

优达学城数据分析师纳米学位 P7

试验设计

度量选择

不变量指标 (Invariant Metrics):

Number of cookies: 实验组和对照组应该是按 Cookies 随机分配得来的, 所以二者应该大致相等。它并不会因为实验而改变, 所以适合作为不变量指标, 作评估指标的话没有意义。

Number of clicks: 在用户点击按钮之前, 他们还未看到实验, 所以实验组和对照组的点击数应该是相等的。它并不会因为实验而改变, 所以适合作为不变量指标, 作评估指标的话没有意义。

Click-through-probability: 在用户点击按钮之前, 他们还未看到实验, 所以实验组和对照组的 CTP 应该是相等的。它并不会因为实验而改变, 所以适合作为不变量指标, 作评估指标的话没有意义。

评估度量 (Evaluation):

Gross conversion: 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。因为按推理, 这一指标会受到实验影响, 所以不适合作为不变量指标。在我们的假设中, 这个指标应该能反映出实验界面对用户最后是否是参与课程的影响, 我们想看看这个界面是否会大幅减少试学的用户, 所以会很有价值。

Retention: 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和课程注册的用户数量。因为按推理, 这一指标会受到实验影响, 所以不适合作为不变量指标。这个指标应该能反映出实验对最后参与课程并坚持下来的比例是否有提高, 对比之前的 Gross conversion 就可以总结出实验的效果怎么样, 是否可以不大幅的减少试学的用户, 同时提高留下来学生的坚持学习的概率。

Net conversion: 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。因为按推理, 这一指标会受到实验影响, 所以不适合作为不变量指标。我们预计, 实验中的设置会减少参加课程的人数, 同时又可以提高学生坚持下来学习的概率, 如果实验中 Net conversion 比对照组的高或者相等, 就说明了实验的益处要超过它减少用户数的害处, 总体上是创造更多收益的。

注: 对于 Number of user-id, 在用户点击尝试之后, 实验页面就会出现, 所以实验组和对照组的数量应该是不同的, 不适合作为不变量指标, 另外, 实验是按 cookie 分配的, 实验组和对照组的 cookie 数量可能不相同, 也就不好区别用户 ID 数量的不同是因为实验影响还是 cookie 的数量的影响, 所以 ID 也不适合作为评估指标。

这次修改后的内容:

本试验的目的是通过该举措减少学习时间不足的同学参加免费试学，这样的话假设这部分同学后面不会继续付费，所以留存率减少，净转化率基本不变，净转化率会适当减少（但如果他们觉得自己时间不够，直接不试学了呢？这样留下来参加学习的学生会投入的，留存率反而会上升。不过就自己来说，确实即使时间不够，有免费的试学机会我最后还是会试一试的，只是在最后不会付费。如果像我这样的学生是大多数，那么这项举措减少的应该是付费了，但没有时间保证完成，最后白交费了而沮丧的学生。我觉得应该引入一个新的指标，完成了所有项目的学生/付费了的学生（但可能需要耗时 6 个月）。

测量标准偏差

根据计算，经过分析估计的每个评估度量的预测标准差如下：

Gross conversion 标准差：0.0202 unit of analysis = Unique Cookie of Clicks

Retention 标准差：0.0549 unit of analysis = Unique User ID

Net conversion 标准差：0.0156 unit of analysis = Unique Cookie of Clicks

Gross conversion 和 Net conversion 的 unit of analysis = unit of diversion，因此它们的分析变异性匹配经验变异性；而 Retention 的 unit of analysis 为唯一的用户 ID，与 unit of diversion（唯一的 Cookie）不一致，所以它的变异性应该不等于经验变异性，应该考虑用经验法则或 A/A 测试等其他方法来评估。

规模

样本数量和支持

计算所需要的访问数后发现，如果想达到效果，Retention 指标的网页浏览量需要达到 474 万条，所以决定放弃这个指标，只使用 Gross Conversion 和 Net Retention 综合其他的指标，所需要的网页访问数量为 685325。

不使用 Bonferroni 校正，因为：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并不是相互独立的，而是相关联的，因此使用 Bonferroni 校正会使得试验结果过于保守。而且，实验中总共只使用了 2 个评价指标，犯第一类错误的概率很小，也没有必要使用 Bonferroni 校正。

