ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΠΕΡΙΓΡΑΦΙΚΗ και ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Εισήγηση 1

Διδάσκων: Δαφέρμος Βασίλειος
ΤΜΗΜΑ ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΣΧΟΛΗΣ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται στην άδεια χρήσης **Creative Commons** και ειδικότερα

Αναφορά – Μη εμπορική Χρήση – Όχι Παράγωγο Έργο 3.0 Ελλάδα

(Attribution – Non Commercial – Non-derivatives 3.0 Greece)

R

[ή επιλογή ενός άλλου από τους έξι συνδυασμούς]

[και αντικατάσταση λογότυπου άδειας όπου αυτό έχει μπει (σελ. 1, σελ. 2 και τελευταία)]

- Εξαιρείται από την ως άνω άδεια υλικό που περιλαμβάνεται στις διαφάνειες του μαθήματος, και υπόκειται σε άλλου τύπου άδεια χρήσης. Η άδεια χρήσης στην οποία υπόκειται το υλικό αυτό αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΔΕΥΤΕΡΑ 24-3-14

ΕΙΣΗΓΗΣΗ 1

ΠΕΡΙΓΡΑΦΙΚΗ ΚΑΙ ΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

- Η έννοια της μεταβλητής
- Είδη μεταβλητών
- Η έννοια της παρατηρησιακής μονάδας
- Πληθυσμός Ενδιαφέροντος, Δείγμα, Μέγεθος Δείγματος,
- Τυχαιότητα, Αντιπροσωπευτικότητα, Σφάλματα.
- Τεχνικές Δειγματοληψίας που υπαινίσσεται ο τρόπος λήψης του Δείγματος

- ΚΛΙΜΑΚΕΣ ΜΕΤΡΗΣΗΣ
- Η κατανόηση των κλιμάκων μέτρησης μέσα από την κωδικοποίηση
- Δείκτες κεντρικής τάσης (measures of central tendency)
- Δείκτες διασποράς (measures of variation) ή μεταβλητότητας (variability) ή διασκόρπισης (dispersion)
- Δείκτες ασυμμετρίας
- Δείκτες ομοιογένειας

====================== =============== ============ ====

- Η έννοια της μεταβλητής

- Αποτελείται από 2 μέρη: το όνομα και την τιμή της. Για παράδειγμα όταν δηλώνουμε: $V1 = 7$ είναι σαφές ότι πρόκειται για μια μεταβλητή της οποίας το όνομα είναι $V1$ και η τιμή που τώρα της αποδίδεται, είναι 7.
-

=============== =================== ============== =======

- Είδη μεταβλητών
- Στις ***ποιοτικές μεταβλητές*** κατατάσσονται το φύλο, (άνδρας ή γυναίκα), η θρησκεία (Χριστιανός, Μουσουλμάνος, κτλ.), ο τόπος διαμονής (αστικός, ημιαστικός, αγροτικός), η οικογενειακή κατάσταση (έγγαμος, άγαμος, διαζευγμένος, χήρος).
- Στις ***ποσοτικές μεταβλητές*** κατατάσσονται η ηλικία, το βάρος, το ύψος, το εισόδημα, ο βαθμός ευφυΐας κτλ.
- Μια ***συνεχής ποσοτική μεταβλητή*** συνήθως κινείται ανάμεσα σε μια ελάχιστη και σε μια μέγιστη τιμή από το σύνολο των πραγματικών αριθμών (real numbers).
- ***Οι διακριτές ποσοτικές μεταβλητές*** δεν διατρέχουν όλες τις δυνατές τιμές ανάμεσα σε ένα ελάχιστο και σε μέγιστο, αλλά λαμβάνουν ***ορισμένες, διακριτές ή απαριθμητές*** τιμές, οι οποίες προκύπτουν από τις μετρήσεις και είναι πάντα ακέραιοι αριθμοί. Για παράδειγμα, μια οικογένεια δεν μπορεί να έχει 2,5 παιδιά.
- ***Θα λέμε ανεξάρτητη*** (independent) μια μεταβλητή, όταν αυτή έχει εισαχθεί από τον ερευνητή για να εκτιμηθεί η επίδρασή της πάνω σε μια άλλη που την χαρακτηρίζομε ως εξαρτημένη. Π.χ. Χρόνος-Πυρετός.

- Θα λέμε ***τυχαία μεταβλητή*** (random variable), τη μεταβλητή της οποίας οι τιμές δεν μπορούν να προσδιορισθούν με ακρίβεια, αλλά μέσω μιας διαδικασίας στην οποία σε κάθε τιμή της μεταβλητής αντιστοιχεί μια τιμή πιθανότητας.
- θα λέμε ***μη τυχαία μεταβλητή*** (fixed variable), τη μεταβλητή της οποίας οι τιμές μπορούν με ακρίβεια να προσδιοριστούν πριν από τη μέτρησή τους. Για παράδειγμα, η ταχύτητα ενός αυτοκινήτου το οποίο κινείται με σταθερή επιτάχυνση είναι γνωστή ανά πάσα στιγμή από γνωστούς μαθηματικούς τύπους.

============== ========== ==== ============== ==== =

- **Τι εννοούμε με τον όρο παρατηρησιακή μονάδα;**
- Αυτό που παρατηρούμε, ως ερευνητές, μέσα σε μια έρευνα, είναι δυνατόν να είναι ένας άνθρωπος, ένα μικρόβιο, ένα πειραματόζωο, ένα πολιτικό υποκείμενο, ένας ασθενής, ένα μαθητευόμενο άτομο, ένα φυτό, ένας αθλητής ή ένα βρέφος. Όλα αυτά ονομάζονται ***παρατηρησιακές μονάδες***, διότι οι επιδόσεις τους (σκόρ), ως προς κάποιο χαρακτηριστικό ή ιδιότητά τους, είναι αυτό που ενδιαφέρει την έρευνά μας.
- ================= ================= ======== =
- **Πληθυσμός Ενδιαφέροντος, Δείγμα, Μέγεθος Δείγματος,**
- **Τυχαιότητα, Αντιπροσωπευτικότητα, Σφάλματα**.

Οι στατιστικομαθηματικές τεχνικές, που θα αναπτύξουμε λεπτομερώς στα επόμενα, αναλαμβάνουν ***το 'πέρασμα' από το δείγμα στον πληθυσμό*** και προσδιορίζουν με ακρίβεια το πιθανό σφάλμα των μετρήσεών μας. Ωστόσο,

- Το δείγμα θα πρέπει να είναι τυχαίο, πράγμα που σημαίνει ότι κάθε στοιχείο (παρατηρησιακή μονάδα) του πληθυσμού, θα πρέπει να έχει ίσες δυνατότητες (πιθανότητα) να συμπεριληφθεί στο δείγμα.
- Το δείγμα θα πρέπει να είναι αντιπροσωπευτικό, που σημαίνει ότι το δείγμα θα πρέπει να έχει τα χαρακτηριστικά και τις ιδιότητες του πληθυσμού από τον οποίο προέρχεται.
- Ο πληθυσμός θα πρέπει να καθορίζεται στην αρχή της κάθε έρευνας με τη μεγαλύτερη δυνατή ακρίβεια. Μόνο τότε, είναι δυνατή η επιλογή αντιπροσωπευτικού δείγματος.
- ================ ============ ==== ======== ====
- **Κάθε συγκεκριμένος τρόπος λήψης ενός δείγματος συνιστά και μια τεχνική δειγματοληψίας,** δηλ. μια επιστημονική μέθοδο συγκρότησης του δείγματός μας από ένα καθορισμένο πληθυσμό. Στις Κοινωνικές Επιστήμες συνήθως χρησιμοποιούνται πέντε τεχνικές δειγματοληψίας ίσων, όπως λέγεται, πιθανοτήτων:
- α) η απλή τυχαία δειγματοληψία,
- β) η συστηματική δειγματοληψία,
- γ) η δειγματοληψία κατά στρώματα,
- δ) η δειγματοληψία κατά ομάδες, και τέλος,
- ε) η δειγματοληψία κατά στάδια.
- *Θα αναφερθούμε σε αυτές τις τεχνικές στα παρακάτω ...*

