Android Free Content and Features Experiment

Results Analysis

Cesaire Tobias

Last Updated: 28 March, 2023

Table of contents

Background

The Android Free content and features experiment was designed to test whether a period of unrestricted usage would be an effective means of activating new users. 'Features' refers to, remove background, ability to upload custom logo, fonts, colour palettes etc while 'Content' refers to, templates, images, videos, graphics, fonts, articles, shapes etc. Note that users where not required to opt-in to join the treatment groups. Three variants were used, control, treatment-7 (7 days free) and treatment-14 (14 days free). The following analysis provides a summary of findings with further details provided for the maximum n day retention rates observed in the experiment participants. For more details on the experiment, see the proposal here. For details, on the experiment outcomes for other metrics, see the Optimizely results here.

Experiment Details

The experiment allocated new users on Android to the treatment groups (treatment-7 and treatment-14) and control for 3 weeks between 09 January 2023 and 01 February 2023. New user allocation was then reduced to 0% while the experiment participants were tracked for a further 6 weeks to 17 March 2023, in order to assess retention rates. The analysis makes use of maximum n day retention.

sample size (n) = 184,290

- control: 61,466 (33.35%)
- treatment-7: 61,499 (33.37%)
- treatment-14: 61,325 (33.28%)

Table 1: Optimizely Results

(a) Treatment-7 vs Control

Metric	Conversions	Conversion Rate	Improvement	Stat-Sig	Conf int.
Project Exported (Avg)	166,065	2.7	49%	>99%	[46.47%;69.3%]
Project Exported (%)	23,064	37.5%	25.75%	>99%	[23.16%;28.55%]
D0 Second Project Exported (%)	7,445	12.11%	46.19%	>99%	[39.15%;52.34%]
Free Trial Upsell Converted (%)	105	0.17%	-69.93%	>99%	[-78.1%;-66.89%]
Application Opened After D0 (%)	19,389	31.53%	26.36%	>99%	[24.07%;29.93%]

(b) Treatment-14 vs Control

Metric	Conversions	Conversion Rate	Improvement	Stat-Sig	Conf int.
Project Exported (Avg)	188,244	3.07	69.38%	>99%	[55.46%;93.47%]
Project Exported (%)	23,376	38.12%	27.81%	>99%	[25.27%;31.02%]
D0 Second Project Exported (%)	7,601	12.39%	49.68%	>99%	[42.34%;55.83%]
Free Trial Upsell Converted (%)	99	0.16%	-71.57%	>99%	[-79.39%;-71.3%]
Application Opened After D0 (%)	19,445	31.71%	27.08%	>99%	[26.1%;30.68%]

Summary Results

In Table 1, `Conversions` for % metrics are unique conversions while for avg metrics they are total conversions. The `Conversion Rate` is therefore, total conversions per user for avg metrics and unique conversions per user for % metrics. `Improvement`, represents the change in conversion rate relative to the control. We see statistically significant improvements in our engagement and activation metrics while upsell conversions significantly underperformed.

