

## 実務と統計(6)

## —効果量 (Cohen の d と Hedges の g)—

山内俊平<sup>†</sup> (豊橋市立看護専門学校 非常勤講師,  
豊橋市食肉衛生検査所 元所長)



## はじめに

実務と統計(5)(79巻3号)で簡単に触れた「効果量」については、近年、「P値」,「差の信頼区間」とあわせて記述することが推奨されています。

効果量は「P値の代わり」ではなく、「役割の違う指標」ですが、単位を持たないため直感的な理解が難しいところがあります。今回、できるだけ“直観で伝わる”ように述べてみたいと思います。

## 効果量

効果量にはさまざまなものがありますが、2群間の差の大きさを示すのが、d 族効果量 (Cohen's d と Hedges' g) と呼ばれるものです。

## [Cohen's d]

効果量 d は、「母平均差 ÷ 共通母標準偏差」ですが、実際には母数は分からないため、これを「標本平均差 ÷ 等分散を仮定した 2 群共通標準偏差 (pooled SD)」によって推定します。仮に、計算値が  $d=1.5$  なら、“2 群の平均差が標準偏差の 1.5 倍 (1.5 個分) 離れている”となります。分母に SD を含むので効果量は無単位で、“同じ 2 群差”でも SD が大きいと、効果量は小さくなります。

効果量は単位を持たないため、大きさのイメージが難しい。Cohen は、d の大きさの目安 (あくまで目安であり、絶対基準ではない) に、 $d=0.2$  (小),  $d=0.5$  (中),  $d=0.8$  (大) の数字を提示しています。これを、d の大きさを示す“表現”として使用してもいいかと思えます。

## 〈例 示〉

$d=0.24$  (小さい効果),  $d=0.72$  (比較的大きい効果)

## 〈d の計算例 (概要)〉

A 群 ( $n=10$ , 平均 2, SD3)

B 群 ( $n=5$ , 平均 4, SD5)

この場合、標本平均差 = 2 です。SD は単純平均ではなく、 $n$  の違いを考慮した“共通 SD (pooled SD)”を統計公式から導きます。計算した pooled SD  $\approx 3.73$  で、効果量は  $d \approx 0.54$  (中程度の効果) となります。

## [Hedges' g]

Hedges' g は、Cohen's d がやや大きく見積もる傾向があるのを補正 ( $n$  が小さい場合 (例: 各群 20 未満など) で補正の影響が出やすい) した指標です。

効果量の記載にあたっては、計算が d か g なのかを明示する必要がありますが、統計的厳密性を考慮し、g を基準に d を併記するのでもいいかと思えます。

## 効果量の役割

2 群差の実務的判断は、基本的には P 値 (差は偶然ではないのか?) や差の信頼区間 (どの程度の差があるのか?) で行います (実務と統計 (1) 78 巻 11 号参照)。

しかし、単位の直接比較が難しいようなケースでは、効果量が、同一研究内での“差の大きさの感覚”を共有する道具となります。

また、単位を持たない効果量は、研究領域や測定尺度が異なる場合でも比較可能です。

## 〈例 示〉

たとえば  $d=0.5$  の効果量は、以下の状況下で同じ意味を持ちます。

- ・ 2 群の平均差 1 kg / 標準偏差 2 kg
- ・ 2 群の平均差 1 g / 標準偏差 2 g
- ・ 2 群の平均差 10 cm / 標準偏差 20 cm
- ・ 2 群の平均差 5°C / 標準偏差 10°C
- ・ 2 群の平均差 3 mmHg / 標準偏差 6 mmHg

測定単位に依存しない効果量は、研究領域にとらわれない共通尺度を提供することができます。これらの理由

<sup>†</sup> 連絡責任者: 山内俊平 (豊橋市立看護専門学校 非常勤講師, 豊橋市食肉衛生検査所 元所長)

から、「P 値」、「差の信頼区間」とあわせて「効果量」を記述することが推奨されます。

## 効果量の計算

### (1) AI 利用

一般に、2 群共通標準偏差 (pooled SD) は、統計ソフトで出力されない場合が多いかと思います。一方、統計公式での計算は、自由度や不偏分散などの知識を要し、初心者にはハードルが高いかも知れません。そんな場合は、対話型 AI を利用するのもいいかと思います (適正な条件提示、単位、入力値、出力の再計算などの確認は必要)。

### (2) 利用例

#### 〈独立 2 群の場合〉

独立 2 群の平均差比較では、近年は“等分散を仮定しない Welch の t 検定”が推奨されることが多い。一方、効果量は、“等分散を仮定した pooled SD による標準化平均差”で推定する方法が、広く用いられます (ただし、分散が極端に異なる場合は別手法もある)。

