2016/11/26@大阪府立大学
中国語教育学会2016年度第2回研究会
【共催】大阪府立大学高等教育開発センター

教育実践論文を書くための統計解析入門
―授業実践を研究に―

水本　篤（関西大学）

本日のおはなし

授業実践を論文や研究発表にしてみたいと考えているが，平均の意味はわかるが，標準偏差と言われると自信がない，ましてや*t*検定や効果量などという言葉は聞いたこともないという方が対象。

発展編や質的研究の詳細

統計解析の基礎的知識

1. 記述統計（平均と標準偏差）
2. 統計的検定
3. 効果量

統計解析の基礎的知識

1. 記述統計（平均と標準偏差）

2. 統計的検定

3. 効果量

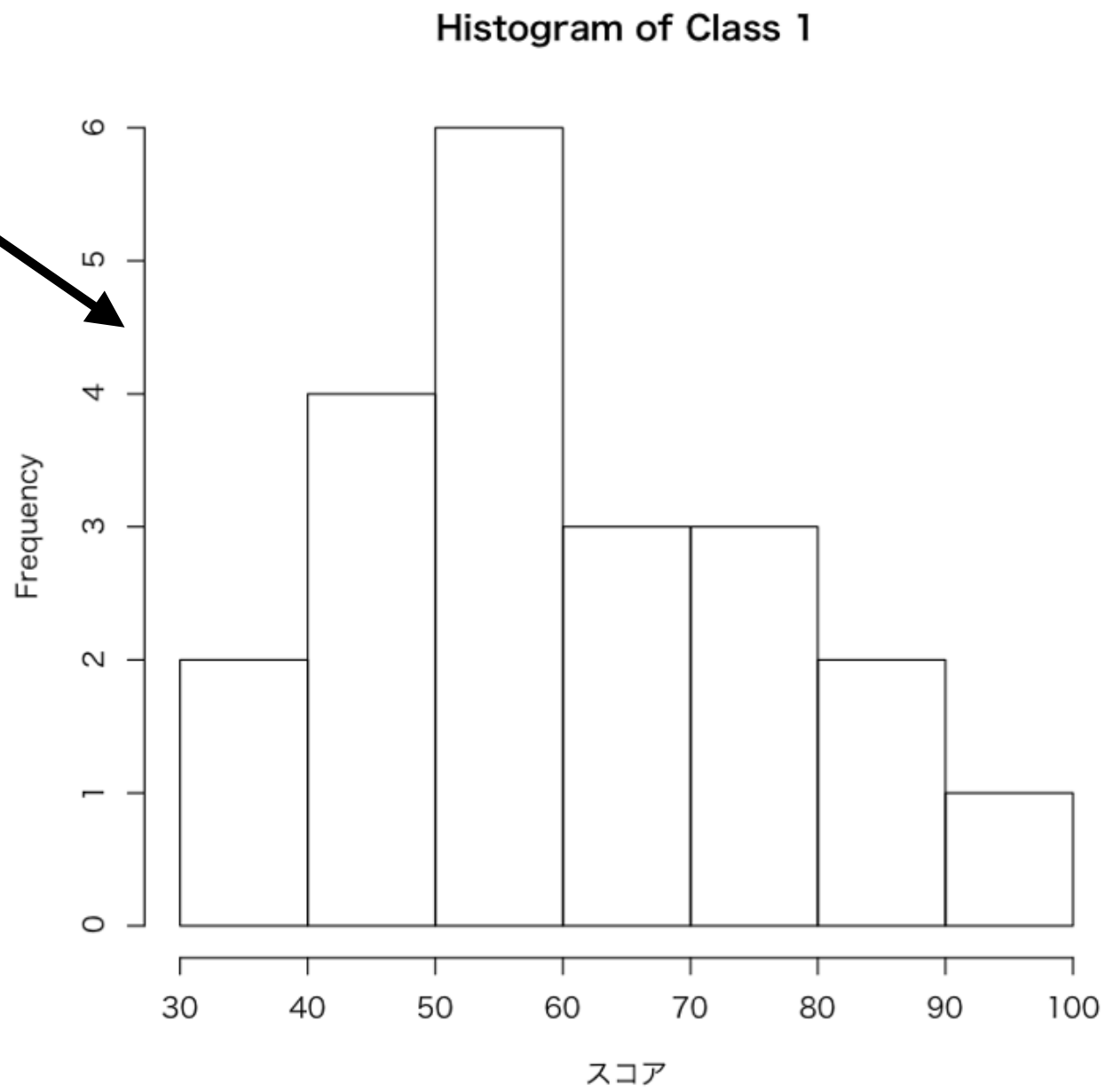
記述統計
(descriptive statistics)

- 平均値（mean: *M*）
- 標準偏差（standard deviation: *SD*）

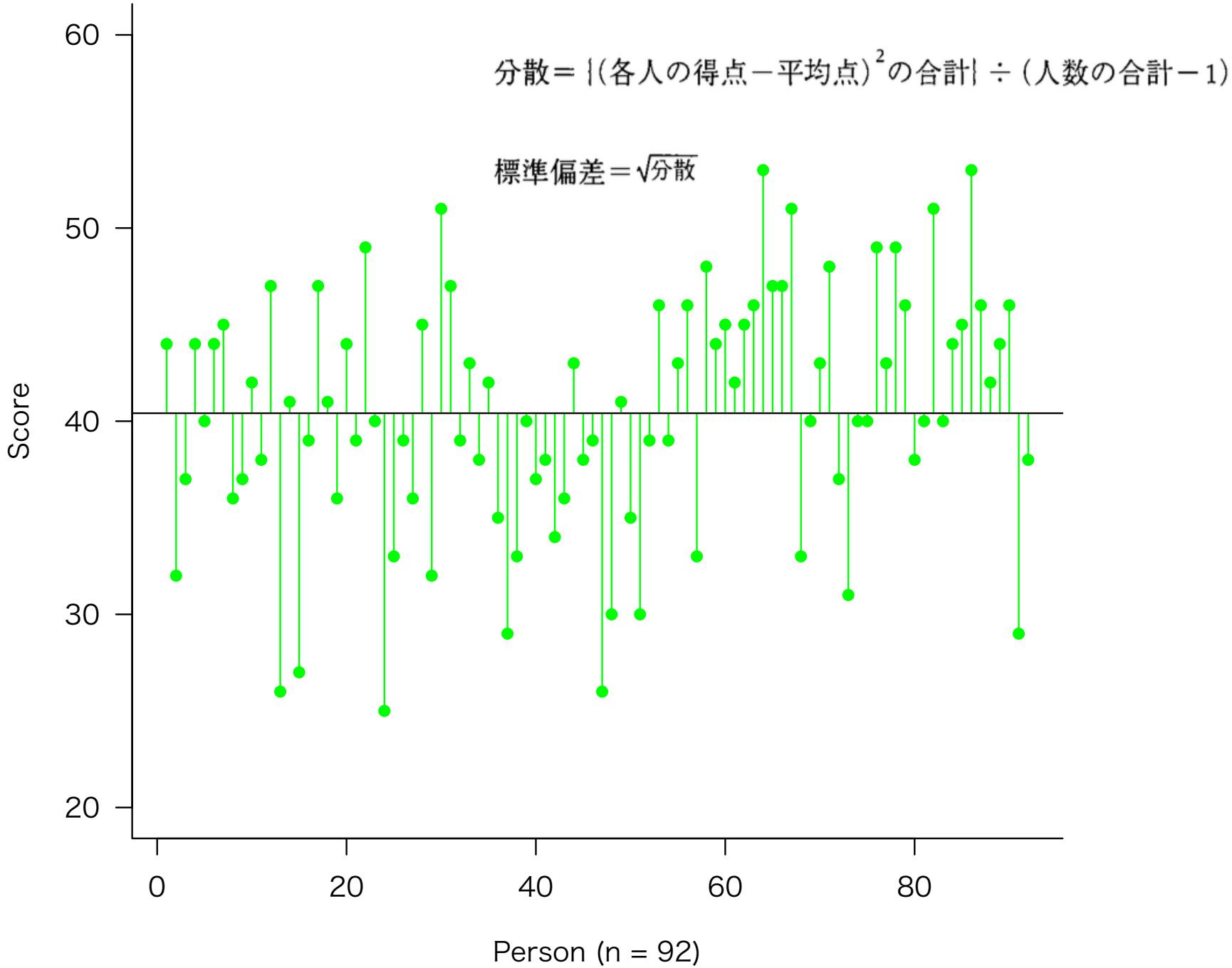
平均値と標準偏差

Group	n	M	SD
1	21	61.33	16.44
2	24	59.79	18.51

	A	B	C
1	氏名	クラス	スコア
2	新倉泰憲	1	50
3	桑原由菜	1	56
4	笠原政信	1	79
5	藤島麻美	1	99
6	武田麻由	1	56
7	本山和徳	1	66
8	沼田哲美	1	67
9	坂野真央	1	81
10	谷本秀子	1	55
11	山内晴菜	1	44
12	黒田真衣	1	45
13	川井強	1	43
14	首藤輝	1	77
15	堀内雄太	1	72
16	古山秀光	1	60
17	村岡素子	1	37
18	都築初太郎	1	39
19	諏訪静雄	1	56
20	小坂夢	1	66
21	前原隆明	1	85
22	門脇沙也香	1	55
23	谷村新次郎	2	22
24	細田若奈	2	100
25	浜真美	2	45
26	大山果音	2	66
27	志賀麗奈	2	77
28	城戸果穂	2	88
29	植松晴美	2	76
30	前川好子	2	79
31	富樫麻里	2	44
32	関谷遥花	2	55
33	小笠原朋美	2	65
34	中島昌彦	2	76
35	新井邦久	2	66
36	森脇亜矢子	2	44
37	高島貴子	2	32
38	小柳健治	2	55
39	黒澤一彦	2	56
40	鹿島沙也佳	2	57
41	森本正毅	2	77
42	河井美雨	2	65
43	秋山杏子	2	40
44	岩瀬保夫	2	41
45	田島孝男	2	49
46	内海更紗	2	60
47			



平均値と分散（標準偏差）

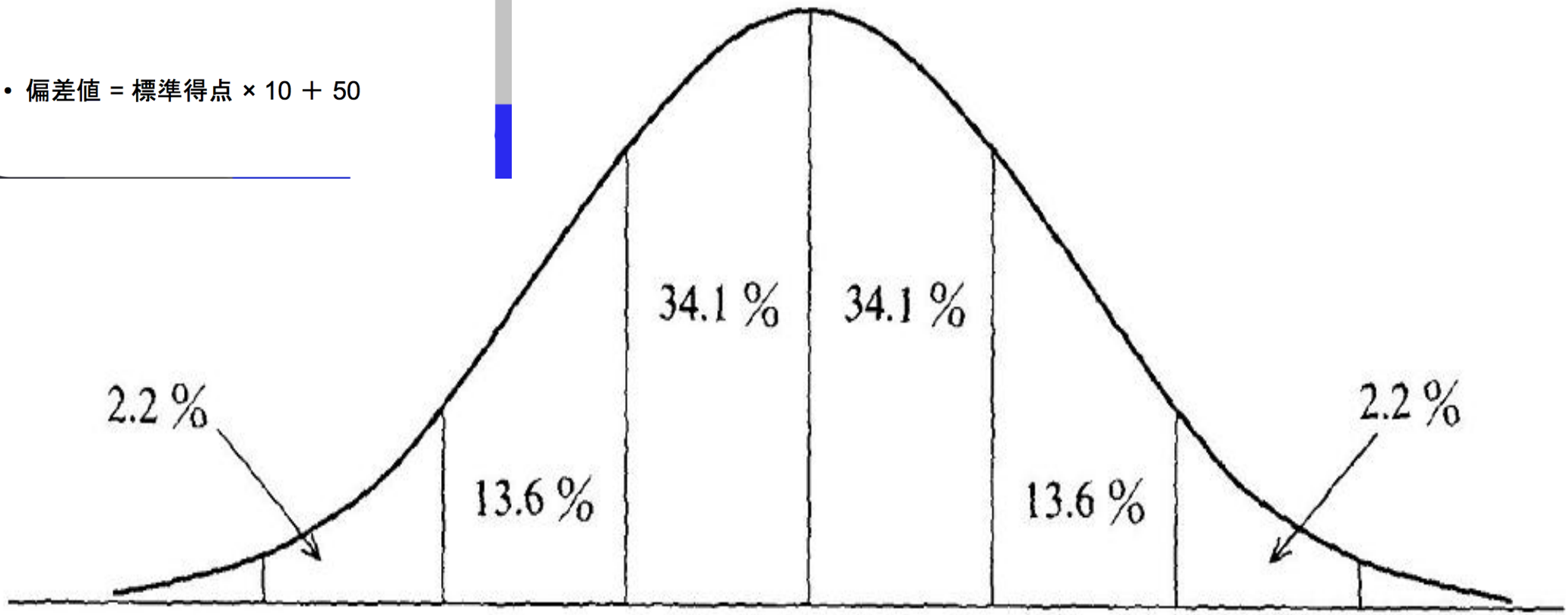

分散＝{(各人の得点－平均点)^{2}の合計} ÷ (人数の合計－1)
標準偏差＝$\sqrt{分散}$
Score
60
50
40
30
20
0
20
40
60
80
Person (n = 92)

No.	変数1(xi)	(xi−平均)・偏差	(xi−平均)^2		
1	5	−4	16		
2	6	−3	9		
3	9	0	0		
4	11	2	4		
5	14	5	25	↓偏差平方和/(データ数[n]−1)	↓√分散
	9	0	54	13.5	3.674
	平均	偏差の和	偏差平方和	(不偏)分散	標準偏差
偏差平方和でデータの散らばりがわかるが					
nが増えればそれだけ値が大きくなる			VAR(B2:B6)	13.5	3.674
↓				↑分散=偏差平方和/(データ数[n]−1)	n−1で計算した標準偏差
分散でそろえる					
↓			VARP(B2:B6)	10.8	3.286
標準偏差で元の単位に戻して考える				↑分散=偏差平方和/データ数(N)	Nで計算した標準偏差

標準得点と偏差値

- 標準(化)得点 (z score)=
  (個人の得点－平均点)/標準偏差

- 偏差値 = 標準得点 × 10 ＋ 50

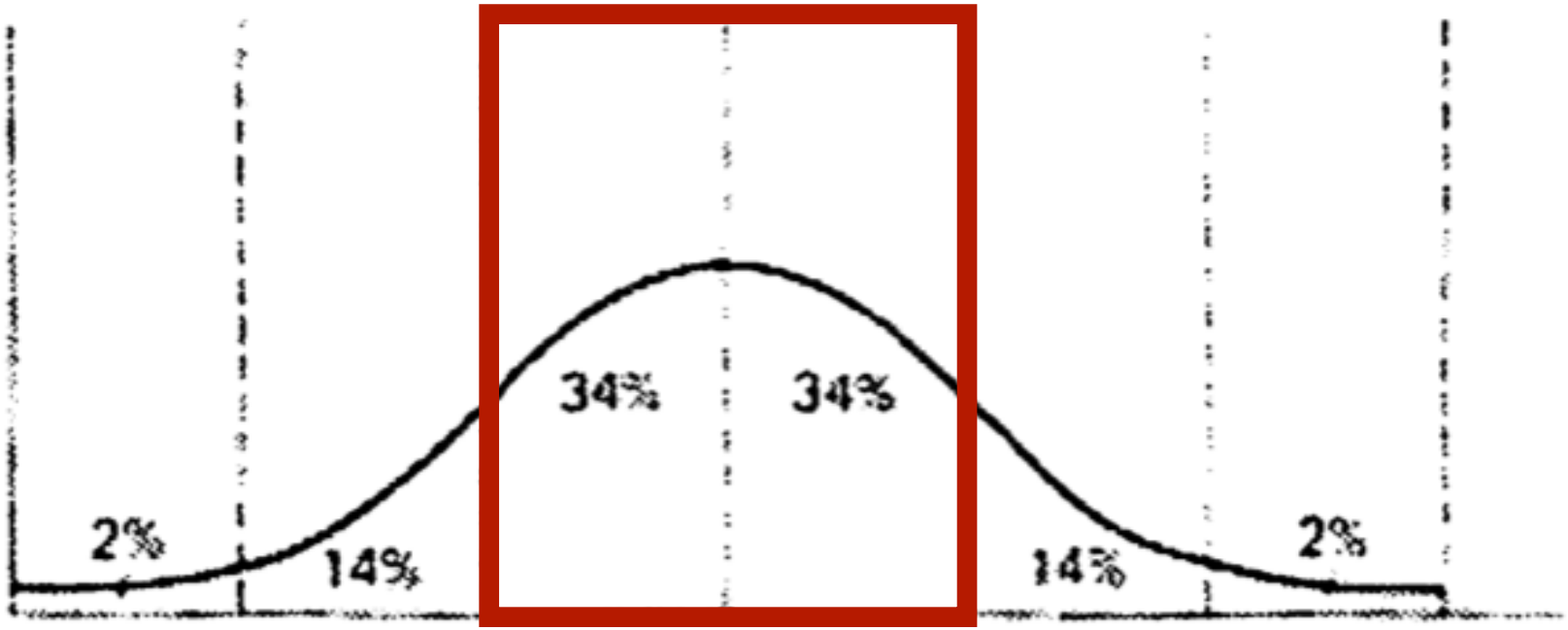


標準偏差	−3	−2	−1	平均	+1	+2	+3
標準得点	−3	−2	−1	0	+1	+2	+3
偏差値	20	30	40	50	60	70	80

図 3-2　標準偏差，標準得点，偏差値と分布の関係

前田 他(2004, p.24)

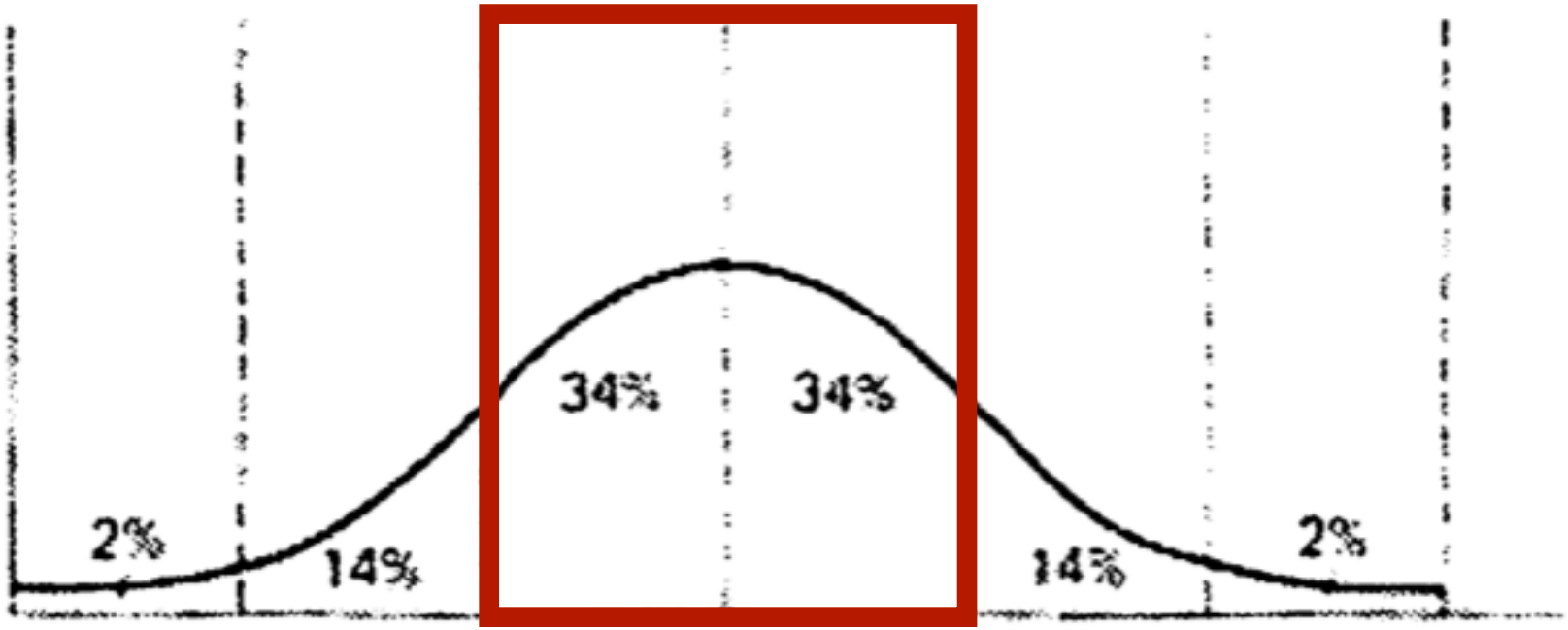
テストなど



Raw Score	0	10	20	30	40	50	60
Standard Deviation	-3 *SD*	-2 *SD*	-1 *SD*	$\bar{x}$	+1 *SD*	+2 *SD*	+3 *SD*
z score	-3	-2	-1	0	+1	+2	+3
T score	20	30	40	50	60	70	80

平均30点，標準偏差10点

質問紙など



Raw Score				2.8			
Standard Deviation	-3 *SD*	-2 *SD*	-1 *SD*	$\bar{x}$	+1 *SD*	+2 *SD*	+3 *SD*
z score	-3	-2	-1	0	+1	+2	+3
T score	20	30	40	50	60	70	80

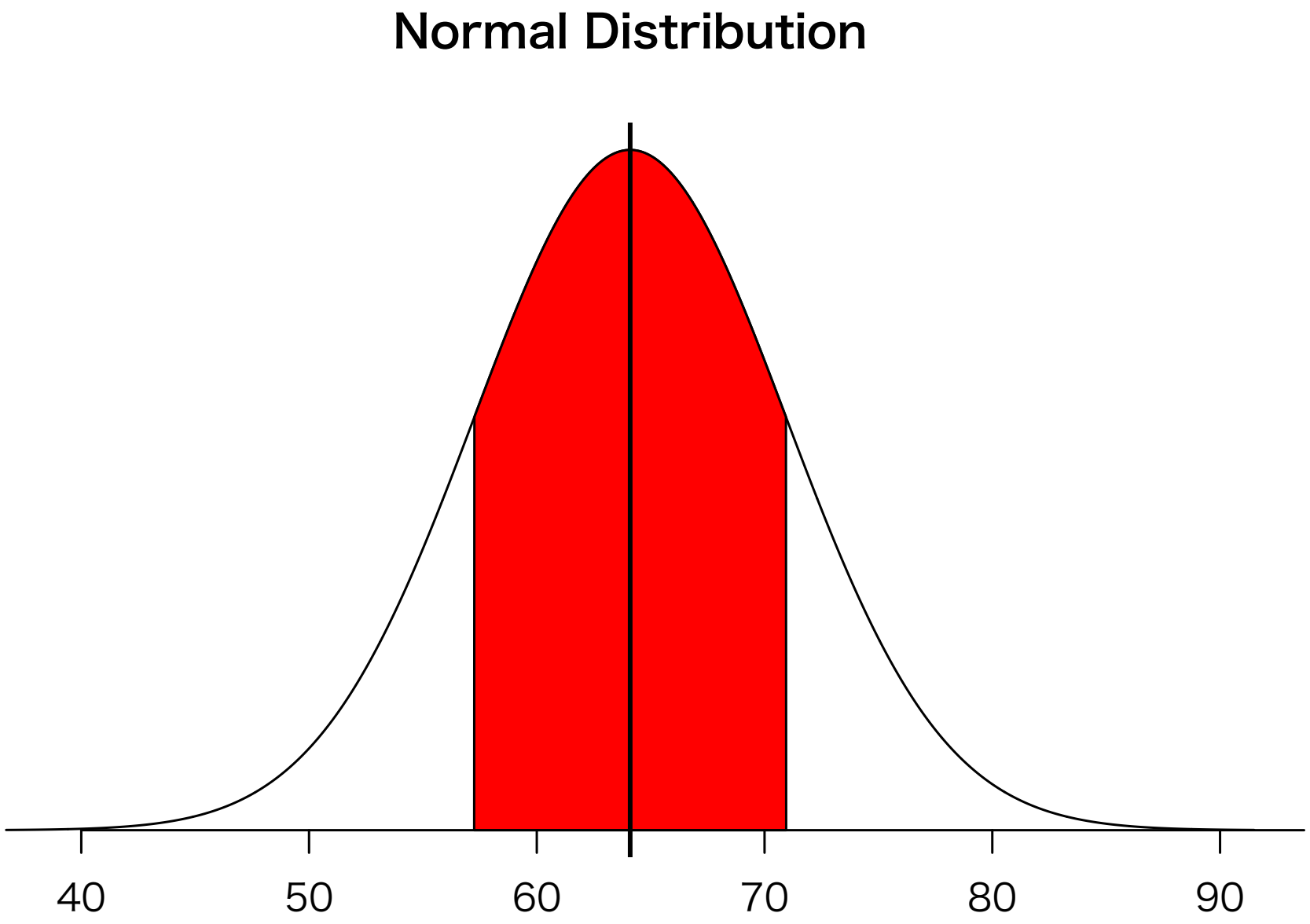
平均2.8点，標準偏差1.2点

どういう分布をイメージしますか？

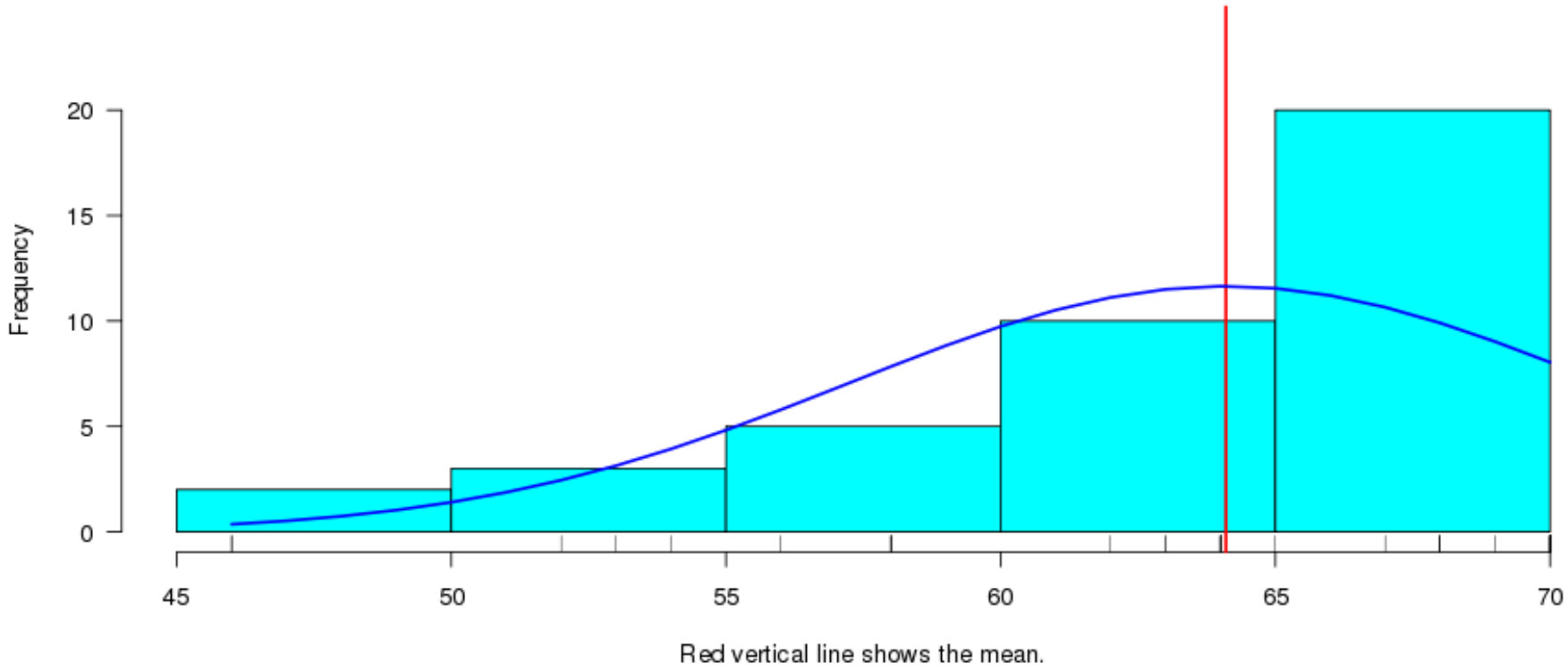
- たとえば，1つのクラス
  - 人数　　40名
  - 平均　　64点
  - 標準偏差　6.9点

「パラメトリック」検定は
この分布を前提にしている。

- **平均　　64点**
- **標準偏差　6.9点**



実際のデータ



「パラメトリック」じゃない！
　＝ノンパラメトリック

- **平均　　64点**
- **標準偏差　6.9点**

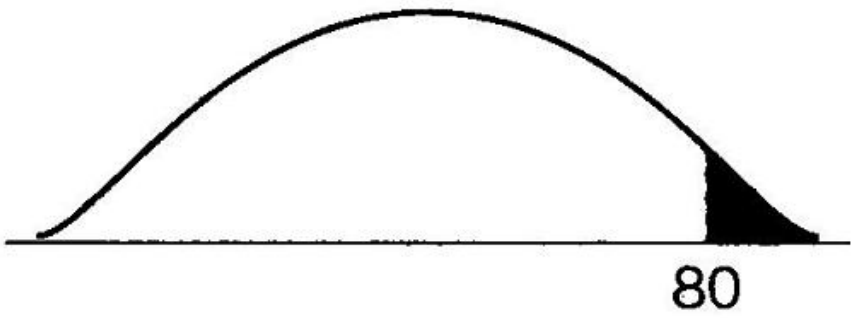
注意点

- 平均と標準偏差で表したデータ
  - 要約としてはわかりやすいが
    すべての情報は保てない
  - 手元のデータの要約というよりは,
    母集団の「きれいな」分布のイメージ

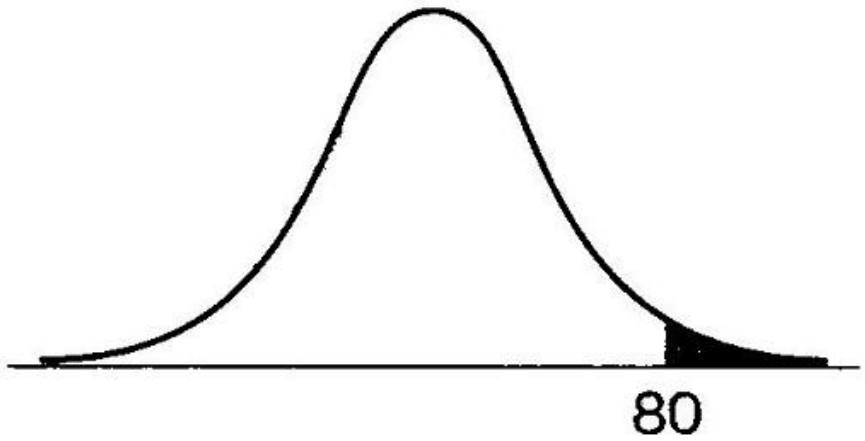
- だから，必ずグラフで確認する！

なぜデータの分布？
なぜ標準偏差が大事？

クラスA



クラスB



統計解析の基礎的知識

1. 記述統計（平均と標準偏差）
2. 統計的検定
3. 効果量

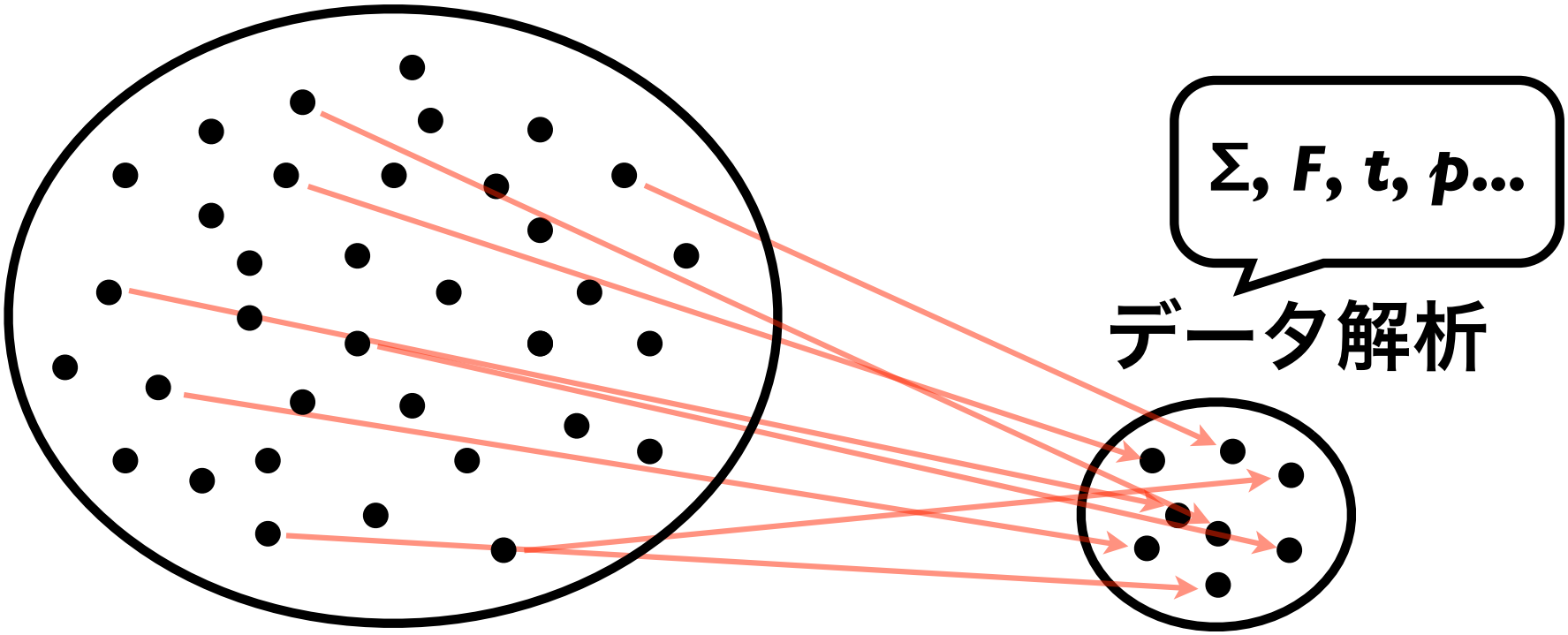
統計解析の基礎的知識

1. 記述統計（平均と標準偏差）

2. 統計的検定

3. 効果量

母集団と標本（推測統計）



母集団

（未知）

推定

・投票所での出口調査
・内閣支持率調査
・TVの視聴率

標　本

（既知）

http://www.urano-ken.com/blog/2013/08/05/let2013-workshop/

*t*検定

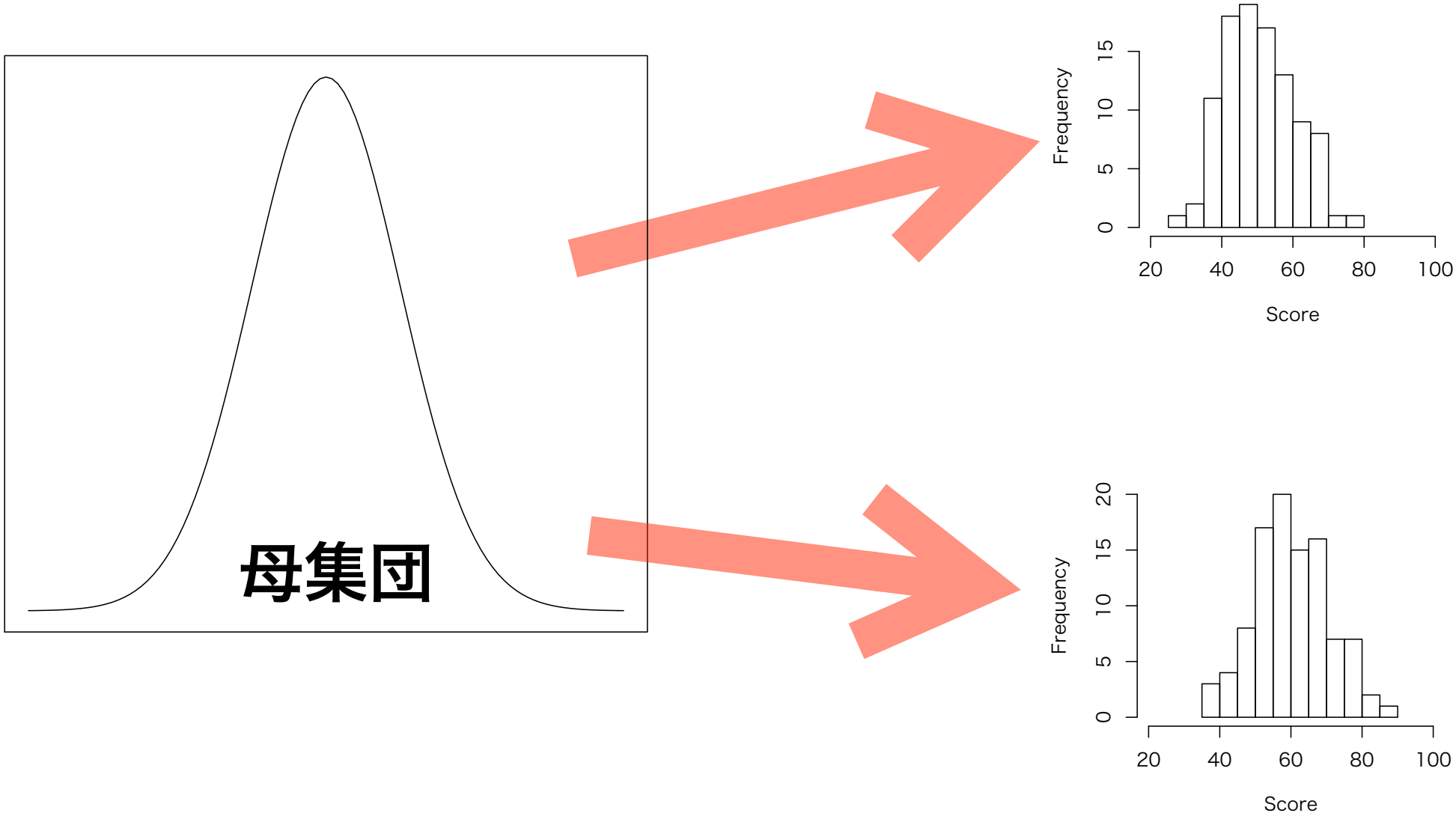
- 特別な指導法を行ったA群と，従来の指導法を行ったB群（もともとの能力は同じとする）が受けた達成度テストの平均点に差があるかを調査したい。
- 平均点，標準偏差（分散）などの記述統計を調べる。
- 点数に差があり，特別な指導法の効果があると言えそう。
- 今回集めたデータだけではなく，母集団を想定して，その平均値にも差があるだろうか？

(t)検定のロジック

- 2つのグループには「差がない」と仮定
- 2つのグループの人数，平均値，標準偏差から t 値を算出
- モデルとして存在する t 分布と標本から得られた t 値を比較し，確率を算出
- 「差がない」確率が $p < .05$ なら有意差あり

2つのグループには「差がない」

= 「同じ母集団から抽出された標本である」



(t)検定のロジック

- 2つのグループには「差がない」と仮定
- 2つのグループの人数，平均値，標準偏差から t 値を算出
- モデルとして存在する t 分布と標本から得られた t 値を比較し，確率を算出
- 「差がない」確率が $p < .05$ なら有意差あり

t 値

等分散の仮定が満たされる場合

＜等分散が仮定できる場合＞

$$t = \frac{\text{群 A の平均値} - \text{群 B の平均値}}{\sqrt{\frac{(\text{群 A 人数} - 1)\text{群 A 分散} + (\text{群 B 人数} - 1)\text{群 B の分散}}{(\text{群 A 人数} + \text{群 B 人数}) - 2}\left(\frac{1}{\text{群 A 人数}} + \frac{1}{\text{群 B 人数}}\right)}}$$

※人数，平均値，分散（標準偏差の2乗）
しか入っていないことに注目！

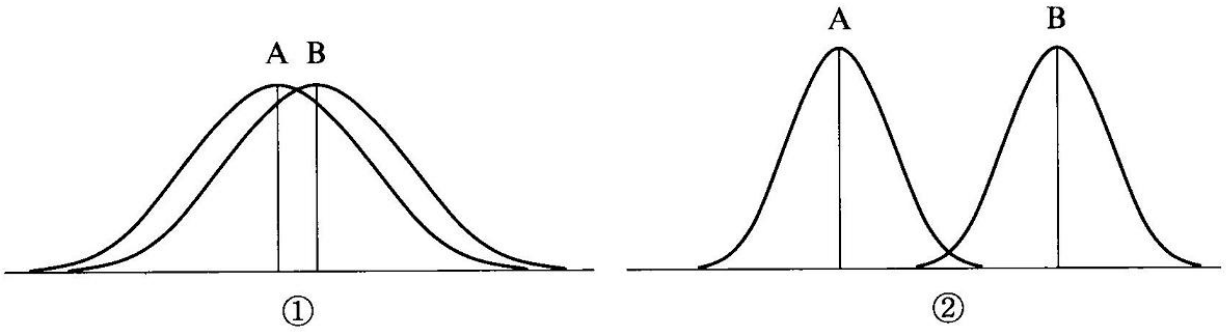
(t)検定のロジック

- 2つのグループには「差がない」と仮定
- 2つのグループの人数，平均値，標準偏差から t 値を算出
- モデルとして存在する t 分布と標本から得られた t 値を比較し，確率を算出
- 「差がない」確率が $p < .05$ なら有意差あり

