



Rパッケージsupport.
CEsとmlogitを利用した選択型コンジョイント分析

メタデータ	言語: Japanese 出版者: 公開日: 2022-04-14 キーワード (Ja): キーワード (En): 作成者: 鈴木, 真 メールアドレス: 所属:
URL	https://doi.org/10.24729/00017652

Rパッケージsupport.CEsとmlogitを利用した 選択型コンジョイント分析

鈴木 真

1. はじめに

我々は生活を送るうえで常に多くの選択場面にさらされている。例えば外出時の服装や目的地への交通手段、飲食店でのメニュー、日用品の購入など我々は選択の連続によって生活が成り立っているとみえる。ミクロ経済学の観点でみると、こうした選択場面において我々は価格やブランド、色、形、大きさなどそれぞれの選択肢にある様々な特性を自身の中で天秤にかけ、最も効用が高くなるものを選択していると考えることができる。このように個人が選択集合の中から代替案を選択する事象について定式化したものを離散選択モデルという(土田, 2010)。

離散選択モデルの基礎となるモデルの1つに、多項ロジット・モデル (Multinomial Logit Model) がある。多項ロジット・モデルは順序のない独立した3つ以上ある選択肢の中から1つの選択肢を選ぶという問題を取り扱うものとして用いられる。選択肢に共通で選択状況ごとに異なる説明変数と選択肢ごとに異なる説明変数によりモデルの呼び方を分けることもあるが、本稿ではまとめて多項ロジット・モデルと呼ぶことにする。

多項ロジット・モデルは選択者がそれぞれの選択肢の持つ複数の属性から選択者が得られる効用を評価することによって決まるということを前提にしており(池尾他, 2010, p.67)、近年においても多数の研究報告で多項ロジット・モデルを使った分析が用いられている(Bediul et al., 2018; Jesper, 2015; Judit et al., 2019; Quan et al., 2019; Tarek & Gustavo, 2019; Yun et al., 2020)。

多項ロジット・モデルによる分析を行うための分析環境として、データ解析環境R(以下、R)がある。Rでは種々のパッケージをインストールすることで、それに応じた分析が可能となる。Rは開発者チーム(R Development Core Team)によって基本パッケージ(base package)が開発・保守されている。さらに様々な研究者等によって開発されたユーティリティソフトウェアやデータ分析に特化した応用ソフトウェア等が無料で公開されている拡張パッケージ群(contributed packages)から構成されている(荒木, 2007)。

各パッケージ群の操作方法についてはCRANサイトで英語によるマニュアルが無料で公開されているものの、英文や統計そのものに不慣れな初学者や大学生などにはこの英文による

マニュアルはRの活用への障壁のひとつとなっている。パッケージによってはマニュアルのような形で実例を交えつつ操作方法を説明する邦文による文献等がある（荒木，2007；合崎 & 西村，2007；中谷，2010；合崎，2015；伊藤，2016；森，2017a；森，2017b）が、数多くあるパッケージの一部に対するものに限られている。

そこで本稿では、Rパッケージのsupport.CEsとmlogitを用いた、多項ロジット・モデルによる分析手順を説明する。特に、マーケティング・サイエンスの分野で消費者の選好構造を分析するために利用されている選択型コンジョイント分析を念頭に置く。選択型コンジョイント分析は、離散選択実験とも呼ばれ、後者の呼び名を推奨する研究者もいるが（Louviere et al., 2010）、本稿ではマーケティング・サイエンスの分野での慣例に従い、選択型コンジョイント分析と呼ぶことにする。第2節では、第3節以降で説明に用いる調査を前提として、多項ロジット・モデルについて簡潔に整理する。第3節では、選択型コンジョイント分析を実施するための準備を、support.CEsパッケージを用いて行う手順を説明する。第4節では、第3節での準備を踏まえて作成した人工データに基づき、mlogitパッケージを用いて推定した結果を説明する。また、想定回答者数の変更が推定結果にどの程度影響するかを検証する。第5節で本稿をまとめる。補論では、第4節の分析に用いる人工データを作成するためのコードを掲載する。

2. 多項ロジット・モデル

鈴木・中山（2020，第4節）に基づき、多項ロジット・モデルについて整理する。選択肢は2つの仮想的な施設と、どちらの施設も選ばないという3つである。選択肢 i ($= 1, 2, 3$) の効用は以下の通り表される。

$$U_i = V_i + \varepsilon_i = \alpha + x_i^t \beta + \varepsilon_i, \quad i = 1, 2 \quad (1)$$

$$U_3 = V_3 + \varepsilon_3 = \varepsilon_3 \quad (2)$$

α は2つの選択肢に共通するベースラインの効用である。選択肢3（選択肢1、2のどちらも選ばない）のベースラインの効用はゼロとする。言い換えると、選択肢3のベースライン効用を基準として α は定義されている。 x_i^t は選択肢 i に対応した説明変数ベクトルである。 t は転置を表す。 β は説明変数ベクトルに対応する係数ベクトルである。 ε_i ($i = 1, 2, 3$) は独立した同一のガンベル分布（第一種極値分布）に従うと仮定する。

ガンベル分布の確率密度関数は2つのパラメータの位置パラメータ μ (> 0) と尺度パラメータ θ (> 0) を用い以下の式で表すことができる（Croissant, 2020）。

$$f(z) = \frac{1}{\theta} e^{-\frac{z-\mu}{\theta}} e^{-e^{-\frac{z-\mu}{\theta}}} \quad (3)$$

累積分布関数は以下の式で表される。

$$P(z < t) = F(t) = \int_{-\infty}^t \frac{1}{\theta} e^{-\frac{z-\mu}{\theta}} e^{-e^{-\frac{z-\mu}{\theta}}} dz = e^{-e^{-\frac{t-\mu}{\theta}}} \quad (4)$$

ガンベル分布の期待値は $E(z) = \mu + \theta\gamma$ (ただし γ はオイラー一定数 $\gamma = 0.577\dots$) である。また、分散 $V(z)$ は $V(z) = \frac{\pi^2}{6}\theta^2$ となる。

ε_i ($i = 1, 2, 3$) が $\mu = 0, \theta = 1$ の独立した同一のガンベル分布 (第一種極値分布) に従い、回答者が ε_i の実現値を既知として、効用を最大化するように選択肢を選ぶことで、 ε_i の実現値を観測できない外部観測者にとって選択肢 i の選択確率は

$$P_i = \frac{e^{V_i}}{e^{V_1} + e^{V_2} + 1}, \quad i = 1, 2 \quad (5)$$

$$P_i = \frac{1}{e^{V_1} + e^{V_2} + 1}, \quad i = 3 \quad (6)$$

となる。

3. 選択型コンジョイント分析を実施するための準備

本節では、選択型コンジョイント分析を実施するための準備を、support.CEs パッケージを用いて行う手順を説明する。

3.1. support.CEs パッケージ

合崎 (2015) が紹介する support.CEs パッケージは選択型コンジョイント分析に必要な選択肢集合の作成が可能である。選択型コンジョイント分析のメリットは、回答者が少数のプロファイルの中から最も好ましいものを 1 つだけ選択するため、実際の購買場面に近いことや回答者への負担が少ないことがあげられる。一方、デメリットとしては、選択肢を構成する属性数が多くなると、設問数が多くなり回答者負担が大きくなることから、無回答やいい加減な回答が増え、結果の信頼性が担保できなくなる。これを避ける方法として、support.CEs パッケージではブロックを設定が可能となっている。全選択肢集合を分割し、ブロックごとに回答者を分けることで回答者の負担を減らすことができる。例えば全選択肢が 18 通りになる設問をブロック 1 とブロック 2 の 2 つのブロック (9 通りの選択肢集合) に設定することで、ブロック 1 とブロック 2 のそれぞれ別の回答者に回答してもらい、それらをまとめたデータを分析することが可能となる。このことによって、回答者負担を軽減し得られたデータの信頼性を担保することが可能となり実務的にも有用なものとなっていると考える。

以降では筆者が作成した仮想の有料老人ホームに対するアンケート調査において、合崎 (2015) を参考にしながら、support.CEs パッケージを用いて準備した手順を説明する。

3.2. 選択肢集合の作成から設問形式への変換まで

support.CEsパッケージを用い選択型コンジョイント分析に必要な設問を設計していくところから始める。要因と各水準は以下の様になっている。水準3を高水準、水準2を中水準、水準1を低水準と設定している。価格以外の要因の具体的イメージについては、鈴木・中山(2020, 表2)を参照してほしい。

