5ο online μάθημα. (5)
8/4/2020

ΕΚΤΙΜΗΣΗ σε Διάστημα

Ποια η αναγκαιότητα;

Γιατί;;

Αναφορά σε συγκεκριμένα παραδείγματα

Διάστημα εμπιστοσύνης

Έστω $X_1, X_2 \dots X_n$ τδ από ένα πληθυσμό με

X συνεχής ← σ.π.π. ή σ.π. → X διακριτή $f(x;\theta)$ $\theta \in \Theta$.

Έστω $L = L(X_1 \dots X_n)$ και $U = U(X_1 \dots X_n)$

το κάτω και άνω άκρο ενός διαστήματος με τιμές

$l(X_1 \dots X_n)$ και $U(X_1 \dots X_n)$

Τι θα πρέπει να έχει για να είναι το διάστημα αποδεκτό;

Συζήτηση:

- Θα πρέπει το διάστημα να έχει όσο γίνεται μικρότερο μήκος (εύρος), δηλαδή θα πρέπει το U να μην απέχει πολύ από L

Επιπλέον θα πρέπει: $P(L < \theta < U) = 1 - \alpha$

δηλαδή το (L, U) να περιέχει το θ ένα μεγάλο ποσοστό φορών $(100(1-\alpha)\%)$

Το ποσοστό αυτό λέγεται **βαθμός εμπιστοσύνης**

Έτσι λέμε ότι έχουμε Δ.Ε. 90% ή 99% ή 95%

(6)

ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΚΑΝΙΚΟ ΠΛΗΘΥΣΜΟ

$X_1 \dots X_n$ τδ $N(\mu, \sigma^2)$

Εστω οτι ενδιαφερομαστε για το μ.

Εχουμε ήδη πει οτι $\bar{X}$ αμερόληπτο και επιπλέον για κανονικό πληθυσμό $\bar{X}$ είναι Α.Ο.Ε.Δ.

ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ για μ

Θα διακρίνουμε δυο περιπτώσεις

1η σ^2 γνωστό

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

με μονο άγνωστο τον μ!!!!

2η σ^2 άγνωστο

τοτε
$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

ΠΡΟΣΟΧΗ: Η κατασκευή Δ.Ε. ξεκινά από την εύρεση μιας σ.σ. που περιέχει το άγνωστο (εδω το μ) και γνωστές ποσότητες, με γνωστή κατανομή

Αρα οταν σ^2 γνωστό ξεκινώ από

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

ενώ όταν σ^2 άγνωστο ξεκινώ από

$$Q = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Έπειτα η ιδέα είναι:

$$P(q_1 \le Q \le q_2) = 1-a$$

ξεκινώντας από των παραπάνω να απομονώσω το μ

ΔΗΛΑΔΗ π.χ

$$P\left(q_1 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le q_2\right) = 1-a \Rightarrow$$

$$P\left(q_1 \frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le q_2 \frac{\sigma}{\sqrt{n}}\right) = 1-a \Rightarrow$$

$$P\left(\bar{X} - q_2 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - q_1 \frac{\sigma}{\sqrt{n}}\right) = 1-a$$

Όμοια

$$P\left(\bar{X} - q_2 \frac{s}{\sqrt{n}} \le \mu \le \bar{X} - q_1 \frac{s}{\sqrt{n}}\right) = 1-a$$

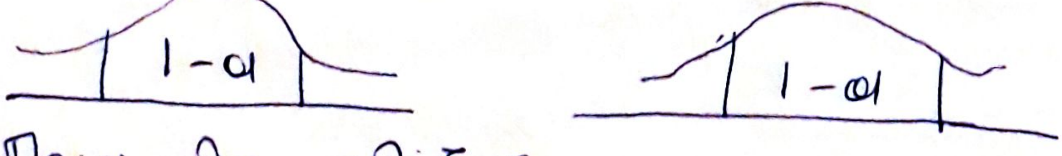
Το θέμα είναι πως προσδιορίζω τα q_1, q_2 έτσι ώστε να ισχύει $P(q_1 \le Z \le q_2) = 1-a$

όταν $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, με $Z \sim N(0,1)$

ή $P(q_1 \le t_{n-1} \le q_2) = 1-a$

όταν $t_{n-1} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$

Πρώτη παρατήρηση: q_1 & q_2 δεν είναι μοναδικά



Ποια θα επιλέξω?

