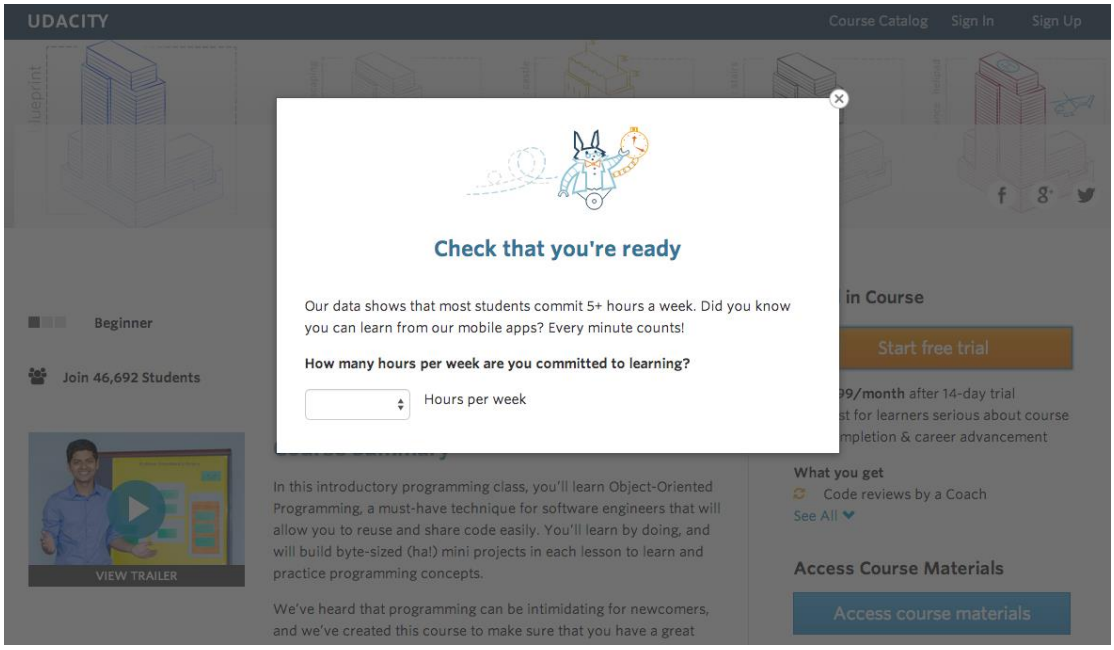
项目：设计 A/B 测试

项目说明

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。 这张截图展示了试验概况：



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

设计控制组和实验组，用来检验询问学习时间的对话框，对继续通过免费试学和最终付费完成课程的学生数量有无影响。

度量选择

任何提及“唯一 cookie”的地方，其唯一性按天决定。（在两个不同日期进行访问的同一个 cookie 将计算两次。）用户 id 自动唯一，因为网站不允许同一个用户 id 参与两次。本实验是在 cookie 的维度，将数据分为控制组和实验组，分组单位是 cookie。

不变度量

Cookie 的数量
解释：查看课程概述页面的唯一 cookie 的数量；该数量属于整体性的，发生在点击开始免费试用按钮事件之前，所以作为不变度量。
点击次数
解释：点击“开始免费试用”按钮的唯一 cookie 的数量；点击按钮后才弹出询问的对话框，发生在控制组和对照组之前，对本实验对该度量无影响，所以可以作为不变度量。
点进概率
解释：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；该度量=点击次数/Cookie 的数量，而由上可知分子分母都为不变度量，所以也可以作为不变度量。

评估度量

总转化率
解释：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；发生在实验之后，预期会减少登录的用户数，本实验对该度量有所影响；该评估度量的分母是 cookie，即分析单位为 cookie，所以总转化率的分析单位与分组单位一致。
留存率
解释：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；发生在实验之后，预期会减少登录的用户数，本实验对该度量有所影响；该评估度量的分母是登录的用户 id 数，即分析单位为 id，所以分析单元和分组单元不一致。
净转化率
解释：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；发生在实验之后，预期会减少登录的用户数，本实验对该度量有所影响；该评估度量的分母是 cookie，即分析单位为 cookie，所以净转化率的分析单位与分组单位一致。

其他度量

*用户 id 的数量：*即参与免费试学的用户数量。这个度量没有作为不变度量和评估度量，是由于实验组和对照组的 cookie 数量不一定相同，两组中用户 ID 数量不同可能是由于实验的影响，也可能是由于两组 cookie 的不同，所以使用用户 ID 数量的区别不能够很好的评估。在一个比例化的评估度量（总转化率）存在的情况下，不选择用户 ID 的数量作为评估度量。

