連　載

リハビリテーション分野の研究で用いられる統計手法

対 馬 栄 輝[1†]
[1]弘前大学大学院保健学研究科

キーワード：データ解析，統計手法，分散分析，級内相関係数

1. はじめに

分散分析 (analysis of variance；ANOVA と略すことが多い) は，3 つ以上の標本または変数の平均差の検定であり，実験のデザインによっては更に多くの手法に細分化される．さらに多重比較法という手法もある．

級内相関係数 (Intra-class correlation；以下，ICC) は，測定における検者内または検者間信頼性を表す係数である．その計算のためには，分散分析で得られた結果を用いる．また，関連した検者内・検者間信頼性係数として，カッパ係数 (κ 係数) という統計値もある．

リハビリテーション分野で用いられる統計的手法は，かなり多岐にわたるが，比較的頻繁に用いられつつも適用に混乱するのは分散分析と ICC またはカッパ係数であろう．本稿では，これらの手法の適用，注意点について解説する．分散分析に限っては，全ての手法を網羅することは困難なので，頻繁に用いられる 2 要因までの分散分析に限定して述べる．

2. 分散分析の基礎事項

分散分析は Fisher により確立された手法で，2 標本または 2 変数の平均差を検定する t 検定はもちろんのこと，回帰分析とも基本的な理論は同一である．分散分析は，他の手法と異なって独特の用語がある．そこで手法を述べる前に，まず基本的な用語を解説しよう．

2.1 要因とは

データをとるために定められた系統的な基準のことであり，因子とも呼ばれる．何らかの変数の差を検定したいとき，ある系統的な基準に従って条件を決めるのが普通である．図 1a は A 町，B 市，C 町に在住する対象者 6 名ずつに対して歩行速度を測った架空のデータ例である．この例では，｛A 町，B 市，C 町｝という地域の違いの要因について 3 群に分け，差を検定することになる．また，図 1b では，地域の要因に加え，さらに ｛40 歳代，60 歳代，80 歳代｝という年代の要因（因子）に着目して 3 群に分けて差を見ようとしている例である．なお，分散分析によって要因に有意な差があったときは，「有意な主効果がみられた」とか「主効果が有意であった」などという．

2.2 水準とは

水準とは，要因内の個々の群のことである（処理ともいう）．上述の例では ｛A 町，B 市，C 町｝ の個々の群，｛40 歳代，60 歳代，80 歳代｝ の個々の群を水準という．

2.3 対応のない要因と対応のある要因とは

「対応のない要因」とは，水準ごとに対象群が異なる要因である．換言すれば，水準がそれぞれ独立した群で構成されている要因である．「対応のある要因」とは，被検者に対して全ての水準を反復測定した要因のことで，対象群は変えずに測定条件（水準）を変えた場合である．図 1a や図 1b の ｛A 町，B 市，C 町｝ または図 1b の ｛40 歳代，60 歳代，80 歳代｝ の各水準は，それぞれが独立した対象群なので対応のない要因となる．他方，図 1c，図 1d の ｛1 週後，2 週後，3 週後｝ の各水準（変数）や，図 1e の ｛上り，下り，平地｝ という路面状況の要因は，同じ被検者に対して条件を変えて反復測定しているので，対応のある要因となる．

2.4 固定要因（固定因子）と変量要因（変量因子）

この用語は特別覚える必要はないが，SPSS（日本 IBM）などの統計ソフトウェア（以下，統計ソフト）で出てくる用語なので，念のために解説しておく．もし，このような用語の出てこない統計ソフトでは，気にする必要はない．固定要因（固定因子とも呼ぶ）とは，簡単にいえば差を検定しようとする要因のことである．通常，分散分析によって差を検定しようとする要因は固定要因である．

変量要因（変量因子）とは，水準の差を検定する必要がない要因である．細かいことをいえばきりがないが，変量要因とは被検者の要因と思えば良い．被検者個人が水準と見なされ，まとめて被検者の要因となる．例えば図 1c の ｛山田，佐藤，鈴木，…｝ という個人（水準）の差が有意であろうがなかろうがどうでも良く，差を検定する必要はない（多くの統計ソフトでは有意確率が出力されるが，無視して良い）．滅多にないと思うが，仮に個人差も検定したいというのであれば，被検者の要因は固定要因になる．救い

2010 年 7 月 28 日受付
† 〒 036-8564 青森県弘前市本町 66-1
弘前大学大学院保健学研究科
対馬　栄輝
Tel: 0172-39-5980　Fax: 0172-39-5981
E-mail: pteiki@cc.hirosaki-u.ac.jp

なのは，変量要因と固定要因を取り違えて解析しても，検定結果は変わらない点である．

2.5 交互作用

交互作用は，2 要因以上の分散分析（例えば図 1b，d，e）で出力される．交互作用が有意だと，結果の解釈が厄介である．図 2 に交互作用の有無に関する 1 例を挙げた．交互作用のない例（図 2a，b）では，各要因の各水準どうしで値の変化が一定している．従って，平均を結んだ線が平行となる．もちろん，これは母集団の平均の話なので，実際のデータでは完全に平行にはならない．交互作用のある例（図 2c，d）では，平均を結んだ線が平行にならない．

交互作用のない図 2a や b のときは，単純に治療 A 群と B 群に差があるか否か（年代要因｛若年群，老年群｝に関わらず），若年群と老年群に差があるか否か（治療要因｛治療 A 群，治療 B 群｝に関わらず）について判定する．

