**BA 830 Experiment Report**

*Cohort A Team 8: Huiying Ba, Danting Huang, Ziqin Ma, Jiao Sun, Senbo Zhang*

**Causal Question of Interest**

Our causality interest originated from the phenomena that a majority of students in the Pardee library often wear earbuds when studying, and we are interested in the effect of listening to music on memory abilities. Given the time and space for a course project, the effect of music long-term memory is difficult to be precisely measured, so we decided to experiment on instant memory. Consequently, our causal question became: does listening to music have any potential causal effect on instant memory?

**Experiment Design**

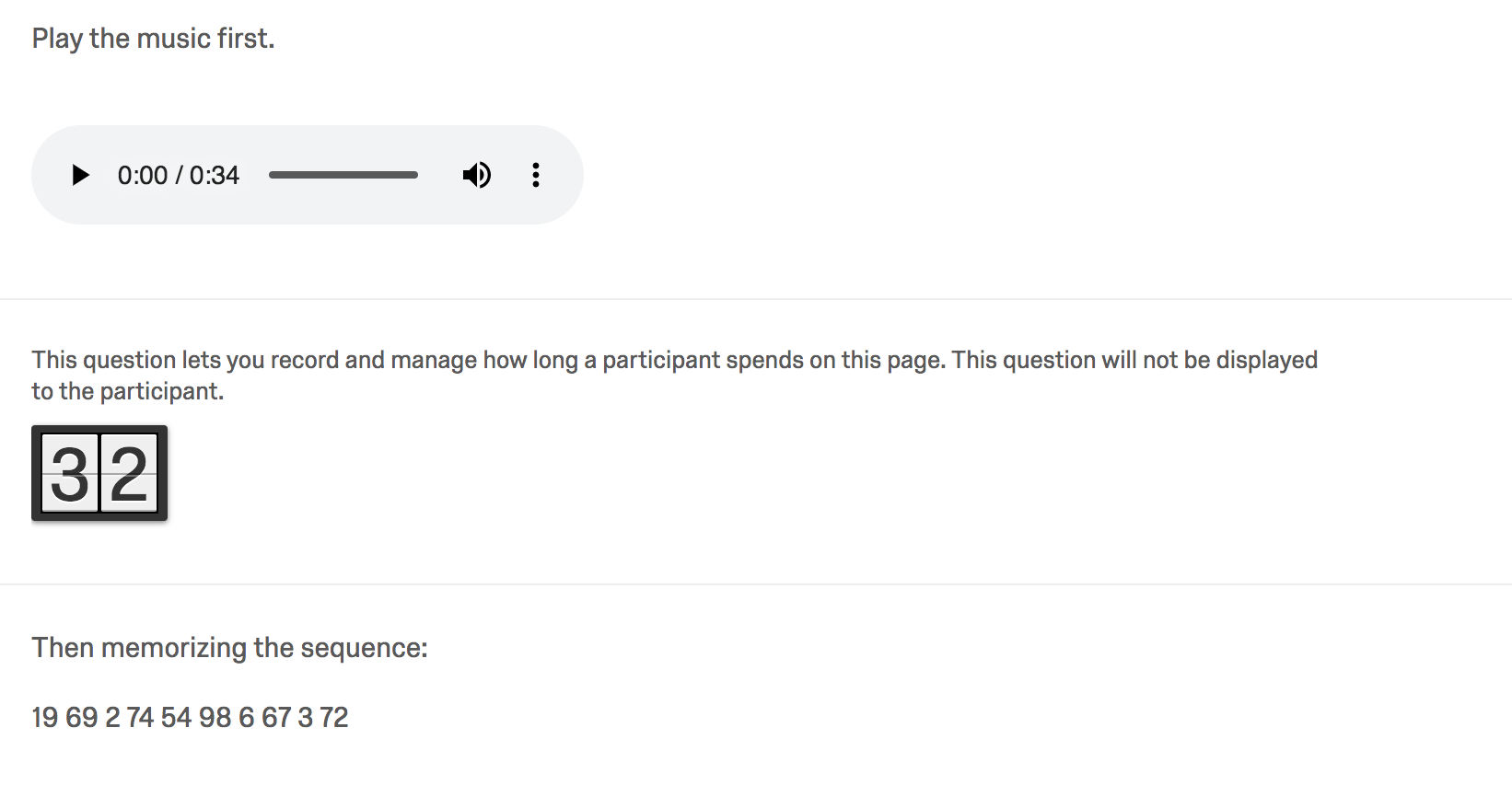
Our experiment design followed two "principles": randomization and simplicity. We always need randomization to make sure that the result would not be influenced by selection bias. We also want to make the process simple, short, and easy to follow, so that participants would not feel intimidated or bothered then refuse to take.

