Feb 26, 2021

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

DEMO\_YEAR (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.

1. Reported foods data

DR1IFF\_YEAR

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.

DR2IFF\_YEAR (this is the same as DR1IFF\_YEAR, but is the second day of intake)

For 1999 and 2001, the data file is named DRXIFF\_1999 or DRXIFF\_2001

1. Overall dietary intake

DR1TOT\_YEAR

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

DR2TOT\_YEAR (this is the same as DR1TOT\_YEAR, but is the second day of intake)

For 1999 and 2001, the data file is named DRXTOT\_1999 or DRXTOT\_2001

1. Foods composition database

DRXFMT\_YEAR

This is a look-up table that gives the name of each reported food in the DRXIFF\_YEAR 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).

1. **Medicaid-Vital Statistics Data**

The Medicaid and Medicare Administration in the State of Delaware (DMMA) examines medical usage and health outcomes of their clients on a regular basis. A report came out in 2014 that stated that individuals with mental illness were twice as likely to die and that they die at a much earlier age (almost 20 years earlier) compared to those without mental illness in the United States. DMMA wanted to know if this was true of their Medicaid population. The DSAMH\_Medicaid dataset is a subset of about 6000 individuals who have Medicaid as their primary insurance and have at least one instance of mental illness. These data were then linked to vital statistics to determine mortality. There were approximately 200 deaths. The dataset and codebook are provided.

QUESTIONS:

DSAMH\_Medicaid Dataset

1. What cause of death codes are most commonly reported in the DSAMH\_Medicaid dataset?
2. Do the cause of death codes cluster into groups by gender? Or by disability status?
3. Are there differences in the patterns of medical care (i.e., time spent in Medicaid, number of medical claims, number of hospital claims, number of emergency department claims, total billed amounts) reported by those that die versus those that didn’t die?

NHANES Dataset

1. What do the nutrient intake profiles of individuals look like? For example, do people who consume more saturated fats consume less fiber?
2. Do nutrient intake profile patterns differ by gender, age, race, education, or poverty level?
3. What do time trends in nutrient patterns look like?
4. How consistent are individuals in their consumption over two days? For example f they are consuming high amounts of fiber on day one, are they also consuming high amounts on day 2?