

# 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支持或验证证书，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。

我们的假设是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学生体验和提高导师为能够完成课程的学生提供支持的能力。

转移单位为 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

#### 不变指标

- Cookie 数量：即查看课程概述页面的唯一 cookie 的数量
  - 该指标的统计发生在变更内容（询问投入时间）之前，因此不会受变更影响
  - 期望对照组和实验组查看课程概述页的 cookie 数一样或没有显著差异，因此作为不变指标
- 点击次数：即点击“开始免费试用”按钮（在免费试用屏幕触犯前发生）的唯一 cookie 的数量
  - 该指标的统计发生在变更内容（询问投入时间）之前，因此不会受变更影响
  - 期望对照组和实验组查看课程概述页的 cookie 数一样或没有显著差异，因此作为不变指标
- 点进概率：即点击“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 的数量所得的结果
  - 因为分子分母都是不变指标，所以相除以后的结果也是不变指标
  - 期望对照组和实验组计算出的点进概率一样或没有显著差异

#### 评估指标

- 总转化率 $d_{min}=0.01$ ：即完成登录并报名参加免费试用的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果

$$P_{\text{总转化率}} = \frac{\text{注册报名免费试学用户数}}{\text{点击“开始免费试学”cookies数}}$$

- 该指标分子的统计发生在变更内容之后，可能会受变更影响
  - 该指标分母的统计发生在变更内容之前，不会受变更影响
  - 综上，该指标可能会受变更影响
  - 按照实验期望，变更内容是为了减少因为没有时间而参与免费试学受挫的学生数量，因此期望分子变小，该指标会变小
- 留存率 $d_{min}=0.01$ ：即在 14 天期限后仍保持参加（并进行了至少一次支付）的用户 id 的数量除以完成登录的用户 id 的数量

$$P_{\text{留存率}} = \frac{\text{试学后付费用户数}}{\text{注册报名免费试学用户数}}$$

- 该指标分子、分母的统计都发生在变更内容之后，都可能会受变更影响
  - 按照实验期望，变更内容是为了减少因为没有时间而参与免费试学受挫的学生数量，且期望不会减少继续学习的学生数，分子不变、分母变小，相应则会期望实验组提高留存率，因此作为评估指标
- 净转化率 $d_{min}=0.0075$ ：即在 14 天期限结束后仍然参加（并至少进行了一次支付）的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果

$$P_{\text{净转化率}} = \frac{\text{试学后付费用户数}}{\text{点击“开始免费试学”cookies数}}$$

- 该指标分子的统计发生在变更内容之后，可能会受变更影响
- 该指标分母的统计发生在变更之前，不会受变更影响
- 按照实验期望，变更内容是为了减少因为没有时间而参与免费试学受挫的学生数量，且期望不会减少继续学习的学生数，分子不变、分母不变，相应则会期望实验组净转化率不变，因此作为评估指标

## 没有选择的指标

- 用户 id 数量 $d_{min}=50$ ：即参与免费试学的用户数量
  - 不被选为不变指标的原因：该指标的统计发生在变更之后，可能会受变更影响
  - 不被选为评估指标的原因：虽然该指标的统计发生在变更之后，会受到影响，本是一个可选择的评估指标。但实验的转移单位是 cookie，若选择这个指标无法确保引流一致性。实验组和对照组的 cookie 数量不一定相同，该指标的变化也许不是受变更影响，而是因为 cookie 数不同导致。由于该指标不能很好地评估变更效果，而总转化率是对用户 id 数量进行归一化的指标，比用户 id 数量更合适作为评估指标，存在总转化的情况下，没有必要选择用户 id 数量作为评估指标

## 测量标准偏差

- 基准值表

基准值名称	基准值	5000 cookies
Unique cookies to view page per day - 每天访问页面的个体 cookies 数	40000	5000
Unique cookies to click "Start free trial" per day - 每天点击“开始免费试学”的个体 cookies 数	3200	400
Enrollments per day - 每天报名免费试学个体数	660	82.5
Click-through-probability on "Start free trial" - “开始免费试学”的点击概率	0.08	0.08
Probability of enrolling, given click - 总转化率	0.20625	0.20625
Probability of payment, given enroll - 留存率	0.53	0.53
Probability of payment, given click - 净转化率	0.1093125	0.1093125

- 总转化率

$$SE_{\text{总转化率}} = \sqrt{\frac{\text{总转化率} * (1 - \text{总转化率})}{\text{点击“开始免费试学”cookie数}}} = \sqrt{\frac{0.20625 * (1 - 0.20625)}{400}} = 0.0202$$

