



Kiểm định giả thuyết

Giảng viên: Hoàng Thị Điệp
Khoa CNTT – Đại học Công Nghệ

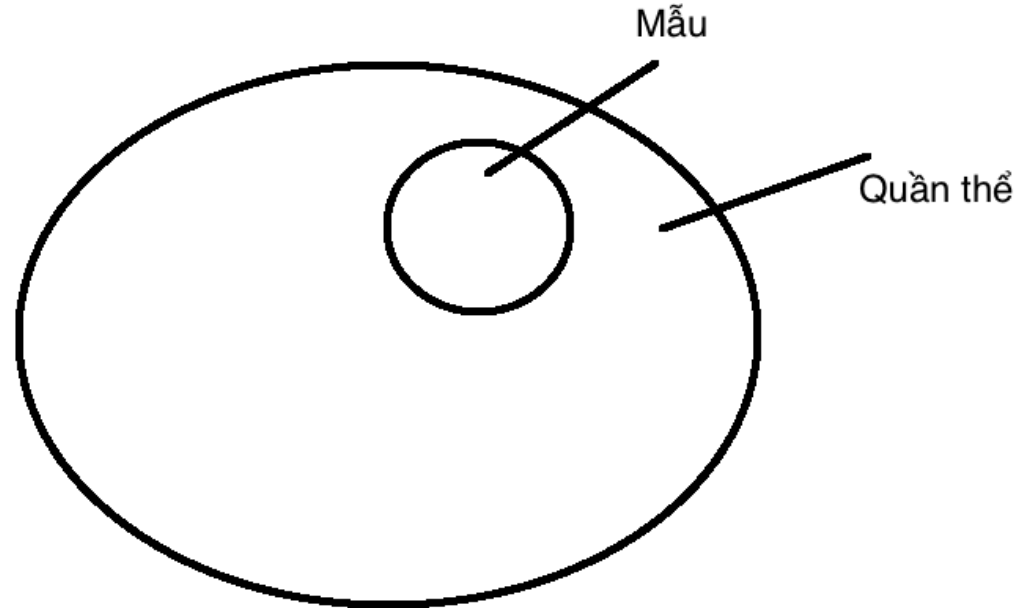
Nội dung

- » Giới thiệu kiểm định giả thuyết
- » Kiểm định giả thuyết về trung bình tổng thể
 - Trường hợp cỡ mẫu lớn
 - Trường hợp cỡ mẫu nhỏ hoặc chưa biết phương sai tổng thể
- » Kiểm định giả thuyết về giá trị xác suất

Ví dụ giả thuyết

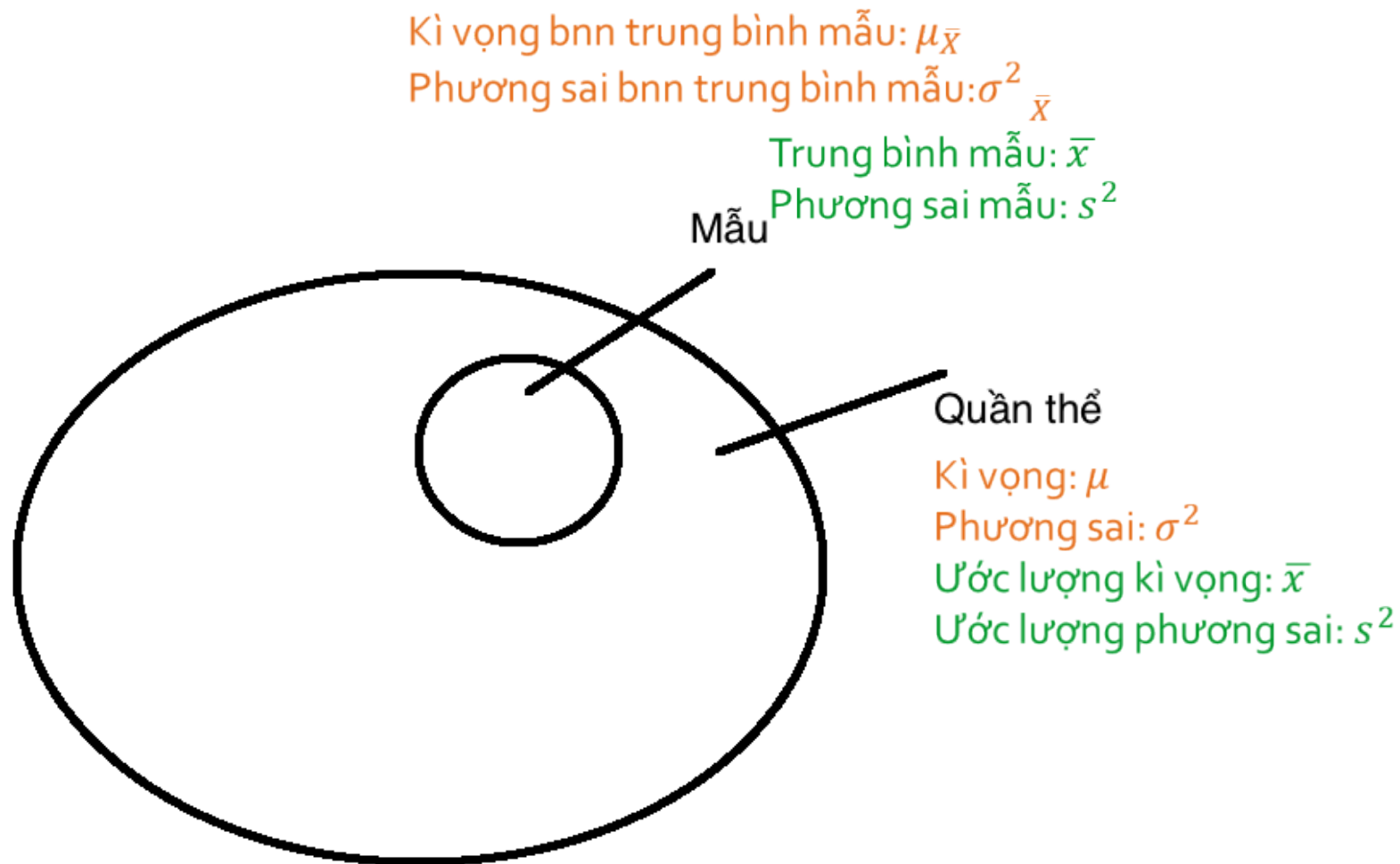
- » Ví dụ 1: Năm 2009 hãng Abbott giới thiệu một công thức sữa mới dành cho em bé dưới một tuổi. Liệu công thức này có tạo ra tăng trọng trung bình **cao hơn** mức 100gram/tháng của công thức sữa được giới thiệu vào 2007.
- » Ví dụ 2: Liệu tiền lương trung bình của công nhân cơ khí tại Bình Dương **có khác** với mức tiền lương trung bình là 2,5 triệu đồng của công nhân cơ khí trên toàn quốc không?
- » Ví dụ 3: Liệu sản lượng bình quân hàng ngày của nhà máy sản xuất thép hiện nay **có khác** với mức sản lượng trung bình 892 tấn/ngày đã ghi nhận cách đây 1 năm.

