信頼性 -相対信頼性と絶対信頼性

国際医療福祉大学保健医療学部理学療法学科　下井俊典

Ⅰ. 級内相関係数（ICC）

1. 信頼性と妥当性

1）信頼性 reliability

テスト結果の正確さについての概念

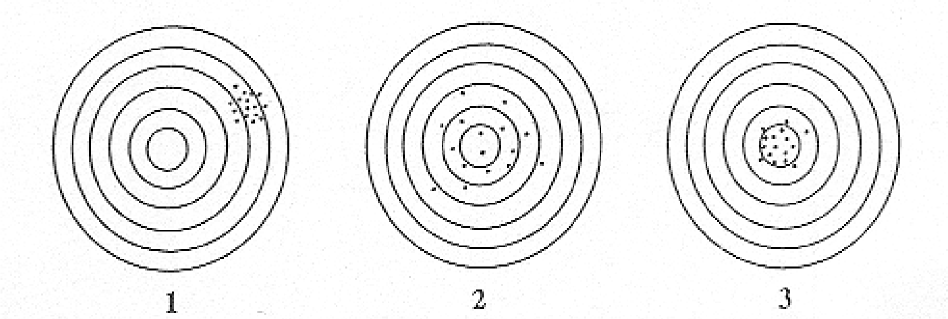
反復して測定したときに同じ結果が得られる程度（再現性・一貫性）

2）妥当性 validity

テストが測定しようとしているものを実際に測定している程度

“ A test is valid if it measures what it purports to measure ”（Kelley, 1927）

Target-Shooting analogyfor validity and reliability（Julius Sim and Peggy Arnell, 1993）



2. 安定性・再現性と一貫性

1）安定性 stability, 再現性

同一被検者に、同一条件で、同一のテストをおこなった場合に、同一（傾向）の結果がでるか

再テスト法 test-retest method、平行テスト法 parallel forms of reliability

2）一貫性 consistency

同一個人に、同様のテストを行った場合に同一傾向の結果がでるか

折半法 split-half correlations、内部一貫性 internal consistency

3. 検者内信頼性と検者間信頼性

1）検者内信頼性 intra-class reliability

同一被検者に期間を空けて同一テストを実施し、複数のテスト結果を比較した時の安定性・一致度（stability over time）

再テスト法 test-retest method

a）測定間隔：基本的には 3 ～ 4 日

b）検者間の違いの要素は考慮しない

c）問題点：測定間隔、練習効果

2）検者間信頼性 inter-class reliability

複数の検者間を通じての安定性・一致度 stability between examiners

a. ランダム効果

「検者間の完全一致」を求める

母集団から無作為抽出された複数の検者を想定

b. 固定効果

「相対的に平行な関係（一致）」

いずれの検者も「a は b より常に 0.5 秒遅い」となれば、信頼性が高いと判断する

異なる母集団から抽出された検者を想定

4. 複数の測定値の「一致（度)」の検討方法

1）連続変数

ピアソンの相関係数（r），級内相関（ICC），最小二乗法による（線形）回帰分析

2）カテゴリー変数

κ係数（κ統計量, kappa coefficient）

偶然によらない一致率の指標

		観察者A		
		陽性	陰性	合計
観察者B	陽性	59	6	65
	陰性	11	166	177
	合計	70	172	242

(59+ 166)/ 242= 93%…「見かけ上の一致率（observed degree of agreement）」

$$\kappa\text{係数（偶然によらない一致率）} = \frac{\text{見かけ上の一致率のうち、偶然によらない一致率}}{\text{全体の一致率のうち、偶然によらない一致率}}$$

