

フリーの統計分析ソフト HAD: 機能の紹介と統計学習・教育, 研究実践における利用方法の提案 An introduction to the statistical free software HAD: Suggestions to improve teaching, learning and practice data analysis

清水裕士^{*1}

Hiroshi Shimizu^{*1}

^{*1} 関西学院大学

^{*1}Kwansei Gakuin University

要約

HADはMicrosoft Excelで動く心理統計分析用のソフトウェアであり, 基礎的な分析から, 多変量解析まで数多くの分析をGUIで行うことができる。HADはWebからソフトウェア, 操作方法についての資料が無償で利用可能であるので, 学生や大学教員などが研究や教育現場で活用することが期待できる。本論文では, HADの分析機能の紹介に加え, HADを統計学習ツール, 統計教育ツールとしての活用方法と機能について紹介した。最後に, HADによる研究活動における探索的分析ツールとしての利用方法の提案と, 利用上の注意点について論じた。

Abstract

HAD is a GUI-based free software program that contains a number of functions for basic statistical and multivariate analysis for use in statistical analysis for psychology. HAD and its user guides are available online, and therefore, teachers and students can utilize the program for teaching and learning. This article introduces the utility of HAD for statistical analysis, as well as for teaching and learning statistics in universities. Moreover, methods of using HAD for research practices are suggested, and important issues in using HAD are discussed.

キーワード

心理統計学, フリーソフトウェア, HAD, 統計学習・統計教育,

Keywords

Psychometrics, Free software, HAD, statistical learning and teaching

1. 統計分析ソフト HAD について

HADとは, 著者が開発した, 心理統計分析用のフリーソフトウェアである。本論文では, 統計分析ソフトとしてのHADの分析機能の紹介, そして統計教育, 統計学習, 研究実践における利用方法について解説をすることを目的とする。

1.1. HADとは

HADは, Microsoft Excel (以下, Excel) で動くソフトウェアであり, Microsoft OfficeがインストールされたWindowsおよびMac OSの環境で実行することができる。本論文執筆現在(2016年2月), WindowsではMicrosoft Office2007以降, MacではMicrosoft Office for Mac 2011以降で動作を確認している。ただし, MacについてはOSのバージョンなどによってはエラーが生じること, また機能が制限されていることから, 基本的にはWindowsによる利用を推奨している。

HADは無償で自由なソフトウェアである。HADのファイル自体は利用が無償であり、誰でも何度でもダウンロードして利用可能である。また、HADのソースコードは閲覧が可能であり¹、自由に修正し再配布を行うことができる。ただし、HADはGNU General Public License (GPL)²のver2以降のライセンスに基づいているので、再配布を行う場合には注意が必要である。加えて、HADの免責事項³にあるように、開発者はその分析結果についていかなる責任も負わないため、分析結果の報告は自己責任で行う必要がある。

HADの開発は著者一人で行っている。開発環境はWindows版のExcelで、2010、2013、2016である。開発言語はExcelに内蔵されているVisual Basic for Applications（以降VBA）を用いている。よって、HADはいくつか基本的な処理（一部の確率密度関数など）はExcelに含まれている関数に依存しているものがあるが、平均値や分散といった統計量に関するほとんどの処理は著者がVBAによって記述した関数を用いている。ただし、構造方程式モデルについてはExcelに内蔵されているソルバーを用いて推定を行っている。

本論文執筆現在、HADはver15.00が最新版であり、データハンドリング、基礎的な統計分析からグラフ作成、そして多変量解析に至るまで広範囲な統計分析が可能である。また、統計教育用の機能もいくつか搭載されており、分析ツールとして以外にも、統計教育ツールとしての活用も可能である。これらについては、それぞれの章で詳述する。

HADの開発は2005年からはじまり、ver3.0が清水・村山・大坊(2006)で発表された。当初は、マルチレベル分析のための補助分析ツールとして利用するためのものであった。現在のよ様な統計分析全般について行えるようになったのはver9.0以降からである。このように、HADは開発当初から統計教育、学習のためのツールとして開発されていたわけではない。機能の追加に伴ってその方向性も大きく変わってきた経緯がある。

ただし本論文ではこれらの開発経緯とは独立に、現状のHADにおける実際の使われ方、そして推奨される使われ方などについて論じる。なお、HADの詳しい使い方についてはHADのWebサイト⁴にマニュアルや解説スライドをアップしているため、本論文では詳しくは触れない。

1.2. HADのダウンロードと起動

HADはWebサイトから自由にダウンロードすることができる。ダウンロードサイトには基本的には最新版のみがアップされているが、場合によっては動作が比較的安定しているバージョンもあわせてアップされていることがある。HADはユーザーからの不具合報告を受けた場合に

