

伝えるための心理統計

大久保街亜・岡田謙介

2012 年（勁草書房）

1 心理統計における新展開

Wilkinson & APA Task Force on Statistical Inference (1999)

- 必要最小限の分析
- 信頼区間の報告
- 効果量の報告
- 分析の前提が満たされているかチェック

Kazdin (1999) による「臨床的意義」の定義

介入的効果の実践的あるいは応用的価値であり、その介入がクライアントあるいはクライアントの周囲の人に、日常生活のレベルではっきりとした違いを与えるか否か。¹

⇒ 有意であるだけでなく、十分な効果量が必要

2 帰無仮説検定

Fisher の有意性検定

H_0 の設定 → p 値 → H_0 を棄却 or 採択

「計算された統計量が偶然と考えられる範囲を超えることはどれくらいあるか？」

Neyman-Pearson による帰無仮説検定

H_0 と $H_1 \rightarrow \alpha \rightarrow H_0$ or H_1 を棄却

2.1 論理における問題点

後件否定

$$P \rightarrow Q \quad (2.1)$$

$$\neg Q \rightarrow \neg P \quad (\text{後件否定 } \bigcirc) \quad (2.2)$$

$$Q \rightarrow P \quad (\text{後件肯定 } \times) \quad (2.3)$$

「差がない」ことを確かめたいこともあるはずだが、検定では H_0 が正しくないとわかるのがせいぜい。

¹Journal of Consulting and Clinical Psychology (1999) 67(3) では、“clinical significance” についての特集号が組まれている。

対立仮説を直接確かめられない

「 H_1 が真 $\rightarrow Q$ 」のとき、「 $Q \rightarrow H_1$ が真」とは言えないので、「 H_1 が偽 $\rightarrow Q$ でない」のとき、「 Q である $\rightarrow H_1$ が真」と言わないといけない

H_0 は常に偽である

Tukey (1991) 「違いは小数点以下のどこかには必ずある」

$P(H_0|D)$ が知りたい (Cohen, 1994)

H_0 は定数と考えるので、この確率は考えられない

2.2 解釈における問題点

有意差 = 重要な差か？

- 1931 年には、Tyler がすでに statistical significance \neq social significance を主張
- p 値は、効果の大小や関係性の強さに関する情報を一切与えない

極端な二分法

- $p = .049$ と $p = .051$ で何が決定的に違うのか？

2.3 手続き上の問題点

- 検定によって、研究内容が限定される。（「差がない」ことを取り上げない）
- 大標本で p 値は小さくなる

2.4 検定を擁護する

- 有意水準という明確な基準がある
- 二値的な判断こそ実験では知りたい (\leftrightarrow CI 重視の立場)

3 効果量と検定統計量との関係 (p.45)

検定統計量は、標本サイズと効果量の関数のかけ算である²。

$$T = f(N) \times g(ES) \quad (3.1)$$

4 効果量の2つの族

Rosenthal (1994) による分類：

- 群間差 (d 族)
- 変数観の関係 (r 族)

²Maxwell & Delaney, 1990; Rosenthal, 1994

4.1 d 族の効果量

母平均 μ_1, μ_2 、2 群に共通の母分散 σ^2 として、

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \quad (4.1)$$

を母集団効果量とする。これを推定するために、以下の方法が提案されている。(分母 σ をどのようにするか。)

Cohen の d

$$d = \frac{M_1 - M_2}{S_p} \quad (4.2)$$

$$= \frac{M_1 - M_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}}} \quad (4.3)$$

ただし、 S_1^2, S_2^2 は各群の標本分散（不偏でない）³。

Hedges の g

$$g = \frac{M_1 - M_2}{s_p} \quad (4.4)$$

$$= \frac{M_1 - M_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \quad (4.5)$$

