Mary Parlier, Rahul Das, Brian Tramuel, Sam Coronado, Finn Isler

Prof. Hector Barriga-Acosta

STAT 2223- Section 004

Team 20: Predicting the Perceived Stress Level of Students at UNCC

Background

With many of us being students pursuing degrees in STEM, the common thread among not only us but our peers always seems to come down to stress. Despite the many differences in schedules, lifestyles, age, and even major- we can usually relate to dealing with a hefty load of stress. Naturally, this topic seemed alluring for collecting real-world data for our final project. In order to avoid a dataset that would yield extreme bias, we could not simply seek to determine the average perceived stress level of all individuals given our small yet statistically significant sample size of at least 50. We determined that for us to have any chance of accurately representing a population within our polling we would need to narrow our scope to students at the University of North Carolina at Charlotte.

Data Collection

Data collection was to be conducted via a google form that would ask students to input mostly numerical data to serve as the values of our potential parameters. As a measure of stress we employed the Likert scale which simply asks the person to express their perceived stress on a scale ranging from 0-5, with 0 being not stressed at all and 5 being extremely stressed. This would become our dependent variable which we aim to predict. The predictors we included in

the survey included age, sex, years spent in college, currently enrolled credit hours, average hours of sleep, average hours worked per week, average hours of weekly exercise, and whether or not the participant is pursuing a STEM major. The qualitative data such as sex and STEM involvement were turned into numerical values for our csv, with sex becoming 0-male, 1-female, 2- non-binary/prefer not to say. STEM involvement was a matter of a simple binary consisting of 0- non-STEM major, and 1- STEM major.

Thanks to our classmates filling out the form and begging our friends and peers to take a second to do the same, we were able to collect 52 responses from UNCC students. We exported this data from the google form to a csv where we were able to clean the data, as two participants entered their responses in a format outside of the dataset.

Data Cleaning

For data cleaning it was minimal due to the clear responses from our participants in the experiment. When the prompt “What year are you” was asked for students, a couple students inputted a decimal point number such as 2.5 or 1-2. Meaning that they are in the middle of their semester. When given the data set, we took the lower number or round down. So for example if a student answers with 2.5 when changed it to 2 as that was the intended response. Another issue we were faced with was when asked for hours of exercises. Instead of answering with a whole number many students answered by giving us a range. When we came across this instance we had the same approach as we did with the year. We took the lower number or rounded down. The last issue we had was a handful of students answer by giving us the written form of a number. For example they answer with “one” instead of 1. The fix for this was straightforward as we manually coveted those instances with the actual number manually. Manual cleaning was

possible due to us only having 51 responses. If the data set was larger, different methods would have been used.

