Assignment 3
(Due: 2020/07/15, 11:59pm)

Note:

- **No late assignment accepted**;

- Submit your assignment in a single PDF file ($name + ID.pdf$) with all R code and outputs; submit it to
  `statistics_sysu@163.com`
  by the deadline;

- Write your assignment in Chinese or English.

Analysis of WCGS Data

Data Source: wcgsdata.csv
Documentation: wcgs.doc
Reference: http://www.epi.umn.edu/cvdepi/study-synopsis/western-collaborative-group-study/

Questions:

1. Run a Cox model with `Age0`, `BMI`, `Sbp0`, `Dbp0`, `Chol0`, `Smoking`, `Dibpat0` as the risk factors, and coronary heart disease (CHD) as the disease under investigation. Note that `BMI` and `Smoking` (indicating smoker or non-smoker) are obtained from Assignment 2. What do you conclude about the effects of these covariates on the occurrence of CHD?

   In particular, what are the hazard ratio and the associated 95% confidence interval of Behaviour Type A compared to Behaviour Type B? Compare your findings to Assignment 2.

2. Perform a goodness of fit test of the PH assumptions for all these covariates, and make your conclusion. We set the significance level $\alpha = 0.05$.

3. Draw log-log plots to check the PH assumptions for all these covariates. Comment on these plots.

4. Perform a stratified Cox analysis, where we stratify the covariates that do not satisfy the PH assumptions (Note: if all covariates satisfy the PH assumptions, then stratify the covariate which has the smallest p-value in the goodness of fit test).

5. We suspect that the effect of behaviour type on the hazard of CHD may be time-dependent. Therefore, we perform an extended Cox analysis, where $Dibpat0 \times t$ is added as an explanatory variable. That is, we fit the extended Cox model:

$$\lambda(t|X(t)) = \lambda_0(t) \exp(\beta^T X(t)),$$

where $X(t) = (Age0, BMI, Sbp0, Dbp0, Chol0, Smoking, Dibpat0, Dibpat0 \times t)^T$. Make your conclusion.

6. Run a log-normal survival model, with `Age0`, `BMI`, `Sbp0`, `Dbp0`, `Chol0`, `Smoking`, `Dibpat0`, as the covariate vector. What do you conclude about the effects of these covariates on the occurrence of CHD? Compare your findings to those in Question 1.

7. Suppose a friend with no CHD history, Friend A, seeks your advice on CHD. He is 55 years old with the following information: he is overweight with BMI equal to 28.5, systolic blood pressure 138 mmHg, diastolic blood pressure 90 mmHg, fasting serum cholesterol 280 mm%, smoker, and his behavior type is A.

Suppose another friend with no CHD history, Friend B, also seeks your advice on CHD. He is 42 years old with the following information: BMI 22, systolic blood pressure 120 mmHg, diastolic blood pressure 80 mmHg, fasting serum cholesterol 180 mm%, non-smoker, and his behavior type is A.

You are always ready to help others. To make your advice to your friends, you fit the Cox model in Question 1 and log-normal survival model in Question 6, and predict the individual survival probabilities based on the fitted models.

According to the Cox model, what are the probabilities that Friend A and Friend B will get CHD in the next 5 years, respectively? What are the probabilities that Friend A and Friend B will get CHD in the next 8 years, respectively?

According to the log-normal survival model, what are the probabilities that Friend A and Friend B will get CHD in the next 5 years, respectively? What are the probabilities that Friend A and Friend B will get CHD in the next 8 years, respectively?