¹ Visual Basic Editorにはパスワードが設定されているが、「simizu706」で解除できる。

² 「GNU General Public License」<http://www.gnu.org/licenses/gpl.html>

³ 「HADとは」<http://norimune.net/696>

⁴ 「統計分析ソフトHAD」<http://norimune.net/had>

頻繁に更新されるため、なるべく最新版を利用することが推奨される。また、HAD には「ソルバーオンバージョン」と「ソルバーオフバージョン」の 2 種類がアップされている。ソルバーオンバージョンは、Excel のソルバーアドオンを参照する設定のもので、オフバージョンは参照されていない。ソルバーオンバージョンでないと構造方程式モデルが実行できないが、それ以外の機能はオンバージョンとオフバージョンには違いがない。ただし、Mac で利用する場合はソルバーが上手く動かないため常にオフバージョンを利用するとよい。

HAD をダウンロードして起動する場合、マクロのセキュリティ設定が最も高い場合はマクロが起動されず、利用することができない。マクロが実行できるセキュリティレベルに変更する必要がある。なお、ソルバーオンバージョンを初めて起動する場合、Excel のソルバーアドオンの参照が上手くいかず、非表示モジュールでコンパイルエラーが生じる場合がある。そのような場合は一度 Excel を閉じてもう一度 HAD を起動すれば、コンパイルエラーは生じなくなる。

1.3. HAD の構成

HAD は、基本的には、データを入力するデータシートと、データ分析を行うための場所であるモデリングシートの 2 つによって構成されている（図 1、図 2）。分析をする上では、この 2 つのシートを主に用いることになる。

	A	B	C	D	E	F	G	H	I
変数名	ID	time1	time2	time3	time4	条件	評価	年齢	
データ読み込み	OBS01	5.02	7.01	8.01	9.09	1 A		21	
	OBS02	6.07	4.04	5.06	6.01	0 A		22	
	OBS03	2	4.07	7.05	9.01	1 A		18	
モデリングシート	OBS04	7.05	5.08	4.07	7.03	0 A		19	
	OBS05	2.01	4.1	6.08	8.05	1 B		19	
	OBS06	2.08	3.07	3.03	3.03	0 B		18	
	OBS07	3.03	5.03	4.05	7.06	1 B			
	OBS08	4.06	6.08	6.02	4.06	0 B		20	
列幅の調整	OBS09	3.06	4.07	6.08	5.04	1 C		18	
	OBS10	3.08	2.09	3.1	2.02	0 C		21	
	OBS11	3.02	3.05	6.1	7.01	1 C		20	
	OBS12	5.07	6.07	2.05	3.05	0 C		20	
数値計算									

図 1 データシート

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	変数名	ID	time1	time2	time3	time4	条件	評価	年齢				
2													
3	分析		選択セルを使用	統制変数を投入	変数を左につめる					● 実数情報		変数の作成	
4										○ 回帰分析			
5										○ 因子分析			
6	データシート											データセット	
7													
8													
9	使用変数	ID											
10													
300	変数情報	フィルタ	値ラベル	ラベル	@変数	コード							
301	ID												
302	time1			1回目									
303	time2			2回目									
304	time3			3回目									
305	time4			4回目									
306	条件		1=実験, 0=統制										
307	評価												
308	年齢			年齢									

図 2 モデリングシート

データシートには図 1 のように、行にサブジェクト、列に変数という形式でデータをセルに入力する。また、B 列にはサブジェクトを区別する「ID 変数」を必ず入力する必要がある。データを入力したら、左上の「データ読み込み」ボタンを押すことで変数の情報などが HAD 内部で処理され、モデリングシートに移動する。

モデリングシートは図 2 のように、分析に使用する変数を指定したり、変数の情報を設定したり、データハンドリングを実行できるボタンが配置されている。モデリングシートでは、最初の状態では隠れているが、「回帰分析」や「因子分析」と書かれたラジオボタンを押すことで多変量解析のモデルを指定するための「モデリングスペース」が表示される。

HAD では、分析結果はすべて別のシートに出力される。シート名は分析の種類ごとにつけられ、同じ分析手法はデフォルトでは上書きされる。ただし、上書きしない設定にすることもできる。また、モデリングシートにある「シート管理」ボタンを押すことで、出力された結果シートの編集、保存、削除などの処理を行うことも可能である。さらに、HAD では各出力に分析コードと呼ばれる、分析法やオプションなどがすべて保存されたコードが表記される。これを用いることで、過去に行った分析と全く同じ分析を実行したり、どのような分析設定で分析したのかを確認できたりすることができるようになっている。

2. HAD の分析機能

HAD の分析機能は、「データハンドリング」、「基礎統計分析」、「多変量解析」の大きく 3 つに分かれている。さらにその中で、データハンドリングは「変数情報の設定」、「変数の作成」、「データセット」の 3 つに分かれるなど、サブカテゴリが存在する。ここでは、3 つの分析機能をそれぞれ簡単に解説しよう。

