Ước Lượng Khoảng
(Confidence Interval)

Khái Quát

- Ước lượng khoảng.
- Phương pháp khi σ^2 biết trước.
- Phương pháp với phân phối Student-t.
- Cách hiểu và giải thích ước lượng khoảng.

Ước Lượng Khoảng (Interval Estimation/ Confidence Interval) (1)

- Thu thập mẫu ngẫu nhiên gồm n quan sát:

$$X_1, X_2, X_3 \dots X_n$$

- Mẫu này được giả định là **I.I.D.** với hàm phân phối chung là $f(x_i)$, có trung bình μ và phương sai σ^2. Đây là hai ẩn số mà chúng ta cần tìm hiểm.
- **Ước lượng khoảng (interval estimation/ confidence interval)** là đi tìm khoảng (phạm vi) với một xác suất nhất định mà ẩn số nằm trong đó.

Ước Lượng Khoảng (Interval Estimation/ Confidence Interval) (2)

- Ví dụ: với độ tin cậy 95% (confidence interval at 95%) thì chiều cao trung bình của nam giới Việt Nam lấy từ một mẫu so sánh là 167cm – 169cm.

Phương Pháp Xác Định Ước Lượng Khoảng

- Từ giả định I.I.D. về mẫu, chung ta có:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

- Chọn mức độ tin cậy (confidence interval) ở mức 95%: $1 - \alpha = 0.95 \rightarrow Z_{\alpha=0.025} = \pm 1.96$

$$\rightarrow \mathrm{P}(-1.96 \leq \mathrm{Z} \leq 1.96) = 0.95 \rightarrow \mathrm{P}\left(-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) = 0.95$$

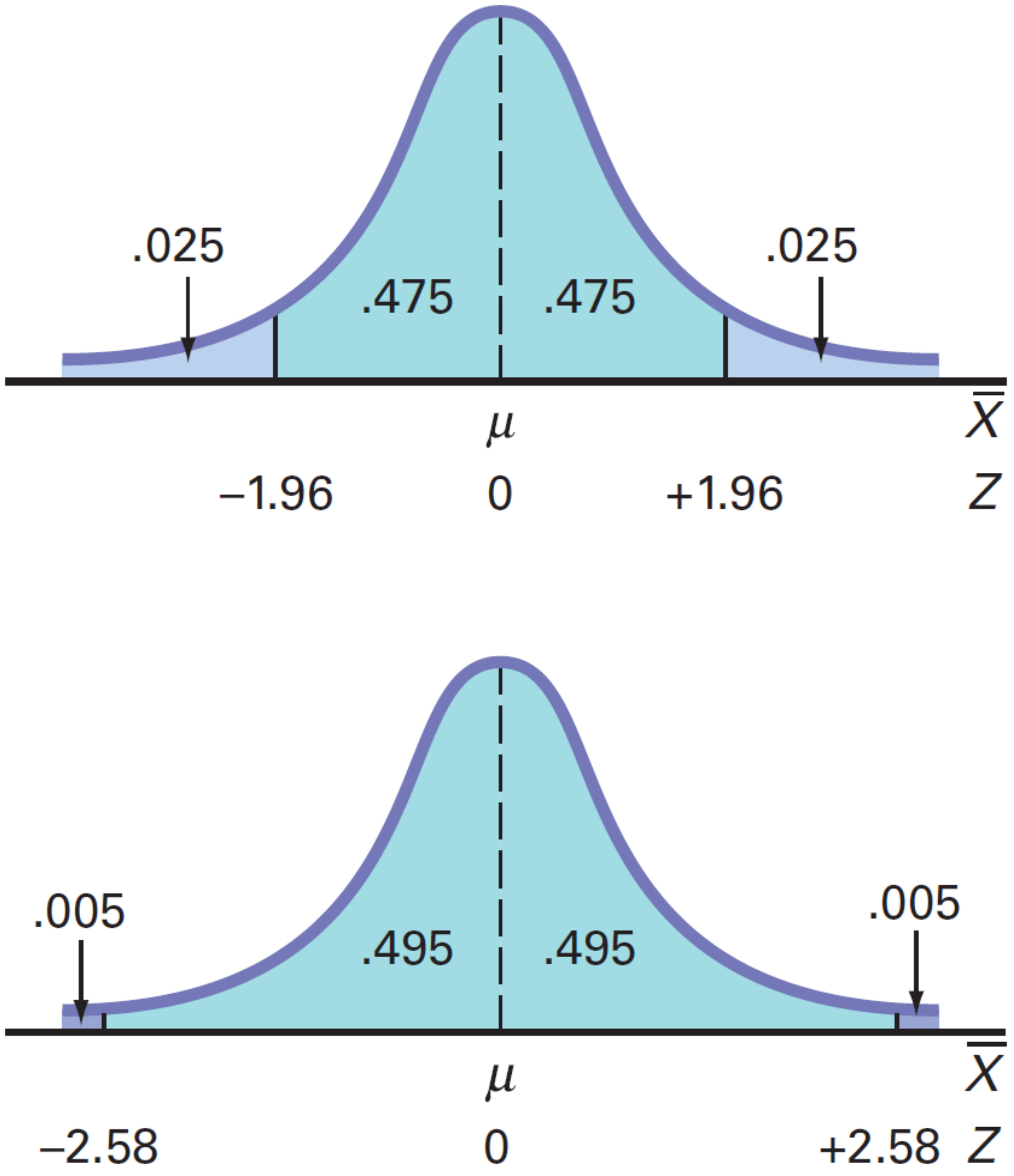
$$\rightarrow \mathrm{P}\left(-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\rightarrow \mathrm{P}\left(1.96\frac{\sigma}{\sqrt{n}} \geq \mu - \bar{X} \geq -1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\rightarrow \mathbf{P}\left(\bar{\boldsymbol{X}} + \mathbf{1.96}\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} \geq \boldsymbol{\mu} \geq \bar{\boldsymbol{X}} - \mathbf{1.96}\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right) = \mathbf{0.95}$$

