

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΧΡΟΝΟΣΗΜΑΣΜΕΝΩΝ, ΑΚΟΛΟΥΘΙΑΚΩΝ, ΣΥΝΘΕΤΩΝ ΤΥΠΩΝ ΔΕΔΟΜΕΝΩΝ

Πανεπιστήμιο Πελοποννήσου

Δομή παρουσίασης

- **Εισαγωγή**
- Βασικές Έννοιες
- Σχετικές μελέτες
- Εφαρμογή Δεδομένων
- Συμπεράσματα

Εισαγωγή

- Μελέτη και προσαρμογή των διάφορων τεχνικών δειγματοληψίας σε σύνθετους τύπους δεδομένων.
- Χρονοσημασμένα, ακολουθιακά δεδομένα.
 - Χρονοσειρές (Timeseries)
 - Τροχιές (Trajectories)
 - Κλικ στον Ιστό (Webclicks)
- Σκοπός: ανάλυση δεδομένων και εξόρυξη γνώσης

Δομή παρουσίασης

- Εισαγωγή
- **Βασικές Έννοιες**
- Σχετικές μελέτες
- Εφαρμογή Δεδομένων
- Συμπεράσματα

Εισαγωγή στη Δειγματοληψία

- Διαδικασία λήψης ενός τμήματος στοιχείων από κάποιο ευρύτερο σύνολο.
- Σκοπός: όσο το δυνατόν πιο ακριβής προσδιορισμός των ιδιοτήτων του πληθυσμού, μελετώντας τα στοιχεία του δείγματος.
- Για την ορθότητα: Τα δεδομένα πρέπει να προέρχονται από το σωστό σύνολο, να μαζευτούν προσεγμένα και να αντιπροσωπεύουν τον πληθυσμό από τον οποίο επιλέχθηκαν.
- Παραδείγματα: πληροφόρηση από ανθρώπους (π.χ. προεκλογική σφυγμομέτρηση), πορείες κινούμενων αντικειμένων στο χωρόχρονο, πλοήγηση χρηστών σε ιστοσελίδες του διαδικτύου κτλ.

Μέθοδοι Δειγματοληψίας (1)

- Δειγματοληψία Πιθανοτήτων: οι παρατηρήσεις του δείγματος επιλέγονται ανεξάρτητα και με ίσες πιθανότητες
 - Απλή Τυχαία ή Απεριόριστη Δειγματοληψία (Simple Random Sampling)
 - Συστηματική Δειγματοληψία (Systematic Random Sampling)
 - Δειγματοληψία με διαστρωμάτωση (Stratified Random Sampling)
 - Δειγματοληψία κατά συστάδες (Cluster Sampling)
 - Πολυσταδιακή Δειγματοληψία (Multistage Random Sampling)

Μέθοδοι Δειγματοληψίας (2)

