

Απλή τυχαία δειγματοληψία: Συμβολισμός

Πληθυσμός:

1. N : μέγεθος πληθυσμού
2. $1, 2, \dots, j, \dots, N$: μονάδες (στοιχεία) πληθυσμού
3. U_j : η τιμή του χαρακτηριστικού U για την j μονάδα του πληθυσμού,
 $j = 1, \dots, N$

Απλή τυχαία δειγματοληψία: Συμβολισμός

Πληθυσμός (συνέχεια):

4. παράμετρος πληθυσμού : κάθε συνάρτηση των $U_1, \dots, U_N$

- $\mu = \frac{1}{N} \sum_{j=1}^N U_j$: μέση τιμή του U στον πληθυσμό

- $\tau = \sum_{j=1}^N U_j$: ολική τιμή του U στον πληθυσμό, ισχύει $\tau = N \mu$

- $\sigma^2 = \frac{1}{N} \sum_{j=1}^N (U_j - \mu)^2$ ή

$$\sigma_0^2 = \frac{1}{N-1} \sum_{j=1}^N (U_j - \mu)^2 : \text{διασπορά των τιμών του } U \text{ στον πληθυσμό,}$$

$$\text{ισχύει } \sigma_0^2 = \frac{N}{N-1} \sigma^2$$

- $p = \frac{\text{πλήθος μονάδων του πληθυσμού που έχουν το χαρακτηριστικό } U}{N}$:
ποσοστό του U στον πληθυσμό

Απλή τυχαία δειγματοληψία: Συμβολισμός

Δείγμα:

1. n : μέγεθος δείγματος
2. $1, 2, \dots, i, \dots, n$: μονάδες του δείγματος
3. X_i : η τιμή του χαρακτηριστικού U για την i μονάδα του δείγματος,
 $i = 1, \dots, n$

Η X_i είναι τυχαία μεταβλητή με δυνατές τιμές $U_1, \dots, U_N$ (η τιμή του X_i θα είναι γνωστή μετά τη λήψη του δείγματος)

π.χ. $N = 8$ $\{1, 2, 3, 4, 5, 6, 7, 8\}$ ο πληθυσμός

$n = 4$ με $\{2, 4, 1, 7\}$ το δείγμα,

άρα $X_1 = U_2, X_2 = U_4, X_3 = U_1, X_4 = U_7$

Απλή τυχαία δειγματοληψία: Συμβολισμός

Δείγμα (συνέχεια):

4. συγκεκριμένες στατιστικές συναρτήσεις στο δείγμα

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$: μέση τιμή του U στο δείγμα (δειγματικός μέσος)

- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$: διασπορά των τιμών του U στο δείγμα
(δειγματική διασπορά)

- $\sum_{i=1}^n X_i$: ολική τιμή του U στο δείγμα

- $\hat{p} = \frac{\text{πλήθος μονάδων του δείγματος που έχουν το χαρακτηριστικό } U}{n}$:
ποσοστό του U στο δείγμα

5. $f = \frac{n}{N}$: δειγματοληπτικό κλάσμα

Απλή τυχαία δειγματοληψία

N διακεκριμένες μονάδες του πληθυσμού, επιλέγουμε ως δείγμα n μονάδες χωρίς επανάθεση και χωρίς να μας ενδιαφέρει η σειρά επιλογής, υπάρχουν $\binom{N}{n}$ διαφορετικά δείγματα

Ορισμός:

Απλή τυχαία δειγματοληψία (ΑΤΔ) είναι η δειγματοληπτική μέθοδος στην οποία καθένα από τα $\binom{N}{n}$ δείγματα έχει την ίδια πιθανότητα επιλογής, δηλαδή $1/\binom{N}{n}$.

Ερώτηση:

Πρακτικά, πώς γίνεται η επιλογή;

Απλή τυχαία δειγματοληψία

Πρόταση 1

Σε ΑΤΔ η πιθανότητα η j μονάδα του πληθυσμού να συμπεριληφθεί στο δείγμα είναι ίση προς το δειγματοληπτικό κλάσμα f για κάθε $j = 1, \dots, N$.

(κάθε μονάδα του πληθυσμού έχει την ίδια πιθανότητα να συμπεριληφθεί στο δείγμα)

Απόδειξη.

