

RcmdrPlugin.catdap (Ver. 1.1.2) マニュアル

1. はじめに

本パッケージは Rcmdr のプラグインであり、カテゴリカルデータ解析のための R パッケージ catdap の解析機能の一部が GUI で利用可能である。catdap パッケージの解析関数としては catdap1() と catdap2() があるが、このプラグインでは catdap2() の解析すなわち一つの着目した項目（目的変数）を固定し最も適切に説明する他の項目（説明変数）の組合せの探索を行うことができる。catdap はカテゴリカルデータだけでなく連続値データの場合も扱えるが、連続値を離散化するために後に説明する観測精度とプール形式といった情報が変数ごとに必要になる。

はじめに R は既にインストールされていることを前提に、準備としての Rcmdr のインストール、続いて catdap と RcmdrPlugin.catdap のインストールについて説明する。R のインストール方法については、RjpWiki, CRAN ホームページ を参照すること。インストール手順は、2021 年 11 月に公開したマニュアルを基に Windows 11 (R-4.6.0) と Ubuntu 24.04 LTS (R-4.5.1) で Rcmdr (バージョン 2.9.5) をインストールして再確認した手順である。

続く解析の進め方は、解析したいデータの読み込み、観測精度とプール形式を記述するメタ情報データセットの作成、解析の実行の順に説明する。最後に、実行例として catdap パッケージのヘルプの例題と同じデータを使用して実行した場合を示す。

2. インストール手順

2.1 Rcmdr のインストール

2.1.1 Windows の場合

```
RGui を起動して R Console のウィンドウで  
> install.packages("Rcmdr", dependencies = TRUE)
```

依存対象 (dependency) 'listenv', 'parallelly', ... もインストール

2.1.2 Linux の場合

```
ターミナルから  
$ sudo apt install cmake  
$ sudo apt install tk-table  
ターミナルから R を起動
```

```
$ R
> install.packages("RcmdrMisc", dependencies = TRUE)
> install.packages("Rcmdr", dependencies = TRUE)
```

2.2 catdap と RcmdrPlugin.catdap のインストール

(注) インストールは catdap, RcmdrPlugin.catdap の順に行うこと.

また, catdap のインストールの前に RColorBrewer パッケージもインストールしておく必要がある.

2.2.1 Windows の場合

```
R Console のメニュー「パッケージ」
-> 「パッケージのインストール...」
-> ファイル選択画面で RColorBrewer を選択
R Console のメニュー「パッケージ」
-> 「ローカルにある zip ファイルからのパッケージのインストール...」
-> ファイル選択画面でダウンロードしたファイル
    catdap_1.3.7-1.zip をクリックして[開く]
```

同様に

```
-> ファイル選択画面でダウンロードしたファイル
    RcmdrPlugin.catdap_1.1.2.zip をクリックして[開く]
```

2.2.2 Linux の場合

```
ターミナルからダウンロードしたフォルダーで R を起動
$ R
> install.packages("RColorBrewer")
> install.packages("catdap_1.3.7.tar.gz", repos=NULL, type="source")
> install.packages("RcmdrPlugin.catdap_1.1.2.tar.gz", repos=NULL,
                    type="source")
```

2.3 RcmdrPlugin.catdap のロード

2.3.1 Windows の場合

```
-> R Console のメニュー「パッケージ」
-> 「パッケージの読み込み...」
-> パッケージ選択画面で RcmdrPlugin.catdap をクリックして「OK」
```

2.3.2 Linux の場合

```
ターミナルから R を起動  
$ R  
> library(RcmdrPlugin.catdap)
```

ロード後, R コマンダーのウィンドウメニューに catdap が表示され, サブメニューは以下のようにになっている.

```
カテゴリカルデータ解析...  
メタ情報...  
▶ 新しいデータセット...  
▶ データセットのロード...
```

3. 解析

3.1 データセットの読み込み

データの操作は R コマンダーのメニュー「データ」から行うが, ここでは解析したいデータが既に csv ファイルとしてある場合と パッケージ内のデータセットの場合についてのみ説明する.

・ csv ファイル(カンマ区切り)の読み込み

```
メニュー「データ」  
->「データのインポート」  
->「テキストファイルまたはクリックボード, URL から...」  
  * 一行目に変数名の場合  
    ->ファイル内に変数名あり をチェック  
  * 変数名がない場合  
    ->ファイル内に変数名あり のチェックをはずす  
  
->フィールドの区切り記号は ◎ カンマ[,] を指定して [✓OK]  
->ファイルの選択画面で既存の csv ファイルを指定し [開く]
```

変数名がない csv ファイルを読み込んだ場合, データセットの変数名は自動的に V1, V2, V3, ... となる.

・ パッケージ内データセットの読み込み

```
メニュー「データ」  
->「パッケージ内のデータ」  
->「アタッチされたパッケージからデータセットを読み込む...」
```

-> パッケージを選択後 データセットを選択して [✓OK]

3.2 メタ情報データセット

次に読み込んだデータセットのメタ情報データセットを作成する。ここでのメタ情報とは各変数に対する観測精度 (accuracy) と プール形式 (pooling) であり、連続変数を離散化するための情報である。

観測精度とは連続変数を離散化するための最小限の幅を示す値であり、データに応じて '1', '0.1', '0.01' などの値を変数毎に指定する。ただし、カテゴリカル変数に対しては '0' を指定する。プール形式は連続変数に対してのみ有効な値であり、'0' または '1' の整数で指定する。

0: 等間隔プーリング (top-down 形式)

1: 不等間隔プーリング (bottom-up 形式)

