

Bios 312: Modern Regression Analysis

April 5, 2012

HW #6

All problems refer to the salary dataset as found on the class web pages. As in previous analyses, only use the data from 1995 for this analysis. This is a continuation of the analysis begun in the lab on 4/4, so you can use that .R script file or .do file as a starting point for this homework.

1. We are interested in making inference about the difference in the mean monthly salary paid to faculty in 1995 according to the year in which faculty obtained their degree and the year in which they started. In all models in this problem, we will appropriately adjust for degree, field, administrative duties, and sex.
 - a. Provide inference about the adjusted association between monthly salary and year of degree (modeled as a linear continuous variable, not adjusted for starting year).
 - b. Provide inference about the adjusted association between monthly salary and starting year (modeled as a linear continuous variable, not adjusted for year of degree).
 - c. Provide inference about the adjusted association between monthly salary and year of degree (modeled as a linear continuous variable, and adjusted for starting year as well as the other variables).
 - d. Provide inference about the adjusted association between monthly salary and starting year (modeled as a linear continuous variable, and adjusted for year of degree as well as the other variables).
 - e. Briefly discuss the scientific relevance between the results obtained in parts a, b and parts c, d of this problem.

Problems 2 – 4 ask you to fit a series of models in which you consider a hierarchy of adjusted analyses for each of three different summary measures. Your response to these problems might be best presented in a table of inference about the adjusted association between monthly salary and sex.

For the benefit of the grader, we will agree on modeling *yrdeg* and *startyr* as linear splines with knots at years 1965, 1975, and 1985. These were obtained in the lab (see fit F).

2. We are interested in making inference about the difference in the mean monthly salary paid to women faculty in 1995 and that paid to men faculty in 1995.
 - a. Report inference regarding the unadjusted comparison of women's and men's salaries.
 - b. Report inference regarding the comparison of women's and men's salaries after adjustment for degree.

- c. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree.
 - d. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, and starting year.
 - e. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, and field.
 - f. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, field, and administrative duties. Save the predicted values of the mean salary for each individual as *fit2*.
 - g. Using model 2.f, test if *field* is associated with salary after adjusting for gender, degree, year of degree, starting year, and administrative duties. Report the p-value and the test used.
3. We are interested in making inference about the ratio of geometric mean monthly salary paid to women faculty in 1995 and that paid to men faculty in 1995.
- a. Report inference regarding the unadjusted comparison of women's and men's salaries.
 - b. Report inference regarding the comparison of women's and men's salaries after adjustment for degree.
 - c. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree.
 - d. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year.
 - e. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, field.
 - f. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, field, administrative duties. Save the predicted values of the geometric mean salary for each individual as *fit3*.
 - g. Using model 3.f, test if *field* is associated with salary after adjusting for gender, degree, year of degree, starting year, and administrative duties. Report the p-value and the test used.
4. We are interested in making inference about the ratio of the mean monthly salary paid to women faculty in 1995 and that paid to men faculty in 1995.
- a. We can "trick" Stata into performing these analyses using Poisson regression with robust standard error estimates. Why does this work? Why is it crucial to use the robust standard errors?

- b. Report inference regarding the unadjusted comparison of women's and men's salaries.
 - c. Report inference regarding the comparison of women's and men's salaries after adjustment for degree.
 - d. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree.
 - e. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year.
 - f. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, field.
 - g. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year, field, administrative duties. Save the predicted values of the mean salary for each individual as *fit5*.
 - h. Using model 4.g, test if *field* is associated with salary after adjusting for gender, degree, year of degree, starting year, and administrative duties. Report the p-value and the test used.
5. Briefly discuss the similarities and differences between the analyses performed in problems 3 – 5. How similar are the predicted values between the models? How different is the inference you would obtain. *A priori*, which set of analyses would you prefer when answering the question regarding sex discrimination?
 6. For the analysis model of your choice (either question 2, 3, or 4), summarize the scientific relevance of the results from the 6 different models used in the comparison of salaries paid to women and men.