$$= \frac{93.0\% - 59.8\%}{100\% - 59.8\%}$$

$$= 0.82$$

κ係数の判定

0 ～ 0.40　：低い一致（poor agreement）

0.41 ～ 0.60　：中等度の一致（moderate）

0.61 ～ 0.80　：かなりの一致（good to fair）

0.81 ～　：高い一致（excellent）

5. ICC の適用方法

1）ICC の種類と適用方法

検者内信頼性　$ICC_{(1,1)}$、$ICC_{(1,k)}$

検者間信頼性

ランダム効果　$ICC_{(2,1)}$ 、$ICC_{(2,k)}$

固定効果　$ICC_{(3,1)}$ 、$ICC_{(3,k)}$

2）一般化可能性理論（generalizability theory）

a. G 研究（generalizability study; 一般化可能性研究）

測定のばらつき（分散）から信頼性を求める

$ICC_{(\chi,1)}$ を用いる

b. D研究（decision study; 決定研究）

G 研究で推定された測定のばらつき（分散）から、適切な測定の使用計画（回数）を求める

繰り返し測定による信頼性の向上の計画

$ICC_{(\chi,k)}$ を用いる

D研究による信頼性の向上の例

G 研究により ICC(1,1)=0.73 が得られた

研究で必要な信頼係数が 0.9 以上の場合に必要な繰り返し測定回数（k）を求める

$$k = \frac{p_1(1-p_2)}{p_2(1-p_1)}$$

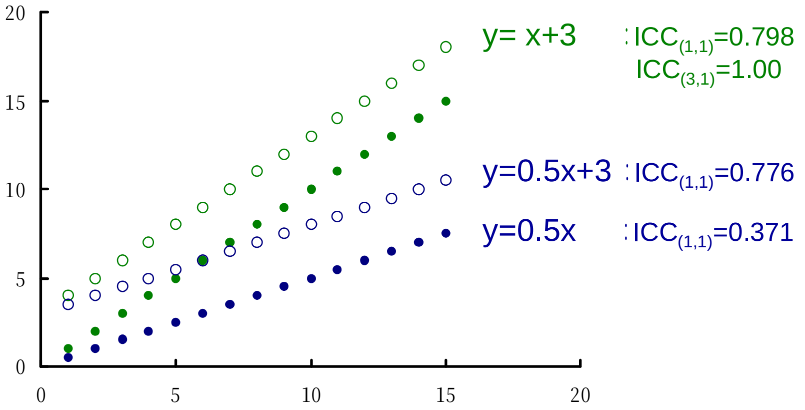
$$k = \frac{0.9\times(1-0.73)}{0.73\times(1-0.9)} = 3.33$$

4 回繰り返し測定した値の平均値を用いる　ICC(1,4)＞ 0.9

6. 相関係数（r）と級内相関係数（ICC）の関係

1）相関係数（r）：2 つの測定値が直線関係にあれば 1 となる

2）ICC



7. 相関係数（信頼係数, r）の大まかな基準

0.81 ～	almost perfect
0.61 ～0.80	substantial
0.41 ～0.60	moderate
0.21 ～0.40	fair
0.0～0.20	slight