测量标准偏差

$$\sigma_p = SE = \sqrt{\frac{p*(1-p)}{n}}$$

事件发生率的标准偏差公式：

标准偏差如下表所示：

metrics	baseline_values	sample_values
Unique cookies to view page per day:	40000	5000
Unique cookies to click "Start free trial" per day:	3200	400
Enrollments per day:	660	82.5
Click-through-probability on "Start free trial":	0.08	
Probability of enrolling, given click:	0.20625	0.020230604
Probability of payment, given enroll:	0.53	0.054949012
Probability of payment, given click	0.1093125	0.015601545

总转化率：期望实验组的该值会**变小**；对话框是建议投入时间不足的人放弃登录免费试用，从而分子减小，而分母是不变度量。

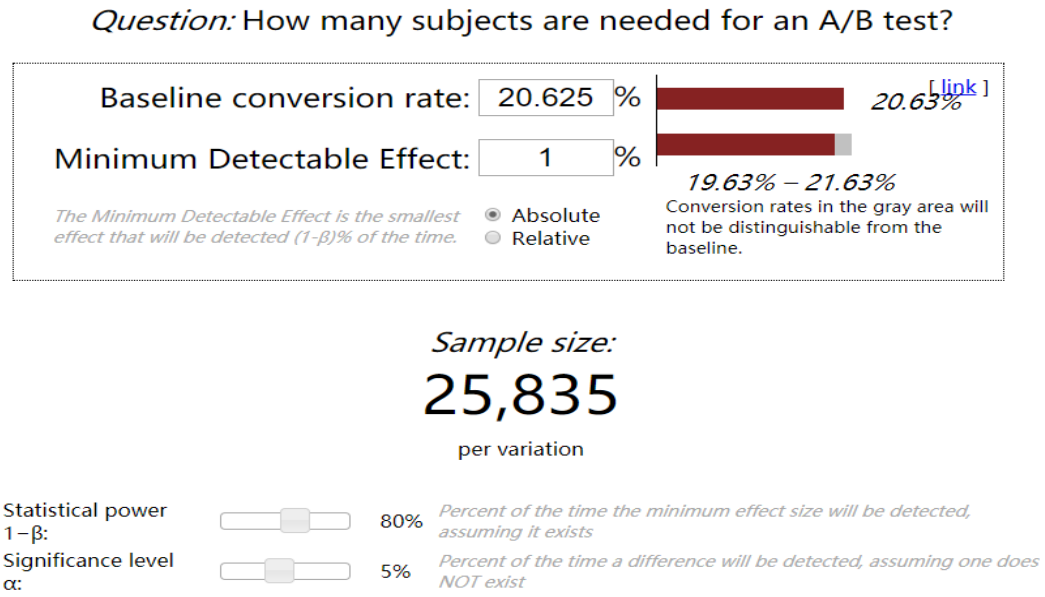
留存率： 期望实验组的该值会**变大**；实验的预期是尽可能减少继续通过免费试学和最终完成课程的学生数量，而分母是所有免费试学的人，对话框是建议投入时间不足的人放弃登录免费试用，分母变小。

净转化率：期望实验组的该值**不降低**；实验的预期是尽可能减少继续通过免费试学和最终完成课程的学生数量，而分母是点击开始免费试学按钮的唯一 cookie 数为不变度量。实验组也是希望通过不影响净转化率结论，来证实“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”。

