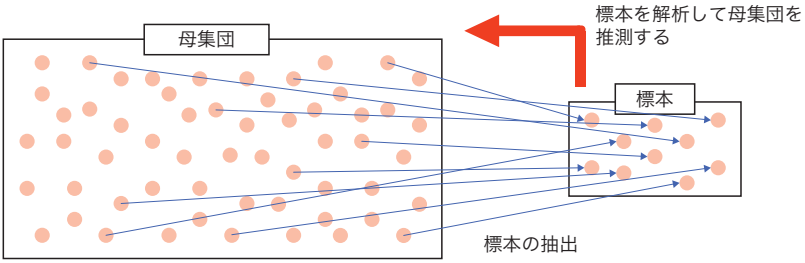
第1章 統計の基礎知識

1 なぜ統計解析が必要なのか？

人間は自分自身の経験に基づいて、感覚的にものごとを判断しがちである。例えばある疾患に対する標準治療薬の有効率が50%であったとする。そこに新薬が登場し、ある医師がその新薬を5人の患者に使ったところ、4人が有効と判定されたとしたら、多くの医師はこれまでの標準治療薬よりも新薬のほうが有効性が高そうだと感じることだろう。しかし、たまたま有効性が出やすい5人に治療が行われたにすぎないかもしれない。同じ疾患を有する患者であったとしても、疾患の細かな分類や進行度、患者の年齢、性別、臓器の状態などによって有効率は左右される。さらに背景の条件が全く同じであったとしても有効率にばらつきは生じる。治療に対して思い入れが強ければ強いほど、治療結果に大きく一喜一憂し、客観的な評価が困難となる。印象に残る結果は感覚的な判断を偏らせてしまう。

統計解析の目的は、前提としてこのような様々なばらつきが存在する状況の中で、限られた**標本（sample）**から**母集団（population）**を**推測**し、より一般的な結論を導き出そうとすることである。**母集団**の定義は状況によって異なるが、例えばある疾患に対する新薬の有効性を評価するのであれば、その疾患を有する全ての患者が母集団となる。統計解析をしていると、目の前にあるデータだけを対象としているような錯覚にとらわれることがあるが、実際に行っていることは、その標本を用いて本当の**母集団の全体像を推定**しようとしているのである（選挙の出口調査による全体の投票数や議席数の予測をイメージすればよい）。

統計解析の目的は、母集団から抽出した標本（サンプル）を用いて解析することによって、母集団を推測することである



2 変数の種類とその要約

❶ 変数の種類

統計解析で扱う主な変数は**連続変数（continuous variable）**、**順序変数（ordinal variable）**、**名義変数（nominal variable）**の3つに分けられる。連続変数は身長、体重など、数値で表される**定量的なデータ**を意味する。順序変数、名義変数はいずれも**質的なデータ（カテゴリー変数、categorical variable）**であるが、順序変数は尿蛋白の（−）、（±）、（+）、（2+）、（3+）や、腫瘍の進行度のステージI、II、III、IVのように順序づけられたものである。一方、名義変数は、性別の男性、女性や、ABO血液型のA、B、O、AB型のように順序の関係がない（男性、女性、あるいは有効、無効のように二値だけを持つ場合は**二値変数**あるいは**二区分変数（binary variable）**とも呼ばれる）。

特殊な変数として医学統計ではしばしば**生存期間**の解析が行われる。正確にいうと、必ずしも生存期間だけを対象とする解析ではなく、ある時点からあるできごと（イベント）が発生するまでの期間（time-to-event variable）の解析であり、死亡がイベントとして定義された場合に生存期間の解析が行われることになる。この解析方法の特徴は、ある時期まで生存していた（あるいはイベントが発生していなかった）ことは知られているが、その後の情報が得られないような場合に**観察打ち切り（censor）**として解析に含めることができる点である。例えば、ある疾患に対して特定の治療を行った後の生存期間を解析する場合に、最終観察時点で生存中の患者の真の生存期間は不明であるが、その時点で打ち切りとして扱うことによって解析に含めることができる。この解析においては、イベントが1回しか発生しないものであることと、打ち切りとなる理由が解析対象のイベントの発生とは無関係であることが必要である。例えば悪性腫瘍に対する化学療法後の生存期間の解析において、打ち切りとなった理由が他院への転院のような場合は、病状が増悪して死期の近づいた患者がしばしばホスピスに転院するという背景が解析上の**偏り（バイアス）**を生じてしまう危険性がある。

❷ 変数の要約、信頼区間

各変数を**要約**して記述する方法はそれぞれの解析のところで詳しく述べるが、まずは全体像を眺めることが重要である。名義変数なら**頻度分布**を、連続変数であれば**散布図**、**ヒストグラム**、**箱ひげ図**などを描いてみる。生存期間を表すためには**Kaplan-Meier曲線**が用いられる。各変数を端的に記述するには、そ

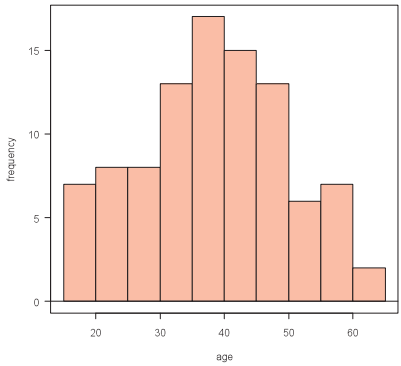
れらを**代表する値**と**信頼区間（confidence interval, CI）**が役に立つ。例えば、有効と無効の二値の名義変数なら**比率**（有効率）とその信頼区間、正規分布に従う連続変数ならその**平均値**とその信頼区間（あるいはばらつきを示したければ平均値と標準偏差）などである。50 人の患者にある治療を行って 30 人が有効、20 人が無効という結果であったとしたら、有効率は 30/50=60% である。この 60% という数値が母集団の有効率に対する**点推定**である。一方、信頼区間の計算は**区間推定**といわれる。母集団から 100 回サンプルを抽出して信頼区間を計算したら 95 回は母集団の真の値（比率、平均値など）を含んでいるという区間が 95% 信頼区間である（似通った表現だが、「母集団の真の比率が 95%

名義変数、連続変数を要約・記述する方法の例

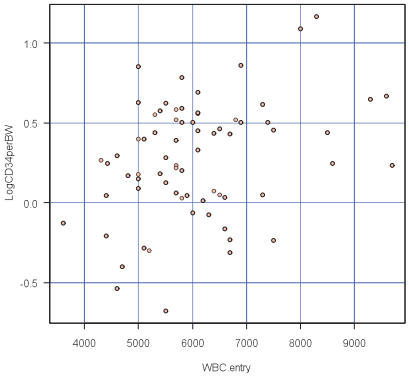
分割表

	CR	NR	PR	合計
A	18	30	13	61
B	30	18	12	60
合計	48	48	25	121

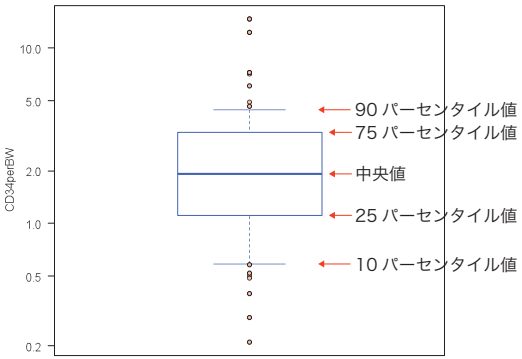
ヒストグラム



散布図



箱ひげ図

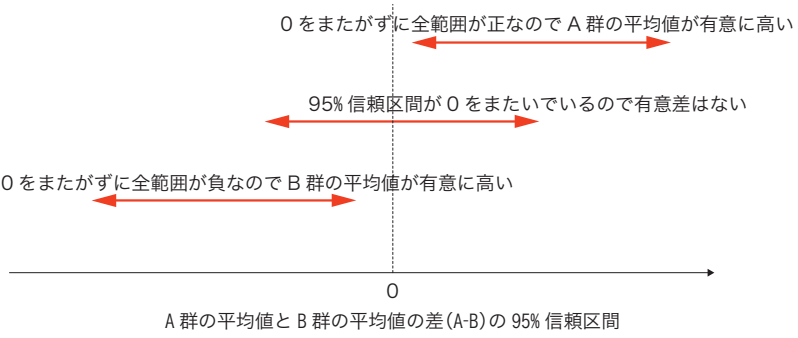


信頼区間の中に含まれる確率が95%」という表現とは異なる。真の母集団の比率は常に一定であり、サンプリングするごとに信頼区間の方が変化するのである)。なお、95%という数値は慣習上しばしば使われているだけであり、状況によっては99%信頼区間や90%信頼区間なども用いられる。*P*値の**有意水準**として慣習的に5%がしばしば用いられていることと同じことである。サンプル数が増えるほど区間推定の幅は狭くなり、点推定値の信頼度も高まる。

3 群間の比較、*P*値とは？

2群を統計学的に比較するには2つの方法がある。これらは同じ統計学的原理と前提に基づいている。1つは2群の差あるいは比の信頼区間を計算することである。2群の差の95%信頼区間が0を含まなければ、あるいは2群の比の95%信頼区間が1を含まなければ有意差があると結論される（これは$P<0.05$に相当する)。もう1つは*P*値を計算（**検定**）することである。まず、*P*値を計算する前に、サンプルが母集団からランダムに抽出されているという前提のもとで**帰無仮説（null hypothesis, H_0）**をたてる。帰無仮説とは、2つの母集団には違いはなく、観察された結果における2群の差は偶然にすぎないという仮説である。*P*値はこの帰無仮説が真である場合に、実際に観察された、あるいはそれ以上の2群の差が観察される確率である。この確率が非常に小さい場合、帰無仮説は正しくないと判断され（**棄却**され)、2群に有意な差があると考える。

2群の差の95%信頼区間による群間比較
矢印の幅が信頼区間を示す。信頼区間が0をまたいでいない場合に有意差があると考える。



*P*値がどれぐらい小さければ有意と判断するかの閾値が**有意水準（significance level, α）**である。αは習慣上0.05（5%）に設定されている（つまり5%ぐらいのエラーは容認せざるを得ないという前提）が、目的に応じて定められるべきであり、状況によっては0.01、0.001などが用いられることもある。*P*値がαよりも小さければ有意と判断するわけであるが、すると帰無仮説が実際には真であるにもかかわらず、それを棄却してしまう過誤（エラー）が生じる確率もαとなる。このような過誤を第Ⅰ種の過誤**（Type Ⅰ error, α error）**という。逆に実際には帰無仮説が偽であるにもかかわらず、これを棄却しない過誤を第Ⅱ種の過誤**（Type Ⅱ error, β error）**という。αの値を小さくすると第Ⅰ種の過誤は減少するが第Ⅱ種の過誤が増加し、逆にαの値を大きくすると、第Ⅰ種の過誤は増加するが、第Ⅱ種の過誤は減少する。両方の過誤を減少させる唯一の方法はより大きいサンプルを集めることである。サンプルサイズが大きくなればβは小さくなり、すなわち統計学的な**検出力（power, 1−β）**は大きくなる。

検定に関連する用語

帰無仮説	比較する母集団の間には差はなく、観察された差は偶然にすぎないという仮説。
*P*値	帰無仮説が真である場合に、実際に観察された、あるいはそれ以上の差が観察される確率。
有意水準（α）	*P*値がどれぐらい小さければ有意と判断するかの閾値。
検出力（1-β）	帰無仮説が偽である場合に、帰無仮説を正しく棄却できる確率。
第Ⅰ種の過誤	帰無仮説が真であるにもかかわらず、これを棄却してしまう誤り（本当は差がないのに有意差があると結論してしまう誤り）。
第Ⅱ種の過誤	帰無仮説が偽であるにもかかわらず、これを棄却しない誤り（本当は差があるのにそれを有意差として検出できない誤り）。

*P*値は観察された差が偶然によるものとして矛盾しないかどうかだけを検討する値であり、実際に観察された差の大きさ（**エフェクトサイズ、effect size**）を判断するものではない。例えば、実質的には意味のないような小さな差であったとしても、サンプルサイズを大きくすることによって有意差として検出される可能性がある（従って、群間の差や比の信頼区間を示す方が情報量が多い）。また、*P*値が小さくなかったとしても、それは有意差を検出することができなかっただけであり、2群に差がないという結論はできない。有意差が検出できなかった原因は単にサンプルサイズが小さかったからかもしれない