KIỂM ĐỊNH GIẢ THUYẾT

GV : Đinh Công Khải – Chương trình Fulbright
Môn: Các Phương Pháp Định Lượng – MPP3

Quy trình kiểm định giả thuyết thống kê

- Xây dựng các *giả thuyết không* và *giả thuyết thay thế*
- Kiểm định *một phía* hay *hai phía* của giả thuyết
- Xác định *trị thống kê kiểm định*
- Xác định *miền chấp nhận* và *miền bác bỏ* giả thuyết H_0
- Kết luận bác bỏ hay không bác bỏ giả thuyết H_0

Xây dựng các giả thuyết

- Giả thuyết không (H_0)
  - Là một phát biểu về tham số của tổng thể
  - Được cho là đúng cho đến khi nó được chứng minh là sai
  - Thường là một tuyên bố bị nghi ngờ
- Giả thuyết thay thế (H_a)
  - Là phát biểu ngược với H_0
  - Được cho là đúng nếu bị H_0 bác bỏ
  - Nhà nghiên cứu mong muốn ủng hộ và chứng minh là đúng

Xây dựng các giả thuyết

- **Ví dụ 1**: Một công thức sữa của hãng Abbott dành cho em bé dưới một tuổi được giới thiệu vào năm 2009 được cho rằng tạo ra tăng trọng trung bình cao hơn mức 100gram/tháng của công thức sữa được giới thiệu vào 2007.
- **Ví dụ 2**: Một quy trình sản xuất bóng đèn đang tạo ra một tỷ lệ bình quân 1% sản phẩm bị lỗi. Nhóm các kỹ sư đang đưa một quy trình mới vào thử nghiệm với hy vọng làm giảm tỷ lệ phế phẩm.
- **Ví dụ 3**: Liệu tiền lương trung bình của công nhân cơ khí tại Bình Dương có khác với mức tiền lương trung bình là 2,5 triệu đồng của công nhân cơ khí trên toàn quốc không?

Xây dựng các giả thuyết

- Tổng kết các dạng của giả thuyết không và giả thuyết thay thế
  - $H_0 : \theta = \theta_0$ hoặc $H_0 : \theta \geq \theta_0$ hoặc $H_0 : \theta \leq \theta_0$
  - $H_a: \theta \neq \theta_0$ $H_a : \theta < \theta_0$ $H_a : \theta > \theta_0$
- Kiểm định giả thuyết sẽ nhằm bác bỏ H_0 hoặc không bác bỏ H_0.

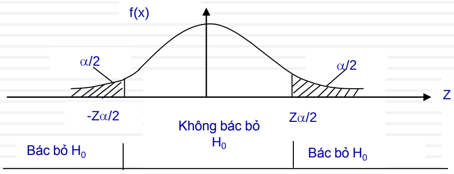
Chú ý: không nên kết luận là *chấp nhận H_0*.

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

- Tổng kết các dạng của giả thuyết không và giả thuyết thay thế về μ.
  - $H_0 : \mu = \mu_0$ hoặc $H_0 : \mu \geq \mu_0$ hoặc $H_0 : \mu \leq \mu_0$
  - $H_a : \mu \neq \mu_0$ $H_a : \mu < \mu_0$ $H_a : \mu > \mu_0$

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

- Kiểm định *2-phía* về trung bình của tổng thể
  - $H_0 : \mu = \mu_0$
  - $H_a : \mu \neq \mu_0$
- Trị kiểm định $$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$
- Dùng α để xác định giá trị tới hạn và quy tắc bác bỏ H_0



Nguồn: Cao Hào Thi (QM-MPP2)

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

- **Phương pháp giá trị tới hạn**
- Bác bỏ H_0 nếu $z < -z_{\alpha/2}$ hoặc $z > z_{\alpha/2}$
- Nếu điều kiện trên không thỏa thì không bác bỏ H_0.
- **Ví dụ 4**: Một nhà máy sản xuất thép ghi nhận sản lượng trong 100 ngày, có trung bình và độ lệch chuẩn của mẫu lần lượt là 880 tấn và 50 tấn. Hãy kiểm định giả thuyết rằng sản lượng bình quân hàng ngày của nhà máy hiện nay khác với mức sản lượng trung bình 892 tấn/ngày đã được ghi nhận các đây 1 năm. Cho biết α = 0,05.

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

- **Phương pháp p_{value}**

  - p_{value} là giá trị nhỏ nhất của α (*được tính từ trị thống kê*) mà qua đó kết quả kiểm định là có ý nghĩa thống kê.