ただし、 s_1^2, s_2^2 は各群の不偏分散⁴。

d と g には

$$g = d \sqrt{\frac{n_1 + n_2 - 2}{n_1 + n_2}} \quad (4.6)$$

という関係がある。

また、 g と検定統計量 t には

$$t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} g \quad (4.7)$$

が成り立つ。

g にもまだ正方向のバイアスがあり、これを補正するために

$$\hat{\delta} = J(n_1 + n_2 - 2)g \quad (4.8)$$

$$\left(J(a) = \frac{\Gamma(a/2)}{\sqrt{a/2} \Gamma((a-1)/2)} \right) \quad (4.9)$$

$$g_{adj} = \left(1 - \frac{3}{4(n_1 + n_2) - g} \right) g \quad (4.10)$$

$$g_{adj}^* = \left(\frac{n_1 + n_2 - 3}{n_1 + n_2 - 2.25} \right) \quad (4.11)$$

などが提案されている⁵。

³当初 Cohen が d として表したのは式 (4.1) の δ のことであり、式 (4.4) の g に相当する量が d_s と表されていた。一方、式 (4.2) を Cohen の提唱した推定量として “Cohen の d ” と定義する文献も多く、混乱を招いている。ここでは、McGrath & Meyer (2006) によって整理・推奨された記法に則っている。

⁴この場合、 s_p^2 は、 σ^2 の不偏推定量になる。 s_p 自体は不偏推定量でないが、 S_p よりもバイアスは小さい。

⁵それぞれ、Hedges & Olkin (1985); Hedges (1981); Hunter & Schmidt (2004) による。

Glass の Δ

介入によって分散が変化することを考慮し（等分散性が満たされない）、統制群の標準偏差 s_2 を用いて

$$\Delta = \frac{M_1 - M_2}{s_2} \quad (4.12)$$

とする⁶。

分母を計算する際に片方のデータしか用いないため、

1. d や g に比べて標本誤差が大きい
2. t から直接計算できず、メタ分析に使いにくい

という欠点がある。

2 群に対応がある場合

- 対応のない場合と同様に d や g を算出する
— 対応の有無に関係なく同尺度上で値を比較できるが、2 群の相関関係は考慮されない
- 差得点を用いて効果量を算出する
— 2 群の相関関係は考慮され、検定統計量との関係からも整合的であるが、対応の有無で値を直接比較できない

差得点を利用する場合、

$$d_D = \frac{M_D}{s_D} \quad (4.13)$$

とする（ s_D^2 は差得点の不偏分散）。

この d_D は、母集団効果量

$$\delta = \frac{\mu_1 - \mu_2}{\sigma \sqrt{2(1 - \rho_{12})}} \quad (4.14)$$

の推定量になる⁷。

4.2 r 族の効果量

4.2.1 相関係数、点双列相関係数

点双列相関係数 r_{pb} は、連続変数を 2 群に分け、各群の比率を q_1, q_2 としデータ全体の（不偏でない）標準偏差を S とすると、

$$r_{pb} = \frac{M_1 - M_2}{S} \sqrt{q_1 q_2} \quad (4.15)$$

が成り立つ。

4.2.2 回帰分析

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_e^2 \quad (4.16)$$

として、

$$\rho^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} \quad (4.17)$$

を決定係数（分散説明率）とする。

⁶Glass, McGaw, & Smith (1981)

⁷ $\rho_{12} = .5$ のときのみ、式 (4.1) と一致する。

4.2.3 1 要因被験者間の分散分析

回帰分析と同様、

$$\sigma_T^2 = \sigma_A^2 + \sigma_E^2 \quad (4.18)$$

として、

$$\eta^2 = \frac{\sigma_A^2}{\sigma_T^2} \quad (4.19)$$

を母集団効果量とする。この推定量として、

$$\hat{\eta}^2 = \frac{SS_A}{SS_T} \quad (4.20)$$

$$\hat{\epsilon}^2 = \frac{df_A(MS_A - MS_E)}{SS_T} \left(= \frac{SS_A - df_A MS_E}{SS_T} \right) \quad (4.21)$$

$$\hat{\omega}^2 = \frac{df_A(MS_A - MS_E)}{SS_T + MS_E} \left(= \frac{SS_A - df_A MS_E}{SS_T + MS_E} \right) \quad (4.22)$$

