EPIB-609: Assignment based on material listed under

"<u>Rates: models</u>"

The <u>learning aims</u> for the readings and the exercise are to review (or if you haven't covered them before, be introduced to) the (frequentist) statistical inference procedures for (i) a single (incidence-density-type) rate (ii) comparisons of 2 such (crude) rates (iii) the (classical) corresponding comparison for stratified data and (iv) the (modern) regression counterparts of (i)-(iii).

    Despite the centrality of counts, time, and (incidence-density-type) rates in epidemiologic research, epidemiologic-data-analysis courses often skip these topics, and favour logistic regression and survival analysis. Moreover, there is the feeling that classical (non-regression) methods for stratified data are old fashioned and not as sophisticated. But sometimes even sophisticated 'types' have trouble explaining their 'black boxes' to journalists and other lay people.

    For those of you who have not covered these topics before, this is a more extensive assignment than some of the others, and needs to be begun well ahead of the class. To compensate for this, some of the answers (and R code) are supplied.

a. Review "Statistical models for inference re rates. 634 Notes. 2pp" before attempting questions 1-3 in the assignment. Compare your answers with the answers supplied in the c634 website.

b. Make a written list of the items/concepts in the Notes or in the answers that you still do not fully understand and bring the list with you to the session: be specific as to what it is that is unclear.

c. Use an 'eye-fit' to obtain estimates of the 6 parameters of the 'model' in Question 4(i) – Denmark data. Put these into a written table.

d. Why would it be better to fit rate models to the observed *counts*, and their corresponding PT *denominators* than to the rates themselves? Put the answer, and the table, in the same document used for b. If you get stuck, you can consult the answers supplied for the 7-parameter and the expanded (3-period) data set, exercise in the c634 website (opposite "Using regression to describe patterns in rates").

634 Assignment on 'Rate Regression'

a. Review "Regression Models for Rates - 1 pp Summary + 5 pp details" and make a written list of the items/concepts that you still do not fully understand; bring the list with you to the session: be specific as to what it is that is unclear.

b. Use your own favourite software to carry out the Analysis of IHD data in C&H Table 22.6. Obtain and include in your document the Observed minus Fitted residuals for the 6 counts under each of the models fitted. {Again, you can consult the answers opposite "Regression models for (incidence) rates" in the c634 website.

c. Compare the value of the summary rate ratio obtained using the classical Mantel-Haenszel estimator with that obtained using the (corresponding) Poisson regression model. Which approach uses more assumptions?

The Danish studies of OCs

a. Describe, in writing, in a way that your RA could follow, how would you set up the datasets for the data-analysis carried out in these Danish studies.

A few (one?) data point(s) and a model (c.f. '3 men and baby'). How small?

a. Make up small datasets (the smaller the better) you could use to illustrate
(i) Least Squares (LS) fitting of a small (the smaller the better) regression model
(ii) Maximum Likelihood (ML) fitting of a small rate-regression model
(you might try the Flash applet: `http://www.epi.mcgill.ca/hanley/2DatapointsAndaModel/`)

2012.08.20