

Βιοστατιστική

Περιγραφικά στατιστικά μέτρα

Ασκήσεις 1

Χαράλαμπος Γναρδέλλης

Τμήμα Ζωικής Παραγωγής Αλιείας και Υδατοκαλλιεργειών
Πανεπιστημίου Πατρών

Άσκηση 1

Το μήκος 27 ψαριών κατά την πρώτη τους αναπαραγωγική περίοδο βρέθηκε να είναι:

15,2	19,0	18,7	29,8	21,3	20,9	21,0	21,7	22,3
22,6	24,2	23,1	23,4	26,8	29,6	19,9	17,5	23,8
27,2	16,9	32,1	22,1	21,9	30,6	18,8	31,1	21,9

Να υπολογιστεί η μέση τιμή η διάμεσος και η τυπική απόκλιση του μήκους των ψαριών.

Διατάσσουμε τις τιμές κατ' αύξουσα σειρά

1	2	3	4	5	6	7	8	9	10	11	12	13
15,2	16,9	17,5	18,7	18,8	19,0	19,9	20,9	21,0	21,3	21,7	21,9	21,9

14	15	16	17	18	19	20	21	22	23	24	25	26
22,1	22,3	22,6	23,1	23,4	23,8	24,2	26,8	27,2	29,6	29,8	30,6	31,1

27

32,1

Επικρατούσα τιμή $M_o = 21,9$

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{27+1}{2} = 14$, $M_d = 22,1$

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{15,2 + 16,9 + 17,5 + \dots + 30,6 + 31,1}{27} = 23,09$

Τυπική απόκλιση

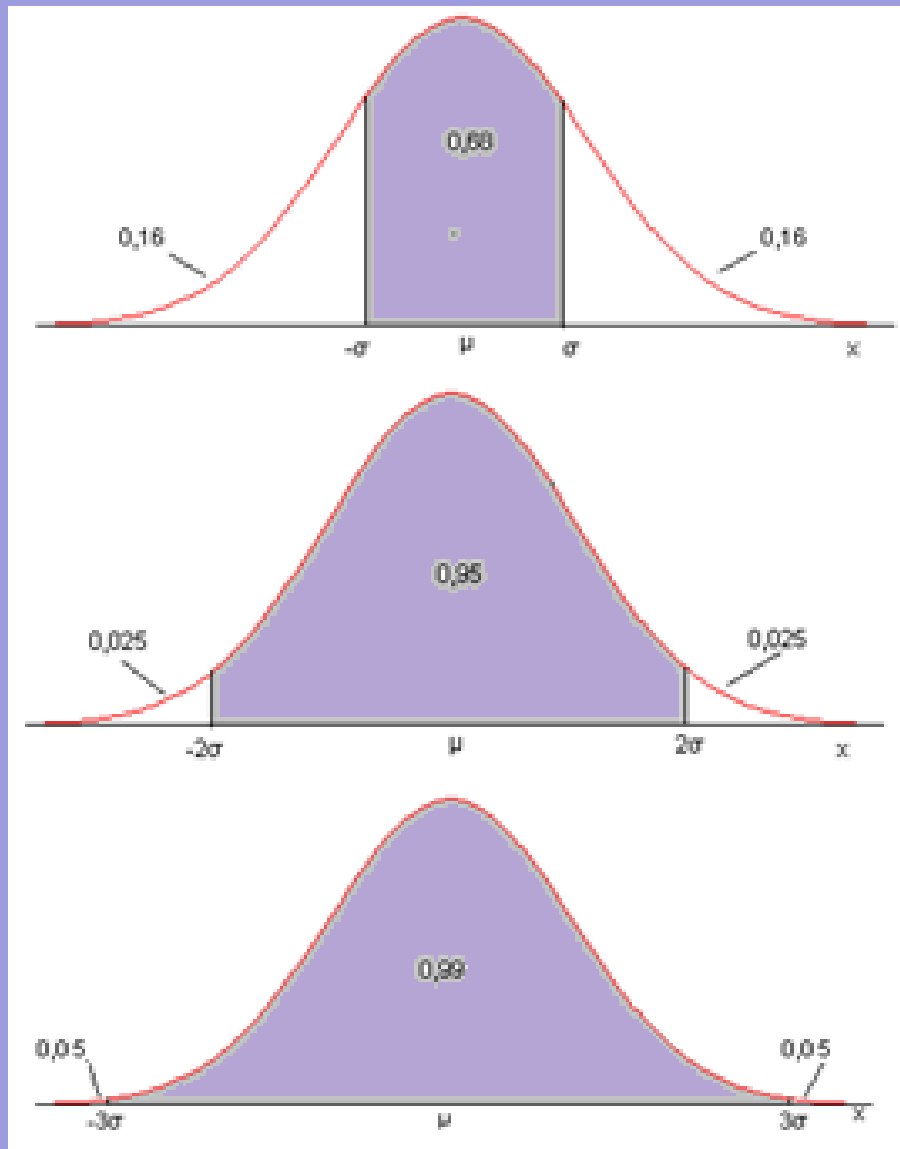
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{(15,2 - 23,09)^2 + (16,9 - 23,09)^2 + (17,5 - 23,09)^2 + \dots + (30,6 - 23,09)^2 + (31,1 - 23,09)^2}{27 - 1}}$$

