Data Final Project

Requirements

1. Choose a topic you are interested in researching. Develop an argument related to your topic that you will attempt to gather data on and test.
2. Find at least one dataset related to your topic that you can analyze. Below, I have provided direct sites where you can find data and explanations of how to find datasets online.
3. Using data from the dataset you acquired, conduct at least one statistical test that we learned this semester. You will need to determine which test is appropriate based on the questions you are interested in answering and the type of data you have access to. Click here to find a summary of the different statistical tests with brief descriptions of what they test and a flowchart to help guide you to the correct test.
4. Complete the written portion of the project following the instructions in below.
5. Create a presentation using PowerPoint, google slides, etc. that shares the highlights of EACH part of the written portion of the project. The presentation should be visually appealing and concise. You should NOT have paragraphs written out on your slides. Instead, you should create slides that resemble the slides I use for my lectures. Although you do not need to include every detail from every section of your paper, you should include the important content from each section. You will ONLY be creating the slides and do NOT have to deliver the presentation.

Paper

This will be the written portion of this projected mentioned in requirement 4 above. Make sure to include each part of each section described below. Each section of the paper has an example provided for you below the description. There is not a specific length requirement for the paper. However, you should focus on including all of the necessary information. Finally, additional research beyond the dataset you use is not required, it is likely necessary. You will notice that in my examples I had to cite three different sources to support my claims. Be sure to provide a source citation when making a claim. Your sources must be cited in-text and in a summary at the end of your paper using a citation format of your choice.

Introduction to Topic/Project

In this section, you will be introducing the class to your project. You will need detail the following elements:

1. Topic Area:
    a. What specifically is your project addressing and trying to research? *For example:*
    You could be interested in demographic differences in academic success rates during the pandemic.
    b. You will start to build the argument you are trying to make. *For example:*
    You could explain that you believe students with a higher socioeconomic status had higher academic success during the pandemic.
    c. As part of your argument, you should address some explanations for this prediction. *For example:*
    Students with a higher socioeconomic status have access to better resources like internet and technology.
    d. You must cite sources to back up these claims. *For example:*
    According to Anderson 2020, students of lower socioeconomic status experienced larger financial deficits than those of higher socioeconomic status which not only limited their access to

resources and materials to be successful in school but also negatively impacted their physical and mental health.

2. Project Details:
    a. Provide formal null and alternative hypotheses for your study. You must have at least 1 null hypothesis and one alternative hypothesis. You may include variations of these hypotheses in the Statistical Analysis section based in the statistical test you choose to conduct but these hypotheses should give the reader a general idea of what your predictions are. *For example:*
        i. *H_0: Academic success for non-white students during the covid-19 pandemic was the same as the academic success for all students during the pandemic.*
        ii. *H_A: Academic success for non-white students during the covid-19 pandemic was the lower than the academic success for all students during the pandemic.*
    b. Describe the sample based on the data you collected and the population to the extent of accessible information.
        i. Sampling Method: In this section, you will need to explain where you accessed the data from, generally what data is available, which data you accessed, and why you selected that data. *For example:*
        I chose to analyze data from El Camino College because the data is easily accessible, represents the area I live, and has a diverse range of demographics. I accessed data using the El Camino College Institutional Research database. The database reflects achievement and success rates for each term using two levels of controls. The first level of controls is the college level which includes the following categories: Division, Department, Course, and Instruction Method. The second level of control is Student Demographics which includes the following categories: Ethnicity, Gender, Age, First Time to ECC. I chose to analyze data from the Fall of 2020 and the Fall of 2019. Although the pandemic began during the Spring of 2020, the Fall of 2020 should provide a better representation of my research questions for three reasons. First, the Fall of 2020 was the first regular term where the pandemic was occurring for the entire term as the pandemic was only occurring for approximately half of Spring of 2020. Second, as noted in the data itself, the data from the Spring of 2020 is heavily skewed because many students were granted emergency withdrawal. Finally, the time period during the Fall of 2020 better illustrates the economic effects. Despite higher unemployment rates during the beginning of the pandemic, the more dire economic impacts became more apparent during the time of the Fall of 2020 term. According to an analysis by the Aspen Institute (2020), public debt grew significantly over the summer and by the end of the summer, one in five renters were at risk of eviction by September 2020. Additionally, I will compare the Fall of 2020 to the Fall of 2019 because terms during the same time period are more comparable than terms of different time periods. A summary of the data can be found here:
        https://app.powerbi.com/view?r=eyJrIjoiNjFiMjkyYzEtNDMwOS00N2E5LTk1YjItNGU2MjMwZTc1NTUwIiwidCI6IjE2YTJhYzEzLTk5YmQtNDA1ZC05YWI2LWIxZmU2YjhkNWJhNiIsImMiOjZ9&pageName=ReportSection
        ii. Sample: In this section, you should provide relevant details about your sample that are available such as demographic information, sample size, etc. *For example:*
        My sample for the Fall of 2020 includes 55, 328 students and my sample for the Fall of 2019 includes 63,664. No College Level Controls were selected. All Student Demographic Controls were left unselected except for ethnicity. With no option to select students based on socioeconomic status, ethnicity is the next best option to measure socioeconomic status. According to the American Psychological Association (2020), those in the non-white demographic tend to have lower socioeconomic status than those in the

