

All questions relate to the question of whether treatment (DES) and index (a measure of disease severity) are prognostic of survival time for prostate cancer. The data is posted on the class web pages (prostaticancer.dta, prostaticancer.dat). We will consider four variables in this lab:

- Time: Time to death or censoring (months)
- Status: Indicator of death (status=0 if subject censored, status=1 if subject died)
- Treatment: Two treatments are considered. Treatment 1 is placebo, Treatment 2 is DES
- Index: Gleason index, a measure of disease severity

Perform analyses to determine whether the distribution of time to relapse differs across groups defined by treatment and index. Some code for Stata and R is provided below.

1. Before looking at the data, we should decide if we are going to robust standard errors or not.
 1. What are the benefits of using robust standard errors over classical PH regression?
 2. What are the benefits of using classical PH regression over PH regression with robust standard errors?
2. Provide suitable descriptive statistics regarding the distribution of time to relapse according to treatment status.
 1. Create a Kaplan-Meier estimate of the survival curves by treatment
 2. From the plot, what is the (approximate) survival time at 30 months for each treatment arm? What is it at 60 months?
3. List the variables time and status for the subjects with treatment==2
 1. Be able to explain what each row of data represents
 2. From these points, calculate (by hand) the Kaplan-Meier estimate of the survival curve for treatment 2. Compare your answer to the plot in 2.1.
4. Perform analysis comparing the instantaneous risk of relapse across groups defined treatment status using the following approaches. Compare the inference obtained from each approach.
 1. The log-rank test. This is a score test
 2. Cox proportional hazard regression using classical standard errors. This will give both a Wald and Likelihood Ratio test
 3. Cox proportional hazard regression using robust standard errors.
5. Perform a proportional hazards regression comparing the instantaneous risk of relapse across groups defined by Gleason index ("index"). Compare the inference obtained from each approach.

1. Cox proportional hazards regression using classical standard errors
2. Cox proportional hazards regression using robust standard errors

Stata notes

“stset time status” will define the outcome

“sts graph, by(treatment)” will create the Kaplan-Meier estimate of the survival curve by treatment group

“sts test treatment” will perform the log rank test

“stcox predictor, [robust]” will run the Cox PH regression model, with or without robust SEs

R notes

Robust standard errors is built into the survival package

- “library(rms)” or “library(survival)” will load this package.

“Surv(obstime, relapse)” defines the outcome, which you can use with the following functions

- “coxph(Surv(time, failure) ~ predictor, robust=TRUE)” Cox PH regression model with robust standard errors
- “coxph(Surv(time, failure) ~ predictor)” Classical Cox PH regression model
- “plot(survfit(Surv(time, failure) ~ predictor))” Plot Kaplan Meier estimate of the survival curve
- “survdiff((Surv(time, failure) ~ predictor))” Log rank test for a categorical predictor

The R survival plot makes it difficult to discern groups by default. For the treatment question, I recommend adding a line type specification (e.g. “lty=2:3”) for the plot and a legend using a second command

- plot(survfit(Surv(time, status) ~ treatment, data=cancer), lty=2:3, xlab="Time", ylab="Survival")
- legend("bottomleft", c("Treatment = 1", "Treatment = 2"), lty=2:3, inset=0.05)