Survival Analysis HW 4

Mario Ibanez

February 16, 2016

4.1)

Using the data reported in section 1.3, find the quantities specified below for the AML low-risk and AML high-risk groups. Note that most of these quantities are worked out in detail in Example 4.2 and its continuations for the ALL group.

a) Estimate the survival functions and their standard deviations for the AML low-risk and AML high-risk groups.

b) Estimate the cumulative hazard rates and their standard errors for the AML low-risk and AML high-risk groups.

c) Provide a crude estimate of the hazard rates for each group baed on the estimates obtained in (b).

I will do parts a), b), and c) all together.

From the data set, the time that is used is the column *t2 - Disease Free Survival Time (Time To Relapse, Death Or End Of Study)* and the event column is *d3 - Disease Free Survival Indicator (1-Dead Or Relapsed, 0-Alive Disease Free)*. We are studying the time until the patient is no longer disease-free, which occurs when the patient either dies or relapses.

The Kaplan-Meier estimates and their estimated standard deviation were straightforward using *R*. The cumulative hazard functions and their standard deviations were calculated in *R* using the equations

$$\hat{H}(t) = \begin{cases} 0 & \text{if } t \leq t_1 \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & \text{if } t_1 \leq t \end{cases}$$

and

$$\sigma_H^2(t) = \sum_{t_i \leq t} \frac{d_i}{Y_i^2}$$

The estimated hazard rates were more difficult to calculate. The textbook mentions rates in terms of events per month. I made a rough estimate of the number of events per month by

$$\text{events per month at } t_i \approx \frac{H_{t_{i+1}} - H_{t_i}}{t_{i+1} - t_i}(30)$$

The fraction gives a very rough estimate of events per day, and then multiplying by 30 gives the corresponding rough estiamte of events per month. A more sophisticated approach could be used.

The estimates for the *ALL* group:

t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\widetilde{H}(t_i)$	$\sigma_H(t_i)$	Events/Month
1	38	1	0.9737	0.0260	0.0263	0.0263	0.0150
55	37	1	0.9474	0.0362	0.0533	0.0377	0.0439
74	36	1	0.9211	0.0437	0.0811	0.0468	0.0714
86	35	1	0.8947	0.0498	0.1097	0.0549	0.0490
104	34	1	0.8684	0.0548	0.1391	0.0623	0.3030
107	33	1	0.8421	0.0592	0.1694	0.0692	0.4688
109	32	1	0.8158	0.0629	0.2007	0.0760	0.9677
110	31	1	0.7895	0.0661	0.2329	0.0825	0.1667
122	30	2	0.7368	0.0714	0.2996	0.0950	0.1531
129	28	1	0.7105	0.0736	0.3353	0.1015	0.0258
172	27	1	0.6842	0.0754	0.3723	0.1081	0.0577
192	26	1	0.6579	0.0770	0.4108	0.1147	0.6000
194	25	1	0.6316	0.0783	0.4508	0.1215	0.0362
230	23	1	0.6041	0.0795	0.4943	0.1290	0.0296
276	22	1	0.5767	0.0805	0.5397	0.1368	0.0255
332	21	1	0.5492	0.0812	0.5873	0.1449	0.0294
383	20	1	0.5217	0.0817	0.6373	0.1532	0.0451
418	19	1	0.4943	0.0819	0.6900	0.1620	0.0347
466	18	1	0.4668	0.0818	0.7455	0.1713	0.0840
487	17	1	0.4394	0.0815	0.8044	0.1811	0.0481
526	16	1	0.4119	0.0809	0.8669	0.1916	0.0258
609	14	1	0.3825	0.0803	0.9383	0.2045	0.0435
662	13	1	0.3531	0.0793	1.0152	0.2185	0.0000

An estimate of the average events per month over the first 526 days for the cumulative hazard function is

$$30 * \frac{0.8669 - 0.0263}{526 - 1} = 0.04803429$$

This provides a much smoother approximation than the rough approximation calcualted each day found in the table.

