

フリーの統計ソフトウェア HAD の多言語化と英語版の作成

清水 裕 士 (社会学部・研究代表者)
三 浦 麻 子 (文学部)
稲 増 一 憲 (社会学部)
小 川 洋 和 (文学部)

要 旨

本研究の目的は、統計ソフトウェア HAD を多言語環境で利用できるようにすることであった。統計ソフトウェア HAD は、研究代表者が開発した Microsoft Excel で動くフリーの統計解析用プログラムである。HAD は基礎的な分析から統計的検定、データの可視化、分散分析や回帰分析、因子分析等の多変量解析を実行することができる。これらの機能はすべて無償で利用でき、すでに本学を含め日本の大学で統計教育や研究に利用されている。現状、HAD は日本語にしか対応しておらず、英語用の OS や他言語環境において動作しない。しかし、近年は日本の大学では留学生が増加し、学生の PC に関わる環境もさまざまである。また、本学はスーパーグローバル大学 (SGU) として留学生にも開かれた統計教育が実施されることが求められていることから、HAD の多言語環境における利用が望まれていると考えられる。そこで本研究では HAD について、1. システムそのものを多言語に対応可能、かつ、英語 OS で動作可能なプログラムに組み替えること、2. HAD 内部の文章をすべて英語に翻訳し、それらに対応させて読み込めるようにすることを目的とした。結果、HAD の多言語化に成功し、英語版 OS での動作を確認した。

はじめに

本報告は、高等教育推進センター2016年度共同研究助成「フリーの統計プログラム HAD の多言語 OS への対応と英語版マニュアル作成」に関するものである。最初に、統計ソフト HAD についての解説を行ったあと、HAD の国際化の必要性を論じる。続いて、HAD のプログラムが依拠する Microsoft Excel の Visual Basic Application の仕様と、国際化を可能にするためのアルゴリズムについてのアイデアを論じる。最後に、英語翻訳作業とその成果について報告し、本研究によってもたらされるだろう統計教育におけるメリットや可能性について展望をまとめる。

なお、このプロジェクトは清水裕士 (社会学部・准教授)、三浦麻子 (文学部・教授)、稲増一憲 (社会学部・准教授)、小川洋和 (文学部・教授) をメンバーとしている。

HAD とは

HAD は、研究代表者が2005年から開発を行っているフリーの統計ソフトウェアである（清水・村山・大坊、2006）。HAD は研究代表者が個人で運営している Web サイト（<http://norimune.net>）から自由にダウンロードできる。当初はマルチレベル分析のための分析補助ツールであったが、2010年以降に基本的な統計解析法が加えられ、2017年現在、商用ソフトと並ぶほどの機能が実装されており、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（表1）。

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。HAD バージョン16現在、Windows では Microsoft Office2010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、MacOS についてはバージョンによっては起動時にエラーが生じて利用できない可能性があること、また機能が一部（構造方程式モデル）制限されていることから、Windows の PC での利用が強く推奨されている。

HAD は無償で自由なソフトウェアである。HAD は利用が無償であり、誰でも何度でも Web サイトからダウンロードして利用可能である。また、HAD の VBA ソースコードは誰でも閲覧が可能であり、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) の ver 2 以降のライセンスに基づいているので、ライセンスが指定するルールに基づいての利用が求められる。

また、統計教育用の機能もいくつか搭載されており、統計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、清水（2016）に詳しいが、本報告でも簡単にその有効性を論じておこう。

表1 HAD で可能な分析手法（回帰分析系）

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	

HAD で可能な分析手法（因子分析系）

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、オプティミ、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

HAD が統計教育ツールとして有効な点として、第一に HAD は Excel で動作することが挙げられる（図 1）。Excel は多くの大学で学内の PC 環境に含められていること、また多くの Windows OS が搭載された PC にインストールされていることから、学生が大学や自宅など環境を選ばずに統計解析を行うことができるというメリットがある。第二に、HAD は無償かつ自由なソフトであるため、学生がいつでも何度でもダウンロードでき、経済的な点での敷居がないことが挙げられる。それによって、学生は大学卒業後も自由に統計解析を行うことができる。第三に、HAD に関する資料は Web（主に研究代表者の Web サイト）から無償で手に入れることができる。また、初心者用のマニュアルも作成されていることから、初学者にとっても敷居が低くなっている。第四に、HAD はアルゴリズム別ではなく、分析目的別に統計手法が分類されていることから、事前の統計知識が少ない学生でも比較的簡単に使いこなすことができる点にある。

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	変数名	teamID	playerID	player_name	team_name	num	league	labor_cost	salary	AVR	ATbase	ATbats	HIT
2													
3	分析		選択セルを使用	統計変数を投入	変数を左につめる					● 変数情報		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	teamID	salary	league									
10													
300	変数情報	フィルタ	値ラベル	ラベル	@変数	コード							
301	teamID												
302	playerID												
303	player_name			選手名									
304	team_name			チーム名									

