

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）

不变指标：

`Cookie`的数量：访问课程概述页面`cookie`

点击次数：点击“开始免费试学”按钮的唯一`cookie`的数量。

点击概率：点击“开始免费试学”`cookies`/课程概述页面`cookie`

评估指标：

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。

净转换率：在14天的期限仍参加课程的用户`id`的数量除以点击了“开始免费试学”按钮的唯一`cookie`的数量所得比率。

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。

区分不变度量和评估度量的主要标准是度量采集于试验之前还是试验之后。如果采集于试验之前，试验组与对照组之间理论上应该没有区别，是不变的。而如果度量发生于试验之后，会受到试验的影响，可以作为评估试验的效果。

因为`cookie`和点击次数，点击概率采集于试验之前，故作为不变指标。

总转化率和净转化率采集于试验之后，故作为评估指标。

User `id`数量和留存率没有被选为不变度量，也没被选为评估度量。

用户`id`数量：它发生于试验之后，会受到试验的影响，因此它是`ok`的评估度量。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`数量的区别不能很好的评估试验效果。

留存率：留存率发生在试验之后，是不错的评估指标。不过经过后续的计算，我们发现他需要过多的页面浏览量和试验运行时间，因此在规定的时间内我们无法采集足够的样本数据，也不适合作为评估度量。

期望的试验结果：

总转化率有统计显著性，显著降低。说明实验组中的部分学生因时间不够而没有登录免费试学环节。

净转化率没有统计显著性，净转化率预期基本不变。

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。）

评估指标的标准偏差:

$$SE = \sqrt{p(1-p)/N}$$

Probability of enrolling, given click: 0.20625

Probability of payment, given click: 0.1093125

Unique cookies to view page per day: 40000

Unique cookies to click "Start free trial" per day: 3200

假设有5000cookie样本大小访问课程概述页面，N（点击“免费试学”数量）=400.

$$SE（总转化率）=0.0202$$

$$SE（净转换率）=0.0156$$

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。

分析估计和经验变异是类似的。总转化率和净转化率的坟墓虽然都是点击“开始免费试学”按钮的唯一cookie的数量，但是“点击是一个动作，是发生在cookie之上的，分析单位是cookie。本试验的转移单位是cookie。因此分析单位=转移单位，是可以确保引流的一致性的，因此分析估计与经验变异类似。

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。）

在分析阶段不使用Bonferroni校正。选择总转化率和净转换率为评估度量。

$$\text{Alpha}=0.05, \text{beta}=0.2$$

总转化率的最小实际显著量 $d_{min}=0.01$ ， $p（总转化率）=0.20625$

在线计算器得25835.

因为对照组和实验组两组，且总转化率基于click， $p（click）=0.08$

$$25835 * 2 / 0.08 = 645875$$

净转化率的最小实际显著量 $d_{min}=0.0075$ ， $p（总转化率）=0.1093$

在线计算器得27413

因为对照组和实验组两组，且总转化率基于click， $p（click）=0.08$

$$27413 * 2 / 0.08 = 685325$$

取所需样本量较大值，所需的页面浏览量为：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。）

选择流量（fraction=1），因为该试验风险较低，没有数据库的改变，且没有道德的风险。

持续时间=页面浏览量/每天的平均页面浏览量

$$685325 / 40000 = 17.1$$

持续时间需要18天

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？

选择流量（fraction=1），因为该试验风险较低，没有数据库的改变，且没有道德的风险。

试验分析

合理性检查

对于每个不变指标，对您在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）

Cookie:

$N(\text{exp}) = 344660$, $N(\text{cont}) = 345543$

$SE = \sqrt{0.5 * 0.5 / (34660 + 345543)} = 0.00060$

$m = Z * SE = 1.96 * 0.00060 = 0.001179$

下限: 0.4988

上限: 0.5012

通过合理性检查。

Number of Click“start free trial”:

$N(\text{exp}) = 28325$, $N(\text{cont}) = 28378$

$SE = \sqrt{0.5 * 0.5 / (28325 + 28378)} = 0.00209$

$m = Z * SE = 1.96 * 0.00209 = 0.004115$

下限: 0.4958

上限: 0.5041

通过合理性检查。

点击概率:

对照组: $N(\text{clicks}) = 28378$, $N(\text{pageviews}) = 345543$

$P(\text{对照组}) = 28378 / 345543 = 0.0821$

$SE = \sqrt{0.082 * (1 - 0.082) / 345543} = 0.00046$

$M = Z * SE = 1.96 * 0.00047 = 0.0009$

上限: 0.0812

下限: 0.0830

$P(\text{实验组}) = 28325 / 344660 = 0.0822$

实验组的概率落入区间，点击概率通过合理性检查。

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。**

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。）

Gross conversion:

注册数 点击数 概率

实验组: 3423 17260 0.1983

对照组: 3785 17293 0.2188

Ppool: 0.2086

SE: $\sqrt{P_{\text{pool}} * (1 - P_{\text{pool}}) * (1/N(\text{实验}) + 1/N(\text{对照}))} = 0.004372$

差异: -0.02055

$M = z * SE$: 0.008568

下限: -0.02912

上限: -0.01198

总转化率: 由于置信区间不包括0, 实验结果具有统计显著性, 且具有实际显著性, 显著减少。

Net conversion:

注册数 点击数 概率

实验组: 1945 17260 0.1127

对照组: 2033 17293 0.11176

P_{pool} : 0.1151

SE: $\sqrt{P_{\text{pool}} * (1 - P_{\text{pool}}) * (1/N(\text{实验}) + 1/N(\text{对照}))} = 0.003434$

差异: -0.0049

$M = z * SE$: 0.0067

下限: -0.0116

上限: 0.0019

净转化率: 实验结果不具有统计显著性。

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。)

Gross conversion:

$N=23$, $P(\text{实验组}) > P(\text{对照组})$ 有4组数据, $P\text{-value}=0.0026$

小于0.05, 总转化率有统计显著性

Net conversion:

$N=23$, $P(\text{实验组}) > P(\text{对照组})$ 有10组数据, $P\text{-value}=0.6776$

大于0.05, 净转化率没有统计显著性。

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。

未使用 Bonferroni 校正。本试验中指标有相关性, 因此使用 Bonferroni 校正过于保守。

根据计算结果, 效应大小的假设检验和符号检验之间没有差异。

建议

提供建议并简要说明你的理由。

不建议启动。总转化率具有统计和实际显著性, 显著减少, 是我们希望看到的结果, 表示减少了因为没有足够的时间而离开免费试学的受挫学生数量。但是净转化率的置信区间包含负数, 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”, 根据此处的计算结果 (-0.0116, 0.0019), 也就是说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0.0075. 因此我们无法说明“降低的程度不大”。所以不建议启动。

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。

后续试验：为付费页面增加解释好处的试验，假设该改动会增加注册学员为辅导服务付费。

测量指标

不变指标：

用户id：注册学员数量。

点击次数：点击“申请辅导”按钮的cookie

评估指标：

注册学员为辅导服务付费的概率： $\text{申请辅导服务的学员数（为辅导付费）} / \text{注册学员数}$ 。

转移单位：用户id。