**Biost 518: Applied Biostatistics II**

**Biost 515: Biostatistics II**

Emerson, Winter 2014

**Homework #8**

March 7, 2014

* 1. The best way to include degree, field, and admin in the model is as dummy variables. This is because these are unordered categorical variables and we would like to model these as separate groups.
	2. We are modeling the mean salary for several groups. Classical linear regression requires the assumption that the variance in salary is the same for all groups. Since we do not know if the variance is equal across groups it would not be safe to make this assumption. Regression using robust standard errors is more likely to be conservative compared to classical linear regression, since it would increase the width of confidence intervals. However, depending on the direction of the data the robust standard error could also be anti-conservative.
	3. Linear regression would be the best method to use for adjusting for year of degree and starting year. All other regression methods require us to categorize the continuous variable start year, which will result in a loss of information. Also, linear regression is the least complicated model in terms of interpretability and answering our scientific question. Unless there was a strong indication the relationship between start year and year of degree was not linear, I would choose the method that is less complicated and easier to interpret.
1. Robust linear regression was used to estimate the difference in mean salary between professors who had different start years of their degree, and who earned their degree in different years. Both of these estimates were yielded after adjusting for degree type, field of study, administrative duties and gender. Corresponding 95% confidence intervals were also calculated. In all models the null hypothesis that there was no difference in salary between professors was tested using the t-test.

	1. Professors with a one year difference in the year of degree earn $89.86 lower compared to professors who earned their degree one year later with the same degree, field of study, administrative duties and gender. It would not be unusual to observe a difference in salary between $98.30 and $81.43 lower for professors with a one year difference in year of degree, with the professors having new degrees earning more. The test of the null hypothesis that professors with a one year difference in degree have the same salary yielded a p-value of <0.0001. Therefore, we reject the null hypothesis and find that professors with a more recent degree tend to have a higher salary compared to professors with a later degree, given they have the same degree, field of study, administrative duties, and gender.
	2. Professors with a one year difference in the starting year of their degree earn $56.88 lower compared to professors who started their degree one later with the same degree, field of study, administrative duties and gender. It would not be unusual to observe a difference in salary between $66.13 to $47.63 lower for professors with a one year difference in start year of degree, with the professors starting later earning more. The test of the null hypothesis that professors with a one year difference in start year of degree have the same salary yielded a p-value of <0.0001. Therefore, we reject the null hypothesis and find that professors with a more recent start year of degree tend to have a higher salary compared to professors with a later start year of degree, given they have the same degree, field of study, administrative duties, and gender
	3. Professors with a one year difference in the year of degree earn $111.96 lower compared to professors who earned their degree one year later with the same degree, start year of degree, field of study, administrative duties and gender. It would not be unusual to observe a difference in salary between $130.58 to $93.34 lower for professors with a one year difference in year of degree, with the professors having new degrees earning more. The test of the null hypothesis that professors with a one year difference in degree have the same salary yielded a p-value of <0.0001. Therefore, we reject the null hypothesis and find that professors with a more recent degree tend to have a higher salary compared to professors with a later degree, given they have the same degree, start year of degree, field of study, administrative duties, and gender.
	4. Professors with a one year difference in the starting year of their degree earn $27.15 higher compared to professors who started their degree one later with the same degree, year of degree, field of study, administrative duties and gender. It would not be unusual to observe a difference in salary between $8.68 to $45.63 higher for professors with a one year difference in start year of degree, with the professors starting later earning more. The test of the null hypothesis that professors with a one year difference in start year of degree have the same salary yielded a p-value of 0.0004. Therefore, we reject the null hypothesis and find that professors with a more recent start year of degree tend to have a lower salary compared to professors with a later start year of degree, given they have the same degree, year of degree field of study, administrative duties, and gender.
	5. Both models (a) and (c) estimate the difference in mean salary between professors with a one year difference in the year they earned. However, the model in (c) holds more covariates fixed so the estimate in mean difference may be more precise, since we are comparing a group of professors that are more similar. The same is true for models (b) and (d) except the difference in mean salary between professors with a one year difference in start year of degree is estimated. Models (a) and (b) are nested models of (c) and (d).
2. Robust linear regression was used to compare the difference in mean salary between male and female professors in 1995. Degree type, year of degree, starting year at UW, field, administrative duties, and rank were all included in the model to adjust for the estimate in mean difference. These covariates were added one by one, creating nested models. Along with the mean estimate, robust 95% confidence intervals were also computed. The null hypothesis, that the salary between male and female professors are equal after controlling for the variables listed above, was tested with a two-sided, t-test. Corresponding values were also reported. The difference in salary is comparing females to males (i.e females have lower mean salary.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covariates** | **Diff. in Salary**  | **t** | **p-value** | **95% CI low** | **95% CI high** |
| Unadjusted | -1335 | -14.04 | < .001 | -1521 | -1148 |
| Degree | -1266 | -13.40 | < .001 | -1452 | -1081 |
| Degree, Year of Degree | -614.1 | -7.17 | < .001 | -782.2 | -446.0 |
| Degree, Yr Degree, Start Year | -614.6 | -7.06 | < .001 | -785.3 | -443.8 |
| Degree, Yr Degree, Str Yr, Field | -420.1 | -5.05 | < .001 | -583.1 | -257.0 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties | -419.7 | -5.17 | < .001 | -580.0 | -260.5 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties, Rank | -280.7 | -4.08 | < .001 | -415.5 | -145.8 |

