BIOST 537: Homework 2

Department of Biostatistics @ University of Washington

Alejandro Hernandez

Winter Quarter 2025

PROBLEM 1.

Consider the following trial data on time until relapse (in months) for patients in remission of acute myelogenous leukemia administered a maintenance chemotherapeutic agent or not:

Maintenance group: 9, 12, 13+, 18, 23, 28+, 31, 34, 45, 45+, 48, 161+.

Control group: 4, 5, 8, 8, 10+, 12, 16+, 23, 27, 30, 38, 43, 45.

(a) Compute the Kaplan-Meier and Nelson-Aalen estimators by filling the following tables for the control and maintenance group.

Table 1: Survival table of remission patients administered maintenance treatment.

Time	At risk (n)	Events (d)	d/n	1-d/n	K-M S(t)	N-A H(t)
9	12	1	0.08	0.92	0.92	0.08
12	11	1	0.09	0.91	0.83	0.17
18	9	1	0.11	0.89	0.74	0.29
23	8	1	0.12	0.88	0.65	0.41
31	6	1	0.17	0.83	0.54	0.58
34	5	1	0.20	0.80	0.43	0.78
45	4	1	0.25	0.75	0.32	1.03
48	2	1	0.50	0.50	0.16	1.53

Table 2: Survival table of remission patients administered a control treatment.

Time	At risk (n)	Events (d)	d/n	1-d/n	K-M S(t)	N-A H(t)
4	13	1	0.08	0.92	0.92	0.08
5	12	1	0.08	0.92	0.85	0.16
8	11	2	0.18	0.82	0.69	0.34
12	8	1	0.12	0.88	0.61	0.47
23	6	1	0.17	0.83	0.50	0.63
27	5	1	0.20	0.80	0.40	0.83
30	4	1	0.25	0.75	0.30	1.08
38	3	1	0.33	0.67	0.20	1.42

Time	At risk (n)	Events (d)	d/n	1-d/n	K-M S(t)	N-A H(t)
43	2	1	0.50	0.50	0.10	1.92
45	1	1	1.00	0.00	0.00	2.92

(b) For each group, what is the estimated probability that no relapse will occur by 36 months?

We fit two nonparametric Kaplan-Meier models of time until relapse for patients in remission of acute myelogenous leukemia, based on whether they were administered a maintenance chemotherapeutic agent or not. Our target is each group's probability of reaching 36 months without relapsing. We estimate this probability to be 0.432 (95% CI*: 0.141-0.698) among the maintenance group and 0.303 (95% CI*: 0.076-0.574) among the control group.

* Confidence intervals were computed using the "complementary log-log" technique, to ensure they fall within the [0,1] interval.

PROBLEM 2.

In this problem, you will revisit the dataset on methadone maintenance for heroin addicts, as studied in Problem Set 1. Please refer to that homework for a description of the relevant variables.

(a) Plot the Kaplan-Meier estimator of the survival function of the time until exit from maintenance along with pointwise 95% confidence intervals. What is the estimated probability that no exit will occur by one year? Provide a 95% confidence interval for your answer.

(b) Provide the estimated median time until exit from maintenance and associated 95% confidence interval by:

  i. scrutinizing values of the Kaplan-Meier estimator and associated confidence intervals (explain how you obtain your answer);

  ii. using the median estimate and confidence intervals provided by the survfit command.

The `survfit` function estimates the median time until exit from maintenance to be 504 days, approximately 1.4 years (95% CI: 394-550 days).

