Assignment 2
(Due: 2023/05/10, 11:59pm)

Note:

- **No late assignment accepted**;
- Write your assignment in Chinese or English. PDF/word file is preferred. Include R code in your assignment.
- Set the significance level $\alpha = 0.01$.

Analysis of the Framingham Heart Study

The Framingham Heart Study, which aimed to unravel the underlying causes of heart disease, started in 1948 with 5,209 participants in the town of Framingham, Massachusetts. Framingham is a longitudinal cohort study, a type of epidemiological study that follows a group of individuals over time to determine the natural history of certain diseases, explore the behavior of those diseases, and identify the factors that might explain their development. Part of the reason Framingham, Massachusetts was picked as the study site was because it was just big enough to provide a sufficient number of individuals for the study, while also small enough to be suited to the community approach of recruiting and effectively following participants over time. Participants underwent physical examinations, gave blood samples for laboratory tests, and provided lifestyle and medical history information at regular intervals. Now a joint project of the NHLBI and Boston University, Framingham has expanded over the years, both in geographical and population scope. Today it includes many grandchildren and spouses in three generations of participants, as well as two cohorts of minority participants (the Framingham Omni Cohorts).

Thanks to Framingham and other studies, we now know that most cardiovascular disease is caused by modifiable risk factors like smoking, high blood pressure, obesity, high cholesterol levels, and physical inactivity. Framingham has equipped researchers and clinicians with a better understanding of how risk factors contribute to the development of chronic diseases. The Framingham Heart Study continues to lead cutting-edge research areas such as genomic, proteomic, and metabolomic biomarkers of cardiovascular disease risk, vascular stiffness, gut microbiome, and cardiopulmonary exercise testing, among others. Even as the Framingham Heart Study turns 70 years old, lessons learned, both old and new, continue to inform not only the way we study health and disease, but also our understanding of who is vulnerable to chronic diseases and why. More

details can be found via

`https://www.framinghamheartstudy.org`

The data **framingham** is a small subset extracted from the Framingham Heart Study, which were collected over a 24 year study for competing risks survival analysis of hypertension and death as a function of smoking. However, this assignment does not involve competing risks.

```
> head(framingham)
  female totchol age   bmi  BPVar heartrte glucose cursmoke outcome time_outcome cigpday
1      0     195  39 26.97 -17.00       80      77        0       0    24.000000       0
2      1     250  46 28.73   1.50       95      76        0       0    24.000000       0
3      0     245  48 25.34   3.75       75      70        1       0    24.000000      20
5      1     285  46 23.10   9.00       85      85        1       1    11.731691      23
7      1     205  63 33.11   0.00       60      85        0       1     6.056126       0
8      1     313  45 21.68 -19.00       79      78        1       1    23.761807      20
```

The data set is explained as follows:

```
Format
A data frame with 2316 rows and 11 columns:

female
Sex of the patient. 1=female, 0=male.

totchol
Total cholesterol of patient at study entry.

age
Age of the patient at study entry.

bmi
Patient body mass index.

BPVar
Average units of systolic and diastolic blood pressure above normal:
((SystolicBP-120)/2) + (DiasystolicBP-80)

heartrte
Patient heartrate taken at study entry.

glucose
Patient blood glucose level.
```

`cursmoke`
`Whether or not the patient was a smoker at the time of study entry.`

`outcome`
`Did the patient die (=1), experience hypertension (=2), or leave the study without experiencing either event (=0).`

`time_outcome`
`The time at which the patient experienced outcome.`

`cigpday`
`Number of cigarettes smoked per day at time of study entry.`

We treat hypertension as the event of interest, and treat death and leaving the study as censoring. We are particularly interested in assessing the effect of smoking on the occurrence of hypertension. Fit the following survival models:

(Model 1) Using time-on-study as time-scale, fit a Cox model with `cursmoke`, `female`, `totchol`, `age`, `age`^{2}, `bmi`, `BPVar`, `heartrte`, `glucose` as risk factors.

(Model 2) Using time-on-study as time-scale, fit a Cox model with `cursmoke`, `female`, `totchol`, `age`, `age`^{2}, `bmi`, `BPVar`, `heartrte`, `glucose`, and the interaction term `cursmoke` $\times$ `female` as risk factors.

(Model 3) Using age as time-scale, fit an appropriate Cox model that accounts for left-truncation with `cursmoke`, `female`, `totchol`, `bmi`, `BPVar`, `heartrte`, `glucose` as risk factors.

Answer the following questions:

1. According to Model 1 and Model 3, what is the adjusted hazard ratio for smoking v.s. non-smoking and its 99% confidence interval? Comment on the results.

2. According to Model 2, what is the adjusted hazard ratio for smoking female v.s. smoking male and its 99% confidence interval? What is the adjusted hazard ratio for nonsmoking female v.s. nonsmoking male and its 99% confidence interval? Comment on the results.

3. Draw the adjusted survival curves using Model 1 for smoking subjects v.s. non-smoking subjects, where all other risk factors are set to the overall means respectively. Comment on the plot.