交互作用のある図 2c や d のときは，上述のようにはいかない．各水準で平均の変化のパターンが違うので，

治療の差を見たいときは，

・若年群の治療 A 群と若年群の治療 B 群の差

・老年群の治療 A 群と老年群の治療 B 群の差

年代の差を見たいときは，

・治療 A 群の若年群と治療 A 群の老年群の差

・治療 B 群の若年群と治療 B 群の老年群の差

という形に細分化して，各水準ごとに差を検定する必要がある．

2.6 分散分析の適用ルール

分散分析を適用するルールとして，

(1) データの尺度は名義尺度以外である

(2) 各水準のデータが正規分布に従う

が満たされてなければならない．つまり，分散分析はパラメトリックな手法である．さらに，対応のある・対応のない要因によって，

(3) 対応のない要因では，各水準のデータのバラツキが同じでなければならない（等分散性の仮定）

(4) 対応のある要因では，水準間の差の分散が等しくなければならない（球形性の仮定）

という前提条件が必要となる．

(1) は解析者が決めなければならず，(2) についてはシャピロ・ウイルク Shapiro-Wilk の検定を用いる．以前はデータのヒストグラムを観察して正規分布に従うかを確認したものだが，シャピロ・ウイルクの検定を使えば客観的に判断できる．正規分布に従わないデータに対しては，代わりと

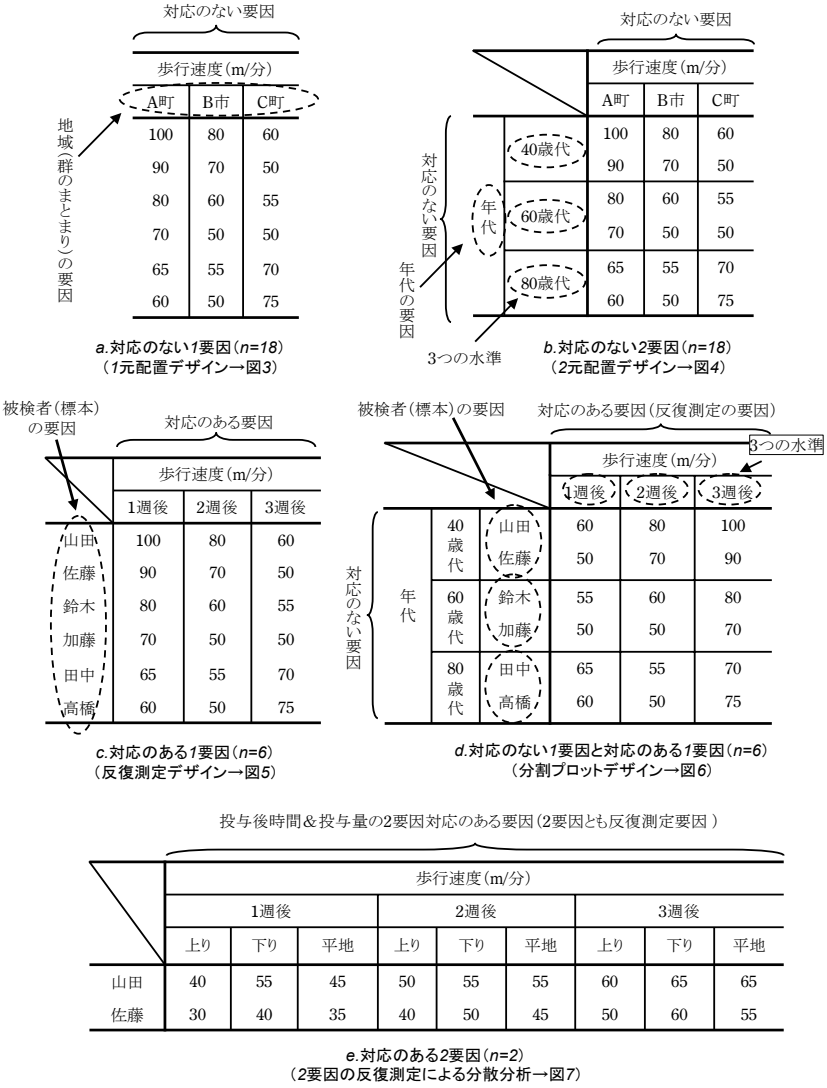


図 1　分散分析の用語と基本的な分類

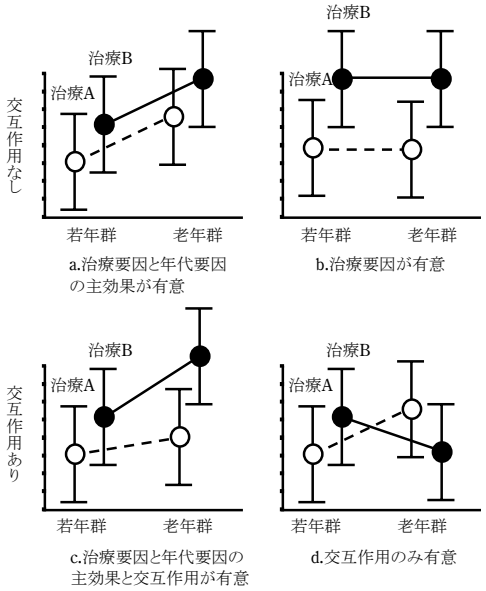


図 2　主効果と交互作用の例

治療要因｛治療 A，治療 B｝と年代要因｛若年群，老年群｝の主効果と交互作用が有意となる組み合わせの例を，平均±標準偏差のエラーバーグラフで表した．

なるノンパラメトリックな手法を適用する．(3) の確認にはレーベン Levene 検定を用いる．等分散性の検定としては，レーベン検定の他に F 検定やバートレット Bartlett 検定といった手法もある．(4) についてはモークリー Mauchly の球形性検定がある．

