HW2

Thomas Kwok

March 7, 2019

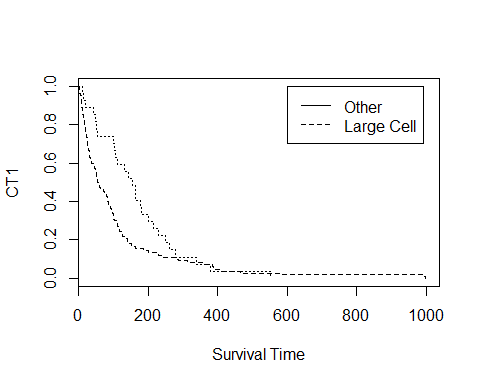
1a.

library(survival)  
#install.packages("survminer")  
library(survminer)

#install.packages("sas7bdat")  
library(sas7bdat)

## Warning: package 'sas7bdat' was built under R version 3.5.2

setwd("C:/Users/thoma/Desktop/Spring 2019/Survival Analysis")  
vets <- read.sas7bdat('vets.sas7bdat')  
q1 <- survfit(Surv(SURVT)~CT1, data=vets)  
plot(q1, lty = 2:3, xlab = "Survival Time", ylab = "CT1")  
legend(600, 1, legend=c("Other", "Large Cell"), lty = 1:2)



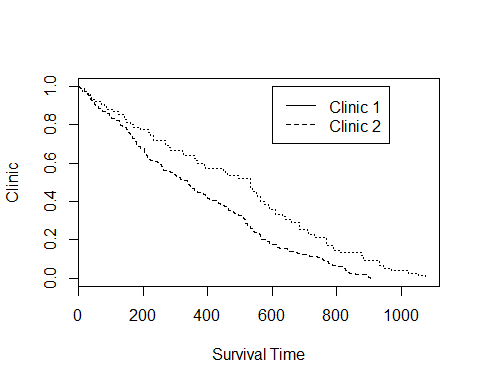
largecell <- survdiff((Surv(SURVT)~CT1==1), data=vets)  
largecell

## Call:  
## survdiff(formula = (Surv(SURVT) ~ CT1 == 1), data = vets)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## CT1 == 1=FALSE 110 110 99.5 1.12 4.26  
## CT1 == 1=TRUE 27 27 37.5 2.96 4.26  
##   
## Chisq= 4.3 on 1 degrees of freedom, p= 0.04

According to the graph, it seems that that the Large Cell survive longer than the Other group. Though one thing to keep in mind as shown on the log-rank test is that there are a lot more observed for the Other than there is for the Large Group. Also the p value for the log-rank test shows there is some significance in the data as it is lower than alpha of 0.5.

2a.

addicts <- read.sas7bdat('addicts.sas7bdat')  
q2 <- survfit(Surv(SURVT)~CLINIC, data=addicts)  
plot(q2, lty = 2:3, xlab = "Survival Time", ylab = "Clinic")  
legend(600, 1, legend=c("Clinic 1", "Clinic 2"), lty = 1:2)

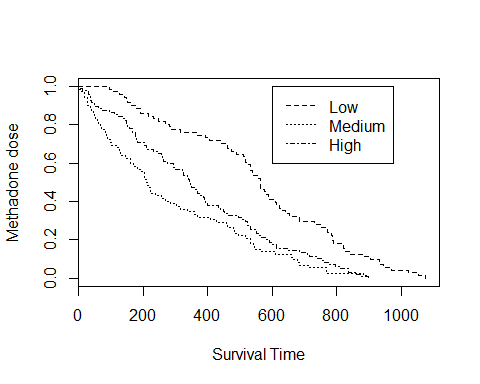


My data shows that the survival time for the patients in Clinic 2 is often better than the survival time for the patients in Clinic 1.

2b. The p-value for the log-rank and Wilcoxon test both are less than alpha 0.05 so it shows that both reject the null hypothesis and the clinics are significantly different. The chi-square for the Log-Rank is higher than the log-rank for Wilcoxon so that means that the data is spread out instead of having more emphasis on the first few data.

2c.

addicts$DOSE[which(addicts$DOSE >= 70)] = 'high'  
addicts$DOSE[which(addicts$DOSE < 55)] = 'low'  
addicts$DOSE[which(addicts$DOSE %in% c(55, 60, 65))] = 'medium'  
q3 <- survfit(Surv(SURVT)~DOSE, data=addicts)  
plot(q3, lty = 2:5, xlab = "Survival Time", ylab = "Methadone dose")  
legend(600, 1, legend=c("Low", "Medium", "High"), lty = 2:5)



The KM curves indicate taht people with higher doses will have the best survival prognosis while low seems to be a bit better than medium.