

ΜΕΤΡΑ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ

Μέτρα Περιγραφικής Στατιστικής

- **Πληθυσμιακοί παράμετροι:** τα αριθμητικά μεγέθη που εκφράζουν τις στατιστικές ιδιότητες ενός πληθυσμού (που προσδιορίζουν / περιγράφουν τη φυσιογνωμία και τη δομή του)
- **Στατιστικά περιγραφικά μέτρα:** τα αριθμητικά μεγέθη που αναφέρονται στις στατιστικές ιδιότητες ενός δείγματος του πληθυσμού

Συνθήκες του Yule για τα στατιστικά μέτρα

- Η τιμή του πρέπει να υπολογίζεται με αντικειμενικό τρόπο και με τη συμμετοχή όλων των παρατηρήσεων
- Το υπολογιζόμενο μέτρο να είναι όσο το δυνατόν λιγότερο ευαίσθητο σε κυμάνσεις της δειγματοληψίας
- Η τιμή του να υπολογίζεται με τον κατά τον δυνατόν απλούστερο τρόπο και να έχει συγκεκριμένη σημασία ώστε να είναι κατανοητή και από μη εξειδικευμένα άτομα

Είδη στατιστικών μέτρων

- **Μέτρα κεντρικής τάσης:** απόκλιση από τη μέση τιμή
- **Μέτρα θέσης:** εντοπισμός της θέσης της κατανομής
- **Μέτρα διασποράς:** διασπορά από τη μέση τιμή
- **Μέτρα ασυμμετρίας:** συμμετρία της κατανομής γύρω από τη μέση τιμή
- **Μέτρα κύρτωσης:** βαθμός κλίσης της κατανομής

Μέτρα κεντρικής τάσης

- Αριθμητικός μέσος
- Γεωμετρικός μέσος
- Αρμονικός μέσος

Αριθμητικός μέσος (1)

➤ Πληθυσμιακός μέσος

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad N: \text{μέγεθος πληθυσμού}$$

➤ Δειγματικός μέσος

$$\mu = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad n: \text{μέγεθος δείγματος}$$

Αριθμητικός μέσος (2)

- Αποτελεί την τιμή την οποία θα είχαν όλα τα δεδομένα αν ήταν ίσα μεταξύ τους
- Στον υπολογισμό συμμετέχουν όλα τα δεδομένα
- Μπορεί να υπολογιστεί με τη γνώση και μόνο του αθροίσματος των τιμών και του πλήθους τους
- Σε κάθε σύνολο δεδομένων ο μέσος είναι ένας και μόνο

Αριθμητικός μέσος (3)

- Επηρεάζεται από έκτροπες παρατηρήσεις
- Δεν είναι δυνατός ο υπολογισμός του σε περίπτωση ποιοτικών δεδομένων
- Έχει τις ίδιες μονάδες μέτρησης με τις τιμές από τις οποίες υπολογίζεται
- Η τιμή του αριθμητικού μέσου βρίσκεται πάντα μεταξύ της ελάχιστης και της μέγιστης τιμής της μεταβλητής

Αριθμητικός μέσος (4)

- Το άθροισμα των αποκλίσεων των τιμών της μεταβλητής από τον αριθμητικό τους μέσο είναι πάντα ίσο με το μηδέν
- Το άθροισμα των τετραγώνων των αποκλίσεων των τιμών της μεταβλητής από μία ποσότητα είναι ελάχιστο όταν οι αποκλίσεις αυτές υπολογίζονται από τον αριθμητικό μέσο των δεδομένων
- Αν η μεταβλητή X έχει μέση τιμή $\bar{x}$ τότε η μεταβλητή $y = ax + b$ έχει μέση τιμή $\bar{y} = a\bar{x} + b$

Αριθμητικός μέσος (5)

➤ Έστω k σύνολα δεδομένων πλήθους $n_1, n_2, \dots, n_k$ και $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ οι αντίστοιχες μέσες τιμές.

Τότε ο αριθμητικός μέσος των τιμών όλων των μετρήσεων είναι ίσος με ($n = n_1 + n_2 + \dots + n_k$)

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n}$$

Σταθμικός αριθμητικός μέσος

Είναι ο αριθμητικός μέσος που υπολογίζεται όχι από το σύνολο των αρχικών δεδομένων αλλά από τα αντίστοιχα ομαδοποιημένα δεδομένα

Υπόθεση: Οι συχνότητες επικεντρώνονται όλες στα κέντρα των αντίστοιχων τάξεων

$$\bar{x} = \frac{\sum_{i=1}^k f_i w_i}{\sum_{i=1}^k f_i} \quad w_i : \text{κεντρική τιμή}$$

Παρατήρηση: Όσο μεγαλύτερο είναι το πλάτος των τάξεων τόσο πιο πολύ απομακρύνεται η υιοθετούμενη υπόθεση από την πραγματικότητα

Γεωμετρικός μέσος (1)

➤ Πληθυσμιακός γεωμετρικός μέσος

$$G = \sqrt[N]{\prod_{i=1}^N x_i} = \sqrt[N]{x_1 x_2 \dots x_N}$$

N: μέγεθος πληθυσμού

➤ Δειγματικός γεωμετρικός μέσος

$$\bar{g} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 x_2 \dots x_n}$$

n: μέγεθος δείγματος

Γεωμετρικός μέσος (2)