The experiment task is memorizing a sequence with 10 integers within 100 in 30 seconds. Treatment and control groups have different sequences, and both were generated randomly from an online number sequence generator. We chose number sequence because people across different backgrounds share the algebraic number system, and no additional skills are required to memorize things, which is a fair game for everyone who participated. In terms of music choice, we searched through several platforms based on three rules: no lyrics, not popular, and distractive enough. Words itself is a distraction other than tunes, and we have no way to detect whether the result comes from the melody or lyrics, especially if the sample has various language backgrounds. The music cannot be popular for the same reason: people may have different levels of familiarity with a broadly-defined "popular" piece, and this could result in a potential bias if we happen to have the sample all extremely familiar with the piece. The music needs to be distractive enough to make sure that the treatment is effective across environments. We ended up with a piece of background music from a Japanese animation in 2010. We cut the first 34 seconds and used that excerpt as of our treatment.

To save time for participants and us, we made the task online by Qualtrics. We created a survey each for treatment and control with the same structure and wording. The survey started with a brief introduction to the task and descriptions, following a page with instructions. Next is the task page, including a media file for treatment, a timer, and a number sequence. Once the timer stops, the site will automatically jump to the answer page with 10 boxes and a 30-second timer. The survey will end with three multiple-choice questions: level of study (undergraduate/graduate/non-student), whether listening to music while studying (yes/no/maybe), and task difficulty.

To ensure the tasks are completed under the same conditions, we highlighted several points in the instruction. First, participants need a set of earbuds for the task and must finish individually. Second, the timer is set to 32 seconds (2 more seconds) to leave participants enough time to click the play button of the media, and they need to click it as quickly as they can.

The most important notice is: *continue the task even if you DON'T hear the sound after clicking.* The task for the control group merely is memorizing a number sequence; putting this task into the survey on the task page, there would be only a sequence and timer. However, we want the only difference between treatment and control is "listening to music" rather than "click and listen to music." So we inserted a media of 34 seconds of silence on the control group's task page, hence the motion to start both surveys is the same.

Treatment Card Treatment Task Page

For the convenience of distribution, we made two types of small cards with QR codes of treatment and control surveys. The target population was students in the Hariri Building. During the three days, group members went to different locations to distribute cards several times a day. The numbers of treatment cards and control cards distributed were the same each day. Before heading out, we mixed and thoroughly shuffled both cards, and each member randomly grabbed a pile of them. Then we gave out cards to randomly picked students and told them we were running a mini-experiment for our course project. To maximally reduce the personal differences among group members, we practiced our introduction, and we cannot choose the assignment given to a specific participant.