- 分析单位和引流单位相同，都是点击“开始免费试学” cookie 数，所以分析估计会和经验变异类似

- 留存率

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

- 分析单位是注册数，引流单位是点击“开始免费试学” cookie 数，因为单位不相同，所以分析估计会和经验变异不同

- 净转化率

$$SE_{\text{净转化率}} = \sqrt{\frac{\text{净转化率} * (1 - \text{净转化率})}{\text{点击“开始免费试学”cookie数}}} = \sqrt{\frac{0.1093125 * (1 - 0.1093125)}{400}} = 0.0156$$

- 分析单位和引流单位相同，都是点击“开始免费试学” cookie 数，所以分析估计会和经验变异类似

## 规模

- 样本数量和功效

- 在分析阶段不会使用 Bonferroni 校正，因为 Bonferroni 校正是用于避免一类错误，适用于检验同一数据集的不同独立指标的假设，而目前评估指标是同一数据集的相关指标，因此不做校正
- 以下所需分析单位由[在线计算器](#)计算得出

$$1 - \beta = 80\%$$

$$\alpha = 5\%$$

- 若以总转换率作为评估指标，计算所需分析单位，即每组所需 cookie 数 = 25835，两组总共所需访问页面数=25835/(400/5000)\*2=645875，所需天数=645875/40000=16.15=17 天
- 若以留存率作为评估指标，计算所需分析单位，即每组所需注册数 = 39115，两组总共所需访问页面数=39115/(82.5/5000)\*2=4741212，所需天数=4741212/40000=118.5=119 天
- 若以净转化率作为评估指标，计算所需分析单位，即每组所需 cookie 数 = 27413，两组总共所需访问页面数=27413/(400/5000)\*2=685325，所需天数=685325/40000=17.13=18 天
- 由于留存率所需时间太长，因此不考虑。总转化率和净转化率的时间均符合要求，选择需要页面较大的进行试验时间计算
- 持续时间和曝光比例
  - 在这里选择转移 70% 的流量到此试验，所需天数=685325/40000/0.7=24.48=25天
  - 该试验的风险较低，因此选择大多数流量，且不选择更小流量是因为避免实验时间太长
- 风险较低的原因如下
  - 不会对个人信息安全造成风险：因为变更只是增加了提醒和引导，而没有额外收集用户个人信息
  - 没有道德伦理风险：提醒页面不会强制用户做出不利于用户的选择
  - 没有数据丢失风险：对于变更的开发而言，增加一个提示页面不涉及后台数据的修改或删除，因此不会因实验导致操作失误令网页崩溃
  - 没有用户体验不适的风险：页面和功能没有做出额外改动，用户仍然可以完成原有的功能，也不需要去适应这个变更

## 试验分析

### 合理性检查

- 实验数据 - Control

Date	Pageviews	Clicks	Enrollments	Payments
Sat, Oct 11	7723	687	134	70
Sun, Oct 12	9102	779	147	70
Mon, Oct 13	10511	909	167	95
Tue, Oct 14	9871	836	156	105
Wed, Oct 15	10014	837	163	64
Thu, Oct 16	9670	823	138	82

Fri, Oct 17	9008	748	146	76
Sat, Oct 18	7434	632	110	70
Sun, Oct 19	8459	691	131	60
Mon, Oct 20	10667	861	165	97
Tue, Oct 21	10660	867	196	105
Wed, Oct 22	9947	838	162	92
Thu, Oct 23	8324	665	127	56
Fri, Oct 24	9434	673	220	122
Sat, Oct 25	8687	691	176	128
Sun, Oct 26	8896	708	161	104
Mon, Oct 27	9535	759	233	124
Tue, Oct 28	9363	736	154	91
Wed, Oct 29	9327	739	196	86
Thu, Oct 30	9345	734	167	75
Fri, Oct 31	8890	706	174	101
Sat, Nov 1	8460	681	156	93
Sun, Nov 2	8836	693	206	67
Mon, Nov 3	9437	788		
Tue, Nov 4	9420	781		
Wed, Nov 5	9570	805		
Thu, Nov 6	9921	830		
Fri, Nov 7	9424	781		
Sat, Nov 8	9010	756		
Sun, Nov 9	9656	825		
Mon, Nov 10	10419	874		
Tue, Nov 11	9880	830		
Wed, Nov 12	10134	801		
Thu, Nov 13	9717	814		
Fri, Nov 14	9192	735		

