**Stats 101A Final Project (Draft)**

By: Sri Chavva, Tem Gareys, Nosung Myung, Jason Filice, Breanna Ramos, and Hayley Todd

**Introduction:**

For this project, we are investigating the relationship between a respondent’s general happiness with twelve other variables including number of household members 18 and older, health, whether they own or rent a home, whether or not they use Instagram, their marriage status, the respondent’s sex, the highest school year the respondent completed, the level of satisfaction with their job, their income, and how many hours per week they typically work. Not only is this an overall interesting project for research purposes, we are motivated to look into this relationship for practical purposes as well, to see what variables in everyday are most likely to affect one’s overall happiness. The general expectation of this project would be that only some, not all, of the predictor variables are statistically significant, so we will be performing different calculations and using various function in R in order to determine which variables will have a significant influence on our data, and which do not. For those that do not, we will remove them from our model. Once we have found significant variables, we can hopefully then find the best possible transformation of the data to apply.

**Methods and Procedures:**

After creating a temporary data set as to not alter the original data, we began by cleaning the data variables. This primarily involved taking values within variables assigned for the purpose of representing a missing data point, and changing them to NAs. Then, by excluding these cases with NAs using complete.cases, we were able to find the p-values of each individual variable in order to determine whether or not those variables would be a significant addition to the full model. After finding these p-values, we determined that the significant variables that should be included are Household, OwnHome, Marital, Sex, Education, and Income. We also used the variance inflation factor (vif) function in R to test the variables multicollinearity.

For our final transformed model, we decided to model the situation as: