

A/B 测试项目

一、试验设计

1、指标选择

(1) 不变指标：

- cookie数量：因为cookie数量是实验前随机分配给对照组和实验组，cookie数量不会因为实验二产生变化，所以可以用来做不变指标；
- 点击次数：因为实验的变量是点击“开始免费试用”按钮后的信息，而点击次数的统计是在免费试用屏幕触犯前发生，所以点击次数并不受实验影响，可以作为不变指标；
- 点进概率：点进概率=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；而点击“开始免费试用”按钮的唯一 cookie 的统计和查看课程概述页面的唯一 cookie 的统计都发生在免费试用屏幕触犯前发生，这两个指标都不受实验影响，所以点击概率可以作为不变指标；

(2) 评估指标：

- 总转化率：“开始免费试用”的提示信息会影响用户是否继续参加，完成登录并报名参加免费试用的用户 id 的数量是受实验影响的，所以总转化率可以用来做评估指标；
- 留存率：用户只有参加免费试用后，才有可能进行支付，实验会影响用户是否参加免费试学，所以同样也会影响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，留存率可以作为评估指标；
- 净转化率：同上，实验会影响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，净转化率可以作为评估指标；

(3) 希望取得的结果：

- 总转化率：我们预期实验组中报名参加免费试用的用户id的数量少于对照组，两组的cookie数量不变，我们希望看到实验组的总转化率相对于对照组有显著下降；
- 留存率：我们预期该试验不会很大程度减少通过免费试学并进行支付的学生数量，但会减少参加免费试学的用户ID的数量，实验组的留存率相对于对照组，分子不变，分母变小，我们希望看到实验组的留存率相对于对照组有显著提高；
- 净转化率：我们预期该试验不会很大程度减少通过免费试学并进行支付的学生数量，且实验组和对对照组分配的cookie数量一致，实验组的净转化率相对于对照组，分子不变，分母不变，我们希望看到实验组的净转化率相对于对照组并无显著差异；又因为我们对最终通过免费试学并进行支付的学生数量

的预期是“不会有显著的减少”，因此净转化率的置信区间可以包含0，但不可以包含负数；

在给定的指标中，用户ID数量既不能做不变指标，又不能做评估指标，原因如下：用户ID数量是指报名参加免费试用的用户数量，该统计行为发生在点击“开始免费试用”按钮之后，受实验变量的影响，所以不能用来做不变指标；另一方面，本次实验的分组是随机分配，对照组和实验组分配到的cookie数量不完全一致，两组的用户ID数量不一致可能是实验影响，也可能是受两组分配的cookie数量不一致影响，所以用户ID数量可以是评估指标，但不是一个好的评估指标

2、测量标准偏差

- 总转化率的标准偏差：0.0202
- 留存率的标准偏差：0.0549
- 净转化率的标准偏差：0.0156

对于总转化率和净转化率两个评估指标，其分组单元和分析单元都是唯一cookie数量，所以我认为这两个评估指标的分析估计与经验变异是类似的；对于留存率评估指标，分析单元为用户ID的数量，分组单元为唯一cookie数量，所以我估计该指标的经验差异会远远高于分析估计得出的差异值；

3、规模

- (1) 是否会在分析阶段使用bonferroni校正：由于总转化率、留存率、净转化率3个评估指标的都与参加免费试学的用户数量有关，关联性极高，如果采用bonferroni校正就会过于保守，所以无需进行bonferroni校正；
- (2) 根据样本容量选择最终评估指标：留存率指标的分析单位为完成登录的用户ID数量（注册量），我们通过样本容量计算器计算出留存率指标所需样本容量为39115，40000的页面浏览量可以产生660的完成登录的用户ID样本，那么如果要产生39115的完成登录的用户ID样本，就需要4741212的页面浏览量；按照每天产生的40000条页面浏览量来计算，就算将其全部投入实验，实验时间也要持续119天；所以使用留存率作为评估指标会导致实验持续时间过长，留存率不适合作为本实验的评估指标；
- (3) 样本数量：685325；我们通过样本容量计算器来估计样本容量，由于我们

最终选取总转化率和净转化率两个指标作为评估指标，我们先计算这两个指标所需要的点击开始按钮的唯一cookie样本容量，总转化率指标所需样本容量为25835.净转化率指标所需样本容量为27413，所以我们选取较大的样本容量（27413）便可满足两个指标的样本量需求；又因为我们有对照组和实验组，所以所需总样本容量为 $27413 \times 2 = 54826$ ，40000的页面浏览量可以产生3200的点击开始按钮的唯一cookie样本，那么如果要产生54826的点击开始按钮的唯一cookie样本，就需要685325的页面浏览量；

4、持续时间和曝光比例

(1) 实验风险评估：

- 道德伦理：该试验旨在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；这对学生来说只是提前打了预防针，并没有造成伤害，所以该试验在道德伦理方面并无风险；
- 好处：试验本身就是在提前给学生设定预期，让学生对整个课程的时间有个了解，并对比自己的时间进行衡量，降低了因进入免费试学后因时间不足而退出课程的概率，对学生来说，是有一定的好处的；
- 数据安全和数据敏感性：该试验的变更部分（点击”开始免费试学“按钮后的提示）并没有收集任何敏感数据，所以该试验不存在数据敏感性风险

(2) 曝光比例：我会将100%的页面浏览量转入此试验，因为该试验对优达学城的用户并没有什么风险，所以为了缩短试验持续时间，可以将100%的页面浏览量转入此试验；

