

优达学城数据分析师纳米学位

A/B 测试项目——Hawli

1. 试验设计

1.1 指标选择

a: 选择不变指标

- **Cookie 的数量:** 即查看课程概述页面的唯一 cookie 的数量。由于 cookie 数量是在试验页面出现之前, 因此不受试验影响。
- **点击次数:** 即点击“开始免费试学”按钮的唯一 cookie 的数量, 由于点击次数在试验之前, 因此不受试验影响。
- **点进概率:** 即点击“开始免费试学”按钮的唯一 cookie 的数量除以查看课程概述页面的唯一 cookie 的数量所得的比率。由于点进概率在试验前, 因此不受试验影响。

b: 选择评估指标

- **总转化率:** 即完成登录并参加免费试学的用户 id 的数量除以点击“开始免费试学”按钮的唯一 cookie 的数量所得的比率。由于用户必须在进行了试验之后才决定是否继续登录并参加免费试学, 因此此试验有可能对其选择产生影响, 可作为评估指标, 期望能够减少总转化率, 让用户更加谨慎选择。
- **留存率:** 即在 14 天的期限过后仍参加课程(因此至少进行了一次付费)的用户 id 数量除以完成登录的用户 id 的数量。由于所有通过 14 天期限且仍参加课程的用户 id 是由登录并参加免费试学的用户 id 留存下来的, 因此都参加过试验, 已筛选掉部分由于时间不足而无法参加试学的学生, 因此选择此指标可以反映出试验的效用, 期望此指标能反应试验后留存率能够提升。
- **净转化率:** 即在 14 天的期限后仍参与课程的用户 id 的数量(因此至少进行了一次付费)除以点击了“开始免费试学”按钮的唯一 cookie 的数量所得的比率。14 天的期限后仍能参加课程, 则为我们目标用户, 此试验的目的就是筛选出最后能通过试用并继续参加课程的用户, 因此此指标为关键指标, 期望通过试验, 净转化率不降低。

c: 未选择指标

用户 id 的数量: 即参与免费试学的用户数量。用户 id 的产生是在试验之后, 因此会受到试验的影响, 不能作为不变指标, 而试验过程中, 用户 id 产生差异的原因, 有可能是实验组和对照组的差异产生, 也有可能是由两组 cookie 的数量不同而产生的影响, 因此其差异不能用来评估试验效果。已有的评估指标已经能够反应试验的差异, 因此不需要用用户 id 来作为评估指标。

1.2 测量标准偏差

总转化率的标准偏差为 0.0202, 分析估计与经验变异类似, 因为分组单位是 cookie, 而分析单位也是 cookie。

留存率的标准偏差为 0.0549, 分析估计与经验变异不一致, 因为分组单位为 cookie, 而分析单位为用户 id, 如果时间允许有必要进行经验估计。

净转化率的标准偏差为 0.0156, 分析估计与经验变异类似, 因为分组单位是 cookie, 而分析单位是 cookie。

1.3 规模

样本数量和功效

样本分析以总转化率和净转化率为评估指标。由于留存率所需评估样本数量过大，从实际试验时间和成本考虑不适宜，而且净转化率已经足够反应测验的假设，因此剔除留存率作为分析指标。

在分析阶段不采用 Bonferroni 校正，因为 Bonferroni 校正主要适用于 n 次独立检验，而选择总转化率和净转化率并不是相互独立的指标，因此使用 Bonferroni 校正会过于保守。

实验正确设计所需要的页面浏览量为 685325。

持续时间和曝光比例

我会选择将 100% 的页面流量转入此试验，需要 18 天完成试验。

选择把全部流量用于试验主要出于以下考虑：一是试验环节目前并未对其他试验造成影响（具体得根据其他试验情况），独立性较强，而且此试验对优达学城来说风险较小，并未影响后台运作与其他风险。另外，由于试验需求流量较大，因此全部投入试验可以减少试验周期，快速得到试验结果，假如试验有成效，可以快速投入使用，增加盈收。

2. 试验分析

合理性检查

对三个不变指标：cookie 的数量和“开始免费试用”的点击数进行合理性检查分别如下：

cookie 的数量，下限 0.4988，上限 0.5012，观察值 0.5006，通过检验；

“开始免费试用”的点击数，下限 0.4959，上限 0.5041，观察值 0.5005，通过检验；

点击概率：下限 0.0812，上限 0.0831，观察值 0.0821，通过检验。

结果分析

效应大小检验

对两个评估指标：总转化率和净转化率进行效应大小检验分别如下：

总转化率，下限 -0.0291，上限 -0.0119，此指标具有统计和实际显著性；

净转化率，下限 -0.0116，上限 0.0019，此指标不具备统计和实际显著性。

符号检验

对两个评估指标：总转化率和净转化率使用每日数据进行符号检验分别如下：

总转化率， $p=0.0026$ ，此指标结果具有统计显著性；

净转化率， $p=0.6776$ ，此指标结果不具有统计显著性。

汇总

在分析过程中，未使用 Bonferroni 校正，因为评估指标并非互相独立，具有一定的相关性，不适合使用此校正方式，而且使用此方式会过于保守。效应大小检验和符号检验之间不存在差异。

3. 建议

从试验分析看来，此试验会降低总转化率，也就是会让用户更谨慎地决定要不要试学，充分考虑自己的时间因素，从而减少学生的挫折感，但净转化率的置信区间存在负数，从计算结果看有很大的概率净转化率会减少，并且有一定的概率净转化率的减少会超过实际显著性

0.0075, 这是我们这个试验不想看到的结果, 因此没办法说明“降低程度不大”。不建议启动此试验。

4. 后续试验

由于前面试验的目的是为了减少受挫并想要提前终止课程的学生数量, 假设是否有足够的时间是影响因素之一, 而根据本人学习课程的过程中发现, 学生的背景也会让学生受挫并想要提前终止学习。

因此, 后续试验会以拟定先修条件, 测试学生的背景, 在主页面增加测试页面, 学生点击开始免费试学之后, 会有一个页面让其进行背景测试, 然后根据测试结果有两个选项, 一是继续登录并试学, 二是建议先修部分免费课程。此试验假设通过背景测试, 可降低学生因背景知识差异而导致学生受挫, 可减少提前终止课程的学生数量, 增加继续付费用户。

测量评估的指标有:

总转化率: 即完成登录并参加免费试学的用户 id 的数量除以点击“开始免费试学”按钮的用户 id 所得的比率。该指标可以让用户根据自己的背景来选择是否进一步进行试学, 期望能够减少因为先修背景不足而受挫的学生。

留存率: 即在 14 天的期限过后仍参加课程(因此至少进行了一次付费)的用户 id 数量除以完成登录的用户 id 的数量。该指标可以看出先修背景是否真的对学生进一步学习有影响, 期望能够通过该指标的分析提高留存率, 也就是减少学生提前终止学习的数量。

净转换率: 即在 14 天的期限后仍参与课程的用户 id 的数量(因此至少进行了一次付费)除以点击了“开始免费试学”按钮的用户 id 所得的比率。该指标显示最终进行课程学习的学生数量, 期望更多的用户能进行试用的后续学习。

本试验采用的分组单位为用户 id, 因为用户 id 作为用户的唯一识别号, 用户不管用任何浏览器或不同平台登录, 都一样的结果, 假如用 cookie, 容易导致用户在同一天时间内换不同的浏览器或平台登录, 发现不一样的页面和结果。因此, 采用用户 id 为分组单元最为合适。

参考资料:

<http://graphpad.com/quickcalcs/binomial2/>