HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

o0o

CHƯƠNG 4: THỐNG KÊ - ƯỚC LƯỢNG THAM SỐ

Bài giảng môn Xác suất Thống kê

Giảng viên: TS. Nguyễn Kiều Linh

Hà Nội, năm 2020

Mục tiêu của Chương 4

Giúp sinh viên

- Hiểu được các khái niệm tổng thể, mẫu, tính được các đặc trưng của mẫu;
- Hiểu và trình bày được khái niệm về ước lượng tham số của biến ngẫu nhiên;
- Phân biệt được các bài toán ước lượng;
- Vận dụng cho các bài toán ước lượng điểm, ước lượng khoảng.

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - **1. Tổng thể**
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

1. Tổng thể

Khái niệm Tổng thể

Khi nghiên cứu các vấn đề về kinh tế - xã hội, cũng như nhiều vấn đề thuộc các lĩnh vực vật lý, sinh vật, quân sự, ... thường dẫn đến khảo sát một hay nhiều dấu hiệu, thể hiện bằng số lượng trên nhiều phần tử. Tập hợp tất cả các phần tử này gọi là tổng thể hay đám đông (population). Số phần tử trong tổng thể có thể là hữu hạn hoặc vô hạn.

1. Tổng thể

Khái niệm Tổng thể

Khi nghiên cứu các vấn đề về kinh tế - xã hội, cũng như nhiều vấn đề thuộc các lính vực vật lý, sinh vật, quân sự, . . . thường dẫn đến khảo sát một hay nhiều dấu hiệu, thể hiện bằng số lượng trên nhiều phần tử. Tập hợp tất cả các phần tử này gọi là tổng thể hay đám đông (population). Số phần tử trong tổng thể có thể là hữu hạn hoặc vô hạn.

1. Tổng thể

Khái niệm Tổng thể

Khi nghiên cứu các vấn đề về kinh tế - xã hội, cũng như nhiều vấn đề thuộc các lính vực vật lý, sinh vật, quân sự, . . . thường dẫn đến khảo sát một hay nhiều dấu hiệu, thể hiện bằng số lượng trên nhiều phần tử. Tập hợp tất cả các phần tử này gọi là tổng thể hay đám đông (population). Số phần tử trong tổng thể có thể là hữu hạn hoặc vô hạn.

1. Tổng thể

Một số ví dụ

- Muốn khảo sát xem chiều cao của thanh niên Việt Nam hiện nay có tăng lên hay không ta phải khảo sát toàn bộ thanh niên Việt Nam (giả sử 40 triệu người) $\Longrightarrow$ Để khảo sát sẽ tốn nhiều thời gian và kinh phí.

1. Tổng thể

Một số ví dụ

- Muốn khảo sát xem chiều cao của thanh niên Việt Nam hiện nay có tăng lên hay không ta phải khảo sát toàn bộ thanh niên Việt Nam (giả sử 40 triệu người) $\Longrightarrow$ Để khảo sát sẽ tốn nhiều thời gian và kinh phí.
- Trong kho có $N = 10000$ bóng đèn. Muốn biết tỉ lệ số bóng đèn hỏng sau một năm bảo quản ta kiểm tra từng bóng. Muốn kiểm tra thì ta cần thắp từng bóng xem có sáng bình thường không? Giả sử kiểm tra có $M = 100$ bóng hỏng. Như vậy tỉ lệ bóng hỏng là $M/N = 0,01$. Một sản phẩm sau khi kiểm tra thì mất phẩm chất, mất tính mới và vì vậy nếu kiểm tra cả kho thì cũng "hỏng" cả kho??? $\Longrightarrow$ Phá vỡ tổng thể nghiên cứu.

1. Tổng thể

Một số ví dụ

- Muốn khảo sát xem chiều cao của thanh niên Việt Nam hiện nay có tăng lên hay không ta phải khảo sát toàn bộ thanh niên Việt Nam (giả sử 40 triệu người) $\Longrightarrow$ Để khảo sát sẽ tốn nhiều thời gian và kinh phí.
- Trong kho có $N = 10000$ bóng đèn. Muốn biết tỉ lệ số bóng đèn hỏng sau một năm bảo quản ta kiểm tra từng bóng. Muốn kiểm tra thì ta cần thắp từng bóng xem có sáng bình thường không? Giả sử kiểm tra có $M = 100$ bóng hỏng. Như vậy tỉ lệ bóng hỏng là $M/N = 0,01$. Một sản phẩm sau khi kiểm tra thì mất phẩm chất, mất tính mới và vì vậy nếu kiểm tra cả kho thì cũng "hỏng" cả kho??? $\Longrightarrow$ Phá vỡ tổng thể nghiên cứu.
- Muốn khảo sát tỉ lệ người bị nhiễm HIV qua tiêm chích ta không thể khảo sát toàn bộ người bị nhiễm HIV, vì ta không thể xác định chính xác là bao nhiêu người bị nhiễm HIV. Do đó ta chỉ biết một phần tổng thể. Ngoài ra số người bị nhiễm HIV mới và bị chết do HIV thay đổi liên tục nên tổng thể thay đổi liên tục $\Longrightarrow$ Không xác định được chính xác tổng thể.

