ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ

Διάλεξη 8η

21/03/2018.

- Στατιστική Συμπερασματολογία (Statistical Inferense)
  - Εκτίμηση Παραμέτρων (Parameter Estimation)
    - Εκτίμηση σε σημείο (Σημειοεκτιμητική) (Point estimation)
    - Εκτίμηση σε διάστημα (Διαστημική Εκτίμηση) (interval Estimation)
  - Έλεγχος Στατιστικών Υποθέσεων (Testing Statistical Hypotheses)

A, B

$n = 100 \begin{cases} \to A \\ \to \text{όχι } A \end{cases}$, $p = P(A)$; || X = αριθμός των A

$X \sim B_{1,n}(n,p)$

$X = x = 70$

$p \approx \frac{70}{100}$

Είναι το $p > \frac{1}{2}$;

▶ Εκτίμηση σε σημείο:

$X_1, \ldots, X_n$ τ.δ. από πληθ(θ)

$\theta =$; || Στατιστική συνάρτηση: $\hat{\theta} = T(X_1, \ldots, X_n)$ λέγεται εκτιμήτρια συνάρτηση και η τιμή της λέγεται εκτιμητής.

$\theta = \mu$; || $\bar{x} \approx \mu$ ||

Αξιολόγηση :

① **Αμεροληψία**: Ένας εκτιμητής του θ, με $\hat{\theta} = T(x_1, \dots, x_n)$ λέγεται αμερόληπτος αν $\forall \theta \in \Theta$, $\boxed{E(\hat{\theta}) = \theta}$

(* Η ιδιότητα της αμεροληψίας είναι μια καλή ιδιότητα, διότι μας δίνει την δυνατότητα προσδιορισμού της "πραγματικής" τιμής του θ)

② **Ελάχιστη διακύμανση** :

Τιμή (μ)

$x_1, \dots, x_n$ και :

Για δείγμα μεγέθους 10:

$$\bar{x} = \frac{1}{n}\sum_1^n x_i = \frac{1}{10}\sum_1^{10} x_i = \bar{x}_{10}$$

$$\hat{\mu} = \bar{x}, \quad E(\hat{\mu}) = E(\bar{x}) = \mu,$$

$$E(\bar{x}_{10}) = \mu.$$

Για δείγμα μεγέθους 20: $\bar{x}_{20} = \frac{1}{20}\sum_1^{20} x_i$,

$$\hat{\mu} = \bar{x}_{20}$$

$$E(\hat{\mu}) = E(\bar{x}_{20}) = \mu.$$

→ Έχω 2 αμερόληπτους : το $\bar{x}_{10}$ και $\bar{x}_{20}$. Ωστόσο ο αμερόληπτος $\bar{x}_{20}$ θα είναι "καλύτερος".

$$\left.\begin{aligned} Var(\bar{x}_{10}) &= \frac{\sigma^2}{10} \\ Var(\bar{x}_{20}) &= \frac{\sigma^2}{20} \end{aligned}\right\} \quad Var(\bar{x}_{20}) < Var(\bar{x}_{10}),$$

καθώς το σ^2 είναι το ίδιο.

(*) Η διακύμανση ενός εκτιμητή είναι **ΠΟΛΥ ΣΗΜΑΝΤΙΚΟ** και είναι μικρή !!

Ισχύει ότι : $Var(\hat{\theta}) = E(\hat{\theta} - \theta)^2$

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

και $E(S^2) = \sigma^2$, $\hat{\sigma}^2 = S^2$.

Παράδειγμα

$$P(X=x) = \frac{1}{3}.$$

Δειγματοληψία με επανατοποθέτηση, $n=2$.