white demographic. As such, I will compare data from student of all ethnicities to the data of students with non-white ethnicities. For the Fall of 2020, 5,587 students were white and 49,741 were non-white. Specific breakdowns of size by ethnicity are as follows: African American 7,194, American Indian/Alaska Native 118, Asian 6,456, Latino 26,750, Pacific Islander 282, Two or More Races 1,419, Unknown or Decline 7,522, and White 5,587. For the Fall of 2019, 6,261 students were white, and 57,403 students were non-white. Specific breakdowns of size by ethnicity are as follows: African American 7,890, American Indian/Alaska Native 120, Asian 7,618, Latino 29,026, Pacific Islander 292, Two or More Races 2,397, Unknown or Decline 10,060, and White 6,261.

iii. Population: In this section, you must describe the population you are pulling your sample from. Additionally, you should provide relevant details about the population you are pulling your sample from that are available such as demographic information, population size, etc. *For example:*
My sample is pulled from the overall population of community college students attending a public institution in the United States. According to Duffin (2021), the population sizes for 2020 and 2019 were 5,584,000 and 5,398,100 respectively. Unlike the sample at El Camino College which has significantly more non-white students than white students (approximately 10% of students are white), the majority of students enrolled in a public community college in the United States are white (approximately 47% are white).

Statistical Analysis

In this section, you should describe, compute, and explain the statistical test(s) you conducted.

1. Variables: Describe your independent and dependent variables. *For Example:*
The independent variable is the ethnicity of a student. The dependent variables are Course Success, completing a course with a grade of A, B, C, or pass, and Course Completion, completing the course without Withdrawing.
2. Descriptive Statistics: Calculate and provide the descriptive statistics for your sample and population (if available). *For example:*
I will first provide the descriptive statistics of my samples, starting with the Fall of 2020. Course Success Rates based on demographic are as follows: African American 56.7%, American Indian/Alaska Native 67.8%, Asian 79.1%, Latino 63.8%, Pacific Islander 53.2%, Two or More Races 71.5 %, Unknown or Decline 69.5%, White 76.6%. The overall Course Success rate is 66.9%. Completion Rates based on demographic are as follows: African American 76.1 %, American Indian/Alaska Native 82.2%, Asian 86.8%, Latino 80.3%, Pacific Islander 69.5%, Two or More Races 84%, Unknown or Decline 85.1%, White 85.4%. The overall Course Completion rate is 81.7%. For the Fall of 2019, Course Success Rates based on demographic are as follows: African American 59.9%, American Indian/Alaska Native 60%, Asian 78.9%, Latino 66.2%, Pacific Islander 53.8%, Two or More Races 74.2%, Unknown or Decline 69.5%, White 79.7%. The overall Course Success rate is 68.8%. Course Completion Rates based on demographic are as follows: African American 79.6%, American Indian/Alaska Native 84.2%, Asian 87.6%, Latino 81.8%, Pacific Islander 78.8%, Two or More Races 84.5%, Unknown or Decline 85.1%, White 88%. The overall Course Completion rate is 83.4%.
3. Statistical Test:
    a. State which test(s) you are using to analyze your data. *For example:*
    Chi-Square Test for Independence.
    b. Explain why you chose that test. *For example:*
    I chose the Chi-Square Test for Independence because both my independent and dependent variables are categorical. Additionally, I chose the Test for Independence because I am testing the relationship between two categorical variables and not just one.

