

メルマガ登録 無料特典

推定ってなに??

著作権について

- e-Book「メルマガ無料特典その2 推定」は、著作権法で保護される著作物です。取り扱いについては、以下の点にご注意ください。
- 「メルマガ無料特典その2 推定」の著作権は、吉田寛輝にあります。著作権者の書面による事前許可なく、「メルマガ無料特典その2 推定」の一部または全部をあらゆる手段(印刷物、ビデオ、音声データ、電子記憶媒体、およびホームページ、ウェブログ等のインターネットメディア、公演、電子メール等での配信など)で一般公開することを禁じます。著作権侵害を行なった場合は、5年以下の懲役または500万円以下の罰金に処せられ、このような侵害行為が法人等の従業員により法人等の業務として行なわれた場合には、当該法人等に1億5千万円以下の罰金が科せられます。(著作権法119条)

「メルマガ無料特典その2 推定」使用契約

- 本契約は、「メルマガ無料特典その2 推定」を購入した個人または法人(以下、甲とする)と著作権者(以下、乙とする)との間の購入および使用に関する条件を定めたものであり、甲と乙の一切の関係に適用します。なお、乙の本契約への同意が得られない場合には、甲は乙に対し「メルマガ無料特典その2 推定」を販売することはできません。
- 第1条(契約の目的) 本契約に基づき、乙が著作権を有する「メルマガ無料特典その2 推定」に含まれる情報において、甲の非独占的使用権を許諾するものです。「メルマガ無料特典その2 推定」は、投資判断の参考となる情報を目的としたものであって、投資勧誘を目的としたものではありません。この内容を利用する者の経済的その他の利益を約束する性質のものではなく、甲はこれを認識し購入するものとしします。
- 第2条(購入手続に関する免責) 甲の入力ミスなどあらゆる要因による注文メールの未着・遅延、およびそれに伴う損害等に対しての責任は、乙は何らの責を負わないものとしします。

「メルマガ無料特典その2 推定」使用契約

- 第3条（乙に許諾される使用权） 甲は、本契約に基づき使用权が許諾された「メルマガ無料特典その2 推定」をいかなる場合であっても他の者に利用させてはならない。ただし、事前に乙の書面による承諾を得た場合に限り、以下の条件をすべて満たす範囲の者に関しては情報を共有することができます。
- 1 ① 住所を同じくしていること。
 - 2 ② 生計を同じくしていること。
 - 3 ③ 3親等以内の親族であること。
- 2 ダウンロードでの購入の場合、乙は、ダウンロードの際使用したコンピュータ（以下単独コンピュータという）でのみ情報の記載されたファイルを閲覧することができます。3 単独コンピュータ以外のコンピュータで情報を閲覧するときは、乙から別途新たに使用权の許諾を取得することを必要とするものとします。
- 4 甲は、第三者に対して、「メルマガ無料特典その2 推定」の本契約に基づく使用权、「メルマガ無料特典その2 推定」の購入によって発生した権利・義務、及び本契約の対象となる情報又はその他の如くなるものも、譲渡したり、移転したりすることはできません。また、本契約には、「メルマガ無料特典その2 推定」の全部又は一部を印刷 又は複製する権利の許諾は含まれておりません。
- 5 甲が本契約に基づく「メルマガ無料特典その2 推定」の全部または一部を複製できる場合は、この情報のバックアップの作成を目的として紙に複製する、またはプログラム用ディスクに複製する場合のみとします。

「メルマガ無料特典その2 推定」使用契約

- 第4条(情報の変更又は改作) 甲は、購入した「メルマガ無料特典その2 推定」を変更、編集および改作することはできません。
- 第5条(契約の解除) 乙により甲が本契約に違反したと判断された場合、乙は何の通告もなく使用契約を解除できるものとします。
- 第6条(禁止事項) 「メルマガ無料特典その2 推定」に含まれる情報は、著作権法によって保護されています。この書面による事前許可なく、「メルマガ無料特典その2 推定」の一部または全部をあらゆる手段(印刷物、ビデオ、音声データ、電子記憶媒体、およびホームページ、ウェブログ等のインターネットメディア、公演、電子メール等での配信など)で一般公開することを禁じます。また、本「メルマガ無料特典その2 推定」を第三者へ譲渡、転売することを禁じます。「メルマガ無料特典その2 推定」の情報は、甲が自ら使用する場合同様にのみ許諾されるものとします。万一、違反行為を発見した場合には弁護士を通じての法的手段によって解決いたします。
- 第7条(損害賠償) 甲が本契約の第6条に違反した場合、本契約の解除にかかわらず、甲は乙に対する違約金として違反件数と販売価格を乗じたものの10倍の金額を支払うものとします。また、公演やインターネット等での公開により、違反件数が特定できない場合は、一律500万円を甲は乙に対して支払うものとします。

「メルマガ無料特典その2 推定」使用契約

- 第8条(責任の範囲) 「メルマガ無料特典その2 推定」に記載の内容は投資の手法として高い有益性があるものと確信していますが、その正確性を乙が保障するものではありません。運用および投資にあたっては、投資者である甲自身の責任において行うものとしします。万一、「メルマガ無料特典その2 推定」に記載の方法に従って甲が投資を行い損害が発生しても、損失金額の補填および弁済その他一切の責任を乙は負わないものとしします。
- 第9条(同意効力の発生時期) 本契約は、甲が乙に対し、ウェブサイト上の規約に対する「同意」をクリックしたときから効力を有するものとしします。

