Φύση

Πληθυσμός

↓

χαρακτηριστικό γνώρισμα
ή ιδιότητα των μελών του πληθυσμού

↓ είναι

(τυχαίες) Μεταβλητή X

↓

∃ μία κατανομή f που περιγράφει την X
Πχ.: $N(\mu,\sigma^2)$

↓ ΠΡΟΒΛΗΜΑ

μ,σ^2 είναι άγνωστες!

Ορισμός

Στατιστική Συμπερασματολογία είναι ο κλάδος της στατιστικής που αναπτύσσει μεθοδολογίες για την προσέγγιση των αγνώστων παραμέτρων της κατανομής που περιγράφει το χαρακτηριστικό γνώρισμα (τ.μ.) του πληθυσμού

- Ⅰ Εκτίμηση σε σημείο
- Ⅱ Εκτίμηση σε διάστημα
- Ⅲ Έλεγχοι Στατιστικών Υποθέσεων - Στατιστικά Τεστ

Καθ' αντιστοιχία:

Στατιστικός

Τυχαίο δείγμα: Ένα αντιπροσωπευτικό υποσύνολο του πληθυσμού

τ.δ.: $X_1,\dots,X_n$ ← Παρατηρήσεις στην X στα n-μέλη του δείγματος
$\bar{X}=\frac{1}{n}(X_1+\dots+X_n)$ και $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})$

Ⅰ Εκτίμηση σε σημείο

- Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό που περιγράφεται από μία κατανομή $f(x,\theta)$ πχ. $N(\mu,\sigma^2)$, $Εκθ(\theta)$, ...
Ειδικότερα, όταν λέμε τ.δ. εννοούμε:
1) n ανεξάρτητες τ.μ. $X_1,\dots,X_n$
2) Οι τ.μ. $X_1,\dots,X_n$ είναι ισόνομες ($X_1\sim f(x,\theta)$, ..., $X_n\sim f(x,\theta)$)
- Η άγνωστη παράμετρος θ ανήκει στον παραμετρικό χώρο $Ⓗ$
$Ⓗ\subseteq\mathbb{R}$ $(\subseteq\mathbb{R}^m)$ δηλ. $\theta\in Ⓗ$

Πχ.1: Έστω τ.δ. από πληθυσμό με κατανομή $f(x,\theta)=\theta e^{-\theta x}$, $x>0$, $\theta>0$
Εκθετική άρα $Ⓗ=(0,+\infty)$

Πχ.2: Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό $N(\mu,\sigma^2)$ μ,σ^2 άγνωστα, $-\infty<\mu<+\infty$, $\sigma^2>0$

• **Στατιστική Συνάρτηση**: κάθε συνάρτηση τ.δ. $X_1, \dots, X_n$
$T = T(X) = T(X_1, \dots, X_n)$ πχ. $\bar{X}, S^2$ κτλ.

• **Εκτιμητής**: Κάθε στατιστική συνάρτηση που χρησιμοποιείται για την εκτίμηση (προσέγγιση) της άγνωστης παραμέτρου θ ή $g(\theta)$

Ποιότητα - Κριτήρια Επιλογής Εκτιμητών - Μέσο τετραγωνικό σφάλμα

Έστω $T = T(X_1, \dots, X_n)$ ένας εκτιμητής της $g(\theta)$, $\theta \in \Theta$ ~~αδυναμίες~~

Κριτήριο ποιότητας του T είναι η $|T - g(\theta)|$ < δυσκολία απόλ. τιμής / εξαρτάται από την τ.δ.

Για να ξεφύγω απ' τις 2 αδυναμίες θεωρώ:

Κριτήριο Επιλογής Εκτιμητών: $E\left(T(X_1, \dots, X_n) - g(\theta)\right)^2$

<u>Ορισμός</u>

Μέσο τετραγωνικό σφάλμα (ΜΤΣ) του εκτιμητή $T(X) = T(X_1, \dots, X_n) = T$ για την εκτίμηση της $g(\theta)$, $\theta \in \Theta$ ονομάζεται

$$MT\Sigma(T, g(\theta)) = E(T - g(\theta))^2$$

<u>Πρόταση</u>

$$MT\Sigma(T, g(\theta)) = Var(T) + (E(T) - g(\theta))^2$$

<u>Απόδειξη</u>

Ισχύει $Var(W) = E(W^2) - (E(W))^2 \Rightarrow E(W^2) = Var(W) + (E(W))^2$

Εφαρμόζοντας αυτή τη σχέση: $MT\Sigma(T, g(\theta)) \overset{\text{ορ}}{=} E(T - g(\theta))^2$

$$= Var(T - g(\theta)) + (E(T - g(\theta)))^2 \xlongequal[E(\alpha W + \beta) = \alpha E(W) + \beta]{Var(\alpha) = 0} Var(T) + (E(T) - g(\theta))^2$$

<u>Πχ</u>

Έστω τ.δ. $X_1, \dots, X_n$ από πληθ. με καταν. $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x > 0$, $\theta > 0$

Έστω σ.σ. $T = T(X_1, \dots, X_n) = \bar{X}$ για την εκτίμηση του θ

Να βρεθεί το $MT\Sigma(T, \theta)$.

