16-03-16

Στατιστική Συμπερασματολογία (Statistical Inference)

Παράδειγμα: A, B υποψήφιοι
p: Πιθανότητα εκλογής του A

X = Αριθμός των ψηφοφόρων του A
$n = 100$ άτομα, {το τυχαίο δείγμα}
$X = x$, $p \approx \frac{x}{n}$ $\left(= \frac{70}{100} \text{ π.χ.}\right)$
$X \sim Bin(n, p)$

→ Εκτίμηση Παραμέτρων
ή
Εκτιμητική
(Parameter Estimation)

- Είναι το $p > \frac{1}{2}$; ⟼ Έλεγχος Στατιστικών Υποθέσεων (Hypothesis Testing)

- Εκτιμητική
  - → Σημειοεκτιμητική ή Εκτίμηση σε σημείο (Point estimation)
  - → Εκτίμηση σε διάστημα ή Διαστήματα Εμπιστοσύνης (Confidence Intervals)

Έστω τυχαίο δείγμα $x_1, \dots, x_n$ από πληθυσμό με μέση τιμή μ.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \mu \approx \bar{X}$$

- Εκτιμήτρια Συνάρτηση ($\bar{X}$, s^2, κλπ)

- $l \leq \mu \leq u$

A

• Συμβολίζουμε με θ ή $g(\theta)$ την άγνωστη παράμετρο και με $\hat{\theta}$ ή $T(x_1, ..., x_n)$ την εκτιμήτρια συνάρτηση

• Σημειοεκτιμητική

① Πώς βρίσκουμε εκτιμήτριες συναρτήσεις;

② Πώς τις αξιολογούμε;

· Ροπές του δείγματος : $\mu = E(x)$, $\bar{x} = m_1$ } $m_1 = E(x)$ ή $\mu = \bar{x}$ (Δηλ. $\hat{\mu} = \bar{x}$)

• Ιδιότητες Αξιολόγησης

① Αμεροληψία: Ένας εκτιμητής $\hat{\theta}$ του θ είναι αμερόληπτος αν $\forall \theta$ ισχύει $E(\hat{\theta}) = \theta$

$x_1, ..., x_{25}, ..., x_{40}$, $\hat{\mu} = \bar{x}_{25} = \frac{1}{25} \sum_{i=1}^{25} x_i$, $E(\bar{x}) = \mu$

$\hat{\mu} = \bar{x}_{40} = \frac{1}{40} \sum_{i=1}^{40} x_i$, $E(\bar{x}) = \mu$

Συνεπώς $\bar{x}$: αμερόληπτος για το μ

$$\sigma^2_{\bar{x}} = Var(\bar{x}) = \frac{\sigma^2}{n}$$

$$\left.\begin{array}{l} \sigma^2_{\bar{x}_{25}} = \frac{\sigma^2}{25} \\ \sigma^2_{\bar{x}_{40}} = \frac{\sigma^2}{40} \end{array}\right\} \sigma^2_{\bar{x}_{25}} > \sigma^2_{\bar{x}_{40}}$$

• Ελάχιστη Διακύμανση

$\sigma^2 = E(x-\mu)^2$ | $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\sigma^2_{\hat{\theta}} = E(\hat{\theta}-\theta)^2$ | $\hat{\sigma}^2 = S^2 \rightarrow E(S^2) = \sigma^2$

Το $E(S'^2) = \frac{n-1}{n}\sigma^2$ δεν είναι αμερόληπτος εκτιμητής

• Διαστήματα Εμπιστοσύνης

Έστω $x_1, \dots, x_n$ τυχαίο δείγμα από πληθυσμό με άγνωστη παράμετρο θ

Έστω το διάστημα $(L, U) = (L(x_1, \dots, x_n), U(x_1, \dots, x_n))$
Το διάστημα (L, U) θα μπορούσε να χρησιμοποιηθεί αν:

① Περιείχε την αληθινή τιμή του θ, ένα μεγάλο "ποσοστό φορών".

② Είχε το (L, U) "ελάχιστο" μήκος

Το ποσοστό φορών που το (L, U) περιέχει το πραγματικό θ λέγεται βαθμός εμπιστοσύνης και δίνεται απ' τη σχέση

$P(L \leq \theta \leq U) = 1-\alpha$, για $0<\alpha<1$

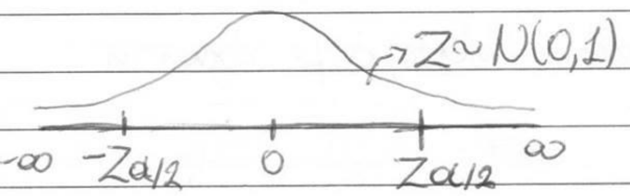
Συνήθως, το $1-\alpha$ θα ισούται με 0,95, 0,99 ή 0,90 (σπάνια)

Παράδειγμα: $100(1-\alpha)\%$ Διάστημα Εμπιστοσύνης (Δ.Ε) για το μ ενός $N(\mu, \sigma^2 = \text{γνωστό})$ πληθυσμού

Έχουμε: $X_1, \dots, X_n$ τυχαίο δείγμα από $N(\mu, \sigma^2 = \text{γνωστό})$

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

$$\hat{\mu} = \bar{X}, \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightsquigarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



$$P\left(-Z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right) = 1-\alpha$$

$$\rightarrow P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \quad , \text{Sn?}$$

$$L(X_1, \dots, X_n) = \bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$U(X_1, \dots, X_n) = \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Το (L, U) θα είναι $(1-\alpha)100\%$ διάστημα εμπιστοσύνης για το μ.

Παράδειγμα 4.1: 7, 6, 0, 8, −9, −4, 0, 1, −9, 1, 2, 7, 0, 6, −6, −5, −1, 6, −2, 4 $\boxed{n = 20}$

$N(\mu, \sigma^2 = 25)$, 95% Διάστημα Εμπιστοσύνης για μ

$$L = \bar{x} - z_{a/2} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} - z_{0,025} \cdot \frac{\sqrt{25}}{\sqrt{20}}$$

$$U = \bar{x} + z_{a/2} \cdot \frac{\sigma}{\sqrt{n}} = \bar{x} + z_{0,025} \cdot \frac{\sqrt{25}}{\sqrt{20}}$$

$\cdot\, z_{0,025} = ?$ $\quad \bar{x} = 0,6$ και $(l, u) = (-1,59,\ 2,79)$

! Δε σημαίνει απαραίτητα ότι $P(-1,59 \le \mu \le 2,79) = 0,95$!