等間隔プーリングは、一定巾の区間に区切って離散化します。ただし、両端の区間は一定巾より短くなる場合がある。不等間隔プーリングは、短い一定巾の区間に区切った後、必要に応じて (AIC の値でより最適と判断された場合) 隣り合う区間を合併するものである。いずれの場合も、 m 区間の境界値を b_1, b_2, \dots, b_{m+1} とした時の i 番目のカテゴリの値の範囲 r_i ($1 \leq i \leq m$) は

$$r_i = [b_i, b_{i+1}) \quad (1 \leq i < m)$$

$$r_m = [b_m, b_{m+1}]$$

となる。

解析したいデータファイルを読み込んだあと、R コマンダーの **データセット:** 欄に解析したいデータセット名がアクティブデータセットとして表示されていることを確認して以下のように作成する。

```
メニュー「catdap」
-> 「メタ情報...」
-> 「新しいデータセット...」
-> 「メタデータセットの作成」ウィンドウが開くので
    データセット名: (デフォルトで <データセット名>.meta)を確認
    または 必要なら変更して [✓OK]

-> 「データエディタ:」ウィンドウが開くので
    データに合った accuracy と pooling に値を設定して [✓OK]
    「メタデータセットを保存しますか?」というダイアログが開くので必要に
    応じて保存
```

メタデータセットを保存した場合は、次回の RcmdrPlugin.catdap をロード後

```
メニュー「catdap」  
->「メタ情報...」  
->「データセットのロード...」
```

を選択して利用できる。

3.3 解析の実行

R コマンダーの **データセット:** 欄に解析したいデータセット名がアクティブデータセットとして表示されていることを確認して、

```
メニュー「catdap」  
->「カテゴリカルデータ解析...」をクリックして解析に進む。
```

(1) データタブ

解析に必要な基本的情報を入力するために使用。

・目的変数を選択

着目したい変数を目的変数として変数リストから選択。

・メタデータセット名:

デフォルトで<データセット名>.meta が表示される。必要に応じて変更する。

(2) オプションタブ

データに欠測値がある場合、変数の数が非常に多い場合、出力を制御したい場合に使用。通常はすべてデフォルトのままで実行できる。

・欠測値の処理

☐ 欠測値処理 欠測値マーク:

欠測値がある場合には、☐ にチェックを入れ欠測値マークに正の値を指定する。
例えば "1000" を指定した場合、値 x のデータは $1000 \leq x < 2000$ ならばタイプ1の欠測値、 $2000 \leq x < 3000$ ならばタイプ2の欠測値、 $1000k \leq x < 1000(k+1)$ ならば k 番目のタイプの欠測値というように扱われる。

・分割表の視覚化

- ☐ モザイクプロット : 水平・垂直交互に分割
- ☒ モザイクプロット(分割は水平方向のみ) : (デフォルト)
- ☐ なし

・ AIC リストのプリント出力レベル

- ◎ すべて出力 : (デフォルト)
- 一部省略

AIC リスト(昇順)はデフォルトで、

```
<< List of single explanatory variables (arranged in ascending order of AIC) >>
<< AIC's of the models with k explanatory variables (k=1,2,...) >>
<< Summary of subsets of explanatory variables >>
```

の 3 種が出力される。一部省略を指定すると 2 番目のリストは出力せず 3 番目の Summary のリストも上位 30 位のみが出力される。

・ 作業領域制御パラメータ

多重クロス表作成のための説明変数の候補数 : デフォルト (変数の数)-1

pa1	: デフォルト 1
pa2	: デフォルト 4
pa3	: デフォルト 10

変数の数が非常に多い場合、作業領域の制御が必要になる場合がある。多重クロス表作成のための説明変数の候補数では、説明変数一つの場合の AIC リストのうち上位何位までを候補とみなすかを指定する。パラメータ pa1, pa2, pa3 は、作業領域不足のためエラーメッセージが出力された場合にそのメッセージにしたがって変更すること。

多重クロス表作成のための説明変数の候補数の値を小さくすれば作業領域不足が解消される場合もあるが、最適モデルが得られないこともある。解析したいデータに応じて候補数の値を小さくするか、パラメータ pa1, pa2, pa3 の値を大きくするか選択すること。オプションタブの設定例は、実行例 4.3 を参照。

[✓OK] をクリックすると解析が実行されるが、目的変数に連続型変数を選択した場合のみ目的変数の区間の数 を指定するダイアログが表示される。連続型目的変数はメタ情報のプーリングの形式にかかわらず等間隔で離散化され、区間の数(カテゴリ数)を指定する必要がある。指定後、再び [✓OK] をクリックする。

4 実行例

4.1 R パッケージ catdap 内データセット HealthData の解析 (8 変数, レコード長 52)

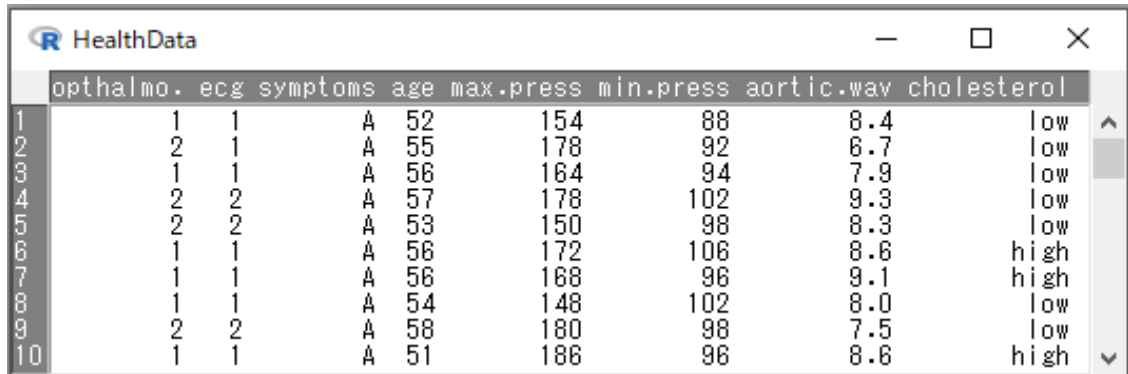
メニュー「データ」

-> 「パッケージ内のデータ」

-> 「アタッチされたパッケージからデータセットを読み込む...」

-> パッケージ catdap のデータセット HealthData を選択して [✓OK]

