**SPSS Dataset:**

600 students participated in a university-wide experiment to test if an educational intervention increased test performance. Half of the students were assigned to be in the control group and half of the students were assigned to be in the experimental group (the experimenters used matching techniques to keep the groups as random as possible). The experimental group was the only group that received the educational intervention. All participants were given exam 1 to assess their initial test-taking ability. Then, for the next 4 weeks, the professors began administering the educational intervention to only the experimental group. All participants then took a midterm. For the last four weeks of class, professors again administered the educational intervention to only the experimental group. All participants then took the final exam. Can it be concluded that the educational intervention worked (e.g., did the experimental group outperform the control group?)?

Variables:

* Gender (1 = male, 2 = female)
* Group (1 = control group, 2 = experimental group)
* 3 exam scores, ranging from 2-100 points (students received two points for writing their name and no extra credit was available). Decimals are possible!
  + Exam 1 at beginning
  + Midterm
  + Final exam
* Accuracy:
  + Include SPSS output and indicate how the data *are not* accurate.
  + Include SPSS output to show that you fixed the accuracy errors by deleting that data.
* Missing data:
  + Include SPSS output that shows you have missing data.
    - MCAR: Missing completely at random: Fill in using linear-trend-at-point for continuous only.
    - MNAR: Missing not completely at random: Delete those who didn’t finish the experiment.
  + Include SPSS output and a description that shows what you did with the missing data.
* Outliers:
  + Univariate:
    - Create z-scores (no output needed).
    - Are there any outliers on each individual exam? (count the number of outlier scores)
      * Exam 1:
      * Midterm:
      * Final Exam:
  + Multivariate:
    - What are the top 5 Mahalanobis scores? (you can list them, rather than output).
    - What are the *df* for your Mahalanobis cutoff?
    - What is the cut off score for your Mahalanobis measure?
    - How many outliers did you have?
    - We’ve got a big sample, delete any outliers.
* Assumptions:
  + Multicollinearity: Include a bivariate correlation table of your continuous measures.
    - Do you have issues with multicollinearity or singularity?
  + Normality:
    - Univariate: Include a SPSS box of skew/kurtosis values for your repeated measures exam scores.
      * Are any of these values problematic?
    - Univariate: Include a SPSS box of the KS test.
      * Are any of these values problematic?
    - Multivariate: Include a SPSS box that shows how you might assess multivariate normality.
      * Do you think you’ve met the assumption for normality?
  + Linearity:
    - Univariate: Include a scattermatrix of all three of the exam scores.
      * Are any of the pairs of combinations bad?
    - Multivariate: Include a SPSS box that shows how you might assess multivariate linearity.
      * Do you think you’ve met the assumption for linearity?
  + Homogeneity:
    - Univariate: how might you test for univariate homogeneity?
    - Multivariate: Include a SPSS box that shows how you might assess multivariate homogeneity.
      * Do you think you’ve met the assumption for homogeneity?
  + Homoscedasticity: Include a SPSS box that shows how you might assess multivariate homoscedasticity.
* Why might we not want to transform data?