What is involved in working with data?

Authored by

stats writer

June 23, 2024

Working with data involves collecting, organizing, analyzing, and interpreting information to derive meaningful insights and make informed decisions. This process requires a combination of technical skills, such as data mining and statistical analysis, as well as critical thinking and problem-solving abilities. It also involves understanding the context and limitations of the data, ensuring its accuracy and reliability, and effectively communicating the findings to stakeholders. Working with data can be a complex and iterative process, requiring attention to detail and the ability to adapt to changing circumstances. It is an essential aspect of many industries and plays a crucial role in driving business growth and innovation.

Part 1: Working with Variables

Most of the time, you'll need to make modifications to your variables before you can analyze your data. These types of modifications can include changing a variable's type from numeric to string (or vice versa), merging the categories of a nominal or ordinal variable, dichotomizing a continuous variable at a cut point, or computing a new summary variable from existing variables. This section will focus on transformations applied to individual variables, particularly recoding and computing new variables.

Computing VariablesThe "Compute Variable" command allows you to create new variables from existing variables by applying formulas. This tutorial shows how the "Compute Variable" command can compute a variable using an equation, a built-in function, or conditional logic.Mean CenteringMean centering refers to a type of variable transformation wherein the average of a variable is subtracted from every observation of that variable. In SPSS, we can use the Aggregate procedure in conjunction with the Compute Variables procedure to mean-center.Recoding (Transforming) VariablesRecoding a variable can be used to transform an existing variable into a different form based on certain criteria. This tutorial covers the "Recode into Different Variable" and "Recode into Same Variable" commands.Automatic RecodeIf you have a string variable and have used blanks to indicate missing values, you may notice that SPSS does not automatically recognize the blank observations as missing. To fix this, you'll need to use Automatic Recode. More broadly, Automatic Recode is also used to quickly convert a string categorical variable into a numeric categorical variable.Rank CasesIn its simplest form, a rank transform converts a set of data values by ordering them from smallest to largest, and then assigning a rank to each value. In SPSS, the Rank Cases procedure can be used to compute the rank transform of a variable.

Part 2: Data Management

Managing a dataset often includes tasks such as sorting data, subsetting data into separate samples, merging multiple sources of data, aggregating of data based on some key indicator, or restructuring a dataset. These types of data management tasks are sometimes called *data cleaning*, *data munging*, or *data wrangling*. This section covers these types of "cleaning" tasks.

Sorting DataSorting a dataset rearranges the rows with respect to one or more variables. Sorting makes it convenient for reading the data; additionally, several SPSS procedures require the data to be sorted in a certain way before the procedure can run. This tutorial discusses how to sort data using the drop-down menus in SPSS.Splitting DataIn SPSS, the "Split File" command can be used to organize statistical results into groups for comparison. Split File is used when you want to run statistical analyses with respect to different groups, but don't necessarily want to separate your data into two different files. This tutorial shows you how to use the Split File command in SPSS and what situations it is useful in.Weighting CasesSometimes, you don't have raw data available to you -- you may only have a frequency table indicating the types of responses and how many times they occurred. Alternatively, you may be working with a dataset that contains a weighting variable. This tutorial shows how to use the Weighting Cases command in SPSS to handle both of these situations.Partitioning Data(Coming soon) The Select Cases procedure is used when you want to create a new dataset by extracting cases from an existing dataset. Unlike the Split File option, Select Cases affects the data itself, rather than the output. This tutorial covers several applications of Select Cases.