Ví dụ:

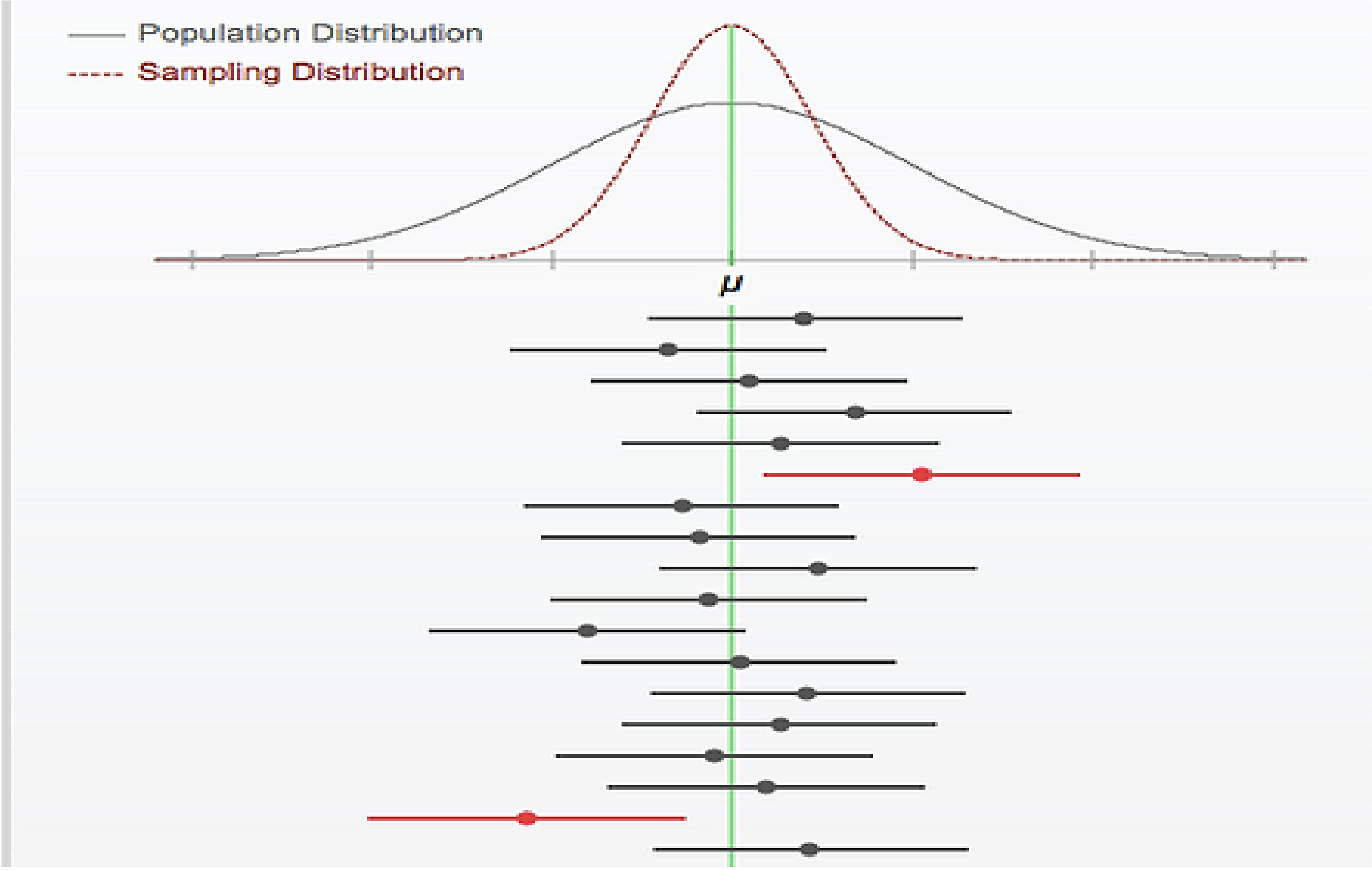
- Giá bán của một cục sạc điện thoại USB-C trên thị trường là một biến ngẫu nhiên có phân phối chuẩn. Một đơn vị nghiên cứu thị trường, khảo sát ngẫu nhiên tại 15 gian hàng trên Shopee và Lazada tìm được giá bán trung bình là 8.5 USD với độ lệch chuẩn là 0.5 USD. Với độ tin cậy 95%, hãy ước lượng khoảng cho giá bán trung bình chung của thiết bị trên thị trường.