	男子 (n =100)	女子 (n =100)	平均値の差
平均値	64.46	67.40	-2.93
標準偏差	5.82	3.98	

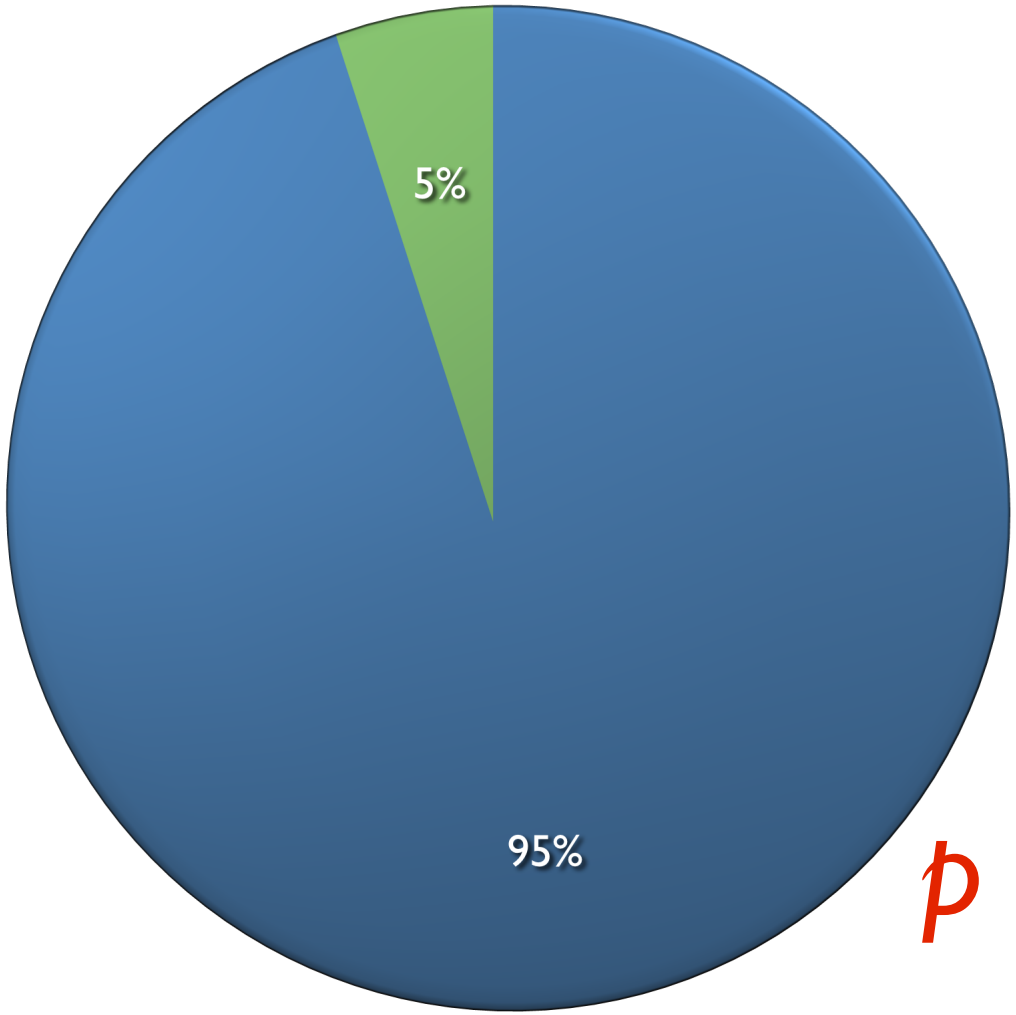
$p < .001$

2群には「差がない」と仮説
↓
この差は統計的に有意な差か？
↓
平均値や標準偏差をもとに
t 値の算出
↓
p 値（確率）が，
有意水準0.05以下か？



p 値は「差がない」確率

p 値は「差がない」確率



$p = .03$（3%）

$p < .05$（0.05以下）

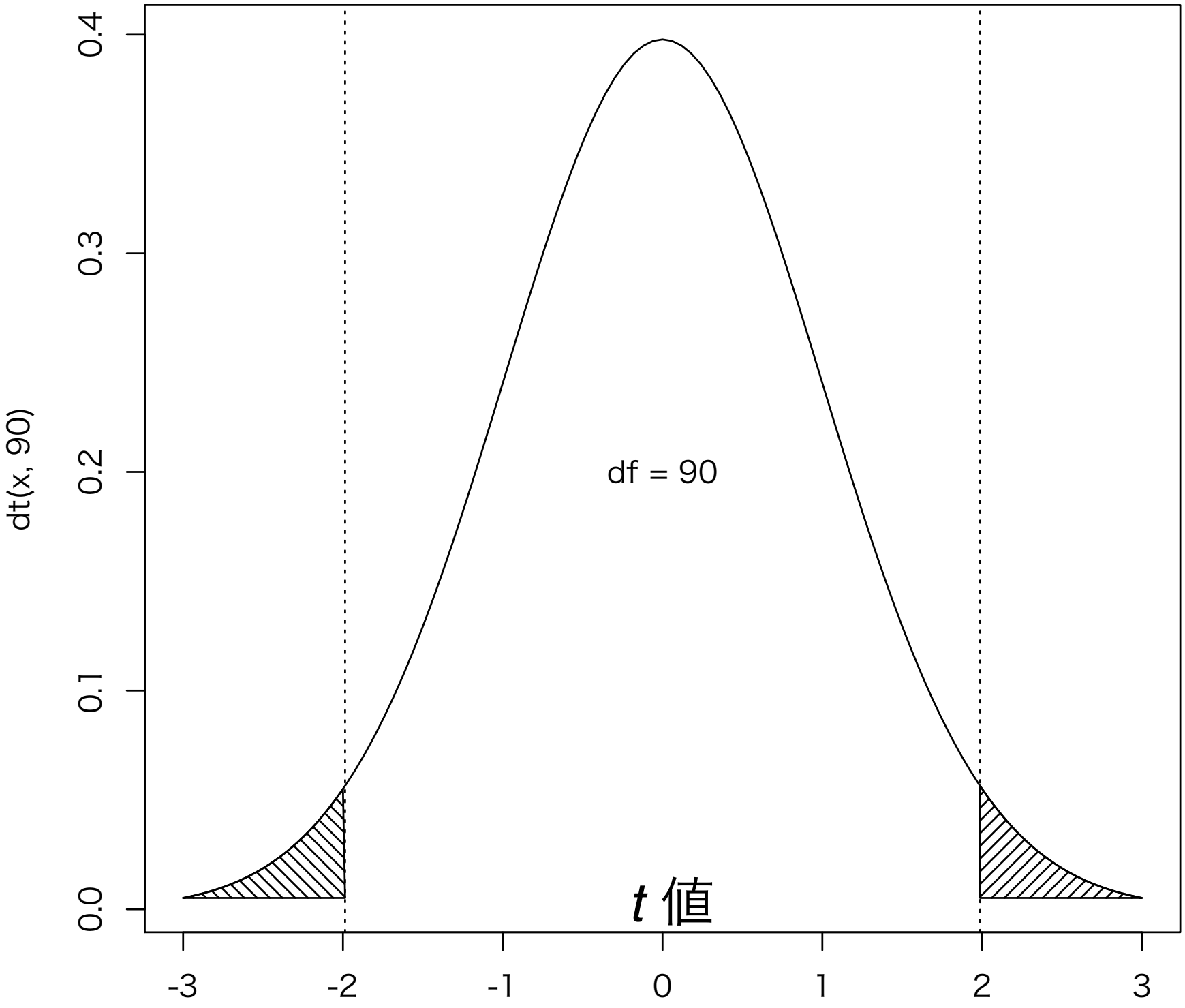
$p < .05$ であれば統計的に有意な差あり

自由度 (degree of freedom: *df*)

- 群1: 女子 44 人
- 群2: 男子 48 人

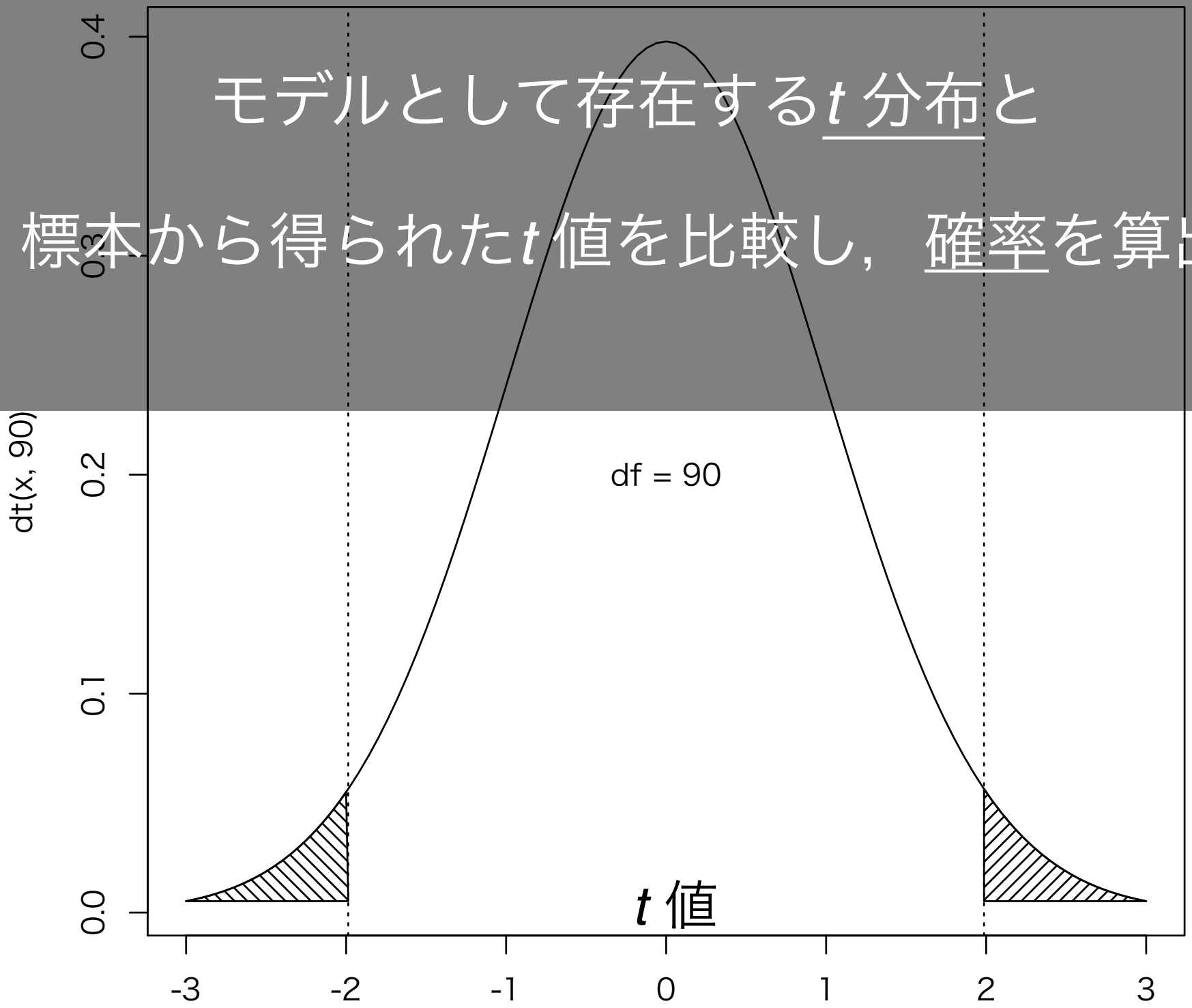
自由度

$$(44-1) + (48-1) = 90$$


0.4
0.3
0.2
0.1
0.0
dt(x, 90)
df = 90
t 値
-3
-2
-1
0
1
2
3

モデルとして存在する*t* 分布と

標本から得られた*t* 値を比較し，確率を算出



結果の見方・報告

$p < .05$（0.05以下）

- $p < .05$ であれば統計的に有意な差あり。
- $p > .05$ であれば統計的に有意な差なし。
- 書き方　$t(90) = 0.09,\ p = .93$

対応（繰り返し）のある *t* 検定

	A	B	E	F
1	学籍番号	氏名	指導前	指導後
2	200703	島田紳助	13	18
3	200704	長澤まさみ	9	16
4	200705	渡辺謙	9	16
5	200706	松嶋菜々子	12	17
6	200707	天海祐希	14	18
7	200708	桑田佳祐	14	18
8	200709	稲垣吾郎	14	18
9	200710	郷ひろみ	11	17
10	200711	水野真紀	10	18
11	200712	宮崎あおい	13	17
12	200713	浜田雅功	13	16
13	200714	ビートたけし	16	17
14	200715	大久保嘉人	9	9
15	200716	森口博子	11	17
16	200717	伊集院光	8	13

同じ学習者が2回
テストを受けて,
その平均点の変化を
比べたいような場合。

ノンパラメトリック検定

- 正規分布を前提としない。
- 正規分布になりにくいケース
  - 間隔尺度ではない（e.g., 4件法アンケート）
  - 人数が少ない，能力・回答が偏っている
- 順位（rank）に変換して検定する
- 「ハンドブック」第8章参照

2グループの点数 => 順位に変換

Wednesday Data

Score	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Group	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E

Sum of Ranks for Alcohol (A) = 59 **Sum of Ranks for Ecstasy (E) = 151**

$$\text{mean rank} = \frac{N(N+1)}{2} = \frac{10 \times 11}{2} = 55$$

We therefore calculate two potential values for W, one for each group:

$$W = \text{sum of ranks} - \text{mean rank}$$
$$W_1 = 59 - 55 = 4$$
$$W_2 = 151 - 55 = 96$$

Field, Miles, & Field (2012, pp. 657-658)

尺度（scale）の種類

1. 名義尺度（nominal scale）

例）男性，女性，血液型，出席番号

2. 順序尺度（ordinal scale）

例）成績（テスト）の順位

ノンパラメトリックなもの

3. 間隔尺度（interval scale）

例）温度，テストの得点

パラメトリックなもの

4. 比率尺度（ratio scale）

例）長さ，重さ

情報量が多くなる

*t*検定をしたいけど
データが正規分布とは言えない場合

- 2つのクラスを比べるような（対応のない）場合
  →マンホイットニーの*U*検定

- 同じテストを2回受けたような（対応のある）場合
  →ウィルコクスンの符号付順位和検定

分散分析

（*ANOVA; analysis of variance)*

- 3つ以上のグループに対して，*t* 検定を繰り返すと第1種の誤りを犯す確率が高くなる。
  → 分散分析

- 平均値間のばらつき（分散）を求めて，
  そのばらつきの大きさが誤差による
  ばらつきよりも大きければ，
  「平均値間に差がある」と推定する。

分散分析で見ているもの

元データ

1

指導法A	指導法B	指導法C	指導法D	
15	13	10	10	
9	8	6	7	
18	8	11	3	
14	12	7	5	
18	7	12	7	全平均
14.8	9.6	9.2	6.4	10

平均行列

2

全平均行列			
10	10	10	10
10	10	10	10
10	10	10	10
10	10	10	10
10	10	10	10

3

群平均行列			
14.8	9.6	9.2	6.4
14.8	9.6	9.2	6.4
14.8	9.6	9.2	6.4
14.8	9.6	9.2	6.4
14.8	9.6	9.2	6.4

偏差（ばらつき）

全体（元全データ-全平均行列）

指導法A	指導法B	指導法C	指導法D
5	3	0	0
−1	−2	−4	−3
8	−2	1	−7
4	2	−3	−5
8	−3	2	−3

群間（3.群平均行列−2.全平均行列）

指導法A	指導法B	指導法C	指導法D
4.8	−0.4	−0.8	−3.6
4.8	−0.4	−0.8	−3.6
4.8	−0.4	−0.8	−3.6
4.8	−0.4	−0.8	−3.6
4.8	−0.4	−0.8	−3.6

効果↑指導法Aを受けることで
全員の点数が4.8点上がった

群内（元データ-3.群平均行列）

指導法A	指導法B	指導法C	指導法D
0.2	3.4	0.8	3.6
−5.8	−1.6	−3.2	0.6
3.2	−1.6	1.8	−3.4
−0.8	2.4	−2.2	−1.4
3.2	−2.6	2.8	0.6

↑群間の効果のあとで、
様々な要因によって生じた変動の大きさ
（群間の効果を取り除いた残り[残差]）

平方（偏差の2乗）

全体^2

指導法A	指導法B	指導法C	指導法D
25	9	0	0
1	4	16	9
64	4	1	49
16	4	9	25
64	9	4	9

（平均に近いほど小さく、離れるほど大きくなる）

群間^2

指導法A	指導法B	指導法C	指導法D
23.04	0.16	0.64	12.96
23.04	0.16	0.64	12.96
23.04	0.16	0.64	12.96
23.04	0.16	0.64	12.96
23.04	0.16	0.64	12.96

（平均に近いほど小さく、離れるほど大きくなる）

群内^2

指導法A	指導法B	指導法C	指導法D
0.04	11.56	0.64	12.96
33.64	2.56	10.24	0.36
10.24	2.56	3.24	11.56
0.64	5.76	4.84	1.96
10.24	6.76	7.84	0.36

（平均に近いほど小さく、離れるほど大きくなる）

平方和

全体平方和
322
データ全体のばらつきの大きさ

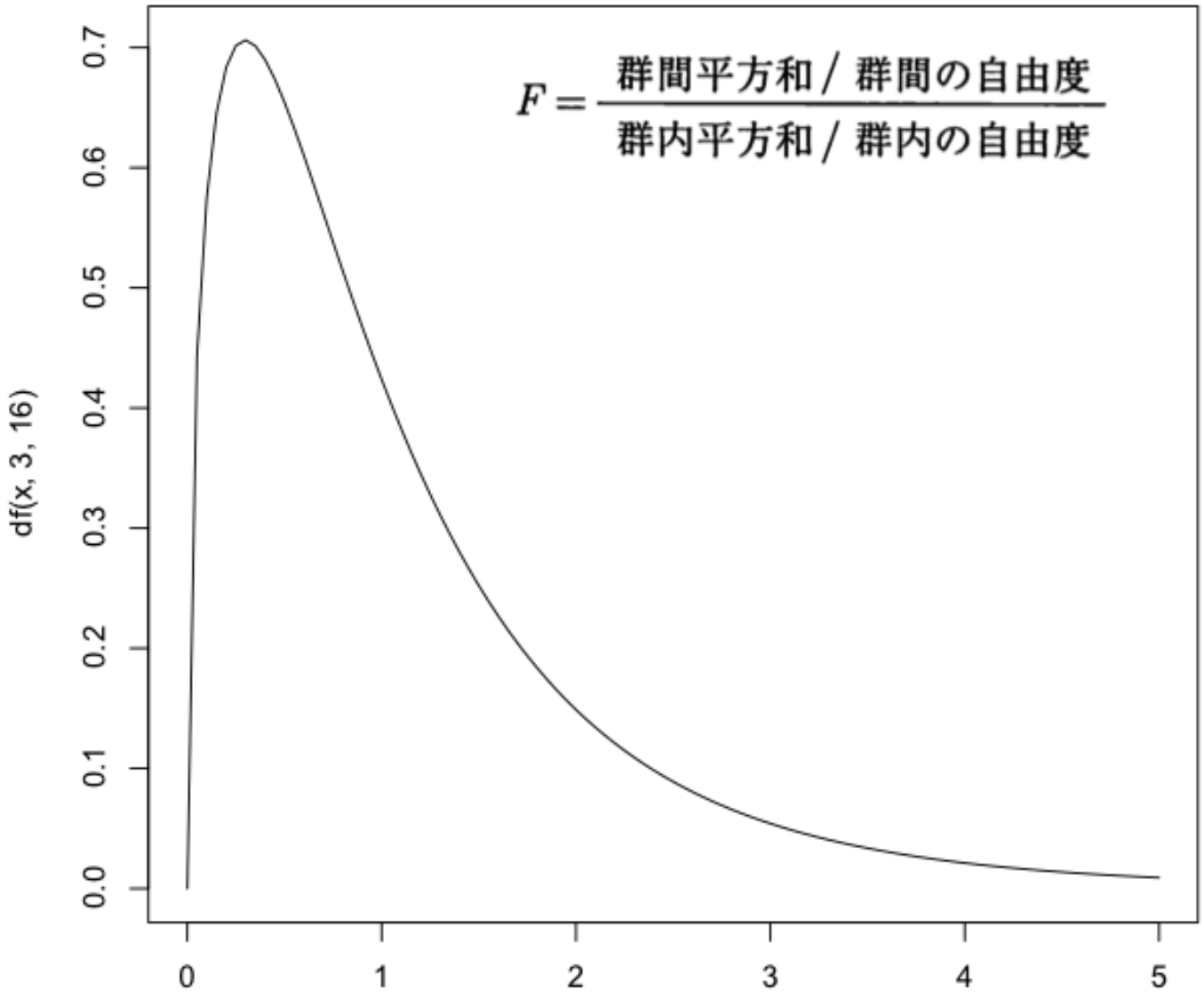
群間平方和
184
群の効果によるばらつき

群内平方和
138
群の効果以外のばらつき（誤差・残差）
residuals

summary

群間平方和
184
+
群内平方和
138
=
全体平方和
322

*F*値


$F = \frac{\text{群間平方和 / 群間の自由度}}{\text{群内平方和 / 群内の自由度}}$
df(x, 3, 16)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
5

対応なし（被験者間要因）

	A	B	C
1	氏名	クラス	VocTotal
2	島田紳助	1	44
3	長澤まさみ	1	32
4	渡辺謙	1	37
5	松嶋菜々子	1	44
6	天海祐希	1	40
7	桑田佳祐	1	44
8	稲垣吾郎	1	45
9	郷ひろみ	1	36
10	水野真紀	1	37
32	ペヨンジュン	2	47
33	石橋貴明	2	39
34	細川茂樹	2	43
35	平原綾香	2	38
36	古谷徹	2	42
37	にしきのあきら	2	35
38	上戸彩	2	29
39	奥菜恵	2	33
40	トータス松本	2	40
41	鈴木紗理奈	2	37
63	加藤晴彦	3	45
64	世良公則	3	46
65	松雪泰子	3	53
66	森高千里	3	47
67	えなりかずき	3	47
68	栗山千明	3	51
69	藤原竜也	3	33
70	石原さとみ	3	40
71	平山あや	3	43
72	安達祐実	3	48

- 3グループ以上の平均点を比べるような場合

対応あり（被験内間要因）

	A	B	C	D
1	氏名	指導前	指導直後	指導3ヶ月後
2	島田紳助	13	18	13
3	長澤まさみ	7	16	9
4	渡辺謙	12	16	9
5	松嶋菜々子	15	17	12
6	天海祐希	8	18	14
7	桑田佳祐	12	18	14
8	稲垣吾郎	13	18	14
9	郷ひろみ	8	17	11
10	水野真紀	9	18	10
11	宮崎あおい	12	17	13