Vì sao cần kiểm định giả thuyết



Nghiên cứu một thuộc tính của quần thể dựa vào 1 tập mẫu.
Sử dụng dữ liệu thu được từ tập mẫu để kiểm định giả thuyết về các thuộc tính của quần thể.

Kí hiệu



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 thuyết về trung bình tổng thể

- » Giả thuyết “có thay đổi”:
 - $H_0: \mu = \mu_0$
 - $H_a: \mu \neq \mu_0$
- » Giả thuyết “thay đổi lớn hơn”:
 - $H_0: \mu = \mu_0$
 - $H_a: \mu > \mu_0$
- » Giả thuyết “thay đổi nhỏ hơn”:
 - $H_0: \mu = \mu_0$
 - $H_a: \mu < \mu_0$
- » Lưu ý: Chúng ta phải bác bỏ H_0 để giả thuyết H_a đúng
- » Trong đó μ_0 là giá trị cho trước.

Các bước kiểm định giả thuyết

- » 1. Thiết lập hai giả thuyết đối lập
- » 2. Xác định mức ý nghĩa
- » 3. Chọn lựa kiểm định phù hợp
- » 4. Tính chỉ số thống kê tương ứng
- » 5. Ra quyết định về giả thuyết không (bác bỏ hay không bác bỏ)
- » 6. Ra một kết luận tổng thể

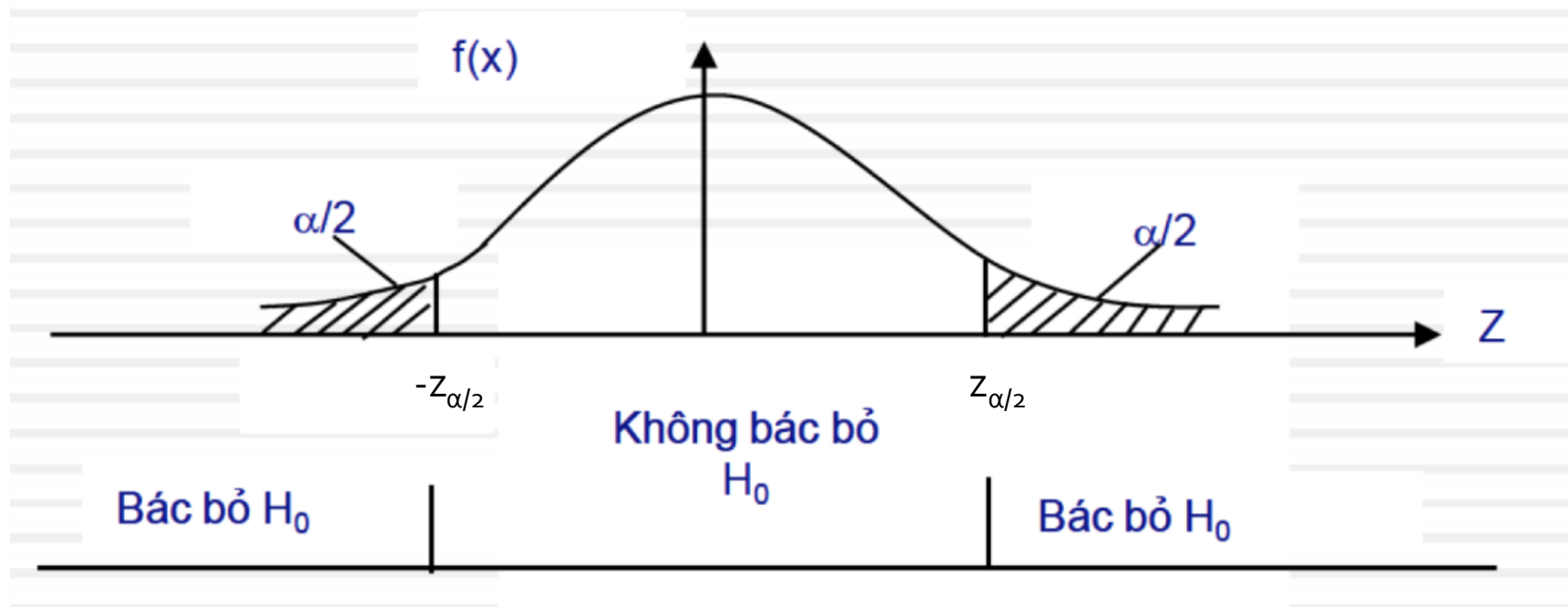
Trường hợp cỡ
mẫu lớn

Kiểm định giả thuyết “có thay đổi”

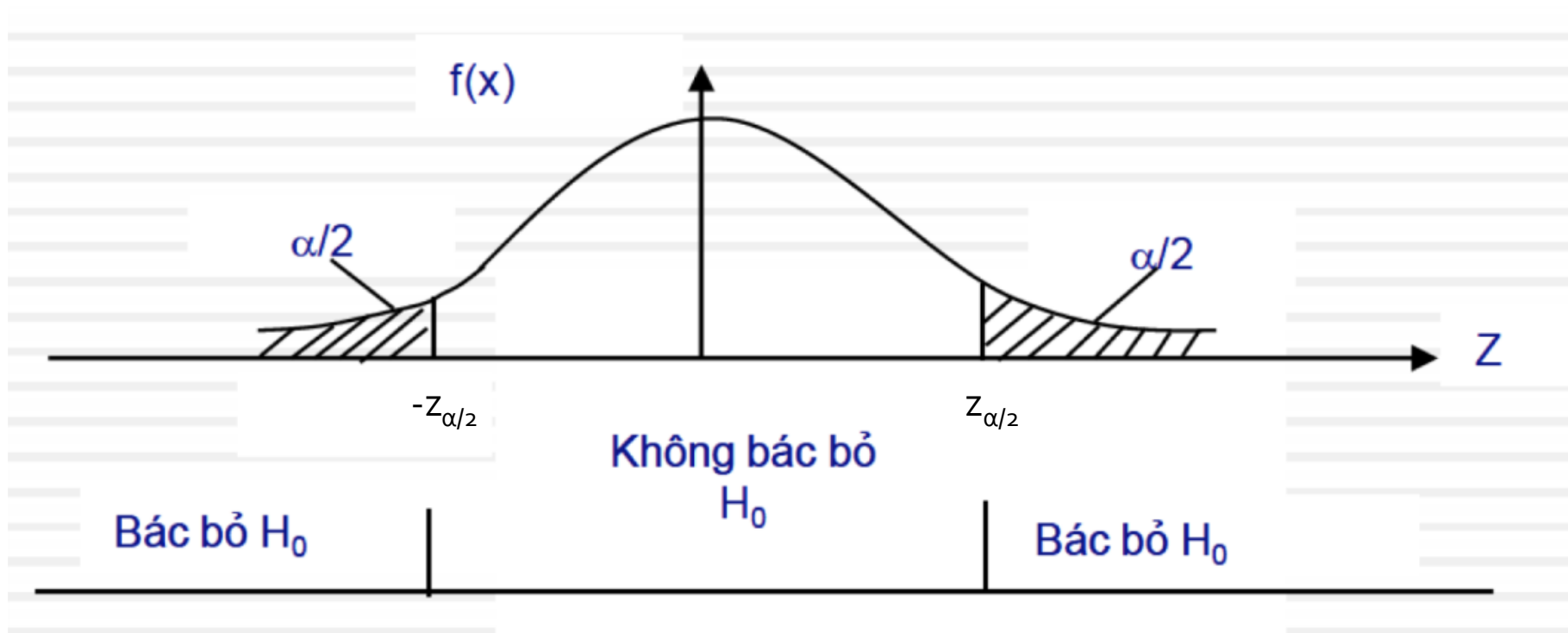
$$H_0: \mu = \mu_0$$

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

- » Đây thường là kiểm thử giả thuyết một số yếu tố thay đổi và làm thay đổi một thuộc tính nào đó của quần thể.
- » Kiểm định 2 phía với α là mức ý nghĩa. Quy tắc bác bỏ H_0 như sau



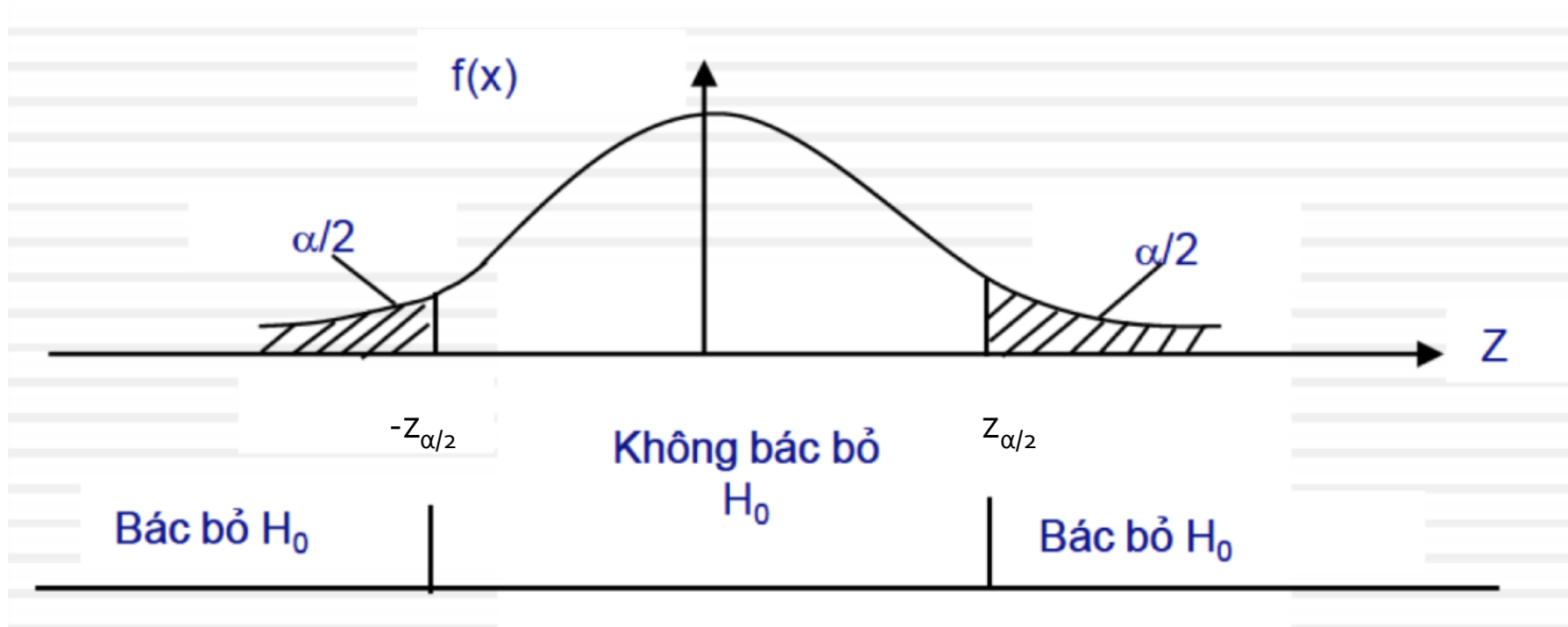
Kiểm định giả thuyết “có thay đổi”



Giá trị kiểm định:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Kiểm định giả thuyết “có thay đổi”



So sánh giá trị kiểm định z với giá trị $-z_{\alpha/2}$ và $z_{\alpha/2}$.
Nếu $z < -z_{\alpha/2}$ hoặc $z > z_{\alpha/2}$ thì bác bỏ giả thuyết H_0 ($\mu = \mu_0$).
Nếu không thì không bác bỏ giả thuyết H_0 .

Bài tập

1. Một người nông dân sử dụng 1 loại phân bón mới cho 1 vườn táo và thu được 2756kg trên 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 kiểm định giả thuyết sản lượng của cây táo có thay đổi bởi loại phân bón này với mức ý nghĩa là 0.05.
 - » 1. Thiết lập hai giả thuyết đối lập
 - » 2. Xác định mức ý nghĩa
 - » 3. Chọn lựa kiểm định phù hợp
 - » 4. Tính chỉ số thống kê tương ứng
 - » 5. Ra quyết định về giả thuyết không (bác bỏ hay không bác bỏ)
 - » 6. Ra một kết luận tổng thể

Bài tập

2. 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/ngày 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 cách đây 1 năm với mức ý nghĩa là 0.05.

- » 1. Thiết lập hai giả thuyết đối lập
- » 2. Xác định mức ý nghĩa
- » 3. Chọn lựa kiểm định phù hợp
- » 4. Tính chỉ số thống kê tương ứng
- » 5. Ra quyết định về giả thuyết không (bác bỏ hay không bác bỏ)
- » 6. Ra một kết luận tổng thể

Bài tập

3. Một nhà máy sản xuất sản phẩm ô tô tuyên bố rằng tuổi thọ trung bình một chiếc ô tô của họ là 29000 dặm. Cơ quan giám định nghi ngờ lời tuyên bố này đã kiểm tra 100 chiếc ô tô và tìm được trung bình mẫu là $\bar{x} = 28000$ dặm với độ lệch tiêu chuẩn là 5000 dặm.

- a) Với mức ý nghĩa $\alpha = 0.05$ cơ quan giám định có bác bỏ được lời quảng cáo của nhà máy nói trên không?
- b) Cũng với câu hỏi trên nhưng với mức ý nghĩa được chọn là $\alpha = 0.02$.

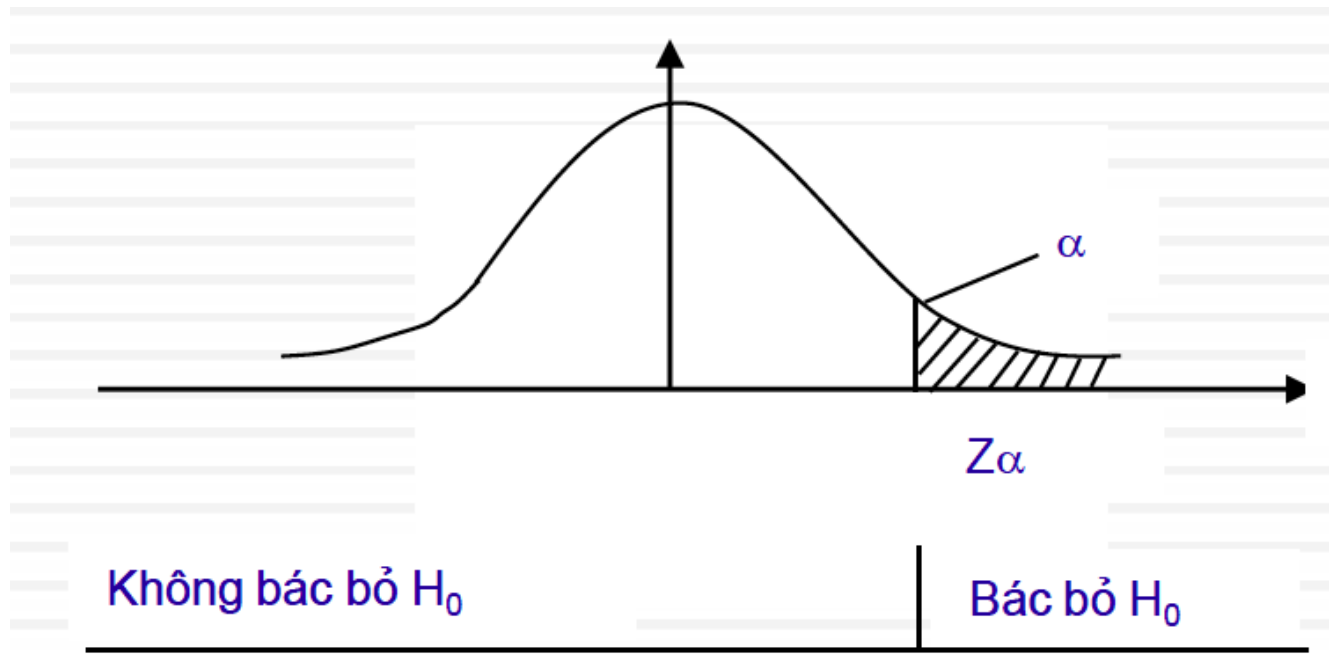
$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Kiểm định giả thuyết “thay đổi lớn hơn”

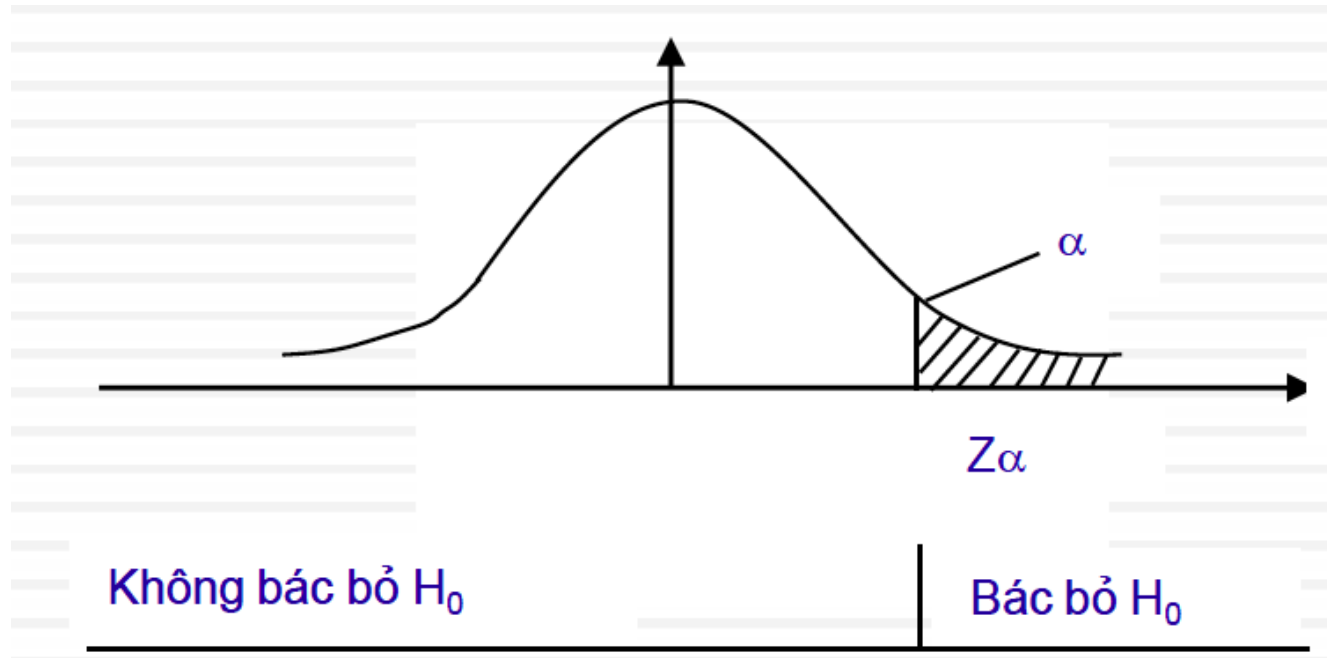
$$H_0: \mu = \mu_0$$

$$H_a: \mu > \mu_0$$

- » Đây thường là kiểm thử giả thuyết một số yếu tố thay đổi dẫn đến thay đổi tăng thêm một thuộc tính nào đó của quần thể. Lưu ý: Chúng ta phải bác bỏ H_0 để giả thuyết H_a đúng
- » Kiểm định 1 phía với α là mức ý nghĩa. Quy tắc bác bỏ H_0 như sau