Απλή τυχαία δειγματοληψία

Πρόταση 2

Σε ΑΤΔ ορίζουμε τις τυχαίες μεταβλητές

$$\delta_j = \begin{cases} 1, & \text{η } j \text{ μονάδα του πληθυσμού συμπεριλαμβάνεται στο δείγμα} \\ 0, & \text{διαφορετικά} \end{cases}$$

Τότε η δ_j είναι Bernoulli με πιθανότητα «επιτυχίας» $P(\delta_j = 1) = f$ και

$$E\delta_j = f,$$

$$\text{Var}\delta_j = f(1 - f)$$

$$\text{Cov}(\delta_{j_1}, \delta_{j_2}) = -f \frac{N - n}{N(N - 1)}, \quad \forall j_1 \neq j_2.$$

(δ_j είναι δείκτρια συνάρτηση της j μονάδας του πληθυσμού, δείχνει αν αυτή συμμετέχει ή όχι στο δείγμα)

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση μέσης τιμής

Εκτίμηση μέσης τιμής: $\mu = \frac{1}{N} \sum_{j=1}^N U_j$

Ως εκτιμητής του μ θα χρησιμοποιηθεί ο δειγματικός μέσος $\bar{X}$.

Πρόταση 3

Σε ΑΤΔ

1. Ο $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ είναι ΑΕ του μ
2. $\text{Var} \bar{X} = \frac{N-n}{N-1} \frac{\sigma^2}{n} = (1-f) \frac{\sigma_0^2}{n}$

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση μέσης τιμής

Παρατηρήσεις:

1. Γνωρίζουμε ότι αν $X_1, \dots, X_n$ είναι ανεξάρτητες και ισόνομες τότε $E\bar{X} = \mu, \text{Var}\bar{X} = \frac{\sigma^2}{n}$
2. Στην ΑΤΔ οι $X_1, \dots, X_n$ είναι ισόνομες αλλά είναι συσχετισμένες, άρα εξαρτημένες.
 $(EX_i = \mu, \text{Var}X_i = \sigma^2, \text{Cov}(X_i, X_j) = -\frac{\sigma^2}{N-1})$
Στην ουσία αυτή η εξάρτηση προκύπτει επειδή η δειγματοληψία γίνεται χωρίς επανάθεση σε πεπερασμένο πληθυσμό.
3. Αν $N \rightarrow \infty$ (δηλ. έχουμε «μεγάλο» πληθυσμό) τότε $\text{Var}\bar{X} \approx \frac{\sigma^2}{n} \approx \frac{\sigma_0^2}{n}$
Οι ποσότητες $\frac{N-n}{N-1}$ και $1 - f$ αναφέρονται ως **διορθωτικοί παράγοντες**, και η διόρθωση που κάνουν είναι λόγω του πεπερασμένου πληθυσμού.
4. Αν $n \ll N$ (δηλ. μέγεθος δείγματος πολύ μικρότερο από το μέγεθος του πληθυσμού) τότε $\frac{n}{N} \approx 0$ οπότε $1 - f \approx 1$.
Πρακτικά, αν $f < 0.05$ τότε θεωρούμε $1 - f \approx 1$.

Απλή τυχαία δειγματοληψία: εκτίμηση διασποράς

Εκτίμηση διασποράς:

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^N (U_j - \mu)^2,$$

$$\sigma_0^2 = \frac{1}{N-1} \sum_{j=1}^N (U_j - \mu)^2$$

Ως εκτιμητής της διασποράς θα χρησιμοποιηθεί η δειγματική διασπορά S^2 .

Πρόταση 4

Σε ΑΤΔ

1. Ο $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ είναι ΑΕ του σ_0^2
2. $\frac{N-1}{N} S^2$ είναι ΑΕ του σ^2

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση διασποράς

Η $\text{Var} \bar{X} = (1 - f) \frac{\sigma_0^2}{n}$ περιέχει την άγνωστη παράμετρο σ_0^2 , οπότε ένας εκτιμητής της είναι $\hat{\text{Var}}(\bar{X}) = (1 - f) \frac{S^2}{n}$.