- 同じ学習者が
  テストを3回以上
  受けたような場合

分散分析をしたいけど
データが正規分布とは言えない場合

- 3つ以上のクラスを比べるような（対応のない）場合
  →クラスカル・ウォリスの順位和検定

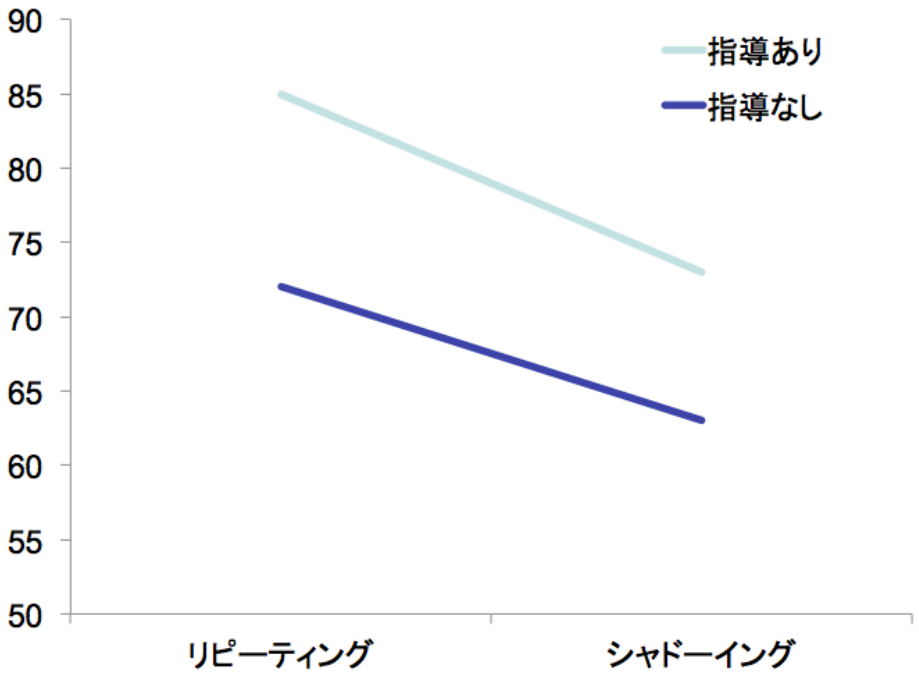
- 同じテストを3回以上受けたような（対応のある）場合
  →フリードマン検定

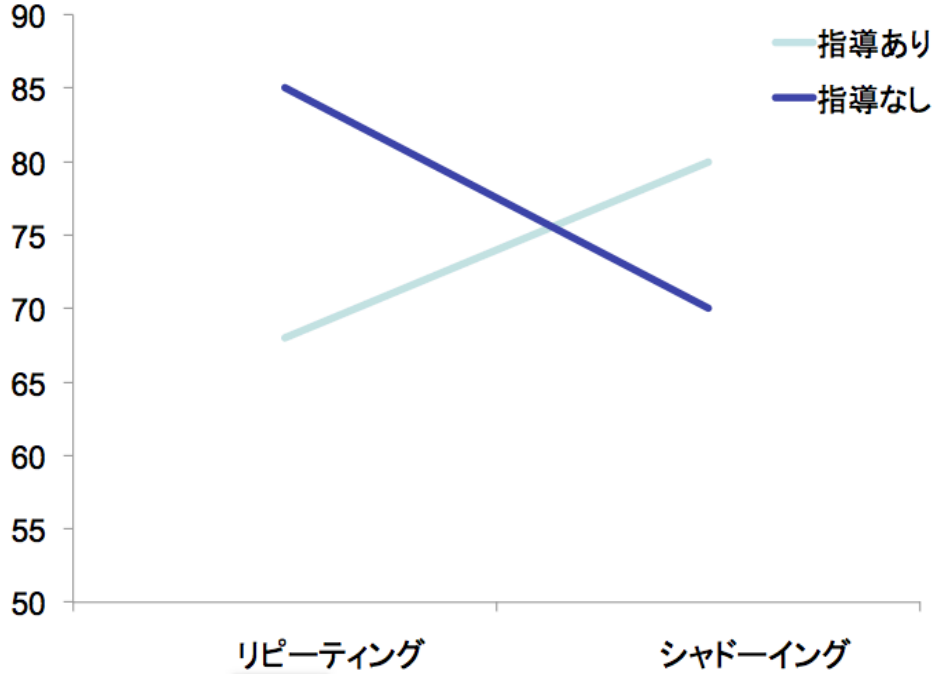
参考：二元配置分散分析

		水準			
要因	指導	あり		なし	
	読み	リピーティング	シャドーイング	リピーティング	シャドーイング
従属変数	得点	85	73	72	63

- 主効果 → 指導の要因，読みの要因
- 交互作用
  → 2つの要因の組み合わせによる影響
- 単純主効果の検定と多重比較

交互作用なし　交互作用あり





参考：カイ二乗検定

度数，比率などの検定。実際に測定される値が期待される値とどれだけ違っているかを調べている。

	オンラインコースを使って自習した	オンラインコースを使って自習しなかった
文系クラス	15人	32人
理系クラス	33人	13人

統計的検定のまとめ

- 手元のデータとモデル（パラメトリック，ノンパラメトリック）とのズレを比較。=> 検定統計量
- 確率的にありえるズレなのか，ありえないズレ（$p < .05$）なのかを見ている。

※「相関分析」とは違います。

統計解析の基礎的知識

1. 記述統計（平均と標準偏差）
2. 統計的検定
3. 効果量

統計解析の基礎的知識

1. 記述統計（平均と標準偏差）

2. 統計的検定

3. 効果量

$p < .05$（0.05以下）

統計的に有意な

statistically significant

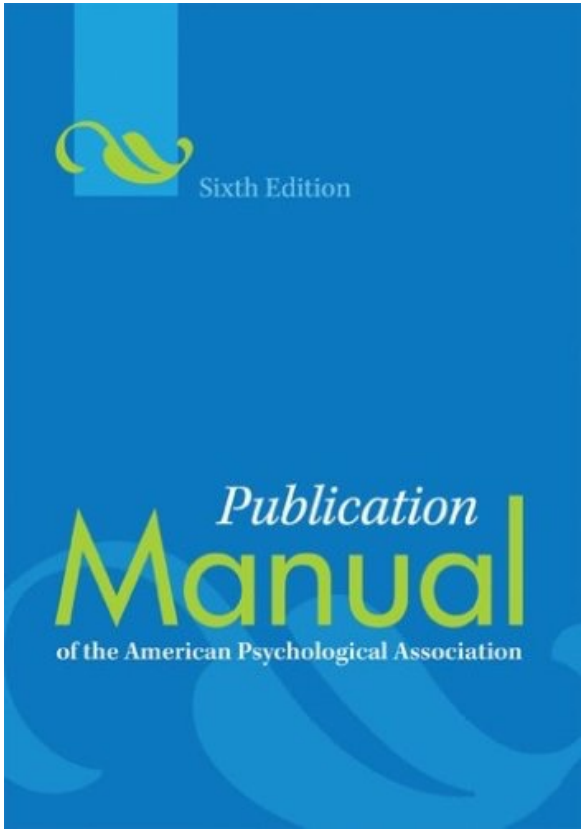
前田・山森（編著）(2004)

「有意差は差の大小については何も語ってくれません。差の大きさを見たいときは，平均値の違いそのものを見ましょう。」（p. 34）

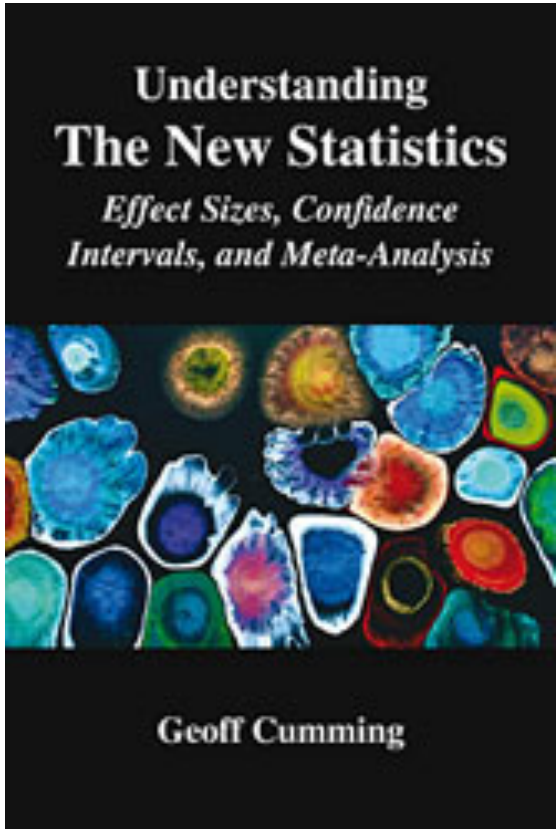
ストップ p 値信仰

「統計改革」

APA 6th (2009)



Cumming (2012)



大久保・岡田 (2009)

比較シミュレーション 1

Group	*M*	*SD*	Diff	*n*	*p* value
Group 1					.374
Group 2					

比較シミュレーション 2

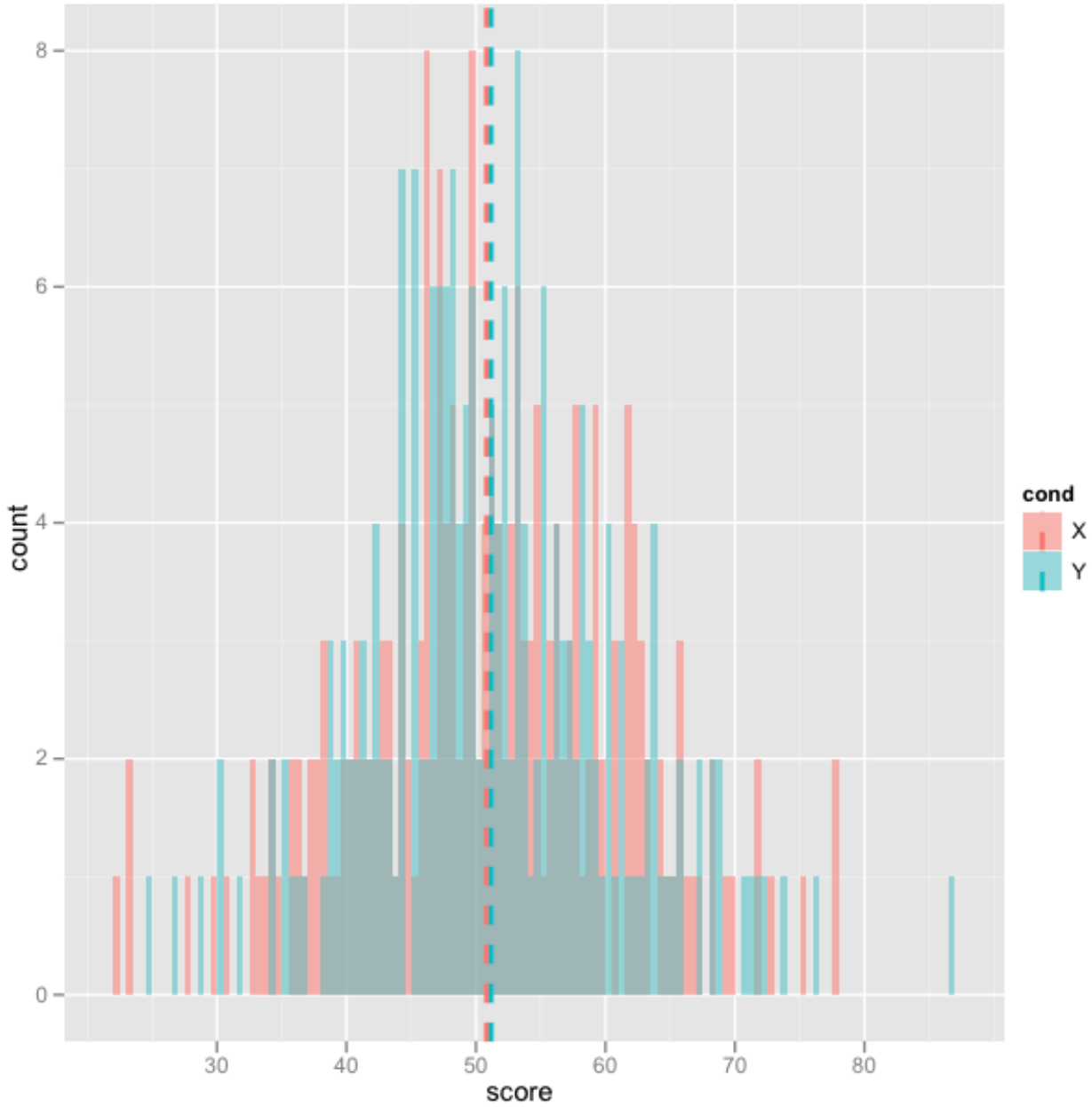
Group	*M*	*SD*	Diff	*n*	*p* value
Group 1					.046
Group 2					

比較シミュレーション 1

Group	*M*	*SD*	Diff	*n*	*p* value
Group 1	50.00	10.00	2.00	20	.374
Group 2	52.00	10.00		20	

比較シミュレーション 2

Group	*M*	*SD*	Diff	*n*	*p* value
Group 1	50.00	10.00	2.00	200	.046
Group 2	52.00	10.00		200	



$N = 400$

$p = .046$

$(p < .05)$

たったの

2点差！

効果量（Effect Size）

- 統計的検定の問題点
  - サンプルサイズが影響。
  - 有意差あり・なしのみの判断。
  - p 値は実質的な差を示さない。

- 効果量（Effect Size）

  - サンプルサイズに影響されない。
  - 効果の大小を示す。
  - 実質的な差を確認できる。

- APA 6th では報告が「不可欠」

各群の人数が等しい場合

$$d = \frac{(\text{実験群の平均} - \text{統制群の平均})}{\sqrt{\dfrac{\text{実験群の標準偏差}^2 + \text{統制群の標準偏差}^2}{2}}}$$

p 値を算出するための検定統計量との関係

検定統計量 = 効果量 (*d*) × サンプルサイズ

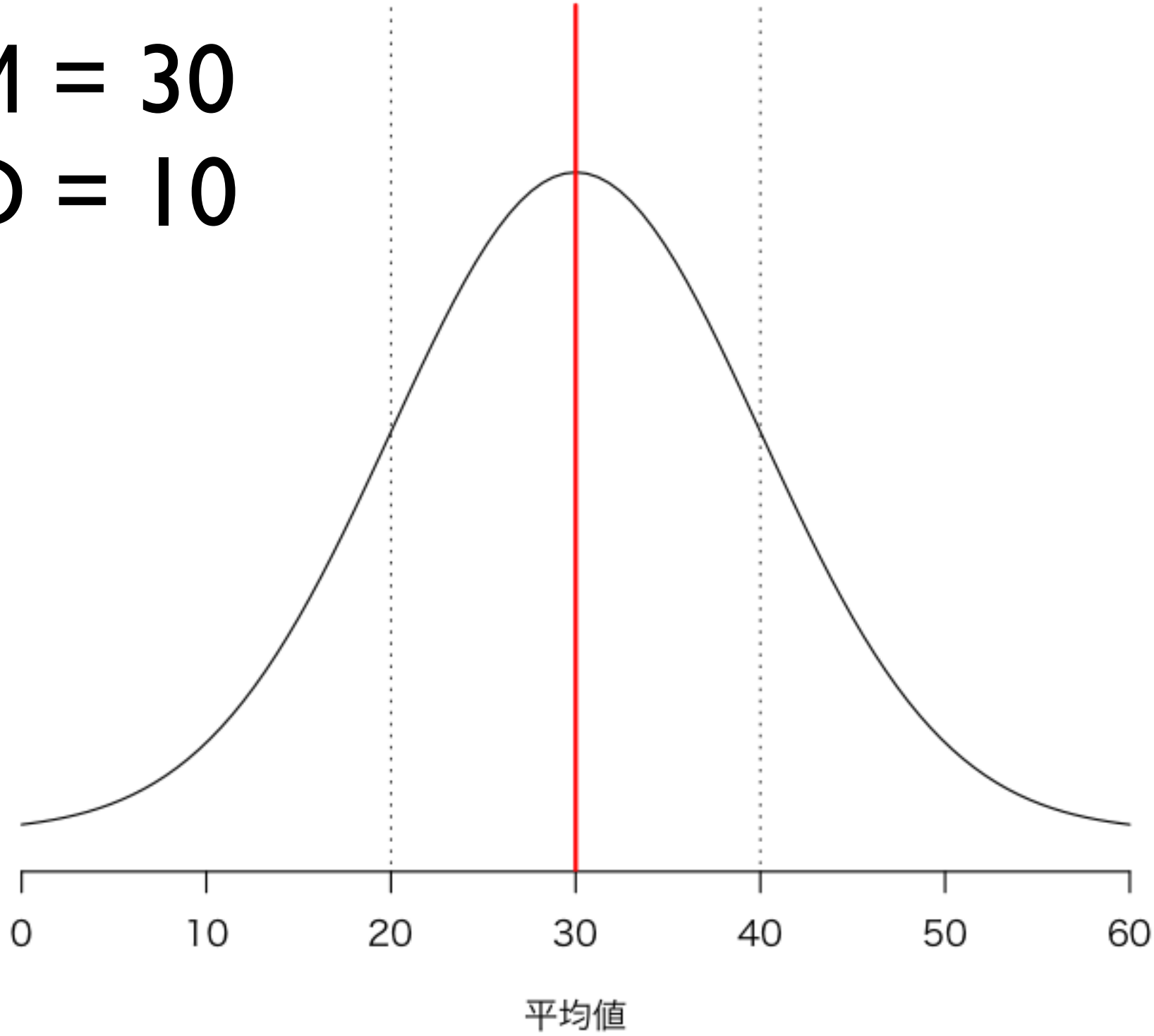
（南風原, 2002, p. 163）

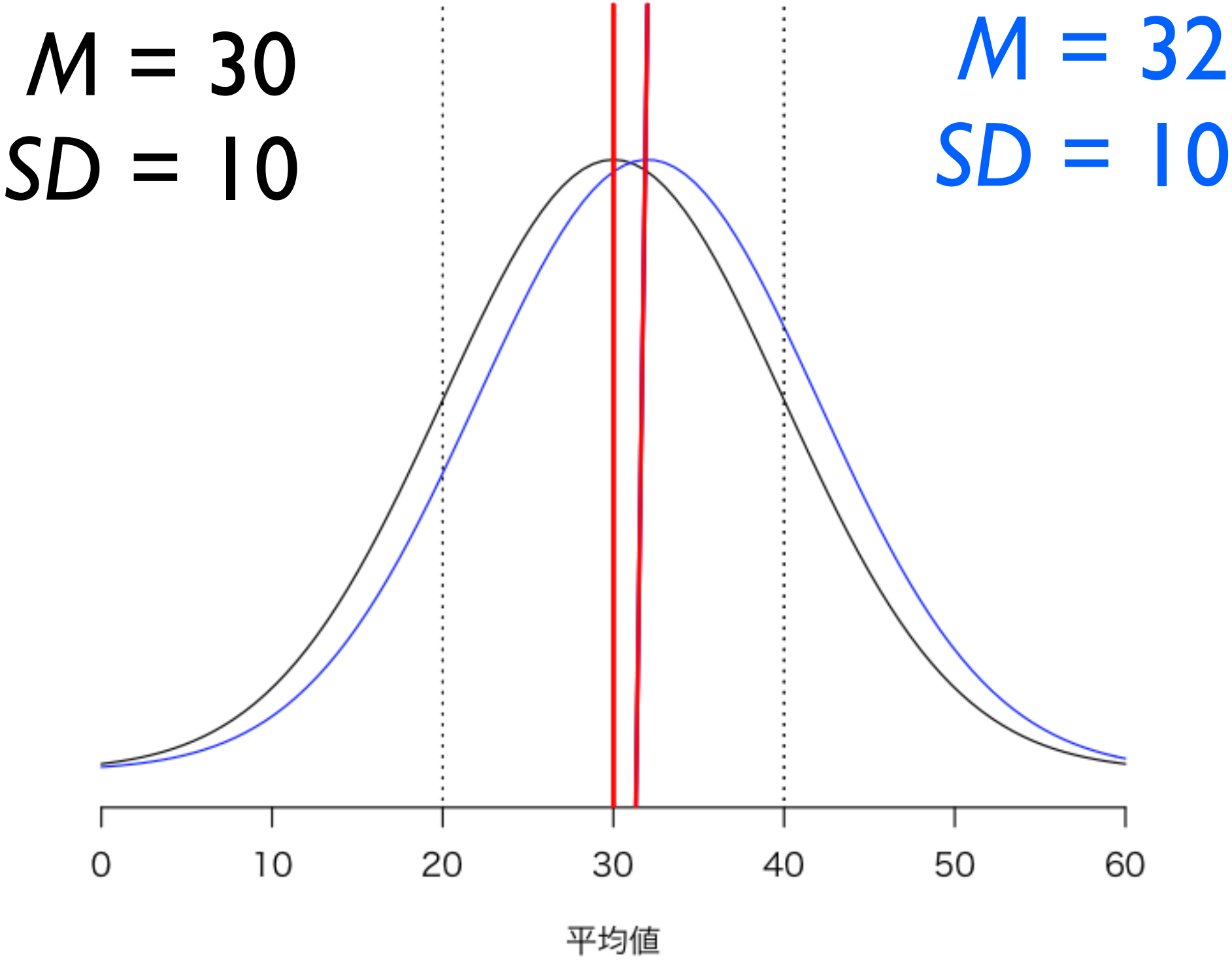
詳しくは浦野(2013)を参照。

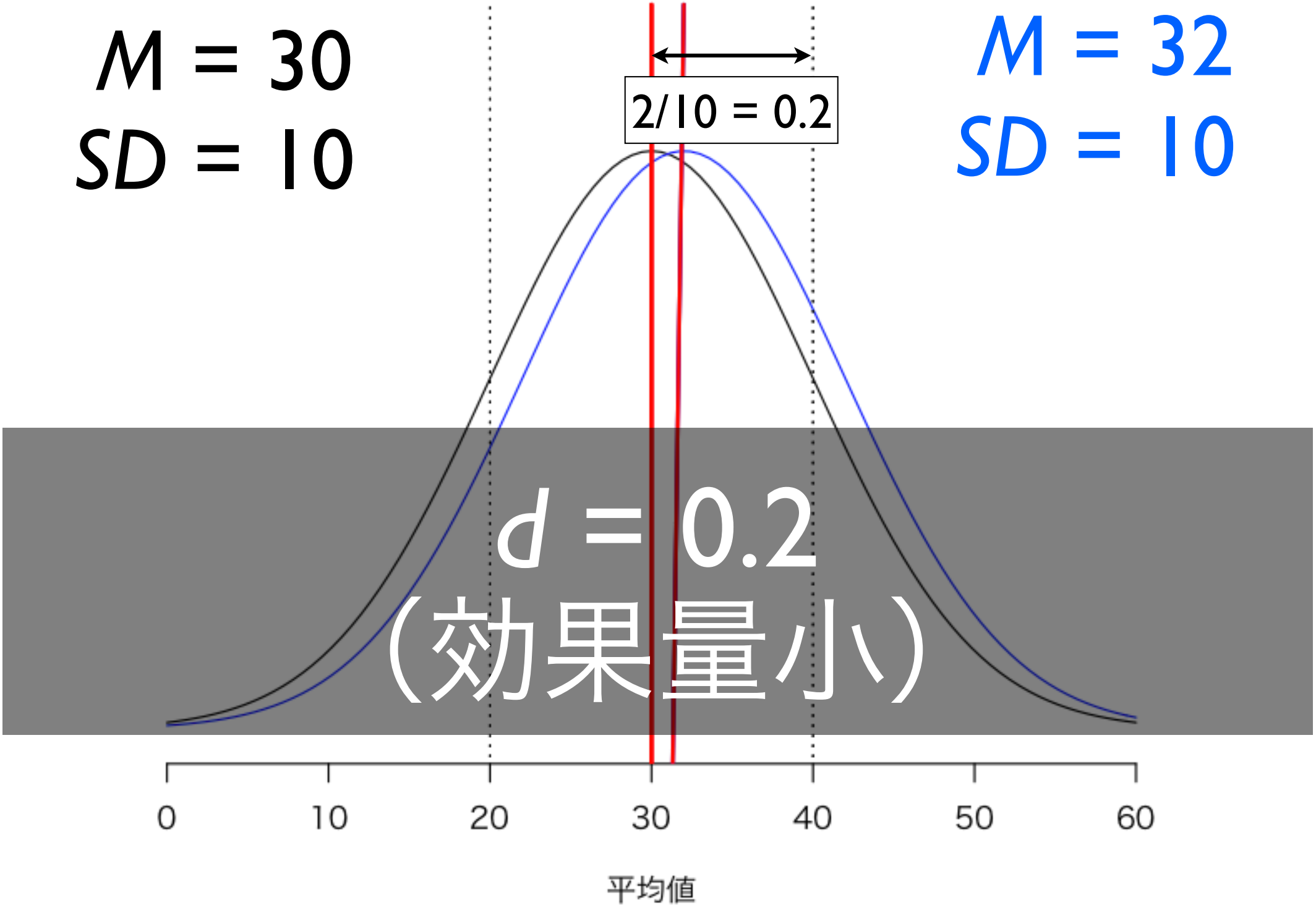
http://www.slideshare.net/uranoken/let2013workshop

