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

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

3

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:

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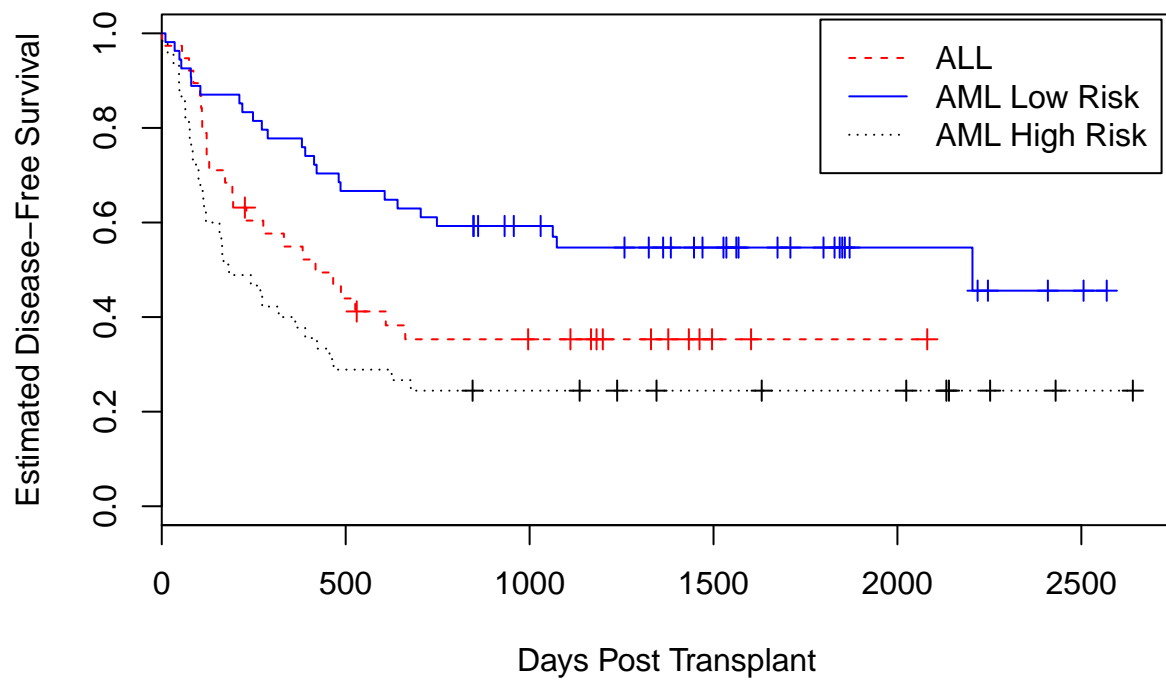
And below is a plot of the estimated cumulative hazard curves for the three groups. Note, these curves are not the Nelson-Aalen estimators $\widetilde{H}(t_i)$. These are the transformations of the Kaplan-Meier estimates: $\hat{H}(t_i) = -ln(\hat{S}(t_i))$. Since the sample sizes are not small and the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

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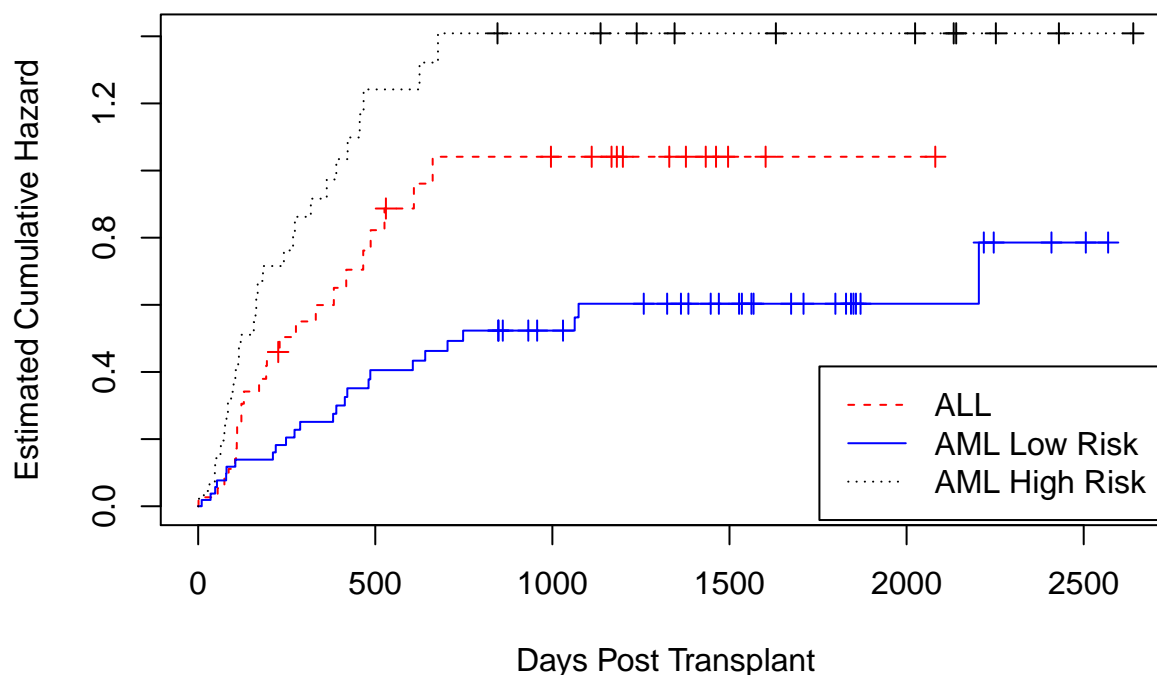
d) Estimate the mean time to death, and find 95% confidence intervals for the mean survival time for both the AML low-risk and AML high-risk groups.

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.

Here is a table of my results:

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 book's 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.

These were calculated with some R code:

ALL group

This is Table 4.7 in the textbook which I recreated for completeness:

time	surv	std.err	linear	log	arcsine
1	0.974	0.026	18.242	3.258	7.674
55	0.947	0.036	12.350	3.607	6.829
74	0.921	0.044	9.625	3.691	6.172
86	0.895	0.050	7.929	3.657	5.609
104	0.868	0.055	6.719	3.557	5.107
107	0.842	0.059	5.783	3.412	4.645
109	0.816	0.063	5.022	3.236	4.214
110	0.789	0.066	4.377	3.036	3.806
122	0.737	0.071	3.316	2.582	3.042
129	0.711	0.074	2.862	2.334	2.679
172	0.684	0.075	2.443	2.074	2.326
192	0.658	0.077	2.052	1.804	1.981
194	0.632	0.078	1.681	1.524	1.642
230	0.604	0.080	1.309	1.220	1.290
276	0.577	0.081	0.952	0.909	0.945
332	0.549	0.081	0.606	0.590	0.604
383	0.522	0.082	0.266	0.263	0.266
418	0.494	0.082	-0.070	-0.070	-0.070
466	0.467	0.082	-0.406	-0.411	-0.405
487	0.439	0.081	-0.744	-0.759	-0.741
526	0.412	0.081	-1.090	-1.114	-1.078
609	0.382	0.080	-1.464	-1.497	-1.437
662	0.353	0.079	-1.853	-1.886	-1.798

The estimated median is 418 days. In the final three columns, we look for values that are between -1.96 and 1.96. The 95% confidence intervals for the median are:

Type	Interval
Linear	$[194, \infty]$
Log	$[192, \infty]$
Arcsine	$[194, \infty]$

AML low-risk

This is the corresponding table for the group *AML low-risk*:

time	surv	std.err	linear	log	arcsine
10	0.981	0.018	26.244	3.613	9.537
35	0.963	0.026	18.014	4.116	8.697
48	0.944	0.031	14.258	4.322	8.046
53	0.926	0.036	11.951	4.395	7.492
79	0.907	0.039	10.328	4.392	6.999
80	0.889	0.043	9.093	4.339	6.548
105	0.870	0.046	8.103	4.251	6.130
211	0.852	0.048	7.278	4.136	5.736
219	0.833	0.051	6.573	4.001	5.362
248	0.815	0.053	5.956	3.849	5.005
272	0.796	0.055	5.406	3.683	4.661
288	0.778	0.057	4.910	3.505	4.328
381	0.759	0.058	4.456	3.317	4.006
390	0.741	0.060	4.037	3.120	3.691
414	0.722	0.061	3.646	2.916	3.384
421	0.704	0.062	3.278	2.703	3.084
481	0.685	0.063	2.930	2.484	2.788
486	0.667	0.064	2.598	2.259	2.497
606	0.648	0.065	2.280	2.029	2.211
641	0.630	0.066	1.973	1.792	1.927
704	0.611	0.066	1.675	1.551	1.647
748	0.593	0.067	1.385	1.304	1.369
1063	0.570	0.068	1.025	0.984	1.019
1074	0.547	0.069	0.681	0.664	0.679
2204	0.456	0.101	-0.436	-0.443	-0.435

The estimated median is 2204. In the final three columns, we look for values that are between -1.96 and 1.96. The 95% confidence intervals for the median are:

Type	Interval
Linear	$[704, \infty]$
Log	$[641, \infty]$
Arcsine	$[641, \infty]$

AML high-risk

This is the corresponding table for the group *AML high-risk*:

time	surv	std.err	linear	log	arcsine
2	0.978	0.022	21.743	3.429	8.530
16	0.956	0.031	14.829	3.852	7.687
32	0.933	0.037	11.653	3.995	7.033
47	0.889	0.047	8.301	3.961	5.978
48	0.867	0.051	7.236	3.861	5.522
63	0.844	0.054	6.375	3.728	5.098
64	0.822	0.057	5.654	3.571	4.698
74	0.800	0.060	5.031	3.393	4.317
76	0.778	0.062	4.482	3.200	3.951
80	0.756	0.064	3.989	2.993	3.599
84	0.733	0.066	3.540	2.775	3.257
93	0.711	0.068	3.125	2.546	2.924
100	0.689	0.069	2.737	2.308	2.599
105	0.667	0.070	2.372	2.063	2.280
113	0.644	0.071	2.024	1.809	1.966
115	0.622	0.072	1.691	1.548	1.657
120	0.600	0.073	1.369	1.281	1.351
157	0.578	0.074	1.056	1.007	1.048
162	0.556	0.074	0.750	0.727	0.747
164	0.533	0.074	0.448	0.441	0.448
168	0.511	0.075	0.149	0.148	0.149
183	0.489	0.075	-0.149	-0.150	-0.149
242	0.467	0.074	-0.448	-0.454	-0.448
268	0.444	0.074	-0.750	-0.764	-0.747
273	0.422	0.074	-1.056	-1.079	-1.048
318	0.400	0.073	-1.369	-1.401	-1.351
363	0.378	0.072	-1.691	-1.728	-1.657
390	0.356	0.071	-2.024	-2.061	-1.966
422	0.333	0.070	-2.372	-2.400	-2.280
456	0.311	0.069	-2.737	-2.745	-2.599
467	0.289	0.068	-3.125	-3.095	-2.924
625	0.267	0.066	-3.540	-3.451	-3.257
677	0.244	0.064	-3.989	-3.812	-3.599

The estimated median is 183 days. In the final three columns, we look for values that are between -1.96 and 1.96. The 95% confidence intervals for the median are:

Type	Interval
Linear	[115, 363]
Log	[113, 363]
Arcsine	[115, 363]

All of these confidence intervals and medians agree with those mentioned in the book.

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.

AML low-risk

For this group, we can get the values we need from the following table (the complete version of this table is actually in the next problem):

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2
288	0.7777778	0.0565750	0.0052910
381	0.7592593	0.0581799	0.0058717
390	0.7407407	0.0596353	0.0064815
400	0.7407407	0.0596353	0.0064815

The values we need are $\hat{S}(300) = 0.7777778$ and $\sigma_S^2(300) = 0.0052910$, or $\sigma_S(300) = 0.07273926$

The table of results is:

Group & Interval Type	Interval
AML low-risk, Log	[0.642, 0.8672]
AML low-risk, Arcsine	[0.6584, 0.8775]

AML high-risk

For this group, we can get the values we need from the following table (the complete version of this table is actually in the next problem):

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2
242	0.4666667	0.0743698	0.0253968
268	0.4444444	0.0740741	0.0277778
273	0.4222222	0.0736283	0.0304094
318	0.4000000	0.0730297	0.0333333
363	0.3777778	0.0722744	0.0366013
390	0.3555556	0.0713576	0.0402778
400	0.3555556	0.0713576	0.0402778

The values we need are $\hat{S}(300) = 0.4222222$ and $\sigma_S^2(300) = 0.0304094$, or $\sigma_S(300) = 0.1743829$

The table of results is:

Group & Interval Type	Interval
AML high-risk Log	[0.2776, 0.5599]
AML high-risk Arcsine	[0.2833, 0.5678]

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.
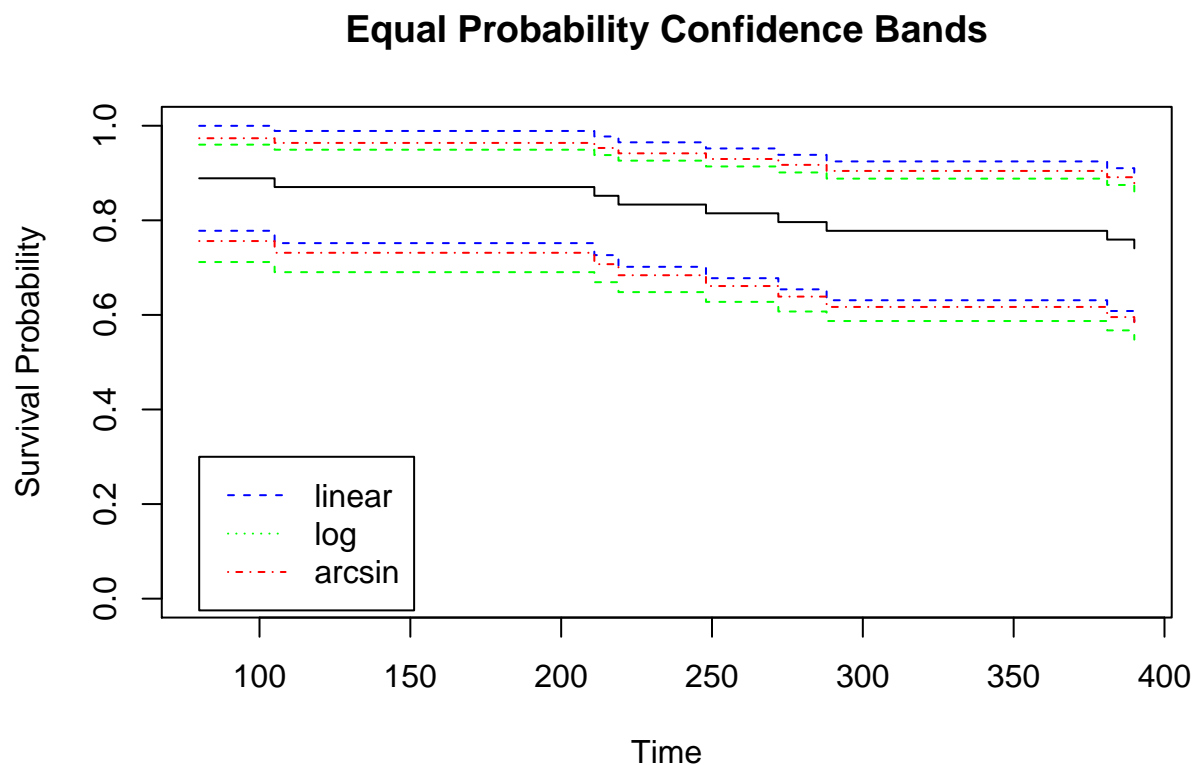
I used a bunch of R code to get these results:

AML low-risk

The values for a_L and a_U are $a_L = 0.1111111$ and $a_U = 0.2592593$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 2.5949$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
100	0.889	0.043	0.002	0.778	1.000	0.712	0.960	0.756	0.974
105	0.870	0.046	0.003	0.752	0.989	0.690	0.949	0.731	0.964
211	0.852	0.048	0.003	0.726	0.977	0.669	0.938	0.707	0.953
219	0.833	0.051	0.004	0.702	0.965	0.648	0.926	0.684	0.942
248	0.815	0.053	0.004	0.678	0.952	0.628	0.914	0.661	0.930
272	0.796	0.055	0.005	0.654	0.939	0.607	0.901	0.639	0.917
288	0.778	0.057	0.005	0.631	0.925	0.587	0.888	0.617	0.904
381	0.759	0.058	0.006	0.608	0.910	0.567	0.875	0.595	0.891
390	0.741	0.060	0.006	0.586	0.895	0.548	0.861	0.574	0.877
400	0.741	0.060	0.006	0.586	0.895	0.548	0.861	0.574	0.877

And below is a plot comparing the three confidence bands:
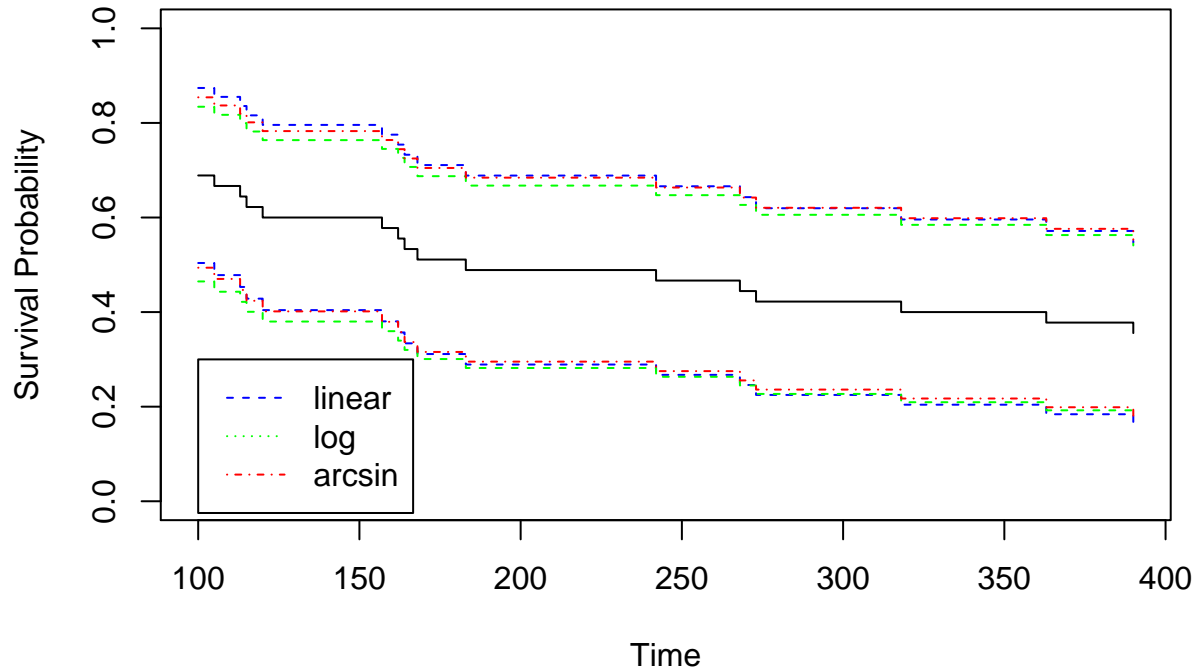

