

优达学城数据分析师纳米学位

A/B 测试项目

邓剑波

试验设计

指标选择

不变指标: Cookie 数量 点击次数 点进概率

评估指标: 总转化率 留存率 净转化率

- Cookie 的数量: 指标为不变指标, cookie 在用户选择“开始免费试用”按钮前已经确定, 试验前后不会发生改变。
- 用户 id 的数量: 此指标不可作为不变指标, 也不能作为评估指标, 试验以 cookie 进行分组, 用户 ID 的数量在试验组和控制组中有可能产生不同的结果, 所以不适合作为不变指标; 另外, 评估指标中总转化率是对用户 id 数量的归一化处理, 同时, 由于 cookies 分组的时候并不完全一致, 所以该指标也不适合作为评估指标。
- 点击次数: 这里定义的点击次数是唯一的点击 cookie 数量, 同 cookie 数量, 在试验前和试验后都不会发生变化。
- 点进概率: 可以作为不变指标, 由于 cookie 已经是唯一的, 且前后一致的, 所以可以作为不变指标。
- 总转化率: 可作为评估指标, 总转化率可以看出控制组和试验组在用户转化上的差别, 发生在用户选择了“是否能够保证 5 小时学习时间”后, 也是 ab 测试希望判别的异同处, 期望看到的结果总转化率的分子减少, 分母不变, 预期会有显著的减少。
- 留存率: 可作为评估指标, 原因同上, 留存率可以反应试验对吸引学生继续学习并且支付是否存在帮助, 期望结果是留存率的分子不变, 分母减少, 预期会有显著的增加。
- 净转化率: 可作为评估指标, 是对“是否能够保证 5 小时学习时间”的测试点的总概述指标, 能够反应测试点在控制组和试验组是否有显著的不用, 期望的结果是净转化率的分子不变, 分母不变, 预期不会有显著的变动。

测量标准偏差

- 总转化率标准差: 0.0202, 这里的“分析单位”是点击了“免费试用”的 cookie, 而“分组单位”也是点击“免费试用”的 cookie, 虽然不可能完全相同, 但是高度相关, 分析估计是正确的, 如果时间允许, 可多收集试验数据。

- 留存率标准差: 0.0549, 这里的“分组单位”是 cookie, 而“分析单位”是完成登陆的用户 ID 数量, 两者不同, 所以分析估计并不是准确的。
- 净转化率标准差: 0.0156, “分析单位”和“分组单位”都是点击了“免费试用”的 cookie, 高度相关, 分析估计是正确的, 如果时间允许, 可收集更多试验数据。

规模

- 样本数量和功效

不使用 bonferroni 校正, 因为 bonferroni 校正后过于谨慎, 使得试验所需样本数量过大, 不适合操作, 选择总转化率和净转化率, 而留存率因为“分组单位”和“分析单位”不同, 会计算出非常大的样本数量需求, 所以不予选择。

最后得出需要 685275 的页面流量进行试验。

- 持续时间和曝光比例

选择 60% 的页面流量进行试验, 所需时间为 29 天。

选择 60% 的曝光比例, 主要原因在于考虑时间成本, 过小的曝光比例会使得试验时间太长, 在 60% 下可以将试验控制在一个月内完成, 较为合理; 同时, 过大的试验比例有可能会影响网页的正常使用, 有可能会有过多的用户受到例如学习效应等因素的影响。

从用户使用习惯上看, 增加了询问是能够满足 5 小时的学习选项, 本质上不会影响真正有需要的用户去完成支付学习课程, 只需重新选择开始免费试用即可, 同时, 由于试验并没有对其他页面进行过大的改动, 用户也不需要经过新的学习过程来适应新的变化, 所以风险可控。

另外, 试验并没有丢后台数据等系统方面进行更改, 所以不存在更高的系统崩溃或者数据丢失的可能性, 涉及支付和录入个人信息的流程上, 控制组和试验组并没有不同, 两者的不同只是在页面的引流上, 所以安全性上也没有更高的风险。

最后, 由于不涉及类似需要进行隐私条款授权等的道德问题, 所以也没有道德风险。

总结, 该试验的风险可控, 并没有突出的风险点。

试验分析

合理性检查

在 95% 置信区间下可得:

- Cookie 的数量: 下限 0.4988, 上限 0.5012, 观测值 0.5006, 通过合理性检查
- 点击次数: 下限 0.4959, 上限 0.5041, 观测值 0.5005, 通过合理性检查
- 点进概率: 下限 0.0812, 上限 0.0830, 观测值 0.0822, 通过合理性检查

结果分析

效应大小检验

不使用 bonferroni 校正, 在 95% 的置信区间下可得:

- 总转化率：下限 -0.0291，上限 -0.0120， $dmin=0.01$ ，具有统计显著性和实验显著性
- 净转化率：下限 -0.0116，上限 0.0019， $dmin=0.0075$ ，不具有统计显著性和实验显著性

符号检验

不使用 bonferroni 校正， $\alpha=0.05$ 下可得：

- 总转化率：在总共 23 天是试验中，试验组高于控制组天数为 4 天，通过网络计算器可得 p 值 0.0026，具有统计显著性
- 净转化率：在总共 23 天是试验中，试验组高于控制组天数为 10 天，通过网络计算器可得 p 值为 0.6776，不具有统计显著性

汇总

没有试用 bonferroni 校正，两个指标相关性太高，使用 bonferroni 校正，将会过于保守。

通过对效应大小校验和符号检验来看，总转化率同时具备统计显著性和实验显著性，而净转化率同时不具有统计显著性和实验显著性。

建议

建议不要上线“免费试学筛选器”。

首先，总转化率由统计显著性，但是试验组的总转化率显著比控制组差，说明试验并没有改进指标，并且试验还使得指标变差。

其次，我们对试验的预期是可以减少注册人数，但是净转化率不会由较大的变动，虽然从试验结果上看，净转化率并没有显著性变动，但是由于净转化率的置信区间是 $(-0.0116, 0.0019)$ ，所以，净转化率的负向变动还是存在较大可能，并且有一定的可能性超过 $dmin=0.0075$ 的变动，这并不是我们希望看到的结果。

所以，“免费试学筛选器”的改动并不适合上线运营。

后续试验

选择调整课程排序，通过对课程难度由浅入深进行重新排序，假设不同的课程排序方式会对课程完成率产生显著性影响。

通过假设，可以选择不变指标：

- 1、注册 ID 总数：作为试验的 unit of diversion，在试验发生前已经确定，不涉及更改。

同时，选择以下指标作为评估指标：

- 1、完成率：支付过课程费用并通过课程的 ID 数除以支付过课程费用的 ID 总数；
- 2、注册支付率：支付过课程费用的 ID 数除以注册 ID 总数。

其中，完成率转移单位为课程付费用户，注册支付率的转移单位为注册用户

选择指标的理由在于，通过对注册支付率可以看出，课程的重新排序，是否能够吸引注册用户付费参与课程，能否和原课程排序产生显著性不同；而完成率可以观察到用户实际的学习体验，是否能够使得用户持续学习完成课程，与原课程排序是否有显著性不同。

