DA6233

Kilger

Exercise #3 Survival Analysis Exercise

Name: Emily Bates

1. You are analyzing a 16 day study of patients who have Krusty the Clown disease. Your data includes a subject ID, their survival time and a status. A status of 0 means they are alive while a status of 1 means they died of the disease.
   1. Construct a survival table similar to the one in the lecture for the data. **You will show your work within the table** – if there is some division like 1/5 then show that like 1/5=.2 and if there is some multiplication then show the numbers being multiplied and the result e.g. .1 \* 4 =0.4
   2. Plot a survival curve for the probabilities you generate in part a.

Here is your data for question #1

|  |  |  |
| --- | --- | --- |
| subject | survival time in days | Status (0=alive, 1=dead) |
| A | 1 | 1 |
| B | 4 | 1 |
| C | 5 | 0 |
| D | 5 | 1 |
| E | 7 | 1 |
| F | 10 | 0 |
| G | 10 | 0 |
| H | 12 | 1 |
| I | 14 | 1 |
| J | 16 | 0 |

Survival Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time | Event | Censored | At Risk | Survival |
| 0 | 0 | 0 | 10 | 1.00 |
| 1 | 1 | 0 | 10 | 9/10 = .90 |
| 2 | 0 | 0 | 9 | 9/10 = .90 |
| 3 | 0 | 0 | 9 | 9/10 = .90 |
| 4 | 1 | 0 | 9 | .9 \* 8/9 = 0.8 |
| 5 | 1 | 1 | 8 | .8 \* 7/8 = 0.7 |
| 6 | 0 | 0 | 6 | 0.7 |
| 7 | 1 | 0 | 6 | .7 \* 5/6 = 0.58 |
| 8 | 0 | 0 | 5 | .58 |
| 9 | 0 | 0 | 5 | .58 |
| 10 | 0 | 2 | 5 | .58 |
| 11 | 0 | 0 | 3 | .58 |
| 12 | 1 | 0 | 3 | .58 \* 2/3 = 0.39 |
| 13 | 0 | 0 | 2 | 0.39 |
| 14 | 1 | 0 | 2 | .39 \* ½ = 0.195 |
| 15 | 0 | 0 | 1 | 0.195 |
| 16 | 0 | 1 | 1 | - |

A graph showing a graph of a survival curve

Description automatically generated

1. You are studying three different new drugs that may help slow the progress of La Traviata disease which compels people to sing opera until they exhaust themselves and die. Do the following:
   1. Draw a survival plot that shows the survival curves for all three drugs.
   2. Test to see if overall there is an effect of any of the drugs on survival taken as a global set.
   3. Compare the survival curves for each of the three drugs with each other (three comparisons) and see if any if the curves are different from each other. Note that you should be sure to adjust for multiple group comparisons.

**Be sure to include your answers to question 2, your output and code!**

A)A graph of a number of patients

Description automatically generated

B) The log rank test shows that there is at least one difference between the three survival curves, as the p-value is 3e-6.

Call: survdiff(formula = Surv(Time, Event) ~ Group, data = df)

N Observed Expected (O-E)^2/E (O-E)^2/V

Group=1 38 24 12.3 11.07 13.66

Group=2 54 25 46.0 9.58 22.50

Group=3 45 34 24.7 3.51 5.04

Chisq= 25.7 on 2 degrees of freedom, p= 3e-06

C) The p-value of 9.8e-7 indicates that we can reject the null hypothesis that there isn’t a difference between group 1 and group 2. The p-value of 0.38 between group 1 and group 3 indicates that we cannot reject the null hypothesis – there isn’t a difference between group 1 and 3. The p-value of 6.1e-5 indicates that we can reject the null hypothesis that there isn’t a difference between group 2 and 3.

Overall, the pairwise test suggests significant differences in survival between group 1 and 2 and between group 2 and 3, but not between group 1 and 3.

Pairwise comparisons using Log-Rank test

data: df and Group

1 2

2 9.8e-7 -

3 0.38 6.1e-5

P value adjustment method: BH

R Code:

install.packages('survminer')

library(survival);library(survminer);library(ggplot2)

#### Question 1b ####

# Survival data df

survival\_data <- data.frame(

Time = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16),

Survival\_Probability = c(1.00, 0.90, 0.90, 0.90, 0.80, 0.70, 0.70, 0.58, 0.58, 0.58, 0.58, 0.58, 0.39, 0.39, 0.195, 0.195, 0.195)

)

# Plotting the survival curve

ggplot(data = survival\_data, aes(x = Time, y = Survival\_Probability)) +

geom\_step() +

labs(x = "Time", y = "Survival Probability", title = "Survival Curve")

#### Question 2 ####

setwd("/Users/emilybates/Documents/Documents - Emily’s MacBook Air/MSDA - Semester 2/Data Driven Decision Making and Design")

df = read.csv('assignment3data.csv', header = TRUE)

# Fit survival curves

survfit\_obj <- survfit(Surv(Time, Event) ~ Group, data = df)

summary(survfit\_obj) #this shows survival tables for each group

# Plot survival curves

ggsurvplot(survfit\_obj, data = df, pval = FALSE, conf.int = FALSE)

# Test for overall effect of drugs on survival (log rank test)

survdiff(Surv(Time, Event) ~ Group, data = df)

# p-value is 3e-06

# Compare survival curves for each drug with each other

pairwise\_survdiff(Surv(Time, Event) ~ Group, data = df)