3. 分散分析の分類

分散分析の基本分類は，対応のある要因と対応のない要因の組み合わせまたは各要因の数によって，5 パターンに分類できる（図 1）．自分の測定したデータが図 1 のどのパターンに該当するかを判断すれば，あとは図 3 ～ 7 に従って統計ソフトに組み込まれている同じ名称の検定手法を選ぶだけである．

図 1a は 3 つ以上の群間の差を見たい場合に用いられる 1 元配置デザインの解析で，さらに要因を 1 つ増やした手法が図 1b の 2 元配置デザインである．もちろん，要因を増やすごとに 3 元配置，4 元配置，…，となる．

図 1c ～図 1e は対応のある（反復測定）要因が存在する．対応のある・ない要因の組み合わせによって 3 つに分類されている．

3.1 1 元配置デザインの解析手順

図 1a のように表せるデータを，ここでは 1 元配置デザインと呼ぶことにする．このデータは，図 3 のような手順で解析する．

1 元配置分散分析は，各水準のデータが正規分布かつ等分散していなければならない．そこでまず，データが正規分布に従うかをシャピロ・ウイルクの検定で確認する．シャピロ・ウイルク検定は全ての水準に対して個々に行い，全ての水準で正規分布に従わないとはいえない結果（$p \geqq 0.05$）が出力されたときは，次に等分散性の検定としてレーベン検定を行う．

レーベン検定は全ての水準に対して一度に行う手法である．レーベン検定の結果で $p \geqq 0.05$ のときは「等分散性が成り立たない，とはいえない」ので 1 元配置分散分析を適用する．仮に，レーベン検定の結果で $p<0.05$ のときは，ウェルチ Welch の補正による 1 元配置分散分析を適用させる．仮にウェルチの補正がプログラムされていない統計ソフトであれば，通常の 1 元配置分散分析を適用するしかない．シャピロ・ウイルク検定によって正規分布に従わないことがわかったなら，クラスカル・ワリス Kruskal-Wallis の検定を適用する．

3.2 2 元配置デザインの解析手順

図 1b のように対応のない 2 要因のデータは 2 元配置デザインとなる．検定の手順は，図 4 に示した．これは 3 元配置以上のデザインでも同様の手順となる．1 元配置デザインからみると，かなり簡素化されているが，これには理由がある．2 元配置デザインでは，ノンパラメトリックな手法が存在しないことと，ウェルチの補正ができないということである．仮に，等分散性の検定を行ったとしても，ウェルチの補正はできないために，無意味である．従って図 1b のようなデータでは，2 元配置分散分析を行うしかない．このように，2 元配置デザインでは，かなり妥協しなければならない点がでてくる．

もともと分散分析には頑健性（ロバストネス）という性質があり，データが「ある程度」は正規分布に従わなくても，妥当な結果を出力する特性を持つ．従って，さほど厳密に前提条件を確認する必要はないと考える．ただし，この「ある程度」についてはどれくらいかは明確にできない．

3.3 反復測定デザインの解析手順

反復測定デザインのデータは図 1c のように，同じ対象に 3 つ以上の条件や測定時間を変えて反復測定した要因（対応のある要因）の差を検定する手法である．検定の手順は図 5 の通りとなる．

対応のある要因の検定なので，事前にモークリーの球形性検定を行うのが特徴である．この検定が有意（$p<0.05$）であったときは，分散分析の結果が有意となりやすいため，自由度を ε 修正した分散分析を行う必要がある．ε 修正の方法としては，主にグリーンハウス・カイザー Greenhouse-Geisser の方法やホイン・フェルト Huynh-Feldt の方法，下限の方法の 3 つが知られているが，よく利用されるのはグリーンハウス・カイザーの ε 修正である．これらの手法は，SPSS のような統計ソフトでは出力されるが，それ以外の統計ソフトでは出力されないものも多い．統計ソフトで出力されない時は，下限による ε 修正が簡単である．1/（要因の水準数－ 1）で求めた値を自由度に乗じる（グリーンハウス・カイザーの保守的検定という）という方法である．この下限による ε 調整で有意な差があるなら，有意差を強く主張できる．しかし，かなり厳しい判定となるので，推奨されていないのも事実である．

3.4 分割プロットデザインの解析手順

図 1d のようなデータは，分割プロットデザインの分散分析の適用となる．図 1d では，｛1 週後，2 週後，3 週後｝の 3 水準で成り立つ対応のある要因が 1 つと，｛40 歳代，60 歳代，80 歳代｝の 3 水準で成り立つ対応のない要因が 1 つの，合計 2 つの要因で構成されるデータ例だが，対応のある要因と，ない要因が組み合わされていれば，要因・水準はいくつあっても分割プロットデザインとなる．これも，ノンパラメトリックな手法は存在しない．また，対応のある要因の差の検定なので，事前にモークリーの球形性検定を行う必要がある．具体的な手順については，図 6 の通りである．なお，対応のある要因の存在する 2 要因以上の分散分析では，モークリーの球形性検定よりはメンドーサ Mendoza の多標本球形検定の方が望ましい．しかし，解析できる統計ソフトは非常に少ないため，現状ではこれが限界である．

