

2019年度秋学期 統計学 第13回

不確かな測定の不確かさを測る — 不偏分散とt分布

浅野 晃
関西大学総合情報学部



ちょっと前回までの復習

正規分布の場合の区間推定

例題

母集団
(受験者全体)

標本 X_1, \dots, X_n をとりだす
サイズ n

標本平均 \bar{X}

母平均 μ

正規分布
と仮定する

母平均 μ の95%信頼区間が
知りたい

(説明の都合です)

母分散 σ^2 がわかっているものとする

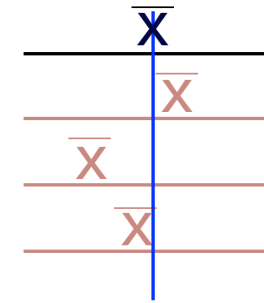
区間推定の考え方

数値をいくつか抽出して標本平均

仮に、何度も抽出したとすると

標本平均の
期待値
= 母平均

標本平均の
分散
= 母分散 \div
標本サイズ



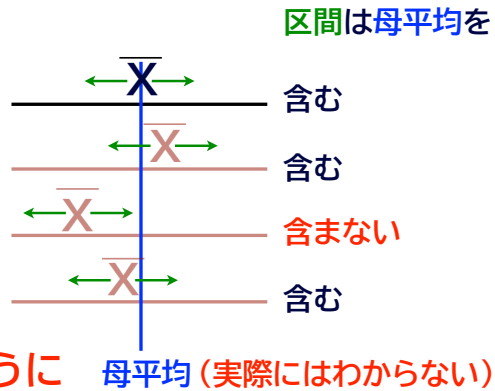
数値1個より
ばらつきが
小さくなる

母平均(実際にはわからない)
のまわりにばらついている

区間推定の考え方

標本平均の左右に区間をつける

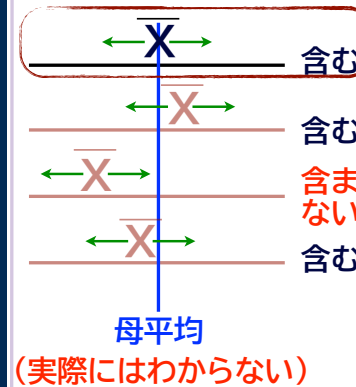
どの回の区間が母平均を含むか・含まないかはわからないが



確率95%で母平均を含むように区間を設定できる

信頼区間

区間は母平均を



確率95%で母平均を含むように計算した区間だから、その1回も含むと信じる

母平均の [信頼係数]95%の [信頼区間] という ([95%信頼区間])

不偏分散💡

正規分布の場合の区間推定

例題

母集団 (受験者全体)

標本 X_1, \dots, X_n をとりだす
サイズ n

標本平均 \bar{X}

母平均 μ

母平均 μ の95%信頼区間を知りたい

正規分布と仮定する

(説明の都合です)

母分散 σ^2 がわかっているものとする

母分散は、ふつうはわからない

母集団全体は調べていないし、
母平均もわからない
(わからないから、
いま推定しようとしている)

それなのに、母分散がわかるはずがない

母分散の「代用品」を、標本を使って計算
できないか。

標本を使って分散を計算

分散 = (偏差)²の平均

(データの各数値) - (データの平均)

標本を使って分散を計算する。

データ: 標本 X_1, \dots, X_n

データの平均: 本当は母平均だが、
わからないので
標本平均 \bar{X} で代用

標本を使って分散を計算

標本を使った分散

$$S^2 = \frac{1}{n} \{ (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \}$$

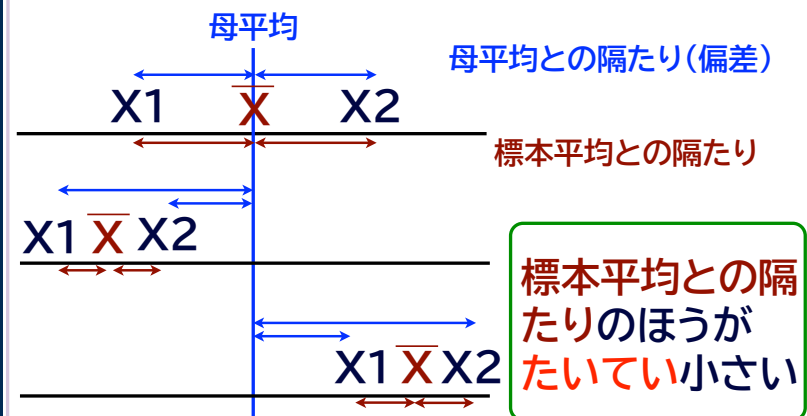
← 標本サイズで割る

分散 = (偏差)²の平均
だから当然だけど...

