

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：Cookie数量，点击数量，点进概率

评估指标：总转化率，存留率和净转换率

Cookie数量：试验的引流单位,在实验组和对照组都拥有相同的分布。

用户ID数量：会受到试学筛选器的影响,用户群体经过筛选,数量会有变化。但是点进概率会是一个更好的评估度量,所以不选择这个指标。

点击数量：试验并没有更改用户点击“开始免费试学”的体验,所以点击量不会发生显著变化,所以这是一个非常好的不变度量。

点进概率：它不会受到试学筛选器的影响,因为查看课程概述和点击“开始免费试学”都在试学筛选器被触发之前执行，所以可以作为不变指标。

总转化率：会受到“免费试学筛选器”的影响,因为用户可能会因为筛选器的提示而放弃免费试学,因此总转化率是一个评估指标。

存留率：同样会受到“免费试学筛选器”的影响,因为用户在通过筛选器的筛选后,继续试学直到付费的比例是有可能变化的,所以也是一个非常好的评估指标。

净转换率：它也会受到筛选器影响，用户行为可能会有不同的表现,也是非常好的评估指标。

对评估度量的期望结果：

总转化率:降低。免费试学筛选器的功能就是筛选用户,将难以完成课程的用户筛去。

存留率:提高。经过免费试学筛选器筛选后的用户更有可能付费完成课程。

净转换率:不降低。从用户群体中筛选掉难以完成的用户,而本身的用户数量没有增加,所以期望是不降低的。

测量标准偏差

总转换率:0.0202， 存留率:0.0549， 净转换率:0.0156

对于总转换率和净转换率,分析估计和经验变异是类似的,分析单位和引流单位相同,都是Cookie。存留率,分析估计和经验变异是不同的,分析单位和引流单位不同,前者是用户ID,后者是Cookie,所以在时间允许的情况下要进行经验估计。

规模

样本数量和功效

在分析阶段不使用Bonferroni校正。需要页面浏览量: 685325 该访问页数是选择“净转换率”指标，利用在线计算器计算得出：

总转化率：通过计算器算出 25835，然后 $25835 \times 2 / 0.08 = 645875$

留存率：通过计算器算出39115，然后 $39115 / 0.20625 = 189648.4848$ 得出点击按钮的cookie数，然

后 $189648.4848 \times 2 / 0.08 = 4741212.121$

净转换率：通过计算器算出 27413，然后 $27413 \times 2 / 0.08 = 685325$

不同的指标会计算出不同的访问页数:选择“存留率”得出的结果为4741212,数字过大会导致试验进行所需时间较长,所以不采用;选择“总转化率”得出的结果为645875,小于685325,所以不采用。

持续时间和曝光比例

选择将100%的流量 的流量转入此试验,需要18天来运行。

该试验的风险等级低,因为用户不太可能因为增加了试学筛选器就大幅度减少或增加付费学习的意愿,所以对数据库和后台服务器影响也不大, 同时也不涉及敏感的用户隐私。

因为该试验风险等级低,可以使用大比例流量进行试验,有助于缩短试验周期。

试验分析

合理性检查

Cookie数量: 期望观察值的置信区间: (0.4988, 0.5012) 实际观察值: 0.5006

是否通过完整性检验: 是

点击数量: 期望观察值的置信区间: (0.4959, 0.5041) 实际观察值: 0.5005

是否通过完整性检验: 是

点进概率: 期望观察值的置信区间: (0.0812, 0.0830) 实际观察值: 0.0822

是否通过完整性检验: 是

结果分析

效应大小检验

总转化率: 置信区间: (-0.0291, -0.0120) 最低敏感度: 0.01 统计显著: 是 实际显著: 是

存留率: 置信区间: (0.0081, 0.0540) 最低敏感度: 0.01 统计显著: 是 实际显著: 否

净转换率: 置信区间: (-0.0116, 0.0019) 最低敏感度: 0.0075 统计显著: 否 实际显著: 否

符号检验

总转换率: p值: 0.0026 统计显著: 是

存留率: p值: 0.6776 统计显著: 否

净转换率: p值: 0.6776 统计显著: 否

汇总

对“存留率”指标而言,符号检验和效应值检验的结果不一致。可能是符号检验的效能低于效应值检验,因为它对样本不作任何假设。又可能该指标在周末的差异具有显著性而在工作日没有,在符号检验中无法体现这一点,从而导致符号检验和效应值检验的结果不符。

试验未使用Bonferroni校正,评估度量只有3个,数量较少,不足以造成过多的假阳性情况。

建议

基于A/B测试的结论,建议是:不启动“试学筛选器”功能。

总转换率降低了,具有统计和实际显著性,即表明试学筛选器确实能使一些无法投入足够时间的用户不选择进行试学,这符合试验期望的结果;

存留率因为没有足够的样本, statistical power不足,不足以作为评估指标;

净转换率不具有统计和实际显著性,它的置信区间是 $(-0.0116, 0.0019)$, 包含负值, 所以净转化率很大概率会减少, 而且有一定概率减少会超过实际显著性0.0075, 因此无法说明降低程度不大, 所以不建议启动新功能。

后续试验

学时提醒 (为了给能充分投入学习的学生更加多的支持,改进学生体验。) :

记录每个试学用户的ID,试用期(2周),记录他们的学习时间,每周尾发送用户他们每周的学习时间,并给出建议(如果投入时间不足5小时一周, 建议免费观看课程而非付费)。

样本分成实验组和对照组,观察不同条件(有无“学时提醒”)下,存留率和净转换率的变化。

引流指标是用户ID,选择不变指标如下:

参加免费试学的用户ID。因为这是引流指标,所以是非常好的不变指标。

点击次数(点击“免费试学”按钮的唯一Cookie数量)。因为点击次数不会受“学时提醒”影响,该试验针对的是参加免费试学的用户,因此这个度量是不变指标。

选择评估指标如下:

存留率:存留率预计会受到“学时提醒”的影响,用户通过知道自己的学习时间,可以判断自己是否适合成为付费用户。因此这个度量可以作为评估指标;

净转换率:由于选择付费的用户数量可能发生变化,那么净转换率也可以成为一个合理的评估指标。