Homework2

Yuki Joyama

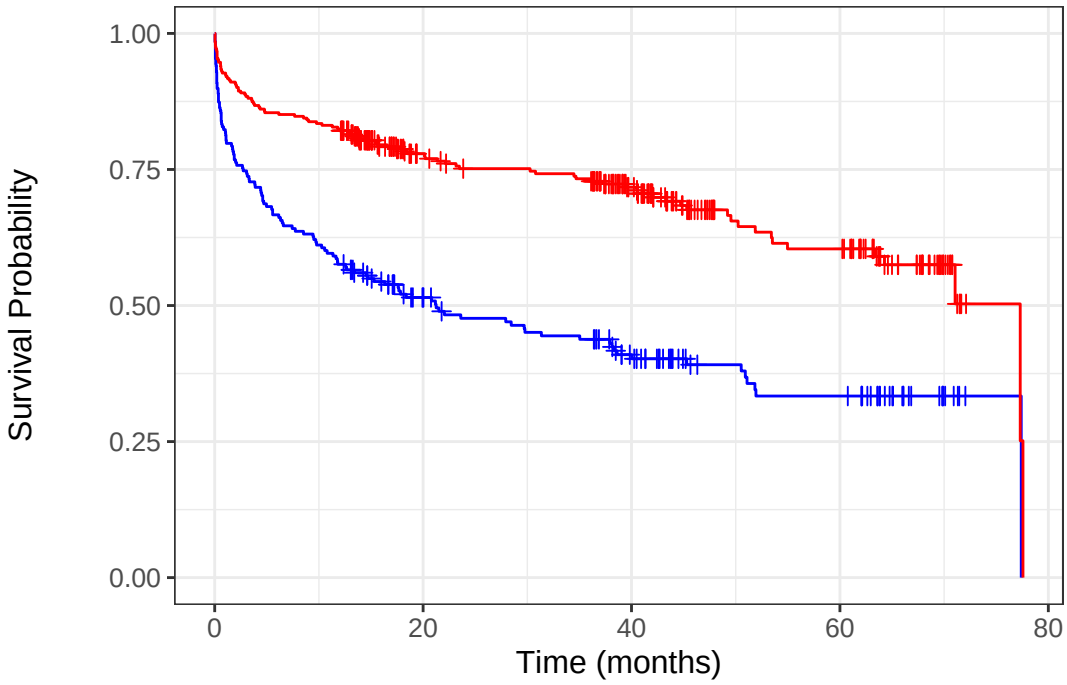
1. Logrank and Score Tests for MI Study

(a) Below is the plot of estimated Kaplan-Meier survival functions for the endpoint of death for those who are obese or overweight ($BMI \geq 25$) vs. those of normal weight (which we will define here as $BMI < 25$):

```
# prepare dataset for plot
df1 = df |>
  mutate(
    obese_ovwt2 = ifelse(bmi >= 25, 1, 0) # redefine obese_ovwt (1: BMI >= 25, 0: BMI < 25)
  )

# fit KM curve
surv <- Surv(df1$dthtime, df1$dthstat)
km <- survfit(surv ~ df1$obese_ovwt2) # compare two groups (obese or overweight vs. normal weight)

# plot the KM plot
km |>
  ggsurvfit() +
  labs(
    x = "Time (months)",
    y = "Survival Probability",
    color = "BMI Category"
  ) +
  scale_color_manual(values = c("blue", "red"), labels = c("BMI < 25", "BMI >= 25")) +
  add_censor_mark(shape = 3, size = 2) +
  add_risktable(risktable_stats = c("n.risk", "cum.censor", "cum.event"))
```



Difference in the censoring patterns between the two BMI groups:

- There is a higher frequency of censoring events throughout the study period in obese or overweight group.
- The censoring events appear to occur at similar time points for both groups, notably around 18 months, 40 months, and 67 months

```
# the number of patients who are overweight or obese (BMI >= 25)
overweight_obese_count <- df |>
  filter(bmi >= 25) |>
  nrow()

# the percentage of overweight or obese patients out of 500
pct_overweight_obese <- (overweight_obese_count / 500) * 100
```

60.4% of the patients out of 500 are either overweight or obese.

(b)
(c)
(d)

2. Cox Model for Myocardial Infarction Study

(a)
(b)
(c)
(d)

3. Model Interpretation - Myocardial Infarction Study

Variable Name	Estimate	s.e.	P-value
Age	0.0500	0.0066	< 0.0001
Heart rate	0.0112	0.0029	0.0001
Diastolic BP	-0.0107	0.0035	0.0024
Sex (0=male, 1=female)	-0.2732	0.1442	0.0581
Congestive heart failure	0.7816	0.1469	< 0.0001
BMI	-0.0453	0.0163	0.0055

Figure 1: Coefficient Estimate Table of Multivariable Model

(a)
(b)
(c)
(d)
(e)
(f)