(c) In this part, you will investigate differences between patients with and without a history of incarceration.

  i. On the same graph, plot the Kaplan-Meier estimator of the survival function of the time until exit from maintenance for patients with a history of incarceration and for patient without.
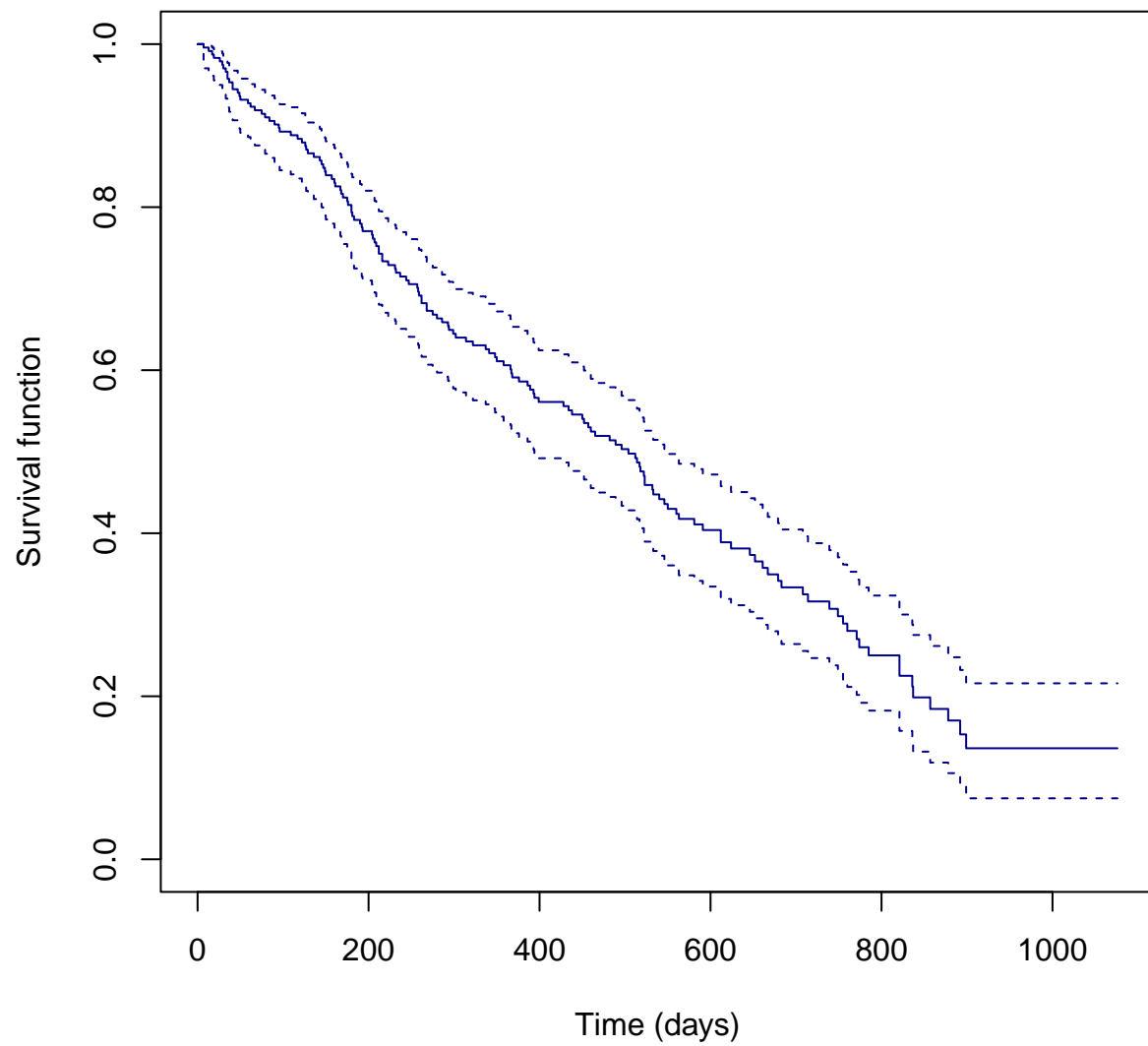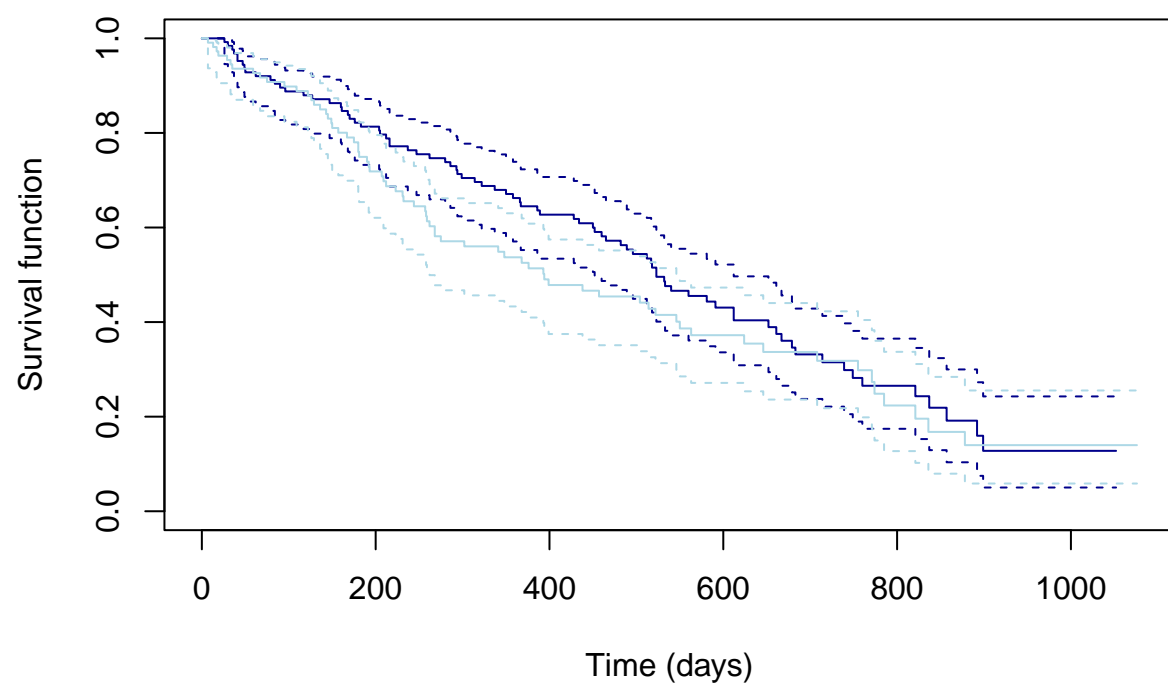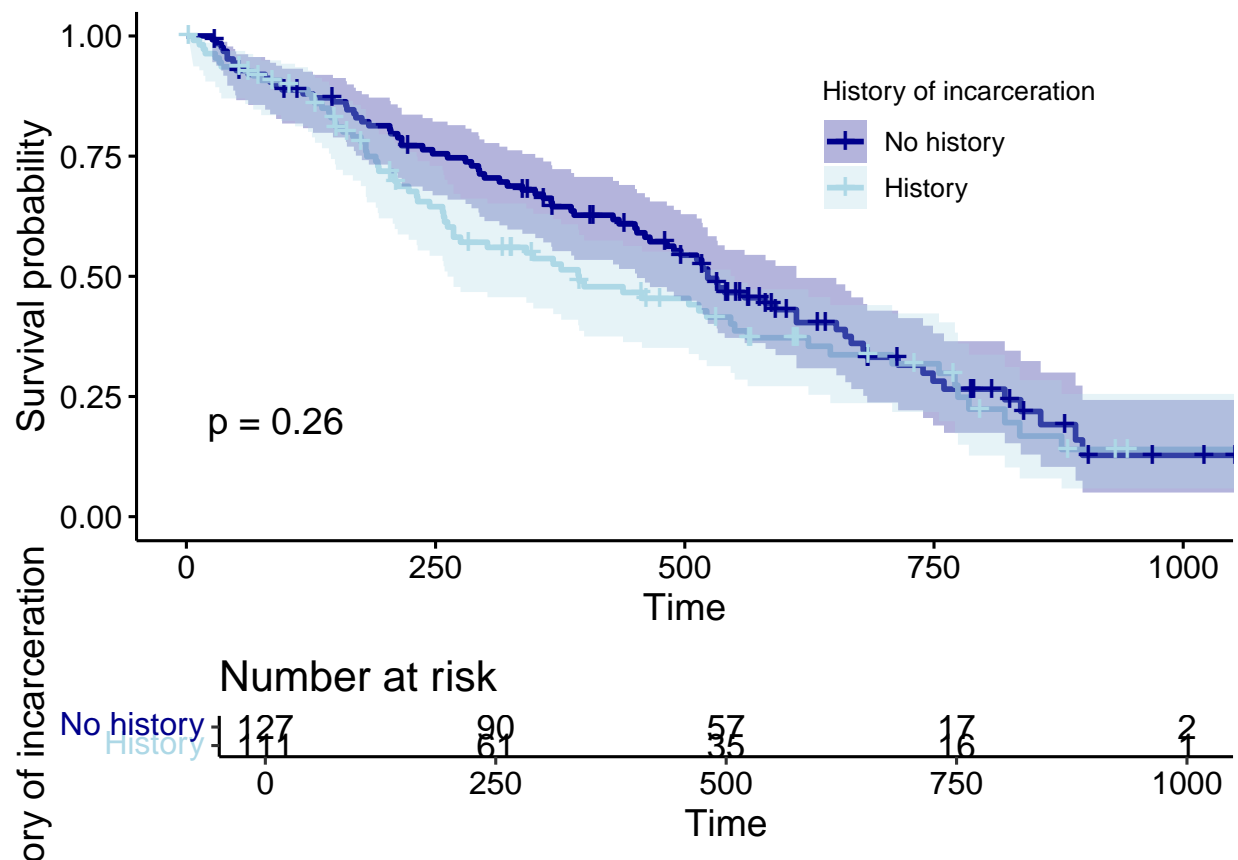
Figure 1: Time from entry to exit from methadone maintenence. Survival is modeled by a nonparametric Kaplan-Meier estimator.

