

A/B 测试项目作业

试验设计

指标选择

可供选择的指标如下：

- cookie 的数量：即访问课程概述页面的唯一 cookie 的数量。（d 最小=3000）
- 用户 id 的数量：即参与免费试学的用户数量。（d 最小 =50）
- 点击次数：即点击“开始免费试学”按钮的唯一 cookie 的数量（在免费试学筛选器触发前发生）。（d 最小 =240）
- 点进概率：即点击“开始免费试学”按钮的唯一 cookie 的数量除以查看课程概述页面的唯一 cookie 的数量所得的比率。（d 最小=0.01）
- 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。（d 最小 =0.01）
- 留存率：即在 14 天的期限过后仍参加课程（因此至少进行了一次付费）的用户 id 数量除以完成登录的用户 id 的数量。（d 最小 =0.01）
- 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。（d 最小 =0.0075）

由于对比试验的区别在于点击“开始免费试学”之后的系统弹框，因此 cookie 数量、点击次数、**点击概率（前面两者比值）**在试验中应该保持不变，而弹框之后产生影响的事件都可能发生变化，包括：用户 id 数量、总转化率、留存率、净转化率。因此，**不变指标**选择为：

- cookie 的数量；
- 点击次数；
- **点击概率。**

评估指标在实验过程中应该发生某种变化，即通过两组数据的对比，寻找弹框这一变化是否造成实在的影响。**用户 id 数量虽然会受到试验的影响，但是由于对照组与试验组 cookie 数量不一定相同，因此用户 id 可能不足以作为试验的评估指标，因而在总转化率（本身就**

是包含用户 id 数量的比值) 存在的情况下, 选择前者即可。所以此处的评估指标选择为:

- 总转化率;
- 留存率;
- 净转化率。

总转化率可能会减小, 弹框提示会过滤掉那些可能在试学期间退出的学员; 留存率可能会增大, 因为留下来的应该都是能够继续学习的学员; 净转化率可能没有大的影响。

测量标准偏差

系统提供的数据包括: 每天的访问 cookie 数量、每天的点击“开始试学”独立 cookie 数量、每天登陆数、点进率、总转化率、留存率、净转化率, 需要注意的是表格中的数据基于的 cookie 数量是 40000 次。实验中我们预计每一组 5000 个 cookie, 所以首先需要重新整理下数据。

$5000/40000 = 0.125$, 这是缩放因子, 因此对于表格中的 3200 点击与 660 次登陆, 可以得到 5000 次 cookie 下对应于 400 次点击与 82.5 次登陆。点击数量与登陆数量认为服从二项分布, 而总转化率、留存率、净转化率认为服从高斯分布, 得到标准差:

$$\sigma = \sqrt{\frac{p(1-p)}{N}}$$

根据数据表格, 有:

$P_{\text{总转化率}} = 0.20625$, $P_{\text{留存率}} = 0.53$, $P_{\text{净转化率}} = 0.1093125$, 故可以得到:

$$P_{\text{总转化率}} = \sqrt{\frac{0.20625 * (1 - 0.20625)}{400}} = 0.0202$$

$$P_{\text{留存率}} = \sqrt{\frac{0.53 * (1 - 0.53)}{82.5}} = 0.0549$$

$$P_{\text{净转化率}} = \sqrt{\frac{0.1093125 * (1 - 0.1093125)}{400}} = 0.0156$$

由于总转化率和净转化率的分组单元与分析单元相同, 因此分析结果与经验结果应该相同; 而留存率的分组单元与分析单元分别是 cookie 与用户 id, 所以分析结果与经验结果应该有所不同。

规模

样本数量和功效

从数据中得到最小可侦测效应分别为：

$$d_{\min}(\text{总转化率}) = 0.01, d_{\min}(\text{留存率}) = 0.01, d_{\min}(\text{净转化率}) = 0.0075。$$

同时 $\alpha = 0.05$, $\beta = 0.2$, 通过在线计算器得到样本数量 25835、39115、27413, 但是还需要转化为页面访问值。根据点击与页面访问的比例 $3200/40000 = 0.08$, 登陆与页面访问比例 $660/40000 = 0.0165$, 有：

$$N_{\text{总转化率}} = \frac{25835}{0.08} * 2 = 645875$$

$$N_{\text{留存率}} = \frac{39115}{0.0165} * 2 = 4741213$$

$$N_{\text{总转化率}} = \frac{27413}{0.08} * 2 = 685325$$

从道理上讲应该选择最大值, 但是由于第二项结果太大, 难以操作, 所以选择 685325 作为结论。没有使用 Bonferroni 校正。

持续时间和曝光比例

由于试验更改内容对用户无害（理由见后文描述），所以选择流量 100% 转入试验。

按照最大的留存率 4741213 个页面访问量, 以及平均每天 40000 页面访问来计算, 总共需要 $4741213/40000 = 118.53$ 约等于 119 天, 时间太长不具可操作性。选择第二大的总转化率 $685325/40000 = 17.13$ 约等于 18 天来作为实验时间。

试验风险评估

该试验无大的风险, 理由如下：

- 1) 即使学员每周学不到五小时, 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、登陆并可能完成继续课程, 不会因此影响用户习惯;
- 2) 页面展示没有过大的改动, 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) 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、用户无法访问等重大问题；
- 4) 试验不会对用户的个人信息安全造成风险，不论网页是否修改，用户在确认参加免费试学时都得输入信用卡信息，很明显系统一定会保护用户的个人信息；
- 5) 试验同样也没有道德上的风险。