Equal Probability Confidence Bands

AML high-risk

The values for a_L and a_U are $a_L = 0.3111111$ and $a_U = 0.6444444$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 2.6811$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
100	0.689	0.069	0.010	0.504	0.874	0.465	0.834	0.494	0.854
105	0.667	0.070	0.011	0.478	0.855	0.443	0.817	0.470	0.837
113	0.644	0.071	0.012	0.453	0.836	0.422	0.800	0.447	0.819
115	0.622	0.072	0.013	0.428	0.816	0.401	0.782	0.424	0.801
120	0.600	0.073	0.015	0.404	0.796	0.380	0.764	0.401	0.783
157	0.578	0.074	0.016	0.380	0.775	0.360	0.745	0.379	0.764
162	0.556	0.074	0.018	0.357	0.754	0.340	0.726	0.358	0.745
164	0.533	0.074	0.019	0.334	0.733	0.320	0.707	0.337	0.725
168	0.511	0.075	0.021	0.311	0.711	0.301	0.687	0.316	0.705
183	0.489	0.075	0.023	0.289	0.689	0.282	0.667	0.295	0.684
242	0.467	0.074	0.025	0.267	0.666	0.263	0.647	0.275	0.663
268	0.444	0.074	0.028	0.246	0.643	0.245	0.627	0.255	0.642
273	0.422	0.074	0.030	0.225	0.620	0.227	0.606	0.236	0.621
318	0.400	0.073	0.033	0.204	0.596	0.209	0.584	0.217	0.599
363	0.378	0.072	0.037	0.184	0.572	0.192	0.563	0.199	0.576
390	0.356	0.071	0.040	0.164	0.547	0.176	0.541	0.181	0.553
400	0.356	0.071	0.040	0.164	0.547	0.176	0.541	0.181	0.553

And below is a plot comparing the three confidence bands:

Equal Probability Confidence Bands

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.
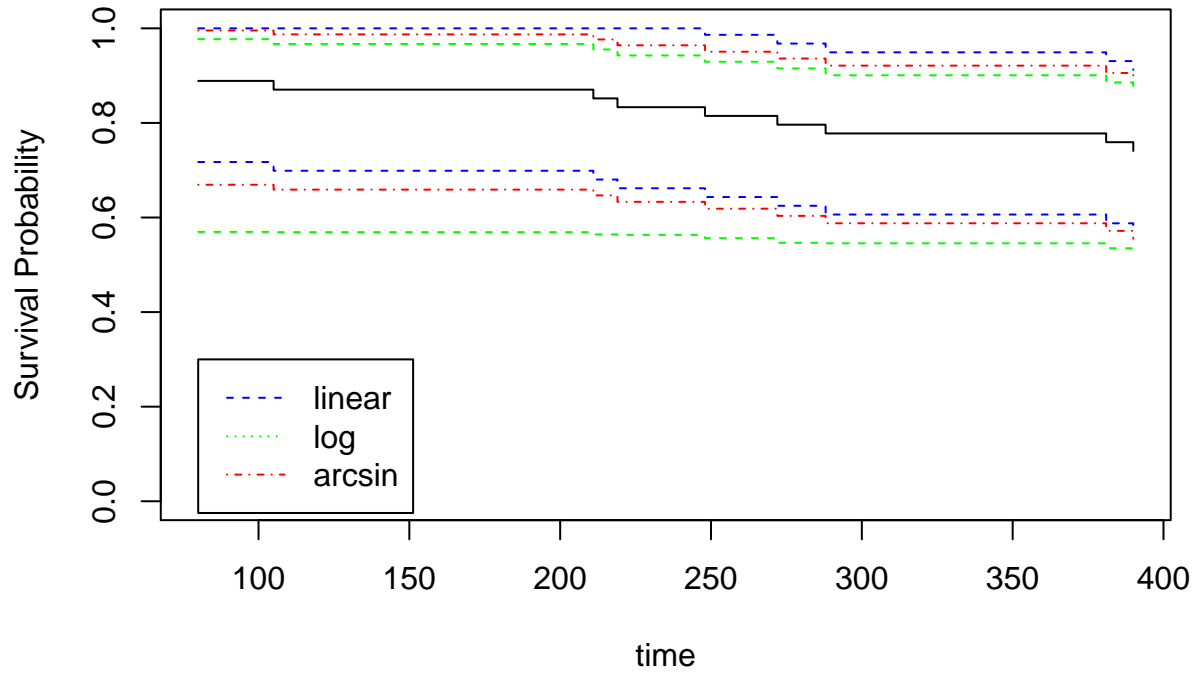
AML low-risk

The values for a_L and a_U are $a_L = 0.1111111$ and $a_U = 0.2592593$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 1.2603$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
100	0.889	0.002	0.717	1.000	0.545	0.977	0.669	0.996
105	0.870	0.003	0.699	1.000	0.563	0.967	0.659	0.987
211	0.852	0.003	0.680	1.000	0.570	0.955	0.647	0.977
219	0.833	0.004	0.662	1.000	0.569	0.943	0.633	0.964
248	0.815	0.004	0.643	0.986	0.564	0.929	0.619	0.951
272	0.796	0.005	0.625	0.968	0.556	0.915	0.603	0.936
288	0.778	0.005	0.606	0.949	0.546	0.901	0.588	0.921
381	0.759	0.006	0.588	0.931	0.535	0.886	0.572	0.906
390	0.741	0.006	0.569	0.912	0.522	0.870	0.555	0.890
400	0.741	0.006	0.569	0.912	0.522	0.870	0.555	0.890

And below is a plot comparing the three confidence bands:

Hall–Wellner Confidence Bands

AML high-risk

The values for a_L and a_U are $a_L = 0.3111111$ and $a_U = 0.6444444$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 1.3155$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
100	0.689	0.010	0.493	0.885	0.449	0.841	0.482	0.862
105	0.667	0.011	0.471	0.863	0.433	0.822	0.462	0.843
113	0.644	0.012	0.448	0.841	0.416	0.803	0.442	0.823
115	0.622	0.013	0.426	0.818	0.398	0.783	0.422	0.803
120	0.600	0.015	0.404	0.796	0.380	0.764	0.401	0.783
157	0.578	0.016	0.382	0.774	0.361	0.744	0.381	0.763
162	0.556	0.018	0.359	0.752	0.342	0.724	0.360	0.742
164	0.533	0.019	0.337	0.729	0.324	0.705	0.340	0.722
168	0.511	0.021	0.315	0.707	0.305	0.685	0.319	0.701
183	0.489	0.023	0.293	0.685	0.286	0.665	0.299	0.681
242	0.467	0.025	0.271	0.663	0.266	0.645	0.278	0.660
268	0.444	0.028	0.248	0.641	0.247	0.625	0.258	0.640
273	0.422	0.030	0.226	0.618	0.228	0.605	0.237	0.619
318	0.400	0.033	0.204	0.596	0.209	0.585	0.217	0.599
363	0.378	0.037	0.182	0.574	0.190	0.565	0.197	0.578
390	0.356	0.040	0.159	0.552	0.172	0.545	0.177	0.558
400	0.356	0.040	0.159	0.552	0.172	0.545	0.177	0.558

And below is a plot comparing the three confidence bands:

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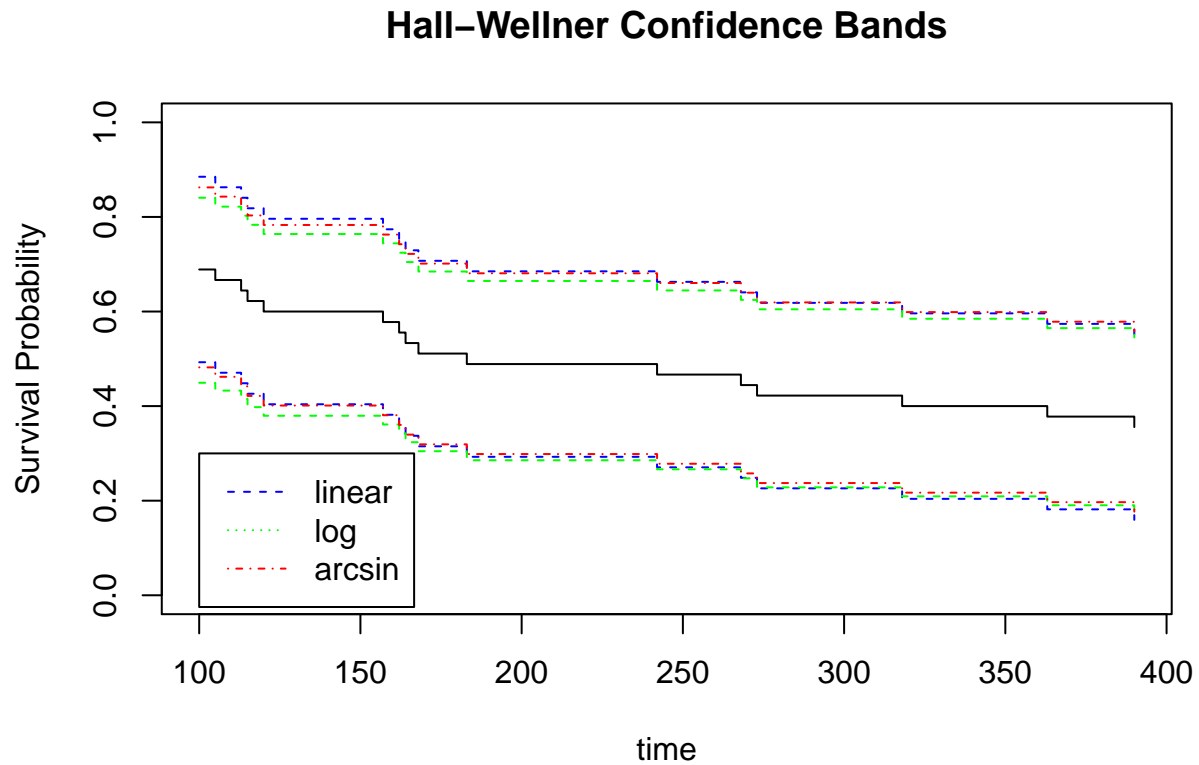
i) Based on the results above and those discussed in Example 4.2 and its continuations, how do the survival experiences of the ALL, AML low risk, and AML high risk groups compare?