Figure 1

Investigators followed a cohort of 238 individuals with heroin use disorder who entered methadone mainte-nance programs in either of two clinics in Sydney, Australia over an 18-month period during the late 1980s. The purpose of the study was to identify factors associated to retention in methadone maintenance. Time from entry into the study to exit from methadone maintenance is the duration of interest. More details can be found in Caplehorn and Bell (1991).

From the 238 individuals enrolled in this study, we fit two non-parametric models of time from entry into the study to exit from methadone maintenance: one for patients from with a history of incarceration and another for patients without a history of incarceration; see Figure 1.

    ii. Does the probability that no exit occurred by 8 months differ significantly between these two groups?

    iii. Based on the log-rank test, does the distribution of time until exit from maintenance differ significantly by history of incarceration?

A log-rank test determined there is insufficient evidence in our sample that time from entry into the study to exit from methadone maintenance differs significantly between patients with and without a history of incarceration at $\alpha = 5\%$ (p=0.3).

    iv. Based on the Wilcoxon-Gehan-Breslow test, does the distribution of time until exit from maintenance differ significantly by history of incarceration?

    v. Plot estimated hazard functions for patients with and without a history of incarceration. Briefly indicate how this plot may inform you regarding the power of the logrank test as well as expected differences in the magnitude of the chi-square statistics from the logrank and Wilcoxon-Gehan-Breslow tests.

