

情報統計 第7回

2022年8月3日 神奈川工科大学



櫻井 望

国立遺伝学研究所
生命情報・DDBJセンター

第6回の補足

検定で
注意すること

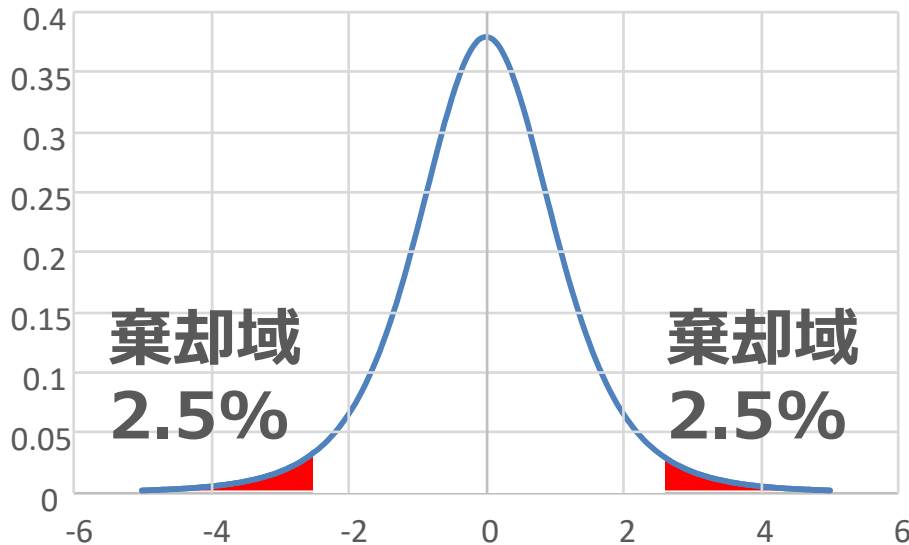
学習目標

- 第6回の補足
 - ✓ 両側検定と片側検定
 - ✓ t 検定のいろいろ
- 検定を考えるときに気をつけたいポイントをおさえる
 - ✓ 検定の間違い
 - ✓ p 値 < 0.05 にとらわれるな！
 - ✓ 多重性の問題、FDR（偽発見率）

第6回の補足

両側検定と片側検定

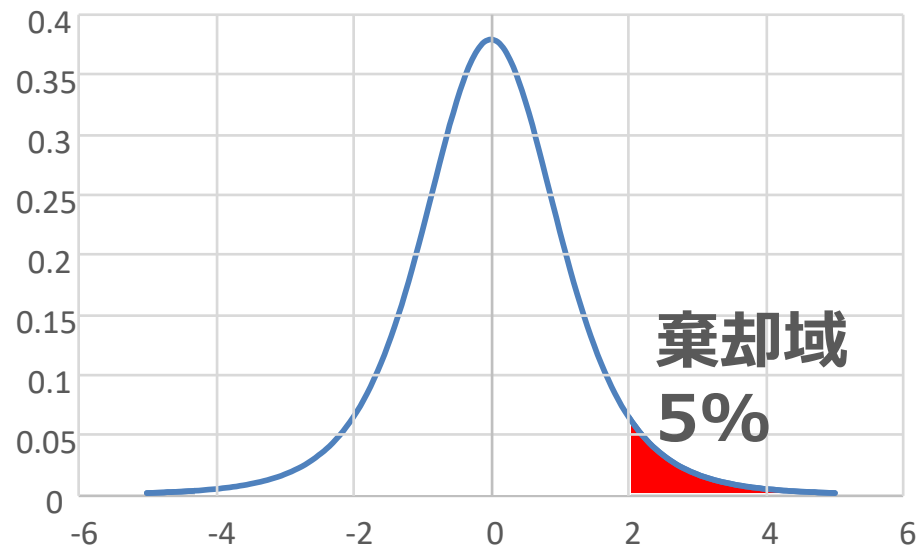
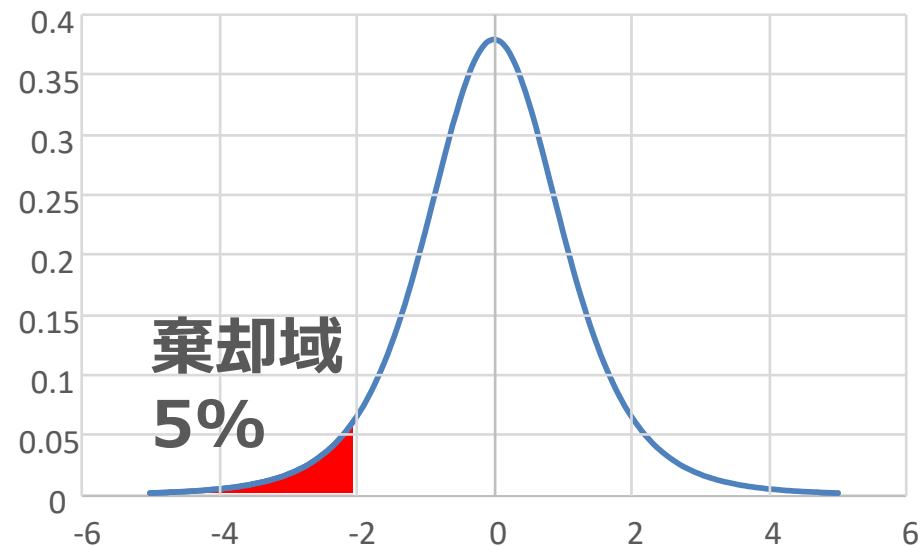
Aさんのカラオケ問題では、有意水準0.05を、その半分の0.025ずつに分け、 t 分布の両側に割り当てて考えました



これは
両側検定
と呼ばれます

片側検定

有意水準を、左右のどちらかにだけ重点配分することもでき、これを片側検定と呼びます。



片側検定をするとき

明らかにどちらかに偏っている場合だけが問題になるような仮説検定をするときは、片側検定を行うことができます

- 例)
- 蛍光灯の寿命は仕様書にある * * 時間よりも短いか？
 - 今年の給料は去年の * * 円よりも上がったか

ただ、有意水準の数字をいくつにするかだけの問題なので、**通常は両側検定で問題ありません**

色々な t 検定

t 検定には、実はいろいろあります。問題にしている群がひとつか二つか、2群の場合はさらに、対応関係があるかないかで分かります。

- 1群の t 検定

母集団の平均値が特定の値であるかどうかの検定

- 2群の t 検定

2つの群の平均値に差があるかどうかの検定

- ✓ 対応のある2群の場合
- ✓ 独立した2群の場合

1群の t 検定

母集団の平均値が、特定の値かどうかを検定します

Aさんのカラオケ平均点が95点かどうかで行ったのは、実は、1群の t 検定です

他の例)

工場のラインで規格どおりに製品が製造されているかどうか？

2群のt検定（対応あり）

「対応がある」とは、例えば以下のような場合です。

介入試験をおこない、試験食の摂取前後で数値を測定した

被験者No.	摂取前	摂取後
1	120	122
2	108	107
3	115	118
4	123	130
5	111	119

被験者ごとに、摂取前（A群）と摂取後（B群）で対応関係があり、知りたいのは、摂取前後で差があるかどうかです。

2群のt検定（対応あり）

実はこの問題は、次の手順で、1群のt検定として処理できます

- 摂取前後の差をとる
- その平均値が0であることを帰無仮説として検定を行う

被験者No.	摂取前	摂取後	摂取前後の差
1	120	122	-2
2	108	107	1
3	115	118	-3
4	123	130	-7
5	111	119	-8

2群の t 検定（独立2群）

実験科学の分野などでよく使われます

例)

- 介入試験で、試験食群とプラセボ群に差があるか？
- 二つのピーナッツ品種で、オレイン酸含量に差があるか？

2群間で、分散が等しいかどうかによって、二つのやり方があります。最近では、分散が等しいかどうかにかかわらず、等しくないことを仮定したウェルチの方法が良く使われます。

2群のt検定（独立2群）

等分散の場合

1群目：標本数 n_1 , 不変標本分散 s_1 , 標本平均 \bar{x}_1

2群目：標本数 n_2 , 不変標本分散 s_2 , 標本平均 \bar{x}_2

プール分散 $s^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2}$

検定統計量 $t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

自由度： $n_1 + n_2 - 2$

帰無仮説： 2群の母集団の平均値は等しい

で、同様に検定できます

参考まで

2群のt検定（独立2群）

等分散が仮定できない場合 **ウェルチの方法**

1群目：標本数 n_1 , 不変標本分散 s_1^2 , 標本平均 \bar{x}_1

2群目：標本数 n_2 , 不変標本分散 s_2^2 , 標本平均 \bar{x}_2

検定統計量
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(近似)自由度
$$v \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$$

帰無仮説：2群の母集団の平均値は等しい

で、同様に検定できます **参考まで**

メッセージ

どんな検定でも

- 検定統計量
- 自由度
- 分布の計算方法

などさえ分かれば、身につけた
ステップで、**自分でできる！**

検定で
注意すること

①

検定の間違い

前提

検定では、
正しくない帰無仮説を棄却して、
対立仮説を採択することが、
主張したいこと（正しい姿）
とします。

検定の二つの間違え

第一種の過誤 偽陽性

本当は間違っていることを、正しいと判定してしまうこと。

[検定では、本当は帰無仮説が正しいのに、間違えだとして棄却してしまうこと]

この過誤を犯す確率は α で表され、実は、その値のことを**有意水準**と呼んでいる。

α : あーわてんぼうのお手つき率

第二種の過誤 偽陰性

本当は正しいことを、誤っていると判定してしまうこと。

[検定では、本当は帰無仮説が**間違え**なのに、正しいとして棄却しないこと]

この過誤を犯す確率は β で表され、 $(1 - \beta)$ 、つまりこの過誤を犯さない確率)を**検出力**と言う。**第二種の過誤をなるべく犯さない** (β が小さい) のが、**よい検定**とされる。

β : ぼーんやりものの見逃し率

		帰無仮説が本当は	
		間違い (正しい姿)	正しい
検定 結果	棄却 する (陽性)	<div> $1 - \beta$ (検出力) </div>	第一種の過誤 偽陽性 α
	棄却 しない (陰性)	第二種の過誤 偽陰性 β	OK


 有意水準 α


第一種の過誤を起こさないように α を下げて厳しく判定すると、 β が増えてしまい、**検出力 ($1 - \beta$) が下がってしまう**。
 うまくバランスのとれた α を設定する必要がある。

少し脇道へ

スクリーニング検査

**の スクリーニング		本当は	
		病気	健康
結果	陽性 (+)	真陽性 True Positive 敏感度	偽陽性 False Positive 偽陽性度
	陰性 (-)	偽陰性 False Negative 偽陰性度	真陰性 True Negative 特異度

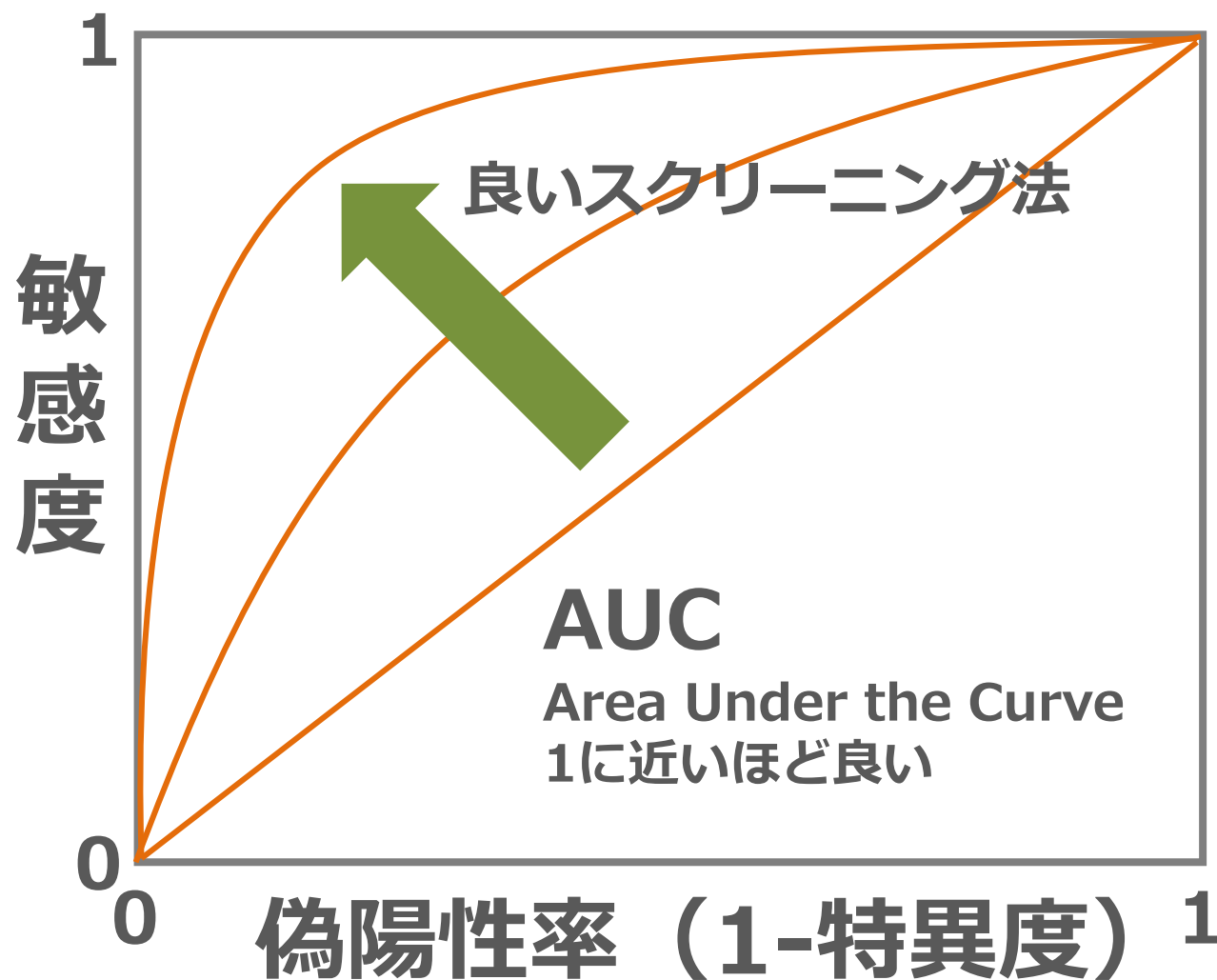
↑
カットオフ値
↓

敏感度を上げたり、偽陽性率を下げたりするためにカットオフ値を調整するのと似ています。

ただし、敏感度を上げるのに、カットオフ値を上げるか下げるかは、スクリーニング検査の方法に依存するので注意！

ROC曲線

Receiver Operating Characteristic curve



カットオフ値を変えたときの偽陽性率と敏感度をプロットしたもの

仮説検定では、何が真に正しいかがわからないため、ROC曲線が描けないことがほとんどです。

ただし、スクリーニング検査と同様に、診断システムの精度評価をする際などには多用されます。

データ解析ではとても重要な考え方です。

検定で
注意すること

②

多重性の問題

**検定は、
繰り返してはいけない**

検定を繰り返すと、誤りが大きくなる

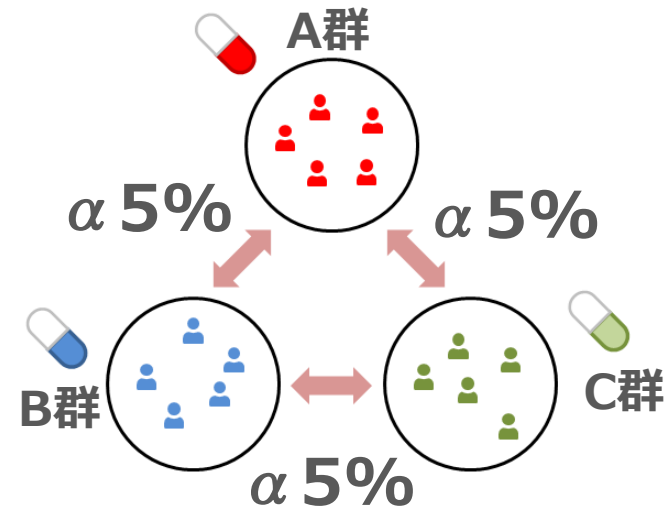
例)

