Bài số 9

BÀI TOÁN ƯỚC LƯỢNG TRUNG BÌNH MẪU

Quy luật phân bố xác suất của các thống kê đặc trưng mẫu phản ánh mối liên hệ chặt chẽ giữa các tham số của mẫu với các tham số của dấu hiệu nghiên cứu tương ứng của tổng thể. Lý thuyết Thống kê sử dụng hai phương pháp sau:

● ***Suy diễn thống kê:*** Nếu đã biết quy luật phân bố xác suất cũng như các tham số đặc trưng của tổng thể thì có thể sử dụng các kết quả trên để suy đoán về tính chất của một mẫu ngẫu nhiên rút ra từ tổng thể đó. Chẳng hạn nếu biết dấu hiệu nghiên cứu X có phân bố chuẩn $n(x;\mu,\sigma)$ thì thống kê $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối chuẩn tắc $n(z;0,1)$.

● ***Quy nạp thống kê:*** Sử dụng các phương pháp thống kê dể từ các đặc trưng mẫu suy ra các đặc trưng của tổng thể.

Chính vì vậy, các phương pháp thống kê giải quyết được nhiều bài toán thực tế, có thể giúp cho các nhà nghiên cứu tìm ra quy luật của tổng thể, giúp các nhà hoạch định chính sách dư đoán sự phát triển trong tương lai, đề ra các quyết định chấp nhận hoặc bác bỏ các giả thuyết nào đó.

Nếu dấu hiệu nghiên cứu của tổng thể có thể xem như một biến ngẫu nhiên và giả sử bằng lý thuyết đã xác định được dạng phân bố xác suất của nó thì vấn đề xác định các tham số của đặc trưng của tổng thể sẽ quy về bài toán xác định các tham số đặc trưng của quy luật phân bố xác suất của biến ngẫu nhiên X. Chẳng hạn, nếu đã biết dấu hiệu nghiên cứu của tổng thể có phân bố chuẩn $n(.;\mu,\sigma)$ thì bài toán đặt ra là phải ước lượng các tham số là kỳ vọng μ và phương sai σ^2, hai tham số này cũng chính là trung bình và phương sai của tổng thể.

Giả sử BNN có tham số θ chưa biết. Ước lượng tham số θ là dựa vào mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ ta đưa ra thống kê $\hat{\theta}$ để ước lượng(dự đoán) θ. Ước lượng gồm:

i.*Ước lượng điểm*: chỉ ra $\theta = \theta_0$ nào đó để ước lượng θ.

ii. *Ước lượng khoảng*: chỉ ra một khoảng $(\hat{\theta}_L, \hat{\theta}_U)$ chứa θ sao cho $P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1-\alpha$ cho trước ($1-\alpha$ goi là độ tin cậy của ước lượng).

I.ƯỚC LƯỢNG ĐIỂM

1. Định nghĩa. Một **ước lượng điểm** cho tham số tổng thể θ là một giá trị đơn $\hat{\theta}$ của một thống kê $\hat{\Theta}$.

Chú ý. Thực chất ta đã dùng một giá trị $\hat{\theta}$ để thay thế cho giá trị của tham số θ chưa biết của tổng thể, thông thường giá trị được chọn này là giá trị cụ thể của một thống kê $\hat{\Theta}$ nào đó của mẫu ngẫu nhiên. Ví dụ như giá trị $\overline{x}$ của thống kê $\overline{X}$, được tính toán từ một mẫu cỡ n, là một ước lượng điểm của tham số trung bình tổng thể μ.

Cùng một mẫu ngẫu nhiên ta có thể xây dựng được nhiều thống kê $\hat{\Theta}$ khác nhau để ước lượng cho tham số tổng thể θ. Một ước lượng cũng có thể có sai số khi ước lượng tham số chung. Đối với một mẫu cụ thể, có thể thu được một ước lượng chính xác hơn của μ bằng cách sử dụng trung vị mẫu $\tilde{X}$ là một ước lượng. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau:

2. Ước lượng không chệch.

Các tính chất kỳ vọng nào của một hàm quyết định "***tốt***" ảnh hưởng đến quyết định lựa chọn ước lượng của chúng ta?

Lấy $\hat{\Theta}$ là một ước lượng có giá trị $\hat{\theta}$ là một ước lượng điểm có tham số chung chưa xác định θ. Chắc chắn chúng ta mong muốn phân bố lấy mẫu của $\hat{\Theta}$ có số trung bình bằng tham số được ước lượng. Một ước lượng có tính chất này được xem là **không chệch.**

Định nghĩa. Một thống kê $\hat{\Theta}$ được xem là **ước lượng không chệch** cho tham số θ nếu:

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta \qquad (*)$$

Ví dụ 1 . Biểu diễn S^2 là một ước lượng không chệch có tham số σ^2.

Giải: + Chúng ta viết

$$\sum_{i=1}^{n}(X_i-\overline{X})^2 = \sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2$$
$$= \sum_{i=1}^{n}(X_i-\mu)^2 - 2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + n(\overline{X}-\mu)^2$$
$$= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2$$

+ Nên

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i-\mu)^2 - nE(\overline{X}-\mu)^2\right]$$
$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\sigma^2_{X_i} - n\sigma^2_{\overline{X}}\right)$$

+ Tuy nhiên,

$$\sigma^2_X = \sigma^2 \quad \text{đối với} \quad i=1, 2, \ldots, n \quad \text{và} \quad \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

+ Vì thế,

$$E(S^2) = \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2.$$

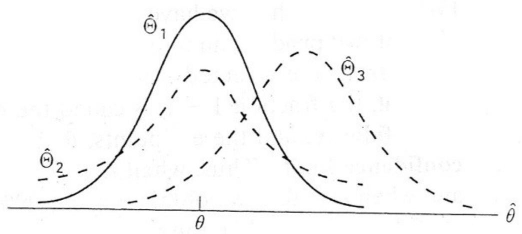
3. Ước lượng hiệu quả.

Điều kiện (*) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của $\hat{\theta}$ bằng giá trị θ. Từng giá trị của $\hat{\theta}$ có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trên là bé nhất.

Nếu $\hat{\Theta}_1$ và $\hat{\Theta}_2$ là hai ước lượng không lệch của cùng tham ẩn chung θ, chúng ta sẽ lựa chọn ước lượng mà phân bố mẫu của nó có phương sai nhỏ hơn. Vì thế, nếu $\sigma^2_{\hat{\Theta}_1} < \sigma^2_{\hat{\Theta}_2}$ chúng ta nói rằng $\hat{\Theta}_1$ là **ước lượng hơn ước lượng $\hat{\Theta}_2$ đối với tham ẩn θ.**

Định nghĩa. Ước lượng không chệch có ***phương sai nhỏ nhất*** so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên được gọi là ***ước lượng hiệu quả***.

Trong Hình vẽ dưới đây, ta biểu diễn các phân bố mẫu của ba ước lượng khác nhau $\hat{\Theta}_1$, $\hat{\Theta}_2$ và $\hat{\Theta}_3$, tất cả đều ước lượng *θ*. Rõ ràng chỉ có $\hat{\Theta}_1$ và $\hat{\Theta}_2$ là không chệch, vì các ước lượng của nó tập trung vào *θ*. Ước lượng $\hat{\Theta}_1$ có phương sai nhỏ hơn $\hat{\Theta}_2$ và sẽ có hiệu quả cao hơn. Vì thế lựa chọn của chúng ta đối với một ước lượng θ, trong số ba bộ được xem xét, sẽ là $\hat{\Theta}_1$.



Chú ý. Đối với các tổng thể chuẩn, chúng ta có thể biểu diễn rằng $\overline{X}$ và $\tilde{X}$ là các ước lượng không chệch của số trung bình tổng thể μ, tuy nhiên phương sai của $\overline{X}$ nhỏ hơn phương sai của $\tilde{X}$. Vì thế, cả hai ước lượng $\bar{x}$ và $\tilde{x}$, trung bình, sẽ bằng số trung bình tổng thể μ, tuy nhiên $\bar{x}$ có thể gần hơn với μ trong một mẫu xác định, vì thế $\overline{X}$ có hiệu quả cao hơn $\tilde{X}$.

Có nhiều tình huống trong đó sẽ thích hợp hơn khi xác định một khoảng trong đó chúng ta kỳ vọng để xác định giá trị của tham số. Khoảng như thế được gọi là **ước lượng khoảng**.

II.ƯỚC LƯỢNG KHOẢNG.

