Στατιστική Επιχειρήσεων Ι

Περιγραφική Στατιστική 2

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

Μέτρα μεταβλητότητας / διασποράς (3)

Ποσοτικές μεταβλητές

- Δειγματική τυπική απόκλιση
  - $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(y_i - \bar{x})^2 n_i} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{k} n_i y_i^2 - n\bar{x}^2\right)}$

Μέτρα μεταβλητότητας / διασποράς (4)

- **Ποσοτικές μεταβλητές**
  - Συντελεστής μεταβλητότητας

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

    - 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, $i=1,2,\ldots,k$ είναι οι κεντρικές τιμές των κλάσεων

Μέτρα λοξότητας και μέτρα κύρτωσης (1)

Ποσοτικές μεταβλητές

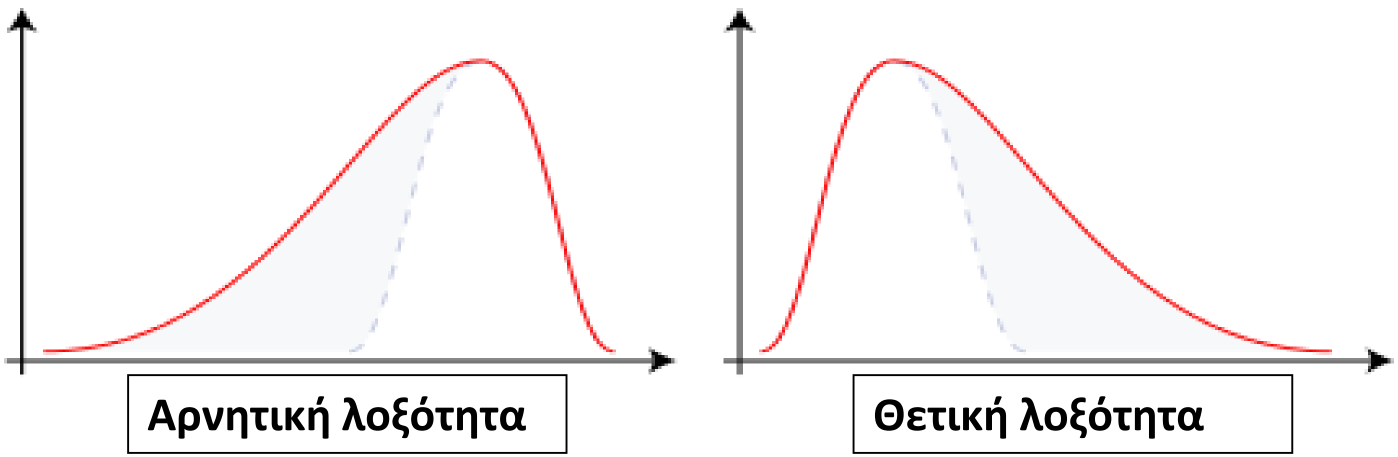
- Συντελεστές ασυμμετρίας του Pearson

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

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

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

(κεντρικές ροπές 2ης και 3ης τάξης)

Όταν $a_3 = 0$ έχουμε **συμμετρία**

Όταν $a_3 > 0$ έχουμε **θετική ασυμμετρία**

Όταν $a_3 < 0$ έχουμε **αρνητική ασυμμετρία**

Μέτρα λοξότητας και μέτρα κύρτωσης (5)

- **Ποσοτικές μεταβλητές**
  - Συντελεστής κύρτωσης του Pearson

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

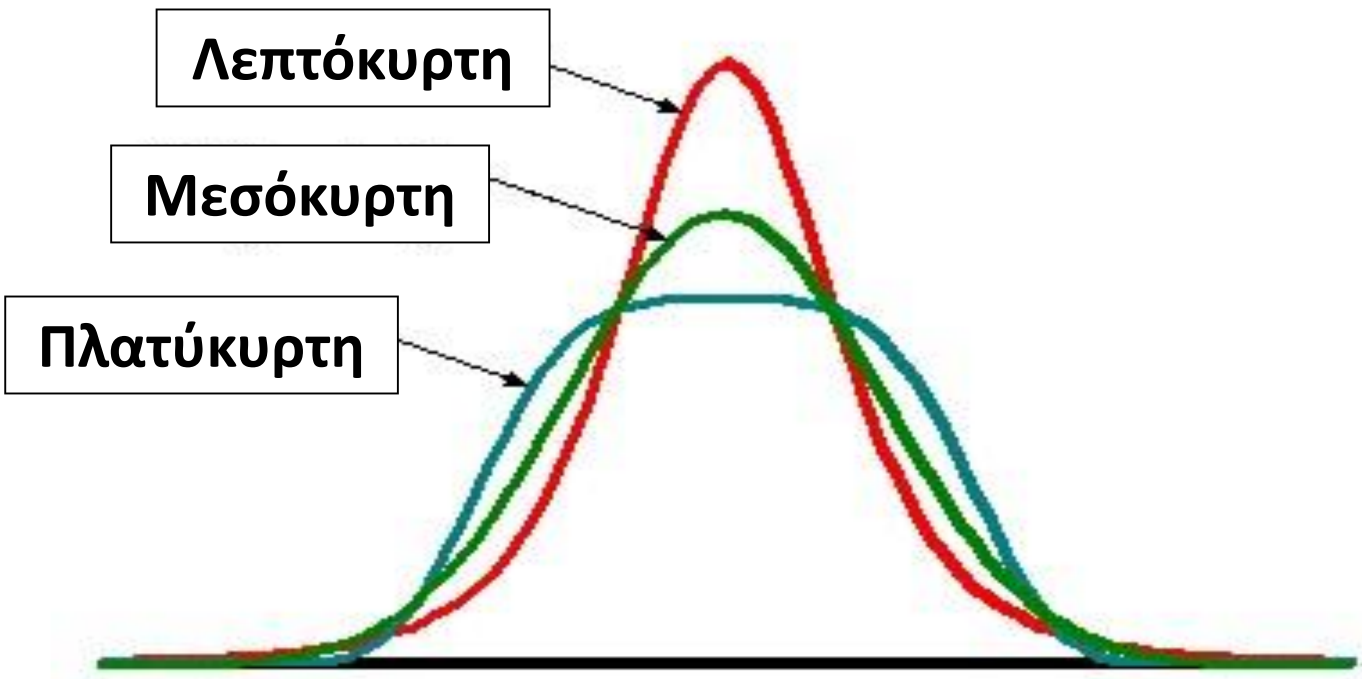
  - $m_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$, $m_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$

(κεντρικές ροπές 2ης και 4ης 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)

- **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 $t_i = ax_i + \beta, i = 1,2, \dots, n$ τότε

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 δείγματος

- Αν η κατανομή του δείγματος προσομοιάζει με μια κανονική κατανομή (έχει κωδωνοειδή μορφή), τότε
  - 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 $1 - \frac{1}{k^2}$