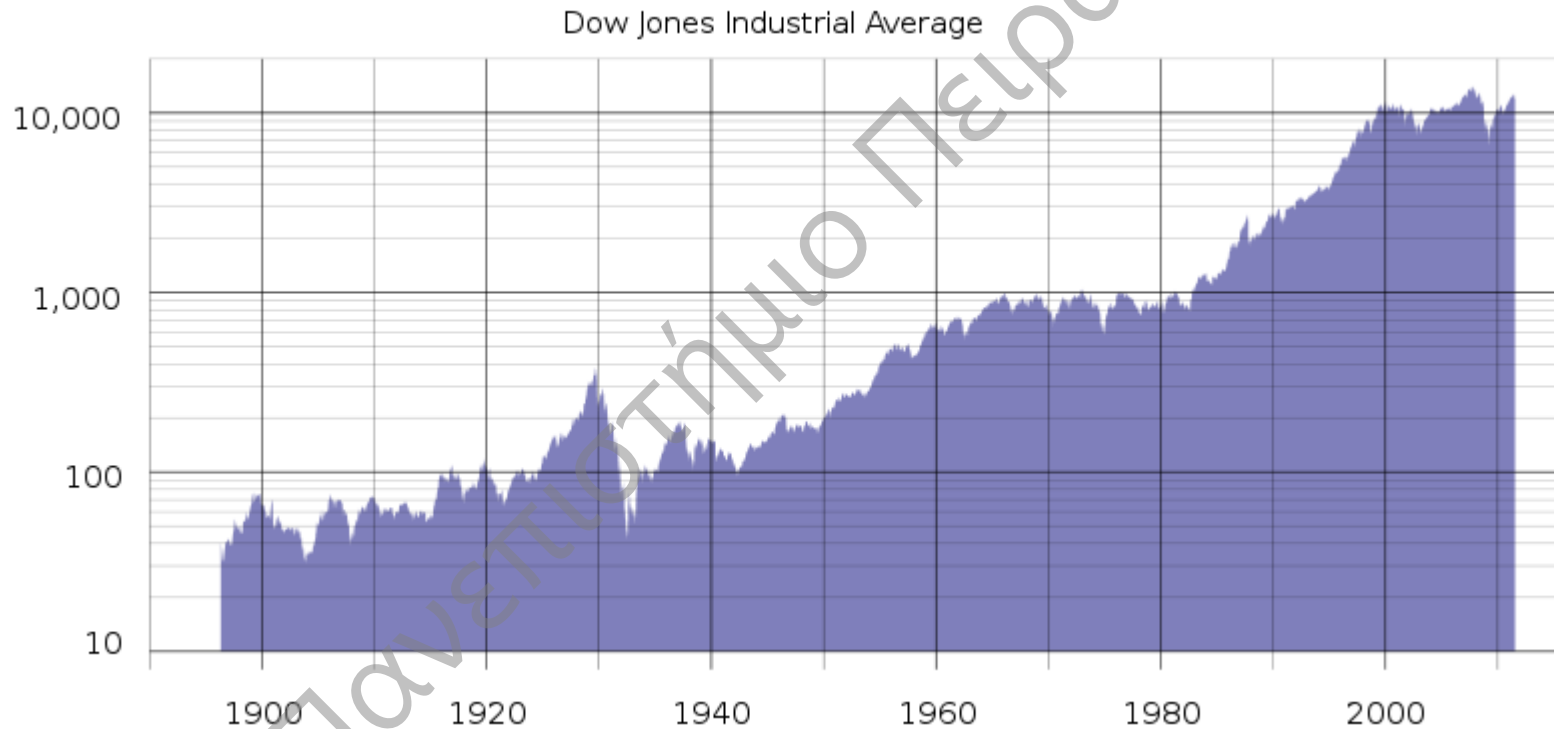
- Δειγματοληψία χωρίς Πιθανότητες: το δείγμα γίνεται με ένα σταθερό και προκαθορισμένο (συστηματικό) τρόπο.
 - Δειγματοληψία ευκολίας ή συμπτωματική δειγματοληψία
 - Δειγματοληψία αναλογίας ή ποσοστιαία δειγματοληψία

Πανεπιστήμιο Πειραιώς

Χρονοσειρές

- Ακολουθίες από σημεία δεδομένων.
- Τα δεδομένα έχουν μια φυσική χρονική διάταξη
- Μετρούνται συνήθως σε διαδοχικές χρονικές στιγμές που ισαπέχουν μεταξύ τους.
- Για την πρόβλεψη μελλοντικών τιμών και καλύτερη κατανόηση του μηχανισμού δημιουργίας των δεδομένων.
- Παραδείγματα: Τιμές κλεισίματος μετοχών.

