HW (Due, 5/31/2017)

Assume that observed data consist of $(U_i, \delta_i, Z_i), i = 1, \cdots, n$, where $U_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$ and Z_i is a covarite vector. Under the PH model, we assume that

$$h(t|Z = z) = h_0(t)e^{\beta_0' z}$$

for the hazard function of T_i given $Z_i = z$, where $h_0(t)$ is a baseline hazard function free of the covariate.

1. Show that Cox model implies that $P(T \geq t|Z = z) = S_0(t)^{\exp(\beta_0' z)}$ for a baseline survival function $S_0(t)$

2. How to estimate $S_0(t)$ based on the Breslow estimator for the baseline cumulative hazard function $H_0(t) = \int_0^t h_0(u)du$?

3. In the "survival library" of R, there is a lung cancer data set. You can access necessary information of the dataset and fit a PH model by typing

```
library(survival)
?lung
coxph(Surv(time, status-1)~age+sex+ph.karno+ph.ecog, data=lung)
```

You may also use the following outputs from "coxph.detail" to construct the Breslow's estimator of the cumulative baseline hazard function.

```
fit=coxph(Surv(time, status-1)~age+sex+ph.karno+ph.ecog, data=lung)
fit.detail=coxph.detail(fit)
fit.detail$time
fit.detail$hazard
```

Let Z be age, sex, ph.karno and ph.ecog. Please estimate the 1-year survival probability for a 60 year old male patient with a ECOG performance score of 0 and a Karnofsky performance score of 80 (by physician) based on the Cox regression model. *(Please note that the baseline hazard function in R is defined as the hazard function of the patient with a covariate vector of $\bar{Z}$, the empirical average of observed Z_is in the entire cohort. R effectively normalizes all the covariates by substracting their means from the individual covariate before model-fitting. For example, instead of using the original "age", R uses $age - mean(age) = age - 62.45$ in fitting the Cox model with this lung dataset.)*