Εισαγωγικά στοιχεία

Φύση

Πληθυσμός ← Τυχαίο Δείγμα

↓

Τ.μ. X

↑

κατανομή π.χ. $X \sim N(\mu, \sigma^2)$

Στατιστικός

Τυχαίο Δείγμα

↓

$X_1, X_2, \dots, X_n$

$\downarrow \quad \downarrow \quad \quad \downarrow$

$x_1, \; x_2, \dots, x_n$

με x_i παριστά την παρατηρηθείσα τιμή της τ.μ. X_i, $i=1,\dots,n$

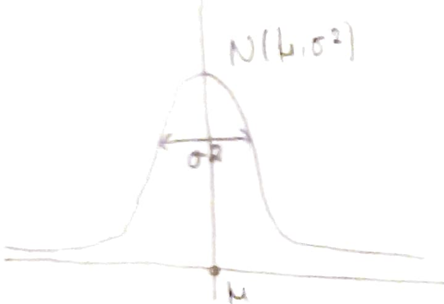
Στατιστική Συμπερασματολογία

Αναπτύσσει μεθοδολογίες για την προσέγγιση των αγνώστων παραμέτρων της κατανομής του πληθυσμού

(I) Εκτίμηση σε σημείο (εκτίμηση = προσέγγιση)

(II) Εκτίμηση σε διάστημα και διαστήματα εμπιστοσύνης

(III) Έλεγχοι στατιστικών υποθέσεων - Στατιστικά τεστ.



Ιστόγραμμα $x_1, \dots, x_n$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Δείχνει τον βαθμό συγκέντρωσης του πολυγώνου γύρω από το $\bar{x}$

Παράμετροι κατανομής ← ΑΓΝΩΣΤΕΣ.

Εκτίμηση σε σημείο

Συμβολισμός - Ορολογία

(I) Ένας πληθυσμός → μια τ.μ. X ή ένα τυχαίο διάνυσμα $X = (X_1, \dots, X_m)$ που περιγράφουν χαρακτηριστικό (-ά) του πληθυσμού. Θεωρούμε μια τ.μ. X

(II) Έστω η τυχαία μεταβλητή X έχει κατανομή με γνωστή συναρτησιακή έκφραση $f(x, \theta)$, αλλά εξαρτάται η f από μια άγνωστη παράμετρο θ. $\theta \in \Theta \subseteq \mathbb{R}^k$. Ο χώρος Θ ονομάζεται παραμετρικός χώρος.

π.χ. (1) Έστω $X \sim Εκθ(\theta)$, τότε $f(x, \theta) = \theta e^{-\theta x}$, $\theta > 0$, $x > 0$, $\theta \in \Theta = (0, \infty) \subseteq \mathbb{R}$

(2) Έστω $X \sim N(\mu, \sigma^2)$ τότε $f(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, $-\infty < x < +\infty$, $-\infty < \mu < +\infty$, $\sigma^2 > 0$

$\theta = (\mu, \sigma^2)$, $\theta \in \Theta = \{(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$

$\Theta \subseteq \mathbb{R} \times \mathbb{R}^+$

(III) Θεωρούμε ένα τυχαίο δείγμα (τ.δ.) μεγέθους n, $X_1, \dots X_n$ από πληθυσμό με κατανομή $f(x,\theta)$ δηλ. n-ανεξάρτητες και ισόνομες τ.μ. $X_1, \dots, X_n$ από την κατανομή $f(x,\theta)$ (δηλ. $X_i \sim f(x,\theta)$). Συμβολίζουμε με τα αντίστοιχα μικρά γράμματα $x_1, \dots, x_n$ τις παρατηρούμενες τιμές των $X_1, X_2, \dots, X_n$ αντίστοιχα.

Το τυχαίο δείγμα $X_1, X_2, \dots X_n$ είναι το μόνο που έχει ο στατιστικός στα χέρια του προκειμένου να μελετήσει τον πληθυσμό $f(x,\theta)$, ή να αναπτύξει συμπερασματολογία (π.χ. εκτίμηση) της άγνωστης παραμέτρου θ.

Στατιστική συνάρτηση (σ.σ.) $T = T(X_1, \dots, X_n)$ είναι κάθε πραγματική συνάρτηση του τ.δ. $X_1, \dots, X_n$ π.χ $T = T(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ή $T = T(X_1, \dots, X_n) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(IV) Εκτιμητής της άγνωστης θ (ή γενικότερα μιας συνάρτησης $g(\theta)$) ονομάζεται κάθε στατιστική συνάρτηση $T = T(X_1, \dots, X_n)$ η οποία προτείνεται για την εκτίμηση (προσέγγιση) της θ ή $g(\theta)$

(V) Εκτίμηση της θ ή $g(\theta)$ λέγεται η τιμή του εκτιμητή $T = (x_1, x_2, \dots x_n)$ στην παρατηρούμενη τιμή $x_1, x_2, \dots, x_n$ του τ.δ. $X_1, X_2, \dots, X_n$.

