1. After loading in data set into SPSS, I used the DESCRIPTIVES and FREQUENCIES commands to look at some exploratory results of my variables (the SUBNR are meaningless). For two different outputs of FREQUENCIES, I generated a histogram and a pie chart detailing the frequency of the different answers for the ADJEXT variable. In terms of anomalies, I found that the maximum value for ADJINT, a scale from 1-4, contained an answer of 6. Because this is a nonsensical answer with no good interpretation, I remove it from further analysis. I also found that the maximum value for SHY was 11 though the scale has a max of 9. In this case, I changed the value to 9 as the participant likely wanted to indicate the maximum. Lastly, I found that one participant, (SUBNR 123) was changed to SUBNR 1235. This was most likely a mistake when it was coded and it was changed back to 123.

2. The distributions of each variable look reasonably normal. I’ve included the qqplot of the variable QUIET, a 9 point scale that shows some minor deviations but is certainly interpretable. The tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk) indicate that they are indeed statistically interpretable). There doesn’t seem to be too many outliers via the box plots (this is probably aided by the fact that there aren’t that many possible answers) with the possible exception of the variable LIVEL. LIVEL has a median of 6 with an IQR or 2. This means that the subject who answered 1 on the LIVEL scale can be considered an outlier (via >3 IQR from median criterion). Aside from that, no other data point fits that criteria. In relation to the bivariate plots, there seems to be a positive linear relationship against EXTRA and ADJEXT. This means that subjects who checked multiple adjectives of extraverted attributes also self reported as being extraverted. The crosstabs output for the EXTRA and ADJEXT relationship support this as well with an informal diagonal relationship between EXTRA and ADJEXT. I also looked at the bivariate relationship between EXTRA and SHY. As predicted, the bivariate plot shows a negative correlation, i.e. the more EXTRA, the less SHY someone is. This is once again supported by the crosstabs output with subjects tending to rate less as SHY if they are high on EXTRA.

3. Because some of the variables deviate a little bit from normality, I COMPUTED two new variables of LIVEL and EXTRA that were the square root and log of themselves. This is because both of these variables showed minor positive skewness in normality and transforming the variables may make it better. I completed the transform by taking the square root and the log of LIVEL and EXTRA to generate 4 new variables EXTRASQRT, EXTRALOG, LIVELSQRT, and LIVELLOG. However, through using EXAMINE, the variables did not become more normal and if anything, were more skewed. Considering these results, future analysis disregarded these 4 transformed variables.

4. For a more aggregate measure of extraversion, I summed up the variables EXTRA and OUTGO into a new variable called EXTRAVERT. Both of these scales are on the same scale (so I’m not weighing one more heavily than the other) and measure extraversion and outgoing. Both of these scales therefore essentially measure the same thing. For a collapsed measure of introversion I summed up QUIET and INTRO into a new variable called INTROVERT since how quiet someone is should certainly positively correlate with how introverted they are. After EXAMINING these new aggregate variables and comparing them to the 4 old ones it looked like it actually might have improved the kurtosis a little bit (brought the mean down closer to 0 compared to the old variables) however, the variables are a little less normal. This is emphasized by the qqplot of EXTRAVERT with more deviations than either EXTRA or OUTGO (shown for comparison) and quantatively confirmed via Kolmogorov-Smirnov and Shapiro-Wilk.