Final Project ——Jingjie Zhang

Jingjie Zhang

2020/12/22

Abstract

For the "The Association between Early Career Informal Mentorship in Academic Collaborations and Junior Author Performance", just like the title, it mainly discuss the relationship of informal mentorship and junior author performance, and it states that the opposite gender mentorship would have affect to the women career. In this report, is to analysis is that relationship exist and evaluate this article through the analysis.

keywords

mentorship, academic collaborations, junior author performance, opposite-gender,The number of years post mentorship

introduction

For this report, we choose option e which to critque the article about the mentorship. In this report, we would critique this article to use the data which he used. We would use a different way to express our own idea about the topic of career information association between the Academic Collaborations and Junior Author Performance. We would use different models to analyse the data. The regression model might give us a new way to express the ideal.

As the article "The Association between Early Career Informal Mentorship in Academic Collaborations and Junior Author Performance" say, the mentorship is really beneficial to the individual careers, and kind of reduce the inequality between the people and let people can fit into a new environment more easily(Alshebil,2020). In addition, this article through the data analysis found that there exists some form of mentorship. In this article, the author analyses the data and give a kind of form which increase the percentage of female mentor would cause the reduction of both female protege and female mentor. Through this analysis, it gives the idea of the opposite sex in the mentorship(Alshebil,2020).

data

The repository includes all the data which use in the article: "The Association between Early Career Informal Mentorship in Academic Collaborations and Junior Author Performance" It contain about three datasets which all have similar variables but in the different year load. All of those datasets collect the "average big shot"; "number of mentors"; "gender" and so on, which important to our analysis of the associate of the information related to career. For the method to select those data is that it select 3 million mentor–protege pairs to do the analysis as random sample. This analysis key feature is to analysis the relationship of mentorships through the analysis of mentor and protege pairs. The strengths of it is that it would be

clear with specific data from pairs. The drawback of it is that it might not be acurate that he data might have the error in different situation. The data in this dataset used in the article and it would provide the numerical data which show the relationship between the informal mentorship and junior author performance. We would use both Data_6_8yearcutoff and Data_7yearcutoff to reduce the error due to the different year load. The dataset mainly about the information of mentor and protege pairs.

```
##  [1] "Disambiguated_ProtegeID"   "AvgBigShot"
##  [3] "AvgBigShotBin"             "AvgHub"
##  [5] "AvgHubBin"                 "Avg_c5"
##  [7] "numMentors"                "numMentorsBin"
##  [9] "ProtegeFirstPubYear"       "ProtegeFirstPubYearBin"
## [11] "Pr0tegeGender"             "ProtegeGenderBin"
## [13] "AffiliationRank"           "AffiliationRankBin"
## [15] "AvgMentorsAcAges"          "AvgMentorsAcAgesBin"
## [17] "NumYearsPostMentorship"    "NumYearsPostMentorshipBin"
## [19] "ProtegeDisciplineID"       "Avg_c10"
## [21] "MaxBigShot"                "MaxBigShotBin"
## [23] "MedianBigShot"             "MedianBigShotBin"
## [25] "MaxHub"                    "MaxHubBin"
## [27] "MedianHub"                 "MedianHubBin"
```

```
##  [1] "Disambiguated_ProtegeID"   "AvgBigShot"
##  [3] "AvgBigShotBin"             "AvgHub"
##  [5] "AvgHubBin"                 "Avg_c5"
##  [7] "numMentors"                "numMentorsBin"
##  [9] "ProtegeFirstPubYear"       "ProtegeFirstPubYearBin"
## [11] "Pr0tegeGender"             "ProtegeGenderBin"
## [13] "AffiliationRank"           "AffiliationRankBin"
## [15] "AvgMentorsAcAges"          "AvgMentorsAcAgesBin"
## [17] "NumYearsPostMentorship"    "NumYearsPostMentorshipBin"
## [19] "ProtegeDisciplineID"
```

model

From the article, it states that the female mentor would affect the female protege, and the opposite gender would affect the relationship between the mentor and protege. Therefore, we use the liner regression model to find the relationship of mentor, average acdemic years and gender of the protege. The liner regression model would be:

$$y_i = \beta_0 + \beta_1 x_{AvgMentorsAcAges,i} + \beta_2 x_{Pr0tegeGender,i} + \epsilon_i$$

Here,
y represents the number of the mentors in this dataset.
β_0 represents the intercept of the model, and is the male with the mentors at average academic age of mentors 0.
β_1 is the coefficient for the AvgMentorsAcAges variablle, for everyone one unit increase in AvgMentorsAcAges, we expect a β_1 increase in the number of the mentors.
β_2 is the coefficient for the Pr0tegeGender variable, for every female, we expect a β_2 increase in the number of the mentors in this dataset.

