**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2014

**Homework #8**

February 28, 2014

* 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.*

I would start modeling by using linear regression comparing log-transformed salary by sex as a binary variable and adjusting for degree, field, admin as unordered 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.*

Classic linear regression analysis assumes homoscedasticity. We cannot assume equal variances within each group and should therefore use robust standard error estimates. This inference will tend to be anti-conservative if we assume that within group variance will increase with sex. However, when we plot salary by sex, we can see that the salary variance within female professors is lower than the variance for male professors. Therefore, I suspect that this inference will be conservative as within group variance decreases with sex (when sex is defined as 1=female).

* 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).*

When we hold year of degree and start year constant, the mean monthly salary for women is $428.30 less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between $588.90 to $267.80 less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

* 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).*

When we hold year of degree and start year constant and model salary as a quadratic continuous variable, the mean monthly salary for women is $428.10 less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between $588.10 to $268.00 less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

* 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).*

When we hold year of degree and start year constant and model year of degree and start year as dummy variables as defined above, the mean monthly salary for women is $447.70 less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between $609.00 to $286.50 less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

* 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).*

When we hold year of degree and start year constant and model year of degree and start year as linear spline variables as defined above, the mean monthly salary for women is $419.70 less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between $579.00 to $260.50 less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

* 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.)*

*C*

When we hold year of degree and start year constant and model salary as a continuous linear variable using Poisson regression, the mean monthly salary for women is 7.34% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 10.31% to 4.89% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

When we hold year of degree and start year constant and model salary as a continuous linear variable, the geometric mean monthly salary for women is 6.53% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 8.87% to 4.13% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

D

When we hold year of degree and start year constant and model salary as a quadratic continuous variable using Poisson regression, the mean monthly salary for women is 7.20% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 9.7% to 4.63% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

When we hold year of degree and start year constant and model salary as a continuous quadratic variable, the geometric mean monthly salary for women is 6.48% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 8.81% to 4.10% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

E

When we hold year of degree and start year constant and model year of degree and start year as dummy variables as defined above using Poisson regression, the mean monthly salary for women is 7.56% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 10.06% to 5% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

When we hold year of degree and start year constant and model year of degree and start year as dummy variables as defined above, the geometric mean monthly salary for women is 6.72% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 9.04% to 4.34% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

F

When we hold year of degree and start year constant and model year of degree and start year as linear spline variables as defined above using Poisson regression, the mean monthly salary for women is 7.11% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 9.59% to 4.56% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

When we hold year of degree and start year constant and model year of degree and start year as linear spline variables as defined above, the geometric mean monthly salary for women is 6.34% less than their male counterparts. It would not be unusual for the difference in women faculty salaries to be between 8.68% to 4.00% less than their male colleagues. These results are highly atypical of what we might expect with no true differences between men and women of the same start year and degree year (p<0.001).

* 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?*

There was agreement of the inferences among all models, where women were consistently found to earn less than their male counterparts. Mean monthly salaries ranged between $419.70 to $447.70 less and the ratio of means were between 6.48% to 8.15% less for female faculty when compared to male faculty after adjusting for year of degree and start year. We know that the mean monthly salary in the sample was $6389.81, so that 6.48% to 8.15% is equal to a mean difference of $514.06 to $520.77.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **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?*

The Z statistic is helpful to assess significant differences in statistical significance of our models. If we suspect a possible multiplicative effect of covariates on salary, a ratio of means model would be appropriate. As our data consists of positive continuous random variables, both Poisson regression and Gaussian GLM will be appropriate, but would favor the Poisson regression as we are comparing risk ratios with binary sex data. As we are not sure if there is a nonlinear relationship, I would avoid using dummy variables in favor of linear splines. Linear splines improved the fit (the Z score moved closer to zero compared with dummy variables). Linear regression is probably appropriate as there was no clear evidence of consistently improved fit with a quadratic relationship between salary and gender (fit did not change significant, based on Z score), but this does not rule out some other nonlinear relationship.

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).*

Compared with faculty with the same degree, field, administrative duties, and sex, the mean monthly salary increases by $88.56 for each additional year that has passed since receiving his or her degree. It would not be unusual if the true mean monthly salary difference was between $80.25 to $96.87 more every year following graduation. These results are highly atypical of what we might expect with no true differences between professors of the same year of degree with the same degree, field, administrative duties, and sex (p<0.001).

