DS-EFA「平均値差検定システム」の効果量について（ver. 1.0.0）

宮城教育大学教職大学院
教授　田端　健人

効果量では「Cohen の d」と「Hedges の g」が有名です。これらは 2 つのサンプル（グループ）間の平均値の差を標準偏差で割って標準化したもので、2 グループの平均値差を比較します。Cohen の d は「記述的効果量」であり、サンプルの 2 群の差を記述した値で、Hedges の g は「推定的（推測的）効果量」であり、サンプルが代表する 2 群の母集団の差を推測する値です。以下の数式からわかるように、サンプル数が大きくなるにつれて、Cohen の d と Hedges の g とは近似・一致します。

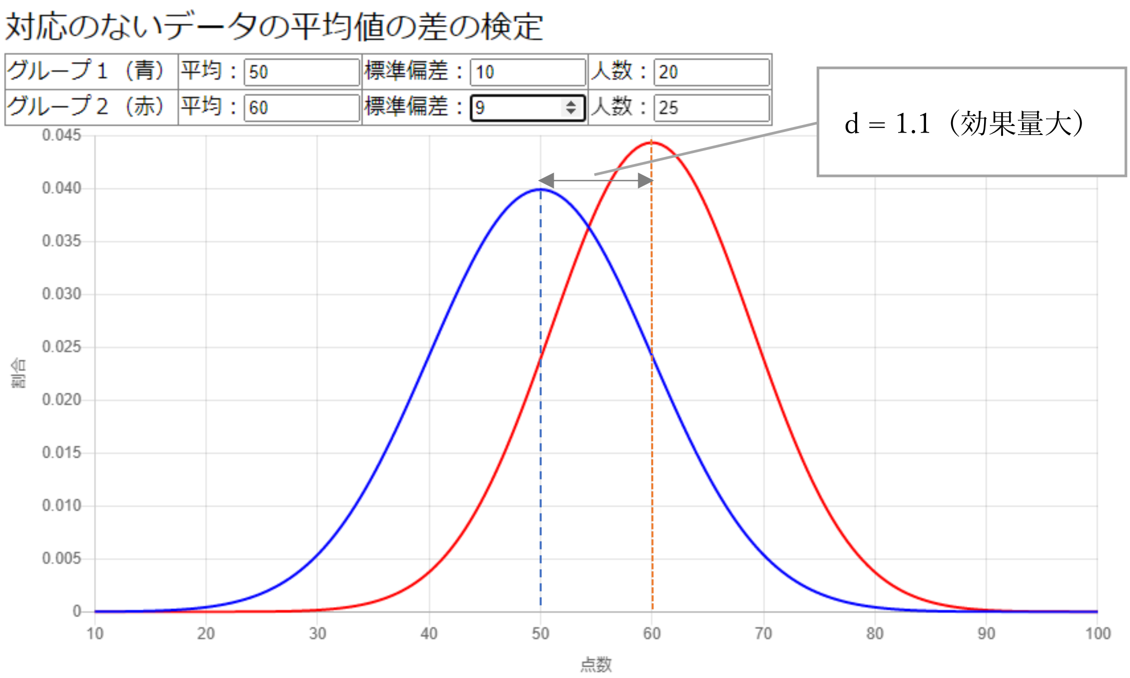
効果量は、「対応のない場合」つまり異なるグループの平均値差を比較する場合と「対応のある場合」つまり同一グループの平均値差を比較する場合で、計算式が異なります。ただ、対応のある場合の効果量については、対応のない場合と同じ計算式を使うという考えもあります（cf.，大久保・岡田，pp. 64-66）。

1． 対応のない場合の効果量

1． 1． 効果量の概略と計算式

まずは対応のない場合を説明します。

私たちが開発した「平均値差検定システム」のグラフで示すなら、効果量は、グループ 1（青）とグループ 2（赤）の平均値の開きの大きさを、両グループの標準偏差を目安に d 値あるいは g 値として評価します。



グループ 1 の平均値M_1を、人数（サンプルサイズ）をn_1、分散をS_1^2（正確には「標本分散」）とし、グループ 2 の平均値をM_1、人数（サンプルサイズ）をn_2、分散をS_2^2とすると、それぞれの効果量は次の式で算出します。ちなみに、「分散」の平方根が「標準偏差」で、