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

2. Mẫu

Do rất khó để khảo sát tổng thể nên người ta chọn ra một tập nhỏ để nghiên cứu và đưa ra quyết định, gọi là Mẫu.

Khái niệm Mẫu

- Mẫu là một tập con của tổng thể và có tính chất tương tự như tổng thể.
- Số phần tử của mẫu được gọi là kích thước mẫu, ký hiệu là n.

2. Mẫu

Do rất khó để khảo sát tổng thể nên người ta chọn ra một tập nhỏ để nghiên cứu và đưa ra quyết định, gọi là Mẫu.

Khái niệm Mẫu

- Mẫu là một tập con của tổng thể và có tính chất tương tự như tổng thể.
- Số phần tử của mẫu được gọi là kích thước mẫu, ký hiệu là n.

2. Mẫu

Do rất khó để khảo sát tổng thể nên người ta chọn ra một tập nhỏ để nghiên cứu và đưa ra quyết định, gọi là Mẫu.

Khái niệm Mẫu

- Mẫu là một tập con của tổng thể và có tính chất tương tự như tổng thể.
- Số phần tử của mẫu được gọi là kích thước mẫu, ký hiệu là n.



Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - **3. Mô hình hóa mẫu ngẫu nhiên**
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

3. Mô hình hóa mẫu ngẫu nhiên

Nhận xét

- Ta nói rằng một mẫu là mẫu ngẫu nhiên nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.

3. Mô hình hóa mẫu ngẫu nhiên

Nhận xét

- Ta nói rằng một mẫu là mẫu ngẫu nhiên nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.
- Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X. Với mỗi mẫu ta chỉ cần quan tâm dấu hiệu nghiên cứu X của mỗi cá thể của mẫu.

3. Mô hình hóa mẫu ngẫu nhiên

Nhận xét

- Ta nói rằng một mẫu là mẫu ngẫu nhiên nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.
- Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X. Với mỗi mẫu ta chỉ cần quan tâm dấu hiệu nghiên cứu X của mỗi cá thể của mẫu.

Ví dụ

Khi cần nghiên cứu chiều cao trung bình của thanh niên trong một vùng nào đó thì với cá thể A được chọn làm mẫu ta chỉ quan tâm về chiều cao của A, tức là dấu hiệu chiều cao X_A, mà không quan tâm đến các đặc trưng khác của cá thể này.

3. Mô hình hóa mẫu ngẫu nhiên

Mẫu ngẫu nhiên

Mẫu ngẫu nhiên kích thước n là một dãy gồm n biến ngẫu nhiên: $X_1, X_2, \ldots, X_n$ độc lập cùng phân bố với X, ký hiệu $W = (X_1, X_2, \ldots, X_n)$, trong đó X_i là dấu hiệu X của phần tử thứ i của mẫu $(i = 1, 2, \ldots, n)$.

3. Mô hình hóa mẫu ngẫu nhiên

Mẫu ngẫu nhiên

Mẫu ngẫu nhiên kích thước n là một dãy gồm n biến ngẫu nhiên: $X_1, X_2, \ldots, X_n$ độc lập cùng phân bố với X, ký hiệu $W = (X_1, X_2, \ldots, X_n)$, trong đó X_i là dấu hiệu X của phần tử thứ i của mẫu $(i = 1, 2, \ldots, n)$.

Giá trị cụ thể của mẫu ngẫu nhiên

- Thực hiện một phép thử đối với mẫu ngẫu nhiên chính là thực hiện một phép thử đối với mỗi thành phần của mẫu.
- Giả sử X_i nhận giá trị x_i $(i = 1, 2, \ldots, n)$, khi đó các giá trị $x_1, x_2, \ldots, x_n$ tạo thành một giá trị cụ thể của mẫu ngẫu nhiên, hay còn gọi là một thể hiện của mẫu ngẫu nhiên, ký hiệu $w = (x_1, x_2, \ldots, x_n)$.

