愛知学院大学心身科学部紀要第12号（101–112）（2016）

〈総説〉

社会行動科学や医学における統計学の誤用・乱用について

千　野　直　仁*

社会行動科学や医学などでは，統計学の誤用・乱用がしばしば起こる．最初に，我々は外れ値や異常値が存在するときのピアソンの偏差積率相関係数の誤用・乱用について議論する． 2 つ目に我々は分散の等質性の検定に続く通常の平均の差の検定の適用時の誤用・乱用について議論する. 3 つ目は，我々が無作為割り付けができないような幾つかの処理水準から成る被験者間要因の効果を分散分析により検定するときの誤用・乱用について述べる． 4 つ目は，複数の平均の差についての全体的仮説が棄却された場合の後続の多くの対比の検定における誤用・乱用について述べる．同じような状況は，要因デザインで全体的交互作用仮説が棄却されたときにも起こる． 5 つ目に我々は，反復測定デザインにおける要因の効果に関する検定の誤用・乱用について述べる．大局的球形仮説や多標本球形仮説の検定は，これらの帰無仮説が棄却されたとき反復測定要因に対応する幾つかの仮説に対応する F 分布がゆがむ可能性があるが，しばしばなされないままである． 6 つ目に我々は因子分析を実際のデータに対して適用するに際しての誤用・乱用について議論する． 7 つ目に我々は重回帰分析やロジスティック回帰分析などの一般化線形モデルの適用に際しての誤用・乱用について述べる．とりわけ，多重共線性の問題は，もし考察中のデータで多重共線性が起これぱそれらのモデルのパラメータ推定値は不安定になるが，しばしばなされないままである．

キーワード：統計学の誤用，多重共線性，外れ値，球形（球面性）仮説

第 1 節　問題

得られたデータに対して何らかの統計学的推定や検定を行うことは，特定の分野を除き社会行動科学の領域から医学・疫学などの自然科学の領域で論文を書く研究者や学生にとってきわめて日常的な活動である．しかし，すべての統計的推定や検定には何らかの前提がある．これらの前提が万が一満たされていないと，推定や検定により得られた結論は全く異なったものとなることがある．この論文では，それらのうち多くの研究者や学生が日常的に使っている統計的推定や検定に見られる誤用・乱用の幾つかをまとめることが目的である．

それらは相関係数の検定，対応のない場合の平均の差の検定，通常の分散分析における無作為割り付けの有無や主効果が棄却された場合の下位検定のやり方や交互作用項の検定方法，反復測定デザイン分散分析，因子分析おける直交解・斜交解の選択や直交因子の因子得点間相関の可能性，重回帰分析やロジスティック回帰分析における多重共線性の問題である．

相関係数の検定に際して注意しなければいけない点の 1 つは, データに**外れ値**（outlier）や**異常値**（abnormal value）が混入している場合である．万が一これらが存在する場合，それを外すかどうかにより検定結果は全く異なることがある．

対応のない場合の平均の差の検定では，通常検定はまず 2 群の分散の等質性の検定を行い，その結果を受けて平均の差の検定方法として適切な統計量を選択する．この場合，ほとんどの統計学の教科書では2 種類の検定を継時的に行う場合の**全体的危険率**（overall level of significance）については無視をして，それぞれの検定を例えば 5 パーセント水準で行うように書か

＊愛知学院大学心身科学部心理学科
（連絡先）〒470–0195 愛知県日進市岩崎町阿良池12　E-mail: chino@dpc.agu.ac.jp

れているが，そのような場合両検定の全体的危険率は少なくとも分散の等質性が成り立つ場合にはインフレをしているので，それを統制するためには正確には個々の検定の危険率は5パーセントでは正しくない（Hogg, 1961; 竹内，1978）．

通常の分散分析では幾つかの点で，誤用・乱用がしばしばみられる．その1つは，本来**無作為割り付け**（random assignment of plots, or randomization）（Fisher, 1926, 1935）ができないような要因，例えば出生順位，学年, 心理検査により判定される患者の特性（例えば，CMIにより得られる患者の神経症傾向に関する4群）などについて，しばしば内外の論文には通常の分散分析を行い要因の効果の有無について考察がなされているが，厳密にはそのような場合，要因の効果であるというのは適切ではなく，高々要因の水準間で平均に差が見られたということにとどめるべきであろう．

通常の分散分析における2つめの誤用・乱用は，主効果や交互作用が有意であったときの下位検定としての**多重比較**（multiple comparison）の方法に見られる．内外の多くの論文では，例えば主効果が有意の場合の下位検定として，**ボンフェロニの手続き**（Bonferroni procedure）（正式名は，**ダンの多重比較手続き**，Dunn's multiple comparison procedure）（Dunn, 1961）を行ったとか, **テューキー法**（Tukey's procedure）（Tukey, 1953）を用いたとあるが，正しくはデータを収集する際に研究者がどのレベルの下位検定の仮説を持っていたかにより下位検定の手続きや方法の中から最適なものを選択する必要がある．

一方，全体的交互作用が有意であった場合，やはり内外の多くの論文では未だに下位検定として**単純主効果**（simple main effects）の検定が行われているが，単純主効果の検定は厳密には全体的交互作用の正確な下位検定にはなっていない（例えば，Kirk, 1982, 1995）．

反復測定デザイン分散分析（repeated measurement design ANOVA）も社会行動科学や医学・疫学などの領域でもよく用いられる方法の1つであるが，この場合の要因の効果の有無の検定のF統計量はデータが多次元空間の中に等方的に広がっているかどうか（**球形あるいは球面性仮説**，sphericity hypothesis）により大きく歪むことが1970年以降知られている（例えば，Huynh & Feldt, 1970; Rouanet & Lépine; 1970）にもかかわらず，この検討がなされていない場合が内外の論文で少なからず見られる．

とりわけ，2要因以上が共に反復測定要因の場合の**大局的球形（球面性）仮説**（global sphericity hypothesis）の検討（Kirk, 1982; Mendoza, Toothaker, & Crain, 1976）や，独立測定要因（実験参加者間要因）と反復測定要因（実験参加者内要因）が混在する分割区画要因デザインでの**多標本球形（球面性）仮説**（multisample sphericity hypothesis）の検討（Harris, 1984; Huynh & Feldt, 1970; Kirk, 1982, Mendoza, 1980; Sugiura & Nagao, 1968），については本邦の論文はおろか欧米のインパクトファクターから見て一流の論文でさえ，論文中にこれらが言及されていないものもあるし，国際的な統計ソフトであるSASやSPSSでさえ，研究者がそれらのプログラムやシンタックスを自分で新たに書かなければ，検討できないのが現状である．

