**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2014

**Homework #8**

February 28, 2014

**Written problems:** To be submitted as a MS-Word compatible file to the class Catalyst dropbox by 9:30 am on Friday, March 7, 2014. See the instructions for peer grading of the homework that are posted on the web pages.

*On this (as all homeworks) Stata / R code and unedited Stata / R output is* ***TOTALLY*** *unacceptable. Instead, prepare a table of statistics gleaned from the Stata output. The table should be appropriate for inclusion in a scientific report, with all statistics rounded to a reasonable number of significant digits. (I am interested in how statistics are used to answer the scientific question.)*

***Unless explicitly told otherwise in the statement of the problem, in all problems requesting “statistical analyses” (either descriptive or inferential), you should present both***

* ***Methods: A brief sentence or paragraph describing the statistical methods you used. This should be using wording suitable for a scientific journal, though it might be a little more detailed. A reader should be able to reproduce your analysis. DO NOT PROVIDE Stata OR R CODE.***
* ***Inference: A paragraph providing full statistical inference in answer to the question. Please see the supplementary document relating to “Reporting Associations” for details.***

All problems refer to the salary dataset as found on the class web pages. This is a very large file, so you need to make sure you have sufficient memory available when you start Stata. Also, it is probably most convenient if you code the variables as numbers, and use labels to make them more understandable. The following file on the Datasets web pages contains commands you might find useful.

http://www.emersonstatistics.com/datasets/initsalary.doc

1. 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. In this problem, we focus on alternative modeling of the variables *yrdeg* and *startyr*. In all models in this problem, we will appropriately adjust for degree, field, administrative duties, and sex. ***(Note that I have provided answers to all parts of this problem except parts a, b and i, which you should answer.)***
   1. In all parts of this problem, in addition to the year of degree and year starting at the UW, you should adjust for the highest degree obtained, field, and administrative duties. What is the best way to model the variables *degree, field,* and *admin*? Briefly justify your answer.

**Ans: because we want to make inference about the difference in the mean monthly salary among women and men, multiple linear regression would be used, and modelling the three variables as categorical variables.**

* 1. In all parts of this problem you should use robust standard error estimates. Briefly explain why inference based on classical linear regression (without robust SE estimates) would be incorrect. Do you think the classical linear regression inference would tend to be conservative or anti-conservative? Justify your answer.

**Ans: The classical linear regression is based on the assumptions that the dependent variable is linearly related to the coefficients of the model and within group variances are the same across groups; however, we didn’t have that evidence to show that linearity and within group variances are the same across the group. The inference from the classical linear regression would tend to anti-conservative, because it uses a smaller standard error.**

* 1. Model *yrdeg* and *startyr* as linear continuous variables. Report the inference you would make for the difference in mean salaries for men and women (a table of the results for parts c, d, e, f, and g will be sufficient).

**Ans: (See table below)**

* 1. Model *yrdeg* and *startyr* as quadratic continuous variables (so linear continuous plus a second order term). Report the inference you would make for the difference in mean salaries for men and women (a table of the results for parts c, d, e, f, and g will be sufficient).

**Ans: (See table below)**

* 1. Model *yrdeg* and *startyr* as dummy variables for groups defined by earlier than 1960, 1960-64, 1965-69, 1970-74, 1975-79, 1980-84, 1985-89, and 1990 or later. Report the inference you would make for the difference in mean salaries for men and women (a table of the results for parts c, d, e, f, and g will be sufficient).

**Ans: (See table below)**

* 1. Model *yrdeg* and *startyr* as linear splines with knots at years 1960, 1965, 1970, 1975, 1980, 1985, and 1990. Report the inference you would make for the difference in mean salaries for men and women (a table of the results for parts c, d, e, f, and g will be sufficient).

**Ans: (See table below)**

* 1. Repeat parts c – f when modeling the ratio of mean salaries across sexes and when modeling the ratio of geometric mean salaries across sexes. These results can be included in the same table.)

**Ans: (See table below)**

* 1. Examine the agreement between the inference about the adjusted association between monthly salary and sex. Did the inference vary substantially across the various models?

