Paul Trimble

Capstone Project—Trauma and Coping Skills

Source and Problem Statement

In this project, we examined data related to trauma experienced and trauma symptoms experienced for 108 students who participated in a study. Some demographic/personal data was included in the data (age, gender, family status, education level) as well as trauma related scores (self reported by the individual according to a measuring instrument survey)—cognitive flexibility score, emotional expression flexibility score, flexibility of coping strategies score, total trauma experienced, current trauma symptoms score, and lifetime trauma symptoms score).
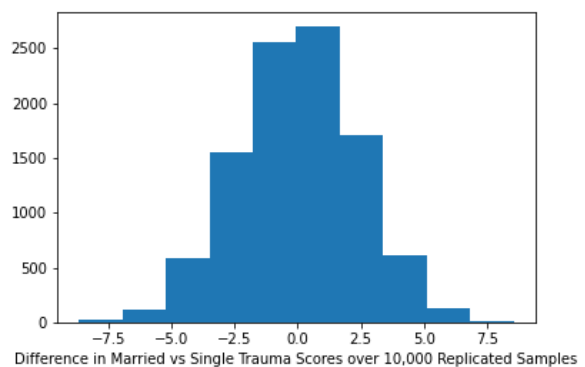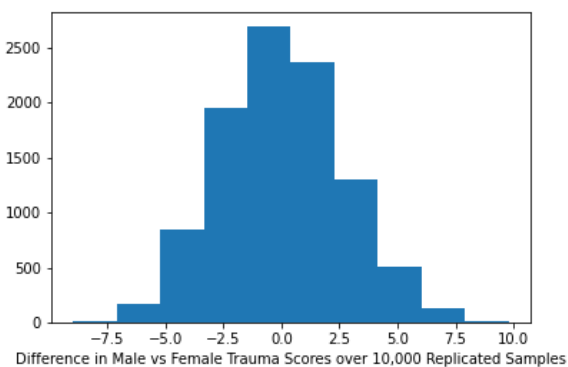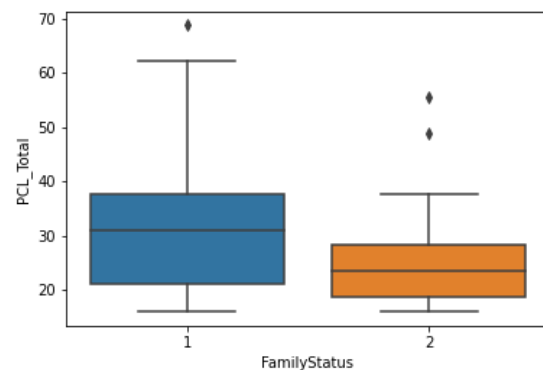
In the original study, the authors explored correlations between the trauma related variables but did not seem to go so far as to attempt to model trauma symptoms based on any of the coping skills and strategies. Since there would be an obvious upside to discovering if equipping people with specific coping skills and strategies could reduce current or long term trauma symptoms, we decided to attempt to model current trauma symptoms as a function of some of the coping strategies (and include demographic features for interest and potential model improvement/understanding). It was thought that if current trauma symptoms could be reduced then also lifetime trauma symptoms would also be reduced, as the cumulative experience of current symptoms over time.

Exploratory Data Analysis/Wrangling

In terms of exploratory data analysis, the data was found to be tidy and clean (no null values or values that did not fit the format for a data column). No dummy variables were needed to be created, but we did convert married status to a 0 or 1 variable (from a variable with values 1 or 2) for better modeling. It was found that many of the coping strategy features were lightly correlated with one another (r values around .3) and trauma experience (r values around .5). This was as expected since the focus of the study as reported focused around the correlation among the variables more than any explanatory model of trauma.

Paul Trimble

Capstone Project—Trauma and Coping Skills

      Education was dropped as an explanatory features as it showed no evidence of contributing to a trauma model (via the visual scatterplot and one feature linear model scored very low). Age was found to be very lightly correlated with trauma (older=less trauma), but with an r^2 of .05 when run as a single feature linear model. Married status and gender both showed mild relationship with trauma score (see box plots below) and both were then tested for significance of effect by running 10,000 permutations of trauma scores to see what kind of differences in average between the two groups for each would result, and then compared to the actual difference. The difference in 6 points between married and single (married=lower scores) was found to be significant at p value .01. The difference between men and women (about 2 points) was not found to be significant (p value .33). Thus age and married status were kept within the features for incorporation into the model but education and gender were not.