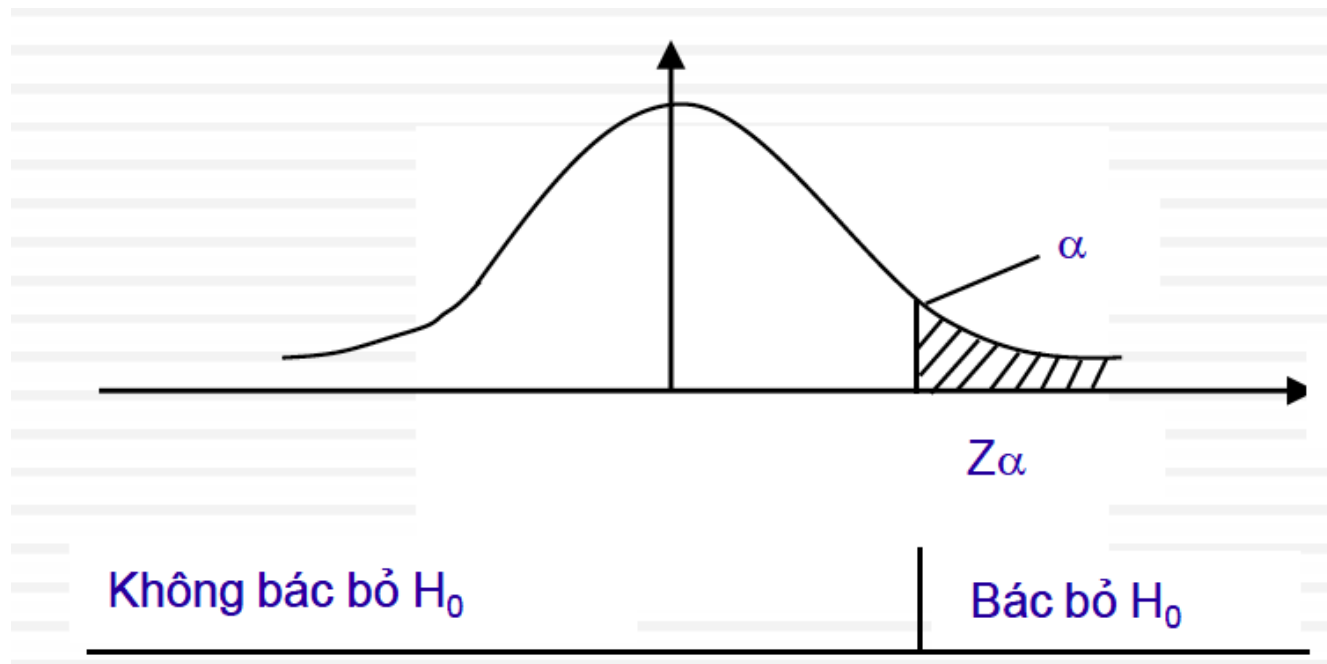
Kiểm định giả thuyết “thay đổi lớn hơn”



Giá trị kiểm định:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Kiểm định giả thuyết “thay đổi lớn hơn”



So sánh giá trị kiểm định z với giá trị z_α .

Nếu $z > z_\alpha$ thì bác bỏ giả thuyết H_0 ($\mu \leq \mu_0$) hay **H_a đúng**.

Nếu không thì không bác bỏ giả thuyết H_0 .

Bài tập

1. 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 kiểm định giả thuyết sản lượng của cây táo tăng lên bởi loại phân bón này với mức ý nghĩa là 0.05.
2. 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 kiểm định 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 với mức ý nghĩa là 0.05.

Bài tập

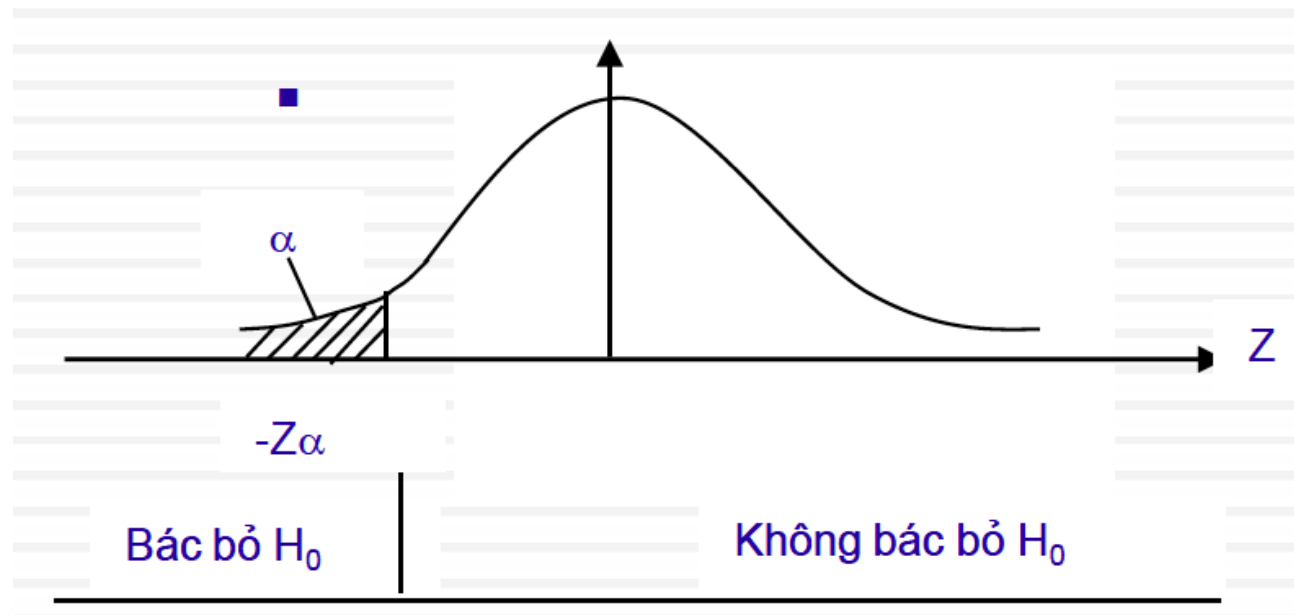
3. Một công ty có hệ thống máy tính có thể xử lý 1200 hóa đơn trong một giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lý trung bình trong 1 giờ là 1260 với độ lệch tiêu chuẩn là 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Kiểm định giả thuyết “thay đổi nhỏ hơn”

