Bài 4: Lý thuyết ước lượng

Vinh Lương

University of Economics and Finance

vinhlx@uef.edu.vn

September 2018

Contents

Khoảng tin cậy cho trung bình μ
- Giới thiệu về ước lượng
- Công thức tìm khoảng tin cậy cho trung bình μ
- Công thức tìm khoảng tin cậy cho trung bình μ khi mẫu lớn và mẫu bé

Khoảng tin cậy cho tỷ lệ p của tổng thể

Bài tập tổng hợp

Contents

Khoảng tin cậy cho trung bình μ

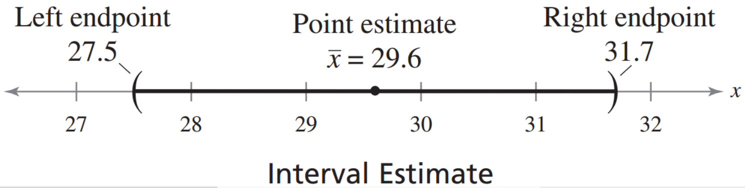
- Giới thiệu về ước lượng
- Công thức tìm khoảng tin cậy cho trung bình μ
- Công thức tìm khoảng tin cậy cho trung bình μ khi mẫu lớn và mẫu bé

Khoảng tin cậy cho tỷ lệ p của tổng thể

Bài tập tổng hợp

Giới thiệu về ước lượng - Estimation Process

- Ước lượng là **quá trình suy luận về thông tin của tổng thể** (không biết trước) khi đã biết thông tin của 1 mẫu ngẫu nhiên.
- Giả sử bạn cần thông tin về tuổi trung bình của toàn bộ người tham dự một lễ hội âm nhạc ngoài trời. Vì số lượng người là rất lớn, ta không thể khảo sát hết được. Ta chọn ra một số người để hỏi tuổi (ví dụ: $n = 100$) và tính ra trung bình là $\bar{x} = 29.6$ tuổi.
- Bạn đi đến kết luận là tuổi trung bình của toàn bộ người sẽ nằm trong khoảng: $29.6 - 2.1 < \mu < 29.6 + 2.1$?
- Hoặc đi đến kết luận là tuổi trung bình của toàn bộ người sẽ nằm trong khoảng: $29.6 - 1 < \mu < 29.6 + 1$?
- Hoặc đi đến kết luận là $\mu = 29.6$?



Point estimation và interval estimation.

- **A point estimate** is a specific numerical value estimate of a parameter. The best point estimate of the population mean μ is the sample mean $\bar{x}$.

For example, **an point estimate** for the average age of all people might be: $\mu = \bar{x} = 29.6$.

- **An interval estimate** of a parameter is an interval or a range of values used to estimate the parameter. This estimate may or may not contain the value of the parameter being estimated.

For example, **an interval estimate** for the average age of all people might be: $29.6 - 2.1 < \mu < 29.6 + 2.1$.

QUIZ: WHICH ONE IS BETTER ?

Giới thiệu về ước lượng

Tại sao ta cần phải đi ước lượng ?

- Giúp bạn đưa ra được **khoảng ước lượng cho tham số tổng thể** phù hợp với mục đích của người làm nghiên cứu.
- Giúp bạn lựa chọn được kích thước mẫu phù hợp.
- Giúp bạn tính toán được xác suất tham số tổng thể sẽ rơi vào khoảng ước lượng.

Khoảng tin cậy cho trung bình μ

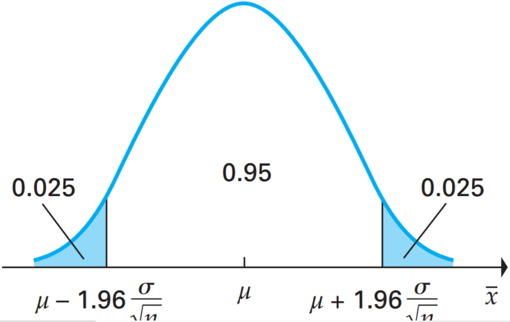
Phần trước chúng ta đã biết: theo định lí giới hạn trung tâm (Central Limit Theorem) thì $\overline{X} \sim N(\mu_{\bar{X}}, \sigma^2_{\bar{X}})$.

