愛知学院大学心身科学部紀要第9号（45–54）（2013）

〈総説〉

行動研究における反復測定デザインANOVAの誤用— I

千　野　直　仁*

1　はじめに

社会・行動科学や医学・疫学の分野では，要因の効果の有無の検討に際して，各サンプルに複数の物差しをあてがう反復測定を行うことがしばしばある．そのようにして得られた得点の分析のためのモデルの1つとしては，いわゆる分散分析における**乱塊法**（randomized block design）や**分割区画型デザイン**（split-plot factorial design）が知られており，例えば前者の場合，物差しの数（水準数）を I，ブロック数を K とすると，当該要因の効果の検定統計量は，つぎのようになる：

$$F=(K-1)\frac{SS_A}{SS_E}. \tag{1}$$

ここで，SS_A 及び SS_E は，x_{ik} を第 i 水準の第 k ブロックの得点として，

$$SS_A=\sum_{i=1}^{I} N\,(\bar{x}_{i\bullet}-\bar{x}_{\bullet\bullet})^2,$$

$$SS_E=\sum_{i=1}^{I}\sum_{k=1}^{K}(x_{is}-\bar{x}_{i\bullet}-\bar{x}_{\bullet s}+\bar{x}_{\bullet\bullet})^2. \tag{2}$$

ここで，注意すべき点は乱塊法そのものは，必ずしも反復測定デザインを意味しないということである．乱塊法では，均質でない実験参加者を主要因子以外のいわゆる局外因子により複数のグループ（ブロックとも言われる）に分けたうえで，各グループごとに実験参加者（複数）を主要因子のいずれかの水準に無作為に割り付ける．乱塊法の模型には，いろいろなものが考えらえるが，通常最もよく用いられる模型は

$$Y_{ik}=\mu+\alpha_i+BL_k+E_{ik},\quad i=1,2,\cdots,I,\ k=1,2,\cdots,K. \tag{3}$$

ここで，Y_{ik} は因子Aの水準 i，かつブロック因子BLの第 k 水準の測定値を実現値とする確率変数，μ は一般平均で定数，α_i は因子Aの（水準 i の）主効果で定数，BL_k はブロック因子の水準 k の効果に関する確率変数，E_{ik} は誤差項を表す確率変数である．このデザインは，式の右辺の誤差項以外の項に母数（定数）α_i と確率変数 BL_k を含むので，**混合模型（モデル）**（mixed model）とも言われる．ここで，各ブロックの実験参加者数が1名だとすると，主要因子の各水準には同一実験参加者が反応することになり，そのような場合に限り，デザインは反復測定デザインとなる．ここで，反復測定デザインを構成する因子（要因，または変数）を，**被験者内要因**（または変数）（the within-subjects factor, または within-subjects variable）ということも多い．また，最近では subject の代わりに participant を充て，**（実験）参加者内要因**（the within-participants factor）などと表記することもある．

ただし，反復測定デザインデータの場合，場合によっては(1)式等のF統計量は歪み，p値のインフレをもたらすことが知られている．この現象をMATLABを用いてわかりやすく図式化したのが，図1である．図1で，青色のF分布は，球形仮説が成り立つ時の，ある自由度のF–分布を描いたものであり，この分布で横軸の値F＝2.2以上である確率の合計を，青色で塗りつぶした部分として表示した．一方，赤色のF–分布は，球形仮説が成り立たない場合の，自由度が縮退したF–分布を描いたものである．また，この赤色の分布で青色の分布と同じ横軸の値以上である確率の合計を，赤色で塗りつぶした部分で表示した．ただし，この場合，その下方の部分は既に青色で塗りつぶされているので，この部分のみ赤色でなく青色で表示されていることに注意されたい．図1からは，球形仮説が成り立たない場合には，成り立つ場合に比べてp値が大きくインフレしていることが一目瞭然である．この図は，千野（1999a）に掲載したものを転載したも

