Bài giảng Xác suất Thống kê và ứng dụng

Nguyễn Thị Nhung

Bộ môn Toán - Đại học THĂNG LONG

Ngày 15 tháng 8 năm 2013

Chương VI

Phân phối của các tham số mẫu

Chương VI

1. Tham số mẫu: định nghĩa và đặc trưng cơ bản
   - Tham số tổng thể và tham số mẫu
   - Định nghĩa trung bình mẫu
   - Định nghĩa tỉ lệ mẫu
   - Định nghĩa phương sai mẫu

2. Phân phối của các tham số mẫu
   - Phân phối của trung bình mẫu
   - Phân phối của tỉ lệ mẫu

Chương VI

1. Tham số mẫu: định nghĩa và đặc trưng cơ bản
   - Tham số tổng thể và tham số mẫu
   - Định nghĩa trung bình mẫu
   - Định nghĩa tỉ lệ mẫu
   - Định nghĩa phương sai mẫu

2. Phân phối của các tham số mẫu
   - Phân phối của trung bình mẫu
   - Phân phối của tỉ lệ mẫu

Nội dung chính trong chương

- Giới thiệu những tham số tổng thể và tham số mẫu cơ bản;
- Giới thiệu về trung bình mẫu: định nghĩa, ví dụ, các đặc trưng cơ bản của trung bình mẫu (kì vọng và phương sai) và phân phối của trung bình mẫu;
- Giới thiệu về tỉ lệ mẫu: định nghĩa, ví dụ, các đặc trưng cơ bản của tỉ lệ mẫu (kì vọng và phương sai) và phân phối của tỉ lệ mẫu;
- Giới thiệu về phương sai mẫu: định nghĩa, ví dụ và hai đặc trưng cơ bản của phương sai mẫu là kì vọng và phương sai.

Yêu cầu đối với sinh viên

- Biết được những tham số tổng thể và tham số mẫu cơ bản được đề cập trong chương trình;
- Nắm được định nghĩa, ví dụ, các đặc trưng cơ bản của trung bình mẫu (kì vọng và phương sai) và phân phối của trung bình mẫu;
- Nắm được định nghĩa, ví dụ, các đặc trưng cơ bản của tỉ lệ mẫu (kì vọng và phương sai) và phân phối của tỉ lệ mẫu;
- Nắm được định nghĩa, ví dụ và hai đặc trưng cơ bản là kì vọng và phương sai của phương sai mẫu.

Bài toán tình huống

Câu hỏi tình huống

Giả sử bạn là một thành viên trong Văn phòng Đoàn của trường Đại học Thăng Long. Do nhiều sinh viên trong trường nhận xét rằng điểm thi môn XSTK rất thấp, tỉ lệ những sinh viên thi đỗ chưa đến 50% nên bạn muốn tìm hiểu xem điểm thi trung bình môn XSTK của sinh viên Thăng Long trong vài năm gần đây là bao nhiêu và tỉ lệ thi qua của môn này là bao nhiêu để viết bài đưa lên mạng.
Bạn phải làm thế nào để giải quyết yêu cầu này?

Bài toán tình huống

Câu hỏi tình huống

- *Bạn không điều tra được toàn bộ điểm XSTK của các sinh viên mà chỉ điều tra được điểm của mẫu gồm vài trăm sinh viên trong vài năm gần đây.*
- *Công thức nào trên mẫu giúp ước tính điểm thi trung bình, tỉ lệ thi đỗ của toàn bộ các sinh viên?*
- *Những công thức này có ưu điểm gì mà bạn lại lựa chọn sử dụng?*
- *Điểm thi của các mẫu sinh viên khác nhau là khác nhau và do đó điểm trung bình, tỉ lệ thi đỗ cũng thay đổi theo mẫu. Tuy những đại lượng này thay đổi nhưng có qui luật gì hay không?*

Bài toán tình huống

Câu hỏi tình huống

- *Bạn không điều tra được toàn bộ điểm XSTK của các sinh viên mà chỉ điều tra được điểm của mẫu gồm vài trăm sinh viên trong vài năm gần đây.*
- *Công thức nào trên mẫu giúp ước tính điểm thi trung bình, tỉ lệ thi đỗ của toàn bộ các sinh viên?*
- *Những công thức này có ưu điểm gì mà bạn lại lựa chọn sử dụng?*
- *Điểm thi của các mẫu sinh viên khác nhau là khác nhau và do đó điểm trung bình, tỉ lệ thi đỗ cũng thay đổi theo mẫu. Tuy những đại lượng này thay đổi nhưng có qui luật gì hay không?*