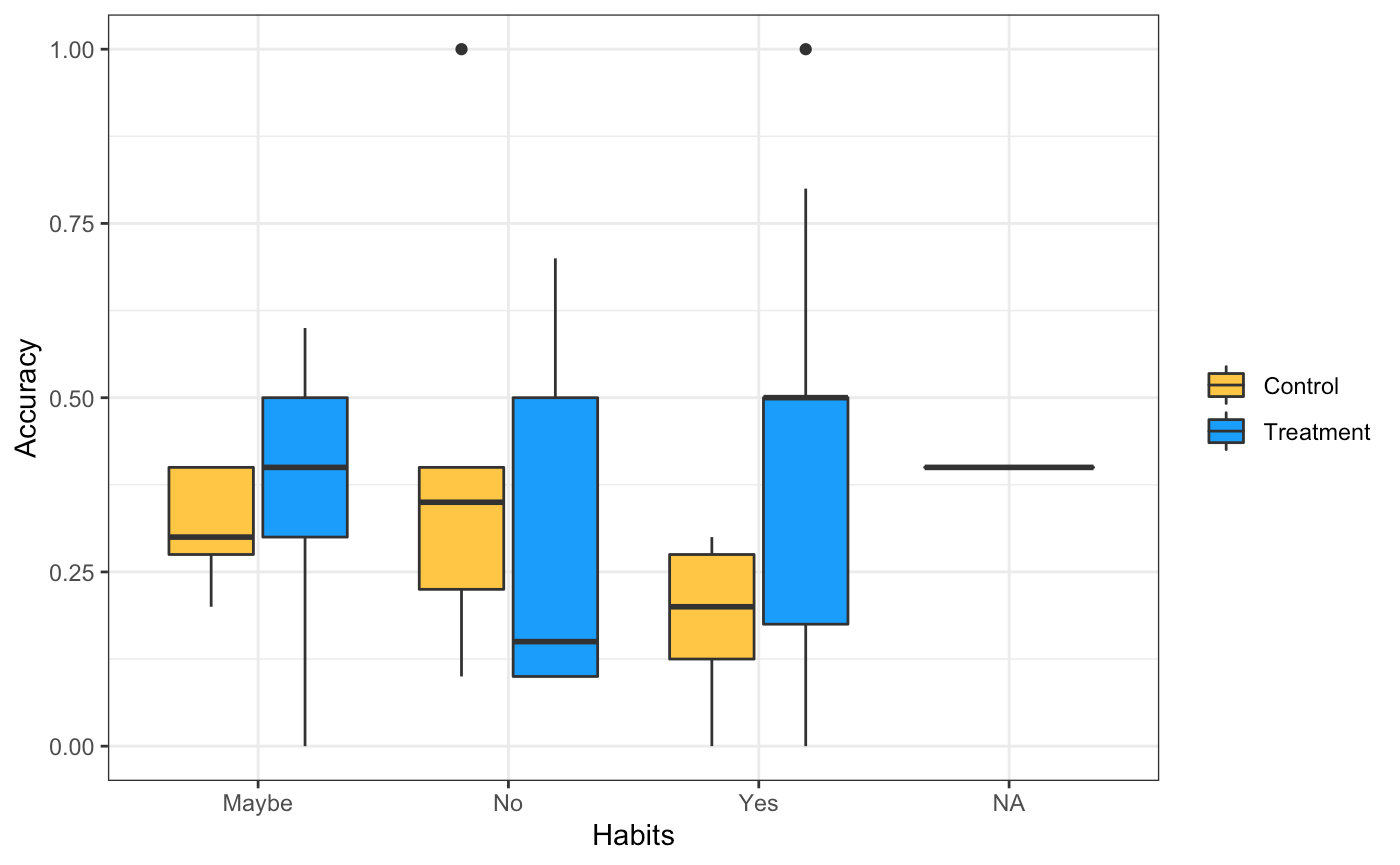
**Data Collected**

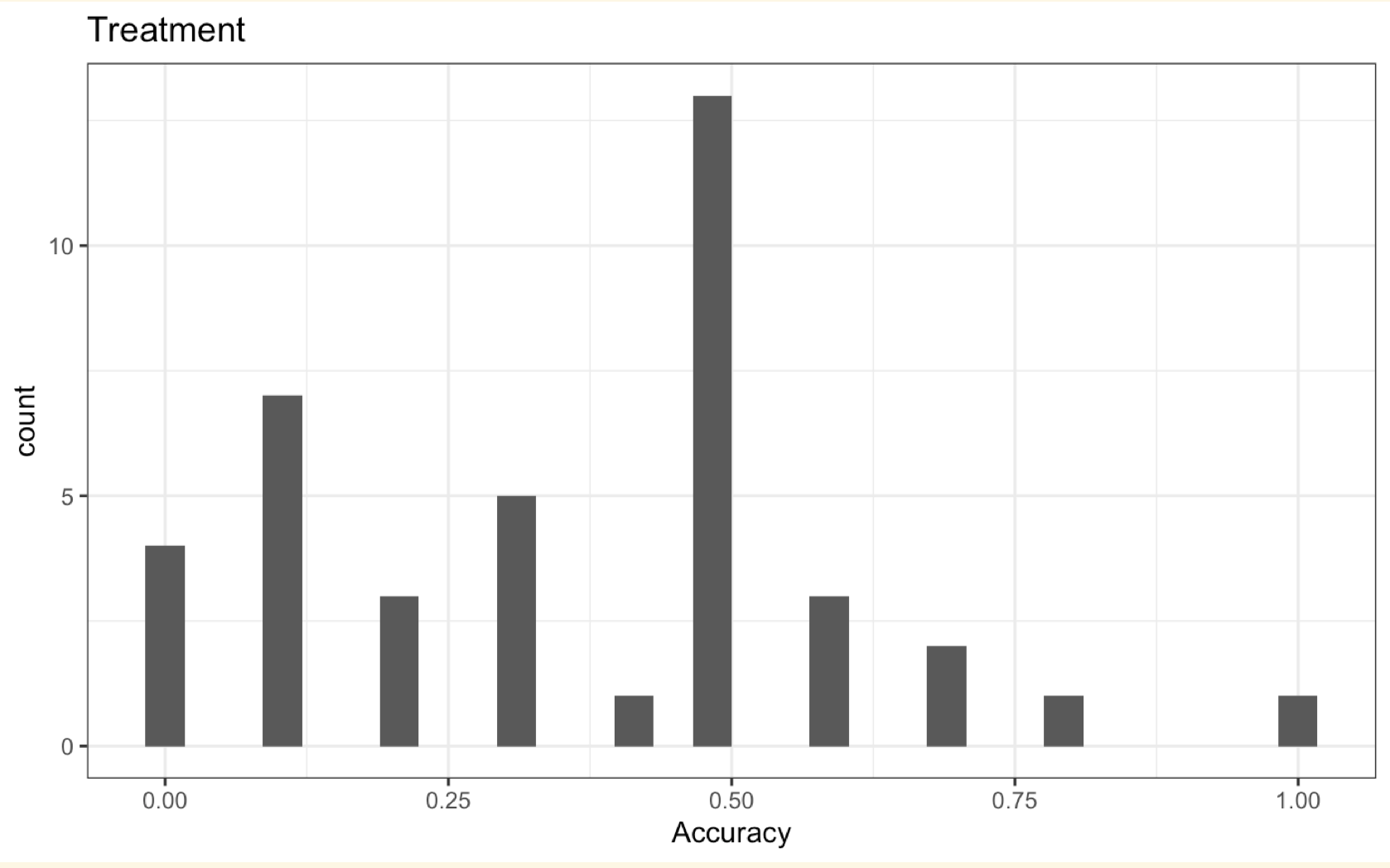
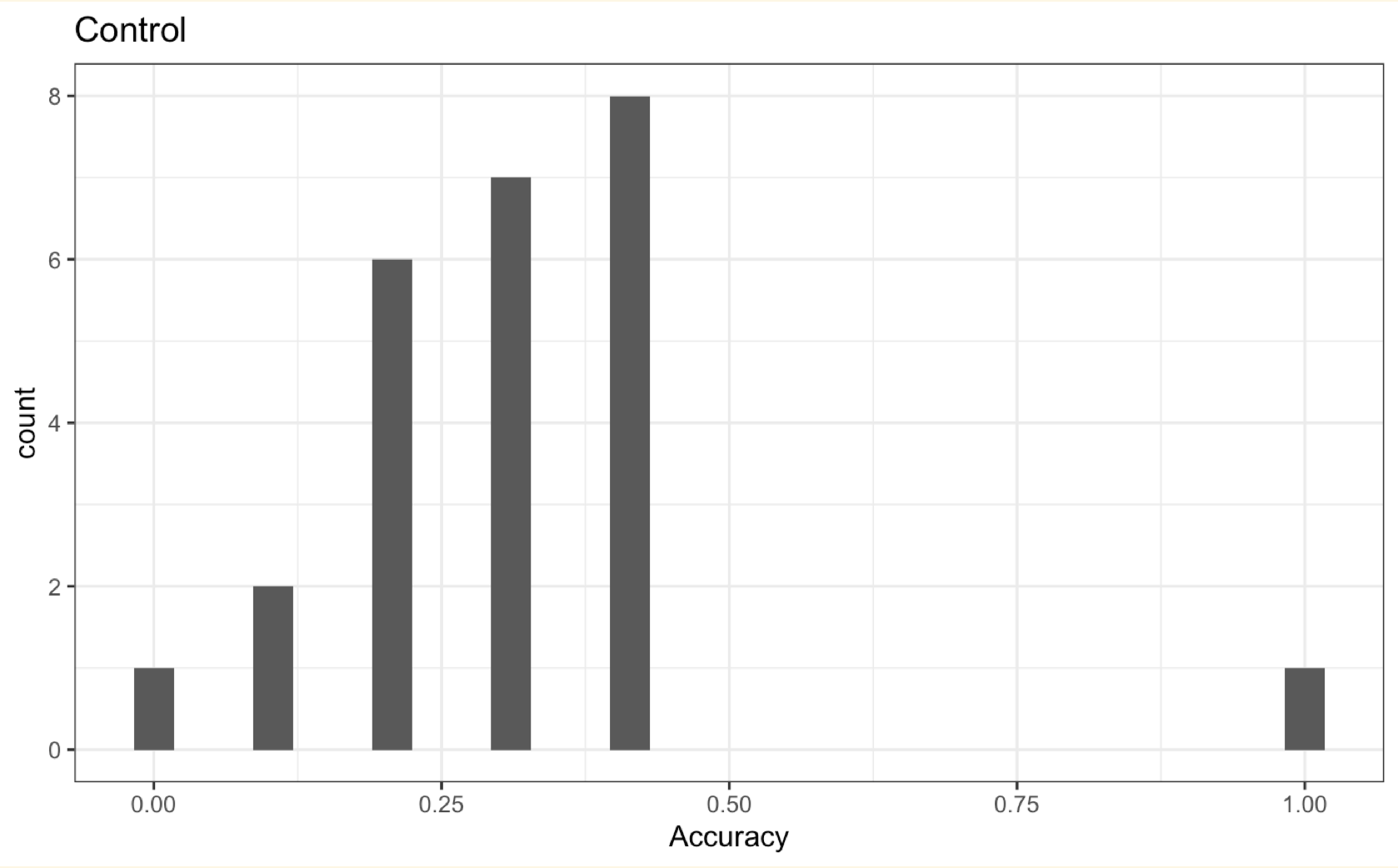
From December 2 to December 4, our team gave out 208 cards in total (104 treatment and 104 control). Until December 9, we collected 101 responses (55 treatment and 46 control). After filtering out those unfinished responses, our sample size was 65, composed of 40 treatments and 25 controls. Since the probability of assigning a treatment or control is the same, the difference between group sizes does not influence our regression analysis.