<u>Λύση</u>

Από πρόταση: $MT\Sigma(\bar{X}, \theta) = Var(\bar{X}) + (E(\bar{X}) - \theta)^2$

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xlongequal{X_i \sim Exp(\frac{1}{\theta})} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1/\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{1}{n} n\theta = \theta$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} Var\left(\sum X_i\right) \qquad \Rightarrow E(\bar{X}) = \theta$$

Αν οι $W_1, \dots, W_n$ είναι ανεξάρτητες τότε: $\boxed{Var\left(\sum \alpha_i W_i\right) = \sum \alpha_i^2 Var(W_i)}$

$$Var(\bar{X}) = \frac{1}{n^2}\sum Var(X_i) = \frac{1}{n^2}\sum \frac{1}{(1/\theta)^2} \Longrightarrow Var(\bar{X}) = \frac{\theta^2}{n}$$

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

Παρατήρηση: Αν $n \to +\infty$ τότε $MT\Sigma(\bar{X}, \theta) = 0$

Ορισμός

Έστω οι εκτιμητές T_1 και T_2 για την εκτίμηση της $g(\theta)$

Ο T_1 λέγεται καλύτερος από τον T_2 με κριτήριο το ΜΤΣ

αν $MT\Sigma(T_1, g(\theta)) \leq MT\Sigma(T_2, g(\theta)) \quad \forall \theta \in \Theta$

Παρατηρήσεις

1) Ο T_2 λέγεται μη αποδεκτός

2) Ένας εκτιμητής λέγεται αποδεκτός αν είναι καλύτερος από οποιονδήποτε άλλο εκτιμητή

Πρόταση

Αποδεικνύεται ότι το ΜΤΣ αναδεικνύει τον καλύτερο εκτιμητή μόνο για την περίπτωση που εκτιμάει σταθερή ποσότητα (δηλ. $g(\theta)=c$, σταθερά)

Αυτό όμως (εκτίμηση σταθ. ποσότητας) δεν έχει πρακτικό ενδιαφέρον

→ Για να άρουμε αυτή την αδυναμία του ΜΤΣ εισάγουμε την έννοια της αμεροληψίας

Ορισμός

Ο εκτιμητής $T=T(X_1,\dots,X_n)$ θα λέγεται αμερόληπτος εκτιμητής της $g(\theta)$ αν $E(T)=g(\theta)$

Ορισμός

Η μεροληψία του εκτιμητή T για την εκτίμηση της $g(\theta)$ ορίζεται: $b(T,g(\theta)) = E(T) - g(\theta)$

Πρόταση

α) $MT\Sigma(T,g(\theta)) = Var(T) + b^2(T,\theta)$

β) Αν ο T αμερόληπτος της $g(\theta)$ τότε $b(T,g(\theta))=0$

γ) Αν ο T αμερόληπτος της $g(\theta)$ τότε $MT\Sigma(T,g(\theta)) = Var(T)$

Η απόδειξη είναι προφανής

Ορισμός

Ο εκτιμητής T λέγεται Αμερόληπτος Ομοιόμορφα Ελάχιστης Διακύμανσης (ΑΟΕΔ) της $g(\theta)$ **αν** ο T είναι αμερόληπτος της $g(\theta)$ και η διακύμανση του είναι η ελάχιστη μεταξύ των διακυμάνσεων όλων των άλλων αμερόληπτων εκτιμητών $g(\theta)$

Παρ1

Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό με μέση τιμή μ και διακύμανση σ^2. Ν.δ.ο. ο $\bar{X}$ αμερόληπτος της μ και $Var(\bar{x})=\frac{\sigma^2}{n}$

Λύση

Αφού $X_1,\dots,X_n$ από πληθ. με μ,σ^2 έχω: $E(X_i)=\mu$ και $Var(X_i)=\sigma^2$ $\forall i=1,\dots,n$

$$E(\bar{x})=E\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n}\sum E(X_i)=\frac{1}{n}\sum\mu=\frac{1}{n}n\mu=\mu$$

$$Var(\bar{x})=Var\left(\frac{1}{n}\sum X_i\right)=\frac{1}{n^2}\sum Var(X_i)=\frac{1}{n^2}\sum\sigma^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

Παρ2

Έστω η τ.δ. $X_1,\dots,X_n$ από πληθ. με σ^2 και μ.

Ν.δ.ο. $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{x})^2$ αμερόληπτος εκτιμητής του σ^2

Λύση

$$E(S^2)=E\left(\frac{1}{n-1}\sum(X_i-\bar{x})^2\right)=E\left[\frac{1}{n-1}\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)\right]$$

$$=\frac{1}{n-1}E\left[\sum X_i^2-2\bar{X}\sum X_i+n\bar{X}^2\right]=$$

$$=\frac{1}{n-1}E\left(\sum X_i^2-2\bar{X}n\bar{X}+n\bar{X}^2\right)=\frac{1}{n-1}E\left[\sum X_i^2-n\bar{X}^2\right]$$

$$=\frac{1}{n-1}\left(\sum E(X_i^2)-nE(\bar{X}^2)\right) \qquad \underline{\underline{Var(W)=E(W^2)-(E(W))^2}}$$

$$=\frac{1}{n-1}\left(\sum\left(Var(X_i)+E(X_i)\right)-n\left(Var(\bar{X})+(E(\bar{X}))^2\right)\right)$$

$$=\frac{1}{n-1}\left(\sum(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)=$$

$$=\frac{1}{n-1}\left(n\sigma^2+n\mu^2-\sigma^2-n\mu^2\right)=\frac{1}{n-1}(n-1)\sigma^2=\sigma^2$$