Анализ данных типа времени жизни

Н.П. Алексеева,

СПбГУ, математико-механический факультет

2020 г.

Данные типа времени жизни

Отказ

Событие, о котором известно, что оно происходит один раз и через некоторое случайное время $\tau > 0$.

Например, 1) "отказ" – это наступление рефлекса выведения молока, τ – время до его наступления; 2) "отказ" – это инфаркт, или инсульт, или смерть, τ – временной интервал до его наступления.

Исходные данные представляют собой времена ожидания отказа $\tau_i, i = 1, \ldots, n$ для каждого из индивидов. Рассмотрим $n = 22$ наблюдения τ – времени ремиссии (в неделях).

τ_i :	1	1	1	2	2	3	4	4	5	5	8
	8	8	8	11	11	12	12	15	17	22	23

Кривая дожития по полным данным

$F(t) = P\{\tau > t\}$ – вероятность того, что отказ наступит после момента времени t.

- d_t количество отказов в момент времени t,
- r_t – количество объектов, у которых отказ наступает не раньше момента t,
- $r_1 = n$.

$$\widehat{F}_t = \frac{r_t - d_t}{n} = \frac{r_{t+1}}{n}.$$

t	r_t	d_t	$r_t - d_t$	$\widehat{F}_t$
0	22	0	22	1
1	22	3	19	$\frac{19}{22}$
2	19	2	17	$\frac{17}{22}$
3	17	1	16	$\frac{16}{22}$

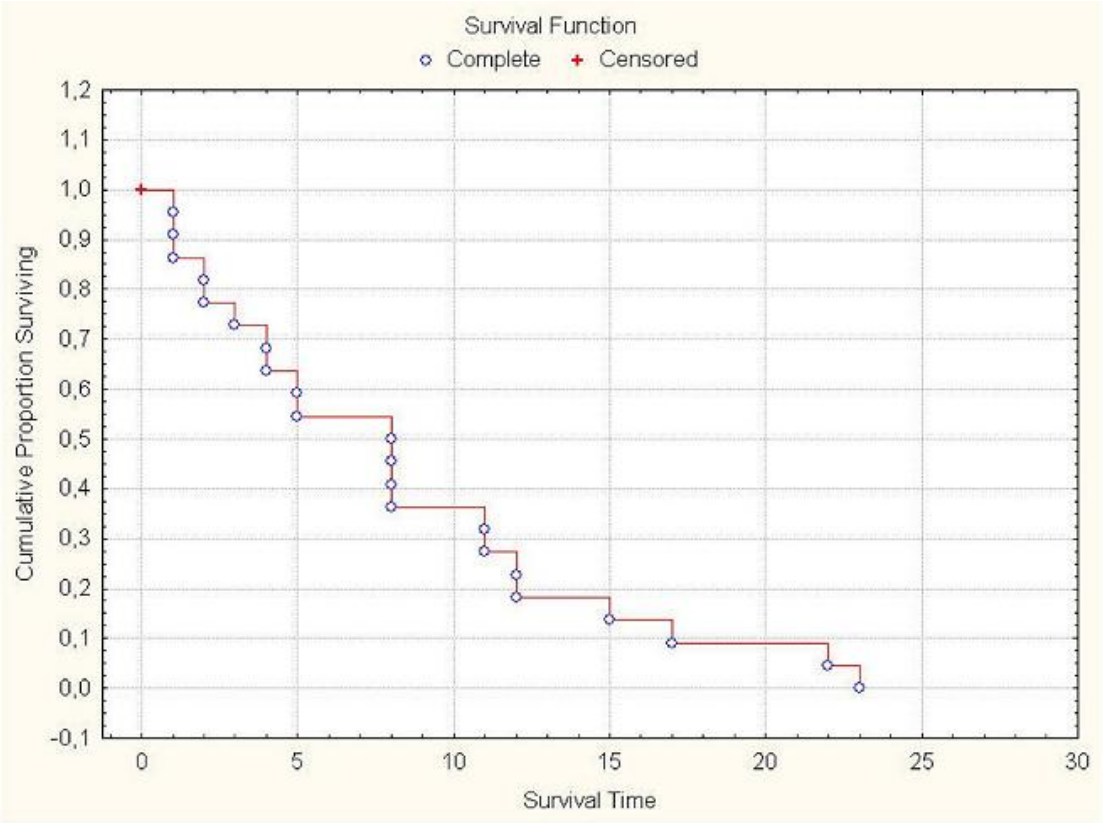


Figure: Кривая дожития по времени ремиссии (без цензурирования).

Требования к данным

Для фиксации данных требуется:

- четко установить начало отсчета,
- определить масштаб времени,
- определить понятие отказа.

Чаще всего стоит задача оценки вероятности дожития или сравниваются времена ожидания отказов в двух или нескольких выборках.

Функция надежности и плотность

Кривую дожития $\mathrm{F(t)} = \mathrm{P}\{\tau > \mathrm{t}\}$ иногда называют функцией надежности.
Ее производная с обратным знаком называется плотностью и обозначается через $\mathrm{f(t)}$.

$$\mathrm{f(t)} = -\mathrm{F}'(\mathrm{t}) = -\lim_{\Delta\to 0}\frac{\mathrm{F(t+\Delta)}-\mathrm{F(t)}}{\Delta} = \lim_{\Delta\to 0}\frac{\mathrm{F(t)}-\mathrm{F(t+\Delta)}}{\Delta} =$$
$$= \lim_{\Delta\to 0}\frac{\mathrm{P}\{\tau > \mathrm{t}\}-\mathrm{P}\{\tau > \mathrm{t}+\Delta\}}{\Delta} = \lim_{\Delta\to 0}\frac{\mathrm{P}\{\mathrm{t} < \tau \le \mathrm{t}+\Delta\}}{\Delta}$$

По смыслу, плотность $\mathrm{f(t)}$ равна вероятности гибели в момент t.