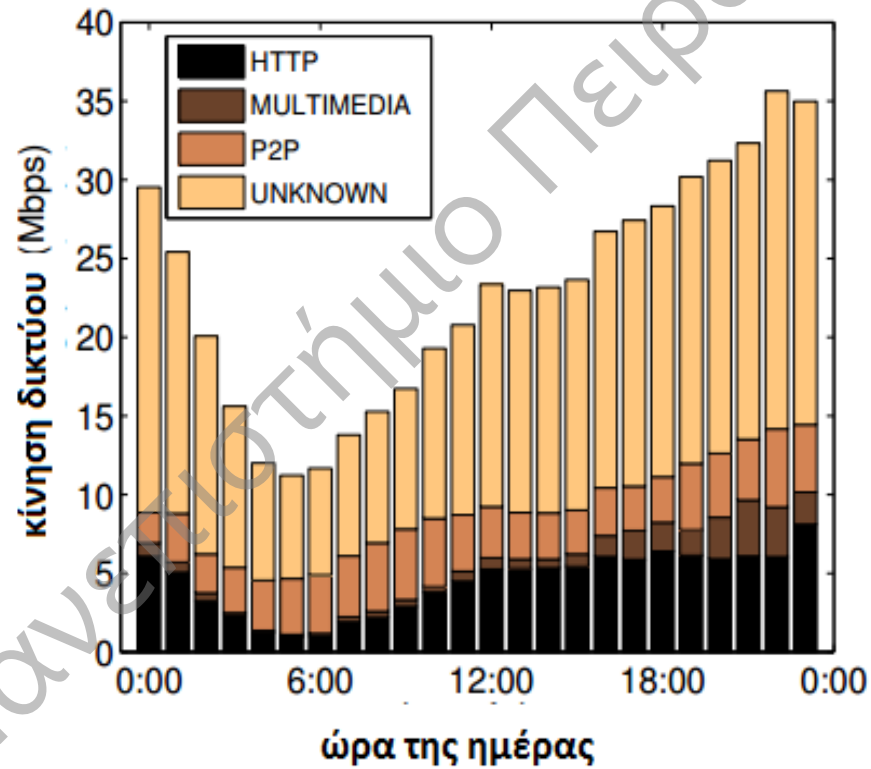
Παράδειγμα Χρονοσειρών



Τροχιές

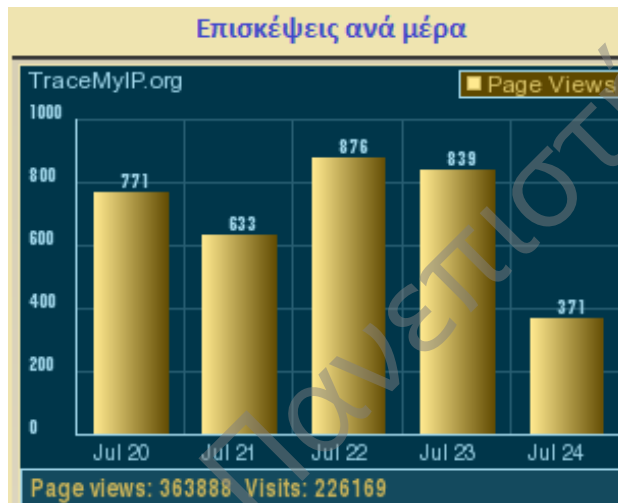
- Διαδρομή που ακολουθεί ένα κινούμενο αντικείμενο μέσα στο χώρο ως μία συνάρτηση του χρόνου
- «αντικείμενα» που η κίνηση τους μπορεί να εντοπιστεί μέσω GPS ή έχουν πάνω τους ένα μηχανισμό GPS μέσω του οποίου καταγράφεται η κίνηση τους.
- «Βάσεις Δεδομένων Τροχιάς» (TD- Trajectory Databases).
- Παραδείγματα: κίνηση ενός χρήστη μέσα στο διαδίκτυο

Παράδειγμα Τροχιών



Webclicks

- Καταγραφή του πλήθους των κλικ που γίνονται σε κάποια συγκεκριμένη ιστοσελίδα από τους διάφορους χρήστες.
- Ανάλυση των κλικ: μέτρηση, συλλογή, ανάλυση και καταγραφή των δεδομένων στον Παγκόσμιο Ιστό για τους σκοπούς της κατανόησης και βελτιστοποίηση της χρήσης web.
- Σε πραγματικό χρόνο ή σε μη πραγματικό χρόνο.