- *ΚΛΙΜΑΚΕΣ ΜΕΤΡΗΣΗΣ*
- **Ονομαστικές ή Κατηγορικές Κλίμακες (nominal scales)**
- **Τακτικές κλίμακες (ordinal scales)**
- **Αριθμητικές ή ισοδιαστημικές κλίμακες (interval scales)**
- **Αναλογικές κλίμακες (ratio scales)**
- ============ =============== ========== ===
- Θα λέμε ***ονομαστικές τις κλίμακες μέτρησης*** στις οποίες η κατάταξη των υποκειμένων γίνεται σε καλά προσδιορισμένες, σαφώς διακρίσιμες μεταξύ τους, ισοδύναμες, και οπωσδήποτε αμοιβαία αποκλειόμενες, αν έχουμε **διχοτομική κλίμακα** μέτρησης, κατηγορίες. Π.χ. Το φύλο, η αστικότητα, η θρησκεία κτλ.
- =========== =================== ======== ====
- Θα λέμε ***τακτικές κλίμακες μέτρησης*** εκείνες στις οποίες η ένταξη των υποκειμένων γίνεται σε κατηγορίες σαφείς, ισοδύναμες, αλλά και διατεταγμένες μεταξύ τους. Π.χ. Η σειρά προτεραιότητας στο ΙΚΑ, η κατάταξη σε κάποιο αγώνισμα-πρώτος, δεύτερος, τρίτος κτλ.
- ================= ============ ========== ====
- Θα λέμε ***αριθμητικές ή (ισο)διαστημικές*** τις κλίμακες μέτρησης στις οποίες τα υποκείμενα εντάσσονται σε σαφώς καθορισμένες, αμοιβαία αποκλειόμενες, διατεταγμένες κατηγορίες, και οι οποίες όμως έχουν και το εξής, επιπλέον χαρακτηριστικό: **χρησιμοποιούν σταθερή μονάδα μέτρησης**. Ο χρόνος (με μονάδες μέτρησης το λεπτό, το δευτερόλεπτο, την ώρα κτλ., η απόσταση, η ηλικία, η θερμοκρασία, σε κλίμακα Celsius ή Fahrenheit κτλ., είναι κλασικά παραδείγματα αριθμητικών κλιμάκων.
- ============= =========== =========== == ====
- Και τέλος, Θα λέμε ***αναλογικές κλίμακες μέτρησης***, εκείνες που διατηρούν όλα τα χαρακτηριστικά των διαστημικών κλιμάκων, και επιπλέον διαθέτουν πραγματικό σημείο αναφοράς το οποίο αντιστοιχεί στο απόλυτο μηδέν. Το σημείο αυτό είναι ***ένα γνήσιο σημείο***, είναι ένα εναρκτήριο σημείο, με την έννοια ότι το χαρακτηριστικό ή η ιδιότητα που μετράει η κλίμακα στο σημείο αυτό, δεν υπάρχει. Κλασικά παραδείγματα αναλογικής κλίμακας είναι η ταχύτητα, η απόλυτη θερμοκρασία (βαθμοί Kelvin), η πίεση του αίματος, το βάρος, η επιτάχυνση, η μάζα κτλ.

================ ====================== ============= ====

- ***Η κατανόηση των κλιμάκων μέτρησης μέσα από την κωδικοποίηση***

  Ας πάρουμε το παρακάτω ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ και ας προσπαθήσουμε μαζί να το κωδικοποιήσουμε:

Telephone number……….. ……
Code Questionnaire………
Code Interviewer.....…….

ΑΝΩΝΥΜΟ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Καλημέρα. Σας παίρνουμε από το τμήμα Πολιτικής Επιστήμης του Π. Κ., κάνουμε μια εργασία στο μάθημα της Κοινωνικής Στατιστικής και θα θέλαμε τη γνώμη σας για το ρόλο των Τραπεζών στη σημερινή συγκυρία. Θα θέλατε να μας απαντήσετε;

=== ====

1)Κατά τη γνώμη σας, οι ΤΡΑΠΕΖΕΣ, <u>σε ποιο βαθμό βοηθούν</u> την οικονομική ανάπτυξη της χώρας; ⌂ καθόλου ⌂ λίγο ⌂ αρκετά ⌂ πολύ ⌂ ΔΓ/ΔΑ

==

2)Εσείς ο ίδιος/α , το Νοικοκυριό σας, ή η Επιχείρησή σας, <u>σε ποιο βαθμό</u> <u>είστε ευχαριστημένος</u> από την πολιτική των ΤΡΑΠΕΖΩΝ απέναντί σας :

⌂ καθόλου ⌂ λίγο ⌂ αρκετά ⌂ πολύ ⌂ ΔΓ/ΔΑ

===

3) Σε ποιό βαθμό πιστεύετε ότι η σημερινή Κυβέρνηση <u>είναι σε θέση να ελέγξει</u> τις ΤΡΑΠΕΖΕΣ; ⌂ καθόλου ⌂ λίγο ⌂ αρκετά ⌂ πολύ ⌂ ΔΓ/ΔΑ

===

4) Εσείς ο ίδιος/α έχετε την αίσθηση ότι οι ΤΡΑΠΕΖΕΣ βάζουν πανωτόκια;

⌂ ΝΑΙ ⌂ ΟΧΙ ⌂ ΔΓ/ΔΑ

============== ============= ================= ======== =====

5) Σε ποιό βαθμό <u>είστε ικανοποιημένος/η</u> από τα μέτρα που έλαβε πρόσφατα η σημερινή Κυβέρνηση για την πρώτη κατοικία; ⌂ καθόλου ⌂ λίγο ⌂ αρκετά ⌂ πολύ

=========== ================== ============ ========= ======

6) Σε ποιό βαθμό πιστεύετε ότι οι ΤΡΑΠΕΖΕΣ πειθαρχούν στις αποφάσεις των Δικαστηρίων; ⌂ καθόλου ⌂ λίγο ⌂ αρκετά ⌂ πολύ ⌂ ΔΓ/ΔΑ

============== ==================== ================= =====

7) Κατά τη γνώμη σας <u>ποιά</u> από τις παρακάτω μελλοντικές Κυβερνήσεις μπορεί να ασκήσει αυστηρό/αποτελεσματικό έλεγχο πάνω στις ΤΡΑΠΕΖΕΣ (μια απάντηση παρακαλώ);

⌂ μια Κυβέρνηση με κορμό τη ΝΔ
⌂ μια Κυβέρνηση με κορμό το ΣΥΡΙΖΑ
⌂ μια Κυβέρνηση της Κεντροαριστεράς
⌂ μια Κυβέρνηση της Χρυσής Αυγής
⌂ μια Κυβέρνηση Εθνικής Ενότητας
⌂ καμία Κυβέρνηση

======================== ================ =================

ΦΥΛΟ: □ άνδρας □ γυναίκα

ΜΟΡΦΩΤΙΚΟ ΕΠΙΠΕΔΟ: □ Στοιχειώδης Εκπαίδευση □ Μέση Εκπαίδευση □ Ανώτερη /Ανώτατη

ΕΠΑΓΓΕΛΜΑ: □ Αγρότης □ Ιδιωτικός Υπάλληλος □ Δημόσιος Υπάλληλος □ Ελεύθερος Επαγγελματίας □ Φοιτητής/τρια □ Συνταξιούχος □ Νοικοκυρά □ Άνεργος

ΗΛΙΚΙΑ :…………..

