Tutorial 2 - Survival Analysis

Problem 1

Given a hazard function $h(t) = c$, derive the corresponding survival function $S(t)$ and probability density function $f(t)$.

Problem 2

Censoring is a missing data problem that is specific to survival analysis. Explain what is meant by a right censored observation that occured at week 18.

Problem 3

Define the Kaplan-Meier estimator of the population survival function. Consider a sample of survival data with event times:

$$6+ \quad 6 \quad 6 \quad 6 \quad 7 \quad 9+ \quad 10+ \quad 10 \quad 11+ \quad 13 \quad 16 \quad 17+ \quad 19+ \quad 20+ \quad 22+,$$

where a + indicates a censored observation. Compute the Kaplan-Meier estimator for these data.

Problem 4

The survival times (in weeks) for a sample of 21 glioblastoma multiforme brain tumour sufferers after a combination of surgery and chemotherapy performed on them were as follows:

$$1 \quad 2 \quad 2 \quad 2 \quad 2+ \quad 6 \quad 8 \quad 8 \quad 9 \quad 9+ \quad 13 \quad 13+ \quad 16 \quad 17 \quad 22+ \quad 25+ \quad 29 \quad 34 \quad 36+ \quad 43+ \quad 45+$$

Table 1: Summary table of survival analysis with some values missing.

Time	Number at Risk	Number Failed	Survival Probability	$\sqrt{\sum_{j=1}^{i} \frac{r_j - s_j}{r_j s_j}}$	SE	95.0% CI Lower	95.0% CI Upper
1	21	1	0.9524	0.0488	0.0465	0.7072	0.9932
2	*	*	******	0.1059	0.0857	0.5689	0.9239
6	16	1	0.7589	0.1240	0.0941	0.5139	0.8920
8	15	2	0.6577	0.1601	0.1053	0.4123	0.8203
9	13	*	0.6071	0.1790	******	0.3650	0.7811
13	11	1	0.5519	******	0.1119	0.3135	0.7375
16	9	1	0.4906	0.2346	0.1151	0.2572	0.6884
17	*	*	******	0.2700	0.1159	0.2058	0.6362
29	5	1	0.3434	0.3505	0.1204	0.1310	0.5701
34	4	1	0.2576	0.4541	0.1170	0.0732	0.4947

Compute the numbers that are missing in the table above.

Problem 5

Two samples of five patients on two different treatments (A and B) were followed until relapse of their disease. The survival times were recorded to the nearest month and were as follows:

Time	3	6	9+	11	14	7	11	16+	17	21+
Treatment	A	A	A	A	A	B	B	B	B	B

Using the log-rank test, decide if there is sufficient evidence to suggest a significant difference in the (population) survival functions for the two treatments.

Problem 6

In a study of peritonitis infection after kidney transplants, the time (from the end of the first episode) until occurrence of a second episode of peritonitis was recorded, as well as whether or not a patient had a low or high immune response. The data are

Low immune response subjects:

1+ 1+ 1+ 1 2+ 2 5 8 9 9 12+ 12 12 14+ 15 17 20 23+ 23 26 30+ 39 40+.

High immune response subjects:

1+ 2 3 3 4+ 6 7+ 14 17 17 23 26 40+ 48+ 48+ 49 89+ 101+ 109+ 117+.

Comment on the plot before you carry out any formal statistical tests. Do you anticipate to find a difference in the survival times between the two immune responses?

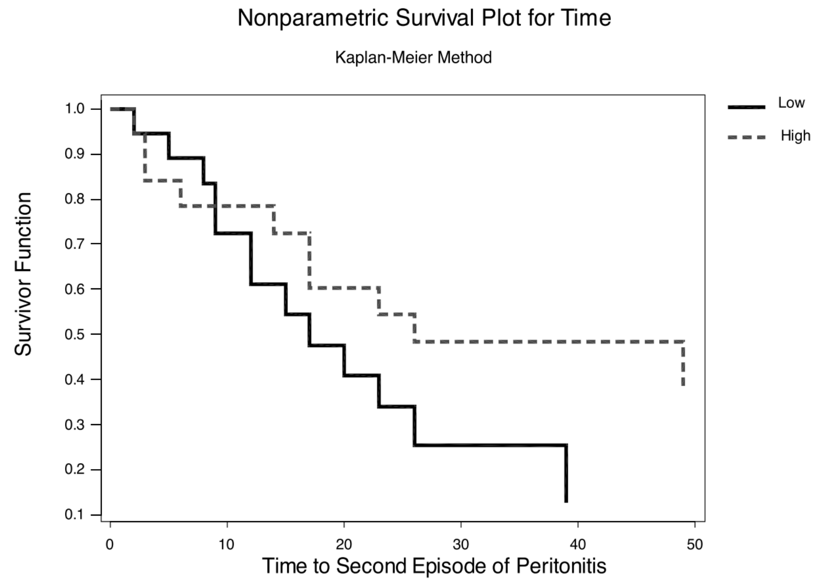

Figure 1: Kaplan-Meier survival curves plotted separately for high and low immune response subjects in a study of peritonitis infection after kidney transplants.

Fill in the asterisks, *, in the table below and carry out the corresponding log-rank test of whether or not the population survival distributions are identical for the two types of immune response.

Table 2: Summary table of log-rank test with some values missing.

$t_{(i)}$	d_{Li}	d_{Hi}	r_{Li}	r_{Hi}	E_{Li}	E_{Hi}
1	1	0	23	20	0.53	0.47
2	1	*	19	19	1.00	1.00
3	0	2	17	*	0.97	1.03
5	1	0	17	15	0.53	0.47
6	*	1	*	15	0.52	*
8	1	0	16	13	0.55	0.45
9	2	0	15	13	*	*
12	2	0	13	13	1.00	1.00
14	0	1	10	13	0.43	0.57
15	1	0	9	12	0.43	0.57
17	1	2	8	12	1.20	1.80
20	1	0	7	10	0.41	0.59
23	1	1	6	10	0.75	1.25
26	1	1	4	9	0.62	1.38
39	1	0	2	8	0.20	0.80
49	0	1	0	5	0.00	1.00
Totals	*	*			*	*

Problem 7

Write down the general form of the proportional hazards model and specify the assumptions required for such a model to hold. Derive the survival function for this model.

Problem 8

A Cox proportional hazards model is fit in `R` to the data from Problem 6 with treatment as its single explanatory variable. Treatment is coded as 0 for treatment A and 1 for treatment B. By default, R chooses treatment A as baseline. This results in the following output:

Variable	β	ese(β)	CI for β
Treatment	-1.25	0.88	$(-2.97, 0.47)$

Using the above output, estimate the hazard ratio for treatment B relative to treatment A and provide a 95% confidence interval for this ratio. What does this suggest to you about the effect of the treatments on the risk of relapse?

Problem 9

In heart transplant surgery, the following factors are thought to influence survival times after operation:

(a) sex of the patient;

(b) age of the patient at operation; and

(c) heart-matching score.

This last variable is an attempt to measure how well the transplanted heart matches the characteristics of the patient's own heart and is scored on a scale of 0 (very poor) to 3 (very good).

Patients within a hospital trust have been randomly allocated to one of four treatment courses after surgery, one of these being the standard regime and the others all being minor modifications of the standard. There is particular interest in whether any of these three modifications provide a general improvement in survival time over the standard.

Write down the hazard function for a proportional hazards model incorporating all these factors, and explain briefly how you would investigate the question of interest.