c. Conduct the test. Include all of the steps for hypothesis testing, including effect size if you reject the null hypothesis, and any supporting content such as tables, calculations, etc. You may use any method you like to conduct the tests (excel, data software, by hand, graphing calculator, etc.). You just must specific which method you chose and the steps you took depending on the method. If the test you are using allows for a written out and mathematical hypothesis, please include both. *For example:*
    i. *Calculation Method: All of the calculations in this study were performed by manually entering formulas in Excel.*
    ii. *Hypothesis Test:*
        1. *Step 1: State the Hypotheses*
            a. *Fall 2020 Course Success*
            H_0: Community College students' ethnicity is independent of course success.
            H_A: Community College students' ethnicity is related to course success.
            b. *Fall 2020 Course Completion*
            H_0: Community College students' ethnicity is independent of course completion.
            H_A: Community College students' ethnicity is related to course completion.
            c. *Fall 2019 Course Success*
            H_0: Community College students' ethnicity is independent of course success.
            H_A: Community College students' ethnicity is related to course success.
            d. *Fall 2019 Course Completion*
            H_0: Community College students' ethnicity is independent of course completion.
            H_A: Community College students' ethnicity is related to course completion.
        2. *Step 2: Find the Critical Value*
            a. *df = (2-1)(2-1) = 1*
            b. *$\alpha = 0.05$*
            c. *Critical Value from Table = 3.841*
        3. *Step 3: Calculate the Test Statistic (see attachments for calculation)*
            a. *Fall 2020 Course Success $\chi^2 = 258.368$*
            b. *Fall 2020 Course Completion $\chi^2 = 76.275$*
            c. *Fall 2019 Course Success $\chi^2 = 477.296$*
            d. *Fall 2019 Course Completion $\chi^2 = 99.288$*
        4. *Step 4: Make the Decision*
            a. *Fall 2020 Course Success:*
            Reject H_0. Based on our data from 55,328 people, we can say that there is a statistically significant relation between a Community College Student's ethnicity and their rate of course success, $\chi^2 (1) = 258.368$, $p < 0.05$.
            b. *Fall 2020 Course Completion:*
            Reject H_0. Based on our data from 55,328 people, we can say that there is a statistically significant relation between a Community College Student's ethnicity and their rate of course completion, $\chi^2 (1) = 76.275$, $p < 0.05$.
            c. *Fall 2019 Course Success:*
            Reject H_0. Based on our data from 63,664 people, we can say that there is

a statistically significant relation between a Community College Student's ethnicity and their rate of course success, χ^2 (1) = 477.296, p < 0.05.

d. *Fall 2019 Course Completion:*
Reject H_0. Based on our data from 63,664 people, we can say that there is a statistically significant relation between a Community College Student's ethnicity and their rate of course success, χ^2 (1) = 99.288, p < 0.05.

5. *Step 5: Calculate the Effect Size*

a. *Fall 2020 Course Success* $V = \sqrt{\frac{258.368}{55{,}328(2-1)}} = 0.07$

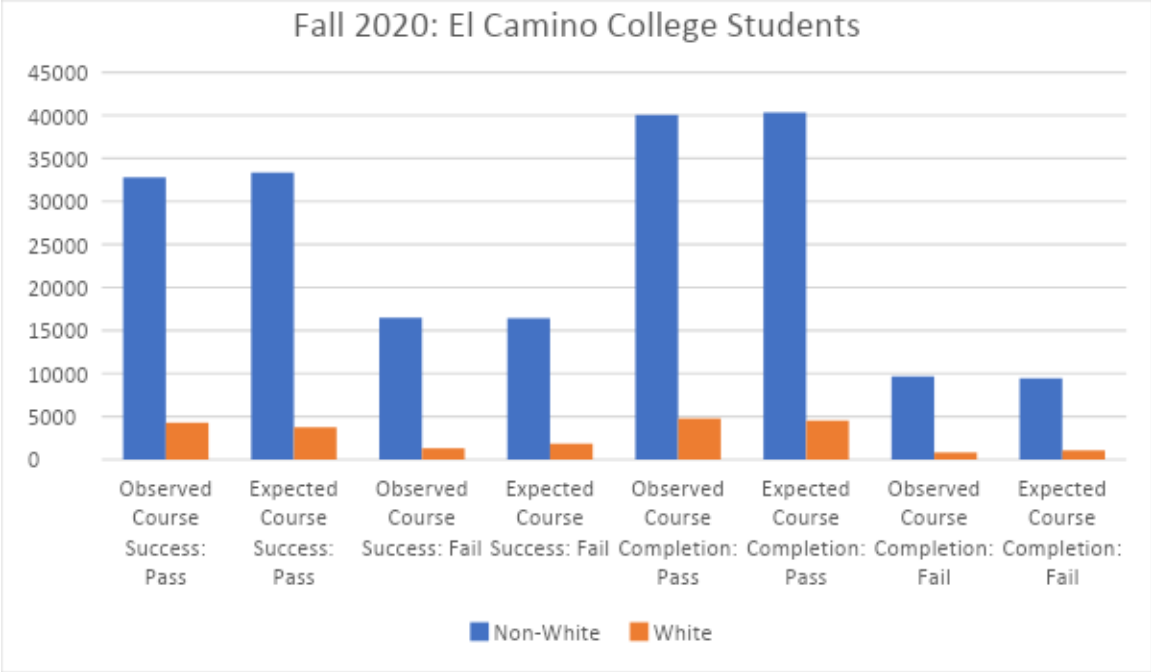
b. *Fall 2020 Course Completion* $V = \sqrt{\frac{76.275}{55{,}328(2-1)}} = 0.04$