Learnings

Activation

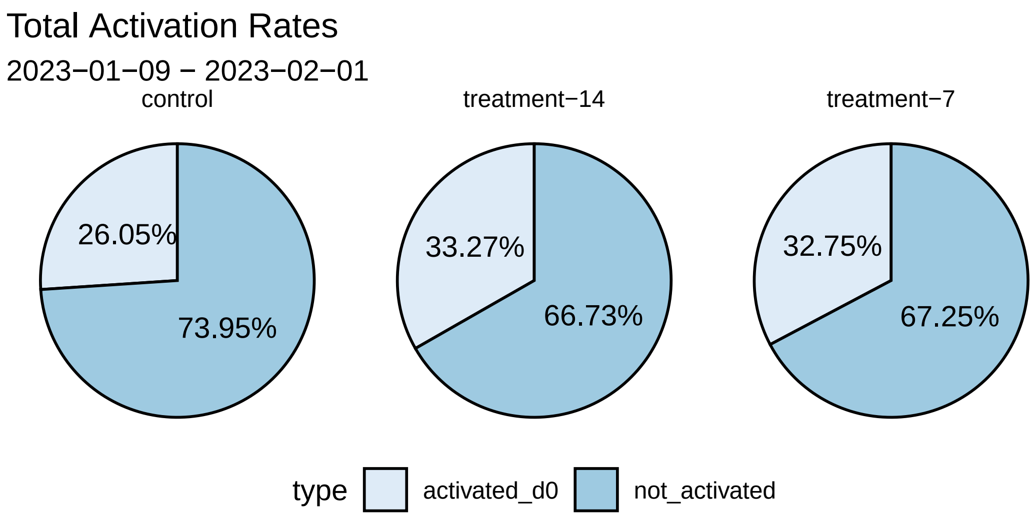
We use `d0 project exports` as our activation metric. For more on activation, please see the knowledge repo document, here.

Figure 1 shows the activation rates as measured by `d0 first project exported`. While we see higher activation rates of ~7ppt for the treatment groups relative to the control, there is little difference between treatment-7 and treatment-14.

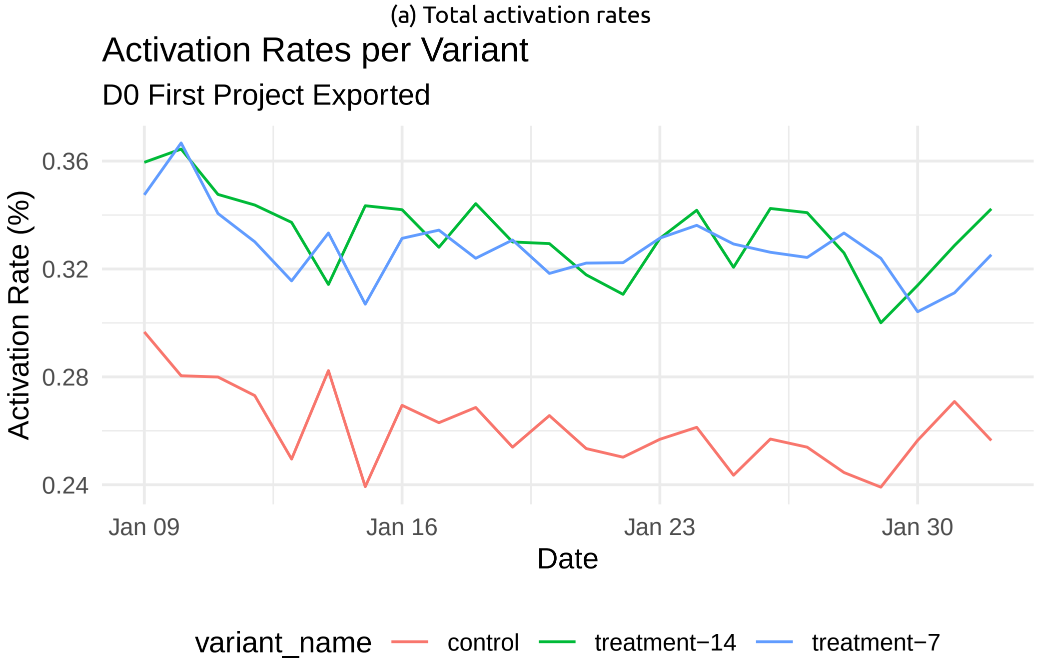
Engagement

Looking at user engagement, we see in Figure 2 that canvas active days is higher in the treatment groups up until ~10 days after new users stopped being allocated to the experiment. For a more detailed view on measuring engagement, see the User Engagement and Engagement Metrics knowledge repo documents.

As expected, we observe higher usage of the pro content elements.



