連載　統計学のキー・ポイント―「検定」に焦点を当てて・3

平均値の差についての検定

Tests for the Difference of Means Between Two Groups

髙木廣文

看　護　研　究

第47巻 第3号 別刷

2014年6月15日発行

KANGO-KENKYU(*The Japanese Journal of Nursing Research*)

Vol.47 No.3/May.–Jun.2014

医学書院

第3回

平均値の差についての検定

Tests for the Difference of Means Between Two Groups

髙木廣文　東邦大学看護学部長

はじめに

新薬と標準薬の効果の比較，ある処理(例えば手術や教育など)の有無による健康への影響の比較など，介入群とコントロール群に分けて，2グループ間での比較を行なう研究は極めて多いと考えられます。特に統計学的分析でよく行なわれる方法は，ある変数の平均値について，2グループ間で違いがあるといえるのかを決めるための検定です。もう少し正確に言うと，調査などで得られたデータをもとにして，帰無仮説「ある変数について2グループ間で母集団の母平均に差はない(2グループの母平均は等しい)」を検定することです。次のような例題をもとに，この検定方法について考えてみたいと思います。

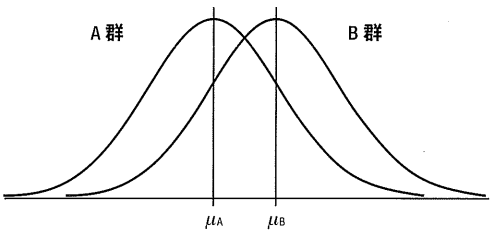
【例題】生活習慣調査を行ない，高塩分の食事をする女性40名(Aグループ)と低塩分の食事をする女性50名(Bグループ)の血圧を調べたとします。収縮期血圧について，**表1**のような結果が得られたとします。ただし，表1の標準偏差は不偏分散の平方根とします。

表1に示した結果から，高塩分食事グループは血圧が低塩分食事グループよりも高いと言えるのでしょうか。

表1　2グループの収縮期血圧

グループ	人数	収縮期血圧(mmHg)	
		平均値	標準偏差
高塩分食事群(A)	40	115	15
低塩分食事群(B)	50	110	10

図1　等分散の場合の2つの母平均値の差の検定

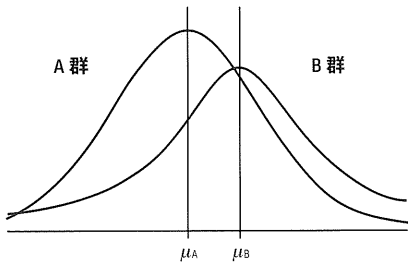


このような問に答えるのが，「独立2群の母平均値の差の検定」と言われる検定方法です。ある変数についての2グループの母平均値の差の検定は，極めてよく用いられている方法ですが，実際の検定では，2グループの母分散が，**(1)等しい場合(等分散の場合)**，**(2)等しくない場合(不等分散の場合)**で，用いる計算方法が異なっています。

(1)の等分散の場合とは，**図1**で示したように，変数の分布の形状は2つのグループで同一なものと考えられています。ただし，2つのグループAとBでは，母平均値に違いがあり，そのため分布全体の位置がAに比べてBでは，右側に移動しているという状態になっています。したがって，

0022-8370/14/¥600/論文/JCOPY

図2　不等分散の場合の2つの母平均値の差の検定

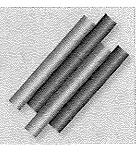


分布の形状は同じなので，分散は2つのグループでは同一，つまり等しいものと仮定されています。このように，一方のグループが他方に比べて全体に一定方向に偏った状態についての検定方法が，等分散の場合の母平均値の差の検定になります。

等分散の場合とは異なり，**図2**に示したように不等分散の場合の検定は，2つのグループの母平均値μ_Aとμ_Bに違いがあるだけではなく，各グループの分布の母分散σ_A^2とσ_B^2にも違いがある場合です。したがってこのような場合には，2つのグループの分布は位置も形状も異なることになります。このような2グループの比較については，分散が異なるということは分布が異なることを意味しますので，2つの母平均値の比較に意味があるのかという議論もあります。しかし，明らかに2グループの分布が異なっている不等分散のような場合でも，2グループの母平均値を比べたいという要望はありえますので，ここでは後述するように「Welchの方法」を説明したいと思います。

以下に，いくつかの検定方法を解説したいと思います。

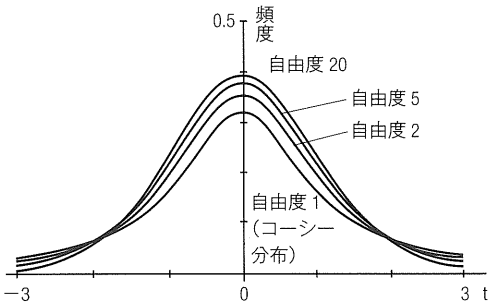
母分散が等しい場合の母平均値の差の検定



なんらかの理由で，2グループの母分散が等しいと仮定できる場合か，後述する「等分散の検定」で有意差が認められない場合に用いる方法です。

ある変数が正規分布に従っており，2グループの分布の母分散は等しいものと仮定します。2グループA，Bの標本数をm，n，平均値を$\bar{x}$，$\bar{y}$とし，不偏分散を$\hat{\sigma}_x^2$，$\hat{\sigma}_y^2$とします。

図3　t分布の自由度と形状



母分散が等しい場合，帰無仮説「2グループの母平均値μ_x，μ_yは等しい；$\mu_x=\mu_y$」の検定統計量は，

$$t_0=\frac{|\bar{x}-\bar{y}|}{\sqrt{\dfrac{(m-1)\hat{\sigma}_x^2+(n-1)\hat{\sigma}_y^2}{m+n-2}\times\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}}$$

によって求められます。

ただし，ここで定義される統計量t_0は，自由度$v=m+n-2$，のt分布に従います。

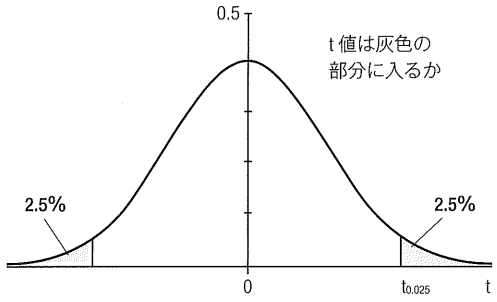
図3に示したように，t分布は自由度によってその形状が異なりますが，標本数が大きくなると，すぐに標準正規分布に近づきます。

実際の検定では，統計量t_0のt分布での両側確率pを求めて，その値が0.05(5%)以下ならば，帰無仮説を棄却し，2グループの母平均値に有意差があるとすればよい訳です。一方，p値が0.05より大きければ，有意差は認められないので，帰無仮説は棄却できず，2グループの母平均値に差があるとは言えなくなります(**図4**)。

先の【例題】について，表1の数値を上記の式に入れると，検定のための統計量は，

$$t_0=\frac{|115-110|}{\sqrt{\dfrac{(40-1)\times15^2+(50-1)\times10^2}{40+50-2}\times\left(\dfrac{1}{40}+\dfrac{1}{50}\right)}}$$
$$=1.891$$

図4　t検定の考え方



と求められます。

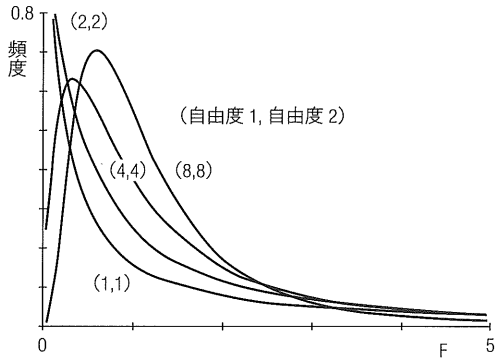
先の統計量 t_0 の自由度は，$v=40+50-2=88$ となり，自由度88のt分布で，$t_0=1.891$ の両側確率は，$p=0.062$ と求められます(当然ですが，専用のソフトを用いてコンピュータで計算します)。6.2%は比較的小さな確率ですが，5%よりは大きくなっています。したがって帰無仮説は棄却できませんので，塩分摂取の多寡による2つのグループAとBでは，収縮期血圧の母平均値に有意差があるとは言えないことになります。

標本平均での5 mmHgの差は小さいとは言えないと思いますが，検定結果から有意差は認められませんでした。生活習慣病などの研究や食事指導などを行なう場合，ちょっと困った結果になったと言えるかもしれません。この問題は最後に考えたいと思います。

等分散の検定

前述のt検定は，2グループの母分散が等しいと仮定できる場合の計算方法でした。実際には，2つの母分散が等しいと仮定できるとは限りません。そのような場合，2つの母分散が等しいと考えてよいかを検定するための方法があります。いわゆる「等分散の検定」と呼ばれる検定方法を用いて，この仮定をしてもよいかを検討するのが一般的です。等分散の検定での帰無仮説は，「2つのグループの母分散は等しい，$\sigma_x^2=\sigma_y^2$」です。

図5　F分布の自由度と形状



等分散の検定の計算は極めて簡単で，2つの不偏分散の比を計算するだけです。すなわち，

$$F_0=\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$$

を求めます。

上記の F_0 は，自由度$(m-1,\ n-1)$のF分布に従います。また，F_0 の逆数である $1/F_0$ は，自由度$(n-1,\ m-1)$のF分布に従います。

F分布は第1自由度と第2自由度の2つの数値により，その分布の形状が大きく異なります(**図5**)。等分散の検定の方法は，両側 2α の有意水準で検定を行なう場合，自由度$(v_1,\ v_2)$のF分布の上側 α 点を $F_\alpha(v_1,\ v_2)$ とすると，

$\hat{\sigma}_x^2>\hat{\sigma}_y^2$ の場合，$F_0>F_\alpha(m-1,\ n-1)$

または，

$\hat{\sigma}_x^2<\hat{\sigma}_y^2$ の場合，$1/F_0>F_\alpha(n-1,\ m-1)$

ならば仮説を棄却

とすればよいことになります。

実際の検定では，$F_0>1$ となるように分子，分母を決めて計算し，自由度$(m-1,\ n-1)$のF分布での F_0 の上側確率(p値)を求め，p値が0.025以下ならば，5%水準で帰無仮説を棄却すればよいでしょう。

帰無仮説が棄却された場合，2グループの分散は等しくない，すなわち不等分散とすればよく，それ以外の場合には帰無仮説を消極的に採択し，

「等分散」を仮定します。

普通の検定では，仮説が棄却できない場合には，態度保留とするのが一般的ですが，等分散の検定では，意志決定として消極的に仮説を受け入れます。これは，気分的には気持ちの悪い検定方法ですが，強いて不等分散と仮定するための根拠もないという理由によります。

先の例題について，「等分散の検定」を行なうと，

$$F_0 = \frac{15^2}{10^2} = \frac{225}{100} = 2.25$$

となります。この統計量 F_0 は自由度(39，49)のF分布に従うので，p値を求めてみますと，上側 $p = 0.0038$(両側 $p = 0.0075$)となります。両側確率は5%より小さいので，等分散の仮説は棄却され，2グループの母分散には有意差があることになります。したがって，2つのグループの母分散は不等分散と考えられますので，前述のt検定は使えないことになります。

Welchの方法（不等分散の場合の母平均値の差の検定）

2グループの母分散が等しくない場合に，2つの母平均値の差を検定する方法は，「Welch(ウェルチ)の方法」と呼ばれています。理論的に考えると，2つのグループの母分散が等しいことを証明するのは，極めて難しいことがわかると思います。したがって，等分散の検定をしなくても，ここで説明するWelchの方法を検定に常に使用するという立場もありうるでしょう。

さて，求めるべき検定統計量は，

$$t_0 = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\hat{\sigma}_x^2}{m} + \frac{\hat{\sigma}_y^2}{n}}}$$

ですが，自由度 v は，

$$v = \frac{\left(\frac{\hat{\sigma}_x^2}{m} + \frac{\hat{\sigma}_y^2}{n}\right)^2}{\frac{\hat{\sigma}_x^4}{m^2(m-1)} + \frac{\hat{\sigma}_y^4}{n^2(n-1)}}$$

となります。

上記の t_0 はt分布に従いますが，自由度 v は必ずしも整数となるとは限らず，小数になることがあります。しかし，実際の検定では自由度が10以上の場合には，小数部分を切り捨てて，整数部分のみを用いればよいでしょう。自由度が10未満の場合には，「小数自由度のt分布表」のような特別な数値表が検定に必要になります。しかし，実際の研究でそのように小さな自由度になるような検定を行なうような研究デザインは，普通はありえないのではないかと思います。

検定方法は，等分散を仮定した場合と全く同様に行ないます。

前述の例題を，ウェルチの検定で行なってみます。まず，検定のための統計量は，

$$t_0 = \frac{|115 - 110|}{\sqrt{\frac{15^2}{40} + \frac{10^2}{50}}} = 1.811$$

となります。等分散を仮定した場合のt検定のt値は1.891でしたので，それよりもやや小さくなっています。

また自由度 v は，

$$t_0 = \frac{\left(\frac{15^2}{40} + \frac{10^2}{50}\right)^2}{\frac{15^4}{40^2 \times (40-1)} + \frac{10^4}{50^2 \times (50-1)}} = 65.1$$

となり，等分散を仮定した場合は88ですので，それよりも幾分か自由度が小さくなっていることがわかります。このp値を求めると，$p = 0.075$，となり5%より大きいので，同様に有意差は認められません。

生活習慣病などの研究を行なっている場合，検

定結果が有意でないと，実際に論文を作成する場合には学術誌への投稿に問題が生じます。また統計学的な有意差がない結果では，通常は投稿しても採択されないのが普通です。一般に，研究を開始する前に，比較する2グループ間の母平均値の差が，検定で有意になるような標本数を求めて，それ以上の標本数で検定を行なう必要があります。したがって，研究デザインとして，どの程度の標本数を設定すればよいかという問題は極めて重要です。

有意となる標本数を求める

前回説明したように，検定が有意になるかどうかは標本数に大きく影響されます。したがって，有意な研究結果を得たければ，研究を開始する前に適切な標本数を求めておく必要があります。2グループの母平均値の差の検定が有意となるように，2グループの標本数を決めるためには以下の情報が必要となります。

(1)検定のための第1種の誤り 2α(有意水準)
(2)検定のための第2種の誤り β，または検出力 $1-\beta$
(3)2グループの母平均値の推定値，または母平均値の差の推定値 Δ
(4)2グループの母分散の推定値 σ^2

通常は，2α は0.05(5％)とし，β は 2α の4倍とします。したがって，有意水準が5％ならば，検出力は80％に設定するのが普通です。あと必要な情報は，母平均値の差の推定値 Δ とその分散の推定値 σ^2 ということになります。

Δ は「効果量 effect size」と呼ばれ，また σ は「変動度 variability」と呼ばれることがあります。2つの変量の比，Δ/σ は「標準化効果量 standardized effect size」と呼ばれることがあります。この値がわからないと，適切な標本数を求めることはできません。

さて，標準正規分布の上側 α 点の値を z_α とします。2グループの全体の標本数を N とし，2グループのそれぞれの標本数を $n_1=pN$，$n_2=qN$，$p+q=1$，とします。このとき，全体の標本数は，

$$N=(1/p+1/q)\times(z_\alpha+z_\beta)^2\times\frac{\sigma^2}{\Delta^2}+\frac{z_\alpha^2}{4}$$

によって近似的に求められます。

有意水準を両側5％，検出力を80％とすると，標準正規分布の上側2.5％点は $z_{0.025}=1.96$，また上側20％点は $z_{0.2}=0.842$ となります。2グループの標本数 n を等しいものとすると，

$$n=15.7\times\frac{\sigma^2}{\Delta^2}+0.5$$

と簡単に近似計算することができます。

表1に示した平均値や分散が正しいものと仮定して，有意水準を両側5％，検出力を80％とした場合に必要となる標本数を求めてみましょう。

まず，2つの母平均値の差の推定値は，$\Delta=115-110=5$(mmHg)となります。また，分散の計算の方法は，等分散を仮定して，プールした分散を計算するのが簡単でしょう。すなわち，t検定の計算式の分母に出てくる式を使います。

$$\hat{\sigma}^2=\frac{(m-1)\hat{\sigma}_x^2+(n-1)\hat{\sigma}_y^2}{m+n-2}=\frac{39\times15^2+49\times10^2}{40+50-2}=155.40$$

となりますので，標本数の計算式に代入すると，

$$n=15.7\times\frac{155.40}{5^2}+0.5=98.1$$

と求められます。この結果から，表1の標本数の2倍程度が必要となります。これは，等分散の検定結果でp値が6.2％であったことを考えると多すぎるように感じるかもしれません。しかし，実際には新規に調査を実施した場合，偶然による変動で，より大きな差になることもあれば，より小

さな差になることもありえます。仮定した母平均値の差や分散が本当であれば，そのような調査による変動も考慮して，検出力80%の場合には，そのような調査を100回繰り返した場合に，80回は有意な結果を得ることができることを示しています。

1回だけの実際の調査結果に基づく検定では，対立仮説は帰無仮説を単純に否定したものなので，検出力が50%の場合に対応しています。

検出力を50%としますと，

$$n = 7.68 \times \frac{155.40}{5^2} + 0.5 = 48.2$$

となり，表1に示した結果が得られたのであれば，A群がほんの少し標本数が多ければ，有意な結果が得られたことになります。しかし，これは後付けの分析ですので，あと10例多く調査していれば有意になったのに，と考えても仕方がないことです。研究計画の開始時点から，適切な標本数を決めるのが，正しい研究デザインのあり方です。

文献

- 髙木廣文，林邦彦(2006)．エビデンスのための看護研究の読み方・進め方．中山書店．pp.45-51.
- 髙木廣文(2009)．ナースのための統計学，第2版．医学書院，pp.50-51.
- 髙木廣文(2014)．統計学のキー・ポイント―「検定」に焦点を当てて　第1回検定の基本的な考え方．看護研究，47(1)，52-58.
- 髙木廣文(2014)．統計学のキー・ポイント―「検定」に焦点を当てて　第2回検定と標本数の関係．看護研究，47(2)，150-154.
- Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G., & Newman, T.B.,(2007). *Designing Clinical Research*(3rd ed.). Lippincott Williams & Wilkins.／木原雅子，木原正博訳(2009)．医学的研究のデザイン，第3版．メディカル・サイエンス・インターナショナル，pp.67-97, 150-154.

たかぎひろふみ●東邦大学看護学部
〒143-015　東京都大田区大森西4-16-20

NURSING BOOK INFORMATION ―― 医学書院

医療におけるヒューマンエラー 第2版
なぜ間違える どう防ぐ

河野龍太郎

●B5　頁200　2014年
定価：本体2,800円＋税
[ISBN978-4-260-01937-8]

なぜ医療事故は減らないのか。それは、事故の見方・考え方が間違っているから。
本書では事故の構造、ヒューマンエラー発生のメカニズム、人間に頼らない対策の立て方を、心理学とヒューマンファクター工学をベースに解説。さらに人間の行動モデルからエラー行動を分析するImSAFERを紹介する。
医療事故のリスク低減のために、事故の見方・考え方を変える1冊。