Μιλάμε για Διαστήματα Ελάχιστου Μήκους όταν q_1, q_2 προσδιορίζονται έτσι ώστε το μήκος: $q_2 - q_1$ να ελαχιστοποιείται. Επιτυγχάνεται πάντα?
Στην περίπτωση του Δ.Ε. για την μέση τιμή αυτό επιτυγχάνεται είτε για σ^2 γνωστή είτε για σ^2 άγνωστη.
Ειδικότερα έχουμε:

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

ή

$$\left[\bar{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}},\ \bar{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Δ.Ε. για τη διαφορά $\mu_x - \mu_y$ κανονικών πληθυσμών

I Ανεξ/τα δείγματα

$$\left.\begin{array}{l} X_1, X_2, \dots X_n \text{ τδ } N(\mu_x, \sigma^2_x) \\ Y_1, Y_2, \dots Y_m \text{ τδ } N(\mu_y, \sigma^2_y) \end{array}\right\} \text{ ανεξ/τα}$$

- σ^2_x, σ^2_y γνωστές

Τότε
$$\left.\begin{array}{l} \bar{X} \sim N(\mu_x, \sigma^2_x/n) \\ \bar{Y} \sim N(\mu_y, \sigma^2_y/m) \end{array}\right\} \text{ ανεξ/εία}$$

$$(\bar{X} - \bar{Y}) \sim N\left(\mu_x - \mu_y, \frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}\right)$$

και τότε: (Βλέπε προηγούμενη διάλεξη)

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}} \sim N(0,1)$$

Επομένως το Δ.Ε. $100(1-\alpha)\%$ για τη διαφορά $\mu_X - \mu_Y$ ελαχίστου μήκους

$$\left[(\bar{X}-\bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \, , \, (\bar{X}-\bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \right]$$

όπου σ_X^2, σ_Y^2 άγνωστες αλλά ίσες.

Τότε $\quad X_1, \dots, X_n \sim N(\mu_X, \sigma^2)$

$\quad\quad\quad Y_1, \dots, Y_m \sim N(\mu_Y, \sigma^2)$

Είναι

$$\left.\begin{aligned} \bar{X} &\sim N\left(\mu_X, \frac{\sigma^2}{n}\right) \\ \bar{Y} &\sim N\left(\mu_Y, \frac{\sigma^2}{m}\right) \end{aligned}\right\} \text{ ανεξαρτησία}$$

$$\frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1)$$

$$\left.\begin{aligned} \frac{(n-1)S_X^2}{\sigma^2} &\sim \chi^2_{n-1} \\ \frac{(m-1)S_Y^2}{\sigma^2} &\sim \chi^2_{m-1} \end{aligned}\right\} \text{ Βλέπε και προηγούμενη διάλεξη.}$$

$$\frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

Άρα

$$t = \frac{\dfrac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \Big/ (n+m-2)}} \sim t_{n+m-2}$$

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

όπου $$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

Επομένως $$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Και το Δ.Ε. ελαχίστου μήκους 100(1-α)% είναι:

$$(\bar{x} - \bar{y}) \pm t_{n+m-2,\, a/2}\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

II Έστω τώρα ότι τα δείγματα εξαρτημένα;

Ερώτηση: Πότε;

Συζήτηση: ΠΡΙΝ - ΜΕΤΑ

$$\left.\begin{array}{lll} X_1, \ldots, X_n & \text{τ.δ.} & N(\mu_X, \sigma_X^2) \\ Y_1, \ldots, Y_n & \text{τ.δ.} & N(\mu_Y, \sigma_Y^2) \end{array}\right\} \text{Εξαρτημένα}$$

Δημιουργώ τις διαφορές

$D_i = X_i - Y_i \quad , \quad i = 1, \ldots, n$

Τότε αυτές είναι τ.δ. από την $N(\mu_X - \mu_Y, \underset{\text{άγνωστη}}{\sigma_D^2})$

Τότε το Δ.Ε. 100(1-α)% ελαχίστου μήκους για το $\mu_D = \mu_X - \mu_Y$ είναι

$$\bar{D} \pm t_{n-1,\, a/2} \frac{S_D}{\sqrt{n}}$$

11

όπου $\bar{D} = \dfrac{\sum_{i=1}^{n} D_i}{n}$

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$$