Biostatistics 513

Homework 6 Solutions

1) a) from Dickson et al (1989) "the initial time point for survival modeling was the date of determination of eligibility for the trials"
b) an event is a death from any cause
c) censoring can be due to end of study (July, 1986; n = 160), liver transplantation (n = 19), or dropout (n = 8)
d) It is likely that censoring due to end of study is uninformative. Censoring due to dropout is unknown … why did they dropout? Fortunately, the number of dropouts is small compared with the number of events (125 deaths) so it shouldn't have much effect. It is also unclear whether or not censoring due to liver transplantation is unrelated to risk of death. If liver transplants were done because these individuals were at particularly high risk of death, then this is informative censoring (that's bad); but if they received liver transplants simply because a liver became available (and they had been waiting a long time) then perhaps this cause of censoring is uninformative (that's good).
e) low is about 15% death; medium is about 50% death; high is about 80% death

2) Here's my table

Time (days) t_i	At Risk R_i	Died d_i	Censored l_i	Failure Probability d_i/R_i	Conditional Survival $(1-d_i/R_i)$	K-M estimate $\prod(1-d_i/R_i)$
42	12	1	0	.0833	.9167	.917
53	11	1	0	.0909	.9091	.833
57	10	1	0	.1000	.9000	.750
63	9	1	0	.1111	.8889	.667
81	8	1	0	.1250	.8750	.583
140	7	1	0	.1428	.8572	.500
176	6	1	0	.1667	.8333	.417
210	5	0	1	0	1	.417
252	4	1	0	.2500	.7500	.312
476	3	0	1	0	1	.312
524	2	1	0	.5000	.5000	.156
1037	1	0	1	0	1	.156

3) a) . summarize age year stage, detail

```
                              age
-------------------------------------------------------------
      Percentiles      Smallest
 1%           41             41
 5%           47             43
10%           49             45       Obs                  90
25%           57             47       Sum of Wgt.          90

50%           65                      Mean           64.61111
                        Largest       Std. Dev.      10.79606
75%           72             82
90%         78.5             84       Variance       116.5549
95%           81             86       Skewness      -.1316666
99%           86             86       Kurtosis       2.243943
```

. tabu year

year of entry	Freq.	Percent	Cum.
70	2	2.22	2.22
71	12	13.33	15.56
72	9	10.00	25.56
73	11	12.22	37.78
74	14	15.56	53.33
75	8	8.89	62.22
76	19	21.11	83.33
77	11	12.22	95.56
78	4	4.44	100.00
Total	90	100.00	

. tabu stage

stage at diagnosis	Freq.	Percent	Cum.
1	33	36.67	36.67
2	17	18.89	55.56
3	27	30.00	85.56
4	13	14.44	100.00
Total	90	100.00	

b) . tabu status stage

	stage at diagnosis				
status	1	2	3	4	Total
alive	18	10	10	2	40
died	15	7	17	11	50
Total	33	17	27	13	90

Inspection of this table suggests that higher stage is associated with greater risk of death.

