

## 教育・心理調査で有用な統計ソフトウェアの活用事例 —初心者が見落としがちな統計知識を中心に—

小野 耕一<sup>1</sup>, 久保 順也<sup>2</sup>, 平 真木夫<sup>2</sup>

<sup>1</sup>宮城教育大学大学院 学校教育専修, <sup>2</sup>宮城教育大学 学校教育講座

本稿の目的は、心理調査に役立つ代表的な統計ソフトウェアの活用事例を紹介し、学生が陥りやすい誤りを指摘することである。アンケート調査の結果は、表計算ソフト EXCEL のデータシートに入力してから、統計ソフトへインポートする。その際、入力方法や変数名の付け方などに若干の注意が必要である。また、ピボットテーブルの機能は、クロス集計を実施する上で非常に便利である。SPSS は、社会調査や心理調査などで広く使用されている統計ソフトである。データのインポートや因子分析に関して、初心者が疑問に思う部分について解説を加えた。また、STATISTICA は、グラフ機能が充実している統計ソフトである。卒業研究に調査を取り入れる学生は、計画段階で統計の種類に関して鳥瞰図を持っていることが望ましい。有意義な研究を行うためには、どのような統計的分析があるかを把握した上で、調査の仮説を立てることが大切である。

キーワード: 統計、心理調査、EXCEL、SPSS、STATISTICA

### 1. はじめに

本稿の目的は、心理調査に役立つ統計ソフトウェアの活用事例を紹介し、統計的分析を行う学生が陥りやすい誤りの実例を示すことである。

近年、パソコンの性能向上と普及はめざましく、大学では多くの学生が自分専用のパソコンを所持するようになった。それに伴って、レポート作成時にワープロソフトや表計算ソフトが使われ、メールで提出することも珍しくなくなった。同様のことは統計ソフトについても言えるだろう。

以前の統計ソフトは、研究機関の汎用機で使われることがほとんどで、価格も著しく高価なものであった。そのため、研究者をはじめとする一部の限られた者だけが扱える敷居の高いものであった。ところが、近年のダウンサイジングによって、個人のパソコンで比較的容易に統計的分析が行えるようになった。統計ソフトは、従来の敷居の高いものから手ごろな分析ツールへと変わりつつある。したがって、統計ソフト自体の利用人口は、

一般のユーザーを中心に増加していると考えられる。

統計ソフトは大変便利な道具である。記述統計量(平均値、標準偏差など)を瞬時に出力することができるし、正しく操作を行えば、複雑な計算結果も数秒で返してくれる。しかし、便利になったことで、かえって統計のブラックボックス化が生じているとの指摘がある[1]。簡単に操作ができるということは、統計に関する知識が充分でなくても、マニュアルに従って適切な操作を行えば、計算結果が出力されるということである。統計ソフトは便利である反面、与えられた指示に対して従順に働くため、ユーザー側が誤った入力をして機械的に計算されてしまう。そのため、統計的に誤った(または、意味のない)結果を論文に記載してしまう恐れがある。

このように統計ソフトは便利であるが、ユーザー側が必要最低限の統計知識を備えていなければ、不適切な分析となることがある。特に卒業論

文で質問紙を使う学生の中には、分析の種類や内容があやふやのまま統計を使っているケースが少なくない。本稿では、統計ソフトの紹介を通して、統計的分析を実施する前提となる基礎知識についても論じていく。これから統計を使う学生には、統計に対する素朴な疑問点を解決した上で、有意義な卒業研究を行ってほしい。

## 2. 表計算ソフト EXCEL の活用

EXCEL(Microsoft 社)は、世界中で最もよく使われている表計算ソフトだと言われている。本学の情報処理センターでは、Macintosh と Windows 双方の OS で使えるようになっている。EXCEL は、便利な関数やショートカットキーを多数持っており、グラフ機能も優れている。心理調査では、アンケート結果を入力するシートとして活用する機会が多いと思われる。ここでは、アンケート集計に役立つ EXCEL の機能を中心に紹介していく。

### 2.1 調査結果の入力

質問紙による調査が終わったら、早速集計することになるのだが、その作業の前にいくつかやらなければならないことがある。まず、入力ミスを防ぐために全ての質問紙に通し番号を記入する。もし、集計後に誤りを見つけたとしても、誤って入力したデータがどの質問紙かを特定できれば、再入力も容易である。全ての分析が終了してから誤りが見つかり、再度分析を行わなければならないという事態は避けたい。

