MTH6157

Week 3 exercise (Kaplan Meier Estimate) Solutions

Q1

(a) K-M suitable here because we have:
- discrete data
- right censoring
- survival which is not age related

(b) the survival function is

$$S(t) = \prod_{tj \le t}(1 - \lambda_j)$$

where $\lambda_j = d_j / n_j$ is the hazard at time t_j days

d_j = number of people first reporting pain-free after t_j days

n_j = risk set at t_j days which allows for censoring c_j where

c_j = number leaving the study at t_j days before reporting pain free

(c) "Survival" here is the continuation of back pain, therefore more successful treatments will lead to lower S(t) values more quickly. We will look to compare the survival functions for the two treatments to see whether one is consistently lower than the other. We need to assess the two survival functions at all durations because we are interested in both the rate of successfully removing pain and the speed with which the treatment works.

(d) we evaluate the Kaplan Meier estimates separately for each treatment

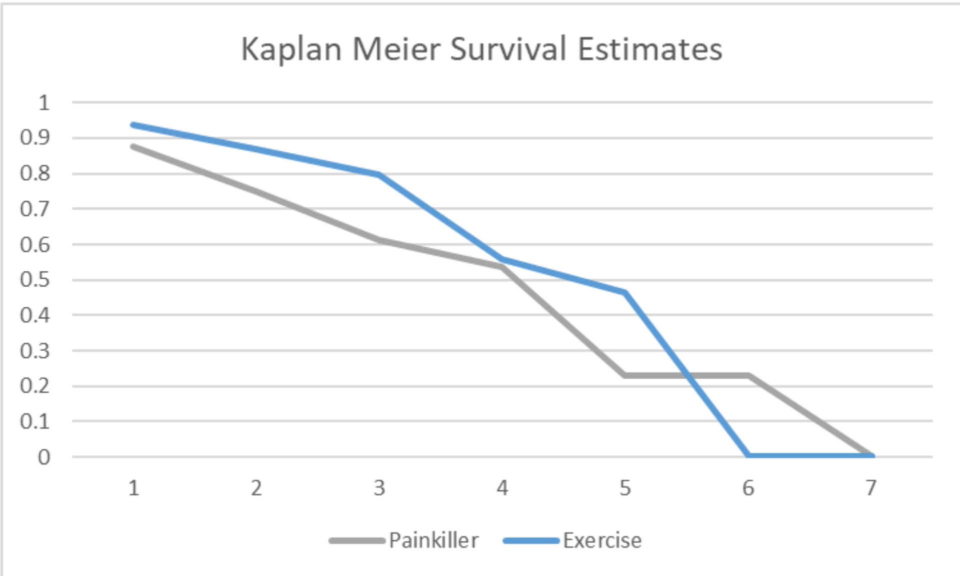
for daily painkiller

t_j	n_j	d_j	c_j	λ_j	$1 - \lambda_j$	S(t)
0	16					
1	16	2	0	1/8	7/8	0.875
2	14	2	1	1/7	6/7	0.75
3	11	2	1	2/11	9/11	0.6136
4	8	1	0	1/8	7/8	0.5369
5	7	4	0	4/7	3/7	0.2301
6	3	0	1	0	1	0.2301
7	2	2	0	1	0	0

for exercise routine

t_j	n_j	d_j	c_j	λ_j	$1 - \lambda_j$	S(t)
0	16					
1	16	1	1	1/16	15/16	0.9375
2	14	1	1	1/14	13/14	0.8705
3	12	1	1	1/12	11/12	0.7980
4	10	3	1	3/10	7/10	0.5586
5	6	1	3	1/6	5/6	0.4655
6	2	2	0	1	0	0
7	0	0	0			0

(e) we begin by graphing the two survival function estimates



conclusions we may make:

- survival here measures time still with back pain therefore we are looking for treatments with lower survival probabilities
- both treatments are successful overall in terms of removing pain and over quite similar timescales
- the shape of the two survival functions is similar but there are differences in gradient at different parts of the week-long trial

- painkiller removes more backpain in the early part of the trial with the survival function lower at t = 1 .. 5. The biggest difference in favour of painkiller occurs after 5 days
- however, for those that persist with the exercises there is a larger drop in survival function at t=6 after which all those in that part of the trial are pain free or have left the trial
- the amount of censoring is high, especially for the exercise patients. This should lead to some caution in interpreting results, especially from t=5 onwards.

(f) ways in which this study might be improved:

- we would benefit from a larger sample
- we need to understand more about the censored data. In particular:
    - what is it about the exercise routine that makes a large proportion of remaining people leave after 5 days?
    - might the censored lives be switching to the other treatment? In particular there seems to be a risk that exercise patients who continue to have pain might start taking painkillers
    - some of the right censoring could also be people simply forgetting to take the daily treatment or do the exercises
- this study does not monitor whether pain returns later having first gone
- there is also some left censoring here:
    - we know all 16 people had back pain at t=0 but not how long prior to the study they had been suffering
    - we do not know what treatment they were taking before t=0
- other covariate data might be relevant in this study

Q2

(a) once again as in Q1 above begin by defining t, n, s, c, λ, S(t) in terms of the pen example and censoring

Then the Kaplan Meier estimate is found from the table below

t	n	d	c	λ	1- λ
0	100				
1	100	2	0	0.020	0.980
2	98	6	2	0.061	0.939
3	90	13	2	0.144	0.856
4	75	9	0	0.120	0.880
5	66	16	3	0.242	0.758

where the hazard is in the λ column [note the question asks for the hazard not for the survival function in part (a)]

(b) Probability a new pen working at the end of the week (assuming 5 day week here) is S(5)

= 0.980 x 0.939 x 0.856 x 0.880 x 0.758

= 0.525