ΝΟΜΟΣ (που ψηφίζετε): □ Χανίων □ Ρεθύμνης □ Ηρακλείου □ Λασιθίου

============================== ========================= ====== = ==
======================== ====================== ================ ===

============================ ======================= =============

ΠΡΟΒΛΗΜΑ 1

========================= ================ ================= ===

Ο Καθηγητής κ. Δεληβοριάς είχε στο σεμινάριό του, 6 φοιτητές και 9 φοιτήτριες. Οι βαθμοί που έλαβαν οι φοιτητές στο τέλος του εξαμήνου ήταν οι εξής:

5,6,6,7,7,8,

ενώ των φοιτητριών ήταν οι εξής:

4,5,6,7,7,7,8,9,10.

Να βρείτε όλους τους δείκτες Κεντρικής τάσης, Διασποράς, Ασυμμετρίας και Ομοιογένειας.

======================= ===================== =================

- **ΔΕΙΚΤΕΣ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ (measures of central tendency)**
- (Μέσος όρος=mean, Διάμεσος=Median, Δεσπόζουσα τιμή=Mode).
- Θα λέμε αριθμητικό μέσο όρο ή μέση τιμή (mean) ενός δείγματος n παρατηρήσεων x_1, x_2, x_3,…,x_n μιας μεταβλητής X και θα το συμβολίζουμε με $\overline{X}$, το πηλίκο του αθροίσματος όλων τιμών της μεταβλητής δια του πλήθους τους:
-
- $$\overline{X} = \frac{f_1x_1 + f_2x_2 + ... + f_kx_k}{f_1 + f_2 + ... + f_k} = \frac{\sum_{i=1}^{k} f_ix_i}{\sum_{i=1}^{k} f_i} \quad (1)$$
- Εφαρμόζουμε στη σχέση (1) τα δεδομένα του Προβλήματος (1) και λαμβάνουμε:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_ix_i}{\sum_{i=1}^{k} f_i} = \frac{1\cdot 4 + 2\cdot 5 + 3\cdot 6 + 5\cdot 7 + 2\cdot 8 + 1\cdot 9 + 1\cdot 10}{1 + 2 + 3 + 5 + 2 + 1 + 1} = \frac{102}{15} = 6,8$$

- ***Ιδιότητες Μέσης τιμής:***

1) Η μέση τιμή είναι ένας δείκτης εύκολα κατανοητός. Οι τιμές του παραδείγματός μας ήταν οι βαθμοί φοιτητών με βάση την γνωστή στο Πανεπιστήμιο κλίμακα 1-10, οπότε εύκολα προκύπτει το συμπέρασμα, ότι πρόκειται για μια ομάδα μάλλον χαμηλών αποδόσεων.
2) Για τον υπολογισμό της μέσης τιμής, όπως κι ίδιοι παρατηρήσαμε στο παράδειγμά μας, έλαβαν μέρος όλες οι τιμές του δείγματός μας. Αυτό κάνει τον εν λόγω δείκτη ‘δημοκρατικότερο’ από όλους τους δείκτες κεντρικής τάσης. Διότι κανένας άλλος δείκτης κεντρικής τάσης, όπως θα δούμε παρακάτω, δεν λαμβάνει υπ’ όψιν του ***όλες μα όλες και κάθε μια με το ιδιαίτερο βάρος (αξία) που έχει,*** τις τιμές του δείγματος από το οποίο προήλθε.
3) Για κάθε ομάδα παρατηρήσεων υπάρχει μία και μόνο μέση τιμή. Δηλ. αυτός ο δείκτης έχει το χαρακτηριστικό της μοναδικότητας.
4) Ωστόσο, η μέση τιμή επηρεάζεται, από ακραίες τιμές (outliers). Για παράδειγμα, αυτό το 10 αν το είχαν οι 5 καλύτεροι φοιτητές του προηγούμενου δείγματος, η μέση τιμή θα επηρεάζονταν σοβαρά (7,33).
5) Η μέση τιμή, όπως είναι λογικό, δεν έχει νόημα για κατηγορικά, αλλά κυρίως ή μόνον για αριθμητικά και γενικά για ποσοτικά δεδομένα.

6) Αν σε κάθε τιμή ενός δείγματος προσθέσουμε τον ίδιο σταθερό αριθμό C, τότε η μέση τιμή του δείγματος αυξάνει κατά C. Δηλ. αν ήταν $\overline{X}$ γίνεται $\overline{X}+C$.
7) Αν κάθε τιμή ενός δείγματος πολλαπλασιαστεί τον ίδιο σταθερό αριθμό C, τότε και η μέση τιμή του δείγματος πολλαπλασιάζεται με C. Δηλ. αν ήταν $\overline{X}$ γίνεται $\overline{X}\cdot C$.
8) Η μέση τιμή έχει γραμμική συμπεριφορά. Αυτό σημαίνει ότι αν $\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$ είναι, αντίστοιχα, οι μέσες τιμές k ομάδων παρατηρήσεων με μεγέθη n_1, n_2,…,n_k, τότε το σύνολο που περιλαμβάνει όλες αυτές τις παρατηρήσεις, δηλ. που έχει μέγεθος $n=n_1+n_2+\ldots+n_k$, έχει μέση τιμή ίση με:

$$\overline{X}=\frac{n_1\overline{X}_1+n_2\overline{X}_2+...+n_k\overline{X}_k}{n_1+n_2+...+n_k} \qquad (3.3)$$

============================ ======================== ======

♦ **Θα λέμε διάμεσο (median) ενός δείγματος n παρατηρήσεων, οι οποίες έχουν διαταχθεί σε αύξουσα σειρά και θα το συμβολίζουμε με Δμ, τη μεσαία παρατήρηση αν ο n είναι περιττός ή το ημιάθροισμα των δύο μεσαίων παρατηρήσεων αν ο n είναι άρτιος.**

Παράδειγμα

Σκόπιμα παίρνουμε πάλι, ως παράδειγμα, το ίδιο δείγμα που πήραμε και στην περίπτωση του μέσου όρου δηλ. το

4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

Προφανώς, αφού οι παρατηρήσεις είναι 15, δηλ. περιττός αριθμός η μεσαία παρατήρηση δηλ, αυτή που έχει δεξιά και αριστερά της ίσο αριθμό παρατηρήσεων, αυτή είναι η διάμεσος, το 7.

Αν λάβουμε το ίδιο δείγμα αλλά χωρίς την τιμή 10 στο τέλος, δηλ. το δείγμα

4, 5, 5, 6, 6, 6, **7**, **7**, 7, 7, 7, 8, 8, 9

Τότε προφανώς το ημιάθροισμα των δύο μεσαίων παρατηρήσεων δηλ. $\frac{7+7}{2}$=7 θα ήταν και πάλι η διάμεσος του δείγματός μας.

Ιδιότητες της διαμέσου

1) Η διάμεσος, όπως έδειξαν και τα προηγούμενα παραδείγματα, δεν επηρεάζεται από ακραίες τιμές.
2) Στον υπολογισμό της τιμής της, όπως έδειξαν και τα προηγούμενα παραδείγματα, η διάμεσος δεν χρησιμοποιεί

όλες τις τιμές του δείγματος. Ως εκ τούτου είναι λιγότερο 'δημοκρατικός' δείκτης, από τη μέση τιμή.