なお、効果量も推定値なので、95% 信頼区間 (95% CI) を求めます。

#### 条件提示例 (Chat GPT (公式サイト))

A 群 (n=25, 平均 4.0, SD 3.5), B 群 (n=20, 平均 2.0, SD 2.0) の独立 2 群を比較します。

Welch の t 検定 (両側) における① P 値, 平均差の 95% 信頼区間 (95% CI), ② 2 群共通標準偏差 (pooled SD), 効果量 (Cohen's d と Hedges' g) の点推定値と 95% 信頼区間 (95% CI) を計算してください。

### 結 果

① P 値 (両側) = 0.021

平均差 (95% CI) = 2.0 (0.32 ~ 3.68)

② pooled SD = 2.93

Cohen's d (95% CI) = 0.68 (0.08 ~ 1.29)

Hedges' g (95% CI) = 0.67 (0.08 ~ 1.26)

### 解 釈

① 平均差 2.0, 95% CI (0.32 ~ 3.68) に帰無仮説 (差 = 0) を含まず,  $P = 0.021 < 0.05$  で統計的に有意

② 効果量 (d, g) は約 0.7 (比較的大きい効果) で, 95% CI に 0 を含まず, 本例では検定結果と整合する (整合しない場合もあり得ます)。

### 〈対応のある 2 群の場合〉

対応のある 2 群の比較の効果量には、複数の算定方法がありますが、ここでは、差の標準偏差を用いる方法 (dz: 「各個人の前後差の平均」÷「個人差の標準偏差」) で計算します。

#### 条件提示例 (Chat GPT (公式サイト))

同一 18 人の前後比較で、個人差の平均が 2.8, 個人差の標準偏差が 5.1 の対応ある 2 群を比較します。Paired・t 検定 (両側) における① P 値, 平均差の 95% 信頼区間 (95% CI), ② 効果量 (Cohen's d と Hedges' g) の点推定値と 95% CI を計算してください。

### 結 果

① P 値 (両側) = 0.033

平均差 (95% CI) = 2.8 (0.26 ~ 5.34)

② Cohen's d (95% CI) = 0.55 (0.05 ~ 1.05)

Hedges' g (95% CI) = 0.53 (0.05 ~ 1.00)

### 解 釈

① 平均差 (大きさ) = 2.8 で, 95% CI に 0 を含まず,  $P = 0.033 < 0.05$  より統計的に有意

②  $d = 0.55$ ,  $g = 0.53$  の中程度の効果で, 95% CI に 0 を含まず, 本例では検定結果と整合する。

### 〈サンプルサイズ (n) 設計の場合〉

サンプルサイズの設計段階では母分散は未知、類似の先行研究などから予想 SD を仮定します。差の大きさは「実務的に価値があると考えられる差」を定義し,  $\alpha = 0.05$ , 検出力 0.8 程度で n を算定するのが一般的です (実務と統計 (2) 78 巻 12 号参照)。

#### 条件提示例 (Chat GPT (公式サイト))

群 1 の SD = 3.5, 群 2 の SD = 4.5 と仮定し, 2 群の平均差 = 3.0 を,  $\alpha = 0.05$  (両側), 検出力 0.8 で検出 (Welch の t 検定) することとします。

① 必要サンプルサイズ n (2 群の n 比 (1:1)), ② 2 群共通 SD と効果量 (Cohen's d と Hedges' g) を計算してください。

### 結 果

①  $n$  ( $\alpha = 0.05$  両側, 検出力 0.8, 1:1 割付) = 各群 29

② 共通 SD = 4.03

Cohen's d = 0.74, Hedges' g = 0.74

## 解 釈

“平均差3.0, 効果量で0.74の比較的大きな効果を狙ってn(各29)を設計した”といえます。EZRのn計算機能にAIで出力した共通SD (pooled SD)=4.03を入力し, nを再計算してみます。

2群の平均値の比較のためのサンプルサイズの計算	
2群間の平均値の差	3
2群共通の標準偏差 (SD)	4.03
$\alpha$ エラー (0.0-1.0)	0.05
検出力 (1- $\beta$ エラー) (0.0-1.0)	0.8
グループ1と2のサンプルサイズの比(1:X)	1

方 法

両側

one-sided

必要サンプルサイズ	計算結果
N1	29
N2	29

各群 n=29 (計 58) の同じ結果が得られます。

## ま と め

効果量は、同一研究内での差の大きさの把握や、他研究との比較・メタ解析などへの橋渡しの道具です。「P値」, 「差の信頼区間 (単位あり)」, 「効果量 (単位なし)」で、それぞれ主たる役割が異なります。3点セットでの記述が推奨されます。

統計検定の目的は“検定結果を統計学的に矛盾なく解釈し、実務や将来研究につなげる”ことです。複雑な理論に基づく計算は、いったん、統計ソフトやAIに委ね、“何を目的に、何を指標に、何を比較し、何が言えて、何が言えないのか”に集中する時間が重要です。