

測定・評価部門

研究法の理解とデータ分析における学生の誤解

鋤柄増根

(名古屋市立大学)

今回の筆者は、測定や統計処理に関心はずっとあったが、それを専門にすることはないと長い間思いながら、測定や統計処理を横目で眺めてきた。そして、中途から方向転換し、測定の分野を専門であると言いだしたし、学生に統計を教える立場にもなった、いわば駆け出しの身である。このような駆け出しの筆者がここ1年の動向をどのような観点から書いたらよいか。

まず、心理学の専門誌に掲載される論文の読者は誰かを考えたとき、我々心理学を専門とするものと同時に、心理学を学ぶ学生、そしてさまざまな現場(教育、医療、福祉など)にいる人たちである。読者はそれぞれに論文を読む目的が異なるであろうし、心理学やデータ処理に関する知識も大きく異なる。したがって、読者が論文に記載してほしいと思う情報も異なるだろうし、論文の理解の程度も異なるだろう。また、著者が訴えたいことと読者の期待の間のずれもあるだろう。さらに、このようなずれに関連する誤解や奇妙な考え方があり、これらは、データ処理についての理解の程度の問題だけではなく、心理学における研究法とその論理の理解のなさに起因する部分が随分あるように思う。

今述べたような問題点を指摘したりその理由を考えながら、2000年7月から2001年6月末までに教育心理学研究、心理学研究、Japanese Psychological Researchに発表された論文(以下では特に断らない限りこの3誌に掲載された論文を指す)を主に見ていきたい。以下で引用した論文は、それぞれ重要な知見を示しており、十分注意されたものであることを理解した上で、例として引用させて頂いた。もし、著者の方々が不快な気持ちになったとしたら、それは本論の筆者の至らなさであるので、どうぞ許していただきたい。

学生のお手本としての論文

心理学を学ぶ学生や心理学の周辺領域にいる人々にとって、掲載された論文の役割が知識を得ることのためだけでなく、卒業研究などのお手本でもある。私の大学で卒論に取り組んでいる学生の姿だけかもしれないが、お手本とした論文のトナーが剥げそうになったコピーを1つだけ片手に握りしめて、そこに書かれている分析を

どれもこれもやろうとしている。ここで問題になるのは、論文にはすべて書かれていないということである。しかし、学生や初学者にとっては、論文に書かれていることがすべてである。したがって、学生や初学者は、論文に現れていないデータ分析、つまり論文で示されている分析をする前になされている予備的な検討や試行錯誤的なデータ分析があることに気づいていないし、その論文がなぜそのような研究法や分析法を採用しているのかについて理解が不十分なまま、卒論を実施してしまうようである。この原因の最大のものは、データを自分の目で見ないということにあると考えられるので、まずその点を述べていく。

データを自分の目で見ること

調査研究では因子分析などの多変量解析を実施することが前提であり、簡単に統計パッケージが使える現状では、数学が苦手な多変量解析の原理もよく分からない学生であっても、多変量解析をやらないわけにはいかないし、やればできてしまい、うまくいけばお手本と同じような結果を手にするができる。極端なことを言えば、データの入力さえすれば、何も分からなくても、専門誌に掲載された論文と同じことができる。

このような現状が、自分のデータを自分の目で見ないで、統計パッケージをとおしてのみ見る傾向を助長しているように思われる。また、既存の統計パッケージでできないことはやらない、あるいはやらなくてよいと考えているようである。

ちょっとした工夫で、随分いろいろなことが分かる例として、情報処理の基本である並べ替えを考えてみたい。行に人、列に項目(できたかできないかの2値をとるとする)という行列があり、それぞれの合計点(個人の得点と項目の通過率)があったとき、個人をその得点で並べ替えることはよく行われる。ここでさらに項目の通過率で並べ替えると、項目がガットマンスケールになっているかどうか比較的に分かる。つまり行の合計と列の合計の両方で並べ替えをすることで、データの構造がある程度はつきりしてくる。ところが統計パッケージ(例えばWindows上で比較よく使われるSPSS)では、行のソートは可能であ

るが、列のソートはできないものが多い。この点は改善すべき点だと思う。ところが、Excelでは可能である。

自分の目でデータを見るための方法として、データをグラフ化することはもっとも重要な方法であろう。簡単なものは散布図やヒストグラムを描くことである。もっとも高度なものには射影追跡といわれるものがある(岩崎, 1990)。このようなデータのグラフ化あるいは図示をもっと気楽にやるべきであるという指摘は、パソコンが普及しだした頃から多くいわれてきているが、統計パッケージが普及することで、データを数値としてみて、グラフ化することがかえってなくなっているようである。グラフ化については、Wainer & Velleman (2001) が、データのグラフ化あるいは図示がデータ解析そのものになることを述べ、グラフ化の復権を主張している。

しかし、このようなグラフ化や記述統計の検討さらには並べ替えのような操作は、予備的な分析に使われても、最終的な分析結果にはなりにくく、専門誌に掲載された論文からはこのような過程の苦勞が、なかなか学生には伝わっていない。したがって、“As soon as you have collected your data, before you compute any statistics, look at your data.” (Wilkinson & TFSI, 1999, p.597) ということのを忘れ、いきなり統計パッケージにデータを打ち込み、解析結果を手にすることが多く、それで論文が書いてしまうことに問題がある。