因子分析は，当初統計学者と（計量）心理学者が1900年代の初頭から提案し発展させてきたものであり（Garnett, 1919−20; Holzinger & Harman, 1937; Spearman, 1904, Thurstone, 1931），近年では心理学の領域にとどまらず多くの社会行動科学の領域から医学・疫学などの自然科学の領域においても用いられている方法である．しかし, これについても, 幾つかの誤用・乱用が見られる．まず, 内外の研究論文を見ていると，現象の性質や理論を顧みず, いつも直交解を用いたり，いつも斜交解を用いたりする研究がしばしばみられるが，かならずしも適切ではない．**直交解**（orthogonal solution）（Hotelling, 1933）か**斜交解**（oblique solution）（例えば，Carroll, 1953; Hendrickson & White, 1964; Holzinger, 1944; Kestelman, 1952; Pinzka & Saunders, 1954; Thurstone, 1945）かの選択はあくまでも現象の性質や理論などから研究者が適切な解法を選択すべきである．

因子分析の教科書にはあまり書かれていない問題に，複数の**回転後の直交共通因子の因子得点間相関**（intercorrelations among rotated orthogonal common factors）に関する問題がある（Guilford & Michael, 1948; Heermann, 1963; Holzinger & Harman, 1941; Thomson, 1948）．因子分析の定番である主因子法による主因子解, それに引き続いてのバリマックス回転は，内外の多くの論文がいまだよく用いているやり方であるが，主因子解とそれに続くバリマックス回転により得られる解はあくまでも直交因子であり，得られる複数の共通因子得点間には定義からは本来無相関であるにもかかわらず，内外の論文ではしばしばそのような因子得点間の相関係数を求め，さらに検定を行い有意である場合懸命にそれらの解釈を試みているものをしばしば目にする（例えば，Fenigstein, 1979; 菅原，1984）．

複数の直交共通因子相互の因子得点間に相関が生じる場合，それらの相関係数の有意性検定を行うと有意である原因は少なくとも 2 通り考えられる．1 つは，そのような場合における因子得点の求め方に問題がある場合であろう．内外の論文でも，因子得点を各因子に負荷の高い項目に絞って，それらの合計点を因子得点の代用物として用いている場合がしばしばみられる（例えば，Fenigstein, 1979; 菅原，1984）が，このようなやり方は決して十分適切とは言えない．この方法では，項目間の相関の情報を一律に扱っている点が不適切である．

もう 1 つの原因は，因子分析モデルそのものに根差したもので，かなり専門的であるが，実は因子分析の方法開発の歴史の中では古くから知られている（例えば，Harman, 1967; Heerman, 1963; Kesterman, 1952; Thomson, 1948; Thurstone, 1935）．また，その中で，Heerman (1963) は直交共通因子間相関が真の意味で直交（無相関）になるような因子得点の推定法などを示している．もっとも，統計ソフトの SAS や SPSS でも Heerman の方法は組み込まれていない．

統計学的方法の誤用・乱用とまでは言えないとしても，しばしば見落とされるのが重回帰分析やロジスティック回帰分析における**共線性**（collinearity）あるいは**多重共線性**（multicollinearity）の問題である（Agresti, 2002; Allison, 2012; Belsley et al., 1980; Tabachnick et al., 2014）．これらの方法の共通点は，何らかの基準となる変数（基準変数，反応変数，従属変数とも呼ばれる）を，それを説明ないし予測するための複数の変数（説明変数，予測変数，独立変数とも呼ばれる）で説明ないし予測しようとする点である．ただし，これらの方法では基準変数は，重回帰分析の場合定量的変数，ロジスティック回帰分析の場合定性的変数(あるいは群)でなければならない．

これらの方法では，説明変数が多くなると同変数間の複雑な相関関係からいわゆる共線性あるいは多重共線性の問題が発生し，偏回帰係数などのパラメータの推定が適切に行われなかったりする可能性が高まる．この問題を避けるには，これらの方法の最終適用に先立ち，説明変数間に共線性が見られないかどうか検討し，もし強い共線性が見られる場合には，あらかじめ共線性を生じる原因となる変数を可能な範囲で除去する必要がある．

最後に，従来の伝統的な多変量解析では誤差分布として正規分布が仮定されてきたが，1970代の初頭にはこの仮定を必ずしも必要としない場合の多くの多変量解析が開発されてきており（McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972），**一般化線形モデル**（generalized linear model）と呼ばれる．これらは一般にモデルの誤差分布に必ずしも正規分布を仮定しないだけでなく，扱う変数も必ずしも定量的変数のみではなく，定性的変数と定量的変数が混在することを許容するもので，社会行動科学領域から医学・疫学領域などの自然科学領域までの多様な現象に対する適用範囲を大きく広げるものである．例えば，重回帰分析で万が一モデルの誤差分布が正規分布から大きく乖離していると判断される場合には，通常の重回帰分析は行わず，一般化線形モデルの中の適切と思われるモデルにより分析を行うべきであろう．

この論文の構成は次のようである. まず第 2 節では，相関係数の検定時の誤用・乱用について述べる．第 3 節では，対応のない場合の平均の差の検定に際して分散の等質性の検定に引き続き平均の差の検定を行う場合の全体的危険率のコントロールの問題について述べる．第 4 節では，通常の分散分析における無作為割り付けの有無と要因の効果の検定の是非，主効果や交互作用効果が有意な場合の下位検定の問題について述べる．第 5 節では，反復測定デザイン分散分析における F 統計量の歪みの可能性とそれに対処するための方法について述べる．第 6 節では，因子分析における回転の問題や，直交共通因子の因子得点間の相関の可能性の問題に触れる．第 7 節では，重回帰分析やロジスティック回帰分析における多重共線性など分析上の注意点ついて触れる．

第２節　相関係数の検定時の誤用・乱用

図 1 は，浮田・横井（1996）の48名の実験参加者に対するクレペリン精神作業検査の結果における休憩後の動揺率と誤答率間の散布図である．この図で A，B はその位置に実験参加者がそれぞれ 1 名，および 2 名いることを表す．この図をみると，明らかに 1 名の実験参加者の動揺率と誤答率の大きさは共に他の成員とかけ離れていることがわかる．この検査の時点ではその原因がわからなかったため, 外れ値とみなされた．

このデータで外れ値に気づかず，両変数間の相関係数を計算すると 0.692 となり，相関係数は 0.1 パーセント水準で統計的に有意とみなされる．しかし，一般に外れ値を見つけた場合には，直ちに外して済ませるのではなく，飛び離れた値になった原因を探るなり，サンプル数を増やすなりして様子をみることが望まれ

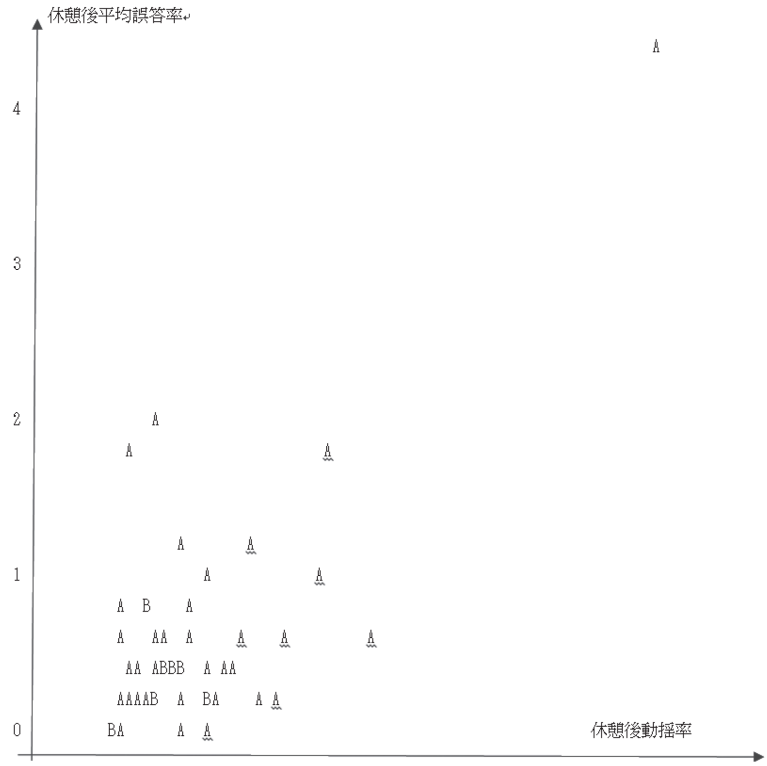

図１．クレペリン精神作業検査における休憩後動揺率と同平均誤答率との散布

る．一方，飛び離れた値となった原因が入力ミスなどはっきりしている場合は，そのような値は外れ値と呼ばず**異常値**（abnormal value）と呼ぶ．実際のデータ解析の場面では異常値に気づかずに相関係数の検定をおこなってしまうことが多いと思われ，その場合にも上記のようなことが起こるので，できれば相関係数の検定に先立ち，当該変数間の散布図を描くことが望まれる．

相関係数については，外れ値，異常値の問題だけでなく，相関係数はあくまでも両変数間の直線的な関係の有無のみを検討するためのものであることに注意が必要である．実際のデータでは時々U字型や逆U字型の散布図が得られることがある．また場合によっては，ドーナツ型の帯状の領域のみにデータが広がっているような散布図が得られることも理論的には可能である．このような場合，相関係数はゼロに近い値，すなわち無相関となるが，当該科学の領域，例えば心理学的にはそのような散布図には実質科学的な意味がある可能性もあろう．そのような２変量間の非線形な関係は，散布図を描かないと見えてこない．ただし，相関係数の検定の前提である，データの分布が２変量正規分布に従っていれば，そのような非線形な関係が得られる可能性は非常に少ないと考えられる．

第３節　対応のない場合の平均の差の検定に際しての全体的危険率の誤用・乱用

第１節で簡単に紹介したように，対応のない場合の平均の差の検定では，通常検定はまず２群の分散の等質性の検定を行い，その結果を受けて平均の差の検定方法として適切な統計量を選択することはよく知られているが，ほとんどの統計学の教科書では２種類の検定を継時的に行う場合の全体的危険率のインフレについて言及していない．筆者の知る限り，これに言及している研究として例えば，Hogg (1961)，永田（2008），および竹内（1978）がある．

なお，千野（2010）は本紀要第６号でこの問題を詳しく紹介しているので，ここでは，以下にそれをかいつまんで紹介するにとどめる．まず，Hogg (1961) によれば，分散の等質性仮説が採択される場合には，等質性の検定も平均の差の検定も，例えば全体的危険率 α^* を0.05にしたければ，個々の検定の危険率をおよそ0.025に，もし α^* を0.01にしたければ，個々の危険率をおよそ0.005に取る必要がある．その理由は，

Hogg が証明しているように，分散の等質性仮説が採択される場合にはその場合の統計量 F と平均の差の検定のための t 統計量が独立となることを用いて，全体的危険率を正確に計算できるからである．

一方，竹内（1978, p. 19）は「ここで，二つの検定の水準は，仮説 $\mu_1=\mu_2$，$\sigma_1=\sigma_2$ の下で，二つの検定でともに仮説が捨てられない確率が，ちょうど $1-\alpha$ になるように定めなければならない」と述べている．竹内はこの知見が誰によるものかを明示していないが，もちろん，その根拠は Hogg (1961) にあることは文献上明らかである．

これに対して，永田（2008, p. 100）は，Hogg の結果を引用しておらず，「この前段階の F 検定は有意水準20％で行う」としている．その理由として，彼は当該 F 検定における第 2 種の過誤を小さくすることをあげている．ただし，この方式では，例えば第 2 段階の平均の差の検定における危険率を 5 ％とすると，2 種類の検定の全体の危険率は，等分散仮説が採択される場合には，およそ24％と大きくインフレすることになる．すなわち，永田方式では，第 1 段階の F 検定における第 2 種の過誤を小さくすることにより，結果的には上記 2 段階で行う全体的検定における第 1 種の過誤は相対的には大きくなってしまうのである．

第４節　通常の分散分析における誤用・乱用

通常の分散分析における誤用・乱用の 1 つは，第 1 節で述べたように，本来無作為割り付けができないような要因について，内外の論文でしばしば通常の分散分析を行い要因の効果の有無について考察がなされることである．その理由を明らかにするためには，**Fisher の原則**（Fisher’s principles）（とりわけ，**Fisher の実験計画法の３原則**）に注意する必要がある（Fisher, 1926, 1935）．Fisher の実験計画法の 3 原則とは，つぎを指す：

