Biost 518: Applied Biostatistics II
Biost 515: Biostatistics II
Emerson, Winter 2014

Homework #8 Key
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.)***

   a. 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.

<u>Ans:</u> (5 pts) Degree and field are both unordered categorical variables, so they should be modeled using dummy variables. Administrative duties is a binary variable, so it can be modeled as is, which is equivalent to using a dummy variable.

b. 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.

<u>Ans:</u> (5 pts) Owing to our tendency to grant raises as a percentage, we would definitely expect heteroscedasticity when modeling differences in means. We would expect to see greater variability in groups tending toward the highest salaries, and owing to historical (at least) trends in hiring and paying women we expect there to be fewer women and for them to be paid less. Hence, when comparing women to "otherwise similar" men, we expect the group with the generally smaller sample size (women) to have lower salaries and, hence, less variable salaries. This would suggest that inference about difference in means from classical linear regression would tend to be conservative. When analyzing ratios of geometric means, there is the possibility that the log transform would correct for the multiplicative raises, etc., but we would not necessarily expect that such a log transform would correct for heteroscedasticity induced by discriminatory practices. So, again, we would expect conservative inference.

(When using inference based on ratios of means, the picture is a bit more complicated because the estimating equations are a bit different. Certainly when using Poisson regression, we need to use robust standard errors, because otherwise the assumption is that the variance is equal to the mean. When using the GLM specifying Gaussian errors with a log link, we would expect the same general tendencies as with classical linear regression, because the data is still heteroscedastic. However, when using the log link, the contributions from the individual subjects will be weighted differently.)

c. 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).

<u>Ans:</u> (0 pts) (See table below)

d. 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).

<u>Ans:</u> (0 pts) (See table below)

e. 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).

<u>Ans:</u> (0 pts) (See table below)

f. 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).

<u>Ans:</u> (0 pts) (See table below)

g. 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.)

<u>Ans:</u> (0 pts) (See table below)

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

<u>Ans:</u> (0 pts) 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:

- **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.**

Table 1: Inference for problems 1c-g.

	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

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

<u>Ans:</u> **(5 pts) We would want to make this decision before fitting the data. To do otherwise might overfit the data and lead to a spuriously low standard error.**

Year of degree and starting year are confounders in our analysis. We are not interested in making inference about these variables, and we would likely not even provide estimates of their associations with salaries. Hence, *a priori* we would like to model those variables in a way that would allow some degree of nonlinearity if we are fearful that their associations with salary are not linear on the appropriate scale. Given the relatively large dataset, we can afford to use linear splines with several knots. Placement of the knots at points that we expect substantial changes in the shape of the curve would be best (e.g., knots surrounding the period of very high inflation during the late 1970's and early 1980's), but lacking detailed knowledge we could just space them evenly. Alternatively we could consider a high order polynomial (say cubic or quartic).

In any case, we should avoid putting too many knots (or polynomial terms) to avoid losing precision (somewhat important) and to avoid possibly adjusting for variables such as rank that for scientific reasons we do not want to adjust for. (Denying promotion is one way of discriminating against women, and there are complicated relationships between year of degree, starting year, rank, and salary. The more terms we put in for year of degree and starting year will allow those terms to model rank somewhat.)

Dummy variables would not be as good an idea, because we do expect a continuous curve rather than a step function. For the same number of covariates added to the model, linear splines would generally provide much greater precision when fitting a continuous nonlinear function.

2. 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.

Table 2: Inference for problems 2a-d.

	Year of Degree			Starting year		
	Estimate	**95% CI**	**P (two-sided)**	**Estimate**	**95% CI**	**P (two-sided)**
Problem 2a	-89.9	(-98.3, -81.4)	<0.0005			
Problem 2b				-56.9	(-66.1, -47.6)	<0.0005
Problem 2c-d	-112.	(-130.6, -93.3)	<0.0005	27.2	(8.7, 45.6)	0.004

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).