このような傾向の中で、自分のデータを自分の方法で見たいこうとすることの重要性を示す論文として酒井(2001)のものがある。酒井はSprangerの価値概念を測定する尺度を作成する中で、因子のまとまりが研究ごとに異なる理由として、Sprangerの各価値概念そのものについての考え方の個人差が関与している可能性を示している。一般的には、尺度構成をするとき、このような個人差を産みだす項目や、うまく因子にまとまらない項目は測定しようとしている構成概念を測定するには不適切であると捨ててしまう。そのようなことをせずに、一般の被験者とは異なる回答をした被験者を追跡して内省報告をていねいに検討する中から、価値概念についての考え方の個人差を検討しているのは、学生のお手本として非常に有用である。もっといえば尺度を作ったらそれで終りというようにやりっ放しにしてない点が評価できる。また、一般的な法則にあてはまらない個人を追跡することは、新たな発見への手掛かりを提供する可能性もあり、非常に重要な点である。

学生の誤解

このようにデータを自分で見て自分の観点や自分なりの工夫で見ようとしなことが、以下に代表されるいく

つかの誤解を学生に持たせるようである。まず始めに、データ分析法の誤用というより、誤用以前の問題、すなわち研究法や推測統計の考え方の不十分な理解からくる奇妙な統計の使い方や推論の仕方を述べ、それらに関係づけながら、卒論のテーマを選択するときに持ちやすい誤解を述べる。

統計処理に関する代表的な誤解

よく体験する統計に関する学生の誤解の主なものを4つあげてみたい。1 相関係数は高いほどよい、2 分散(標準偏差)は小さければよい、3 分布は正規分布でなければならぬ、4 p 値は小さければよい、という4つの誤解である。

相関係数は高いほどよい

1番目の誤解は、当然のことながら、ちょっと考えれば、おかしい場合があることは分かるものである。例えば、テストにおいて2つの項目間の相関係数が1.0であれば、まったく同じことを質問しているのだから、どちらか一方のみを質問すればよいことになる。実際には.9くらいの相関であっても、心理学における測定誤差の大きさを考えればほぼ1.0と同じと考えてもいい。

一般的にある概念をより明確に測定したり、あることをよりよく説明するために、測定変数を増やす。これは、今までの測定変数群では測定できなかったものが測定できるという仮定に基づいている。つまり、測定変数を増やすことで、得る情報を増加させることにその目的があるといえる。2つの変数の間に相関があることは、片方の変数の値が分かれば、もう片方の変数の値をある程度予測できるということである。逆に、相関がないということは、2つの変数のいずれが分かっても、もう片方の変数についての予測はできないということである。したがって、相関がない2つの変数から得られる情報量は、相関がある場合に得られる情報量より多いといえるので、あまりに相関が高い場合、情報は少しも追加されないことになる。ただし、今まで関係があるかどうか分からなかった変数間に高い相関があることを見いだすのは重要なことである。

ここで、問題となるのがKline (1994, Pp.128-129) がテスト項目の作成に関して指摘しているパラフレーズの問題である。つまり同じ内容の言い換えの問題である。もっとも極端なものは「私は明るいです」に対して「私は明るくありません」のように否定肯定の関係にある項目である。もう少しありそうな項目対では「私は人前で話すのが苦にならない」と「私は人前で話すのは苦痛です」のようなものである。このような項目が複数あると、こ

これらの項目が相互に高い相関をもち、因子分析をしたときに、見掛け上の因子として抽出されることになる。

例えば、黒田・桜井 (2001) の目標指向性尺度の項目で「自分の性格の悪いところを友だちにみられないようにしています」「自分の性格のいやなところをかくそうと努力しています」「自分のいやなところを友だちに知られないように努力しています」などは、パラフレーズの可能性がある。藤井 (2001) の心理的対処の仕方における「相手が何を考えているのかが気になって動けなくなる」と「相手はこう考えるのではないかと、あれこれ思い悩んでしまう」などがパラフレーズの可能性としてあげられる。

実際にパラフレーズかどうかを決めるのは非常に難しいし、なんらかの目的(例えば、回答の一貫性のチェックなど)がある場合には、パラフレーズされた項目が必要になることがある。しかし、特に意図されていない場合は、項目間の相関が異常に高い、例えば.9以上あるようなら、パラフレーズの可能性を疑うべきであろう。したがって、項目間相関を紙面が許すなら載せる必要があると思う。

分散に関する誤解

「分散は小さければよい」という考えは、個人差測定という観点からいえば、個人差成分による測定値の変動は大きい方が、個人をよりよく弁別することが可能になることから必ずしも正しいといえないことが分かるだろう。一方、実験研究は個人差を誤差とみなすので個人差による分散は小さい方がよいと考えられる。いずれの場合も測定誤差は小さい方がよいことはいうまでもない。この個人差測定と実験研究の両者の発想の違いを混乱していることに、「分散は小さければよい」という誤解の主な原因があるといえる。

「分布は正規分布でなければならない」という誤解に関しては、「…単峰型の分布であることから、いずれも正規分布とみなした」(佐々木, 2001, p.76)という記述を読んだ学生、あるいは初学者は「なるほど正規分布とみなせてよかったのだ」という印象をもつはずである。もちろん、著者はそんなつもりは全くないのだが。

この印象は、心理テストの観点からいえば、テスト得点の分布はそのテストが測定しようとしている構成概念(潜在特性)の分布と同じではないことをよく理解していないことからくるものである。テスト得点の分布は、そのテストの項目の選択の仕方によって、大きく変わってくる。理論的には、こちらの思いどおりのテスト得点分布を得ることが可能である。臨床診断のように弁別のための目的であれば、あるカットオフ得点の位置に谷ができるような2峰型の分布が望ましいかもしれない。また、能力あるいは特性値の高いところでの個人の弁別がはっ

