

优达学城数据分析师纳米学位 P7

试验设计

度量选择

列出你将在此用作不变度量和评估度量的度量。

答：

- Cookie 的数量：即查看课程概述页面的唯一 cookie 的数量。

我选择cookie作为不变度量。cookie是总体规模指标，分组单元是cookie，cookie被随机的分配到实验组和对照组，各组的cookie数量大致相等。

- 用户 id 的数量：即报名参加免费试用的用户数量。

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。

用户id的数量不适合作为不变量度量。如果这项变化有效果，那么实验组参加免费使用的用户数会减少。另外，用户id的数量也不适合作为评估度量，应该用报名参加免费试用的用户数量除以点击“开始免费试用”按钮来衡量转化率而不是一个绝对数值（报名参加免费试用的用户数）。分子的增减可能是有分母本身的增减带来的。

- 点击次数：即点击“开始免费试用”按钮（在免费试用屏幕触犯前发生）的唯一 cookie 的数量。

我选择点击次数作为不变度量，实验组中点击“开始免费试学”按钮的次数应该与对照组大致相同。只有当实验组中用户点击“开始免费试学”，系统才回跳出消息，所以跳出消息之前的用户行为不受影响。

- 点进概率：即点击“开始免费试用”按钮的唯一 cookie 除以查看课程概述页面的唯一 cookie 的数量所得的结果。

点进概率，也可以作为不变度量。原因同上，跳出消息之前的用户行为不受影响，所以实验组和对照组的点击概率应该大致相同。

- 总转化率：即完成登录并报名参加免费试用的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果。

我选择总转化率作为评估变量。如果这一项系统变化能告知学生所需投入时间，为其设定明确的期望，那么会有一部分原本打算点击“开始免费试用”的学员因时间原因，而不选择免费试用。从而实验组的总转化率会小于对照组。

- 留存率：即在 14 天期限后仍保持参加（并进行了至少一次支付）的用户 id 的数量除以完成登录的用户 id 的数量。

留存率不适合做不变度量。登录的用户id的数据量可能变小。

我选择把留存率作为一个评估度量。

试验的预期是减少因学习时间不足受挫的学生的数量，但是原本付费的学生预期不减少，使用留存率可以帮助判断出最终付费的学生比例是否会增加。实验组的留存率应该有所增加。

- 净转化率：即在 14 天期限结束后仍然参加（并至少进行了一次支付）的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果。

我认为净转化率是一个评估变量。

本试验还有一个假设是，实验组中那些看到系统消息后选择放弃免费试用的学生，本来就不会在 14 天试用期限后仍然参加，因此这个系统变化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。结合前面提到的点击“开始免费适用”按钮的 cookie 数量也是不变的，所以实验组的净转化率应该不低于对照组。

测量标准偏差

列出你的每个评估度量的标准偏差。

答：

总转化率：0.0202

留存率：0.0549

净转化率：0.0156

总转化率和净转化率的分析单元都是点击按钮的 cookie id，和引流单元一致，所以我认为对标准偏差的分析估计与经验估计是类似的。

但是，留存率的分析单元是登录用户 id，如果有足够的时间，留存率的标准偏差应该用经验估计。

规模

样本数量和支持

说明你是否会在分析阶段使用 Bonferroni 校正，并给出你适当开展试验所需的支持网页访问数。

答：

不会在分析阶段使用 Bonferroni 校正。

网页访问数量 ($\alpha = 0.05$, $\beta = 0.2$)：685325。由于留存率所需的网页访问数过多

（4741212，即使每天的流量都用于试验，也需要 119 天才能完成试验），本试验中不考察留存率的变化及显著性，用总转化率所需网页访问数（645875）和净转化率所需网页访问数（685325）的较大值作为试验的网页访问数。

持续时间和风险暴露

说明你会将哪一部分流量转入此试验，以及鉴于此条件，你需要多少天来运行试验。

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？

答：
选择0.8（80%）的流量转入此试验，需要22天运行（向上取整）。

选择部分流量的主要原因是可以同时进行其他试验，还可以减少节假日对流量的影响。
本试验对优达学城无太大风险。原因如下：

- 1. 即使有新系统通知，学生仍然可以选择免费试学及后续课程。不会因每周投入时间少于5小时而被禁止免费试学，所以不会对学员产生太多的负面影响。
- 2. 新增的系统通知，并不涉及用户的信息安全问题。
- 3. 这个系统通知是“一扫而过”的消息提示，应该不会使用户的体验变差。
- 4. 也不涉及道德上的风险。