＊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

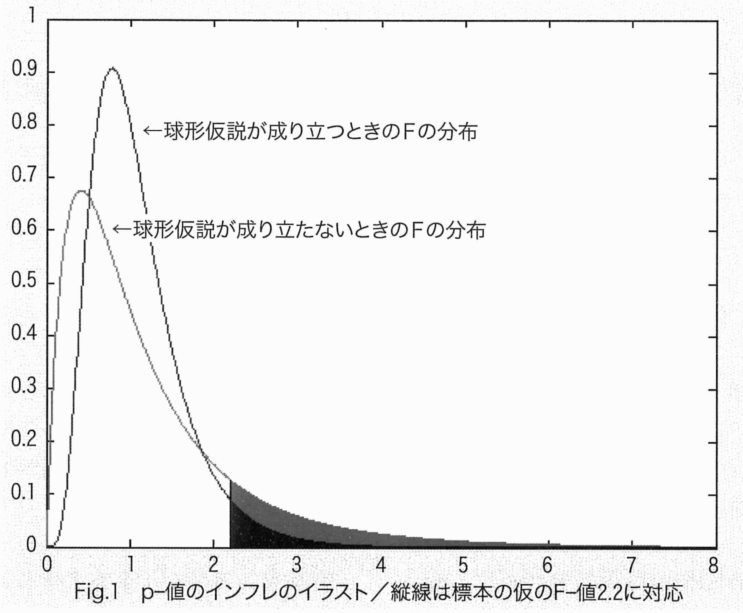

図 1　反復測定デザイン ANOVA の F–統計量の歪みに伴う p 値のインフレ

のである．

上記のような p 値のインフレを避けるためには，同統計量の歪みの有無の検定を行う必要があり，モクリーの球形（球面性）検定（Mauchly's sphericity test）(Mauchly, 1940) が知られている．本邦では，これを球面性検定と訳している場合が多い（藤越, 2009；石村, 1997；SAS, Version 9.3, SPSS, Version 21）が，筆者は球形検定と訳している（千野，1993, 1994, 1995；1999；2003a；2003b）．その理由については，後の節でふれることとし，以降特別の理由がない場合は球形検定と書くこととする．

ここで，モクリーの検定は，もともとは分散分析の文脈ではなく一般の多変量データがどの方向にも等方的に広がっていることを帰無仮説，すなわち，

$$H_0 : \mathbf{\Sigma}_p = \sigma^2\, \boldsymbol{I}_p, \tag{4}$$

とするものである．ここで，$\mathbf{\Sigma}_p$ は，多変量データ間の $p \times p$ 母共分散行列（population covariance matrix）であり，I_p は p 次の単位行列を指す．すなわち，

$$\mathbf{\Sigma}_p = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \cdots & \sigma_p^2 \end{pmatrix}. \tag{5}$$

一方，分散分析の文脈でのモクリーの検定は，これを正規直交変換後の反復測定変数間共分散行列，すなわち $\boldsymbol{M}\mathbf{\Sigma}\boldsymbol{M}^t$ に対して，次式を帰無仮説とするものである：

$$H_0 : \boldsymbol{M}\,\mathbf{\Sigma}_p\, \boldsymbol{M}^t = c^2\, \boldsymbol{I}_m. \tag{6}$$

ここで，$\boldsymbol{M}$ は $m \times p$ 正規直交対比行列であり，この行列の各行は p 個の反復測定変数に対する対比の係数を表す．また，c は未知の正定数であるとする．(6) 式の帰無仮説は，分散分析の文脈における最も単純な球形仮説である．

ここで，もし反復測定要因ばかりを複数個仮定するようなデザイン，すなわち反復測定要因デザイン（repeated measures factorial designs）の場合には，図 2 の手順により，大局的球形仮説（global sphericity hypothesis）及び局所的球形仮説（local sphericity hypothesis）の順に，2 種類の球形仮説の検定が必要となる（例えば，Chino, 1999; Crowder & Hand, 1990; Kirk, 1982）．これらは，それぞれ全体的循環性（overall circularity），局所的循環性（local circularity）と呼ばれることもある（Kirk, 1982; Mendoza, Toothaker, & Crain, 1976）．大局的球形仮説と局所的球形仮説の使い分けのポイントは，両仮説の違いにより，分散分析の構造模型を加算的モデル（additive model）と非加算的モデル（non-additive model）のいずれかに使い分けする必要がある点である．

一方，独立測度と反復測度が混在するデザインでは，上記とは異なる球形（球面性）仮説の検定が必要であ

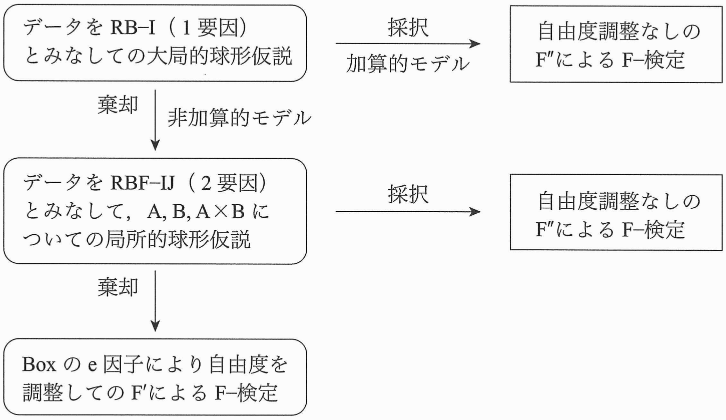

図 2　2 要因反復測定デザイン分散分析における球形検定の手順

る．たとえば，**2処理分割区画要因デザイン**（two-treatment split plot factorial design）で反復測度型のサンプリングを行う場合には，独立測度としての処理Aの各水準に異なる観測単位（実験参加者）群を無作為に割り付けるので，水準ごとで反復測度となる処理Bの水準間の相関関係は異なるものになる可能性がある．ここで，しばしば，反復測度要因を被験者内（あるいは実験参加者内）要因と表記するのに対して，独立測度要因の方は**被験者間要因**（between-subjects factor），あるいは**実験参加者間要因**（between-participants factor）と呼ぶ．その言い方からすると，分割区画型の反復測定デザインは，被験者間要因と被験者内要因（もしくは，実験参加者間要因と実験参加者内要因）が混在するデザインとも言える．