逆に「標準偏差」を 2 乗すると「分散」になります。

なお以下の Cohen の d と Hedges の g の計算式は、大久保・岡田（2012）の 50 頁、55-56 頁を参照しました。ただ値がマイナスをとらないよう、絶対値としています。

Cohen の d の計算式は、2 群の平均値差を「プールした標準偏差（S_p）」で割ります。

$$d = \frac{|M_1 - M_2|}{S_p}$$

$$S_p = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}}$$

Hedges の g の計算式は、「標本分散（S^2）」（大文字の S）ではなく、「不偏分散（s^2）」（小文字の s）を使います。「標本」とはサンプル（グループ）のことです。

$$g = \frac{|M_1 - M_2|}{s_p}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

１．２．グループの平均と標準偏差の出し方～エクセルの活用～

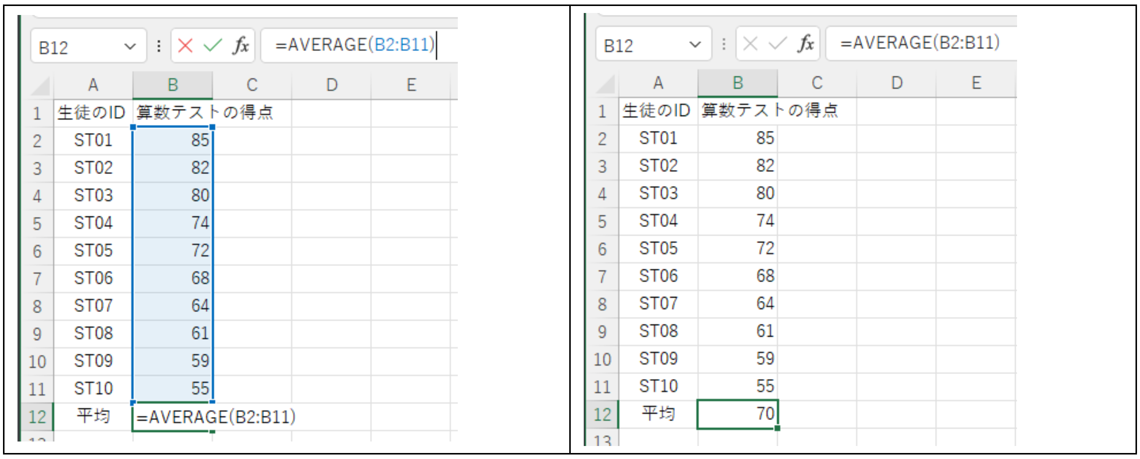
ですので、いずれの場合も、①「各グループの平均」②「標準偏差（標本分散あるいは不偏分散の平方根）」③「人数」という 3 つの統計量がわかってなくてはなりません。グループの平均と標準偏差は、エクセルに元データがあれば、関数で簡単に計算できます。エクセルでグループの平均と標準偏差を出すやり方は、以下になります。架空のデータセットを例に説明します。

まずエクセルで、次のような形式のデータセットを作成します。実際には、テストスコアの数値は自校の生徒のデータになります。スコアは学力テストでも、アンケートの得点でも同じ扱いです。

	A	B	C	D
1	生徒のID	算数テストの得点		
2	ST01	85		
3	ST02	82		
4	ST03	80		
5	ST04	74		
6	ST05	72		
7	ST06	68		
8	ST07	64		
9	ST08	61		
10	ST09	59		
11	ST10	55		
12				
13				

12 行目の B 列（B12）で平均値を計算するエクセルの関数 AVERAGE を入れます。半角英数で「=ave」と入力すると、自動的に関数の候補が出てきますので、AVERAGE 関数を指定