3. Mô hình hóa mẫu ngẫu nhiên

Ví dụ

- Gọi X là số chấm xuất hiện khi tung con xúc xắc cân đối, khi đó X là một biến ngẫu nhiên.
- Giả sử tung con xúc xắc 3 lần, gọi X_i là số chấm cuất hiện trong lần tung thứ $i(i = 1, 2, 3)$ thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân bố xác suất với X. Vậy mẫu ngẫu nhiên kích thước 3, $W = (X_1, X_2, X_3)$.
- Thực hiện một phép thử đối với mẫu ngẫu nhiên này tức là tung con xúc xắc 3 lần. Giả sử lần thứ nhất được 2 chấm, lần thứ hai được 5 chấm, lần ba được 3 chấm thì $w = (2, 5, 3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W.

3. Mô hình hóa mẫu ngẫu nhiên

Ví dụ

- Gọi X là số chấm xuất hiện khi tung con xúc xắc cân đối, khi đó X là một biến ngẫu nhiên.
- Giả sử tung con xúc xắc 3 lần, gọi X_i là số chấm cuất hiện trong lần tung thứ $i(i = 1, 2, 3)$ thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân bố xác suất với X. Vậy mẫu ngẫu nhiên kích thước 3, $W = (X_1, X_2, X_3)$.
- Thực hiện một phép thử đối với mẫu ngẫu nhiên này tức là tung con xúc xắc 3 lần. Giả sử lần thứ nhất được 2 chấm, lần thứ hai được 5 chấm, lần ba được 3 chấm thì $w = (2, 5, 3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W.

Câu hỏi

Làm thế nào để chọn được mẫu có tính chất tương tự như tổng thể, các kết luận của mẫu có thể dùng cho tổng thể?

Nội dung

4. Một số cách chọn mẫu cơ bản

Dạng liệt kê

- **Chọn mẫu ngẫu nhiên có hoàn lại:** Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Lặp lại như thế n lần ta được một mẫu có hoàn lại gồm n phần tử.

4. Một số cách chọn mẫu cơ bản

Dạng liệt kê

- **Chọn mẫu ngẫu nhiên có hoàn lại:** Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Lặp lại như thế n lần ta được một mẫu có hoàn lại gồm n phần tử.
- **Chọn mẫu ngẫu nhiên không hoàn lại:** Lấy ngẫu nhiên một phần tử của tổng thể và khảo sát rồi để qua một bên, không trả lại tổng thẻ. Sau đó lấy ngẫu nhiên một phần tử khác. Lặp lại như thế n lần ta được một mẫu n phần tử không hoàn lại.

4. Một số cách chọn mẫu cơ bản

Dạng liệt kê

- **Chọn mẫu ngẫu nhiên có hoàn lại:** Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Lặp lại như thế n lần ta được một mẫu có hoàn lại gồm n phần tử.
- **Chọn mẫu ngẫu nhiên không hoàn lại:** Lấy ngẫu nhiên một phần tử của tổng thể và khảo sát rồi để qua một bên, không trả lại tổng thẻ. Sau đó lấy ngẫu nhiên một phần tử khác. Lặp lại như thế n lần ta được một mẫu n phần tử không hoàn lại.
- **Chọn mẫu phân nhóm:** Đầu tiên ta chia tập nền thành các nhóm tương đối thuần nhất, từ mỗi nhóm đó chọn ra một mẫu ngẫu nhiên. Tập hợp tất cả các mẫu đó cho ta một mẫu phân nhóm. Phương pháp này dùng khi tập nền có những sai khác lớn giữa các cá thể. Hạn chế là phụ thuộc vào việc chia nhóm.

4. Một số cách chọn mẫu cơ bản

Dạng liệt kê

- **Chọn mẫu ngẫu nhiên có hoàn lại:** Lấy ngẫu nhiên 1 phần tử từ tổng thể và khảo sát nó. Sau đó trả phần tử đó lại tổng thể trước khi lấy 1 phần tử khác. Lặp lại như thế n lần ta được một mẫu có hoàn lại gồm n phần tử.
- **Chọn mẫu ngẫu nhiên không hoàn lại:** Lấy ngẫu nhiên một phần tử của tổng thể và khảo sát rồi để qua một bên, không trả lại tổng thẻ. Sau đó lấy ngẫu nhiên một phần tử khác. Lặp lại như thế n lần ta được một mẫu n phần tử không hoàn lại.
- **Chọn mẫu phân nhóm:** Đầu tiên ta chia tập nền thành các nhóm tương đối thuần nhất, từ mỗi nhóm đó chọn ra một mẫu ngẫu nhiên. Tập hợp tất cả các mẫu đó cho ta một mẫu phân nhóm. Phương pháp này dùng khi tập nền có những sai khác lớn giữa các cá thể. Hạn chế là phụ thuộc vào việc chia nhóm.
- **Chọn mẫu có suy luận:** Dựa trên ý kiến của chuyên gia về đối tượng nghiên cứu để chọn mẫu.