次に入力用のフォーマットを作成することになるが、1 行目の変数名の長さは半角 8 文字か全角 4 文字以内のアルファベット表記とする。統計ソフトでは変数名の長さ制限があることが多く、バージョンによっては日本語の変数名がつけられないことがある。したがって、EXCEL に入力する段階で、アルファベットを使って端的に表現することが望ましいだろう。

その後、入力する際には、セパレーターでウィンドウを分割して、表の異なる範囲を同時に表示

すると見やすい。右スクロールバー上端の分割バーにマウスポインタを合わせ、両向き矢印になったら下方向にドラッグする。また、よくある失敗として、数字の半角と全角を間違えて入力してしまうことがある。半角数字は数字として、全角数字は文字として認識されてしまうため特に注意が必要である。上記のような EXCEL の入力例は、学校評価のアンケートや児童生徒の実態調査などで集計作業をする際に役立つはずである。

### 2.2 ピボットテーブル

EXCEL の優れた機能の一つにピボットテーブルがある。ピボットテーブルは、ドラッグ&ドロップ操作によって様々な要因を組み合わせることができるため、クロス集計に大変便利である。以下に簡単な用法を提示する。

表の中のセルを選択し、[データ]→[ピボットテーブルとピボットグラフレポート]をクリックする。[ピボットテーブル/ピボットグラフウィザード]が表示されたら、そのまま[次へ]をクリックする。使用するデータの範囲が表示されるので、そのまま良ければ[次へ]をクリックし、[新規のワークシート]をオンにしてから[完了]を押すと、空白のピボットテーブルが作成される(図 1)。

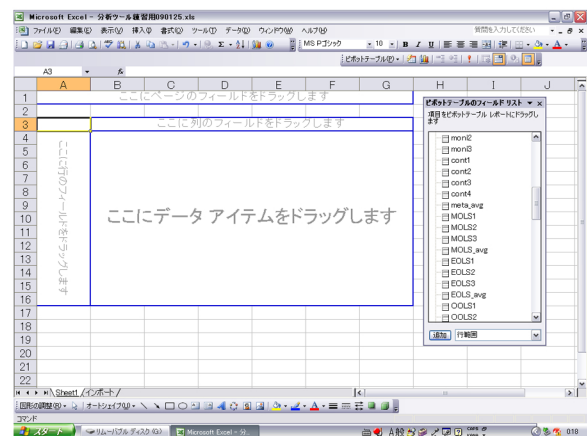


図 1 ピボットテーブル

また、表示される数値を変更する場合には、変更したいセルを選択し、[ピボットテーブル]→[フィールドの設定]から平均値や個数を選択す

ると良いだろう。このようにピボットテーブルを活用すれば、比較的容易にクロス集計ができる。たとえば、学校でアンケート調査を実施した場合には、ピボットテーブルを活用して資料を作成すると効率的だと言えよう。

### 3. 簡単だけれど重要な記述統計の話

統計は大きく記述統計と推測統計に分けることができる。記述統計は得られたデータを要約してわかりやすく表現するためのものであり、推測統計はデータに関係や差が見られるかどうかを調べるものである。ここでは、統計的分析の前提となる記述統計について重要なトピックを示す。統計的分析は、データの形式によって適用できる分析が異なっているのである。

#### 3.1 平均値

平均値は日常生活で最も馴染みの深い統計値である。試験結果やアンケート集計と言え、真つ先に平均値を思いうかべるのではないだろうか。しかし、得られたデータの種類によっては、平均値を求めてもあまり意味のない(適さない)ものが存在する。その判断を誤れば、意味のない分析結果を求めることになってしまう。

##### 例1: カテゴリカルデータの平均値

データの入力段階で男性を1、女性を2とした後、これらの平均を求めても意味はない。

##### 例2: データが正規分布と見なせない場合

極めて多量のデータを集めれば、データは正規分布の形に近づくとされている。しかし、データによっては分布の形が正規分布と見なせない場合がある。たとえば、「勉強するときはノートをまとめ直す」という4件法の質問項目に対して、「はい: 4点」(11人)、「どちらかというとはい: 3点」(3人)、「どちらかと言うといいえ: 2点」(2人)、「いいえ: 1点」(15人)という結果が得られたとする。このピークが2つあるデータについて平均値を求めたとしても、その平均値が何を指しているかは不明確である。つまり、求めた平均値が代表値だとは言えないのである。した

