Udacity Data Analyst Nano Degree

A/B TESTING

Template Fromat

This template can be used to organize the answers to your final project. Items that should be copied from your answer to the quiz are shown in blue.

Experiment Design

Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here. (These should

be the same metrics you chose in the "Choosing Invariant Metrics" and "Choosing Evaluation

Metrics" quizzes.)

Choosing Invariant Metrics

Invariant Metrics：

1. **Number of cookies**: The number of unique cookies for viewing the course overview page.
2. **Number of clicks**: The number of unique cookies that click the "Start Free Trial" button (occurs before the free trial screen is violated)
3. **Click-through-probability**: The number of unique cookies on the "Start Free Trial" button divided by the number of unique cookies on the course overview page.

#Reason：First of all, the units of diversion in this experiment is Unique Cookie, so the number of cookies both in CONT group and EXP group should be the same or similar. That was because this experiment is prompted when you click the ‘Start Free Trial’ button, thus before the prompt button comes out, the Number of clicks and Click-Through-probability should be similar, In summary, these three metrics should remain unchanged.

Choosing Evaluation Metrics

1. **Gross conversion**: The ratio of the number of user IDs who have logged in and participated in the free trial divided by the number of unique cookies that clicked on the "Start Free Trial" button. (dmin =0.01)
2. **Net conversion**：The number of user ids who still participated (and paid at least once) after the end of the 14-day period divided by the number of unique cookies that clicked on the "Start Free Trial" button. (dmin=0.0075)

For each metric, explain both why you did or did not use it as an invariant metric and why you

did or did not use it as an evaluation metric. Also, state what results you will look for in your

evaluation metrics in order to launch the experiment.

#Reason: The purpose of this experiment is to reduce the number of students who leave free teaching because they don't have enough time, so as to increase the proportion of students who have completed trial learning after starting free teaching and stayed (referred to as paid). Therefore, the total conversion rate and net conversion rate can be used as good evaluation indicators. If the results are statistically significant, the total conversion rate of the experimental group should be lower than the control group, because some students in the experimental group opt out after seeing the prompt. But the net conversion rate examines the number of people who clicked to start the free trial and the number of people who completed the project. If the experimental hypothesis is correct, the net conversion rate should remain unchanged. At the same time, the project description believes that this experiment will not greatly reduce the number of students who continue to pass the free trial and finally complete the course (grouped by user id). It is certain that the number of user-ids will be reduced but the reduction will change a lot, may not have statistical significant. so the number of user-ids cannot be a good basis for evaluation. The retention rate is because it is found that too many page views are required in the subsequent calculation process, and the experiment time is too long, so it is not selected as an evaluation metric**.**

Measuring Standard Deviation

List the standard deviation of each of your evaluation metrics. (These should be the answers

from the "Calculating standard deviation" quiz.)

For each of your evaluation metrics, indicate whether you think the analytic estimate would be

comparable to the the empirical variability, or whether you expect them to be different (in which

case it might be worth doing an empirical estimate if there is time). Briefly give your reasoning in

each case.

#Reason: The data in excel is based on 40,000 unique cookies per day, but 5000 per day is given in the project, so it needs to be converted proportionally. Only 400 of the 5000 unique cookies clicked on Start free trial, so here N=400.



Standard Error

Formula: SE = (p\*(1-p)/N)\*\*0.5

1. Gross conversion

SE = (0.2063\*(1-0.2063)/400)^0.5 = **0.0202**

1. Net conversion

SE = (0.1093\*(1-0.1093)/400)^0.5 = **0.0156**

Sizing

Number of Samples vs. Power

It is given alpha = 0.05 and beta = 0.2, round the answer to the nearest integer.

Indicate whether you will use the Bonferroni correction during your analysis phase, and give the

number of pageviews you will need to power you experiment appropriately. (These should be

the answers from the "Calculating Number of Pageviews" quiz.)

