ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ, Τιμές z, Λοξότητα, Κυρτότητα

Κεφάλαιο 6

Αντώνης Κ. Τραυλός (B.A., M.A., Ph.D.)
Καθηγητής
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
Σχολή Επιστημών Ανθρώπινης Κίνησης και Ποιότητας Ζωής
Τμήμα Οργάνωσης και Διαχείρισης Αθλητισμού

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

- Τα μέτρα διασποράς συνθέτουν με τα μέτρα θέσης το βασικό συνταγολόγιο της περιγραφής στατιστικής, ενώ αποτελούν και δομικά στοιχεία της επαγωγικής στατιστικής.
- Τα μέτρα διασποράς είναι:
  - το **αριθμητικό εύρος (Ε),**
  - το (ημι-) **ενδοτεταρτημοριακό εύρος (Q_I),**
  - **απόκλιση,**
  - η **διακύμανση ή διασπορά (S^2),**
  - η **τυπική απόκλιση (S)**
  - **Τυπικό σφάλμα του μέσου (SE_M)** και
  - ο **δείκτης μεταβλητότητας (CV).**

Αριθμητικό Εύρος (Ε)

- **Εύρος** μιας ομάδας αριθμών Χ είναι η διαφορά μεταξύ του μεγαλύτερου και του μικρότερου αριθμού

$$E = X_{max} - X_{min}$$

- Για **παράδειγμα,** το εύρος των αριθμών 2 & 3 είναι
  - Ε=3-2=1,
- των αριθμών 1, 3 & 7 είναι
  - Ε=7-1=6,
- των αριθμών 1, 3, 7 & 8 είναι
  - Ε=8-1=7 και
- των αριθμών 2, 5, 9,14 & 20 είναι
  - Ε=20-2=18.

3

Αριθμητικό Εύρος (Ε)

- Με τον ίδιο τρόπο μπορεί να υπολογιστεί εύκολα και το εύρος μιας μεγαλύτερης σειράς αριθμών ή γενικά μιας κατανομής αριθμητικών τιμών:
  - **αφαιρούμε από τη μέγιστη τιμή την ελάχιστη**.
- Το αποτέλεσμα δείχνει μόνο σε *τι έκταση* (σε μονάδες) της κλίμακας μέτρησης είναι διασπαρμένες οι τιμές της κατανομής.
  - Δηλαδή εκφράζει το μέγεθος του *διαστήματος μεταξύ των ορίων της κατανομής* και για το λόγο αυτό επηρεάζεται μόνο από τα μεγέθη των δύο οριακών τιμών της, χωρίς όμως να απεικονίζει τη μορφή της διασποράς μέσα στα όρια αυτά.

4

Ενδοτεταρτημοριακό Εύρος (Q_I)

- Το **ενδοτεταρτημοριακό εύρος** είναι το μισό της απόστασης μεταξύ του $1^{ου}$ (Q_1,) και του $3^{ου}$ (Q_3) τεταρτημορίου μιας κατανομής:

$$Q_I = (Q_3 - Q_1) = C_{75} - C_{25}$$

- Το $1^{ο}$ τεταρτημόριο αντιστοιχεί στο $25^{ο}$ εκατοστημόριο ($Q_1 = C_{25}$), ενώ το $3^{ο}$ τεταρτημόριο αντιστοιχεί στο $75^{ο}$ εκατοστημόριο ($Q_3 = C_{75}$).

Παράδειγμα:

- οι **11** τιμές 5, 7, **7**, 9, 9, 12, 12, 12, **14**, 16, 16 έχουν $Q_1 = 7$, δηλαδή την (N + 1) * 0.25 = 12 * 0.25 = $3^{η}$ σε μέγεθος τιμή,
- $Q_3 = 14$, δηλαδή την (N + 1) * 0.75 = 12 * 0.75 = $9^{η}$ σε μέγεθος τιμή και
- ενδοτεταρτημοριακό εύρος $Q_I = Q_3 - Q_2 = 14 - 7 = 7$.

