

## Lab 8

Psychology 319 (GCM)

*Instructions.* Work through the lab, saving the output as you go. If you work in Microsoft Word, you can easily copy any graph to Word via the clipboard. Numerical output may also be copied easily by highlighting, moving it to the clipboard, then copying into Word. However, you should format R output in TrueType Courier New font so that it is *monospaced*. Output from this lab is to be handed in by Friday, April 23. Your output file should be named `LAST_FIRST_LAB8.DOC`, where `LAST` is your last name, and `FIRST` is your first name. Any additional files should have the same naming scheme, except the file extension should be correct. You may add any description text you wish after `LAB8` in the file name.

*Preamble.* In today's lab, we examine a survival analysis that includes a time-invariant covariate as a possible source of confounding in evaluating the impact of a treatment.

### 1 Introduction

Freireich, et al. (1963) examined the effect of a drug, 6-Mercaptopurine, on the duration of steroid-induced remissions in acute leukemia. 21 subjects received the drug, and 21 paired subjects received a placebo. Data are shown below.

## EXAMPLE

### Leukemia Remission Data

Group 1( $n = 21$ )		Group 2( $n = 21$ )	
$t(\text{weeks})$	log WBC	$t(\text{weeks})$	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

+ denotes censored observation

## 2 Lab Assignment

1. Set up a data file with an arbitrary `id`, `time`, `treatment`, `censored`, and `wbc` as the variables.
2. Break the time line into 5 “periods” of 5 weeks each, except for the last period. (Period 1 is weeks 1–5, Period 2 is weeks 6–10, but Period 5

is anyone leaving the study after 20 weeks.)

3. Add the `period` variable to the data. Hint: The file should look something like this in the first 11 lines

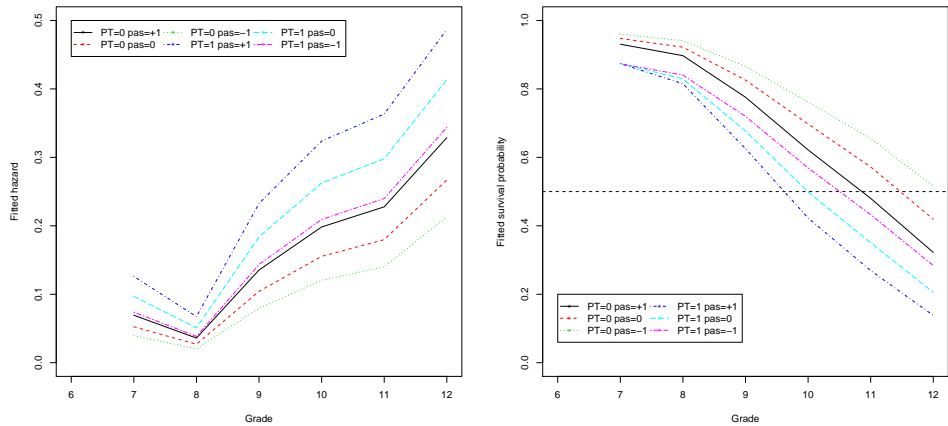
```
id,t,wbc,censored,drug,period
1,6,2.31,0,1,2
2,6,4.06,0,1,2
3,6,3.28,0,1,2
4,7,4.43,0,1,2
5,10,2.96,0,1,2
6,13,2.88,0,1,3
7,16,3.6,0,1,4
8,22,2.32,0,1,5
9,23,2.57,0,1,5
10,6,3.2,1,1,2
```

4. Take a look at the data.
5. Plot the estimated hazard and survival functions for:
  - (a) All subjects
  - (b) Just the subjects in the drug (experimental) group
  - (c) Just the subjects in the placebo (control) group

6. Write R code to construct a person-period data file from the raw data file. (If you get stumped, contact me for help.) Your person-period data should have 119 records. The first 10 records of the file should look like this

```
"id","time","event","drug","wbc"
1,1,0,1,2.31
1,2,1,1,2.31
2,1,0,1,4.06
2,2,1,1,4.06
3,1,0,1,3.28
3,2,1,1,3.28
4,1,0,1,4.43
4,2,1,1,4.43
5,1,0,1,2.96
5,2,1,1,2.96
```

7. Fit the following models to the data using the logistic link:
  - (a) A fully general model using dummy variables for each `time` period
  - (b) A fully general model with `drug` as a covariate
  - (c) A fully general model with `wbc` as a covariate
  - (d) A fully general model with both `drug` and `wbc` as covariates
8. Perform deviance tests and AIC comparisons, and justify on the basis of the results that `drug` and `wbc` are both significant predictors.
9. Pick a set of reasonable “high” and “low” values for `wbc`. (Hint: I suggest using integer values that are close to the 10th and 90th percentile.) Then, compute and plot prototypical fitted hazard and survival curves like the ones we discussed in lecture (see graphs below), using your high and low values for `wbc`, and values of 0 and 1 for `drug`. (Don’t bother to plot for a medium `wbc` value. This way you will have only 4 lines and the plot will be much easier to interpret than those from lecture shown



below.)

10. Comment on the results as portrayed in the hazard and survivor plots.
11. Now, fit a quadratic model for time with `wbc` and `drug` included as covariates. Compare the fit of this model to the completely general one.
12. Compute and plot the fitted hazard and survivor curves for the quadratic model. How do they compare with the ones yielded by the completely general specification?