Sat, Nov 15	8630	743		
Sun, Nov 16	8970	722		

- 实验数据 - Experiment

Date	Pageviews	Clicks	Enrollments	Payments
Sat, Oct 11	7716	686	105	34
Sun, Oct 12	9288	785	116	91
Mon, Oct 13	10480	884	145	79
Tue, Oct 14	9867	827	138	92
Wed, Oct 15	9793	832	140	94
Thu, Oct 16	9500	788	129	61
Fri, Oct 17	9088	780	127	44
Sat, Oct 18	7664	652	94	62
Sun, Oct 19	8434	697	120	77
Mon, Oct 20	10496	860	153	98
Tue, Oct 21	10551	864	143	71
Wed, Oct 22	9737	801	128	70
Thu, Oct 23	8176	642	122	68
Fri, Oct 24	9402	697	194	94
Sat, Oct 25	8669	669	127	81
Sun, Oct 26	8881	693	153	101
Mon, Oct 27	9655	771	213	119
Tue, Oct 28	9396	736	162	120
Wed, Oct 29	9262	727	201	96
Thu, Oct 30	9308	728	207	67
Fri, Oct 31	8715	722	182	123
Sat, Nov 1	8448	695	142	100
Sun, Nov 2	8836	724	182	103
Mon, Nov 3	9359	789		
Tue, Nov 4	9427	743		

Wed, Nov 5	9633	808		
Thu, Nov 6	9842	831		
Fri, Nov 7	9272	767		
Sat, Nov 8	8969	760		
Sun, Nov 9	9697	850		
Mon, Nov 10	10445	851		
Tue, Nov 11	9931	831		
Wed, Nov 12	10042	802		
Thu, Nov 13	9721	829		
Fri, Nov 14	9304	770		
Sat, Nov 15	8668	724		
Sun, Nov 16	8988	710		

- 统计数据进行合理性检查

- 选择置信度 95%， $z^* = 1.96$
- Cookie 数量：即查看课程概述页面的唯一 cookie 的数量

$$p = p_{cnt-cookie} = 0.5$$

$$N_{cookie} = N_{cnt-cookie} + N_{exp-cookie} = 345543 + 344660 = 690203$$

$$SE_{cookie} = \sqrt{\frac{p * (1 - p)}{N_{click}}} = \sqrt{\frac{0.5 * (1 - 0.5)}{690203}} = 0.0006$$

$$\text{置信区间下限} = p - z^* * SE_{cookie} = 0.5 - 1.96 * 0.0006 = 0.4988$$

$$\text{置信区间上限} = p + z^* * SE_{cookie} = 0.5 + 1.96 * 0.0006 = 0.5012$$

$$\text{观测值} = \frac{N_{cnt-cookie}}{N_{cookie}} = \frac{345543}{690203} = 0.5006$$

∴ 观测值在置信区间内，符合期望 ∴ 该不变指标通过合理性检查

- 点击次数：即点击“开始免费试用”按钮（在免费试用屏幕触犯前发生）的唯一 cookie 的数量

$$p = p_{cnt-click} = 0.5$$

$$N_{click} = N_{cnt-click} + N_{exp-click} = 28378 + 28325 = 56703$$

$$SE_{click} = \sqrt{\frac{p * (1 - p)}{N_{click}}} = \sqrt{\frac{0.5 * (1 - 0.5)}{56703}} = 0.0021$$

$$\text{置信区间下限} = p - z^* * SE_{click} = 0.5 - 1.96 * 0.0021 = 0.4959$$

$$\text{置信区间上限} = p + z^* * SE_{click} = 0.5 + 1.96 * 0.0021 = 0.5041$$

$$\text{观测值} = \frac{N_{cnt-click}}{N_{click}} = \frac{28378}{56703} = 0.5005$$

∴ 观测值在置信区间内，符合期望 ∴ 该不变指标通过合理性检查

- 点进概率：即点击“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 的数量所得的结果

$$p = p_{cnt-CTP} = \frac{N_{cnt-click}}{N_{cnt-cookie}} = \frac{28378}{345543} = 0.0821$$

$$N_{CTP} = N_{cnt-cookie} = 345543$$

$$SE_{CTP} = \sqrt{\frac{p * (1 - p)}{N_{CTP}}} = \sqrt{\frac{0.0821 * (1 - 0.0821)}{345543}} = 0.000467$$

$$\text{置信区间下限} = p - z^* * SE_{CTP} = 0.0821 - 1.96 * 0.000467 = 0.0812$$

$$\text{置信区间上限} = p + z^* * SE_{CTP} = 0.0821 + 1.96 * 0.000467 = 0.0830$$

$$\text{观测值} = \frac{N_{exp-click}}{N_{exp-cookie}} = \frac{28325}{344660} = 0.0822$$

∴ 观测值在置信区间内，符合期望 ∴ 该不变指标通过合理性检查

## 结果分析

- 效应大小检验
  - 选择 95% 置信区间
  - 统计值

	控制组	实验组	total
参加免费试学（23天）	3785	3423	7208
点击次数（23天）	17293	17260	34553
支付用户数（23天）	2033	1945	3978