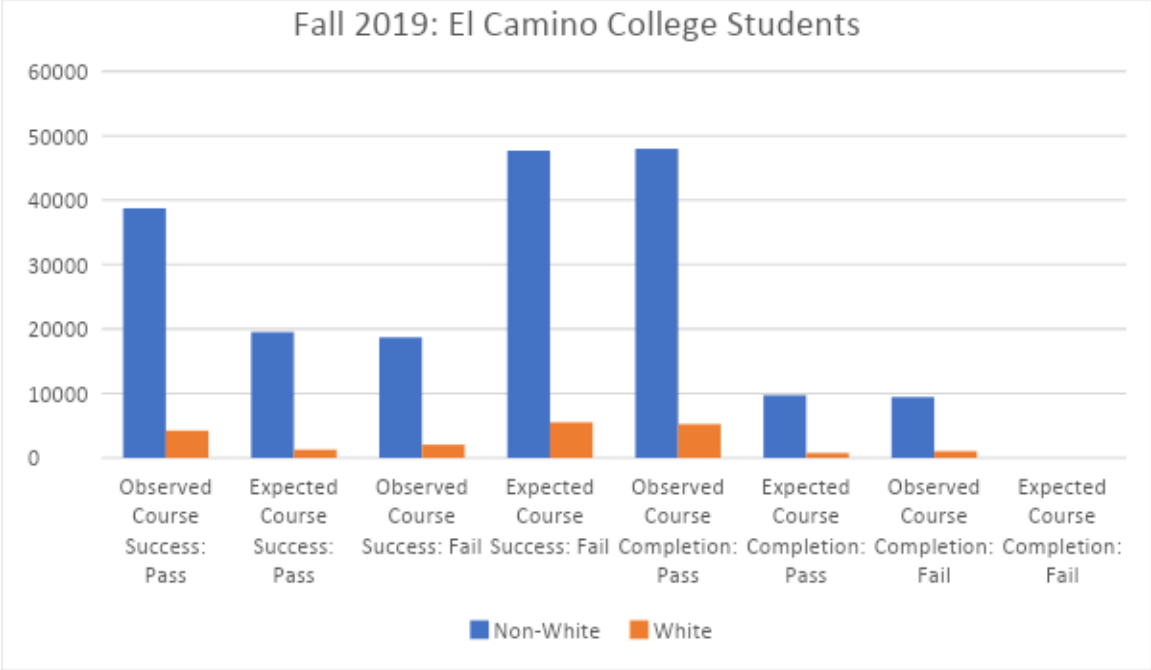
c. *Fall 2019 Course Success* $V = \sqrt{\frac{477.296}{63{,}664(2-1)}} = 0.09$

d. *Fall 2019 Course Completion* $V = \sqrt{\frac{99.288}{63{,}664(2-1)}} = 0.04$

e. *The statistically significant relation between the variables was small for each test.*

d. Data Visualization: Include visual representations of your data using the method that corresponds with the test you conducted. Make sure all identifying information like labels and/or keys are included. *For example:*





