Final Project:

Complex Survey Design:

Dataset: NHIS 2023 Data (Sample Adult Interview)

Dataset is available here:

https://www.cdc.gov/nchs/nhis/documentation/2023-nhis.html

Codebook Available here:
https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2023/adult-codebook.pdf

Using the dataset provided (NHIS 2023 Dataset)—conduct an analysis of one of the following outcome variables:

- Ever had hypertension (HYPEV_A) (or questions on those who have it)
- Ever had high cholesterol (CHLEV_A) (or questions on those who have it)
- Ever had asthma (ASEV_A) (or questions on those who have it)
- Ever had cancer (CANEV_A)
- Repetitive strain injuries in last 3 months (REPSTRAIN_A) (or questions for those who had them)

I am open to additional outcomes upon request. Just discuss it with me (must be able to be made binary).

Create a binary variable of yes vs. no for your outcome (missing for Refused/Not Ascertained, Don't Know, Missing)

As part of your analysis, you should:

a. Identify a primary outcome and primary exposure variable of interest.
b. Develop a scientific research question answered with a statistical model (adjusted model).

   You should select at least 5 variables other than your outcome variable; 1 of 5 is primary exposure; Recommend using only categorical variables or categorizing them.

   A single variable includes all categories within it.

c. Analyze the data using appropriate complex survey design procedures--creating a descriptive table 1 and display model results in a table.

- You probably want to create a survey design object and subset the complex survey design down to a subpopulation of interest. I recommend that you eliminate missing data on outcome variable.
- Table 1 must contain all 6 variables. Display Overall column, Yes/No columns for binary outcome variable; within row percentages. Display raw N, weighted N, and w%.
- Perform significance tests for table 1.
- Model results table should contain odds ratios, p-values for an adjusted model minimum.

d. Summarize your research findings (methods and results/discussion sections). You do not need a thorough literature review.

Place your tables in the body of your report. ***This report should be written for a general public health audience like it would for a scientific manuscript. Do not use bulletpoints or incomplete sentences. Correct grammar, spelling etc. are expected.***

Grading Criteria:

Criteria	Details	Point total possible
R/Rstudio Code	Code provided and annotated; code will be rerun to ensure it works. Data management was performed adequately to allow variables to be well-labeled.	1 point
Analysis Procedures	Correct modeling techniques are used (accounting for survey design etc.). Descriptive statistics obtained correctly; modeling results obtained correctly Correct approaches are used throughout and match what is reported in the report.	2 points
Tables	Table 1 is presented formally. It contains all 6 variables, and statistical tests. Appropriate statistics are included and calculated. Modeling results are displayed appropriately. This must include the final adjusted model but models for unadjusted models may also be present. Table captions must be present. All tables should be well organized and look professional.	2 points
Methods	This section should include the following subsections: **Dataset/Complex Survey Description:** • Provide a description of the dataset used for analysis • Provide a description of the sampling strategy and what the sample is designed to represent • Provide a statement describing how survey weights were created (what they were based on). • Cite NHIS literature here for this (on dataset). • Indicate any further inclusion/exclusion criteria	2 points

	that you use in your analysis • Define how you defined your primary outcome variable • Define other variables used in your analysis (do not say 1 was coded as this etc. write it like you would for a scientific journal). **Statistical Methods:** • Describe the descriptive statistical analyses performed (including mentioning statistical tests) • Indicate that analyses accounted for complex survey design. • Describe the modeling strategy you used and any steps you used to build your final adjusted model. • Report types of standard errors used for analyses. • Indicate that you used R (cite survey package)	
Results/Discussion	Must include the following: • Table 1 • Describe results in the table 1 • Modeling Results Table • Interpret the modeling results for a naive public health audience (no statistical jargon like null hypothesis). Make sure you answer your scientific research question • Indicate implications of these findings (discuss why this is important). • Identify limitations and assumptions of your statistical modeling strategy.	3 points
TOTAL		**10 points**

I will run your R code. If evidence exists that you plagiarized (including had AI generate your whole code for you; example: code does not work or generate the results you report), this will

result in a 0 on the assignment. Make sure you cite your sources. Interpretations/writing should be written in your own words!

Note: the package tableone is not allowed. Use the commands as taught in class via the survey package for all calculations.