2.1. データハンドリング

HAD のデータハンドリングは、「変数情報の設定」、「変数の作成」、「データセット」の 3 つに分かれている。

「変数情報の設定」は、モデリングシートの下部にある変数情報のスペース（図 2 の 300 行目以降）で、変数のフィルタ、値ラベル、変数ラベルの設定を行える。フィルタとは変数の値にあわせて分析に使うサブジェクトを選択したり、変数の値ごとに分析結果をグループ分けして出力したりするなどの機能のことである。値ラベルの設定では、変数の値にそれぞれラベルをつける（たとえば、性別を区別する変数に「男性」、「女性」というラベルをつける）もので、変数ラベルは変数名以外のラベルをつける（たとえば sex という名の変数に「性別」というラベルをつける）ことができる。値ラベルを設定すると、平均値の差の検定のような群分けする分析の結果に、値ラベルが自動的に反映される。

「変数の作成」は、モデリングシート上部にある「変数を作成」ボタンから利用する。この

機能は、変数を合成・平均して新しい変数を作成したり、尺度変換、数値演算などを行って変数を変換したり、ダミー変数を作成するものである。

「データセット」は、モデリングシート上部の「データセット」ボタンから利用する。ここには、一部の変数だけのデータセットの出力、リストワイズ削除したデータセットの作成、複数の変数を指定して縦並び（long 型）のデータセットの作成、といった機能が含まれている。他にも、乱数データを生成するなど可能である。

2.2. 基礎統計分析

HAD では、統計分析をするための方法には 2 種類ある。一つは基礎統計分析、もう一つは多変量解析である。基礎統計分析はモデリングシート上部にある「分析」ボタンを押し、チェックボックスをチェックして OK ボタンを押すだけで実行することができる。多変量解析については後の節で解説する。

基礎統計分析には、大きく分けて 4 つの分類がある。図 3 のように、「データの要約」、「差の検定」、「変数間の関連性」、「マルチレベル分析」にカテゴリが分けられている。



図 3 HAD の基礎統計分析機能

「データの要約」には、データの代表値の計算や、データからグラフを作成するための機能が含まれている。たとえば平均値や標準偏差の計算、ヒストグラムや箱ひげ図を作成する、散布図やバブルチャートを作成する、などである。また、グラフは、設定を変えることで散布図に予測直線を引く、ヒストグラムに正規近似曲線を加える、箱ひげ図にドットプロットやバイオリンプロットを追加する、といったように出力を変化させることができる。

「差の検定」は、平均値の差の検定を代表とする、各種検定を実行する。順位の差の検定は、いわゆるノンパラメトリック検定を各種含んでおり、対応のあり・なし、そして変数のタイプによって自動的に分析手法を判断して出力する。

「変数間の関連性」は、変数間の共分散、相関係数、順位相関係数、ポリコリック相関係数（たとえば豊田, 1998）など、各種の相関係数の計算、尺度の α 係数を代表とする信頼性係数の計算、多重クロス表の出力と対数線形モデルを含んでいる。

最後に「マルチレベル分析」は、マルチレベル分析を実行する前に確認することが多い、級内相関係数、マルチレベル相関係数、グループごとの回帰係数と散布図を出力する。ただし、この「マルチレベル分析」に含まれている分析法を利用するときは、B 列に集団を区別する ID 変数を入力しておく必要がある。

2.3. 多変量解析：回帰分析系

HAD では、チェックボックスで簡単に実行できる基礎統計分析とは別に、複雑なモデルを指定する多変量解析の機能もある。多変量解析は、大きく分けてデータの予測をする回帰分析系と、データの縮約や分類を行う因子分析系に分かれている。この 2 つの多変量解析の機能は、上述したように、モデリングスペースと呼ばれるモデルの詳細を選択するシートを開いて利用する。この節ではまず回帰分析系について解説する。回帰分析系は、さらに 4 つのサブカテゴリに分かれており、「回帰分析」、「分散分析」、「一般化線形モデル」、「階層線形モデル」が含まれている。表 1 は、それぞれのサブカテゴリにどのような分析が含まれているかを示している。

「回帰分析」カテゴリには、重回帰分析以外に、判別分析や多変量回帰分析などが含まれている。また、外れ値への重みを小さくするロバスト回帰分析(Huber, 1973)や、正規分布の仮定からの逸脱を補正する頑健標準誤差、ブートストラップ法なども実行可能である。

表 1 回帰分析系の分析手法一覧

カテゴリ	実行できる分析手法	
回帰分析	重回帰分析 ロバスト回帰分析 判別分析 多変量回帰分析（正準相関分析）	
分散分析	分散分析（一般線形モデル） 対数線形モデル	
一般化線形モデル	連続 順序 カウント 割合 打ち切り 名義 混合分布	正規分布・ガンマ分布・対数正規分布 順序多項分布 ポアソン分布・負の二項分布 二項分布・ベータ二項分布 打ち切り正規分布 多項分布 混合正規分布
階層線形モデル	階層線形モデル	

「分散分析」カテゴリでは、共変量を含む 5 要因までの参加者内・間要因計画の分散分析を実行する。さらに要因ごとに単純効果分析や多重比較などを自動的に行う。多重比較はボンフェローニ法、ホルム法、シェーファー法などが選択することができる。HAD における分散分析の特徴は、単純主効果の検定において誤差項を要因でプールしたものをを用いるか、水準で別の誤差項を用いるかを選択することが可能な点である。その他、効果量や平均値のパターンを示すグラフも出力される。

「一般化線形モデル」カテゴリには、従属変数がさまざまなタイプの分布に従う場合にでも対応できるように、多くの線形モデルが包括されている。HAD では、分布名によって分析を選択するのではなく、従属変数の性質によって分類されている。例えば、正規分布、ガンマ分布、対数正規分布などは連続変数としてカテゴリ化されている。同様に、順序尺度の場合、名義尺度の場合、カウントデータの場合、割合（比率）データの場合というように、データの性質に合わせて選択できるようになっている。

「階層線形モデル」カテゴリでは、変数効果を含む回帰分析である階層線形モデル（Bryk & Raudenbush, 1992, 清水, 2014）を実行する。HAD では、階層線形モデルを実行する場合に用いられるいくつかの事前処理（集団平均中心化、全体平均中心化など）を自動的に行ってくれる。ただし、HAD の階層線形モデルでは、他のソフトウェアの一般線形混合モデルのように残差共分散構造を選択したり、制限付き最尤法による推定を実行したりすることはできない。これらの分析については次節で述べるように、他のソフトウェアと連携することでカバーしている。

これらに加えて、HAD では、重回帰分析、一般化線形モデル、階層線形モデルについては調整分析と媒介分析、ステップワイズ法による変数選択法を実行することができる。調整分析では自動的に説明変数を中心化して交互作用項を作成し、調整変数の特定の点での単純効果分析、グラフの作成も自動で行う。媒介分析では、間接効果の検定としてソベル検定やブートストラップ信頼区間の利用が可能であり、また調整媒介分析（媒介効果が別の変数で調整されることを検討する分析）も実行可能である。ただし、一般化線形モデルと階層線形モデルではノンパラメトリックブートストラップ法は利用することができない。変数選択法は重回帰分析と一般化線形モデルではステップワイズ法以外に、階層的投入法も選択できる。

2.4. 多変量解析：因子分析系

因子分析系には、「因子分析」、「クラスタ分析」、「数量化分析」、「構造方程式モデル」の 4 つのサブカテゴリが含まれている（表 2）。

「因子分析」カテゴリでは、因子分析を実行する。共通性の推定方法として、最尤法、最小二乗法、反復主因子法、主成分法、ポリコリック相関行列を用いた重みつき最小二乗法（HAD 上では「カテゴリカル」と表記）が用意されている。また因子軸の回転方法として直交回転と

表2 因子分析系の分析手法一覧

カテゴリ	実行できる分析手法
因子分析	因子分析 抽出法：最尤法・最小二乗法・反復主因子法・主成分法・重みつき最小二乗法 回転法：バリマックス, プロマックス, オブリミン, 独立クラスタ, プロクラステス カテゴリカル因子分析 項目反応理論
クラスタ分析	階層クラスタ分析 ウォード法・群平均法・最長距離法 非階層クラスタ分析 k-means法・改良k-means法 混合正規分布モデル 潜在クラス分析 潜在ランク分析
数量化分析	等質性分析（多重対応分析あるいは数量化Ⅲ類）
構造方程式モデル	確認的因子分析 平均・共分散構造分析 探索的因子分析 マルチレベル構造方程式モデル

してオーソマックス回転（バリマックス回転を含む一般化された直交回転族）、斜交回転としてプロマックス回転、直接オブリミン回転、独立クラスタ回転がある。また、仮説検証型の回転方法としてプロクラステス回転も用意されている。

また推定方法で「カテゴリカル」を選択し、一因子解を実行すると項目反応理論に関するパラメータ推定を行う。このとき、識別力や閾値母数、テスト情報関数、項目特性曲線などが出力される。