統計の二つの意味（特典その1再掲）

■統：まとめる

□ = 要約する

□例：A製薬の100人の社員の身長の平均値を求める

推定といえます

■計：計画する

□ = 検定する

□例：A製薬の100人の身長平均値が、165cm以上かどうかを判断する

検定といえます

この資料でみなさんにわかっていただきたいこと

1.推定が何をしているかを理解してもらおうこと

推定：データを要約することからスタート！

■そもそもデータ = Dataとは？

- Data = 数値や特性の集まり
- Datumの複数形

■臨床試験では？

- 被験者ごとに収集されるデータ
 - ✓例：身長、体重、性別、腫瘍の大きさ、など



数値を一つ一つ眺めても、何も分からない！

一つ一つのデータはばらつく

■女子大生の身長（10人分）

身長の測定結果(cm)									
1	2	3	4	5	6	7	8	9	10
155	156	160	163	165	162	158	167	163	159

■個人の身長を眺めても、女子大生全体の平均値は不明

□なぜか？ → 一人一人の身長はバラついているため

論文や学会発表にある集計表

	A群	B群
	N=291	N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288
Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

Callout boxes and arrows:

- 標準偏差 (Standard Deviation) points to the SD row.
- 標準誤差 (Standard Error) points to the SEM row.
- 中央値 (Median) points to the Median row.
- 平均値 (Mean) points to the Mean row.

ここからは、連続量を中心にお話しします

■推定：データを要約すること

□代表的な要約の指標（**要約統計量**といいます）

✓平均値、中央値、標準偏差、標準誤差、95%信頼区間（、最大値、最小値）

■これからのお話

1. 怖い話：平均値しか知らない学生

2. スッキリ!：偏差値って？

3. 点と線：信頼区間はなぜ必要か？

ここからは、連続量を中心にお話しします

■推定：データを要約すること

□代表的な要約の指標（**要約統計量**といいます）

✓平均値、中央値、標準偏差、標準誤差、95%信頼区間（、最大値、最小値）

■これからのお話

1. 怖い話：平均値しか知らない学生

2. スッキリ!：偏差値って？

3. 点と線：信頼区間はなぜ必要か？

論文や学会発表にある集計表

	A群	B群
	N=291	N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288
Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

平均値

中央値

統計に対する皮肉2

■統計学 = 平均値で議論しているイメージ

■Statistics has been described as the science which tells you that if you lie with your head in the oven and your feet in the refrigerator, on average you'll be comfortably warm



質問！（再掲）

- 規模（社員10人）、業務内容が同じ製薬会社
 - 唯一異なるのは、平均年収！



A製薬：平均年収 **500**万円



B製薬：平均年収 **900**万円

どちらの会社に入りたいですか??

覚えていただきたい指標：中央値

■データ数が奇数個（例：1,4,5,6,8）の場合

□5つのデータの真ん中に当たる3番目のデータ

1 → 4 → 5 → 6 → 8

■データの数が偶数個（例：1,5,6,8）の場合

□4つのデータの間際に当たるデータ（2番目と3番目の平均値）

$$\frac{5 + 6}{2} = 5.5$$

中央値の特徴：外れ値の影響を受けない

年収データ公開！

■各社の年収の中央値を算出してみましょう！



A製薬、平均値：500万円、中央値：_____万円

450	465	475	475	495	505	525	525	535	550
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



B製薬、平均値：900万円、中央値：_____万円

300	350	375	380	380	380	400	430	455	5550
-----	-----	-----	-----	-----	-----	-----	-----	-----	------