3) Η διάμεσος δεν έχει νόημα για ποιοτικά δεδομένα.
4) Η διάμεσος, όπως και η μέση τιμή, υπολογίζεται εύκολα.
5) Η διάμεσος έχει μοναδικότητα για κάθε ομάδα παρατηρήσεων.
6) Η διάμεσος είναι, όπως και η μέση τιμή, εύκολα κατανοητή.
7) Αν σε κάθε τιμή ενός δείγματος προσθέσουμε τον ίδιο σταθερό αριθμό C, τότε η διάμεσος του δείγματος αυξάνει κατά C.
8) Αν κάθε τιμή ενός δείγματος πολλαπλασιαστεί τον ίδιο σταθερό αριθμό C, τότε και η διάμεσος του δείγματος πολλαπλασιάζεται με C.
9) Η διάμεσος, πολύ σπάνια ή καθόλου απαντάται σε απλές ή σύνθετες στατιστικές αναλύσεις.

===================== =========== ========= ====

♦Θα λέμε δεσπόζουσα ή επικρατούσα τιμή (mode), ενός δείγματος n παρατηρήσεων και θα τη συμβολίζουμε με $\Delta_{\sigma\pi}$, την παρατήρηση που εμφανίζεται με τη μεγαλύτερη συχνότητα.

Σκόπιμα παίρνουμε πάλι, ως παράδειγμα, το ίδιο δείγμα που πήραμε και στην περίπτωση του μέσου όρου και της διαμέσου δηλ. το

4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 9, 10

Ιδιότητες της δεσπόζουσας τιμής

- Η δεσπόζουσα τιμή, δεν επηρεάζεται από ακραίες τιμές
- Η δεσπόζουσα τιμή, όπως έδειξαν και τα προηγούμενα παραδείγματα, είναι δυνατόν να υπολογισθεί, ακόμη κι όταν έχουμε ελλιπή δεδομένα.
- Στη διαμόρφωση της τιμής της, η δεσπόζουσα τιμή, δεν χρησιμοποιεί όλες τις τιμές από το δείγμα που προέρχεται, αλλά τις λιγότερες, συγκριτικά με τους άλλους δύο δείκτες κεντρικής τάσης. Ως εκ τούτου, είναι ο λιγότερο 'δημοκρατικός' δείκτης, από τους όλους δείκτες κεντρικής τάσης.
- Η δεσπόζουσα τιμή δεν έχει πάντα μοναδικότητα. Για παράδειγμα, το δείγμα 2,3,4,4,4,7,7,7,9,9 έχει δύο δεσπόζουσες τιμές: το 4 και το 7. Το SPSS αποκαλεί μια τέτοια περίπτωση δείγματος multiple Mode.
- Η δεσπόζουσα τιμή ορίζεται και σε ποιοτικά και σε ποσοτικά δεδομένα.
- Η δεσπόζουσα τιμή υπολογίζεται και είναι πολύ εύκολα κατανοητή.

- Αν σε κάθε τιμή ενός δείγματος προσθέσουμε τον ίδιο σταθερό αριθμό C, τότε και η δεσπόζουσα τιμή του δείγματος αυξάνει κατά C.
- Αν κάθε τιμή ενός δείγματος πολλαπλασιαστεί τον ίδιο σταθερό αριθμό C, τότε και η δεσπόζουσα τιμή του δείγματος πολλαπλασιάζεται με C.
- Η δεσπόζουσα τιμή πολύ σπάνια ή καθόλου απαντάται σε απλές ή σύνθετες στατιστικές αναλύσεις.

================== ===================== ==========

ΔΕΙΚΤΕΣ ΔΙΑΣΠΟΡΑΣ (measures of variation) ή μεταβλητότητας (variability), ή διασκόρπισης (dispersion)

Όπως μαρτυρεί και το όνομά τους, οι δείκτες αυτοί δείχνουν κατά πόσο ‘απλώνεται’ μια κατανομή δεδομένων, κατά πόσον δηλαδή διασπείρονται οι τιμές της .

♦ **Θα λέμε διασπορά ή διακύμανση, ή μέση τετραγωνική απόκλιση (variance) n παρατηρήσεων x_1, x_2, x_3,…,x_n μιας μεταβλητής X και θα τη συμβολίζουμε με s^2, τον αριθμό που προκύπτει από τους δύο παρακάτω τύπους:**

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \quad \text{αν} \quad n < 30 \qquad (2)$$

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} \quad \text{αν} \quad n > 30 \qquad (3)$$

Ωστόσο, αν έχουμε μια άλλη κατανομή δεδομένων την x_1, x_2, x_3,…, x_k στην οποία όμως η παρατήρηση x_1 έχει συχνότητα εμφάνισης f_1, η x_2 συχνότητα εμφάνισης f_2, …, η x_k συχνότητα εμφάνισης f_k , τότε για τη διασπορά ισχύουν οι σχέσεις:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \bar{X})^2}{\sum_{i=1}^{k} f_i - 1} \qquad \text{αν} \ \sum_{i=1}^{k} f_i < 30 \qquad (4)$$

$$s^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{\sum_{i=1}^{k} f_i} \qquad \text{αν} \quad \sum_{i=1}^{k} f_i > 30 \qquad (5)$$

Παράδειγμα

Θα επιλέξουμε, σκόπιμα και πάλι, το ίδιο παράδειγμα, δηλ. το:

4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 9, 10

Η μέση τιμή γι αυτά τα δεδομένα είναι, όπως είδαμε $\overline{X}$ **=6,8**

Προφανώς η σχέση η οποία πρέπει να εφαρμόσουμε είναι η (4) από την οποία προκύπτει:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{\sum_{i=1}^{k} f_i - 1} =$$

$$\frac{1(4-6,8)^2 + 2(5-6,8)^2 + 3(6-6,8)^2 + 5(7-6,8)^2 + 2(8-6,8)^2 + 1(9-6,8)^2 + 1(10-6,8)^2}{1+2+3+5+2+1+1-1} =$$

$$= \frac{34,398}{14} = 2,457$$

Ιδιότητες της διασποράς

1) Η διασπορά είναι ένα μέτρο που μας δείχνει πόσο πολύ απέχουν δηλ συγκεντρώνονται ή απομακρύνονται οι τιμές μιας κατανομής από τη μέση τιμή. Έτσι, αν οι τιμές μιας μεταβλητής δεν διαφέρουν πολύ από τη μέση τιμή, τότε είναι σαφές ότι η διασπορά είναι μικρή. Αντίθετα, όταν οι τιμές διαφέρουν πολύ από τη μέση τιμή, τις βλέπουμε δηλ., να διασκορπίζονται σε μεγάλη απόσταση, εκατέρωθεν τη μέσης τιμής, τότε η διασπορά είναι μεγάλη.
2) Η διασπορά είναι μια αξιόπιστη παράμετρος μεταβλητότητας των δεδομένων μας.
3) Η διασπορά δεν έχει τις ίδιες μονάδες μέτρησης με τη μέση τιμή, ούτε και με τις ίδιες τις παρατηρήσεις μας. Για την ακρίβεια η διασπορά εκφράζεται σε μια μονάδα που είναι το τετράγωνο της μονάδας μέτρησης του χαρακτηριστικού.

4) Αν σε κάθε τιμή ενός δείγματος προσθέσουμε τον ίδιο σταθερό αριθμό C, η διασπορά του δείγματος παραμένει αμετάβλητη.
5) Αν κάθε τιμή ενός δείγματος πολλαπλασιαστεί με τον ίδιο σταθερό αριθμό C, τότε η διασπορά του δείγματος πολλαπλασιάζεται με C^2.