Câu hỏi tình huống

- *Bạn không điều tra được toàn bộ điểm XSTK của các sinh viên mà chỉ điều tra được điểm của mẫu gồm vài trăm sinh viên trong vài năm gần đây.*
- *Công thức nào trên mẫu giúp ước tính điểm thi trung bình, tỉ lệ thi đỗ của toàn bộ các sinh viên?*
- *Những công thức này có ưu điểm gì mà bạn lại lựa chọn sử dụng?*
- *Điểm thi của các mẫu sinh viên khác nhau là khác nhau và do đó điểm trung bình, tỉ lệ thi đỗ cũng thay đổi theo mẫu. Tuy những đại lượng này thay đổi nhưng có qui luật gì hay không?*

Câu hỏi tình huống

- *Bạn không điều tra được toàn bộ điểm XSTK của các sinh viên mà chỉ điều tra được điểm của mẫu gồm vài trăm sinh viên trong vài năm gần đây.*
- *Công thức nào trên mẫu giúp ước tính điểm thi trung bình, tỉ lệ thi đỗ của toàn bộ các sinh viên?*
- *Những công thức này có ưu điểm gì mà bạn lại lựa chọn sử dụng?*
- *Điểm thi của các mẫu sinh viên khác nhau là khác nhau và do đó điểm trung bình, tỉ lệ thi đỗ cũng thay đổi theo mẫu. Tuy những đại lượng này thay đổi nhưng có qui luật gì hay không?*

Nội dung trình bày

1. Tham số mẫu: định nghĩa và đặc trưng cơ bản
   - Tham số tổng thể và tham số mẫu
   - Định nghĩa trung bình mẫu
   - Định nghĩa tỉ lệ mẫu
   - Định nghĩa phương sai mẫu

2. Phân phối của các tham số mẫu
   - Phân phối của trung bình mẫu
   - Phân phối của tỉ lệ mẫu

Tham số tổng thể và tham số mẫu

- Tham số tổng thể là đặc trưng của tổng thể dùng để mô tả những đặc tính của tổng thể như: Trung bình, trung vị, mode, phương sai, độ lệch chuẩn,...
- Tham số mẫu là đặc trưng của mẫu dùng để mô tả những đặc tính của mẫu như: Trung bình mẫu, phương sai mẫu, độ lệch chuẩn mẫu,...

Tham số tổng thể và tham số mẫu

- Tham số tổng thể là đặc trưng của tổng thể dùng để mô tả những đặc tính của tổng thể như: Trung bình, trung vị, mode, phương sai, độ lệch chuẩn,...
- Tham số mẫu là đặc trưng của mẫu dùng để mô tả những đặc tính của mẫu như: Trung bình mẫu, phương sai mẫu, độ lệch chuẩn mẫu,...

Tham số tổng thể và tham số mẫu

Trong thống kê suy diễn ta dùng các tham số mẫu để đưa ra những ước lượng về các tham số của tổng thể:

- Trung bình tổng thể được ước lượng từ trung bình mẫu;
- Phương sai tổng thể được ước lượng từ phương sai mẫu;
- Tỉ lệ tổng thể được ước lượng từ tỉ lệ mẫu.

Nội dung trình bày

1. Tham số mẫu: định nghĩa và đặc trưng cơ bản
   - Tham số tổng thể và tham số mẫu
   - Định nghĩa trung bình mẫu
   - Định nghĩa tỉ lệ mẫu
   - Định nghĩa phương sai mẫu
2. Phân phối của các tham số mẫu
   - Phân phối của trung bình mẫu
   - Phân phối của tỉ lệ mẫu

Ví dụ tính trung bình mẫu

Giả sử ta có một tổng thể nhỏ gồm $N = 8$ phần tử:

$$54, 55, 59, 63, 64, 68, 69, 70.$$

Chọn mọi mẫu (có hoàn lại) gồm $n = 2$ phần tử từ tổng thể trên, ta được tất cả 64 mẫu. Bảng sau liệt kê 64 mẫu cùng với trung bình mẫu tương ứng:

Ví dụ tính trung bình mẫu

Mẫu	TBMẫu	Mẫu	TBMẫu	Mẫu	TBMẫu	Mẫu	TBMẫu
(54,54)	54.0	(59,54)	56.5	(64,54)	59.0	(69,54)	61.5
(54,55)	54.5	(59,55)	57.0	(64,55)	59.5	(69,55)	62.0
(54,59)	56.5	(59,59)	59.0	(64,59)	61.5	(69,59)	64.0
(54,63)	58.5	(59,63)	61.0	(64,63)	63.5	(69,63)	66.0
(54,64)	59.0	(59,64)	61.5	(64,64)	64.0	(69,64)	66.5
(54,68)	61.0	(59,68)	63.5	(64,68)	66.0	(69,68)	68.5
(54,69)	61.5	(59,69)	64.0	(64,69)	66.5	(69,69)	69.0
(54,70)	62.0	(59,70)	64.5	(64,70)	67.0	(69,70)	69.5
(55,54)	54.5	(63,54)	58.5	(68,54)	61.0	(70,54)	62.0
(55,55)	55.0	(63,55)	59.0	(68,55)	61.5	(70,55)	62.5
(55,59)	57.0	(63,59)	61.0	(68,59)	63.5	(70,59)	64.5
(55,63)	59.0	(63,63)	63.0	(68,63)	65.5	(70,63)	66.5
(55,64)	59.5	(63,64)	63.5	(68,64)	66.0	(70,64)	67.0
(55,68)	61.5	(63,68)	65.5	(68,68)	68.0	(70,68)	69.0
(55,69)	62.0	(63,69)	66.0	(68,69)	68.5	(70,69)	69.5
(55,70)	62.5	(63,70)	66.5	(68,70)	69.0	(70,70)	70.0

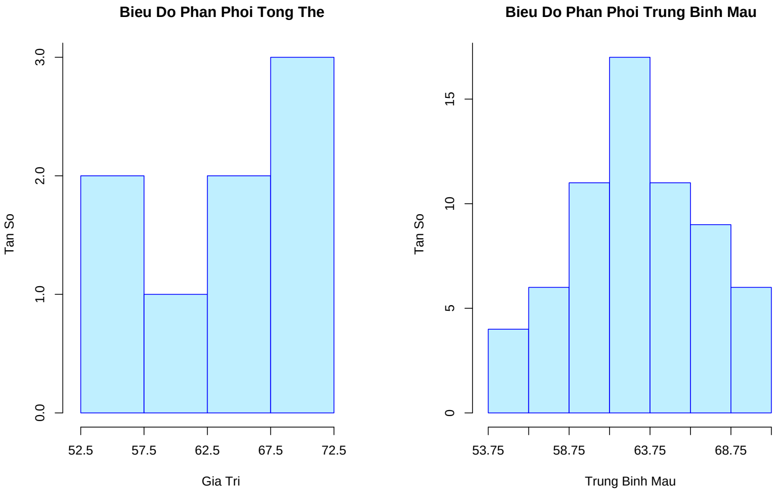
Ví dụ tính trung bình mẫu

Các giá trị trung bình của 64 mẫu gồm 2 phần tử được chọn từ tổng thể có phân phối xác suất cho trong bảng dưới đây:

Giá trị	54	54.5	55	56.5	57	58.5	59	59.5	61	61.5	62	62.5
Xác suất	$\frac{1}{64}$	$\frac{2}{64}$	$\frac{1}{64}$	$\frac{2}{64}$	$\frac{2}{64}$	$\frac{2}{64}$	$\frac{5}{64}$	$\frac{2}{64}$	$\frac{4}{64}$	$\frac{6}{64}$	$\frac{4}{64}$	$\frac{2}{64}$

Giá trị	63	63.5	64	64.5	65.5	66	66.5	67	68	68.5	69	69.5	70
Xác suất	$\frac{1}{64}$	$\frac{4}{64}$	$\frac{3}{64}$	$\frac{2}{64}$	$\frac{2}{64}$	$\frac{4}{64}$	$\frac{4}{64}$	$\frac{2}{64}$	$\frac{1}{64}$	$\frac{2}{64}$	$\frac{3}{64}$	$\frac{2}{64}$	$\frac{1}{64}$

Minh họa hình học



Xây dựng định nghĩa trung bình mẫu

- Nêu cách tính trung bình của từng mẫu trong 64 mẫu ở ví dụ trên.
- Xây dựng công thức tính trung bình mẫu trong ví dụ trên.

Xây dựng định nghĩa trung bình mẫu

- Nêu cách tính trung bình của từng mẫu trong 64 mẫu ở ví dụ trên.
- Xây dựng công thức tính trung bình mẫu trong ví dụ trên.

