PSTAT 175 Project Proposal

Phillip Kim, Allison Tam, Jenny Gao

11/4/2018

1.

The event of interest of this dataset is the survival times for patients until a recurrence of bladder cancer is found. In this data set we are given the outcome variable or the survival time for each patient (inttime), the corresponding censoring status (event), and several other variables. The variables NUM and SIZE are exposure variables of interest, as they are possible confounders of the survival time variable. The 9 variables included in our data are:

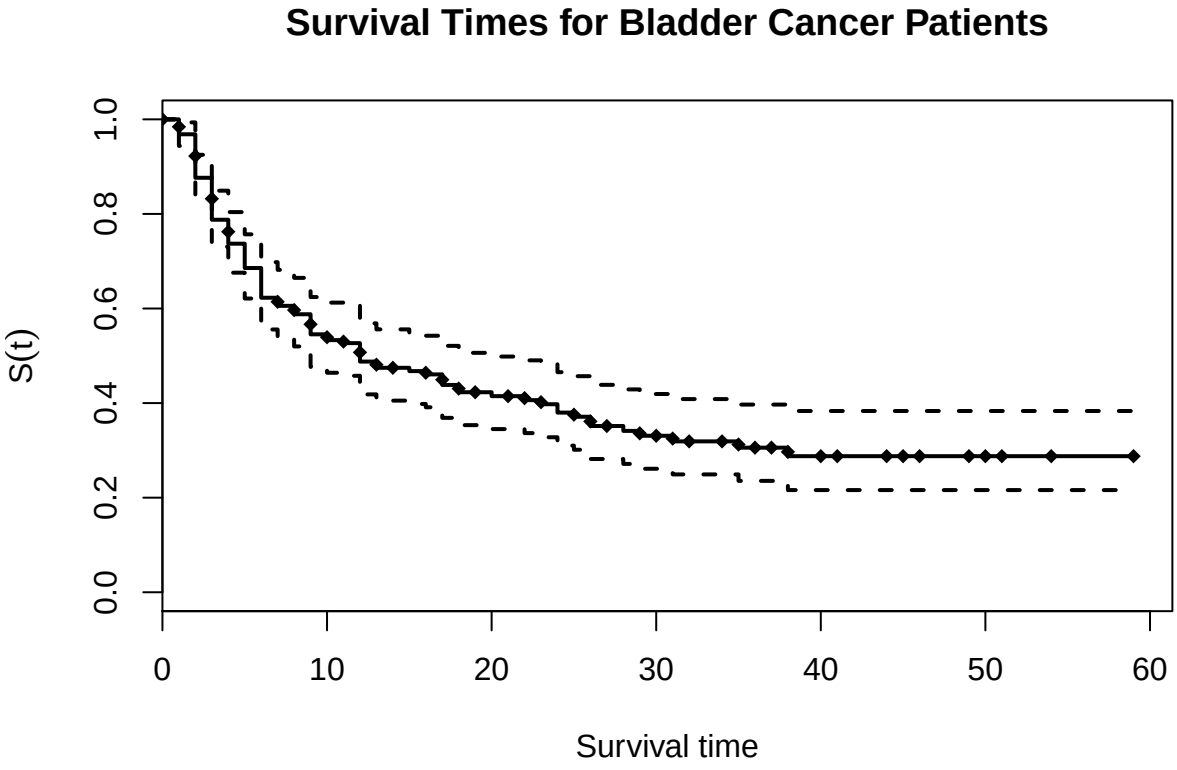
ID: subject ID Event: 0 = censored, 1 = recurrence of cancer found (event of failure is recurrence of cancer) Interval: number of data lines for each subject (each event triggers additional line) Inttime: length of observation period Start: month in which observation starts (unless the data is left-censored) Stop: month in which observation ends (unless the data is right-censored) Tx: different treatment groups (2 groups), one group is treated with thiotepa (tx=1) and one group is the control group (tx=0) Num: initial number of tumors Size: initial size of the tumor (cm)

The scientific questions we are trying to answer are: What is the survival time for patients before experiencing recurrence of bladder cancer? Do initial number of tumors affect the survival time of patients? Do initial size of tumors affect the survival time of patients? Does the thiotepa treatment affect the survival times of patients?

2.

```
library("survival")
data.bladder <- read.table("bladder.txt", skip = 12, header = TRUE)
```

```
kmfit = survfit(Surv(INTTIME, EVENT)~1, data=data.bladder)
plot(kmfit, main="Survival Times for Bladder Cancer Patients",
     xlab="Survival time", ylab=expression(hat(S)(t)),lwd=2,mark.time = TRUE,mark=18)
```



3.

We would look at survival functions for the control group and the group treated with thiotepa. We would also look at the different survival rates at different time periods for the two groups. There are extra variables in our data, and in particular, the number of tumors that a patient has can possibly affect survivability more than the length of the patient's observation time. We would also use hypothesis testing to see if the survivability functions for both groups (control and treatment) are equal to each other.

Log rank test to determine the differences in survival probabilities for the control group and the treated group Cox PH regression to model a non-time-changing linear function of the covariates (NUM and SIZE). Using the Likelihood Ratio Test, we can use model selection to determine which covariates are significant to the linear model.