Seems to me that the *AML high-risk* group is in the most danger of an early relapse or death. The *AML low-risk* group seems to be most well off in terms of not dying or relapsing.

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.

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

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

Like the previous problem, I'll do parts (a), (b) and (c) all together. And I'll be using some of the same functions/code I used for the previous problem:

Diploid Group

t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\tilde{H}(t_i)$	$\sigma_H(t_i)$
1	28	1	0.9643	0.0351	0.0357	0.0357
3	27	1	0.9286	0.0487	0.0728	0.0515
4	26	1	0.8929	0.0585	0.1112	0.0642
5	25	2	0.8214	0.0724	0.1912	0.0856
8	23	1	0.7857	0.0775	0.2347	0.0960
12	21	1	0.7483	0.0824	0.2823	0.1072
13	20	1	0.7109	0.0863	0.3323	0.1183
18	19	1	0.6735	0.0895	0.3849	0.1294
23	18	1	0.6361	0.0921	0.4405	0.1409
26	17	1	0.5986	0.0939	0.4993	0.1526
27	16	1	0.5612	0.0952	0.5618	0.1649
30	15	1	0.5238	0.0959	0.6285	0.1779
42	14	1	0.4864	0.0961	0.6999	0.1917
56	13	1	0.4490	0.0957	0.7768	0.2066
62	12	1	0.4116	0.0948	0.8602	0.2227
69	10	1	0.3704	0.0938	0.9602	0.2442
104	8	2	0.2778	0.0904	1.2102	0.3014
112	5	1	0.2222	0.0877	1.4102	0.3618
129	4	1	0.1667	0.0815	1.6602	0.4397
181	2	1	0.0833	0.0717	2.1602	0.6659

Aneuploid Group

| t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\widetilde{H}(t_i)$	$\sigma_H(t_i)$
1	52	1	0.9808	0.0190	0.0192	0.0192
3	51	2	0.9423	0.0323	0.0584	0.0337
4	49	1	0.9231	0.0370	0.0789	0.0394
10	48	1	0.9038	0.0409	0.0997	0.0446
13	47	2	0.8654	0.0473	0.1422	0.0538
16	45	2	0.8269	0.0525	0.1867	0.0623
24	43	1	0.8077	0.0547	0.2099	0.0665
26	42	1	0.7885	0.0566	0.2338	0.0706
27	41	1	0.7692	0.0584	0.2581	0.0747
28	40	1	0.7500	0.0600	0.2831	0.0788
30	39	2	0.7115	0.0628	0.3344	0.0867
32	37	1	0.6923	0.0640	0.3615	0.0909
41	36	1	0.6731	0.0651	0.3892	0.0950
51	35	1	0.6538	0.0660	0.4178	0.0992
65	33	1	0.6340	0.0669	0.4481	0.1037
67	32	1	0.6142	0.0677	0.4794	0.1083
70	31	1	0.5944	0.0683	0.5116	0.1130
72	30	1	0.5746	0.0689	0.5449	0.1179
73	29	1	0.5548	0.0693	0.5794	0.1228
77	27	1	0.5342	0.0697	0.6165	0.1283
91	19	1	0.5061	0.0715	0.6691	0.1386
93	18	1	0.4780	0.0728	0.7247	0.1494
96	16	1	0.4481	0.0741	0.7872	0.1619
100	14	1	0.4161	0.0754	0.8586	0.1770
104	12	1	0.3814	0.0767	0.9419	0.1956
157	5	1	0.3051	0.0918	1.1419	0.2797
167	4	1	0.2289	0.0954	1.3919	0.3752

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

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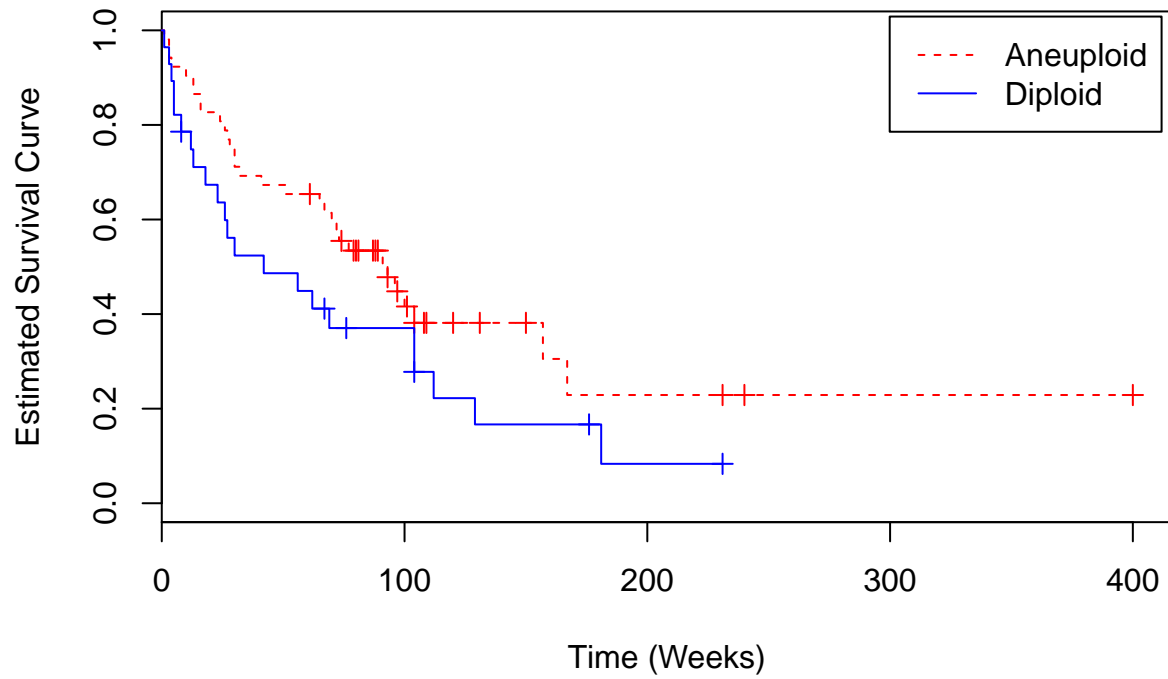
And below is a plot of the estimated cumulative hazard curves for the two groups. Note, these curves are not the Nelson-Aalen estimators $\widetilde{H}(t_i)$. These are the transformations of the Kaplan-Meier estimates: $\hat{H}(t_i) = -ln(\hat{S}(t_i))$. Since the sample sizes are not small and the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

Estimated Cumulative Hazard Curves
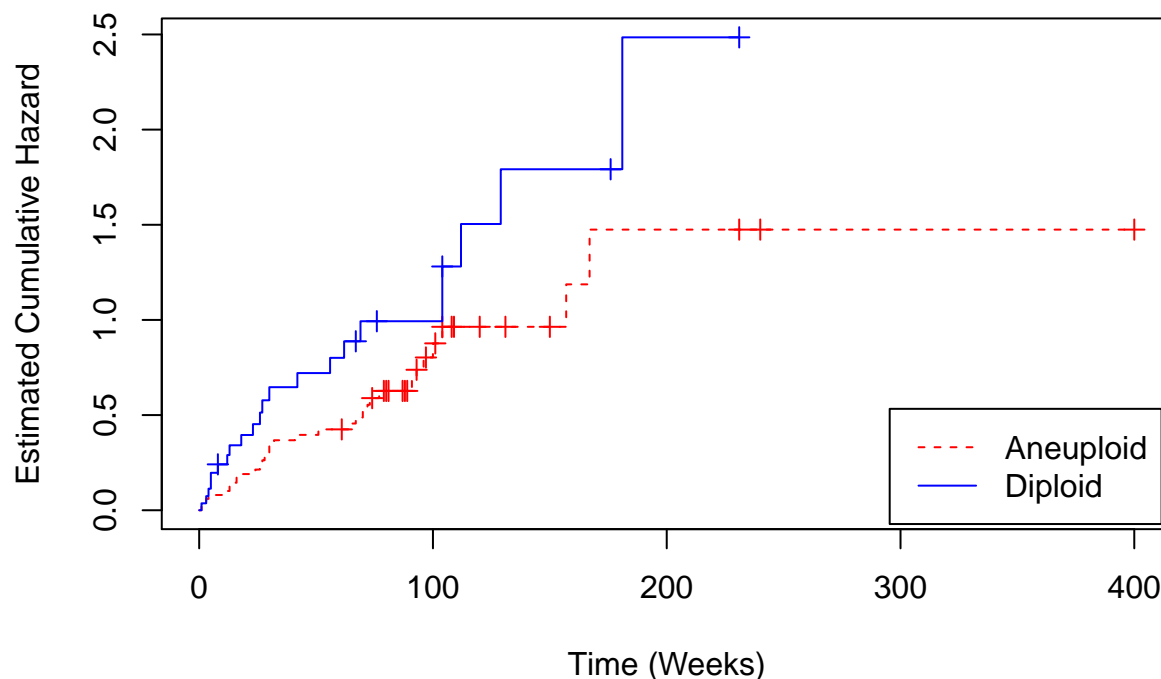
Both cumulative hazard curves appear somewhat linear for the first 180 weeks approximately. Using this time span, we can approximate the hazard rate for each group during this time span. The units will be in events per week:

Estimated hazard rate for diploid group:

$$\frac{2.16 - 0.0357}{181 - 1} = 0.01180167 \text{ events per week}$$

Estimated hazard rate for aneuploid group:

$$\frac{1.3919 - 0.0192}{167 - 1} = 0.008269277 \text{ events per week}$$

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.

I did this problem similar to the same part of the previous problem. In this problem I will estimate up until 250 weeks.

Here is a table of my results:

Group	Mean Restricted to 250 weeks	Standard Error	95% Confidence Interval
Aneuploid	112.1	14.6	[83.484, 140.716]
Diploid	73.6	15.4	[43.416, 103.784]

Again, these are just estimations.

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.

Diploid Group

time	surv	std.err	linear	log	arcsine
1	0.964	0.035	13.239	2.947	6.300
3	0.929	0.049	8.806	3.161	5.449
4	0.893	0.058	6.721	3.135	4.783
5	0.821	0.072	4.441	2.812	3.695
8	0.786	0.078	3.685	2.580	3.219
12	0.748	0.082	3.014	2.295	2.738
13	0.711	0.086	2.442	1.991	2.286
18	0.673	0.090	1.937	1.670	1.856
23	0.636	0.092	1.478	1.333	1.440
26	0.599	0.094	1.050	0.984	1.036
27	0.561	0.095	0.643	0.621	0.640
30	0.524	0.096	0.248	0.245	0.248
42	0.486	0.096	-0.142	-0.142	-0.142
56	0.449	0.096	-0.533	-0.542	-0.531
62	0.412	0.095	-0.933	-0.954	-0.923
69	0.370	0.094	-1.381	-1.410	-1.350
104	0.278	0.090	-2.459	-2.418	-2.283
112	0.222	0.088	-3.166	-2.951	-2.791
129	0.167	0.082	-4.089	-3.479	-3.336
181	0.083	0.072	-5.815	-3.690	-3.800

The estimated median is 42 weeks. In the final three columns, we look for values that are between -1.96 and 1.96. The 95% confidence intervals for the median are:

Type	Interval
Linear	[18, 69]
Log	[18, 69]
Arcsine	[18, 69]

Aneuploid Group

time	surv	std.err	linear	log	arcsine
1	0.981	0.019	25.244	3.575	9.321
3	0.942	0.032	13.680	4.254	7.829
4	0.923	0.037	11.449	4.316	7.274
10	0.904	0.041	9.878	4.303	6.780
13	0.865	0.047	7.720	4.143	5.909
16	0.827	0.052	6.232	3.876	5.139
24	0.808	0.055	5.630	3.716	4.780
26	0.788	0.057	5.093	3.542	4.435
27	0.769	0.058	4.608	3.356	4.100
28	0.750	0.060	4.163	3.160	3.776
30	0.712	0.063	3.367	2.742	3.150
32	0.692	0.064	3.005	2.521	2.847
41	0.673	0.065	2.661	2.294	2.549
51	0.654	0.066	2.332	2.061	2.255
65	0.634	0.067	2.004	1.812	1.955
67	0.614	0.068	1.688	1.558	1.658
70	0.594	0.068	1.382	1.299	1.365
72	0.575	0.069	1.083	1.035	1.075
73	0.555	0.069	0.791	0.767	0.788
77	0.534	0.070	0.491	0.483	0.490
91	0.506	0.071	0.086	0.085	0.086
93	0.478	0.073	-0.302	-0.305	-0.302
96	0.448	0.074	-0.700	-0.712	-0.697
100	0.416	0.075	-1.112	-1.137	-1.101
104	0.381	0.077	-1.546	-1.580	-1.516
157	0.305	0.092	-2.123	-2.123	-2.009
167	0.229	0.095	-2.842	-2.670	-2.524

The estimated median is 93 weeks. In the final three columns, we look for values that are between -1.96 and 1.96. The 95% confidence intervals for the median are:

Type	Interval
Linear	[67, 104]
Log	[65, 104]
Arcsine	[65, 104]

f) Using the log and arcsine-square root transformed confidence intervals, find a 95% confidence interval for the survival function, at 26 weeks, for both the diploid and aneuploid groups.

Diploid

For this group, we can get the values we need from the following table (the complete version of this table is actually in the next problem):

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2
10	0.786	0.078	0.010
12	0.748	0.082	0.012
13	0.711	0.086	0.015
18	0.673	0.090	0.018
23	0.636	0.092	0.021
26	0.599	0.094	0.025
27	0.561	0.095	0.029
30	0.524	0.096	0.034

The values we need are $\hat{S}(26) = 0.599$ and $\sigma_S^2(26) = 0.025$, or $\sigma_S(26) = 0.1581139$.

The table of results is:

Group & Interval Type	Interval
Diploid, Log	[0.3913, 0.7558]
Diploid, Arcsine	[0.4108, 0.7732]

Aneuploid

For this group, we can get the values we need from the following table (the complete version of this table is actually in the next problem):

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2
10	0.904	0.041	0.002
13	0.865	0.047	0.003
16	0.827	0.052	0.004
24	0.808	0.055	0.005
26	0.788	0.057	0.005
27	0.769	0.058	0.006
28	0.750	0.060	0.006
30	0.712	0.063	0.008

The values we need are $\hat{S}(26) = 0.788$ and $\sigma_S^2(26) = 0.005$, or $\sigma_S(26) = 0.07071068$.

The table of results is:

Group & Interval Type	Interval
Aneuploid, Log	[0.6529, 0.8753]
Aneuploid, Arcsine	[0.6699, 0.8857]

g) Using the linear formula, the log-transformed formula, and the arcsine-square root formula, find the 95% EP confidence band for the survival function for both the diploid and aneuploid groups over the range 13-104 weeks.
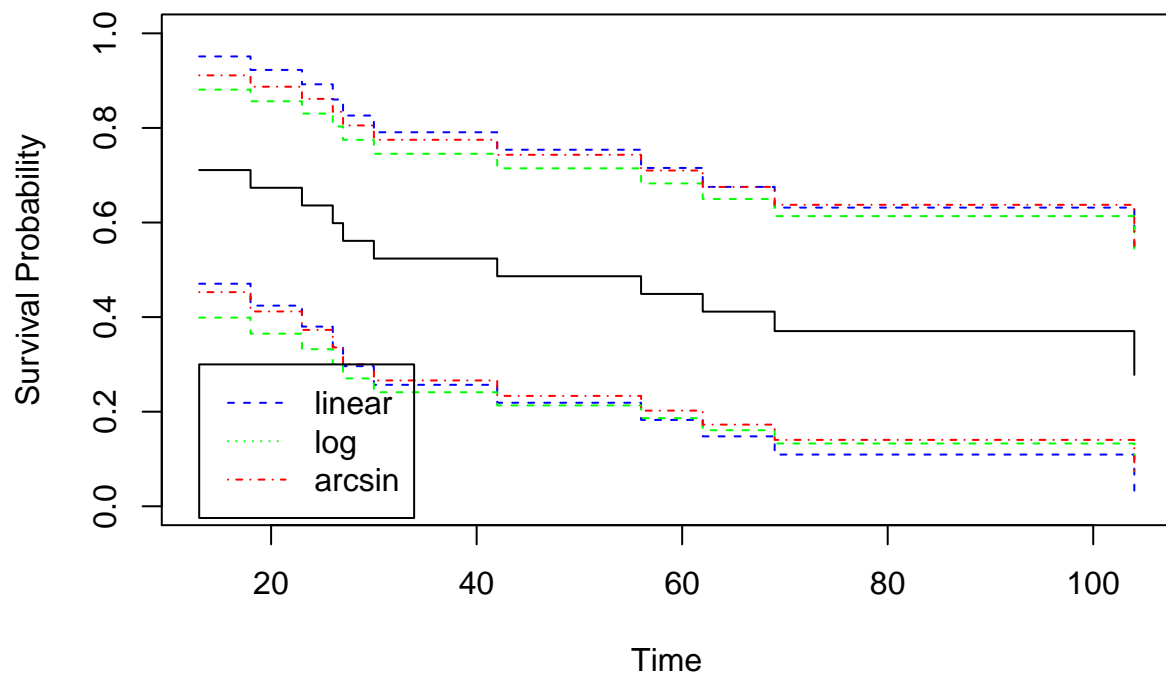
I used a bunch of R code to get these results:

Diploid

The values for a_L and a_U are $a_L = 0.2923251$ and $a_U = 0.7476421$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 2.7831$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
13	0.711	0.086	0.015	0.471	0.951	0.399	0.881	0.453	0.911
18	0.673	0.090	0.018	0.424	0.923	0.365	0.856	0.412	0.887
23	0.636	0.092	0.021	0.380	0.892	0.332	0.830	0.373	0.861
26	0.599	0.094	0.025	0.337	0.860	0.301	0.803	0.336	0.834
27	0.561	0.095	0.029	0.296	0.826	0.270	0.775	0.300	0.805
30	0.524	0.096	0.034	0.257	0.791	0.241	0.745	0.266	0.775
42	0.486	0.096	0.039	0.219	0.754	0.213	0.715	0.233	0.743
56	0.449	0.096	0.045	0.183	0.715	0.186	0.683	0.202	0.710
62	0.412	0.095	0.053	0.148	0.675	0.161	0.650	0.173	0.675
69	0.370	0.094	0.064	0.109	0.631	0.133	0.614	0.140	0.637
104	0.278	0.090	0.106	0.026	0.529	0.075	0.532	0.073	0.550
104	0.278	0.090	0.106	0.026	0.529	0.075	0.532	0.073	0.550

And below is a plot comparing the three confidence bands:

Equal Probability Confidence Bands

Legend:
- linear (blue dashed)
- log (green dotted)
- arcsin (red dash-dot)

Y-axis: Survival Probability

X-axis: Time

Aneuploid

The values for a_L and a_U are $a_L = 0.1346154$ and $a_U = 0.6777492$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 2.8784$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

t_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
13	0.865	0.047	0.003	0.729	1.000	0.651	0.953	0.704	0.969
16	0.827	0.052	0.004	0.676	0.978	0.608	0.930	0.654	0.948
24	0.808	0.055	0.005	0.650	0.965	0.588	0.918	0.630	0.937
26	0.788	0.057	0.005	0.625	0.951	0.567	0.905	0.607	0.925
27	0.769	0.058	0.006	0.601	0.937	0.547	0.892	0.584	0.912
28	0.750	0.060	0.006	0.577	0.923	0.527	0.879	0.562	0.899
30	0.712	0.063	0.008	0.531	0.892	0.488	0.851	0.519	0.871
32	0.692	0.064	0.009	0.508	0.877	0.468	0.837	0.498	0.857
41	0.673	0.065	0.009	0.486	0.860	0.450	0.822	0.477	0.842
51	0.654	0.066	0.010	0.464	0.844	0.431	0.807	0.457	0.827
65	0.634	0.067	0.011	0.442	0.827	0.412	0.791	0.436	0.811
67	0.614	0.068	0.012	0.419	0.809	0.393	0.775	0.416	0.795
70	0.594	0.068	0.013	0.398	0.791	0.374	0.759	0.395	0.778
72	0.575	0.069	0.014	0.376	0.773	0.356	0.743	0.376	0.762
73	0.555	0.069	0.016	0.355	0.754	0.338	0.726	0.356	0.745
77	0.534	0.070	0.017	0.334	0.735	0.319	0.709	0.336	0.727
91	0.506	0.071	0.020	0.300	0.712	0.290	0.687	0.306	0.706
93	0.478	0.073	0.023	0.268	0.688	0.263	0.665	0.276	0.683
96	0.448	0.074	0.027	0.235	0.662	0.234	0.642	0.246	0.660
100	0.416	0.075	0.033	0.199	0.633	0.204	0.617	0.214	0.634
104	0.381	0.077	0.040	0.161	0.602	0.173	0.589	0.180	0.607
104	0.381	0.077	0.040	0.161	0.602	0.173	0.589	0.180	0.607

And below is a plot comparing the three confidence bands:

Equal Probability Confidence Bands

(y-axis) Survival Probability

linear
log
arcsin

(x-axis) Time
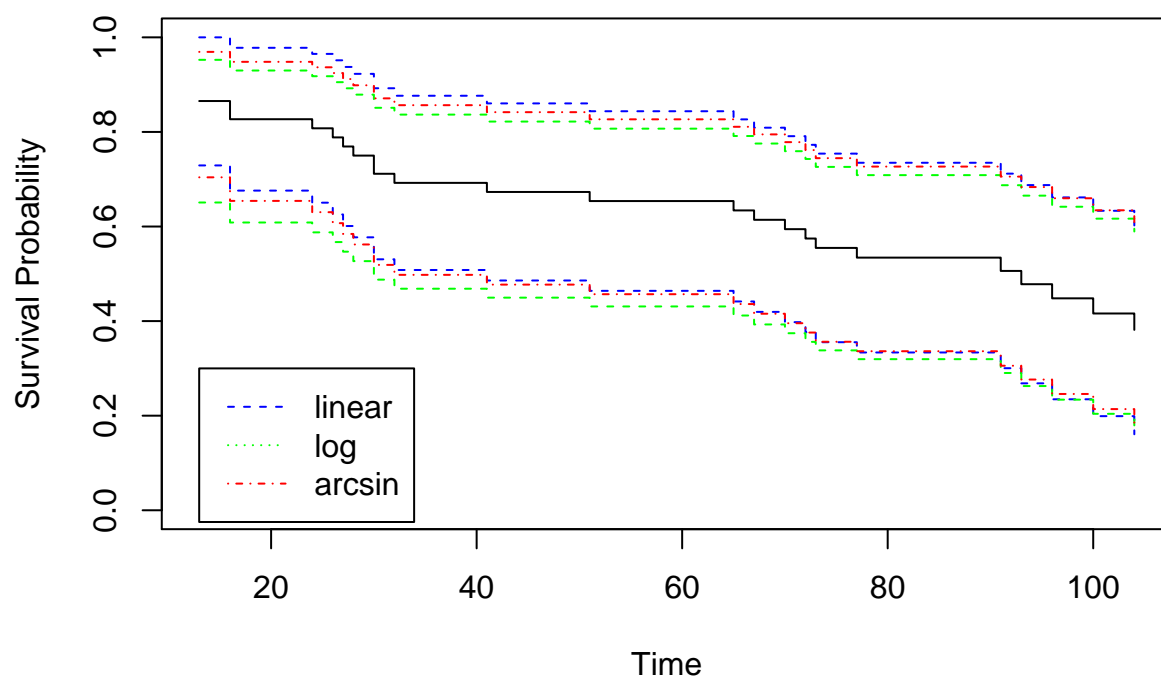
h) Using the linear formula, the log-transformed formula, and the arcsine-square root formula, find a 95% HW confidence band for the survival function for both the diploid and aneuploid groups over the range 0 to 104 weeks.
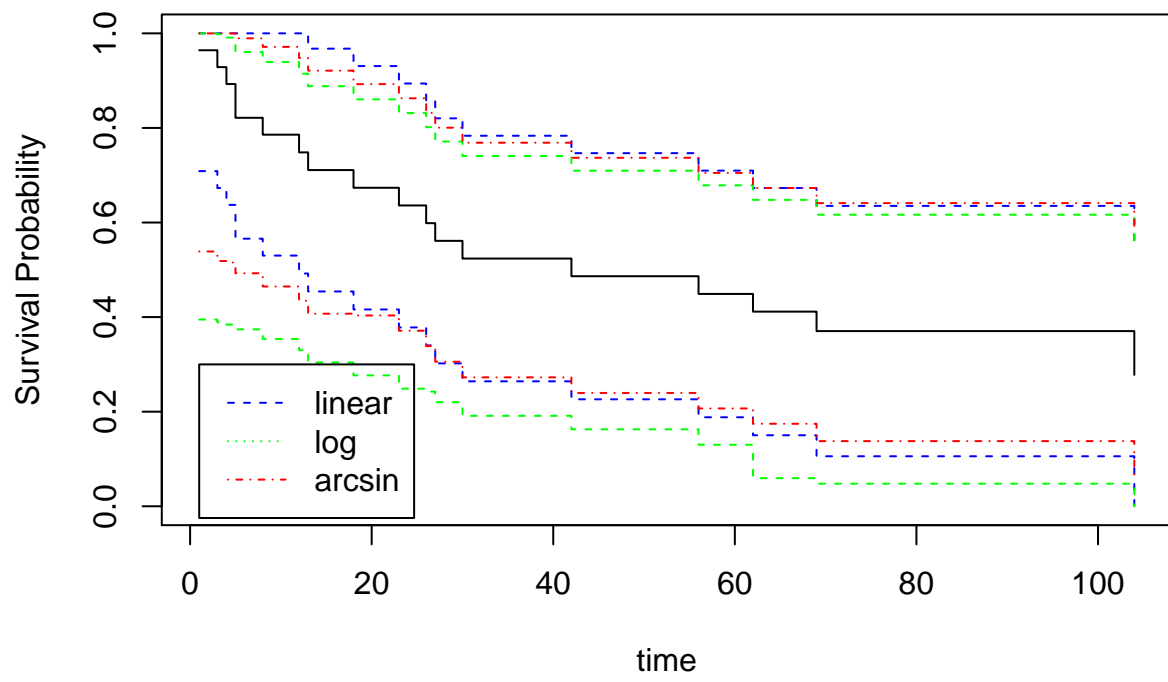
Diploid

The values for a_L and a_U are $a_L = 0.0357143$ and $a_U = 0.7476421$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 1.3525$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

Note: I made one small adjustment and found the confidence band over the range 1 to 104 to avoid a computer error, it should not affect the output very much.