「クラスタ分析」カテゴリでは、参加者は変数をグループ分けするための分析手法が含まれている。クラスタ分析としては、階層クラスタと非階層クラスタの2種類がある。階層クラスタ分析ではウォード法、群平均法、最長距離法を選択でき、またデンドログラムの出力を行う。非階層クラスタ分析では、k-means法と改良k-means法（豊田・池原, 2011）の2つがある。なお、階層クラスタ分析を選択した場合は、サブジェクトではなく変数のクラスタリングを行うことも可能である。

さらに、混合正規分布モデル（Everitt & Hand, 1981; 豊田, 2000）や潜在クラス分析（Lazarsfeld, P. F. & Henry, N. W., 1968; 豊田, 2000）といった、ファジークラスタリング⁵の手法も利用できる。クラスタ分析とファジークラスタリングモデルでは、HADの出力の多くは共通しており、サブジェクトを分類するという目的に合った出力を行っている点がHADの特徴である。この特徴については3章でも触れる。

⁵ サブジェクトを1つのクラスタに固定して所属させるのではなく、それぞれのクラスタに確率的に所属すると想定するモデルのことである。

「数量化分析」は、質的変数を量的次元に縮約する分析手法である等質性分析（あるいは数量化Ⅲ類）を実行する。また、数量化分析ではクロス表を直接入力するように指定することで、対応分析も実行可能である。

「構造方程式モデル」では、通常の構造方程式モデル（たとえば豊田, 1998）以外に、確認的因子分析、探索的因子分析、マルチレベル構造方程式モデル（Muthén, 1994; 清水, 2014）を実行する。推定方法としては最尤法と一般化最小二乗法が用意されている。また、構造方程式モデルでは完全情報最尤法を用いた欠損データの分析も可能である。構造方程式モデルで探索的因子分析を実行する意味は、この欠損データの分析があることによる。

2.5. 他のソフトウェアとの連携

上記に紹介した機能はすべて VBA によって書かれたコードに基づいて計算を行うが、より高度な分析になると、プログラミングの困難さと Excel の計算速度の問題から、HAD では実装が難しいものもある。それらの分析手法については、他の統計ソフトウェアで簡単に実行できるようにコードを自動的に生成する機能によって対応している。

一般化線形混合モデル（たとえば久保, 2012）については、SAS の GLIMMIX プロシージャ用の SAS コードを自動的に作成する HAD2glimmix がある。またそれ以外に、著者が開発した R 用の一般化線形混合モデルをベイズ推定するためのパッケージ、glmmstan⁶のコードを自動的に作成する HAD2glmmstan という代筆機能がある。

また、ver15 以降では、基礎統計分析について R で同様の分析を実行するためのコードを出力する、「HAD2R」がある。これについては次章で詳述する。

さらに HAD には、テキストマイニング用の分析機能も備わっている。これはテキストマイニング用のフリーソフトである TTM (Tiny Text Miner: 松村・三浦, 2014) との連携機能「TTM2HAD」である⁷。TTM2HAD には、TTM で単語分解されたデータを読み込み、テキストデータに対する要約や、主成分分析やクラスタ分析、数量化分析などの多変量解析を実行する機能が搭載されている。そして、テキストマイニングによって数量化されたデータは、そのまま HAD で他の変数との関連などを検討することもできる。

3. 統計学習・教育ツールとしての HAD

上記のように、HAD にはさまざまな分析機能が用意されている。また一方で、HAD はただデータ分析をする以外にも、統計の学習ツール、教育ツールとしての利用価値があると著者は考

⁶ glmmstan は github 上にアップされている。 <https://github.com/norimune/glmmstan>

⁷ なお、HAD と連携しているのは Excel で動くバージョンである「ExcelTTM」のほうである。以下の URL からダウンロードすることができる。 <http://mtmr.jp/excelttm/>

えている。本章では、HADの統計学習・教育利用方法について紹介する。

3.1. 統計学習ツールとしての利用

まず、HADの統計学習ツールとしての長所をいくつか挙げておこう。それは、1. HADがExcelで動くこと、2. Webに使い方の資料があり、自由に閲覧できること、3. 分析法がアルゴリズムではなく分析目的で分類されているため、高度な分析もシームレスに利用・学習できること、の3つがあると考えられる。なお、この節で論じられている統計学習とは、大学生や大学院生が、大学の授業以外の場所や時間で、独自に学習をすることを指している。授業内における学習については、次の統計教育についての節で述べる。

Excelで動くHAD 統計学の初学者の多くは大学生や院生であることが多いが、大学に行かないと商用の統計分析が利用できない状況では、十分な統計学習の環境にあるとは言えないだろう。特に長期休みには、大学自体が利用できない場合もある。学生が主体的に学習するうえで、自身が所有するPCで自由に統計分析をできる環境があるほうが望ましいといえる。

