2019年度春学期　統計学　　第12回

分布の平均を推測する ― 区間推定

浅野　晃
関西大学総合情報学部

ちょっと前回までの復習🤔

「統計的推測」とは

調べたい集団の，データ全体を調べられるか？

日本男性全員の身長を調べられるか？

データの一部を調べて
度数分布を推測する

いや，せめて平均や分散を推測する

統計的推測

母平均の推定

母集団
（日本男性全体）

母平均 μ

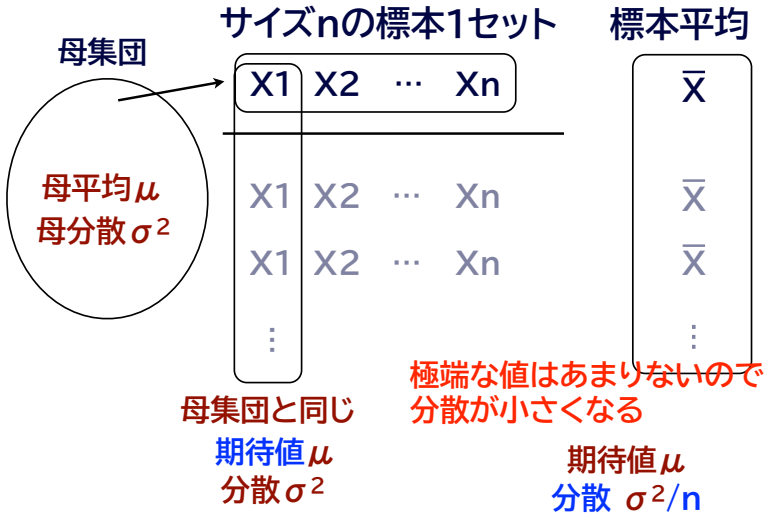
標本として数値を
いくつか取り出して，
それらの平均

［標本平均］

標本平均は母平均に
近い値になるか？

母平均が知りたい
が，日本男性全員は調べられない

母平均の推定



母平均の推定

いま1回だけ計算した標本平均も，
おそらく，ほぼ母平均に近い値だろう

どのくらい近い？

どのくらいの確率で？
はずれる確率は？

正規分布モデル

世の中には，[正規分布モデル]で表せるような母集団分布がたくさんある

長さの測定値の分布
センター試験の成績の分布 …

[中心極限定理]

母集団のばらつきの原因が
無数の独立な原因の和のとき，
母集団分布は概ね正規分布になる

正規分布の性質1

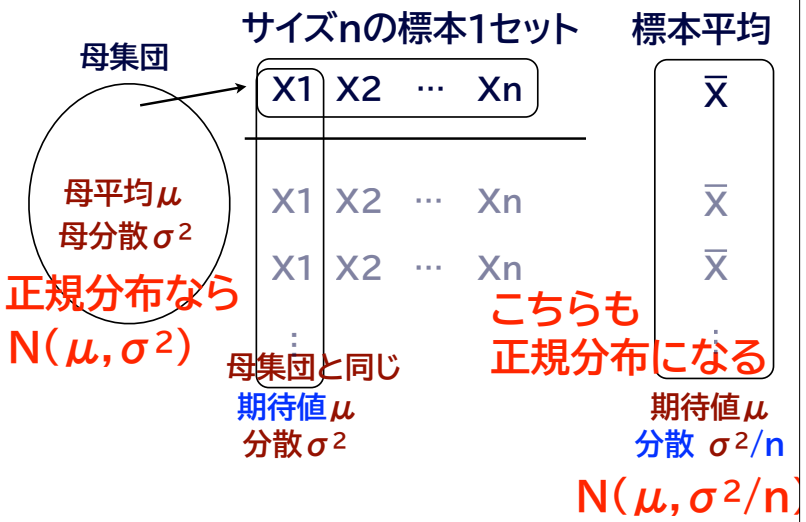
確率変数XがN(μ, σ^2)にしたがう とき
$(X-\mu)/\sigma$ は N(0, 1)にしたがう

