Data Analysis for Omaha Girls Rock

**FA 21**

Name, NU-ID

Abstract. Provide a short overview (a couple of sentences) that describe your project.

1. INTRODUCTION

Since 2011, Omaha Girls Rock’s mission is to empower youth to find their unique voice through music education, performance and creativity. Our goal is to cultivate strong, contributing members of our community. Their future starts with Omaha Girls Rock. Omaha Girls Rock's primary role is to create opportunity for cultural expression and equitable access to the arts. Our work is to develop voices who want to do good in their community by allowing them to identify their sense of self through music and other artistic outlets.   Our programs use music education: songwriting, technical skill building, forming a band, and historical impact of women in music to accomplish our goals. Here are the programs and ways in which we work:

* **Summer Camps for ages 4-16: (this is the program that was assessed and provided the data for your class!).**Divided into two week-long sessions, OGR’s summer camp program serves approximately 200 youth per year. The camp is an intensive, six-day program that includes instrument instruction, workshops, songwriting and band practice, live performances by visiting artists, and a final showcase concert. Volunteer musicians provide campers with over an hour of instrument instruction each day in small groups organized by skill level and age. At the beginning of camp, the girls form bands. During their daily band practice, they write music and lyrics together with assistance from band coaches. Campers also attend workshops on topics ranging from songwriting to diversity and self-esteem. At the end of camp, the bands each perform their original song at a local music venue in front of their families and the community.
* **OGR Academy: ages 8-14**
  + After School Program with Collective for Youth, Kids Can, Completely Kids and Omaha Public Schools(grades 2-8)l:  With thoughtful presentations and engaging activities, students learn about and celebrate historical women in music and beyond, but are also empowered to use what they've learned to celebrate themselves. Students are able to synthesize and apply their learning to develop a greater sense of self which will reflect in an improved self confidence.  In addition, the gain leadership skills through participation in thoughtful discussions, collaboration and cooperation in small group projects.
  + Instrument Instruction:  Using the historical impacts of women in music, students receive a comprehensive approach to instrument instruction.  Students receive 16 weeks of of one on one virtual instruction, 1 hour a week.  The focus on impacts women have made throughout history, specifically those that have been left out of or break through oppressions in the historical narrative, resonate with students so they are able to feel represented and further the strengthening of their voice and technical skill base in their instrument.

In this course project, you are going to help Omaha Girls Rock improve their program by providing sophisticated statistical data analysis for a set of research questions. Specifically, OGR conducts pre- and post-participation surveys using a range of questions to measure the impact of their program on participants. Along with the Likert-scale data from the question, for each participant they collect additional demographic information, e.g., age, ethnicity, and ZIP code.

The main goals of this course project are to a) analyze various question answers to identify where participants significantly improved over time, and b) identify which factors influence the change/improvement over time.

**Team 1, 2: Measuring differences over time, accounting for race/ethnicity of participants**

You are going to work on a project that focuses on two questions.

**Research question 1**: Are there significant differences in answer ratings for various outcomes in the younger group and the older group?

Outcomes are the ratings per question, often measured on a scale from 1-7, etc. This will help OGR to better understand which aspects significantly improve over time, and identify aspects that don’t show a significant change yet.

* Hypothesis test about the mean scores for each question between the two age groups
* What questions have a statistically significant difference between the two age groups?
* Different results for pre-test and post-test (do all years)
  + Population 1: Older Group
  + Population 2: Younger Group
    - H0: Older Group = Younger Group
    - Ha: Older Group! = Younger Group

**Research question 2**: Does the race/ethnicity of participants influence the change in scores for various questions over time?

Instead of simply looking at the change in scores over time, you will also include the race/ethnicity in your considerations. Do outcomes (e.g., changes in scores) differ based on the race/ethnicity of the students? Is there a significant change for one ethnicity, but not the other?

* Multiple Paired Samples t-test for each race/ethnicity in the data

Open question: As a team, think about what other relevant questions can be studied with the given data and propose 1-2 such additional questions.

**Note: Team 1 should work on Questions 1-14, Team 2 on Questions 15 – 25 plus the three last questions.**

**Team 3, 4: Measuring differences over time, accounting for age of participants**

You are going to work on a project that focuses on two questions.

**Research question 1:** Are there significant differences in answer ratings for various outcomes in the younger group and the older group?

Outcomes are the ratings per question, often measured on a scale from 1-7, etc. This will help OGR to better understand which aspects significantly improve over time, and identify aspects that don’t show a significant change yet.

