"Introduction to Statistics" Evaluation

Nhanes data

The National Health and Nutrition Examination Survey (NHANES) is a survey conducted as part of a research program of the National Center for Health Statistics (NCHS) to assess and track the health and nutritional status of adults and children in the United States. The survey combines demographic, socioeconomic, dietary, and health-related data. The data provided here are from NHANES prior to 2006.

Objectives

Each student will work on a different question among:

1. Prevalence of obesity: analysis of body mass index distribution in the population and in different demographic and ethnic groups, and exploration of factors associated with obesity.

2. Prevalence of hypertension in the population, different demographic and ethnic groups, and exploration of factors associated with hypertension. Hypertension may be defined from the diastolic and systolic blood pressure values.

3. Analysis of cholesterol levels in the population, different demographic and ethnic groups, and exploration of factors associated with cholesterol.

4. Prevalence of osteoporosis in the population and in different demographic and ethnic groups, and exploration of factors associated with obesity.

Evaluation

- The evaluation will be based on a summary report of the analyses conducted to meet the objective of the study. This document will be organized into the following main parts:

  1. an introduction presenting the context of the study, the structure of the sample (sample size, variables measured), and the biological questions motivating the study.

  2. the main lines of the analysis (descriptive statistics, confidence intervals, tests, ...) by focusing on the most relevant elements. The main results may be presented in numerical form or in the form of graphs or tables. All results included in the report must be commented on and interpreted.

  3. a conclusion: synthesis of the results, review of the encountered difficulties, perspectives.

- All `R` scripts are to be submitted with the report. The programs should have minimal comments and must be executable by only changing the working directory. Reports for which the R scripts do not comply with these instructions will not be evaluated. It is possible, but optional, to choose `Markdown` to combine the report with the R program in the final file you will submit.

- Report are individual and are due by **December 17th, 8pm**. They will be sent to `maud.delattre@ inrae.fr`.

List of variables

1. In `Nhanes1.csv`:

   - `SEQN` : respondent sequence number
   - `SEX` : gender (1 : Male, 2 : Female)

- **AGEYR** : age in years at screening
- **RACE** : race (1 : Mexican American, 2 : Other Hispanic, 3 : Non-Hispanic White, 4 : Non-Hispanic Black, 5 : Non-Hispanic Asian, 6 : Other Race – Including Multi-Racial)
- **PIR** : ratio of family income to poverty (0 to 4.99 : range of values, 5 : value greater than or equal to 5)
- **SALTTAB** : how often add salt to food at table (1 : rarely, 2 : occasionally, 3 : very often, 7 : refused, 9 : don't know)
- **SALTPREP** : salt used in preparation ? (1 : never, 2 : rarely, 3 : occasionally, 4 : very often, 9 : don't know)
- **TKCAL** : energy (daily aggregated, kcal)
- **TPROT** : protein (daily aggregated, gm)
- **TCARB** : carbohydrate (daily aggregated, gm)
- **TSUGR** : total sugars (daily aggregated, gm)
- **TFIBE** : dietary fiber (daily aggregated, gm)
- **TTFAT** : total fat (daily aggregated, gm)
- **TSFAT** : total saturated fatty acids (daily
- aggregated, gm)
- **TMFAT** : total monounsaturated fatty acids (daily aggregated, gm)
- **TPFAT** : total polyunsaturated fatty acids (daily aggregated, gm)
- **TCHOL** : cholesterol (daily aggregated, mg)
- **BPXSY** : Systolic Blood pres (mm Hg)
- **BPXDI** : Diastolic Blood pres (mm Hg)
- **WT** : weight (kg)
- **BMI** : Body Mass Index (kg/m**2)

2. In `Nhanes2.csv`:

- **SEQN**: respondent sequence number
- **GENDR**: gender (1 : Male, 2 : Female)
- **AGEYR**: age in years at screening
- **RACE**: race (1 : Mexican American, 2 : Other Hispanic, 3 : Non-Hispanic White, 4 : Non-Hispanic Black, 5 : Non-Hispanic Asian, 6 : Other Race – Including Multi-Racial)
- **BMXWT**: weight (kg)
- **BMXBMI**: Body Mass Index (kg/m**2)
- **TKCAL**: energy (daily aggregated, kcal)
- **TSFAT**: total saturated fatty acids (daily aggregated, gm)
- **TCALC**: calcium (mg)
- **d_milk**: number of milk cup equivalent
- **treatOSTEO**: osteoporosis status (1 : ever treated for osteoporosis, 2 : never told had osteoporosis or never treated for osteoporosis)
- **DAILYAVG**: average daily calcium supplement amount
- **ANYCALSUP**: used any supplement calcium