ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ Δ.Ε.

Ορισμός Έστω $x_1, x_2, \ldots, x_n$ τυχαίο δείγμα από πληθυσμό του οποίου δε γνωρίζουμε κάποια παράμετρο θ. Το διάστημα $(A, B) = (\hat{\theta} - \hat{e}, \hat{\theta} + \hat{e})$, όπου $\hat{\theta}$ η αντίστοιχη δειγματική παράμετρος και $\hat{e}$ το δειγματοληπτικό σφάλμα ή ακρίβεια της εκτίμησης (που εξαρτώνται από το δείγμα), για το οποίο $P(A < \theta < B) = 1 - a$, $0 < a < 1$, ονομάζεται $100 \cdot (1-a)\%$ διάστημα εμπιστοσύνης Δ.Ε. για την παράμετρο θ. Ο αριθμός $A = \hat{\theta} - \hat{e}$ ονομάζεται $100 \cdot (1-a)\%$ κάτω όριο εμπιστοσύνης για την πληθυσμιακή παράμετρο θ, ενώ ο $B = \hat{\theta} + \hat{e}$ ονομάζεται $100 \cdot (1-a)\%$ άνω όριο εμπιστοσύνης για την πληθυσμιακή παράμετρο θ.

Η πιθανότητα a ονομάζεται πιθανότητα σφάλματος ή επίπεδο στατιστικής σημαντικότητας, ενώ η συμπληρωματική πιθανότητα

$1-a$ ονομάζεται συντελεστής ή επίπεδο εμπιστοσύνης σ.ε.

Σημείωση 1 (i) Η ερμηνεία του συντελεστή εμπιστοσύνης είναι η εξής: σ.ε. $1-a$ σημαίνει ότι αν πάρουμε π.χ. 100 διαφορετικά δείγματα, από τον ίδιο πληθυσμό, που έχουν το ίδιο μέγεθος, τότε κάθε δείγμα θα δώσει ένα διαφορετικό δ.ε. αλλά $100 \cdot (1-a)$ από αυτά θα περιέχουν την άγνωστη παράμετρο θ. Έτσι η πιθανότητα ένα τυχαία επιλεγμένο δ.ε. να περιέχει την παράμετρο θ είναι $1-a$.

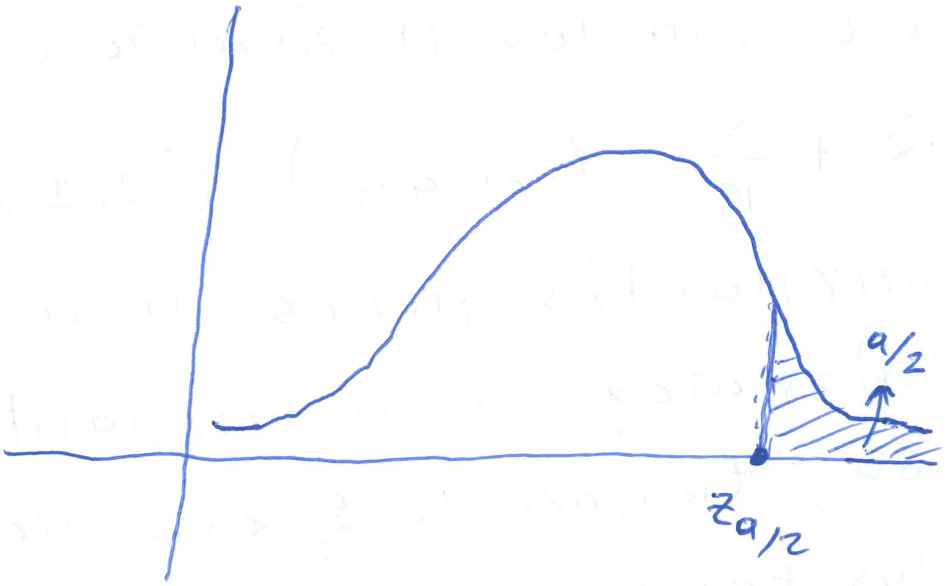
(ii) Όσο μειώνεται η πιθανότητα σφάλματος a, τόσο αυξάνεται ο σ.ε. $1-a$, τόσο μειώνεται το πλάτος του δ.ε., τόσο αυξάνεται η ακρίβεια της εκτίμησης.

