Because we did not have a homework assignment on survival analysis and weighting methods, I have provided a few basic problems for you to use as practice for the final. Problems on these topics will be more conceptual than calculational given the limited amount of time we spent on the material. In particular I wouldn’t expect you to be able to do all the calculations for the survival curve in Problem 3 other than as a bonus problem but I thought it would be useful for you to have an example for future reference and you should definitely understand the conceptual issues associated with this problem. Note that some of the previous year’s finals do include a little more in the way of survival calculations because I got in an extra lecture on the topic those years—just focus on the conceptual pieces from those problems.

(1) Survival Analysis Basics:

(a) Explain what the term censoring means, some different ways in which it can occur, and why it is an important issue in analyses of time to event or survival data.

(b) Give an intuitive description of the key quantities in survival analysis (the survival function, S(t), the cumulative death function, F(t), the death density, f(t) and the hazard function, h(t) = f(t)/S(t)).

(c) Describe the two major categories of approaches to estimating the survival quantities above.

(d) Explain briefly the ideas behind the accelerated failure time model and the Cox proportional hazards model for incorporating covariates into a survival model.

(2) Weighted Analysis Basics:

(a) Give four examples of situations in which you might want to perform a weighted analysis.

(b) Describe briefly some of the ways in which you could incorporate weights into an estimation procedure.

(c) Define as carefully as you can what a propensity score is and give two examples of when and how you might want to use it.

(3) Survival Analysis Example: The numbers below represent the survival times in years for two groups with a rare disease. The values with (+) are censored. Use the values to answer the following questions. (Note: You should know how to do the calculations in this problem manually but it’s much easier in STATA or SAS. In STATA the relevant commands are under the “sts” header. You use “sts set” to tell STATA what your time and censoring variables are. You use “sts graph, by(group)” to get survival curves, separated by groups if you like. You use “sts test time group, logrank” to obtain the log rank test. I would never make the calculations this messy on the exam—I might simply give you the tables of calculations like the one on the class handout and ask you to explain how one of the values was computed....)

Group 0: 1.33+, 2.21, 2.80+, 3.92, 5.44+, 9.99, 12.01, 12.07, 17.31, 28.65
Group 1: 0.122+, 0.43, .64, 1.58, 2.02, 3.08, 3.62+, 4.33, 5.52+, 11.86

(a) Show how to find the Kaplan-Meier estimates of the survival curves for these two groups and sketch the first few values. Does it appear that one group does better than the other? (The complete curves are shown
in the accompanying graphics file for your reference.)

(b) Give the estimated 3-year survival probabilities, \( S_1(3) \) and \( S_2(3) \) based on these data along with the corresponding confidence intervals.

(c) Estimate the median survival time for people in these two groups.

(d) Test whether there is a difference in 3-year survival probability for these two groups using (a) a Z-test and (b) a confidence interval for the difference between the survival curves.

(e) The printout below shows the log-rank test comparing these two groups. Explain what hypotheses are being tested and give your real-world conclusions.

Log-rank test for equality of survivor functions

<table>
<thead>
<tr>
<th>group</th>
<th>Events</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed</td>
<td>expected</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>10.47</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>3.53</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>14.00</td>
</tr>
</tbody>
</table>

\[ \text{chi}^2(1) = 5.51 \]
\[ \text{Pr}>\text{chi}^2 = 0.0189 \]

(f) Suppose none of the observations had been censored. What would your estimates of the 3-year survival probabilities have been and why?