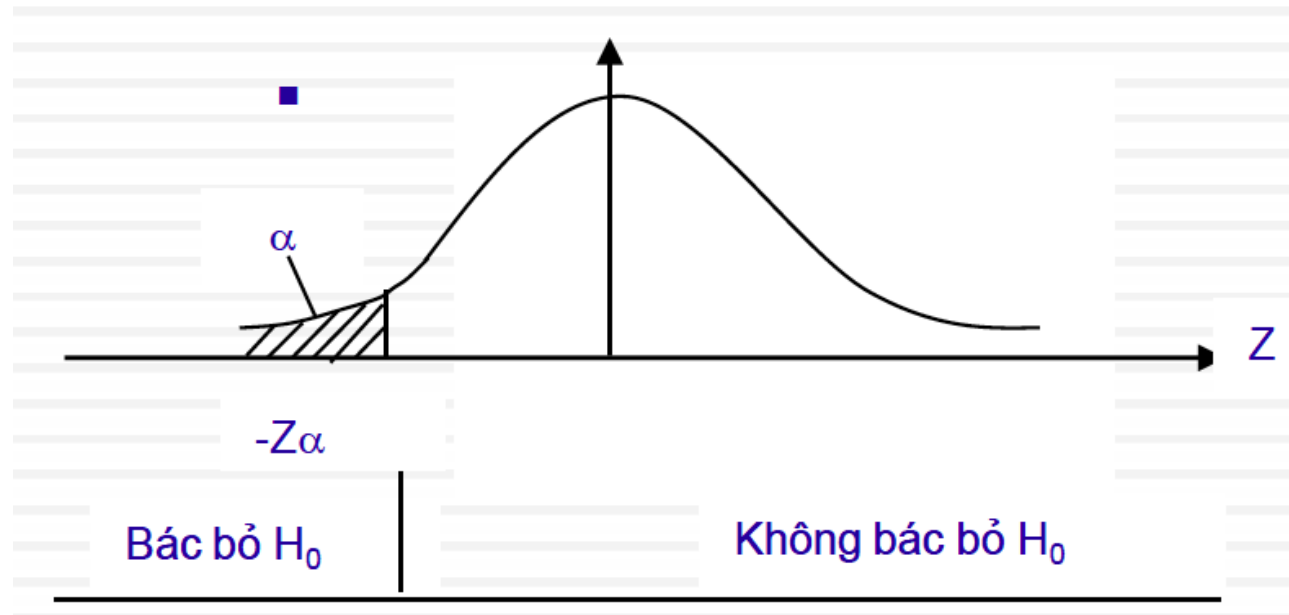
$$H_0: \mu = \mu_0$$

$$H_a: \mu < \mu_0$$

- » Đây thường là kiểm thử giả thuyết một số yếu tố thay đổi dẫn đến thay đổi nhỏ đi một thuộc tính nào đó của quần thể. Lưu ý: Chúng ta phải bác bỏ H_0 để giả thuyết H_a đúng
- » Kiểm định 1 phía với α là mức ý nghĩa. Quy tắc bác bỏ H_0 như sau



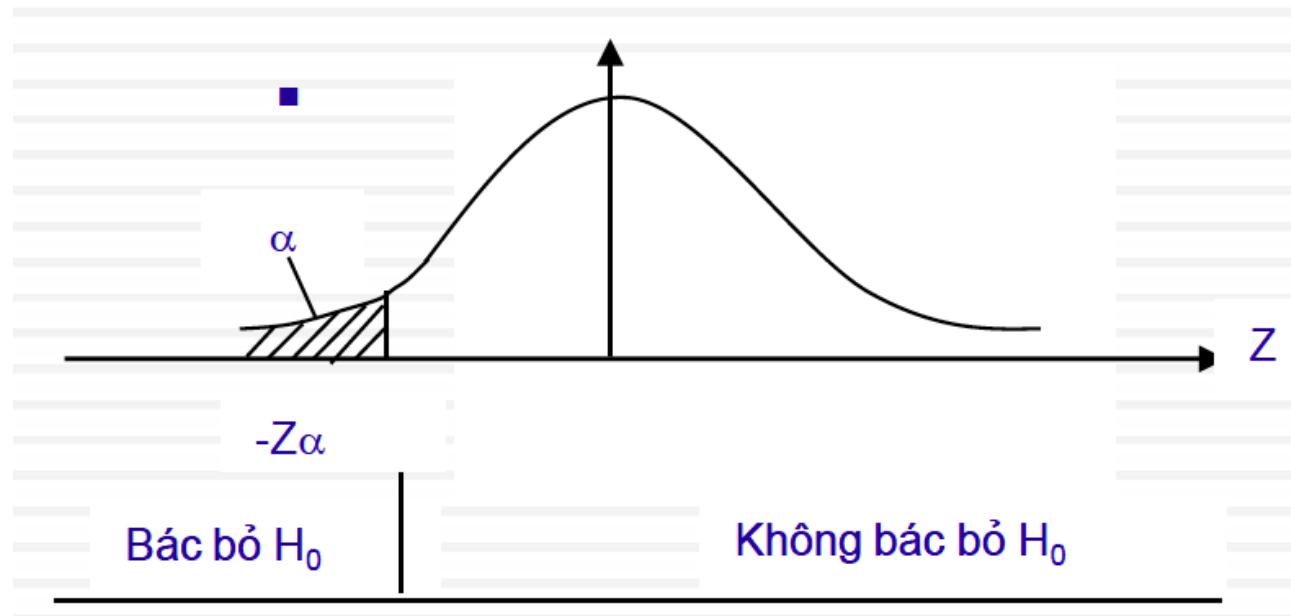
Kiểm định giả thuyết “thay đổi nhỏ hơn”



Giá trị kiểm định:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Kiểm định giả thuyết “thay đổi nhỏ hơn”



So sánh giá trị kiểm định z với giá trị $-z_\alpha$.

