Session 2: Lecture notes exercise solutions

Exercise 2.1
Create a suitable table and obtain Kaplan-Meier estimates of the survivor function (a) in the whole group, (b) separately by IM status. The Kaplan-Meier plots are shown in Figure 2.3 - check that you can recreate these from your table.

The results for the whole group are shown below (in Stata and R output). The 'Survivor function' column (Stata) and the 'survival' column (R) give the Kaplan-Meier estimates.

```
. sts list

         failure _d:  death
   analysis time _t:  time
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
19	45	1	0	0.9778	0.0220	0.8525	0.9968
22	44	1	0	0.9556	0.0307	0.8338	0.9887
23	43	1	0	0.9333	0.0372	0.8074	0.9780
25	42	1	0	0.9111	0.0424	0.7803	0.9657
30	41	1	0	0.8889	0.0468	0.7534	0.9522
34	40	1	0	0.8667	0.0507	0.7271	0.9378
37	39	1	0	0.8444	0.0540	0.7012	0.9226
38	38	1	0	0.8222	0.0570	0.6758	0.9068
42	37	1	0	0.8000	0.0596	0.6509	0.8905
46	36	1	0	0.7778	0.0620	0.6264	0.8737
47	35	1	0	0.7556	0.0641	0.6022	0.8565
51	34	1	0	0.7333	0.0659	0.5785	0.8388
56	33	1	0	0.7111	0.0676	0.5551	0.8208
57	32	1	0	0.6889	0.0690	0.5320	0.8025
61	31	1	0	0.6667	0.0703	0.5092	0.7838
66	30	1	0	0.6444	0.0714	0.4867	0.7648
67	29	1	0	0.6222	0.0723	0.4646	0.7455
73	28	1	0	0.6000	0.0730	0.4427	0.7260
74	27	1	0	0.5778	0.0736	0.4211	0.7061
77	26	1	0	0.5556	0.0741	0.3998	0.6860
78	25	1	0	0.5333	0.0744	0.3787	0.6657
86	24	1	0	0.5111	0.0745	0.3579	0.6450
89	23	1	0	0.4889	0.0745	0.3374	0.6241
115	22	1	0	0.4667	0.0744	0.3172	0.6030
122	21	0	1	0.4667	0.0744	0.3172	0.6030
123	20	0	1	0.4667	0.0744	0.3172	0.6030

```
   130       19       0       2              0.4667    0.0744      0.3172    0.6030
   133       17       0       1              0.4667    0.0744      0.3172    0.6030
   134       16       0       1              0.4667    0.0744      0.3172    0.6030
   136       15       0       1              0.4667    0.0744      0.3172    0.6030
   141       14       0       1              0.4667    0.0744      0.3172    0.6030
   143       13       0       1              0.4667    0.0744      0.3172    0.6030
   144       12       0       1              0.4667    0.0744      0.3172    0.6030
   148       11       0       1              0.4667    0.0744      0.3172    0.6030
   151       10       0       1              0.4667    0.0744      0.3172    0.6030
   152        9       0       1              0.4667    0.0744      0.3172    0.6030
   153        8       0       1              0.4667    0.0744      0.3172    0.6030
   154        7       0       1              0.4667    0.0744      0.3172    0.6030
   156        6       0       1              0.4667    0.0744      0.3172    0.6030
   162        5       0       1              0.4667    0.0744      0.3172    0.6030
   164        4       0       1              0.4667    0.0744      0.3172    0.6030
   165        3       0       1              0.4667    0.0744      0.3172    0.6030
   182        2       0       1              0.4667    0.0744      0.3172    0.6030
   189        1       0       1              0.4667    0.0744      0.3172    0.6030
--------------------------------------------------------------------------------

