

Στρωματοποιημένα δειγματοληψία

Η ιδέα της στρωματοποίησης (stratification) στη δειγματοληψία:

1. Ο πληθυσμός χωρίζεται σε υποπληθυσμούς βάσει κάποιου κριτηρίου. Οι υποπληθυσμοί (ομάδες) που δημιουργούνται αναφέρονται ως **στρώματα**, ενώ η διαδικασία χωρισμού του πληθυσμού σε στρώματα λέγεται **στρωματοποίηση**.
2. Από κάθε στρώμα επιλέγεται ένα δείγμα μονάδων και τα δείγματα (από στρώμα σε στρώμα) είναι **ανεξάρτητα** μεταξύ τους.
3. Το συνολικό μέγεθος δείγματος είναι το άθροισμα των επιμέρους δειγματικών μεγεθών.

Πλεονεκτήματα:

1. Επιτυγχάνεται εκπροσώπηση στο συνολικό δείγμα όλων των τυχόν διαφορετικών «τάσεων» (κατηγοριών) του πληθυσμού, δηλαδή το δείγμα είναι περισσότερο αντιπροσωπευτικό (με αποτέλεσμα καλύτερη εκτίμηση).
2. Συγκέντρωση πληροφοριών και συνεπώς εκτίμηση (συμπεράσματα) για κάθε στρώμα χωριστά (επιπλέον, φυσικά, του συνολικού πληθυσμού).
3. Πιο εύκολη διεξαγωγή έρευνας σε ορισμένες περιπτώσεις και χαμηλότερο κόστος.

Στρωματοποιημένα δειγματοληψία

Παραδείγματα:

Ανάλογα με τη φύση της εκάστοτε έρευνας, π.χ.

- έρευνα για συμπεριφορά κατοίκων για κοινωνικά ή οικονομικά χαρακτηριστικά,
στρώματα: αστικές, ημιαστικές, αγροτικές περιοχές
- έρευνα για συμπεριφορά κατοίκων για κοινωνικά ή οικονομικά χαρακτηριστικά,
στρώματα: χαμηλού, μεσαίου, υψηλού εισοδήματος
- έρευνα για παιδαγωγικές δραστηριότητες,
στρώματα: δημόσια, ιδιωτικά σχολεία
- έρευνα για παιδαγωγικές δραστηριότητες,
στρώματα: δημοτικό, γυμνάσιο, λύκειο

Στρωματοποιημένη τυχαία δειγματοληψία

Ορισμός:

Στρωματοποιημένη τυχαία δειγματοληψία (ΣΤΔ) είναι η στρωματοποιημένη δειγματοληψία στην οποία από κάθε στρώμα το δείγμα επιλέγεται με ΑΤΔ.

Πληθυσμός:

1. N : μέγεθος πληθυσμού
2. N_ℓ : μέγεθος πληθυσμού στο στρώμα ℓ , $\ell = 1, \dots, L$, όπου L είναι το πλήθος στρωμάτων
ισχύει $N = \sum_{\ell=1}^L N_\ell$
3. $W_\ell = \frac{N_\ell}{N}$: βάρος του στρώματος ℓ στον πληθυσμό
ισχύει $0 < W_\ell < 1$ και $\sum_{\ell=1}^L W_\ell = 1$
4. $U_{\ell j}$: η τιμή του χαρακτηριστικού U για τη j μονάδα του στρώματος ℓ , $j = 1, \dots, N_\ell$, $\ell = 1, \dots, L$
5. μ_ℓ : μέση τιμή του U στο στρώμα ℓ , δηλαδή $\mu_\ell = \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} U_{\ell j}$
 σ_ℓ^2 ή $\sigma_{0\ell}^2$: διασπορά των τιμών του U στο στρώμα ℓ , δηλαδή
 $\sigma_\ell^2 = \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (U_{\ell j} - \mu_\ell)^2$ ή $\sigma_{0\ell}^2 = \frac{1}{N_\ell - 1} \sum_{j=1}^{N_\ell} (U_{\ell j} - \mu_\ell)^2$
 τ_ℓ : ολική τιμή του U στο στρώμα ℓ , δηλαδή $\tau_\ell = \sum_{j=1}^{N_\ell} U_{\ell j} = N_\ell \mu_\ell$
 p_ℓ : ποσοστό του U στο στρώμα ℓ

