Chương 5
KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

BỘ MÔN TOÁN ỨNG DỤNG[1]

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
ĐẠI HỌC BÁCH KHOA HÀ NỘI

SAMI.HUST – 2023

[1] Phòng BIS.201-D3.5

5.3. SO SÁNH

Bài toán

Bài toán 4

Giả sử X_1 và X_2 là hai biến ngẫu nhiên gốc cùng mô tả về một đặc trưng thống kê và được xét trên hai tổng thể và giả sử $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, trong đó, $E(X_i) = \mu_i$, $i = 1, 2$ chưa biết. Từ X_1 và X_2, xây dựng hai mẫu ngẫu nhiên tương ứng $W_{X_1} = (X_{11}, X_{12}, \ldots, X_{1n_1})$ kích thước n_1 và $W_{X_2} = (X_{21}, X_{22}, \ldots, X_{2n_2})$ kích thước n_2. Bài toán đặt ra là cần so sánh hai kỳ vọng μ_1 với μ_2 dựa trên các mẫu quan sát $W_{x_1} = (x_{11}, x_{12}, \ldots, x_{1n_1})$ và $W_{x_2} = (x_{21}, x_{22}, \ldots, x_{2n_2})$.

✍ **Các cặp giả thuyết**

Ta cần kiểm định giả thuyết so sánh hai kỳ vọng ở một trong ba dạng của cặp giả thuyết sau (Δ_0 là số đã biết):

1. $H_0 : \mu_1 - \mu_2 = \Delta_0;\ H_1 : \mu_1 - \mu_2 \neq \Delta_0$
2. $H_0 : \mu_1 - \mu_2 = \Delta_0;\ H_1 : \mu_1 - \mu_2 > \Delta_0$
3. $H_0 : \mu_1 - \mu_2 = \Delta_0;\ H_1 : \mu_1 - \mu_2 < \Delta_0$

So sánh hai kỳ vọng, hai phương sai đã biết

Phân phối mẫu

- Giả sử hai biến ngẫu nhiên gốc X_1 và X_2 là độc lập.
- Với $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$ và $\overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$, thì

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0;1). \tag{18}$$

So sánh hai kỳ vọng, hai phương sai đã biết

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}. \tag{19}$$

   Nếu giả thuyết $H_0 : \mu_1 - \mu_2 = \Delta_0$ là đúng, thì $Z_0 \sim \mathcal{N}(0; 1)$.

3. Miền bác bỏ giả thuyết H_0 được xác định phụ thuộc vào đối thuyết H_1.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 \neq \Delta_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$(z_\alpha; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 < \Delta_0$	$(-\infty; -z_\alpha)$

So sánh hai kỳ vọng, hai phương sai đã biết

Các bước tiến hành (tiếp theo)

4. Từ hai mẫu cụ thể $W_{x_1} = (x_{11}, x_{12}, \ldots, x_{1n_1})$ và $W_{x_2} = (x_{21}, x_{22}, \ldots, x_{2n_2})$, tính giá trị quan sát của tiêu chuẩn kiểm định

$$z_0 = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}. \tag{20}$$

5. Kiểm tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

So sánh hai kỳ vọng, hai phương sai đã biết

Ví dụ 11

Một nhà phát triển sản phẩm quan tâm đến việc giảm thời gian khô của sơn lót. Hai công thức của sơn được thử nghiệm. Công thức I là công thức tiêu chuẩn và Công thức II có thành phần làm khô mới nên hy vọng sẽ làm giảm thời gian làm khô. Từ kinh nghiệm cho thấy thời gian làm khô của sơn lót có luật phân phối chuẩn với độ lệch chuẩn là 8 phút và nó không bị ảnh hưởng khi bổ sung thành phần làm khô mới. Mười mẫu vật được sơn theo Công thức I và 10 mẫu khác được sơn theo Công thức II theo thứ tự ngẫu nhiên cho kết quả thời gian khô trung bình của hai mẫu lần lượt là $\overline{x}_1 = 121$ phút và $\overline{x}_2 = 112$ phút. Nhà phát triển sản phẩm có thể rút ra kết luận gì về hiệu quả của thành phần mới với mức ý nghĩa $\alpha = 0,05$?

Giải. Gọi X_1 (phút), X_2 (phút) lần lượt là các biến ngẫu nhiên chỉ thời gian làm khô của sơn sản xuất theo Công thức I và II, X_1 và X_2 độc lập và $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, $i = 1, 2$ với $\sigma_1 = \sigma_2 = 8$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trong trường hợp hai phương sai đã biết và $\Delta_0 = 0$.

So sánh hai kỳ vọng, hai phương sai đã biết

- Cặp giả thuyết cần kiểm định $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$.
- Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

  Nếu giả thuyết H_0 là đúng thì $Z_0 \sim \mathcal{N}(0;1)$.

