

25/04/2020

Στατιστική Ι

7η Διάλεξη

(Περιγραφική Στατιστική)

Μέτρα μεταβλητότητας / διασποράς (1)

○ Ποσοτικές μεταβλητές

- Εύρος

$$R = x_{max} - x_{min}$$

- Ενδοτεταρτημοριακό εύρος

$$Q_3 - Q_1$$

Μέτρα μεταβλητότητας / διασποράς (2)

○ Ποσοτικές μεταβλητές

▪ Δειγματική διακύμανση

$$\begin{aligned} \bullet s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ &= \frac{1}{n-1} \sum_{i=1}^k (y_i - \bar{x})^2 n_i = \\ &= \frac{1}{n-1} \left(\sum_{i=1}^k n_i y_i^2 - n\bar{x}^2 \right) \end{aligned}$$

Μέτρα μεταβλητότητας / διασποράς (3)

○ Ποσοτικές μεταβλητές

- Δειγματική τυπική απόκλιση

$$\begin{aligned}
 \bullet s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \\
 &= \sqrt{\frac{1}{n-1} (\sum_{i=1}^n x_i^2 - n\bar{x}^2)} = \\
 &= \sqrt{\frac{1}{n-1} \sum_{i=1}^k (y_i - \bar{x})^2 n_i} = \\
 &= \sqrt{\frac{1}{n-1} (\sum_{i=1}^k n_i y_i^2 - n\bar{x}^2)}
 \end{aligned}$$

Μέτρα μεταβλητότητας / διασποράς (4)

ο Ποσοτικές μεταβλητές

- Συντελεστής μεταβλητότητας

$$CV = \frac{s}{|\bar{x}|} 100\%$$

- Εκφράζει την τυπική απόκλιση των τιμών του δείγματος ως ποσοστό του μέσου τους (μέτρο σχετικής μεταβλητότητας)
- Μπορεί να χρησιμοποιηθεί ως:
 1. Μέτρο σύγκρισης της μεταβλητότητας δύο ή περισσότερων δειγμάτων που έχουν διαφορετικούς μέσους
 2. Μέτρο ομοιογένειας ενός δείγματος (αν $CV < 10\%$) το δείγμα θεωρείται ομοιογενές)

Μέτρα μεταβλητότητας / διασποράς (5)

○ Ποσοτικές μεταβλητές

- Σε όλους τους παραπάνω τύπους, αν έχει γίνει ομαδοποίηση των τιμών του δείγματος σε k κλάσεις τα $y_i, i=1,2,\dots,k$ είναι οι κεντρικές τιμές των κλάσεων

Μέτρα λοξότητας και μέτρα κύρτωσης (1)

ο Ποσοτικές μεταβλητές

- Συντελεστές ασυμμετρίας του Pearson

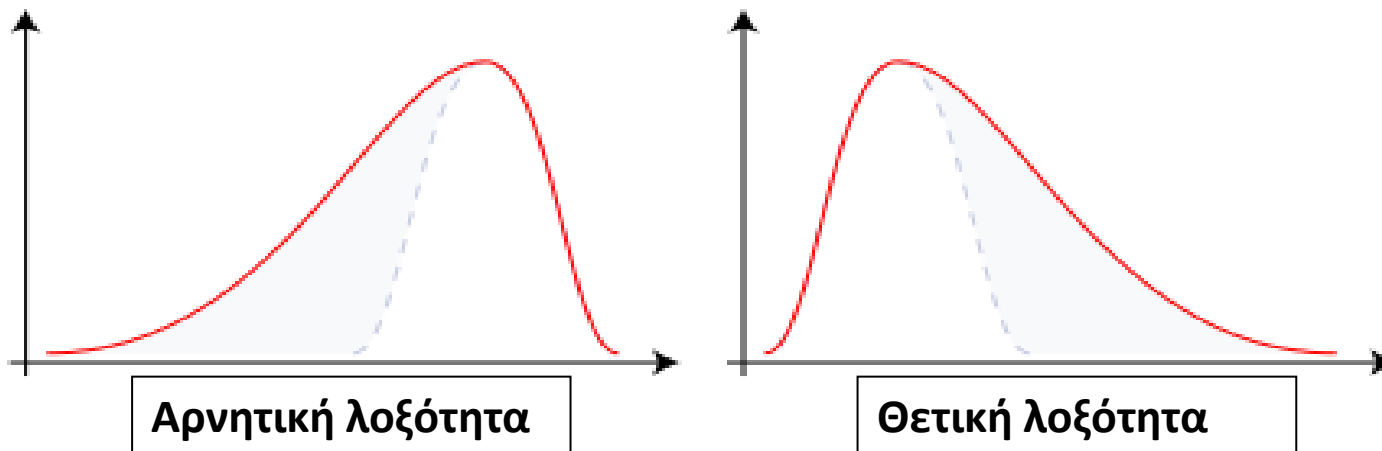
$$1. \gamma_1 = \frac{\bar{x} - M_0}{s}$$