が提案されている⁸。

$\hat{\epsilon}^2$ は $\hat{\eta}^2$ と比較して分子の $df_A MS_E$ 分だけ小さく、 $\hat{\omega}^2$ は $\hat{\epsilon}^2$ と比較して分母の MS_E 分だけ小さい値になる。

$$\hat{\omega}^2 \leq \hat{\epsilon}^2 \leq \hat{\eta}^2 \quad (4.23)$$

だが、 $\hat{\omega}^2$ と $\hat{\eta}^2$ の差はかなり小さい。

$\hat{\omega}^2$ は、母分散 $\sigma_T^2, \sigma_A^2, \sigma_E^2$ それぞれを平方和ではなく、より適切な推定量で推定して置き換える⁹。

メモ

平均平方の期待値については、Kirk (2013) で再確認

4.2.4 1 要因被験者内の分散分析

$$\sigma_T^2 = \sigma_A^2 + \sigma_S^2 + \sigma_E^2 \quad (4.24)$$

のように分散成分が3つ以上に分かれる場合、関心のある分散を σ_μ^2 と表し、

$$\eta^2 = \frac{\sigma_\mu^2}{\sigma_T^2} \quad (4.25)$$

$$\eta_p^2 = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2} \quad (4.26)$$

の2種類の効果量を考えることができる（決定係数と偏決定係数）。

この場合の η^2 および η_p^2 についても、 SS を用いて $\hat{\eta}^2$ や $\hat{\eta}_p^2$ を推定量とすることもできるし、分散成分の推定量 $\hat{\sigma}_A^2, \hat{\sigma}_S^2, \hat{\sigma}_E^2$ からバイアスを補正した推定量 $\hat{\omega}^2, \hat{\omega}_p^2$ を用いることもできる。

4.2.5 2 要因被験者間の分散分析

決定係数、偏決定係数の推定量として SS を用いる限り、1 要因の場合と値が変わることはないが¹⁰、母分散の推定量自体は変化するので、 $\hat{\omega}^2, \hat{\omega}_p^2$ の値は1 要因の場合と変化する。

メモ

本書では、各要因の分散説明率を決定係数としているが、回帰分析との整合性の観点からすれば、モデルで説明できる部分全体を分子に取ったものを決定係数と言い、各要因の追加による決定係数の増加は部分決定係数と呼ぶはず。南風原 (2014) の指摘するように部分決定係数は検定力を直接規定せず、決定係数と偏決定係数は検定力を直接規定するので、この区別は曖昧にしない方がよい。

メモ

η^2 は要因の数が増えると小さな値となる、とあるが、必ずしもそうではない。南風原 (2014) p.96 で述べられている例のように、母集団を変えずに比較すべき要因が増えていくだけでは η^2 は変化しない（少なくともバランスデザインでは）。

4.2.6 一般化効果量 (Olejnik & Algina, 2003; Bakeman, 2005)

被験者内デザインでは、被験者間デザインよりも η_p^2 が必然的に大きくなる¹¹ので、偏決定係数の値を比較できない。

データを変動させる要因を

1. 操作された要因の分散
2. 個人差の分散（操作されていない要因に由来する分散の全体）

に大別し、

$$\eta_G^2 = \frac{\sigma_\mu^2}{I \times \sigma_\mu^2 + \sigma_{id}^2} \quad (4.27)$$

を一般化効果量とする。

- σ_μ^2 : 関心のある要因の分散
- I : 関心のある要因が操作された要因なら 1、そうでなければ（個人差の成分さえ含んでいれば）0
- σ_{id}^2 : 操作されていない要因の分散の総和