Δείγμα:

1. n : μέγεθος δείγματος
2. n_ℓ : μέγεθος δείγματος στο στρώμα $\ell, \ell = 1, \dots, L$,
ισχύει $n = \sum_{\ell=1}^L n_\ell$
3. $w_\ell = \frac{n_\ell}{n}$: βάρος του στρώματος ℓ στο δείγμα
ισχύει $0 < w_\ell < 1$ και $\sum_{\ell=1}^L w_\ell = 1$
4. $X_{\ell i}$: η τιμή του χαρακτηριστικού U για τη i μονάδα του δείγματος στο στρώμα $\ell, i = 1, \dots, n_\ell, \ell = 1, \dots, L$
5. $\bar{X}_\ell = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} X_{\ell i}$: δειγματικός μέσος στο στρώμα ℓ
 $S_\ell^2 = \frac{1}{n_\ell - 1} \sum_{i=1}^{n_\ell} (X_{\ell i} - \bar{X}_\ell)^2$: δειγματική διασπορά στο στρώμα ℓ

ΣΤΔ: εκτίμηση παραμέτρων

Πρόταση 1

Σε ΣΤΔ

$$1. \mu = \frac{1}{N} \sum_{\ell=1}^L \sum_{j=1}^{N_{\ell}} U_{\ell j} = \sum_{\ell=1}^L W_{\ell} \mu_{\ell}$$

δηλαδή η μέση τιμή του πληθυσμού είναι κυρτός συνδυασμός των μ_{ℓ} με συντελεστές W_{ℓ}

$$2. \bar{X} = \frac{1}{n} \sum_{\ell=1}^L \sum_{i=1}^{n_{\ell}} X_{\ell i} = \sum_{\ell=1}^L w_{\ell} \bar{X}_{\ell}$$

δηλαδή ο δειγματικός μέσος είναι κυρτός συνδυασμός των $\bar{X}_{\ell}$ με συντελεστές w_{ℓ}

Απόδειξη.

ΣΤΔ: εκτίμηση παραμέτρων

Εκτίμηση μέσης τιμής: $\mu = \sum_{\ell=1}^L W_{\ell} \mu_{\ell}$

Ως εκτιμητής του μ θα χρησιμοποιηθεί ο **στρωματοποιημένος δειγματικός μέσος**

$$\bar{X}_{st} = \sum_{\ell=1}^L W_{\ell} \bar{X}_{\ell}$$

Πρόταση 2

Σε ΣΤΔ

1. Ο $\bar{X}_{st}$ είναι ΑΕ του μ .
2. Ο $\bar{X}$ **δεν** είναι ΑΕ του μ , εκτός εάν $W_{\ell} = w_{\ell}, \forall \ell = 1, \dots, L$.
Στην περίπτωση αυτή έχουμε $\bar{X} \equiv \bar{X}_{st}$.

Απόδειξη.

ΣΤΔ: εκτίμηση παραμέτρων

Ερμηνεία της σχέσης $W_\ell = w_\ell, \forall \ell = 1, \dots, L$:

$$w_\ell = W_\ell, \forall \ell \Leftrightarrow \frac{n_\ell}{n} = \frac{N_\ell}{N}, \forall \ell \Leftrightarrow \frac{n_\ell}{N_\ell} = \frac{n}{N}, \forall \ell \Leftrightarrow f_\ell = f, \forall \ell$$

δηλαδή όλα τα δειγματοληπτικά κλάσματα είναι ίδια.

$$w_\ell = W_\ell, \forall \ell \Leftrightarrow n_\ell = \frac{n}{N} N_\ell, \forall \ell \Leftrightarrow n_\ell \propto N_\ell, \forall \ell$$

δηλαδή αναλογική κατανομή των δειγματικών μεγεθών ανά στρώμα.

