Statistical Analysis of the Omaha Girls Rock Program with a Focus on the Age Group and Race/Ethnicity of Participants

Joshua Ellis, Alex Daly, Satwik Chandra

ABSTRACT

Omaha Girls Rock, a local non-profit that puts on program for young local girls in the area, conducts a survey before and after the program to see if the girls have developed their social abilities through the course of their participation. The purpose of this study is to investigate the effects that the age group and race/ethnicity of a participant has on the way they will answer survey questions about themselves as well as the effectiveness of the program on those participants. Specifically, are the answer ratings in the survey different between the older group and the younger group? Additionally, does the overall effectiveness of the OGR program differ between participants of various races/ethnicities? It was found that there is no difference in answer ratings between age groups, meaning that neither group scores significantly higher/lower than the other. However, there are subtle differences in delta between the pre- and post-test for different age group. The tests on race/ethnicity suggested no difference in delta scores on a per-question basis, however overall delta scores (not question specific) do suggest that there is a difference in mean delta scores between groups of race/ethnicity. We did find that there were some differences in how the girls answered the questions on the post survey based on their racial identity than should also be examined more.

1 INTRODUCTION

Omaha Girls Rock is a local non-profit organization that seeks to use music to grow the young girls in the community. By teaching them music and giving them outlets to perform, they seek to create identities of self. As part of this program, the organization conducts surveys before and after the girls participate in the program to see how their skills have improved. These questions ask the girls to rate their likelihood of doing various activities, such as wearing the kind of clothes they like even if they are different from others to making friends their age, on a scale of 1-7 (1 being the least likely to 7 being the most). Additionally, there were three statements where they were to either agree or disagree to a varying degree. We want to analyze these results to make recommendations on how Omaha Girls Rock can improve their programs.

First, we would like to see if there was any significant difference in how the girls answered the questions before and after their program based on if they were in the older group (10-12) or the younger group (13-16). We believe that the girls should show growth for each response, regardless of what age group they are a part of. This information would tell them which age groups they need to focus on in building the skills that are associated with that given question. For example, if girls in the younger age group showed that significantly they did not improve or was worse at being able to express their feelings to another kid, the organization could work with the girls 10-12 on how to express themselves more.

Second, we would like to examine the results to see if there were any significant differences between how the girls answered the survey based on their race. Again, we expect there to be no significant differences in how the girls answer each question regarding their race. If we do find a significant difference though, Omaha Girls Rock would then know how to better cater their activities to the community. No organization wants to show that a specific group is being treated unequally compared to others. This will let them know of any blind spots they might have and work to correct them.

2 DATA DESCRIPTION

2.1 THE DATA

The data in the dataset pertains to a survey of girls from Omaha, ages 8-14, before and after participation in the OGR Academy offered by the Omaha Girls Rock Organization. Specifically, OGR conducts pre and post participation surveys using a range of questions to measure the impact of their program on participants. For each participant, additional information is collected, such as age, ethnicity, and ZIP code. For this analysis, we focus on survey questions 15 – 25 and ending questions 1,2, and 3.

2.2 VARIABLES

The following table describes the data from the survey.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Data Type |
| client | Unique identifying code of a participant | Nominal |
| age\_group | The age group of the participant {Older Group or Younger Group} | Nominal |
| year | The year of the test {2018, 2019} | Nominal |
| test\_type | The stage of the test {Pre-Test or Post-Test} | Nominal |
| age | The age of the participants | Ratio |
| years\_at\_camp | The number of years a participant has been in the Camp | Ratio |
| race/ethnicity | The race/ethnicity of the participant {Caucasian, Black, Multi-Racial, Other} | Nominal |
| zip\_code | The zip code of where the participant lives | Nominal |

2.3 DATA PROCESSING

After familiarizing ourselves with the raw survey data, we noticed some data quality errors, such as missing data, different representations of the same field (i.e., 1 vs 1st and 2 vs 2nd in the same field), class imbalance (race/ethnicity), and non-numeric answers. We underwent a data processing step before deciding to do any analysis on the data, so that we are only working with easy-to-use and clean data. Below describes each data quality issue and how it was corrected.

***Many Tables***

Instead of having 8 datasets to do analysis on, all the data was combined to a single table, where fields such as age\_group, year, and test\_type were added as new fields. These new fields allowed us to identify what each data observation represents (i.e., Older Group, 2019, Pre-Test). We can filter on these fields to subset the data in any way we want, while only needing to read from and work with a single table.