それに対してHADはExcelで動くソフトウェアであるので、PCにExcelがインストールされてさえいれば、誰でもどこでも統計分析を行うことができる。学生の多くはWindowsかMac OSおよびMicrosoft Officeが入ったPCを所有していることから、自宅でも統計分析が行える環境を提供することができる。

また、多くの大学では初年度でExcelの使い方の基礎を教えているため、すでに馴染みにあるソフトウェアによって統計分析が行えることも学習者にとって心理的な壁が少なくなるというメリットもある。統計学習においては、統計分析以外のつまずき（ソフトウェアの操作やPCの操作）が極力ない状態で取り組むことが望ましいと考えられるため、HADによる統計学習の効率は高くなると考えられる。

Webの資料の活用 HADはWebサイトにソフトウェアそのものの使い方に加え、分析法ごとに操作方法を開示したスライドがアップされている⁸。そのため、学習者は書籍を購入せずとも、Webを用いて無償で操作方法を学習することができる。またHADの資料は、著者が作成したもの以外にも、学生教育向けにいくつかの大学でマニュアルが作成されている⁹。

また、HADの操作方法については、開発者である著者に対して匿名の質問を行うこともできるようになっている¹⁰。いつもすぐに返事が来るわけではないが、Web上の資料では解決できない問題を解決したい場合、HADの不具合などを見つけた場合は相談してもらえるとよい。

⁸ 「HADの使い方」 <http://norimune.net/713>

⁹ たとえば広島大学大学院教育学研究科の中島健一郎氏のWebサイトがある。

<http://kennakashi.jimdo.com/授業資料/>

¹⁰ ask.fm という質問サービスを利用している。 <https://ask.fm/simizu706>

シームレスな学習 最後に、分析手法がアルゴリズムによる分類ではなく、使用目的によって分類されている点も、統計学習において有効であると思われる。例えば、上述したようにクラスタ分析と混合正規分布モデルは、アルゴリズムは全く異なるが、分析目的としては「サブジェクトを分類する」という似た性質を持っている。比較的学習が容易なクラスタ分析の応用として混合正規分布を使ってみる、といったようにシームレスに統計手法に慣れていくことが可能な点が HAD で統計を学習するうえでも有効な点である。

また、商用ソフトウェアではアルゴリズムが異なると結果の出力が全く違うものになっているものもあるが、それでは学習者がそのたびに心理的にも知識的にも壁を感じてしまうだろう。HAD では、たとえば回帰分析と一般化線形モデル、階層線形モデルは「データを予測する」という点では同じであるため、極力モデルの指定方法と出力は似たものになるようにデザインされている。よって、重回帰分析を理解すれば、あとは少しモデルの指定を変えるだけで一般化線形モデルや階層線形モデルの結果が解釈可能ようになってきているのである。

このように、HAD では学生が統計分析について自習をするためのすぐれた環境を提供するツールであるといえよう。

他のソフトウェアとの併用 ただし、HAD は著者が一人で開発していることもあり、常に計算間違いが生じるリスクにさらされている点には十分注意しておく必要がある。HAD だけで学習を完結するのではなく、他のフリーソフトウェアも使えるようになることも重要である。また、統計の学習において異なるソフトウェアの分析結果がどのように一致し、異なるのかを知ることは大きな効果があると思われる。なぜなら、統計分析には様々なオプションがあり、その設定によって結果がいかに変化するかを知ることは統計手法を使いこなす上で必須であるからである。HAD による分析に慣れてきたら、他のソフトウェアと併用することが望ましいだろう。

HAD には HAD2R という、基礎統計分析について同様の分析が実行可能な R コードを出力する機能がある。R は無償で自由な統計解析用のフリーソフトウェアであり、またマルチプラットフォームである。そのため、いつでも誰でも利用可能であるという点は HAD よりも強い。ただし、R のコードを書くのは慣れが必要なため、最初は HAD2R を利用することで R の習熟がスムーズになるかもしれない。また、HAD に入力したデータセットを R にスムーズに読み込めるよう、csv ファイルを指定したディレクトリに作成する機能もある。これらを使うことで、PC 知識が十分でない学生も R で容易に分析できるようになることが期待できる。

3.2. 統計教育ツールとしての利用

続いて、HAD の統計教育ツールとしての特徴を論じる。2015 年度現在でも、著者の勤務校をはじめ、いくつかの大学ですでに HAD を用いた授業を行っているという報告を受けている。また、関西学院大学では、社会心理学研究センター主催で HAD を用いた統計教育に向けた FD 研

修も行われている¹¹。

HAD を大学の統計教育で用いる場合のメリットとして、1. 学生全員に同じ分析環境を提供できる、2. 分析結果に図表が出力され、編集しやすい、3. サンプルデータを簡単に作成できる、といったことが挙げられる。