$$2. \gamma_2 = \frac{3(\bar{x} - \delta)}{s}$$

- Αν $\gamma_1 = \gamma_2 = 0$, τότε η κατανομή είναι συμμετρική (αφού $\bar{x} = M_0 = \delta$)

Μέτρα λοξότητας και μέτρα κύρτωσης (2)

- Ποσοτικές μεταβλητές



- Αν η κατανομή μίας μεταβλητής έχει **αρνητική ασυμμετρία (λοξότητα)** τότε
δειγματικός μέσος < διάμεσος < επικρατούσα τιμή
- αν είναι συμμετρική τότε
δειγματικός μέσος = διάμεσος = επικρατούσα τιμή
- ενώ αν έχει **θετική ασυμμετρία (λοξότητα)** τότε
επικρατούσα τιμή < διάμεσος < δειγματικός μέσος

Μέτρα λοξότητας και μέτρα κύρτωσης (3)

○ Ποσοτικές μεταβλητές

- Συντελεστής ασυμμετρίας του Bowley

$$S_A = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

- Αν $S_A = 0$ η κατανομή είναι συμμετρική (αφού $(Q_1 - Q_2) + (Q_3 - Q_2) = 0$)

Μέτρα λοξότητας και μέτρα κύρτωσης (4)

● Ποσοτικές μεταβλητές

- Συντελεστής ασυμμετρίας

$$\alpha_3 = \frac{m_3}{(m_2)^{3/2}}$$

$$m_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad m_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}$$

(κεντρικές ροπές 2^{ης} και 3^{ης} τάξης)

Όταν $\alpha_3 = 0$ έχουμε συμμετρία

Όταν $\alpha_3 > 0$ έχουμε θετική ασυμμετρία

Όταν $\alpha_3 < 0$ έχουμε αρνητική ασυμμετρία

Μέτρα λοξότητας και μέτρα κύρτωσης (5)

ο Ποσοτικές μεταβλητές

- Συντελεστής κύρτωσης του Pearson

$$\beta_2 = \frac{m_4}{(m_2)^2}$$

- $m_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$, $m_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}$

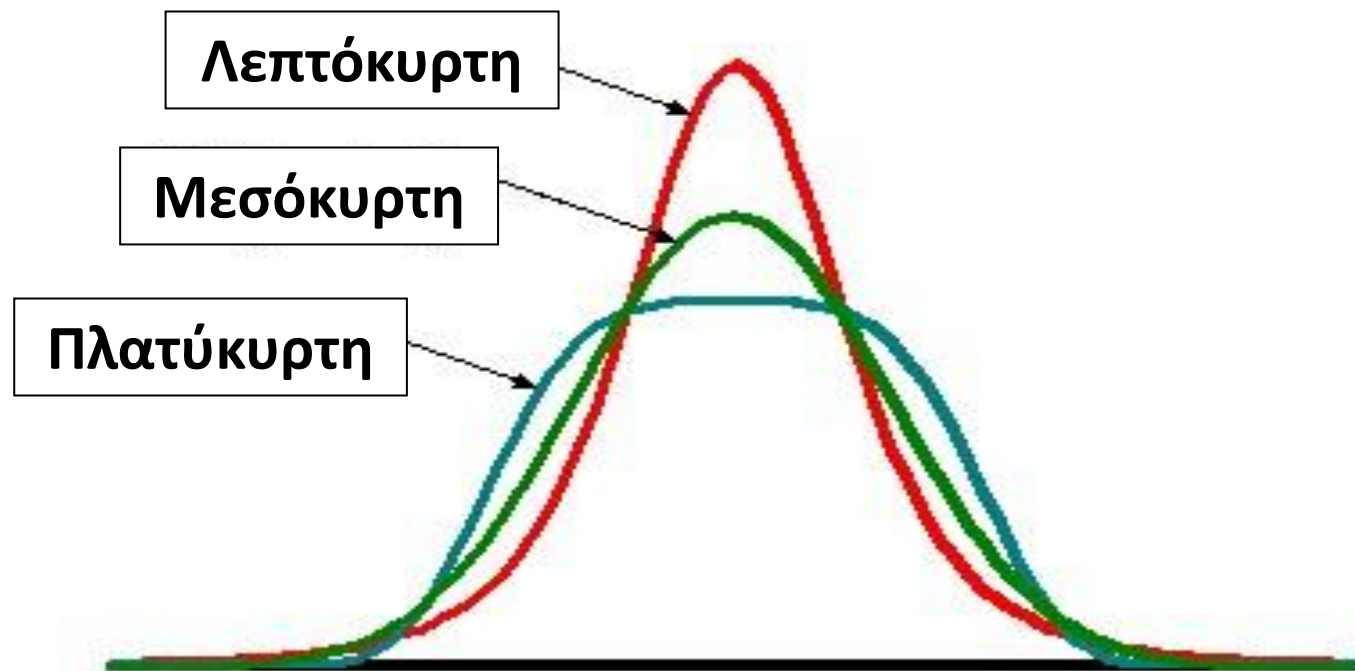
(κεντρικές ροπές 2^{ης} και 4^{ης} τάξης)