Định nghĩa trung bình mẫu

Định nghĩa

Giả sử ta chọn một mẫu ngẫu nhiên gồm n phần tử từ tổng thể. Gọi $X_i, i = \overline{1, n}$ tương ứng là các biến ngẫu nhiên chỉ phần tử thứ i trong mẫu. Khi đó trung bình mẫu, kí hiệu là $\overline{X}$, là biến ngẫu nhiên xác định bởi công thức:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Nhận xét: Nếu tổng thể có trung bình μ và phương sai σ^2 thì $X_1, X_2, \ldots, X_n$ cùng có trung bình μ và phương sai σ^2.

Định nghĩa trung bình mẫu

Định nghĩa

Giả sử ta chọn một mẫu ngẫu nhiên gồm n phần tử từ tổng thể. Gọi $X_i, i = \overline{1, n}$ tương ứng là các biến ngẫu nhiên chỉ phần tử thứ i trong mẫu. Khi đó trung bình mẫu, kí hiệu là $\overline{X}$, là biến ngẫu nhiên xác định bởi công thức:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Nhận xét: Nếu tổng thể có trung bình μ và phương sai σ^2 thì $X_1, X_2, \ldots, X_n$ cùng có trung bình μ và phương sai σ^2.

Kì vọng của trung bình mẫu

Ta có

$$E(\overline{X}) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n)$$
$$= \frac{1}{n}(E(X_1) + E(X_2) + \ldots + E(X_n))$$
$$= \frac{1}{n}(n\mu) = \mu.$$

Như vậy, kì vọng của biến ngẫu nhiên trung bình mẫu bằng đúng trung bình tổng thể.

Ví dụ: Trong ví dụ trên ta có

$$\mu_X = \frac{54 + 55 + 59 + 63 + 64 + 68 + 69 + 70}{8} = 62.75.$$

và kì vọng của trung bình mẫu $\overline{X}$ là:

$$E(\overline{X}) = 54 \times \frac{1}{64} + 54.5 \times \frac{2}{64} + \ldots + 69.5\frac{2}{64} + 70 \times \frac{1}{64} = 62.75.$$

Như vậy $E(\overline{X}) = \mu$.

Độ lệch chuẩn của trung bình mẫu

- Khi lấy mẫu có lặp lại hoặc không lặp lại nhưng số phần tử của mẫu n rất nhỏ so với số phần tử của tổng thể N ($n < 0.05N$), các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ là độc lập. Ta có

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n) = n\sigma^2.$$

Do đó

$$V(\overline{X}) = \frac{1}{n^2} V(X_1 + X_2 + \ldots + X_n) = \frac{\sigma^2}{n}.$$

Vậy độ lệch chuẩn của $\overline{X}$ được cho bởi: $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

Nhận xét:

- Phương sai (độ lệch chuẩn) của trung bình mẫu nhỏ hơn phương sai (độ lệch chuẩn) của tổng thể. Như vậy, các giá trị của trung bình mẫu ít phân tán hơn các giá trị của tổng thể.
- Khi cỡ mẫu tăng lên, phương sai (độ lệch chuẩn) của trung bình mẫu giảm đi, điều này có nghĩa là sẽ có một khả năng lớn hơn các trung bình mẫu lấy được có giá trị gần đúng với trung bình thực của tổng thể.

Ví dụ

Ví dụ: Lấy lại các số liệu đầu trong việc chọn có hoàn lại 64 mẫu gồm n = 2 phần tử từ tổng thể gồm N = 8 phần tử. Ta có

- Độ lệch chuẩn của tổng thể là

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (55 - 62.75)^2 + \ldots + (70 - 62.75)^2}{8}} = 5.82559.$$

- Độ lệch chuẩn của trung bình mẫu là

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum(\bar{x}_i - \mu^2_{\overline{X}})}{64}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (54.5 - 62.75)^2 + \ldots + (70 - 62.75)^2}{64}} = 4.119314.$$

- Do chọn mẫu là có hoàn lại nên $\sigma_{\overline{X}} = 4.119314 = 5.82559/\sqrt{2} = \sigma/\sqrt{2}$.

Ví dụ

Ví dụ: Lấy lại các số liệu đầu trong việc chọn có hoàn lại 64 mẫu gồm n = 2 phần tử từ tổng thể gồm N = 8 phần tử. Ta có

- Độ lệch chuẩn của tổng thể là

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (55 - 62.75)^2 + \ldots + (70 - 62.75)^2}{8}} = 5.82559.$$

- Độ lệch chuẩn của trung bình mẫu là

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum(\bar{x}_i - \mu^2_{\overline{X}})}{64}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (54.5 - 62.75)^2 + \ldots + (70 - 62.75)^2}{64}} = 4.119314.$$

- Do chọn mẫu là có hoàn lại nên $\sigma_{\overline{X}} = 4.119314 = 5.82559/\sqrt{2} = \sigma/\sqrt{2}$.

