I. Scholarly Article Report

The way this world functions is very much centered around interaction, attachment and interdependency among all forms of living beings and social constructs. In an effort to model those relationships, mathematicians have introduced the concept of covariance, which essentially measures the extent to which two variables vary in tandem with one another. The sign of the covariance reflects whether the two variables exhibit similar or opposite trends in behaviors ("Covariance," 2021). In this section we report on the application of covariance in the field of behavioral science, drawing on the study "Ethnic Differences Among Correlates of Physical Activity in Women" (Mâsse & Anderson, 2003).

The study set out to explore the variations on five different constructs in relation to physical activities of two minority groups, which specifically are the value of physical activity, normative modeling, perceived barriers, outcome expectations, and self-efficacy . Other socio-economic indicators like education and income were also accounted for. The sample included 260 female residents of the greater Houston area, aged between 40 and 70. The project employed a mixed-method approach which combines survey-based and interview-based research. With the five correlates detailed above as dependent variables and all possible correlations among ethnicity, education and income as independent variables, the researchers performed a multivariate analysis of covariance (MANCOVA), followed by univariate analysis of variance (ANOVA), to address the key research question: "How do ethnic differences influence physical activities in women?". Specifically, the correlations, means, and standard deviations for the five correlates of physical activity are summarized in the table below:

	1	2	3	4	5	African-American Women Mean (SD)	Hispanic Women Mean (SD)
1. Normative modeling	1.00					3.51 (0.58)	3.64 (0.72)
2. Beliefs	0.27*	1.00				4.50 (0.58)	4.61 (0.58)
3. Outcome expectation	0.36*	0.63*	1.00			4.26 (0.55)	4.43 (0.63)
4. Perceived barriers	−0.21*	−0.29*	−0.41*	1.00		2.25 (0.63)	2.15 (0.68)
5. Self-efficacy	0.27*	0.37*	0.46*	−0.53*	1.00	3.53 (0.81)	3.63 (0.82)

† Means are on a scale from 1 to 5, with 5 representing high norms, beliefs, outcome expectation, perceived barriers, and self-efficacy.
* Significant at $p < .01$.

(Mâsse & Anderson, 2003, 359)

According to this table, the association between ethnicities, education, income and five correlates of physical activity was confirmed.

II. Additional Data Analysis

Introduction

In this section, we aim to explore the correlation between education and four factors, specifically race, gender, income and employment status, using the same application of ANOVA as the literature discussed above together with the Chi-squared Test. It is commonly believed that one's educational level is determined by the social and economic capital they possess. However, a growing scholarship points out that other immutable demographic factors, including race and gender, aslo play an integral role in one's academic pursuit. Education has emerged as not merely a vector for social mobility, but also as a tool for which the elite preserve their wealth, so we hope to paint a better picture of whether the U.S. is moving towards a more equitable society. Ultimately, we hope to shed light on how the institutionalized inequalities across different spectrums serve to further perpetuate the differential privileges when it comes to education, so that we can better tackle them in our striving towards a fairer society.

To accomplish this goal, we are going to use the National Health and Nutrition Examination Survey from 2009-2012. This dataset consists of 76 different variables, categorized into demographics data, dietary data, examination data, and laboratory data. The portion we are focused on is the demographics data, which contains four variables we want to analyze. The education variable has five different possible values: eighth grade, ninth through eleventh grade, high school, some college, or college grad, with the latter denoting the highest level of education that the individuals who participated in the survey held. As a followup question, we want to ask: how does the level of education relate to race, gender, wealth, employment status?

We will analyze the relationships among the variables by conducting hypothesis testing on how each of the variables relates to education. Specifically, we will need to find p-values, linear regression models, and correlation coefficients to determine the significance of the relationship holistically. Interpreting the output variables individually, we hope to create a multiple linear regression line to demonstrate how all chosen factors together contribute to an individual's education level.

Ultimately, we are hoping that our society as a whole can get a better understanding of the barriers facing marginalized groups in their pursuit of education, which will in turn affect well-informed policy-making processes.

