Exercise 5: Checking for Outliers and Normal Distribution

Author: Leilani Madrigal

Department of Kinesiology

California State University, Long Beach

Long Beach, CA 90840

Email: [Leilani.madrigal@csulb.edu](mailto:Leilani.madrigal@csulb.edu)

**Note to the Instructor**: The data set used in this exercise (**Sport Injury and Anxiety**) is data that was collected and put together for the purpose of this exercise module. This exercise uses EXPLORE to check for normal distribution of associated variables.  A good reference on using SPSS is SPSS for Windows Version 23.0 A Basic Tutorial by Linda Fiddler, John Korey, Edward Nelson (Editor), and Elizabeth Nelson.  You have permission to use this exercise and to revise it to fit your needs.  Please send a copy of any revision to the author. Included with this exercise (as separate files) are the SPSS file necessary to carry out the exercise (SPSS file\_scales), answers for the exercises (Exercise 5\_answers) and the SPSS output for the exercises (Exercise 5\_Output). Please contact the author for additional information.

Attached are files for this exercise:

- SPSS file\_scales (SPSS file)

- Exercise 5\_answers (MS Word document)

- Exercise 5\_output (SPSS output file)

**Goals of Exercise**

The goal of this exercise is to learn to check for outliers of main scales that were previously created in Exercise 4. Additionally, students will check that their data is normally distributed by checking for kurtosis, skewness, and histograms of the main scales.

**Checking for normal distribution and outliers**

Using the SPSS file\_scales.sav, we will now go through the steps to check for normal distribution and outliers. In terms of normal distribution, we will be using ***histograms, outliers, skewness, and kurtosis***. There are other ways to check for normal distribution but these will be the ones focused on in this exercise. In a normal distribution, 68% of values will lie between ± 1 standard deviation of the mean, 95.5% of values will lie between ±2 standard deviations of the mean, and 99.7% of values will lie between ±3 standard deviations. Kurtosis measures the peakness or flatness of the distribution, a value near 0 indicates a shape close to normal, (+) = more peaked, while (-) = flatter than normal. A value between ±1 is good and ±2 is acceptable. Skewness measures to what extent a distribution of values deviates from symmetry around the mean. Zero equals symmetric. Positive skewness means greater number of smaller values. Negative skewness means greater number of larger values. Values between ±1 is good and ±2 is acceptable.

Open your SPSS file, click “Analyze” 🡪” Descriptive Statistics” 🡪”Explore”. Then move your scale scores to the dependent list. Let’s start with using “TSK score”. Move it to the Dependent list by clicking on the “🡪”. Next click on the “Statistics” box on the right. Notice that descriptives is already checked, go ahead and check the “outliers” box as well and then click “Continue”. Next click on the “Plots” box on the right. Here you will want to click on the “Histogram” box, then click “continue” 🡪 “ok”.

Your output will begin with the case processing summary, but you can scroll down to the Descriptives chart. Notice on this chart, you are given the skewness and kurtosis values (-0.42, and 0.39, respectively). These fall within good ranges. If you continue to scroll down, you will see the histogram of the scale. This is done as a visual inspection so you can see the shape of the data distribution. In looking at the histogram for TSK score, this looks ok (a little flushed to the right, but overall nice distribution of scores). Continue to scroll down and at the very bottom is boxplot. This will show if there are any outliers in the data set. It appears that there are three outliers. There are different steps you can do with outliers, I always go back to the dataset and ensure that number was entered correctly. If it is correct, you may want to remove the outlier before running analysis. The objective of this lesson is just to be able to identify if your data set has outliers. See *Using Multivariate Statistics* by Tabachnick and Fidell for information on handling outliers.

**Exercise 1.** Now you try checking the normal distribution and outliers for RSSIQ: Renewed Perspective and Return Concerns and SAS2: Worry, Concentration Disruption, and Somatic Anxiety. Again for the purpose of this lesson, we are only checking if data is normally distributed in our scale scores. Check to see if data is normally distributed by running histograms, testing for skewness and kurtosis, and checking for outliers.

Use the table below to organize your information you found and indicate if the data is normally distributed for the two scale scores.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Skewness | Kurtosis | Outliers | Normally Distributed  (Yes or No) |
| TSK-Score | -0.42 | 0.09 | 14, 16, 92 | Yes |
| RSSIQ – Return Concerns |  |  |  |  |
| RSSIQ – renewed perspective |  |  |  |  |
| SAS2-worry |  |  |  |  |
| SAS2-concentration disruption |  |  |  |  |
| SAS2-somatic anxiety |  |  |  |  |