

## データの形

### 1 データの形式

いきなりデータの形といわれても何のことか分からない方も多いであろう。正しく解析を行うにはデータの形を正しく把握する事が第一歩となる。しかし、データの形などという概念は普段の生活ではあまり意識をする事のない物なので、これに関しては以下の説明を読んで頂き、このような物であると覚えてしまうのが一番である。

我々が手にするデータには様々な形のデータがある。臨床検査分野では、定量検査、定性検査など多種のデータが扱われるが、データの形（属性）によって使われる統計手法が異なる。形の異なる物同士で統計計算を行なっても信頼できる結果は出てはこないからで。統計解析を行なう際、データの属性の把握は適用できる手法を選択するのに必須の知識である。

一般的にはデータは次の3つの属性に分類される。

#### 1. 間隔尺度（比尺度を含む意見もある）

対象は温度、検査データなど物理的特性を表すデータが該当し。量的特性をその順序だけでなくその距離も明確にして測定する時に得られるデータと定義される。この定義は言葉だけでは判りづらいが数直線を考えると理解しやすい。例えば温度データが0度、12度、27度、38度と有る。まず0度を数直線上の任意の場所に起き、そして残りのデータを数直線上に置いてゆく。こう並べると4つのデータの順序、データ間の距離関係が一目で分かる。このように定義できる物を間隔尺度と定義される。我々、検査技師が扱う定量的検査項目はこれに含まれる。間隔尺度は離散変数、連続変数どちらも採りうるが連続変数であることが多い。

#### 2. 順序尺度

量的特性をその順序関係に依って数値や分類名を割り当てて測定するときに得られるデータが該当する。例として、治療効果の判定では著効、有効、無効の分類や競技での1位、2位、3位といった具合に固体間の順序（大小）関係のみが定義される物で、割り振られた数値はとびとびの値、離散変数を取る。定性的検査項目の陰性、陽性という判定はこの順序尺度に当てはまる。

#### 3. 分類尺度（名義尺度）

ある属性によって数値、分類名を当てはめて分類するときに得られるデータが該当する。これは各固体の大小関係が定義されない（というか大小関係が問われない分類といえる）。コードNoによる病因分類、男女の分類、喫煙者と非喫煙者の分類などがこの分類尺度に含まれる（これらの分類はその群比較において大小関係が無意味であるということが分かるだろうか）、この分類（名義）尺度はカテゴリ変数（定性的変数）を採るという特徴がある。

また、実際の使用については、男女の分類データを集める際は一般にM/Fの分類で集めると思うが統計処理段階ではMを1にFを2に置き換えて集計処理を行なう。喫煙者・非喫煙者であれば1・0の数値を当てる事が多い。（どのような数字を当てはめるかは個人の自由だが計算が簡単になるように0ないし、1から当てはめるのが一般的だろう。喫煙者・非喫煙者の様に独立した事象であれば非喫煙者に0を当てると喫煙が0（ゼロ、無い）と解釈することができるので見た目の判読性が高まるというメリットもある）文字列型のデータには数値を対応させることで統計計算を行ないやすくすることが行なわれる（これをコーディングと呼ぶ）

次ページに一覧表としてまとめた。

	変数の種類	データ例	説明
間隔尺度	連続変数 or 離散変数	温度, 身長 生化学定量検査 物理的特性量 等	順序 (大小)、距離関係が定義される
順序尺度	離散変数	尿潜血 (定性) -, +, ++, +++ 治療効果の判定 無効, 有効, 著効 競技順位 1着, 2着, 3着 等	順序 (大小) 関係のみが定義される
分類尺度	カテゴリ変数	性別 男, 女 病名 喫煙・非喫煙者の分類 等	順序 (大小) 関係が定義されない

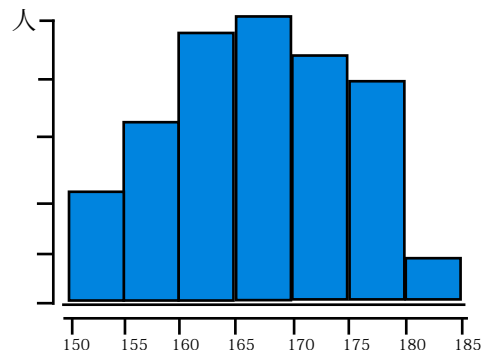
それぞれの尺度は分類→順序→間隔尺度の順で上位の存在と定義され、分類尺度を順序尺度で置き換えることも可能である。しかし、置き換えるには、それぞれの定義を満たす必要がある。例えば、喫煙と非喫煙の形でカテゴリ分けをすると分類尺度となることはお分かりと思うが、では非喫煙と喫煙者を10本以下/日・20本以上/日のようにカテゴリ分けするとどうだろうか？この場合、非喫煙者は喫煙本数が0と読み替えられるから、カテゴリは次のように0本・10本以下・20本以下と読み替えられる。喫煙の有るか無いかだけでなく、喫煙本数の大小関係を定義することが出来るので順序尺度として扱う必要が出てくる。そのため使用する解析手法も前者と後者では異なる物となるので、尺度の当てはめは実験計画の前提条件とも関連するので注意が必要である。

離散変数は整数値をとり連続変数は実数値を採るのが定義であり。自分のデータがどちらに属するか分からない時は得られたデータをグラフ化してみれば良い。

連続変数の場合 ヒストグラムを作成してみる。  
(注意、横軸が連続していることがヒストグラムの必須要件である)

例 身長をグラフ化

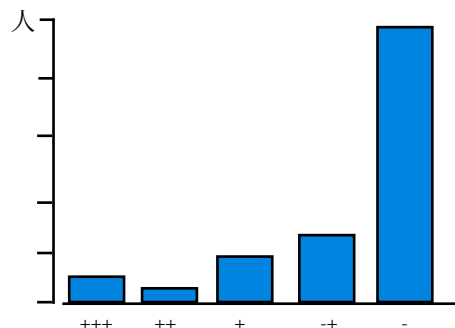
横軸は150cmから185cmまでの間連続した数値となっている  
グラフ化の都合上5cm間隔でデータを区切っているが、プロットされたグラフは隙間を開けずに描画されている。



離散変数・カテゴリ変数の場合  
度数分布表、積み上げ棒グラフなどを作成してみる。

例 尿潜血結果のグラフ化

カテゴリ・変数ごとの度数を情報として使う。連続変数でのヒストグラムと異なり各変数ごとの関係は連続していないので横軸に空白があることが分かる。



グラフの定義から、連続変数と離散・カテゴリ変数のグラフは同じ形に描画してはいけないということが分かる

# 統計的仮説検定とは？その基本原理

右図のように群AとBがあり、この二つの中心位置がずれている。これを見て直感的にその差を意味のあるものと感じたとする。しかし、データはわずか4件ずつであり、その差は偶然のものであるとも考えられる。統計的仮説検定は、その直感に客観性を与える手段として使われる。

検定をするとき最も大切なことは「差がある」という仮説を前提としては検定できないということである。「差がある」という仮説を置くには、二群の差の大きさがあらかじめわかっていないと前提条件とはならない。そしてその差を知るには全数調査を行なう以外には手段が無く、全数調査は事実上不可能であるし、もし行なったとしたらその結果は仮説ではなく「事実」であるので検定を行なう必要が無く、「差がある」という前提で検定を行なうことは論理的に矛盾していることになる。

そこで、「差がある」ではなく「差がない」という仮説を前提において、その仮説に何らかの矛盾があれば「差がある」という最初の仮説を正しいと判断し、逆に、明らかな矛盾が見いだせないときはその判定を保留するという証明方法をとることで矛盾を回避するのが統計的仮説検定の考え方である。

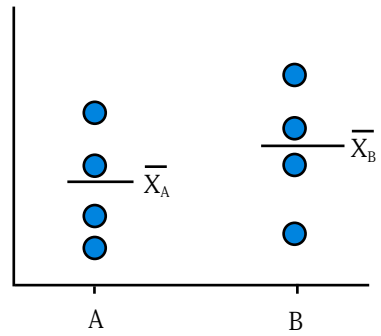
この説明は言葉のトリックのように聞こえるかもしれないが、原理自体は背理法（反証法）によって成立している。証明のために命題の反証を検証する、その反証に矛盾があればその反証を捨てて証明を行うというもので、反証に相当する「差がない」という仮説は、本来は無に帰するものとして「帰無仮説」 $H_0$ 、最初の命題「差がある」という仮説は「対立仮説」 $H_1$ と定義される。

## 検定の流れ方

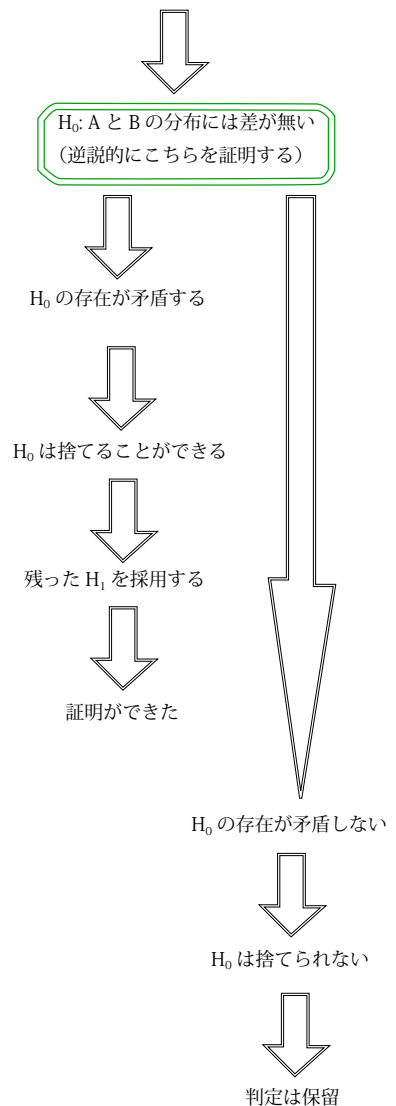
- 1 母集団に関する帰無仮説と対立仮説（両側 or 片側検定）を設定
- 2 検定統計量を選ぶ
- 3 有意水準 $\alpha$ を設定する
- 4 データを収集し、そのデータから検定統計量 $X$ を求める
- 5 帰無仮説の差が無いという仮説から $X$ が生じる確率 $P$ を求める
- 6  $P$ が有意水準 $\alpha$ よりも大きければ $H_0$ を棄却できず判定を保留  
 $P$ が有意水準 $\alpha$ よりも小さければ $H_0$ を棄却し $H_1$ を採用する

検定において1から3の手順は実験計画を立てる段階で設定を行なうべき物で、3までの手順が決まれば検定手法も設定が可能となるので集まったデータを設定した検定手法によって解析するという流れになる。

5で求めた確率 $P$ は帰無仮説 $H_0$ が起こりうるであろうという仮説の存在確率（%）を表す。従って確率 $P$ が有意水準 $\alpha$ より小さいければ $H_0$ が成立する確立が低いことを意味し $H_0$ を捨て（棄却）、対立仮説 $H_1$ を採用することになる。 $H_1$ を採用できればAとBには差があるということになるが、そうでない場合は「差がない」ではなく、「差があるとはいえない」「有意な差を見ることは出来なかった」と控えめな言い方しておくべきである。一般的にデータ数を増やしてゆくほど、わずかな差を検出できる可能性が増加してゆくため



$H_1$ : AとBの分布には差がある  
(証明したいのはこちら)



では、データを増やせば検出力が上がるのなら標本数はどの程度あれば「差がある」という望みの結果が出るのか、標本数は多ければ多いほど良いのではないのかという疑問も持たれるであろう、実験計画条件が適切に設定されている前提で後述する t 検定において、その t 検定の頑強制から 50 から 100 件程度の標本数で検定の精度は満たされるとする考えがあり。これ以上の標本数で検定を行なっても全く問題は無いが、徒に標本数が多すぎると「差が無い」という結果を望んでいるのに「差がある」という結果が出てくる事も十分想定される。一般的に t 検定に限らずその他の検定手法においても数百件くらいが標本数の上限ではないかと考える。もちろん疫学調査等においてはその標本数が数千件単位になることも珍しくは無いが、それは対象とするものを含むデータを信頼に足るだけ集めようとするれば（ある病気の罹患率など）当然それぐらいの標本数が必要とされるからである。

また、検定で求めているものは存在確率であり。結果として「有意差あり」となっても差がどのくらいあるのかについては一切求めてはいない点にも注意しておくべきである。確率の大きさとデータの差の大きさは無関係であるので、統計的仮説検定は仮説の妥当性を計算しているだけで実際のデータについて何かを論じているのでは無い。「木を見て森を見ず」ではないが統計的仮説検定の結果だけに頼ってはならない、実験計画とデータから見出される解釈と各人が属する領域での解釈と合わせての判断が求められることを忘れてはならない。