<u>Ans:</u> (5 pts – I provide full inference below, but it was acceptable for this part to only provide the estimates in the table above)

The data were analyzed using a linear regression model of monthly salaries on year of degree as a continuous variable with adjustment for dummy variables modeling degree, field, administrative duties, and sex. Inference included 95% confidence intervals (CI) and two-sided P values computed using the robust standard errors derived using the Huber-White sandwich estimator. We find a highly statistically significant trend toward lower salaries for faculty members who received their highest degree most recently (P < 0.0005). We estimate that the average monthly salary is $89.90 lower for every additional year difference in the year of highest degree when comparing groups having similar degrees, fields, administrative duties, and sex. Based on the 95% CI, the observed results are not unusual when the true trend would be for average monthly salaries that were anywhere between $81.40 to $98.30 lower for each year difference in year of degree.

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).

<u>Ans:</u> (5 pts – I provide full inference below, but it was acceptable for this part to only provide the estimates in the table above)

The data were analyzed using a linear regression model of monthly salaries on starting year as a continuous variable with adjustment for dummy variables modeling degree, field, administrative duties, and sex. Inference included 95% confidence intervals (CI) and two-sided P values computed using the robust standard errors derived using the Huber-White sandwich estimator. We find a highly statistically significant trend toward lower salaries for faculty members who started working at the university most recently (P < 0.0005). We estimate that the average monthly salary is $56.90 lower for every additional year difference in the starting when comparing groups having similar degrees, fields, administrative duties, and sex. Based on the 95% CI, the observed results are not unusual when the true trend would be for average monthly salaries that were anywhere between $47.60 to $66.10 lower for each year difference in starting year.

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).

<u>Ans:</u> (5 pts – I provide full inference below, but it was acceptable for this part to only provide the estimates in the table above)

The data were analyzed using a linear regression model of monthly salaries on year of degree as a continuous variable with adjustment for starting year modeled continuously and dummy variables modeling degree, field, administrative duties, and sex. Inference included 95% confidence intervals (CI) and two-sided P values computed using the robust standard errors derived using the Huber-White sandwich estimator. We find a highly statistically significant trend toward lower salaries for faculty

members who received their highest degree most recently (P < 0.0005). We estimate that the average monthly salary is $112 lower for every additional year difference in the year of highest degree when comparing a groups having similar starting years, degrees, fields, administrative duties, and sex. Based on the 95% CI, the observed results are not unusual when the true trend would be for average monthly salaries that were anywhere between $93 to $131 lower for each year difference in year of degree.

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).

<u>Ans:</u> (5 pts – I provide full inference below, but it was acceptable for this part to only provide the estimates in the table above)

The data were analyzed using a linear regression model of monthly salaries on starting year as a continuous variable with adjustment for year of degree modeled continuously and dummy variables modeling degree, field, administrative duties, and sex. Inference included 95% confidence intervals (CI) and two-sided P values computed using the robust standard errors derived using the Huber-White sandwich estimator. We find a highly statistically significant trend toward higher salaries for faculty members who started working at the university most recently (P = 0.004) compared to otherwise similar faculty. We estimate that the average monthly salary is $27.20 higher for every additional year difference in the starting when comparing groups having similar degrees, year of obtaining that degree, fields, administrative duties, and sex. Based on the 95% CI, the observed results are not unusual when the true trend would be for average monthly salaries that were anywhere between $8.70 to $45.60 higher for each year difference in starting year.

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

<u>Ans:</u> (5 pts) Year of degree and starting year are highly correlated in this dataset (r = 0.79). Year of degree is a strong surrogate for years of experience, which should be highly predictive of salary with greater experience (earlier year of degree). Starting year measures experience at this university, which is probably not quite as important a predictor as would be total experience. However, when starting year is modeled without year of degree, the high correlation between those variables will mean that starting year will tend to model the effect of total experience on salary and thus similarly show a tendency to lower salaries with later start years.

When both year of degree and starting year are modeled, the scientific meaning of the variables changes. The year of degree coefficient will reflect the strong association between salaries and total experience. After adjustment for year of degree, however, the coefficient for starting year will reflect salary differential between people hired soon after getting their degree (as is the case with most faculty) and people hired later (perhaps when recruiting department chairs or highly productive researchers). Hence, the change in sign of the coefficient adjusted for year of degree.

