Statistical Intersectionality of Male Occupational Inheritance

Richard Paul Yim

University of California, Davis

UC DAVIS DEPARTMENT OF STATISTICS

Introduction

What is intersectionality? Intersectionality is the idea that social identities such as race, class and ethnicity, intersect to impact an individual's privileges and social experiences (e.g., discrimination, biases).

How does intersectionality relate in the general population? We ask this question with respect to occupational inheritance between fathers and sons. In particular, we take a look at census data collect from males in the continental United States from 1962 to 1973 to answer this question.

Data

What does the data look like? The data set that we use consists of occupations of father and son with classes "farmer," "unskilled," "skilled" and "professional." We turn these characteristics into 6 variables total, using farmer occupation as the baseline for both father and son, while turning the other classes into dummy variables. Additionally we have two binary covariates of race, being black or not, coded 1 or 0, respectively; and nonintact, corresponding to whether the family was disrupted (e.g., divorce or parent death), coded 1 or 0, respectively.

Dimension? The data set used in our analysis consists of $n = 21,107$ observations with 9 total covariates including an intercept column of ones, or

$$\mathbf{X} = \left[\mathbf{1}\ \mathbf{X}'\right].$$

where $\mathbf{X}'$ corresponding to a $n \times 8$ matrix, where 8 is the number of main effects. Additionally, our response variable, $\mathbf{y}$, will be the counts/frequency of observations belong to an *intersection.*

Methods

Generalized Linear Model Generalized linear models GLMs are regression models that linearly relate covariates $\mathbf{X}$ the expected values of a response $\mathbf{y}$ according to what is called a link function, g (monotone, differentiable). In particular, our final model has the form

$$\log(\mathbb{E}(Y|X^{(2)}) = X^{(2)}\beta^{(2)},$$

where our final model includes all second interactions including the main effects, and our new design matrix in our final model denoted $\mathbf{X}^{(2)}$ has dimension $n \times 31$, with corresponding parameter vector $\beta^{(2)}$. This particular model is known as a Poisson regression model, which is used when when the response variable $\mathbf{y}$ is distributed as integer count values. GLMs and their coefficients are also useful because of their intepretability and asymptotic properties (Lindeberg-Feller Central Limit Theorem, least squares and maximum likelihood estimation).

How good is the model really? Our final model was achieved by looking at both model residual deviances and the Bayesian information criterion (BIC) which penalizes both model complexity and model performance, with a bias toward more parsimonious models.

- BIC went from 4868.912 in initial main effects model to 636.7593 in full second order interaction model!
- At a statistically significant level the residual deviance went down from 41330.49 in the main effects model to 110.03 in the full second order interaction model! (Using χ^2 test and $\alpha = 0.01$.)

Results

We ask and answer three hypotheses according to our Poisson regression model:

1. Is there any effect from the predictors on population count? Or

$$H_0 : \sum_{i=1}^{31} |\beta_i| = 0 \text{ v.s. } H_1 : \sum_{i=1}^{31} |\beta_i| > 0$$

Yes! At an $\alpha = 0.01$ we have nonzero coefficients such that the sum of absolute values are nonzero at a statistically significant level.