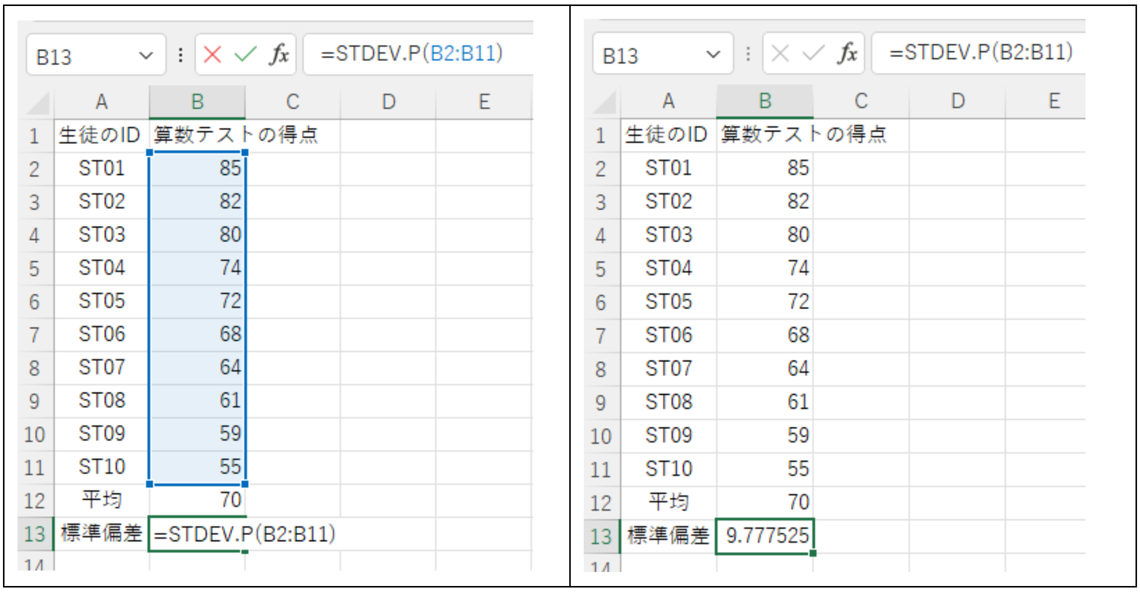
します。この関数の括弧の中に、B 列の 2 行目（B2）のセルから、B 列 11 行目（B11）を指定します（図左）。



これを実行すると、平均値 70 が表示されます（図右）。

同様に標準偏差（標本分散の平方根）を、エクセルの STDEV.P 関数で計算します。半角英数で「=std」と入力すれば候補が出るのでそれを選択し、セルを範囲指定します。

Hedges の g の場合は、不偏分散とその平方根の不偏標準偏差なので、STDEV.S 関数か、STDEV 関数を選択します。



標準偏差（S）は 9.78 と計算されています。

これらの数値がわかれば、ウェブアプリ「平均値差検定システム」に、2 グループの平均、STDEV.P 関数で求めの標準偏差、人数を入力します。そうするとシステムが、効果量（Cohen の d と Hedges の g）を出力します。

ちなみに、この値の 2 乗が分散ですので、エクセルの B14 のセルに=9.78^2 と入力しエンターキーを押すと、95.65 という値を返してくれます。またエクセルでは VAR.P 関数でデータセットから直接分散を計算できるので、B15 セルにこの関数を入れ、カッコ内に B2 から B11 を指定し、実行すると B14 と同じ値を返してくれます。

B15 | =VAR.P(B2:B11)

	A	B	C	D	E
1	生徒のID	算数テストの得点			
2	ST01	85			
3	ST02	82			
4	ST03	80			
5	ST04	74			
6	ST05	72			
7	ST06	68			
8	ST07	64			
9	ST08	61			
10	ST09	59			
11	ST10	55			
12	平均	70			
13	標準偏差	9.777525			
14	分散	95.6484	<-- 「=9.78^2」の計算値		
15	分散	95.6	<-- VAR.P関数での計算値		

１．３．補足的解説～標準偏差と分散とは何か？～

ここからは、統計の基礎基本の理解を進める補助的な解説です。

そもそも標準偏差（S）とか分散（S^2）とは、何なのでしょうか。これらはデータのばらつきを意味します。それぞれの生徒のスコアが、グループの平均値とどれほど差があるかを全体として評価する指標です。各生徒のスコアと平均との差（「偏差」）を 2 乗し、全員分の偏差の合計を人数で割り、グループの偏差の平均をとると、これが分散（標本分散）になります。母集団を推定する不偏分散の場合は、人数から 1 を引いた数で割ります。分散の平方根が標準偏差です。先のデータセットで計算すれば、以下の表になります。