- 価格 (Price) : 3水準 (1.0, 1.5, 2.0) 単位10万円
- 職員の配置について (Staff) : 3水準
- 健康管理について (HC) : 3水準
- サービスについて (Service) : 2水準
- 介護の質について (QC) : 3水準

まず、全選択肢集合を作成する(図1)。用いる関数はrotation.design()であり、それぞれの要因と水準を設定していく。3水準のものは中水準を、2水準のものは低水準を基準としている。例えば、職員の配置について(Staff)は中水準Staff2が基準である。サービスについて(Service)は低水準S0が基準である。

```
library(support.CEs)
#全選択肢集合の作成
NS = rotation.design(attribute.names=
                      list(Staff=c("Staff2","Staff1","Staff3"),
                           HC=c("HC2","HC1","HC3"),Service=c("S0","Service"),
                           QC=c("QC2","QC1","QC3"),Price=c(1.0,1.5,2.0)),
                      nalternatives=2,nblocks=2,row.renames=FALSE,randomize=TRUE,seed=123)
```

図1 全選択肢集合を作成するためのコード

nalternativesは選択型コンジョイント分析における1問あたりの選択肢の数を設定する。ここでは1問につき2つの選択肢を設定している。nblocksは先述したブロック数の設定である。注意すべきことは全質問数で割り切れる整数に設定する必要があることである。この関数内では無作為で選択肢の組み合わせを作るために乱数を用いているが、その結果に再現性を持たせるために引数seedを用いる。ここではseed=123としている。以上の関数と引数による結果をNSに格納する(図2)。

NS

Choice sets:

alternative 1 in each choice set

	BLOCK	QES	ALT	Staff	HC	Service	QC	Price
6	1	1	1	Staff1	HC3	S0	QC1	1
12	1	2	1	Staff3	HC1	Service	QC2	1.5
15	1	3	1	Staff3	HC3	S0	QC3	2
14	1	4	1	Staff2	HC3	Service	QC3	1.5
11	1	5	1	Staff2	HC2	S0	QC2	1
7	1	6	1	Staff2	HC3	Service	QC1	2
5	1	7	1	Staff2	HC2	S0	QC1	1.5
9	1	8	1	Staff1	HC1	S0	QC1	1.5
1	1	9	1	Staff1	HC3	Service	QC2	1.5
16	2	1	1	Staff3	HC1	Service	QC1	1
18	2	2	1	Staff3	HC2	S0	QC3	1.5
3	2	3	1	Staff2	HC1	S0	QC2	2
4	2	4	1	Staff2	HC1	Service	QC3	1
13	2	5	1	Staff3	HC2	Service	QC1	2
10	2	6	1	Staff1	HC2	Service	QC2	2
17	2	7	1	Staff3	HC3	S0	QC2	1
8	2	8	1	Staff1	HC1	S0	QC3	2
2	2	9	1	Staff1	HC2	Service	QC3	1

alternative 2 in each choice set

	BLOCK	QES	ALT	Staff	HC	Service	QC	Price
18	1	1	2	Staff2	HC1	Service	QC2	2
8	1	2	2	Staff3	HC3	Service	QC2	1
1	1	3	2	Staff3	HC2	S0	QC1	2
17	1	4	2	Staff2	HC2	Service	QC1	1.5
10	1	5	2	Staff3	HC1	S0	QC1	1
11	1	6	2	Staff1	HC1	Service	QC1	1.5
16	1	7	2	Staff2	HC3	S0	QC3	1.5

(続く)

9	1	8	2	Staff3	HC3	Service	QC3	2
6	1	9	2	Staff3	HC2	Service	QC3	1.5
7	2	1	2	Staff1	HC2	S0	QC3	1
5	2	2	2	Staff1	HC1	Service	QC3	2
2	2	3	2	Staff3	HC1	S0	QC2	1.5
4	2	4	2	Staff1	HC3	S0	QC2	1.5
15	2	5	2	Staff2	HC2	Service	QC2	1
14	2	6	2	Staff1	HC2	S0	QC2	2
3	2	7	2	Staff1	HC3	Service	QC1	1
13	2	8	2	Staff2	HC1	S0	QC3	1
12	2	9	2	Staff2	HC3	S0	QC1	2

Candidate design:

	A	B	C	D	E
1	2	3	2	1	2
2	2	1	2	3	1
3	1	2	1	1	3
4	1	2	2	3	1
5	1	1	1	2	2
6	2	3	1	2	1
7	1	3	2	2	3
8	2	2	1	3	3
9	2	2	1	2	2
10	2	1	2	1	3
11	1	1	1	1	1
12	3	2	2	1	2
13	3	1	2	2	3
14	1	3	2	3	2
15	3	3	1	3	3
16	3	2	2	2	1
17	3	3	1	1	1
18	3	1	1	3	2

class=design, type= oa

(続く)

```

Design information:
number of blocks = 2
number of questions per block = 9
number of alternatives per choice set = 2
number of attributes per alternative = 5

```

図2 全選択肢集合の出力結果

alternative 1とalternative 2を組み合わせることで選択型コンジョイント分析が可能となる。つまり、1問目はalternative 1の6とalternative 2の18が提示されることとなる。Candidate designはalternative 1を生成するために使用した直交配列表である。

さらに、関数questionnaire() を利用することで、より分かりやすく質問を提示するための準備をすることが可能となる (図3)。

```
questionnaire(choice.experiment.design = NS)
```

```
Block 1
```

```
Question 1
```

```

      alt.1   alt.2
Staff  "Staff1" "Staff2"
HC     "HC3"   "HC1"
Service "S0"   "Service"
QC     "QC1"   "QC2"
Price  "1"     "2"

```

(中略)

```
Question 9
```

```

      alt.1   alt.2
Staff  "Staff1" "Staff3"
HC     "HC3"   "HC2"
Service "Service" "Service"
QC     "QC2"   "QC3"
Price  "1.5"   "1.5"

```

(続く)

Block 2

Question 1

	alt.1	alt.2
Staff	"Staff3"	"Staff1"
HC	"HC1"	"HC2"
Service	"Service"	"S0"
QC	"QC1"	"QC3"
Price	"1"	"1"

(中略)

Question 9

	alt.1	alt.2
Staff	"Staff1"	"Staff2"
HC	"HC2"	"HC3"
Service	"Service"	"S0"
QC	"QC3"	"QC1"
Price	"1"	"2"

図3 全選択肢集合の質問形式への変換のためのコードと出力

3.3. 計画行列の作成

最後に関数`make.design.matrix()`を用いて計画行列を作成する。関数内の引数は先に作成したNSのデータセットを割り当て、引数`optout`で選択外オプションの有無を指定する。この調査では「どちらも選ばない」も選択肢とするので、TRUEを割り当てている。次に、選択肢集合に含まれる属性を、質的属性と量的属性に分けて、それぞれ引数`categorical.attributes`と引数`continuous.attributes`に文字列ベクトルとして割り当てる。引数`unlabeled`は、選択肢に固有の名称を持つラベル型(FALSE)か選択肢に固有の名称を持たない非ラベル型(TRUE)かを設定する。作成するデータでは、設定した属性以外に固有の特徴がない非ラベル型の選択肢であることから、TRUEを割り当てている。以上の計画行列をオブジェクトDatasetに格納する(図4)。

#全選択肢集合の計画行列への変換

```
Dataset = make.design.matrix(choice.experiment.design = NS,
  optout = TRUE,
  categorical.attributes = c("Staff", "HC", "Service", "QC"),
  continuous.attributes = c("Price"), unlabeled = TRUE)
```

Dataset

	BLOCK	QES	ALT	ASC	Staff1	Staff3	HC1	HC3	Service	QC1	QC3	Price
1	1	1	1	1	1	0	0	1	0	1	0	1.0
2	1	1	2	1	0	0	1	0	1	0	0	2.0
3	1	1	3	0	0	0	0	0	0	0	0	0.0
(中略)												
25	1	9	1	1	1	0	0	1	1	0	0	1.5
26	1	9	2	1	0	1	0	0	1	0	1	1.5
27	1	9	3	0	0	0	0	0	0	0	0	0.0
28	2	1	1	1	0	1	1	0	1	1	0	1.0
29	2	1	2	1	1	0	0	0	0	0	1	1.0
30	2	1	3	0	0	0	0	0	0	0	0	0.0
(中略)												
52	2	9	1	1	1	0	0	0	1	0	1	1.0
53	2	9	2	1	0	0	0	1	0	1	0	2.0
54	2	9	3	0	0	0	0	0	0	0	0	0.0

