Biostatistics 515

Homework 7

3/7/14

1. The best way to model the variables *degree, field, and admin* is to use dummy variables as they are all nominal categorical variables.
2. Inference based on classical linear regression without robust SE estimates assume equal variances within groups. There is insufficient support to assume that the variation is the same across all groups. Classical linear regression (without robust SE estimates) would tend to be anti-conservative as the average confidence interval length will tend to be smaller than the linear regression model assuming robust SE estimates.
3. In a real situation, I would choose to adjust for the variables ‘year of degree’ and ‘starting year’ by modeling them with linear splines. Although linear splines lose some information through grouping observations, there may be some gain to relaxing the assumption of a linear relationship between salary and *year of degree* and *starting year*. There is no prior knowledge to support modeling the variables as quadratic functions, and there is not a convincing reason for using dummy variables to model these variables.
4. Methods: We use linear regression to evaluate the association between year of degree (modeled as a linear continuous variable) on salary, adjusting for the covariates degree, field, administrative duties, and sex by modeling each using dummy variables. We confine our analysis to the data collected for year 1995. Wald-based confidence intervals are computed using robust standard error estimates.

Inference: The estimated slope coefficient for year of degree is -89.87. This means that the model predicts that a population with a degree one year is associated with a decrease in average monthly salary of $89.87 when degree, field, administrative duties, and sex are held constant. This estimate is highly statistically significant (P<0.001), so with high confidence we reject the null hypothesis of no association between starting salary and the year a candidate was granted their degree. The 95% confidence interval suggests that the observed difference in salary by *year of degree* is not unusual if the true difference in monthly salary was anywhere between $98.30 lower and $81.43 lower, with the group having their degree one year later tending toward a lower monthly salary.

1. Methods: We use linear regression to evaluate the association between *start year* (modeled as a linear continuous variable) on salary, adjusting for the covariates degree, field, administrative duties, and sex by modeling each using dummy variables. We confine our analysis to the data collected for year 1995. Wald-based confidence intervals are computed using robust standard error estimates.

Inference: The estimated slope coefficient for start year is -56.88. This means that the model estimates that each successive year added to the start year is associated with a decrease in average monthly salary of $56.88 when degree, field, administrative duties, and sex are held constant. This estimate is highly statistically significant (P<0.001), so with high confidence we reject the null hypothesis of no association between starting salary and the year a candidate started their position. The 95% confidence interval suggests that the observed difference in salary by *year of degree* is not unusual if the true difference in monthly salary was anywhere between $66.13 lower and $47.63 lower, with the group starting their position one year later tending toward a lower monthly salary.

1. Methods: We use linear regression to evaluate the association between *start year* (modeled as a linear continuous variable) on salary, adjusting for degree, field, administrative duties, and sex using dummy variables and adjusting for year of degree as a continuous variable. We confine our analysis to the data collected for year 1995. Wald-based confidence intervals are computed using robust standard error estimates.

Inference: The slope coefficient for the *start year* is -111.96. We estimate that groups differing in starting year by one year (with the same degree, field, administrative duties, sex, and year of degree) are estimated to have a $111.96 difference in monthly salary, with the group starting in the later year having a lower salary. The 95% confidence interval suggests that the observed difference in estimated monthly salary in groups differing by one starting year is not unusual if the true difference is anywhere between $130.58 lower and $93.34 lower, with the group starting at a later year tending to have a lower monthly salary.

1. Methods: We use linear regression to evaluate the association between *year of degree* (modeled as a linear continuous variable) on salary, adjusting for degree, field, administrative duties, and sex using dummy variables and adjusting for starting year as a continuous variable. We confine our analysis to the data collected for year 1995. Wald-based confidence intervals are computed using robust standard error estimates.

Inference: The slope coefficient for *year of degree* is 27.15. We estimate that groups differing the year they were granted their degree by one year (with the same degree, field, administrative duties, sex, and starting year) are estimated to have a $27.15 difference in monthly salary, with the group with a later year of degree having a higher salary. The 95% confidence interval suggests that the observed difference in estimated monthly salary in groups differing by one year in when their degree was granted is not unusual if the true difference is anywhere between $8.68 higher and $45.63 higher, with the group with a degree granted at a later year tending to have a higher monthly salary.

