# 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.)

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.

假设是不会显着减少继续通过免费试用期并最终完成课程的学生人数。

不变指标：

Cookie数量是不变指标，这是由于udacity要增加的实验的界面是在学生点击开始免费试用后才出现的，所以对本身网页的访问量不会有改变。

点击次数是不变指标是因为，点击后才会出现实验的界面，所以不会影响点击次数，点击次数不会发生变化。

由于上面两个指标不会改变，点击率也不会改变。

评估指标：  
总转化率：完成结账并注册免费试用的用户id数量 / 点击按钮的cookie数量，由于添加的实验会向将每周投入学习时间少于5小时的用户推荐选择免费访问课程材料，所以会分流一些用户到免费访问课程材料，所以分子会变小。然而分母在本实验中不会改变所以期望总转化率会减小。

净转化率：14天期限后保持注册的用户id数 / 点击按钮的cookie数量。实验希望推荐投入学习时间大于5小时的用户参加免费试用，他们有更长的学习时间所以他们在14天后继续保持注册学习的概率会增加，所以分子会变大，净转化率可能也增加。

无关指标：

留存率：净转化率 / 总转化率，分母会减小，分子的数量在比例上应该会增加，但是分子的数量基于分母，所以留存率总体的改变很难比较。

点击次数：点击开始免费注册按钮的唯一cookie数量，不使用时因为总cookie可能变化导致点击按钮的cookie数量也变化，无法作为不变指标。同时也不是作为目标的评价指标。

## 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.

|  |  |  |
| --- | --- | --- |
| Unique cookies to view page per day cookie总数 | NC | 40000 |
| Unique cookies to click “Start free trial” per day 点击按钮的cookie数 | Ncb | 3200 |
| Enrollments per day 报名参加免费试用的用户数量 | NE | 660 |
| Click-through-probability on “Start free trial” 点击率 | CTR | 0.08 |
| Probability of enrolling, given click 总转化率 | PGC | 0.20625 |
| Probability of payment, given enroll 留存率 | PR | 0.53 |
| Probability of payment, given click 净转化率 | PNC | 0.1093125 |
| Total number of cookies 假设样本有5000个cookie | N | 5000 |

对于总转化率

PGC = 0.20625

NGC = 5000 \* CTR = 400

标准差SEGC = 0.0202

对于净转化率

PNC = 0.1093125

NNC = 5000 \* CTR = 400

标准差SENC  = 0.0156

两个分析估计与经验估计都不会有差异，是由于它们的分析单元都是cookie与实验的分组单元应该相似。 分析单元指的就是指标的分母。

## Sizing

### Number of Samples vs. Power

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.)

总转化率的样本大小NS = 25835，然而总转化率的分母是点击按钮的总数，所以还要通过点击率来计算出cookie的总数。并且由于分两个组所以要乘以二。总数N = NS / CTR \* 2 = 645875

对于净转化率的样本大小计算方式与总转化率一样，NS = 27413所以总数N = NS / CTR \* 2 = 685325

因此样本大小取更大的值，为685325浏览量.

### 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.)

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

假设我们不需要同时进行其他实验，所以可以将每天最大的流量都给本实验。每天都流量为40000，计算685325 / 40000后得出需要18天

# 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 day-by-day data. **Do not proceed to the rest of the analysis unless all sanity checks pass.**

对三个不变指标进行检测

|  |  |  |
| --- | --- | --- |
|  | Pageview | Click |
| Ncontrol | 345543 | 28378 |
| Nexperiment | 344660 | 28325 |
| Ntotal | 690203 | 56703 |
| P | 0.5 | 0.5 |
| P^ | 0.5006 | 0.5005 |
| SE | 0.0006 | 0.0021 |
| Margin | 1.96 \* 0.0006 = 0.0012 | 1.96 \* 0.0021 = 0.0041 |

Pageview的 95% confidence interval是[0.4988,0.5012],p^为0.5006在区间内。

click的 95% confidence interval是[0.4959,0.5041]，p^为0.5005在区间内。

对于点击率

Pcon = 0.082126， Pexp = 0.082182

d^ = Pxep – Pcon = 0.000056

Ppool = 56703 / 690203 = 0.082154

SE\*\*2 = Ppool(1 - Ppool) \* （1 / Ncontrol + 1 / Nexperiment）, SE = 0.000661

Margin = 1.96 \* SE = 0.00013

H0: d = 0 真正的差额等于零 ，那么d^ 为N(0,SE)的正态分布

If d^ > margin or d^ < -margin, reject H0并认为差别具有统计显著性。

因为d^ = 0.000056 ,所以不能拒绝原假设，也就是承认差额等于零。

所以在合理性检查中三个不变指标都通过了测试。

## 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.)

数据：

|  |  |  |  |
| --- | --- | --- | --- |
|  | Click | Enrollment | Payment |
| Control | 17293 | 3785 | 2033 |
| Experiment | 17260 | 3423 | 1945 |

|  |  |  |
| --- | --- | --- |
|  | 总转换率 | 净转化率 |
| Pcontrol | 0.2189 | 0.1176 |
| Pexperiment | 0.1983 | 0.1127 |
| d^ | -0.0206 | -0.0049 |
| Ppool | 0.2086 | 0.1151 |
| SEpool | 0.0044 | 0.0034 |
| Margin = SEpool \* 1.96 | 0.0086 | 0.0067 |
| Confidence interval:  [d^ - margin, d^ + margin] | [-0.0292, -0.012] | [-0.0116, 0.0018] |
| Statistical significant  如果区间内不包含0则是 | 是 | 否 |
| dmin | 0.01 | 0.0075 |
| Practically significant  dmin不在区间内则是 | 是 | 否 |

如果置信区间不包含 0，即可以确信发生了变化，则该指标具有统计显着性，所以实验组的总转换率显著低于控制组，在统计上和实际上都显著。而经转换率在两组之间不具有显著性的区别。

## Recommendation

Make a recommendation and briefly describe your reasoning.

对于三个不变指标的完整行检测，三个指标都通过了检测，这说明除了实验的改变之外，控制组和实验组没有收到其他因素影响，同时也表示本次实验之后的结果是有效的。对评估指标总转化率和净转化率的检验结果表示，总转化率的变化是显著的，与预期一致，推荐时间不足的人去参与免费访问课程材料导致注册免费试用的用户减少，从而导致总转化率降低。而净转化率是不显著的，并且置信区间包括负值，这表示起初的对该指标的预期不成立，而且净转化率还有下降的可能性。因此根据本次实验的结果建议不要启动这个改变。

# Follow-Up Experiment

Give a high-level description of the follow up experiment you would run, what your hypothesis would be, what metrics you would want to measure, what your unit of diversion would be, and your reasoning for these choices.

接下来的实验可能希望对净转化率进一步研究，净转化率：14天期限后保持注册的用户id数 / 点击按钮的cookie数量。如果想要提高净转化率要提高14天免费试用后还保持注册的用户数量。而在14天期间用户可能退出的主要原因是内容过难，用户感到受挫。因此设置本实验的内容为，在每个阶段结束后增加课程难度的评价界面，如果用户选择困难则推荐相关前置课程提示学生可以通过学习这个课程更好的理解本课程；如果选择其他则不推荐。

因为实验发生在注册后，所以分组单位使用用户id。

评估指标为净转化率。

不变指标为用户id的数量。

Zhibo Liu