An alternative explanation based on technical statistical behavior is also possible, however. Owing to the high correlation between year of degree and starting year, it is possible that modeling the two together is allowing a more flexible model of a relationship

that is truly only related to year of degree. *(When using linear splines with several knots, however, this latter explanation is less likely to be a problem.)*

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.

3. 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.

<u>Ans:</u> (20 pts total for having the estimates, CI, and either Z statistics or comparable P values (all are < 0.0001) in the table shown below)

   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 at UW.

   e. Report inference regarding the comparison of women’s and men’s salaries after adjustment for degree, year of degree, starting year at UW, field.

   f. 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*.

   g. 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.

4. 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.

<u>Ans:</u> (20 pts total for having the estimates, CI, and either Z statistics or comparable P values (all are < 0.0001) in the table shown below. The ratio of geometric means must be provided. It is not acceptable to have provided only the log geometric mean ratios.

Note that Stata provides slightly different robust SE estimates when using `glm` than when using `regress`. Hence, students who used the GLM command might have CI that differ in the third significant digit (or so).)

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 at UW.

e. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year at UW, field.

f. 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.*

g. 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.

5. 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.

<u>Ans:</u> (20 pts total for having the estimates, CI, and either Z statistics or comparable P values (all are < 0.0001) in the table shown below. Note that I provided inference based on both the Poisson and the GLM regression models. Either is acceptable <u>providing</u> the student made clear which was used.)

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 at UW.

e. Report inference regarding the comparison of women's and men's salaries after adjustment for degree, year of degree, starting year at UW, field.

f. 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*.

g. 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 3: Inference for problems 3, 4 and 5.

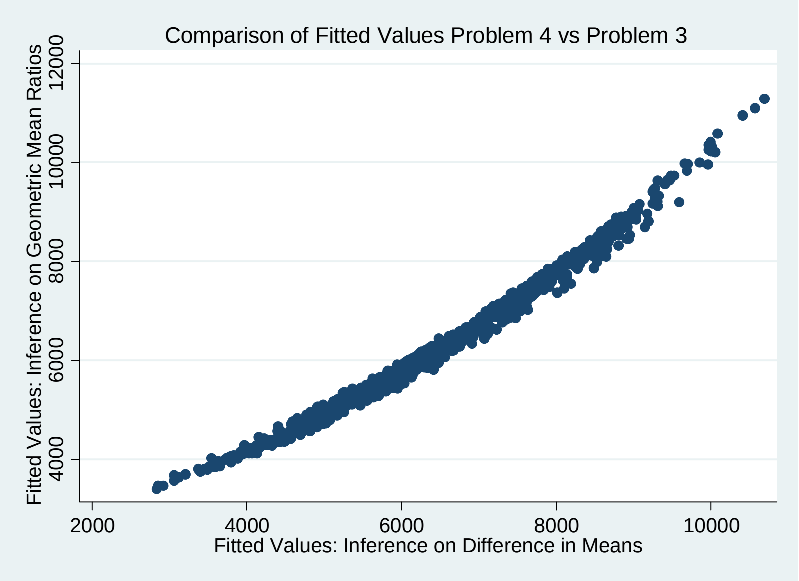
	Comparison of Females' Salaries to Males' Salaries		
	Estimate	**95% CI**	**Z**
Problem 3: Difference in Means			
Unadjusted	-1,335	(-1,521, -1,148)	-14.04
Adjusted for Degree	-1,266	(-1,452, -1,081)	-13.40
+ Year of degree	-614	(-782, -446)	-7.17
+ Starting year	-615	(-785, -444)	-7.06
+ Field	-420	(-583, -257)	-5.05
+ Admin duties	-420	(-579, -260)	-5.17
+ Rank	-281	(-416, -146)	-4.08
Problem 4: Ratio of Geometric Means			
Unadjusted	0.812	(0.788, 0.837)	-13.73
Adjusted for Degree	0.820	(0.796, 0.845)	-13.09
+ Year of degree	0.909	(0.885, 0.934)	-6.99
+ Starting year	0.909	(0.885, 0.933)	-6.98
+ Field	0.936	(0.913, 0.960)	-5.06
+ Admin duties	0.936	(0.913, 0.960)	-5.17
+ Rank	0.957	(0.938, 0.978)	-4.08
Problem 5: Ratio of Means (Poisson regression)			
Unadjusted	0.802	(0.777, 0.828)	-13.58
Adjusted for Degree	0.811	(0.785, 0.837)	-12.98
+ Year of degree	0.901	(0.875, 0.927)	-7.09
+ Starting year	0.901	(0.875, 0.928)	-7.01
+ Field	0.929	(0.903, 0.955)	-5.22
+ Admin duties	0.929	(0.904, 0.954)	-5.34
+ Rank	0.951	(0.930, 0.973)	-4.30
Problem 5: Ratio of Means (GLM – Gaussian with log link)			
Unadjusted	0.802	(0.777, 0.828)	-13.58
Adjusted for Degree	0.810	(0.784, 0.836)	-12.99
+ Year of degree	0.898	(0.872, 0.925)	-7.12
+ Starting year	0.896	(0.869, 0.924)	-7.04
+ Field	0.925	(0.899, 0.952)	-5.26
+ Admin duties	0.924	(0.899, 0.951)	-5.49
+ Rank	0.951	(0.928, 0.974)	-4.15