（Landis,1977）

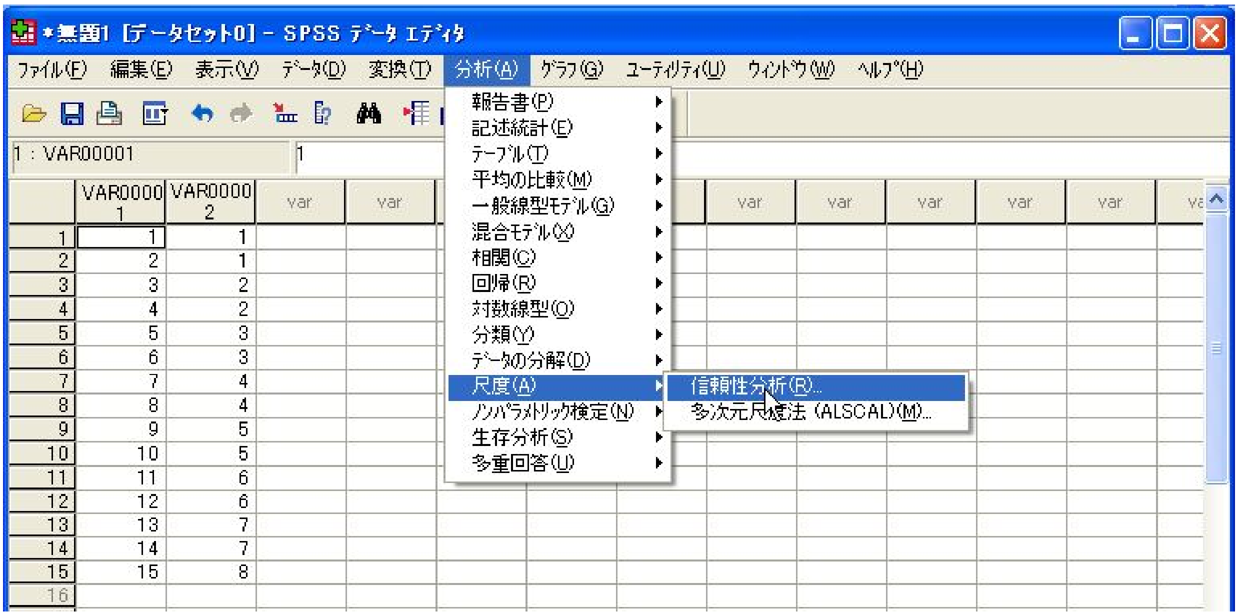
0.9～	great	優秀
0.8～	good	良好
0.7～	fair (OK)	普通
0.6～	possible	可能
～0.6	re-work	要再考

（桑原, 1993）

8. SPSS を用いた ICC の算出　-SPSS 14.0J-

1）データの入力

2）ツールバー「分析」－「尺度」－「信頼性分析」

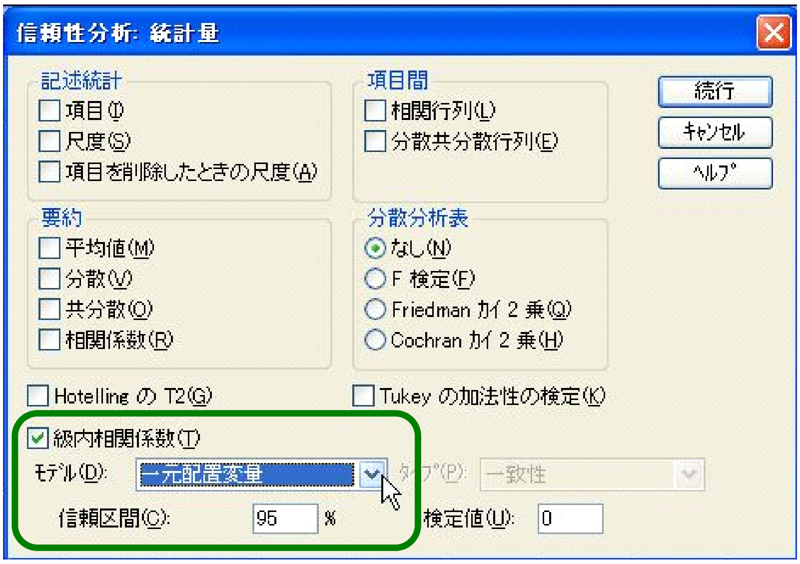


3）「信頼性分析」ウィンドウ

a. 対象とする変数を「項目」ボックスに移動させる

b. 「統計」をクリック

4）「信頼性分析- 統計量」ウィンドウ



a. 「級内相関係数」を選択

b. モデル

ICC(1,1), (1,k)：「一元配置変量」

ICC(2,1), (2,k)：「二元配置変量」

ICC(3,1), (3,k)：「二元配置混合」

c. 「続行」をクリック

5）結果

級内相関係数

	級内相関	95% 信頼区間		真の値 0 を使用した F 検定			
		下限	上限	値 (判別分析)	df1	df2	有意確率
単一測定値	.371	-.141	.731	2.177	14	15	.073
平均測定値	.541	-.328	.844	2.177	14	15	.073

人的効果が変量であるときの一元変量効果モデル。

「単一測定値」：ICC(1, 1)

「平均測定値」：ICC(1, k)　この例では、k=2

Ⅱ. 級内相関係数の限界

1. 誤差：“ error ” と ” bias ”