================== ===================== ===========

♦ **Θα λέμε τυπική απόκλιση (standard deviation) ενός δείγματος, τη θετική τετραγωνική ρίζα της διασποράς. Δηλ. ισχύει:**

$$s = \sqrt{s^2} \qquad (6)$$

Παράδειγμα

Αν πάρουμε το προηγούμενο δείγμα παρατηρήσεων δηλ. το:

4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 9, 10

δεδομένου ότι έχουμε υπολογίσει γι' αυτό τη διασπορά, είναι εύκολο να υπολογίσουμε άμεσα και την τυπική του απόκλιση. Πράγματι, από τη σχέση (6) έχουμε:

$$s = \sqrt{s^2} = \sqrt{2{,}457} = 1{,}568$$

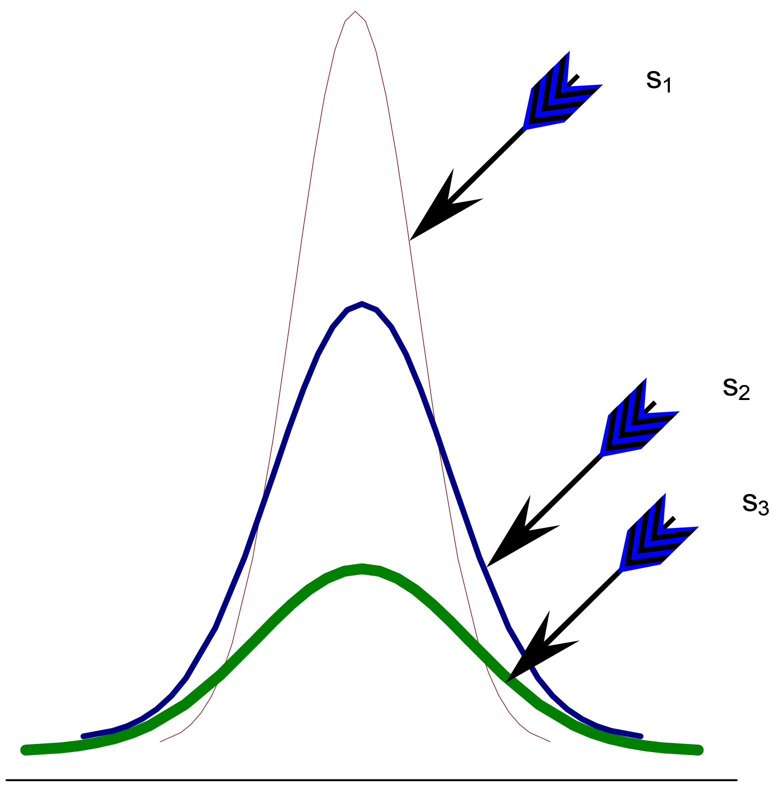
Ιδιότητες της τυπικής απόκλισης

- Η τυπική απόκλιση είναι το πιο αξιόπιστο και πιο χρήσιμο μέτρο από όλα τα μέτρα μεταβλητότητας.
- Η τυπική απόκλιση αποτελεί **το σταθερότερο δείκτη μεταβλητότητας**, πράγμα που σημαίνει ότι αν από ένα πληθυσμό λάβουμε πολλά όμοια δείγματα και υπολογίσουμε για κάθε ένα από αυτά όλους τους δείκτες μεταβλητότητας, τότε θα διαπιστώσουμε ότι οι τυπικές αποκλίσεις όλων αυτών των δειγμάτων, θα διαφέρουν μεταξύ τους, λιγότερο από ότι θα διαφέρουν μεταξύ τους οι τιμές των άλλων δεικτών μεταβλητότητας.
- Η τυπική απόκλιση διαθέτει ίδιες μονάδες μέτρησης, τόσο με τη μέση τιμή, όσο και με τις ίδιες τις παρατηρήσεις μας.
- Αν σε κάθε τιμή ενός δείγματος προσθέσουμε τον ίδιο σταθερό αριθμό C, η τυπική απόκλιση του δείγματος παραμένει αμετάβλητη.

- Αν κάθε τιμή ενός δείγματος πολλαπλασιαστεί με τον ίδιο σταθερό αριθμό C, τότε η τυπική απόκλιση του δείγματος πολλαπλασιάζεται με την απόλυτη τιμή αυτού του αριθμού $|C|$.

============= ============== ============ ======== =

Η αναπαραστατική ερμηνεία της τυπικής απόκλισης



Σχ. 3.0

================== ==================== ============ ======= ==

♦ **Θα λέμε εύρος(range) ενός δείγματος και θα το συμβολίζουμε με R, τη διαφορά ανάμεσα στη μεγαλύτερη και στη μικρότερη τιμή του. Δηλ. αν min είναι η μικρότερη και max είναι η μεγαλύτερη τιμή ενός δείγματος, τότε για το εύρος του θα έχουμε: R=max-min.**

Παράδειγμα

Λαμβάνουμε και πάλι το ίδιο δείγμα παρατηρήσεων, το:

4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

Τότε, προφανώς R=max-min=10-4=6

Ιδιότητες του εύρους

- Το εύρος είναι ένα κατανοητό και εύκολα υπολογίσιμο μέτρο μεταβλητότητας.
- Το εύρος δεν είναι ένας αξιόπιστος δείκτης διασποράς, καθώς βασίζεται σε δύο μόνο ακραίες παρατηρήσεις.

====================== =========== ======= ====== ==

♦ **Θα λέμε τεταρτημόρια (Quartiles), τρεις τιμές, ή αλλιώς, τρία σημεία, τα οποία διαιρούν την κατανομή του δείγματός μας σε τέσσερα ίσα τμήματα. Το πρώτο τεταρτημόριο (Q_1), είναι το σημείο της κατανομής (σημείο πάνω στον οριζόντιο άξονα, φυσικά), κάτω από το οποίο βρίσκεται το 25 % των τιμών της κατανομής. Το δεύτερο τεταρτημόριο (Q_2), είναι το σημείο κάτω (αριστερά) από το οποίο βρίσκεται το 50% των τιμών της κατανομής. Άρα ισούται με τη διάμεσο δηλ. έχουμε Q_2=Δμ . Το τρίτο τεταρτημόριο είναι το σημείο της κατανομής, κάτω από το οποίο βρίσκεται το 75% των παρατηρήσεών μας.**

Η διαφορά Q_3-Q_1 ονομάζεται ενδοτεταρτημοριακό εύρος (interquartile range) και προφανώς αναφέρεται στο 50 % των μεσαίων τιμών της κατανομής.

====================== ========== ========= ====== = =

♦ **Θα λέμε εκατοστημόρια (Percentiles), τις τιμές P_1, P_2, …,P_{99} , οι οποίες χωρίζουν την κατανομή του δείγματός μας σε 100 ίσα μέρη. Το P_k εκατοστημόριο επομένως, θα είναι η τιμή για την οποία το k% των παρατηρήσεων είναι μικρότερες του P_k , ενώ το (100-k) % των παρατηρήσεων είναι μεγαλύτερες του P_k .**

Με βάση τα παραπάνω τα εκατοστημόρια P_{25}, P_{50}, P_{75} θα συμπίπτουν με τα τεταρτημόρια Q_1, Q_2, Q_3, και βέβαια το εκατοστημόριο P_{50} θα συμπίπτει με τη διάμεσο.

Παραδείγματα

Ας λάβουμε ξανά το αρχικό μας δείγμα, το:

4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

→ Είναι σαφές ότι P_{50}=Δμ=7.

→ Επίσης είναι σαφές ότι η διάμεσος των παρατηρήσεων αριστερά της διαμέσου, δηλ. των παρατηρήσεων 4, 5, 5, **6**, 6, 6, 7 είναι το 6, επομένως $P_{25}=Q_3=6$.

