

# Kiểm định giả thuyết

PGS.TS. Lê Sỹ Vinh

# Các loại giả thuyết

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

**Kiểm định giả thuyết nhằm mục đích bác bỏ hoặc không bác bỏ  $H_0$ .  
*Không nên kết luận là chấp nhận  $H_0$ .***

# Xây dựng giả thiết trung bình tổng thể (mẫu lớn)

- Giả thuyết “có thay đổi”:  $H_0: \mu = \mu_0$   
 $H_a: \mu \neq \mu_0$

- Giả thuyết “thay đổi lớn hơn”:  $H_0: \mu \leq \mu_0$   
 **$H_a: \mu > \mu_0$**

*Lưu ý: Chúng ta phải bác bỏ  $H_0$  để giả thuyết  $H_a$  đúng*

- Giả thuyết “thay đổi nhỏ hơn”:  $H_0: \mu \geq \mu_0$   
 **$H_a: \mu < \mu_0$**

*Lưu ý: Chúng ta phải bác bỏ  $H_0$  để giả thuyết  $H_a$  đúng*

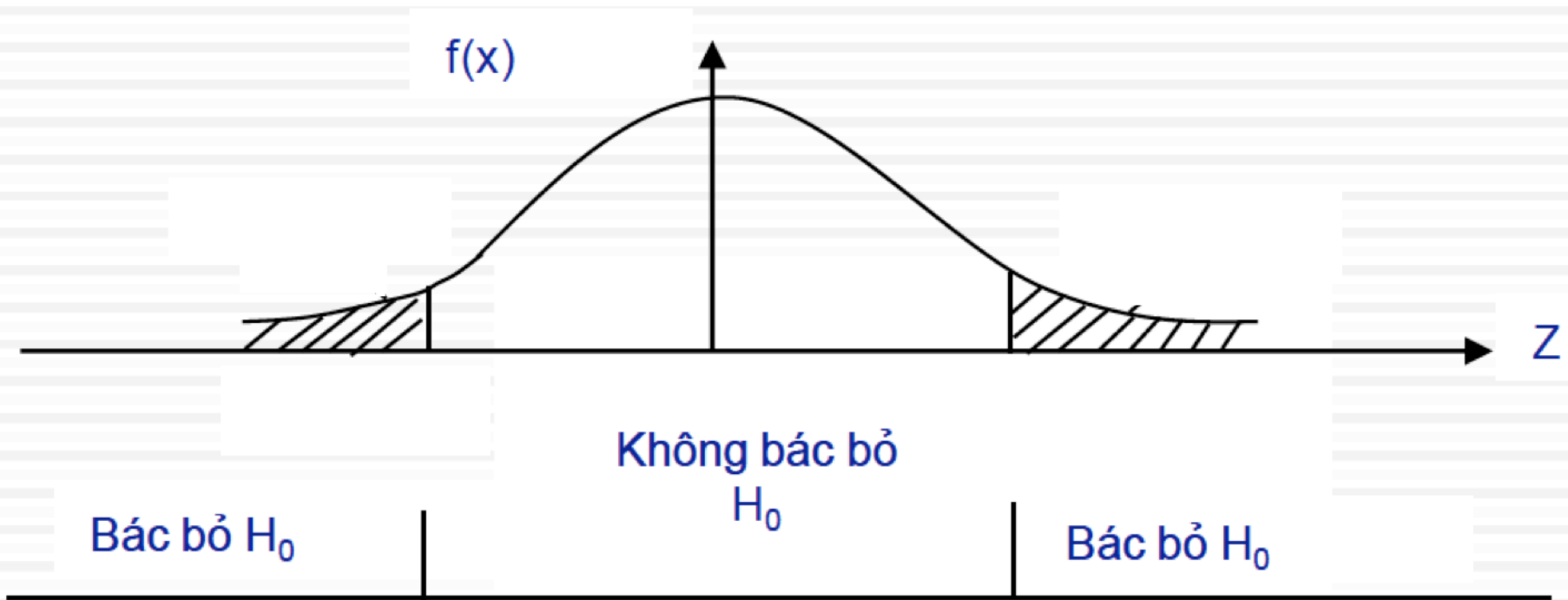
Trong đó  $\mu_0$  là giá trị cho trước.

