P8131 HW9

Renjie Wei rw2844

Problem 1

Determine the survival and density functions for a continuous survival time variable with hazard function

$$h(x) = \frac{2x}{(1+x^2)}$$

Answers

The cumulative harzard funciton is

$$H(t) = \int_0^t h(x)dx$$
$$= \int_0^t \frac{2x}{(1+x^2)}$$
$$= \log(1+t^2)$$

The survival function is

$$S(t) = \exp(-H(t))$$
$$= \exp(-\log(1+t^2))$$
$$= \frac{1}{1+t^2}$$

$$F(t) = 1 - S(t)$$
$$= \frac{t^2}{1+t^2}$$

Hence, the density function is

$$f(t) = F'(t)$$
$$= (1 - \frac{1}{1+t^2})'$$
$$= \frac{2t}{(1+t^2)^2}$$

Problem 2

a

Find the Kaplan-Meier estimate of the survival function

Answers

The Kaplan-Meier estimator of survival function is

$$\hat{S}(t) = \prod_{i=1}^{k}(1 - \hat{\lambda}_i) = \prod_{i=1}^{k}(1 - \frac{d_i}{n_i})$$

b

Find the Nelson-Aalen estimate of the cumulative hazard function

Answers

The Nelson-Aalen estimate of the cumulative hazard function is

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

c

Find the Fleming-Harrington estimate of the survival function

Answers

The Fleming-Harrington estimate of the survival function $S(t)$ is

$$\exp(-\tilde{H}(t))$$

The three estimator above can be shown in the table

Table 1: Table of the survival data 1,2,2,4+,5+,6,7+,8+,9+,10+

t_i	n_i	d_i	c_i	$\hat{\lambda}_i$	$\hat{S}(t)$	$\tilde{H}(t)$	$\exp(-\tilde{H}(t))$
1	10	1	0	0.100	0.90	0.100	0.905
2	9	2	0	0.222	0.70	0.322	0.725
4	7	0	1	0.000	0.70	0.322	0.725
5	6	0	1	0.000	0.70	0.322	0.725
6	5	1	0	0.200	0.56	0.522	0.593
7	4	0	1	0.000	0.56	0.522	0.593
8	3	0	1	0.000	0.56	0.522	0.593
9	2	0	1	0.000	0.56	0.522	0.593
10	1	0	1	0.000	0.56	0.522	0.593

Note:

$\hat{S}(t)$: Kaplan-Meier estimator of survival function

$\tilde{H}(t)$: Nelson-Aalen estimate of the cumulative hazard function

$\exp(-\tilde{H}(t))$: Fleming-Harrington estimate of the survival function

Problem 3

Use the tongue data in the R package KMsurv. For each tumor type (aneuploidy and diploid), plot the Kaplan-Meier curve of survival function and its pointwise 95% confidence intervals (using the log transformation). What are the estimated 1-year survival rate and 95% CI?