この結果、データセット: の欄は [HealthData](#) となり [データセットを表示] をクリックすると以下のようなデータが表示される。



	opthalgo.	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
1	1	1	A	52	154	88	8.4	low
2	2	1	A	55	178	92	6.7	low
3	1	1	A	56	164	94	7.9	low
4	2	2	A	57	178	102	9.3	low
5	2	2	A	53	150	98	8.3	low
6	1	1	A	56	172	106	8.6	high
7	1	1	A	56	168	96	9.1	high
8	1	1	A	54	148	102	8.0	low
9	2	2	A	58	180	98	7.5	low
10	1	1	A	51	186	96	8.6	high

メニュー「catdap」

-> 「メタ情報...」

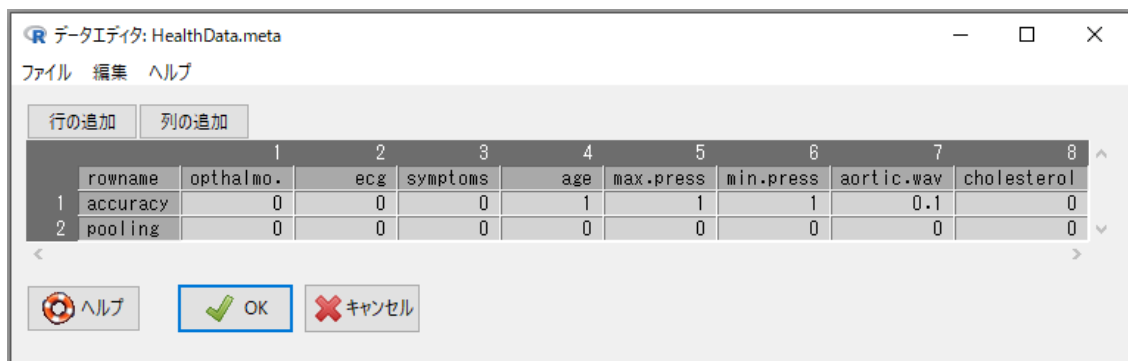
-> 「新しいデータセット...」

-> 「メタデータセットの作成」ウィンドウが開くので

データセット名: HealthData.meta (デフォルト)を確認

または 必要ならば変更して [✓OK]

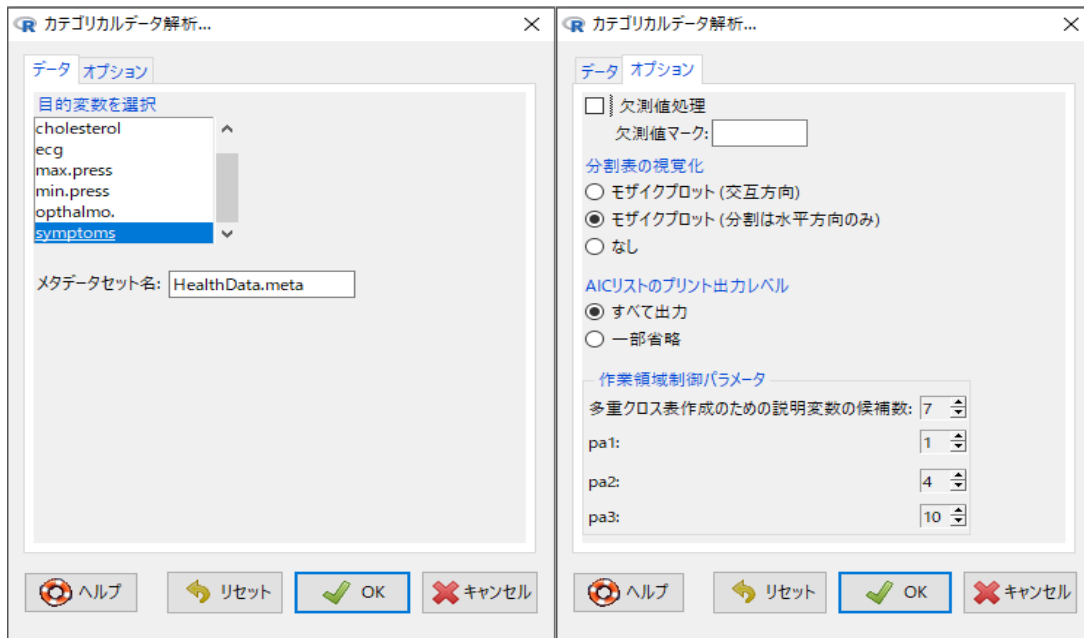
「データエディタ: HealthData.meta」ウィンドウが開くので、データに合わせ以下のように設定。



	rowname	1	2	3	4	5	6	7	8
		opthalgo.	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
1	accuracy	0	0	0	1	1	1	0.1	0
2	pooling	0	0	0	0	0	0	0	0