1 Δ.ε. για τον πληθυσμιακό μέσο μ με γνωστή διακύμανση σ^2 (κανονικός πληθυσμός)

Πρόταση 1 Ένα $100 \cdot (1-a)\%$ δ.ε. για τον μ είναι το

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{a/2}, \ \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{a/2}\right) \quad (1)$$

όπου $\bar{x}$ ο δειγματικός μέσος, n το μέγεθος του δείγματος, $z_{a/2}$ το $\frac{a}{2}$ άνω ποσοστιαίο σημείο της $N(0,1)$ δηλ. το σημείο για το οποίο $P(Z > z_{a/2}) = \frac{a}{2}$ ή ισοδύναμα $P(Z \le z_{a/2}) = 1 - \frac{a}{2}$ αφού οι πίνακες της $N(0,1)$ αφορούν τιμές της συνάρτησης κατανομής.



Σημείωση 2 (i) Η ποσότητα $e = \frac{\sigma}{\sqrt{n}} \cdot z_{a/2}$ είναι η ακρίβεια της μέτρησης. Η ποσότητα $\frac{\sigma}{\sqrt{n}}$ καλείται και τυπικό σφάλμα της δειγματοληπτικής κατανομής του μέσου. (Είναι η δειγματική τυπική απόκλιση)

(ii) Ο τύπος (1) ισχύει και για δείγματα προερχόμενα από πληθυσμό που δεν ακολουθεί κανονική κατανομή αρκεί μέγεθος δείγματος $n \geq 30$ οπότε και εφαρμόζεται το Κεντρικό Οριακό Θεώρημα Κ.Ο.Θ.

② Δ.ε. για τον πληθυσμιακό μέσο μ με άγνωστη διακύμανση σ^2 (κανονικός πληθυσμός)

(i) Πρόταση 2.1 | $n < 30$ (μικρό δείγμα) Ένα $100\cdot(1-a)\%$ δ.ε. για τον μ είναι το $\left(\bar{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1, a/2}, \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1, a/2}\right)$ (2.1), όπου $\bar{x}$ ο δειγματικός μέσος, n το μέγεθος του δείγματος, s η δειγματική τυπική απόκλιση, $t_{n-1, a/2}$ το $\frac{a}{2}$ άνω ποσοστιαίο σημείο της κατανομής Student με $n-1$ βαθμούς ελευθερίας t_{n-1}, δηλ. το σημείο για το οποίο $P(t_{n-1} > t_{n-1, a/2}) = \frac{a}{2} = P(t_{n-1} < -t_{n-1, a/2})$ που υπολογίζεται απευθείας από τους πίνακες της κατανομής Student (μονοκατάληκτο κριτήριο).

(ii) $n \geq 30$. (μεγάλο δείγμα) Ένα $100\cdot(1-a)\%$ δ.ε. για τον μ είναι το $\left(\bar{x} - \frac{s}{\sqrt{n}} \cdot z_{a/2}, \bar{x} + \frac{s}{\sqrt{n}} \cdot z_{a/2}\right.$

(2.2)

<u>Σημείωση 3</u> Ο τύπος (2.2) ισχύει και για δείγματα προερχόμενα από μη κανονικό πληθυσμό λόγω του Κ.Ο.Θ.

③ <u>Δ.Ε. ποσοστού p</u>

<u>Πρόταση 3</u> Ένα $100\cdot(1-a)\%$ δ.ε. για το ποσοστό p ενός πληθυσμού με κάποιο συγκεκριμένο χαρακτηριστικό είναι το $(\hat{p}-z_{a/2}\cdot s_{\hat{p}}\ ,\ \hat{p}+z_{a/2}\cdot s_{\hat{p}})$ (3), όπου $\hat{p}$ το δειγματικό ποσοστό και $s_{\hat{p}}=\sqrt{\dfrac{\hat{p}\cdot(1-\hat{p})}{n}}$ που ονομάζεται και <u>τυπικό σφάλμα της δειγματοληπτικής κατανομής του ποσοστού</u>.