がって、データに複数のピークが見られる場合には、平均値と標準偏差を使うt検定や分散分析は避け、正規分布を仮定しないノンパラメトリック検定である $\chi^2$ 検定や直接確率計算法を選択すべきである。

#### 3.2 尺度

質問紙で用いられる尺度には、名義尺度(属性データ)、順序尺度(順番はあるが等間隔ではない)、間隔尺度(数値の差の間隔が等しいと見なせる)、比尺度(温度計のように原点がある)がある。心理調査では間隔尺度を使うことが多く、社会調査では名義尺度や順序尺度が多いように見受けられる。使用する尺度の種類によって、分析方法はある程度限定されてしまうので、質問紙を作成する段階でどのような分析を行うかについて見通しを立てておくことが必要である。

## 4. SPSSの利用

SPSSは、1968年に米国スタンフォード大学で開発された社会調査用の統計ソフトである。SPSSは、マーケティングや心理調査、医療関係の調査で活躍しており、パソコンで最も多く使用されていると言われる。

#### 4.1 インポート

初心者がはじめにつまずくのがデータをインポートする作業である。複数の方法があるが、ここではEXCELファイルをそのままインポートする方法を解説する。ただし、使用するのはSPSS ver11.0である。

SPSSを起動させ、[データに入力]を選ぶ。[ファイル]→[テキストデータの読み込み]を選択した後、ファイルの種類をEXCELに指定し、保存場所にあるEXCELファイルを選択する。最後に[データの最初の行から読み込む]にチェックが入っていることを確認し[OK]をクリックする。インストールされているSPSSのバージョンによっては、EXCELの形式(たとえば、EXCEL2007)が対応していない場合がある

ので、適切なバージョンに落として保存する必要がある。

#### 4.2 因子分析に関する疑問

文献に記載された尺度や自作した質問項目を使用して調査を実施する場合に、因子分析を行うことがある。因子分析とは、類似した項目をある基準によって分類する統計手法である。たとえば、一般的に言われるような「理系能力」や「文系能力」などは、実際に測定されない(測ることのできない)データである。しかし、私たちは数学や国語の試験成績などから、自分の適性を文系や理系だと判断することがあるだろう。こうした試験の点数は、実際に測定される(測ることのできる)データである[2]。前者を潜在変数、後者を観測変数と呼ぶ。

この例によれば、数学や理科の点数は、「理系能力」から強く影響を受けており、逆に国語や社会の点数は、「文系能力」から強く影響を受けていると考えられる。因子分析は、観測変数である「各教科の点数」が、潜在変数である「文理能力」から”どのくらい影響を受けているか”を因子負荷量という数値で表現する。そして、その得られた数値の絶対値が1に近いほど強く影響を受けているということになる。

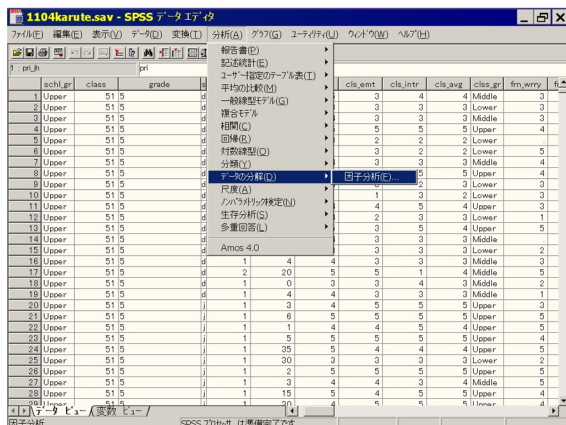


図2 SPSS の分析画面

本稿では、因子分析の具体的な操作方法は解説書に譲り、因子分析を実施する上での注意点に限定して話を進めたい。

#### ① 因子分析はどこで終わりにするべきか

因子分析は、分析者自身も想定していなかった共通因子が、自動的にあぶり出される魔法のような分析手法ではない。すなわち、非常に恣意的な分析方法だと言えるだろう。因子分析の結果は、観測変数の選び方によって異なるため、分析者がある程度明確な仮説を持っている必要がある。そして納得のいく結果が得られた時点で分析を終わりにする。因子分析は、一つの判断基準として活用するものである。

