

What are the insights and findings revealed through the process of Exploring Data?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are the insights and findings revealed through the process of Exploring Data?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149463>

Exploring data is a structured process of collecting, organizing, and analyzing information to gain insights and reveal findings. This process involves identifying patterns, trends, and relationships within the data, which can provide valuable insights and understanding of a particular topic or phenomenon. Through the exploration of data, researchers can uncover new knowledge, validate existing theories, and make informed decisions. The findings revealed through this process can have practical applications in various fields such as business, healthcare, and education. By delving deep into the data, we can gain a better understanding of complex issues and make data-driven decisions that can lead to positive outcomes. Exploring data is a powerful tool that can uncover valuable insights and help us make sense of the world around us.

Introduction

Before doing any kind of statistical testing or model building, you should always examine your data using summary statistics and graphs. This process is called *exploratory data analysis*, and it's a crucial part of every research project. Exploratory data analysis is about "getting to know" your data: which values are typical, which values are unusual; where is it centered, how spread out is it; what are its extremes. More importantly, it's an opportunity to identify and correct any problems in your data that would affect the conclusions you draw from your analysis.

How do we "get to know" our data? The answer is different depending on whether our variables are numeric or categorical. In this section, we'll demonstrate which statistics and SPSS procedures to use for both types of data.

Part 1: Descriptive Statistics for Continuous Variables

When summarizing a quantitative (continuous/interval/ratio) variable, we are typically interested in things like:

How many observations were there? How many cases had missing values? (N valid; N missing) Where is the "center" of the data? (Mean, median) Where are the "benchmarks" of the data? (Quartiles, percentiles) How spread out is the data? (Standard deviation/variance) What are the extremes of the data? (Minimum, maximum; Outliers) What is the "shape" of the distribution? Is it symmetric or asymmetric? Are the values mostly clustered about the mean, or are there many values in the "tails" of the distribution? (Skewness, kurtosis)

In Part 1, we discuss how to explore quantitative (continuous/interval/ratio scale) data using the *Descriptives*, *Compare Means*, *Explore*, and *Frequencies* procedures. Each of these procedures offers different strengths for summarizing continuous variables. The *Descriptives* and *Frequencies* commands provide summary statistics for an entire sample, while the *Explore* and *Compare Means* commands can produce descriptive statistics for subsets of the sample.

Descriptives Descriptives (Analyze > Descriptive Statistics > Descriptives) is best to obtain quick summaries of numeric variables, or to compare several numeric variables side-by-side. Compare Means Compare Means (Analyze > Descriptive Statistics > Descriptives) is best used when you want to summarize several numeric variables across the categories of a nominal or ordinal variable. It is especially useful for summarizing numeric variables simultaneously across multiple factors. Explore Explore (Analyze > Descriptive Statistics > Explore) is best used to deeply investigate a single numeric variable, with or without a categorical grouping variable. It can produce a large number of descriptive statistics, as well as confidence intervals, normality tests, and plots. Frequencies Part I (Continuous Variables) Frequencies (Analyze > Descriptive Statistics > Frequencies) is typically used to analyze categorical variables, but can also be used to obtain percentile statistics that aren't otherwise included in the Descriptives, Compare Means, or Explore procedures.

Part 2: Descriptive Statistics for Categorical Variables

When summarizing qualitative (nominal or ordinal) variables, we are typically interested in things like:

How many cases were in each category? (Counts) What proportion of the cases were in each category? (Percentage, valid percent, cumulative percent) What was the most frequently occurring category (i.e., the category with the most observations)? (Mode)

In Part 2, we describe how to obtain descriptive statistics for categorical variables using the Frequencies and Crosstabs procedures.

Frequencies Part II (Categorical Variables) Frequencies (Analyze > Descriptive Statistics > Frequencies) is primarily used to create frequency tables, bar charts, and pie charts for a single categorical variable. Crosstabs The Crosstabs procedure (Analyze > Descriptive Statistics > Crosstabs) is used to create contingency tables, which describe the interaction between two categorical variables. This tutorial covers the descriptive statistics aspects of the Crosstabs procedure, including and row, column, and total percents. Multiple Response Sets / Working with "Check All That Apply" Survey Data Check-all-that-apply questions on surveys are recorded as a set of binary indicator variables for each checkbox option. Frequency tables and crosstabs alone don't capture the dependent nature of this data -- and that's where Multiple Response Sets come in.

Sample Data Files

Our tutorials reference a dataset called "sample" in many examples. If you'd like to download the sample dataset to work through the examples, choose one of the files below:

Data definitions (*.pdf)

Data - Comma delimited (*.csv)

Data - Tab delimited (*.txt)

Data - Excel format (*.xlsx)

Data - SAS format (*.sas7bdat)

Data - SPSS format (*.sav)

SPSS Syntax (*.sps)

Syntax to add variable labels, value labels, set variable types, and compute several recoded variables used in later tutorials. SAS Syntax (*.sas)

Syntax to read the CSV-format sample data and set variable labels and formats/value labels.

ARABPSYCHOLOGY.COM