- error：偶然誤差（random error）
  - （生物学的）個体差（biological variation）
  - 測定誤差（measurement error）
- bias：系統誤差（systematic bias）
  - 固定誤差（fixed bias）
  - 比例誤差（proportional bias）

2. 系統誤差（systematic bias）

1）特定方向に出現する「乖離」

偶然誤差（error）：不定方向に出現する「誤差」

2）種類

a. 固定誤差（fixed bias）

真の値に関わらず、特定方向に一定の幅で生じる「乖離」

b. 比例誤差（proportional bias）

真の値に比例して増減する、特定方向に生じる「乖離」

3）系統誤差の問題点と対策

a. 問題点

a）繰り返し測定で克服できない

双方向に乖離する偶然誤差は、繰り返し測定で克服できる

b）試験の計画・実施段階で混入した系統誤差に対しては、検定・推定は原則的に無力

b. 対策

a）実験デザインの検討による系統誤差の回避（フィッシャーの 3 原則）

例）rondomization, double blind

b）どの種類の誤差が、どの程度混入しているかの検出が必要

3. ICC、相関係数の限界

・ICC(x,k)でわからないこと：繰り返し測定で精度が向上できるのは、「偶然誤差」のみがある場合

・臨床応用を前提とするならば、2 つの測定値間に差がないことの証明

どの種類のばらつき（誤差、乖離）が、どの程度あるか、
その「ばらつき」は臨床応用上、問題ないか

Ⅲ. 絶対信頼性

1. 相対信頼性と絶対信頼性

1）相対信頼性

・2 群の測定値の一貫性を、相関係数により判断するもの

・単位をもたない

・手法：相関係数、ICC

2）絶対信頼性

・2 群の測定値の変動性を、測定値と同じ単位であらわしたもの

・測定値と同じ単位をもつ

・手法

Bland-Altman 分析
測定の標準誤差(standard error of measurement; SEM)
最小可検変化量(minimal detectable change; MDC)