全員に同じ分析環境を提供 大学における統計教育において、商用ソフトウェアの場合、アカウント数が学生の人数分用意できない場合もある。HAD は Excel が入っていれば動くため、大抵の大学の環境では学生全員に同じ分析をさせることができる。また、授業の課題も大学以外でも取り組めるといったメリットもある。

ただし、大学によっては Excel のマクロを使えないようにしている場合もあるかもしれない。その場合は、HAD をテンプレートファイルとして保存し、システム管理者にブックのテンプレートとして使えるように設定してもらうとよい。詳しい設定方法については Web サイト¹²を参照してもらいたい。テンプレート化することによって、1. 学生が授業ごとに HAD をダウンロードしなくてもよい、2. マクロのセキュリティ設定に依存しない、3. あらかじめサンプルデータを入力しておけばデータの再配布も必要ない、といったメリットがある。

分析結果に図表が出力される HAD はいくつかの分析については、グラフを分析結果に合わせて表示する。たとえば分散分析であれば、各群の平均や検定統計量に加えて、平均パターンを示す棒グラフを自動的に作成して出力する(図4)。このようにプログラムが自動的に結果の理解を助ける図表を出力するため、結果が理解しやすいことに加え、その後の復習の助けにもなる。

また、HAD が出力する図表は Excel を用いているため、学生でも簡単に編集することができる。よって、レポート作成時にも、統計以外の知識やスキルがなくても図表の作成や編集が可能であるため、統計の学習そのものにより時間と労力を費やせるだろう。

サンプルデータの作成 HAD では、多変量正規分布に従う乱数を簡単に生成することができる。

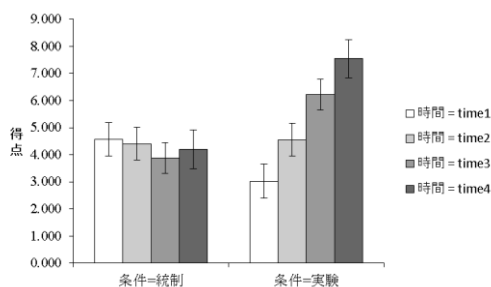


図4 HAD で出力される2要因分散分析の交互作用効果のグラフの例

¹¹ 「統計への苦手意識をなくす教育：フリーソフト HAD を利用した教授法の習得」

http://www.kg-rcsp.com/news/news_kg-rcsp6th/

¹² 「HAD を大学の授業などで使いたい場合」 <http://norimune.net/1678>

乱数の生成方法には2種類あり、1. 指定した母平均と母共分散行列をもつ正規乱数を生成する、2. 指定したものと全く同じ平均値と共分散行列を持つ正規乱数を生成する、というものがある。前者は得られるデータが指定した共分散行列と一致するとは限らない。後者は、指定した共分散行列と全く同じになる。

また、共分散行列の指定方法も、1. 直接入力する、2. 入力されているデータの共分散行列を指定する、という2つがある。これらを組み合わせることで、「教員が実際に研究で収集したデータと全く同じ平均値と共分散行列を持つ乱数」を生成することができるため、学生にリアリティのあるデータを再配布することも可能である。

HADでは正規分布以外にも、ポアソン分布、二項分布、負の二項分布、ガンマ分布といった指数分布族からの乱数も生成する機能がある。一般化線形モデルのような正規分布外の確率分布を用いたモデリングを教育する際にも活用することができるだろう。

また上記以外にも、標本抽出理論を視覚的に解説するための「サンプリング」機能がある。これは、ある大きさの標本をたくさん抽出することで標本分布が正規分布に漸近的に近づくことをデモンストレーションするものである。この機能を用いることで、区間推定や統計的検定のメカニズムについて直観的な理解を助けるだろう。

4. HAD を用いた研究実践

多くの研究者は、信頼できるソフトウェアを用いて研究成果を報告している。分析結果を報告するうえで、途中に計算間違いが含まれていると、研究の主張も間違えてしまうことになるからである。

HADは著者個人が開発していることから、どうしても計算間違いが含まれている可能性は否定できない。よって、HADの結果を直接報告することは全く推奨しない。最終的な分析結果は必ず信頼できるソフトウェアで検算なりをしてから報告をしてほしい。

しかし一方で、HADは結果とともに図表も表示されるため、分析結果の概要を知るのに便利である。HADをうまくデータ分析に組み入れることで、より効率的に、よりよい分析の選択ができるかもしれない。本章ではこのような研究実践への利用可能性について論じておこう。

4.1. 探索的分析用ツールとしての特徴

HADは、他のソフトウェアでは実行に時間がかかる分析法をより手軽に実行するための工夫がなされている。たとえば、重回帰分析の交互作用効果の検討や、分散分析の単純効果分析や多重比較、因子分析結果に基づいた尺度得点の計算、階層線形モデルにおける中心化などが挙げられる。また、HADは上述のように、分析結果を簡単に知るための図表が同時に出力される。このことから、他のソフトウェアに比べてよりスピーディーに分析が実行でき、またより効率的に分析結果の概要を知ることができるというメリットがある。

