

优达学城数据分析师纳米学位

A/B 测试项目

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。

不变指标

(1) Number of cookies

定义：查看课程概述页面的唯一 cookie 的数量。

因为测试的是首页中的链接页面，故不会影响到首页情况，也就不会影响 cookies 值。

期望结果：不改变。

(2) Number of clicks

定义：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。

因为首页没有改变，cookie 不变，点击次数也不变。

期望结果：不改变。

(3) Click-through-probability

定义：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。

因为点击次数不变，cookie 不变，点进概率也不变。

期望结果：不改变。

评估指标

(1) Gross conversion

定义：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。

测试进行了一个筛选，可能会影响完成登录并参加免费试用的用户 id 数量，故选择此指标。

由于 Gross conversion 是一个比值，下面通过分析，分子与分母可能的变化，来预估期望结果。

分子：完成登录并报名参加免费试用的用户 id 的数量

由于测试会对询问用户每周学习时间，然后给用户多了一个不选择开始免费试用的按钮（也就是“访问课程资料”），相当于起了一个分流的作用。也就是说，以前只有一个选项，

现在这个选项变成了两个，又假设进来的用户数不变，且这两个按钮均会有用户选择，那么这时候选择原来一个选项的用户就会变少（就是测试时候的免费试用按钮），那么分子就可能会变小。

分母：点击“开始免费试用”按钮的唯一 cookie 的数量

这里试验预期的分母是不变的

期望结果：由于分子变小，分母不变，故期望结果是变小。

(2) Retention

定义：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。

测试会对付费用户和完成免费试用的用户数量产生影响，故选为评估指标。

分子：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

即付费用户数量不变

分母：完成登录的用户 id 的数量

通过筛选（会影响到登录），登录用户数量相比之前会有所减少

期望结果：分子不变，分母减小，故期望结果是变大。

(3) Net conversion

定义：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。

测试会对付费用户产生影响，因此选为评估指标

分子：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

即付费用户数量不变

分母：点击“开始免费试用”按钮的唯一 cookie 的数量

点击“开始免费试学”按钮的用户数量并未改变，改变的是“访问课程资料”的点击数。

期望结果：分子不变，分母不变，故期望结果是不变。

未选择指标

Number of user-ids

根据实验设计，用户 id 数量有可能变化，因此不能作为不变指标。但是，也不能作为评估指标，因为如果实验组和对照组的用户 id 变化很大，而实验组和对照组的点击 cookie 数量又可能不同，测试带来的变化就不一定是因为试验本身而带来的效果了，即不能准确评估实验效果。

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。

(1) Gross conversion

标准偏差: 0.0202

Gross conversion 以 cookie 数量为分母, 也是转移的单位, 转移单位等于分析单位, 所以分析估计与经验变异类似

(2) Retention

标准偏差: 0.0549

Retention 以“登录用户数”为分母, 与转移单位 cookie 不相似, 分析单位和转移单位不相似, 所以分析估计与经验变异不同

(3) Net conversion

标准偏差: 0.0156

Net conversion 是以 cookie 数量作为分母, 也是转移的单位, 转移单位等于分析单位, 所以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。)

不使用 Bonferroni 校正, 因为试验中 Gross conversion 和 Net conversion 不是独立, 只是相关关系, 使用 Bonferroni 校正会使得实验结果过于保守, 可能导致实验效果不太显著。

所需页面浏览量: 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。)

曝光比例: 0.9

持续时间: 20

这里, 我选择 0.9, 只是因为周期相对较短

相对于曝光比例 0.2, 需要 86 天,

曝光比例 0.5 需要 35 天,

曝光比例 0.8 需要 22 天。

曝光比例 0.9 需要 20 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?

原因:

(1) 周期不能过长

只有 20 天, 较短

(2) 风险与成本不大

1. 对于用户

只是询问用户每周所化学习时间, 为非敏感数据, 不会对用户的身心造成影响, 也不涉及道

德伦理问题。

2.对于优达学城网站
不涉及网站及后台，数据库的架构等关键节点，运营成本不高

试验分析

合理性检查

对于每个不变指标，对你在 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）

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。**

	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"/>
• Retention				<input type="checkbox"/>
• Net conversion				<input type="checkbox"/>

以上是我直接截取小测试答案中的图（下同），由于以上三个观察值都介于最小值和最大值之间，故以上三个指标均通过合理性检查。

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。）

Effect Size Tests

For each of your evaluation metrics, compute a confidence interval around the difference.