きりするような分布、すなわち高いところに尾をひくような分布が必要かもしれない。当然であるが、正規分布すると仮定される潜在特性の分布をテスト得点が再現してほしい場合もあるだろう。このようなことは、項目の選択しだいで、原理的には可能である。この点を可能にする理論的な根拠が、項目反応理論 (IRT: Item Response Theory) である。

p 値は小さければよい

p 値は小さければよいという誤解としてまとめたが、有意性検定の考え方の理解不足からくる誤解にはいくつかのものがある。Cohen (1994) を始めとする多くの論文が指摘している。そして最近では Nickerson (2000) がその誤解をリストアップして論じているので、言い尽くされていることではあるが、その代表的なものをあげておきたい。

最近になっておそらく増えてきただろうと思われる初歩的なものとして、 p 値が計算の結果得られた F 値や t 値そのものの生起確率を表していると考えられる誤解である。実際には、計算で得られた F 値や t 値(片側検定とすれば)以上の値をとる確率である。多くの統計パッケージが F 値や t 値について p 値を正確に計算するので、このような誤解が生まれるのだと思う。

これに関連して、この1年の論文では見られなかったが、統計パッケージが計算した p 値を記載するだけで、有意水準を記載していないことがある。もちろんこれで十分といえるが、事前に有意水準を決めた上で検定がなされるという有意性検定の論理を忘れてしまった結果としてこのような記載をしている可能性があるとするれば、これは重要な問題である。このことは、ときどき目にする「1%水準では有意でなかったが、5%水準では有意になった」という記述にも現れている。

これらの誤解は、 p の値が効果の大きさ (effect size: 平均値の差や相関係数から推測される処理効果の大きさ) を表している、すなわち p 値が小さければ効果の大きさが大きいと考えてしまうという有意性検定に関する最大の誤解を助長している。例えば、2つの相関のいずれが大きいかを、相関が0でないという帰無仮説のもとで得られた p 値を比較して、こちらの相関の方が大きいという結論をくだすようなことにそれは現れている。

p 値が効果の大きさを表していないのは、相関係数の有意性検定の p 値が標本の大きさに依存することを思いだせば理解できる。また、同じ平均値の差であっても、誤差分散の大きさや標本の大きさによっても異なってくる。このように、 p 値は実質的な効果の大きさだけでは決まらないものである。

したがって、APA (2001) の執筆要綱は効果の大きさや区間推定の結果を提供することを勧めている。しかし、この1年間に掲載された論文には、1つもこの区間推定や効果の大きさを提供している論文はなかった。区間推定を実施しない理由の1つには、区間推定をしてもそれをどのように解釈すればよいのか、ほとんどの統計の教科書に記述がないことに原因があるかもしれない。この解釈は以下のように率直に考えればよいだろう。例えば相関係数であれば95%の信頼区間に0が入っていない、つまり95%の信頼区間が正の範囲あるいは負の範囲のみかかっている場合は、相関係数は統計的に0でないと考えればよいだろう。また、差の比較なら、95%の信頼区間に0が入っていないければ、どちらかの群の平均が大きいと考えればよいであろう。

効果の大きさの報告が必要な研究は、Chow(1996)が研究目的を以下の4つに分けているうちの2番目のものである。1番目はある現象を説明する理論を構成する研究、2番目が実用性研究であり、ある処理(教育方法など)の実用性を検討するものである。3番目が、すでに有効であることが確かめられている理論のもとで、ある現象や特性が現れるかを検討する研究である。最後が一般化研究といわれ、ある理論が他の対象や状況でも有効かどうかを検討するものである。

Chow (1996) の研究目的の2番目に分類される実用性を確かめる研究では、現実場面に研究成果を適用しようとしているので、ここで重要なのはどれだけの効果があるのかである。したがって、必ず効果の大きさが分かるような情報を報告する必要があるはずである。そして、その効果を産みだすためにかかる費用、すなわち費用対効用を検討することで、研究成果を現実場面に利用するかどうか決めることになる。教育心理学会の構成員は従来のように大学の教員だけでなく、現職の小中高の教員が増加している現状から考えると、何か教育現場に還元あるいは応用可能な知見を提供できる研究があると思う。

ただし、すべての研究に、区間推定や効果の大きさの報告が必要ではないといえる。研究目的の1, 3, 4番目に分類される実験の場合には、データとして得られたものがランダムな変動で説明できるかできないかの2分法的な決定ができる有意性検定のみで十分といえる(Chow, 1996)。つまり、ある理論の予測を支持する結果かどうかは、平均値の差の大きさとは関係ないと考えられることはできる。平均値の差が大きければその理論がより強く支持されるということではないはずである。

この1年の論文を、Chow(1996)の研究目的にしたがって分類したところ、2番目の実用性研究はつぎのような

数になった。なお、この分類は筆者がひとりできっと目的などを読んで分類したもので、それほど正確なものでないことをお断りしておく。心理学研究では、49編中1編もなかった。ただ、玉井・中島・北口・今田(2001)のものが恐怖症の治療に言及していることから、無理をすればこの分類に入れることができる程度であった。Japanese Psychological Researchでは24編中2編が、教育心理学研究では42編中6編がこの分類であった。実用性を確かめる研究に分類できた。ただし、実践研究は対象としなかった。岩男(2001)、山崎(2001)、佐々木(2001)、竹内・犬上・石原・福田(2000)、岡安・高山(2000)、桑原(2000)をこの分類に入れた。これらはいずれも、臨床場面や教育場面への応用を視野に入れた議論をしているので、効果の大きさあるいはそれを計算できる情報を記載するべきである。