Answers

The Kaplan-Meier curve of aneuploid tumor survival function and its pointwise 95% confidence intervals
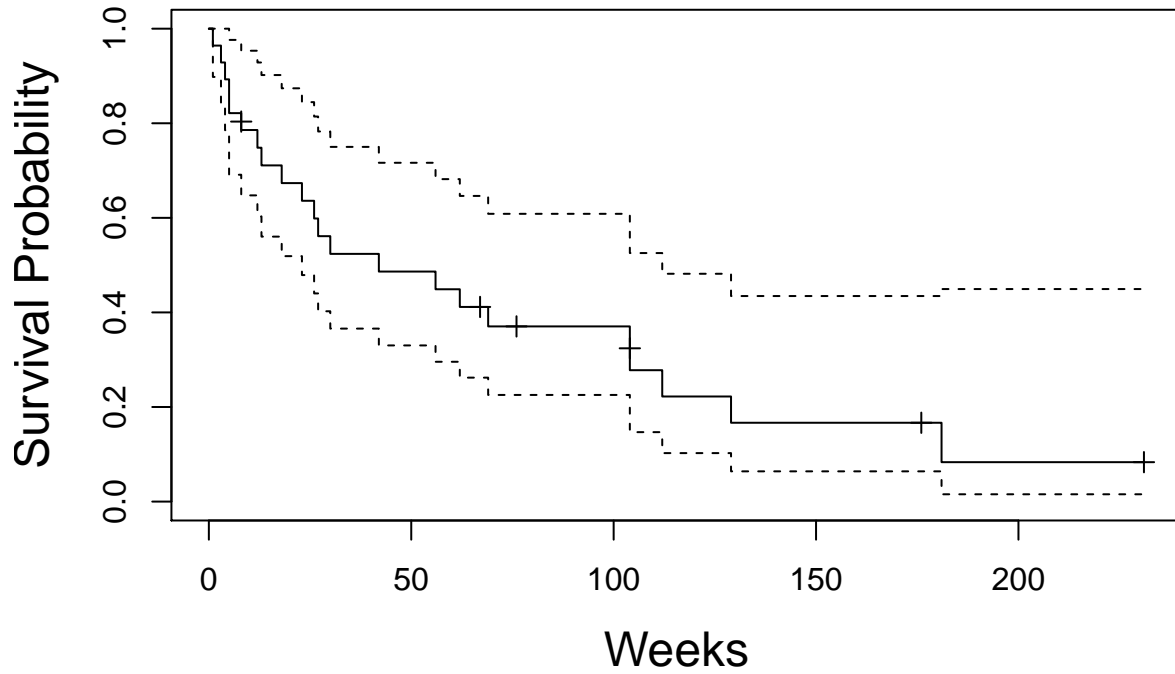
K–M curve of aneuploid tumor
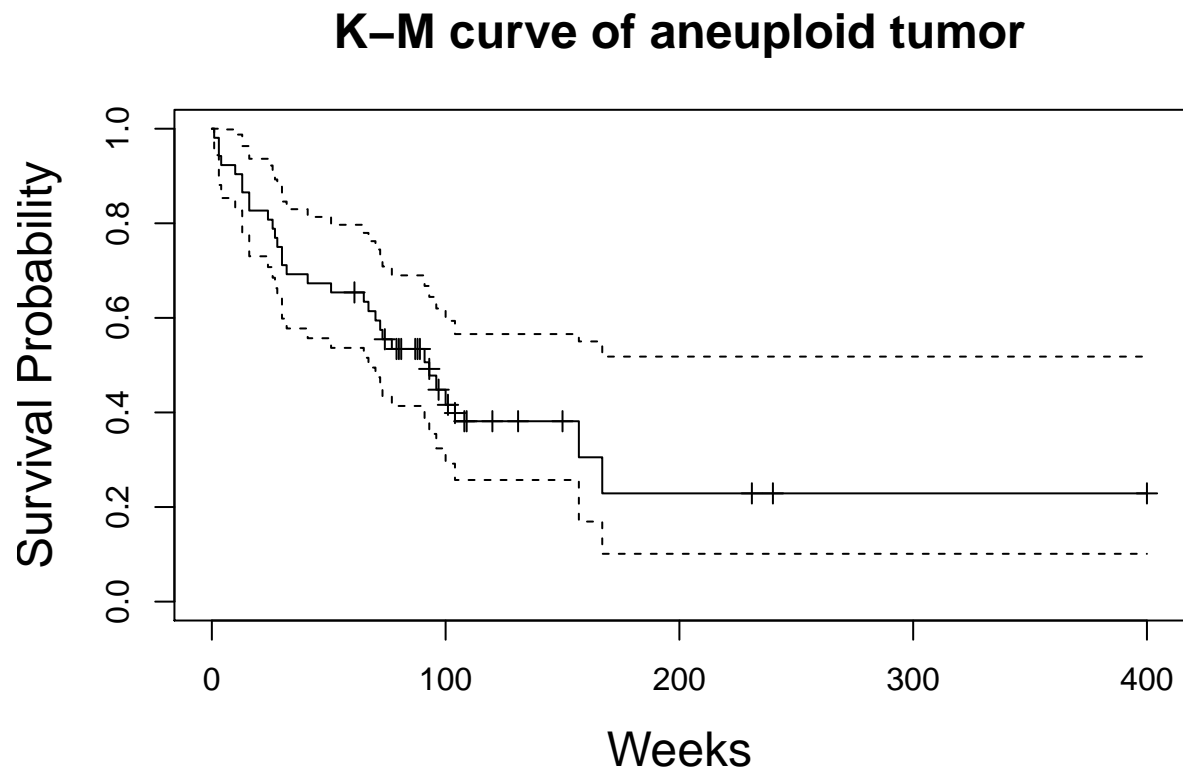
The Kaplan-Meier curve of diploid tumor survival function and its pointwise 95% confidence intervals

