1. 統計改革と検定力

大久保・岡田(2012). 第一章 心理統計における新展開：統計改革がはじまった

■ Wilkinson & APA Task Force on Statistical Interference (1999)

データの分析において、

（１）必要最小限の分析を選択する

（２）信頼区間（CI）を使用した区間推定を行う

（３）重要な知見あるいはp値を報告するときは必ず効果量を報告する

（４）分析の前提が満たされていることを示す

ことがガイドラインとして設定された。

■ APA 第5版では効果量などの報告を義務付けしておらず、シンプルでわかりやすいガイドラインの必要性が叫ばれ、APA 第6版において CI、効果量の表記について具体例とともに示された。

「帰無仮説分析は統計的分析のはじまりに過ぎず、効果量や CI などを合わせて記載することが結果を適切に報告するために必要である(American Psycological Association, 2009, p.33)」

Historically, researchers in psychology have relied heavily on null hypothesis statistical significance testing (NHST) as a starting point for many (but not all) of its analytic approaches. APA stresses that NHST is but a starting point and that additional reporting elements such as effect size, confidence intervals, and extensive description are needed to convey the most complete meaning of the results.

「検定結果を完全に報告すること、並びに効果量や CI などの適切な推定値を報告することは APA が発行する論文誌においては<u>最低限の要求</u>である(American Psycological Association, 2009, p.33)」

… complete reporting of all tested hypotheses and estimates of appropriate effect sizes and confidence intervals are the minimum expectations for all APA journals.

➢ p 値の報告のみでは統計的分析の正確さに疑問が残る。APA manual からは、検定力について特に論じられていないようだが、検定力分析によって実験のコストや手間を下げることができる。

大久保・岡田(2012). 第五章 検定力：研究の信頼性と経済性を高めるために

■ 「検定力は帰無仮説が偽であるときに正しく帰無仮説を棄却する確率」。適切な検定力があるなら（サンプルに）差があるときに差があることを正しく判断できる。

→「信頼に足る研究を行いたいなら、研究の計画段階で適切な検定力のあるデザインを考えなくてはならない(p. 149)」

■ 適切な検定力は研究の信頼性の面だけではなく、経済面・実用面への効果もある。被験者が多いと分析の手間が非常にかかるような場合、非常に特殊なケースの被験者を募って実験を行う場合などの必要サンプルサイズを事前に推定できるため。

■ 「どんな判断でも、第1種の過誤と第2種の過誤をおかしてしまう可能性がある。」

→ネットオークションで最安値だと判断し入札したら、実は相場と変わらなかった（差がある[最安値だ]と判断したのに違ったため、<u>第1種の過誤</u>）。

→相場と変わらなく安値感がないと思っていたのに、よく調べたらお買い得だった（差がないと判断したのに差があった：<u>第2種の過誤</u>）。

■ 有意水準だけに着目すると第2種の過誤についてほとんど情報が得られない。第1種の過誤と第2種の過誤はそれぞれ別物なので、双方を考慮する必要がある。有意水準ばかりにこだわって判断をしていると、正しい判断ができていない可能性を含むことになる。

■ 有意水準は慣例的に 5%に固定されており、効果量は実際に分析してみないとわからない。サンプルサイズのみが研究者が能動的に動かせる唯一のパラメータになる。

- 効果量と有意水準が等しければ、サンプルサイズが大きくなるほど検定力が高まる。サンプルサイズが小さくなれば、検定力は低くなる。
- 高すぎる検定力は、ほんとうにわずかな差（効果）でも高い確率で検出されてしまう原因になる。ほとんど無意味と言っていい小さな差も検出してしまう。
- 低すぎる検定力の場合、本当は存在する差を検出できなくなる。低い検定力は医学関係では特に敬遠される（患者への不利益が生じる可能性が高いため)。
- Cohen (1988)は、行動科学や社会科学の研究では第1種の過誤が第2種の過誤よりも重大な問題であるとし、第1種の過誤を第2種の過誤より4倍の慎重さで操作すべきとした。第1種の過誤：α = .05 の場合、第2種の過誤：β = .20 となる。検定力(1 – β)のとき、第一種の過誤 a： 第2種の過誤 b = 0.05：0.20 であるため、1 – 0.2 = 0.8 となり、検定力が 0.80 を基準と考える。
- α = βであるべき、という分野もあり、そこではβも.05 に設定され、検定力を 0.95 とすることもある。
- Kline (2004)は最低限の検定力は 0.5 としており、これを下回る場合は問題があるとしている。これは丁半博打の状態なので、分析をかける意義がなくなるためである。
- Cohen の基準である、Power = 0.80 を下回った場合でも、特殊な実験結果であったり、サンプルを多くとれないような研究の場合は、多少この基準を下回っても許容してよいのかもしれない。

2. 対応なしの t 検定

2.1. G*Power 3 による事前検定

- Test family を「t test」に、Statistical test を「Means: Difference between two independent means (two groups)、Type of power analysis を「A priori: Compute required sample size – given α, power, and effect size」に設定し、Input Parameters の Tail(s)を「Two」にする。

- Determine をクリックし、効果量を測定する。事前検定なので、先行研究に効果量が記載されていればその数値をそのまま使えばよいが、載っていない場合はグループ 1 とグループ 2 の平均値と標準偏差を n1 = n2 の欄にそれぞれ情報を挿入し、「Calculate and transfer to main window」をクリックする。
  ＊効果量 d の目安、効果量小：(0.2), 効果量中：(0.5), 効果量大：(0.8)
- 今回は、グループ 1 の平均値を 16.99、グループ 2 の平均値を 15.24、グループ 1 の標準偏差を 2.90、グループ 2 の標準偏差を 3.19 としました（それぞれの数字は適当です)。
- この作業で Effect size がメイン画面に挿入されるので、あとはα error prob に 0.05、Power (1 – β error prob) に 0.80 (画面上では 0.95 になっていますが、0.80 といれてください)、Allocation ratio N2/ N1 に 2 つのグループの n の比率を記入します。
- すると、各群に必要なサンプルサイズが表示され (Sample size group1, 2, Total sample size)、Actual power が計算されます。

1.2. G*Power 3 による t 検定（事後検定）

- 「test family」、「Statistical test」、「Tail(s)」の部分は事前検定と同じく、「t test」、「Means: Difference between two independent means (two groups)」、「Two」とする。「Type of power analysis」を「Post hoc: ~」とし、Input Parameters に必要な情報を入力する。
- 平井(2012)の対応のない t 検定のデータを使用し分析を行ったところ、検定力が 1.00（理論上の上限）となった。
  →これは検定力が高すぎたため、些細な差も含まれている可能性が高いと思われる。

2. 対応のある一元配置分散分析

2.1. G*Power 3 による対応のある一元配置分散分析（事前検定）

- 「Test family」を「F tests」に、「Statistical test」を「ANOVA: Repeated measures, within factors」と選択し、「Type of power analysis」は事前検定なので「A periori」とする。
- 分散分析の効果量は f で求められる。その水準は、効果量小：0.10、効果量中：0.25、効果量大：0.40 が目安になる。

  ＊偏 η 2 乗値 ≠ f なので、注意が必要。（効果量 f は $f = \frac{グループ間平方和}{グループ内平方和}$ で求められる。

- 効果量を直接計算するため、「Determine =>」から Direct の「Partial η^2」にわかっている効果量を入れ、「Calculate and transfer to main window」をクリック。α, β共に固定値なので、次に「Number of groups」に適当な数字（一元配置の場合、1 を入力)。「Number of measures」には繰り返した測定の数（水準数）を入れる（今回は 3 回繰り返したと想定したため、3 を代入)。
- 「Corr among rep measures」では先行研究や予備実験で相関が分かった場合にはそれを入れ、わからなければ 0.5 としておく。「Nonsphericition correction」は球面性の仮定についての数字を入れる。球面性の仮定が満たされていれば 1 を、満たされていない場合は、$\frac{1}{水準数-1}$を入れればよい。
- 今回の分析の結果、効果的なサンプルサイズは 1 群につき、10 であるとみなすことができる。