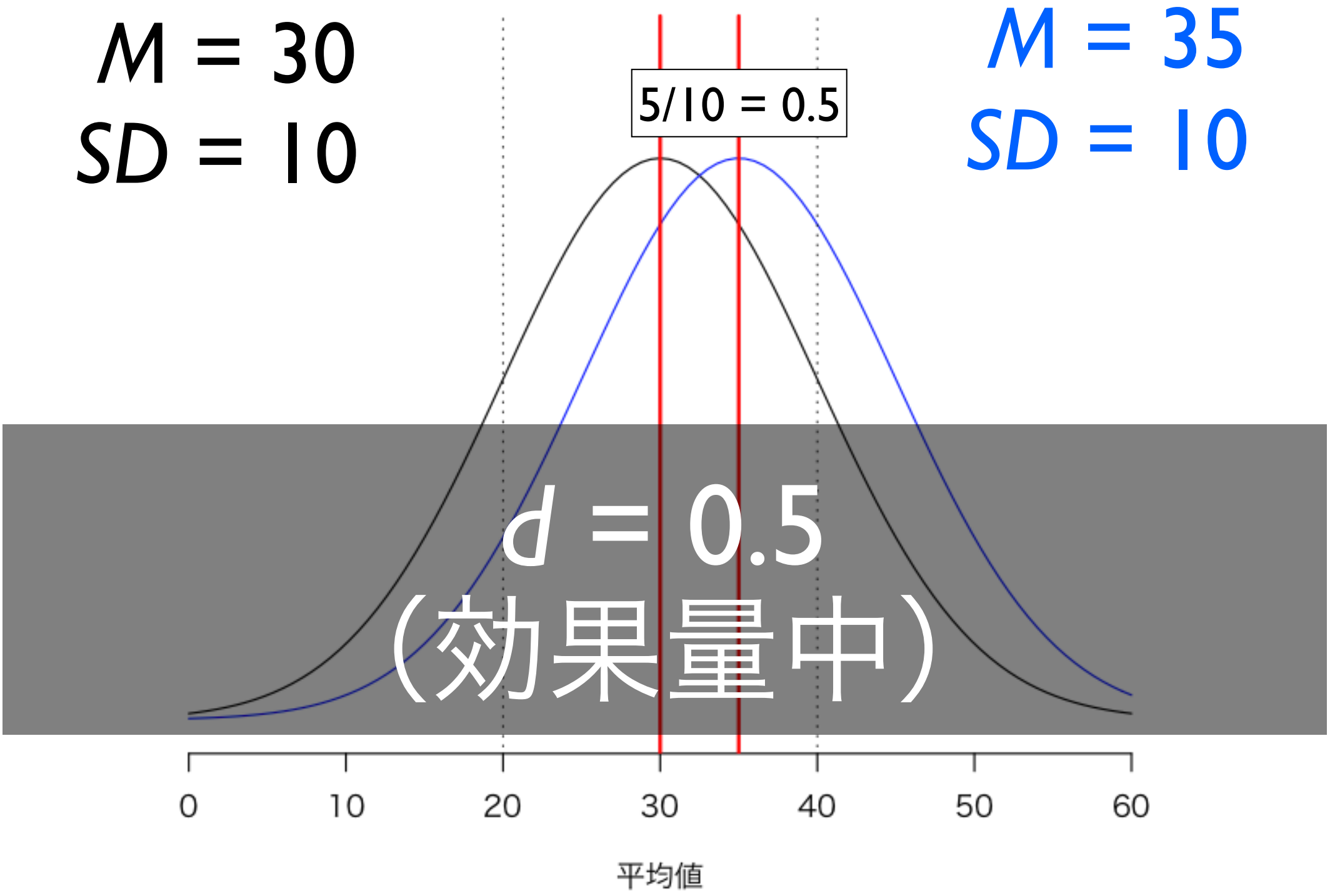
$M = 30$

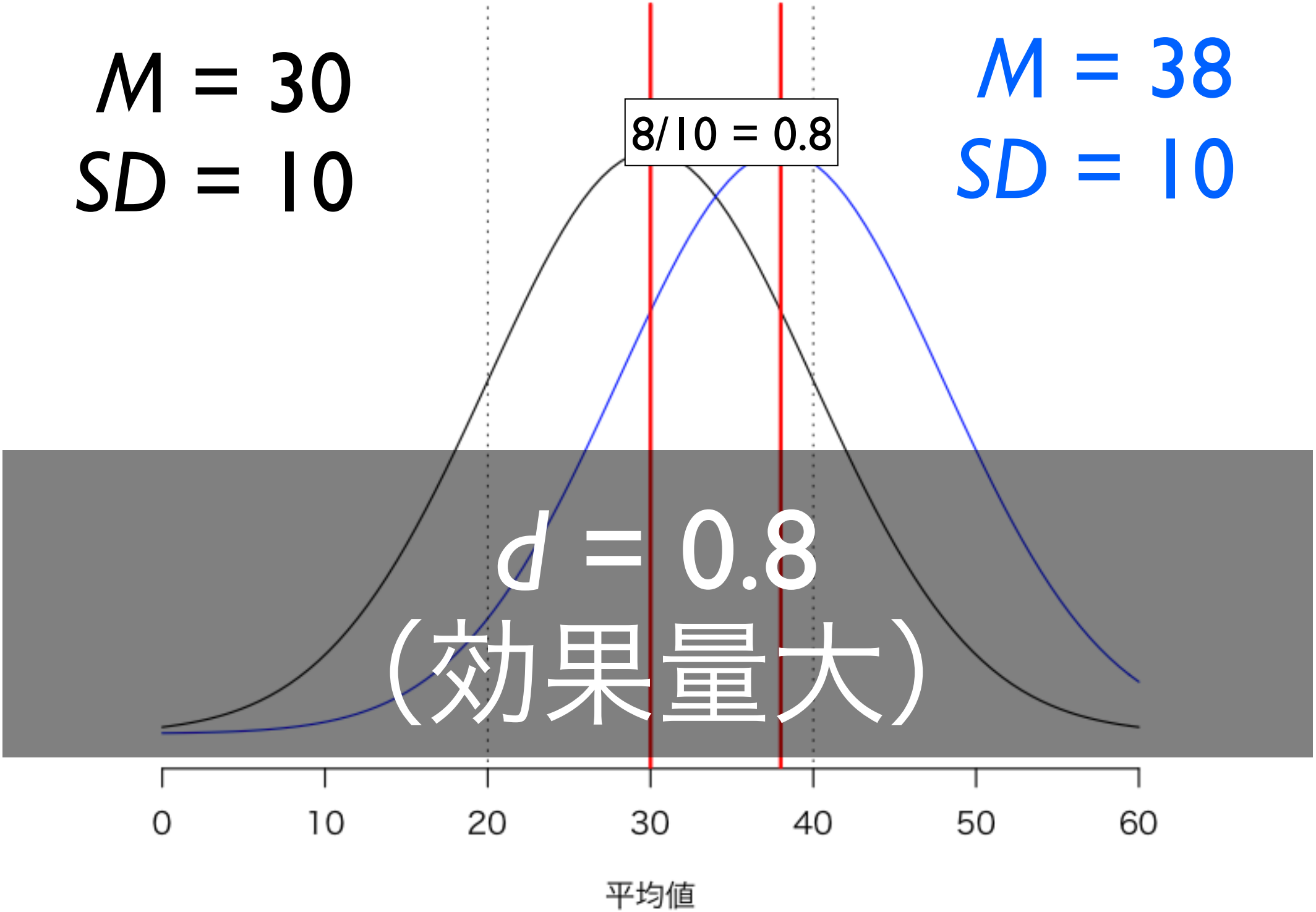
$SD = 10$




M = 30
SD = 10
M = 32
SD = 10
0
10
20
30
40
50
60
平均値


M = 30
SD = 10
2/10 = 0.2
M = 32
SD = 10
d = 0.2
（効果量小）
0
10
20
30
40
50
60
平均値


M = 30
SD = 10
M = 35
SD = 10
5/10 = 0.5
d = 0.5
（効果量中）
0
10
20
30
40
50
60
平均値


M = 30
SD = 10
M = 38
SD = 10
8/10 = 0.8
d = 0.8
(効果量大)
0
10
20
30
40
50
60
平均値

比較シミュレーション 1

効果量 $d = 0.2$

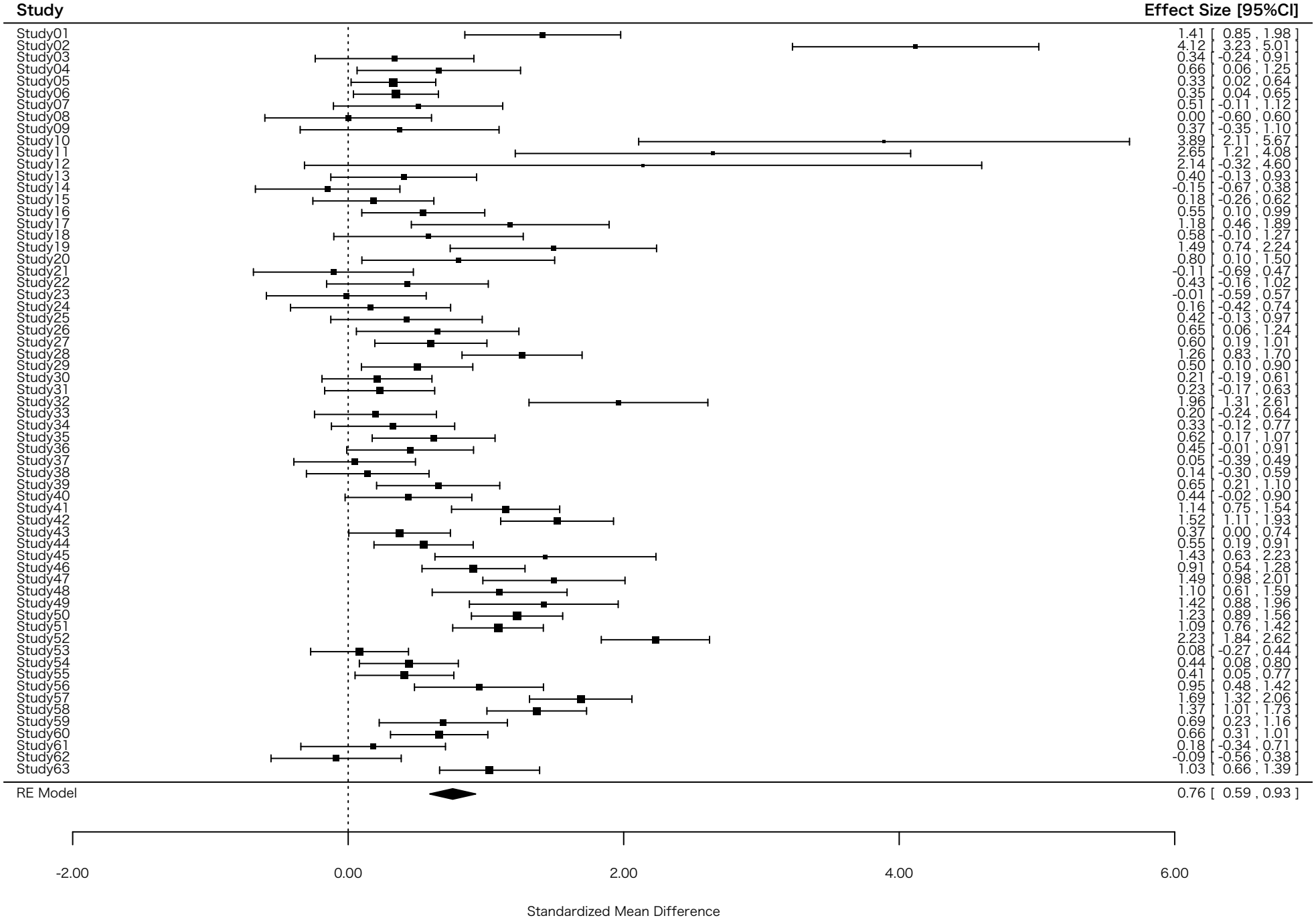
Group	M	SD	Diff	n	p value
Group 1	50.00	10.00	2.00	20	.374
Group 2	52.00	10.00		20	

比較シミュレーション 2

効果量 $d = 0.2$

Group	M	SD	Diff	n	p value
Group 1	50.00	10.00	2.00	200	.046
Group 2	52.00	10.00		200	

効果量の統合＝メタ分析


Study
Effect Size [95%CI]
Study01 1.41 [0.85 , 1.98]
Study02 4.12 [3.23 , 5.01]
Study03 0.34 [-0.24 , 0.91]
Study04 0.66 [0.06 , 1.25]
Study05 0.33 [0.02 , 0.64]
Study06 0.35 [0.04 , 0.65]
Study07 0.51 [-0.11 , 1.12]
Study08 0.00 [-0.60 , 0.60]
Study09 0.37 [-0.35 , 1.10]
Study10 3.89 [2.11 , 5.67]
Study11 2.65 [1.21 , 4.08]
Study12 2.14 [-0.32 , 4.60]
Study13 0.40 [-0.13 , 0.93]
Study14 -0.15 [-0.67 , 0.38]
Study15 0.18 [-0.26 , 0.62]
Study16 0.55 [0.10 , 0.99]
Study17 1.18 [0.46 , 1.89]
Study18 0.58 [-0.10 , 1.27]
Study19 1.49 [0.74 , 2.24]
Study20 0.80 [0.10 , 1.50]
Study21 -0.11 [-0.69 , 0.47]
Study22 0.43 [-0.16 , 1.02]
Study23 -0.01 [-0.59 , 0.57]
Study24 0.16 [-0.42 , 0.74]
Study25 0.42 [-0.13 , 0.97]
Study26 0.65 [0.06 , 1.24]
Study27 0.60 [0.19 , 1.01]
Study28 1.26 [0.83 , 1.70]
Study29 0.50 [0.10 , 0.90]
Study30 0.21 [-0.19 , 0.61]
Study31 0.23 [-0.17 , 0.63]
Study32 1.96 [1.31 , 2.61]
Study33 0.20 [-0.24 , 0.64]
Study34 0.33 [-0.12 , 0.77]
Study35 0.62 [0.17 , 1.07]
Study36 0.45 [-0.01 , 0.91]
Study37 0.05 [-0.39 , 0.49]
Study38 0.14 [-0.30 , 0.59]
Study39 0.65 [0.21 , 1.10]
Study40 0.44 [-0.02 , 0.90]
Study41 1.14 [0.75 , 1.54]
Study42 1.52 [1.11 , 1.93]
Study43 0.37 [0.00 , 0.74]
Study44 0.55 [0.19 , 0.91]
Study45 1.43 [0.63 , 2.23]
Study46 0.91 [0.54 , 1.28]
Study47 1.49 [0.98 , 2.01]
Study48 1.10 [0.61 , 1.59]
Study49 1.42 [0.88 , 1.96]
Study50 1.23 [0.89 , 1.56]
Study51 1.09 [0.76 , 1.42]
Study52 2.23 [1.84 , 2.62]
Study53 0.08 [-0.27 , 0.44]
Study54 0.44 [0.08 , 0.80]
Study55 0.41 [0.05 , 0.77]
Study56 0.95 [0.48 , 1.42]
Study57 1.69 [1.32 , 2.06]
Study58 1.37 [1.01 , 1.73]
Study59 0.69 [0.23 , 1.16]
Study60 0.66 [0.31 , 1.01]
Study61 0.18 [-0.34 , 0.71]
Study62 -0.09 [-0.56 , 0.38]
Study63 1.03 [0.66 , 1.39]
RE Model 0.76 [0.59 , 0.93]
-2.00
0.00
2.00
4.00
6.00
Standardized Mean Difference

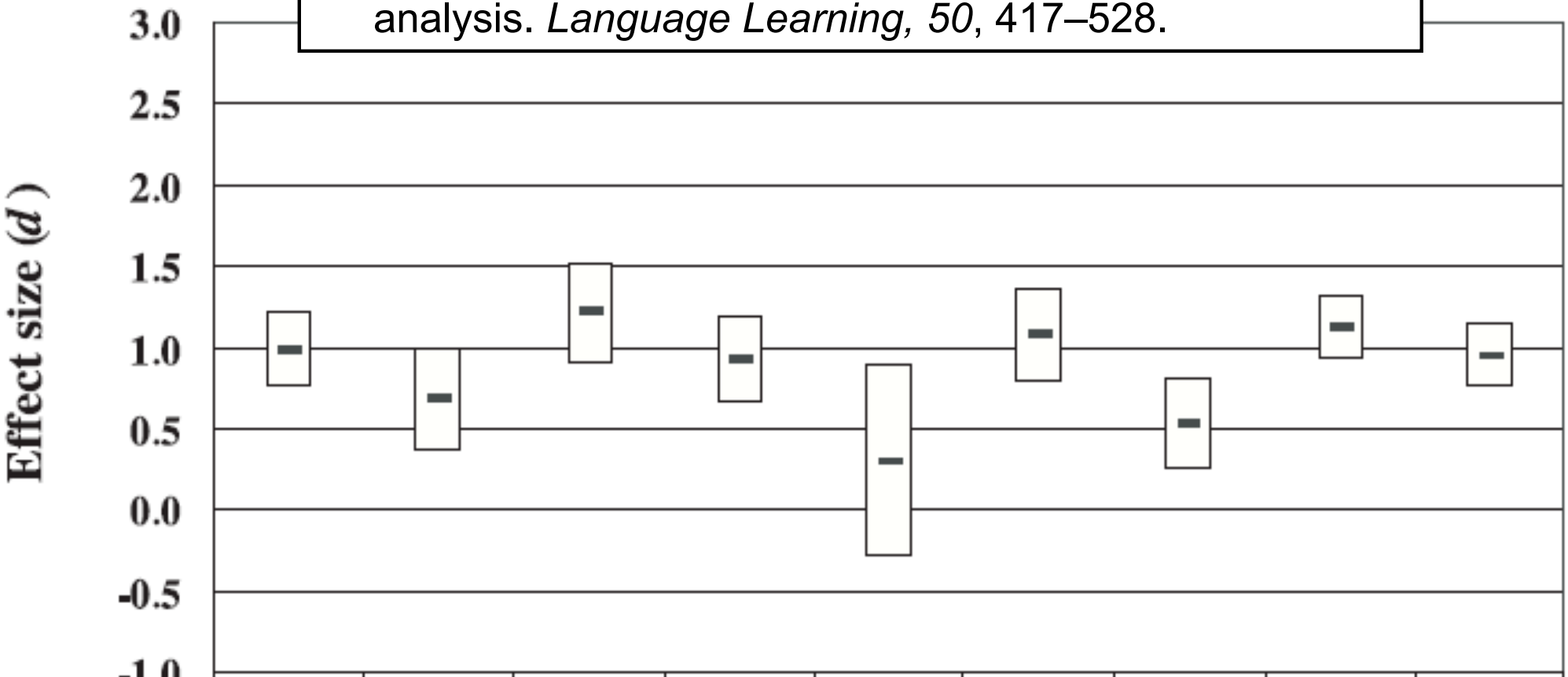
Plonsky and Oswald(2014)

"L2 field-specific benchmarks"

- d = 0.40, small effect
- d = 0.70, medium effect
- d = 1.00, large effect

*For pre-post and within-group contrasts:
d = 0.60 (small), 1.00 (medium), 1.40 (large)

Norris, M. J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.