.025
.475
.475
.025
μ
X̄
−1.96
0
+1.96
Z
.005
.495
.495
.005
μ
X̄
−2.58
0
+2.58
Z

Cách Hiểu và Giải Thích Ước Lượng Khoảng

- Thực hiện lấy mẫu rất nhiều lần (k lần tổng cộng).
- Khi chọn $1 - \alpha = 95\%$, tỷ lệ số lần μ rơi vào khoảng tin cậy là 95% c của tổng số lần lấy mẫu k.
- Phát biểu “xác suất của giá bán thiết bị nằm trong khoảng \$8.25 tới \$8.75 là 95%” có chính xác?



Phân Phối Student's *t* (1)

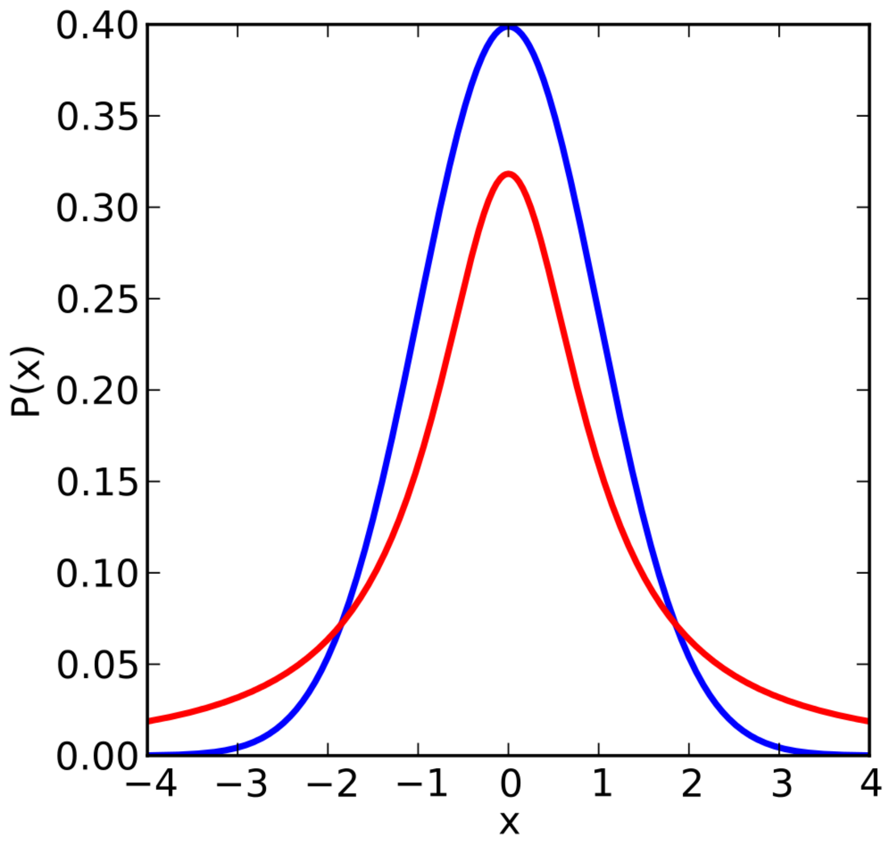
- Trường hợp σ^2 được nhà nghiên cứu biết trước là không thực tế.
- Chúng ta phải dùng ước lượng không chệch S^2 cho phương sai σ^2, khi đó chúng ta không còn sử dụng phân phối chuẩn, mà sử dụng **phân phối Student's *t* (the Student's t distribution)** hay **phân phối *t*** như sau:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