Difference in Male vs Female Trauma Scores over 10,000 Replicated Samples

Difference in Married vs Single Trauma Scores over 10,000 Replicated Samples

Paul Trimble
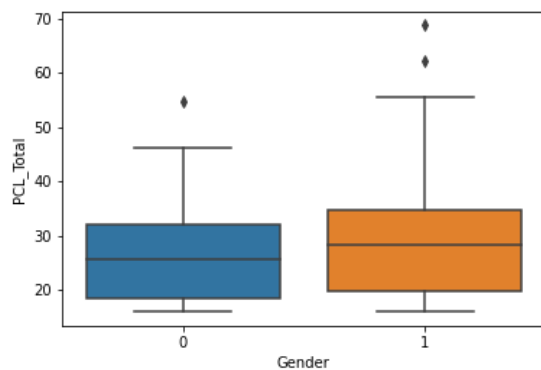
Capstone Project—Trauma and Coping Skills

At this point, we were ready to move forward with age and married status, along with the three coping skills features as the explanatory variables, and current trauma score as the response variable.

Modeling and Model Evaluation

After splitting the data into training and testing partitions, a linear regression was run on the training data. Then the results were applied to the testing data. The results were underwhelming. The training r^2 was .37 and testing .015. For thoroughness cross validation was done as well for linear regression and the results were similar—average r^2 was .11. In addition, a simple scatterplot of predicted vs actual trauma scores showed a very wide and random scatter, further contributing to the sense that our feature variables did little to explain trauma scores.

Since there was clearly collinearity among the variables included in the data, and the scores were not all of equal scale (some had much larger ranges), we also performed a Lasso and Ridge regression, which includes scaling the features before analysis and allows for more distinction among the variables as to which is contributing more to the model.

Before the Lasso regression was performed, we needed to scale the data and establish a hyperparameter value (alpha). This was achieved using cross validation in two stages, zeroing in on the apparent best choice of .616. With this hyperparameter value set, the Lasso regression was performed. The results were r^2 of .367 (training) and .006 (testing). This translates to very little predictive power of trauma via our model features. Of note, cognitive flexibility and flexibility of emotional expression (features 2 and 3 in visual below) were the features with the largest coefficients (contributions).

The ridge regression followed a very similar process. Cross validation hyperparameter search in two stages yielded an alpha of 23.43. Scaled data were analyzed and the model yielded r^2 of .361 and .039 (training and testing, respectively). Again, cognitive flexibility and flexibility of emotional expression (features 2 and 3 in the visual below) were the features with the largest coefficients (contributions).
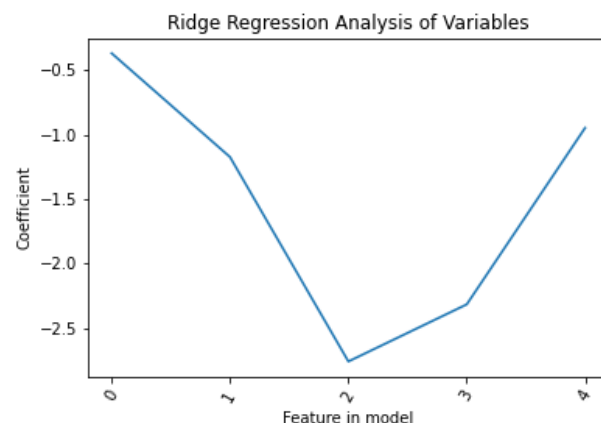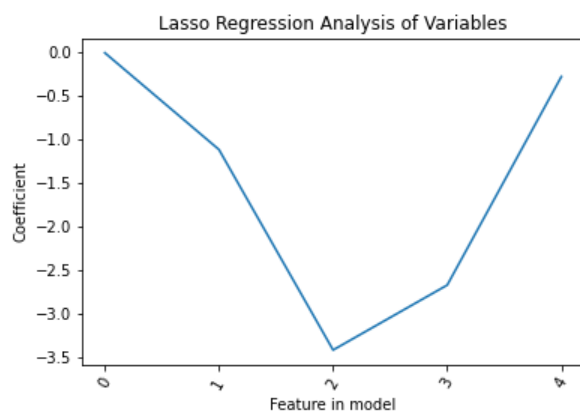
Paul Trimble

Capstone Project—Trauma and Coping Skills

The headline takeaway from this modeling process is that none of the models does a good job of explaining trauma as a function of these skills or coping strategies; perhaps there is simply too much spread among the trauma and coping skill levels among the subjects to allow for such modeling. This does not necessarily mean that improving coping skills could not lower trauma experience scores, but the data at hand would not allow us to draw such conclusions. Perhaps in then end this is why the original authors chose to place their focus on the correlations among the various coping skill levels, since more could be stated with mathematical confidence in that area.

If forced to choose a model, I would choose the Ridge regression, since it takes into account collinearity and differently scaled variables better than linear regression, and has a slightly better testing score than the Lasso regression. It must be noted, however, that this is far from an endorsement of the model, since the main takeaway from the process was that this data does not provide us with a powerful or accurate explanatory model.

Of course, a much better source of data for this sort of modeling would be a controlled experiment, where coping skills strategies were applied to subjects, who then had a before and after trauma symptoms score. This could have as many subject groups as treatments or combinations of treatments (skills training). This approach could lead to much more confident analysis with respect to cause and effect, enabling the development of a model with powerful and accurate explanatory power.