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - **5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên**
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

- **Liệt kê:** Các số liệu thu được ta ghi lại thành dãy số liệu: $x_1, x_2, \ldots, x_n$.

5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

- **Liệt kê:** Các số liệu thu được ta ghi lại thành dãy số liệu: $x_1, x_2, \ldots, x_n$.
- **Bảng phân bố tần số:** Từ một mẫu cụ thể của mẫu ngẫu nhiên kích thước n của X ta sắp xếp các giá trị của mẫu cụ thể theo thứ tự tăng dần, giả sử giá trị x_i xuất hiện với tần số n_i, $i = 1, 2, \ldots, k$. Khi đó ta có thể biểu diễn mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm sau

Giá trị	x_1	x_2	$\ldots$	x_k
Tần số	n_1	n_2	$\ldots$	n_k

trong đó: $n_1 + n_2 + \ldots + n_k = n$.

5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

- **Bảng phân bố tần suất:** Ký hiệu $f_i = \dfrac{n_i}{n}$ và gọi là tần suất của x_i. Ta có bảng phân bố tần suất thực nghiệm tương ứng của X.

Giá trị	x_1	x_2	$\ldots$	x_k
Tần số	f_1	f_2	$\ldots$	f_k

5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên

- **Bảng phân bố ghép lớp:** Dữ liệu thu được nhận giá trị trong khoảng (a, b). Chia (a, b) thành k miền con bởi các điểm chia: $a_0 = a < a_1 < a_2 < \ldots < a_{k-1} < a_k = b$.

Giá trị	$(a_0 - a_1]$	$(a_1 - a_2]$	$\ldots$	$(a_{k-1} - a_k]$
Tần số	n_1	n_2	$\ldots$	n_k

Thông thường độ dài các khoảng chia bằng nhau.

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

1. Định nghĩa thống kê

Định nghĩa thống kê

Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu. Thống kê của mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ có dạng:

$$T = T(X_1, X_2, \dots, X_n).$$

Nhận xét

- Như vậy thống kê cũng là một biến ngẫu nhiên tuân theo một quy luật phân bố xác suất nhất định và có các tham số đặc trưng như kỳ vọng ET, phương sai DT,.... Mặt khác, khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, \dots, x_n)$ thì T cũng nhận một giá trị cụ thể còn gọi là giá trị quan sát của thống kê

$$T_{qs} = T(x_1, x_2, \dots, x_n).$$

- Các thống kê cùng với quy luật phân bố xác suất của chúng là cơ sở để suy rộng các thông tin của mẫu cho dấu hiệu nghiên cứu của tổng thể.

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - **2. Các đặc trưng của mẫu ngẫu nhiên**
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

2.1. Trung bình mẫu

Trung bình mẫu

- Trung bình mẫu của mẫu ngẫu nhiên $W = (X_1, X_2, \ldots, X_n)$ của biến ngẫu nhiên gốc X được định nghĩa và ký hiệu

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Giá trị trung bình mẫu cụ thể của mẫu ngẫu nhiên cụ thể là

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- Giả sử dấu hiệu nghiên cứu X có kỳ vọng và phương sai hữu hạn, áp dụng các công thức tính kỳ vọng và phương sai của tổng các biến ngẫu nhiên độc lập ta có

$$E\bar{X} = EX;\ D\bar{X} = \frac{DX}{n}.$$

Phương sai mẫu

- Phương sai mẫu $\widehat{S}^2$

$$\widehat{S}^2 = \frac{1}{n}X_i^2 - (\bar{X})^2, \; E\widehat{S} = \frac{n-1}{n}DX.$$

Để kỳ vọng của phương sai mẫu trùng với phương sai của biến ngẫu nhiên gốc ta cần hiệu chỉnh như sau.