Ποιότητα εκτιμητών - Κριτήρια Επιλογής

Έστω τ.δ. $X_1, \dots, X_n$ από πληθυσμό με κατανομή $f(x,\theta)$, $\theta \in \Theta \subseteq \mathbb{R}$ και έστω $T = T(X_1, \dots, X_n)$ εκτιμητής της $g(\theta)$, $\theta \in \Theta$.

Ποιότητα του εκτιμητή $T \to |T - g(\theta)|$

Προβλήματα
- → Η τ.μ. δύσχρηστη
- → Η τ.μ. εξαρτάται από το τ.δ. $X_1, X_2, \dots X_n$ (σοβαρή αδυναμία)

Τα προβλήματα ξεπερνιούνται αν θεωρήσουμε την μέση απόσταση $E[T - g(\theta)]^2$

Η ποσότητα εκφράζει την ποιότητα του εκτιμητή. Μέσο Τετραγωνικό Σφάλμα.

ΟΡΙΣΜΟΣ Το μέσο τετραγωνικό σφάλμα του εκτιμητή $T=T(\underline{X})=T(X_1,\dots,X_n)$ για την εκτίμηση της $g(\theta)$ ορίζεται: $MT\Sigma(T,g(\theta))=E[T(X)-g(\theta)]^2$.

ΠΡΟΤΑΣΗ: $MT\Sigma(T,g(\theta))=Var(T)+[E(T)-g(\theta)]^2,\ \forall\theta\in\Theta$.

Απόδ.:

$$MT\Sigma(T,g(\theta))\overset{\text{ορ}}{=}E[T(X)-g(\theta)]^2\ \xlongequal{Var(W)=E(W^2)-(EW)^2}\ Var(T(\underline{X})-g(\theta))^2+[E(T(\underline{X})-g(\theta))]^2$$

$$\xlongequal[E(aW+b)]{Var(a)=0}\ Var(T(\underline{X}))+[E(T(X))-g(\theta)]^2$$

Παράδειγμα: Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό Εκθ$(1/\theta)$ που σ.π.π.:

$f(x,\theta)=\frac{1}{\theta}e^{-x/\theta}$, $x>0,\ \theta>0$. Έστω η σ.σ. $T=T(X_1,\dots,X_n)=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ για την εκτίμηση της θ. Να βρεθεί το Μ.Τ.Σ(T,θ)

Λύση

$$MT\Sigma(\bar{x},\theta)=Var(\bar{x})+[E(\bar{x})-\theta]^2$$

$$E(\bar{x})=E\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(x_i)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\frac{1}{\theta}}=\frac{1}{n}\sum_{i=1}^{n}\theta=\frac{1}{n}n\theta=\theta$$

$$\left.\begin{array}{l}
Var(\bar{x})=Var\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}Var\left(\sum_{i=1}^{n}x_i\right)\\
\text{Ισχύει } Var\left(\sum_{i=1}^{n}a_iw_i\right)=\sum_{i=1}^{n}a_i^2Var(w_i)+\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\,cov(w_i,w_j)\\
\text{Αν } w_1,\dots,w_n \text{ ανεξάρτητες } cov=0\\
Var\left(\sum_{i=1}^{n}a_iw_i\right)=\sum_{i=1}^{n}a_i^2Var(w_i)
\end{array}\right\}\Rightarrow$$

$$\Rightarrow Var(\bar{x})=\frac{1}{n^2}\sum_{i=1}^{n}Var(x_i)=\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\left(\frac{1}{\theta^2}\right)}=\frac{\theta^2}{n}$$

Επομένως $MT\Sigma(\bar{x},\theta)=\frac{\theta^2}{n}+(\theta-\theta)^2=\frac{\theta^2}{n}$

Παρατήρηση: Μ.Τ.Σ $(\hat{x}, \theta) \xrightarrow[n\to\infty]{} 0$

Αν ΜΤΣ ενός εκτ. < ΜΤΣ ενός άλλου ο $1^{ος}$ είναι καλύτερος

ΟΡΙΣΜΟΣ: Αποδεκτός εκτιμητής.

Για δύο εκτιμητές $T_1 = T_1(x)$, $T_2 = T_2(x)$ της θ, ο εκτιμητής T_1 ορίζεται να είναι καλύτερος από τον T_2 με κριτήριο το ΜΤΣ, αν και μόνο αν:

(1) ΜΤΣ $(T_1, \theta) \leq$ ΜΤΣ (T_2, θ) $\forall\, \theta \in \Theta$.

(2) ΜΤΣ $(T_1, \theta) <$ ΜΤΣ (T_2, θ) για τουλάχιστον ένα $\theta \in \Theta$

Ο εκτιμητής T_2 λέγεται μη-αποδεκτός εκτιμητής.