3.5 2 要因の反復測定デザインの解析手順

図 1e のデータは，2 要因の反復測定の分散分析の適用となる．具体的な手順については，図 7 の通りとなる．これも，ノンパラメトリックな手法は存在せず，事前にモークリーの球形性検定（これも望ましくはメンドーサ Mendoza の多標本球形検定）を行う．3 要因以上でも同様の手順となる．

4. 多重比較法

4.1 適用について

分散分析またはそれに代わるノンパラメトリックな手法によって，有意な差があった（主効果が有意であった）要因については「要因全体に有意な差がある」となる．どの水準とどの水準に有意な差があるかまでは言及できない．更に，どの水準とどの水準に差があるかを知るためには，次に続く検定として水準どうしの差を多重比較法で検定する必要がある．この「分散分析で有意な差があった後に多重比較で検定する」という手続きを post-hoc 検定という．多重比較法の代わりに，対応のある t 検定，ウィルコクソン Wilcoxon 検定，2 標本 t 検定やマン・ホイットニー Mann-Whitney 検定といった 2 群・2 変数の差を検定する手法は使えない．

パラメトリックな手法としての多重比較法は，テューキー Tukey 法やシェフェ Sheffe 法が推奨される．テューキー法に比較してシェフェ法は，有意な差が出にくい特徴を持つ．ウェルチの補正による 1 元配置分散分析の後に行う多重比較法としては，ゲームス・ハウェル Games-Howell 法を適用する．ゲームス・ハウェル法は，ウェルチの補正によるテューキー法と考えられる．

クラスカル・ワリスの検定（図 3）やフリードマン Friedman の検定（図 5）といった，ノンパラメトリックな手法の場合も，要因全体に有意な差があることを検定するに過ぎない．post-hoc 検定として，例えばノンパラメトリックな手法の多重比較法であるスティール・ドゥワス Steel-Dwass 法が適用となる．

4.2 適用上の問題について

多くの多重比較法と分散分析は，検定統計量が異なるので，post-hoc 検定では，判定の矛盾が生じる．分散分析で有意な差があるのにテューキー法では有意な差がない，またはその逆が起こるという矛盾である．シェフェ法は，分散分析と同じ検定統計量なので，この問題は生じない．また，post-hoc 検定では検定の多重性の問題が存在する．シミュレーション実験では，5％有意水準を保つことが出来ないことを確認している [1]．これらの問題は理論的に明白である [2]．

解決法は post-hoc 検定を行わないことである．つまり水準間の差を見たいのであれば，分散分析は行わずに最初から多重比較法を適用すれば良い．しかしながら post-hoc 検定は，計算機の発達していなかった時代の名残として，リハビリテーション領域の研究はおろか医学全般の研究でも慣習的に行われている現状である．

反復測定要因の水準どうしの比較に多重比較法を適用することにも問題がある．多重比較法は，対応のない要因の水準（独立な群どうし）を比較する手法である．そのために，対応のある水準どうしの検定では，帰無仮説の下で有意水準を保てないという問題が生じる．筆者が R2.8.1（CRAN，free software）を用いて簡易的に行ったシミュレーションでは差が出難くなる性質を見出している．現状の統計ソフトでは，テューキー法やシェフェ法を適用させるのが限界であろう．どうしても解決したいなら，ボンフェローニ Bonfferoni の不等式に基づいたボンフェローニ法や，ボンフェローニ法を応用したシェイファー Shaffer 法の適用が正確である．

4.3 ボンフェローニ法

手順は以下の通りとする．

1) k 個の水準の組み合わせで，データの型に応じて，対応のある t 検定，ウィルコクソン検定，2 標本 t 検定，マン・ホイットニー検定といった 2 群・2 変数の差の検定法を適用する．
2) i=｛k ×（k − 1)｝÷ 2 を求める．
3) 1) の検定結果の p 値に i を乗ずる．
4) p × i 値が，有意（p<0.05）かどうかを確認する．

例として，1 元配置分散分析で 3 つの水準の差が有意であった図 1a のようなデータを考えよう．

1) まず，全ての水準の組み合わせで 2 標本 t 検定を行う．検定の結果は，A 町と B 市の検定で p12 ＝ 0.063，A 町と C 町の検定で p13 ＝ 0.044，B 市と C 町の検定で p23 ＝ 0.901 であった．
2) i ＝（3 × 2）÷ 2 ＝ 3
3) p12 ＝ 0.063 × 3 ＝ 0.189，p13 ＝ 0.044 × 3 ＝ 0.132，p23 ＝ 0.901 × 3 ≧ 1 となる．P23 は 1 を超えてしまっている．
4) 有意水準を 5％とすると，全てにおいて有意な差はない．