- Với $\alpha = 0,05$, $z_\alpha = z_{0,05} = 1,645$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (1,645; +\infty).$$

- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n_1 = n_2 = 10$, $\overline{x}_1 = 121$, $\overline{x}_2 = 112$ và $\sigma_1 = \sigma_2 = 8$,

$$z_0 = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{121 - 112}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} \simeq 2,5156.$$

- Vì $z_0 = 2,5156 \in W_\alpha$ nên có cơ sở bác bỏ giả thuyết H_0. Như vậy, với dữ liệu thực nghiệm đã cho, việc thêm thành phần làm khô mới đã làm giảm thời gian làm khô của loại sơn lót này với mức ý nghĩa 5%.

So sánh hai kỳ vọng, hai mẫu kích thước lớn

Phân phối mẫu
Nếu $n_1 \geq 30$ và $n_2 \geq 30$ thì thống kê

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

So sánh hai kỳ vọng, hai mẫu kích thước lớn

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định:

$$Z_0 = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \tag{21}$$

   Nếu giả thuyết $H_0 : \mu_1 - \mu_2 = \Delta_0$ là đúng và nếu $n_1 \geq 30$ và $n_2 \geq 30$ thì thống kê Z_0 trong (21) có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

3. Miền bác bỏ giả thuyết H_0 được xác định cho ba trường hợp như sau.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 \neq \Delta_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$(z_\alpha; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 < \Delta_0$	$(-\infty; -z_\alpha)$

So sánh hai kỳ vọng, hai mẫu kích thước lớn

Các bước tiến hành (tiếp theo)

4. Từ hai mẫu cụ thể $W_{x_1} = (x_{11}, x_{12}, \ldots, x_{1n_1})$ và $W_{x_2} = (x_{21}, x_{22}, \ldots, x_{2n_2})$, tính giá trị quan sát của tiêu chuẩn kiểm định

$$z_0 = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}. \tag{22}$$

5. Kiểm tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

So sánh hai kỳ vọng, hai mẫu kích thước lớn

Ví dụ 12

Hai máy tự động dùng để cắt những thanh kim loại do cùng một kỹ thuật viên phụ trách và căn chỉnh. Từ mỗi máy lấy ra 31 thanh kim loại để kiểm tra và thu được kết quả sau:

- Máy I: Trung bình mẫu là 12 cm, độ lệch chuẩn hiệu chỉnh mẫu là 1,2 cm.
- Máy II: Trung bình mẫu là 12,3 cm, độ lệch chuẩn hiệu chỉnh mẫu là 1,4 cm.

Với mức ý nghĩa $\alpha = 0,01$ có thể cho rằng chiều dài của các thanh kim loại do Máy I và II sản xuất là như nhau hay không?

Giải. Gọi X_1 (cm), X_2 (cm) lần lượt là chiều dài các thanh kim loại do Máy I và II sản xuất, X_1 và X_2 là độc lập. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên trường hợp mẫu kích thước $n_1 = n_2 = 31 > 30$, $\Delta_0 = 0$.
Cặp giả thuyết cần kiểm định là $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.

So sánh hai kỳ vọng, hai mẫu kích thước lớn

- Chọn tiêu chuẩn kiểm định
$$Z_0 = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}.$$
Vì $n_1 = n_2 = 31 > 30$ và nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0;1)$.
- Với $\alpha = 0,01$, $z_{\alpha/2} = z_{0,005} = 2,576$. Miền bác bỏ giả thuyết H_0 là
$$W_\alpha = (-\infty; -2,576) \cup (2,576; +\infty).$$
- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n_1 = n_2 = 31$, $\overline{x}_1 = 12$, $s_1 = 1,2$, $\overline{x}_2 = 12,3$ và $s_2 = 1,4$:
$$z_0 = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{12 - 12,3}{\sqrt{\dfrac{1,44}{31} + \dfrac{1,96}{31}}} = -0,9059.$$
- Vì $z_0 = -0,9059 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Có thể xem chiều dài các thanh kim loại do hai nhà máy sản xuất là như nhau với mức ý nghĩa 1%.

So sánh hai kỳ vọng, hai phương sai chưa biết

✍ **Giả thiết**

Ta xét bài toán kiểm định giả thuyết hai mẫu về kỳ vọng trong trường hợp có giả thiết dưới đây.

($\mathcal{H}$) Hai biến ngẫu nhiên gốc có phân phối chuẩn với phương sai chưa biết nhưng được giả thiết là chúng bằng nhau, tức là $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Phân phối mẫu

- Với giả thiết $(\mathcal{H})$ thì

$$V(\overline{X}_1 - \overline{X}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

- Một ước lượng gộp của σ^2 được định nghĩa là

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}. \tag{23}$$

So sánh hai kỳ vọng, hai phương sai chưa biết

Phân phối mẫu (tiếp theo)

- Thống kê

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  có phân phối chuẩn tắc $\mathcal{N}(0;1)$.

- Do σ chưa biết, nên ta thay σ bởi S_p và với giả thiết $(\mathcal{H})$ thì thống kê

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{24}$$

  có phân phối Student với $n_1 + n_2 - 2$ bậc tự do.

