

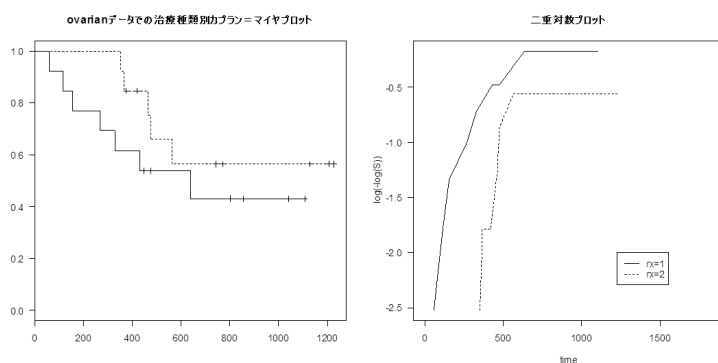
医学情報処理演習第13回課題回答例

課題

まず他の共変量を無視して、治療種類ごとに Kaplan-Meier 法で生存曲線を求め、二重対数プロットもしてみるため、下枠内を実行する。

```
it13-ans-2006.R(1)
require(survival)
loglogplot <- function(X) {
  S <- X$surv
  T <- X$time
  G <- X$ntimes.strata
  GG <- names(X$strata)
  GX <- rep(GG,G)
  xr <- c(0,max(T)*1.5)
  yr <- c(log(-log(max(S))),log(-log(min(S))))
  plot(T[GX==GG[1]],log(-log(S[GX==GG[1]])),type="l",lty=1,xlim=xr,ylim=yr,
       xlab="time",ylab="log(-log(S))",main="二重対数プロット")
  for (i in 2:length(GG)) {
    lines(T[GX==GG[i]],log(-log(S[GX==GG[i]])),lty=i)
  }
  legend(max(T),-2,legend=GG,lty=1:length(GG))
}
#
print(KM <- survfit(Surv(futime,fustat)~rx,data=ovarian))
win.metafile("it13-ans-2006.emf",width=16,height=8,pointsize=14)
par(family="sans",las=1,mfrow=c(1,2))
plot(KM,main="ovarian データでの治療種類別 Kaplan-Meier プロット",lty=1:2)
loglogplot(KM)
dev.off()
```

すると、カレント作業ディレクトリに it13-ans-2006.emf という名前のウィンドウズ・メタファイル形式のファイルとして、下図ができる。治療種類 2 の方が治療種類 1 よりも生存時間が長いことがわかる。Kaplan-Meier 法による生存時間の中央値は、治療種類 1 では 638 日（95%信頼区間は 268 日から無限大）、治療種類 2 では無限大（95%信頼区間は 475 日から無限大）である。また、二重対数プロットは、ほぼ平行に見える。



そこで、次の枠内のように、治療種類、年齢、残留疾病の有無を共変量としたコックス回帰を実行する。モデルのデータのばらつきの半分弱を説明し ($R^2 = 0.475$), 有意に当てはまっているが (Rao のスコア検定で $\chi^2 = 20.8, d.f. = 3, p = 0.00012$ など), 生存時間に対して有意に影響しているのは年齢だけであり (ハザード比 1.137, 95%信頼区間は 1.036-1.25, $p=0.0066$), 残留疾病の有無と治療種類の影響は有意ではなかった。

it13-ans-2006.R(2)

```
res <- coxph(Surv(futime, fustat)~rx+age+factor(resid.ds), data=ovarian)
summary(res)
```

したがって、年齢と残留疾病の有無を共変量として調整したときに、この治療種類 1 と 2 の違いは、卵巣がん患者の生存時間に有意差をもたらすとはいえないという結論になる。

なお、<http://phi.med.gunma-u.ac.jp/medstat/it13-ans-2006-2.R> を使うと、2 群別々に共変量として年齢と残留疾病の有無を調整してコックス回帰したときのベースラインハザードを使った二重対数プロットを描くことができるので参考にされたい。