2. Is there a positive effect with the interaction covariate between a son and father being both professionals? Or

$$H_0 : \beta_{sP:fP} \leq 0 \text{ v.s. } H_1 : \beta_{sP:fP} > 0.$$

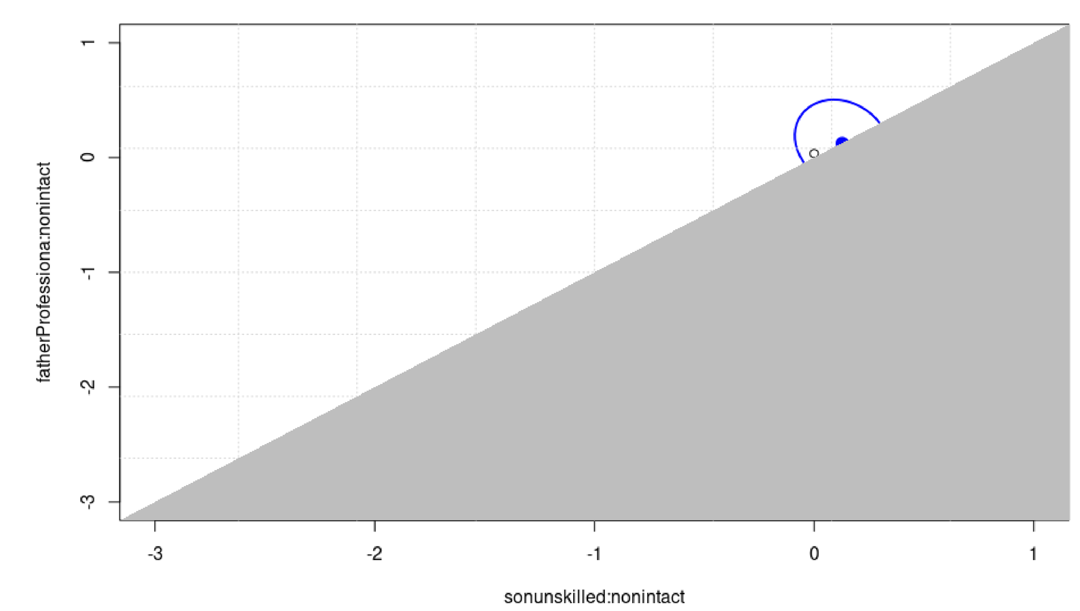
Yes! Again, at $\alpha = 0.01$, we are able to reject the null hypothesis and see that the slope parameter for interaction between father and son both being professionals as positive.

3. Is the parameter for fatherProfessional greater than the parameter for sonUnskilled given an interaction of family disruption? or

$$H_0 : \beta_{fP:nonintact} \leq \beta_{sU:nonintact} \text{ v.s. } H_1 : \beta_{fU:nonintact} > \beta_{sP:nonintact}.$$

No! At an $\alpha = 0.01$, we find that the corresponding confidence ellipsoid intersects the null hypothesis parameter domain in Figure 1.

Figure 1. 99% confidence ellipsoid; intersection with null parameter domain, can't reject!



Conclusion

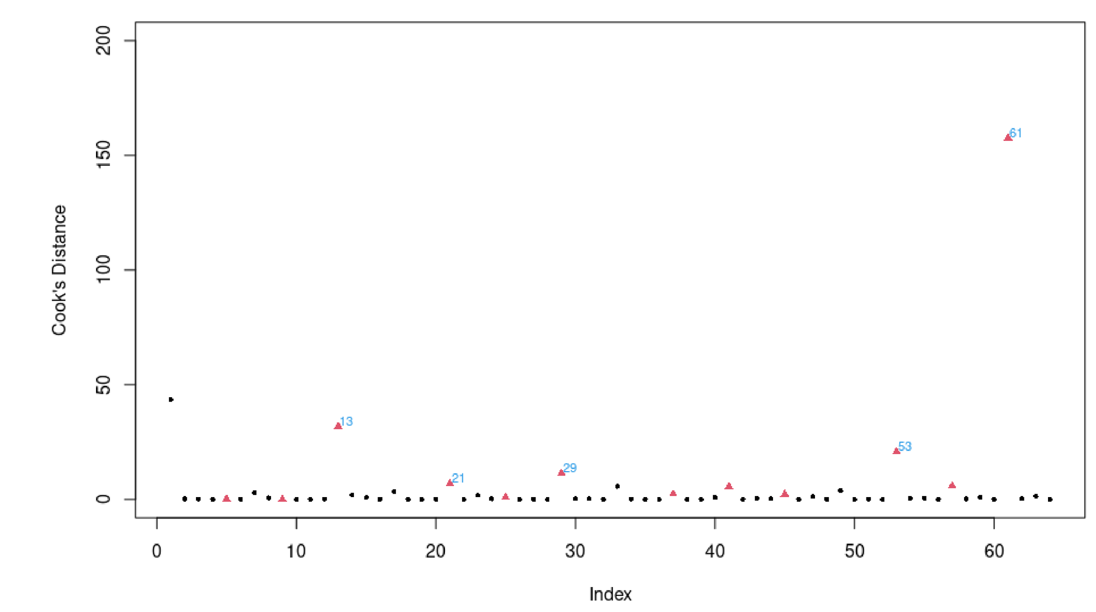
From our results we can conclude that there indeed is an overall regression effect from our result in (1). WE see that in (2) there is a positive association between fathers and sons being professionals, which relates to our outlier result: professional father son pairs are much more frequently occurring in the population than other occupation pairs. Finally, as an example test (3) shows that given an interaction of nonintact, or disrupted family, effects of both a father being a professional is at most the same effect as a son being unskilled (again, over a disrupted family interaction).