t_i	$\hat{S}(t_i)$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
1	0.964	0.001	0.709	1.000	0.000	1.000	0.407	1.000
3	0.929	0.003	0.673	1.000	0.048	0.998	0.519	1.000
4	0.893	0.004	0.637	1.000	0.242	0.991	0.539	1.000
5	0.821	0.008	0.566	1.000	0.384	0.960	0.515	0.990
8	0.786	0.010	0.530	1.000	0.395	0.939	0.493	0.971
12	0.748	0.012	0.492	1.000	0.389	0.915	0.465	0.948
13	0.711	0.015	0.454	0.968	0.374	0.888	0.435	0.921
18	0.673	0.018	0.416	0.931	0.354	0.860	0.403	0.893
23	0.636	0.021	0.378	0.894	0.330	0.831	0.371	0.863
26	0.599	0.025	0.340	0.857	0.304	0.802	0.339	0.832
27	0.561	0.029	0.302	0.820	0.277	0.771	0.306	0.801
30	0.524	0.034	0.264	0.783	0.249	0.741	0.273	0.769
42	0.486	0.039	0.226	0.747	0.220	0.710	0.240	0.737
56	0.449	0.045	0.188	0.710	0.191	0.679	0.207	0.705
62	0.412	0.053	0.150	0.673	0.163	0.648	0.174	0.673
69	0.370	0.064	0.106	0.635	0.130	0.617	0.138	0.641
104	0.278	0.106	0.000	0.559	0.059	0.559	0.057	0.583
104	0.278	0.106	0.000	0.559	0.059	0.559	0.057	0.583

And below is a plot comparing the three confidence bands:

Hall−Wellner Confidence Bands

Aneuploid

The values for a_L and a_U are $a_L = 0.0192308$ and $a_U = 0.6777492$. These were rounded to the nearest values found in the table. From the table, this means that $c_{0.05} = 1.3434$. Then some behind the scenes R code prints out a table similar to the table from the textbook. Note, "L" and "R" mean the left and right endpoints of the interval.

Note: I made one small adjustment and found the confidence band over the range 1 to 104 to avoid a computer error, it should not affect the output very much.

t_i	$\hat{S}(t_i)$	σ_S^2	Linear L	Linear R	Log L	Log R	Arcsine L	Arcsine R
1	0.981	0.000	0.794	1.000	0.000	1.000	0.468	1.000
3	0.942	0.001	0.756	1.000	0.191	0.998	0.641	1.000
4	0.923	0.002	0.737	1.000	0.369	0.994	0.652	1.000
10	0.904	0.002	0.718	1.000	0.460	0.987	0.652	1.000
13	0.865	0.003	0.679	1.000	0.527	0.968	0.635	0.989
16	0.827	0.004	0.641	1.000	0.537	0.944	0.609	0.967
24	0.808	0.005	0.621	0.994	0.533	0.930	0.595	0.953
26	0.788	0.005	0.602	0.975	0.526	0.916	0.579	0.939
27	0.769	0.006	0.583	0.956	0.517	0.901	0.563	0.924
28	0.750	0.006	0.564	0.936	0.506	0.886	0.547	0.908
30	0.712	0.008	0.525	0.898	0.480	0.854	0.513	0.875
32	0.692	0.009	0.506	0.879	0.466	0.838	0.496	0.858
41	0.673	0.009	0.487	0.859	0.451	0.821	0.478	0.841
51	0.654	0.010	0.468	0.840	0.436	0.805	0.461	0.824
65	0.634	0.011	0.448	0.820	0.419	0.787	0.442	0.806
67	0.614	0.012	0.428	0.801	0.403	0.770	0.424	0.788
70	0.594	0.013	0.408	0.781	0.386	0.753	0.405	0.770
72	0.575	0.014	0.388	0.762	0.369	0.735	0.387	0.752
73	0.555	0.016	0.368	0.742	0.352	0.717	0.368	0.734
77	0.534	0.017	0.347	0.722	0.334	0.699	0.349	0.715
91	0.506	0.020	0.314	0.698	0.305	0.677	0.318	0.693
93	0.478	0.023	0.281	0.675	0.276	0.655	0.288	0.671
96	0.448	0.027	0.246	0.650	0.244	0.633	0.256	0.649
100	0.416	0.033	0.206	0.626	0.210	0.611	0.220	0.627
104	0.381	0.040	0.161	0.602	0.173	0.589	0.180	0.606
104	0.381	0.040	0.161	0.602	0.173	0.589	0.180	0.606

And below is a plot comparing the three confidence bands:

Hall−Wellner Confidence Bands
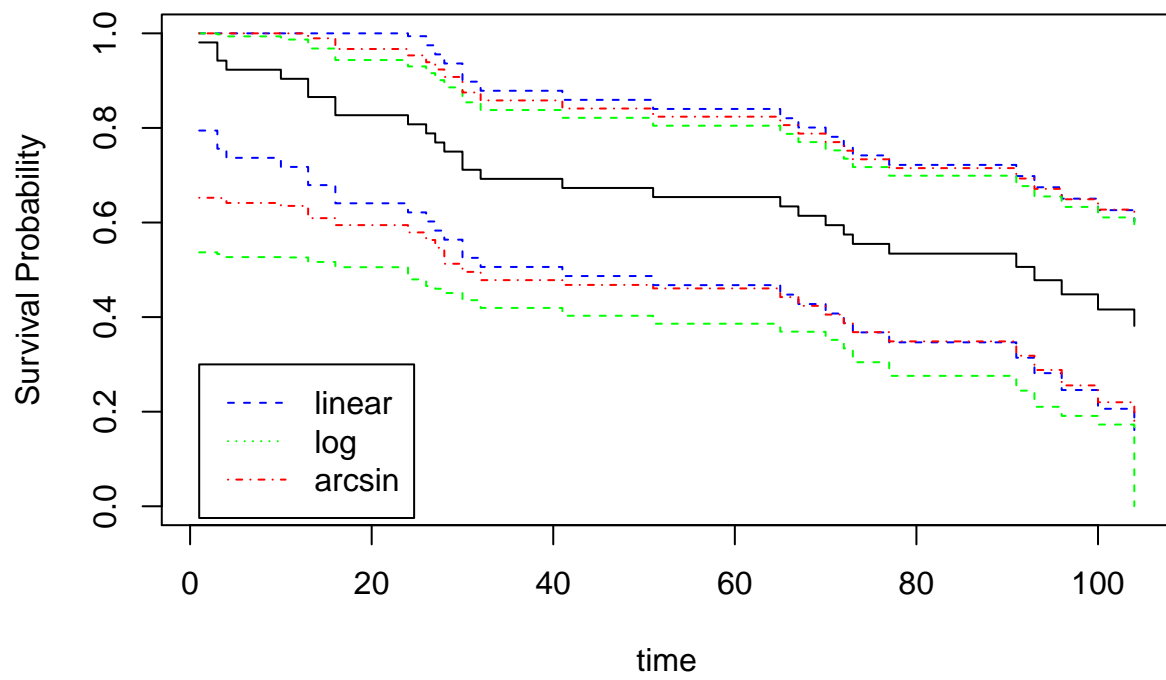
4.4)

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.

t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\widetilde{H}(t_i)$	$\sigma_H(t_i)$
22	25	1	0.9600	0.0392	0.0400	0.0400
27	24	1	0.9200	0.0543	0.0817	0.0578
50	23	1	0.8800	0.0650	0.1251	0.0723
68	22	1	0.8400	0.0733	0.1706	0.0854
134	17	1	0.7906	0.0840	0.2294	0.1037
136	16	1	0.7412	0.0922	0.2919	0.1211
139	15	1	0.6918	0.0984	0.3586	0.1382
144	14	1	0.6424	0.1030	0.4300	0.1556
193	12	1	0.5888	0.1074	0.5134	0.1765
203	10	1	0.5299	0.1117	0.6134	0.2029
210	9	1	0.4711	0.1137	0.7245	0.2313
248	6	1	0.3925	0.1188	0.8911	0.2851
304	3	1	0.2617	0.1330	1.2245	0.4386

Below is a plot of the Kaplan-Meier Curve:

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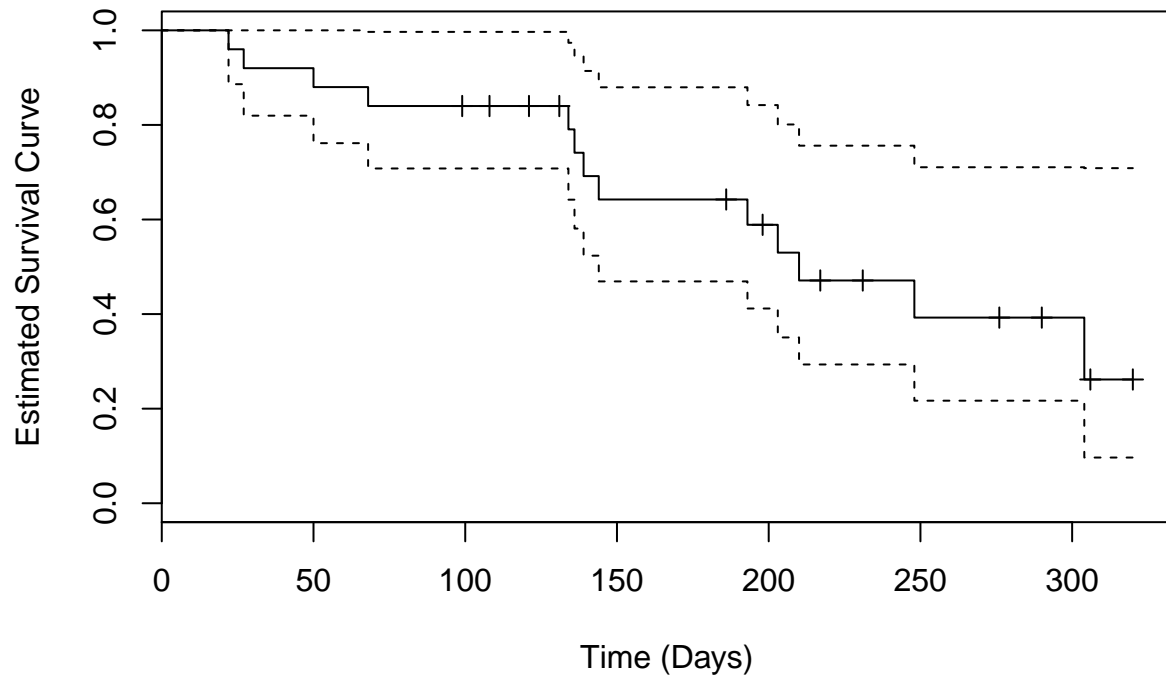
b) Use the Brown, Hollandar, and Kowar technique (Practical Note 3 of section 4.2) to complete the right-hand tail of the product-limit estimate found in part a.

