

2015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



1

ちょっと前回までの復習

2

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

2015

3

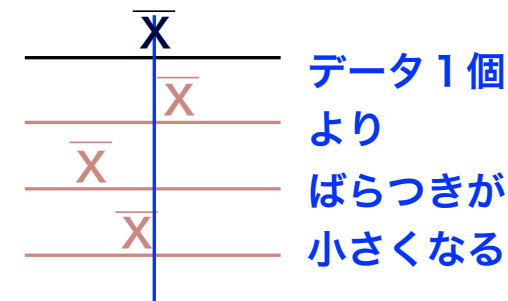
区間推定の考え方

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



データ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

2015

4

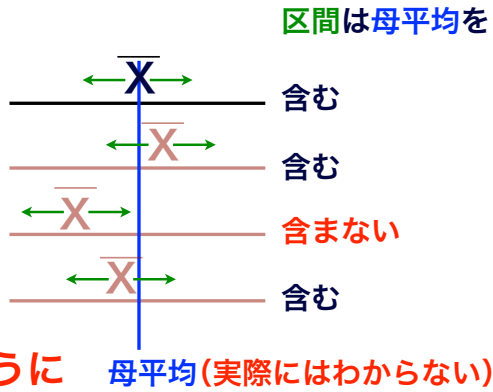
区間推定の考え方

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

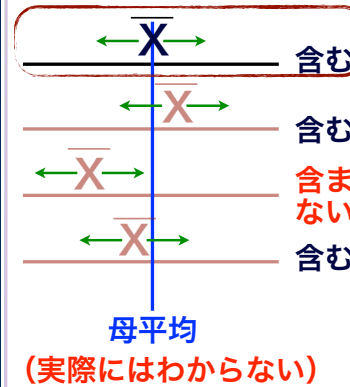


2015

5

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])

A.Asano, Kansai Univ.

2015

6

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

A.Asano, Kansai Univ.

2015

8

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。

2015

9

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

2015

10

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…

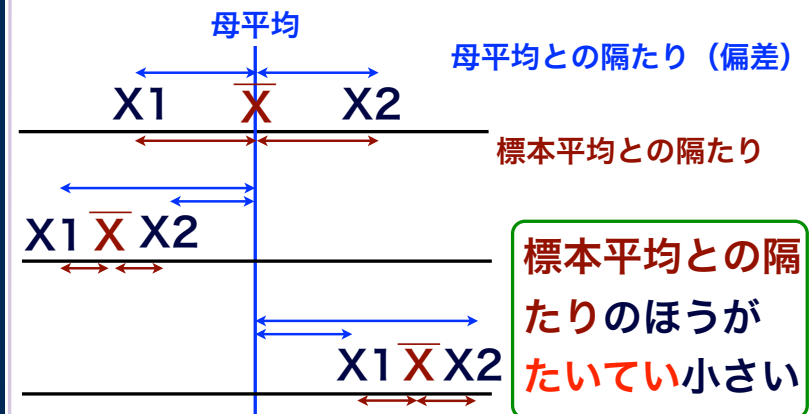
本当にこれでいいの？

2015

11

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



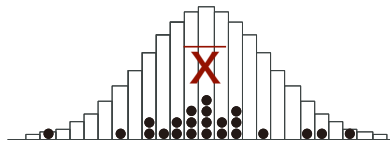
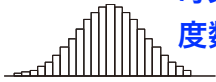
2015

