

RExcel: ExcelでRを使う (6)

(独)農業・食品産業技術総合研究機構
農村工学研究所農村計画部主任研究員

合崎 英男 (Aizaki Hideo)

■2000年3月北海道大学大学院農学研究科博士後期課程修了。博士(農学)。農林水産省農業研究センター研究員、農業工学研究所研究員、同主任研究官を経て、06年4月より現職。専門分野は農業経済学(主に環境配慮や食品安全性に関する意思決定分析)。



1. はじめに

RExcelシリーズ第6回では、前号に引き続きRとExcel VBA(以下、VBA)を利用したマクロ・プログラムを試作します。

今号は、コンジョイント分析の質問への回答結果を分析するプログラムです。前号のプロファイル作成プログラムを利用してプロファイルを作成・出力したシートに回答結果を入力し、分析を実行します。

プログラムは、前号と同じく(助)統計情報研究開発センターのWebサイト(<http://www.sinfonica.or.jp/>)内の[刊行物]>[ESTRELA]>[参考]から、テキストファイルとしてダウンロードできます。入力作業を省きたい方は、同ファイルをご利用下さい。プログラムの入手/入力方法の違いにかかわらず、本プログラムについてもRと同様に、利用者の自己責任のもとでご利用下さい。著者及び(助)統計情報研究開発センターともに、本プログラムの利用に関して一切の責任を負いません。

2. 分析方法の概要

本プログラムでは、「順位付け」あるいは「評定」によるコンジョイント分析の質問への回答結果を、Rにデフォルトでインストールされるstatsパッケージ内の関数lm(線形回帰分析)によって分析します。各属性変数(説明変数)には、その属性に含まれる水準を「1」からの整数値として表現したデータを用いますが、Rの関数factorを利用して(Rでいうところの)因子に変換し、さらにRの関数contrastsによって各属性変数(因子)に含まれる(各水準に対応した)係数推定値の和がゼロになるように制約(対比)を課しています(対比については文献[7]の第5章)。各属性の重要度は、当該属性に含まれる係数推定値の最大値と最小値の差(効用差)を計算した上で、すべての効用差の合計で各効用差を除して100を乗じることで求めています。

なお、本プログラムとは別のアプローチを紹介しているR関連の文献[4][9]もありますので、関心のある方はご参照下さい。

図 1 回答結果分析プログラム(1)

```

1行 Option Explicit
2行
3行 Sub AnalyzingResponses()
4行
5行     Dim NumA As Integer, NumP As Integer, NumL() As Integer
6行     Dim MaxL As Integer, SumL As Integer, rTab As Integer
7行     Dim i As Integer, j As Integer, k As Integer, l As Integer
8行     Dim MaxU As Double, MinU As Double, SumU As Double
9行     Dim DifU() As Double, cmdstr As String
10行
11行     Range("B2").CurrentRegion.Name = "TabAL"
12行     NumA = Range("TabAL").Rows.Count - 1
13行     NumP = Range(Cells(NumA + 5, 2), _
14行                 Cells(NumA + 4, 2).End(xlDown)).Rows.Count
15行     Range(Cells(NumA + 4, 2), _
16行           Cells(NumA + NumP + 4, NumA + 2)).Name = "DataSet"
17行
18行     MaxL = Range("TabAL").Columns.Count - 1
19行     ReDim NumL(1 To NumA)
20行     For i = 1 To NumA
21行         NumL(i) = Application.WorksheetFunction.CountA _
22行                 (Range(Cells(i + 1, 2), Cells(i + 1, MaxL + 1)))
23行     Next
24行

```

3. プログラムの解説

(1) 各種情報の取り込み

図 1 は、Excel のシート上に整理した各種情報を取り込む作業を実行する部分のプログラムです。

属性・水準を整理したセル範囲から属性数を求める一方 (11 ~ 12 行)、プロファイル集合が入力されている範囲からプロファイル数を算出しています (13 ~ 14 行)。それらの情報を利用して、プロファイルと回答結果が入力されているセル範囲を、名前で参照できるように設定しています (15 ~ 16 行)。さらに、水準数の最大値を求めた上で、各属性の水準数を保存しています (18 ~ 23 行)。

