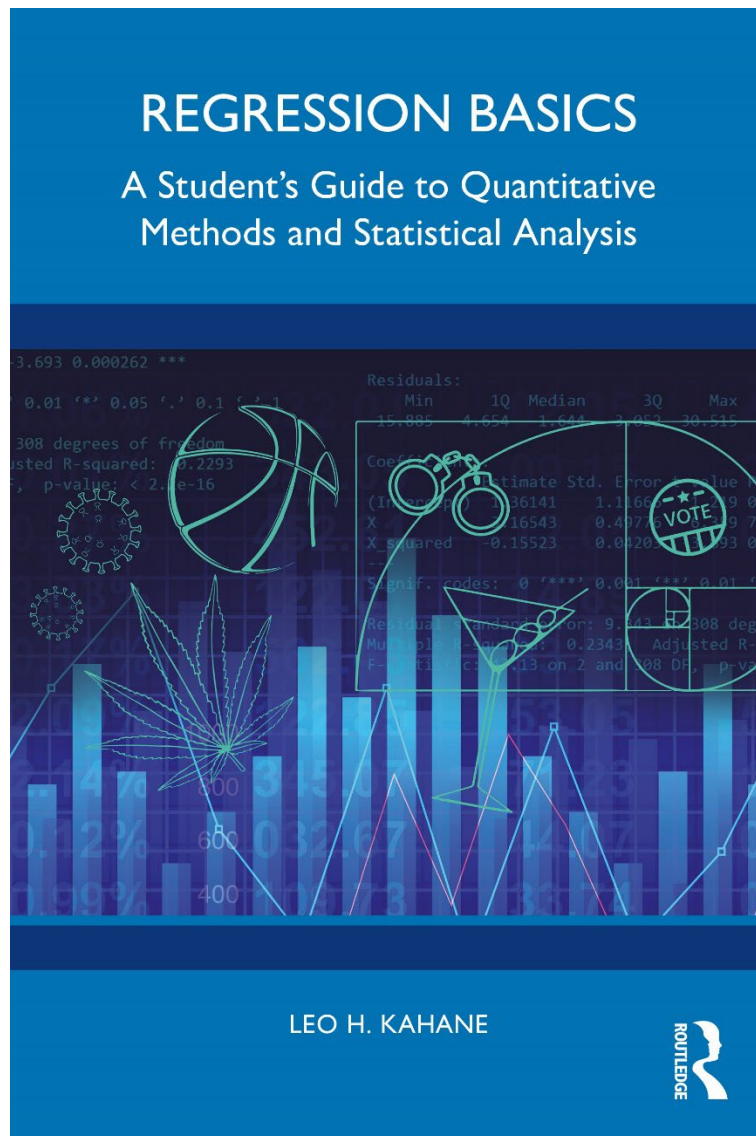


# Instructions for Using SPSS®, Stata®, and R



## ***Introduction***

The purpose of this online appendix is to provide the readers of *Regression Basics: A Student's Guide to Quantitative Methods and Statistical Analysis* (third edition) with basic instructions that will allow them to reproduce the results shown in the book, as well as complete the end-of-chapter problems. For each software used in the book (SPSS, Stata, and R) the instructions show the reader how to:

- Read a data set into the software.
- Transform variables (e.g., compute natural logs or squared values, etc.), and create dummy variables.
- Produce summary (descriptive) statistics and correlations for variables.
- Estimate basic (OLS) linear regressions.
- Save predicted (a.k.a., fitted) values and residuals.
- Create a scatter plot of two variables, (including a fitted OLS regression line).
- Test for heteroskedasticity using the Breusch-Pagan method.
- Estimate OLS regressions with robust standard errors.
- Compute Variance Inflation Factors (VIFs), used to identify high multicollinearity.
- Produce the Durbin-Watson statistic to test for first-order autocorrelation.
- Produce a connected graph for times series variables (including residuals).
- Produce Prais-Winsten estimates to correct for autocorrelation.
- Produce studentized residuals and leverage measures.
- Produce publication-quality tables in R and Stata.

It should be noted that these instructions are *not* designed to be a comprehensive instructions manual on how to use these software programs. SPSS, Stata, and R are *very* powerful programs that can do much more than what is described here. To learn how to take greater advantage of these software programs the reader is directed to more comprehensive books/resources. For example, the following books may be helpful:

*How to Use SPSS®: A Step-By-Step Guide to Analysis and Interpretation*, 11<sup>th</sup> Edition, 2020, by Brian C. Cronk, published by Routledge (available [here](#)).

*A Gentle Introduction to Stata, Revised*, 6<sup>th</sup> Edition, 2022, by Alan C. Acock, published by Routledge (available [here](#)).

*Introduction to R for Social Scientists: A Tidy Programming Approach*, 2021, by Ryan Kennedy, Philip D. Waggoner, published by Routledge (available [here](#)).

### ***Differences Between SPSS, Stata, and R***

Each of these three programs can carry out all the empirical work presented in the book. Of the three programs, SPSS is probably the simplest to learn as nearly all the computations shown in the book can be completed using drop-down menus in SPSS. While writing program code (using “syntax” files) that carry out computations in SPSS is certainly possible, in my experience with new users of SPSS, the simpler ‘point-and-click’ approach is sufficient. As such, the instructions for using SPSS (with only a couple of exceptions) will be accomplished using the drop-down menus.

Stata also has drop-down menus where users can carry out most computations. However, in my experience, most instructors using Stata in a course will teach their students how to write code to

carry out computations. Code can either be submitted in a line-by-line approach submitted on a command line, or submitted as a collection of commands submitted all at once using a text file called a ‘do-file’. Thus, for Stata, we will focus on carrying out computations via coding.

As for R, computations are only carried out using code. Like Stata, this can be done by submitting commands one at a time on a command line, or by submitting collections of commands using a text file called an ‘R script’. In my experience, most instructors teach students how to use R with scripts files, and we will follow this approach in this appendix. While this may make for a steeper ‘learning curve’ in R, the payoff is that by learning some basic coding, users will be able to take advantage of the ‘object-oriented’ feature of R that makes it distinct from Stata and SPSS. Another major difference between R and the other two programs is that R is ‘open source’ and is free to use. Lastly, of the three programs, R has much greater capabilities for data visualization.

One final difference to note between these three programs is that both the Stata and R community of users are continuously contributing to the growing number of new programs available to users that can carry out new procedures and make various computations much easier.

These differences aside, *all* three programs are *very* powerful analytical tools that can be used to explore data and answer empirical questions with them.