5

Υπολογισμός του Ενδοτεταρτημοριακό Εύρος (Q_I) σε απλή κατανομή

- **Παράδειγμα**:
- η απλή κατανομή των 115 τιμών Βάρους (πίνακας εργασίας) έχει Q_1 = 72 Kg, την (N + 1) * 0.25 = 115 * 0.25 = $29^{η}$ σε μέγεθος τιμή (δηλαδή το Βάρος που αντιστοιχεί στην $9^{η}$ κλάση),
- Q_3 = 81 Kg, την (N + 1) * 0.75 = 115 * 0.75 = $87^{η}$ σε μέγεθος τιμή (δηλαδή η τιμή Χ που αντιστοιχεί στην 18η κλάση), και ενδοτεταρτημοριακό εύρος $Q_I = Q_3 - Q_1 = 81 - 72 = 9$.

ΒΑΡΟΣ

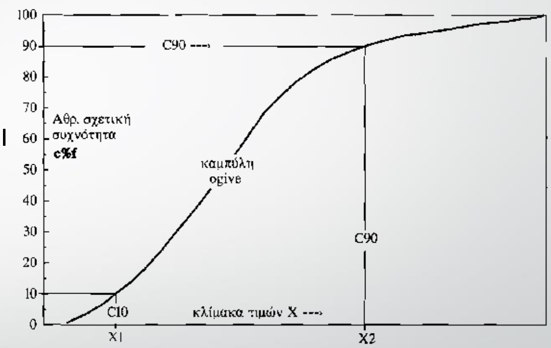
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	64	1	,9	,9	,9
	65	3	2,6	2,6	3,5
	67	1	,9	,9	4,3
	68	3	2,6	2,6	7,0
	69	5	4,3	4,3	11,3
	70	5	4,3	4,3	15,7
	71	6	5,2	5,2	20,9
	72	6	5,2	5,2	26,1
	73	2	1,7	1,7	27,8
	74	8	7,0	7,0	34,8
	75	13	11,3	11,3	46,1
	76	7	6,1	6,1	52,2
	77	9	7,8	7,8	60,0
	78	3	2,6	2,6	62,6
	79	10	8,7	8,7	71,3
	80	3	2,6	2,6	73,9
	81	7	6,1	6,1	80,0
	82	5	4,3	4,3	84,3
	83	4	3,5	3,5	87,8
	84	4	3,5	3,5	91,3
	85	5	4,3	4,3	95,7
	86	1	,9	,9	96,5
	87	1	,9	,9	97,4
	88	2	1,7	1,7	99,1
	90	1	,9	,9	100,0
	Total	115	100,0	100,0	

Ενδο-Εκατοστημοριακό Εύρος (C_I)

- Το ενδο-εκατοστημοριακό εύρος είναι η απόσταση μεταξύ του 10ου (C_{10}) και του 90ού (C_{90}) εκατοστημορίου της κατανομής:

$$C_I = C_{90} - C_{10}$$

- Ο τρόπος εύρεσης του C_I μιας κατανομής είναι όμοιος με αυτόν του Q_I. Χρησιμοποιείται η καμπύλη (ogive) της αθροιστικής σχετικής συχνότητας (c%f) σύμφωνα με τη διαδικασία που απεικονίζεται στην εικόνα 6.2.
- Με αφετηρία τις θέσεις των εκατοστημορίων C_{10} & C_{90} στον κατακόρυφο άξονα χαράζουμε δύο οριζόντιες γραμμές που τέμνουν την καμπύλη σε δύο σημεία. Από τα σημεία αυτά χαράζουμε δύο κατακόρυφες γραμμές που τέμνουν τον (οριζόντιο) άξονα στα σημεία : $X_1 = C_{10}$ & $X_2 = C_{90}$.