2. 絶対信頼性による信頼性の検討手順

1）研究デザインの検討（フィッシャーの 3 原則）

a. 反復

b. 無作為化

c. 局所管理

2）系統誤差の有無・種類の確認：Bland-Altman 分析（Bland and Altman, 1986）

a. 特徴

2 つの測定値の比較から

①測定値間に含まれる誤差の種類の特定

②臨床応用の際の許容範囲（LOA: limits of agreement）の算出

が可能

※ 2 つの測定値

2 種類の方法による変数（平行テスト法、併存妥当性）

2 名の検者による 2 変数（検者間信頼性）

1 名の検者による 2 変数（検者内信頼性）

b. 手順

a）“ Bland-Altman plot ” の作成

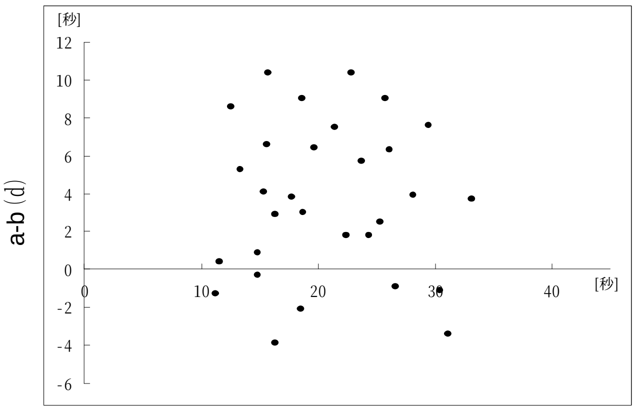
y 軸：2 つの測定値の差（d）

x 軸：2 つの測定値の平均

※真の値が不明な場合、2 つの測定値の平均が真の値の 予測値として最適

（Bland and Altman, 1986）

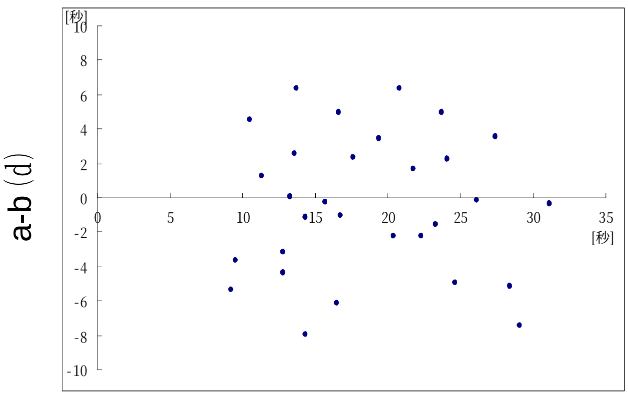
$$A_{(x,y)} = \left(\frac{a+b}{2}, a-b \right)$$



(a+b)/2

b）2 つの測定値間に偶然誤差しか存在しない場合

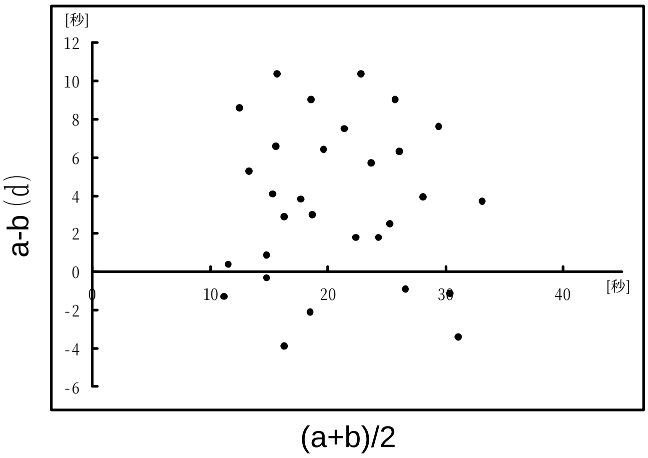
真の値（2 つの測定値の平均）にかかわらず、2 測定値は誤差（2 測定値の差; d）が y 軸 +/-双方向に均等に分布



(a+b)/2

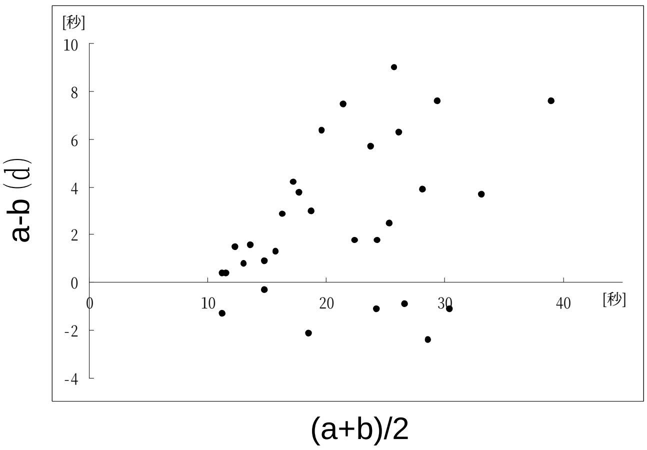
c）固定誤差が存在する場合

真の値（2 つの測定値の平均）にかかわらず、誤差（2 測定値の差; d）が特定方向（y 軸+/-いずれかの方向）に乖離



d）比例誤差が存在する場合

真の値（2 つの測定値の平均）の増加に伴い、誤差（2 測定値の差; d）も増加する（右に開いた扇形の散布図）



c. 系統誤差の統計学的な検出方法

a）固定誤差の検出

・2 つの測定値の差（d）の平均（d）が 0（ゼロ）であるという帰無仮説を棄却する

・d の 95%信頼区間（95%CI）が 0（ゼロ）を含まない

d の 95%CI： $\bar{d} \pm t \times \sqrt{\dfrac{SD_d}{n}}$

※標本数（n）、d の標準偏差（SD_d）、自由度 n-1 の t 値

例 1

	A	B	C	D	E
1		1回目の測定値（a）	2回目の測定値（b）	(a+b)/2	a-b (d)
2	被験者1	25.1	17.6	21.4	7.5
3	被験者2	23.3	21.5	22.4	1.8
4	被験者3	33.2	25.6	29.4	7.6
5	被験者4	20.9	10.5	15.7	10.4
6	被験者5	22.8	16.4	19.6	6.4