The estimates for the *AML low-risk* group:

t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\widetilde{H}(t_i)$	$\sigma_H(t_i)$	Events/Month
10	54	1	0.9815	0.0183	0.0185	0.0185	0.0226
35	53	1	0.9630	0.0257	0.0374	0.0264	0.0444
48	52	1	0.9444	0.0312	0.0566	0.0327	0.1176
53	51	1	0.9259	0.0356	0.0762	0.0381	0.0231
79	50	1	0.9074	0.0394	0.0962	0.0430	0.6122
80	49	1	0.8889	0.0428	0.1166	0.0476	0.0250
105	48	1	0.8704	0.0457	0.1375	0.0520	0.0060
211	47	1	0.8519	0.0483	0.1587	0.0562	0.0815
219	46	1	0.8333	0.0507	0.1805	0.0602	0.0230
248	45	1	0.8148	0.0529	0.2027	0.0642	0.0284
272	44	1	0.7963	0.0548	0.2254	0.0681	0.0436
288	43	1	0.7778	0.0566	0.2487	0.0720	0.0077
381	42	1	0.7593	0.0582	0.2725	0.0758	0.0813
390	41	1	0.7407	0.0596	0.2969	0.0796	0.0312
414	40	1	0.7222	0.0610	0.3219	0.0835	0.1099
421	39	1	0.7037	0.0621	0.3475	0.0873	0.0132
481	38	1	0.6852	0.0632	0.3738	0.0912	0.1622
486	37	1	0.6667	0.0642	0.4009	0.0951	0.0069
606	36	1	0.6481	0.0650	0.4286	0.0991	0.0245
641	35	1	0.6296	0.0657	0.4572	0.1031	0.0140
704	34	1	0.6111	0.0663	0.4866	0.1072	0.0207
748	33	1	0.5926	0.0669	0.5169	0.1114	0.0037
1063	26	1	0.5698	0.0681	0.5554	0.1179	0.1091
1074	25	1	0.5470	0.0691	0.5954	0.1245	0.0044
2204	6	1	0.4558	0.1012	0.7621	0.2080	0.0000

An estimate of the average events per month over the first 600 days for the cumulative hazard function is

$$30 * \frac{0.4286 - 0.0185}{606 - 10} = 0.02064262$$

The estimates for the *AML high-risk* group:

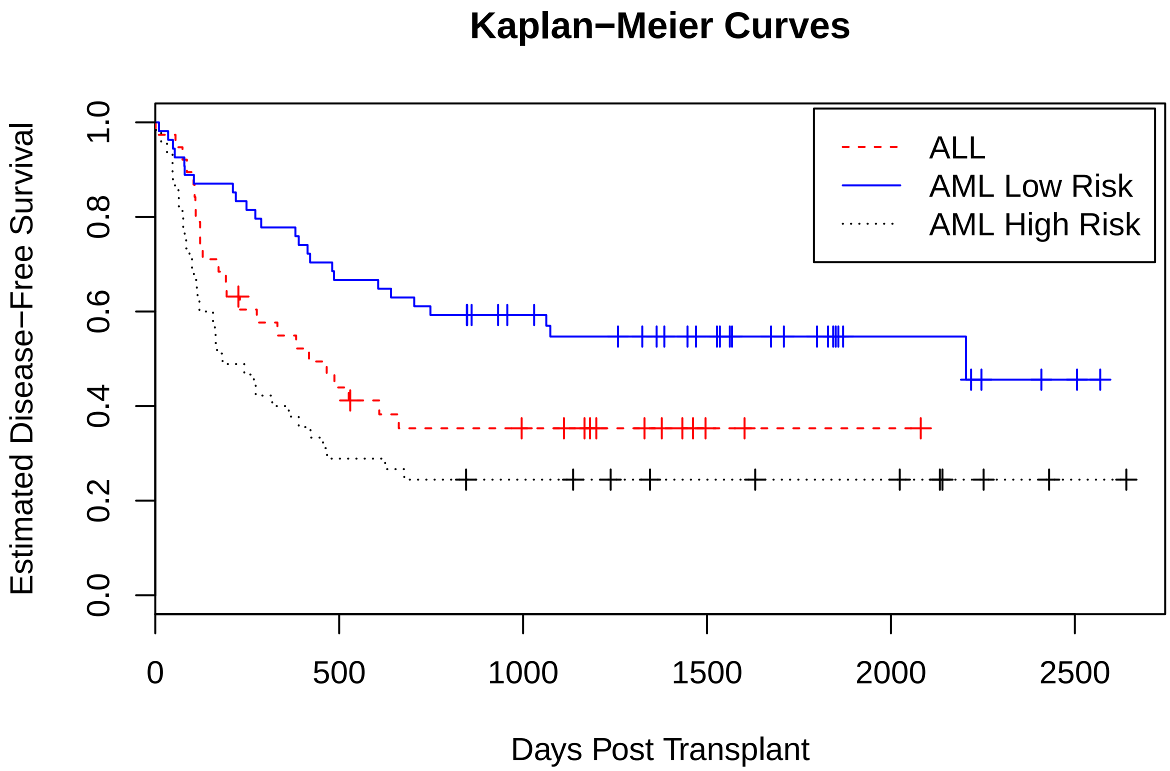
t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\widetilde{H}(t_i)$	$\sigma_H(t_i)$	Events/Month
2	45	1	0.9778	0.0220	0.0222	0.0222	0.0487
16	44	1	0.9556	0.0307	0.0449	0.0318	0.0436
32	43	1	0.9333	0.0372	0.0682	0.0394	0.0952
47	42	2	0.8889	0.0468	0.1158	0.0518	0.7500
48	40	1	0.8667	0.0507	0.1408	0.0575	0.0513
63	39	1	0.8444	0.0540	0.1665	0.0630	0.7895
64	38	1	0.8222	0.0570	0.1928	0.0683	0.0811
74	37	1	0.8000	0.0596	0.2198	0.0734	0.4167
76	36	1	0.7778	0.0620	0.2476	0.0785	0.2143
80	35	1	0.7556	0.0641	0.2762	0.0835	0.2206
84	34	1	0.7333	0.0659	0.3056	0.0886	0.1010
93	33	1	0.7111	0.0676	0.3359	0.0936	0.1339
100	32	1	0.6889	0.0690	0.3671	0.0987	0.1935
105	31	1	0.6667	0.0703	0.3994	0.1038	0.1250
113	30	1	0.6444	0.0714	0.4327	0.1090	0.5172
115	29	1	0.6222	0.0723	0.4672	0.1144	0.2143
120	28	1	0.6000	0.0730	0.5029	0.1198	0.0300
157	27	1	0.5778	0.0736	0.5399	0.1254	0.2308
162	26	1	0.5556	0.0741	0.5784	0.1312	0.6000
164	25	1	0.5333	0.0744	0.6184	0.1371	0.3125
168	24	1	0.5111	0.0745	0.6601	0.1433	0.0870
183	23	1	0.4889	0.0745	0.7036	0.1498	0.0231
242	22	1	0.4667	0.0744	0.7490	0.1565	0.0549
268	21	1	0.4444	0.0741	0.7966	0.1636	0.3000
273	20	1	0.4222	0.0736	0.8466	0.1711	0.0351
318	19	1	0.4000	0.0730	0.8993	0.1790	0.0370
363	18	1	0.3778	0.0723	0.9548	0.1874	0.0654
390	17	1	0.3556	0.0714	1.0136	0.1964	0.0586
422	16	1	0.3333	0.0703	1.0761	0.2061	0.0588
456	15	1	0.3111	0.0690	1.1428	0.2166	0.1948
467	14	1	0.2889	0.0676	1.2142	0.2281	0.0146
625	13	1	0.2667	0.0659	1.2912	0.2407	0.0481
677	12	1	0.2444	0.0641	1.3745	0.2547	0.0000

