

优达学城数据分析师纳米学位

A/B 测试项目

试验概述：免费试学筛选器

在进行此试验时，优达学城当前的主页上有两个选项：“开始免费试学”和“访问课程资料”。

如果学生点击“开始免费试学”，系统将要求他们输入信用卡信息，然后他们将进入付费课程版本的免费试学。14 天后，将对他们自动收费，除非他们在此期限结束前取消试用。若学生点击“访问课程材料”，他们将能够观看视频和免费进行小测试，但是他们不会获得导师指导支持或验证证书，无法提交最终项目来获取反馈。

在此试验中，优达学城测试了一项变化，如果学生点击“开始免费试学”，系统会问他们有多少时间投入到这个课程中。如果学生表示每周 5 小时或更多，将按常规程序进行登录。如果他们表示一周不到 5 小时，将出现一条消息说明优达学城的课程通常需要更多的时间投入才能成功完成，并建议学生可免费访问课程资料。在这里，学生可选择继续进行免费试学，或免费访问课程资料。这张截图展示了试验概况。

我们的假设是这会为学生预先设定明确的期望，从而减少因为没有足够的时间而离开免费试学，并因此受挫的学生数量，同时不会在很大程度上减少继续通过免费试学和最终完成课程的学生数量。如果这个假设最后为真，优达学城将改进整体学生体验和提高导师为能够完成课程的学生提供支持的能力。

转移单位为 cookie，尽管学生参加的是免费试学，但在登录后他们的用户 id 便被跟踪。同一个用户 id 不能两次参加免费试学。对于不参加免费试学的用户，他们的用户 id 不会在试验中被跟踪，即使他们在访问课程概述页面时登录了网站。

试验设计

不变指标

- cookie 的数量：即访问课程概述页面的唯一 cookie 的数量。（d 最小=3000）（不会受实验的影响，用户在被提问前访问课程概述页面）
- 点击次数：即点击“开始免费试学”按钮的唯一 cookie 的数量（在免费试学筛选器触发前发生，不受实验影响）。（d 最小 =240）
- 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的唯一 cookie 的数量所得的比率（d 最小 =0.01）（不会受实验的影响）

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。（d 最小 =0.01）（会受实验的影响，实验组中参加免费试学的学员数量可能会下降，因为某些学员不能够每周学习 5 小时以上）
- 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。（d 最小 =0.01）（会受实验的影响，实验组中的留存率比对照组增多，因为他们注意到了每周要学习 5 小时以上，在试用过程中会更好的学习效果）
- 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”按钮的唯一 cookie 的数量所得的比率。（d 最小

=0.0075) (我们假设学习时间不足的同学后续都不会付费，因此实验中对净转化率的期望是不会减少，即置信区间可以包含 0 和大于 0，但是不可以包含负数。)

用户 id 数量发生于试验之后，会受到试验的影响，因此可以选做一个评估度量。但是，由于实验组和对照组的 cookie 数量不一定相同，也就是说两组中用户 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。

测量标准偏差

下表包含了三个度量的基准值的粗略估计

Unique cookies to view page per day:	40000
Unique cookies to click "Start free trial" per day:	3200
Enrollments per day:	660
Click-through-probability on "Start free trial":	0.08
Probability of enrolling, given click:	0.20625
Probability of payment, given enroll:	0.53
Probability of payment, given click	0.1093125

有 5000 个 cookie 样本大小访问课程概述页面的情况下

标准偏差估计：(根据二项式分布计算)