→ Ακόμη, είναι σαφές ότι η διάμεσος των παρατηρήσεων δεξιά της διαμέσου, δηλ. των παρατηρήσεων 7, 7, 7, **8**, 8, 9, 10 είναι το 8, οπότε $P_{75}=Q_3=8$.

=========== ============ ========== ======= ====

♦ **Θα λέμε τυπικό σφάλμα του μέσου όρου (standard error of mean) και θα το συμβολίζουμε με $s_{\bar{X}}$ το πηλίκο της τυπικής απόκλισης του δείγματος, προς την τετραγωνική ρίζα του μεγέθους του δείγματος. Δηλ.**

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} \qquad \textbf{(7)}$$

Παράδειγμα

Ας λάβουμε και πάλι το ίδιο δείγμα, δηλ. το:

4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

Επειδή η τυπική απόκλιση γι αυτό το δείγμα έχει υπολογισθεί και είναι s=1,568 και το μέγεθός του είναι n=15 , η σχέση (7) μας δίνει:

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{1{,}568}{\sqrt{15}} = \frac{1{,}568}{3{,}88} = 0{,}405$$

========== =========== ============= ==============

♦ **Θα λέμε στρεβλότητα ή λοξότητα (skewness), μιας κατανομής x_1, x_2, x_3,…, x_k στην οποία όμως η παρατήρηση x_1 έχει συχνότητα εμφάνισης f_1, η x_2 συχνότητα εμφάνισης f_2, …, η x_k συχνότητα εμφάνισης f_k , το μέγεθος που ορίζεται από τη σχέση:**

$$SK = \frac{\dfrac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^3}{\sum_{i=1}^{k} f_i}}{s^3} \qquad \textbf{(8)}$$

Η στρεβλότητα, είναι ένα μέτρο που μας δείχνει την έκταση στην οποία μια κατανομή τιμών αποκλίνει από τη συμμετρία, γύρω από το μέσο όρο. Έτσι:

▹ Αν σε μια κατανομή έχουμε SK > 0 ⇒ η κατανομή μας είναι ασύμμετρη δεξιά (positive skewed), δηλ. έχει την ουρά της στα δεξιά (βλ. Σχ.3.1). Αυτό σημαίνει ότι στα δεξιά υπάρχει έλλειμμα τιμών, ενώ προς την άλλη μεριά, δηλ. προς τα αριστερά, υπάρχει πλεόνασμα τιμών. Ας μη λησμονούμε ωστόσο, ότι στα δεξιά είναι οι υψηλού μεγέθους τιμές (***μεγάλης αξίας***), ενώ προς τα αριστερά οι χαμηλού μεγέθους (***μικρής αξίας***) τιμές του δείγματός μας.

▹Αν σε μια κατανομή έχουμε SK < 0 ⇒ η κατανομή μας είναι ασύμμετρη αριστερά (negative skewed), δηλ. έχει την ουρά της στα αριστερά (βλ. Σχ.3.2).

▹ Αν σε μια κατανομή έχουμε SK = 0 ⇒ η κατανομή μας δεν είναι ασύμμετρη ούτε δεξιά, ούτε αριστερά, δηλ. δεν έχει ουρά, οπότε είναι μια συμμετρική κατανομή (βλ. Σχ.3.3).

Σχόλιο: Αν υποθέσουμε ότι ένας ερευνητής εκτελεί ψυχομετρικά πειράματα, τι θα μπορούσε να σημαίνει μια τιμή για παράδειγμα SK=0,86 ; Οι Darren and Mallery (2001), ισχυρίζονται ότι για τη στρεβλότητα μια τιμή ± 1, θεωρείται εξαιρετική για ψυχομετρικά πειράματα, ενώ μια τιμή ± 2 είναι σε αρκετές περιπτώσεις αποδεκτή. Ωστόσο, εμείς δεν αυτό δεν μπορούμε να το λάβουμε ως κανόνα για τις τιμές της στρεβλότητας. Για κάθε συγκεκριμένη εφαρμογή θεωρούμε ότι είναι καλό να αναζητούμε την προϊστορία του πράγματος, δηλ. να ανατρέχουμε στη βιβλιογραφία του θέματος που ερευνούμε.

Παράδειγμα

Αν έχουμε το ίδιο δείγμα δηλ. το

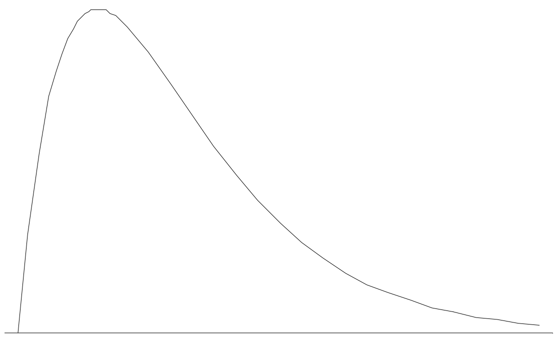
4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

τότε επειδή s=1,568 η σχέση (8) δίνει:

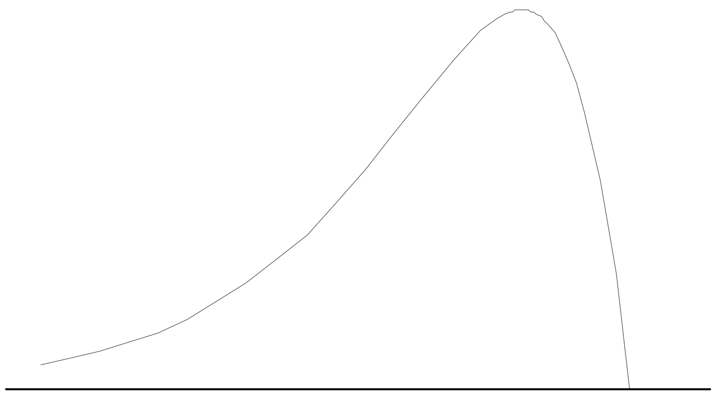
$$SK = \frac{\dfrac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^3}{\sum_{i=1}^{k} f_i}}{s^3} =$$

$$= \frac{\dfrac{1(4-6,8)^3 + 2(5-6,8)^3 + 3(6-6,8)^3 + 5(7-6,8)^3 + 2(8-6,8)^3 + 1(9-6,8)^3 + 1(10-6,8)^3}{1+2+3+5+2+1+1}}{(1,568)^3} =$$

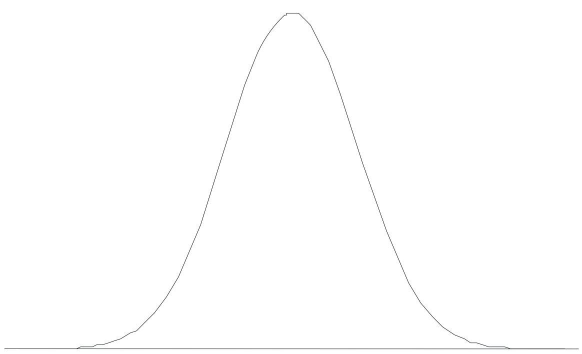
$= \dfrac{0,9715}{3,85} = 0,252 > 0 \Rightarrow$ η κατανομή μας είναι ασύμμετρη δεξιά, δηλ. έχει την ουρά της δεξιά.



Σχ. 3.1



Σχ. 3.2



Σχ. 3.3

Ιδιότητες της στρεβλότητας

- Αν σε κάθε τιμή μιας κατανομής προσθέσουμε τον ίδιο σταθερό αριθμό C, η στρεβλότητα της κατανομής παραμένει αμετάβλητη.
- Αν κάθε τιμή μιας κατανομής την πολλαπλασιάσουμε με τον ίδιο σταθερό αριθμό C, η στρεβλότητα της κατανομής παραμένει αμετάβλητη.

