

医学情報処理演習第13回「生存時間解析」課題解答例 2008年1月21日

生データの様子を見ておくことは常に必要なので、まず年齢分布を次のコードにより集計する。年齢は連続変数なので、`hist()` を `plot=F` オプションをつけて実行する。`right=F` オプションをつけると、日本で通常用いられている区切りである「以上、未満」の集計ができる。

```
kadail3ans-2007.R(1)

library(MASS)
library(survival)
attach(Aids2)
sex <- relevel(sex,2)
hist(age,plot=F,right=F)
```

年齢 5 歳階級別患者数は下表の通りである。やや右裾を引いた分布になっていることと、0 歳から 5 歳までの患者が 5 歳から 10 歳までの患者より多いことがわかる。

表 1. 1991 年までのオーストラリアにおける年齢階級別 AIDS 患者数

年齢階級	患者数
0 歳以上 5 歳未満	10
5 歳以上 10 歳未満	6
10 歳以上 15 歳未満	9
15 歳以上 20 歳未満	14
20 歳以上 25 歳未満	118
25 歳以上 30 歳未満	466
30 歳以上 35 歳未満	568
35 歳以上 40 歳未満	575
40 歳以上 45 歳未満	472
45 歳以上 50 歳未満	295
50 歳以上 55 歳未満	144
55 歳以上 60 歳未満	89
60 歳以上 65 歳未満	34
65 歳以上 70 歳未満	29
70 歳以上 75 歳未満	9
75 歳以上 80 歳未満	3
80 歳以上 85 歳未満	2

次に、年齢の影響をひとまず考えずに、それぞれの変数のカテゴリ別にカプランマイヤ推定をする。次の枠内のコードで生存曲線の描画と生存時間の中央値の推定がなされる。

```
kadail3ans-2007.R(2)

time <- Surv(death-diag,as.integer(status)-1)
par(las=1,mfrow=c(1,3),family="Japan1GothicBBB",cex=0.8)
plot(km1 <- survfit(time~sex), main="カプランマイヤ法による\nAIDS 患者生存曲線 (性別)", xlab="生存日数 (日)", lty=1:2)
legend(max(death-diag)*0.6,1,lty=1:2,legend=levels(sex))
plot(km2 <- survfit(time~state), main="カプランマイヤ法による\nAIDS 患者生存曲線 (州別)", xlab="生存日数 (日)", lty=1:4)
legend(max(death-diag)*0.6,1,lty=1:4,legend=levels(state))
plot(km3 <- survfit(time~T.categ), main="カプランマイヤ法による\nAIDS 患者生存曲線 (感染経路別)",
      xlab="生存日数 (日)", lty=1:8, col=(1:8 %% 7)+1)
legend(max(death-diag)*0.6,1,lty=1:8,col=(1:8 %% 7)+1,legend=levels(T.categ))
km1; km2; km3
```

結果は次の図表のようにまとめられる。また、この結果から、各カテゴリ別の患者数がわかる。男女別患者数は、女性 89 人、男性 2754 人で、男性が圧倒的に多い。州別患者数は、首都特別区を含むニューサウスウェールズ州が 1780 人で最も多く、ビクトリア州の 588 人、クイーンズランド州の 226 人、その他の 249 人となっている。感染経路別患者数は、同性間の性行為が 2465 人で圧倒的に多く、同性間の性行為をしていてかつ薬物注射針共用 72 人、薬物注射針共用のみが 48 人、異性間の性行為で感染した人が 41 人、血液病または血液凝固異常で血液製剤により感染した人が 46 人、血液感染が 94 人、母子垂直感染が 7 人、その他が 70 人となっている。

