7.1 We consider the duration of breast feeding of women following birth. The `breastfeeding` data contains the following variables:

- `duration`: duration of breast feeding (in weeks)
- `delta`: indicator variable, `1` for completed breastfeeding, `0` for right-censored data
- `race`: race of mother, one of white (`1`), black (`2`), or other (`3`)
- `poverty`: is mother revenue below poverty line? either yes (`1`) or no (`0`)
- `smoke`: smoking status of mother at birth of child, either yes (`1`) or no (`0`)
- `alcohol`: alcohol-drinking status of mother at birth of child, either yes (`1`) or no (`0`)
- `agemth` age of mother at birth of child
- `ybirth` year of child's birth
- `yschool`: years of school of the mother
- `pc3mth`: binary indicator, `0` if mother sought prenatal care in first three months of pregnancy, `1` otherwise.

(a) Estimate and plot the survival curves for the two levels of the smoker status variable `smoke` using the Kaplan–Meier estimator. Comment on the plotted survival curves.

(b) Based on survival curve estimated using Kaplan–Meier's estimator, what is the estimated probability that a mother who is a non-smoker will breastfeed for more than 36 weeks? What about for a mother who is a smoker?

(c) What is the median and mean number of weeks that a mother who is a non-smoker will breastfeed? What about for a mother who is a smoker?

(d) Test whether the survival curves for non-smoker and smoker mothers are the same.

(e) Fit a Cox proportional hazards model to evaluate the impact of the poverty status, smoker status, age of the mother and years of schooling on the hazard. Write down the equation of the estimated hazard function and interpret the estimated $\widehat{\boldsymbol{\beta}}$ parameters.

7.2 A shoe store in Montreal wishes to know how long it takes before products are sold. The dataset `shoes` contains the following variables:

- `status`: categorical variable, `0` for sales, `1` if the article is still in stock, `2` if destocked.
- `time`: storage time of the article (in months).
- `price`: sale price, rounded to the nearest dollar.
- `gender`: binary variable for gender, `0` for men shoes, `1` for women shoes.

Our ojective is to estimate the survival time of items in stock.

(a) What does censoring represent in this example?

(b) Estimate the survival function of the stocking time using Kaplan–Meier estimator and report the estimated quartiles of the survival time.

(c) Fit a Cox proportional hazard model for stocking time as a function of shoes gender and sale price. and report the estimated coefficients, $\widehat{\boldsymbol{\beta}}$. Which of the covariates impacts hazard of staying in store, if any?

Compiled 02/12/2020 at 15:11