Statistical Analysis of OGR Program with a Focus on Age and Race

**SP 19**

Joshua Ellis (18862025), Alex daly (46488761), satwik chandra ()

ABSTRACT

Omaha Girls Rock, a local non-profit that puts on programs for young local girls in the area, has run surveys before and after their programs to see if the girls have developed their social abilities through the course of their participation. We took that data to see if there was any significance difference in how the girls’ answers changed before and after the program, first based on which age group they were in using matched sample test and then based on their racial identity using one-way anova. We found there was no significance difference detected to how the girls answered the questions, regardless of their age, but we did see that a couple questions were significant different based on their racial identity. This second observation should be examined further by Omaha Girls Rock to determine why and what can be done about it.

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

|  |  |  |
| --- | --- | --- |
| 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

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

3 DESCRIPTION OF STATISTICAL TESTS AND RESULTS

3.1 RESEARCH QUESTION I

The first question to be analyzed asks: Are there significant differences in answer ratings for various outcomes between the younger group and the older group? This question will help OGR 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.

***Statistical Method***

We are interested in comparing the scores between the pre-test and the post-test (delta score) by the age group of participants. We want to know which questions (if any) significantly improved after the OGR program between the younger group and older group participants. Since we are seeking a positive improvement where $\mu\_d > 0$ we conducted a right-tailed dependent (matched-sample) t-test at a 5% level of significance. This test is run for each age group and question combination resulting in 28 different statistical tests (2 groups x 14 questions).

***Hypotheses***

The objective of this test is to determine whether the scores of a particular question increased for each age group. Therefore, we are seeking for evidence that the program increased survey scores. Since we are conducting a right-tailed test, the alternative hypothesis will take on a “greater than” condition where the null hypothesis will take on a “less than or equal to” condition.

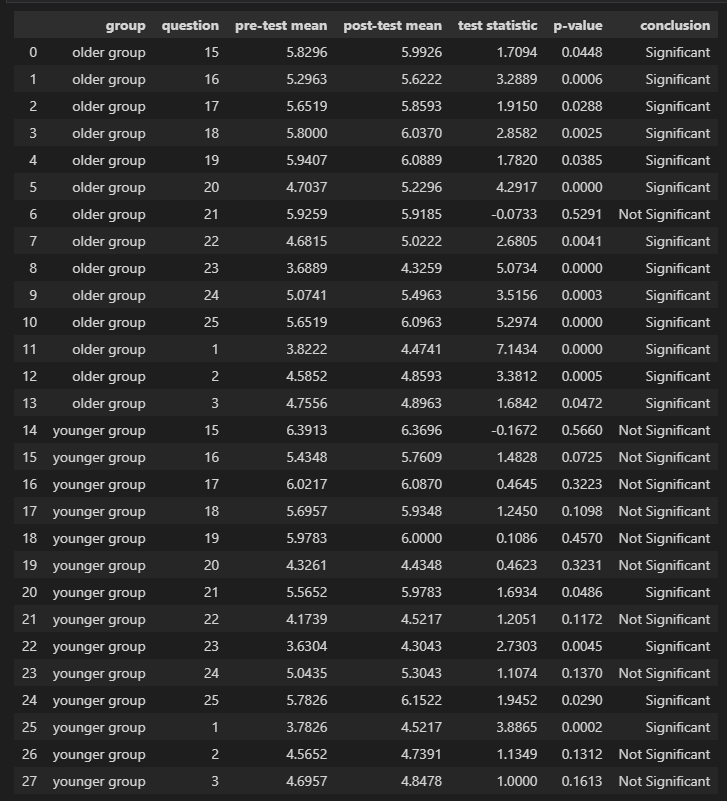
$H\_0: \mu\_d \le 0$

$H\_a: \mu\_d > 0$

***Results***

The results of the 28 tests are summarized into a table. This table consists of group, question, pre-test mean, post-test mean, test statistic, p-value, conclusion. Here the table shows there is only one question (question 21) in older group where p-value is greater than 0.05. Except question 21 all the p-values are less than or equal to 0.05. In the case of younger group, questions 1,21,23,25 is significant i.e., p-value for those is less than or equal to 0.05. Barring them, all the p-values in the younger group are greater than 0.05.

A key take-away is that the program seems to be more effective for participants in the older group. In the older group, 85.7% of questions in the survey had a statistically significant increase at a 5% level of significance between the pre-test and the post-test, only questions 2 and 21 did not have a statistically significant improvement. Comparatively, in the younger group, only 28.6% of the questions had a statistically significant improvement between the pre-test and post-test.



***Conclusion***

The results from the test show there is a complete contrast between the older group and younger group. In terms of the older group there is an almost all the questions where there is a significant difference between pre-test and post-test mean. So, the null hypothesis is rejected. In the younger group, there is not much difference between the pre-test and post-test except for a few questions. Hence the null hypothesis is accepted but not completely.

3.1 RESEARCH QUESTION II

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

What is your overall conclusion?