(a) Total activation rates



(b) Activation rate time series

Figure 1: Activation Rates

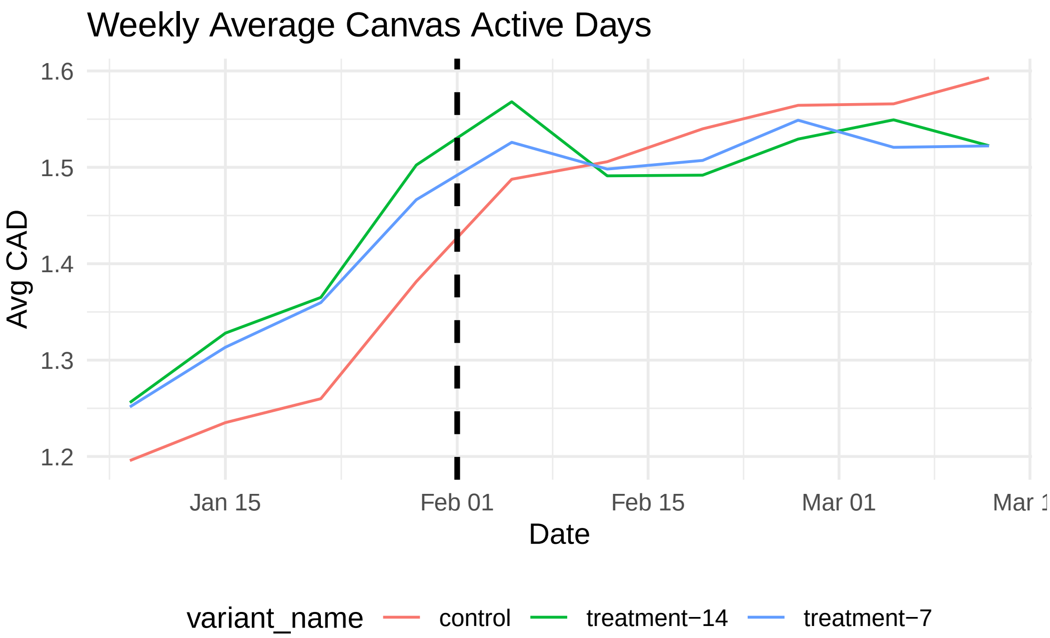


Figure 2: Weekly average Canvas Active Days

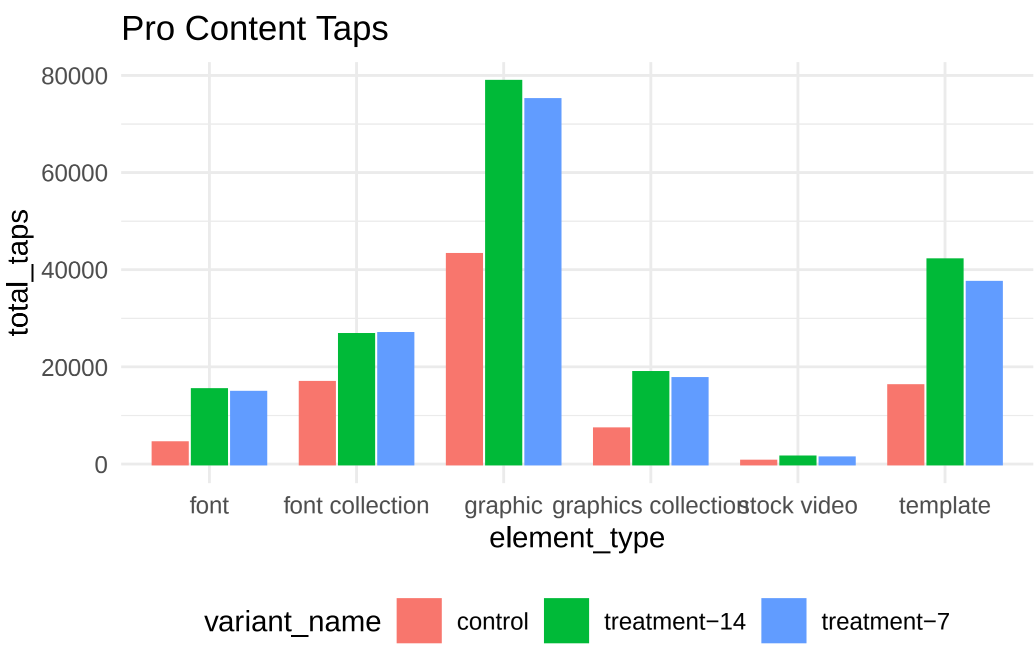


Figure 3 shows new user average project exports before and after starting the experiment in an attempt to measure the impact of the intervention on engagement. The analysis is based on the work presented in Pre-Post Experiment Analysis.

Revenue

An important consideration when giving away free access is to assess the revenue implications. `monthly recurring revenue` (MRR) is used as the metric. A similar methodology used in previous Black Friday experiments has been used.

Revenue Results

period	realised_subscribers	potential_subscribers	realised_mrr	potential_mrr	diff_subs_pct	diff_mrr_pct
monthly	205	245.6	$1,033.7$	$1,420.1$	-16.5	-27.2

annual	27	158.1	58.0	632.4	−82.9	−90.8

Retention Analysis

M1 retention is the primary business success metric in this experiment. To evaluate M1 unbounded retention[1] for new users on Android we refer to the techniques proposed in Experimentation: User Retention Testing. We will make use of visual inspection and bootstrap testing.

Visual Inspection

Any experiment with retention as a primary metric should use the `application opened after d0` metric. This allows us to use Optimizely's stats engine to test significance.

In the event that significance is detected, visual inspection of unbounded retention should follow in order to provide an indication of the nature of the retention, viz., quick convergence, potential convergence or uniform improvement. In this experiment we find that the treatment outperforms the control in percentage of users retained and that the relative difference remains fairly stable at an average difference of 19.89% (see Figure 5).