(3,3), (3,4), (3,5), (4,3), (4,4), (4,5), (5,3), (5,4), (5,5)

$\bar{X}=\bar{x}$: 3   3.5   4   3.5   4   4.5   4   4.5   5

$\bar{X}=\bar{x}$: 3 , 3.5 , 4 , 4.5 , 5

$P_{\bar{X}}(i)$: 1/9 , 2/9 , 3/9 , 2/9 , 1/9

$\mu_{\bar{x}} = E(\bar{x}) = 4 \,(=\mu)$

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{\sigma^2}{2} = \frac{2/3}{2} = \frac{1}{3}$

$$\boxed{\hat{\mu} = \bar{x}}$$

Διαστήματα Εμπιστοσύνης:

$X_1, \dots, X_n$ τ.δ. από την $f(\theta)$

$L = L(X_1, \dots, X_n)$, $U = U(X_1, \dots, X_n)$, $[L, U]$.

1. "Ελάχιστο" μήκος
2. Περιέχει το άγνωστο θ, δηλαδή την αληθινή τιμή, ένα μεγάλο "ποσοστό" φορών.

$$P\left(L(X_1, \dots, X_n) \le \theta \le U(X_1, \dots, X_n)\right) = 1-\alpha$$

όπου $\alpha = 0.01,\ 0.05,\ 0.10$ 0.95

Το $(1-\alpha)100\%$: Βαθμός εμπιστοσύνης του διαστήματος.

Εφαρμογή:

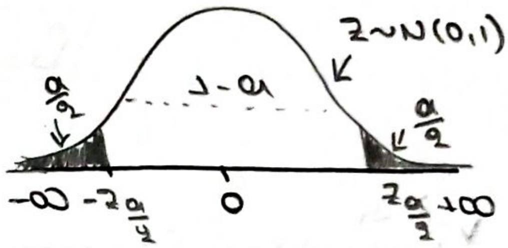
(1-α)100% Δ.Ε. για το μ ενός $N(\mu, \sigma^2$ γνωστό$)$ με βάση $x_1, \dots, x_n$ τ.δ. από τον πληθυσμό αυτό

$$\hat{\mu} = \bar{x}.\ \left(E(\hat{\mu}) = E(\bar{x}) = \mu\right) \text{ και } \bar{x} = \frac{1}{n}\sum_{1}^{n} x_i$$

$$P(L \le \mu \le U) = 1-\alpha$$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightsquigarrow \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

δηλαδή το $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$



$$P(L \le \mu \le U) = 1-\alpha$$

$$\Downarrow$$

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\frac{\alpha}{2}}\right) = 1-\alpha \rightsquigarrow P\left(-\bar{x} - z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightsquigarrow P\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \rightsquigarrow$$

$$\rightsquigarrow \underbrace{\boxed{L = \bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}},\ \boxed{U = \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}}$$

$\uparrow$

(1-α)100% Δ.Ε. για τ.μ. $[L, U]$

$N(\mu, \sigma^2$ άγνωστο$)$

(1-α)100% Δ.Ε. για το μ: $\quad \dfrac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1},\ P\left(-t_{\frac{\alpha}{2}, n-1} \le \dfrac{\bar{x}-\mu}{s/\sqrt{n}} \le t_{\frac{\alpha}{2}, n-1}\right) = 1-\alpha.$

$$[L, U] = \left[\bar{x} - t_{\frac{\alpha}{2}, n-1}\cdot\frac{s}{\sqrt{n}},\ \bar{x} + t_{\frac{\alpha}{2}, n-1}\cdot\frac{s}{\sqrt{n}}\right]$$

Παράδειγμα 4.1:

1, 6, 0, 8, -2, -4, 0, 1, -2, 1.
2, 7, 0, 6, -6, -5, -1, 6, -2, 4 | $n=20$

$N(\mu, \sigma^2 = 25)$, 95% Δ.Ε. για το μ, $\alpha = 0{,}05$.

$$L = \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \bar{X} - z_{0.025} \cdot \frac{5}{\sqrt{20}} \rightsquigarrow l = -1.59$$

$$U = \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \bar{X} + z_{0.025} \frac{5}{\sqrt{20}} \rightsquigarrow u = 2.79$$

$[l, u] = [-1.59, 2.79]$ **και** $P(-1.59 \le \mu \le 2.79) \neq 0.95$

$$= \begin{cases} 1, & \text{αν } \mu \in [\ ] \\ 0, & \text{αν } \mu \notin [\ ] \end{cases}$$