STAT 4008 Assignment 2 **Due: Mar. 5, 2021**

1. Let $\hat{S}(t)$ be the Kaplan-Meier estimator of the survival function and

$$\sigma_S^2(t) = \sum_{j|t_j<t} \frac{d_j}{n_j(n_j - d_j)}$$

(a) Show that the approximate variance of $\arcsin\left[\sqrt{\hat{S}(t_0)}\right]$ is

$$\frac{1}{4}\sigma_S^2(t_0)\frac{\hat{S}(t_0)}{1-\hat{S}(t_0)}$$

(b) Hence, the $100(1-\alpha)\%$ confidence interval for $S(t_0)$, based on this transformation, is given as

$$\sin^2\left\{\max\left[0, \arcsin\left(\hat{S}(t_0)^{1/2}\right) - 0.5z_{1-\alpha/2}\sigma_S(t_0)\left(\frac{\hat{S}(t_0)}{1-\hat{S}(t_0)}\right)^{1/2}\right]\right\}$$
$$\leq S(t_0) \leq$$
$$\sin^2\left\{\min\left[\frac{\pi}{2}, \arcsin\left(\hat{S}(t_0)^{1/2}\right) + 0.5z_{1-\alpha/2}\sigma_S(t_0)\left(\frac{\hat{S}(t_0)}{1-\hat{S}(t_0)}\right)^{1/2}\right]\right\}$$

2. Consider the following data set:

Time	10	11	12	13	14	15	16	17	18	19
Number of items failed	3	2	0	4	5	3	2	1	1	2
Number of items (right) censored	0	1	2	1	1	0	3	2	4	2

Assume the data are from a population with survival function $S(t)$

(a) write down the likelihood function of $S(t)$.

(b) Estimate the survival function using the Kaplan-Meier method.

(c) Estimate the survival function using the Nelson-Aalen method.

(d) Estimate the mean survival time and its standard error.

(e) Give all three 95% confidence intervals for $S(14)$ discussed in the class.

Please give details in this question, i.e., use no computer package.

3. Consider the following right-censored sample:

$$2, 4, 4, 4+, 5, 6+, 7, 7+, 8+$$

Estimate the mean survival time and its standard error.

4. A study was conducted on the effects of ploidy on the prognosis of patients with cancers of the mouth. Patients were selected who had a paraffin-embedded sample of the cancerous tissue taken at the time of surgery. Follow-up survival data was obtained on each patient. The tissue samples were examined using a flow cytometer to determined if the tumor had an aneuploid (abnormal) or diploid (normal) DNA profile. The data are in the following table. Times are in weeks.

Aneuploid Tumors:
Death Times: 1, 3, 3, 4, 10, 13, 13, 16, 16, 24, 26, 27, 28, 30, 30, 32, 41, 51, 65, 67, 70, 72, 73, 77, 91, 93, 96, 100, 104, 157, 167
Censored Observations: 61, 74, 79, 80, 81, 87, 87, 88, 89, 97, 101, 104, 108, 109, 120, 131, 150, 231, 240, 400
Diploid Tumors:
Death Times: 1, 3, 4, 5, 5, 8, 23, 26, 27, 30, 42, 56, 62, 69, 104, 104, 112, 129, 181
Censored Observations: 8, 67, 76, 104, 176, 231

   (a) Estimate the survival functions and their standard error for both the diploid and aneuploid groups.

   (b) Estimate the cumulative hazard rates and their standard error for both the diploid and aneuploid groups.

   (c) Provide an estimate of the mean time to death, and find a 95% confidence interval for the mean survival time for both the diploid and aneuploid groups.

   (d) Provide an estimate of the median time to death, and find a 95% confidence interval for the median survival time for both the diploid and aneuploid groups.

5. A data set is given in "ass2q5" from Blackboard. It contains "Times(in Years)" in Column 1, "Censor" in Column 2 where 0 indicates death and 1 indicates alive.

   (a) Estimate the survival function and its standard error.

   (b) Estimate the survival function at $t = 5$ and its standard error.

   (c) Provide an estimate of the median time to death, and find a 95% confidence interval for the median survival time.

6. Suppose you are given the following data set:

   $1, 6+, 5-, 3, (7, 9], 2-, (3, 4], 4, 5+$
   +: right censored; -: left censored

   Find the estimate of the survival function.

7. Consider a hypothetical study of the mortality experience of diabetics. Thirty diabetic subjects are recruited at a clinic and followed until death or the end of study. The subject's age at entry into the study and their age at the end of study or death are given in the table below. Of interest is estimating the survival curve for a 60- or for a 70-year-old diabetic.

   (a) Since the diabetics needed to survive long enough from birth until the study began, the data is left-truncated. Construct a table showing the number of subjects at risk, Y, as a function of age.

(b) Estimate the conditional survival function for the age of death of a diabetic patient who has survived to age 60.

(c) Estimate the conditional survival function for the age of death of a diabetic patient who has survived to age 65.

Entry Age	Exit Age	Death Indicator	Entry Age	Exit Age	Death Indicator
58	60	1	67	70	1
58	63	1	67	77	1
59	69	0	67	69	1
60	62	1	68	72	1
60	65	1	69	79	0
61	72	0	69	72	1
61	69	0	69	70	1
62	73	0	70	76	0
62	66	1	70	71	1
62	65	1	70	78	0
63	68	1	71	79	0
63	74	0	72	76	1
64	71	1	72	73	1
66	68	1	73	80	0
66	69	1	73	74	1

8. A study was performed to estimate the distribution of incubation times of individuals known to have a sexually transmitted disease (STD). Twenty five patients with a confirmed diagnosis of STD at a clinic were identified on June 1, 1996. All subjects had been sexually active with a partner who had a confirmed diagnosis of a STD at some point after January 1, 1993 (hence $\tau = 42$ months). For each subject the date of the first encounter to the clinical confirmation of the STD diagnosis. Based on this right truncated sample, compute an estimate of the probability that the infection period is less than x months conditional on the infection period's being less than 42 months.

Date of First Encounter	Months From 1/93 to Encounter	Time (in months) until STD Diagnosed in Clinic
2/93	2	30
4/93	4	27
7/93	7	25
2/94	14	19
8/94	20	18
6/94	18	17
8/93	8	16
1/94	13	16
5/94	17	15
2/95	26	15
8/94	20	15
3/94	15	13
11/94	23	13
5/93	5	12
4/94	16	11
3/94	15	9
11/93	11	8
6/93	9	8
9/95	33	8
4/93	4	7
8/93	8	6
11/95	35	6
10/93	10	6
12/95	36	4
1/95	25	4