1．**反復**（repetition）
同一水準には 2 回以上の標本の反復（繰り返し）が必要である．
2．**無作為化**（randomization, or random assignment of plots）
標本の各水準への割り付けは，無作為化しなければならない．
3．**局所管理**（local control）
標本全体の均一化を達成することが難しい時に，それにかかわる副次的な因子に対しては複数の水準を設定し，その水準内では局所的に均一化を達成させるとよい（これは，通常ブロック因子を導入することにより達成される）．

これらのうちの第 2 原則が，無作為化の原理である．これを達成するための最も単純なデザインは完全無作為化デザインであり，2 要因以上の場合完全無作為化要因デザインと呼ばれ，心理学の領域等ではこれまで標準的なデザインとしてよく知られている．

いずれにせよ，第 1 節で指摘した出生順位，学年，心理検査によるクライアントの判定結果などは，分散分析を行う時点で既に実験参加者の属性としてどの水準に属するかが決まってしまっており，無作為割り付けができない．そのような場合，分散分析により F 統計量から当該要因に対する帰無仮説が棄却されたとしても，その原因を要因の効果と判定することは慎重になるべきである．なぜならば，そのような場合，各水準の平均の違いには，たまたま実験に参加した者が当該要因以外の要因に関して特定の水準に偏っていた可能性を十分排除できないからである．無作為割り付けは，そのような可能性を可能な限り排除するために工夫された仕掛けである．そのような工夫ができない場合，我々が言えることは高々当該要因の水準間のどこかに差が見られた，ということである．

つぎに，分散分析で主効果や交互作用が有意である場合の下位検定のやり方について内外の多くの論文では，例えば主効果が有意の時，実験の前に研究者がどのような下位検定を想定していたかにかかわらず，ボンフェルニの手続きを行ったとか，テューキー法を用いたとしているが，例えば，ボンフェルニの手続きはそのような下位検定が実験の前にきちんと計画されており，なおかつ複数の対比（contrasts）が直交していないものを含む場合，テューキー法は当該下位検定の対比が実験の前にはあらかじめ計画されておらず，複数の対比が必ずしも直交していない場合に適用すべきものである．

もう 1 つの問題点は，全体的交互作用が統計的に有意であるときの下位検定のやり方に関するものである．内外の多くの研究でも，よく用いられているのが単純主効果の検定である．この検定は，2 要因の交互作用に限定したとして，一方の要因の各水準ごとに他方の要因の主効果の有無を検討する方法である．しかし，このやり方では，F 統計量には交互作用項だけではなく主効果の効果も混入してしまっており，厳密な下位検定とは言えない．というのは，単純主効果の効

果に関わる平方和は，交互作用平方和と他方の要因の主効果の平方和の双方から成り立っているからである（Kirk, 1982, pp. 366–367）．正確な全体的交互作用の下位検定には，**処理・対比交互作用**（treatment-contrast interaction）や**対比・対比交互作用**（contrast-contrast interaction）の検定（Kirk, 1982, 1995）がなされるべきであろう．

第５節　反復測定デザイン分散分析における検定の誤用・乱用

まず，最も単純な 1 要因反復測定デザイン分散分析における F 統計量の分布の歪みの可能性について図示したのが，図 2 である．ここでは，球形仮説が成り立つ場合，すなわち F に歪みがない場合の自由度 $v_1=12$，$v_2=27$ の F 分布を青色のカーブで，球形仮説が成り立たない場合の歪みのある F 分布を赤色のカーブで示している．仮にこの場合の F 値が2.2であったとしたときの両統計量における p 値（赤色及び青色で塗りつぶした領域の面積）を比較すると明らかなように，球形仮説が成り立っていないときには p 値は大きくインフレする可能性が見て取れる．もし, 球形（球面性）検定を行わずこのようなインフレを無視して結論を下すと，場合によっては要因の主効果の関する帰無仮説は採択ではなく棄却されてしまうようなことも起こりうる．

同様なことが，複数の要因が共に反復測定要因の場合には大局的球形仮説の検討を怠ると，また独立測定要因と反復測定要因が混在する場合には多標本球形仮説の検討を怠ると，起こりうる．とりわけ，２要因以上の反復測定要因デザインの場合，SAS や SPSS でさえ，プログラムやシンタックスを追加しないと，これらの検定を出力しないことも災いしてか，内外のレベルの高い論文でさえ，この種の検討をしていない場合が多い．これらの概要については, 例えば繁桝・柳井・森（1999, 2008）の反復測定デザインに関する項（Q39, Q40, Q41）を，詳細については千野（1993, 1994, 1995）を参照されたい．

第６節　因子分析における統計の誤用・乱用

因子分析における統計の誤用・乱用について述べる

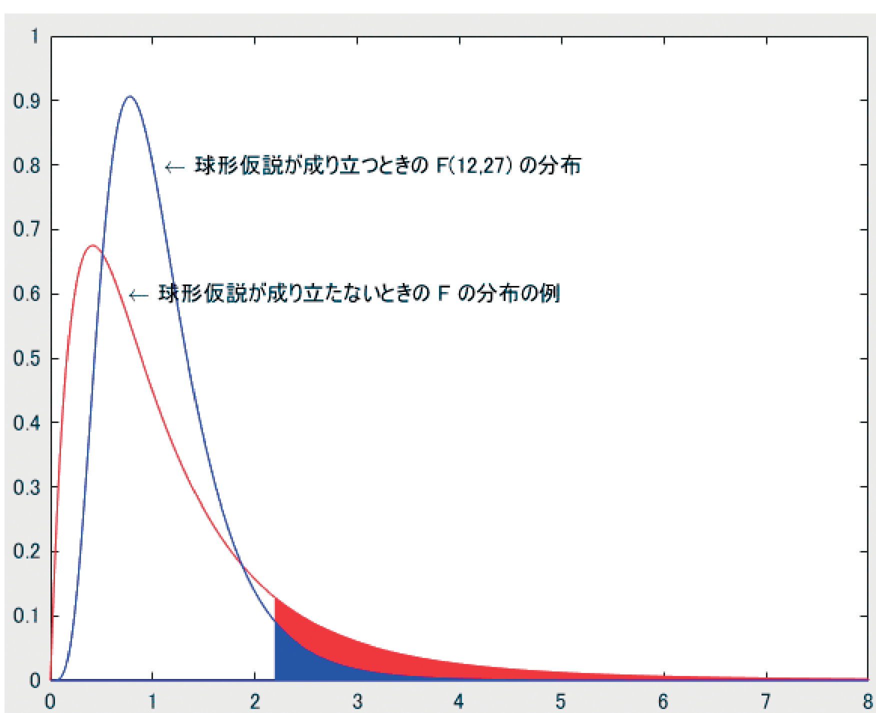

図 2．球形（球面性）仮説の成否による F 統計量の分布の歪みの有無

（図中，青色のカーブは球形仮説が成り立つ場合，赤色のカーブは成り立たない場合を示す．また，図中塗りつぶされた領域の左端に対応する横軸の値は，仮定された F 値で2.2である）