Πρόταση 5

Σε ΑΤΔ

1. $\hat{\text{Var}}(\bar{X}) = (1 - f) \frac{S^2}{n}$ είναι ΑΕ της $\text{Var} \bar{X}$
2. $\hat{\Sigma}(\bar{X}) = \sqrt{\hat{\text{Var}}(\bar{X})} = \sqrt{(1 - f) \frac{S^2}{n}}$ είναι εκτιμητής του $\Sigma(\bar{X})$ αλλά όχι αμερόληπτος
3. ΑΔΕ με σ.ε. $1 - \alpha$ για το μ είναι $\bar{X} \pm z_{\alpha/2} \sqrt{(1 - f) \frac{S^2}{n}}$

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση ολικής τιμής

Εκτίμηση ολικής τιμής: $\tau = \sum_{j=1}^N U_j = N\mu$

Ως εκτιμητής του τ θα χρησιμοποιηθεί ο $\hat{\tau} = N\bar{X}$.

Πρόταση 6

Σε ΑΤΔ

1. Ο $\hat{\tau} = N\bar{X}$ είναι ΑΕ του τ
2. $\text{Var}\hat{\tau} = N^2(1-f)\frac{\sigma_0^2}{n}$ ($= N^2\text{Var}\bar{X}$)
3. $\text{ΤΣ}(\hat{\tau}) = N\sqrt{(1-f)\frac{\sigma_0^2}{n}}$ ($= N \cdot \text{ΤΣ}(\bar{X})$)
4. $\hat{\text{Var}}\hat{\tau} = N^2(1-f)\frac{S^2}{n}$ είναι ΑΕ της $\text{Var}\hat{\tau}$
5. $\hat{\text{ΤΣ}}(\hat{\tau}) = N\sqrt{(1-f)\frac{S^2}{n}}$ ($= N \cdot \hat{\text{ΤΣ}}(\bar{X})$)
6. ΑΔΕ με σ.ε. $1 - \alpha$ για το τ είναι $\hat{\tau} \pm z_{\alpha/2}\hat{\text{ΤΣ}}(\hat{\tau}) \equiv N\bar{X} \pm z_{\alpha/2}N\hat{\text{ΤΣ}}(\bar{X})$

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση ποσοστού

Εκτίμηση ποσοστού: p

Το p είναι **ειδική περίπτωση** μέσης τιμής.

Ορίζοντας ως

$$U_j = \begin{cases} 1, & \text{η } j \text{ μονάδα του πληθυσμού έχει το χαρακτηριστικό } U, \\ 0, & \text{διαφορετικά,} \end{cases}$$

$$j = 1, \dots, N,$$

τότε

$$\mu = \frac{1}{N} \sum_{j=1}^N U_j = \frac{1}{N} \times \{\text{πλήθος μονάδων πληθυσμού που έχουν το } U\} = p$$

$$\begin{aligned} \sigma^2 &= \frac{1}{N} \sum_{j=1}^N (U_j - \mu)^2 = \frac{1}{N} \left\{ \sum_{j=1}^N U_j^2 - N\mu^2 \right\} = \frac{1}{N} \left\{ \sum_{j=1}^N U_j - N\mu^2 \right\} \\ &= \mu - \mu^2 = p(1 - p) \end{aligned}$$

Απλή τυχαία δειγματοληψία: εκτίμηση ποσοστού

Θεωρούμε δείγμα $X_1, \dots, X_n$ από ΑΤΔ με

$$X_i = \begin{cases} 1, & \text{η } i \text{ μονάδα του δείγματος έχει το χαρακτηριστικό } U, \\ 0, & \text{διαφορετικά,} \end{cases}$$
$$i = 1, \dots, n,$$

τότε

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \times \{\text{πλήθος μονάδων δείγματος που έχουν το } U\} = \hat{p}$$

Απλή τυχαία δειγματοληψία: εκτίμηση ποσοστού

Πρόταση 7

Σε ΑΤΔ

1. Ο $\hat{p}$ είναι ΑΕ του p
2. $\text{Var } \hat{p} = \frac{N-n}{N-1} \frac{p(1-p)}{n} = (1-f) \frac{1}{n} \frac{N}{N-1} p(1-p)$
3. $\text{TS}(\hat{p}) = \sqrt{\frac{N-n}{N-1} \frac{p(1-p)}{n}} = \sqrt{(1-f) \frac{1}{n} \frac{N}{N-1} p(1-p)}$
4. $\hat{\text{Var}} \hat{p} = (1-f) \frac{\hat{p}(1-\hat{p})}{n-1}$ είναι ΑΕ της $\text{Var} \hat{p}$
5. $\hat{\text{TS}}(\hat{p}) = \sqrt{(1-f) \frac{\hat{p}(1-\hat{p})}{n-1}}$
6. ΑΔΕ με σ.ε. $1 - \alpha$ για το p είναι $\hat{p} \pm z_{\alpha/2} \hat{\text{TS}}(\hat{p})$

Απόδειξη.