12

標本平均を用いた偏差

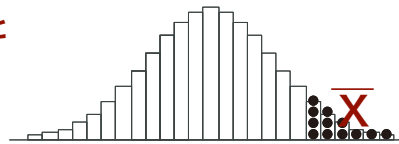
別の説明

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



2015

13

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？

2015

14

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？

2015

15

「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と？

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

2015

16

不偏分散を用いた区間推定

17

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

2015

18

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 $N(\mu, \sigma^2/n)$

2015

19

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

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

2015

20

母分散はわからない

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

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

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

不偏分散

何分布にしたがう？

2015

21

t分布

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

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

2015

22

この例題は

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、

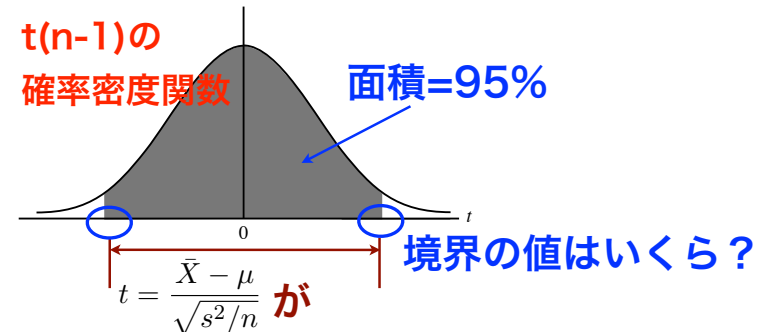
2015

23

t分布を用いた区間推定

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

$t(n-1)$

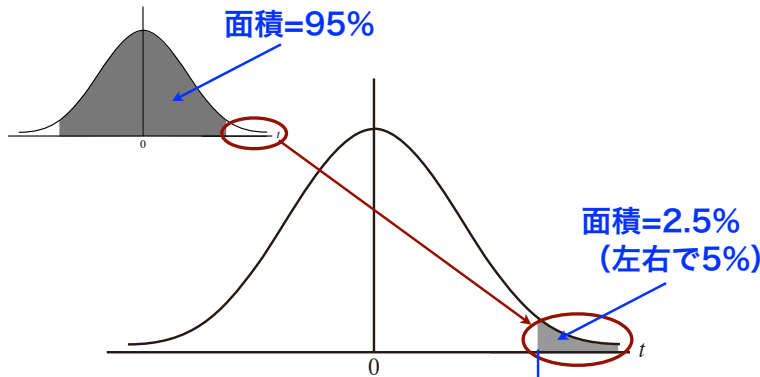


この区間に入っている確率=95%とすると

2015

24

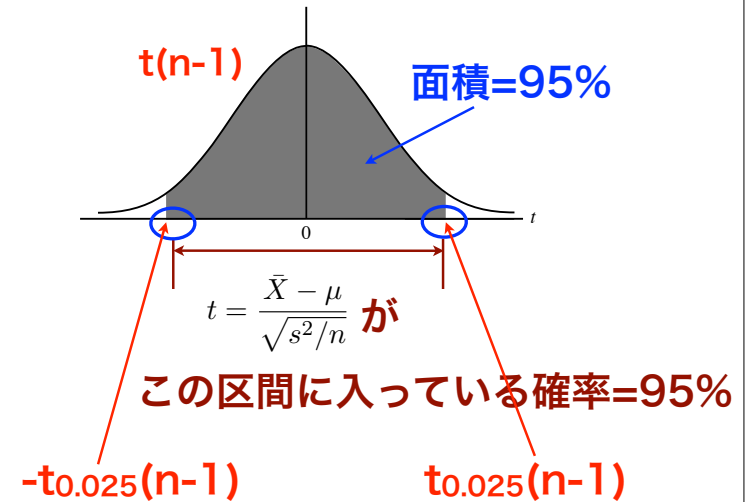
t分布を用いた区間推定



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] という

25

t分布を用いた区間推定



26

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

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

27

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%

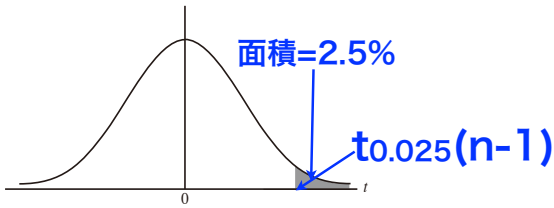
信頼区間の

上限

上側2.5%点 $t_{0.025}(n-1)$ は?

28

t分布表



パーセントの値 **0.025**

自由度	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	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248
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 $n-1 = 9$

$t_{0.025}(9)=2.262$

2015

t分布を用いた区間推定

例題では

$$t_{0.025}(10-1)=2.262$$

標本平均=50

不偏分散=25

標本サイズ=10

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

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])

A.Asano, Kansai Univ.

2015

前回の例題と比較

どちらも

標本平均=50

標本サイズ=10

母平均の

95%信頼区間は

母分散=25 のとき

[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

A.Asano, Kansai Univ.

2015