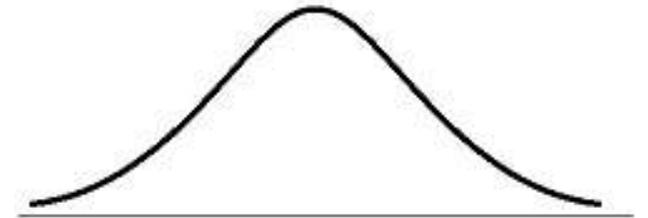
平均値と中央値のどちらが、要約として適切か？

平均値と中央値の使い分け

■平均値での要約が適切な状況

□データが左右対称の分布に従っていること

✓例：身長は左右対称な分布（正規分布）に従うことが知られている

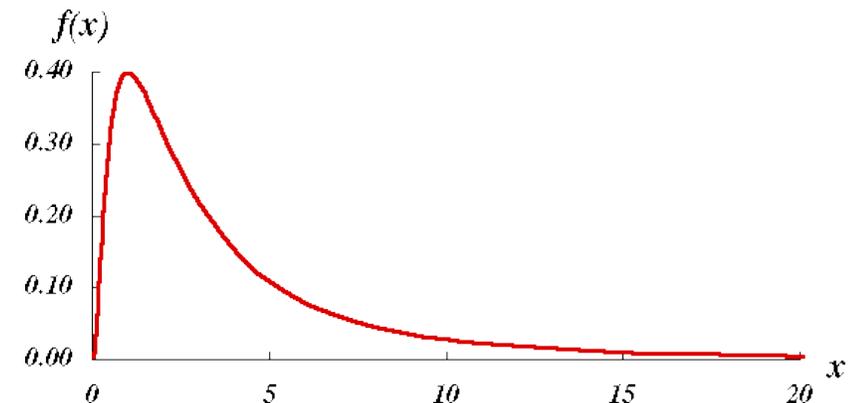


■中央値での要約を検討した方が良い状況

□データの分布が左右対称でない時

✓例：血液検査（AST,ALT）などは分布が右に裾を引く

✓これらは、少数例が非常に高い値のことがある



ここからは、連続量を中心に話します

■推定：データを要約すること

□代表的な要約の指標（**要約統計量**といいます）

✓ 平均値、中央値、**標準偏差**、**標準誤差**、95%信頼区間（、最大値、最小値）

■これからのお話

1. 怖い話：平均値しか知らない学生

2. スッキリ!：偏差値って？

3. 点と線：信頼区間はなぜ必要か？

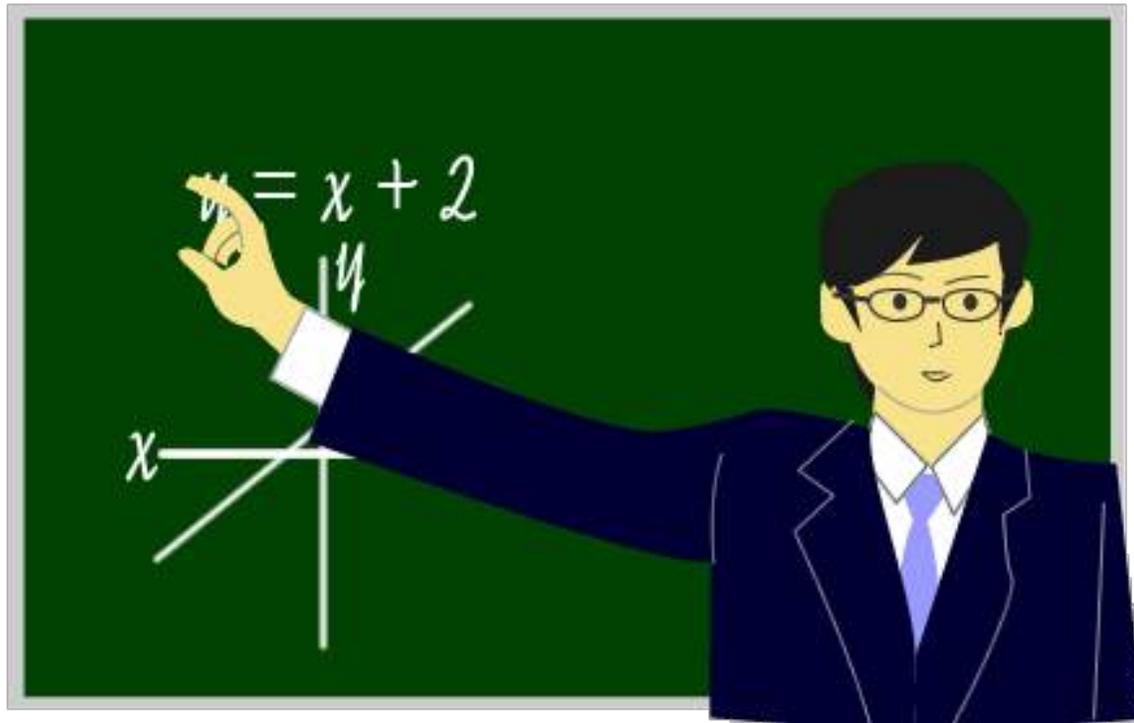
論文や学会発表にある集計表

	A群	B群
	N=291	N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288
Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

標準偏差

標準誤差

覚えていますか？偏差値？



あの大学の偏差値は63くらい！

この大学は偏差値55程度でも大丈夫かも！

偏差値の不思議

- ひろき君は数学と英語のテストを受けました
- 答案が返却され、確認すると…

	数学	英語
クラスの平均点	50点	50点
ひろき君の点数	70点	70点
偏差値	70	60



一つ一つのデータは、ばらつく

■女子大生の身長（10人分）

身長の測定結果(cm)									
1	2	3	4	5	6	7	8	9	10
155	156	160	163	165	162	158	167	163	159

