ΚΕΦΑΛΑΙΟ 6 - ΕΚΤΙΜΗΤΙΚΗ

Βασική αρχή στατιστικής:

Αποτέλεσμα = Πραγματική τιμή + Σφάλμα

Προσπαθούμε να εκτιμήσουμε τις παραμέτρους ενός πληθυσμού (π.χ τη μέση τιμή και τη διασπορά) επιλέγοντας ένα δείγμα και κοιτώντας την αντίστοιχη παράμετρο σε αυτό.

Συγκεκριμένα:

- Έχουμε έναν πληθυσμό που η κατανομή είναι γνωστή ή άγνωστη αλλά δεν ξέρουμε τις ακριβείς παραμέτρους.
- Επιλέγουμε ένα τυχαίο δείγμα $x_1, x_2, \dots x_n$

Μέχρι να δούμε τις ακριβείς τιμές, τα x_i θεωρούνται τυχαίες μεταβλητέ

- Επιλέγουμε τις παραμέτρους του πληθυσμού που θα εκτιμήσουμε

* - Επιλέγουμε μια συνάρτηση = εκτιμήτρια) εκτιμητής των $x_1, x_2 \dots x_n$ που δεν περιέχει την άγνωστη παράμετρο

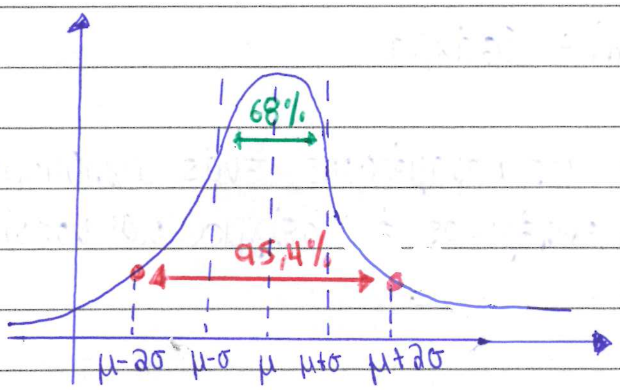
Απο την εκτιμήτρια θέλουμε

1) Να παίρνει τιμές κοντά στην πραγματική τιμή της παραμέτρου με μεγάλ πιθανότητα $\rightarrow$ η μέση τιμή της = άγνωστη παράμετρος

2) Οι τιμές της συνάρτησης να μην απέχουν πολύ από την μέση τιμή → μικρή διασπορά

Ορισμός: • Έστω Τ μια εκτιμήτρια της άγνωστης παραμέτρου θ
• Η Τ λέγεται αμερόληπτη αν $E(T) = \theta$
• Το σφάλμα της Τ είναι $b(T) = E(T) - \theta$

- Το τυπικό σφάλμα της Τ είναι η τυπική της απόκλιση
(Τυπικό σφάλμα = 68% σφάλμα περιθωρίου)
- Το 95,4% σφάλμα περιθωρίου είναι το διπλάσιο της τυπικής απόκλισης της Τ



Εμείς θα ασχοληθούμε με εκτίμηση:
- μέσης τιμής
- διασποράς
- ποσοστού

Γενικά
Αν έχουμε αμερόληπτες εκτιμήτριες επιλέγουμε αυτήν με την μικρότερη διασπορά

Παράδειγμα: Έστω δείγμα X_1, X_2, X_3, X_4 με $E(X_i) = \mu$ και $V(X_i) = \sigma^2$
Ποια είναι η καταλληλότερη εκτιμήτρια του μ από τις:

• $T_1 = \frac{X_1+X_2+X_3}{3}$ • $T_2 = 2X_1 - 2X_2$ • $T_3 = 3X_1 - X_2$ • $T_4 = X_1 + X_2 - \mu$

— Η T_4 απορρίπτεται επειδή περιέχει την άγνωστη παράμετρο μ

Δίνονται: $E(aX + bY) = aE(X) + bE(Y)$

$V(aX+bY) = a^2 V(X) + b^2 V(Y)$

• $E(T_1) = E\left(\frac{X_1+X_2+X_3}{3}\right) = \frac{E(X_1)+E(X_2)+E(X_3)}{3} = \frac{\mu+\mu+\mu}{3} = \mu$ (αμερόληπτη)

• $E(T_2) = E(2X_1 - X_2) = 2E(X_1) - E(X_2) = 2\mu - \mu = \mu$ (αμερόληπτη)

• $E(T_3) = E(3X_1 - X_2) = 3E(X_1) - E(X_2) = 3\mu - \mu = 2\mu$ → δεν είναι αμερόληπτη

