Assignment: Analysis of Weighted Survey Data

Background

This assignment is based on the 2022 National Health Interview Survey. For those who are interested, a full description of the 2022 survey ("2022 NHIS Survey Description") as well as the related weighting procedures ("2022 NHIS Paradata File") are provided on the course website. The ASCII files were downloaded and prepared for analysis.

This assignment is based on the "sample adult" file. This file was derived from data collected from one adult household member randomly selected from each surveyed household. The "2022 Sample Adult Codebook" (available on the course website) provides details on all of the variables available in the complete file.

The dataset that will be used for this assignment (NHIS 2022 Adult Data.csv) has been cut down from the complete file, and contains the following variables:

hhx	Household identifier	
wtfa_a	Final annual weight	
region	Household region	1=Northeast 2=Midwest 3=South 4=West
pstrat	Stratum for public use	
ppsu	Primary Sampling Unit for public use	
agep_a	Age of sample adult	

The following variables were created from existing variables:

sex	Created from the original sex_a	0=Female 1=Male
maxed	<=High School *vs*. Some College+, Created from the original maxeducp_a	0=High School graduate or less 1=Any college or more
hypertension	Ever had hypertension, created from hypev_a	0=No 1=Yes
partner	Partner living in home, created from marital_a	0=Not married or living together 1=Married or Living together

Statistical Program

We are proving the dataset in csv format. Feel free to use the program of your choosing. All numbers have been checked, and any numbers you are asked to provide in this assignment are identical (to the level of granularity needed for this assignment) in R and Stata. If you choose to use another statistical program instead, unfortunately we can't guarantee that the numbers will correspond exactly.

Missing Data

For ease of use, we will be ignoring the missing values for this assignment. Please report population proportions and other analytic measures without attempting to account for the missing values in any way.

Assignment Note

Make sure your software knows the design specifications for this weighted dataset so that the numbers are analyzed appropriately. The following variables are important:

Variable Name	Variable Description
pstrat	Stratum
ppsu	Primary Sampling Unit
wtfa_a	Weight (final annual weight)

As the NHIS material explains, the pstrat and ppsu variables are actually simplified version of the true sample design variables (to preserve confidentiality of participants). Much of the design details are not publicly released, and the NHIS analysts use various clustering, collapsing and partitioning techniques to create the pstrat and ppsu variables we use.

Questions 1-4: (8 points)

Fill in the unweighted and weighted proportions (not percents, accurate to 3 decimal places) of individuals reporting having every had hypertension vs. having normal blood pressure.

Hypertension	Unweighted Proportion	Weighted Proportion
Normal Blood Pressure	0.632	0.680
Hypertension	0.368	0.320

Question 5-6 (4 points):

Using weighted data, of all individuals reporting ever having had hypertension, what proportion reported having a partner (married/partnered) (give the proportion and standard error, again accurate to 3 decimal places).

Proportion: 0.618

Standard Error: 0.006

Report the Design Effect (DEFF) for this estimate: 1.432

Question 7-8: (4 points)

Using weighted data, among those reporting ever having had hypertension, what is the mean age of males (give the mean and standard error, again accurate to 3 decimal places).

Mean: **59.116**

Standard Error: **0.266**

Question 9-10: (6 points)

Using weighted data, visually depict the proportions of individuals reporting having a high school education or less *vs.* at least some college education within groups of those reporting having Normal Blood Pressure *vs.* Hypertension.

In 2-3 sentences, describe the message conveyed in the graph:

Among individuals who reported having normal blood pressure, the proportion of individuals had a high school education or less is 0.222, smaller than the proportion of at least some college education (0.778). In contrast, among those reporting hypertension, the proportion of individuals had a high school education or less was 0.291, while 0.709 had at least some college education. This indicates there are more individuals with lower education in the hypertension group.

Question 11-13: (8 points)

You have decided that the optimal estimates from your survey are derived from regression estimates using weighted data. The outcome of interest for this analysis is "Ever having had Hypertension", with "Hypertension" (as opposed to "Normal Blood Pressure") being considered suboptimal. You hypothesize that having a partner in the house (married or living together) is protective against hypertension as you are more likely to motivate each other to follow healthy living habits. Calculate the odds ratio for having Hypertension among those reporting being married or living with a partner, controlling for sex as a binary variable (male vs female), region (as a categorical variable), and maximum education as a binary variable (high school or less vs some college or more). Answer the following questions (to an accuracy of 3 decimal places):

Odds Ratio: **1.126**
p value:
     0.000241

In 2-3 sentences, compare the interpretation you make from this analysis with what you would have concluded had you done your analysis without accounting for weights in your data. Give the unweighted OR and p value in your answer.

When we use the weighted data, the risk of ever having had hypertension for people have a partner in the house is 1.126 times those without a partner, with p value less than 0.05 after control for sex, region, maximum education. This result doesn't support the hypothesize and shows that people having a partner are more likely to have hypertension.

When we use the unweighted data, the unweighted OR for hypertension is 0.866 with P value 3.35e-08. The odds for hypertension among those report having a partner is 0.866 compared to those without a partner, with a p value equals to 3.35e-8. This result support the hypothesize that having a partner in the house is protective against hypertension