Nếu $z < -z_\alpha$ thì bác bỏ giả thuyết H_0 ($\mu \geq \mu_0$), tức là **Ha đúng**.

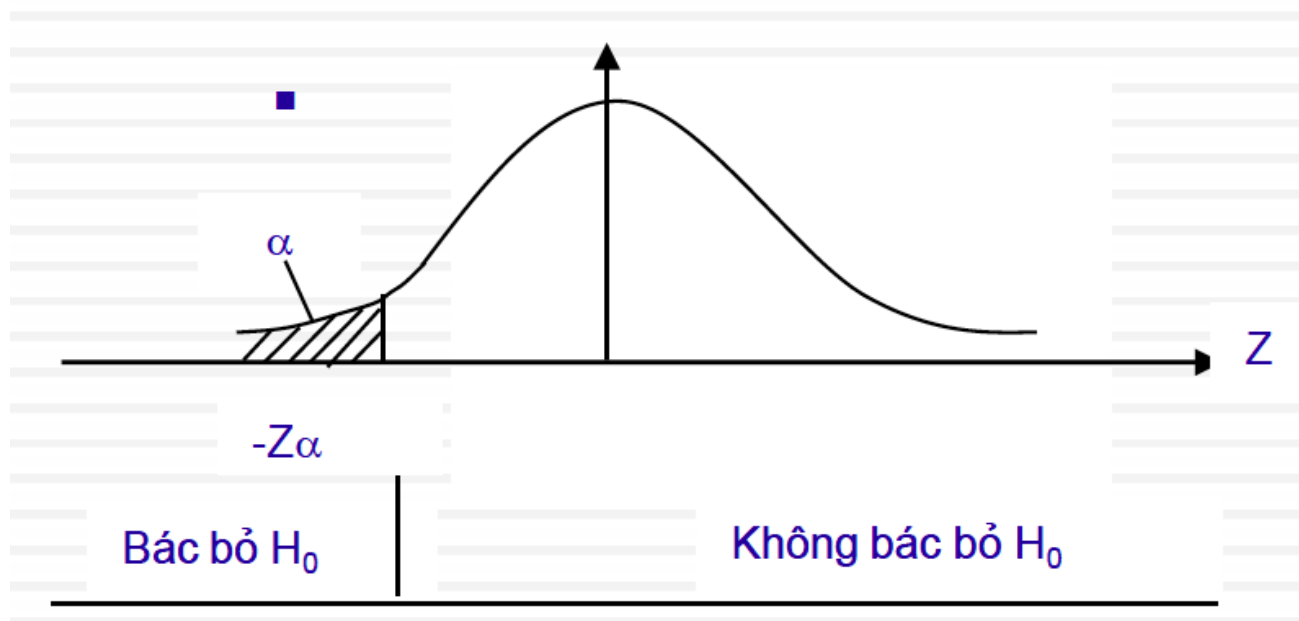
Nếu không thì không bác bỏ giả thuyết H_0 .

Bài tập

1. 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. Hãy kiểm định giả thuyết sản lượng của cây táo bị giảm đi do loại phân bón này với mức ý nghĩa là 0.05.
2. 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. Hãy kiểm định 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 với mức ý nghĩa là 0.05.

Bài tập

3. Một phương pháp ăn kiêng được quảng cáo rằng sẽ làm giảm trọng lượng ít nhất là 45 pound trong 6 tháng. Một mẫu gồm 36 người theo chế độ ăn kiêng này giảm trọng lượng trung bình là 35 pound với độ lệch tiêu chuẩn là 20 pound. Với mức ý nghĩa $\alpha = 0.025$, hãy nhận định xem quảng cáo trên có nói quá không?



Trường hợp cỡ
mẫu nhỏ hoặc
chưa biết
phương sai
tổng thể

Phương sai chưa biết

» Nếu phương sai của quần thể chưa biết, có thể được ước lượng bằng phương sai mẫu.

- Trung bình mẫu:

$$\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$$

- Phương sai mẫu:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

» Nếu phương sai của quần thể chưa biết và tập mẫu nhỏ ($n < 30$)

- \bar{X} có phân bố Student với $(n-1)$ bậc tự do; kì vọng μ và phương sai $\sigma_{\bar{x}}^2 = s^2/n$.

Bài tập

1. Một nhóm nghiên cứu công bố rằng trung bình một người vào siêu thị A tiêu hết 140 ngàn đồng. Chọn một mẫu ngẫu nhiên gồm 50 người mua hàng, ta tính được số tiền trung bình họ tiêu là 154 nghìn với độ lệch tiêu chuẩn là 62 nghìn. Với mức ý nghĩa 0.02 hãy kiểm tra xem công bố của nhóm nghiên cứu có đúng hay không?
2. Một công ty có hệ thống máy tính có thể xử lý 1200 hóa đơn trong một giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lý trung bình trong 1 giờ là 1260 với độ lệch tiêu chuẩn là 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Bài tập

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

3. Một công ty sản xuất pin tuyên bố rằng pin của họ có tuổi thọ trung bình là 21.5 giờ. Một cơ quan kiểm tra chất lượng kiểm tra 6 chiếc pin của công ty và thu được số liệu sau đây về tuổi thọ của 6 chiếc pin này:

19, 18, 22, 20, 16, 25.

Kết quả này có xác nhận là quảng cáo của công ty là đúng hay không? Mức ý nghĩa được chọn là 5%.

4. Một bản nghiên cứu thông báo rằng mức tiêu dùng hàng tháng của một sinh viên là 420 nghìn đồng. Để kiểm tra, người ta chọn ngẫu nhiên 16 sinh viên và tìm được trung bình mỗi tháng họ tiêu 442 nghìn đồng với độ lệch tiêu chuẩn là 60 nghìn đồng.

Với mức ý nghĩa 5% nhận định xem kết luận của bản thông báo có thấp hơn sự thật hay không?

Đề thi mẫu

- » Một nhà sản xuất tuyên bố rằng loại pin mới được cải tiến của ông ta tuổi thọ dài hơn loại pin cũ. Biết rằng, loại pin cũ có tuổi thọ tuân theo phân bố chuẩn với kì vọng toán là 150 giờ và phương sai là 16 giờ². Để kiểm tra, người ta đo tuổi thọ của 9 pin loại mới được chọn một cách ngẫu nhiên và tính được trung bình mẫu là 153 giờ. Giả sử rằng phương sai của loại pin mới không thay đổi so với loại pin cũ.