[✓OK]をクリックすると「メタデータセットを保存しますか?」というダイアログが開くので必要に応じて保存する。データセット: [HealthData.meta](#) となっているところをクリックして、[HealthData](#) をアクティブにする。

メニュー「catdap」
 ->「カテゴリカルデータ解析...」
 を選択して解析を進める。
 目的変数に symptoms を選択した以外はデフォルトのまま。



出力 (途中省略)

> HealthData.meta

	optharmo.	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
accuracy	0	0	0	1	1	1	0.1	0
pooling	0	0	0	0	0	0	0.0	0

> acc <- c(0,0,0,1,1,1,0.1,0)

> pooling <- c(2,2,2,0,0,0,0,2)

> Catdap2(HealthData, pool=pooling, response.name="symptoms", accuracy=acc,
 + nvar=8, pa1=1, pa2=4, pa3=10, print.level=0, plot=1)

:

<< Summary of subsets of explanatory variables >>

Response variable : symptoms

Explanatory variables	Number of Categories of exp. var.	A I C	Difference of A I C	Weight

1	max.press aortic.wav	4	-11.19	0.00	1.00
2	aortic.wav min.press	4	-10.90	0.29	0.87
3	max.press aortic.wav min.press	8	-8.96	2.23	0.33
4	max.press cholesterol	4	-8.76	2.43	0.30
5	max.press aortic.wav cholesterol	8	-7.29	3.90	0.14
6	max.press	2	-7.10	4.09	0.13
7	aortic.wav	2	-6.25	4.95	0.08
8	max.press aortic.wav ecg	8	-6.11	5.09	0.08
9	max.press ecg	4	-6.06	5.13	0.08
10	min.press	2	-4.46	6.73	0.03
	:				

<< Contingency table constructed by the best subset of explanatory variables >>

X(1) : symptoms

X(2) : max.press

X(3) : aortic.wav

X X response variable X(1)

(2) (3) 1 2 Total

1 1 4 (22.2) 14 (77.8) 18 (100.0)

1 2 10 (62.5) 6 (37.5) 16 (100.0)

2 1 6 (66.7) 3 (33.3) 9 (100.0)

2 2 9 (100.0) 0 (0.0) 9 (100.0)

Total 29 (55.8) 23 (44.2) 52 (100.0)

<Note>

X(1) : symptoms

category	variable value
----------	----------------

1	A
---	---

2	B
---	---

X(2) : max.press

category	value range
----------	-------------

1	9.80000e+01 - 1.65500e+02
---	---------------------------

2	1.65500e+02 - 2.16000e+02
---	---------------------------

X(3) : aortic.wav

category	value range
----------	-------------

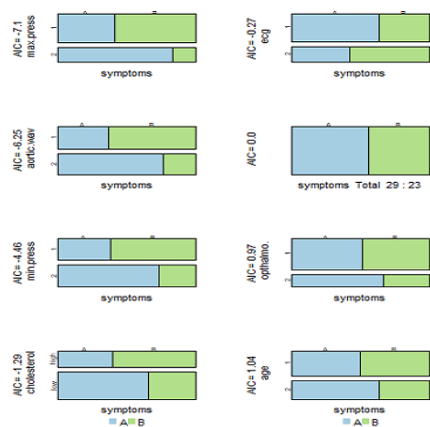
1	6.30000e+00 - 8.25000e+00
---	---------------------------

2	8.25000e+00 - 1.02000e+01
---	---------------------------

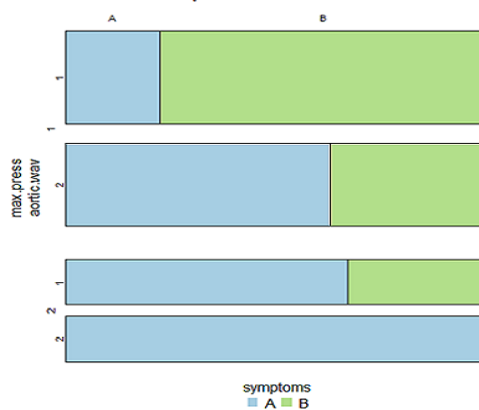
AIC = -11.19

base AIC = 73.39

Single Explanatory Models in ascending order of AIC



Minimum AIC Model of the Response Variable Distribution



* 補足

R スクリプトや出力には、カテゴリカル変数に対して pooling=2 が設定される。

RcmdrPlugin.catdap では accuracy=0 を優先し自動的に設定されるので、ユーザーは特に設定する必要はない。

4.2 R パッケージ datasets 内のデータセット iris の解析

メニュー「データ」

-> 「パッケージ内のデータ」

-> 「アタッチされたパッケージからデータセットを読み込む...」

-> パッケージ datasets のデータセット iris を選択して [✓OK]

この結果、データセット: の欄は [iris](#) となり [データセットを表示] をクリックすると以下のようなデータが表示される。

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa

4.1 の場合と同様にメタデータセット iris.meta を作成する。

データエディタ: iris.meta

ファイル編集ヘルプ

行の追加列の追加

		1	2	3	4	5
1	rowname	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	accuracy	0.1	0.1	0.1	0.1	0
2	pooling	0	0	0	0	0

ヘルプ

OK

キャンセル

データセット: [iris.meta](#) となっているところをクリックして, [iris](#) をアクティブにする。

メニュー「catdap」

-> 「カテゴリカルデータ解析...」

を選択した場合の設定例と実行結果を以下に示す。目的変数に Petal.Width 選択する以外はデフォルトのまま。

データ オプション

目的変数を選択

Petal.Length
Petal.Width
Sepal.Length
Sepal.Width
Species

メタデータセット名: iris.meta

ヘルプ リセット OK キャンセル

データ オプション

☐ 欠測値処理
欠測値マーク:

分割表の視覚化

☐ モザイクプロット (交互方向)
☒ モザイクプロット (分割は水平方向のみ)
☐ なし

AICリストのプリント出力レベル

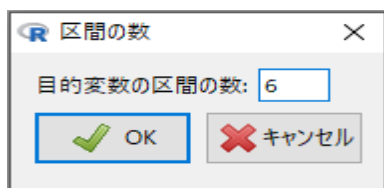
☒ すべて出力
☐ 一部省略

作業領域制御パラメータ

多重クロス表作成のための説明変数の候補数: 4
pa1: 1
pa2: 4
pa3: 10

ヘルプ リセット OK キャンセル

[✓OK]をクリックすると Petal.Width は連続型変数なので区間の数を指定するためのダイアログが開くので、区間の数を入力して再び[✓OK]をクリックする。



出力 (途中省略)

```
> iris.meta
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
accuracy	0.1	0.1	0.1	0.1	0
pooling	0.0	0.0	0.0	0.0	0

```
> acc <- c(0.1,0.1,0.1,0.1,0)
```

```
> pooling <- c(0,0,0,-6,2)
```

```
> Catdap2(iris, pool = pooling, response.name = "Petal.Width", accuracy = acc,
+ nvar=5, pa1 = 1, pa2 = 4, pa3 = 10, print.level = 0, plot = 1)
:
```

<< Summary of subsets of explanatory variables >>

Response variable : Petal.Width

	Explanatory variables	Number of Categories of exp. var.	A I C	Difference of AIC	Weight
1	Species	3	-243.00	0.00	1.00
2	Species	9	-224.98	18.02	0.00
	Petal.Length				
3	Petal.Length	3	-224.37	18.63	0.00
4	Species	6	-220.77	22.23	0.00
	Sepal.Width				
5	Species	6	-210.53	32.47	0.00
	Sepal.Length				
6	Petal.Length	6	-202.23	40.77	0.00

	Sepal.Length				
7	Species	18	-190.39	52.62	0.00
	Petal.Length				
	Sepal.Length				
8	Species	18	-178.68	64.32	0.00
	Petal.Length				
	Sepal.Width				
9	Sepal.Length	2	-109.37	133.63	0.00
10	Sepal.Width	2	-59.28	183.72	0.00
11	- - -	0	0.00	243.00	0.00

<< Contingency table constructed by the best subset of explanatory variables >>

X(1) : Petal.Width

X(2) : Species

X		response variable X(1)					
(2)	1	2	3	4	5	6	Total
1	48 (96.0)	2 (4.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	50 (100.0)
2	0 (0.0)	0 (0.0)	15 (30.0)	33 (66.0)	2 (4.0)	0 (0.0)	50 (100.0)
3	0 (0.0)	0 (0.0)	0 (0.0)	4 (8.0)	23 (46.0)	23 (46.0)	50 (100.0)
Total							
	48 (32.0)	2 (1.3)	15 (10.0)	37 (24.7)	25 (16.7)	23 (15.3)	150 (100.0)

<Note>

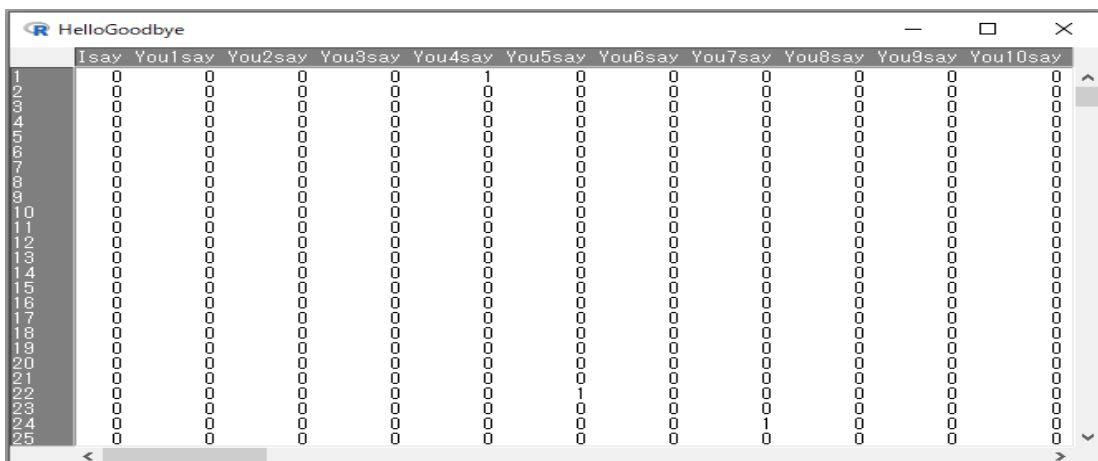
X(1) : Petal.Width

category	value range
1	1.00000e-01 - 5.00000e-01
2	5.00000e-01 - 9.00000e-01
3	9.00000e-01 - 1.30000e+00
4	1.30000e+00 - 1.70000e+00
5	1.70000e+00 - 2.10000e+00
6	2.10000e+00 - 2.50000e+00

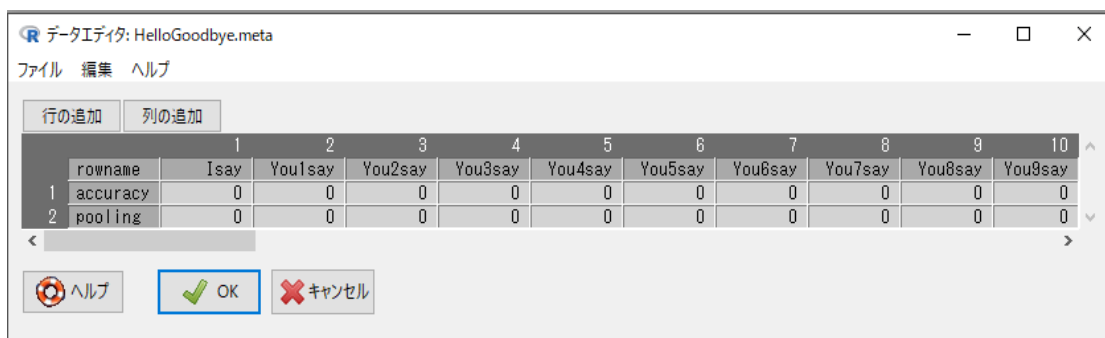
X(2) : Species

category	variable value