Εικ. 6.2 - Γραφικός υπολογισμός του εύρους C_I κατανομής

7

Ενδο-Εκατοστημοριακό Εύρος (C_I)

- Ο δείκτης Q_I αποκλείει τις τιμές X που ανήκουν στα δύο ακραία δεκατημόρια της κατανομής (συνολικά 20% των X).
- Είναι κατάλληλος για μια πρώτη εκτίμηση της διασποράς κατανομών, στις οποίες εμφανίζονται *ακραίες* τιμές.
- Η χρήση του C_I εξυπηρετεί την κατάληξη σε μια πιο ρεαλιστική έκφραση του εύρους της ολικής διασποράς της κατανομής, χωρίς τον επηρεασμό των "ανεπιθύμητων" ακραίων τιμών.

8

Απόκλιση (deviation, d, χ, y)

- Η απόκλιση (deviation) αποτελεί θεμελιώδες μέτρο διασποράς, καθότι αφορά ατομικές τιμές Χ και όχι το σύνολο των Ν τιμών μιας κατανομής Χ. Ουσιαστικά, η απόκλιση αποτελεί το "κύτταρο" της εφαρμοσμένης στατιστικής, αφού πάνω σ' αυτήν την θεμελιώδη έννοια της διασποράς βασίζονται τα μέτρα διασποράς και οι αναλύσεις ANOVA και Παλινδρόμηση.
- Η απόκλιση είναι η διαφορά της τιμής από τον μέσο: d = τιμή - μέσος.

Απόκλιση d = Τιμή - Μέσος

- Παράδειγμα: οι τιμές 2, 5, 9, 14, 20 έχουν μέσο 10 και αποκλίσεις

(2-10) = -8 (5-10) = -5 (9-10) = -1 (14-10) = 4 (20-10) = 10.

9

Απόκλιση (deviation, d, χ, y)

- Η απόκλιση (d) δείχνει πόσο απέχει κάθε τιμή Χ από τον μέσο των τιμών.
- Όταν μεταβλητή είναι η Χ η απόκλιση (d) είναι: $\chi = X - X$
- Όταν μεταβλητή είναι η Υ η απόκλιση (d) είναι: $y = Y - Y$
- **Στο κεφάλαιο 4 είδαμε ότι ο μέσος (Μ) έχει 2 πολύ σπουδαίες ιδιότητες:**
- (α) το άθροισμα των αποκλίσεων (d=X-M) είναι μηδέν (Σd=0) και (β) το άθροισμά των τετραγώνων των αποκλίσεων είναι ελάχιστο (Σd^2=min).

10

Απόκλιση (deviation, d, χ, y)

- Για την ομάδα τιμών 2, 5, 9, 14, 20 έχουμε:

Σd = $\Sigma(X - M) = (2 - 10) + (5 - 10) + (9 - 10) + (14 - 10) + (20 - 10) =$

$= (-8) + (-5) + (-1) + (4) + (10) = 0$

$\mathbf{\Sigma d^2 =} \Sigma(X - M)^2 = (2 - 10)^2 + (5 - 10)^2 + (9 - 10)^2 + (14 - 10)^2 + (20 - 10)^2 =$

$= (-8)^2 + (-5)^2 + (-1)^2 + (4)^2 + (10)^2 =$

$= 64 + 25 + 1 + 16 + 100 = 206 =$ ελάχιστο (min).

- Το $\mathbf{\Sigma d^2}$ ή $\mathbf{\Sigma \chi^2}$ ή $\mathbf{\Sigma y^2}$ λέγεται άθροισμα τετραγώνων (sums of squares, SS) και αποτελεί κομβικό στατιστικό μέγεθος τόσο στην περιγραφική (descriptive) όσο και στην επαγωγική (inferential) στατιστική, όπως π.χ. στην Ανάλυση Διασποράς (ANOVA) και στην Παλινδρόμηση (Regression).

11

Διασπορά ή Διακύμανση (S^2)

- Η **διακύμανση** ή **διασπορά** μιας **σειράς i = 1, 2,..., N αριθμών** X_i δίνεται από τον τύπο:

$$S^2 = \frac{\sum (X - \bar{X})^2}{N} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_N - \bar{X})^2}{N} = \frac{\sum d^2}{N}$$

- και δείχνει τη μέση *τετραγωνική* απόσταση των αριθμών από τον μέσο τους.

