Final Written Report (NHANES)

Ulises Bustos

1. Background/Motivation for the Study

Obesity is a leading U.S. risk factor for type 2 diabetes through insulin resistance (Centers for Disease Control and Prevention [CDC], 2023). The NHANES dataset, maintained by the CDC, provides nationally representative health data suitable for examining relationships between BMI and chronic disease outcomes (CDC, 2023).

2. Research Question and Hypothesis

Research Question: Is there an association between BMI category (Normal vs Obese) and diabetes prevalence among U.S. adults?

Null Hypothesis (H_0):

$$H_0: p_{\text{obese}} = p_{\text{normal}}$$

Alternative Hypothesis (H_1):

$$H_1: p_{\text{obese}} > p_{\text{normal}}$$

3. Data Description and Exploratory Data Analysis

3.1 Setup & Data

To answer the research question, I filter the NHANES dataset to include only adults (Age ≥ 18). I then create a categorical variable for BMI with two groups: *Normal weight (18.5–24.9)* and *Obese (≥30)*. Individuals classified as underweight or overweight are excluded to simplify the comparison. The Diabetes variable is recoded as a binary factor with levels "Yes" and "No." This cleaned dataset is then used for summary statistics and visualizations to describe the sample and explore associations between BMI classification and diabetes prevalence.

3.2 Data Cleaning

I next summarize the analytic sample by reporting the total sample size, distribution of BMI (typical values and spread), and diabetes prevalence overall and within BMI groups. These descriptive summaries provide initial evidence for the hypothesized association between obesity and diabetes and satisfy the Module 5 "Data Overview" requirement (sample size, means/SD, counts/proportions).

3.3 Descriptive Statistics

```
$sample_size
[1] 4833

$bmi_summary
# A tibble: 1 × 7
  mean_BMI sd_BMI min_BMI q1_BMI median_BMI q3_BMI max_BMI
     <dbl>  <dbl>   <dbl>  <dbl>      <dbl>  <dbl>   <dbl>
1     29.7   7.89    18.5   22.9       30.6   34.4    81.2

$bmi_groups
# A tibble: 2 × 3
  bmi_grp                n  prop
  <fct>              <int> <dbl>
1 Normal (18.5–24.9)  2198 0.455
2 Obese (≥30)         2635 0.545

$overall_diabetes_prev
[1] 0.109042

$diabetes_prev_by_bmi
# A tibble: 2 × 4
  bmi_grp                n diabetics prevalence
  <fct>              <int>     <int>      <dbl>
1 Normal (18.5–24.9)  2198       100     0.0455
2 Obese (≥30)         2635       427     0.162
```

The analytic dataset included **N = 4833** adults. BMI was **29.7 ± 7.9 kg/m^2**, with a median of **30.6** (IQR **22.9–34.4**) and a range of **18.5–81.2**. The BMI group sizes were **Normal weight** (n = 2198, 45.5%) and **Obese** (n = 2635, 54.5%). Overall diabetes prevalence was **10.9%**. Within BMI groups, diabetes prevalence was **4.5%** among normal-weight adults versus **16.2%** among obese adults, providing initial descriptive support for the hypothesis.

3.4 Visualizations and Tables

3.4.1 Visualization A — Diabetes Prevalence by BMI Group

I visualize diabetes prevalence by BMI group to directly compare proportions across the two categories required by the research question.