図4 全選択肢集合の計画行列への変換のためのコードと出力

作成したDatasetにおいて、BLOCKでブロック1、2を区別する。QESは設問1から設問9を区別する。3行ごとに1回に提示される選択肢の情報が含まれている。ALTが3の行は「どちらも選ばない」という選択肢に対応している。このDatasetと同一の内容のデータを回答者数にあわせて複製する（行を増やす）ことで、分析に用いる説明変数データとすることができる。

次節では、人工データを用いて分析した結果を説明する。なお、人工データの作成方法は補論に掲載する。

4. 人工データを使った分析

mlogitパッケージを用いて多項ロジット・モデルによる推定を行う。用いるデータは3.3で作成した計画行列Datasetに基づき、補論において説明したように、効用関数のパラメータ（部分効用）の真値を設定して、回答結果を人工的に作成してまとめたDatasetAllである。

4.1. mlogitパッケージを用いた推定までの手順

選択型コンジョイント分析によるアンケートを実施したと想定した200人分の回答結果を用いる。なお、補論における人工データ作成での繰り返し回数は100だが、2つのブロックに分けたと想定しているため、回答者数にすると200人分となる。

mlogitパッケージで多項ロジット・モデルを分析するために関数mlogit()を用いる。ただし、関数mlogit()で分析するための型に成形する必要がある。成形のためにはmlogitパッケージを読み込んだ際に同時に読み込まれるdfidxパッケージに含まれる関数dfidx()を用いる（図5）。mlogitパッケージの古いバージョンでは、そのパッケージ内の関数mlogit.data()を使って、成形する必要があったが、確認した新しいバージョン（Version 1.1-1）のmlogitパッケージではこの関数は廃止予定（deprecated）とされているので注意が必要である。

関数dfidx()ではまず成形するデータセットDatasetAllを指定し、関数mlogit()で分析する際に用いるインデックスを引数idxで指定している。データセット全体に対する質問の通し番号を指すインデックスと各質問における選択肢を指すインデックスを指定する必要があるが、DatasetAllには対応する列がないため、NAとしている。引数choiceはどれを選択したかを示すインデックスを指定し、shapeはデータセットの型を指定する。今回のデータセット（DatasetAll）は1回の選択機会の情報が3行分使用するlong型である。これに対して1行に1回の選択機会の情報が含まれるものがwide型である。

```
library(mlogit)
Loading required package: dfidx
DatasetAll = dfidx(DatasetAll, idx = list(NA, "ALT"),
                  choice="RES", shape="long")
```

図5 関数mlogit()での推定前に必要なデータ変換のためのコード

この処理を行った後、関数mlogit()で分析することが可能となる（図6）。関数mlogit()内ではモデル式にあたる引数formulaと分析用データセットを指定している。変数RESは

TRUEもしくはFALSEからどの選択肢を選んだかを表すことになる。チルダ (~) の右側は選択行動に影響を及ぼした変数であり、最後に関数dfidx() で生成したデータセット DatasetAllを指定する。なお、チルダの直後に「-1」があるが、これは3つの選択肢全てに

```
ResultAll = mlogit(RES ~ -1+ASC + Staff1 + Staff3 + HC1+HC3+Service+QC1
                   +QC3 + Price ,DatasetAll)
```

```
print(summary(ResultAll),digits = 3)
```

(中略)

Frequencies of alternatives:choice

```
      1      2      3
0.403 0.390 0.207
```

nr method

4 iterations, 0h:0m:0s

g'(-H)^-1g = 0.00165

successive function values within tolerance limits

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z)	
ASC	0.8633	0.1717	5.03	4.9e-07	***
Staff1	-0.2557	0.0891	-2.87	0.0041	**
Staff3	0.5935	0.0953	6.23	4.7e-10	***
HC1	-0.4161	0.0963	-4.32	1.5e-05	***
HC3	0.3744	0.0877	4.27	2.0e-05	***
Service	0.4900	0.0829	5.91	3.4e-09	***
QC1	-0.4693	0.0950	-4.94	7.9e-07	***
QC3	0.5175	0.0964	5.37	8.1e-08	***
Price	-0.4286	0.0940	-4.56	5.1e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1710

図6 関数mlogit() で多項ロジット・モデルの推定のためのコードと出力

共通の定数項を設定しないように付けている。alternative 1とalternative 2にだけに表れる定数項（ベースラインの効用）のために別の変数ASCを設定している。

