

Ασκήσεις II (ΣΤΔ)

1. Ο παρακάτω πίνακας περιέχει τους πληθυσμούς (σε χιλιάδες) 64 μεγάλων πόλεων των ΗΠΑ σε απογραφή του 1930. Οι πόλεις έχουν χωριστεί σε δύο στρώματα, το πρώτο περιέχει τις 16 μεγαλύτερες και το δεύτερο τις υπόλοιπες 48.

στρώμα	$U_{\ell j}$	$\sum_j U_{\ell j}$	$\sum_j U_{\ell j}^2$
1	900, 822, 781, 805, 670, 1238, 573, 634, 578, 487, 442, 451, 459, 464, 400, 366	10 070	7 145 450
2	364, 317, 328, 302, 288, 291, 253, 291, 308, 272, 284, 255, 270, 214, 195, 260, 209, 183, 163, 253, 232, 260, 201, 147, 292, 164, 143, 169, 139, 170, 150, 143, 113, 115, 123, 154, 140, 119, 130, 127, 100, 107, 114, 111, 163, 116, 122, 134	9 498	2 141 720

Για δείγμα μεγέθους 24 πόλεων να υπολογισθεί το τυπικό σφάλμα του εκτιμητή του συνολικού πληθυσμού

- με ΑΤΔ,
- με ΣΤΔ και αναλογική κατανομή των δειγματικών μεγεθών,
- με ΣΤΔ και ίσα δειγματικά μεγέθη.

Ποιά από τις τρεις μεθόδους παρέχει την μεγαλύτερη ακρίβεια; Ποιά είναι τα δειγματικά μεγέθη που προκύπτουν για ΣΤΔ με κατανομή Neyman;

2. Για ένα δείγμα 1000 ατόμων που επιλέχθηκε με ΣΤΔ και αναλογική κατανομή δειγματικών μεγεθών από μια περιοχή που χωρίζεται σε τρία στρώματα, ο παρακάτω πίνακας δίνει, ανά στρώμα, το μέσο εισόδημα (σε χιλιάδες ευρώ), την αντίστοιχη διασπορά, το ποσοστό των ατόμων του δείγματος με εισόδημα κάτω των 3000 ευρώ. Επίσης, δίνονται οι αντίστοιχες ποσότητες για ολόκληρο το δείγμα, καθώς και τα βάρη των στρωμάτων για τον πληθυσμό.

στρώμα	W_{ℓ}	$\bar{X}_{\ell}$	S_{ℓ}^2	$\hat{p}_{\ell}$
1	0.35	3.1	4	0.54
2	0.55	3.9	11	0.39
3	0.10	7.8	128	0.24
για ολόκληρο το δείγμα		4.01	21.98	0.43

Να απαντήσετε στα παρακάτω ερωτήματα αγνοώντας τις διορθώσεις για πεπεραμένους πληθυσμούς.

- Χρησιμοποιώντας ΑΤΔ να εκτιμηθεί το μέσο εισόδημα του πληθυσμού και το ποσοστό των ατόμων με εισόδημα κάτω των 3000 ευρώ. Να δώσετε τις εκτιμήσεις των διασπορών τους.
- Χρησιμοποιώντας ΣΤΔ με αναλογική κατανομή δειγματικών μεγεθών να απαντήσετε το (α). Ποιές είναι οι σχετικές βελτιώσεις των διαπορών των εκτιμητών σε σχέση με το (α);

γ. Να βρεθεί η κατανομή των δειγματικών μεγεθών για τα τρία στρώματα έτσι ώστε να ελαχιστοποιείται η διασπορά του εκτιμητή του πληθυσμιακού μέσου και του πληθυσμιακού ποσοστού, αντίστοιχα. Να υπολογίσετε (εκτιμήσετε) τις διασπορές αυτών των εκτιμητών και να τις συγκρίνετε με αυτές των ερωτημάτων (α) και (β).