Future studies should naturally survey new census data to account for the digital age and data deluge, and also perhaps include other tangentially related family disruption events not restricted to parent status such as presence/absence or divorce (i.e., more covariates can be surveyed for).

Outliers

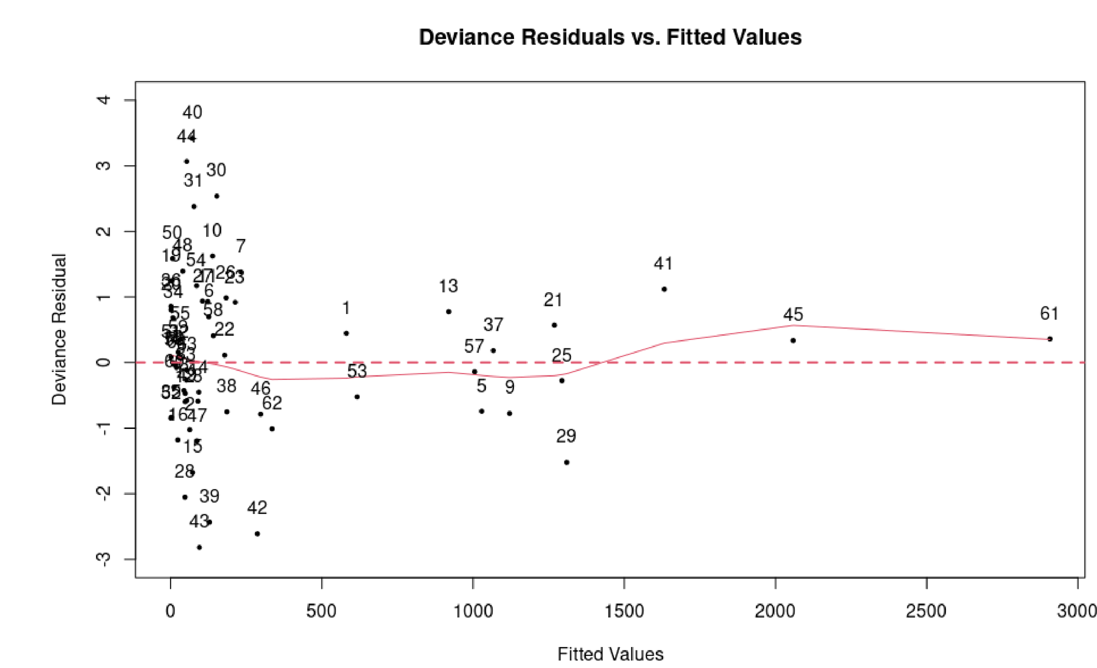
An outlier among us? We use both leverage analysis and Cook's distance to find data points that may affect both model performance in a significant way. Figure 2 with red dots indicates data points with high leverage effect on our parameter vector and with considerable Cook's distance. (Professional Father and son, not black, not nonintact.)

Figure 2. Cook's distance plot with red dots indicating high leverage points.



Goodness-of-fit

Does the model fit well? We show a plot below of the deviance residuals against the fitted values and see that we have good fit since the smoothing spline is fairly linear and nonzero.



Acknowledgements

Thanks to Han and Professor Müller for the teaching so far in the quarter. Thank you UC Davis Math Department for funding.

References

1. https://grodri.github.io/glms/datasets/mobility
2. http://lib.stat.cmu.edu/datasets/socmob
3. T.J. Biblarz, A.E. Raftery. *The Effects of Family Disruption on Social Mobility.* American Sociological Review. Vol. 58, No. 1 (Feb., 1993), pp. 97-109.

rpyim@ucdavis.edu