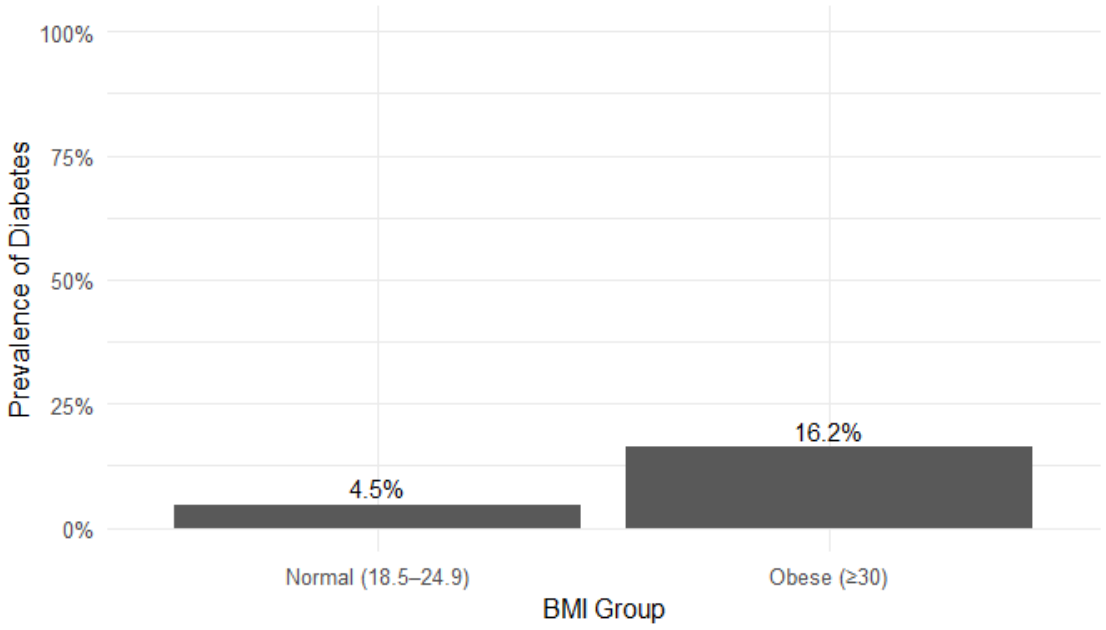

Figure 1: Diabetes Prevalence by BMI Group (Adults ≥18)

Interpretation: The obese group shows a higher prevalence of diabetes than the normal-weight group, consistent with the stated hypothesis and the descriptive tables.

3.4.2 Visualization B — BMI Distribution by Diabetes Status

I also examine BMI distributions by diabetes status to contextualize differences in central tendency and spread.

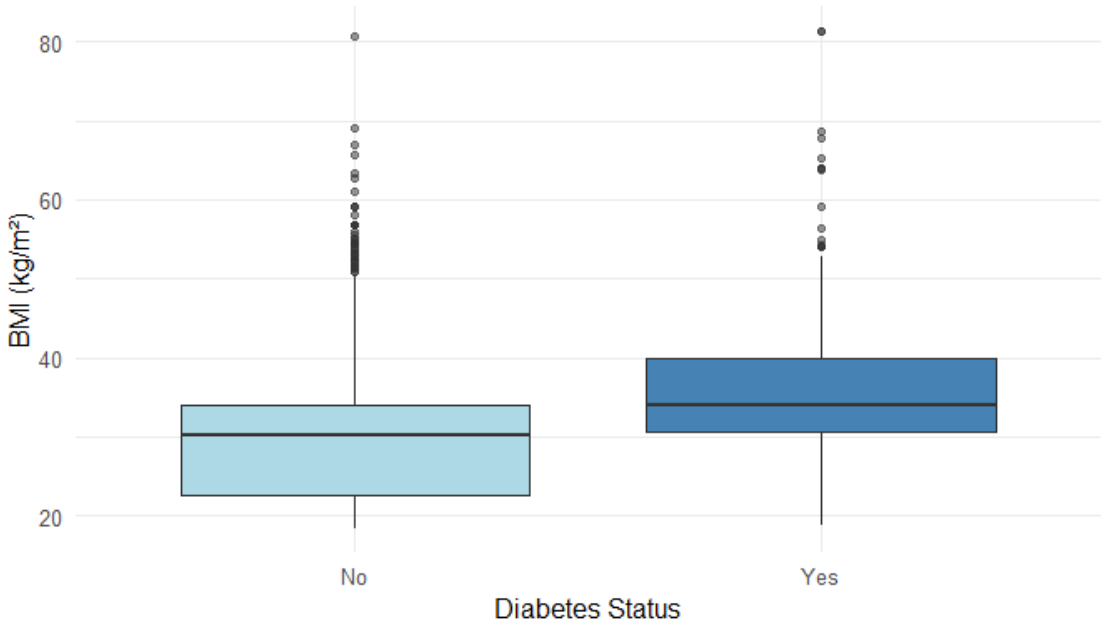

Figure 2: BMI Distribution by Diabetes Status (Adults ≥18)

Interpretation: Median BMI is higher among adults with diabetes, and the upper quartile extends into obese ranges, which is consistent with the expected direction of the association.

3.4.3 Nicely Formatted Prevalence Table

```
# A tibble: 2 × 4
  `BMI Group`            N `Diabetes Cases` Prevalence
  <fct>              <int>            <int> <chr>
1 Normal (18.5–24.9)  2198              100 4.5%
2 Obese (≥30)         2635              427 16.2%
```

Interpretation: The table summarizes sample size, diabetes cases, and prevalence within each BMI group and mirrors the differences visualized above.

4.1 Statistical Methodology

To compare diabetes prevalence between BMI categories (Normal vs Obese), a two-sample test for equality of proportions was used. This test is appropriate for binary outcome data (Diabetes = Yes/No) across two independent groups.

4.1.1 BMI Distribution Check

Before conducting inferential tests the BMI distribution needs to be examined for normality and to detect any possible skewness or extreme values. The process verifies statistical test requirements and shows how different groups relate to each other.

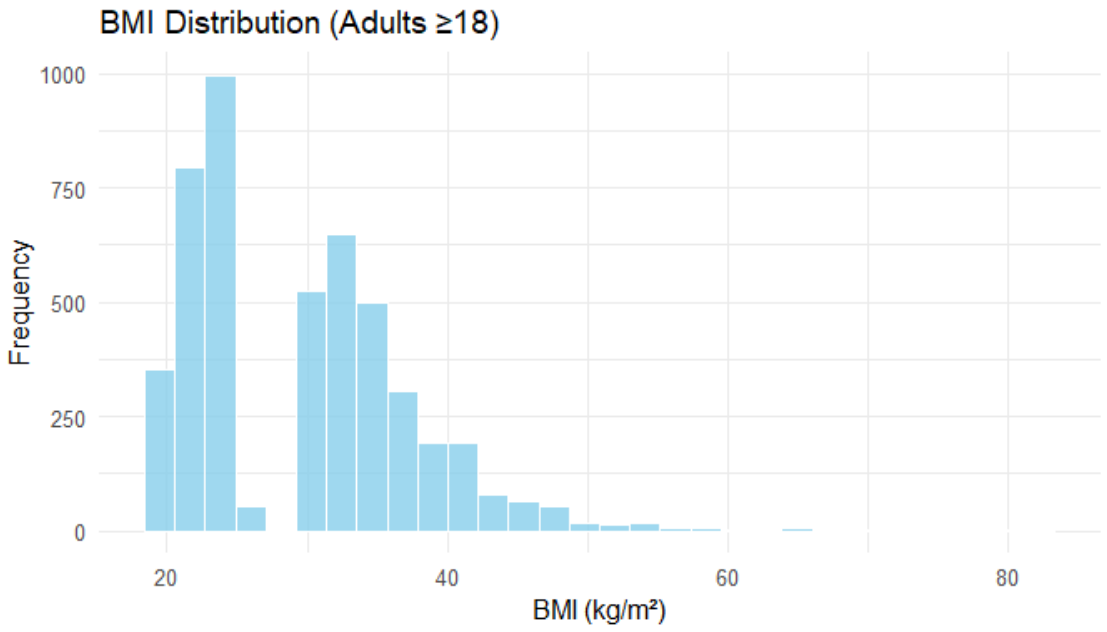

Figure 3: Histogram of BMI Distribution among U.S. Adults (≥18 years)

Interpretation: The histogram displays a right-skewed pattern in BMI values which proves that most adults maintain normal weight levels but only a few people carry extreme BMI measurements. The data set lacks perfect normality but the Normal and Obese groups remain comparable because BMI values were divided into categories for the Chi-square test.

4.2 Test Results

```
    2-sample test for equality of proportions with continuity correction

data:  tbl
X-squared = 166.37, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.09958755 0.13351931
sample estimates:
   prop 1    prop 2
0.9545041 0.8379507
```

The two-sample test for equality of proportions demonstrated a significant difference in diabetes prevalence between BMI categories according to $\chi^2(1, N = 4833) = 166.37$ and $p < 0.001$. The difference in proportions between obesity and diabetes incidence showed a 95% confidence interval from 0.10 to 0.13 which provides strong evidence that obesity is associated with a major increase in diabetes risk. The research data showed obesity-related diabetes occurred in 16.2% of obese people but only 4.5% of people who maintained normal weight. We reject the null hypothesis because the data provide strong evidence that obesity is associated with higher rates of diabetes

5. Conclusions

The statistical evaluation proves that diabetes occurs at a higher rate in obese adults than in normal weight individuals which matches the findings of earlier studies. The results support the alternative hypothesis which predicts that obesity leads to an elevated risk of developing diabetes. Key take-aways: maintaining healthy BMI may reduce diabetes risk; NHANES data effectively demonstrates this public-health relationship. The research supports the need for public health programs which focus on weight control to decrease diabetes risk among American adults.

References

Centers for Disease Control and Prevention (CDC). (2023). National Health and Nutrition Examination Survey (NHANES). U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/nhanes/ Centers for Disease Control and Prevention (CDC). (2023). Overweight and Obesity. https://www.cdc.gov/obesity/index.html