(d) Repeat (c) but substituting history of incarceration by methadone dosage dichotomized at 60mg/day (i.e., compare the subpopulation of patients administered more than 60mg/day of methadone to the subpopulation of patients administered no more than 60 mg/day).
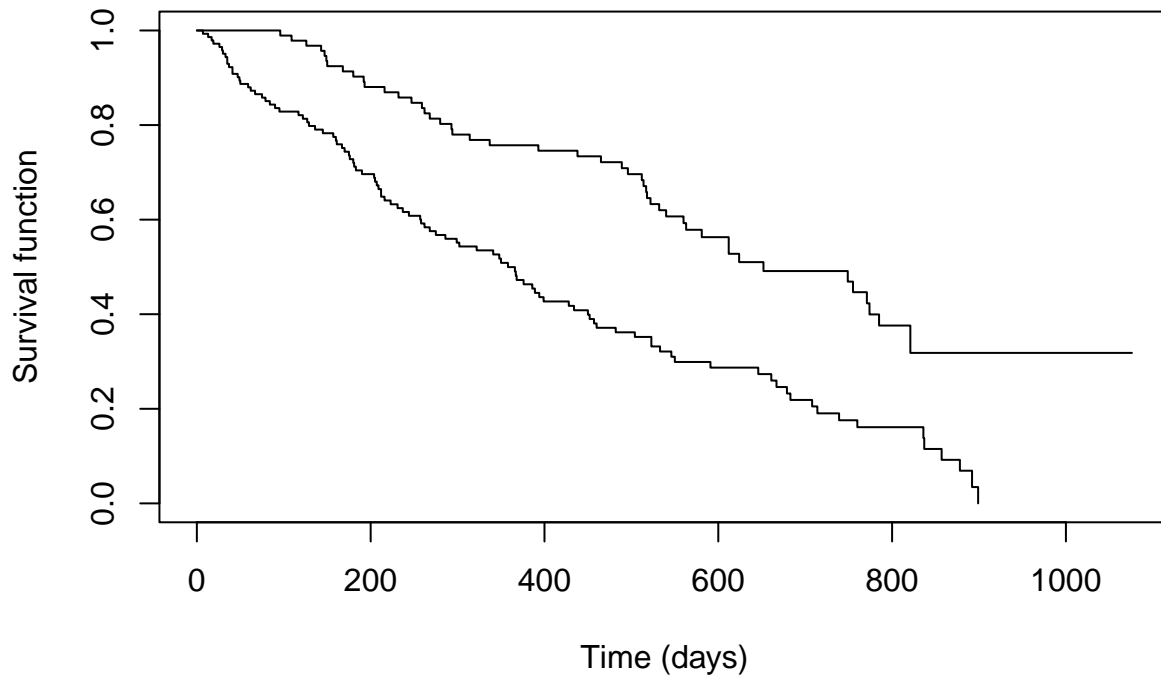
(e) Based on a stratified log-rank test, does the time until exit from maintenance differ by history of previous incarceration adjusting for clinic membership? State explicitly what the null and alternative hypotheses are and contrast with what they are in a standard log-rank test.

(f) What is the estimated median residual time until exit from maintenance at 4, 8 and 12 months? Calculate these estimates using only values of the Kaplan-Meier estimator. Verify your answer using the R function provided for this purpose, and obtain 95% confidence intervals to accompany your estimates.

End of report. Code appendix begins on the next page.