1. Mô tả. Các ước lượng điểm có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch khá nhiều so với giá trị tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó, khi kích thước mẫu bé người ta thường dung ước lượng khoảng.

Một ước lượng khoảng của một tham số tổng thể θ là một khoảng $\hat{\theta}_L < \theta < \hat{\theta}_U$, trong đó $\hat{\theta}_L$ và $\hat{\theta}_U$ phụ thuộc vào giá trị của thống kê $\hat{\Theta}$ đối với một mẫu xác định và phân bố mẫu của $\hat{\Theta}$.

Các mẫu khác nhau sẽ sinh ra các giá trị $\hat{\Theta}$ khác nhau và vì thế ta nhận được các giá trị khác nhau của $\hat{\theta}_L$ và $\hat{\theta}_U$: và đây cũng chính là các giá trị của các biến ngẫu nhiên tương ứng $\hat{\Theta}_L$ và $\hat{\Theta}_U$ .Từ phân bố mẫu của $\hat{\Theta}$, ta có thể xác định được $\hat{\theta}_L$ và $\hat{\theta}_U$ sao cho:

$$P\,(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1-\alpha$$

tức là, với $0<\alpha<1$, chúng ta có xác suất của một lựa chọn mẫu ngẫu nhiên sinh ra một khoảng chứa θ là $1-\alpha$.

+ Khoảng $\hat{\theta}_L < \theta < \hat{\theta}_U$ được tính toán từ mẫu được chọn và được gọi là ***khoảng tin cậy*** $(1-\alpha).100\%$,

+ Đại lượng $1-\alpha$: gọi là ***hệ số tin cậy*** hay ***độ tin cậy***

+ Các điểm cuối $\hat{\theta}_L$ và $\hat{\theta}_U$: tương ứng là các ***giới hạn tin cậy dưới*** và ***giới hạn tin cậytrên*** .

+ $\left|\hat{\theta}_U - \hat{\theta}_L\right|$ gọi là độ dài khoảng tin cậy.

Do đó: khi $\alpha = 0{,}05$ chúng ta có khoảng tin cậy 95% .

khi $\alpha = 0{,}01$ chúng ta thu được khoảng tin cậy rộng hơn bằng 99%.

2. Ước lượng trung bình.

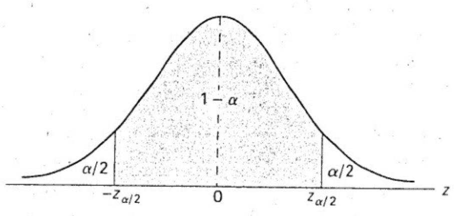
Giả sử trung bình của tổng thể $\mu = E(X)$ chưa biết. Ta tìm khoảng (μ_1, μ_2) chứa μ sao cho: $P(\mu_1 < \mu < \mu_2) = 1-\alpha$ với $1-\alpha$ là độ tin cậy cho trước.

Trường hợp 1. ***Khoảng tin cậy của*** μ***; khi biết*** σ

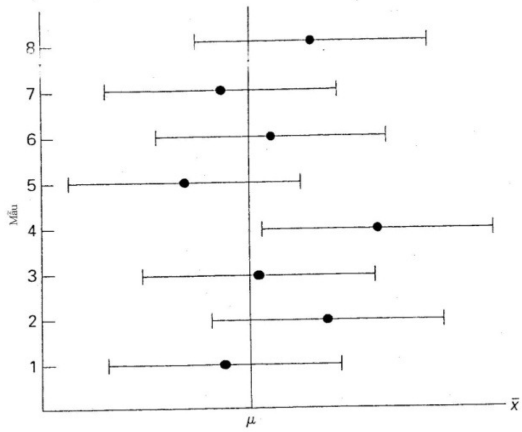
Nếu $\bar{x}$ là số trung bình của một mẫu ngẫu nhiên kích thước n trong một tổng thể có ***phương sai đã biết*** σ^2, một khoảng tin cậy $(1-\alpha)\%$ đối với μ được xác định bằng

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

trong đó $z_{\alpha/2}$ là giá trị tạo nên một diện tích $\alpha/2$ sang bên phía phải của nó, tức $P(Z > z_{\alpha/2}) = \dfrac{\alpha}{2}$.