■個人の身長を眺めても、女子大生全体の平均値は不明

□なぜか？ → 一人一人の身長はバラついているため

■個々のデータがどのくらいばらつくかにも興味がある

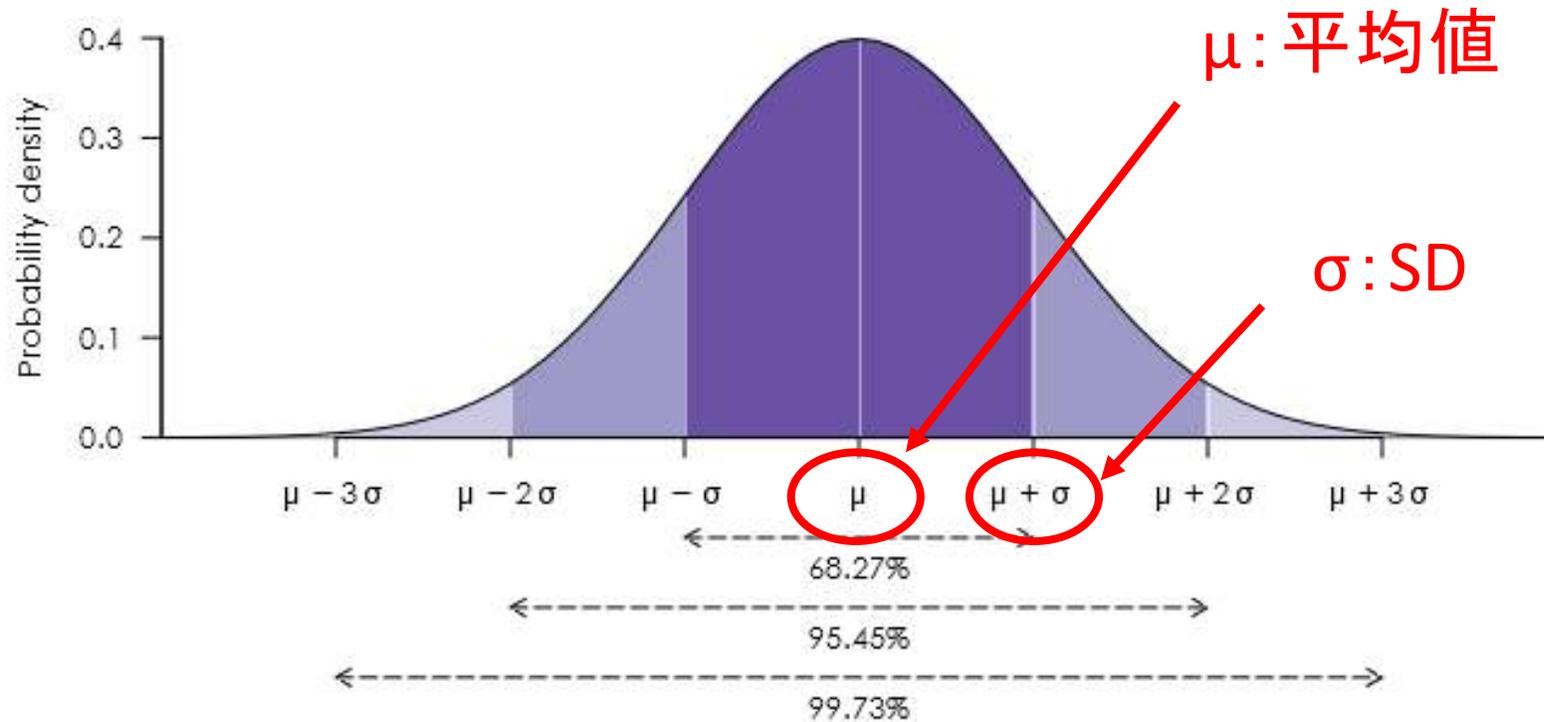
個々のデータのばらつきの指標：標準偏差(SD)