ΣΤΔ: εκτίμηση παραμέτρων

Πρόταση 3

Σε ΣΤΔ

1. Ο $\bar{X}_{st}$ είναι ΑΕ του μ .

$$2. \text{Var} \bar{X}_{st} = \sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\sigma_{0\ell}^2}{n_{\ell}}$$

$$3. \text{ΤΣ}(\bar{X}_{st}) = \sqrt{\sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\sigma_{0\ell}^2}{n_{\ell}}}$$

(συνεχίζεται)

ΣΤΔ: εκτίμηση παραμέτρων

4. Ο $\hat{\text{Var}}\bar{X}_{st} = \sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell})\frac{S_{\ell}^2}{n_{\ell}}$ είναι ΑΕ της $\text{Var}\bar{X}_{st}$

5. $\hat{\Sigma}(\bar{X}_{st}) = \sqrt{\sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell})\frac{S_{\ell}^2}{n_{\ell}}}$ είναι εκτιμητής του $\Sigma(\bar{X}_{st})$

6. ΑΔΕ με σ.ε. $1 - \alpha$ για το $\mu : \bar{X}_{st} \pm z_{\alpha/2} \sqrt{\sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell})\frac{S_{\ell}^2}{n_{\ell}}}$

Απόδειξη.

ΣΤΔ: εκτίμηση παραμέτρων

Παρατήρηση:

$$\text{Var} \bar{X}_{st} = \sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\sigma_{0\ell}^2}{n_{\ell}}$$

Εάν $\sigma_{0\ell}^2$ είναι «μικρή» για κάθε ℓ τότε $\text{Var} \bar{X}_{st}$ είναι «μικρή»,

δηλαδή εάν οι τιμές του U σε κάθε στρώμα χωριστά δεν διαφέρουν πολύ μεταξύ τους (ομοιογένεια σε κάθε στρώμα) τότε έχουμε «μεγάλη» ακρίβεια στην εκτίμηση του μ .

Αρχή της στρωματοποίησης:

Ο καθορισμός των στρωμάτων θα πρέπει να γίνεται με τέτοιο τρόπο ώστε κάθε στρώμα να αποτελείται από ομοιογενείς ως προς το χαρακτηριστικό U μονάδες.

ΣΤΔ: εκτίμηση παραμέτρων

Εκτίμηση ολικής τιμής: $\tau = \sum_{\ell=1}^L \sum_{j=1}^{N_{\ell}} U_{\ell j} = N\mu$

Ως εκτιμητής του $\tau = N\mu$ θα χρησιμοποιηθεί ο

$$\hat{\tau}_{st} = N\bar{X}_{st}$$

Πρόταση 4

Σε ΣΤΔ

1. Ο $\hat{\tau}_{st}$ είναι ΑΕ του τ .

2. $\text{Var}\hat{\tau}_{st} = N^2 \sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\sigma_{0\ell}^2}{n_{\ell}}$

3. $\text{TS}(\hat{\tau}_{st}) = N \sqrt{\sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\sigma_{0\ell}^2}{n_{\ell}}}$

(συνεχίζεται)

ΣΤΔ: εκτίμηση παραμέτρων

4. Ο $\hat{\text{Var}}\hat{\tau}_{st} = N^2 \sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{S_{\ell}^2}{n_{\ell}}$ είναι ΑΕ της $\text{Var}\hat{\tau}_{st}$

5. $\hat{\Sigma}(\hat{\tau}_{st}) = N \sqrt{\sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{S_{\ell}^2}{n_{\ell}}}$ είναι εκτιμητής του $\text{TS}(\hat{\tau}_{st})$

6. ΑΔΕ με σ.ε. $1 - \alpha$ για το τ : $\hat{\tau}_{st} \pm z_{\alpha/2} N \sqrt{\sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{S_{\ell}^2}{n_{\ell}}}$

Απόδειξη.

Παρατήρηση:

Ένας άλλος εκτιμητής του $\tau = N\mu$ είναι ο

$$\hat{\tau} = N\bar{X}$$

ο οποίος όμως προφανώς στη ΣΤΔ δεν είναι ΑΕ του τ .

ΣΤΔ: εκτίμηση παραμέτρων

Εκτίμηση ποσοστού: p

Θεωρώντας

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

έχουμε

$$p_{\ell} = \frac{1}{N_{\ell}} \sum_{j=1}^{N_{\ell}} U_{\ell j} = \text{ποσοστό στο στρώμα } \ell \quad (\text{ειδική περίπτωση του } \mu_{\ell})$$

$$p = \frac{1}{N} \sum_{\ell=1}^L \sum_{j=1}^{N_{\ell}} U_{\ell j} = \sum_{\ell=1}^L W_{\ell} p_{\ell} \quad (\text{ειδική περίπτωση του } \mu)$$

Ως εκτιμητής του p θα χρησιμοποιηθεί ο

$$\hat{p}_{st} = \sum_{\ell=1}^L W_{\ell} \hat{p}_{\ell}$$

όπου $\hat{p}_{\ell} = \bar{X}_{\ell}$.