Функции интенсивности и риска

Условная вероятность гибели в момент времени t выражается функцией интенсивности h(t) вида

$$\text{Риск } h(t) = \lim_{\Delta \to 0} \frac{P\{t < \tau \le t + \Delta | \tau > t\}}{\Delta} = \lim_{\Delta \to 0} \frac{P\{t < \tau \le t + \Delta\}}{\Delta P\{\tau > t\}} = \frac{f(t)}{F(t)},$$

$$\text{или } \; h(t) = \frac{-F'(t)}{F(t)} = (-\ln F(t))' = H'(t)\,.$$

Функция риска: $H(t) = -\ln F(t)$, отсюда $F(t) = e^{-H(t)}$. Поскольку $F(0) = 1$, то $H(0) = 0$. Тогда учитывая, что $H'(t) = h(t)$, получаем:

$$H(t) = H(t) - H(0) = \int_0^t h(u)du \;\; \text{и} \; F(t) = e^{-\int_0^t h(u)du}.$$

Постоянный риск

Если риск (интенсивность) является постоянным $h(u) = \varrho$, то функция надежности убывает по экспоненте: $F(t) = e^{-\varrho t}$.

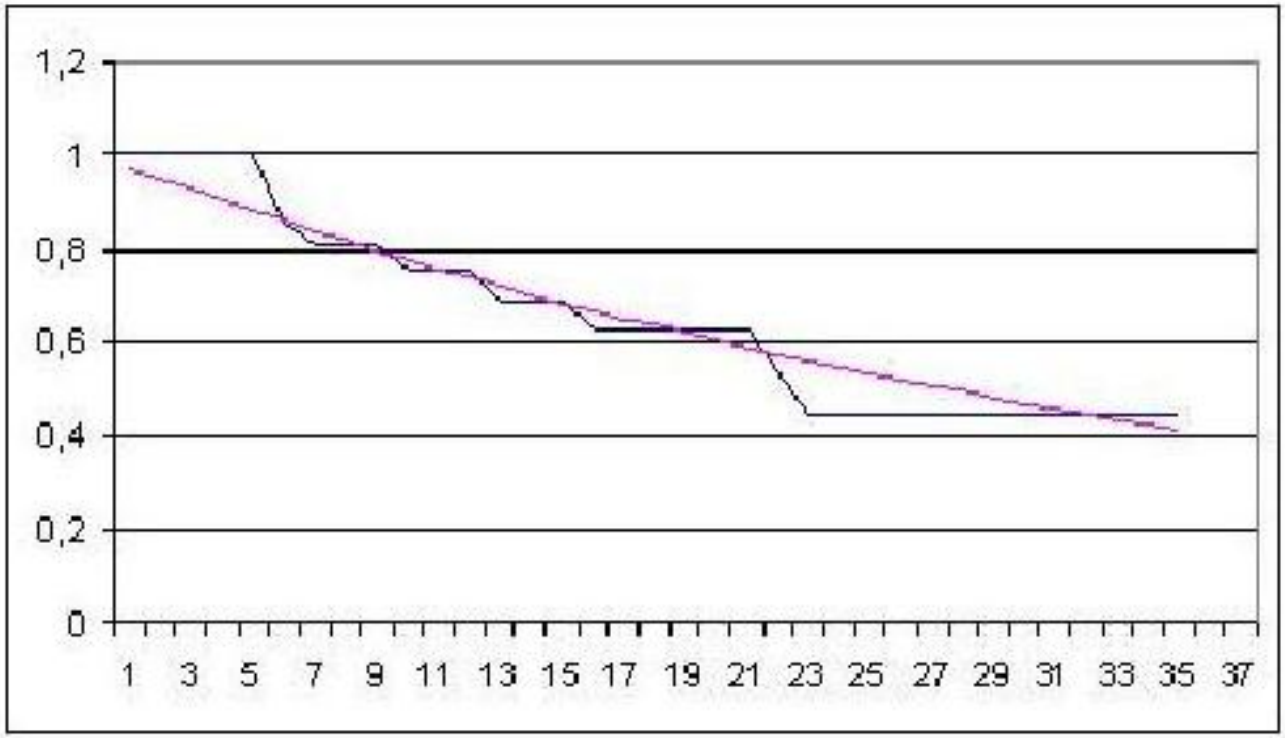


Figure: Кривые дожития времени ремиссии с цензурированием и ее экспоненциальная модель).

Цензурирование или неполное наблюдение за временем ожидания отказа.

Известно только то, что до некоторого времени t отказ еще не наступил.
Например, больной состоял на медицинском учете, а потом уехал в другой город, и о нем больше ничего не известно, или у животного фиксируемый рефлекс не наступил за установленное время эксперимента. При этом наблюдения записываются в виде пары признаков: в одном указывается или время наступления отказа, или время цензурирования, второй признак является индикатором цензурирования.

Цензурирование

Если обозначить через τ_i время ожидания отказа, c_i – момент цензурирования, то исходные данные можно представить в виде:

$$(X_i, V_i), \quad \text{где } X_i = \min(\tau_i, c_i), \quad i = 1, \ldots, n,$$

$$V_i = \begin{cases} 0, & \text{если } \tau_i \le c_i \quad (\text{отказ }), \\ 1, & \text{если } \tau_i > c_i \quad (\text{цензурирование}). \end{cases}$$

j	1	2	3	4	5	6	7	8	9	10
X_j	6	6	6	6	7	9	10	10	11	13
V_j	1	0	0	0	0	1	1	0	1	0

Оценка Каплана-Мейера

- d_t количество отказов в момент времени t,
- r_t – количество объектов, у которых отказ наступает не раньше момента t, $r_1 = n$,
- p_i вероятность того, что при условии дожить до момента i отказ в этот момент так и не наступил,
- $h_i = \frac{d_i}{r_i}$ условную вероятность погибнуть в момент i.

$$\widehat{p}_i = \frac{r_i - d_i}{r_i} = 1 - \frac{d_i}{r_i},$$

$$P\{\tau > t\} = P\{\tau > t-1\}P\{\tau > t | \tau > t-1\} = P\{\tau > t-1\}p_t$$

Так как $P\{\tau > 0\} = 1$, то $P\{\tau > t\} = p_1 p_2 \dots p_t$. Следовательно,

$$\widehat{P}\{\tau > t\} = \frac{r_1 - d_1}{r_1}\frac{r_2 - d_2}{r_2}\dots\frac{r_t - d_t}{r_t} = \prod_{i=1}^{t}(1 - h_i)\ . \qquad (1)$$

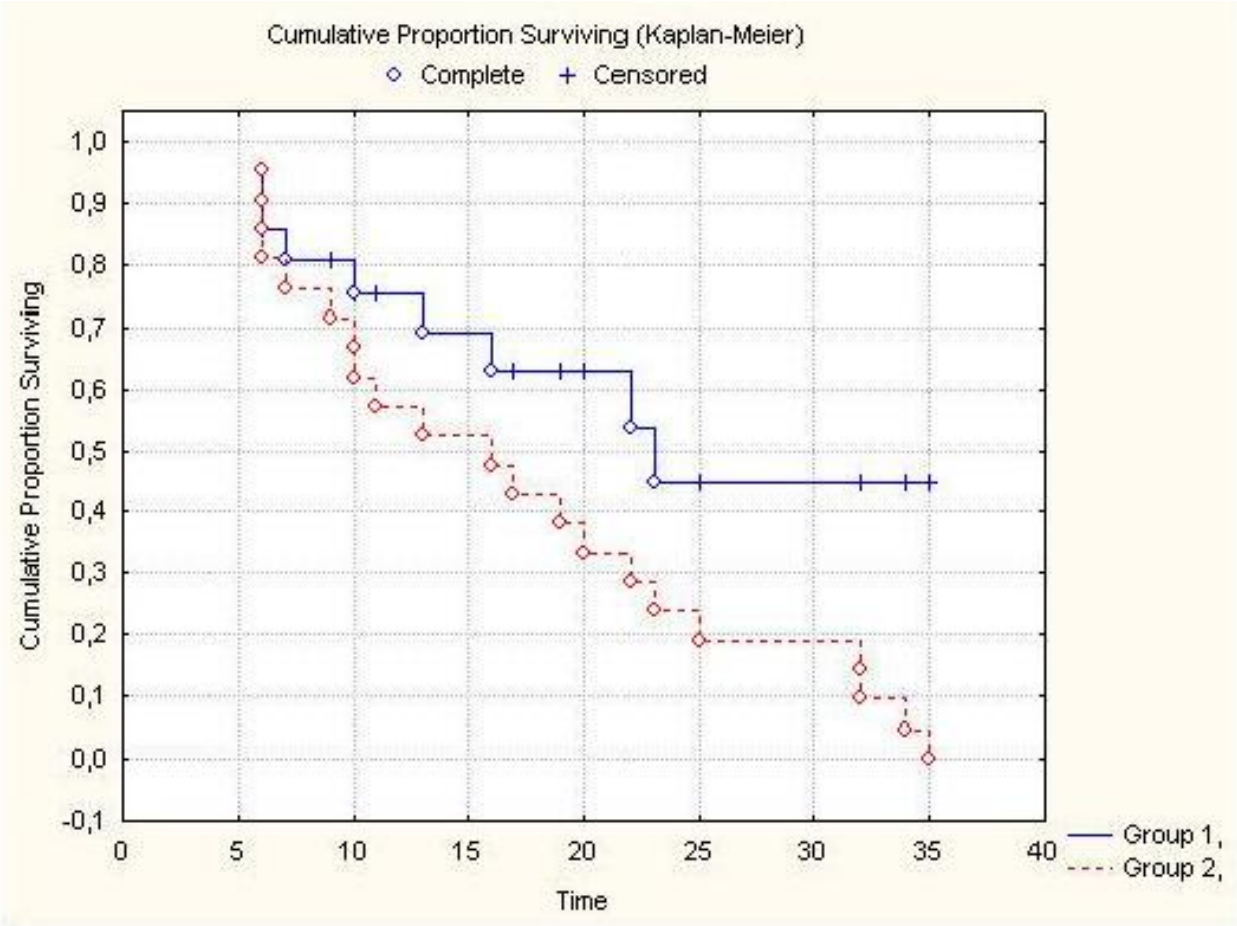


Figure: Кривые дожития времени ремиссии с цензурированием и без цензурирования.

В StatSoft анализ данных типа времени жизни осуществляется при помощи программы Statistics - Advanced Linear - Nonlinear Models - Survivel Analysis.

Выбираем Kaplan and Meier product – limit method. Данные должны быть представлены по крайней мере в двух столбцах. В одном из них должны быть указаны наработки до отказа или до цензурирования, обозначим его, например, через X, во втором – индикатор C цензурирования. В Variables указываем требуемые переменные. После активизации признака C появляются окна, в которых нужно указать коды цензурирования. Для получения графика, подобного рис.1, нужно во вкладке Kaplan – Meier plots выбрать Survivel times vs.cum. proption surviving.

Вычисление характеристик кривой дожития в R

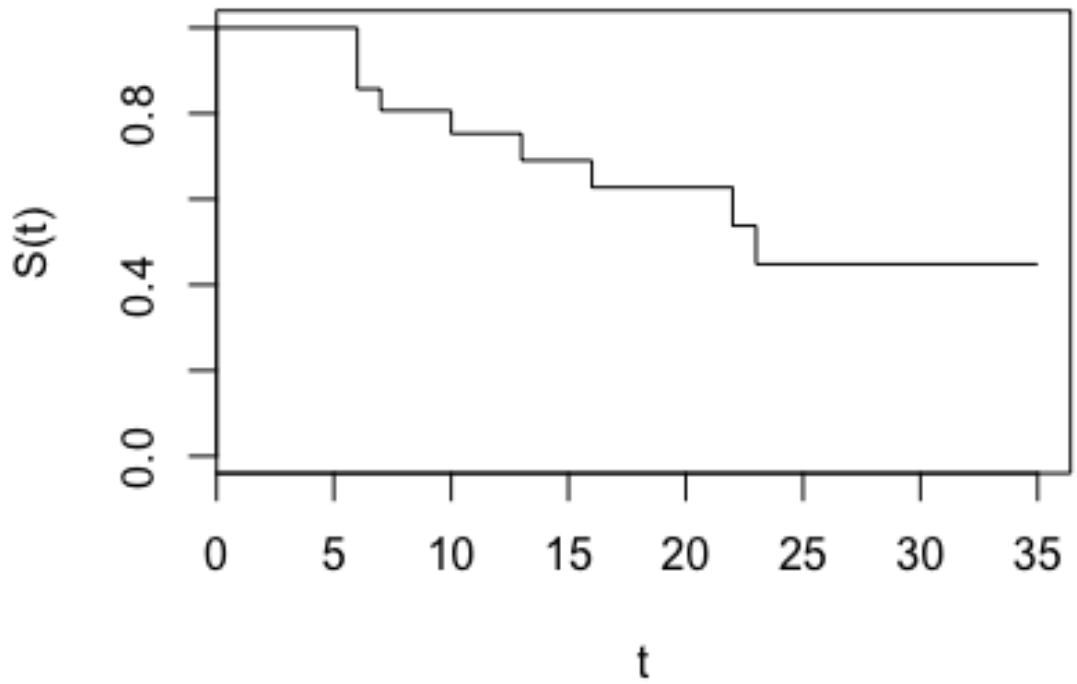
	time	c
1	6	1
2	6	0
3	6	0
4	6	0
5	7	0
6	9	1
⋮	⋮	⋮

```
fitKM < −survfit(Surv(time, event = 1 − c) ∼ 1, data = data)
print(fitKM, print.rmean = TRUE)
```

n	events	∗rmean	∗se(rmean)	median	0.95LCL	0.95UCL
21.00	9.00	23.29	2.83	23.00	16.00	NA