**Research question 2**: Does the age of participants influence the change in scores for various questions over time?

Instead of simply looking at the change in scores over time, you will also include the age in your considerations. Do outcomes (e.g., changes in scores) differ based on the age of the students? E.g., do younger participants on average see higher or lower average improvements than older participants?

Open question: As a team, think about what other relevant questions can be studied with the given data and propose 1-2 such additional questions.

**Note: Team 3 should work on Questions 1-14, Team 4 on Questions 15 – 25 plus the three last questions.**

**Group 4: Add one more research question**

**Team 5: Measuring differences over time, considering returning participants**

You are going to work on a project that focuses on two questions. Your team will work on a modified dataset that you will derive from the 2018 and 2019 datasets. Specifically, you will identify the participants that participated in both years, and use this set of participants in your analysis. In this case, per participant you have 4 measures: pre-survey 2018, post-survey 2018, pre-survey 2019, and post-survey 2019.

**Research question 1:** Are there significant differences in answer ratings for various outcomes in the younger group and the older group, when comparing the post-survey 2018 with the pre-survey 2018?

Outcomes are the ratings per question, often measured on a scale from 1-7, etc. This will help OGR to better understand which aspects significantly improve over time, and identify aspects that don’t show a significant change yet.

**Research question 2:** Are there significant differences in answer ratings for various outcomes in the younger group and the older group, when comparing the post-survey 2019 with the pre-survey 2019?

Open question: As a team, think about what other relevant questions can be studied with the given data and propose 1-2 such additional questions.

**Group 5: Add one more research question**

**Team 6: Identifying factors that increase or decrease the improvement probability (Questions 1-14)**

You are going to work on a project that focuses on two questions. Your team will conduct a logistic regression to identify relevant factors that influence whether participants are more or less likely to show improvement in ratings for the considered questions. You will build a logistic regression model for each of questions 1-14.

**Research question 1:** Considering the 2018 pre- and post-survey results for the various questions, are there significant factors that increase or decrease the chances of improvements?

For this question, calculate the change in scores/ratings per student and question. Convert positive changes to a 1, and negative changes (or no change) to a 0. Then, use this 0/1 outcome per question to run logistic regressions, using the demographic variables as inputs.

**Research question 2:** Considering the 2019 pre- and post-survey results for the various questions, are there significant factors that increase or decrease the chances of improvements?

This is similar to the previous question, however now you use the 2019 data. Convert the differences to a 0/1 variable again and conduct logistic regression to identify differences.

Open question: As a team, think about what other relevant questions can be studied with the given data and propose 1-2 such additional questions.

**Group 6: Add one more research question**

**Team 7: Identifying factors that increase or decrease the improvement probability (Questions 15-25, plus last three questions)**

You are going to work on a project that focuses on two questions. Your team will conduct a logistic regression to identify relevant factors that influence whether participants are more or less likely to show improvement in ratings for the considered questions. You will build a logistic regression model for each of questions 15-25, plus the last three questions in the excel sheet.

**Research question 1:** Considering the 2018 pre- and post-survey results for the various questions, are there significant factors that increase or decrease the chances of improvements?

For this question, calculate the change in scores/ratings per student and question. Convert positive changes to a 1, and negative changes (or no change) to a 0. Then, use this 0/1 outcome per question to run logistic regressions, using the demographic variables as inputs.

**Research question 2:** Considering the 2019 pre- and post-survey results for the various questions, are there significant factors that increase or decrease the chances of improvements?

This is similar to the previous question, however now you use the 2019 data. Convert the differences to a 0/1 variable again and conduct logistic regression to identify differences.

Open question: As a team, think about what other relevant questions can be studied with the given data and propose 1-2 such additional questions.

**Group 7: Add two more research questions**

1. DATA DESCRIPTION

What data will be used?

Following data could be provided:

* Student demographics (age, race/ethnicity, ZIP code, times participated in camp, etc)
* Answers (scores/ ratings) to 25 core and 3 additional questions

Descriptive (numerical) statistics (how are the variables distributed, number of overall observations, correlations, etc.)?

Graphical (exploratory) description of the dataset: obvious differences between groups?

1. OUTLOOK OF PLANNED METHODS (ONLY FOR MILESTONE 1)

How do you want to proceed to answer the research questions?

Which methods do you plan to use?

1. DESCRIPTION OF STATISTICAL TESTS AND RESULTS

Description of statistical methods:

* Which methods did you use?
* What are the respective hypotheses?

Method evaluation:

* What are the results of your statistical hypothesis tests?
* What do the results mean for your research questions?

1. SUMMARY AND CONCLUSION

What is your overall conclusion/recommendation?

1. REFERENCES

Include references where appropriate.