1. Robust linear regression was used to compare the ratio in geometric mean salary between male and female professors in 1995. The log of the data was taken to compute this and the estimates from linear regression were exponentiated to calculate geometric mean salary ratio comparing men and women. Degree type, year of degree, starting year at UW, field, administrative duties, and rank were all included in the model to adjust for the estimate in ratio of means. These covariates were added one by one, creating nested models. Along with the ratio of geometric mean estimate, robust 95% confidence intervals were also computed. The null hypothesis, that the mean salary between male and female professors are equal (ratio of means is 1) after controlling for the variables listed above, was tested with a two-sided, t-test. Corresponding p-values were also reported.

Exponentiating the terms below will give the estimate of the ratio of geometric means comparing females to males and the corresponding confidence intervals.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covariates** | **Log of****Ratio of Geometric Mean** | **t** | **p-value** | **95% CI low** | **95% CI high** |
| Unadjusted | -0.2082 | -13.73 | < .001 | -0.2380 | -0.1785 |
| Degree | -0.1980 | -13.09 | < .001 | -0.2277 | -0.1683 |
| Degree, Year of Degree | -0.0954 | -6.99 | < .001 | -0.1222 | -0.0687 |
| Degree, Yr Degree, Start Year | -0.0958 | -6.98 | < .001 | -0.1227 | -0.6884 |
| Degree, Yr Degree, Str Yr, Field | -0.0659 | -5.06 | < .001 | -0.0914 | -0.0403 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties | -0.0658 | -5.17 | < .001 | -0.0908 | -0.0409 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties, Rank | -0.0435 | -4.08 | < .001 | -0.0644 | -0.0226 |

1. Robust Poisson regression was used to estimate the ratio of mean salary of women compare to men in 1995. Degree type, year of degree, starting year at UW, field, administrative duties, and rank were all included in the model to adjust for the estimate in ratio of means. These covariates were added one by one, creating nested models. Along with the ratio of mean estimate, robust 95% confidence intervals were also computed. The null hypothesis, that the mean salary between male and female professors are equal (ratio of means is 1) after controlling for the variables listed above, was tested with a two-sided, t-test. Corresponding p-values were also reported.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Adjustment** | **Ratio of Mean Estimate** | **Z** | **p-value** | **95% CI low** | **95% CI high** |
| Unadjusted | 0.8017 | -13.58 | < .001 | 0.7765 | 0.8277 |
| Degree | 0.8097 | -12.99 | < .001 | 0.7844 | 0.8359 |
| Degree, Year of Degree | 0.8981 | -7.12 | < .001 | 0.8719 | 0.9251 |
| Degree, Yr Degree, Start Year | 0.8964 | -7.04 | < .001 | 0.8695 | 0.9241 |
| Degree, Yr Degree, Str Yr, Field | 0.9251 | -5.26 | < .001 | 0.8986 | 0.9523 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties | 0.9245 | -5.49 | < .001 | 0.8989 | 0.9501 |
| Degree, Yr Degree, Str Yr, Field, Admin Duties, Rank | 0.9507 | -4.15 | < .001 | 0.9283 | 0.9736 |

1. The plot below shows the predicted values from each of the models above, by gender. The predicted values for males are generally higher compared to those for females. Among each gender there does not appear to be a large difference in the fit between the three models fitted.

****

1. A priori, I would choose the linear regression model to compare the difference in means between men and women. I would control for the variables degree, year of degree, starting year, administrative duties, and rank.

Methods: Robust linear regression was used to compare the difference in mean salary between male and female professors in 1995. Degree type, year of degree, starting year at UW, field, administrative duties, and rank were all included in the model to adjust for the estimate in mean difference. Along with the mean estimate, robust 95% confidence intervals were also computed. The null hypothesis, that the salary between male and female professors are equal after controlling for the variables listed above, was tested with a two-sided, t-test. Corresponding p-values were also reported

In 1995, females were found to have a mean salary $280.70 lower compared to male professors with the same degree, year of degree, starting year, administrative duties and rank. It would not be unusual to observe a difference in mean salary between $415.50 and $145.80 lower for women compared to men. The test of the null hypothesis that men and women have the same mean salary after adjusting for the variables listed above is less than 0.001. Therefore, we reject the null hypothesis and can say with confidence that give the same degree, year of degree, starting year, administrative duties, and rank female professors have a lower salary on average compared to men.