**Ans: The following table provides the regression parameter estimates for the predictor indicating female sex, its Z statistic, its two-sided P value, and its 95% CI for the alternative methods of modeling year of degree and starting year. A few comments are in order**

* **In all cases, the linear splines provided the best fit to the data in the sense that adding the linear splines to each of the other models proved to be statistically significant. Adding the dummy variables to the model that included the linear splines did not improve the fit. I do not recommend doing this sort of testing unless your question was about the form of the relationship (e.g., linear vs nonlinear). My point here is that the linear splines did seem to model the true relationship with salary better when I was modeling sex, field, degree, and administrative duties.**
* **When modeling year of degree and start year as quadratic functions, I could not statistically establish nonlinearity in the linear regression model of the difference of means. When considering ratios of means or geometric means, I could detect the nonlinearity of either the year of degree or starting year when testing them combined, but because the terms are so correlated, I could not ensure that both were nonlinear when adjusting for the other.**
* **When modeling year of degree and start year as dummy variables or linear splines, there tended to be statistically significant departures from linearity for each variable separately and combined.**
* **Note that I included the Z statistic in this table only because the results were so strikingly statistically significant, that is only through looking at the Z statistic that we can assess whether there were any substantial differences (there were not).**
* **Note the similarity in ratios across all methods of modeling year of degree and start years and across the summary measures (means or geometric means).**
* **I provided inference about ratios of means using both Poisson regression and the generalized linear model when assuming Gaussian data with a log link. I prefer the Poisson regression, though this really only makes a big difference when looking at risk ratios with binary data. In that case, I *highly* recommend using Poisson regression rather than the generalized linear model with the binomial family and the log link. With means of positive continous random variables Poisson regression or the Gaussian GLM will both tend to behave okay.**
* **Lastly, the difference in means is of course a very different scale than the ratios of means or geometric means. But if you consider that the mean monthly salary for the entire sample was $6,389.81, the difference in means of about $420 is about 7% of the overall mean. So all models are giving quite similar answers.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Z** | **P Value** | **95% CI low** | **95% CI high** |
| *Difference in Means* | | | | | |
| **Linear** | -428.3 | -5.23 | < .0001 | -588.9 | -267.8 |
| **Quadratic** | -428.1 | -5.25 | < .0001 | -588.1 | -268.0 |
| **Dummy** | -447.7 | -5.45 | < .0001 | -609.0 | -286.5 |
| **Splines** | -419.7 | -5.17 | < .0001 | -579.0 | -260.5 |
| *Ratio of Means (Poisson)* | | | | | |
| **Linear** | 0.9266 | -5.42 | < .0001 | 0.9014 | 0.9525 |
| **Quadratic** | 0.9280 | -5.36 | < .0001 | 0.9030 | 0.9537 |
| **Dummy** | 0.9244 | -5.63 | < .0001 | 0.8994 | 0.9500 |
| **Splines** | 0.9289 | -5.34 | < .0001 | 0.9041 | 0.9544 |
| *Ratio of Means (GLM)* | | | | | |
| **Linear** | 0.9227 | -5.55 | < .0001 | 0.8969 | 0.9493 |
| **Quadratic** | 0.9246 | -5.43 | < .0001 | 0.8988 | 0.9511 |
| **Dummy** | 0.9185 | -5.83 | < .0001 | 0.8926 | 0.9451 |
| **Splines** | 0.9245 | -5.49 | < .0001 | 0.8989 | 0.9508 |
| *Ratio of Geometric Means* | | | | | |
| **Linear** | 0.9347 | -5.22 | < .0001 | 0.9113 | 0.9587 |
| **Quadratic** | 0.9352 | -5.22 | < .0001 | 0.9119 | 0.9590 |
| **Dummy** | 0.9328 | -5.42 | < .0001 | 0.9096 | 0.9566 |
| **Splines** | 0.9363 | -5.17 | < .0001 | 0.9132 | 0.9600 |

* 1. In a real situation, how would choose among the alternative methods for adjusting for year of degree and starting year?

**Ans: I would choose the method in f, model year of degree and starting year as linear splines. Because this method provided the best fit to the data.**

1. We are interested in making inference about the difference in the mean monthly salary paid to faculty according to the year in which faculty obtained their degree and the year in which they started at UW. In all models in this problem, we will appropriately adjust for degree, field, administrative duties, and sex.
   1. Provide inference about the adjusted association between monthly salary and year of degree (modeled as a linear continuous variable, not adjusted for starting year).