试验分析

合理性检查

对于每个不变度量，对你期望观察到的值、实际观察的值及度量是否通过合理性检查给出95% 置信区间。
对于任何未通过的合理性检查，根据每日数据解释你猜测的最可能的原因。在所有合理性检查通过前，不要开始其他分析工作。

答：
下表是3个不变度量、95%的置信区间的上下边界值、实际观察值。3个不变度量都通过健全性检验。

不变度量	下限	上限	观察值	通过/未通过
Cookie 的数量	0.4988	0.5011	0.5006	通过
点击次数	0.4959	0.5041	0.5005	通过
“开始免费试用” 的点进概率	-0.0013	0.0013	-0.0001	通过

结果分析效应大小检验

对于每个评估度量，对试验和对照组之间的差异给出 95% 置信区间。说明每个度量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。）

答：
下表是2个评估度量、95%的置信区间的上下边界值（由于本试验的网页浏览数量未满足留存率所需的网页访问数，本试验中不考察留存率的变化及显著性）。
总转化率有统计显著性，置信区间不包含0，并且不包含实际显著性边界。
净转化率无统计显著性，置信区间包含0，包含显著性边界。

评估度量	下限	上限	统计显著性	实际显著性
总转化率	-0.0291	-0.0120	有	有
净转化率	-0.0116	0.0019	无	无

符号检验

对于每个评估度量，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。）

答：

下表是2个评估度量符号检验的p值。

在符号检验中，总转化率有统计显著性，净转化率无统计显著性。

评估度量	p-value	统计显著性
总转化率	0.0026	有
净转化率	0.6776	无

汇总

如果你选择了多个评估度量，你将需要决定是否使用 Bonferroni 校正。在决定时，记住你发起此试验想要获得的结果。

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。

答：

不会在分析阶段使用Bonferroni矫正。Bonferroni校正主要适用于n次独立检验。但是本试验中的总转化率和净转化率是相关联的，用户的支付行为发生在免费适用之后。因此使用Bonferroni校正会使得试验结果过于保守。

建议

提供建议并简要说明你的理由。

答：

我不建议实施这项变化。

总转化率衡量点击“开始免费试用”按钮的学生中有多少比例的学生会参加免费试用。通过分析，可知总转化率具有统计和实际显著性，并且通过了符号检验，也就是说我有95%的把握这项变化对用户有影响，总转化率会降低（置信区间是负）。

但是，净转化率的置信区间包含了负数。也就是说，此项变化如果实施，有一定的概率使得净转化率降低。本试验的预期是原本会付费学习的学生数不会因这项变化而减少，净转化率降低不是我们希望得到的。

我认为是否开展更深入的探究，需要和业务相关人员进行协商，试验是否有显著产出，以及后续试验的成本大小。

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，你将测量哪些度量，你的转移单位将是什么，以及做出这些选择的理由。

答：

在试验中，在学生课程页面中增加一个明显的悬浮窗，告知学生如果学习上的问题可以咨询导师。需要对一段时间内参加课程的学生进行为期几个月的学习进度跟踪。这段时间主要取决于样本数量和每日可用流量。跟踪观察时间有两种选择：1.大致等于（或略大于）课程普遍完成时长；2.考虑到完整的学习周期太长，如果有现成的业务结论表明，学生提前终止课程大多发生在学习前期或中期，观察时间也可以选择学习前期或中期时长。

假设：提供导师指导的实验组中，提前终止课程的学生比例要少于没有提供导师指导的对照组。

不变度量：用户id数量（参加正式课程的用户数量），试验中的变化不影响参加正式课程的用户数量。

评估度量：辍学率（提前终止课程的用户id数量除以参加正式课程的用户id数量），辍学率可以衡量提供导师指导是否能减少学生提前终止课程。

引流单元：用户id。不使用cookie，因为学习过程中学生可能会切换设备，如从浏览器到手机客户端，但是不同设备上的用户行为都属于一个用户。