

试验设计

度量选择

Cookie 的数量：变化发生在访问页面之前，没有收到实验的影响，分组单元是 cookie，可随机分布，这是不变指标

用户 ID 数量：参与免费试学的用户数量，明显发生在测试之后，并非不变指标，可作为评估指标。但由于分组单元是 cookie，控制组和试验组按照 cookie 随机分配，可能有误差，用户 ID 不同很可能由 cookie 数量不同导致，不适合做评估指标。

点击次数：变化发生在访问页面之前，没有收到实验的影响，此 cookie 仍可以随机分配，为不变指标

点进概率：点击次数/cookie 数量，不变指标

总转化率：变化发生在测试之后，评估指标

留存率：变化发生在测试之后，评估指标

净转化率：变化发生在测试之后，评估指标

希望看到的结果：预期减少时间不足的同学注册，即希望看到总转化率减少，留存率增大，净转化率不变。

变异性

测量标准偏差

$$SE = \sqrt{q \cdot (1-q) / N}$$

假设有 5000 个 pageview 已知点击率为 0.08，因此点击人数为 $5000 \cdot 0.08$ ；

点击人数中付费概率为 0.53，故 5000 个 pageview 中付费人数为 $5000 \cdot 0.08 \cdot 0.53$

$$\text{总转化率: } \sqrt{0.21 \cdot (1-0.21) / (5000 \cdot 0.08)} = 0.0202$$

留存率: $\sqrt{0.53*(1-0.53)/(5000*0.08*0.53)}=0.0549$

净转化率: $\sqrt{0.1093*(1-0.1093)/(5000*0.08)}=0.0156$

经验变异与实际变异分析

分组单元: cookie

总转化率分析单元: 点击“开始免费试学”按钮的唯一 cookie 的数量, 与分组单元一致, 分析变异性更可能与经验变异性一致

留存率分析单元: 登录的用户 id 的数量, 分析变异性可能与经验变异性不一致, 导致结果不准确。

净转化率分析单元: 点击“开始免费试学”按钮的唯一 cookie 的数量, 分析变异性更可能与经验变异性一致。

规模

样本数量和支持

不使用 Bonferroni 校正: 计算总转化率、留存率、净转化率三个指标之间是相互关联的, 使用 Bonferroni 校正会导致结果保守, 故不使用。

根据在线计算器计算所需样本数量:

总转化率: 输入 $d_{min}=1\%$; $\alpha=5\%$; $1-B=80\%$, 转化率: 0.20835, 得到样本数量为 25839, 此为单组样本量, 设置控制组与实验组, 故总量: $25839*2=51678$, 此为点击人数的样本量, 要求出页面浏览量的样本数: $51678/0.8=64597.5$

留存率: 输入 $d_{min}=1\%$; $\alpha=5\%$; $1-B=80\%$, 转化率: 0.53, 得到样本数量为 39115, 此为单组样本量, 设置控制组与实验组, 故总量: $39115*2=78230$, 此为点击人数的样本量, 要求出页面浏览量的样本数: $78230/0.8/0.53=47412.12$

净转化率: 输入 $d_{min}=0.75\%$; $\alpha=5\%$; $1-B=80\%$, 转化率: 0.1093, 得到样本数量为 27411, 此为单组样本量, 设置控制组与实验组, 故总量: $27411*2=54822$, 此为点击人数的样本量, 要求出页面浏览量的样本数: $54822/0.8=68527.5$

选择最大样本量为 4741212，可满足上述所有指标所需样本量

风险等级、选择流量的原因

风险等级低，由于：

- a. 伦理道德：在网站做出提示需要多长时间学习的更改并不能对用户造成身体、情感、心理等的影响，小于最低风险。
- b. 益处：本次研究的目的在于减少注册用户，进而可进行精准定位用户提高付费人员数目
- c. 数据敏感性：用 cookie 分组，给出的数据仅计算各项总数，未联系到个人，无个人可识别性，敏感性低。
- d. 数据库安全：对数据库系统和信息方面都不会造成很大影响。

选择全部流量的原因：

- a. 本次测试不会对网站造成重大更改，风险较低
- b. 测试 18 天时间已能大体分清工作日和休息日区别
- c. 无需进行其他实验。

持续时间

已知每天浏览量为 40000，则若使用上述全部流量，分别需要持续的天数为：

总转化率：645975/40000=17

留存率：4741212/40000=119

净转化率：685275/40000=18

由于所选择的 4741212 的样本量持续时间为 119 天，A/Btest 一般不超过 30 天，故舍弃此度量，使用第二个样本量：685275，持续时间为 18 天。

完整性检查

查看数据可以看到第 25 条记录开始，enrollment 和 payment 无数据，因此分析时只考虑此行之前的 23 行数据，其余空缺数据直接剔除。

判断选中的不变量是否不变

1. Cookie 数量，为总计值，可按照 0.5 概率计算出置信区间，再查看试验组概率是否在置信区间范围之内。

按照 cookie 随机分配，因此

$$Se = \sqrt{0.5 * 0.5 / (n1 + n2)} = 0.0006018$$

$$M = 1.96 * se = 0.00118$$

置信区间：[0.4988, 0.5012]