$\bar{d} = 3.6$　$SD_d = 4.1$　$t = 2.045$　より　$3.6 \pm 2.045 \times \sqrt{\dfrac{(4.1)^2}{30}}$

$= 3.6 \pm 1.5$

$\therefore 2.1 \sim 5.1$

b）比例誤差の検出

・Bland-Altman plot の傾きが 0（ゼロ）であるという帰無仮説を棄却する（回帰分析）

・Bland-Altman plot の相関係数（r）が 0（ゼロ）であるという帰無仮説を棄却する（無相関検定）

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

例 2

	A	B	C	D	E
1		1回目の測定値（a）	2回目の測定値（b）	(a+b)/2	a-b (d)
2	被験者1	25.1	17.6	21.4	7.5
3	被験者2	27.4	29.8	28.6	-2.4
4	被験者3	42.8	35.2	39.0	7.6
5	被験者4	23.3	21.5	22.4	1.8
6	被験者5	33.2	25.6	29.4	7.6

n = 30, r = 0.37 より

$$t = 0.37\sqrt{\frac{30-2}{1-(0.37)^2}}$$

$$= 2.107$$

自由度 28, 有意水準 5%の t 値：2.048

3）誤差の範囲の推定

a. 加算誤差のみが認められた場合：誤差の許容範囲（limits of agreement; LOA）

2 つの測定値の差の平均（d）、d の標準偏差（SDd）、 95%信頼区間の z 値（1.96）、標本から得られた LOA の標準誤差（SELOA）、自由度 n-1 の t 値より

$$(\bar{d} - 1.96 \times SD_d) + t \times SE_{LOA} \quad \sim \quad (\bar{d} + 1.96 \times SD_d) - t \times SE_{LOA}$$

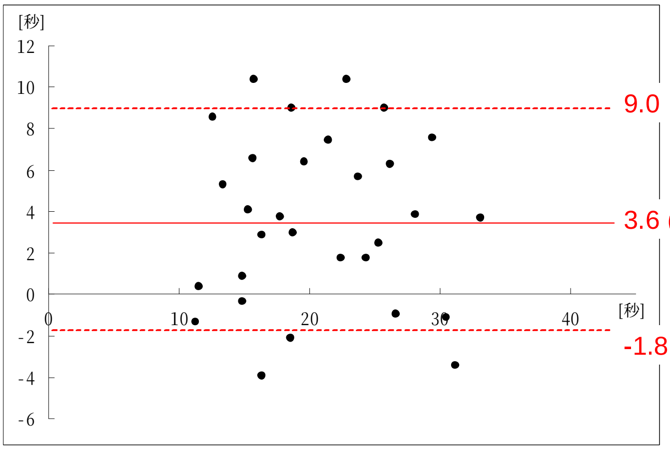
SE_{LOA} の概算値

$$SE_{LOA} = \sqrt{\frac{3SD_d^2}{n}}$$

例 1 では、

$$(\bar{d} - 1.96 \times SD_d) + t \times SE_{LOA}$$
$$= (3.6 - 1.96 \times 4.1) + 2.045 \times \sqrt{\frac{3 \times (4.1)^2}{30}}$$
$$= -1.8$$

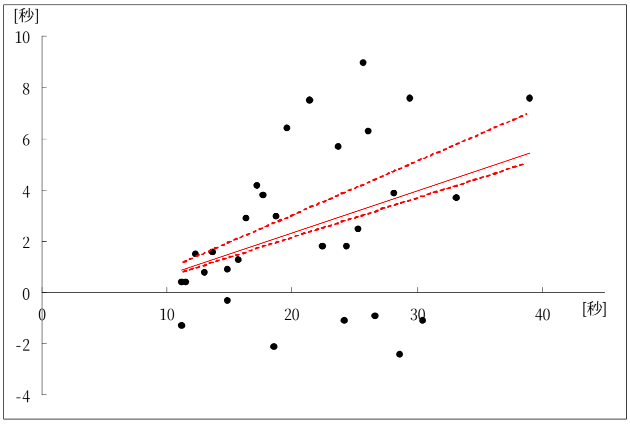
$$(\bar{d} + 1.96 \times SD_d) - t \times SE_{LOA}$$
$$= (3.6 + 1.96 \times 4.1) - 2.045 \times \sqrt{\frac{3 \times (4.1)^2}{30}}$$
$$= 9.0$$