- Αν $\beta_2 = 3$ η κατανομή είναι **μεσόκυρτη**
- Αν $\beta_2 < 3$ η κατανομή είναι **πλατύκυρτη**
- Αν $\beta_2 > 3$ η κατανομή είναι **λεπτόκυρτη**

Μέτρα λοξότητας και μέτρα κύρτωσης (6)

ο Ποσοτικές μεταβλητές

- Συντελεστής κύρτωσης του Pearson



Ποιοτικές μεταβλητές

- Από τα αριθμητικά περιγραφικά μέτρα ορίζεται (και έχει νόημα) μόνο η κορυφή της κατανομής

Γραμμικός μετασχηματισμός των παρατηρήσεων / δεδομένων

• Αν $t_i = ax_i + \beta, i = 1, 2, \dots, n$ ΤΟΤΕ

1. $\bar{t} = a\bar{x} + \beta$

2. $s_t^2 = a^2 s_x^2$

3. $s_t = |a|s_x$

4. $\delta_t = a\delta_x + \beta$

5. $M_{0t} = aM_{0x} + \beta$

Εμπειρικός κανόνας κατανομής δειγματος

- Αν η κατανομή του δείγματος προσομοιάζει με μια κανονική κατανομή (έχει κωδωνοειδή μορφή), ΤΟΤΕ
 - Στο διάστημα $(\bar{x} - s, \bar{x} + s)$ βρίσκεται περίπου το 68% των παρατηρήσεων
 - Στο διάστημα $(\bar{x} - 2s, \bar{x} + 2s)$ βρίσκεται περίπου το 95% των παρατηρήσεων
 - Στο διάστημα $(\bar{x} - 3s, \bar{x} + 3s)$ βρίσκονται περίπου όλες οι παρατηρήσεις (πάνω από το 99%)

Η ανισότητα Chebyshev

- Το ποσοστό των τιμών του δείγματος που βρίσκονται στο διάστημα $(\bar{x} - ks, \bar{x} + ks)$ είναι τουλάχιστον $1 - \frac{1}{k^2}$

Παράδειγμα 1^ο (1)

- Έστω τα παρακάτω στοιχεία για το μηνιαίο οικογενειακό εισόδημα από ένα δείγμα του πληθυσμού μιας χώρας

y_i	n_i	f_i	N_i	F_i
1000	1	0.05	1	0.05
1200	3	0.15	4	0.20
1250	1	0.05	5	0.25
1400	4	0.20	9	0.45
1450	2	0.10	11	0.55
1600	4	0.20	15	0.75
1800	3	0.15	18	0.90
2000	2	0.10	20	1.00
Σύνολο	20	1.00		

Παράδειγμα 1^ο (2)

- Να υπολογιστούν
 1. Δειγματικός μέσος
 2. Δειγματική διακύμανση
 3. Δειγματική τυπική απόκλιση
 4. Συντελεστής μεταβλητότητας του δείγματος
 5. Κορυφή του δείγματος
 6. Διάμεσος του δείγματος
 7. 1^ο και 3^ο τεταρτημόριο του δείγματος
 8. Ενδοτεταρτημοριακό εύρος του δείγματος
- Να κατασκευαστεί το θηκόγραμμα του δείγματος χρησιμοποιώντας για την εύρεση της ανώτερης και κατώτερης οριακής τιμής το ανώτερο και κατώτερο εσωτερικό φράγμα, αντίστοιχα
- Υπάρχουν ακραίες τιμές στο δείγμα;