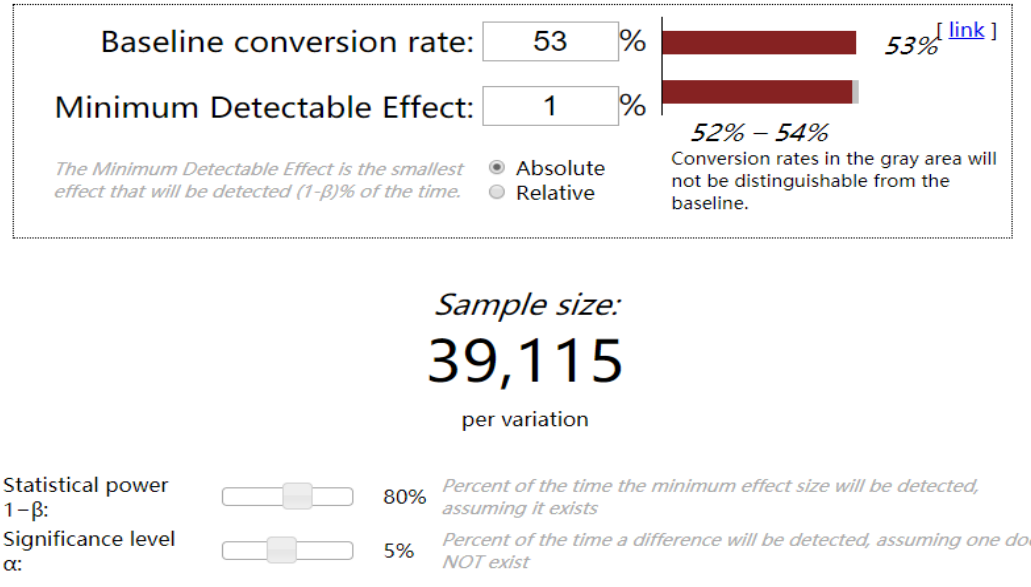
规模

样本数量和支持

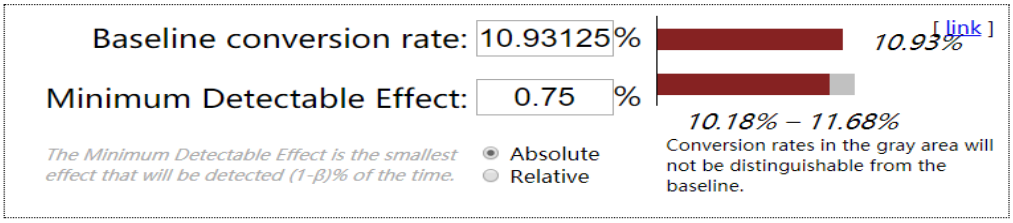
在 ABTest 的分析阶段，不使用 Bonferroni 校正，因为选择的指标具有一定的相关性。
开展试验所需的支持网页访问数: (http://www.evanmiller.org/ab-testing/sample-size.html)
tool: α=5%，1−β=80%
总转化率（$d_{最小}$ =0.01） Sample size=25835

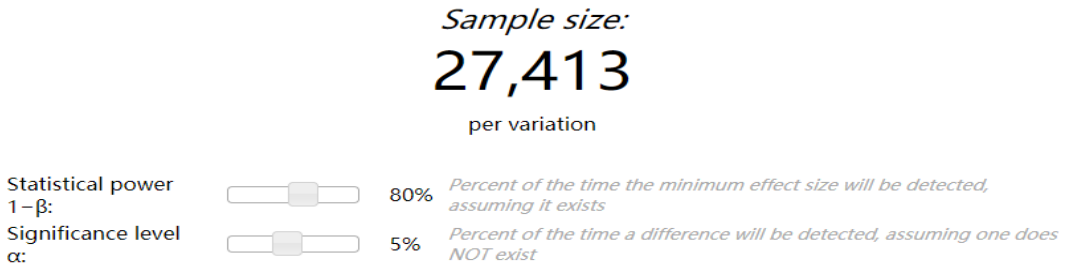


留存率（$d_{最小}$ =0.01）Sample size =39115



净转换率（$d_{最小}$ =0.0075）Sample size =27413





持续时间和风险暴露

Baseline- metrics:

Unique cookies to view page per day: 40000

Click-through-probability on "Start free trial": 0.08

Enrollments per day: 660

选择 80%的流量转入实验: 0.8

总转化率 Sample size=25835

(点击按钮)页面浏览量=(25835/0.08)*2=645875

持续时间(天)=645875/40000/0.8=20

留存率 Sample size=39115

(用户登录)页面浏览量=(39115/(660/40000))*2=4741212

持续时间(天)=4741212/40000/0.8=147.5

净转换率 Sample size=27413

(点击按钮)页面浏览量=(27413/0.08)*2=685325

持续时间(天)=685325/40000/0.8=21.25

选择 685325 的页面浏览量，需要 22 天来运行该实验；4741212 的页面浏览量的情况下，所需要的实验周期太长为 148 天，留存率需要的时间过长，不再把留存率作为评估度量。

试验分析

该实验的风险等级较低，点击免费试学按钮后弹出的对话框，不会对用户造成身体或者精神方面的伤害，而收集到某用户是否有足够时间学习的数据对用户的敏感程度小，远低于用户的 id、身份证识别号、手机号码等数据的敏感程度。对优达学城的影响可能就是某种程度上减少注册量，不排除有用户学习免费试用后能下决心安排时间付费学习。

