

## 医学情報処理演習第 13 回「生存時間解析」2008 年 1 月 21 日

### 今回の課題

MASS ライブラリに含まれているデータフレーム Aids2 は、1991 年 7 月 1 日以前にオーストラリアで AIDS と診断された患者についてのデータである。2843 人について次の枠内の変数を含む。

state グループ化された患者の出身州（"NSW"=ニューサウスウェールズは"ACT"=オーストラリア首都区別地域を含み,"other"は"WA"=西オーストラリア,"SA"=南オーストラリア,"NT"=ノーザンテリトリー,"TAS"=タスマニアを含む）。因子型, 4 水準（第 1 が"NSW", 第 2 が"Other", 第 3 が"QLD"=クイーンズランド, 第 4 が"VIC"=ヴィクトリア）。

sex 性別。因子型, 2 水準（第 1 が"F", 第 2 が"M"）。

diag 診断日。1970 年 1 月 1 日からの通日。整数型。ヘルプには (Julian) date と書かれているが、通常、ユリウス通日 (Julian date) といえは紀元前 4713 年 1 月 1 日正午からの経過日数を指すので、これは厳密にはユリウス通日ではない。

death 死亡日あるいは観察打ち切り日。1970 年 1 月 1 日からの通日。整数型。ヘルプには (Julian) date と書かれているが、通常、ユリウス通日 (Julian date) といえは紀元前 4713 年 1 月 1 日正午からの経過日数を指すので、これは厳密にはユリウス通日ではない。

status 観察終了時の状態。因子型, 2 水準（第 1 水準は"A"で生存, 第 2 水準は"D"で死亡）。

T.categ 自己申告による感染経路のカテゴリ（このデータの元になった Australian National Centre in HIV Epidemiology and Clinical Research の調査では"Exposure categories"として大別して 5 つのコード, それぞれの下にいくつかのサブカテゴリが割り振られているが、このデータでは hs (同性愛), hsid (同性愛かつ薬物注射針共用), id (薬物注射針共用), het (異性愛), haem (血友病または血液凝固異常), blood (血液), mother (母子垂直感染), other (その他) のどれかである。因子型。8 水準（順序は上の通り）。

age 診断時年齢（歳）。整数型。

このデータで AIDS と診断された患者の死亡ハザードに対して、出身州による差があるか、性差があるか、感染経路による差があるかをコックス回帰によって検討せよ。ただし年齢の影響を調整すること。統計処理の結果と、考察する文章と学籍番号を Wordなどで打って印刷し、署名して提出すること。

なお、データフレーム内の各変数を使えるようにして生存時間オブジェクトを作るまでのコードは、以下のようなになる。

```
kadai13-2007.R  
library(MASS)  
library(survival)  
attach(Aids2)  
time <- Surv(death-diag,as.integer(status)-1)
```