Hãy kiểm định giả thiết $H_0: \mu = 150$ và đối thiết $H_1: \mu > 150$ với mức ý nghĩa 1%.

Biết $z_{0.05} = 1.645$; $z_{0.025} = 1.96$; $z_{0.01} = 2.326$; $z_{0.005} = 2.576$

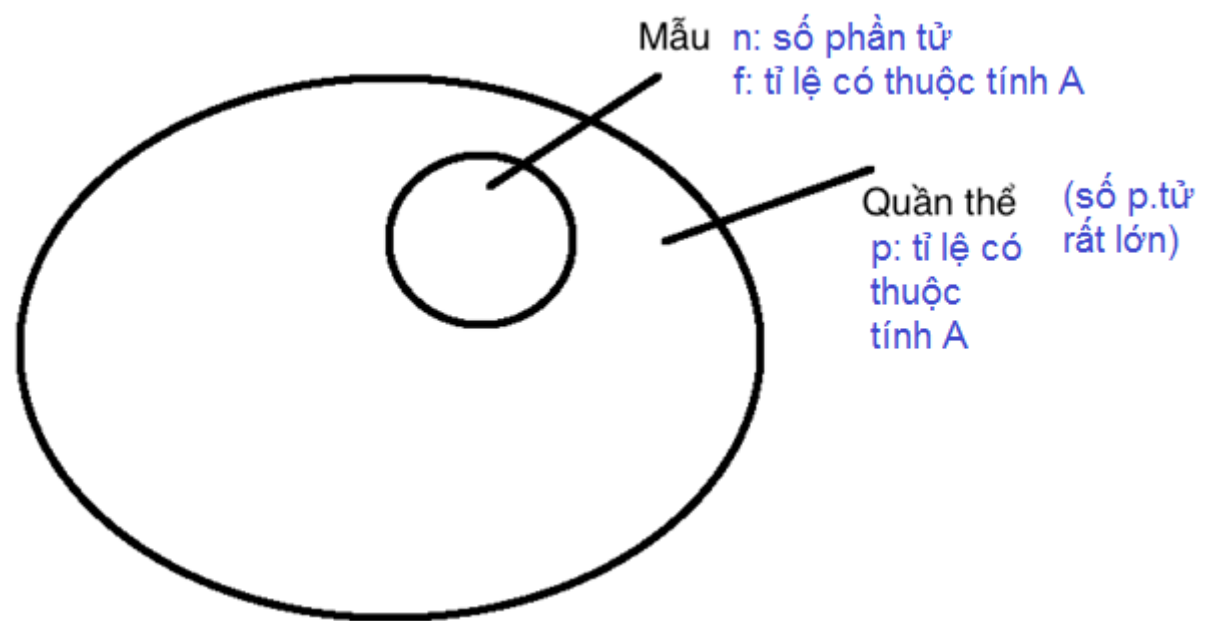
$t_{0.05;8} = 1.860$; $t_{0.025;8} = 2.306$; $t_{0.01;8} = 2.896$; $t_{0.005;8} = 3.355$

Nội dung

- » Giới thiệu kiểm định giả thuyết
- » Kiểm định giả thuyết về trung bình tổng thể
 - Trường hợp cỡ mẫu lớn
 - Trường hợp cỡ mẫu nhỏ hoặc chưa biết phương sai tổng thể
- » Kiểm định giả thuyết về giá trị xác suất

Kiểm định giả thuyết về giá trị xác suất

- » Nghiên cứu một quần thể mà mỗi cá thể có thể có hoặc không có một thuộc tính A nào đó.
 - p là tỉ lệ cá thể có thuộc tính A trong quần thể
 - $f = k/n$ là tỉ lệ (tần suất) cá thể có thuộc tính A trong mẫu nghiên cứu
- » Câu hỏi: Ta muốn kiểm định giả thuyết liên quan đến p dựa vào tần suất f .
- » Nhắc lại: Tần suất F là một ĐLNN có phân bố xấp xỉ phân bố chuẩn với kì vọng $EF = p$ và phương sai $DF = p(1-p)/n$ với điều kiện $np > 5$ và $n(1-p) > 5$.
- » **Trong bài toán ước lượng**, do không biết p , cho nên DF có thể được xấp xỉ bằng
$$DF = f(1-f)/n \text{ với điều kiện } nf > 10 \text{ và } n(1-f) > 10.$$
- » **Trong bài toán kiểm định**, $H_0: p = p_0$, ta giả sử H_0 đúng cho nên DF có thể được xấp xỉ bằng
$$DF = p_0(1-p_0)/n \text{ với điều kiện } np_0 > 5 \text{ và } n(1-p_0) > 5.$$



Bài tập

1. Một đảng chính trị trong một cuộc bầu cử tổng thống ở Mỹ tuyên bố rằng 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của họ. Chọn ngẫu nhiên 200 cử tri để thăm dò ý kiến cho thấy 80 người trong số đó tuyên bố sẽ bỏ phiếu cho A.

Hãy kiểm thử dự đoán của đảng trên với mức ý nghĩa

- a) 2%
- b) 5%
- c) 1%

Kết luận: Ta chưa đủ cơ sở để bác bỏ tuyên bố rằng 45% cử tri sẽ bỏ phiếu cho ứng cử viên A.

Bài tập

2. Một báo cáo nói rằng 18% gia đình ở thành phố A có máy tính cá nhân ở nhà. Để kiểm tra, người ta chọn ngẫu nhiên 80 gia đình trong thành phố có trẻ em đang đi học và thấy rằng có 22 gia đình có máy tính.
- a) Với mức ý nghĩa 2% hãy kiểm định xem liệu trong các gia đình có trẻ em đang đi học, tỉ lệ gia đình có máy tính có cao hơn tỉ lệ chung hay không.
 - b) Hãy kiểm thử với mức ý nghĩa 5%
 - c) Hãy kiểm thử với mức ý nghĩa 1%

Bài tập

3. Một công ty A sản xuất bánh kẹo tuyên bố rằng $\frac{2}{3}$ số trẻ em thích ăn bánh của công ty. Trong một mẫu gồm 100 trẻ em được hỏi, có 55 em tỏ ra thích bánh của công ty A.

- a) Với mức ý nghĩa 5%, số liệu nói trên có chứng tỏ là tuyên bố của công ty là hơi quá hay không?
- b) Hãy kiểm thử với mức ý nghĩa 2%.
- c) Hãy kiểm thử với mức ý nghĩa 1%.