⇒ Από τις αμερόληπτες καλύτερη είναι αυτή με τη μικρότερη διασπορά

• $V(T_1) = V\left(\frac{X_1+X_2+X_3}{3}\right) = \frac{V(X_1)+V(X_2)+V(X_3)}{9} = \frac{\sigma^2+\sigma^2+\sigma^2}{9} = \frac{3}{9}\sigma^2$

• $V(T_2) = V(2X_1 - X_2) = 2^2 V(X_1) + (-1)^2 V(X_2) = 4\sigma^2 + \sigma^2 = 5\sigma^2$

⇒ Η T_1 είναι η καλύτερη εκτιμήτρια

Θεώρημα: Έστω $X_1, X_2 \dots X_n$ τυχαίο δείγμα από οποιοδήποτε πληθυσμό με μέση τιμή μ και διασπορά σ^2

1) Η $\bar{X} = \frac{X_1 + \dots X_n}{n}$ έχει μέση τιμή μ (δηλ. είναι αμερόληπτη εκτιμήτρια του μ)

2) Η $S^2 = \frac{1}{n-1}(\Sigma x_i^2 - n\bar{X}^2)$ έχει μέση τιμή σ^2 (δηλ. είναι αμερόληπτη εκτιμήτρια του σ^2)

$E(S^2) = \mu$

Άγνωση Παραμ.	Εκτιμήτρια	Τυπικό σφάλμα
• μ	$\bar{X} = \frac{x_1 + x_2 \dots x_n}{n}$	$\sigma / \sqrt{n}$ *
• σ^2	$S^2 = \frac{1}{n-1}(\Sigma x_i^2 - n\bar{x}^2)$	
• p	X/n	$\sqrt{\frac{\frac{X}{n} \cdot \frac{n-X}{n}}{n}}$

όπου X το πλήθος των ατόμων του δείγματος με το επιθυμητό χαρακτηριστικ

* Αν το σ είναι άγνωστο, χρησιμοποιούμε το $\frac{S}{\sqrt{n}}$ όπου $S = \sqrt{S^2}$

Παράδειγμα: Θέλουμε να εκτιμήσουμε το ποσοστό των φοιτητών που έχουν Instagram.
Ρωτήσαμε 43 φοιτητές και πήραμε καταφατική απάντηση από τους 41
• Ποια είναι η εκτίμηση για το πραγματικό ποσοστό;
• Ποιο το τυπικό σφάλμα της εκτίμησης;

→ $\frac{X}{n} = \frac{41}{43} \cong 0,95 = 95\%$

→ Τυπικό σφάλμα εκτίμησης: $\sqrt{\frac{\frac{x}{n} \cdot \frac{n-x}{n}}{n}} = \sqrt{\frac{0,95 \cdot 0,05}{43}} \cong 0,032$

95,4% σφάλμα περιθωρίου είναι $2 \cdot 0,032 = 0,064$

Παράδειγμα

Ενας δάσκαλος θέλει να εκτιμήσει τη δυσκολία ενος διαγωνίσματος. Το δίνει σε 5 μαθητές οι οποίοι γράφουν 18, 16, 11, 20, 7

- Ποια αναμένεται να είναι η μέση τιμή των βαθμών των μαθητών;
- Ποιο το τυπικό σφάλμα της εκτίμησης;

→ Εκτίμηση μέσης τιμής : $\bar{X} = \frac{18+16+11+20+7}{5} = 15$

→ Τυπικό σφάλμα: $\frac{s}{\sqrt{n}} = \frac{5}{\sqrt{5}} = \sqrt{5}$

$$s^2 = \frac{18^2+16^2+11^2+20^2+7^2-2\cdot 15^2}{4} = 25 \Rightarrow s = \sqrt{25} = 5$$

ΕΚΤΙΜΗΤΙΚΗ

Εκτίμηση $\mu \rightarrow \bar{x}$
Εκτίμηση $\sigma^2 \rightarrow S^2$
Εκτίμηση $p \rightarrow \frac{x}{n} = \hat{p}$

Διαστήματα Εμπιστοσύνης (Δ.Ε)

Οι εκτιμήτριες που είδαμε, δίνουν έναν αριθμό για εκτίμηση

Προτιμάμε ένα διάστημα που με μεγάλη πιθανότητα περιέχει την άγνωστη παράμετρο

Κατασκευή Δ.Ε

1) Έχουμε άγνωστη παράμετρο θ
2) Επιλέγουμε εκτιμήτρια $\hat{\theta}$
3) Επιλέγουμε επίπεδο σημαντικότητας ή συντελεστή εμπιστοσύνης $(1-a)\%$
4) Κατασκευάζουμε διάστημα $[\hat{\theta}-d, \hat{\theta}+d]$, για κάποιον αριθμό d ώστε $P(\theta \in [\hat{\theta}-d, \hat{\theta}+d]) = 1-a$

Παράδειγμα

Έστω ότι έχουμε πληθυσμό με κανονική κατανομή στην οποία γνωρίζουμε το σ^2 αλλά όχι το μ.
Θέλουμε να κατασκευάσουμε Δ.Ε με συντελεστή $1-a$

Άγνωστη παράμετρος : μ
Παίρνουμε τυχαίο δείγμα $x_1, x_2, \dots x_n$ και χρησιμοποιούμε την εκτιμήτρια: $\bar{x} = \frac{x_1+x_2+\dots x_n}{n}$

Συντελεστής : $1-a$

Ψάχνουμε αριθμό d ώστε $P(\mu \in [\bar{X}-d, \bar{X}+d]) = 1-a$

Κάθε $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ (από το θεώρημα)

$$\Rightarrow \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

- $P(\mu \in [\bar{X}-d, \bar{X}+d]) = 1-a$

$$\Rightarrow P(\bar{X}-d \le \mu \le \bar{X}+d) = 1-a$$
$$\Rightarrow P(-d \le \mu-\bar{X} \le d) = 1-a$$
$$\Rightarrow P(d \ge \bar{X}-\mu \ge -d) = 1-a$$
$$\Rightarrow P(-d \le \bar{X}-\mu \le d) = 1-a$$

$$\Rightarrow P\left(\frac{-d}{\sqrt{\frac{\sigma^2}{n}}} \le \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \le \frac{d}{\sqrt{\sigma^2/n}}\right) = 1-a$$

$$\Rightarrow P\left(\frac{-d}{\sqrt{\sigma^2/n}} \le Z_1 \le \frac{d}{\sqrt{\sigma^2/n}}\right) = 1-a \Rightarrow \Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) - \Phi\left(\frac{-d}{\sqrt{\sigma^2/n}}\right) = 1-a$$

$$\Rightarrow \Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) - \left(1-\Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right)\right) = 1-a \Rightarrow 2\Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) - 1 = 1-a$$

$$\Rightarrow 2\Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) = 2-a \Rightarrow \Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) = 1-\frac{a}{2}$$

Έστω $z_{\frac{a}{2}}$ ο αριθμός για τον οποίο $\Phi(z_{\frac{a}{2}}) = 1-\frac{a}{2}$

Άρα $\Phi\left(\frac{d}{\sqrt{\sigma^2/n}}\right) = \Phi(z_{a/2})$

$$\Rightarrow \frac{d}{\sqrt{\sigma^2/n}} = z_{a/2} \Rightarrow \boxed{d = z_{a/2}\sqrt{\frac{\sigma^2}{n}}}$$

Άρα το Δ.Ε είναι

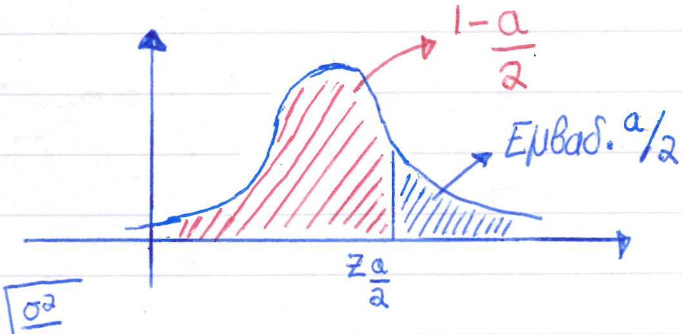
$$\left[\bar{X} - z_{\frac{a}{2}}\sqrt{\frac{\sigma^2}{n}},\ \bar{X} + z_{\frac{a}{2}}\sqrt{\frac{\sigma^2}{n}}\right] \text{ ή } \bar{X} \pm z_{a/2}\sqrt{\frac{\sigma^2}{n}}$$

Τι είναι το $z_{a/2}$

- Ορίζεται από τη σχέση $\Phi\left(z_{\frac{a}{2}}\right) = 1 - \frac{a}{2}$
- Άρα το βρίσκουμε από τον πίνακα καν. κατανομής ή θα δίνεται ξεχωριστός πίνακας

