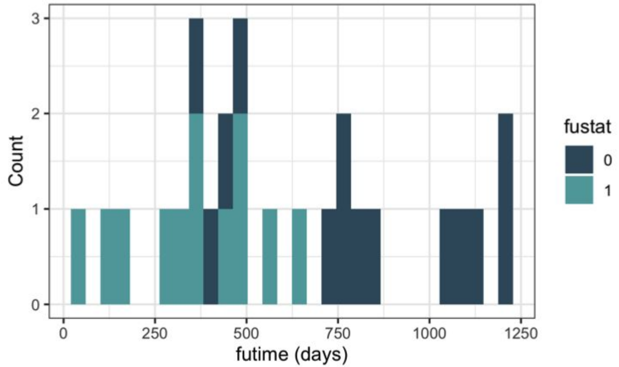
Chapter 11: Survival Data and the Kaplan-Meier Curve

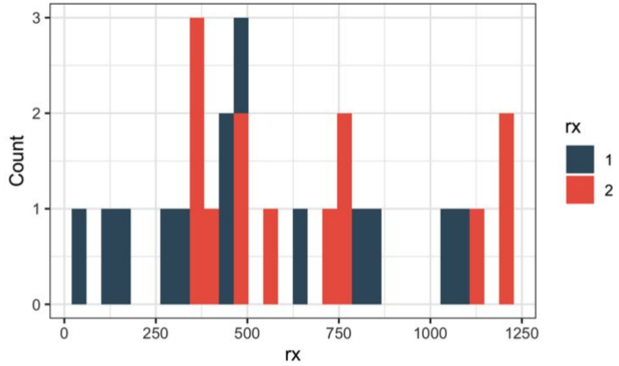
Modern Clinical Data Science
Chapter Guides
Bethany Percha, Instructor

How to Use this Guide

- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 15 min, max) while following along in the notes

- `futime`: The number of days from enrollment in the study until death or censoring, whichever came first
- `fustat`: An indicator of death (1) or censoring (0)
- `age`: The patient's age in years at the time of treatment administration
- `resid.ds`: Residual disease present at the time of treatment administration (1 = no, 2 = yes)
- `rx`: Treatment group (1 = cyclophosphamide, 2 = cyclophosphamide + adriamycin)
- `ecog.ps`: A measure of performance score or functional status at the time of treatment administration, using the Eastern Cooperative Oncology Group's (ECOG) scale. It ranges from 0 (fully functional) to 4 (completely disabled). Level 4 subjects are usually considered too ill to enter a randomized trial such as this. The patients in this dataset are all at Levels 1 and 2.



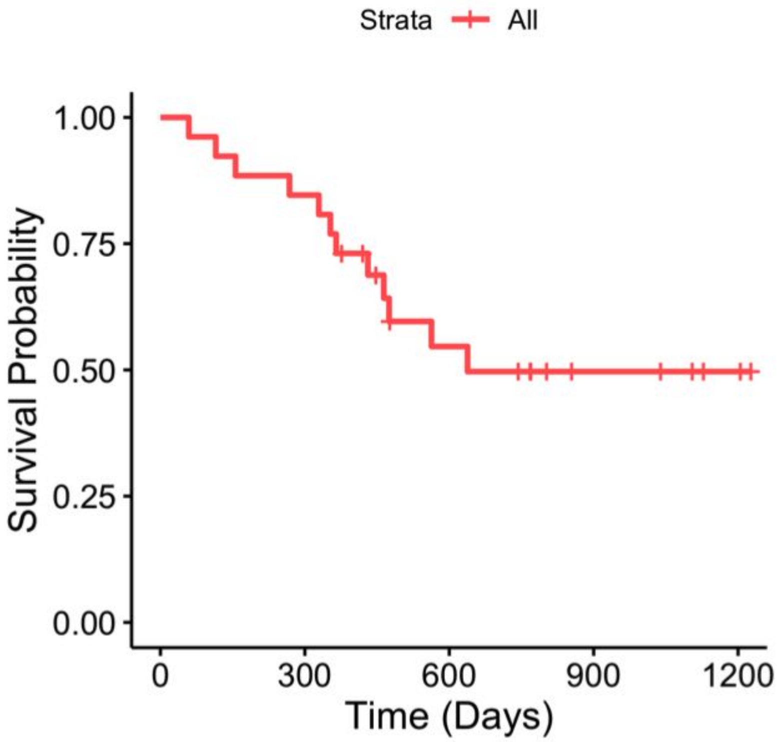


Right censoring: A situation that arises when the event of interest has not occurred by the end of the follow-up period. This may be because (a) the study itself ends, (b) a patient is lost to follow-up during the study period, or (c) a patient experiences a different event that makes further follow-up impossible[2].

Survival: Also called the **survival function** or **survival probability** and abbreviated $S(t)$, this is the probability that an individual survives to time t (i.e., does not experience the event by time t).

Hazard: Usually denoted by $h(t)$ or $\lambda(t)$, this is the probability that an individual who has not yet experienced the event at time t experiences it at that exact time. In other words, it is the instantaneous event rate for an individual who has already survived to time t.

