2023年8月25-27日
遺伝統計学・夏の学校　講義実習資料

統計学入門

大阪大学大学院医学系研究科 遺伝統計学
東京大学大学院医学系研究科 遺伝情報学
理化学研究所生命医科学研究センター　システム遺伝学チーム

http://www.sg.med.osaka-u.ac.jp/index.html

講義の概要

統計学入門

① 統計学について

② 母集団と標本集団

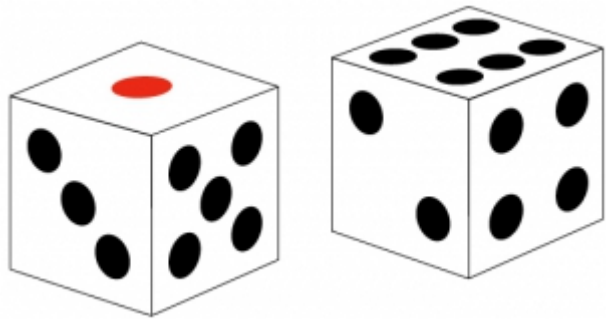
③ 帰無仮説とP値

④ 統計検定手法

① 統計学について

- 世の中には、「**統計は難しい**」という苦手意識が蔓延しているようです。
- 確かに、統計学を構成する数学的理論を理解するのは難しいです。
- でも、ツールとしての統計学を適切に使いこなしたり、統計解析の結果を適切に解釈することは、そこまで難しくなく、むしろ**必須のスキル**です。

① 統計学について



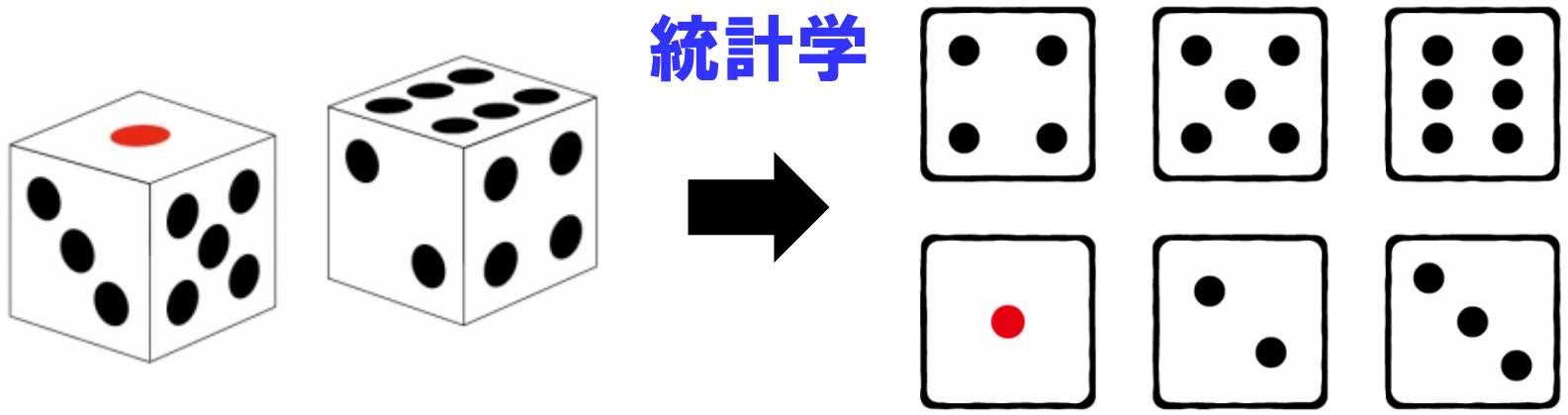
現象：サイコロ
調査：サイコロを振る
数量：さいころを振って出た目

統計学

サイコロを振って出た目を検討して、サイコロについて知る。

- 調査によって得られた数量で現象を把握することを統計といいます。
  （調査によって得られた数量そのものを、統計ということもあります。）
- 統計に関する学問や検討方法が、統計学です。
- 「サイコロ」という現象を把握するために、「サイコロを振る」という調査を行い、「サイコロを振って出た目」という数量を調べる、のは統計学です。

① 統計学について

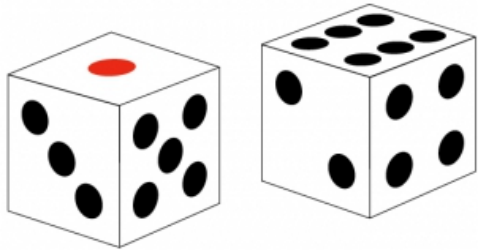


・サイコロを振ると、１から6までの目が同じ**確率**で出ます。

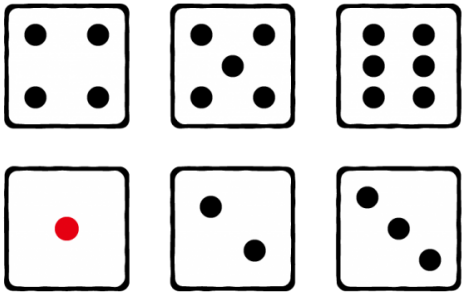
・同じ確率であることは、サイコロの物理的な構造に基づく解釈というより、**これまでの観測結果に基づく経験**から得られています。

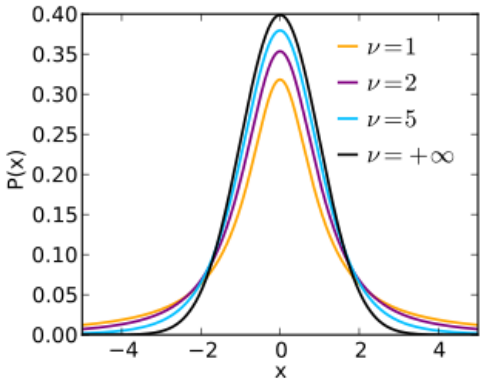
・つまり、**統計学による成果**ということができます。

① 統計学について



統計学





- 世の中には沢山の現象があり、サイコロのように簡単にはわかりません。
- 「この現象の統計はどのような分布をとるのか？」という問いに答えるため、統計学が発展してきました。
- 例えば、有名なt検定で用いるt分布は、ビールの品質管理の過程で発見されました。

① 統計学について

統計学

- 数理統計学
- ベイズ統計学
- 統計力学
- 恒星統計学
- 経済統計学
- 保険統計学
- 極値統計学
- テキストマイニング
- 統計的因果推論
- 生物統計学
- 医療統計学
- **遺伝統計学**

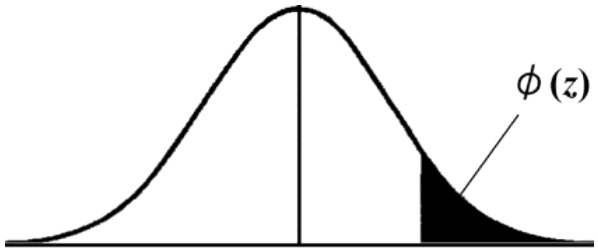
・データ解析を行う統計学には、様々な学問分野が含まれています。

・ゲノムデータの解析に特化した統計学として、遺伝統計学があります。

① 統計学について

数理統計学

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



ベイズ統計学

ベイズ論者

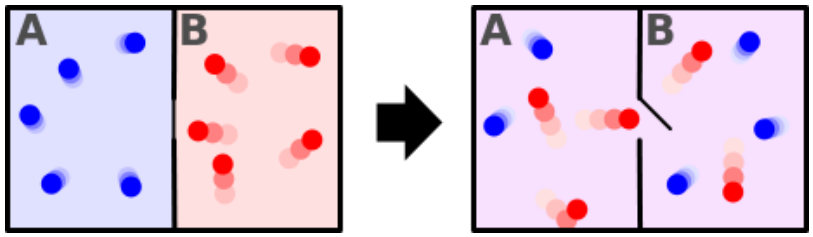
P値

頻度論者

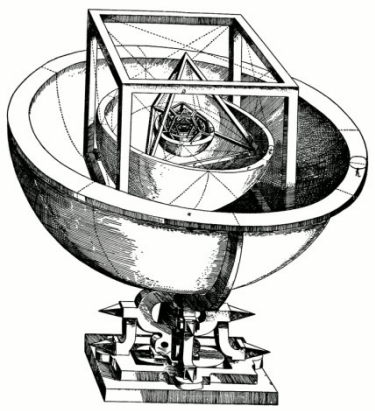
- **数理統計学**は、数学的理論に基づき統計を記述する学問分野です。
- 統計学の根幹となる分野で、多数の重要な数式で記述されます。
- **ベイズ統計学**は、事前分布を仮定するベイズ理論に基づく統計学です。
- ベイス統計学では、いわゆる”P値”を(あまり)計算しません。
- ベイズ論者と、相対する頻度論者の間には、長きに渡る論争があります。

① 統計学について

統計力学



恒星統計学



- **統計力学**は、微視的な物理法則に基づき巨視的な現象を説明する学問です。**統計物理学**や**統計熱力学**とも呼ばれます。
- エントロピーや熱平衡の話も、統計学と関連があったわけです。
- **恒星統計学**は、観測された天体の位置・運動・法則に基づき、恒星を研究する学問です。
- **ケプラーの法則**や**万有引力**の発見をもたらしたのは、ティコ・ブラーエによる膨大な天体観測記録でした。

① 統計学について

経済統計学

保険統計学

- **経済統計学**は、経済活動の指標を理論的に説明・予測する学問です。
- 経済活動の複雑化・IT化・グローバル化に伴い進歩を続けています。経済統計学の専門家には、**ノーベル賞**（経済学賞）の受賞者もいます。
- **保険統計学**は、保険が商品として成立するために、保険の対象となるイベントの発生リスクを適切に推定するための学問です。**保険料は統計学により決められている**、という面があったわけです。

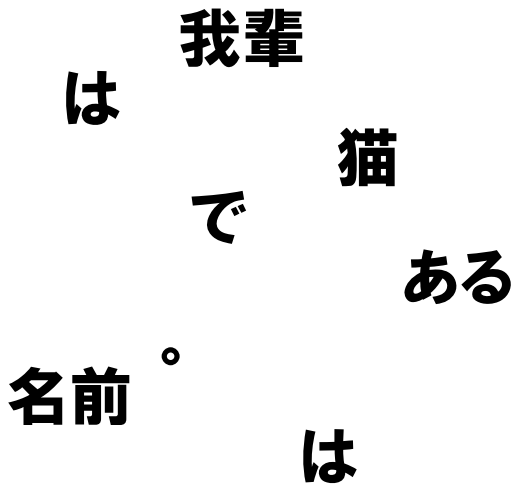
① 統計学について

極値統計学

Event	Men		
	Endpoint	Standard error	World record
Running			
100 m	9.29	.39	9.74
110/100-m hurdles	12.38	.35	12.88
200 m	18.63	.88	19.32
400 m	—	—	43.18
800 m	1:39.65	1.44	1:41.11
1,500 m	3:22.63	3.31	3:26.00
10,000 m	—	—	26:17.53
Marathon	2:04:06	57	2:04:26

(Einmahl JHJ et al. ***J Am Stat Assoc*** 2008)

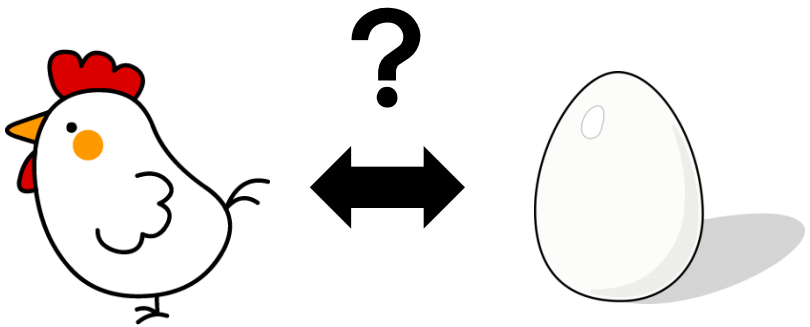
テキストマイニング



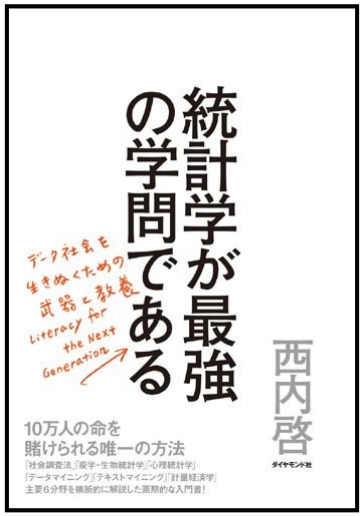
- **極値統計学**は、最大値や最小値が漸近的に従う分布を研究します。
- **陸上競技の世界記録**を検討した結果、男子100m走の記録は、9.29秒程度まで短縮出来る余地があると推定されています（2008年時）。
- **テキストマイニング**は、自然言語の文章を対象としたデータ解析です。
- 文章を単語や文節で区切り、出現頻度や相関関係、傾向、時系列の解析や、**文章の作者の推定**などを実施します。**SNS**も対象になります。

① 統計学について

統計的因果推論



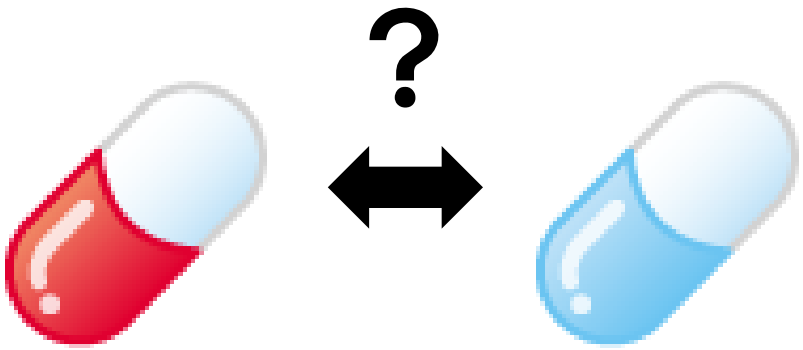
生物統計学



- **統計的因果推論**は、複数の現象の間の**因果関係**を明らかにして、見かけ上の類似(相関)との区別を行う、理論的な学問です。
- 観測された現象から因果関係を推定するのは難しいですが、重要です。
- **生物統計学**は、生物学領域で観測される現象に対する統計解析や、それを通じた**生命現象の理解**を目的とする学問です。
- 生命科学研究の計画デザインや結果の解釈において重要です。

① 統計学について

医療統計学

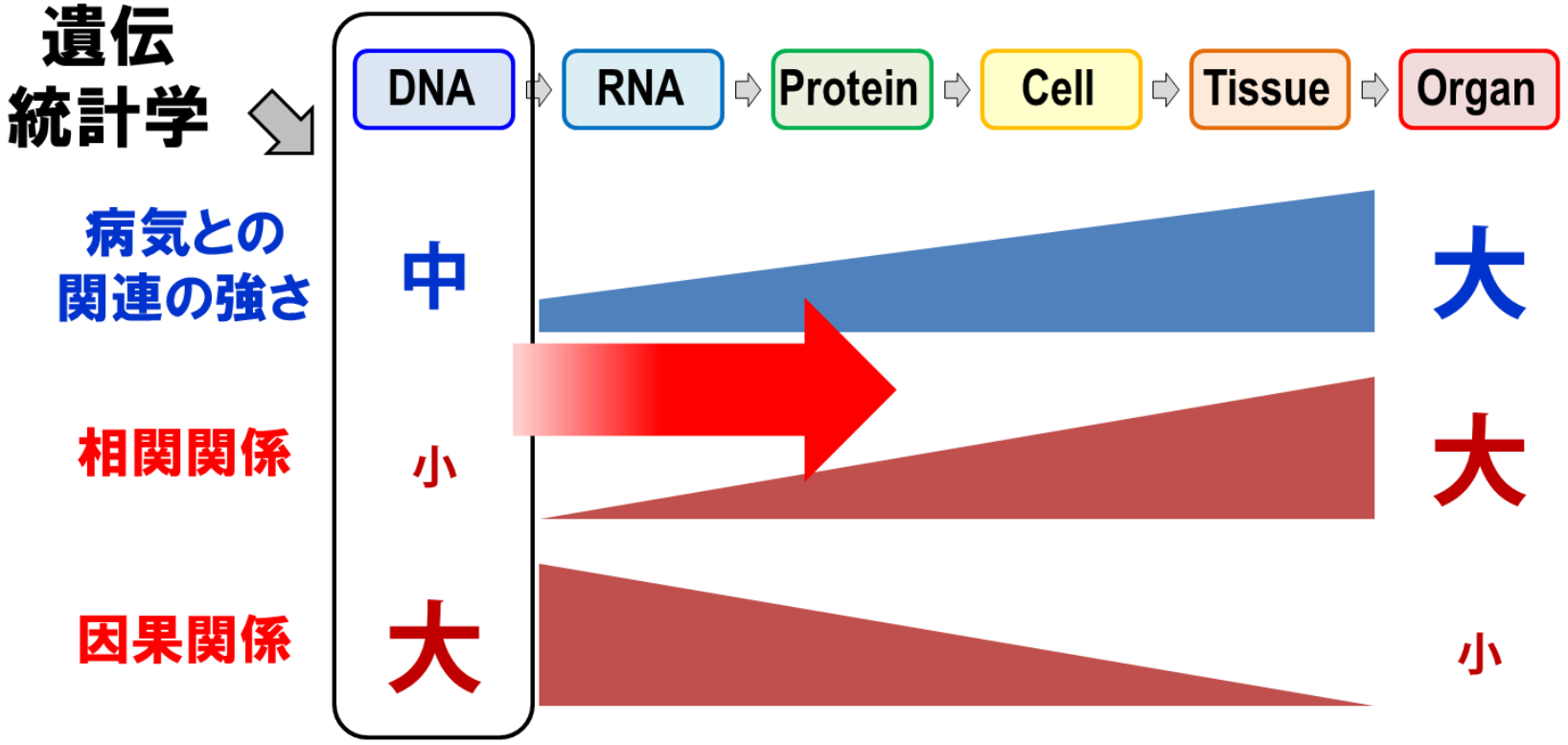


遺伝統計学

- **医療統計学**は、医学研究における統計解析を専門とした学問です。
- EBM（Evidence Based Medicine）や**臨床研究**を支えている学問です。
- **遺伝統計学**は、**遺伝情報**と**形質情報**の関わりを評価する学問です。
- 遺伝情報と形質情報の間に、**因果関係が担保されている**のが特徴です。
- ゲノムビッグデータ時代に突入した現在、重要性が増しています。

① 統計学について

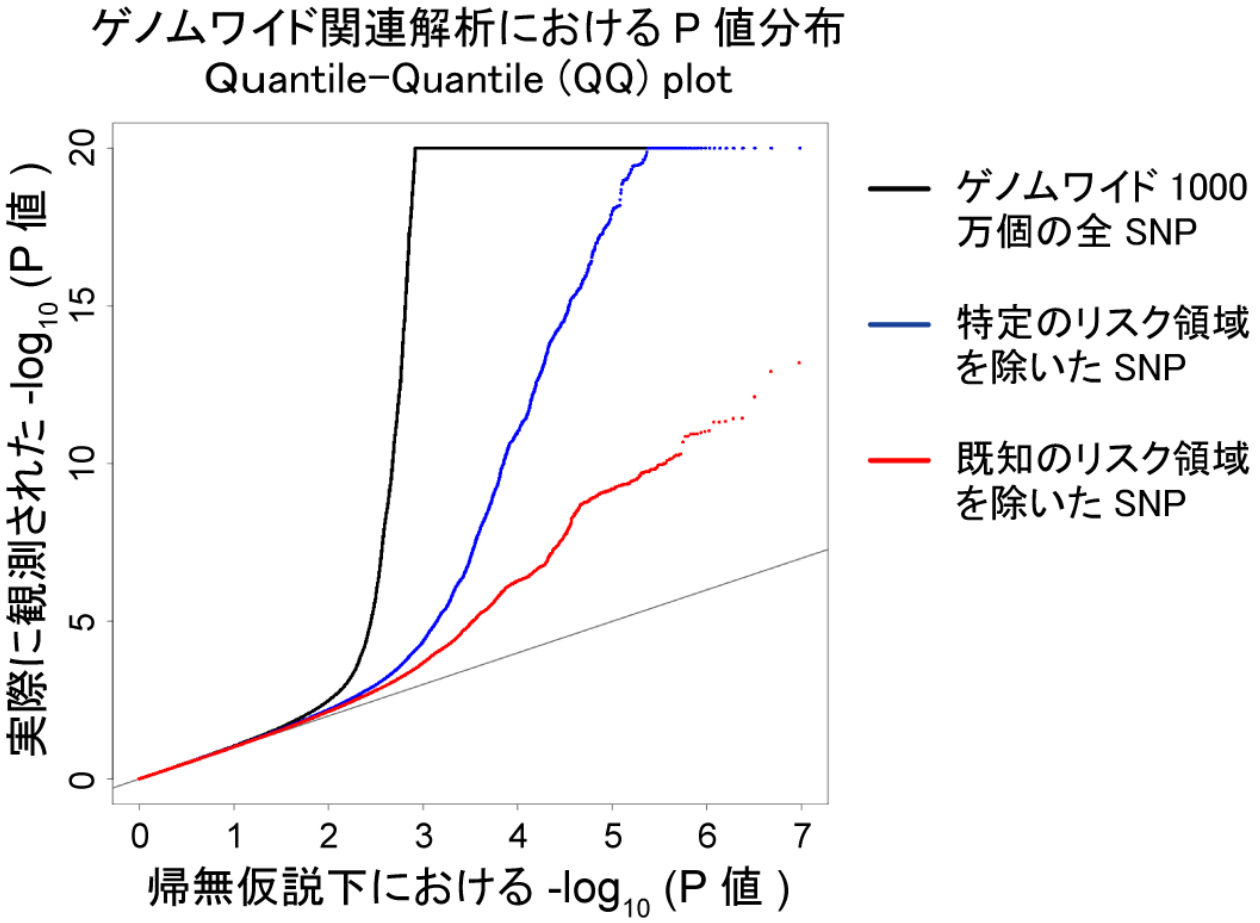
遺伝統計学の特徴：その①



- 統計学で最も難しいこと(の一つ)は、因果関係の証明です。
- 遺伝統計学では、遺伝情報 → 形質情報の因果関係が(比較的)担保されています。そのため結果の応用性・再現性が高いことが特徴です。
- ヒトゲノム情報の充実に伴い、医学における重要性が高まっています。

① 統計学について

遺伝統計学の特徴：その②

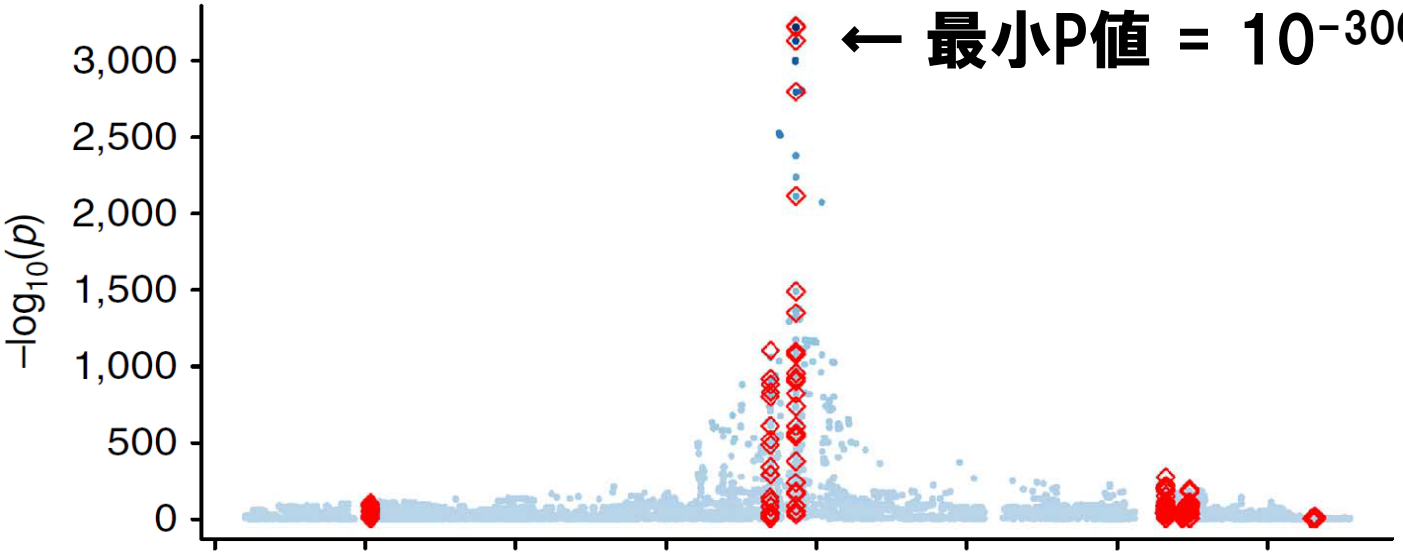


- 遺伝統計学では、数千万～数億もの**膨大な数の検定**を実施します。
- 得られたP値がどんな分布をとるのか？という検討が重要になります。
- 多重検定の問題も重要になります。

① 統計学について

遺伝統計学の特徴：その③

MHC領域における強直性脊椎炎リスク解析結果



通常の関数数値計算で求められる最小P値 = 10^{-15}
通常のプログラミングで扱える最小数値 = 10^{-300}

- 遺伝統計学では、計算限界以下のとてつもなく小さいP値を計算します。
- 有意水準も、$\alpha = 0.05$よりずっと小さくなります。
- ゲノムワイド関連解析における有意水準は、$\alpha = 5.0 \times 10^{-8}$です。

① 統計学について

遺伝統計学の特徴:その④

人が読める
データ

人が読めない
データ

- ヒトゲノム配列解読技術の進歩に伴い、遺伝統計学の分野では、とてつもなく大きなデータを解析対象とする必要性に迫られています。
- 人が読みとれないファイル形式にデータを圧縮保存して解析を行います。

講義の概要

統計学入門

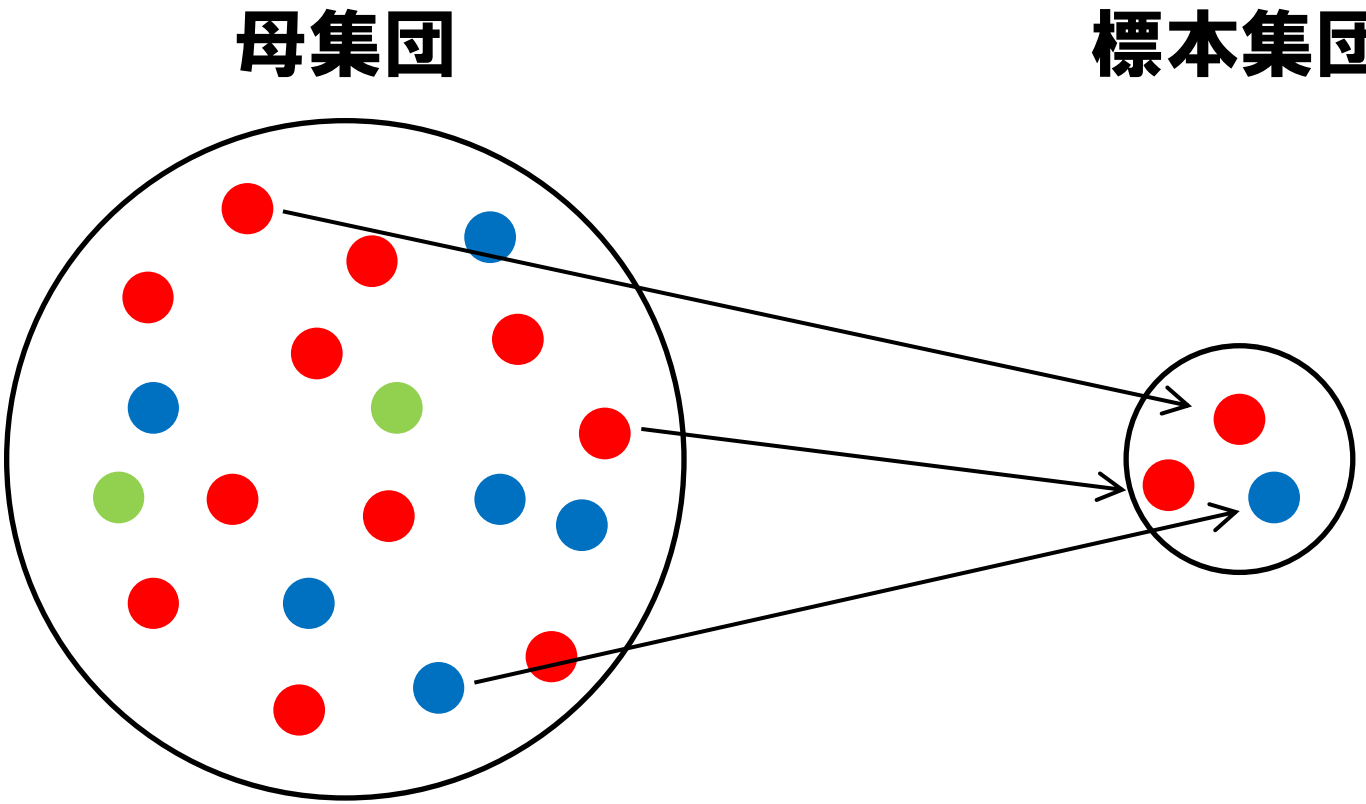
① 統計学について

② 母集団と標本集団

③ 帰無仮説とP値

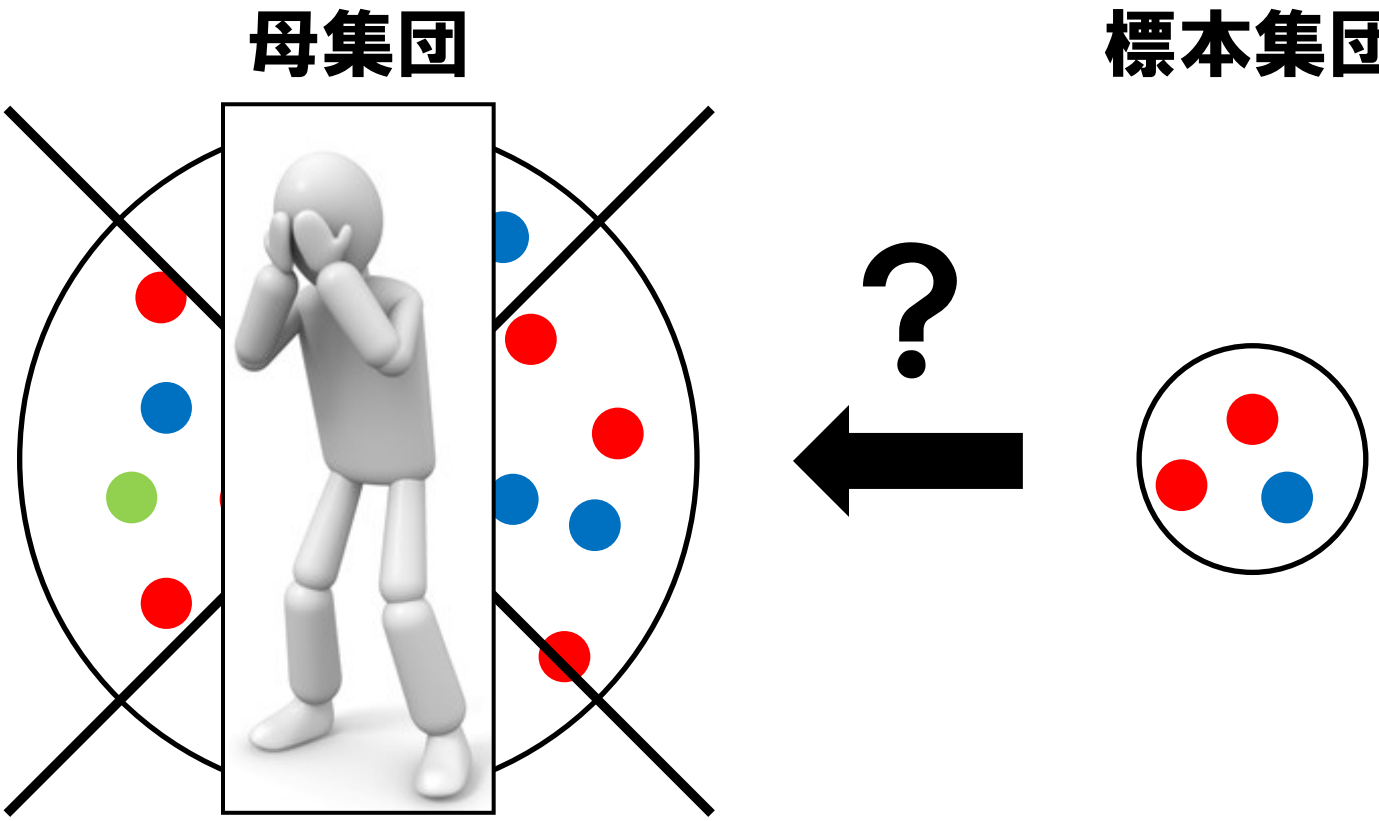
④ 統計検定手法

② 母集団と標本集団



- **母集団**とは、現象を構成するデータ全体の集まりのことです。
- **標本集団**とは、母集団から取り出されて観測されたデータの集まりです。
- **標本集団は母集団の部分集合**、と捉えることもできます。

② 母集団と標本集団

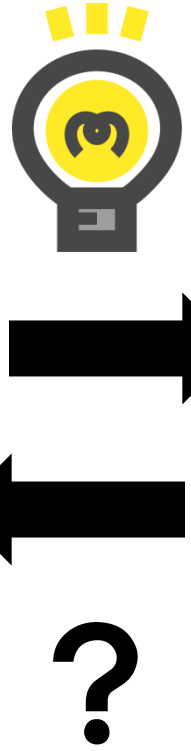


- 母集団の全容を観測することは(現実的に)不可能な場合が多いです。
- つまり、**我々が観測できるのは標本集団だけ**です。
- 替わりに、**標本集団の情報を手がかりに母集団を推定**します。
- でも、どうやって部分的な情報から全体を推定すればいいのでしょう？

② 母集団と標本集団

標本集団

1, 2, 2, 3, 3, 3,
4, 4, 4, 4

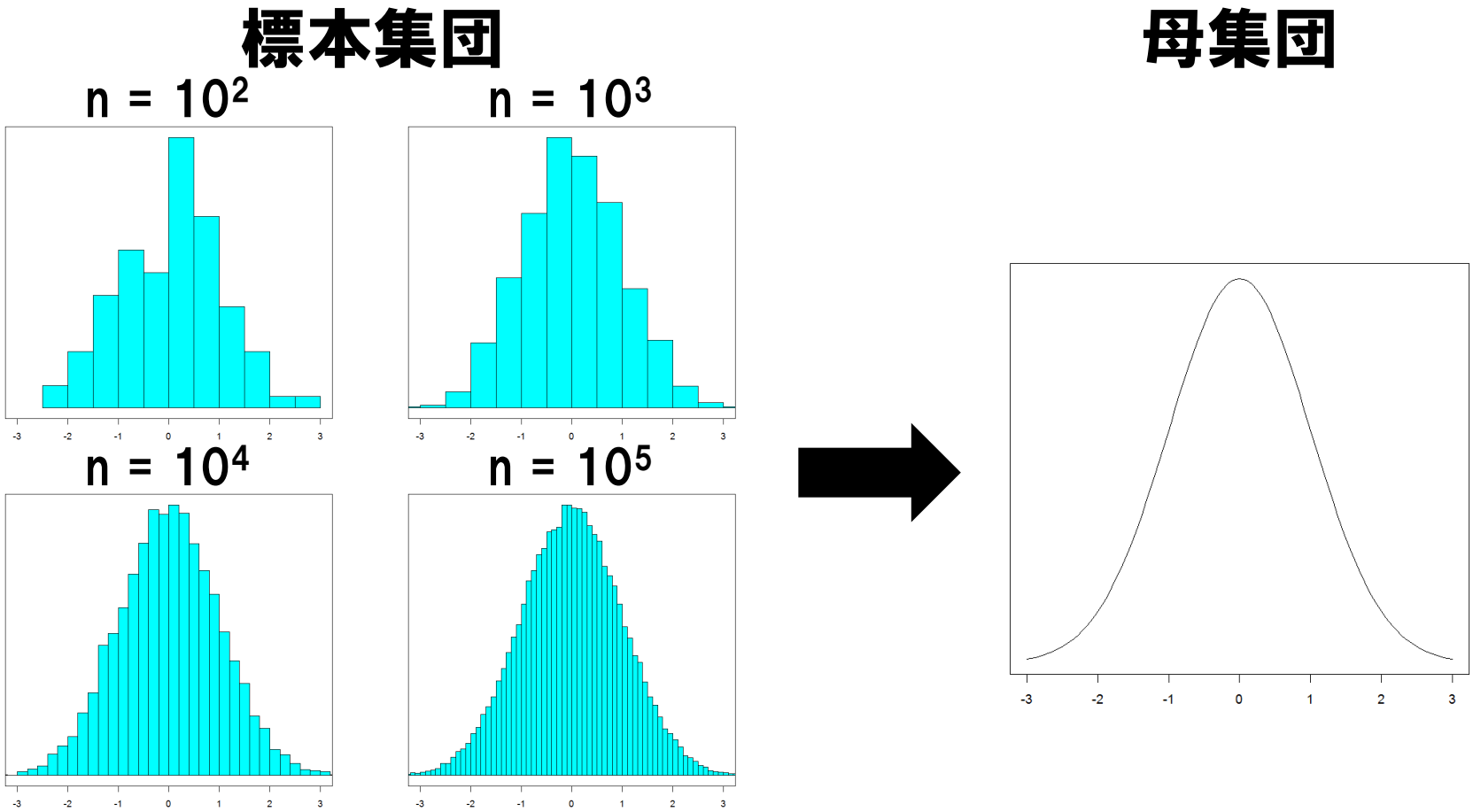


母集団

平均値：3
標準偏差：1.054
分散：1.111
中央値：3
最頻値：4

・第一の方法は、**母集団の特徴を表す数値**を調べることです。
・**平均値**、**標準偏差**、**分散**、**中央値**、**最頻値**、等の代表値が対象です。
・しかし、代表値だけでは母集団の全容はわからないことが多いです。

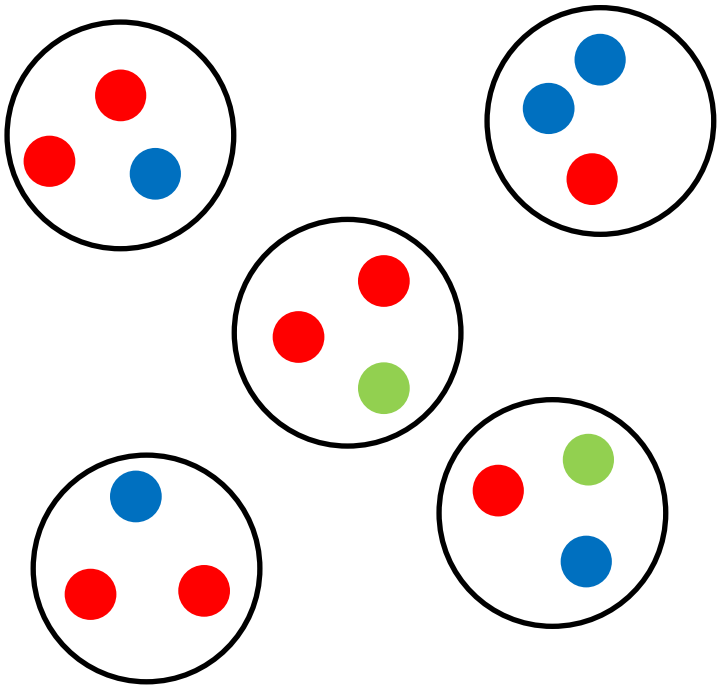
② 母集団と標本集団



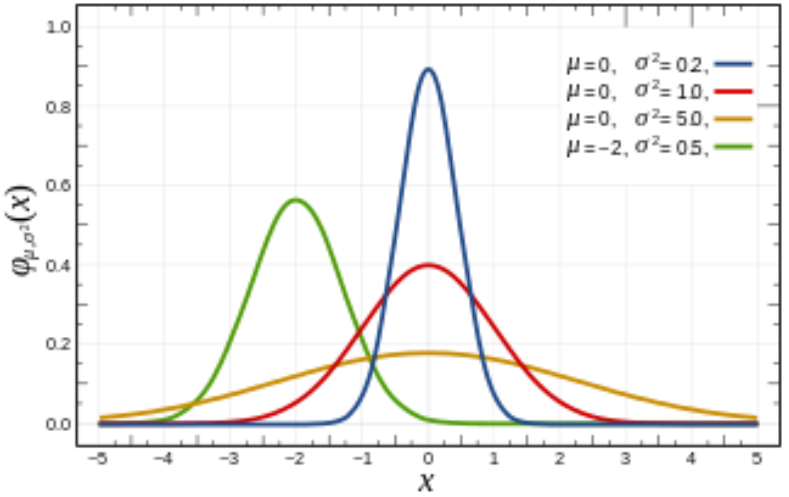
- 第二の方法は、**標本数を増やす**ことです。
- 標本数が増加すると、標本集団の分布は母集団に近づいていきます。
- 例えば、標本数が増えるにつれ、標本集団の平均値は、母集団の平均値に近づいていくことが知られています（**大数の法則**）。

② 母集団と標本集団

標本集団



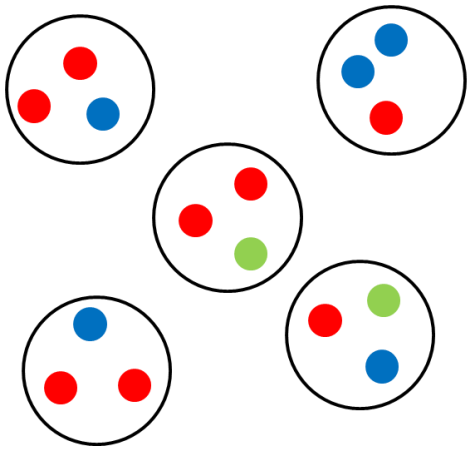
母集団



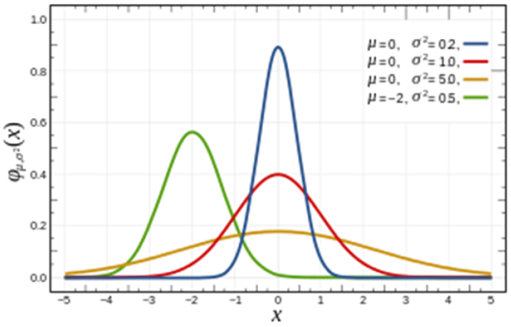
- 第三の方法は、**理論的に母集団を推定**することです。
- 特定条件下でのサンプリングにより得られた標本集団については、**数理的に母集団の分布(=数式)を定義**することができます。

② 母集団と標本集団

標本集団



母集団



標本集団から母集団を知るためのアプローチ

①：代表値を計算する。

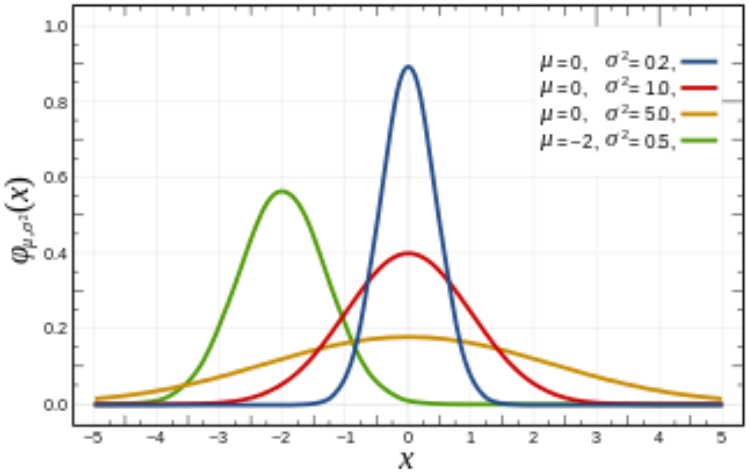
②：標本数を増やす。

③：理論的な分布を理解する。

・代表的な理論分布について、幾つか眺めてみましょう。

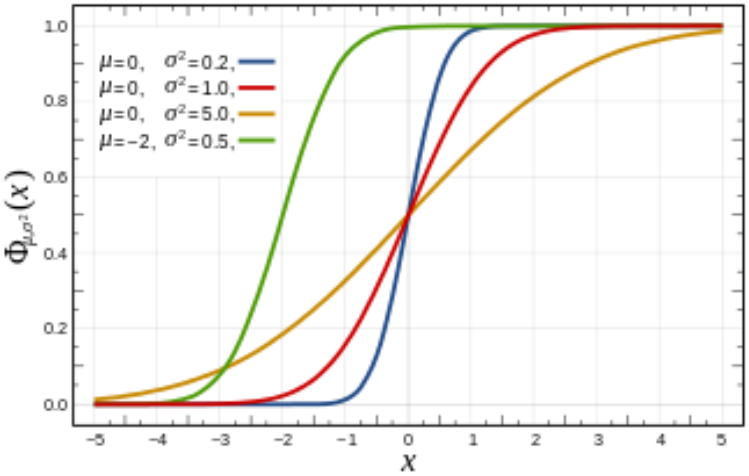
② 母集団と標本集団

確率密度関数



積分

累積分布関数

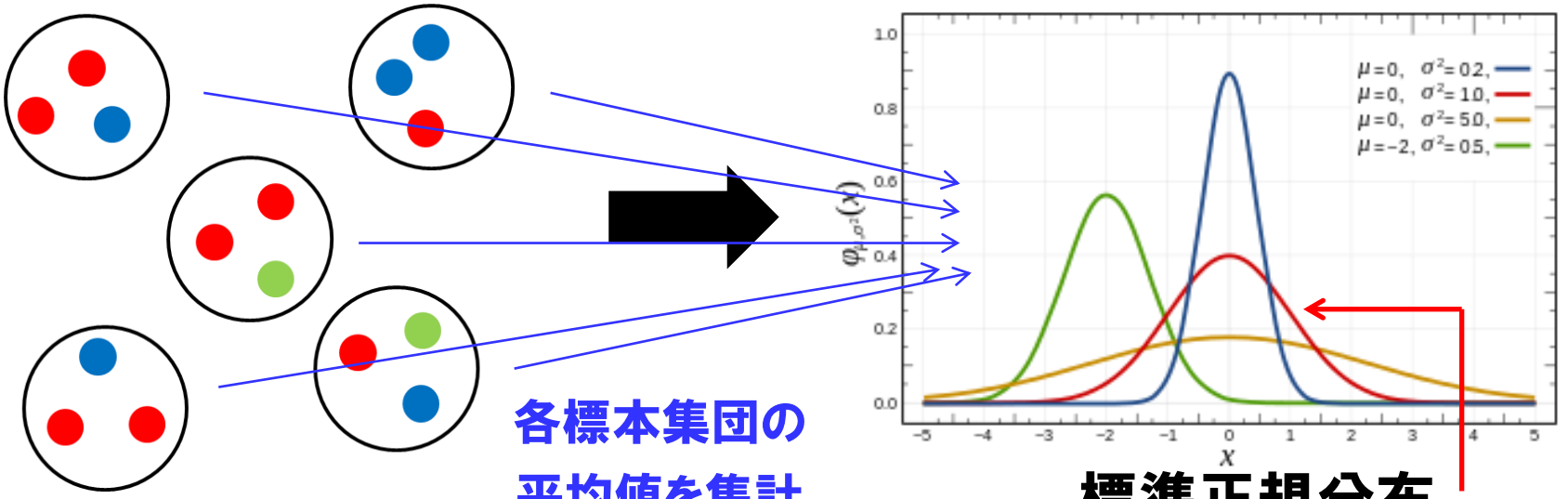


- 分布を表す代表的な数式には、**確率密度関数**と、その積分結果である**累積分布関数**の２種類があります。
- 確率密度関数は、数値で表された現象が生じる確率で、全ての現象が生じる確率を足すと1になります。
- 結果として、累積分布関数は最小値0から最大値1へと、値が一方方向に増えていく形をとります。

② 母集団と標本集団

標本集団

母集団



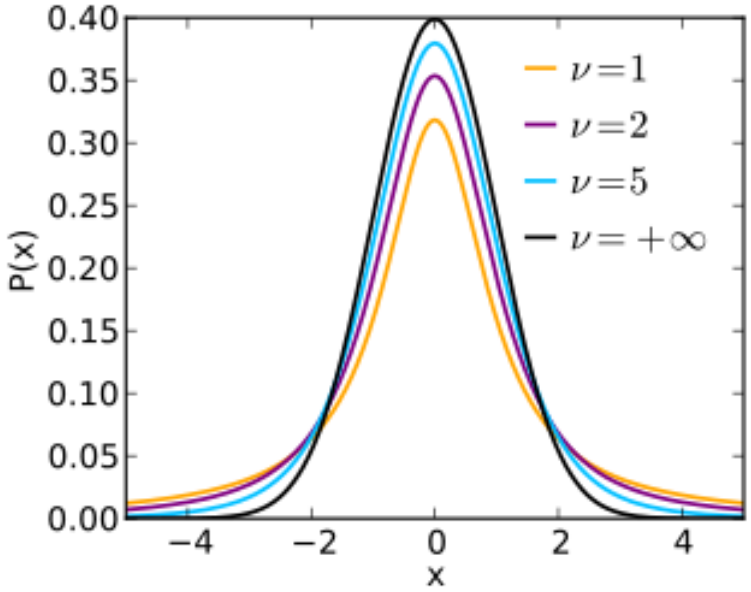
各標本集団の
平均値を集計

標準正規分布
（平均=0、標準偏差=1）

- 一番有名な分布は、**正規分布**です。
- 統計学の根幹を成す、重要な分布です。
- 特に、平均値=0、標準偏差=1、の時に、**標準正規分布**と呼びます。
- 標本数が増えるにつれ、標本集団の平均値と母集団の平均値の差は、正規分布に近づいていくことが知られています（**中心極限定理**）。

② 母集団と標本集団

t分布



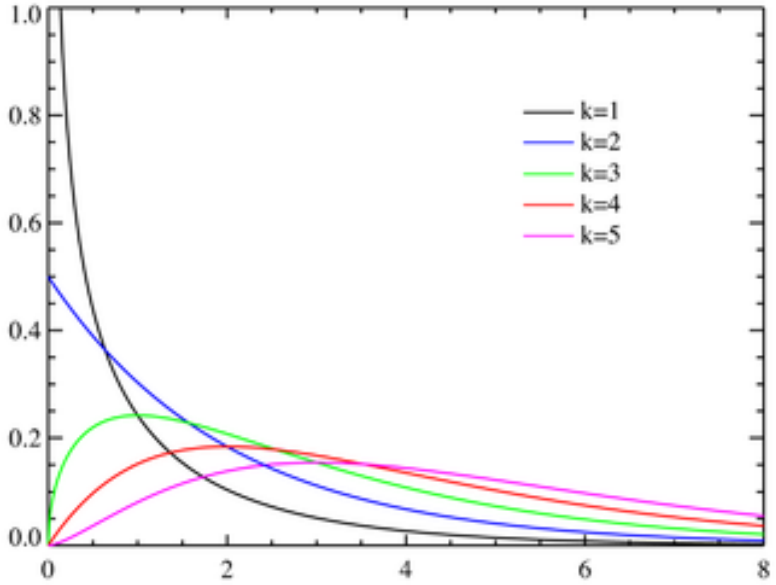
$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

- t分布は、正規分布に従う母集団から得られた標本数が少ないときに、用いられる分布です。
- t検定に際して用いられます。
- t分布と正規分布は近い関係にあり、標本数が十分に多い時、t分布は正規分布に近づく事が知られています。

② 母集団と標本集団

カイ二乗分布

正規分布×正規分布
↓
自由度1のカイ二乗分布



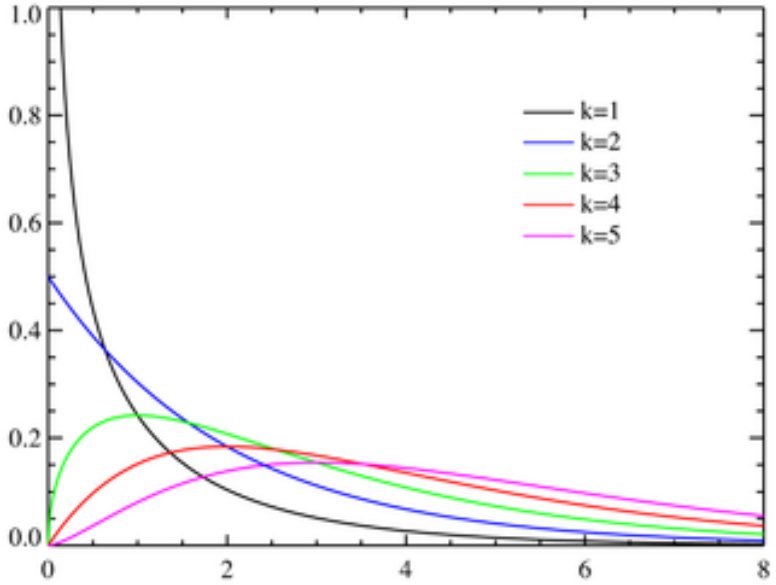
$$\frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}$$

- カイ二乗分布は、ある現象において観測された数値と、理論上期待される数値の差(の二乗の和)が従う分布です。
- カイ二乗検定に際して用いられます。
- カイ二乗分布と正規分布は近い関係にあり、正規分布を二乗するとカイ二乗分布になります。

② 母集団と標本集団

カイ二乗分布

自由度1のカイ二乗分布
＋
自由度1のカイ二乗分布
↓
自由度2のカイ二乗分布



$$\frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}$$

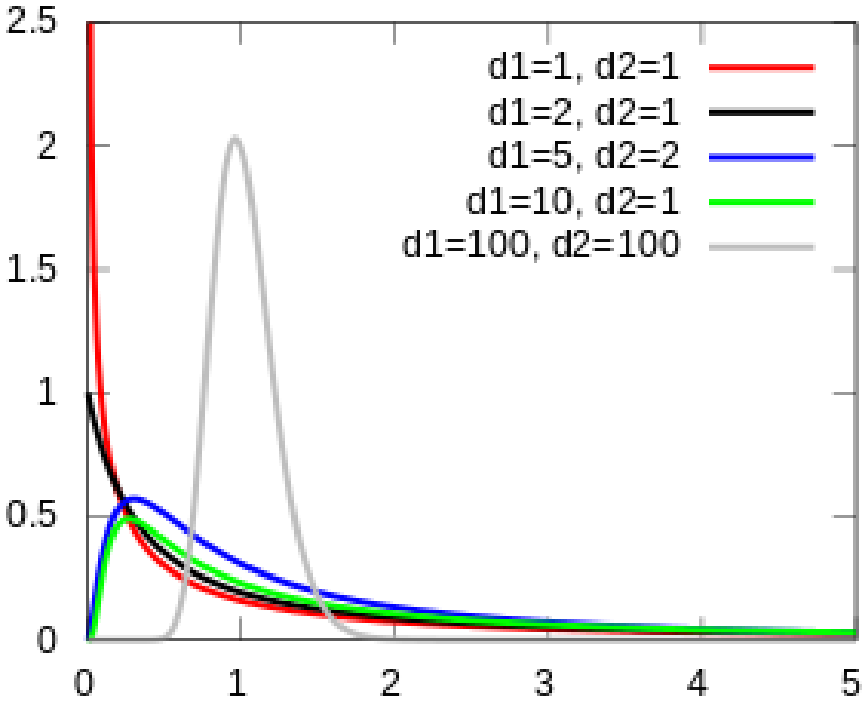
自由度

- カイ二乗分布には、自由度（degree of freedom）という概念があります。
- 観測値を自動的に決定するのに必要なパラメーター数が、自由度です。
- カイ二乗分布の形は、自由度に応じて変化してきます。
- 自由度1のカイ二乗分布と、独立な自由度1のカイ二乗分布をあわせると、自由度2のカイ二乗分布になります。

② 母集団と標本集団

F分布

カイ二乗分布/カイ二乗分布
↓
F分布

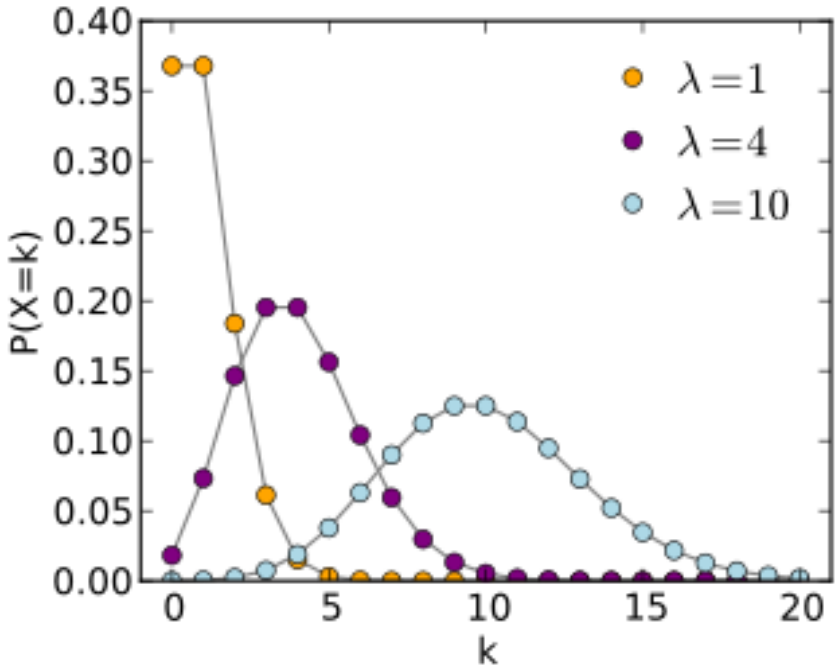


$$\frac{\sqrt{\frac{(d_1\,x)^{d_1}\;\;d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$$

- **F分布**は、独立な2つのカイ二乗分布の比で表される分布です。
- **等分散の検定**や**分散分析**に際して用いられます。

② 母集団と標本集団

ポアソン分布



$$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$$

- **ポアソン分布**は、発生確率の低い事象が、単位時間あたりに生じる数の分布です。
- **離散的**な事象を扱うため、確率密度関数はギザギザになります。
- ポアソン分布の初めての適用例は、**行軍中に馬に蹴られて死亡する兵士の数**でした。

② 母集団と標本集団

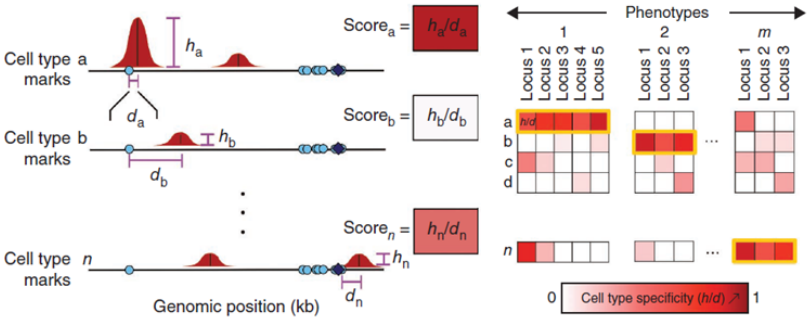
離散一様分布、二項分布、ポアソン分布、負の二項分布、ベルヌーイ分布、超幾何分布、多項分布、ゼータ分布、ジップ分布、連続一様分布、正規分布、標準正規分布、対数正規分布、多変量正規分布、指数分布、t分布、カイ2乗分布、ガンマ分布、ベータ分布、ディリクレ分布、F分布、コーシー分布、アーラン分布、三角分布、ラプラス分布、レイリー分布、ロジスティック分布、ミゼスフィッシャー分布、ミンコフスキー分布、レヴィ分布、ガンベル分布、パレート分布、ワイブル分布

- これまでの研究により、沢山の理論的な分布が知られています。
- 同時に、理論的に解けない分布が存在することも知られています。
- 色々と調べてみると、面白いかもしれません。

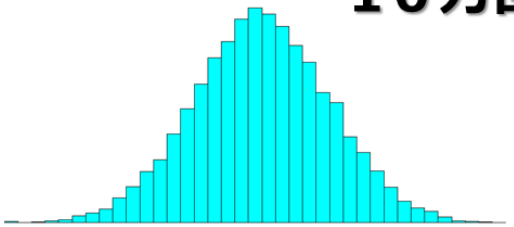
② 母集団と標本集団

遺伝統計学の特徴：その⑤

遺伝統計学



経験的に
１０万回計算

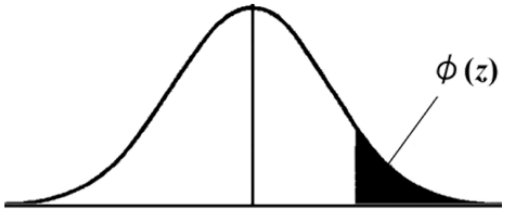


数理統計学

正規分布の数式

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

理論的に
積分計算



- 一方で、遺伝統計学では、検定統計量が従う分布を理論的に定義できないことが多いです。
- スパコン等を使って、シミュレーションで標本数を増やすことにより、経験的に分布を求める例が多いです。

講義の概要

統計学入門

① 統計学について

② 母集団と標本集団

③ 帰無仮説とP値

④ 統計検定手法

③ 帰無仮説とP値

命題：知りたいこと

帰無仮説：否定したい仮説

対立仮説：否定したい仮説とは反対の仮説

↓

帰無仮説が**棄却**される場合、対立仮説が成立する

- **仮説検定**（＝統計的仮説検定、統計検定）とは、数値や分布に関する**命題**をについて、確からしいか検証することです。
- 否定したい仮説＝**帰無仮説**（きむかせつ、Null Hypothesis、H0）を立てて、それが**棄却**されるとき、対する**対立仮説**（Alternative Hypothesis、H1）が成立する、という論理に基づきます。

③ 帰無仮説とP値

命題：薬Aは疾患Bに効果があるのか？

帰無仮説：薬Aは疾患Bに効果がない

対立仮説：薬Aは疾患Bに効果がない、わけではない

↓

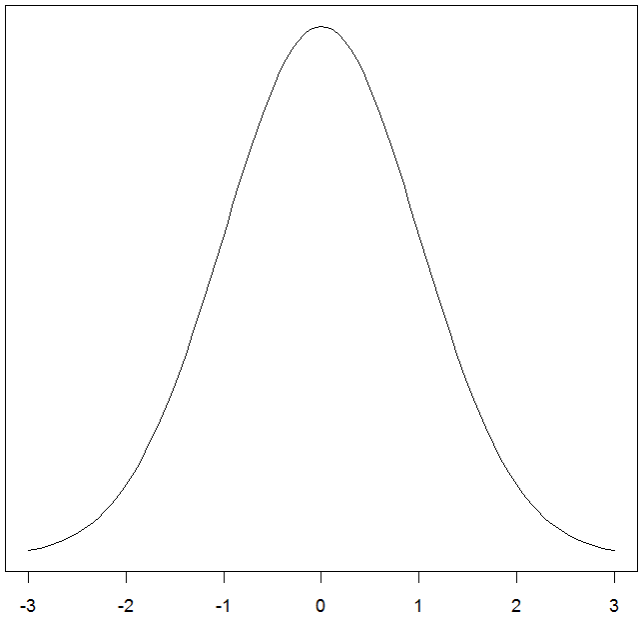
帰無仮説が**棄却**される場合、対立仮説が成立し、即ち「薬Aは疾患Bに効果がある」ということになります。

- **帰無仮説**には、効果が無い、差が無い、因果関係が無い、など、**主張したいことの反対の仮説**が相当します。

③ 帰無仮説とP値

統計量：薬A服用時の、疾患Bの改善値

帰無仮説下での統計量分布

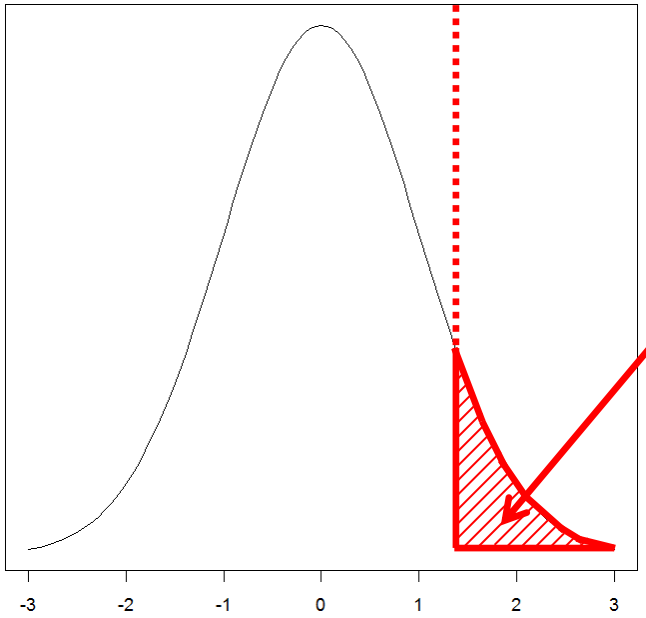


（むしろ悪化）→（大きく改善）

- 帰無仮説が成立すると仮定した場合、著しく稀な観測値が得られたとしたら、帰無仮説は正しくない（＝棄却すべき）と考えられます。
- そのため、**帰無仮説下で観測値（＝統計量）が従う分布**を検討します。

③ 帰無仮説とP値

今回観測した結果



今回観測した結果
より稀な事象
↓
確率の総和＝P値

（むしろ悪化）→（大きく改善）

- 帰無仮説下で統計量が従う分布において、今回観測した結果がどれだけ稀であったかを定量します。
- 分布上で、今回よりもっと稀な事象の発生確率の和を求めます。
- この値が、P値になります。

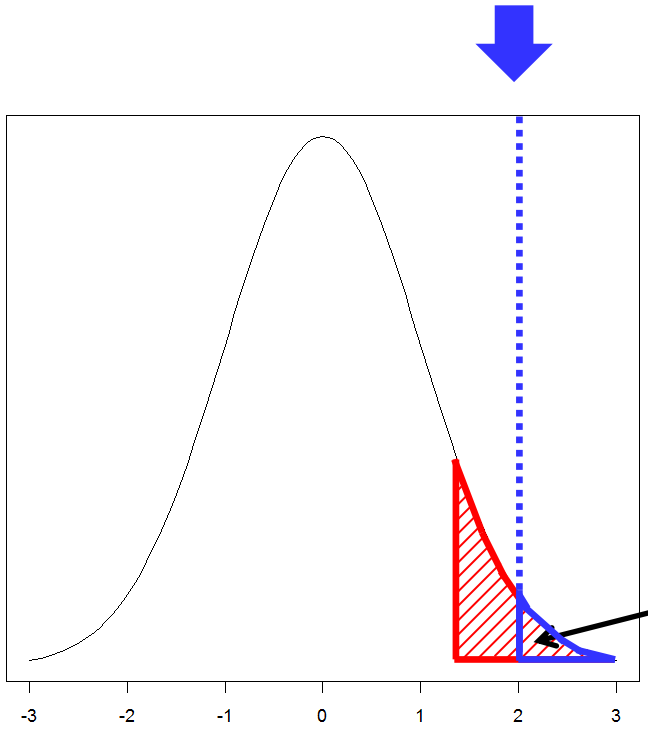
③ 帰無仮説とP値

帰無仮説下でも2回に1回は観察される程度の稀さ：P = 0.5
帰無仮説下でも5回に1回は観察される程度の稀さ：P = 0.2
帰無仮説下でも10回に1回は観察される程度の稀さ：P = 0.1
帰無仮説下でも20回に1回は観察される程度の稀さ：P = 0.05
帰無仮説下でも50回に1回は観察される程度の稀さ：P = 0.02
帰無仮説下でも100回に1回は観察される程度の稀さ：P = 0.01
⋮

- P値が小さいほど稀な結果を観測しているため、棄却されやすくなります。
- 1つの目安として、P値が0.05以下の際に棄却することができ、即ち「統計学的に有意である」と記述されてます。
- P値=0.05とは、たとえ帰無仮説に従っていても、約20回に1回は偶然観測される程度の稀さ、と言う意味です。

③ 帰無仮説とP値

棄却域と採択域の境界



P値>有意水準
なので帰無仮説
は棄却されず

有意水準

- **このように、帰無仮説が有意に棄却されるか、されないかを決めるP値の閾値を、有意水準(α)といいます。**
- **生物統計や医学統計では、有意水準を0.05に設定する例が多いです。**

③ 帰無仮説とP値

$$P < 0.05/1{,}000{,}000 = 5.0\times10^{-8}$$

ボンフェローニ補正：有意水準を検定数で割る
（ゲノムワイド関連解析の場合、n = 1,000,000）

- 1つの帰無仮説に対して多数の検定を繰り返すと、有意水準を下回る結果を観測する確率が高くなるため、補正する必要があります。
- 多重検定補正といい、有意水準を検定数で割った値とする、ボンフェローニ補正が有名です。
- ゲノムワイド関連解析の有意水準：$\alpha = 5.0\times10^{-8}$は、ゲノム全体に100万箇所相当の独立なSNPが存在する、という前提に基づきます。

（100万箇所は目安であり、実際には人種集団に依存して個数が異なります）

③ 帰無仮説とP値

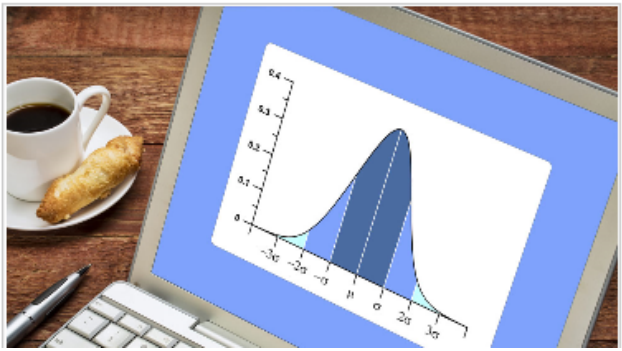
*p*値の誤用の蔓延に米国統計学会が警告

Nature ダイジェスト Vol. 13 No. 6 | doi : 10.1038/ndigest.2016.160612
原文: *Nature* (2016-03-10) | doi: 10.1038/nature.2016.19503 | EN Statisticians issue warning on p values

Monya Baker

科学者による*p*値の誤用を止めるため、米国統計学会（ASA）が異例の声明を出した。

2016年3月8日、米国統計学会（American Statistical Association; ASA）は、科学者の間で*p*値の誤用が蔓延していることが、多くの研究成果を再現できないものにする一因になっていると警告した。ASAは、科学的証拠の強さを判断するために広く用いられている*p*値について、「*p*値では、仮説が真であるか否か、あるいは、結果が重要であるか否かの判断はできない」として、その利用法の指針を発表するという異例の動きに出た。



- 有意水準は１つの基準であり、必ずしも0.05である必要はありません。
- 「有意水準は絶対的な基準」、「有意水準を満たした結果が全て正しい」という解釈は間違っています。
- P値は便利ですが、このように批判もあります。
- 2016年には、米国統計学会がP値の誤用に警鐘を鳴らしました。

③ 帰無仮説とP値

REPRODUCIBILITY

P-value shake-up proposed

Big names in statistics recommend tightening threshold for significance in biomedical science.

BY DALMEET SINGH CHAWLA

Science is in the throes of a reproducibility crisis, and researchers, funders and publishers are increasingly worried that the scholarly literature is littered with unreliable results. Now, a group of 72 prominent researchers is targeting what they say is one cause of the problem: weak statistical standards of evidence for claiming new discoveries.

In many disciplines, the significance of findings is judged by *P* values. They are used to test (and dismiss) a 'null hypothesis', which generally posits that the effect being tested for doesn't exist. The smaller the *P* value that is found for a set of results, the less likely it is that the results are purely due to chance. Results are deemed 'statistically significant' when this value is below 0.05.

But many scientists worry that this threshold has caused too many false positives to appear in the literature, a problem exacerbated by a practice called *P* hacking, in which researchers gather data without first creating a hypothesis to test, and then look for patterns in the results that can be reported as statistically significant.

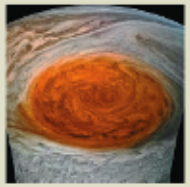
So, in a provocative manuscript posted on the PsyArXiv preprint server on 22 July, researchers argue that *P*-value thresholds should be lowered to 0.005 for the social and biomedical sciences (D. Benjamin *et al.* Preprint at PsyArXiv http://osf.io/preprints/psyarxiv/mky9j; 2017). The final paper is set to be published in *Nature Human Behaviour*.

"Researchers just don't realize how weak the evidence is when the *P* value is 0.05," says Daniel Benjamin, one of the paper's co-lead authors and an economist at the University of Southern California in Los Angeles. He thinks

MORE ONLINE

NASA/IPL-CALTECH/SWIRI/MSSS/JASON MAJOR

IMAGES OF THE MONTH



July's sharpest science shots — as selected by *Nature*'s photo team go.nature.com/2vdlt0g

TOP NEWS

- Why astronomers reluctantly announced a possible exomoon discovery go.nature.com/2tueikq
- Landslide triggered rare Greenland mega-tsunami go.nature.com/2hjtsce
- Clues emerge in mystery of flickering quasars go.nature.com/2tutftt

NATURE PODCAST

The first flower; gene editing human embryos; and the quest for antimatter nature.com/nature/podcast

- 「有意水準の閾値が0.05というのは緩すぎる。$P < 0.005$へと厳しくすべきである」、という意見も出てきました。

(Chawla DS. ***Nature*** 2017)

③ 帰無仮説とP値

P値

・一方で、統計解析の結果に基づき適切な判断を下すためには、一定の共通した基準で結果を解釈する必要があり、P値は有用です。

・P値を適切に解釈することで、統計学を有効利用して下さい。

講義の概要

統計学入門

① 統計学について

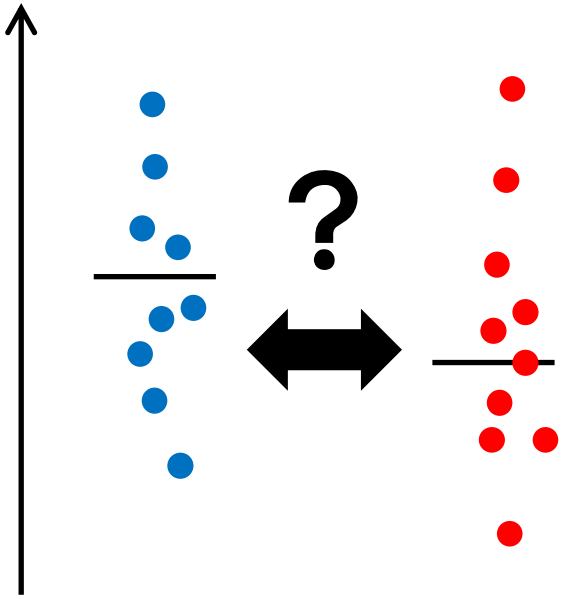
② 母集団と標本集団

③ 帰無仮説とP値

④ 統計検定手法

④ 統計検定手法

t検定



チェックポイント

- 2群間の対応関係の有無？
- 2群間の等分散の有無？
- 正規分布に従う？

- t検定は、2つの母集団がいずれも正規分布に従うと仮定したうえで、2群の平均値が等しいかどうかを検定する統計検定です。
- 2群の間の対応関係の有無や、2群の間に等分散が仮定できるか、により、t検定の種類が変わります。
- 2つの母集団が正規分布に従わない場合は、別の検定手法になります。

④ 統計検定手法

カイ二乗検定

	改善有	改善無	計
治療群	150	100	250
非治療群	250	200	450
計	400	300	700

治療群と非治療群における改善率の差：

カイ二乗値 = 1.1212、P値 = 0.29

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}},$$

1つが決まると残り

3つが自動的に決まる

	改善有	改善無	計
治療群	150		250
非治療群			450
計	400	300	700

自由度=1

	改善有	改善無	計
治療群	150	100	250
非治療群	250	200	450
計	400	300	700

- カイ二乗検定は、観測された度数分布と理論分布の差を、カイ二乗分布に基づき検定します。
- 分割表における独立性の検定が有名です。
- 自由度の概念があります。
- 2×2の分割表の場合、周辺度数が固定された条件下では、内部の数字が1つ決まると残り3つが自動的に決まります（自由度＝1）。

④ 統計検定手法

	改善有	改善無	計
治療群	50		250
非治療群			450
計	400	300	700

→

	改善有	改善無	計
治療群	50	200	250
非治療群	350	100	450
計	400	300	700

	改善有	改善無	計
治療群	100		250
非治療群			450
計	400	300	700

→

	改善有	改善無	計
治療群	100	150	250
非治療群	300	150	450
計	400	300	700

	改善有	改善無	計
治療群	150		250
非治療群			450
計	400	300	700

→

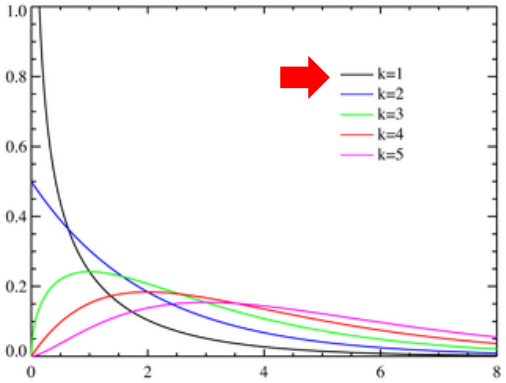
	改善有	改善無	計
治療群	150	100	250
非治療群	250	200	450
計	400	300	700

	改善有	改善無	計
治療群	200		250
非治療群			450
計	400	300	700

→

	改善有	改善無	計
治療群	200	50	250
非治療群	200	250	450
計	400	300	700

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}},$$



- カイ二乗検定は、観測された度数分布と理論分布の差を表す統計量が、近似的にカイ二乗分布に従うことを利用しています。
- 実際、２×２分割表の周辺度数を固定した上で、（超幾何分布に従う生起確率に基づき）内部度数を変化させると、得られたカイ二乗値の分布は、自由度1のカイ二乗分布に従うことが確認できます。

④ 統計検定手法

カイ二乗検定

度数の値が大きい

	改善有	改善無	計
治療群	150	100	250
非治療群	250	200	450
計	400	300	700

➡ カイ二乗検定

度数の値が小さい

	改善有	改善無	計
治療群	1	1	2
非治療群	3	2	5
計	4	3	7

➡ 正確確率検定

- 分割表の度数（内部の数字）が小さい時、近似が不正確になるため（＝統計量の分布がカイ二乗分布に従わなくなる）、フィッシャーの正確確率検定の使用や、イエイツの補正が推奨されます。
- 1つの目安として、「度数が5以下」という指標が知られています。

④ 統計検定手法

	変異有	変異無	計
患者群	2	14	16
健常者群	0	61	61
計	2	75	77

- 線型回帰 ： P値 = 0.0047
- カイ二乗検定 ： P値 = 0.0051
- 正確確率検定(Mid-P値) ： P値 = 0.021
- フィッシャー正確確率検定 ： P値 = 0.041

----------（有意水準 = 0.05）----------

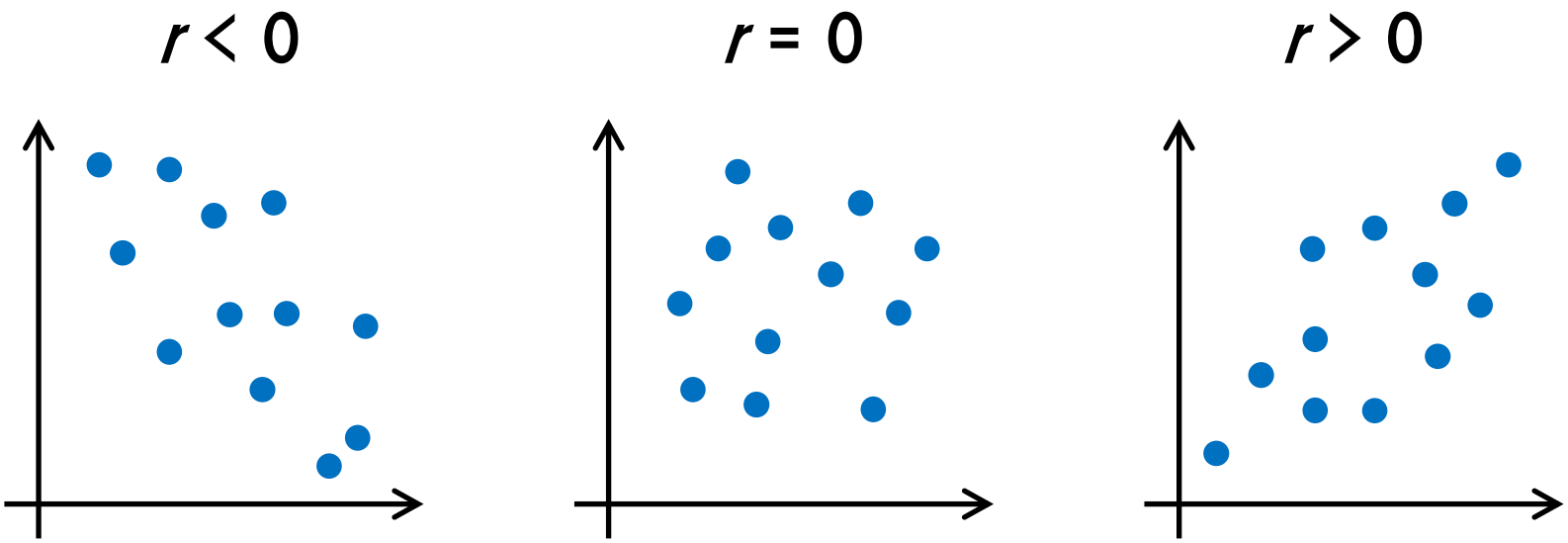
- カイ二乗検定＋イエイツ補正 ： P値 = 0.056
- ロジスティック回帰 ： P値 = 0.99

- 別の言い方をすると、同じ観測結果(分割表)であっても、**統計検定手法が異なれば異なるP値が得られる**、ということです。
- 色々試した上で、「P値がなるべく小さく(or 大きく)なる検定手法を選ぶ」のは、解釈にバイアスを生じる原因となるので、**絶対ダメ**です。
- 設定した帰無仮説下において、**適切なtype I errorを示す検定**を選ぶ必要があります。

(Okada Y et al. ***JAMA Opthalmol*** 2017)

④ 統計検定手法

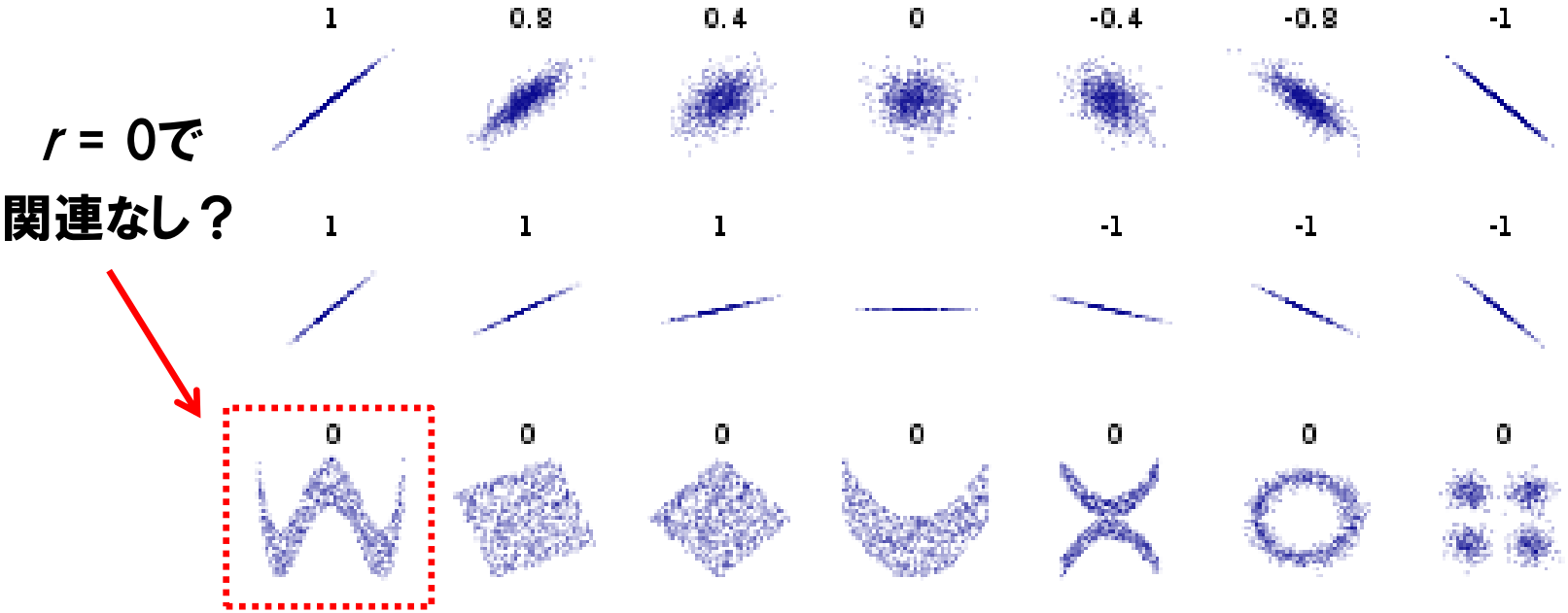
相関係数



- **相関係数**（*r*）は、対応のある2群の間の相関関係（関連の強さ・非独立性）を示す統計学の指標です。
- **−1から1**の間を取り、1に近い時は正の相関、0の時に相関なし、−1の時に負の相関、を意味します。
- 相関関係は因果関係とは異なり、**原因と結果を区別しません。**

④ 統計検定手法

相関係数

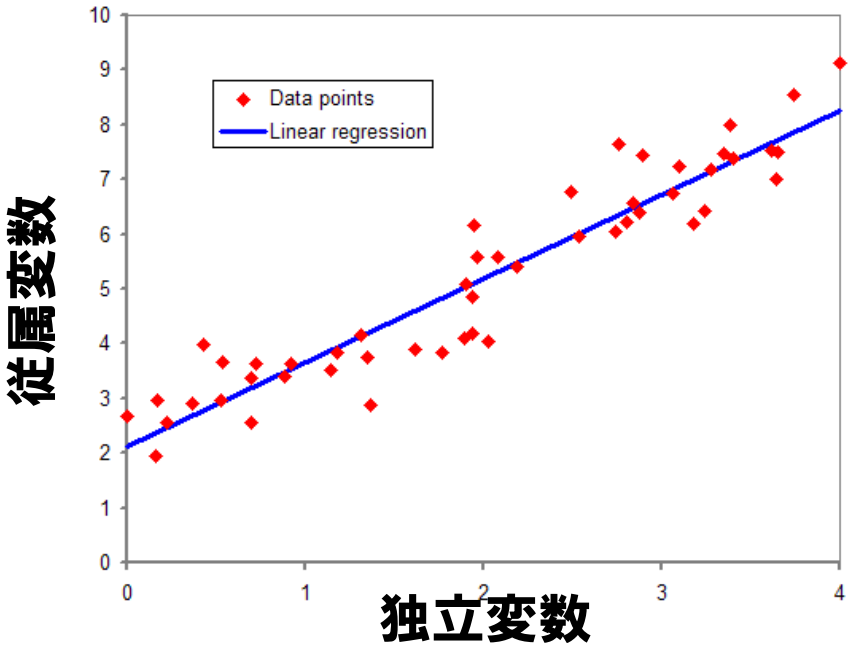


- **相関係数**は、２群の間の相関関係を、直線関係に基づき判断します。
- ２群の間に直線関係以外の関係性がある場合や、直線関係の傾きは、相関係数の値に反映されません。
- 確認も兼ねて、相関係数を計算する前に、一度、**２群の値をグラフにプロットしてみると良い**でしょう。

(Wikipediaより転載)

④ 統計検定手法

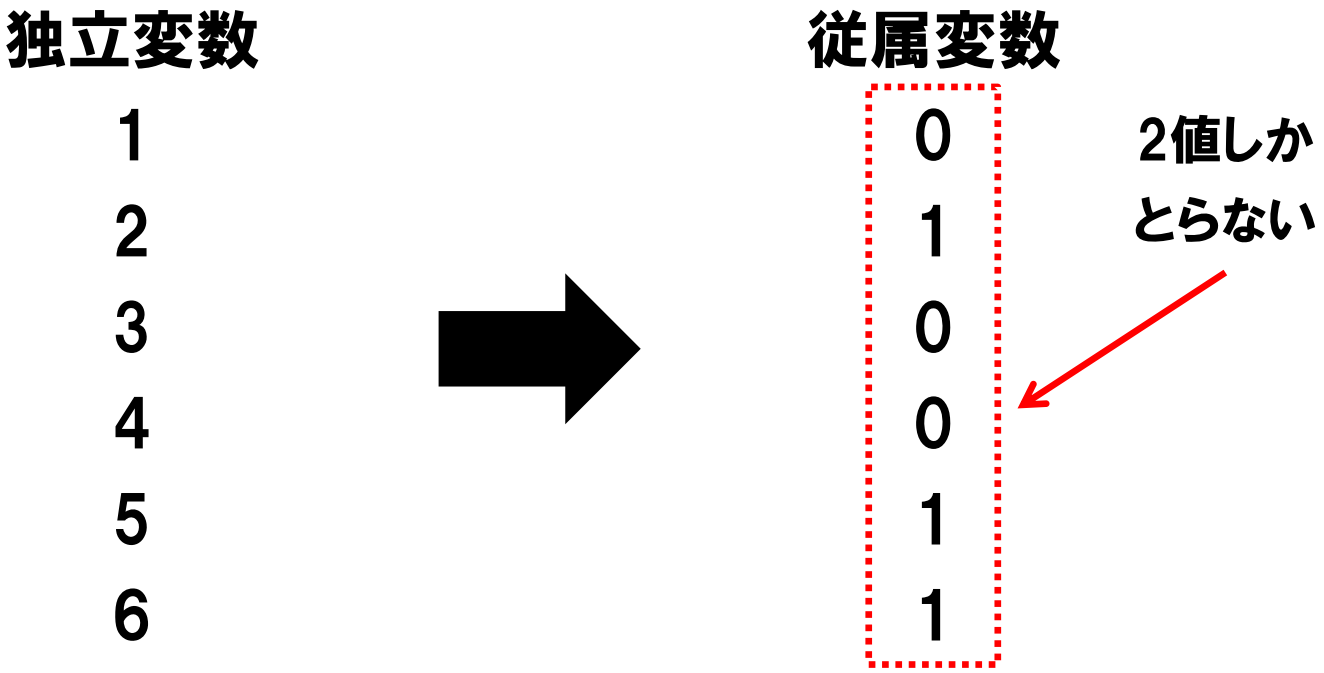
線形回帰



- **線形回帰**は、対応のある2群である**従属変数**（応答変数）と**独立変数**（説明変数）について、独立変数が従属変数をどれだけ説明できるかを評価します。
- 独立変数が一つの場合は**単回帰**、複数存在する場合は**多重回帰**といいます。

④ 統計検定手法

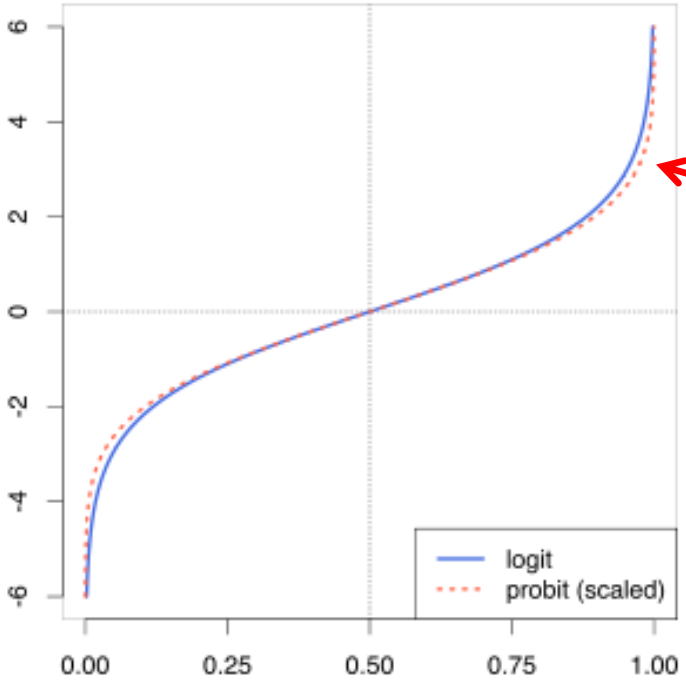
ロジスティック回帰



- ロジスティック回帰は、「あり」、「なし」のように従属変数が2値を取る場合に採用される回帰モデルです。
- 患者群と対照群、治療群と非治療群、というふうに、医療統計学では2値で表される結果の検討を行う例が多く、重宝されています。

④ 統計検定手法

線形回帰とロジスティック回帰

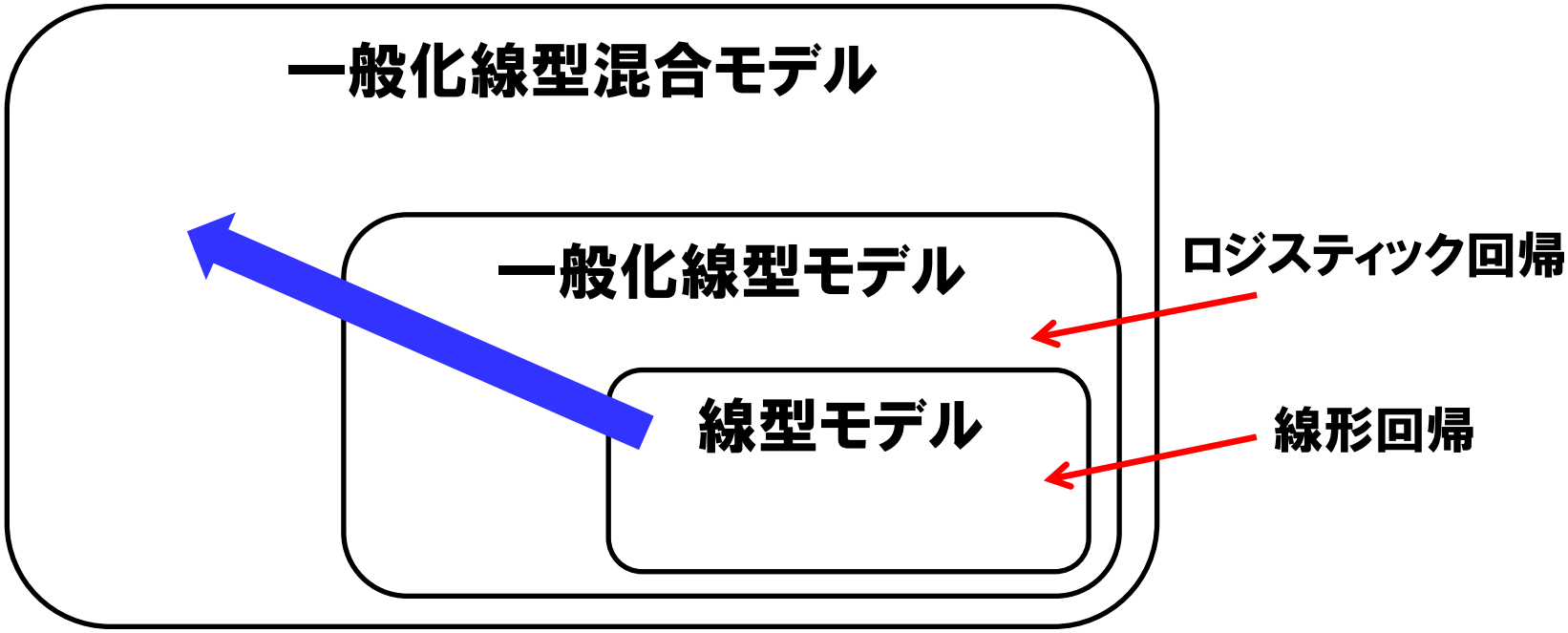


よく似てる
（けどちょっと違う）

- 線形回帰がプロビット曲線(≒正規分布)を扱う一方で、ロジスティック回帰はロジット曲線を扱います。
- ロジット曲線はプロビット曲線とは異なりますが、形がよく似ているため、同様に使われています。

④ 統計検定手法

線形回帰とロジスティック回帰



- 線形回帰とロジスティック回帰は、理論的にも近い関係にあります。
- 線型モデル → 一般化線型モデル → 一般化線型混合モデルと、階層的にモデルが拡張されていきます。

④ 統計検定手法

1標本t検定、2標本t検定、対応のある2標本t検定、一元配置分散分析、2元配置分散分析、F検定、バートレット検定、ルビーン検定、マクネマー検定、コックランQ検定、マンホイットニー検定、クラスカル・ウォリス検定、フリードマン検定、ウィルコクソン検定、コルモゴロフ・スミノルフ検定、符号検定、カイ二乗検定、フィッシャー正確確率検定、MidP値、トレンド検定、マクネマー検定、コックラン・アーミテージ検定、マンテルヘンツェル検定、ヨンクヒール検定、スミルノフ・グラブス検定、ピアソン積率相関係数、ケンドール積率相関係数、スピアマン順位相関係数、線形回帰、重回帰、偏相関係数、ロジスティック回帰、非線形回帰、一般線型化モデル、一般化線型混合モデル、主成分分析、クラスター分析、カプランマイヤー生存曲線、ログランク検定、コックス比例ハザードモデル

- これまでの研究により、沢山の統計検定手法が提唱されています。
- 数十年前に開発されて埋もれていた統計検定が、ビッグデータ時代に入り発掘され、脚光を浴びている、なんて例もあります。

（例：SNPリスク解析におけるコックラン・アーミテージ検定）

終わりに

・統計学の初歩的な概念と、統計検定の簡単な説明と使い方を、なぞってみました。

・個別の検定手法については、必要に応じて調べてみて下さい。

・統計学に関わる数式を理解するのは確かに難しいですが、概念を把握して適切に解釈することは、そこまで難解ではありません。

・統計ソフトの発達により、誰でも統計解析が実施できるようになった反面、適切な解釈が追いついていない面があります。

・P値も含め、統計学を正しく使いこなせるようにしていきましょう。