メモ

1 要因分散分析について、対応ありとなしの場合を考えてみる。要因 A が操作された要因で、かつ関心のある要因である場合、

対応なし

$$SS_T = SS_A + SS_E \quad (4.28)$$

であるから、

$$\eta_G^2 = \frac{SS_A}{1 \times SS_A + SS_E} \quad (4.29)$$

$$= \frac{SS_A}{SS_A + SS_E} \quad (4.30)$$

となる。

⁸後者 2 つは、それぞれ Ezekiel (1930); Hays (1994)。

⁹Searle, Casella, & McCulloch (2006) などにまとめ。

¹⁰バランスデザインであれば、のほず。アンバランスデザインだと要因を加えることで SS も変わる。

¹¹被験者効果に関する分散を分母から取り除くことができる。

対応あり

$$SS_T = SS_A + \underbrace{SS_S + SS_E^*}_{SS_E} \quad (4.31)$$

であるから、

$$\hat{\eta}_G^2 = \frac{SS_A}{1 \times SS_A + SS_S + SS_E^*} \quad (4.32)$$

$$= \frac{SS_A}{SS_A + SS_E} \quad (4.33)$$

となる。

対応ありとなしの場合で、データ自体が同じであれば $\hat{\eta}_G^2$ の値自体は変わらない¹²

4.3 Cohen の U_1, U_2, U_3

平均だけ異なる 2 つの分布について、重なり具合を効果の大きさとして評価したもの。

U_1 分布が重ならない X 軸の範囲の面積割合 $0 \leq U_1 \leq 1$

U_2 分布の高さが交わる X 軸の値より大きい（小さい）範囲の面積割合 $0.5 \leq U_2 \leq 1$

U_3 高群の中央値よりも小さな範囲の低群の面積割合 $0.5 \leq U_3 \leq 1$

4.4 ノンパラメトリックな効果量

4.4.1 優越確率

母集団において 2 群の値に同順はないものとする、第 2 群より第 1 群の個体が大きな値を取る確率（優越確率）

$$p_s = p_{1>2} \quad (4.34)$$

を定義できる¹³。

実際のデータでは同順位が発生するので、そのペアについては $p_{1>2}, p_{1<2}$ に半数ずつ割り当てるとして、

$$\hat{p}_s = \hat{p}_{1>2} + 0.5\hat{p}_{1=2} \quad (4.35)$$

として算出する。これは、Mann-Whitney の検定統計量 U に等しい。

母集団においても同順を認めるとすると、

$$\delta_s = p_{1>2} - p_{1<2} \quad (4.36)$$

を定義でき、この推定量 d_s を Cliff の d と呼ぶ。

2 群に対応がある場合、ペアごとに第 1 群の 2 群での大小を比較して、その数を ω とすれば、

$$\hat{p}_{s_{dep}} = \frac{\omega}{N} \quad (4.37)$$

とすればよい。

4.5 単純効果量

標準化しない、もとの単位のままの効果量を単純効果量¹⁴とよび、APA マニュアルでも認められている。

¹²偏決定係数は、それぞれ $SS_A/(SS_A + SS_E)$ と $SS_A/(SS_A + SS_E^*)$ になる。

¹³Grissom (1994)

¹⁴Frick (1999)

5 信頼区間

5.1 母平均の信頼区間

5.1.1 Loftus & Masson (1994); Masson & Loftus (2003) の方法

分散分析における誤差項を利用した母平均の信頼区間。分散の等質性が満たされている必要がある。

独立測定の場合

$$CI = M_k \pm t_{critical} \times \sqrt{\frac{MS_E}{n}} \quad (5.1)$$

ただし、 $t_{critical}$ は誤差の自由度において算出する。

反復測定の場合

$$CI = M_k \pm t_{critical} \times \sqrt{\frac{MS_{SA}}{N}} \quad (5.2)$$

Jarmasz & Hollands (2009) は、被験者数だけでなく、繰り返しの数を考慮して

$$CI = M_k \pm t_{critical} \times \sqrt{\frac{MS_{SA}}{N \times \frac{L}{r}}} \quad (5.3)$$

