Bài giảng Xác suất Thống kê và ứng dụng

Phan Thanh Hồng

Bộ môn Toán-Đại học THĂNG LONG

Ngày 10 tháng 9 năm 2009

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần VI

Phân phối của các tham số mẫu

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần VI

1. Phân phối của trung bình mẫu
2. Phân phối của tỷ lệ mẫu

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phần VI

1. Phân phối của trung bình mẫu

2. Phân phối của tỷ lệ mẫu

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu

Giả sử ta cần chọn một mẫu ngẫu nhiên cỡ n từ một tổng thể có trung bình μ và độ lệch chuẩn σ. Trước khi tiến hành chọn mẫu có thể xảy ra nhiều trường hợp khác nhau cho mẫu quan sát và có nhiều trường hợp xảy ra cho trung bình của mẫu quan sát được. Như vậy, trước khi tiến hành chọn mẫu, trung bình mẫu $\overline{X}$ là một biến ngẫu nhiên .

Định nghĩa

Phân phối chọn mẫu của trung bình mẫu $\overline{X}$ *là phân phối xác suất của tổng thể tất cả các trung bình mẫu có thể có cỡ* n.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu

Ví dụ: Xét tổng thể gồm 4 người mà tuổi của họ tạo thành một tổng thể các quan sát sau $\{18, 20, 22, 23\}$.

Tổng thể này có trung bình $\mu = \dfrac{18 + 20 + 22 + 23}{4} = 20.75.$

Giả sử ta chọn ra một mẫu gồm 2 người không lặp lại từ tổng thể trên, có 6 trường hợp như sau:

Mẫu	Trung bình mẫu
18,20	19
18,22	20
18,23	20.5
20,22	21
20,23	21.5
22,23	22.5

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu

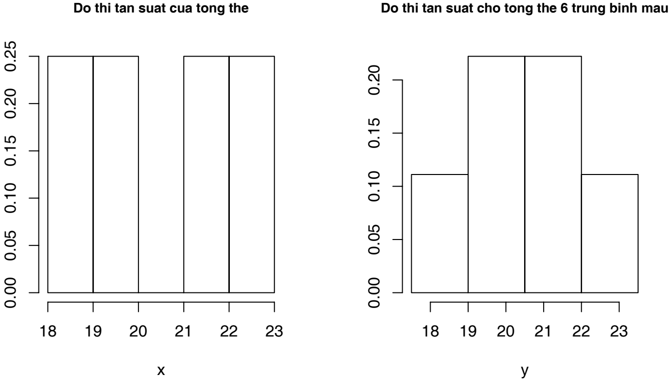
Ta có trung bình của các trung bình mẫu

$$\mu_{\overline{X}} = \frac{19 + 20 + 20.5 + 21 + 21.5 + 22.5}{6} = 20.75$$

giá trị này đúng bằng trung bình μ của tổng thể.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu



Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Độ lệch chuẩn của của trung bình mẫu

Người ta chứng minh được rằng trong trường hợp tổng thể ban đầu có vô hạn phần tử thì độ lệch chuẩn của trung bình mẫu $\sigma_{\overline{X}}$ có giá trị bằng $\sigma/\sqrt{n}$.
Nếu không lưu ý gì phần này ta chỉ xét những tổng thể có vô hạn phần tử.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

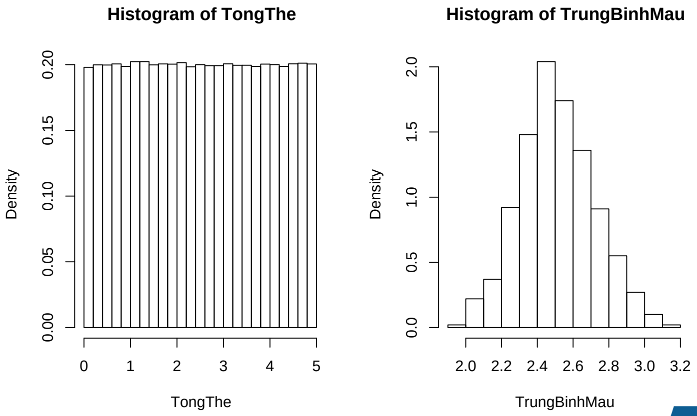
Phân phối của trung bình mẫu

Cho tổng thể có trung bình μ và độ lệch chuẩn σ. Khi đó tổng thể gồm tất cả các trung bình mẫu cỡ n rút ra từ tổng thể trên:

1. Có phân phối chuẩn nếu tổng thể ban đầu có phân phối chuẩn.
2. Có trung bình $\mu_{\overline{X}} = \mu$
3. Có độ lệch chuẩn $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ và phương sai $\sigma^2_{\overline{X}} = \sigma^2/n$ (Công thức này chính xác khi tổng thể ban đầu có vô hạn phần tử, và gần chính xác khi số phần tử của tổng thể này gấp trên 20 lần cỡ mẫu n).

