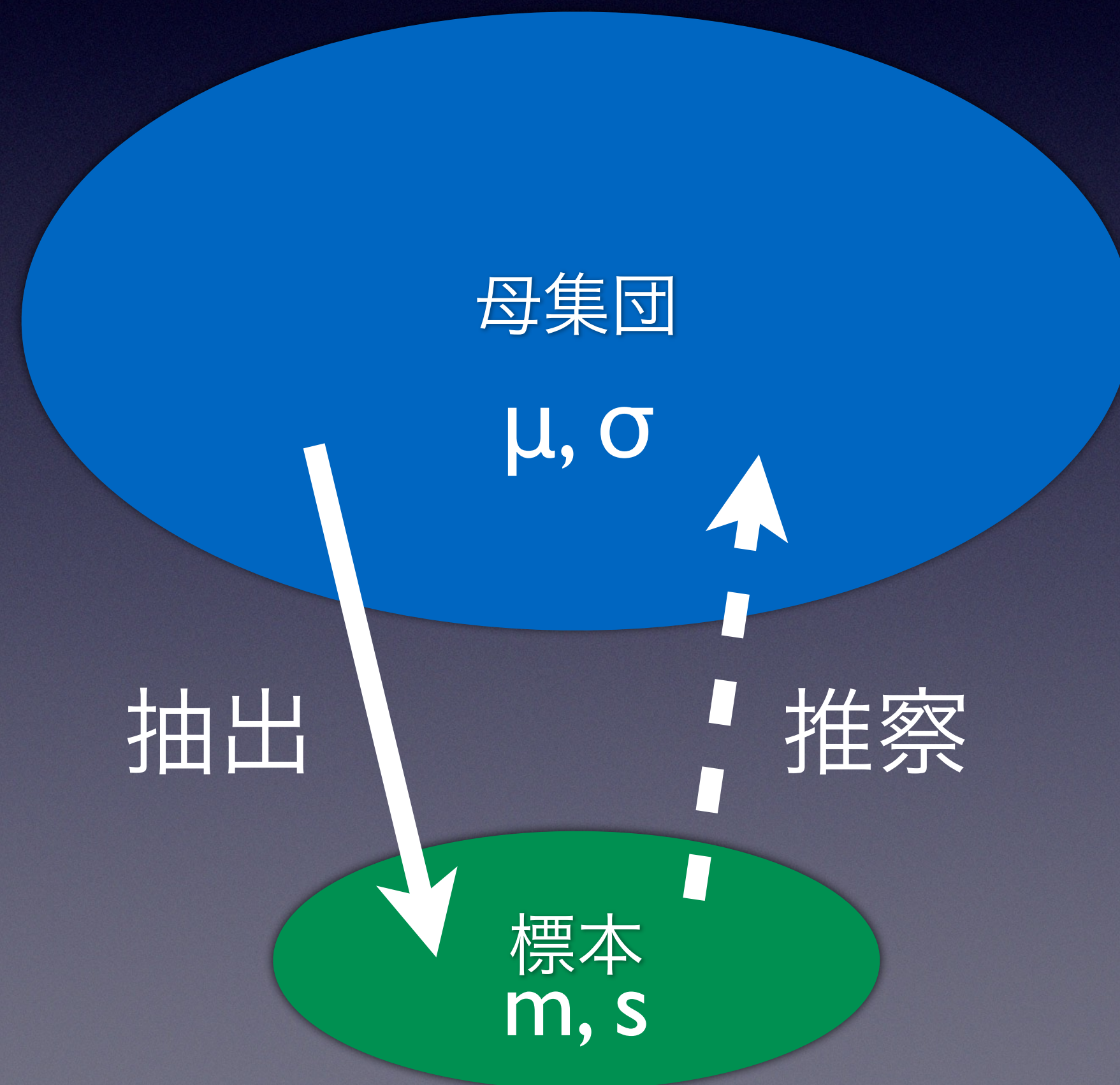


検 定

標本

- 統計の基本：少ないサンプル（標本）で元（全体）を推定



日本人全体の分布
母平均： μ 母分散： σ^2

ランダムにサンプルした分布
標本平均： m 標本分散： s^2

なぜ標本？

- 母平均，母分散は，一般的には簡単に求めることは無理！

日本人の平均身長は？

全員の調査は無理.
サンプルから母集団を推定する.

ある種の疑問

推定する値には
偏りがあるのでは？

偏りのない推定量

- 不偏推定量という。平均は偏りなし。分散は偏りあり。

不偏分散： $n-1$ で割る。

(偏りがなくなるように $n-1$ で割る)

母分散： n で割る。

(n が小さいと偏りが大きい)

平均など
計算

	セット1	セット2	セット3	セット4	セット5	セット6	セット7	セット8	セット9	セット10		
	0.292862859	0.192280118	0.298170761	0.961517878	0.870424714	0.687494851	0.091478915	0.638821884	0.333183591	0.121809438		
	0.807271641	0.629187979	0.766882219	0.828012921	0.167082683	0.335302226	0.673023353	0.711798657	0.552210765	0.500000000		
	0.267396936	0.908540846	0.421670396	0.366337766	0.133115805	0.770307286	0.551299706	0.13699061	0.719495253	0.000000000		
	0.568295419	0.177439853	0.135135948	0.111244078	0.65095788	0.279356857	0.80227509	0.995913906	0.492455823	0.000000000		
	0.578404067	0.958912007	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.075310719	0.900000000		
	0.245042665	0.793218862	0.289000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.754134527	0.000000000		
	0.672192664	0.064032566	0.730000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.912023379	0.000000000		
	0.984286388	0.61731474	0.364000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.610281399	0.800000000		
	0.271767557	0.514851321	0.983000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.358367813	0.200000000		
	0.374814836	0.192468067	0.473000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.537136668	0.700000000		
	0.184492633	0.573512331	0.763000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.992495776	0.068858987		
	0.096518284	0.827338803	0.966000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.365000329	0.431156641		
	0.730052116	0.05912724	0.142000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.499805031	0.617018171		
	0.234706713	0.255902907	0.056000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.373898229	0.43920589		
	0.609839565	0.321940928	0.151111239	0.786536575	0.282025746	0.962778592	0.963375287	0.458776964	0.500299368	0.544643356		
	0.184054151	0.534919279	0.927239663	0.697382349	0.37094995	0.05075963	0.211845922	0.779137861	0.519195472	0.398324723		
	0.868459901	0.15770128	0.758895821	0.132283695	0.251161112	0.867213727	0.368413216	0.967838333	0.570797243	0.59670038		
	0.018688909	0.472806961	0.195522654	0.231736453	0.14070369	0.520178632	0.745166291	0.019542534	0.639044208	0.42835983		
	0.469276614	0.217910181	0.466714335	0.504784205	0.215131752	0.503630048	0.052543674	0.697908679	0.94228836	0.868644206		
	0.529536687	0.8286803	0.899914423	0.299698362	0.421650671	0.315709402	0.41627624	0.898070962	0.728740642	0.153179345		
											10セットの平均	
セットごとの平均	0.44939803	0.464904328	0.489757867	0.539121077	0.427346038	0.536095601	0.464313273	0.506193445	0.57380823	0.471476836	0.492241472	
偏差平方和	1.414747165	1.657846547	2.086299839	1.962308006	1.435947241	1.641430599	1.627522048	1.812249322	0.967479134	1.568716334		
不偏分散(÷19)	0.074460377	0.087255081	0.109805255	0.103279369	0.075576171	0.086391084	0.085659055	0.095381543	0.050919954	0.082564018	0.085129191	
母分散(÷20)	0.070737358	0.082892327	0.104314992	0.0981154	0.071797362	0.08207153	0.081376102	0.090612466	0.048373957	0.078435817	0.080872731	
理論値:	平均	0.5										
	分散	0.083333333		不偏分散との差	0.001795857							
				母分散との差	0.002460602							
	全部から求めた	0.082770761		全部から求めた	0.000562573							
	偏差平方和	16.55415214										

乱数を発生
=rand()
で[0.0, 1.0]の一様乱数

=devsq
で偏差平方和を計算
(varは使わない)

	セット1	セット2	セット3	セット4	セット5	セット6	セット7	セット8	セット9	セット10		
	0.292862859	0.192280118	0.298170761	0.961517878	0.870424714	0.687494851	0.091478915	0.638821884	0.333183591	0.121809438		
	0.807271641	0.629187979	0.766882219	0.828012921	0.167082683	0.335302226	0.673023353	0.711798657	0.552210765	0.5		
	0.267396936	0.908540846	0.421670396	0.366337766	0.133115805	0.770307286	0.551299706	0.13699061	0.719495253	0		

ある確率分布から生まれる x の分散は、

c 上限

=devsq
で偏差平方和を計算
(=分散ではない)

結果として
不偏分散で計算して推定した分散の値が理論値に近い

平均な
計算

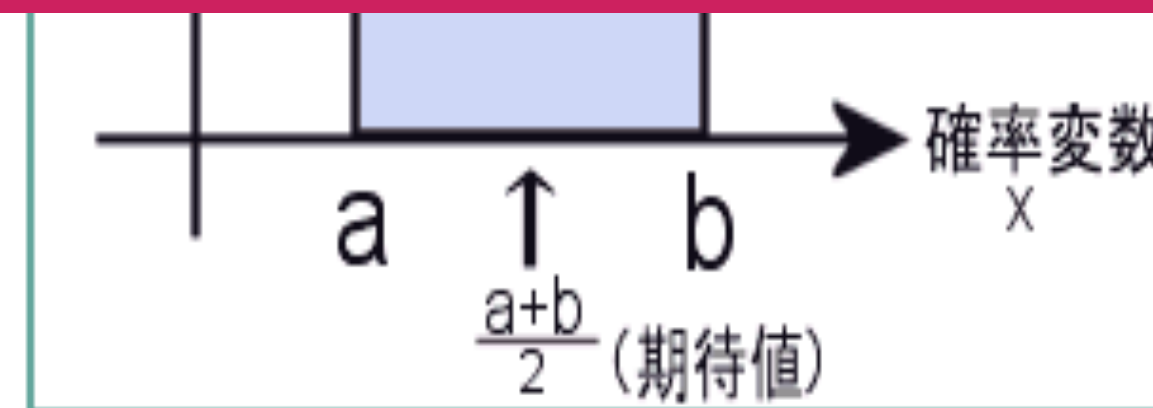
セットごとの平均
偏差平方和
不偏分散(÷19)
母分散(÷20)

理論値:

$$\text{一様分布の分散} = \frac{(b-a)^2}{12}$$

[0,1] から発生する一様乱数の場合は、

$$a=0, b=1 \text{ を代入して分散} = \frac{1}{12} = 0.083333\dots$$



71476836	0.492241472
68716334	
82564018	0.085129191
78435817	0.080872731

エクセル講座になるので

やりませんが

自分ではやってみてください

なぜ分散の推定にこだわる？

- 分散は何をしめしていたか？

データのばらつきを表す基本的な値

なぜ分散の推定にこだわる？

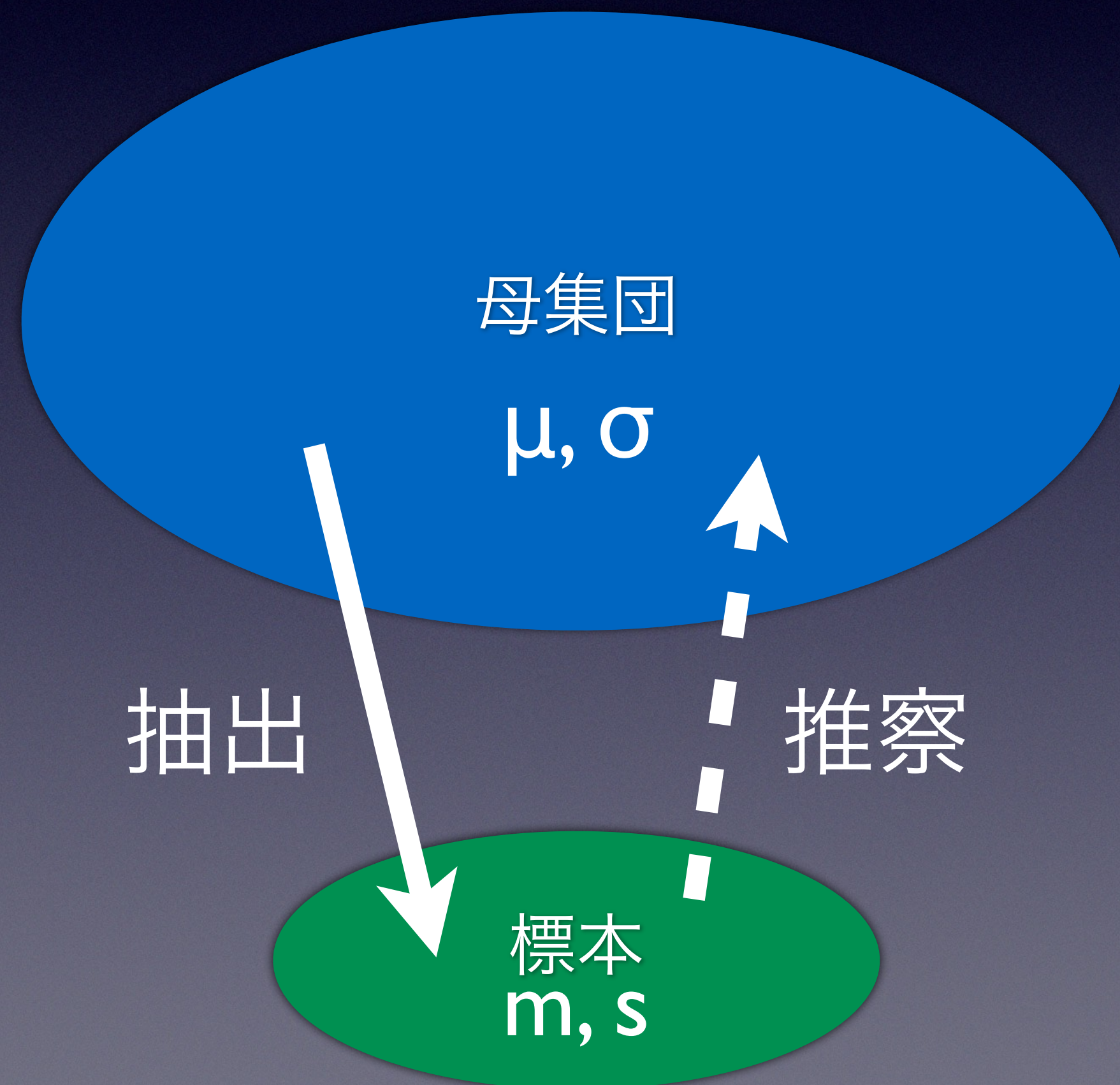
- 分散は何をしめしていたか？

データのばらつきを表す基本的な値

この値がぶれると全部がおしやか

標本

- 統計の基本：少ないサンプル（標本）で元（全体）を推定

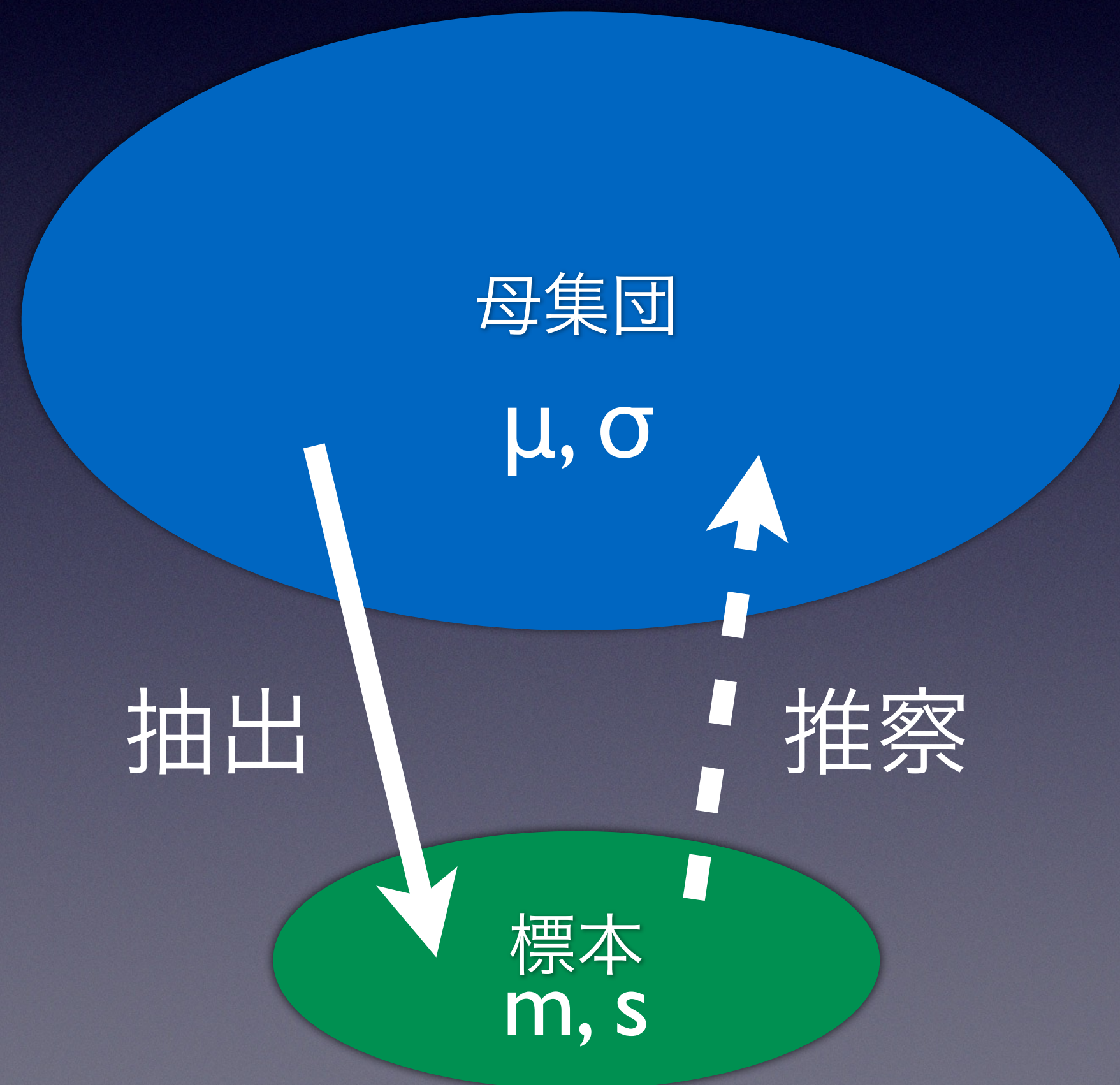


日本人全体の分布
母平均： μ 母分散： σ^2

ランダムにサンプルした分布
標本平均： m 標本分散： s^2

標本

- 統計の基本：少ないサンプル（標本）で元（全体）を推定



日本人全体の分布
母平均： μ 母分散： σ^2

ランダムにサンプルした分布
標本平均： m 標本分散： s^2

標本

- 統計の基本：少ないサンプル（標本）で元（全体）を推定



母平均の推定

標本の大きさが大きいときには（通常 $n > 30$ くらい）

母平均： μ の信頼区間は

95%信頼区間 $m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$

99%信頼区間 $m - 2.58 \sigma / \sqrt{n} \leq \mu \leq m + 2.58 \sigma / \sqrt{n}$

m : 標本平均

σ : 母標準偏差

矛盾

母平均を知りたいのに母分散の値を使う必要がある？

→ 母分散の代わりにその推定量として

標本の不偏分散を用いる

(標準偏差² = 分散)

演習

- 大量の魚から100匹の標本を抽出して測定したところ、標本平均：150g、標本標準偏差：30 gであった。母平均の95%信頼区間を求めなさい。



解答例

- 母分散（母標準偏差）が分からないので、標本標準偏差を代わりに用いる。計算方法は、不偏分散のやり方。
- $m = 150$, $\sigma = 30$, $n = 100$

$$\text{95\%信頼区間} \quad m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$$

$$150 - 1.96 \cdot 30 / \sqrt{100} \leq \mu \leq 150 + 1.96 \cdot 30 / \sqrt{100}$$

$$150 - 1.96 \cdot 30 / 10 \leq \mu \leq 150 + 1.96 \cdot 30 / 10$$

$$150 - 1.96 \cdot 3 \leq \mu \leq 150 + 1.96 \cdot 3$$

$$144.1 \leq \mu \leq 155.9$$

解答例

95%の確率で母平均が144.1 [g]から155.9 [g]の間にある

- $m = 150$, $\sigma = 30$, $n = 100$

$$95\% \text{信頼区間} \quad m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$$

$$150 - 1.96 \cdot 30 / \sqrt{100} \leq \mu \leq 150 + 1.96 \cdot 30 / \sqrt{100}$$

$$150 - 1.96 \cdot 30 / 10 \leq \mu \leq 150 + 1.96 \cdot 30 / 10$$

$$150 - 1.96 \cdot 3 \leq \mu \leq 150 + 1.96 \cdot 3$$

$$144.1 \leq \mu \leq 155.9$$

演習

- 別の磯へ釣りに出かけました.
- 連れた大量の魚から100匹の標本を抽出して測定したところ, 標本平均: 155g, 標本標準偏差: 30 gであった. 母平均の95%信頼区間を求めなさい.