前に，**主成分**（principal components）あるいは**成分分析**（component analysis）（Hotelling, 1933; Pearson, 1901）と，伝統的な**探索的因子分析**（exploratory factor analysis）についての**基本的因子分析モデル**（basic factor analysis model）（あるいは，サーストンの**多因子模型**，Thurston’s multiple factor model）（Thurstone, 1931）の紹介を行う．

まず，主成分分析モデルは，m変数のN人からなるデータが得られているとして，第 i サンプルの第 j 変数の得点を z_{ij} とするとき，

$$z_{ij} = a_{j1}F_{i1} + a_{j2}F_{i2} + \cdots + a_{jm}F_{im}, \quad \mathrm{i} = 1,2,\ldots,\mathrm{N},\ \ \mathrm{j} = 1,2,\ldots,\mathrm{m}. \tag{1}$$

ここで，成分分析では観測得点はm個の成分 $F_{i1}, F_{i2}, \ldots, F_{im}$ およびそれに対する重み $a_{j1}, a_{j2}, \ldots, a_{jm}$（成分ウエイト，あるいは因子負荷量）から成り立つと仮定する．モデルでは成分は各観測変数の数，すなわちm個仮定しているが，実用的な観点からはm変数の得点の分散を可能なかぎり多く説明できるような少数の成分のみを考える．成分分析では，観測得点 z_{ij} は素点を用いる場合と基準化得点（平均ゼロ，分散１）のいずれかを用いる．素点を用いる場合は，全変数の関連は**共分散行列**（covariance matrix）**C**で，基準化得点を用いる場合は，同関連は**相関行列R**（correlation matrix）で，表される．

m個の成分と成分ウエイトは，行列CまたはRの**固有値問題**（eigenvalue problem）として代数的に解くことができ，CまたはRの性質からm個の成分間は相互に無相関（幾何学的には**直交**，orthogonal）である．

一方，因子分析の基本モデルは，

$$z_{ij} = a_{j1}F_{i1} + a_{j2}F_{i2} + \cdots + a_{jm}F_{ir} + b_jU_{ij}, \quad \mathrm{i} = 1,2,\ldots,\mathrm{N},\ \ \mathrm{j} = 1,2,\ldots,\mathrm{m}. \tag{2}$$

である．ここで，因子分析では観測得点は r ＋ 1 個の因子の値あるいは因子得点（factor scores）$F_{i1}, F_{i2}, \ldots, F_{ir}, U_{ij}$ およびそれに対する重み $a_{j1}, a_{j2}, \ldots, a_{jm}, b_j$（因子負荷量）から成り立つと仮定する．因子分析では2種類の因子を仮定しており，最初の r 因子は**共通因子**（common factors），最後の 1 因子は**独自因子**（unique factor）と呼ばれる．したがって，より正確には $F_{i1}, F_{i2}, \ldots, F_{ir}$ は共通因子得点，$a_{j1}, a_{j2}, \ldots, a_{jm}$ は共通因子負荷量と呼ばれ，$U_{ij}\ (j = 1,2,\ldots,m)$ は独自因子得点，$b_j\ (j = 1,2,\ldots,m)$ は独自因子負荷量と呼ばれる．独自因子はさらに**特殊因子**（specific factor）と誤差に分解されることもある．共通因子得点間は無相関（幾何学的には直交, orthogonal）であっても相関があっても（幾何学的には**斜交**，oblique factors）よい．一方，独自因子得点間は無相関が仮定される．また，共通因子得点と独自因子得点間も無相関が仮定される．(2) 式のモデルは**確認的因子分析**（confirmatory factor analysis），別名**共分散構造分析**（covariance structure analysis あるいは structural equation model, **SEM**）と区別する意味では探索的因子分析あるいは**古典的因子分析**（classical factor analysis）とも呼ばれる．

古典的因子分析の解法には最小 2 乗法を用いる主因子解と最尤法を用いる最尤解が知られている．主因子法による主因子解は，**因子の寄与**（contribution of factor）が与えられたデータの条件下で最大になるような因子を少数個（ r 個）抽出する．一方, 最尤解は, モデルの誤差分布を仮定し, データの**尤度**（likelihood）が最大になるような因子を少数個（ r 個）抽出する．

成分分析と因子分析は，うえに述べたように共通点が多く，成分分析における成分ウエイトはしばしば因子負荷量と呼ばれ，因子分析と混同される場合が多いが，独自因子，特殊因子などを含むかどうかで大きく異なる．

うえの成分分析および因子分析モデルの議論から明らかなように，因子分析モデルからは共通因子間では直交因子の場合因子間相関はすべてゼロであるが，実際にデータを手にして主因子解，バリマックス回転を行い (2) 式のモデルの共通因子負荷量を求め，つぎに共通因子の因子得点 $F_{i1}, F_{i2}, \ldots, F_{ir}$ を第 1 章でふれた簡便法ではなく**最小２乗推定量**（least-squares estimators）あるいは**重回帰推定量**（multiple-regression estimators）により推定しても，成分分析の場合の主成分得点とは異なり，共通因子の因子得点間は，必ずしも無相関とはならないことが，第 1 章で指摘したように古くから知られている．

ただし, この問題に対処するための方法については, Heermann (1963) が既に 1，2 提案している．１つの方法は，直交共通因子間をすべて無相関にする方法であるが，あらかじめ仮定した真の共通因子得点とその推定量間に必ずしも一対一の対応が得られない．一方，もう 1 つの方法では，真の共通因子得点とその推定量間に一対一の対応が得られるが，直交共通因子推定量間は必ずしも無相関にはならない．

いずれにせよ，Heermann の研究結果からは，因子分析でわれわれが直交共通因子をデータから推定する場合には，モデルの性質から（成分分析とは異なり）完全無欠な直交共通因子得点を推定することはむずか