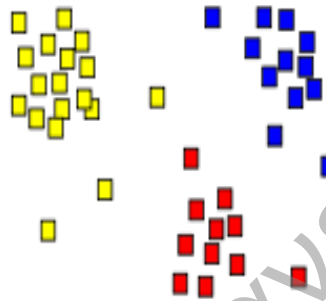


Παράδειγμα Webclicks:

Διάγραμμα που παρουσιάζει αποτελέσματα από την καταμέτρηση των κλικ ανά ημέρα

Ομαδοποίηση (Clustering)

- Τεχνική για την εξόρυξη δεδομένων.
- Που χρησιμοποιείται για να τοποθετήσει τα στοιχεία δεδομένων σε σχετικές ομάδες (clusters).
- Τα αντικείμενα στην ίδια ομάδα είναι πιο παρόμοια μεταξύ τους παρά με άλλα αντικείμενα σε άλλες ομάδες.
- Εφαρμόζεται στα δεδομένα αξιολογώντας τις τεχνικές δειγματοληψίας.



Παράδειγμα Ομαδοποίησης:

Το αποτέλεσμα της ανάλυσης ομαδοποίησης (clustering) εμφανίζεται μέσω του χρωματισμού των τετραγώνων σε τρεις συστάδες.

Δομή παρουσίασης

- Εισαγωγή
- Βασικές Έννοιες
- **Σχετικές μελέτες**
- Εφαρμογή Δεδομένων
- Συμπεράσματα

Σχετικές Μελέτες (1)

- Μελέτες βασισμένες στις μεθόδους δειγματοληψίας τροχιών
- Τμηματοποίηση και δειγματοληψία της τροχιάς ενός κινούμενου αντικειμένου:
 - Η καταγραφή και παρακολούθηση της τροχιάς ενός αντικειμένου χρησιμοποιείται σε καθημερινή βάση, με την χρήση του GPS (Global Positioning System).
 - Μέθοδος τμηματοποίησης και δειγματοληψίας της τροχιάς βασισμένη στην αντιπροσωπευτικότητα των υπό-τροχιών.
 - Σκοπός: εύρεση των πιο αντιπροσωπευτικών τροχιών του κινούμενου αντικειμένου έτσι ώστε να μπορέσει να γίνει κατανόηση, έρευνα, ανάλυση και περιήγηση στο χωροχρονικό περιεχόμενο.

Σχετικές Μελέτες (2)

- Δειγματοληψία τροχιάς για άμεση παρατήρηση της κυκλοφορίας:
 - Μέτρηση της κυκλοφορίας.
 - Σκοπός: έλεγχος και την μηχανική λειτουργία.
 - Παράδειγμα: τα μονοπάτια που ακολουθούνται από πακέτα μεταξύ εισόδου και εξόδου της διεύθυνσης του Διαδικτύου.

- Δειγματοληψία τροχιάς χωρίς επίβλεψη:
 - Ανακάλυψη προτύπων συμπεριφοράς κινούμενων αντικειμένων.
 - Δειγματοληψία τροχιών - προσέγγιση που αποτελείται από τρία βασικά βήματα.
 1. Συμβολική αναπαράσταση των τροχιών.
 2. Μέθοδος αναπαράστασης κάθε τροχιάς μέσω μίας συνεχόμενης συνάρτησης που περιγράφει την αντιπροσωπευτικότητα του κάθε συστατικού της.
 3. Αυτόματη μέθοδος για τη δειγματοληψία της τροχιάς (T-Sampling)