π.χ

a	0,1	0,05	0,02	0,01
$z_{a/2}$	1,645	1,96	2,33	2,575



95% ΔΕ για το μ $\bar{X} = \pm z_{\frac{a}{2}}\sqrt{\frac{\sigma^2}{n}}$

$a = 0{,}05$, $z_{a/2} = z_{0{,}025} = 1{,}96$

Γενικά έχουμε 2 περιπτώσεις για κατασκευή Δ.Ε

A) Μεγάλα δείγματα (≥ 30)
B) Μικρά δείγματα

A- Μεγάλα δείγματα

Λόγω του ΚΟΘ έχουμε πάντα κανονική κατανομή

- ΔΕ για το μ με γνωστό σ^2 $\bar{X} \pm z_{a/2}\sqrt{\frac{\sigma^2}{n}}$
- ΔΕ για το μ με άγνωστο σ^2 $\bar{X} \pm z_{a/2}\sqrt{\frac{s^2}{n}}$

• Δ.Ε για ποσοστό p $\frac{x}{n} \pm z_{a/2} \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$

Παράδειγμα

• Θέλουμε ΔΕ για τη μέση κατανάλωση φλιτζανιών καφέ ανά μήνα των φοιτητών του Π.Κ

• Επιλέγουμε δείγμα 64 φοιτητών, στο οποίο βρέθηκε μέση τιμή 59,5 και διασπορά 16

Άγνωστη παράμετρος: μ

Εκτιμήτρια: $\bar{X} = 59,5$

Άγνωστο σ^2, $S^2 = 16$

Δ.Ε για μ με άγνωστο σ^2 $\bar{X} \pm z_{a/2}\sqrt{\frac{S^2}{n}}$

90% Δ.Ε: $a = 0,1 \Rightarrow z_{a/2} = 1,645$

$59,5 \pm 1,645\sqrt{\frac{16}{64}} = [58,6775 , 60,3225]$

95% Δ.Ε: $a = 0,05 \Rightarrow z_{a/2} = 1,96$

$59,5 \pm 1,96\sqrt{\frac{16}{64}} = [58,52 , 60,48]$

99% Δ.Ε: $a = 0,01 \Rightarrow z_{a/2} = 2,575$

$59,5 \pm 2,575\sqrt{16/64} = [58,2125 , 60,7885]$

Ελάχιστο μέγεθος δείγματος

Θέλουμε να κατασκευάσουμε $(1-a)\%$ Δ.Ε για την άγνωστη παράμετρο με μήκος (εύρος) το πολύ δ και ψάχνουμε το μέγεθος του δείγματος

$\underset{a}{|}\!\!-\!\!\underset{b}{|}$ μήκος $= b - a$

Παράδειγμα:

$(1-a)\%$ ΔΕ για μ με γνωστό σ^2 $\quad \bar{X} \pm z_{a/2}\sqrt{\frac{\sigma^2}{n}}$

Πρέπει $\delta \geq \bar{x} + z_{a/2}\sqrt{\sigma^2/n} - \left(\bar{x} - z_{a/2}\sqrt{\frac{\sigma^2}{n}}\right)$

$\delta \geq \cancel{\bar{x}} + z_{a/2}\sqrt{\sigma^2/n} - \cancel{\bar{x}} + z_{a/2}\sqrt{\sigma^2/n} \quad \delta \geq 2 z_{a/2}\frac{\sigma}{\sqrt{n}}$

$\sqrt{n} \geq 2 z_{a/2}\frac{\sigma}{\delta} \Rightarrow n \geq \left(\frac{2\sigma z_{a/2}}{\delta}\right)^2$

Γενικός τύπος: $(1-a)\%$ ΔΕ μήκους δ για:

- μ με γνωστό σ^2: $n \geq \left(\frac{2\sigma z_{a/2}}{\delta}\right)^2$
- μ με άγνωστο σ^2: $n \geq \left(\frac{2 s z_{a/2}}{\delta}\right)^2$
- p: $n \geq 4\hat{p}(1-\hat{p})\left(\frac{z_{a/2}}{\delta}\right)^2$,

όπου $\hat{p} = \frac{x}{n}$

Παράδειγμα:

Να βρεθεί το μέγεθος του δείγματος που απαιτείται ώστε το 99% ΔΕ για το μέσο εισόδημα των νοικοκυριών να έχει εύρος €50 αν είναι γνωστό ότι η τυπική απόκλιση είναι €600

Τι γίνεται αν αλλάξουμε το εύρος σε €25

ΔΕ για μ με γνωστό σ^2

$$n \geq \left(\frac{2\sigma z_{a/2}}{\delta}\right)^2 = \left(\frac{2 \cdot 600 \cdot 1,96}{50}\right)^2 = 2212$$

