

GB740 Group Project Memo

Topic: A Mechanism Test for Instagram's Push Notifications

Group 1: Alec Peters, Sarah Link, Yuzhe Zhao

Target Organization, Business Objective, and Profit Model

Organization: Instagram is an online photo and video sharing social networking app with over 500 million daily active users.¹ People can upload photos or videos to their service and share them with their followers or with a select group of friends. They can also view, comment and like posts shared by their friends on Instagram.²

Profit Model: Instagram generates substantially all of its revenue from advertising by displaying ad products.³ Marketers pay for ad products either directly or through their relationships with advertising agencies or resellers, based on the number of impressions delivered (cost-per-impressions bidding model, CPM) or the number of actions, such as clicks, taken by users (cost-per-click bidding model, CPC).⁴

Business Objective: (sorted by priority, from the most important to less important)⁵

- Mission: Give people the power to build community and bring the world closer together.⁶
- For long-term development: Improve user experience and increase user engagement.
- For revenue: Increase the profit from ads delivered.

Brief Introduction of Experiment Design

Experiment Category: Mechanism test.

Experiment Purpose: Push notifications are an important tool to help Instagram improve user engagement and deliver more ads to users. The purpose of this experiment is to improve the efficiency of push notifications from the aspect of notification content. We want to know what push content has better performance in improving user engagement and ads revenue. We could assign the one that has better performance more weight in Instagram's push notification algorithm in the future.

¹ Aslam, S. (2020, October 28). Instagram by the Numbers: Stats, Demographics & Fun Facts. Retrieved November 27, 2020, from <https://www.omnicoreagency.com/instagram-statistics/>

² Instagram, Inc. (2020). Instagram Help Center. Retrieved November 27, 2020, from <https://help.instagram.com/424737657584573>

³ Facebook, Inc. (2020, January 30). Facebook, FORM 10-K, For the fiscal year ended December 31, 2019. Retrieved November 27, 2020, from <http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf>

⁴ How Much Does It Cost to Advertise on Instagram? (n.d.). Retrieved November 27, 2020, from <https://www.webfx.com/social-media/how-much-does-it-cost-to-advertise-on-instagram.html>

⁵ According to "We frequently make product and investment decisions that may not prioritize short-term financial results if we believe that the decisions are consistent with our mission and benefit the aggregate user experience and will thereby improve our financial performance over the long term.", Facebook, Inc. (2020, January 30). Facebook, FORM 10-K, For the fiscal year ended December 31, 2019. Retrieved November 27, 2020, from <http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf>

⁶ Facebook, Inc. (2020, January 30). Facebook, FORM 10-K, For the fiscal year ended December 31, 2019. Retrieved November 27, 2020, from <http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf>

Target Population: All Instagram users in the United States that are following over 20 people (include at least one person who has over 5k followers and at least one person who follows them back) with their notifications settings turned on.

Unit of Analysis: Individual

Data Acquisition: All data can be acquired from Instagram's API system.

Randomization Plan and Balance Check Sheet: We will select users in our target population based on their information in the API system, then randomly assign users into four groups. These users will be selected at the beginning of experiment. We will check means and distributions of users' age, gender, location, and interests to make sure the four groups are balanced.

Control/Treatment Groups:

For users in the control group, they will receive notifications directly linked to them, such as someone likes your post/someone comments on your post/someone follows you on Instagram.

For users in the treatment group 1-3, beside notifications that directly linked to them, they will receive 30% more additional special pushes per day.

Different treatment groups receive different special pushes,

- Treatment group 1: <Friend: people follow each other> just posted a photo/story/video.
- Treatment group 2: <Influencer: people with more than 5K followers> just posted a photo/story/video.
- Treatment group 3: <Friend A> and <Friend B> followed <Friend C> on Instagram. See their posts.

Outcome Metrics:

- Average amount of time that each user spends on Instagram (Primary Metric)
- Average revenue per user (ARPU)

We use a combination of user engagement and advertisement effects as the outcome metrics according to the business objective of Instagram that we've mentioned above. We choose the "average amount of time each user spent on Instagram" as the priority and choose the "average revenue per user (ARPU)" as the second priority for several reasons. First, sometimes an increase in the short-term advertisement gains is at the cost of the user engagements. Because the business objective of Instagram is putting the long-term development first, we choose the average amount of time as the priority. Second, we chose to make ARPU as a second metrics as there could be confounding variables from different advertisers' activity. We will have a business discussion after seeing the results.

The time horizon for measuring the effect would be one week. This time period is long enough to make sure that the change in average amount of time that each user spend on Instagram is due to the change of push contents.

Sample Size:

The sample size was calculated based on the following statistics and assumptions:

- The average time people spend on Instagram is 6.183 hours per week.⁷

⁷ According to data from Similar Web (<https://www.similarweb.com/>), the average time people spend on Instagram in 2019 is 53 minutes per day. If we assume people use Instagram every day, users will spend $53 \times 7 / 60 = 6.183$ hours on Instagram per week.