Nếu ta biết trước μ và σ thì:

$P(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$

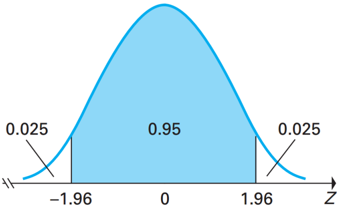
Hoặc nếu như ta coi như chưa biết μ thì:

$P(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$



Khoảng tin cậy cho trung bình μ

Proof: Ta đặt $Z = \dfrac{\overline{X} - \mu}{\sigma}$ thì Z có phân phối chuẩn tắc $N(0,1)$



$$
\begin{aligned}
1 - \alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \\
&= P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)
\end{aligned}
$$

Khoảng tin cậy cho trung bình μ

TRƯỜNG HỢP 1:
Phương pháp tìm khoảng tin cậy cho trung bình μ
(Khi biết σ của tổng thể)

- Với độ tin cậy là 95 % thì:

$$\boxed{\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}}$$

- Tổng quát với độ tin cậy $(1-\alpha)$ bất kì thì:

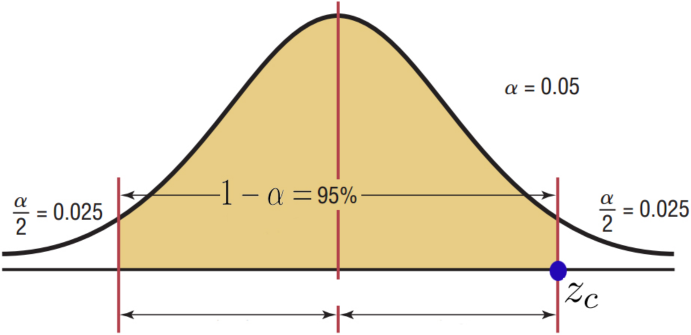
$$\boxed{\overline{x} - z_c\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_c\frac{\sigma}{\sqrt{n}}}$$

- Trong đó z_c được gọi là **giá trị tới hạn** tương ứng với độ tin cậy $(1-\alpha)$ cho trước.
- Khoảng tin cậy càng rộng thì độ tin cậy càng cao.
- Các giá trị độ tin cậy hay sử dụng là: **90%, 95% và 99%**

Khoảng tin cậy cho trung bình μ

Cách tính giá trị tới hạn z_c:

- Giá trị z_c được tính bằng cách sử dụng phân phối chuẩn tắc $Z \sim N(0,1)$
- Excel: $z_c = NORM.S.INV(1-\alpha/2)$



Độ tin cậy $(1-\alpha)$	90%	95%	99%
z_c	1.64	1.96	2.58

Example 1

Giả sử thời gian mua sắm của 1 người bất kì tại một trung tâm mua sắm có phân phối chuẩn với độ lệch chuẩn được biết là 20 phút. (dựa vào dữ liệu quá khứ)
Để ước lượng thời gian mua sắm trung bình trong năm nay. Người ta lấy một mẫu gồm 64 người và tính toán ra thời gian mua sắm trung bình của 64 người đó là 75 phút.
Với độ tin cậy 95%, hãy tìm khoảng tin cậy của thời gian mua sắm trung bình của toàn bộ tổng thể.

Giải:

- Gọi μ là thời gian shopping trung bình của tổng thể.
- Công thức ước lượng cho μ khi đã biết σ là:

$$\boxed{\overline{x} - z_c \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_c \frac{\sigma}{\sqrt{n}}}$$

- Trong đó z_c là giá trị tới hạn tương ứng với độ tin cậy 95%
- $1 - \alpha = 95\%$ Suy ra: $z_c = 1.96$
- Suy ra:

$$75 - 1.96 \times \frac{20}{\sqrt{64}} < \mu < 75 + 1.96 \times \frac{20}{\sqrt{64}}$$

- Suy ra: $70.1 < \mu < 79.9$
- Kết luận: Với độ tin cậy 95% thì thời gian shopping trung bình của tổng thể được ước lượng nằm trong khoảng từ 70.1 phút đến 79.9 phút.