$$= 4,5$$

• Άσκηση 2

- Αν το πηλίκο νοημοσύνης (IQ) των ατόμων ενός πληθυσμού είναι κανονικά κατανομημένο με μέση τιμή 100 και τυπική απόκλιση 10, ποια αναλογία του πληθυσμού θα έχει πηλίκο νοημοσύνης:
 - Α) μεγαλύτερο από 130
 - Β) μικρότερο από 90
 - Γ) μεταξύ 120 και 130
 - Δ) μεταξύ 90 και 110



- $\mu=100, \sigma=10$
- $x > 130$, $x > 100+30$, $x > \mu+3\sigma$ ποσοστό 0,5%
- $x < 90$, $x < 100-10$, $x < \mu-\sigma$ ποσοστό 16%
- $120 < x < 130$, $100+20 < x < 100+30$, $\mu+2\sigma < x < \mu+3\sigma$
ποσοστό $2,5\% - 0,5\% = 2\%$
- $90 < x < 110$, $100-10 < x < 100+10$, $\mu-\sigma < x < \mu+\sigma$
ποσοστό 68%

Άσκηση 3

Στην Άσκηση 1 να υπολογιστεί ο συντελεστής μεταβλητότητας του μήκους των ψαριών.

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{4,5}{23,09} \times 100 = 19,49$$

Άσκηση 4

Σε μία ομάδα ατόμων βρέθηκε μέση τιμή ουρίας αίματος 40 mg/dl και τυπική απόκλιση 3 mg/dl. Στην ίδια ομάδα ατόμων βρέθηκε μέση τιμή χοληστερόλης του ορού του αίματος 220 mg/dl και τυπική απόκλιση 10 mg/dl. Ποιο από τα δύο χαρακτηριστικά διακυμαίνεται περισσότερο και πόσο περισσότερο;

$$CV_1 = \frac{s_1}{\bar{x}_1} \cdot 100 = \frac{3}{40} \times 100 = 7,5\%$$

$$CV_2 = \frac{s_2}{\bar{x}_2} \cdot 100 = \frac{10}{220} \times 100 = 4,5\%$$

Άρα

$$CV_1 - CV_2 = 7,5 - 4,5 = 3\%$$

Η κατανομή της ουρίας διακυμαίνεται κατά 3% περισσότερο απ' ότι η κατανομή της χοληστερόλης.

Άσκηση 5

Στην Άσκηση 1 να υπολογιστεί ο συντελεστής ασυμμετρίας του Pearson για το μήκος των ψαριών και να οριστεί το είδος της ασυμμετρίας που εμφανίζει η κατανομή του μήκους.

$$s_k = \frac{3(\bar{x} - M_d)}{s} = \frac{3(23,09 - 22,1)}{4,5} = 0,66$$

Η κατανομή είναι θετικά ασύμμετρη.

Άσκηση 6

Προκειμένου να ελεγχθεί η συγκέντρωση του αζώτου στα δένδρα μίας περιοχής, προσδιορίστηκε η μέση περιεκτικότητα σε άζωτο των δένδρων ενός δείγματος 36 περιβολιών. Τα αποτελέσματα συνοψίζονται στον παρακάτω πίνακα :

2,0968	2,8220	2,1739	1,9928	2,2194	3,0926
2,4685	2,5198	2,7983	2,0961	2,9216	2,1997
1,7486	2,7741	2,8241	2,6691	3,0521	2,9263
2,9367	1,9762	2,3821	2,6456	2,7678	1,8488
1,6850	2,7043	2,6814	2,0596	2,3597	2,2783
2,7507	2,4259	2,3936	2,5464	1,8049	1,9629

Αφού στρογγυλοποιηθούν οι τιμές της συγκέντρωσης του αζώτου με ακρίβεια δύο δεκαδικών ψηφίων, να υπολογιστεί η μέση τιμή, η διάμεσος, η τυπική απόκλιση, ο συντελεστής μεταβλητότητας και ο συντελεστής ασυμμετρίας του Pearson.

