

Data Summary:

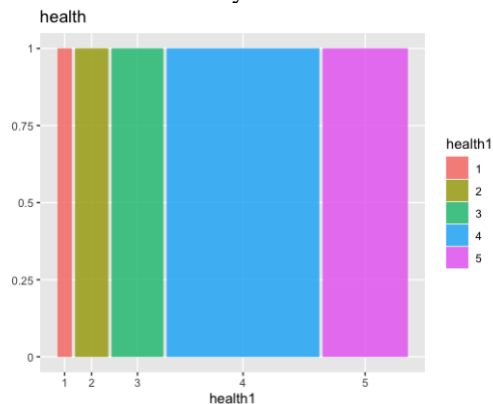
The original data has 20 variables and 886 observations in total. In order to calculate and test for correlation, I have chosen three variables which are part, health, and stud_h. Part is a binary variable containing 1 and 0 values. 1 means that the participant has a partner, while 0 means that he or she does not have a partner. Health is an ordinal variable, ranging from 1 to 5. 1 means the participant is very dissatisfied with his or her health, and 5 means the participant is very satisfied with his or her health. Stud_h is a continuous variable. Each of these variables has 886 observations. I also created a new variable called variablesselected containing the chosen variables by cleaning the original data set. In order to visualize the variables, I transformed the variables of health and part to factor variables. I believe that the health status of medical students can be affected by the variation of their study hours, and the influence of their partners may affect their mental health. Thus, I want to test and calculate the correlation among these three variables.

Planning:

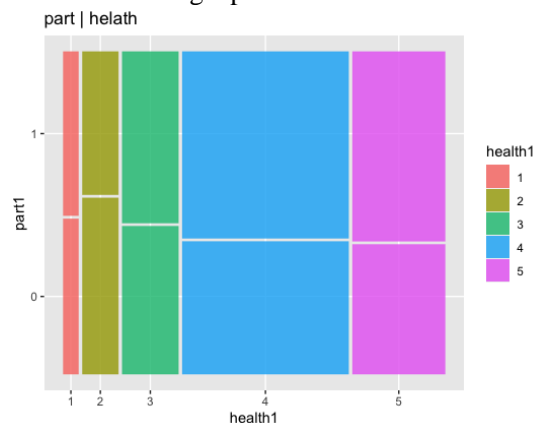
My assumption is that as study hours should be negatively correlated with health, meaning that as study hours increase, health should be indicated to be the level of dissatisfaction. The other assumption is that participants with partner should be relatively more satisfied with their health. In addition, I also tested partial correlation between health and study hours, controlling for the effect of partner. Because of the characters of the variables which are binary and ordinal, I tried to visualize them by using mosaic and jitter plot for intuitively checking the correlation. In terms to test the correlation, given that the variables of health and partner are binary and ordinal data, I chose the tests of spearman and Kendall. In the appendix, I also bootstrapped the dataset to calculate the confidence interval.

Analysis:

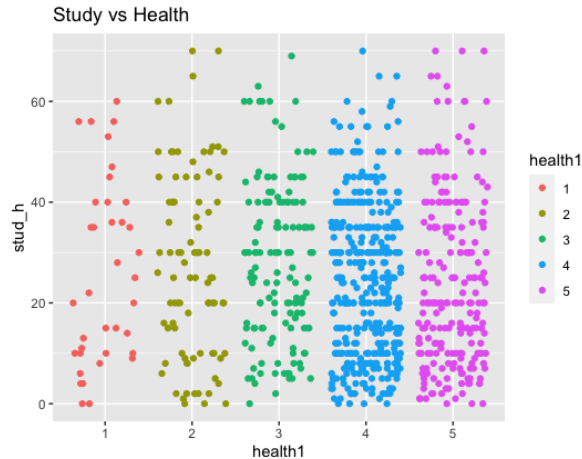
First, I tried to visualize the health status of selected participants. As the graph shows, most of the participants have satisfied and very satisfied health conditions.



And then, I tried to show the relation between partner and health. From the graph, we can see that for those people having a partner, they tend to be more positive towards their health at the level of being satisfied and very satisfied. In addition, at the level of being dissatisfied, people having no partners occupy relatively higher than those having a partner.



Furthermore, I also plotted the graph between study hours and health. From the graph, it illustrates that, at the level of being satisfied and very satisfied, people are more gathered around studying below forty hours, while at the level of being dissatisfied, the number of people is more spread between 0 to 60 hours. It means that most people who are at least satisfied spend less than hours to study per week. For those people who are dissatisfied, it could be other reasons instead of longer study hours, affecting their health status.



In addition, according to the Kendall's test, the correlation between partner and health is 0.06731945. It means that, although having a partner may indicate a more satisfied health, the correlation is very weak. I also used Spearman's test to test the correlation between study hours and health. The number is around -0.08, meaning that less study hours may generate a good health; however, the correlation is very weak.

In terms of partial correlation, I controlled the effect of having a partner, and used Spearman's test to test the partial correlation between health and study hours. The number is around -0.066. From -0.08 to -0.066, we can see that although it still indicates that study hours and health have a weak negative relation, when controlling the effect of having a partner, the relation became weaker. It means that partnership may indeed influence health. In addition, the p-value is approaching 0.05, which may suggest that the data is not very significant.

Conclusion:

Although my findings and the statistic show that partner and study hours indeed affect health status, the relations tend to be relatively weak. In terms of partner and health, by visualizing them and doing Kendall's test, participants having a partner tend to be more satisfied with their health, though the correlation is not strong. In terms of study hours and health, people having dissatisfied health status, not necessarily study longer hours, while, people having more satisfied health, generally spend less than forty hours per week to study. It indicates that there might be other factors which affect health. In addition, by examining partial correlation, when controlling the effect of partner, the correlation between study hours and health became weaker. Regarding additional reasons, one research suggests that positive lifestyle may influence the health of medical students and they are more willing to counsel their health status (Brehm et al., 2016).

Appendix:

Bootstrapping the data of variablesselected:

Intervals :

Level	Normal	Basic
-------	--------	-------

95%	(-0.1180, -0.0095)	(-0.1208, -0.0116)
-----	---------------------	---------------------

Level	Percentile	BCa
-------	------------	-----

95%	(-0.1167, -0.0075)	(-0.1150, -0.0061)
-----	---------------------	---------------------

Calculations and Intervals on Original Scale

References

- Brehm, B. J., Summer, S. S., Khoury, J. C., Filak, A. T., Lieberman, M. A., & Heubi, J. E. (2016). *Health status and lifestyle habits of US Medical Students: A Longitudinal Study*. Annals of medical and health sciences research. Retrieved February 18, 2023, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5423333/>