しいことがわかる．しかし，だからといって直交共通因子得点の推定値間の相関係数の検定を行って相関関係が有意であったとしても，内外の研究の一部でおこなわれているような心理学的解釈まで無理に行うことは，避けるべきであろう．もっとも，筆者の経験ではSAS を用いて直交共通因子得点間の相関係数を検定しても統計的に有意になることはまれである．たぶん，内外の文献上でこれらの幾つかが有意となる場合，因子得点の推定法として最小 2 乗推定量などを用いるのではなく，第 1 章で述べた簡便法を用いていることによる場合が多いのではなかろうか．

第７節　重回帰分析やロジスティック回帰分析における分析上の注意点

まず，重回帰分析では，基準となる定量的変数（**基準変数**，criterion variable; **反応変数**，response variable; **従属変数**，dependent variables などと呼ばれる）を Y_i，これを説明ないし予測するための m 個の定量的変数 $x_{i1}, x_{i2}, \dots, x_{im}$（**説明変数**，explanatory variables; **予測変数**，predictor variables; **独立変数**，independent variables）でもって説明ないし予測しようとするもので，つぎのようなモデルを用いる：

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} + E_i, \quad \mathrm{i} = 1, 2, \dots, \mathrm{N}. \tag{3}$$

ここで，通常の重回帰分析では (3) 式の大文字で表記した変数 Y_i および E_i は確率変数を，また小文字で表記した変数は定数を表す．一方，β_0 及び各説明変数の重み係数 $\beta_1, \beta_1, \dots, \beta_m$ は，$\mathrm{m} \geqq 2$ の時，**偏回帰係数**（partial regression coefficients）と呼ばれる．また，E_i は**誤差項**（error terms）で，平均ゼロ，分散 σ^2 の正規分布に従うと仮定される．N はサンプル数である．

ここで，Y_i の実現値 y_i を N 個観測したとする．このとき，データの組 $(y_i, x_{i1}, x_{i2}, \dots, x_{im})$，$\mathrm{i} = 1, 2, \dots, \mathrm{N}$ を観測したものとして，サンプル全体にわたり誤差を最小にするような偏回帰係数の推定値は，行列を用いればつぎのように書ける：

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{X}^t y_B. \tag{4}$$

ここで，$\widehat{\boldsymbol{\beta}}$ は偏回帰係数ベクトル $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)^t$ の推定値 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$ を要素とするベクトルで，説明変数のデータ行列 $\boldsymbol{X}$ と，基準変数の N 個のサンプルの観測値のそれぞれからその平均を引いたベクトル $y_B = (y_1 - \bar{y},\ y_2 - \bar{y},\ \dots,\ y_N - \bar{y})^t$ により計算される．$\bar{y}$ は，N 個の基準変数の観測値の平均を表す．一方，N 行 m 列の行列 $\boldsymbol{X}$ は各サンプルの m 個の説明変数の値のそれぞれから各変数の平均を引いた値からなる行列であり，つぎのように書ける：

$$\boldsymbol{X} = \begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1m} - \bar{x}_m \\ \vdots & \ddots & \vdots \\ x_{N1} - \bar{x}_1 & \cdots & x_{Nm} - \bar{x}_m \end{pmatrix}. \tag{5}$$

また，モデルの定数項 β_0 の推定値 $\hat{\beta}_0$ は，$\beta_1, \beta_2, \dots, \beta_m$ の推定値 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$ が決まると，それらを用いてつぎのように書ける：

$$\hat{\beta}_0 = \bar{y} - \textstyle\sum_{j=1}^{m} \hat{\beta}_j \bar{x}_j. \tag{6}$$

ここで，**多重共線性**或いは**共線性**に対する正確な定義は，Belsley, Kuh, and Welsch (1980) によれば，文献上では完全には確立していない．彼らによれば，文字通りに言えば，2 変数のデータベクトルが同一線上に位置するならば，2 変数は共線性がある，という．多変数に拡張すれば，k 変数のデータベクトルが k より低い次元の部分空間内に位置するならば，k 変数は共線性がある，という．ただし，実際のデータでは，この状態そのものが起こることはまれであり，これに近い状態が起こっている可能性がある．

多重共線性が生起していると，重回帰分析では偏回帰係数に影響が現れ, その正確な推定ができなくなる．その理由はつぎの通りである．

まず，Belsley, Kuh, and Welsch (1980) によれば，多重共線性の有無は全サンプル（N 人）の全説明変数（m 変数）の観測値から成る (5) 式の行列 $\boldsymbol{X}$ の**特異値**（singular value）が m 個あるとし，それらを $\mu_1, \mu_2, \cdots, \mu_m$ と書き，それらのうちの最大のものを μ_{max} とすれば，次式で定義される m 個の**条件指標**（condition index）の値を検討することにより推論できる：

$$\eta_k = \frac{\mu_{max}}{\mu_k}, \qquad \mathrm{k} = 1, 2, \dots, \mathrm{m}. \tag{7}$$

ここで，任意の行列 $\boldsymbol{X}$ の特異値は，一般に $\boldsymbol{X}\boldsymbol{X}^t$ あるいは $\boldsymbol{X}^t\boldsymbol{X}$ の**固有値**（eigenvalues）の平方根として求めることができる．条件指標は説明変数の数だけ計算されるが，条件指標の値が最大となるのは行列 $\boldsymbol{X}$ の特異値のうち (7) 式の分子の最大値 μ_{max} に対して同分母のそれが最小値となる場合である．さらに，行列 $\boldsymbol{X}$ の特異値のうちの最小なものが小さな値であると，対応する行列 $\boldsymbol{X}^t\boldsymbol{X}$ の最小固有値も小さくなる．そのような場合，(4) 式で示される偏回帰係数の推定式の右辺の逆行列の計算は，スカラーの逆数の計算の場合ゼロに近い値で割り算をすることに近い状況が生起するた

め，偏回帰係数の推定が不安定になる．Belsley, Kuh, and Welsch (1980) によれば，条件指標の値が 5 から 10 あたりまでならば**弱い従属性**（week dependency）（言い換えれば**弱い共線性**）が，同じく同じく 30 から 100 あたりであれば強い関係（言い換えれば, **強い共線性**）があるといえる．