	FonF (k=43)	FonF implicit (k=18)	FonF explicit (k=25)	FonFS (k=55)	FonFS implicit (k=11)	FonFS explicit (k=44)	All implicit (k=29)	All explicit (k=69)	Overall (k=98)
CI upper (95%)	1.22	1.00	1.53	1.19	0.89	1.36	0.82	1.33	1.14
Mean	1.00	0.69	1.22	0.93	0.31	1.08	0.54	1.13	0.96
CI lower (95%)	0.78	0.38	0.91	0.67	-0.27	0.80	0.26	0.93	0.78

L2 分野でのメタ分析

- Task-based instruction

  $d = 0.92$ (Keck et al., 2006)

- Strategy Instruction

  $d = 0.49$ (Plonsky, 2008)

- CALL over non-CALL

  $d = 0.35$ (Grgurović et al., 2013)

langtest.jp

Web App

- Cronbach's Coefficient Alpha (mirror)
- Basic Statistics Calculator (mirror)
- Learning by Doing Stats (t-test Tutorial) (mirror)
- Comparing Two Independent Samples (mirror)
- Comparing Paired Samples (mirror)
- Effect Size Calculator 1 (Means, SDs, Ns) (mirror)
- ANOVA (mirror)

http://langtest.jp/

効果量と信頼区間も
計算してくれる

効果量計算シートは
信頼区間の算出なし

d の計算

1. 各群の人数が同じ場合

↓ここにそれぞれの値を入れてください ↓自由度と *t* 値は *d* の計算のために記入する必要ありません

	平均 (Mean)	標準偏差 (SD)	自由度	*t* 値
実験群/Test A	59.75	8.91		
統制群/Test B	50.5	11.38		

	↓これを報告する	効果量の目安
効果量(*d*)	0.91	効果量大

$$d = \frac{(\text{実験群の平均} - \text{統制群の平均})}{\sqrt{\frac{\text{実験群の標準偏差}^2 + \text{統制群の標準偏差}^2}{2}}}$$

http://www.mizumot.com/stats/effectsize.xls

http://www.mizumot.com/stats/effectsize.xls

ノンパラメトリック検定の効果量の計算

群の数	用いる代表値	データの対応	検定方法
2群	中央値・順位など	なし	マン・ホイットニーの*U*検定
		あり	ウィルコクスンの符号付順位和検定
3群以上		なし	クラスカル・ウォリスの順位和検定
		あり	フリードマン検定

出所:『英語教師のための教育データ分析入門』p.61

r の計算

(*Z* と被験者数から計算)

↓ここにそれぞれの値を入れてください

Z	2.20
人数	80

	↓これを報告する	効果量の目安
効果量(*r*)	.25	効果量小

【例】

SPSSのアウトプット

	VAR00001
Mann-Whitney の U	4.000
Wilcoxon の W	59.000
Z	-3.484
漸近有意確率 (両側)	.000
正確有意確率 [2x(片側有意確率)]	.000

統計的検定は必要ない？

- Yes and no.
  - サンプリングの影響
  - *p* 値は再現性がない（unreliable）
  - 効果量を言い訳に使ってない？

- 効果量とその信頼区間を併せて報告する。

前田・山森（編著）(2004)

「統計処理を行うということは，ある実体を見いだすことではないのです。あくまでも，研究者が取ったデータに対して，確率的にどうなのか，もしくはデータの傾向くらいしか言えないのです。結果の記述の際には，データと分析結果に対して忠実に記述する慎重さが求められます。」（p. 173）

→ 効果量，信頼区間で情報付加

The Basic and Applied Social Psychology

http://www.tandfonline.com/doi/abs/10.1080/01973533.2015.1012991#.Vb3tuJPtlBd

BASIC AND APPLIED SOCIAL PSYCHOLOGY, 37:1–2, 2015
Copyright © Taylor & Francis Group, LLC
ISSN: 0197-3533 print/1532-4834 online
DOI: 10.1080/01973533.2015.1012991

Routledge
Taylor & Francis Group

*p*値（帰無仮説検定）禁止！

Editorial

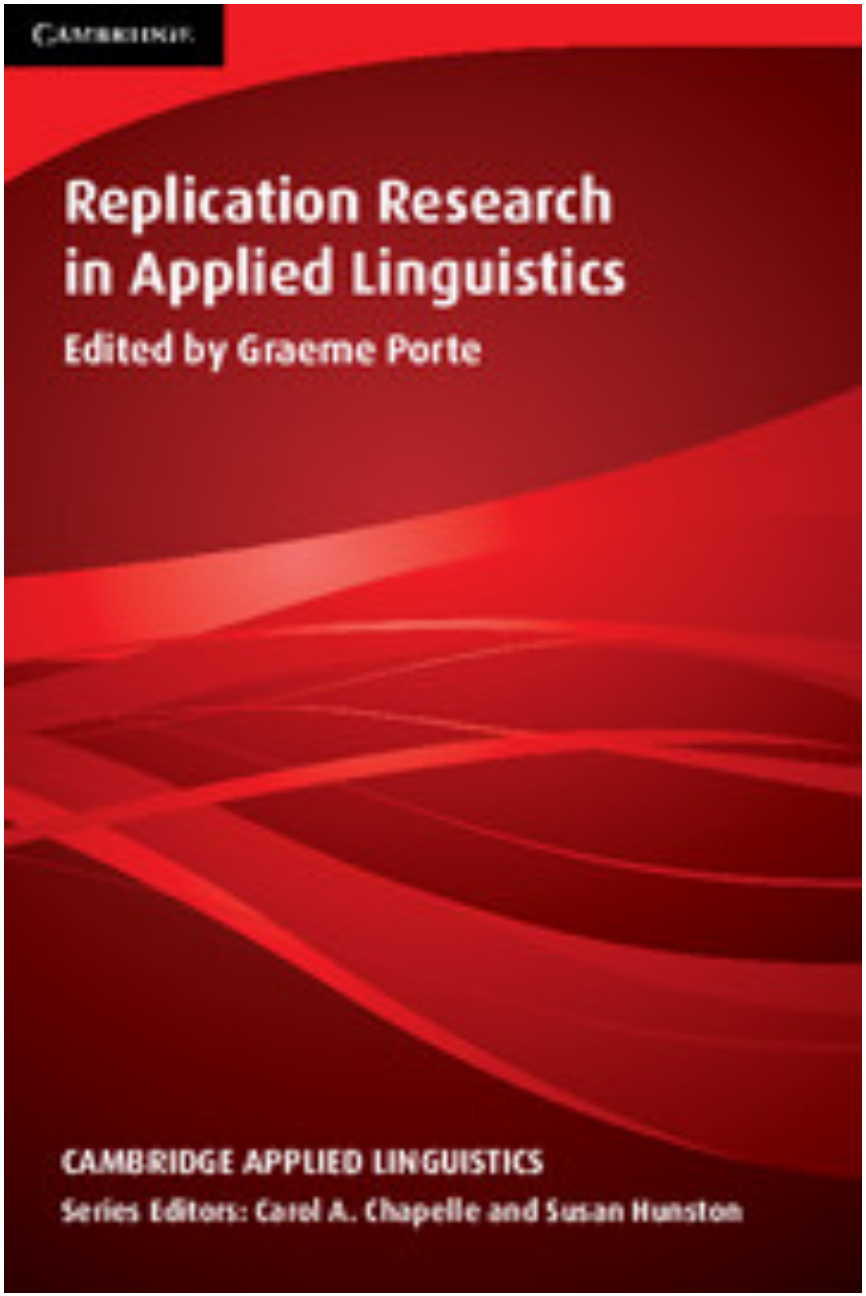
David Trafimow and Michael Marks
New Mexico State University

The *Basic and Applied Social Psychology* (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP.

a strong case for rejecting it, confidence intervals do not provide a strong case for concluding that the population parameter of interest is likely to be within the stated interval. Therefore, confidence intervals also are banned from BASP.

Bayesian procedures are more interesting. The usual problem with Bayesian procedures is that they depend on some sort of Laplacian assumption to generate numbers where none exist. The Laplacian assumption is that

Porte (2012)


CAMBRIDGE
Replication Research in Applied Linguistics
Edited by Graeme Porte
CAMBRIDGE APPLIED LINGUISTICS
Series Editors: Carol A. Chapelle and Susan Hunston

1回の研究で断言できることはほぼない！

- p 値は再現性を考えるとひどい指標。
- 実質的な差や効果は「効果量」。
- 先行研究からどの程度の効果があるか推測する。（もしくはメタ分析を行う）

Replication や
メタ分析に
必要な情報を書く

前田・山森（編著）(2004)

「必要な情報はきちんと書く。
情報は追試できるように書く。
読者にわかりやすく書く。」
(p. 172)

déjà-vu

「結果は図3に示した。プリテストとポストテストの間で8.2点の得点上昇があり，統計的にも有意な上昇であった（$t(11) = 9.108, p < .01$, 両側検定）。」

「ダメ。ゼッタイ。」

- 平均・標準偏差の記載なし。
- 人数・総数が不明。
- 信頼性係数などの報告なし。
- *p* 値のみの報告。（* がたくさん。）

（分析の）追試に必要な情報

- サンプルサイズ，平均，標準偏差
- 相関係数（対応ありデータ，SEMなど）
- 信頼性係数（平均への回帰，相関の希薄化の修正など）

具体例

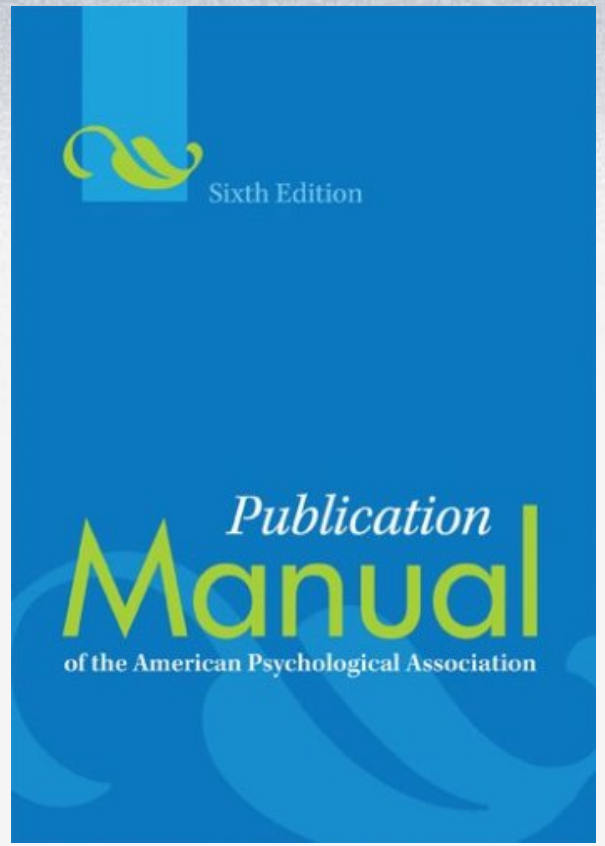
ある国際誌掲載論文の一部
何かがおかしい気がする...

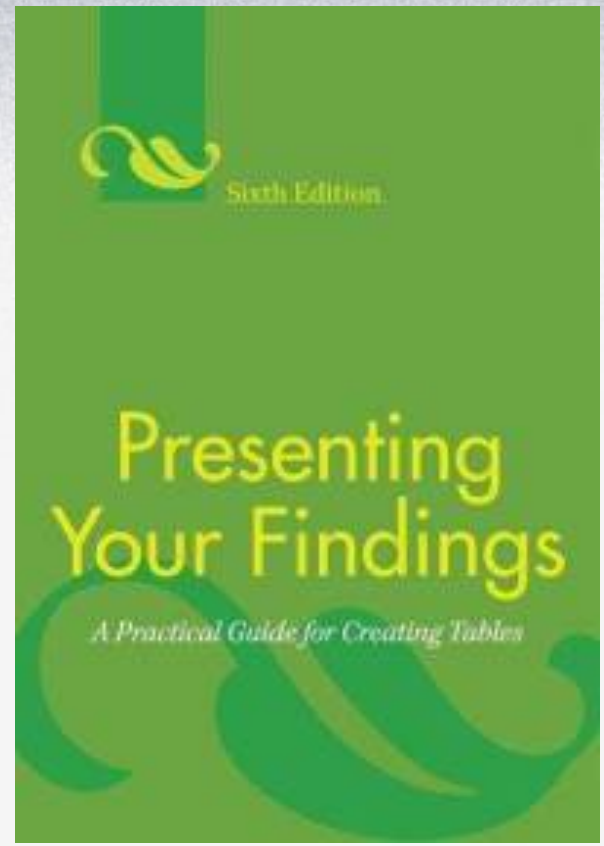
	N	Term exam overall mean	SD	Independent t-test $p < 0.005$
Experimental group	11	**78.91**	8.42	$P = 0.000$
Control group	11	**76.82**	7.41	

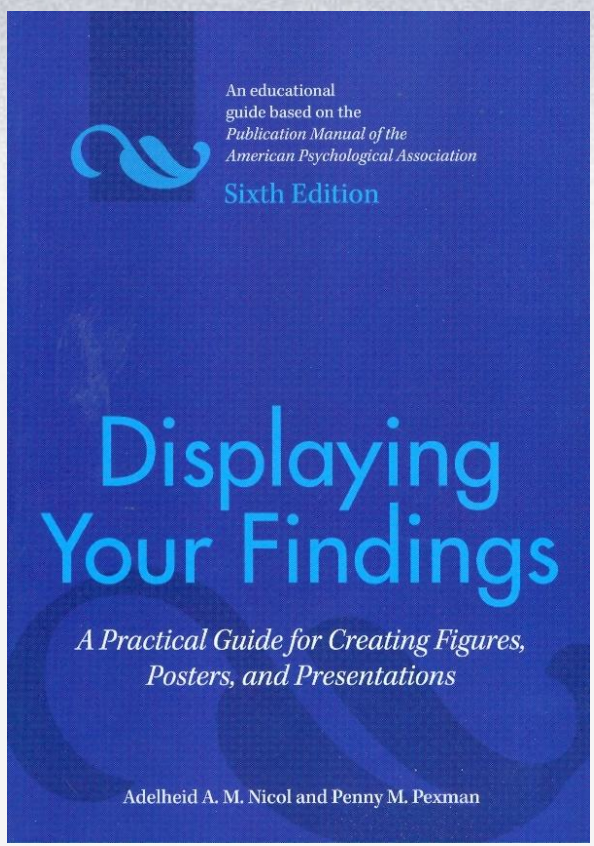
青木繁伸先生作成のRコードで追試（http://aoki2.si.gunma-u.ac.jp/R/my-t-test.html）

ウエルチの方法による，二群の平均値の差の検定

n1 = 11, mean1 = 78.91, variance1 = 70.8964
n2 = 11, mean2 = 76.82, variance2 = 54.9081
t = 0.618, df = 19.682, p-value = 0.5437



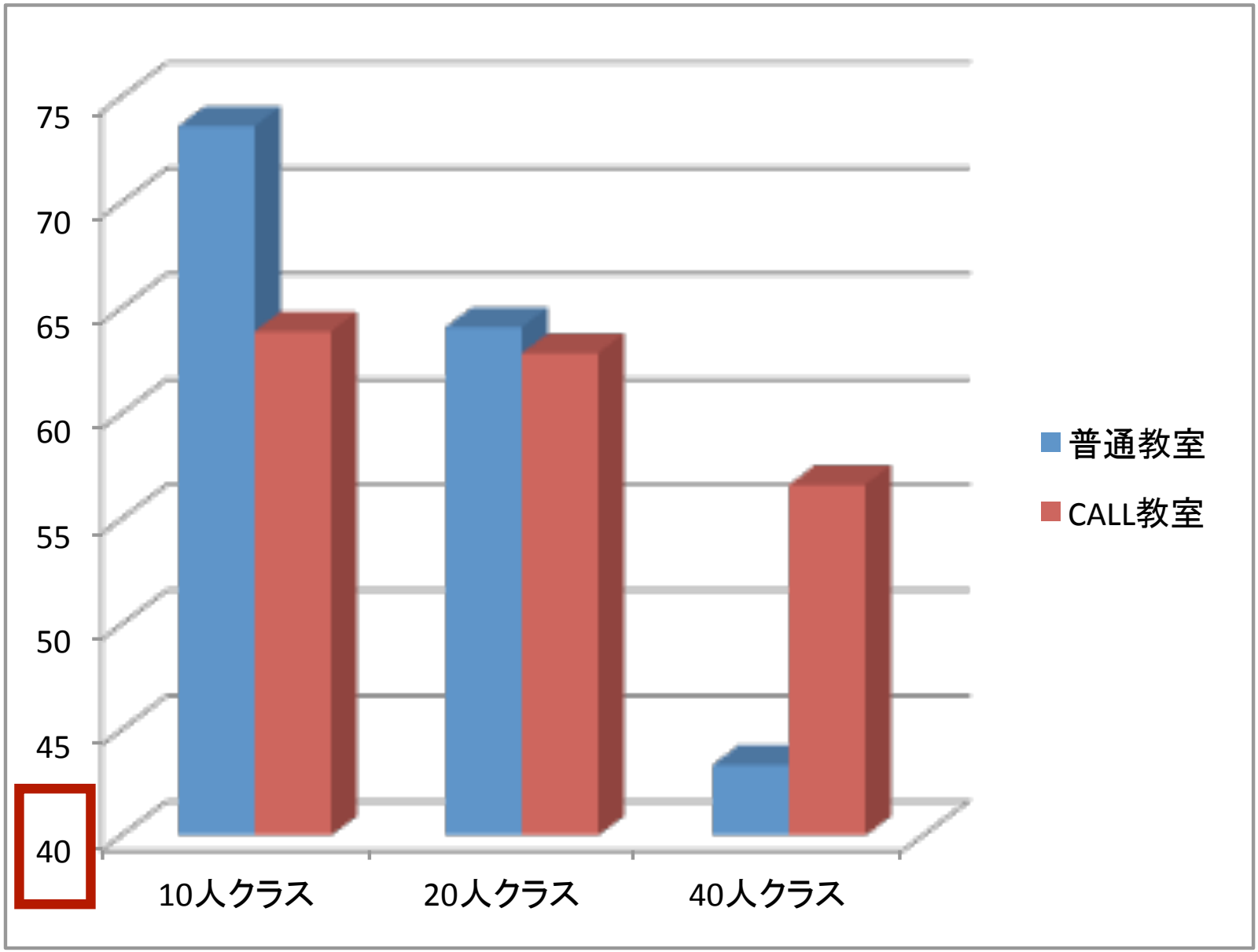




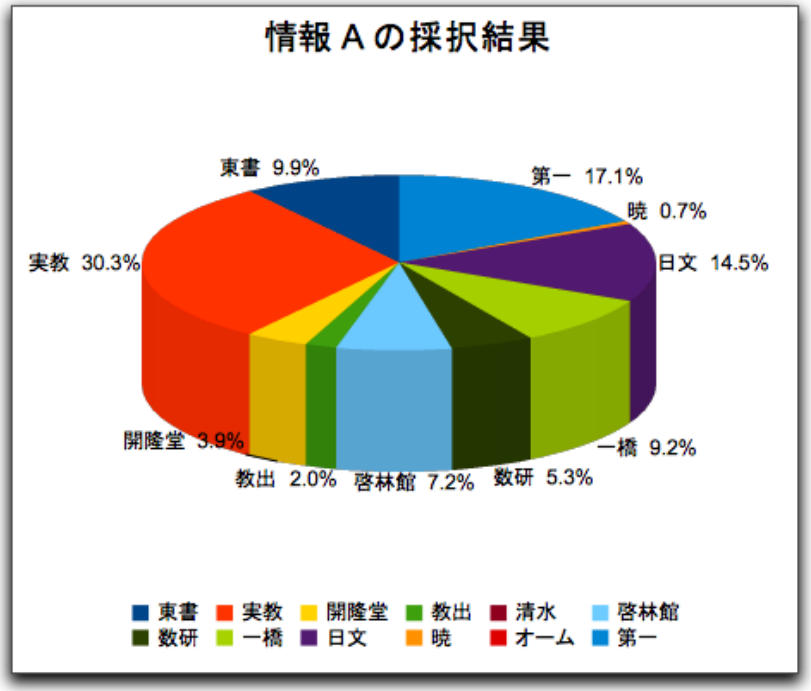
図表の作成

「ハンドブック」第22章も参照

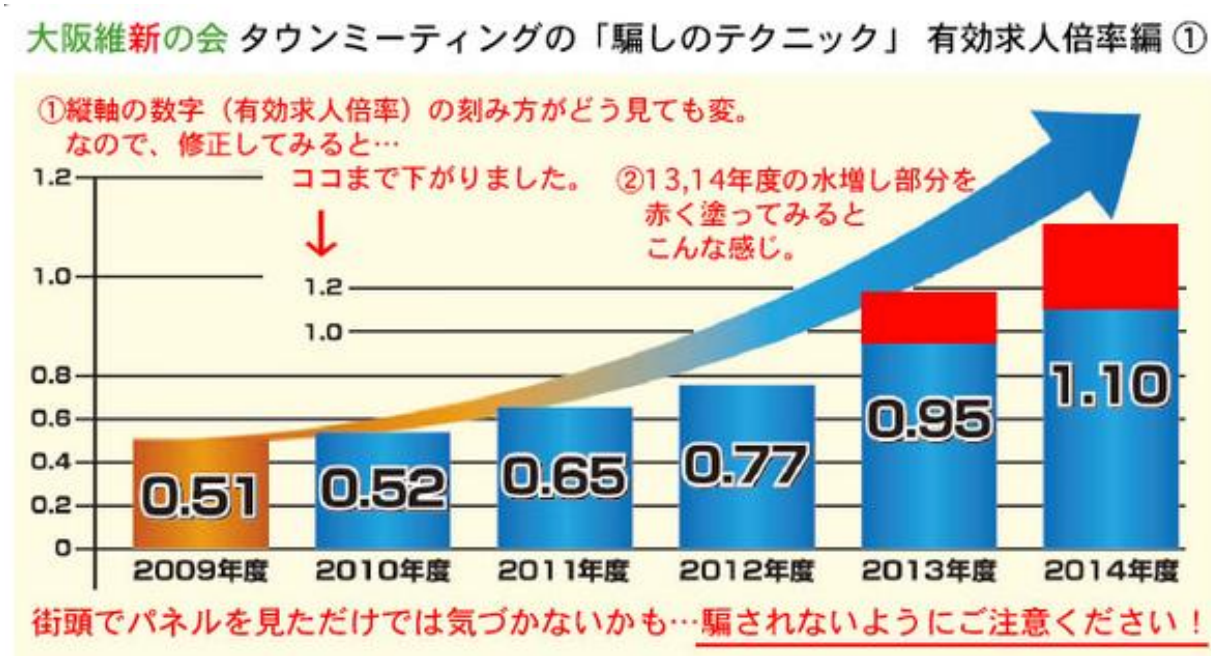
まずい図の例



まずい図の例



http://www.clas.kitasato-u.ac.jp/~fujiwara/infoScienceA/chooseGraph/chooseGraph.html