ΔΕ με εύρος 25

$$n \geq \left(\frac{2\sigma z_{a/2}}{\delta}\right)^2 = \left(\frac{2 \cdot 600 \cdot 1,96}{25}\right)^2 = 8848$$

Τι κάναμε;

Διαστήματα Εμπιστοσύνης

Α) Μεγάλα δείγματα (≥ 30)

Ελάχιστο μέγεθος δειγμ. για ΔΕ μήκους δ

ΔΕ μ με γνωστό σ^2 $\quad \bar{x} \pm z_{a/2}\sqrt{\frac{\sigma^2}{n}}$ $\quad n \geq \left(\frac{2\sigma z_{a/2}}{\delta}\right)^2$

ΔΕ μ με άγνωστο σ^2 $\quad \bar{x} \pm z_{a/2}\sqrt{s^2/n}$ $\quad n \geq \left(\frac{2s z_{a/2}}{\delta}\right)^2$

ΔΕ για ποσοστό p $\quad \frac{x}{n} \pm z_{a/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$ $\quad n \geq 4\hat{p}(1-\hat{p})\left(\frac{z_{a/2}}{\delta}\right)^2$

Δ.Ε για δύο πληθυσμούς

Θέλουμε να συγκρίνουμε τις άγνωστες παραμέτρους δύο ανεξάρτητων πληθυσμών

1) Διαφορά μέσων

Έστω ότι ο πρώτος πληθυσμός έχει μέση τιμή μ_1 και ο δεύτερος πληθ. μ_2

Άγνωστη παράμετρος : $\mu_1 - \mu_2$

Παίρνουμε δείγματα $x_1, x_2, \dots x_{n_1}$ από τον πρώτο πληθυσμό και $y_1, y_2 \dots y_{n_2}$ από τον δεύτερο πληθυσμό

Εκτιμήτρια : $\bar{X} - \bar{Y}$ όπου $\bar{X} = \dfrac{x_1 + x_2 + \dots x_{n_1}}{n_1}$ και $\bar{Y} = \dfrac{y_1 + y_2 + \dots y_{n_2}}{n_2}$

$(1-a)\%$ ΔΕ για τα $\mu_1 - \mu_2$:

- με σ_1^2, σ_2^2 γνωστά: $\bar{X} - \bar{Y} \pm z_{a/2} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

- με σ_1^2, σ_2^2 άγνωστα: $\bar{X} - \bar{Y} \pm z_{a/2} \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$

όπου $S_1^2 = \dfrac{1}{n_1 - 1} \left(\Sigma x_i^2 - n_1 \bar{X}^2 \right)$

$$S_2^2 = \frac{1}{n_2 - 1} \left(\Sigma y_i^2 - n_2 \bar{Y}^2 \right)$$

Παράδειγμα

Θέλουμε να συγκρίνουμε τις ταχύτητες ενσύρματων και ασύρματων δικτύων

Στείλαμε 200 αρχεία ενσύρματα:
μέσος χρόνος: 1,92 msec
τυπική απόκλιση: 5,21 msec

Στείλαμε 121 αρχεία ασύρματα
μέσος χρόνος: 10,51 msec
τυπική απόκλιση: 8,22 msec.

Να κατασκευαστεί 90% ΔΕ για τη διαφορά των μέσων χρόνων αποστολής

Έστω μ_1 ο μέσος χρόνος αποστολής με ενσύρματα δίκτυα
Έστω μ_2 " " " " ασύρματα δίκτυα

Πήραμε δείγματα μεγέθους 200 και 121 αντίστοιχα

$\bar{X} = 1,92$ msec και $\bar{Y} = 10,51$ msec

σ_1^2, σ_2^2 άγνωστα $\quad S_1 = 5,21$ msec $\quad a = 10\% = 0,1$
$S_2 = 8,22$ msec $\quad z_{a/2} = z_{0,05} = 1,64$

$$\text{Δ.Ε}: \bar{X} - \bar{Y} \pm z_{a/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 1,92 - 10,51 \pm 1,64\sqrt{\frac{5,21^2}{200} + \frac{8,22^2}{121}}$$

$= [-9,95, -7,23]$

② Διαφορά ποσοστών

Θέλουμε να συγκρίνουμε τα ποσοστά των ατόμων με κάποιο χαρακτηριστικ σε δύο πληθυσμούς

Άγνωστη παράμετρος: $p_1 - p_2$

Εκτιμήτρια: $\frac{X}{n_1} - \frac{Y}{n_2} = \hat{p}_1 - \hat{p}_2$ όπου:

X = # ατόμων στο 1ο δείγμα με το χαρακτηριστικό

Y = # ατόμων » 2ο δείγμα » » ».

(1-α)% ΔΕ για το $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 \pm z_{a/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Παράδειγμα

Θέλουμε να συγκρίνουμε το ποσοστό ελαττωματικών προϊόντων που παράγουν δύο μηχανήματα.

Επιλέξαμε 400 προϊόντα από το πρώτο, 16 βρέθηκαν ελαττωματικά
» 300 προϊόντα από το δεύτερο, 24 » » ».

Να βρεθεί 99% Δ.Ε για τη διαφορά των ποσοστών των ελαττωματικών προϊόντων.

$(n_1 = 400)$ $(n_2 = 300)$

$\hat{p}_1 = \frac{16}{400} = 0{,}04$, $\hat{p}_2 = \frac{24}{300} = 0{,}08$, $a = 1\% = 0{,}01$

$z_{a/2} = z_{0{,}005} = 2{,}58$

ΔΕ: $\hat{p}_1 - \hat{p}_2 \pm z_{a/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$= 0{,}04 - 0{,}08 \pm 2{,}58 \sqrt{\frac{0{,}04 \cdot 0{,}96}{400} + \frac{0{,}08 \cdot 0{,}92}{300}} =$

$[-0{,}087, 0{,}0076]$

Β- Μικρά Δείγματα

Πλέον δεν μπορούμε να χρησιμοποιήσουμε το ΚΟΘ. Θα πρέπει να δίνεται ότι ο πληθυσμός ακολουθεί κανονική κατανομή.

Β1- ΔΕ για μ με, σ^2 γνωστό

Έχουμε πληθυσμό με κανονική κατανομή με άγνωστο μ και γνωστό σ^2
Θέλουμε $(1-a)\%$ ΔΕ για το μ.

Παίρνουμε ένα δείγμα $x_1, x_2 \dots x_n$ $(n<30)$ $\bar{X} = \dfrac{x_1+x_2+\dots x_n}{n}$

Άγνωστη παράμετρος: μ
Εκτιμήτρια: $\bar{X}$
ΔΕ για μ με γνωστό σ^2 $\bar{X} \pm z_{a/2}\sqrt{\sigma^2/n}$ (ίδιος τύπος για μεγάλα δείγματα)

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

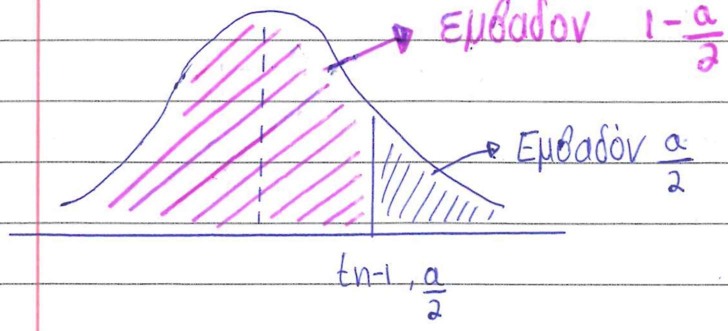
Β2- ΔΕ για μ με σ^2 άγνωστο

Εχουμε πληθυσμό με κανονική κατανομή και άγνωστα μ, σ^2

Παίρνουμε δείγμα $x_1, x_2, \dots x_n$ $(n<30)$ $\bar{x} = \dfrac{x_1+x_2+\dots x_n}{n}$

$S^2 = \dfrac{1}{n-1}(\Sigma x_i^2 - n\bar{X}^2)$ $\dfrac{\bar{X}-\mu}{S}$ δεν ακολουθεί κανονική κατανομή

Το $\frac{\bar{X}-\mu}{s}$ ακολουθεί κατανομή student με n-1 βαθμούς ελευθερίας (t_{n-1})



ΔΕ για μ με άγνωστο $σ^2$

$$\boxed{\bar{X} \pm t_{n-1,a/2}\sqrt{\frac{s^2}{n}}}$$

Παράδειγμα

Ενας αρχαιολόγος πήρε δείγμα 8 γυναικών της Μινωικής περιόδου. Βρήκε μέση τιμή 155 cm και τυπική απόκλιση 15 cm. Να βρεθεί 95% ΔΕ για το μέσο ύψος των γυναικών αυτής της περιόδου.
(Το ύψος ακολουθεί κανονική κατανομή)

ΔΕ για μ με άγνωστο $σ^2$ $\bar{X}=155$ cm, $s=15$ cm

$a=5\%=0,05$

$t_{n-1,\frac{a}{2}} = t_{7,\,0,025} = 2,365$

$$\bar{x} \pm t_{n-1,\frac{a}{2}}\sqrt{\frac{s^2}{n}} = 155 \pm 2,365\sqrt{\frac{15^2}{8}} = [142,46\ ;\ 167,54]$$

Διαστήματα Εμπιστοσύνης

Β - Μικρά Δείγματα

Β1 - ΔΕ για μ με γνωστό σ^2 $\bar{x} \pm z_{a/2} \frac{\sigma}{\sqrt{n}}$

Β2 - ΔΕ για μ με άγνωστο σ^2 $\bar{x} \pm t_{n-1,\frac{a}{2}} \frac{s}{\sqrt{n}}$

κατανομή student

Β3 - ΔΕ για διαφορά μέσων

Έχουμε δύο πληθυσμούς με κανονική κατανομή $N(\mu_1, \sigma_1^2)$ και $N(\mu_2, \sigma_2^2)$ αντίστοιχα (μ_1, μ_2 άγνωστα)

Παίρνουμε δείγματα

$x_1, x_2 \dots x_{n_1}$ από τον 1ο πληθυσμό

$y_1, y_2 \dots y_{n_2}$ από τον 2ο πληθυσμό

$$\bar{X} = \frac{x_1 + x_2 + \dots x_{n_1}}{n_1} \qquad \bar{Y} = \frac{y_1 + y_2 + \dots y_{n_2}}{n_2}$$

$$S_1 = \frac{1}{n_1 - 1}\left(\Sigma x_i^2 - n_1 \bar{x}^2\right) \qquad S_2 = \frac{1}{n_2 - 1}\left(\Sigma y_i^2 - n_2 \bar{y}^2\right)$$

Πρώτη περίπτωση: σ_1^2, σ_2^2 γνωστά

Το $(1-a)\%$ για το $\mu_1-\mu_2$ είναι: $\boxed{\bar{X}-\bar{Y} \pm z_{a/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

Δεύτερη περίπτωση:

σ_1^2, σ_2^2 άγνωστα αλλά $\sigma_1^2=\sigma_2^2$

Ορίζουμε την ποσότητα: $S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

Το $(1-a)\%$ για το $\mu_1-\mu_2$ είναι: $\boxed{\bar{X}-\bar{Y} \pm t_{n_1+n_2-2,\frac{a}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Παρατήρηση:

Αν $0,5 \leq \frac{S_1}{S_2} \leq 2$ τότε θα θεωρούμε ότι $\sigma_1^2=\sigma_2^2$

Παράδειγμα

Οι χρόνοι εξυπηρέτησης δύο υπαλλήλων μιας τράπεζας ακολουθούν κανονική κατανομή $N(\mu_1, 40^2)$ και $N(\mu_2, 40^2)$ αντίστοιχα.
Πήραμε δείγμα χρόνων εξυπηρέτησης σε sec:

1η υπάλληλος: 234, 99, 234, 174, 107, 173, 172
2η υπάλληλος: 105, 194, 77, 33, 159, 150, 167, 127, 169, 166

Να κατασκευαστεί 95% ΔΕ για τη διαφορά των μέσων χρόνων εξυπηρέτησης

ΔΕ για $\mu_1-\mu_2$ με γνωστό σ_1^2, σ_2^2, μικρό δείγμα

$\bar{X}-\bar{Y} \pm z_{a/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

$\bar{X} = \frac{234+99+234+174+107+173+172}{7} = 172,625$

$\bar{Y} = \frac{105+194+77+33+159+150+167+127+169+166}{10} = 134,7$

$a=0,05$, $z_{a/2} = z_{0,025} = 1,96$

$\sigma_1^2 = 40^2 = \sigma_2^2$

$n_1=7$, $n_2=10$

Αρα το ΔΕ είναι: $172,625 - 134,7 \pm 1,96\sqrt{\frac{40^2}{7}+\frac{40^2}{10}} = [0,73 , 75,11]$

Παραδειγμα:

Αρχαιολόγος πήρε δείγμα ανθρώπων από την Μινωική περίοδο

7 ανδρών με μέσο ύψος 165 cm και τυπ. αποκλ. 17 cm
8 γυναικών με μέσο ύψος 155 cm και τυπ. απ. 15 cm

Να δοθεί 99% ΔΕ για την διαφορά του μέσου ύψους ανδρων-γυναικών
(Το ύψος ακολουθεί κανονική κατανομή)

