

Практика №6 Анализ данных типа времени жизни.

Долаева А.Р., г.20.М04-мм

26/05/2021

Данные о больных глиомой из ВМА (base_short.xls).

karnovsky - индекс Карновского. Общее состояние онкологических больных (0-100%).

Умиравший (10%), состояние нормальное, жалоб нет (100%).

diagnosis:

0 стадия — опухоль расположена в одном слое, и не прорастает за его пределы.

1 стадия — ограниченный очаг (до 2 см).

2 стадия — опухоль имеет те же размеры, но поразила регионарный лимфоузел (ближайший к очагу).

3 стадия — размер новообразования увеличивается до 5 и более см, прорастает в окружающие ткани и дает множественные метастазы в лимфоузлы.

4 стадия — опухоль любого размера, которая имеет как минимум один метастаз в отдаленном органе.

time - параметр для обозначения пройденного времени (дни): от дня операции либо до момента выбывания из наблюдений, либо до смерти больного.

c - цензурирование: 0 - больной умер, 1 - выбыл из наблюдения.

1. Построить оценку Каплана-Мейера с доверительным интервалом.

Первые 10 строк анализируемых данных:

```
data <- read.table(file = "base_short.csv", header = TRUE, sep = ";")
data<-na.omit(data[, -1])
data[1:10,]
```

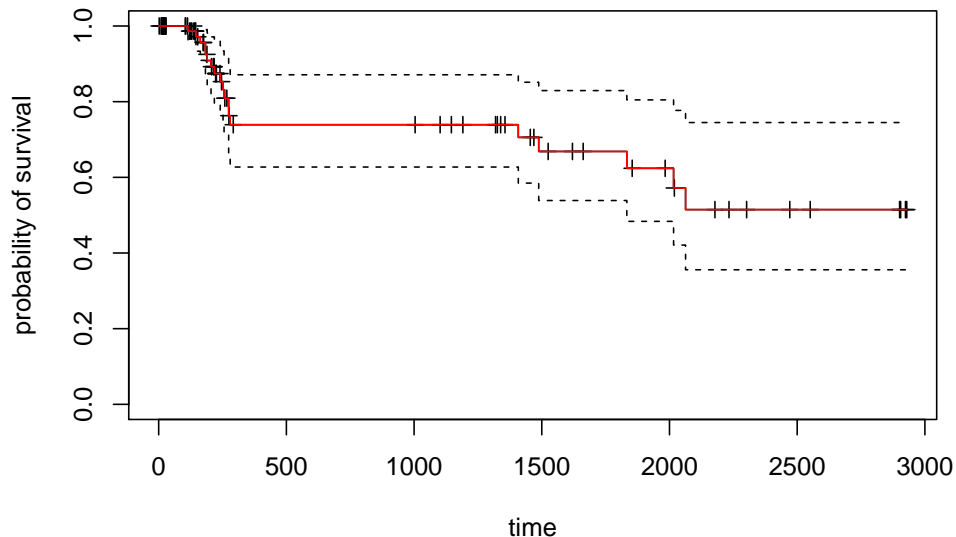
```
##      diagnosis karnovsky c time
## 1           3         70 1  207
## 4           3         90 1  240
## 7           3         40 1   26
## 9           3         90 1   21
## 15          3         80 1   14
## 17          4         70 1  223
## 20          2         90 1   10
## 21          4         80 1   16
## 23          3         60 1   12
## 29          4         60 1   12
```

Построение кривой дожития Каплана-Мейера по одной группе:

событием (event) в Surv является отказ (наблюдалось ли событие (1) или нет (0)), а не цензурирование, поэтому переопределяем данные: event=1-c.

~1 для наблюдаемых событий

```
fitCS <- survfit(Surv(time, event=1-c)~1, data=data)
plot(fitCS, mark.time = TRUE, conf.int = TRUE, xscale = 1,
     xlab = 'time', ylab = 'probability of survival')
lines(fitCS, conf.int = FALSE, col='red')
```



На графике показана вероятность того, что отказ (с) наступит после* момента времени time (t).