**Intercept: Mean of salary for male subjects with year of degree in 1900 and in arts field, with no administrative duties, and other degree is 11636.5. There is no scientific relevance here, because there were no such people in our sample.**

**Female slope: mean of monthly salary is 420.2 lower for females compared to males with similar year of degree, and degree, field, and administrative duties. (95% CI: 579.3 lower to 261.2 lower for females compared to males) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between females and males having similar year of degree, and degree, field, and administrative duties: P < 0.0005.**

**Admin slope: mean of monthly salary is 1443.8 higher for subjects with administrative duties compared to subjects without administrative duties, with similar year of degree, and degree, field, and sex. (95% CI: 1169.3 higher to 1718.2 higher for subjects with administrative duties compared to subjects without administrative duties) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between subjects with administrative duties compared to subjects without administrative duties, having similar year of degree, and degree, field, and sex: P < 0.0005.**

**Field slope: mean of monthly salary is 783.8 higher other field compared to arts field (95% CI: 605.1 higher to 962.4 higher for other field compared to arts field), and 1953.1 for professional field compared to arts field (95% CI: 1712.6 higher to 2193.7 higher for professional field compared to arts filed), with similar year of degree, and degree, sex, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different field having similar year of degree, and degree, sex, and administrative duties: P < 0.0005.**

**Degree slope: mean of monthly salary is 685.1 higher for subjects with PhD degree compared to subjects with other degree (95% CI: 438.6 higher to 931.6 higher for PhD degree compared to other degree), and 1000.7 higher for subjects with professional degree compared to other degree (95% CI: 612.2 higher to 1389.1 higher for professional degree compared to other degree), with similar year of degree, and sex, field, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different degrees having similar year of degree, and sex, field, and administrative duties: P < 0.0005.**

**Year of degree slope: mean of monthly salary is 89.9 lower for each year difference in year of degree between two groups with similar sex, and degree, field, and administrative duties, with later year getting the degree having a lower salary. (95% CI: 98.3 lower to 81.4 lower for each year difference in year of degree) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in year of degree, but having similar sex, and degree, field, and administrative duties: P < 0.0005.**

* 1. Provide inference about the adjusted association between monthly salary and starting year (modeled as a linear continuous variable, not adjusted for year of degree).

**Intercept: Mean of salary for male subjects with starting year in 1900 and in arts field, with no administrative duties, and other degree is 9741.5. There is no scientific relevance here, because there were no such people in our sample.**

**Female slope: mean of monthly salary is 704.5 lower for females compared to males with similar starting year, and degree, field, and administrative duties. (95% CI: 879.7 lower to 529.2 lower for females compared to males) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between females and males having similar starting year, and degree, field, and administrative duties: P < 0.0005.**

**Admin slope: mean of monthly salary is 1611.0 higher for subjects with administrative duties compared to subjects without administrative duties, with similar starting year, and degree, field, and sex. (95% CI: 1322.4 higher to 1899.7 higher for subjects with administrative duties compared to subjects without administrative duties) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between subjects with administrative duties compared to subjects without administrative duties, having similar starting year, and degree, field, and sex: P < 0.0005.**

**Field slope: mean of monthly salary is 885.1 higher other field compared to arts field (95% CI: 693.6 higher to 1076.6 higher for other field compared to arts field), and 1988.0 for professional field compared to arts field (95% CI: 1724.9 higher to 2251.2 higher for professional field compared to arts filed), with similar starting year, and degree, sex, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different field having similar starting year, and degree, sex, and administrative duties: P < 0.0005.**

**Degree slope: mean of monthly salary is 279.5 higher for subjects with PhD degree compared to subjects with other degree (95% CI: 32.3 higher to 526.7 higher for PhD degree compared to other degree), and 911.0 higher for subjects with professional degree compared to other degree (95% CI: 497.2 higher to 1324.7 higher for professional degree compared to other degree), with similar starting year, and sex, field, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different degrees having similar starting year, and sex, field, and administrative duties: P < 0.0005.**

**Startyr slope: mean of monthly salary is 56.9 lower for each year difference in starting year between two groups with similar sex, and degree, field, and administrative duties, with later starting year with a lower salary (95% CI: 66.1 higher to 47.6 higher for each year difference in starting year) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in starting year, but having similar sex, and degree, field, and administrative duties: P < 0.0005.**

* 1. 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).

