Lab 4 - Experiments with Online Advertising

Digital and Social Media Strategies

Fall 2024

Assignment Information

This formative assignment is designed to help you build and revise your knowledge of incrementality experiments in digital advertising. This material was discussed in Lecture 4.

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-04-adexp.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:

- Analyse data from field experiments and geo-experiments that investigate the effectiveness of online display ads and draw managerially relevant conclusions
- Explain the advantages and disadvantages of public service advertising as a control group
- Describe how a Ghost Ads experiment improves on the PSA framework

Case Study 1: Analysing Display Advertising Effectiveness

You have been provided with user level data from an anonymous company that contains information on a field experiment with online display advertising. The company ran an experiment that randomly allocated users who would usually see their display ad to one of two conditions:

1. The “ad” condition, where users see the company’s online display advertising
    - This is the *treatment* group
2. The “psa” condition, where users see a Public Service Advertisement (PSA) instead of the company’s ad.
    - This is the *control* group

The company’s goal from the experiment is to evaluate whether their online display ad is effective, as measured by conversions which are tracked at the user level. Specifically they want to know whether their own ad generates more conversions than the PSA ad.

Online Display Advertising:

Online display advertising is a form of online keyword advertising that promotes a brand message visually using text, logos, videos, photographs, or other graphics. The experiment was run via Google Display Network, which reaches 90% of Internet users worldwide, across millions of websites, news pages, blogs, and Google sites like Gmail and YouTube. Online Display ads can build brand awareness, drive clicks and traffic to your website, increase conversions and sales, and encourage repeat buyers. An example of such an advertisement is shown in Figure 1.

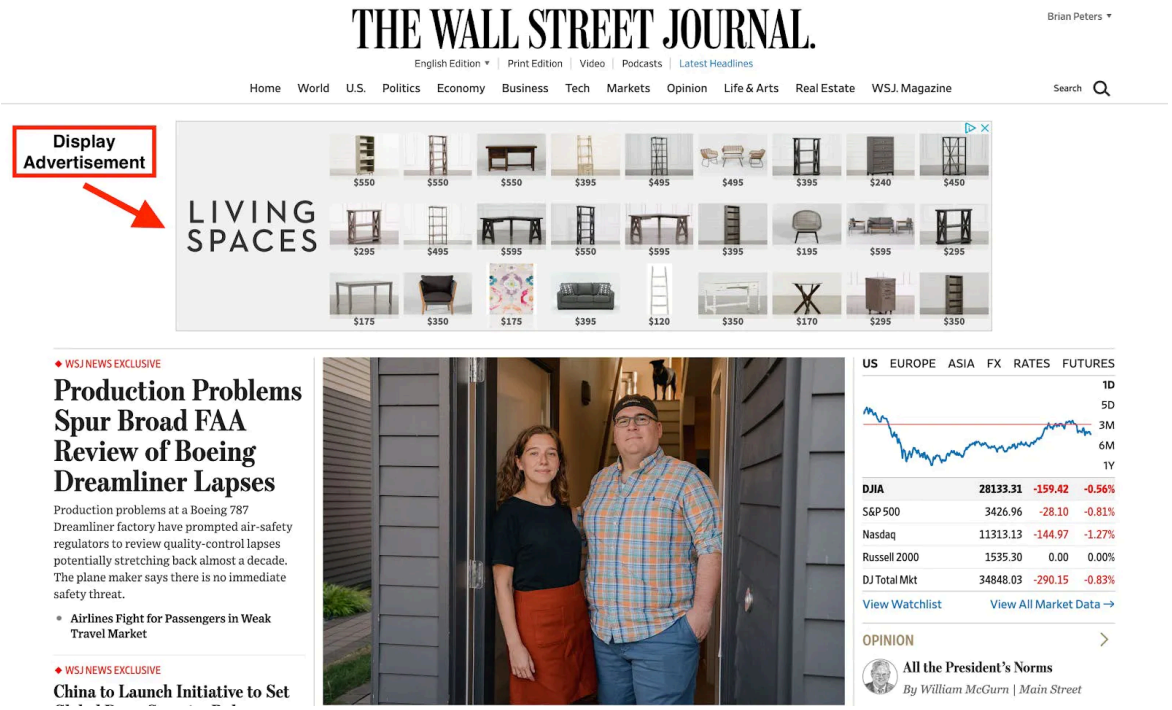

Figure 1: Example of Online Display Advertising

Public Service Ad (PSA) Experiments:

The public service ad methodology is a Randomised Control Trial where users are split into two groups: (i) a control condition, and (ii) a treatment condition. In the treatment condition, users see the company's ad as usual. However, in the control group users receive an alternative (though still "real") ad. The control group sees public service announcement (PSA) ads - ads that help raise social awareness like red cross banners or don't drink and drive ads. By serving these real, alternative ads, we obtain the information on which users within the control group would have been exposed to the company's ad if the experiment did not allocate them to see PSA ads. This process is illustrated in Figure 2.

Figure 2: The PSA Experiment

Data & Variables:

You have been given access to a dataset that contains information at the user level. The variables included in the data are:

- `user_id` - a unique identifier that identifies one device (which we assume is one user).
- `test_group` - specifies whether the user saw the company ad, "`ad`", or the public service advertisement, "`psa`"
- `converted` - did the user convert?
    - 0: No, 1: Yes
- `total_ads` - The total number of times a user saw the ad they were exposed to

Setup for Working Through the Exercises:

To work through the exercises you'll need to have done the following:

- Open RStudio
- Inside RStudio, create a new project using the folder "lab-04-adexp"
- Open the script "case_study_01.R"
- Installed any additional R packages necessary for this assignment

Exercise 1.
Load the dataset called “display_ads_student.csv” which is located in the “data/” directory.

Exercise 2.
Report how many unique users there are in the dataset. Is it equal to the number of rows in the data?

Exercise 3.
Report how many users are in each treatment group.

Exercise 4.
A. How many users saw the public service ads? How many users saw the standard display ad?
B. What percentage of users saw the public service ads? What percentage of users saw the standard display ad

Exercise 5.
How many users converted in each of the treatment conditions?

Exercise 6.
What percentage of users converted in each of the treatment conditions?

Exercise 7.
Conduct a t-test to evaluate whether users who saw the standard display ad have a higher conversion rate than those that saw the public service ad.

HINTS:
- Write down the null and alternative hypotheses to help you use the correct code
- Use the 5% significance level to decide whether to reject the null hypothesis

Exercise 8.
Run the following linear regression:

$$converted_i = \beta_0 + \beta_1 1[Saw\ Display\ Ad]_i + \varepsilon_i$$

Where $1[Saw\ Display\ Ad]_i$ is a variable that takes the value 1 if a user saw the display ad and takes the value 0 if a user saw a Public Service Announcement Ad.

A. Report the estimated regression coefficients
B. Interpret the regression coefficient $\hat{\beta}_1$.
C. Is $\hat{\beta}_1$ statistically significant at the 5% significance level?

Exercise 9.
Use the answers from Exercises 7 and 8 to conclude whether display advertising is an effective form of advertising for this company.

Exercise 10.

A. Why might Public Service Advertising not be a good control condition for evaluating the effects of a display advertising campaign?
B. Can you propose a better control condition? Explain the methodology and why it improves on Public Service Advertising.

Case Study 2: Analysing Search Engine Ad Effectiveness

An anonymous company is trying to evaluate the effectiveness of their search engine marketing on sales revenue. The company adopts the same framework as E-Bay's experiments described in the paper authored by Nosko and Tadelis The company ran an experiment in the USA that stopped search engine marketing in some regions of the country (called Designated Market Areas, or DMAs for short) while it kept its search engine ads on in other regions.

Your task is to evaluate whether stopping search engine advertising led to a reduction in weekly sales revenue in regions where the ads were stopped compared to regions where the ads stayed turned on.

Search Engine Marketing:
Search engine marketing refers to the display of a company's advertisements in the paid or sponsored sections of the search engine result page. When a user inputs a query on a search engine, two main types of results are shown: the advertising results and the organic (or natural) search results. Any website could be listed as an organic result, whereas only sites who have active advertising accounts appear in the advertising results. Figure 3 shows examples of paid and organic search results for a given google search query

