

Research Methods in Management course | Final Project | How gender influence on academic performance.

Leung Yiu Fung, Yu Xinning, Saloev Shahzod, Malysheva Elena, Zavalina Anna, Krukovskii Boris, Trofimova Iлона, Svetosh Ksenia, Vinogradov Danila, Mukhataev Dmitriy, Sokolovskiy Vladislav

^aSchool of Economics and Management Saint-Petersburg

Abstract

In this paper, we investigated the question of the influence of a student's gender on his or her performance in a Portuguese language course. Our work included data transformations, data visualizations, the use of statistical tests, as well as regressions. As a result of our work, we cannot firmly state that gender affects the performance of students, but in the process of the study we observed certain patterns in the dependence of these variables.

Keywords: Education, Academic Performance, Gender

1. Introduction

Education is an important factor in the development of a country and the success of its citizens. Kappe[1] claims that academic, scientific and technological strengths have become increasingly important. Stephenson[2] defined one of the main goals in higher education is giving students the confidence and ability to take responsibility for their own continuing personal. Therefore, classifying students' performance and identifying areas where they need improvement can help ensure that they are prepared for exams and future courses. The question of the gender gap has been studied for decades. This gender gap can be seen, for example, in physical indicators, such as achievements and records in sports, in wage inequality, and in student achievement. Researchers in different countries have documented and researched the factors that influence how students perform differently according to their gender. Ghazvini[3] discovered gender difference existing in performance attained in school subjects of Literature and Mathematics and didn't count to boy and girls motivation. Ranjeeth[4] used gender-wise performance analysis based on different parameters. From Ranjeeth's researched showed the relation of parental involvement will be change the academic performance, Griffith [5] also has the same result as Renjeeth.

In this report, our team wants to study the influence of a student's gender on his or her performance in a Portuguese course. Our results are related to the extensive literature in which this issue has been studied. As noted earlier, studies generally show that girls have higher grades. However, there are studies that contradict them. This paper wants to contribute to the study of this issue.

2. Setting

Within our report, the statistical test showed the significance of differences in academic performance between boys and girls. Also, when constructing a linear regression with only the variables responsible for academic performance, this dependence was confirmed and allowed the difference to be expressed numerically. However, when control variables were added to the regression, the situation changed, with the result that gender as a factor affecting academic performance lost its significance at the 99% confidence interval level.

*Corresponding author

The following will be a more detailed description of the data, our analysis, results, and discussion of our findings.

3. Data and empirical strategy

The data from Kaggle[6] includes 649 observations and 33 variables. The data includes many variables including student gender, age, family size, parental education, study time per week, frequency of alcohol consumption, and many others. For the purpose of our report, since we planned to study the effect of gender on student performance, the same-name variable gender is taken as the variable of gender, which includes 383 females and 266 males. And as a proxy variable to study student performance, we take the student's final grade for the course ranging from 0 to 19 with a mean of 12.25 for females and 11.41 for males, the cumulative mean being 11.91. As a result of histogram building it was found that there were zero values among the final grades, which also show themselves as outliers when the boxplot is constructed.

It was decided to exclude these zeros because, first of all, they distort our normal distribution and decrease the average, as well as based on the data it is unclear the nature of these zeros, whether a student made an honest attempt to pass the exam and got a 0, or just did not come and got a 0. Thus, the inclusion of those who did not show up could have a negative impact on the result. Also, based on the boxplot, the data have different variances, from which it was decided to use Welch's t-test rather than The independent samples t-test, because Welch's test does not make the assumption of equal variances between samples. After excluding the null values, 634 observations remained. These included 376 females and 258 males. The mean for males became 12.48 for females and 11.76 for males, and the cumulative mean was 12.19. So there information about variables after null excluding is represented in fig.1.

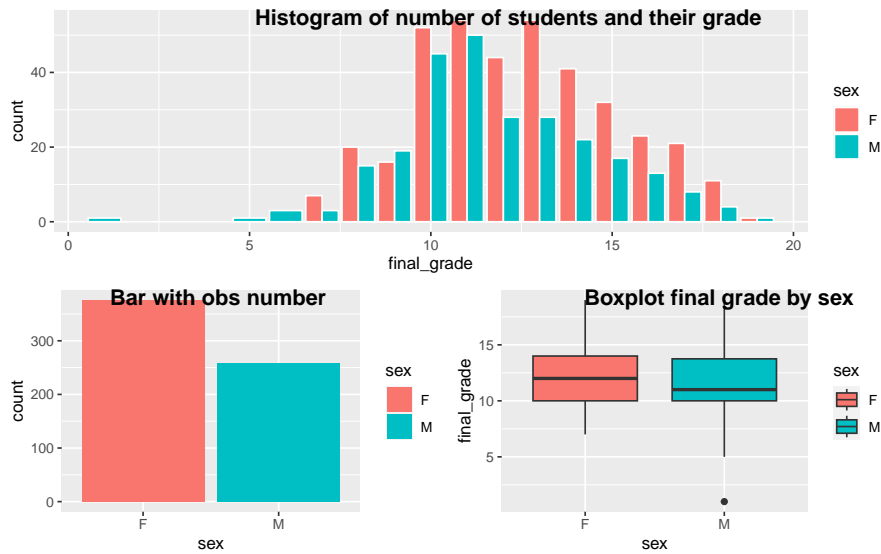


Figure 1: Descriptive figures

4. Result

The test conducted showed that we can reject null hypothesis about the equality of average achievement values between students of different genders on the 95% confidence level. Thus, there is a significant difference at the level of 95% in mean final grade between male and females. Also, a statistical test confirmed at the 99% significance level that females have higher scores than males.

```
##
## Welch Two Sample t-test
##
## data: final_grade by sex
## t = 3.3383, df = 549.72, p-value = 0.0009001
## alternative hypothesis: true difference in means between group F and group M is not equal to 0
## 95 percent confidence interval:
## 0.2970362 1.1463499
## sample estimates:
## mean in group F mean in group M
## 12.48138 11.75969
```

```
##
## Welch Two Sample t-test
##
## data: final_grade by sex
## t = 3.3383, df = 549.72, p-value = 0.0004501
## alternative hypothesis: true difference in means between group F and group M is greater than 0
## 95 percent confidence interval:
## 0.3654941 Inf
## sample estimates:
## mean in group F mean in group M
## 12.48138 11.75969
```

