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

不变指标如下：

1. Number of cookies：试验不影响登录网站的cookie数量
2. Number of clicks：首页未改变，点击发生在试验改变部分的前面。
3. Click-through-probalility：Number of cookies和Number of clicks不变，点击概率也不会改变。

评估指标如下：

1. Gross conversion：因试验改变的部分是如果学生能投入的时间小于每周5小时，则弹出消息说明需要更多的投入时间，并建议访问免费的课程资料。这个改变会影响到总的转化率，因此可以作为评估指标。

期望结果是减少，因期望能投入的时间小于每周5小时学生去浏览免费的课程资料，报名参加免费试用的用户id的数量会减少，而总的点击“开始免费试用”按钮的唯一 cookie的数量不变。

1. Retention：因试验的改变期望能投入的时间小于每周5小时学生去浏览免费的课程资料，会影响留存率，因此可以作为评估指标。

期望结果是增加，因期望能投入的时间小于每周5小时学生去浏览免费的课程资料，完成登录的用户id的数量会减少，而期望最终通过免费试学并最终完成课程的学生数量不变，所以结果是增加。

1. Net conversion：需要验证试验是否会影响到净转化率，因此可以作为评估指标。

期望结果是不变。根据试验预期，付费用户不变，点击“开始免费试学”按钮的唯一cookie数量不变，因此结果不变。

用户id数量：

它发生于试验之后，会受到试验的影响，因此它是一个ok的评估度量。但由于实验组和对照组的cookie数量不一定相同，也就是说两组中用户ID数量不同可能是由于实验的影响，也可能是由于两组cookie的不同。所以使用用户ID数量的区别不能够很好的评估试验的效果。在一个比例化的评估度量（总转化率）存在的情况下，我们可以不选择用户ID的数量作为评估度量。

测量标准偏差

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： 0.0202

p = 0.20625，N= 5000\*0.08 = 400，Standard deviation = sqrt(0.20625 \* (1-0.20625) / 400) = 0.0202

Gross conversion是以cookie数量作为分母，也是试验的转移单位。分析单位等于转移单位，表明分析估计与经验变异类似。

1. Retention: 0.0549

P = 0.53，N= 5000\*0.08\*0.20625 = 82.5，Standard deviation = sqrt(0.53 \*(1-0. 53)/ 82.5) = 0.0549

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

1. Net conversion: 0.0156

p = 0.1093125，N= 5000\*0.08 = 400，Standard deviation = sqrt(0.1093125 \* (1-0.1093125)/ 400) = 0.0156

Net conversion是以cookie数量作为分母，也是试验的转移单位。分析单位等于转移单位，表明分析估计与经验变异类似；

规模

样本数量和功效

不使用Bonferroni校正；

通过在线计算器http://www.evanmiller.org/ab-testing/sample-size.html计算：

1. Gross conversion: 645875

baseline conversion rate:20.625%, Minimum detectable effect:1%, Sample size: 25835

实验组页面浏览量：25835/0.08 =322937.5

总页面浏览量：322937.5 \* 2 = 645875

1. Retention：

baseline conversion rate:53%, Minimum detectable effect:1%, Sample size: 39115

实验组页面浏览量：39115/0.20625/0.08 =2370606

总页面浏览量：2370606 \* 2 = 4741212

1. Net conversion:

baseline conversion rate:10.93125%, Minimum detectable effect:0.75%, Sample size: 27413

实验组页面浏览量：27413/0.08 =342662.5

总页面浏览量：342662.5 \* 2 = 685325

所以，取其中的最大值，为4741212。

持续时间和暴光比例

Number of pageviews: 由于4741212太大，这里选择小一点的总转化率需要的浏览量685325。

风险分析：该试验并没有太大的风险

* 1. 即使学生每周学不到五小时，他们只是被页面的变更提醒引导到了另外的一个页面，如果今后有需要学生仍然可以进入免费试学、登录并继续完成课程，不会因此而影响用户使用网站的习惯；
  2. 没有在页面展示上有过大的改动，不会对用户产生感情上的冲击，用户也不需要花长时间去适应页面的改变；
  3. 该试验没有关于数据库及后台的改变，不用担心数据的丢失及由于后台的失误导致网页崩溃、用户无法访问网页等大问题；
  4. 此试验也不会对用户的个人信息安全造成风险，因为不论网页是否增加了提醒，用户在确认参加免费试学时都得输入信用卡信息，而很明显系统一定会保护用户的个人信息；
  5. 该试验同样也没有道德上的风险。

Fraction of traffic exposed: 选择0.8。因经过上面的风险分析，该试验并没有太大的风险。考虑到实验周期，选择0.8。

Length of experiment: 685325/(40000\*0.8) = 21.42，取整为22天。

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：

控制组Pageviews总量：345543

实验组Pageviews总量：344660

Pageviews总量：345543+344660=690203

Cookie分布概率：0.5

SE = sqrt(0.5\*(1-0.5)/690203) = 0.0006018

m = SE \* 1.96 = 0.0011795

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

观察值 = 345543/ 690203 = 0.5006

因此通过合理性检查。

1. Number of clicks：

控制组clicks总量：28378

实验组clicks总量：28325

clicks总量：28378+28325=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

因此通过合理性检查。

1. Click-through-probalility:

控制组概率：28378/345543 = 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 = 28325/344660 = 0.0822

因此通过合理性检查。

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。）

不使用Bonferroni校正

1. Gross conversion：

控制组Clicks: 17293

实验组Clicks: 17260

控制组Enrolment：3785

实验组Enrolment：3423

控制组Gross conversion：0.2189

实验组Gross conversion：0.1983

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

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

m = SE \* 1.96 = 0.008569

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

1. Net conversion：

控制组Clicks: 17293

实验组Clicks: 17260

控制组Payment：2033

实验组Payment：1945

控制组Gross conversion：0.1176

实验组Gross conversion：0.1127

总的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。）

不使用Bonferroni校正

通过在线计算器计算https://www.graphpad.com/quickcalcs/binomial1.cfm

1. Gross conversion：

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；

1. Net conversion：

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。

不使用Bonferroni校正，因本试验中的Gross conversion和Net conversion不是独立的，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Gross conversion具有统计和实际显著性，且为负，说明试验会减少完成登录并报名参加免费试用的学生数量。

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。

为了减少同学提前终止学习，可在鼓励同学参加一对一辅导。在课程页面增加与导师对话的按钮，当同学开始免费课程时，系统就自动为每个同学分配一个导师，同学可在课程页面与导师对话，咨询问题或接受一对一辅导。

假设：

假设增加该功能后，会减少提前终止学习的同学，增加留存率。

度量选择：

不变度量：

Number of cookies：试验不影响登录网站的cookie数量

评估度量：

留存率：试验会减少提前终止学习的同学，让更多的同学完成免费课程并参加付费课程，留存率是很好的评估指标。

转移单位：

user id：

此测试发生在用户登录后，会记录用户的user id，user id是合适的转移单位。

优达学城

2016年9月