という改良案を出している¹⁵。

メモ

表 4.1 (p.128) のデータを用いて、裸足条件の母平均の CI を算出してみる。

通常の t 検定にもとづく方法

$M = 8.7, SD = 0.5, n = 10, t(9)_{critical} = 2.26$ なので、

$$CI = 8.7 \pm \frac{0.5}{\sqrt{10}} \times 2.26 = [8.34, 9.06] \quad (5.4)$$

となる。

3 群を独立測定とみなした場合

$M = 8.7, MS_E = 0.32, n = 10, t(27)_{critical} = 2.05$ なので、

$$CI = 8.7 \pm \sqrt{\frac{0.32}{10}} \times 2.05 = [8.33, 9.07] \quad (5.5)$$

となる。

¹⁵ L は全反復測定要因の水準数の積、 r は要因 A の水準数。

3 群を反復測定とみなした場合

$M = 8.7, MS_{AS} = 0.06, N = 10, t(18)_{critical} = 2.10$ なので、

$$CI = 8.7 \pm \sqrt{\frac{0.06}{10}} \times 2.10 = [8.54, 8.86] \quad (5.6)$$

となる。

この場合は、3 群が独立である限り、通常の t 検定にもとづく方法を用いても、Loftus & Masson (1994) の分散分析にもとづく方法を用いても、結果にほとんど違いはなかった¹⁶。

さらに、2 要因以上で対応がある場合には誤差平均平方をどのように指定するのかという問題が出てくる。もし誤差平均平方間に 2 倍を超える差がなければ、これらをプールして用いる。例えば、対応のある 2 要因の場合、

$$MS_{SAB} = \frac{SS_{SA} + SS_{SB} + SS_{SAB}}{df_{SA} + df_{SB} + df_{SAB}} \quad (5.7)$$

ということ。

メモ

対応のある 2 要因では、

$$SS_T = SS_S + SS_A + SS_{AS} + SS_B + SS_{BS} + SS_{AB} + SS_{ABS} \quad (5.8)$$

と平方和が分割される。

対応ありの場合、分散の等質性に加えて球面性の仮定が満たされていないといけないことにも注意。

混合計画の場合

対応ありの要因の誤差と対応なしの要因の誤差をプールすることはできない。

メモ

対応あり要因となし要因の交互作用については、Hollands & Jarmasz (2010) が誤差平均平方のプールを提案しているという。

2 要因混合計画で A が対応なし、B が対応ありとすると、平方和は

$$SS_T = SS_A + SS_S + SS_B + SS_{AB} + SS_{BS} \quad (5.9)$$

と分割される。この場合、主効果 A は SS_S が、主効果 B と交互作用 AB は SS_{BS} が誤差平方和となる。これらの平均平方 MS_S と MS_{BS} をプールするということらしい。

5.2 単回帰分析の信頼区間

残差標準偏差

$$s_{res} = \sqrt{\frac{(n-1)(s_y^2 - b^2 s_x^2)}{n-2}} \quad (5.10)$$

$$= \sqrt{\frac{(n-1)s_y^2(1 - r_{xy}^2)}{n-2}} \quad (5.11)$$

¹⁶独立測定の場合、どのような条件下なら分散分析にもとづく方法が有利なのだろうか？

信頼区間

傾きと切片の標準誤差は、

$$SE(b) = \frac{s_{res}}{s_x \sqrt{n-1}} \quad (5.12)$$

$$SE(a) = s_{res} \times \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}} \quad (5.13)$$

を利用して、それぞれの信頼区間を求めることができる。

ある値 x_0 に対応する y 値の信頼区間は、標準誤差

$$SE(y_{fit}) = s_{res} \times \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \quad (5.14)$$