解答例

- 母分散（母標準偏差）が分からないので，標本標準偏差を代わりに用いる．計算方法は，不偏分散のやり方．
- $m = 155$, $\sigma = 30$, $n = 100$

$$\text{95\%信頼区間} \quad m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$$

$$155 - 1.96 \cdot 30 / \sqrt{100} \leq \mu \leq 155 + 1.96 \cdot 30 / \sqrt{100}$$

$$155 - 1.96 \cdot 30 / 10 \leq \mu \leq 155 + 1.96 \cdot 30 / 10$$

$$155 - 1.96 \cdot 3 \leq \mu \leq 155 + 1.96 \cdot 3$$

$$149.1 \leq \mu \leq 160.9$$

解答例

95%の確率で母平均が149.1 [g]から160.9 [g]の間にある

- $m = 155$, $\sigma = 30$, $n = 100$

$$95\% \text{信頼区間} \quad m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$$

$$155 - 1.96 \cdot 30 / \sqrt{100} \leq \mu \leq 155 + 1.96 \cdot 30 / \sqrt{100}$$

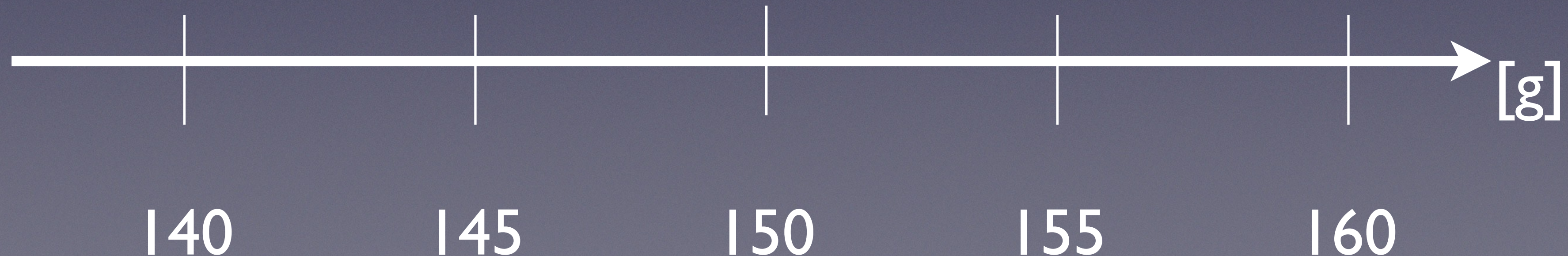
$$155 - 1.96 \cdot 30 / 10 \leq \mu \leq 155 + 1.96 \cdot 30 / 10$$

$$155 - 1.96 \cdot 3 \leq \mu \leq 155 + 1.96 \cdot 3$$

$$149.1 \leq \mu \leq 160.9$$

区間推定から検定へ

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]
(いずれも95%信頼区間)



区間推定から検定へ

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]
(いずれも95%信頼区間)



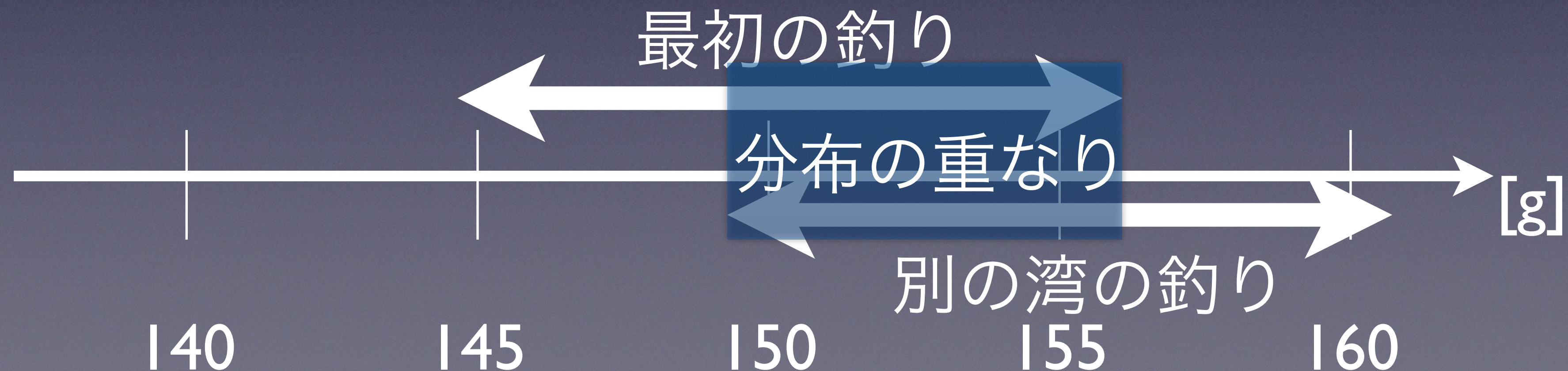
区間推定から検定へ

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]
(いずれも95%信頼区間)



区間推定から検定へ

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]
(いずれも95%信頼区間)

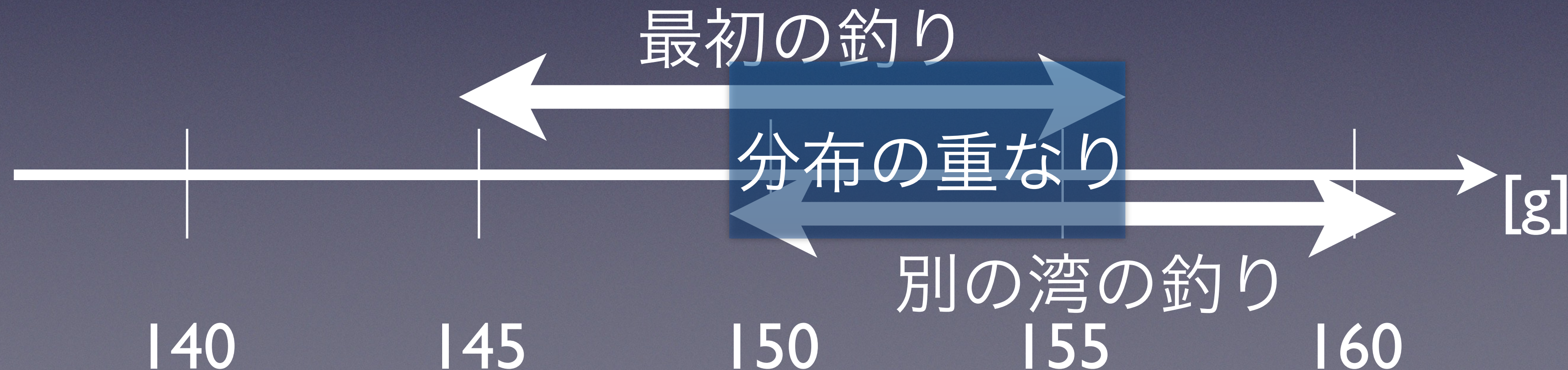


区間推定から検定へ

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]

2つの差はあるとは言えない
(できごとが起こる範囲が重なる)

(いずれも95%信頼区間)



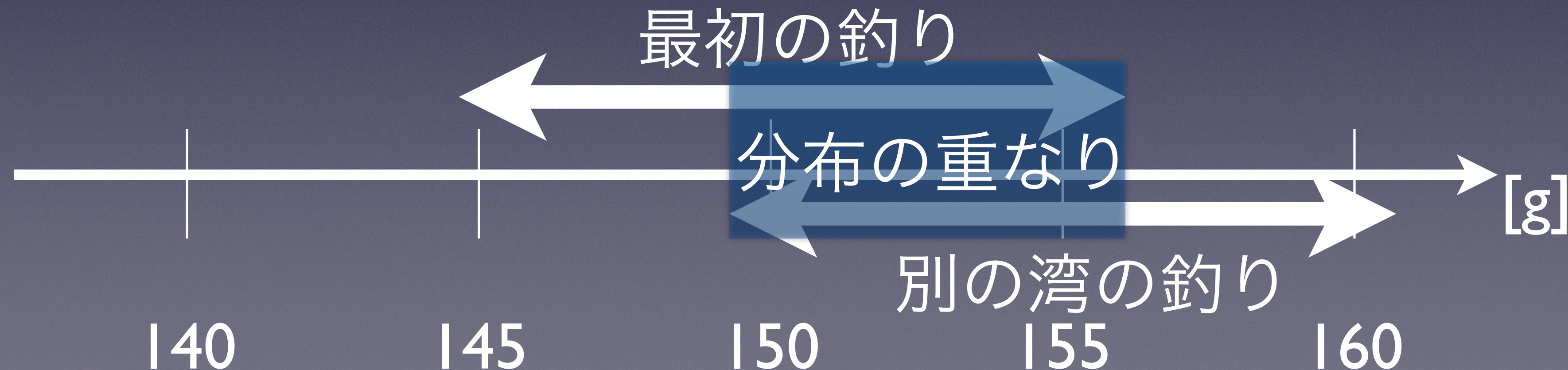
区間推定から

対応のない
2群間の母平均の検定

- 最初の釣りの成果 : $n=100, m=150, \sigma=30$ [144.1, 155.9]
- 別の湾の成果 : $n=100, m=155, \sigma=30$ [149.1, 160.9]

2つの差はあるとは言えない
(できごとが起こる範囲が重なる)

(いずれも95%信頼区間)



演習

- 麹菌Aを使ったとき、100回の醸造で一定量の精米から1回あたり平均20gのアルコールができた。
- 麹菌Bを使ったとき、100回の醸造で同量の精米から1回あたり平均24gのアルコールができた。
- 別の調査で、それぞれのアルコール生成量には8gの標準偏差があることが分かっている。
- 麹菌A, Bに違いがあるといえるか？



95%信頼区間

$$m - 1.96 \sigma / \sqrt{n} \leq \mu \leq m + 1.96 \sigma / \sqrt{n}$$

麹菌A	$20 - 1.96 \cdot 8 / \sqrt{100}$	$\leq$	μ	$\leq$	$20 + 1.96 \cdot 8 / \sqrt{100}$
	$20 - 1.96 \cdot 0.8$	$\leq$	μ	$\leq$	$20 + 1.96 \cdot 0.8$
	18.2	$\leq$	μ	$\leq$	21.6

麹菌B	$24 - 1.96 \cdot 8 / \sqrt{100}$	$\leq$	μ	$\leq$	$24 + 1.96 \cdot 8 / \sqrt{100}$
	$24 - 1.96 \cdot 0.8$	$\leq$	μ	$\leq$	$24 + 1.96 \cdot 0.8$
	22.4	$\leq$	μ	$\leq$	25.6

対応がある場合には？

- 酸素の缶で補給しながら登った場合と、補給なしで登った場合で、差がでるか？
- 同一人物でそれぞれ計測すれば対応あり。



対応あり

	酸素缶なし	酸素缶あり	Δ
登山者 1	202min	190min	
登山者 2	198min	182min	
登山者 3	188min	191min	
登山者 4	244min	205min	
登山者 5	232min	215min	
登山者 6	209min	189min	
登山者 7	192min	181min	
平 均			

差の母平均で検討できる：

差の母平均の信頼区間に 0 が含まれれば差があるとは言えない

エクセルでやってみる

登山者	酸素缶なし	酸素缶あり			
1	202	190	=TTEST(C3:C9,D3:D9,2,1) TTEST(配列1, 配列2, 尾部, 検定の種類)		
2	198	182			
3	188	191			
4	244	205			
5	232	215			
6	209	189			
7	192	181			

関数：ttest
確率が返る

尾部：

- 1 片側検定
- 2 両側検定

検定の種類：

- 1 対をなすデータのt検定
- 2 等分散を仮定した2標本の検定
- 3 非等分散を仮定した2標本の検定

データとは？ 質 か 量？

		比較方法	例
質的 データ	名義尺度 (Nominal scale) 分類のための符号. 同一性のみを示す	最頻値	品番／型番
	順序尺度 (Ordered scale) 比較のための符号. 同一性と順序を示す	中央値	スコア (A, B, C...) , ステージ
量的 データ	間隔尺度 (Interval scale) 間隔の関係を説明する. 0が定まらない.	平均と	温度
	比率尺度 (Ratio Scale) 比の関係を説明する. 0が定まる.	標準偏差	体重, 圧力, 連続確信度

順序尺度か比率尺度か？

- 10段階評価を1 cm間隔でマーク 順序尺度 > 比率尺度
質的データ
- 100段階評価を1 mm間隔でマーク 比率尺度 > 順序尺度
量的データ

質の検定 量の検定

質の検定 ノンパラメトリック検定

データの分布に依存しない検定方法.

データを順位に置き換えて, 平均順位を比較する.

量の検定 パラメトリック検定

データの分布に依存する検定方法.

データの平均値を直接比較する.

分布

分散

対象とする変数

ノンパラメトリック
(質の検定)

分布に
依存しない

等分散で
なくとも
よい

順序変数 (離散), 対数
0 が定まらない数
(体温, pH)

パラメトリック
(量の検定)

正規分布

等分散

間隔尺度, 比率尺度
(連続変数)

AUCの検定か、 確信度から得られる分布の検定か？

	量の検定 parametric test	質の検定 non-parametric test
対応のない 2 群の検定	スチューデントt検定 Student's t-test	マン・ホイットニー検定 Mann-Whitney's U test (Wilcoxon rank sum test)
対応のある 2 群の検定	対応のあるt検定 Paired t-test	ウィルコクソン符号付き 順位検定 Wilcoxon signed-rank test

ROC曲線下面積とその検定

面積は量のデータ. しかし, 観察者の数は少ない!

方針：AUCの変化を調べる

(もう一度)

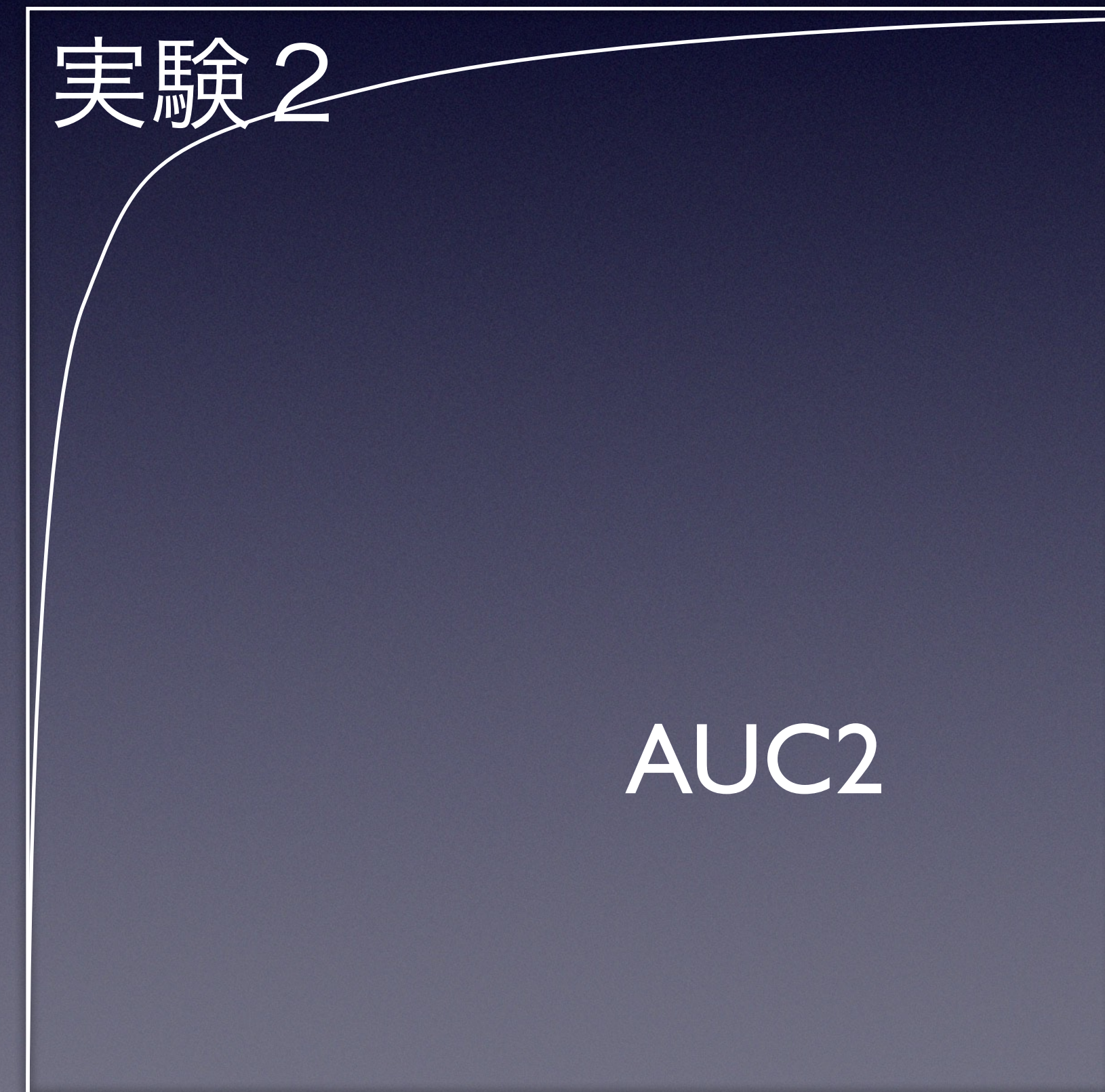
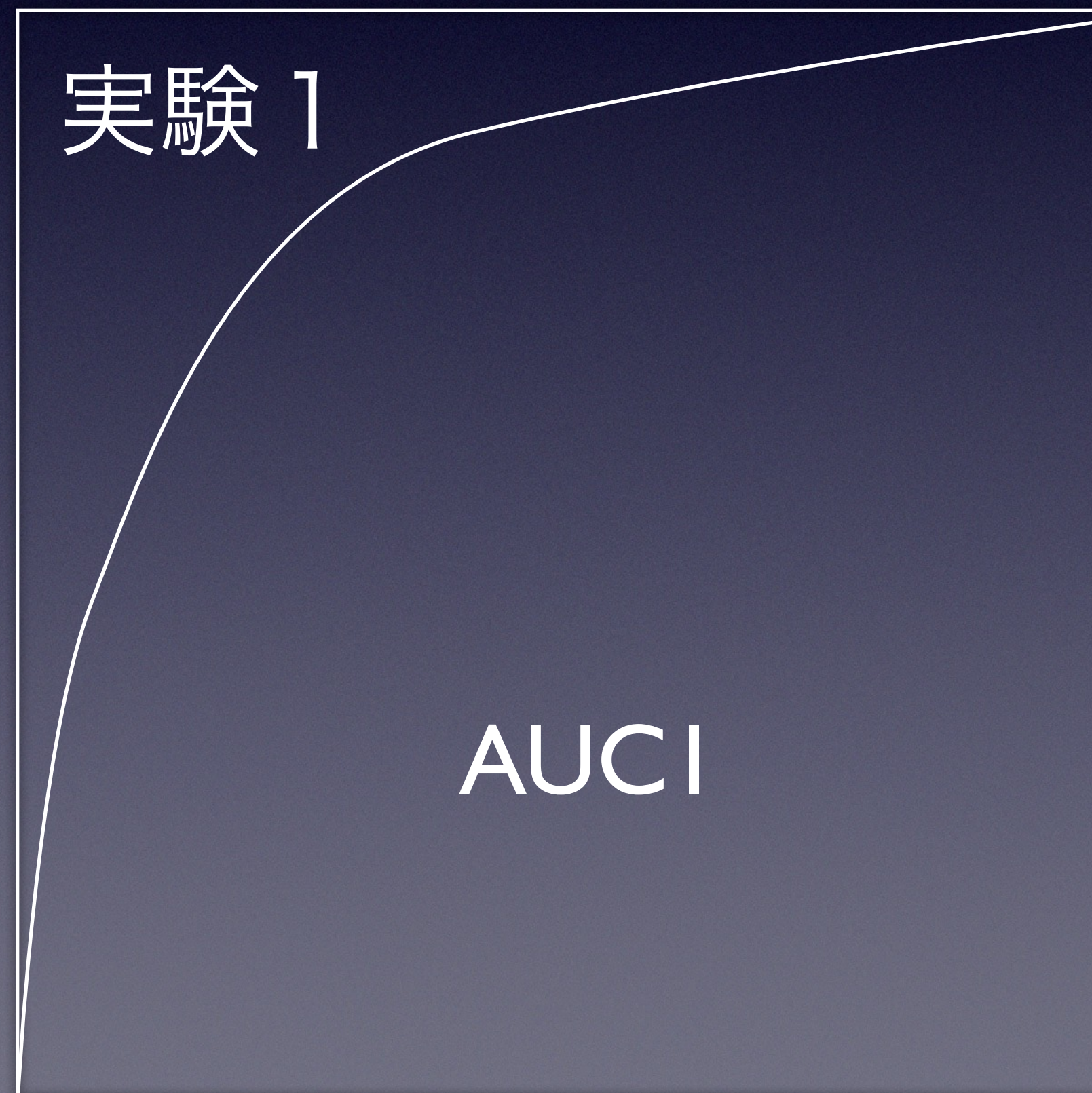
- システム1とシステム2の結果について調べる：2群
- 同一の観察者が実験してその差を調べる：同一の2群の差
- 面積なので「量」の検定でやりたい

1人の観察者の結果を比較する

2つのAUCを比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる

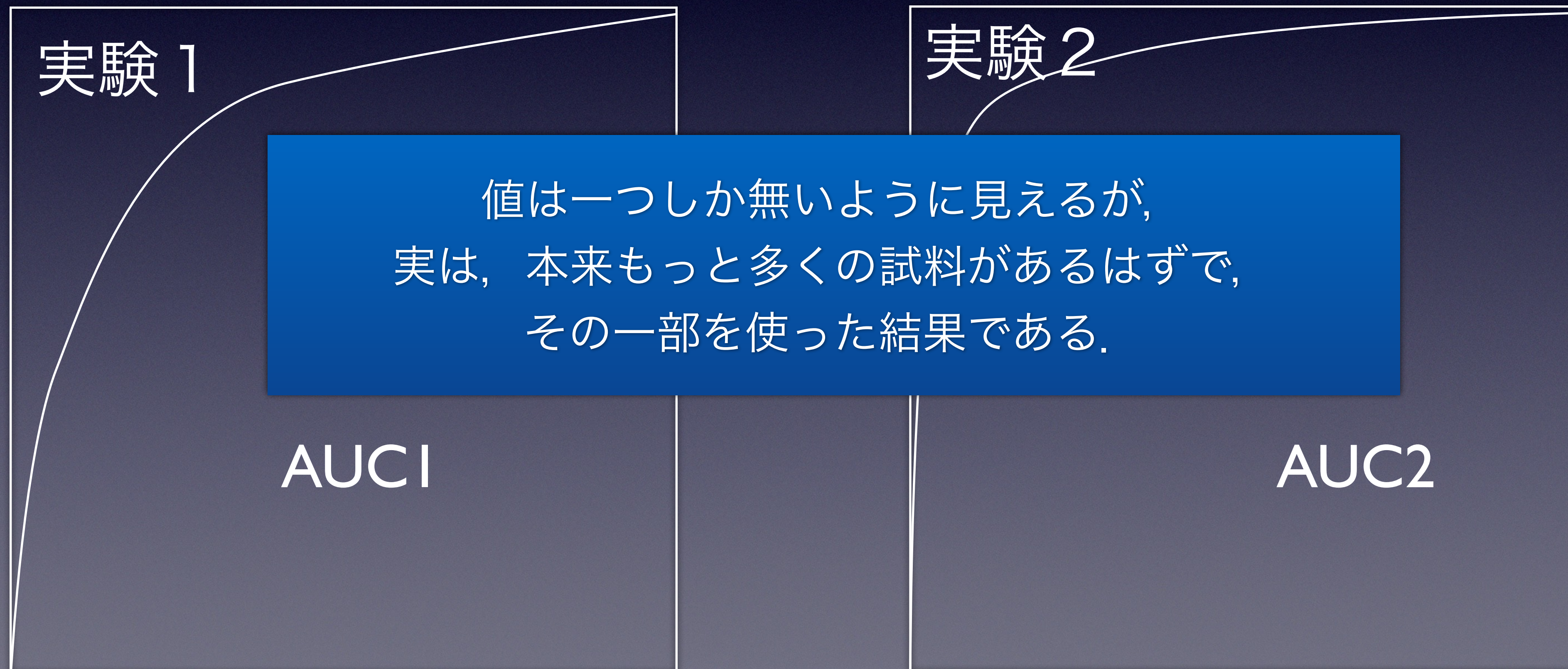


1人の観察者の結果を比較する

2つのAUCを比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



1人の観察者の結果を比較する

すなわち、母集団から標本を作り、
そこで得られたAUCであるので、このAUCには変動が含まれる。
したがって、このAUCは標本AUCとも言える。
ならば、この標本AUCから母AUCを推定（区間推定）してやろう

実験 1

AUC1

実験 2

AUC2

値は一つしか無いように見えるが、
実は、本来もっと多くの試料があるはずで、
その一部を使った結果である。

1人の観察者の結果を比較する

すなわち、母集団から標本を作り、
そこで得られたAUCであるので、このAUCには変動が含まれる。
したがって、このAUCは標本AUCとも言える。
ならば、この標本AUCから母AUCを推定（区間推定）してやろう

実験 1

実験 2

値は一つしか無いように見えるが、
実は、本来もっと多くの試料があるはずで、
その一部を使った結果である。

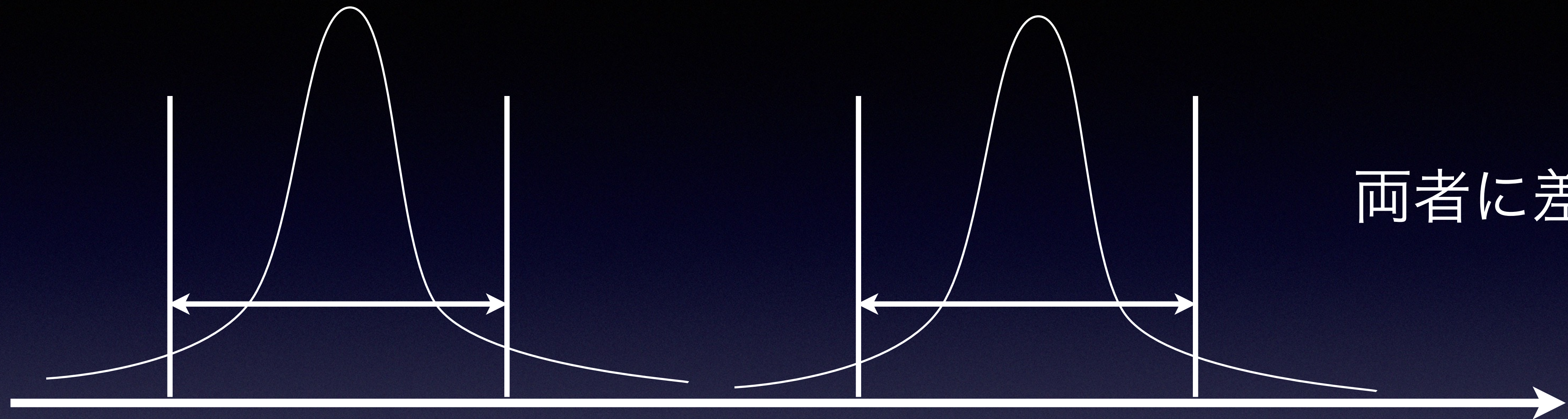
AUC1

AUC2

そのためには、確信度を様々な方法でリサンプリングして変動を調べて、
精密に、母AUC（実は、AUCの平均の分布）を定めよう

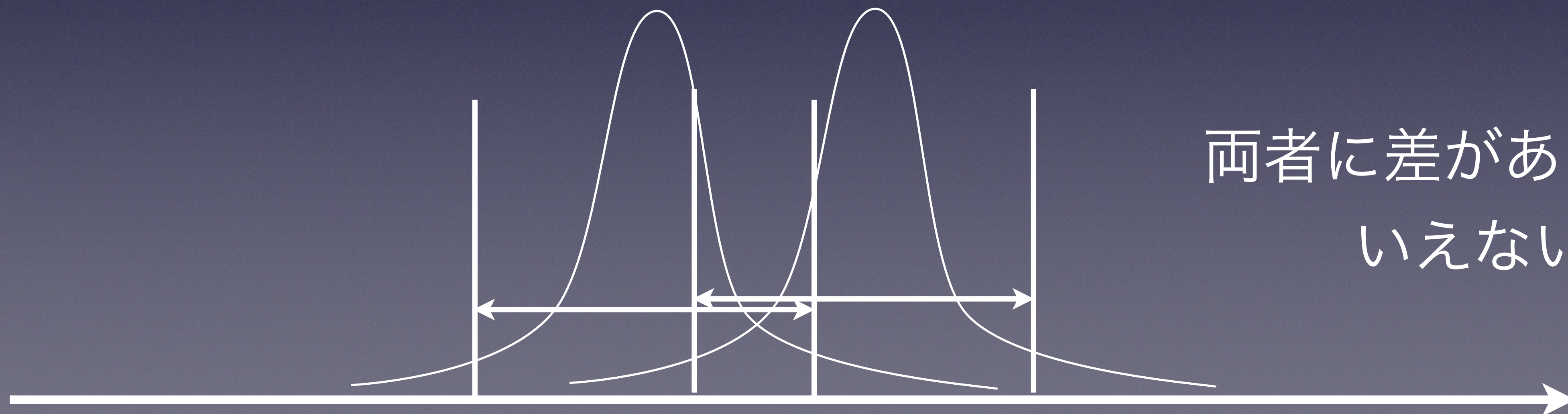
95%信頼区間

95%信頼区間



両者に差があり

95%信頼区間



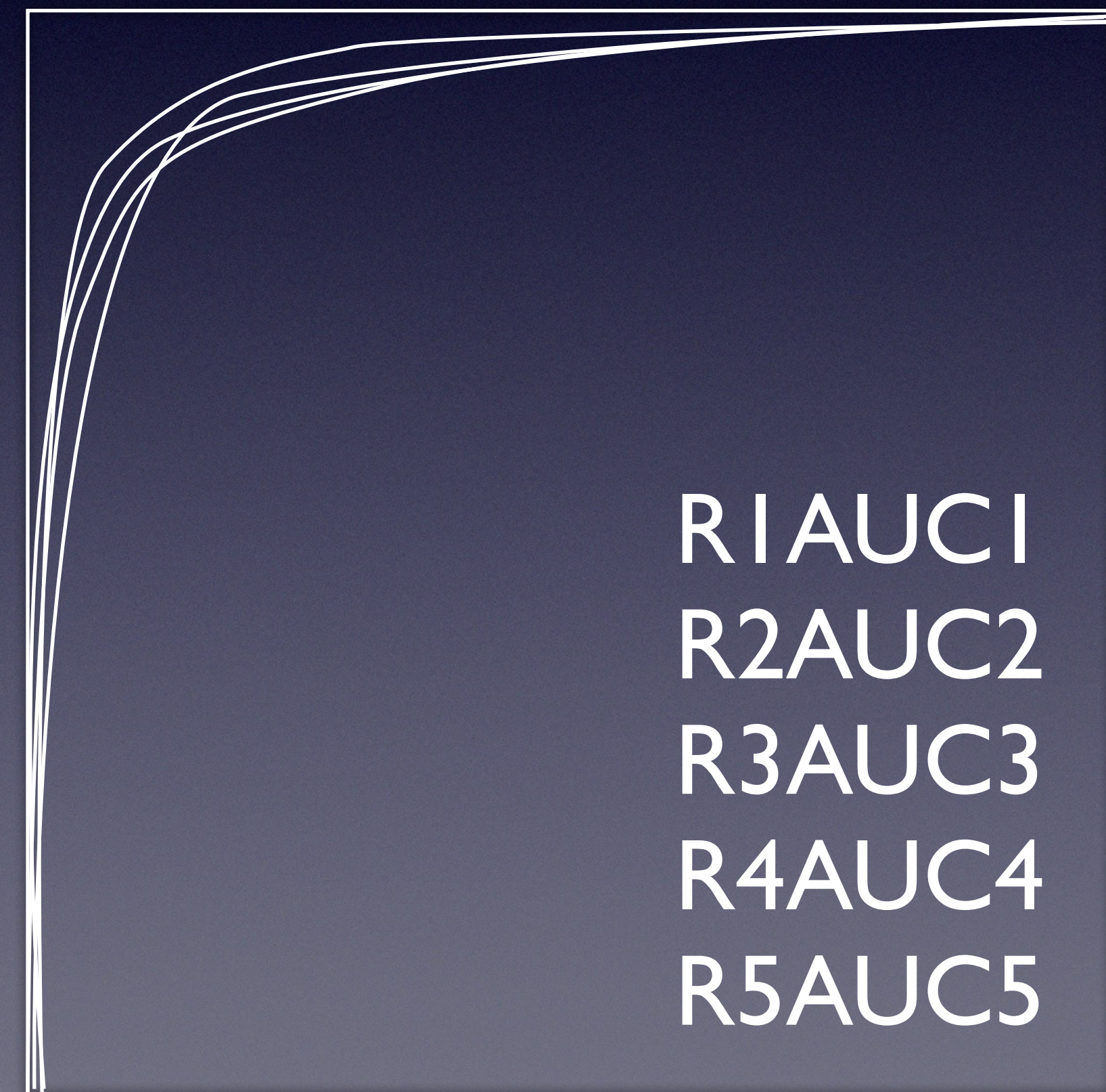
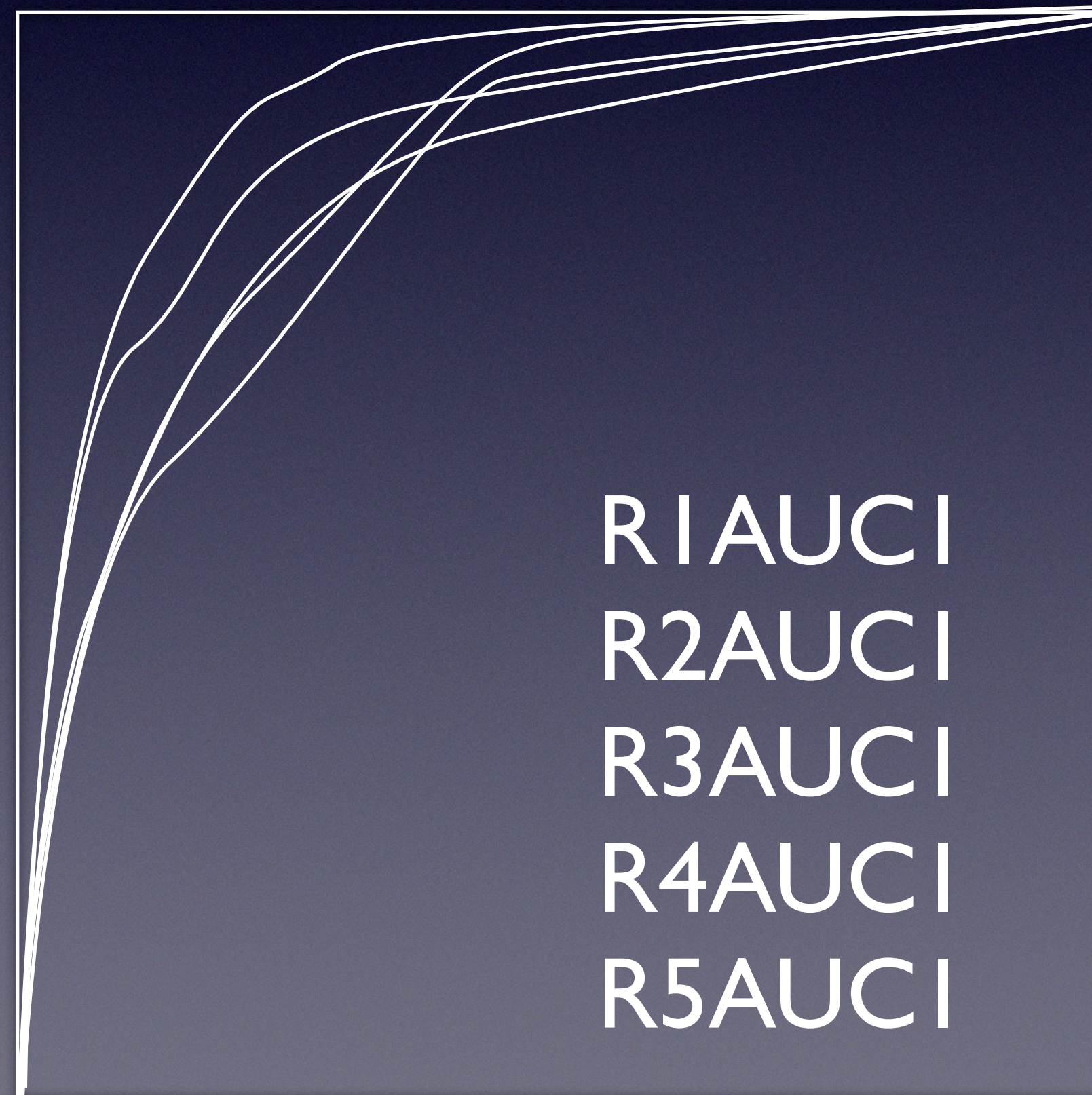
両者に差があるとはいえない

複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる

5人の結果からそれぞれのAUCの平均を求める

R1AUC1
R2AUC1
R3AUC1
R4AUC1
R5AUC1

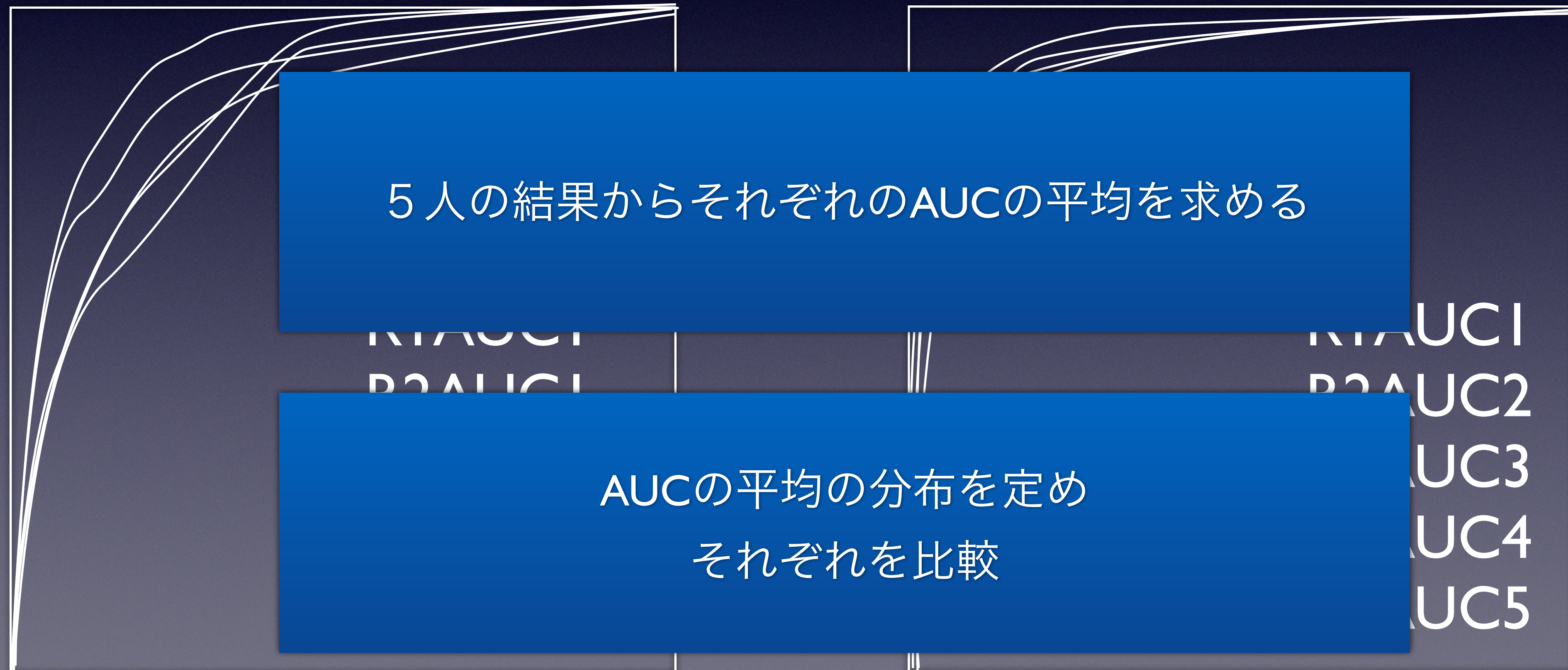
R1AUC1
R2AUC2
R3AUC3
R4AUC4
R5AUC5

複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つ

→それぞれ

それでは,

それをパラメトリックに処理する？

ノンパラメトリックに処理する？

る

5人の結果からそれぞれのAUCの平均を求める

AUCの平均の分布を定め

それぞれを比較

AUC1

AUC2

AUC3

AUC4

AUC5

AUCを検定する

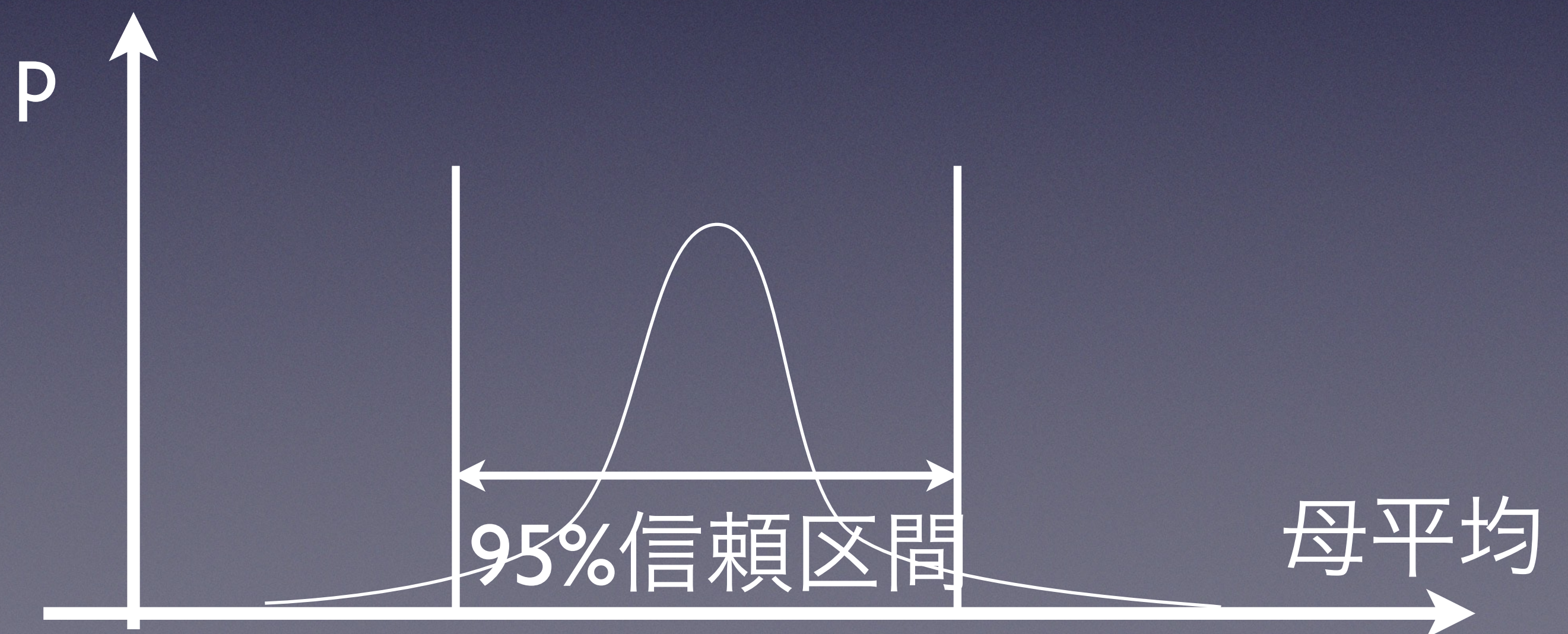
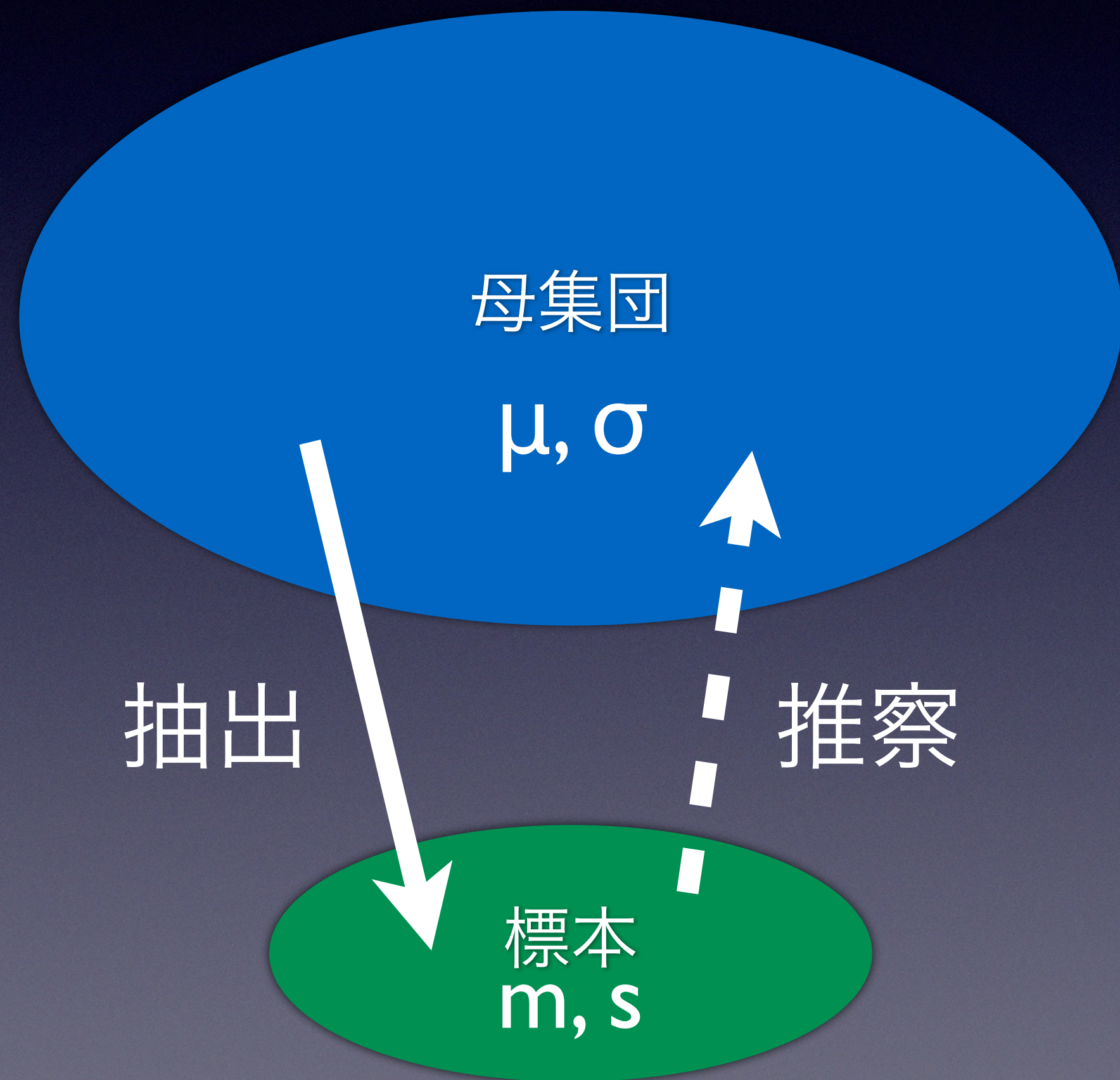
t検定：母平均の検定

標本から母集団の平均（母平均）を推定

→ 母平均の信頼区間を求める

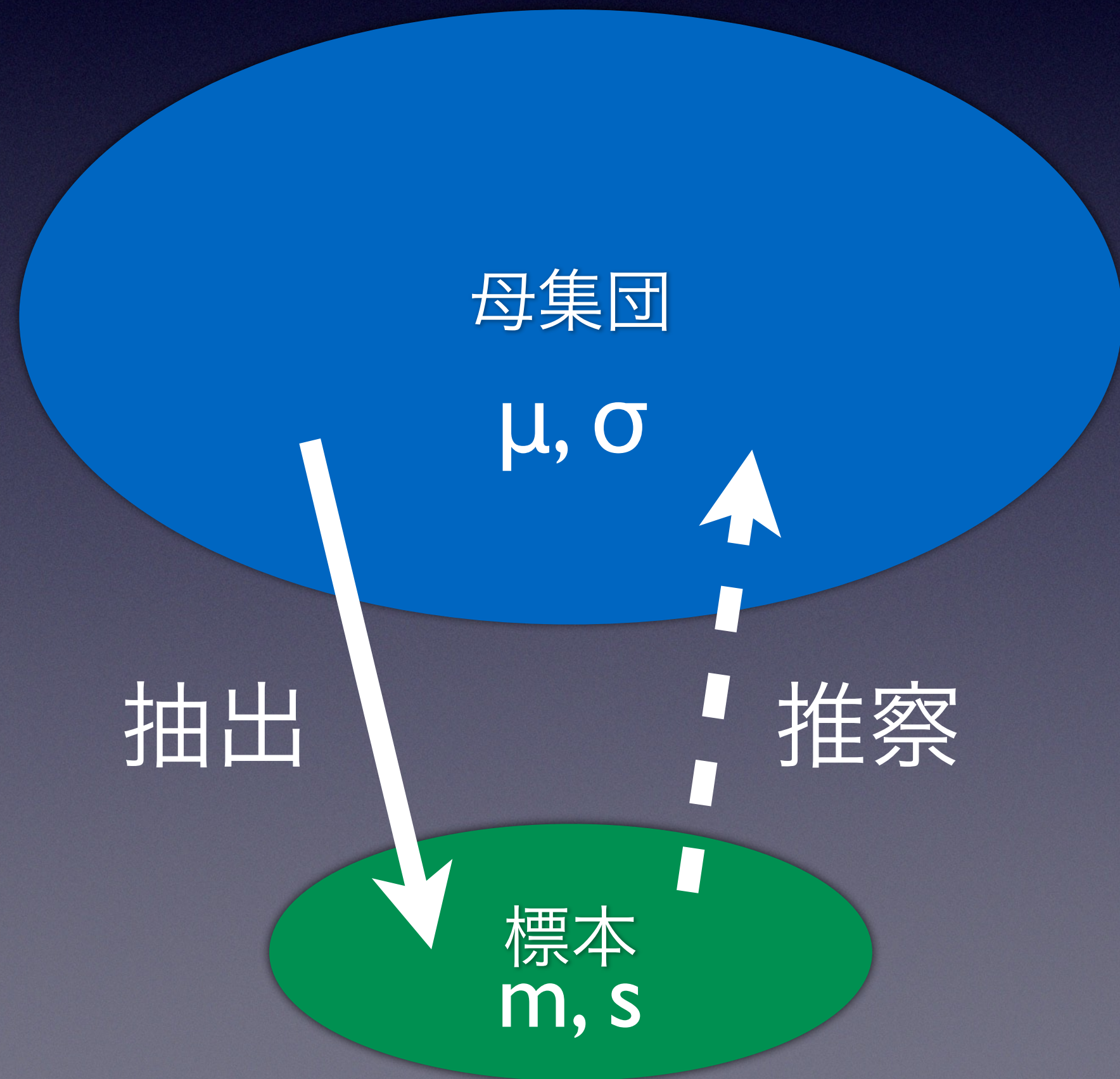
95%信頼区間

→ 95%の確率で母平均は
その区間に存在する



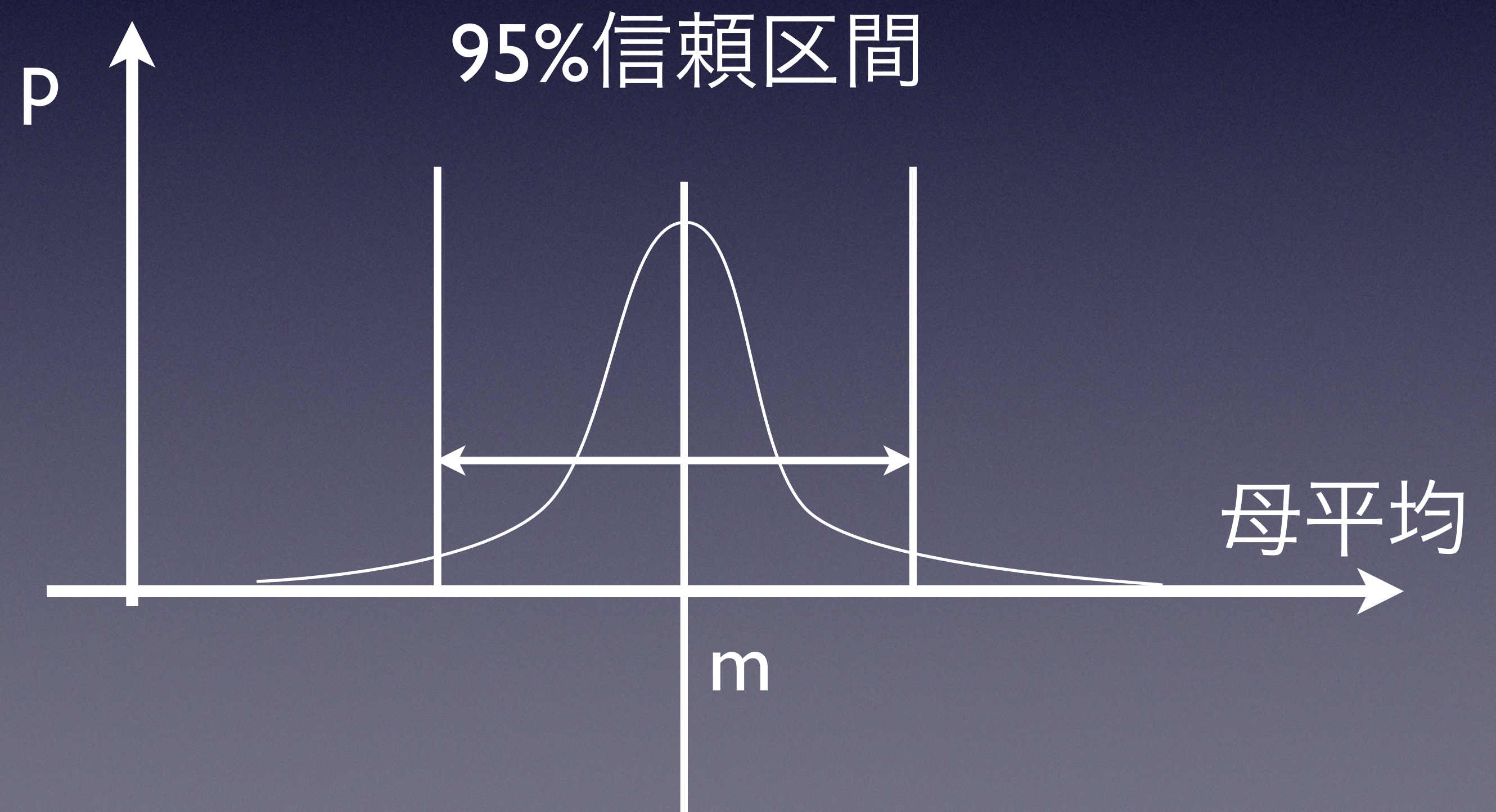
区間推定

平均の分布をパラメトリック
に表現する

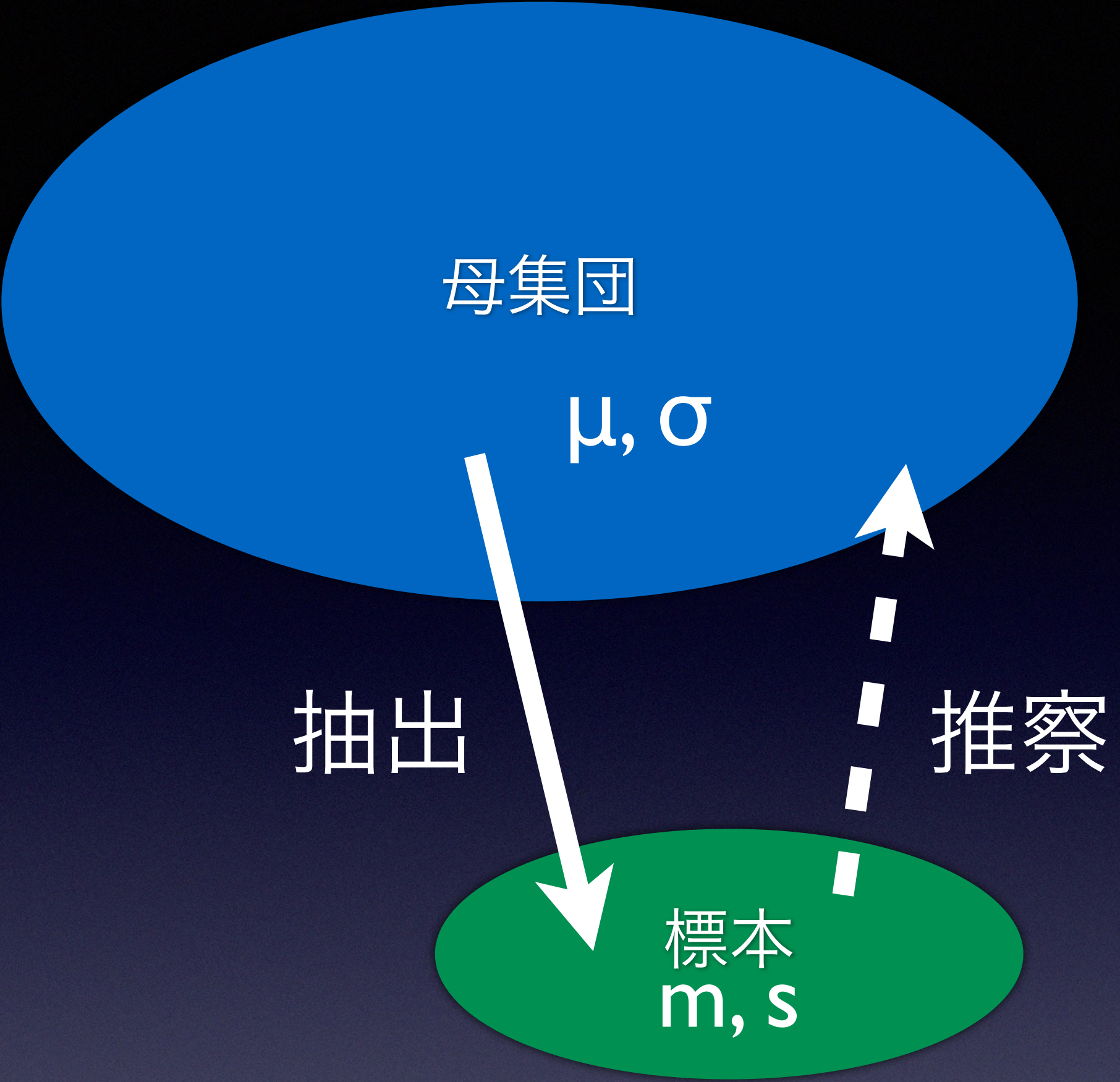


平均値の分布は必ず正規分布になる
(中心極限定理)

なので, その分布は m と s で表せる

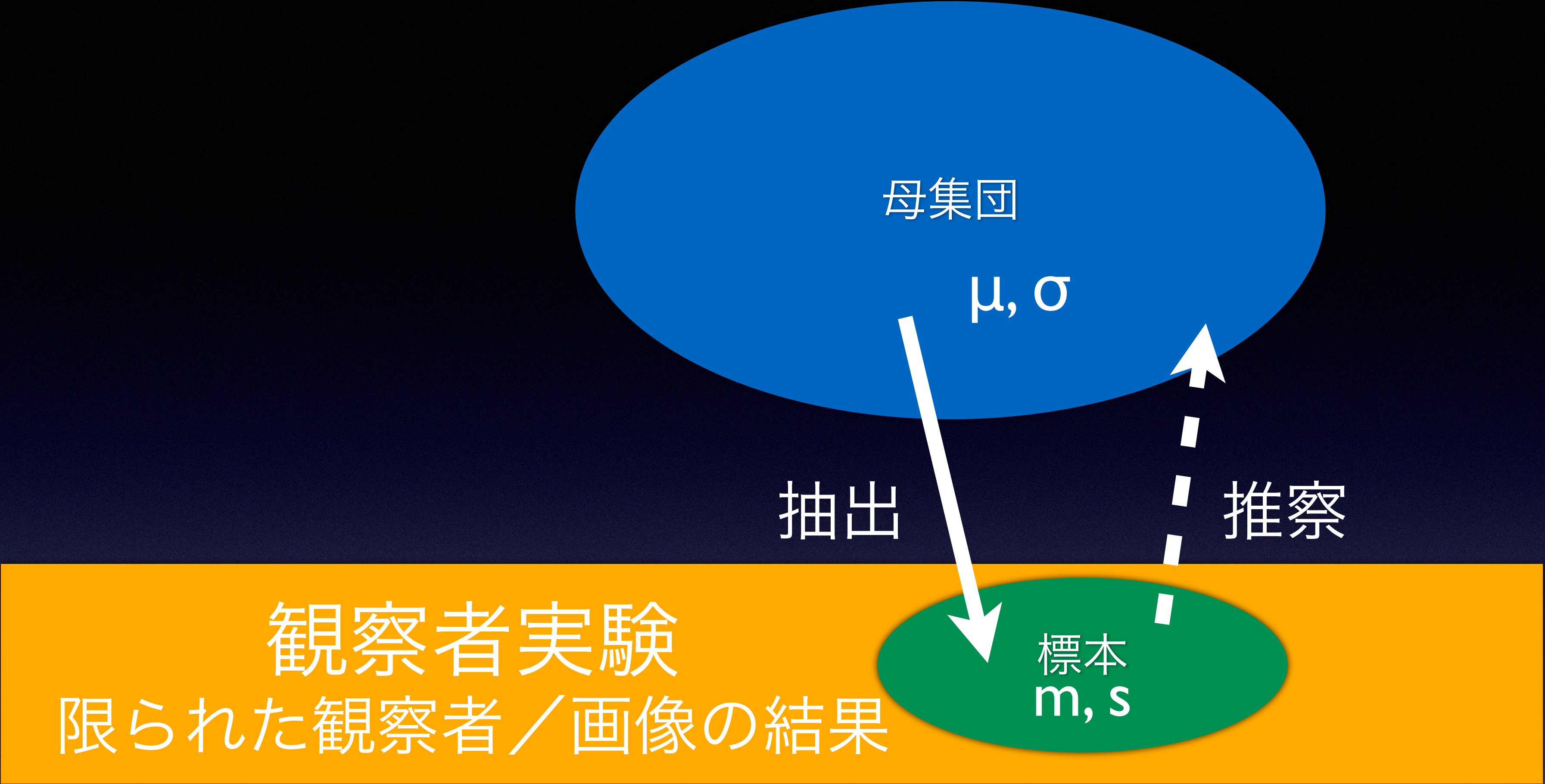


対応あり



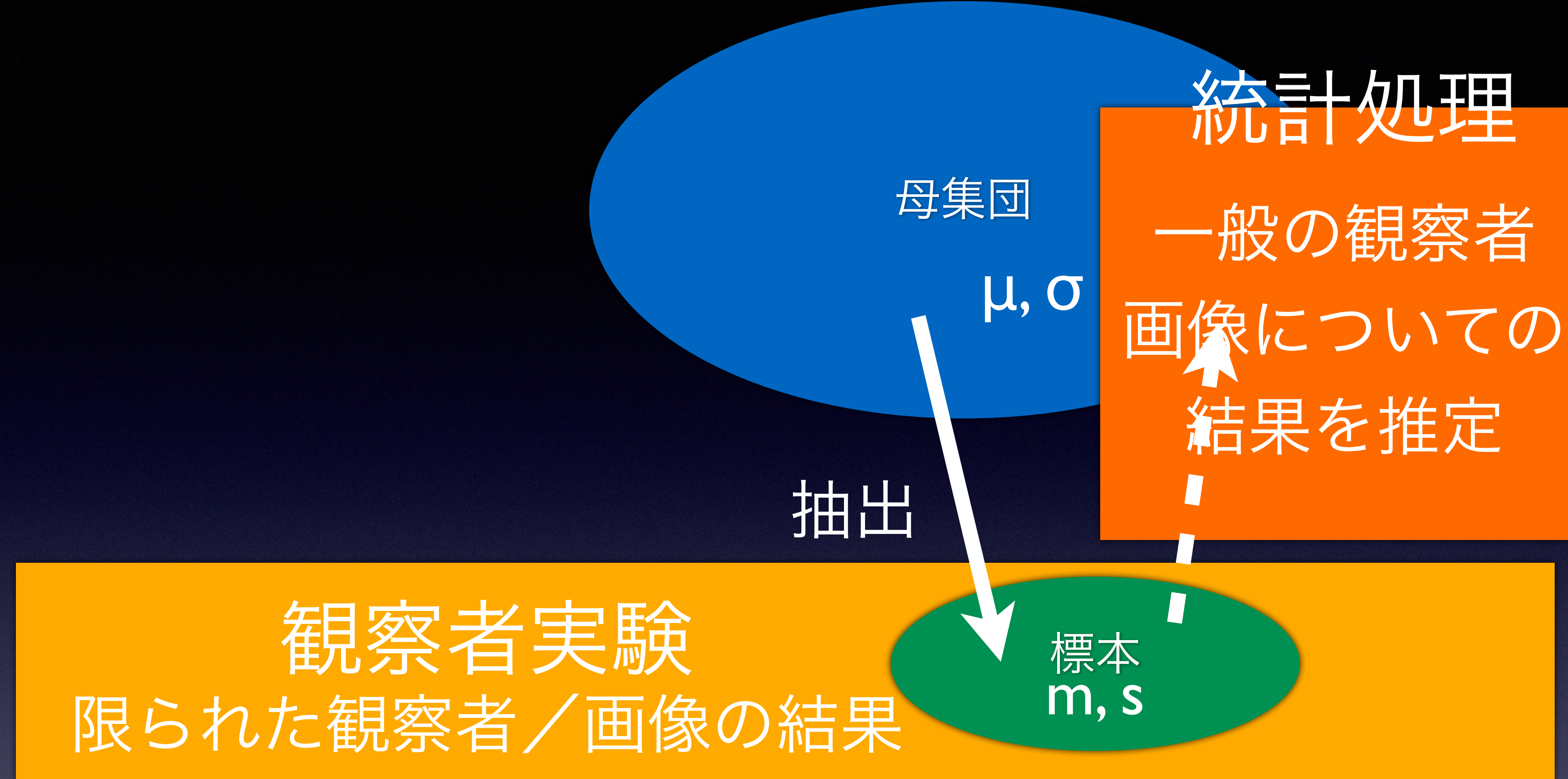
Reader	AUC SystemA	AUC SystemB	dAUC
1	0.83	0.9	-0.07
2	0.7	0.65	0.05
3	0.9	0.71	0.19
4	0.85	0.74	0.11
5	0.95	0.83	0.12
Average	0.846	0.766	0.08

対応あり



Reader	AUC SystemA	AUC SystemB	dAUC
1	0.83	0.9	-0.07
2	0.7	0.65	0.05
3	0.9	0.71	0.19
4	0.85	0.74	0.11
5	0.95	0.83	0.12
Average	0.846	0.766	0.08

対応あり

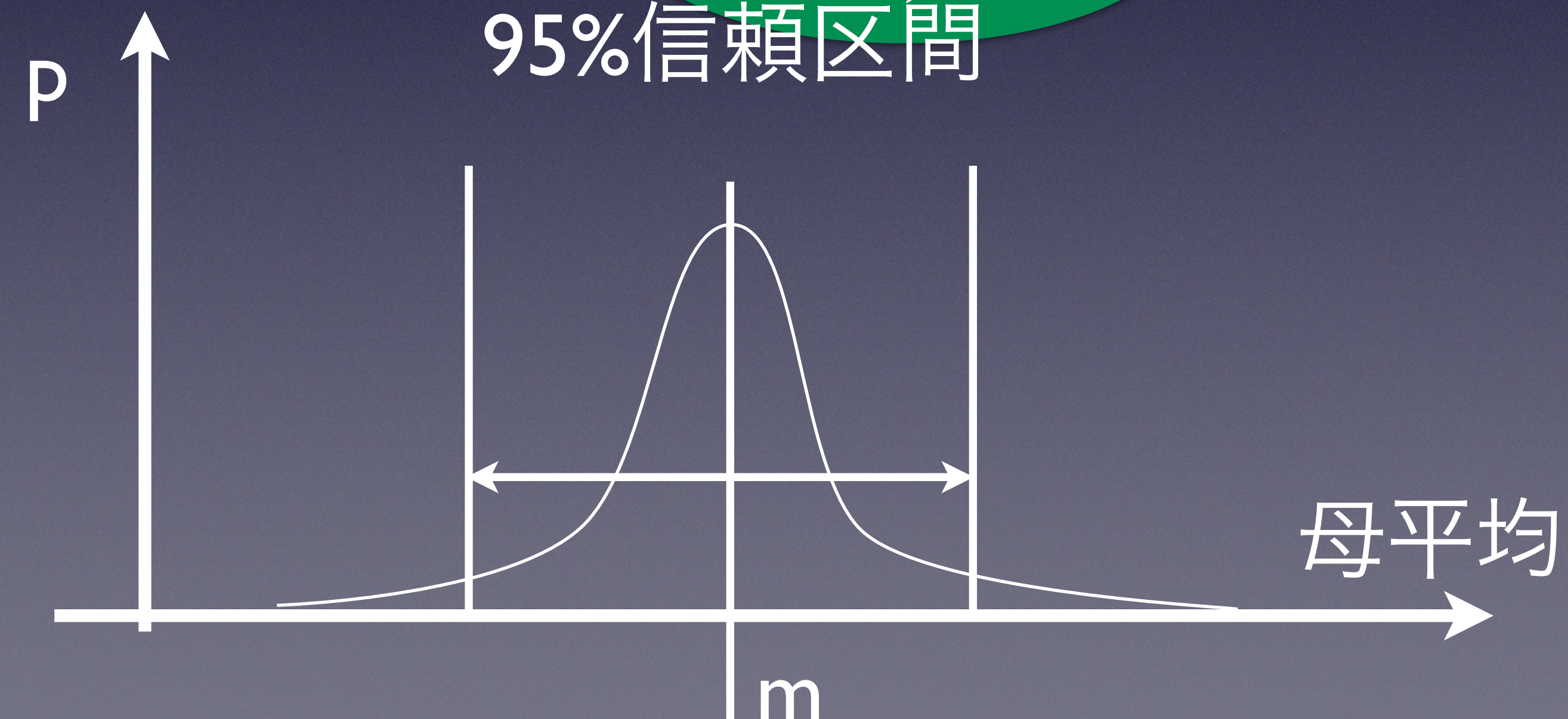


Reader	AUC SystemA	AUC SystemB	dAUC
1	0.83	0.9	-0.07
2	0.7	0.65	0.05
3	0.9	0.71	0.19
4	0.85	0.74	0.11
5	0.95	0.83	0.12
Average	0.846	0.766	0.08

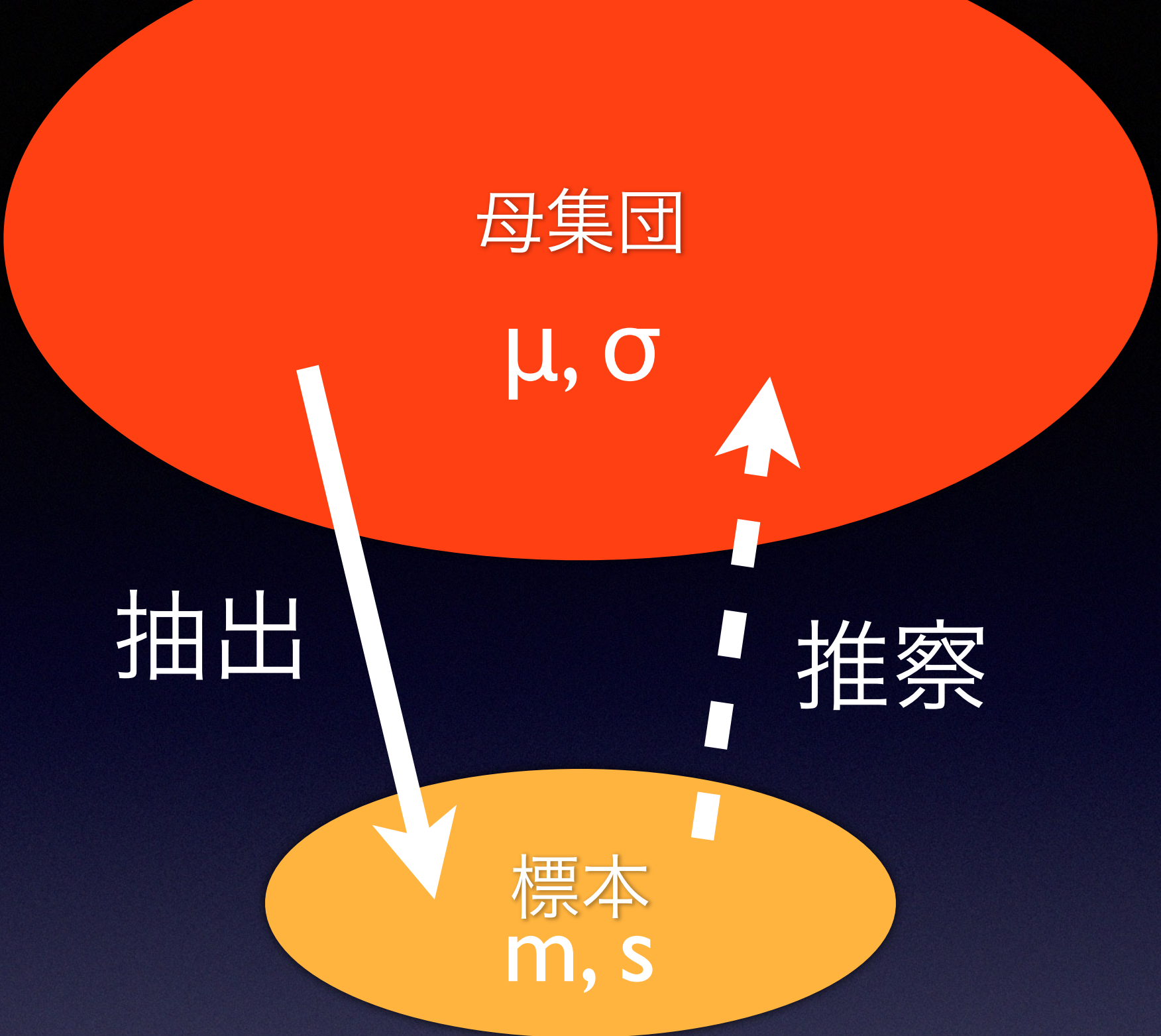
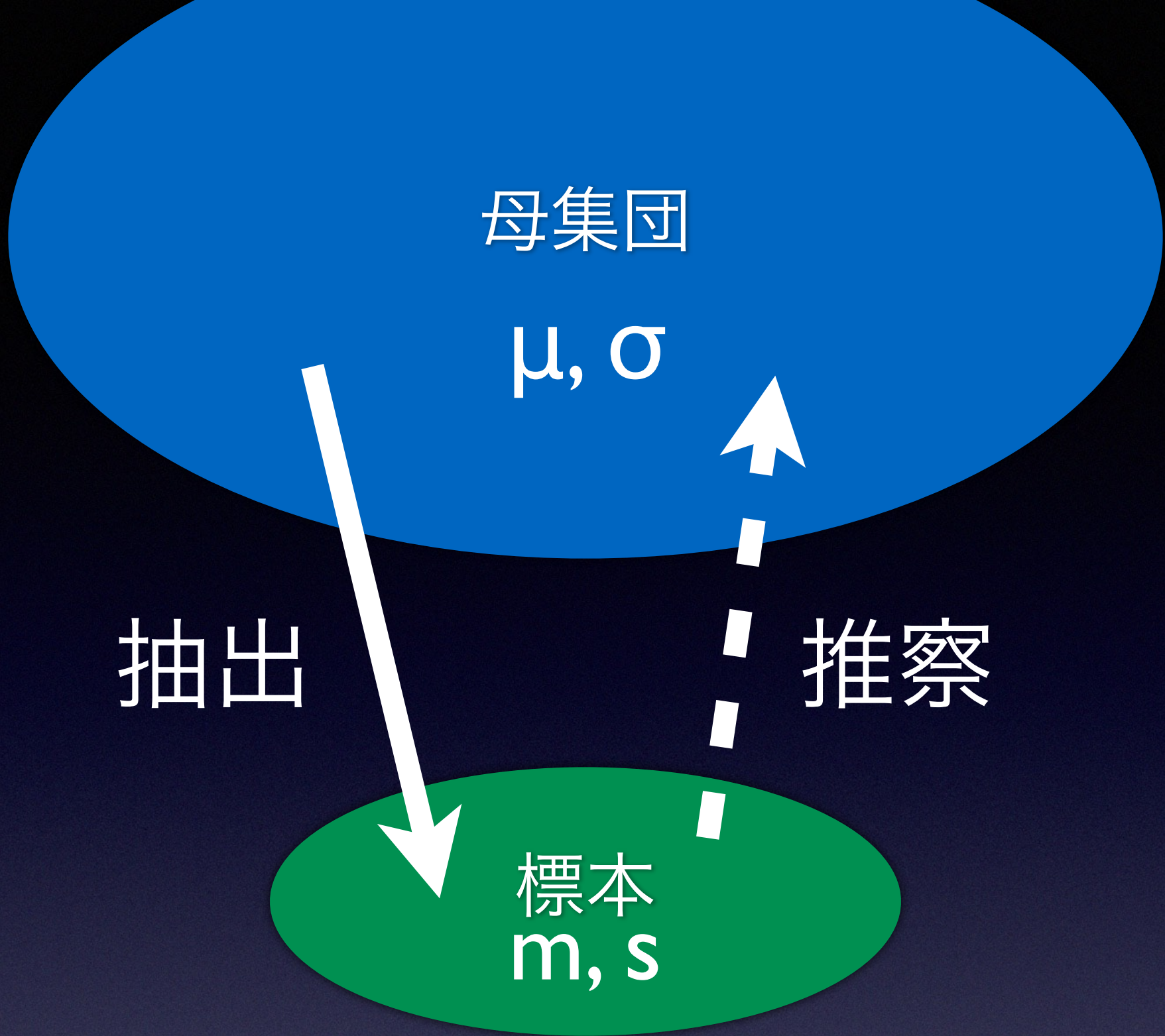
対応あり

95%信頼区間にゼロが含まれるか否か

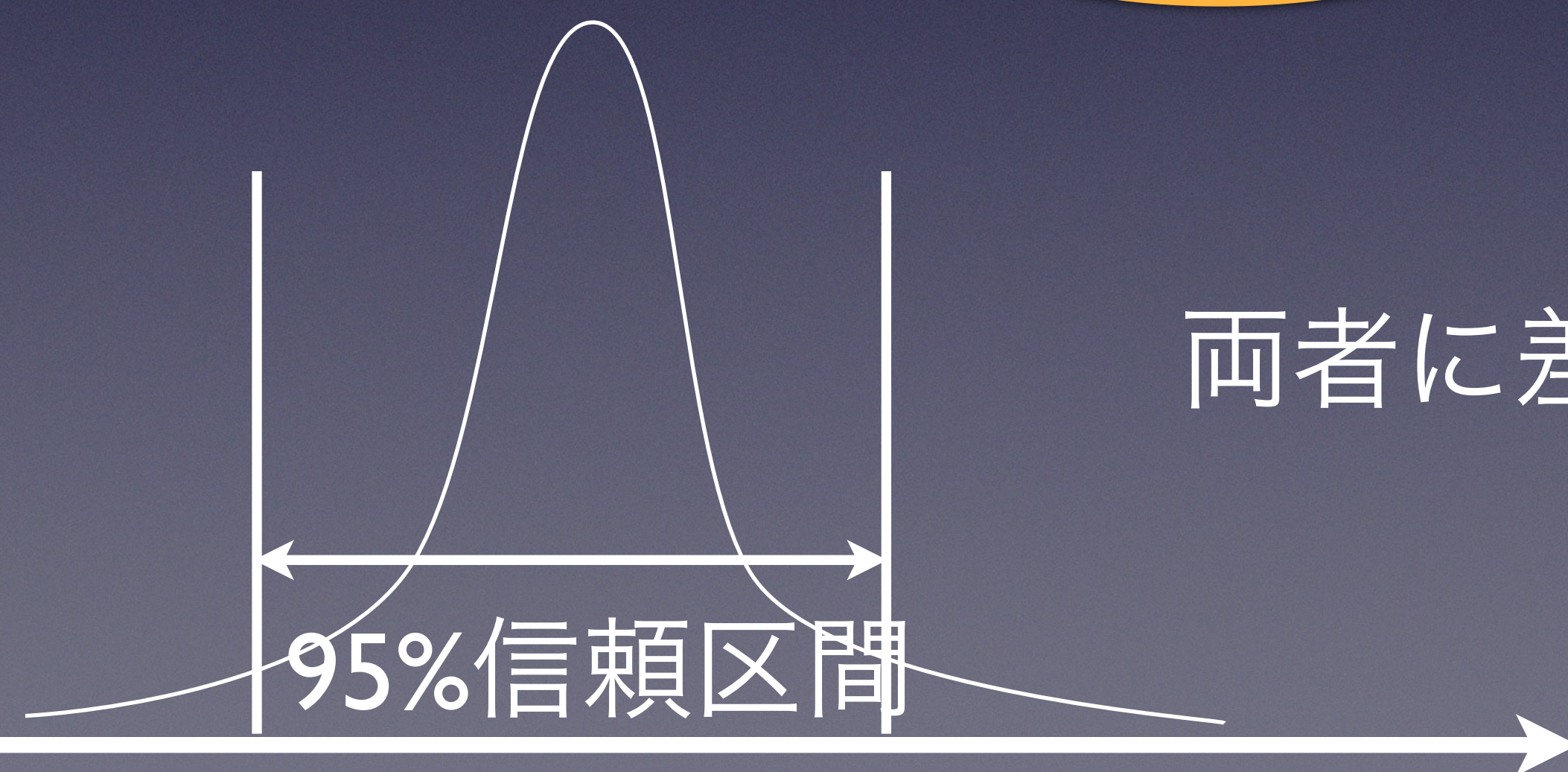
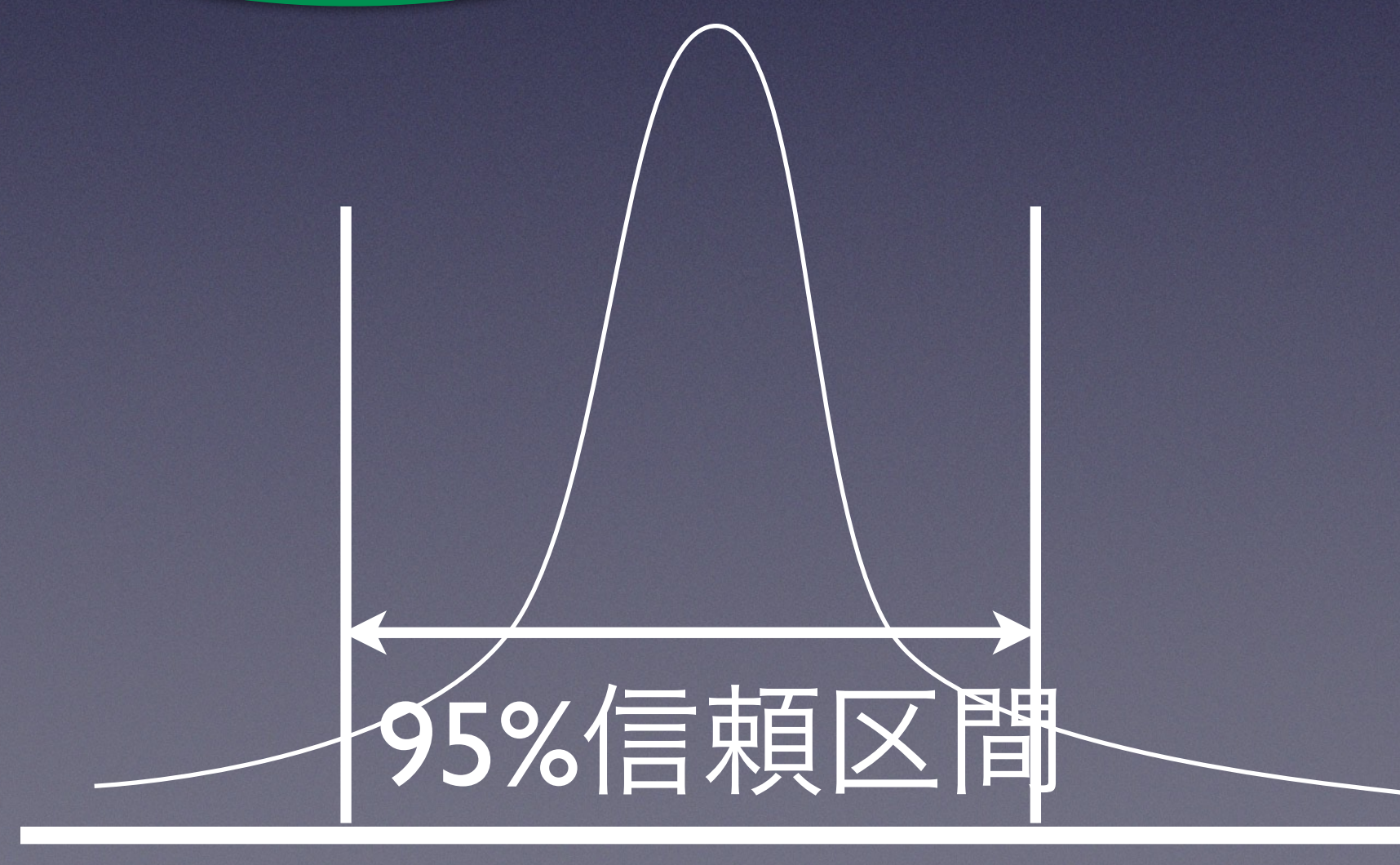
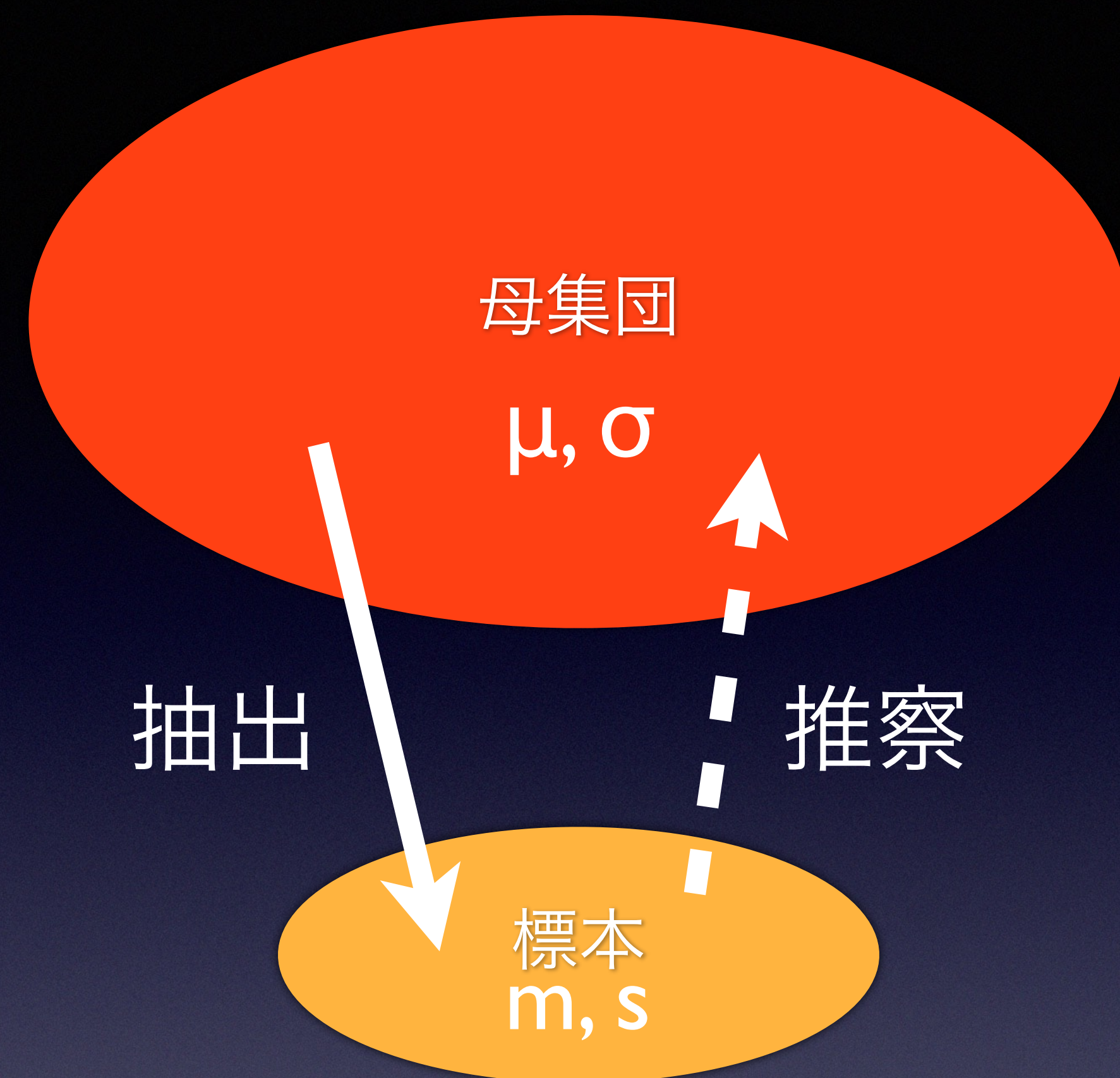
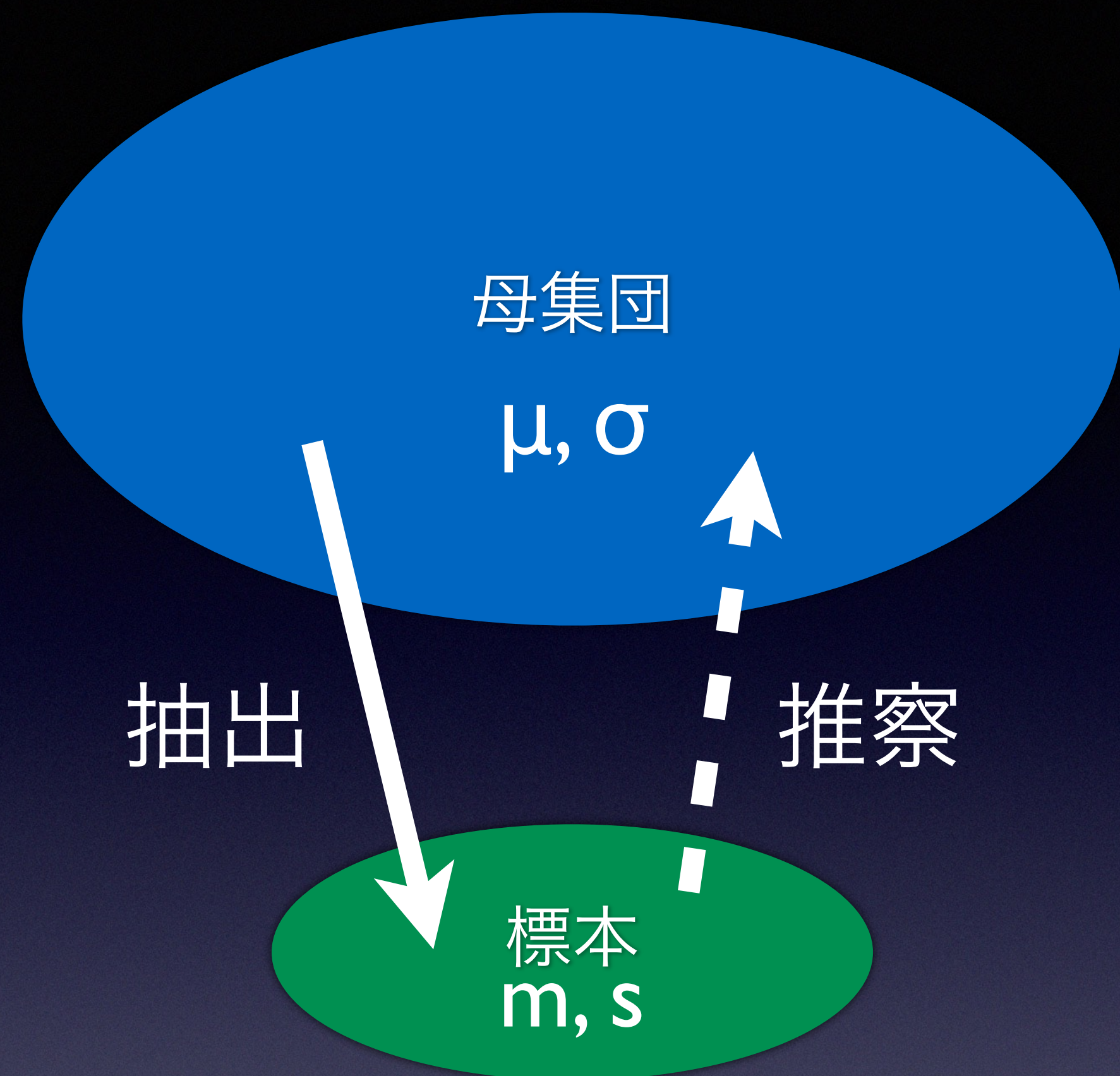
AUC SystemA	AUC SystemB	dAUC
0.83	0.9	-0.07
0.7	0.65	0.05
0.9	0.71	0.19
0.85	0.74	0.11
0.95	0.83	0.12
0.846	0.766	0.08



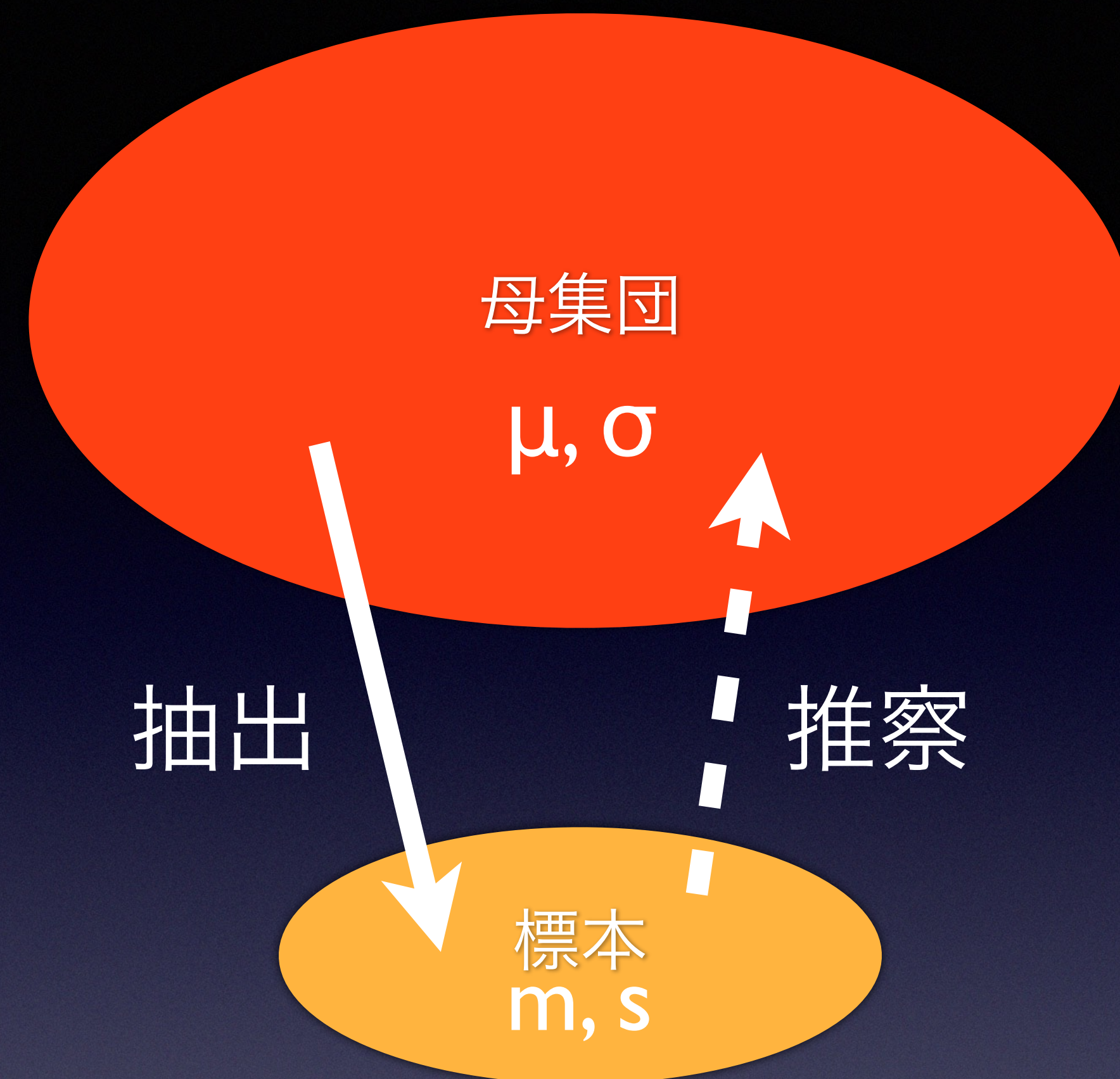
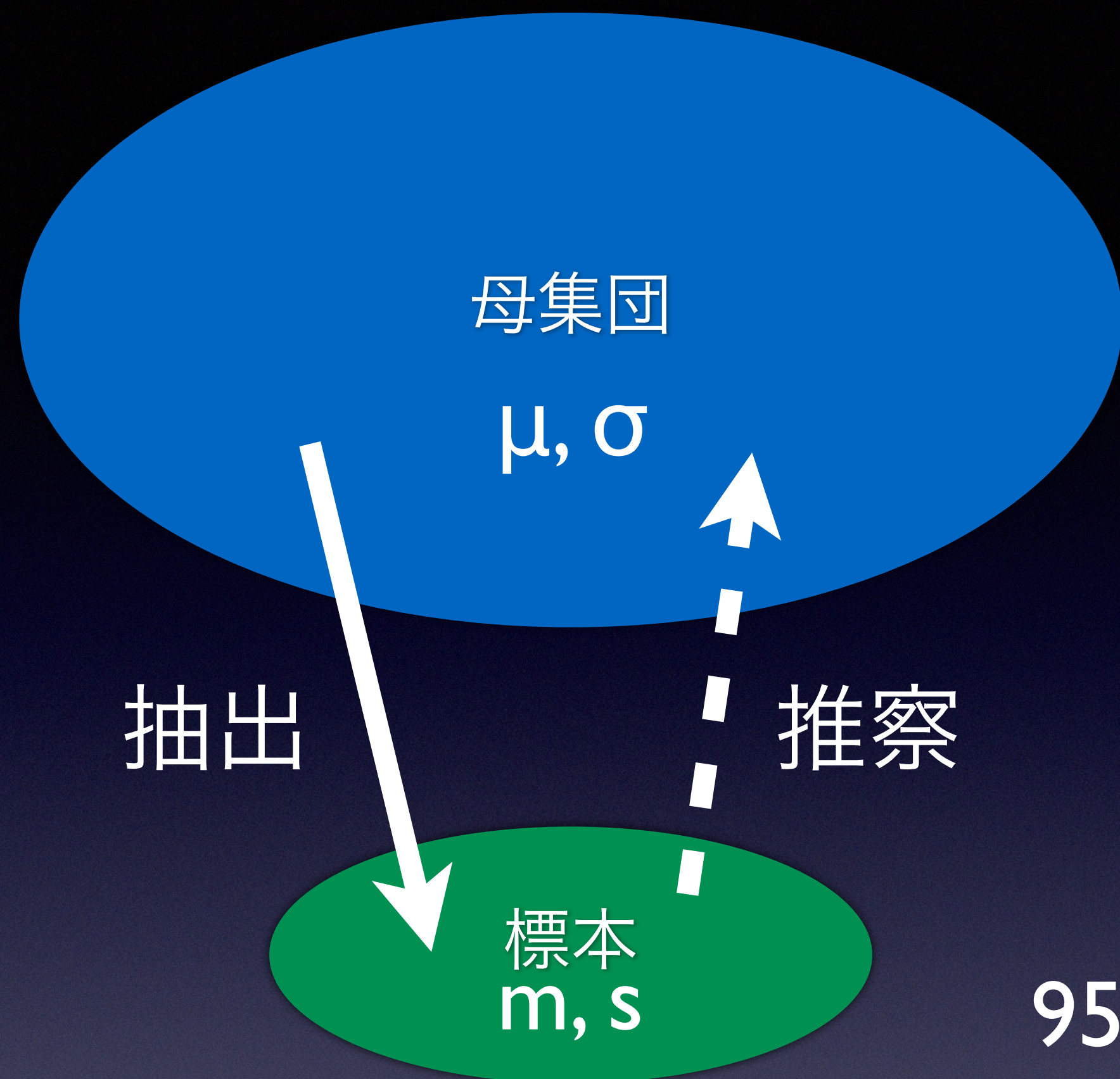
対応なし



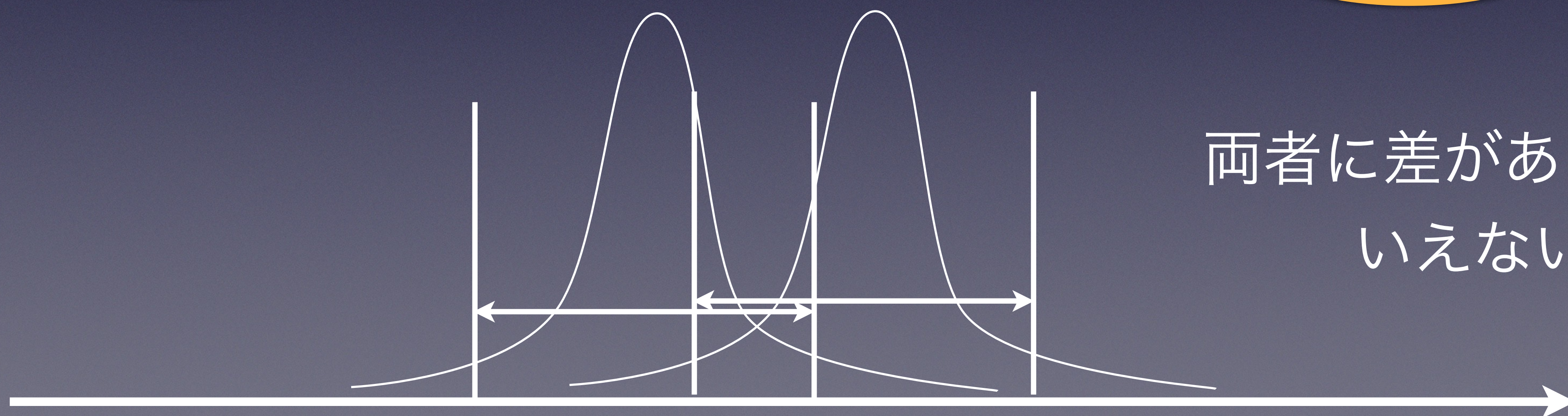
Reader	AUC SystemA	AUC SystemB
1	0.83	—
2	0.7	—
3	0.9	—
4	0.85	—
5	0.95	—
6	—	0.9
7	—	0.65
8	—	0.71
9	—	0.74
10	—	0.83
Average	0.846	0.766



両者に差があり



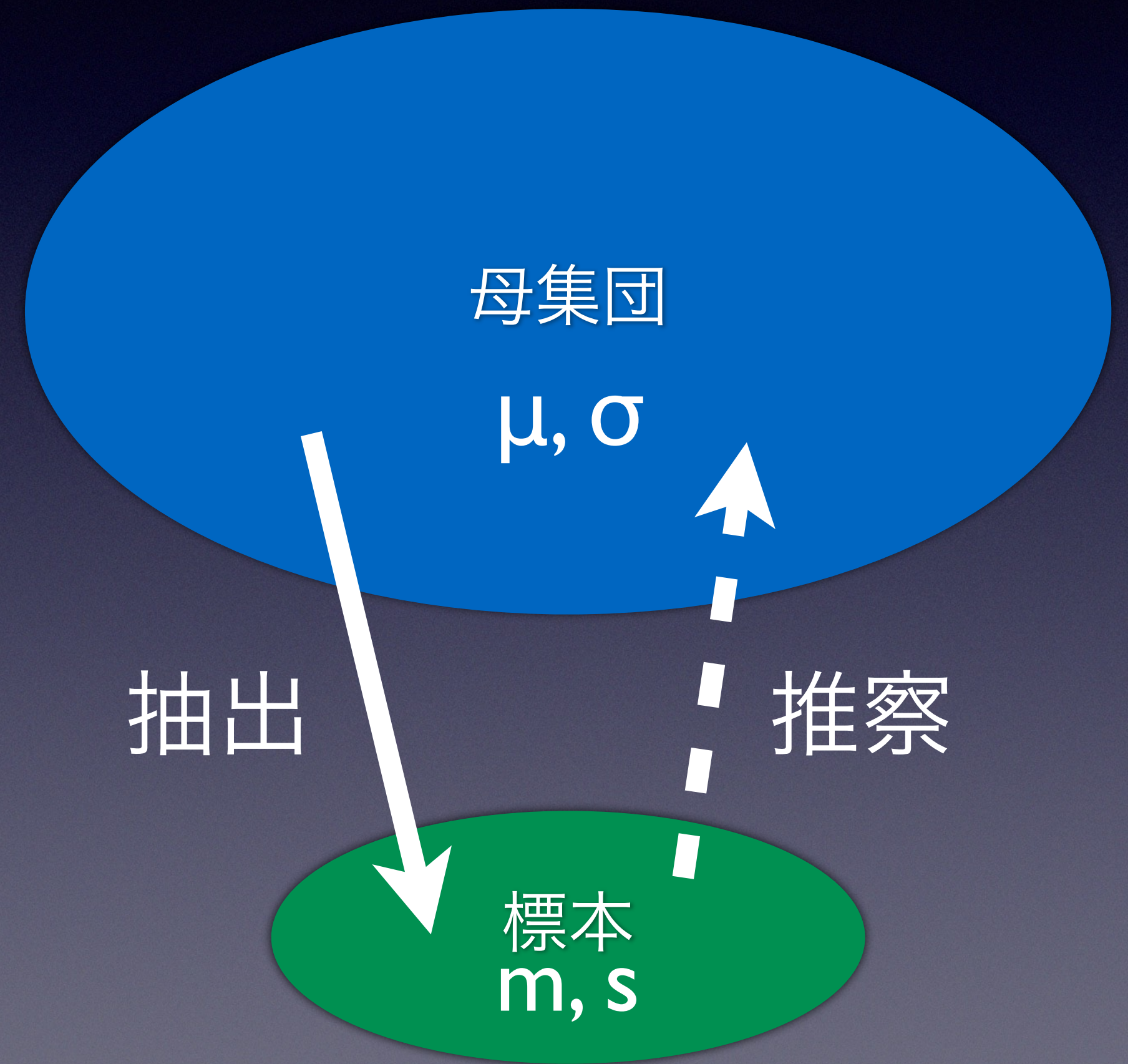
95%信頼区間



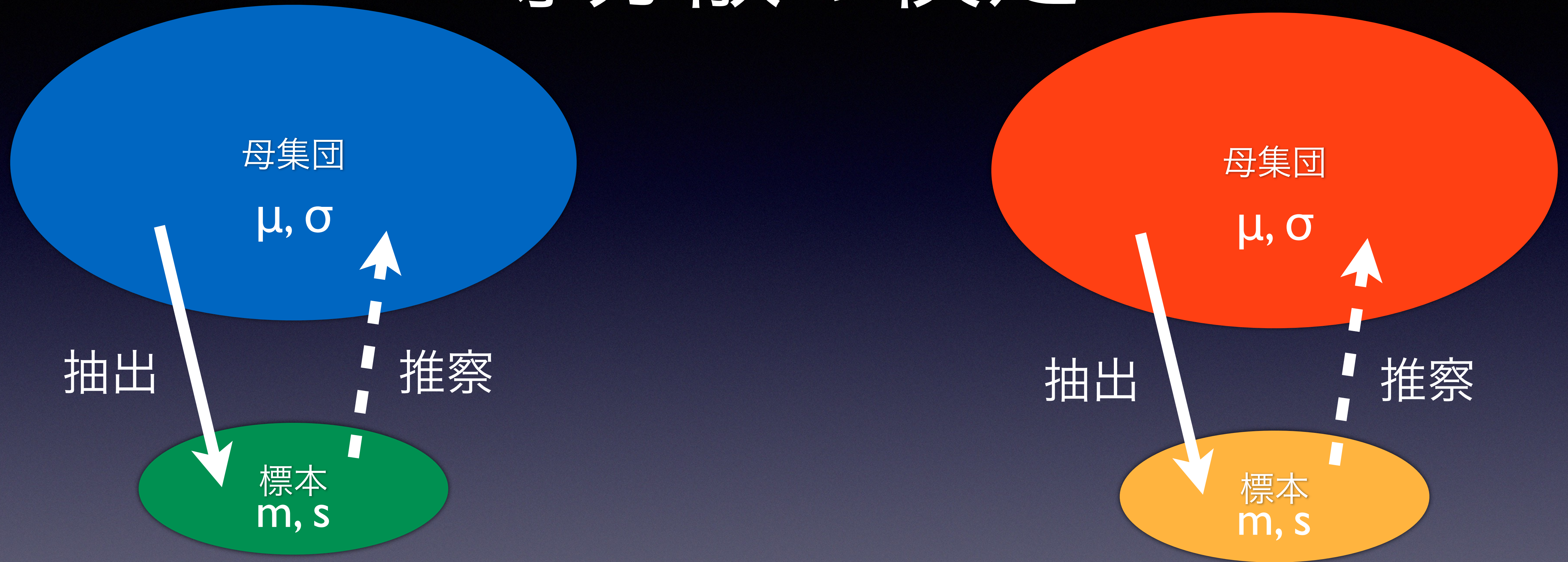
両者に差があるとはいえない

正規性の検定は？

- 標本の分布は正規分布になる場合とない場合があります。
- しかし、標本の平均の分布（ややこしい）は必ず正規分布になります（これが中心極限定理：証明済み）
- したがって、標本の分布の正規性の検定は不要と考えます



等分散の検定



両方の分散がちがうかどうかの検定
(母分散の区間推定を行う：F検定)

t検定：まとめ

- サンプル数が二桁あればt検定でよいと思う
- 2群で「対応あり」か「なし」かは最も重要
- 等分散を確認する：等分散か否かで方法が違う
- 正規性の検定は趣味：元データが自然データならば正規性高

質の検定？ 量の検定？

ノンパラメトリックな検定の理解を深める

DBM-MRMCに関する理解を深める

データとは？ 質 か 量？

		比較方法	例
質的 データ	名義尺度 (Nominal scale) 分類のための符号. 同一性のみを示す	最頻値	品番／型番
	順序尺度 (Ordered scale) 比較のための符号. 同一性と順序を示す	中央値	スコア (A, B, C...) , ステージ
量的 データ	間隔尺度 (Interval scale) 間隔の関係を説明する. 0が定まらない.	平均と	温度
	比率尺度 (Ratio Scale) 比の関係を説明する. 0が定まる.	標準偏差	体重, 圧力, 連続確信度

順序尺度か比率尺度か？

- 10段階評価を1 cm間隔でマーク 順序尺度 > 比率尺度
質的データ
- 100段階評価を1 mm間隔でマーク 比率尺度 > 順序尺度
量的データ

質の検定 量の検定

質の検定 ノンパラメトリック検定

データの分布に依存しない検定方法.

データを順位に置き換えて, 平均順位を比較する.

量の検定 パラメトリック検定

データの分布に依存する検定方法.

データの平均値を直接比較する.

分布

分散

対象とする変数

ノンパラメトリック
(質の検定)

分布に
依存しない

等分散で
なくとも
よい

順序変数 (離散), 対数
0 が定まらない数
(体温, pH)

パラメトリック
(量の検定)

正規分布

等分散

間隔尺度, 比率尺度
(連続変数)

代表的な検定

- 2群の違いを調べる
 - システム間の違い
 - モダリティ間の違い

2群の何を比較する？

代表的な検定

- 2群の違いを調べる
 - システム間の違い
 - モダリティ間の違い

2群の何を比較する？

平均

分散

代表値

順序

代表的な検定

- 2群の違いを調べる
システム間の違い
モダリティ間の違い

2群の何を比較する？

平均

分散

パラメトリック
(古典的?)

代表値

順序

代表的な検定

- 2群の違いを調べる
システム間の違い
モダリティ間の違い

2群の何を比較する？

平均

分散

パラメトリック
(古典的?)

代表値

順序

ノンパラメトリック

代表的な検定

- 2群の違いを調べる
システム間の違い
モダリティ間の違い

2群の何を比較する？

十分な数があるとコッチ

平均

分散

パラメトリック
(古典的?)

代表値

順序

ノンパラメトリック

代表的な検定

- 2群の違いを調べる
システム間の違い
モダリティ間の違い

2群の何を比較する？

十分な数があるとコッチ

平均

分散

パラメトリック
(古典的?)

数が少ないとコッチ

代表値

順序

ノンパラメトリック

2群は独立か同一か？

ちょっと
復習

10人の患者に睡眠薬を飲ませた。その結果、睡眠時間がそれぞれ変化した。

TABLE I.

Patient	Controls (no hypnotic)		0.6 mg. L-Hyoscyamine HBr.			0.6 mg. L-Hyoscyne HBr.			0.6 mg. R-Hyoscyne HBr.		
	No. of obser- vations	Average hours of sleep	No. of obser- vations	Average hours of sleep	Increase over controls	No. of obser- vations	Average hours of sleep	Increase over controls	No. of obser- vations	Average hours of sleep	Increase over controls
1	9	0.6	6	1.3	0.7	6	2.5	1.9	6	2.1	1.5
2	9	3.0	6	1.4	-1.6	6	3.8	0.8	6	4.4	1.4
3	8	4.7	6	4.5	-0.2	6	5.8	1.1	6	4.7	0.0
4	9	5.5	3	4.3	-1.2	3	5.6	0.1	3	4.8	-0.7
5	9	6.2	3	6.1	-0.1	3	6.1	-0.1	3	6.7	0.5
6	8	3.2	4	6.6	3.4	3	7.6	4.4	3	8.3	5.1
7	8	2.5	3	6.2	3.7	3	8.0	5.5	3	8.2	5.7
8	7	2.8	6	3.6	0.8	6	4.4	1.6	5	4.3	1.5
9	8	1.1	5	1.1	0.0	6	5.7	4.6	5	5.8	4.7
10	9	2.9	5	4.9	2.0	5	6.3	3.4	6	6.4	3.5
11	—	—	2	6.3	—	2	6.8	—	2	7.3	—

2群は独立か同一か？

ちょっと
復習

10人の患者に睡眠薬を飲ませた。その結果、睡眠時間がそれぞれ変化した。

同一の群で比較
(対応あり : paired)

TABLE I.

Patient	Controls (no hypnotic)		0.6 mg. L-Hyoscyamine HBr.			0.6 mg. L-Hyoscyne HBr.			0.6 mg. R-Hyoscyne HBr.		
	No. of obser- vations	Average hours of sleep	No. of obser- vations	Average hours of sleep	Increase over controls	No. of obser- vations	Average hours of sleep	Increase over controls	No. of obser- vations	Average hours of sleep	Increase over controls
1	9	0.6	6	1.3	0.7	6	2.5	1.9	6	2.1	1.5
2	9	3.0	6	1.4	-1.6	6	3.8	0.8	6	4.4	1.4
3	8	4.7	6	4.5	-0.2	6	5.8	1.1	6	4.7	0.0
4	9	5.5	3	4.3	-1.2	3	5.6	0.1	3	4.8	-0.7
5	9	6.2	3	6.1	-0.1	3	6.1	-0.1	3	6.7	0.5
6	8	3.2	4	6.6	3.4	3	7.6	4.4	3	8.3	5.1
7	8	2.5	3	6.2	3.7	3	8.0	5.5	3	8.2	5.7
8	7	2.8	6	3.6	0.8	6	4.4	1.6	5	4.3	1.5
9	8	1.1	5	1.1	0.0	6	5.7	4.6	5	5.8	4.7
10	9	2.9	5	4.9	2.0	5	6.3	3.4	6	6.4	3.5
11	—	—	2	6.3	—	2	6.8	—	2	7.3	—

2群は独立か同一か？

2組の患者に痛みの
レベルを報告しても
らった。

Data on shoulder tip pain after

I.D. No.	Active treatment	Sex	Age (years)
101	Y	F	64
102	Y	M	41
103	Y	F	77
104	Y	F	54
105	Y	F	66
106	Y	M	56
107	Y	M	81
108	Y	F	24
109	Y	F	56
110	Y	F	29
111	Y	M	65
112	Y	F	68
113	Y	M	77
114	Y	M	35
115	Y	M	66

Table 2

Data on shoulder tip pain after laparoscopic cholecystectomy (Jor)

I.D. No.	Active treatment	Sex	Age (years)	Pain scores (1 = low)		
				Time 1	Time 2	Time 3
101	Y	F	64	1	1	1
102	Y	M	41	3	2	1
103	Y	F	77	3	2	2
104	Y	F	54	1	1	1
105	Y	F	66	1	1	1
106	Y	M	56	1	2	1
107	Y	M	81	1	3	2
108	Y	F	24	2	2	1
109	Y	F	56	1	1	1
110	Y	F	29	3	1	1
111	Y	M	65	1	1	1
112	Y	F	68	2	1	1
113	Y	M	77	1	2	2
114	Y	M	35	3	1	1
115	Y	M	66	2	1	1
116	Y	F	70	1	1	1
117	Y	M	79	1	1	1
118	Y	F	65	2	1	1
119	Y	F	61	4	4	2
120	Y	F	67	4	4	4
121	Y	F	32	1	1	1
122	Y	F	33	1	1	1
201	N	F	20	5	2	3
202	N	F	50	1	5	3
203	N	F	40	4	4	4
204	N	M	54	4	4	4
205	N	M	34	2	3	4
206	N	F	34	3	4	3
207	N	M	56	3	3	4
208	N	F	82	1	1	1
209	N	M	56	1	1	1
210	N	M	52	1	5	5
211	N	F	65	1	3	2
212	N	F	53	2	2	3
213	N	F	40	2	2	1
214	N	F	58	1	1	1
215	N	M	63	1	1	1
216	N	F	41	5	5	5
217	N	M	72	3	3	3
218	N	F	60	5	4	4
219	N	M	61	1	3	3

2群は独立か同一か？

2組の患者に痛みの
レベルを報告しても
らった。

独立した群
(対応なし)

Data on shoulder tip pain after

I.D. No.	Active treatment	Sex	Age (years)
101	Y	F	64
102	Y	M	41
103	Y	F	77
104	Y	F	54
105	Y	F	66
106	Y	M	56
107	Y	M	81
108	Y	F	24
109	Y	F	56
110	Y	F	29
111	Y	M	65
112	Y	F	68
113	Y	M	77
114	Y	M	35
115	Y	M	66

Table 2

Data on shoulder tip pain after laparoscopic cholecystectomy (Jor)

I.D. No.	Active treatment	Sex	Age (years)	Pain scores (1 = low)		
				Time 1	Time 2	Time 3
101	Y	F	64	1	1	1
102	Y	M	41	3	2	1
103	Y	F	77	3	2	2
104	Y	F	54	1	1	1
105	Y	F	66	1	1	1
106	Y	M	56	1	2	1
107	Y	M	81	1	3	2
108	Y	F	24	2	2	1
109	Y	F	56	1	1	1
110	Y	F	29	3	1	1
111	Y	M	65	1	1	1
112	Y	F	68	2	1	1
113	Y	M	77	1	2	2
114	Y	M	35	3	1	1
115	Y	M	66	2	1	1
116	Y	F	70	1	1	1
117	Y	M	79	1	1	1
118	Y	F	65	2	1	1
119	Y	F	61	4	4	2
120	Y	F	67	4	4	4
121	Y	F	32	1	1	1
122	Y	F	33	1	1	1
201	N	F	20	5	2	3
202	N	F	50	1	5	3
203	N	F	40	4	4	4
204	N	M	54	4	4	4
205	N	M	34	2	3	4
206	N	F	34	3	4	3
207	N	M	56	3	3	4
208	N	F	82	1	1	1
209	N	M	56	1	1	1
210	N	M	52	1	5	5
211	N	F	65	1	3	2
212	N	F	53	2	2	3
213	N	F	40	2	2	1
214	N	F	58	1	1	1
215	N	M	63	1	1	1
216	N	F	41	5	5	5
217	N	M	72	3	3	3
218	N	F	60	5	4	4
219	N	M	61	1	3	3

観察者実験では？

観察者実験では？

確信度は、質的データか？ 量的データか？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、質的データか？ 量的データか？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、質的データか？ 量的データか？

質的データを量として考えられるか？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、質的データか？ 量的データか？

質的データを量として考えられるか？

それはいつ？ どうやって？

対応あり

Reader	AUC SystemA	AUC SystemB	dAUC
1	0.83	0.9	-0.07
2	0.7	0.65	0.05
3	0.9	0.71	0.19
4	0.85	0.74	0.11
5	0.95	0.83	0.12
Average	0.846	0.766	0.08

対応なし

Reader	AUC SystemA	AUC SystemB
1	0.83	—
2	0.7	—
3	0.9	—
4	0.85	—
5	0.95	—
6	—	0.9
7	—	0.65
8	—	0.71
9	—	0.74
10	—	0.83
Average	0.846	0.766

実際のデータで
やってみましょう

原の方針

- 20人位いるので、 t 検定
- 同じ観察者ではないので、対応なしの t 検定
- 等分散の検定もしなきゃ
- 念のため、Wilcoxonもやっとくかな. . .

2群の比較

パラメトリック
平均, 分散
(量の検定)

ノンパラメトリック
代表値, 順序
(質の検定)

同一の群 対応のあるt検定, F検定

Wilcoxon符号付き順位検定

独立した群 t検定, F検定

Wilcoxon-Mann-Whitney検定
(U検定, Wilcoxon順位和
検定も同じ意味)

ノンパラメトリックな検定 なぜ必要なのか？

(あくまで原の私見です)

サンプル数が少ない

- 十分な数がないと，正規分布に従うかどうか分からない
- t 分布でもよいが，数が少ないと信頼区間が広がる
- 逆に，数が多ければ（30以上くらい？，100くらい）どれでやっても一緒（順位を数値のように読むことも可能）

符号検定

(符号の数を検定する)

同一の 2 群の代表値の比較

符号検定 (sign test)

ノンパラメトリックな検定の基本

- 数値の変化を「差」で表す
- 差には正負があるはず
- 正負の符号の出方の確率を計算して、検定とする考え方

薬を飲んだら睡眠時間が増えた

- 10人の患者にある睡眠薬を飲ませた.
- 睡眠時間がそれぞれ変化しました（9人増えた）

1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4

10人の変化：正負

- 10人の正負の変化の「符号の付き方」は全部で $2^{10} = 1024$ 通り
- すべて正になる： ${}_{10}C_0 = 1$ 通り
- 一つだけ正になる： ${}_{10}C_1 = 10$ 通り／一つだけ負：10通り

そうすると、この現象は

1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4

1024通りのうちの10通りのできごとが起きた
すなわち、10/1024のできごとが起きている

ただし、符号の正負は勝手に計算の順序で決まっただけ

計算すべき確率

1.9, 0.8, 1.1, 0.1, **-0.1**, 4.4, 5.5, 1.6, 4.6, 3.4

一つだけ符号が違う, もしくはすべての符号が同じ
になる確率を計算する必要がある

一つだけ違う:	正	10通り	負	10通り
すべて同じ:	正	1通り	負	1通り

すなわち, $p = 22/1024 = 0.021$

分布を推定していない

- t検定やカイ二乗検定のように、母平均などの分布を仮定していない。
- 純粹に確率的な組み合わせを計算して、その場合の数を算出して確率を計算する。

観察者実験では？



観察者実験では？

確信度は、**質的データ**か？ 量的データか？



観察者実験では？

確信度は、**質的データ**か？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？



観察者実験では？

確信度は、**質的データ**か？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？



観察者実験では？

確信度は、**質的データ**か？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、**質的データ**か？ 量的データか？

観察者実験では？

確信度は、**質的データ**か？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、**質的データ**か？ 量的データか？

質的データを量として考えられるか？

観察者実験では？

確信度は、**質的データ**か？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、**質的データ**か？ 量的データか？

質的データを量として考えられるか？

それはいつ？ どうやって？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、質的データか？ 量的データか？

質的データを量として考えられるか？

それはいつ？ どうやって？

観察者実験では？

確信度は、質的データか？ 量的データか？

5段階評定の結果は質的データ？ 量的データ？

連続確信度法の結果は？

AUCは、質的データか？ 量的データか？

質的データを量として考えられるか？

それはいつ？ どうやって？

2群の比較

パラメトリック
平均, 分散
(量の検定)

ノンパラメトリック
代表値, 順序
(質の検定)

同一の群	対応のあるt検定, F検定	Wilcoxon符号付き順位検定
------	---------------	------------------

独立した群	t検定, F検定	Wilcoxon-Mann-Whitney検定 (U検定, Wilcoxon順位和 検定も同じ意味)
-------	----------	--

Wilcoxonの符号付き順位検定

(Wilcoxonの符号付き順位和検定とも呼ばれます)

同一の2群の代表値の比較

Wilcoxonの符号付き順位検定

基本的な考え方

- 値の大きさの順序にのみ意味があるのならば、検定方法が違
う（質の検定）
- 順序（＝順位）を変化させる要因を考える

薬を飲んだら睡眠時間が増えた

- 10人の患者にある睡眠薬を飲ませた.
- 睡眠時間がそれぞれ変化しました（9人増えた）

1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4

Wilcoxonの符号付き順位検定

実測値： 1.9, 0.8, 1.1, 0.1, **-0.1**, 4.4, 5.5, 1.6, 4.6, 3.4

ソート： **-0.1**, 0.1, 0.8, 1.1, 1.6, 1.9, 3.4, 4.4, 4.6, 5.5

絶対値： **0.1**, 0.1, 0.8, 1.1, 1.6, 1.9, 3.4, 4.4, 4.6, 5.5

順位： **1.5**, 1.5, 3, 4, 5, 6, 7, 8, 9, 10

符号付順位： **-1.5**, 1.5, 3, 4, 5, 6, 7, 8, 9, 10

符号付き順位を考える

符号付順位 : **-1.5**, 1.5, 3, 4, 5, 6, 7, 8, 9, 10

順位和 : $-1.5 + 1.5 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 52$

符号付き順位の要素それぞれに、符号がつく可能性がある。

それが、順位和をどのくらい変動させるか、すべての組み合わせを考える。

その中で、順位和の絶対値が52を超える場合を数え確率を求める。

Wilcoxonの符号付き順位検定はWilcoxonの符号付き順位和検定とも呼ばれます。

順位和を作る組み合わせ

順位和： $-1.5+1.5+3+4+5+6+7+8+9+10=52$

52以上になる場合：

全部正, 1.5の一つがマイナス：3通り

-52以下になる場合：

全部負, 1.5の一つがプラス：3通り

順位和の絶対値が、偶然 52 を 超える確率を求める

- 10個の数値に符号の付け方は1024通り（2の10乗）
- そのうち、6通りのできごとが起こった
- したがって、この確率： p は $p = 6/1024 = 0.0059$

Wilcoxonの符号付き順位検定は
Wilcoxonの符号付き順位和検定とも呼ばれます。

Wilcoxon-Mann-Whitney検定

(U検定, Wilcoxonの順位和検定とも呼ばれます)

独立した2群の代表値の比較

演習

- 大量の魚から100匹の標本を抽出して測定したところ、標本平均：150g、標本標準偏差：30 gであった。母平均の95%信頼区間を求めなさい。



演習

- 別の磯へ釣りに出かけました.
- 連れた大量の魚から100匹の標本を抽出して測定したところ、標本平均：155g、標本標準偏差：30 gであった。母平均の95%信頼区間を求めなさい.



これだけ釣れれば
パラメトリック検定

演習

- ある磯では4匹釣れました
- 別の磯では5匹釣れました



釣れなければ
ノンパラメトリック検定

独立した2群の値

群1 : 1.2, 1.5, 1.8, 2.6

群2 : 1.3, 1.9, 2.9, 3.1, 3.9

これを全部並べて順位をつける

1.2, 1.3, 1.5, 1.8, 1.9, 2.6, 2.9, 3.1, 3.9

1, 2, 3, 4, 5, 6, 7, 8, 9

それぞれの群の順位和を求める

群1 : $R1 = 1+3+4+6 = 14$

群2 : $R2 = 2+5+7+8+9=31$

統計検定量：Uを求める

$$\text{群 1} : R1 = 1+3+4+6 = 14$$

$$\text{群 2} : R2 = 2+5+7+8+9=31$$

$$\text{群 1 の数} : n1=4, \text{ 群 2 の数} : n2=5$$

群 1 のU：

$$U1 = n1 \times n2 + n1(n1 + 1)/2 - R1 = 4 \times 5 + 4(4+1)/2 - 14 = 20 + 20/2 - 14 = 16$$

群 2 のU：

$$U2 = n1 \times n2 + n2(n2 + 1)/2 - R2 = 4 \times 5 + 5(5+1)/2 - 31 = 20 + 30/2 - 31 = 4$$

統計検定量：U0を定める

U1とU2の値の小さい方

したがって， $U0=4$

U検定の表を見る

$n1=4, N2=5, U0=4$ を見る

		両側検定		片側検定	
<u>n1</u>	<u>n2</u>	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
	<u>49</u>	<u>23</u>	<u>12</u>	<u>30</u>	<u>16</u>
	<u>50</u>	<u>24</u>	<u>12</u>	<u>31</u>	<u>16</u>
<u>4</u>	<u>4</u>	<u>0</u>	<u>---</u>	<u>1</u>	<u>---</u>
	<u>5</u>	<u>1</u>	<u>---</u>	<u>2</u>	<u>0</u>
	<u>6</u>	<u>2</u>	<u>0</u>	<u>3</u>	<u>1</u>
	<u>7</u>	<u>3</u>	<u>0</u>	<u>4</u>	<u>1</u>
	<u>8</u>	<u>4</u>	<u>1</u>	<u>5</u>	<u>2</u>

両側検定の棄却限界値は 1 .
これとU0を比較する.

比較した結果

- U0>棄却限界値ならば差があるとは言えない
- U0≤棄却限界値ならば差があり

U検定の表を見る

$n1=4, N2=5, U0=4$ を見る

n1	n2	両側検定		片側検定	
		$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
	49	23	12	30	16
	50	24	12	31	16
4	4	0	---	1	---
	5	1	---	2	0
	6	2	0	3	1
	7	3	0	4	1
	8	4	1	5	2

両側検定の棄却限界値は 1 .
これと $U0$ を比較する.

比較した結果

$U0 > \text{棄却限界値}$ ならば差があるとは言えない

$U0 \leq \text{棄却限界値}$ ならば差があり

U検定の表を見る

$n1=4, N2=5, U0=4$ を見る

		両側検定		片側検定	
<u>n1</u>	<u>n2</u>	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
	49	23	12	30	16
	50	24	12	31	16
4	4	0	---	1	---
	5	1	---	2	0
	6	2	0	3	1
	7	3	0	4	1
	8	4	1	5	2

両側検定の棄却限界値は 1 .
これとU0を比較する.

この例では,
 $U0=4>1$ なので
「差があるとは言えない」

比較した結果

- $U0>$ 棄却限界値ならば差があるとは言えない
- $U0\leq$ 棄却限界値ならば差があり

ROC曲線下面積とその検定

面積は量のデータ. しかし, 観察者の数は少ない!

方針：AUCの変化を調べる

(もう一度)

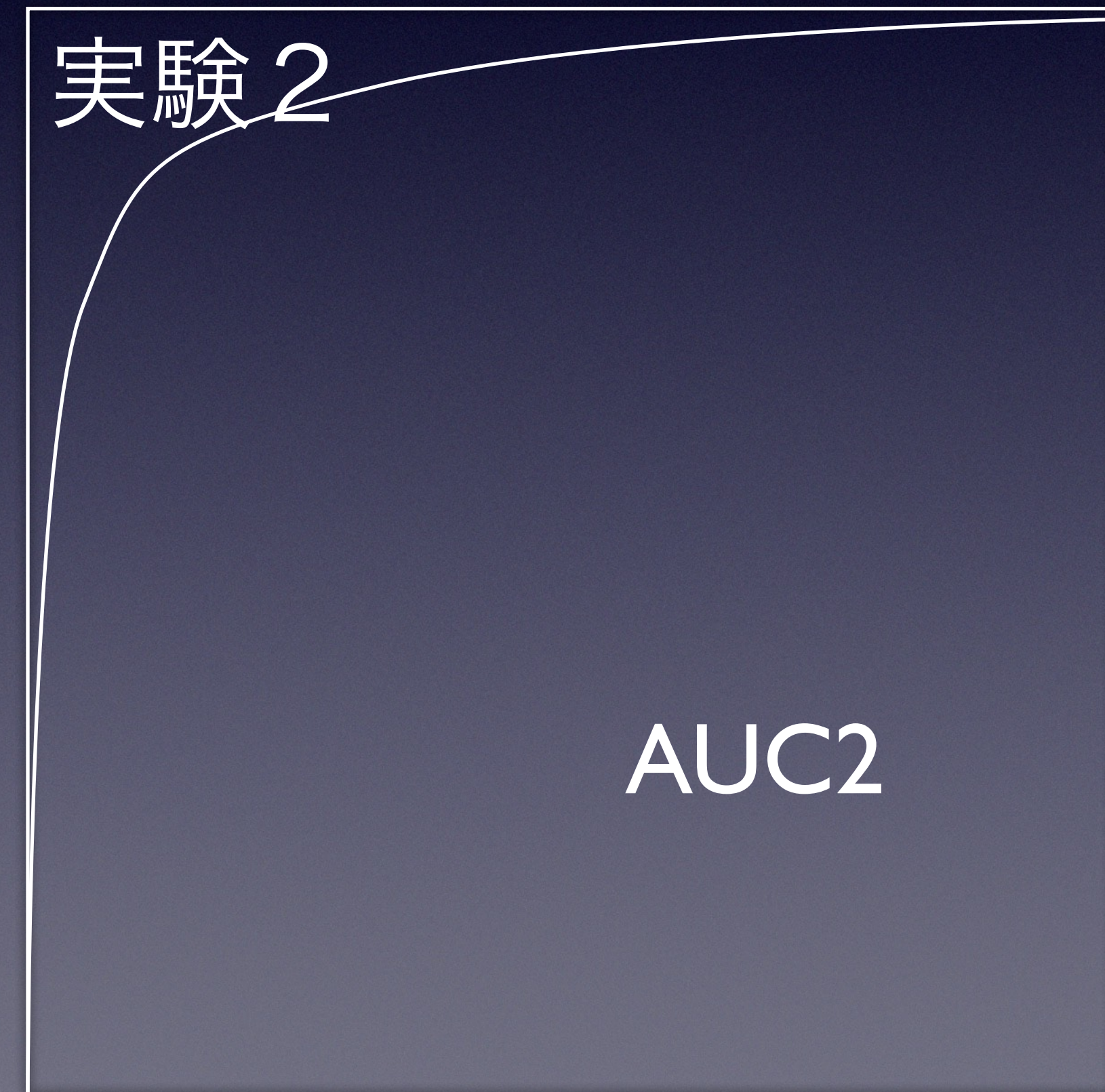
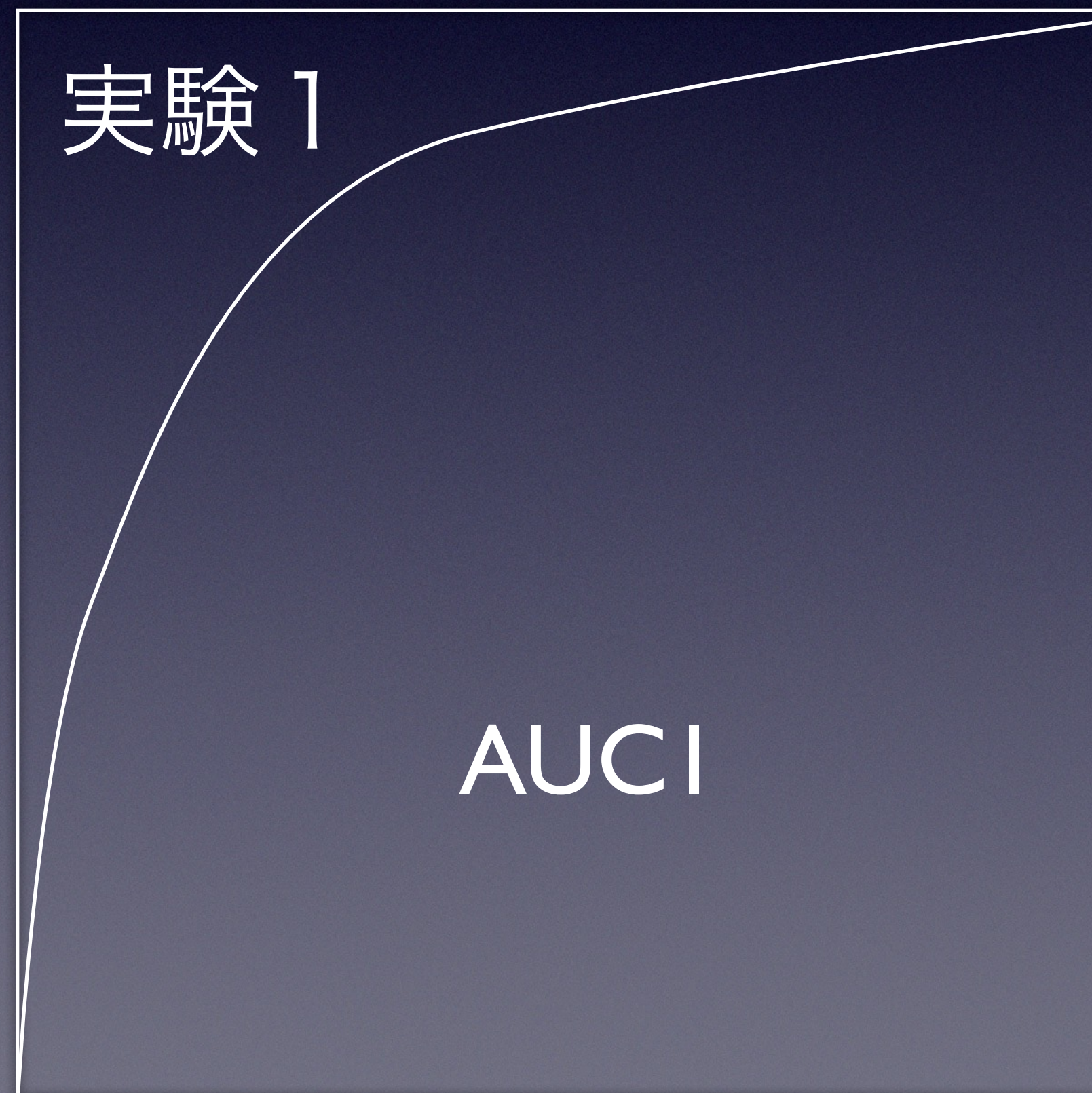
- システム1とシステム2の結果について調べる：2群
- 同一の観察者が実験してその差を調べる：同一の2群の差
- 面積なので「量」の検定でやりたい

