

「Excel」→「js-STAR」→「R」で  
 初歩の統計処理を学ぶ講座  
 <番外編>



## 統計的仮説検定の原理

記述統計と推測統計  
 母集団と標本集団  
 背理法  
 帰無仮説と対立仮説  
 出現確率・有意水準  
 効果量・検出力



## 記述統計と推測統計の関連

### 記述統計

情報の集約を目的とした統計処理

>「**代表値**」「**散布度**」「**関連度**」の算出



### 推測統計

現在ある情報から新しい情報を推測しようとする統計処理

>各種の「**検定法**」など

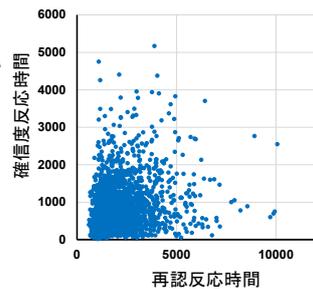


## 記述統計 散布図

### 散布図グラフ

全体の様相を知るにはよい  
 「ローデータ」と呼ぶ

>しかし、データを  
**集約**していない



- わかることも多いが「結局どうということ？」  
 >人間は同時にたくさんのデータを把握できない



## 統計的な情報の集約「代表値」

- ローデータの持つ情報をできるだけ多く含む指標が望ましい  
 >「代表値」と呼ぶ

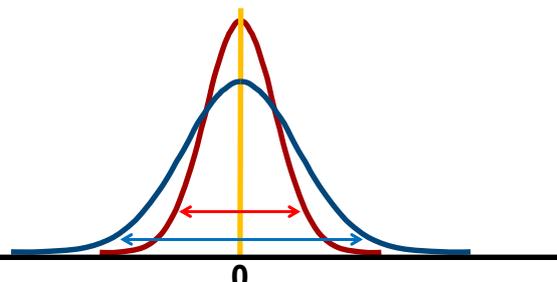
- 最頻値(モード)(名義尺度以上)
- 中央値(メディアン)(順序尺度以上)
- 相加平均(間隔尺度以上)
- 相乗平均・調和平均(比率尺度以上)



- 心理学的測定法では「相加平均」が多い



## 代表値だけでよいのか？

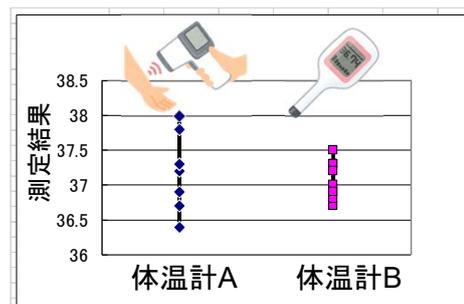


平均値(代表値)だけでは  
 この二つのグラフの違いを伝えられない

「**散らばりの程度**」を指標として含める必要がある 「**散布度**」

## 散布度と測定の精度

- ばらつきが**小さい**ほうが「精度が高い」



## 代表値と散布度の意味

- **代表値**  
 > その集団をもっともよく表現できる値
- **散布度** = 「代表値の**精度**」を示す  
 > どのくらい信頼できる代表値なのか  
 → 散らばり具合が小さい方が信頼できそう
- 調査や実験では「**散布度**」情報が大事  
 研究がどの程度の**精度**で行われたのかわることができるため



## 散布度の例:「分散」

データの「ちらばり(散布度)」を示す情報が必要

散布度の情報として  
 「一つ一つのデータと“全体平均値”との差(偏差)」を累積( $\Sigma$ )する

> ただし、そのまま加算すると定義上0になるため  
 偏差を「2乗」してからデータ個数で割った値を採用

これを「**標本分散**」(variance)と呼ぶ

## 標準偏差(SD)

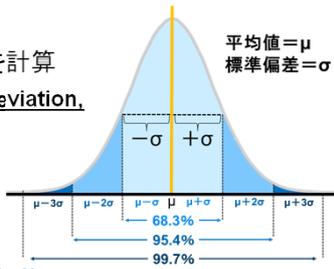
- 分散値は「二乗」されたデータが含まれるため意味がつかみにくい

**分散の平方根**を計算  
 = 「**標準偏差**」(Standard Deviation, SD あるいは  $\sigma$ )

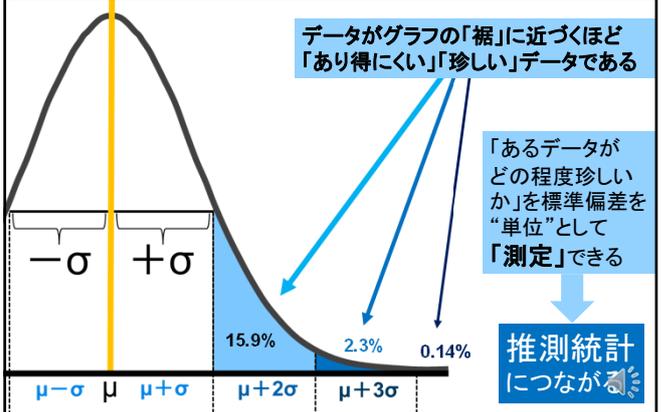
- 標準偏差の特徴  
 (1)元の測定値と同じ単位  
 (2)数学的な特徴がある

例: 中心から±1標準偏差以内に(理論上)全体のデータの約68%が含まれる

標準偏差は「**データのちらばりを計るものさし**」として使える

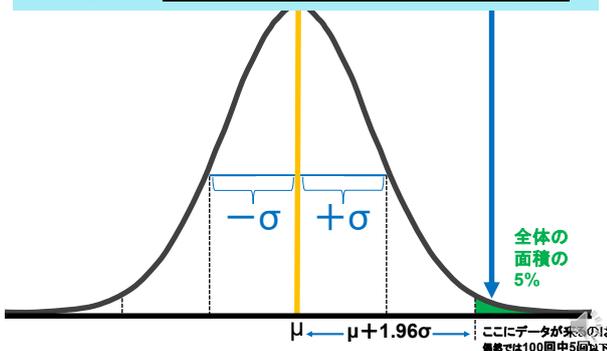


## 標準偏差を“単位”とした「珍しさ」の定義



## 面積での考え方

面積で「**5%**」の領域に注目(1.96SD分右の位置)  
 偶然には**100回に5回以下しか起きない領域**



## 母集団と標本集団

### 母集団

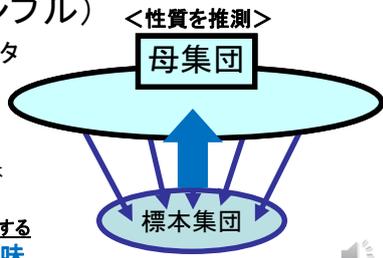
- 本来対象としたい(仮想的な)全体
- 例: 人類全体のデータ

### 標本集団(サンプル) <性質を推測>

- 実際にとれたデータ
- 例: 参加者のデータ

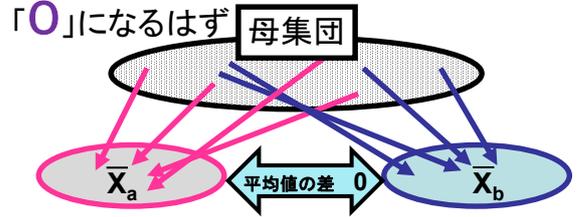
- 母集団である人類全体のデータを得ることはできない
- そのため現実で得られた標本(参加者データ)から

母集団の性質を推測する  
これが推測統計の意味



## 同一母集団からのサンプリングの性質

- 2つのグループが元が同じ母集団(分布)からのサンプリングである場合
- 2つのグループの平均値の間には(本来は)差がないはず
- データ同士の引き算をすると「平均的に」



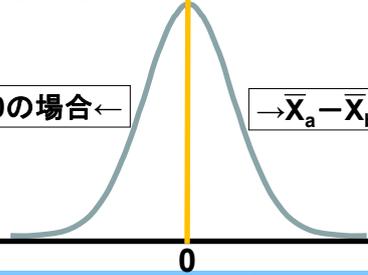
しかし毎回「完全に0」になるわけではないことに注意

## 同一母集団からのサンプリングの性質

$\bar{X}_a - \bar{X}_b$ の値は0を中心として散らばる

$\bar{X}_a - \bar{X}_b < 0$ の場合 ←

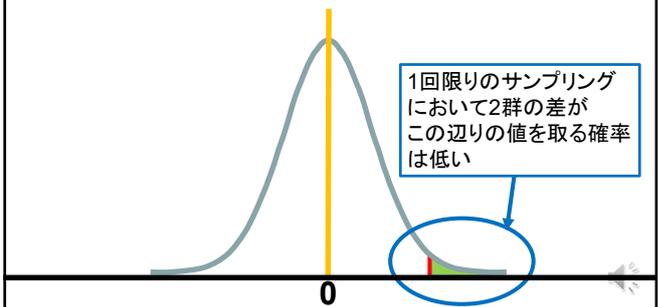
→  $\bar{X}_a - \bar{X}_b > 0$ の場合



一つの母集団からランダムサンプリングを行って「2群の平均値の差」をとることを「多数回」繰り返すとその差の値は「0を中心とした正規分布」をとる

## 偶然による差異?

- ◆ 偶然の揺らぎによって、同じグループからのサンプリングでも「ある程度の平均値の差」が認められることが「普通」
- どの程度の差であれば「普通」ではないのか?



## 偶然からの「ずれ」を測る基本的な考え方

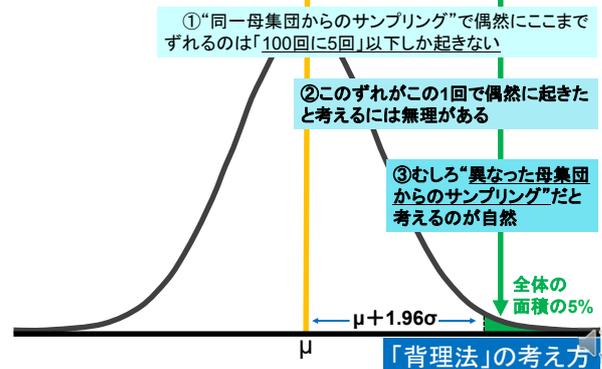
$$\text{偶然の生起から離れている程度} = \frac{\text{2群の平均値の差}}{\text{(2群を込みにした)標準偏差}}$$

- 標準偏差を「単位」として
- 2群の差の大きさが「偶然」から離れている程度を測定する

>これが0に近いほど偶然である割合が高く  
値が大きくなるほど偶然とは考えにくくなる

## 面積での考え方

2群の平均値の差の出現確率がこの領域になった



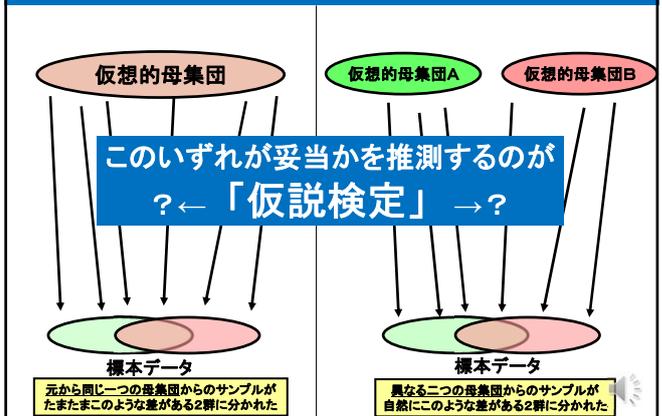
「背理法」の考え方

## 統計検定の目的＝母集団の識別

- 差が“偶然程度”であるということ  
「同じ母集団からのサンプリングだろう」  
元々2群のデータの中に「**差はない**」とみなす  
↓  
どちらの解釈が妥当なのか？
- 差が“偶然程度の大きさではない”ということ  
同じ母集団からのサンプリングとするには無理がある  
異なった母集団からのサンプリングと考えるのが自然  
2群のデータの中に「**差がある**」とみなす(仮に)  
厳密には「**差がないとは言えない**」という考え方



## 2つの仮説の図示



## 統計的仮説検定の原理

### • 背理法

- 「差があるか否か」を確認する手法として  
推測統計では「**背理法**」を用いる
- まずはじめに「差がない」ことを仮定し  
その仮説が否定されたことをもって  
「**差がないとは言えない**」ことを述べる  
間接的に「差がある」と主張する



(注意: 背理法が適用できるためには2つの事象が「排反事象」である前提)

## 背理法の仮説検定への適用

- 背理法とは
  - 否定したい仮説を否定することで  
本当に成立させたい仮説を成立させる論法
- 統計的仮説検定における「**帰無仮説**」と「**対立仮説**」
  - 帰無仮説( $H_0$ ): 否定したい仮説
    - 「実験の結果、両群の平均値に差はない」  
同じ母集団からのサンプリング
  - 対立仮説( $H_1$ ): 本当に証明したい仮説
    - 「実験の結果、両群の平均値には差がある」  
異なる母集団からのサンプリング



## 帰無仮説と対立仮説

- 帰無仮説(否定されるべき仮説)  
 $H_0$  「薬に効果はない」  
(2群の間に体温の差は統計的に認められない)  
これを否定することで
- 対立仮説  
 $H_1$  「薬に効果がある」  
(2群の間に体温の差が統計的に認められる)



こちらを間接的に主張したい

## 統計的仮説検定の2種の誤り

統計的仮説検定の2種類の誤り

- (1)「帰無仮説が正しいのに棄却してしまう」誤り  
=「誤って対立仮説を採用してしまう」誤り

### 第1種のエラー(α)

- これは「有意水準」で制限をかける

- (2)「対立仮説が正しいのに棄却してしまう」誤り  
=「誤って帰無仮説を採用してしまう」誤り

### 第2種のエラー(β)

- これは有意水準だけでは統制できない
- 帰無仮説と対立仮説の「分布の重複」が問題になるため

## 「有意水準」α

- 統計的検定において、ある統計値が算出された場合にどの基準で「差がある(偶然ではない)と言える」と判断するのか

> 厳密な数学的規定はない

- 経験的に「5%」or「1%」という値を利用する  
これを「有意水準」と呼ぶ  
「歴史的なお約束」の値

- 統計値の「出現確率p」がこの有意水準を下回る

「差がないとはいえない」と考える



## 「出現確率」p

- 確率“p”で表現される

- その結果が「偶然だったらどの程度の確率で生じるのか」を示したもの(帰無仮説の元でのデータの“自然な”発生確率)
- これが有意水準αよりも小さい場合には「その現象がそのとき偶然に生じていた」ととらえるよりも「何らかの理由で必然的に生じていた」と解釈するほうが自然

- 記述例 “ $p=0.023$ ” (最初の“0.”は略されることが多い)  
「両群の母集団が同じ場合にこの差が偶然に生じる確率は2.3%である。」



## 参考:「有意確率」という“用語”

- 「有意確率」なる用語は使わない方が良い



- 統計検定は「出現確率」pと「有意水準」αの比較によってなされる
- 「有意確率」なる用語は上記二つの概念を混同している
  - 「出現確率」は純粋な計算結果の数値: 人間の判断は関係ない
  - 「有意水準」は統計運用者が決めた“お約束”の値: 数学的な基準は本来ない
- 両者は全く別の概念

- 数表を利用していた過去の統計検定の手続きの際に「出現確率」と「有意水準」をまとめた書き方が利用されていたことの名残ではないかと推測される
- 例:  $t(20)=4.22, p<0.05$  ←このpの部分は本来は「得られた統計量の出現確率」を書くべきである。だが、当時はコンピュータが研究者の手元になかったために統計量の出現確率の計算が難しく、代用として「ある統計量の出現確率をあらかじめ計算しておいた数表」を利用して検定を行った。
- その数表の制約(有意水準として定めていた5%の出現確率値しか掲載されていない)の関係で、「ある統計量の出現確率が有意水準である5%以下であること」しかわからない＆書けない時代が長かった(当時は必要十分だった)。そのため、この“ $p<0.05$ ”という記述は「出現確率」と「有意水準」が混ざった意味(「この統計量の出現確率は有意水準以下である」となっており、その結果として「有意確率」という謎の用語?で表現していたのではないかと推測される)
- 実際に「有意確率」なる言葉の意味は「出現確率」として説明しているサイトもあれば、「有意水準」として説明しているサイトもあり、世間的にも混乱している。そもそも不要な用語であるため、使うべきではない

## 「検出力」 $1-\beta$

- 「第2種のエラー(β)」とは
  - 実際には差があるのに
    - 対立仮説を棄却して
    - 帰無仮説を誤って採択する 確率



- 「検出力( $1-\beta$ )」とは
  - 実際に差がある状態で
    - 帰無仮説を棄却して
    - 対立仮説を正しく採択する 確率

**0.8程度が妥当** 「検出力」を高める条件設定が望ましいとされる

## 統計的仮説検定による判断の構造

		＜真実＞	
		差が「ある」 (元が異なる母集団)	差が「ない」 (元が同じ母集団)
＜仮説＞	対立仮説(H <sub>1</sub> ) 「差がある」を採択	$1-\beta$ 検出力	$\alpha$ 危険率 (有意水準)
	帰無仮説(H <sub>0</sub> ) 「差がない」を採択	$\beta$ 誤棄却率	$1-\alpha$ 正棄却率

### 仮説検定の原理 帰無仮説の分布

- 2変数の「差」の問題のため、帰無仮説側は常に「0」が仮定される(=差がないという仮説)
- 実際には確率変数として「平均値0の分布」を取る
- 有意水準 $\alpha$ と統計値の出現確率がこの帰無仮説の分布のもとに算出される

本当は帰無仮説が正しいのに対立仮説を採用するという事態は出来るだけ避けたい

そうしたことが生じる確率を $\alpha$ 以下で抑えたい(通常 $\alpha=0.05$ )

▲部分の面積を全体の中の $\alpha$ に押さえる基準を「有意水準」として設定する

### 仮説検定の原理 対立仮説の分布

- 対立仮説は実際に得られたデータからもたらされる
- 対立仮説の分布の仮説採用の可否の判断を切り替えるための**判定基準**が第2種のエラーの割合( $\beta$ )と検出力(検出確率)( $1-\beta$ )を分ける境界となる

本当は対立仮説が正しいのに誤って帰無仮説を採用するという事態も出来れば避けたい

ある**判定基準**の下でそうしたことが生じる確率は $\beta$ と置かれる

対立仮説を正しく採用する確率 =  $1-\beta$

### 「検出力(検定力)」と「効果量」

- 検出力の判定基準は帰無仮説分布の有意水準点
- この値よりも大きな対立仮説分布の面積が「検出力」
- 分布間の距離が「効果量」

### 「効果量」

- 実験操作の大きさ
  - 帰無仮説と対立仮説のグラフ上の「距離」に相当
  - 効果が明確な実験結果かどうか

**効果量 = 両群の平均値の差 / 両群を込みにした標準偏差**

- 統計的な検定の有意性とは別の軸での評価
  - サンプル数を増やしすぎると統計的には有意だが実質的には無意味な結果が出てしまうことも
- 効果量は実験操作の“実質的な意味”を示す
- 検定手法によっていろいろな値を利用する
  - js-STARの分散分析では偏 $\eta^2$ 乗

### 「検出力(検定力)」と「効果量」

検出力を上げるためには: 分布の重なりを小さくする

→(1)分布間の距離を大きくする ← 効果量の増加

### 「検出力(検定力)」と「効果量」

検出力を上げるためには: 分布の重なりを小さくする

→(2)分布の散布度を小さくする

(a) データ数を増やす  
散布度の式の分母に参加者数が含まれる

(b) 実験精度を上げる  
参加者による偶然誤差が小さくなるように

散布度の減少

αの位置が帰無仮説の「0」に近づく

効果量は変わらないが検出力は増加する

## 無価値な有意差

統計的には“有意”  
だが効果量 0

実質的には  
価値がないに等しい  
のでは？

データ数が多すぎると  
統計的検定は意味を失う

この結果に実質的な  
意味があるのか  
検討の必要がある

0x

0.1

## 実験前の計画

- 事前にデータ分布がほぼわかっている場合

先行研究などがある場合

- 代表値と散布度がだいたいわかっている
- 効果量がだいたいわかっている
- 検出力もだいたいわかっている

ような実験の「追試」を行う

- この条件が満たされている場合

研究に必要な「検出力(1-β)」「危険率(α)」「効果量」を  
あらかじめ確認し それを満たすために必要な「サンプル数」を  
**実験前に計算しておくことが一応可能**

効率の良い研究が可能

- ただし「新しい研究」では分布等が未知なため  
不可能(多くの当ゼミの研究に当てはまる)

割とやってみないとわからない

「Excel」→「js-STAR」→「R」で  
初歩の統計処理を学ぶ講座  
＜番外編＞



## 統計的仮説検定の原理

記述統計と推測統計  
母集団と標本集団  
背理法  
帰無仮説と対立仮説  
出現確率・有意水準  
効果量・検出力

<完>

