Homework2

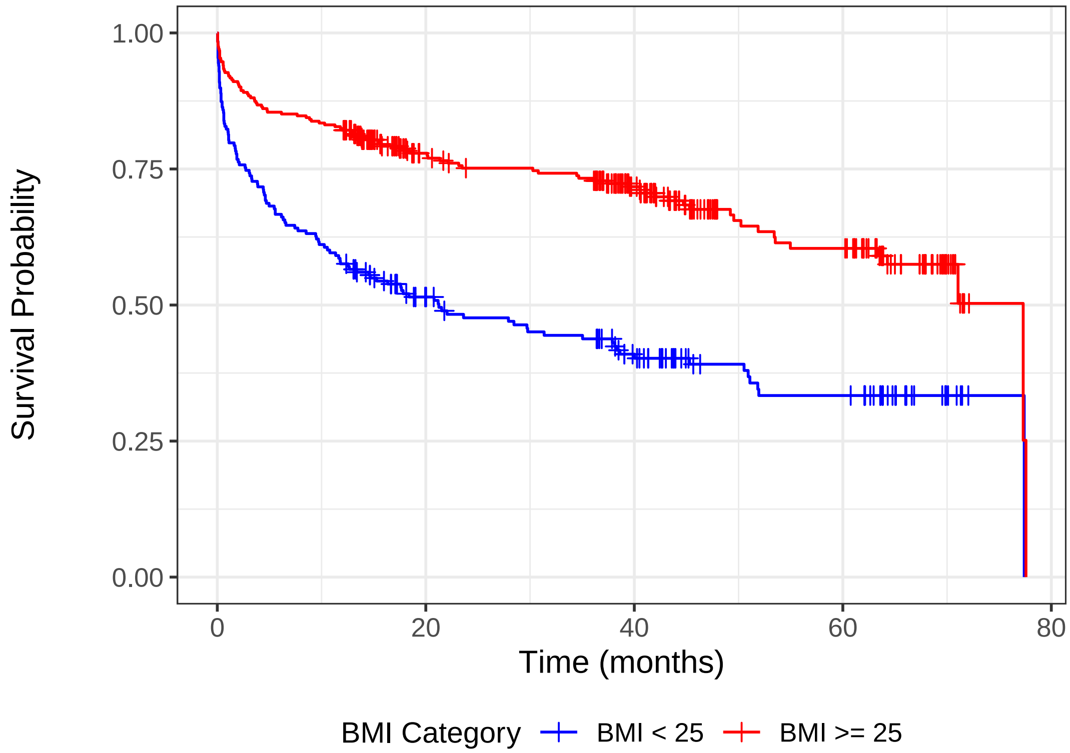
Yuki Joyama

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 $BMI < 25$):

```r
# fit KM curve
surv <- Surv(df$dthtime, df$dthstat)
km <- survfit(surv ~ df$obese_ovwt, df) # 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"))
```



df$obese_ovwt=0

At Risk	198	83	54	29	0
Censored	0	20	33	51	79
Events	0	95	111	118	119

df$obese_ovwt=1

At Risk	302	172	123	59	0
Censored	0	65	101	152	206
Events	0	65	78	91	96

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) I will implement log-rank test and Wilcoxon test using CMH approach. Below is the output from SAS.

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for dthtime over Strata

Rank Statistics		
obese_ovwt	Log-Rank	Wilcoxon
0	44.682	17295
1	-44.682	-17295

Covariance Matrix for the Log-Rank Statistics		
obese_ovwt	0	1
0	48.1884	-48.1884
1	-48.1884	48.1884

Covariance Matrix for the Wilcoxon Statistics		
obese_ovwt	0	1
0	6858079	-6858079
1	-6858079	6858079

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	41.4301	1	<.0001
Wilcoxon	43.6153	1	<.0001
-2Log(LR)	47.4160	1	<.0001

$\chi^2_{MH} = 41.4301$ (p-value <0.0001)
$\chi^2_{W} = 43.6153$ (p-value <0.0001)

The both tests suggest that there is an association between whether being overweight/obese and death. The log-rank test gives equal weight to all time points, while the Wilcoxon test gives more weight to early events. Given the KM curve, there is more differences in survival in early stage, so we can assume that Wilcoxon test would yield a larger test statistic.

(c)

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for dthtime over Strata

Rank Statistics	
obese_ovwt	Fleming
0	44.682
1	-44.682

Covariance Matrix for the Fleming Statistics		
obese_ovwt	0	1
0	48.1884	-48.1884
1	-48.1884	48.1884

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Fleming(0,0)	41.4301	1	<.0001

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for dthtime over Strata

Rank Statistics	
obese_ovwt	Fleming
0	36.014
1	-36.014

Covariance Matrix for the Fleming Statistics		
obese_ovwt	0	1
0	30.4221	-30.4221
1	-30.4221	30.4221

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Fleming(1,0)	42.6330	1	<.0001

The LIFETEST Procedure

Testing Homogeneity of Survival Curves for dthtime over Strata

Rank Statistics	
obese_ovwt	Fleming
0	6.399
1	-6.399

Covariance Matrix for the Fleming Statistics		
obese_ovwt	0	1
0	1.36298	-1.36298
1	-1.36298	1.36298

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Fleming(1,1)	30.0436	1	<.0001

p=0, q=0: $\chi^2_{FH} = 41.4301$ (p-value <0.0001)
p=1, q=0: $\chi^2_{FH} = 42.6330$ (p-value <0.0001)
p=1, q=1: $\chi^2_{FH} = 30.0436$ (p-value <0.0001)

When p=0, q=0, the test statistic is equal to that of log-rank test. When p=1, q=0, the test is similar to the Peto-Prentice test and it is closer to the test statistic of Gehan's Wilcoxon test in (b). When p=1, q=1, the Fleming-Harrington test gives more weight to events happening around the median time. The smaller test statistic indicates fewer mid-term differences between two BMI groups.
If most events happen earlier, we expect p=1, q=0 to be more powerful than log-rank test, and if most events happen around the middle of the study period, we expect p=1, q=1 to be more powerful. Fleming-Harrington may be less powerful compared to log-rank test when most of the events occur later in the follow-up period.

(d) I will use a logistic regression model to test whether the proportions of deaths during follow-up differ for those who are over-weight/obese versus those of normal weight.

```
##
## Call:
## glm(formula = dthstat ~ obese_ovwt, family = binomial, data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.4097     0.1451   2.823  0.00476 **
## obese_ovwt   -1.1732     0.1906  -6.155 7.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 683.31  on 499  degrees of freedom
## Residual deviance: 644.01  on 498  degrees of freedom
## AIC: 648.01
##
## Number of Fisher Scoring iterations: 4
```

The result indicates that being over-weight/obese has approximately 69.06% lower odds of death compared to non-obese individuals when adjusting for no other variables, and this is statistically significant.
This conclusion (being over-weight/obese have somewhat positive effect on mortality) is similar to the results of survival analysis in (b) and (c). Survival analysis may become more powerful when there are considerable number of censoring in the study subject and the data is highly skewed.

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

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

4. Impact of Ties on Cox Model Estimation and Testing