Кривая дожития - $F(t) = P\{\tau > t\}$

τ - переменные из data, time - моменты времени (ось абсцисс).

* здесь должно быть “после”

```
print(fitCS, print.rmean=TRUE)
```

```
## Call: survfit(formula = Surv(time, event = 1 - c) ~ 1, data = data)
##
##          n      events      *rmean *se(rmean)      median      0.95LCL      0.95UCL
##        121         19       1973       175         NA         1833         NA
##      * restricted mean with upper limit = 2929
```

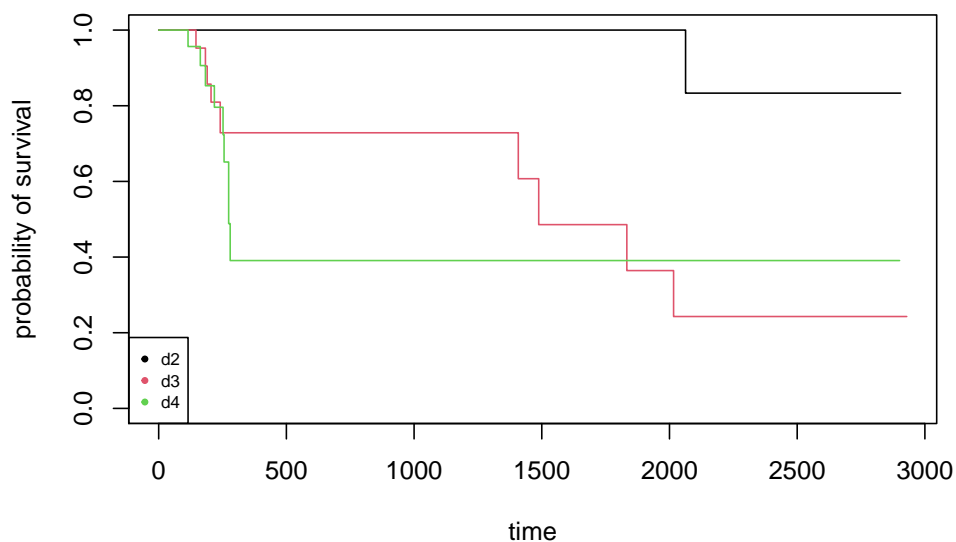
1973 - среднее времени дожития*.

* среднее для одной группы

2. Вычислить интенсивность постоянного риска. Сравнить две кривые дожития.

Кривые дожития по для стадий 2, 3, 4 соответственно:

```
fitCS <- survfit(Surv(time, event=1-c)~diagnosis, data=data)
plot(fitCS, mark.time = FALSE, conf.int = FALSE, xscale = 1,
     xlab = 'time', ylab = 'probability of survival', col=seq(3))
legend('bottomleft', c('d2', 'd3', 'd4'), col=seq(3), cex=0.7, pch=20)
```



Исходя из графика кривых дожития, можно сделать следующий вывод.

Черная кривая, соответствующая второй группе расположена выше кривых, вырисованных для третьей и четвертой групп. Пациенты со 2 стадией обладают более высокой вероятностью дожития, чем пациенты других стадий.

Для сравнения кривых дожития используются характеристики медианы дожития и среднего времени дожития.

```
print(fitCS, print.rmean=TRUE)
```

```
## Call: survfit(formula = Surv(time, event = 1 - c) ~ diagnosis, data = data)
##
##              n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## diagnosis=2  46      1  2785      132      NA      NA      NA
## diagnosis=3  37      9  1584      277  1488    1408      NA
## diagnosis=4  38      9  1289      371   274    256      NA
##      * restricted mean with upper limit = 2929
```

Как видно из следующей таблицы, наименьшее время дожития (1289) наблюдается при 4 стадии.