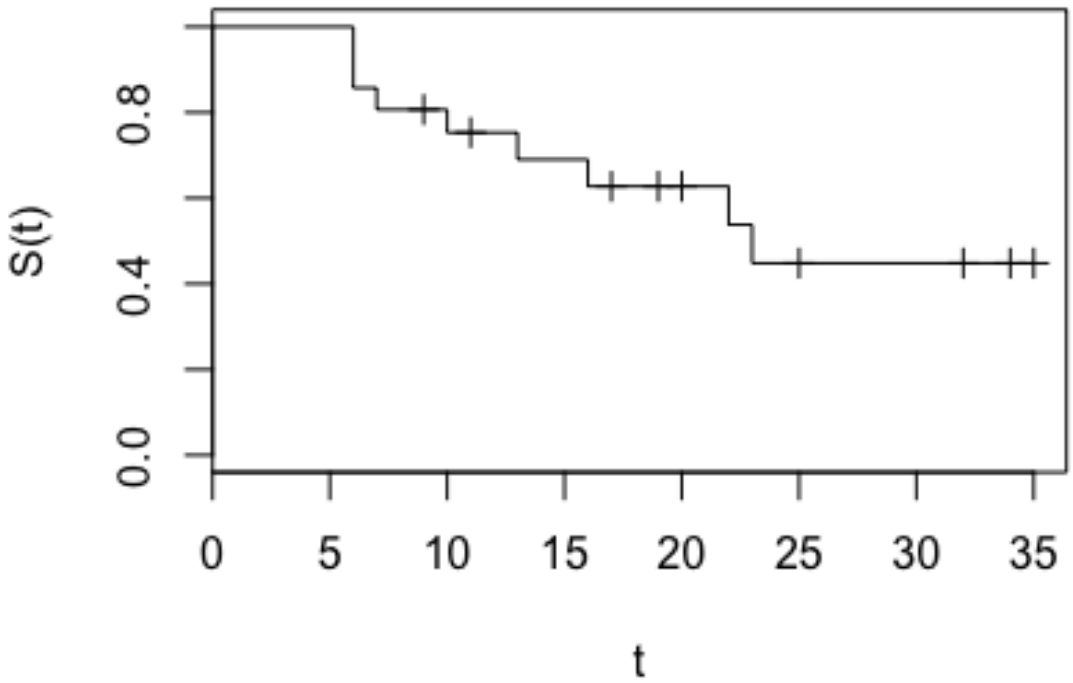
Построение кривой дожития в R

$$\text{fitKM} < -\text{survfit}(\text{Surv}(\text{time}, \text{event} = 1 - \text{c}) \sim \ 1, \text{data} = \text{data})$$
$$\text{plot}(\text{fitKM}, \text{mark.time} = \text{FALSE}, \text{col} = 1, \text{conf.int} = \text{FALSE}, \text{xlab} = \text{"t"}, \text{ylab} = \text{"S(t)"})$$



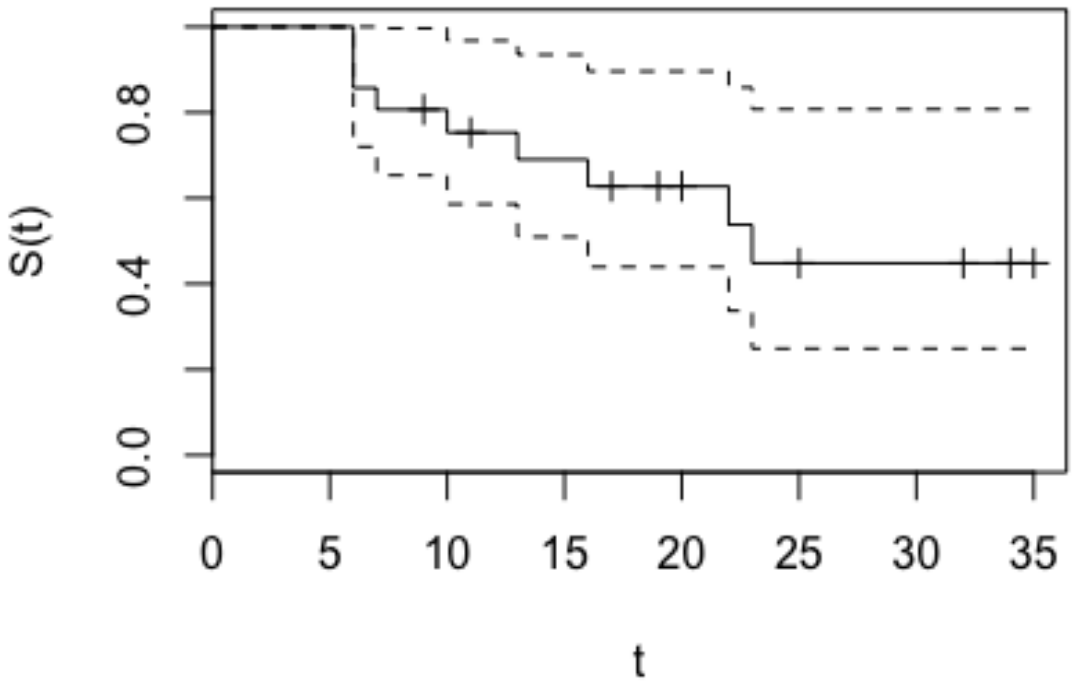
Построение кривой дожития в R с указанием цензурирования

```
fitKM <- survfit(Surv(time, event = 1 - c) ~ 1, data = data)
plot(fitKM, mark.time = TRUE, col = 1, conf.int = FALSE, xlab =
"t", ylab = "S(t)")
```



