FIRST YEAR EATS: A FIRST ATTEMPT AT COMBATING FOOD INSECURITY ON COLLEGE CAMPUSES

An Undergraduate Research Scholars Thesis

by

ALEX PETERS

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. Alan Dabney
Dr. Suma Datta

May 2020

Major: B.S. Statistics

TABLE OF CONTENTS

ABSTRACT

First Year Eats: A First Attempt at Combating Food Insecurity on College Campuses

Alex Peters
Department of Statistics
Texas A&M University

Research Advisor: Dr. Alan Dabney
Department of Statistics
Texas A&M University

Research Advisor: Dr. Suma Datta
Department of Biochemistry and Biophysics
Texas A&M University

First Year Eats (FYE) is a new program on Texas A&M's campus dedicated to teaching freshmen how to cook from the comfort of their dorms while providing them with the resources necessary to do so. To assess the impact of FYE on students, we compare the Fall semester final GPA's of students who participated with a fitting control group of students who did not participate. We estimate the impact of FYE on student GPA's for various demographic subsets of students using a bootstrap linear regression on a host of indicator variables including an indicator for involvement in FYE. The results show that FYE participants have a higher average GPA than non-participants across all demographic strata. This effect is found statistically insignificant, but we do find a significant association between FYE participation and differential GPA across income strata. We speculate on causes of this and points of future research, as well as make suggestions to the FYE team regarding potential improvements to the program going forward.

ACKNOWLEDGMENTS

I would like to thank my research advisors, Dr. Alan Dabney and Dr. Suma Datta, for giving me the opportunity to work on this incredible project with them, and for being so instructive throughout this process. I would also like to thank the academic advisor for the Dept. of Statistics, Alyssa Brigham, for making me aware of this project opportunity in the first place; she was instrumental in my involvement with First Year Eats.

Finally, I want to thank all of my family, friends, teachers, and professors who have helped set me up for success in this project and beyond; particularly my Mother and Father, who have always encouraged me in my pursuits and have always reminded me to do everything to the best of my abilities.

NOMENCLATURE

FYE	First Year Eats
GPA	Grade Point Average
LAUNCH	Program at Texas A&M University that houses FYE
Learning Community	General categorization of many on-campus programs designed to help incoming freshmen make the transition to college life in a variety of ways, often based on need
Food Insecurity	The lack of consistent access to food over an extended period of time due to systematic reasons, such as financial inability, or lack of open time to eat

CHAPTER I

INTRODUCTION

It seems to be well-known and highly intuitive that having consistent access to food is a vital contributor to a students academic and overall well-being. In fact, much work has been done on the importance of food availability and quality in primary and secondary school systems. We have learned that food insecurity is correlated with low math and reading scores in K-12 students, and that a lack of food on the table can reduce a child's chances of graduating from high school (Alaimo, Olsen, & Frongillo, 2001). The research on food insecurity of primary students has inspired great progress in the role that schools play in keeping students fed. However, one is surprised to find that the state of such research on university campuses was all but absent until fairly recently.

Food Insecurity on College Campuses

College students are just like everyone else; they need food, shelter, relationships, and goals, among other things. It seems obvious that lower-level needs like food and shelter need to be tended to before asking someone with such needs to excel in a college environment.

Unfortunately, this is simply not the reality of many students. One study showed that average food insecurity among college students across the nation is an alarming 42%, while another found food insecurity among college students in Illinois as high as 84% (Bruening et al., 2017; Morris et al., 2016). These disturbing numbers are beginning to uncover a problem that is impossible to shy away from: many students are at an immense disadvantage due to inconsistent access to food, and many of them will go on to leave the university space entirely because of it (Woerden, 2018).

The Goals of First Year Eats

This is where First Year Eats (FYE) becomes relevant. FYE is categorized as a Learning Community and incoming freshmen are enrolled accordingly; in this case, enrolled students are those who are projected to be at greater risk of food insecurity at some point in the semester, and the program is also open to any who would like to invest their time in learning to cook. It is hoped that this program will help these students stay academically and socially focused by providing them a rudimentary cooking skillset and the resources with which to keep food on the table.

In the best case scenario, we project that the students in FYE will have higher-on-average GPA's compared to other freshmen enrolled in Learning Communities unassociated with FYE. If this can be well-established, it is hoped that this program will become a staple at Texas A&M and a model for other universities, in an effort to reduce college hunger rates and positively contribute to overall student success across the country.

An Important Caveat

It is of the upmost importance to be very clear about the sampling methodology here; our sample is *not* a simple random sample from the population. Ours is a need-based sample, and in some cases, participants are self-selected. Because of this, the strength of our conclusions is limited to statements about *associative relationships.* These can give an indication of where to look for true *causal relationships,* but we cannot infer one from the other. Do keep this in mind in reading what follows.

CHAPTER II

METHODOLOGY

FYE consists of a variety of classes and seminars intended to help freshman students learn to cook easy meals that they can make themselves. The effectiveness of the program as a whole will be based on the effectiveness of these classes and resources and the secure environment that these things create for participating students. To determine whether FYE is effective, we will need to look at data from the food pantry stocked as part of FYE and the academic data of the participants in comparison to non-participating students. With both of these datasets in hand, we will be able to draw informative conclusions on preliminary effects that FYE has on the students, and use those conclusions to drive the future of the program.

Data Collection: Pantry Data

The first major chapter of this project consists of building a web application that the LAUNCH team will use to analyze changes in food usage over time. Because FYE is providing a free-of-charge pantry for participant students to take whatever they like, it is of interest to catalog the changes in usage of various ingredients, as well as how those changes correlate with the content of the cooking classes given to the students. It is hoped that we will see positive change in the usage of ingredients that constitute a recipe being taught, as well as flat or negative change in the usage of ready-made foods and snack items. A change like this would show that, at minimum, the classes being given are actually having an effect on the students involved in the program. However, the direction or magnitude of that effect cannot be determined from this data alone, so we will need more information to draw concrete conclusions.

Data Collection: Student Data

After the food pantry application is (more or less) complete, we will turn our attention to the actual academic data from participants that we will have on hand after the end of the fall semester. The students participating in this program are all already members of on-campus Learning Communities in addition to FYE, so it is most natural to look for a comparison in students also participating in Learning Communities but not in FYE. So, in order to discern the effect that FYE may have on GPA, we will be comparing all members of FYE (treatment) with all other learning community participants (control). This will allow a fair comparison of GPA's and should give us a relatively unbiased look at the true association of FYE with student GPA.

Student Family Income Estimation

One factor of interest in our analysis is whether or not a student's income level changes the effect that FYE has on that student's GPA. However, not every Learning Community is required to report a student's income level, so there is quite a bit of missing data here. Fortunately, where we do not have income level, we do have zip codes. Zip codes can be fairly accurate proxies for average income estimates, and get more accurate with the inclusion of information on ethnicity. With this in mind, we turn our attention to the data available from the United States Census Bureau (US Census Bureau, 2019). We obtained a dataset from this website that contains 4 variables: zip code, the median household income estimate for that zip code, the median household income estimate for Hispanic or Latino identifying families, and the median household income estimate for White only identifying families. Asian or African American students will be estimated using the overall median household income estimate, because the data on these demographics in many Texas counties is quite sparse. Using this information, we can estimate student income level based on their zip code and ethnicity. For any student that was missing an income indicator and was not missing a zip code, we estimated their income with the median income estimate from their zip code by ethnicity. If their ethnicity was not White only or Hispanic or Latino, the median household income estimate was used to estimate their income. With this, we now have income estimates for all but 3 of the students in our dataset, and we can proceed with our analysis.

Data Analysis & Variable Selection

In the beginning of our modeling, we will attempt to identify which variables out of the 14 given provide the most predictive power for GPA. Although our best predictive power would come from having all variables in the model, it is best to find some optimal balance between predictive accuracy and model complexity. Generally, all variables will yield some strength towards a prediction, but only a handful will carry the bulk of the predictive power. In other words, most variables in a dataset are not very useful for making predictions, so we would like to simplify our model as much as possible by removing all of the extraneous variables.

Importantly, the kind of variable selection here is a bit different than textbook variable selection. While many basic regression problems involve quantitative predictors and desire the highest possible prediction accuracy, we are working with categorical data and want a model that is interpretable and realistic. Because we are trying to discern the particular association that FYE has with GPA, a model that is not readily interpretable will not be so useful. So, though it may be true that some particular subset of variables provides the optimal complexity-accuracy balance, we have the further constraint that the purpose of our model is not to predict GPA outright, but to associate GPA with a handful of variables that are sensible and yield an intuitive insight into the kinds of factors that strongly affect GPA.

A Potential Regression Model

To select which variables should be in our model, we went through many different permutations of subsets of variables to see which ones carried the strongest statistical significance. Then, to maintain the interpretability of our model as discussed above, we consulted with Dr. Suma Datta, a genetics expert with substantial knowledge of food insecurity and the kinds of things that would be expected to affect GPA. By her account, GPA is strongly affected by things such as ethnicity and income status, and very weakly affected by other demographic factors like gender. With this in mind, we went back to the models that we had produced in order to find the one that struck an optimal balance between predictive power, simplicity, and interpretability. Again, it is important to remember that although the textbook case of linear regression is interested solely in predictive power, the purpose of this model is not necessarily to predict GPA given some information on a student. Models that are built to optimize for predictive power are often blackboxes, and are therefore very difficult to deconstruct in terms of understanding exactly how much a particular factor is contributing to the prediction. Instead, we are interested in a model that best explains reality, and makes the direct effect that a particular factor has on GPA as transparent and relatable to the real world as possible. In the end, we have decided on the following set of seven indicator (yes/no) variables:

- Indicator for membership in the College of Engineering (x_1)
- Indicator for membership in the College of Science (x_2)
- Indicator for \$40,000 - \$60,000 income strata (x_3)

- Indicator for \$60,000 - \$80,000 income strata (x_4)
- Indicator for \$80,000 - \$100,000 income strata (x_5)
- Indicator for \$100,000+ income strata (x_6)
- Indicator for membership in FYE (x_7)

Again, these variables were selected via tests of statistical significance, measures of predictive power on GPA, and sensibility of interpretation. So, the following regression model will be used in the statistical analysis of our data:

$$GPA_i = \beta_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \epsilon_i,$$

where ϵ is a non-parametric error term with mean 0 and variance estimated by bootstrap resampling, and $i \in \{1,2,\ldots n = 436\}$.

Now, it is very important to clarify the interpretation of this worrisome equation. Recall that each of $x_1 \ldots x_7$ is a yes/no variable, so each will take value 0 if the student is not a member of this factor, or value 1 if the student is a member of this factor. The values of $\beta_0 \ldots \beta_7$ will be chosen through bootstrap linear regression, which selects the values of these coefficients that force our model to best fit the data. The interpretation of these values can best be described with a couple of examples: suppose we would like to predict the GPA of a student who is not a science or engineering, is in the lowest (0 - \$40,000) income strata, and is not in FYE. Then all of our $x_1 \ldots x_7$ variables will be equal to zero, and our GPA prediction would be

$$GPA = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0) + \beta_4(0) + \beta_5(0) + \beta_6(0) + \beta_7(0) = \beta_0$$

So, the value of β_0 is essentially a ‘baseline’ GPA to compare all other factors against, and this value will be the average GPA of all students who are not members of any of the selected factors.

Next, consider a student who is an engineering student, in the $60,000 - $80,000 income strata, and is a member of FYE. Our GPA prediction would now be

$$GPA = \beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(0) + \beta_4(1) + \beta_5(0) + \beta_6(0) + \beta_7(1) = \beta_0 + \beta_2 + \beta_4 + \beta_7$$

Now, our prediction is the sum of four components; one from our baseline and three from each of the factors that this student is a member of.

The idea here is that each of $\beta_1 \ldots \beta_7$ is interpreted as the average change to a students GPA due to membership in a particular factor. So, β_1 represents the average change in GPA that results from being a member of the College of Science, and so on. A positive value represents a positive change in GPA, and a negative value represents a negative change in GPA.

After all of this, we can say that our precise goal here is to determine the true value of β_7, or the average change in GPA that results from being a member of FYE. We would like two things of this value: statistical significance and positivity. Positivity is desirable because it indicates that the average effect that FYE has on a students GPA is beneficial rather than detrimental. Statistical significance is desirable because it indicates that the average effect that FYE has on a student's GPA is real, and not just an artifact of our dataset. If both of these things hold, then we have strong evidence from our dataset that FYE has an average positive association with student GPA.

If one or both of these things fails to hold, it will become important to explore possible reasons they do not hold, or to propose new ideas and strategies to help propel FYE forward in an effort to eradicate food insecurity on college campuses. This will be covered in the next section, after the implementation and analysis of the described regression model.