控制组： $p1 = n1 / (n1 + n2) = 0.5006$ ，在置信区间范围之内

实验组： $p2 = n2 / (n1 + n2) = 0.4994$ ，在置信区间范围之内

此指标通过完整性检测。

2. “开始免费试用”的点入概率

按照 cookie 随机分配，因此

$$Se = \sqrt{0.5 * 0.5 / (n1 + n2)} = 0.0021$$

$$M = 1.96 * se = 0.004116$$

置信区间：[0.4959, 0.5041]

控制组： $p1 = n1 / (n1 + n2) = 0.5005$ ，在置信区间范围之内

实验组： $p2 = n2 / (n1 + n2) = 0.4995$ ，在置信区间范围之内

此指标通过完整性检测。

3. “开始免费试用”的点入概率：方法 a. 计算控制组的置信区间，查看试验组是否在其范围内；方法 b. 计算试验组的置信区间，查看控制组是否在其范围内；方法 c. 假设

差异 $d=0$ ，计算合并概率 p_{pool} 、合并标准差 se_{pool} ，误差界限 m ，即求出置信区间，，
即当 $abs(\text{实际差异}) > abs(m)$ ，具有显著差异。

$$p_{pool} = (x_1 + x_2) / (n_1 + n_2) = (28325 + 282378) / (344660 + 345543) = 0.0822$$

$$se_{pool} = \sqrt{0.0822 * (1 - 0.0822) / (n_1 + n_2)} = 0.0006$$

$$m = 1.96 * se_{pool} = 0.001296$$

置信区间： $[-0.001296, 0.001296]$

$$d = (28325 / 344660) - (28378 / 345543) = 0.00005663$$

d 在置信区间范围之内，通过完整性检测。

结果分析效应大小检验

1. 总转化率

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

$$p_{pool} = (3423 + 3785) / (17260 + 17293) = 0.2086$$

$$se_{pool} = \sqrt{p_{pool} * (1 - p_{pool}) * (1 / 17260 + 1 / 17293)} = 0.004372$$

$$m = 1.96 * se_{pool} = 0.008568$$

差异值 d 在 $[-m, m]$ 范围之外，因此具有统计显著性。

$$\text{置信区间下限: } d - m = -0.02912$$

$$\text{置信区间上限: } d + m = -0.01197$$

$$\text{置信区间: } [-0.02912, -0.01197]$$

$$\text{实际显著区间为 } [-0.01, 0.01]$$

因此置信区间全部在实际区间左侧，因此具有实际显著统计。

2. 留存率：计算实际所需样本时不符合需求，因此去除此指标，不进行计算。

3. 净转化率

$$d=1945/17260-2033/17293=-0.00487$$

$$ppool=(1945+2033)/(17260+17293)=0.1151$$

$$sepool=\sqrt{ppool*(1-ppool)*(1/17260+1/17293)}=0.003434$$

$$m=1.96*sepool=0.006731$$

差异值 d 在 $[-m,m]$ 范围之内，因此不具有统计显著性。

置信区间下限： $d-m=-0.0116$

置信区间上限： $d+m=0.001857$

置信区间： $[-0.0116, 0.001857]$

实际显著区间为 $[-0.0075, 0.075]$

因此置信区间在实际区间之内，因此不具有实际显著统计。

符号检验

1. 总转化率： $enrollment/clicks$

- 计算每行的总转化率
- 计算试验组总转化率大于控制组总转化率的数目，可看为成功数目
- 使用在线计算器算出 p 。(输入总数目 23，成功数目 4，概率 0.5)

得到 $p=0.0026$ ， $p < \alpha$ ；具有显著性差异

2. 净转化率： $payment/clicks$

计算每行的净转化率

计算试验组总转化率大于控制组总转化率的数目，可看为成功数目

使用在线计算器算出 p 。(输入总数目 23，成功数目 10，概率 0.5)

得到 $p=0.6776$ ， $p > \alpha$ ；不具有显著性差异

汇总

不使用 Bonferroni 校正:计算总转化率、留存率、净转化率三个指标之间是相互关联的, 使用 Bonferroni 校正会导致结果保守。

效应大小假设检验和符号检验得出一致结果。

建议

1.通过在点击“开始免费试学”后出现提示的实验可以看出试验组总转化率与控制组总转化率有显著性差异,但此项差异表现为负值,即此项更改的减少了转化率,符合预期效果。

2.净转化率不具有显著性差异, $[-0.0116, 0.001857]$, 置信区间包括负数, 是我们不愿意看到的结果。

综上, 此更改不建议发布。

后续试验

建议在学生每完成一个项目或一个章节时, 系统弹出鼓励消息:“加油! 离成功还有一步!”“很棒!”等等等。预计付费人员增加, 为网站增加财政收入。

假设: 净转化率增加

不变指标: 页面浏览量、点击人数、点击率

评估指标: 净转化率

转移单位: 用户 ID。已经在学习过程中的学生已经有了固定的 ID, 如果使用 cookie 作为转移单位并不能确保分组的准确性。