Построение кривой дожития в R с указанием цензурирования и с доверительными интервалами

```
fitKM <- survfit(Surv(time, event = 1 - c) ~ 1, data = data)
plot(fitKM, mark.time = TRUE, col = 1, conf.int = TRUE, xlab =
"t", ylab = "S(t)")
```



Оценка параметра постоянного риска

Функцией правдоподобия $L(x_1, \ldots, x_n)$ или вероятность появления данной выборки $x_i = \min(t_i, c_i),\ i = 1, \ldots, n$.

$$L(x_1, \ldots, x_n) = \prod_{t_i} f(t_i, \varrho) \prod_{c_i} F(c_i, \varrho)\,,$$

$$\ln L(x_1, \ldots, x_n) = \sum_{t_i} \ln f(t_i, \varrho) + \sum_{c_i} \ln F(c_i, \varrho)\,.$$

Поскольку $h(t) = \frac{f(t)}{F(t)}$ и $f(t) = h(t)F(t)$, то

$$\ln L(x_1, \ldots, x_n) = \sum_{t_i} \ln h(t_i, \varrho) + \sum_{t_i} \ln F(t_i, \varrho) + \sum_{c_i} \ln F(c_i, \varrho) =$$

$$= \sum_{t_i} \ln h(t_i, \varrho) + \sum_{x_i} \ln F(x_i, \varrho) = \sum_{t_i} \ln h(t_i, \varrho) - \sum_{x_i} H(x_i, \varrho)\,.$$

Если риск постоянный, то $h(t_i, \varrho) = \varrho$, $H(t) = \varrho t$. Следовательно,

$$\ln L(x_1, \ldots, x_n) = d \ln \varrho - \varrho \sum_{x_i} t\,,$$

где d означает общее число отказов, $\sum\limits_{x_i} t$ – суммарную наработку.

Оценка параметра постоянного риска

$$\ln \mathrm{L}(\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}) = \mathrm{d} \ln \varrho - \varrho \sum_{\mathrm{x}_\mathrm{i}} \mathrm{t}\,,$$

$$\frac{\partial \ln \mathrm{L}}{\partial \varrho} = \frac{\mathrm{d}}{\varrho} - \sum_{\mathrm{x}_\mathrm{i}} \mathrm{t} = 0 \iff \varrho = \frac{\mathrm{d}}{\sum\limits_{\mathrm{x}_\mathrm{i}} \mathrm{t}}\,.$$

Таким образом, параметр экспоненциальной модели кривой дожития вычисляется как отношения числа отказов к суммарной наработке. В нашем примере $\widehat{\varrho} = \frac{9}{359} = 0.025$.

Доверительные интервалы в случае постоянного риска

При отсутствии цензурирования $\frac{2d\varrho}{\hat{\varrho}}$ имеет распределение хи-квадрат с 2d степенями свободы. Действительно, суммарная наработка имеет гамма распределения

$$\sum_{i=1}^{d} x_i \sim \gamma(\varrho, d), \quad \frac{1}{d}\sum_{i=1}^{d} x_i \sim \gamma(d\varrho, d),$$

$$2d\varrho\left(\frac{1}{d}\sum_{i=1}^{d} x_i\right) = \frac{2d\varrho}{\hat{\varrho}} \sim \gamma(d\varrho\left(\frac{1}{2d\varrho}\right), d) = \gamma\left(\frac{1}{2}, \frac{2d}{2}\right) = \chi^2(2d)\,.$$

Доверительные интервалы для кривых дожития

При незначительном влиянии цензурированных данных используется формула для вычисления дисперсии

$$\mathbb{D}(\hat{F}_j) = \frac{\hat{F}_j^2(1-\hat{F}_j)}{r_{j+1}}.$$

Пусть n_0 число нецензурированных данных, $\hat{F}_j = \frac{r_{j+1}}{n_0}$, значение $n_0 = \frac{r_{j+1}}{\hat{F}_j}$ подставляем в выражение для дисперсии биномиальной частоты $\hat{F}_j$ в n_0 испытаниях:

$$\mathbb{D}(\hat{F}_j) = \frac{\hat{F}_j(1-\hat{F}_j)}{n_0} = \frac{\hat{F}_j^2(1-\hat{F}_j)}{r_{j+1}}.$$

Формула Гринвуда

$$D(\hat{F}_t) = \hat{F}_t^2 \sum_{j=1}^{t} \frac{d_j}{r_j(r_j - d_j)}$$

Если p_j ($\hat{p}_j = 1 - \frac{d_j}{r_j}$) - вероятность выжить, то число выживших $s_j = r_j - d_j$ имеет биномиальное распределение $\mathcal{B}(p_j, r_j)$. Обозначим через $F_t = \mathbb{E}\hat{F}_t$, где $\hat{F}_t = \prod\limits_{j=1}^{t} \frac{s_j}{r_j}$. Очевидно, что $\mathbb{E}\left(\frac{s_j}{r_j}\right) = p_j$, $F_t = \prod_{j=1}^{t} p_j$, следовательно,

$$\frac{\hat{F}_t}{F_t} = \prod_{j=1}^{t} \frac{s_j}{r_j p_j} = \prod_{j=1}^{t} \left(1 + \frac{s_j - r_j p_j}{r_j p_j}\right) \approx 1 + \sum_{j=1}^{t} \frac{s_j - r_j p_j}{r_j p_j}$$

$$\mathbb{D}\left(\frac{\hat{F}_j}{F_j}\right) = \sum_{j=1}^{t} \frac{r_j p_j (1 - p_j)}{(r_j p_j)^2} = \sum_{j=1}^{t} \frac{1 - p_j}{r_j p_j} \approx \sum_{j=1}^{t} \frac{1 - \hat{p}_j}{r_j \hat{p}_j},$$

$$\mathbb{D}\hat{F}_t = {F_t}^2 \sum_{j=1}^{t} \frac{1 - \hat{p}_j}{r_j \hat{p}_j} \approx \hat{F}_t^2 \sum_{j=1}^{t} \frac{1 - \hat{p}_j}{r_j \hat{p}_j} = \hat{F}_t^2 \sum_{j=1}^{t} \frac{d_j}{r_j(r_j - d_j)}.$$

Регрессионная модель Кокса

$$F(t) = e^{-\int_0^t h(u)du}$$

Рассматривается модель пропорциональных интенсивностей

$$h(t, z) = \psi(z, \beta)h_0(t) = e^{\beta^T z}h_0(t)\,,$$

в которой для любого объекта вектор поясняющих переменных z не зависит от времени. Будем считать, что базовая функция интенсивности $h_0(t)$ не известна. Обозначим через $\tau_1 < \tau_2 < \ldots < \tau_n$ упорядоченные моменты наработок до отказов у n объектов, T_j – номер объекта, отказавшего в момент τ_j, то есть

$$T_j = i \;\Leftrightarrow\; t_i = \tau_j\,.$$

Введем обозначение для множества риска – множества объектов, действующего непосредственно перед j-м отказом: $R(\tau_j) = \{i : t_i \geq \tau_j\}$, r_j – размер этого множества.
Пусть известна полная история до момента τ_j

$$H_{j-1} = \{\tau_1, \tau_2, \ldots, \tau_{j-1}; i_1, i_2, \ldots, i_{j-1}\}\,,$$

и один из объектов множества риска $R(\tau_j)$ отказал в момент τ_j.

