**Chapter 10 Dataset Codebook**

# ch10\_crc\_screening Dataset

The data come from BRFSS 2014. There are 250 observations in the dataset, each with a unique id and a value for both the female and screened variables.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| female | 1 = Yes  0 = No | Female sex indicator |
| screened | 1 = Yes  0 = No | Met USPSTF guidelines for colorectal cancer screening |

# ch10\_demographics Dataset

The data are from BRFSS 2014. There are 250 subjects in the dataset, each with a unique id. Gender is a two-level variable with options male and female. Race has three levels: Black, White, and other race. Each subject is either in the active treatment group or in the placebo group.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| gender | “Female”  “Male” | Gender of Subject |
| race | “Black”  “White”  “Other” | Race of subject |
| treatment | “Active”  “Placebo” | Treatment group |

# ch10\_diabetes Dataset

500 subjects from the COH study are in the ch10\_diabetes dataset. Each participant has a value of yes or no for self-reported diabetes and yes or no for an indication of diabetes on the medical record.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| self\_diabetes | 1 = Yes  0 = No | Self-reported having diabetes |
| mr\_diabetes | 1 = Yes  0 = No | Medical record indicates diabetes |

# ch10\_edu\_healthcare Dataset

The data come from NHANES 2013 – 2014. There are 349 subjects in the ch10\_edu\_healthcare dataset, each with a unique id. Education is a four level categorical variable with levels less than high school, high school graduate or GED, some college or associate’s degree, or college graduate or above. There is also a variable indicating if the subject has a routine place for healthcare.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| education | 1 = Less than high school  2 = High school graduate or GED  3 = Some college or associate’s degree  4 = College graduate or above | Education level |
| healthcare | 1 = Yes  0 = No | Has routine place for healthcare |

# ch10\_mh\_bullying Dataset

There are 450 subjects from the YRBS 2013 in the ch10\_mh\_bullying dataset. Each has a value for the unique identifier. There is a variable called sad indicating whether or not the student felt so sad or hopeless almost every day for two weeks or more in a row in the past year that he or she stopped doing some usual activities. There is also an indicator of whether or not he or she was bullied on school property in the past year.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| sad | 1 = Yes  0 = No | Felt so sad or hopeless almost every day for two weeks or more in a row in the past year that you stopped doing some usual activities |
| bullied | 1 = Yes  0 = No | Bullied on school property in the past year |