Table 2: Estimated 1-year survival rate and 95% CI of different tumor types

Tumor Type	1-year survival rate	lower 95% CI	upper 95% CI
aneuploid tumor	0.654	0.537	0.797
diploid tumor	0.486	0.330	0.716

K–M curve of diploid tumor

The table below summarizes the estimated 1-year survival rate and 95% CI.

Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = F, message = F)
q2_tab <- data.frame(
    ti = c(1,2,4,5,6,7,8,9,10),
    ni = c(10,9,7,6,5,4,3,2,1),
    di = c(1,2,0,0,1,0,0,0,0),
    ci = c(0,0,1,1,0,1,1,1,1)
)


q2_tab$lambda <- ifelse(q2_tab$ni == 0, 0, q2_tab$di/q2_tab$ni)


for (i in 1:nrow(q2_tab)){
    q2_tab$St[i] = prod(1 - q2_tab$lambda[1:i])
}
```

```r
for (i in 1:nrow(q2_tab)){
    q2_tab$Ht[i] = sum(q2_tab$lambda[1:i])
}

q2_tab$expHt <- exp(-q2_tab$Ht)
library(tidyverse)
names(q2_tab) = c(
                "$t_i$",
                "$n_i$",
                "$d_i$",
                "$c_i$",
                "$\\hat{\\lambda_i}$",
                "$\\hat{S(t)}$",
                "$\\tilde{H}(t)$",
                "$\\exp(-\\tilde{H}(t))$"
            )
q2_tab %>% knitr::kable(
            digits = 3,
            escape = FALSE,
            row.names = FALSE,
            caption = "Table of the survival data 1,2,2,4+,5+,6,7+,8+,9+,10+"
            ) %>% kableExtra::footnote(general =
                                    c(
                        "$\\\\\\hat{S(t)}$: Kaplan-Meier estimator of survival function ",
                        "$\\\\\\tilde{H(t)}$: Nelson-Aalen estimate of the cumulative hazard functio
                        "$\\\\\\exp(-\\\\\\tilde{H(t)})$: Fleming-Harrington estimate of the survival
                            ),
                            escape = F)
require(KMsurv)
require(survival)
data(tongue)
aneuploid = tongue %>% filter(type == 1) %>% mutate(survt = Surv(time, delta, type = "right"))
diploid = tongue %>% filter(type == 2) %>% mutate(survt = Surv(time, delta, type = "right"))
KM_aneuploid = survfit(Surv(time, delta) ~ 1, data = aneuploid, conf.type = "log")
plot(KM_aneuploid, conf.int = TRUE, mark.time = TRUE, xlab = "Weeks", ylab = "Survival Probability", ma
KM_diploid = survfit(Surv(time, delta) ~ 1, data = diploid, conf.type = "log")
plot(KM_diploid, conf.int = TRUE, mark.time = TRUE, xlab = "Weeks", ylab = "Survival Probability", main
sum_an <- summary(KM_aneuploid, time = c(52))
sum_di <- summary(KM_diploid, time = c(52))
one_year_surv <- data.frame(
    tumor = c("aneuploid tumor", "diploid tumor"),
    rate = c(sum_an$surv, sum_di$surv),
    lower = c(sum_an$lower, sum_di$lower),
    upper = c(sum_an$upper, sum_di$upper)
)

one_year_surv %>% knitr::kable(digits = 3,
                            caption = "Estimated 1-year survival rate and $95\\%$ CI of different tum
                            escape = F,
                            row.names = F,
                            col.names = c(
                                "Tumor Type",
                                "1-year survival rate",
```

```
        "lower $95\\\%$ CI",
        "upper $95\\\%$ CI"
),
format = "latex"
)
```