# Data Exam Two

## 36-402 Advanced Methods for Data Analysis

### Due Friday, May 5, 2023 at 4pm EDT

Medical care for serious diseases can be very expensive, and in countries without socialized medical systems it can impose serious hardships on patients. Ideally, patients would get regular check-ups (or "general health examinations," GHEs) so that serious conditions could be detected and treated early, before they cause serious problems. This would save money and improve public health.

However, there are many possible obstacles to getting everyone to go to regular check-ups. They might be too expensive, or too difficult to schedule; some people may not trust doctors or believe that check-ups have any value; or some people may have had bad experiences when they previously had check-ups.

Public health researchers in Vietnam wanted to determine what obstacles prevented widespread use of regular check-ups. They conducted an interview survey in Hanoi and Hung Yen, Vietnam, by traveling to "secondary schools, hospitals, companies, government agencies and randomly selected households in Hanoi" and interviewing people in person for about 10–15 minutes. This dataset contains the valid responses from that survey.[1]

## Relevant Variables

Each row of the data represents one respondent to the survey. There were 2,068 valid responses.

The response variable is `HadExam`, which is 1 if the respondent had a check-up in the past 12 months, and 0 otherwise.

### Demographic and Personal Variables

`id` A unique ID for each survey respondent

`place` The location where this respondent was surveyed. Many locations only have a few respondents, so you should not include this variable in your models unless specifically necessary

---

[1]Original source: Q-H Vuong (2017). Survey data on Vietnamese propensity to attend periodic general health examinations. *Scientific Data* 4 170142.

Age  Age of the survey respondent (years)

Sex  Sex of the survey respondent

Jobstt  Job status: stable, unstable, student, retired, homemaker, or other

height  Height (cm)

weight  Weight (kg)

BMI  Body mass index

HealthIns  Whether the respondent has health insurance

## Value and Quality of Medical Service

Next, there are variables measuring the respondent's beliefs about the value of check-ups and the quality of the medical service:

Wsttime  Does the respondent believe check-ups are a waste of time?

Wstmon  Does the respondent believe check-ups are a waste of money?

Lessbelqual  Does the respondent have little faith in the quality of medical service?

NotImp  Does the respondent believe check-ups are not urgent or important?

Tangibles  Perceived quality of medical equipment and personnel at check-ups. Scale of 1 to 5, 1=lowest quality

Empathy  Perceived thoughtfulness and sense of responsibility of medical staff, scale of 1 to 5

SuitFreq  How often respondent believes check-ups should be done (6m = every 6 months, 12m = every 12 months, 18m = every 18 months, g18m = less often than every 18 months)

## Quality of Information

Next are variables about how the respondent rates the information they receive in check-ups:

SuffInfo  Respondent's rating of the sufficiency of information they received in check-ups, 1 to 5 scale

AttractInfo  Respondent's rating of the attractiveness of information they received in check-ups, 1 to 5 scale

`ImpressInfo` Respondent's rating of the impressiveness of information they received in check-ups, 1 to 5 scale

`PopularInfo` Respondent's rating of the popularity of information they received in check-ups, 1 to 5 scale

## Your Goals

Suppose that the Vietnamese Ministry of Health wants to increase the number of people who get annual health exams. They could send reminders or offer incentives to anyone who gets an exam. But the Ministry does not know *why* people do not sign up for exams, and understanding why could help them make their public relations campaigns more effective. (For example, if people tend to think check-ups are a waste of money, their advertising should emphasize how cheap and effective they are; if people don't believe medical service is high-quality, their advertising should emphasize the expertise and training of doctors.)

The Assistant Minister of Health hence has three questions for you:

- Overall, how do people rate the value and quality of medical service, and the quality of information they receive in check-ups? Give us some summaries of these variables so we know how to improve.

- What factors appear to make a person less likely to get a check-up every twelve months? Find the most important factors that could help us design our advertising, and give us some measure of how important they are.

- The Ministry of Health is considering a marketing campaign focusing on the quality of information patients receive in check-ups. Does the evidence suggest this is an important predictor of whether patients get check-ups, and does this differ between people with and without health insurance?

Yes, the Assistant Minister's questions are somewhat vague, but this tends to happen with government officials. The Rmd template provides scaffolding by suggesting steps you can take to answer these questions. Your analysis should answer the questions posed in the Rmd template, and use the results to answer the Assistant Minister of Health's questions in the Introduction and Conclusion.

Your task is to build **interpretable, theoretically valid models** to address these research questions. Note that our focus is inference, not prediction. Specifically, you must address the questions in the R Markdown template along with all other issues necessary for describing and justifying a statistically valid analysis.