CHAPTER III

RESULTS

With all of our data in, we can now begin to answer our research questions. There are a variety of different conclusions that we could draw from what we have observed so far, and we will cover those in the next section. First, we must lay out all of our data, in many forms, from many angles. This will help us in understanding the true impact that FYE has had on our students, and give us clues on how to improve the program in the coming years. In the following subsections, we will explore our data; first with graphics, then with concrete statistical methods.

Summarizing our Data

First, we want to present a table that summarizes our data, to help the reader get an idea of the demographic breakdown of this set of students, how many are participating in FYE, and how many live in our control group. This is a very important point of realism for the rest of our analysis, because the statistical power of the conclusions that can be drawn here depend very heavily on the number of students in a particular group. So, if we want to make conclusions about the effect of FYE on high income students, for example, we will need to see a relatively large number of students (between 40 and 60, at least) within that category in order to have any hope of drawing significant conclusions about that particular demographic. Sample size always plays an integral role in all statistical analyses, so we will break these numbers down by different demographics in Table 1, shown below, to help the reader get a sense of the structure of our sample and an intuition for why some groups are particularly difficult to do analyses on.

Table 1. Summary of the number of students within various factors in our data

		FYE	Control	Total
Gender	**Female**	98	195	293
	Male	28	115	143
	Total	126	310	436
Ethnicity	**Hispanic/Latino**	110	206	316
	Not Hispanic/Latino	0	104	104
	White Only	7	0	7
	Black Only	4	0	4
	Asian Only	4	0	4
	Other	1	0	1
	Total	126	310	436
College	**Engineering**	21	77	98
	Science	16	28	44
	Other	89	205	294
	Total	126	310	436
First-Gen	**Yes**	121	250	371
	No	5	60	65
	Total	126	310	436
Income	**0 - 40,000**	108	172	280
	40,000 - 60,000	16	76	92
	60,000 - 80,000	0	38	38
	80,000 - 100,000	0	15	15
	100,000+	2	6	8
	N/A	0	3	3
	Total	126	310	436

Using Table 1 as a reference, we can see the sizes of various demographic breakdowns of our data, as well as how those demographic factors intersect with FYE. This table alone holds a lot of information as to why we selected a particular set of variables to build our model with, so it's worth explaining some of what we see here.

First, you'll notice that this table contains 5 rows labelled 'Total', and each of those rows indicates that we have a total of 126 students participating in FYE, 310 students not participating in FYE, and 436 total students in our dataset. These repeated rows serve as a good way to check that the numbers in the various demographic breakdowns actually add up to 126 and 310 respectively, telling us that our demographic breakdowns have not left any students out.

Now, take a look at the breakdowns by ethnicity and income. You'll notice that, in several cases, either the number of students in the control or the number of students in FYE within this category is zero (as in the White Only ethnic demographic, or the 80,000 - 100,000 income range, respectively). This is important because it tells us that it will be impossible to distinguish an effect that FYE may have on that particular demographic from our data. Now, because of this, we would like to separate those factors from FYE in our model, and we do this by letting all of ethnicity, income, and FYE be predictors in our model. If this were not done, we may end up with a model that gives a false impression of the true impact FYE has on GPA, because our estimate of FYE's impact would be inseparable from the fact that several demographics do not contain a control group or a test group at all. This could skew our results in the direction of how well the students in such demographics perform, without allowing us to differentiate between membership and non-membership in FYE (in the end, the inclusion of ethnicity turned out to have a statistically insignificant impact on GPA, so it was dropped from the model. The impact that FYE has on ethnicity will instead be analyzed in a different way, explained below).

Next, the demographic subsets of students that have a reasonable number of students within both the FYE group and the control group will provide a good basis of comparison for how FYE affects GPA. In order to tease this out, we will not include these factors (such as male vs female) in our model, but will rather fit the regression model described in the previous section for each of these different demographic categories. This will allow us to directly compare the impact that FYE has on underrepresented vs. non-underrepresented students, for example (here, underrepresented students are defined as students identifying as Hispanic/Latino or Black Only, this categorization serves as an amalgamated proxy for ethnicity). Any difference here will be seen not within one model, but between the same model fit on different subsets of our data.

We will soon provide a graphic as well as a host of examples to help you understand exactly what this means, but for now, we will start by looking at the preliminary statistical results from our data.

Initial Results

We will start our deep dive into the FYE data by presenting several graphs that will develop an expectation and intuition for the concrete results in the following subsection. Please note that not all observed difference between sample means will actually be statistically significant and the following figures are being presented only as a precursor to the statistical analysis, to aid in our understanding of the results. In figures 1, 2, and 3, given below, we present our data first broken up by membership in FYE, then further broken up by gender and then by low income status. These kinds of plots are quite useful in telling us what kinds of things to look for and explain in our deeper analysis.

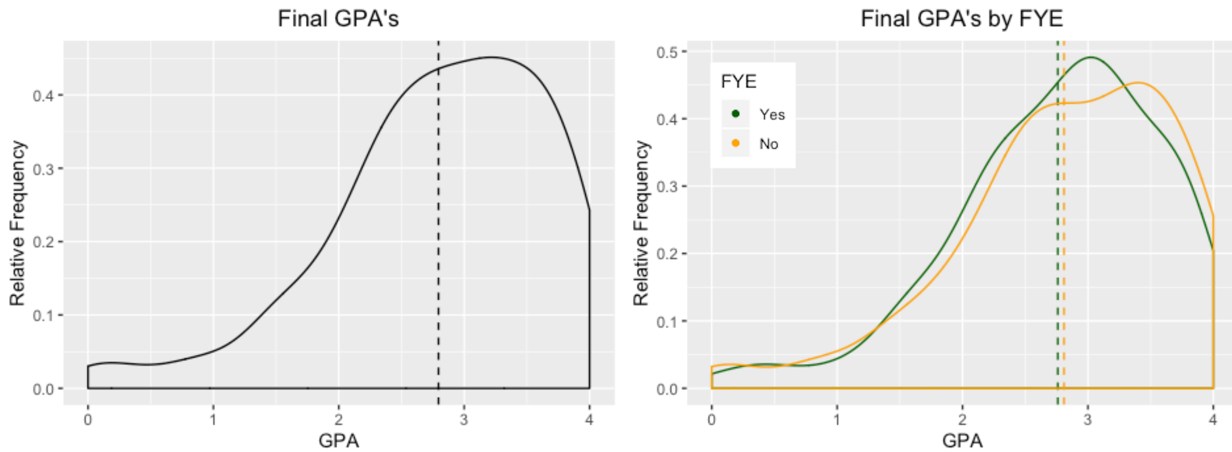


Figure 1. Overall GPA alongside GPA of FYE vs. non-FYE Students

First, we would like to get a bird's eye glimpse at the overall effect FYE may be having on our students. Figure 1 does exactly this; the plot on the left shows a *kernel density estimate* (read: smooth histogram) of the overall GPA *distribution* for our students. The kernel density estimate is a continuous analog of a histogram, so the x-axis represents all possible student GPA's (from 0 to 4), and the y-axis represents the frequency with which any particular GPA occurs in our data. The mean of the data, represented by a vertical dashed line, is just above 2.5. There is not much surprising about this plot; however, it is of interest to note that the FYE group has a slightly lower GPA average than the non-FYE group, though this difference does not yet look to be statistically significant.