Khoảng tin cậy cho trung bình μ

TRƯỜNG HỢP 2:
Phương pháp tìm khoảng tin cậy cho trung bình μ
(Khi KHÔNG biết σ của tổng thể)

- **Ta thay thế σ bằng giá trị $\hat{s}$ của mẫu.**
- Nếu mẫu lớn $n \geq 30$ thì ta dùng z_c:

$$\boxed{\overline{x} - z_c \frac{\hat{s}}{\sqrt{n}} < \mu < \overline{x} + z_c \frac{\hat{s}}{\sqrt{n}}}$$

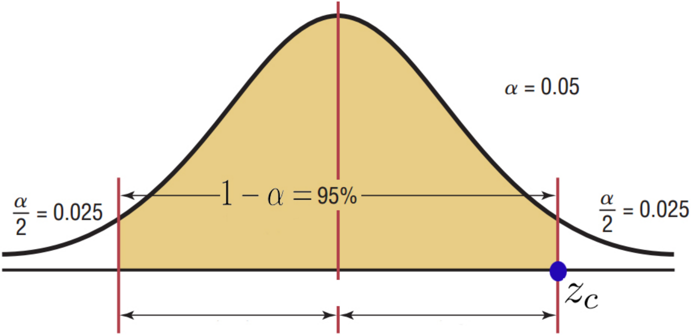
- Nếu mẫu nhỏ $n < 30$ thì ta dùng t_c thay cho z_c

$$\boxed{\overline{x} - t_c \frac{\hat{s}}{\sqrt{n}} < \mu < \overline{x} + t_c \frac{\hat{s}}{\sqrt{n}}}$$

- **Giá trị t_c được tra bằng Phân Phối Student với (n-1) bậc tự do.**

Khoảng tin cậy cho trung bình μ

Cách tìm t_c khi mẫu bé: Tra bảng Student với (n-1) bậc tự do



Độ tin cậy $1-\alpha$	90%	95%	99%
Bậc tự do $(n-1)$	24	24	24
t_c	1.71	2.06	2.8

Sai số epsilon trong ước lượng

- **Độ chính xác ε (Margin of Error): Là đại lượng cho biết độ rộng hẹp của khoảng tin cậy.**
- Như trên ta có:

$$\varepsilon = z_c \frac{\sigma}{\sqrt{n}}$$

- Hoặc khi không biết σ thì:

$$\varepsilon = z_c \frac{\hat{s}}{\sqrt{n}}$$

- Muốn độ chính xác càng nhỏ thì n càng phải to
- Do đó ta có thể thay đổi kích thước mẫu để đạt được độ chính xác như mong muốn.

Example 2

Để ước lượng tuổi thọ trung bình của một loại bóng đèn, người ta kiểm tra ngẫu 50 bóng và tính được tuổi thọ trung bình của chúng là X = 1200 giờ với độ lệch tiêu chuẩn mẫu điều chỉnh 26.5 giờ.

1. Hãy ước lượng tuổi thọ trung bình của bóng đèn với độ tin cậy là 95%; giả thiết rằng tuổi thọ của bóng đèn có phân phối chuẩn.
2. Ta phải chọn kích thước mẫu tối thiểu n bằng bao nhiêu để với độ tin cậy 95%, sai lệch của phép ước lượng tuổi thọ trung bình sẽ không vượt quá 2 giờ.

Giải:

- Gọi μ là tuổi thọ trung bình của toàn bộ bóng đèn.
- Công thức ước lượng cho μ khi đã biết $\hat{s}$ là:

$$\boxed{\overline{x} - z_c\frac{\hat{s}}{\sqrt{n}} < \mu < \overline{x} + z_c\frac{\hat{s}}{\sqrt{n}}}$$

- Mẫu lớn nên ta dùng z_c
- Trong đó z_c là giá trị tới hạn tương ứng với độ tin cậy 95%
- $1 - \alpha = 95\%$ Suy ra: $z_c = 1.96$
- Suy ra:

$$1200 - 1.96 \times \frac{26.5}{\sqrt{50}} < \mu < 1200 + 1.96 \times \frac{26.5}{\sqrt{50}}$$

- Suy ra: $1192.65 < \mu < 1207.35$
- Kết luận: Với độ tin cậy 95% thì tuổi thọ trung bình của toàn bộ bóng đèn được ước lượng nằm trong khoảng từ 1192.65 giờ đến 1207.35 giờ.
- Cho $\varepsilon \leq 2$ suy ra được $n \geq 674.44$, kích thước mẫu tối thiểu n=675

Example 3

Dưới đây là bảng khảo sát thời gian ngủ trung bình mỗi ngày của từng người trong 25 người được khảo sát.(đơn vị : giờ)
(The Macmillan Visual Almanac, 1996, USA)
a. Tính thời gian ngủ trung bình mỗi ngày của mẫu 25 người trên.
b. Giả sử trong tổng thể toàn nước Mỹ thì thời gian ngủ trung bình mỗi ngày của 1 người tuân theo phân phối chuẩn. Với độ tin cậy 95%, hãy ước lượng thời gian ngủ trung bình mỗi ngày μ của tổng thể là toàn bộ người Mỹ.