- There are 82.81% Instagram's users use Instagram more than once per week.⁸

The average amount of time people spend on Instagram per week were assumed fits in normal distribution. The mean of this distribution is 6.183. The standard deviation of this distribution is calculated based on the second assumption, which is 6.53. See Chart 1.

We want to be able to detect at least a 1% change on the average amount of time each user spent on Instagram at the significant level equals to 0.05 and the power equals to 0.8. Therefore, the sample size for each control/treatment group is 174,133. Given Instagram has more than 140 million users in the United States⁹ and the cost of running this experiment is relatively low, running this experiment with this sample size is feasible. See Chart 2.

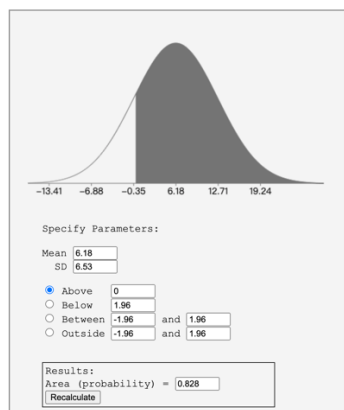


Chart 1: Assumption of Distribution

AREA UNDER THE NORMAL DISTRIBUTION

Instructions

1. Specify the mean and standard deviation.
2. Indicate whether you want to find the area above a certain value, below a certain value, between two values, or outside two values.
3. Indicate the value(s).
4. Hit tab, return, or the "recalculate button."

The area will be shaded and the size of the area will be shown at the bottom.

• ☒ Calculate Sample Size (for specified Power)

• ☐ Calculate Power (for specified Sample Size)

Enter a value for mu1:

Enter a value for mu2:

Enter a value for sigma:

• ☐ 1 Sided Test

• ☒ 2 Sided Test

Enter a value for alpha (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

Chart 2: Sample Size Calculation

How Results will be Used

We will first select treatment groups that have positive ATEs at a 5% significant level (which means the notification contents in those treatment groups have positive influence on Instagram), then compare these treatment groups with each other to see whether there is a significant difference among these groups and rank them based on outcome metrics. The result of this mechanism test could be used in lots of areas. For example, we will let the treatment group with higher average treatment effect have more weight in Instagram's push notification algorithm to help Instagram improve its push notifications system. When looking at results we could decide to look at heterogeneous treatment effects for different sub-groups based on number of followers and age. Having more followers would lead to more notifications overall and different age groups use Instagram in different ways. However, we won't know if we can break the results into sub-groups until we see results, because we don't know the range of the target population's characteristics.

Although sometimes the improvement in the average amount of time that users spend on Instagram might be small, we still would not ignore it and we may have a business discussion to decide whether to change in the future. This is because of two reasons: first, Instagram has a large amount of users, so even a small improvement in push notification efficiency would create a large amount of influence and revenue.

⁸ According to data from Statista (<https://www.statista.com/>), there are more than 82.81% of people who have used Instagram using Instagram more than once a week.

⁹ Clement, J. (2020, October 29). Instagram: Users by country. Retrieved November 28, 2020, from <https://www.statista.com/statistics/578364/countries-with-most-instagram-users/>

Second, compared to the revenue gained from the improvement of push notifications, the cost of improving the push notification algorithm is relatively low – so a small improvement would still bring profit to Instagram.

Limitations and Potential Failure Points

- a) The sample size may be not enough for several reasons:
 - The statistics and assumptions used to calculate the sample size are based on two assumptions that may not be true, which may lead to different sample size:
 - Users receive notifications generated from the current push notification algorithm, which contains both notifications that directly link to users and notifications with some other contents. But because the content in the control group only have content directly linked to users, the average amount of time users spend on Instagram in control group would be smaller than it in current push notification and we may need more sample size.
 - The distribution of average time users spend on Instagram per week fits in normal distribution. However, the real distribution may be different from the assumed distribution.
 - We not only compared the control group with treatment groups, but also compared treatment groups with each other. Our assumed sample size is large enough to detect at least a 1% change between control group and treatment groups, but it may be not large enough to detect the change among different treatment groups.
- b) There are some non-compliance problems cause by several reasons:
 - There may have some “false” and “duplicated” accounts in our sample that may not accept our treatments.
 - Some users in the sample may not look at the pushes before deleting and thus may not accept the treatment.
- c) There are some attrition problems:
 - Some users in the sample may turn off notification in the middle of experiment. Hence, we would not receive feedback from those users. This might lead to differential attrition as the treatment group might disproportionately turn off notifications since they are receiving more push notifications.
- d) We may have different treatment effects as we scale up these notifications to users in countries outside the United States due to the differences in language and culture.
- e) We also will have scaling up issues if the notification settings are applied to users who fall outside our target population criteria (no notifications enabled in settings or few followers). For those users the treatment effects might be different.