由上面的分析，该试验需要 80%的流量转入此实验对优达学成风险不大，因为对话框只是根据询问学习时间做不同的指引，对于真正准备好付费学习的学生来说只是多了一个友好的提示；而对于暂时没准备好学习的同学引流到其他地方，比如其他的入门课程等；从另一方面来说更加人性化，能够提示该课程所需要的学习时间从而保证学习的效果。

合理性检查

详细计算请查看 p3+Baseline+Values.xlsx(Sheet:Check)，以下是表格及公式截图

Control_Pageview	Experiment_Pageview	p=0.5	z-critical=1.96 (α=0.05)	Lower bound	Upper bound	Observed	Passes
345543	344660	0.000601841	0.001179608	0.498820392	0.501179608	0.5006397	yes
Control_Clicks	Experiment_Clicks						
28378	28325	0.002099747	0.004115504	0.495884496	0.504115504	0.5004673	yes
Click-through-probability on "Start free trial":							
^d	P ^pool	SE pool					
-0.000057	0.082154091	0.000661061	0.001295679	-0.001295679	0.001295679	-0.000057	yes

$$\hat{d} = p\hat{}_{exp} - p\hat{}_{cont}$$

$$p\hat{}_{pool} = \frac{X_{cont}+X_{exp}}{N_{cont}+N_{exp}}$$

$$SE_{pool} = \sqrt{p\hat{}_{pool} * (1 - p\hat{}_{pool})(\frac{1}{N_{cont}} + \frac{1}{N_{exp}})}$$

$$\sigma_p = SE = \sqrt{\frac{p*(1-p)}{n}}$$

合理性检查通过：

95%的置信区间下，从表格的数据可以看出，控制组和实验组的页面的浏览量、点击量、点击率相当，分析可以继续下去。

结果分析效应大小检验

对于每个评估度量，对试验和对照组之间的差异给出 95% 置信区间。说明每个度量是否具有统计和实际显著性。注：**留存率**的实验的时间周期过长，后续的分析将其忽略。

详细计算请查看 p3+Baseline+Values.xlsx(Sheet: Effect Size Tests)，以下是表格及公式截图

	click	enroll	pay
Control	17293	3785	2033
Experiment	17260	3423	1945

总转化率=enroll/click

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

^d(exp-cont)	P ^pool	SE pool	z-critical=1.96 (α=0.05)	Lower bound	Upper bound
-0.02055487	0.208607067	0.004372	0.008568484	-0.029123358	-0.011986391

净转化率=pay/click

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c

^d(exp-cont)	P ^pool	SE pool	z-critical=1.96 (α=0.05)	Lower bound	Upper bound
-0.00487372	0.115127485	0.003434	0.006730902	-0.011604624	0.001857179

$$\hat{d} = p\hat{}_{exp} - p\hat{}_{cont}$$

$$p\hat{}_{pool} = \frac{X_{cont}+X_{exp}}{N_{cont}+N_{exp}}$$

$$SE_{pool} = \sqrt{p\hat{}_{pool} * (1 - p\hat{}_{pool})(\frac{1}{N_{cont}} + \frac{1}{N_{exp}})}$$