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu



Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của trung bình mẫu

Ví dụ: Cho một tổng thể có phân phối chuẩn với trung bình $\mu = 45$ và độ lệch chuẩn $\sigma = 8$. Tính xác suất để $\overline{X}$ của một mẫu cỡ 25 rút từ tổng thể sai khác trung bình tổng thể không quá 4 đơn vị.

Ta cần tính $P(-4 < \overline{X} - \mu < 4)$.

Ta có

$$\begin{aligned} P(-4 < \overline{X} - \mu < 4) &= P\left(\frac{-4}{\sigma/\sqrt{n}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{4}{\sigma/\sqrt{n}}\right) \\ &= P\left(\frac{-4}{8/5} < \frac{\overline{X} - \mu}{8/5} < \frac{4}{8/5}\right) \\ &= P(-2.5 < Z < 2.5) = 0.988 \qquad (1) \end{aligned}$$

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Ví dụ: Cho một tổng thể có phân phối chuẩn với trung bình $\mu = 45$ và độ lệch chuẩn $\sigma = 8$. Tính xác suất để $\overline{X}$ của một mẫu cỡ 25 rút từ tổng thể sai khác trung bình tổng thể không quá 4 đơn vị.
Ta cần tính $P(-4 < \overline{X} - \mu < 4)$.
Ta có

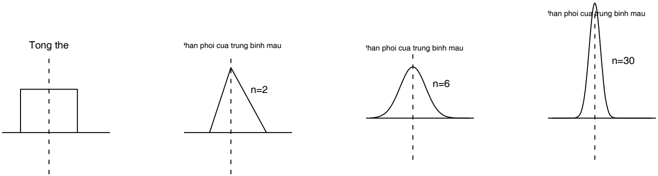
$$\begin{aligned} P(-4 < \overline{X} - \mu < 4) &= P\left(\frac{-4}{\sigma/\sqrt{n}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{4}{\sigma/\sqrt{n}}\right) \\ &= P\left(\frac{-4}{8/5} < \frac{\overline{X} - \mu}{8/5} < \frac{4}{8/5}\right) \\ &= P(-2.5 < Z < 2.5) = 0.988 \end{aligned} \tag{1}$$

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Định lý giới hạn trung tâm

Định lí

Nếu cỡ mẫu n đủ lớn thì phân phối của trung bình mẫu xấp xỉ phân phối chuẩn không kể tổng thể có phân phối nào.



Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Định lý giới hạn trung tâm

$n \geq 30$ được coi là đủ lớn, và n càng lớn phân phối của trung bình mẫu càng gần với phân phối chuẩn.

Ví dụ: Một mẫu nhẫu nhiên cỡ $n = 100$ rút ra từ tổng thể có trung bình 12.5 và độ lệch chuẩn 5.5. Tính xác suất để trung bình mẫu nằm trong phạm vi 12.25 đến 13.

Do $n = 100 \geq 30$ có thể áp dụng định lý giới hạn trung tâm cho biến ngẫu nhiên $\overline{X}$, với $\mu_{\overline{X}} = 12.5$, $\sigma_{\overline{X}} = 5.5/\sqrt{100} = 0.55$.

Ta cần tính xác suất $P(12.25 < \overline{X} < 13)$.

$$\begin{aligned} P(12.25 < \overline{X} < 13) &= P\left(\frac{12.25 - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{13 - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right) \\ &= P\left(\frac{12.25 - 12.5}{0.55} < Z < \frac{13 - 12.5}{0.55}\right) \\ &= P(-0.45 < Z < 0.91) = 0.4922 \end{aligned}$$

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Định lý giới hạn trung tâm

$n \geq 30$ được coi là đủ lớn, và n càng lớn phân phối của trung bình mẫu càng gần với phân phối chuẩn.

Ví dụ: Một mẫu nhẫu nhiên cỡ $n = 100$ rút ra từ tổng thể có trung bình 12.5 và độ lệch chuẩn 5.5. Tính xác suất để trung bình mẫu nằm trong phạm vi 12.25 đến 13.

Do $n = 100 \geq 30$ có thể áp dụng định lý giới hạn trung tâm cho biến ngẫu nhiên $\overline{X}$, với $\mu_{\overline{X}} = 12.5$, $\sigma_{\overline{X}} = 5.5/\sqrt{100} = 0.55$.