図6では分析結果をresultAllに格納しているため、その要約を関数summary()で返すことができる。表示桁数を調整するために、関数print()も併用している。出力結果をみると、Frequencies of alternativesで各選択肢の選択割合を表している。Coefficients以下が推定結果である。推定値は全て有意水準1%で0と統計的に有意な差がある。推定に使った多項ロジット・モデルの設定が正しいという前提の下で、200人分のデータを集めれば、定めた要因の効果が検証できることが分かる。ただし、設定したパラメータの真値からの乖離は、最も大きいStaff3で0.19である。McFadden (1984, p.1409) が示すように、多項ロジット・モデルの最尤法に基づく推定量は、一致性と漸近正規性を持つが、1回の推定では、パラメータの推定値が真値から乖離する程度の判断ができない。

4.2. 回答者数を変更した場合の比較

ここまでの結果を踏まえて、人工データを繰り返し作成して、推定結果を検証する。回答者数の想定を2000人、400人、200人、100人とする。補論のコードではHを1000、200、100、50と設定することに対応する。それぞれの設定で、ガンベル分布から発生させる乱数を変えた上で、1000回繰り返し人工データを作成して推定した結果の要約を表1に掲載する。この表にはパラメータの真値と、真値からの乖離を根平均二乗誤差（Root Mean Square Error、以下RMSE）で測った値も掲載している。

どの想定回答者数においても推定値の平均と真値との乖離は小さい。ASCで0.009未満、その他の変数では最大で0.011（想定回答者数100人におけるQC1）である。最尤推定量では一般に不偏性は保証されないが、今回の人工データ・シミュレーションでは、想定回答者数100人としても、推定値の平均と真値に大きな乖離はない。

ただし、個別の推定では、パラメータの真値との乖離の程度は、想定回答者数によって変化する。そのことは、RMSEや分位点の値に表れている。RMSEはASCがその他の変数よりも大きくなる傾向にある。想定回答者数を2000人とする、RMSEは0.055であるが、想定回答者数を100人とする、0.263となる。ASC以外では、想定回答者数100人でも0.15未満である。想定回答者数200人とする、ASC以外では、0.10未満である。分位点でみると、25%分位点から75%分位点がASC以外では真値からの乖離が±0.70の範囲に収まっている。

表1の結果を踏まえると、5つの要因（4つは3水準、1つは2水準）を設定し、価格の要因を間隔尺度で設定する調査において、1人あたり9問の設問に答えてもらう調査であれば、回答者の選好が同質であるという前提の下で、少なくとも200人の回答者を確保できれば、実態を反映する妥当な結果が得られることが示唆された。

