Bài 2. Ước lượng điểm

KHÁI NIỆM CHUNG VỀ ƯỚC LƯỢNG

- Ước lượng là phỏng đoán một giá trị chưa biết của tổng thể dựa vào quan sát trên mẫu lấy ra từ tổng thể đó. Thông thường, ta cần ước lượng về trung bình, tỉ lệ, phương sai, hệ số tương quan của tổng thể.

- Có hai hình thức ước lượng:
  - ***Ước lượng điểm***: kết quả cần ước lượng được cho bởi một trị số.
  - ***Ước lượng khoảng***: kết quả cần ước lượng được cho bởi một khoảng.

- Ước lượng điểm có ưu điểm là cho ta một giá trị cụ thể, có thể dùng để tính các kết quả khác, nhưng nhược điểm là không cho biết sai số của ước lượng.
  Ước lượng khoảng thì ngược lại.

§2. ƯỚC LƯỢNG ĐIỂM
(tham khảo)

Cho mẫu độc lập $X_1,\ldots,X_n$ có hàm mật độ phụ thuộc vào tham số θ cần ước lượng (θ có thể là trung bình, phương sai, tỉ lệ,…). Gọi $T = T(X_1,\ldots,X_n)$ là thống kê chỉ phụ thuộc vào $X_1,\ldots,X_n$, không phụ thuộc vào θ.

2.1. Ước lượng đúng

- Ta nói $T = T(X_1,\ldots,X_n)$ là ước lượng đúng (hay ước lượng không chệch) của θ nếu $ET = \theta$.
- Khi $ET \neq \theta$, ta nói T là ước lượng không đúng của θ:
  - $ET < \theta$, ta nói ước lượng thiếu;
  - $ET > \theta$, ta nói ước lượng thừa.

2.2. So sánh các ước lượng

a) Ước lượng ít phân tán

- Gọi T_1, T_2 là hai ước lượng đúng của θ.

  Ta nói T_1 ***ít phân tán hơn*** T_2 nếu $Var(T_1) \leq Var(T_2)$.

- Khi T_1 ít phân tán hơn T_2, ta nói T_1 ***tốt hơn*** T_2.

  Nghĩa là, khi dùng T_1 để ước lượng θ ta nhận được sai số ước lượng ít hơn so với dùng T_2.

b) Ước lượng tốt nhất

- **Định nghĩa**

  Thống kê T được gọi là ước lượng tốt nhất của θ nếu T là ước lượng đúng và ít phân tán nhất.

VD. Giả sử chiều cao X của người Việt Nam có phân phối chuẩn $N(\mu;\, \sigma^2)$. Quan sát mẫu $X_1, \ldots, X_n$ để ước lượng chiều cao trung bình μ. Xét các thống kê sau:

$$T_1 = X_1,\ T_2 = \frac{X_1 + 2X_2}{3} \text{ và } T_3 = \frac{X_1 + \ldots + X_n}{n}.$$

- Đánh giá ước lượng đúng:

$$E(T_1) = E(X_1) = \mu,$$

$$E(T_2) = E\left(\frac{X_1 + 2X_2}{3}\right) = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu,$$

$$E(T_3) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\mu + \ldots + \mu}{n} = \mu.$$

- Đánh giá độ phân tán:

$$Var(T_1) = Var(X_1) = \sigma^2,$$

$$Var(T_2) = Var\left(\frac{X_1 + 2X_2}{3}\right) = \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2 = \frac{5}{9}\sigma^2,$$

$$Var(T_3) = Var\left(\frac{X_1 + \ldots + X_2}{n}\right) = \frac{\sigma^2 + \ldots + \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Vậy khi n lớn thì T_3 là ước lượng tốt nhất.

• Bất đẳng thức Rao – Cramer

Giả sử biến ngẫu nhiên X có hàm mật độ $f(x,\theta)$ phụ thuộc vào tham số θ.

Gọi tin lượng Fisher của X là:

$$I(\theta) = E\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2.$$

Nếu thống kê T là ước lượng đúng của θ thì:

$$VarT \geq \frac{1}{n.I(\theta)}.$$

Vậy T thỏa $VarT = \dfrac{1}{n.I(\theta)}$ là ước lượng tốt nhất.

……………………………………………………………………