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{2,10 + 2,82 + 2,17 + \dots + 1,80 + 1,96}{36} = 2,43$

Τυπική απόκλιση

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \sqrt{\frac{(2,10 - 2,43)^2 + (2,82 - 2,43)^2 + (2,17 - 2,43)^2 + \dots + (1,80 - 2,43)^2 + (1,96 - 2,43)^2}{36 - 1}}$$
$$= 0,397$$

Υπολογισμός Διαμέσου

1 1,69	2 1,75	3 1,80	4 1,85	5 1,96	6 1,98	7 1,99	8 2,06	9 2,10
10 2,10	11 2,17	12 2,20	13 2,22	14 2,28	15 2,36	16 2,38	17 2,39	18 2,43
19 2,47	20 2,52	21 2,55	22 2,65	23 2,67	24 2,68	25 2,70	26 2,75	27 2,77
28 2,77	29 2,80	30 2,82	31 2,82	32 2,92	33 2,93	34 2,94	35 3,05	36 3,09

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση

$$\frac{n+1}{2} = \frac{36+1}{2} = 18,5$$

$$M_d = \frac{2,43 + 2,47}{2} = 2,45$$

- Υπολογισμός συντελεστού μεταβλητότητας

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{0,397}{2,43} \times 100 = 16,3\%$$

- Υπολογισμός συντελεστού ασυμμετρίας

$$s_k = \frac{3(\bar{x} - M_d)}{s} = \frac{3(2,43 - 2,45)}{0,397} = -0,15$$

Άσκηση 7

Σε δύο ομάδες εκτρεφόμενων ψαριών μιας ιχθυοτροφικής μονάδας, δόθηκε διαφορετικό, σε περιεκτικότητα πρωτεϊνών, είδος διατροφής, για δύο μήνες. Στην ομάδα A δόθηκε διατροφή υψηλής περιεκτικότητας σε πρωτεΐνες, ενώ στην ομάδα B χαμηλής περιεκτικότητας. Το βάρος (σε γραμμάρια) που κέρδισαν στο διάστημα αυτό τα ψάρια ήταν:

Ομάδα A

126, 108, 98, 133, 138, 105, 136, 110, 115, 135, 94,
122, 130, 125, 133, 150, 145, 201, 133

Ομάδα B

57, 138, 121, 112, 75, 108, 130, 85, 89, 165, 170, 87, 94

Για την ομάδα A να υπολογιστεί η μέση τιμή, η διάμεσος και η επικρατούσα τιμή.

Για την ομάδα B να υπολογιστεί η μέση απόκλιση και η τυπική απόκλιση.

Πρώτη ομάδα

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{126 + 108 + 98 + \dots + 201 + 133}{19} = 128,3$

Διάμεσος

1	2	3	4	5	6	7	8	9	10	11	12	13
94	98	105	108	110	115	122	125	126	130	133	133	133
14	15	16	17	18	19							
135	136	138	145	150	201							

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{19+1}{2} = 10$

$$M_d = 130$$

Επικρατούσα τιμή $M_0 = 133$

Δεύτερη ομάδα

Τυπική απόκλιση

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \sqrt{\frac{(57 - 128,3)^2 + (138 - 128,3)^2 + (121 - 128,3)^2 \dots + (87 - 128,3)^2 + (94 - 128,3)^2}{13 - 1}} = 33,9$$

Μέση απόκλιση

$$s = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{|57 - 128,3| + |138 - 128,3| + |121 - 128,3| \dots + |87 - 128,3| + |94 - 128,3|}{13} =$$
$$\frac{71,3 + 9,7 + 7,3 + \dots + 41,3 + 34,3}{13} = 32,04$$

Άσκηση 8

Για τις χώρες του Οργανισμού Οικονομικής Συνεργασίας και Αναπτύξεως (ΟΟΣΑ) δίδεται το προσδόκιμο επιβίωσης για τα δύο φύλα, η παιδική θνησιμότητα (αριθμός θανάτων παιδιών σε 1000 γεννήσεις) και ο μέσος αριθμός παιδιών ανά οικογένεια.

A) Να υπολογιστεί η διάμεσος του Προσδόκιμου επιβίωσης των ανδρών και των γυναικών.

B) Να υπολογιστεί ο συντελεστής ασυμμετρίας του Pearson της κατανομής του αριθμού των Παιδιών ανά οικογένεια. Να ορίσετε το είδος της ασυμμετρίας που εμφανίζει η κατανομή .

Γ) Να υπολογιστεί ο συντελεστής μεταβλητότητας της Παιδικής θνησιμότητας.