So sánh hai kỳ vọng, hai phương sai chưa biết

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định:

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}. \tag{25}$$

   Nếu giả thuyết $H_0 : \mu_1 - \mu_2 = \Delta_0$ là đúng và nếu giả thiết $(\mathcal{H})$ được thỏa mãn thì $T_0 \sim t(n_1 + n_2 - 2)$.

3. Miền bác bỏ giả thuyết H_0 được xác định phụ thuộc vào đối thuyết H_1.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 \neq \Delta_0$	$(-\infty; -t_{\alpha/2;n_1+n_2-2}) \cup (t_{\alpha/2;n_1+n_2-2}; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$(t_{\alpha;n_1+n_2-2}; +\infty)$
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 < \Delta_0$	$(-\infty; -t_{\alpha;n_1+n_2-2})$

trong đó, $t_{\alpha/2;n_1+n_2-2}$ và $t_{\alpha;n_1+n_2-2}$ là các giá trị tới hạn của phân phối Student mức $\alpha/2$ và α tương ứng với $n_1 + n_2 - 2$ bậc tự do.

So sánh hai kỳ vọng, hai phương sai chưa biết

Các bước tiến hành (tiếp theo)

4. Từ hai mẫu cụ thể $W_{x_1} = (x_{11}, x_{12}, \dots, x_{1n_1})$ và $W_{x_2} = (x_{21}, x_{22}, \dots, x_{2n_2})$, tính giá trị quan sát của tiêu chuẩn kiểm định

$$t_0 = \frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \tag{26}$$

ở đây, s_p là ước lượng cho độ lệch chuẩn chung σ và

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

5. Kiểm tra xem t_0 có thuộc W_α hay không để kết luận:

- Nếu $t_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $t_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

So sánh hai kỳ vọng, hai phương sai chưa biết

Ví dụ 13

Hai chất xúc tác đang được phân tích để xác định xem chúng ảnh hưởng như thế nào đến năng suất trung bình của một quá trình sản xuất sản phẩm trong một nhà máy. Chất xúc tác I hiện đang được sử dụng. Chất xúc tác II được chấp nhận vì nó rẻ hơn, miễn là nó không làm thay đổi năng suất trung bình của quá trình sản xuất sản phẩm. Kết quả thử nghiệm trong nhà máy được cho như sau.

- Năng suất khi dùng chất xúc tác I: 91,50; 94,18; 92,18; 95,39; 91,79; 89,07; 94,72; 89,21
- Năng suất khi dùng chất xúc tác II: 89,19; 90,95; 90,46; 93,21; 97,19; 97,04; 91,07; 92,75

Có sự khác biệt nào giữa năng suất trung bình của quá trình sản xuất sản phẩm khi sử dụng hai loại chất xúc tác này không? Sử dụng mức ý nghĩa $\alpha = 0,05$ và giả sử năng suất của quá trình sản xuất có phân phối chuẩn với các phương sai bằng nhau.

So sánh hai kỳ vọng, hai phương sai chưa biết

Giải. Gọi X_1, X_2 tương ứng là năng suất của quá trình sản xuất sản phẩm khi dùng chất xúc tác I và II. Theo đầu bài, X_1 và X_2 độc lập và $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, $i = 1, 2$. Đây là bài toán so sánh hai kỳ vọng μ_1 và μ_2 của hai biến ngẫu nhiên có phân phối chuẩn trong trường hợp phương sai σ_1^2 và σ_2^2 chưa biết, nhưng được giả thiết là bằng nhau. Ta muốn biết liệu $\mu_1 - \mu_2 = 0$ hay không.

- Cặp giả thuyết cần kiểm định $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$.
- Chọn tiêu chuẩn kiểm định

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}.$$

  Nếu giả thuyết $H_0 : \mu_1 = \mu_2$ là đúng thì $T_0 \sim t(n_1 + n_2 - 2)$.
- Với $\alpha = 5\%$, $t_{\alpha/2;n_1+n_2-2} = t_{0,025;14} = 2,145$ suy ra miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -2,145) \cup (2,145; +\infty).$$

So sánh hai kỳ vọng, hai phương sai chưa biết

- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n_1 = n_2 = 8$, $\overline{x}_1 = 92,255$, $\overline{x}_2 = 92,733$, $s_1 = 2,385$ và $s_2 = 2,983$:

$$s_p^2 = \frac{7 \times (2,39)^2 + 7 \times (2,98)^2}{8+8-2} = 7,293$$

và

$$t_0 = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{92,255 - 92,733}{\sqrt{7,30}\sqrt{\frac{1}{8} + \frac{1}{8}}} = -0,354.$$

- Vì $t_0 = -0,354 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ H_0, nghĩa là chưa có đủ bằng chứng để khẳng định sử dụng chất xúc tác II cho năng suất trung bình khác với việc sử dụng chất xúc tác I.