表1 回答者数を変更した場合の推定結果

回答者数		ASC	Staff1	Staff3	HC1	HC3	Service	QC1	QC3	Price	
	真値	1.0	-0.4	0.4	-0.4	0.4	0.4	-0.4	0.4	-0.4	
2000	平均	1.003	-0.399	0.400	-0.400	0.400	0.399	-0.401	0.399	-0.401	
	根平均二乗誤差	0.055	0.028	0.030	0.031	0.027	0.025	0.029	0.031	0.029	
	0%	0.827	-0.480	0.311	-0.488	0.311	0.322	-0.496	0.290	-0.502	
	分	25%	0.965	-0.418	0.380	-0.421	0.382	0.382	-0.421	0.379	-0.422
	位	50%	1.002	-0.399	0.400	-0.400	0.399	0.400	-0.400	0.400	-0.401
	点	75%	1.041	-0.380	0.421	-0.377	0.418	0.417	-0.381	0.419	-0.380
	100%	1.182	-0.300	0.480	-0.306	0.485	0.468	-0.315	0.490	-0.312	
400	平均	1.003	-0.403	0.398	-0.399	0.402	0.399	-0.403	0.401	-0.400	
	根平均二乗誤差	0.125	0.064	0.068	0.071	0.064	0.059	0.068	0.069	0.066	
	0%	0.591	-0.590	0.179	-0.628	0.191	0.203	-0.604	0.197	-0.600	
	分	25%	0.918	-0.448	0.353	-0.445	0.359	0.360	-0.448	0.354	-0.445
	位	50%	1.003	-0.405	0.398	-0.400	0.399	0.400	-0.405	0.402	-0.400
	点	75%	1.087	-0.357	0.444	-0.350	0.445	0.439	-0.358	0.447	-0.357
	100%	1.389	-0.199	0.606	-0.139	0.655	0.570	-0.202	0.616	-0.177	
200	平均	1.003	-0.404	0.400	-0.402	0.401	0.401	-0.406	0.400	-0.399	
	根平均二乗誤差	0.179	0.090	0.096	0.098	0.086	0.083	0.096	0.101	0.094	
	0%	0.496	-0.712	0.087	-0.699	0.141	0.135	-0.755	0.110	-0.675	
	分	25%	0.875	-0.464	0.330	-0.470	0.342	0.344	-0.472	0.333	-0.463
	位	50%	1.002	-0.404	0.402	-0.404	0.401	0.400	-0.401	0.400	-0.396
	点	75%	1.123	-0.348	0.461	-0.332	0.457	0.457	-0.341	0.464	-0.336
	100%	1.650	-0.118	0.720	-0.090	0.723	0.631	-0.095	0.729	-0.080	
100	平均	1.009	-0.402	0.403	-0.406	0.402	0.404	-0.411	0.395	-0.403	
	根平均二乗誤差	0.263	0.130	0.132	0.142	0.123	0.115	0.136	0.141	0.136	
	0%	0.212	-0.801	-0.055	-0.803	-0.014	0.057	-0.913	-0.025	-0.962	
	分	25%	0.824	-0.488	0.316	-0.498	0.317	0.325	-0.502	0.302	-0.495
	位	50%	1.000	-0.403	0.402	-0.399	0.398	0.405	-0.408	0.395	-0.404
	点	75%	1.193	-0.316	0.490	-0.312	0.483	0.483	-0.321	0.494	-0.310
	100%	1.781	-0.035	0.871	0.110	0.786	0.770	0.048	0.855	0.003	

5. おわりに

本稿では、Rパッケージのsupport.CEsとmlogitを用いた、多項ロジット・モデルによる分析手順を説明した。特に、マーケティング・サイエンスの分野で消費者の選好構造を分析するために利用されている選択型コンジョイント分析を念頭に置いた。本文中の図や補論に掲載したRコードを、本稿の結果の再現や、独自に選択型コンジョイント分析を実施するための参考情報としてほしい。なお、mlogitパッケージは、多項ロジット・モデルのみではなく入れ子ロジット・モデル (nested logit model) や混合ロジット・モデル (mixed logit model) など、様々な離散選択モデルの推定のために用いることができる。Croissant (2020) を参考にして、離散選択モデルを用いた分析に活用してほしい。

補論

第3節で作成した計画行列Datasetを使って、第4節の分析に用いる人工データを作成するためのコードを掲載する。

```
#ガンベル分布からの乱数を発生させる関数を持つパッケージevdの読み込み
library(evd)
#係数 真の値の設定
B = c(1,-0.4,0.4,-0.4,0.4,0.4,-0.4,0.4,-0.4)
names(B) =colnames(Dataset)[4:12]
B
      ASC  Staff1  Staff3      HC1      HC3 Service      QC1      QC3  Price
      1.0   -0.4    0.4   -0.4    0.4    0.4   -0.4    0.4   -0.4
#回答者数の設定 Hの値を変更することで回答者数を変えた分析が可能になる。
#(2つのブロックを設定しているため)
#200人が回答していると解釈する。
H = 100

#説明変数に該当する列の抽出
X = Dataset[,4:12]

#保存先の設定
DatasetAll = NULL

for(h in 1:H){
  #確定効用
  V = as.matrix(X) %%% B
  Dataset$V = V

  #1行に1回分の選択肢の確定効用を整理
  V = matrix(V,byrow=TRUE,ncol=3)

  #確率効用
  set.seed(h)
```

```
e = rgumbel(18*3,loc=0,scale=1)
Dataset$e = e

#1行に1回分の選択枝の確率効用を整理
e = matrix(e,byrow=TRUE,ncol=3)

#全体効用の計算
U = V + e
U = data.frame(U)
colnames(U) = c("U1","U2","U3")

#全体効用の最大値を抽出
maxU = apply(U,1,max)
U$maxU = maxU

#全体効用が最大となる選択枝を識別
U$RES1 = ifelse(U$U1==U$maxU,1,0)
U$RES2 = ifelse(U$U2==U$maxU,1,0)
U$RES3 = ifelse(U$U3==U$maxU,1,0)

#Datasetへの保存
tempU = U[,c("U1","U2","U3")]
tempU = t(tempU)
dim(tempU) = c(54,1)
Dataset$U = tempU
tempRES = U[,c("RES1","RES2","RES3")]
RES = t(tempRES)
dim(RES) = c(54,1)
Dataset$RES = RES

#DatasetAllに追加
DatasetAll = rbind(DatasetAll,Dataset)
}
```


