Exercises 11 · Survival models

Due Monday, May 2, 2016

(1) PREDIMED

Revisit the "predimed.csv" data from the PREDIMED study:

> Estruch R, Ros E, Salas-Salvado J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med 2013; 368:1279-1290. Full text available at `http://www.nejm.org/doi/full/10.1056/NEJMoa1200303`

The data set has the following variables:

- group: a factor with levels Control MedDiet + Nuts MedDiet + VOO. Intervention group.
- sex: a factor with levels Male Female.
- age: Age.
- smoke: a factor with levels Never Current Former. Smoking status.
- bmi: Body mass index.
- waist: Waist circumference.
- wth: Waist-to-height ratio.
- htn: a factor with levels No Yes. Hypertension.
- diab: a factor with levels No Yes. Type-2 diabetes.
- hyperchol a factor with levels No Yes. Dyslipidemia.
- famhist: a factor with levels No Yes. Family history of premature CHD.
- hormo: a factor with levels No Yes. Hormone-replacement therapy.
- p14: Mediterrean Diet Adherence score.
- toevent: follow-up time to main event (years).
- event a factor with levels No Yes. Yes means AMI, stroke, or CV Death. No means that the observation was right-censored at the time given in the toevent variable.

Your task is to explain and replicate Table 3. Specifically, what do the parts of the table labeled "Hazard ratio for each Mediterranean diet vs. control (95% CI)" and "Hazard ratio for Mediterranean diets combined vs. control (95% CI)" mean? Can you replicate these numbers using the data at hand? (All three rows: unadjusted, multivariate adjusted 1, and multivariable adjusted 2.) Explain your replication approach.

Two notes here. First, focus only the primary endpoint part of the table; your data set doesn't distinguish between the different types of endpoints, Second, your data set has only about 85% of the people in the full data set, so it is possible that your numbers won't match exactly. Just do your best for as close a replication as you can.