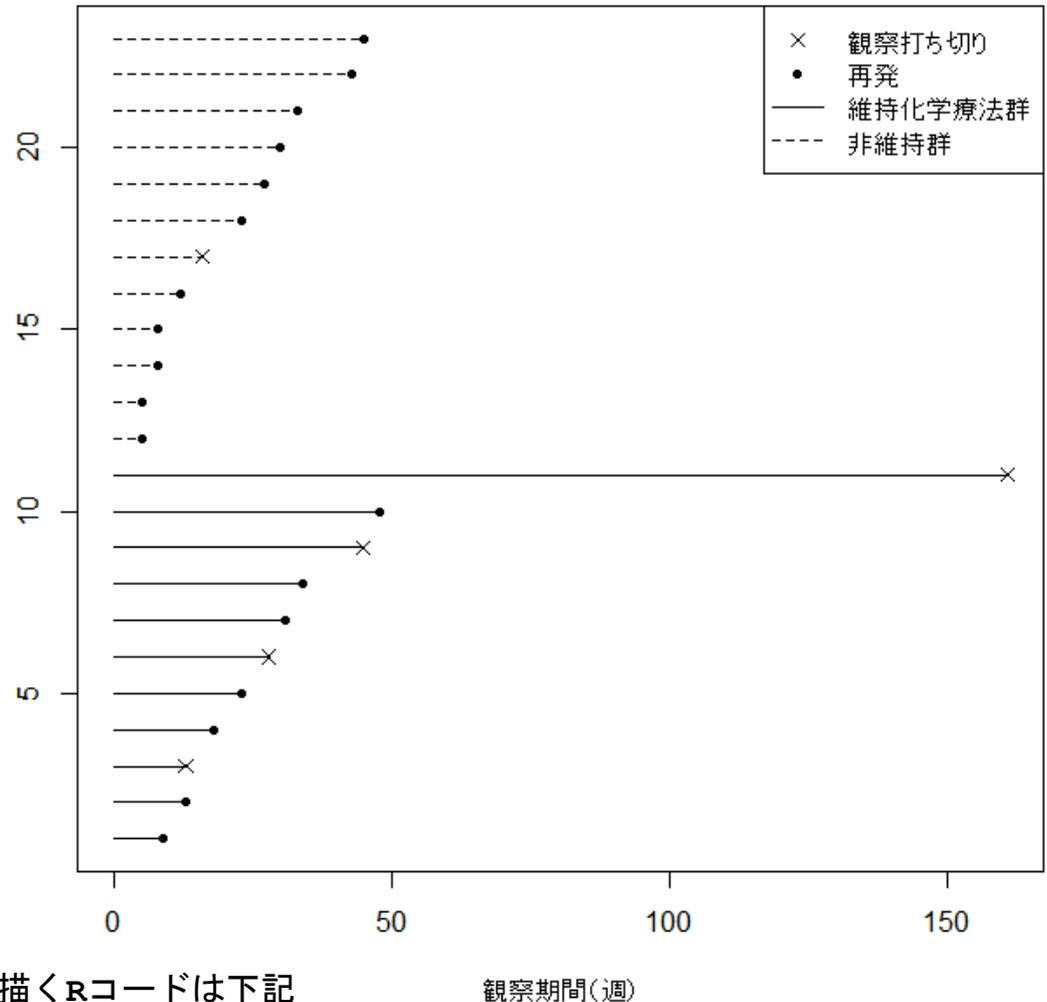


生存時間解析

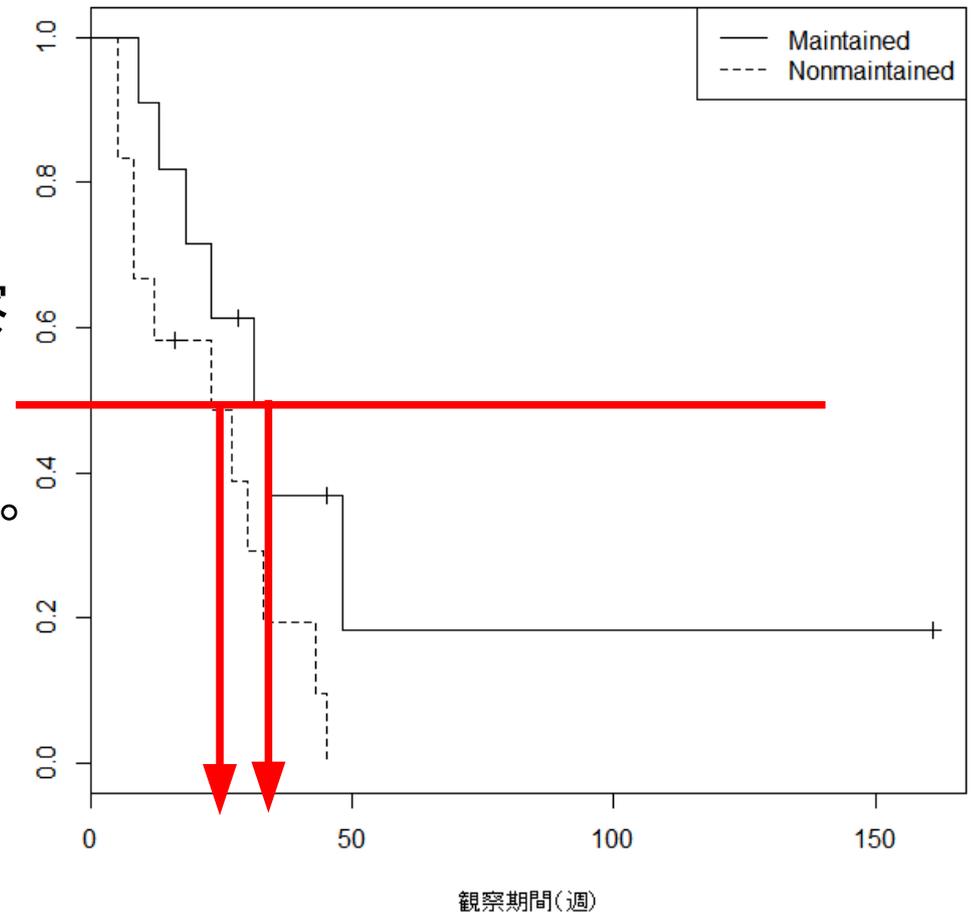
- イベント発生までの時間を分析する
- イベント発生の有無より、時間の分布を調べる方が情報量が多い
- 「イベントがまだ発生していない」状態の扱いが鍵
 - 観察終了時まではイベントが発生していないので、イベント発生までの時間が長めな人である可能性が高い
 - 単純に除去するとイベント発生までの時間が過小評価される
 - 右側打ち切りレコードとして扱う！



```
上図を描くRコードは下記
library(survival)
plot(1:23~seq(0,max(time),len=23),type="n",data=leukemia,
      xlab="観察期間(週)",ylab="")
segments(rep(0,23),1:23,leukemia$time,1:23,
         lty=ifelse(leukemia$x=="Maintained",1,2))
points(leukemia$time,1:23,
       pch=ifelse(leukemia$status==0,4,20))
legend("topright",pch=c(4,20,NA,NA),lty=c(NA,NA,1,2),
       legend=c("観察打ち切り","再発","維持化学療法群","非維持群"))
```

Kaplan-Meier法

- 右側打ち切りを考慮した生存時間の中央値(カプラン=マイヤ推定量)を求める方法
- 時点0では生存率が1。各死亡イベントが起こった時点ごとに、その時点で観察中の人数でその時点の死亡数を割った値を1から引いた値を掛けていくと、順次その時点までの生残率が得られる。
- 死亡イベントと死亡イベントの間で打ち切りがあると観察中人数が減る。
- 生残率が0.5を横切った時点が生存時間の中央値(維持群31週, 非維持群23週)。



- 先のam1データでのコードは下記

```
library(survival) # ライブラリ呼び出しが必須  