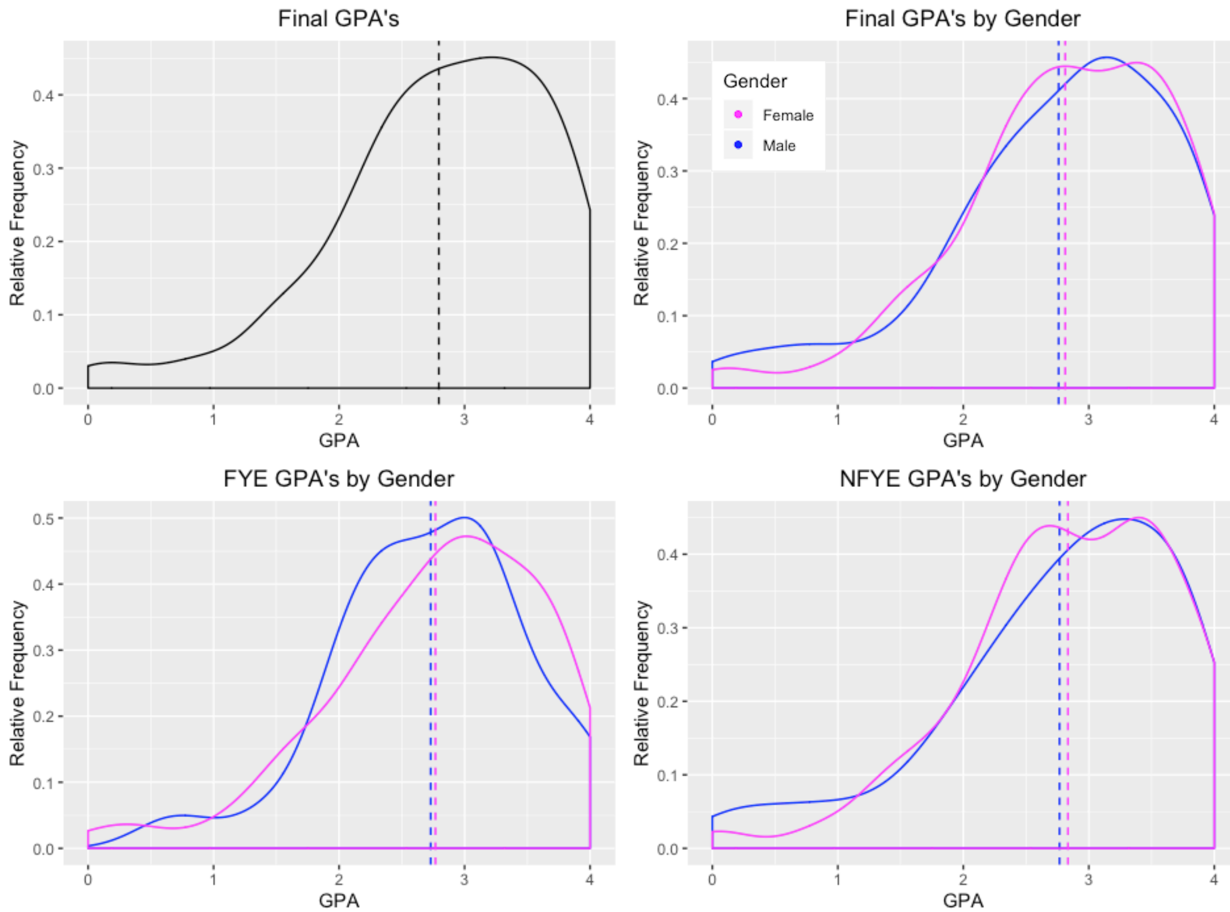


Figure 2. Overall GPA alongside GPA by gender, then factored by FYE

Figure 2 shows the same overall GPA distribution, now accompanied with plots factored by gender, where the blue plots belong to male students and the purple plots belong to female students. The top right plot only factors students by gender, whereas the bottom two plots are further broken down by participation in FYE. It is worth noting that the difference between male and female GPA's is almost non-existent, contrary to some research. This could be due to the fact that the students in our sample are not average college students, but Learning Community participants. These students (particularly females) may excel in their academic performance relative to others.