Απλή τυχαία δειγματοληψία: εκτίμηση ποσοστού

Παρατήρηση:

Για μεγάλο πληθυσμό ή αγνοώντας το διορθωτικό παράγοντα έχουμε

$$1. \text{Var } \hat{p} = \frac{N-n}{N-1} \frac{p(1-p)}{n} \approx \frac{p(1-p)}{n}$$

$$2. \text{TS}(\hat{p}) \approx \sqrt{\frac{p(1-p)}{n}}$$

$$3. \hat{\text{Var}} \hat{p} = (1 - f) \frac{\hat{p}(1-\hat{p})}{n-1} \approx \frac{\hat{p}(1-\hat{p})}{n-1}$$

$$4. \hat{\text{TS}}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$

$$5. \text{ΑΔΕ με σ.ε. } 1 - \alpha \text{ για το } p \text{ είναι } \hat{p} \pm z_{\alpha/2} \hat{\text{TS}}(\hat{p}) \approx \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$$

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Πρόβλημα:

Εύρεση του μεγέθους δείγματος n έτσι ώστε να έχουμε ελεγχόμενα «μικρό» σφάλμα εκτίμησης.

Ερώτημα:

Πώς μεταφράζεται το «μικρό» σφάλμα στην εκτίμηση μιας παραμέτρου θ ;

Αν $\hat{\theta}$ είναι εκτιμητής του θ , θα μπορούσαμε να ελέγξουμε το σφάλμα εκτίμησης ως εξής.

Α' τρόπος: Θεωρώντας $\mathbb{P}(|\hat{\theta} - \theta| \leq v_0) = v_0$, όπου v_0 είναι μία δεδομένη μικρή τιμή, π.χ. 1%

Β' τρόπος: Θεωρώντας $\mathbb{P}(|\hat{\theta} - \theta| \leq b) = 1 - \alpha$, όπου $1 - \alpha$, b είναι δεδομένες τιμές, δηλαδή με πιθανότητα $1 - \alpha$, το σφάλμα στην εκτίμηση να μην ξεπερνά την τιμή b .

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Εκτίμηση μέσης τιμής:

Α' τρόπος: $\text{TS}(\bar{X}) = v_0$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{\sigma_0^2}{v_0^2}$$

Β' τρόπος: $P(|\bar{X} - \mu| \leq b) = 1 - \alpha$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{z_{\alpha/2}^2 \sigma_0^2}{b^2}$$

Σημειώνεται ότι z_α δηλώνει το (άνω) α -ποσοστιαίο σημείο μιας τυπικής κανονικής τυχαίας μεταβλητής, $Z \sim \mathcal{N}(0, 1)$, δηλαδή $P(Z > z_\alpha) = \alpha$.

Σχόλιο (ερμηνεία του n_0).

Αν $N \rightarrow \infty$ (δηλ. «μεγάλος» πληθυσμός) τότε $n \approx n_0$.

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Εκτίμηση ολικής τιμής:

Α' τρόπος: $\mathbb{E}(\hat{\tau}) = v_0$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{N^2 \sigma_0^2}{v_0^2}$$

Β' τρόπος: $P(|\hat{\tau} - \tau| \leq b) = 1 - \alpha$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{N^2 z_{\alpha/2}^2 \sigma_0^2}{b^2}$$

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Εκτίμηση ποσοστού:

Α' τρόπος: $\Sigma(\hat{p}) = v_0$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{\sigma_0^2}{v_0^2} = \frac{N}{N-1} \frac{p(1-p)}{v_0^2}$$

Β' τρόπος: $P(|\hat{p} - p| \leq b) = 1 - \alpha$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{z_{\alpha/2}^2 \sigma_0^2}{b^2} = \frac{N}{N-1} \frac{p(1-p) z_{\alpha/2}^2}{b^2}$$