**Intercept: Mean of salary for male subjects with year of degree in 1900, and starting year in 1900, and in arts field, with no administrative duties, and other degree is 11060.0. There is no scientific relevance here, because there were no such people in our sample.**

**Female slope: mean of monthly salary is 428.3 lower for females compared to males with similar year of degree, starting year, and degree, field, and administrative duties. (95% CI: 588.9 lower to 267.8 lower for females compared to males) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between females and males having similar starting year, year of degree, and degree, field, and administrative duties: P < 0.0005.**

**Admin slope: mean of monthly salary is 1451.3 higher for subjects with administrative duties compared to subjects without administrative duties, with similar year of degree, starting year, and degree, field, and sex. (95% CI: 1180.3 higher to 1722.4 higher for subjects with administrative duties compared to subjects without administrative duties) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between subjects with administrative duties compared to subjects without administrative duties, having similar starting year, year of degree, and degree, field, and sex: P < 0.0005.**

**Field slope: mean of monthly salary is 735.1 higher other field compared to arts field (95% CI: 559.8 higher to 910.4 higher for other field compared to arts field), and 1906.2 for professional field compared to arts field (95% CI: 1669.7 higher to 2142.7 higher for professional field compared to arts filed), with similar year of degree, starting year, and degree, sex, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different field having similar year of degree, starting year, and degree, sex, and administrative duties: P < 0.0005.**

**Degree slope: mean of monthly salary is 797.8 higher for subjects with PhD degree compared to subjects with other degree (95% CI: 538.4 higher to 1057.2 higher for PhD degree compared to other degree), and 1044.2 higher for subjects with professional degree compared to other degree (95% CI: 651.1 higher to 1437.3 higher for professional degree compared to other degree), with similar year of degree, starting year, and sex, field, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different degrees having similar year of degree, starting year, and sex, field, and administrative duties: P < 0.0005.**

**Startyr slope: mean of monthly salary is 27.2 higher for each year difference in starting year between two groups with similar year of degree, sex, and degree, field, and administrative duties, with later starting year with a higher salary. (95% CI: 8.7 higher lower to 45.6 higher for each year difference in starting year) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in starting year, but having similar year of degree, sex, and degree, field, and administrative duties: P = 0.004.**

**Year of degree slope: mean of monthly salary is 112.0 lower for each year difference in year of degree between two groups with similar starting year, sex, and degree, field, and administrative duties, with later year getting the degree having a lower salary. (95% CI: 130.6 lower to 93.3 lower for each year difference in year of degree) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in year of degree, but having similar starting year, sex, and degree, field, and administrative duties: P < 0.0005.**

* 1. 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).

**Intercept: Mean of salary for male subjects with year of degree in 1900, and starting year in 1900, and in arts field, with no administrative duties, and other degree is 11060.0. There is no scientific relevance here, because there were no such people in our sample.**

**Female slope: mean of monthly salary is 428.3 lower for females compared to males with similar year of degree, starting year, and degree, field, and administrative duties. (95% CI: 588.9 lower to 267.8 lower for females compared to males) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between females and males having similar starting year, year of degree, and degree, field, and administrative duties: P < 0.0005.**

**Admin slope: mean of monthly salary is 1451.3 higher for subjects with administrative duties compared to subjects without administrative duties, with similar year of degree, starting year, and degree, field, and sex. (95% CI: 1180.3 higher to 1722.4 higher for subjects with administrative duties compared to subjects without administrative duties) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between subjects with administrative duties compared to subjects without administrative duties, having similar starting year, year of degree, and degree, field, and sex: P < 0.0005.**

**Field slope: mean of monthly salary is 735.1 higher other field compared to arts field (95% CI: 559.8 higher to 910.4 higher for other field compared to arts field), and 1906.2 for professional field compared to arts field (95% CI: 1669.7 higher to 2142.7 higher for professional field compared to arts filed), with similar year of degree, starting year, and degree, sex, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different field having similar year of degree, starting year, and degree, sex, and administrative duties: P < 0.0005.**

**Degree slope: mean of monthly salary is 797.8 higher for subjects with PhD degree compared to subjects with other degree (95% CI: 538.4 higher to 1057.2 higher for PhD degree compared to other degree), and 1044.2 higher for subjects with professional degree compared to other degree (95% CI: 651.1 higher to 1437.3 higher for professional degree compared to other degree), with similar year of degree, starting year, and sex, field, and administrative duties. Both of these results are highly atypical of what we might expect with no true difference in the mean of monthly salary between different degrees having similar year of degree, starting year, and sex, field, and administrative duties: P < 0.0005.**

