα μέτρα με τις πληροφορίες που περιέχονται στις ακραίες τιμές των δεδομένων δίνοντας έτσι μια πληρέστερη εικόνα της κατανομής τους.

- Το Διάγραμμα Πλαισίου – Απολήξεων χρησιμοποιεί πέντε χαρακτηριστικές τιμές των δεδομένων:
  - X_{min}: Ελάχιστη τιμή των δεδομένων
  - Q_1: Πρώτο τεταρτημόριο
  - M: Διάμεσος
  - Q_3: Τρίτο τεταρτημόριο
  - X_{max} : Μέγιστη τιμή των δεδομένων



Some References (for your interest)

- Duckworth, F. (2001). A role for statistics in international cricket. *Teaching Statistics* **23**(2), 38–44.
- Matthews, R. (2000). Storks deliver babies(*p*=0.008).*Teaching Statistics* **22** (2), 36–8.
- Cai, J. (1998), "Exploring Students' Conceptual Understanding of the Averaging Algorithm," *School Science and*
- *Mathematics, 98, 93-98.*
- Cooper, L. (2002), An Assessment of Prospective Secondary Mathematics Teachers' Preparedness to Teach Statistics,"
- *Dissertation Abstracts International, 64 (01), 89A. (University Microfilms No. 3078386).*
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007), *Guidelines for*
- *Assessment and Instruction in Statistics Education (GAISE) Report. Alexandria, VA: American Statistical Association.*
- Friel, S. (1998), "Teaching Statistics: What's Average?" In L. J. Morrow (Ed.) *The Teaching and Learning of*
- *Algorithms in School Mathematics (pp. 208-217). Reston, VA: National Council of Teachers of Mathematics.*
- Friel, S. & Bright, G. (1995), "Graph Knowledge: Understanding How Students Interpret Data Using Graphs,"
- Columbus, OH: Paper presented at the Annual Meeting of the North American Chapter of the International Group for
- the Psychology of Mathematics Education.
- Groth, R. & Bergner, J. (2006), "Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean,
- Median, and Mode," *Mathematical Thinking and Learning, 8(1), 37-63.*

MBA Χίος
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
διοικώ σημαίνει αποφασίζω

Που απευθύνομαι?????

Ελένη Γάκη,
Τηλ. 2271035161
Email: e.gaki@aegean.gr
Υλικό μαθήματος
https://eclass.aegean.gr/