1. The data in table D.2 on page 365 in Collett’s book, gives times to first recurrence of a tumor in bladder cancer patients. These data come from a randomized clinical trial conducted by the Veteran’s Administration cooperative Urological Research Group. Patients with superficial bladder tumors first had their tumors removed. They were next randomized to either a placebo or a chemotherapeutic agent called thiotepa. The initial number of tumors and the diameter of the largest tumor in each patient were recorded at the time of randomization to treatment. The response is time to first recurrence of a tumor, in months. Times for patients who had not experienced a recurrence by the end of the follow-up period are right censored.

The data are posted in the file bladder.dat in the assignment section of the course web page. SAS code for reading the file and fitting a particular proportional hazards model is posted in the file bladder1.sas. S-Plus code for reading the data into a data frame and fitting a particular proportional hazards model is posted in the file bladder1.ssc. You should be able to modify one of files to obtain answers to the following questions. There are 86 lines in the data file, one for each subject, with the values for six variables on each line in the following order.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time</th>
<th>Status</th>
<th>Treat</th>
<th>Init</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient identification code (1-86)</td>
<td>Time to first recurrence of a bladder tumor (months)</td>
<td>Censoring Status (0 = censored, 1 = actual time)</td>
<td>Treatment group (1 = placebo, 2 = thiotepa)</td>
<td>Initial number of tumors</td>
<td>Diameter of largest initial tumor (in cm)</td>
</tr>
</tbody>
</table>

(a) First fit a Cox proportional hazards model using the treatment factor as the only covariate. To do this, create a new variable.

\[ x_1 = \begin{cases} 
0 & \text{placebo} \\
1 & \text{treated with thiotepa} 
\end{cases} \]
Then the hazard function for the i-th individual is

\[ h_i(t) = h_0(t)e^{\beta X_i} \]

(i) Report the partial likelihood estimates of \( \beta_1 \) and its standard errors.

(ii) Report estimates of \( \exp(\beta_1) \) and interpret this quantity in the context of this bladder cancer study.

(b) Fit the following Cox proportional hazards model:

\[ h_i(t) = h_0(t)e^{\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}} \]

where \( X_{i1} \) is the initial number of tumors and \( X_{i3} \) is the diameter of the largest initial tumor for the i-th patient.

(i) Report partial likelihood estimates of \( \beta_1, \beta_2, \beta_3 \) and their standard errors.

(ii) After adjusting for the number of initial tumors and the diameter of the largest initial tumor, is there a significant treatment effect? Use \( a = .05 \) for your Type I error level.

(iii) Report estimates of \( \exp(\beta_1), \exp(\beta_2) \) and \( \exp(\beta_3) \) and interpret these quantities in the context of this bladder cancer study.

(iv) Construct a 95% confidence interval for \( \exp(\beta_1) \) and compare it with the corresponding confidence interval obtained from fitting a Weibull proportional hazards model.

(c) Using a Cox proportional hazards model, is there any evidence of interaction between the treatment factor and either the number of initial tumors or the diameter of the largest initial tumor? If so, search for and report results from the model you think best describes these data. Interpret the effects of treatment with thiotepa for the model you select.

2. A study is being planned to examine the possible effects of the anti-viral drug, interferon, to delay infections in patients with chronic granulomatous disease (CGD). CGD is a group of inherited disorders of the immune system that render patients vulnerable to certain infections. CGD patients will be randomized to either a placebo or the interferon treatment. The end-point of interest in this study is time to first serious infection. Past experience with untreated CGD patients suggests that survival probabilities for the placebo group will approximately be as follows:
Patients will be recruited for during a 120 day accrual period. This will be followed by a 240 day follow-up period. What is the number of patients needed to have power .90 of observing a significance difference between the estimated survivor functions using a log-rank test at the $\alpha = 0.05$ level of significance? Assume that patients are recruited uniformly across time during the accrual period and 50% of the patients are randomized to each treatment group. CGD patients who do not exhibit an infection by the end of the follow-up period have right censored times. In addition, assume an exponential distribution for random loss to follow-up with rate 0.02 patients per day for each treatment group.

(b) Repeat part (a) for the Wilcoxon test.
(c) Repeat part (a) for the log-rank and Wilcoxon tests with the follow-up time extended to 360 days. Add a line with \( S(t) = 0.05 \) at \( t = 480 \) days to both sets of projected survival times.

(d) Repeat part (a) for the log-rank and Wilcoxon tests with the original follow-up time of 240 days, but change the accrual time to 60 days. Change the last line of each set of projected survival times from 360 to 300 days with survival probabilities of 0.25 for each set.

(e) Which is the better thing to do, shortening the accrual time or lengthening the follow-up time? Explain.

3. Consider the survival probabilities for placebo treated patients that precede part (a) of problem 2. Now consider a proportional hazards alternative for the interferon treated patients, i.e., the hazard function for the interferon treated patients is \( h(t) = 0.8h_0(t) \) where \( h_0(t) \) is the hazard function for the CGD patients who received the placebo. For this alternative, treatment with interferon is assumed to provide a 20% reduction in the hazard function at any time point.

(a) Using the survival probabilities for the placebo patients, compute survival probabilities at 0, 30, 60, 90, 120, 180, 240, 360 days for patients treated with interferon.

(b) Use the probabilities computed in part (a) and the SURVPOW macro to compute the number of CGD patients needed to achieve 90% power for demonstrating a significant difference in the survival distributions of CGD patients treated with interferon and placebo. Do this for both the log-rank and Wilcoxon tests. As in problem 2, assume an accrual period of 120 days and a 240 day follow-up period. Assume an exponential distribution for loss to follow-up with a rate of 0.02 patients per day. Use an \( \alpha = 0.05 \) significance level for each test.

4. Starting with the relationship \( h(t) = f(t)/S(t) \) that we derived in class from the definition of the hazard function \( h(t) \), show that if the hazard function satisfies the proportionality criterion \( h(t) = h_0(t)g(x, \beta) \), for some function \( g(x, \beta) \), then \( S(t) = [S_0(t)]^{g(x, \beta)} \).