(2) R での分析の実行

図 2 は、Excel から R ヘデータを転送し、分析を実行する部分のプログラムです。

R と Excel を結びつけ (25 行)、Excel 上で DataSet と名付けた範囲を R ヘデータフレーム mydf として転送しています (27 行)。データフレームに含まれる説明変数の名前を “x+ 数字” (数字は「1」からの整数) に変更した上で (28 ~ 29 行)、被説明変数と因子に変換した説明変数を新たなデータフレームに保管しています (30 ~ 34 行)。各属性に含まれる水準の係数推定値の和がそれぞれゼロになる制約を課した上で回帰分析を実行し、その結果を保存しています (35 ~ 37 行)。

(3) 分析結果の Excel への転送

図 3 は、分析結果を R から Excel へ転送・整理する作業を実行する部分に該当します。

分析結果を表示 (転送) する Excel のシート上の範囲 (表) に表頭項目を設定した上で (39 ~ 46 行)、係数推定値等を転送しています (48

図2 回答結果分析プログラム(2)

```

24行
25行 Rinterface.StartRServer
26行
27行 Rinterface.PutDataframe "mydf", Range("DataSet")
28行 Rinterface.RRun "xs <- paste('x', 1:(ncol(mydf)-1), sep='')"
29行 Rinterface.RRun "names(mydf) <- c(xs, 'y')"
30行 Rinterface.RRun "mydff <- subset(mydf, select = y)"
31行 For i = 1 To NumA
32行     cmdstr = "mydff$x" & i & "<- factor(mydff$x" & i & ")"
33行     Rinterface.RRun cmdstr
34行 Next
35行 cmdstr = "options(contrasts = c('contr.sum', 'contr.sum'))"
36行 Rinterface.RRun cmdstr
37行 Rinterface.RRun "lmout <- summary(lm(y~., data = mydff))"
38行

```

図3 回答結果分析プログラム(3)

```

38行
39行 rTab = NumA + NumP + 6
40行
41行 Cells(rTab, 3).Value = "係数推定値"
42行 Cells(rTab, 4).Value = "標準誤差"
43行 Cells(rTab, 5).Value = "t値"
44行 Cells(rTab, 6).Value = "p値"
45行 Cells(rTab, 7).Value = "効用差"
46行 Cells(rTab, 8).Value = "重要度"
47行
48行 k = 2
49行 l = 2
50行 Cells(rTab + 1, 1).Value = "定数"
51行 Cells(rTab + 1, 3).Name = "Const"
52行 Rinterface.GetArray "t(lmout$coef[1,])", Range("Const")
53行 For i = 1 To NumA
54行     Cells(rTab + k, 1).Value = Cells(i + 1, 1).Value
55行     k = k + 1
56行     For j = 1 To NumL(i) - 1
57行         Cells(rTab + k, 2).Value = Cells(i + 1, j + 1).Value
58行         cmdstr = "t(lmout$coef[" & l & ",])"
59行         Cells(rTab + k, 3).Name = "Var"
60行         Rinterface.GetArray cmdstr, Range("Var")
61行         k = k + 1
62行         l = l + 1
63行     Next
64行     Cells(rTab + k, 2).Value = Cells(i + 1, NumL(i) + 1).Value
65行     Range(Cells(rTab + k - NumL(i) + 1, 3), _
66行         Cells(rTab + k - 1, 3)).Name = "Coef"
67行     Cells(rTab + k, 3).Value = _
68行         -1 * Application.WorksheetFunction.Sum(Range("Coef"))
69行     k = k + 1
70行 Next
71行

```

図 4 回答結果分析プログラム(4)

```

71行
72行 ReDim DifU(1 To NumA)
73行 SumL = 0
74行 For i = 1 To NumA
75行     j = rTab + 1 + i + SumL
76行     MaxU = WorksheetFunction.Max _
77行         (Range(Cells(j + 1, 3), Cells(j + NumL(i), 3)))
78行     MinU = WorksheetFunction.Min _
79行         (Range(Cells(j + 1, 3), Cells(j + NumL(i), 3)))
80行     DifU(i) = MaxU - MinU
81行     Cells(j, 7).Value = DifU(i)
82行     SumL = SumL + NumL(i)
83行 Next
84行
85行 SumL = 0
86行 SumU = WorksheetFunction.Sum(DifU)
87行 For i = 1 To NumA
88行     j = rTab + 1 + i + SumL
89行     Cells(j, 8).Value = DifU(i) / SumU * 100
90行     SumL = SumL + NumL(i)
91行 Next
92行
93行 i = rTab + SumL + NumA
94行 Cells(i + 3, 1).Value = "決定係数"
95行 Cells(i + 4, 1).Name = "RS"
96行 Rinterface.GetArray "lmout$r.squared", Range("RS")
97行 Cells(i + 3, 2).Value = "自由度調整済み決定係数"
98行 Cells(i + 4, 2).Name = "ARS"
99行 Rinterface.GetArray "lmout$adj.r.squared", Range("ARS")
100行
101行 Rinterface.StopRServer
102行
103行 End Sub