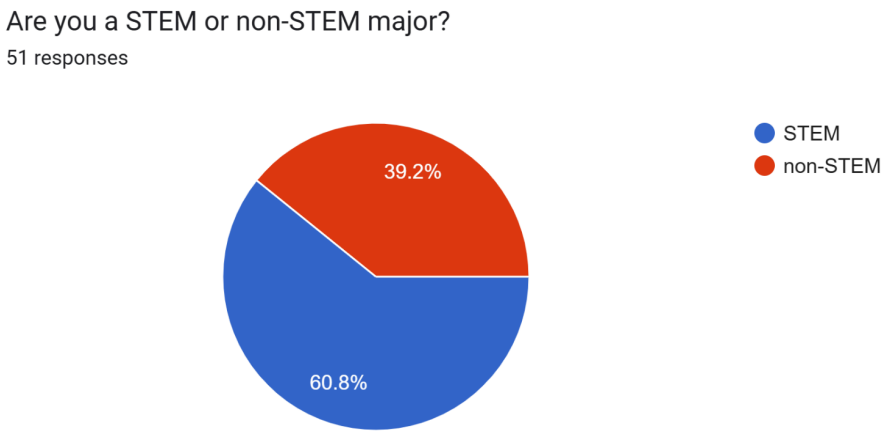
Data Exploration

The SAS System

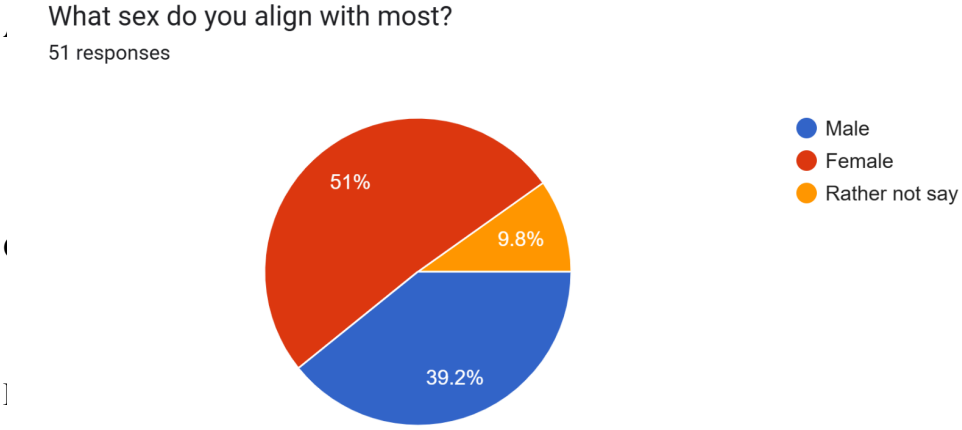
The MEANS Procedure

Variable	Mean	Std Dev	Minimum	Maximum
age	20.4509804	1.3009800	18.0000000	24.0000000
year	2.7450980	1.1973828	0	6.0000000
credit_hours	13.9803922	2.9427212	6.0000000	20.0000000
work	13.8431373	11.4653784	0	35.0000000
sleep	6.3725490	1.2955429	4.0000000	10.0000000
exercise	4.7254902	5.0480825	0	30.0000000
stress	3.8039216	0.9801961	1.0000000	5.0000000

Distribution of stem and non stem students



Gender Ratio of Respondents



Data Analysis

Linear Regression on Entire Population

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	7.27618	1.03945	1.10	0.3825
Error	43	40.76304	0.94798		
Corrected Total	50	48.03922			

Root MSE	0.97364	R-Square	0.1515
Dependent Mean	3.80392	Adj R-Sq	0.0133
Coeff Var	25.59573		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	8.08516	3.62115	2.23	0.0308
age	1	-0.26093	0.18669	-1.40	0.1694
year	1	0.35498	0.20267	1.75	0.0870
credit_hours	1	0.02277	0.04903	0.46	0.6447
work	1	0.00903	0.01263	0.71	0.4787
sleep	1	-0.06672	0.12291	-0.54	0.5900
exercise	1	-0.03988	0.03130	-1.27	0.2094
sex	1	0.35564	0.23290	1.53	0.1341

We first tested the entire data set with multi linear regression. The results we found were not statistically significant for any of the variables we had in our survey. Meaning that there are no statistically significant factors that can predict stress for the entire student body of UNCC.

Variable	P value
Age	.1558
Year	.1184
Gender	.1341
Sleep	.5900
Credit Hours	.6447
Work	.4787

We then tested each response variable individually with stress to find any relations. Here are the results below

Variables	P Value
Age	.9573
Year	.3746
Credit Hours	.9306
Work	.6334
Sleep	.4902
Exercise	.4007

None of these variables where significant even when tested individually with stress

After these disappointing results we narrowed down the group to women in stem and the results were more significant and useful.

Data Analysis for Women in STEM

Simple Regression for Women in STEM

The following model is showing simple regression with sleep and its effect on stress

The REG Procedure
Model: MODEL1
Dependent Variable: stress

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.59223	5.59223	12.22	0.0036
Error	14	6.40777	0.45770		
Corrected Total	15	12.00000			

Root MSE	0.67653	R-Square	0.4660
Dependent Mean	4.00000	Adj R-Sq	0.4279
Coeff Var	16.91334		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.08738	0.89930	7.88	<.0001
sleep	1	-0.46602	0.13332	-3.50	0.0036

Hypothesis Testing:

Null Hypothesis: There is no correlation between sleep and stress

Alternative Hypothesis: There is a correlation between sleep and stress

With a p value of .0036 and the threshold of .05. We can **reject the null hypothesis** and say that for women in STEM there is a correlation between sleep and stress.

We then looked deeper into the group of women that are stem students. Are there any other factors that can be a predictor of stress

Multi Linear Regression Analysis for Women in STEM

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	8.84292	2.21073	7.70	0.0033
Error	11	3.15708	0.28701		
Corrected Total	15	12.00000			

Root MSE	0.53573	R-Square	0.7369
Dependent Mean	4.00000	Adj R-Sq	0.6412
Coeff Var	13.39326		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.85614	3.81378	-0.49	0.6360
credit_hours	1	0.67617	0.26733	2.53	0.0280
sleep	1	1.22309	0.60056	2.04	0.0665
sleep_year	1	-0.03784	0.01950	-1.94	0.0784
credit_hours_sleep	1	-0.11871	0.04206	-2.82	0.0166

For the following model above we looked into how the amount of credit hours, sleep and the interaction of sleep and year along with the interaction with sleep and credit hours affect stress. Meaning are there more complicated relationships between the amount of credit hours and sleep that can affect stress? Is there a relationship between the amount of sleep you get and college year status (freshman, sophomore junior, senior) affect the amount of stress you have?

From the following model we can see that every factor is either statistically significant, or approaching statistical significance.

Conclusion

In conclusion we did not find any predictors when we ran linear regression on the entire population. We then narrowed and tried different combinations of groups to find any sub populations that have a reliable predictor to stress. We found that Women in STEM have a correlation between sleep and the amount of stress they have. There is also a significant correlation between the amount of credit hours, sleep, the amount of sleep they get along with credit hours and the amount of sleep they get and their academic year to stress.