Ví dụ

Ví dụ: Lấy lại các số liệu đầu trong việc chọn có hoàn lại 64 mẫu gồm n = 2 phần tử từ tổng thể gồm N = 8 phần tử. Ta có

- Độ lệch chuẩn của tổng thể là

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (55 - 62.75)^2 + \ldots + (70 - 62.75)^2}{8}} = 5.82559.$$

- Độ lệch chuẩn của trung bình mẫu là

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum(\bar{x}_i - \mu_{\overline{X}}^2)}{64}}$$
$$= \sqrt{\frac{(54 - 62.75)^2 + (54.5 - 62.75)^2 + \ldots + (70 - 62.75)^2}{64}} = 4.119314.$$

- Do chọn mẫu là có hoàn lại nên $\sigma_{\overline{X}} = 4.119314 = 5.82559/\sqrt{2} = \sigma/\sqrt{2}$.

Độ lệch chuẩn của trung bình mẫu

- Khi chọn mẫu không lặp lại từ một tổng thể và số phần tử của mẫu n không quá nhỏ so với số phần tử của tổng thể N ($n > 0.05N$), các biến ngẫu nhiên $X_1, X_2, \dots, X_n$ không độc lập với nhau nữa. Trong trường hợp này ta phải sử dụng thừa số điều chỉnh tổng thể hữu hạn: $FPC = \sqrt{\dfrac{N-n}{N-1}}$ nhân thêm vào độ lệch chuẩn của trung bình mẫu:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}.$$

Nhận xét: Do $FPC < 1$ nên yếu tố hiệu chỉnh sẽ làm độ lệch chuẩn của trung bình mẫu nhỏ hơn, điều này làm tăng khả năng lớn hơn các trung bình mẫu lấy được có giá trị gần đúng với trung bình thực của tổng thể, tức là phản ánh tổng thể tốt hơn so với trường hợp mẫu nhỏ.

Độ lệch chuẩn của trung bình mẫu

- Khi chọn mẫu không lặp lại từ một tổng thể và số phần tử của mẫu n không quá nhỏ so với số phần tử của tổng thể N ($n > 0.05N$), các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ không độc lập với nhau nữa. Trong trường hợp này ta phải sử dụng thừa số điều chỉnh tổng thể hữu hạn: $FPC = \sqrt{\dfrac{N-n}{N-1}}$ nhân thêm vào độ lệch chuẩn của trung bình mẫu:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}.$$

Nhận xét: Do $FPC < 1$ nên yếu tố hiệu chỉnh sẽ làm độ lệch chuẩn của trung bình mẫu nhỏ hơn, điều này làm tăng khả năng lớn hơn các trung bình mẫu lấy được có giá trị gần đúng với trung bình thực của tổng thể, tức là phản ánh tổng thể tốt hơn so với trường hợp mẫu nhỏ.

Câu hỏi ôn luyện

Câu hỏi: Giả sử tổng thể có trung bình μ và phương sai σ^2. Hãy nêu lại công thức tính $E(\overline{X})$ và $V(\overline{X})$

Nội dung trình bày

1. Tham số mẫu: định nghĩa và đặc trưng cơ bản
   - Tham số tổng thể và tham số mẫu
   - Định nghĩa trung bình mẫu
   - Định nghĩa tỉ lệ mẫu
   - Định nghĩa phương sai mẫu
2. Phân phối của các tham số mẫu
   - Phân phối của trung bình mẫu
   - Phân phối của tỉ lệ mẫu

Phân phối của tỉ lệ mẫu

Định nghĩa

Cho X là biến ngẫu nhiên chỉ số lần thành công trong n lần thử nghiệm độc lập với xác suất thành công trong mỗi lần thử là p. Khi đó tỉ lệ mẫu, kí hiệu là p_X *là biến ngẫu nhiên được xác định bằng công thức:* $P_X = \dfrac{X}{n}$.

Kì vọng của tỉ lệ mẫu

Do X tuân theo phân phối nhị thức nên $E(X) = np$ và $V(X) = np(1-p)$. Từ đó ta có

- $E(P_X) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}np = p$;
- Vậy kì vọng của tỉ lệ mẫu đúng bằng tỉ lệ tổng thể.

Kì vọng của tỉ lệ mẫu

Do X tuân theo phân phối nhị thức nên $E(X) = np$ và $V(X) = np(1-p)$. Từ đó ta có

- $E(P_X) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}np = p$;
- Vậy kì vọng của tỉ lệ mẫu đúng bằng tỉ lệ tổng thể.