を利用して求める。

予測区間

ある値 x_0 に対応する y 値の予測区間は、標準誤差

$$SE(y_{pred}) = s_{res} \times \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \quad (5.15)$$

を利用して求める。

メモ

y 値の信頼区間・予測区間とも、 x_0 が平均 \bar{x} から離れるにつれて広くなる。また、予測区間は信頼区間よりも広い。
 y 値の信頼区間とは、100 回同じことを繰り返せば $x = x_0$ に対応する y の予測値は 95% 程度そこに含まれるだろう区間のことで、予測区間とは、 $x = x_0$ であるサンプルを 100 個得れば、その y 値の 95% は程度そこに含まれるだろう区間。
 y 値の信頼区間とは、結局のところ回帰直線の信頼区間と同じになるような気がするのだが、違うのだろうか？

R による信頼区間と予測区間の算出

R では、`predict` に `lm` のオブジェクトを入れることで、信頼区間と予測区間（正確にはその上限と下限の x に対する値）を得ることができる。R のデータ `cars` を用いた作図例は以下の通り。

```
head(cars) # データセット cars を利用
plot(cars) # 散布図
fit.cars <- lm(dist~speed, data=cars) # 単回帰分析
summary(fit.cars) # 単回帰分析の結果
plot(cars); abline(fit.cars) # 散布図に回帰直線を引く
ci <- predict(fit.cars, interval="confidence", level=0.95) # 95%CI
ci <- as.data.frame(ci)
head(ci,3)
pi <- predict(fit.cars, newdata=data.frame(speed=cars$speed),
  interval="prediction", level=0.95) # PI
pi <- as.data.frame(pi)
head(pi,3)
```

ちなみに、このデータで a と b の分散共分散行列を $(X'X)^{-1}s_{res}^2$ によって推定し、 a と b の相関係数を求めると、 $-.947$ となる。

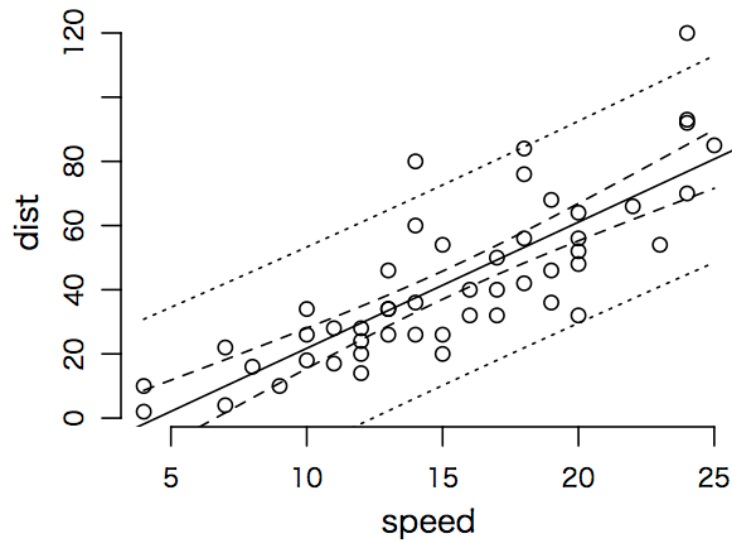


Figure 5.1: 信頼区間と予測区間 (cars データ)

5.3 効果量の信頼区間

Cohen の d

標準誤差

$$SE_d = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2 - 2)}} \quad (5.16)$$

と $t_{critical}$ を用いて求める。

メモ

- 点推定値として g などを用いていても、CI は d を用いて算出するのか？
- t の自由度は？また、正規分布で求める場合もあるが？

6 検定力

6.1 5-80 ルール

社会科学の研究においては、第1種の誤りの方が第2種の誤りよりも深刻である。よって、後者に対して前者を4倍の厳しさで臨む¹⁷。

一方で、双方の誤りを同等に扱うべきであるとする研究者もいる¹⁸。

また、先行研究において効果量が高いとわかっている場合は、第2種の誤りの方を深刻と考えるべきとする研究者もいる¹⁹。

¹⁷Cohen (1988)