6.9	5.5	6	7.6	7.6
7.8	7.1	8.6	7	6.9
7.3	7.2	7.7	6.6	5.8
6.8	6	7.1	6.5	7.2
7.6	6.2	6.5	5.3	6.7

Khoảng tin cậy cho trung bình μ

Kích thước mẫu	$n =$	25
Độ tin cậy	$1 - \alpha =$	95%
Trung bình mẫu	$\bar{x} =$	6.86
Độ lệch chuẩn mẫu hiệu chỉnh	$\hat{s} =$	0.7773

Gọi μ là thời gian ngủ trung bình của tổng thể.
Vì mẫu nhỏ $n < 30$ nên ta dùng công thức:

$$\boxed{\overline{x} - t_c \frac{\hat{s}}{\sqrt{n}} < \mu < \overline{x} + t_c \frac{\hat{s}}{\sqrt{n}}}$$

Giá trị tới hạn $t_c = 2.06$ (n=25 thì số bậc tự do là d.f=24)
Suy ra: $6.54 < \mu < 7.18$
Kết luận: với độ tin cậy 95%, thời gian ngủ trung bình của người dân Mỹ được ước lượng nằm trong khoảng từ 6.54 giờ đến 7.18 giờ

Example 4

Bài 6.8: Thời gian phản ứng của một người đối với một tác động kích thích nào đó được ghi nhận với 5 kết quả như sau:
0,28; 0,30; 0,27; 0,33; 0,31 giây.
Tìm khoảng tin cậy của thời gian phản ứng trung bình thực sự của tổng thể, với độ tin cậy là:

1. $1 - \alpha = 95\%$.
2. $1 - \alpha = 99\%$.

Khoảng tin cậy cho trung bình μ

Giải:

Kích thước mẫu	$n =$	5
Độ tin cậy	$1-\alpha =$	95%
Trung bình mẫu	$\bar{x} =$	0.298
Độ lệch chuẩn mẫu hiệu chỉnh	$\hat{s} =$	0.0238

Gọi μ là thời gian phản ứng trung bình thực sự của tổng thể.
Vì mẫu nhỏ $n < 30$ nên ta dùng công thức:

$$\boxed{\overline{x} - t_c\frac{\hat{s}}{\sqrt{n}} < \mu < \overline{x} + t_c\frac{\hat{s}}{\sqrt{n}})}$$

Giá trị tới hạn $t_c = 2.776$ (tra bảng Student với d.f=4)
Suy ra: $\mathbf{0.2683} < \mu < \mathbf{0.3276}$
Kết luận: với độ tin cậy 95%, thời gian phản ứng trung bình của người đó được ước lượng nằm trong khoảng từ 268.3ms đến 327.6ms.

Exercise 1

(mẫu nhỏ) Trọng lượng của những vỉ thuốc do một công ty dược sản xuất có phân phối chuẩn. Một mẫu ngẫu nhiên gồm 10 vỉ thuốc có trọng lượng trung bình là 4.87mg với độ lệch chuẩn hiệu chỉnh là 0.038mg. Với độ tin cậy 95%, hãy tìm khoảng tin cậy của trọng lượng trung bình các vỉ thuốc do công ty sản xuất.

Exercise 2

(mẫu lớn) Một mẫu gồm 250 đầu đinh tán do một công ty sản xuất có đường kính trung bình và độ lệch chuẩn lần lượt là 0,72642 cm và 0,00058 cm. Tìm khoảng tin cậy đường kính trung bình của các đầu đinh tán do công ty này sản xuất, với độ tin cậy la:

1. *99%.*
2. *95%.*

Exercise 3

Giả sử tuổi thọ bóng đèn của công ty A là đại lượng ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là $\sigma = 100$ *giờ. Lấy một mẫu gồm 50 bóng đèn và kiểm tra tuổi thọ của 50 bóng đó. Ta tính được tuổi thọ trung bình là 1250 giờ. Với độ tin cậy 90%, hãy tìm khoảng tin cậy của tuổi thọ trung bình của toàn bộ bóng đèn do công ty này sản xuất.*

Khoảng tin cậy cho trung bình μ

Exercise 4

A college admissions director wishes to estimate the mean age of all students currently enrolled. In a random sample of 20 students, the mean age is found to be 22.9 years. From past studies, the standard deviation is known to be 1.5 years, and the population is normally distributed. Construct a 90% confidence interval of the population mean age.