1. The analyses in parts (a) and (c) both evaluated the association between start year and estimated monthly salary, except the model in part (c) adjusted the estimated effect of start year by the year of degree. Similarly, the analyses in parts (b) and (d) evaluated the association between year of degree and monthly salary, except the model in part (d) adjusted for starting year. The difference is scientifically relevant since adjusting for either variable changes the estimated effect of the predictor of interest. In the case of *year of degree*, the estimated effect changes from being negatively associated with salary to being positively associated with salary after adjusting for *start year.*
2. Inference on the difference in mean monthly salary paid to women faculty in 1995 and men faculty in 1995.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **adjusted for** | **Estimated difference** | **t-statistic** | **p-value** | **95% CI, low** | **95% CI, high** |
| **none** | -1334.73 | -14.04 | <0.001 | -1521.18 | -1148.29 |
| **degree** | -1266.15 | -13.40 | <0.001 | -1451.56 | -1080.75 |
| **Year of degree** | -614.12 | -7.17 | <0.001 | -782.24 | -446.02 |
| **Start year** | -614.58 | -7.06 | <0.001 | -785.31 | -443.85 |
| **Field** | -420.05 | -5.05 | <0.001 | -583.12 | -256.99 |
| **Administrative duties** | -419.73 | -5.17 | <0.001 | -578.99 | -260.47 |
| **Rank** | -280.66 | -4.08 | <0.001 | -415.52 | -145.81 |

Table 1: Hierarchical model of the difference in mean monthly salary paid to faculty in year 1995 across gender. The top row is the unadjusted model, and the subsequent rows adjust the estimated difference in mean salary across gender including the variable listed on the left and all preceding variables.

1. Inference on the ratio of geometric mean monthly salary paid to women faculty in 1995 and men faculty in 1995.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **adjusted for** | **Estimated ratio** | **t-statistic** | **p-value** | **95% CI, low** | **95% CI, high** |
| **none** | 0.812 | -13.73 | <0.001 | 0.788 | 0.837 |
| **degree** | 0.820 | -13.09 | <0.001 | 0.796 | 0.845 |
| **Year of degree** | 0.909 | -6.99 | <0.001 | 0.885 | 0.934 |
| **Start year** | 0.909 | -6.98 | <0.001 | 0.884 | 0.934 |
| **Field** | 0.936 | -5.06 | <0.001 | 0.913 | 0.961 |
| **Administrative duties** | 0.936 | -5.17 | <0.001 | 0.913 | 0.960 |
| **Rank** | 0.957 | -4.08 | <0.001 | 0.938 | 0.978 |

Table 2: Hierarchical model of the ratio of geometric mean monthly salary paid to faculty in year 1995 across gender. The top row is the unadjusted model, and the subsequent rows adjust the estimated ratio of geometric mean salary across gender including the variable listed on the left and all preceding variables.

1. Inference on the ratio of mean monthly salary paid to women faculty in 1995 and men faculty in 1995.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **adjusted for** | **Estimated ratio** | **Z-statistic** | **p-value** | **95% CI, low** | **95% CI, high** |
| **none** | 0.802 | -13.58 | <0.001 | 0.777 | 0.828 |
| **degree** | 0.810 | -12.99 | <0.001 | 0.784 | 0.836 |
| **Year of degree** | 0.898 | -7.12 | <0.001 | 0.872 | 0.925 |
| **Start year** | 0.896 | -7.04 | <0.001 | 0.870 | 0.924 |
| **Field** | 0.925 | -5.26 | <0.001 | 0.899 | 0.952 |
| **Administrative duties** | 0.924 | -5.49 | <0.001 | 0.899 | 0.951 |
| **Rank** | 0.951 | -4.19 | <0.001 | 0.928 | 0.974 |

Table 3: Hierarchical model of the ratio of mean monthly salary paid to faculty in year 1995 across gender. The top row is the unadjusted model, and the subsequent rows adjust the estimated ratio of geometric mean salary across gender including the variable listed on the left and all preceding variables.

1. Below is a graph of the predicted monthly salaries against the actual monthly salaries for each of the models in problems (3-5) above. The analyses all yielded the same conclusions about the statistical significance of the additional covariates in the hierarchical models, although the estimated t-statistics changed slightly.



1. I would have chosen the analysis model using linear regression with robust standard errors as I am interested in the absolute difference in monthly salary between women and men and not the ratio of the two salaries. For this, we created a hierarchical model to assess the association between gender and salary with sequential adjustment for degree, year of degree, starting year, field, administrative duties, and rank. When not adjusting for any covariates, women are estimated to be paid $1334.73 less than men in 1995. Adjusting for the degree earned reduced the difference to $1266.15. Adjusting for the year the degree was earned reduced the estimated difference by over 50% to $614.12. After adjusting for the year the candidate’s degree was earned, adjusting for start year did not change the estimated difference. After adjusting for field, women were paid on average $420.05 compared to their male counterparts with similar start year, year of degree, and degree granted. Adjusting for administrative duties did not markedly change the difference in the mean salary earned for women compared to men.