#### ② 因子得点と素点どちらを使うべきか

因子数が確定したら、調査の仮説に基づいて尺度間の相関を求めたり平均の差を求めたりする。そのときに、因子得点と素点のどちらを使うべきかと言った疑問がある。両者は高い相関があることから、どちらを使っても良いと言われている。因子得点を用いた方が正確ではあるが、統計に詳しくない一般人に向けて使うには不親切である。”調査結果を誰に見せるのか”という観点から、選択するべきである。

### 5. STATISTICA の利用

STATISTICA は、元々パソコン市場をターゲットに開発されたソフトウェアであり、グラフ機能が SPSS に比べて扱いやすく編集が容易である。したがって、グラフによる直感的な理解が有効となる分析方法の場合に、分析結果を効果的に表示することができる。

#### 5.1 インポート

STATISTICA も、EXCEL からインポートすることが可能である。しかし、所属研究室の端末にインストールされている STATISTICA のバージョンは 2000 年版であるため、現在の EXCEL(97-2007) のファイル形式には対応していない。STATISTICA 2000 が対応している EXCEL のファイル形式は ver4.0 までである。

このように、研究室で用意されている作業環境が常に最新のものであることを期待することは一般的に難しいのではないだろうか。したがって、様々

な状況に対応できるためにも、ファイルのインポート・エクスポートに関して汎用性の高い方法を知っておく必要があるだろう。たとえば一太郎とワード、Star Suite の Writer など、ワープロソフトによってファイルの保存形式は異なるように、統計ソフト、表計算ソフトでもファイルの保存形式がアプリケーションごとに異なっている。このとき、ワープロソフトで汎用性の高いミニマムなファイル形式はアスキーファイルであるが、表計算ソフトでは CSV 形式がこれに相当している。CSV とは Comma Separated Value の頭文字をとったものである。その名の通り数値をカンマで区切ったファイルである。

アプリケーションごとにファイルの保存形式が異なるし、同一のアプリケーションでもバージョンアップによる上位互換が保証されないことがありうることを考えると、データのバックアップファイルはより汎用性の高い CSV フォーマットで保存しておくことが望ましいと言えるであろう。実際に、CSV フォーマットのファイルは(保存されている文字コードにもよるが)メモ帳などで編集することが可能である。

なお、CSV ではセパレーター(数値を区切るルール、記号)がカンマであるが、カンマの代わりにタブで数値を区切ることもよく見られる。EXCEL では、[名前をつけて保存]の画面で[ファイルの種類]のオプションの中に CSV(カンマ区切り)を選ぶ。

STATISTICA からこのファイルを利用するときには、まず始めに STATISTICA を起動させる。次に、[分析方法の選択]の画面が表示されるので、これから実施する分析の種類を選択し、[切替]をクリックする。データ画面が表示されたら [インポート]→[クイック]を選択し、保存した CSV ファイルを呼び出す。

## 5.2 グラフの保存

STATISTICA の魅力はグラフ機能の充実である。特に分散分析(3群以上の平均値を比較する)は、交互作用を確認する上で、グラフによる結果

の表示は非常に有効である。分散分析を実行してグラフが得られたら、[メタファイル形式で保存]を選択し保存をする。この形式で保存をしておけば、Word に貼り付けることができるので便利である。

