HW3 (Due, 4/28/2017)

1. In R, "survfit" under the R package "survival" can be used to construct the KM estimator with right censored survival data. Try the "survfit" function for estimating the survival function using the "pbc" data in which you only need to use "time" ($\min(T, C)$) and "status" ($\delta = I(T < C) = I(\text{status} == 2)$). You may type "?pbc" to find more detailed explanation about this dataset. The "plot" and "summary" functions are very useful to examine the results. To use the functions in "survival" package, you will need to load the package first by typing

```
library(survival)

fit=survfit(Surv(time, (status==2))~1, data=pbc)

summary(fit)
```

   Try to obtain the median survival time and its 95% confidence interval by the output from the "summary" function. (For this problem, you don't need to hand in any output.)

2. Can you try to write R-script to estimate the area under the KM curve between 0 and τ for a given time point τ? (For this problem, you don't need to hand in your code. But it is a good exercise for your R programming.)

3. In R, "survdiff" under the R package "survival" can be used to perform the logrank test. Try the "survdiff" function to compare the survival distribution between D-penicillmain and placebo arms in pbc data, where "trt=1, 2" and "NA" representing D-penicillmain, placebo and not randomised, respectively. You also may use the command "plot" coupled with "survfit" to examine the differences in survival function graphically, e.g.

```
survdiff(Surv(time, (status==2))~trt, subset=(is.na(trt)==F),
data=pbc)

fit=survfit(Surv(time, (status==2))~trt, subset=(is.na(trt)==F),
data=pbc)

plot(fit, col=1:2, xlab="time(days)", ylab="survival probability")
```

The function "survdiff" has an argument ρ, which specifies the weight function $w(t) = \hat{S}(t)^{\rho}$, where $\hat{S}(t)$ is the KM estimator for the survival function with pooled samples. When $\rho = 0$, "survdiff" performs the regular logrank test. When $\rho = 1$, "survdiff" performs the weighted logrank test similar to the generalized Wilcoxon test (why?). Try different ρ in "survdiff" function. (For this problem, you don't need to hand in any output.)

4. If we expect that the survival time from two groups follow the Weibull distribution with the same shape parameter but possible different scale parameters, which ρ value should we use in the aforementioned "survdiff" function and why?