

## ΦΡΟΝΤΙΣΤΗΡΙΟ #5

Διαστήματα εμπιστοσύνης για μέσες τιμές

- #1 Σε ένα τ.δ. 70 παρατηρήσεων από έναν πληθυσμό που ακολουθεί την κανονική κατανομή, ο αριθμητικός μέσος είναι 26.2 και η τυπική απόκλιση είναι 4.1 (αφερόληπτη <sup>επιμέτρησα</sup>).
- (α) Κατασκευάστε ένα 95% δ.ε. για τη μέση τιμή του πληθυσμού  $\mu$ .
- (β) Κατασκευάστε ένα 99% δ.ε. για τη μέση τιμή του πληθυσμού  $\mu$ .
- (γ) Πως μεταβάλλεται το εύρος του δ.ε. όταν ο συντελεστής εμπιστοσύνης αυξάνεται ενώ το μέγεθος του δείγματος παραμένει σταθερό;
- (δ) Δώστε μία ερμηνεία για τη φράση "είμαστε 95% σίγουροι ότι το διάστημα περιέχει την τιμή  $\mu$ ".
- (ε) Θα ήταν αξιόπιστα τα δ.ε. στα δύο πρώτα ερωτήματα αν η κατανομή του πληθυσμού δεν ήταν κανονική;
- (στ) Κατασκευάστε ένα 95% δ.ε. για τον μέσο  $\mu$ , για την περίπτωση όπου το μέγεθος του δείγματος είναι 20.

Λύση (α) Έχουμε  $n = 70$ ,  $\bar{X}_{70} = 26.2$ ,  $S^* = 4.1$

Ισχύει ότι  $\bar{X}_n \overset{\text{Κ.Ο.Θ}}{\sim} N\left(\mu, \frac{4.1^2}{70}\right)$

Ένα 95% δ.ε. για τον παράμετρο  $\mu$  είναι:

$$26.2 \pm \underset{\text{"}}{1.96} \cdot \frac{4.1}{\sqrt{70}} = 26.2 \pm 0.96 \Rightarrow (25.23, 27.16)$$

$$Z_{1-\frac{0.05}{2}} = Z_{0.975}$$

(β) Παρόμοια, θα έχουμε ένα 99% δ.ε. για το  $\mu$ :

$$26.2 \pm 2.57 \cdot \frac{4.1}{\sqrt{70}} = 26.2 \pm 1.25$$

$$Z_{1 - \frac{0.01}{2}} = Z_{0.99} \Rightarrow (24.95, 27.45)$$

(δ) Το δ.ε. θα είναι ευρύτερο (θα έχει μεγαλύτερο μήκος).

(δ) Αν κατασκευάσουμε έναν μεγάλο αριθμό δειγμάτων ίδιου μεγέθους από τον πληθυσμό, και για κάθε δείγμα λάβουμε ένα αντίστοιχο δ.ε. για την παράμετρο  $\mu$ , 95% αυτών των δ.ε. θα περιέχουν την άγνωστη σε μας αλλά σταθερή τιμή  $\mu$  της μέσης τιμής του πληθυσμού.

(ε) Τα δ.ε. θα ήταν αξιόπιστα διότι τα μεγέθη των δειγμάτων είναι "μεγάλα" (αρκετά).

(στ) Ένα 95% δ.ε. για την παράμετρο  $\mu$ :

(Μικρό μέγεθος δείγματος  $n=20$ )

$$26.2 \pm t_{19, \frac{0.05}{2}} \cdot \frac{4.1}{\sqrt{20}} \quad (\text{Χρήση της } t_{n-1} \text{-κατανομής})$$

$$26.2 \pm 2.093 \frac{4.1}{\sqrt{20}} \Rightarrow (25.175, 27.225).$$

#2 Δύο εταιρείες που παρέχουν παρόμοιες υπηρεσίες συγχρίνονται όσον αφορά τον χρόνο ανταπόκρισης (σε ημέρες) σε παράπονα. Ένα τ.δ. από πρόσφατα παράπονα σε αυτές τις εταιρείες έδωσε τα ακόλουθα αποτελέσματα:

|            | Μέγεθος δείγματος | Μέσος δείγματος | Τυπική απόκλιση δείγματος |
|------------|-------------------|-----------------|---------------------------|
| Εταιρεία Α | 12                | 8,5             | 3,6                       |
| Εταιρεία Β | 10                | 4,8             | 2,1                       |

Δώστε ένα 95% δ.ε. για την πραγματική διαφορά στους -3- χρόνους ανταπόκρισης. Χρησιμοποιήστε το διάστημα αυτό για να αποφασίσετε αν μία εταιρεία είναι πραγματικά πιο γρήγορη από την άλλη.

Λύση

Ζητάμε ένα 95% δ.ε. για τη διαφορά  $\mu_A - \mu_B$  στους μέσους χρόνους ανταπόκρισης στους πληθυσμούς των δύο εταιρειών Α και Β. Η διαφορά στους μέσους  $\mu_A, \mu_B$  μπορεί να επιτηρηθεί από τη δειγματική διαφορά  $\bar{X}_A - \bar{X}_B$ .

Η τυπική απόκλιση της διαφοράς  $\bar{X}_A - \bar{X}_B$  είναι:

$$\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

Επειδή  $\sigma_A^2, \sigma_B^2$  είναι άγνωστες και επειδή τα μεγέθη των δειγμάτων είναι μικρά, η τυπική απόκλιση της διαφοράς  $\bar{X}_A - \bar{X}_B$

είναι:  $S \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$

όπου  $S = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}}$

(pooled standard deviation)

Ένα 95% δ.ε. για τη διαφορά  $\mu_A - \mu_B$  είναι:

$$\bar{X}_A - \bar{X}_B \pm t_{n_A + n_B - 2; \frac{0.05}{2}} \cdot S \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$8.5 - 4.8 \pm t_{12 + 10 - 2, \frac{0.05}{2}} \cdot S \sqrt{\frac{1}{12} + \frac{1}{10}}$$

Είναι:  $t_{20, \frac{0.05}{2}} = 2.085$

$$\text{και } S = \sqrt{\frac{(12-1)3.6^2 + (10-1)2.1^2}{12+10-2}} = 3.0187$$

Τελικά, ένα 95% δ.ε. για τη διαφορά  $\mu_A - \mu_B$

$$\text{είναι: } 1.005 < \mu_A - \mu_B < 6.394$$

Επειδή το 0 δεν περιέχεται στο διάστημα εμπιστοσύνης, μπορούμε να πούμε ότι υπάρχουν ενδείξεις ότι η εταιρεία Β είναι γρηγορότερη (ταχύτερη) στις ανταποκρίσεις της στα παράπονα των πελατών της (διότι  $\mu_A - \mu_B > 0$ ).

#3 Ο χρόνος  $X$  <sup>(σε λεπτά)</sup> που χρειάζεται ένα ταξί για να πάει από το

Σύνταγμα στο αεροδρόμιο στη διάρκεια της νύχτας ακολουθεί την  $N(22, 4.14)$ .

(α) Να υπολογισθεί το διάστημα στο οποίο θα βρισκεται ο χρόνος μιας τέτοιας διαδρομής με πιθανότητα 99%.

(β) Να υπολογισθεί το διάστημα στο οποίο θα βρισκεται ο μέσος χρόνος που κάνει ένα ταξί σε 8 τέτοιες κούρσες.

(γ) Σε τυχαίο δείγμα  $n=25$  δρομολογίων που έγιναν στην ίδια διαδρομή υπολογίστηκε μέσος χρόνος  $\bar{x}=16$  λεπτά. Αν η διακύμανση του χρόνου μπορεί να θεωρηθεί ίση με  $\sigma^2=4.14$ , να ευριχθεί δ.ε. με πιθανότητα 99% για το μέσο χρόνο της διαδρομής.

Λύση (α) Το 3ητούμενο δ.ε. προκύπτει από τη σχέση:

$$\mu - \sigma z_{1-\alpha/2} < X < \mu + \sigma z_{1-\alpha/2}$$

$$\text{Για } 1-\alpha=0.99 \Rightarrow \alpha=0.01 \Rightarrow$$

$$\Rightarrow \frac{\alpha}{2}=0.005 \Rightarrow 1-\frac{\alpha}{2}=0.995$$

$$[\text{διότι } X \sim N(\mu, \sigma^2)]$$

$$\Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Έχουμε:  $z_{1-\alpha/2} = z_{0.995} = 2.58$ .

Επίσης  $\sigma = \sqrt{4.14} = 2.03$ . Συνεπώς, το ζητούμενο διάστημα στο οποίο θα βρίσκεται ο χρόνος μιας τέτοιας διαδρομής με πιθανότητα 99% είναι:

$$22 - 2.58(2.03) < X < 22 + 2.58(2.03)$$

$$\Rightarrow 16.7626 < X < 27.2374$$

(β) Ισχύει  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow \bar{X} \sim N(22, 0.5175)$

Στην περίπτωση αυτή έχουμε γνωστή διακύμανση, οπότε ανεξάρτητα από το μέγεθος του δείγματος, μπορούμε να γράψουμε:

$$\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu - \bar{X} \leq z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq -\bar{X} \leq -\mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \geq \bar{X} \geq \mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}},$$

δηλαδή επιτυγχάνουμε το δ.ε. της μέσης τιμής του δείγματος  $\bar{X}$  γνωρίζοντας τη μέση τιμή του πληθυσμού  $\mu$ .

