



Lab 3 - Analysing Data From Field Experiments

Digital and Social Media Strategies

Fall 2024

Assignment Information

This formative assignment is designed to help you build and revise your knowledge of how to form (answerable) strategic questions from an experiment and analyse data from field experiments to answer them. This material was discussed in Lecture 3.

This question document, the dataset for this assignment, any additional information about the data and an R script for you to write your code are available on Canvas as a zip file "lab-03-experiments.zip." Download this repository and unzip on your computer in a location where you are keeping files for this class.

We will provide solutions to the coding parts of the assignment via Canvas on Friday after the final Lab Section of the week has concluded.

You do not need to submit this assignment for grading.

Learning Goals

By the end of this assignment, you will be able to:

- Identify the business question that a field experiment is designed to answer
- Use descriptive and inferential statistics to analyse data from a field experiment
- Interpret statistical output from analysing field experiment data
- Use statistical output to write a summary detailing an experiment's design and summarising the findings to draw a managerially relevant conclusion

Case Study: Field Experiments with Cookie Cats

[Cookie Cats](#) is a hugely popular mobile puzzle game developed by [Tactile Entertainment](#). It's a classic "connect three"-style puzzle game where the player must connect tiles of the same color to clear the board and win the level.



As players progress through the levels of the game, they will occasionally encounter gates that force them to wait a non-trivial amount of time or make an in-app purchase to progress. In addition to driving in-app purchases, these gates serve the purpose of giving players an enforced break from playing the game, hopefully resulting in the player's enjoyment of the game being increased and prolonged.



The location of these gates is something the firm is uncertain about. Initially the first gate was placed at level 30, but the firm ran an A/B-test where they moved the first gate in Cookie Cats from level 30 to level 40. In particular, we will look at the impact on player retention and the amount of levels they complete. Before we get to that, a key step is understanding the data structure.

Data & Variables

The data we have is from players that installed the game while the A/B-test was running. Each player who installed the game were randomly assigned to one of two conditions:

- `gate_30`: where the gate was placed before playing level 30, the control condition
- `gate_40`: where the gate was placed before playing level 40, the treatment condition

The variables included in the data are:

- `userid` - a unique number that identifies each player.
- `version` - whether the player was put in the control group (`gate_30` - a gate at level 30) or the treatment with the moved gate (`gate_40` - a gate at level 40).
- `sum_gamerounds` - the number of game rounds played by the player during the first 14 days after install.
- `retention_1` - did the player come back and play **1 day** after installing?
 - labelled as True or False
- `retention_7` - did the player come back and play **7 days** after installing?
 - labelled as True or False

When a player installed the game, he or she was randomly assigned to either `gate_30` or `gate_40`

Part A: Identifying the Research Problem

Question 1.

What is the strategic question the creators of Cookie Cats are trying to answer with this A/B Test? (Max. 2 sentences)

Question 2.

What outcome variables might the creators be interested in including in the analysis? Explain your motivation for including each. (Max. 4 sentences)

Question 3.

Why might the addition of gates with waiting times to proceed lead to a player's enjoyment being increased? Is there any way this could lead to the opposite outcome? (Max. 4 sentences)

Part B: Data Descriptives & Data Visualisation

Open RStudio and use the script "field_experiment.R" to load the data set. Write the code that answers the questions into this script.

Question 4.

How many players downloaded CookieCats during the A/B test?

Question 5.

How many players downloaded the game but did not complete any level in the first 14 days? Drop these consumers from your data set.

Question 6.

Are there any players that played the game substantially more than others? If so, drop them from your data set you began to modify in Question 5.

Hint: what is the mean number of `sum_gamelevels`? Are there players completing more than 10,000 levels?

Hint 2: Use the dataset that solves Question 6 for the remainder of the assignment.

Question 7.

What proportion of players are retained 1 day after installing for each treatment? What about after 7 days? To answer this question, complete the table below by filling in the missing

numbers. When filling out the table, use percentages (75% rather than 0.75) and round to two decimal places (0.755361 becomes 75.36%)

