



Data Management Checklist

- Assigning Data Manager (Go – To Person)**
- List of Project Personnel**
 - Principal Investigators
 - Graduate students, RA, etc.
- Description of the Project**
 - When and How data are being collected
- Organizing measurement materials**
 - Double check current version with the published version
 - Note all modifications on current version used
 - References for measurements
- Where to save the data**
 - Set up folder structure
 - Set up initial data file structure and format
 - CSV
 - SPSS
 - Mplus (.dat)
 - SATA (.dta)
 - Other
- Data entry log (if applicable)**
 - Who is responsible for Which proportion of the data entry
 - Timeline for data entry
 - Schedule and manage double checking data entry
- Data backup schedule**
 - Storage for backup data files and related materials
- Development of an user friendly codebook**
 - Communicate with PIs, statisticians and data users to know the plans and needs for future analysis
- Description of Missing Data**
- Basic data cleaning plan**
 - Screen for anomaly
 - Basic data distribution
 - Aggregate items into scale scores
 - Generate cutoffs for clinical scales
 - Remove identifiable information from the dataset
 - Other
- Secured data transfer plan**
- Data dissemination plan (if applicable)**



Basic Data Cleaning

- **Let's be friends! – Get to know your data**
 1. Are all variables in? – Screen through Variable View
 2. Are they ordered correctly in the dataset? – Screen through Variable View
 3. Are all variables labeled correctly? – Screen through “Label” in Variable View
 4. Are missing values defined? – Screen through “Missing” in Variable View
 5. Are they all within the normal range? – Run “Descriptives” at the item level
- **“Productivity is never an accident.” – Think ahead**
 1. Communicate with all data users.
 2. What kind of analysis will be conducted using these data?
 3. What software will be used to conduct the analysis?
 4. Missing data analysis will be conducted at which level? Item or Scale?
- **Let the game begin! – Basic cleaning**
 1. Create reverse scores – Use “recode” or “compute” function
 2. Create Scale scores – Mean or Total according to publication
 3. Generate clinical cutoff scores for scales – Use “recode”, “IF” or “DO IF” functions
 4. Compute Cronbach's α for each scale, divided by groups if applicable (men vs. women)
 5. Compile a reader friendly codebook
 6. Descriptive analysis – Use “Explore” to run Outliers, Percentiles, Normal distribution, Kurtosis, Skewness
 7. Calculate mean, SD of all scales, divided by groups if applicable (men vs. women)
 8. Missing data analysis – “NMISS” function or “Descriptives”
 9. If missing data are imputed, repeat Step 1 to Step 8
 10. Back up your dataset *before* and *after* the imputation
- **Houston: We are ready for takeoff. – Detach ID info**
 1. Sort by ID variable
 2. Separate ID from the rest of the dataset
 3. Use scrabble ID variables as matching variables – use “random.org” String variable generator
 4. Save ID info in separate files and be kept by different people (i.e., PIs, supervisors and project manager) to increase security