# Homework 8

*Aji John*

*Winter Quarter 2018*

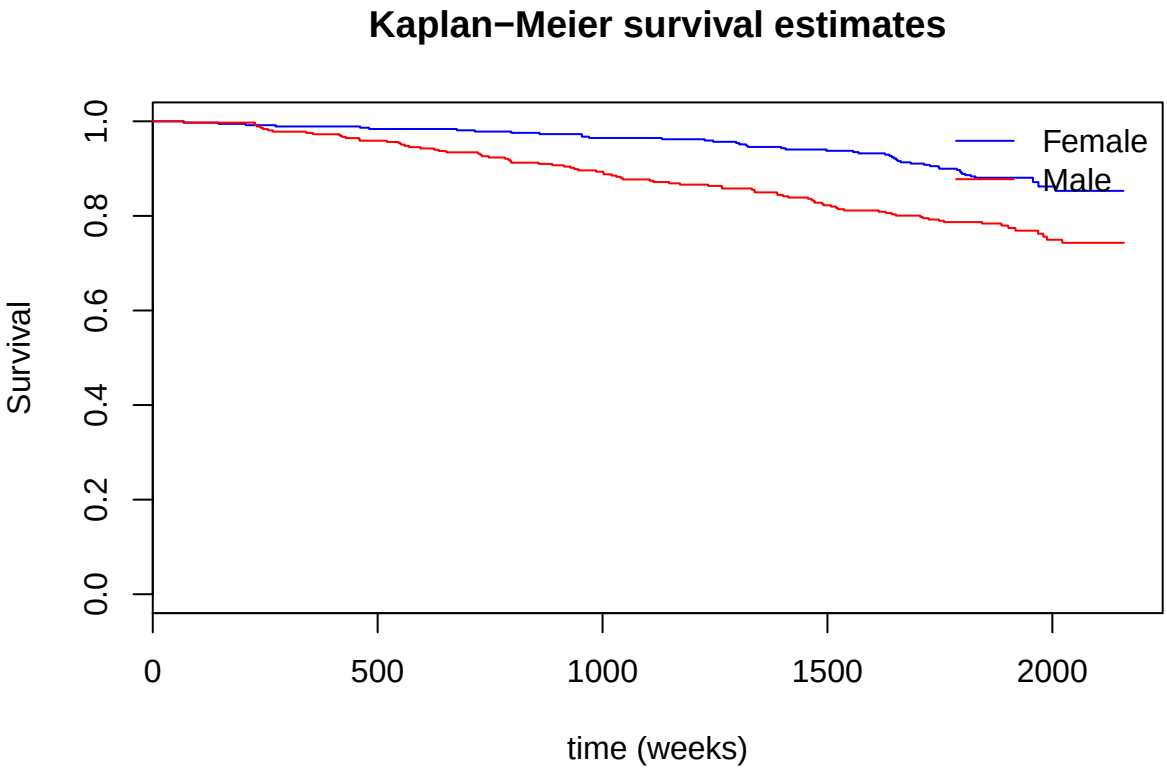
We are interested in assessing associations among mortality, creatinine, age, sex, and smoking behavior in a population of generally healthy elderly subjects in four U.S. communities

## Question 1

Suppose we are interested in any association between risk of all-cause mortality and sex using the Kaplan-Meier estimator of the survival function. Estimate survival functions for the two sex groups using the Kaplan-Meier estimator.

### (a)

Provide a plot with the Kaplan-Meier estimated survival functions for the two sex groups. The two Kaplan-Meier curves should appear on the same plot. Also briefly comment on any differences/similarity of the survival curves.




Lower survival rate in males in the censoring period than females.

### (b)

Is there an association between risk of all-cause mortality and sex based on the Kaplan-Meier survival estimates? Explain and provide appropriate statistical evidence supporting your reasoning.

Based on KM estimates, we find that there is an association between sex and all-cause mortality. We support it by doing an Chi-sq test betwwen the survival rate and two sex groups, and our test suggested it as significant(p-value $< 0.001$).

## Question 2

Now suppose we are interested in any association between risk of all-cause mortality and sex using a Cox proportional hazards regression model. Perform a Cox proportional hazards regression analysis of risk of mortality with sex as a predictor.

### (a)

Provide an interpretation of the exponentiated slope for sex in your proportional hazards regression model.

It is the hazard ratio which comes to be 1.962ˆsex. Males have instantaneous event rate 1.962 times higher (96.2% higher) than females.

### (b)

Provide full inference for an association between risk of mortality and sex from the Cox proportional hazards regression model in part a.