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}$$



Question 11.1

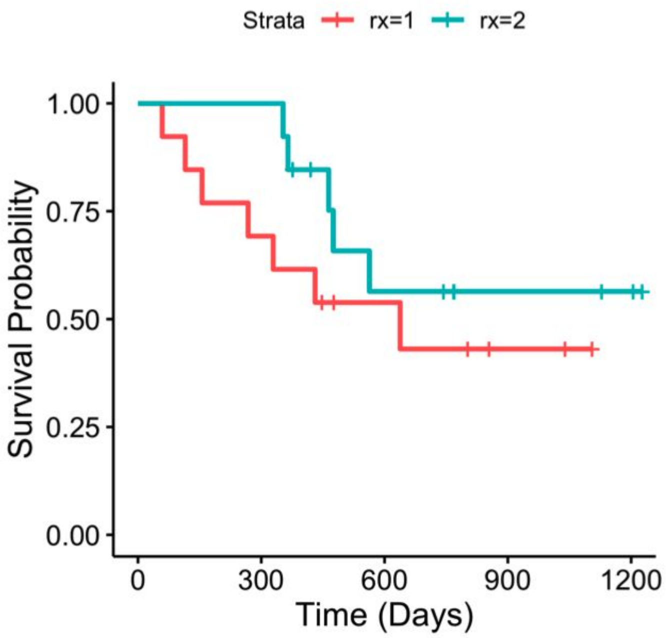
Here are the raw data from treatment group 1 of the `ovarian` dataset. Using these data, fill in the remaining cells of the table below.

	rx	futime	fustat
1	1	59	1
2	1	115	1
3	1	156	1
4	1	268	1
5	1	329	1
6	1	431	1
7	1	448	0
8	1	477	0
9	1	638	1
10	1	803	0
11	1	855	0
12	1	1040	0
13	1	1106	0

j	t_j	n_j	d_j	$\hat{S}(t_j)$	Calculation
0	0	13	0	1.000	$\frac{13-0}{13}$
1	59	13	1	0.923	$\hat{S}(t_0)\left(\frac{13-1}{13}\right)$
2	115	12	1	0.846	$\hat{S}(t_1)\left(\frac{12-1}{12}\right)$
3	156	11	1	0.769	$\hat{S}(t_2)\left(\frac{11-1}{11}\right)$
4	268	10	1	0.692	$\hat{S}(t_3)\left(\frac{10-1}{10}\right)$
5	329	9	1	0.615	$\hat{S}(t_4)\left(\frac{9-1}{9}\right)$
6	431	8	1	0.538	$\hat{S}(t_5)\left(\frac{8-1}{8}\right)$
7	448	7	0	0.538	$\hat{S}(t_6)\left(\frac{7-0}{7}\right)$
8	477	6	0	0.538	$\hat{S}(t_7)\left(\frac{6-0}{6}\right)$
9	638	5	1	0.431	$\hat{S}(t_8)\left(\frac{5-1}{5}\right)$
10	803	4	0	0.431	$\hat{S}(t_9)\left(\frac{4-0}{4}\right)$
11	855	3	0	0.431	$\hat{S}(t_10)\left(\frac{3-0}{3}\right)$
12	1040	2	0	0.431	$\hat{S}(t_11)\left(\frac{2-0}{2}\right)$
13	1106	1	0	0.431	$\hat{S}(t_12)\left(\frac{1-0}{1}\right)$

Question 11.2

Based solely on the Kaplan-Meier curves for the two treatment groups, which treatment appears to prolong survival more effectively?



The Kaplan-Meier estimator makes three important assumptions:

1. The probability of censoring is unrelated to the outcome of interest.
2. The survival probabilities are the same for participants recruited at different times during the study (e.g., circumstances that could alter the survival, such as treatments, do not change over calendar time).
3. The events occurred at exactly the times specified.

Question 11.3

What is one way each of these assumptions could be violated?

Question 11.4

Here are the data for treatment group 2 of the `ovarian` dataset. Perform the calculations of $\hat{S}(t_j)$ for $j = 0, \ldots, 13$, starting with $t_0 = 0$. Draw the Kaplan-Meier curve, adding symbols for the censoring events.

	rx	futime	fustat
1	2	353	1
2	2	365	1
3	2	377	0
4	2	421	0
5	2	464	1
6	2	475	1
7	2	563	1
8	2	744	0
9	2	769	0
10	2	770	0
11	2	1129	0
12	2	1206	0
13	2	1227	0

$\left(\frac{13-1}{13}\right) = 0.923$

$0.923\left(\frac{12-1}{12}\right) = 0.846$

$0.846\left(\frac{11}{11}\right) = 0.846$

$0.846\left(\frac{11-1}{11}\right) = 0.769$

$0.769\left(\frac{10-1}{10}\right) = 0.692$

$0.692\left(\frac{9-1}{9}\right) = 0.615$

$0.615\left(\frac{8-0}{8}\right) = 0.615$

$\rightarrow 0.615$ b/c all censored