Our team had a thought on non-compliance. As mentioned, both treatment and control groups had some incomplete responses: some participants just scanned the code and never moved on. Initially, we tended to define them as non-complier, since they scanned the code, but then we had errors in regression computing CACE. Later we realized that the true treatment in our experiment is clicking the music on the task page from treatment survey, not receiving the code nor scanned. It might happen that some participants clicked the music but left the answer boxes blank; in this case, they are non-compliers. To check that, we examined the raw data. All incomplete responses stayed at the instruction pages, and links remained unclicked. We then considered all responses are compliant.

**Summary Statistics**

The raw data from Qualtrics had a lot of redundant columns. We deleted those and reformatted the spreadsheets by adding the correct answers to calculate the accuracy rate as our dependent variable. Before running regressions, we first created a boxplot to see the relation between participants study habit and accuracy:

In the control group, the median accuracy rate of participants who listen to music while studying is lower than those who do not listen to music when learning. In contrast, the rate of participants who have the habit in the treatment group is higher in median than those who do not. We guessed one possible explanation: those with the habit performed better in treatment, while those who do not have the habit performed better in control.

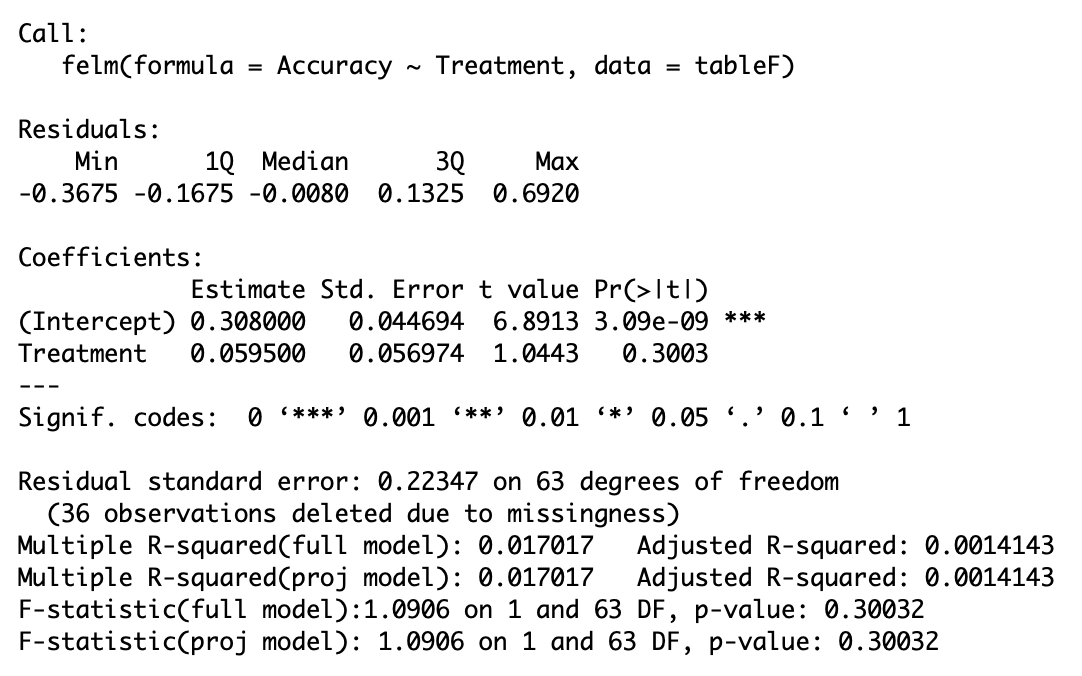


We also plotted the distribution of accuracy in both groups. The accuracy rate of most people (almost 15 people) in the treatment group is almost 50%, so we can speculate that it is normal for most people to remember half of the numbers. However, one person got the accuracy of 100%. We can guess that music can promote instant memory. Nevertheless, it does not rule out the possibility of cheating. About four people have a correct rate of 0. We checked the dataset and found that they did not answer any questions.

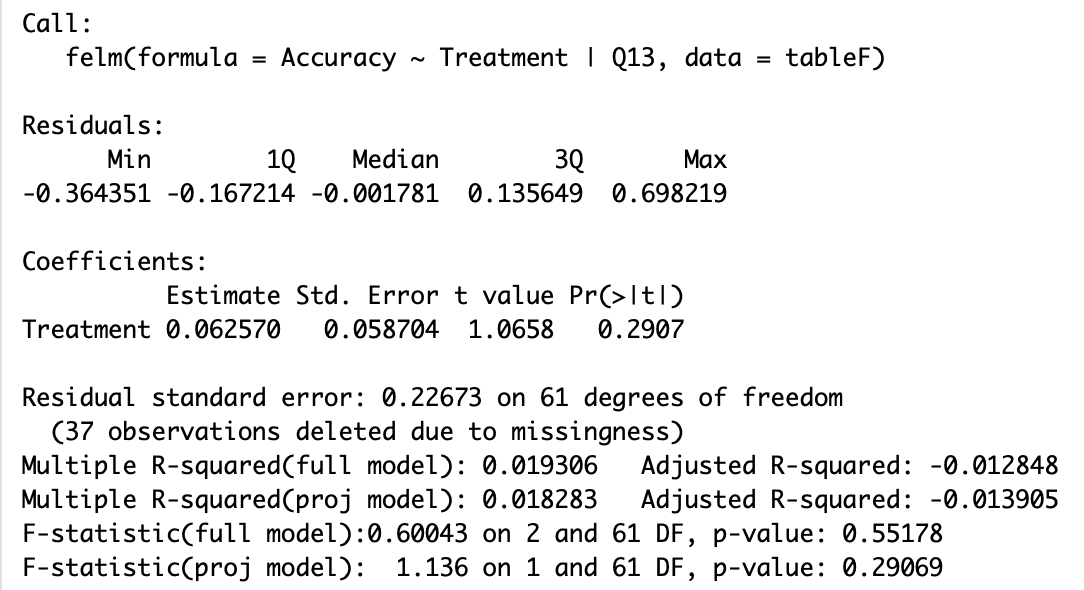
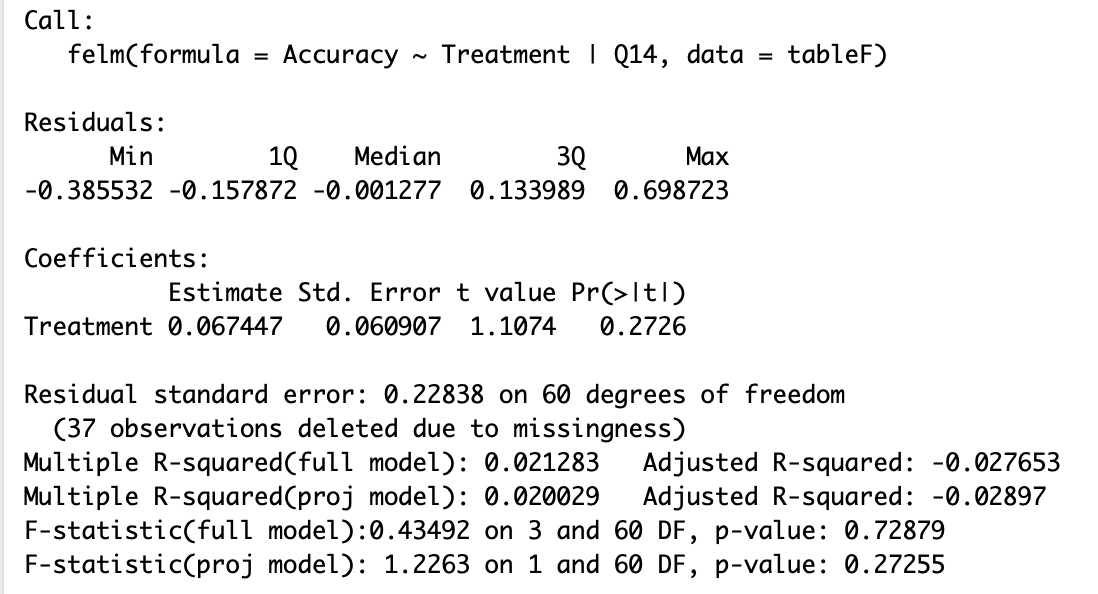
For the control group, the accuracy rate of most people (8 people) in the control group is almost 40%, followed by 30% and 20%. Compared with the treatment group, we found that the correct rate of most people in the treatment group was higher (10%) than that in the control group. We initially believe that listening to music can promote memory to some extent. There is one person who memorized all the numbers in 30 seconds. We treated this person as the outliner.