Экспоненциальной модели кривой дожития $\varrho = \frac{d}{\sum_i t_i}$.
(число отказов (1-с) на суммарную наработку (time))

Оценка постоянного риска:

```
rho<-sum(1-data$c)/sum(data[,4])
rho
```

```
## [1] 0.0002892065
```

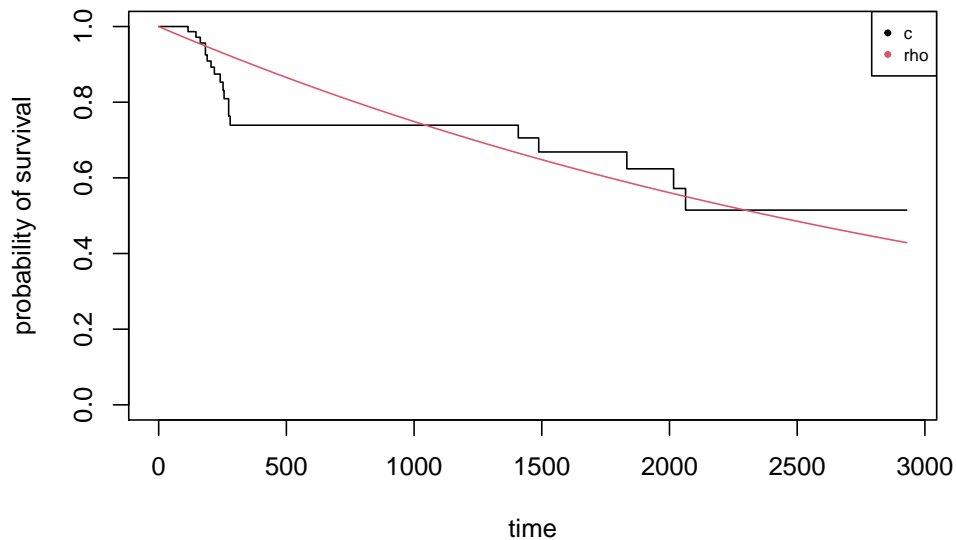
Графическое представление оценки:

```
fitCS <- survfit(Surv(time, event=1-c)~1, data=data)
plot(fitCS, mark.time = FALSE, conf.int = FALSE, xscale = 1,
```

```

xlab = 'time', ylab = 'probability of survival', col=1)
curve(exp(-rho*x), from=0, to=max(data$time), add=TRUE, col=2)
legend('topright', c('c', 'rho'), col=seq(2), cex=0.7, pch=20)

```



Если риск (интенсивность) является постоянным $h(t) = \frac{-(F(t))'}{F(t)} = \varrho$, то функция надежности (линия дожития) убывает по экспоненте $F(t) = e^{-\varrho t}$.

* пропустила формулы

Лог-ранговый критерий ($\rho=0$):

```
survdif(Surv(time, event=1-c)~diagnosis, data=data, rho=0)
```

```

## Call:
## survdif(formula = Surv(time, event = 1 - c) ~ diagnosis, data = data,
##       rho = 0)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## diagnosis=2 46         1    9.32      7.43    15.20
## diagnosis=3 37         9    5.19      2.80     3.90
## diagnosis=4 38         9    4.49      4.54     6.14
##
##  Chisq= 15.4  on 2 degrees of freedom, p= 4e-04

```

Значимое отличие групп (pvalue = 0,0004)

Попарное сравнение с помощью критерия Гехана-Вилкоксона:

```

for (i in c(2,3,4)) {
p<- with( subset(data, data$diagnosis!=i, select = c(time, diagnosis, c)),
  gehan.test(time, 1-c, diagnosis))$p.value;

```

```
d<-c(2,3,4)
d<-d[!(d %in% i)]
print(paste("diagnosis -",list(d), "pvalue", round(p, 4)));
}
```

```
## [1] "diagnosis - c(3, 4) pvalue 0.6104"
## [1] "diagnosis - c(2, 4) pvalue 3e-04"
## [1] "diagnosis - c(2, 3) pvalue 0.0013"
```

pvalue 0.6104 > 0.05 при сравнении 3 и 4 групп, что может говорить о случайном различии между ними. При сравнении 2 и 3, а также 2 и 4 групп pvalue < 0.05. Значит можем сделать вывод, что группа с диагнозом 2 стадии значительно отличается от третьей и четвертой групп.

