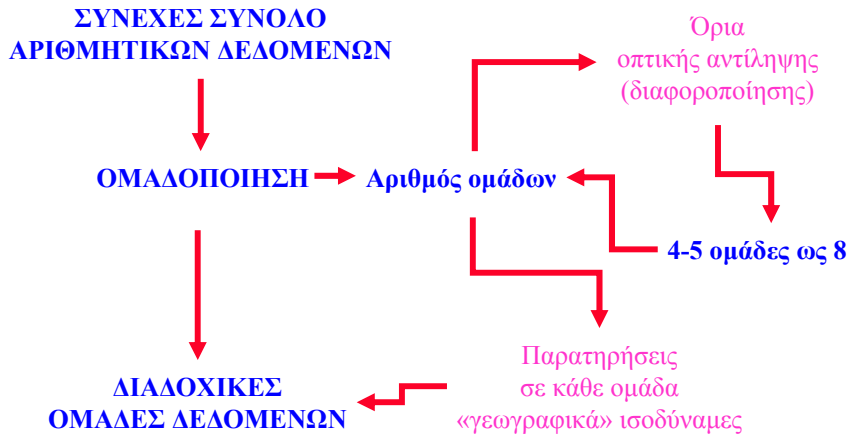


Ομαδοποίηση ποσοτικών δεδομένων



- Ομοιογένεια τιμών σε κάθε ομάδα
- Σημαντική διαφοροποίηση μεταξύ των ομάδων

Προσδιορισμός ορίων ομάδων - μέθοδοι ομαδοποίησης

ΜΕΘΟΔΟΙ ΟΜΑΔΟΠΟΙΗΣΗΣ

ΕΝΔΟΓΕΝΗ
ΚΡΙΤΗΡΙΑ

ΚΑΤΑΝΟΜΗ
ΤΙΜΩΝ
ΔΕΔΟΜΕΝΩΝ

ΕΞΩΓΕΝΗ
ΚΡΙΤΗΡΙΑ

ΚΡΙΣΙΜΕΣ
ΤΙΜΕΣ

π.χ. εισόδημα που
χαρακτηρίζει το
επίπεδο πτώχευσης

Μέθοδοι:

- Σταθερές τιμές ή ίσα διαστήματα
- Συστηματικά άνισα διαστήματα
- Ακανόνιστα ή μεταβαλλόμενα διαστήματα

Σταθερές τιμές ή ίσα διαστήματα

- *Μέθοδος ίσων διαστημάτων*

Υποδιαίρεση του εύρους μεταξύ χαμηλότερης και υψηλότερης τιμής σε επιθυμητό αριθμό ομάδων με ίσα διαστήματα

- *Μέθοδος παραμέτρων κανονικής κατανομής*

Εφαρμογή των παραμέτρων κανονικής κατανομής, τα όρια των διαστημάτων ορίζονται με βάση τη μέση τιμή και την τυπική απόκλιση του συνόλου των δεδομένων

- *Μέθοδος κανονικής τμηματοποίησης (quantiles)*

Υποδιαίρεση του συνόλου των δεδομένων σε ομάδες με ίσο αριθμό μελών ταξινομώντας τα δεδομένα κατά αύξουσα ή φθίνουσα σειρά

- *Μέθοδος ίσων διαστημάτων εμβαδού*

Υποδιαίρεση του συνόλου των δεδομένων σε ομάδες των οποίων τα εμβαδά των περιοχών είναι ισοδύναμα

Συστηματικά άνισα διαστήματα

Μέθοδοι εφαρμογής αναγωγικών σχέσεων ακολουθιών ή γεωμετρικών προόδων:

- Αύξουσες με σταθερό ρυθμό
- Αύξουσες με αυξανόμενο ρυθμό
- Αύξουσες με φθίνοντα ρυθμό
- Φθίνουσες με σταθερό ρυθμό
- Φθίνουσες με αυξανόμενο ρυθμό
- Φθίνουσες με φθίνοντα ρυθμό

$$a_n = a_{n-1} + (n-1)\omega$$

a: όρια διαστημάτων
n: αριθμός ομάδων
ω: εύρος διαστημάτων

$$b_n = b_1 \omega^{n-1}$$

b: όρια διαστημάτων
n: αριθμός ομάδων
ω: εύρος διαστημάτων

Ακανόνιστα ή μεταβαλλόμενα διαστήματα

ΟΜΑΔΟΠΟΙΗΣΗ ΒΑΣΙΣΜΕΝΗ ΣΕ ΔΙΑΓΡΑΜΜΑΤΑ

- Ιστόγραμμα
- Διάγραμμα συχνοτήτων
- Κλινογραφική καμπύλη
- Διάγραμμα αθροιστικών συχνοτήτων

ΟΜΑΔΟΠΟΙΗΣΗ ΒΑΣΙΣΜΕΝΗ ΣΕ ΕΠΑΝΑΛΗΠΤΙΚΕΣ ΜΕΘΟΔΟΥΣ

Εφαρμογή στατιστικών κριτηρίων:

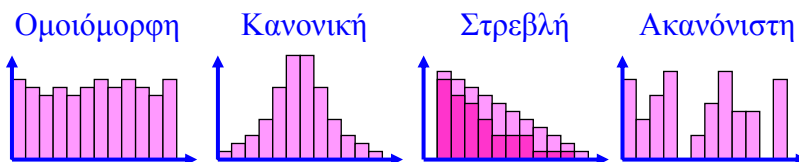
- Βέλτιστη προσαρμογή της μεταβλητότητας (Goodness of the variance fit - GVF)
Ελαχιστοποίηση των τετραγωνικών αποκλίσεων του μέσου όρου κάθε ομάδας
- Βέλτιστη προσαρμογή της απόλυτης απόκλισης (Goodness of absolute deviation fit - GADF)
Μεγιστοποίηση της βέλτιστης προσαρμογής των απολύτων αποκλίσεων

Μεθοδολογία ομαδοποίησης δεδομένων

Προσδιορισμός ορίων διαστημάτων:

- Καθορισμός κρίσιμων τιμών
- Στρογγύλευση ορίων διαστημάτων
- Σε σειρές θεματικών χαρτών κοινά όρια διαστημάτων
- Αξιοποίηση διαγραμμάτων στον προσδιορισμό των ορίων των διαστημάτων (διάγραμμα συχνοτήτων - ιστόγραμμα)

Κατανομή διαγράμματος συχνοτήτων



Επιλογή μεθόδου ομαδοποίησης

