**60-Minute Data Cleaning Seminar Outline**

**Title: "Data Cleaning 101: Preparing Qualtrics Data in SPSS"**

**Introduction (5 minutes)**

1. **Welcome and Objectives**
   * Overview of the session goals:
     + Understand key principles of data cleaning.
     + Learn practical steps to clean Qualtrics data in SPSS.
     + Prepare data for meaningful analysis.
   * Why data cleaning is critical:
     + Accuracy in results.
     + Reproducibility.
     + Avoiding errors in interpretation.

**Part 1: Understanding the Dataset (10 minutes)**

**1.1 Importing and Exploring Data**

* Demonstration: Open Qualtrics .sav file in SPSS.
* Key observations:
  + Variable names: sometimes are autogenerated (e.g., Q1, Q2).
  + Variable labels: descriptive text from survey questions.
  + Change variable names to match labels where needed
* **Hands-on:** Review and familiarize yourself with variable names and labels.

**1.2 Comparing Variable Labels and Names**

* Check for mismatches or unclear labels:
  + Are the labels descriptive enough for analysis?
  + Do variable names align with the labels?
* Discuss best practices for renaming variables:
  + Use clear, short, consistent names (e.g., Age, Gender, Satisfaction).

**Part 2: Selecting Variables of Interest (10 minutes)**

**2.1 Identifying Key Variables**

* Criteria for selection:
  + Variables related to your research questions.
  + Demographics or covariates.
* Removing irrelevant variables (e.g., metadata columns from Qualtrics like ResponseId, StartDate).

**2.2 Practical Exercise**

* Demonstrate filtering and retaining specific variables in SPSS using the **Select Cases** or **Variable View** tab.

**Part 3: Cleaning the Data (20 minutes)**

**3.1 Identifying and Handling Missing Data**

* Types of missing data:
  + System missing vs. user-defined missing values (e.g., “-99”).
* Approaches to handle missing data:
  + Remove cases with excessive missingness.
  + Impute or recode missing values.

**3.2 Ensuring Consistency**

* Correct inconsistent coding:
  + E.g., recode "Male" / "male" / "M" into a single category.
* Standardizing response scales (e.g., reversing Likert items).

**3.3 Cleaning Variable Values**

* Recode Qualtrics exports (e.g., 1 = Strongly Disagree, 5 = Strongly Agree).
* Handling open-ended responses:
  + Identify and fix typos or irrelevant entries.
* Practical tip: Use SPSS syntax for efficient recoding.

**Part 4: Viewing Descriptive Statistics (10 minutes)**

**4.1 Running Descriptives**

* Generate frequencies for categorical variables (e.g., gender, ethnicity).
* Check means, medians, and distributions for numeric variables.
* Identify outliers and strange patterns.

**4.2 Visualizing the Data**

* Use SPSS Charts:
  + Histograms for numeric data.
  + Bar charts for categorical data.

**Part 5: Preparing for Analysis (10 minutes)**

**5.1 Creating New Variables**

* Compute scale scores (e.g., average of items for a subscale).
* Create dummy variables for categorical predictors if needed.

**5.2 Saving the Cleaned Dataset**

* Save as a new SPSS file to preserve the raw data.
* Document changes:
  + Maintain a data cleaning log for transparency and reproducibility.

**Conclusion and Q&A (5 minutes)**

1. **Recap Key Principles:**
   * Importance of variable review, selection, and cleaning.
   * Role of descriptive statistics in identifying issues.
   * Preparing clean, analysis-ready datasets.
2. **Open for Questions and Discussion**

**Supplemental Materials**

1. Step-by-step SPSS syntax examples.
2. Sample messy Qualtrics .sav dataset for practice.
3. Checklist for data cleaning.

Would you like help with any specific content or materials for the seminar?