図 1 HAD の画面例

これらの利点から、HAD はすでに複数の大学で統計教育のための標準ソフトとして活用されている実績がある (清水、2016)。

HAD の多言語化の意義

HAD はもともと研究代表者本人のための個人的なソフトウェアであったが、その後日本の大学生を対象とした統計教育ツールを目的とした開発が行われてきた。後述するように、もともと HAD は日本人の利用のみを想定してきたため、その開発仕様が日本語対応の OS、かつ、日本語対応の Office での動作のみが開発対象となっていた。

しかし、統計教育は日本人学生だけではなく、留学生も当然その対象となる。また、上記のように統計分析環境は大学の PC だけではなく、個人所有の PC で動作することが望まれることから、留学生が個人で所有する PC での動作確認もニーズとして高まってきている。研究代表者が所属する社会学部では、中国や韓国の留学生も多く、院生は個人の所有する PC で分析することも多いため、HAD を用いた統計教育が部分的なものにならざるを得なかった点が限界点として見えてきていた。

加えて、関西学院大学はスーパーグローバル大学 (SGU) に選出され、より多言語に開かれた統計教育の実施が求められてきている。留学生の増加による多言語化のニーズの高まりはもちろんのこと、日本にいた学生が留学先の PC 環境でこれまで学習してきた統計分析を実施できることも重要な検討事項である。

そこで学生への統計教育をサポートするツールとしての HAD に多言語化機能を追加することは、本学のみならず日本の大学における統計教育の国際化を促進する意味で非常に重要な意義があるといえる。

HAD の開発環境と多言語化への工夫

HAD の開発は研究代表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。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。

当初 HAD は、多言語化は想定していなかったため、メッセージや結果の出力で表示される日本語の文章はソースコードに直接入力されていた。そのため、日本語が搭載されていない OS で VBA をコンパイルするとエラーとなってしまい不具合があった。つまり、日本語が含まれていない OS では動作しない仕様となっていたのである。

多言語化を実現するためには、ソースコード中の日本語をすべて外に吐き出し、外部から文章を読み込んで出力する必要がある。そのため、すべての日本語文章を Excel ファイル内のシートの各セルに保存しておき、ソースコードからそれらを読み出して出力する仕様に変更した。

続いて多言語化を実装するために、言語設定を切り替えるごとに文章を読み込むシートを切り替える仕様とした。たとえば日本語設定の場合は「\$Japanese\$」シートから文章を読み込むが、英語設定の場合は「\$English\$」のシートから読み込むということである。これによって、言語

	A	B	C	D	E
1	基本システム用語	メッセージボックス	出力	ステータスバー	コントロール
2	データ	統計分析ソフト HAD	※Startに入力するためのinput	HAD実行中...	データ読み込み
3	モデリング	HADファイルをアクティブにしてください	要約統計量	HAD実行中...	モデリングシート
4	ソルバー アドイン	エラー内容:	箱ひげ図	HAD実行中...	分析列幅の調整
5	シート管理	変数名を選択してください。	ヒストグラム	HAD実行中...	分析数値計算
6	HADを起動	変数より多い変数の使用は分析できません。	度数分布	HAD実行中...	因子HAD2R
7	変数の投入	分析コードを選択してください。	正規性の検定	HAD実行中...	クラ TTMHAD
8	変数の作成	分析コードが正しくありません。	散布図	HAD実行中...	等差分析
9	分析コードを実行	分析コードが不連続です。	群ごとの統計量	HAD実行中...	構造化データシート
10	分析コードをHADに反映	モデリングシートの設定が初期化されます。	カーネル密度推定	HAD実行中...	出力使用変数
11	記述統計量	初期化しますか？	クロス集計表	HAD実行中...	力付 選択セルを使用
12	ヒストグラム	セルを選択してください。	平均値の検定	HAD実行中...	多重統計変数を投入
13	メインデータセットとして利用	データに数式エラーが含まれています。	多重クロス表	HAD実行中...	対数変数を左につめる
14	変数をデータセットに追加	数式エラーを欠損値に変換しますか？	主成分分析	HAD実行中...	対数変数情報
15	分析コード	データに空白セルが含まれています。	対応のない平均順位の差の検定	HAD実行中...	シミュレーション実行/回帰分析
16	変数情報	データの欠損値や変数の数を確認してください。	対応のある平均順位の差の検定	HAD実行中...	分散因子分析
17	使用変数:	データにスペースのみのセル(変数名	ピアソンの積率相関係数	HAD実行中...	分類変数の作成

図2 外部に出力された日本語文章

間の切り替えが容易であるだけでなく、言語の種類を増やすのも文章を保存したシートを増やすだけで良くなるため、今後の拡張性が高くなることもこの方法の利点であるといえる。

日本語文章の英訳

多言語化機能の実装によって、メッセージや出力の文章、ユーザーフォームのCaptionをファイル内のシートに保存することができた。あとは、これらの日本語を英語に翻訳し、英語用のシートに保存することで英語版のHADとして使用することができる。

そこで統計学に素養がある日本語と英語のバイリンガル1名(文学部総合心理科学科の学生)に翻訳を依頼した。作業はすべての日本語文字列(24629文字)を提出し、メッセージ用文章は逐次訳を、ユーザーフォームは日本語と同じ長さになるよう一部意識を、出力の文章は統計用語に対応するよう留意しながら翻訳作業を行うよう求めた。