Δ) Ποια μεταβλητή διακυμαίνεται περισσότερο: η Παιδική θνησιμότητα ή ο αριθμός των Παιδιών ανά οικογένεια και πόσο περισσότερο;

Χώρες ΟΟΣΑ	Προσδόκημο επιβίωσης γυναικών	Προσδόκημο επιβίωσης ανδρών	Παιδική θνησιμότητα	Παιδιά ανά οικογένεια
Αυστραλία	85	80	3,7	1,8
Αυστρία	84	79	3,5	1,5
Βέλγιο	84	78	3,9	1,7
Καναδάς	85	79	4,9	1,6
Δανία	81	77	4,4	1,7
Φινλανδία	84	78	2,3	1,7
Γαλλία	85	79	3,9	1,9
Γερμανία	83	78	3,8	1,5
Ελλάδα	83	78	3,8	1,3
Ισλανδία	85	81	2,1	1,8
Ιταλία	85	79	3,3	1,4
Ιρλανδία	83	78	3,6	1,9
Ολλανδία	83	79	3,8	1,7
Ν. Ζηλανδία	83	79	5,4	2,0
Νορβηγία	84	80	2,6	1,7
Πορτογαλία	83	76	3,5	1,3
Ισπανία	85	79	3,3	1,3
Σουηδία	84	80	2,9	1,9
Ελβετία	85	80	4,1	1,5
Βρετανία	83	78	4,3	1,8
Η.Π.Α.	82	77	6,5	1,8

Προσδόκιμο επιβίωσης γυναικών

1	2	3	4	5	6	7	8	9	10	11	12	13
81	82	83	83	83	83	83	83	83	84	84	84	84
14	15	16	17	18	19	20	21					
84	85	85	85	85	85	85	85					

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{21+1}{2} = 11$

$$M_d = 84$$

Προσδόκιμο επιβίωσης ανδρών

1	2	3	4	5	6	7	8	9	10	11	12	13
76	77	77	78	78	78	78	78	78	79	79	79	79
14	15	16	17	18	19	20	21					
79	79	79	80	80	80	80	81					

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{21+1}{2} = 11$

$$M_d = 79$$

Αριθμός παιδιών ανά οικογένεια

1	2	3	4	5	6	7	8	9	10	11	12	13
1,3	1,3	1,3	1,5	1,5	1,5	1,7	1,7	1,7	1,7	1,7	1,7	1,8
14	15	16	17	18	19	20	21					
1,8	1,8	1,8	1,9	1,9	1,9	1,9	2					

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{21+1}{2} = 11$

$$M_d = 1,7$$

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1,3 + 1,3 + 1,3 + \dots + 1,9 + 2}{21} = 1,69$

Αριθμός παιδιών ανά οικογένεια

Τυπική απόκλιση

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \sqrt{\frac{(1,3-1,69)^2 + (1,3-1,69)^2 + (1,3-1,69)^2 \dots + (1,9-1,69)^2 + (2-1,69)^2}{21-1}} = 0,21$$

Συντελεστής ασυμμετρίας του Pearson

$$S_k = \frac{3(\bar{x} - M_d)}{s} = \frac{3(1,69 - 1,7)}{0,21} = -0,14$$

Παιδική θνησιμότητα

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{3,7 + 3,5 + 3,9 + \dots + 4,3 + 6,5}{21} = 3,79$

Τυπική απόκλιση

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \sqrt{\frac{(3,7 - 3,79)^2 + (3,5 - 3,79)^2 + (3,9 - 3,79)^2 + \dots + (4,3 - 3,79)^2 + (6,5 - 3,79)^2}{21 - 1}} = 1,00$$

Συντελεστής μεταβλητότητας

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{1,00}{3,79} \cdot 100 = 26,4\%$$

Παιδική Θνησιμότητα και Αριθμός παιδιών ανά οικογένεια

Παιδική Θνησιμότητα

$$CV_1 = \frac{s}{\bar{x}} \cdot 100 = \frac{1,00}{3,79} \cdot 100 = 26,4\%$$

Αριθμός παιδιών ανά οικογένεια

$$CV_2 = \frac{s}{\bar{x}} \cdot 100 = \frac{0,21}{1,69} \cdot 100 = 12,4\%$$

$$CV_1 - CV_2 = 26,4 - 12,4 = 14\%$$

Άρα η Παιδική Θνησιμότητα διακυμαίνεται κατά 14% περισσότερο από τον Αριθμό παιδιών ανά οικογένεια

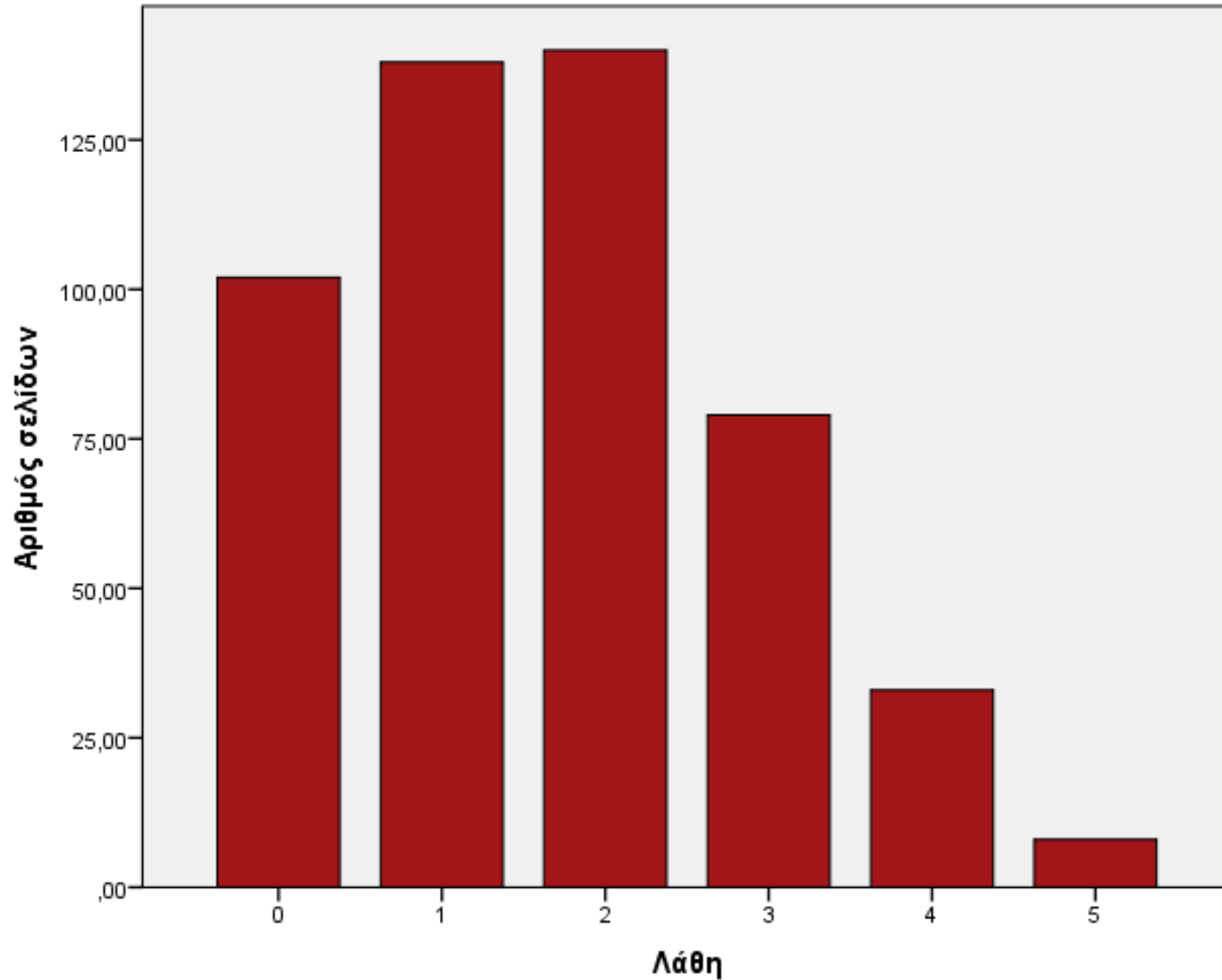
Άσκηση 9

Δίδεται η κατανομή των λαθών ανά σελίδα που βρέθηκαν κατά τη διόρθωση ενός βιβλίου 500 σελίδων:

Λάθη	Σελίδες
0	102
1	138
2	140
3	79
4	33
5	8
Σύνολο	500

Να κατασκευαστεί το ραβδόγραμμα της κατανομής των λαθών και να υπολογιστεί η μέση τιμή και η τυπική απόκλιση της κατανομής αυτής.

Ραβδόγραμμα της κατανομής των λαθών



Μέση τιμή

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{102 \cdot 0 + 138 \cdot 1 + 140 \cdot 2 + 79 \cdot 3 + 33 \cdot 4 + 8 \cdot 5}{500} = \frac{0 + 138 + 280 + 237 + 132 + 40}{500} = 1,65$$

Τυπική απόκλιση

$$\begin{aligned} s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{k=1}^6 f_k (x_k - \bar{x})^2} \\ &= \sqrt{\frac{102 \cdot (0 - 1,65)^2 + 138 \cdot (1 - 1,65)^2 + 140 \cdot (2 - 1,65)^2 + 79 \cdot (3 - 1,65)^2 + 33 \cdot (4 - 1,65)^2 + 8 \cdot (5 - 1,65)^2}{500 - 1}} \\ &= \sqrt{\frac{102 \cdot (-1,65)^2 + 138 \cdot (-0,35)^2 + 140 \cdot (0,35)^2 + 79 \cdot (1,35)^2 + 33 \cdot (2,35)^2 + 8 \cdot (3,35)^2}{500 - 1}} = 0,99 \end{aligned}$$

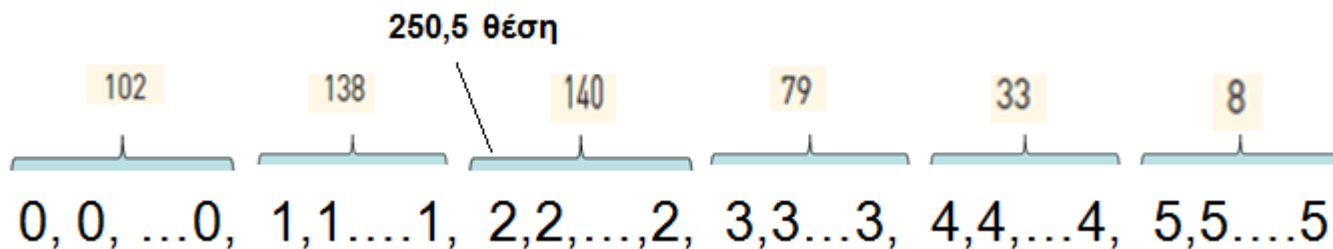
Άσκηση 10

Στα δεδομένα της προηγούμενης άσκησης να υπολογιστεί ο συντελεστής μεταβλητότητας και ο συντελεστής ασυμμετρίας του Pearson.

Συντελεστής μεταβλητότητας

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{0,99}{1,65} \cdot 100 = 60\%$$

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{500+1}{2} = 250,5$



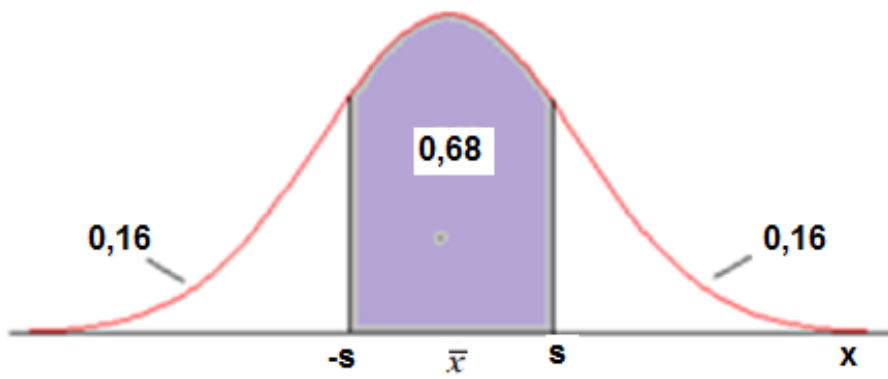
$$M_d = 2$$

Συντελεστής ασυμμετρίας του Pearson

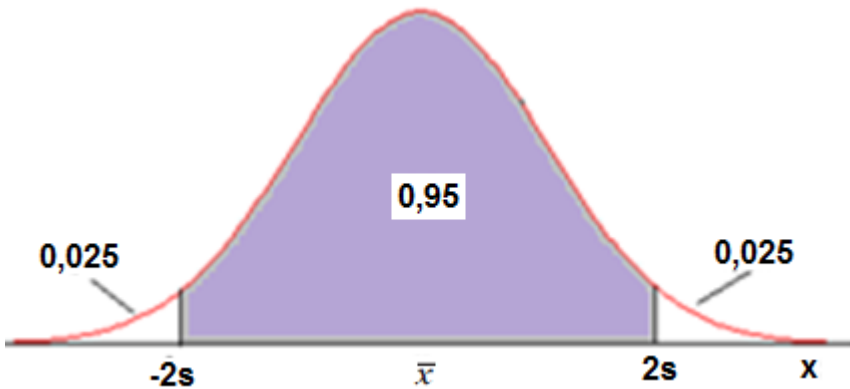
$$S_k = \frac{3(\bar{x} - M_d)}{s} = \frac{3(1,65 - 2)}{0,99} = -1,06$$

Άσκηση 11

Σε μια ομάδα ασθενών με στεφανιαία νόσο βρέθηκε ότι οι τιμές της χοληστερόλης του ορού ακολουθούν κανονική κατανομή. Διαπιστώθηκε ότι 16% των ασθενών της ομάδας είχαν επίπεδα χοληστερόλης κάτω από 182 mg/100ml, ενώ 2,5% είχαν τιμές πάνω από 359 mg/100ml. Ποια είναι η μέση τιμή και ποια η τυπική απόκλιση της κατανομής των τιμών της χοληστερόλης.