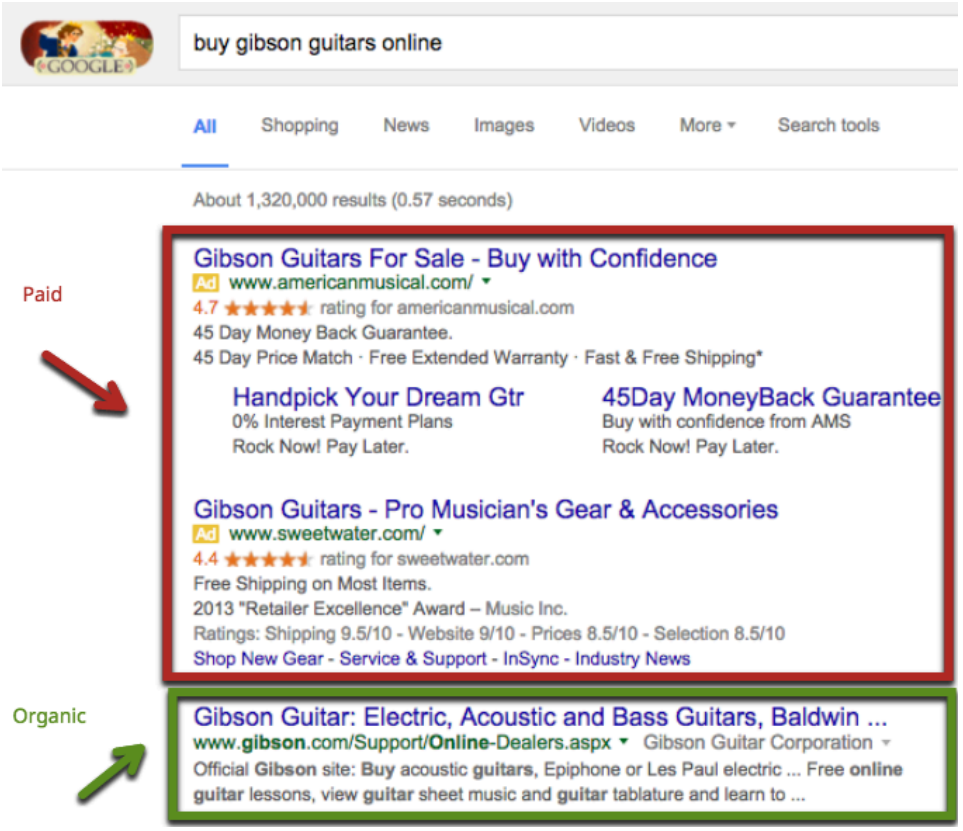

Figure 3: Search Engine Marketing

Data & Variables:

You have been given access to a dataset that contains information at the DMA-date level. This means that for each region you have access to daily revenue data and some additional variables. The dates included in the data include dates before and after the experiment has begun. This means it is possible to use differences in revenue both between DMAs in the "treatment vs control" conditions and "before vs after" the experiment begins in your analysis. The variables included in the data are:

- `date` - the date that the data is being recorded
- `dma` - specifies which DMA (i.e. which region) the data are coming from
- `treatment` - is the DMA part of the treatment group or control group?
    - 0: Control Group - DMAS where search stays turned on
    - 1: Treatment Group - DMAs where search will be turned off
- `treatment_period` - has the turning off of search started yet for the treatment group?
    - 0: No - search is on in both treatment and control
    - 1: Yes - search has been turned off in the treatment group (occurs on 2012-05-22)
- `revenue` - revenue for the DMA on a given date

Setup for Working Through the Exercise:

To work through the exercises you'll need to have done the following:

- Open RStudio
- Inside RStudio, use the RStudio Project associated with the folder "lab-4-adexp"[1]
- Open the script "case_study_02.R"
- Installed any additional R packages necessary for this assignment

Exercise 1.
What is search engine marketing? Explain why it could be an effective form of advertising for a company.

Exercise 2.
Provide a brief overview of the experiments and results in the Nosko and Tadelis paper.

Exercise 3.
Now, let's move on to analysing our data.

Load the dataset called "paid_search_student.csv" which is located in the "data/" directory.

Exercise 4.
Our analysis is going to use weekly revenue as the outcome variable of interest. Let's compute weekly revenue for each DMA. Before proceeding make sure you run the code that creates the weekly data.

[1] This assumes you have completed the exercises from Case Study 1. If you have not, refer to instructions there to create a new RStudio Project.

Before we perform our statistical analysis, we want to visualise how weekly revenue evolves over time in the treatment group (where search ads stay on) and control group (where search is turned off).