<u>Σημείωση 4</u> Ο παραπάνω τύπος ισχύει για δείγματα μεγέθους $n \geq 30$ ανεξαρτήτως κατανομής πληθυσμού.

<u>ΠΑΡΑΔΕΙΓΜΑΤΑ</u>

① Από έναν κανονικό πληθυσμό με διακύμανση $\sigma^2=25$ παίρνουμε ένα δείγμα μεγέθους $n=16$ και υπολογίζουμε τον δειγματικό μέσο

$\bar{x}=14,5$. Θέλουμε να προσδιορίσουμε ένα διάστημα τιμών στο οποίο να περιέχεται η πληθυσμιακή μέση τιμή με εμπιστοσύνη 90%.

Λύση | $1-a=0,90 \Rightarrow a=1-0,90=0,10 \Rightarrow \frac{a}{2}=0,05$. Καθώς είναι γνωστή η πληθυσμιακή διακύμανση θα χρησιμοποιήσουμε τον τύπο (1). Έχουμε ότι $P(Z>z_{0,05})=0,05 \Rightarrow P(Z\leq z_{0,05})=1-0,05=0,95$. Από τους πίνακες της $N(0,1)$ υπολογίζουμε, κατά προσέγγιση, ότι $z_{0,05}=1,65$. Επιπλέον $\sqrt{n}=\sqrt{16}=4$, $\sigma=\sqrt{\sigma^2}=\sqrt{25}=5$, $\bar{x}=14,5$. Έτσι το ζητούμενο δ.ε. είναι το $\left(14,5-\frac{5}{4}\cdot 1,65,\ 14,5+\frac{5}{4}\cdot 1,65\right)=(12,44,\ 16,56)$

---

2 Από έναν κανονικό πληθυσμό με άγνωστη διασπορά παίρνουμε ένα δείγμα μεγέθους $n=16$ και υπολογίζουμε τον δειγματικό μέσο $\bar{x}=14,5$ και τη δειγματική τυπική απόκλιση $s=5$. Να βρεθεί ένα δ.ε. του πληθυσμιακού μέσου μ με εμπιστοσύνη 90%

1)64 $1-a=0,90 \Rightarrow a=1-0,90=0,10$ $\Rightarrow \frac{a}{2}=0,05$. Καθώς δεν γνωρίζουμε την πληθυσμιακή τυπική απόκλιση θα χρησιμοποιήσουμε τον τύπο (2.1) αφού και $n=16<30$. Από τους πίνακες της κατανομής Student έχουμε ότι αν $P(t_{15}>t_{15,0,05})=0,05$, τότε $t_{15,0,05}=1,753$. Άρα το ζητούμενο δ.ε. είναι το $\left(14,5-\frac{5}{4}\cdot 1,753,\ 14,5+\frac{5}{4}\cdot 1,753\right)=(12,31,\ 16,69)$

---

3 (σελ. 132-133 βιβλίο) Ένας φοιτητής διεξάγει έρευνα για το μέσο επίπεδο των ετήσιων αμοιβών των εργαζομένων σε μια χρηματιστηριακή εταιρία. Για τον λόγο αυτό επιλέγει ένα δείγμα $n=40$ εργαζομένων της εταιρίας και καταγράφει τις ετήσιες αποδοχές τους. Ο δειγματικός μέσος προκύπτει $\bar{x}=29.500$€ και η δειγματική τυπική απόκλιση $s=16.689$€. Να βρεθεί ένα 95% δ.ε για τον μέσο όρο των αμοιβών όλων των εργαζομένων στην εταιρία