===================== ===================== ====

♦ **Θα λέμε κύρτωση (kurtosis), μιας κατανομής $x_1, x_2, x_3, \ldots, x_k$ στην οποία όμως η παρατήρηση x_1 έχει συχνότητα εμφάνισης f_1, η x_2 συχνότητα εμφάνισης f_2, …, η x_k συχνότητα εμφάνισης f_k , το μέγεθος που ορίζεται από τη σχέση:**

$$KU = \frac{\dfrac{\sum_{i=1}^{k} f_i (X_i - \bar{X})^4}{\sum_{i=1}^{k} f_i}}{s^4} - 3 \qquad \textbf{(9)}$$

Σημείωση: Το –3 στον τύπο 3.13, ίσως ξαφνιάσει τον αναγνώστη, που ίσως πιστέψει ότι πρόκειται για τυπογραφικό λάθος. Όχι δεν πρόκειται για λάθος. Ο τύπος 3.13 είναι ο τύπος που χρησιμοποιεί ο μεγάλος στατιστικός R. Fisher. Μας βολεύει και τον χρησιμοποιούμε κι εμείς. Αυτόν υιοθετεί και το SPSS και το σημαντικότερό του πλεονέκτημα είναι ότι, όπως θα δούμε παρακάτω, οι τιμές που αυτός παράγει, είναι εύκολα ερμηνεύσιμες.

Η κύρτωση είναι ένα μέτρο που μας πληροφορεί για το βαθμό συγκέντρωσης των τιμών γύρω από το ‘κέντρο’ ή το ‘μέσον’ της. Έτσι,

▹ Αν σε μια κατανομή έχουμε KU > 0 ⇒ η κατανομή μας ονομάζεται οξύκυρτη ή λεπτόκυρτη (leptokurtic), δηλ. πάρα πολλές τιμές της έχουν συγκεντρωθεί στο ‘κέντρο’ της (βλ.

Σχ.3.4). Αυτό σημαίνει ότι στην ‘περιφέρεια’ της κατανομής υπάρχει έλλειμμα τιμών.

▹Αν σε μια κατανομή έχουμε KU < 0 ⇒ η κατανομή μας ονομάζεται πλατύκυρτη (platykurtic), δηλ. πολύ λίγες τιμές της υπάρχουν στο κέντρο της, ενώ στην ‘περιφέρειά της, υπάρχουν πάρα πολλές τιμές (βλ. Σχ.3.5).

▹ Αν σε μια κατανομή έχουμε KU = 0 ⇒ η κατανομή μας ονομάζεται μεσόκυρτη (mesokurtic) και δεν παρουσιάζει ούτε μεγάλες ούτε μικρές συγκεντρώσεις στο κέντρο (βλ. Σχ.3.6). Με άλλα λόγια είναι μια κανονική κατανομή, αν βέβαια δεν παρουσιάζει και πρόβλημα συμμετρίας.

Παράδειγμα

Αν έχουμε το ίδιο δείγμα δηλ. το

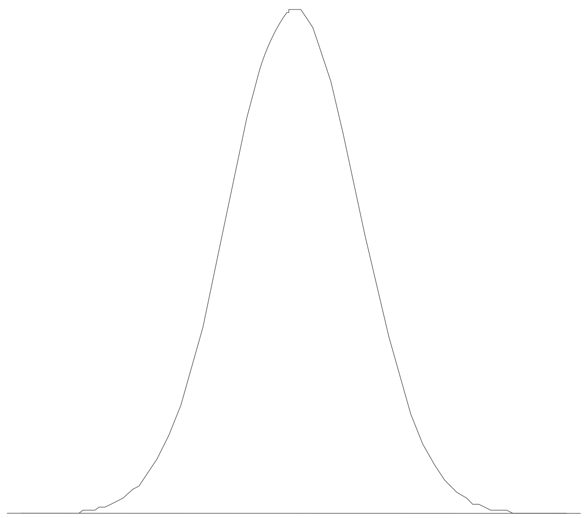
4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

τότε επειδή s=1,568 η σχέση (9) δίνει:

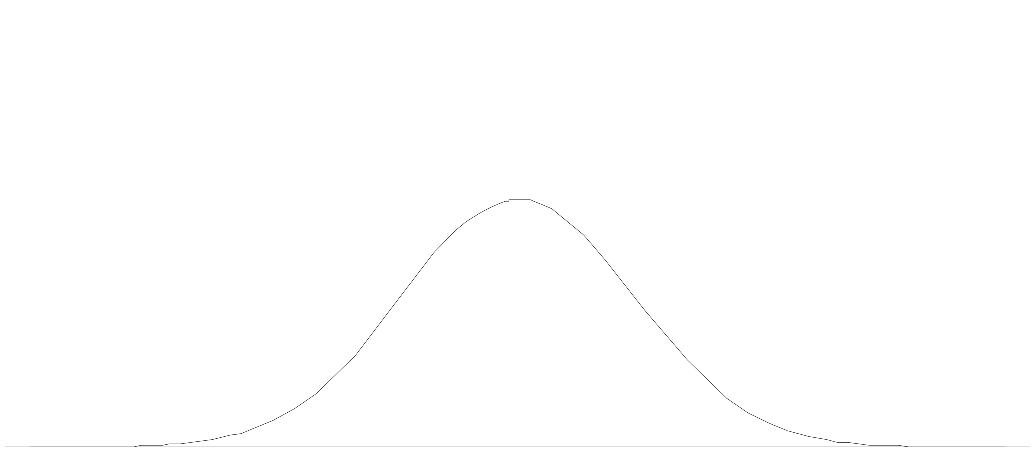
$$KU = \frac{\dfrac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^4}{\sum_{i=1}^{k} f_i}}{s^4} - 3 =$$

$$= \frac{\dfrac{1(4-6,8)^4 + 2(5-6,8)^4 + 3(6-6,8)^4 + 5(7-6,8)^4 + 2(8-6,8)^4 + 1(9-6,8)^4 + 1(10-6,8)^4}{1+2+3+5+2+1+1}}{(1,568)^4} - 3 =$$

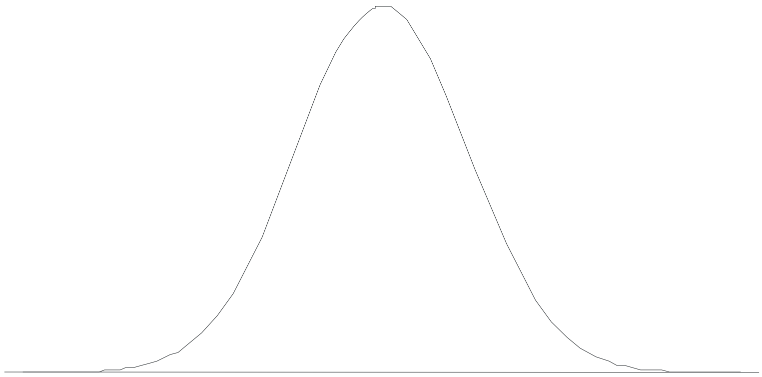
=0,165 > 0 ⇒ η κατανομή μας είναι οξύκυρτη, δηλ. παρουσιάζει τάσεις συγκέντρωσης των τιμών της στο κέντρο της.