ΣΤΔ: εκτίμηση παραμέτρων

Πρόταση 5

Σε ΣΤΔ

1. Ο $\hat{p}_{st}$ είναι ΑΕ του p .

$$2. \text{Var}\hat{p}_{st} = \sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\frac{N_{\ell}}{N_{\ell}-1} p_{\ell} (1 - p_{\ell})}{n_{\ell}}$$

$$3. \text{TS}(\hat{p}_{st}) = \sqrt{\sum_{\ell=1}^L W_{\ell}^2 (1 - f_{\ell}) \frac{\frac{N_{\ell}}{N_{\ell}-1} p_{\ell} (1 - p_{\ell})}{n_{\ell}}}$$

(συνεχίζεται)

ΣΤΔ: εκτίμηση παραμέτρων

4. Ο $\widehat{\text{Var}}\hat{p}_{st} = \sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell}) \frac{\hat{p}_{\ell}(1 - \hat{p}_{\ell})}{n_{\ell} - 1}$ είναι ΑΕ της $\text{Var}\hat{p}_{st}$

5. $\hat{\Sigma}(\hat{p}_{st}) = \sqrt{\sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell}) \frac{\hat{p}_{\ell}(1 - \hat{p}_{\ell})}{n_{\ell} - 1}}$ είναι εκτιμητής του $\Sigma(\hat{p}_{st})$

6. ΑΔΕ με σ.ε. $1 - \alpha$ για το $p : \hat{p}_{st} \pm z_{\alpha/2} \sqrt{\sum_{\ell=1}^L W_{\ell}^2(1 - f_{\ell}) \frac{\hat{p}_{\ell}(1 - \hat{p}_{\ell})}{n_{\ell} - 1}}$

Απόδειξη.

ΣΤΔ: εύρεση μεγέθους δείγματος

I. Επιλογή μεγέθους δείγματος n

Πρόβλημα:

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

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

Α' τρόπος: Θεωρώντας $\mathbb{P}(\hat{\theta}) = v_0$, 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 .

ΣΤΔ: εύρεση μεγέθους δείγματος

I. Επιλογή μεγέθους δείγματος n (συνέχεια)

Εκτίμηση μέσης τιμής:

Α' τρόπος: $\mathbb{P}(\bar{X}_{st}) = v_0$ και θεωρώντας w_ℓ γνωστά (προκαθορισμένα), τότε

$$n = \frac{\sum_{\ell=1}^L \frac{W_\ell^2}{w_\ell} \sigma_{0\ell}^2}{\sum_{\ell=1}^L \frac{W_\ell^2}{N_\ell} \sigma_{0\ell}^2 + v_0^2} \quad (1)$$

Ειδική περίπτωση: Εάν θεωρήσουμε αναλογική κατανομή, δηλ. $w_\ell = W_\ell, \forall \ell$, τότε

$$n = \frac{1}{\frac{1}{N} + \frac{1}{n_0}}, \quad \text{όπου } n_0 = \frac{\sum_{\ell=1}^L W_\ell \sigma_{0\ell}^2}{v_0^2} \quad (2)$$

Β' τρόπος: $\mathbb{P}(|\bar{X}_{st} - \mu| \leq b) = 1 - \alpha$, τότε το πρόβλημα ανάγεται στον Α' τρόπο και προκύπτουν οι σχέσεις (1) και (2), αντίστοιχα με $v_0^2 = \frac{b^2}{z_{\alpha/2}^2}$

ΣΤΔ: εύρεση μεγέθους δείγματος

1. Επιλογή μεγέθους δείγματος n (συνέχεια)

Εκτίμηση ολικής τιμής:

Α' τρόπος: $\mathbb{P}(\hat{\tau}_{st}) = v_1$ και θεωρώντας w_ℓ γνωστά (προκαθορισμένα), τότε επειδή

$$\mathbb{P}(\hat{\tau}_{st}) = v_1 \Leftrightarrow \mathbb{P}(\bar{X}_{st}) = \frac{v_1}{N} = v_0,$$