Analyzing Data

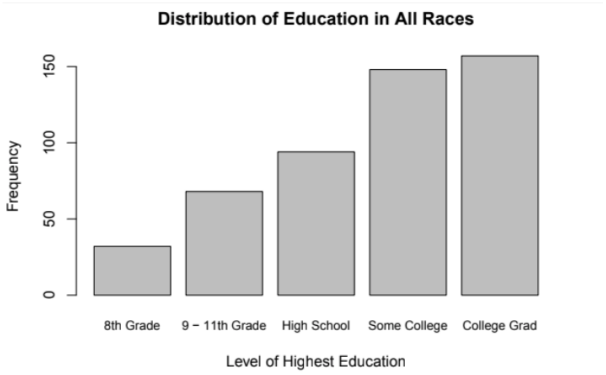
Before conducting any tests, we took a sample of 500 adults from the National Health and Nutrition Examination Survey, as younger individuals would not have completed their education. Also, we wanted to be able to generalize our findings to the population without having to use the numbers from every single person.

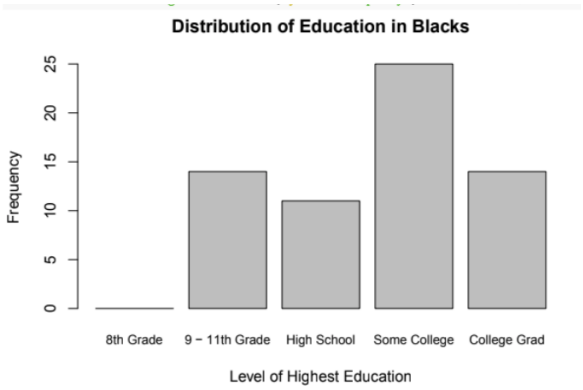
The first test conducted with the sample we took from the National Health and Nutrition Examination Survey was the distribution of education based on race. The education was split into five groups based on when people's highest level of education: up to 8th grade, 9-11th grade, high school, some college, and college grad. The data did not specify beyond college. For this specific test, it was decided that a Chi-squared test of independence would be the most effective for determining whether race was a factor in the level of education achieved. The Chi-squared *Independence*, n.d.). There are three conditions to this test: data values from a random simple sample from the population of interest, two categorical variables, and at least five expected values for the two combined variables. While the last case was not achieved in the category of blacks with 8th grade education and Mexicans with some college or college graduate education, they were more indicative of the greater trends found and while noted, were ignored. In this case, we also plotted the results in a bar graph for each in order to visualize the differences. The graph titled "Distribution of Education In All Races" is also our established baseline of education level in the sample. As shown, there are more people with each higher respective education level: there are 32 at the 8th grade level, 68 at the 9th-11th grade level, 94 at the high school graduate level, 148 at the some college level, and 157 at the college graduate level, for a total of 499 in the sample. One sample response described their education as NA, so they were omitted from the sample.

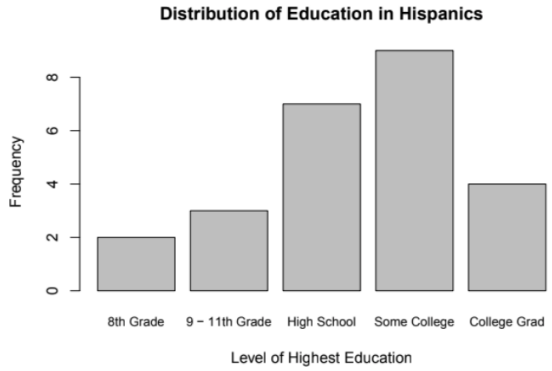
From the plotted bar graphs, it is relatively obvious that there seems to be some relationship between race and education level, so it was predicted that there would be a very small p-value. The Chi-square testing confirmed the result: it was smaller than $2.2e^{16}$, which is significantly lower than the chosen p-value of .05. The test also gives a statistic that is a single number corresponding to the amount of difference between the observed results and the expected results if the hypothesis holds true; that is, if there was no relationship between the two categorical variables. In this case, the Chi-square statistic was 167.28, conveying an extreme difference between a no-relationship data set and the one that was sampled. Another important variable that influenced the result was the degrees of freedom, which was 16 here. It is an estimate of the number of independent pieces of information that went into the estimate, and is normally calculated by [(number of categories) - 1]. In the case of independence, the degrees of freedom variable equals [(number of rows - 1) * (number of columns - 1)].