12

Διασπορά ή Διακύμανση (S^2)

Παράδειγμα: οι 5 τιμές 2, 5, 9, 14, 20 έχουν μέσο M = 10

άθροισμα τετραγώνων

$$\Sigma d^2 = (2-10)^2 + (5-10)^2 + (9-10)^2 + (14-10)^2 + (20-10)^2$$

$$= (-8)^2 + (-5)^2 + (-1)^2 + (4)^2 + (10)^2$$

$$= (64 + 25 + 1 + 16 + 100) = 206,$$

και διασπορά $S^2 = \Sigma d^2 / N = 206 / 5 = 41.2.$

Τυπική Απόκλιση (standard deviation, S)

$$S = \sqrt{\delta\iota\alpha\sigma\pi o\rho\acute{\alpha}} = \sqrt{\frac{\sum d^2}{N}} = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$$

Αμερόληπτη Διασπορά και Τυπική απόκλιση

- Η αμερόληπτη (unbiased) εκτίμηση τής πληθυσμιακής διασποράς (σ^2) γίνεται με βαθμούς ελευθερίας (degrees of freedom) df = N - 1.
- Η αμερόληπτη διασπορά (s^2) και τυπική απόκλιση (s) είναι

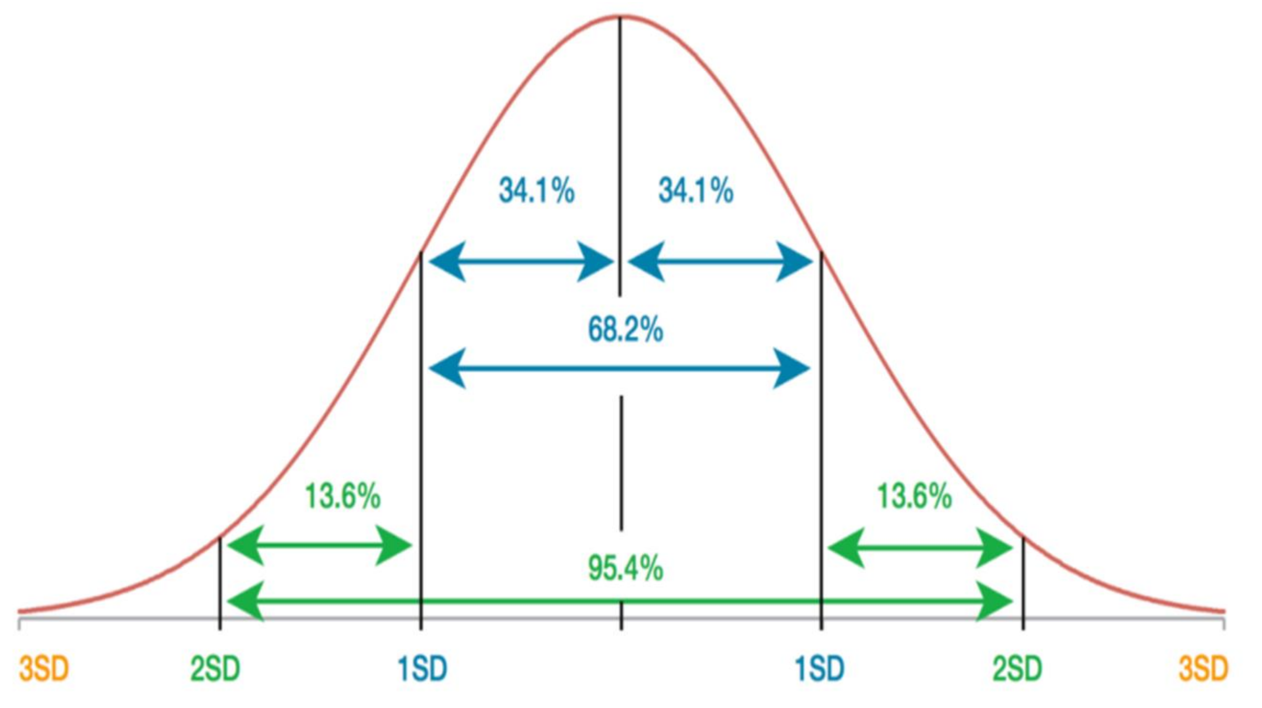
$$s^2 = \frac{SS_x}{df} = \frac{\sum (x-\bar{x})^2}{N-1} = \frac{\sum d^2}{N-1} = \quad \rightarrow \quad s = \sqrt{\delta\iota\alpha\sigma\pi o\rho\acute{\alpha}} = \sqrt{s^2}$$

Παράδειγμα: οι τιμές 2, 5, 9, 14, 20 έχουν μέσο M = 10, $\Sigma d^2 = 206$, **αμερόληπτη διασπορά s^2** $= \Sigma d^2/(N-1) = 206 / 4 = 51.5$.

αμερόληπτη τυπική απόκλιση s $= \sqrt{s^2} = \sqrt{51.5} = 7.1764 \sim 7.18$.

Τα προγράμματα ανάλυσης EXCEL, SPSS, SAS κ.ά. δίνουν την αμερόληπτη διασπορά και τυπική απόκλιση, δηλαδή μετά από τη διόρθωση N-1 = df.

14



Τυπικό σφάλμα του μέσου (standard error of the mean, SE_M)

- Το τυπικό σφάλμα του μέσου (Μ) δείγματος Ν τιμών είναι

$$SE_M = \frac{\text{τυπική απόκλιση}}{\sqrt{\text{μέγεθος δείγματος}}} = \frac{S}{\sqrt{N}}$$

- όπου S η τυπική απόκλιση
- Παράδειγμα: οι 5 τιμές 2, 5, 9, 14, 20 έχουν
  - μέσο M = 10,
  - διασπορά $S^2 = 41.2$,
  - τυπική απόκλιση $S = \sqrt{S^2} = \sqrt{41.2} = 6.42$ και
  - τυπικό σφάλμα του μέσου
  - $SE_M = S/\sqrt{N} = 6.42 / \sqrt{5} = 2.8711$,

16

Συντελεστής Μεταβλητότητας (CV)

- Ο **συντελεστής μεταβλητότητας (coefficient of variation, CV)** μιας σειράς i=1, 2,..., N αριθμών X_i είναι:

$$CV = \frac{\text{Τυπική Απόκλιση}}{\text{Μέσος}} = \frac{SD}{M} = \frac{S}{\overline{X}} * 100$$

- και δείχνει τι ποσοστό του αριθμητικού μέσου αποτελεί η τυπική απόκλιση.
- Για **παράδειγμα,** ο δείκτης μεταβλητότητας των αριθμών 2, 5, 9, 14 & 20, με αριθμητικό μέσο 10 και τυπική απόκλιση 6.42, είναι 64.20. Η τιμή αυτή δείχνει ότι η τυπική απόκλιση αποτελεί το 64.2% του αριθμητικού μέσου: ***σχετικά υψηλή ανομοιογένεια (μεταβλητότητα) τιμών***.

$$\text{συντελεστής μεταβλητότητας:}\ CV = \frac{S}{\overline{X}}100 = \frac{6.42}{10}100 = 64.2\%$$

Συντελεστής Μεταβλητότητας (CV)

- Ο CV αποτελεί *μέτρο σχετικής διασποράς* και δείχνει **την εκατοστιαία (%) έκταση της διασποράς σε σχέση με τον αριθμητικό μέσο**. Αυτή η ιδιότητα τον καθιστά χρήσιμο σε **περιπτώσεις περιγραφικής αξιολόγησης** της διασποράς μιας κατανομής προκειμένου να διαπιστωθεί η ομοιογένεια της.
  - Ακόμα είναι χρήσιμος για συγκρίσεις δύο ή περισσότερων μεταβλητών διαφορετικών μονάδων μέτρησης.
- Επισημαίνεται, όμως, ότι ο CV παρουσιάζει το σοβαρό μειονέκτημα να *υπερτιμά την υπάρχουσα διασπορά όσο ο μέσος τείνει στο μηδέν,* όπως συμβαίνει όταν στην κατανομή υπάρχουν και αρνητικές τιμές.

18

Συντελεστής Μεταβλητότητας (CV) Π.χ.

- Για **παράδειγμα,** οι τιμές -1, 2, -2, 5, -3, με αριθμητικό μέσο 0.2, έχουν διασπορά 8.58 και τυπική απόκλιση 2.93.

διασπορά $S^2 = [\,(-1-0.2)^2+(2-0.2)^2+(-2-0.2)^2+(5-0.2)^2+(-3-0.2)^2\,] \,/\, 5 = 8.58$

τυπική απόκλιση $S = \sqrt{(S^2)} = \sqrt{(8.58)} = \mathbf{2.93}$ **και**

συντελεστή μεταβλητότητας $CV = \frac{2.93}{0.2}100 = 14.65*100 = 1465\%$.

- Η τιμή αυτή δείχνει υπερβολική σχετική διασπορά ίση με 1465%, που φυσικά δεν έχει καμία σχέση με την υπάρχουσα απόλυτη διασπορά, όπως αυτή εκφράζεται αντικειμενικότερα με την τυπική απόκλιση S=2.93.

Ιδιότητες S_x

Πρόσθεση Σταθεράς C στις Αρχικές τιμές X

- Όταν μια σταθερά C προστεθεί στις αρχικές τιμές X μιας κατανομής, η τυπική απόκλιση της κατανομής δεν αλλάζει: $S_{(X+C)} = S_x$.

Πολλαπλασιασμός Αρχικών Τιμών X με Σταθερά C

- Όταν οι αρχικές τιμές X μιας κατανομής πολλαπλασιαστούν με μια σταθερά C, τότε και η τυπική απόκλιση της κατανομής πολλαπλασιάζεται με την απόλυτη τιμή της σταθεράς C : $S_{(X*C)} = |C|\, S_x$.

Τυπικές τιμές (βαθμοί) z (standard score z)

- Ο δείκτης z (z-score) μιας τιμής X είναι ο λόγος της απόκλισης (d, χ) προς την τυπική απόκλιση (S) προς τον μέσο:

$$\text{τυπική τιμή } z = \frac{\text{απόκλιση}}{\text{τυπική απόκλιση}} = \frac{X - \overline{X}}{S} = \frac{d}{S}$$

- & δείχνει τι μέρος της τυπικής απόκλισης είναι η απόκλιση μιας τιμής X από τον αριθμητικό μέσο της κατανομής στην οποία ανήκει.

21

Τυπικές τιμές (βαθμοί) z (standard score z)

Παράδειγμα: οι τιμές 2, 5, 9, 14 & 20, με μέσο M=10 και **τυπική απόκλιση** S=6.42, δίνουν τις τιμές z = d / S :

η τιμή 2 έχει z = (2 - 10) / 6.42 = -8 / 6.42 = -1.2461 ≈ -1.25

η τιμή 5 έχει z = (5 - 10) / 6.42 = -5 / 6.42 = -0.7788 ≈ -0.78

η τιμή 9 έχει z = (9 - 10) / 6.42 = -1 / 6.42 = -0.1558 ≈ -0.16

η τιμή 14 έχει z = (14 - 10) / 6.42 = 4 / 6.42 = 0.6231 ≈ 0.62

η τιμή 20 έχει z = (20 - 10) / 6.42 = 10 / 6.42 = 1.5576 ≈ 1.55

Το **άθροισμα** των τιμών z σε κάθε δείγμα (κατανομή) είναι πάντα **Σz = 0.**

Στο **παράδειγμα Σz** = -1.2461 + -0.7788 + -0.1558 + 0.6231 + 1.5576 = 0.