- Phương sai mẫu có hiệu chỉnh S^2

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

- Áp dụng các công thức tính kỳ vọng và phương sai của tổng các biến ngẫu nhiên độc lập ta có

$$ES^2 = DX.$$

2.3. Độ lệch tiêu chuẩn

Độ lệch tiêu chuẩn

Độ lệch tiêu chuẩn S

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

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - **3. Cách tính giá trị của các đặc trưng mẫu**
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - 1. Phương pháp ước lượng bằng khoảng tin cậy
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

3. Cách tính giá trị của các thống kê thông dụng qua mẫu cụ thể (đặc trưng mẫu)

Cách tính giá trị của các đặc trưng mẫu

- Nếu mẫu được cho bởi bảng giá trị

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \tag{1}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} n_i x_i\right)^2\right] \tag{2}$$

3. Cách tính giá trị của các thống kê thông dụng qua mẫu cụ thể (đặc trưng mẫu)

Lập bảng tính toán

x_i	n_i	$n_i x_i$	$n_i x_i^2$
x_1	n_1	$n_1 x_i$	$n_1 x_1^2$
...	...	...	...
x_k	n_k	$n_k x_k$	$n_k x_k^2$
(1)	$n = \sum_{i=1}^{k} n_i$ (2)	$\sum_{i=1}^{k} n_i x_i$ (3)	$\sum_{i=1}^{k} n_i x_i^2$ (4)

Khi đó

$$\bar{x} = \frac{(3)}{(2)}; \quad s^2 = \frac{1}{n-1}\left((4) - \frac{(3)^2}{(2)}\right) \tag{3}$$

3. Cách tính giá trị của các đặc trưng mẫu

Ví dụ

Số xe hơi bán được trung bình trong một tuần ở mỗi đại lý trong 45 đại lý cho bởi

Số xe hơi bán	1	2	3	4	5	6
Số đại lý	15	12	9	5	3	1

Tính $\bar{x}, s$.

Lời giải

Lập bảng tính toán:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
1	15	15	15
2	12	24	28
3	9	27	81
4	5	20	80
5	3	15	75
6	1	6	36
$\sum$	$n = 45$	107	335

Khi đó

$$\bar{x} = \frac{107}{45} \simeq 2,38; \quad s^2 = \frac{1}{44}\left(335 - \frac{107^2}{45}\right) = \frac{3626}{1980};$$

$$s = \sqrt{s^2} \simeq 1,35.$$

3. Cách tính giá trị của các đặc trưng mẫu

- Nếu mẫu được cho bởi bảng giá trị

Giá trị	$(a_0 - a_1]$	$(a_1 - a_2]$	$\ldots$	$(a_{k-1} - a_k]$
Tần số	n_1	n_2	$\ldots$	n_k

khi đó

$$\bar{x} = x_0 + h\bar{u} = x_0 + \frac{h}{n}\sum_{i=1}^{k} n_i u_i \tag{4}$$

$$s^2 = h^2 s_u^2 = \frac{h^2}{n-1}\left[\sum_{i=1}^{k} n_i u_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} n_i u_i\right)^2\right] \tag{5}$$

trong đó, $u_i = \dfrac{x_i - x_0}{h}$. Ta thường chọn x_i là các điểm giữa của khoảng thứ i, $x_0 = x_i$ ứng với n_i lớn nhất, h là độ rộng của các khoảng.

3. Cách tính giá trị của các đặc trưng mẫu

Lập bảng tính toán

$a_i - a_{i-1}$	x_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$
$a_0 - a_1$	x_1	u_1	n_1	$n_1 x_i$	$n_1 x_1^2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$a_k - a_{k-1}$	x_k	u_k	n_k	$n_k x_k$	$n_k x_k^2$
$\sum$			$n = \sum_{i=1}^{k} n_i \quad (4)$	$\sum_{i=1}^{k} n_i x_i \quad (5)$	$\sum_{i=1}^{k} n_i x_i^2 \quad (6)$

Khi đó

$$\bar{u} = \frac{(5)}{(4)}; \quad s_u^2 = \frac{1}{n-1}\left((6) - \frac{(5)^2}{(4)}\right) \tag{6}$$

3. Cách tính giá trị của các đặc trưng mẫu

Ví dụ

Cân 25 bao gạo ta thu được các số liệu cho ở bảng sau

Trọng lượng	48-48,5	48,5 - 49	49-49,5	49,5 - 50	50-50,5
Số bao	2	5	10	6	2

Tính $\bar{x}, s$.

