|  |
| --- |
| **CSDE 502 Winter 2020**  **Assignment 8**  **Introduction to Add Health**  **Instructor: Phil Hurvitz**  [**phurvitz@uw.edu**](mailto:phurvitz@uw.edu) |

**Due Date: 2021-03-04 09:00**

**Explanation:**

In this assignment, there are two parts. The first part will continue the variable creation we started in assignments 1 and 2. The second part will analyze summary measures generated from the Knowledge Quiz section of the In-home Questionnaire from Add Health.

1. More variable creation

Imagine a new variable: *multirace*, using the following value definitions:

1 = one race, White

2 = one race, not White

3 = two races, includes White

4 = two races, both non-White

5 = three or more races, includes White

6 = three or more races, all non-White

9 = any race missing (White, Black/African American, American Indian, Asian, other)

1.1 Fill in the codes for the hypothetical cases below.

Table 1: Race coding

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **white** | **black** | **ai** | **asian** | **raceother** | **multirace** |
| 1 | 0 | 0 | 0 | 0 |  |
| 0 | 1 | 0 | 0 | 0 |  |
| 1 | 0 | 0 | 1 | 0 |  |
| 0 | 1 | 1 | 0 | 0 |  |
| 1 | 1 | 0 | 1 | 0 |  |
| 0 | 1 | 0 | 0 | 1 |  |
| 0 | 1 | 1 | 0 | 1 |  |
| 1 | 0 | 1 | 0 | 0 |  |
| 1 | 1 | 1 | 0 | 1 |  |
| 6 | 1 | 8 | 1 | 6 |  |

You can paste this text into your R Markdown file

1.2 Create an R data frame matching Table 1 with the code:

dat <- structure(list(white = c(1L, 0L, 1L, 0L, 1L, 0L, 0L, 1L, 1L,

6L), black = c(0L, 1L, 0L, 1L, 1L, 1L, 1L, 0L, 1L, 1L), ai = c(0L,

0L, 0L, 1L, 0L, 0L, 1L, 1L, 1L, 8L), asian = c(0L, 0L, 1L, 0L,

1L, 0L, 0L, 0L, 0L, 1L), raceother = c(0L, 0L, 0L, 0L, 0L, 1L,

1L, 0L, 1L, 6L), multirace = c(NA, NA, NA, NA, NA, NA, NA, NA,

NA, NA)), class = "data.frame", row.names = c(NA, -10L))

Using this data frame, what command(s) could list cases that checked more than one race?

1.3 How many cases checked more than one race? Include the R code that produced the answer to this question.

1.4 Write R code to create the *multirace* variable (paste below).

*Hint: You may want to create another variable, numrace, that counts the number of races.*

1.5 Translate this code into R (shown in your well-commented .R file). Use [AHwave1\_v3.rds](https://canvas.uw.edu/files/61620368/download?download_frd=1) from the Canvas site.

1.6 Label the *multirace* variable as well as its values using labelled::var\_label() and labelled::labelled() (shown in your well-commented .R file).

1.7 Paste below a contingency table of the *multirace* variable. Make sure that the values are labelled so the table is readable, and also include any missing values.

2. Review part B of each of the answers (i.e., H1KQ1B ... H1KQ10B) to the Knowledge Quiz (Section 19 of the Add Health questionnaire, documented in INH19PUB.PDF). The 10 questions each ask: “How confident are you that your answer is correct?”

2.1 Write R code that creates a single summary variable named *kqconfidence*, with a larger number representing the respondent being more confident across all questions (scale of 0 to 3 for each individual question; *kqconfidence* will be the sum for each subject across the 10 questions). Note that any observations with value 7 (age less than 15) should be removed from the data frame, and values 6, 8, and 9 should be coded as NA (i.e., missing) for the purposes of scoring confidence. Document your code so that the reader knows how you scored the scale and how you handled missing values. Make sure to label the new variable. (Include your well-documented .R file.)

2.2 Create and paste below a contingency table from *kqconfidence* with raw counts, percentages, and cumulative percentages. Put code to do this in your .R file.

2.3 **[BONUS]** For each subject there were zero to 10 “missing” answers to each of the 10 component questions. We would like to know what this distribution is. Paste below a table that shows the count of subjects for each unique value of the count of missing questions, and include code in your .R file.

2.4 For each possible value of the Knowledge Quiz Part A *score* (from Assignment 3), what is the mean *kqconfidence* level? (Paste results below and put code in your .R file.)

2.5 **[BONUS]** For each respondent, create two different confidence scores: a confidence score for the items answered correctly and a confidence score for the items answered incorrectly. How many respondents are more confident when answering incorrectly?