	A	B	C	D	E	F
1	生徒のID	算数テストの得点（x）	偏差（x - M）	偏差の2乗（x－M）²		
2	ST01	85	15	225		
3	ST02	82	12	144		
4	ST03	80	10	100		
5	ST04	74	4	16		
6	ST05	72	2	4		
7	ST06	68	-2	4		
8	ST07	64	-6	36		
9	ST08	61	-9	81		
10	ST09	59	-11	121		
11	ST10	55	-15	225		
12	平均（M）	70	偏差の合計	偏差の2乗の合計		
13			Σ（x-M）	Σ（x-M）²		
14			0	956		9.777525

偏差の 2 乗を合計した D14 セルの値 956 を人数（n=10）で割って平均値を出せば、95.6 となり、先に計算した分散の値になります。この平方根をとれば（エクセルでは SQRT 関数を使い「=SQRT(95.6)」と入力）、標準偏差 9.78 となります。

上記データセットの分散（標本分散）の計算式は、

$$s^2 = \frac{(85-70)^2 + (82-70)^2 + (80-70)^2 + (74-70)^2 + (72-70)^2 + (68-70)^2 + (64-70)^2 + (61-70)^2 + (59-70)^2 + (55-70)^2}{10}$$

となります。分子が（個人スコア − 平均値）^{2}の合計なので、総和を表す記号Σ（シグマ）を用いて表記を変えると、分散の計算は次の一般式になります。

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - M)^2}{n}$$

この式は、次のようにも表記できます。

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2$$

これを平方根で開けば標準偏差ですので、標準偏差の一般式は次になります。

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Hedges の g で使う不偏分散（小文字の s^2）の一般式は、サンプル数（n）から 1 を引いた数で割ります。

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - M)^2}{n - 1}$$

それゆえ、Hedges の g で使う不偏標準偏差の一般式は、

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - M)^2}$$

になります。

１．４．対応のない 2 グループの効果量の計算例

では架空のデータで、対応のない 2 グループの効果量を計算してみましょう。

先に紹介したグループを 6 年 1 組としましょう。6 年 2 組は効果的な学習（例えば、対話と探究学習）を取り入れ、平均点が 1 組よりも上がったという架空の設定です。

データセットは次のようになります。人数、平均、標準偏差、不偏標準偏差を入れています。

生徒のID	算数テストの得点
ST101	85
ST102	82
ST103	80
ST104	74
ST105	72
ST106	68
ST107	64
ST108	61
ST109	59
ST110	55
ST201	87
ST202	89
ST203	85
ST204	80
ST205	79
ST206	75
ST207	76
ST208	74
ST209	65
ST210	63

	人数（n）	平均（μ）	標準偏差（S）	不偏標準偏差（s）
6年1組	10	70	9.78	10.31
6年2組	10	77.3	8.21	8.65

Cohen の d の計算：

$$d = \frac{|70 - 77.3|}{S_p}$$

$$S_p = \sqrt{\frac{10(9.78)^2 + 10(8.21)^2}{10 + 10}}$$

$$S_p = \sqrt{\frac{956.48 + 674.00}{20}} = \sqrt{\frac{1630.48}{20}} = \sqrt{81.52} = 9.03$$

$$d = \frac{|-7.3|}{9.03} = 0.808$$

Hedges の g の計算：

$$g = \frac{|70 - 77.3|}{s_p}$$

$$s_p = \sqrt{\frac{(10-1)(10.31)^2 + (10-1)(8.65)^2}{10 + 10 - 2}}$$

$$s_p = \sqrt{\frac{9 * 106.30 + 9 * 74.82}{18}} = \sqrt{\frac{1630.08}{18}} = \sqrt{90.56} = 9.52$$

