优达学城数据分析师纳米学位

A/B 测试项目

说明：[点击此处下载此文档的英文版本](https://s3.cn-north-1.amazonaws.com.cn/static-documents/nd002/FinalProjectTemplate.pdf)。

模板格式

此模板可用于组织你最终项目的答案。应从你的答案中复制到小测试中的项用蓝色显示。

试验设计

指标选择

列出你将在项目中使用的不变指标和评估指标。（这些应与你在“选择不变指标”和“选择评估指标”小测试中使用的指标一样）

Choosing Invariant Metrics

不变指标：

1. Cookie 的数量**Number of cookies**：即查看课程概述页面的唯一 cookie 的数量。
2. 点击次数**Number of clicks**：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。
3. 点进概率**Click-through-probability**：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。

#Reason：首先这个实验的units of diversion 分组单位为唯一cookie，因此对于cont组合exp组的cookie数目应保持一致或差不多。因为该实验是在点击开始免费使用按钮会才会出现提示（也就是该实验变量，是否应该加这块提示），所以在提示按钮出来之前，cont组和exp组的Number of clicks 以及 Click-through-probability 应该类似。综上这三点应该保持不变，作为我们的Invariant Metrics。

Choosing Evaluation Metrics

选择评估指标:

1. 总转化率 **Gross conversion**：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。（ dmin =0.01）
2. 净转化率 **Net conversion**：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。（dmin=0.0075）

对于每个指标，解释你为什么使用或不使用它作为不变指标或评估指标。此外，说明你期望从评估指标中获得什么样的试验结果。

#Reason: 这实验的目的是减少因为没有足够时间而离开免费教学的学生数量，从而增加在开始免费教学后完成试学并且留存下来(指付费)的比例。因此总转化率以及净转化率可以作为很好的评估指标，若结果具有统计显著性，则实验组的总转化率应该小于对照组，因为实验组的同学会有一部分在看到提示后选择退出。但是净转化率考察的是在点击开始免费试学后的人数与完成项目的人数，若实验假设正确，净转化率应该不变。同时项目说明认为这个实验不会很大程度减少继续通过免费试学以及最终完成课程的学生数量(以user id分组)，可以肯定的是number of user-ids会减少但是减少的变化很多，不一定具有统计显著性，因此number of user-ids不能很好的作为评估的依据。**留存率**则是因为在后续的计算过程中发现所需的页面浏览数量过多，并导致实验时间过长，故不选择作为评估指标。

测量标准偏差

列出你的每个评估指标的标准偏差。（这些应是来自“计算标准偏差”小测试中的答案。）

对于每个评估指标，说明你是否认为分析估计与经验变异是类似还是不同（如果不同，在时间允许的情况下将有必要进行经验估计）。简要说明每个情况的理由。

#Reason: excel中的数据基于40000个unique cookies per day，题目中给定5000per day，因此需要按比例转换，5000个unique cookies只有400个点击了Start free trial，因此这里的N=400.



标准偏差Standard Error

计算公式：SE = (p\*(1-p)/N)\*\*0.5

1. 总转化率 Gross conversion

SE = (0.2063\*(1-0.2063)/400)^0.5 = **0.0202**

1. 净转化率 Net conversion

SE = (0.1093\*(1-0.1093)/400)^0.5 = **0.0156**

规模

样本数量和功效

题目给定使用 alpha = 0.05，beta = 0.2。将答案约为最近的整数。

说明你是否会在分析阶段使用 Bonferroni 校正，并给出实验正确设计所需的页面浏览量。（这些应是来自“计算页面浏览量”小测试中的答案。）

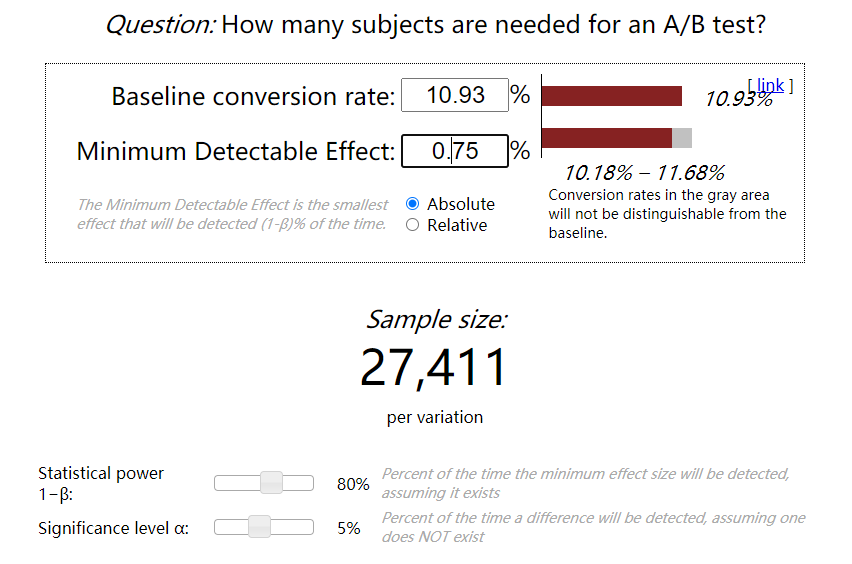
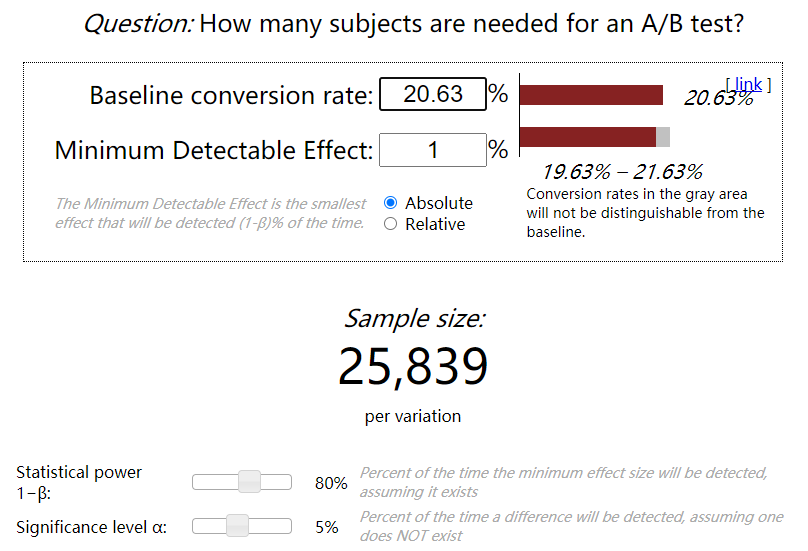
样本规模计算使用[在线样本计算器](https://www.evanmiller.org/ab-testing/sample-size.html)