3. Εάν το κόστος μιας ΣΤΔ είναι της μορφής

$$C = c_0 + \sum_{\ell=1}^L c_{\ell} \sqrt{n_{\ell}},$$

όπου τα c_0 και c_{ℓ} είναι γνωστά, αποδείξτε ότι το μέγεθος του δείγματος για το στρώμα ℓ , $\ell = 1, \dots, L$, που απαιτείται για την ελαχιστοποίηση της διασποράς του στρωματοποιημένου δειγματικού μέσου είναι ανάλογο με $(W_{\ell}^2 \sigma_{0\ell}^2 / c_{\ell})^{2/3}$. Στη συνέχεια, χρησιμοποιώντας τα δεδομένα του παρακάτω πίνακα, βρείτε ποιά θα είναι η κατανομή των δειγματικών μεγεθών στα τρία στρώματα ενός δείγματος μεγέθους $n = 1000$.

στρώμα	W_{ℓ}	$\sigma_{0\ell}$	c_{ℓ}
1	0.4	4	1
2	0.3	5	2
3	0.3	6	4

Ποιά θα ήταν η κατανομή εάν είχαμε γραμμικό κόστος $C = c_0 + \sum_{\ell=1}^L c_{\ell} n_{\ell}$;

4. Για να εκτιμηθεί το ποσοστό τηλεθέασης μιας εκπομπής ο συνολικός πληθυσμός 3100 οικογενειών χωρίζεται σε τρία στρώματα με 1550, 620, 930 οικογένειες αντίστοιχα και επιλέγεται στρωματοποιημένο τυχαίο δείγμα 400 οικογενειών με αναλογική κατανομή δειγματικών μεγεθών στα τρία στρώματα. Ο αριθμός των οικογενειών στο δείγμα που παρακολουθούν την εκπομπή βρέθηκε 160, 20, 60, αντίστοιχα.

α. Να εκτιμηθεί το ποσοστό τηλεθέασης και να κατασκευασθεί ΑΔΕ με σ.ε. 95%.

β. Ας υποθέσουμε ότι τα παραπάνω δεδομένα συγκεντρώθηκαν πέρυσι και ότι ο συνολικός πληθυσμός και τα στρώματα παραμένουν τα ίδια και φέτος. Η έρευνα θα πραγματοποιηθεί τηλεφωνικά φέτος, οπότε το κόστος ανά δειγματοληπτική μονάδα μπορεί να θεωρηθεί το ίδιο. Να εκτιμηθεί το μέγεθος του δείγματος που απαιτείται ώστε να ελαχιστοποιηθεί το κόστος με το τυπικό σφάλμα στην εκτίμηση του ποσοστού τηλεθέασης να είναι 1%.

γ. Ας υποθέσουμε ότι, αντί της ΣΤΔ, επιλέγεται με ΑΤΔ δείγμα μεγέθους 400 οικογενειών από τις οποίες οι 240 παρακολουθούν την εκπομπή. Να εκτιμηθεί το ποσοστό τηλεθέασης και να διαπιστωθεί ποιά από τις δύο μεθόδους είναι πιο ακριβής.

5. Ένας προμηθευτής ειδών supermarket ενδιαφέρεται να προσθέσει ένα νέο προϊόν σε αυτά που ήδη εφοδιάζει 4 αλυσίδες supermarket. Οι αριθμοί καταστημάτων των αλυσίδων είναι 24, 36, 30, 30, αντίστοιχα. Για να αποφασίσει εάν τελικά θα προσθέσει το νέο προϊόν χρησιμοποιεί δειγματοληπτικά 20 συνολικά καταστήματα με ΣΤΔ, θεωρώντας τις αλυσίδες ως στρώματα και επιλέγοντας τα μεγέθη δειγμάτων από κάθε αλυσίδα αναλογικά με τον αριθμό των καταστημάτων της. Μετά από ένα μήνα διάθεσης του προϊόντος στα 20 επιλεγμένα καταστήματα συγκεντρώθηκαν τα εξής στοιχεία

	αλυσίδα 1	αλυσίδα 2	αλυσίδα 3	αλυσίδα 4
μέση είσπραξη (σε χιλιάδες ευρώ)	99	100	98	100
διασπορά εισπράξεων	78.70	55.60	39.50	112.50

ενώ η συνολική διασπορά εισπράξεων (στα 20 καταστήματα) είναι 59.8.