https://twitter.com/otouta/status/537616602259156992

「なんかExcelの3Dグラフかっこいいからやってみました」

サンプルをまねる

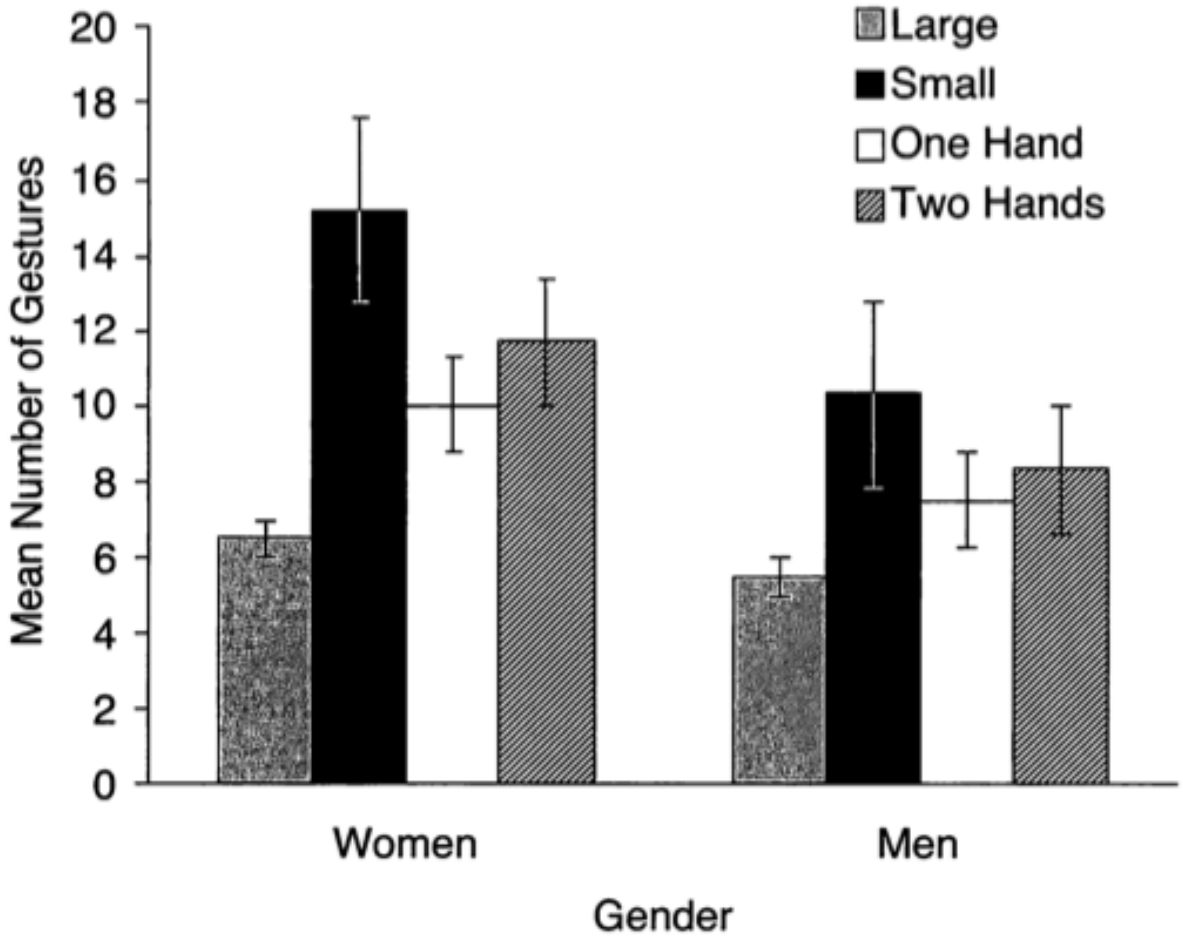
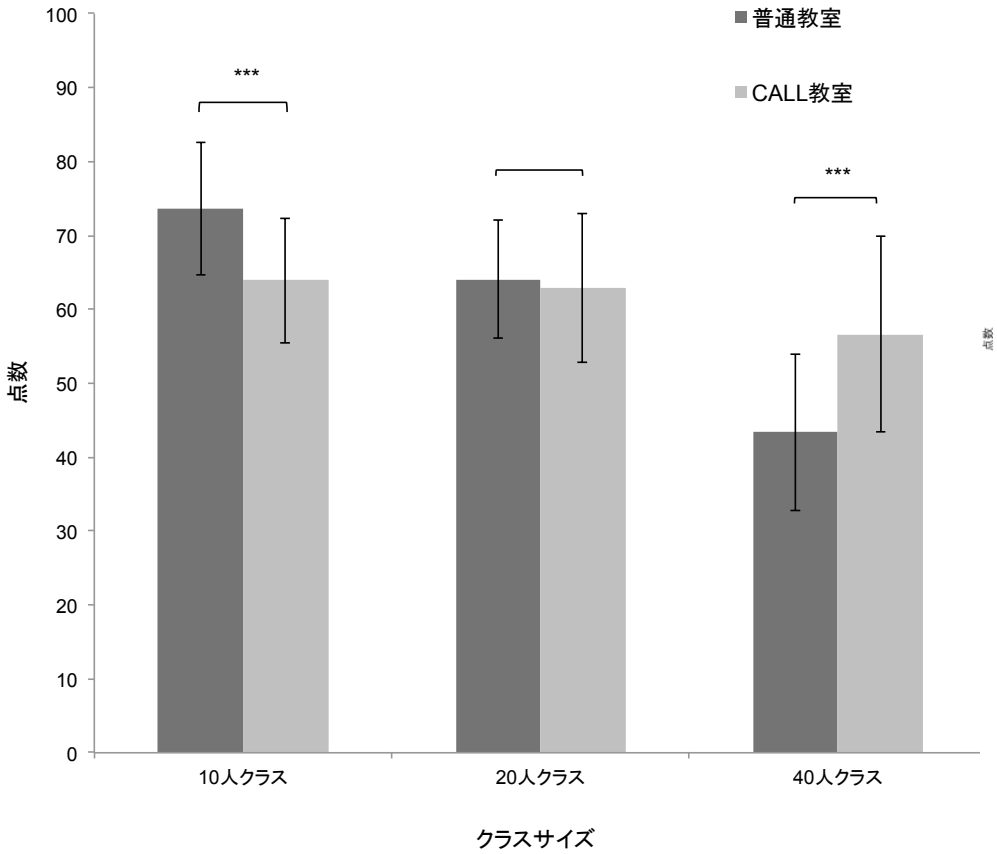
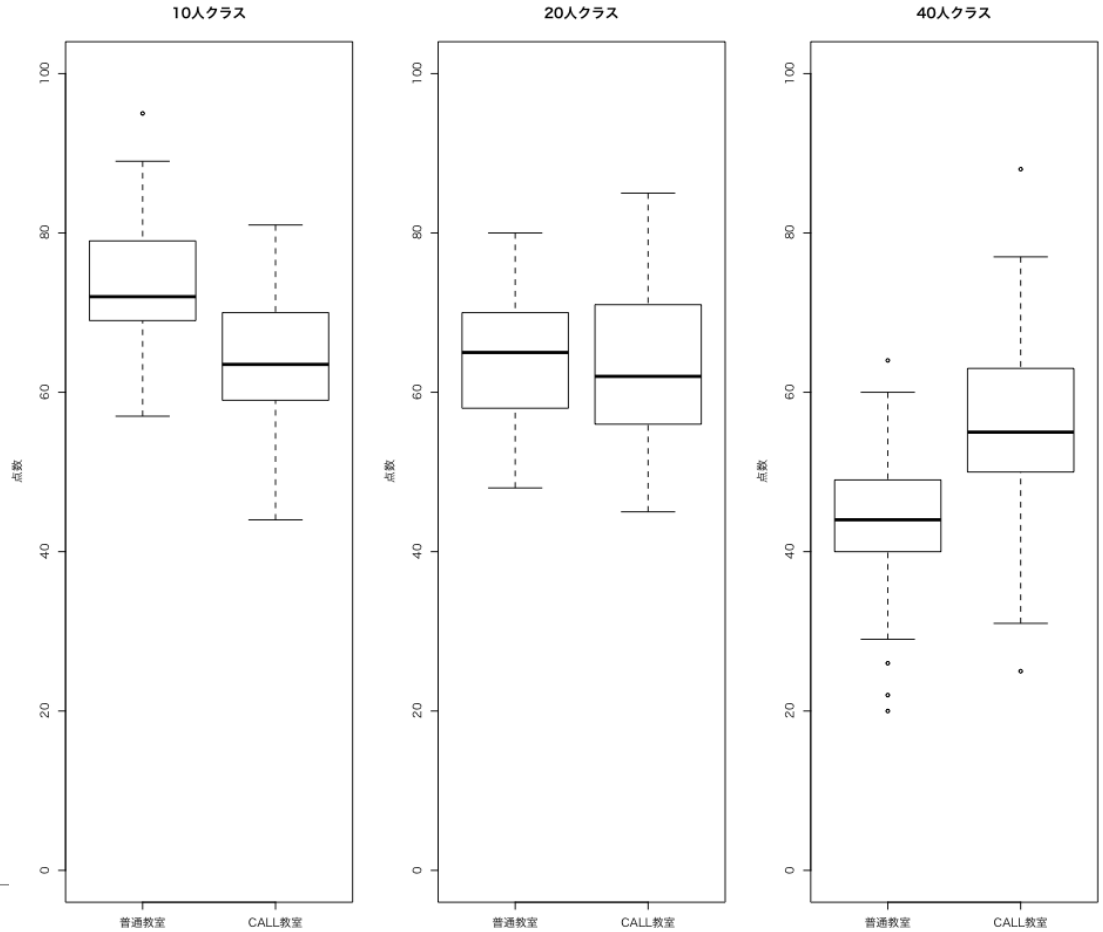


Figure X. Average number of times (±*SE*) different types of hand gestures were used by women (*n* = 20) and men (*n* = 20) when telling a joke over a 5-min period.

Figure 2.4 (p. 19) in Nicol, A. A. M., & Pexman, P. M. (2010) *Displaying your findings: A practical guide for creating figures, posters, and presentations* (6th ed.). Washington, DC: American Psychological Association.

「隠れる」「隠される」情報がない

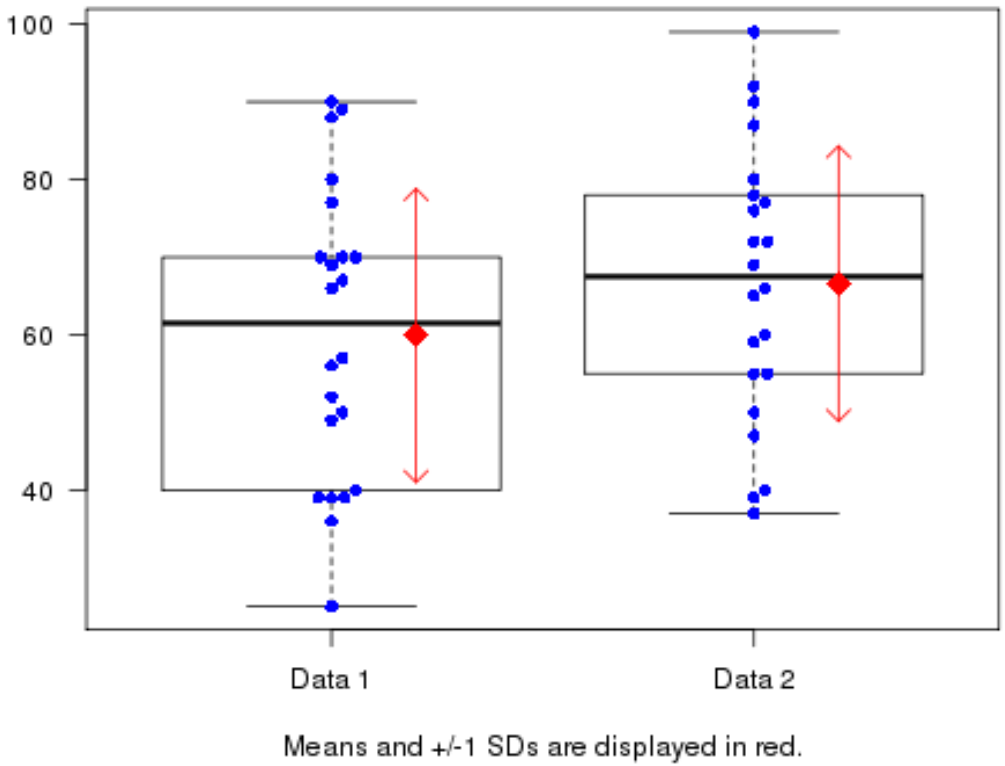


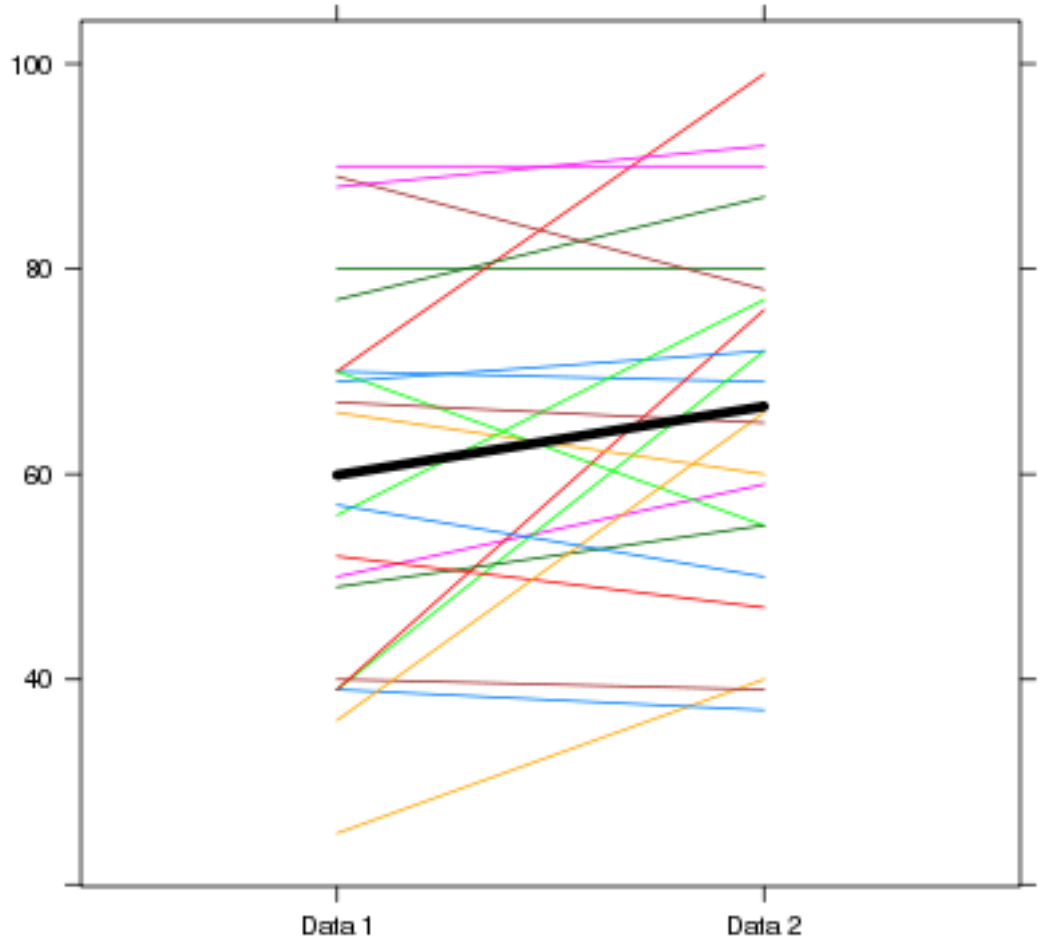


Not so good

Better

個別のデータも見せる

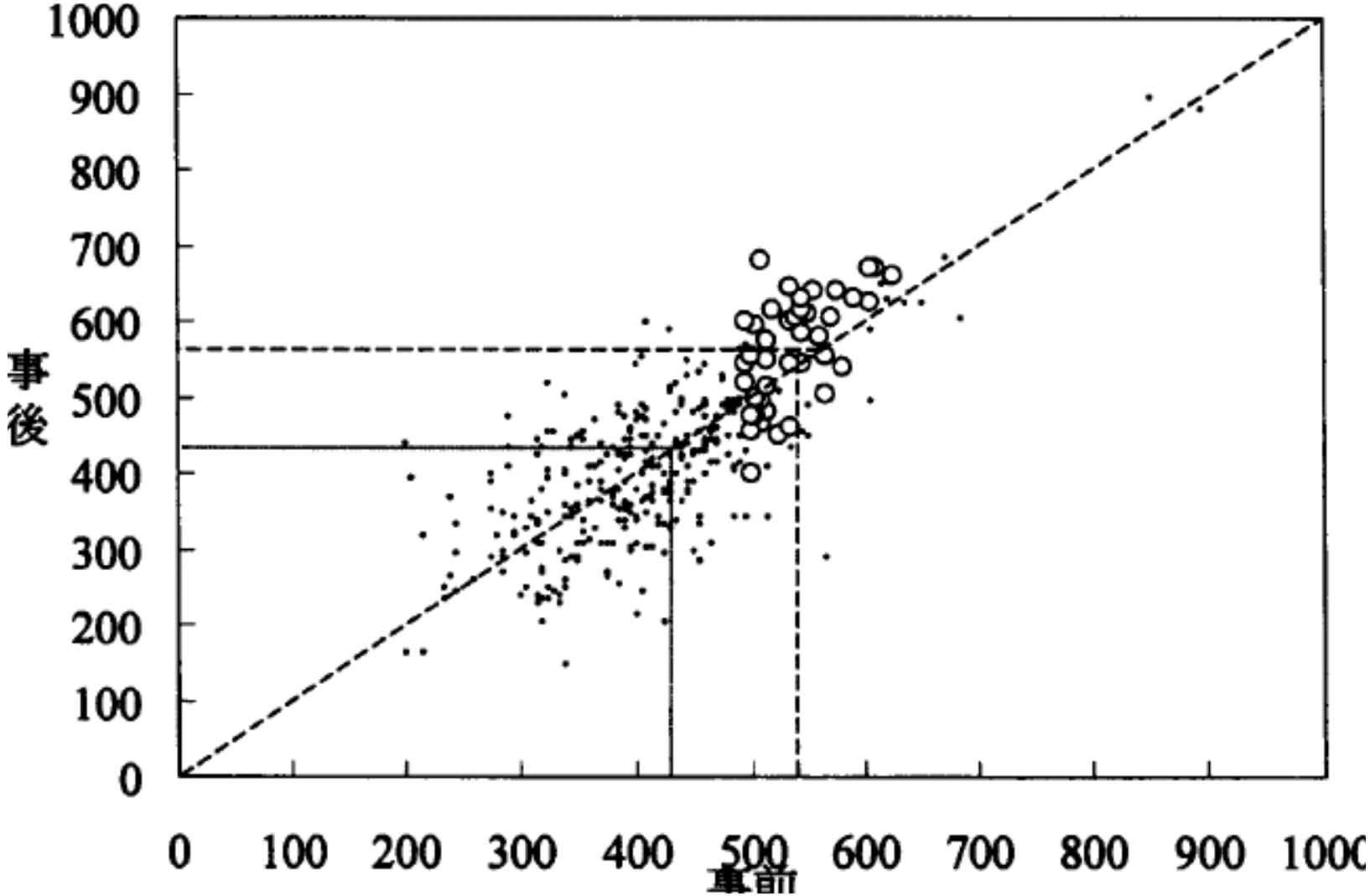




Much better

論旨に合った図示方法

前田啓朗 (2008). WBTを援用した授業で成功した学習者・
成功しなかった学習者. *ARELE*, *19*, 253–262.



もう一つの"売り"

外国語
教育研究
ハンドブック
―研究手法のより良い理解のために―
竹内 理 水本 篤
外国語教育学や
第二言語習得の論文において
多く使われている手法を網羅的に紹介し、
実際に論文を書くときに活用できるよう、
わかりやすく丁寧に指南！

http://mizumot.com/handbook

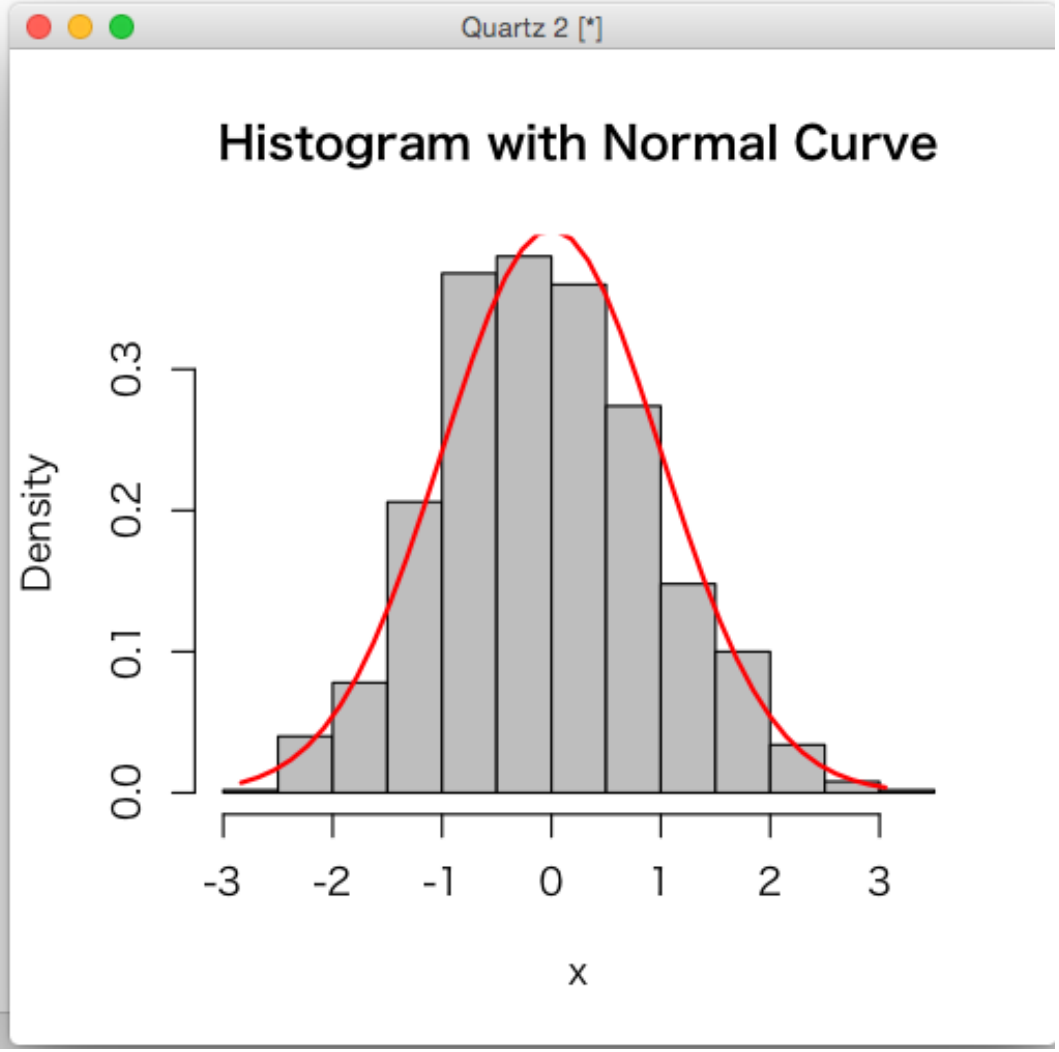
書籍で使用したデータと分析方法

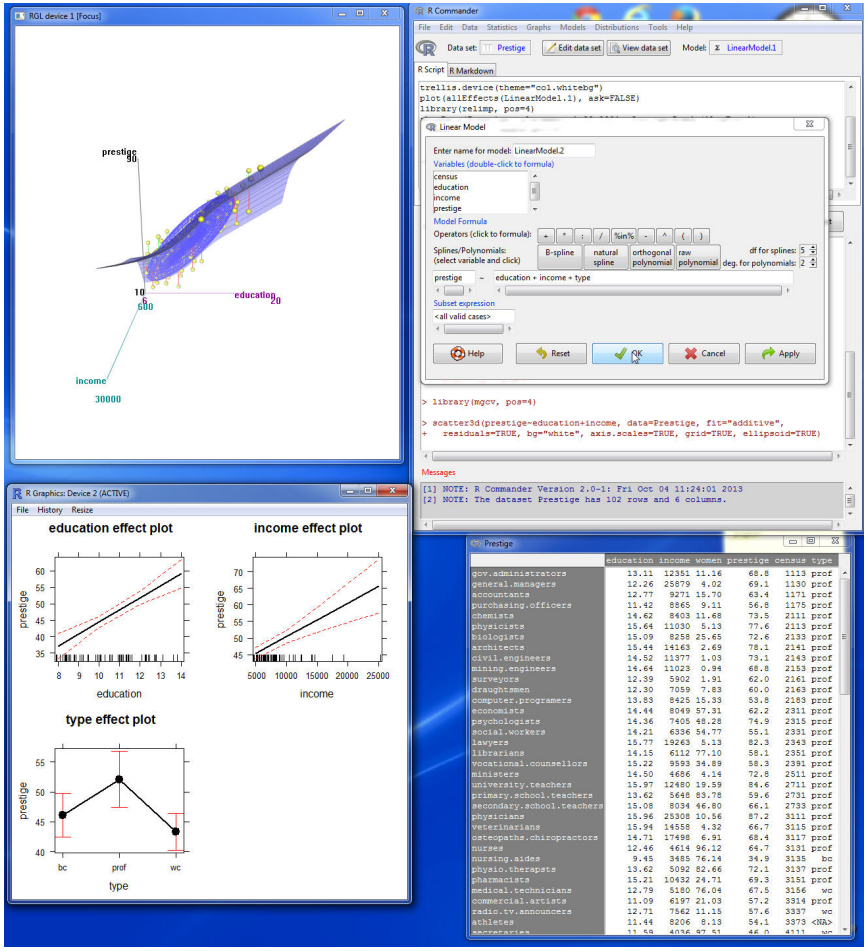
- MS Excel（できるものだけ）
- SPSS
- フリーのデータ解析環境R

