Oct 10, 2105

Public Health Data Sets

1. **NHANES DATA**

The National Health and Nutrition Examination Survey (NHANES) is a cross-sectional survey that is conducted every two years in the United States. As part of the survey, individuals are asked to complete a demographics questionnaire and a 24 hour dietary recall. 24-hour dietary recalls have shown to be a valid and reliable method for describing usual dietary intakes of a population. The datasets provided contain demographic data and 24 hour dietary recall data for all individuals sampled in NHANES for the years 1999 thru 2011. In the early years (1999 and 2001) only 1 day of dietary information was collected. Later on, it was shown that validity can be improved by collecting 2 days of dietary data. So, for years 2003 thru 2011, you will see 2 days of dietary data.

For each year, individuals have a unique identifier called a sequence number. All files for that year can be linked by the sequence number (variable name: SEQN). Every year, different participants are sampled, so you cannot track the same person over time. But you can track the average population intakes over time. For example, these data are used to examine trends in total energy intake over time (measure in kilocalories) to see if the trends in increasing obesity rates correspond to similar trends in increasing energy intakes.

Currently, questions exist as to what the nutrient intake profiles of individuals look like? For example, do people who consume more saturated fats consume less fiber? Does this pattern differ by gender? Do nutrient profiles look different for older vs. younger individuals? What do time trends in nutrient patterns look like? The data used to answer these questions are the following NHANES datasets:

DATASETS

1. Demographics Data

YEAR\_DEMO (where YEAR is 1999, 2001, 2003, 2005, 2007, 2009 or 2011)

These are the demographic data for each participant. There is one line of data per participant. It includes age, race/ethnicity, gender, citizen status, marital status, income, poverty income ratio, number of people in the household, education, and length of time in the US.

Documentation on this data can be found at <https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DEMO_C.htm>

1. Reported foods data

YEAR\_DR1IFF

This is a data set that contains nutrient information for each food the subject reports. There are multiple lines of data for each subject and the number of lines per subject varies by the number of foods each subject eats. For example, if participant 1 ate 13 foods, then participant 1 will have 13 lines of data. If participant 2 ate 2 foods, then participant 2 will only have 2 lines of data.

YEAR\_DR2IFF (this is the same as YEAR\_DR1IFF, but is the second day of intake)

For 1999 and 2001, the data file is named 1999\_DRXIFF or 2001\_DRXIFF

Documentation on this data can be found at <https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DR1IFF_C.htm>

1. Overall dietary intake

YEAR\_DR1TOT

This is a dataset that contains the average nutrient intakes for each participant. There is one line of data per subject.

YEAR\_DR2TOT (this is the same as YEAR\_DR1TOT, but is the second day of intake)

For 1999 and 2001, the data file is named 1999\_DRXTOT or 2001\_DRXTOT

Documentation on this data can be found at <https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DR1TOT_C.htm>

1. Foods composition database

YEAR\_DRXFCD

This is a look-up table that gives the name of each reported food in the YEAR\_DRXIFF dataset. In the individual foods dataset, the variable name is DRDIFDCD. You will see that this has a numeric code. The look-up table has the label for each code.

For example, if the dataset has DRDIFDCD with a value of 57348000, then this look up table tells you that the food was frosted corn flakes, NFS (where NFS means any brand besides Kellogg).

Documentation on this data can be found at <https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DRXFCD_C.htm>

1. Medical Condition Data

YEAR\_MCQ

This is a dataset that contains answers to survey questions related to medical conditions. Each participant’s responses are recorded in a single line of the data. There are multiple columns that correspond to different survey questions. For example, column MCQ010 contains answers to “Have you ever been told you have asthma?”

Documentation on this data can be found at <https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/MCQ_C.htm>