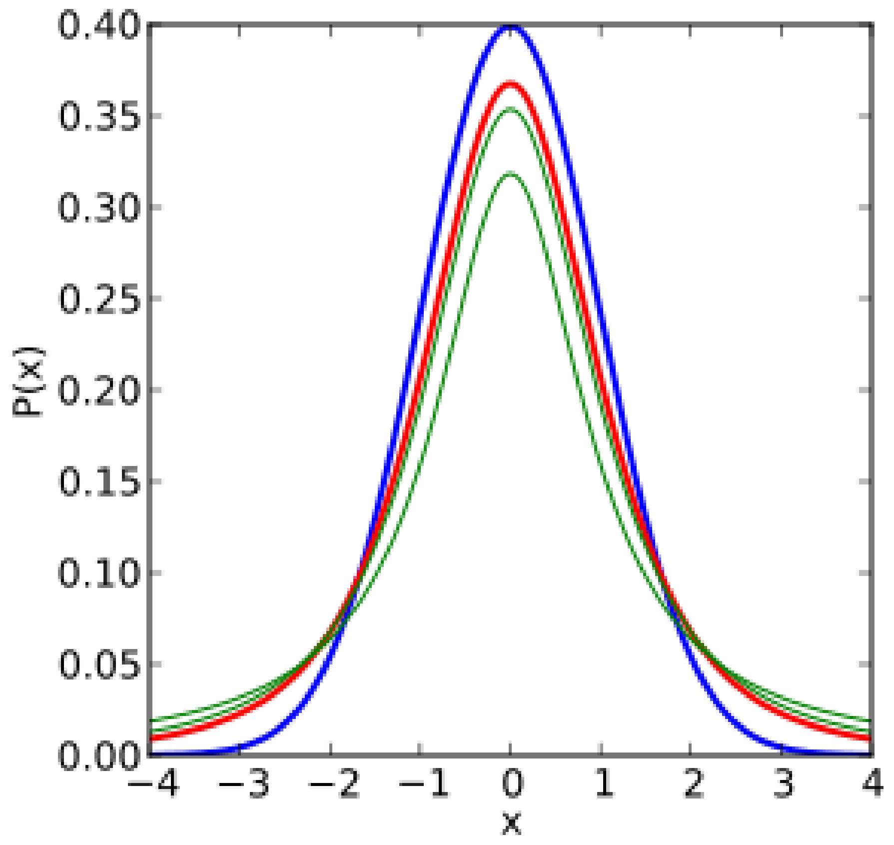
với $n - 1$ là số bậc tự do (degree of free – d.o.f.) và

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Phân Phối Student-t: khi n tăng thì phân phối Student-t xấp xỉ dần tới phân phối chuẩn



Màu xanh dương: phôi phối chuẩn.
Màu đỏ: phân phối t với 1 bậc tự do.



Màu xanh dương: phôi phối chuẩn.
Màu đỏ: phân phối t với 3 bậc tự do.
Màu xanh lá cây: phối phối t với 1 và 2 bậc tự do.

Phân Phối Student's t (2)

- Bắt đầu từ:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

- Biến đổi tương tự với trường hợp σ^2 không phải là ẩn số:

$$\to P\left(\bar{X} + t_{n-1,\,\alpha/2}\frac{S}{\sqrt{n}} \geq \mu \geq \bar{X} - t_{n-1,\,\alpha/2}\frac{S}{\sqrt{n}}\right) = 0.95$$

với $1 - \alpha = 0.95$, $t_{n-1,\,\alpha/2}$ phụ thuộc vào giá trị n và α.

Ví dụ:

- Khảo sát chiều cao 101 sinh viên nam của một lớp thuộc ĐHQG TP.HCM. Chiều cao trung bình của các sinh viên là 1.70m với độ lệch chuẩn là 8cm. Với mức độ tin cậy 99%, hãy xác định khoảng tin cậy cho chiều cao trung bình của toàn bộ sinh viên ĐHQG.