***Missing Data***

There were some instances where a question did not have a score for a given observation. Since the data we are working with is ordinal, we replaced and missing data with the mode of that question. Since the amount of missing data only makes up < 1% of the total dataset, we do not risk altering the true results of the survey.

***Participant ID***

In the raw survey data, there is a client ID field, however in 2018 the ID was the name of the participant and in 2019 the ID was a number. To standardize this, each participant in 2018 was assigned a unique number based on their name. With a standardized Participant ID, it is now easier to compare pre- and post- test results.

***Data Representation***

In the *Years at Camp* field, some survey results include the superscript of the year, (1st, 2nd, 3rd), while others did not (1, 2, 3). Since it is appropriate to cast this field as an integer, all superscripts were removed and set to an integer type.

***Class Imbalance***

At first look into the race/ethnicity field, there were 17 unique groups of race/ethnicity. There was no standardized way to collect this data in the survey, therefore there were differences in the way a race is referenced and the order of races if an individual is multi-racial. During data processing, a cleaning procedure is done to reclassify race/ethnicity by standardizing all these different factors. This left us with 12 standardized unique groups.

However, classes were still highly unbalanced: 250 Caucasian, 53 Black, and the remaining 10 groups all have less than 10 data points. To overcome this, race/ethnicity was further grouped into 4 primary groups: Caucasian, Black, Multi-Racial, and Other.

***Non-numeric answers***

The last three questions in the survey do not contain numeric answers, instead they are categorized as strings with 6 levels from strongly disagree to strongly agree. We used a numeric dictionary ranging from 1 to 6 to enumerate these values into numbers to aid numerical analysis.

2.4 DESCRIPTIVE STATISTICAL ANALYSIS

Before progressing on to running statistical tests on our data, we must perform some descriptive statistics to describe and summarize the OGR survey data. This process will help us better understand the data we are working with when testing statistical models.

2.4.1 CATEGORICAL COUNT PLOTS ABOUT RACE AND AGE GROUPS

Chart, bar chart

Description automatically generatedThe statistical tests that are performed in later sections are focused on quantifying the differences in answer rating between categories of Race as well as Age Groups. The following plots describe the number of observations within each group in the survey.

In the *Count of Students by Race/Ethnicity* plot, we can see that the partipants are grouped into four different races/ethnicities: Caucasian, Black, Multi-Racial, and Other. It is clear that an imbalance exists between the four groups of race. Caucasian’s make up approximately half of all partipants in the program in both years; though in 2019 there is a large drop in caucasian participants.

The second plot, *Count of Participants by Age Group,* depicts the number of participants by age group. We do not have the same imbalance as metioned before, which supports a “fair” analysis between the two groups. The plot does show a decrease in participants in the older group in 2019 vs 2018, which might suggest that more students “Graduated” from the program in 2018 than new participants joining in 2019.

Chart, bar chart

Description automatically generated

2.4.2 NUMERICAL DISTRIBUTIONS

Text

Description automatically generatedThe below statistical table and plots depict the mean answer rating and standard deviation across all questions in the suvey for 2018 and 2019. It is important to note that the last thre questions (1, 2, and 3) are scored differently than question 15 – 25. The last thre questions range from 1 – 6 and were originally categorized from strongly disagree (1) to strongly agree (6). Due to this difference, we observe a bimodel distribution in the mean answer ratings, as shown in the *Mean Scores for all Questions* KDE plot.

2.4.3 SUMMARY OF ANSWER RATINGS BETWEEN PRE-TESTS AND POST-TESTS

Chart, box and whisker chart

Description automatically generatedThe box plot below shows the distribution of answer ratings between the pre-test and post-test survey results.

2.4.4 CORRELATION AMONG ANSWER RATINGS

Chart, treemap chart

Description automatically generated

**3 STATISTICAL TESTS AND RESULTS**

**3.1 RESEARCH TOPIC I**

The first topic to be analyzed asks: Are there significant differences in Answer Rating for various outcomes in the younger group and the older group? This topic will help the OGR program better understand which questions significantly improve over time and identify the questions that don’t show a significant change, dependent on the age group of participants.

For this research topic, we will seek to answer the following questions:

1. Are the mean answer ratings different between the older group and the younger group of participants?
2. Does the effectiveness of the Omaha Girls Rock Program differ between age groups?
   1. Which survey questions (per age group) have a significant increase in answer ratings after the OGR Program?
   2. Are there differences in the mean **delta score** between age groups for each question?