本当にこれでいいの？

標本平均を用いた偏差

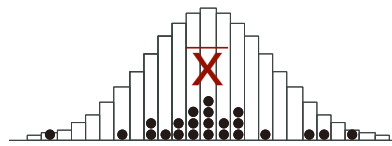
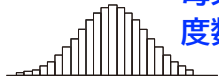
標本サイズ $n=2$ とする 標本は X_1, X_2



標本平均を用いた偏差

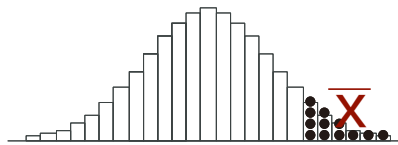
別の説明

母集団の
度数分布



これなら
「標本平均との隔たり」と
「母平均との隔たり」は
かわらない

こんなふうに偏っていると
「標本平均との隔たり」
のほうが小さい



不偏分散

母平均との隔たりよりも
標本平均との隔たりのほうが
たいてい小さい

標本平均との隔たりを使って分散を計算
すると、母分散よりもたいてい小さめ
になる

では、計算のときに少し大きめ
にしておけば？

不偏分散

計算のときに少し大きめにする

$$s^2 = \frac{1}{n-1} \{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2\}$$

(標本サイズ - 1)で割る

これを不偏分散(不偏標本分散)といい、
母分散の代用に用いる

「不偏」とは？

「不偏」とは？

標本平均との隔たりを使って分散を計算
すると、母分散よりもたいてい小さめになる

計算のときに少し大きめにする？

母分散と一致するわけではないが
母分散より大きくも小さくも
平等にはずれる

「不偏」とは「えこひいきしない」こと

不偏分散を用いた区間推定💡

正規分布の場合の区間推定

前回の例題

ある試験の点数の分布は正規分布であるとします。

この試験の受験者から、10人からなる標本を無作為抽出して、この人たちの点数を平均したところ50点でした。

この試験の受験者全体の標準偏差が5点であるとわかっているとき、受験者全体の平均点の95%信頼区間を求めてください。

正規分布の場合の区間推定

前の例題

母集団
(受験者全体)

標本 X_1, \dots, X_n をとりだす
サイズ n

標本平均 \bar{X}

母平均 μ

母平均 μ の95%信頼区間が
知りたい

正規分布
と仮定する

(説明の都合です)

母分散 σ^2 がわかっているものとする

正規分布の場合の区間推定

考え方

標本は、母集団分布と同じ確率分布にしたがう
正規分布 $N(\mu, \sigma^2)$

標本平均は、やはり正規分布にしたがうが、分散が $1/n$ になる [性質2]
正規分布 $N(\mu, \sigma^2/n)$

正規分布の場合の区間推定

考え方

標本は、母集団分布と同じ確率分布にしたがう
正規分布 $N(\mu, \sigma^2)$

\bar{X}
標本平均は、やはり正規分布にしたがうが、分散が
 $1/n$ になる 正規分布 $N(\mu, \sigma^2/n)$ [性質2]

正規分布の[性質1]により

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \text{ は標準正規分布にしたがう } N(0, 1)$$

本当は、母分散はわからない

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \text{ は標準正規分布にしたがう } N(0, 1)$$

本当は母分散はわからない

不偏分散で代用する

$$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \text{ 不偏分散}$$

何分布にしたがう？

t分布

t統計量 $t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}}$ は

自由度 $(n-1)$ のt分布にしたがう
 $t(n-1)$

(「スチューデントのt分布」という)

発見者ウィリアム・ゴセットのペンネーム

正規分布(母分散不明)の場合の区間推定

テキストの例題

ある試験の点数の分布は正規分布であるとしてます。

この試験の受験者から、10人からなる標本を無作為抽出して、この人たちの点数を平均したところ50点でした。

この10人の不偏分散が 5^2 点であるとき、受験者全体の平均点の95%信頼区間を求めてください。

前は
「受験者全体の標準偏差が5点であるとわかっている」

今回の例題は

母集団 (受験者全体) 母平均 μ 正規分布と仮定する

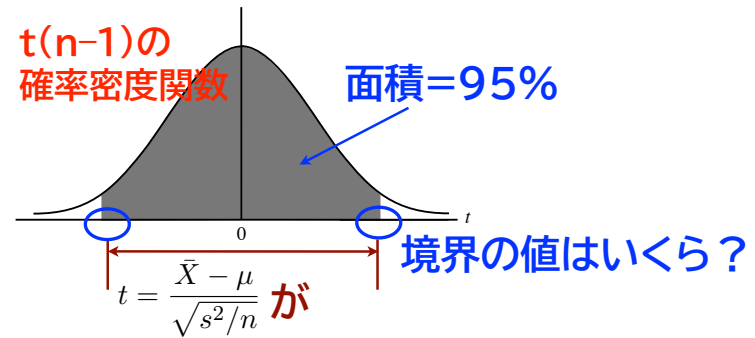
標本 X_1, \dots, X_n をとりだす サイズ n 標本平均 \bar{X}