Build OLS regression with final_grade as dependent and sex as independent variable. Conducting also Breush-Pagan test test to check heteroscedasticity.

```
##
## =====
## Dependent variable:
## -----
## final_grade
## -----
## sexM -0.722***
## (0.216)
##
## Constant 12.481***
## (0.138)
##
## -----
## Observations 634
## R2 0.017
## Adjusted R2 0.016
## Residual Std. Error 2.670 (df = 632)
## F Statistic 11.175*** (df = 1; 632)
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01
```

```
##
## studentized Breusch-Pagan test
##
## data: model_sex
## BP = 0.015832, df = 1, p-value = 0.8999
```

Regression shows that male students receive a lower final grade than female students. OLS regression showed that the coefficient is significant at the 0.01 level. Considering the presented regression coefficient sexM for the core independent variable which is statistically significant at 99% confidence level, there is enough evidence to suggest that sex influences the final grade. The negative coefficient for sexM in the regression output indicated that being a Male is associated with decrease in final grade.

From the Breush-Pagan test there is enough sufficient evidence to suggest the existence of heteroscedasticity problem at 95% confidence level. In order to facilitate this problem the robust option is going to be used.

As control variables we take the binary variable “higher” reflecting the student’s desire for higher education. If a student wants to get a college degree, he or she most likely understands that he or she needs a high score in school in order to get in. Thus, he or she tries to do better in school and his or her academic performance is higher. The studytime variable reflects the number of study hours per week, which also affects a student’s performance, because the more often a student studies, the higher his or her grade point average is likely to be. We also add as a control a change of absences, which reflects the number of missed classes. In theory, the more classes a student misses, the worse his or her performance, because he or she misses some of the knowledge. We also use the Variance Inflation Factor to check for multicollinearity.

```
##          GVIF Df GVIF^(1/(2*Df))
## sex      1.067875  1      1.033380
## higher   1.067797  1      1.033342
## absences  1.036742  1      1.018205
## studytime 1.131585  3      1.020817

##
## =====
##                               Dependent variable:
##                               -----
##                               final_grade
##                               (1)          (2)
## -----
## sexM                        -0.731***    -0.413*
##                               (0.233)      (0.216)
##
## higheryes                    2.431***
##                               (0.351)
##
## absences                    -0.079***
##                               (0.022)
##
## studytime2 to 5 hours        0.686***
##                               (0.244)
##
## studytime5 to 10 hours      1.417***
##                               (0.333)
##
## studytimemore than 10 hours  1.169**
##                               (0.477)
##
## Constant                    12.426***    9.829***
##                               (0.149)      (0.383)
##
## -----
## Observations                634          634
```

```
## Residual Std. Error      2.512 (df = 632) 2.490 (df = 627)
## =====
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

As a result from the VIF we cannot suggest the problem of multicollinearity. Regression with controls shows that the male students receive a lower final grade than female students only statistically significant at 90% confidence level. So there is no sufficient evidence to state that male students receive a lower grade than female student. It can be seen that final mark is also influenced by the amount of time spent by a student to study and the number of absences. There is statistically significant and positive influence of incentive to get higher education on the final mark. On average, a student who wants to get higher education have 2.431 higher final mark holding other variables constant.

5. Conclusion

In this study we examined the difference in academic performance between male and female students. As part of our report, we use a statistical test to determine the significance of differences in grades between students. We also use a number of variables as controls to get more representative results. To get more information about the effect of gender on the expected mean score and to account for the possible problem of heteroscedasticity, we use a robust regression version. Thus, we conclude that there is no conclusive evidence that there are differences in mean scores between students of different genders. However, we also found that a student's desire for higher education, time spent per week studying a discipline significantly affects a student's academic performance, and skipping classes, in turn, negatively. Our results have certain limitations, such as the non-normal distribution of test absences alternately. This report can be used to further develop and explore the topic of the influence of factors on student achievement.

Nevertheless, in this report we have tried to use the tools that we have had time to learn during the short course Research Methods in Management and make the best of what we could.

References

- [1] Kappe, Rutger, and Henk Van Der Flier. "Predicting academic success in higher education: what's more important than being smart?." *European Journal of Psychology of Education* 27.4 (2012): 605-619.
- [2] Stephenson, John. "The concept of capability and its importance in higher education." *Capability and quality in higher education*. Routledge, 2013. 1-13.
- [3] Ghazvini, Sayid Dabbagh, and Milad Khajehpour. "Gender differences in factors affecting academic performance of high school students." *Procedia-Social and Behavioral Sciences* 15 (2011): 1040-1045.
- [4] Ranjeeth, S., T. P. Latchoumi, and P. Victor Paul. "Role of gender on academic performance based on different parameters: Data from secondary school education." *Data in brief* 29 (2020): 105257.
- [5] Griffith, James. "Relation of parental involvement, empowerment, and school traits to student academic performance." *The Journal of educational research* 90.1 (1996): 33-41.
- [6] Data: <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>