  - Cách thức sử dụng p_{value} để kiểm định giả thuyết

Bác bỏ H_0 nếu $p_{value} < \alpha$

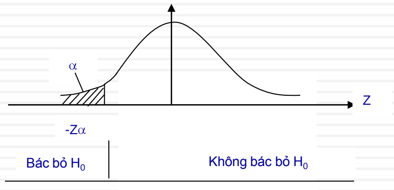
Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

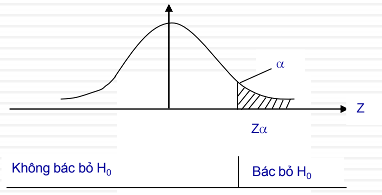
- Kiểm định *1-phía* về trung bình của tổng thể

  - $H_0: \mu \geq \mu_0$ hoặc $H_0: \mu \leq \mu_0$
  - $H_a: \mu < \mu_0$ $H_a: \mu > \mu_0$

- Trị kiểm định $$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

- Dùng α để xác định giá trị tới hạn và quy tắc bác bỏ H_0





Nguồn: Cao Hào Thi (QM-MPP2)

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

Trường hợp 1

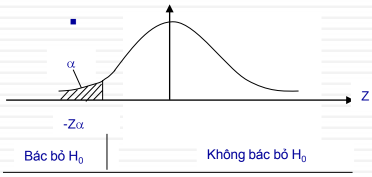
- $H_0: \mu \geq \mu_0$
- $H_a: \mu < \mu_0$

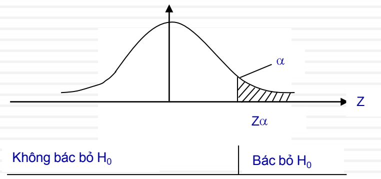
hoặc

Trường hợp 2

$H_0: \mu \leq \mu_0$

$H_a: \mu > \mu_0$





- Phương pháp trị giới hạn: Bác bỏ H_0 nếu

  $z < -z_\alpha$ $\qquad$ $z > z_\alpha$

- Phương pháp p_{value} $\quad p < \alpha$

Kiểm định giả thuyết về trung bình của tổng thể (mẫu lớn)

- **Ví dụ 5**: Một mẫu ngẫu nhiên gồm n=35 quan sát từ một tổng thể tạo ra một số trung bình bằng 2,4 và độ lệch chuẩn bằng với 0,29. Giả định bạn mong muốn chứng minh rằng số trung bình tổng thể μ vượt quá 2.3. Kiểm định giả thuyết với $\alpha = 0.05$?

Kiểm định giả thuyết về trung bình của tổng thể (mẫu nhỏ)

- Khi *cỡ mẫu là nhỏ* hoặc *độ lệch chuẩn của tổng thể là chưa biết* thì việc kiểm định giả thuyết về trung bình tổng thể dựa trên trị thống kê kiểm định.

$$t = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

- Cách thức kiểm định
- Kiểm định 2 phía: Bác bỏ H_0 nếu $t < -t_{\alpha/2}$ hoặc $t > t_{\alpha/2}$
- Kiểm định 1phía: Bác bỏ H_0 nếu $t < -t_{\alpha}$ (TH1); hoặc $t > t_{\alpha}$ (TH2)
- Ví dụ: Kiểm định giả thuyết trong ví dụ 5 với n= 20 và $\alpha = 0.05$?

Các sai lầm khi kiểm định thống kê

- Có 2 loại sai lầm:
- *Sai lầm loại I* là sai lầm của việc bác bỏ H_0 khi nó đúng
- *Sai lầm loại II* là sai lầm của việc không bác bỏ H_0 khi nó sai.
- Các kết luận đúng và sai trong kiểm định giả thuyết

		Giả thuyết H_0	
		H_0 đúng	**H_0 sai**
Kết luận	**Bác bỏ H_0**	Sai lầm loại I	Kết luận đúng
	Không bác bỏ H_0	Kết luận đúng	Sai lầm loại II

Các sai lầm khi kiểm định thống kê

- **Sai lầm loại I (α)**
- α là xác suất của việc bác bỏ H_0 khi nó đúng. Do đó, α là xác suất bác bỏ sai H_0.
- α được coi là mức ý nghĩa của kiểm định
- **Sai lầm loại II (β)**
- β là xác suất của việc không bác bỏ H_0 khi nó sai.

