

## 第4回 データを分析する (2)

宮井 信行 (和歌山県立医科大学大学院 教授)

### 1. 検定の基礎

#### 1) 検定とは

研究テーマが設定されると、その研究の概念上の対象集団 (母集団) が決まります。研究において母集団の構成員すべてを調査 (全数調査) できれば、その結果は最も正確なものとなりますが、多くの場合は、母集団から実施可能な規模の集団 (標本) を選び出して調査 (標本調査) を行い、そのデータの解析結果に基づいて母集団の姿を推測する方法をとります。このときの統計解析が「推計統計」で、「推定」と「検定」の2種類に分けられます。このうち、「検定 (統計的仮説検定)」とは、標本で観察された平均や割合の差または変数間の関連性が、母集団にも当てはまるかどうかを数学的な確率論を用いて検証することです。

#### 2) 帰無仮説と対立仮説

標本で観察された群間の差や変数間の関連性が母集団にも当てはまるのか、それとも標本

でのみ偶然に起こっただけなのかを直接的に証明することは不可能です。そこで、検定では、母集団についての仮説を設定し、その仮説が成り立つかどうかを検証します (仮説検定)。このとき設定する仮説が「帰無仮説」と「対立仮説」です。帰無仮説は、例えば、差の検定の場合には、母集団に「差がない (等しい)」という仮説で、対立仮説は、実際に立証したい事柄、すなわち帰無仮説とは反対の「差がある」という仮説になります。

#### 3) 検定の手順

差の検定では、最初に、母集団の平均や割合に「差がない」という帰無仮説を設定し、その仮説が成り立っているとの仮定のもとで、標本において観察された群間の差が偶然に起こりえる確率がどの程度かを計算します。この確率を有意確率 (p 値) といいます。そして、p 値が極めて小さいときは (例えば、5%や1%)、帰無仮説がそもそも誤っていたと判断し、「差が

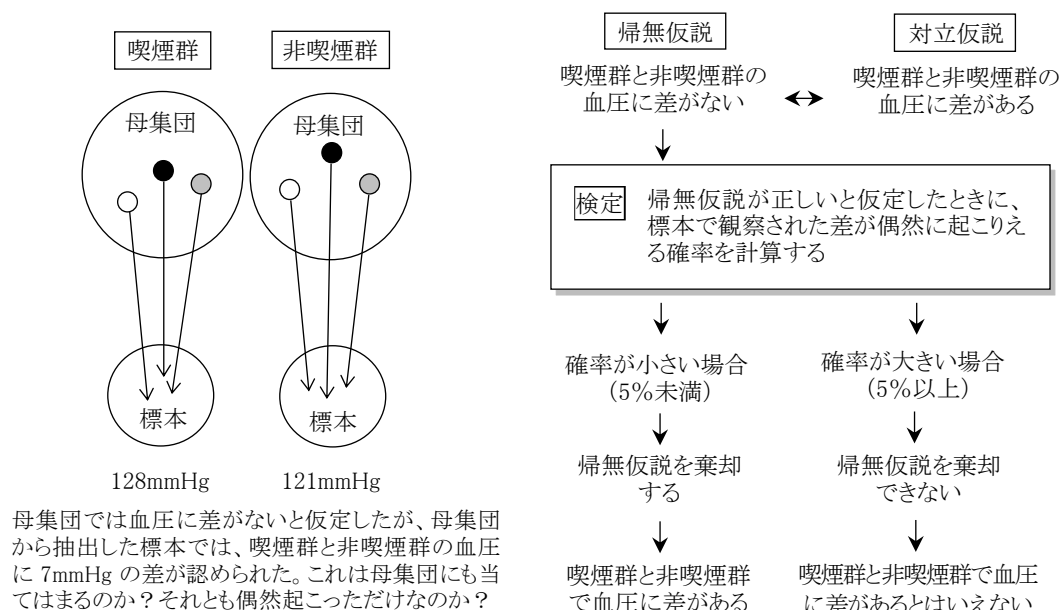


図1. 仮説検定の考え方と手順

ある」という対立仮説が正しいと考えます。つまり、検定では、数学的な確率論によって、帰無仮説を否定することで対立仮説を採択するという方法をとります。

## 2. 検定結果の解釈と注意

### 1) 有意水準と仮説の棄却・採択

帰無仮説を棄却するか否かを判断する基準を「有意水準」といい、通常、5% (0.05), 1% (0.01), 0.1% (0.001) が用いられます。検定では、有意水準よりも計算した p 値が小さいときは、帰無仮説を棄却して、「差がある」と判定します。一方、p 値が大きいときは、帰無仮説を棄却しないこととなりますが、これは、帰無仮説が正しいということではなく、標本における差は母集団で差があるためなのか、それとも偶然に起こっただけなのかが判断できないこととなります。したがって、この場合は、母集団の平均や割合に「差があるとはいえない」という判定になります。帰無仮説を採用して「差がない」または「等しい」という結論にははいけません。

### 2) 検定における誤り

母集団の平均や割合に差があるかどうかの真実は1つですが、検定では確率によって「差がある」か「差があるとはいえない」かの判断をするため、どちらに判定しても誤り（過誤）を犯す危険は免れません。検定における誤りには、「第1種の過誤（ $\alpha$ エラー）」と「第2種の過誤（ $\beta$ エラー）」があります。 $\alpha$ エラーは、帰無仮説が正しいのにこれを否定して棄却してしまう過誤、すなわち、母集団では「差がない」

のに検定で「差がある」と判定する誤りのことです。一方、 $\beta$ エラーは、帰無仮説が誤りであるのにこれを否定せず棄却しない過誤で、母集団では「差がある」のに検定で「差があるとはいえない」と判定する誤りになります。

検定で誤りが起こるか否かは、有意水準や標本サイズ（対象の人数）に依存し、 $\alpha$ エラーを犯す危険を小さくするために有意水準を小さくすれば、それは同時に、 $\beta$ エラーを犯す危険を大きくすることになります。また、標本サイズを大きくすると $\alpha$ エラーが起こりやすくなり、逆に、小さすぎると $\beta$ エラーの可能性が高くなります。

### 3) 統計学的有意性と臨床的有意性

統計学的有意性とは、「その結果が仮定した分布のもとでは起こりそうもない」ことを意味するもので、このことは、臨床的（または生物学的）な有意性とは必ずしも一致しません。つまり、統計学的に有意な差が認められたとしても、それが実質的に意味のない僅かな差であれば、その価値はほとんどないこととなります。検定の結果がすべてではなく、その差がどの程度意味があるかも含めて結果を判断する必要があります。

これに関連して、検定で有意な差が認められた場合に、それが有意水準5%よりも1%のときのほうが群間の差が大きいというような、p 値の解釈に関する誤解もみられます。しかし、実際の差の程度と p 値は無関係で、p 値の大小によって差の程度を判断することはできません。差の程度を評価するための指標には、信頼区間や効果量などがあります。データの記述的な解析においては、信頼区間を使用した区間推定を行うこと、また、重要な知見や p 値を報告する際には、効果量を併記することが望ましいといえます。

### 3. 検定方法の選択

標本で観察された平均や割合の差が母集団にも当てはまるかどうかを検証するための検定方法には多種多様なものがあり、何を知れた

表 1. 母集団の真の姿と仮説検定の結果

		母集団の真の姿	
		差がない	差がある
検定の結果	差がない (帰無仮説)	正しい判定 $1-\alpha$	第2種の過誤 $\beta$
	差がある (対立仮説)	第1種の過誤 $\alpha$ (=有意水準)	正しい判定 $1-\beta$ (=検出力)

表 2. 差の検定方法の一覧

	連続変数		カテゴリー変数
	パラメトリック	ノンパラメトリック	
対応のある差 (2 変数)	対応のある t 検定	Wilcoxon の符号付き 順位検定	McNemar 検定
独立した標本の差 (2 標本)	2 標本 t 検定 (Welch の t 検定)	Mann-Whitney の U 検定	$\chi^2$ 検定
対応のある差 (3 変数以上)	反復測定の分散分析 多重比較法	Friedman 検定 多重比較法	Cochran Q 検定
独立した標本の差 (3 標本以上)	1 元配置分散分析 多重比較法	Kruskal-Wallis 検定 多重比較法	$m \times n$ $\chi^2$ 検定

いかという目的とともに、分析するデータについての変数の種類、分布の正規性や等分散性、対応の有無、比較する標本または変数の数などによって使い分けることになります。

### 1) 変数の種類

分析する変数の種類が、連続変数（量的な特性のデータ）であるか、カテゴリー変数（質的な特性のデータ）であるかによって、分布の傾向や特徴を要約する特性値が異なるため、それに応じて検定方法も違ってくることになります。連続変数では平均値や中央値が、カテゴリー変数では割合や比が分布の代表値となります。差の検定を行うときは、まず、分析するデータが連続変数なのか、カテゴリー変数なのかを確認します。

### 2) 分布の正規性

連続変数のデータの数値は、左右対称の釣形鐘型の分布（正規分布）をとることが多いですが、分布の裾が右（高値側）に長かったり、左

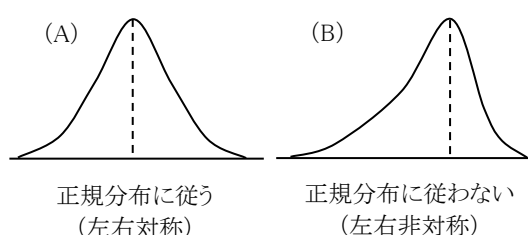


図 2. 連続変数のデータの分布

（低値側）に長かったりといった左右非対称の歪んだ分布（非正規分布）を示すこともあります。連続変数の差の検定方法は、母集団の分布が正規分布であることを仮定して行う方法（パラメトリック検定）と、分布を規定せずに行う方法（ノンパラメトリック検定）の 2 種類に分けられます。

通常、正規分布に従うデータの場合は、平均値を分布の代表値とすることができるので、パラメトリック検定を適用します。一方、正規分布に従わず、平均値が代表値として相応しくないときは、ノンパラメトリック検定を適用することになります。

ノンパラメトリック検定は、正規分布に従うデータにも適用することができますが、そのときは、 $\beta$ エラーを犯す可能性が高まるなどのいくつかの問題が生じます。また、本来はパラメトリック検定を適用すべきデータであるにも拘わらず、標本サイズが小さいためにノンパラメトリック検定を適用するという誤用も比較的多くみられます。

### 3) 対応の有無

比較するデータが独立関係（対応がない）にあるか、従属関係（対応がある）にあるかによっても検定方法が違ってきます。独立関係とは比較するデータが属性の異なる複数の群から得られている場合であり、従属関係とは 1 つの群に対して条件や時間を変えて 2 回以上繰り返

し収集されたデータのことを示します。例えば、平均の差を検定する場合、従属関係にある2変数を比較するときは、対応のあるt検定を適用し、独立関係にある2群を比較したいときは、独立2標本のt検定を適用することになります。

#### 4) 等分散性

独立した2群の平均の差を検定する際には、比較する群のデータの分散が等しいかどうかによって異なる検定方法を用います。F検定などの等分散性の検定で、帰無仮説が棄却されず等分散を仮定することができるときは、通常の2標本t検定を適用しますが、仮定できないときには、ウェルチの補正による2標本t検定を適用することになります。

#### 5) 比較する群または変数の数

独立した群の場合は、それが2群か3群以上であるかによって、また、対応のある変数の場合には、前後で2回測定したのか、それとも3回以上反復して測定したのかによって検定方法が異なります。3群（または3変数）以上の

差の検定では、分散分析や多重比較法などの複雑な手法が適用されますが、選択にあたっての基本的な考え方は2群（または2変数）の差の検定のとおりです。

多重比較法は、3群以上の差の検定において、どの群の組合せに差があるかを知るために行う解析ですが、これにもさまざまな手法があり、分布の正規性や等分散性が仮定できるかどうかなどによって使い分けることになります。なお、3群以上の差の比較において、2群の差の検定を繰り返すと、検定多重性（多重比較）の問題が生じるため避けなければいけません。

これまでに述べてきた差の検定はすべて扱う要因が1つでしたが、分散分析は、差をみたい要因の数によって、2元配置（2要因）や3元配置（3要因）などの種類があります。2要因以上の分散分析においては、各要因の群間に差（主効果）があるかどうかに加えて、要因間に相乗効果や相殺効果（交互作用）がみられるかどうかを検証します。

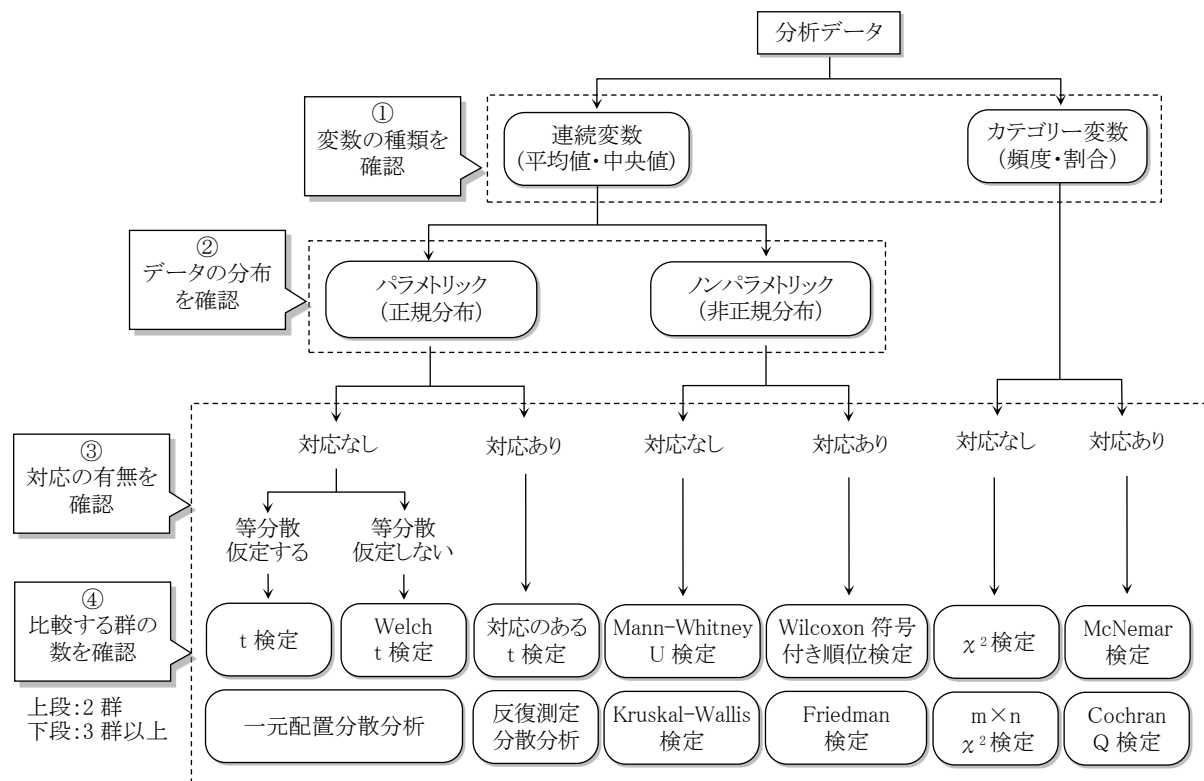


図3. 差の検定方法を選択するためのフローチャート

藤林和俊 著：看護研究をはじめのための統計と臨床疫学、p78、学研メディカル秀潤社、2013 から改変引用