1	setosa

base AIC = 210.27



4.1 の場合と同様にメタデータセット HelloGoodbye.meta を作成する。



	1	2	3	4	5	6	7	8	9	10
1	accuracy	0	0	0	0	0	0	0	0	0
2	pooling	0	0	0	0	0	0	0	0	0

データセット: [HelloGoodbye.meta](#) となっているところをクリックして, [HelloGoodbye](#) をアクティブにする。

メニュー「catdap」

-> 「カテゴリカルデータ解析...」

を選択後, 目的変数に Isay を選択して解析を進める。

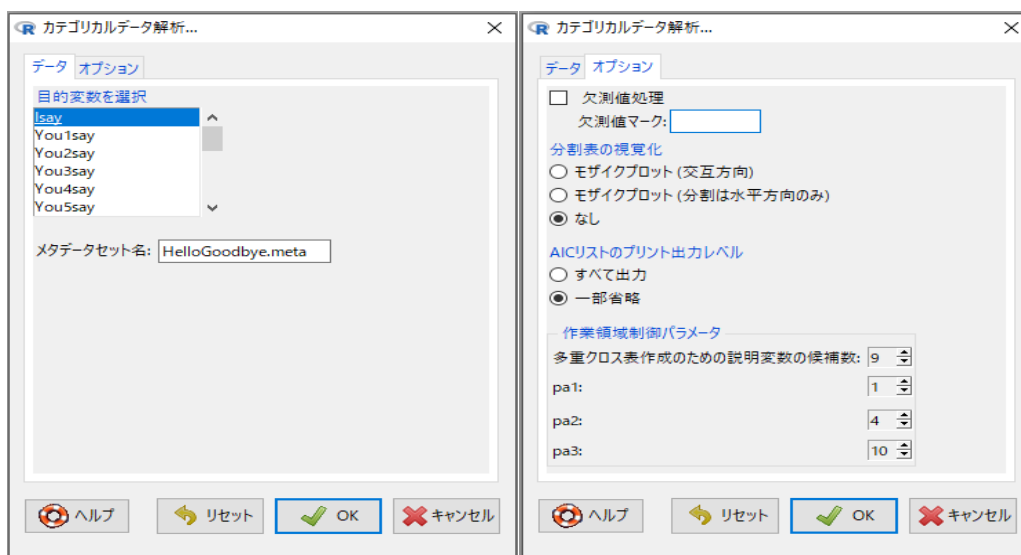
このデータは変数の数が多くレコード長も長いバイナリデータなので, オプションタブのデフォルト値の変更した方がよい場合がある。実行時間が長くなり作業領域が不足してエラーメッセージが出力されたりする場合もあるので, 多重クロス表作成のための説明変数の候補数の値はデフォルトより小さくした方がよい。この例では最大で 10 変数の組合せまで考えることとし

分割表の視覚化 ◎ なし

AIC リストのプリント出力レベル ◎ 一部省略

多重クロス表作成のための説明変数の候補数: 9

とした。



	Explanatory variables	Number of Categories of exp. var.	A I C	Difference of AIC	Weight
1	You14say	2	-37.80	0.00	1.00
2	You26say	2	-24.91	12.88	0.00
3	You49say	2	-20.37	17.42	0.00
4	You27say	2	-9.63	28.16	0.00
5	You4say	2	-9.52	28.28	0.00
6	You40say	2	-4.85	32.95	0.00
7	You23say	2	-0.80	37.00	0.00
8	You10say	2	-0.66	37.13	0.00

9	You39say	2	-0.58	37.21	0.00
10	You19say	2	-0.52	37.28	0.00
	:				

<< Summary of subsets of explanatory variables >>

Response variable : Isay

	Explanatory variables	Number of Categories of exp. var.	A I C	Difference of AIC	Weight
1	You14say	2	-37.80	0.00	1.00
1	You14say You49say You26say You4say You27say You40say You23say You10say You39say	512	-99.62	0.00	1.00
2	You14say You49say You26say You4say You27say You40say You23say You10say	256	-98.87	0.75	0.69
3	You14say You49say You26say You4say You27say You40say You23say You39say :	256	-98.83	0.79	0.67

<< Contingency table constructed by the best subset of explanatory variables >>

X(1) : Isay

X(2) : You14say

X(3) : You49say

X(4) : You26say

X(5) : You4say

X(6) : You27say

X(7) : You40say

X(8) : You23say

X(9) : You10say

X(10) : You39say

X	X	X	X	X	X	X	X	X	response variable X(1)		
(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	1	2	Total

1	1	1	1	1	1	1	1	1	7473 (98.8)	93 (1.2)	7566 (100.0)
1	1	1	1	1	1	1	1	2	120 (96.8)	4 (3.2)	124 (100.0)
1	1	1	1	1	1	1	1	2	36 (94.7)	2 (5.3)	38 (100.0)
1	1	1	1	1	1	1	1	2	0 (0.0)	0 (0.0)	0 (0.0)
1	1	1	1	1	1	1	2	1	497 (99.6)	2 (0.4)	499 (100.0)
1	1	1	1	1	1	1	2	1	0 (0.0)	0 (0.0)	0 (0.0)
1	1	1	1	1	1	1	2	2	0 (0.0)	0 (0.0)	0 (0.0)