- Η τιμή του γεωμετρικού μέσου εκφράζεται σε μονάδες μέτρησης της μεταβλητής
- Δεν μπορεί να υπολογιστεί αν υπάρχει έστω και μία μέτρηση με μηδενική ή αρνητική τιμή
- Ο λογάριθμος του γεωμετρικού μέσου είναι ίσος με τον αριθμητικό μέσο των λογαρίθμων των τιμών των διαθέσιμων παρατηρήσεων

$$\log G = \frac{1}{N} (\log x_1 + \log x_1 + \dots + \log x_N) = \frac{\sum_{i=1}^N \log x_i}{N}$$

Γεωμετρικός μέσος (3)

- Το γινόμενο των πηλίκων $u_i = x_i / G$ είναι ίσο με τη μονάδα ανεξάρτητα από το πλήθος των παρατηρήσεων και τις τιμές των δεδομένων

$$\prod_{i=1}^N u_i = \prod_{i=1}^N \left(\frac{x_i}{G} \right) = 1$$

- Το άθροισμα των τετραγώνων των λογαρίθμων των λόγων u_i γίνεται ελάχιστο μόνο όταν οι λόγοι αυτοί υπολογίζονται με βάση το γεωμετρικό μέσο

$$\sum_{i=1}^N \log \left(\frac{x_i}{a} \right)^2 \rightarrow \min \text{imum} \Leftrightarrow a = G$$

Σταθμικός γεωμετρικός μέσος

- Σε αναλογία με τον σταθμικό αριθμητικό μέσο, ο σταθμικός γεωμετρικός μέσος ενός συνόλου n παρατηρήσεων ομαδοποιημένων σε μία κατανομή συχνοτήτων με k τάξεις δίδεται από τον ακόλουθο τύπο

$$\bar{g} = \sqrt[n]{\frac{w_1 w_1 \dots w_1}{f_1} \cdot \frac{w_2 w_2 \dots w_2}{f_2} \dots \frac{w_k w_k \dots w_k}{f_k}}$$

- Ισοδύναμα

$$\log(\bar{g}) = \frac{\sum_{i=1}^k f_i \log w_i}{\sum_{i=1}^k f_i}$$

Αρμονικός μέσος (1)

- Πληθυσμιακός αρμονικός μέσος

$$\frac{1}{H} = \frac{\sum_{i=1}^N 1/x_i}{N} \Rightarrow H = \frac{N}{\sum_{i=1}^N 1/x_i}$$

N: μέγεθος πληθυσμού

- Δειγματικός αρμονικός μέσος

$$\frac{1}{H} = \frac{\sum_{i=1}^n 1/x_i}{n} \Rightarrow H = \frac{n}{\sum_{i=1}^n 1/x_i}$$

n: μέγεθος δείγματος

Αρμονικός μέσος (2)

- Η τιμή του γεωμετρικού μέσου εκφράζεται σε μονάδες μέτρησης της μεταβλητής
- Δεν μπορεί να υπολογιστεί αν υπάρχει έστω και μία μέτρηση με μηδενική ή αρνητική τιμή
- Η τιμή του αρμονικού μέσου είναι μία και μοναδική

Αρμονικός μέσος (3)

- Το άθροισμα των αποκλίσεων των αντίστροφων τιμών της μεταβλητής από τον αντίστροφο του αρμονικού μέσου είναι πάντα ίσο με μηδέν

$$\sum_{i=1}^N u_i = \sum_{i=1}^N \left(\frac{1}{x_i} - \frac{1}{H} \right) = 0$$

- Το άθροισμα των τετραγώνων των αρμονικών αποκλίσεων γίνεται ελάχιστο όταν οι αποκλίσεις αυτές υπολογίζονται από τον αρμονικό μέσο μέσο των μετρήσεων

$$\sum_{i=1}^N \left(\frac{1}{x_i} - \frac{1}{a} \right)^2 \rightarrow \min \text{imum} \Leftrightarrow a = H$$

Σταθμικός αρμονικός μέσος

- Σε αναλογία με τους προηγούμενους ορισμούς,

$$\frac{1}{\bar{h}} = \frac{f_1 \frac{1}{w_1} + f_2 \frac{1}{w_2} + \dots + f_k \frac{1}{w_k}}{f_1 + f_2 + \dots + f_k}$$

- Ισοδύναμα,

$$\bar{h} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n f_i \cdot 1/w_i} = \frac{n}{\sum_{i=1}^n f_i / w_i}$$

Σύγκριση των τριών μέτρων κεντρικής τάσης

- Μεταξύ των τριών μέτρων κεντρικής τάσης, ο αριθμητικός μέσος επηρεάζεται περισσότερο από τους άλλους δύο από ακραίες τιμές των δεδομένων
- Για το ίδιο σύνολο μετρήσεων η προκύπτουσα τιμή του αριθμητικού μέσου είναι η μεγαλύτερη μεταξύ των τριών και ειδικότερα,

$$\bar{x} > \bar{g} > \bar{h}$$

Παρατηρήσεις για τους Σταθμικούς Μέσους

- Οι τιμές των σταθμικών μέσων (αριθμητικού, γεωμετρικού και αρμονικού) δεν επηρεάζονται αν χρησιμοποιηθούν οι σχετικές αντί των απολύτων συχνотήτων
- Οι τιμές των σταθμικών μέσων επηρεάζονται αν αλλάξει η ομαδοποίηση των δεδομένων, δηλαδή αν μεταβληθεί το πλήθος των τάξεων και συνεπώς το πλάτος και τα κέντρα των τάξεων
- Οι σταθμικοί μέσοι ικανοποιούν τις ιδιότητες και διατηρούν τα χαρακτηριστικά των αντίστοιχων αστάθμητων
- Οι σταθμικοί μέσοι δεν μπορούν να υπολογιστούν στην περίπτωση που η κατανομή είναι ανοικτή