Phương sai của tỉ lệ mẫu

- Khi lấy mẫu có lặp lại hoặc không lặp lại nhưng số phần tử của mẫu n rất nhỏ so với số phần tử của tổng thể N ($n < 0.05N$), ta có

$$V(P_X) = V(\frac{X}{n}) = \frac{1}{n^2}V(X) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

- Khi chọn mẫu không lặp lại từ một tổng thể và số phần tử của mẫu n rất không quá nhỏ so với số phần tử của tổng thể N ($n > 0.05N$), ta sử dụng thừa số điều chỉnh tổng thể hữu hạn: $FPC = \frac{N-n}{N-1}$ nhân thêm vào phương sai của tỉ lệ mẫu:

$$V(P_X) = \frac{p(1-p)}{n}\frac{N-n}{N-1}.$$

Phương sai của tỉ lệ mẫu

- Khi lấy mẫu có lặp lại hoặc không lặp lại nhưng số phần tử của mẫu n rất nhỏ so với số phần tử của tổng thể N ($n < 0.05N$), ta có

$$V(P_X) = V(\frac{X}{n}) = \frac{1}{n^2}V(X) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

- Khi chọn mẫu không lặp lại từ một tổng thể và số phần tử của mẫu n rất không quá nhỏ so với số phần tử của tổng thể N ($n > 0.05N$), ta sử dụng thừa số điều chỉnh tổng thể hữu hạn: $FPC = \dfrac{N-n}{N-1}$ nhân thêm vào phương sai của tỉ lệ mẫu:

$$V(P_X) = \frac{p(1-p)}{n}\frac{N-n}{N-1}.$$

Độ lệch chuẩn của tỉ lệ mẫu

- Độ lệch chuẩn của tỉ lệ mẫu σ_{P_X} là căn bậc hai của phương sai $V(P_X)$ và được cho bởi:
  - $\sigma_{P_X} = \sqrt{\dfrac{p(1-p)}{n}}$, nếu cỡ mẫu nhỏ so với tổng thể;
  - $\sigma_{P_X} = \sqrt{\dfrac{p(1-p)}{n}}\sqrt{\dfrac{N-n}{N-1}}$, nếu cỡ mẫu không nhỏ so với tổng thể.

Nội dung trình bày

Định nghĩa phương sai mẫu

Định nghĩa

Cho $\overline{X}$ *là trung bình mẫu của mẫu ngẫu nhiên* $(X_1, X_2, \ldots, X_n)$. *Phương sai mẫu, kí hiệu là* S_X^2 *là biến ngẫu nhiên được xác định bởi công thức:*

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

Kì vọng và phương sai của phương sai mẫu

Giả sử σ_X^2 là phương sai của tổng thể, khi đó

- Nếu cỡ mẫu n là nhỏ so với số phần tử N của tổng thể thì $E(S_X^2) = \sigma_X^2$;
- Nếu tổng thể tuân theo phân phối chuẩn thì $V(S_X^2) = \dfrac{2\sigma_X^4}{n-1}$.

Kì vọng và phương sai của phương sai mẫu

Giả sử σ_X^2 là phương sai của tổng thể, khi đó

- Nếu cỡ mẫu n là nhỏ so với số phần tử N của tổng thể thì $E(S_X^2) = \sigma_X^2$;
- Nếu tổng thể tuân theo phân phối chuẩn thì $V(S_X^2) = \dfrac{2\sigma_X^4}{n-1}$.

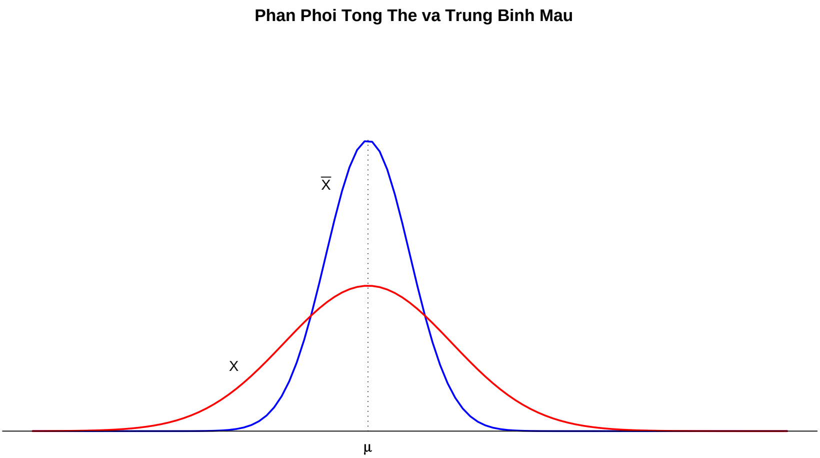
Nội dung trình bày

Chọn mẫu từ tổng thể có phân phối chuẩn

Định lí

Nếu mẫu được chọn ra từ tổng thể tuân theo phân phối chuẩn $N(\mu, \sigma^2)$ *thì trung bình mẫu* $\overline{X}$ *cũng là biến ngẫu nhiên tuân theo phân phối chuẩn* $N(\mu, \sigma^2_{\overline{X}})$.