*3.1.1 COMPARE ANSWER RATINGS BETWEEN AGE GROUPS*

**Chart

Description automatically generated**Are the mean answer ratings different between the older group and the younger group of participants? Answering this question will allow the Omaha Girl Rock Program to understand if participants of different age group (younger and older) inherently answer the survey questions differently. If a significant difference is determined, OGR could further adjust the program based on the age of a participant.

Figure 1: Distributions of Answer Ratings by Age Group

**The Statistical Test**

We are interested in comparing the answer rating among all questions, years, and tests between the two age groups, older & younger. Specifically, we want to test if there is a statistically significant difference between the answer ratings between the older group and the younger group. The appropriate test that we will conduct is a **95%** **two-tailed independent t-test.**

Hypotheses:

Test Statistic:

**Results and Conclusion**

There is no significant difference, t = 0.49, p = 0.621, despite the older group (M = 5.252, SD = 1.59) providing slightly higher answer ratings compared to the younger group (M = 5.229, SD = 1.703).

There is not enough evidence in the data to reject the null hypothesis. Therefore, we are unable to claim that there is any statistically significant difference in the mean answer ratings between the younger group and the older group.

3.1.2 COMPARE PRE-TEST AND POST-TEST SCORES BETWEEN AGE GROUPS

For the second comparison, we are interested in discovering which survey questions (per age group) have a significant increase in answer ratings after the OGR Program? The results will inform us if the Omaha Girls Rock Program is effect at increasing the answer ratings per question and age group.

**The Statistical Test**

Since we are seeking statistical evidence for an increase in answer rating in the post-test from the pre-test, we will be using **a right-tailed dependent (matched-sample) t-test at a 5% level of significance**. We will perform this test 28 different times for each combination of age group and question (2 groups \* 14 question). Conducting the test this way will show if a particular question for a particular age group is significance. For example, is there a significant improvement on question 15 for the older age group? What about for the younger age group? Which age group has a higher number of significant increases?... These are the questions this analysis will answer.

Hypothesis

Test Statistic

**Where:**

**Results**

The table below displays the results of the 28 individual dependent t-tests, where each row is one test. The survey data is iteratively filtered and tested for each combination of age group and question.

Table

Description automatically generated

Conclusion

Many of the questions have a statistically significant increase after the Omaha Girls Rock Program and only question 17 was not significant for both groups. The older group had an increase in 12 out of the 14 questions, whereas the younger group had a significant increase in 9 out of the 14 questions. From these tests, it appears that OGR is slightly more effective for participants in the older group.

3.1.3 COMPARE PRE-TEST AND POST-TEST SCORES BETWEEN AGE GROUPS

**3.2 COMPARISONS BETWEEEN RACE/ETHNICITY GROUPS**

The first topic to be analyzed asks: Are there significant differences in Answer Rating for various outcomes based on their racial identity/ethnicity? Again, this is to help the OGR program understand where differences may lie in the effectiveness of their programs based on a girl’s racial identity so that they can better prioritize on how to make the programs equally beneficial for all.

For this research topic, we will seek to answer the following questions:

1. For each survey question, is there a significant difference in delta scores between each racial identity?
2. Does race, the survey question, and the interaction of race and the survey question have a significant effect on the delta scores of the OGR participants?
3. For survey questions groups by personality traits asked about, is there a significant difference in the delta scores between each racial identity?

*3.2.1 ONE-WAY ANOVA (PER QUESTION)*

For each question, is there a significant difference in the delta scores between each race/ethnicity? With the results of this analysis, we should be able to tell if there are any social scenarios where different groups of girls are improving more or less than others. This will inform the OGR program on which scenarios and demographics they might want to be more mindful about to assist.

**Statistical Test**

As we are examining more than 2 populations when we are looking at the population means of the delta changes in survey answers, we are using a **One-way ANOVA test at a 95% level of significance** for each survey question, running it a total of 14 times. This will test us if the delta means for at least one of the populations significantly differs from another and for which survey question(s) that is occurring.

**Hypothesis**

**Results**

Table

Description automatically generated

**Conclusion**

Surprisingly, we found that none of the population means of the delta changes differed for any of the questions. We had expected there to be at least one question where it had. Since we were unable to reject our null hypothesis for any survey question, it appears that each of the girls saw similar amounts of improvements, regardless of their racial identity/ethnicity.

*3.2.2 TWO-WAY ANOVA (RACE + QUESTION)*

Does race, the survey question, and the interaction of race and the survey question have a significant effect on the delta scores of the OGR participants? We wanted to double check our results from the One-way ANOVA tests, so by checking if the interaction of race and the survey was significant, we would be able to either check if the results lined up or if we had made a mistake.