**Startyr slope: mean of monthly salary is 27.2 higher for each year difference in starting year between two groups with similar year of degree, sex, and degree, field, and administrative duties, with later starting year with a higher salary. (95% CI: 8.7 higher lower to 45.6 higher for each year difference in starting year) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in starting year, but having similar year of degree, sex, and degree, field, and administrative duties: P = 0.004.**

**Year of degree slope: mean of monthly salary is 112.0 lower for each year difference in year of degree between two groups with similar starting year, sex, and degree, field, and administrative duties, with later year getting the degree having a lower salary. (95% CI: 130.6 lower to 93.3 lower for each year difference in year of degree) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between two groups with each difference in year of degree, but having similar starting year, sex, and degree, field, and administrative duties: P < 0.0005.**

* 1. Briefly discuss the scientific relevance between the results obtained in parts a,b and parts c,d of this problem.

**The results in part a and b are different in the intercepts, slopes estimates, and their according inference. Because they are based on different models; and one is answering the association between salary and year of degree, and another one is answering the association between salary and starting year. When coefficient of starting year parameter answered how the association between the starting year and the salary, with later starting year tending to be with a higher salary. On the other hand, the coefficient of year of degree showed that with later year of degree, the salary was lower. When we looked at the model used in part b, all the coefficients of other parameter decreased, which mean that the starting year had stronger associated with salary compared to year of degree.**

**The results in part c. and d. are the same, including the estimates and the confidence intervals and the p values. The intercepts in all of these parts are all no scientific relevance.**

Problems 3 – 5 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 graders, we will agree on modeling *yrdeg* and *startyr* as linear splines as computed in problem 1f.

1. 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.
   1. Report inference regarding the unadjusted comparison of women’s and men’s salaries.
   2. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree.
   3. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree.
   4. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW.
   5. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field.
   6. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties. Save the predicted values of the mean salary for each individual as *fit3.*
   7. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank.

**Table: Difference in the mean monthly salary paid to females compared to that paid to males**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **P Value** | **95% CI low** | **95% CI high** |
| **unadjusted** | -1334.7 | < .0001 | -1521.2 | -1148.3 |
| **adjustment for degree** | - 1266.2 | < .0001 | -1451.6 | -1080.8 |
| **adjustment for degree, year of degree** | -614.1 | < .0001 | -782.2 | -446.0 |
| **adjustment for degree, year of degree, starting year** | -614.6 | < .0001 | -614.6 | -443.8 |
| **adjustment for degree, year of degree, starting year, field** | -420.1 | < .0001 | -583.1 | -257.0 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties** | -419.7 | < .0001 | -579.0 | -260.5 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank** | -280.7 | < .0001 | -415.5 | -145.8 |

1. 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.
   1. Report inference regarding the unadjusted comparison of women’s and men’s salaries.
   2. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree.
   3. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree.
   4. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW.
   5. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field.
   6. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties. Save the predicted values of the geometric mean salary for each individual as *fit4.*
   7. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank.

**Table: Difference in the ratio of geometric mean monthly salary paid to females compared to that paid to males**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **P Value** | **95% CI low** | **95% CI high** |
| **unadjusted** | 0.812 | < .0001 | 0.796 | 0.845 |
| **adjustment for degree** | 0.820 | < .0001 | 0.885 | 0.934 |
| **adjustment for degree, year of degree** | 0.909 | < .0001 | 0.885 | 0.933 |
| **adjustment for degree, year of degree, starting year** | 0.909 | < .0001 | 0.913 | 0.960 |
| **adjustment for degree, year of degree, starting year, field** | 0.936 | < .0001 | 0.913 | 0.960 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties** | 0.936 | < .0001 | 0.938 | 0.978 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank** | 0.957 | < .0001 | 0.796 | 0.845 |

1. 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. You can use Poisson regression (with the irr option to get exponentiated parameter estimates), or you can use a generalized linear model with a log link. Stata has a regression function “glm” that allows the specification of a log link function. Hence, you can fit the regression for part a using the command

glm salary female if year==95, link(log) robust

Parameter estimates will be interpretable as the log mean (intercept) and log mean ratio (slope). (glm stands for “generalized linear model” and it includes as special cases linear regression, logistic regression, and Poisson regression. By default, it presumes the data are continuous and models the mean according to the value of the link function.) By specifying the “eform” option, it will return the exponentiated parameter estimates.