Για  $1-\alpha = 0.99 \Leftrightarrow \alpha = 0.01 \Leftrightarrow \frac{\alpha}{2} = 0.005 \Leftrightarrow$

$\Leftrightarrow 1 - \frac{\alpha}{2} = 0.995$  έχουμε  $z_{1-\alpha/2} = z_{0.995} = 2.58$

Έχουμε επίσης  $\sigma_{\bar{X}} = \sqrt{0.5175} = 0.7193$ , οπότε το διάστημα στο οποίο θα βρίσκεται ο μέσος χρόνος

που δάνει ένα ταξί σε  $n=8$  μούσες είναι:

-6-

$$22 - 2.58(0.7194) < \bar{X} < 22 + 2.58(0.7194)$$

$$\Rightarrow 20.144 < \bar{X} < 23.856$$

(δ) Το ζητούμενο δ.ε. υποδίδεται ως εξής:

$$\bar{X} - z_{1-\alpha/2} \sigma_{\bar{X}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \sigma_{\bar{X}}$$

δηλ. στην περίπτωση αυτή ευτιμούμε τη μέση τιμή του πληθυσμού  $\mu$  συγκρίνοντας τη μέση του δείγματος  $\bar{X}$ .

Από τα δεδομένα του προβλήματος έχουμε:

$$z_{1-\alpha/2} = z_{0.995} = 2.58 \text{ και}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.035}{\sqrt{25}} = 0.407$$

Συνεπώς, το ζητούμενο δ.ε. είναι:

$$16 - 2.58(0.407) \leq \mu \leq 16 + 2.58(0.407)$$

$$\Rightarrow 16 - 1.05 \leq \mu \leq 16 + 1.05$$

$$\Rightarrow 14.95 \leq \mu \leq 17.05$$

#4 Ένα σύστημα εξάσκησής στη χρήση πινάκων της τυποποιημένης κανονικής κατανομής υποτίθεται ότι βελτιώνει την απόδοση των φοιτητών. Προκειμένου να ελεγχθεί αυτό πέντε τυχαία επιλεγμένοι φοιτητές από μία μεθήκη τάξη στατιστικής ψάχνουν μία τιμή, επιλεγμένη στην τύχη, σε ένα τέτοιο πίνακα. Οι χρόνοι τους (σε δευτερόλεπτα) καταγράφονται. Ύστερα από μία περίοδο εφαρμογής του προγράμματος το πείραμα επαναλαμβάνεται. Τα αποτελέσματα είναι:

Πριν 16 15 16 13 15  
Μετά 13 15 14 12 13

Υπολογίστε ένα 99% δ.ε. για τη μέση μείωση στον χρόνο εύρεσης των τιμών εξαιτίας του προγράμματος. Τι ενδείξεις υπάρχουν για το πρόγραμμα εξάσκησης;

Λύση Έχουμε ζεύγη παρατηρήσεων.

Πριν(B) 16 15 16 13 15  
Μετά(A) 13 15 14 12 13  
B-A 3 0 2 1 2

$$\bar{d} = \frac{3+0+2+1+2}{5} = \frac{8}{5} = 1.6$$

$$\text{Τυπική απόκλιση} = \frac{\sum_{i=1}^5 (d_i - \bar{d})^2}{5-1} = \frac{\sum_{i=1}^5 d_i^2 - 5\bar{d}^2}{5-1} = \text{S.D.}$$

$$= 1.14018$$

Ένα 99% δ.ε. για τη (πραγματική) διαφορά των μέσων  $\mu_A - \mu_B$  είναι:  $\bar{d} \pm t_{5-1, 0.01/2} \frac{\text{S.D.}}{\sqrt{5}}$

$$\Rightarrow 1.6 \pm 4.604 \cdot 0.5099 \Rightarrow (-0.748, 3.948)$$

Το δ.ε. περιέχει το 0. Υπάρχουν ενδείξεις ότι δεν φαίνεται να έχουμε στατιστικά σημαντική διαφορά μέσους χρόνων εύρεσης των τιμών πριν και μετά το πρόγραμμα εξάσκησης.

Διαστήματα εμπιστοσύνης για ποσοστά

#5 Ένα τ.δ. από 120 άτομα επιλέγεται από έναν μεγάλο πληθυσμό. Πενήντα από αυτά τα άτομα έχουν ένα συγκεκριμένο χαρακτηριστικό. Βρείτε ένα 95% διάστημα

εμπιστοσύνης για το ποσοστό αυτού του χαρακτηριστικού στον πληθυσμό χρησιμοποιώντας την προσέγγιση της κανονικής κατανομής.

Λύση

Έχουμε  $n=120$ ,  $X=50$

Η αναλογία  $p$  του πληθυσμού εκτιμάται από τη δειγματική αναλογία  $\frac{X}{n} = \frac{50}{120} = \frac{5}{12} = 0.417$

Ένα 95% δ.ε. για την αναλογία  $p$  είναι:

$$\frac{5}{12} \pm 1.96 \sqrt{\frac{\frac{5}{12} \cdot \frac{7}{12}}{120}}$$

όπου το εκτιμώμενο τυπικό σφάλμα είναι:

$$\sqrt{\frac{\frac{5}{12} \left(1 - \frac{5}{12}\right)}{120}} = 0.045$$

Άρα,  $0.4166 \pm 1.96 \cdot 0.045 \Rightarrow (0.3284, 0.5048)$ .

Σ