謝辞

指導教員である中山雄司教授に本稿の内容、特に関連する参考文献や補論での人工データ作成に関して助言を得たことを感謝します。

参考文献

- [1] Bedilu, Demissie., Hussien, H. Komicha., Adem Kedir., Sisay, Asef. (2018). Camel Milk Marketing Channel Choices for Enhancing Competitiveness in Eastern Ethiopia: Multinomial Logit Approach. *Global Journal of Management and Business*. 18(5), 31-39.
- [2] Croissant, Yves. (2020). mlogit: Random Utility Models in R. *Journal of Statistical Software*, 95(11), 1-41.
- [3] Jesper, N, Wulff. (2015). Interpreting Results From the Multinomial Logit Model: Demonstrated by Foreign Market Entry. *Organizational Research Methods*, 18(2), 300-325.
- [4] Judit, Guimera, Busquets., Eduardo, Alonso., Antony, D. Evans. (2017). Air itinerary shares estimation using multinomial logit models. *Transportation Planning and Technology*. 41(1), 1-14.
- [5] Louviere, J. J., Flynn, T. N. and Carson, R. T. (2010). Discrete Choice Experiments Are Not Conjoint Analysis, *Journal of Choice Modelling*, 3(3), 57-72.
- [6] McFadden, D. L. (1984). Econometric Analysis of Qualitative Response Models. *Handbook of Econometrics*, Volume 2, 1395-1457.
- [7] Quan, Hoang, Vuong., Thi, Hanh,Vu., Quang, Hung, Doan.,Manh, Tung, Hoe. (2019). Determinants of Vietnamese footwear exporting firms' market selection: A multinomial logistic analysis of panel data. *Heliyon*, 5, e02582
- [8] Tarek Abdallah., Gustavo Vulcano. (2019). Demand Estimation under the Multinomial Logit Model from Sales Transaction Data. *Manufacturing and Service Operations Management*, 1-35.
- [9] Yun, Hui, Lin., Yuan, Wang., Dongdong, He., Loo, Hay, Lee. (2020). Last-mile delivery: Optimal locker location under multinomial logit choice model. *Transportation Research Part E Logistics and Transportation Review*, 142, 102059.
- [10] 合崎英男. (2015). Rパッケージsupport.CEsとsurvivalを利用した離散選択実験の実施手順. 北海道大学農経論叢, 70, 1-16.
- [11] 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, 151-173.
- [12] 荒木孝治. (2007). フリーソフトウェアのソーシャルライフ —データ解析環境Rの拡張パッケージ—. 関西大学商学論集, 51(6), 1-12.
- [13] 池尾恭一・青木幸弘・南知恵子・井上哲浩. (2010). 『マーケティング』. 有斐閣.
- [14] 伊藤宏樹. (2016). 状態空間モデルの実行方法と実行環境の比較. 日本生態学会誌, 66, 361-374.

- [15] 鈴木真, 中山雄司. (2020). 有料老人ホームに対する選好構造分析：年代別分析. School of Economics, Osaka Prefecture University Discussion Paper New Series, 2020(3), 1-33.
- [16] 土田尚弘. (2010). マーケティング・サイエンスにおける離散選択モデルの展望. 経営と制度, 8, 63-91.
- [17] 中谷朋昭. (2010). データ解析環境Rにおける多変量GARCHモデルの取り扱いについて — ccgarchパッケージの利用一. 北海道大学農経論叢, 65, 61-68.
- [18] 森大輔. (2017a). 質的比較分析(QCA)のソフトの使用方法—fs/QCAとRのQCA/SetMethodsパッケージ(1). 熊本法学, 140, 250-209.
- [19] 森大輔. (2017b). 質的比較分析(QCA)のソフトの使用方法—fs/QCAとRのQCA/SetMethodsパッケージ(2). 熊本法学, 141, 388-348.