Σχ. 3.4



Σχ. 3.5



Σχ. 3.6

Ιδιότητες της κύρτωσης

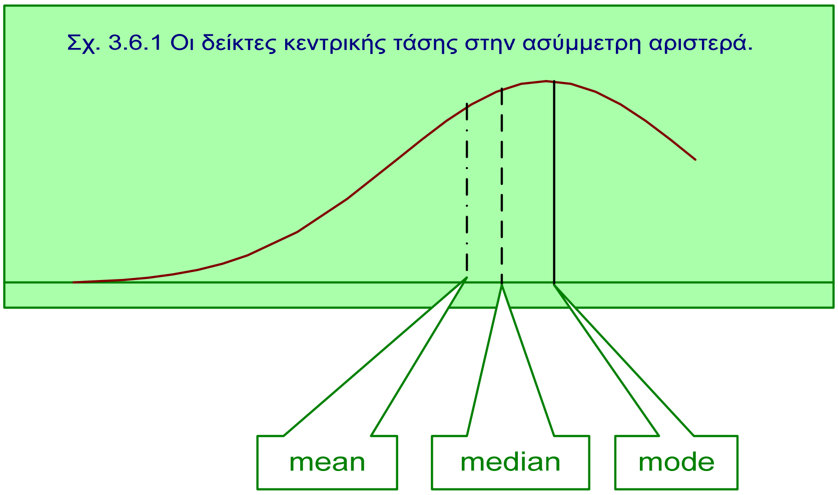
- Αν σε κάθε τιμή μιας κατανομής προσθέσουμε τον ίδιο σταθερό αριθμό C, η κύρτωση της κατανομής παραμένει αμετάβλητη.
- Αν κάθε τιμή μιας κατανομής την πολλαπλασιάσουμε με τον ίδιο σταθερό αριθμό C, η κύρτωση της κατανομής παραμένει αμετάβλητη.

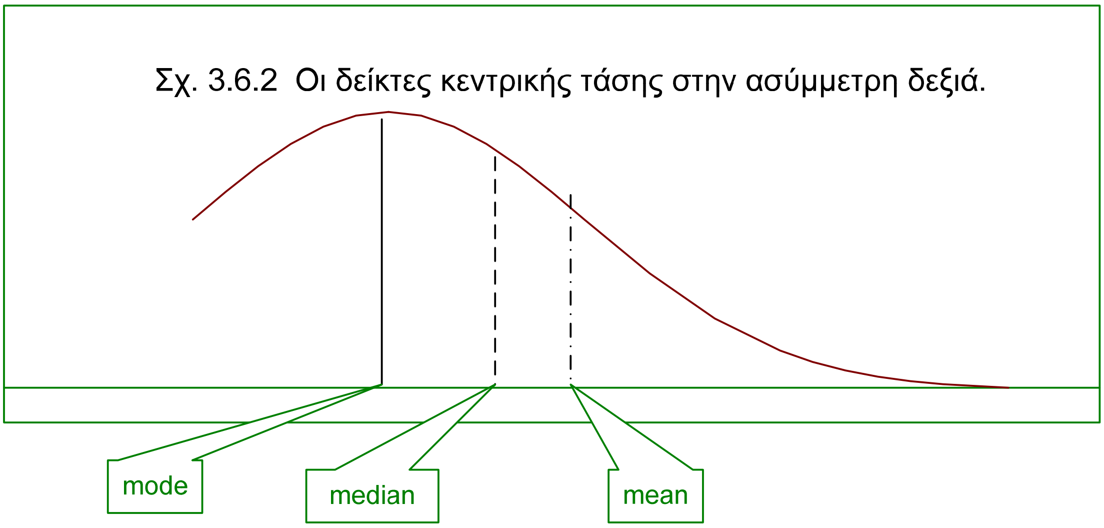
================= ============== ============= ====

============== ==================== ======== =======

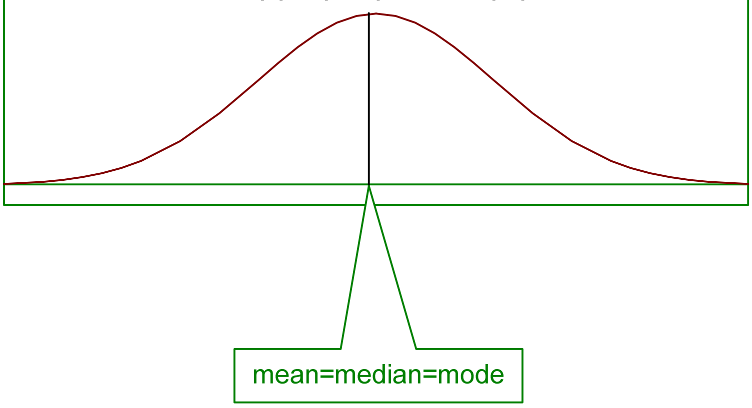
Γραφική Αναπαράσταση των δεικτών κεντρικής τάσης σε ασύμμετρες και συμμετρικές κατανομές

Τα παρακάτω σχήματα, Σχ.3.6.1, Σχ.3.6.2 και Σχ.3.6.3 δείχνουν τις θέσεις της μέσης τιμής, της διαμέσου και της δεσπόζουσας τιμής, όταν η κατανομή μας είναι ασύμμετρη αριστερά (Σχ.3.6.1), όταν η κατανομή μας είναι ασύμμετρη δεξιά (Σχ.3.6.2), και τέλος, όταν η κατανομή μας είναι συμμετρική (Σχ.3.6.3).





Σχ. 3.6.3 Οι δείκτες κεντρικής τάσης στη συμμετρική κατανομή.



Ωστόσο, πως είναι δυνατόν να θυμάται ο αναγνώστης τις θέσεις των 3 δεικτών κεντρικής τάσης, όταν η κατανομή αλλάζει, όταν δηλ. από ασύμμετρη δεξιά, γίνεται ασύμμετρη αριστερά, ή τέλος όταν παύει να είναι ασύμμετρη;

Η γνώση της θεωρίας μας οδηγεί σε δύο ή τρεις ασφαλιστικές δικλίδες:

■ Η μέση τιμή (mean) είναι πάντα προς το μέρος της ουράς της κατανομής.

■ Η διάμεσος έχει πάντα εκατέρωθεν αυτής τους άλλους δύο δείκτες κεντρικής τάσης.

■ Η κατακόρυφος που ξεκινά από το υψηλότερο σημείο της καμπύλης, προφανώς διέρχεται από το σημείο που βρίσκεται η δεσπόζουσα τιμή, πάνω στον οριζόντιο άξονα Χ.

ΔΕΙΚΤΕΣ ΟΜΟΙΟΓΕΝΕΙΑΣ

Η ομοιογένεια των δεδομένων μας μετριέται με το συντελεστή μεταβλητότητας.

♦ **Θα λέμε συντελεστή μεταβλητότητας (coefficient variation) και θα τον συμβολίζουμε με CV, το λόγο της τυπικής απόκλισης προς το δειγματικό μέσο. Συμβολικά:**

$$CV = \frac{s}{\bar{x}} \qquad \textbf{(10)}$$

Ο συντελεστής μεταβλητότητας, προφανώς είναι ανεξάρτητος από μονάδες μέτρησης, γιατί όπως βλέπουμε στον τύπο του και έχουμε αναφέρει στα προηγούμενα, τόσο ο αριθμητής όσο και ο παρονομαστής, μετριώνται με τις ίδιες μονάδες μέτρησης.

Ένα δείγμα θα χαρακτηρίζεται ομοιογενές, αν η τιμή του δείκτη CV δεν ξεπερνά το 10%.

Παράδειγμα

Αν λάβουμε και πάλι το ίδιο δείγμα, δηλ. το:

4, 5, 5, 6, 6, 6, 7, **7**, 7, 7, 7, 8, 8, 9, 10

αφού s=1,568 και $\bar{x}$=6,8 $\Rightarrow CV = \frac{s}{\bar{x}} = \frac{1,568}{6,8} = 0,23 > 0,10 \Rightarrow$ **Δεν υπάρχει ομοιογένεια στο δείγμα μας.**

Τέλος Ενότητας

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης