

プロフェッショナル統計分析ワークショップ
応用コース 3: パワーアナリシス-G * Power によるサンプルサイズ計算

Q & A

このコースを受講いただいた皆様からの質問の一部とインストラクターからの回答を紹介致します。受講後はお気軽にご質問ください。

(インストラクター: 佐々木亮)



頁	質問	佐々木の回答・コメント
2-3	Power Analysis の意義や目的について、より詳しいご説明をいただければ幸いです(「30 あるいは 25 が必要、できれば 100、もしくはそれより多ければ多いほど良い」というのが、サンプルサイズに関する最低の考え方ですが、例えば「30 あるいは 25」が満たされることがあらかじめ分かっている場合でも、G * Power を用いてサンプルサイズを決定した方がベターなのでしょうか?それは何故?)	<p>最小のサンプルサイズとして経験的に用いられている 25 あるいは 30 という思想と、G*Power による Power Analysis の思想は寄って立つ考え方が違いますね。それがどうつながるのかは誰も解説していないと思うので私がトライしてみます。</p> <p>25 あるいは 30 は「母集団を適切に代表すると言えるサンプル数は最小でいくらか」という思想です。母集団とサンプル集団の特徴がほぼ同じかどうかを判断するのに、(i)平均値と(ii)標準偏差を見ますね。ここで、(i)母集団の平均値とサンプル集団の平均値は自動的にほぼ同じするのに対して、(ii)母集団の標準偏差よりサンプル集団の標準偏差は自動的に少し狭くなります。その狭さというか差を無視していいと言われるのが 25 あるいは 30 です。そこから最低のサンプル数は 25 あるいは 30 ということになります。25 あるいは 30 あれば、母集団とサンプル集団で(i)平均値と(ii)標準偏差がほぼ同じになる、つまりサンプル集団は母集団を適切に代表している、と言えるわけです。</p>

		<p>一方、Power Analysis は「2つのサンプル集団の平均値差が偶然では起こり得ないほど大きな差だと言えるのは、二つのサンプル集団のサンプル数がいくつの場合か」という思想です。それを、Effect size と有意水準（本当は差がないのに偶然で差があると出てしまうのは 100 中何回現れるか）から逆算して得た値が Power Analysis から得たサンプル数となります。</p> <p>この両方の思想をつなげるとどうなるかと言うと、2つのサンプル集団のそれぞれが最低でも 25 あるいは 30 以上（とると、2つのサンプル集団がそれぞれの背後になる母集団を適切に代表していることになる）で、あとはサンプル集団の大きければ大きい程「その背後にある2つの母集団の平均値差があると判定しやすくなる」ということになります。</p>
4	<p>T 検定で「両側の p 値」をチェックするということですが、この両側 (Two-tails) とは何を意味しますか。また「相当の確信がない限り、両側 (Two-tails) を選ぶのが無難」と書かれていますが、「相当の確信」とは例えばどのようなケースが考えられますか。</p>	<p>Tails は、正規分布などの分布図の端が何個あるかですね。通常は左と右にありますから2個あります。2群の t 検定では、「効果があった」、「効果がなかった」を判断しますが、「効果があった」には「+の効果があった」「-の効果があった」と両方ありますね。それがプラスマイナス2標準偏差の外側の領域で左と右の2か所あるわけです。</p> <p>これを「+の効果があった」と「効果がなかった」だけで検定する場合に、TwoではなくOneを選ぶわけです。しかし、私の講義を受けて「+の効果があった」と「効果がなかった」だけを想定していいでしょうか？万が一のことを考えて「-の効果があった」という場合も考えに入れるのが適切でしょう。</p> <p>「相当の確信」とは私の講義は、「+の効果があった、あるいは効果がなかった」という場合のみだと確信できる場合です。つまり私の講義には「マイナスの効果はあり得ない」と自信を持って言える場合です。しかし、通常は「-の効果」の場合もあり得ると思うので、Twoと選ぶというわけです。</p> <p>実際に「-の効果」となった研修の場合を見たことがあり、それは Political Economy という私の同僚の講義でしたが、講義後のアンケートの記述には「今回の講義を聞いてますます分からなくなった」と書いてありました。（私のこのワークショップの講義がそうでないことを祈りますが、そうなる可能性は否定できないので事後アンケートを取ったら、私はやっぱり Two-tails を前提としますね。）</p>
9-10	<p>（全体として、G*Power の操作方法や結果の見方の後、別添 1、別添 2 の内容説明をしていただけると有難かったです。）</p> <p>(1) 確率サンプリングと (2) 非確率サンプリングの違いは何ですか。</p>	<p>「確率サンプリング」は途中でサイコロを使う場合で、「非確率サンプリング」は担当者が自分でこれが代表例だと選ぶ場合です。前者はサイコロの代わりに現在は、エクセルの Random Number Generator を使うことが多いでしょう。一方、後者は恣意的（意図的）になりがちです。</p>
9	<p>「体系的な方法」は、「まず無作為に一つサンプルを選び…」とありますが、その意図は何でし</p>	<p>スタートのサンプル1個を選ぶということですね。最初の1個を無作為に選ばずに、特定の1個（例えば100個のうちの第1個目）を選ぶということも広く行われていると思います。しかし母集団を用意した</p>

	<p>ようか。何故、「(すべて) 等間隔でサンプルを選んでいく」方法ではダメなのでしょう。</p>	<p>人が、(最初の印象が肝心とばかりに) 第1個目に一番いい個体を持ってきている可能性があるため、第1個目を選ばずに、無作為(ランダム)にひとつ選ぶというのが勧められるというわけです。</p>
9	<p>「階層化法」と「クラスター法」でいうところの「階層」と「クラスター」の違いは何ですか。</p>	<p>「階層化法」の例は次の場合です。1000人の所得を高・中・低に分けると200人・500人・300人に分けられるので、サンプルとして20人・50人・30人を選ぶと、つまり母集団の比率でそのまま縮小して、高・中・低の各階層からサンプル集団を選ぶということになります。</p> <p>クラスター法は、合計1000人の住民が住むA地区、B地区、C地区、D地区・・・Z地区のうちから、C地区とL地区のみ選んで、その2つの地区の住民を全員サンプルとすると100人になる、といえます。明らかに前者の「階層化法」の方が優れていますね。ただし後者は交通費など費用を節約できるでしょうが、あまり勧められません。</p>
10	<p>「最大類似事例法・最大相違事例法」を用いる場合、「なぜ、類似としたか」「なぜ、相違としたか」の理由説明は、どの程度、報告書や論文に書く必要がありますか。</p>	<p>論文ではほとんど説明されることはなくて、「複数の専門家が専門的見地から選んだ」という断り書きを書きますね。根拠はその専門家の専門的知見です。どれほど「専門家」という言葉を信じるかにかかっていますね。</p>
10	<p>「代表事例法」を用いる場合、「(選んだサンプルが) 何故、代表的な事例か」、そして「何故、サンプルとして選んで調べることが有益だとあらかじめ分かっている事例か」の理由説明は、どの程度、報告書や論文に書く必要がありますか。</p>	<p>これは論文では書く必要があるでしょう。「通常より状況が悪い学校なので、学ぶことが多いと専門家として判断した」といった具合ですね。やはり「専門家」という言葉で押し切るという感じですね。</p>
10	<p>「雪だるま法」はどのようなケースで用いることが多いですか。</p>	<p>これはじつはたいへん頻りに用いられています。つまり一人インタビューして、友達を紹介してもらうという場合です。友達のネットワークを利用(悪用)しているとも言えるでしょう。弱点としては似たような人が紹介されるということと、紹介した(された)友達に気を使って回答してしまうということが挙げられます。利点としては例えば、インフォーマルな犯罪者集団のメンバーをインタビュー対象とする時などはサンプルの収集方法はこれしかありません。</p>
11	<p>(1) ランダム・サンプリングは、「単純無作為法」(テキスト p9) のことでしょうか。</p>	<p>そうです。</p>
12	<p>(3) 現場サンプリングは、「コンビニエンス法」(テキスト p10) のことでしょうか。</p>	<p>この二つは違います。現場サンプリングは、母集団と同じ比率で縮小してサンプルを選びますね。コンビニエンス法は、手近な例を選ぶ(例えば自分の電話帳に載っている友達から選ぶ)ということ勧められませんね。</p>

<p>13</p>	<p>Effect size (効果幅)の説明がありました。「大・中・小」が判断できて便利だし、現場で利用価値が高いと思いました。この「大・中・小」はどのような根拠によるのでしょうか。</p>	<p>そもそも Effect size (効果幅)は、ニューヨーク大学の Jacob COHEN (1923-1988)がとりまとめて提案したものです。Effect size (効果幅)とは、「2群のサンプルの平均値の差は標準偏差の何割か」を計算するという事です。判断基準は、大: ES = 0.8、中: ES = 0.5、小: ES = 0.2 です (出所 Cohen 1988)。</p> <p>彼はフィッシャーに取って代わろうとしました。その試みは半分成功して、半分成功していません。彼はその 1988 年の書籍でこう書いています。</p> <p>『統計学的有意の検定 (Fisher の遺物) の強調は、結果の統計学的有意への注意を喚起するが、一方で効果幅 (Effect size) への注意を阻害している。私は、慣例として (as a convention)、定性的な形容詞である“小”、“中”、“大”を提案する。この運用は多くの危険を伴う。これらの定義は曖昧で (arbitrary)、このような定性的な概念である“大”は、時に絶対的と理解され、時に相対的に理解されるので、誤解を生むであろう。しかし、正当化のためにいくつかの主張が提示できる。</p> <p>第一に、全ての慣例は曖昧である。そして、全ての慣例は間違っていて使われ、それゆえ、慣例としての位置づけが乱用されている。例えば、行動科学と生物化学のほとんどの分野で帰無仮説を否定するのに、非公式ではあるが、0.05(5%)の基準が慣例として、用いられている。しかし曖昧だとしても、慣例は、理性的な人々にとっては理性的なものである。』(p.12)</p> <p>*****</p> <p>私がこの解説を読んだ感想は次のとおりです。第一に、いきなり「全ては曖昧だ！」と言いついてはいるわけです。曖昧に見えるが明確に説明せよ、と言われて、「全ては曖昧なんだ！」と最初に開き直るとは、よく言い切りましたね。</p> <p>第二に、0.05(5%)が慣例として用いられている、だから慣例でもいいのだ、としていますね。この0.05(5%)の基準はフィッシャーが 1929 年に提案して普及したものです¹。コーエン先生はフィッシャーを過去の遺物(legacy)と呼んでそれを乗り越えようとしているのに、そのフィッシャーを自分の主張の根拠に持ってくるとは、素直に驚嘆します。</p> <p>最後に第三として、「理性的な人々にとっては理性的なものである」と言っていますね。これは「分かる人には分かる」と言っているわけです。私たちも口論して負けそうになると、「分かる人には分かる。僕はそう信じてるよ。」と言ったりするわけです。つまり、捨て台詞で、論争に負けた時に言うセリフなわけですよね。それを言うてはおしまいよ、とも言えるでしょう。</p>
-----------	--	--

¹ 「慣例として、偶然によって生じるのが 20 回の試行のうち 1 回未満という程度であれば、結果は有意であると判断する。」 Fisher R.A., (1929). "The statistical Methods in Psychological Research" In *Proceeding of the Society for Psychological Research*,39, p.112.

一方で、Effect Size (効果幅) は、最近はわりと広く使われるようになってきました。そして、判断の基準である $d=0.2, 0.5, 0.8$ について、今はまだ (Source: Cohen 1988) と皆書いています。一方、有意判定で用いられる $p=0.05(5\%)$ についてはもはや誰も (Source: Fisher 1929) とは書きませんね。おそらくそのうちコーエン先生を出所として挙げる人もいなくなるでしょう、そして常識となるでしょう。歴史は繰り返すと言ったところでしょうか。そしてさらに将来、コーエン先生も乗り越えられていくのでしょうか。

なお、学会誌によって、 p 値のほかに Effect Size (効果幅) を明記することを要求する学会誌もありますし、逆に p 値のみでよくてそれ以上は求めないという学会誌もあります。私も過去に報告書に p のほかに Effect Size (効果幅) も書いたことがあります。そうしたところ、「小とはなんだ。小でも重要な効果であることもある！」とクライアントに言われて議論が激しくなり、「じゃあ Effect Size (効果幅) の記載はやめておきましょう。有意かどうかの p だけで十分ではないですか」と提案して、「まあ、それでいい」と皆で合意したということがありました。道は遠いですね。

ご質問を寄せていただいた皆様、たいへんありがとうございました。この場を借りて御礼申し上げます。今後も随時追加していきたいと存じます。

(最終更新日) 2021/08/22

イラスト：いらすとや (<https://www.irasutoya.com/>)

(著作権の範囲内で使用させていただきました。)

このワークショップの全体構成は以下のようになっています。

- 『プロフェッショナル統計分析ワークショップ』(基礎コース) <https://www.idcj.jp/seminar/statistical-analysis-workshop.html>
講師：佐々木亮 (ウェスタンミシガン大学評価学博士)、高木桂一 (スタンフォード大学社会学博士)
- 『プロフェッショナル統計分析ワークショップ』(応用コース) <https://www.idcj.jp/pickup/grow/statistical-analysis-workshop-advanced.html>
インストラクター：佐々木亮 (ウェスタンミシガン大学評価学博士)
応用コース1：インパクト評価の最新テクニック (DID, PSM, IV)
応用コース2：構造方程式モデリング (SEM: Structural Equation Modeling)
応用コース3：インパクト評価のためのサンプルサイズの計算
応用コース4：メタ分析 (システマティック・レビュー) の計算