The largest observed time in this set of data is 320 days and it is a censored observation. Thus, the estimate of the curve after $t_{max} = 320$ is:
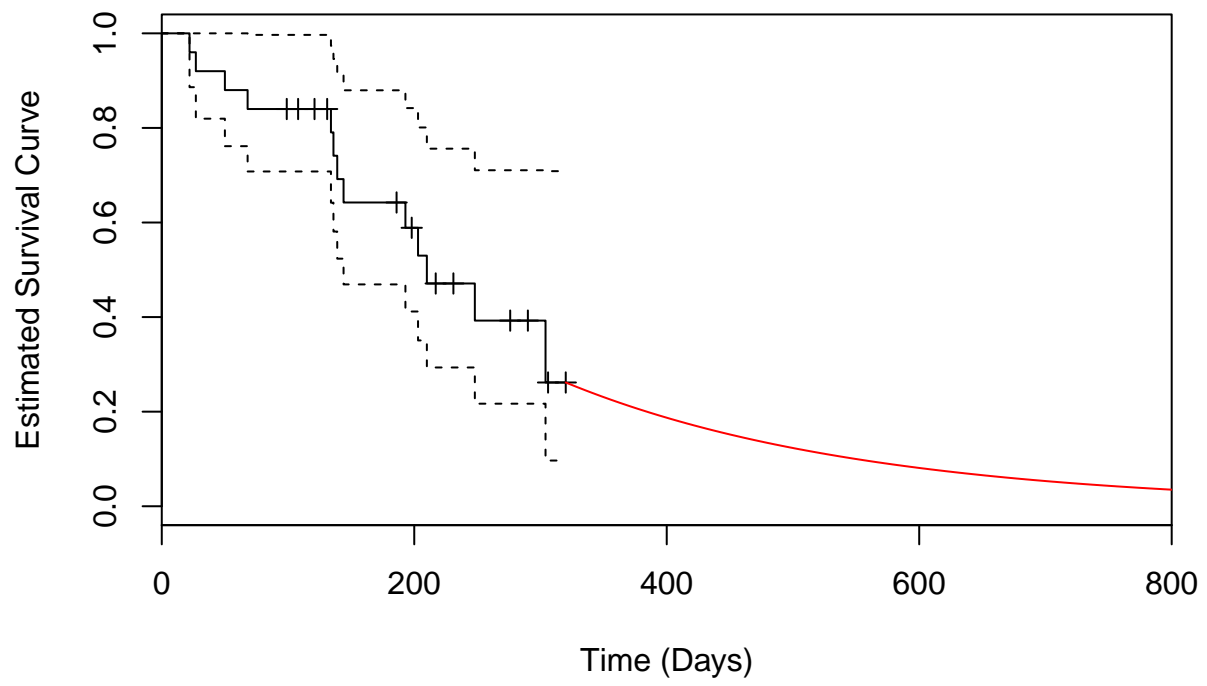
$$\hat{S}(t) = e^{t \frac{ln(\hat{S}(320))}{320}}$$

Since $\hat{S}(320) = 0.2617$, this simplifies to:

$$\hat{S}(t) = exp(-0.004189239(t)) \text{ for } t \geq t_{max}$$

Below is the same plot from above but with the complete right hand tail.

Kaplan−Meier Curve

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.

t_i	Y_i	d_i	$\hat{S}(t_i)$	$\sqrt{\hat{V}[\hat{S}(t_i)]}$	$\tilde{H}(t_i)$	$\sigma_H(t_i)$
22	25	1	0.96	0.0392	0.0400	0.0400
27	24	1	0.92	0.0543	0.0817	0.0578
50	23	1	0.88	0.0650	0.1251	0.0723
68	22	1	0.84	0.0733	0.1706	0.0854
101	21	1	0.80	0.0800	0.2182	0.0978
134	20	1	0.76	0.0854	0.2682	0.1098
136	19	1	0.72	0.0898	0.3209	0.1218
139	18	1	0.68	0.0933	0.3764	0.1339
144	17	1	0.64	0.0960	0.4352	0.1462
191	16	1	0.60	0.0980	0.4977	0.1590
193	15	1	0.56	0.0993	0.5644	0.1724
203	14	1	0.52	0.0999	0.6358	0.1866
210	13	1	0.48	0.0999	0.7127	0.2019
224	12	1	0.44	0.0993	0.7961	0.2184
248	11	1	0.40	0.0980	0.8870	0.2365
256	10	1	0.36	0.0960	0.9870	0.2568
304	9	1	0.32	0.0933	1.0981	0.2798
308	8	1	0.28	0.0898	1.2231	0.3065
441	7	1	0.24	0.0854	1.3660	0.3381
482	6	1	0.20	0.0800	1.5326	0.3770
511	5	1	0.16	0.0733	1.7326	0.4267
561	4	1	0.12	0.0650	1.9826	0.4946
580	3	1	0.08	0.0543	2.3160	0.5964
683	2	1	0.04	0.0392	2.8160	0.7783
724	1	1	0.00	NaN	3.8160	1.2672

Below is a plot of the Kaplan-Meier Curve:

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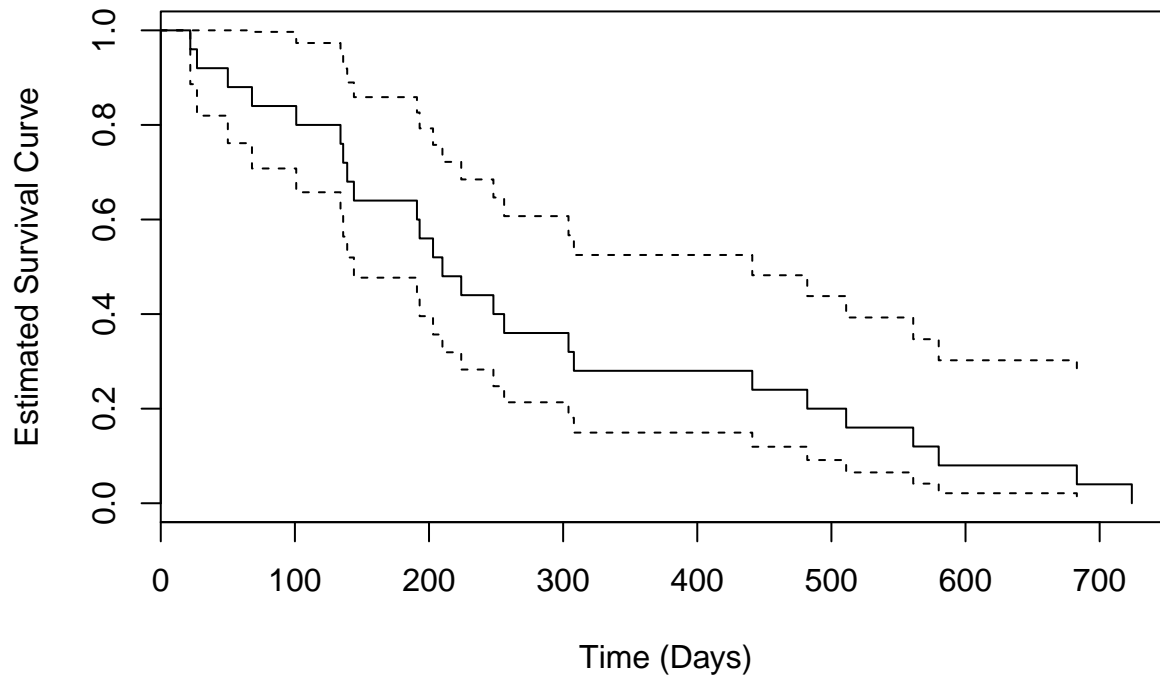
d) Estimate the mean time to death, restricted to 724 days, based on the product-limit estimator found in part a.

Here is a table of my results:

Group	Mean Restricted to 724 Days	Standard Error	Confidence Interval
Lung cancer	317.8	68	[184.52, 451.08]

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.

Here is a table of my results:

Group	Mean Restricted to 724 Days	Standard Error	Confidence Interval
Lung cancer	277.4	40.4	[198.216, 356.584]

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.

The mean is 277.6 and the standard error is 41.2664109. These values are almost identical to the values found in part (e).