- α. Να εκτιμηθεί η μέση είσπραξη ανά μήνα, καθώς και το τυπικό σφάλμα της.
- β. Ας υποθέσουμε ότι, αντί της ΣΤΔ, ο προμηθευτής χρησιμοποιεί όλα τα καταστήματα και των 4 αλυσίδων ως ένα πληθυσμό, επιλέγει ΑΤΔ 20 καταστημάτων και προκύπτουν τα ίδια 20 όπως στην ΣΤΔ. Να εκτιμηθεί η μέση είσπραξη και το τυπικό σφάλμα της και να διαπιστωθεί ποιά από τις δύο μεθόδους είναι πιο ακριβής.
- γ. Μπορείτε να προτείνετε έναν εναλλακτικό και ενδεχομένως πιο αποτελεσματικό τρόπο στρωματοποίησης του πληθυσμού των καταστημάτων;

6. Σε δειγματοληπτική έρευνα με ΣΤΔ που έγινε σε μία πόλη για να διαπιστωθεί εάν έπρεπε να μεταφερθεί ένας παιδικός σταθμός από το κέντρο στα περίχωρα της πόλης, ρωτήθηκαν οικογένειες που είχαν τουλάχιστον ένα παιδί κάτω των έξι ετών. Οι οικογένειες αυτές χωρίστηκαν σε 4 κατηγορίες ανάλογα με το εάν είχαν ή όχι παιδί στον παιδικό σταθμό και εάν έμεναν στο κέντρο ή τα περίχωρα της πόλης. Ο αριθμός των οικογενειών καθώς και το κόστος δειγματοληψίας σε κάθε κατηγορία δίνονται στον ακόλουθο πίνακα,

	περίχωρα χωρίς παιδί στο σταθμό	περίχωρα με παιδί στο σταθμό	κέντρο χωρίς παιδί στο σταθμό	κέντρο με παιδί στο σταθμό
στρώμα	1	2	3	4
N_ℓ	97	43	145	68
c_ℓ	4	4	9	9

ενώ από προηγούμενη μελέτη βρέθηκε ότι (προσεγγιστικά) το 90% των οικογενειών με παιδί στον παιδικό σταθμό και το 50% των οικογενειών χωρίς παιδί στον παιδικό σταθμό θα χρησιμοποιούσε το σταθμό μετά τη μεταφορά του ανεξάρτητα εάν μένει κέντρο ή περίχωρα.

- α. Θεωρώντας αναλογική κατανομή, βρείτε το μέγεθος δείγματος και την κατανομή του στα στρώματα για την εκτίμηση του ποσοστού των οικογενειών με περιθώριο σφάλματος 0.05 και εμπιστοσύνη 95%.
- β. Εάν οι εκτιμήσεις των ποσοστών των οικογενειών που θα χρησιμοποιήσουν τον καινούργιο παιδικό σταθμό ανά στρώμα είναι

$$\hat{p}_1 = 0.60, \quad \hat{p}_2 = 0.93, \quad \hat{p}_3 = 0.53, \quad \hat{p}_4 = 0.87,$$

εκτιμήστε το ποσοστό και δώστε ΑΔΕ με σ.ε. 95%.

- γ. Εάν το κόστος δειγματοληψίας είναι $C = 400$ χρηματικές μονάδες, βρείτε τη βέλτιστη κατανομή δειγματικών μεγεθών για την ελαχιστοποίηση της διασποράς του εκτιμητή του ποσοστού.