しかし、実用性を検討していると分類された論文にも必要な情報が載せられていない場合が多い。例えば、漢字学習におけるイメージ媒介方略の有効性を見だし、非漢字圏日本語学習者の漢字学習との関わりを考察している桑原(2000)でも、イメージ媒介方略とそれを利用しない単純なりハサル方略の各条件での平均と標準偏差が必要であるが、その値は直接示されていない。今後このような情報は積極的に記載すべきではないだろうか。

また、メタ分析をする上でも基礎統計量や効果の大きさは必要な情報である。メタ分析を利用したレビュー論文が日本ではまだあまり見られないが、Mullen(1989)を翻訳したものが出版されたので、この様な分析を利用する研究者が増えるであろう。

さて、 p 値に関する誤解のもっとも重要な点は、 p 値は $P(D|H_0)$ であることが理解されていない点である。つまり、 p 値は、帰無仮説が正しいという仮定のもとで、そのデータ (D) が得られる確率であり、そのデータ (D) が得られたときに帰無仮説が正しい確率 $P(H_0|D)$ ではないことである。ましてや p 値は $1 - P(D|H_1)$ でもないし $1 - P(H_1|D)$ でもない。つまり、1 から p 値を引いた値が対立仮説 (H_1) の正しさではないということである。

さらに、有意性検定では、帰無仮説を棄却したからといって特定の対立仮説を支持したことを意味しない。対立仮説は無数にありうるし、有意性検定で直接問題にするのはあくまでも帰無仮説である。これは研究法における、研究仮説、実験仮説、統計仮説の区別の曖昧さからきている。実験仮説では「実験群の平均が統制群より高い」と予測しているにもかかわらず、有意性検定が直接検証する統計仮説は「実験群と統制群の間に平均値に差がない」(帰無仮説)である。この両者の違いは大きいにもかかわらず、これをあまり意識せずに結論をくだ

すことが多い。

また、帰無仮説(H_0)が正しいなら、何度も実験を繰り返せば、そのデータ(D)が起きる確率であると考えすることは一定の条件下でのみ成り立つということである(Nickerson, 2000)。「まったく同じ実験条件」で実験を繰り返すことが可能であるとするなら、そのように考えることは可能であるようである。化学の物質合成の実験では、まったく同じ温度で圧力でというように「まったく同じ実験条件」を再現することが可能であるが、心理学実験の場合には統制不可能な部分をもった人間が条件の中にはいてくるので、「まったく同じ実験条件」を設定することは不可能であるといえる。この点は、物理や化学の研究と異なる心理学研究特有の再現可能性における困難さである。また、心理学実験の場合、実験群と統制群との間に平均値に差があることは再現することができるが、何点の差があったかまでは再現できないのが普通である。

分散分析の誤用と実験仮説

研究法そのものの理解不足からくると思われる、かなりよく見られる分散分析の奇妙な使い方もしくは誤用について述べていく。

学生は、 t 検定も1要因分散分析も群間の平均値の差が統計的に有意かどうかを検定するものであると理解していることが多い。しかし、 t 検定も分散分析も同じ従属変数すなわち同じ物差上の変数の平均値の差を検定するものであり、反応時間と誤反応率とを比較するものではない。つまり、何でもいからいくつもある平均値の違いを統計的に検定する道具ではない。比較できないものを比較してはならない。

したがって、因子分析の結果得られた因子から構成された複数の尺度値の平均の差異を検定するのに利用するのは適切ではない、もっといえば比較しようとしてはならない。例えば、榎本(2000)が、青年期の友人関係において重視している欲求を比較するのに、3つの欲求、すなわち親和欲求、相互尊重欲求、同調欲求の各平均の違いを、一要因の分散分析を実施して検討している。因子分析の直交回転によって得られた3つの欲求は比較可能なものではないといえる。では、このような場合にどういう検定をすればよいのかと質問されるかもしれないが、比較できないものを比較する方法はないとしかいえないのではない。このような誤用は学生には当然よく見られる点である。これは結局統計の理解の問題ではなく、研究法の理解の問題であるといえる。

つぎに分散分析における多重比較について、実験仮説のたて方との関連で述べておきたい。この1年間の論文で分散分析を実施したものは76編あった。そのうち要因

配置が 2×2 のように多重比較が必要のない場合と多重比較をしていない論文を除いた42の論文がなんらかの下位検定を実施している。この中で多重比較を行ったとしか記述のないものが8編もあったのは問題であろう。また、これは単純なミスであろうが、分散分析の結果に自由度の記載のないものが1編(辻井, 2001)あった。このような情報は必ず必要となるので、忘れないで記載してほしい。

さて、一般に分散分析の後の多重比較は、事後(post hoc)検定であり、主効果あるいは交互作用が有意になった後に、さらに分析をするためのものである。要するに、どこに差があるのか分からないが、とにかく分散分析を実施したらなんらかの効果が有意になりました。では、さらに細かく条件間の比較をして、差のあったところを探しましょうというやり方である。この方法は探索的な場合にはよいが、条件間にどのような差があるのかが明確になっている実験仮説を検証するために実験を実施する場合にはあまり有効な方法とはいえない。このような場合には、コントラスト比較(Rosenthal & Rosnow, 1985; Rosenthal, Rosnow & Rubin, 1999)を使うべきである。コントラスト比較のもっともよく利用されるものは傾向検定といわれるものである。この傾向検定については、後藤・大野木・中澤(2000)の「心理学マニュアル 要因計画法」に詳しい。