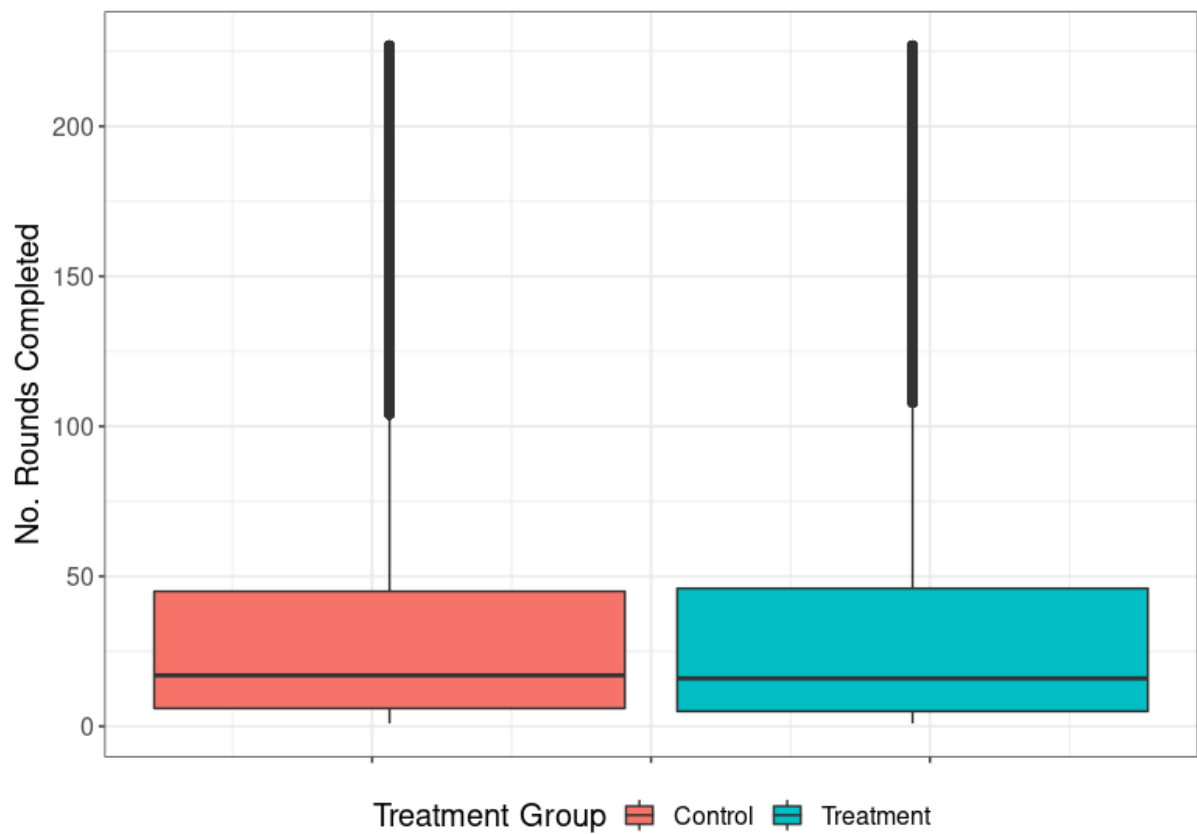
	One Day Retention			Seven Day Retention		
	Number Retained	Number Not Retained	Proportion Retained	Number Retained	Number Not Retained	Proportion Retained
gate_30						
gate_40						

Question 8.

What is the average number of rounds played for users in the gate_30 treatment? What about in the gate_40 treatment?

Question 9.

Create a box plot that compares the distribution of the total levels completed between the two treatments. Your plot should closely resemble the following, but not match it exactly:



Question 10.

Summarise the main observations you can make from Questions 6 through 8. (Max 3 sentences)

Part C: Statistical Analysis

Next, we turn to analysing the data to draw concrete conclusions via hypothesis tests.

Be sure to include the following information when testing a hypothesis:

- Null and Alternative hypotheses
- Level of significance
- Value of the test statistic
- p-value
- A decision whether to reject / fail to reject the null hypothesis

Question 11.

Is the one day retention rate different between the two treatments?

Hint: Use a two sample test of proportions to answer this question.

Question 12.

Is the seven day retention rate in the gate_30 treatment greater than the gate_40 treatment?

Hint: Use a two sample test of proportions to answer this question.

Question 13.

Is the average number of levels completed over 14 days different between the two treatments?

Hint: use a two sample test of independent means to answer this question.

Question 14.

Estimate the following [linear regression](#) model:

$$\log(\text{Sum Game Rounds})_i = \beta_0 + \beta_1 \text{Treatment Group}_i + \varepsilon_i$$

Test the hypothesis that $\hat{\beta}_1 \neq 0$. Interpret the regression coefficient and the output from the test using language that can be understood by a manager with little statistical knowledge.

Part D: Managerial Implications

Question 15.

Summarise the main takeaways from your analysis in Part C. (Max 4. sentences)

Question 16.

At which level should CookieCats place a gate? Explain why. (Max 3. sentences)