Code Appendix

```r
# Clear environment
rm(list=ls())

# Setup options
knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE, results='hide')
options(knitr.kable.NA = '-', digits = 2)
labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "allcode")]
# Load relevant packages
library(survival)  # survival model
library(dplyr)  # data manipulation
library(knitr)  # pretty tables
library(ggplot2)  # data visualization
library(survminer)

# Load data
methadone <- read.csv("../data/methadone.csv")

# Handle missing data

#### Question 1 ####

# Maintenance group survival times
mtnc_surv <- survival::Surv(
  time = c(9,12,13,18,23,28,31,34,45,45,48,161),
  event = c(1,1,0,1,1,0,1,1,1,0,1,0))
# Control group survival times
ctrl_surv <- survival::Surv(
  time = c(4,5,8,8,10,12,16,23,27,30,38,43,45),
  event = c(1,1,1,1,0,1,0,1,1,1,1,1,1))

# Survival table for treated
mtnc_fit <- survival::survfit(mtnc_surv ~ 1, conf.type = "log-log")
mtnc_surv_tab <- data.frame(summary(mtnc_fit)[c(2,3,4,6,8)])
mtnc_surv_tab$"d/n" <- with(mtnc_surv_tab, n.event / n.risk)
mtnc_surv_tab$"1-d/n" <- 1 - mtnc_surv_tab$"d/n"

mtnc_surv_tab[,c(1,2,3,6,7,4,5)] %>%
  knitr::kable(
    col.names = c("Time", "At risk (n)", "Events (d)", "d/n", "1-d/n", "K-M S(t)",
                  "N-A H(t)"),
    caption = "Survival table of remission patients administered maintenance treatment.")

# Survival table for controls
ctrl_fit <- survfit(ctrl_surv ~ 1, conf.type = "log-log")
ctrl_surv_tab <- data.frame(summary(ctrl_fit)[c(2,3,4,6,8)])
ctrl_surv_tab$"d/n" <- with(ctrl_surv_tab, n.event / n.risk)
ctrl_surv_tab$"1-d/n" <- 1 - ctrl_surv_tab$"d/n"

ctrl_surv_tab[,c(1,2,3,6,7,4,5)] %>%
  kable(
```

```r
                col.names = c("Time", "At risk (n)", "Events (d)", "d/n", "1-d/n", "K-M S(t)",
                              "N-A H(t)"),
          caption = "Survival table of remission patients administered a control treatment.")

# Maintenance group
summary(mtnc_fit, times = 36)
# Control group
summary(ctrl_fit, times = 36)


#### QUESTION 2 ####
# Plot K-M estimator of survival
meth_surv <- with(methadone, Surv(time, event))
meth_fit <- survfit(meth_surv ~ 1, data = methadone, conf.type = "log-log")

# With base R
plot(meth_fit, conf.int = TRUE, col = "darkblue",
     ylab = "Survival function", xlab = "Time (days)")

# Estimate survival to one year
summary(meth_fit, times = 365)
# # With ggplot
# survminer::ggsurvplot(meth_fit, data=methadone, ggtheme=theme_bw(),
#                       conf.int=TRUE, xlab="", ylab="")
# ggsurvplot(meth_fit, data=methadone, ggtheme=theme_bw(),
#            risk.table = TRUE, conf.int=TRUE, xlab="", ylab="")

# Facet
# survminer::ggsurvplot_facet(meth_fit,
#                             facet.by=c("clinic"), # in data, not model
#             # facet.by is only for fit models that already have some covariate
#                             data=methadone)

# Other arguments
# censor = TRUE
# legend.labs = c("Survival Estimate")
# legend.title = ""
# legend = c(0.8, 0.8)
# palette = "purple"

# Median estimate by survfit
meth_fit # complementary log-log CI

median(meth_surv) # untransformed CI
quantile(meth_fit, prob = 0.5) # using clog-log
## Plot KM estimator for patients with and without a history of incarceration
meth_fit2 <- survfit(meth_surv ~ prison, conf.type = "log-log", data = methadone)
meth_fit2 %>%
  plot(xlab = "Time (days)", ylab = "Survival function", conf.int = TRUE,
       col = c("darkblue", "lightblue"))

gg <- ggsurvplot(meth_fit2,
                 data = methadone,
           conf.int = TRUE,
```

```
        pval = TRUE,
        risk.table = TRUE,
        legend.title = "History of incarceration",
        legend.labs = c("No history", "History"),
        legend = c(0.75,0.75),
        theme = theme_bw(),
        palette = c("darkblue", "lightblue"))

gg

# gg + facet_wrap(~ clinic)

# Compare survival probability at 8 months
summary(meth_fit2, times = 8*30)

# Log-rank test of significance of prison coefficient
survival::survdiff(meth_surv ~ prison, data = methadone)




# Plot KM estimator for patients with dosage above versus below 60 mg/day
methadone$dose.bin <- ifelse(methadone$dose > 60, 1, 0)

fit3 <- survfit(meth_surv ~ dose.bin, conf.type = "log-log", data = methadone)
fit3 %>% plot(xlab = "Time (days)", ylab = "Survival function")

# Compare survival probability at 8 months
summary(fit3, times = 8*30)

# Log-rank test of significance of dosage coefficient
survdiff(meth_surv ~ dose.bin, data = methadone)

# Stratified log-rank test of significance of prison coefficient adjusting for clinic
survdiff(meth_surv ~ dose.bin + strata(clinic), data = methadone)
```

End of document.