

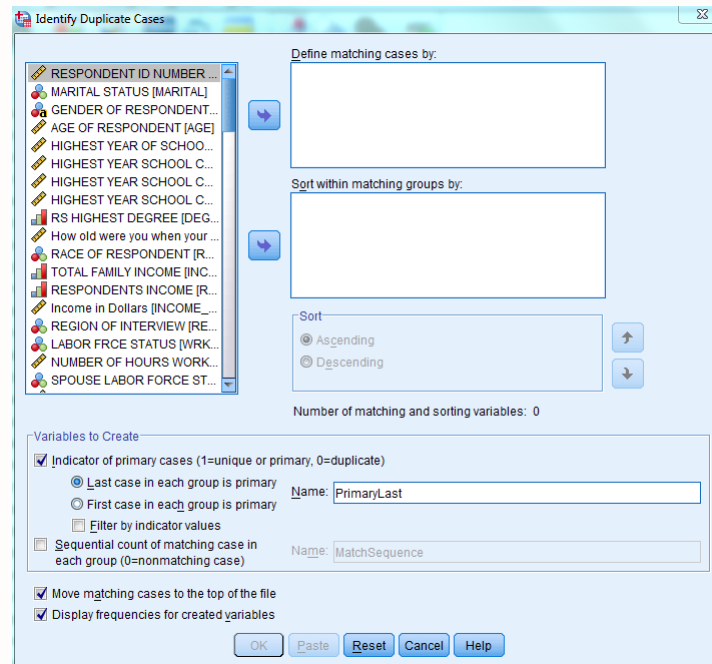
# Identifying Duplicate Cases in IBM SPSS Statistics

# Duplicate Cases

- On cleaning a dataset, one of your first steps should be to identify possible duplicate cases
- Duplicate cases may occur for three reasons:
  - (1) data entry errors
  - (2) multiple cases that share a common primary ID value but have different secondary ID values
  - (3) multiple cases represent the same case but with different values for variables other than those that identify the case
- The **Identify Duplicate Cases** feature enables you to find duplicate cases using almost any method, and allows you to decide whether to identify primary or duplicate cases

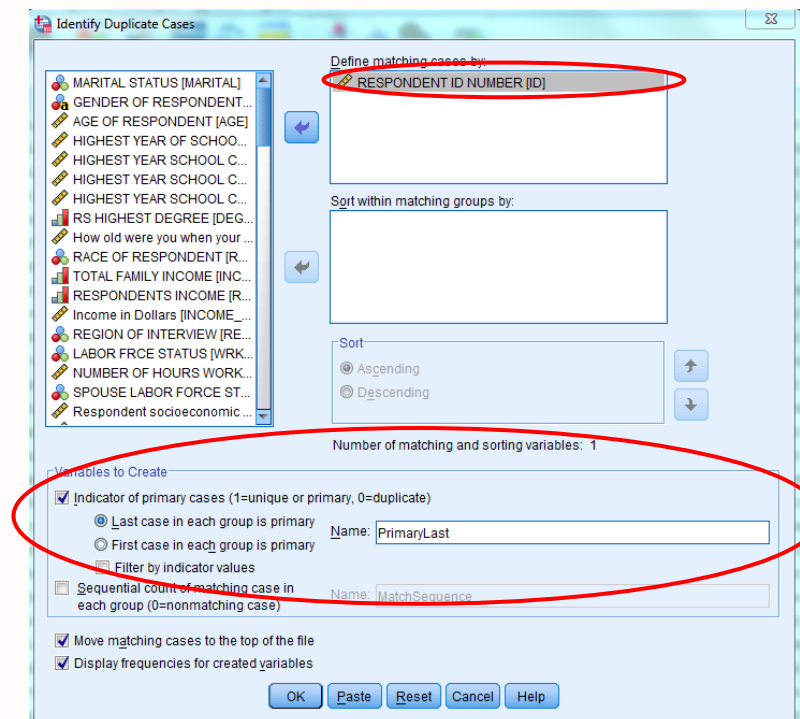
# Identify Duplicate Cases

- To identify and flag duplicate cases:
  - Select **Data** from the menu
  - Select **Identify Duplicate Cases**
  - This opens the **Identify Duplicate Cases** Dialog Box



# Identify Duplicate Cases

- Select one or more variables that identify matching cases and move them to the Define matching cases by box
- Select an appropriate option in the **Variables to Create** section



# Identify Duplicate Cases

- Finally, select one or more variables to sort cases, or automatically filter the duplicate cases, so they won't be included in reports, charts, or calculations of statistics

**For more Tech Tips  
visit**

[www.presidion.com/ibm-spss-technical-tips](http://www.presidion.com/ibm-spss-technical-tips)

[www.presidion.com](http://www.presidion.com)

**Talk to us**

✉ [info@presidion.com](mailto:info@presidion.com) ☎ +44 (0)208 757 8820 (UK) ☎ +353 (0)1 415 0234 (IRL)