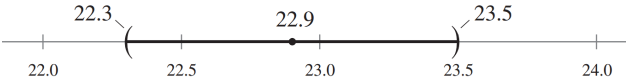
Left Endpoint

$$\bar{x} - E \approx 22.9 - 0.6 = 22.3$$

Right Endpoint

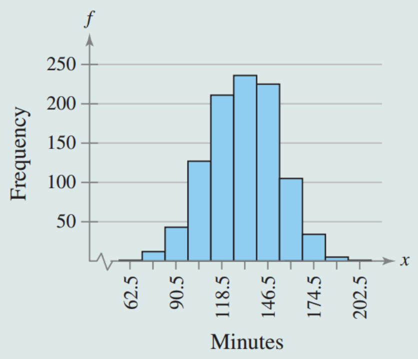
$$\bar{x} + E \approx 22.9 + 0.6 = 23.5$$

$$22.3 < \mu < 23.5$$



Exercise 5

A survey of a random sample of 1000 smartphone owners found that the mean daily time spent communicating on a smartphone was 131.4 minutes. From previous studies, it is assumed that the population standard deviation is 21.2 minutes.
What would be 95% confidence interval for the population mean daily time spent communicating on a smartphone?



Exercise 6

(Distance Traveled to Work)
A recent study of 28 employees of XYZ company showed that the mean of the distance they traveled to work was 14.3 miles. The standard deviation of the sample mean was 2 miles.

1. *Find the 95% confidence interval of the true mean.*
2. *If a manager wanted to be sure that most of his employees would not be late, how much time would he suggest they allow for the commute if the average speed were 30 miles per hour?*

ĐS: $13.5 < \mu < 15.1$ and about 30 minutes.

Exercise 7

The data represent a sample of the number of home fires started by candles for the past several years. (Data are from the National Fire Protection Association.)
5900 ; 6090 ; 6310 ; 7160 ; 8440 ; 9930
Find the 99% confidence interval for the mean number of home fires started by candles each year.

Solution: Between 4785.2 and 9297.6

Contents

Khoảng tin cậy cho trung bình μ

Khoảng tin cậy cho tỷ lệ p của tổng thể

Bài tập tổng hợp

Khoảng tin cậy cho tỷ lệ p của tổng thể

Điều kiện:

- Gọi p là tỷ lệ của tổng thể (ta không biết)
- f là tỷ lệ của 1 mẫu ngẫu nhiên. (ta tính toán ra được)
- Ta chỉ khảo sát kích thước mẫu lớn n≥30

Khi đó công thức tìm khoảng tin cậy của tỷ lệ tổng thể p là:

$$\boxed{f - z_c.\sqrt{\frac{f(1-f)}{n}} < p < f + z_c.\sqrt{\frac{f(1-f)}{n}}}$$

- Trong đó z_c được gọi là **giá trị tới hạn** tương ứng với độ tin cậy $(1-\alpha)$ cho trước.
- Excel: $z_c = NORM.S.INV(1-\alpha/2)$

Example 5

Thăm dò ý kiến của 100 cử tri được chọn ngẫu nhiên tại một địa phương cho thấy có 80% trong số này ủng hộ ứng cử viên A.
Với độ tin cậy 99%, hãy ước lượng tỷ lệ của tất cả các cử tri tại địa phương này ủng hộ ứng cử viên A.

Khoảng tin cậy cho tỷ lệ p của tổng thể

- Gọi p tỷ lệ các cử tri bầu cho A tại địa phương.
- $f = 80\%$ là tỷ lệ cử tri bầu cho A
- $n = 100$ người được khảo sát.
- Công thức tìm khoảng tin cậy của tỷ lệ tổng thể p là:

$$\boxed{f - z_c.\sqrt{\frac{f(1-f)}{n}} < p < f + z_c.\sqrt{\frac{f(1-f)}{n}}}$$

- Ứng với độ tin cậy 99%, suy ra: $z_c = 2.58$
- Sai số $\varepsilon = z_c.\sqrt{\frac{f(1-f)}{n}} = 10.3\%$
- Khoảng tin cậy: $69.7\% < p < 90.3\%$
- Kết luận : Với độ tin cậy 99%, tỷ lệ của tất cả các cử tri tại địa phương này ủng hộ ứng cử viên A được ước lượng trong khoảng 69.7% đến 90.3% .

