

2条件間の比較における効果サイズ

蔵 富 恵

Cohen's d effect size for two groups

Kei Kuratomi

要旨

近年、心理学の研究論文において、帰無仮説有意性検定 (NHST: Null Hypothesis Significance Testing) による有意差判断だけではなく、差の大きさを示す効果サイズの報告も求められ始めている。効果サイズは、帰無仮説有意性検定のような単なる差の有無だけを示すのではなく、条件間にどの程度の差があるのかを示すことができる。本論文では、複数ある効果サイズのうち、2条件間の差の大きさを示すCohen's d についてまとめる。Cohen's d は、平均値と標準偏差を用いて算出することができ、解釈が容易であることから、多くの研究で用いられている。しかし、Cohen's d は、実験計画に応じて、様々な算出方法が、これまでに考案されてきた。本論文では、参加者間計画および参加者内計画におけるCohen's d の算出方法を概説し、実際のデータを用いてCohen's d を算出した。その結果、ほとんどの式において同等の値が得られることは明らかとなったものの、研究報告時には、使用した効果サイズの算出方法を明記する必要がある。

キー・ワード：効果サイズ, Cohen's d , 実験計画

1. 差の大きさを示す効果サイズの必要性

近年、心理学の研究論文において、有意確率を用いた帰無仮説有意性検定 (NHST: Null Hypothesis Significance Testing) だけではなく、効果サイズの記載が求められる。最新の *Publication Manual of the American Psychological Association* (American Psychological Association, 2009) にも、推測統計を記載する際には、検定統計量、自由度、統計量が得られる確率 (正確な p 値) だけではなく、効果サイズと方向を含むことが求められている。そこで、本論文では、 t 検定において最もよく使用される効果サイズ、Cohen's d について紹介をする。

これまで帰無仮説有意性検定を用いた多くの研究では有意確率5%という基準によって差があるのか差がないのかという結論を求めてきた。しかし、この基準には問題があり、 $p = .049$ と $p = .051$ のように直感的には差がなさそうな有意確率でも、

5%水準を用いることによって、後者は有意差なしという結論として扱われる。したがって、数値的には、ほとんど差がないにも関わらず、帰無仮説有意性検定に基づく結論が変わる (ただし、多くの場合は傾向とする)。そこで、 p 値を報告する際には、不等号 (<, >) によって p 値を表現するのではなく、その値をイコール (=) によって明記することが求められる。しかし、注意しなければならないのは、有意水準は差があるということしか示しておらず、どの程度の差があるのかといった指標ではないという点である (大久保・岡田, 2012)。それに対して、どの程度の差があるかを示す指標が、効果サイズである。

効果サイズの必要性は数十年前から言われてきた (e.g., Cohen, 1988)。それにも関わらず、大久保 (2009) によれば、日本基礎心理学会発行の基礎心理学研究 (1982年から2008年対象) において、 p 値の報告が2000年代になるとほぼすべて

の論文で報告されているにも関わらず、効果サイズの報告はほとんどされていなかった。さらに、 t 検定（論文全体の24.70%）を用いた論文に至っては、効果量の報告は一つもなかった。

効果サイズの報告が求められる背景には、認知心理学の発展と同様に、コンピュータの発展が一つの要因である。コンピュータやインターネットの発展により、論文がPDFファイルとして容易に手に入れられることや、オンラインジャーナルの増加によって、様々な分野における研究蓄積が爆発的に増加した。その結果、先行研究で明らかになった知見を吟味、整理するために、メタ分析が行われるようになった。このようなメタ分析の際、実験手続きや人数が異なることから、各研究で得られた結果を標準化するために効果サイズが使用される。

Lakens (2013) は効果サイズを報告することの有用性を以下の3点にまとめている。1) 従属変数を測定するのに使用された単位に関わらず、理解されている標準的な測度において報告された効果の大きさを示すことができ、2) 研究間の標準的な効果サイズを比較することによって、メタ分析の結論を引き出せる。そして、3) 先行研究から得られた効果サイズは、検定力分析に用いることができるため、新たな研究計画の際に、必要な実験参加者数などを算出することができる。このように、帰無仮説有意性検定結果に加え、効果サイズを報告することは、今後の研究の発展にもつながるのである。

