**Module 3 Problem Set**

**GAF, Consumer Satisfaction, and Type of Clinical Agency (Public or Private)**

A researcher wants to know if mental health clients of private versus public service agencies differ on Global Assessment of Functioning (GAF) scores and on Satisfaction with Services (Satisfaction). She has collected data for 34 clients from a private agency and for 47 clients of a public agency.

**Directions:**

Use the SPSS data file for Module 3 (located in Topic Materials) to answer the following questions:

1. What is the independent variable in this study? What are the dependent variables?
2. The first step for the researcher will be to clean and screen the data. Please do this for the researcher and report your findings. Be sure to check it for possible coding errors, as well as complete the screening of the data to see if the data meet assumptions for parametric tests. Did you find any errors that the researcher made when setting up the SPSS data file (check the variable view)? If so, what did you find? How did you correct it?

         HINT:

Yes, one of the variables is incorrectly listed as scale.

1. Were there missing values on any of the variables? If so, what might you do for those for the independent variable? What about those for each of the dependent variables? Explain your reasoning.

HINTS:

* Yes, each variable has some missing data. Describe how many (and % of all) are missing on each variable.
* When considering what to do about the missing values on each variable, consider if you really can guess what agency a person came from. Next, for the continuous variables, consider (1) what % of values are missing (if more than 5% are missing, what might this mean?); (2) is there a pattern to the missing scores? Include information from the Output file of your SPSS Explore analyses to provide specific number and % of missing values on each of the dependent variables. Based on this, what recommendation would you make for what to do about the missing values?

1. Did you find any outliers on the dependent variables that were due to errors of coding? If so, what and why? How would you correct an error of coding?3

HINT:

One of the outliers on one continuous variable clearly is a coding error. Which one is that? What would be the best way to handle that outlier?

1. How might you deal with outliers that are not due to coding errors? Explain your reasoning.

HINT:

Use the information you have from your Output file from your Explore analyses to describe the outliers (e.g. how many outliers are there on each continuous variable; do they fall above and/or below the mean). What are ways to handle outliers on the continuous variables? Might there be some arguments against deleting outliers? What are these?

1. Check the descriptive statistics, histograms, stem-and-leaf plots, and the tests for normality that you obtained from your analyses (see box to check in "Plots" when using Explore to analyze descriptive statistics of your data). Considering the skewness and kurtosis values, as well as the Shapiro-Wilk's results (preferred for small sample sizes), did the distribution of scores on either of the dependent variables violate the assumption of normality? How can you tell from the information you obtained from your analyses?

HINTS:

* First, you can look at your histograms and stem-and-leaf plots to see if you observe marked skewness or other indicators of differences between the distribution of scores from the normal distribution.
* Next, you can inspect the computed values for skewness and kurtosis for your variables from your analyses. Report these values in your answer for the continuous dependent variables? Which ones are greater than + 1.0? What does having a skewness or kurtosis value that is greater than + 1.0 tell you about normality? Then, discuss what having these kinds of values tell you about the normality of the distribution of scores on that variable.
* Next, look at the Shapiro-Wilks’ tests of normality that you ran. Results with p < .001 or less indicate a violation of the normality assumption using this type of evaluation.

1. If in #6, you identified any distributions that violate the assumption of normality, what are some options you might use to try to correct the distribution to get closer to normality? (You do not need to do these steps. Just describe them.)
2. Write a sample result section, discussing your data screening activity.