$$\bar{x} - s = 182$$



$$\bar{x} + 2s = 359$$

$$\bar{x} - s = 182$$

$$\bar{x} + 2s = 359$$

$$3s = 177 \Leftrightarrow s = \frac{177}{3} = 59$$

$$\bar{x} - 59 = 182$$

$$\bar{x} = 182 + 59 = 241$$

Άρα

$$\bar{x} = 241$$

$$s = 59$$

Άσκηση 12

Να αποδειχθεί ότι

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Δηλαδή το άθροισμα των αποκλίσεων ενός συνόλου μετρήσεων από τη μέση τιμή τους είναι 0

Λύση

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \Leftrightarrow \sum_{i=1}^n x_i = n \cdot \bar{x}$$

Επομένως, το άθροισμα των αποκλίσεων ενός συνόλου μετρήσεων από τη μέση τιμή τους είναι 0

Άσκηση 13

Κατά τον προσδιορισμό της περιεκτικότητας σε οξυγόνο του νερού μιας λίμνης (σε mg/L) έγιναν δειγματοληψίες από 10 σημεία της λίμνης. Οι τιμές που προέκυψαν ήταν :

11,1 10,8 10,3 11,7 10,2 9,9 10,8 12,1 10,7 11,3

Να υπολογιστεί ο συντελεστής ασυμμετρίας του Pearson και να χαρακτηριστεί η κατανομή των τιμών του οξυγόνου ως προς την ασυμμετρία που εμφανίζουν.

Λύση

Οξυγόνο (mg/L)

1	2	3	4	5	6	7	8	9	10
9,9	10,2	10,3	10,7	10,8	10,8	11,1	11,3	11,7	12,1

Διάμεσος, ο αριθμός που καταλαμβάνει τη θέση $\frac{n+1}{2} = \frac{10+1}{2} = 5,5$

$$M_d = 10,8$$

Μέση τιμή $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{9,9 + 10,2 + 10,3 + \dots + 11,7 + 12,1}{10} = 10,89$

Τυπική απόκλιση

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= \sqrt{\frac{(9,9 - 10,89)^2 + (10,2 - 10,89)^2 + (10,3 - 10,89)^2 + \dots + (11,7 - 10,89)^2 + (12,1 - 10,89)^2}{10 - 1}} = 0,682$$

Συντελεστής ασυμμετρίας του Pearson

$$S_k = \frac{3(\bar{x} - M_d)}{s} = \frac{3(10,89 - 10,8)}{0,682} = 0,40$$

Η κατανομή εμφανίζει θετική ασυμμετρία

Άσκηση 14

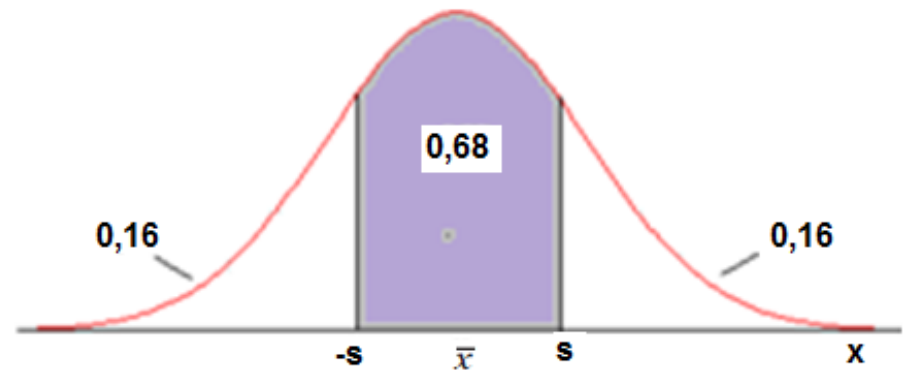
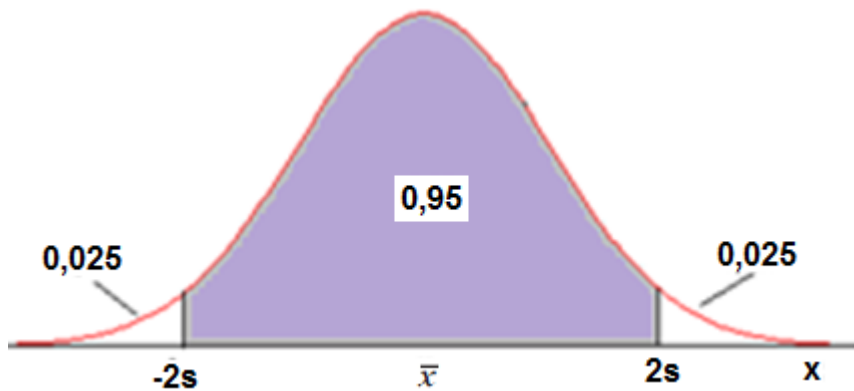
Η μέση τιμή βάρους σε ένα δείγμα ψαριών είναι 256 g και η τυπική απόκλιση 22 g. Στο ίδιο δείγμα, η μέση τιμή μήκους των ψαριών είναι 21 cm και η τυπική απόκλιση 6 cm.

α) Θεωρώντας ότι η κατανομή του βάρους των ψαριών είναι κανονική, προσδιορίστε το ποσοστό των ψαριών με βάρος μικρότερο των 234 γραμμαρίων και μεγαλύτερο των 212 γραμμαρίων

β) Ποιο από τα δύο χαρακτηριστικά (βάρος, μήκος) διακυμαίνεται περισσότερο και πόσο περισσότερο;

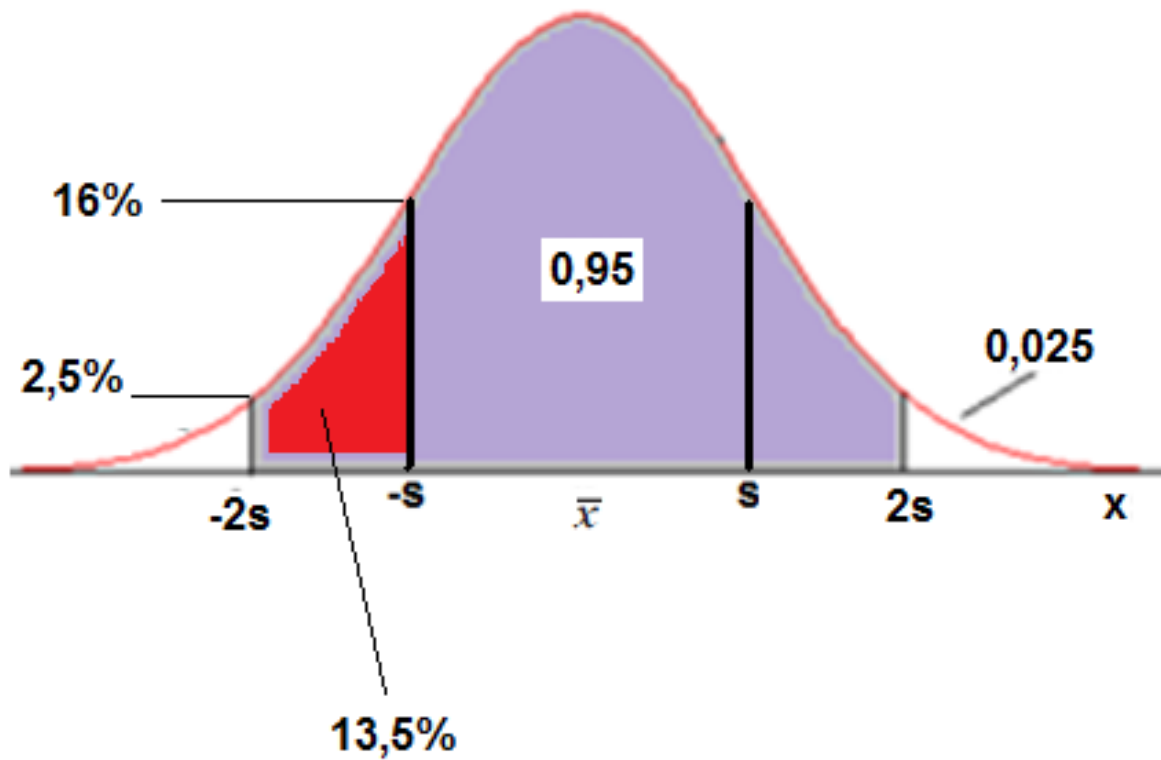
Λύση

- α) $\bar{x} = 256$ $s=22$
- $x > 212$, $x > 256-2 \cdot (22)$, $x > \bar{x} - 2s$
- $x < 234$, $x < 256-22$, $x < \bar{x} - s$
- $212 < x < 234$, $256-2 \cdot (22) < x < 256-22$, $\bar{x} - 2s < x < \bar{x} - s$
16% - 2,5% = 13,5%



$$256 - 2 \cdot (22) < x < 256 - 22$$

$$212 < x < 234$$



Σύγκριση μεταβλητότητας βάρους και μήκους

Βάρος $CV_1 = \frac{s}{\bar{x}} \cdot 100 = \frac{22}{256} \cdot 100 = 8,6\%$

Μήκος $CV_2 = \frac{s}{\bar{x}} \cdot 100 = \frac{6}{21} \cdot 100 = 28,6\%$

$$CV_2 - CV_1 = 28,6 - 8,6 = 20\%$$

Το μήκος διακυμαίνεται περισσότερο, κατά 20%