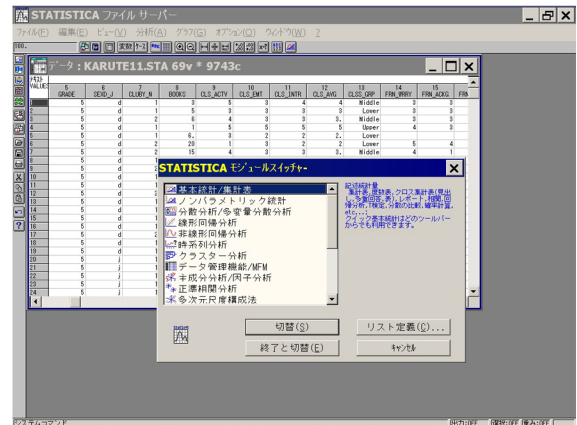


図3 STATISTICA の分析画面

## 6. 卒業論文における統計

本学の情報処理センターは、新システムへの移行により、EXCEL for 統計がインストールされた Windows 端末が第3演習室の30台用意される。それに伴って、課題や卒業研究で統計を使う学生が増加すると考えられる。しかし、パソコンの性能やソフトの機能が便利になったことで、かえって統計のブラックボックス化が生じてしまう可能性がある。統計ソフトは、実験計画を指定してしまえば、瞬時に結果を算出してくれる。したがって、ユーザー自身が正しい統計の知識を身に付けた上で分析を行わないと、何を示したい分析なのかが不明確となってしまうだろう。

筆者は大学院生という立場から、学部生に統計のアドバイスを行ってきた。その中で見られた統計の誤用に関して、具体的な例をいくつか示すことで、これから研究に取り組む学生の参考になると考えている。

### 6.1 仮説の重要性

質問紙を用いた心理調査では、日常の観察から得られた人間の行動を量的データとして収集する。そこには証明したい事柄である仮説が存在し

ているはずである。しかし、学部学生の中には仮説が曖昧な状態で調査を実施してしまう者がいる。一般に研究を目的とした調査は、調査を実施する時点で分析方法が定まっているものである。したがって、調査データが集まってから分析方法を決めるのでは、苦勞して収集したデータを十分に活かせるようになる可能性がある。たとえば、どんな調理器具や調味料があるかわからないのに、作る料理だけが決まっているのはおかしいだろう。このように統計を使った調査を計画する段階で、分析手法のイメージができていないとまずいと言えるのではないだろうか。

## 6.2 因果関係

因果関係が逆転している実験計画を見ることがある。統計的分析では、従属変数(目的変数)と独立変数(説明変数)の区別があり、従属変数とは独立変数から影響を受ける変数である。しかし、気をつけないとその変数が逆になっていることがある。たとえば、“夜更かしをする子どもは不健康である”という関係があったとしよう。不健康の原因は夜更かしだと推察できるが、夜更かしの原因が不健康だとは言えないだろう。

## 6.3 統計的分析のルール

分散分析(3群以上の平均値を比較する)について相談を受けた。グラフを見たところ、交互作用が見られていてもおかしくない様相であった。ところが、注意深く出力結果を読んだところ、2人の群と110人の群を比較していることがわかった。実験計画自体は興味深いものであったが、分散分析の適用できるルールからはみ出している。100人の群の平均値は、ある程度その集団の特徴を指し示していると言えるだろう。しかし、2人の群について平均値を求めたところで、その平均値が群の代表値かどうかは何とも言えない。残念ながら誤差の影響が著しいために統計的に有意な値が得られなかった。ここで特筆すべきは、統計ソフトは機械的に計算結果を求めてしまうということである。つまり、ユーザー側が正しい知

識を持って操作を行わなければ、誤った分析結果を算出することになりかねない。くれぐれも注意する必要があるだろう。

また、分散分析によって明らかになるのは、“どこかに差がある”ということであり、実際にどの群間に差が生じているのかは、下位検定である多重比較を実施しなければ分からない。代表的な下位検定には、TukeyのHSD法による多重比較がある。

## 7. おわりに

本稿では、統計ソフトウェアの活用事例を示すと共に、統計を活用するために必要な知識や条件について多くの紙数を割いた。トピックによっては、当たり前だと言われるような内容も含まれている。しかしながら、コンピュータやソフトウェアが便利になったことで、統計知識は見過ごされてしまう傾向にある。本稿は、そうした“初心者が見落としがちな統計知識”を再度確認してもらうことが目的であった。これから統計を使う人には、正しい統計知識とセンスを持って、卒業研究や調査に取りかかってもらいたいと願っている。

## 8. 文献

- [1] 新村秀一：日本における統計ソフトの過去・現在・未来、計算機統計学、第10巻、第1号、pp.37-60(1998).
- [2] 小塩真司：SPSSとAmosによる心理・調査データ解析—因子分析・共分散構造分析まで—、pp.106-107、東京図書(2004).