この問題は、 t 検定において、両側検定をするか片側検定をするかの選択と同じことである。つまり、2群の平均値の差の方向が予測できなければ、「実験群の平均値と統制群の平均値のどちらが高いか分からないが、とにかく差がある」という仮説を検証するために、両側検定で t 検定をすることになる。しかし、「実験群の方が統制群より平均値が高い」というような方向性のある実験仮説があるのなら、片側検定をすればよい。

t 検定の両側検定に対応するものが、普通に行われている分散分析であり、オムニバス分散分析といわれる(ところで、分散分析に片側検定がないのを不思議に思った方はいないだろうか)。オムニバス分散分析は、どこに差があるか分からないけれどとにかく差があるということを検定するわけだから、すべての平均値間の比較パターンを考慮に入れた分析をすることになる。したがって、狙った差のパターンのみを検定可能なコントラスト比較(t 検定における片側検定に対応)をすることは、差を検出しやすくなるので、是非利用すべきである。有意な結果が得られなくて投稿をあきらめた研究も生き返るかもしれない。

ただし、このコントラスト比較をするためには根拠の十分な実験仮説を事前にたてられ、条件間の差のパターンが十分な根拠をもって予測できることが必要である。こ

のような要件は実験研究では十分成り立つのではないだろうか。実際に方向性のある予測をたてている研究はいくつもあった。例えば、中本(2000)や高岡(2000)などは、平均値がどの条件で高くなるなど、特定のパターンを予測しているので、このようなコントラスト比較をすることが可能である。他の論文でも目的でかなり突っ込んだ議論をしており、方向性のある予測が可能であるにもかかわらず、「関連があるかどうか」あるいは「平均値に差があるかどうか」を検討するという謙虚な予測しか述べないことがある。もう少し積極的な予測を述べたことをしてもいいのではという印象をもった。また、一般に相関の有意性検定も、ほとんどの研究が無批判に両側検定を利用しているが、今述べたように予測があるのなら、片側検定であってもかまわないはずである。

また、要因が連続変量と考えられる場合(図形の複雑さや刺激の輝度レベルなど)は連続変量である独立変数と連続変量である従属変数との関数関係を知ることが主眼であることが多い。したがって、傾向検定を実施すべきである。多重比較をただけでは関数の形は特定できないが、傾向検定を実施することで関数関係が明確になる。この場合、傾向検定が可能ないように要因のレベルは等間隔に条件設定しなければならない。つまり、実験を始める前に分析方法を決め、それに合わせて実験条件を設定しなければならない。このように、実験の要因配置計画とその後の分散分析の方法とは密接に関連しているものであり、何でもいから実験をした後で、どんな分析をすればよいか考えるものではないことをよく理解すべきである。ましてや、たまたまあるいくつかの平均値の差を検定する道具として分散分析があるのではない。

このように、実験計画と分散分析に代表されるデータ分析法とは密接な関連にあり、方法論について十分な理解を必要とするものである。つまり、目的、仮説、条件設定、実験操作、結果の分析が相互に緊密に結びついていることが十分理解されていない。また、アイデアを実験にのせるまでの過程も非常に重要である。したがって、適当に実験をしておいて、さあどうやって分析をするのかと質問されても困る。このような方法論、特に実験法の適切な書籍として小牧(2000)のものがあげられる。また、同じ小牧(1995)は日本の教科書ではあまり触れられない分散分析の要因が変動要因(random factor)であるか固定要因(fixed factor)であるかについて詳しい説明がある。

独立変数と従属変数の区別

上で述べたような研究法における理解の問題は、独立変数と従属変数の区別の曖昧さにも現れてくる。特に調

査法を使った卒論ではこの曖昧さをよく目にする。

独立変数は原因であり、従属変数は結果であると考えられると、独立変数は従属変数より時間的に先行すべきである。この時間関係は、実験研究では、独立変数は実験条件として実験者が事前に設定し、実験結果として従属変数である測定値が得られる。したがって、原因(独立変数)が時間的に先行し、結果(従属変数)がそれに続いて起きるという因果関係における時間関係が、実験研究では再現されている。さらに、従属変数は測定されるものであり、独立変数は実験操作によって実験者が設定するものであり、研究計画上も研究者の意識の上でも、全く異なるものであることが明瞭である。

ところが、ほとんどの調査研究では、従属変数と独立変数が同じ方法(調査)によって同時にデータとして得られる。例えば、親の養育態度(独立変数)が大学生の自己効力感(従属変数)にどのように影響しているかを検討するとき、いずれの変数も1つにまとめられた冊子を使って調査される。したがって、独立変数と従属変数の時間関係や別のものであるということが研究方法の中で曖昧になってしまっている。このことが、学生に何が従属変数であり何が独立変数であるのか、そして、この両者を区別しなければならないことに対する意識が希薄になっている。

このようなことを考えると、学部学生の指導としては、たとえ調査法でしかできないようなテーマに強い興味を持っている学生に対してであっても、いきなり調査法で卒論を書かすのではなく、なるべく実験をさせるようにした方が、独立変数と従属変数の区別を明確に意識させるという意味ではよいといえる。

研究をする目的が統計的有意差を出すことか

専門紙に掲載される(すなわち採択される)論文は、特別な場合を除いて、有意差もしくは有意な関係があるような予測を立て、統計的に有意な結果を得ている。このような統計的に有意な結果が必要である理由はつぎの2点になるだろう。

まず、差異がないことを積極的に主張する研究は大変難しい。この点は、図書館で探している本がないことを証明することは難しいが、あることを示すのは簡単であることに対応している。条件間に有意差が得られなかった場合に、標本数が少ない、実験条件がうまく統制されていない、測定の信頼性が低いなどによって検定力が低くなっている研究計画でないことを証明しないと、はっきり有意差がなかったと主張できない。つぎに、2つの対立する理論があったときに、それらが同じ予測をするのなら、それらの理論間に違いはないのだから、別の理

論を考える意味がないことになる。このように差のないという予測そのものがあまり意味のないことが多い。

しかし、有意な結果が得られなかったときの、研究指導は検討すべきである。なぜ有意な結果が得られなかったかを検討させるべきである。ところが、学生はなんでもいいから有意な結果を今あるデータから得ようとすることがある。特に、調査研究であると、相関係数がいくつもあるのでどれか1つぐらいは有意になったりしている。したがって、有意になった部分で話を作ってしまうようなことがある。これは、有意性検定を同時に大量にやれば、まったくランダムなデータであっても、いくつかは有意な結果が得られるということに対する理解のなさからきている問題である。実際に、この1年の論文でも、岡安・高山(2000)がボンフェロニの修正をしているだけで、このような点にあまり注意が払われていないのは残念である。

また、調査研究で有意な相関が得られなかったとき、学生は被験者の上位4分の1を上位群、下位4分の1を下位群として、群間に有意な平均差が得られるだろうと分析のやり直しをしようとする。しかし、相関が有意にならなかった場合は、たいてい群分けしてもその平均の間に有意な差はでてこない。むしろ、中位の2分の1の被験者を捨てるわけだからデータの情報量をわざわざ減らしていることになる。

このように尺度値で群分けすることについては実践的な問題がある。2つ以上の尺度値でHH群HL群などと群分けしていくと、ある群の人数が思いもかけず少なくなってしまうことがある。なんとなく均等に人がいそうな気がしているが、そうでもない。1000人位の被験者がいても、3つの尺度で群分けをしてくると、相当少ない人数になる群ができて、群の人数が非常にアンバランスになることがある。

このようなデータに分散分析を実施したとき、SPSSは何の警告も出さないうで結果を出してしまう。極端な場合は、セルの人数が0のところがあっても平気で計算してしまう。もちろん、計算可能であることは分かるが、特に警告も出さないうで結果を出すのは、初学者にとってあまりにも不親切だと思う。

また、尺度値で群分けすることは、Meehl(1967)が被験者の特性に関する変数と実験者が操作する変数に分けたことに対応する問題でもある。前者の被験者変数による群分けは、分ける基準となる測度によっては測定誤差が大きく、要因配置計画における要因となりにくい。例えば、男女のように明確に区別できる指標なら被験者変数は意味があるが、自尊感情のような指標では、測定誤差がかなりあり、一般の実験条件のように明確にその条

件設定ができたことにならない場合が多いといえる。

基礎統計量の表示

最近の論文は、いきなり因子分析の結果から書き始めてあるものも多い。この1年でも、因子分析あるいは主成分分析を使った研究は、教育心理学研究で17編、心理学研究で9編あり、それぞれ12編、8編が結果を因子分析から書き始めていた。著者たちはちゃんと基礎統計量を算出し、それらを検討しているが、紙面の量的な問題でしかたがなく論文には載せていないことを我々は分かっている。しかし、学生はこれを読んで卒論を書くとき、やはり因子分析から書いてしまい、項目ごとの平均などの基礎データを求めもしないことになる。

この問題は先の統計パッケージを通してしかデータを見ない傾向とあいまって、因子分析をする過程には相関係数を求めている過程があることを忘れさせている。項目が評定尺度のとき特定の評定段階に回答者が偏っていることで不当に相関係数が低くなることもある。したがって、各項目ごとの評定値の分布状況を把握しておくことは重要である。また、一般の統計パッケージがこの相関係数を計算するときにピアソンの相関係数を計算しており、他の相関係数を選択できないのは困ることがある。回答が2値の場合にテトラコリック相関を使いたくてもできないパッケージがほとんどである。

実はこのような基礎データは、論文のデータを再分析したりメタ分析をする上でも大変重要なものである。また、実験研究では、その多くがデータを図で示しているので、読んでいる方も分かりやすくよいのであるが、そのデータを再分析しようとするときに平均や分散が正確に分からず困ることがある。

このような基礎データは、掲載誌の紙面の量的な都合ですべて載せることは無理なので、なんらかの形で学会がデータベースとして持つことができないだろうか。それを、インターネット経由でアクセスすることが可能になっているとよいかもしれない。もちろん著作権などを配慮する必要があるのでIDやパスワードが必要になるのは当然であるが。

複雑な分析より簡単な分析

お手本とする論文と同じ分析をしようとするのはいいのだが、自分の理解を超えた難しい分析をやろうとしようとしても多々ある。また、初学者は論文がそのような複雑な分析をやっているのを見ると、それをしないとなんらかの結論が導けないかのような錯覚をしてしまう。しかし、複雑なデータ処理より簡単なものが有効なこともあるし、簡単な分析で十分なこともある。

例えば、宇井・松井・福富(2001)が数量化Ⅲ類の POSA を使用して、尺度値のパターンと学年変化を示しているが、 χ^2 検定のみでも同様の結論が導けるのではないか。宇井らの結論を明確に言うためには縦断的な研究も必要であろう。もちろん、このような分析をすることでデータそのものから、ボトムアップ的に結論づけることが可能になることは理解できる。しかし、初学者はこのような分析をしなければ、結論がくだせないかのような錯覚をもってしまう。

このように新しく複雑な分析方法が普及してくればくほど、多くの学生は複雑なものをしなければという強迫的な観念をもつことになり、生半可な知識で分析をし、本来やらねばならない自分の目で自分のデータを見て、それを自分なりの工夫でまとめていくということを、どんどんしなくなっていくことになるのだろう。