これまでいくつかの効果サイズが考案されている（大別すると d 族と r 族があるが、詳しくは、大久保・岡田, 2012を参照）。例えば、相関係数も効果サイズの一つである。本論文では、その中でも、2条件の平均値と標準偏差によって簡便に算出できるCohen's d について、いくつかの算出方法を紹介する。

2. 2条件間で用いられる効果サイズ

2条件間の効果サイズとして代表的なものにCohen's d がある。これは、単位に依存しないため、研究間で効果の比較としても使うこともできる。Cohen's d は、式(1)によって算出するこ

とができる (Cohen, 1988)。

$$d = \frac{M_A - M_B}{\sigma} \dots (1)$$

分子は、統制群 (M_A) と実験群 (M_B) の平均値の差を表し、分母は、二つの条件の標準偏差の平均 (σ) を表している (吉田, 1998)。つまり、平均値の差が標準偏差の何個分離れているかを示している。従って、Cohen's d の取り得る範囲は、0 (統制群と実験群に差がない) から無限となる。

本論文では、研究計画に応じて、具体的にCohen's d がどのように算出されるのかをLaken (2013)に基づいて紹介する。具体的な効果サイズを算出するため、2要因実験参加者内計画 (適合性×一致試行出現確率視野) で行われた蔵富・吉崎 (2010) のデータを用いる。これは、刺激が左右いずれかの視野に呈示されるフランカー課題 (Eriksen & Eriksen, 1974) を用いている。フランカー課題とは、文字列の中央の文字を同定する課題で、一般的に、一致試行 (例えば, “XXX XX”) に比べて不一致試行 (例えば, “XXNXX”) の反応時間が遅延し、その差は適合性効果と呼ばれる。重要な操作は、一方の視野の一致試行出現確率が75%とし、もう一方の視野のそれを25%としたことである。適合性効果が一致試行出現確率によってどのように影響しているのかに注目している。表1は、蔵富・吉崎 (2010) によって得

表1 各参加者の条件毎の適合性効果量 (ms)
(蔵富・吉崎, 2010)

ID	75%視野	25%視野
1	115	16
2	59	17
3	31	16
4	45	37
5	45	25
6	60	45
7	34	30
8	37	24
9	42	25
10	64	27
11	19	22
12	43	49
13	47	37
14	34	17
15	60	62
16	29	11

られた16名の適合性効果量である。表2は、表1のデータをもとに、条件別の平均値、不偏標準偏差、標本標準偏差、2条件間の相関係数を算出したものである。

本論文でも、適合性効果が一致試行出現確率視野によってどの程度異なるのかについて、Cohen's d を使って検討する。蔵富・吉崎(2010)では一致試行出現確率視野の要因は参加者内要因であったが、本論文では参加者間要因と参加者内要因で算出方法の異なるCohen's d によって、具体的にどのような値が算出され、どの程度値が異なるのかを示す。

表2 各条件の平均、標準偏差、相関係数
(蔵富・吉崎, 2010)

	75%視野	25%視野
平均	47.7	28.8
標本標準偏差	21.3	13.4
不偏標準偏差	22.0	13.8
相関係数	.10	

3. 参加者間計画におけるCohen's d

独立した2群間の効果サイズを求める際は、一般的に、式(2)が用いられる(大久保・岡田, 2012; 吉田, 1998)。

$$d = \frac{M_A - M_B}{\sqrt{\frac{n_A SD_A^2 + n_B SD_B^2}{n_A + n_B}}} \dots (2)$$

