A/B 测试项目

试验概述：免费试学筛选器

在进行此试验时，优达学城当前的主页上有两个选项：“开始免费试学”和“访问课程资料”。如果学生点击“开始免费试学”，系统将要求他们输入信用卡信息，然后他们将进入付费课程版本的免费试学。 14 天后，将对他们自动收费，除非他们在此期限结束前取消试用。若学生点击“访问课程材料”，他们将能够观看视频和免费进行小测试，但是他们不会获得导师指导或认证证书，无法提交最终项目来获取反馈。在此试验中，优达学城测试了一项变化，如果学生点击“开始免费试学”，系统会问他们有多少。时间投入到这个课程中。如果学生表示每周 5 小时或更多，将按常规程序进行登录。如果他们表示一周不到 5 小时，将出现一条消息说明优达学城的课程通常需要更多的时间投入才能成功完成，并建议学生可免费访问课程资料。在这里，学生可选择继续进行免费试学，或免费访问课程资料。 这张截图 展示了试验概况。我们假设这会为学生预先设定明确的期望，从而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，同时不会在很大程度上减少继续通过免费试学和最终完成课程的学生数量。如果这个假设最后为真，优达学城将改进整体的学生体验和提升导师为能够完成课程的学生提供帮助支持的能力。分组单位为 cookie ，尽管学生参加的是免费试学，但在登录后他们的用户 id 便被跟踪。同一个用户 id 不能两次参加免费试学。对于不参加免费试学的用户，他们的用户 id 不会在试验中被跟踪，即使他们在访问课程概述页面时登录了网站。

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. Cookie 的数量(Number of cookies):
   Cookie 是分组单元，所以一定是不变指标。不会影响到首页的情况，所以不会影响 cookie 数量。期望结果：不变
2. 点击次数(Number of clicks)
   首页没有改变，并且在免费试学筛选器出发前发生，cookie 不变，点击次数也不会改变。期望结果：不变
3. 点击概率(Click-through-probability)
   即点击“开始免费试学”按钮的唯一 cookie 数量/查看课程概述页面的唯一 cookie 数量所得比率。在免费嗜血筛选器出发前发生，点击次数不变，cookie 数量不变，点击概率也不变。期望结果：不变

评估指标：

1. 总转化率(Gross conversion):
   即完成登陆并参加免费试学的用户 id 的数量/点击“开始免费试学”按钮的唯一 cookie 的数量所得的比率。因为提醒用户每周需要最少 5 小时的学习时间，这会影响完成

登陆并参加免费试学的用户 id 的数量，数量可能会减少。实验预期是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的数量。分子变小，分析单元不变。

期望结果：减小

2. 留存率(Retention):

即在 14 天的 期限过后仍参加课程的用户 id 数量/完成登陆的用户 id 数量。因为提醒用户每周 5 小时学习的时间，实验预期是通过筛选器达到尽可能少的流失免费试学期限过后仍参加课程的用户 id 数量，所以会对该比率产生影响。

期望结果：增大

3. 净转化率(Net conversion):

即在 14 天的期限后仍参与课程的用户 id 的数量/点击了“开始免费试学”按钮的唯一 cookie 的数量的比率。实验预期 14 天的期限后仍参与的付费的用户不变。分母不变。

期望结果：不变

无关指标:

用户 id:

用户 id 注册发生于试验之后，会收到试验影响，可以作为一个指标，但是实验组和对照组的点击 cookie 数量可能不同，所以无法确定力量足中用户 id 不同是由于实验影像还是由于两组 cookie 数量不用造成的。所以使用用户 id 不能很好的评估实验效果，所以不考虑。

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. **总转化率：**

P=0.20625

N = 5000*0.08 = 400

Std dev = =SQRT((0.20625*(1-0.20625))/3200) = 0.0202

总转化率以 cookie 数量作为分母是分析单位，也是分组单元分组单元和分析单元相同，分析估计值与经验差异性类似。

2. **留存率：**

P=0.53

N = 5000*(660/40000) = 82.5

Std dev = =SQRT((0.53*(1- 0.53))/82.5 = 0.0549

留存率是以登录用户 id 作为分析单位，不变指标 cookie 作为分流单元，所以分析估计值与经验变异性不相似。

3. **净转化率：**

p = 0.1093125

N = 400

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

净转化率以 cookie 数量作为分母，不变指标 cookie 作为分流单元，分析估计值与经验变异性相似。

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。）

此测试中的总转化率和净转化率非独立，是相关联的，使用 Bonferroni 校正会使得实验结果过于保守，不考虑使用。

通过在线计算器 http://www.evanmiller.org/ab-testing/sample-size.html 计算样本量。在线计算器算出的结果是满足条件所需要的分析单位的个数，即在总转化率和净转化率中计算出的是 cookie 数量。留存率计算出的是完成登录的用户 id 数量，所以还需转化成 cookie 数量再转化为页面浏览量。

所需 cookie 数量 = 页面浏览量*0.08 (CTP)

测试已给出 alpha = 0.05, beta = 0.2

1. **总转化率:**
   d_{min} = 0.01
   样本数量 = 25,835
   页面浏览量 = (25835/0.08)*2 = 645875
2. 留存率：
   d_{min} = 0.01
   样本数量 = 39,115
   页面浏览量 = (39115/0.20625/0.08)*2 = 4741212
3. **净转化率:**
   d_{min} = 0.0075
   样本数量 = 27,413
   页面浏览量 = (27,413/0.08)*2 = 685325