So sánh cặp

- Nếu hai biến ngẫu nhiên gốc X_1 và X_2 cùng xét về một đặc trưng thống kê và cùng được quan sát trên một tổng thể xác định và được giả thiết là $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, $i = 1, 2$.
- Quan sát đồng thời về cặp hai biến ngẫu nhiên gốc (X_1, X_2) ta có một mẫu ngẫu nhiên kích thước n là $W_{(X_1,X_2)} = \big((X_{11}, X_{21}), (X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})\big)$. Khi đó, bài toán kiểm định giả thuyết hai mẫu so sánh hai kỳ vọng của hai biến ngẫu nhiên gốc X_1 và X_2 cùng được xác định trên một tổng thể có thể được chuyển về bài toán kiểm định giả thuyết một mẫu về kỳ vọng đối với biến ngẫu nhiên gốc

$$Z = X_1 - X_2.$$

- Thật vậy, từ giả thiết suy ra $Z \sim \mathcal{N}(\mu_Z; \sigma_Z^2)$ với

$$\mu_Z = E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2.$$

So sánh cặp

- Do đó, từ cặp giả thuyết hai mẫu so sánh hai kỳ vọng ban đầu

$$\begin{cases} H_0 : \mu_1 - \mu_2 = \Delta_0, \\ H_1 : \mu_1 - \mu_2 \neq \Delta_0 \text{ hoặc } \mu_1 - \mu_2 > \Delta_0 \text{ hoặc } \mu_1 - \mu_2 < \Delta_0 \end{cases}$$

ta chuyển về cặp giả thuyết một mẫu về kỳ vọng của biến ngẫu nhiên Z tương ứng

$$\begin{cases} H_0 : \mu_Z = \Delta_0, \\ H_1 : \mu_Z \neq \Delta_0 \text{ hoặc } \mu_Z > \Delta_0 \text{ hoặc } \mu_Z < \Delta_0 \end{cases}$$

So sánh cặp

Tiêu chuẩn kiểm định và miền bác bỏ

- Tiêu chuẩn kiểm định

$$T_0 = \frac{\overline{Z} - \Delta_0}{S_Z/\sqrt{n}} \sim t(n-1). \tag{27}$$

- Miền bác bỏ giả thuyết H_0:

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu_Z = \Delta_0$	$\mu_Z \neq \Delta_0$	$(-\infty; -t_{\alpha/2;n-1}) \cup (t_{\alpha/2;n-1}; +\infty)$
$\mu_Z = \Delta_0$	$\mu_Z > \Delta_0$	$(t_{\alpha;n-1}; +\infty)$
$\mu_Z = \Delta_0$	$\mu_Z < \Delta_0$	$(-\infty; -t_{\alpha;n-1})$

- Với mẫu cụ thể $W_{x_1,x_2} = \{(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})\}$, từ $z_i = x_{1i} - x_{2i}$, $i = 1, 2, \ldots, n$ ta nhận được mẫu mới $W_z = (z_1, z_2, \ldots, z_n)$. Từ đó, tính được giá trị quan sát của tiêu chuẩn kiểm định.

So sánh cặp

Ví dụ 14

Để khảo sát tác dụng tăng sản lượng hoa màu A của việc bón thêm một loại phân đạm B ta so sánh năng suất hoa màu A trung bình trên một héc-ta khi bón phân B với năng suất hoa màu A trung bình trên một héc-ta khi không bón phân B. Gọi μ_1 là năng suất hoa màu A trung bình trên một héc-ta khi bón phân B, μ_2 là năng suất hoa màu A trung bình trên một héc-ta khi không bón phân B. Khi đó ta có bài toán kiểm định giả thuyết $H_0 : \mu_1 = \mu_2$ với đối thuyết $H_1 : \mu_1 > \mu_2$. Để giải quyết bài toán này, ta sẽ bố trí thí nghiệm theo hai cách.

So sánh cặp

Ví dụ 14 (tiếp theo)

1. Cách 1. Chọn n thửa ruộng thí nghiệm diện tích 1 héc-ta, trồng hoa màu A bón phân B và n mảnh đất thí nghiệm diện tích 1 héc-ta, trồng hoa màu A không bón phân B. Ngoài việc bón phân, tất cả các yếu tố canh tác khác đều như nhau. Cuối vụ ta thu được sản lượng của n thửa ruộng có bón phân là $W_{x_1} = (x_{11}, x_{12}, \ldots, x_{1n})$ và sản lượng của n thửa ruộng không bón phân là $W_{x_2} = (x_{21}, x_{22}, \ldots, x_{2n})$. Ta sử dụng phương pháp so sánh hai kỳ vọng với hai mẫu độc lập. Chú ý rằng nếu $n_1 < 30$ và $n_2 < 30$ thì ta cần thêm điều kiện là giả thiết $(\mathcal{H})$ thỏa mãn.

2. Cách 2. Chọn n thửa ruộng thí nghiệm diện tích 2 héc-ta, chia mỗi thửa làm 2 mảnh, mỗi mảnh diện tích 1 héc-ta. Một mảnh bón phân B và một mảnh không bón phân B. Cuối vụ thu hoạch ta có sản lượng của hai nửa tương ứng là $W_{x_1} = (x_{11}, x_{12}, \ldots, x_{1n})$ và $W_{x_2} = (x_{21}, x_{22}, \ldots, x_{2n})$. Khi đó, ta sẽ sử dụng phương pháp so sánh từng cặp. Cách này có ưu điểm hơn cách 1 ở chỗ: không cần đến giả thiết $(\mathcal{H})$, Ngoài ra, nó sẽ cho kết quả chính xác hơn vì đã loại bỏ được các nhân tố ngoại lai ảnh hưởng tới giá trị trung bình. Điều kiện canh tác trên mảnh không bón phân và có bón phân gần như đồng nhất, chỉ khác nhau ở việc có bón phân B hay không.