προκύπτουν οι σχέσεις (1) και (2), αντίστοιχα με $v_0 = \frac{v_1}{N}$

Β' τρόπος: $\mathbb{P}(|\hat{\tau}_{st} - \tau| \leq b_1) = 1 - \alpha$, τότε επειδή

$$\mathbb{P}(|\hat{\tau}_{st} - \tau| \leq b_1) = 1 - \alpha \Leftrightarrow \mathbb{P}(|\bar{X}_{st} - \mu| \leq \frac{b_1}{N}) = 1 - \alpha$$

προκύπτουν οι σχέσεις (1) και (2), αντίστοιχα με $v_0^2 = \frac{b_1^2}{N^2 z_{\alpha/2}^2}$

ΣΤΔ: εύρεση μεγέθους δείγματος

Ι. Επιλογή μεγέθους δείγματος n (συνέχεια)

Εκτίμηση ποσοστού:

Το ποσοστό p αποτελεί ειδική περίπτωση μέσης τιμής μ , οπότε προκύπτουν οι σχέσεις (1) και (2), αντίστοιχα με

$$\begin{aligned}\sigma_{0\ell}^2 &= \frac{N_\ell}{N_\ell - 1} p_\ell (1 - p_\ell) \\ &\approx p_\ell (1 - p_\ell) \quad \text{για } N_\ell \text{ «μεγάλο»}\end{aligned}$$

ΣΤΔ: εύρεση μεγέθους δείγματος

II. Επιλογή των n_ℓ με δεδομένο συνολικό μέγεθος δείγματος n

A. Αναλογική κατανομή:

$$n_\ell = \frac{n}{N} N_\ell, \quad \forall \ell$$

B. Βέλτιστη κατανομή Neyman:

Θέλουμε να βρούμε τα n_ℓ έτσι ώστε να ελαχιστοποιείται η διασπορά του εκτιμητή μας για την υπό μελέτη παράμετρο με $\sum_{\ell=1}^L n_\ell = n$ (όπου n δεδομένο)

ΣΤΔ: εύρεση μεγέθους δείγματος

II. Επιλογή των n_ℓ με δεδομένο συνολικό μέγεθος δείγματος n (συνέχεια)

B. Βέλτιστη κατανομή Neyman:

Εκτίμηση μέσης τιμής:

$\min_{n_1, \dots, n_L} \text{Var}(\bar{X}_{st})$ υπό τη συνθήκη $\sum_{\ell=1}^L n_\ell = n$, τότε

$$\begin{aligned} n_\ell &= n \frac{W_\ell \sigma_{0\ell}}{\sum_{\ell=1}^L W_\ell \sigma_{0\ell}} \\ &= n \frac{N_\ell \sigma_{0\ell}}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell}} \end{aligned} \tag{3}$$

Παρατήρηση:

Είναι $n_\ell \propto N_\ell \sigma_{0\ell}$, δηλαδή η βέλτιστη κατανομή Neyman λαμβάνει υπόψη όχι μόνο το μέγεθος του πληθυσμού στο στρώμα ℓ , αλλά και τη διασπορά στο στρώμα ℓ .

ΣΤΔ: εύρεση μεγέθους δείγματος

II. Επιλογή των n_ℓ με δεδομένο συνολικό μέγεθος δείγματος n (συνέχεια)

B. Βέλτιστη κατανομή Neyman:

Εκτίμηση ολικής τιμής:

$\min_{n_1, \dots, n_L} \text{Var}(\hat{\tau}_{st})$ υπό τη συνθήκη $\sum_{\ell=1}^L n_\ell = n$, τότε επειδή
 $\text{Var}(\hat{\tau}_{st}) = N^2 \text{Var}(\bar{X}_{st})$, προκύπτει η (3).

Εκτίμηση ποσοστού:

$\min_{n_1, \dots, n_L} \text{Var}(\hat{p}_{st})$ υπό τη συνθήκη $\sum_{\ell=1}^L n_\ell = n$, τότε επειδή το ποσοστό
 p είναι ειδική περίπτωση του μ , προκύπτει η (3) με

$$\begin{aligned}\sigma_{0\ell}^2 &= \frac{N_\ell}{N_\ell - 1} p_\ell (1 - p_\ell) \\ &\approx p_\ell (1 - p_\ell) \quad \text{για } N_\ell \text{ «μεγάλο»}\end{aligned}$$