A. Compute the average weekly revenue for each week for the treatment and control groups.
B. Plot this data as a line chart that shows how average weekly revenue changes week to week across both groups. Include a line that marks where the experiment of turning search off for some DMAs begins.

Exercise 5.
When describing the experiment to a friend who is new to analysing data from experiments, they suggest the following:

"To estimate the effect of search ads in this context, you only need to focus on data after for the time <u>after</u> search ads have been turned off for some DMAs"

Let's use this advice in what follows:

A. Create a dataset of data on weekly revenue per DMA that only includes rows of data from when the experiment begins. We will use this data for the remainder of this exercise.
B. Use a t-test to test whether weekly revenue is higher in the DMAs where search is not turned off compared to their counterparts where search is turned off. Use the 5% significance level when making conclusions.
C. Run the following linear regression:

$$weekly\ revenue_{it} = \beta_0 + \beta_1 1[Search\ Turned\ Off]_i + \varepsilon_{it}$$

Where $1[Search\ Turned\ Off]_i$ is a variable that takes the value 1 if search advertising has been turned off in a DMA and takes the value 0 if search advertising stays on.

Interpret the regression coefficient $\hat{\beta}_1$, and conclude whether search engine advertising is effective.

D. From your analysis above, does it appear that search engine advertising is effective? Explain your answer.
E. Do your conclusions in (D) match up with the graphical evidence constructed in Exercise 5? Explain your answer.

Exercise 6.
Instead of following our friend's advice, we'll next try to utilise data from before and after the experiment begins for both treatment and control groups. The analysis you will perform is known as "Difference in DIfferences."

A. Our first step will be to compute the average weekly revenues for treatment and control groups in the weeks before and after the experiment begins. Fill in the values of average weekly revenue to complete the following table:

	Before Experiment Begins	**After Experiment Begins**
Control Group		
Treatment Group		

B. Compute the differences across the rows of the table constructed in (A), specifically subtract the average revenue before the experiment from average revenue after the experiment. This is the "first difference."
C. Subtract the difference from the control group from the difference in the treatment group. This is the "second difference".
D. Your answer to (C) is your "difference in differences" estimate. Can you conclude that the turning off search engine advertising off has a large effect on revenue?

Exercise 7.
The Difference in Differences estimate in Exercise 7 can be computed via linear regression. Run the following linear regression using the weekly revenue data:

$$weekly\ revenue_{it} = \beta_1 TreatmentGroup_i + \beta_2 After_t + \delta TreatmentGroup_i * After_t + \varepsilon_{it}$$

Where

- $TreatmentGroup_i$ takes the value 1 if a DMA is in treatment group (search turned off), and is zero otherwise
- $After_t$ takes the value 1 for weeks after the experiment begins and is zero otherwise
- $Treatment\ Group_i * After_t$ is the multiplication of the two variables defined above.

A. Estimate this linear regression in R.
B. Report the regression coefficients.
C. Link the Estimates back to the table you constructed in Exercise 7.
D. Interpret the regression coefficient which reports the difference in differences estimate of the effect of search advertising
E. At the 5% significance level, do search engine ads appear to increase revenue?

Exercise 8.
Looking back at the figure in Exercise 5, why might the DMAs that turned search engine advertising off have different weekly revenue than those that don't?

HINT: Think about the company's incentives and their belief about advertising effectiveness when deciding which markets they will keep advertising in.

Exercise 9.

A. Why don't the level differences in weekly revenue across treatment and control groups invalidate the experiment design?
B. What assumption(s) is needed for the difference in difference analysis to produce a correct estimate of the effectiveness of search engine advertising?

Exercise 10.

Re-estimate the linear regression in Exercise 8 using the log of weekly revenue rather than the level. That is, run the following regression:

$$log(weekly\ revenue)_{it} = \beta_1 TreatmentGroup_i + \beta_2 After_t + \delta Treatment\ Group_i * After_t + \varepsilon_{it}$$

A. Estimate this linear regression in R.
B. Report the regression coefficients.
C. Interpret the regression coefficient which reports the difference in differences estimate of the effect of search advertising

Exercise 11.

Write a short one paragraph summary that you can present to a manager that:

- Explains the experiment conducted and the strategic question it aims to answer
- Highlights the findings of the correct statistical analysis
- Provides implications for improving the firm's advertising strategy based on the results you find