Pre-SEM Workflow

Quality Control Steps for new SEM Data Sets

# Identify SEM Variables

Use the variable codes tab to identify SEM variables

Add these variable names to the SEMvars.csv file under header “PARK”

Add the variable descriptions to SEMvars-CODES.xls file under new table titled “PARK”

# Variable Specific QC

### ID:

Verify that all ID values are unique. No duplicates

### Zip:

Identify all blank values and change to “NA”.

Identify all values where respondent answered don’t know, and change values to “DK”.

All variables are case sensitive, so for surveys where “don’t know” responses are recorded as “dk”, they must be changed to “DK”.

Create a new column of zeros next to “zip” with the header “zip\_int”. Identify observations where respondent wrote in an international country name. For these observations, change “zip\_int” from zero to one.

### Local:

The local variable needs to be binary and equal: **Local { 1: local party, 0: otherwise }**

See PARK codebook for values and change if necessary.

## Expenditures

### DKexp:

This needs to be binary and equal: : **DKexp { 1: Don’t know expenditures, 0: otherwise }**

### expOtherSpec:

Generate a column of zeros adjacent to “expOtherSpec” with the header:

“expOtherCleaned”

Generate a column of “NA” ‘s adjacent to “expOtherCleaned” with the header:

“expOtherCleaned\_Note”

### expOtherCleaned:

For observations where respondent entered a value > 0 into “expOther”, use “expOtherSpec” to identify which expenditure group this value may be added to. Add this value to the corresponding expenditure group. For this observation, change “expOtherCleaned” from zero to 1.

### expOtherCleaned\_Note:

For the observations where “expOther” was added to an expenditure group, write a note briefly identifying the expenditure group and amount that was added.

### New Expenditure Column:

After sorting thought “expOtherSpec”, if a non-trivial number of observations write in an expenditure group which is not already identified, considered generating this expenditure variable. For example, if >5% of observations write in “marina fees” or “dock fees”, it is advisable to generate this expenditure column. If this is done, add the variable to SEMvars.csv under both the PARK variables and SEMvars columns (case sensitive).

## DaysPark:

If there is no variable present which corresponds to “daysPark”, generate a column of “NA”’s with the header “daysPark” and follow the steps below:

### NightsLocalArea:

Generate a column adjacent to “daysPark” with the header “nightsLocalArea”. The values in “nightsLocalArea” is the sum across the accommodation variables for overnight visitors, and “nightsLocalArea” equals 1 for day visitors (i.e. overnight == 1). These variables are identified in SEMvars.csv.

Now set “daysPark” equal to “nightsLocalArea”

### DaysLA:

Generate a column with the header “daysLA” adjacent to the set of accommodation variables. “daysLA” is “nightsLocalArea” + 1.

## HoursPark:

If there is no variable present which corresponds to “hoursPark”, generate a column of “NA”’s with the header “hoursPark”.

These values are to be filled with the mean hours visit for the PARK from IRMA.gov

### Entries:

If there is no variable present which corresponds to “entries”, generate a column with the header “entries” and is equal to “daysLA”

## DK Variables

Use the PARK variables codebook along with SEMvars.csv to identify variables where possible answers are “don’t know”. Where applicable, change the values in these variables to “DK” or “NA” rather than a numeric response such as “99”. Note: these variables are case sensitive.