Lab C Dataset Codebook

# labc\_diet\_weight Dataset

The National Health and Nutrition Examination Survey (NHANES) asks, “In general, how healthy is your overall diet?” The response is a 7 level categorical variable ranging from “Excellent” to “Poor” with a “Don’t know” and “Refused” option. Those who reported having an excellent, very good, or good diet were classified as eating a healthy diet (good\_diet = 1). Respondents that selected “Fair” or “Poor” were given a value of good\_diet = 0. The body weight of each respondent is also measured during the survey and is captured in the variable weight. The sample in the dataset labc\_diet\_weight is additionally limited to those who selected “Male” as their gender. There are 216 observations, each with a unique id.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| good\_diet | 1 = Yes  0 = No | Indicator of good diet |
| weight | Numeric | Weight in kg |

# labc\_breast\_cancer Dataset

The data was simulated based on a data report from the CDC using NHIS data that 0.9% of Black women have ever been diagnosed with breast cancer. Thus 90 of the 10,000 “women” in the sample have a value of cancer = 1 indiciating that they have had breast cancer at some point. There is a unique id for each “woman.”

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| cancer | 1 = Yes  2 = No | Ever had breast cancer indicator |

# labc\_smoking\_bw Dataset

The National Health and Nutrition Examination Survey (NHANES) asks, “How much did he/she weigh at birth?” to caregivers of respondents age 15 and younger. Responses were recorded in pounds and ounces. Ounces were converted to fractions of a pound to create the variable birth\_weight (birth weight in pounds). The survey also asks, “Did the respondent’s biological mother smoke at any time while she was pregnant with him/her?” The response options are “Yes,” “No,” “Don’t know,” or refused to answer the question. The sample in the dataset labc\_smoking\_bw is limited to records where the mother was reported to be a smoker (mother\_smoked = 1). There are 100 records in the sample.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| id | Numeric | Unique identifier |
| mother\_smoked | 1 = Yes  2 = No | Mother smoked while pregnant indicator |
| birth\_weight | Numeric | Birth weight in pounds |