ここで，独立測度の水準ごとの反復測度となる処理Bの J 次の水準間母共分散行列を，$\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_I$，とし，$\boldsymbol{I}_{J-1}$ を $J-1$ 次の単位行列とすれば，処理B（反復測度要因）及び A×B についての F 比が正確な F–分布に従うための条件は，つぎが成り立つことであることを，Huyhn and Feldt（1970）が証明した：

$$H_0 : \boldsymbol{M}\boldsymbol{\Sigma}_1\boldsymbol{M}^t = \boldsymbol{M}\boldsymbol{\Sigma}_2\boldsymbol{M}^t = \cdots = \boldsymbol{M}\boldsymbol{\Sigma}_I\boldsymbol{M}^t = \sigma^2\,\boldsymbol{I}_{J-1}. \tag{7}$$

うえの仮説は，**多標本球形仮説**（multi-sample sphericity hypothesis）と呼ばれる．彼らは，この仮説がつぎの 2 ステップで検定できることを示している：

$$H_0^{(1)} : \boldsymbol{M}\boldsymbol{\Sigma}_1\boldsymbol{M}^t = \boldsymbol{M}\boldsymbol{\Sigma}_2\boldsymbol{M}^t = \cdots = \boldsymbol{M}\boldsymbol{\Sigma}_I\boldsymbol{M}^t, \tag{8}$$

及び

$$H_0^{(2)} : \boldsymbol{M}\boldsymbol{\Sigma}_{pooled}\boldsymbol{M}^t = \sigma^2\,\boldsymbol{I}_{J-1}. \tag{9}$$

ここで，$\boldsymbol{M}$ は $J-1$ 行 J 列の直交正規対比行列である．また，$\boldsymbol{\Sigma}_{pooled}$ は，

$$\boldsymbol{\Sigma}_{pooled} = \frac{1}{\sum_{i=1}^{I}(N_i-1)}\sum_{i=1}^{I}(N_i-1)\boldsymbol{\Sigma}_i, \tag{10}$$

である．

一方，Mendoza（1980）は，多標本球形仮定が 1 ステップで検定できることを示した．この方法は，**1ステップ多標本球形検定**（a single-step multisample sphericity test）と言える．この方法の尤度比基準の概要については，例えば千野（1999）を参照のこと．Mendoza（1980）の多標本球形検定は，実はその12年前に Sugiura and Nagao（1968）が，それを含むより一般的な形で提案している．明らかに Mendoza は，Sugiura らの論文を見落していたことになる．

これまで紹介した各種球形検定は，すべて従属変数が 1 変数の場合の検定である．これに対して，従属変数が複数存在する場合も，しばしば出くわす．とりわけ，各従属変数が同一単位では測ることができない（non-commensurable）場合には，**2重多変量モデル**（doubly multivariate model, 略して DMM），または**多変量混合模型**（multivariate mixed model, MMM）を用いるのがよい（Bock, 1975; Boik, 1988; Thomas, 1983; Timm, 1980）．この場合の球形仮説は，**多変量球形仮説**（multivariate sphericity hypothesis）と呼ばれ，つぎのように書ける：

$$\underset{pq\times pq}{\boldsymbol{\Sigma}} = \boldsymbol{I}_q \otimes \boldsymbol{\Upsilon}_p. \tag{11}$$

ここで，$\boldsymbol{\Upsilon}_p$ は，適当な p 次の正定値行列である．また，$\boldsymbol{I}_q$ は q 次の単位行列で，q はもとの p 変量 t 時点のデータを変換して最終的に p 変量データにするため

の変換行列の階数（rank）である．記号⊗は，クロネッカー積（Kronecker product）を表す．この仮説は，当初 Thomas (1983) が MMM 分析が妥当であるための十分条件としていたのに対して，Boik (1988) が必要十分条件であることを示している．この方法の理論的概要については，例えば千野（1994）を参照されたい．

千野（1993, 1994, 1995）は，上記を含む反復測定デザインモデルによる ANOVA のみならず MANOVA・GMANOVA モデルやそれらによる多くの反復測定デザイン分散分析の理論と問題点について，まとめている．千野は，さらに千野（1999, 2003a, 2003b）でも，反復測定デザインと現状の問題点について紹介している．それにも拘わらず，現状では心理学研究など主要な学会誌に掲載されている論文で，モクリーの検定を無視しているものが数多くみられる．