Figure 4 shows that the treatment groups outperform the control, with little difference between treatment-7 and treatment-14. Through visual inspection we can be confident that there is an increase in D0 retention (confirmed in Optimizely with the `application opened after d0` event).

Taking the visual inspection a step further, Figure 5 quantifies the differences between the curves for the control and treatment-7. We see a relatively stable, positive difference which would indicate improved retention as a result of the treatment. In the next section we make use of bootstrapping as a more rigorous quantitative approach to our assessment.

Bootstrapping

Bootstrapping is a procedure that re-samples a single dataset, in this instance, the sample collected over the first 3 weeks of the experiment. This single sample is treated as one of many random samples that the experiment could have collected. We construct a sampling distribution from the simulated datasets created by drawing repeatedly from our actual data (10^4 times in this analysis).

The process works as follows:

1. We assume that there is no difference between the control and the treatment. Our null hypothesis would be to say that they [control and treatment] are from the same distribution.
2. With this assumption, we can simulate the difference we'd see from random sub-samples of the control and treatment users and create a probability distribution from this.
3. We then compare this to the actual difference we observed in the experiment. If it's sufficiently unlikely that the observed difference falls within the sampling distribution, then we reject the null hypothesis.

Figure 6 illustrates the distribution of our collected sample for the control and treatment-7. We observe both an increase in the mean as well as a wider inter-quartile range in the treatment. The treatment mean is 2.42 percentage points higher than the control.

Table 3: T-test results

[1] Whether the user opened the app any time between day 30 and 59.