* 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).*

Compared with faculty with the same degree, field, administrative duties, and sex, the mean monthly salary increases by $56.93 for each additional year he or she has worked at UW. It would not be unusual if the true mean monthly salary difference was between $47.68 to $66.19 more with each year of experience. These results are highly atypical of what we might expect with no true differences between professors of the same start year of the same degree, field, administrative duties, and sex (p<0.001).

* 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).*

Compared with faculty with the same degree, field, administrative duties, sex, and starting year, the mean monthly salary increases by $106.94 for each additional year that has passed since receiving his or her degree. It would not be unusual if the true mean monthly salary difference was between $89.01 to $124.87 more every year following graduation. These results are highly atypical of what we might expect with no true differences between professors of the same year of degree with the same degree, field, administrative duties, sex and starting year (p<0.001).

* 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).*

Compared with faculty with the same degree, field, administrative duties, sex, and year of degree, the mean monthly salary decreases by $23.17 for each additional year he or she has worked at UW. It would not be unusual if the true mean monthly salary difference was between $5.18 to $41.15 less with each year of experience. These results are atypical of what we might expect with no true differences between professors of the same start year of the same degree, field, administrative duties, sex, and year of degree (p=0.012).

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

In parts a and b, we see that the longer a professor has been employed at the UW or out of school, the more he or she makes. The Z scores are quite high and suggest strong statistical significance. In parts c and d, we observe a paradoxical effect that more recent employment is associated with a higher mean monthly salary of about $23 among professors of the same field, degree, administrative responsibilities, sex, and years since graduation. On the other hand, when comparing professors of the same field, degree, admin duties, sex, and who started in the same year, professors earn $106.94 more per year since he or she graduated. The Z scores are much smaller (closer to zero) with the additional adjustments in c and d, suggesting that these may not be improving the fit of the model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Z** | **P Value** | **95% CI low** | **95% CI high** |
| *Difference in Means* |
| **Year of degree** | -88.56 | -20.90 | < .0001 | -96.87 | -80.25 |
| **Start year** | -56.93 | -12.06 | < .0001 | -66.19 | -47.68 |
| **Year of degree\*** | -106.94 | -11.70 | < .0001 | -124.87 | -89.01 |
| **Start year\*\*** | 23.17 | 2.53 | 0 .012 | 5.18 | 41.15 |

\*adjusted for start year
\*\*adjusted for year of degree

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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Z** | **P Value** | **95% CI low** | **95% CI high** |
| *Difference in Means between female and male faculty* |
| **Female, unadjusted** | -1334.73 | -14.04 | < .0001 | -1521.18 | -1148.29 |
| **Female, adjusted for degree** | -1262.20 | -11.38 | < .0001 | -1479.74 | -1044.65 |
| **Female, adjusted for degree, year of degree** | -638.82 | -6.15 | < .0001 | -842.64 | -435.01 |
| **Female, adjusted for degree, year of degree, start year** | -650.22 | -6.28 | <0.001 | -853.21 | -447.22 |
| **Female, adjusted for degree, year of degree, start year, field** | -453.26 | -4.57 | <0.001 | -647.70 | -258.82 |
| **Female, adjusted for degree, year of degree, start year, field, admin** | -450.98 | -4.72 | <0.001 | -638.26 | -263.71 |
| **Female, adjusted for degree, year of degree, start year, field, admin, rank** | -318.07 | -3.65 | <0.001 | -489.22 | -146.91 |

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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Z** | **P Value** | **95% CI low** | **95% CI high** |
| *Geometric Means ratio* |
| **Female, unadjusted** | 0.812 | -12.56 | < .0001 | 0.786 | 0.839 |
| **Female, adjusted for degree** | 0.821 | -12.02 | < .0001 | 0.795 | 0.848 |
| **Female, adjusted for degree, year of degree** | 0.905 | -6.60 | < .0001 | 0.878 | 0.932 |
| **Female, adjusted for degree, year of degree, start year** | 0.904 | -6.68 | <0.001 | 0.877 | 0.931 |
| **Female, adjusted for degree, year of degree, start year, field** | 0.932 | -4.92 | <0.001 | 0.906 | 0.958 |
| **Female, adjusted for degree, year of degree, start year, field, admin** | 0.932 | -5.08 | <0.001 | 0.907 | 0.958 |
| **Female, adjusted for degree, year of degree, start year, field, admin, rank** | 0.953 | -3.93 | <0.001 | 0.931 | 0.976 |