例えば，*Japanese Psychological Research* 論文の最新号 Vol. 55に掲載されている 9 篇の論文のうち，7 篇は分散分析デザインを用いている．そのうちの 1 篇は MANOVA，1 篇は完全無作為化デザイン，1 篇はノンパラメトリック検定を行っているが，残りの 4 篇は反復測定デザイン分散分析を行っている．ちなみに，これら 4 篇のデザインは，1 要因から 3 要因までの反復測定デザインが 3 篇と，分割区画型反復測定デザインが 1 篇から成る．しかし，4 篇とも球形検定に対する言及が全く見られず，さらに万が一 F 分布に歪みが生じる場合の自由度の修正の記述も全くない．

実は，このような現状は日本特有の現象でもないし，また最近の傾向でもない．反復測定デザイン分散分析の効果の有無の検定統計量 F が歪まないための必要十分条件が上記 (6) 式の帰無仮説であることが，1970年に 2 つの論文，すなわち Rouanet and Lépine（1970）と Huynh and Feldt（1970）により証明される前の1960年代までは，帰無仮説のもとで F 比が正確な F–分布に従うための条件として，対称性仮定（symmetry assumption）が知られていたが，それが F 比の正確な F–分布に従うための必要条件なのか十分条件なのかについて，統計学者の間でさえ定説がなかったという（Rouanet & Lépine, 1970）．

ここで，対称性仮定は複合対称性（compound symmetry）と呼ばれることもあるが（Crowder & Hand, 1990; Kirk, 1982），一般に反復測度の水準数を p として，上記の水準間母共分散行列 $\mathbf{\Sigma}_p$ について，

$$\sigma_1^2=\sigma_2^2=\cdots=\sigma_p^2, \tag{12}$$

および

$$\sigma_{12}=\sigma_{13}=\cdots=\sigma_{1p}=\sigma_{23}=\sigma_{24}=\cdots=\sigma_{p-1,\,p}, \tag{13}$$

が成り立つことをいう．なお，これら 2 式が成り立つ行列は S 型行列（a type S matrix）と呼ばれることがある（Huynh & Feldt, 1970）．

そのような状況下で，既に千野（1995）でも指摘したように，1950年代から1970年代の初等に欧米の学会を中心に球形仮説からの乖離についての全体的 F 検定の頑健性についての重要な論文の幾つかに楽観的見解，すなわち，テストサイズ（第 1 種の過誤）のインフレは，例えば名目サイズ 5 %の場合，高々10%ぐらいであるという論文が発表された（例えば，Box, 1954; Collier et al., 1967; Keppel, 1973）．本邦においては例えば橘（1986）なども，そのような楽観的な見解を述べている．しかし，Box (1954b) や Collier et al. (1967) の原著を見れば，彼らの見解は特殊な反復測度間相関構造や限られた反復要因の水準数に基づいていることがわかる．千野（1995）は，そのことを SAS プログラムによる簡単なシミュレーションにより明らかにしている．図 3 は，これを再掲載したものである．

図 3 で，横軸は反復測定要因の水準数を，縦軸は名目サイズ 5 %の場合の，球形仮説からの乖離が最大の場合（Box/Greenhouse-Geisser ϵ の下限値，すなわち保守的検定にあたる）の理論的な p 値を表す．また，図中 B なる記号は Box（1954b）のシミュレーションにおける p 値を，記号 C は Collier et al.（1967）のシミュレーションにおける p 値を，それぞれ表す．さらに表中の数値 1 から 5 は，サンプルサイズが順に 3，5，10，20，30であることを示す．この図を見れば，上記 Box や Collier et al. のシミュレーションによる見解が，如何に特殊な反復測度間相関構造や限られた反復要因の水準数に基づいているかが，明らかであろう．さらに，たとえ名義サイズが 5 %であろうと，球形仮説が成り立たない場合には，p 値のインフレは10%どころではなく，場合により30%を超えることが起こりうることも明らかである．研究者が，このような場合に球形検定を無視するならば，場合によっては本来ならば棄却できない（反復測定要因の）主効果や交互作用仮説が棄却されてしまうことになり，分析の結論が逆転してしまい，きわめて危険である．

もっとも，図 3 のシミュレーション結果の横軸の値，すなわち反復測定要因の水準数が10を超えるようなケースは，単一要因では数少ないであろう．しかし，既に見たように，反復測定要因が複数存在する場合，

