

統計検定を理解せずに使っている人のために II

池田郁男

東北大学大学院農学研究科

今回は、母集団、標本、母分散、母標準偏差、標本分散、標本標準偏差、不偏分散、不偏標準偏差について主に記述した。これらの理解が曖昧な場合は、再度前回をお読みいただきたい。なお、図番号は、前回からの通し番号である。まず、図の内容をざっと眺めてから文章をお読みいただきたい。

パラメトリック検定（図15）

パラメトリック検定は母集団のデータが正規分布していると仮定する。平均値や分散などのパラメータを使う

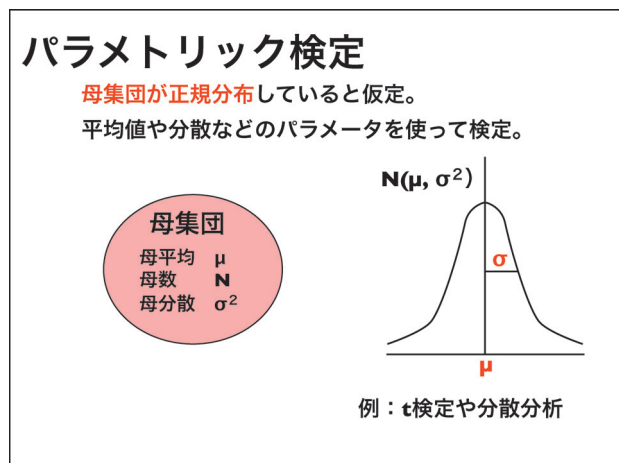


図15 ■ パラメトリック検定

ことからこの名称がある。パラメトリック検定の例としては、Studentの t 検定や分散分析などがある。前回と同様、母集団の総数 N 、母平均は μ 、母分散は σ^2 と定義する。母平均 μ 、母分散 σ^2 の正規分布を $N(\mu, \sigma^2)$ と記述することとする。

パラメトリック検定の基本：母集団から取り出した n 個の標本の平均値の分布を考える（図16）

かなり込み入った話になるので、注意して読んでほしい。じっくり読まないで難しいが、ここを理解できないと標準誤差SEと t 検定は理解できない。ここでは、母

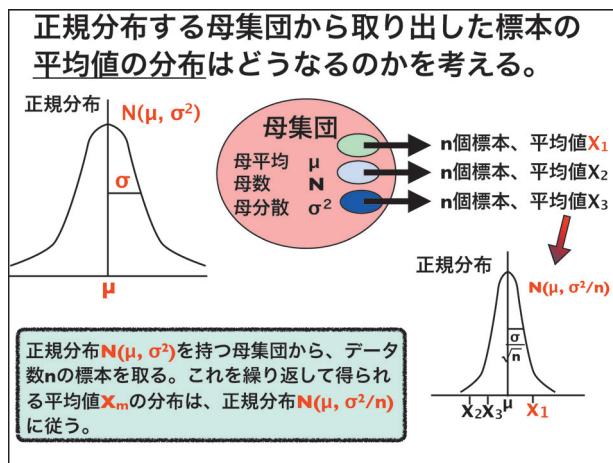


図16 ■ パラメトリック検定の基本

集団から取り出した n 個の標本の平均値の分布を考える。標本データの分布ではないので注意してほしい。

図16右上のように、正規分布している母集団から n 個の標本を取り出し平均値 X_1 を計算する。これを何度も繰り返すとそれぞれ平均値 X_1, X_2, X_3, \dots が得られる。このたくさん得られた平均値を一つの別の母集団と考えて分布を調べると、最初の母集団と同様に正規分布することが知られている。しかし、最初の母集団の正規分布 $N(\mu, \sigma^2)$ とはバラツキ方、すなわち σ^2 が異なり、 $1/n$ だけ小さいバラツキ方 σ^2/n になることがすでにわかっている。つまり、 $N(\mu, \sigma^2/n)$ の正規分布となる（図16右下グラフ）。すなわち、母標準偏差は σ/\sqrt{n} である*1。ここで例として、ある一つの平均値 X_1 の位置を右下の正規分布グラフに赤字で示している。

正規分布を標準正規分布に変換する（図17）

正規分布は μ と σ^2 が違えば分布が異なるので、いろいろな母集団でいちいち異なる正規分布を用いて考えるのは面倒である。そこで、母平均 μ が0（ゼロ）で、 σ^2 が 1^2 の正規分布を標準正規分布と定義し、それぞれの異なる正規分布を標準正規分布に変換して考えることに決められている。そうすれば、どのような平均値と分布の母集団でも標準正規分布で考えることができるのである。図17の右下のグラフが標準正規分布である*2。

そこで、先ほどの平均値 X_1 （図17左上グラフ）を標準正規分布の値に変換することとする。まず、標準正規分布では母平均はゼロであるから、母平均 μ をゼロとするためには、 μ で引いてやればよい。すなわち、 $(X_1 - \mu)$ である。一方、標準正規分布では母標準偏差は1であるから、 σ/\sqrt{n} を1に変換するためには、 σ/\sqrt{n} で割ってやればよい。すなわち、平均値 X_1 は、標準正規分布の値に変換すると（変換した値を Z_1 とすると）、

$$Z_1 = \frac{(X_1 - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

で表すことができる（図17中央式）。そうすると、 X_1 は図17右下の標準正規分布グラフ中の Z_1 に変換される。

*1 実験動物の母集団を例にとれば、そのなかからたとえば6匹を標本として取り出して、あるパラメータの平均値を取る。この操作を何度も繰り返して得られる多くの平均値の分布は、母集団データのバラツキよりも小さくなることは、何となく予想できるであろう。

*2 前回にも述べたが、本セミナーのすべての図は筆者が模式的に作成したものであり正確なものではない。

正規分布を標準正規分布にデータ変換する！

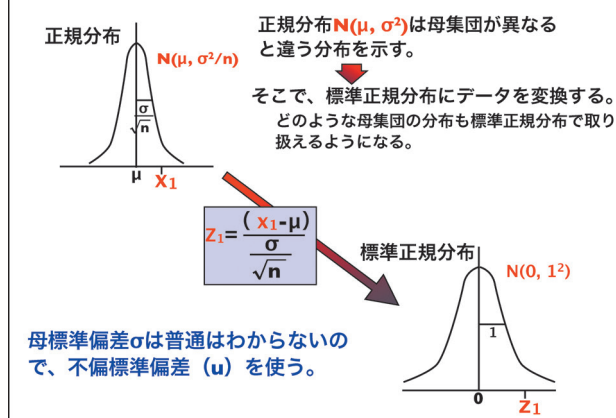


図17 ■ 標準正規分布へのデータ変換

母標準偏差 σ は普通はわからないので、不偏標準偏差 (u) を使う。

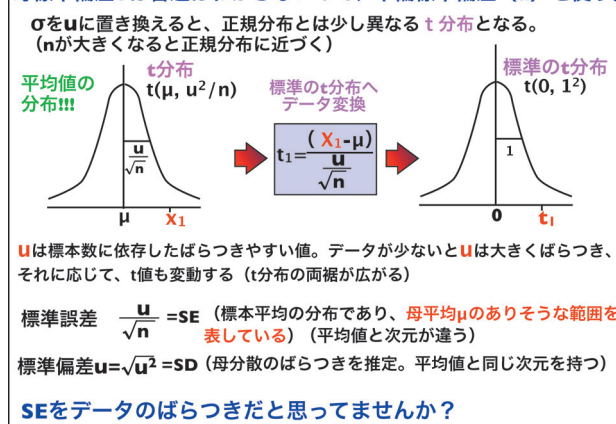


図18 ■ 母標準偏差 σ の不偏標準偏差 u への置き換え

母標準偏差 σ はわからないので、不偏標準偏差 u に置き換える（図18）

ここまで、正規分布している母集団から取り出した標本データの平均値の分布は正規分布することを述べてきたが（図17左上、これを $N(\mu, \sigma^2/n)$ とした）、バラツキは σ/\sqrt{n} であった。ここで、母標準偏差 σ は母集団の情報であり、通常われわれは知ることができない値である。前回記述したように、標本データから母標準偏差 σ を推定する値として得られるのが不偏標準偏差 u である。そこで、母標準偏差 σ を不偏標準偏差 u に置き換える（図18左上図）。 σ を u に置き換えると、正規分布とは少し異なる分布になることがすでにわかっている。この分布は t 分布と呼ばれており、正規分布 $N(\mu, \sigma^2/n)$ に対して、 $t(\mu, u^2/n)$ で表す（図18左上グラフ）。

t 分布も正規分布同様、 μ と u^2/n の違いで異なる分布を示すので、標準 t 分布、すなわち、 μ がゼロで u^2/n が 1^2 の t 分布 ($t(0, 1^2)$) を用いると、あらゆる t 分布に対応

することができる (図18右上グラフ)。

ここで、平均値 X_1 を標準 t 分布に当てはめるため、標準正規分布の場合と同様のデータ変換を行う (図18中央上)。標準正規分布の場合と同様に計算すると、平均値 X_1 は以下の式により t_1 へと変換される (図18の中央式および右上グラフ)。

$$t_1 = \frac{(X_1 - \mu)}{\frac{u}{\sqrt{n}}}$$

不偏標準偏差 u は母標準偏差 σ を推定している。しかし、標本数が少ないと、標本データを取り出して u を計算することを何度も繰り返した場合、 u は σ に比較してバラツキが大きくなることが予想される。一方、標本数をたくさん取ると、 u は σ に対してあまりバラつかないと考えられる。すなわち u は標本数が少ないと大きくばらつくことから、正規分布に比べて t 分布の裾野が相対的に広がることとなる。一方、標本数が多くなると、 u のバラツキは σ に近くなるので、 t 分布の裾野が狭くなり、 u が σ と一致すれば正規分布になるのである。

この項では標本データではなく、平均値の分布やバラツキを考えてきたことから、上の式の子母である $\frac{u}{\sqrt{n}}$ は平均値がどのようなバラツキ方をしているかを示しており、標準誤差 (standard error; SE) と呼ばれる。したがって、SE は母平均 μ がどこにあるのか、すなわち、母平均のありそうな範囲を表していることになる。SE を標本データのバラツキの一種と思っている方が多いかもしれないが、それは間違いであり、SE には標本データのバラツキの意味は全くない！ そもそも SE の単位 (次元) は平均値のそれとは異なっている。 t_1 の式の子母が SE であるから、上の式は以下のようにも書ける。

$$t_1 = \frac{(X_1 - \mu)}{SE}$$

平均値 \pm SE は何を意味しているのか？ (図19)

多くの読者は、データ表示に平均値 \pm SE を用いているはずである。標本データの平均値は母平均 μ を推定する値と位置づけられる。そして、SE は母平均 μ のありそうな範囲を示しており、図19にあるように、平均値 \pm SE は母平均 μ が68%存在することが期待される範囲である。平均値 \pm 2SE では、96%存在することが期待される範囲を示す。SE の式を見ると、SE は n が大きくなればなるほど、小さくなることがわかる (図19右上)。これは、 n が増えれば増えるほど、母平均 μ の存在する範囲が絞られることを意味している。 n が増

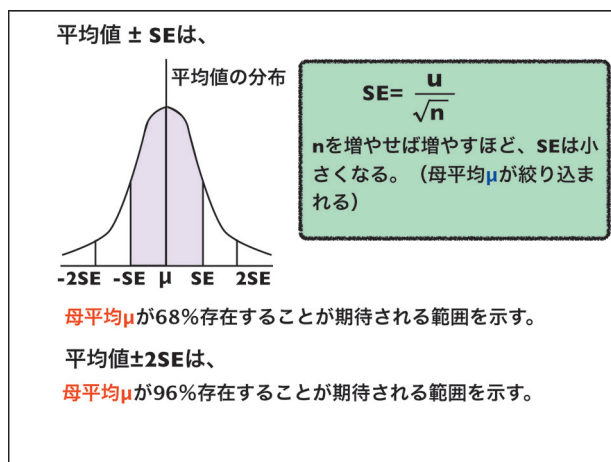


図19 ■ 平均値 \pm SE の意味

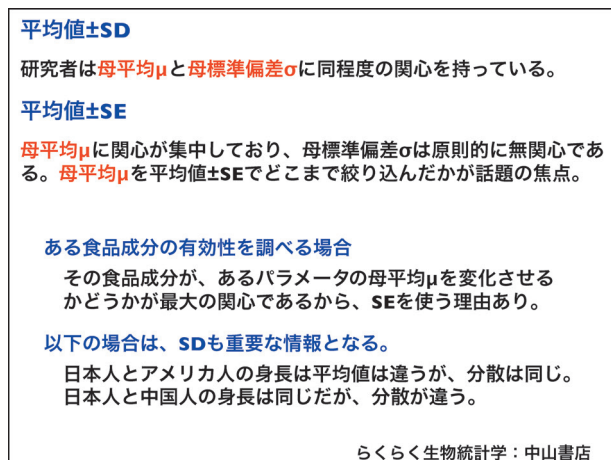


図20 ■ 平均値 \pm SD と平均値 \pm SE の違い

えて、母集団の総数 N に近くなればなるほど、 μ が絞られていくことは感覚的に理解できるのではないだろうか。

平均値 \pm SD と平均値 \pm SE はどう違うのか？ (図20)

平均値 \pm SD については、前回の図13でも記述したように、研究者は母平均 μ と母標準偏差 σ の両方に興味があるはずである。それでは、平均値 \pm SE ではどうだろうか？ 図20に示すように、平均値 \pm SE はあくまでも母平均 μ のありそうな範囲を示していることから、母平均 μ に関心が集中しているのであり、母標準偏差 σ には原則的に関心がないことになる。母平均 μ を SE によりどこまで絞り込んだかが話題の焦点である (この点は『らくらく生物統計学』⁽¹⁾ に詳しく述べられている)。実例を図20下側に示している。ある機能性成分が母平均 μ を変化させるかどうかに関心があるのであれ

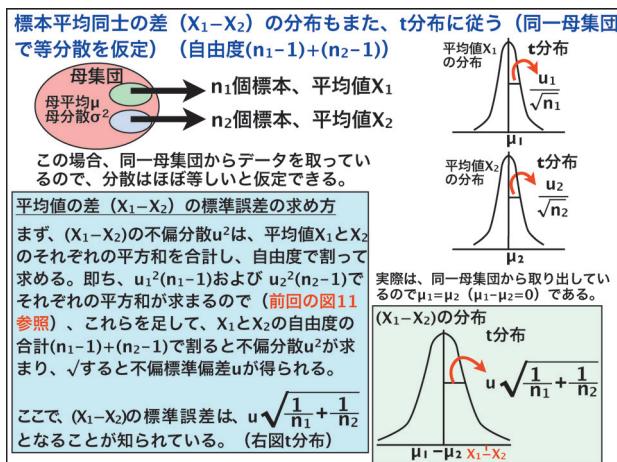


図21 ■ Student の t 検定

ば、SEを使えばよい。おそらく、食品成分の有効性を調べる研究を行っている研究者のほとんどは、平均値にしか興味がないのではないだろうか？ その場合、SEを使っても何らの問題もない。図20最下に示すような例では、SDが適している。しかし、SEとSDを上記の観点から使い分けている研究者がどれくらい存在するであろうか？ 幸いにも、投稿論文査読者もSEとSDの使い分けを理解していない場合が多いことから、SE、SDのどちらを使用しても論文査読の可否に影響しない現実は存在するが…。

Studentの t 検定 (図21)

Studentの t 検定を知らない人はいないであらう。2群のパラメトリック検定であり、 t 分布を用いることからその名がある。詳しくは後述するが、 t 検定には対応のない独立2群の検定 (unpaired t test) と対応のある関連2群の検定 (paired t test) がある。ここでは、unpaired t testを例として詳述する。

ある正規分布する母集団から標本データを n_1 および n_2 個取り出し、平均値 X_1 と X_2 を求める。母集団から n 個の標本を取って平均値 X を得たとして、これを無限に繰り返したときの平均値の分布は、不偏標準偏差 u を用いると t 分布することはすでに述べた (図21右上グラフ)。ここで、平均値 X_1 と X_2 に差があるかどうかを考える。そこで、平均値の差 ($X_1 - X_2$) を求めることとして、この操作を何度も繰り返すと、平均値の差もまた正規分布するのである (不偏標準偏差 u を用いると t 分布する。図21右下グラフ)。この時の不偏分散、不偏標準偏差、標準誤差の求め方は、図21の左下枠内に記述しているが、難しければ読み飛ばしてよい*3。

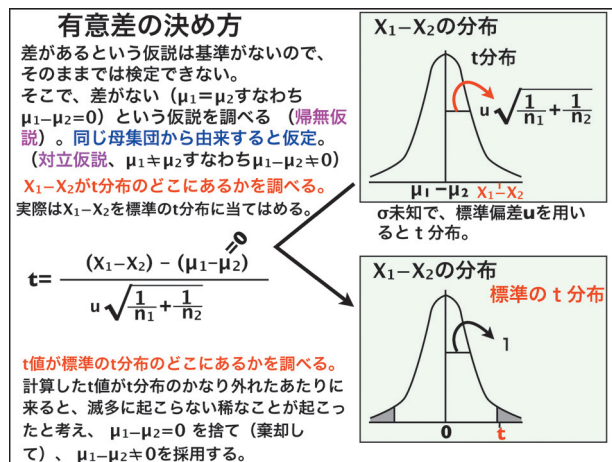


図22 ■ 有意差の決め方

Studentの t 検定 有意差の決め方 (図22)

いま、ある二つの平均値の差 ($X_1 - X_2$) (仮に $X_1 > X_2$ とし記述する) が図22右上グラフの赤字で示す位置にあるとする。 X_1 と X_2 に差があるという仮説は基準がないのでそのままでは検定できない (図22左上)。そこで、統計検定では差がないという仮説、すなわち、母平均 μ_1 と μ_2 に差がない $\mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$) という仮説を考えることになっている。このような仮説は帰無仮説と呼ばれる。これに対して、 $\mu_1 - \mu_2 \neq 0$ という仮説 (すなわち $\mu_1 \neq \mu_2$) は対立仮説と呼ばれる。すでに、2つの平均値の差 ($X_1 - X_2$) の分布は不偏標準偏差 u を用いれば t 分布することは述べたとおりである。 $X_1 > X_2$ と仮定して、もし X_1 と X_2 に大きな差があれば、 $X_1 - X_2$ は t 分布の中央の $\mu_1 - \mu_2$ からかなり離れるはずである (同一母集団から考えてきたので本来 $\mu_1 = \mu_2$ すなわち $\mu_1 - \mu_2 = 0$ である)。

標準正規分布の項で述べたことと同様に、 t 分布は標準の t 分布に変換したほうが便利である。そこで、 $(\mu_1 - \mu_2)$ をゼロに変換するためには、 $(X_1 - X_2)$ から $(\mu_1 - \mu_2)$ を差し引けばよい (図22左中央式)。また、標準誤差を1にするためには、

$$u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

*3 この標準誤差の式の理解が難しい場合は、実際に、ある正規分布する母集団を作成し、そこから標本データ n_1 および n_2 個を取りだし、その平均値 X_1 と X_2 の差 ($X_1 - X_2$) を求める。これを何度も繰り返し、不偏標準偏差 u を用いて t 分布を作成すると、その標準誤差は $u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ になってしまうと考えればよい。

で割り算すればよい。そうすると以下の式が出現し、 $(X_1 - X_2)$ は t に変換される (図22左中央式)。

$$t = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ここで、 μ_1 と μ_2 は差がない、すなわち $\mu_1 - \mu_2 = 0$ と仮定したので、上記の式は以下のようにになる。

$$t = \frac{(X_1 - X_2)}{u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

標準 t 分布における t 値の位置を図22右下グラフに赤字で示している。ここで、 t 値が t 分布のどの位置にあるかを考える。 X_1 と X_2 はもともと同一母集団から取り出した標本の平均値から考えてきたので、その差 $(X_1 - X_2)$ はゼロの回りに分布するはずである。もし $(X_1 - X_2)$ がゼロから遠く離れた値になったとすると、それは起こりうることはあるが稀にしか起こらない。そこで、 t 値が t 分布のかなり外れたあたりにくると、それは減多に起こらない稀なことが起こったと考えることにして、 $\mu_1 - \mu_2 = 0$ という帰無仮説を棄て (棄却して)、対立仮説である $\mu_1 - \mu_2 \neq 0$ を採用することにするというのが、 t 検定における「差がある」という決め方となっている。言い方を変えると、「差がない」ことから考えていくと、「差がない」と言うにはあまりにも外れているから「差がある」ことにしようという考え方である。

減多に起こらない稀な確率の決め方 (図23)

それでは、減多に起こらない稀なこと (稀な確率) をどの程度にとればよいのであろうか? 図23の上に書いているように、一般的に t 分布全体の両端あるいは片端の5%または1% ($p = 0.05$ または $p = 0.01$) とすることに決められている。これを有意水準と呼ぶ。平均値 X_1 と X_2 でどちらが大きな値になるのかわからない場合は、 $(X_1 - X_2)$ はプラスになるかマイナスになるかわからないので、 t 分布の左右両端のそれぞれ2.5%ずつあるいは0.5%ずつに t 値が入ってくると、稀であると判定するのである。しかし、減多に起こらない稀なことといっても、もともと差がないことから考えてきたのであり、差があると断言するのは危険である。そこで、 p は危険率とも呼ばれる。記述する際には、「有意水準 (危険率) 5% (1%) 以下で有意差あり」と書く。

標準誤差 $u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ の式を見ると、 n が大きくなれ

減多に起こらない稀な確率をどの程度にとるか?

一般的に、5%または1% ($p = 0.05, 0.01$) とする (有意水準)。 t 分布の左右2.5%、0.5%に入ってくると、稀であると判定する。しかし、稀でも起こりうることであるから、帰無仮説 ($\mu_1 - \mu_2 = 0$) は正しいのかもしれない。従って、 p を危険率とも呼ぶ。

有意水準 (危険率) 5%以下で
有意差ありと記述する。

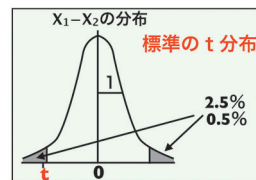


図23 ■ 有意水準 (危険率)

ばなるほど、標準誤差の値が小さくなり、 $X_1 > X_2$ と仮定すると、 t 値は大きくなることがわかる。 t 値が大きくなると、 t 分布の端のほうに寄るので、有意となる可能性が高まる。したがって、 t 検定では、標本データ数 n を増やせば有意差は出やすい。微妙な差しかない、あるいは、バラツキが大きいことが最初からわかっているパラメータの場合に、標本データ数を増やすことは有益な手段である。

図23左下に書いているように、危険率5%は全く同じ実験を20回行ったときに1回は有意差がないのに有意差があると判定される可能性のある確率である。それは1回目に起こるかもしれないし、20回目に起こるのかもしれない。実験研究では、普通は1回しか実験しないので、もしも1回目に起これば、本当は有意差がないのに有意差ありと判定されることになる。研究者はこのことを念頭に入れておかなければならない。1回の実験だけで、実験データを論文化する危うさはここにある。

対立仮説の立て方で検定結果は異なる! (図24)

これまで2群の実験で、あるパラメータを測定するとき、帰無仮説として母平均 μ_1 と μ_2 は差がない、すなわち $\mu_1 = \mu_2$ とした。 μ_1 と μ_2 はどちらが大きな値になるかは、通常はわからない。そこで、対立仮説は $\mu_1 \neq \mu_2$ すなわち $\mu_1 - \mu_2 \neq 0$ とした。標本データの平均値 X_1 と X_2 で考えると、 $(X_1 - X_2)$ はプラスかマイナスかわからない。したがって、 t 分布の両端の2.5%ずつの範囲に入れば、危険率5%で有意差ありとする。このような検定を両側検定と呼ぶ (図24)。未知の機能性成分の影響を調べる場合は、両側検定とすべきである。しかし、前もってどちらかが大きいという十分な情報がある場合

対立仮説の立て方は重要

$\mu_1 - \mu_2 \neq 0$ この場合どちらが大きいかわからないので、両側検定 ($\mu_1 \neq \mu_2$) となる。

$\mu_1 > \mu_2$ どちらかが大きい (小さい) と立てると片側検定となる。
 $\mu_1 < \mu_2$

未知の成分の機能性を調べる場合、両側検定が妥当。

前もって、 $\mu_1 > \mu_2$ あるいは $\mu_1 < \mu_2$ であるという情報がある場合は、片側検定もありうる。

例えば、植物ステロール含有機能性食品を開発する場合

植物ステロールの血漿コレステロール濃度低下作用は十分立証されているので、「植物ステロール摂取群で血漿コレステロール濃度は低い」と対立仮説を立てることが可能 → 片側検定となる！
 片側検定のほうが、有意差は出やすい。

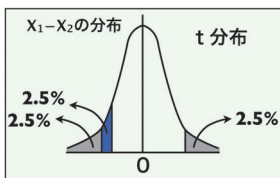


図24 ■ 対立仮説の立て方

は、 $\mu_1 > \mu_2$ あるいは $\mu_1 < \mu_2$ と対立仮説を立てることも可能である。具体的な例を図24左下に記述している。もし、 $\mu_1 < \mu_2$ と対立仮説を立てることができれば、標本データ ($X_1 - X_2$) はマイナスとなることが最初から期待できることから、図24の t 分布グラフのマイナス側だけを考えればよいので、マイナス側に5%を設定できる。このような検定を片側検定と呼ぶ。片側検定では両側検定よりも有意差がでやすくなることがわかる (図24右下図)。片側検定を使えるのは十分な証明のある場合であるが、一般的研究では十分な立証がないから研究を行っている場合が多いので、最新の研究において片側検定を用いることはほとんどない。なお、ある試験物質で1, 2回実験を行ってみて両側検定で有意な影響が得られたからといって、それ以降の少し条件を変えた実験において、片側検定を用いて有意差を得やすくとした勝手な判断は許されることではない。最新研究の原著論文において、両側検定か片側検定かの記述はしないが、常識的には両側検定を行っているはずである。

2群のパラメトリック検定の流れ (図25)

これまで2群のパラメトリック検定として、対応のない独立2群の検定 unpaired t test を述べてきたが、図21で述べたとおり、 t 検定には対応のある関連2群の検定 (paired t test) がある (図25)。unpaired t test では等分散である必要があるので、まず標本データの等分散性の検定を行い、等分散であれば unpaired t test を行う。等分散でない場合は Welch の検定を行う (後述する)。統計アプリでは、等分散性の検定、および、unpaired t test と Welch 検定の両者の結果が同時に表示される場合があり、意味を知らないとどちらの検定結果を

統計検定の流れ

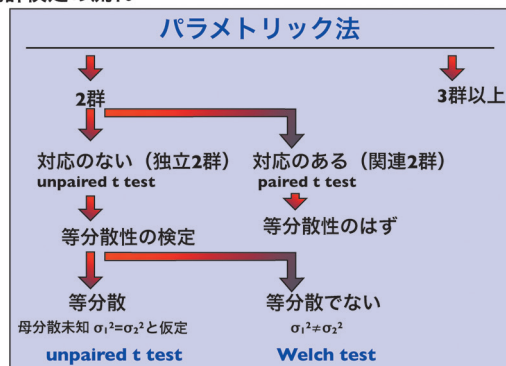


図25 ■ 2群のパラメトリック検定の流れ

対応のない独立2群の差の検定 (unpaired t test)

例：独立した2群のネズミで、対照食と試験食を与えて比較する場合。

(対照食と試験食ではネズミが同じではない)

対応のある関連2群の差の検定 (paired t test)

例：同じネズミで、食品成分投与前と投与後の血圧を比較する場合。

(投与前と投与後は同じネズミ)

paired t testを行うべき場合に、unpaired t testを行うと、有意差が出にくい！

図26 ■ unpaired と paired t testの違い

採用してよいかわからないはずである。そこで、 t 検定の結果だけを見て、Welchの検定を無視するといったことが起こるので要注意である。

unpaired と paired t testはどう違うのか？ (図26)

わかりやすい例を図26に記述した。2群の別々のネズミで試験すると、ネズミは対応していないので独立した2群の検定 (unpaired t test) である。同じネズミで投与前と投与後の比較をする場合は、対応があるので paired t test となる。

paired t testの考え方は、これまで述べてきた unpaired と基本的には同じであるので、ここでは詳細には述べない。簡単に記すと、paired t testの場合には、同じネズミからのデータなので、投与後のデータ x_1 と投与前のデータ x_2 を個々のネズミで比較できる。そこで、仮に6匹のネズミがいるとして、それぞれのネズミで d

$=x_1 - x_2$ を計算する。投与前と投与後で大きな変化がなければ d はゼロの前後に分布することになり、何らかの影響があれば、 d はゼロから離れてくる。パラメトリック検定であるからもとの母集団は正規分布と仮定しており、 d もまた正規分布する（不偏標準偏差 u を用いれば t 分布する）ことが知られている。そこで、6匹の d の平均値と不偏標準偏差を計算して、母平均 $\mu=0$ の標準 t 分布に当てはめ、 d の平均値が t 分布の両端の2.5%に入れば、 $p<0.05$ で差があると考えるのである（図23参照）。

対応のある関連2群の実験は、同じネズミで比較し合うので、全く異なるネズミで比較する独立2群の実験よりも信頼性の高いデータが得られる^{*4}。したがって、pairedのほうがunpaired t testよりも有意差が得られやすい。paired t testで行うべきところをunpairedで行うと、せっかくの有意差を見落とすことになるので注意が必要である。

等分散性の検定（F検定）とは何か？（図27）

t 検定は、もともとと同じ母集団から取り出した標本データから考えられたものであり、分散はおおよそ同じ（等分散）であるはずである。したがって、等分散でない t 検定を使えない。そこで等分散かどうかを判断するために、等分散性の検定が行われる。実際には、母分散はわからないので、取り出した標本から計算した不偏分散 u^2 を用いる（前回図11参照）。図27にあるように、母集団から標本を2組取り出し、 u_1^2 と u_2^2 を計算する。等分散かどうかは、 $u_1^2 = u_2^2$ であるかどうかを考えればよい。もし、 $u_1^2 = u_2^2$ であれば、 u_1^2/u_2^2 は1となる。例によって、このような計算操作を何度も繰り返しプロットすると、ある分布が出現する（図27右下グラフ）。分散がほぼ等しければ、 u_1^2/u_2^2 は1を中心に分布し、1から離れるほど少なくなるはずである。この分布は**F分布**と呼ばれる。そこで、もし u_1^2/u_2^2 の値がグラフの F_1 あたりにあるとすると等分散であると考え、もし、1から大きく離れた場合（グラフでは F_2 や F_3 ）、等分散と考えるにはあまりにも離れており、無理があると判定する。この場合も、全体の5%を基準に考えて、F分布の両端それぞれ2.5%に u_1^2/u_2^2 が入った場合、等分散ではないと判断する。

しかし、標本数 n が少ないと不偏分散 u^2 のバラツキが

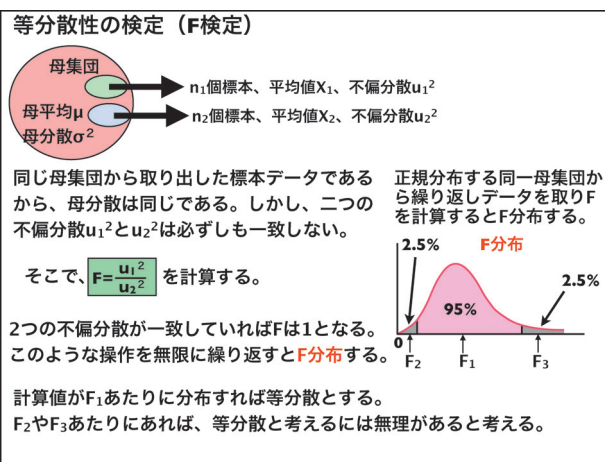


図27 ■ 等分散性の検定

大きくなるため、等分散かどうかの判断が難しくなる。等分散かどうかの判断は、一般的にデータ数が30以上必要とされる。これらのことから、データ数が少ない場合は等分散性の検定はあてにできないのである。統計アプリの検定結果で、 t 検定とWelch検定ともに有意差の評価が同じであれば問題ないが、異なる場合はどちらを採用するのが妥当か考えなければならない。判断に迷う場合は、結局は、等分散性の検定の結果に従って、 t 検定かWelch検定かを選ぶのが良いかもしれない。しかし、意図した結果と異なる場合は悩むことになる^{*5}。こういう場合は、後述するノンパラメトリック検定の結果を考慮することも一つの選択肢である。

等分散でない場合（不等分散）の考え方（図28）

等分散でない場合は、母集団が同一ではなく別々の母集団に由来するデータと考えられる（図28左上）。ここで、別々の母集団由来のデータであっても、平均値の差 $X_1 - X_2$ は正規分布することが知られている。したがって、結局は、図21の等分散の場合と t 検定の考え方は同じになる。しかし、 t の計算式の分母である標準誤差の計算式が以下ようになる。

$$\sqrt{\frac{u_1^2}{n_1} + \frac{u_2^2}{n_2}}$$

図21で同一母集団の場合は、2群の合計の不偏分散 u^2 を計算したが、母集団が異なる場合は合計することでは

^{*4} 同じネズミからのデータなので、母分散は同じと考え、等分散性の検定は行わない（図25）。

^{*5} 検定結果の判断が分かれるデータは明確な結論が得られたとは言いがたいことから、真実がどこにあるのかを見極める必要がある。

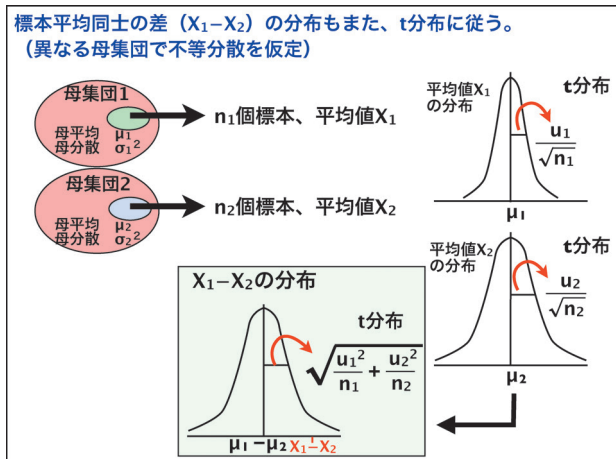


図28 ■ 等分散でない場合（不等分散）の考え方

きないので、別々に計算して足していることがわかる。したがって t の計算式は、以下の式となる。これが、Welchの検定法の基本的考え方である。

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{u_1^2}{n_1} + \frac{u_2^2}{n_2}}}$$

ノンパラメトリック検定 (図29)

パラメトリック検定は、母集団のデータが正規分布していることが前提で組み立てられていることはすでに述べてきた。ノンパラメトリック検定は、母集団の分布は正規分布する必要がなく、平均値や分散などのパラメータは使用しない (図29)。等分散にならない場合や外れ値*6がある場合でも利用することができる。

前回図5で述べたが、データ数が少ない場合は、母集団が正規分布しているかどうかは判然としないため、パラメトリック検定にこだわる必要はない。ノンパラメトリック検定は応用範囲が広い。したがって、筆者は悩ましいデータが得られた場合にパラメトリック検定とノンパラメトリック検定を併用して、結果を考えることにしている。

2群間のノンパラメトリック検定 (図30)

ノンパラメトリック検定でも、2群間および3群以上

*6 ほかの値から大きく外れた値のことである。研究においてはよくお目にかかり、どう処理するか悩む場合が多い。統計検定における外れ値の取り扱いについては、次回に記述する。

ノンパラメトリック検定

母集団の分布によらない検定（正規分布する必要なし）。平均値や分散などのパラメータは使用しない。

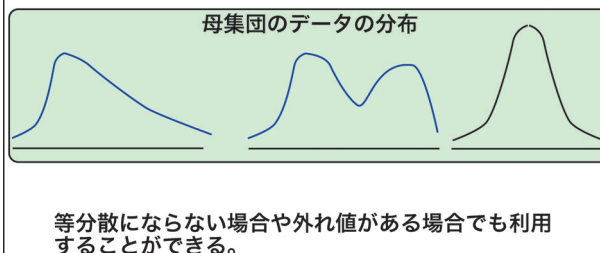


図29 ■ ノンパラメトリック検定

統計検定の流れ

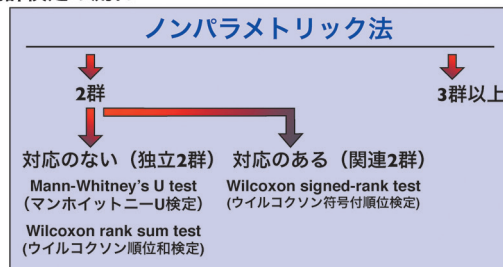


図30 ■ ノンパラメトリック検定の流れ

の検定法が存在する。図30にはよく用いられる2群間の検定法を挙げている。

順位和検定の考え方 (1) (図31)

平均値も分散も使わないで、どのように検定するのであろうか？ ノンパラメトリック検定はいくつかあるが、研究者はそれらすべてを把握する必要はないと筆者は考える。研究者はノンパラメトリック検定がどのような原理で検定しているのかを最低限知っておけば、統計検定に利用したほうが有利かどうかを判断できる。本セミナーでは、よく用いられるウィルコクソンの順位和検定に基づいて説明する。

順位和検定の基本はデータの並べ替えである。図31の①を見て欲しい。A群とB群にそれぞれ6個のデータがあり、それらのデータを小さい値から大きい値に向かって順に並べている。そして、A群、B群両方を一緒

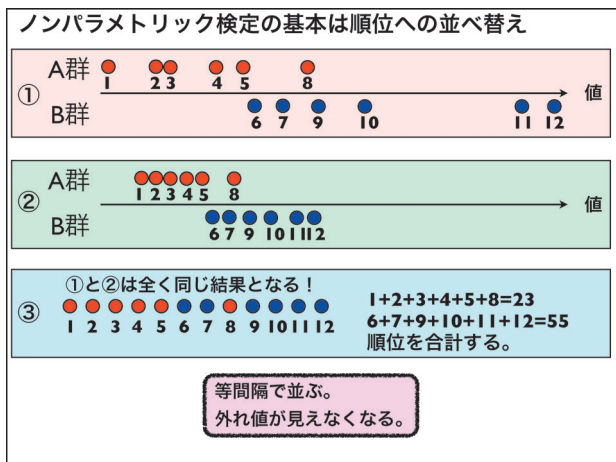


図31 ■ 順位和検定の考え方 (1)

にして、小さいほうから順番に順位をつける。そうすると1から12番までの番号がつく。順位和検定では、この順位の数字を用いて検定する。したがって、平均値も分散も関係なくなることがわかるであろう。たとえば、12番のデータがどんなに右方向へかけ離れた大きな値でも、12番は12番である。

次に、図31の②を見て欲しい。①に比べて、データのバラツキがかなり小さい2群のデータである。この場合も①と同様に順位をつけると、①と②は全く同じ順位になることがわかる。①と②の12個のデータを順位どおりに並べたのが③である。このように、ノンパラメトリック検定では、データのバラツキは関係なくなり、等間隔で並ぶことおよび外れ値が見えなくなることが特徴である (図31下)。

A群(赤)の順位を合計すると23が得られる。B群(青)の順位を合計すると55が得られる。このような値を順位和と呼ぶ。これらの数字が果たして差がある数字かを検証することになる。

順位和検定の考え方 (2) (図32)

ここで、順位和の組み合わせがどれくらいあるかを考える。まず、順位和の最小値は、たとえば、赤が1から6であり合計21、青は自動的に7から12となり、合計57で最大値となる (図32上)。この組み合わせは1通りしかない。ここで、赤が決まれば、青は自動的に決まるので、赤だけを考えることとする。そこで次に、赤の6が7に替わったとすると、1, 2, 3, 4, 5, 7となり合計22であり、この組み合わせも1通りしかない。次に、赤が1, 2, 3, 4, 5, 8となると合計23であるが、合計が23になる組み合わせは、1, 2, 3, 4, 6, 7もありうる。したがって、2

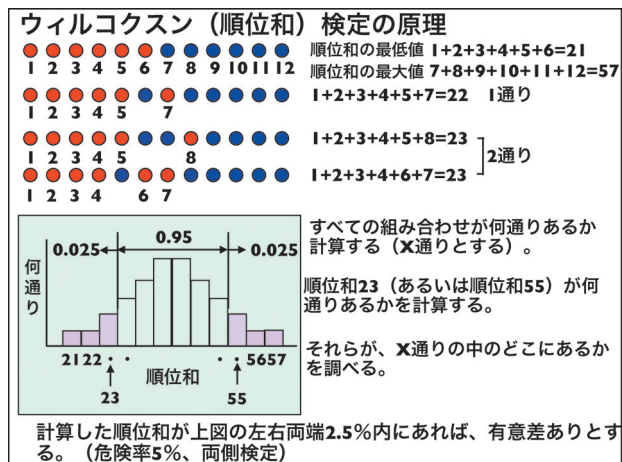


図32 ■ 順位和検定の考え方 (2)

通りあることになる。このような考え方で、合計が24, 25...となる組み合わせが何通りあるかを最大値の57まで計算していく。それぞれの合計値に対して組み合わせが何通りあるかを棒グラフにすると、図32左下のようなグラフが得られる。次に、すべての組み合わせが何通りあるかを合計する (ここではX通りとする)。そこで、図31の③で示した例である順位和23がこのグラフのどこにあるかを考える。

X通りの全体を100として順位和23の2通りが、たとえば、両端の2.5%以内に入った場合、有意差があると判定する (有意水準5%、両側検定の場合)。この判定は、青の場合の順位和55で考えても、グラフでの位置は左右逆側になるが、同じ結果を得ることができる (スライド32左下グラフ)。パラメトリック検定で説明したので詳しくは述べないが、対立仮説として $A > B$ が仮定できれば、片側検定もありうる。これが順位和検定で有意差を求める原理である^{*7}。

ノンパラメトリック検定の利点、欠点 (図33)

ノンパラメトリック検定の原理を知ると、多くの方は大きくばらついたデータでも、そうでなくても検定結果が同じになることに対して、騙されたような気になるのではないだろうか？ 私も最初はそうであった。図33に利点、欠点を示している。母集団が正規分布する場合

^{*7} われわれはデータを表示する場合に、平均値±標準誤差あるいは標準偏差を書く。これらはパラメトリック検定のためのパラメータであり、ノンパラメトリック検定において、これらを表記することは意味がなく、よくよく考えてみると奇妙である。しかし、論文や学会発表では慣習的にこの表記が行われている場合がほとんどである。

ノンパラメトリック検定の利点、欠点

母集団の分布に神経質になる必要がない。

母集団が正規分布の場合、検出力が落ちる。
しかし、その場合でも検出力の低下は意外と少ない。

外れ値がある場合は、有効性が高い可能性あり。
(安易に、棄却検定をしないこと)

パラメトリック検定だけでなく、ノンパラメトリック検定の利用も考えよう！

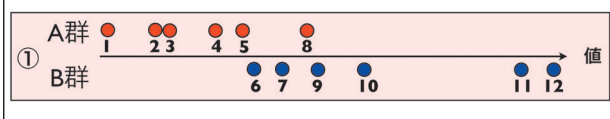


図33 ■ ノンパラメトリック検定の利点、欠点

は多少検出力が落ちるらしいが、経験的に、たとえば、 $n=6$ 程度で t 検定と順位和検定を平行して行ってみると、ほとんどの場合、検定結果は同じになり、分布の偏った場合にのみ異なる結果となる。したがって、研究者はパラメトリックとノンパラメトリック検定を上手く利用することを考えたほうが良い。

また、外れ値があるとき、棄却検定で外すことが可能になる場合もあるが、安易に外すことには問題があるので、外さずにノンパラメトリック検定を行ってみる価値はある。ノンパラメトリック検定でも有意差が得られなければ、あきらめることができる。外れ値の取り扱いについては次回に述べるが、思い悩むデータがあるときは、まずは、パラメトリック検定とノンパラメトリック検定、さらには棄却検定の結果を見てから考えてみてはどうだろうか？

なお、上述の順位和検定の例は、対応のない独立2群での検定であるウイルコクソンの順位和検定であり、 t 検定では unpaired t test に相当する。ノンパラメトリック検定でも、対応のある関連2群での検定 (paired t test に相当) がありうる。ここでは詳細に述べないが、「対応のある」の考え方は、パラメトリックの paired t

testと同様である(図26)。これは、ウイルコクソンの順位和検定に対して、ウイルコクソン符号付順位和検定が一般的である(詳しくは『すぐわかる統計解析』⁽²⁾を参照されたい)。類似した名称であるので間違わないことである。なお、統計アプリによっては、ウイルコクソン順位検定→対応なしか対応ありにチェックといった手順となっている。これは上記の2つの検定法に対応する。

おわりに

次回最終回は、多重比較(3群以上の検定)、一元配置分散分析、二元配置分散分析を中心に記述する。研究論文や学会発表において、3群以上ある試験で多重比較を用いず不適切に t 検定を繰り返している例は減少してきているが、まだ完全にはなくなっていない。また、二元配置分散分析時の群間比較は、正しい方法で行われているほうが少ないように思われ大きく混乱している。ほとんど文献を引用していないが、主に以下の統計書を総合して記述している^(1~4)。

文献

- 1) 足立堅一：“らくらく生物統計学”，中山書店，1998.
- 2) 石村貞夫：“すぐわかる統計解析”，東京図書，1993.
- 3) 市原清志：“バイオサイエンスの統計学”，南江堂，1990.
- 4) 石村貞夫：“分散分析のはなし”，東京図書，1992.

プロフィール



池田 郁男 (Ikuko IKEDA)

＜略歴＞1975年九州大学農学部食糧化学工学科卒業／九州大学大学院農学研究科博士後期課程修了，農学博士，九州大学農学部助手，助教授を経て2005年より東北大学大学院農学研究科教授＜研究テーマ＞食品成分および栄養素の肥満，動脈硬化症予防作用に関する研究，ステロール吸収機構に関する研究＜趣味＞バドミントン