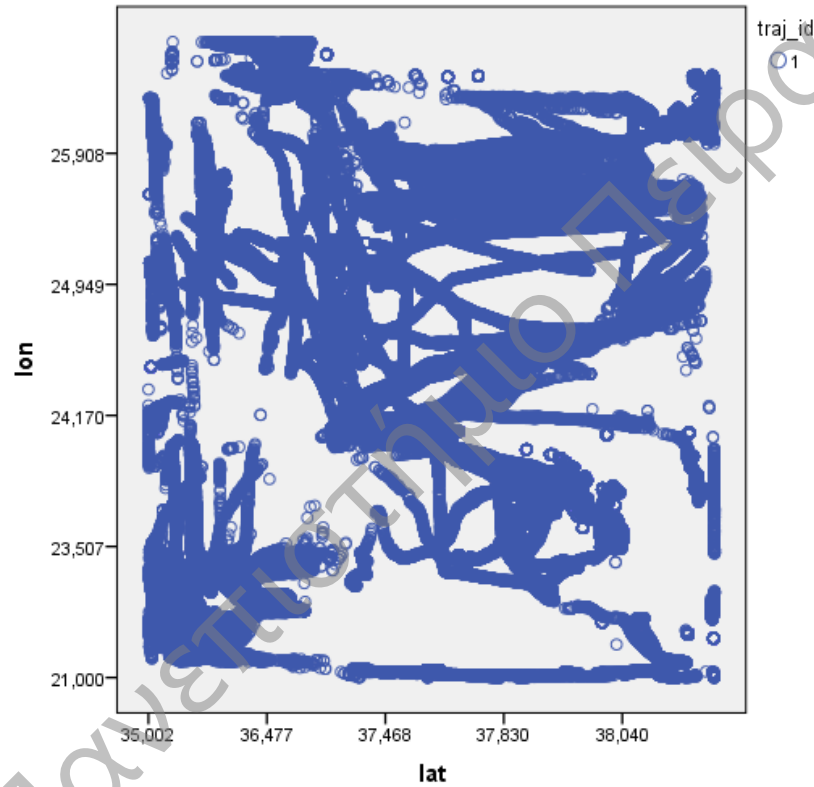
Δομή παρουσίασης

- Εισαγωγή
- Βασικές Έννοιες
- Σχετικές μελέτες
- **Εφαρμογή Δεδομένων**
- Συμπεράσματα