このときに用いられる SD はサンプル数で割った標本分散を使用した標準偏差である (i.e., エクセルではSTDEVP関数となる)。それゆえ、式(2)は、記述的な効果量であると言える(大久保・岡田, 2012)。表2の数値を用いて、75%視野と25%視野における適合性効果の差分をCohen's d によって算出するため、式(2)に値を代入すると、 $d = (47.7 - 28.8) / \sqrt{[(16 \times 21.3^2 + 16 \times 13.4^2) / (16 + 16)]} = 1.06$ となる。

また、Cohen (1988) は、式(3)を用いて、効果サイズを算出している。

$$d = \frac{M_A - M_B}{\sqrt{\frac{(n_A - 1)SD_A^2 + (n_B - 1)SD_B^2}{n_A + n_B - 2}}} \dots (3)$$

このときの SD はサンプル数-1で割った不偏

分散を使用した不偏標準偏差である (i.e., エクセルではSTDEV関数となる)。この式では、式(2)とは異なり、母集団における推測的な効果量を算出している(大久保・岡田, 2012)。従って、表2における75%視野条件の適合性効果の効果サイズは、 $d = (47.7 - 28.8) / \sqrt{[(16 - 1) \times 22.0^2 + (16 - 1) \times 13.8^2] / (16 + 16 - 2)} = 1.03$ となる。

注意しなければならないのが、式(3)のことは、Cohen's g やHedges's g と呼ぶこともある。このように式が同一にも関わらず文献によって呼び方が異なることが、効果サイズを報告する上で、混乱を招く原因となっている(Lakens, 2013; 大久保・岡田, 2012)。

式(3)は、 t 値と各群の人数が明らかな場合は、式(4)のように簡便に算出することもできる。

$$d = t \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \dots (4)$$

蔵富・吉崎(2010)で行われた実際の実験とは異なるが、75%視野と25%視野を参加者間計画だと仮定し、対応のない t 検定を行うと、 $t = 2.91$ となる。この t 値を式(4)に代入すると、 $d = 2.91 \times \sqrt{[(1/16) + (1/16)]} = 1.03$ となり、式(3)で算出した値と同じになる。

このように、式(3)および式(4)に比べて、式(2)によるCohen's d は、大きく算出される。これは、式(2)は、標準偏差に標本標準偏差を用いることから、記述的な効果量を算出し、式(3, 4)は、不偏標準偏差を用いることから推測的な効果量を算出している(大久保・岡田, 2012)ためである。どちらの式を用いるかは、その目的に応じて選択する必要がある。

4. 参加者内計画におけるCohen's d

参加者内計画の場合、2条件間のデータが対応するため、相関を考慮する必要がある。そこで、Rosenthal (1991) は、単なる2条件間の平均値だけではなく、2条件間の相関係数を考慮したCohen's d を考案している。これは、 t 値と人数を用いて式(5)のように算出することができる。

$$d = \frac{t}{\sqrt{n}} \dots (5)$$

例えば、表1における t 検定の結果は、 $t = 3.05$ となるため、 $d = 3.05/\sqrt{16} = 0.76$ となる。

一方で、実験デザインに依存せず、Cohen's d は同じ計算方法によって導かれるべきとする、メタ分析の立場もある (Borenstein, Hedges, Higgins, & Rothstein, 2009; 水本・竹内, 2008)。水本・竹内 (2008) や Cumming (2012) は、式 (6) のように、単純に両群の標準偏差の平均を分母として扱っている。これは、参加者内計画では、2条件間で人数が同数となることから、式 (2) や式 (3) を簡略化したものである。Cumming (2012) は、このときの標準偏差は不偏標準偏差を使用していることから、推測的な効果サイズを示している。ただし、標準偏差に標本標準偏差を代入することによって、記述的な効果サイズを算出することもできる。

$$d = \frac{M_A - M_B}{\sqrt{\frac{SD_A^2 + SD_B^2}{2}}} \dots (6)$$

実際、式 (6) によって、表1の75%視野条件の標本標準偏差と不偏標準偏差を用いて適合性効果の効果サイズをそれぞれ算出すると、記述的な効果量は、 $d = (47.7 - 28.8) / \sqrt{[(21.3^2 + 13.4^2) / 2]} = 1.06$ と式 (2) によって算出された値と同等に、推測的な効果量は、 $d = (47.7 - 28.8) / \sqrt{[(22.0^2 + 13.8^2) / 2]} = 1.03$ と式 (3) によって算出された値と同等になる。