#### 有意水準について

一般的に 5%、1% 有意水準として覚えられている方も多いと思うが、この 5%、1% は絶対的に定義されているものではない。この当たりの数値であれば第一種、第二種の過誤の点から妥協できるだろうとの経験則から出発しているが、特殊な分布型を持つ標本では全く異なる数値が良いとされる考えもある。この有意水準は研究者がそれぞれ決めることが出来るが、普通は先行して研究があればその数値に従うのが無難であり、先行研究が無い場合でもよほど特殊な事情が無い限りは 5%、1% の有意水準値を用いておくのが普通である。

## パラメトリック検定・ノンパラメトリック検定

様々な場面で目にする事の多いパラメトリック・ノンパラメトリック。これはデータの形に応じて適応できる検定方法が異なるために用いられる用語で以下の様に定義される。

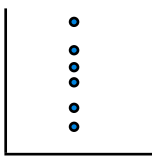
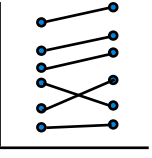
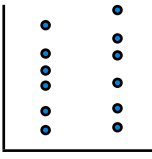
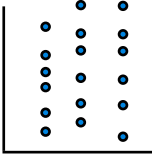
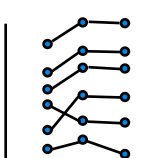
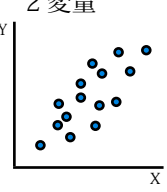
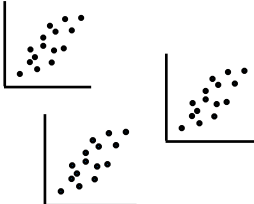
	適応データ形式	母集団分布形式
パラメトリック検定	間隔尺度	正規分布を前提
ノンパラメトリック検定	分類尺度・順序尺度	特定の分布型を前提としない

パラメトリックとはパラメータ(parameter)から由来する言葉で母数を意味する。パラメトリック検定とは母数(母集団)の分布に何らかの仮定をおく検定を指し t 検定、F 検定がこれに該当する。対してノンパラメトリック検定とは母数を仮定しない検定を指すが、厳密にはこの区分はそれほど明確ではないのが実情である。例えば分類尺度を用いる  $\chi^2$  検定は母集団分布に特定の仮定をおかないのでノンパラメトリック検定になるが、その統計量が  $\chi^2$  分布に従うには十分なデータ数が必要となる。一方、正規近似する順位和検定もデータ数が多いことが必要であり、検定の本質は何なのかという問題になってくる。そのため一般的には表の定義は踏まえながら量的変数を扱う時、その量の情報は使わずに大小関係、つまり順位情報だけを扱う分析法をノンパラメトリックな検定（解析法）としても構わないとされている。

t 検定、F 検定のように母集団分布に過程を置く検定法の場合、集められたデータがその母集団分布に従うかどうかの検証が前段階として必要になる。その検証をクリアしてから t 検定、F 検定を行なう

次ページにデータの形式別にどのような検定法が当てはまるか、図を掲載する。

データ形式と尺度による検定手法の選択例

データ形式	間隔尺度	順序尺度	分類尺度
1 標本 	平均値の検定 分布の歪度・尖度		比率の検定 二項検定 ポアソン検定 $\chi^2$ 適合度検定
関連 2 標本 	一標本 t 検定	Wilcoxon 符号順位和検定	符号検定
独立 2 標本 	二標本 t 検定 等分散の検定	Mann-Whitney 検定	2 x 2 分割表 $\chi^2$ 独立性の検定 Fisher の直接確率 比率の差の検定
独立多標本 	一元配置分散分析 Bartlett 検定	Kruskal-Walis 検定	1 x m 分割表 $\chi^2$ 独立性の検定
関連多標本 	二元配置分散分析 Bartlett 検定	Friedman 検定 Kendall の一致係数	
2 変量 	回帰係数の検定 相関係数 重相関係数	Spearman 順位相関係数	$\phi$ 係数 クラメールの C 係数 ユールの連関係数
多変量 	偏相関係数 重相関係数	Kendall 一致係数	

全ての検定法の説明を行なうことはできないが、良く使われる検定法についてこの後説明をする。

# t 検定

検定の説明を行うのによく知られている t 検定を使って順を追って、統計的仮説検定を説明したい。

t 検定は間隔尺度データである二標本の平均値の差の検定に用いられる。順序・分類尺度では t 検定は使われない。二標本には独立した二標本と関連した二標本があり、対応のない t 検定、対応のある t 検定と呼ばれることが多い。そして、t 検定を行うにはそれが成立する前提条件を満たす必要がある。その条件は以下の通りである。

- 1 標本の抽出が無作為に行われていること（無作為抽出）
- 2 母集団の分布が正規分布に従っていること（正規性）
- 3 2つの母集団の分散が等質であること（分散の等質性）

1の無作為抽出は実験計画設計時に考慮すべき物なので、得られたデータが2と3を満たすことをどう確認するのだが、まず2の正規性に従う確認は、幾つかの方法でデータの正規性を確認する方法がある。なお、統計専用ソフトであれば正規性の確認をするためのメニューがあると思うがどの方法を用いているかは筆者には分からないので各人で確認していただきたい。

Excelで正規性の検定を行うには、簡単な方法が無いが Jarque-Bera 式を使って確認ができる。データから尖度「kurt」と歪度「skew」を Excel 関数から算出し、以下の式に代入する

$$\text{Jarque-Bera 検定統計量} = n/6*(s^2+(k-3)^2/4)$$

で求められた統計量がおおよそ 6 より大きければ有意水準 5% で正規分布と見なしても構わない。

R ではシャピロ=ウィルク (Shapiro-Wilk) の検定と呼ばれる方法を使う (\* hokanim)、以下の関数で実行可能である `shapiro.test(X)`

(X) は検定を行うデータ列を指定する。この式で計算を行うと p-value で有意確率が計算されるので、0.05 より大きければ正規性に従うという帰無仮説が棄却されないので正規分布に従っているという判断を支持することができる。

(\*) 他にもコルモゴロフ-スミルノフ Kolmogorov-Smirnov 検定も知られているがこちらはノンパラメトリック手法であり、正規分布を仮定しているシャピロ=ウィルク検定の方が検出力が高い

次のデータは成人男性の体重データであるが、Excel と R でそれぞれ正規性の検定を行ってみる

体重 50.5 58.0 47.5 53.0 54.5 61.0 56.5 65.5 56.0 53.0 54.0 56.0 51.0 59.0 44.0 53.0  
62.5 55.0 64.5 55.0 67.0 70.5 46.5 63.0 51.0 44.5 57.5 64.0

Excel では Jarque-Bera 検定統計量 =  $n/6*(s^2+(k-3)^2/4)$  の必要数値をまず求める。n は標本数なので 28、s が歪度で =SKEW(データ範囲) から 0.0785。k が尖度で =KURT(データ範囲) から -0.4183 となり、これらを代入すると 13.6609 となる。この統計量は 6 よりも大きいので正規分布と見なして良い。

R では `shapiro.test(Dataset$WT)` と打つと以下のように計算される

```
Shapiro-Wilk normality test
data: Dataset$WT
W = 0.9799, p-value = 0.8473
```

この p-value=0.84 は 0.05 よりも相当大きいので Jarque-Bera と同じように正規性があるという帰無仮説を支持できるので、この成人男性の体重は正規分布をしていると判断される。

3の分散の等質性は2つ以上の標本データの分散を調べることで検証できる。Excelでは分析ツールにF検定、Rでは `var.test()` で調べることができる。

例題として 自動血糖測定機の在来機と新規機を使って同一検体を測った時のデータを使い、それぞれの測定値の分散が等しいか検討してみる。データ数は 59 件で Excel ではワークシートにデータを入力した後、分析ツールから F 検定を選ぶと以下のような結果が表示される。

D	OLD	NEW			
1	91	90			F-検定 : 2 標本を使った分散の検定
2	97	99			
3	102	104			変数 1 変数 2
4	86	91			平均 96.8 98.51667
5	92	93			分散 170.3661017 179.3048
6	104	109			観測数 60 60
7	83	87			自由度 59 59
8	102	104			観測された分散比 0.950148014
9	106	110			P(F<=f) 両側 0.422478723
10	101	105			F 境界値 両側 0.649368559
11	99	102			
12	97	99			
13	84	86			

Rにも同様にデータを入れて var.test を行なうと以下のようになる

```

R Console
50 50 122 125
51 51 112 112
52 52 91 91
53 53 92 96
54 54 78 78
55 55 93 93
56 56 90 88
57 57 82 82
58 58 80 79
59 59 93 93
60 60 98 97
> var.test(df$OLD,df$NEW)

      F test to compare two variances

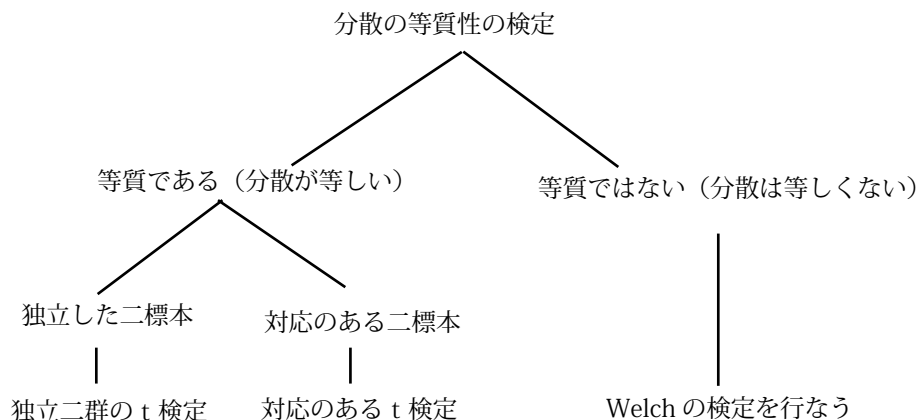
data:  df$OLD and df$NEW
F = 0.9501, num df = 59, denom df = 59, p-value = 0.845
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5675467 1.5906732
sample estimates:
ratio of variances
      0.950148

>

```

ここで見るべきところは、ExcelのF検定であればセルを黄色に塗った部分のP ( $F \leq f$ ) 両側となっているところの数値で、Rであればp-valueを見る。これが分散が等質かどうか(等分散しているか)を見る数値である。ところでExcelが0.4224でRが0.845と表示されているが、この部分が前回の資料でも書いたExcelのバグである、筆者の持っているExcel2000ではこの0.4224というのは片側確率の数値でこれを二倍すれば0.8448とRで計算された両側確率と同じになる、両側確率が0.84と0.05より十分に大きいので二つのデータの分散は等しいと判断できる。

では、分散が等しくなければどうするか? 等分散でなければwelchの検定と呼ばれる方法で検定を行なう事ができ、まとめると以下のようなになる



Excelの分析ツールでは「独立二群のt検定」が「t検定:等分散を仮定した二標本による検定」、「対応のあるt検定」が「t検定:一対の標本による平均の検定」、「Welchの検定」が「t検定:分散が等しくないと仮定した二標本による検定」にそれぞれ対応する。

Rでは「独立二群のt検定」が `t.test(x1,x2,var.equal=TRUE)`、「対応のあるt検定」が `t.test(x1,x2,paired=TRUE)`、「Welchの検定」が `t.test(x1,x2,var=FALSE)` に対応している。

#### 補足

t検定を成立させる前提条件から、正規性の検定、等分散の検定を成立させるには(そうならないと判っているものは除いて)無作為抽出が上手くできるか否かに掛かっている事が分かると思う。とはいえ、この前提条件は極端な場合を除きたてまえ上のもので、差の検出力を考慮しなければそれほど厳密に考えなくても良いとされる。

何故か?

小標本の場合

厳密に正規性・等分散性を検定しても、小数であるが故に検定をパスする。

大標本の場合

こちらでは厳密に検定を行うと僅かな差から正規性・等分散性の偏りが検出されるため、殆ど全てのデータは正規分布から偏っているとされ、t検定が適用できなくなるが

#### ●非正規分布だが分散は等しいとき

中心極限定理から平均値の差は近似的に正規分布する。合成分散  $s^2$  も正確に求められるのでt値は正規分布で近似可能となり、t検定の適用が正当化出来る。

#### ●非正規分布で分散も等しくないとき

この状態が極端であればやはりt検定は適用できない。ただ、データ数が十分に大きく、標本分散から母分散をほぼ正確に推定出来る場合は、正規検定で2群の平均値の差を調べられる。



このように、殆どの場合で前提条件のチェックを、さして厳密にする必要が無いことになってしまう。しかし、それぞれのパターンにおける解析方法はあるわけだし、前提条件の検定が疑わしい結果となったのであれば、どうして前提条件を満たさないのか、当初の計画が間違いではなかったのかといった検証をおこなって見るべきであろう。見直しの結果、それでも t 検定を行うしかなければそれはそれで正当化が出来ると思う。

## t 検定の実際

前提条件の説明に大きくページを割いてしまったがここからは例題を使って t 検定の行うことで、統計学的仮説検定がどのように行われるのかを追ってみたい。

### 例題

	1	2	3	4	5	6	7	8	9	10
早朝	0.98	0.87	1.12	1.34	0.88	0.91	1.04	1.21	1.17	1.09
夕方	1.03	0.78	1.04	1.52	0.97	0.84	1.32	1.12	1.09	1.32

10 人の協力者から早朝と夕方採血をして血中 CRE 値を測定した。早朝時と夕方では CRE 値に有意差があるかを検定したい。

- まず同じ人物から早朝と夕方に採血をして得られたデータであるので「対応のある二標本」となる。
- 帰無仮説と対立仮説を設定する。  
 帰無仮説  $H_0$  : CRE 値の変化は無い (二群の差の平均値には差がない)  
 対立仮説  $H_1$  : CRE 値は変化がある (二群の差の平均値には差がある)
- 検定統計量の選択  
 ここでは差の平均値  $\bar{d}$ 、標準偏差  $S_d$  を求める。  
 $\bar{d}$  を標準誤差  $S_d / \sqrt{n}$  で標準化して t 値を求める。
- 確率を求める。  
 帰無仮説  $H_0$  に基づく  $\bar{d}$  の理論分布から  $|\bar{d}| \geq \bar{d}$  となる確率 p を求める
- 判定  
 $p \geq \alpha$   $H_0$  を棄却できない  
 $p < \alpha$   $H_0$  を棄却し、 $H_1$  を採用

検定統計量をどうやって求めるのかと思われるであろう、実際これぐらいの標本数であれば手計算でも十分計算が可能であるが、これら統計量の選択と確率を求める部分の計算は統計計算関数で一気に行うことが出来るので計算式は省略させて頂いて、Excel と R に計算させてみたい。

Excel では分析ツールの「t 検定：一対の標本による平均の検定」を実行すると次のような結果になる。

	J	K	L	M	N	O	P
	AM	PM		t-検定：一対の標本による平均の検定ツール			
1	0.98	1.03					
2	0.87	0.78			変数 1	変数 2	
3	1.12	1.04		平均	1.061	1.103	
4	1.34	1.52		分散	0.023921	0.052112	
5	0.88	0.97		観測数	10	10	
6	0.91	0.84		ピアソン相関	0.7769		
7	1.04	1.32		仮説平均との差異	0		
8	1.21	1.12		自由度	9		
9	1.17	1.09		t	-0.91276		
10	1.09	1.32		P(T<=t) 片側	0.192584		
				t 境界値 片側	1.833114		
				P(T<=t) 両側	0.385169		
				t 境界値 両側	2.262159		

R では `t.test (x1,x2,paired=TRUE)` でデータを指定すれば、以下ようになる。

## Paired t-test

```
data: Dataset$AM and Dataset$PM
t = -0.9128, df = 9, p-value = 0.3852
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.14609201  0.06209201
sample estimates:
mean of the differences
                -0.042
```

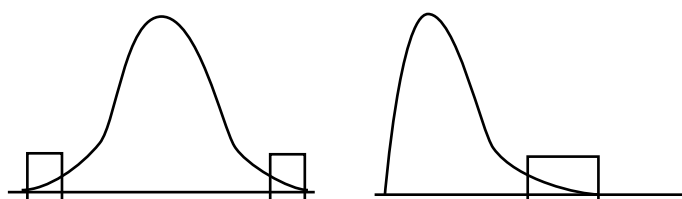
検定統計量は  $t$  値で表される数値で  $-0.912$  で、確率は  $P(T \leq t)$  両側 (Excel)、 $p$ -value (R) で表示される  $0.3852$  が最終的に知りたい確率値である  $5\%$  の有意水準で見ると  $0.3852$  は  $0.05$  (有意水準  $5\%$ ) よりも大きいので帰無仮説  $H_0$  を棄却することはできない。つまり二つのデータからは朝と夕方で CRE 値に差があるとは言えない。という結果になる。

もし、この  $p$ -value が  $0.0011$  のように  $0.05$  よりも小さな数値になったときは二つの標本が同じ様な状況になる確率がかなり小さいことを意味し、そのような状況になることはなかなか考えにくいということで帰無仮説  $H_0$  を棄却して (つまり帰無仮説は無かったものとして) 対立仮説  $H_1$  を採用する。この場合の対立仮説は「二標本の平均値には差がある」なので「二つの標本データは同じ物とは言えない」、「 $5\%$  有意水準において有意の差を認める」というような言い方をする。 $P$  値は定められた有意水準において有意差があるかないか (帰無仮説を棄却できるか否か?) を二者択一するための値であり、 $P$  値の差が有意差の大きさを表していない点を理解しておく必要がある。

この例題では対応のある  $t$  検定で行ったが対応のない二標本であっても検定方法の指定を正しく行えば出てくる  $P$  値の評価方法は同じである。(計算方法が異なるので  $p$  値自体は異なる) また、 $t$  検定だけではなく他の検定方法も得られる  $P$  値が有意水準を越えるか超えないかを評価することは同じであるので統計学的仮説検定の本質は、 $t$  検定を行なってみれば容易に実感できると思う。ただ、検定が難しいと感じさせる大きな要因は比べるデータの選択、適切な方法の選択が慣れないと非常に分かりづらい点ではないかと思う。

「対応のある」、「ない」の使い分けだが、計算自体は「対応のある」データを「対応のない」 $t$  検定式に代入して計算することは可能だが「対応がある」状況で得られたデータは「対応のある二標本」として扱わなければならない。

両側確率と片側確率のどちらを使えば良いのかと疑問を持つ方もおられると思うが、標本母集団の分布型に従うのが基本である。正規分布は下図のように両側に裾を引くので四角で囲ったように両裾に確率分布が存在するので両側確率を使うのが望ましいが、隣の  $F$  分布 ( $F$  検定ではない) では確率分布は片側だけしか存在しない。この場合は両側は無意味なので片側確率を使えばよろしい。



## 第2回のまとめ

### 1 データ形式

- データには幾つかの形があり 間隔尺度、順序尺度、分類尺度に分けられる。
- データの形に応じて適用できる統計学的手法が異なる。

### 2 統計学的仮説検定

- 比較するデータが属すると仮定した母集団が同じ物であるかどうかを検証する手法。
- 検証すべき仮説は比較する「母集団は同じ物」である、「異なる母集団」という仮説の検証ではない。
- 検証仮説の妥当性を確率を使用して判断する。
- あらかじめ設定している有意水準より検定統計量の生ずる確率が大きければ母集団に差は見られない、有意水準より小さければ二つの母集団が同じだと確率的に考えずらいということで、母集団に差があると判断する。
- 統計学的仮説検定は決して有意差を出すためのものではないし、有意差が出るような検定方法もまた存在しない。
- この結果はあくまでも計算に使用したデータのみ有効な結果であって、(データから想定される母集団に対しての結果) 結果を対象データ群、全てに普遍化してしまうことは危険、データ数が増えれば異なる結果が出てくることは稀ではない。(概ね結果はその母集団に反映されるが、必ずしも一つの検定の結果が母集団性質を100%語っているわけではない、何事にも例外があるし、自分の結果がその例外である可能性もあるので、解釈に対して大胆さと謙虚さのバランスが求められる。)
- 結果の最終判断は統計学優先ではなく、自分の専門分野の知識に立脚して判断を行なう事が望ましい。統計学的な有意差と学問的(臨床的)有意差は同じではない。実質的に無意味な細かな差を論じても、それはただの自己満足に過ぎない。学問的に意義があるかどうかを判断する材料として統計学的仮説検定を用いているのだから、最終判断は学問的に意義があるかどうかの点から判断すべきである。

### 3 パラメトリック・ノンパラメトリック検定

- パラメトリック検定は正規分布に従うと仮定できる間隔尺度のデータに対して用いられる。
- 正規分布ではない分布型や間隔尺度以外のデータ型はノンパラメトリック検定として扱う。

### 4 t 検定

- 二標本の平均値の差の検定方法である
- 標本が等分散であることが必要。
- 対応のない(独立した)二標本と、対応のある(関連した)二標本とでは計算方法が異なり。同じデータでも結果が異なるので、方法を取り違えてはいけない。
- 標本が等分散しなければWelchの検定と呼ばれる方法で検定が可能。