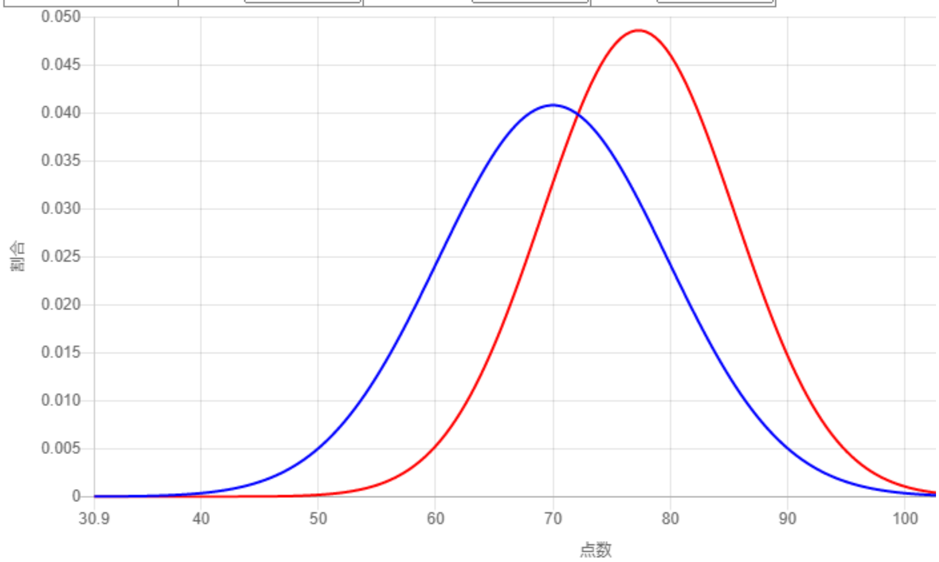
$$g = \frac{|-7.3|}{9.52} = 0.767$$

システムの出力と照合してみましょう。

システムの「標準偏差」にはエクセルの STDEV.P 関数で出力される標準偏差を入力してください。

対応のないデータの平均値の差の検定

グループ１（青）	平均：70	標準偏差：9.78	人数：10
グループ２（赤）	平均：77.3	標準偏差：8.21	人数：10



検定結果
平均の差(Deviation of mean values)： 7.30
効果量Cohenのd(Effect Size Cohen's d)：0.808
効果量Hedgesのg(Effect Size Hedges' g)：0.767
検定統計量t値(t-value)： 1.808
P値(P[T<=t])： 0.08836
95%信頼区間の下限(CI lower limit)：-1.183
95%信頼区間の上限(CI upper limit)：15.783
有意水準５%で帰無仮説は、棄却されない（差はない）

Cohen の d も、Hedges の g も、上記計算結果とシステムの出力が一致しました！

１．５．対応のない場合のＲスクリプト

大久保・岡田（2012）の 193 頁からの転載です。「架空のサンプルデータ」は本解説で紹介したデータです。「絶対値」での出力は田端が加筆しました。

```
x1 <- c(85, 82, 80, 74, 72, 68, 64, 61, 59, 55)  # 架空のサンプルデータ
x2 <- c(87, 89, 85, 80, 79, 75, 76, 74, 65, 63)

# 要約統計量
m1 <- mean(x1)
```

```
m2 <- mean(x2)
n1 <- length(x1)
n2 <- length(x2)
s1 <- sd(x1)
s2 <- sd(x2)

# Cohen's d
Sp <- sqrt(((n1-1)*s1^2 + (n2-1)*s2^2)/(n1+n2))
d <- (m1 - m2)/Sp
d

# 絶対値での d の出力
d <- sqrt((m1 - m2)^2)/Sp
d

# Hedges' g
sp <- sqrt(((n1-1)*s1^2 + (n2-1)*s2^2)/(n1+n2-2))
g <- (m1 - m2)/sp
g

# 絶対値での g の出力
g <- sqrt((m1 - m2)^2)/sp
g
```

２．対応のある場合の効果量

２．１．「差得点の効果量」

対応のある場合、例えば同じ 6 年 1 組の 4 月と 7 月の算数実力テストの平均値差の効果量の場合、2 つのやり方があります。1 つは対応のない倍と同じにするやり方で、もう 1 つは個々人の点差から効果量を求めるやり方です（cf.，大久保・岡田，pp. 64−66)。私たちのシステムには「対応のないデータ」として後者の計算をします。