Условная вероятность того, что это был i-й объект, равна

$$p_j(i|i_1, i_2, \ldots, i_{j-1}) = \frac{h_i(\tau_j)}{\sum\limits_{k \in R(\tau_j)} h_k(\tau_j)} = \frac{\psi(i)}{\sum\limits_{k \in R(\tau_j)} \psi(k)},$$

где $\psi(i) = \psi(z_i, \beta)$ есть множитель, влияющий на интенсивность i-го объекта. По цепному правилу для условных вероятностей получаем выражение для функции правдоподобия или вероятность истории H_n

$$p(i_1, \ldots, i_n) = \prod_{j=1}^{n} p_j(i_j|i_1, i_2, \ldots, i_{j-1}) = \prod_{j=1}^{n} \frac{\psi(i_j)}{\sum\limits_{k \in R(\tau_j)} \psi(k)}.$$

При наличии цензурирования произведение осуществляется по отказам:

$$lik = \prod_{i \in D} \frac{\psi(i)}{\sum\limits_{k \in R_i} \psi(k)},$$

где через D обозначено множество отказавших объектов, через R_i соответствующее i множество риска.

При $\psi(z;\beta) = e^{\beta^T z}$ для оценивания параметров вектора β логарифм функции правдоподобия

$$\ln(lik) = \sum_{i\in D}\left(\ln\psi(i) - \ln\sum_{k\in R(\tau_j)}\psi(k)\right) =$$

$$= \sum_{i\in D}\left(\beta^T z_i - \ln\sum_{k\in R(\tau_j)} e^{\beta^T z_k}\right)$$

дифференцируют по параметрам, приравнивают производные к нулю и решают систему уравнений:

$$\sum_{i\in D}(z_{i\nu} - A_{i\nu}(\beta)) = 0,\ \nu = 1,2,\ldots,K,$$

где $z_{i\nu}$ – значение ν-й из K компонент поясняющей переменной z для i-го объекта,

$$A_{i\nu}(\beta) = \frac{\sum\limits_{k\in R_i} z_{k\nu}e^{\beta^T z_k}}{\sum\limits_{k\in R_i} e^{\beta^T z_k}}$$

взвешенное среднее переменной $z_{i\nu}$ по соответствующему множеству риска.

В случае двух выборок имеется одна $(\mathrm{K} = 1)$ поясняющая переменная z, которая принимает два значения: 0, если объект принадлежит нулевой выборке, и 1, если принадлежит первой выборке.

$$\psi(\mathrm{i}) = \begin{cases} \mathrm{e}^{\beta}, & \text{i-ый объект из группы 1} \\ 1, & \text{i-ый объект из группы 0} \end{cases}$$

Обозначим через d_1 количество отказов в первой выборке, через $\mathrm{r}_{0\mathrm{i}}$ и $\mathrm{r}_{1\mathrm{i}}$ объемы множеств риска для каждой выборки в момент i. Логарифм функции правдоподобия равен

$$\ln(\mathrm{lik}) = \sum_{\mathrm{i}\in\mathrm{D}} \left(\ln\psi(\mathrm{i}) - \ln(\sum_{\mathrm{k}\in\mathrm{R_i}} \psi(\mathrm{k})) \right) = \beta \mathrm{d}_1 - \sum_{\mathrm{i}\in\mathrm{D}} \ln(\mathrm{r}_{1\mathrm{i}}\mathrm{e}^{\beta} + \mathrm{r}_{0\mathrm{i}}),$$

$$\frac{\mathrm{d}\ln(\mathrm{lik})}{\mathrm{d}\beta} = \mathrm{d}_1 - \sum_{\mathrm{i}\in\mathrm{D}} \frac{\mathrm{r}_{1\mathrm{i}}\mathrm{e}^{\beta}}{\mathrm{r}_{1\mathrm{i}}\mathrm{e}^{\beta} + \mathrm{r}_{0\mathrm{i}}} = 0.$$

Из последнего уравнения находится оценка параметра $\widehat{\beta}$.

Сравнение двух групп по кривым дожития

Для упрощения рассмотрим нецензурированные данные. Обозначим через d_{0i} и d_{1i} количество отказов в группе 0 и 1 в момент времени i, через $d_i = d_{0i} + d_{1i}$ число отказов в обеих группах. Пусть r_{0i} и r_{1i} – количество доживших до момента i.

время	группа 0		группа 1		всего		Peto Cox		Log-rank
i	r_{0i}	d_{0i}	r_{1i}	d_{1i}	r_i	d_i	$\frac{d_i r_{0i}}{r_i}$	$\frac{d_i r_{1i}}{r_i}$	$\frac{d_i r_{0i} r_{1i}(r_i - d_i)}{r_i^2(r_i-1)}$
1	3	1	3	0	6	1	0.5	0.5	0.250.25
2	2	1	3	0	5	1	0.4	0.6	0.24
3	1	0	3	1	4	1	0.25	0.75	0.1875
4	1	0	2	1	3	1	$\frac{1}{3}$	$\frac{2}{3}$	0.25
5	1	1	1	0	2	1	0.5	0.5	1.15
6	0	0	1	1	1	1	0	1	
сумма		3		3			1.98	4.02	1.15
		d_0		d_1			E_0	E_1	D

Сравнение двух групп по кривым дожития

Обозначим через E_0 и E_1 математические ожидания числа отказов

$$d_0 = \sum_{i=1}^{n} d_{0i}, \quad d_1 = \sum_{i=1}^{n} d_{1i}.$$

При $\beta = 0$ ожидаемые числа отказов как суммы математических ожиданий гипергеометрических случайных величин равны

$$\widehat{E}_0 = \sum_{i=1}^{n} \frac{d_i r_{0i}}{r_i}, \quad \widehat{E}_1 = \sum_{i=1}^{n} \frac{d_i r_{1i}}{r_i}.$$

$r_i = r_{0i} + r_{1i}$ – две группы по аналогии с красно-белыми шарами в урновой схеме. Умерло $d_i = d_{0i} + d_{1i}$, среднее число умерших в группе 0 равно $\frac{d_i r_{0i}}{r_i}$, а в группе 1 равно $\frac{d_i r_{1i}}{r_i}$. Среднее гипергеометрической величины равно $\frac{nM}{N}$, где N общее число шаров, M число красных и n число вынимаемых шаров.

