

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

列出你将在项目中使用的不变指标和评估指标。（这些应与你在“选择不不变指标”和“选择评估指标”小测试中使用的指标一样）

对于每个指标，解释你为什么使用或不使用它作为不变指标或评估指标。此外，说明你期望从评估指标中获得什么样的试验结果。

1. 不变指标：

a) Number of cookies（Cookie 的数量）：

访问课程概述页面的唯一 cookie 的数量。cookie 是分组单元，且在访问者看到页面更改之前，cookies 已经生成了，因此是不变指标。

期望结果：不变；

b) Number of clicks（点击次数）：

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）。首页没有任何改变，在点击“开始免费试学”按钮前时，访问者还看不到该试验的变化。

期望结果：不变；

c) Click-through-probability（点进概率）：

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 的数量所得的比率点击次数。点击次数不变，cookie 的数量不变，点进概率也不变。

期望结果：不变

2. 评估指标：

a) Gross conversion（总转化率）：

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。因为该试验组与对照组不同点在于，系统会问用户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。而提醒用户每周 5 小时学习时间，可能会影响完成登录并参加免费试学的用户 id 的数量，这也是想通过 A/B 测试来研究的问题。因此，Gross conversion 可以作为评估指标。

期望结果：减小；部分学生会因为没有足够的时间而离开免费试学，因此总转化率期望是减小。

b) Retention（留存率）：

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。因为提醒用户每周 5 小时学习时间，会对付费用户和完成免费试用的用户的比例产生影响，是个很好的评估指标。此指标在后续计算页面浏览量中得出的值太大而舍弃。

期望结果：增大；因为根据预期，期望最终通过免费试学和最终完成课程的学生数量不变，而登录的用户 id 数量减少，留存率会增大。

c) Net conversion (净转化率)：

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。因为提醒用户每周 5 小时学习时间，会对付费用户产生影响，因此净转化率是个评估指标。

期望结果：不变；因为根据实验预期，付费用户影响不大，且点击“开始免费试学”按钮的唯一 cookie 数量不变，所以净转化率的期望结果不变。

3. 无关指标：

Number of user-ids (用户 id 的数量)：

根据实验设计，用户 id 数量有可能变化，因此不能是不变指标；同时，也不是好的评估指标，因为实验组和控制组的点击 cookie 数量可能不同，所以测试出来的变化不一定仅仅是实验的影响，也肯能是由于两组的 cookie 不同。

测量标准偏差

列出你的每个评估指标的标准偏差。(这些应是来自“计算标准偏差”小测试中的答案。)

对于每个评估指标，说明你是否认为分析估计与经验变异是类似还是不同（如果不同，在时间允许的情况下将有必要进行经验估计）。简要说明每个情况的理由。

1. Gross conversion (总转化率)：0.0202

总转化率是基于 cookie 数量，也是分组单位。因此分组单位等于分析单位，分析估计值与经验变异类似；

2. Retention (留存率)：0.0549

留存率是基于完成登录的用户 id 的数量，与分组单位 cookie 不相似，分析单位和分组单位不相似，所以分析估计值与经验变异不相似；

3. Net conversion (净转化率)：0.0156

净转化率是基于 cookie 数量，也是分组单位。因此分组单位等于分析单位，分析估计值与经验变异类似；

规模

样本数量和功效

说明你是否会在分析阶段使用 Bonferroni 校正，并给出实验正确设计所需的页面浏览量。(这些应是来自“计算页面浏览量”小测试中的答案。)

不使用 Bonferroni 校正，因为本次试验的指标总转化率和净转化率具有高度相关性，并非独立的，使用 Bonferroni 校正会影响试验结果，使结果过于保守。

Gross conversion(总转化率)计算得出 645875，Retention(留存率)计算得出 4741212，

Net conversion(净转化率)计算得出 685325，由于留存率的值太大，舍弃留存率，因此最终所需的浏览量为 685325。

持续时间和曝光比例

说明你会将多少百分比的页面流量转入此试验，以及鉴于此条件，你需要多少天来运行试验。(这些应是来自“选择持续时间和曝光”小测试中的答案。)

