統計的仮説検定

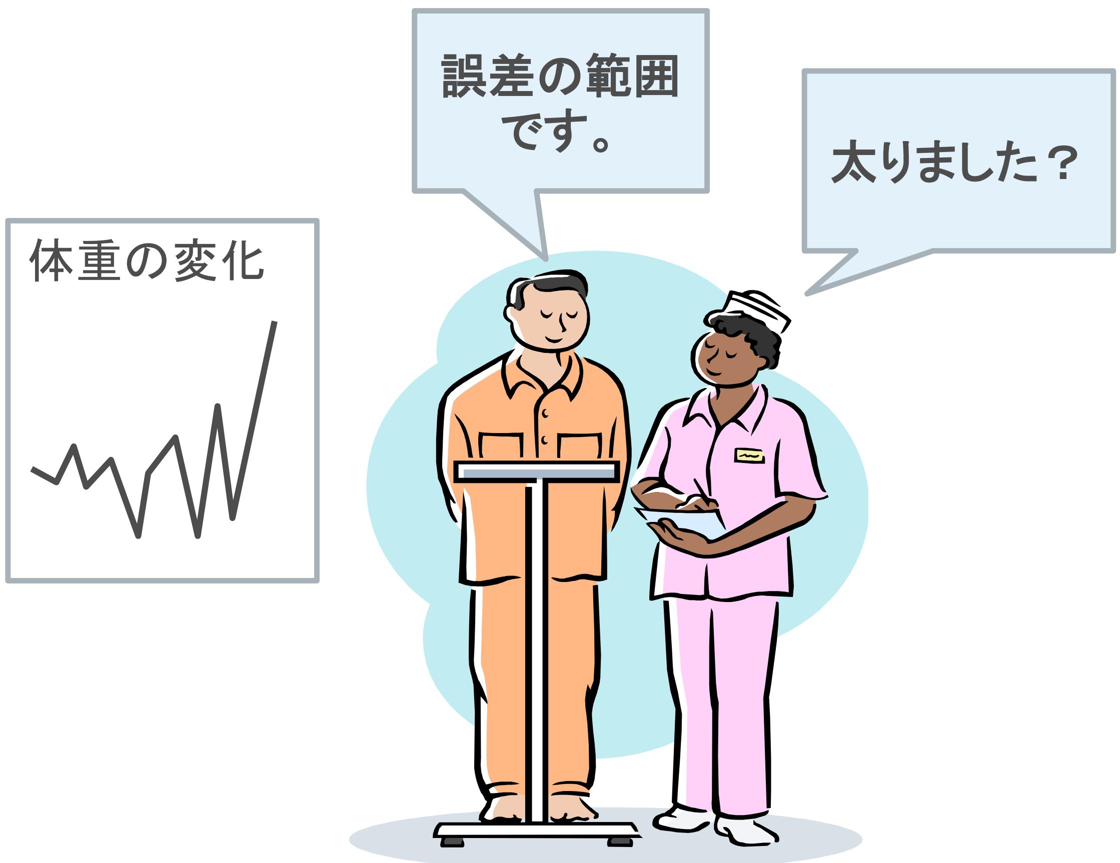
ATR脳情報解析研究所　計算脳イメージング研究室 室長
理研革新知能統合研究センター　チームリーダ
CINET　客員研究員
大阪大学院　生命機能研究科　客員準教授

山下　宙人

参考文献

統計学とは、

ばらつきを伴う情報を、客観的に分析・評価する学問



統計学の二つの機能

機能1. 少数の情報から、全体を推し量る･･･推論する

部分から全体を推し量る機能

検定：標本を基準に母集団の特徴や状態について何らかの仮説を設け、その妥当性を確率論的に検証する方法

例）先月の雨量は例年どおりとみなしてよいか

推定：標本の特徴（特徴量）から、母集団の特徴（母数）を推し量る方法

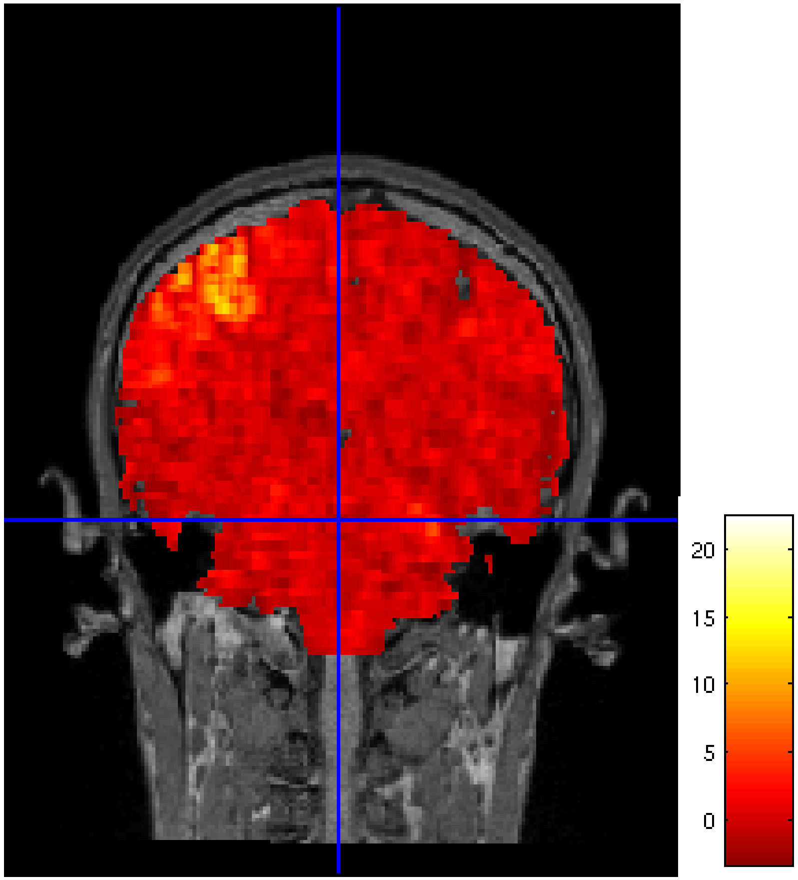
機能2. 情報の基本構造を明らかにする･･･解析する

標本から得られた情報を数量的に要約したり分類する機能

例）50m走、1500m走、ボール投げ、走り幅跳びの結果から生徒の体力を評価したい。

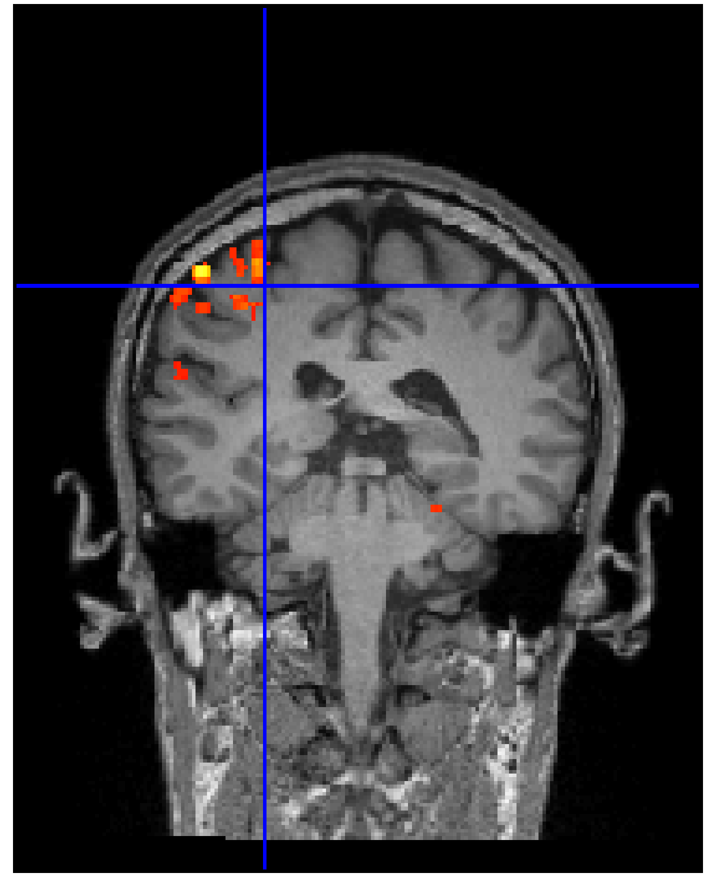
ヒト脳機能イメージングによる仮説検定：ブレインマッピング

T-Map



統計検定
（閾値処理）

Activation map

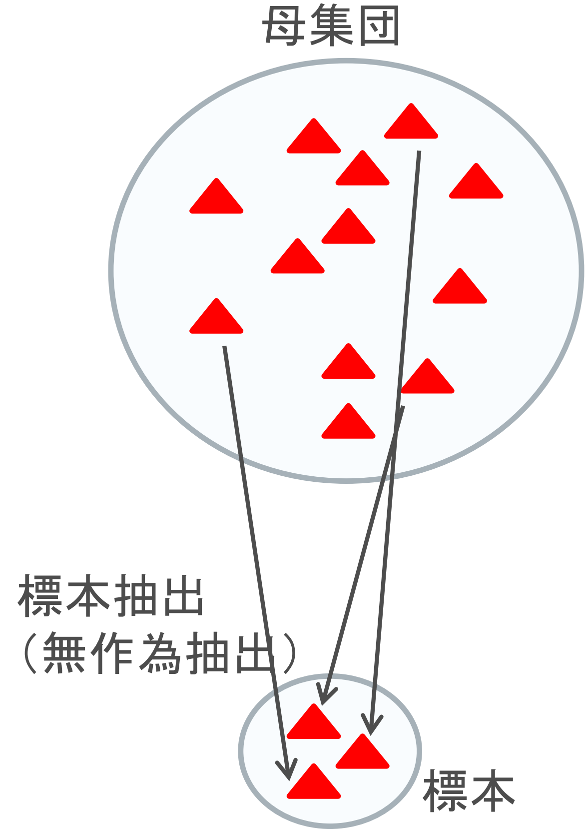


基本概念とキーワード

母集団 : 調べる対象となる有限または無限の個体の全集合

標本 : 母集団を調べるため、そこから取り出した個体の集合

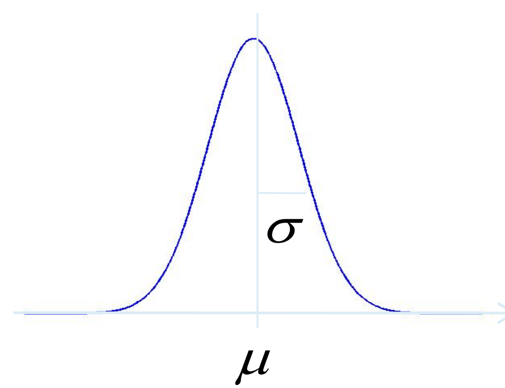
母数 : 母集団の特徴を要約した値。（平均や分散など）



平均 $\mu = E[X]$

分散 $\sigma^2 = E[(X-\mu)^2]$

標準偏差 σ



講義内容

1. 統計的仮説検定
2. 同時多重仮説検定
3. まとめ

統計的仮説検定とは

ゆらぎのあるデータから、ある仮説が正しいといってよいかどうかを統計学的・確率論的に判断すること。

科学的発見において結果の信頼性を客観的に評価する方法（の1つ）

統計的仮説検定の基本は反証の論理

「差がある、A≠B」という仮説の証明

→

「差がない、A=B」という仮説が確率論的に矛盾することを証明

統計的仮説検定：方法

1. 反証すべき仮説(帰無仮説)と対立仮説を立てる。

2. 帰無仮説の元では、**今観測されるデータが起こる可能性が低いこと**を言う。

3. 帰無仮説が低い確率で起こったとは考えずに、帰無仮説ではない仮説(対立仮説)からデータが生成された結論する　(帰無仮説の棄却)。

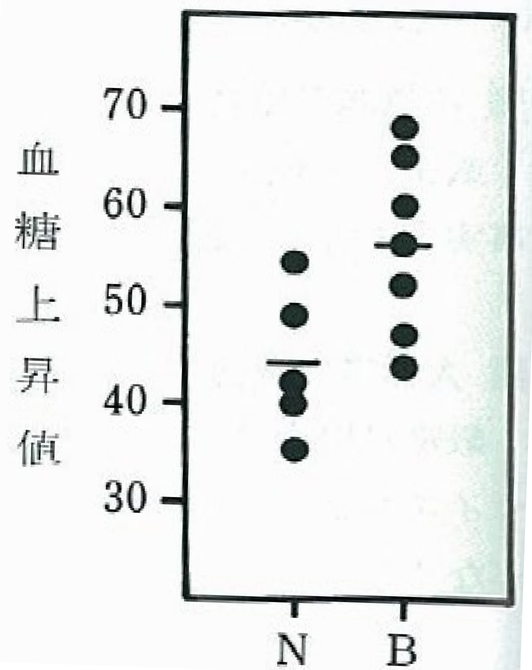
統計的仮説検定の例１

2つの被験者群に対して、負荷後の血糖上昇値を求めた.
両群間に差があると考えてよいか.

N群 x_1	B群 x_2
54	68
49	65
42	60
40	56
35	52
	47
	44

データを要約すると

	N群	B群
データ数	5	7
平均値	44	56
標準偏差	7.5	9.0



1. 仮説の設定: 帰無仮説H0: $\mu_1 = \mu_2$ 対立仮説H1: $\mu_1 \neq \mu_2$

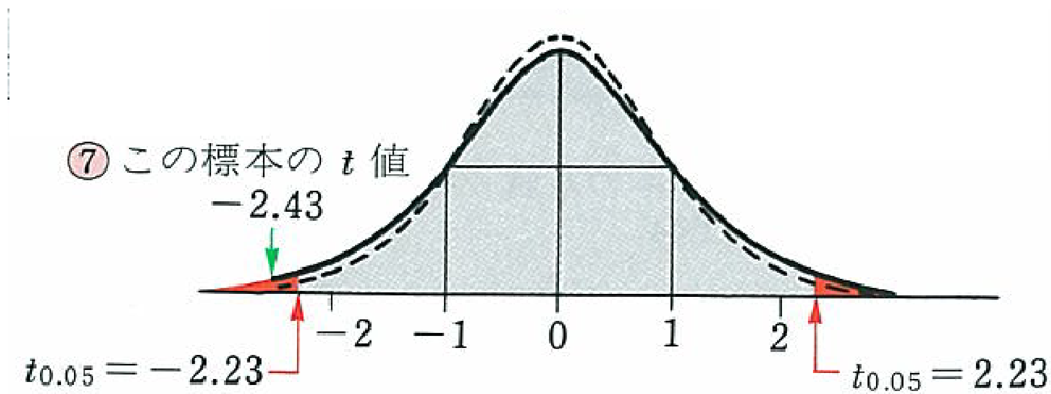
2. 統計量を求める:

$$\overline{x_1} - \overline{x_2} = 44 - 56 = -12 \qquad s = \sqrt{\frac{7.5^2(5-1) + 9^2(7-1)}{5+7-2}} = 8.43$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-12}{8.43\sqrt{\frac{1}{5} + \frac{1}{7}}} = -2.43$$

3. めずらしいか判断する:

有意水準 0.05, 自由度の10の t 値は $t_{0.05} = \pm 2.23$



例2：比率の検定（二項検定）

サイコロを5回投げたところ、1の目が3回でた。このサイコロは偏っていると考えてよいか。

例2：比率の検定（二項検定）

サイコロを5回投げたところ、1の目が3回でた。このサイコロは偏っていると考えてよいか。

設問
サイコロは偏っている？

↓

（1）仮説の設定
偏っていない　H0:p=1/6
（ H1:p≠1/6 ）

↓

（2）統計量Xを求める
1の目のでた回数r=3

↓

（3）確率を求める
H0のとき、r≧3の確率Pを求める

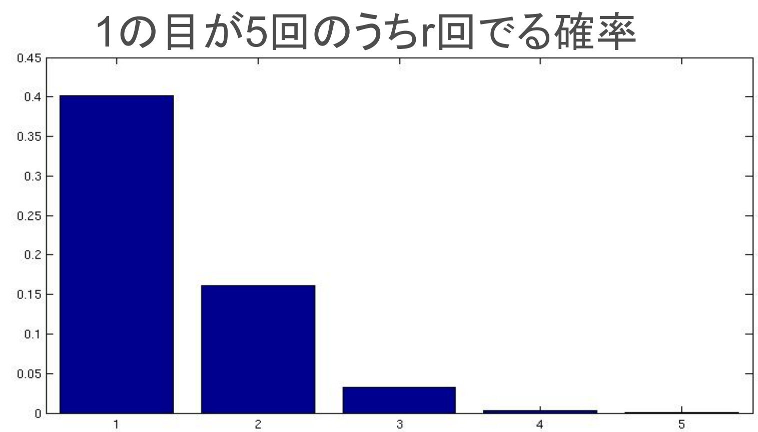
↓

（4）判定
P<0.05
→ H0を棄却し、H1を採用
（サイコロは偏っている）

（3）確率を求める

1の目がn回のうちr回でる確率

$$P_r = {}_nC_r \times \left(\frac{1}{6}\right)^r \times \left(\frac{5}{6}\right)^{n-r}$$



$$P = P_3 + P_4 + P_5 = 0.0351 < 0.05$$

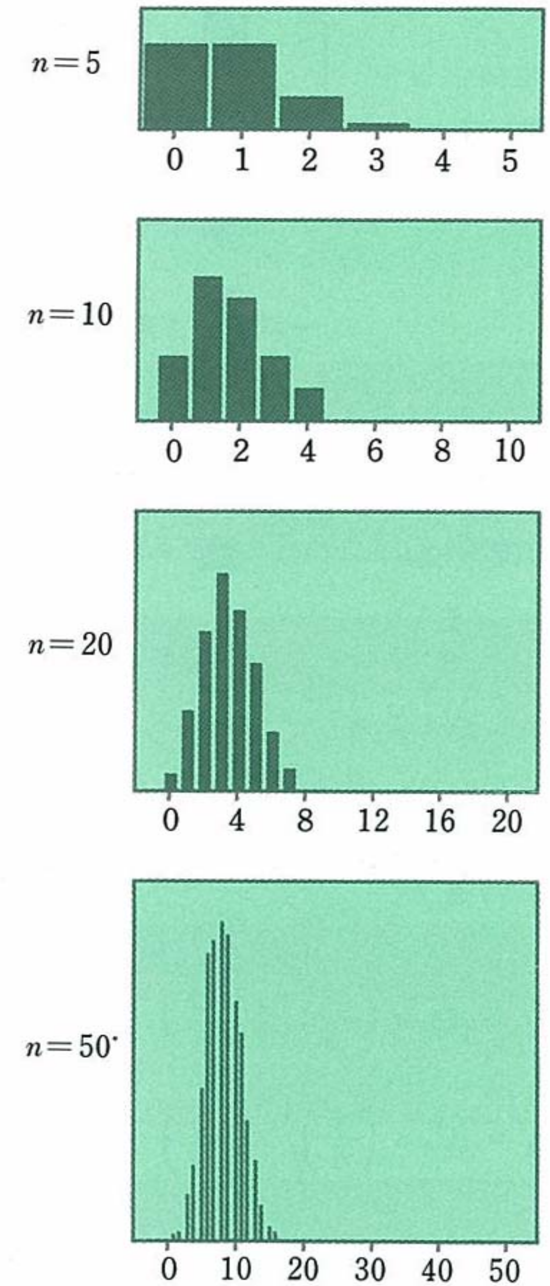
関連知識

ある事象の生じる比率がpのとき、n回試行してr回
その事象が生じる確率は、

$$P_r = {}_nC_r \times p^r \times (1-p)^{n-r}$$

で求まり、これを二項確率と呼ぶ。

試行数nを増やしていくと二項分布は次第に正規分布に
近づく。



例3：平均値の検定

あるクラスの身長の分布は、全国水準と違うか？

全国の10歳女子の身長の平均 $\mu = 140$ cm
標準偏差 $\sigma = 5$ cm

あるクラス25人の身長の平均 $\bar{x} = 137$

例3：平均値の検定

あるクラスの身長の分布は、全国水準と違うか？

設問
全国水準と違う？

↓

（1）仮説の設定
全国平均と同じ　$H_0:\mu=140$
（$H_1:\mu\neq 140$）

↓

（2）統計量Xを求める
クラスの平均値　$\bar{x}=137$

↓

（3）確率を求める
H_0のとき、$\bar{x}=137$以上極端になる確率P

↓

（4）判定
$P<0.05$
→ H_0を棄却し、H_1を採用

全国水準と違う

全国の10歳女子の身長の平均　$\mu = 140$ cm
標準偏差　$\sigma = 5$ cm

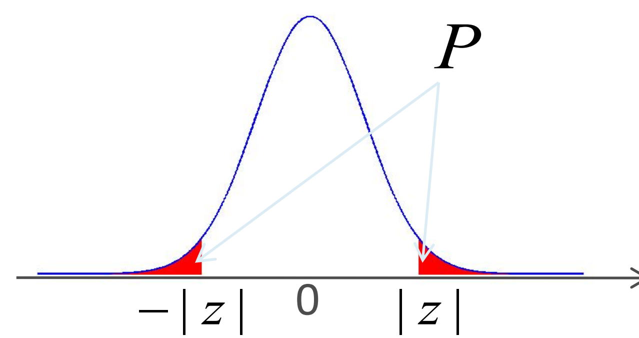
あるクラス25人の身長の平均　$\bar{x} = 137$

（3）確率を求める

H_0のとき、$\bar{x}$は正規分布$N(140, \frac{5^2}{25})$に従うはず。

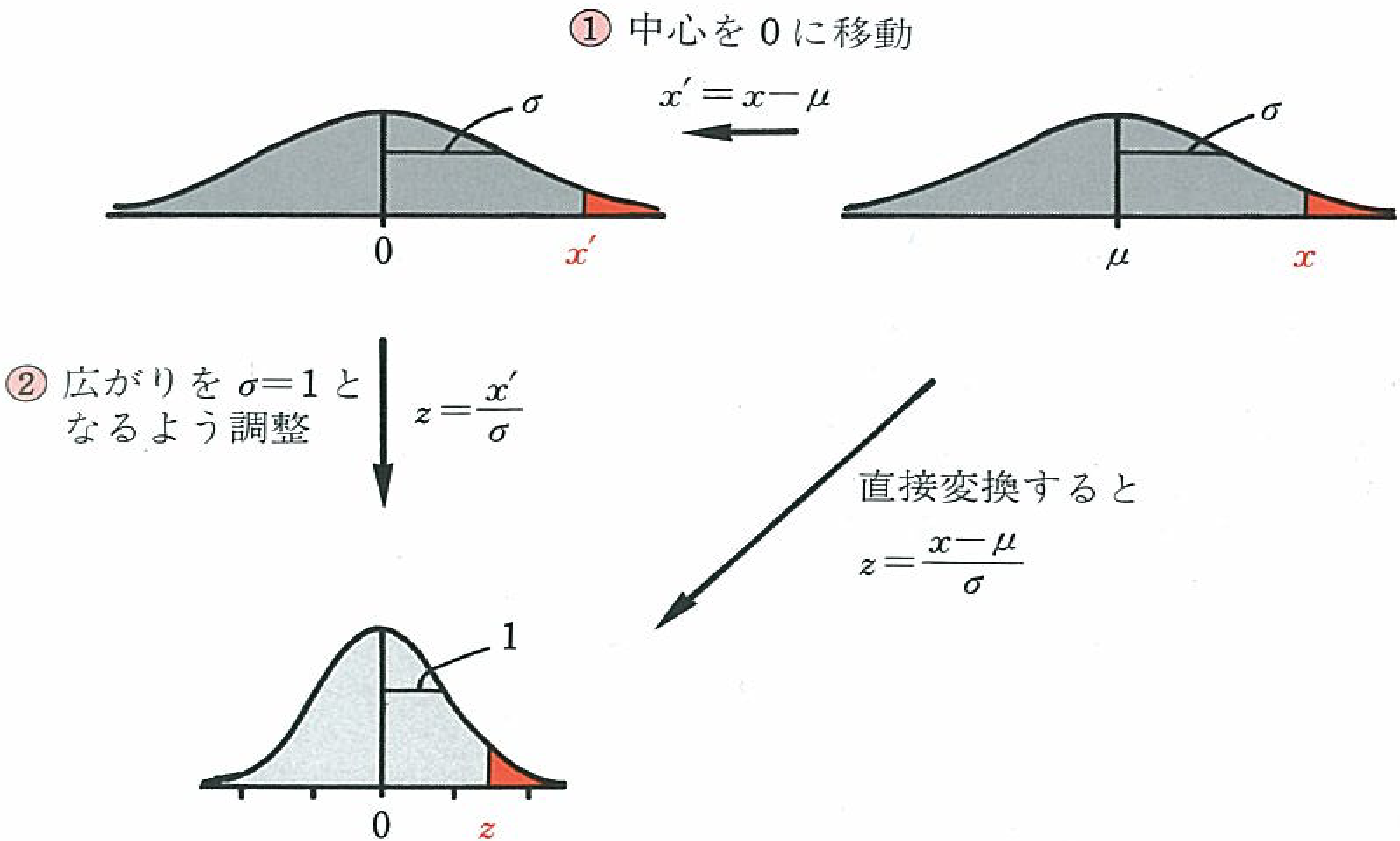
$z = \dfrac{\bar{x}-140}{5/\sqrt{25}}$　は正規分布$N(0,1)$に従うはず。

$$P = 2\times\int_{|z|}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})dx$$
$$= 0.0027 < 0.05$$



関連知識　データの標準化

一般正規分布 $N(\mu, \sigma^2)$ の値 x を、標準正規分布 $N(0,1^2)$ の値 z へ変換



仮説検定の作り方：統計量・統計量の分布・棄却域

- 統計量　：　データからそれらを要約する値への関数

  Ex.　$T:(x_{11},x_{12},\cdots,x_{15},x_{21},x_{22},\cdots,x_{27})\to t$

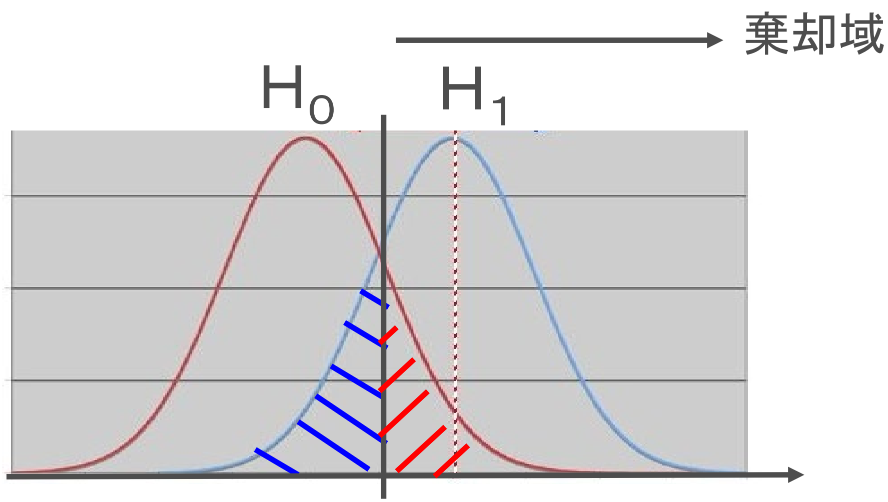
- 統計量の分布

  Ex.　t分布

- 棄却域　：　統計量の定義域上で、仮説が正しくないと判断する領域

  Ex.　{t : t > 2.23 or t < -2.23}

Type I error (第1種の過誤) と Power (検出力)



α:第1種の過誤

β:第2種の過誤

Power(検出力) = 1-β

	H_0 採択	H_0 棄却
H_0 真	1-α	α 第1種の過誤
H_1 真	β 第2種の過誤	1-β 検出力

- 第1種の過誤と検出力はトレードオフの関係
- 良い検定とはαを一定に保ち、検出力が高い検定。

ベストな検定の作り方：Neymann-Pearsonの補題

X_1・・X_nをf(x；Θ)からのサンプルとする。
帰無仮説・対立仮説

$$H_0 : \Theta = \Theta_0 \quad H_1 : \Theta = \Theta_1$$

で与えられる検定を行いたい。
この時、尤度比

$$L(\Theta_1 ; x)/L(\Theta_0 ; x) \geqq A(\alpha)$$

によって定まる検定が**有意水準αの検定で最強力な検定**となる。

単純帰無仮説・単純対立仮説の元では
尤度比検定が最強力検定となる。

中間まとめ

- 仮説検定は反証の論理
- 仮説検定の手続き
- 第1種の過誤と第2種の過誤

Question :
なぜ科学の分野では、第1種の過誤 (擬陽性) のコントロールが重視されるのか？

講義内容

1. 統計的仮説検定
2. 同時多重仮説検定
3. まとめ

同時多重仮説検定とは

複数の検定を同時に扱う検定 (Multiple Simultaneous Hypothetical Testing)

– fMRIにおいて各ボクセルごとに活動があるかを検定

– Microarray データにおいて、遺伝子ごとに発現量を検定

個別の検定で見るとめずらしいイベントでも、“検定全体”で考えたらめずらしくない。
“検定全体”でめずらしさをコントロールするには？

例：同時多重仮説検定

10万円もらえる当たる確率が5%のくじがある。
Aさん、Bさん・・・、Tさんの20人がくじを引いた結果、Kさんが当たった。

「Kさんが1回だけくじを引いたら当たった」

と言うと、めずらしいこと（確率5%で起こったこと）に感じる。

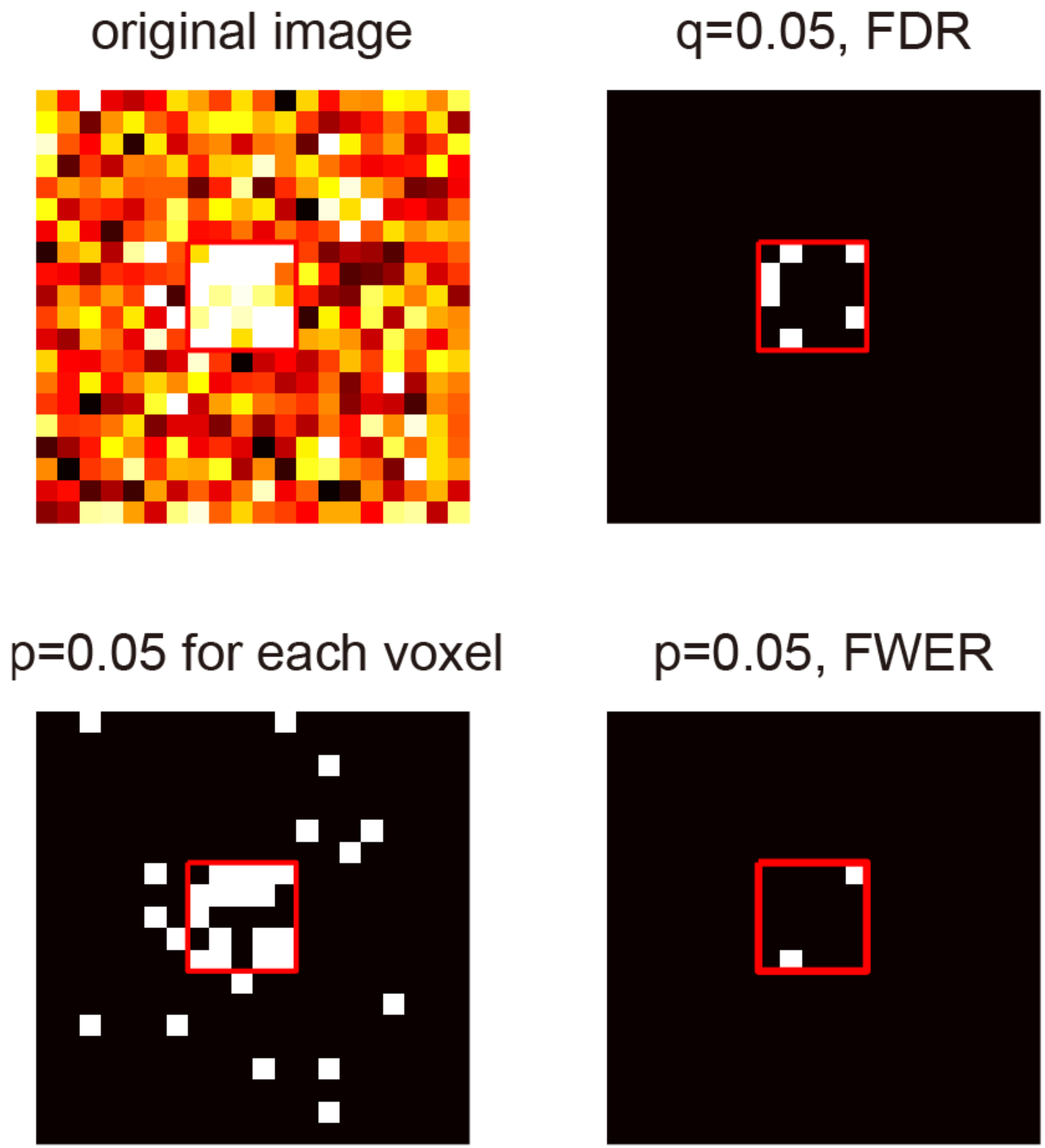
しかし、実際20人がくじを引いたことから考えれば、

「20人もくじを引けば誰かが当たる確率は0.6415であり、
誰か1人は当たりくじを引くだろう。そして、それがたまたまKさんだった。」

ということにすぎず、めずらしいことではない。

永田・吉田「統計的多重比較法の基礎」から改編

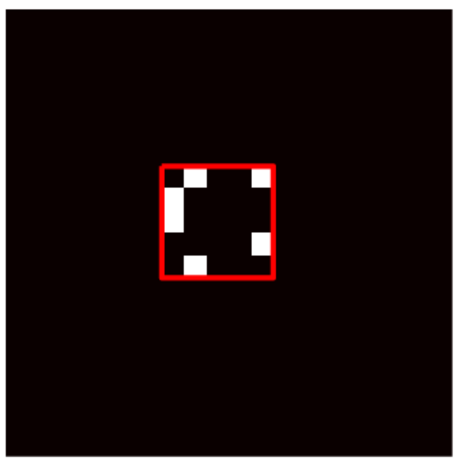
例２：同時多重仮説検定　画像シミュレーション



Family-wise error rate と False discovery rate

	H0 : Retained (not detected)	H0 : Rejected (detected)	
No activity	N0|0	N1|0	M0
Activity	N0|1	N1|1	M1
	D0	D1	

q=0.05, FDR



N0|0=375, N1|0=0
N0|1=18, N1|1=7

Family-wise error rate と False discovery rate

	H0 : Retained (not detected)	H0 : Rejected (detected)	
No activity	N0|0	N1|0	M0
Activity	N0|1	N1|1	M1
	D0	D1	

Family-Wise Error Rate : FWER = P($N_{1|0}$ >= 1)

“検定全体”で少なくとも1つは疑陽性が起こる確率

False Discovery Rate : FDR = E($N_{1|0}$ /D1) (= 0 if D1=0)

棄却された仮説のうち、間違って棄却された仮説の割合

FWERをコントロールする方法：Bonferroniの方法

FWERを抑えるもっとも基本的な方法

N個の同時多重検定を考える。
個々の検定エラー率を　v = α/N と設定すれば、FWER は　α　以下になる。

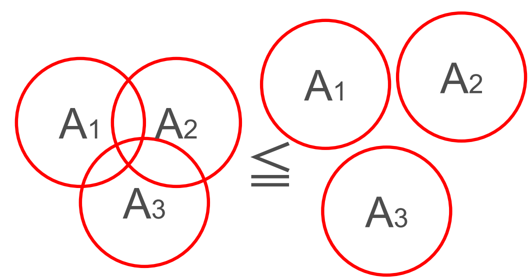
証明：　Bonferroni不等式より

$$\mathrm{FWER} = \; P\left\{\bigcup_i \{T_i \geq u_v\} \,|H_0\right\} \leq \sum_i P\{T_i \geq u_v|H_0\} = Nv$$

v：個々の検定のエラー率

Bonferroni 不等式

$$\mathrm{P}\left\{\bigcup_i A_i\right\} \leq \sum_i \mathrm{P}\{A_i\}$$



FWERをコントロールする方法：その他の方法

検定統計量の最大値の分布を利用して閾値を決める.

- ***Random field methods***
  - 最大値分布をトポロジカルに近似.
- ***Permutation methods***
  - 最大値分布をリサンプリング(=順列並び替え)により近似.

Assumption	Methods
画像が滑らかでない	Bonferroni methods
画像が滑らか	Random Field Theory
No assumption & computational power	Permutation methods

FWERの問題

検定結果が保守的過ぎる。
多重性が高くなると、新しい発見がほとんど検出されない

False discovery rate (FDR)：FWERに代わる新基準

J. R. Statist. Soc. B (1995)
57, *No.* 1, *pp.* 289–300

Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing

By YOAV BENJAMINI† and YOSEF HOCHBERG

Tel Aviv University, Israel

FDR=ポジティブと検出された個々の検定のうち、擬陽性である検定の割合
FDRの期待値をコントロールする閾値を決める。

Example

100個の同時検定

FDR=0.1 のもとで検出された
検定が10個あった

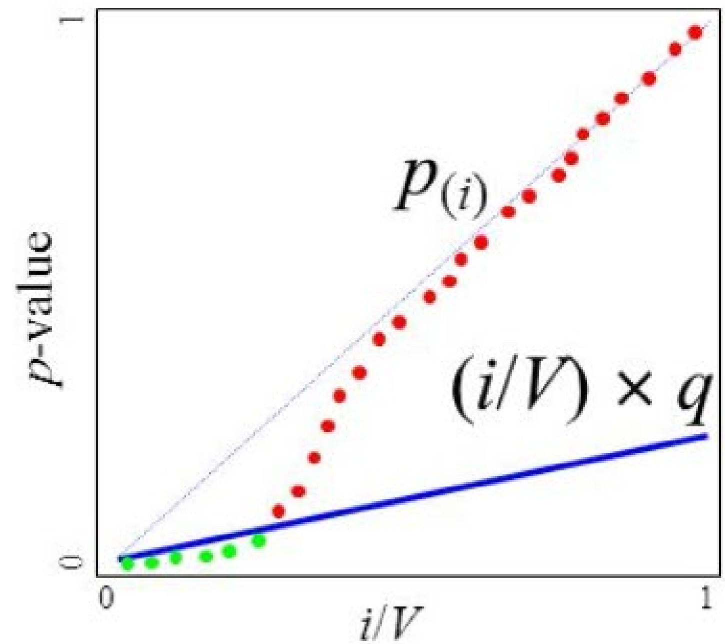
10個のうち平均して
1個の擬陽性が含まれる

FDRをコントロールする方法 ： BH法

- Select desired limit q on FDR
- Order p-values, $p_{(1)} \leq p_{(2)} \leq \ \ldots \leq p_{(V)}$
- Let r be largest i such that

$$p_{(i)} \leq \ (i/V) \times q$$

- Reject all null hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$.



i/V = proportion of all selected voxels

Benjamini & Hochberg, *JRSS-B* (1995) 57:289-300

FDRをコントロールする方法

- BH法（Benjamini and Hochberg 1995)
  Independent or Positive Regression Depenedency on Subset
- BY法（Benjamini and Yekutieli 2001)
  Arbitrary Covariance Structure
- Positive FDR (Storey 2002)

講義内容

1. 統計的仮説検定
2. 同時多重仮説検定
3. まとめ

Class まとめ

- 仮説検定は反証の論理
- 仮説検定の手続き
- 第1種の過誤と第2種の過誤
- 同時多重比較検定
- 同時多重比較検定の擬陽性をコントロールする方法
    - Family-wise error rate
    - False discovery rate