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Παρατηρήσεις:

1. $n < n_0$
2. n, n_0 είναι **αύξουσες** συναρτήσεις ως προς σ_0^2 , οπότε για «μεγάλη» διασπορά του χαρακτηριστικού U στον πληθυσμό χρειαζόμαστε μεγάλο μέγεθος δείγματος
3. n, n_0 είναι **φθίνουσες** συναρτήσεις ως προς b , οπότε για «μικρό» b χρειαζόμαστε μεγάλο μέγεθος δείγματος
4. n, n_0 είναι **αύξουσες** συναρτήσεις ως προς $z_{\alpha/2}$, οπότε για «μεγάλο» σ.ε. $1 - \alpha$ χρειαζόμαστε μεγάλο μέγεθος δείγματος

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Παρατηρήσεις (συνέχεια):

5. Στην περίπτωση της εκτίμησης ποσοστού, ειδικότερα:

α. για «μεγάλο» N έχουμε

$$n \approx n_0 = \begin{cases} \frac{p(1-p)}{v_0^2}, & \text{για } T\hat{\Sigma}(\hat{p}) = v_0, \\ \frac{p(1-p)z_{\alpha/2}^2}{b^2}, & \text{για } P(|\hat{p} - p| \leq b) = 1 - \alpha. \end{cases}$$

β. επειδή $\max_{0 < p < 1} p(1-p) = 1/4$ έχουμε

$$n < n_0 \leq \begin{cases} \frac{N}{N-1} \frac{1}{4v_0^2} \overset{N \text{ «μεγάλο»}}{\approx} \frac{1}{4v_0^2} = n_0^*, \\ \frac{N}{N-1} \frac{z_{\alpha/2}^2}{4b^2} \overset{N \text{ «μεγάλο»}}{\approx} \frac{z_{\alpha/2}^2}{4b^2} = n_0^*, \end{cases}$$

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Παρατηρήσεις (συνέχεια):

5. Στην περίπτωση της εκτίμησης ποσοστού, ειδικότερα:

γ. Στον ακόλουθο πίνακα ενδεικτικά φαίνεται πως μεταβάλλονται οι τιμές των n_0 και n_0^* για διάφορες τιμές των p και v_0 , α , b .

p	v_0	$b (\approx 2v_0)$	$n_0 = \frac{p(1-p)}{v_0^2}$ ή $n_0 = \frac{4p(1-p)}{b^2}$	$n_0^* = \frac{1}{4v_0^2}$ ή $n_0^* = \frac{1}{b^2}$
0.40	1%	2%	2400	2500
0.40	0.5%	1%	9600	10000
0.25	1%	2%	1875	2500
0.25	0.5%	1%	7500	10000
0.06	0.5%	1%	2256	10000
0.06	0.2%	0.4%	14100	62500
0.03	0.5%	1%	1164	10000
0.03	0.1%	0.2%	29100	250000

Σημείωση. Στον παραπάνω πίνακα θεωρούμε $1 - \alpha = 0.95$, δηλαδή $z_{\alpha/2} = z_{0.025} \approx 1.96 \approx 2 \Rightarrow z_{\alpha/2}^2 \approx 4$, και επειδή το ΑΔΕ με σ.ε. $1 - \alpha$ για το p είναι

$$\hat{p} \pm z_{\alpha/2} \hat{\Sigma}(\hat{p}) = \hat{p} \pm z_{\alpha/2} v_0, \text{ δηλαδή } P(|\hat{p} - p| \leq z_{\alpha/2} v_0) \approx 1 - \alpha,$$

θεωρώντας ως τιμή του $b \approx 2v_0$ ενοποιούνται τα αποτελέσματα των δύο τρόπων στην παραπάνω παρουσίαση.