**Treatment Effect**

Our null hypothesis states that listening to music has no causal effects on people's instant memory. We first built a felm model between accuracy and treatment. From the summary, the *p*-value of the treatment is larger than any reasonable level of statistical significance, so it is insignificant and indicates weak evidence against the null hypothesis.



We then considered two variables, the participants are either undergraduate or graduate and their habits of listening to music when doing homework, to be fixed effects, and reran felm function to check if they can increase the precision of the estimate.



As a result, the standard errors of both models are not reduced but increased. Thus, both variables we chose are not tested as fixed effects for the model. Therefore, we failed to reject the null hypothesis and concluded that listening to music has no effects on instant memory.

**Limitations**

We sent out the survey by distributing QR codes in which way we expect participants can follow the instructions to complete the task, especially those highlights. The team would have managed these critical factors in the experiment if we could experiment in person with participants, though in reality, we barely “lost” control over these. This is the primary source of limitations and drawbacks of our design.

First, we could not prevent possible “cheating” during the experiment process, though it is unlikely. To get a 100% accuracy is pretty easy by writing down the sequence in the first 30 seconds and input them in the boxes. Spillover may also occur when people in the same group did the experiment together. Although we gave out the codes separately, participants could do it in groups with friends who received the card in the same location. Since we only have two sequences, people could have copied others’ answers. In the case of using earbuds, if people in treatment did the experiment in the same table with control without earbuds, control would hear the music and receive the treatment, thus influencing the control result.

The time to memorize may be slightly different among participants. We intentionally left 2 extra seconds for them to load the media and play; however, if someone did not read the instructions carefully or had trouble loading the music (for example, internet connection), they might have less time to memorize, and their performance might be influenced. A related media issue is the 32-second silence in the control task. Though we have informed that in the instruction, participants in the control group may still be confused about the silence after clicking or even distracted. As a result, they might have less time available to memorize.

In the ideal case where the team ran this experiment in person, the types of earbuds should also be controlled, and both groups need to use the same type to listen. In this practice, participant’s earbuds varied in brand and function in noise reduction, and this variation may bias the result. Moreover, the environment and volume when completing the piece, which were not controlled this time, could bring some noise to the result.

In terms of statistical power, our sample only has 65 observations, which is not large enough to capture the causal effect of the treatment. The regression indicates insignificance; even if we got a significant p-value, we would still need to be cautious to conclude the existence of the causal effect.

**Conclusion**

To test the causal effect of listening to music on instant memory, we ran an experiment on memorizing a number sequence with treatment. Although we observed some patterns in summary analysis, the regression result cannot infer any causality. Our experiment design could be further improved on some noticeable limitations.