- MS Excel（できるものだけ）
- SPSS
- フリーのデータ解析環境R

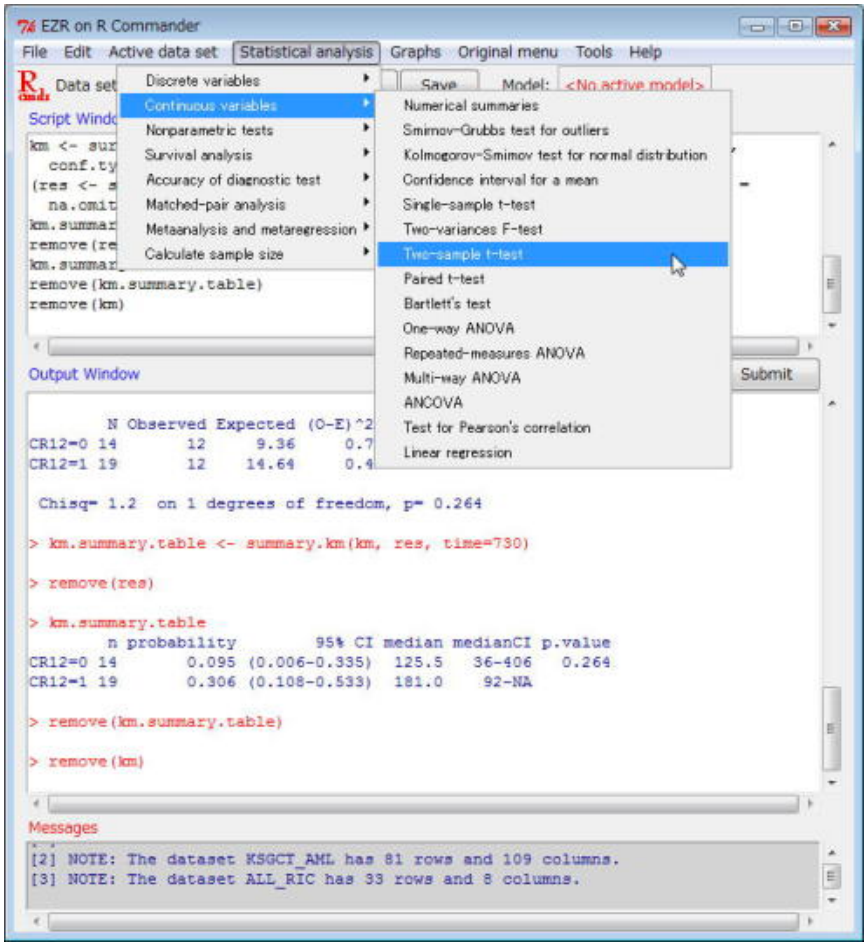
ただ... RはCLI

```
> x <- rnorm(1000)
> h<-hist(x, breaks=20, freq=FALSE,col="grey", xlab="x",main="Histogram with Normal Curve")
> xfit<-seq(min(x),max(x),length=40)
> yfit<-dnorm(xfit)
> lines(xfit, yfit, col="red", lwd=2)
>
```





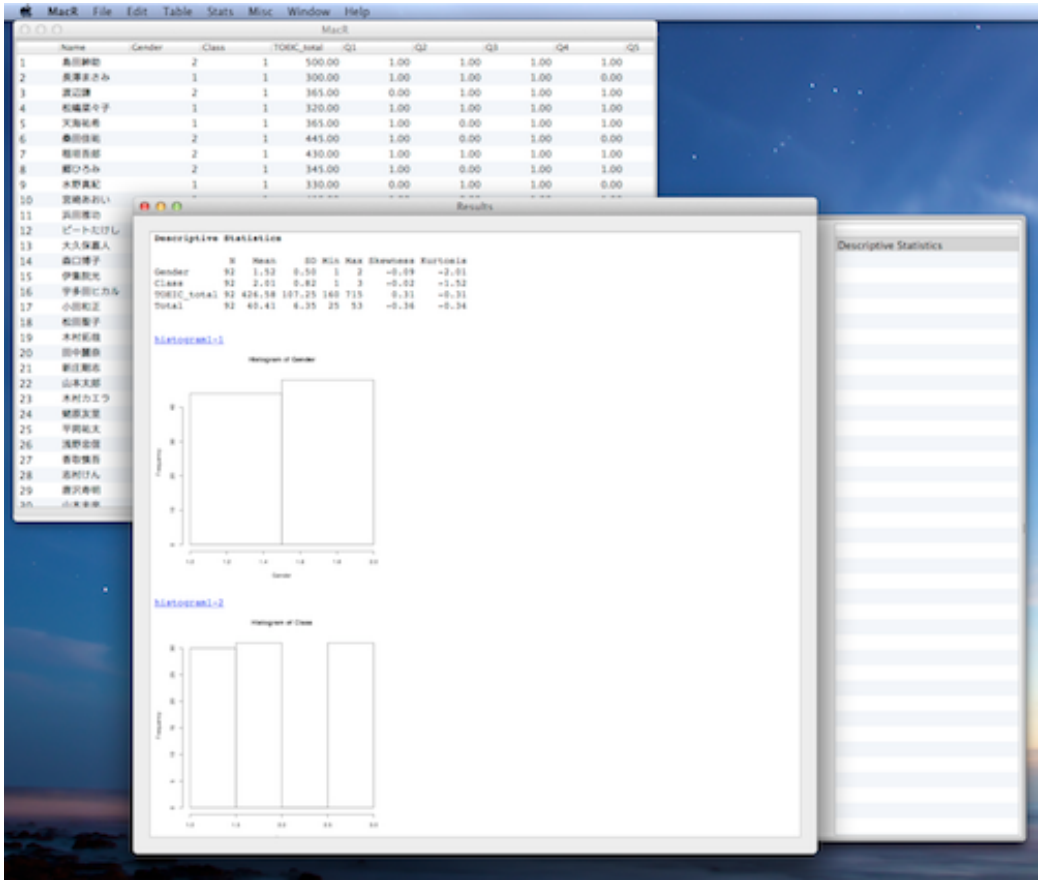
http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Rcmdr-screenshot.html



http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmedEN.html

RをGUIで利用できる

R Commander（EZR）など



https://sites.google.com/site/casualmacr/home

RをGUIで利用できる

Mac用アプリのMacR

さらに一歩進んで便利（というか楽）なのがWebアプリケーション

便利なWebアプリの例

PythonCGIによる統計計算

こんにちは統計学

—χ2乗検定・t検定・U検定・分散分析・相関係数—

プログラミング言語PythonをCGIに使った統計計算プログラムです。
とくに難しい設定は必要なく、ただデータを貼り付けるだけです。
心理検定などで統計計算が必要な方は、お試しになってみてください。

どんなときに、どんな統計分析を使うのか？

- 尺度の違いによる検定の種類（リッカート尺度@Wikipedia）
- パラメトリックな手法とノンパラメトリックな手法@群馬大学社会情報学部・青木繁信先生

ここでできる統計計算

〈基本統計計算〉

- 合計・平均・分散・標準偏差の計算

〈パラメトリック手法〉

http://www.m-sugaya.jp/python/

http://www.hju.ac.jp/~kiriki/anova4/index.html

便利なWebアプリの例

ANOVA4 on the Web

ANOVA4 on the Web について

Copyright (c) 2002 Kiriki Kenshi

新しい初期画面を準備しました（JavaScript使用）。お試しください。 → index_js.html
Excel上のデータを貼り付けるための説明ページを掲載しました。→ Excelデータの貼り付け方

要因計画を設定してください。 [ANVファイルを指定して実行]

《要因計画》

コメント:

要因の数: 被験者間要因(b)= 0 / 被験者内要因(w)= 0 --> [b＋wは４以内]

確定

js-STAR 2012

release 2.0.6j

Programing by Satoshi Tanaka & nappa(Hiroyuki Nakano)

copyright© all rights reserved,1998-2013

便利なWebアプリの例

★彡お知らせ

- What's new!
- 動作確認・バグ状況
- Tips

★彡各種分析ツール

度数の分析

- 1×2表(正確二項検定)
- 1×2表：母比率不等
- 1×j表(カイ二乗検定)
- 2×2表(Fisher's exact test)
- i×j表(カイ二乗検定)
- 2×2×k表(層化解析)
- i×j×k表(モデル選択)
- i×j×k×L表(4元モデル)
- 自動集計検定2×2

t検定(Rプログラム)

- t検定(参加者間)
- t検定(参加者内)

TOP

すばやいデータ分析を可能にする、ブラウザで動く、フリーの統計ソフト！

js-STAR 2012（以下js-STAR）は、わかりやすいインターフェースとかんたんな操作により、驚くほどすばやくデータ分析ができる、無償の統計ソフトです。

ブラウザ上で動作するため、WindowsでもMacでも使用できます。（※ただしInternet Explorer7と8をご使用の方はTips 7の設定、Internet Explorer10をご使用の方はTips 10の設定が必要です。）

表計算ソフトや統計ソフトRとの連携も充実しており、js-STARのみではできないデータ管理や、より高度な分析も可能にしています。

たとえば、表計算ソフトのデータは、テキストエリアやグリッドに、かんたんに貼り付けることができます。また、一部のツールでは、分析に必要なRプログラムを出力することができます。

やさしい解説書はありますか？

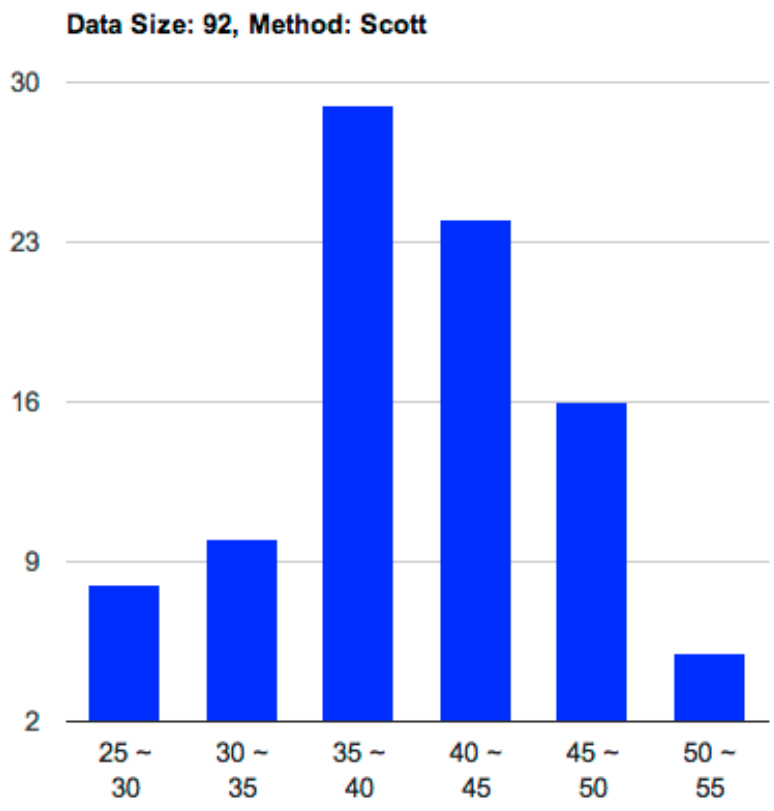
js-STARを利用した初学者向けの独習本

初学者向けとして、js-STARの基本操作と分析結果の読み方を解説した書籍が、技術評論社より発刊されています。

http://www.kisnet.or.jp/nappa/software/star/

こういうのを作りたかった

便利なWebアプリの例



http://hoxom-hist.appspot.com/hist.html

便利なWebアプリの例

こういうのを作りたかった

:: Cronbach Alpha - Free Statistics Software (Calculator) ::

Like Share 28

All rights reserved. The non-commercial (academic) use of this software is free of charge. The only thing that I ask in return is to cite this software when results are used in publication.

This free online software (calculator) computes the Cronbach alpha statistics for a set of items that are believed to represent a latent variable (construct).

Enter (or paste) a matrix (table) containing all data (time) series. Every column represents a different variable and must be delimited by a space or Tab. Every row represents a period in time (or category) and must be delimited by hard returns. The easiest way to enter data is to copy and paste a block of spreadsheet cells. Please, do not use commas or spaces to seperate groups of digits!

Send output to:

Browser Blue - Charts White

Data X (click to load default data)

7	7	7	5
5	5	5	5
6	5	4	4
4	5	5	5
5	5	5	5
6	6	6	7
7	4	7	7
6	5	5	6
6	7	7	6
6	5	5	6
5	2	4	6
5	5	6	6
4	5	4	6
6	6	6	6
6	6	7	7

Names of X columns:

Q2 Q9 Q16 Q23

Sample Range:
(leave blank to include all observations)

From:

Compute

http://www.wessa.net/rwasp_cronbach.wasp

普段R（やSPSS）でやってること

- csvやxlsなどで元データを準備
- Rにデータを読み込む
- パッケージの関数を使って分析

langtest.jp

langtest.jp

Web App

Cronbach's Coefficient Alpha (mirror)

Basic Statistics Calculator (mirror)

Learning by Doing Stats (t-test Tutorial) (mirror)

Comparing Two Independent Samples (mirror)

Comparing Paired Samples (mirror)

Effect Size Calculator 1 (Means, SDs, Ns) (mirror)

ANOVA (mirror)

Non-parametric Tests (mirror)

Correlation (mirror)

Regression Analysis (mirror)

Chi-square Test (mirror)

- ハンドブックの量的チャプターのサンプルを使用して再現できる。
- アウトプットの見方がわかる
- 自分でも簡単に分析できる。
- グラフを充実させている。
- Excelのデータをコピペするだけ。

Basic Statistics Calculator

Data input:

44
32
37
44
40
44
45
36
37
42
38
47
26
41
27
39
47
41
36
44
39
49
40
25
33
39
36
45
32
51

newlines, spaces, commas, or tabs.

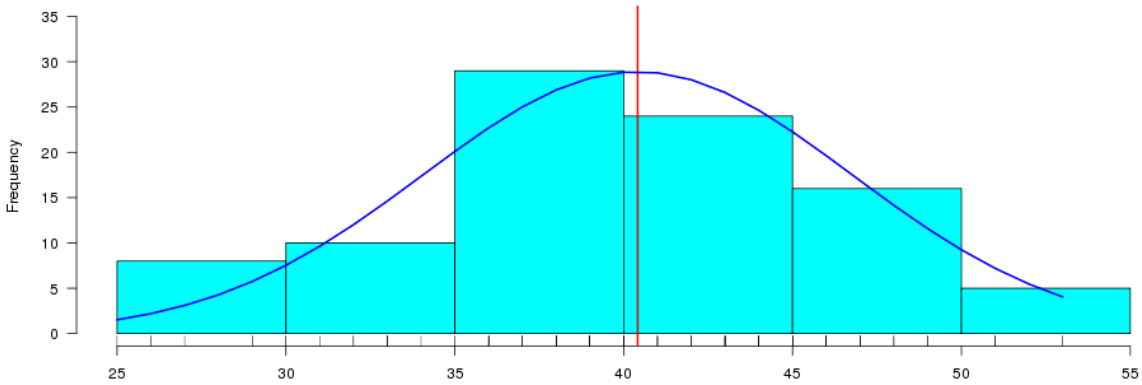
Main About

Basic statistics

```
  n  mean   sd median trimmed  mad min max range  skew kurtosis   se
 92 40.41 6.35     40   40.69 5.93  25  53    28 -0.36    -0.34 0.66
```

Histogram

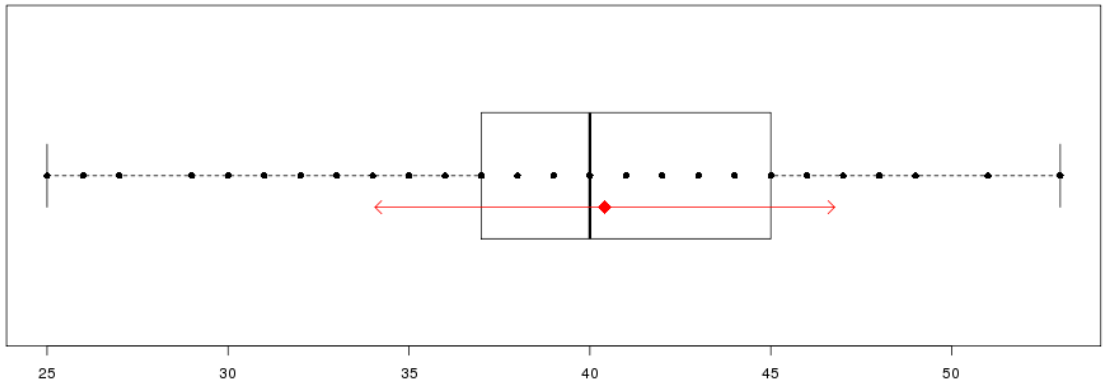
Download the plot as pdf



Red vertical line shows the mean.

Box plot with individual data points

Download the plot as pdf



Mean and +/-1 SD are displayed in red.

ここにExcelから
データをコピペするだけ

Cronbach's Coefficient Alpha

Main About

Option:

☑ **The input data includes variable names (header).**

Note: Input values must be separated by tabs. Copy and paste from Excel/Numbers.

	Item1	Item2	Item3	Item4
1				
2	2	3	3	3
3	3	3	4	4
4	4	4	3	4
5	5	4	3	4
6	3	4	2	4
7	3	3	4	3
8	4	3	4	4
9	3	3	2	2
10	4	5	5	5
11	2	2	1	2
12	4	3	4	3
13	3	4	3	3
14	3	4	4	3
15	3	4	3	4
16	5	5	5	4

行列もいける。

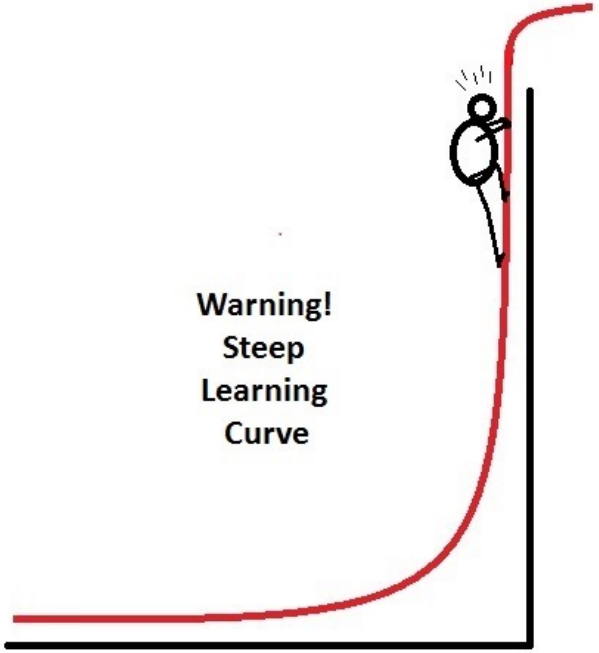
Basic statistics of the scale (test)

```
[[1]]
          n mean   sd median trimmed  mad  min  max range  skew kurtosis  se
Average  15 3.45 0.77    3.5    3.48 0.37 1.75 4.75     3 -0.22    -0.09 0.2
```

注意点

- 誰でもできる... だけに危険。
- ドキュメンテーションがない。
- サーバでRを走らせているので少し重い。
- 自由度ゼロ（要望が有り次第改善予定）。
- コードが残らないので再現性に乏しい。

対象と目的



- 学部生，修士課程の院生，入門者
  「ハンドブック」の分析をハンズオンで実行し，
  卒論，修論の分析で利用。

- 博士課程の院生，量的研究を行う研究者
  分析方法の確認，コードを見て自分でRを使う。
  （langtest だけでは不十分と感じるはずなので）

http://mizumot.com/handbook

http://langtest.jp/

langtest.jp

Web App

Cronbach's Coefficient Alpha (mirror)

Basic Statistics Calculator (mirror)

Learning by Doing Stats (t-test Tutorial) (mirror)

Comparing Two Independent Samples (mirror)

Comparing Paired Samples (mirror)

Hands-on Session

http://mizumot.com/ws/cle2016.xls

1. 平均と標準偏差

2. 対応のない*t*検定，対応のある*t*検定
   （結果の見方，解釈と効果量の確認）

3. 分散分析（説明のみ）

4. ノンパラメトリック検定