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Estimate** | **Z** | **P Value** | **95% CI low** | **95% CI high** |
| *Ratio of Means (Poisson)* |
| **Female, unadjusted** | 0.802 | -13.58 | < .0001 | 0.777 | 0.828 |
| **Female, adjusted for degree** | 0.811 | -12.94 | < .0001 | 0.785 | 0.837 |
| **Female, adjusted for degree, year of degree** | 0.895 | -7.39 | < .0001 | 0.869 | 0.922 |
| **Female, adjusted for degree, year of degree, start year** | 0.893 | -7.45 | <0.001 | 0.867 | 0.920 |
| **Female, adjusted for degree, year of degree, start year, field** | 0.922 | -5.57 | <0.001 | 0.896 | 0.949 |
| **Female, adjusted for degree, year of degree, start year, field, admin** | 0.923 | -5.56 | <0.001 | 0.898 | 0.949 |
| **Female, adjusted for degree, year of degree, start year, field, admin, rank** | 0.946 | -4.65 | <0.001 | 0.925 | 0.969 |

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?*

All three analyses share general trends that the estimated effect of female sex on monthly salaries decreases as we adjust for more variables. Also, the confidence intervals become narrower, particularly in the Poisson and geometric mean ratio analyses as we adjust for more variables. Likewise, the RMSE decreases with more covariates in the linear and geometric mean models, suggesting an increase in precision of the effect of female sex. Also, the addition of adjusting for start year after already adjusting for year of degree did not seem to add much precision, and this was consistent across analyses. Similarly, adding administrative duties after already adjusting for field did not add much precision and the estimates were largely unchanged.

The first analysis estimates the mean monthly salary while the other two analyses estimate the mean ratio or geometric mean ratio of female to male salaries. The Poisson and geometric mean ratio estimates are quite consistent (as illustrated in the table below, and women’s salaries range from 80-81% up to 95% of male salaries. This is a larger estimated contribution of sex on salary disparities than estimated in question and accounts to nearly a $900 pay difference between the sexes. This is lower than the unadjusted difference in the linear regression model from question 3 of almost $1400, but higher than the estimated $400 difference in question 1.



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.*

METHODS:
The estimated monthly salary mean ratio in 1995 was compared between male and female faculty members using a Poisson regression model using sex as a binary variable. We adjusted for the type of degree, professional field, rank (assistant, associate, or full professor), and year their degrees were awarded which were modeled as linear splines in 5 year intervals between 1960 to 1990. Statistical inference was based on the Wald statistic computed from the Huber-White sandwich estimator from the regression slope parameter and its robust standard error with a two-sided p value and 95% confidence interval computed assuming normal distribution for the Poisson regression parameter estimates.

RESULTS:
Of the 1,597 faculty members, 25.61% or 409 were women. Most faculty worked in “other” fields (66.81%) with 13.78% employed in the school of arts and sciences and 19.41% employed by professional schools. The majority of faculty were full professors, 845 or 52.91%, 27.36% were associate professors, and 19.72% were assistant professors. Approximately 85.1% or 719 of all full professors were male. This contrasts with the distribution among assistant professors by gender, where there was a roughly equal distribution with 53.1% of assistant professors were male and 46.9% were female.

The monthly salary mean ratio between men and women was 19.8% lower for female faculty at baseline, before adjusting for other variables. It would not be unusual for the mean ratio to be between 22.3% to 17.2% lower for female faculty, and we reject the null hypothesis that there is no difference in the mean salary ratio based on sex (p<0.001). When comparing faculty members with the same degree, working in the same field, of the same rank, and who were awarded their degrees in the same year, the monthly female faculty salary was 5.22% lower than it was for male faculty. We say this with a high degree of statistical confidence (p<0.001) and estimate that it would not be unusual if the true mean ratio was between 7.36% to 2.84% less for women than it is for men.