Minh họa hình học



Chọn mẫu từ tổng thể không theo phân phối chuẩn

Định lí (Định lí giới hạn trung tâm)

Cho $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có cùng trung bình μ và phương sai hữu hạn σ^2. Khi đó với cỡ mẫu n đủ lớn thì phân phối của biến ngẫu nhiên

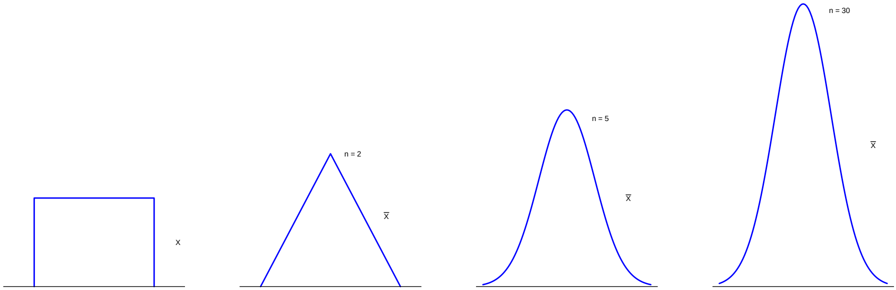
$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

xấp xỉ phân phối chuẩn với bất kì phân phối của tổng thể.

Hệ quả

Khi n đủ lớn, biến ngẫu nhiên trung bình mẫu $\overline{X}$ xấp xỉ phân phối chuẩn.

Minh họa định lí giới hạn trung tâm



Kết luận chung về phân phối của trung bình mẫu

- Nếu tổng thể có phân phối chuẩn thì phân phối của trung bình mẫu $\overline{X}$ cũng là phân phối chuẩn với mọi cỡ mẫu n;
- Khi kích thước mẫu khá lớn ($n \geqslant 30$) thì phân phối mẫu sẽ xấp xỉ phân phối chuẩn với bất kì phân phối của tổng thể;
- Nếu hình dáng của phân phối tổng thể là khá đối xứng, phân phối mẫu sẽ xấp xỉ phân phối chuẩn nếu cỡ mẫu $n \geqslant 15$.

Kết luận chung về phân phối của trung bình mẫu

- Nếu tổng thể có phân phối chuẩn thì phân phối của trung bình mẫu $\overline{X}$ cũng là phân phối chuẩn với mọi cỡ mẫu n;
- Khi kích thước mẫu khá lớn ($n \geqslant 30$) thì phân phối mẫu sẽ xấp xỉ phân phối chuẩn với bất kì phân phối của tổng thể;
- Nếu hình dáng của phân phối tổng thể là khá đối xứng, phân phối mẫu sẽ xấp xỉ phân phối chuẩn nếu cỡ mẫu $n \geqslant 15$.

Kết luận chung về phân phối của trung bình mẫu

- Nếu tổng thể có phân phối chuẩn thì phân phối của trung bình mẫu $\overline{X}$ cũng là phân phối chuẩn với mọi cỡ mẫu n;
- Khi kích thước mẫu khá lớn ($n \geqslant 30$) thì phân phối mẫu sẽ xấp xỉ phân phối chuẩn với bất kì phân phối của tổng thể;
- Nếu hình dáng của phân phối tổng thể là khá đối xứng, phân phối mẫu sẽ xấp xỉ phân phối chuẩn nếu cỡ mẫu $n \geqslant 15$.

Ví dụ

Bài toán

Một ngân hàng ước tính và nhận thấy rằng những tài khoản tiết kiệm cá nhân tuân theo phân phối chuẩn với trung bình 100 triệu và độ lệch chuẩn 15 triệu. Nếu ngân hàng chọn một mẫu ngẫu nhiên gồm 50 tài khoản, tính xác suất để trung bình mẫu sẽ nằm từ 98 triệu đến 103 triệu.

