ADD IN MEANINGFUL TITLE HERE

STA304 - Fall 2025 - Assignment 1

A1-GROUP 40: ISLA CHEONG, VANESSA GABRIELA KWOK, MIA SUNG, CAROL XU

1 Introduction

Mentorship is a resource that aids students in finishing high school and getting ready for higher education, especially in science, technology, engineering, and mathematics (STEM) areas. STEMBuddies is a group that connects high school students with mentors, organizes workshops, offers programs centered on post-secondary education, and supplies free resources to decrease obstacles for underrepresented communities. To develop an optimized handbook that supports students both academically and personally, it is essential to collect feedback from the high school students so that their needs and preferences can be assessed.

This report outlines the design and analysis of a survey created to assess the needs of high school students and their preferences. By using best practices in survey design [1], the survey provides actionable insights for STEMBuddies to improve mentorship accessibility, inclusivity, and their overall impact.

2 Survey Showcasing

2.1 Survey Description

Our survey, hosted on https://forms.gle/sWqTQ1HCXPrjkk6BA, is designed to gather insights from high school students regarding their academic and personal needs, as well as their preferences for mentorship programs. The survey consists of the following sections:

1. **Introduction**: A brief description of STEMBuddies and purpose of the survey.

2. **Core Questions**: Demographic questions, needs-assessment questions, and program feedback prompts.
3. **Closing Message**: Thanks participants and provides contact information.

2.2 Testing My Survey

Each question was designed to minimize ambiguity and cognitive load. For example, we avoided double-barreled items and ensured that the survey was focused and concise [1]. We piloted the survey with three peers and incorporated feedback to refine clarity and ease of responsiveness.

2.3 Showcasing Relevant Question(s)

A key survey question is showcased below:

Which of the following resources would be most valuable to you in the STEMBuddies program? (Select up to three options)

- Academic Tutoring
- College Application Guidance
- Career Counseling
- Mental Health Support
- Time Management Skills
- Financial Literacy
- Other: _____

We choose this question because it provides insight into what our target population may be seeking in the handbook and helps us better understand their needs.

Benefits: Multiple-choice with "Other" allows structured yet flexible responses. Drawbacks: May still miss nuanced needs not anticipated in listed options.

3 Procedure

3.1 Proposed Sampling Procedure

To collect survey feedback, we propose a stratified random sampling procedure targeting high school students across Canada, with the individual student serving as the sampling

unit. The survey will be distributed to students via partnerships with high schools, provincial boards, mailing lists, STEM clubs, outreach initiatives, and educational organizations that cater to diverse student populations in order to acquire the sampling frame. This method guarantees representation across various strata such as geographic area, school category (public or private), and socioeconomic level. The survey can be conducted in-class with teacher oversight to boost response rates or online using unique survey links shared through schools and partner organizations.The sampling frame is limited but feasible. Errors may arise from coverage bias (because some schools or students may not be included in the frame), non-response bias (students that are absent or less engaged choose not to participate) and limited internet access in rural or low-income areas.

3.2 Simulation

We simulate survey responses for 100 students to mirror the showcased resource-preference question. Each simulated respondent selects up to three resources from the options provided in the survey. This process ensures that the simulated dataset reflects the structure and response style of the actual survey, while still being reproducible through the use of a fixed random seed. By generating these responses, we create a realistic dataset that can be analyzed to demonstrate the kinds of insights STEMBuddies could obtain if the survey were administered in practice.

4 Data

4.1 Data Cleaning and Processing

The dataset consists of simulated survey responses from 100 high school students. Each student was asked to select up to three resources they would find most valuable from a list that included options such as Academic Tutoring, College Application Guidance, and Financial Literacy.

To prepare the data for analysis, we reshaped the responses into a long format, where each selected resource for a student appeared as a separate entry. This helps us to summarize selections across all students. Next, we converted the data into a binary format, creating separate indicator variables for each resource. In this structure, a value of 1 indicates that a student selected the resource, and 0 indicates they did not.

The main variables are the binary indicators for each resource, showing whether a student selected it. This cleaned format allows for consistent, reproducible analysis of overall resource preferences.

4.2 Variables of Interest

The main analysis focuses on the survey question: "Which of the following resources would be most valuable to you in the STEMBuddies program? (Select up to three options)." Each resource is represented as a binary-coded variable (0 = not selected, 1 = selected). The key variables analyzed are: Academic Tutoring, College Application Guidance, Career Counseling, Mental Health Support, Time Management Skills, Financial Literacy, and Other. These variables form the core of the main analysis, allowing us to calculate the frequency and proportion of students selecting each resource, visualize student preferences, and compute confidence intervals to quantify the uncertainty around these proportions.