c) . tabu year stage

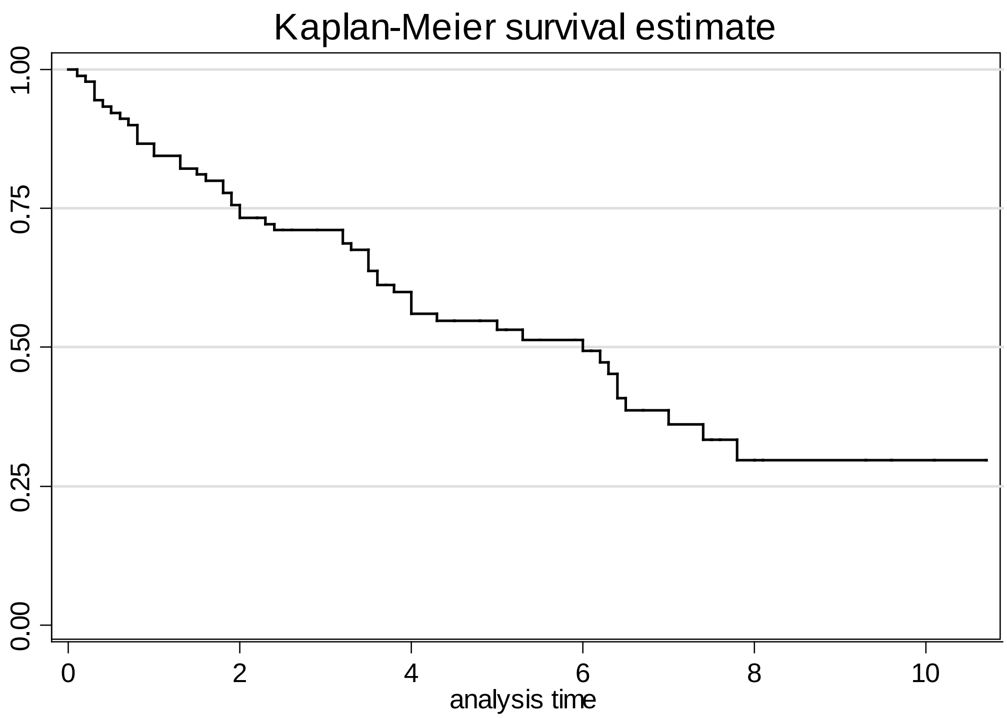
year of	stage at diagnosis				
entry	1	2	3	4	Total
70	2	0	0	0	2
71	6	2	3	1	12
72	2	1	5	1	9
73	5	3	2	1	11
74	6	1	6	1	14
75	5	1	1	1	8
76	4	3	8	4	19
77	2	4	2	3	11
78	1	2	0	1	4
Total	33	17	27	13	90

I don't see any obvious relationship between year and stage - i.e. no secular trends.

```
. sort stage
. by stage: summarize age

__  Variable |       Obs        Mean    Std. Dev.       Min        Max
-----------------------------------------------------------------------------
-> stage = 1
         age |        33    64.18182    11.25959         43         86

-> stage = 2
         age |        17    64.82353    10.64328         47         86

-> stage = 3
         age |        27    63.81481    10.29203         49         82

-> stage = 4
         age |        13    67.07692     11.7151         41         84
```

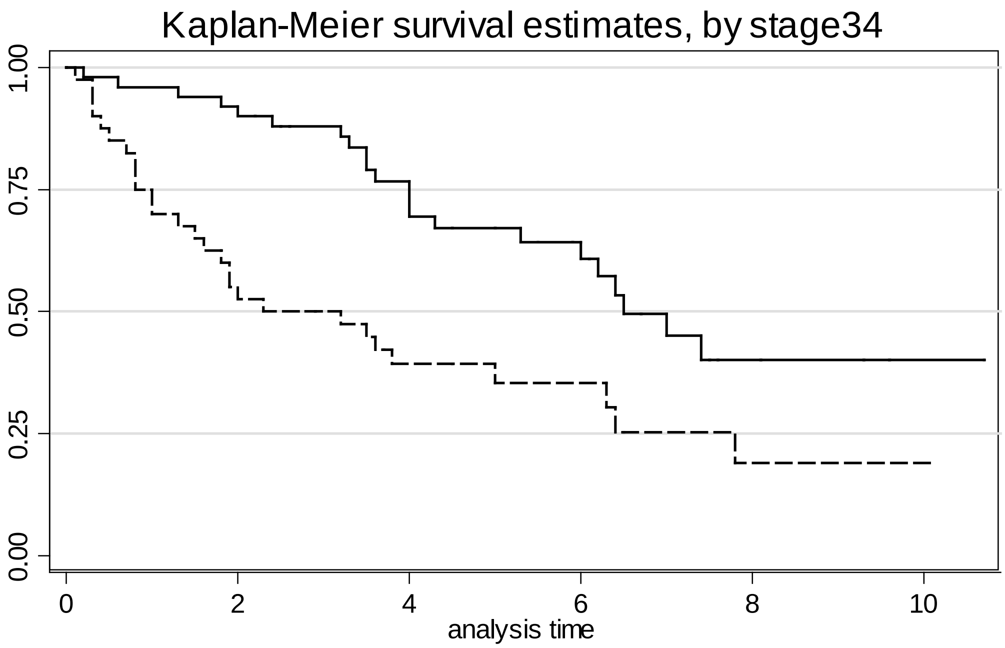
Also, no obvious relationship between age and stage.