留存率算出来需要页面浏览量为 4741212，数值太大，相当于每天 4 弯浏览量的话是要需要进行 119 天，时间较长，不考虑留存率。从总转化率和净转化率中取其中较大值 685325，所以使用净转化率需要的页面浏览量为 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？

风险分析:

1. 此实验不会对参与者的身体、心理、情感、社会和经济方面造成影响，基本不涉及道德伦理问题。
2. 试验以 cookie 分流， 匿名数据不携带用户身份信息，不属于敏感数据。
3. 对网站来说，不涉及网站后台、数据库的架构等关键点，对数据库的安全遭到威胁的风险较低。

按经验分析，此实验风险较小，可以给出 100%的流量。

根据给出的数据，每天浏览页面的唯一 cookie 量是 40000，所以总量是 645875，

645875/40000 ≈ 18 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。**

Ncont = 345543　　Nexp = 344660

Xcont = 28378　　Xexpt = 28325

1. Cookie 数量:

SE_{pool} = =SQRT(0.5*0.5/(345543+344660)) = 0.0006

m = 1.96 * 0.0006 = 0.0012

CI = (0.4988, 0.5012)

观察值 = 345543/(345543+28378) = 0.5006

在置信范围内，通过检查

2. 点击量:

SE_{pool} = SQRT(0.5*0.5/(28378+28325)) = 0.0021

m = 1.96 * 0.0021 = 0.0041

CI = (0.4959, 0.5041)

观察值 = 28378/(28378+ 28325) = 0.5005

在置信范围内，通过检查

3. CTP:

Pcont = 28378/ 345543 = 0.0821

Pexp = 28325/ 344660 = 0.0822

SE = SQRT(0.0821*(1-0.0821)*(1/345543)) = 0.0005

m = 1.96 * 0.0005 = 0.0009

CI = (0.0812, 0.0830)

观察值为实验组的 P 值 = 0.0822

在置信区间内，通过检查

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。）

	点击量	入学	付款
对照组	17293	3785	2033
实验组	17260	3423	1945

1. 总转换率:

对照组 = 3785/17293 = 0.2189

实验组 = 3423/17260 = 0.1983

d =0.1983- 0.2189 = -0.0206

P_{pool} = (3785 + 3423) / (28378 + 28325) = 0.2086

SE_{pool} = SQRT(0.2086*(1-0.2086)*(1/17293+1/17260)) = 0.0044

m = 1.96 * 0.0044 = 0.0086

CI = -0.0206 ± 0.0086 = (-0.0291, -0.0120)

$(-d_{min}, d_{min})$ = (-0.01, 0.01)

置信区间不包含 0，可以确信它是会变的，这个指标具有统计显著性。置信区间不包含实际显著性边界，所以可以确信变化是有用的，具有实际显著性。

2. 净转化率:

对照组 = 2033/17293 = 0.1176

实验组 = 1945/17260 = 0.1127

d = 0.1127 - 0.1176 = -0.0049

P_{pool} = (2033 + 1945)/(17293+ 17260) = 0.1151

SE_{pool} = SQRT(0.1151*(1-0.1151)*(1/17293+1/17260)) = 0.0034

m = 1.96 * 0.0034 = 0.0067

CI = -0.0049 ± 0.0067 = (-0.0116, 0.0019)

$(-d_{min}, d_{min})$ = (-0.0075, 0.0075)

置信区间包含 0，确信它是不会变的，这个指标不具有统计显著性。置信区间包含实际实际显著性边界，所以不可以确信变化是有用的， 置信区间含有负数部分，说明净转化率有一定的几率会下降，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。）

该检验需要使用在线计算器 https://www.graphpad.com/quickcalcs/binomial2/

1. **总转化率:**

   实验次数: 23

   成功次数(总转换率 $_{exp}$ – 总转换率 $_{cont}$ > 0) = 4

   概率 = 0.5

   双尾 P 值 = 0.0026

   P 值小于 alpa 水平 0.025 概率， 统计学显著。

2. **净转化率:**

   实验次数: 23

   成功次数(总转换率 exp – 总转换率 cont > 0) = 10

   概率 = 0.5

   双尾 P 值 = 0.6776

   P 值大于 alpa 水平 0.025，统计学不显著。

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。

效应大小假设检验和符号检验之间不存在差异。

建议

不建议启动试验
因为总转化率具有统计和实际显著性，并且< 0，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。
净转化率的置信区间含有负数部分，说明净转化率有一定的几率会减少并且会超过实际显著性的 0.0075。
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。

优达学城是一个教育机构。DAND 入门项目是的门槛低不等于没有要求，对于想进入项目学习的同学，应该对其标注需要进入项目的要求，如果达不到，建议学前辅导达到入门要求。
假设：
对想进入项目学习的学生设立要求，对于达不到入门要求的做出提醒，期望减少因为在试学期间因为能力不够学习起来困难而退出的学生。

指标选择:

1. 不变指标:
   用户 id:
   用户试学登陆后，相互独立且唯一，不会影响到其他用户 id。

2. 评估指标:
   a. 留存率
      试验旨在测试最终留下继续学习的用户概率，所以留存率可以成为评估指标。
1. 转移单位:
   a. 用户 id:
      测试发生在用户登陆并且发生付费行为后，所以 id 可以作为转移单位。

参考资料：
https://discussions.youdaxue.com/t/ab/55953
https://www.graphpad.com/quickcalcs/binomial2/
http://www.evanmiller.org/ab-testing/sample-size.html
https://baike.baidu.com/item/Bonferroni%E6%A0%A1%E6%AD%A3/8646271?fr=aladdin
https://discussions.youdaxue.com/t/47-7-8-9/42329/2
https://discussions.youdaxue.com/t/47-5/43240/6