Λύση | $1-a=0,95 \Rightarrow a=1-0,95=0,05$ $\Rightarrow \frac{a}{2}=0,025$. Επειδή δεν γνωρίζουμε την πληθυσμιακή διασπορά σ^2 και το δείγμα είναι μεγάλο ($n=40>30$) θα χρησιμοποιήσουμε τον τύπο (7.2). Έχουμε ότι $P(Z>z_{0,025})=0,025 \Rightarrow P(Z \leq z_{0,025})=1-0,025$ $=0,975$. Από τους πίνακες της $N(0,1)$ βρίσκουμε ότι $z_{0,025}=1,96$. Επιπλέον $\sqrt{n}=\sqrt{40} \simeq 6,32$, οπότε το ζητούμενο δ.ε. είναι το $\left(29.500-\frac{16.689}{6,32}\cdot 1,96,\ 29.500+\frac{16.689}{6,32}\cdot 1,96\right)=(24.324,30,\ 34.675,70)$

---

(4) (σελ. 134 βιβλίο) Αναφερόμενοι στα δεδομένα της προηγούμενης άσκησης, υποθέτουμε επιπλέον ότι στο δείγμα των 40 εργαζομένων περιέχονται 22 γυναίκες. Να εκτιμηθεί το δ.ε. στο οποίο κυμαίνεται το ποσοστό των εργαζομένων γυναικών στην εταιρία με πιθανότητα 95%.

Λύση | $1-a=0,95 \Rightarrow a=1-0,95=0,05 \Rightarrow$ $\frac{a}{2}=0,025 \Rightarrow z_{0,025}=1,96$. Επειδή αναζη-

Τώρα με δ.ε. για ποσοστό θα χρησιμοποιήσουμε τον τύπο (3). Το δειγματικό ποσοστό είναι $\hat{p} = \frac{22}{40} = 0,55$ και $s_{\hat{p}} = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0,55\cdot(1-0,55)}{40}}$ $= \sqrt{6,1875\cdot 10^{-3}} \simeq 0,079$. Άρα το ζητούμενο δ.ε. είναι το $(0,55-1,96\cdot 0,079,\ 0,55+1,96\cdot 0,079) = (0,40,\ 0,70)$ ή $(40\%,\ 70\%)$

---

5 (σελ. 137 βιβλίο) Ο διευθυντής Μάρκετινγκ μιας εταιρίας κινητής τηλεφωνίας θέλει να εκτιμήσει τη μέση διάρκεια των συνδιαλέξεων συγκεκριμένης ομάδας συνδρομητών (π.χ. ελεύθεροι επαγγελματίες ηλικίας 25-30 ετών) με απώτερο σκοπό την προώθηση ενός νέου είδους συμβολαίου κινητής τηλεφωνίας για νέους επαγγελματίες. Η επιθυμητή ακρίβεια της εκτίμησης του δ.ε. του μέσου είναι $e = \pm 2$ δευτερόλεπτα σε επίπεδο εμπιστοσύνης 95%. Παλαιότερη έρευνα με παρόμοιο αντικείμενο, αλλά για άλλη κατηγορία συνδρομητών, είχε δώσει εκτίμηση της τυπικής απόκλισης $s = 30$ δευτερόλεπτα. Τι μέγεθος δείγματος χρειάζεται για να έχουμε την επιθυμητή ακρίβεια;

Λύση | $1-a=0,95 \Rightarrow a=1-0,95=0,05$ $\Rightarrow \frac{a}{2}=0,025$. Άρα $z_{0,025}=1,96$. Εφόσον δεν γνωρίζουμε την πληθυσμιακή διασπορά και τα δείγματα σε αυτές τις έρευνες είναι συνήθως μεγάλα θα χρησιμοποιήσουμε τον τύπο (2.2) κατά τα γνωστά $e=\frac{s}{\sqrt{n}}\cdot z_{a/2} \Rightarrow \sqrt{n}=\frac{s}{e}\cdot z_{a/2} \Rightarrow n=\left(\frac{s}{e}\cdot z_{a/2}\right)^2$ $=\left(\frac{30}{2}\cdot 1,96\right)^2=864,36$. Άρα αν $n=865$ τότε με πιθανότητα 95% το δ.ε. που θα προκύψει θα περιέχει τον πληθυσμιακό μέσο με ακρίβεια ±2 sec.

