**Is there a statistically significant difference in average height between Sophomore and Junior Statistics students at Longwood University?**

By: Hailey Marin

**Introduction**

Throughout this paper I have displayed the data of height for both the sample of sophomore and Junior class at Longwood University. I have discussed how I selected my samples and designed my graphs while displaying the mean and standard deviation of the samples as well. The graph also displays important values that are determined. I used Microsoft excel spreadsheets to collect my samples and there is an explanation for that within my data collection section. Sophomore participant and Junior participant samples consisted of 20 and I used the data collection method of simple random sample to ensure that I had a potential representative population. This project highlights for you whether or not there is a statistically significant difference in average height between Sophomore and Junior Statistic students at Longwood University. From the test that I ran I concluded that the junior class has a greater height average than the sophomore class at Longwood University. In the data analysis section, I go on to include how I came to this conclusion as well as the significance of it.

**Data Collection**

The populations in this study consist of 20 random Sophomores who participated in the study and 20 random Juniors who participated in the study. The variable we are comparing in this study is the Height of the participants. In this study the height is quantitative data as the people are recording the data in number values. To obtain my simple random sample of 20 from the sophomore participants and junior participants I used the excel commands. To make this repeatable I will give you step by step instructions on how to create the samples. First go into your spreadsheet and you are going to want to create two new sheets to display your data for both classes. Then go into the original with all of the survey answers and copy and paste the section of Juniors and Sophomores into their designated sheet. In column E of each sheet Row 1 you are going to want to put the title Rand Samp as a title for the columns. Now it is time to actually create your sample, you are going to want go to the cell E2 and insert the equation =RANDBETWEEN(A2, A84). This will only be the case for the juniors as there is only 83 juniors who responded to the survey but as for the sophomores you second A value will be 154 because 153 sophomores responded to the survey. You are going to want to repeat that same equation down row E until you reach row 21 as that will ensure you have 20 random samples. If you find that there is multiple of some numbers, you can refresh the list a few times to ensure that there are no repeats. Then you are going to want and highlight your values, copy them and then go under paste options and select paste values. This will ensure that the numbers do not change on you when you are manipulating the data for your graphs. Overall this study is an observational study as the definition of an observational study is to observe certain outcomes and have no influence over them at all. With this study we were asking participants to answer honestly and that is what we assume we achieved.

**Data Description**

Sophomore Sample

|  |  |
| --- | --- |
|  | Height |
| Mean | 66.1 |
| Standard Deviation | 3.73 |

Junior Sample

|  |  |
| --- | --- |
|  | Height |
| Mean | 69.65 |
| Standard Deviation | 5.46 |

The height for the Juniors and Sophomores is similar within the end of the graphs as both of them when looking at heights that are greater than 74 inches there is a significantly smaller amount. Some differences between the data is amongst the height for each group, surprisingly the sophomore class had a higher mean height than the junior class. But the standard deviation was greater with the sophomore class meaning that there is a potential outlier in the data causing the mean to be dragged higher. As you can see from the graphs below, with Graph 1 you can see the data of height for the sophomores is approximately symmetric which means that the data displayed is closer to the population data height. Unlike the sophomore data, the junior heights (graph 2) is rightly skewed, this could indicate that this sample is not representative of the population. Another possible reason for these differences is there could have been a bigger pool of sophomores to get a random sample from rather than a smaller number of juniors. This could lead to more of a representative sample from the juniors because they are smaller in numbers and less variation.

**Data Analysis**

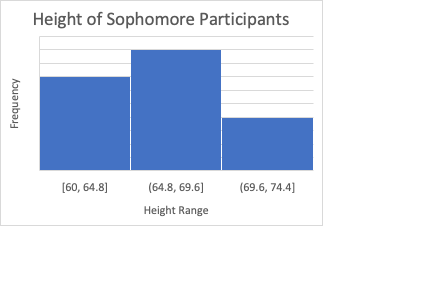
The conditions that are to be met before running a hypothesis test are as follows:

* The data is a random sample from the population
* The sampling distribution of x̅ is approximately normal
* Individual observations can be considered independent

All of these conditions are met as the sample was obtained randomly from the population, the sampling distribution is approximately normal, and the observations are independent as they are all separate individuals. The sample size is large enough that it is greater than 10 but smaller than the population as a whole. The original population is also normally distributed, so this supports that the sample mean is approximately normal as well. The null hypothesis of this test is that there is no difference in height between the junior and sophomore class. The alternative hypothesis of this test is that the junior class has a greater height than the sophomore class. The significance level that I will be using to determine whether or not to accept or reject the hypothesis will be .05. The test statistic equals 2.401 and the p-value of this test is .011023. To determine these values, I used my calculator and ran a 2-sample T test by using the calculator I minimalized the opportunity for mistakes rather than by doing it by hand. We do have a significant result at 𝛼 = .05 level. In the context of this scenario we reject the null hypothesis and accept the alternative as there is a statistically significant difference at the .05 level in which the Junior class has a greater average height than the Sophomore class at Longwood University. Overall from my test that I ran it is clear that the Junior class at Longwood University has a greater height average than the sophomore class at Longwood University as the sample was representative of the population.

**Appendix**

*Graph 1*



*Graph 2*