¹⁸Cashen & Geiser (2004); Lipsey (1998)

¹⁹Ellis (2010)

6.2 最低限の検定力

検定力は最低でも 0.5 を超えるべきであるとされるが²⁰、社会科学の研究では大多数が検定力 0.5 に満たないとされる²¹。

6.3 ソフトウェア

R や G*Power²²など。

6.4 事後検定力分析

有意でなかったケースについて事後検定力分析を行うべきであるとする雑誌が多数あるが、データから検定力を計算しても大した情報は得られず、時間の無駄である²³。

p 値と事後検定力には対応があり、 $p = .05$ なら事後検定力は 0.5 となる。

7 メタ分析

「メタ分析」という語を発案した Glass による研究の 3 段階：

1. 1 次分析
2. 2 次分析
3. メタ分析

8 ベイズ統計学

8.1 ベイズファクター

Jeffreys (1935, 1961) による「科学的検定」

H_0 と H_1 が排反であるとする、ベイズの定理より

$$\frac{P(H_0|D)}{P(H_1|D)} = \underbrace{\frac{P(D|H_0)}{P(D|H_1)}}_{\text{Bayes Factor}} \times \frac{P(H_0)}{P(H_1)} \quad (8.1)$$

を導ける。つまり、

$$\text{Bayes Factor} = \frac{\text{PreOdds}}{\text{PostOdds}} \quad (8.2)$$

とも表せ、データを得ることによって、仮説の確からしさがどれだけ変化したかを評価する量となる。

Jeffreys (1961) による判定基準

*** < .01 < ** < .1 < * < .31251 < x < 3.2 < * < 10 < ** < 100 < ***
決定的 < 強い < かなり < あまりなし < かなり < 強い < 決定的

²⁰Kline (2004)

²¹Sedlmeier & Gigerenzer (1989)

²²Faul et al. (2007; 2009)

²³Ellis (2010); Hoenig & Heisey (2001)

8.2 対応のある 2 標本の平均の比較

帰無仮説の設定

$$H_0 : X_D \sim N(0, \sigma^2) \quad (8.3)$$

$$H_1 : X_D \sim N(\mu_D, \sigma^2) \quad (8.4)$$

事前分布の設定

- μ_D と σ^2 それぞれに事前分布を設定
- 母集団効果量 $\delta = \frac{\mu_D}{\sigma}$ に事前分布を設定²⁴

$$H_0 : \delta = 0 \quad (8.5)$$

対立仮説について、標準化された指標 δ が完全に無情報な一様分布に従うと考えるのはムリがあるので、

$$H_1 : \delta \sim N(0, \sigma_\delta^2) \quad (8.6)$$

$$\sigma_\delta^2 \sim \chi^{-2}(1) \quad (8.7)$$

超事前分布の設定については Zellner & Siou (1980) に従う。

ベイズファクターの計算は積分を含むため複雑なのだが、Rouder, et al. (2009) が web プログラムを公開している²⁵。

注意点

- 事前分布の設定によって結果が大きく変わりうる（事前情報がない場合の Jeffreys の事前分布や、無情報事前分布など、唯一の方法はない。）
- Hoijsink et al. (2008) の情報仮説によるアプローチも有望

9 p_{rep}

同じ研究がもう一度行われたときに、今回の研究から得られたのと同じ符号の効果量が得られる確率²⁶。

要は、今回得られた効果量を g 、将来得られる効果量を g' とすると、

$$P_{rep} = Pr(g' > 0 | g) \text{ when } g > 0 \quad (9.1)$$

効果量の符号に関する予測確率であり、仮説検定の枠組みとは基本的に関係がない²⁷。

9.1 Killeen (2005a) による p_{rep}

片側検定における p 値を p_1 と表して、

$$p_{rep} = \Phi_N \left[\frac{\Phi_N^{-1}(1 - p_1)}{\sqrt{2}} \right] \quad (9.2)$$

$$\approx \left[1 + \left(\frac{p_1}{1 - p_1} \right)^{2/3} \right]^{-1} \quad (9.3)$$