相関係数が小さく、2条件間の標準偏差の差が大きいときを除いて、一般的に式 (6) は、式 (2) および式 (3) に近似した値が算出される。また、 t 値を用いた式 (5) の場合、相関係数が

0.5以上の場合、式 (6) よりも大きく算出され、反対に相関係数が0.5よりも小さいときには、式 (6) よりも小さく算出される (水本・竹内, 2010; Morris & DeShon, 2002)。実際、表2からも明らかなように、相関係数は、 $r = .10$ となり、式 (6) で算出した $d = 1.06$ よりも、式 (5) で算出した値 ($d = 0.76$) が小さくなっている。それゆえ、相関を考慮した式 (5) は、過大評価することから、Lakens (2013) は、実験参加者内計画の事態には、式 (6) を推奨している。

6. Cohen's d の解釈

Cohen's d は、式 (1) から明らかなように、 $d = 1$ の場合、二つの平均値が2条件の平均された標準偏差の一つ分離れていることを意味する。このCohen's d の大きさの基準はCohen (1988) によって、小 ($d = 0.2$)、中 ($d = 0.5$)、大 ($d = 0.8$) と定められている。図1、表3からもわかるように、 $d = 0.2$ は、 M_B の57.93%が M_A の平均値よりも高い人ことを示し、データ全体の92.03%が重複していることを意味している。 $d = 0.5$ は、 M_B の69.15%が、 $d = 0.8$ は78.81%が M_A の平均値よりも高いことを示し、さらにそれぞれ80.26%、68.92%のデータが重複していることを示している。このような値に関しては、Web上で確認することもできる (Magnusson, 2014)。従って、式 (6) に基づけば、表1の適合性効果

表3 Cohen's d と2条件のデータ分布

Cohen's d	0.2	0.5	0.8
2条件間でデータが重複する割合	92.03%	80.26%	68.92%
1条件のデータがもう1条件の平均値以上になる割合	57.93%	69.15%	78.81%

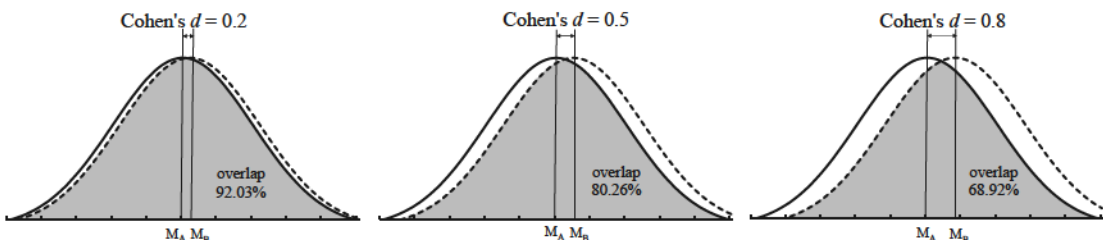


図1 Cohen (1988) の基準におけるCohen's d Overlapは2群間の重複の割合を表す

の記述的な効果サイズは、 $d = 1.06$ となるため、75%視野の適合性効果の85.54%のデータが、25%視野の適合性効果よりも大きいことを示している。

これらの効果サイズの解釈に基づけば、平均値、標準偏差が記載された先行研究も効果サイズを算出することによって、差の大きさを解釈することができる。例えば、長年引用され続けているストロープ効果 (Stroop, 1935) を最初に示した研究 (実験2) では、インクの色と色名单語の意味が異なる事態のインクの色読み (不一致条件) は、平均110.3秒 ($SD = 18.8$)、インクの色と色名单語の意味が同じ事態のインクの色読み (一致条件) は、平均63.3秒 ($SD = 10.8$) であった。これらの平均値と標準偏差を用いて効果サイズ (ただし、記述的か推測的かは明らかではない) を式 (6) によって算出すると、 $d = 3.07$ となる。従って、不一致条件の読み上げ時間は、一致条件の平均読み上げ時間よりも、99.89%のデータで遅く、12.55%のデータが重複している。このような効果サイズの大きい現象は、再現性も高くなる。