持续时间和风险暴露

我将会使用全部流量转入此试验，因为实验中需要 68 万多条网页方位，为了避免时间（节日、季节、假期）变化等对实验的影响，我想尽快将数据收全。本次试验可能会降低用户参与课程的数量，但按常理来说影响不会很大，也不会对用户造成伤害，即使采用全部流量也有 50% 的用户体验丝毫未变，不会影响所有人。本实验没有道德和伦理上的风险，试验中也仅仅改变了浏览的页面，对于后台中的数据并没有修改，在数据安全上没有威胁。总共需要 18 天的时间收集数据。

试验分析

合理性检查

不变量	Lower bound	Upper bound	Observed	Passes
Number of cookies	0.4988	0.5012	0.5006	Yes
Number of clicks on "Start free trial"	0.4959	0.5041	0.5005	Yes
Click-through-probability on "Start free trial"	-0.0013	0.0013	0	Yes

如上表所示，所有不变量的观测值都在相应的置信区间内，通过了合理性检查。

结果分析效应大小检验

Evaluation metrics	Lower bound	Upper bound	Statistical significance	Practical significance
Gross conversion	-0.0291	-0.012	Yes	Yes
Net conversion	-0.0116	0.0019	No	No

Gross conversion 的置信区间不包括 0, 而且最小的差异都有 0.012, 大于 $d_{min} = 0.01$, 所以统计效果和实际应用效果都显著。

Net conversion 的置信区间包括 0, 而且最小的差异为 0.0019, 小于 $d_{min} = 0.0075$, 所以统计效果和实际应用效果都不显著。

符号检验

Gross conversion 的符号检验 p-value 为 0.0026, 小于 $\alpha = 0.05$, 结果显著。

Net conversion 的符号检验 p-value 为 0.6776, 大于 $\alpha = 0.05$, 结果不显著。

汇总

不使用 Bonferroni 校正，因为：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并不是相互独立的，而是相关联的，因此使用 Bonferroni 校正会使得试验结果过于保守。而且，实验中总共只使用了 2 个评价指标，犯第一类错误的概率很小，也没有必要使用 Bonferroni 校正。

本次实验中，效应大小假设检验和符号检验之间并未出现差异。

建议

实验中的界面会降低用户的总转化率（Gross conversion），大约在 1%-3% 之间。净转化率（Net conversion）不显著，但是其置信区间中包含负数，我们认为这种界面设置会有降低净转化率的风险。综合考虑，我认为这种新的界面并没有带来足够的好处，并具有风险，不建议实行。

后续试验

我想设计一个新实验来测试“在规定时间内学完课程，并完成所有项目，Udacity 提供 50% 的学费返还”这种政策是否可以提高课程参与人数和最后完成课程的比例，因为心理学中，相比收益，人们更在乎自己的损失，我们希望这种将减少费用和学习投入结合起来的方法可以更好的激励人们，从而改善他们的完成情况。为了让用户了解这一政策，我们可以在课程浏览界面上标注这一信息，也就是说用户在点击“Start Now”按钮之前就可以看到，这样也可以使信息传递到更多的用户。点击“Start Now”之后会要求用户登录，并进行 2 周内的免费试用。

转移单位：

选择唯一 Cookie，因为如果选择唯一的用户 ID，对于很多用户，我们无法做到用户点击“Start Now”之前进行分流，也就是说很难保证对照组和实验组的平均分配，而 Cookie 能做到这一点，而且也能大致对应上每一个人。

不变度量：

Number of cookies: 实验组和对照组应该是按 Cookies 随机分配得来的，所以二者应该大致相等。它并不会因为实验而改变，所以适合作为不变量指标，作评估指标的话没有意义。

Number of clicks: 在用户点击按钮之前，他们还未看到实验，所以实验组和对照组的点击数应该是相等的。它并不会因为实验而改变，所以适合作为不变量指标，作评估指标的话没有意义。

Click-through-probability: 在用户点击按钮之前，他们还未看到实验，所以实验组和对照组的 CTP 应该是相等的。它并不会因为实验而改变，所以适合作为不变量指标，作评估指标的话没有意义。

评估指标：

Gross conversion: 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。按推理，实验提高用户积极性，从而提高总的转化率。

Retention: 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和课程注册的用户数量。按推理，实验能够增加用户完成的动力，因此提高用户的留存率。

Hours spent every week: 用户每周花在网上这门课的学习时间（按小时计算）。按照推理，实验能够提高用户完成学位的意愿，因此他们会更加投入，每周花费更多的时间。

Net conversion: 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。我们很好奇实验是否能提高用户的净转化率，因为按推理试验后该指标的分母也会增加。