NHANES Complex Tabular Report

Merhawi Kidane

2025-09-27

Complex Tabular Report of NHANES Dataset

Introduction

The National Health and Nutrition Examination Survey (NHANES), conducted by the U.S. Centers for Disease Control and Prevention (CDC) through the National Center for Health Statistics (NCHS), is designed to assess the health and nutritional status of the U.S. population using interviews, physical examinations, and laboratory testing. The survey produces nationally representative data that inform public health research, policy, and program development[1]. For this analysis, we used data from three release cycles (2013–2014, 2015–2016, and 2017–2018), which are originally organized into two main groups: demographics and sample weights, covering information such as gender, age, race and Hispanic origin, education, marital status, country of birth, citizenship, and household/family income; and blood pressure measurements, including three consecutive systolic and diastolic readings, pulse, and related methodological information. Only a subset of variables from these two groups was selected and merged into a single cleaned dataset. During this process, the wave variable was created to consolidate the three survey cycles and subsequently renamed wave_id, with three year categories labeled 2015, 2016, and 2017, representing the original survey cycles. This allows each observation to be traced back to its original wave while combining the data from all three cycles into a cohesive dataset. In this report, we have generated tables summarizing key variables from the cleaned dataset to illustrate sample characteristics and blood pressure patterns across the survey waves.

1.Sample Size and Male Distribution by Wave

This table shows the total number of participants per wave, the number of males, and the percentage of males in each wave.

Table 1: Sample Size and Male Distribution by Wave

Wave	Total	Male Count	Male %
2015	10175	5003	**49.2**
2016	9971	4892	**49.1**
2017	9254	4557	**49.2**

Table 1 summarizes the sample size of participants in the NHANES dataset for each of the three survey waves (2013–2014, 2015–2016, and 2017–2018). It reports the total number of participants per wave as well as the count and percentage of male participants. Across all three waves, the proportion of male participants remains consistent at approximately 49%, reflecting an almost equal gender distribution within the survey population.

2.Ethnicity Distribution by Wave

The dataset contains two ethnicity variables: ethnicity_1 (RIDRETH1) and ethnicity_2 (RIDRETH3) [1]. We chose ethnicity_2 because it provides a more detailed classification, separating Non-Hispanic Asian participants into their own category, unlike ethnicity_1 which groups them as "Other Race". This allows for a more precise analysis.

Table 2: Ethnicity distribution by wave (number and percentage of participants).

Wave	Asian	Black	Mexican	Other	Other Hispanic	White
2015	1074 (10.6%)	2267 (22.3%)	1730 (17%)	470 (4.6%)	960 (9.4%)	**3674 (36.1\%)**
2016	1042 (10.5%)	2129 (21.4%)	1921 (19.3%)	505 (5.1%)	1308 (13.1%)	**3066 (30.7\%)**
2017	1168 (12.6%)	2115 (22.9%)	1367 (14.8%)	634 (6.9%)	820 (8.9%)	**3150 (34\%)**

Table 2 presents the distribution of NHANES participants by self-reported ethnicity across the three survey waves (2013–2014, 2015–2016, and 2017–2018). White participants consistently made up the largest proportion, although their share decreased slightly from 36.1% in 2015 to 30.7% in 2016, then increased to 34% in 2017. The proportion of Black participants remained relatively stable at around 22–23% across all waves. For Mexican and Other Hispanic groups, the percentage of Mexican participants rose from 17% in 2015 to 19.3% in 2016 before dropping to 14.8% in 2017, while "Other Hispanic" participants increased from 9.4% to 13.1% in 2016 and then decreased to 8.9% in 2017. Asian participants showed a gradual increase from 10.6% to 12.6%, and the "Other" category also increased slightly from 4.6% to 6.9% by 2017. Overall, the table highlights the diversity of the NHANES sample and the small shifts in ethnic composition across the three survey waves.

3.Summary Statistics of Average Systolic Blood Pressure by Wave

This table provides minimum, median, mean, maximum, and standard deviation of average systolic blood pressure for each NHANES wave.

Average Systolic Blood Pressure by Wave

To prepare the data for analysis (the next table), a new variable representing the average systolic blood pressure (SBP) was created for each participant. This was computed as the mean of the four available SBP measurements, ignoring missing values so that the average could still be calculated if at least one reading was present. This step ensures that each participant has a single, representative SBP value for subsequent summary statistics and tables.

Table 3: Summary statistics of average systolic blood pressure by wave.

Wave	Min	Median	Mean	Max	SD
2015	64.7	115.3	**118.3**	228.7	18.1
2016	74.0	117.3	**120.4**	231.3	18.4
2017	72.7	118.0	**121.7**	234.0	20.3

Table 3 summarizes the distribution of average systolic blood pressure across the three NHANES survey waves. Median and mean systolic pressures remain relatively stable over time, hovering around 115–118 mmHg for the median and 118–122 mmHg for the mean. The minimum values rise slightly from 64.7 mmHg in the 2015 wave to over 72 mmHg in 2017, while the maximum recorded pressures are consistently high (228–234 mmHg). Standard deviations (18–20 mmHg) indicate a comparable spread of values across waves. Overall, the figures show that the distribution of systolic blood pressure is very similar across the three survey cycles, with only small differences from year to year.

Conclusion

In summary, the cleaned NHANES dataset provides a clear overview of the study population across three survey waves (2015, 2016, 2017). Table 1 shows that the overall sample size is fairly consistent across waves, with a slight decrease in 2016. Table 2 highlights the distribution of ethnic groups, with Non-Hispanic White participants representing the largest proportion in all three waves and the other ethnic groups remaining relatively stable. Table 3 presents the summary statistics of average systolic blood pressure, demonstrating a similar distribution across waves with minor variations in mean and median values. Overall, the data distributions across waves appear consistent, supporting comparability for analyses across the three survey cycles.

Packages used

The tables in this report were generated using the kableExtra package in R, which allows for advanced table formatting and LaTeX integration [2]. Other packages used for data manipulation, such as dplyr, have been employed in previous exercises and are not cited separately here.

References

[1] Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey (NHANES)*. https://www.cdc.gov/nchs/nhanes/index.htm. Data from the 2013–2018 NHANES survey cycles. Accessed 2025-09-28. 2013.

[2] Hao Zhu. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.3.4. 2025. URL: https://haozhu233.github.io/kableExtra/.