依頼した翻訳について、研究代表者が統計的な専門用語についていくつか修正を行った。またユーザーインターフェースに合った長さになるように一部文章の長さを調整した。これらの英文をHAD内のシートに保存することで、表示言語を英語に切り替えたときにHADの表示がすべて英語に切り替えることができるようになった。

英語版HADの成果

英語版HADはHADの「設定」ボタンを押して開くユーザーフォームから、言語の設定を変更することで利用できる(図3)。また、英語版に切り替えた場合のソフトウェアは図4のとおりである。図4上はHADのモデリングシート(分析計画を入力する画面)、図4下は分析法を選択するユーザーフォームである。

また、英語版HADを英語版OS、英語版Microsoft Officeで起動するかを確認したところ、データの読み込みや分析を無事行うことができた。これにより、日本語対応していないOSでもHADの利用可能であることが示された。

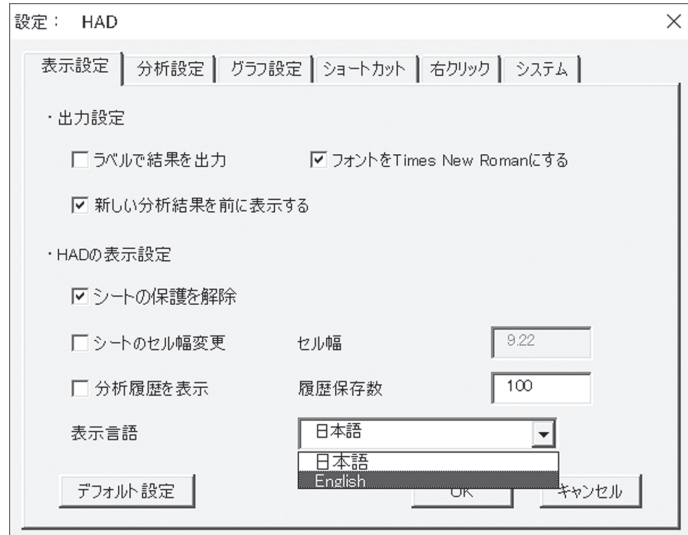


図3 表示言語の切り替え

	A	B	C	D	E	F	G	H	I	J	K	L	M
Variables	teamID	playerID	player_name	team_name	num	league	labor_cost	salary	AVR	ATbase	ATbats	HIT	
Analysis			Input variable	Control variable			Move left			<input checked="" type="radio"/> Variables		Create variable	
DATA Sheet										<input type="radio"/> Regression			
										<input type="radio"/> Factor Analysis		Data set	
UseVariable	teamID	salary	league										
Variables	Filter	Value	Labels	@variable	コード								
teamID													

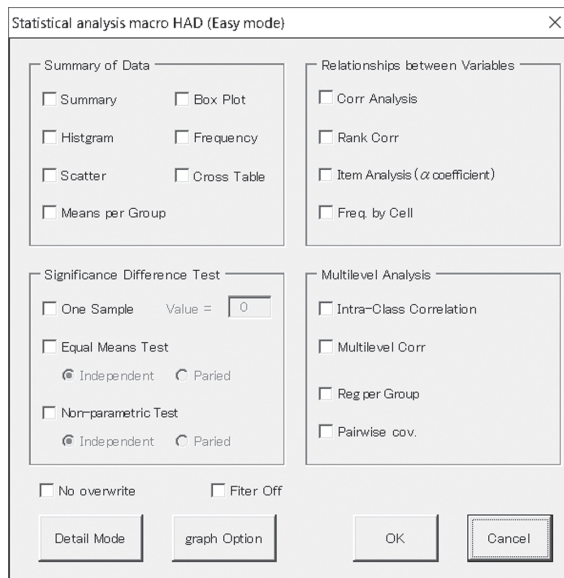


図4 英語版の表示

多言語化・英語版 HAD の今後の展開

本研究によって、統計ソフト HAD の多言語化および英語版の作成が達成された。この成果は今後の統計教育における国際化を大きく貢献することが期待される。本学では SPSS などの商用ソフトウェアが学生でも利用可能であるが、個人の PC にインストールする場合は安価であるとはいえず有償である。それに対し、HAD は Excel さえあれば言語環境を選ばず統計分析を行うことができる。

ただし、HAD にもまだ統計教育の国際化においていくつか制限がある。第一に英語版のマニュアルの不在が挙げられる。日本語の資料は豊富ではあるがこれらは未だ英語への翻訳が行われていない。今後の研究の課題である。第二に、中国語版や韓国語版に対応していない点である。日本の大学にいる留学生の多くはアジアから来た学生である。よって、統計教育の国際化を進めていくためには、これらの言語への対応が不可欠である。これらの限界点を解決することによって、HAD は留学生にとっても有効な統計ソフトとして貢献すると考えられる。

引用文献

清水裕士 (2016). 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 メディア・情報・コミュニケーション研究、1、59-73.

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.