David Jobin – Project Proposal for STA130 course project

Analysis # 1

I plan to compare the PSYCH\_social\_interactions\_anxiety\_scale\_one\_on\_one variable to the LONELY\_others\_aware variable. The PSYCH\_social\_interactions\_anxiety\_scale\_one\_on\_one variable measures the amount of anxiety the survey responder would feel if they were to have a one-on-one social interaction with a stranger. The LONELY\_others\_aware variable asks the responder if they feel generally recognized by others. I’ll visualize the variables with a 3d histogram. I will perform bootstrapping to generate a confidence interval speaking to the hypothesis that people who feel lonely have a higher level of anxiety when talking to strangers, since they would have had issues making new connections with others, explaining their lack of social connections. I will bootstrap the entire sample of people surveyed, measuring their social anxiety in 1-on-1s, and compare the mean of people who answered they were lonely to the confidence interval I generate.

Analysis # 2

My next analysis will look at the relationship between WORK\_online\_learning and LIFESTYLE\_time\_use\_balance\_exercising. I’ll visualize the variables with a 3d histogram. The former asks the respondent how much of his or her learning during COVID was done online, and the latter measures whether the respondent feels he or she gets enough exercise in. My hypothesis is that people who had more of their education online disrupted their exercise habits for longer, and that correlates with a lower value for LIFESTYLE\_time\_use\_balance\_exercising. I will test this through a linear regression, and studying the correlation value.

Analysis # 3

My last analysis will handle the connection between GEO\_rural\_urban, i.e. what kind of community the respondents live in, and KNOWLEDGE\_morbidity, whether the respondents knew that social connections are negatively correlated to early death. I’ll visualize the data with violin plots and bar plots of the means, sorted categorically by the living location. My hypothesis is that the region you live in does not affect your knowledge of this, and I’ll test this formally, with the constructed population where each living situation has the same mean knowledge and simulate samples to find a p-value. Then I will determine whether to throw out my assumption as a rejected null hypothesis.