d)



At t = 6 the KM estimate of S(t) is 0.49. This is the estimated proportion of individuals with laryngeal cancer surviving at time 6 years following first treatment.

e)



[redacted]

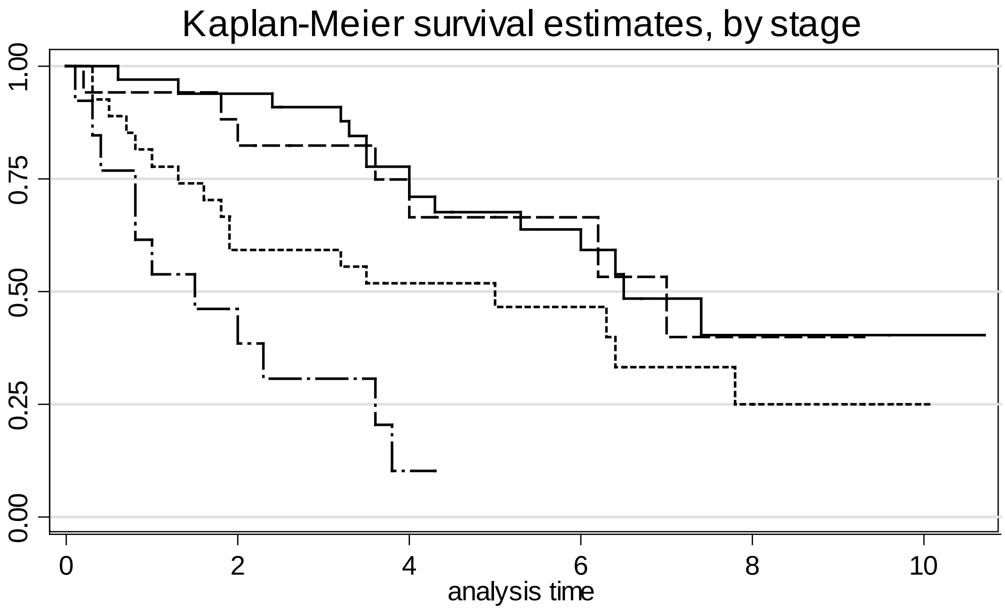
It appears as though the prognosis is better for those diagnosed with stage 1 and 2 disease (i.e. stage34=0).

f)
The hypotheses are

Ho: $S_{1/2}(t) = S_{3/4}(t)$ (survival is the same in the two groups)
Ha: $S_{1/2}(t) \neq S_{3/4}(t)$ (survival is not the same in the two groups)

```
. sts test stage34

         failure _d:  status
   analysis time _t:  time


Log-rank test for equality of survivor functions

        |   Events         Events
stage34 |  observed       expected
--------+-------------------------
0       |        22          32.58
1       |        28          17.42
--------+-------------------------
Total   |        50          50.00

              chi2(1) =      10.13
              Pr>chi2 =     0.0015
```

I conclude that the survivor functions are significantly different for states 1/2 vs 3/4. From the observed versus expected numbers we can see that survival is relatively higher in the low stage groups and relatively lower in the high stage groups g)



[redacted]

h) Again use the log-rank test with null hypothesis

Ho: $S_1(t) = S_2(t) = S_3(t) = S_4(t)$
Ha: at least one $\neq$

Results are

```
. sts test stage

         failure _d:  status
   analysis time _t:  time


Log-rank test for equality of survivor functions

      |   Events         Events
stage |  observed       expected
------+-------------------------
1     |        15          22.57
2     |         7          10.01
3     |        17          14.08
4     |        11           3.34
------+-------------------------
Total |        50          50.00

            chi2(3) =      22.76
            Pr>chi2 =     0.0000
```

And we conclude that there is a highly significant difference in survival between stages. Note that this chi-square test has 3 df (one less than the number of groups)