Representation of female students in STEM majors across time at Grinnell College

Background

According to the World Economic Forum there is an underrepresentation of females in STEM with only 16% of female students graduating from STEM majors as opposed to 30% for male students[1]. We wanted to observe whether female students are underrepresented in STEM majors at Grinnell College, a liberal arts college that prides itself for its progressive ideology and promotion of diversity across various academic disciplines. Moreover, we wanted to examine how these trends change over time and what they might be in the future. By creating a multiple regression model that predicts the percentage of female students with regards to major and year as independent variables, we can not only assess the strengths of these predictors, but also gain insight into disciplines with lower percentages of female students and discover the rate of change in female participation across various stem majors. This analysis can help the college understand which academic departments are underrepresented in female participation and which majors are showing a slow female student growth rate, thereby advising the college's decisions on dispersion of resources to promote women in STEM subjects.

Methods

Data Collection

Our data set comprised of 11794 cases from 1986 to 2019. Each case was a student with their graduation year, major(s), sex and ethnicity. This data was obtained from the Analytics and Institutional Research at Grinnell College.

Variable Creation

Since our purpose was to analyse the percentage of women in STEM fields across time, we stratified the data to create a subset of every STEM major based. Within every respective subset of a major, we tallied the frequency of men and women for each year to calculate the percentage of women for each year. After repeating this process for every major, we compiled every major and its years with their respective percentage of female students into one single sheet. In doing so, we reduced our 11794 cases to 208 observations.[2] Each observation included the year, major and the percentage of female students. Thereby we had organized our data with year and major as our explanatory variables and percentage female as our response variable.

Analytic Methods

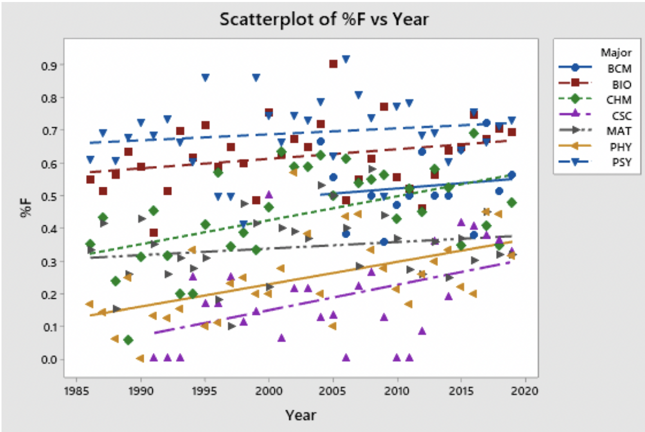
Firstly, we sought to analyse the relationship of time (continuous) and major (categorical) on percentage of female students (quantitative). Secondly, we wanted to assess the interaction between year and major to question its interdependence, and finally we wanted to predict future trends. For these reasons we decided to use a multiple linear regression model with interactions between our independent variables. Since our model predicts the percentage of females in a certain major in a given year, the aforementioned double counting of double majors does not impact the model, as no student is counted twice within one single major.

Results

---

[1] "Gender Parity and Human Capital." *Global Gender Gap Report 2016*, World Economic Forum, reports.weforum.org/global-gender-gap-report-2016/gender-parity-and-human-capital/.

[2] This process involved double counting double majors, as they would be represented twice when stratifying into subsets by major.



Fi ercentage female across time.

As seen in Fig.1[3], the residuals of the model are normally distributed, thus the assumptions of the model have been met. Our regression model can be visualized through the scatterplot in Fig.2.Our first question was whether certain majors have an underrepresentation of female students. We can observe in Fig. 2 that the regression lines of majors like Physics, Computer Science and Math are consistently below 0.5 indicating less than 50% of students in these majors are women.

Regression Equation

Major	
BCM	%F = -5.9 + 0.00319 Year
BIO	%F = -5.25 + 0.00293 Year
CHM	%F = -14.19 + 0.00731 Year
CSC	%F = -15.40 + 0.00777 Year
MAT	%F = -3.68 + 0.00201 Year
PHY	%F = -13.43 + 0.00683 Year
PSY	%F = -3.00 + 0.00184 Year

Fig.5: Regression Model Equation

By using interactions our model allows for each major to have a unique slope or rate of change thus enabling us to answer our second question: Whether STEM majors have a positive growth rate in percentage of women and how do growth rates differ across majors? Evidently, the positive coefficient of Year for every individual major in the regression equation (i.e the second coefficient) in Fig.5 indicates that the female participation within each of these majors is increasing every year.

We have the highest growth in Computer Science, Chemistry and Physics in that order. All of these majors observe an increase of approximately 0.7% female majors every year. We can test these observations by analysing the coefficient table in Fig.3[4] where we chose Math to be the reference model due to its flat slope and it placement within the middle of other regression lines in Fig.2[5]. By equating the significance level to 0.05, we conclude that the coefficients for the interactions between year and major are statistically significant for Chemistry (p=0.026), Physics (p=0.042) and Computer Science (p=0.041) when compared to the reference category of Math.

By running an ANOVA test on our regression model we observed a P-Value of 0.067 for Major as a predictor, and 0.078 for the Year*Major interaction acting as predictor, showing moderate evidence that using Major and the Major*Year interaction helped improve the fit of the model.

Lastly our model had a reasonably high R^2 value of 72.83% indicating that the model explains 72.83% of all variability in the observed data.

Discussion

As mentioned earlier, we sought to answer the following questions. Are women underrepresented in certain fields? Is there a positive growth rate in female participation in

[3] Refer to the Appendix
[4] Refer to the Appendix
[5] Refer to the Appendix

STEM majors and how does this growth rate vary across majors? And finally, what can we expect if these trends hold true for the future?