# Kiểm định giả thuyết “có thay đổi”

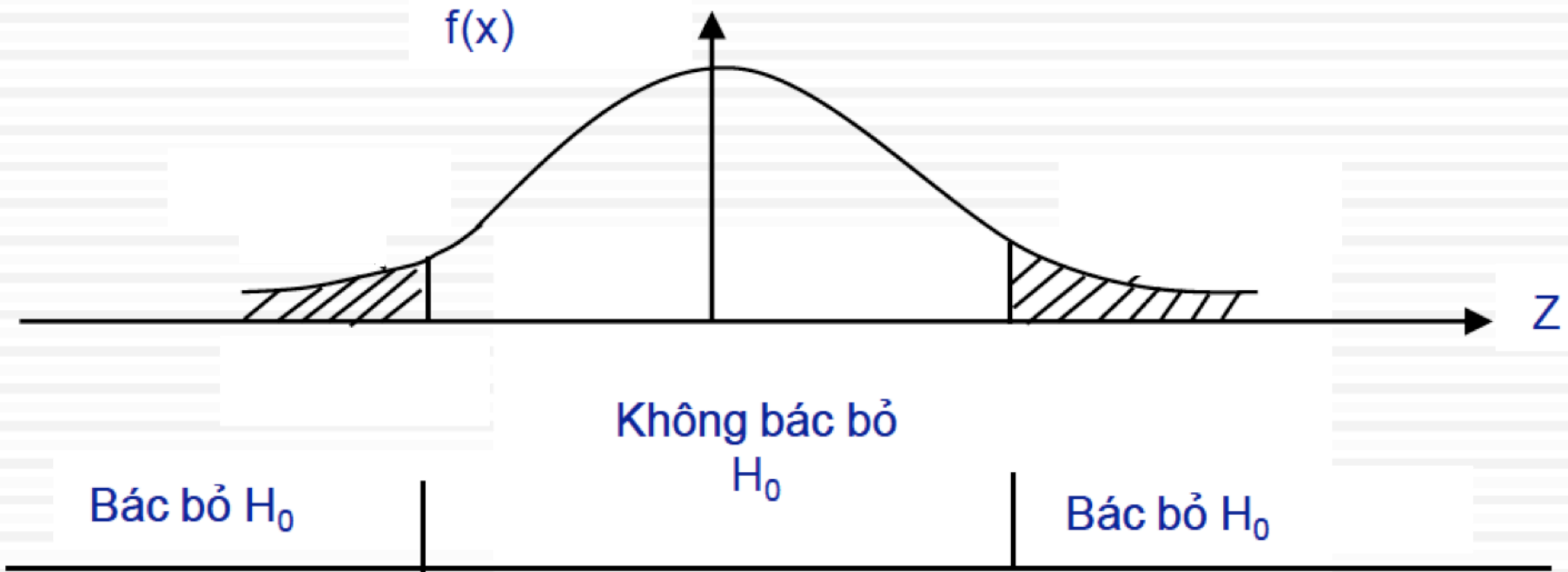
$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

P-value là *mức ý nghĩa nhỏ nhất mà vẫn có thể* bác bỏ được  $H_0$ .



# Kiểm định giả thuyết “có thay đổi”



Tính giá trị 
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$P\text{-value} = P(z < -|Z|) + P(z > |Z|) = 2 \times P(z > |Z|)$$

# Kiểm định giả thuyết “có thay đổi”

- Một người nông dân sử dụng 1 loại phân bón mới cho 1 vườn táo và thu được 2756kg trên một 100 cây. Biết rằng mức trung bình khi chưa sử dụng loại phân bón mới này là 26kg/1 cây với độ lệch chuẩn là 5,2kg. Hãy tìm giá trị p-value của giả thuyết loại phân bón mới làm thay đổi năng suất.
- Sau khi thay đổi giám đốc mới, 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 tìm giá trị p-value của 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 cách đây 1 năm.

# Kiểm định giả thuyết “thay đổi lớn hơn”

$H_0: \mu \leq \mu_0$

$H_a: \mu > \mu_0$

Cách tính giá trị p-value để bác bỏ  $H_0$ .

# Kiểm định giả thuyết “thay đổi lớn hơn”

- Một người nông dân sử dụng 1 loại phân bón mới cho 1 vườn táo và thu được 2956kg trên một 100 cây. Biết rằng mức trung bình khi chưa sử dụng loại phân bón mới này là 26kg/1 cây với độ lệch chuẩn là 5,2kg. Hãy tính p-value của giả thuyết sản lượng của cây táo tăng lên bởi loại phân bón này.
- Sau khi thay đổi giám đốc mới, nhà máy sản xuất thép ghi nhận sản lượng trong 100 ngày, có trung bình và độ lệch chuẩn của mẫu lần lượt là 980 tấn và 50 tấn. Hãy tính p-value của giả thuyết rằng sản lượng bình quân hàng ngày của nhà máy hiện tăng hơn so với mức sản lượng trung bình 892 tấn/ngày đã được ghi nhận cách đây 1 năm.



# Kiểm định giả thuyết “thay đổi nhỏ hơn”

$$H_0: \mu \geq \mu_0$$

$$H_a: \mu < \mu_0$$

Cách tính p-value?

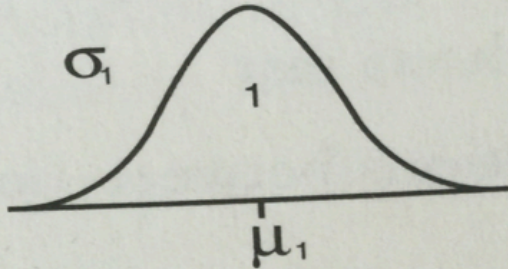
# Kiểm định giả thuyết “thay đổi nhỏ hơn”

- Một người nông dân sử dụng 1 loại phân bón mới cho 1 vườn táo và thu được 2356kg trên một 100 cây. Biết rằng mức trung bình khi chưa sử dụng loại phân bón mới này là 26kg/1 cây với độ lệch chuẩn là 5,2kg. Tính p-value của giả thuyết sản lượng của cây táo bị giảm đi do loại phân bón này.
- Sau khi thay đổi giám đốc mới, nhà máy sản xuất thép ghi nhận sản lượng trong 100 ngày, có trung bình và độ lệch chuẩn của mẫu lần lượt là 820 tấn và 50 tấn. Tính p-value của giả thuyết rằng sản lượng bình quân hàng ngày của nhà máy hiện tại giảm đi so với mức sản lượng trung bình 892 tấn/ngày đã được ghi nhận cách đây 1 năm.

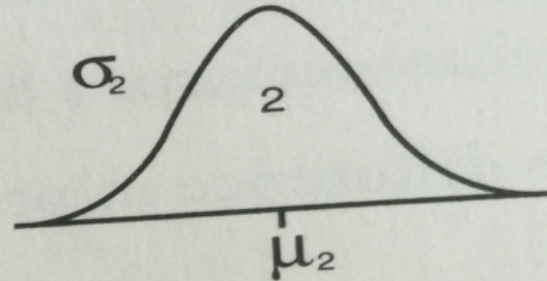
# Kiểm định giả thuyết với 2 mẫu

So sánh sự khác biệt giữa hai quần thể. Ví dụ như kiểm thử xem có sự khác biệt về điểm thi môn xác suất thống kê giữa sinh viên nữ và sinh viên nam?

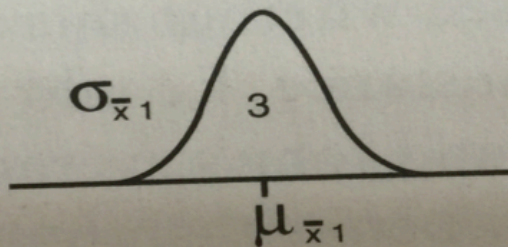
Quần thể 1



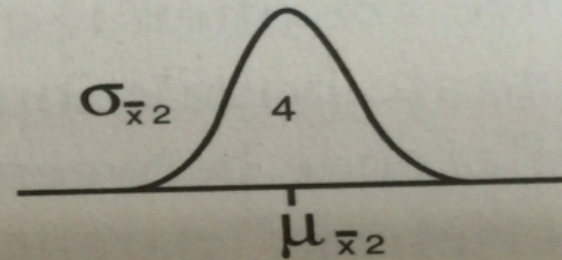
Quần thể 2



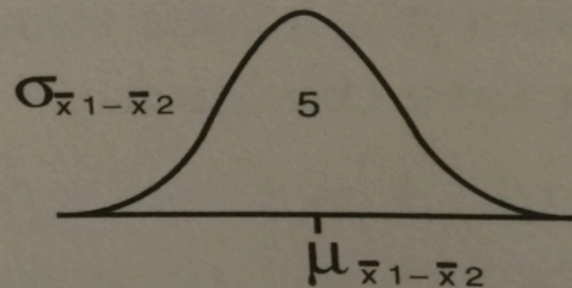
Phân bố của trung bình mẫu 1



Phân bố của trung bình mẫu 2



Phân bố của sự khác biệt giữa trung bình mẫu 1 và mẫu 2



# Kiểm định giả thuyết với 2 mẫu

Biến ngẫu nhiên về khác biệt giữa mẫu 1 và mẫu 2 sẽ có phân bố chuẩn với

Kì vọng:  $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2}$

Phương sai:  $\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

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

- $n_1$ , và  $n_2$  là kích thước của mẫu từ quần thể 1 và quần thể 2

- $\sigma_1$  và  $\sigma_2$  là độ lệch chuẩn của quần thể 1 và quần thể 2.

Lưu ý:  $\sigma_1$  và  $\sigma_2$  có thể được lượng từ độ lệch chuẩn của mẫu lấy từ quần thể 1 và quần thể 2 nếu kích thước mẫu  $\geq 30$ .

# Kiểm định giả thuyết trung bình của hai quần thể khác nhau

Kiểm định giả thuyết trung bình của quần thể 1 và quần thể 2 khác nhau:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Tính giá trị kiểm định z:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

# Kiểm định giả thuyết trung bình của hai quần thể khác nhau

Quần thể	Điểm	Độ lệch chuẩn	Kích thước mẫu
Nam	6.5	0.5	50
Nữ	7.0	0.7	30

Kiểm định giả thuyết không có sự khác biệt giữa điểm của các bạn nam và các nữ.

# Kiểm định giả thuyết trung bình của hai quần thể

Quần thể	Lương	Độ lệch chuẩn	Kích thước mẫu
Công ty A	12t	1.5	100
Công ty B	8.5t	1.2	64

Kiểm định giả thuyết lương công ty A hơn lương công ty B là 3 triệu.