Khoảng tin cậy cho tỷ lệ p của tổng thể

Example 6

Khảo sát một mẫu gồm 100 người, cho thấy có 25 người thuận tay trái.

1. Với độ tin cậy 95%, hãy cho biết khoảng tin cậy của tỷ lệ người thuận tay trái của tổng thể.
2. Tìm n để chúng ta có độ chính xác không quá 5 %.

- $n = 100$ và $f = 25\%$
- Ứng với độ tin cậy 95%, suy ra: $z_c = 1.96$
- Sai số $\varepsilon = z_c.\sqrt{\dfrac{f(1-f)}{n}} = 8.487\%$
- Khoảng tin cậy: $16.51\% < p < 33.49\%$
- Ta cần tìm n sao cho sai số $\varepsilon = z_c.\sqrt{\dfrac{f(1-f)}{n}} \leq 5\%$
- Suy ra $n \geq$ .....

Exercise 8

Bài 6.10 : *Một lô hàng chứa nhiều sản phẩm loại A và sản phẩm loại B. Lấy một mẫu ngẫu nhiên gồm 60 sản phẩm được chọn có hoàn lại từ lô hàng này thì thấy có 70% sản phẩm loại A.*

1. *Tìm khoảng tin cậy của tỷ lệ các sản phẩm loại A thực sự có trong lô hàng, với độ tin cậy 95%, 99%.*
2. *Hãy tìm kích thước mẫu n sao cho tỷ lệ thực và tỷ lệ mẫu không khác nhau quá 5%, với các độ tin cậy lần lượt là 95%, 99%.*

Exercise 9

DVD Players A survey of 85 families showed that 36 owned at least one DVD player.

1. *Find the 99% confidence interval of the true proportion of families who own at least one DVD player.*
2. *If another survey in a different location found that the proportion of families who owned at least one DVD player was 0.52, would you consider that the proportion of families in this area was larger than in the area where the original survey was done?*

Contents

Khoảng tin cậy cho trung bình μ

Khoảng tin cậy cho tỷ lệ p của tổng thể

Bài tập tổng hợp

Exercise 10

Điểm thi môn XSTK của 1 mẫu gồm 50 sinh viên được tổng hợp lại thành bảng sau:

1. *Với độ tin cậy 95%, hãy ước lượng điểm trung bình môn XSTK của toàn thể sinh viên.*
2. *Học sinh dưới 44 điểm sẽ bị rớt môn XSTK. Với độ tin cậy 90%, hãy ước lượng tỷ lệ sinh viên rớt môn XSTK của tổng thể.*

Class	Frequency, f
26–34	2
35–43	5
44–52	12
53–61	18
62–70	11
71–79	1
80–88	1

$51.83 < \mu < 57.84$

$5.93\% < p < 22.07\%$

BÀI TẬP TỔNG HỢP

Exercise 11

Khảo sát chiều cao của một số nam sinh viên, ta có bảng:

1. *Với độ tin cậy 99%, hãy ước lượng chiều cao trung bình của toàn thể sinh viên.*
2. *Những sinh viên nam có chiều cao từ 1m75 được cho là sinh viên có chiều cao tốt. Với độ tin cậy 90%, hãy ước lượng tỷ lệ sinh viên có chiều cao tốt của tổng thể.*

Tâm lớp x_i	Chiều cao (cm) *(khoảng lớp)*	Số SV *(tần số m_i)*
154,5	151–158	5
162,5	159–166	18
170,5	167–174	41
178,5	175–182	27
186,5	183–190	8

Tài liệu tham khảo

Phạm Đắc Thắng, Nguyễn Công Trí, Đoàn Thiện Ngân
Giáo trình Xác suất thống kê

Murray R. Spiegel, John Schiller, R. Alu Srinivasan
Schaum's Outlines - Probability and Statistics

Anderson, Sweeney, Williams
Statistics for Business and Economics 11 edition

Ron Larson, Betsy Farber
Elementary Statistics, Picturing the world 6 edition