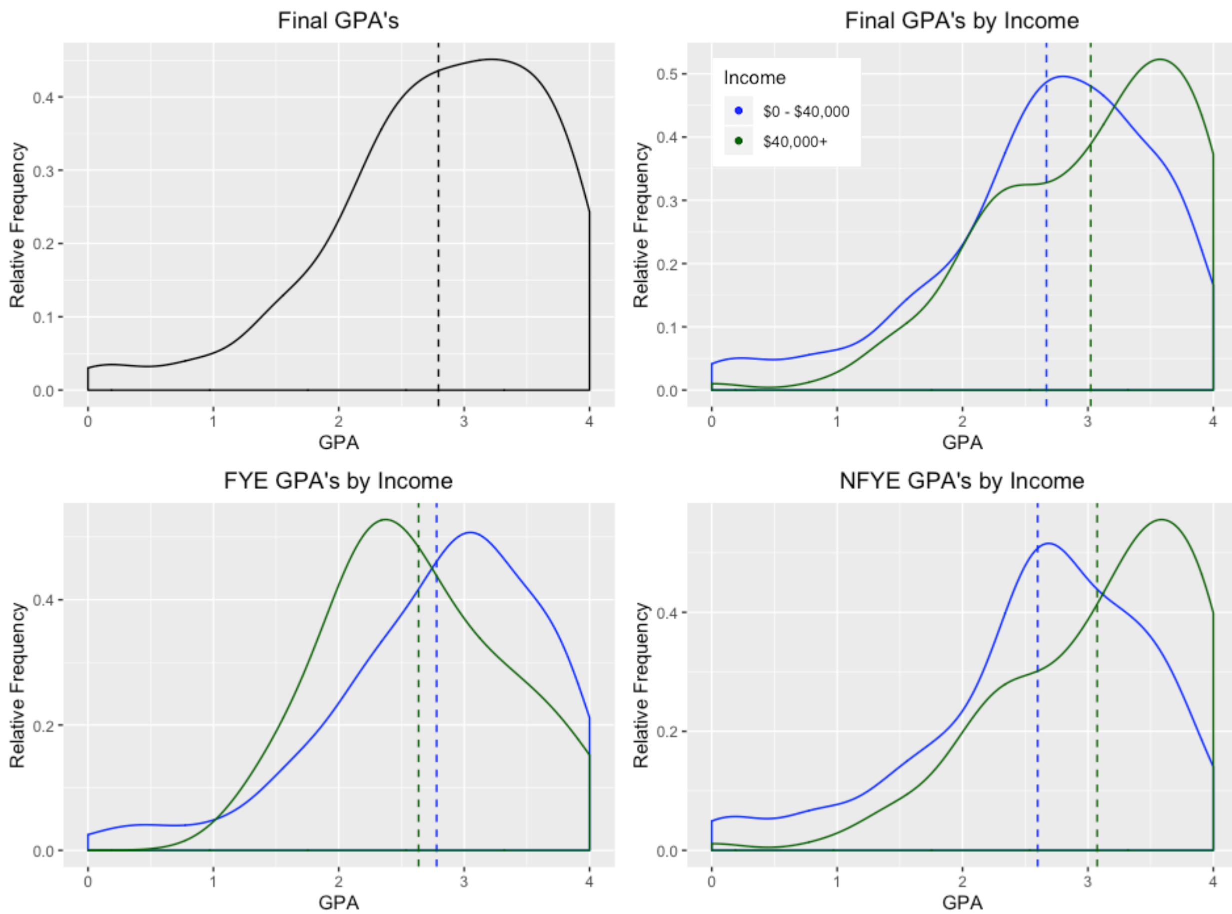


Figure 3. Overall GPA alongside GPA by income stratification, then factored by FYE

Arguably the most fascinating of these plots is Figure 3, shown above. As expected from previous research, the gap between low income students and non-low income students is significant, with low income students, represented by the blue distribution, trailing pretty far behind the others, represented with the green distribution. However, when we control for participation in FYE, we see something very unexpected: the low income students who are in FYE have a significantly higher average GPA than the non-low income FYE participants. This difference absolutely deserves further investigation, if it indeed turns out to be statistically significant.

We hope that these graphs will leave you with some kind of intuition for which factors affect GPA. In particular, keep these in mind through the next section, because they again help substantiate our decision to choose a particular subset of variables for our model.

Statistical Modeling and Analysis

Now that we have gone through the exploratory stage of data analysis, it is time to present the results in a more rigorous, scientific manner. First, recall our linear regression model used to estimate the average effect that college, income, and FYE indicators have on GPA:

$$GPA_i = \beta_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \epsilon_i,$$

where each of $\beta_1 \ldots \beta_7$ indicates the average change in GPA due to being a member of the respective group indicated by $x_1 \ldots x_7$ (refer to page 9). To develop this model using our data, we implemented the classic least squares regression technique to estimate the true values of $\beta_1 \ldots \beta_7$, which we will call $\hat{\beta}_1 \ldots \hat{\beta}_7$. It is important to bear in mind that these quantities are *estimates* of the true effects that these factors have on a general student population. Because we would like to extrapolate any conclusions we make to future cohorts of Learning Communities, or even similar programs at other universities, we will treat our dataset as a sample from a broader population and do the analysis accordingly.

The statistical significance of these estimates will be computed using the bootstrap resampling technique (Efron et al., 1994). This method is a very useful work-around for data where we cannot assume that the things we are trying to estimate are normally distributed.

In the Figure 4 presented below, we will present the findings of our analysis in what is hopefully a relatively uncomplicated way. We will take the time to explain exactly what it means, and how it answers our research questions.

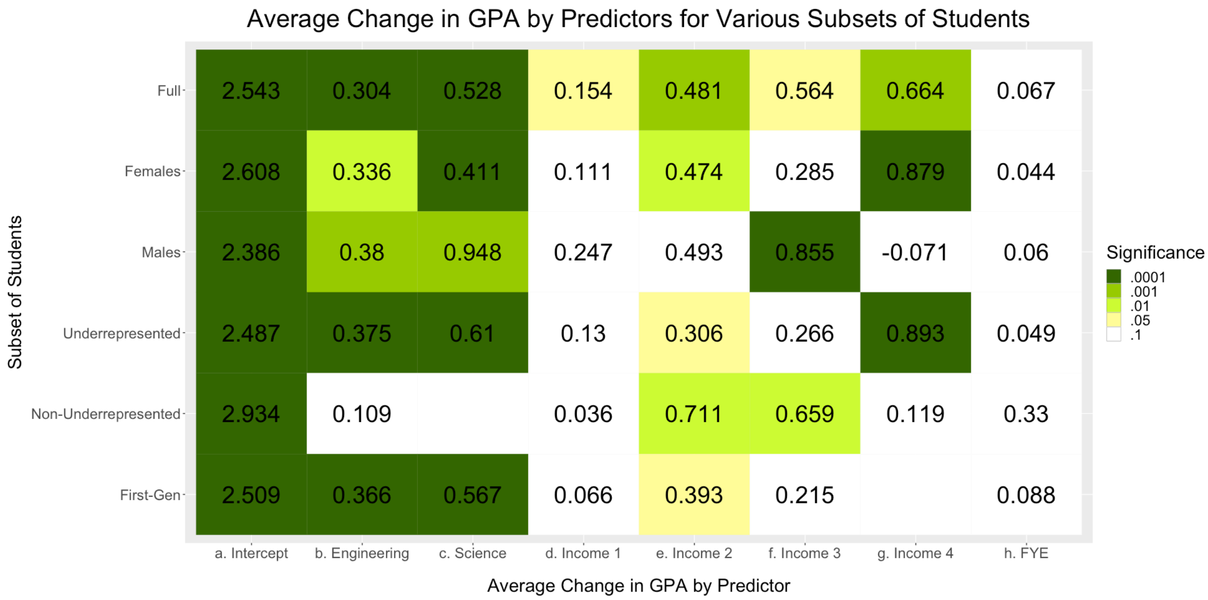


Figure 4. Results from a bootstrap linear regression on GPA for various subsets of students

Though this may seem a bit cumbersome, Figure 4 contains a lot of information about our data and the effects that particular factors have on GPA. First, observe that on the left of this graph, we have six different subsets of students, including the top row of the table, which are the results for our entire dataset. The motivation here is to examine the differences in the effect that each of the seven predictors has on different subsets of students. It is of interest to see the difference in how FYE affects males and females, for instance. Next, observe that on the bottom of this graph, we have an 'intercept' term followed by the seven predictors that we used to construct our model. Finally, observe that each cell is shaded to indicate its statistical significance, with a darker cell being more statistically significant that a lighter cell, and an unshaded cell being insignificant.

