

Problem Statement

Organization Name: The U.S. National Cancer Institute (NCI)

Dataset Name: Health Information National Trends Survey (HINTS5, Cycle 3)

Difficulty: ⚙️ ⚙️ ⚙️ ⚙️

Level 4: Participants with advanced data analysis skills.

The problem statement is open-ended and require multitudes of analytical perspectives and visualizations. Statistical modeling is highly recommended.

Background

The U.S. National Cancer Institute (NCI) has been conducting the Health Information National Trends Survey (HINTS) since 2003 to learn about U.S. adults' cancer-related perceptions and knowledge, their health behaviors, and their health-related information access, needs, seeking, and use. This survey is administered every few years to civilian, non-institutionalized adults in the U.S.

Questions

1. Are there any relationships between people's self-rated health (question F1) and their health-related information behavior [e.g., whether they have ever looked up health information using any type of source (A1);
 - a. what type of source they turned to the most recent time they looked for health information (A2);
 - b. the barriers they faced while searching for health information (A4);
 - c. where they would turn if they had a strong need to get health information (A7);
 - d. whether they have used a computer, smartphone, or electronic means to look for health information (B5a);
 - e. and whether they have shared health information on social networking sites, such as Facebook or Twitter (B14b)]?
2. How do people's information-related self-efficacy (A5) and their health-related self-efficacy (F2 & N4) relate to
 - a. their use of social media for health information seeking or sharing (B14),
 - b. their self-reported health status (F1),
 - c. their avoidance of doctors (F3),
 - d. and their perceived social support (F5)?
3. How does feeling depressed or anxious (F11) relate to people's health-related information seeking [see RQ 1 above for suggestions of relevant survey questions]? For example, do people who feel depressed or anxious report encountering more barriers during their search for health information?

4. How do people's trust profiles (which can be calculated based on their responses to question A6) relate to:
 - a. (1) various demographic factors [e.g., age (O1),
 - b. educational attainment (O3),
 - c. income (O12),
 - d. race (O6),
 - e. sexual orientation (O7)];
 - f. (2) whether or not respondents have used a computer, smartphone, or other electronic means to look for health information for themselves or someone else during the past year (B5a);
 - g. (3) respondents' perceptions regarding the healthcare they have received over the past year (C3 & C6);
 - h. and (4) their participation in cancer screenings (L1 – L4)?
5. How do respondents who reported having health and wellness apps on their tablet or smartphone (B7) differ from those who do not have these apps based on factors such as:
 - a. (1) information-related self-efficacy (A5);
 - b. (2) health-related self-efficacy (F2 & N4);
 - c. (3) Self-reported health status (F1);
 - d. (4) barriers they faced while searching for health information (A4)
 - e. (5) their perceptions regarding the healthcare they have received over the past year (C3 & C6);
 - f. (6) their participation in cancer screenings (L1 – L4)?

Data Considerations

The most recent HINTS dataset available is HINTS 5, Cycle 3. This dataset is dated January 2020, but the data was actually collected between January 22 and April 30, 2019. The file includes data from a total of 5,247 completed questionnaires and 191 partially completed questionnaires, for a total of 5,438 questionnaires.

This dataset, along with supporting documentation (pdf files), can be downloaded in either SPSS, STATA, or SAS format here: <https://hints.cancer.gov/data/download-data.aspx>. The excel version is attached along with the supporting documentations, including:

- *HINTS5_Cycle3_Annotated_Instrument_English.pdf*: This is the actual survey instrument. It shows not only the specific questions respondents answered, but also how each question has been labeled in the data file and how the associated response options have been coded. For example, question A5, "Overall, how confident are you that you could get advice or information about health or medical topics if you needed it?" has been labeled "ConfidentGetHealthInf" in the data file, and the response options have been

coded as follows: 1 = Completely confident; 2 = Very confident; 3 = Somewhat confident; 4 = A little confident; 5 = Not confident at all.

- *HINTS 5 Cycle 3 Public Codebook.pdf*: This codebook provides a detailed description of the contents of the data file, including for each question, the full text of the question (and associated variable label and format) and the response options (and the associated codes). It also includes descriptive statistics for each question/response option, based on both unweighted and weighted sample sizes.
- *HINTS5 Cycle 3_Methodology Report.pdf*: This report details the methods used to conduct the survey and to prepare the resulting data for statistical analysis.