So sánh cặp

Ví dụ 15

Để so sánh hai chế độ chăm bón cho một loại cây trồng A, trên 8 mảnh ruộng người ta chia mỗi mảnh thành hai nửa: nửa thứ nhất áp dụng phương pháp chăm bón I, nửa thứ hai theo phương pháp chăm bón II. Sau khi thu hoạch ta được số liệu về năng suất loại cây trồng A như sau:

Mảnh	1	2	3	4	5	6	7	8
Năng suất nửa thứ nhất	5	20	16	22	24	14	18	20
Năng suất nửa thứ hai	15	22	14	25	29	16	20	24

Đánh giá xem hai chế độ chăm bón có giống nhau không với mức ý nghĩa 1%. Biết rằng năng suất loại cây trồng A có phân phối chuẩn.

So sánh cặp

Giải. Gọi X_1, X_2 lần lượt là năng suất loại cây trồng A ở nửa thứ nhất, thứ hai (sử dụng hai phương pháp chăm bón tương ứng). Ta có X_1 và X_2 có phân phối chuẩn $\mathcal{N}(\mu_1; \sigma_1^2)$, $\mathcal{N}(\mu_2; \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn cho ở dạng cặp. Sử dụng phương pháp so sánh từng cặp. Thiết lập biến ngẫu nhiên $Z = X_1 - X_2$, ký hiệu $E(Z) = \mu_Z$.

- Giả thuyết $H_0 : \mu_Z = \mu_1 - \mu_2 = 0$, đối thuyết $H_1 : \mu_Z = \mu_1 - \mu_2 \neq 0$.
- Chọn tiêu chuẩn kiểm định

$$T_0 = \frac{\overline{Z} - 0}{S_Z/\sqrt{n}}.$$

  Nếu giả thuyết $H_0 : \mu_Z = 0$ là đúng thì $T_0 \sim t(7)$.
- Với $\alpha = 0,01$, $t_{0,005;7} = 3,499$, suy ra miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -3,499) \cup (3,499; +\infty).$$

So sánh cặp

- Lập hiệu $z_i = x_i - y_i$, $i = 1, 2, \ldots, 8$.

Mảnh	1	2	3	4	5	6	7	8
z_i	-10	-2	2	-3	-5	-2	-2	-4

Tính được $\overline{z} = -3,25$, $s_z = 3,412$, suy ra giá trị quan sát của tiêu chuẩn kiểm định

$$t_0 = \frac{\overline{z} - 0}{s_z/\sqrt{n}} = \frac{-3,25}{3,412}\sqrt{8} \simeq -2,694.$$

- Vì $t_0 = -2,694 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0, hay có thể xem hai phương pháp chăm bón cho kết quả như nhau với mức ý nghĩa 1%.

5.3. SO SÁNH

Bài toán

Bài toán 5

Cho hai tổng thể với hai biến ngẫu nhiên gốc $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$, $i = 1, 2$, trong đó $V(X_i) = \sigma_i^2$, $i = 1, 2$ chưa biết. Hãy so sánh hai phương sai này.

✍ **Cặp giả thuyết**

Tùy theo yêu cầu của bài toán kiểm định, các cặp giả thuyết cần kiểm định là

1. $H_0 : \sigma_1^2 = \sigma_2^2;\ H_1 : \sigma_1^2 \neq \sigma_2^2$
2. $H_0 : \sigma_1^2 = \sigma_2^2;\ H_1 : \sigma_1^2 > \sigma_2^2$
3. $H_0 : \sigma_1^2 = \sigma_2^2;\ H_1 : \sigma_1^2 < \sigma_2^2$

Phân phối mẫu

Phân phối Fisher

- Từ hai biến ngẫu nhiên gốc, ta lập hai mẫu ngẫu nhiên độc lập $W_{X_1} = (X_{11}, X_{12}, \dots, X_{1n_1})$ kích thước n_1 và $W_{X_2} = (X_{21}, X_{22}, \dots, X_{2n_2})$ kích thước n_2. Khi đó, thống kê

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

  có phân phối Fisher với $n_1 - 1$ và $n_2 - 1$ bậc tự do.
- Nếu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ là đúng thì $F_0 = S_1^2/S_2^2$ có phân phối Fisher với $n_1 - 1$ và $n_2 - 1$ bậc tự do.