Lời giải

Ta lập bảng tính toán sau:

$a_i - a_{i-1}$	x_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$
48-48,5	48,25	-2	2	-4	8
48,5-49	48,75	-1	5	-5	5
49-49,5	49,25	0	10	0	0
49,5-50	49,75	1	6	6	6
50-50,5	50,25	2	2	4	8
$\sum$			25	1	27

Khi đó

$$\bar{x} = x_0 + h\bar{u} = 49,25 + 0,5\frac{1}{25} = 49,27;$$

$$s^2 = h^2 s_u^2 = \frac{(0,5)^2}{24}\left(27 - \frac{1^2}{25}\right) = \frac{674}{2400}.$$

$$s = \sqrt{s^2} \simeq 0,53.$$

Nội dung

1. **BÀI 1. TỔNG THỂ VÀ MẪU**
   - 1. Tổng thể
   - 2. Mẫu
   - 3. Mô hình hóa mẫu ngẫu nhiên
   - 4. Một số cách chọn mẫu cơ bản
   - 5. Một số phương pháp mô tả giá trị của mẫu ngẫu nhiên
2. **BÀI 2. THỐNG KÊ VÀ CÁC ĐẶC TRƯNG CỦA MẪU NGẪU NHIÊN**
   - 1. Định nghĩa thống kê
   - 2. Các đặc trưng của mẫu ngẫu nhiên
   - 3. Cách tính giá trị của các đặc trưng mẫu
3. **BÀI 3. ƯỚC LƯỢNG THAM SỐ**
   - **1. Phương pháp ước lượng bằng khoảng tin cậy**
   - 2. Khoảng tin cậy cho kỳ vọng
   - 3. Ước lượng tỷ lệ

1. Phương pháp ước lượng bằng khoảng tin cậy

Ước lượng khoảng

Giả sử chưa biết đặc trưng θ nào đó của biến ngẫu nhiên X. Ước lượng khoảng của θ là chỉ ra một khoảng số (g_1, g_2) nào đó chứa θ, tức là có thể ước lượng $g_1 \le \theta \le g_2$.

Phương pháp ước lượng bằng khoảng tin cậy

Để ước lượng khoảng tin cậy của tham số θ của biến ngẫu nhiên X, từ biến ngẫu nhiên này ta lập mẫu $W_X = (X_1, X_2, \ldots, X_n)$ cỡ n. Chọn thống kê $G(X, \theta)$ sao cho mặc dù chưa biết giá trị của θ thì quy luật phân phối của G vẫn hoàn toàn xác định. Do đó với xác suất α khá bé ta tìm được $P(g_1 \le \theta \le g_2) = 1 - \alpha$. Khi đó,

- (g_1, g_2) được gọi là khoảng tin cậy của θ với độ tin cậy $\beta = 1 - \alpha$;
- $\beta = 1 - \alpha$ được gọi là độ tin cậy của ước lượng;

Phương pháp ước lượng bằng khoảng tin cậy (tiếp theo)

- Do α khá nhỏ nên $\beta = 1 - \alpha$ khá lớn (thông thường yêu cầu $\beta = 1 - \alpha \geq 0,95$). Khi đó, biến cố $(g_1 \leq \theta \leq g_2)$ hầu như chắc chắn xảy ra trong một phép thử.
- Thực hiện một phép thử đối với mẫu ngẫu nhiên W_X ta thu được mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, từ đó tính được các giá trị của g_1, g_2.
- Như vậy có thể kết luận với độ tin cậy $\beta = 1 - \alpha$ tham số θ nằm trong khoảng (g_1, g_2).

Nội dung

2. Khoảng tin cậy cho kỳ vọng

Bài toán

Giả sử biến ngẫu nhiên X tuân theo luật phân bố chuẩn $N(\mu, \delta^2)$ với kỳ vọng $E(X) = \mu$ chưa biết. Cần ước lượng $E(X)$. Ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ và xét các trường hợp sau:

- Trường hợp 1: Đã biết phương sai $DX = \delta^2$;
- Trường hợp 2: Chưa biết phương sai $DX = \delta^2$, cỡ mẫu $n \geq 30$;
- Trường hợp 3: Chưa biết phương sai $DX = \delta^2$, cỡ mẫu $n < 30$.

2.1. Trường hợp 1: Đã biết phương sai $DX = \delta^2$

- Chọn thống kê: $U = \dfrac{\bar{X} - \mu}{\delta}\sqrt{n}$. Khi đó $U \sim N(0, 1)$.
- Khoảng tin cậy của μ với độ tin cậy $\beta = 1 - \alpha$ là

$$(\bar{x} - \epsilon, \bar{x} + \epsilon)$$

trong đó, $\epsilon = u_{\frac{\alpha}{2}} \dfrac{\delta}{\sqrt{n}}$, với $u_{\frac{\alpha}{2}}$ là giá trị tới hạn chuẩn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc.