Now, we want to explain the precise interpretation of Figure 4, because it may not be immediately clear. This is essential to understanding our results, so we will carefully walk through a couple of examples to get a good idea of what this table really says.

Interpreting the Analysis

To best explain the contents of Figure 4, we will start with some examples.

Ex. 1. Suppose we want to predict the GPA of a student who is not in the College of Engineering or Science, is in the lowest (zeroth) income stratification, and is not in FYE. Additionally, suppose that we are not given any other demographic information about the student.

Because we are not given any information about which of the five demographic subsets this student belongs to, it is most reasonable to base our prediction off of the results for our entire dataset. Thus, we are going to consider only the first row of Figure 4, given below:

	a. Intercept	b. Engineering	c. Science	d. Income 1	e. Income 2	f. Income 3	g. Income 4	h. FYE
Full	2.543	0.304	0.528	0.154	0.481	0.564	0.664	0.067

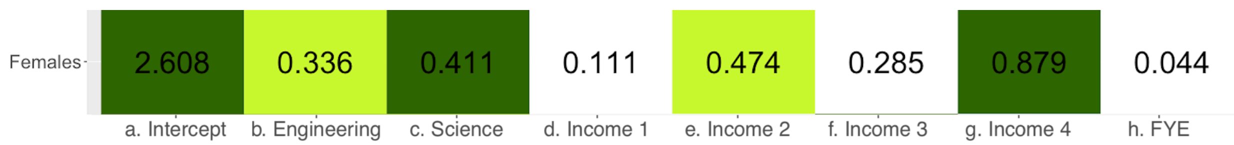
Now, we have assumed that this student is not a member of any of the seven categories (Engineering through FYE), so (as explained on page 10), we set each of these values to zero. Then, this table reduces to:

	a. Intercept
Full	2.543

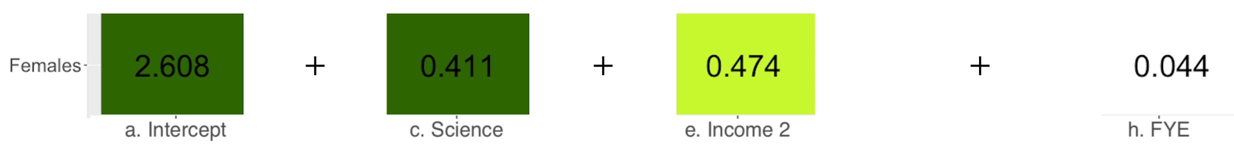
So, our predicted GPA for this student would be exactly 2.543 (95% PI [.913, 4.175]).

From this example, we see that this 'intercept' term roughly acts as a baseline GPA prediction for a student who is not a member of any of the seven categories that we used to build our model. In fact, this prediction is simply the average GPA of all students in our dataset who are not engineering or science students, in the lowest income stratification, and not in FYE.

Ex. 2. Next, suppose we want to predict the GPA of a female student who is in the College of Science, comes from the second income stratification, and is in FYE. Since this student is female, we will consider only the second row of the table:



We have assumed that this student is a science student, is in the second income stratification, and is in FYE. So, we will only include those particular variables in our prediction, and the rest will be set to zero, and we get:



So, our predicted GPA for this student would be 2.608 + .411 + .474 + .044 = 3.537 (95% PI [1.891, 5.188]). Notice that the final term is included in our prediction, even though the effect of that factor has been deemed statistically insignificant.

Hopefully you can see that each of the groups that this student is a member of contributes something to her predicted GPA. In that way, all columns of the table (aside from the intercept, discussed earlier) indicate the change that being a member of that group produces in predicted GPA. Positive values indicate that membership in a particular group has a positive impact on GPA, and negative values indicate a negative impact on GPA.

With these things in mind, let us turn our attention back to our research questions to see what information we can extract from Figure 4.

Findings and Their Interpretation

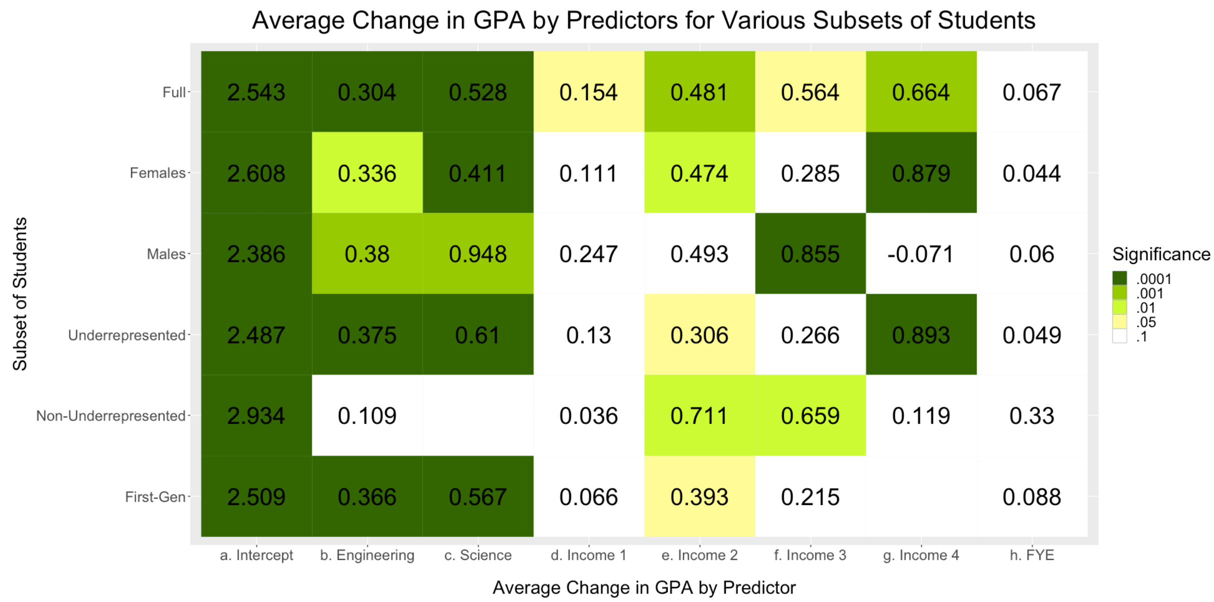


Figure 4. Repeated for ease of reference, having explained its interpretation

Recall that our aim is to determine whether the FYE program has a positive, significant association with the GPA's of the students involved. In Figure 4, the effect that FYE has on GPA is displayed in the rightmost column of the table.

