Assignment 4
(Due: 2021/06/30, 11:59pm)

Note:

- **No late assignment is accepted**;
- Write your assignment in Chinese or English.

Questions:

1. Analysis of rhDNase data.

   Fuchs et al. (1994) reported on a double-blind randomized multicenter clinical trial designed to assess the effect of rhDNase, a recombinant deoxyribonuclease I enzyme, versus placebo on the occurrence of respiratory exacerbations among patients with cystic fibrosis. The rhDNase operates by digesting the extracellular DNA released by leukocytes that accumulate in the lung as a result of bacterial infection, and so it was expected that aerosol administration of rhDNase would reduce the incidence of exacerbations.

   Data on the occurrence and resolution of all exacerbations were recorded over approximately 169 days of followup for 645 patients in this trial. Part of the data is given in the following Table for the first 20 patients. We include a patient identifier, the treatment assignment (T) (1 = rhDNase, 0 = placebo), two baseline measurements of forced expiratory volume (FEV1 and FEV2) reflecting lung capacity, and the date of randomization. In addition, the number of days from randomization to the beginning (B) of the exacerbations is recorded, as well as the day on which treatment for each exacerbation ended (E) and patients became at risk of a new exacerbation. Therefore, for patient number 589302, the first exacerbation began 8 days after randomization and antibiotic therapy for this exacerbation ended 22 days after randomization. The patient then remained at risk until the second exacerbation, which began 63 days after randomization, and became at risk again after therapy ended on day 88; the patient

Table D.2. Data from rhDNase study for first 20 subjects.

					Onset and Resolution Times						
					1st		2nd		3rd		Cens.
ID	T	FEV_1	FEV_2	Rand. Date	B	E	B	E	B	E	Time
493301	1	28.8	28.1	20/03/1992							168
493303	1	64.0	63.0	24/03/1992							169
493305	0	67.2	68.7	24/03/1992	65	75					168
493309	1	57.6	56.5	26/03/1992							168
493310	0	57.6	56.3	24/03/1992							171
493311	1	25.6	25.3	27/03/1992							166
493312	0	86.4	85.4	27/03/1992							168
493313	0	32.0	32.4	28/03/1992	90	104					166
589301	1	86.4	86.0	27/02/1992							169
589302	0	28.8	29.2	06/03/1992	8	22	63	88			169
589303	0	112.0	110.7	28/02/1992	60	74	83	124			169
589305	0	70.4	71.7	04/03/1992	50	68					169
589307	1	96.0	94.5	05/03/1992							169
589309	0	44.8	44.6	05/03/1992	99	114					169
589310	1	70.4	70.1	06/03/1992	35	64	71	108			169
589311	1	54.4	53.8	11/03/1992							169
589312	0	73.6	73.2	12/03/1992	8	13					196
589313	1	96.0	97.2	12/03/1992							169
589314	0	105.6	107.0	12/03/1992							169
589316	1	80.0	79.4	19/03/1992							167

did not have another exacerbation over the remainder of followup which ended on day 169.

The data frame is as follows.

```
> rhDNase <- read.table("rhDNase.txt", header=F)
> setNames(rhDNase,c("id","trt","fev1","fev2","start","stop","status","etype",
+ "enum","enum1","enum2"))[1:18,]
       id trt fev1 fev2 start stop status etype enum enum1 enum2
1  493301   1 28.8 28.1     0  168      0     1    1     1     0
2  493303   1 64.0 63.0     0  169      0     1    1     1     0
3  493305   0 67.2 68.7     0   65      1     1    1     1     0
4  493305   0 67.2 68.7    65   75      1     2    2     1     1
5  493305   0 67.2 68.7    75  168      0     1    3     2     1
6  493309   1 57.6 56.5     0  168      0     1    1     1     0
7  493310   0 57.6 56.3     0  171      0     1    1     1     0
8  493311   1 25.6 25.3     0  166      0     1    1     1     0
9  493312   0 86.4 85.4     0  168      0     1    1     1     0
10 493313   0 32.0 32.4     0   90      1     1    1     1     0
11 493313   0 32.0 32.4    90  104      1     2    2     1     1
12 493313   0 32.0 32.4   104  166      0     1    3     2     1
13 589301   1 86.4 86.0     0  169      0     1    1     1     0
14 589302   0 28.8 29.2     0    8      1     1    1     1     0
15 589302   0 28.8 29.2     8   22      1     2    2     1     1
16 589302   0 28.8 29.2    22   63      1     1    3     2     1
17 589302   0 28.8 29.2    63   88      1     2    4     2     2
18 589302   0 28.8 29.2    88  169      0     1    5     3     2
```

Here `id` is the patient ID number, `trt` equals 1 for patients receiving rhDNase and 0 if they receive placebo, and `fev1` and `fev2` are the forced respiratory volume measured twice at randomization. The variable `start` is the start of a period indicating when subjects become "at risk" for a new exacerbation and `stop` is the time of the new exacerbation or a censoring time. The `status` variable equals 1 if `stop` is a recurrence time of new exacerbation and equals 0 if it is a censoring time (i.e. the end of followup). The `etype` variable indicates the nature of the event time recorded in `stop`; specifically, if etype=1 then `stop` corresponds to the onset of an exacerbation (or censoring) and if etype=2, `stop` corresponds to the time of a resolution of an exacerbation (or censoring).

The `enum` variable simply records the cumulative number of lines in the data frame for each individual, `enum1` the cumulative number of exacerbation-free periods. Ignore `enum2`.

(a) Use the renewal model (gap time model in KK2012):

$$\lambda_k(w|X) = \lambda_{0k}(w)\exp(X^T\beta)$$

with a subset of the data satisfying `etype`=1 (because the data line with `etype`=2 does not contribute to the gap time analysis), to assess the effect of rhDNase on reducing the risk of a new exacerbation. Here, X should include two covariates: `trt` and the average of `fev1` and `fev2` (you may name it as `fev`). Interpret your result.

(b) The Poisson model (CP model/stratified CP model in KK2012) is much more difficult to fit to this data set because of the `etype`=2 data. Try to modify the data set and fit the stratified CP model to assess the effect of rhDNase on reducing the risk of the recurrence of exacerbation.

2. Analysis of hospital data: Impact of pneumonia status on admission on intensive care unit mortality

The data set `sir.adm` comes with the `mvna` package. Briefly, 747 intensive care unit patients are included in `sir.adm`. Competing endpoints are discharge from the unit and death on the unit. `pneu` informs on a patients pneumonia status on admission, 1 for pneumonia present on admission, and 0 for no pneumonia. A patients status at the end of the observation period is contained in `status`, 1 for discharge (alive), 2 for death, 0 for patients censored before end of unit stay. A patients length of stay is in `time`.

The aim of the present analysis is to study the impact of pneumonia present on admission on unit mortality. As pneumonia is a severe illness, we should expect more patients dying with pneumonia than without. Death is the event of interest and discharge is the competing event.

(a) Fit a cause-specific hazard model with `pneu`, `age`, `sex` as the covariates, for the two events, respectively. Furthermore, apply the `crr` function of the R package `cmprsk`, and fit a Fine and Gray model (subdistribution hazard model) for the two events, respectively. Interpret your result.

(b) Draw a plot of two cumulative incidence curves (also in the R package `cmprsk`) for the two events (no covariates). Interpret your result.

```
> data(sir.adm)
> head(sir.adm)

    id pneu status time      age sex
1   41    0      1    4 75.34153   F
2  395    0      1   24 19.17380   M
3  710    1      1   37 61.56568   M
4 3138    0      1    8 57.88038   F
5 3154    0      1    3 39.00639   M
6 3178    0      1   24 70.27762   M
```