Two-sample T-test

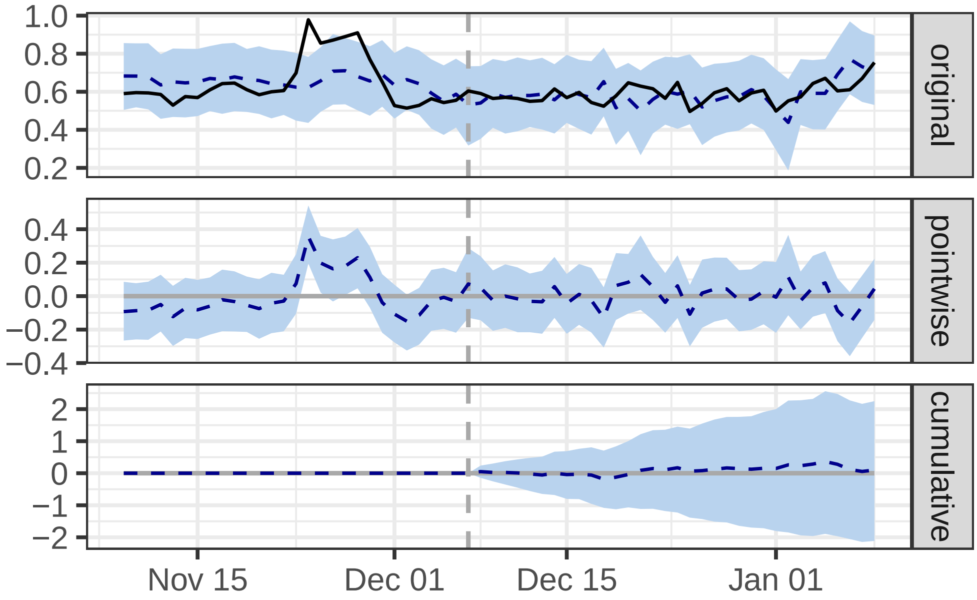
alpha: 90%

control	treatment-7	diff	n1	n2	test_stat	p	conf.low	conf.high
11.69	14.11	-2.42	67.00	67.00	-1.57	0.12	-4.99	0.14

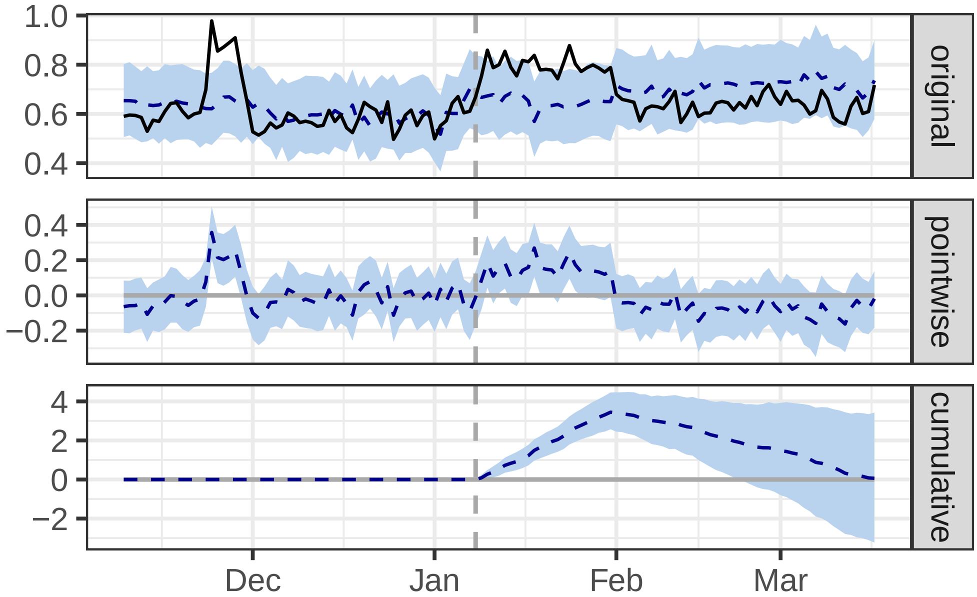
Figure 7 shows the observed difference in the means of the control and treatment-7 against the bootstrapped distribution. The Achieved Significance Level (ASL)[2] is at 0.94 and we see treatment-7 outperforming the control.

Next Steps

[2] Indicates the percentage of the distribution that is more than the observed value.



