

优达学城数据分析师纳米学位

A/B 测试项目

说明: [点击此处下载此文档的英文版本](#)。

模板格式

此模板可用于组织你最终项目的答案。应从你的答案中复制到小测试中的项用蓝色显示。

试验设计

指标选择

不变指标:

1. **cookie 的数量:** 即访问课程概述页面的唯一 cookie 的数量 (d 最小=3000)。因为它不会受实验的影响, 用户在被提问前就已经访问课程概述页面。
2. **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) (d 最小=240)。因为它不受实验的影响。
3. **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)。因为它也不受实验影响。

评估指标:

1. **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)。因为它受实验影响, 期望实验组中的参加免费试学的学院数量下降, 即期望实验组总转化率比控制组的下降。
2. **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)。因为它受实验影响, 期望实验组中留存率比控制组增多。
3. **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)。因为它受实验影响, 期望实验组的净转换率与控制组的无显著性不同, 因为筛选器只会影响参加免费试学的人数, 对付费行为影响不大。

用户 id 的数量是在实验之后被采集的度量, 受到实验的影响; 但是实验组和控制组的 cookie 不一定相同, 导致用户 id 数量的不同可能会因为 cookie 的不同引起, 所以不能将它作为评估指标。

测量标准偏差

通过标准偏差的计算公式 $SE = \frac{\sqrt{\hat{p}*(1-\hat{p})}}{\sqrt{N}}$, 在有 5000 个 cookie 样本大小访问课程概述页面的情况下, 计算测量标准偏差:

		SE
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	0.0202
Probability of payment, given enroll:	0.53	0.0549
Probability of payment, given click	0.1093125	0.0156

总转化率和净转化率的分析估计和经验变异是类似的,因为这两个指标的分析单位和引流单位是唯一 cookie; 而留存率的分析估计与经验变异是不同的,因为其分析单位是用户 id 数量, 但引流单位是 cookie 数量。

规模

样本数量和功效

在分析阶段我不使用 Bonferroni 校正, 因为总转化率和净转化率并不独立, 在这里使用 Bonferroni 校正会过于保守, 它适用于多次独立实验。

使用 $\alpha=0.05$, $\beta=0.2$, 通过在线工具, 计算出所需样本容量是:

1. 总转化率: $2 \times 25835 \times 40000 / 3200 = 645875$
2. 留存率: $2 \times 39115 \times 40000 / 660 = 4741212$
3. 净转化率: $2 \times 27413 \times 40000 / 3200 = 685325$

取最大的样本容量, 所以总共需要收集的网页浏览数是 4741212。

持续时间和曝光比例

为了减少实验时间, 将会用 100% 的页面流量转入此实验, 总共需要收集的页面浏览数是 4741212, 所以需要 $4741212 / 40000 = 118.5303$ 天, 即大约需要 118 天来进行实验。可见时间太长了, 不适合选取它作为评估指标。所以选取第二大的样本容量 685325, 其需要 $685325 / 40000 = 17.1131$ 天, 即大约需要 18 天来进行实验, 这个时间比较合理。

实验风险很小:

1. 实验对页面展示的改动很小, 仅仅是在学生点击“开始免费试学”后多弹出一个询问框, 不会对用户产生感情上的冲击, 用户也不需要花长时间去适应页面的改动。
2. 实验不会阻止学生进行免费试学, 不会影响用户的使用习惯。
3. 实验没有记录额外的个人信息, 不会对用户个人信息安全造成风险。
4. 实验没有后台和数据库的改变, 不会导致网站崩溃用户无法访问等大问题。
5. 实验没有道德风险。

试验分析

合理性检查

各不变指标的 95% 置信区间、实际观察值和是否通过合理性检查情况如下:

指标	置信区间下限	置信区间上限	观察值	是否通过
----	--------	--------	-----	------

cookie 的数量	0.4988	0.5012	0.5006	是
点击次数	0.4959	0.5041	0.5005	是
点进概率	-0.0013	0.0013	0.0001	是