Các mẫu khác nhau sẽ sinh ra các giá trị khác nhau của $\bar{x}$ và vì thế sinh ra các ước lượng khoảng tin cậy khác nhau của tham số μ như biểu diễn trong Hình 2. Các điểm hình tròn ở tâm mỗi khoảng biểu diễn vị trí của ước lượng điểm $\bar{x}$ cho mỗi mẫu ngẫu nhiên.



Hình 2. Các ước lượng khoảng của μ cho các mẫu khác nhau

Ví dụ 2. Hàm lượng kẽm trung bình thu hồi được từ một mẫu các giá trị đo kẽm tại 36 điểm đo khác nhau được xác định là 2,6g/mili lít. Xác định các khoảng tin cậy 95% và 99% cho mật độ kẽm trung bình ở sông. Giả thiết độ lệch tiêu chuẩn tổng thể là 0,3.

Giải. + Ước lượng điểm của μ là $\bar{x}$=2,6.

+ Giá trị z sinh ra một diện tích 0,025 sang bên phải và vì thế sinh ra một diện tích 0,975 sang bên trái, là $z_{0,025} = 1,96$ (Bảng A.3).

+ Vì thế, khoảng tin cậy 95% là:

$$2,6-(1,96)(\frac{0,3}{\sqrt{36}})<\mu<26+(1,96)(\frac{0,3}{\sqrt{36}}),$$

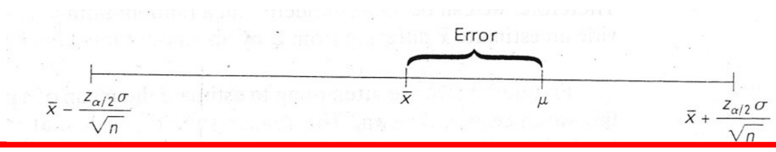
hay là $2,50<\mu<2,70$.

+ Để xác định khoảng tin cậy 95%, chúng ta tìm giá trị z sinh ra một diện tích 0,005 sang bên phải và 0,995 sang bên trái. Vì vậy, sử dụng Bảng A.3 ta được: $z_{0,005} = 2,575$, và khoảng tin cậy 99% là:

$$2,6-(2,575)(\frac{0,3}{\sqrt{36}})<\mu<2,6+(2,575)(\frac{0,3}{\sqrt{36}})$$

hay là: $2,47<\mu<2,73$.

Nhận xét. Nếu μ *l*à giá trị tâm của khoảng , thì khi đó $\bar{x}$ ước lượng μ không bị lỗi. Tuy nhiên, hầu hết thì $\bar{x}$ sẽ không chính xác bằng μ và ước lượng điểm có lỗi. Cỡ sai số sẽ là giá trị tuyệt đối có chênh lệch giữa μ và $\bar{x}$ và chúng ta có thể đạt đến độ tin cậy (1-α)% rằng độ chênh lệch này sẽ không quá $z_{\alpha/2}\sigma/\sqrt{n}$.



Định lý 1. Nếu $\bar{x}$ được sử dụng để ước lượng μ, khi đó với độ tin cậy $(1-\alpha)$ ta có sai số sẽ không vượt quá $z_{\alpha/2}\sigma / \sqrt{n}$.

Trong Ví dụ 2., chúng ta 95% tin cậy rằng số trung bình mẫu $\bar{x}$=2,6 khác số trung bình chân thực μ theo một lượng nhỏ hơn 0,1 và 99% tin cậy rằng độ chênh lệch nhỏ hơn 0,13.
Thông thường, chúng ta đều muốn biết mẫu cần lớn như thế nào để đảm bảo sai số khi ước lượng μ sẽ nhỏ hơn một lượng e cụ thể. Theo Định lý 1, ta phải chọn n sao cho $z_{\alpha/2}\sigma / \sqrt{n} = e$. Giải đẳng thức này thu được công thức sau đây của n.

