Inequality of Success in Online Music Markets

The Music Lab Project is a web based randomized experiment designed to investigate the role of social influence in the market success of songs. This project examines whether songs becomes "hits" not only because of their inherent *musical* qualities but also because of a *social* process: when people hear that a song is popular, they become more inclined to like the song themselves. This exercise is in part based on:

Salganik, Matthew J., Peter Sheridan Dodds and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311(5762): 854-856.

Two experiments were conducted in this study using a web platform called *Music Lab* where users listen to music, rate it, and optionally download songs of their choice. For this exercise, we use a subset of the original data and focus on a smaller number of treatment conditions. In **Experiment 1**, users were shown a list of previously unknown songs of unknown music bands. They were randomly assigned to one of two conditions: *Independent* and *Social influence*. In the *Independent* condition, users decide which songs to listen to based on their titles and the names of the bands alone. After listening to songs, they were asked to rate each of them (from one star for "I hate it", to five stars for "I love it"). Finally, they have an option to download any of the songs they listened to. In the *Social Influence* condition, everything is identical to the *Independent* condition except that users are provided with additional information about the number of times each song was downloaded by other users.

Experiment 2 is similar to **Experiment 1** except that under the *Social Influence* condition, the songs were ordered according to the number of downloads (For the sake of simplicity, we will ignore a minor difference in the way in which the *Independent* condition was administered). Thus, although the information provided to users is identical under the *Social Influence* condition, in **Experiment 2** this information is presented visually in a different manner on the website. Note that while the randomization of treatment assignment was done within each experiment, no randomization was used when assigning each user to one of the two experiments because the experiments were conducted sequentially.

The researchers hypothesize that the existence of social influence contributes to the inequality of success in music markets. According to this hypothesis, we expect the degree of inequality to be greater under the *Social Influence* condition than the *Independent* condition within each experiment. We will analyze a portion of the original data. The names and descriptions of variables in each data set are shown below. Note that it is impossible to connect individual users to the songs they listened to or downloaded from these data sets.

1. Data sets about songs: `songs1.csv` for **Experiment 1** and `songs2.csv` for **Experiment 2**

Name	Description
`song_id`	Song id
`listen_soc`	Number of times each song was listened to by users in the *Social Influence* condition
`listen_indep`	Number of times each song was listened to by users in the *Independent* condition
`down_soc`	Number of times each song was downloaded by users in the *Social Influence* condition
`down_indep`	Number of times each song was downloaded by users in the *Independent* condition

2. Data sets about users: `users1.csv` for **Experiment 1** and `users2.csv` for **Experiment 2**

Name	Description
`id`	User id
`world_id`	`1` if assigned to the *Social Influence* condition, and `9` if assigned to the *Independence* condition

Name	Description
`country_code`	Code for user's country of residence
`country`	String for user's country of residence
`web`	User's ability to use the world wide web
`visit`	User's frequency of internet visits to consult about music or concerts
`purchase`	`1` if user purchased a song in the past as a result after listening to it on the web, and `0` otherwise

Question 1

Within each experminent, compute the proportion of users assigned separately for the *Social Influence* and *Independent* condition. Summarize the results as a table of proportions for each experiment.

##Answer 1

```
## Load user data for each experiment
users1 <- read.csv("data/users1.csv", header = TRUE)
users2 <- read.csv("data/users2.csv", header = TRUE)
## Recode the variable with informative labels
users1$world_id[users1$world_id == 1] <- "Social Influence"
users1$world_id[users1$world_id == 9] <- "Independent"
users2$world_id[users2$world_id == 1] <- "Social Influence"
users2$world_id[users2$world_id == 9] <- "Independent"
## Generate the proportion of users in each condition
p_exp1 <- prop.table(table(users1$world_id))
p_exp1
```

```
##
##      Independent Social Influence
##        0.6724218        0.3275782
```

```
p_exp2 <- prop.table(table(users2$world_id))
p_exp2
```

```
##
##      Independent Social Influence
##        0.6772834        0.3227166
```

In **Experiment 1**, the proportion of users assigned to the *Social Influence* condition was 0.328, while 0.672 where assigned to the *Independent* condition. Similarly, in **Experiment 2**, the proportion of users assigned to the *Social Influence* condition where 0.323, while 0.677 where assigned to the *Independent* condition. In both experiments, about 2/3 are assigned to the *Social Influence* condition.

Question 2

Within each experiment, compute the *average number of downloads per user* separately for the treatment and control conditions. Note that the number of users is different between the conditions. Comment on the differences across the two conditions within each experiment. Repeat the same using the number of times each song was listened to.

Question 3

We examine the main hypothesis of the study by investigating whether social influence increases the inequality of success in music markets. We measure inequality using the Gini coefficient, which will be covered in Chapter 3 of *QSS* in detail. The Gini coefficient ranges from 0 (most equal) to 1 (most unequal). In the current context, the coefficient is equal to 0 if every song has the same number of downloads whereas it is

equal to 1 if all users download the same song. To compute this measure, we can use the `ineq()` function available in the **ineq** package. Within each experiment, compute the Gini coefficient separately for the *Social Influence* and *Independent* conditions. Interpret the results in light of the hypothesis. Repeat the same analysis using the number of times each song was listened to.

Question 4

Within each experiment, compare the characteristics of users between the *Social Influence* and *Independent* conditions. In particular, compare the mean values of `web`, `visit`, and `purchase` variables. Interpret the results in light of the internal validity of the conclusions you draw for each study in the previous question.

Question 5

Compute the difference in the estimated average effect of the *Social Influence* condition on inequality of success between the two experiments. Under the experimental design of this study, does this between-study comparison have as much internal validity as the within-study comparison you conducted in Question 3? Why or Why not? Do the data provide any information regarding the internal validity of this between-study comparison?