22

Τυπικές τιμές (βαθμοί) z (standard score z)

- Οι z-τιμές N τιμών X έχουν μέσο **M=0** και τυπική απόκλιση **S=1**.
- Λόγω της ιδιότητας αυτής χρησιμοποιούνται για τη σύγκριση τιμών από διαφορετικές μεταβλητές.
- Π.χ. το Βάρος (kg) και ο χρόνος (min) είναι μη συγκρίσιμα μεγέθη, αλλά αν μετασχηματισθούν σε τιμές z μπορούν να συγκριθούν, αφού έτσι έχουν θα έχουν αναχθεί στην κοινή κλίμακα "τυπικών αποκλίσεων" που είναι ανεξάρτητη των αρχικών κλιμάκων μέτρησης (Kg, min).

23

Λοξότητα (skewness, Sk)

- Η λοξότητα της κατανομής N τιμών X με μέσο M είναι

$$Sk = \frac{N}{(N-1)(N-2)} \sum z^3$$

και έχει τυπικό σφάλμα $SE_{sk} = \sqrt{\frac{6N(N-1)}{(N-2)(N+1)(N+3)}}$

- όπου z = (X - M) / s και s η αμερόληπτη τυπική απόκλιση.
- Ο δείκτης αυτός δείχνει πόσο λοξή (skewed, μη συμμετρική) είναι η κατανομή σε σχέση με την κανονική (normal).

24

Παράδειγμα: οι 5 τιμές 2, 5, 9, 14, 20 έχουν μέσο M = 10,

αμερόληπτη τυπική απόκλιση $s = \surd(\Sigma d^2/(N-1)) = \surd(206/4) = 7.1764$

τιμές z (υπολογισμένες με την αμερόληπτη τυπική απόκλιση s)

για την 2 $z = (2 - 10) / 7.1764 = -8 / 7.1764 = -1.1148$
για την 5 $z = (5 - 10) / 7.1764 = -5 / 7.1764 = -0.6967$
για την 9 $z = (9 - 10) / 7.1764 = -1 / 7.1764 = -0.1394$
για την 14 $z = (14 - 10) / 7.1764 = 4 / 7.1764 = 0.5574$
για την 20 $z = (20 - 10) / 7.1764 = 10 / 7.1764 = 1.3935$, και

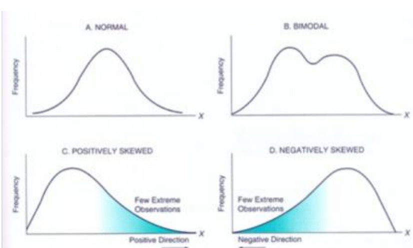
λοξότητα (Sk) (υπολογισμένης και στον πίνακα 6.1)

$$Sk = \frac{5}{(5-1)(5-2)}\left[(-1.1148)^3 + (-0.6967)^3 + (-0.1394)^3 + (0.5574)^3 + (1.3935)^3\right]$$

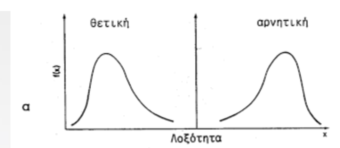
$$= \frac{5}{12}[(-1.3854) + (-0.3382) + (-0.0027) + (0.1732) + (2.7058)] = 0.4803$$

με τυπικό σφάλμα $SE_{Sk} = \sqrt{\frac{6 * 5(5-1)}{(5-2)(5+1)(5+3)}} = 0.9129$

25



Λοξότητα (skewness, Sk)



- Η **θετική λοξότητα** δείχνει κατανομή λοξή αριστερά και η **αρνητική** λοξή δεξιά.
- Το πηλίκο «λοξότητα / τυπικό σφάλμα» = Sk / (SE_{Sk}) = 0.48 / 0.91 = 0.53 είναι μέσα στα αποδεκτά όρια του διαστήματος ± 2 (SE_{Sk}) ή ακόμα *καλύτερα* του διαστήματος ±1 (SE_{Sk}) και δείχνει ότι η κατανομή προσεγγίζει καλά τη συμμετρία (symmetry).
- Τιμές λοξότητας κοντά στο 0 δείχνουν συμμετρία και > 2 τυπικά σφάλματα μεγάλη ασυμμετρία και επομένως μεγάλη απόκλιση από την κανονικότητα (normality).