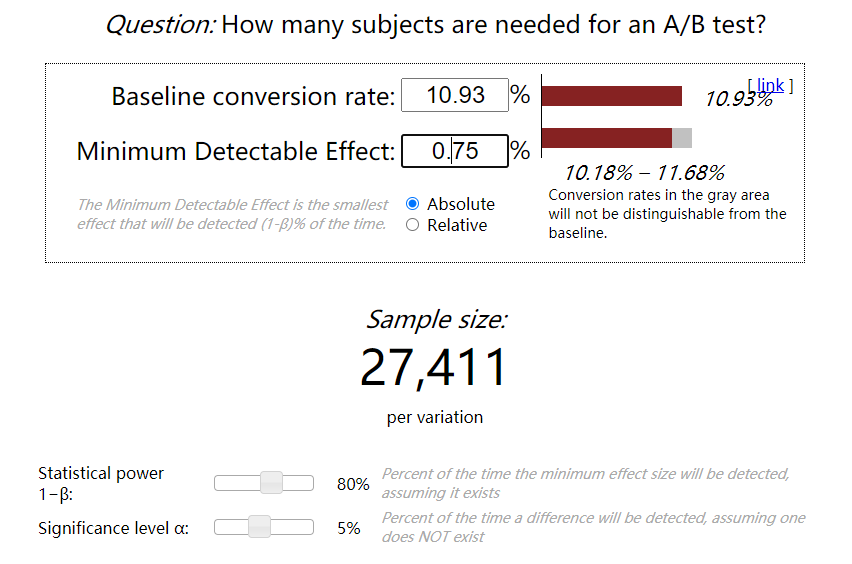
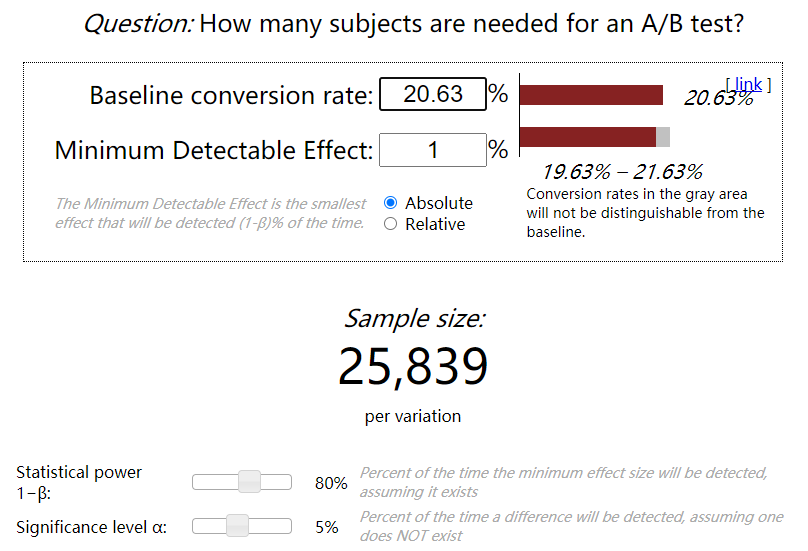
Sample size calculation using [Online sample calculator](https://www.evanmiller.org/ab-testing/sample-size.html)

1. Gross conversion

The base rate of Gross conversion is 0.2063, dmin=0.01, setα=0.05, 1-β=0.8, Thus we get the minimum unique pageviews is **25839。**

1. Net conversion

The base rate of Net conversion is 0.1093, dmin=0.0075, setα=0.05, 1-β=0.8, Thus we get the minimum unique pageviews is **27411。**



## Note: If you use retention rate as an evaluation indicator, 39511 unique pageviews are required, which is too large.

Duration vs. Exposure

Indicate what fraction of traffic you would divert to this experiment and, given this, how many

days you would need to run the experiment. (These should be the answers from the "Choosing

Duration and Exposure" quiz.)



The above calculation of the two evaluation metrics (evaluation metrics) to reach the dmin, the minimum amount of unique cookies required to click the Start free trial button (Start free trial), and then calculate the total amount of unique cookies required according to the above table (On this basis, the click probability is divided by 0.08 and then multiplied by 2. Here we need to consider that the test is divided into the control group and the test group). According to the larger ones, we think that a total of 685,275 unique cookies are needed.

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment

would be for Udacity?