このような分析の高度化と複雑化は、今や心理学においてなくてはならない因子分析においてもみられる。確証的因子分析の進歩は大きいといえるであろう。それと同時に因子の斜交回転の利用頻度も多くなっている。また、柳井(2000)も「まず、斜交回転を行い、因子間の相関が低い場合に限って直交回転をすべき」(Pp.97-98)としている。この議論が、確証的因子分析において正しいことは認めるが、初学者が一般に使用する探索的因子分析の場合にこのようなことを勧めるのはどうであろう。

Nunnally & Bernstein (1994) は、探索的因子分析では直交回転の方を中程度によいと考えている。その理由として、1 直交回転は数学的に単純である、2 因子の数や種類について、直交回転による結論と斜交回転によるものがよく似ていると多く事例が示している、3 直交回転よりも斜交回転によって簡単にだまされる。最後の点は、斜交回転が数学的にはデータをよく説明するので、データへの適合がよいことから、斜交回転の方がよりよいと簡単に思ってしまう点を指している。

探索的因子分析は、データの探索的な分析であるから、まず、データから得られる情報量が最大になるような結果を求めるべきである。先に述べたように、相関がない因子から得られる情報量は、相関があるときより多いことを思いだせば、直交回転すなわち因子間に相関がないという仮定の下で結果を分析した方がより多くの情報が得られることになる。

複雑な条件設定

テーマ選択あるいは研究計画をたてる段階で、学生は研究をどんどん複雑にしてしまうことがよくある。これは、卒論になんらかの独自性が要求されたとき、学生は、お手本とした論文が扱っていない変数をいれたり、研究

計画をさらに複雑にすることで対応しようとする。この傾向は、変数を増やしても、先行研究の知見はそのまま再現され、さらに何か新しい知見が加算されるという勘違いに基づいている。

具体的には、重回帰分析で変数を増やしたり、投入する変数群を変えると、それまで有意になっていた変数の回帰係数が有意でなくなってしまうことがあることから、勘違いであることが分かる。また、重回帰分析におけるステップワイズ法による変数選択において、つぎに投入される変数はそれまでに投入された変数群にとって最適のものであり、今投入された変数が残りの変数の中でもっともよく従属変数を説明する保証はないという問題点 (Thompson, 1995) からいえる点である。

また、尺度を集めた心理測定尺度集 1~3 (堀, 2001) やインターネットで検索可能な尺度のデータベースがあり、尺度を簡単に手にすることができるので、調査研究では、変数を増やすことは比較的簡単である。こうして尺度が決まれば、被験者さえ集めればデータは簡単にとることができる。ただし、最近は調査実施における倫理の問題などで学校現場での被験者を集めることが難しい状況にあり、それほど標本数はとれない。そこで、それほど多くない大学生標本で調査が行われ、変数のみ増えるという卒論になってしまう。

このように大学生を被験者にしたとき、測定内容(例えばストレス)によってはその分散がかなり小さくなる可能性がある。また、安定した結果を得るためには、変数の数の少なくとも 3 倍、理想的には 5, 6 倍の被験者を必要とすることを考えると、このような研究は、たいいてい思うような結果が出ないことになる。

最後に、近年盛んに研究されており、学生も卒論に取り上げることが多いストレス研究を例にあげて、学生はどんな研究計画をたてるのかシミュレートしてみよう。

まず、ストレス研究において取り上げるべき変数は大きく 3 つの変数、すなわち、ストレス反応、ストレスサ、緩衝変数であり、この緩衝変数はさらに、ストレスへの対処行動(コーピング)、個人特性(性格や欲求)、サポートに分けられ、その 5 つの変数が必要であることを概論書などでまず知る。

そして、この 1 年のストレスに関連した 6 件の研究(田中・中澤・中澤, 2000; 岡安・高山, 2000; 佐々木, 2000; 杉浦, 2001; 加藤, 2001; 周, 2000)を見て、今述べた 5 つの変数をすべて入れた研究も潜在構造分析をしている研究もないことを知る。そこで、学生は、卒論にすべての変数を入れ、潜在構造分析をしようとする。研究計画段階で各変数についてもいくつかの側面があり、それぞれを測定するのに適切な尺度も違うことが分かってくる。さらに、

変数がかなり複雑に関連していることが分かってくるので、卒論の締め切りの時間を考えていくと、結局、明確な予測が立てられないという中途半端なままで、何でもいからたくさんの変数をとっておけば、何か結果がでるだろうと考えてしまうことになる。そして、データを統計パッケージに入れて、マウスを適当にクリックすれば、何か新たな発見やとても重要な知見が得られると勘違いして研究を行うことになる。

実際には、ストレスを引き起こしたりそれを緩和したりする要因はおそらく手に負えないくらい複雑であり、それを適切に解析する手法も少ないと考えられる。1つには、交互作用をうまく扱うことである。先にあげた、6件の研究を見てみると、杉浦(2001)が性格特性によってとるコーピングスタイルが変わるという交互作用を検討している。ここで、ストレス研究でよく利用される重回帰分析や潜在構造分析において、このような交互作用を変数に入れた分析(Jaccard, Turrisi & Wan, 1990)を試みることをしてもよいだろう。あるいはこのような交互作用を直接測定してしまうことが可能ならばしてもよいであろう。また、変数の間に循環的な関係、相互に促進的に影響を与えたり抑制的な影響を与えたりするときどのように分析すればよいのだろうか。これをするには学生には時間も能力もなく、実際には有意になった結果のみで、なんとかして卒論を書いてしまうことになる。