Κυρτότητα (kurtosis, Ku)

- Η κυρτότητα της κατανομής N τιμών X με μέσο M είναι

$$Ku = \frac{N(N+1)}{(N-1)(N-2)(N-3)}\left(\sum z^4\right) - \frac{3(N-1)^2}{(N-2)(N-3)}$$

και έχει **τυπικό σφάλμα** $SE_{Ku} = \sqrt{\frac{4(N^2-1)(SE\,sk)^2}{(N-3)(N+5)}}$

- όπου z = (X - M)/s & **s** η **αμερόληπτη τυπική απόκλιση**.
- Ο δείκτης κυρτότητας (Ku, κύρτωσης) δείχνει πόσο λεπτόκυρτη (leptokurtic) ή πλατύκυρτη (platykurtic) είναι η κατανομή μιας μεταβλητής.

27

Παράδειγμα: οι 5 τιμές 2, 5, 9, 14, 20 έχουν μέσο M = 10,

αμερόληπτη τυπική απόκλιση s = 7.1764 και αντίστοιχες

τιμές z (υπολογισμένες με την **αμερόληπτη τυπική απόκλιση** στο υποκεφάλαιο 6.9)

-1.1148, -0.6967, -0.1394, 0.5574, 1.3935

κυρτότητα (Ku) (υπολογισμένης και στον πίνακα 6.1)

$$Ku = \frac{5(5+1)}{(5-1)(5-2)(5-3)}\begin{bmatrix}(-1.1148)^4 + (-0.6967)^4 + (-0.1394)^4 \\ +(0.5574)^4 + (1.3935)^4\end{bmatrix} - \frac{3(5-1)^2}{(5-2)(5-3)} =$$

$$\frac{30}{24}[1.5443 + 0.2356 + 0.0004 + 0.0965 + 3.7704] - \frac{36}{6} = -0.9409$$

με **τυπικό σφάλμα** $SE_{Ku} = \sqrt{\frac{4(5^2-1)(0.9129)^2}{(5-3)(5+5)}} = 2$

Κυρτότητα (kurtosis, Ku)

- Η **θετική κυρτότητα** δείχνει λεπτή κατανομή και η **αρνητική** πλατιά κατανομή.
- Το πηλίκο «κυρτότητα / τυπικό σφάλμα» = Ku / (SE_{ku}) = -0.94 / 2 = 0.47 είναι μέσα στα αποδεκτά όρια του διαστήματος ± 2 (SE_{ku}) ή ακόμα καλύτερα του διαστήματος ±1 (SE_{ku}) και δείχνει ότι η κατανομή προσεγγίζει τη μέση κυρτότητα (κανονικότητα).
- Τιμές κυρτότητας κοντά στο 0 δείχνουν μέση κυρτότητα (κανονικότητα) και > 2 σημαντική απόκλιση από την κανονικότητα.

29

Παράδειγμα 6.1 – Μέτρα διασποράς: *ομάδα 5 τιμών X.*

Πίνακας 6.1 - Υπολογισμός δεικτών διασποράς σε ομάδα αριθμών

#	X	$d = X - \overline{X}$	d^2	$d / s = z$	z^3	z^4
1	2	-8	64	-1.11477	-1.38535	1.5443
2	5	-5	25	-0.69673	-0.33822	0.2356
3	9	-1	1	-0.13935	-0.00271	0.0004
4	14	4	16	0.557386	0.17317	0.0965
5	20	10	100	1.393466	2.70576	3.7704
Σ =	50	0	206	0	1.15265	5.6473

= παρατήρηση (observation), μέγεθος δείγματος (sample size) N = 5.

30

Στο επόμενο μάθημα και εργαστήριο θα κάνουμε στην πράξη (με το SPSS) ότι μάθαμε μέχρι τώρα. Μελετήστε πολύ καλά την ύλη (κυρίως από τις διαφάνειες του μαθήματος και το βοηθητικό υλικό που σας έδωσα).

Σας ευχαριστώ για την προσοχή σας

ΑΚΤ

31