Week 2:
Data Entry

The Steps of Data Analysis

6 Week Webinar Class

Presented By

William Bannon, PhD

StatsWhisperer

New York, New York

Data Entry

Data Entry is a critical process within a quantitative research study. Within a quantitative research study, data are traditionally collected by staff via interviews with study participants where hard copies (actual paper) of study surveys are completed. From this point, the completed study surveys are brought back to the research office where the data from the surveys are coded, entered, cleaned, and recoded if necessary. Within the webinar this week, this process is simulated using a short study survey. Please see the webinar video presentations, as well as *The 7 Steps of Data Analysis* text for greater detail. For a synopsis of this process, please see the text below, which presents the key points covered this week in crib note form.

Coding the Data

Each item on the study survey must be assigned a *code*, which will include a *variable name*, as well as *numbers for any assigned responses* to the item. After these codes are assigned, the first step will have been taken toward entering study data into the software database (e.g., the SPSS database). Please see pages 62-67 of *The 7 Steps of Data Analysis* text for an overview of this process.

The data dictionary

Each study must incorporate a data dictionary, which lists the codes assigned to each survey item. The data dictionary will list each variable name assigned to each survey item, as it appears in the software database. Please see pages 63-67 of *The 7 Steps of Data Analysis* text to view a sample data dictionary, as well as instructions on how to create such a document.

Entering the Data

For our purposes, entering the data is defined as a two-step process of *Creating the Software Database* and *Entering Survey Data* from the hard copies of the study surveys into the software database. Please see pages 68-73 of *The 7 Steps of Data Analysis* text for greater detail.

Creating the Software Database

The webinar videos and our text provide a step by step instruction on how to create a database in SPSS (please see pages 68-71 of *The 7 Steps of Data Analysis* text). Within these materials we see that the most critical columns within the database are the *Name*, *Type*, *Label*, *Values*, and *Missing* columns. These columns are critical because if they

are not completed correctly, the statistical analysis will be compromised or not possible. The Remaining Columns in SPSS are the *Width*, *Decimals*, *Columns*, *Align*, *Measure*, and *Role* columns.

Entering survey data

Once the software database is created, the data may then be entered from the hard copies of the study survey. Please see pages 72-73 of *The 7 Steps of Data Analysis* text for detail regarding data entry.

Once all the data are entered from the completed survey hard copies, they will be ready for *cleaning* and then *recoding* as described in the next two sections.

Cleaning the data

Cleaning the data refers to the process of reviewing the data in your database to detect, correct, or remove incorrect, inconsistent, or inaccurate values, which are commonly referred to as *dirty data*. These dirty data have several causes such as data entry errors, corruption in transmission or storage, or invalid responses given by study participants. Data cleaning is tremendously important, as this process is our means of being certain that the data present in the software database actually reflect the responses given by study participants via data collection.

Pages 73-79 of *The 7 Steps of Data Analysis* text describe essential methods used to clean data include *Referencing Survey Hard Copies* (page 74) to the data that have been entered into the software database, *Examining the Variables* (pages 75-77), and examining *Violations in Logic* (pages 77-79).

Recoding Variables

Recoding variables generally refers to changing a variable in some fundamental way to suit the purpose of the study. For example, if letters (e.g., A, B, C) were given as responses to an item, the text needs to be recoded into numbers to facilitate a statistical analysis. Pages 79-85 of *The 7 Steps of Data Analysis* text illustrates two of the most common techniques used to recode variables, which are *Reverse Coding Variables* (pages 79-84) and *Dummy-Coding Variables* (pages 84-85).

In Summary

In this section we see that the process of data entry within a quantitative study is quite involved. Specifically, this process involves, coding the data (as well as recording those codes in a data dictionary), creating a software database (e.g., an SPSS database), entering survey data, cleaning the data, and recoding the data as needed. This process is a prime example of the *Russian Doll Effect in Analysis* (please see pages 58-59 of *The 7 Steps of Data Analysis* text), where the term *Data Entry* sounds like one task, but is actually an involved process comprised of many smaller subtasks.

It is critical to view data entry as a pivotal step in data analysis. In short, a study may incorporate near perfect methods (i.e., study design, data gathering), which may be followed by a flawless data analysis. However, if the data from the study surveys are entered into the software database incorrectly, there is practically no hope that the data provided by the study participants will be reflected in the study findings. Therefore, in order to produce a legitimate and effective data analysis, sound data entry is essential.