必ず覚えていただきたい分布：正規分布

■ 平均値を中心に左右対称の分布

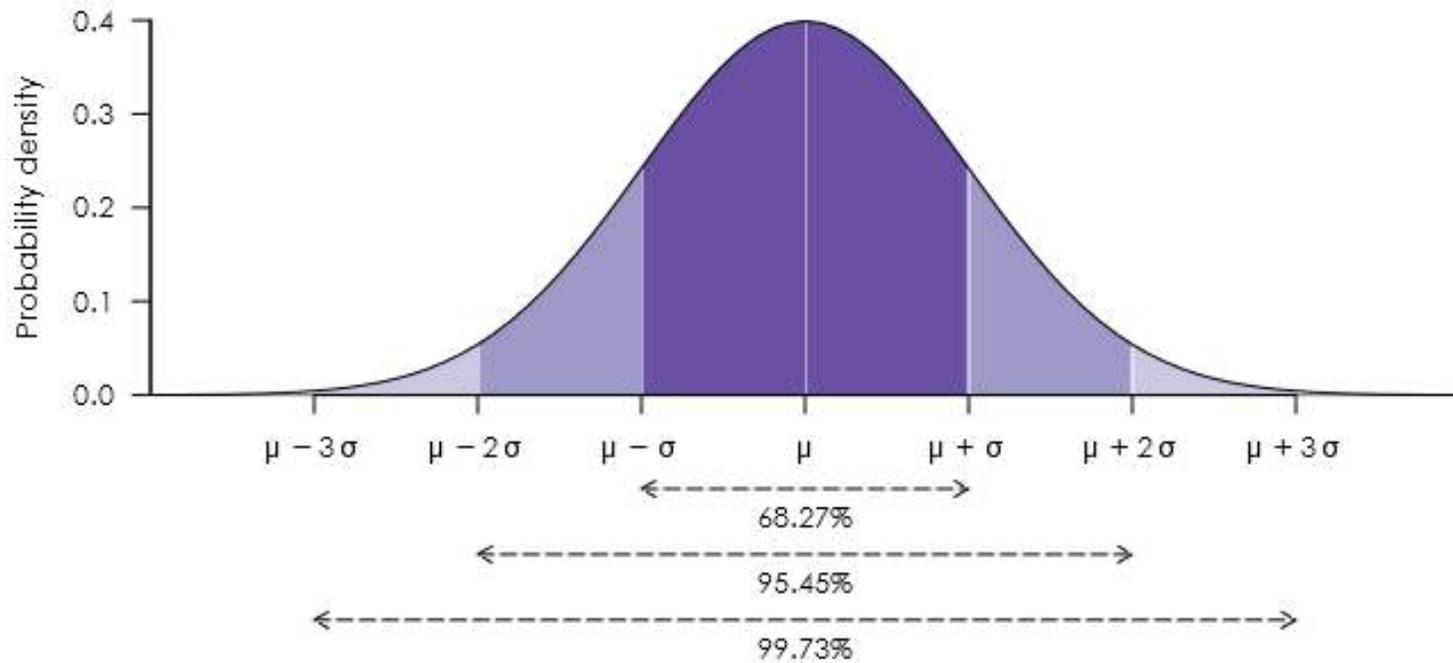
□ 身長や体重といったデータの多くは正規分布に従うと言われている

✓ 個々のデータを表す時は平均値と標準偏差（SD）で形が決まる

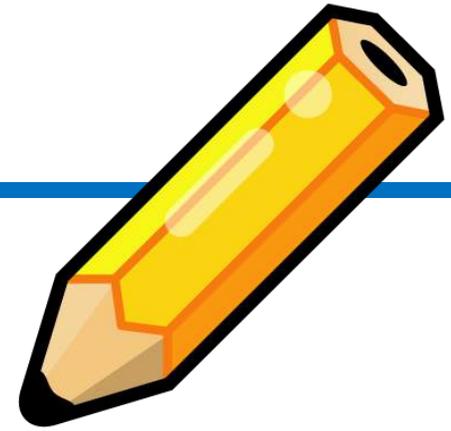


正規分布の性質

- 平均値±SDの区間に68.3%のデータが含まれる
 - 約3分の2
- 平均値±2*SDの区間に95.5%のデータが含まれる
 - 平均値±1.96*SDの区間に95%のデータが含まれる



データを読み解いてみましょう



	A群	B群
	N=291	N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288

母集団の約68%が

~

cm

に含まれる

Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

母集団の約95%が

~

cm

に含まれる

正規分布の性質：平均値とSD

■数学のテスト結果の分布

□ひろき君の偏差値：70

数学のSD
10点



70点

■英語のテスト結果の分布（青）

□ひろき君の偏差値：60

英語のSD
20点



50点 70点

標準偏差 → 偏差値

■センター試験などのテスト結果 → 正規分布に従う

■偏差値

□平均点を偏差値50として、SDの分だけ点数が上がると10加算する

□このため、ひろき君のテストの点数が70点であったとき

✓平均点が50点、標準偏差が10点の数学では、偏差値は70になる

■テストごとにばらつき（標準偏差）は当然異なる

□標準偏差が**大きい** = 低い点～高い点まで広く分布している（**英語**の状況）

□標準偏差が**小さい** = 平均点付近に多く分布している（**数学**の状況）

標準偏差とよく似た言葉：標準誤差

■10人の身長を思い出してみましょう

10人全員の身長の平均値: 160.8cm									
1	2	3	4	5	6	7	8	9	10
155	156	160	163	165	162	158	167	163	159

■4人の平均値の計算結果を書いてみてください

	自分の結果	Aさん	Bさん	Cさん
平均値				

知りたいのは女子大生全体（母集団）の平均値

■推定精度を上げるためにはどうすれば良いだろう？

女子大学生全体(122万人)



<母集団>

女子大学生の一部



<標本>

抽出

推定

抽出する人数が増えたら？

- A製薬に勤めている男性の身長の平均値を知りたい
 - 簡単のために、A製薬では男性が10,000人働いているとする

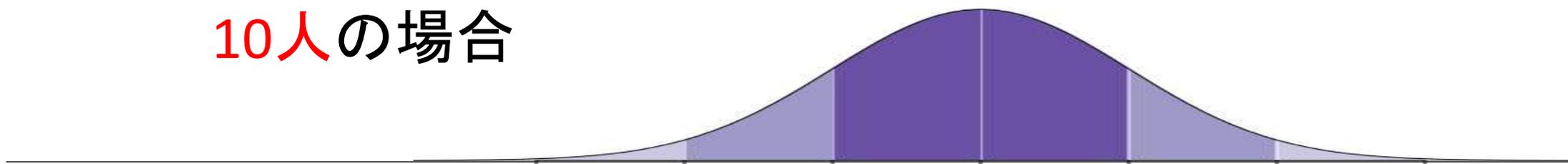
- 規模を変えて、A製薬の男性の身長を推測した
 1. 勤めている10人の身長を測り、平均値を求めた
 2. 勤めている100人の身長を測り、平均値を求めた
 3. 勤めている1000人の身長を測り、平均値を求めた

どの調査結果が1万人の身長の平均値(真値、母平均)に近いだろうか？

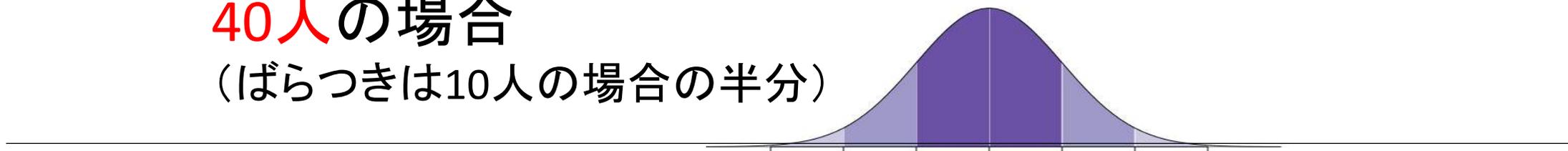
平均値の推定精度 = 標準誤差 (SE)

- 標準誤差は、抽出した人数が多いほど、小さくなる（推定精度が上がる）

10人の場合



40人の場合
(ばらつきは10人の場合の半分)



標準偏差 (SD) と標準誤差 (SE)

■標準偏差 (SD: Standard Deviation)

- 個々の値のばらつきを表す (例: 10,000人のデータのばらつき)

■標準誤差 (SE: Standard Error)

- 推定値の推定精度 (推定値のばらつき) を表す

- 被験者数 (データ) が増えるほど小さくなる →推定精度Up!

