心理学における実験計画の手引き

〔その２〕実験結果の評価と結論

～におい・かおりに関わる心理学的研究で気をつけることは？～

　このリーフレットは、におい・かおりに関わる研究の中で、ヒトを対象として何らかの測定・評価を行う際に必要とされる基本的知識をまとめたものです。
　心理実験の初心者向けに易しく解説しています。皆さんの研究の一助となれば幸いです。

小林　剛史・山本　晃輔・小峯　裕己
におい・かおり環境協会技術課（重岡久美子・中辻　康・石井　進）

2025 年 1 月 10 日

公益社団法人　におい・かおり環境協会

３．評価（統計解析）の留意点

心理学における実験計画の手引き〔その１〕に引き続き、〔その２〕では、心理学実験の結果を評価する際に用いる主な統計解析や、解析結果から導かれる解釈の留意点などについて解説します。

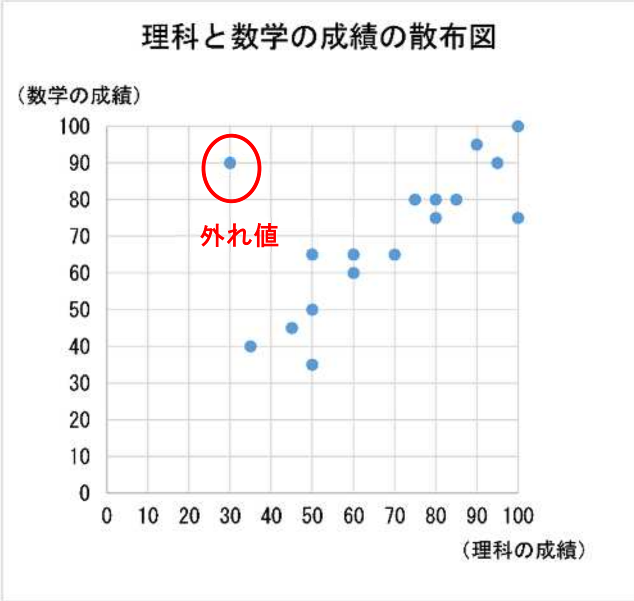
３．１　統計解析で誤解されやすい統計用語

■　母集団と標本データ

調査や研究の対象となる集団全体、本来知りたいと思っている集団全体のことを「母集団」といいます。一方、母集団の情報を推測するために選ばれた母集団の一部で、観測、測定の対象とする集団を「標本」といいます。標本で得られた結果をそのまま母集団に当てはめると誤差や偏りを生じる可能性があります。調査や研究で得られたデータを用いて母集団の特性を統計解析により明らかにする必要があります。

■　散布図

標本が持つ２つの特性値に着目し、縦軸に一方の特性値、横軸にもう一方の特性値を取り、標本の値をプロットした図を散布図と呼びます。二つの変数の間の関連を知る上で、まず散布図を描くことが重要です。散布図によって、変数間の関係性や傾向が視覚的にわかりやすく示されます。さらに「外れ値（はずれち）」と呼ばれるデータポイントも明らかになります。外れ値はたとえ１データであっても、その値が極端である場合には統計解析に著しい影響を与えます。しかし、単なる異常値として切り捨ててよいとは言い切れません。外れ値の出現には特定の原因や理由がある可能性があるため、その背景を検討することが必要でしょう。



■　検定・有意水準

検定とは、母集団の特徴について、標本に基づいて何らかの仮説を立て、その仮説が正しいと見なせるか否かを確率論に基づく方法を用いて検証することです。

検定では、まず検証の結果として棄却（否定）されることを目的とする仮説を立てます。この仮説を帰無仮説と呼びます。帰無仮説を否定することにより、帰無仮説と対立する主張が成立する可能性を明らかにする手順を踏みます。帰無仮説を否定するか否かは、仮説の下で得られる実験データの確率と、予め設定した有意水準（危険率）を比較することで判定します。有意水準とは、帰無仮説が誤りであると判断する（帰無仮説を棄却する）基準となる確率のことです。5%以下の確率で生じる現象は、非常にまれなことであると考えられるため、多くの場合、有意水準を 0.05（5%）に設定します。また有意水準は、0.05 だけでなく 0.01 もよく使われています。たとえば、有意水準を 0.05 に設定した場合、帰無仮説の下で 5%以下の確率でしか生じないデータが観測されたとき、その仮説を棄却することになります。これは、「観測された結果が偶然ではなく、何らかの理由による」と考えられるからです。

実験データと帰無仮説が矛盾すると判断された場合、統計学では「有意（siginificant）である」と表現します。有意とは平たく言えば、観察された結果が偶然でなく、何らかの理由によって起こったということを指していると解釈できます。

■　片側検定・両側検定

帰無仮説が棄却されたときに採択される仮説を対立仮説と呼びます。帰無仮説の内容が同じでも、異なる内容の対立仮説を立てることができます。

例えば、「薬 A 中の成分 B の含有量は 100mg である。」という帰無仮説に対して、次の 3 つの対立仮説を立てることができます。

ア：薬 A 中の成分 B の含有量は 100mg ではない。
イ：薬 A 中の成分 B の含有量は 100mg より多い。
ウ：薬 A 中の成分 B の含有量は 100mg より少ない。

アに対応する検定は、成分 B の含有量が 100mg であるか否かを調べる検定です。この方法では、含有量が 100mg より多い場合と少ない場合のどちらも帰無仮説の棄却理由に含まれます。

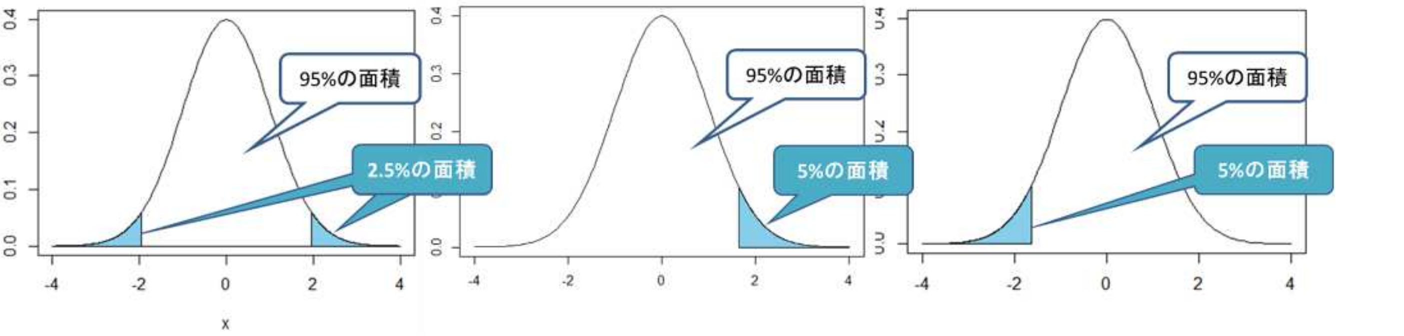
イは成分 B の含有量が 100mg より多いかどうかを調べるための方法であり、この場合 100mg より少ないか否かは問いません。

ウは逆に成分 B の含有量が 100mg より少ないかどうかを調べる検定であり、多いか否かは考慮しません。

アのような検定方法を**両側検定**、イやウのような検定方法を**片側検定**と呼びます（リーフレット〔その１〕も参照）。

有意水準を 5%とした場合、両側検定では有意水準が分散され、片側検定では特定の方向に集中します。

両側検定と片側検定の有意水準を示すと、下図のようになります。



ア　両側検定（設定値か否か）　　　　イ　片側検定（多いか否か）　　　ウ　片側検定（少ないか否か）

https://bellcurve.jp/statistics/course/9319.html?srsltid=AfmBOooQiXA_K4hNmyRwKrK_4yQQeWGjOEVk_AitLkNLG3qrV7mYIjGm

■　推定・信頼区間

推定とは、**母集団の特性値（平均や分散など）を標本のデータから統計学的に推測すること**で、点推定と区間推定の 2 種類があります。点推定では 1 つの具体的な値を推定し、区間推定では一定の区間（幅）を用いて値を推定します。

点推定は、標本データを用いて、母集団の平均や分散などの特性値を単一の値として推定する方法です。

区間推定は、母集団が正規分布に従うと仮定できる場合に、標本データを用いて母平均などの推定量を、1 つの値ではなく一定の範囲（区間）で推定する方法です。推定する区間を**信頼区間**と呼びます。この区間は、「90%信頼区間」「95%信頼区間」「99%信頼区間」などで表されます。例えば「95%信頼区間」で母平均（母集団の平均）の区間推定を行うということは、「母集団から標本を取り出して、その標本から母平均の 95%信頼区間を求める」という作業を 100 回実施したとき、95 回程度はその信頼区間内に母平均が含まれる」と考えられる、ということです。

もっとわかりやすく説明しましょう。

日本人全員の平均身長（＝母平均）が 170cm であるとします。このときに、ランダムに選ばれた 100 人の身長から 95%信頼区間を算出する実験を 100 回行います。その結果、次のように

1 個目の実験：150　$\leqq \mu \leqq$　175

2 個目の実験：162　$\leqq \mu \leqq$　172

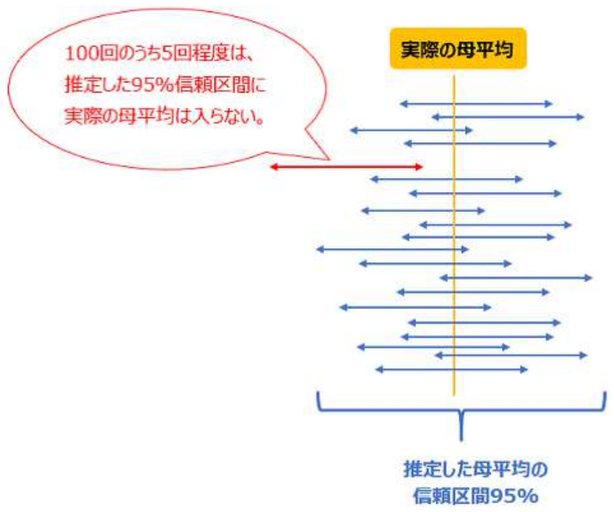
3 個目の実験：171　$\leqq \mu \leqq$　179

⋮

100 個目の実験：145　$\leqq \mu \leqq$　180

というような信頼区間が算出されるとします。

これら 100 個の信頼区間のうち、5 個くらいは母平均である 170 を含まない範囲になっていると考えられます。



https://www.stat.go.jp/naruhodo/11_tokusei/kentei.html

■　自由度

n 個の標本から求めた標本平均値において、n-1 個のデータが既知であれば、残りの 1 つのデータは自動的に値が決定されます。これは、標本平均が既知であるため、n-1 個のデータの合計から最後の 1 つのデー

タが計算で求められるためです。n-1 個のデータまでは自由に値を与えることができるので、この n-1 を自由度と呼びます。

■　母集団の統計値と標本の統計値

母集団の平均、標準偏差などの統計値は、「母平均」、「母標準偏差」のように、統計値を意味する語句の前に「母」という語句をつけて表現します。

一方、標本の平均、標準偏差などの統計値は、「標本平均」、「標本標準偏差」のように、統計値を意味する語句の前に「標本」という語句をつけて表現します。

また、母集団の分布が正規分布である場合、その母集団のことを「正規母集団」と呼びます。

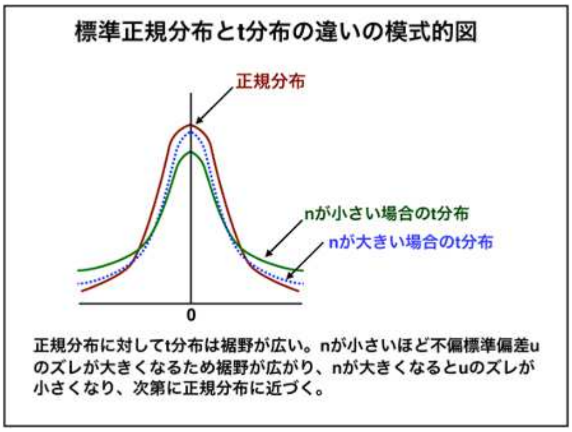
３．２　t 検定とは

■　t 分布

多くの検定では、標本の分布は正規分布に従うという前提があります。しかし、標本数が少ない場合、この前提が成り立たず、標本の分布は正規分布ではなく、よく似た形状を持つ「t 分布」に従うと考えられます。そのため、標本数が少ない場合には、t 分布を用いて検定を行います。

t 分布は、母標準偏差が未知である場合に使用されます。観測値が正規母集団から得られたと仮定して、母平均と標本平均の差を、標本標準偏差を用いて標準化（平均が 0、標準偏差が 1 となるように数値を変換すること）した値の分布を指します。

t 分布は、標本の平均値と標準偏差が分かれば計算可能であるため、平均値に関する仮説の検定で広く用いられています。特に、標本数が少ない場合や母標準偏差が分からない場合に有用です。



https://katosei.jsbba.or.jp/view_html.php?aid=1196

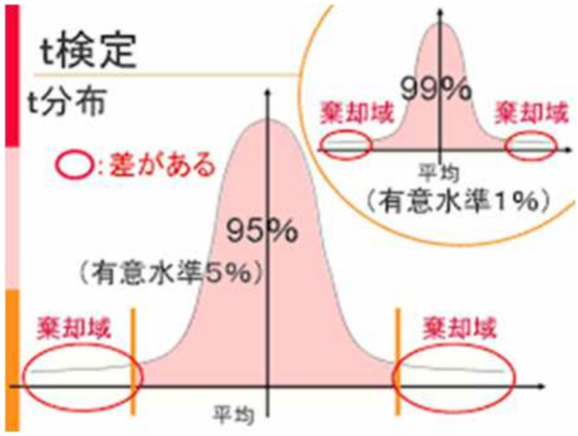
■　t 検定

2 つの独立した母集団が存在し、それぞれの母集団から抽出した標本の平均に差があるかどうかを検定することを「t 検定」といいます。

t 検定は、標本サイズが小さい場合（一般的に 10 以下）、データが t 分布に従うと仮定して、平均値等の有意差を検定する方法です。

なお、2 つのデータが「対応のあるデータ」か「対応のないデータ」かによって、検定に使用する統計量やその算出方法が異なります（リーフレット〔その１〕も参照）。

「対応のないデータ（対応なし）」とは、異なる対象から抽出された 2 つの標本を指します。「対応のあるデータ（対応あり）」とは、同一の対象から抽出された「対」となる 2 つの標本を指します。



https://analytics.creaman.com/2018/12/18/t%E6%A4%9C%E5%AE%9A%E3%81%A8%E3%81%AF/

対応なしの場合の帰無仮説は、「母平均が等しい」となります。
対応ありの場合の帰無仮説は「対応のあるデータの差の平均は 0 である」となります。
対応なしの場合、標本データからの t 値（絶対値）の算定式は、
　　絶対値（標本平均－母平均）／（標本標準偏差／平方根（サンプル数））
対応ありの場合の標本データからの t 値（絶対値）の算定式は、
　　絶対値（条件 1 の標本平均－条件 2 の標本平均）／（標本標準偏差／平方根（サンプル数））
で求めることが出来ます。

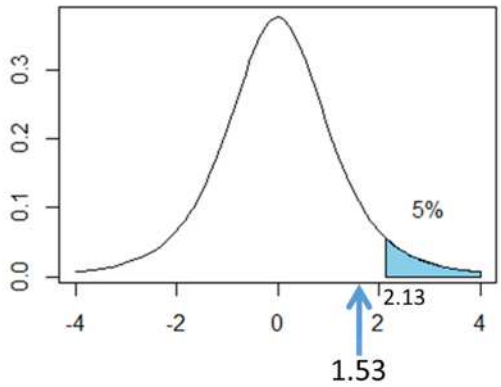
なお、自由度と有意水準が判れば、その時の棄却域の境界値は t 分布表から求めることができます。

それでは、事例に基づいて、具体的に色々な検定方法を解説します。

■　対応があるｔ検定とは

対応があるｔ検定とは、同一の対象から得られた２つのデータ群の差を検定する手法です。

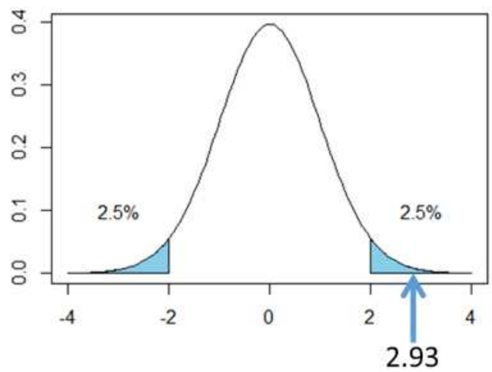
実験参加者５人の薬の投与前後の血圧データを用いて、投薬によって血圧が下がったかどうかを検定する例を解説します。この場合は測定した血圧のデータは対応があるデータとなるため、測定した血圧の２条件の差が０かどうかを検定します。また、この検定では血圧の低下だけを問題としているので、片側検定を行います。



帰無仮説は「投薬前後の血圧は等しい（投薬によって血圧は下がらなかった）。」とします。対立仮説は「投薬によって血圧は下がった。」となります。有意水準は 0.05 と設定します。自由度はデータ数−1 なので、5−1=4 となります。

標本データを用いて算定したｔ値は約 1.53 でした。このｔ値を、自由度４のｔ分布における 5%棄却域境界値と比較します。自由度４のｔ分布における 5%棄却域境界値は 2.13 です。算定したｔ値（1.53）はこれを下回っています。ｔ値が棄却域に入っていないため、「有意水準 5%において、帰無仮説は棄却されない」ということになります。したがって、「投薬によって血圧が下がったとは言えない」という結論になります。なお、同じ実験参加者による投薬前後の２つのデータを使用しているため、この実験は**実験参加者内計画**に基づくものです。

■　対応がないｔ検定とは



対応がないｔ検定とは、異なる対象から得られた２つのデータ群の平均値の差を検定する方法です。

ある学校の１組と２組の算数テストの平均点を比較して、それぞれの平均値に差があるかどうかを検定する例を解説します。この場合、平均点に差があるかどうかを調べることが目的であるため、両側検定を用います。

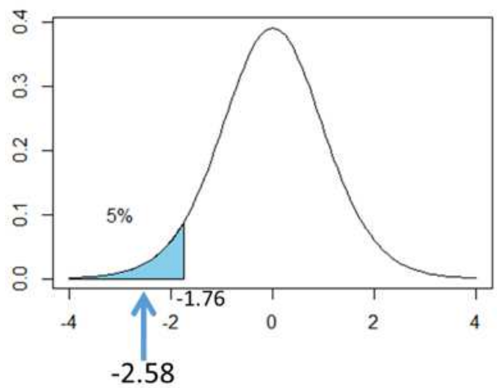
１組の生徒 30 人の平均点は 75 点、標準偏差は５点、２組の生徒 32 人の平均点は 70 点、標準偏差は８点でした。

帰無仮説は**「1 組の算数のテストの平均点と２組の算数のテストの平均点に有意な差はない。」**、対立仮説は「1 組と２組の算数のテストの平均点には有意な差がある。」となります。有意水準を 0.05 に設定します。

標本データから算定したｔ値は約 2.93 でした。自由度はそれぞれの標本サイズから計算され、(30−1)+(32−1)=60 となります。自由度 60 のｔ分布表を見ると、有意水準 0.05 を両側検定で分配すると片側 0.025 となり、その棄却域の境界値は 2.00 です。算出したｔ値（2.93）はこの境界値を上回っており、棄却域に含まれることから、「有意水準 5%の両側検定において、帰無仮説は棄却され、対立仮説が採択される」と判断できます。つまり、「1 組と２組の算数のテストの平均点には有意な差があり、１組の平均値の方が高い。」と結論付けられます。

なお、この検定では異なる参加者のデータを取得、解析しているので、この検定は**実験参加者間計画**に基づくものとなります。

■　母平均の検定



あるメーカーの電球は寿命が 2,000 時間（この値が母平均）であると宣伝しています。この性能が本当であるかを検定するため、様々なお店で合計 15 個購入し、テストを行ったところ、寿命の平均値（標本平均）は 1,900 時間、標本標準偏差は 150 時間でした。

この結果を基に、この電球の寿命はメーカーの宣伝どおりの性能であるか検定します。

帰無仮説は「この電球の寿命は 2,000 時間である」と設定します。電球の寿命が 2,000 時間よりも長い場合は特に問題視しない

ため、対立仮説は「この電球の寿命は2,000時間よりも短い」とします。有意水準は0.05と設定します。自由度は標本数から1を引いて15-1=14となります。

標本データを用いて算定したt値は約-2.58でした。この値を、自由度14のt分布に基づく棄却域の境界値と比較します。有意水準0.05の片側検定における自由度14のt分布での棄却域の境界値-1.76でした。算定されたt値（-2.58）は、この境界値を下回っています。従ってt値は棄却域に含まれることになります。この結果、「有意水準5%において、帰無仮説は棄却され、対立仮説が採択される」と判断できます。こうして、「この電球の寿命は2,000時間より短い。」と結論付けられます。

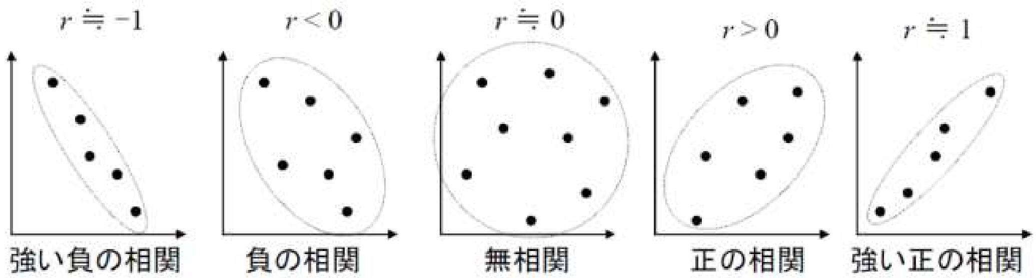
３．３　相関分析は慎重に

■　相関分析

標本を用いた統計解析として最初に行うことが多いのが「相関分析」です。

相関分析とは、２つの変数の相互関係を統計的に分析する手法です。具体的には、ある変数が変化したときに、もう一方の変数がその変化に応じてどのように変化するかを分析します。このような関係を「相関関係」と呼びます。

相関分析では、標本を用いて計算される標本相関係数を求めます。この相関係数は、2つの変量間に線形関係がどの程度あるかを数値で示しており、−1から＋1の値をとります。値が＋1に近いほど正の相関が強く、値が−1に近いほど負の相関が強いことを意味します。一方、値が0に近い場合、2つの変量の間に線形関係がほとんどないと解釈されます。



https://www3.cuc.ac.jp/~nagaoka/2011/ouyou/10/expr/index.html

■　相関係数の検定

標本で相関が認められる場合、母集団でも同様に相関があるか否かを確認する必要があります。この手法を相関係数の検定と呼びます。この検定の際には、帰無仮説として、「母集団の相関係数は0である（つまり、母集団では相関がない）。」と設定するのが一般的です。検定の結果、帰無仮説が棄却される場合、母集団の 2 つの特性値の間の相関係数は 0 であるとは言えないことになり、相関がある可能性が高いと判断されます。

相関係数の検定結果は、たとえば、「有意水準5%では、標本相関係数が有意である」というように表現します。通常、有意水準として1%、5%の値が採用されます。有意水準の値が小さいほど、厳しい判断基準を採用していることを意味し、検定もより厳密なものになります。

多くの統計解析ソフトでは、標本相関係数を算定する際に、同時に相関係数の検定結果まで出力されるようになっています。そのため、標本データから相関の有無を判断する手順が効率化されています。

■　相関係数の区間推定

標本相関係数を用いて母集団の相関係数を推定する際、点推定だけでなく区間推定することも必要です。ただし、母集団の相関係数を区間推定する具体的な方法は非常に複雑な数式を伴うため、ここでは途中の計算過程を省略し、結論を示す図表を用いて説明します。

標本のサイズと標本相関係数から、母集団の相関係数の信頼区間を求めることができます。図中の線分は標本サイズに対応しています。

例えば、標本のサイズが100で標本相関係数がr=0.7の場合、母集団の相関係数の95%信頼区間は0.60～0.80となります。同様に、99%信頼区間では0.41～0.74です。このように、信頼区間は設定された信頼度（95%や99%など）と標本サイズによって決まります。

また、**標本サイズが小さいほど**、**母集団の相関係数の下限値と上限値の幅が大きくなる**傾向があります。これは、標本サイズが小さい場合、標本相関係数に基づいて母集団の相関係数を推定する際の不確実性が高くなるためです。この結果から、標本サイズが推定精度に大きな影響を与えることが分かります。

信頼区間を求める際には、専用の計算ツールや統計解析ソフトを活用することで、迅速かつ正確な推定が可能です。これらのソフトウェアは、信頼区間の幅や精度に関する傾向も視覚的に示すことができるため、実務的な解析において有効です。

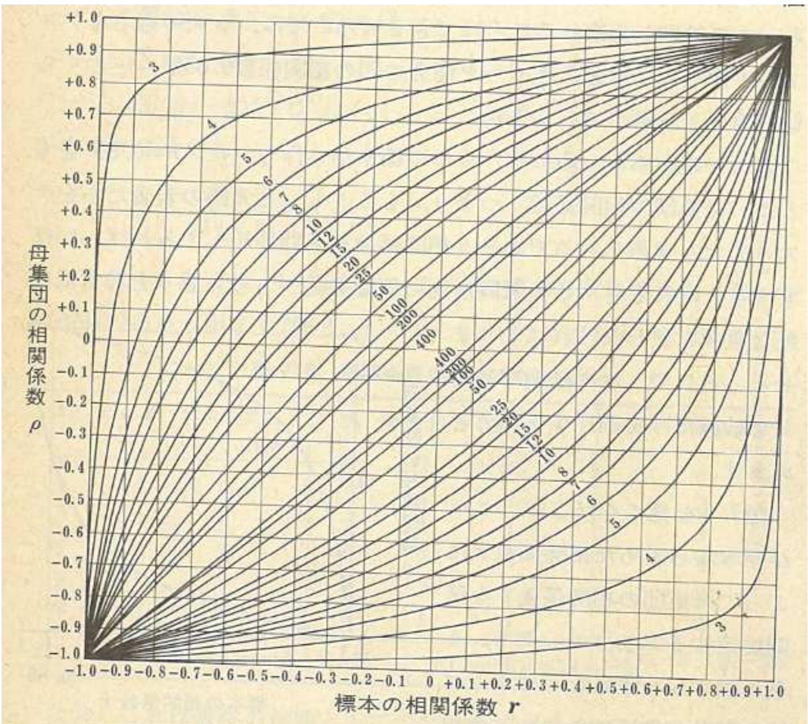


図 7.5 母集団の相関係数を区間推定する
(95%信頼区間)

大村　平：統計解析のはなし、p.212、日科技連出版社、1982 年 11 月

■　有意な相関関係と判断する相関係数の基準は？

相関分析を行い、有意水準 1%で 0.18 という相関係数が得られたとしても、この相関係数は非常に小さいため、「相関はほとんど見られない」と解釈されます。

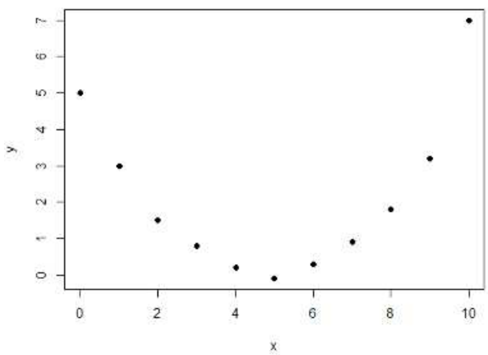
相関係数の値の解釈については、概ね以下の様に分類されます。

0～0.3 未満：ほぼ無関係
0.3～0.5 未満：非常に弱い相関
0.5～0.7 未満：相関がある
0.7～0.9 未満：強い相関
0.9 以上：非常に強い相関

この分類に基づけば、相関係数 0.18 は「ほぼ無関係」と評価される範囲内であると判断されます。

ただし、右に示した散布図のように、2 つの変量間の相関係数が 0.18 程度であっても、変量間に放物線のような 2 次的な関係が認められる場合があります。この場合、相関係数は 2 つの要素の直線的な相関関係の強弱を表すものであるため、線形ではない相関関係の強弱は正しく表すことができないことに注意してください。2 次関数で表せるような非線形な関係では、相関係数が低い値を示していても、両者の間に強い関係が存在する可能性があります。

このような理由からも、相関分析を行うときには、まず、散布図を描くことが基本です。

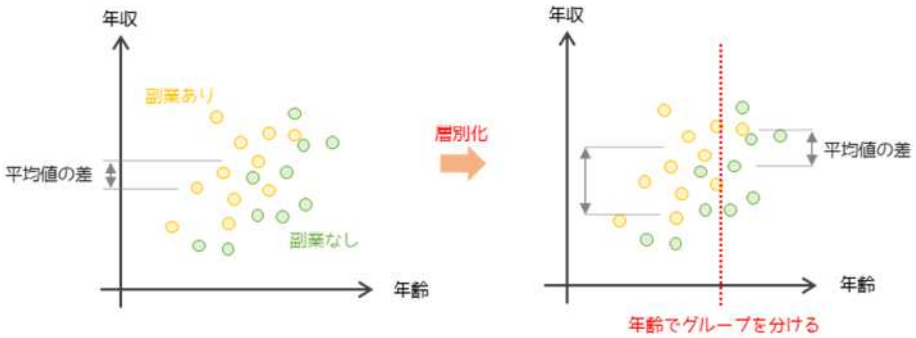


■　母標本の特性を正しく把握して層別相関分析を

同じ傾向を持つと仮定したグループが、実際には異なる傾向を複数持っている場合、「層別相関」という現象が見られることがあります。このような場合、グループ全体での相関分析だけでは十分な情報が得られないため、より精確な分析手法として「層別相関分析」が必要となります。

例えば、夏の気候条件が同時期のエアコンの需要や売上に与える影響は、地域ごとに異なる可能性があります。夏に湿度の高い地域ではエアコンの売上が増加する一方、乾燥した地域における需要は湿度の高い地域における需要よりも低いかもしれません。この場合、気温とエアコンの需要の相関分析を行うと、散布図が 2 つのプロット群に分かれる、あるいは 2 つのプロット群が一部オーバーラップする可能性があります。このような場合、層別相関分析を用いて、地域ごとに気温とエアコン需要の関係を評価する方が、より妥当な分析方法となります。

もう少し、具体例を挙げて説明します。年収と年齢の関係を明らかにしようとする研究を考えます。この研究で、副業をしている人と、していない人が標本内に混在していると仮定します。散布図から推定できるように、両者の平均値の間には差がありそうです。この場合、対応がない t 検定を行い、有意差を明らかにすることができれば、副業の有無が年収に影響を及ぼしていると考えられます。ただし、この検定では、年齢による年収の増加という要因が考慮されていません。副業の有無による年収の差違を明確にするためには、年齢による年収の増加の影響を排除する必要があると言えます。そこで、年齢により標本を分けて（標本を年齢で層別化して）、それぞれの層で副業の有無と年収との関係を相関分析することが有効です。こうすることで、年齢の影響を排除した上で、副業の有無による年収の差をより明確化することができます。



（年齢によるグループ分けをしていない）　　　（年齢によるグループ分けをした）

https://corvus-window.com/whats_stratified-analysis/

３．４　３水準以上は分散分析を

■　３水準以上の場合 t 検定は不適

「かおりＡ、かおりＢ、かおりＣ、無臭の 4 つのにおい条件（第 1 要因）における注意維持課題の成績を検討する。」という実験を考えます。この実験に、年齢層（若齢者層と高齢者層）という第 2 要因を追加し、比較を行う場合を考えてみましょう。1 つ目の要因は 4 種類のかおり条件（無臭も含む）、2 つ目の要因は年齢層であり、これらが独立変数として設定されます。なお、従属変数は注意維持課題の成績とします。

この実験では、2 つの要因（かおり条件と年齢層）について、いずれも参加者間計画に基づいて分析を行います。例えば、それぞれ 20 名の参加者を設定した場合、若齢層でかおりＡを嗅ぎながら注意維持課題を行う 20 名、若齢層でかおりＢを嗅ぎながら注意維持課題を行う 20 名、以下かおりＣと無臭の条件で各 20 名、そして、高齢者層でも同様に各条件で 20 名ずつが設定されます。合計で 8 群となり、参加者総数は 20×8＝160 名になります。

以上の 8 群を 2 群ずつ抽出した場合、28 通り（${}_8C_2$）の組み合わせがあります。しかし、多くの公表されている論文では、これら全ての 2 群の組み合わせについて t 検定を行うケースが意外にも散見されます。この場合、それぞれの母集団の特性を考慮して、年齢層の違いや、かおり条件による差異を検定するだけにとどまり、注意維持課題の成績に対する 2 つの独立変数（要因）の相互影響を明らかにしていません。そのため、このような分析方法は実験計画に適切に対応した分析とは言えません。

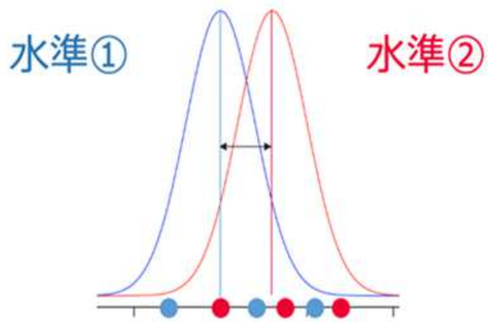
ここで適切な分析は、二元配置分散分析（参加者間計画）です。1 つ目の要因を「かおり条件」とし、その水準を 4 種類（かおりＡ、かおりＢ、かおりＣ、無臭）に設定します。2 つ目の要因は「年齢層とし、その水準を 2 種類（若齢者層と高齢者層）に設定します。

一方で、もし、4つのかおり条件（かおりＡから無臭まで）を同じ参加者が全て経験している場合は、「かおり条件」は参加者内計画、「年齢層」は参加者間計画となります。この場合、適切な分析方法は、二元配置分散分析（混合計画）です。混合計画では、参加者間計画と参加者内計画の両方を考慮した計算方法が用いられます。

参加者間計画、参加者内計画（t検定では対応なし、対応あり）、混合計画では、それぞれ分析の計算方法が異なるため、適切な方法を選択することが重要です。実験の設計に応じて正しい分析手法を選択しないと、結果の解釈を誤る可能性があります。計画に基づいた分析手法を慎重に選択しましょう。

■ https://toketarou.com/anova/**分散分析はF分布を用いて**

分散分析とは、比較する群が3つ以上である場合や、1つのデータに複数の要因（水準）が含まれる場合に、母平均の差を検定するための統計解析手法です。この手法では、群ごとのデータのばらつき（分散）を基にF分布を用いて検定を行います。



https://toketarou.com/anova/

分散分析の基本的な考え方は、「群間のばらつき（水準同士のばらつき）」と「群内のばらつき（標本のばらつき）」を比較し、その比率に基づいて母平均に差があるかどうかを明らかにすることです。右図のように「水準同士のばらつき」に比べて「標本のばらつき」が大きい場合、標本だけを観察しても母平均に差があるか否かを結論づけることができません。そこで、分散分析では群間のばらつきと群内のばらつきの比率（F値）を計算し、これに基づいて検定を行います。

F分布とは、2つの群の分散の比に基づいて計算されるF値が従う確率分布です。F分布の形状は自由度により異なります。そのため、検定を行う際には、2つの自由度（分子の自由度mと分母の自由度n）を指定する必要があります。検定は、これらの自由度に基づいた$F(m,n)$の分布を用いて実施します。

分散分析は、母平均の差を評価する際に3群以上や複数の要因を含むデータセットに対して非常に有効な手法であり、t検定の拡張版とも言えます。複雑なデータ構造を正確に解析するためには、分散分析を正しく理解し、適切に適用することが重要です。

■ **一元配置分散分析**

あるコンビニの店舗で、商品Aを陳列棚の上から1段目、2段目、3段目のいずれかに並べた場合、1日に売れる個数に差があるかを検定します。この分析では、一元配置分散分析を用いて検定を行います。無作為に選んだ9店舗の1日の売上個数をまとめた結果が、右表に示されています。

1段目	2段目	3段目
15	18	15
21	24	16
18	21	14

https://toketarou.com/anova/

帰無仮説を「段数による売上げ個数の差異はない。」と設定します。一方、対立仮説は、「段数3つのうち少なくとも1つで売上個数が異なる。」とします。

標本データに基づき計算した水準間平方和、平均平方、および残差平方和からF値を求めたところ、その値は4.26でした。この場合のF分布の自由度は、**分子の自由度（水準間自由度）が、水準の数が3つ（3段）ですので3-1=2**、**分母自由度（残差自由度）が、標本の総数9から水準の数3を引いた9-3=6**になります。

算出したF値4.26をF分布における$F(2,6)$の上側5%点である約5.14を比較します。$4.26<5.14$であるため、棄却域には含まれません。よって、「帰無仮説を棄却することはできない。」と判断されます。この結果より、「母平均に差があるとは言えない、すなわち、段数による売上げ個数の差は認められない。」という結論になります。

なお、水準間平方和、残差平方和、平均平方の定義ですが、以下のとおりです。

・水準間平方和（Sum of Squares Between, SSB）
　各水準の平均値と全体の平均値の差を基に計算される平方和で、水準間のばらつきを表します。

・残差平方和（Sum of Squares Within, SSW）
　各データ値とその所属する水準の平均値の差を基に計算される平方和で、群内のばらつきを表します。

	肥料 100g	肥料 200g	肥料 300g	肥料 400g
土 A	14.5	16.5	17.8	18.1
	15.1	16.1	19	20.2
	14.1	15	15.2	17.2
土 B	16.2	18.6	21.7	23.6
	15.3	16.9	20.5	24.9
	17.5	18.6	19.4	25.5

・平均平方（Mean Square, MS）

水準間平方和または残差平方和を、それぞれの自由度で割った値です。分散分析では、水準間平均平方と残差平均平方の比率（F 値）を算出して検定を行います。

この例題では、1 段目の平均は 18、2 段目は 21、3 段目は 15 です。全体の平均は 18 となります。

水準間平方和(SSB)

$3(18-18)^2+3(21-18)^2+3(15-18)^2=54$

残差平方和(SSW)

$(15-18)^2+(21-18)^2+(18-18)2+(18-21)^2+(24-18)^2+(21-21)^2+(15-15)^2+(16-15)^2+(14-15)^2=38$

平均平方(MS)

水準間平方和／水準の自由度=54/2＝27

残差の平均平方（MSW）

残差平方和／標本の自由度＝38/6≒6.33

F 値

水準の平均平方／残差の平均平方=27/6.33≒4.26　　　　となります。

■ 二元配置分散分析

二元配置分散分析は、2 つの要因がデータに与える影響を分析する手法で、それぞれの要因ごとの水準間で平均値の差を検定するとともに、2 つの要因間の交互作用効果を検討することができます。

右表は、肥料の量（4 パターン）と土壌の種類（2 パターン）という 2 つの要因が、作物の収量に与える影響を調べた実験結果です。

このデータを用いて、肥料の量、土壌の種類、肥料の量×土壌の種類（交互作用）のそれぞれにおいて、収量の平均値に差があるかを二元配置分散分析で検定します。

・帰無仮説：

各要因の水準間で作物の収量の平均値は等しい。

・対立仮説：

少なくとも 1 つの水準間で収量の平均値に差がある。

	肥料 100g	肥料 200g	肥料 300g	肥料 400g
土 A	14.5	16.5	17.8	18.1
	15.1	16.1	19	20.2
	14.1	15	15.2	17.2
土 B	16.2	18.6	21.7	23.6
	15.3	16.9	20.5	24.9
	17.5	18.6	19.4	25.5

【自由度の計算】

・全体の自由度：データの総数から 1 を引く。

24-1=23

・肥料の量に関する自由度：水準数から 1 を引く。

4-1=3

・土壌に関する自由度：水準数から 1 を引く。

2-1=1

・交互作用に関する自由度：2 つの要因の水準数から 1 を引き、それらを掛け合わせる。

(4-1)×(2-1)＝3

肥料の量と土壌の種類の 2 つの要因が組み合わさることで初めて現れる効果を交互作用と呼びます。

・残差の自由度：全体の自由度から上記の自由度（肥料の量に関する自由度・土壌に関する自由度・肥料の量）を引く。

23-3-1-3=16

以上に基づき、肥料の量に関する検定では F 分布表の F(3,16)を、土壌に関する検定では F(1,16)を、肥料の量×土壌の種類(交互作用)に関する検定では F(3,16)を参照します。

【F 値による検定】

F 値（標本データから算出した F 値、以下「f 値」という）は、各要因の平均平方／残差の平均平方で算出されます。

・肥料の量：f 値 29.49　＞　有意水準 5%の F 値 3.239

→帰無仮説は棄却され、「肥料の量によって収量の平均値に差がある。」と結論付けられます。

・土壌の種類：f 値 46.33　＞　有意水準 5%の F 値 4.494

→帰無仮説は棄却され、「土の種類によって収量の平均値に差がある。」と結論付けられます。

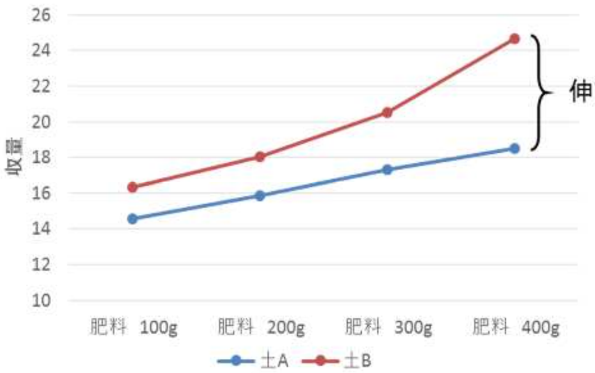
・肥料の量×土壌の種類（交互作用）：f 値 4.14　＞　有意水準 5%の F 値 3.239

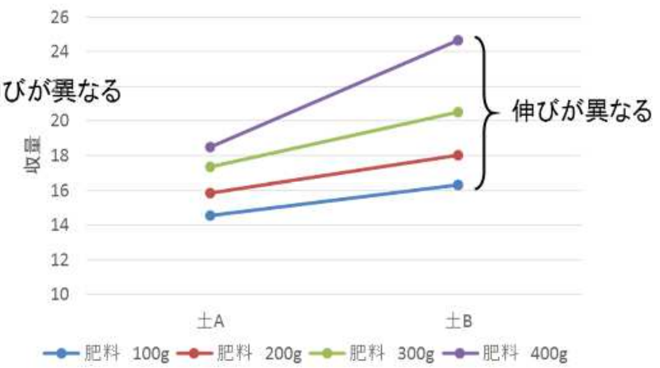
→帰無仮説は棄却され、「土壌の種類により、肥料の量が収量へ及ぼす効果が異なる。」と結論付けられます。

肥料の量×土壌の種類による交互作用を理解するために、土壌の種類×収量の平均値のグラフを下図に示します（肥料の量×収量の平均値のグラフでも可）。土 A も土 B も肥料の量が増えるにつれて収量が増加します。しかし、土 B のほうが土 A よりも収量の増加幅が大きいことが分かります。これは、肥料の量に対する収量の変化が土壌の種類によって異なることを示し、この現象を「交互作用効果」と呼びます。交互作用が有意な場合、「収量に対する肥料の量と土の種類の効果は一様ではない」と結論付けられます。

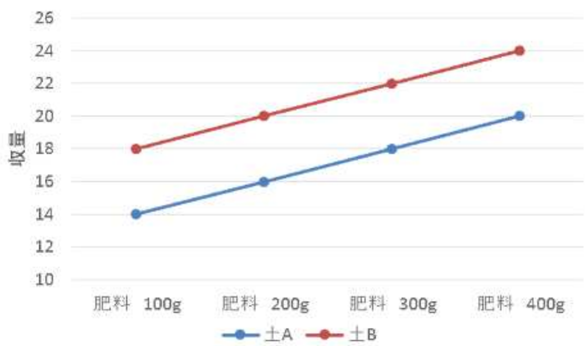
一方、交互作用が有意ではない場合、交互作用は図のように平行な線となります。これは、「肥料の量が増えるに連れて収量が増加するものの、その増加幅は土の種類によって変わらない」ことを意味します。

この場合、「肥料の量×土壌の種類」に関する帰無仮説（「肥料の量と土壌の種類は相互に影響を及ぼさない」）を棄却しないという検定結果が得られます。つまり、「肥料の量と土壌の効果は独立しており、相互に影響を及ぼしていない」と結論付けられます。





https://bellcurve.jp/statistics/course/10090.html



■　多元配置分散分析について

3 つ以上の要因を含むデータを分析する必要がある場合、適切な方法は多元配置分散分析です。この手法では、複数の要因の主効果や交互作用効果を同時に検討することが可能です。ただし、交互作用の解釈が増えるため、分析結果の解釈はより複雑かつ難解になります。

■　分散分析の主要な結果表記は「有意差（ゆういさ）」ではない？

分散分析の適切な検定結果の表記としては、

- **「〇要因の主効果が有意である」**
- **「〇要因の主効果は有意でない」**
- **「〇要因と△要因の交互作用が有意である」**

という表現を用います。散見される誤りとして、「～の要因に有意差が見られた」という表記があります。これは適切な表現ではありませんので注意してください。

分散分析の結果に基づいて多重比較を行った場合には、その後の多重比較で「ＡとＢの間に有意差がみられる」という表現が用いられることがあります。これは多重比較における特定の水準間の差を説明するための表現であり、分散分析そのものの検定結果では用いられません。

【二元配置分散分析の結果例】

前述の「肥料の量と土壌の種類による収量の違い」に関する二次元分散分析の結果を、適切に記述した例を示します。

・肥料の量の主効果は有意である。

・土壌の種類の主効果は有意である。

・肥料の量と土壌の種類の交互作用が有意である。

■　要因数の多い分散分析でも交互作用が有意でなければ、多重比較が容易

「世代（若齢層、高齢層）」、「においの種類（A、B、C、D）」、「においの濃度（高、中、低）」の要因の主効果が全て有意であり、これら 3 要因の間の交互作用が全て有意ではなかったとします。

さらに多重比較で、次のような結果が得られたとします：

① 世代により感覚強度評定値が、　若齢層　＞　高齢層

② においの種類により感覚強度評定値が、　Ａ　＞　Ｂ　＞　Ｃ　＞　Ｄ

③ におい濃度により感覚強度評定値が、　高　＞　中　＞　低

すなわち、

1）若齢であるほど感覚強度が高い。

2）においが高濃度であるほど感覚強度評定値が高い。

3）においの種類についてはＡが最も高い評定値であり、Ｂ、Ｃ、Ｄと順に低くなる。

これらの解釈は、要因間の交互作用を考慮する必要がないため、各要因の効果を個別に評価できることに起因します。

■　要因数の多い分散分析で交互作用が有意である場合

交互作用が有意になる場合は、結果の解釈が非常に複雑になるケースが多くあります。

上の事例で、例えば、

・「世代」および「濃度」の主効果は有意である。

・「においの種類」の主効果は有意ではない。

・「世代」と「においの種類」の交互作用が有意である。

という結果が得られたとします。

この場合、世代によって、どのにおいの感覚強度評定値が高くなるのか、あるいは低くなるのか、一定の傾向が見られないことが示唆されます。

具体例として、高齢層の「においＣ」に対する感覚強度評定値が若齢層の評定より高い一方で、「においＤ」では逆に若齢層の評定が高いという結果が考えられます。要因間の複雑な相互影響を十分に理解しないまま解釈を試みると、結果が妥当であるかどうかを判断するのが困難になります。

交互作用が全て有意になってしまう場合、解釈の多様性は一気に増加し、もはや何が妥当な解釈なのかは、まぐれあたりのような確率になってしまいます。すなわち、結果が偶然性に左右される確率が高まることが示唆されます。

以上の事由から、**分散分析を想定した実験は、要因の数を増やしすぎないこと、一般的には、2 要因くらいまでの実験計画が推奨されます**。3 要因を設定した場合は、交互作用が有意になると解釈の多様性が高まるリスクを十分認識し、慎重に実験を進める必要があります。複雑な交互作用を回避するためには、明確な仮説を立て、要因間の関係性を適切にモデル化しておくと良いでしょう。

４．　結論（考察）の留意点

■　多くの研究は演繹的な研究

研究には、**帰納的アプローチ**と**演繹的アプローチ**の 2 つの方法があります。演繹的アプローチは**仮説検証型アプローチ**と呼ぶこともできます。

通常、科学論文誌に掲載されるのは後者で、仮説が明確に提起され、それが検証されるというスタイルです。しかし、まず、ある現象をつぶさに観察し、帰納的に情報を蓄積した上で、そこから仮説を提起していくことが自然な流れでと言えるでしょう。その意味で、研究者がこの 2 つのアプローチを行き来しながら研究を進めるべきであることについては、多くの専門家が同意するところでしょう。

ここで、当然のように、「～すれば・・・なる。」という仮説は、基本的に**「因果関係（いんがかんけい）」の推論**に基づくものです。ヒトは言語の発達によって、現象の背後に因果関係を見いだし、法則性の仮説を立ててしまう癖があるようです。こうした思考の癖は多くの発見を生む原動力となる一方で、注意が必要です。特に、**相関関係を因果関係と錯誤**してしまう危険性が伴います。人間が因果関係を推論する特性については、「分離脳研究」で明らかにされています。この研究では、脳が情報を補完し因果関係を推論する際の特性が示されています。具体的には、脳の左右半球が分離された状態でも、ヒトは一方の情報を基に補完的な物語を作り、

因果関係を推測します。こうした特性が、人間の仮説形成能力の基盤であると同時に、誤った推論のリスクを内包していることが分かります。

■　相関関係と因果関係とは別物

統計分析において、相関関係と因果関係は異なる概念であることを理解することが重要です。以下に具体例を挙げて説明します。

都内Ａ市にあるアイスクリーム屋は順調に売り上げを伸ばしており、**毎月の売り上げは、線形に上昇しているとします。**このデータと、Ａ市にある幼稚**園児の身長に関する毎月の推移データ**を相関分析にかけてみましょう。

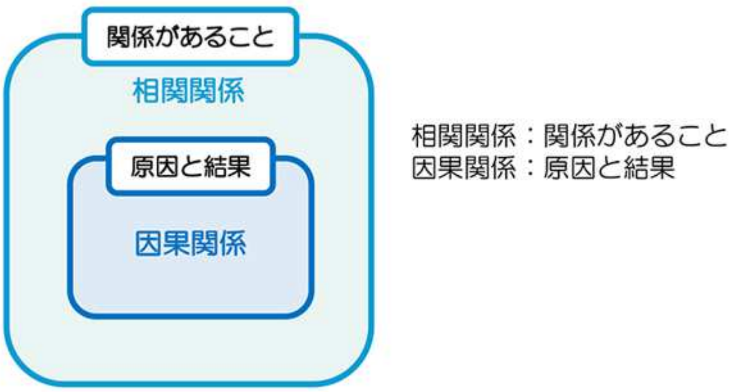
この場合、**両者とも月毎に値が上昇しているので、非常に高い相関係数**が得られます。さらに、t 検定や分散分析、回帰分析を用いても、有意な効果が検出されるでしょう。

しかし、この相関が意味するのは、両者が「同じような増加傾向を持つ」という事実だけです。アイスクリームをたくさん食べた幼稚園児の身長が高くなったと結論付けられないばかりか、園児の身長が高くなり食欲が増した結果、アイスクリームをたくさん食べるようになったと結論付けられるわけでもありません。

そもそも、これらのデータは因果関係を検証するために収集されたものではなく、単に月ごとに増加する現象の増減傾向を調べたに過ぎません。相関関係とは、2 つの事象の間に、片方が高いとき他方も高い、あるいは片方が高いとき他方が低いという関係性を示すものです。**事実関係を作り出す因果応報のからくりや事実関係の裏に潜む要因**などは、全く関与していない数字だけを取り扱っているに過ぎないことを肝に銘じる必要があります。

因果関係が存在する場合には相関が高くなることがありますが、逆に**相関が高いことが必ずしも因果関係の存在を意味する訳ではありません。**さらに、因果関係を想定した統計的手法（例：回帰分析や分散分析）を用いた場合でも、それによって因果関係が立証されたと考えるのは適切ではありません。因果関係を想定するのは研究者自身の仮説であり、統計的分析手法そのものではないからです。

回帰分析や分散分析で有意な結果が得られたとしても、それらは基本的に相関関係を示すものに過ぎないことを覚えておく必要があります。統計的手法を正しく理解し、因果関係を示すためには、適切なデータ収集と設計が不可欠です。



https://ez2understand.ifi.u-tokyo.ac.jp/terms/terms_24/

■　因果関係が示唆される条件とは

相関関係が見られた場合、因果関係を立証するためには、以下のような条件を満たす必要があります。

・時系列の確認：

時系列においてどちらの事象が「先」に起こったかを明らかにすることが重要です。例えば、二酸化炭素濃度の上昇が地球温暖化に先んじて発生した場合、二酸化炭素濃度の増加が地球温暖化の原因である可能性が示唆されます。一方で、二酸化炭素濃度の上昇に先んじて地球温暖化が起こった痕跡があれば、両者の因果関係の仮定は成立しません。

・理論的な関連付け：

因果関係を仮定するには、両者を結びつける理論的な根拠が必要です。大気中の二酸化炭素濃度と地球温暖化の事例に関しては、**理論的な関連付けをすることができます**。地球の表面は大気を透過した太陽の光（短波長の電磁波）により暖まり、表面の熱は逆に赤外線（長波長の電磁波）として宇宙空間に放射されます。二酸化炭素が温室効果ガスであることは既に明らかにされていますが、温室効果ガスは赤外線（長波長の電磁波）を吸収して、地球表面へ再放射する性質があるため、地表の熱が宇宙空間へ放射され難くなります。このため、地表を暖める働きが強くなり、地表面付近の温度が上昇する地球温暖化が起こる可能性が高いと言えます。

実際、大気中での高さ方向の赤外線の遣り取りについて地球の気温が決まることを計算モデルで明らかにし、大気中の二酸化炭素濃度が 2 倍に増えると地表付近の温度が 2℃程度上昇するという計算結果が得られています。この計算結果は、理論的には二酸化炭素濃度の増加が地球温暖化をもたらす、すなわち因果関係の可能性を裏付けるものとなります。

・他の影響要因の除去：

因果関係を明確にするためには、他の可能性を除外する必要があります。例えば、地球温暖化を説明する際には、**二酸化炭素濃度以外の影響要因の存在についての検討**も重要です。これには、太陽活動、氷河期と間氷期のサイクルといった自然要因が挙げられます。大気中の二酸化炭素濃度の上昇を考慮しない自然要因のみの気候シミュレーションモデルでは、1000 年間、大幅な気温の上昇が認められませんでした。一方、太陽活動, 火山の噴火などの自然要因と人為要因（温室効果ガス等を考慮）を入力したシミュレーションでは、現在観測されている気温上昇を説明する結果が得られました。

このシミュレーション結果から、現在の地球温暖化には人為要因が強く関与していることが示唆されます。

・データの普遍性：

因果関係を立証するには、観測結果が特定の状況や場所に限られないこと、つまりデータの普遍性も重要です。先述のアイスクリーム屋の売上げと幼稚園児の身長との関係について、A 市に限らず、東京都全域において同様の現象が観測される場合、両者の間には何らかの関係がある可能性を検討する価値があるかもしれません。

また、ある一人の研究者の統計解析の結果だけでなく、**複数の研究者が同じような統計解析結果**を得ている場合、その因果関係の可能性はさらに高まります。

例えば、地域や集団を対象に、病気の原因と考えられる要因と病気の発生の関連性を統計的に調査する疫学調査では、データの普遍性に基づいて疾病とその原因との因果関係を明らかにしています。

5. 結びに

におい・かおりに関わる研究の中で、ヒトを対象として何らかの測定・評価を行う際に必要とされる基本的知識を、心理実験の初心者を想定してできるだけ易しく解説しました。

心理学的アプローチにハードルを感じることなく取り組んでいただきたいという主旨で、できる限り平易な表現に努めました。にもかかわらず、心理学者ゆえの性分ゆえ、理屈っぽく分かりにくい箇所があったかと思います。この点については、自戒も込めて認識しています。

ヒトの行動の研究、心理学的検討においては、繊細かつ周到な実験計画をはじめ、**多岐にわたる留意事項**への配慮が必要です。本リーフレットを通して、こうしたポイントの重要性についてご理解いただけたのではないかと思います。

独創的な実験計画は、ヒトの特徴的な傾向の解明に寄与します。また、巧みな実験統制は、ヒトの興味深い特徴を明瞭に「炙り出す」手段となります。このリーフレットで解説した内容を参考に、実験計画を立案していただければ幸いです。

もし、内容に関して理解が難しい箇所があれば、関連する参考書を掲載しますので、是非、一読ください。

それでも疑問が残る場合には、**におい・かおり環境協会**までご連絡ください。このリーフレットは完成形ではありません。皆さまからのご意見、ご質問に基づいて、さらに改訂を重ねていく予定です。

　統計学に関する参考書はたくさん出版されていますが、難しい数式を前面に出さずに判りやすく丁寧に解説しているという評判の書籍を紹介します。

石井俊全：意味がわかる統計学、ペレ出版、2012 年 1 月
涌井貞美：意味がわかる統計解析、ペレ出版、2013 年 2 月
石井俊全：意味がわかる多変量解析、ペレ出版、2014 年 6 月