Ta cần tính xác suất $P(12.25 < \overline{X} < 13)$.

$$\begin{aligned}
P(12.25 < \overline{X} < 13) &= P\left(\frac{12.25 - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{13 - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right) \\
&= P\left(\frac{12.25 - 12.5}{0.55} < Z < \frac{13 - 12.5}{0.55}\right) \\
&= P(-0.45 < Z < 0.91) = 0.4922
\end{aligned}$$

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Ước lượng không chệch và ước lượng tốt nhất

Định nghĩa

Phân phối chọn mẫu của một thống kê mẫu là phân phối xác suất của tổng thể tất cả các giá trị quan sát được của thống kê mẫu đó.

Định nghĩa

Một thống kê mẫu là một ước lượng không chệch của một tham số tổng thể nếu trung bình của thống kê mẫu đó bằng chính trung bình tổng thể

Trung bình mẫu $\overline{X}$, phương sai mẫu S^2 là những ước lượng không chệch của trung bình và phương sai tổng thể, nhưng độ lệch chuẩn mẫu không phải là ước lượng không chệch của độ lệch chuẩn tổng thể.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Ước lượng không chệch và ước lượng tốt nhất

Định nghĩa

Phân phối chọn mẫu của một thống kê mẫu là phân phối xác suất của tổng thể tất cả các giá trị quan sát được của thống kê mẫu đó.

Định nghĩa

Một thống kê mẫu là một ước lượng không chệch của một tham số tổng thể nếu trung bình của thống kê mẫu đó bằng chính trung bình tổng thể

Trung bình mẫu $\overline{X}$, phương sai mẫu S^2 là những ước lượng không chệch của trung bình và phương sai tổng thể, nhưng độ lệch chuẩn mẫu không phải là ước lượng không chệch của độ lệch chuẩn tổng thể.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Định nghĩa

Phân phối chọn mẫu của một thống kê mẫu là phân phối xác suất của tổng thể tất cả các giá trị quan sát được của thống kê mẫu đó.

Định nghĩa

Một thống kê mẫu là một ước lượng không chệch của một tham số tổng thể nếu trung bình của thống kê mẫu đó bằng chính trung bình tổng thể

Trung bình mẫu $\overline{X}$, phương sai mẫu S^2 là những ước lượng không chệch của trung bình và phương sai tổng thể, nhưng độ lệch chuẩn mẫu không phải là ước lượng không chệch của độ lệch chuẩn tổng thể.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Có nhiều thống kê mẫu cùng là ước lượng không chệch cho một tham số tổng thể, trong đó thống kê mẫu có phương sai nhỏ nhất được gọi là ước lượng không chệch tốt nhất của tham số tổng thể.

Trung bình mẫu $\overline{X}$, phương sai mẫu S^2 là những ước lượng tốt nhất của trung bình và phương sai tổng thể.

Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Ước lượng không chệch và ước lượng tốt nhất

Có nhiều thống kê mẫu cùng là ước lượng không chệch cho một tham số tổng thể, trong đó thống kê mẫu có phương sai nhỏ nhất được gọi là ước lượng không chệch tốt nhất của tham số tổng thể.

Trung bình mẫu $\overline{X}$, phương sai mẫu S^2 là những ước lượng tốt nhất của trung bình và phương sai tổng thể.

Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Xét một tổng thể và ta quan tâm tới tỷ lệ số phần tử của tổng thể có một tính chất T nào đó, ký hiệu tỷ lệ này là p thì p là một tham số tổng thể. Trên thực tế khi số phần tử của tổng thể lớn khó có thể xác định p. Chúng ta sử dụng mẫu những quan sát để ước lượng cho p.

Định nghĩa

Với một mẫu cụ thể rút ra từ một tổng thể, tỷ lệ mẫu, ký hiệu là $\hat{p}$ được xác định bởi tỷ số $\frac{k}{n}$, trong đó n là cỡ mẫu, k là số quan sát trong mẫu có tính chất T.
Và $\hat{p}$ là một ước lượng điểm cho tỷ lệ tổng thể p.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Xét một tổng thể và ta quan tâm tới tỷ lệ số phần tử của tổng thể có một tính chất T nào đó, ký hiệu tỷ lệ này là p thì p là một tham số tổng thể. Trên thực tế khi số phần tử của tổng thể lớn khó có thể xác định p. Chúng ta sử dụng mẫu những quan sát để ước lượng cho p.

Định nghĩa

Với một mẫu cụ thể rút ra từ một tổng thể, tỷ lệ mẫu, ký hiệu là $\hat{p}$ được xác định bởi tỷ số $\frac{k}{n}$, trong đó n là cỡ mẫu, k là số quan sát trong mẫu có tính chất T.
Và $\hat{p}$ là một ước lượng điểm cho tỷ lệ tổng thể p.