「標準得点」と同じ

変換しても，
やはり正規分布になる

N(0,1)を[標準正規分布]という

正規分布の性質2



区間推定

区間推定

母平均を推測する。その答え方が,

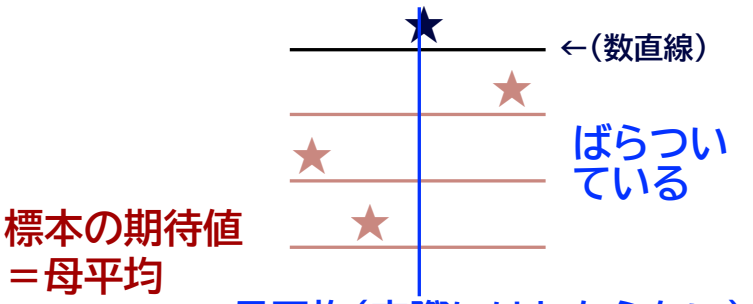
母平均は, 50 から 60 の間にあると推測する。

この推測が当たっている確率は95%である。

区間推定の考え方

標本として数値をひとつだけ抽出

仮に, 何度も抽出したとすると



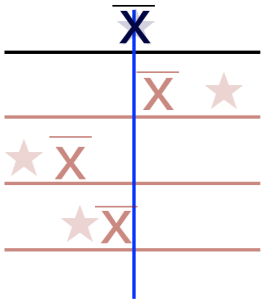
区間推定の考え方

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÷
標本サイズ



数値1個
のときより

ばらつきが
小さくなる

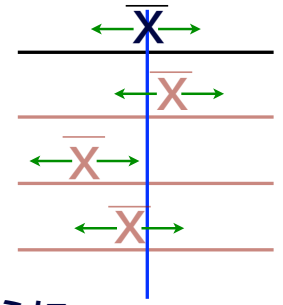
母平均（実際にはわからない）
のまわりにばらついている

区間推定の考え方

標本平均の左右に区間をつける

区間は母平均を

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

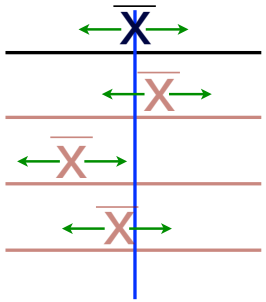


確率95%で
母平均を含むように
区間を設定できる

母平均（実際にはわからない）

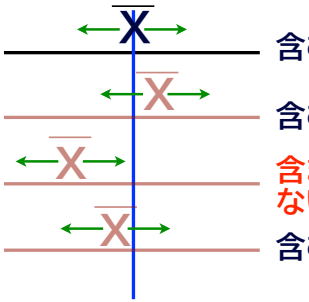
区間推定の考え方

確率95%で母平均を含むように
区間を設定できる



標本平均のばらつきは
小さくなっているので
区間の幅は
そこそこ狭くてよい

区間推定の考え方

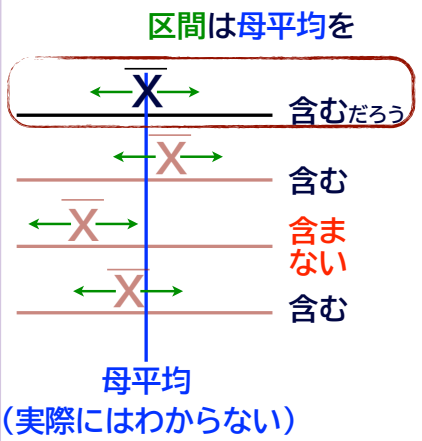


実際には，
標本平均と区間は
1度しか計算しない

その区間が，母平均を
含むかどうかは
実際にはわからない

しかし，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

信頼区間



95%という大きな確率で母平均を含むように設定した区間だから，その1回でも含むと信じる

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

「信頼」ということば

私は，予言者です。
私の予言は，確率95%で当たります💡

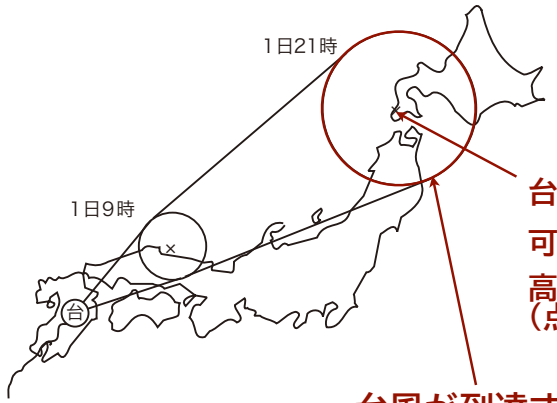
いまから，来年おきることを予言します。「来年は、…」

このひとつの予言が正しいかどうかはわからない

しかし，十分多くの予言をすれば95%は当たるのだから，この予言も信じる価値はある

これが「信頼」

台風情報と信頼区間



台風が到達する可能性がもっとも高いところ（点推定）

台風が到達する確率が70%であるような円
＝区間推定

正規分布の場合の信頼区間の計算

正規分布の場合の区間推定

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

例題

標本X1, ... Xnをとりだす
サイズn

母集団
（受験者全体）

標本平均 $\overline{X}$

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

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(μ, σ^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(μ, σ^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(μ, σ^2)

$\overline{X}$
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(μ, σ^2/n) [性質2]

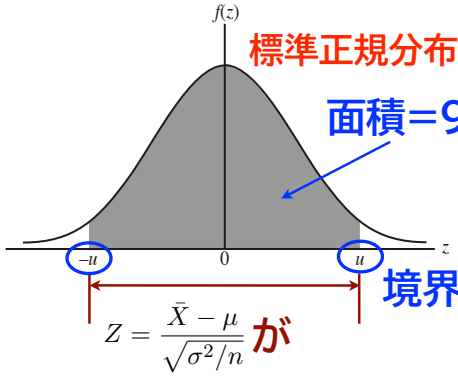
正規分布の[性質1]により

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

は標準正規分布にしたがう
N(0, 1)

