

検定

牧野 泰才

平成 20 年 5 月 13 日

1 はじめに

人間の感覚・知覚などを評価する場合、心理物理実験が行われる。これは、物理的に客観的な量(光の波長、音の周波数、圧力や変位など)を、人間が主観的にどのように知覚するか、という対応関係を明らかにすることを目的とした実験である。

人間の主観的な評価であるから、ある刺激 A を加えた場合の評価が常に同じになるわけではなく、例えば 10 段階評価であれば、4 点あたりを平均に 2 点から 7 点の間でばらつく、といったような結果が得られる。

それでは、この刺激 A とは異なる刺激 B が与えられたとき、その結果を見て「刺激 A とは異なる刺激」として知覚されているのか、「刺激 A と似た刺激」と知覚されているのかは、どのようにして判断すべきか? ここで登場するのが検定の考え方である。

検定の数学的な意味を解説するのも重要ではあるが、現在は様々なツールがあり、自分で全て手計算で検定を行うことはほとんど無い。そこで本講では、数式等の厳密な話は教科書に任せ、もっと定性的に検定の考え方のエッセンスを紹介する。実際に Excel を用いて検定を行い、各値がそれぞれ何を意味するかを理解し、ツールとして使いこなせることを目的とする。

2 検定の考え方

検定のおおまかな流れを以下に示す。

- 1) 比較したい 2 つ以上の分布に差が無いと仮定する
- 2) 適当な指標を計算する
- 3) その指標の数値が起こりうる確率を求める
- 4-1) 確率が低ければ、分布に差が無いという仮定が間違っている 分布に差がある
- 4-2) 確率が高ければ、分布に差が無いという仮定が正しい 分布に差は無い

1) の仮定は「帰無仮説」と言われる。大抵の場合、実際に示したいのは「2 つの分布に差がある」ということであるが、検定においてはまず、「2 つの分布に差が無い」と仮定する。そして、この仮定が正しくなければ「2 つの分布に差が無いとは言えない」=「2 つの分布に差がある」と考えるのである。

2) の「適当な指標」は、それぞれ比較したい対象によって異なる。例えば、よく利用される t 検定では、2 つの分布の平均の差を分散で除したような指標を用いる。ここで、母集団に対し正規分布であるという仮定をおくことで、その指標が統計的に従うべき分布 (t 分布) が算出される。

これにより、その指標の数値が起こりうる確率が算出され、その結果を元に、帰無仮説を採用するか否かを判断する。一般的に利用される判断基準は 3 種類で、それぞれ確率が 5% 未満、1% 未満、0.1% 未満か否かを判定基準 (有意水準) とする。

2.1 パラメトリック検定とノンパラメトリック検定

検定には大きく分けて2つの方法がある。1つはパラメトリックな方法であり、もう1つはノンパラメトリックな方法である。パラメトリックな方法は、母集団の分布に対して正規分布であるという仮定をおく方法である。一方ノンパラメトリックな方法では、母集団の分布型に対して一切の仮定を設けない。

注意が必要なのは、パラメトリックな方法の場合、次節に示す尺度水準のうちの名義尺度、順序尺度については適用できない点である。名義、順序の両尺度により評価がなされている場合には、ノンパラメトリックな検定を行う必要がある。

2.2 尺度水準

データには、名義、順序、間隔、比率の4つの尺度がある。

名義尺度

「はい」と「いいえ」、「男」と「女」など、単に分けることにのみ使っているデータ。例えば刺激が「あった」か「なかった」かなどは名義尺度である。

順序尺度

「1位」「2位」「3位」や「大変良い」「良い」「ふつう」「悪い」「大変悪い」のように、順序関係で分類されるもの。名義尺度の変数は数直線上に並べることは出来ないが、順序尺度の変数は並べることが可能。その間の差は定義されない。したがって、中央値には意味があるが、平均値には意味が無い。

間隔尺度

温度など、測定値の間の差が意味を持つが、比は意味を持たないもの。例えば10と20の差は意味があるが、2倍の温度であるわけではない。

比率尺度

長さや重さなど、測定間の距離とともに、その比率も意味を持つもの。真の0を持つデータであるとも言える。

尺度は比率、間隔、順序、名義の順で上位に位置し、上位の尺度は下位の尺度の性質も備える。前節で述べたように、パラメトリックな方法は、上位の間隔、比率尺度にのみ適用可能であり、順序、名義尺度には適用できない。というのも、パラメトリックな方法の場合には、正規分布の平均、分散を考えるからであり、名義、順序尺度はその平均値が意味を成さないからである。なお統計では比率、間隔尺度は同じ手法で扱えるため、それほど厳密に区別する必要は無い。

2.3 検定方法対応表

<http://aoki2.si.gunma-u.ac.jp/lecture/Kentei/nonpara.html>

に、尺度と比較したい統計量、その際の検定方法についてまとめた表があったので転載する。

	パラメトリックな手法	ノンパラメトリックな手法
対象とする統計量	平均値	代表値
	分散	散布度
	積率相関係数	関連性係数, 順位相関係数
		度数
尺度水準	間隔、比率	名義、順序、間隔、比率
母集団の分布型	正規分布を仮定	不問
	等分散性を仮定	
標本サイズ	小さすぎたはいけない	不問