ΣΤΔ: εύρεση μεγέθους δείγματος

III. Επιλογή των n_ℓ λαμβάνοντας υπόψη το κόστος της δειγματοληψίας

Ας θεωρήσουμε το κόστος της δειγματοληψίας ως γραμμική συνάρτηση των n_ℓ

c_ℓ : κόστος ανά μονάδα δείγματος από το στρώμα ℓ , οπότε $c_\ell \cdot n_\ell$ είναι το κόστος για στρώμα ℓ

c_0 : γενικό κόστος δειγματοληψίας (ανεξάρτητο από τα n_ℓ)

Το συνολικό κόστος είναι $C = c_0 + \sum_{\ell=1}^L c_\ell n_\ell$

Προβλήματα:

- A. Θέλουμε να βρούμε τα n_ℓ έτσι ώστε να ελαχιστοποιείται η διασπορά του εκτιμητή μας για την υπό μελέτη παράμετρο με δεδομένο κόστος C .
- B. Θέλουμε να βρούμε τα n_ℓ έτσι ώστε να ελαχιστοποιείται το κόστος C με δεδομένη διασπορά του εκτιμητή μας για την υπό μελέτη παράμετρο.

ΣΤΔ: εύρεση μεγέθους δείγματος

III. Επιλογή των n_ℓ λαμβάνοντας υπόψη το κόστος της δειγματοληψίας
(συνέχεια)

Εκτίμηση μέσης τιμής:

Α' πρόβλημα: $\min_{n_1, \dots, n_L} \text{Var}(\bar{X}_{st})$ υπό τη συνθήκη $C = c_0 + \sum_{\ell=1}^L c_\ell n_\ell$ με C, c_0, c_ℓ δεδομένα, τότε

$$\begin{aligned} n_\ell &= \frac{C - c_0}{\sum_{\ell=1}^L W_\ell \sigma_{0\ell} \sqrt{c_\ell}} \frac{W_\ell \sigma_{0\ell}}{\sqrt{c_\ell}} \\ &= \frac{C - c_0}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell} \sqrt{c_\ell}} \frac{N_\ell \sigma_{0\ell}}{\sqrt{c_\ell}} \end{aligned} \quad (4)$$

ενώ

$$n = \sum_{\ell=1}^L n_\ell = \frac{C - c_0}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell} \sqrt{c_\ell}} \sum_{\ell=1}^L \frac{N_\ell \sigma_{0\ell}}{\sqrt{c_\ell}}$$

Ειδική περίπτωση: Εάν θεωρήσουμε $c_\ell = c, \forall \ell$, τότε προκύπτει

$$n_\ell = n \frac{N_\ell \sigma_{0\ell}}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell}} \quad (\text{βέλτιστη κατανομή Neyman})$$

ΣΤΔ: εύρεση μεγέθους δείγματος

III. Επιλογή των n_ℓ λαμβάνοντας υπόψη το κόστος της δειγματοληψίας
(συνέχεια)

Εκτίμηση μέσης τιμής:

Β' πρόβλημα: $\min_{n_1, \dots, n_L} \left\{ c_0 + \sum_{\ell=1}^L c_\ell n_\ell \right\}$ υπό τη συνθήκη $\text{Var}(\bar{X}_{st}) = v_0^2$ με
 v_0 δεδομένο, τότε

$$\begin{aligned} n_\ell &= \frac{\sum_{\ell=1}^L W_\ell \sigma_{0\ell} \sqrt{c_\ell}}{\frac{1}{N} \sum_{\ell=1}^L W_\ell \sigma_{0\ell}^2 + v_0^2} \frac{W_\ell \sigma_{0\ell}}{\sqrt{c_\ell}} \\ &= \frac{\sum_{\ell=1}^L N_\ell \sigma_{0\ell} \sqrt{c_\ell}}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell}^2 + N^2 v_0^2} \frac{N_\ell \sigma_{0\ell}}{\sqrt{c_\ell}} \end{aligned} \quad (5)$$

ενώ

$$n = \sum_{\ell=1}^L n_\ell = \frac{\sum_{\ell=1}^L N_\ell \sigma_{0\ell} \sqrt{c_\ell}}{\sum_{\ell=1}^L N_\ell \sigma_{0\ell}^2 + N^2 v_0^2} \sum_{\ell=1}^L \frac{N_\ell \sigma_{0\ell}}{\sqrt{c_\ell}}$$

Ειδική περίπτωση: Εάν θεωρήσουμε $c_\ell = c, \forall \ell$, τότε προκύπτει
για τα n_ℓ βέλτιστη κατανομή **Neyman**.