$$SE = \frac{SD}{\sqrt{\text{被験者数}}}$$

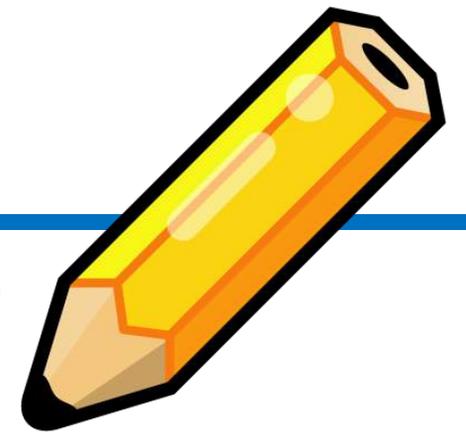
標準誤差 < 標準偏差

	A群 N=291	B群 N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288
Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

標準偏差

標準誤差

確認クイズ！！



■以下の場合、SDとSEのどちらを使えば良いでしょうか？

□Q1：被験者の身長がどの程度ばらついているのか

□Q2：データから出された身長の平均値が、どの程度信頼できるのか

ここからは、連続量を中心に話します

■推定：データを要約すること

□代表的な要約の指標（**要約統計量**といいます）

✓平均値、中央値、標準偏差、標準誤差、**95%信頼区間**（、最大値、最小値）

■これからのお話

1. 怖い話：平均値しか知らない学生

2. スッキリ!：偏差値って？

3. 点と線：信頼区間はなぜ必要か？

真値と推定値は一致しましたか？

■10人の身長を思い出してみましょう

10人全員の身長の平均値: 160.8cm									
1	2	3	4	5	6	7	8	9	10
155	156	160	163	165	162	158	167	163	159