---

⑥ (σελ. 138-139 βιβλίο) Αναφερόμενοι στα δεδομένα της προηγούμενης άσκησης, υποθέτουμε επιπλέον ότι ο διευθυντής Μάρκετινγκ της εταιρίας θέλει να εκτιμήσει και το ποσοστό των συνδιαλέξεων που αφορούν επαγγελματικές μόνο κλήσεις. Η επιθυμητή ακρίβεια της εκτίμησης του δ.ε. του ποσοστού είναι $e=\pm 0,01$ ή ±1% σε επίπεδο εμπιστοσύνης 95%.

(i) Τι μέγεθος δείγματος χρειάζεται για να έχουμε την επιθυμητή ακρίβεια όταν:

(α) Δεν δίνεται άλλη πληροφορία

(β) Από προηγούμενη έρευνα είναι γνωστό ότι το ποσοστό των επαγγελματικών κλήσεων είναι περίπου 80%.

(ii) Αν μετά από την ολοκλήρωση της έρευνας προέκυψε ποσοστό επαγγελματικών κλήσεων 80%, να υπολογισθεί η ακρίβεια της εκτίμησης.

Λύση $1-a=0,95 \Rightarrow a=1-0,95=0,05 \Rightarrow$ $\frac{a}{2}=0,025$. Δεδομένου ότι εργαζόμαστε με ποσοστά θα χρησιμοποιήσουμε τον τύπο (3). Ισχύει ότι $e=z_{a/2}\cdot s_{\hat{p}}=z_{a/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \Rightarrow$

$$\frac{e}{z_{a/2}}=\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \Rightarrow \left(\frac{e}{z_{a/2}}\right)^2=\frac{\hat{p}\cdot(1-\hat{p})}{n} \Rightarrow$$

$$n=\frac{\hat{p}\cdot(1-\hat{p})}{\left(\frac{e}{z_{a/2}}\right)^2}$$

(i) (α) $e=0,01$, $z_{a/2}=z_{0,025}=1,96$ και όταν δεν έχουμε πληροφορία για το δειγματικό ποσοστό, όπως εδώ, θα θέτουμε $\hat{p}=\frac{1}{2}=0,5$

διότι αυτή η τιμή μεγιστοποιεί το $\hat{p}\cdot(1-\hat{p})$ (και συνεπώς και το $s_{\hat{p}}$).

Έτσι προκύπτει ότι $n = \dfrac{0,5\cdot 0,5}{(5,102\cdot 10^{-3})^2} \simeq$

$$\frac{0,25}{26,03\cdot 10^{-6}} = \frac{25\cdot 10^{-2}}{26,03\cdot 10^{-6}} = \frac{25}{26,03}\cdot\frac{1}{10^{-4}}$$

$$\simeq 0,96\cdot 10^4 = 9.600$$

(B). Θεωρούμε $\hat{p}=0,8$, οπότε

$$n = \frac{0,8\cdot(1-0,8)}{\left(\frac{0,01}{1,96}\right)^2} = \frac{0,8\cdot 0,2}{26,03\cdot 10^{-6}} = \frac{0,16}{26,03\cdot 10^{-6}}$$

$$= \frac{16\cdot 10^{-2}}{26,03\cdot 10^{-6}} = \frac{16}{26,03}\cdot 10^4 \simeq 6.100$$

(ii) Θέτουμε πάλι $\hat{p}=0,8$ επειδή όμως η γνώση αυτή προέκυψε μετά την ολοκλήρωση της έρευνας θεωρούμε $n=9.600$ (και όχι 6.100), οπότε $e = z_{a/2}\cdot\sqrt{\dfrac{\hat{p}\cdot(1-\hat{p})}{n}}$

$$= 1,96\cdot\sqrt{\frac{0,8\cdot(1-0,8)}{9.600}} = 1,96\cdot\sqrt{\frac{0,8\cdot 0,2}{9.600}} =$$

$$1,96\cdot\sqrt{\frac{0,16}{9.600}} \simeq 1,96\cdot\sqrt{1,6667\cdot 10^{-5}} \simeq$$

$$1,96\cdot 4,0825\cdot 10^{-3} = 8,0017\cdot 10^{-3} \simeq 0,008 < 0,01$$