# PSY 792F – SAS Assignment 1

# Building a SAS Dataset for Analysis

Objectives: Build a permanent SAS dataset with variables from the HRS Tracker File to examine demographic characteristics for respondents in the 2012 wave of the HRS and additional variables to facilitate analysis with some 2014 wave variables. Report basic sample characteristics

## Write a SAS program to do the following:

1. Read in “Tracker” dataset
2. Demonstrate that dataset was read in correctly
3. Obtain the following variables:

* HHID (Identifier)
* PN (Identifier)
* NSUBHH (Identifier)
* OSUBHH (Identifier)
* GENDER
* RACE
* HISPANIC
* Education (SCHLYRS, HIDEGREE)
* NCOUPLE
* OCOUPLE
* NAGE
* BIRTHYR
* BIRTHMO
* NIWMONTH
* NIWYEAR
* NIWWAVE (Sample indicator – in 2012 wave)
* OIWWAVE (Sample indicator – in 2014 wave)
* NRESCODE
* ORESCODE
* NLBWGTR
* NWGTR
* OWGTR

1. Run descriptive statistics to examine the following demographics characteristics for respondents who participated in the 2012 wave:

Frequencies:

* Age
* Gender
* Education
* Race
* Hispanic
* Coupleness

Means:

* Age

1. Recode extraneous values where needed
2. Create dummy coded variable for FEMALE, coupleness
3. Determine the range of time during which data were collected in the 2012 and 2014 waves
4. Save a permanent SAS dataset with only the variables needed for your analysis
5. Run PROC CONTENTS to check file

## HRS Codebooks:

Tracker file:

<http://hrsonline.isr.umich.edu/modules/meta/tracker/codebook/trk2014tr_ri.htm>

## What to turn in:

1. SAS Syntax file
2. SAS Output file
3. SAS Log
4. Worksheet with answers to the following questions:
5. How many participants are in the original Tracker file?

38,183

1. How many variables are in the original Tracker file?

444

1. How many respondents participated in the 2012 wave?

38,183

1. Of those who participated in 2012, report descriptive statistics for the following sample demographic characteristics. (Percentages, means as appropriate)

* Age range in sample
  + 20 - 103
* Age

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analysis Variable : AGE AGE AT 2012 INTERVIEW | | | | |
| N | Mean | Std Dev | Minimum | Maximum |
| 20554 | 66.8069962 | 11.5938949 | 20.0000000 | 103.0000000 |

* Gender

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GENDER | | | | |
| GENDER | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Male | 9156 | 41.83 | 9156 | 41.83 |
| Female | 12735 | 58.17 | 21891 | 100.00 |

* Race

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RACE/ETHNICITY | | | | |
| RACE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| White/Caucasian | 15800 | 72.38 | 15800 | 72.38 |
| Black or African America | 4221 | 19.34 | 20021 | 91.72 |
| Other | 1807 | 8.28 | 21828 | 100.00 |

* Hispanicity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HISPANICITY TYPE | | | | |
| HISPANIC | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Hispanic, Mexican | 1738 | 7.95 | 1738 | 7.95 |
| Hispanic, Other | 1139 | 5.21 | 2877 | 13.16 |
| Hispanic, type unknown | 35 | 0.16 | 2912 | 13.32 |
| Non-Hispanic | 18948 | 86.68 | 21860 | 100.00 |
| Missing = 31 | | | | |

* Education, including % who have a college degree

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HIGHEST DEGREE OF EDUCATION | | | | |
| DEGREE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| No Degree | 4379 | 21.34 | 4379 | 21.34 |
| GED | 1139 | 5.55 | 5518 | 26.89 |
| High School Diploma | 10275 | 50.07 | 15793 | 76.95 |
| Two Year College Degree | 2822 | 13.75 | 18615 | 90.70 |
| Master Degree | 1447 | 7.05 | 20062 | 97.75 |
| Professional Degree (Ph.D., M.D., J.D.) | 461 | 2.25 | 20523 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DEGREE2 | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No College Degree | 15793 | 76.95 | 15793 | 76.95 |
| Obtained Associates Degree or Higher | 4730 | 23.05 | 20523 | 100.00 |

* Coupleness

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2012 WHETHER COUPLED OR PARTNERED | | | | |
| COUPLED | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Yes | 13600 | 62.13 | 13600 | 62.13 |
| No | 8291 | 37.87 | 21891 | 100.00 |

1. How many participants have any missing data on the demographic characteristics listed above?

Age missing = 1,337

Gender Missing = 0

Race Missing = 63

Hispanicity Missing = 31

Highest Degree of education missing = 1,368

Coupled Missing = 0

1. How do the demographic characteristics change when the sample weights are applied?

The actual N’s and frequencies change to a large degree. Although the means, and percentages change as well, they are to a smaller degree of change. For example, the Mean, after weights added changed from 66.8 to 65.71. Additionally, the original percentages of males and females changed from 41.8% & 58.1% respectively to 45.95% and 54.05% respectively. Similar patterns can be seen across the other variables with weights added as well.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analysis Variable : NAGE AGE AT 2012 INTERVIEW | | | | |
| N | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 20554 | 65.7183091 | 668.6660619 | 20.0000000 | 103.0000000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GENDER | | | | |
| GENDER | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Male | 41680606 | 45.95 | 41680606 | 45.95 |
| Female | 49018154 | 54.05 | 90698760 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RACE/ETHNICITY | | | | |
| RACE | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| White | 75281931 | 83.13 | 75281931 | 83.13 |
| Black | 9254875 | 10.22 | 84536806 | 93.35 |
| Other | 6019746 | 6.65 | 90556552 | 100.00 |
| Frequency Missing = 142208 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HISPANICITY TYPE | | | | |
| HISPANIC | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Hispanic, Mexican | 4888149 | 5.39 | 4888149 | 5.39 |
| Hispanic, Other | 2824334 | 3.12 | 7712483 | 8.51 |
| Hispanic, type unknown | 82681 | 0.09 | 7795164 | 8.60 |
| Non-Hispanic | 82813225 | 91.40 | 90608389 | 100.00 |
| Frequency Missing = 90371 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DEGREE | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| No College Degree | 58922368 | 69.87 | 58922368 | 69.87 |
| Associates Degree or Higher | 25405982 | 30.13 | 84328350 | 100.00 |
| Frequency Missing = 6370410 | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2012 WHETHER COUPLED OR PARTNERED | | | | |
| COUPLED | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| Yes | 58950214 | 65.00 | 58950214 | 65.00 |
| No | 31748546 | 35.00 | 90698760 | 100.00 |

1. What were some of the challenges you faced in completing this assignment, if any?

I really wasn’t sure how to separate some of the variables. Otherwise, everything worked out great!