■4人の平均値の計算結果を書いてみてください

	自分の結果	Aさん	Bさん	Cさん
平均値				

95%信頼区間 → 区間推定とといいます

■推定値は大きく分けて二つ！

推定値

```
graph TD; A[推定値] --- B[点推定]; A --- C[区間推定];
```

点推定

平均値 : Mean
中央値 : Median

区間推定

平均値の信頼区間 :
Confidence Interval of Mean

よくある試験結果（奏功割合）

■推定には2種類ある

- 点推定：1つの値のみを用いて真の値を推定
- 区間推定：推定値のばらつきを考慮し、真の値のある範囲を推定

■ある薬剤の試験結果

- 奏功割合は75.0%、その95%信頼区間は34.9%~96.8%であった



区間推定はなぜ必要か？

- コインを投げて表が出る確率の真の値 = 50%とする

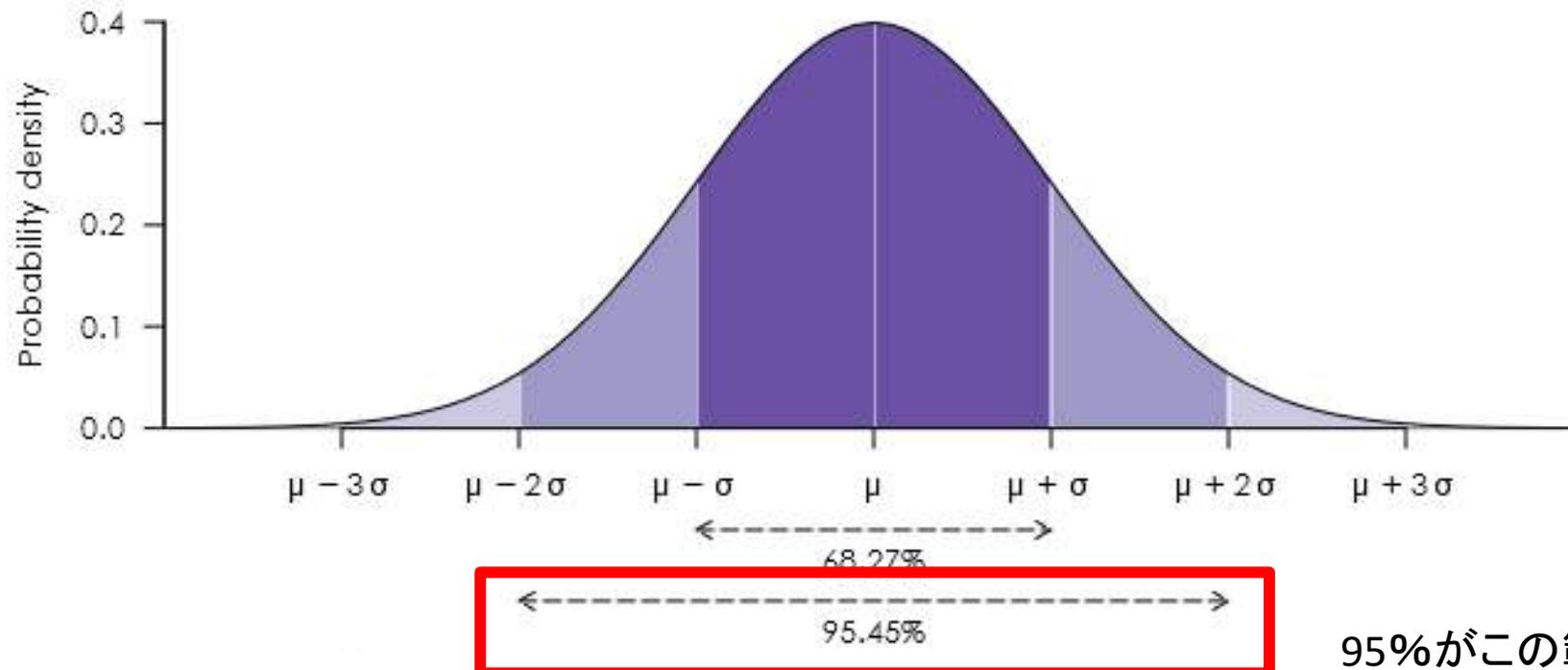


- コインを3回投げる実験を考える
 - この実験の結果得られる表が出る確率の点推定値
 - ✓ 0%, 33%, 67%, 100%の4通りのみ

真の値(表が出る確率=50%)を推定することが不可能！

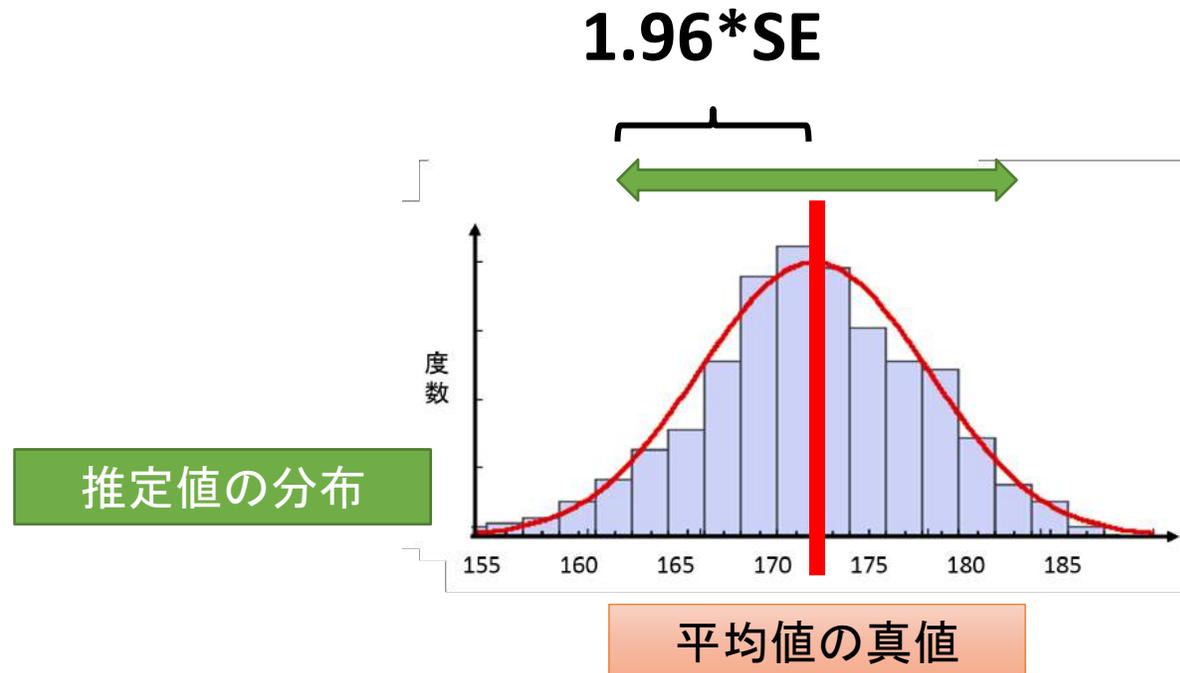
正規分布の性質（再掲）

- 平均値 $\pm 2 * SD$ の区間に95.5%のデータが含まれる
 - 平均値 $\pm 1.96 * SD$ の区間に95%のデータが含まれる



点推定した値 → 真値とは一致しない

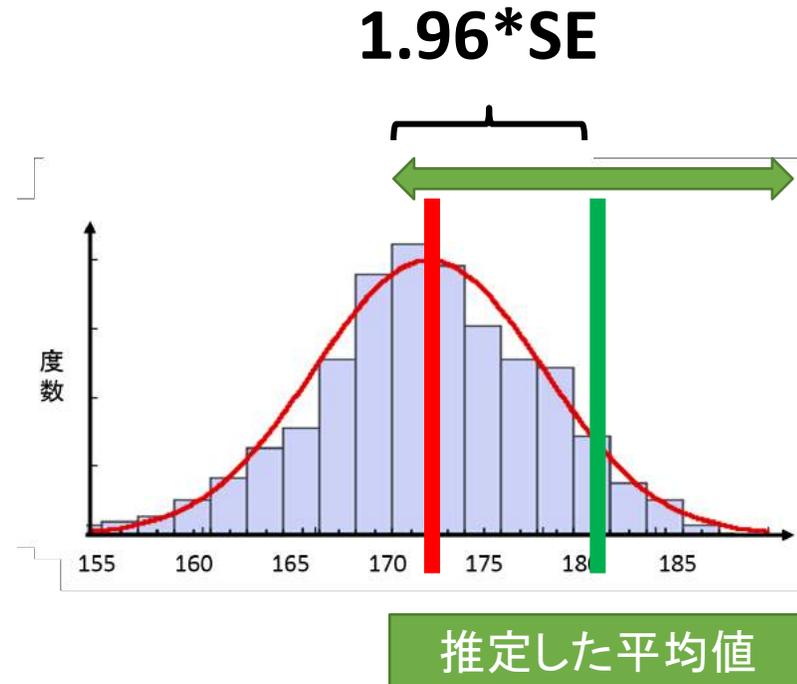
- ある程度の幅をもって「この辺に真値がありそう」という場所を考える
- 推定値のばらつき：標準誤差（SE）で表す
 - 推定した平均値のうち95%は、**平均値の真値 $\pm 1.96 * SE$** の範囲にある



推定した平均値を基準にすると

■ 推定した平均値 $\pm 1.96*SE$ の区間

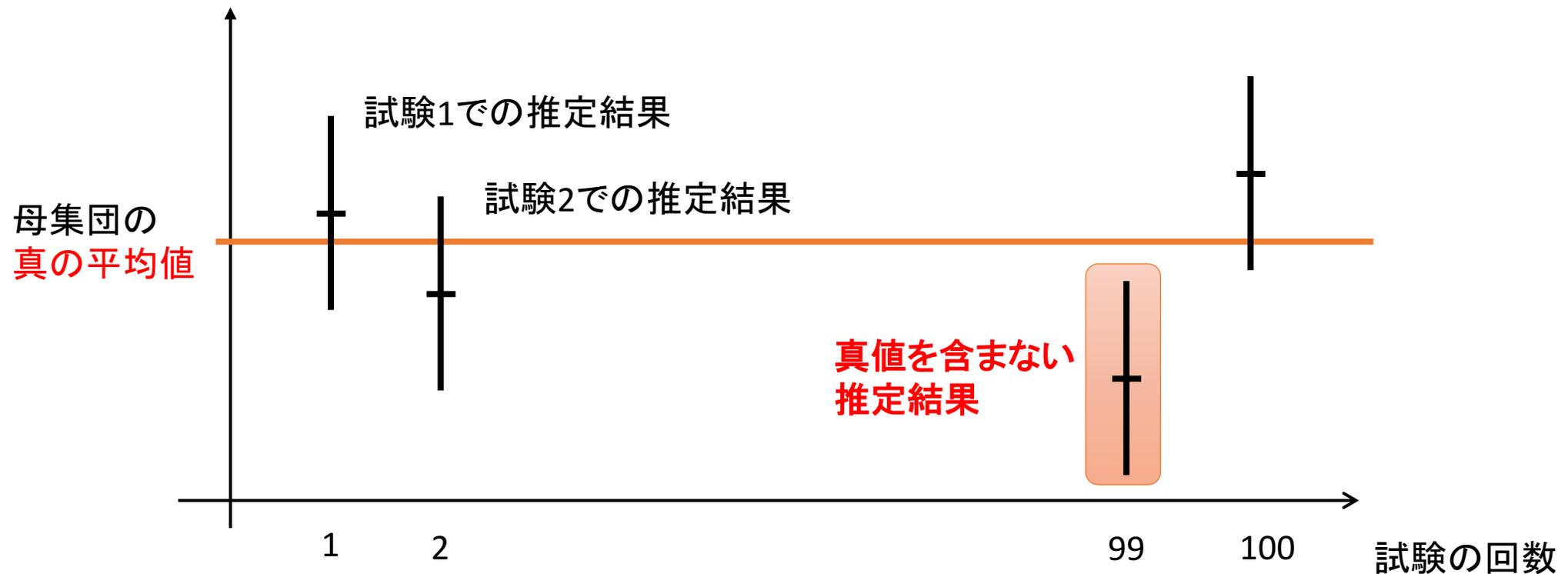
□ 100回中95回は平均値の真値を含んでいるはず → 95%信頼区間



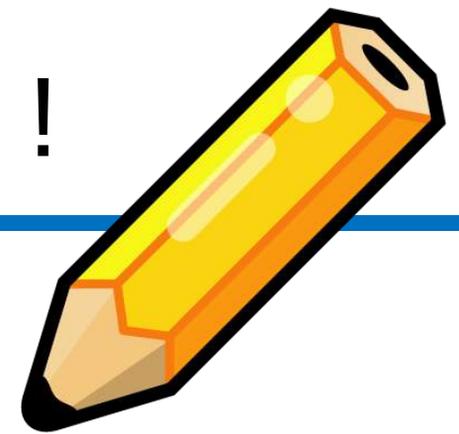
注意：95%信頼区間の解釈

■Q：解釈で正しいのはどっち？

1. 95%信頼区間の中に、95%の確率で真値がある
2. 100回中95回の95%信頼区間は真値を含んでいる



データを読み解いてみましょう！



	A群	B群
	N=291	N=288
Sex		
Male	157 (54%)	170 (59%)
Female	134 (46%)	118 (41%)
n	291	288

Height in cm		
Mean	167.7	168.1
SD	9.49	9.37
SEM	0.56	0.55
Median	168.1	168.3
Min-Max	140 - 192	144 - 189
n	291	288

点推定値

_____ cm

95%信頼区間

_____ ~ _____ cm

まとめ (1/2)

- 平均値の他に、中央値をぜひ知ってください
- 標準**偏差**は個々のデータの、標準誤差は推定地のばらつきを表す
 - 成績 = **偏差**値は、学生の中での自分の順位を表すもの
- 95%信頼区間 = 「100回試験をしたら95回は真値を含んでいる」
 - 点推定と区間推定の考え方の違いを感じてください

まとめ (2/2)

■ 「推定」の重要ポイント！

□重要ポイント1：データは目的の集団を表しているか

- ✓ 偏った対象から集めてきたデータは無意味

□重要ポイント2：試験から得られる**推定値** = 要約統計量

- ✓ **ばらつきを持つ**結果のうち、偶然出てきた結果でしかない
- ✓ 結果が得られたらどのように判断を行うのか