```

～70行)。係数推定値にはゼロ和制約が課されているため、各属性の最後の水準（j水準属性なら「第j水準」）に該当する推定値が、関数lmの出力には含まれていません。そこでゼロ和制約から、当該推定値をExcel上で計算・表示しています（65～68行）。

(4) 重要度の計算と決定係数の転送

図4は、各属性の重要度の計算と決定係数を転送するコードを示します。

各属性の水準の係数推定値からそれぞれ最大値と最小値を探索し、その差を「効用差」列に

表示させています（72～83行）。すべての属性の効用差を求めたところで、各属性の重要度を計算・表示しています（85～91行）。最後に、（自由度調整済み）決定係数をRからExcelへ転送・表示しています（93～99行）。

4. 実行例

前号と同じ3つの属性a、b、cから構成される商品を対象としたコンジョイント分析の実行例を考えます。属性aとbは2水準属性、属性cは4水準属性です。

本プログラムでは、水準を数字で表現したブ

図5 プログラムの実行結果例

	A	B	C	D	E	F	G	H	I
1	属性	第1水準	第2水準	第3水準	第4水準				
2	a	a1	a2						
3	b	b1	b2						
4	c	c1	c2	c3	c4				
5									
6	oa								
7		A	B	C	回答		a	b	c
8		1	2	2	1	6	1 a2	b2	c1
9		2	1	2	4	4	2 a1	b2	c4
10		3	1	2	2	7	3 a1	b2	c2
11		4	2	1	2	5	4 a2	b1	c2
12		5	2	1	4	1	5 a2	b1	c4
13		6	1	1	1	8	6 a1	b1	c1
14		7	2	2	3	2	7 a2	b2	c3
15		8	1	1	3	3	8 a1	b1	c3
16									
17			係数推定値	標準誤差	t値	p値	効用差	重要度	
18	定数		4.5	0.176777	25.45584	0.00154			
19	a						2	28.57143	
20		a1	1	0.176777	5.656854	0.029857			
21		a2	-1						
22	b						0.5	7.142857	
23		b1	-0.25	0.176777	-1.41421	0.292893			
24		b2	0.25						
25	c						4.5	64.28571	
26		c1	2.5	0.306186	8.164966	0.014671			
27		c2	1.5	0.306186	4.898979	0.039231			
28		c3	-2	0.306186	-6.53197	0.022644			
29		c4	-2						
30									
31	決定係数	自由度調整済み決定係数							
32	0.988095	0.958333							

ロファイル集合の右側の列に、回答結果を入力します（図5のE列）。プロフィールに対する評価は、「順位付け」と「評定」のどちらかですが、ここでは「順位付け」とします。今回は8枚のプロファイルがありますので、1～8位の順位が付けられます。重複順位はないとして、Excelに入力する「回答」は、1位を「8」、2位を「7」、…、8位を「1」と、高順位ほど大きい値を割り当てます。

使用するプロフィール集合は、前号と同じ条件で作成したものです。しかし、前号のプログラムは、実行状況によって異なるプロフィール集合を出力する可能性があります。そこで、ご自身の前号のプログラムの実行結果（プロフィール集合）と、図5のプロファイル集合を比較し、異なる部分があるときには、図5のプロファ

イル集合に一致するように、ご自身のプロフィール集合を修正して下さい。この確認・修正作業の後、図5のセルE7～E15の変数名とデータを入力して下さい。

以上の準備が完了したところで、Excelのメニューから「ツール」→「マクロ」→「マクロ」と進み、「マクロ名」として「AnalyzingResponses」を選択して[実行]ボタンを押します。

本プログラムでは、ダイアログボックスを通じた各種情報の入力はありません。必要な情報を自動的に収集した上で分析を開始し、プロフィール集合の下方のセル範囲に結果を出力します。