依据定义：如果置信区间不包含 0，这个指标具有统计显著性。如果置信区间不包含实际显著性边界。

总转化率具有统计显著性：置信区间都是负值，不包含 0 即在同一侧；

总转化率具有实践显著性：置信区间不包含显著性的边界，绝对值>d.min=0.01；

净转化率不具有统计和实践的显著性：置信区间包含 0，即在 0 的两侧，无法说明问题，有降低的情况超出了预期。

符号检验

对于每个评估度量，使用每日数据进行符号检验，然后报告符号检验的 p 值以及结果是否具有统计显著性。

详细计算请查看 p3+Baseline+Values.xlsx(Sheet: Sign Tests)，以下是表格截图

Date	Control		Experiment		C-E	C-E
	总转化率	净转化率	总转化率	净转化率	diff_总转化率	diff_净转化率
Sat, Oct 11	0.195051	0.101892	0.153061	0.049563	0.041989722	0.052329603
Sun, Oct 12	0.188703	0.089859	0.147771	0.115924	0.040932765	-0.026064774
Mon, Oct 13	0.183718	0.10451	0.164027	0.089367	0.019691223	0.015143935
Tue, Oct 14	0.186603	0.125598	0.166868	0.111245	0.019734673	0.014352621
Wed, Oct 15	0.194743	0.076464	0.168269	0.112981	0.026473899	-0.036517209
Thu, Oct 16	0.167679	0.099635	0.163706	0.077411	0.003973639	0.022224312
Fri, Oct 17	0.195187	0.101604	0.162821	0.05641	0.032366653	0.045194022
Sat, Oct 18	0.174051	0.110759	0.144172	0.095092	0.029878854	0.015667469
Sun, Oct 19	0.18958	0.086831	0.172166	0.110473	0.017413891	-0.023642778
Mon, Oct 20	0.191638	0.11266	0.177907	0.113953	0.013730654	-0.00129379
Tue, Oct 21	0.226067	0.121107	0.165509	0.082176	0.060557638	0.038931341
Wed, Oct 22	0.193317	0.109785	0.1598	0.087391	0.033517173	0.022394441
Thu, Oct 23	0.190977	0.084211	0.190031	0.105919	0.000946291	-0.021708477
Fri, Oct 24	0.326895	0.181278	0.278336	0.134864	0.048558778	0.046414159
Sat, Oct 25	0.254703	0.185239	0.189836	0.121076	0.064867753	0.064162551
Sun, Oct 26	0.227401	0.146893	0.220779	0.145743	0.006621909	0.00114951
Mon, Oct 27	0.306983	0.163373	0.276265	0.154345	0.030718281	0.009027853
Tue, Oct 28	0.209239	0.123641	0.220109	0.163043	-0.010869565	-0.039402174
Wed, Oct 29	0.265223	0.116373	0.276479	0.13205	-0.011255405	-0.015676041
Thu, Oct 30	0.22752	0.10218	0.284341	0.092033	-0.056820223	0.010146869
Fri, Oct 31	0.246459	0.143059	0.252078	0.17036	-0.005618639	-0.027300621
Sat, Nov 1	0.229075	0.136564	0.204317	0.143885	0.024758343	-0.007321015
Sun, Nov 2	0.297258	0.096681	0.251381	0.142265	0.045877082	-0.045584097
23 days						

总转化率预期是实验组变小的，未达到预期是要统计出 C-E < 0 的天数；净转化率预期是不会减少，未达到预期是要统计出 C-E>=0：

	总转化率 （预期变小）	净转化率 （预期是不会减少）
未达到预期的天数	4	13
双尾检验 P-value	0.0026	0.6776

汇总

使用每日数据对每个评估度量运行符号检验，这里并没有使用 Bonferroni 校正法，因为控制组和对照组有一定的相关性，成此消彼长的关系。

通过符号检验发现，总转化率未达到预期的天数所占比例小，出现的概率小；而净转化率未达预期的天数所占比例大，出现的概率大。

差异的原因可能是因为点击“开始免费试学”然后登录这件事情是一次性的，当天可以完成，而付费在 14 天内任意的区间都可以进行，付费跟弹窗询问后再登录的事件关系不紧密。

建议

该试验不建议实施。由之前的分析，用来验证总转化率是有效的，说明了实验组中的弹窗询问学习的时间，会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。但是，但是净转化率的置信区间包含负数，也就是说在投入人力和时间成本之后，进行该试验之后净转化率可能会减少，实验开始时的期望是该值不降低，即置信区间完全在 0 的右侧。

后续试验

总转化率=免费试学用户 id 数/点击免费试学 cookie 数，该度量验证虽然有效，但似乎是结论又是明显的。因为产生点击免费使用按钮 cookie 时，该事件可能是随机的，也可能是有目的性的，弹框询问学习时间，很大程度上会减少随机行为的继续，会让用户思考自己的学习时间是否足够，避免了低效率的免费试学。但是我们不期望净转化率出现降低的情况，希望从更多角度去探索其他度量。

后续实验想从其他方面入手，实验概括如下：

实验组：点击免费试学并成功登录后，会收到邮件说明在 7 天内学费可以分期付款，7 天后直至免费试学结束期间需要一次性付清学费。

控制组：点击免费试学并成功登录后，免费试学期间，学费需要一次性付清。

分组单位：实验分析的是从用户免费试学登录到付款直接的数据，分组单位是用户 id；

不变度量：点击免费试学并成功登录的用户 id 数，发生在实验之前；

评估度量：从点击免费试学并成功登录到付款的时间间隔均值，发生在实验之后；

评估度量：用户付款 id 数除以免费试学并成功登录的用户 id 数，发生在实验之后；

实验理由：可以用来验证分期付款是否更能在免费试学后尽快决定付费学习；2 可以用来验证分期付款是否导致付费用户的增加。

优达学城 徐海浪

2017 年 6 月 28 日