(a) AA Test



(b) Pre-Post Results

Figure 3: Causal Impact Analysis

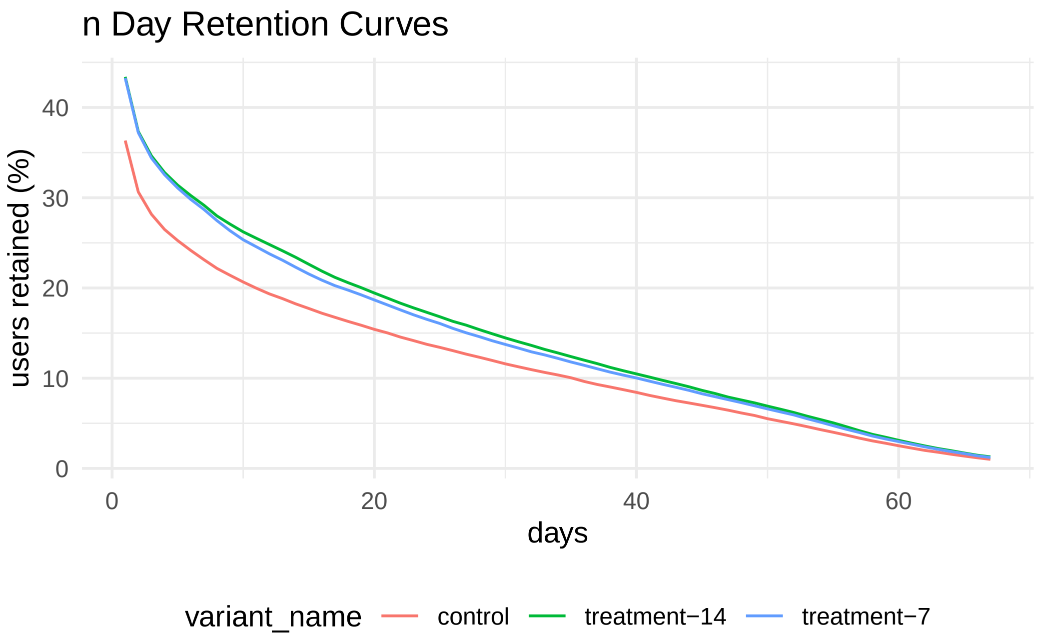
