Sampling Assignment Two Final Version

The goal of this assignment is to practice developing and evaluating sampling materials.

**Part A - Survey Design:**

Select one of the scenarios below and design a survey to meet the need(s) outlined in the prompt.

1. In two to three sentences, describe the purpose of your survey
2. Describe your target population, sampling frame, sampling units, and overall sampling strategy.
3. Write a 5-10 question survey to address your chosen scenario below.

You work in the Human Resources Department at a large tech company. Over the past few months, the company has been experiencing a high turnover rate across many of its departments, specifically within the entry- and lower-level positions. The company wishes to understand why this turnover is happening, and what changes need to occur to improve employee satisfaction.

**Response:**

Since my background is in biotech, I will assume the company is a US-based biotech company.

**Purpose**: This survey is designed to assess the employees’ satisfaction with their work at the company and identify ways in which the company can better support employees. The survey will ask demographic questions and questions about common sources of employee satisfaction. All questions are optional and all data will remain confidential. Raw data will only be viewed by HR.

**Target population**: Entire company.

**Sampling Frame**: A specific percentage of the company. Will try to collect as large a sample size as the HR team can manage. (If they HR team can survey the entire company, they will do so). Will use a stratified sampling approach. Divide the employees into strata based on type of position (entry level or introductory, mid-level, senior-level, C-suite / executive leadership). Within each stratum, only sample a set percentage of employees, such as 20%. For the entry level or introductory stratum, collect extra samples (perhaps send surveys to 30-50% of that stratum), since they are a particular focus of this study group. This population can be weighted differently when discussing data about the company as a whole.

**Sampling units**: First stage sampling units are strata within the company based on job level. The second stage sampling units are randomly selected individuals within each stratum.

**Strategy**: Survey will be online with a link distributed by email, and participants will be allowed to take the survey in private (at home) or at work. At work, they will be given private space to take the survey if desired, as taking the survey in a public or shared office can be uncomfortable. All employees at this company know how to use a computer and the internet (it is a job requirement) and have access to a computer, so with email and internet are appropriate to administer the survey. However, a paper copy of the survey can be administered upon request.

When analyzing data, responses will be analyzed collectively. They will also be analyzed for each demographic group listed in part I and for each job level stratum within the company. In particular, HR will examine survey data for any strong discrepancies between entry-level / introductory employees and those in more senior roles. Numerical data will be collected, but HR will also pay attention to any common themes observed in the “open response” sections of the survey.

If imputation is needed, will input any missing answers based on responses from employees with similar demographic details.

**Survey:**

1. **Demographic Information**

All questions are optional. Responses are confidential and raw data will only be seen by HR. If HR already has this demographic information on file for each employee and it can be linked to their survey responses, then the demographic section of this survey can be omitted.

1. Age:
2. Gender:
3. Highest educational degree completed:
   1. High school, Associate, or Equivalent
   2. Some college education, but no degree
   3. Bachelors
   4. Masters / Ph.D.
   5. Other / NA
4. Job title:
5. Years of employment at company:
6. Do you identify as LGBTQIA?
7. Yes
8. No
9. Prefer not to respond
10. Do you identify as a person with a disability?
11. Yes
12. No
13. Prefer not to respond
14. Do you identify as neurodivergent?
    1. Yes
    2. No
    3. Prefer not to respond
15. With which racial groups do you identify (select all that apply):
16. White
17. Black or African American
18. American Indian or Alaska Native
19. Asian
20. Native Hawaiian or Other Pacific Islander
21. None of the Above
22. Prefer not to respond.
23. Do you identify as Hispanic, Latino, or Latinx?
    1. Yes
    2. No
    3. Prefer not to respond
24. If you wish, please elaborate on your responses above:

The racial and ethnic groups mentioned in these questions are those used in the US Census: <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html>

Note: the questions in the non-demographic sections of the survey were inspired by my own experience and a recently published American Job Quality Study: <https://www.gallup.com/analytics/691241/american-job-quality-study.aspx>

For the questions in the sections below, please rate the statements as **Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree.**

1. **Financial Well Being**
2. I am fairly compensated for my work.
3. The company offers competitive benefits
4. My compensation affords a good quality of life in my geographical area.
5. Provided the economy does not change and my performance is adequate, I can work for this company as long as I like.
6. I can afford to take time off when I need it.
7. (Free response): If you wish, please elaborate on your responses above:
8. **Work Structure and Autonomy**
9. I work a reasonable number of hours.
10. I feel micromanaged.
11. My work schedule is predictable.
12. I have enough time to get everything done at work.
13. I have flexibility in where I complete work (ex. in-office, work from home, travel, etc.).
14. (Free response): If you wish, please elaborate on your responses above:
15. **Workplace Culture and Safety**
16. I get along with my coworkers
17. I get along with my manager
18. I feel physically safe at work.
19. I feel psychologically safe at work
20. I have experienced health issues because of my work.
21. (Free response): If you wish, please elaborate on your responses above:
22. **Growth and Development Opportunities**
23. My assignments at work are appropriate to my level of expertise.
24. I receive feedback that helps me improve
25. The company supports the professional development of employees.
26. The company establishes clear pathways to promotion.
27. After I started, I received proper on-the-job training for my role
28. (Free response): If you wish, please elaborate on your responses above:
29. **Agency and Voice**
30. My ideas are heard and respected.
31. Management cares about the well-being of employees.
32. I am included in important workplace decisions.
33. I have the same opportunities as my peers.
34. My work is meaningful.
35. (Free response): If you wish, please elaborate on your responses above:

**Part B - Survey Evaluation:**

For the **Canadian General Social Survey on Giving, Volunteering, and Participating, 2018 (cycle 33)**, conducted by Statistics Canada find any and all available documentation for the data gathered and identify and describe the survey features indicated below.

