**Day 1: Data Processing**

**Session 1: Introduction to Stata and Data Import**

* **Objectives:** Familiarize participants with Stata, its interface, and basic commands for data import.
* **Topics:**
  + Introduction to Stata software and its interface
  + Navigating the Stata interface
  + Importing datasets (CSV, Excel, text files)
  + Using the import command
  + Initial exploration of datasets using list, describe, and summarize

**Exercise:**

stata

\* Import CSV file

import delimited "path\_to\_your\_file.csv", clear

\* Explore the dataset

list in 1/10

describe

summarize

**Session 2: Data Inspection and Basic Cleaning**

* **Objectives:** Perform initial data inspection and basic cleaning tasks.
* **Topics:**
  + Viewing data with browse and edit
  + Inspecting data using summarize, tabulate, and codebook
  + Identifying and handling missing values
  + Basic transformations (creating new variables, handling missing values with mvdecode and replace)

**Exercise:**

stata

\* Check for missing values

misstable summarize

\* Replace missing values

replace varname = 0 if missing(varname)

**Session 3: Data Types and Variable Management**

* **Objectives:** Understand data types and manage variables effectively.
* **Topics:**
  + Understanding variable types (numeric, string, etc.)
  + Converting variable types using destring and tostring
  + Encoding string variables
  + Labeling variables and values using label and labmask
  + Renaming variables with rename
  + Generating and modifying variables using generate and replace

**Exercise:**

stata

\* Convert string to numeric

destring varname, replace

\* Label variable and values

label variable varname "Description"

label define lblname 1 "Category 1" 2 "Category 2"

label values varname lblname

**Session 4: Sorting and Filtering Data**

* **Objectives:** Learn how to sort and filter data for analysis.
* **Topics:**
  + Sorting data with sort and gsort
  + Filtering data using keep and drop
  + Subsetting data using if conditions
  + Creating new datasets with preserve and restore

**Exercise:**

stata

\* Sort data by a variable

sort varname

\* Filter data

keep if varname == "condition"

drop if varname == "condition"

**Day 2: Data Cleaning**

**Session 5: Handling Outliers and Duplicates**

* **Objectives:** Identify and manage outliers and duplicate records.
* **Topics:**
  + Identifying outliers using statistical methods (e.g., z-scores)
  + Handling outliers with winsorize or truncate
  + Detecting duplicates with duplicates report and duplicates list
  + Removing duplicates with duplicates drop

**Exercise:**

stata

\* Identify outliers using z-scores

gen z\_score = (varname - mean(varname)) / sd(varname)

\* Drop duplicates

duplicates report

duplicates drop varlist, force

**Session 6: Data Transformation and Recoding**

* **Objectives:** Perform advanced data transformation and recoding.
* **Topics:**
  + Recoding variables with recode
  + Using egen for complex variable creation
  + Transforming data (e.g., log transformations)
  + Creating interaction terms

**Exercise:**

stata

\* Recode variable

recode varname (1/10=1) (11/20=2)

\* Generate interaction terms

gen interaction = var1 \* var2

**Session 7: Merging and Appending Datasets**

* **Objectives:** Combine datasets through merging and appending.
* **Topics:**
  + Merging datasets with merge
  + Understanding different merge types (1:1, 1:m, m:1, m:m)
  + Appending datasets with append
  + Handling merge conflicts and mismatches

**Exercise:**

stata

\* Merge datasets

merge 1:1 id using "other\_dataset.dta"

\* Append datasets

append using "other\_dataset.dta"

**Collapse**

**Session 8: Reshaping Data**

* **Objectives:** Learn how to reshape data for different analyses.
* **Topics:**
  + Reshaping data from wide to long format with reshape long
  + Reshaping data from long to wide format with reshape wide
  + Practical examples of data reshaping for time series and panel data

**Exercise:**

stata

\* Reshape data from wide to long

reshape long varprefix, i(id) j(time)

\* Reshape data from long to wide

reshape wide varprefix, i(id) j(time)

**Day 3: Data Analysis**

**Basics of Stata programming**

**Loop**

**Program (function) making**

**Local and global variables**

**Session 9: Introduction to Hypothesis Testing**

* **Objectives:** Understand the basics of hypothesis testing.
* **Topics:**
  + Formulating hypotheses
  + Types of errors (Type I and Type II)
  + Common hypothesis tests (t-test, chi-square test)
  + Interpreting p-values and test statistics

**Exercise:**

stata

\* Perform a t-test

ttest varname, by(groupvar)

\* Perform a chi-square test

tabulate var1 var2, chi2

**Session 10: Introduction to Regression Analysis**

* **Objectives:** Learn the basics of regression analysis.
* **Topics:**
  + Simple linear regression
  + Multiple linear regression
  + Interpreting coefficients and R-squared
  + Assumptions of regression analysis

**Exercise:**

stata

\* Simple linear regression

regress outcome\_var predictor\_var

\* Multiple linear regression

regress outcome\_var predictor\_var1 predictor\_var2

**Session 11: Advanced Regression Techniques**

* **Objectives:** Explore advanced regression techniques.
* **Topics:**
  + Logistic regression
  + Interaction terms in regression
  + Dealing with multicollinearity
  + Model diagnostics and validation

**Exercise:**

stata

\* Logistic regression

logit outcome\_var predictor\_var1 predictor\_var2

\* Interaction terms

regress outcome\_var c.var1##c.var2

**Session 12: Time Series Analysis**

* **Objectives:** Introduction to time series analysis and techniques.
* **Topics:**
  + Basics of time series data
  + Autocorrelation and stationarity
  + ARIMA models
  + Forecasting with time series data

**Exercise:**

stata

\* Set time variable

tsset timevar

\* Fit an ARIMA model

arima outcome\_var, arima(p,d,q)

\* Forecasting

predict forecast, dynamic(t+1)

**Materials and Resources:**

* **Presentation Slides:** Detailed slides covering each session’s topics.
* **Practice Datasets:** Sample datasets for hands-on exercises.
* **Do-Files:** Example do-files demonstrating various data cleaning and analysis techniques.
* **Reading Materials:** Supplementary materials and references for further reading.

This outline provides a comprehensive and structured approach to training on data processing, cleaning, and analysis using Stata. Feel free to adjust the sessions based on the specific needs and prior knowledge of the participants. If you need further details or specific content for any session, let me know!