Sleep Study Analysis Report

Introduction

The study on students' health and wellness has obtained a sample of 253 colleges students from a liberal arts college in the northeastern United States. For evaluating the college's mental health services and the prevalence of mental health issues, this analysis report will focus on the following two questions: 1) What is an estimate for the average combined depressed, anxious, stress score (DAS score) for this population of college students. 2) What is an estimate for the proportion of college students who pulled all nighters for this population? Do the majority of college students in this population pull all nighters?

We used bootstrap, Central Limit Theorem, hypothesis statement and the p-value to conclude that with 95% confidence, the average DAS score for this population of college students is between 18.04 and 22.09 as well as the proportion of college students who pulled all nighters for this population is between 0.09 and 0.18. And, we recommend that the study would better to get more students' questionnaires for getting more reasonable conclusion.

Exploratory data analysis

Explore the DASScore variable

After we explored the DASScore variable, we can get the following summary statistics (See table 1) and the histogram chart (See figure 1).

Minimum	1st quantile	Median	Mean	3rd quantile	Maximum
0.00	7.00	16.00	20.04	28.00	82.00

Table 1

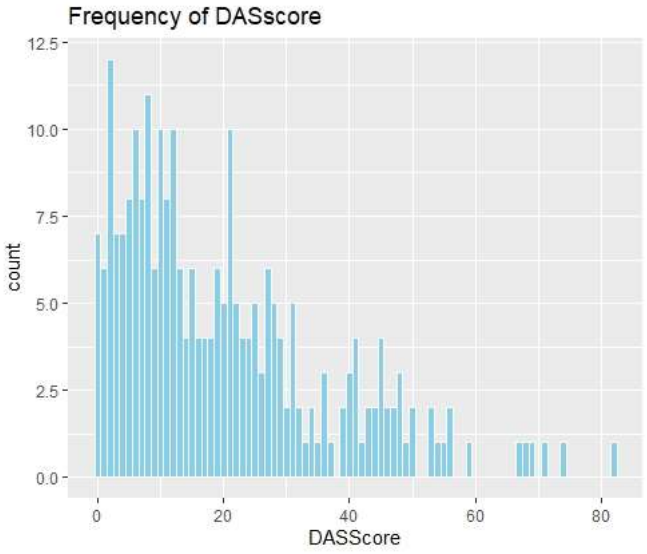

Figure 1

The DASScore variable in histogram chart shows a right-skewed shape. The median of the DASScore variable in this sample is 16 and the mean is 20.04. The maximum is 82.

Explore the AllNighter variable

The AllNighter variable is a categorical variable, so let's look at the frequency of all the categories.

AllNighter	Frequency	Proportion
No	219	0.87
Yes	34	0.13

Table 2

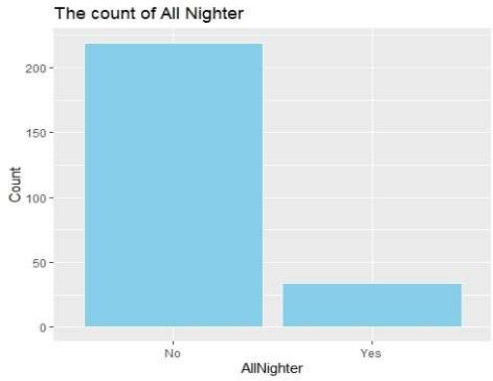

Figure 2

The table and figure of AllNighter variable show that in this sample 87% of students in this sample did not pull all nighters and 13% of students in this sample pulled all nighters before.

Group investigation

We investigated whether there are differences of DAS scores among students who are early risers or night owls or neither of them. We grouped the LarkOwl variable which is a categorical variable with 3 categories (Early riser, night owl and neither) and observed the average DAS score in different groups. We found that the average DAS score in night owl group is the highest, the average DAS score in group that is neither early riser nor night owl is the lowest and the average DAS score in early riser group is in the middle. (See table 3 and figure 3)

LarkOwl	AvgDAS
Early riser	20.17
Neither	19.80
Night Owl	20.71

Table 3

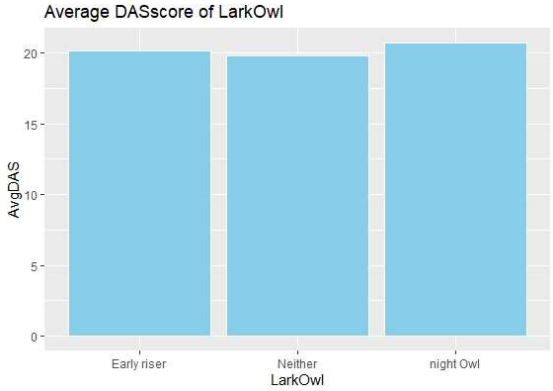

Figure 3

Inference

For answering the question, what is an estimate for the average combined depressed, anxious, stressed score (DAS score) for this population of college students, we simulated a 95% confidence interval with the bootstrap simulations based on the sample of 253 students. We are 95% sure that the plausible value for the DAS score is between 18.03 and 22.10 in this population of college students. To make sure the result, we also applied the Central Limit Theorem*. We would conclude that we are 95% sure that the average combined depressed, anxious, stressed score for this population of college students is between 18 and 22.

For answering the question, what is an estimate for the proportion of college students who pulled all nighters for this population, we applied the Central Limit Theorem. We assumed the independence and checked success/failure conditions firstly. With the conditions met, we calculated that the 95% confidence interval is between 0.09 and 0.18. So, we are 95% sure that the proportion of college students who pulled all nighters for this population is between 0.09 and 0.18.

Then, we conducted a hypothesis test to find out whether the majority of college students in this population pull all nighters. We assumed that the majority of college students meant the proportion of college students who pulled all nighters is greater than 50%. So, we stated that the null hypothesis is 0.5 and the alternative hypothesis is less than 50%. We would test the percentage of current sample proportion (0.13) will happen when the normal distribution is based on the null hypothesis (0.5). Our significant level would be 0.01. The low-tail p-value result is $1.43*10^{-31}$ which is really small that we can reject the null hypothesis. Therefore, if the p-naught is greater than 0.5, we could say the p-value would be much smaller. So, we would reject the null hypothesis and conclude that the majority of the students in this population do not pull all nighters.

Conclusion

We can conclude that if we assume the sample is random, we are 95% sure that the average combined depressed, anxious, stressed score for this population of college students is between 18 and 22.

Also, if we assume the sample is random, we are 95% sure that the proportion of college students who pulled all nighters is between 0.09 and 0.18.

Based on our assumption that the sample is random, we can reject that the majority of college students pull all nighters. However, the sample came from a specific liberal arts college and from a specific area, northeastern, so that we can not exclude some sample bias in this sample. On the other hand, the sample size is only 253, it is really such a small sample size that it can not be considered to be a reasonably and comfortably big enough sample.

However, due to the small sample size, we would also recommend the study to go on getting more questionnaires from this liberal arts college.

*: Central Limit Theorem method. We assumed the sample is random and checked Normality (the sample size is greater than 30 and without extreme outliers). With the conditions met, we calculated the 95% confidence interval is (17.99, 22.08).