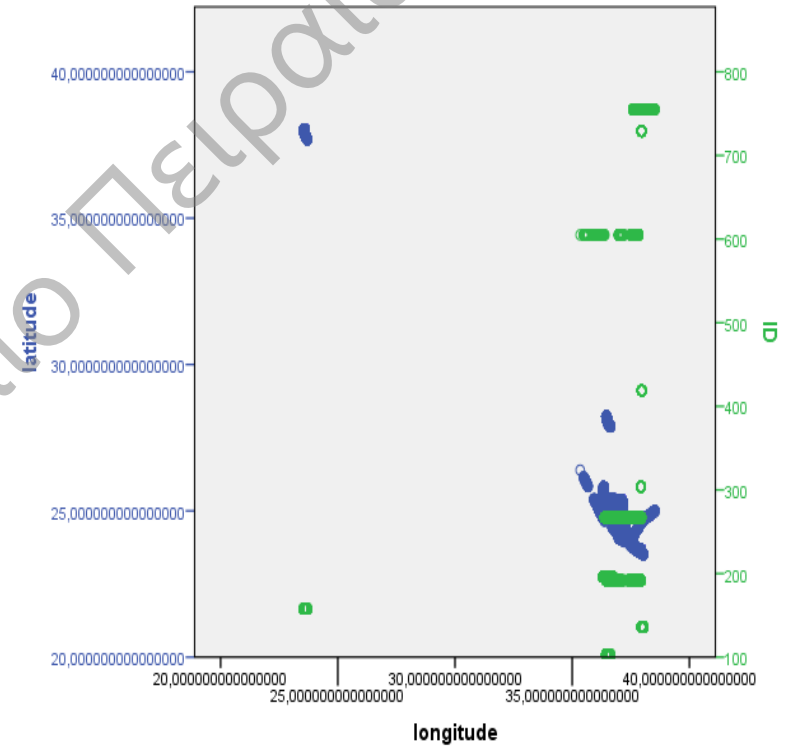
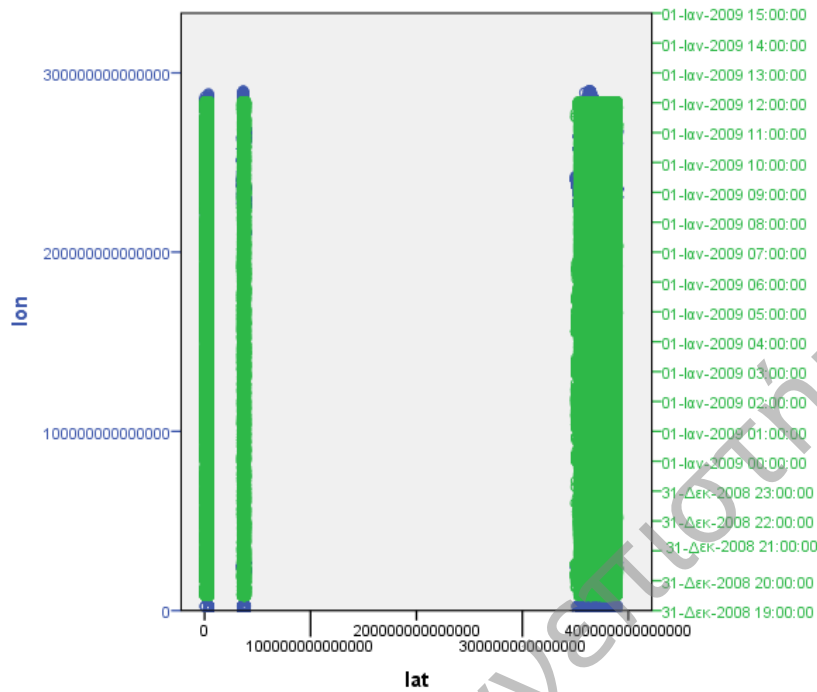
Τροχιές Πλοίων

- Δυσδιάστατα Timeseries.
- Τα δεδομένα έχουν συλλεχθεί από την IMIS Hellas S.A.» και έχει δωριθεί στο «Infolab» για σκοπούς ερευνών
- Κάθε εγγραφή στο σύνολο δεδομένων έχει τη μορφή «obj_id, traj_id, t, lon, lat».
- Αναπαράσταση της διαδρομής των σκαφών για τον βέλτιστο και ασφαλή σχεδιασμό της διαδρομής των πλοίων μέσα στο Αιγαίο

