Methods in Epidemiologic Research

Sample Problems

Chapter 4 – Measures of Disease Frequency

Preparation

In Chapter 19 data on factors affecting survival after admission to hospital with myocardial infarction are analysed using survival analysis methods. For the following exercises, we will focus on survival over a 5 year period following admission.

The first thing you will have to do is generate three new variables.

- Create a 0/1 variable which indicates whether or not the individual died during the 5 years following admission.
- Create a variable indicating the amount of time that the person was under observation. This will be 5 years if the person survived for the full period. If they died, or were lost to follow up (ie "censored") before the end of the 5 year period, it will be the time until death of censoring. When computing this variable, take the following 2 issues into consideration.
  - Before computing this variable, add 1 to the original survival time variable (-surv_mi-) because people who died on the day of admission were listed as having a "survival" of 0 days, but be want them to have a "time under observation" of 1 day.
  - Convert the time from days to years (by dividing by 365.25). This is just to avoid having very small values when estimating rates.
- Generate a 0/1 variable indicating whether or not the person survived the full 5 year period.

Questions

1. First determine how many individuals died, survived or were censored (withdrawals) during the 5 year period.
2. Estimate the 5-year risk of death in 2 ways.
   (a) First, assume (incorrectly) that all individuals were followed for the full 5 year period (ie no withdrawals).
   (b) Now adjust the estimate to take into account that there were quite a few withdrawals.
   (c) Discuss whether or not this is a good estimate of the 5-year risk of death.
3. Estimate the annual incidence rate of deaths in two ways.
   (a) First, do an approximate calculation based on the assumption that withdrawals took place ½ way through the follow up period
   (b) Second, perform an exact calculation based on each person's actual time under observation.
   (c) Discuss whether or not this is a good estimate of the annual incidence rate. Is this estimate a good way to express the frequency of deaths of people in this population? If not, what is its main limitation?
4. Estimate both the 1-year and 5-year risks of death based on the incidence rate computed above.

5. It makes no sense to talk about the "prevalence of death" so instead, compute the prevalence of obesity (defined as bmi>30) among patients admitted to hospital.
6. Compute the standard error (SE) and confidence interval (CI) for the crude 5-year risk estimate from 2(a). Compute both approximate and exact estimates. Are they similar? If so, why? If not, why not?
7. Compute the SE and CI for the annual incidence rate estimate from 3(b).
8. We will now compare the incidence rates across hospitals. However, first create a 4-level categorical variable for -age- with age divided into 4 groups (0-59.99, 60-74.99, 75-84.99 and >=85).
   (a) Determine if the distribution of age categories appears to vary across hospitals (ie do some hospitals tend to see younger or older patients than other hospitals).
   (b) Determine if the crude 5-year risk of death (ignoring withdrawals) appears to differ across hospitals.
   (c) Compute age-standardized risks for the 10 hospitals. Is there more, or less, variation between the crude or age-standardized risks?