Лог-ранговый критерий

Итак в каждый момент времени общее множество рисков имеет мощность $r_i = r_{0i} + r_{1i}$, всего отказов d_i. Случайное число отказов ξ из 0-группы подчиняется гипергеометрическому распределению:

$$P\{\xi = d_{0i}\} = \frac{C_{r_{0i}}^{d_{0i}} C_{r_{1i}}^{d_{1i}}}{C_{r_i}^{d_i}}$$

с математическим ожиданием $E_0 = \mathbb{E}\xi = \frac{d_i r_{0i}}{r_i}$ и дисперсией $D\xi = d_i \frac{r_{0i} r_{1i}(r_i - d_i)}{r_i^2 (r_i - 1)}$. Лог-ранговый критерий проверки однородности основан на асимптотической нормальности числа отказов d_0, которое в случае однородных выборок имеет нормальное распределение со средним E_0 и дисперсией

$$D = \sum_{i=1}^{n} d_i \frac{r_{0i} r_{1i}(r_i - d_i)}{r_i^2 (r_i - 1)}.$$

В случае однородных кривых дожития статистика лог-рангового критерия вида

$$\chi^2 = \frac{(d_0 - E_0)^2}{D} \sim \chi^2(1),$$

В частности, $\chi^2 = 0.904$, $p = 0.34$, выборки однородны по дожитию.

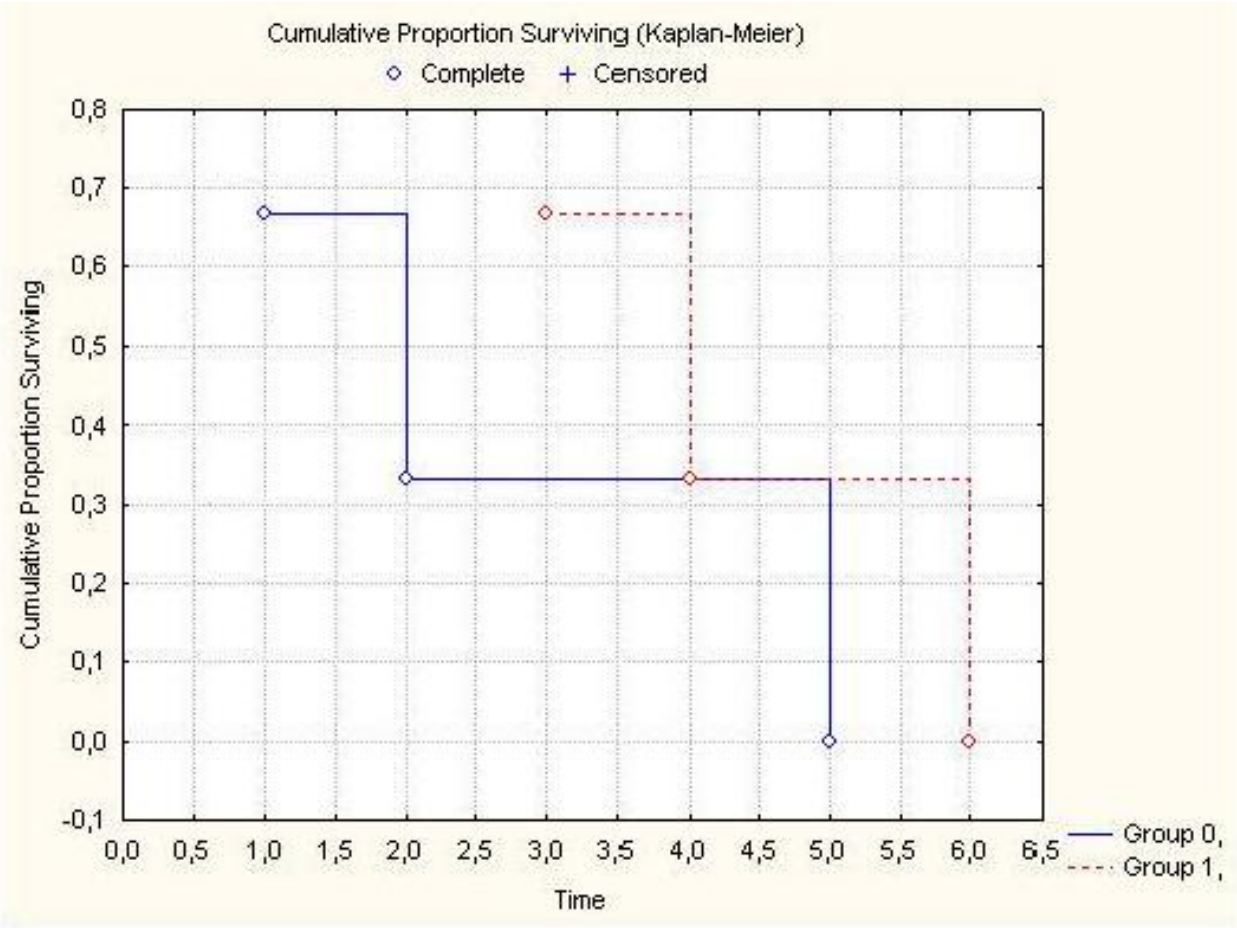


Figure: Сравнение кривых дожития, p = 0.34.