7. まとめ

t 検定において最もよく用いられる効果サイズにはCohen's d がある。数多くある算出方法の中から、Lakens (2013) によってまとめられた方法を中心に紹介した。実験計画によって算出方法が異なることや、同じ式でも効果サイズの名称が異なることによって、いずれの式を用いるべきか、混乱を招いている (Lakens, 2013; 大久保・岡田, 2012)。このことを踏まえ、効果サイズを使って議論する際は、どのような式によってCohen's d を算出したか、その出典となる書籍、論文を記載することが推奨される (Cumming, in press; 水本・竹内, 2010; 大久保・岡田, 2012)。また、計算に関しては、Web上で行えるものも多数存在する (e.g., Becker, 2000; Ellis, 2009; Wiseheart, 2013) が、それぞれのWebサイトがどのような算出方法を用いているのかは確認する必要がある。

本論文では、Cohen's d について、考案されている算出方法について、Lakens (2013) に基づ

いて紹介した。近年では、Cumming (2012) や大久保・岡田 (2012) のような心理学者による心理学者のための統計の書籍も数多く出版されるようになった。また、新しい効果サイズの考え方や取扱方法についての議論は、今後も続くと考えられる。われわれ心理学者は今後も、新しい科学的な知見を供給するために、統計的手法についてしっかりした知識を得る必要があるだろう。

付 記

この論文をまとめるにあたり、コメントをいただいた愛知淑徳大学の吉崎一人先生に感謝の意を表します。

引用文献

- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed). Washington, DC: Author.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons.
- Becker, L. A. (2000, March 20). Effect size calculators [Web site]. Retrieved from <http://www.uccs.edu/~lbecker>
- Cohen, J. (1988). *Statistical Power analysis for the Behavioral Sciences*. New York: Routledge Academic.
- Cumming, G. (2012). *Understanding the New Statistics: Effect sizes, confidence intervals, and Meta-Analysis*. New York: Routledge.
- Cumming, G. (in press). The new statistics: Why and how. *Psychological Science*.
- Ellis, P. D. (2009, January 9). Effect size calculators [Web site]. Retrieved from <http://www.polyu.edu.hk/mm/ effectsizefaqs/calculator/calculator.html>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects

- of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143-149.
- 蔵富恵・吉崎一人 (2010). 一致試行の出現確率と呈示視野が適合性効果に及ぼす影響 日本認知心理学会第8会大会発表論文集, 120.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, **4**, 863.
- Magnusson, K. (2014, February 3). Interpreting Cohen's *d* effect size: An interactive visualization [Web site]. Retrieved from <http://rpsychologist.com/d3/cohend>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, **7**, 105-125.
- 水本篤・竹内理 (2010). 効果量と検定力分析入門ー統計的検定を正しく使うためにー『より良い外国語教育研究のための方法』外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会2010年度報告論集, 47-73.
- 大久保街重 (2009). 日本における統計改革ー基礎心理学研究を資料としてー 基礎心理学研究, **28**, 88-93.
- 大久保街重・岡田謙介 (2012). 伝えるための心理統計 効果量・信頼区間・検定力 勁草書房
- 小野寺孝義・菱村豊 (2005). 文科系学生のための新統計学 ナカニシヤ出版
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: SAGE Publications, Incorporated.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.
- Wiseheart, M. (2013, December 2). Effect size calculator [Web site]. Retrieved from <http://www.cognitiveflexibility.org/effectsize/>
- 吉田寿夫 (1998). 本当にわかりやすいすぐ大切なことが書いてあるごく初歩の統計の本 北大路書房