基于上述考虑，不必考虑转移流量问题。

试验分析

合理性检查

之前选择的不变指标包括：cookie 数量、点击次数，期望分配概率为 $p = 0.5$ ，即应该平均分配到对照组与试验组，假设分布为正态分布，计算 $1 - \alpha = 95\%$ 置信度下的置信区间：

$$\sigma_1 = \sqrt{\frac{0.5 * (1 - 0.5)}{345545 + 344660}} = 0.000602$$

$$m = 1.96 * 0.000602 = 0.00118$$

$$CI = [0.5 - 0.00118, 0.5 + 0.00118] = [0.4988, 0.5012]$$

$$P1(\text{实际}) = 0.5006$$

同理可得：

$$\sigma_2 = \sqrt{\frac{0.5 * (1 - 0.5)}{28378 + 28325}} = 0.0021$$

$$m = 1.96 * 0.0021 = 0.00412$$

$$CI = [0.5 - 0.00412, 0.5 + 0.00412] = [0.4959, 0.5041]$$

$$P2(\text{实际}) = 0.5005$$

因此，两个指标都通过了合理性检验。

结果分析

效应大小检验

对于总转化率：

$$R_{\text{对照组}} = 0.21887, R_{\text{试验组}} = 0.19832, d = -0.02055$$

$$\text{Var}_{\text{对照组}} = 0.21887 * (1 - 0.21887) / 17293 = 9.886e-6$$

$$\text{Var}_{\text{试验组}} = 0.19832 \cdot (1 - 0.19832) / 17260 = 9.211\text{e-}6$$

$$\text{Var}_d = \text{Var}_{\text{对照组}} + \text{Var}_{\text{试验组}} = 1.9097\text{e-}5$$

$$\sigma_d = 0.00437$$

$$m = 1.96 \cdot \sigma_d = 8.5652\text{e-}3$$

$$\text{CI} = [d - m, d + m] = [-0.0291, -0.012]$$

由于 $d_{\min} = -0.01$ 不在 CI 范围内，0 也不在 CI 范围内，所以总转化率既是统计显著，也是实际显著。

对于净转化率：

$$R_{\text{对照组}} = 0.11756, R_{\text{试验组}} = 0.11269, d = -0.00487$$

$$\text{Var}_{\text{对照组}} = 0.11756 \cdot (1 - 0.11756) / 17293 = 5.999\text{e-}6$$

$$\text{Var}_{\text{试验组}} = 0.11269 \cdot (1 - 0.11269) / 17260 = 5.793\text{e-}6$$

$$\text{Var}_d = \text{Var}_{\text{对照组}} + \text{Var}_{\text{试验组}} = 1.1792\text{e-}5$$

$$\sigma_d = 0.00343$$

$$m = 1.96 \cdot \sigma_d = 0.00672$$

$$\text{CI} = [d - m, d + m] = [-0.016, 0.0019]$$

由于 $d_{\min} = -0.0075$ 在 CI 范围内，0 也在 CI 范围内，所以总转化率既不是统计显著，也不是实际显著。

符号检验

测试指标有 23 组对比值，对于**总转化率**，其中有 4 组值呈现“大”，19 组呈现“小”，对应概率 p 值为 0.0026，与独立指标的 α 值（Bonferroni 校正）0.025 比较远小，因此具有统计显著性。而**净转化率**“大”的比例为 10/23，对应 p 值为 0.6776，说明该指标不具有统计显著性。

汇总

此处没有使用 Bonferroni 校正，因为希望考核每个单独指标是否都满足显著性。

分析结果显示，总转化率既满足统计显著性，也满足实际显著性，而净转化率两者都不满足。与之前我们的预期相比，总转换率降低了约 2 个百分点（ $d = -0.02055$ ），这说明弹框提示确实会过滤掉那些可能在试学期间退出的学员；而净转化率降低了约 0.5 个百分点（ $d =$

-0.00487)，说明弹框对于 14 天的期限后仍参与课程的用户比例呈现出负面影响。

建议

总转化率具有统计和实际显著性，与预期相符；净转化率置信区间[-0.016, 0.0019]包含负数，说明有一定可能性净转化率会减小，且有一定可能性净转化率减小程度超过 0.0075，因此无法说明“降低程度不大”。所以建议对试验设计重新审视，暂时不要发布。

后续试验

后续试验可以考虑对于时间不够的学员给出弹框信息，要求增加学习时间；如果投入时间满足要求，则不弹框提示。这应该会使得那些原本投入时间不充分的学员更加努力，从而提高整体的通过率。可以将留存率指标作为测试指标，将学员分配到对照组和测试组，用户 id 应该作为不变指标。

参考：

<https://discussions.youdaxue.com/t/topic/36225>

<https://discussions.youdaxue.com/t/topic/42144>

<https://discussions.youdaxue.com/t/final-project-quiz-calculating-standard-deviation/38092>

<https://discussions.youdaxue.com/t/topic/32932>

<https://discussions.youdaxue.com/t/topic/46664>

<https://discussions.youdaxue.com/t/47-5/43240>

<https://discussions.youdaxue.com/t/topic/43307>

<https://discussions.youdaxue.com/t/topic/32504>

<http://discussions.youdaxue.com/t/topic/37373>