Table 1: Linear regression output

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.450	0.012	383.190	0
AvgMentorsAcAges	0.008	0.000	17.947	0
Pr0tegeGendermale	-0.505	0.010	-50.806	0

Table 2: Linear regression output

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.853	0.010	395.495	0
AvgMentorsAcAges	0.008	0.000	21.543	0
Pr0tegeGendermale	-0.467	0.008	-56.109	0

Also, we would like use the table to see that the avg_c5 and avgc_10 with the protege gender(Daniel,2020). Just like below

```
##
## Table: Table of Avg_c5 and Avg_c10 by Pr0tegeGender
##
## |                           | female (N=343252) | male (N=711960)  | Total (N=1055212) | p value|
## |:--------------------------|:-----------------:|:----------------:|:-----------------:|-------:|
## |**Avg_c5**                 |                   |                  |                   | < 0.001|
## |   Mean (SD) |  11.010 (23.201)  | 10.005 (20.560)  |  10.332 (21.460)  |        |
## |   Range     | 0.000 - 2126.000  | 0.000 - 2412.000 | 0.000 - 2412.000  |        |
## |**Avg_c10**                |                   |                  |                   | < 0.001|
## |   Mean (SD) |  15.960 (36.462)  | 14.980 (32.829)  |  15.299 (34.057)  |        |
## |   Range     | 0.000 - 4290.000  | 0.000 - 4531.000 | 0.000 - 4531.000  |        |

##
## Table: Table of Avg_c5 by Pr0tegeGender
##
## |                           | female (N=348437) | male (N=686058)  | Total (N=1034495) | p value|
## |:--------------------------|:-----------------:|:----------------:|:-----------------:|-------:|
## |**Avg_c5**                 |                   |                  |                   | < 0.001|
## |   Mean (SD) |  11.170 (24.072)  | 10.066 (21.232)  |  10.438 (22.235)  |        |
## |   Range     | 0.000 - 2731.000  | 0.000 - 3026.500 | 0.000 - 3026.500  |        |
```

result

From abouve table of the liner regression model, we found out that the year load would not really affect the result and relationship, both 6-8 and 7 year load show similar result. From those table we found out the number of mentor have the relationship with the pretege gender and a little with the avaerage acdemic year. From the table of liner regression model, we could states that with 1 incerase in avaerage acdemic year, the number of mentors would increase 0.008, and with 1 increase in protege gender is male, the number of mentors would decrease 0.505. Which follow the article statement,the opposite gender would affect mentor and protege and mentor would decrease through the proportion increase of male protege.

discussion

summary

This artcle state most about the dataset and give a right direction of the analysis. The informal mentorship related to the junior author and academic collaborations.

conclusion

crtique the article

From this article, it states that the mentorship relation might related to the gender of mentor and protege. Through the analysis in the report, it seems to have a relationship. This article has a clear structure and framework. It contains every part which a report needs. Moreover, it gives a clear abstract to open the analysis in this report(Alshebil,2020). However, the analysis in this report does not prove the statement which it gives in the abstract. In this article, the analysis more about the percentage of the experience of mentor and protege. In some way, it might need more about the data analysis with the aim of this article. Perhaps, create a model to see the relation. The mentor gender affect only state at abstract, but not provide any analysis.

crtique the data

For the data used in this article. It has some drawback, Thus it has the protege gender, but not the mentor gender. The article, states that it might be related to the female mentors, but the dataset does not provide the gender of the mentor. For the character NumYearsPostMentorship, ProtegeFirstPubYear, AvgMentorsAcAges, and so on. Those characters in this dataset seem too large. Those characters are mostly related to age and time, but the number of those be too large which against common sense(Daniel,2020).

weakness

There are several weaknesses. First, different schools or organizations have a different method to add up. We need to figurate it out. Next, we need to add more variables to increase the model prediction. we can add more variable which might be related to the number of mentors. Finally, the increase in number, not stately related to the mentorship relationship.

next step

The result of the analysis would be completed in the next several years, with the time pass, we would get more data to analyze the relationships of the mentorship and it might change in different year and different year load.

citation

AlShebli, B., Makovi, K., & Rahwan, T. (2020, November 17). The association between early career informal mentorship in academic collaborations and junior author performance. Nature Communications. https://www.nature.com/articles/s41467-020-19723-8?error=cookies_not_supported&code=5bd4d63f-3451-46e8-8935-dff1f69b66f5#Abs1

Daniel E. (2020, November 21). Retrieved December 22, 2020, from https://danieleweeks.github.io/Mentorship/

Hullman, J. (2020, November 19). Are female scientists worse mentors? This study pretends to know « Statistical Modeling, Causal Inference, and Social Science. Are Female Scientists Worse Mentors? This Study Pretends to Know. https://statmodeling.stat.columbia.edu/2020/11/19/are-female-scientists-worse-mentors-this-study-pretends-to-know/