

STA130 Project Proposal

Research Question #1

Question: Did the new registration system implemented on March 4, 2024 impacted the engagement of their members?

Which variable(s) will you use to explore this question?

x-variable:

- member_start_year (integer): Year in which the member started with Wellspring
- member_start_month (integer): Month in which the member started with Wellspring
- registration_group (categorical, derived variable): A derived binary variable indicating whether a member registered before (old system) or after (new system) March 4, 2024

y-variable: number_of_present_service_deliveries (integer): Represents the total number of services in which the member was marked as "Present" (all-time total, not just 2023-24, so will not match with list of Service Deliveries)

Which method of analysis do you propose using?

Two-sample hypothesis: We will compare the mean number of present service deliveries between members who registered before March 4, 2024 (old system) and those who registered after (new system). This test will help determine whether the new registration system significantly increased engagement. H_0 (Null Hypothesis): There is no significant difference in the mean number of present service deliveries between members who registered before and after March 4, 2024. H_1 (Alternative Hypothesis): Members who registered after March 4, 2024, attended significantly more (or fewer) events than those who registered before.

Visual Representation:

we can create a Boxplot with: X-axis be the registration_group (Old System vs. New System). Y-axis be the number_of_present_service_deliveries (Total attendance of each member). This boxplot will compare the distribution of attendance between members who registered before and after March 4, 2024. Each box represents the interquartile range (IQR), with the median attendance clearly labeled inside each box.

Relevance to Wellspring:

This analysis allows Wellspring to quantitatively assess whether the new registration system led to a meaningful increase in member engagement. If the new system significantly increased attendance, Wellspring can confirm that the update was effective and use this evidence to inform future registration improvements. If no significant difference is found, or if engagement decreased, Wellspring may need to explore additional factors affecting attendance, such as program offerings, scheduling, or outreach strategies, to further enhance member participation.

Research Question #2

Question: What background information about members (i.e. age, gender) significantly predicts their participation in events?

Which variable(s) will you use to explore this question?

x-variable:

- Age (integer): Age of member
- gender (character): Gender of member

y-variable: number_of_present_service_deliveries (integer): It reflects the total number of services in which the member was marked as "Present" (all-time total, not just 2023-24, so will not match with list of Service Deliveries)

Which method of analysis do you propose using?

Linear regression- A model was developed using linear regression to explore the effect of each variable on activity participation. In our linear regression model, the response variable is a numeric variable, such as number_of_present_service_deliveries, which represents the number of times a member has attended an event. Predictors would include both numeric (e.g., age) and categorical (e.g., gender) variables to analyze the impact of each factor on event participation.

Visual Representation:

Use a scatter plot and regression lines to show model predictions. For the scatter plot, x-axis is the age and y-axis is the number_of_present_service_deliveries. Each point represents a single member's total attendance plotted against their. Different colors to represent Gender (e.g., blue for male, red for female, etc.). Each dot on the scatterplot represents a single member, plotted based on their age and participation level. A linear regression line will be added to indicate the general trend (e.g., does participation increase or decrease with age?). If gender is an important factor, we can include separate regression lines for males and females. If participation appears clustered at certain ages (e.g., younger members participating more), this pattern will be visible in the scatter plot.

Relevance to Wellspring:

Using this linear regression analysis, Wellspring was able to gain a clear understanding of which members' background characteristics, such as age or gender, were associated with high engagement, allowing them to design personalized and targeted promotional services. This analysis also helps them identify groups that are less engaged, so they can develop targeted interventions to improve overall service participation. For example, if younger members participate less, targeted outreach efforts may be needed to increase their involvement. If the analysis shows that male members participate significantly less than females, Wellspring can adjust its outreach efforts to better engage male members.

Research Question #3

Question: Is Wellspring's membership age distribution balanced?

Which variable(s) will you use to explore this question?

age (integer): Represents the member's age. Using this variable, we will create a binary variable that classifies members as "young members" (age < 40) and "non-young members" to calculate the proportion of young members.

age group (categorical, derived variable): A derived binary variable indicating whether a member is younger than 40 or older than 40 years

Which method of analysis do you propose using?

One-sample proportion test – We will use a one-sample proportion test to determine whether the proportion of young members is equal to 50%. This method will help assess whether the proportion of young members (defined as those under 40 years old) is equal to the proportion of elder members. H_0 (Null Hypothesis): The proportion of young members is equal to 50% (i.e., the age distribution is balanced). H_1 (Alternative Hypothesis): The proportion of young members is not equal to 50% (i.e., there is an imbalance in the age distribution).

Visual Representation: we can create a bar chart that has x-axis be the age group (Young Members vs. Non-Young Members) and the Y-axis be the percentage of total members in each age group. Each bar represents the percentage of young vs. non-young members, and a dashed

horizontal reference line at 50% can indicate the expected balanced distribution. Bars will be color-coded (e.g., blue for young members, red for non-young members) and labeled with the exact percentages for clarity.

Relevance to Wellspring: If the proportion of young members is significantly lower than 50%, Wellspring may need to enhance its outreach efforts to attract younger individuals, such as adjusting marketing strategies or introducing programs tailored to younger demographics. If the proportion of young members is significantly higher than 50%, it could indicate that older individuals face barriers to participation, prompting Wellspring to investigate potential obstacles such as accessibility, program relevance, or communication strategies.