4.3 Figures

Below is a bar plot of frequencies:

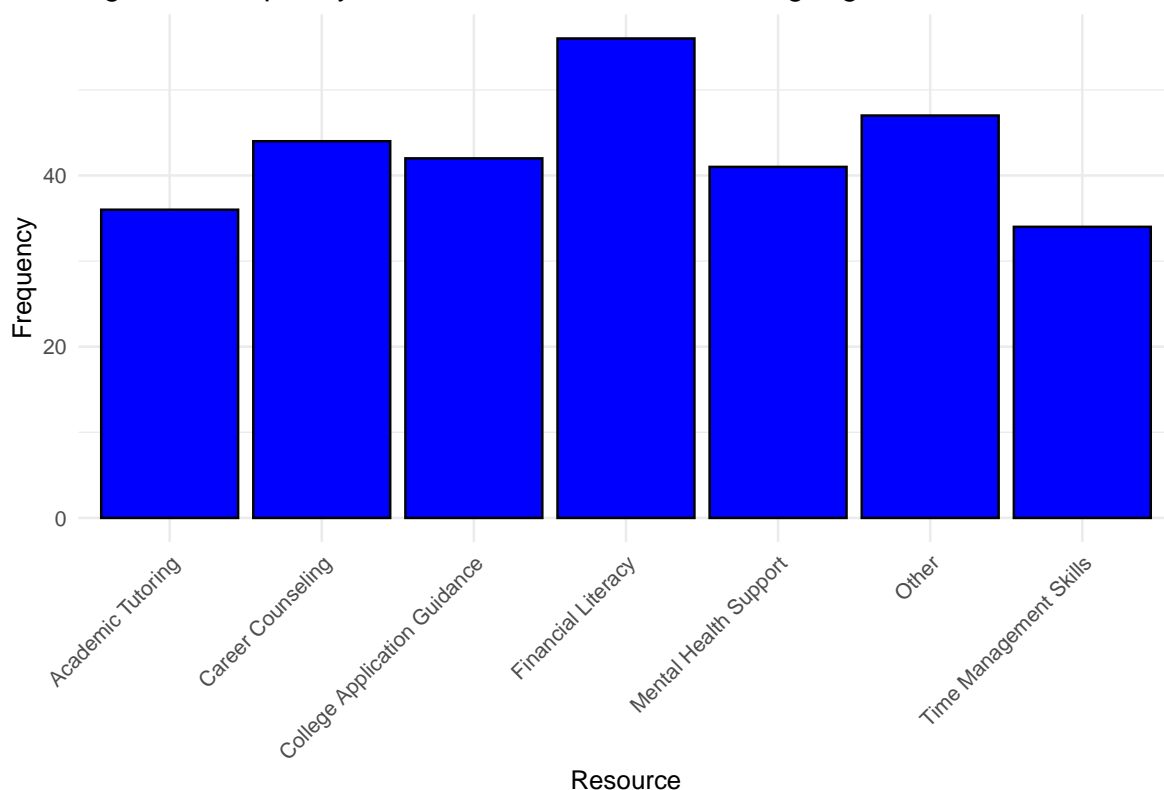**Figure 1: Frequency of Resource Preferences Among High School Students**

Figure 1 shows the number of students who selected each resource from the survey. The x-axis lists all possible resources, while the y-axis represents the number of students selecting each option. This figure clearly highlights which resources are most frequently requested, providing STEMBuddies with a visual summary of student preferences. For example, taller bars indicate higher demand for a particular resource, helping to prioritize support areas in the handbook.

4.4 Data Summary for Confidence Interval Analysis

5 Methods

To estimate the proportion of students who prefer a specific resource, we will use the following formula for a 95% confidence interval for a population proportion:

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}$ is the sample proportion, z^* is the critical value from the standard normal distribution (z-score) corresponding to the desired confidence level (1.96 for 95% confidence), and n is the sample size [2]. The value under the square root represents the standard error, reflecting the variability expected from repeated student sampling, and multiplying it by z^* gives the margin of error. This method assumes a large enough sample size to support the normal approximation and that student responses are independent of each other. With these conditions, a 95% confidence interval indicates that if the survey were repeated many times, about 95% of the constructed intervals would capture the true population proportion.

6 Results

Table 1: Table 1: The proportion and 95% confidence interval of students preferring "Academic Tutoring".

Resource	Proportion of Students	95% Confidence Interval
Academic Tutoring	0.30	(0.21, 0.39)

Interpretation: For example, 30% of students selected academic tutoring as a valuable resource, with a 95% CI of approximately (21%, 39%).

7 Generative AI Statement

Generative AI (ChatGPT) was used to help structure this report, generate simulation code, and ensure clarity of statistical explanations. All simulated data generation, results, and interpretation were conducted independently by the authors.

8 Bibliography

1. *Creating Effective Surveys: Best Practices in Survey Design.* Evidence Use in Education. (2021). Regional Educational Laboratory West. https://eric.ed.gov/?id=ED619797.

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: April 4, 1991)

9 Appendix

9.1 Survey questions

Below is a well formatted copy of all of my survey questions (along with the opening and closing messages/statements).

9.2 Simulated Survey Data

Below is a `glimpse()` or `head()` of my simulated data (this should be the data that looks like the inputs from your survey). This is the only place in the report where you can show your raw R code and output.

```
head(sim_data)
```

```
# A tibble: 6 x 4
# Rowwise:
     id resource1                     resource2               resource3
  <int> <chr>                         <chr>                   <chr>
1     1 Mental Health Support         Academic Tutoring       Financial Literacy
2     2 College Application Guidance  Other                   Mental Health Support
3     3 Financial Literacy            Other                   Time Management Skil~
4     4 Time Management Skills        Mental Health Support   Academic Tutoring
5     5 Mental Health Support         Academic Tutoring       Time Management Skil~
6     6 Other                         Financial Literacy      College Application ~
```

9.3 Supplementary Materials

Any other supplementary items you wish to include (e.g., more plots, derivations, etc.) can be in this sub section.