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。

1. 不变指标：

（1） Number of cookies（Cookie 的数量）：

cookie 是分组单元，必须是不变指标，而且测试的是首页的子页面，不会影响首页的情况，也就不会影响 cookie 数量；

期望结果：不变；

（2） Number of clicks（点击次数）：

首页没有任何改变，cookie 不变，点击次数也不变；

期望结果：不变；

（3） Click-through-probalility（点进概率）：

点击次数不变，cookie 不变，点进概率也不变；

期望结果：不变

2.评估指标：

（1）Gross conversion（总转化率）：

因为提醒用户每周 5 小时学习时间，可能会影响完成登录并参加免费试学的用户 id 的数量，这也是我们想通过 A/B 测试来研究的问题。因此，是个很好的评估指标。

期望结果：减小；

实验预期是减少因为没有足够的时间而离开免费试学，并因此受挫的学生数量，即：分子会减小，分母不变，因此总转化率期望是减小；

（2）Retention（留存率）：

因为提醒用户每周 5 小时学习时间，会对付费用户 和 完成免费试用的用户的比例产生影响，是个很好的评估指标；

期望结果：增大；

因为根据预期，期望最终通过免费试学和最终完成课程的学生数量不变，而登录的用户 id 数量减少，根据公式，留存率会增大；

（3） **Net conversion（净转化率）：**

为提醒用户每周 5 小时学习时间，会对付费用户产生影响，因此净转化率是个很好的评估指标；

期望结果：不变；

因为根据实验预期，付费用户不变，点击“开始免费试学”按钮的唯一 cookie 数量不变，所以净转化率的期望结果不变；

3.无关指标：

（1）Number of user-ids（用户 id 的数量）：

根据实验设计，用户 id 数量有可能变化，因此不能是不变指标；

同时，也不是好的评估指标，因为如果实验组和控制组的用户 id 变化很大，但是因为实验组和控制组的点击 cookie 数量可能有不同，所以测试出来的变化不一定仅仅是因为试验，因此不能准确地评估试验的效果。

怎么样才能更好地评估呢？我们可以将用户 id 的数量除以点击的 cookie 的数量（进行归一化操作），只比较比例，这样就可以消除 cookie 数量不同的影响了。归一化操作得到的度量也就是总转化率，在总转化率存在的条件下，我们可以不选择用户 id 作为评估度量。

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。

标准偏差根据公式： sqrt(p * (1-p) / N) 计算

（1） Gross conversion（**总转化率**）：

p = 0.20625（给出）

N= 5000*0.08 = 400

Std dev = sqrt（0.20625 * （1-0.20625） / 400） = 0.0202

总转化率是以 cookie 数量作为分母，也是转移的单位。转移单位等于分析单位，标明分析估计值与经验变异类似；

（2） Retention（**留存率**）：

p = 0.53（给出）

N= 5000*0.08*0.20625 = 82.5

Std dev = sqrt（0.53 * （1-0. 53） / 82.5） = 0.0549

留存率是以“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，所以分分析估计值与经验变异不相似；

（3） Net conversion（**净转化率**）：

p = 0.1093125（给出）

N= 5000*0.08 = 400

Std dev = sqrt（0.1093125 * （1-0.1093125） / 400） = 0.0156

净转化率是以 cookie 数量作为分母，也是转移的单位。转移单位等于分析单位，标明分析估计值与经验变异类似；

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；

因为本试验中总转化率和净转化率并非是独立的，而是相关联的，使用 Bonferroni 校正会使试验结果过于保守；

页面浏览量通过在线计算器(http://www.evanmiller.org/ab-testing/sample-size.html)计算样本量（alpha = 0.05，beta = 0.2），再转化为实验组的页面浏览量，再乘以 2 得出总的页面浏览量：

（1） Gross conversion（**总转化率**）：

参数：baseline conversion rate:20.625%(给出),

Minimum detectable effect:1%(给出)，

样本数量：25835

实验组页面浏览量：25835/0.08 =322938

总页面浏览量：322938 * 2 = 645875

（2） Retention（留存率）：

参数：baseline conversion rate:53%(给出),

Minimum detectable effect:1%(给出)，

样本数量：39115

实验组页面浏览量：39115/0.20625/0.08 =2370606

总页面浏览量：2370606 * 2 = 4741212

（3） Net conversion（净转化率）：

参数：baseline conversion rate:10.93%(给出),

Minimum detectable effect:0.75%(给出)，

样本数量：27413

实验组页面浏览量：27413/0.08 =342663

总页面浏览量：342663 * 2 = 685325

取其中较大值，但是留存率算出来需要的页面浏览量 474 万，相对于每天 4 万的页面浏览量太大了，所以舍弃掉 留存率这个指标；

在总转化率和净转化率算出的取较大值，最后使用 净转化率需要的页面浏览量 **685325**；

持续时间和暴光比例

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。）

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？

曝光比例是 0.8；

因为，曝光流量比例主要是根据对实验的风险容忍度决定的，还要考虑到实验的周期不能太长；

从风险上来说，

（1） 本试验只是询问用户每周能投入多少时间学习，不会对用户的身体，心理等造成不良影响，不涉及道德伦理问题；

（2） 收集投入学习时间的数据，不具有个人识别性的信息，不是敏感数据；

（3） 对网站来说，不涉及网站及后台，数据库的架构等关键节点，对数据库安全没有影响；

因此试验风险较小，可以考虑给出 50%--100%的流量。

从实验周期上来说，每天页面总流量是 4 万，需要的总流量是 68.5 万，分别计算需要的时间，50%流量需要 35 天，100%流量需要 18 天；

综合考虑风险容忍度和试验周期，我选择了曝光 80%的流量，持续时间 22 天的方案；

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。**

（1） Number of cookies（Cookie 的数量）：

控制组页面总量：345543

实验组页面总量：344660

页面总量：690203

Cookie 分布概率：0.5

SE = sqrt(0.5*(1-0.5)/(345543+344660)) = 0.0006018

m = SE * 1.96 = 0.0011796

置信区间=[0.5-m, 0.5+m] = [0.4988, 0.5012]

