Data project 2: Game developer A/B testing

Due: Monday March 7 2022 at 11:59PM (note the longer deadline so that we've covered two-sample tests).

You may use either R or Python to complete this project. Your submission should consist of a single pdf file, and your figures should be created using R or Python. I recommend that you create your submission using RMarkdown, since this is an easy way to embed figures into a pdf document. A template has been provided for you in the file data_project2_template.Rmd. Try to avoid having raw code *output* show in your pdf file. If printing a table, for example, format it using `kable()` or `gt()`.

You need to show your code for problems 2 and 3 (and you can show your code for problem 1 if you want to, but you don't need to). Your pdf submission containing both your answers and code (where relevant) should be submitted to "Data project 2" on gradescope no later than 11:59PM on the due date. You do not need to make a separate code submission for this project.

Introduction

For this project, you will be working as a data scientist working for a popular mobile game app. Currently, players face a challenge once they get to level 10, but the app developers are considering moving the challenge to take place instead at level 15. They thus created two slightly different versions of the game:

1. Version A: original version of the game which has a challenge at level 10
2. Version B: new version of the game which has a challenge at level 15

Your task is to decide whether or not to make the switch from version A to version B based on playtime and player retention. To answer this question, you run the following A/B test: everyone that installed the app over the one-week period during which you were running your experiment was randomly placed in either group A (the version they installed was the original version of the game, version A), or group B (the version they installed was the new version of the game, version B).

For each user, you collected data (over the 14-day period following installation) on the number of game rounds played, whether they used the app one day after installation, and whether they used the app 7 days after installation. This data is contained within `game_data.csv`.

In this folder, you will find the following files:

1. `game_data.csv`: the dataset containing the player information collected over the 14-day period after installation of the game. The columns include:

- `id`: a unique identifier for each player
- `group`: whether the player got version A or version B
- `games_played`: the number of games played by the player over the 14-days following installation
- `retained_1`: whether the player played the game 1 day after installation
- `retained_7`: whether the player played the game 7 days after installation

2. `data_project2_template.Rmd`: a template that you can (but are not required to) use for completing this project

Problem 1: Exploring the data

a. [2 points] Produce a data visualization that displays the distribution of games played (ignoring the `group` variable). If the original version of your plot is not very informative, try to modify it in some way such that you are actually conveying some *interesting* information.

b. [2 points] Produce a data visualization that compares the number of games played across each group (A/B). Based on your plot, would you recommend one version of the game over the other?

c. [1 point] Compute the proportion of players separately in groups A and B that continued to play the game **1 day** after installation. Compute the proportion of players separately in groups A and B that continued to play the game **7 days** after installation.

Problem 2: Bootstrapped estimates

You need to **show your code** for each of the following questions.

a. [2 points] Compute a classical (theoretical) 95% confidence interval for the average number of games played in the first 14 days (ignoring group). Show your code.

b. [3 points] Generate 1,000 non-parametric bootstrapped estimates of the average number of games played in the first 14 days (ignoring group). Present a histogram of your bootstrap estimates and compute a 95% bootstrapped confidence interval for the average number of games played in the first days. How does your bootstrap confidence interval compare with the classical confidence interval from the previous question? Show your code.

c. [2 points] Generate 1,000 non-parametric bootstrapped estimates of the proportion, $\hat{p}_{day1}$, of retention at day 1 separately for groups A and B. Present a visualization that compares the distribution of this estimate for each group. Based on your plot, which version of the game would you recommend? Show your code.

d. [2 points] Generate 1,000 non-parametric bootstrapped estimates of the proportion, $\hat{p}_{day7}$, of retention at day 7 separately for groups A and B. Present a visualization that compares the distribution of this estimate for each group. Based on your plot, which version of the game would you recommend? Does your recommendation agree with your conclusion in part c of this question? If so, do you feel that these findings are more or less compelling than the 1-day findings (justify your answer). Show your code.

Problem 3: Hypothesis tests

For each of the following problems: **Compute your hypothesis test manually** (i.e. by writing code in which you define the sample mean, and sample standard deviation and using the formulas from class) and **check your results using an inbuilt test**. **Show your working**: write down your null and alternative hypotheses, the format of your test statistic, the distribution of your test statistic, and the format of your p-value. **Show your code**

a. [2 points] Your coworker claims that the average player plays 52 games in the 14 days after they install the app (that's almost 4 games per day). Your boss disagrees and claims that the number of games played is lower than that. Conduct a hypothesis test to your coworker's claim against your boss's claim.

b. [1 points] Write down a response to your coworker from the previous question, informing them how they should interpret both the average game play data and the result of the hypothesis test. Hint: recall the distribution of number of players and do some critical thinking.

c. [2 points] Conduct a hypothesis test that tests whether there is a difference in the number of games played in the first 14 days across groups A and B. Explicitly state which test you are conducting. Justify whether or not you should use a pooled variance estimator. Does your answer agree with your conclusion from the Figure you produced in Problem 1.2?

d. [2 points] Conduct a hypothesis test that tests whether there is a difference in the proportion of players that are retained after 1 day across groups A and B.

e. [2 points] Conduct a hypothesis test that tests whether there is a difference in the proportion of players that are retained after 7 days across groups A and B.

f. [1 point] Based on all of the analyses you have conducted, would you recommend sticking with version A of the game where the challenge remains at level 10, or switching to version B where the challenge is moved to level 15?