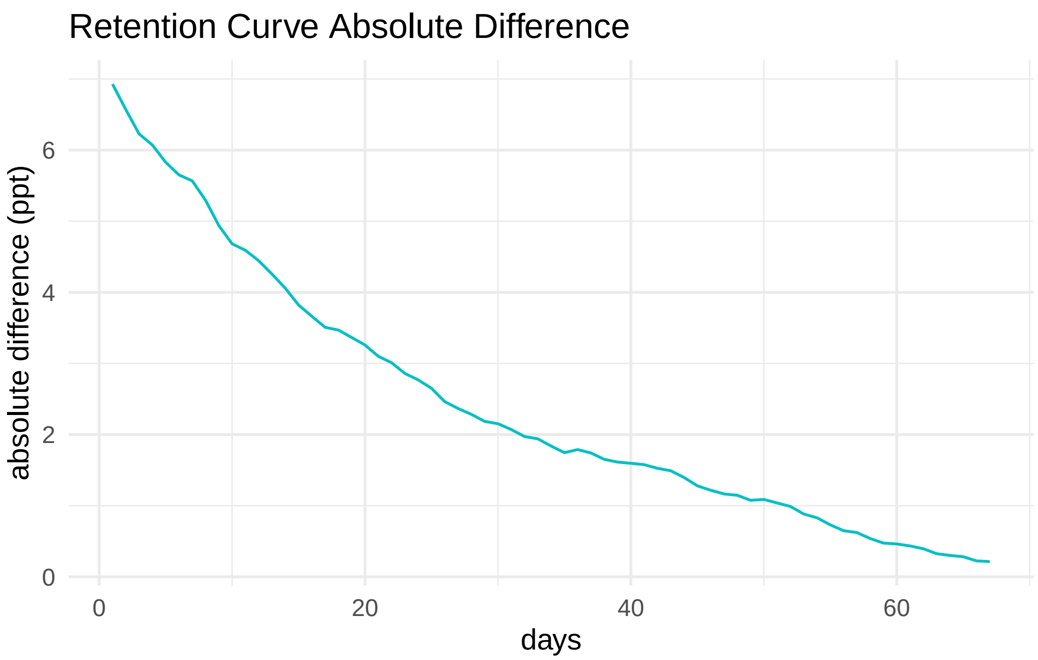
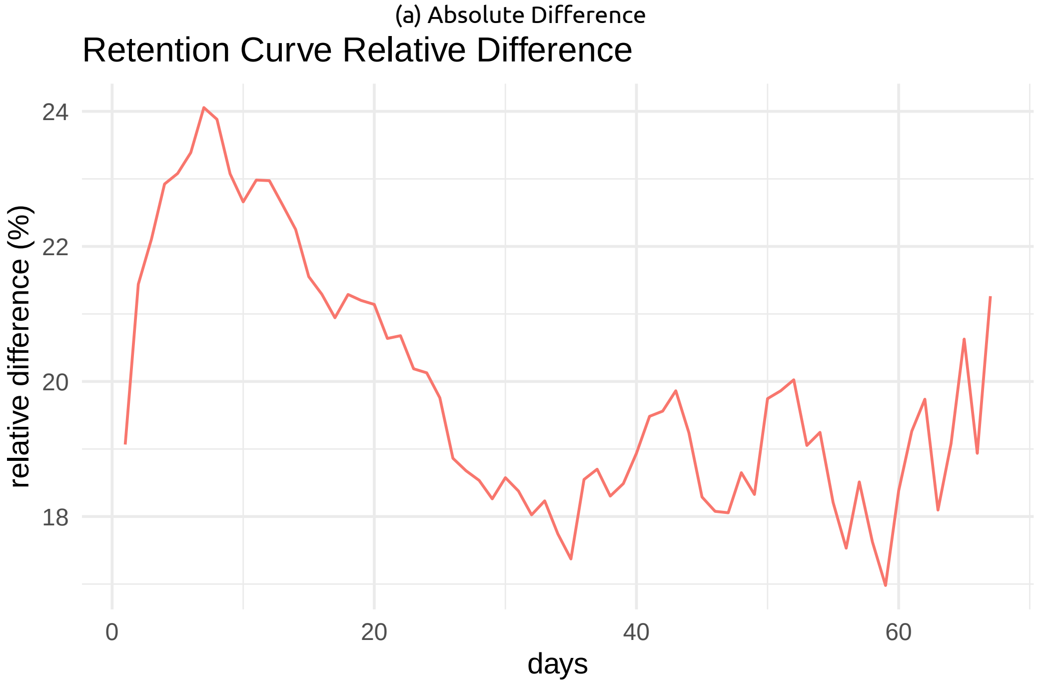


