HW #4 due on 12/18 (Friday)

1. **Model Selection**

Data KM 1.13: Time to hospitalized pneumonia in young children

Data were gathered from 3,470 annual personal interviews conducted for the National Longitudinal Survey of Youth from 1979 to 1986 to study whether the mother’s feeding choice (breast fed and never breast fed) protected the infant against hospitalized pneumonia in the first year of life.

Variables included in the data are:

chldage: age in months child had pneumonia (This is the time to event variable)  
hospital: hospitalized pneumonia indicator

mthage: mother’s age in years at infant’s birth

urban: environment of the mother (1: urban; 0: rural)

alcohol: mother’s alcohol use per month by the mother (0: no drinks; 1:less than one drink; 2: 1~2 drinks; 3: 3~4 drinks; 4: more than 4 drinks)

smoke: mother’s cigarette use (0: none; 1: <1 pack/day; 2: ≥1 pack/day)

region: region of country (1: North East; 2: North Central; 3: South; 4: West)

poverty: mother’s poverty status (1: Yes; 0, No)

bweight: infant’s birthweight

race: mother’s race (1: White; 2: Black; 3: Other)

education: mother’s years of education

nsibs: number of siblings of infant

**wmonth: months the child was weaned**

**sfmonth: months the child on solid food**

**~~agepn: age child in the hospital for pneumonia, months~~**

First use wmonth to create breast feeding indicator (breast feeding: wmonth>0; no breast feeding: wmonth=0) and perform analysis to answer the following questions.

1. (10%) Perform appropriate descriptive analysis to summarize mother’s and infant’s characteristics by feeding choice. Do you notice any difference in the characteristics between two feeding choices?
2. (5%) Unadjusted analysis: Fit a Cox PH model to time to hospitalized pneumonia with the indicator of breast feeding as the only covariate in the model and then interpret the results.
3. (10%) Adjusted analysis: Fit a Cox PH model with the indicator of breast feeding as the covariate adjusting for other risk factors (except sfmonth and agepn) one at a time and then interpret the results.
4. (15%) Perform variable and model selections with the indicator of breast feeding always kept in the model. Don’t forget to specify the approach and criteria you choose for the selection.
5. (15%) Write a short paragraph to summarize your findings from the above analyses.

II. **Clinical trial of locally unresectable gastric cancer**

The objective of the trail was comparing overall survival between chemotherapy alone and combination of chemotherapy and radiotherapy.

Dataset: gastric.sas7bdat (This is SAS dataset you can directly use, not need to read in)

Variables

Time (1st column): time to death

Status (2nd column): death indicator (1: death; 0: alive)

Chemo (3rd column): chemotherapy indicator (1: chemo alone; 0: combination)

Gender: 0-male 1-female

Logincome: log(income), scale has been changed so don’t judge whether it makes sense

or not

Use this dataset to answer the following questions.

1. (10%) Summarize time to death, i.e. so-called overall survival (OS), by treatment group (i.e. chemotherapy alone and combination treatment). Do you notice anything interesting regarding the Kaplan-Meier survival curves and smoothed hazard curves?
2. (5%) Compare OS between treatment groups using Cox proportional hazards regression. Report the hazard ratio and the associated 95% CI and then interpret the results. (Include all covariates)
3. (5%) Conduct model diagnostics of the model in Q2 to examine the overall fit of the model.
4. (5%) It is important to check the proportional hazard assumption while fitting a Cox proportional hazards model. Perform an appropriate analysis to confirm that the hazard rates for the two treatment groups are not proportional. (Only test for chemo variable)
5. (10%) Use a proper method to address the issue found in (4) and analyze the treatment effects. Please write down the partial likelihood for this method. (Only address the potential problem associated with chemo)
6. (5%) Can one perform the likelihood ratio test to compare Q2 and Q5 models? If Yes, perform the test. If no, explain and then propose a criterion and conduct the comparison.
7. (5%) Please examine the assumption of the method used in (5).