Пример счета в R

	time	dead	gr
1	15.00	1.00	1.00
2	18.00	1.00	1.00
3	19.00	1.00	1.00
4	19.00	1.00	1.00
5	20.00	1.00	1.00
6	16.00	0.00	2.00
7	18.00	0.00	2.00
8	20.00	0.00	2.00
9	23.00	1.00	2.00
10	24.00	0.00	2.00

Характеристики кривых

$sft < -survfit(Surv(time, event = dead) \sim gr, data = data);$
$print(sft, print.rmean = TRUE)$

	n	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
gr=1	5	5	18.2	0.769	19	18	NA
gr=2	5	1	22.0	0.000	23	23	NA

Лог-ранговый критерий

$survdiff(Surv(time, event = dead) \sim gr, data = data, rho = 0)$

$Chisq = 7 \text{ on } 1 \text{ degrees of freedom}, \ p = 0.00836$

Критерий Гехана-Вилкоксона

$with(data, gehan.test(time, dead, gr))$

$statistic = -2.368057, \ p - value = 0.0178818$

Построение графиков

plot(sft,col=seq(2))

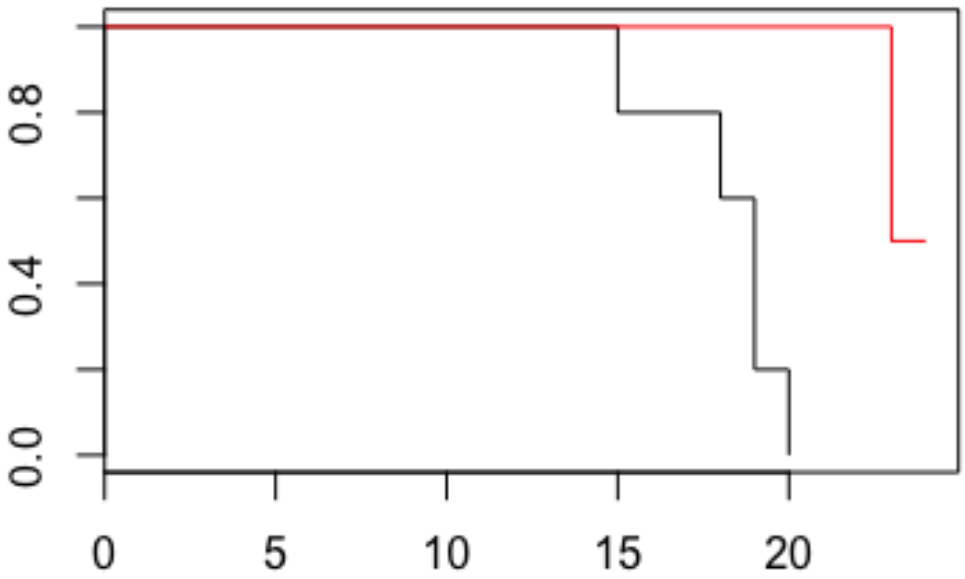


Figure:

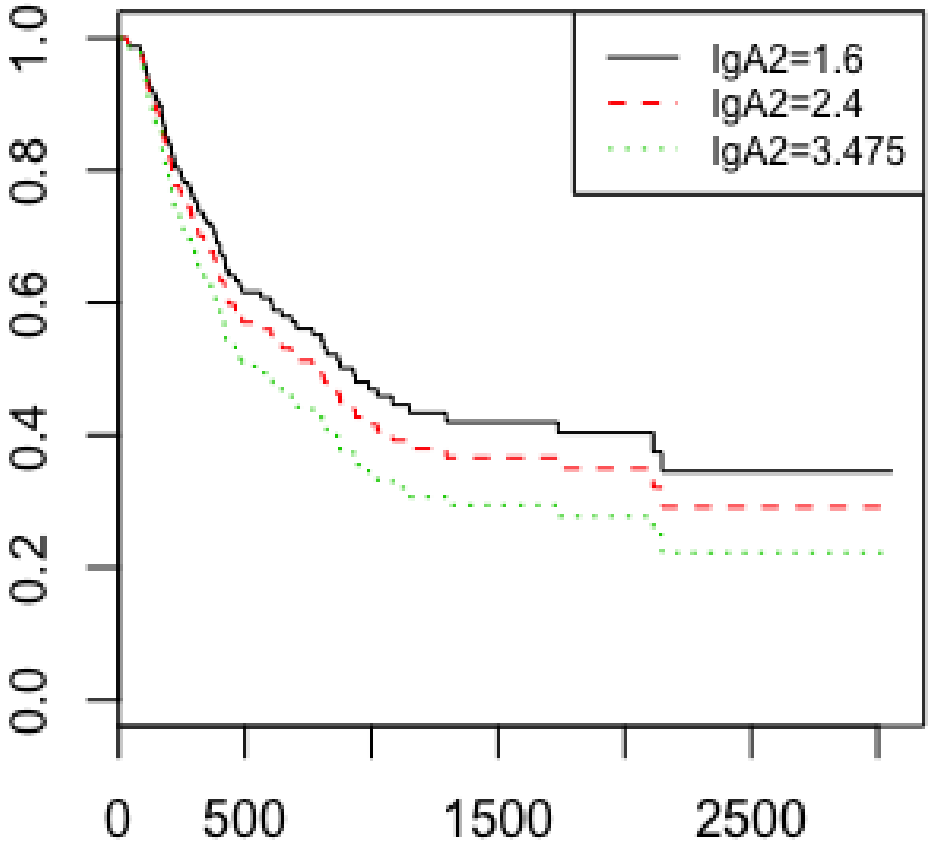
Регрессионная модель Кокса в R

$$\text{fit} < -\text{coxph}(\text{Surv}(\text{Duration.of.Reanimation}, \text{Outcome}) \sim \text{PLT1} + \text{PLT2}, \text{data} = \text{dat})$$

$$\text{summary}(\text{fit})$$

Регрессионная модель Кокса в R

```
points <- read.csv("points5.csv", sep = ";", dec = ",")
attach(points)
fitCox <- coxph(Surv(dni.zhil, 1 - zhiv) ~  IgA0 + IgA1 + IgA2, points)
qq <- quantile(IgA2, na.rm = TRUE)
df <- data.frame(rep(mean(IgA0, na.rm = TRUE), 3),
rep(mean(IgA1, na.rm = TRUE), 3), qq[c(2 : 4)])
colnames(df) <- c("IgA0","IgA1","IgA2")
fit <- survfit(fitCox, newdata = df)
plot(fit, conf.int = FALSE, lty = seq(3), col = seq(3))
legend('topright', paste("IgA2", qq[c(2 : 4)], sep = " = "), lty = seq(3),
col = seq(3), cex = 0.75)
detach(points)
```


IgA2=1.6
IgA2=2.4
IgA2=3.475
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1500
2500