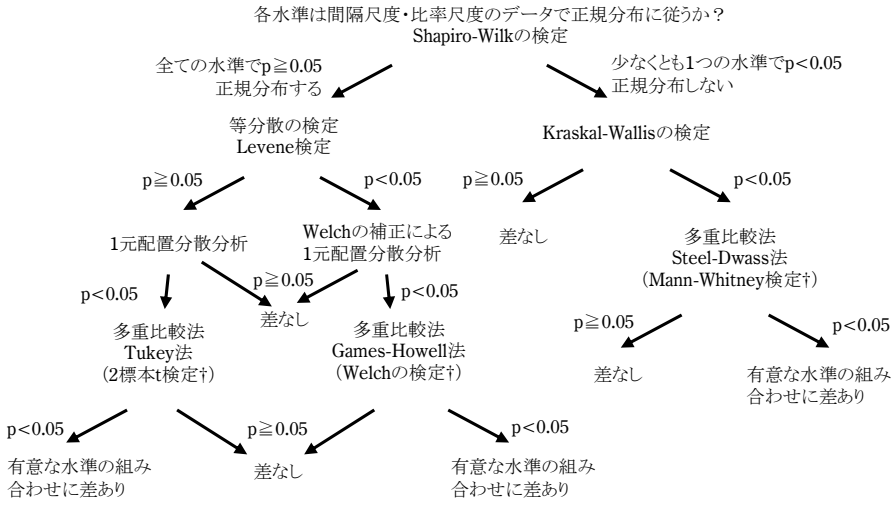
となる．ボンフェローニ法は水準の数が多くなるほど，差が出にくいという欠点を持つ．そこで，この欠点を補うためのシェイファー法を推奨する．

4.4 シェイファー法

ボンフェローニの不等式を基にしつつ，ボンフェローニ法のような無駄な p 値の引き下げを行わない手法である．MSRB（Modified Sequentially Rejective Bonferroni）法とも呼ばれる．ボンフェローニ法よりも推奨される方法である．ただし，シェイファー法をプログラムした統計ソフトは皆無に等しい．また判定手順が，やや複雑である．しかし，計算そのものは電卓や Excel などでも簡単にできる．計算手順は紙面の都合もあって解説できないが，永田ら 2) が参考となる．また，筆者の Web[3] に計算の助けとなる簡単な Excel ファイルを掲載している．