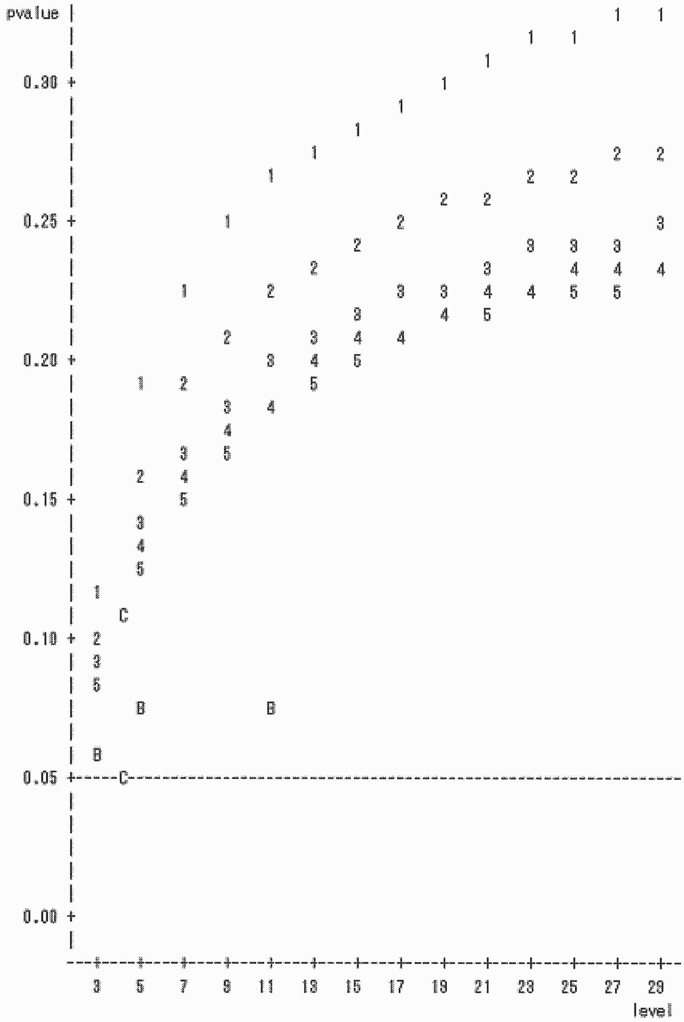

図 3　反復測定デザイン ANOVA における危険率のインフレーション（千野(1995)より再掲載）

図 3 から明らかなように，われわれはまず複数要因から成る反復測定デザインデータを 1 要因にいわば“潰して”，データを 1 要因反復測定デザインデータとみなすことにより大局的球形仮説の検定を行う必要があることに注意しよう．この場合の反復測定要因の水準数は，3 要因反復測定デザインで第 1 要因から第 3 要因までの水準数が例えば順に，2，3，3 であるとすれば，2×3×3＝18となり，簡単に水準数は10を超える場合がありうるのである．

しかし，一方では千野（1995）でもふれたように，欧米では1960年代以降反復測定デザイン ANOVA の問題点を指摘する論文（例えば，Hertzog & Rovine, 1985; Jennings & Wood; Keselman & Logan, 1980; Latour & Miniard, 1983; MaCall & Appelbaum, 1973; Wilson, 1967）や，同デザインの正しい使い方を学会の編集方針として掲げる学会誌も見られる（Jennings, 1987; Vasey & Thayer, 1987）．後者の Psychophysiology 誌で Vasey and Thayer (1987) は，心理生理学的研究（psychophysiological research）では反復測定デザインは最もよく知られたデザインであるが F–比の歪みの可能性に言及した論文が少ないことを，Psychophysiology 誌の第12巻で84%（Jennings & Wood, 1976の指摘を引用），第21巻と22巻で50%以上が無視していることを指摘して，憂慮している．また，Psychophysiology 誌の編集方針を要約した Jennings (1987) は，ANOVA のみならず，場合によっては MANOVA（より正確には，GMANOVA）を用いるべきことも指摘している．

それでは，同雑誌の最新号としての2013年度分では，この事態はどれほど改善されているであろうか．現時点での最新の 3 号分である No. 4 から No. 6 をチ

ェックしたところ，3号分合計27篇の論文のうち，乱塊法型あるいは分割区画型反復測定 ANOVA を用いている論文は14本（全体の約52%）で, それらのうち，球形検定とそれに引き続く Greenhouse-Geisser ϵ への言及のあるものは9篇であるのに対して，いずれに対しても言及のない論文は5篇あった．その結果，球形検定を無視しているとみられる論文は，36%である．この数字は，他の内外の行動科学雑誌，とりわけ心理学雑誌の現状でのそれに比べ，たぶん際立って低い数値ではなかろうか．その原因として考えられるのは，Psychophysiology 誌が，同雑誌の編集方針として球形検定の必要性を説いたことであろう．

もっとも，上記 Psychophysiology 誌の最新の3号分の球形検定に言及のある9篇についても，詳しく見てみると問題がないわけではない．それは，これらの論文で用いられている反復測定デザインの多くは，要因デザイン（factorial design）であり，その場合，第1節でも指摘したように，乱塊要因デザインでは大局的球形仮説および局所的球形仮説の検定が，また分割区画型デザインでは多標本球形仮説の検定の2ステップ検定を行うのであればその第1ステップとしての (8) 式の検定，すなわち被験者間要因の各水準の共分散行列を正規直交化したうえでの（同行列(複数)の）共分散行列の等質性の検定が必要なのだが，9篇共，これらの検定に関する言及が皆無である．たぶん，その原因としては，Jennings（1987）及び Vasey and Thayer（1987）による学会の編集方針の中に，これらの重要な2種類の球形仮説に対する言及がなされていないことにあるのではなかろうか．また，残念ながら，現時点では国際的な統計ソフトである SAS も SPSS も，用意されている repeated 文のみでは, 大局的球形検定や, 多標本球形検定の第1段階の検定を行わないので，ユーザがそれらの検定を SAS プログラムや SPSS シンタックスを用いて追加しないといけないことも，もう1つの原因となっていると思われる．少なくとも，SAS プログラムについては，千野のウエブサイトを参照されたい．