In either case, make clear which analysis method you used.

* 1. Report inference regarding the unadjusted comparison of women’s and men’s salaries.
  2. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree.
  3. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree.
  4. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW.
  5. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field.
  6. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties. Save the predicted values of the mean salary for each individual as *fit5.*
  7. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank.

**Table: Difference in the ratio of the mean monthly salary paid to females compared to that paid to males, using Poisson regression**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **P Value** | **95% CI low** | **95% CI high** |
| **unadjusted** | 0.811 | < .0001 | 0.785 | 0.837 |
| **adjustment for degree** | 0.811 | < .0001 | 0.785 | 0.837 |
| **adjustment for degree, year of degree** | 0.901 | < .0001 | 0.875 | 0.927 |
| **adjustment for degree, year of degree, starting year** | 0.900 | < .0001 | 0.875 | 0.927 |
| **adjustment for degree, year of degree, starting year, field** | 0.929 | < .0001 | 0.903 | 0.955 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties** | 0.929 | < .0001 | 0.904 | 0.954 |
| **adjustment for degree, year of degree, starting year at UW, field, administrative duties, rank** | 0.951 | < .0001 | 0.930 | 0.973 |

1. 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?

**The p-value of three models are all significant, and the estimates of three models also provide similar results (the difference in mean is about 0.934 in ratio scale, considering that the mean monthly salary for the entire sample was $6,389.81.) However, the model with the difference in the ratio of mean provides the different t statics, which means lower precision in this model.**

**The means of predicted values are close to each other in three model, with values around 6572-6735. The predicted value from the linear regression has the smallest mean, and the one from geometric mean has the largest mean. The standard deviations are also close to each (1177-1231), with a smallest value with a geometric mean model and largest with Poisson regression.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimates** | **T statics** | **P Value** | **95% CI low** | **95% CI high** |
| **Difference in the mean** | -419.7 | -5.17 | < .0001 | -579.0 | -260.5 |
| **Difference in the ratio of geometric mean** | 0.936 | -5.17 | < .0001 | 0.938 | 0.978 |
| **Difference in the ratio of the mean** | 0.929 | -5.34 | < .0001 | 0.904 | 0.954 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. Dev. | Min | Max | Range | 25th quartile | 75th quartile | IQR |
| fit3 | 19792 | 6735.1 | 1190.369 | 2868.8 | 10556.1 | 7687.3 | 5954.8 | 7438.6 | 1483.8 |
| expfit4 | 19792 | 6572.7 | 1177.087 | 3396.6 | 11380.5 | 7983.9 | 5822.6 | 7214.0 | 1391.4 |
| fit5 | 19792 | 6740.2 | 1231.21 | 3394.0 | 11732.0 | 8338.0 | 5942.8 | 7421.3 | 1478.5 |

**Figures for fit3, exp(fit4), and fit5 (in order)**



1. For the analysis model that you would have chosen *a priori*, summarize the scientific relevance of the single model that you think would best reflect any discrimination against women in awarding salaries. Give a formal report of your methods and results.

**Although the model of difference of mean and geometric mean provide the same precision, I would choose the model of robust linear regression for the difference in the mean monthly salary is simpler and easier to interpret. In the analysis, I would adjust the variables of adjustment for degree, year of degree, starting year at UW, field, and administrative duties because they all influenced how much a person would receive. I wouldn’t adjust rank, for it could be a mediator in the analysis.**

**Statistical Methods: The mean of monthly salary was compared between women and men using a linear regression which has adjustment for degree, year of degree, starting year at UW, field, and administrative duties. In this model, the variables of year of degree and starting year were modeled as linear splines with knots at years 1960, 1965, 1970, 1975, 1980, 1985, and 1990. Confidence intervals and two-sided p values were computed using Wald statistics based on the Huber-White sandwich estimator. No subjects were missing data for either variable.**

**Inferential results: Data was available on 1597 subjects, of whom 409 were women and 1188 were men. The mean of monthly salary is 419.7 lower for females compared to males with similar degree, year of degree, starting year at UW, field, administrative duties, and rank. (95% CI: 579.0 lower to 260.5 lower for females compared to males) These results are highly atypical of what we might expect with no true difference in the mean of monthly salary between females and males having similar degree, year of degree, starting year at UW, field, and administrative duties: P < 0.0005.**