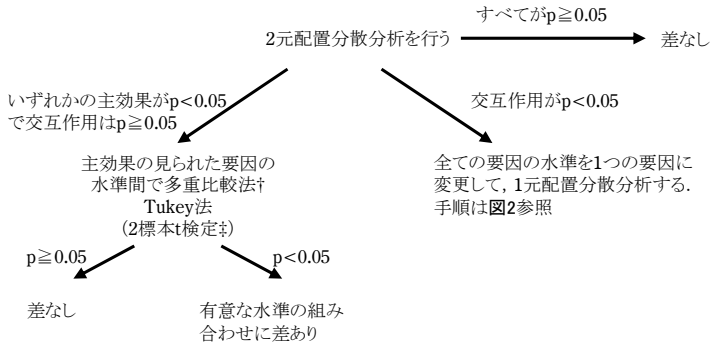
5. 検者間・検者内信頼性としての信頼性係数

検者内・検者間信頼性係数は，パラメトリックな手法と



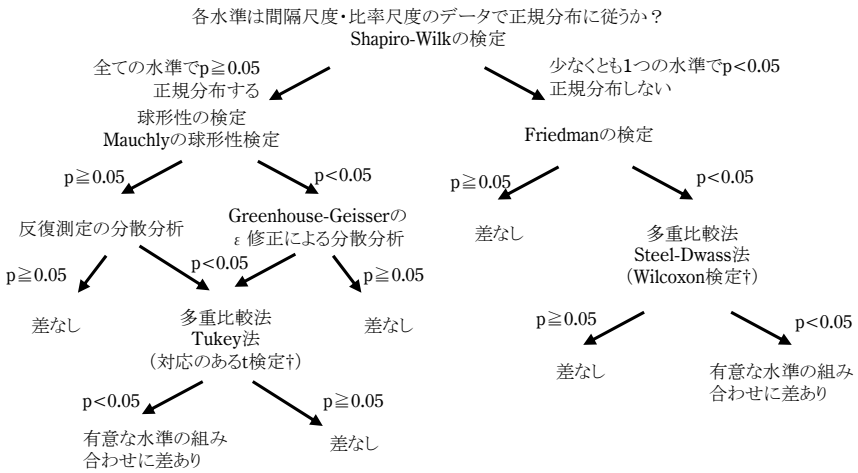
† Bonfferoni法またはShaffer法を行う場合

図 3　1 元配置デザインの解析手順



† 水準数が2つしかない要因は，分散分析の時点で終了
‡ Bonfferoni法またはShafferの方法を行う場合

図 4　2 元配置デザインの解析手順



† Bonfferoniの方法またはShafferの方法を行う場合

図 5　反復測定デザインの解析手順

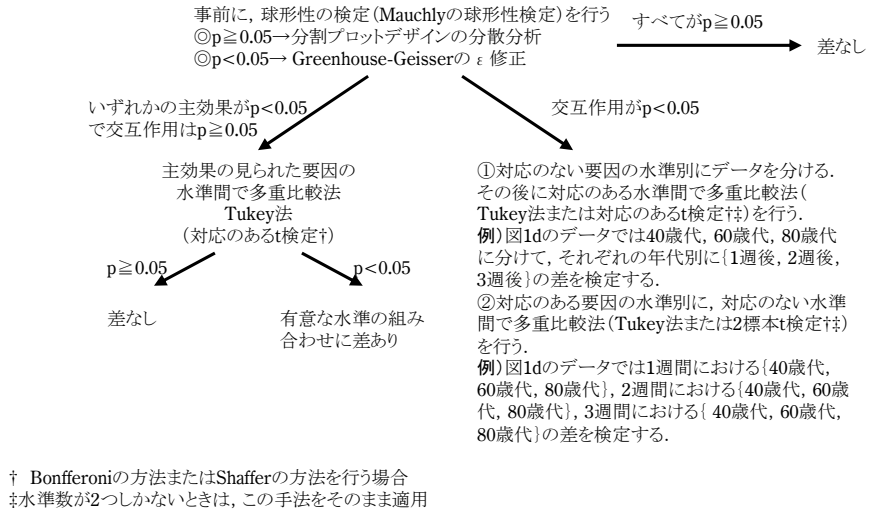


図 6　分割プロットデザインの解析手順

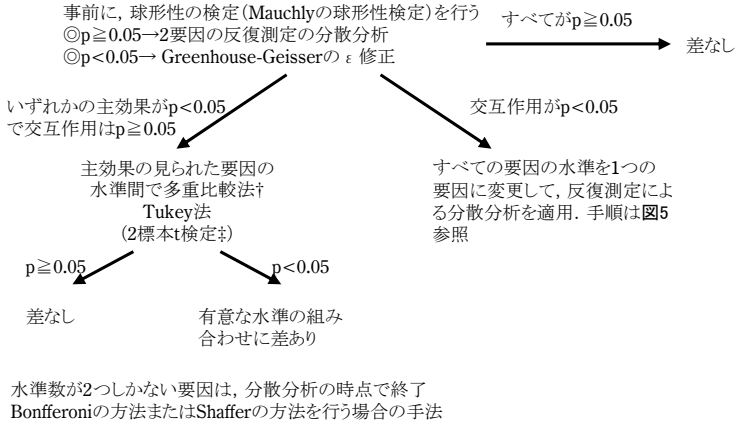


図 7　2 要因の反復測定デザインの解析手順

ノンパラメトリックな手法に分けられる．

パラメトリックな手法としては，ICC がある．ノンパラメトリックな手法としてはカッパ係数（κ 係数）がよく用いられる．

5.1 ICC の分類

ICC は，分散分析の理論を応用した信頼性係数であり，正規分布に従う比率・間隔尺度のデータに適用となる．

ICC は 0 ～ 1 の範囲をとり，相関係数のように 1 に近づくほど信頼性が高いと解釈する．まれに負の値をとることはあるが，そのときは 0 として考える．

ICC が高い値を示すときは，データ間の相関が高くて，かつ平均に差がないことを意味していると思えば良い．ゆえに，信頼性を調べるために，相関係数のみや，差の検定のみからの検討は不十分となる．ICC には様々なタイプがあるが，多くの統計ソフトにプログラムされて良く用いられるものは，Shrout ら [4] の分類に従った ICC である．その分類では大きく 3 つの公式に分けられる．

5.1.1 ICC(1，1)

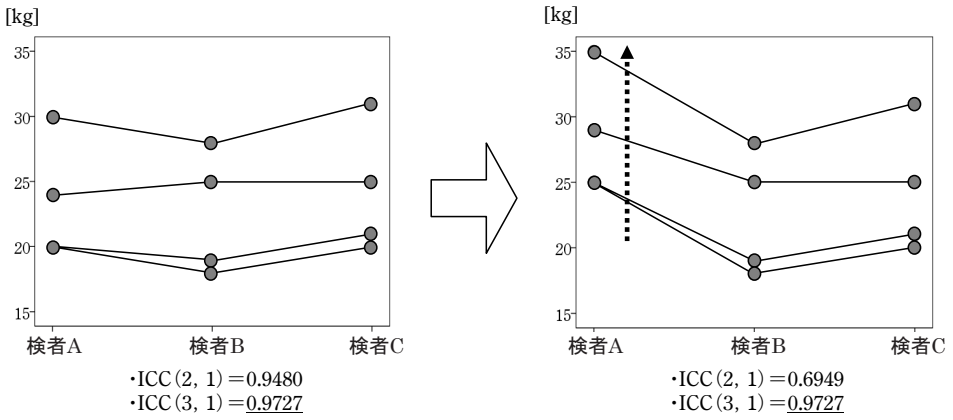
複数の被検者を対象として，1 人の検者が 2 回以上繰り返し測定したときの信頼性，すなわち検者内信頼性を知るためには ICC(1，1) を用いる．

5.1.2 ICC(2，1)

複数の被検者を対象として，2 名以上の検者が 1 回ずつ測定したときの信頼性，つまり検者間信頼性を知るために ICC(2，1) を用いる．

5.1.3 ICC(3，1)

ICC(3，1) は，複数の被検者を対象として，2 名以上の検者が 1 回ずつ測定したときの検者間信頼性を知るものである．検者間信頼性を知るという目的では ICC(2，1) と同様である．しかし，ICC(3，1) は，測定の精度というよりは整合性を確かめるものであり（図 8），リハビリテーションや医学の領域では使用することは少ないと思われる．手法の目的をよくわきまえた上で適用しなければならない．



a.表1bまたはcのデータ　　b.検者Aだけが全ての被検者で5kg大きく測定

図 8　ICC（2、1）と ICC（3、1）の比較

表 1b または c のデータで，ICC（2，1）と ICC（3，1）を求めた（a のグラフ）．次に，検者 A のデータだけ +5 となるように定数分大きくして（b のグラフ），再び ICC（2，1）と ICC（3，1）を求めた．ICC（2，1）は変化するが，ICC（3，1）は不変である．つまり ICC（3，1）は，検者間の絶対的な差は考慮せず，相対的な値の散らばり（いわば相関係数のようなもの）を表している．