An estimate of the average events per month over the first 600 days for the cumulative hazard function is

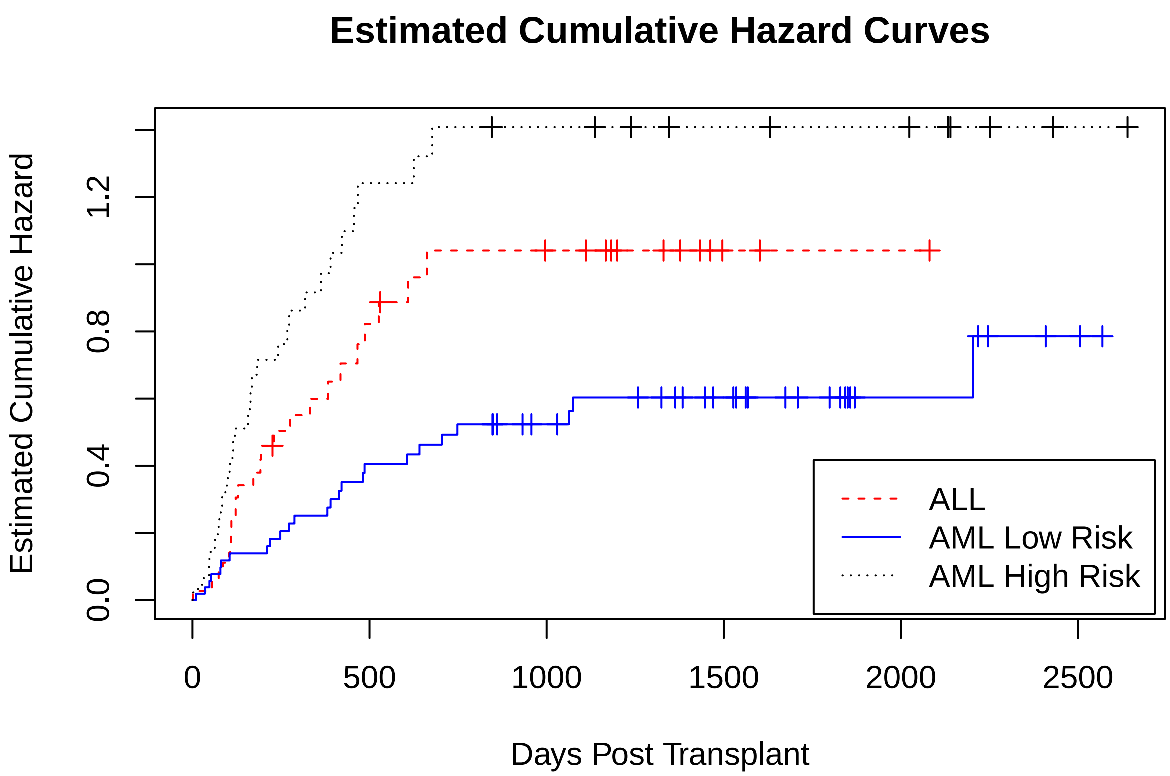
$$30 * \frac{1.2912 - 0.0222}{625 - 2} = 0.06110754$$

These estimates of 0.06 events per month for *AML high-risk*, 0.02 events per month for *AML low-risk*, and 0.048 events per month for *ALL* agree with what was mentioned in the textbook.

Below is a plot of the Kaplan-Meier estimates for the three groups:



And below is a plot of the estimated cumulative hazard curves for the three groups. Note, these curves are not the Nelson-Aalen estimators $\widetilde{H}(t_i)$. These are the transformations of the Kaplan-Meier estimates: $\hat{H}(t_i) = -ln(\hat{S}(t_i))$. Since the sample sizes are not small and the estimators are asymptotically equivalent, these curves should look approximately like the Nelson-Aalen curves.



d) Estimate the mean time to death, and find 95% confidence intervals for the mean survival time for both the AML low-risk and AML high-risk groups.

For these calculations, I used the *stepfun()* function in *R* which creates a function object of the survival curves. I then passed these to the function *integrate()*. The way these step functions were created, their final value is whatever the value is at the last observation. I followed the book's example and integrated these curves from day = 0 to day = 2081 so that each mean can be compared with each other. My results were close to the books answers.

Putting these results into a table:

Disease Group	Mean Restricted to 2081 Days	Standard Error	95% Confidence Interval
ALL	899.5	146.1	[613.144, 1185.856]
AML low-risk	1314.9	119.4	[1080.876, 1548.924]
AML high-risk	661.4	122.9	[420.516, 902.284]

These values are similar to the books values:

Disease Group	*Mean Restricted to 2081 days*	*Standard Error*	*95% Confidence Interval*
ALL	899.3 days	150.3 days	606.6–1193.9 days
AML low risk	1315.2 days	118.8 days	1082.4–1548.0 days
AML high risk	655.67 days	122.9 days	414.8–896.5 days

e) Work out estimates of the median time to death and find 95% confidence intervals for the median survival time for both the AML low-risk and AML high-risk groups using the linear, log-transformed and arcsine-square root formulas.

dfgbfg

f) Find 95% confidence intervals for the survival functions at 300 days post transplant for both the AML low-risk and AML high-risk groups, using the log and arcsine-square root transformed formulas.

dfbfgb

g) Find 95% EP confidence bands for the survival functions over the range 100–400 days post-transplant for both the AML low risk and AML high risk groups using the linear, log-transformed, and arcsine-transformed formulas.

dfgbfgb

h) Find 95% HW confidence bands for the survival functions over the range 100–400 days post-transplant for both the AML low risk and AML high risk groups using the linear, log-transformed, and arcsine-transformed formulas.

fgbdfg

i) Based on the results above and those discussed in Example 4.2 and its continuations, how do the survival experiences of the ALL, AML low risk, and AML high risk groups compare?

dfghdfgh

4.2)

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.

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

dfghdfgh

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

dfghfdg

c) Provide a crude estimate of the hazard rate for each group based on the estimates obtained in the exercise.

dfghdfh

d) 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.

dfhfgh

e) Estimate the median time to death, and find a 95% confidence interval for the median survival time for both the diploid and aneuploid groups.

dfghfgh

4.4)

The following table contains data on the survival times of twenty-five patients with inoperative lung cancer entered on a study between November 1, 1979 and December 23, 1979. Complete follow-up was obtained on all patients so that the exact date of death was known. The study had on interim analysis conducted on March 3, 1980, when only 13 patients had died.

a) Estimate the survival function based on the available sample information at the time of the interim analysis on 3/31/80. Provide an estimate of the standard error of your estimate.

dfghfgh

b) Use the Brown, Hollandar, and Kowar technique (Practical Note 3 of section 4.2) to complete the right-hand tail of the product-limit estimate found in part a.

dfhdffh

c) Using the complete follow-up on each patient, compute the estimate of the survival function and an estimate of its standard error. Compare this estimate to that found in part a.

dfghfdgh

d) Estimate the mean time to death, restricted to 724 days, based on the product-limit estimator found in part a.

dfghdfh

e) Estimate the mean time to death by finding the area under the survival curve found in part c. Find the standard error of your estimate.

dfghdfh

f) Compute the usual estimate of the time to death based on complete follow-up data by finding the arithmetic mean of the complete follow-up data. Find the standard error of this estimate in the usual way as the sample standard deviation divided by the square root of the sample size. Compare your answers to those obtained in part e.

dfhdfh

Code

4.1)

```
# Code for 4.1)
```

4.2)

```
# Code for 4.2)
```

4.4)

```
# Code for 4.4)
```