页面浏览量：685325

曝光的流量部分：1

试验持续时间：18

说明你选择所转移流量部分的原因。你认为此试验对优达学城来说有多大风险？

在该试验中，系统会问实验组的用户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。只是询问用户能投入的学习时间，不会对用户的生理和心理产生不良影响，不涉及伦理问题。同时学习时间不属于敏感信息，不涉及隐私问题。对优达网站来说，该试验未造成过多的改变，未增加网站的负担，不需要对系统进行大幅改造。

考虑到试验周期，仅将部分流量分流至该试验，会导致周期变长，增加不可控因素。因此综合风险和试验周期，可将全部流量用于该试验。

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。（这些应是来自“合理性检查”小测试中的答案）

对于任何未通过的合理性检查，根据每日数据解释你觉得最有可能的原因。在所有合理性检查通过前，不要开始其他分析工作。

1. Number of cookies（Cookie 的数量）：
置信区间=[0.4988, 0.5012]
观察值=0.5006
通过合理性检查
2. Number of clicks（点击次数）：
置信区间=[0.4959, 0.5041]
观察值=0.5005
通过合理性检查
3. Click-through-probability（点进概率）：
置信区间=[0.0812, 0.0830]
观察值=0.0821824
通过合理性检查

结果分析

效应大小检验

对于每个评估指标，对试验和对照组之间的差异给出 95% 置信区间。说明每个指标是否具有统计和实际显著性。（这些应是来自“效应大小检验”小测试的答案。）

1. Gross conversion（总转化率）：

置信区间=[-0.0291, -0.0120]

置信区间不包括 0，具有统计显著性；

置信区间不包含 d_min，具有实际显著性

2. Net conversion（净转化率）：

置信区间=[-0.01160, 0.001857]

置信区间包括 0，不具有统计显著性；

置信区间包含 d_min（+/- 0.0075），不具有实际显著性；

符号检验

对于每个评估指标，使用每日数据进行符号检验，然后报告符号检验的 p 值以及结果是否具有统计显著性。（这些应是“符号检验”小测试中的答案。）

1. Gross conversion（总转化率）：

成功数量：4

试验次数：23

概率：0.5

双尾 P 值：0.0026

双尾 P 值 0.0026 小于 alpha 水平 0.025，具有统计显著性；

2. Net conversion（净转化率）：

成功数量：10

试验次数：23

概率：0.5

双尾 P 值：0.6776

双尾 P 值 0.6776 大于 alpha 水平 0.025，不具有统计显著性；

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。

不使用 Bonferroni 校正，因为本次试验的指标总转化率和净转化率具有高度相关性，并非是独立的，使用 Bonferroni 校正会影响试验结果，使结果过于保守。

建议

提供建议并简要说明你的理由。

不建议发布免费试学筛选器。

总转化率具有负面的统计显著性和实际显著性，说明试验会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，符合试验预期。

净转化率不具有实际显著性和统计显著性，其置信区间包含负数，置信区间的含义是“我们有 95% 的信心试验结果会落在这个区间”。根据置信区间(-0.0116, 0.0019)，即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。因此此试验可能“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”。

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。

为减少受挫并想要提前终止课程的学生数量，针对中国用户，我设计一个试验：在课程中提供更多的中文资料来辅助学习，而不是以英文资料为主。

假设提供更多的中文资料，会提高用户的付费率。

不变指标：参与免费试学的用户 ID 数量。此试验关注的是用户的付费学习，因此选择用户 ID 数量作为不变指标。

评估指标：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。试验可能会影响最终付费用户数量，而留存率反映注册用户进一步成为付费用户的比例，是个非常好的评估指标。

转移单位：用户 ID。此试验关注的是已经点击“开始免费试学”并注册的用户，ID 会被跟踪，因此选择用户 id 作为转移单位，统计用户的行为。

优达学城

2016 年 9 月