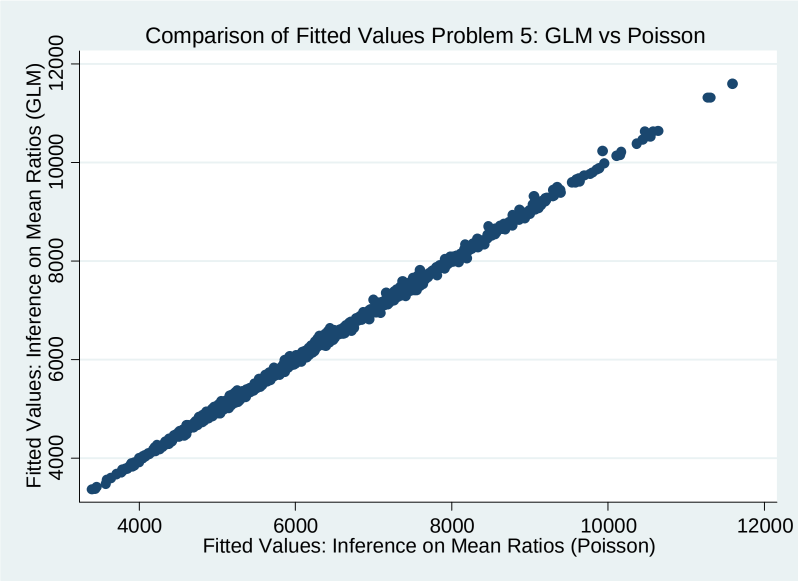
6. Briefly discuss the similarities and differences between the analyses performed in problems 3 – 5. How similar are the predicted values between the models? How different is the inference you would obtain?

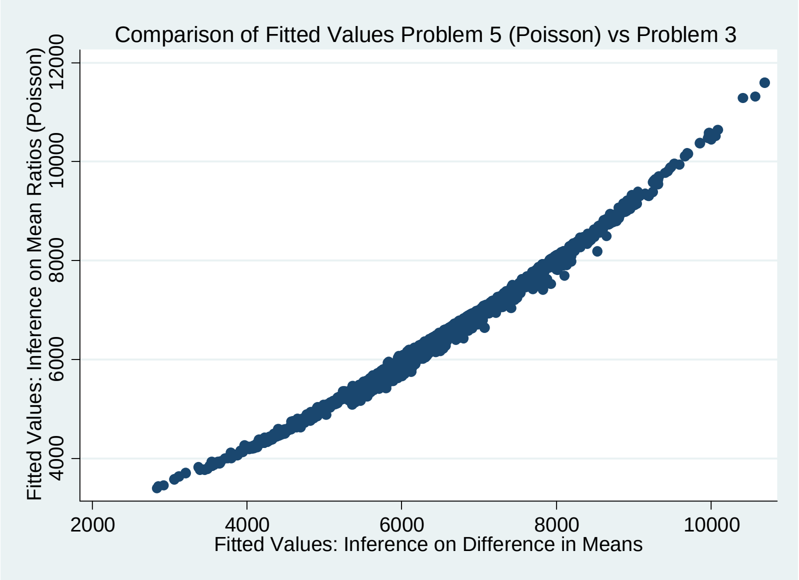
<u>Ans:</u> (10 pts) As seen in Table 3, the estimated ratios of geometric means and ratios of means agree to within 0.01 across all analyses, with quite similar inference as measured by Z statistics (the P values are too small to make for useful display). When considering the difference in means, we can consider those difference relative to the overall mean monthly salary of $6,389.81. We then find that the difference in estimated means corresponds closely to the ratios. For instance, using the model adjusting for all variables, the estimate difference in monthly salaries between the sexes is $281, which is 4.4% of $6,389.81. This is comparable to the percentage difference in geometric mean (4.3%) or percentage difference in means when computed using a log link (4.9%).

