

R パッケージ `catdap` とその拡張

1. `catdap` パッケージ

CATDAP (CATegorical Data Analysis Program)は、統計数理研究所で開発された最適な分割表（クロス表）の探索のためのプログラムであり、CATDAP-01 と CATDAP-02 がある。CATDAP-01 は、カテゴリカルな（質的）データに対し二次元分割表の比較を行い各変数の間の関係の深さを検出するプログラムである。全ての変数がカテゴリカルであることが前提となる。CATDAP-02 は、ひとつの着目した項目（目的変数）を固定し他の項目（説明変数）の組合せで多次元分割表を作り、目的変数の分布の違いを最も適切に説明する説明変数組合せを探索する。このプログラムは量的な項目も適当に区分してカテゴリカルなデータに帰着することにより、変数が量的か質的かにかかわらず適用できる。いずれも最適な説明変数の選択には、AIC (Akaike Information Criterion) が使われている。

R パッケージ `catdap` は、これらを基に作成された関数 `catdap1()` と `catdap2()` とからなるが、`catdap2()` では以下の機能を追加した (石黒[5]) :

- ・ Base AIC (=説明変数なしモデルの AIC) の利用
- ・ 連続値目的変数に適用
- ・ 目的変数, 説明変数に欠測が含まれるデータに適用

Base AIC の利用により、このプログラムであてはめられたモデルと、目的変数に正規分布を仮定する回帰モデル あるいは ロジスティック回帰モデルとの AIC の比較が可能になっている。

`catdap.zip` を解凍すると、`catdap` パッケージの使い方およびソースファイルとバイナリファイルがある。

2. `catdap2ext` パッケージ

さらに論理的にあり得る全てのタイプのデータに対応可能なプログラム (石黒[7]) が提案され、それを基に `catdap2()` に以下の二つの機能

- ・ 角度データ (circular data) の扱い
- ・ 条件付き確率のグラフィカル出力

を追加した拡張版が `catdap2ext` パッケージである。

`catdap2ext.zip` を解凍すると、`catdap2ext` パッケージの使い方と実行例、ソースファイルとバイナリファイルがある。

3. RcmdrPlugin.catdap パッケージ

catdap パッケージの `catdap2()` の機能を, R コマンドの GUI から利用可能にしたのが RcmdrPlugin.catdap パッケージである.

RcmdrPlugin.catdap.zip を解凍すると, RcmdrPlugin.catdap パッケージの使い方と実行例, ソースファイルとバイナリファイルがある.

4. 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>
- [7] 石黒 真木夫 (2021) CATDAP 機能強化プラン. <http://hdl.handle.net/10787/00034178>