

Συστηματική Δειγματοληψία

Περιγραφή:

Από πληθυσμό μεγέθους N θέλουμε να πάρουμε δείγμα μεγέθους n με **Συστηματική Δειγματοληψία** (ΣΔ).

- Καθορίζεται φυσικός αριθμός k , συνήθως $k = \left\lceil \frac{N}{n} \right\rceil$. Το k ονομάζεται **βήμα** της ΣΔ.
- Επιλέγεται τυχαία αριθμός r , $r \in \{1, 2, \dots, k\}$. Το r ονομάζεται **τυχαία αρχή** της ΣΔ.

Τότε οι μονάδες του δείγματος είναι οι

$$r, r + k, r + 2k, \dots, r + (n - 1)k$$

Παρατήρηση:

- Η τυχαία αρχή r καθορίζει και τις υπόλοιπες μονάδες του δείγματος.
- Η δειγματοληψία αυτή λέγεται και 1-κάθε- k Συστηματική Δειγματοληψία.
- Υπάρχουν k δυνατά (συστηματικά) δείγματα και καθένα έχει πιθανότητα επιλογής $1/k$.

Συστηματική δειγματοληψία

Έστω $N = nk$:

(δηλ. το μέγεθος πληθυσμού N είναι πολλαπλάσιο του μεγέθους του δείγματος n)

π.χ. $N = 16, n = 4$, τότε $k = \left\lceil \frac{16}{4} \right\rceil = 4$

r	δυνατά δείγματα
1	1, 5, 9, 13
2	2, 6, 10, 14
3	3, 7, 11, 15
4	4, 8, 12, 16

Συστηματική δειγματοληψία

Έστω $N \neq nk$:

(δηλ. το μέγεθος πληθυσμού N δεν είναι πολλαπλάσιο του μεγέθους του δείγματος n)

π.χ. $N = 17, n = 4$, τότε $k = \left\lceil \frac{17}{4} \right\rceil = 4$

r	δυνατά δείγματα
1	1, 5, 9, 13
2	2, 6, 10, 14
3	3, 7, 11, 15
4	4, 8, 12, 16

Προσοχή, η 17η μονάδα του πληθυσμού δεν θα συμπεριληφθεί σε δείγμα ποτέ!

Για να ξεπεράσουμε το παραπάνω πρόβλημα, μπορούμε να αυξήσουμε το βήμα σε $k = 5$, οπότε $r \in \{1, 2, 3, 4, 5\}$.

r	δυνατά δείγματα
1	1, 6, 11, 16
2	2, 7, 12, 17
3	3, 8, 13, 18 $\rightarrow 1$
4	4, 9, 14, 19 $\rightarrow 2$
5	5, 10, 15, 20 $\rightarrow 3$

Πέραν της 17ης μονάδα (τελευταίας) του πληθυσμού ξεκινούμε από την αρχή.

Συστηματική δειγματοληψία

Εκτίμηση της μέσης τιμής:

Πρόταση 1

Έστω $N = nk$. Στη ΣΔ ο δειγματικός μέσος $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ είναι ΑΕ του μ .

Απόδειξη.

Πρόταση 2

Έστω $N = nk$. Στη ΣΔ είναι

$$\text{Var}(\bar{X}) = \sigma^2 - \frac{k(n-1)}{N} \sigma_{wsy}^2$$

όπου σ_{wsy}^2 (within systematic) εκφράζει το μέσο όρο των διασπορών από κάθε συστηματικό δείγμα, δηλαδή,

$$\sigma_{wsy}^2 = \frac{1}{k} \sum_{r=1}^k \sigma_{0r}^2 = \frac{1}{k(n-1)} \sum_{r=1}^k \sum_{j=1}^n (U_{r+(j-1)k} - \mu_r)^2,$$

$$\mu_r = \frac{1}{n} \sum_{j=1}^n U_{r+(j-1)k}.$$

Απόδειξη.

Συστηματική δειγματοληψία

Πρόταση 3 (Σύγκριση ΣΔ με ΑΤΔ)

Έστω $N = nk$. Είναι $\text{Var}_{\Sigma\Delta}(\bar{X}) < \text{Var}_{\text{ΑΤΔ}}(\bar{X})$ αν και μόνο αν $\sigma_{wsy}^2 > \sigma_0^2$.

Απόδειξη.

Παρατήρηση:

Από την παραπάνω πρόταση γίνεται φανερό ότι η ΣΔ παρέχει μεγαλύτερη ακρίβεια από την ΑΤΔ εάν το σ_{wsy}^2 είναι «μεγάλο», δηλαδή αρκεί οι διασπορές μέσα στα συστηματικά δείγματα, σ_{0r}^2 , να είναι «μεγάλες». Αυτό σημαίνει ότι οι μονάδες που αποτελούν κάθε συστηματικό δείγμα θα πρέπει να είναι **ανομοιογενείς**.