ΣΤΔ: εύρεση μεγέθους δείγματος

III. Επιλογή των n_ℓ λαμβάνοντας υπόψη το κόστος της δειγματοληψίας (συνέχεια)

Εκτίμηση ολικής τιμής:

Α' πρόβλημα: $\min_{n_1, \dots, n_L} \text{Var}(\hat{\tau}_{st})$ υπό τη συνθήκη $C = c_0 + \sum_{\ell=1}^L c_\ell n_\ell$ με C, c_0, c_ℓ δεδομένα, τότε επειδή $\text{Var}(\hat{\tau}_{st}) = N^2 \text{Var}(\bar{X}_{st})$ προκύπτει η σχέση (4) για τα n_ℓ (λόγω ισοδυναμίας των δύο προβλημάτων).

Β' πρόβλημα: $\min_{n_1, \dots, n_L} \left\{ c_0 + \sum_{\ell=1}^L c_\ell n_\ell \right\}$ υπό τη συνθήκη $\text{Var}(\hat{\tau}_{st}) = v_1^2$ με v_1 δεδομένο, τότε επειδή $\text{Var}(\hat{\tau}_{st}) = v_1^2 \Leftrightarrow \text{Var}(\bar{X}_{st}) = \frac{v_1^2}{N^2} = v_0^2$ προκύπτει η σχέση (5) για τα n_ℓ (λόγω ισοδυναμίας των δύο προβλημάτων) με $v_0^2 = \frac{v_1^2}{N^2}$.

ΣΤΔ: εύρεση μεγέθους δείγματος

III. Επιλογή των n_ℓ λαμβάνοντας υπόψη το κόστος της δειγματοληψίας
(συνέχεια)

Εκτίμηση ποσοστού:

Α' πρόβλημα: $\min_{n_1, \dots, n_L} \text{Var}(\hat{p}_{st})$ υπό τη συνθήκη $C = c_0 + \sum_{\ell=1}^L c_\ell n_\ell$ με C, c_0, c_ℓ δεδομένα, τότε επειδή το ποσοστό p είναι ειδική περίπτωση του μ , προκύπτει η σχέση (4) για τα n_ℓ με

$$\sigma_{0\ell}^2 = \frac{N_\ell}{N_\ell - 1} p_\ell (1 - p_\ell) \\ \approx p_\ell (1 - p_\ell) \quad \text{για } N_\ell \text{ «μεγάλο»}$$

Β' πρόβλημα: $\min_{n_1, \dots, n_L} \left\{ c_0 + \sum_{\ell=1}^L c_\ell n_\ell \right\}$ υπό τη συνθήκη $\text{Var}(\hat{p}_{st}) = v_0^2$ με v_0 δεδομένο, τότε επειδή το ποσοστό p είναι ειδική περίπτωση του μ , προκύπτει η σχέση (5) για τα n_ℓ με

$$\sigma_{0\ell}^2 = \frac{N_\ell}{N_\ell - 1} p_\ell (1 - p_\ell) \\ \approx p_\ell (1 - p_\ell) \quad \text{για } N_\ell \text{ «μεγάλο»}$$

Σύγκριση ΣΤΔ και ΑΤΔ

Ας συμβολίσουμε

$\bar{X}_{st,opt}$ ο στρωματοποιημένος δειγματικός μέσος παίρνοντας βέλτιστη κατανομή Neyman για τα n_ℓ

$\bar{X}_{st,prop}$ ο στρωματοποιημένος δειγματικός μέσος παίρνοντας αναλογική κατανομή για τα n_ℓ

$\bar{X}_{AT\Delta}$ ο δειγματικός μέσος με ΑΤΔ

Προφανώς, $\text{Var}(\bar{X}_{st,opt}) \leq \text{Var}(\bar{X}_{st,prop})$ εξ ορισμού.

Πρόταση 6

Εάν ο όρος $1/N$ μπορεί να αγνοηθεί (δηλ. $\frac{1}{N} \approx 0$), τότε $\text{Var}(\bar{X}_{st,opt}) \leq \text{Var}(\bar{X}_{AT\Delta})$.

Απόδειξη.