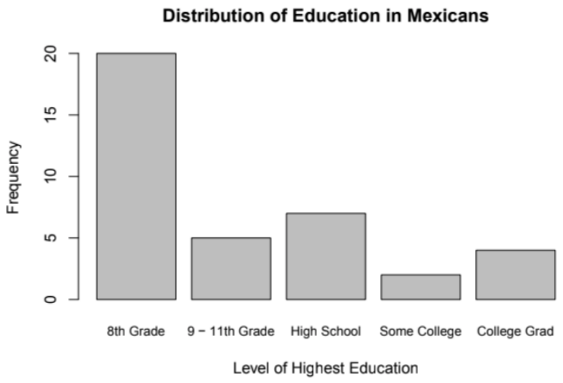
More specifically, it is clear that the majority of whites are college graduates, whereas the majority of blacks and hispanics only have some college education, and the majority of

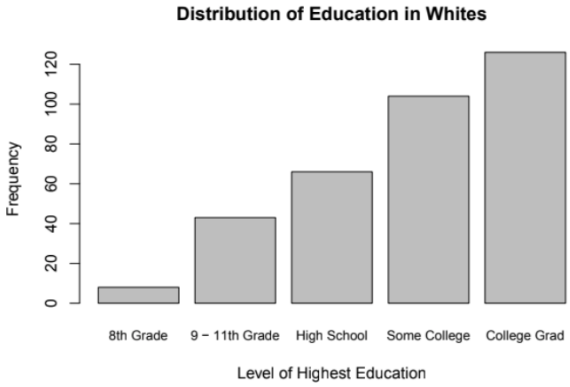
Mexicans have less than high school education. The data explains the vast differences in education level and consequently wealth in today's society based on race.





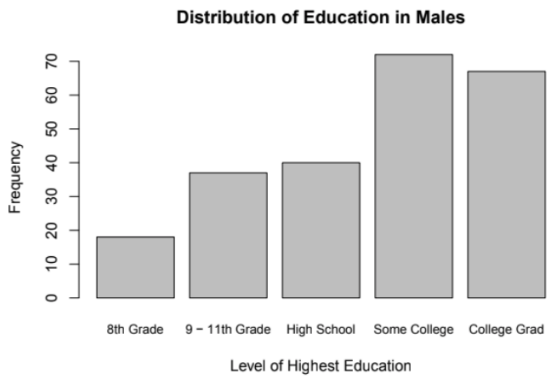


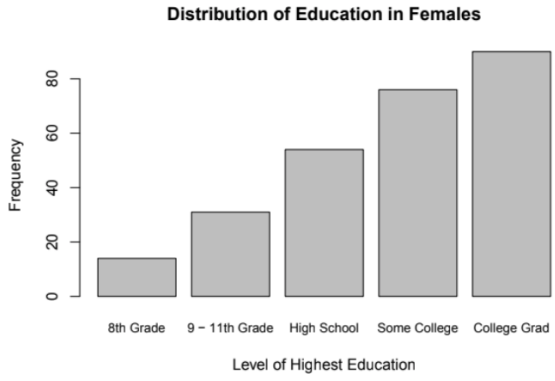




The second test conducted was the distribution of education between males and females. Even though females have always been discriminated against in the classroom and in education, as females were only allowed in the classroom of colleges starting the 1900s, modern data suggests that women even outnumber men in classrooms today in higher education (depending on concentration). Thus the test was conducted to see if females still trailed males in terms of education levels. A Chi-squared test of independence was calculated. The Chi-square statistic was 4.6843, the degrees of freedom was 4, and the p-value was 0.3213, compared to an alpha

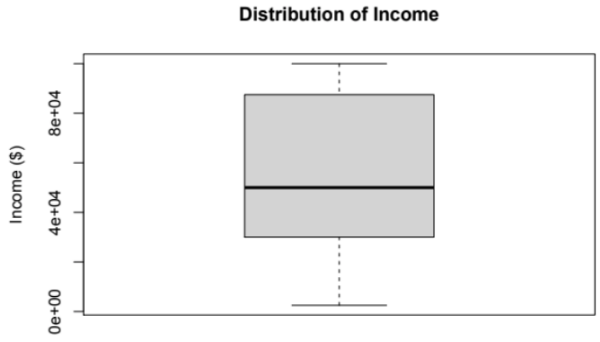
level of 0.05. Therefore the sampled data did not show any relationship between gender and education. This conclusion made sense given the strides that have been made in the classroom and higher education regarding gender equality. It is important to note that the sample only offered male and female options in the gender category, but this oversimplifies many of the complexities of people who identify as non-binary. This simplistic survey categorization is representative of a larger failure to fully consider the realities of minorities in different categories, and an important consideration that should be addressed and reflected upon.