Định lý 2. Nếu $\bar{x}$ được sử dụng là một ước lượng của μ, khi đó với độ tin cậy $(1-\alpha)$ ta nói rằng sai số sẽ không vượt quá một lượng cụ thể *e* khi kích thước là:

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

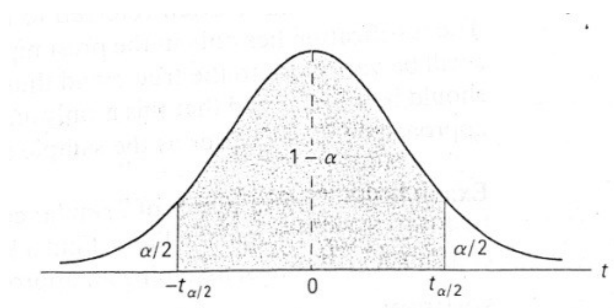
theo quy tắc làm tròn đến toàn bộ số tiếp theo.

Theo nguyên tắc này, chúng ta có thể chắc chắn rằng độ tin cậy không bao giờ được thấp dưới (1-α)%.

Ví dụ 3. Trong Ví dụ 2, một mẫu cần lớn bao nhiêu nếu chúng ta muốn tin cậy 95% rằng ước lượng μ giảm nhỏ hơn 0,05?

Giả. + Độ lệch chuẩn tổng thể là $\sigma = 0,3$. Khi đó, theo Định lý 2: $n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138,3$

+ Vì thế, chúng ta có thể tin cậy 95% rằng một mẫu ngẫu nhiên có cỡ 139 sẽ cho ước lượng $\bar{x}$ khác μ một lượng nhỏ hơn 0,05.



Trường hợp 2: ***Khoảng tin cậy cho μ khi chưa biết σ***

Nếu $\bar{x}$ và s là số trung bình và độ lệch chuẩn của mẫu ngẫu nhiên được rút ra từ biến ngẫu nhiên của chuẩn có phương sai σ^2 **chưa xác định**, khoảng tin cậy $\left(1-\alpha\right)$ cho μ là:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

trong đó $t_{\alpha/2}$ là giá trị t với $v = n-1$ bậc tự do, sinh ra một diện tích bằng $\alpha/2$ bên phía phải của nó, tức là tức $P(T > t_{\alpha/2}) = \dfrac{\alpha}{2}$.

Chú ý. + Đối với trường hợp σ đã biết, chúng ta sử dụng định lý giới hạn trung tâm.
+ Đối với σ chưa biết, chúng ta sử dụng phân phối lấy mẫu của biến ngẫu nhiên *T*.
+ Tìm $t_{\alpha/2}$ thông qua Bảng A4.

Ví dụ 4. Các hàm lượng của 7 container axit sulfuric là 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 và 9.6 lít. Tìm khoảng tin cậy 95% cho giá trị trung bình của tất cả các container đó, giả sử có phân phối chuẩn ước lượng.

Giải. + Trung bình mẫu và độ lệch chuẩn là:

$$\bar{x} = 10.0 \text{ và } s = 0,283$$

+ Sử dụng Bảng A.4, chúng ta xác định được $t_{0.025} = 2,447$ đối với $v = 6$ bậc tự do.

+ Vì thế khoảng tin cậy 95% cho μ là:

$$10.0 - (2.477)(\frac{0.283}{\sqrt{7}}) < \mu < 10.0 + (2.447)(\frac{0.283}{\sqrt{7}})$$

tức là: $9,74 < \mu < 10,26$.

3. Ước lượng hiệu hai kỳ vọng.

Nếu chúng ta có hai tổng thể có các giá trị trung bình μ_1 và μ_2, các phương sai σ_1^2 và σ_2^2, ước lượng điểm về hiệu giữa μ_1 và μ_2 được sinh ra bởi thống kê $\overline{X}_1 - \overline{X}_2$.

Mục tiêu ta cần thiết lập được khoảng tin cậy $(1-\alpha)\%$ đối với $\mu_1 - \mu_2$.

Trường hợp 1: ***Khoảng tin cậy cho $\mu_1 - \mu_2$ khi biết σ_1^2 và σ_2^2***

Nếu $\bar{x}_1$ và $\bar{x}_2$ là các giá trị trung bình của các **mẫu ngẫu nhiên độc lập** có kích thước n_1 và n_2 từ các tổng thể có các phương sai đã biết σ_1^2 và σ_2^2, khoảng tin cậy $\left(1-\alpha\right)$ đối với $\mu_1 - \mu_2$ là:

$$(\bar{x}_2 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 + \mu_2 < (\bar{x}_1 + \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

trong đó $z_{\alpha/2}$ được xác định bởi $P(Z > z_{\alpha/2}) = \dfrac{\alpha}{2}$.