1	1	1	1	1	1	1	2	2	0 (0.0)	0 (0.0)	0 (0.0)
1	1	1	1	1	1	2	1	1	10 (83.3)	2 (16.7)	12 (100.0)
1	1	1	1	1	1	2	1	1	0 (0.0)	0 (0.0)	0 (0.0)
:											

Total									13773 (98.7)	181 (1.3)	13954 (100.0)

<Note>

X(1) : Isay

category

variable value

1

0

2

1

X(2) : You14say

category


variable value

	1	0
	2	1
X(3) : You49say		
	category	variable value
	1	0
	2	1
:		
AIC = -99.62		
base AIC = 1934.54		

4.4 R パッケージ catdap 内データセット MissingHealthData の解析 (8 変数, レコード長 52, 欠測値あり)

- メニュー「データ」
- > 「パッケージ内のデータ」
- > 「アタッチされたパッケージからデータセットを読み込む...」
- > パッケージ catdap のデータセット MissingHealthData を選択して [✓OK]

この結果、データセット: の欄は [MissingHealthData](#) となり [データセットを表示] をクリックすると以下のようなデータが表示される。これは 4.1 HealthData の max.press と min.press に欠測値 300 がある場合のデータである。



The screenshot shows the R console window titled "MissingHealthData". It displays the output of a command, likely `print(data)`, showing a data table with 10 rows and 8 columns. The columns are labeled: `opthalgo`, `ecg`, `symptoms`, `age`, `max.press`, `min.press`, `aortic.wav`, and `cholesterol`. The data is as follows:

	opthalgo	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
1	1	1	A	52	300	300	8.4	low
2	2	1	A	55	300	300	6.7	low
3	1	1	A	56	300	300	7.9	low
4	2	2	A	57	300	300	9.3	low
5	2	2	A	53	300	300	8.3	low
6	1	1	A	56	172	106	8.6	high
7	1	1	A	56	168	96	9.1	high
8	1	1	A	54	148	102	8.0	low
9	2	2	A	58	180	98	7.5	low
10	1	1	A	51	186	96	8.6	high

4.1 の場合と同様にメタデータセット MissingHealthData.meta を作成する.

The screenshot shows a software window titled "データエディタ: MissingHealthData.meta". It contains a table with the following data:

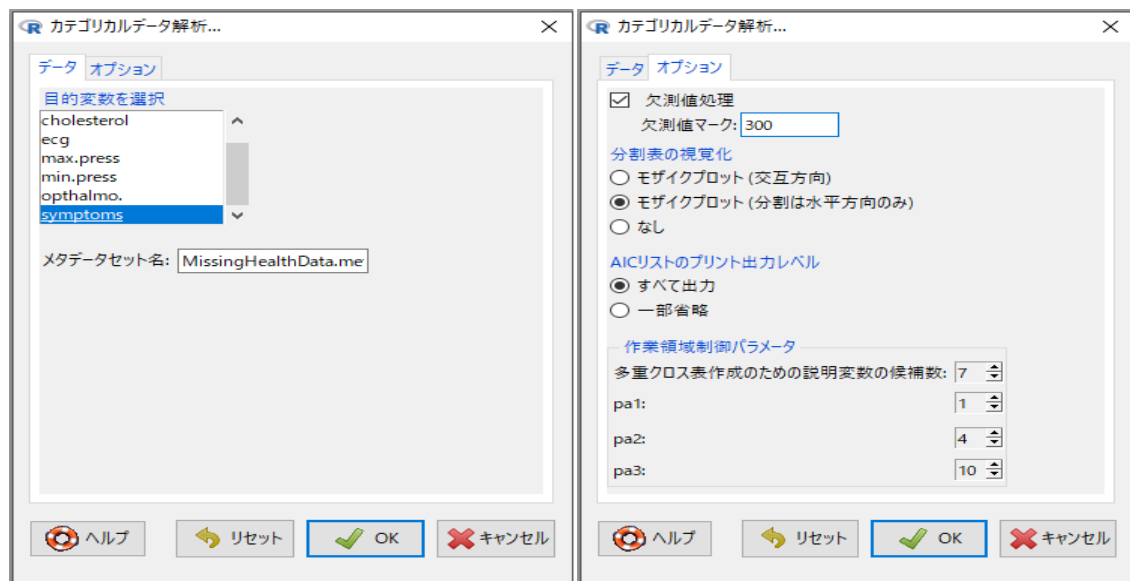
		1	2	3	4	5	6	7	8
	rowname	opthamo.	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
1	accuracy	0	0	0	1	1	1	0.1	0
2	pooling	0	0	0	0	0	0	0	0

At the bottom of the window, there are three buttons: "ヘルプ" (Help), "OK" (with a green checkmark icon), and "キャンセル" (Cancel, with a red X icon).

データセット: [MissingHealthData.meta](#) となっているところをクリックして,
[MissingHealthData](#) をアクティブにする.

メニュー「catdap」
->「カテゴリカルデータ解析...」

を選択して解析を進める. 目的変数に symptoms を選択, 欠測値処理を ✓ チェックして
欠測値マークに '300' を入力, それ以外はデフォルトのままとする.



出力 (途中省略)

```
> MissingHealthData.meta
```

	optharmo.	ecg	symptoms	age	max.press	min.press	aortic.wav	cholesterol
accuracy	0	0	0	1	1	1	0.1	0
pooling	0	0	0	0	0	0	0.0	0

```
> acc <- c(0,0,0,1,1,1,0.1,0)
```

```
> pooling <- c(2,2,2,0,0,0,0,2)
```

```
> Catdap2(MissingHealthData, pool=pooling, response.name="symptoms",  
+ accuracy=acc, nvar=8, missingmark=300, pa1=1, pa2=4, pa3=10, print.level=0,  
+ plot=1)  
:
```

```
<< Summary of subsets of explanatory variables >>
```

```
Response variable : symptoms
```

```
-----
```

	Number of Explanatory variables	Categories of exp. var.	A I C	Difference of AIC	Weight
1	aortic.wav min.press	6	-10.58	0.00	1.00
2	max.press aortic.wav	6	-10.23	0.35	0.84
3	max.press	3	-9.54	1.04	0.59
4	max.press cholesterol	6	-9.18	1.40	0.50
5	max.press aortic.wav min.press	18	-8.51	2.08	0.35
6	max.press min.press	9	-7.75	2.83	0.24
7	min.press	3	-6.49	4.09	0.13
8	aortic.wav	2	-6.25	4.34	0.11
9	max.press ecg	6	-5.68	4.90	0.09
10	aortic.wav min.press cholesterol :	12	-5.20	5.38	0.07

<< Contingency table constructed by the best subset of explanatory variables >>

X(1) : symptoms

X(2) : aortic.wav

X(3) : min.press

X X		response variable X(1)		
(2)	(3)	1	2	Total
1	1	1 (8.3)	11 (91.7)	12 (100.0)
1	2	7 (53.8)	6 (46.2)	13 (100.0)
1	3	2 (100.0)	0 (0.0)	2 (100.0)

2	1	7 (58.3)	5 (41.7)	12 (100.0)
2	2	9 (90.0)	1 (10.0)	10 (100.0)
2	3	3 (100.0)	0 (0.0)	3 (100.0)

Total		29 (55.8)	23 (44.2)	52 (100.0)
-------	--	-------------	-------------	--------------

<Note>

X(1) : symptoms

category	variable value
1	A
2	B

X(2) : aortic.wav

category	value range
1	6.30000e+00 - 8.25000e+00
2	8.25000e+00 - 1.02000e+01

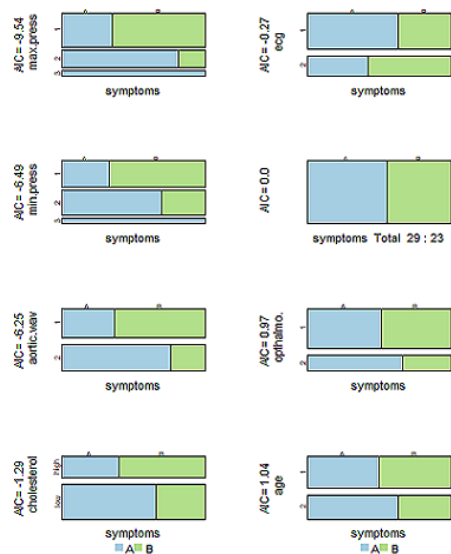
X(3) : min.press

category	value range
1	5.60000e+01 - 9.35000e+01
2	9.35000e+01 - 1.20000e+02
3	missing of type 1

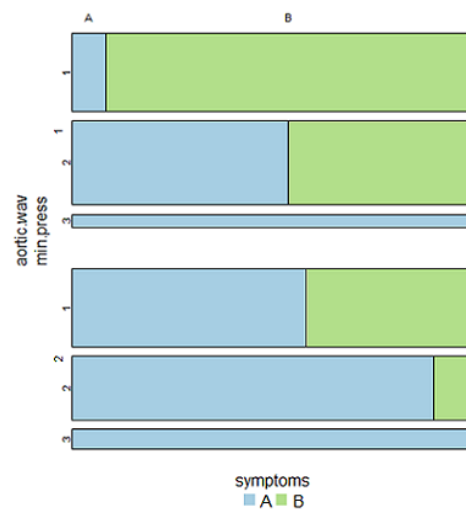
AIC = -10.58

base AIC = 73.39

Single Explanatory Models in ascending order of AIC



Minimum AIC Model of the Response Variable Distribution



5. 参考文献

- [1] Y.Sakamoto and H.Akaike (1978). Analysis of Cross-Classified Data by AIC. Ann. Inst. Statist. Math., 30, pp.185-197.
- [2] K.Katsura and Y.Sakamoto (1980). A Categorical Data Analysis Program Package, Computer Science Monographs, No.14. The Institute of Statistical Mathematics, Tokyo.
- [3] Y.Sakamoto (1985). Categorical Data Analysis by AIC, Kluwer Academic publishers.
- [4] (株) NTT データ数理システム (2015). 情報量統計学的データ可視化ツール.
<http://hdl.handle.net/10787/3614>
- [5] 石黒 真木夫 (2016) CATDAP マニュアル. <http://hdl.handle.net/10787/3821>
- [6] 石黒 真木夫 (2016) 統計モデル可視化. <http://hdl.handle.net/10787/3823>