From proportional hazards regression analysis, we estimate that for two sex groups, the risk of death is 96.2% higher in males. This estimate is highly statistically significant ($P < .001$). A 95% CI suggests that this observation is not unusual if male group might have risk of death that was anywhere from 37.4% higher to 180% higher than the females.

### (c)

Compare the risk of mortality and sex association results from the Cox proportional hazards regression model to the association results in problem 1 obtained using the Kaplan-Meier method. Briefly discuss any differences in assumptions between the two methods.

The results between KM and Cox seem to be in agreement, the risk of death is higher in males than in females.

## Question 3

Now conduct a Cox proportional hazards regression analysis for risk of all-cause mortality with both sex and age at the time of study enrollment included as predictors.

## (a)

Provide an interpretation of the exponentiated slope for sex in your Cox proportional hazards regression model.

It is the hazard ratio which comes to be 1.903ˆsex. Males have instantaneous event rate 1.903 times higher (90.3% higher) than females.

## (b)

Provide an interpretation of the exponentiated slope for age in your Cox proportional hazards regression model.

It is the hazard ratio which comes to be 1.067ˆage Older age group have instantaneous event rate 1.067 times higher (6.7% higher) than younger group.

## (c)

Provide full inference for an association between risk of all-cause mortality and sex with the Cox proportional hazards regression model.

From proportional hazards regression analysis, we estimate that for two sex groups, the risk of death is 90.3% higher in males. This estimate is highly statistically significant ($P < .001$). A 95% CI suggests that this observation is not unusual if male group might have risk of death that was anywhere from 33.2% higher to 171.8% higher than the females.

## (d)

Provide full inference for an association between risk of all-cause mortality and age with the Cox proportional hazards regression model.

From proportional hazards regression analysis, we estimate that for two age groups who differ by 1 year, the risk of death is 6.7% higher in older age group This estimate is highly statistically significant ($P < .001$). A 95% CI suggests that this observation is not unusual if the older age group might have risk of death that was anywhere from 4% higher to 9.5% higher than the younger age group.

## (e)

Does age at the time of study enrollment confound the association between risk of all-cause mortality and sex? Explain and provide evidence to support your reasoning.

On doing ANOVA test with null model as 'Death' given age, and alternate model as 'Death' given age and sex, we find that age has significant effect the all cause mortality(p-value $< 0.001$ ). Furthermore, if age is assumed to confound the association between all-cause mortality and sex, it should have significant associations with both, but what we find is contrary, death and age are associated as it is found to be significant(no surprise! - p-value $<0.001$), and innately, age and sex are not associated(neverthless, p-value 0.4). Given the additional analysis, we can say with certainity that age at the time of enrollment did not confound the association, rather it contributed as a precision.

# Question 4

Now perform a Cox proportional hazards regression analysis of all-cause mortality with creatinine, age, sex, and indicator of ever smoked included as predictors.

## (a)

Provide full inference for an association between risk of all-cause mortality and creatinine with the proportional hazards regression model.

From proportional hazards regression analysis, we estimate that for two groups who differ by 1 unit(mg/DL) change in CRT year, the risk of death is 248.6% higher in the group that has the higher CRT. This estimate is highly statistically significant ($P < .001$). A 95% CI suggests that this observation is not unusual if the group with higher CRT might have risk of death that was anywhere from 127.33% higher to 434.7% higher than the group with lower CRT.

## (b)

Provide full inference for an association between risk of all-cause mortality and smoking with the proportional hazards regression model.

From proportional hazards regression analysis, we estimate that for two smoking groups(never smoked vs smoked), the risk of death is 33.5% higher in the group that has smoked. This estimate is found to be not statistically significant ($P = 0.12$). A 95% CI suggests that this observation is not unusual if the group that smoked might have risk of death that was anywhere from -7.98% lower to 93.6% higher than the group that has never smoked.

# Question 5

Now perform a Cox proportional hazards regression analysis and provide inference on whether sex modifies an age and ever smoked adjusted association between mortality and creatinine. Explain and provide full inference supporting your reasoning.

(Rephrased) Now perform a Cox proportional hazards regression analysis and provide inference on whether sex modifies the association between mortality and creatinine in a model that controls for age and ever smoked.

On doing ANOVA with null model assuming sex irrelevant(i.e. having mortality studied with respect to CRT, age and ever smoked indicator), and alternate model explaining it by sex(i.e. having mortality studied with respect to CRT, sex,age and ever smoked indicator), we find evidence that sex does not seem to modify the association between all cause mortality and CRT when controlled by age and smoking behaviour(ever smoked or not), p-value is found to be .13.