**Statistical Test**

To check the interaction or race and the questions, we ran a **Two-way ANOVA test at a 95% level of significance**. This would show us three things: if the girls’ racial identities significantly impacted the delta scores, if the questions themselves significantly affected the scores, and if the combination of the racial identity and the survey question had any significance.

**Hypothesis**

**Results**

Table

Description automatically generated

**Conclusion**

From the results, we can see that the girl’s racial identities and the survey questions themselves significantly affected the delta scores to the survey. The interaction of the racial identities and the survey question, however, was not significant. This was in line with our one-way ANOVA tests, making it appear that the girl’s racial identity had no significant impact on how their responses changed.

*3.2.3 ONE-WAY ANOVA (PER PERSONALITY TRAIT)*

The second question that our analysis focuses on is: Does the race/ethnicity of participants influence the change in scores for various questions over time? Answering this question will help OGR better understand if their program has differentiating outcomes for participants based on the race/ethnicity of a participant. For each question in the survey, we will analyze the differences in the delta between the pre-test and post-test scores between each race/ethnicity group.

After the processing of the survey data, we narrowed down the number of race/ethnicity groups from ~12 groups (many with very little observations) to just 4 groups (Caucasian, Black, Multi-Racial, Other).

* Method Used
  + One-way Anova
  + Fisher’s LSD

Because we are working with multiple populations with one population for each ethnicity we are examining, we decided to use a One-Way Anova for each of the questions with a significance level of 95% to see if there were any questions where the population means were not equal. This was calculated by seeing if their p-value for running Anova was less than 0.05. If we did find this, this meant that at least two of the populations had a significantly different response to that question. To determine which populations these were, we then ran Fisher’s LSD procedure. For each pair of populations, the populations were determined to be significantly different if the absolute value was greater than the critical value.

* Hypotheses
  + H\_0: \mu\_caucasian = ….
  + H\_a: Not all population means are equal

Since we are using One-Way Anova, our null hypothesis is that all the population means for a given question are equal. This means that there is no significant difference between how the girls responded regarding their racial identity. The alternative hypothesis is that not all the population means are equal, meaning that racial identity can affect how the girls responded to that question.

* Results
  + Table/Visuals
  + Body Text

Our first table shows all of the p-values that were calculated for each question. Only two questions have a p-value that was less than our significance level, 0.05: questions 21 and 23. For these two, we had to reject the null hypothesis, meaning that at least one population is not equal. For the rest, we are unable to reject the null hypothesis, meaning that there is no significant difference in the girls’ responses regarding their racial identity.

This second table is our results of having run Fisher’s LSD for question 21, asking how likely they would be to ask someone over to their house on a Saturday. There are two pairs where the populations were significant, the Caucasian and black populations and the multi-racial and black populations.

This third table is from when we ran Fisher’s LSD for question 23, asking how likely the girls would be to go to a party where you are sure you won’t know any of the kids. There are two pairs where the populations had significant differences again. The Caucasian and black populations differed for this question, similarly as they did in question 21. The black population also differed from our population of girls where there were not enough data points to form their own specific population.

* Conclusion/Meaning
  + Text

In terms of racial identity, we were unable to reject out null hypothesis for most of the survey questions, meaning that participation in the Omaha Girls Rock does increase their comfortability in doing various social activities, regardless of their ethnicity. For the two questions where there were significant differences, in terms of the distributions of changes in answers, it appears that black girls who attended the programs had greater increases in their comfortability in performing these tasks. More investigation should take place to analysis differences between the groups and the effectiveness of the program across all demographics.

4 SUMMARY AND CONCLUSION

Looking at our research analysis, we have found that there is no substantial difference in the survey scores or program effectiveness for the participants between the two age groups. This is a great finding as it signifies that everyone is equally growing through participating in the programs and no one is being treated unequally because of their age. In examining the data through their racial identities, we see only minimal differences in how they answered, and there is a possibility that the data was skewed because of outliers due to the small population sizes we were working with. Further analysis is recommended, especially with larger population sizes to be able to make either substantiate what we have found or to disprove it.

Finally, we do want to point out that even if the girls have developed their social skills by similar intervals, that does not mean that they are all at the same skill level. Girls in the black populations had lower answer ratings for questions regarding their growth mindsets but higher for questions asking how extroverted and agreeable they were. Again, more analysis is recommended to collaborate our findings, but these should also be considered so that all girls who participate in Omaha Girls Rock programs can leave on level skill levels to take forth into their futures.