> btrial.km<-survfit(Surv(time=time,event=death)~1,data=breastcancer)
> summary(btrial.km,censored = T)
Call: survfit(formula = Surv(time = time, event = death) ~ 1, data = breastcancer)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   19     45       1    0.978  0.0220        0.936        1.000
   22     44       1    0.956  0.0307        0.897        1.000
   23     43       1    0.933  0.0372        0.863        1.000
   25     42       1    0.911  0.0424        0.832        0.998
   30     41       1    0.889  0.0468        0.802        0.986
   34     40       1    0.867  0.0507        0.773        0.972
   37     39       1    0.844  0.0540        0.745        0.957
   38     38       1    0.822  0.0570        0.718        0.942
   42     37       1    0.800  0.0596        0.691        0.926
   46     36       1    0.778  0.0620        0.665        0.909
   47     35       1    0.756  0.0641        0.640        0.892
   51     34       1    0.733  0.0659        0.615        0.875
   56     33       1    0.711  0.0676        0.590        0.857
   57     32       1    0.689  0.0690        0.566        0.838
   61     31       1    0.667  0.0703        0.542        0.820
   66     30       1    0.644  0.0714        0.519        0.801
   67     29       1    0.622  0.0723        0.496        0.781
   73     28       1    0.600  0.0730        0.473        0.762
   74     27       1    0.578  0.0736        0.450        0.742
   77     26       1    0.556  0.0741        0.428        0.721
```

78	25	1	0.533	0.0744	0.406	0.701
86	24	1	0.511	0.0745	0.384	0.680
89	23	1	0.489	0.0745	0.363	0.659
115	22	1	0.467	0.0744	0.341	0.638
122	21	0	0.467	0.0744	0.341	0.638
123	20	0	0.467	0.0744	0.341	0.638
130	19	0	0.467	0.0744	0.341	0.638
133	17	0	0.467	0.0744	0.341	0.638
134	16	0	0.467	0.0744	0.341	0.638
136	15	0	0.467	0.0744	0.341	0.638
141	14	0	0.467	0.0744	0.341	0.638
143	13	0	0.467	0.0744	0.341	0.638
144	12	0	0.467	0.0744	0.341	0.638
148	11	0	0.467	0.0744	0.341	0.638
151	10	0	0.467	0.0744	0.341	0.638
152	9	0	0.467	0.0744	0.341	0.638
153	8	0	0.467	0.0744	0.341	0.638
154	7	0	0.467	0.0744	0.341	0.638
156	6	0	0.467	0.0744	0.341	0.638
162	5	0	0.467	0.0744	0.341	0.638
164	4	0	0.467	0.0744	0.341	0.638
165	3	0	0.467	0.0744	0.341	0.638
182	2	0	0.467	0.0744	0.341	0.638
189	1	0	0.467	0.0744	0.341	0.638

Exercise 2.2
(a) Complete table 2.5 and compare your results with those shown graphically in Figure 2.5. (b) What was the probability of surviving 5 years or more? What was the probability of surviving 2.5 years or more?

In the first row you should get $p_1 = 456/2418 = 0.188$, giving the survivor function 0.814.

In the second row you should get $p_2 = 226/(1962 - 39/2) = 0.116$, giving the survivor function 0.720 (after multiplying (1-0.116) by 0.814).

etc.

The probability of surviving 5 years or more is 0.521 - see the survivor function in row 5 of your table.

The probability of surviving 2.5 years or more is 0.720 - note that the survivor function is flat at that time.

Exercise 2.3
State the null hypothesis for the log rank test applied to the breast cancer example. Interpret the above results from performing the log rank test.

The null hypothesis is that the survivor curves in the two exposure groups (IM positive

and negative) are the same. The log rank test statistic is 5.49. Comparing this with the χ^2_1 distribution gives a p-value of 0.0191. There is quite strong evidence against the null hypothesis.

Bonus exercise: Here is how to derive the result for a 95% CI given in equation (2.17)

The 95% confidence interval for $\log\{-\log S(t)\}$ is such that

$$\log\{-\log \hat{S}(t)\} - 1.96v(t) < \log\{-\log S(t)\} < \log\{-\log \hat{S}(t)\} + 1.96v(t).$$

Taking the exponential gives:

$$\{-\log \hat{S}(t)\}\exp(-1.96v(t)) < -\log S(t) < \{-\log \hat{S}(t)\}\exp(1.96v(t)).$$

We now multiply everything by -1, giving:

$$\log\{\hat{S}(t)\}\exp(1.96v(t)) < \log S(t) < \log\{\hat{S}(t)\}\exp(-1.96v(t)).$$

Finally, take the exponential again:

$$\hat{S}(t)^{\exp(1.96v(t))} < S(t) < \hat{S}(t)^{\exp(-1.96v(t))}.$$