表 1　握力を測定したデータ例

a.1人の検者が4人の被検者を3回繰り返し測定

		1回目	2回目	3回目
被検者	a	20	19	21
	b	24	25	25
	c	30	28	31
	d	20	18	20

→検者内信頼性：ICC(1, 1)の適用

b.3人の検者が4人の被検者を1回ずつ測定

		検者		
		A	B	C
被検者	a	20	19	21
	b	24	25	25
	c	30	28	31
	d	20	18	20

→検者間信頼性：ICC(2, 1)の適用

c.3人の検者が4人の被検者を1回ずつ測定

		検者		
		A	B	C
被検者	a	20	19	21
	b	24	25	25
	c	30	28	31
	d	20	18	20

→検者間信頼性：ICC(3, 1)の適用
※ただし，相関の概念だけで，検者どうしの平均差は無視

5.2 ICC を適用する際の注意

上述した ICC(1，1)，ICC(2，1)，ICC(3，1) は，測定法の信頼性を示す．高い信頼性とは，どれくらいかと問われると，現状では明確に回答できないが，諸家 [5,6] の意見をまとめると 0.7 ～ 0.8 以上を示したときに信頼性が高いと考えて良さそうである．

ところで，上述の ICC で求めた値は，それぞれ 1 人の検者で 1 回測定したときの信頼性を求めている．仮にある測定法 A が ICC(1，1) で 0.7 以上を示したときは，1 人の検者で 1 回測定すれば良い．もし 0.7 未満であったときは，信頼性を高めなければならない．信頼性の理論では，検者内信頼性を高めるために複数回測定した値の平均を使う方法と，検者間信頼性を高めるために複数の検者で測定した値を平均して使う方法を考える．

それでは一定の信頼性を保証するために，最低何回繰り返して測定した値の平均を使ったらよいとか，最低何名の検者で測定した平均を使ったらよいという具体的方法を述べよう．この計算にはスペアマン・ブラウンの公式，

$k = (\rho\ 1\ (1 - \rho\ 2)) / (\rho\ 2\ (1 - \rho\ 1))$

を用いる．ここで，ρ 1 は期待する ICC の値，ρ 2 は実際に求められた ICC 値である．

例えば，ある測定法の ICC(1，1) が 0.65（＝ρ 2）だったとする．0.65 では低いので，数回測定した平均を用いて，0.9（＝ρ 1）まで信頼性を高めたい．その際には，

$k = (0.9 \times (1 - 0.65)) / (0.65 \times (1 - 0.9)) = 4.846\cdots$

と計算され，4.8 回以上，実際的には 5 回以上繰り返して測定した平均をデータとすれば，0.9 以上の信頼性を確保出来るはずである．

ICC を解釈する上での大きな問題は，信頼性係数の範囲制約性である．ICC は，被検者の個人差が大きいデータでは，検者の個人差や誤差が相対的に小さくなって ICC が高くなる性質を持つ．例えば握力測定の ICC を求めるとき，握力の強い者から弱い者までを幅広く対象にすれば，ICC を高くすることが可能である．このような範囲制約性を考慮するためには，測定の標準誤差（standard error of measurement；以下，SEM）も参考にすると良い[7]．これだけだとイメージが湧き難いので，例を挙げてみよう．

表 2 は様々にデータを変えて ICC と SEM の値を求めた例である．表 2a のデータに対して被検者 a の値だけを定数分増加させると（表 2b），ICC は高くなる．これが範囲制約性である．しかし SEM は変化しない．表 2c は測定の 1 回目のみを定数分だけ増加させている．ここでは表 2a と比較して ICC（3，1）以外は低くなる．また，表 2a ～ c までは，データを系統的に変化させているので相対的な値のバラツキは一定であり，ゆえに SEM は変わらない．つまり，ICC が高い測定と低い測定を比較するとして，SEM に差がないときは，単に値の大きいまたは小さい被検者が存在するだけという可能性がある．ただし，表 2c のように特定の測定回の値が全体的に高い，つまり信頼性が低い場合もあり得るので，一概にはいえないことに注意する．

次に，被検者 a の 3 回目の値だけを段階的に増やしてみる（表 2d →表 2e）．これは信頼性を低くしていることになる．このとき表 2a と比較して，ICC は低く SEM は高く，または ICC は高く SEM は低くなっていく．この状態であれば信頼性が高い，低いと比較できる．

また，表 2e のデータ全体の単位を 1/10 に小さくした表 2f では，ICC は変化しないが SEM は小さくなる．データのバラツキが極めて小さくなると（表 2g），ICC は 1 となり，