Απλή τυχαία δειγματοληψία: εύρεση μεγέθους δείγματος

Παρατηρήσεις (συνέχεια):

5. Στην περίπτωση της εκτίμησης ποσοστού, ειδικότερα:

δ. *Πρακτικό συμπέρασμα για την εύρεση του μεγέθους δείγματος στην εκτίμηση ποσοστού:*

Στην πράξη το p αντικαθίσταται στους αντίστοιχους τύπους με κάποια εκτίμησή του, ενώ

εάν δεν γνωρίζουμε κάτι για το p , τότε συνήθως παίρνουμε ως εκτίμησή του το 50%, δηλαδή την τιμή που μεγιστοποιεί το $p(1 - p)$ και κατά συνέπεια δίνει μέγεθος δείγματος μεγαλύτερο του απαιτούμενου (**συντηρητική προσέγγιση**), π.χ. στον προηγούμενο πίνακα, συγκρίνετε πως διαφοροποιούνται οι αντίστοιχες τιμές των n_0 και n_0^* (το n_0^* αντιστοιχεί σε πιο συντηρητική προσέγγιση θεωρώντας στην ουσία ότι $p = 0.5$), ειδικά όσο η τιμή του p απομακρύνεται από την τιμή $p = 0.5$,

εάν είναι γνωστό ότι το p κυμαίνεται σε ένα διάστημα, τότε θα πάρουμε ως εκτίμησή του την τιμή που μεγιστοποιεί το $p(1 - p)$, π.χ. εάν είναι γνωστό ότι το p κυμαίνεται μεταξύ

- 25% και 35%, θα πάρουμε ως εκτίμησή του το 35%,
- 60% και 80% θα πάρουμε ως εκτίμησή του το 60%,
- 40% και 60% θα πάρουμε ως εκτίμησή του το 50%.

Απλή τυχαία δειγματοληψία: εκτίμηση σε υποπληθυσμούς

Έστω ότι ένας πληθυσμός μεγέθους N χωρίζεται σε q υποπληθυσμούς

υποπληθυσμός 1		υποπληθυσμός k		υποπληθυσμός q
μεγέθους N_1	...	μεγέθους N_k	...	μεγέθους N_q
$U_{11} \dots U_{1N_1}$		$U_{k1} \dots U_{kN_k}$		$U_{q1} \dots U_{qN_q}$

όπου U_{kj} είναι η τιμή του χαρακτηριστικού U για την j μονάδα του υποπληθυσμού k .

Πρόβλημα:

Μας ενδιαφέρει να εκτιμήσουμε παραμέτρους του υποπληθυσμού k όταν δεν υπάρχει δειγματοληπτικό πλαίσιο για αυτόν αλλά υπάρχει για ολόκληρο τον πληθυσμό.

Παράδειγμα:

Θέλουμε να εκτιμηθεί η ολική κατανάλωση (σε λίτρα) πετρελαίου θέρμανσης των νοικοκυριών μιας περιοχής που χρησιμοποιούν πετρέλαιο για τη θέρμανσή τους. Επειδή δεν μπορούμε να γνωρίζουμε εκ των προτέρων ποια νοικοκυριά χρησιμοποιούν πετρέλαιο για τη θέρμανσή τους, το απλό τυχαίο δείγμα επιλέγεται από λίστα νοικοκυριών της περιοχής που είναι διαθέσιμη.

Απλή τυχαία δειγματοληψία: εκτίμηση σε υποπληθυσμούς

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} U_{kj}$$

$$\tau_k = \sum_{j=1}^{N_k} U_{kj}$$

$$p_k = \frac{\text{\#μονάδων του υποπληθυσμού } k \text{ με συγκεκριμένη ιδιότητα}}{N_k}$$

$$\pi_k = \text{\#μονάδων του υποπληθυσμού } k \text{ με συγκεκριμένη ιδιότητα}$$

Αντικείμενο μελέτης:

Με βάση απλό τυχαίο δείγμα μεγέθους n από ολόκληρο τον πληθυσμό θέλουμε να εκτιμήσουμε τις παραμέτρους $\mu_k, \tau_k, p_k, \pi_k$ του υποπληθυσμού k .