1人の観察者の結果を比較する

2つのAUCを比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる

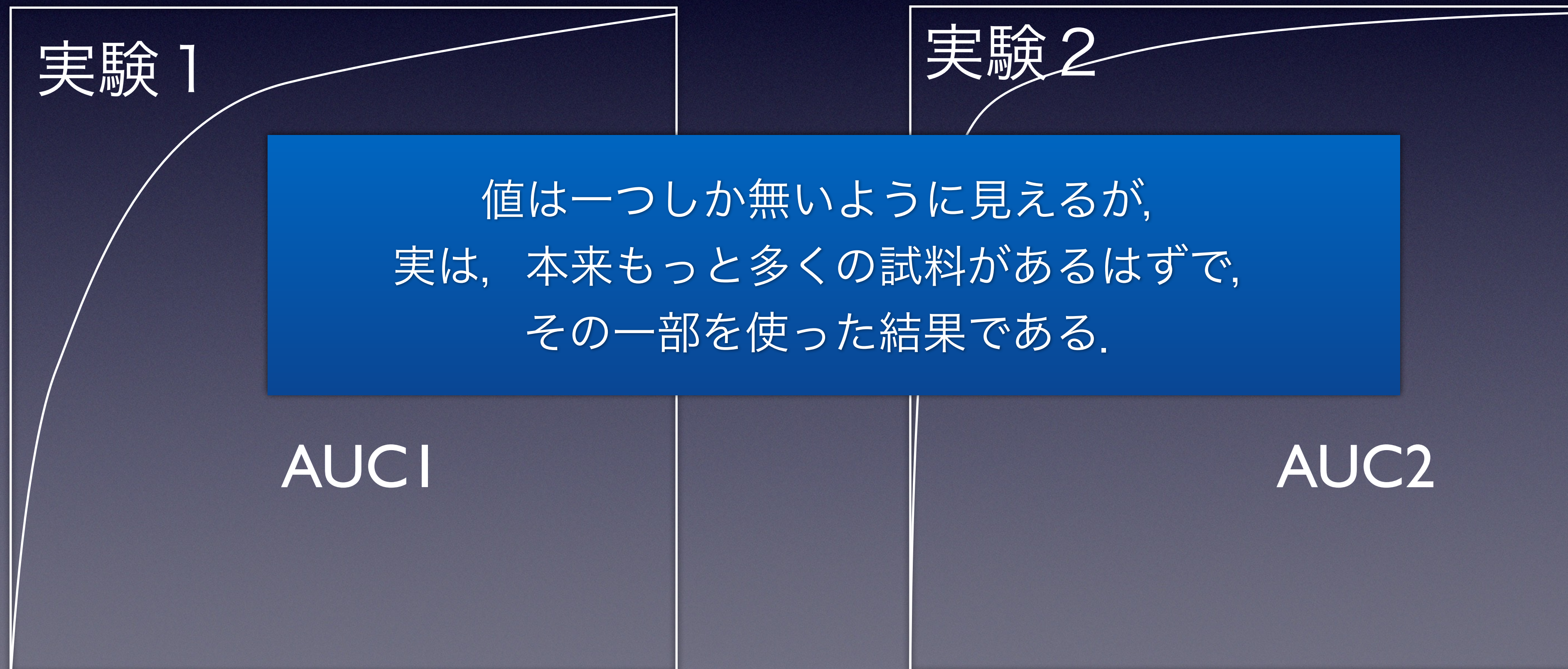


1人の観察者の結果を比較する

2つのAUCを比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



1人の観察者の結果を比較する

すなわち、母集団から標本を作り、
そこで得られたAUCであるので、このAUCには変動が含まれる。
したがって、このAUCは標本AUCとも言える。
ならば、この標本AUCから母AUCを推定（区間推定）してやろう

実験 1

AUC1

実験 2

AUC2

値は一つしか無いように見えるが、
実は、本来もっと多くの試料があるはずで、
その一部を使った結果である。

1人の観察者の結果を比較する

すなわち、母集団から標本を作り、
そこで得られたAUCであるので、このAUCには変動が含まれる。
したがって、このAUCは標本AUCとも言える。
ならば、この標本AUCから母AUCを推定（区間推定）してやろう

実験 1

実験 2

値は一つしか無いように見えるが、
実は、本来もっと多くの試料があるはずで、
その一部を使った結果である。

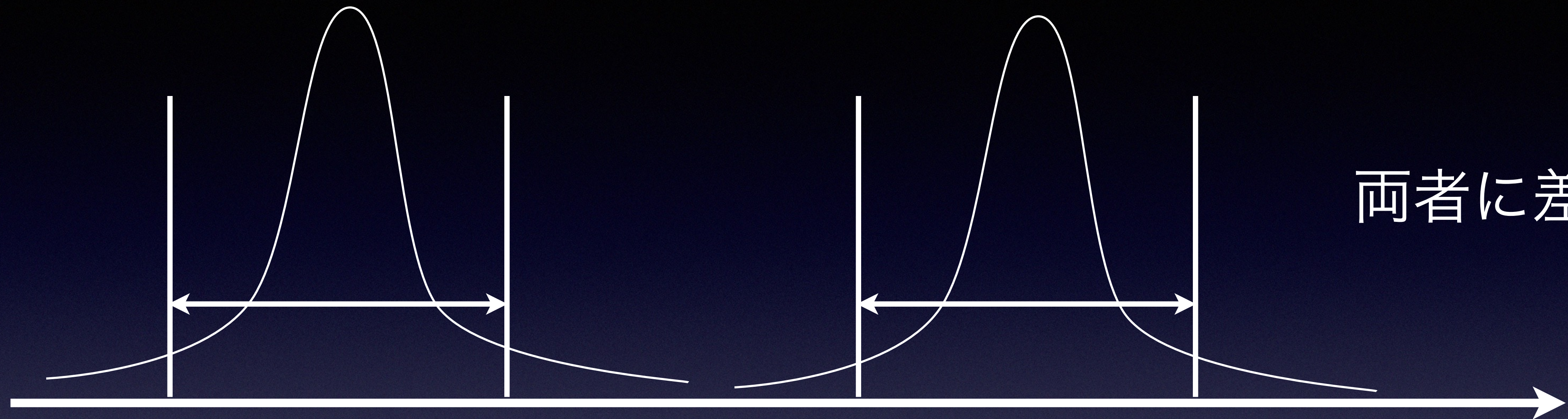
AUC1

AUC2

そのためには、確信度を様々方法でリサンプリングして変動を調べて、
精密に、母AUC（実は、AUCの平均の分布）を定めよう

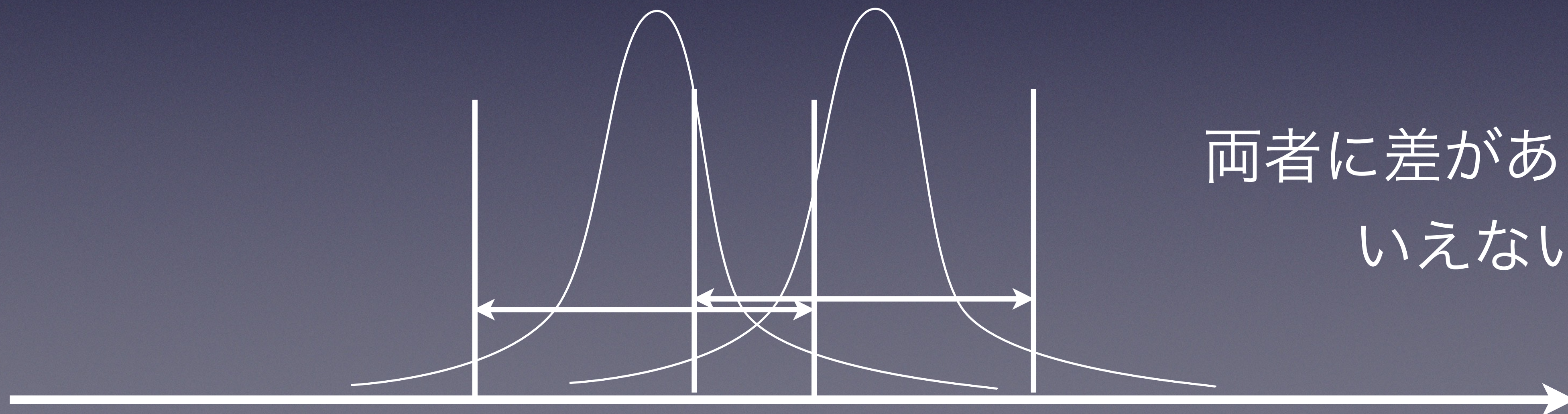
95%信頼区間

95%信頼区間



両者に差があり

95%信頼区間



両者に差があるとはいえない

どの検定方針を採用するか？

パラメトリック
平均, 分散
(量の検定)

ノンパラメトリック
代表値, 順序
(質の検定)

同一の群 対応のあるt検定, F検定

Wilcoxon符号付き順位検定

独立した群 t検定, F検定

Wilcoxon-Mann-Whitney検定
(U検定, Wilcoxon順位和
検定も同じ意味)

どの検定方針を採用するか？

パラメトリック
平均, 分散
(量の検定)

ノンパラメトリック
代表値, 順序
(質の検定)

同一の群 対応のあるt検定, F検定

Wilcoxon符号付き順位検定

独立した群 t検定, F検定

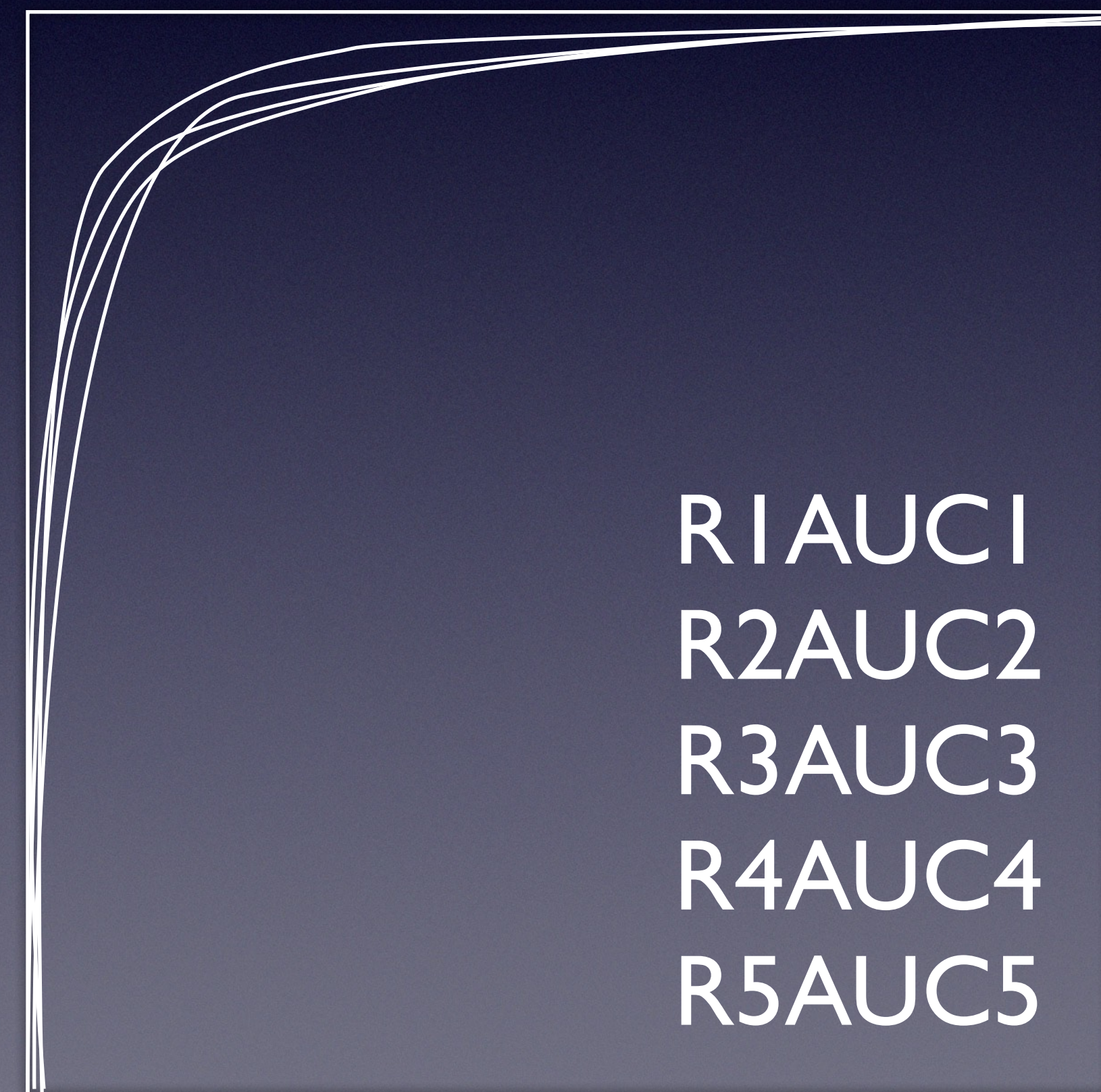
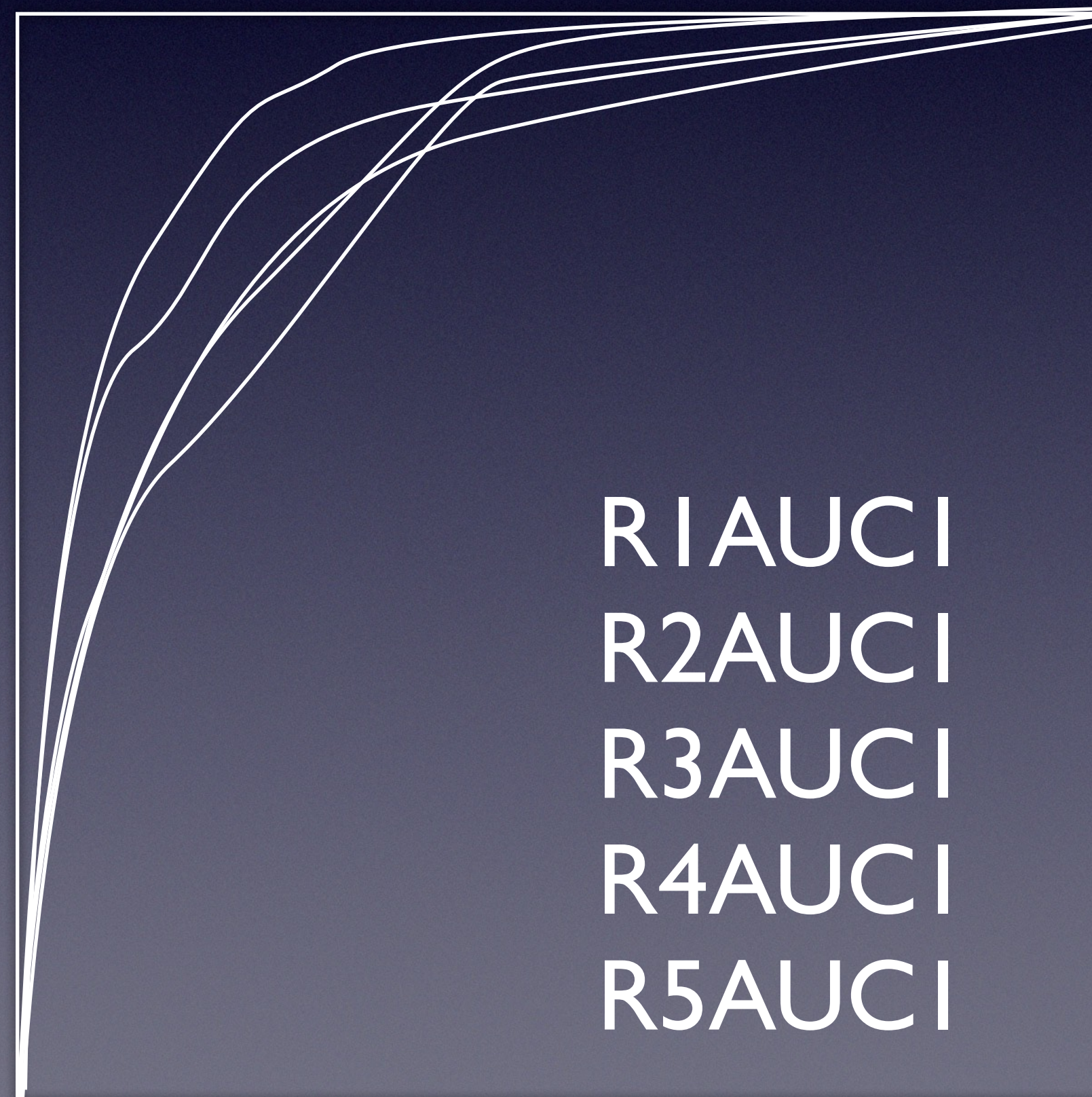
Wilcoxon-Mann-Whitney検定
(U検定, Wilcoxon順位和
検定も同じ意味)

複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる

5人の結果からそれぞれのAUCの平均を求める

R1AUC1
R2AUC1
R3AUC1
R4AUC1
R5AUC1

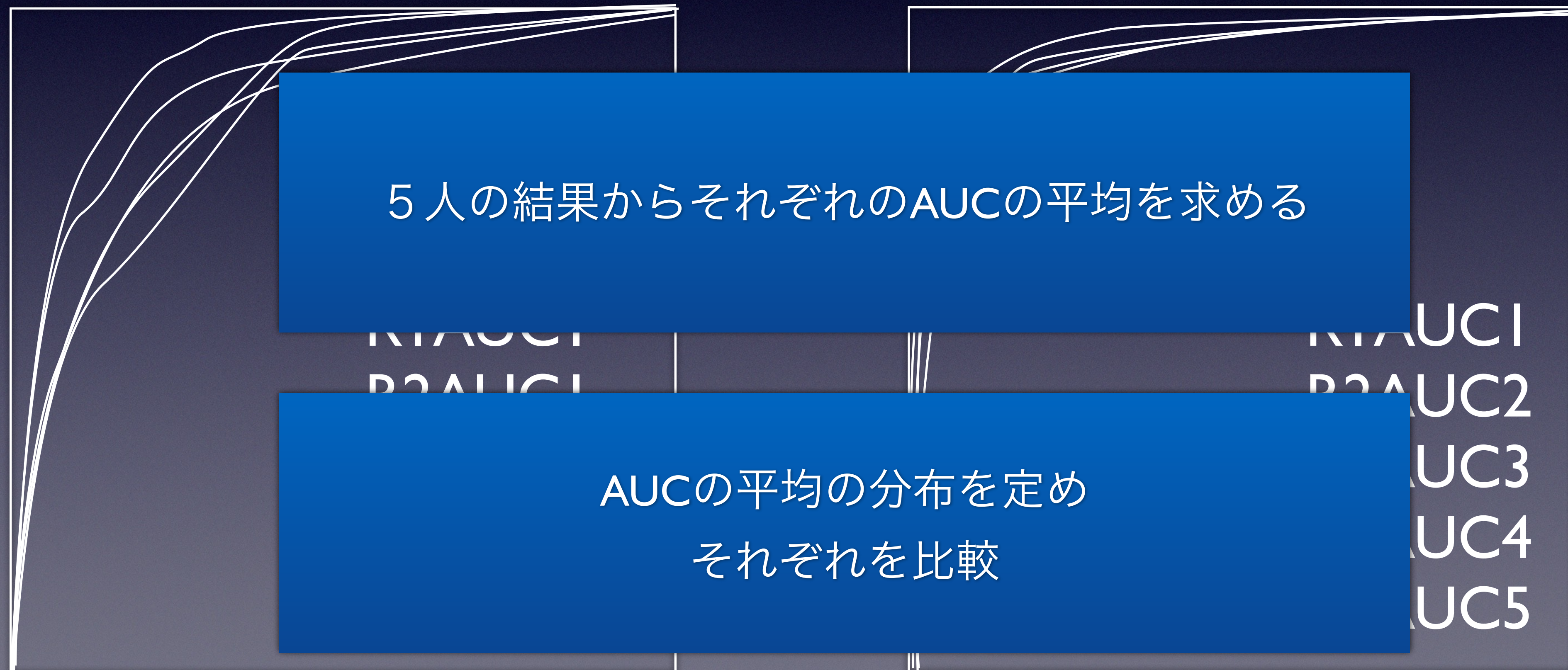
R1AUC1
R2AUC2
R3AUC3
R4AUC4
R5AUC5

複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つの実験で得られたAUCの変動を調べる

→それぞれの実験においてAUCの信頼区間を調べる



複数の観測者を比較する

2つのAUCの平均（違うときもあり）を比較する

2つ

→それぞれ

それでは、

それをパラメトリックに処理する？

ノンパラメトリックに処理する？

る

5人の結果からそれぞれのAUCの平均を求める

AUCの平均の分布を定め

それぞれを比較

AUC1

AUC2

AUC3

AUC4

AUC5