表 2　様々にデータを変えたときの ICC と SEM の変化

a.表1と同じ握力のデータ

		1回目	2回目	3回目
被検者	a	20	19	21
	b	24	25	25
	c	30	28	31
	d	20	18	20

・ICC(1, 1)　ρ ＝0.9475
・ICC(2, 1)　ρ ＝0.9480
・ICC(3, 1)　ρ ＝0.9727
・SEM＝0.799

b.被検者aの値だけを+30kgにしたとき

		1回目	2回目	3回目	
被検者	a	50	49	51	←a+30kg
	b	24	25	25	
	c	30	28	31	
	d	20	18	20	

・ICC(1, 1)　ρ ＝0.9931
・ICC(2, 1)　ρ ＝0.9931
・ICC(3, 1)　ρ ＝0.9965
・SEM＝0.799

c.1回目のみ+30kg

		1回目	2回目	3回目
被検者	a	50	19	21
	b	54	25	25
	c	60	28	31
	d	50	18	20

↑1回目+30kg

・ICC(1, 1)　ρ ＝-0.3471
・ICC(2, 1)　ρ ＝0.0698
・ICC(3, 1)　ρ ＝0.9727
・SEM＝0.799

d．aに対して被検者aの3回目だけを+5としたとき

		1回目	2回目	3回目
被検者	a	20	19	26
	b	24	25	25
	c	30	28	31
	d	20	18	20

・ICC(1, 1)　ρ ＝0.8002
・ICC(2, 1)　ρ ＝0.8047
・ICC(3, 1)　ρ ＝0.8629
・SEM＝1.7321

e．aに対して被検者aの3回目だけを+10としたとき

		1回目	2回目	3回目
被検者	a	20	19	31
	b	24	25	25
	c	30	28	31
	d	20	18	20

・ICC(1, 1)　ρ ＝0.5394
・ICC(2, 1)　ρ ＝0.5539
・ICC(3, 1)　ρ ＝0.6116
・SEM＝3.0867

f．eのデータを1/10にしたとき

		1回目	2回目	3回目
被検者	a	2.0	1.9	3.1
	b	2.4	2.5	2.5
	c	3.0	2.8	3.1
	d	2.0	1.8	2.0

・ICC(1, 1)　ρ ＝0.5394
・ICC(2, 1)　ρ ＝0.5539
・ICC(3, 1)　ρ ＝0.6116
・SEM＝0.3087

g.値のバラツキが小さく信頼性の高い例

		1回目	2回目	3回目
被検者	a	30	30	30
	b	30	30	30
	c	30	30	30
	d	30.1	30.1	30.1

・ICC(1, 1)　ρ ＝1.0000
・ICC(2, 1)　ρ ＝1.0000
・ICC(3, 1)　ρ ＝1.0000
・SEM＝0.000

h.値のバラツキが小さいのに信頼性が低い例

		1回目	2回目	3回目
被検者	a	30	30.1	30
	b	30	30	30.1
	c	30	30.1	30
	d	30	30	30

・ICC(1, 1)　ρ ＝-0.2857
・ICC(2, 1)　ρ ＝-0.2857
・ICC(3, 1)　ρ ＝-0.2857
・SEM＝0.0500

SEM も 0 に近くなる．ただし，データのバラツキが極めて小さく SEM が小さくても，ICC はかなり低い値になってしまうこともある（表 2h）．こうしたことから，ICC の値だけを見て信用するのは危険であり，同時に SEM や生データを観察することも必要である．

5.3 カッパ係数

カッパ係数は，間隔・比尺度のデータが正規分布しないときや順序・名義尺度のデータに対して適用される信頼性係数である．この値も ICC と同様に 0 ～ 1 の範囲をとり，0.7 ～ 0.8 以上を示したときに信頼性が高いと考えて良い．検者間信頼性も検者内信頼性も，同じカッパ係数で計算する． カッパ係数に関しても，係数値だけを見て高低を検討するには限界があり，ICC と同様に生データも観察しておくことが必要である．

5.4 検者内・検者間信頼性係数の活用について

信頼性係数の適用上の注意点については，上述の他にも様々あるので，文献[9]も参考とされたい．

ICC の SEM は統計ソフトで求められないことが多い．また，カッパ係数については，統計ソフトによっては 3 回以上の繰り返し測定に対する検者内信頼性や，3 名以上の検者間信頼性は求めることができないものもある．その際には，フリーソフト R の利用を薦める．R については，著者の web[8]でも紹介しているので参考にして頂きたい．

参考文献

1) 対馬栄輝：理学療法の研究における多重比較法の適用について，東北理学療法学，13，30-37，（2001）．
2) 永田靖，吉田道弘：統計的多重比較法の基礎，サイエンティスト社，（1997）．
3) http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/text.html
4) Shrout PE,Fleiss JL：Intraclass correlations:uses in assessing rater reliability，Psychological Bulletin,86，420-428，（1979）．
5) Portney LG，Watkins MP：Foundations of clinical research-Applications to practice-，505-516，Appleton & Lange，USA，（1993）．
6) 桑原洋一，斎藤俊弘，稲垣義明：検者内および検者間の Reliability（再現性，信頼性）の検討，呼と循 41，945-952，（1993）．
7) Stratford PW, Goldsmith CH：Use of the error as a reliability index of interest: an applied example using elbow flexor strength data，Phys Ther，77，745-750，（1997）．
8) http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/S/
9) 対馬栄輝：理学療法の研究における信頼性係数の適用について，理学療法科学，17(3)，181-187，（2002）．