#Reason：Here we need transfer all 40000 traffic per day for experiment, Days = 685275/40000 ≈ 18days

#Risk：

There is not much risk in this test: (from the analysis of the csdn, mainly starting from the four principles of abtesting: risk, benefit, alternatives, data sensitivity and ethical aspects)

1. The tracked data exists in the form of cookies, and subsequent logins and payments will not change due to events.
2. The website has not undergone major changes, and users do not need a long time to adapt to the new browsing page.
3. The database has not been changed, and the risk of data leakage has not changed
4. The change is only a page reminder, and it does not cause difficulties for users.
5. The test does not involve moral hazard.

Experiment Analysis

Sanity Checks

For each of your invariant metrics, give the 95% confidence interval for the value you expect to

observe, the actual observed value, and whether the metric passes your sanity check. (These

should be the answers from the "Sanity Checks" quiz.)

For any sanity check that did not pass, explain your best guess as to what went wrong based on

the day­by­day data. ​Do not proceed to the rest of the analysis unless all sanity checks

pass.

# There are three Invariant Metrics, the number of cookies, the number of clicks, and the click-through probability(CTP)



Sanity Checks：

The number of cookies and the number of clicks should conform to a random binomial distribution, which means that the expected probability should be 0.5. Here we calculate the confidence interval (lesson5.6) with a 95% confidence level, z\*=1.96.

Calculation Process: The calculation of pageviews and clicks are easy. The value of P\_expected is 0.5, then through the formula “SE = (P\*(1-P)/N)^0.5” to get the Standard Error, SE times the Z\*, we get the Margin. Then get the confidence interval centered on P\_expected, and finally check whether our observation probability P\_hat is within this interval.

CTP calculation, here we use the cont group as our expected probability, and use the exp group as our observation probability. CTP\_P\_expected=28378/345543=0.082125814, CTP\_P\_hat=28325/344660=0.082182441, the latter calculation method is consistent with the first two items.



#Conclusion: all pass the sanity check.

Result Analysis

Effect Size Tests

For each of your evaluation metrics, give a 95% confidence interval around the difference

between the experiment and control groups. Indicate whether each metric is statistically and

practically significant. (These should be the answers from the "Effect Size Tests" quiz.)

Evaluation Metrics:

1. Gross conversion
2. Net conversion



Here we sort out the changes in the Gross conversion rate and Net conversion rate of the control group and the experimental group, and then analyze the statistical significance and practical significance. Note: Statistical significance means that the confidence interval of dhat does not include 0, and practical significance means that the confidence interval of dhat does not include dmin.





#Reason：Through the analysis of the significance of the Gross conversion and the Net conversion, we found that the change of the total conversion rate is statistically significant and practically significant, while the change in the net conversion rate is not statistically significant and practically significant, which is in line with our previous assumptions. This measure will reduce the total conversion rate, but at the same time will not significantly affect the net conversion rate.

Sign Tests

For each of your evaluation metrics, do a sign test using the day­by­day data, and report the

p­value of the sign test and whether the result is statistically significant. (These should be the

answers from the "Sign Tests" quiz.)

Use this link to calculate the two-tailed probability: <http://graphpad.com/quickcalcs/binomial1.cfm>

# Note: If the two-tailed probability P<α, it is statistically significant; otherwise, it does not have。

1. Gross conversion
2. Net conversion



Through statistics, we found that within 23 days, the Gross conversion(GC) of the experimental group was less than that of the control group in 23 days, and the Net conversion(NC) of the experimental group was less than that of the control group in 13 days. The results are brought to Sign Check The website calculates. The results are as follows:



# Conclusion: This is consistent with the result of our effect size test! The Gross conversion is statistically significant, while the Net conversion is not statistically significant.

Summary

State whether you used the Bonferroni correction, and explain why or why not. If there are any

discrepancies between the effect size hypothesis tests and the sign tests, describe the

discrepancy and why you think it arose.

#Reason：Bonferroni correction is not used, because the two evaluation metrics are not independent of each other but closely related. If Bonferroni is used, it will be too conservative.

Recommendation

Make a recommendation and briefly describe your reasoning.

A: This experiment can be carried out. At present, the difference between the experimental group and the control group is in line with our assumptions before the experiment. Udacity will improve the overall student experience and enhance the tutors to provide help and support for students who can complete the course ability.

Yicheng Jin

2021.8