Απλή τυχαία δειγματοληψία: εκτίμηση σε υποπληθυσμούς

Έστω ΑΤΔ μεγέθους n από ολόκληρο τον πληθυσμό μεγέθους N

$$\begin{array}{ccc} \text{υποπληθυσμός } 1 & \text{υποπληθυσμός } k & \text{υποπληθυσμός } q \\ \text{μεγέθος δείγματος } n_1 & \dots & \text{μεγέθος δείγματος } n_k & \dots & \text{μεγέθος δείγματος } n_q \\ X_{11} \dots X_{1n_1} & & X_{k1} \dots X_{kn_k} & & X_{q1} \dots X_{qn_q} \end{array}$$

όπου X_{ki} είναι η τιμή του χαρακτηριστικού U για την i μονάδα του δείγματος που ανήκει στον υποπληθυσμό k . Τονίζεται ότι μόνο εκ των υστέρων διαπιστώνουμε το μέγεθος δείγματος n_k που έχει επιλεγεί από τον υποπληθυσμό k , και ισχύει $n = \sum_{\ell=1}^q n_{\ell}$.

Επομένως, το μέγεθος δείγματος n_k από τον υποπληθυσμό k είναι τ.μ. και ακολουθεί υπεργεωμετρική κατανομή με συνάρτηση πιθανότητας

$$P(n_k = x) = \frac{\binom{N_k}{x} \binom{N-N_k}{n-x}}{\binom{N}{n}}, \quad \max\{0, n+N_k-N\} \leq x \leq \min\{n, N_k\},$$

με μέση τιμή

$$En_k = n \frac{N_k}{N}.$$

Απλή τυχαία δειγματοληψία: εκτίμηση σε υποπληθυσμούς

Εκτίμηση παραμέτρων:

πaráμετρος

εκτιμητής

μ_k

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ki}$$

τ_k

$$\hat{\tau}_k = N_k \bar{X}_k, \quad \text{εάν } N_k \text{ γνωστό}$$

$$\hat{\tau}_k = \frac{n_k}{n} N \bar{X}_k = \frac{N}{n} \sum_{i=1}^{n_k} X_{ki}, \quad \text{εάν } N_k \text{ άγνωστο}$$

p_k

$$\hat{p}_k = \frac{\text{\#μονάδων του δείγματος από τον υποπληθυσμό } k \text{ με συγκεκριμένη ιδιότητα}}{n_k}$$

π_k

$$\hat{\pi}_k = N_k \hat{p}_k, \quad \text{εάν } N_k \text{ γνωστό}$$

$$\hat{\pi}_k = \frac{n_k}{n} N \hat{p}_k, \quad \text{εάν } N_k \text{ άγνωστο}$$

Πρόταση 8

Οι παραπάνω εκτιμητές είναι ΑΕ των αντίστοιχων παραμέτρων.

Απόδειξη. Επειδή n_k είναι τυχαία μεταβλητή, εφαρμόζουμε την ιδιότητα $E(X) = E[E(X | Y)]$.

Απλή τυχαία δειγματοληψία: εκτίμηση σε υποπληθυσμούς

Παράδειγμα:

Για τον έλεγχο παραγωγής ενός αντικειμένου ελήφθη απλό τυχαίο δείγμα 20 κιβωτίων από ένα σωρό 200 κιβωτίων. Οκτώ κιβώτια που περιέχουν

4, 6, 4, 5, 8, 5, 5, 6

αντικείμενα το καθένα βρέθηκαν να έχουν τουλάχιστον ένα ελαττωματικό αντικείμενο. Να εκτιμηθεί ο ολικός αριθμός αντικειμένων στα κιβώτια που περιέχουν τουλάχιστον ένα ελαττωματικό αντικείμενο.

Λύση. Αποτελεί πρόβλημα εκτίμησης ολικής τιμής σε υποπληθυσμό με ΑΤΔ, όπου ο πληθυσμός είναι τα 200 κιβώτια, με δύο υποπληθυσμούς

- υποπληθυσμός 1: τα κιβώτια που περιέχουν τουλάχιστον ένα ελαττωματικό αντικείμενο, και
- υποπληθυσμός 2: τα κιβώτια που δεν περιέχουν ελαττωματικό αντικείμενο.

Θέλουμε να εκτιμήσουμε την ολική τιμή του χαρακτηριστικού U : αριθμός αντικειμένων στα κιβώτια, στον υποπληθυσμό 1. Συνεπώς (επειδή το μέγεθος υποπληθυσμού N_1 είναι άγνωστο)

$$\hat{\tau}_1 = \frac{N}{n} \sum_{i=1}^{n_1} X_{1i} = \frac{200}{20} \{4 + 6 + 4 + 5 + 8 + 5 + 5 + 6\} = 430.$$