

优达学城数据分析师纳米学位

A/B 测试项目

试验设计

指标选择

不变指标：不变指标是指在对照组和实验组都不变的指标。因此我们选择不变指标为：Cookie 的数量、点击次数、点进概率。评估指标为：总转化率、留存率、净转化率。

Cookie 的数量、点击次数和点进概率三个指标都强调了唯一性，并且用户登录后他们的用户 id 便被跟踪不会有重复测量，因此这三个指标不应受变更的影响。所以将其选为不变指标。

考虑到我们需要测量试验是否会减少由于时间原因无法完成课程的学生数量，同时不会大量减少继续通过免费试学和最终完成课程的学生数量。因此需要得出试验前后即实验组和对照组中总转化率、留存率、净转化率是否真的有差异。

实验组和对照组的用户 ID 数量不同可能是由于实验的影响，也可能是由于两组 cookie 的不同。所以使用用户 ID 数量的区别不能够很好的评估试验的效果。在一个比例化的评估度量（总转化率）存在的情况下，我们可以不选择用户 ID 的数量作为评估度量。

- Cookie 的数量：即查看课程概述页面的唯一 Cookie 的数量。
- 点击次数：即点击“开始免费试用”按钮（在免费试用屏幕触犯前发生）的唯一 cookie 的数量。
- 点进概率：即点击“开始免费试用”按钮的唯一 cookie 除以查看课程概述页面的唯一 cookie 的数量所得的结果。
- 用户 ID 的数量：即报名参加免费试用的用户数量
- 总转化率：即完成登录并报名参加免费试用的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果。
- 留存率：即在 14 天期限后仍保持参加（并进行了至少一次支付）的用户 id 的数量除以完成登录的用户 id 的数量。
- 净转化率：即在 14 天期限结束后仍然参加（并至少进行了一次支付）的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果。

期望从评估指标中获得的试验结果：总转化率降低，即减少由于时间原因无法完成课程并因此受挫的学生数量。留存率有显著的提高，净转化率不会大量减少。

测量标准偏差

$$s = \sqrt{\frac{P*(1-p)}{n}}$$

标准偏差：

总转化率	留存率	净转化率
0.0202	0.0549	0.0156

总转化率和净转化率的分析标准偏差与经验标准差相同，因为分析单元和分组单元相同，都是 cookie。留存率的分析单元是用户 ID，分组单元是 cookie，因此留存率的分析标准偏差与经验标准差可能不相同。

规模

样本数量和功效

在分析阶段不会使用 Bonferroni 校正

计算页面浏览量: <http://www.evanmiller.org/ab-testing/sample-size.html>

页面浏览量:

总转化率	留存率	净转化率
645875	4,741,212	685,325

持续时间和曝光比例

根据上面对三个评估指标的页面浏览量分别进行计算所需天数:

总页面浏览量/每天的页面浏览量 (40000)

所需天数:

总转化率	留存率	净转化率
17	119	18

显然留存率需要的天数过长,不在我们考虑的范围之内,而总转化率和净转化率分别需要 17, 18 天,是个不错的选择,在此我们选择净转化率。因为净转化率的分组单元是 cookie,该试验针对的是用户是否有时间,不涉及用户隐私,没有收集敏感信息。而且对本来有时间学习的用户的付费行为没有影响。实验没有涉及到数据库及后台的改动,不会造成数据丢失等风险。

所以选择曝光比例 100%

试验分析

合理性检查

对于每个不变指标,对你在 95%置信区间下期望观察到的值、实际观察的值及指标是否通过合理性检查给出结论。

不变指标	下限	上限	观察值	是否通过
Cookie 的数量	.4988	.5012	.5006	通过
点击次数	.4959	.5041	.5005	通过
点进概率	.0812	.0830	.0822	通过

结果分析

效应大小检验

没有使用 Bonferroni 校正法

评估指标	下限	上限	有无统计和实际显著性
总转化率	-0.0291	-0.012	有
净转化率	-0.0116	.0019	没有

符号检验

使用 了 Bonferroni 校正法

评估指标	P 值	是否有统计性显著意义
总转化率	0.0026	是

净转化率	0.6776	否
------	--------	---

汇总

说明你是否使用了 Bonferroni 校正，并解释原因。若效应大小假设检验和符号检验之间存在任何差异，描述差异并说明你认为导致差异的原因是什么。

在我的分析中没有使用 Bonferroni 校正，因为 Bonferroni 更正是一种在多个独立指标比较中使用的方法，总转化率和净转化率有较强的相关性。

建议

提供建议并简要说明你的理由。

通过上述分析结果，不建议使用该筛选器，虽然该实验减少了参加免费试用的时间不足的学生数量，但是净转化率告诉我们筛选器对继续免费试用的学生数量产生了负面影响，减少了付款的学生人数。

后续试验

对你会开展的后续试验进行概括说明，你的假设会是什么，你将测量哪些指标，你的转移单位将是什么，以及做出这些选择的理由。

将随机抽取的学生进行 24 小时追踪，每天如果不能达到固定的学习时间，就会对其发出提醒，每天在固定的时间进行提醒学习时间是否达标，如果没有达标提出建议。

假设这样会提高学生的学习积极性。

测量没有被跟踪的学生和被跟踪的学生在一周到两周内的学习进度是否有区别。

我将使用 user-id 作为转移的单位，仅限注册学生，将用户 id 的数量作为不变度量标准，将项目的通过率作为评估度量标准。

注册学生人数不应该在实验组或对照组间变化，因此将它作为不变量度量，项目的通过率应该随实验组和对照组改变，因此可以用作评估指标。

优达学城

2016 年 9 月