

MUSI 8803/4843

Assignment 7: Statistics Planning

Group #: 6

Group member names:

Sile Yin

Brittney Allen

Karl Hiner

In this assignment, I would like you to work with your final project group members to carefully consider all the variables you will have, how you are measuring them, and what statistical tests you could use to evaluate your hypotheses.

1. (1 point) List all hypotheses that you still plan to test as a part of your final project:

Our only hypothesis is: The proportion of I-V-vi-IV quadgrams, or any of its rotations (henceforth referred to as 1564-R), in the Billboard Top-100, will trend downward from 2016 to 2022.

2. (1 point). For each hypothesis, list an appropriate null hypothesis (note: this should relate to your statistical test).

Null hypothesis: There is no correlation between the year and the ratio of the use 1564-R, in Billboard Top-100 songs.

3. (1 point). For each hypothesis, list what are you ultimately trying to demonstrate. (E.g., differences between groups on some variable? Relations between variables? Difference between a sample and a hypothetical population?)

We are looking for a relationship between variables (year and the proportion of 1564-R).

4.

- (a) (1 point) For each hypothesis, list what you believe to be the outcome variable(s) and predictor variable(s). Recall that in some cases it is possible to 'flip' the variables with no real damage to the outcome (e.g., with correlation or linear regression) and also that in some cases there may only be one variable observed with no independent variables.

- (b) (3 points) For each outcome and predictor variable(s), list:

- what it is (e.g., counts, proportion, score, measurement, category, etc.)
- whether it can/should be treated as discrete or continuous
- what type of measurement scale it uses (e.g., interval, ratio, ordinal, nominal)

- If the data collected is likely to fit a known distribution or not?
- how many 'levels' it has (if appropriate).

Predictor variable: **Song release year**

- Categorical
- Discrete
- Interval
- Interpreting this as the number of songs released per year charting on Billboard, we will assume the population will be normally distributed.
- Although we could consider our year intervals as each being one level, we aren't formulating our question in this way.

Outcome: **Proportion of I-V-vi-IV quadgrams, or any rotations, in the year**

- Proportion
- Continuous (a real number in $[0, 1]$, per year)
- Ratio
- Since the source of our measured chord ratio will arise from many true/false outcomes (for each chord, it either is or isn't one of 1564-R), we would expect that each year would have an expected value for the ratio following a beta distribution.
- Since our outcome variable is continuous, it has no levels.

5. (3 points) Based on your answers to the above questions, list all the possible statistical tests you *could* use to test each of your relevant hypotheses (including multiple tests if appropriate), and then decide amongst your group which test you think would be the best choice. Defend your choice.

Since our hypothesis concerns only whether or not the ratio of progressions is either increasing or decreasing (without regard to the magnitude), we plan to use the Mann-Kendall statistical test. This test evaluates data collected over time for consistently increasing or decreasing trends.

The only other possible choice of statistical test we could think of was Linear Regression. Once we learned about the Mann-Kendall test, however, it seemed more appropriate than Linear Regression since our hypotheses isn't concerned with whether or not a potential trend is linear, but only whether or not it's increasing over an interval of time.