From this, we can answer our research questions on positivity and significance. First, all of the FYE coefficients are positive; therefore, the students in our sample participating in FYE had a higher average GPA than the students that did not participate in FYE. Additionally, this positivity holds even when students are broken down into demographic subgroups. It looks like the impact of FYE on students considered 'non-underrepresented' (students identifying as strictly caucasian) may be higher than the rest, but a test on equality of regression coefficients shows this to be false (p = .318).

Second, none of the FYE coefficients are statistically significant. This means that, despite the positivity that we observe, we do not have the statistical evidence to conclude that FYE has a positive association with GPA on average. This could be due to a variety of reasons, such as sample size or true lack of impact, but the important take away is that our data do not provide us the justification to extrapolate FYE's observed positive impact on GPA to future cohorts of students at Texas A&M or otherwise. In a nutshell, FYE had a positive association with the GPA's of this particular group of students, but we cannot conclude that this impact will translate to any other group of students.

Finally, one of our primary research questions was not only whether or not FYE has a general effect on GPA, but whether it has a *differential* effect across different groups of students, say by ethnicity or gender. We can see from Figure 4 that the FYE coefficients are insignificant across all subsets of students, but this graphic doesn't exactly tell us whether there is a true difference between the effect of FYE on underrepresented (Hispanic/Latino or Black Only) students and the effect on all other students. If there were a true difference here, we could conclude that FYE affects minority ethnic groups differently from others, and this would be of substantial interest in the coming year. However, a statistical test on the difference between these coefficients reveals no significant difference. This means that the effect that FYE had on males was no different than females, and similarly for underrepresented ethnic groups vs. others.

Now, you may recall the figures 1, 2, and 3 presented on pages 16, 17, and 18. Our model tells us that FYE does not have a significant association with average GPA, nor with the average difference between male and female GPA's, so figures 1 and 2 provide nothing of interest. However, Figure 3 presented on page 18 looked quite remarkable, and we noted that we wanted to test this difference explicitly. We saw that the GPA gap between low-income students and all others seemed to change very drastically when participation in FYE was taken into account. We performed a hypothesis test on a difference in means, where one mean was the average difference in GPA between low-income and other students *not* in FYE, and another mean was the average difference in GPA between low-income and other students *in* FYE. As it turns out, this difference is statistically significant ($p = .012$) and tells us that FYE has a significant positive association with the difference in GPA between low-income students and all others. This difference was not revealed by our regression model because both FYE and income stratifications were predictor variables in the model. Now that we know this difference is significant, a future point of research will be to use a more complicated model for trying to capture this relationship exactly.

It may be of general interest to note other columns of Figure 4. In particular, there is a consensus that engineering and science disciplines add a GPA *penalty* to a student's GPA (Tomkin et al., 2016). However, our data show quite the opposite. In contrast to established literature, students in engineering and science in our dataset have a significantly higher GPA than the average student; as much as .948 GPA boost for males in the College of Science. These effects are also highly statistically significant, so they are at least noteworthy. Given that our dataset consists of students in Learning Communities, those communities may be particularly effective at equipping science and engineering students to do well academically. It could also be the case that the kind of students selected to participate in learning communities may be uniquely poised to do well in science and engineering. This may be an interesting point of future research

We also establish that family income has a substantial impact on student GPA; this is less surprising and has been demonstrated before (Betts, 1999).

This analysis could be made much more complex; there are ways to include the possibility of relationships within our predictors, techniques beyond bootstrap linear regression, many different tests to perform, and so on. These may be good starting points for future research on this program, and we will require that any future model explain the difference found in GPA for low-income and other students caused by FYE. However, in an effort to keep our analysis intuitive, explanatory, and efficient, these more complicated ideas were left out for the time being.

CHAPTER IV

CONCLUSION

Despite all of this work, the jury is still out on whether or not FYE truly has an impact on GPA, across different groups of students over time. This was almost anticipated from the beginning; as for any observational study, decisive results that clear us of any future work are not the expectation. Things will become more clear as more data rolls in, and we will be able to make data-informed statements next year that would be entirely unwarranted right now.

Discussion

Here, we want to take time to examine exactly what we see in our results, what we can conclude and what we cannot, and explore possible reasons for what we see in the data.

First, the general insignificance of the FYE coefficients deserves investigation. It may be the case that FYE really does have a positive impact on GPA, but that effect is rather small; on the order of a tenth of a letter grade. In fact, if we were able to extrapolate from our dataset to the general population of students, this is precisely what we would conclude. However, a smaller effect requires a large sample to make statistical tests powerful enough to detect it. At small to medium sample sizes, like we have, small effects can be very difficult to discern from random noise. As sample size increases, however, that random noise drops, and can eventually drop below the level of the small effect we're looking for, revealing it and allowing us to decisively conclude that it is real. Until then, such effects remain hidden.

Now, we must turn our attention to the relationship that has been established between FYE and GPA differentials across income strata. It is important to clarify that we *cannot* conclude that FYE has a significant impact on the GPA's of low-income students. However, we *can* conclude that membership in FYE has a significant positive association with the *difference* between the GPA's of low-income students and non-low-income students. In particular, FYE looks to close the GPA gap between low-income students and other students. This is quite remarkable, because if the effect of FYE on the students were non-existent, we would not be observing such a significant difference here. To be frank, FYE seems to have leveled the playing field for these students with respect to their income stratification, and that is incredible. It could very well be that the students who benefit most from FYE are the students that come from lower-income families and do not have access to the kinds of resources that other students have. For them, FYE looks to play an essential role in their academic well-being. This is incredibly encouraging, as it decisively shows that FYE has some non-zero effect on students, even though this effect is not necessarily on GPA directly. This is really the punchline result of this study, and future work here will require deep investigation into this surprising relationship.