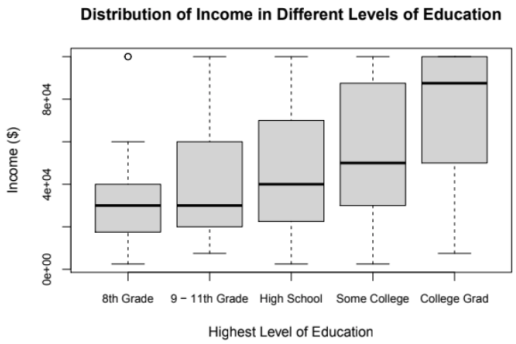




The third test conducted was to determine the relationship between income and education. Although it seems obvious that income is correlated to a higher education, there is still information to be gained from studying the differences. In this case, a scatter plot was used as opposed to bar graphs in order to better compare the differences in median income, but also the 25th and 75th percentiles, and the maximums and minimums to go further in depth because income distribution is also important: for example, if the median is closer to the maximum and 75th percentile compared to the minimum and 25th percentile, we know that there is a more equitable distribution of money, which points to less income equality. In this case, the results of the entire sample, seen in the first graph titled “Distribution of Income,” pointed to a median of $50,000, the minimum was $2500, first quartile of $30,000, mean of $57,849, third quartile of $87,500, and maximum of $100,000. Instead of completing a Chi-square test of independence, pairwise t-test (with pooled standard deviation) was used. Essentially, the test would compare each education level to every other education level to determine whether the difference in education could reasonably predict a difference in income in this sample. The table below details the p-values found for each pairwise t-test. It is interesting to note that there was no statistical importance in the p-value between three pairs, 8th grade to 9th-11th grade, 8th grade to high school, and high school to 9th-11th grade. Thus it can be inferred that there is no statistical difference in income between until at least some college is completed. However, this does not mean that it is recommended to give up before 8th grade. Furthermore, the p-value between 8th

grade to high school is 0.05572, which is very close to the chosen alpha level of .05, so more analysis is needed for a more definitive answer, perhaps with different samples. The p-values then got increasingly large, suggesting that it is very important to complete at least some college to see an increase in income. The lowest p-values were between college grads and other levels: even the p-value between some college and college grad was significantly under 0.05 ($5.9e^{-7}$). Consequently, it can be inferred that graduating college makes a big difference from just completing some of college.

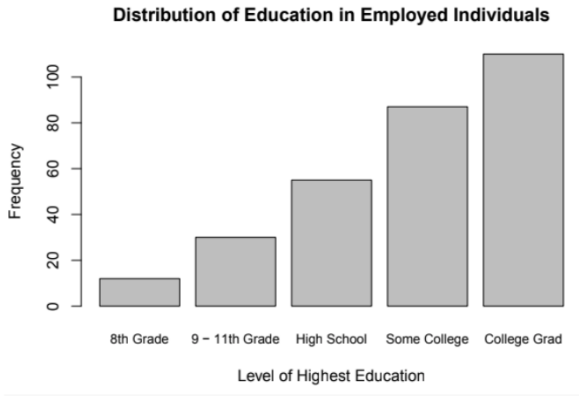


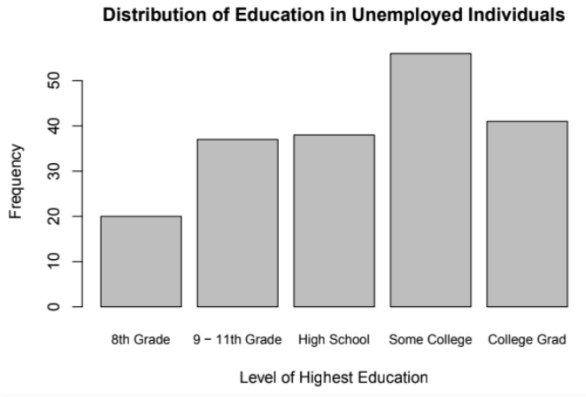



```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  nhanes$HHIncomeMid and nhanes$Education
##
##               8th Grade 9 - 11th Grade High School Some College
## 9 - 11th Grade 0.12296   -             -           -
## High School    0.05572   0.53319       -           -
## Some College   0.00029   0.03522       0.06117     -
## College Grad   1.7e-11   1.2e-10       3.0e-11     5.9e-07
##
## P value adjustment method: holm
```

The fourth test was completed to determine the statistical differences in the education of employed and unemployed individuals. The test relates to job security and whether people are more likely to be unemployed because they do not have the indicators that employers look for (such as a college degree). A more complex point of discussion that goes beyond the scope of this project is that in recent years, the most wealthy have been making much more of their income off capital gains instead of labor; they are no longer working to make most of their income, but making it off of their financial holdings such as real estate, stocks, and art, so it would seem that less of the most wealthy would be employed. However, this test focused on the difference between employed and unemployed individuals with a Chi-square test of independence. The Chi-square statistic was 27.223, the degrees of freedom was 8, and then p-value was .000647, suggesting that there was a statistical difference between the two categories of employed and unemployed. This conclusion is further backed by the plotted bar graphs. Between the distribution of education in employed and unemployed individuals, it is evident that the most drastic change occurs in terms of increases for the 9th-11th grade and some college categories. There is a less drastic increase for 8th grade and high school, and a decrease for

college graduates. These differences make sense in that employers do look for a degree as a signal of competency, and completing some more education only to drop out presumably seems like a signal of incompetency to employers. Job security is strongest for college graduates.





III. New Mathematical Concept

In this section, we review how we made use of the Chi-squared Test and ANOVA for our project. Chi-squared test for association is used when we have categorical data as the response variable with multiple groups as the explanatory variables. It allows us to see if there is a significant association between the response variable and the status in a certain group. Specifically, we employed the following formula:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

with three conditions, specifically values from sample, two categorical variables, five expected values for each entry. We set the Alpha level at .05. To make the process easier, we first made a table out of the data sampled, and then proceeded to apply the formula with expected values as independent variables.

On the other hand, ANOVA testing is used when we have numerical data as the response variable and we want to test its association to the explanatory groups. An ANOVA test is defined to be a method for finding out whether findings/results are statistically significant (*ANOVA Test*, n.d.). In all of our testing, the explanatory groups were the different levels of highest education reported by the surveyees. Since we mainly drew on one independent variable in our separate testings, only one-way ANOVA was applicable.

Depending on what type of data we had, whether it was categorical or numerical, we were able to switch back and forth between the chi-squared test and ANOVA test. Therefore, we were able to test whether the association that race, gender, wealth, and employment status had on the level of education was significant or not.

IV. Conclusion

Although none of the results in regards to education and other societal factors were shocking, the project gave important insight towards the specific relationships that existed and shifted based on small changes. For example, race plays a larger role than gender in the level of education one receives. It was also found that completion of an educational level is important for job security, most likely during more difficult times when unemployment is high. While the project was based on one sample and more research would need to be conducted in order to more definitively come to these conclusions, a smaller project always helps direct larger projects in terms of the questions that should be asked or further explored. One question that would be interesting to look into is the influence of implied discrimination in education and how that affects one's experience. Rather than just looking into race and how others' perception affects one's education, looking into how the individual is affected and their consequent educational decisions differ could be immensely important. An interesting part two of the project could be to relate educational level and health - people who are more educated generally have more resources, both financially and time-wise, to dedicate towards improving physical and mental health. Seeing these results could encourage more health and educational policy changes in the future.

References

ANOVA Test: Definition, Types, Examples. (n.d.). Statistics How To. Retrieved May 14, 2021, from https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/

Chi-Square Test of Independence. (n.d.). Retrieved May 10, 2021, from https://www.jmp.com/en_ca/statistics-knowledge-portal/chi-square-test/chi-square-test-of-independence.html

https://en.wikipedia.org/w/index.php?title=Covariance&oldid=1012044200

Covariance. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Covariance&oldid=1012044200

Mâsse, L. C., & Anderson, C. B. (2003). Ethnic Differences among Correlates of Physical Activity in Women. *American Journal of Health Promotion*, *17*(6), 357–360. https://doi.org/10.4278/0890-1171-17.6.357