Ví dụ 5. Tiến hành một thí nghiệm với hai loại động cơ A và B để so sánh số dặm đi dược trên mỗi gallon xăng. Trong 50 thí nghiệm đã được tiến hành có sử dụng loại động cơ A và 75 thí nghiệm được tiến hành cho động cơ B. Xăng sử dụng và các điều kiện khác không đổi. Lượng tiêu thụ trung bình đối với động cơ A là 36 dặm mỗi gallon và đối với loại máy B là 42 dặm mỗi gallon. Xác định một khoảng tin cậy 96% trên $\mu_B - \mu_A$, trong đó μ_B và μ_A là lượng tiêu thụ chuẩn tổng thể đối với máy B và A. Giả thiết rằng độ lệch chuẩn tổng thể là 6 và 8 cho lần lượt máy A và B.

Giải: + Ước lượng điểm của $\mu_B - \mu_A$ là $\overline{x}_B - \overline{x}_A = 42 - 36 = 6$.

+ Xác định $z_{0,02} = 2,05$ từ Bảng A.3. Vì thế, khoảng tin cậy 96% là

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_1 + \mu_2 < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}}$$

tức là: $3,43 < \mu_B - \mu_A < 8,57$.

Trường hợp 2: Khoảng tin cậy cho $\mu_1 - \mu_2$; $\sigma_1^2 = \sigma_2^2$ chưa biết $\sigma_1; \sigma_2$.

Nếu $\overline{x}_1$ và $\overline{x}_2$ là các giá trị trung bình của các **mẫu ngẫu nhiên độc lập** kích thước n_1 và n_2, từ các tổng thể chuẩn ước lượng có các phương sai chưa biết nhưng cân bằng, một khoảng tin cậy (1-α)% cho $\mu_1 - \mu_2$ được xác định bằng

$$(\overline{x}_1 - \overline{x}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

trong đó $t_{\alpha/2}$ là giá trị t với $v = n_1 + n_2 - 2$ bậc tự do, sinh ra một diện tích $\alpha / 2$ sang bên phải, và

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

Ví dụ 6. Hai trạm lấy mẫu độc lập được lựa chọn cho việc nghiên cứu, một được đặt tại hạ lưu tính từ điểm xả của mỏ axit và trạm còn lại được đặt tại thượng lưu. Đối với 12 mẫu được thu thập hàng tháng tại trạm hạ lưu, danh mục đa dạng loài có giá trị trung bình $\overline{x}_1 = 3.11$ và độ lệch chuẩn $s_1 = 0.771$, trong khi đó 10 mẫu thu thập hàng tháng tại trạm đầu nguồn có giá trị danh mục trung bình $\overline{x}_2 = 2.04$ và độ lệch chuẩn $s_2 = 0.448$. Tìm một khoảng tin cậy 90% cho độ lệch giữa các kỳ vọng tổng thể và cho hai trạm này, giả thiết tổng thể được phân bố chuẩn có các phương sai bằng nhau.

Giải. + Gọi μ_1 và μ_2 biểu diễn các kỳ vọng tổng thể cho danh mục đa dạng loài ở trạm đầu nguồn và hạ lưu. Chúng ta muốn xác định một khoảng tin cậy 90% cho μ_1-μ_2.

+ Ước lượng điểm μ_1-μ_2 của chúng ta là

$$\overline{x}_1 - \overline{x}_2 = 3,11 - 2,04 = 1,07$$

+ Ước lượng chung của s_p^2 của phương sai σ^2 là

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(11)(0.771^2)+(9)(0.448^2)}{12+10-2}$$

Lấy căn bình phương, chúng ta thu được s_p =0.646. Sử dụng α=0.1, chúng ta xác định được trong Bảng A.4 rằng $t_{0.05} = 1.725$ đối với $v = n_1 + n_2 - 2 = 20$ bậc tự do. Vì thế, khoảng tin cậy 90% của μ_1-μ_2 là:

$$1.07-(1.725)(0.646)\sqrt{\frac{1}{12}+\frac{1}{10}} < \mu_1\text{-}\mu_2 < 1.07+(1.725)(0.646)\sqrt{\frac{1}{12}+\frac{1}{10}}$$

được rút gọn thành 0,593<μ_1-μ_2<1,547.