(KM <- survfit(Surv(time, status) ~ x, data=leukemia))  
plot(KM, lty=1:2, xlab="観察期間(週)")  
legend("topright", lty=1:2,  
      legend=names(table(leukemia$x)))
```
- 群分けがないときはsurvfit()の中を~xではなく~1と書く

ログランク検定

- 「2群の生存時間に差が無い」帰無仮説の検定。
- 生存時間の順序の情報を用いる
- amlデータで、維持群と非維持群で白血病再発までの時間に差が無いという帰無仮説をログランク検定するコードは,
`library(survival)`
`survdiff(Surv(time, status) ~ x, data=leukemia)`

- 結果は

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
x=Maintained	11	7	10.69	1.27	3.4
x=Nonmaintained	12	11	7.31	1.86	3.4

Chisq= 3.4 on 1 degrees of freedom, p= 0.0653

- p値が0.05より大きいので、有意水準5%で有意な差があるとはいえない。

コックス回帰

- 比例ハザードモデルともいう
- 仮想的な「基準人」の生存曲線を計算し、個々の人のイベント発生ハザードは基準人のハザードに比例するという「比例ハザード性」を仮定する
- 生存曲線の関数形は仮定しないのでセミパラメトリックなモデルと言われる
- 詳細は説明しないが、コードは下記

```
library(survival)
res<-coxph(Surv(time,status)~x, data=leukemia)
summary(res)
plot(survfit(res)) # 基準生存曲線と95%信頼区間
# 比例ハザード性の確認は二重対数プロットが平行ならOK
KM <- survfit(Surv(time,status)~x, data=leukemia)
plot(KM,fun=function(y){log(-log(y))},lty=1:2)
```