1 回目テスト得点（x_1）から 2 回目テスト得点（x_2）を引いた各自の得点差

$$x_D = x_1 - x_2$$

を全員分足し合わせて人数で割るとグループの平均値（M_D）が出ます。これを得点差の不偏標準偏差（s_D）で割ると対応のあるデータの効果量（d_D）になります。

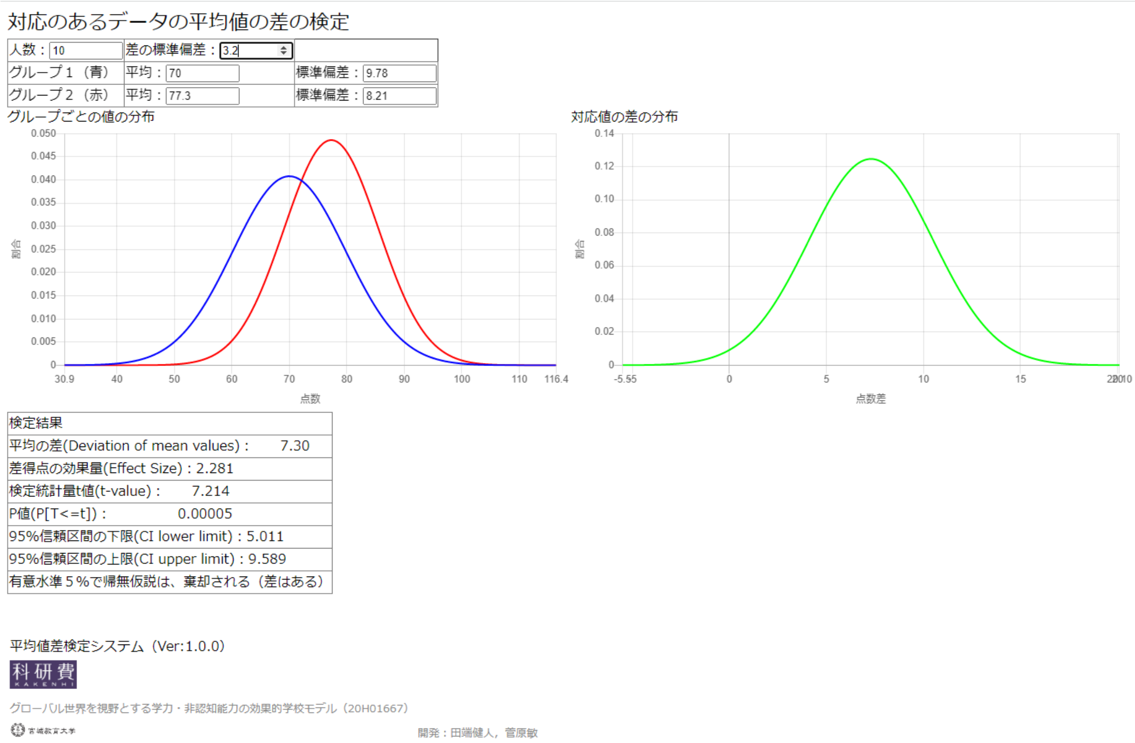
$$d_D = \frac{M_D}{s_D}$$

架空のデータでやってみましょう。

	A	B	C	D	E	F	G	H
1	生徒のID	算数テストの得点 1回目 x1	算数テストの得点 2回目 x2	1回目と2回目の差 x1 - x2				
2	ST101	85	87	-2				
3	ST102	82	89	-7				
4	ST103	80	85	-5				
5	ST104	74	80	-6				
6	ST105	72	79	-7				
7	ST106	68	75	-7				
8	ST107	64	76	-12				
9	ST108	61	74	-13				
10	ST109	59	65	-6				
11	ST110	55	63	-8				
12				-73	<-- 差の合計			
13				-7.3	<-- 差の平均			
14				3.20	<-- 差の不偏標準偏差（STDEV.S関数）			

M_Dが−7.3、s_Dが 3.20 なので、割り算をすると$d_D = -2.28$となります。2 回目の方が得点が高いので、$d_D = 2.28$とするのがよいでしょう。

システムと照合してみましょう。



システム出力「差得点の効果量（Effect Size)」2.28 と一致しました！

２．２．「差得点の効果量」のＲスクリプト

大久保・岡田（2012）の 194 頁からの転載です。

```
# 対応のある 2 群の平均値の比較
xd <- x1 - x2
md <- mean(xd)
sd <- sd(xd)
dd <- md/sd
dd
```

引用文献

大久保街亜・岡田謙介．2021『伝えるための心理統計』勁草書房.