The answer to the first question is yes. Majors such as Physics, Computer Science and Math consistently underrepresent women with less than 35% of students being female. Similarly, the answer to our second question is also yes. Evidently, through the use of interactions, it is observable that every major has a positive growth rate of women. Moreover, these growth rates vary differently across majors. For example, we can observe that Physics, Computer Science and Chemistry add roughly 0.7% more women each year, however Math only exhibits 0.2% increase in women each year. Lastly, if these results hold true, we can expect very welcoming changes in the STEM division. For example, in 1986, Chemistry comprised of only ~30% female. However, with a growth rate of 0.7% per year, in the span of 33 years, Chemistry has grown to have a majority of ~55% in 2019. This can be attributed to the College's emphasis on the interdisciplinary nature of Biology and Chemistry and the subsequent introduction of Biological Chemistry as a new major. Drawing on this example, currently in 2019, we have close to a 30% female participation rate in Physics and Computer Science akin Chemistry in 1986. Therefore, if Physics and Computer Science continue to grow at their current rate of ~0.7% each year, we can expect these majors to be balanced within the next 35 years if not earlier. Additionally with the boom in the technology sector, the promotion of diversity through the Women Techmakers Scholars Program, The Grace Hopper Celebration of Women in Computing, and Grinnell College's push to hire more female CS professors, tutors and mentors, the time horizon for female balanced departments may be expedited. Unfortunately, Math as a major has not seen such changes. With a 0.2% growth rate of female majors in the past 33 years female participation has increased by a mere 7% from 31% to 38%. This stagnation should encourage the college to promote diversity within the Math department.

The scope of our project is extensive. This analysis can be replicated for every major at the College. Furthermore, with a larger sample of students from various universities and colleges across the US, this research can be used to draw statistically significant conclusions of the greater population (i.e. Every US college and university). However, Grinnell College's uniqueness in its ideology, geographical location and student demographics limits it from being a representative sample of the population. Additionally, due to various constraints, we were unable to include concentrations into the model. Another limitation is that we ignored several cases of female students with extensive classes in a certain department lacking a major. A more accurate analysis would involve the study of the make up of every single class from every department over a longer time period.

To conclude, Grinnell College's efforts to promote women in STEM can be deemed successful with every STEM major increasing the representation of female students over time. Furthermore, with changes in professional and academic landscapes and the greater push for women in STEM, Grinnell College may just pioneer the representation of women in this predominantly male division, as a result, improving gender equality at the college.

References

"Gender Parity and Human Capital." *Global Gender Gap Report 2016*, World Economic Forum, reports.weforum.org/global-gender-gap-report-2016/gender-parity-and-human-capital/.

Appendix

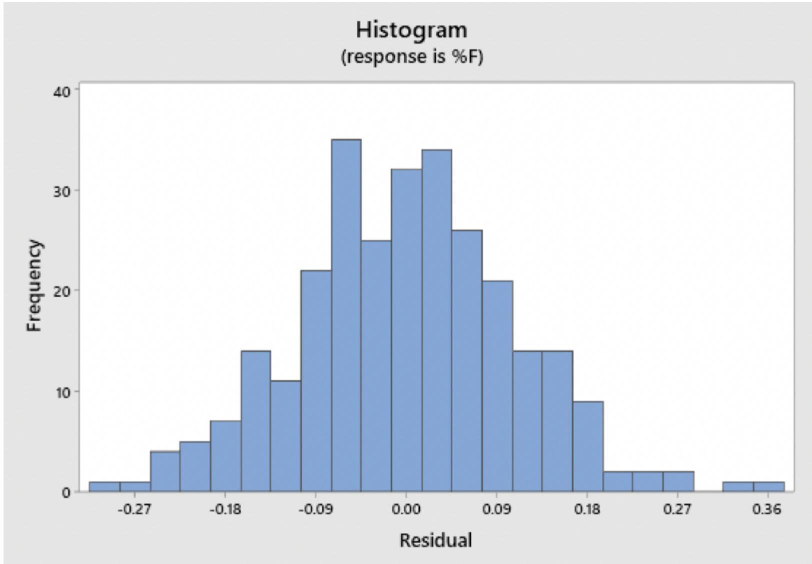


Fig.1: Distribution of the model's residuals.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-3.68	2.73	-1.35	0.179	
Year	0.00201	0.00136	1.47	0.142	4.06
Major					
BCM	-2.2	12.3	-0.18	0.858	188891.06
BIO	-1.57	3.86	-0.41	0.684	63304.82
CHM	-10.52	4.73	-2.22	0.027	54986.86
CSC	-11.72	5.62	-2.09	0.038	67492.33
PHY	-9.75	4.73	-2.06	0.040	54986.86
PSY	0.68	4.73	0.14	0.886	54986.86
Year*Major					
BCM	0.00118	0.00613	0.19	0.848	188974.27
BIO	0.00092	0.00193	0.48	0.632	63305.30
CHM	0.00530	0.00236	2.25	0.026	54987.04
CSC	0.00577	0.00280	2.06	0.041	67532.20
PHY	0.00482	0.00236	2.04	0.042	54987.04
PSY	-0.00016	0.00236	-0.07	0.945	54987.04

Fig.3: Table of coefficients of the model

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	13	8.7684	0.674493	55.46	0.000
Year	1	0.0264	0.026359	2.17	0.142
Major	6	0.1455	0.024256	1.99	0.067
Year*Major	6	0.1398	0.023303	1.92	0.078
Error	269	3.2717	0.012162		
Lack-of-Fit	201	3.2717	0.016277	*	*
Pure Error	68	0.0000	0.000000		
Total	282	12.0401			

Fig.4: ANOVA of the multiple regression model.