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của tỷ lệ mẫu

Trước khi tiến hành chọn mẫu giá trị của $\hat{p}$ là chưa xác định, và thay đổi theo mẫu được chọn. Khi đó ta có biến ngẫu nhiên $\hat{P}$. Phân phối của $\hat{P}$ gọi là phân phối chọn mẫu của tỷ lệ mẫu.

Tổng thể tất cả các quan sát về tỷ lệ mẫu:

1. Có phân phối xấp xỉ chuẩn nếu cỡ mẫu n lớn.
2. Có trung bình $\mu_{\hat{P}} = p$
3. Có phương sai $\sigma^2_{\hat{P}} = \dfrac{p(1-p)}{n}$ (Công thức này chính xác khi tổng thể ban đầu có vô hạn phần tử, và gần chính xác khi số phần tử của tổng thể này gấp trên 20 lần cỡ mẫu n).

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Phân phối của tỷ lệ mẫu

Trước khi tiến hành chọn mẫu giá trị của $\hat{p}$ là chưa xác định, và thay đổi theo mẫu được chọn. Khi đó ta có biến ngẫu nhiên $\hat{P}$. Phân phối của $\hat{P}$ gọi là phân phối chọn mẫu của tỷ lệ mẫu.

Tổng thể tất cả các quan sát về tỷ lệ mẫu:

1. Có phân phối xấp xỉ chuẩn nếu cỡ mẫu n lớn.
2. Có trung bình $\mu_{\hat{P}} = p$
3. Có phương sai $\sigma^2_{\hat{P}} = \dfrac{p(1-p)}{n}$ (Công thức này chính xác khi tổng thể ban đầu có vô hạn phần tử, và gần chính xác khi số phần tử của tổng thể này gấp trên 20 lần cỡ mẫu n).

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Chú ý:

1. n được coi là lớn nếu np và $n(1-p)$ đều ≥ 5
2. $\hat{P}$ là một ước không chệch (và tốt nhất) của p

Đại học ThăngLong
Khương Trung, Thanh Xuân, Hà Nội
Tel.: 04.8587346 , Fax.: 04.5636775

Ví dụ: Theo sở điện lực, có khoảng 30% số nhà cũ không an toàn về dây điện. Một mẫu ngẫu nhiên gồm 250 nhà được chọn ra, tìm xác suất để tỷ lệ số nhà không an toàn về dây điện trong mẫu nằm trong phạm vi từ 25% đến 35%.

Ta có $p = 0.3$, $n = 250$, $\sigma^2_{\hat{P}} = \dfrac{p(1-p)}{n} = 0.00084$. Ta cần tính $P(0.25 < \hat{P} < 0.35)$. Do $np = 75$, $n(1-p) = 175 > 5$ nên phân phối của $\hat{P}$ xấp xỉ phân phối chuẩn, áp dụng phương pháp chuẩn hóa dữ liệu ta có

$$
\begin{aligned}
P(0.25 < \hat{P} < 0.35) &= P\Big(\frac{0.25 - p}{\sigma_{\hat{P}}} < \frac{\hat{P} - p}{\sigma_{\hat{P}}} < \frac{0.35 - p}{\sigma_{\hat{P}}}\Big) \\
&= P\Big(\frac{0.25 - 0.3}{\sqrt{0.00084}} < Z < \frac{0.35 - 0.3}{\sqrt{0.00084}}\Big) \\
&= P(-1,72 < Z < 1.72) = 0.9146
\end{aligned}
$$

Ví dụ: Theo sở điện lực, có khoảng 30% số nhà cũ không an toàn về dây điện. Một mẫu ngẫu nhiên gồm 250 nhà được chọn ra, tìm xác suất để tỷ lệ số nhà không an toàn về dây điện trong mẫu nằm trong phạm vi từ 25% đến 35%.

Ta có $p = 0.3$, $n = 250$, $\sigma^2_{\hat{P}} = \dfrac{p(1-p)}{n} = 0.00084$. Ta cần tính $P(0.25 < \hat{P} < 0.35)$. Do $np = 75$, $n(1-p) = 175 > 5$ nên phân phối của $\hat{P}$ xấp xỉ phân phối chuẩn, áp dụng phương pháp chuẩn hóa dữ liệu ta có

$$
\begin{aligned}
P(0.25 < \hat{P} < 0.35) &= P\left(\frac{0.25 - p}{\sigma_{\hat{P}}} < \frac{\hat{P} - p}{\sigma_{\hat{P}}} < \frac{0.35 - p}{\sigma_{\hat{P}}}\right) \\
&= P\left(\frac{0.25 - 0.3}{\sqrt{0.00084}} < Z < \frac{0.35 - 0.3}{\sqrt{0.00084}}\right) \\
&= P(-1,72 < Z < 1.72) = 0.9146
\end{aligned}
$$