……………………………………………………………

§1. KHÁI NIỆM VỀ KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

1.1. Khái niệm chung

- Mô hình tổng quát của bài toán kiểm định là: ta nêu lên hai mệnh đề trái ngược nhau, một mệnh đề được gọi là *giả thuyết* H và mệnh đề còn lại được gọi là *nghịch thuyết* (hay đối thuyết) $\overline{H}$.
- Giải quyết một bài toán kiểm định là: bằng cách dựa vào quan sát mẫu, ta nêu lên một quy tắc hành động, ta chấp nhận giả thuyết H hay bác bỏ giả thuyết H.

- Khi ta chấp nhận giả thuyết H, nghĩa là ta tin rằng H đúng; khi bác bỏ H, nghĩa là ta tin rằng H sai. Do chỉ dựa trên một mẫu quan sát ngẫu nhiên, nên ta không thể khẳng định chắc chắn điều gì cho tổng thể.
- Trong chương này, ta chỉ xét loại kiểm định tham số (so sánh đặc trưng với 1 số, so sánh hai đặc trưng của hai tổng thể).

1.2. Các loại sai lầm trong kiểm định

Khi thực hiện kiểm định giả thuyết, ta dựa vào quan sát ngẫu nhiên một số trường hợp rồi suy rộng ra cho tổng thể. Sự suy rộng này có khi đúng, có khi sai.

Thống kê học phân biệt 2 loại sai lầm sau:

a) Sai lầm loại I

- Sai lầm loại 1 là loại sai lầm mà ta phạm phải trong việc bác bỏ giả thuyết H khi H đúng.
- Xác suất của việc bác bỏ H khi H đúng là xác suất của sai lầm loại 1 và được ký hiệu là α.

b) Sai lầm loại II

- Sai lầm loại 2 là loại sai lầm mà ta phạm phải trong việc chấp nhận giả thuyết H khi H sai.
- Xác suất của việc chấp nhận giả thuyết H khi H sai là xác suất của sai lầm loại 2 và được ký hiệu là β.

c) Mối liên hệ giữa hai loại sai lầm

- Khi thực hiện kiểm định, ta luôn muốn xác suất phạm phải sai lầm càng ít càng tốt. Tuy nhiên, nếu hạ thấp α thì β sẽ tăng lên và ngược lại.

Trong thực tế, giữa hai loại sai lầm này, loại nào tác hại hơn thì ta nên tránh.

- Trong thống kê, người ta quy ước rằng sai lầm loại 1 tác hại hơn loại 2 nên cần tránh hơn. Do đó, ta chỉ xét các phép kiểm định có α không vượt quá một giá trị ấn định trước, thông thường là 1%; 3%; 5%;…
  Giá trị α còn được gọi là ***mức ý nghĩa*** của kiểm định.

1.3. Cơ sở lý thuyết của kiểm định

- Để giải quyết bài toán kiểm định, ta quan sát mẫu ngẫu nhiên $X_1,\ldots,X_n$ và đưa ra giả thuyết H.
- Từ mẫu trên, ta chọn thống kê $T = f(X_1,\ldots,X_n;\, \theta_0)$ sao cho nếu khi H đúng thì phân phối xác suất của T hoàn toàn xác định.
- Với mức ý nghĩa α, ta tìm được khoảng tin cậy (hay khoảng ước lượng) $[a;\, b]$ cho T ở độ tin cậy $1-\alpha$.

  Khi đó:
  - nếu $t \in [a;\, b]$ thì ta chấp nhận giả thuyết H;
  - nếu $t \notin [a;\, b]$ thì ta bác bỏ giả thuyết H.

• Nếu hàm mật độ của T đối xứng qua trục Oy thì ta chọn khoảng đối xứng $[-t_\alpha;\, t_\alpha]$, với:

$$P(T \le -t_\alpha) = P(T \ge t_\alpha) = \frac{\alpha}{2}.$$

Vậy, khi xét nửa bên phải của trục Oy thì ta được:

- nếu $t \le t_\alpha$ thì ta chấp nhận giả thuyết H;
- nếu $t > t_\alpha$ thì ta bác bỏ giả thuyết H.

• Nếu hàm mật độ của T không đối xứng qua trục Oy thì ta chọn khoảng tin cậy $[0;\, C]$, với $P(T \ge C) = \alpha$.

- Nếu $t \le C$ thì ta chấp nhận giả thuyết H, và
- nếu $t > C$ thì ta bác bỏ giả thuyết H.

………………………………………………………………

§2. KIỂM ĐỊNH SO SÁNH ĐẶC TRƯNG CỦA TỔNG THỂ VỚI MỘT SỐ

2.1. Kiểm định so sánh trung bình với một số

Với số μ_0 cho trước, ta đặt giả thuyết $H : \mu = \mu_0$.

a) *Trường hợp* 1. Với $n \geq 30,\ \sigma^2$ đã biết.

• Từ mức ý nghĩa $\alpha \Rightarrow \dfrac{1-\alpha}{2} = \varphi(t_\alpha) \xrightarrow{B} t_\alpha$.

• Tính giá trị thống kê $t = \dfrac{|\overline{x} - \mu_0|}{\sigma}\sqrt{n}$.

• Nếu $t \leq t_\alpha$ thì ta chấp nhận H, nghĩa là $\mu = \mu_0$;
nếu $t > t_\alpha$ thì ta bác bỏ H, nghĩa là $\mu \neq \mu_0$.

b) *Trường hợp* 2. Với $n \geq 30,\ \sigma^2$ chưa biết.

Ta làm như trường hợp 1 nhưng thay σ bằng s.

c) *Trường hợp* 3. Với $n < 30,\ \sigma^2$ đã biết và

X ***có phân phối chuẩn***, ta làm như trường hợp 1.

d) *Trường hợp* 4. Với $n < 30,\ \sigma^2$ chưa biết và

X ***có phân phối chuẩn***.

• Từ cỡ mẫu n và mức ý nghĩa $\alpha \xrightarrow{\text{tra bảng } C} t_\alpha^{n-1}$.

• Tính giá trị thống kê $t = \dfrac{\left|\overline{x} - \mu_0\right|}{s}\sqrt{n}$.

• Nếu $t \leq t_\alpha^{n-1}$ thì ta chấp nhận giả thuyết H;

$t > t_\alpha^{n-1}$ thì ta bác bỏ giả thuyết H.

Chú ý

Trong tất cả các trường hợp bác bỏ, ta so sánh $\overline{x}$ và μ_0:

- Nếu $\overline{x} > \mu_0$ thì ta kết luận $\mu > \mu_0$.
- Nếu $\overline{x} < \mu_0$ thì ta kết luận $\mu < \mu_0$.

VD 1. Sở Điện lực A báo cáo rằng: trung bình một hộ hàng tháng phải trả 250 ngàn đồng tiền điện, với độ lệch chuẩn là 20 ngàn. Người ta khảo sát ngẫu nhiên 500 hộ thì tính được trung bình hàng tháng một hộ trả 252 ngàn đồng tiền điện.

Trong kiểm định giả thuyết H: "**trung bình một hộ phải trả hàng tháng là 250 ngàn đồng tiền điện**" với mức ý nghĩa $\alpha = 1\%$, hãy cho biết giá trị thống kê t và kết luận ?