Biostats 518 Homework #8

* 1. I think it is best to model degree, field, and admin in a categorical manner. We would need to use dummy variables in our model. This is not ideal, but we need to do this if we want to make categorical variables work in regression.
	2. Classical linear regression assumes that the variables are homoscedastic (the variances are equal). This isn’t necessarily true. Since in this dataset, the variables are much more likely heteroscedastic, it is important to create a model that reflects that. Whether or not classical linear regression would be more or less conservative depends on the data. If the data has a smaller variance than normal, the linear model could be anticonservative. However, if the data has a larger variance than normal, the linear model could be conservative.
	3. Scott answered already
	4. Scott answered already
	5. Scott answered already
	6. Scott answered already
	7. Scott answered already
	8. Scott answered already
	9. I would choose between alternative methods for adjusting for year of degree and starting year by creating a table of descriptive statistics and analyzing whether there is a relationship between those variables and the outcome. I would also take the sample size, nature of the data (numerical, categorical, etc), and the scientific question at hand into account.
	10. Salary = 6737.639 – 27.955(yrdeg)
		1. According to this model, the earlier the degree was earned, the higher the salary. The 95% CI for the coefficient for degree earned is between -30.938 and -24.973 and the standard error is 1.521. The 95% CI for the intercept is between 6512.882 and 6962.395 and the standard error is 114.667.
	11. Salary = 4027.549 + 9.126(startyr)
		1. The later a person starts their position, the higher their salary. The 95% CI for the coefficient for starting year is between 5.990 and 12.262 and the standard error is 1.599. The 95% CI for the intercept is between 3784.431 and 4270.667 and the standard error is 124.034.
	12. Salary = -17973.96 + 234.402(year)+ 795.866(rank)+ 586.9466(field)+ 447.1503(degree)-275.7143(female)+ 788.7137(admin)+ 26.38614(startyr)-51.64041(yrdeg)
		1. It looks like the factors that make the most impact on salary are (in order): rank, admin, field, degree, sex, year, the year the degree was earned, and the current year. It seems like females tend to make $275.71 dollars less and the 95% CI is between $243.810 and $307.618.
	13. Same as (c ) above
	14. In the above questions, we systematically approached the idea of sex-based discrimination for salaries. We first investigated salary as a function of highest degree earned, then as a function of highest degree earned and starting year, etc. We find that each factor we add effects monthly salary, but that being female tends to drop monthly salary by about $275. We see that sex does make a difference in monthly salary, but we aren’t sure if it makes a significant difference.
	15. There seems to be a significant difference between the salaries of men and women. Women seem to make $419.7 less than men and the 95% CI is between $260 lower and $570 lower. The p-value is less than 0.0001.
	16. Highest degree earned seems to make a major difference. It amounts to roughly $447 with a 95% CI between $411 and $482. Although it is close in value to the point estimate that sex makes, the confidence interval has much tighter bounds to we are more certain of the effect degree has over sex.
	17. Contrary to the highest degree earned making a difference, the year the degree was earned seems to make little difference. Older degrees tend to make $51 less in salary with a 95% CI between $56 and $47.
	18. In addition to the comments above, start year seems to make little difference as a function of salary as well. It is estimated that those with later start years make $26 less with a 95% CI between $22 and $30.
	19. Field seems to make a large difference in the salary earned. The point estimate is $586 with a 95% CI between $563 and $610. This is a much larger difference than anything we’ve seen thus far, including sex.
	20. Those who have administrative duties also seem to make more money than those who don’t. The point estimate is $788 more with a 95% CI between $731 and $845. This is now the largest difference maker.
	21. Rank also seems to make a major difference in salary. In fact, it is the largest coefficient in the model with a point estimate of $795 with a 95% CI between $768 and $823.
	22. There seems to be a significant difference between the salaries of men and women. Women seem to make 7% less than men and the 95% CI is between 4% lower and 9% lower. The p-value is less than 0.0001.
	23. Highest degree earned seems to make a major difference. It amounts to roughly 0.0917 with a 95% CI between 0.0847539 and 0.0983866. Although it is close in value to the point estimate that sex makes, the confidence interval has much tighter bounds to we are more certain of the effect degree has over sex.
	24. Contrary to the highest degree earned making a difference, the year the degree was earned seems to make little difference. Older degrees tend to make -.0082937 in salary with a 95% CI between -0.0090343 -0.0075531.
	25. In addition to the comments above, start year seems to make little difference as a function of salary as well. It is estimated that those with later start years make .0049719 less with a 95% CI between .0042876 and .0056561.
	26. Field seems to make a large difference in the salary earned. The point estimate is .1221737 with a 95% CI between .1175455 and .1268018. This is a much larger difference than anything we’ve seen thus far, including sex.
	27. Those who have administrative duties also seem to make more money than those who don’t. The point estimate is .141953 more with a 95% CI between .1329698 and .1509363. This is now the largest difference maker.
	28. Rank also seems to make a major difference in salary. In fact, it is the largest coefficient in the model with a point estimate of .2001514 with a 95% CI between .1947298 and .205573.
1. I used a generalized linear model with log link for this question
	1. There seems to be a significant difference between the salaries of men and women. Women seem to make $318 less than men and the 95% CI is between $457.6063 lower and $178.5282 lower. The p-value is less than 0.0001.
	2. Highest degree earned seems to make a major difference. It amounts to roughly $464 with a 95% CI between $294.3992 lower and $633.6668 lower. The range in the 95% CI is very large.
	3. Contrary to the highest degree earned making a difference, the year the degree was earned seems to make little difference. Older degrees tend to make $56 less in salary with a 95% CI between $72 lower and $39 lower.
	4. In addition to the comments above, start year seems to make little difference as a function of salary as well. It is estimated that those with later start years make $38 less with a 95% CI between $21 and $54.
	5. Field seems to make a large difference in the salary earned. The point estimate is $877 with a 95% CI between $767 and $987.
	6. Those who have administrative duties also seem to make more money than those who don’t. The point estimate is $ 1165.607 more with a 95% CI between $911 and $1419.
	7. Rank also seems to make a major difference in salary. In fact, it is the largest coefficient in the model with a point estimate of $ 1184.08 with a 95% CI between $1064 and $1303.
2. The models may have been different between problems 1-3, but the conclusions I reached were the same. However, the GLM model had different results than the linear regression or poisson model, which were essentially identical (with the poisson model being ratios instead of absolute numbers). The inferences I obtained were identical in all models, but this could have been different if the conclusion was not as clean cut.
	1. Statistical method: We will create a poisson regression model to compare the relative risk ratio of salaries between men and women. The model will be adjusted for highest degree earned, year highest degree was attained, current year, field of study, rank, and whether or not administrative duties are part of a professor’s responsibilities. A 95% CI will be generated for each variable in the model.
	2. Inference: Based on the model, women do not seem to be discriminated against for salaries. Rank and administrative duties seem to make the biggest difference in salaries, followed by field of study and highest degree earned. Sex does make a difference, but it accounts for less than most other variables. In context, it seems as though the impact of sex is not significant compared to the impact of other factors. At the end of the day, rather than investigating whether or not there is a significant difference in starting salary between men and women, an interesting question would be to investigate whether or not there is a difference in raises between men and women.