多重共線性の生起の有無を検討するには，統計ソフト，例えば SAS の重回帰分析で多重共線性のオプションを指定すると，偏回帰係数の推定と検定の出力のつぎに共線性の診断結果が出力される．そこには，上記固有値の大きい順に，固有値及び条件指標の値等が出力されているので，最後の説明変数（固有値の最小の変数）の条件指標を見て，その値がうえの値の範囲のどこに入っているかを見て，共線性の有無を診断することができる．なお，説明変数に定性的な変数が混在する場合は，あらかじめそのような変数についてはダミー変数化した新たな変数をプログラムで定義したうえで，多重共線性を指定して重回帰分析を行えばよい．

つぎに，それではどの変数（複数）が多重共線性の原因になっているかを見つければよいであろうか．これを行うための 1 つの方法は，**分散インフレ因子**（variance inflation factors）（あるいは**分散拡大因子**）（VIF と略記される）を計算することである．この指標は，上記条件指標と同様に，説明変数ごとに計算され，次式で定義される：

$$VIF_k = \frac{1}{1-R_k^2}, \qquad k = 1,2,\dots,m. \tag{8}$$

ここで，この式の右辺の R^2_k は，k 番目の説明変数をそれ以外の説明変数で説明した時の重相関係数を指す．

(8) 式から明らかなように，分散インフレ因子の値が大きくなる説明変数は，その変数をそれ以外の説明変数で説明した時の説明力が大きな変数であり，その変数は他の説明変数の線形和で近似できることを示しているため，そのような説明変数は多重共線性に関与している可能性があると見れる．例えば，統計ソフト SAS の REG procedure では，多重共線性の出力を指示すると, 重回帰分析の「パラメータの推定値」の項に, 偏回帰係数の推定値，標準誤差，t 値，p–値などに引き続いて,「分散拡大」の項が出力されるので，最大条件指標の値が相対的に大きい場合，この項をみて分散拡大の値の大きい変数を除去するのが当該データの多重共線性に対応する 1 つの方法である．

なお, ここで注意したいのは, SAS の REG procedure では多重共線性の検討を直接指示するパラメータとして collin と collinoint の 2 種類が用意されており，(7) 式で与えらえる条件指標は後者の collinoint を指定した時に出力されるが，<u>collin を指定すると条件指標は説明変数＋ 1 個出力される</u>点である．この場合，条件指標の最初に出力されるのはモデルの y 切片に関わる．さらにこれに対応する固有値の値は説明変数に対応する固有値より通常かなり大きな値となる．両者のどちらを選択するかは，現象の性質やモデルとて y 切片を含めるかどうかに依存する．

ロジスティック回帰分析，さらにはそれを含む一般化線形モデルの場合にも，重回帰分析の場合と同様な問題が存在する（例えば，Agresti, 2002, p. 212; Allison, 2012, p. 60）．Allison によれば，多重共線性の影響は幸運にも共線性を持つ変数（一般に複数）にのみ現れるという．

それでは，重回帰分析以外の一般線形モデルにおける多重共線性を検討するにはどうすればよいであろうか．例えば SAS によるロジスティック回帰分析手続きである LOGISTIC procedure は，同重回帰分析 REG procedure のような多重共線性検討のためのオプションである tol，vif，collin，collinoint を持ち合わせていない．しかし，<u>多重共線性は説明変数の特性であり基準変数のそれではない</u>ため（例えば，Allison, p. 60），多重共線性の有無を検討するためには基準変数は定性的変数であることを無視して当該データに対してロジスティック回帰分析に先立ち REG procedure を適用すればよい．

もちろん，ロジスティック回帰分析の場合，説明変数には定性的変数も含まれていてよいが，定性的変数も混在する説明変数の場合，重回帰分析を適用する前に SAS プログラムを用いて定性的変数をすべてダミー変数化しておく必要がある．SPSS の場合にはシンタックスを用いて同様なプログラムを書けばよい．Tabachnick and Fidell (2014) には多くの多変量解析や時系列解析のための SAS プログラムおよび IBM SPSS シンタックスとその適用例が掲載されている．

第８節　結語

社会行動科学の領域から医学・疫学などの自然科学領域では，得られたデータに対して何らかの統計学的検定を行うことが多い．とりわけ，相関係数の検定，対応のない場合の 2 群間の平均の差の検定, 通常の（独立測定要因，あるいは実験参加者間要因についての）分散分析，反復測定デザイン分散分析（実験参加者内

要因についての分散分析），因子分析，重回帰分析的手法（重回帰分析，ロジスティック回帰分析，それらを一般化した一般化線形モデル）などは，内外の研究論文で日常的に使われている．しかし，これらの方法については，内外の論文を問わず，必ずしも適切な使い方がなされているとは言えず，むしろ誤用・乱用がしばしばなされている．

それらは，相関係数の外れ値や異常値の問題，対応のない 2 群間の平均の差の検定では分散の等質性の検定と平均の差の検定を含めた全体的危険率の問題，通常の分散分析では無作為割り付けの有無，反復測定デザイン分散分析では球形（球面性）仮説の成否と p 値のインフレの問題，因子分析では直交共通因子の因子得点の非直交性の問題，重回帰分析やロジスティック回帰分析などの一般化線形モデルでの多重共線性の問題にかかわる．

一般的に言って，どのような統計的検定や推定を行うに際しても，それぞれの方法には必ず検定や推定が適切に行えるための前提条件や仮定が存在するが，手にしたデータがそれらの前提や仮定を満たしていないと，正しい結論が得られないことになる．それにもかかわらず，内外の審査員付きのレベルの高い論文でさえ，しばしば誤用・乱用が見られる 1 つの原因として考えられるのは，審査付き論文あるいは欧米のインパクトファクターの高い論文で一度そのような誤用・乱用がなされると，あたかもそのような統計的検定や推定が正しいとの錯覚に陥る研究者の心理にあるのではなかろうか．科学的な研究に携わるものに求められるのは，如何に権威のある研究者であろうが論文であろうが完璧なものはなく，どこかにミスや問題があるという態度を持ち続けることではなかろうか．とりわけ，統計的検定や推定を行うに際しては，審査付き論文に出ている方法なので使ってよいというような安易な考えではなく，手にしたデータはそれぞれの統計的方法の前提条件や仮定をほんとうに満たしているのかどうかを常に疑いの目でみてチェックする必要があろう．

引用文献

Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. New Jersey: Wiley-Interscience.

Allison, P. D. (2012). *Logistic Regression Using SAS—Theory and Application*. Cary: SAS Institute Inc.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Carroll, J. B. (1953). An analytical solution for approximating simple structure in Factor analysis. *Psychometrika*, **18**, 23–37.