上記以外の分野でも最近筆者が気づいたものの1つが認知心理学，とりわけワーキングメモリー関連の研究における反復測定デザインの使用上の問題点である．今回はこの分野での内外の研究をきちんとチェックできてはいないが，象徴的な論文を3つ上げる．1つは，膨大なワーキングメモリー研究の先駆けの1つと思われる Levy (1971) 及び Baddeley and Scott (1971) の論文であり，他方は最近の本邦における同領域の論文である．

まず，Levy 論文では4要因反復測定デザインで，水準数は2×3×2×7となっており，総水準数は84である．これに対して，第1実験では被験者は24名，第2実験では2×2×2×9となっており，総水準数は72で，被験者は32名である．一般に，すべての要因が反復測度の場合のデザインでは，大局的球形仮説も含めた検定をきちんと行うには，被験者数は総水準数以上必要である．Levy 論文では，その意味でまず被験者数が全く不足しており，きちんとした球形検定は不可能であるだけでなく，実際の検定でも特定の球形仮説への言及でさえなされていない．一方，Baddeley et al. 論文でのデザインは，2要因反復測定デザインで，水準は2×5で，総水準数は10であるのに対して，被験者数は40名なので，球形仮説のきちんとした検討は理論上は可能であるが，彼らの論文にはやはり球形仮説への言及さえない．もっとも，これらの論文は1971年のものなので，球形検定を彼らや論文を審査した審査者が知らなくてもやむを得ないかもしれない．

一方，同様なワーキングメモリー関連の論文として心理学研究に掲載されている水野（1996）論文では，3つの実験とも反復測定デザイン ANOVA を行っているがいずれの場合も球形仮説の検定への言及が皆無である．例えば，第1実験での主要な分析は2要因反復測定デザイン ANOVA で，水準数は4×7であるが，被験者は24名でしかない．したがって，大局的球形検定を含めたきちんとした分散分析を行うことも不可能である．

上記のような反復測定デザインの現状を踏まえ，筆者は1990年代中頃にまとめた一連の反復測定デザイン分散分析の理論と方法のレビュー（千野, 1993, 1994, 1995）に引き続き，今回から数回でこれまでにきちんと紹介してこなかった反復測定デザイン分散分析の理論をまとめ，また現状で内外の主要な統計ソフトの現状と問題点を指摘することとする．今回は，紙面の都合上，つぎの節で sphericity test の日本語訳の問題点について指摘するにとどめる．

2　sphericity test の日本語訳の現状と問題点

これまで，本邦では sphericity test に対する日本語訳として幾つかのものが知られている．千野（1993, 2003a, 2003b）が指摘しているように，この検定は等方性検定，球状性検定，球面性検定，球形検定などと

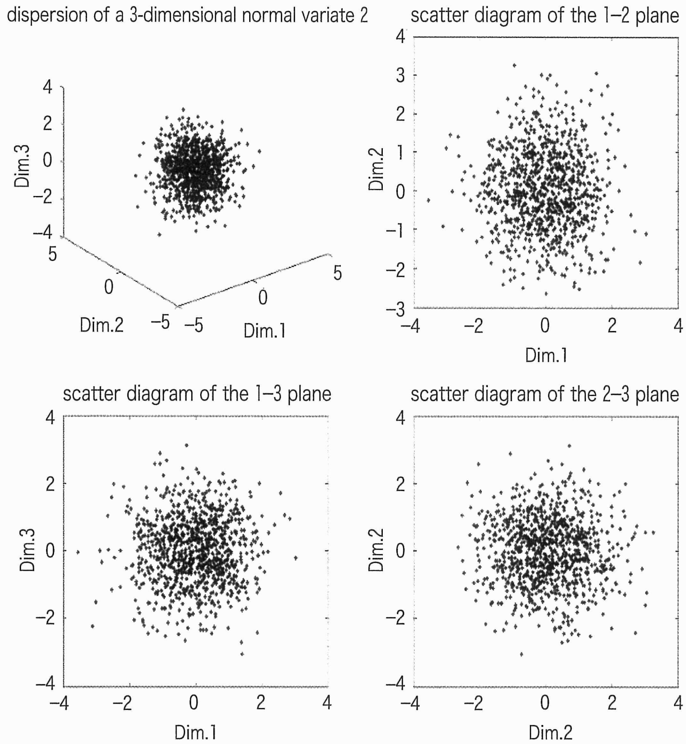

図 4　Mauchly (1940) の (4) 式の球形仮説が成り立つ場合の 3 変量データの散布図（再掲載）

訳されてきた．また，前節で指摘したように，最近ではこれらの中の「球面性検定」なる訳が定着しつつある．この節では，なぜ sphericity test の訳として「球面性検定」があまり適切ではないかを説明する．