$$SD = \sqrt{p^{\wedge} \times (1 - p^{\wedge}) / N}$$

总转化率：

$$p^{\wedge} = 0.20625$$

$$N = 5,000 \times (3,200 / 40,000) = 400$$

$$SD = \sqrt{0.20625 \times (1 - 0.20625) / 400} = 0.0202$$

留存率：

$$p^{\wedge} = 0.53$$

$$N = 5,000 \times (660 / 40,000) = 82.5$$

$$SD = \sqrt{0.53 \times (1 - 0.53) / 82.5} = 0.0549$$

净转化率：

$$p^{\wedge} = 0.1093125$$

$$N = 5,000 \times (3,200 / 40,000) = 400$$

$$SD = \sqrt{0.1093125 \times (1 - 0.1093125) / 400} = 0.0156$$

在此实验中，转移单位为 cookie，是经验估算中的分母。在计算总转化率和净转换率时，标准方差的分析估算也是以 cookie 作为分母的，因此我预期分析估计与经验变异是类似的。对于留存率，计算单位是登录的用户 id 数而不是 cookie 数，所以我预期分析估计与经验变异是不同的。

规模

样本数量和功效

使用 $\alpha = 0.05$, $\beta = 0.2$ 。我根据此[在线工具](#)，计算出所需样本容量。

计算实验组和控制组所需网页浏览数：

总转化率: $2 \times 25835 \times 40000 \div 3200 = 645875$

留存率: $2 \times 39115 \times 40000 \div 660 = 4741212$

净转化率: $2 \times 27413 \times 40000 \div 3200 = 685325$

在分析阶段我不使用 Bonferroni 校正, 总共需要收集的网页浏览数: 4741212

持续时间和曝光比例

总共需要收集的网页浏览数 4741212

测试所有 3 个指标:

天数(100%的流量) = $4741212 \div 40000 = 118.5303$

即使使用 100%的流量, 测试所有 3 个指标(总转化率、留存率、净转化率)需要 118 天, 对于实验来说时间太长了。减少流量更会增加实验时间。因此, 必须减少测试指标, 我只选择 2 个指标: 总转化率和净转化率

网页浏览数: 685325

测试总转化率和净转化率:

Days (100% 的流量) = $685325 \div 40000 = 17.133125$

Days (50% 的流量) = $685325 \div 40000 \div 0.5 = 34.26625$

对于总转化率和净转化率实验, 我们需要 18 天时间, 每天 100%的流量。这个时间比起 118 天更合理。但是为了更好的分析用户的行为, 我们决定每天使用 50%的流量, 将时间增加到 35 天。

由于本实验并不会要求用户输入其他敏感信息, 也不会使用户有受到伤害的风险, 风险较低。同时, 考虑到商业风险较低, 本实验可以让用户更好的计划和评估他们的学习时间。理论上讲, 支付的用户数量可能会因为免费试用的人减少而减少, 但目前这仅仅是一种假设。我们主要考虑的因素是不浪费学生的时间和导师资源。需要指出的是, 使用 50%的流量不会对商业利益有很大影响。

试验分析

合理性检查

95%置信区间, Z 值等于 1.96

Cookie 的数量:

$p^{\wedge} = 0.5$

$N = N_{\text{con}} + N_{\text{exp}} = 345,543 + 344,660 = 690,203$

$SE = \sqrt{p^{\wedge} \times (1 - p^{\wedge}) / 690,203} = 0.00060184$

$m = 1.96 \times 0.00060184 = 0.0011796$

置信区间: [0.4988, 0.5012]

观察值: 0.5006

观察值处于置信区间内

“开始免费试用”的点击数:

$p^{\wedge} = 0.5$

$N = N_{\text{con}} + N_{\text{exp}} = 28,378 + 28,325 = 56,703$

$SE = \sqrt{p^{\wedge} \times (1 - p^{\wedge}) / 56,703} = 0.0020997$

$m = 1.96 \times 0.0020997 = 0.0041155$

置信区间: [0.4959, 0.5041]

观察值: 0.5005

观察值处于置信区间内

“开始免费试用”的点入概率:

$$p^{\wedge}_{\text{pool}} = (28,378 + 28,325) / (345,543 + 344,660) = 0.08216$$

$$SE_{\text{pool}} = \sqrt{0.08216 \times (1 - 0.08216) \times (1/28,378 + 1/28,325)} = 0.00066106$$

$$m^{\wedge} = 1.96 \times 0.00066106 = 0.0012956776$$

$$d^{\wedge} = 28,325/344,660 - 28,378/345,543 = 0.000056627$$

置信区间: [-0.0013, 0.0013]