検定目的	パラメトリック	ノンパラメトリック	
		名義尺度	順序尺度以上
適合度		χ^2 検定	χ^2 検定 1 標本コルモゴロフ・スミルノフ検定
独立性	相関係数の検定	χ^2 検定 フィッシャーの正確確率検定	χ^2 検定 フィッシャーの正確確率検定
比率の差		χ^2 検定 フィッシャーの正確確率検定 マクネマー検定 コ克兰の Q 検定	χ^2 検定 フィッシャーの正確確率検定 マクネマー検定 コ克兰の Q 検定
母比率		二項検定	二項検定
対応のない 2 標本の 代表値の差	平均値の差の t 検定		マン・ホイットニーの U 検定 2 標本コルモゴロフ・スミルノフ検定 ファン・デル・ワーデン検定 中央値検定
対応のある 2 標本の 代表値の差	平均値の差の t 検定		符号検定 符号付順位和検定
対応のない K 標本の 代表値の差	一元配置分散分析		クラスカル・ウォリス検定 中央値検定
対応のある K 標本の 代表値の差	乱塊法		フリードマンの検定

3 t 検定

3.1 概略

パラメトリックな手法の例として、t 検定について概説する。t 検定は既に少し触れたが、2 つの分布の平均が同じか否かを判定する場合に利用する検定である。間隔尺度以上の尺度水準に対し、母集団の分布が正規分布であることを仮定できる場合に対して適用可能である。

t 検定は、まず大きく 2 つの場合に分けられる。1 つはデータに対応関係がある場合、もう 1 つは無い場合であり、それぞれ処理が異なる。対応関係がある場合というのは、例えば同じ被験者が異なる 2 つの実験条件下において、長さの評価を行ったような場合である。A という刺激の場合と、B という刺激の場合で、同じ被験者のデータが対になって得られているのであれば、「それらが対を成すデータである」という前提の下で検定を行ったほうが、より正確な結果となる。

対応関係が無い場合、さらに 2 つの状況に分類することが出来る。それは、各分布の分散が同じとみなせるか、否かの 2 つである。そもそも、帰無仮説で仮定するのは、2 つの分布が同じ母集団からのサンプルであり、結果として同じ平均・分散を持つという状況である。この仮定に基づいて、実験による結果から母集団の分散を推定するわけであるが、実際のデータの分散が大きく異なる場合には、この前提が成り立たないため、異なる処理を行う必要がある。

なお、この等分散性を検証するために、F 検定が利用される。同じ正規母集団から取り出された 2 つのサンプル群の不偏分散の比が、F 分布に従うということを利用して行う検定である。

3.2 実際に行うこと

EXCELにおいては、t検定を行うための関数としてTTEST関数が用意されている。この関数は4つの引数を必要とする。はじめの2つは、比較したい2つの分布の数値データ列である。残りの2つのうちの1つは上記の3つの状態をそれぞれ指定する。すなわち「対をなすデータのt検定」「等分散の2標本を対象とするt検定」「非等分散の2標本を対象とするt検定」の3種類のうち、どの条件で解析するかを指定する。

最後の1つのパラメータは、両側検定か片側検定かを指定するものである。t検定において、帰無仮説が棄却される条件は2つある。1つはAの方がBよりも大きいために差が生じる場合。もう1つはBの方がAよりも大きいために差が生じる場合である。片側検定は、上記の条件の片方だけを満たすかどうかを検定する方法であり、両側検定はどちらの場合も考慮される。同じ有意水準の場合、片側検定の方が条件が緩くなるため、片側検定においては有意差が認められたが、両側検定においては有意差が得られないという場合も生じる。検定を行う際には、結果を見てからではなく、自身の実験内容と照らし合わせて、予め片側検定か両側検定かを決定して解析しなければならない。

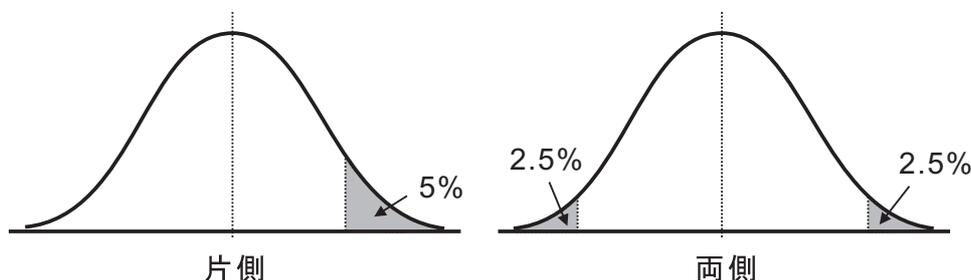


図 1: 片側検定と両側検定

3.3 数値例

例えば以下のようなテスト結果があったとする。このとき、両者の分布に差があるかどうかを検定したい。

	A	B	C	D	E	F	G	H	I	J
国語	70	65	35	50	70	80	65	70	65	60
数学	80	60	55	70	85	95	70	85	70	70

そこでまずはじめに、F検定により等分散を仮定できるかどうかを確認する。適当なセルにおいて、
=FTEST(配列 1, 配列 2)

とすると、2つのデータ間の分散が等しいかどうかの検定が行える。今回の場合 0.9408 という結果が得られる。これは「2つの分布が等しい分散を持つ」という帰無仮説に対して、それが生じうる確率が 94%であるということの意味する。すなわち、両者の分布は同じ分散を持つと考えて差し支えない。なお、実際の分散の値は 156.67 と 148.89 である。

そこで次に、「対をなすデータのt検定」を行う。すなわち、A~Jは各被験者を表すラベルであると考え、テストの点数にそれぞれ対応関係があるとする。

=TTEST(配列 1, 配列 2, 2 (両側検定), 1 (対応関係あり))

として計算すれば、“0.0015”という数値が得られる。これは、個別に対応のある国語と数学の点数に対して、差が無いと仮定して同じ母集団からデータを抽出してきた場合に、このような分布になる可能性は 0.15%程度である。

る、ということを示している。つまり、5%水準において有意な差があると言える。このテストに関して言えば、国語の方が数学よりも難しかったということが分かる。

では、対応の無い検定を行った場合はどうなるだろうか？ つまり、A~Jは被験者の名前を表すのではなく、各テストの結果から無作為に10個ずつデータを取ってきたと考えた場合である。この場合、

=TTEST(配列1, 配列2, 2 (両側検定), 2 (対応関係なし等分散))

として計算すると、“0.062”という結果が得られた。これは5%有意水準で帰無仮説が採用されるという結果になっている。つまり、2つの分布に差が無いというのがこの検定での結論になる。

このように、同じt検定でも対応の有無によって結果が異なることがある。これが何を意味するかを知るには、対応のあるt検定と、対応の無いt検定とで、それぞれ何をしているのかを知る必要がある。

定性的に述べると、対応のある場合には、“対応するデータ同士の差を算出し、その差の平均”を検定統計量として用いる。一方、対応の無い場合には、“データ系列の平均値の差”を検定統計量として用いる。直感的な話をすれば、平均を取って均されてしまったデータの差を比較するよりも、個別に差を取りその平均値が0に近いかな否かで比較した場合の方が、検出力が高くなるということである。

3.4 注意点

以上のように、t検定を適用可能な条件というのがあり、またt検定でも3つの状況があるということが分かった。更に注意すべきは、t検定は2つの分布の比較にしか用いることができないという点である。3つ以上の分布には、分散分析等を利用する必要がある。

この理由は、以下のようにして理解できる。A,B,Cの3つの標本間のt検定を独立に行った場合「少なくとも1つのペアに差が生じる確率」は、「3つ全てに差が生じない確率」を1から引いたものになる。有意水準を5%とすると、

$$\begin{aligned} P &= 1 - (A - B \text{ 間に差がない} : 0.95) * (B - C \text{ 間に差がない} : 0.95) * (C - A \text{ 間に差がない} : 0.95) \\ &= 1 - 0.857 = 0.143 \end{aligned}$$

となってしまう。すなわち、各t検定では5%の有意水準で比較しているにも関わらず、最終的な評価は14.3%の水準で評価してしまうことになっている。よって、3つ以上の分布の比較には、分散分析を採用しなければならない。

なお、分散分析の場合には、1つのデータ群内の分散と、トータルなデータ群の分散の比を、統計検定量として用いる。全ての分布が同じであれば、データ群内の分散とトータルな分散とは、同じ値になるため1に近づき、分布が大きく異なれば、その値は1から大きく離れる。このときの検定量の従う確率分布は、先ほど等分散性の検定に用いたF分布になる。これを用いて検定を行う。(要望があれば詳しくは次回)

最後に、パラメトリックな手法がメジャーなようなイメージがあるが、パラメトリックな手法だからといって、必ずしもノンパラメトリックな手法よりも検定力が高いわけではない。これまでパラメトリックな手法が主に使われてきた理由には、単純に計算にかかるコストの問題という面もある。というのも、パラメトリックな方法では分布を正規分布で近似するため、平均と分散という少ないデータで、全体を扱うことが出来るからである。現在のように高速に計算可能なコンピュータが使えるのであれば、より検出力が高く、実験を評価するのに適した手法を採用すべきである。(ノンパラメトリックな方法は[1]に詳しい)

参考文献

- [1] 粕谷 英一: “生物学を学ぶ人のための統計のはなし～きみにも出せる有意差～,” 文一総合出版, 1998.
- [2] 向後 千春, 富永 敦子: “統計学がわかる,” 技術評論社, 2007.
- [3] 東京大学教養学部統計学教室 編: “基礎統計学 1 統計学入門,” 東京大学出版, 1991.

その他ウェブサイト多数。主に以下。

<http://aoki2.si.gunma-u.ac.jp/lecture/Kentei/nonpara.html>

<http://hosho.ees.hokudai.ac.jp/kazuki/base.html>

http://homepage2.nifty.com/nandemoarchive/suite_and_kente/t.htm

<http://www.iac21.com/ryoukata.htm>