ΔΕ για $\mu_1-\mu_2$ με άγνωστα σ_1^2, σ_2^2, μικρό δείγμα

$\frac{S_1}{S_2} = \frac{17}{15} = 1,13$

Εφόσον $0,5 \leq 1,13 \leq 2$ υποθέτουμε ότι $\sigma_1^2 = \sigma_2^2$

$\bar{X}-\bar{Y} \pm t_{n_1+n_2-2,\frac{a}{2}} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$\bar{X}=165$

$\bar{Y}=155$

$a=0,01$

$\Rightarrow t_{n_1+n_2-2,\frac{a}{2}} = t_{13,0,005} = 3,012$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{6\cdot 17^2+7\cdot 15^2}{13}} \cong 15,954$$

Άρα το ΔΕ είναι: $165-155 \pm 3,012 \cdot 15,954\sqrt{\frac{1}{7}+\frac{1}{8}}$

$= [-14,892 , 34,8917]$

Παράδειγμα

Ένας ιδιοκτήτης καταστήματος ενδιαφέρεται για τη μέση κατανάλωση των πελατών του. Πήρε δείγμα 16 αγορών $x_1, x_2 \dots x_{16}$ και προέκυψε ότι $\sum_{i=1}^{16} x_i = 800$ και $\sum_{i=1}^{16} x_i^2 = 40240$

α) Να βρεθεί η μέση τιμή, η διασπορά και η τυπική απόκλιση των 16 αγορών.

β) Να κατασκευαστεί 90% και 95% ΔΕ για τη μέση κατανάλωση, αν γνωρίζουμε ότι ακολουθεί κανονική κατανομή.

α) $\bar{X} = \frac{\Sigma x_i}{n} = \frac{800}{6} = 50$

$S^2 = \frac{1}{n-1}(\Sigma x_i^2 - n\bar{x}^2) = \frac{1}{15}(40240 - 16\cdot 50^2) = 16$

$S = \sqrt{S^2} = 4$

β) ΔΕ για μ με άγνωστο σ^2, μικρό δείγμα

$\bar{x} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$

90% $\alpha = 0,1$
$t_{15,\,0,05} = 1,753$ $\Big\}$ $50 \pm 1,753\,\frac{4}{\sqrt{16}} = [48,247 , 51,753]$

95% ΔΕ: $a=0,05$, $t_{n-1,\frac{a}{2}} = t_{15,0,025} = 2,131$

$\Rightarrow 50 \pm 2,131 \frac{4}{\sqrt{16}} = [47,869 , 52,131]$

γ) Σωστό / Λάθος / Δεν γνωρίζω (χωρίς πράξεις)
Το ποσό των € 52 ανήκει στο:
1) 87% ΔΕ για τη μέση κατανάλωση
2) 97% ΔΕ " " "
3) 92% ΔΕ " " "

i) Λ γιατί το 52 δεν ανήκει στο 90% ΔΕ
(όσο μικραίνει το ποσοστό, τόσο στενεύει το διάστημα)
ii) Σ γιατί το 52 ανήκει στο 95% ΔΕ
iii) Δεν γνωρίζω (χωρίς πράξεις)

δ) Από το δεύτερο κατάστημα, πήραμε δείγμα 36 αγορών και ο μέσος όρος βρέθηκε € 40 και η τυπική απόκλιση €12

Να κατασκευαστεί 99% ΔΕ για τη μέση κατανάλωση από το 2ο κατάστημα

ΔΕ για μ με άγνωστο $σ^2$, μεγάλο δείγμα
$\bar{y} \pm z_{a/2} \frac{s'}{\sqrt{n}}$

$\bar{y}=40$, $s'=12$, $a=0,01$ $z_{a/2=0,005} = 2,576$

$\Rightarrow 40 \pm 2,576 \cdot \frac{12}{\sqrt{36}}$: $[34,848 , 45,152]$

ε) Να υπολογιστεί το ελάχιστο μέγεθος του δείγματος ώστε το 99 % ΔΕ για τη μέση κατανάλωση του 2ου καταστήματος να έχει μήκος το πολύ €5.

Μήκος του ΔΕ

$$\bar{y} + z_{a/2} \frac{s}{\sqrt{n}} - \left(\bar{y} - z_{a/2} \frac{s}{\sqrt{n}} \right) = \frac{2 s z_{a/2}}{\sqrt{n}}$$

$$\Rightarrow \frac{2 s z_{a/2}}{\sqrt{n}} \leq 5 \Rightarrow n \geq \left(\frac{2 s z_{a/2}}{5} \right)^2$$

$$\Rightarrow n \geq \left(\frac{2 \cdot 12 \cdot 2,576}{5} \right)^2 \Rightarrow \boxed{n \geq 152,89}$$

Άρα πρέπει να πάρουμε δείγμα 153 ατόμων