3. Применить регрессионную модель Кокса для выявления факторов, значимо влияющих на дожитие.

Зависимость индекса Карновского на дожитие:

```
fitCS<-coxph(Surv(time, event=1-c)~diagnosis+karnovsky, data=data)
fitCS
```

```
## Call:
## coxph(formula = Surv(time, event = 1 - c) ~ diagnosis + karnovsky,
##       data = data)
##
##               coef exp(coef) se(coef)      z      p
## diagnosis  0.83696   2.30933  0.33862  2.472 0.0134
## karnovsky -0.05012   0.95112  0.02811 -1.783 0.0746
##
## Likelihood ratio test=16.26  on 2 df, p=0.0002941
## n= 121, number of events= 19
```

p = 0.0002 < 0.5 различие совместного влияния diagnosis и karnovsky на дожитие значимо.

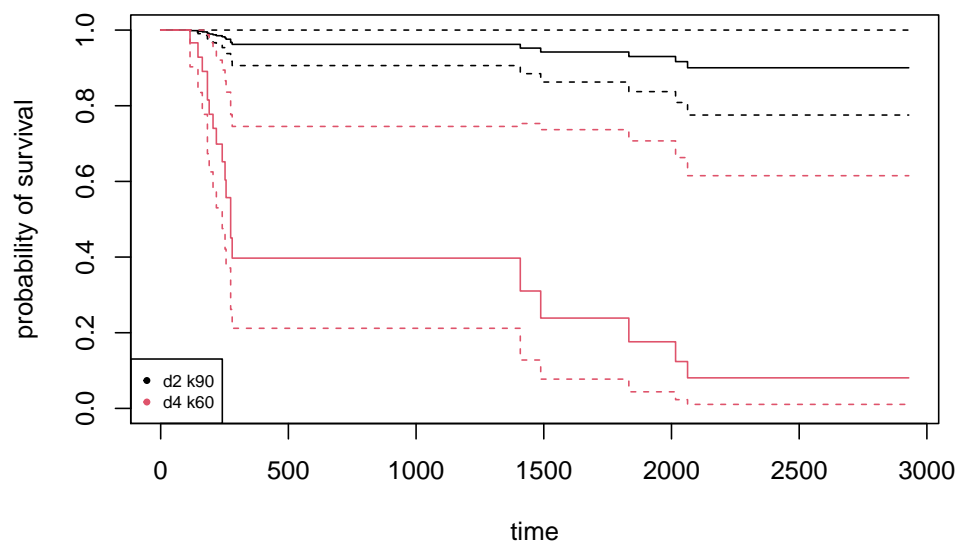
Коэффициент при переменной diagnosis положителен (0.837), чем выше значение диагноза, тем меньше вероятность дожития. pvalue = 0.013 < 0.05, значимое отличие от нуля.

При переменной karnovsky отрицателен (-0.05), чем ниже индекс Карновского, тем меньше вероятность дожития. pvalue = 0.07 > 0.05, влияние индекса karnovsky может быть случайным.

У пациентов с меньшей стадией и более высоким индексом Карновского показатели дожития выше.

Ниже приведено графическое обоснование результатов анализа.

```
plot(survfit(fitCS, newdata = data.frame(diagnosis=2, karnovsky=90)),
     mark.time=FALSE,col=1,conf.int=TRUE,
     xscale = 1, xlab = 'time', ylab = 'probability of survival',)
lines(survfit(fitCS, newdata = data.frame(diagnosis=4, karnovsky=60)),
     mark.time=FALSE, conf.int = TRUE, col=2,
     xscale = 1, xlab = 'time', ylab = 'probability of survival',)
legend('bottomleft', c('d2 k90', 'd4 k60'), col=seq(2), cex=0.7, pch=20)
```



На графике черным непрерывным цветом выделена линия дожития при диагнозе второй стадии рака и с высоким индексом Карновского. Линия расположена на верхней части графика, что говорит о высокой вероятности дожития.

Красным непрерывным цветом - диагноз четвертая стадия рака и низкий индекс Карновского. Линия расположена на нижней части графика - низкая вероятность дожития.