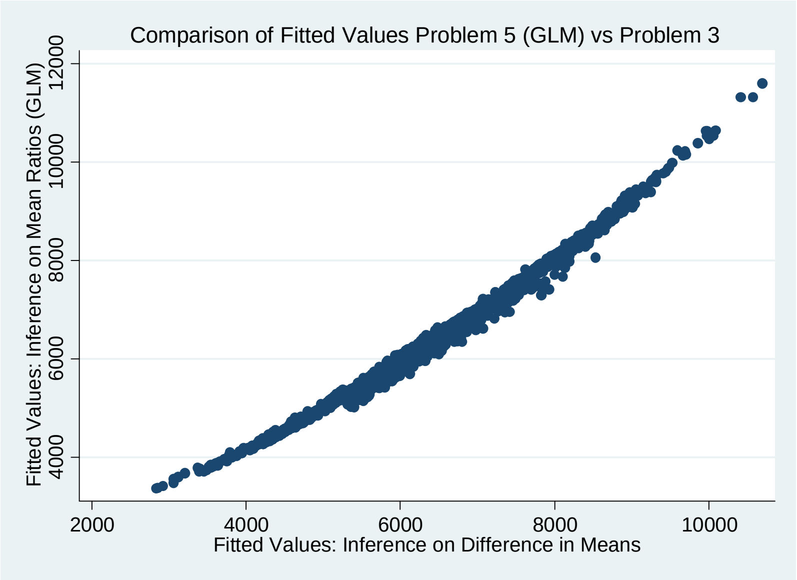
This close agreement is not unexpected given that all salaries are fairly distant from 0 and do not vary over too many orders of magnitude (doublings). The figures given below compare the fitted values from the various models. They do not agree perfectly, but certainly provide similar estimates.

(The takehome message is that it sometimes does not make too much difference. Certainly in this problem we could have predicted that there would be little difference in our conclusion. So that would argue that we do not go wrong just choosing a model a priori and living with that choice. The possible danger of searching through multiple models far outweighs any possible benefits in added precision you might get by fishing for the best model.)









7. 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.

Ans: ***(10 pts – No more than 0 points should be awarded if the chosen model included rank, because promotion is a likely mediator of discrimination. No more than 5 points should be awarded if the chosen model did not include degree and year of degree, because these reflect historical practices that are not of as much interest in assessing current discriminatory practices. It is acceptable to me to choose difference in means, ratios of geometric means, or ratios of means, so long as the student describes the methods and results correctly.)***

Ratios of geometric mean salaries were compared between women and men using a linear regression model of log transformed monthly salaries on an indicator of female sex with adjustment for highest degree obtained (PhD, professional, other), year of that degree, year starting at the university, field (professional, fine arts, other), and an indicator of having administrative duties. Field and degree were modeled using dummy variables, while year of degree and starting year were modeled using linear splines with knots every 5 years between 1960 and 1990). Inference included 95% confidence intervals (CI) and two-sided P values computed assuming an approximate normal distribution for regression coefficients and using the robust standard errors derived using the Huber-White sandwich estimator.

We find a highly statistically significant trend P < 0.0001) toward lower geometric mean salaries for female faculty members compared to males having similar degrees, year of degree, starting year, field, and administrative duties. We estimate that the geometric mean monthly salary is 6.4% lower for females compared to otherwise similar males. Based on the 95% CI, the observed results are not unusual when the true trend would be for geometric mean monthly salaries that were anywhere between 4.0% to 8.7% lower for females compared to males.