观察值 = 344660 / 690203 = 0.5006

通过合理性检查

（2） Number of clicks（点击次数）：

控制组总量：28378

实验组总量：28325

总量：56703

Cookie 分布概率：0.5

SE = sqrt(0.5*(1-0.5)/(28378 +28325)) = 0.0021

m = SE * 1.96 = 0.0041

置信区间=[0.5-m, 0.5+m] = [0.4959, 0.5041]

观察值 = 28378 / 56703 = 0.5005

通过合理性检查

（3） Click-through-probalility（点进概率）：

控制组概率：0.0821258

SE = sqrt(0.0821258 * (1-0.0821258) / 344660) = 0.000468

m = SE * 1.96 = 0.00092

置信区间=[0.0821258-m, 0.0821258+m] = [0.0812, 0.0830]

观察值 = 0.0821824

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。）

（1） Gross conversion（**总转化率**）：

	Control 控制组	Experiment 实验组
Clicks 点击	17293	17260
Enrolment 登录	3785	3423

Gross conversion（**总转化率**）	0.2189	0.1983

Pooled Probability=（3785+3423）/ (17293+17260) = 0.2086

SE = sqrt(0.2086 * (1-0.2086) / (1/17293 + 1/17260) = 0.004372

m = SE * 1.96 = 0.008568

d = 3423/17260 - 3785/17293 = -0.02055

置信区间=[-0.02055-m, -0.02055+m] = [-0.0291, -0.0120]

置信区间不包括 0，具有统计显著性；

置信区间不包含 d_min，具有实际显著性

（2） Net conversion（**净转化率**）：

	Control 控制组	Experiment 实验组
Clicks 点击	17293	17260
Payment 付费	2033	1945
Gross conversion（**总转化率**）	0.1176	0.1127

Pooled Probability=（2033+1945）/ (17293+17260) = 0.1151

SE = sqrt(0.1151 * (1 - 0.1151) / (1/17293 + 1/17260) = 0.003434

m = SE * 1.96 = 0.006731

d = 1945/17260 - 2033/17293 = -0.004874

置信区间=[-0.004874-m, -0.004874+m] = [-0.01160, 0.001857]

置信区间包括 0，不具有统计显著性；

置信区间包含 d_min（+/- 0.0075），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。）

利用在线计算器计算：http://graphpad.com/quickcalcs/binomial1.cfm

（1） Gross conversion（**总转化率**）：

成功数量：4

试验次数：23

概率：0.5

双尾 P 值：0.0026

双尾 P 值 0.0026 小于 alpha 水平 0.025，具有统计显著性；

（2） Net conversion（**净转化率**）：

成功数量：10

试验次数：23

概率：0.5

双尾 P 值：0.6776

双尾 P 值 0.6776 大于 alpha 水平 0.025，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。

没有使用 Bonferroni 校正，因为本试验中的总转化率和净转化率不是独立的，是高度关联的，使用会使得实验结果过于保守；

建议

提供建议并简要说明你的理由。

不建议启动试验；

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的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，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)，也就是说有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性 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。

试验概述：

说一下我自己的亲身体验，从我开始知道优达学城，到付费报名数据分析师，中间有大概 6 个月的时间，期间有体验过纳米学位试学，但当时有些内容根本看不懂，编程之前也没有学习过，很懵，也不知道该怎么学，学习路径怎样，学习后的就业问题等等，都不清楚。

试用到期后我也就没有付费，后来就开始自己摸索这学习编程，找就业方面的资料等，直到我看了一个数据分析师毕业学员的分享，才真正解答了我的疑惑，也就报名正式开始学习了。

我总结下来，问题是，试学时，在课程中遇到问题需要解答的时候，需要到论坛发帖（我习惯了淘宝式的即时客服，对于论坛发帖，还要等好久才能知道结果，我是拒绝的），没有方便的即时沟通方式，这会大大影响到用户体验和付费转化。

此实验，可以测试一项变化，在第一次开始学习纳米课程时，提示"已经为您配置了专属学习导师，是否有问题咨询？"的提示框，点击"咨询"，立即与课程导师建立在线对话，有问题可以随时问。同时在纳米学位课程学习页面的右侧有"咨询导师"的按钮，点击直接在线咨询导师；

假设：我假设这会即时解答学员的问题，个性化提供学习建议和设计学习方案，提高学员最终付费率。

度量选择：

1. 不变度量：

（1）用户 id 的数量：

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，因此选择用户 id 做不变度量；

2.评估度量：

（1）留存率；

试验可能会影响最终付费用户数量，留存率是个很好的评估指标；

转移单位：

（1）用户 id：

此测试发生在用户登录后，id 会被跟踪，用户 id 是合适的转移单位；

优达学城

2016 年 9 月