

第7章のデータとプログラムについての説明

関西大学 高井啓二

ここでは、第7章の7.1.2のLittleの検定と、7.1.3のHausmanの検定で使ったデータとプログラムの、説明を与える。

7.1.2の例7.2のデータ `apple.csv` とプログラム `Little.R` について

データ `apple.csv` は、二つの興味ある変数の列と二つの欠測指標の列から成る。このデータをもとに、Littleの検定を行う。Littleの検定では、正規分布モデルのパラメータの推定が必要である。その推定には、第3章で用いた `emexample1.R` を使う。推定したパラメータを用いて、Littleの検定を行う。

プログラム `Little.R` の冒頭では `apple.csv` と `emexample.R` (ともに第3章で利用)を用いて、パラメータを推定している。ただし、第3章の `emexample.R` では、欠測指標行列は観測される変数に対応する要素を0、欠測する変数に対応する要素を1としているが、Littleの検定を計算するためのプログラムでは、欠測指標行列 `R` は1のとき観測、0のとき欠測としている。

<使用例> *現在のフォルダが”Chapter7”であるとする。

```
source("Little.R")
```

```
# データ, EM アルゴリズム, Little の統計量の読み込み
```

```
d2 #検定統計量の算出
```

7.1.3の例7.4のデータ `apple.csv` とプログラム `hausman.R` について

この例で使用するデータは上の例と同じ `apple.csv` である。上のプログラムの実行に続いて実行する場合にはデータの読み込みは不要であるが、このプログラム単独で動かす場合には、データを読み込まねばならない。検定統計量を計算するだけなので、プログラムはいたって単純である。

`hausman.stat(R,D)` : ハウスマン統計量を計算する。引数の `R` は欠測指標行列、`D` はデータである。`R` の要素は対応する `D` の要素が観測されているとき1、欠測するとき0である。`D` の欠測部分には適当な値として1000を入れてあるが、この値は結果に影響を及ぼさない(最終的には0をかけて消してしまうので結果に影響を及ぼさない)。戻り値は、検定統計値である。

<使用例 1> *現在のフォルダが”Chapter7”であるとし，上の `source(“Little.R”)` が実行されているものとする．このときには，以下のようにすれば良い．

```
source(“hausman.R”) #プログラムの読み込み
hausman.stat(R,D) #検定統計量の計算
```

<使用例 2> *現在のフォルダが”Chapter7”であるとし，上の `source(“Little.R”)` は実行していないものとする．このときには，データを読み込むことから始める．

```
tmp.dat <- read.csv(“apple.csv”,header=F) #データの読み込み
tmp.dat <- as.matrix(tmp.dat)
D <- tmp.dat[,1:2]
R <- 1 - tmp.dat[,3:4] #1 と 0 の役割を交換
source(“hausman.R”) #プログラムの読み込み
hausman.stat(R,D) #検定統計量の計算
```