Estimating R Anxiety Level Distribution Among Students At the University Using MRP*

Yiqu Ding

2020-12-10

Abstract

In this report, I look at different factors that affect a studebt's anxiety level towards using R. The data is obtained from onlien survey responses which contained variables that could influence anxiety levels. I then run a multilevel regression on the sample and post-stratify them using a simulated student census. After the post-stratification, the results tells us…(.)

0.1 Keywords

MRP, R, Psychology, Education

1 Introduction

Relatively speaking, the science of statistics is a new discipline. In 1998, the public image of statistics was poor, and almost nobody knows what statisticians do. Now, statistics is an essential tool for nearly all millennial industries. Accompanied by technological improvements in computers, R has become a necessary tool for all statistical practitioners; that makes the teaching and training of R extremely important.

I am interested in the distribution of R anxiety levels among students. In this report, I will show the method to estimate the anxiety distribution among students using a sample I collected from the University of Toronto. I run a multilevel regression on my sample and then post-stratify the results using a simulated student census to get population estimates.

The results of the analysis can be useful in many ways. The university can periodically conduct this analysis to keep track of teaching results; generally, this method should be solid for similar reports in any other university. Students can use the information as a threshold to understand where they stand among their peers. It is also possible to study the effect of a treatment such as data camp using this approach, which potentially saves cost for the department.

In paper we run our analysis in `R` (R Core Team 2020). We also use the `tidyverse` which was written by Wickham et al. (2019).

2 Data

The data is of students enrolled in a program under the statistics department and are in their third year of study. (Figure 1).

*Code and data are available at: https://github.com/dding33/STA304-PS5.

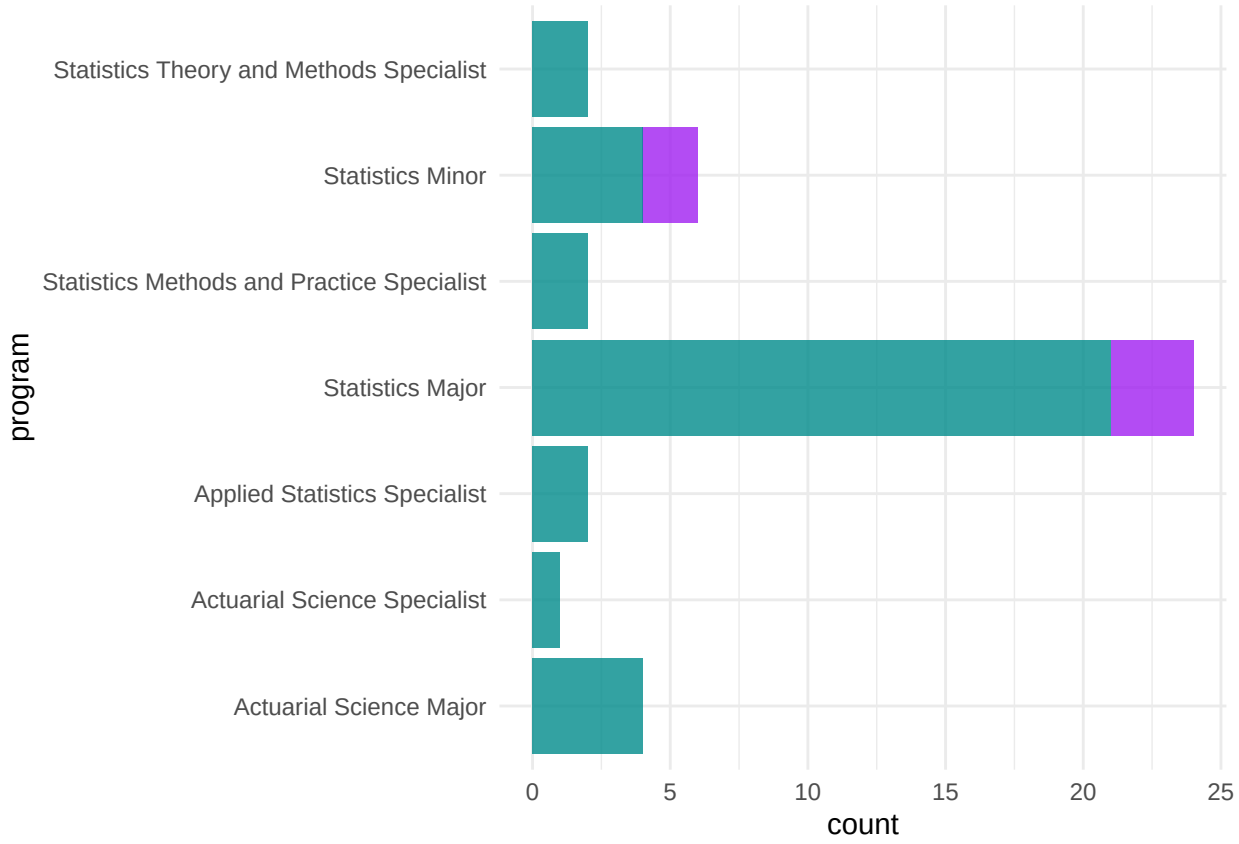

Figure 1: Distribution of Programs

3 Model

$$\hat{y}_S^{PS} = \frac{\sum N_j \hat{y}_j}{\sum N_j} \tag{1}$$

We get our post-stratification estimates by equation (1).

4 Results

5 Discussion

Appendix

References

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.