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định

$$F_0 = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1; n_2 - 1). \tag{28}$$

3. Miền bác bỏ giả thuyết H_0 phụ thuộc vào đối thuyết H_1.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$(0; F_{1-\alpha/2;n_1-1;n_2-1}) \cup (F_{\alpha/2;n_1-1;n_2-1}; +\infty)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$(F_{\alpha;n_1-1;n_2-1}; +\infty)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$(0; F_{1-\alpha;n_1-1;n_2-1})$

trong đó, $F_{\alpha/2;n_1-1;n_2-1}$, $F_{1-\alpha/2;n_1-1;n_2-1}$, $F_{\alpha;n_1-1;n_2-1}$ và $F_{1-\alpha;n_1-1;n_2-1}$ là các giá trị tới hạn của phân phối Fisher mức $\alpha/2$, $1-\alpha/2$, α và $1-\alpha$ tương ứng vớ $n_1 - 1$ và $n_2 - 1$ bậc tự do.

Các bước tiến hành

4. Từ các mẫu cụ thể $W_{x_1} = (x_{11}, x_{12}, \dots, x_{1n_1})$ và $W_{x_2} = (x_{21}, x_{22}, \dots, x_{2n_2})$ ta tính giá trị quan sát của tiêu chuẩn kiểm định
$$f_0 = \frac{s_1^2}{s_2^2}.$$

5. Kiểm tra xem f_0 có thuộc W_α hay không để kết luận.

- Nếu $f_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $f_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Các bước tiến hành

Ví dụ 16

Thử nghiệm hai mẫu đạn của hai công ty khác nhau, người ta đo tốc độ xuất phát của đạn X_1, X_2 khi súng phát hỏa. Số liệu thử nghiệm của mẫu thứ nhất là $n_1 = 10$, $\overline{x}_1 = 1210$, $s_1^2 = 2500$ và của mẫu thứ hai là $n_2 = 10$, $\overline{x}_2 = 1175$, $s_2^2 = 3600$. Với mức ý nghĩa 5% có thể kết luận gì về chất lượng của hai mẫu đạn trên? Biết rằng tốc độ xuất phát của đạn của hai công ty là các biến ngẫu nhiên có phân phối chuẩn.

Các bước tiến hành

Giải. Đây là bài toán so sánh hai giá trị trung bình của hai tổng thể phân phối chuẩn trường hợp hai phương sai $V(X_1) = \sigma_1^2$ và $V(X_2) = \sigma_2^2$ chưa biết, mẫu cỡ $n_1 = 10$ và $n_2 = 10$ nhỏ hơn 30.

- Trước hết ta kiểm tra xem X_1 và X_2 có cùng phương sai hay không. Tức là kiểm tra cặp giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$, đối thuyết $H_1 : \sigma_1^2 < \sigma_2^2$.

  Tra bảng giá trị tới hạn phân phối Fisher, ta được $F_{0,025;9;9} = 4,03$, suy ra miền bác bỏ giả thuyết H_0 là $W_\alpha = (4,03; +\infty)$.

  Vì $f_0 = \dfrac{s_2^2}{s_1^2} = \dfrac{3600}{2500} \simeq 1,41 \notin W_\alpha$ nên giả thuyết về hai phương sai của X_1 và X_2 bằng nhau chấp nhận được.

Các bước tiến hành

- Bây giờ ta kiểm định giả thuyết $H_0 : \mu_1 = \mu_2$ với thuyết đối $H_1 : \mu_1 \neq \mu_2$ với giả thiết $(\mathcal{H})$ được thỏa mãn. Sử dụng tiêu chuẩn kiểm định (25) với giả thuyết H_0 đúng.

  Với $\alpha = 0,05$, $t_{0,025;18} = 2,101$ suy ra miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -2,101) \cup (2,101; +\infty).$$

  Tính giá trị quan sát của tiêu chuẩn kiểm định

$$t_0 = \frac{1210 - 1175}{\sqrt{\frac{9 \times 2500 + 9 \times 3600}{10 + 10 - 2}\left(\frac{1}{10} + \frac{1}{10}\right)}} \simeq 1,42.$$

  Vì $t_0 \notin W_\alpha$ nên chấp nhận giả thuyết H_0, nghĩa là chất lượng của hai mẫu đạn là giống nhau với mức ý nghĩa 5%.

5.3. SO SÁNH

Bài toán

Bài toán 6

Xét hai tổng thể $\mathcal{X}$ và $\mathcal{Y}$ và một dấu hiệu A nào đó, trong đó, mỗi cá thể của tổng thể đó có thể có hay không có dấu hiệu này. Gọi p_1 và p_2 tương ứng là tỷ lệ các cá thể có dấu hiệu A của tổng thể thứ nhất và tổng thể thứ hai. Hai tỷ lệ này chưa biết. Bài toán đặt ra là hãy so sánh p_1 với p_2.

✍ **Cặp giả thuyết**

Các cặp giả thuyết cần kiểm định cho bài toán này như sau.