2.1. Trường hợp 1: Đã biết phương sai $DX = \delta^2$

Ví dụ

Khối lượng sản phẩm là biến ngẫu nhiên có phân bố chuẩn với độ lệch tiêu chuẩn $\delta = 1$. Cân 25 sản phẩm ta thu được kết quả sau:

Khối lượng	18	19	20	21
Số sản phẩm	3	5	15	2

Hãy ước lượng khối lượng trung bình của sản phẩm bằng khoảng tin cậy với độ tin cậy $\beta = 95\%$.

2.1. Trường hợp 1: Đã biết phương sai $DX = \delta^2$

Lời giải

Ta có $\bar{x} = \dfrac{18.3 + 19.5 + 20.15 + 21.2}{25} = 19,64.$

Độ tin cậy $\beta = 1 - \alpha = 0,95$, suy ra $\alpha = 0,05$, $\dfrac{\alpha}{2} = 0,025.$

Tra bảng ta được $u_{\frac{\alpha}{2}} = 1,96$. Độ chính xác của ước lượng

$$\epsilon = u_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} = 1,96 \frac{1}{\sqrt{25}} = 0,392.$$

Khoảng tin cậy cho khối lượng trung bình của sản phẩm

$$(\bar{x} - \epsilon, \bar{x} + \epsilon) = (19,248; 20,032).$$

2.2. Trường hợp 2: Chưa biết $DX = \delta^2$, $n \geq 30$

Trường hợp 2: Chưa biết phương sai $DX = \delta^2$, $n \geq 30$

- Do chưa biết δ nên ta thay thế bằng S. Chọn thống kê: $T = \dfrac{\bar{X} - \mu}{S}\sqrt{n}$. Khi đó $T \sim N(0, 1)$.
- Khoảng tin cậy của μ với độ tin cậy $\beta = 1 - \alpha$ là

$$(\bar{x} - \epsilon, \bar{x} + \epsilon)$$

trong đó, $\epsilon = u_{\frac{\alpha}{2}} \dfrac{s}{\sqrt{n}}$, với $u_{\frac{\alpha}{2}}$ là giá trị tới hạn chuẩn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc.

2.2. Trường hợp 2: Chưa biết $DX = \delta^2$, $n \geq 30$

Ví dụ

Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên có phân bố chuẩn. Cân thử 100 sản phẩm loại này ta thu được kết quả

Trọng lượng (g)	40-42	42-44	44-46	46-48	48 - 50	50 -52
Số sản phẩm	7	13	25	35	15	5

với độ tin cậy 95%, hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.

Giải

Thực hiện phép đổi biến $u_i = \dfrac{x_i - 47}{2}$ với $h = 2$ và $x_0 = 47$. Ta có bảng tính toán sau

$a_i - a_{i-1}$	x_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$
40-42	41	-3	7	-21	63
42-44	43	-2	13	-26	52
44-46	45	-1	25	-25	25
46-48	47	0	35	0	0
48-50	49	1	15	15	15
50-52	51	2	5	10	20

Khi đó

$$\bar{x} = x_0 + h\bar{u} = 47 + 2\frac{-47}{100} = 46,06 \tag{7}$$

$$s^2 = h^2 s_u^2 = 2^2 \frac{1}{99}\left(175 - \frac{(-47)^2}{100}\right) = \frac{15291}{2475}. \tag{8}$$

$$s = \sqrt{s^2} \simeq 2,49. \tag{9}$$

2.2. Trường hợp 2: Chưa biết $DX = \delta^2$, $n \geq 30$

Giải

Độ tin cậy 95%, suy ra $1 - \alpha = 0,95$ hay $\alpha = 0,5$. Khi đó $\dfrac{\alpha}{2} = 0,025$.

Tra bảng ta được $u_{\frac{\alpha}{2}} = 1,96$.

Độ chính xác của ước lượng là

$$\epsilon = u_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 1,96 \frac{2,49}{10} = 0,49.$$

Khoảng tin cậy của trọng lượng trung bình của loại sản phẩm

$$(\bar{x} - \epsilon, \bar{x} + \epsilon) = (45,57; 46,55).$$

2.3. Trường hợp 3: Chưa biết $DX = \delta^2$, $n < 30$

Trường hợp 3: Chưa biết phương sai $DX = \delta^2$, $n < 30$

- Giống như TH 2, do chưa biết δ nên ta thay thế bằng S. Chọn thống kê: $T = \dfrac{\bar{X} - \mu}{S}\sqrt{n}$. Khi đó $T \sim T^{(n-1)}$.
- Khoảng tin cậy của μ với độ tin cậy $\beta = 1 - \alpha$ là

$$(\bar{x} - \epsilon, \bar{x} + \epsilon)$$

trong đó, $\epsilon = t_{\frac{\alpha}{2}}(n-1)\dfrac{s}{\sqrt{n}}$, với $t_{\frac{\alpha}{2}}(n-1)$ là giá trị tới hạn chuẩn mức $\dfrac{\alpha}{2}$ của phân bố Student với $n-1$ bậc tự do.