まず，sphericity という英単語の語根としての sphere は，例えば Cambridge Advanced Learner’s Dictionary (2012) で調べると “an object shaped like a round ball” となっており,「丸いボール」ではあっても「球の表面」を連想する「球面」ではない．千野 (2013) は, 最近, これに関わる議論を MATLAB の出力を用いて説明している．ここでは，それらの一部を再掲載する：

まず, 図 4 の 4 つの散布図のうち, 上部左側の図は, 平均がすべてゼロで変数相互間もすべて無相関なる 1,000個の 3 変数からなる正規乱数を発生させ，3 次元空間上にプロットしたものである．それに対して，4 つの散布図の残りの 3 つは，当該 3 変数の 3 次元空間への同時プロットではなく，いわばそれを Dim.1–Dim.2 平面，Dim.1–Dim.3 平面，Dim.2–Dim.3 平面に順に投影した 2 次元の散布図である．反復測定変数が 3 変数から成る多変量データで, (4) 式の Mauchly (1940) の球形仮説が成り立つ場合，そのようなデータは左上の図のような 3 次元空間上に等方的に広がっており，

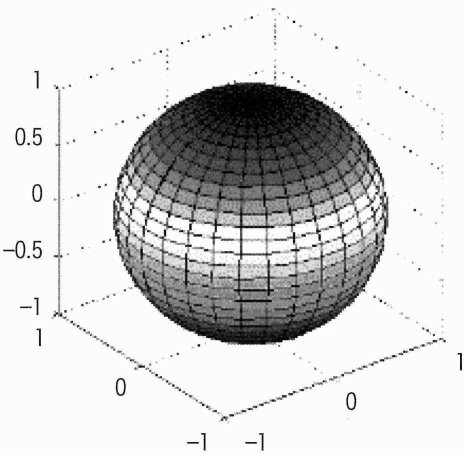

図 5　ボールの表面としての 3 次元球面のイラスト（再掲載）

その大雑把な全体像はデータがボール状に広がっていることを意味する．そのような 3 次元空間上でのデータの散布状態を，3 つの 2 次元平面上に投影させると，図 4 のそれ以外の 3 つの平面上の散布図となることが見て取れる．これらの図から，ボール状に広がった 3 次元データは，決して球の表面にのみ広がっているのではなく，ボールの内側全体に広がっていることがわ

かる.

これに対して，図 5 はデータが球の表面にのみ広がっているイメージをボールの表面を色づけして示したものである．つまり，sphericity hypothesis を「球面性仮定」と訳すと，そのイメージは主として球の表面を連想することが多いと思われるので，Mauchly の sphericity hypothesis を球面性仮定と訳すことは，少し適切さを欠くのでは，というのが筆者の主張である．

3 討 論

社会・行動科学や医学・疫学の分野では，研究対象となる現象に対して同時あるいは経時的に多くの物差しをあてがい，現象の差異や変化とその原因を分散分析を用いて検討することが多い．そのようなデザインによる分析方法は，反復測定デザイン分散分析と呼ばれ，場合によっては，要因の効果を見るための統計指標である F–統計量に歪みを生じさせる．このことは，結果として，検定の過誤率のインフレをまねく．その結果，インフレを考慮せず各種反復測定要因がらみの要因の効果の検定で，本来ならば要因の効果はないにも関わらず，あると誤った判断をする可能性がある．

これを避けるためには，もとは分散分析のための仮説としてではなく，一般の多変量データに対する 1 つの仮説として Mauchly（1940）により提案された sphericity test を分散分析用に修正したものを，反復測定デザインデータの分散分析に先立ち行う必要があるが，1970年に 4 名の研究者により提案された 2 つの論文及びその後の当該検定の発展にもかかわらず，未だに内外の主要な論文でさえ，この検定が適用されない場合が多いことをみてきた．

当論文では，また sphericity test の邦訳として最近定着しつつある「球面性検定」なる訳が，sphere の本来の意味からも，Mauchly の提案した sphericity hypothesis の意味からも，十分適切とは言えないことを述べた．

References

Baddeley, A. D. & Scott, D. (1971). Word frequency and the unit sequence interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **10**, 35–40.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: Wiley.

Boik, R. J. (1988). The mixed model for multivariate repeated measures: Validity conditions andan approximate test. *Psychometrika*, **53**, 469–486.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of varianceproblems–II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics*, **25**, 484–498.

Cambridge University Press (2008). *Cambridge Advanced Learner's Dictionary*. 3rd Ed. Cambridge: Cambridge University Press.