1. $H_0 : p_1 = p_2;\ H_1 : p_1 \neq p_2$
2. $H_0 : p_1 = p_2;\ H_1 : p_1 > p_2$
3. $H_0 : p_1 = p_2; H_1 : p_1 < p_2$

Phân phối mẫu

- Từ hai tổng thể, lập hai mẫu ngẫu nhiên độc lập kích thước n_1 và n_2 và gọi X_1, X_2 lần lượt là số cá thể có dấu hiệu A trong mẫu thứ nhất và thứ hai.
- Giả sử xấp xỉ chuẩn cho phân phối nhị thức được áp dụng cho mỗi tổng thể. Khi đó, các ước lượng điểm $\widehat{P}_1 = X_1/n_1$ và $\widehat{P}_2 = X_2/n_2$ có phân phối xấp xỉ phân phối chuẩn $\mathcal{N}(p_1; p_1(1-p_1)/n_1)$ và $\mathcal{N}(p_2; p_2(1-p_2)/n_2)$. Suy ra $\widehat{P}_1 - \widehat{P}_2$ xấp xỉ phân phối chuẩn với kỳ vọng

$$\mu_{\widehat{P}_1 - \widehat{P}_2} = p_1 - p_2$$

  và phương sai

$$\sigma^2_{\widehat{P}_1 - \widehat{P}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

- Khi đó, thống kê

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

  xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

Phân phối mẫu

- Nếu giả thuyết $H_0 : p_1 = p_2$ là đúng và giả sử $p_1 = p_2 = p$ thì

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - 0}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

  xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0;1)$.

- Một ước lượng điểm của tham số tổng hợp p là

$$\widehat{P} = \frac{X_1 + X_2}{n_1 + n_2}.$$

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định:

$$Z_0 = \frac{(\widehat{P}_1 - \widehat{P}_2)}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{29}$$

   Nếu giả thuyết $H_0 : p_1 = p_2$ đúng và nếu $n_1\widehat{p}_1 \geq 5$, $n_1(1-\widehat{p}_1) \geq 5$, $n_2\widehat{p}_2 \geq 5$ và $n_2(1-\widehat{p}_2) \geq 5$ thì $Z_0 \sim^{\text{xx}} \mathcal{N}(0;1)$.

3. Miền bác bỏ giả thuyết H_0 được xác định phụ thuộc vào đối thuyết H_1

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$p_1 = p_2$	$p_1 \neq p_2$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$p_1 = p_2$	$p_1 > p_2$	$(z_\alpha; +\infty)$
$p_1 = p_2$	$p_1 < p_2$	$(-\infty; -z_\alpha)$

Các bước tiến hành

4. Từ mẫu cụ thể, ta tính giá trị quan sát của tiêu chuẩn kiểm định:

$$z_0 = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{30}$$

với $\widehat{p}_1 = \dfrac{x_1}{n_1}$, $\widehat{p}_2 = \dfrac{x_2}{n_2}$, $\widehat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$.

5. Kiểm tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Các bước tiến hành

Ví dụ 17

Từ kho đồ hộp thứ nhất lấy ngẫu nhiên 1000 hộp kiểm tra thấy có 20 hộp bị hỏng. Từ kho đồ hộp thứ hai lấy ngẫu nhiên 900 hộp kiểm tra thấy 30 hộp bị hỏng. Hỏi chất lượng bảo quản của hai kho có thực sự giống nhau hay không với mức ý nghĩa 5%.

Giải. Gọi p_1, p_2 lần lượt là tỷ lệ hộp hỏng ở kho đồ hộp thứ nhất và thứ hai. Đây là bài toán so sánh hai tỷ lệ. Kiểm tra điều kiện $n_1\widehat{p}_1 = 20 > 5$, $n_1(1-\widehat{p}_1) = 980 > 5$, $n_2\widehat{p}_2 = 30 > 5$ và $n_2(1-\widehat{p}_2) = 870 > 5$.

- Cặp giả thuyết cần kiểm định $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.
- Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0;1)$.

Các bước tiến hành

- Với $\alpha = 0,05$, $z_{\alpha/2} = z_{0,025} = 1,96$. Miền bác bỏ giả thuyết H_0 là $W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty)$.
- Tính giá trị quan sát của tiêu chuẩn kiểm định với

$$\widehat{p}_1 = \frac{x_1}{n_1} = \frac{20}{1000} = \frac{1}{50}; \quad \widehat{p}_2 = \frac{x_2}{n_2} = \frac{30}{900} = \frac{1}{30};$$
$$\widehat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{20 + 30}{1000 + 900} = \frac{1}{38}.$$

Suy ra

$$z_0 = \frac{1/50 - 1/30}{\sqrt{1/38 \times (1 - 1/38)\left(\frac{1}{1000} + \frac{1}{900}\right)}} \simeq -1,8129.$$

- Vì $z_0 = -1,8129 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0, nghĩa là có thể xem chất lượng bảo quản của hai kho hàng là như nhau với mức ý nghĩa 5%.

Các bước tiến hành

Ví dụ 18

Một bệnh viện điều trị loại bệnh A theo hai phương pháp. Sau một thời gian thấy kết quả như sau:

- Trong 102 bệnh nhân điều trị phương pháp I có 82 bệnh nhân khỏi bệnh.
- Trong 98 bệnh nhân điều trị phương pháp II có 69 bệnh nhân khỏi bệnh.