- 总转化率：即完成登录并报名参加免费试用的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果。

$$\text{合并概率 } P_{pool} = \frac{N_{enroll}}{N_{click}} = \frac{7208}{34553} = 0.20861$$

$$\text{合并标准误差 } SE_{pool} = \sqrt{(1 - P_{pool}) * P_{pool} * (\frac{1}{N_{cnt-click}} + \frac{1}{N_{exp-click}})}$$

$$= \sqrt{(1 - 0.20861) * 0.20861 * (\frac{1}{17293} + \frac{1}{17260})}$$

$$= 0.0044372$$

$$\text{差异 } d = P_{exp} - P_{cnt} = \frac{N_{exp-enroll}}{N_{exp-click}} - \frac{N_{cnt-enroll}}{N_{cnt-click}} = \frac{3423}{17260} - \frac{3785}{17293} = -0.020555$$

$$\text{置信区间下限} = d - z^* * SE_{pool} = -0.020555 - 1.96 * 0.004372 = -0.029124$$

$$\text{置信区间上限} = d + z^* * SE_{pool} = -0.020555 + 1.96 * 0.004372 = -0.011986$$

∴ 置信区间不包含 0 ∴ 具有统计显著性

∴ 置信区间和  $[-d_{min}, d_{min}]$  ( $d_{min} = 0.01$ ) 没有交集 ∴ 具有实际显著性

- 净转化率：即在 14 天期限结束后仍然参加（并至少进行了一次支付）的用户 id 的数量除以点击“开始免费试用”按钮的唯一 cookie 的数量所得的结果



$$\text{合并概率 } P_{pool} = \frac{N_{pay}}{N_{click}} = \frac{3978}{34553} = 0.115127$$

$$\begin{aligned}\text{合并标准误差 } SE_{pool} &= \sqrt{(1 - P_{pool}) * P_{pool} * (\frac{1}{N_{cnt-click}} + \frac{1}{N_{exp-click}})} \\ &= \sqrt{(1 - 0.115127) * 0.115127 * (\frac{1}{17293} + \frac{1}{17260})} \\ &= 0.003434\end{aligned}$$

$$\text{差异 } d = P_{exp} - P_{cnt} = \frac{N_{exp-pay}}{N_{exp-click}} - \frac{N_{cnt-pay}}{N_{cnt-click}} = \frac{1945}{17260} - \frac{2033}{17293} = -0.004874$$

$$\text{置信区间下限} = d - z^* * SE_{pool} = -0.004874 - 1.96 * 0.003434 = -0.011605$$

$$\text{置信区间上限} = d + z^* * SE_{pool} = -0.004874 + 1.96 * 0.003434 = 0.001857$$

∵ 置信区间包含 0 ∴ 不具有统计显著性

∵ 置信区间和  $[-d_{min}, d_{min}]$  ( $d_{min} = 0.0075$ ) 有交集 ∴ 不具有实际显著性

- 符号检验

- 以下通过[在线计算器](#)计算
- 总转化率

总天数 = 23

成功天数 = 19

假设概率 = 0.5

计算结果： $p_{two-tail} = 0.0026$

∵ 结果小于假设概率 ∴ 具有统计显著性

- 净转化率

总天数 = 23

成功天数 = 13

假设概率 = 0.5

计算结果： $p_{two-tail} = 0.6776$

∵ 结果大于假设概率 ∴ 不具有统计显著性

- 汇总

- 没有使用 Bonferroni 校正，因为 Bonferroni 校正是用于避免一类错误，适用于检验同一数据集的不同独立指标的假设，而目前评估指标是同一数据集的相关指标，因此不做校正
- 效应大小假设检验和符号检验的结果一致，符合预期

## 建议

- 虽然总转化率具有统计和实际显著性，符合实验预期，变更能够减少原本时间不够但进行免费试学的用户，但因为净转化率的置信区间 (-0.0116, 0.0019) 包含负数，也就是有很大概率净转化率会降低，且有一定概率净转化率的降低会超过实际显著性 0.0075，与实验预期“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”不符。因此综合分析结果不建议启动。

## 后续试验

- 实验概述：为了提高用户体验，减少付费后但最终不能按期完成项目毕业的用户数量，且不影响能正常毕业的用户，在7天免费试学期间，用户访问课程时提示用户，昨天是否按照推荐进度完成了课程，如果没有，提示存在不能毕业的风险以及7天内可以申请退款。
- 假设：增加变更提示后，付费7天试学后因为没有时间而不能按其毕业的用户会减少
- 不变指标：免费试学用户。因为变更发生在报名免费试学后，不会影响免费试学用户
- 测量指标：毕业率=按期毕业用户/免费试学用户、退款率=申请退款用户/免费试学用户。根据假设，免费试学用户不变，增加了变更后，没有按推荐进度完成课程的用户可能会申请退款，退款率就会提高，而毕业率应该保持不变
- 转移单位：免费试学用户。根据测量指标，分析单位是免费试学用户，因此匹配的转移单位选择是免费试学用户