Figure 4: Retention Curves



(a) Absolute Difference



(b) Relative Difference

Figure 5: Differences in user retention over time

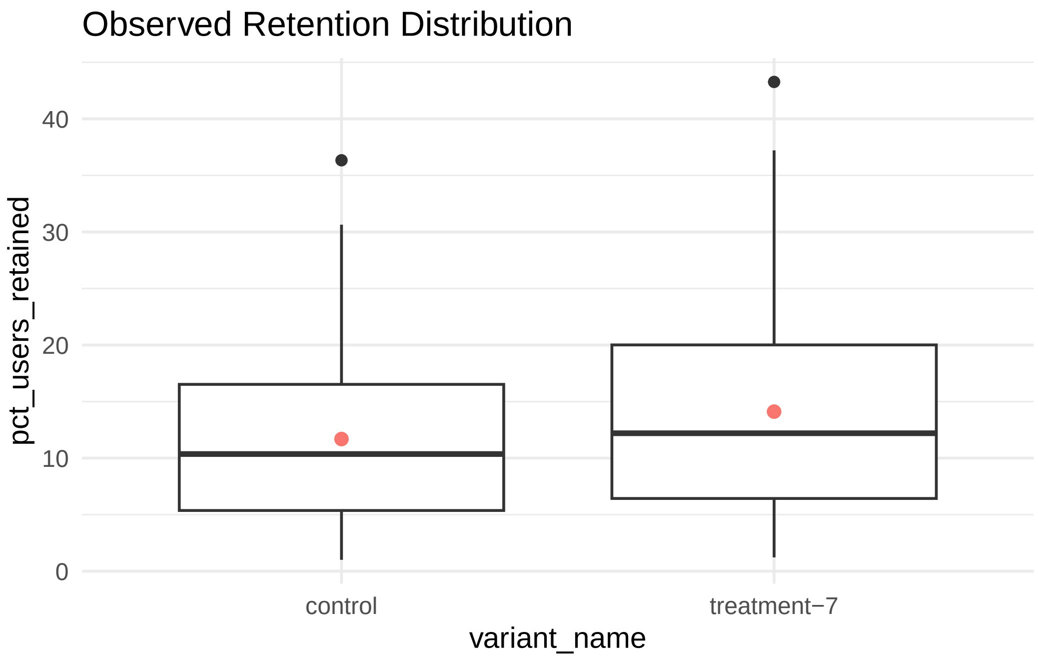


Figure 6: Observed Retention Distribution

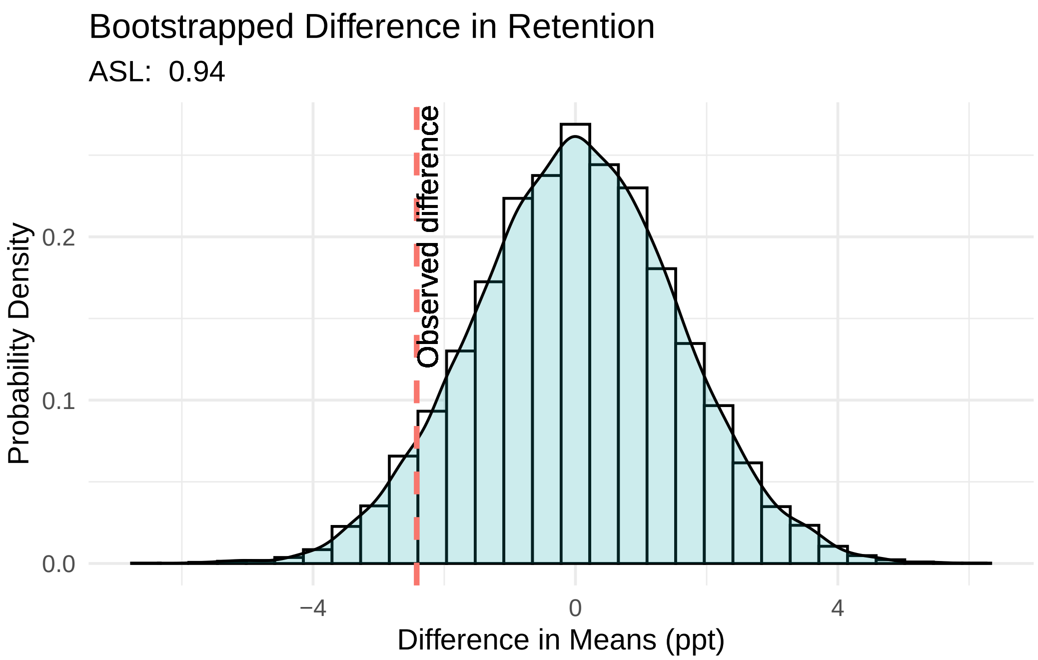


Figure 7: Bootstrapped Distribution