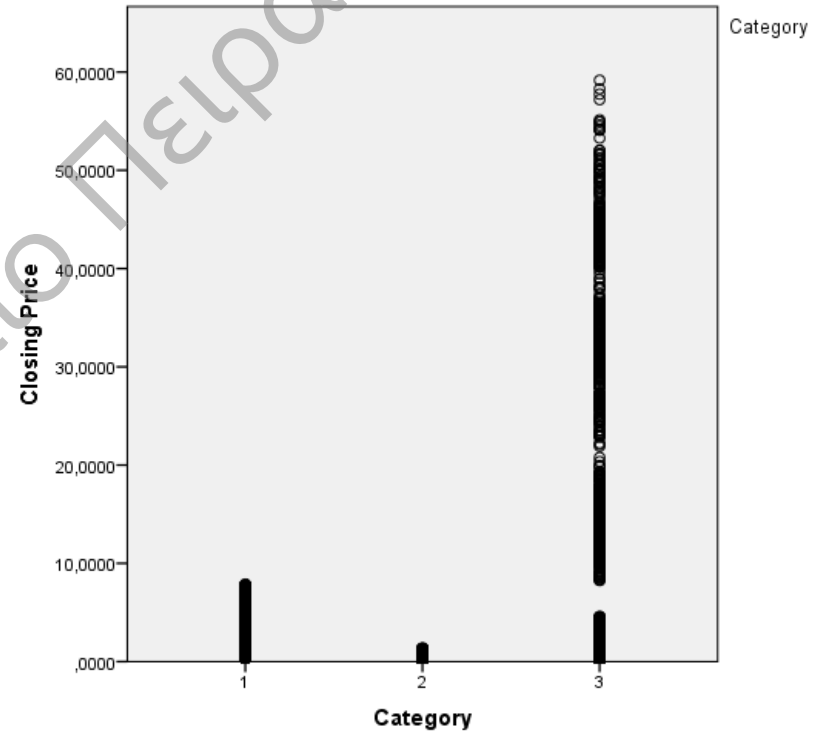
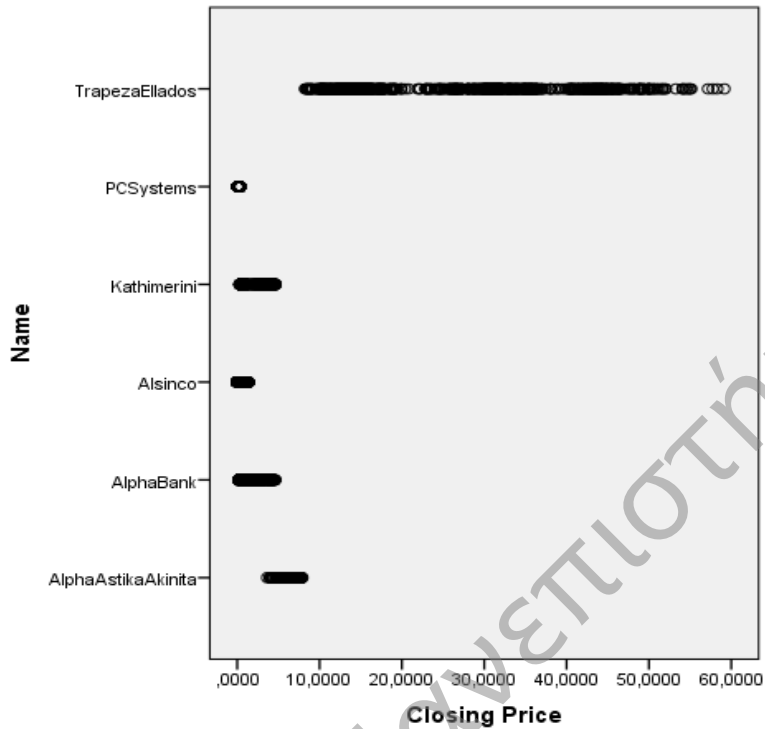
Τροχιές Πλοίων - Γράφημα



Τροχιές Πλοίων - Γράφημα



Τιμές Κλεισίματος Μετοχών



Δομή παρουσίασης

- Εισαγωγή
- Βασικές Έννοιες
- Σχετικές μελέτες
- Εφαρμογή Δεδομένων
- **Συμπεράσματα**

Συμπεράσματα

- Διαδικασία Δειγματοληψίας:
 - σημαντική για την μελέτη και την παρακολούθηση συνόλων .
 - βοηθάει στην εξαγωγή γενικών συμπερασμάτων,
 - προβλέπει μελλοντικές συμπεριφορές,
 - προτρέπει ανεπιθύμητα γεγονότα.
- Δεν μπορούν όλες οι τεχνικές να είναι κατάλληλες σε κάθε περίπτωση.
- Η επιλογή των διάφορων τεχνικών δειγματοληψίας σε σύνθετους τύπους δεδομένων πρέπει να γίνει προσεκτικά ανάλογα με το:
 - Ζητούμενο.
 - Δείγμα
 - Πώς είναι η δομή των δεδομένων που θα εξετασθούν.

Ερωτήσεις