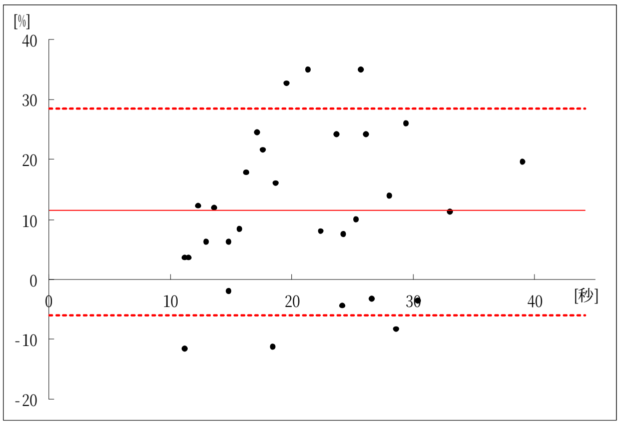


b. 比例誤差が認められた場合：相対軸プロット（percent difference plot）

y 値：x 値（対応する 2 測定値の平均、(a+b)/2）の相対値（%d）

$$A'_{(x,y)} = \left(\frac{a+b}{2}, \frac{a-b}{\left(\frac{a+b}{2}\right)} \right)$$





例 2 では

	A	B	C	D	E	F
1		1回目の測定値（a）	2回目の測定値（b）	(a+b)/2	a-b (d)	%d
2	被験者1	25.1	17.6	21.4	7.5	35.0
3	被験者2	27.4	29.8	28.6	-2.4	-8.4
4	被験者3	42.8	35.2	39.0	7.6	19.5
5	被験者4	23.3	21.5	22.4	1.8	8.0
6	被験者5	33.2	25.6	29.4	7.6	25.9

$\overline{\%d} = 11.1 \quad SD_{\%d} = 13.2$ より

$\therefore LOA: -6.2 \sim 28.4$

c. 系統誤差が認められなかった場合（偶然誤差のみが認められた場合）

最小可検変化量（minimal detectable change; MDC）

$$MDC = SEM \times 1.96 \times \sqrt{2}$$

SEM（測定の標準誤差, standard error of measurement）

$$SEM = \frac{SD_d}{\sqrt{2}}$$

$$\therefore MDC_{95} = 1.96 \times SD_d$$

例 3

	A	B	C	D	E
1		1回目の測定値（a）	2回目の測定値（b）	(a+b)/2	a-b (d)
2	被験者1	21.1	17.6	19.4	3.5
3	被験者2	19.3	21.5	20.4	-2.2
4	被験者3	29.2	25.6	27.4	3.6
5	被験者4	16.9	10.5	13.7	6.4
6	被験者5	18.8	16.4	17.6	2.4

$SD_d = 4.1$ より $\quad MDC_{95} = 1.96 \times 4.1$

$= 8.0$

4）Bland-Altman 分析の特徴

a. 長所

a）系統誤差の検出

b）LOA の算出

c）検者内・検者間信頼性、平行テスト法（信頼性、基準関連妥当性）に利用可能

b. 短所

a）基本的には 2 種類の方法・2 名の検者による測定値しか比較・検討は不可能

b）単位の異なる測定方法の比較・検討は不可能

c）固定誤差と比例誤差の厳密な検出力に劣る

例）固定誤差と比例誤差の両方が混在する場合

IV. 参考文献

下井研究室ホームページ　http://shimoi.iuhw.ac.jp/