**Lab D Dataset Codebook**

# labd\_metro\_areas Dataset

The data come from the American Community Survey (ACS) 2010 – 2014 estimates. There are 388 metropolitan statistical areas in the dataset. Each have a character variable with the name of the metro area, a numeric variable with the unemployment rate, and an indicator of whether or not the area is a minority majority.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| metro\_name | Character | Name of metro area |
| unemployment\_rate | Numeric | Unemployment rate (%) |
| minority\_majority | 1 = Yes  0 = No | Indicator of metro area with > 50% minority residents |

# labd\_measles\_vaccinations Dataset

Data from the WHO’s Global Health Data Repository. The percentage of one-year-olds with measles immunization in 2000 and in 2014 is shown for 19 countries.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| country | Character | Country name |
| coverage\_2000 | Numeric | Percentage of one-year-olds with measles immunization in 2000 |
| coverage\_2014 | Numeric | Percentage of one-year-olds with measles immunization in 2014 |
| coverage\_difference | Numeric | Difference in measles immunization coverage from 2000 to 2014 (= coverage\_2014 – coverage\_2000) |

# labd\_measles\_vaccinations\_2 Dataset

Data from the WHO’s Global Health Data Repository. The percentage of one-year-olds with measles immunization in 2000 and in 2014 is shown for a sample of 19 countries. This is the same data that is in labd\_measles\_vaccinations, but in a different format. There are 38 lines in the dataset (two for each country).

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| year | Numeric, 2000 or 2014 | Year |
| vaccination\_coverage | Numeric | Percentage of one-year-olds with measles immunization |

# labd\_birth\_weights Dataset

The National Health and Nutrition Examination Survey (NHANES) asks caregivers of respondents age 2 to 15, “Has a doctor or health professional ever told you that he/she was overweight?” Possible responses are “Yes,” “No,” “Don’t know,” or refused to answer the question. Those who answered yes to the question were coded overweight\_kid = 1, and those who answered no were given a value of overweight\_kid = 0. The caregivers are also asked, “How much did he/she weigh at birth?” Responses were recorded in pounds and ounces, and the value was converted to pounds and fractions of a pound in the variable birth\_weight. The sample in the dataset labd\_birth\_weights is limited to those who had non-missing birth weights and who answered “Yes” or “No” to the question about being overweight as a child.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| birth\_weight | Numeric | Birth weight in pounds |
| overweight\_kid | 1 = Yes  0 = No | Indicator for doctor or health professional has told caregiver that child is overweight |

# labd\_vpa\_pulse Dataset

Same data as used in Problem 9.3, but with “yes” and “no” changed to a 0/1 coding scheme. The National Health and Nutrition Examination Survey System (NHANES) asks, “Do you do any vigorous-intensity sports, fitness, or recreational activities that cause large increases in breathing or heart rate like running or basketball for at least 10 minutes continuously?” Possible responses are “Yes,” “No,” “Don’t know,” or refused to answer the question. Those who reported engaging in a vigorous activity were coded as vpa = 1, and those who said they do not participate in VPA were coded as vpa = 0. The examination component of the survey measures the 60 second pulse rates of the respondents, and this information is contained in the variable pulse\_rate. The data given in the problem comes from a random sample of 32 adults who had measured pulse rates and answered “Yes” or “No” to the VPA question.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| pulse | Numeric | 60 second pulse rate |
| vpa | 1 = Yes  0 = No | Indicator for reports engaging in VPA |

# labd\_premature\_death Dataset

From the County Health Rankings & Roadmap datasets on Social Explorer. Data was pulled from 2013 and 2016 for all counties in the US, and a sample is in the dataset. The variables rate\_difference is the difference between the rate in the two years.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Values** | **Notes** |
| county | Character | County name |
| premature\_death\_rate\_2013 | Numeric | Number of deaths among residents under the age of 75 per 100,000 population in 2013 |
| premature\_death\_rate\_2016 | Numeric | Number of deaths among residents under the age of 75 per 100,000 population in 2016 |
| rate\_difference | Numeric | Difference in premature death rates from 2013 to 2016 ( = premature\_death\_rate\_2016 – premature\_death\_rate\_2013) |