P8110: Applied Regression II

Homework #1 Solution [10 points]

NOTE: Please do all calculations by hand for this assignment. SHOW ALL WORK.

1. Listed below are values of survival time in years for 10 patients from the WHAS100 study. I have changed some event data to censored data for homework practice. The censored observations are denoted by a "+" as a superscript. The data are $0.4, 1.2, 1.2^+, 3.4^+, 4.3, 5.0, 5.0, 5.0^+, 6.1^+, 7.1$.

   (a) Using Kaplan-Meier method to estimate the survival function for these 10 patients (show a table with 6 columns: j, t_j, time interval, n_j, d_j, $1 - \frac{d_j}{n_j}$, and $\hat{S}(t)$). [4 points]

Table 1: Kaplan-Meier estimates of the survival function

j	t_j	For t in	n_j	d_j	$1-\frac{d_j}{n_j}$	$\hat{S}(t)$
0	0.0	[0,0.4)	10	0	1	1
1	0.4	[0.4,1.2)	10	1	0.9	0.9
2	1.2	[1.2,4.3)	9	1	0.889	0.8
3	4.3	[4.3,5.0)	6	1	0.833	0.667
4	5.0	[5.0,7.1)	5	2	0.6	0.4
5	7.1	[7.1,∞)	1	1	0	0

   (b) Sketch the Kaplan-Meier curve. [2 points]

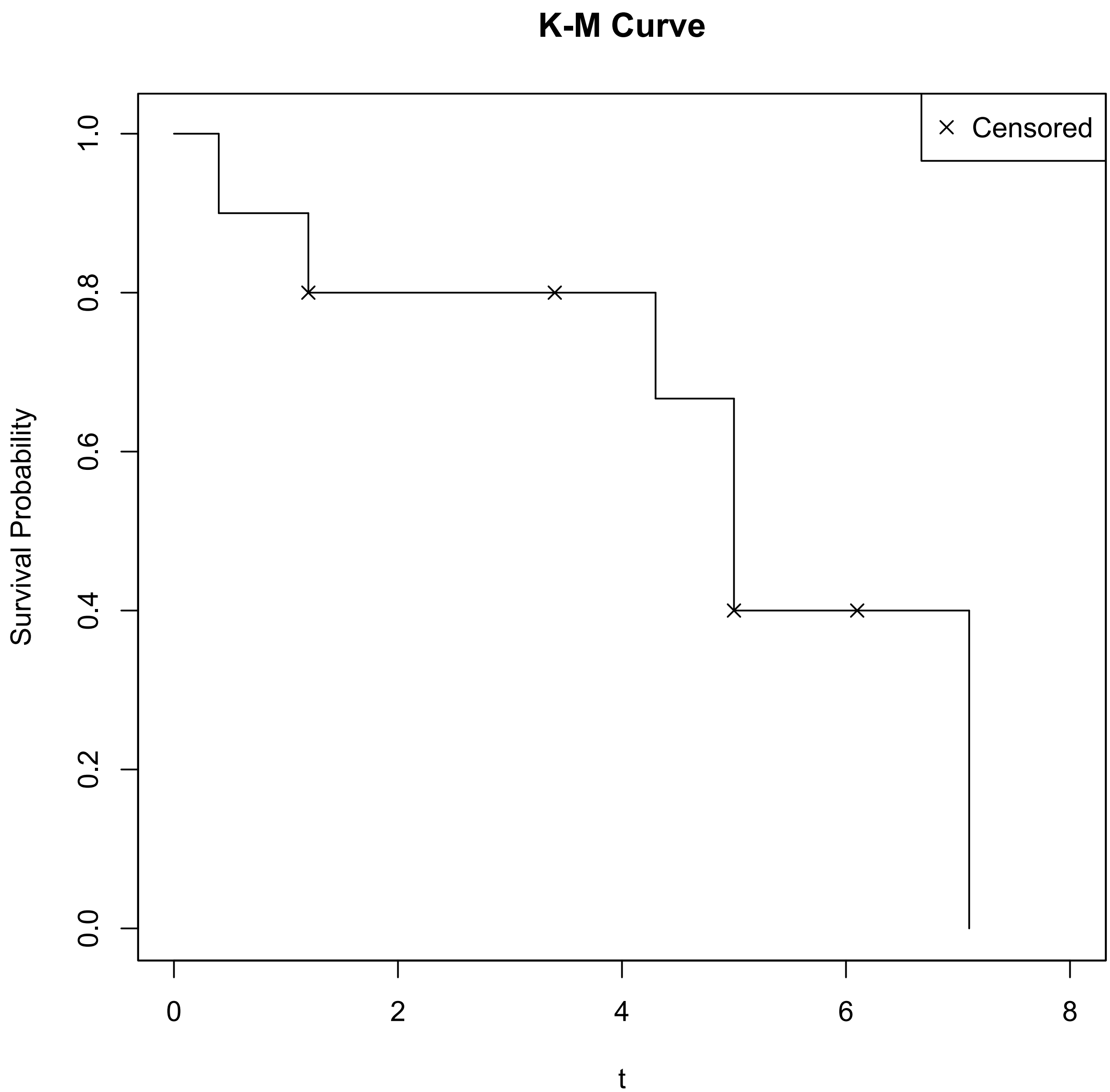

Figure 1: The survival curve

(c) Give an estimate of the probability of survival at 5 years. Provide its 95% confidence interval. Interpret both the estimate and 95% CI. [4 points]

Given that 5 belongs to the time interval [5.0,7.1), the probability of sur-

vival at 5 years is $\hat{S}(5) = 0.4$.

Transform $\hat{S}(t)$ *to* $\log[-\log\{\hat{S}(t)\}]$, *then* $\log[-\log\{\hat{S}(5)\}] = -0.087$.

Its variance can be derived as

$$\begin{aligned}
\widehat{var}(\log[-\log\{\hat{S}(5)\}]) &= \frac{1}{(\log(\hat{S}(5)))^2}\left(\sum_{t_j \le 5} \frac{d_j}{n_j(n_j - d_j)}\right) \\
&= \frac{1}{(\log(0.4))^2}\left(\frac{1}{10 \times 9} + \frac{1}{9 \times 8} + \frac{1}{6 \times 5} + \frac{2}{5 \times 3}\right) \\
&= 0.228
\end{aligned}$$

The 95% confidence interval for $\log[-\log\{S(5)\}]$ *is*

$$\begin{aligned}
\log(-\log(\hat{S}(5))) \pm 1.96\sqrt{\widehat{var}(\log(-\log(\hat{S}(5))))} &= -0.087 \pm 1.96 \times 0.478 \\
&= [-1.024, 0.849]
\end{aligned}$$

Therefore 95% confidence interval for $S(5)$ *is*

$$[\exp(-\exp(0.849)), \exp(-\exp(-1.024))] = [0.097, 0.698]$$

Interpretation: The probability that a patient can survive more than 5 years is estimated to be 0.4 (95% CI: 0.097 – 0.698).