

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

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）

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。

本实验中选择的不不变指标有：

1. Cookie 的数量：即查看课程概述页面的唯一 cookie 的数量，由于查看课程概述页面是发生在实验变化因素之前的，因此该指标应该是不不变指标。
2. 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。点击‘开始免费试用’按钮也是发生在实验变化因素之前的，因此可以作为不变指标。
3. 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。理由同上

本实验中选择的评价指标有：

1. 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。由于用户在点击‘开始免费试用’按钮之后用户的界面就会有变化，而这种变化会直接导致了登录并报名参加免费试用的用户 id 数量不同。也就是会影响到转化率。
2. 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。通过该指标可以用于说明对界面做的这种变化是否能够提升用户的留存率。

在这个过程中之所以没有选择用户 id 数量作为评价指标是由于用户 id 数量的差异不能反映出课程的整体转化率，而本实验比较在意的是转化率的改变。没有采用留存率作为评价指标是觉得净转化率一定程度上可以反映留存率，而且采用留存率计算得到的页面浏览量较大，因此不考虑这一变量。

希望取得的试验结果：

由于改动之后，一些自认为时间不够的同学会放弃登录和参加免费试学，因此实验组的总转化率应该低于控制组的总转化率，我们预期最终付费的学生数量不会发生太大的变化（预期参加免费试学的同学会减少，但是最终付费的同学数量不会发生太大的变化）因此推测净转化率预期不会发生太大的变化。

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。）

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。

对总转化率和净转化率分别计算标准偏差，所得结果如下：

标准误差就是SE	
Se-gross	0.020230604
se-net	0.015601545

分析估计和经验变异是类似的，对于总转化率和净转化率均采用 cookie 作为分组单元，因此这两个评估变量就是符合分析变异性匹配经验变异性的要求。

规模

样本数量和功效

实验过程中没有采用 Bonferroni 校正，采用净转化率和总转化率作为评估指标，对其样本数量的计算过程如下：

总转化率：0.20625， $d_{min} = 1\%$ 可推出 click on start free 为 25835，进而计算出 paper view 的数量为 $25835 \times 2 / 0.08 = 645875$

净转化率：0.1093125， $d_{min} = 0.75\%$ 可推出 click on start free 为 27413，进而计算出 paper view 的数量为 685325，选择此数量作为其最终结果。

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。）

持续时间和曝光比例

实验中计划将 60% 的流量转移到此实验当中，需要 35 天时间来进行实验，计算方法如下：
 $(685325 / 40000) / 0.6 = 29$

之所以采用 60% 进行流量的转移，是由于考虑到该种变化可能影响不是很大，所以可以将较多的流量用于测试。

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。）

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 60% 的流量进行测试，个人认为对优达学城的风险不大。其原因可以从以下几个角度说明：

1. 道德角度：实验只是对登录界面进行了简单改动，并未涉及到道德问题
2. 敏感数据：该种改动不涉及到对用户个人信息的收集，只是添加了一个提醒而已
3. 数据库安全：后台并未做改动，只是在前端进行了变动而已，不用担心数据库会崩溃，或者由于该改动导致的用户无法登陆的情况
4. 用户习惯：改动不是太大，不会让用户在心理产生太大的不适感，不会改变用户的学习

习惯

试验分析

合理性检查

完整性检查计算结果如下：结果显示所有评估量均通过检查。

Sanity checks

For each metric that you chose as an invariant metric, compute a 95% confidence interval for the value you expect to observe. Enter the upper and lower bounds, and the observed value, all to 4 decimal places. Check the box if the metric passes your sanity check.

	Lower bound	Upper bound	Observed	Passes
• Number of cookies	0.4988	0.5012	0.5006	<input checked="" type="checkbox"/>
• Number of user-ids				<input type="checkbox"/>
• Number of clicks on "Start free trial"	0.4959	0.5041	0.5005	<input checked="" type="checkbox"/>
• Click-through-probability on "Start free trial"	0.0812	0.0830	0.0822	<input checked="" type="checkbox"/>
• Gross conversion				<input type="checkbox"/>

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。）

其计算过程中用到的公式如下：

合并概率	注册人数和/点击人数和
合并标准误差	根号（（1-合并概率）×合并概率× （1/控制组点击人数+1/实验组点击人数））
差异	实验注册/实验点击 - 控制注册/控制点击

Excel 计算过程如下：

		有效大小检验				
		click	enroll	payment	gross	net
cont		17293	3785	2033	0.218874689	0.117562019
exper		17260	3423	1945	0.198319815	0.112688297
					Ppool	
					0.208607067	0.115127485
					Seppool	
					0.004371675	0.003434134
					D	
					-0.020554875	-0.004873723
					ME1.96	
					0.008568484	0.006730902
					-0.011986391	-0.011604624
					-0.029123358	0.001857179

从结果中可以看出，对于总转化率其置信区间为【-0.0291, -0.0120】，在 0 的一侧，因此认为有统计显著性，同时不包括【-0.01,0.01】所以具有实际显著性。但是对于净转化率，其置信区间为【-0.0116, 0.0018】，包含 0 点且和【-0.01,0.01】有重叠，因此不具有统计显著性和实际显著性。

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。）

符号检测结果：

对于 gross conversion 实验次数为 23 次，其中积极成果为 4 次，通过在线计算器计算可得到双尾 p 值为 0.026，小于 0.05 具有统计显著性，而对于 net conversion 而言，积极成果为 10 次，计算在线计算器计算可得到双尾 p 值为 0.6776，大于 0.05，因此不具有统计显著性。

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。

本实验中没有使用 Bonferroni 校正，这是由于 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非是独立的，而是相关联的，因此使用 Bonferroni 校正会使得试验结果过于保守

建议

提供建议并简要说明你的理由。

建议 launch 这个改动，从总转化率上看，其具有统计显著性和实际显著性，认为可以实施这种改动。

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。

为了提高注册学生的课程完成度，考虑添加一个邮件提醒的功能，在学生注册后即可提示是否添加邮件提醒功能并且留下电子邮箱，提示说道一旦系统检测你一段时间后没有进行学习

那么 udacity 将会给你发一封提醒邮件以提醒你多久没有学习了等。

我们假设这个邮件提醒功能可以让学生意识到学习进度下降或者是感觉到了羞愧感从而激发学生的学习动力，提高课程完成率，缩短课程完成周期：

评估指标： 平均每个项目的完成周期（即从开始每一个项目到通过该项目所用时间），完成纳米学位的周期（即从注册到最终拿到证书所用时间，这个评估变量的实验周期可能比较长），学生上课的时间总数（比如实验组的每周上课时间是不是显著比对照组高），学生的论坛发帖数量（用于说明参与度是否有提高）

不变量： 注册的学生数量，这是发生在发邮件提醒功能之前的

分组变量： 选择用户 id 作为分组变量，用户 id 基本对应着一个人，直接跟踪不同 id 完成课程项目过程中的表现即可。

优达学城

2016 年 9 月