Data Analysis of Wellspring Member Engagement

Thea Guan, Cindy Yang, Rachel Pham, TUT0104, and 0104-D

Introduction

▶ Wellspring is a Canada-wide network of charities offering free cancer support programs to individuals at any stage of their cancer journey.

▶ This project aims to evaluate the impact of their simplified registration system, implemented on March 4, 2024, on member engagement, such as event attendance rates or time to first event.

▶ The analysis will be tailored to an audience of Wellspring employees, thus presenting clear, practical, and visually accessible results is critical in order to support their mission of improving cancer care services.

Question 1: Introduction

▶ Research question - Is Wellspring's membership age distribution balanced?

▶ For this question, the variables which will be utilized within this method would be the '**age_years**' variable within the data set.

▶ From the **age_years** variables, a new binary variable will be created, which classifies members as "young members" (age < 40) and "elder members" to calculate the proportion of young members among all membership.

▶ A **one-sample proportion test** would be used to determine whether the proportion of young members (defined as those under 40 years old) is equal to the proportion of elder members.
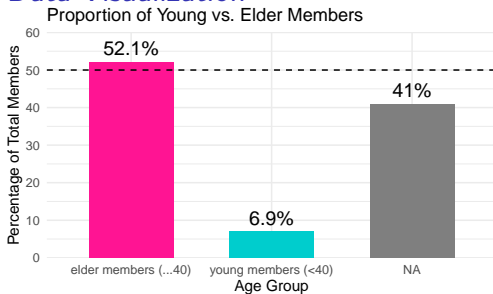
Data summary

Figure 1: Proportion of Young vs. Elder Members

▶ See Figure 1, elder Members (40) make up 52.1% of members, while only 6.9% are under 40. 41% of age data is missing, which may affect the reliability of these results.

▶ The age imbalance suggests Wellspring's services are more used by older individuals.

Question 1: Statistical Analysis

▶ H0 (Null Hypothesis): The proportion of young members is equal to 50% (i.e., the age distribution is balanced).

▶ Ha (Alternative Hypothesis): The proportion of young members is not equal to 50% (i.e., there is an imbalance in the age distribution).

▶ The test statistic and the p-value can be estimated from a graph based on multiple repetitions of simulated samples assuming the null hypothesis is true.

▶ The test included **10,000 simulations** to determine the distribution of the proportion of young members under the assumption that the true proportion is 0.5.

Question 1: Results

▶ The **p-value** was found to be **0**, which is less than 0.001, we conclude that we have **very strong evidence against the null hypothesis** that the proportion of young and elder members is 50% each. Wellspring's membership is heavily skewed toward elder members, with few young members.

▶ Further research is suggested to investigate possible causes behind the underrepresentation of younger individuals at Wellspring, and whether targeted outreach or program redesign could help attract a more age-diverse membership.

▶ However, if the missing data is random, the analysis likely reflects the true membership. But if younger people are more likely to omit their age, the imbalance may be overestimated.

▶ Research question - **Is the proportion of the first listed property a travel ad, equal to 50%?**

▶

Question 2: Data Visualization

The above bar graph shows the frequency of the first listing being a Travel Ad or not. It can be observed that out of the sample data, 600 of the first listings were not Travel Ads, while the rest 400 were Travel Ads.

Question 2: Statistical Analysis

▶ Null Hypothesis (H_0): Among all the searches on the Expedia website that span the period from 2021-06-01 to 2021-07-31, the proportion of the first listings which are advertised in travel ads is equal to 50%.

▶

Question 2: Results

▶ The actual proportion of first listings advertised in a travel ad was found to be 0.419.

▶

Question 3: Introduction

▶ The variables member_start_year, member_start_month were used to determine the registration date for each member.

▶ Taking the minimum of the variables delivery_year, delivery_month, delivery_day, as well as filtering by Present in attendance_status were used to determine the date of the member's first attended event.

▶ The new variable system_change was created to differentiate between if a member registered before or after the system change date, and grouped accordingly.
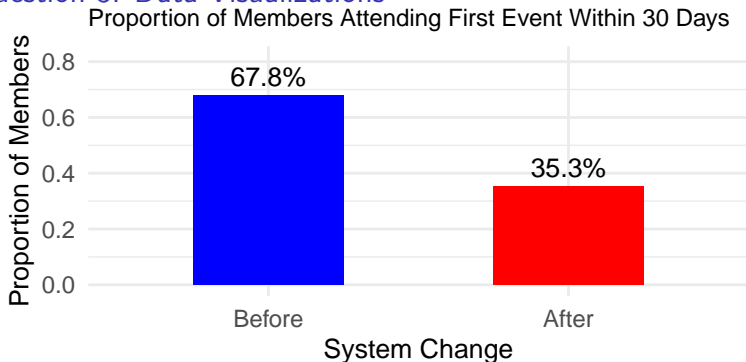
Question 3: Data Visualizations

Figure 2: Proportion of Members Attending First Event Within 30 Days

See Figure 3, the bar graph shows the proportion of members attending their first event within 30 days. Before the registration system change, 67.8% of members attended their first event within 30 days, while only 35.5% did after the change.

Question 3: Statistical Analysis

▶ Null hypothesis (H0): The proportion of members attending their first event within 30 days of registration is the same before and after the registration system change.

▶ Alternative hypothesis (Ha): The proportion of members attending their first event within 30 days changed after the registration system change.

▶ To assess the significance of the observed difference in proportions between the two groups, we performed a permutation test with 10,000 random permutations. This included randomly shuffling the system_change labels to simulate the null hypothesis, and recalculating the difference in proportions for each permutation to build the null distribution of differences.

Question 3: Results

▶ The actual difference between the proportions was
-0.32564096245674. This indicates a 32.6 percentage point
decrease in participation within the first 30 days after the
system change.

The p-value was found to be 0.

Since the p-value is less than 0.05, we conclude that we have
very strong evidence against the null hypothesis.

We conclude that the system change had a statistically
significant negative effect on member participation.

Conclusion

▶ The significant underrepresentation of young members suggests a potential gap in Wellspring's outreach or service appeal. Adapting programs to better engage younger individuals could help build a more balanced and sustainable membership base.

▶

▶ The proportion of members attending their first event within 30 days of registration decreased by 32.6 percentage points after the change, with a p-value of 0, indicating there were lower early participation rates among members

Limitations

▶ For Q1, a key limitation is the high proportion of missing age data, which may bias the results if not missing at random. Additional variables such as gender, location, or membership activity would help provide a more complete picture of engagement across age groups.

▶

▶ Since there was no member_start_day indicating the day of the month which Wellspring patients registered, we had to assume that everyone registered on the first day of the month for consistency. The accuracy of findigns for Q3 could be improved upon if we had access to a member_start_day variable.

References and Acknowledgements