つまり、 p_{rep} は p_1 の関数である²⁸。

²⁴Rouder, et al. (2009)

²⁵<http://pcl.missouri.edu/bayesfactor>

²⁶Killeen (2005a) によって提案されたが、誤りや説明不足があり混乱を招いた。

²⁷なお、多くの研究者が p 値を予測確率と誤って考えていることを示した研究もある (Oakes, 1986)。標本理論における予測確率については、Miller (2009) を参照。

²⁸仮説検定の枠組みではないといっても、数値としては関係がある。

注意点

- 片側検定の p 値を用いている
- 母分散を既知とみなしている²⁹

Psychological Methods に p_{rep} の特集あり (2010 年)。

片側検定の p 値と p_{rep} との関係

```
> ### p_rep と p 値の関係 (Killeen, 2005a)
> # 正確な値
> p.rep <- function(p1){
+   prep <- pnorm(q=(qnorm(p=1-p1)/sqrt(2)))
+   return(prepare)
+ }
> p.rep(p1=c(0.5,0.1,0.05,0.01))
[1] 0.5000000 0.8175834 0.8776029 0.9500127
> # 近似値
> p.rep.approx <- function(p1){
+   prep.approx <- 1/(1+(p1/(1-p1))^(2/3))
+   return(prep.approx)
+ }
> p.rep.approx(p1=c(0.5,0.1,0.05,0.01))
[1] 0.5000000 0.8122682 0.8768529 0.9553581
```

$p_1 = .1$ とあまり小さくなくとも、 $p_{rep} > .8$ と大きな値になる。

9.2 Lecoutre et al. (2010) による p_{rep}

$$p_{rep} = \Phi_{t(df)}\left(\frac{|t|}{\sqrt{2}}\right) \quad (9.4)$$

- 片側検定、両側検定の区別なく使える
- 母分散を未知としている

→ 1 要因分散分析への拡張も

R で p_{rep} の算出

R に付属のデータ `PlantGrowth` の統制群と実験群 1 のデータを用いて、Lecoutre et al. (2010) による p_{rep} を計算してみる。

```
> ### p_rep (Lecoutre et al., 2010)
> data(PlantGrowth) # 植物の成長データ
> data01 <- subset(PlantGrowth, group!="trt2") # 2 群だけ取り出す
> data01$group <- droplevels(data01$group) # 水準 trt2 を削除
> (fit.ttest <- t.test(weight~group, var.equal=T, data=data01)) # t 検定
```

Two Sample t-test

```
data: weight by group
t = 1.1913, df = 18, p-value = 0.249
alternative hypothesis: true difference in means is not equal to 0
```

²⁹ということは、 t 検定の結果や g とあわせて報告するのは矛盾している。

```

95 percent confidence interval:
  -0.2833003  1.0253003
sample estimates:
mean in group ctrl mean in group trt1
          5.032          4.661

> t.value <- as.numeric(fit.ttest$statistic) # t 値
> df <- fit.ttest$parameter                # 自由度
> library(effsize)
> cohen.d(weight~group, data=data01)       # Hedges の g

Cohen's d

d estimate: 0.5327478 (medium)
95 percent confidence interval:
      inf      sup
-0.4769679  1.5424635
> (prep <- pt(q=abs(t.value)/sqrt(2),df=df)) # p_rep
[1] 0.7946769

```

この例の場合、検定結果自体は有意ではないが ($t_{(18)} = 1.19, p = .249$)、効果量の符号の再現性については $p_{rep} = .795$ と高いと評価できそうである。