Lời giải

Ví dụ

Chú ý: Do $X \backsim N(100, (\frac{15}{\sqrt{50}})^2)$, nên xác suất $P(98 < \overline{X} < 103)$ được tính như sau:

```
> pnorm(103, 100, 15/sqrt(50)) - pnorm(98, 100, 15/sqrt(50))
0.7484611
```

Nội dung trình bày

Phân phối của tỉ lệ mẫu

Xét n lần phép thử và xác suất thành công trong mỗi lần thử là p. Gọi X là biến ngẫu nhiên chỉ số lần thành công trong n lần thử nghiệm, X_i là biến ngẫu nhiên chỉ số lần thành công ở lần thử thứ i thì $X = X_1 + X_2 + \ldots + X_n$ và $X_i, i = 1, \ldots, n$ là các biến ngẫu nhiên có cùng trung bình p và phương sai p(1-p). Theo định lí giới hạn trung tâm tỉ lệ mẫu

$$P_X = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

xấp xỉ phân phối chuẩn khi n đủ lớn.

Kết luận về phân phối của tỉ lệ mẫu

Định lí

Nếu $np \geqslant 5$ *và* $n(1-p) \geqslant 5$ *thì tỉ lệ mẫu* P_X *xấp xỉ phân phối chuẩn* $N(E(P_X), V(P_X))$.

Ví dụ

Bài toán

Theo trung tâm hỗ trợ sinh viên thì có 60% sinh viên hiện theo học đại học ở một thành phố muốn tìm việc làm ngoài giờ học. Một mẫu ngẫu nhiên gồm 205 sinh viên được chọn ra. Tính xác suất để trong mẫu có trên 70% sinh viên muốn tìm việc ngoài giờ học.

Lời giải

Ví dụ

Chú ý: Do $P_X \backsim N(0.6, \dfrac{0.6(1-0.6)}{205})$ nên để tính $P(P_X > 0.7)$ trong R ta thực hiện:

```
> 1 - pnorm(0.7, mean = 0.6, sd = sqrt(0.6*(1-0.6)/205))
0.001735538
```

Bài toán

Một cơ sở sản xuất bánh. Trọng lượng trung bình của một cái bánh là 20g với độ lệch chuẩn là 0.6g. Giả sử phân phối trọng lượng của bánh là chuẩn. Bạn mua 4 cái bánh (xem như một biến ngẫu nhiên gồm 4 phần tử)

a. *Tìm độ lệch chuẩn của trọng lượng trung bình mẫu.*
b. *Tìm xác suất để trọng lượng trung bình của 4 cái bánh này nhỏ hơn 19.7.*
c. *Tìm xác suất để trọng lượng trung bình của 4 cái bánh này lớn hơn 20.6.*
d. *Tìm xác suất để trọng lượng trung bình của 4 cái bánh này ở khoảng 19.5 đến 20.5.*
e. *Chọn ngẫu nhiên hai trong 4 chiếc bánh này. Tìm xác suất để trọng lượng trung bình của chúng ở trong khoảng từ 19.5g đến 20.5g.*

Bài tập ôn luyện

Bài toán

Theo sở vụ thuế, 75% số biên lai nộp thuế sẽ được khấu trừ (chẳng hạn 10%). Chọn ngẫu nhiên 100 biên lai thu thuế.

a. *Tìm trung bình của tỉ lệ mẫu của biên lai sẽ được khấu trừ.*
b. *Tìm phương sai và độ lệch chuẩn của tỉ lệ mẫu.*
c. *Tìm xác suất để tỉ lệ mẫu vượt 0.8.*
d. *Với xác suất 0.25 thì tỉ lệ mẫu thua tỉ lệ tổng thể bao nhiêu?*
e. *Với xác suất 0.25 thì tỉ lệ mẫu sai biệt tỉ lệ tổng thể bao nhiêu?*

Bài tập tự học

- Trình bày định nghĩa và phân phối của trung bình mẫu;
- Trình bày định nghĩa và phân phối của tỉ lệ mẫu;
- Định nghĩa về phương sai mẫu và các đặc trưng của phương sai mẫu.

Bài tập tự học

- Trình bày định nghĩa và phân phối của trung bình mẫu;
- Trình bày định nghĩa và phân phối của tỉ lệ mẫu;
- Định nghĩa về phương sai mẫu và các đặc trưng của phương sai mẫu.

Bài tập tự học

- Trình bày định nghĩa và phân phối của trung bình mẫu;
- Trình bày định nghĩa và phân phối của tỉ lệ mẫu;
- Định nghĩa về phương sai mẫu và các đặc trưng của phương sai mẫu.