千野直仁 (1993). 反復測定デザイン概説—その 1　愛知学院大学文学部紀要　**23**, 223–235.

千野直仁 (1994). 反復測定デザイン概説—その 2　球形検定とその周辺についての批判的レビュー　愛知学院大学文学部紀要　**24**, 103–119.

千野直仁 (1995). 教育や心理の分野における ANOVA, MANOVA, GMANOVA 適用上の問題点　愛知学院大学文学部紀要　**25**, 71–96.

千野直仁 (2010). 統計的独立性とその周辺 (1)　愛知学院大学心身科学部紀要　**6**, 119–128.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, **56**, 52–64.

Fenigstein, A. (1979). Self-consciousness, self-attention, and social interaction. *Journal of Personality and Social Psychology*, **37**, 75–86.

Fisher, R. A. (1926). The arrangement of field experiment. *Journal of the Ministry of Agriculture of Great Britain*, **33**, 503–513.

Fisher, R. A. (1935). *The Design of Experiment*. Edinburgh: Oliver and Boyd Ltd.

Garnett, J. C. M. (1919–1920). On certain independent factors in mental measurement. *Proceedings of the Royal Society A*, **96**, 91–111.

Guilford, J. P., & Michael, W. B. (1948). Approaches to univocal factor scores. *Psychometrika*, **13**, 1–22.

Harman, H. H. (1967). *Modern Factor Analysis* (2nd ed. Revised). Chicago: The University of Chicago Press.

Harris, P. (1984). An alternative test for multisample sphericity. *Psychometrika*, **49**, 273–275.

Heermann, E. F. (1963). Univocal or orthogonal estimators of orthogonal factors. *Psychometrika*, **28**, 161–172.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique Simple structure. *British Journal of Statistical Psychology*, **17**, 65–70.

Hogg, R. V. (1961). On the resolution of statistical hypotheses. *Journal of the American Statistical Association*, **56**, 978–989.

Holzinger, K. J. (1944). A simple method of factor analysis. *Psychometrika*, **9**, 257–262.

Holzinger, K. J., & Harman, H. H. (1937). Relationships between factors obtained from certain analyses. *Journal of Educational Psychology*, **28**, 321–345.

Holzinger, K. J., & Harman, H. H. (1941). *Factor Analysis*. Chicago: University of Chicago Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417–441, 498–520.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact

F-distributions. *Journal of the American Statistical Association*, **65**, 1582–1589.

Kestelman, H. (1952). The fundamental equation of factor analysis. *The British Journal of Psychology, Statistical Section*, **5**, 1–6.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. Monterey: Brooks/Cole.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Monterey: Brooks/Cole.

McCullagh, P., & Nelder FRS, J. A. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.

Mendoza, J. L. (1980). A significance test for multisample sphericity. *Psychometrika*, **45**, 495–481.

Mendoza, J. L., Toothaker, L. E., & Crain, B. R., (1976). Necessary and sufficient conditions for F ratios in the L×J × K factorial designs with two repeated factors. *Journal of the American Statistical Association*, **71**, 992–993.

永田靖（2008）．入門統計解析法　日科技連出版社

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, **135**, 370–384.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559–572.

Pinzka, C., & Saunderrs, D. R. (1954). Analytic rotation to simple structure. II: Extension to anoblique solution. *Research bulletin*, RB–54–31. Princeton, N. J.: Educational Testing Service.

Rouanet, H., and Lépine, D. (1970). Comparison between treatments in a repeated- measurement design: ANOVA and Multivariate methods. *The British Journal of Mathematical and Statistical Psychology*, **23**, 147–163.

繁桝算男・柳井晴夫・森敏昭編（1999）．Q ＆ A で知る統計データ解析（第 1 版）サイエンス社

繁桝算男・柳井晴夫・森敏昭編（2008）．Q ＆ A で知る統計データ解析（第 2 版）サイエンス社

菅原健介（1984）．自意識尺度（self-consciousness scale）日本語版作成の試み　心理学研究　**55**, 184–188.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, **15**, 201–293.

Sugiura, N., & Nagao, H. (1968). Unbiasedness of some test criteria for the equality of one or two covariance matrices. *The Annals of Mathematical Statistics*, **39**, 1686–1692.

Tabachnick, B. G., Fidell, L. S. (2014). *Using Multivariate Statistics* (6th ed.) Harlow: Pearson.

竹内啓（1978）．数理統計学の方法的基礎　東洋経済新報社

Thomson, G. (1948). *Multiple Factor Analysis*. Chicago: University of Chicago Press.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, **38**, 406–427.

Thurstone, L. L. (1935). *The Vector of Mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1945). A multiple group method of factoring the correlation matrix. *Psychometrika*, **10**, 73–78.

Tukey, J. W. (1953). *The problem of multiple comparisons*. Dittoed manuscript of 396 pages, Department of Statistics, Princeton university.

浮田恵司・横井次郎（1996）．科学研究費報告書分析データ

最終版平成28年 9 月28日

Bulletin of The Faculty of Psychological & Physical Science, No. 12, 101−112, 2016

Misuse and Abuse of Statistics in the Social and Behavioral Sciences, and Medical Science

Naohito CHINO

Abstract

Misuse and/or abuse of statistics can often occur in research in the social and Behavioral sciences, medical science and so on. We discuss it in the following seven cases:

1. In testing Pearson's product moment correlation coefficient when outliers and/or abnormal values exist.
2. In applying the usual test of difference in means followed by the test of homogeneity of variances.
3. In testing effects of between-subject factors with a few treatment levels to which we can't randomly assign subjects in ANOVA.
4. In testing various contrasts as the subsequent tests after the overall hypothesis of equality of means or is rejected. Similar situations occur when the overall or omnibus interaction hypothesis is rejected in factorial designs.
5. In testing effects of factors in repeated measurement designs. Tests for sphericity hypotheses such as global sphericity hypothesis and multisample sphericity hypothesis are often left aside, although there is a possibility that the F distributions corresponding to some hypotheses about repeated measurement factors are distorted.
6. In the applications of factor analysis to empirical data.
7. In the applications of the generalized linear models such as multiple regression, logistic regression and so on. Especially, the problem of multicollinearity in these methods are often left aside, although parameter estimates for these models will be unstable if the multicollinearity occurs in the data under consideration.

Keywords: misuse of statistics, multicollinearity, outlier, sphericity hypotheses