The Japanese Association for Behavior Analysis(JABA)

単一事例実験データへの統計的方法の適用

山田 剛史
東京大学大学院教育学研究科

1. 本講演の目的

行動分析の研究対象は，主に単一の個体の行動である．単一事例実験計画(single-case experimental design)はひとつのケースから妥当な推論を引き出すための方法として，行動分析の領域においても広く利用されている．単一事例実験計画で得られたデータの分析は，主にデータをグラフ化して目で見て判断する，視覚的判断(visual inspection)という方法によって行われてきた．しかし，この視覚的判断という方法は，方法が客観的でない，他の評定者・方法との分析結果が一致しない，といった問題も指摘されている．そこで，視覚的判断に代わるもの，あるいはこれを補佐するものとして，単一事例実験データへの統計的方法の適用が提案されている．本講演では，こうして提案された統計的手法のうち，代表的なものを紹介し，その長所と問題点を述べる．

2. データにおける系列依存性(serial dependency)

シングルケースデータは，基本的にひとつの被験体について繰り返し測定を行った時系列データである．こうした時系列に沿って測定された結果には，相互に相関関係を示す傾向がある．つまり，時間軸に沿って集められたデータでは，ある時点での被験体のデータから同じ系列の次の時点におけるその被験体の状態をある程度予測できてしまうのである．そのようなデータは系列依存性があるデータといわれる．この系列依存性が，シングルケースデータの特徴であり，データの分析を困難にする原因となっている．データに存在する系列依存性の程度は，そのデータの自己相関(Autocorrelation)を調べることで推定できる．ラグ1の自己相関がもっともよく系列依存性を表すので，データの系列依存性の評価には，ラグ1の自己相関が有意な数値を示すかどうかを調べればよい(Kazdin,1982,1984)．有意な数値が得られたときは，データに系列依存性があるということになる．データに系列依存性がある場合，視覚的判断を行う際のバイアスになるし，データ相互間の独立性の仮定を満たさなくなるので，パラメトリックな統計的検定の利用は深刻な問題となる(Busk &Marascuilo,1992;Kazdin,1984).

3. ティ検定・分散分析

単一事例実験データを分析する統計的検定として初めに考えられたのが，グループ比較研究でなじみの深い分散分析法の適用である．しかし，これには直ちに激しい批判が起こった．前述のように，単一事例実験データは系列依存性を持つ可能性があり，系列依存性のあるデータは分散分析の前提条件から著しく逸脱する．すなわち，誤差の独立性の仮定を満たさなくなる．よって，単一事例実験データへ分散分析の適用は適切ではないとされた．

それでも，シングルケースデータへ分散分析を適用できるよう，いくつかの提案がなされた．

Gentile, Roden & Klein(1972)は，系列依存性を緩和する手続きとして，ABABデザインにおいて，隣り合わないフェーズであり同一の条件である，2つのA期と2つのB期をそれぞれ合計し，それらを比較する方法を提案した．Kazdin(1982,1984)は，系列依存性をラグ1の自己相関が有意になるかどうかによって判定し，有意にならなければティ検定やエフ検定を適用できるとした．Huitema(1985,1986a)は自らの調査をもとに，シングルケースデータにおける自己相関は有意なものではないからティ検定やエフ検定を利用してもよいと主張した．

しかし，こうした提案も厳しく批判されている．Gentile et al.の方法を用いても系列依存性の問題は解消されないし，ラグ1の自己相関が低い場合でも，もっと大きなラグの自己相関は高いかもしれない．そうした場合，分散分析を用いても第1種の誤りと第2種の誤りを統制することはできない(Hartmann,

NII-Electronic Library Service

1974; Kratochwill et al. 1974; Michael, 1974; Thoresen & Elashoff, 1974). また，Kazdin と Huitema の意見については次のような批判がある．有意にならないがゼロでない自己相関が存在するときは，分散分析の結果は大きく歪む．小さいサンプルサイズで自己相関の有意性検定を行っても，検定力が低く有意にならない．しかし，有意にならないからといって，シングルケースデータにおける自己相関の存在を否定することはできない(Busk & Marascuilo,1992).

4. 時系列分析(interrupted time series analysis)

単一事例実験データへの分散分析の適用は理論的に大きな問題があった．そこで，代わりに提案されたのが時系列分析(ITSA)である(e.g, Box & Jenkins, 1970;Glass et al.,1975;Hartman et al., 1980). この方法が有効なのは，データに系列依存性があっても分析を行うことができる点にある．それは，ARIMA モデルや ARMA モデルなどの時系列モデルを当てはめることにより，データから系列依存性の影響を取り除いてから検定を行うことができるためである．時系列分析の問題点は，その理論が複雑で難しいことと，適切なモデルフィットのためにはかなり多くのデータポイントが必要になることである．また，複雑なモデルほど実際のデータをよく表現するが，パラメタが多くなると，推定に必要なデータ数が飛躍的に増える．

Box & Jenkins(1970)は各フェーズに最低 50～100 個のデータがあることが望ましいとしている．同様に，Glass et al.(1975)は 50～75 個のデータが必要と述べている．最近の研究では，Crosbie(1993)がより少ないデータで適用できる時系列分析(ITSACORR)を提案しているが，それでも第 1 種の誤りを統制し，適切な検定力を得るためには各フェーズに 25 個以上のデータが必要であるとしている．

5. C 統計による処理効果の検定

時系列分析におけるデータポイントの問題を解消する方法として，Tryon(1982)は C 統計による処理効果の簡便な検定法を提案した．この方法では，各フェーズに最低 8 個のデータがあれば検定を行うことができる．C 統計の利用には批判もある(Blumberg, 1984; Crosbie,1989)が，一方で，ベースライン期に傾向がないデータに対して適用する場合には有効な手法となりうるという意見もある(河合,1996; 河合ら,1987; 桑田,1993b). その方法とは次のような手続きをとる．

まず，ベースライン期のデータから，標準化した C 統計量である Z の値を求める．この Z が有意になるかどうかを Young(1941)の表を用いて確認する．有意にならなかったら，ベースラインにトレンド(上昇や下降といったベースラインにおける傾向)はないとみなす．次に，ベースライン期と処理期の両方のデータを使って同様に Z を算出，その有意性検定を行う．ここで，有意になったら，処理の効果ありと判断する．問題はベースライン期のデータのトレンドが有意になった場合である．この場合，次のステップの解釈が困難になることを Tryon 自身も認めている．

6. ランダマイゼーション検定

シングルケースデータへのパラメトリックな統計的検定の適用については，理論的，実用的に問題があった．そこで，データ相互の独立性の仮定，分布の仮定を必要としないノンパラメトリックな統計的検定をシングルケースデータに適用することが考えられた．特に多くの研究者から推奨されてきたのがランダマイゼーション検定である(Busk & Marascuilo,1992;Edgington,1967,1987,1992; Levin,Marascuilo & Hubert, 1978; Wampold & Furlong, 1981).

ランダマイゼーション検定の理論は非常にシンプルで，条件をランダムに振り分けることだけを前提としている．この検定の目的は，処理の効果がないという仮説のもとで実験から得られた統計量が，可能なランダム振り分けの組み合わせの，それぞれで算出される統計量と比べてどれだけ極端な値かを明らかにすることにある．通常の検定と違って，母集団からのデータのランダム抽出を前提としないし，系列依存性も問題にしない．様々な単一事例実験計画のデザインに適用できるように，種々のランダマイゼーション検定が開発されている．

NII-Electronic Library Service