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

千野直仁（1994）．反復測度デザイン概説 —その 2　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（1999a）．反復測定(測度)分散分析／基礎と応用 2.1.3節 大局的球形仮定と局所的仮定　1999年 3 月 5 日〈http://www.agu.ac.jp/~chino/anova/chapter2/sec2-1-3.html〉（2013年 7 月18日）

千野直仁（1999b）．分散分析 Q39–41.　繁桝算男・柳井晴夫・森敏昭編著　Q&A で知る統計データ解析 Dos and Don'ts　サイエンス社　pp. 76–84.

千野直仁（2003a）．分散分析 Q39–41.　繁桝算男・柳井晴夫・森敏昭編著　Q&A で知る統計データ解析 Dos and Don'ts 第 2 版　サイエンス社　pp. 75–84.

千野直仁（2003b）．反復測定データの分析とその周辺　教育心理学年報 **42**, 107–118.

千野直仁（2013）．反復測定(測度)分散分析／基礎と応用 2.1.1節　対称性仮定と球形仮定　2013年 5 月22日〈http://www.agu.ac.jp/~chino/anova/chapter2/sec2-1-1.html〉（2013年 7 月25日）

Collier, R. O., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika*, **32**, 339–353.

Crowder, M. J., & Hand, D. J. (1990). *Analysis of Repeated Measures*. London: Chapman and Hall.

Hertzog, C., & Rovine, M. (1985). Repeated-measures analysis of variance in developmental research: Selected issues. *Child Development*, **56**, 787–809.

藤越康祝（2009）．経時データ解析の数理　朝倉書店

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.

IBM Corporation (1989). *IBM SPSS Statistics, Version 21*. New York: IBM Corporation.

石村貞夫（1997）．SPSS による分散分析と多重比較の手順　東京図書

Jennings, J. R. (1987). Methodology Editorial policy on analysis of variance with repeated measures. *Psychophysiology,* **24**, 474–475.

Jennings, J. R., & Wood, C. C. (1976). The ε-adjustment procedure for repeated-measures analyses of variance. *Psychophysiology*, **13**, 277–278.

Keppel, G. (1973). *Design and analysis: Researcher's handbook*. Englewood Cliffs: Prentice-Hall.

Keselman, H. J., & Rogan, J. C. (1980). Repeated measures F tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology*, **17**, 499–503.

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

Latour, S. A., & Miniard, P. W. (1983). The misuse of repeated measures analysis in marketing research. *Journal of Marketing Research*, **20**, 45–57.

Levy, B. A. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **10**, 123–132.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development*, **44**, 401–415.

Mauchly, J. W. (1940). Significance test for sphericity of a normal *n*-variate distribution. *The Annals of Mathematical Statistics*, **11**, 204–209.

Mendoza, J. L. (1980). A significance test for multisample sphericity. *Psychometrika*, **45**, 495–498.

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

水野りか (1996). 視空間スケッチパッドへの長期記憶の影響 —親近性効果の新たな要因の検討— 心理学研究 **67**, 359–366.

Rouanet, H., & 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.

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.

SAS Institute (2002). *SAS software 9.3*. Cary: SAS Institute Inc.

橘敏明 (1986). 医学・教育学・心理学にみられる統計的検定の誤用と弊害 医療図書出版

Thomas, D. R. (1983). Univariate repeated measures techniques applied to multivariate data. *Psychometrika*, **48**, 451–464.

Timm, N. H. (1980). Multivariate analysis of variance of repeated measures. In P. R. Krishnaiah (Ed.), *Handbook of statistics*, Vol. 1 (pp. 41–87). New York: North Holland.

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in Psychophysiology: A multivariate solution. *Psychophysiology*, **24**, 497–486.

Wilson, R. S. (1967). Analysis of automatic reaction pattern. *Psychophysiology*, **4**, 125–142.

最終版平成25年7月29日受理

Bulletin of The Faculty of Psychological & Physical Science, No. 9, 45–54, 2013

The Misuses of Repeated Measures Design ANOVA in Behavioral Research—I

Naohito CHINO

Abstract

A brief survey is made on the repeated measures design ANOVA. It is pointed out that the misuses of repeated measures design ANOVA have still been frequently observed in behavioral research, typical misuse being to ignore the problem of sphericity, which had repeatedly been pointed out elsewhere (i.e., Jennings, 1987; Vasey & Thayer, 1987). It is also pointed out that the global sphericity hypothesis and the multisample sphericity hypothesis have also been neglected in applying factorial designs with repeated measures in most cases. Finally, the inappropriateness of the Japanese name, "Kyumensei kentei", for the sphericity test, is suggested by depicting a three-dimensional scatter diagram as well as the corresponding two-dimensional scatter diagrams of the artificial repeated measures design data under the Mauchly's sphericity hypothesis.

Key words and phrases: repeated measures design ANOVA, sphericity test, global sphericity hypothesis, local sphericity hypothesis, multisample sphericity hypothesis