Hỏi có phải phương pháp I điều trị tốt hơn phương pháp II hai hay không với mức ý nghĩa 5%.

Giải. Gọi p_1, p_2 lần lượt là tỷ lệ bệnh nhân khỏi bệnh khi điều trị bệnh A bằng phương pháp I và II. Đây là bài toán so sánh hai tỷ lệ. Kiểm tra điều kiện $n_1\widehat{p}_1 = 82 > 5$, $n_1(1-\widehat{p}_1) = 20 > 5$, $n_2\widehat{p}_2 = 69 > 5$ và $n_2(1-\widehat{p}_2) = 29 > 5$.

- Giả thuyết $H_0 : p_1 = p_2$, đối thuyết $H_1 : p_1 > p_2$.
- Tiêu chuẩn kiểm định

$$Z_0 = \frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}.$$

  Nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0;1)$.

Các bước tiến hành

- Với $\alpha = 0,05$, $z_\alpha = z_{0,05} = 1,645$. Miền bác bỏ giả thuyết H_0 là
$$W_\alpha = (1,645; +\infty).$$

- Tính giá trị quan sát của tiêu chuẩn kiểm định với
$$\widehat{p}_1 = \frac{x_1}{n_1} = \frac{82}{102} = \frac{41}{51}; \quad \widehat{p}_2 = \frac{x_2}{n_2} = \frac{69}{98};$$
$$\widehat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{82 + 69}{102 + 98} = \frac{151}{200}.$$

  Suy ra
$$z_0 = \frac{41/51 - 69/98}{\sqrt{151/200 \times (1 - 151/200)\left(\frac{1}{102} + \frac{1}{98}\right)}} \simeq 1,6411.$$

- Vì $z_0 = 1,6411 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0, nghĩa là chưa thể xem phương pháp I điều trị tốt hơn phương pháp II với mức ý nghĩa 5%.

5.3. SO SÁNH

Bài tập

Bài 8

Người ta muốn so sánh tuổi thọ của một loại pin (trong điều kiện sử dụng liên tục) được sản xuất bởi hai công nghệ khác nhau. Biết rằng độ lệch chuẩn tuổi thọ của pin được sản xuất từ công nghệ thứ nhất và công nghệ thứ hai tương ứng là 120 giờ và 125 giờ. Ứng với mỗi công nghệ, người ta thử nghiệm 50 pin, thu được tuổi thọ trung bình tương ứng là 264 giờ và 245 giờ. Với mức ý nghĩa 5%, tuổi thọ trung bình của các pin được sản xuất từ hai công nghệ trên có khác nhau không?

Bài 9

Trong một cuộc thi về khả năng tính toán ở một trường học với số lượng lớn học sinh, một mẫu gồm 50 học sinh nữ có điểm số trung bình là 64 và độ lệch chuẩn là 4; một mẫu gồm 50 học sinh nam có điểm số trung bình là 60 và độ lệch chuẩn là 2. Có thể cho rằng khả năng tính toán của học sinh nữ trong trường này tốt hơn học sinh nam không, kết luận với mức ý nghĩa 5%.

Bài tập

Bài 10

Sản lượng sữa của mỗi giống bò sữa là biến ngẫu nhiên có phân phối chuẩn. Theo dõi sản lượng sữa (kg/con/ngày) của hai giống bò sữa được kết quả như sau:

Sản lượng sữa A (X kg)	[8-9)	[9-10)	[10-11)	[11-12)	[12-13)	[13-14)	[14-15)
Số con	1	4	7	9	6	3	2

Sản lượng sữa B (Y kg)	[8-9)	[9-10)	[10-11)	[11-12)	[12-13)	[13-14)
Số con	3	3	5	4	3	2

Giả thiết rằng hai biến ngẫu nhiên X và Y có phương sai là như nhau. Với mức ý nghĩa $\alpha = 0,05$, có thể coi sản lượng sữa trung bình của hai giống bò là như nhau không?

Bài tập

Bài 11

Hai loại trái cây có năng suất trung bình xấp xỉ nhau nhưng mức độ phân tán về năng suất có thể khác nhau. Để kiểm tra điều đó, người ta tiến hành lấy hai mẫu của hai loại trái cây thu được kết quả như sau:
Loại A: $n_1 = 41; s_1^2 = 11,41$.
Loại B: $n_2 = 30; s_1^2 = 6,52$.
Biết rằng năng suất của hai loại trái cây là các biến ngẫu nhiên có phân phối chuẩn, với mức ý nghĩa 5% có thể kết luận mức độ phân tán về năng suất của hai loại trái cây khác nhau hay không?

Bài 12

Có ý kiến cho rằng tỷ lệ phế phẩm do máy thứ I sản xuất ít hơn tỷ lệ này do máy thứ II sản xuất. Người ta kiểm tra 100 sản phẩm do máy I sản xuất thì thấy có 8 phế phẩm, kiểm tra 200 sản phẩm do máy II sản xuất thì thấy có 12 phế phẩm. Với mức ý nghĩa 5%, hãy kết luận về ý kiến trên.