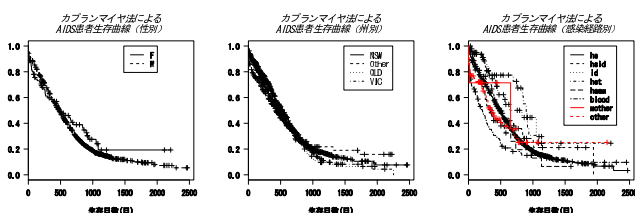


表 2. カテゴリ別生存日数のカプラン = マイヤ推定

カテゴリ	人数	死亡数	生存日数		
			中央値	95%CI 下限	95%CI 上限
男性	2754	1708	479	449	500
女性	89	53	506	303	850
ニューサウスウェールズ州	1780	1116	462	435	492
ビクトリア州	588	355	511	475	586
クイーンズランド州	226	148	420	352	551
その他の州	249	142	524	442	617
同性間の性行為	2465	1532	485	457	507
同性間の性行為 + 薬物注射針共用	72	45	492	396	826
薬物注射針共用	48	19	831	462	∞
異性間の性行為	41	17	896	842	∞
血友病や血液凝固異常の治療	46	29	339	256	734
血液	94	76	215	138	299
母子垂直感染	7	3	654	15	∞
その他の感染経路	70	40	368	275	697

最後に、次の枠内のコードにより、年齢の影響を調整して、各カテゴリ変数の生存時間への影響を調べるコックス回帰を行う。

kadail3ans-2007.R(3)

```
par(las=1,mfrow=c(1,3),family="Japan1GothicBBB",cex=0.8)
plot(survfit(res1 <- coxph(time~sex+age)),
     main="性別と年齢を共変量とする\nAIDS 患者ベースラインハザード\nによる生存曲線", xlab="生存日数(日)")
plot(survfit(res2 <- coxph(time~state+age)),
     main="州と年齢を共変量とする\nAIDS 患者ベースラインハザード\nによる生存曲線", xlab="生存日数(日)")
plot(survfit(res3 <- coxph(time~T.categ+age)),
     main="感染経路と年齢を共変量\nとする AIDS 患者ベースライン\nハザードによる生存曲線", xlab="生存日数(日)")
summary(res1); summary(res2); summary(res3)
detach(Aids2)
```

ベースラインハザードによる生存曲線は、どのカテゴリ変数を共変量にした場合でも大差ないので図示は省略する。コックス回帰の結果は、次の表のようにまとめることができる。モデルの適合は、いずれのモデルにおいても有意だった (Rao のスコア検定などで) が、 R^2 の値は非常に小さく、あまり良いモデルとはいえない。性別や州は AIDS 患者の死亡ハザードには有意に影響しないが、感染経路については、患者の大多数を占める同性間性行為による感染に比べて、薬物注射針共用や異性間性行為による感染で死亡ハザードが有意に低く、血液感染の場合に死亡ハザードが有意に高くなるのがわかった。

表 3. 年齢を共変量として調整した、性別、州、感染経路の AIDS 死亡ハザードへの影響

共変量カテゴリ	ハザード比			有意確率
	推定値	95%CI 下限	95%CI 上限	
性別モデル、性別は男性をリファレンスとした場合のハザード比、 $R^2 = 0.013$				
女性	0.901	0.685	1.18	0.45
年齢	1.02	1.010	1.02	8.2×10^{-10}
州モデル、州はニューサウスウェールズをリファレンスとした場合のハザード比、 $R^2 = 0.015$				
ビクトリア	0.965	0.856	1.09	0.56
クイーンズランド	1.137	0.958	1.35	0.14
その他の州	0.879	0.738	1.05	0.15
年齢	1.015	1.010	1.02	1.1×10^{-9}
感染経路モデル、感染経路は同性間性行為をリファレンスとした場合のハザード比、 $R^2 = 0.023$				
同性間性行為 + 薬物注射針共有	0.924	0.686	1.244	0.60
薬物注射針共有	0.618	0.392	0.973	0.038
異性間性行為	0.493	0.306	0.795	0.0037
血友病や血液凝固異常の治療	1.405	0.972	2.033	0.071
血液	1.455	1.148	1.845	0.0019
母子垂直感染	1.329	0.423	4.173	0.63
その他	1.084	0.791	1.485	0.62
年齢	1.013	1.008	1.018	9.2×10^{-8}