3つの薬A, B, Cを与えた群で、差がなかったかどうかを、
A-B, B-C, C-A投与群間で α 5%で検定する。
3つの薬に差がないことを主張したい。

1回の検定で差がないという結果になる確率は0.95。

3回の検定でどれもが差がない結果となる確率は、0.95の3乗で、0.86。

どこかで有意な差が出てしまう確率は、 $1 - 0.86 = 0.14$ 。



数打てば当たる状況！

多重比較のための 検定法を使う

Tukey（チューキー）の
多重比較検定など

Bonferroniの補正

有意水準 α を繰り返す検定の数で割り、それを有意水準として用いる

例)

$\alpha = 0.05$ で3回検定を繰り返す場合、

$$\alpha' = 0.05 \div 3 = 0.0167$$

を代わりに用いる

全体の α （お手つき率）が決して水準を超えないように、むりやり α を引き下げるので、第二種の過誤の率（見逃し率） β が上がってしまう恐れがある。

False Discovery Rate (FDR, 偽発見率)を調整する

ある程度 α が上がるのを許容しながら、 β を小さく抑える方法。

		帰無仮説が本当は		
		間違い(正しい姿)	正しい	計
検定結果	棄却する(陽)	s	v α 偽陽性	R
	棄却しない(陰)	t β 偽陰性	u	N-R
計		N-n	n	N

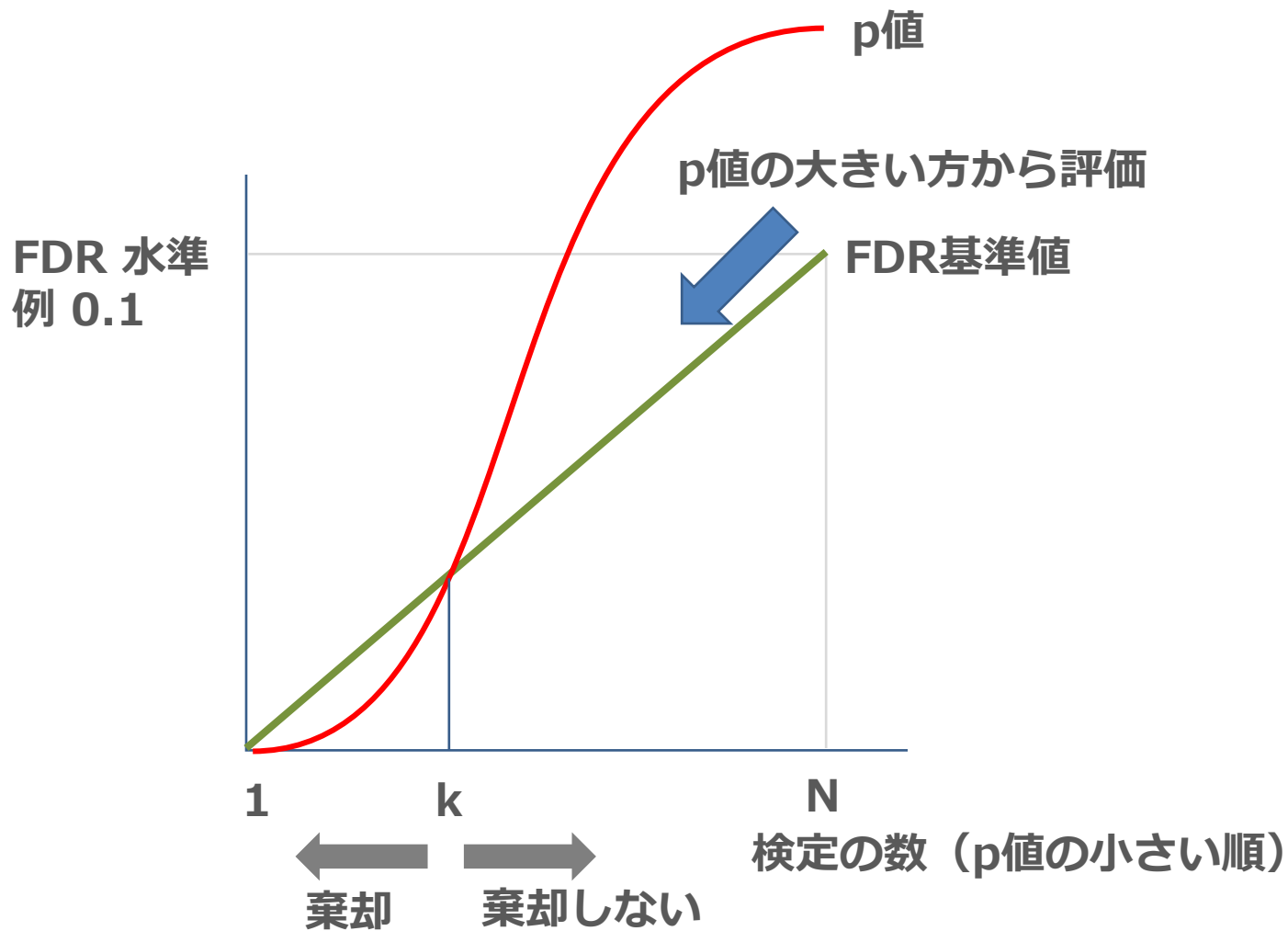
FDR $q = v/R$ 棄却したもののうち、偽陽性の率
これを、一定水準例えば0.05にする方法

FDR調整の手順

Benjamini & HochbergのFDR調整方法（BH法）（1995年に発表）
その後いろんな改良法が考案された。

- ① N個の検定結果について、p値の小さい順に並べる。
この時の順番を、 $i = 1$ 番目からN番目とする。
- ② $i = N$ （p値が一番大きいもの）とする。
- ③ $q \times i / N$ を計算する。
これが、もとのp値以上であれば、 $k = i$ として、④に進む。
もとのp値を下回れば、 $i = i - 1$ として、③を繰り返す。
 $i = 1$ に達したら、どの検定の帰無仮説も棄却しないものとする。
- ④ $i = 1$ から k までの検定の帰無仮説を棄却する

FDRのイメージ



FDR調整のイメージ

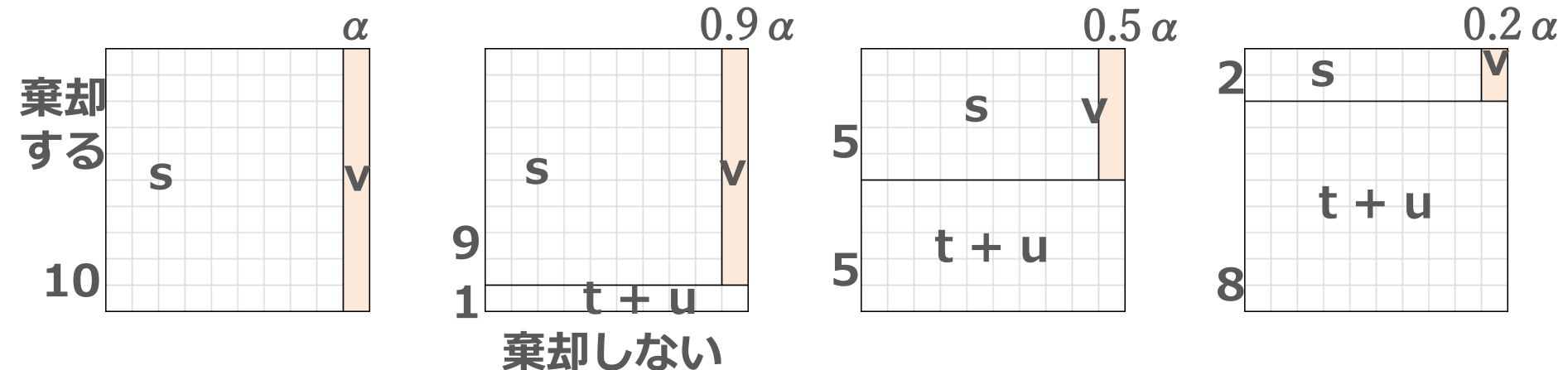
by 櫻井

p値は、検定を繰り返したときに誤る確率でもあるので、複数回検定を繰り返したときに、最大のp値が有意水準 α を下回っているなら、すべての検定が十分有意であると判断してもよいものとする（甘い）。

10回検定し、FDRを0.1に制御したいとする。

10回を全部棄却したとき、FDRを0.1以下にするには、 α は0.1でよい。
1回分を棄却しないとすると、残り9回のFDRを0.1にするには、 α は0.1 * 9/10に設定する必要がある。

以下同様、棄却する検定の数が減るほどに、 α を小さく調整する。



検定で
注意すること

③

p値 < 0.05
にとらわれるな！

有意水準 α としてよく使われる0.05
という数字に、特に深い意味はない

起こりにくい確率のひとつの基準と
して使われているだけ

アメリカ統計学会の声明

Wasserstein and Lazar (2016) The American statistician 70: 129-133 Editorial

Wasserstein et al (2019) The American Statistician 73 (S1): 1-19 Editorial

- p値が特定の値以下だったことで「統計的に有意であった」と言うてはいけない
- それよりも、p値そのものを提示する
- p値は、仮説が正しい確率を測るものではない

など

2016年の声明の日本語訳が読める

<http://www.biometrics.gr.jp/>



一般社団法人
日本計量生物学会
The Biometric Society of Japan

[HOME](#) [学会について](#) [お知らせ](#) [ニュースレター](#) [学会誌](#)

[計量生物学の未来に向けて](#) [試験統計家認定制度](#)

No.60～69

No. タイトル

- 61 [研究不正と研究環境](#) 井上永介(昭和大学)
- 60 [計量生物学徒としてHTAに貢献する](#) 萩原康博(東京大学大学院医学系研究科)

No.50～59

No. タイトル

- 59 [真実がわからない中で過去からの学びをどう活かすか](#) 坂巻頼太郎(横浜市立大学)
- 58 [計量生物学を理解したいと思って毎日挑戦しています.](#) 長島健悟(統計数理研究所)
- 57 [これからの計量生物学の発展を担う生物統計家の育成](#) 安藤宗司(東京理科大学)
- 56 [一教員として貢献できること](#) 高橋佳苗(大阪市立大学)
- 55 [ベースラインハザードから思うこと](#) 横田 勲(北海道大学)
- 54 [放射線疫学と日本人のコホートを追跡する日米共同研究機関](#) 三角宗近(放射線影響研究所)
- 53 [実務の現場から:食品・栄養研究にも活用される生物統計学の専門性](#) 高田理浩(味の素株式会社)
- 52 [異分野,異文化の接点から](#) 島津秀康(英国ラフバラ大学)
- 51 [統計学を学んで](#) 奥井 佑(九州大学)
- 50 [教育・指導への感謝と未来への還元](#) 井桁正亮(兵庫医科大学)

[> トップページ](#)

[> 学会について](#)

[> お知らせ](#)

[> ニュースレター](#)

[> 学会誌](#)

[> 計量生物学の未来に向けて](#)

[> 試験統計家認定制度](#)

[> 臨床研究に関する日本計量生物学会声明](#)

[> 統計家の行動基準](#)

[> 統計家の行動基準\(英語版\)](#)

[> 統計的有意性とP値に関するASA声明](#)

[> メーリングリスト](#)

[> 当会へのお問合せ](#)

やってはいけない不正行為

- t 検定で有意にならなかったのに、有意になる検定方法を試して、マン・ホイットニーのU検定を採用した
- サンプルサイズを調整した



p値ハッキング

情報統計 第8回

2022年8月3日 神奈川工科大学



櫻井 望

国立遺伝学研究所
生命情報・DDBJセンター

そのほかの検定

明日の準備

自習 (課題準備)

学習目標

ほかの検定方法を把握する

- ✓ パラメトリック検定と
ノンパラメトリック検定の違い
- ✓ χ^2 （カイ）二乗検定

パラメトリック検定

- 分布を用いる
- 正規分布に従うとか、等分散性があるとか、何かしらの前提条件が必要

ノンパラメトリック検定

- 分布を用いない
- 前提条件がない
- データを並び替えて検定する

例えば2群の差の検定

パラメトリック検定

対応ない場合

2群の t 検定

対応ある場合

対応ある1群の t 検定

ノンパラメトリック検定

対応ない場合

マンホイットニーのU検定

対応ある場合

ウィルコクソンの符号付き
順位和検定

分割表による検定

- カイ二乗検定
- フィッシャーの正確確率検定

	ゲームが好き	ゲームそれほどでも	合計
朝食を食べる			
朝食を食べない			
合計			

など