Trường hợp 3: **Khoảng tin cậy μ1-μ2 với $\sigma_1 \neq \sigma_2$ và chưa biết** σ_1, σ_2

Nếu $\bar{x}_1$ và s_1^2, $\bar{x}_2$ và s_2^2 là các số trung bình và phương sai của các mẫu độc lập kích thước nhỏ n_1 và n_2 rút ra từ các phân phối xấp xỉ chuẩn với các phương sai chưa biết và không bằng nhau, khoảng tin cậy xấp xỉ $(1-\alpha)$ ước lượng cho μ_1-μ_2 là

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1\text{-}\mu_2 < (\bar{x}_1-\bar{x}_2)+t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

trong đó $t_{\alpha/2}$ là giá trị *t* với

$$v = \frac{(s_1^2/n + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1\text{-}2)]+[((s_2^2/n_2)^2/(n_2\text{-}2)]}$$

bậc tự do, sinh ra một diện tích α/2 bên phía phải.

Ví dụ 7. Khoa động vật học thuộc Học viện bách khoa Virginia và Đại học tổng hợp bang đã tiến hành một nghiên cứu để ước lượng độ chênh lệch về lượng của hóa chất orthophospho được đo tại hai trạm khác nhau trên sông James. Orthophospho được đo bằng mili gam trên mỗi lít. 15 mẫu được thu thập tại trạm 1 và 12 mẫu được thu thập tại trạm 2. 15 mẫu thu tại trạm có hàm lượng orthophospho trung bình là 3.84 mili gam mỗi lít và độ lệch chuẩn 3.07 mili gam mỗi lít, trong khi đó 12 mẫu của trạm 2 có hàm lượng trung bình 1.49 mili gam trên mỗi lít và độ lệch chuẩn 0.80 mili gam mỗi lít. Tìm một khoảng tin cậy 95% cho độ lệch trong các hàm lượng orthophospho trung bình chân thực tại hai trạm này, giả thiết rằng các quan sát lấy từ các tổng thể chuẩn có phương sai khác nhau.

Giải. + Đối với trạm 1, chúng ta có $\bar{x}_1 = 3,84$, s_1=3,07, và $n_1 = 15$.

+ Đối với trạm 2, $\bar{x}_2 = 1,49$, s_2 = 0,80, và $n_2 = 12$.

+ Chúng ta cần xác định một khoảng tin cậy 95% cho $\mu_1 - \mu_2$.

+ Vì các phương sai tổng thể được giả thiết không bằng nhau, cho nên chúng ta chỉ có thể xác định được một khoảng tin cậy ước lượng 95% dựa trên phân phối *t* có *v* các bậc tự do, trong đó

$$\nu = \frac{\left(3,07^2/15+0,80^2/12\right)^2}{\left[\left(3,07^2/15\right)^2/14\right]+\left[\left(0,80^2/12\right)^2/11\right]} = 16,3 \simeq 16$$

+ Ước lượng điểm cho $\mu_1 - \mu_2$ là:

$$\bar{x}_1 - \bar{x}_2 = 3,84 - 1,49 = 2,35$$

+ Dùng $\alpha = 0,05$, chúng ta xác định được trong Bảng A.4 : $t_{0,025} = 2,120$ với $\nu = 16$ bậc tự do.

+ Vì vậy, khoảng tin cậy 95% cho $\mu_1 - \mu_2$ là:

$$2,35 - 2,120\sqrt{\frac{3,07^2}{15} + \frac{0,80^2}{12}} < \mu_1 - \mu_2 < 2,35 + 2,120\sqrt{\frac{3,07^2}{15} + \frac{0,80^2}{12}}$$

Tức là 0,60 < $\mu_1 - \mu_2$ < 4,10. Vì thế, chúng ta tin cậy 95% rằng khoảng từ 0.60 đến 4.10 miligam trên mỗi lít có độ chênh lệch các hàm lượng orthothosphorus trung bình chân thực đối với hai vị trí này.

Về nhà:

Tự đọc: Mục 9.5; 9.6; 9.8

Bài tập: Tr. **278; 290**

Đọc trước các Mục từ 9.9 đến 9.12 chuẩn bị cho **Bài số 10 :**

Bài toán ước lượng tỷ lệ