これらの機能は、研究者がデータを収集してすぐに分析結果の概要を知りたい場合、複数の分析手法の結果を比較して最適な方法を選択したい場合などに有効であると思われる。これらのHADの長所によって分析時間を短縮が可能なため、研究計画や結果の報告のための準備により多くの時間を割けるようになるだろう。

また、HADは心理統計の範囲とはいえ、いくつか高度な統計手法も実行可能となっている。具体的には、階層線形モデル、混合正規分布モデル、項目反応理論、マルチレベル構造方程式モデルなどがある。これらの統計手法は商用ソフトでも専用のものを使わなければならない場合が多く、データの受け渡しの作業が煩雑になることが多い。HADでは、これらの分析を同じデータセットですぐに実行可能なため、最適なモデルに到達するまでの時間が短くすむ上に、研究者自身の統計手法の習得にも役立つというメリットがある。また、人力で行う手間を極力少なくすることから、ヒューマンエラーも回避できる。

このように、HADを最終的な分析結果報告に使わずとも、その最終的な結果にたどり着くまでの探索的な分析において、分析時間の短縮やヒューマンエラー回避といった点からも、利用する価値はあるといえる。ただし、繰り返しになるが、最終的な分析結果の報告は他のソフトウェアで検算してから行うことを推奨する。

4.2. 共同研究における分析結果の共有

複数の研究者とともに共同研究をする場合、リサーチミーティングで分析結果をいち早く共有したいと思うことがあるだろう。しかし、分析やその結果の整理に時間がかかるといった理由で、「分析結果については後日に共有する」、といったことになることも少なくない。

HADは、このような、「とりあえず分析結果を共有したいとき」のツールとして有用である。それは、1. GUIのためすぐに分析に取りかかることができる、2. 分析結果に図表が出力されるため分析の解釈が容易である、3. 一通りの分析手法がそろっているため、思いつく分析手法を試すことができる、といった点が挙げられる。また、分析結果を送る際も、Microsoft Officeを持っている共同研究者ならば確認、編集、再分析が可能である点も便利である。

5. まとめ

本論文では、フリーの統計ソフトウェアであるHADについて、その概要と分析機能の紹介、統計学習・教育ツールとしての利用、そして研究実践における活用方法について解説した。

HADはユーザーからの不具合やプログラムミスの報告を受けた場合、すぐ修正するようにしている。そのため、数年前のバージョンに比べてかなり不具合は少なくなっている。しかし、すでに述べたようにHADの分析結果は常に信頼できるとは限らない。また、HADの免責事項にもあるように、分析結果について開発者はその責任を負わない。よって、分析結果を報告する場合は自己責任で行う必要がある点には注意が必要である。

これらの注意点を踏まえれば、本論文で挙げたような HAD のいくつかのメリットをユーザーは無償で利用することができる。ぜひ研究や大学教育で活用してほしい。

謝辞

HAD の開発にあたり、広島大学総合科学部社会心理学研究室の卒業生をはじめ、その他数多くの HAD ユーザーの助力を得ました。また、Mac 版の HAD の開発では、滋慶医療科学大学院大学医療管理学研究科講師の岡耕平氏、関西学院大学文学部教授の三浦麻子氏、小川洋和氏の多大な助力を得ました。この場を借りて感謝申し上げます。

引用文献

- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Everitt, B.S. & Hand, D.J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, 1(5), 799–821.
- 久保拓弥 (2012). データ解析のための統計モデリング入門 一般化線形モデル・階層ベイズモデル・MCMC 岩波書店
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- 松村真宏・三浦麻子 (2014). 人文・社会科学のためのテキストマイニング[改訂新版] 誠信書房
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376–398. doi:[10.1177/0049124194022003006](https://doi.org/10.1177/0049124194022003006)
- 清水裕士 (2014). 個人と集団のマルチレベル分析 ナカニシヤ出版
- 清水裕士・村山綾・大坊郁夫 (2006). 集団コミュニケーションにおける相互依存性の分析(1) コミュニケーションデータへの階層的データ分析の適用 電子情報通信学会技術研究報告, 106(146), 1–6.
- 豊田秀樹 (1998). 共分散構造分析 入門編 朝倉書店
- 豊田秀樹 (2000). 共分散構造分析 応用編 朝倉書店
- 豊田秀樹・池原一哉 (2011). 変数間の関係性を考慮してクラスター数を決定する k-means 法の改良 心理学研究, 82(1), 32–40. doi:[10.4992/jipsy.82.32](https://doi.org/10.4992/jipsy.82.32)