引用文献

- American Psychological Association 2001 *Publication manual of the American Psychological Association*. 5th ed. Washington, DC : American Psychological Association.
- Chow, S. L. 1996 *Statistical significance : Rationale, validity and utility*. London : Sage.
- Cohen, J. 1994 The earth is round ($p < .05$). *American Psychologist*, **49**, 997-1003.
- 榎本淳子 2000 青年期の友人関係における欲求と感情・活動との関連 教育心理学研究, **48**, 444-453.
- 藤井恭子 2001 青年期の友人関係における山アラシ・ジレンマの分析 教育心理学研究, **49**, 146-155.
- 後藤宗理・大野木裕明・中澤 潤(編著) 2000 心理学マニュアル 要因計画法 北大路書房
- 堀 洋道(監修)山本真理子(編) 2001 心理測定尺度集1 サイエンス社
- 堀 洋道(監修)吉田富二雄(編) 2001 心理測定尺度集2 サイエンス社
- 堀 洋道(監修)松井 豊(編) 2001 心理測定尺度集3 サイエンス社
- 岩男卓実 2001 文章生成における階層的概念地図作成の効果 教育心理学研究, **49**, 11-20.
- 岩崎 学 1990 射影追跡と多変量データ解析 柳井晴夫・岩坪秀一・石塚智一編 人間行動の計量分析 東京大学出版会
- Jaccard, J., Turrisi, R., & Wan, C. K. 1990 *Interaction effects in multiple regression*. Newbury Park, CA : Sage.
- 加藤 司 2001 コーピングの柔軟性と抑うつ傾向との関係 心理学研究, **72**, 57-63.
- Kline, P. 1994 *An easy guide to factor analysis*. London UK : Routledge.
- 小牧純爾 1995 データ分析法要説—分散分析法を中心に ナカニシヤ出版
- 小牧純爾 2000 心理学実験の理論と計画 ナカニシヤ出版
- 黒田祐二・桜井茂男 2001 中学生の友人関係場面における目標志向性と抑うつとの関係 教育心理学研究, **49**, 129-136.
- 桑原陽子 2000 非漢字圏日本語学習者の漢字学習におけるイメージ媒介方略の有効性—漢字と英語単語の対連合学習課題による検討— 教育心理学研究, **48**, 389-399.
- Meehl, P. E. 1967 Theory-testing in psychology and physics : A methodological paradox. *Philosophy of Science*, **34**, 103-115.
- Mullen, B. 1989 *Advanced BASIC Meta-Analysis*. Mahwah, NJ: LEA (ミューレン, B. 小野寺孝義(訳) 2000 基礎から学ぶメタ分析 ナカニシヤ出版)
- 中本敬子 2000 上下の方向づけのメタファーに関する実験的検討—ストループ的課題を用いて— 心理学研究, **71**, 408-414.
- Nickerson, R. S. 2000 Null hypothesis significance testing : A review of an old and continuing controversy. *Psychological Methods*, **5**, 241-301.
- Nunnally, J. C., & Bernstein, I. H. 1994 *Psychometric theory*. 3rd ed. New York NY: McGraw-Hill.
- 岡安孝弘・高山 巖 2000 中学校におけるいじめ被害者および加害者の心理的ストレス 教育心理学研究, **48**, 410-421.
- Rosenthal, R., & Rosnow, R. L. 1985 *Contrast analysis : Focused comparisons in the analysis of variance*. Cambridge : Cambridge University Press.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. 2000 *Contrasts and effect sizes in behavioral research : A correlational approach*. Cambridge : Cambridge

University Press.

- 酒井恵子 2001 価値概念の個人差とその背景—価値尺度作成課題による検討 教育心理学研究, **49**, 102-111.
- 佐々木万丈 2001 中学生用体育学習ストレスコーピング尺度 (SCS-PE) の開発と標準化 教育心理学研究, **49**, 69-80.
- 杉浦義典 2001 ストレス事態に関する思考の抑制困難性と関連する対処方略—情報回避・情報収集・解決策産出と心配— 教育心理学研究, **49**, 186-197.
- 高岡昌子 2000 性格特性語の記憶に及ぼす印象形成の効果—検索過程の違いによる検討 心理学研究, **71**, 187-196.
- 竹内朋香・犬上 牧・石原金由・福田一彦 2000 大学生における睡眠習慣尺度の構成および睡眠パタンの分類 教育心理学研究, **48**, 294-304.
- 玉井紀子・中島定彦・北口勝也・今田 寛 2001 消去された恐怖反応の文脈変化による再出現—ラットの条件性摂水抑制事態での検討 心理学研究, **71**, 493-497.
- 田中佑子・中澤 潤・中澤小百合 2000 単身赴任の長期化が母親のストレスに与える影響—横断的研究・縦断的研究を通じて 心理学研究, **71**, 370-378.
- Thompson, B. 1995 Stepwise regression and stepwise discriminant analysis need not apply here : A guidelines editorial. *Educational and Psychological Measurement*, **55**, 525-534.
- 辻井岳雄 2001 顔の典型性が全体処理に及ぼす効果 心理学研究, **72**, 64-68.
- 宇井美代子・松井 豊・福富 譲 2001 女子高校生における性役割態度の変化過程 心理学研究, **72**, 95-103.
- Wainer, H., & Velleman, P. F. 2001 Statistical graphics : Mapping the pathways of science. *Annual Review of Psychology*, **52**, 305-335.
- Wilkinson, L., & the Task Force on Statistical Inference 1999 Statistical methods in psychology journals. *American Psychologist*, **54**, 549-604.
- 山崎晃男 2001 「教訓」の提示または産出による類推的問題解決の促進 教育心理学研究, **49**, 21-30.
- 柳井晴夫 2000 因子分析法の利用をめぐる問題点を中心にして 教育心理学年報, **39**, 96-108.
- 周 玉慧 2000 ソーシャル・サポート獲得方策リストの作成 心理学研究, **71**, 234-240.