Did you use the Bonferroni correction? ☐ Yes ☒ No

	Lower bound	Upper bound	Statistical significance	Practical significance
• Number of cookies	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Number of user-ids	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Number of clicks on "Start free trial"	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Click-through-probability on "Start free trial"	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Gross conversion	-0.0291	-0.0120	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
• Retention	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Net conversion	-0.0116	0.0019	<input type="checkbox"/>	<input type="checkbox"/>

这里, Gross conversion 的置信区间不包括 0, 故具有统计显著性, 而其不包含 $d_{\min}(\pm 0.01)$, 故具有实际显著性;

而 Net conversion 的置信区间包含 0, 故不具有统计显著性, 包含了 $d_{\min}(\pm 0.0075)$, 故不具有实际显著性。

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。)

Sign Tests

Run a sign test on each of your evaluation metrics using the day-by-data data. Enter each p-value, and indicate whether each result is statistically significant.

Did you use the Bonferroni correction? ☐ Yes ☒ No

	p-value	Statistical significance
• Number of cookies	<input type="text"/>	<input type="checkbox"/>
• Number of user-ids	<input type="text"/>	<input type="checkbox"/>
• Number of clicks on "Start free trial"	<input type="text"/>	<input type="checkbox"/>
• Click-through-probability on "Start free trial"	<input type="text"/>	<input type="checkbox"/>
• Gross conversion	0.0026	<input checked="" type="checkbox"/>
• Retention	<input type="text"/>	<input type="checkbox"/>
• Net conversion	0.6776	<input type="checkbox"/>

Gross conversion, 双尾 p 值为 0.0026, 小于 α 水平 0.025, 故具有统计显著性

Net conversion, 双尾 p 值为 0.6776, 大于 α 水平 0.025, 故不具有统计显著性

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。

没使用 Bonferroni 校正，因为试验中 Gross conversion 和 Net conversion 不是独立，只是相关关系，使用 Bonferroni 校正会使得实验结果过于保守，可能导致实验效果不太显著。

建议

提供建议并简要说明你的理由。

建议：不启动试验

因为总转化率具有统计和实际显著性，且为负，说明试验会减少因为没有足够时间而离开免费试学，并因此受挫的学生数量，这个符合试验预期。

但是净转化率的置信区间包含负数，区间为 $(-0.0116, 0.0019)$ ，净转化率有更大的概率会落入 $(-0.0116, 0)$ 区间，也就是净转化率有更大的概率会为负，并且有一定的概率使得净转化率的减少超过实际显著性 0.0075。故无法说明“降低的程度不大”。所以不建议启动试验。

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。

试验概述：

现在 Udacity（中国）对于纳米学位项目，已通过微信建群来联络，就我目前所知分为两类。

一类是未付费群，这里是接纳有意向购入优达学城某个课程的新人群。

一类是付费群，这里是工作人员，还有付费学员汇聚的群，负责解答学员的课程和非课程问题，但有时也会遇到提问过了好久才有人解答这类问题的情况。

前者中的部分新人会流向付费群。已经完成了一个筛选分流。

但是新人是如何加入微信群呢？

他需要在了解到这个微信群的二维码位置，扫描加入。

如果他没发现呢？而优达学城目前并没有一个人工引导过程，网站没有在线客服，对于一个新来尝试的人，他注册了账号登录，可能他没有发现这个微信群二维码位置，但也无人咨询，也就看看，关闭了主页。也就是说，缺少一个在线即时反馈机制。

同时，对于，付费用户，学习中遇到问题，直接在线询问人工客服，问题也可即时解决。

假设：

假设人工客服，能够更为即时解答学员的问题，可以提高学员的最终付费率。

度量选择：

1. 不变度量

用户 id 数量：由于此试验是在用户点击试学，并且登录 id 后，不会影响到登录用户数量，这里可选择用户 id 作为不变度量。

2. 评估度量

留存率：试验可能会影响到最终付费用户数量，留存率可以作为评估指标。

转移单位：

用户 id：因为此测试发生在用户登录后，id 会被跟踪，故用户 id 是合适的转移单位。