  β = P(Không bác bỏ H_0 | H_0 sai)

  1- β = P(Bác bỏ H_0 | H_0 sai) = Năng lực của kiểm định
- *α càng nhỏ thì β càng lớn*

Kiểm định giả thuyết về sự khác biệt giữa 2 trung bình của tổng thể

- Giả thuyết không H_0: $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = D_0$
- Giả thuyết thay thế H_a: $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq D_0$
- Trị thống kê kiểm định

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sigma_{(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

- Nếu cỡ mẫu nhỏ

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sigma_{(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Kiểm định giả thuyết về sự khác biệt giữa 2 trung bình của tổng thể

- Ví dụ 6: Một công ty đang sử dụng một phương án trả lương mới là doanh số bán hàng cộng hoa hồng cho đội ngũ bán hàng của mình. Công ty muốn so sánh các kỳ vọng lương hàng năm của các nhân viên bán hàng nam và nữ của mình theo kế hoạch mới này. Các mẫu ngẫu nhiên gồm n_1= 40 đại diện bán hàng nữ và n_2= 40 đại diện bán hàng nam được yêu cầu dự báo về thu nhập hàng năm của mình theo kế hoạch mới này. Các số trung bình và độ lệch chuẩn của mẫu lần lượt là: đối với nhóm nữ $31.083, $2322; đối với nhóm nam $29.745, $2.569. Liệu dữ liệu này có cung cấp đủ bằng chứng cho thấy rằng có sự khác biệt về thu nhập hàng năm được kỳ vọng giữa nhân viên nam và nữ? (α=0,05).

Kiểm định giả thuyết về sự khác biệt cặp giữa 2 trung bình của tổng thể

- Một nghiên cứu nhằm kiểm định sự khác biệt về số Km mà các nhân viên y tế đi hiện trường giữa kế hoạch làm việc 6 ngày/tuần và 5 ngày/tuần. Số liệu được thu thập cho 6 nhân viên trong 1 năm như sau:

Tên nhân viên	6 ngày/tuần	5 ngày/tuần	Chênh lệch
A	8089	6392	1697
B	7724	6112	1612
C	7505	6177	1328
D	4592	3281	1311
E	8107	4997	3110
F	3807	3362	445
Trung bình	6637	5053	1583
Độ lệch chuẩn	1751	1302	869

Kiểm định giả thuyết về sự khác biệt cặp giữa 2 trung bình của tổng thể

- Kiểm định khác biệt cặp cho (μ_1 - μ_2 = μ_d)
- Giả thuyết không H_0: $\mu_d = 0$
- Giả thuyết thay thế H_a: $\mu_d \neq 0$ (hoặc $\mu_d > 0$ hoặc $\mu_d < 0$)
- Trị thống kê kiểm định

$$t = \frac{\bar{d} - 0}{s_d \sqrt{n}}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n - 1}}$$

Kiểm định giả thuyết về tỷ lệ của tổng thể

- Gọi p : tỷ lệ của tổng thể

  p_0 : giá trị cụ thể của giả thuyết đối với tỉ lệ của tổng thể
- **Giả thuyết**
  - $H_0 : p = p$ $\quad H_o : p \geq p_0$ $\quad H : p \leq p_0$
  - $H_a : p \neq p_0$ $\quad H_a : p < p_0$ $\quad H : p > p_0$
- **Trị thống kê kiểm định**

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Kiểm định giả thuyết về sự khác biệt giữa 2 tỷ lệ nhị thức của tổng thể

- Giả thuyết không H_0: $(p_1 – p_2) = D_0$.
- Giả thuyết thay thế
- $H_a : (p_1 – p_2) \neq D_0$ hoặc $H_a : (p_1 – p_2) > D_0$ hoặc $H_a : (p_1 – p_2) < D_0$.

- **Trị thống kê kiểm định**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}} = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Kiểm định giả thuyết về sự khác biệt giữa 2 tỷ lệ nhị thức của tổng thể

- **Ví dụ 7**: Một người quản lý bệnh viện nghi ngờ rằng trễ hạn trong việc thanh toán các hóa đơn viện phí đã gia tăng trong năm qua. Hồ sơ lưu trữ của bệnh viện cho thấy rằng các hóa đơn của 48 trong số 1284 người nhập viện trong tháng 4 đã trễ hạn trong hơn 90 ngày. Con số này so với 34 trong 1002 người nhập viện trong cùng tháng này năm trước đó. Liệu những dữ liệu này có cung cấp đủ bằng chứng để cho thấy có sự gia tăng trong tỷ lệ trễ hạn thanh toán vượt quá 90 ngày không? Hãy kiểm định giả thuyết với α= 0.1?