Excelのシート上に出力される分析結果は、「係数推定値 (Estimate)」「標準誤差 (Std.

error)」「t値(t value)」「p値(Pr(>|t|))」「効用差」「重要度」「決定係数 (Multiple R-squared)」「自由度調整済み決定係数 (Adjusted R-squared)」です。ただし、カッコ内は、関数 `lm` がデフォルトで出力する際の表記で、Excel では表示されません。

係数推定値は、属性ごとにまとめられ、各属性の最終水準の係数推定値はゼロ和制約を利用して計算しています。たとえば、4水準属性 `c` であれば、水準 `c1`、`c2`、`c3` の係数推定値の合計値の符号を逆転させることで、水準 `c4` の係数推定値を求めています。

効用差と重要度は、それぞれ属性ごとに1つの値として計算されるため、属性名を配置している行に表示しています。効用差は、その属性に含まれる水準の係数推定値のうち、最大値と最小値の差として求めています。たとえば、4水準属性 `c` では、最大の係数推定値 2.5 と最小の係数推定値 -2 の差である 4.5 となります。すべての効用差の合計で各効用差を除し、さらに 100 を乗じて得た値が、各属性の重要度になります。ここでは効用差の合計が 7 ですから、重要度は属性 `a` が 28.6、属性 `b` が 7.1、属性 `c` が 64.3 となっています。

5. おわりに

半年間にわたって連載 RExcel シリーズにお付き合いいただきありがとうございます。少しでもお役に立てる情報を提供できていれば、私としても大変うれしく思います。

最後になりましたが、同志社大学の金明哲先生には、本シリーズ執筆のご縁を結んでいただきましたこと、記して感謝申し上げます。

* 参考情報

実証経済学のさまざまな分野で活用されてきているコンジョイント分析に類似の手法として、選択実験 / 選択型コンジョイント分析がありますが、プロファイルの作成や回答結果の分析などの方法が、通常のコンジョイント分析と異なります ([1][5][6][10][11][12])。R で選択実験を試みたい方は、文献 [3] が参考になると思います。ただし、同文献では選択肢集合の作成に AlgDesign パッケージ [13] の利用を紹介していますが、現在では DoE.base パッケージ [8] を利用した方が効率的と思います。また、RExcel を利用した選択実験向けアプリケーション [2] もあります。興味のある方は農村工学研究所の Web サイト (<http://nkk.naro.affrc.go.jp/>) 内の [お役立ち情報] > [データベース・プログラム] > [選択実験 / 選択型コンジョイント分析向けアプリケーション及び関連情報] をご参照下さい。

* 参考文献

- [1] 合崎英男 (2005) : 農業・農村の計画評価 : 農林統計協会 .
- [2] 合崎英男 (2009) : R を活用した選択実験向け選択肢集合の作成およびデータ分析用アプリケーションの開発 : 行動計量学, 36(1), pp.35-46.
- [3] 合崎英男・西村和志 (2007) : データ解析環境 R による選択型コンジョイント分析入門 : 農村工学研究所技報, 206, pp.151-173.
- [4] 朝野熙彦編著 (2008) : R によるマーケティング・シミュレーション : 同友館 .
- [5] Bateman, I. J., et al. (2002) : Economic Valuation with Stated Preference Techniques : Edward Elger.
- [6] 出村克彦・山本康貴・吉田謙太郎編著 (2008) : 農業環境の経済評価 : 北海道大学出版会 .
- [7] 船尾暢男 (2008) : R Commander ハンドブック : オーム社 .
- [8] Groemping, U. (2010) : DoE.base: R package (<http://www.r-project.org/>) .
- [9] 神田範明監修 / 石川朋雄・小久保雄介・池畑政志著 (2009) : 商品企画のための統計分析 : オーム社 .
- [10] Kanninen, B. (Ed.) (2007) : Valuing Environmental Amenities Using Stated Choice Studies : Springer.
- [11] 栗山浩一・庄子康編著 (2005) : 環境と観光の経済評価 : 勁草書房 .
- [12] Louviere, J. J., Hensher, D. A., and Swait, J.D. (2000) : Stated Choice Methods : Cambridge University Press.
- [13] Wheeler, B. (2006) : AlgDesign: R-package (<http://www.r-project.org/>) .