(3) 持续时间：整个试验需要685325的页面浏览量，而优达学城每天可以产生40000条页面浏览量，所以整个试验持续时间是18天；

二、试验分析

1、合理性检查

(1) 唯一cookie数量指标：我们期望分配到两组的cookie数量相同，也就是每组的分配比例为0.5，计算得出期望的分配比例的95%的置信区间为[0.4988,0.5012]，实际上控制组的唯一cookie数量的分配比例为0.4994，并无实际显著差异，该指标通过完整性检查；

(2) “开始免费试用”点击数指标：我们期望两组的点击数量相同，也就是每组点击数量所占比例为0.5，计算得出期望的点击数所占比例的95%的置信区间为[0.4959,0.5041]，实际上控制组的“开始免费试用”点击次数的分配比例为0.4995，并无实际显著差异，该指标通过完整性检查；

- (3) “开始免费试用”的点入概率：我们期望对照组和实验组的“开始免费试用”点入概率一致，也就是两组的点入概率之差为0，我们计算得出期望的两组点入概率只差的95%的置信区间为 $[-0.0013, 0.0013]$ ，实际上对照组和实验组的点入概率只差为0.000056627，并无实际显著差异，该指标通过完整性检查；

三、结果分析

1、效应大小检验

- (1) 总转化率指标：根据实验组和对照组的总转化率的差值，我们计算出两组之间的总转化率的差值的95%的置信区间为 $[-0.0291, -0.0120]$ ，该置信区间不包含0，所以改评估指标具有统计显著性；同样，该区间与实际显著性边界 $[-0.01, 0.01]$ 没有交集，所以该评估指标具有实际显著性；
- (2) 净转化率指标：根据实验组和对照组的净转化率的差值，我们计算出两组之间的净转化率的差值的95%的置信区间为 $[-0.0116, 0.0018]$ ，该置信区间包含0，所以该评估指标不具有统计显著性；同样，该区间与实际显著性最小区间 $[-0.0075, 0.0075]$ 有交集，所以该评估指标不具有实际显著性；

2、符号检验

- (1) 总转化率指标：通过符号检验所得双尾P值为0.0026，小于给定的95%置信水平的 α 值0.05，表明该结果的出现不具有随机性，所以对于总转化率指标来说，符号检验和效应检验的结果一致，都具有统计显著性；
- (2) 净转化率指标：通过符号检验所得双尾P值为0.6776，大于给定的95%置信水平的 α 值0.05，表明该结果的出现具有随机性，所以对于净转化率指标来说，符号检验和效应检验的结果一致，都不具有统计显著性；

3、汇总

在效应大小检验和符号检验中，我并没有使用bonferroni校正，因为总转化率、净转化率2个评估指标的都与参加免费试学的用户数量有关，关联性极高，如果采用bonferroni校正就会过于保守，所以无需进行bonferroni校正；

四、建议

1. 我们选择的两个评估指标：总转化率和净转化率，其中净转化率指标的检测

结果与我们的预期不符，该指标置信区间为 $[-0.0116, 0.0018]$ ，我们预期实验组的净转化率相对于对照组要不减小，单算出来的置信区间有负值，说明有很大可能性实验组的净转化率要小于对照组的净转化率，并且有一定的概率净转化率的减少会超过实际显著性的最小值0.0075，所以我们无法保证不会很大程度减少通过免费试学并进行支付的学生数量，与实验目的相违背，不建议启动该试验；

2. 为了达到我们实验最初的目的，我认为应该进行后续试验：后续试验测试一项变化点，学生在点击“开始免费试学”按钮后，系统会弹出一张测试问卷（测试题目都为选择题，测试学生的基础能力），学生花3分钟填完后提交，系统根据测试分数估计出该学生完成该课程至少需要的总时间、每个部分需要的时间，将分析结果呈现给学生后，学生可以选择“可以完成”进入免费试学，或者选择“完成有难度”退出免费试学；

五、后续试验

1. **试验内容：**后续试验测试一项变化点，学生在点击“开始免费试学”按钮后，系统会弹出一张测试问卷（测试题目都为选择题，测试学生的基础能力），学生花3分钟填完后提交，系统根据测试分数估计出该学生完成该课程至少需要的总时间、每个部分需要的时间，将分析结果呈现给学生后，学生可以选择“可以完成”进入免费试学，或者选择“完成有难度”退出免费试学；
2. **试验假设：**我们通过提前测试学生对于该课程的基础知识了解的情况，将学生的基础能力转化为完成该课程所需时间（所需时间与基础能力成反比），让学生对自己的能力以及课程耗时有个预期，做出符合自己的选择；如果这个假设最后为真，我们将能够有效减少受挫并想要提前终止课程的学生数量；
3. **测量指标的选择：**
 - (1) 总转化率（即完成登录并参加免费试学的用户 id 的数量除以点击“开始免费试学”按钮的唯一 cookie 的数量所得的比率）：测试问卷会影响学生是否参加免费试学，完成登录并报名参加免费试用的用户 id 的数量是受实验影响的，所以总转化率可以用来做测量指标；
 - (2) 净转换率（即在 14 天的期限后仍参与课程的用户 id 的数量(因此至少进行了一次付费)除以点击了“开始免费试学”按钮的唯一cookie 的数量所得的比率）：测试问卷会影响学生是否参加免费试学，所以也会影响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，净转化率可以作为测量指标；
4. **转移单位：**后续试验中的转移单位为cookies，因为我们追踪的是学生是否

参加免费试学以及是否进行付费，如果使用用户ID作为转移单位，那么没有参加免费试学的用户ID就不会被记录，大大减少了我们的样本量，所以最优的转移单位是cookie；