Année 2019-2020

Survival and longitudinal data analysis

Exercise 8.11 of Klein and Moeschberger 2005

Exercice 1

We study data gathered from annual personal interviews conducted for the National Longitudinal Survey of Youth (NLSY) from 1979 through 1986. This data was used to study whether or not the mother's feeding choice protected the infant against hospitalized pneumonia in the first year of life. Ages of young children at the time they were hospitalized with pneumonia were recorded as well as the observed ages of those infants that were not hospitalized with pneumonia during the study period. The data `pneumon` is available in R package `KMsurv`. Use the discrete method for handling ties in the following.

1. Check if the variables have been correcly imported, especially the factors.
2. Plot the Kaplan-Meier estimator for the survival function of the age at pneumonia. Give an estimation and a confidence interval for the probability for a newborn of not having developed pneumonia at 6 months.
3. Construct a dummy variable $Z = 1$ if infants were breast fed at birth, 0 if infants were never breast fed, and test the hypothesis $\mathcal{H}_0$: there is not difference in distributions of age at first pneumonia between child whether were breast fed or not.
4. : Test the hypothesis $\mathcal{H}_0 : \beta^\star_{\text{breastf}} = 0$, i.e., the survival functions for the two types of breast feeding are equal, using the likelihood ratio, and Wald tests. Find the estimate of $\beta^\star_{\text{breastf}}$, $\hat{\beta}^\star_{\text{breastf}}$, its standard error, and the relative risk.
5. Also available in the data set is information on other factors that may be associated with the timing of hospitalized pneumonia. These factors are age of the mother at the infant's birth, rural-urban environment of the mother, use of alcohol by the mother (no drinks, less than one drink, 1-2 drinks, 3-4 drinks, or more than 4 drinks per month), mother's cigarette use (none, less than 1 pack/day, 1 or more pack/day), region of country (northeast, north central, south, or west), birthweight of infant (less the 5.5 lbs or 5.5 lbs or more), poverty status of mother (yes/no), race of mother (white, black, or other), or number of siblings of infant. Test the hypothesis that the times to hospitalized pneumonia are the same for the two feeding groups adjusting for each of these factors in a separate model using the Wald test.
6. Since one is primarily interested in comparing the two types of breast feeding, interest will center upon building a model with the view of testing the particular comparison of interest adjusting for the other non controllable fixed covariates in question 4. Build such a model using the Wald test.
7. In the final model, predict the probability of not having developed pneumonia at 6 months for a newborn whith covariates

```
mthage urban alcohol smoke region poverty bweight race education nsibs
    27     1       3     0      2       1       0    1        12     1
wmonth sfmonth agepn
     0       0     4
```

References

[KM05] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.