因为这三个指标的观察值都处于置信区间内，所以都通过合理性检查。

其中检查 cookie 的数量和点击次数的合理性用到以下公式：

$$\hat{p} = \frac{N_{con}}{N_{con} + N_{experiment}}$$

$$SE = \sqrt{\frac{0.5 * 0.5}{N_{con} + N_{experiment}}}$$

$$m = 1.96 * SE$$

$$x_{0,1} = 0.5 \pm m$$

检查点进概率的合理性用到以下公式：

$$\hat{p}_{pool} = \frac{X_{con} + X_{exper}}{N_{con} + N_{exper}}$$

$$SE_{pool} = \sqrt{\hat{p}_{pool} * (1 - \hat{p}_{pool}) * \left(\frac{1}{N_{con}} + \frac{1}{N_{exper}} \right)}$$

$$\hat{d} = \frac{X_{exper}}{N_{exper}} - \frac{X_{con}}{N_{con}}$$

$$m = 1.96 * SE_{pool}$$

$$x_{0,1} = 0 \pm m$$

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。）

各评估指标 95%置信区间、实际观察值和是否具有统计和实际显著性情况如下：

指标	置信区间下限	置信区间上限	统计显著性	实际显著性
总转化率	-0.0291	-0.0120	是	是
净转化率	-0.0116	0.0019	否	否

总转化率的置信区间不包括 0，具有统计显著性，而且不包含实际显著性边界($d_{min} = 0.01$)，具有实际显著性；

净转化率的置信区间包括 0，不具有统计显著性，而且也包含了实际显著性边界($d_{min} = 0.0075$)，不具有实际显著性。

其中检查以上 2 个指标的显著性用到以下公式：

$$\hat{p}_{pool} = \frac{X_{con} + X_{exper}}{N_{con} + N_{exper}}$$

$$SE_{\text{pool}} = \sqrt{\hat{p}_{\text{pool}} * (1 - \hat{p}_{\text{pool}}) * \left(\frac{1}{N_{\text{con}}} + \frac{1}{N_{\text{exper}}} \right)}$$

$$\hat{d} = \frac{X_{\text{exper}}}{N_{\text{exper}}} - \frac{X_{\text{con}}}{N_{\text{con}}}$$

$$m = 1.96 * SE_{\text{pool}}$$

$$x_{0,1} = \hat{d} \pm m$$

符号检验

对于每个评估指标，使用每日数据进行符号检验，利用在线符号检验计算器，检验情况如下：

指标	P 值	统计显著性
总转化率	0.0026	是
净转化率	0.6776	否

对于总转化率：

总数是 23，实验组小于控制组的数量是 19，双尾 p 值<0.05，结果具有统计显著性；

对于净转化率：

总数是 23，实验组小于控制组的数量是 13，双尾 p 值>0.05，结果不具有统计显著性。

汇总

我没有使用 Bonferroni 校正，因为评估度量的两个指标，总转化率和净转化率不是独立，具有相关性，若有 Bonferroni 校正会过于保守。Bonferroni 校正适用于多次独立实验。

效应大小检验和符号检验表明了实验组与控制组的总转化率存在显著性差异，即增加筛选器会显著减少总转化率。而实验组与控制组的净转化率不存在显著性差异，增加筛选器不会对净转化率带来变化。

建议

我建议不发布此项实验。

因为虽然实验组的总转化率会显著较少，这符合我们的预期，增加筛选器可以较少登陆参加免费试学的学员数目，节省了免费使用的成本。但净转化率没有显著性的变化，而且其置信区间包含负值，即意味着转化率有可能减少，网站会失去一部分免费试学后留下来的学员，所以这实验达不到我们的预期，我们本来预期实验的净转化率不会减少。

后续试验

后续实验概述：

免费试听部分课程。

实验说明：

尝试开放部分课程作为免费试学，当用户点击免费试学，看到的课程只是部分内容，想查看

全部内容则需要付费。

假设：

总转化率会不会显著减少，净转化率显著增加。

测量指标不变：

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