Even with this remarkable result, it is still possible that most of FYE's impact on students is not on GPA, but on other aspects of a students life. We proposed that food, shelter, relationships, and goals are integral parts to being successful at a university, and it could be that most of the effect of FYE is not obvious, but is instead tied more deeply to relationships and goals. It may be that the programs implemented by FYE provide a unique social environment for the students involved that is difficult to quantify, or that the confidence in being able to cook your own food increases general well-being and physical health, but not necessarily GPA. The opportunity to build relationships with a host of peers who are in the same circumstances as yourself is a very necessary component to learning to be a college student. Freshman year is a time of great change and uncertainty, and programs like this could really be helping students ease their minds, if only just a little. The knowledge and ability to make your own food can act as incentives for setting new nutritional goals and striving to improve your general life skills. These things would be exceedingly difficult to measure, but nonetheless an incredibly profound impact of FYE. We can and will speculate on possible ways to uncover these effects, but for the moment, they principally cannot be discerned from our dataset.

As the final part of this work, we would like to give some suggestions to the FYE and / university administration, concerning ways to improve upon the program, enhance the experience for incoming freshman, and collect more detailed data in an effort to draw definitive conclusions about what role FYE is playing in the lives of our freshmen.

Suggestions to First Year Eats

In an effort to help FYE progress towards their goal of eradicating food insecurity on Texas A&M's campus, we are going to lay out a list of recommendations for the program. We hope that these will help guide the program towards success in the coming years.

1. Entry and Exit Surveys

Our stated research goal was to discover whether or not FYE has an impact on student GPA. However, GPA is a one-dimensional piece of information, and students are complex, high-dimensional human beings. This being the case, it may be wise to consider adding entry and exit surveys to the program, as a way of quantifying how the students perceive themselves and their environment, and how that might change as a function of membership in FYE. We imagine these would look something like 5-10 question surveys, where each question looks something like:

On a scale from 1 to 5, please rate your confidence in being self-sustainable on campus

On a scale from 1 to 5, please rate your overall mental and physical health

...

These kinds of questions put an emphasis not on the student's academic performance, but on their self-esteem and perceived adaptation to the college environment. There are many, many things that can affect GPA, but FYE targets a very particular facet of a student's life and it would be easiest to measure its effects through questions that directly target a student's relationship to food. The best way to produce or deliver this kind of survey lives within a psychology department, so it would be remiss not to consult a psychologist here. We believe that this could really illuminate the kind of effects that FYE is having on freshmen, and provide our statisticians with more data to draw conclusions from.

2. Food Stocking Options

Next, we think that the students should have some say in exactly what kind of food is available in the open pantry. This is where the applet that we discussed on page 6 may be useful, as it may inform FYE as to which foods are and are not being consumed. Based off of this data, it may be wise to poll participating students, asking whether or not a particular item should be taken off the shelves or stocked heavy-handedly. We know that such polls can be answered with a bias towards what the subjects think the administration wants to hear, so a poll could be constructed to reveal a students food preferences indirectly, such as asking what they enjoy most at dining halls. We believe this interactivity would increase retention, participation, and overall effectiveness of the program.

3. Tracking Students Post Participation

Since all of the students in this program live in dorms on campus, there may be a certain standard of living and eating that is shared among students in those circumstances. However, as students move off campus to apartments or houses, the need to be self reliant will drastically increase, and the variability in how well/consistently students feed themselves will also drastically increase. This means that it is possible that the effect that membership in FYE has is minimal as a freshman, but much more pronounced as a sophomore, and so on. While we do not currently have the scope to keep track of this data, it may be worth considering an *entirely optional, opt out at anytime* way for the FYE staff to keep collecting data on students that have passed through the program, particularly GPA data and potentially data from the surveys mentioned earlier. Again, it would be of upmost importance to make this opt in, and allow the students to opt out at a moments notice, because they should be the gatekeepers of their data and should not be required to continue to give up their data for even a second longer than they want to. It would also be essential to be very clear with them on exactly what kind of data you will continue to collect and what to expect from it. This kind of continued tracking would open the possibility to seeing the long term effects of FYE that would remain entirely hidden without it.

If these suggestions were implemented into the FYE program, we could make improvements and determine outcomes at a much faster rate, compared to not having them at all. If these things could be implemented into all Learning Communities, that would be even better.

It may not be an easy task, but we absolutely believe that it would, in the end, make the programs much more effective and enjoyable for all of the students involved.

Final Words

After all of this, we cannot wait to see where FYE goes next. The people running the program are its lifeblood, and they want absolutely nothing but the best for the students involved. The ethos of this program is wonderful, and the goals that they have set are enormous and noble. It has been an immense pleasure to contribute to the program in this way, and we hope to see this initial groundwork used as a platform to propel the program into the future. All statistics aside; if this program makes a real difference in the life of even one student, then all the time and energy spent on this project is worth it. If we could see more selfless initiatives like this, this world would become a much better place for us all.

REFERENCES

Alaimo, K, et al. “Food Insufficiency and American School-Aged Children's Cognitive, Academic, and Psychosocial Development.” Pediatrics, U.S. National Library of Medicine, July 2001

Betts, J, Morell. “The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects.” The Journal of Human Resources, vol. 34, no. 2, 1999

Bruening, Meredith, et al. “The Struggle Is Real: A Systematic Review of Food Insecurity on Postsecondary Education Campuses.” Arizona State University, Elsevier USA, 2017

Efron, Bradley, and Robert Tibshirani. An Introduction to the Bootstrap. New York: Chapman & Hall, 1994

Morris, Mary, et al. “The Prevalence of Food Security and Insecurity Among Illinois University Students.” Journal of Nutrition Education and Behavior, U.S. National Library of Medicine, June 2016

Tomkin, J, et al. "A methodological refinement for studying the STEM grade-point penalty," 2016 IEEE Frontiers in Education Conference (FIE), Erie, PA, USA, 2016

United States Census Bureau, (2018). “Median Income in the Past 12 Months.” Retrieved from https://data.census.gov/cedsci/table?t=Income%20%28Households,%20Families,%20Individuals%29&tid=ACSST1Y2018.S1903&hidePreview=false&vintage=2018

Woerden, Irene, et al. “Food Insecurity Negatively Impacts Academic Performance.” Wiley Online Library, John Wiley & Sons, Ltd, 26 Nov 2018