1. **Sample type:**

The sampling units were identified through a combination of stratification and random sampling.

Surveyors identified 27 strata based on geographic locations. (Several strata are also census metropolitan areas). Within each stratum, they identified groups of phone numbers that formed a “household” by comparing phone numbers to address records found in census data. Households were randomly selected from each stratum. Within each household, one respondent was chosen at random.

1. **Sample size**:

In one document about the survey, the expected sample size was reported to be around 20,000 respondents.(**“**A field sample of approximatively 50,000 units was used. Among them, about 40,000 invitation letters to the electronic questionnaire were sent to selected households across Canada. A completion of 24,000 questionnaires was expected.”).

In a different document about the survey, the reported target sample size was 20,000 respondents and the actual sample size was lower, at 16,149 respondents, excluding “rejected” respondents.

1. **Target population**:

The target population for the GSS Giving, volunteering and participating includes all persons 15 years of age and older living in the ten provinces of Canada. It excludes full-time (residing for more than six months) residents of institutions.

Of note, the method of identifying sampling units involved using phone numbers, so people without a consistent phone number may have been inadvertently excluded.

1. **Sampling frame**:

The sampling frame was identified through multiple stages. First, geographic strata were designed from census data on provinces and metropolitan areas. Within each stratum, the primary sampling units were clusters of phone numbers linked (through census data) to the same address and labelled as a “household.” Households were chosen at random from each stratum. The final sampling units were the randomly selected individuals within each household.

The final sampling frame, therefore, is the individuals selected for the study after the above selection processes (one person per identified household within the strata).

1. **Survey mode(s)**:

The 2018 version of the survey featured an internet option for the first time. Invitations to the electronic survey were sent in the mail.  Data were collected directly from survey respondents either through the electronic questionnaire or CATI (computer assisted telephone interviewing). No proxy reporting was allowed. The respondents could choose between French and English.

1. **Timeline**:

* A pilot survey was conducted February 19, 2018 to March 30, 2018
* The main survey was conducted September 04, 2018 to December 28, 2018.

1. **Response rate**: The overall response rate was 41.9%.
2. **Weights**:

The person weight calculation (number of people represented by each respondent) was accomplished using a multi-step process based on multiple criteria:

* 1. Initially, a weight per record (“household” estimated from telephone numbers) was established by taking the total number of records in each stratum in the sampling frame divided by the total number of records sampled in each stratum.
  2. This initial weight was adjusted to account for the fact that some households had more phone numbers associated them and were thus more likely to be selected for the study. The resulting value was defined as the household weight.
  3. The weights for responding households were adjusted to account for non-responding telephone numbers. Auxiliary information, if available, was used to model the non-responding households’ propensity to respond. This stage of weighting resulting in a number called “Factor 1”.
  4. The weight of the person surveyed within a household was adjusted so that they could represent all eligible respondents (those aged 15 years or older) within the household. This weight was = initial household weight \* Factor 1 \* Number of eligible household members.
  5. The weight was further adjusted to account for the fact that some respondents were “rejected” as non-volunteers.
  6. Person weights were adjusted several times using a raking ratio procedure. This step is meant to ensure alignment between the study sample and external reference totals such as income distribution, population per geographic stratum, and age-sex groups per province.

1. **Data processing**:

Errors were detected using the social survey processing environment (SSPE) processing steps and guidelines. Such steps included internal automatic and manual checks to verify the accuracy and consistency of data. Some specific examples described include checks on family relationships, comparisons of stated age to birthdate, and flow edits to ensure respondents followed the correct path through the survey. The CATI system also had error detection built in. All survey records were subjected to computer edits throughout.

In some cases, to better estimate sampling variability, bootstrap weights were used.

1. **Cleaning, imputation, etc.** :

Imputation was carried out by imputing missing answers for a recipient respondent with the answers from a donor respondent deemed similar. The “similarity” of the respondents was determined through a nine-step scoring mechanism involving multiple variables. If an appropriate donor could not be identified, a mean was used.

1. **Sources of error**:

Other than sampling error, the major sources of error mentioned by the study authors are imperfect coverage and non-response.

1. **Limitations, known biases, etc.**:

One major limitation is that the study would only capture people with a working phone number that had been recorded in the most recent census.

Households without telephones were also excluded from the survey population. Survey estimates were adjusted (weighted) to represent all persons in the target population, including those not covered by the survey frame.

1. **Link to documentation and any additional sources used**:
   1. <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=796234#a1>
   2. <https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&a=1&&lang=en&Item_Id=1183690>
   3. <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=425224>
   4. <https://volunteer.ca/data-on-giving-volunteering-and-participating-in-canada/>
   5. <https://www150.statcan.gc.ca/n1/en/catalogue/45250011>
   6. <https://www150.statcan.gc.ca/n1/pub/45-25-0001/cat5/c33_2018.zip>