Discussion

This section will wrap up your study and function as your conclusion. You should explain your findings in a non-statistical way (think of how you could explain so a common person would understand). Reflect back on your intention of the study described in the Introduction to Project section. Were you able to address your initial research interests? If yes, explain how your findings address your initial research interests. If no, explain why you were unable to address your initial research interests and what you could do differently if you were to continue this study. Did your statistical tests and analysis uncover any other noteworthy findings or suggest an area for further research? Discuss your answer to that question. What would you recommend based on your findings? You may briefly discuss possible solutions or actions that should be taken. You may also discuss further research with different samples, different types of data, different variables, etc. *For example: Based on analysis of El Camino College students from the Fall of 2020 and the Fall of 2019, there is a relationship between ethnicity and rates of student achievement (course success and course completion). Although the relationship between the ethnicity and rates of student achievement are significant in both the Fall of 2020 and 2019, the relationship was actually slightly stronger in Fall of 2019, before the pandemic. In both terms, the relationship between ethnicity and course success is stronger than the relationship between ethnicity and course completion. Overall, my findings provided support for the following arguments. First, ethnicity has an impact on students' course success and course completion. I did not conduct a test to who was affected more by the pandemic, whites or non-whites, both the data from the Fall of 2020 and the Fall of 2019 illustrate that white students have higher rates of student course success and course completion. Second, ethnicity had less of an impact on rates of success and completion in the Fall of 2020, during the pandemic, than during the Fall of 2019. However, ethnicity does have a significant impact on rates of success and completion. The decrease in significance in the Fall of 2020 and the Fall of 2019 could be explained by supplemental resources and services provided by institutions as well has leniency from instructors.*

I was unable to fully address my initial research interests. First, the data I analyzed reflects ethnicity and not necessarily socioeconomic status. Although the two are often related, I cannot verify this based on the data I accessed. Second, I could not conduct a test of significance to determine whether success and completion rates of whites or non-whites were impacted more. I can make predictions based on the descriptive statistics, but I am unable to make support those predictions with inferential statistics. This is because the data I was able to

access only contains categorical variables. If I had access to other types of data, I could have conducted a repeated measures test and independent samples test.

For this project, I would recommend further research using data directly related to socioeconomic status as well has acquiring data that would allow for additional types of statistical tests. Additionally, data should be sampled from different institutions across the country. My data only represents one institution in one location and may not accurately represent the overall population of community college students in the United States. While not the intended focus of this study, the data do support a relationship between ethnicity and student success and completion. Based on this relationship, institutions should consider and implement strategies to decrease the ethnic achievement gap.

Source Summary

In this section, you will provide a summary for any sources you used in a citation format of your choice (e.g. MLA, APA, Chicago, Etc.). *For example:*

Anderson, G. (2020, September 16). More pandemic consequences for underrepresented students. *Inside Higher Ed*. https://www.insidehighered.com/news/2020/09/16/low-income-and-students-color-greatest-need-pandemic-relief.

Duffin, E. (2021, May 3). Undergraduate enrollment in 2-year colleges, by ethnicity U.S. 2018. https://www.statista.com/statistics/421137/undergraduate-enrollment-in-2-year-colleges-in-the-us-by-ethnicity/#statisticContainer.

Ethnic and racial minorities & socioeconomic status. (2017, July). https://www.apa.org/pi/ses/resources/publications/minorities.

Tables

In this section, you will attach any tables that you created. *For example:*

Observed				
		Course Success		
FA 2020		Yes	No	Total
Ethnicity	Non-White	32801	16940	49741
	White	4280	1307	5587
	Total	37081	18247	55328

		Course Completion		
FA 2020		Yes	No	Total
Ethnicity	Non-White	40077	9664	49741
	White	4771	816	5587
	Total	44848	10480	55328

FA 2019		Course Success		
		Yes	No	Total
Ethnicity	Non-White	37927	19476	57403
	White	4990	1271	6261
	Total	42917	20747	63664

FA 2019		Course Completion		
		Yes	No	Total
Ethnicity	Non-White	47694	9709	57403
	White	5510	751	6261
	Total	53203	10461	63664

Expected

FA 2020		Course Success		
		Yes	No	Total
Ethnicity	Non-White	33336.575	16404.425	49741
	White	3744.42501	1842.57499	5587
	Total	37081	18247	55328

FA 2020		Course Completion		
		Yes	No	Total
Ethnicity	Non-White	40319.565	9421.435	49741
	White	4528.767	1058.233	5587
	Total	44848	10480	55328

FA 2019		Course Success		
		Yes	No	Total
Ethnicity	Non-White	38696	18707	57403
	White	4221	2040	6261
	Total	42917	20747	63664

FA 2019		Course Completion		
		Yes	No	Total
Ethnicity	Non-White	47971	9432	57403
	White	5232	1029	6261
	Total	53203	10461	63664