2.3. Trường hợp 3: Chưa biết $DX = \delta^2$, $n < 30$

Ví dụ

Để ước lượng tuổi thọ trung bình một loại sản phẩm người ta chọn ra 26 sản phẩm và thu được kết quả sau:

Tuổi thọ (giờ)	190	195	198	200	204	205
Số sản phẩm	5	4	2	8	6	1

Giả sử tuổi thọ sản phẩm tuân theo phân phối chuẩn, hãy ước lượng tuổi thọ trung bình của sản phẩm trên với độ tin cậy 95%.

2.3. Trường hợp 3: Chưa biết $DX = \delta^2$, $n < 30$

Giải

Ta lập bảng tính toán và tính được $\bar{x} = 198,27$ và $s = 5,103$.
Độ tin cậy 95%, suy ra $1 - \alpha = 0,95$ hay $\alpha = 0,5$. Khi đó $\dfrac{\alpha}{2} = 0,025$.
Tra bảng ta được $t_{\frac{\alpha}{2}}(25) = 2,06$.
Độ chính xác của ước lượng là

$$\epsilon = t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} = 2,06\frac{5,103}{\sqrt{26}} \simeq 2,06.$$

Khoảng tin cậy về tuổi thọ trung bình của sản phẩm là

$$(\bar{x} - \epsilon, \bar{x} + \epsilon) = (196,21; 200,33).$$

Nội dung

3. Ước lượng tỷ lệ

- Xác suất xảy ra biến cố A của tổng thể là p. Thông thường không biết p. Cần ước lượng khoảng tin cậy của p.
- Thực hiện n phép thử độc lập, cùng điều kiện, trong đó có m phép thử xảy ra A. Khi đó tần số xuất hiện A: $f = \dfrac{m}{n}$. Khi đó khoảng tin cậy cho tỷ lệ phần tử mang tính chất A với độ tin cậy $\beta = 1 - \alpha$ là
$$(f - \epsilon, f + \epsilon)$$
trong đó, $\epsilon = u_{\frac{\alpha}{2}} \sqrt{\dfrac{f(1-f)}{n}}$, với $u_{\frac{\alpha}{2}}$ là giá trị tới hạn chuẩn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc.
- Khoảng tin cậy được xây dựng khi kích thước mẫu đủ lớn thoả mãn $nf \geq 10$ và $n(1-f) \geq 10$.

3. Ước lượng tỷ lệ

Ví dụ

Người ta lấy ngẫu nhiên từ một lô hàng ra 200 sản phẩm thì thấy có 182 sản phẩm đạt yêu cầu chất lượng.
a) Với độ tin cậy 95%, hãy ước lượng tỷ lệ sản phẩm đạt yêu cầu chất lượng của lô hàng.
b) Giả sử lô hàng có 6000 sản phẩm, với độ tin cậy 95%, hãy ước lượng số sản phẩm đạt yêu cầu của lô hàng.

3. Ước lượng tỷ lệ

Lời giải

a) Ta có $n = 200,\ f = \dfrac{182}{200} = 0,91.$
Kiểm tra điều kiện của n và f: $nf = 182 > 10$ và $n(1-f) = 18 > 10$ (thoả mãn). Vì độ tin cậy 95%, suy ra $1 - \alpha = 0,95$ hay $\alpha = 0,5$. Khi đó $\dfrac{\alpha}{2} = 0,025$. Tra bảng ta được $u_{0,025} = 1,96$.
Độ chính xác của ước lượng là

$$\epsilon = u_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} = 1,96\sqrt{\frac{0,91(1-0,91)}{200}} \simeq 0,04.$$

Khoảng tin cậy cho tỷ lệ sản phẩm đạt yêu cầu chất lượng của lô hàng là

$$(f-\epsilon, f+\epsilon) = (0,87; 0,95).$$

3. Ước lượng tỷ lệ

Lời giải

b) Gọi N là sản phẩm đạt yêu cầu của cả lô hàng. Khi đó tỷ lệ sản phẩm đạt yêu cầu chất lượng của lô hàng là $\dfrac{N}{6000}$. Theo câu a) ta có

$$0,87 < \frac{N}{6000} < 0,95 \Leftrightarrow 5220 < N < 5700.$$