母平均 μ の95%信頼区間を知りたい

不偏分散 s^2 で代用 母分散 σ^2 がわからないので、

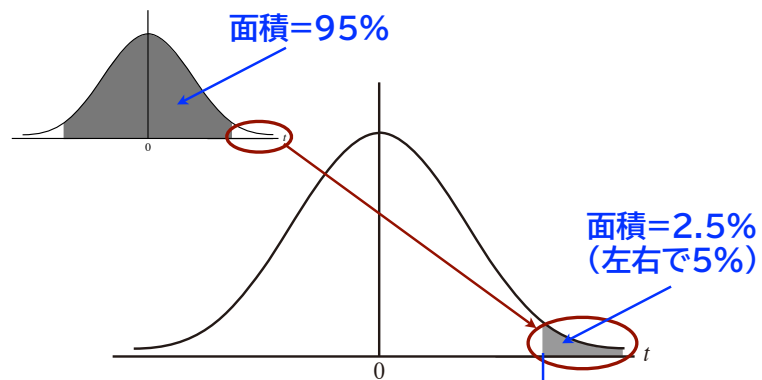
t分布を用いた区間推定

$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}}$ は自由度 $(n-1)$ の t 分布にしたがう $t(n-1)$



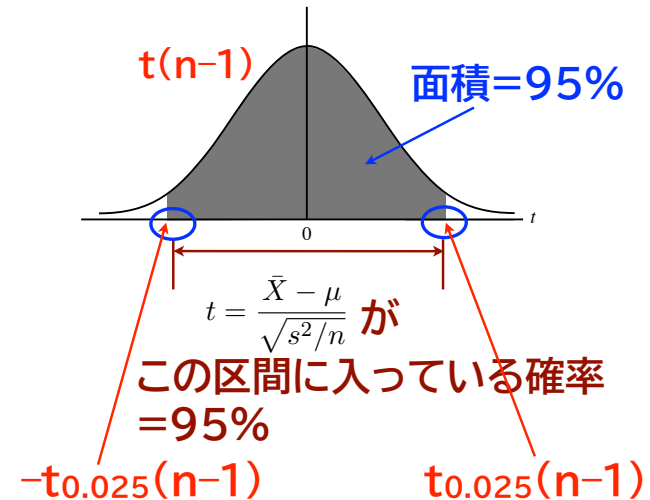
この区間に入っている確率=95%とすると

t分布を用いた区間推定



境界の値は自由度によってちがうので $t_{0.025}(n-1)$ としておく [上側2.5%点]

t分布を用いた区間推定



t分布を用いた区間推定

$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}}$ が $-t_{0.025(n-1)}$ と $t_{0.025(n-1)}$ の間に入っている確率が95%

式で書くと

$$P\left(-t_{0.025(n-1)} \leq \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \leq t_{0.025(n-1)}\right) = 0.95$$

μ の式に直すと

$$P\left(\bar{X} - t_{0.025(n-1)}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + t_{0.025(n-1)}\sqrt{\frac{s^2}{n}}\right) = 0.95$$

t分布を用いた区間推定

例題では

標本平均=50 不偏分散=25 標本サイズ=10

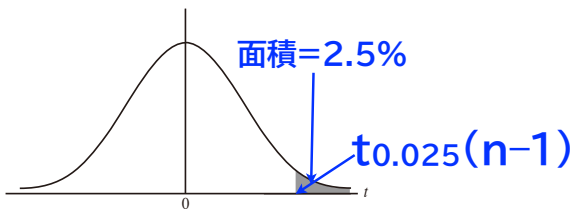
$$P\left(\bar{X} - t_{0.025(n-1)}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + t_{0.025(n-1)}\sqrt{\frac{s^2}{n}}\right) = 0.95$$

μ の95%
信頼区間の
下限

μ の95%
信頼区間の
上限

上側2.5%点 $t_{0.025(n-1)}$ は?

t分布表



パーセントの値 0.025

	0.40	0.30	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	0.3249	0.7265	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567
2	0.2887	0.6172	0.8165	1.0607	1.3856	2.9200	4.3027	6.9646	9.9248	15.8585
3	0.2767	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.2707	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.2672	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.2648	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.2632	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.2619	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.2610	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498

例題では $n-1 = 9$ $t_{0.025(9)} = 2.262$

t分布を用いた区間推定

例題では

標本平均=50 不偏分散=25 標本サイズ=10

$$P\left(\bar{X} - t_{0.025(n-1)}\sqrt{\frac{s^2}{n}} \leq \mu \leq \bar{X} + t_{0.025(n-1)}\sqrt{\frac{s^2}{n}}\right) = 0.95$$

μ の95%
信頼区間の
下限

μ の95%
信頼区間の
上限

計算すると、例題の答は
「46.4以上53.6以下」([46.4, 53.6])

前回の例題と比較

どちらも

標本平均=50
標本サイズ=10

母分散=25 のとき

母平均の
95%信頼区間
は[46.9,
53.1]

不偏分散=25 のとき [46.4, 53.6]

不偏分散は、母分散の推定量なので、
不確か → 信頼区間が広い