正規分布の場合の区間推定

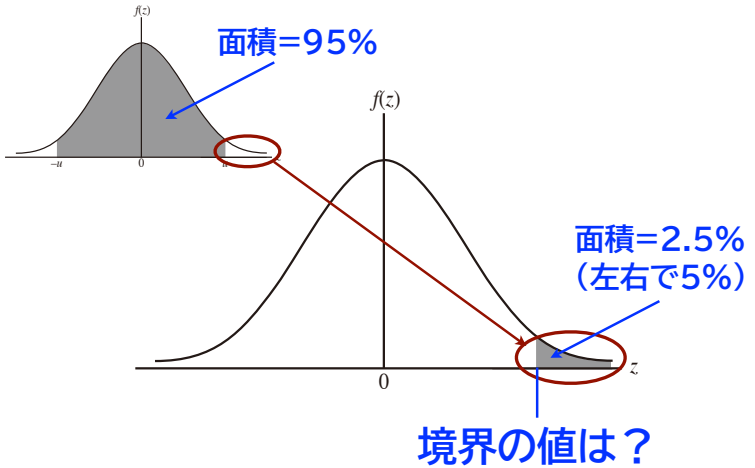
$Z = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ は標準正規分布にしたがう



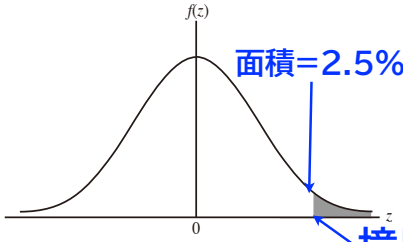
$Z = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ が

この区間に入っている確率=95%とすると

正規分布の場合の区間推定



正規分布の場合の区間推定



うまいぐあいに，正規分布表で

0.06

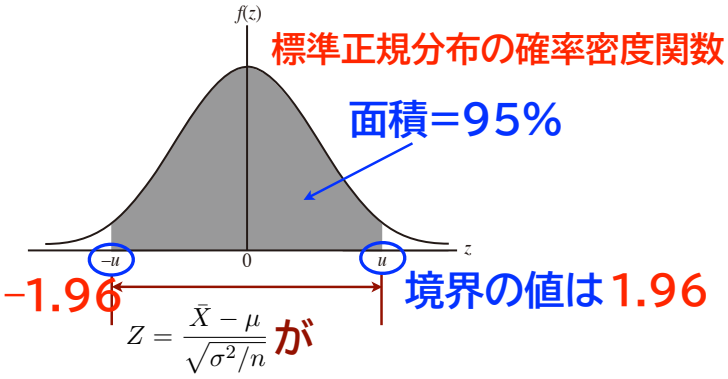
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
...										
1.9	0.028717	0.028067	0.027429	0.026803	0.026190	0.025588	0.024998	0.024419	0.023852	0.023295

1.9

0.024998

正規分布の場合の区間推定

$Z = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ は標準正規分布にしたがう



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

この区間に入っている確率=95%とすると

正規分布の場合の区間推定

$Z = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ が −1.96と1.96の間に入っている確率が95%

式で書くと

$$P(-1.96 \leqq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leqq 1.96) = 0.95$$

μの式に直すと

$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leqq \mu \leqq \bar{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

正規分布の場合の区間推定

例題では

標本平均=50　母分散=25　標本サイズ=10

$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leqq \mu \leqq \bar{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

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「46.9以上53.1以下」　[46.9, 53.1]

信頼区間の答え方

$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leqq \mu \leqq \bar{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

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 [46.9, 53.1]

これを

$$P(46.9 \leqq \mu \leqq 53.1) = 0.95$$

と書いてはいけないの？

だめです🤦　なぜ？

信頼区間の答え方

$$P(46.9 \leqq \mu \leqq 53.1) = 0.95$$

と書いてはいけないの？　だめ。

確率の式なのに，()内にランダムなものが入っていないから

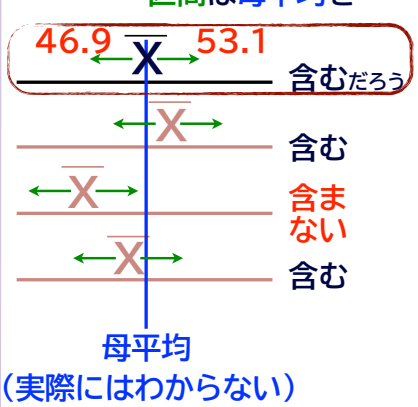
$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leqq \mu \leqq \bar{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

こう書いたとき，

μ(母平均)はランダムではない

ランダムなのは $\bar{X}$(標本平均)

これを思いだしてください



実際には，
標本平均と区間は
1度しか計算しない

その区間が，母平均を
含むかどうかは
実際にはわからない

しかし，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

区間推定についての注意

$$P(\bar{X} - 1.96\sqrt{\sigma^2/n} \leqq \mu \leqq \bar{X} + 1.96\sqrt{\sigma^2/n}) = 0.95$$

1. 母集団の大きさは関係ない

復元抽出なら，母集団分布は
標本抽出によって変化しない

2.「95%」を選ぶ根拠はない

「確率5%なら，推測がはずれて
失敗しても，まあいいか」と思っ
ているだけ