观察值: 0.0001

观察值处于置信区间内

所有的不变指标都通过了合理性检查

结果分析

效应大小检验

总转化率:

$$d^{\wedge} = (3,423 / 17,260) - (3,785 / 17,293) = -0.02055$$

$$p^{\wedge}_{\text{pool}} = (3,423 + 3,785) / (17,260 + 17,293) = 0.208607$$

$$SE_{\text{pool}} = 0.00437$$

$$m = 0.0085685$$

置信区间: [-0.0291, -0.012]

置信区间不包含 0, 具有统计显著性

置信区间不包含实际显著性边界($d_{\min}=0.01$), 具有实际显著性

净转化率:

$$d^{\wedge} = (1,945 / 17,260) - (2,033 / 17,293) = -0.0048737$$

$$p^{\wedge}_{\text{pool}} = (1,945 + 2,033) / (17,260 + 17,293) = 0.115127$$

$$SE_{\text{pool}} = 0.0034341$$

$$m = 0.0067309$$

置信区间: [-0.0116, 0.0018]

置信区间包含 0, 不具有统计显著性

置信区间包含实际显著性边界($d_{\min}=0.0075$), 不具有实际显著性

符号检验

使用[在线工具](#), 进行符号检验

总转化率:

总数: 23

实验组大于对照组数: 4

The two-tail P value is 0.0026

对于 $\alpha = 0.05$, $P_{\text{value}} < \alpha$, 结果具有统计学显著性

净转化率:

总数: 23

实验组大于对照组数: 10

The two-tail P value is 0.6776

对于 $\alpha = 0.05$, $P_{\text{value}} > \alpha$, 结果不具有统计学显著性

汇总

效应大小检验和符号检验都表明, 网站的改变将会在统计学上和实际上显著减少总转化率, 但是不会在统计学上和实际上显著影响净转化率。

我没有使用 Bonferroni 校正进行分析, 因为评估度量总转化率和净转化率之间有相关性, 所以才造成使用 Bonferroni 校正过于保守, 两者并不是并列关系。

此外, 如果要进行 Bonferroni 校正的话, 除了判断指标是否相互独立, 我们还要判断几个评估度量之间是与还是或的关系。(我们期望这几个指标同时满足还是满足其中之一) 如果指标之间是相互独立的, 但是得出最终结论时是与的关系, 这样我们进行 Bonferroni 校正也会过于保守。在这里我们的实验要求总转化率和净转化率指标同时满足期望。

建议

我的建议是不发布此项实验改变, 因为:

- 在实验中, 实验组的总转化率如我们预期的那样减少了, 并且具有统计学显著性和实际显著性。这会为我们节省免费试用中的成本。
- 但是, 我们预期净转化率不应当减少。实验没有达到我们的预期, 改变不具有统计学显著性和实际显著性。同时, 置信区间中包含负值, 也就意味着净转化率有减少的可能, 网站会失去一部分免费试用后留下来的用户, 这将会减少网站的收入, 带来商业风险。

后续试验

我建议以下对实验做以下两项改变:

- 将每周花 5 小时学习变成非强制性的要求, 仅仅作为提示和推荐信息, 这样可以避免流失一部分不需要每周花 5 小时学习也能毕业的学员。
- 对不能花 5 小时学习的学员提供等级测试, 通过的仍然可以入学, 这样不会流失能力高的学员

我的假设是

- 总转化率将会显著减少
- 净转化率不会减少

实验测试指标不变

- 转移单位: cookie.
- 不变指标: cookie 的数量、点击数、点击概率
- 评估指标: 总转化率和净转化率

参考资料

[A/B Test](#)

[A/B Test](#)

[符号测试](#)

[样本容量计算](#)

[符号测试在线计算](#)