1. 总转化率 Gross conversion

总转化率的基准转化率为0.2063，dmin=0.01，取α=0.05,1-β=0.8，得到最小的unique pageviews为**25839。**

1. 净转化率 Net conversion

净转化率的基准转化率为0.1093，dmin=0.0075，取α=0.05,1-β=0.8，得到最小的unique pageviews为**27411。**



##注意：如果使用留存率作为评价指标，需要39511个unique pageviews, 过大了

持续时间和暴光比例

说明你会将多少百分比的页面流量转入此试验，以及鉴于此条件，你需要多少天来运行试验。（这些应是来自“选择持续时间和曝光”小测试中的答案。）



上面计算了这两项评估指标(evaluation metrics)若要达到dmin的话最少需要的点击开始免费学习按钮(Start free trial)的的unique cookies的量，再根据上表计算出总共需要的unique cookies的量（在此基础上除以点击概率0.08再乘以2，这里需要考虑测试分为对照组和试验组），按照大的我们认为总共需要685275条unique cookies左右。

说明你选择所转移流量部分的原因。你认为此试验对优达学城来说有多大风险？

#Reason：这里我们将每天40000条流量全部转移进行实验，这样的话每天有40000条ucs，一共需要685275/40000 ≈ 18days

#风险：

此试验并不存在太大风险：（来自于csdn大神分析, 主要从abtesting的四个原则出发：风险，益处，备选方案，数据敏感性以及伦理方面）

1. 被追踪的数据是以Cookie的形式存在，而后续的登录和付费并不会因事件发生而变化。
2. 网站并没有大幅度的改动，用户不需要很长时间去适应新的浏览页面。
3. 并未对数据库进行改动，数据泄露的风向未发生变化
4. 改动仅为一个页面提醒，并使用户造成使用上的困难。
5. 该试验不涉及道德风险。

试验分析

合理性检查

对于每个不变指标，对你在95%置信区间下期望观察到的值、实际观察的值及指标是否通过合理性检查给出结论。（这些应是来自“合理性检查”小测试中的答案）

对于任何未通过的合理性检查，根据每日数据解释你觉得最有可能的原因。**在所有合理性检查通过前，不要开始其他分析工作。**

#不变指标一共三个，cookies的数量，点击次数，点进概率(CTP)



完整性检查：

对于cookies的数量和clicks点击次数应该符合随机的二项分布，也就是说期望的概率应该为0.5，这里我们计算以95%的置信度计算置信区间(lesson5.6)，z\*=1.96。

计算过程：pageviews和clicks的计算比较简单，P\_expected的值都为0.5，同时通过公式SE = (P\*(1-P)/N)^0.5计算出Standard Error，之后乘以Z\*得到Margin，之后得到以P\_expected为中心的置信区间，然后查看我们的观测概率P\_hat是否处在这个区间内。

CTP的计算，这里用cont组作为我们的期望概率，使用exp组作为我们的观测概率。因此CTP\_P\_expected=28378/345543=0.082125814, CTP\_P\_hat=28325/344660=0.082182441。后面的计算方法与前两项一致。



#结论：可以看到三个不变指标都通过了sanity check。

结果分析

效应大小检验

对于每个评估指标，对试验和对照组之间的差异给出 95% 置信区间。说明每个指标是否具有统计和实际显著性。（这些应是来自“效应大小检验”小测试的答案。）

评估指标：

1. 总转化率 Gross conversion
2. 净转化率 Net conversion



这里我们整理了对照组与实验组在总转化率以及净转化率上的变化情况，接下来进行统计显著性以及实践显著性的分析。注：统计显著性指dhat的置信区间不包括0，实践显著性指dhat的置信区间不包括dmin。





#Reason：通过对总转化率以及净转化率的显著性分析，我们发现总转化率的变化具有统计显著性以及实践显著性，而净转化率的变化并不具有统计显著性以及实践显著性，这与我们之前的假设比较符合，这个措施会让总转化率下降，但同时不会明显的影响净转化率。

符号检验

对于每个评估指标，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。（这些应是“符号检验”小测试中的答案。）

使用该链接计算双尾概率：<http://graphpad.com/quickcalcs/binomial1.cfm>

#注意：若双尾概率P<α，则具有统计显著性；反之，则不具有。

1. 总转化率 Gross conversion
2. 净转化率 Net conversion



通过统计，我们发现在23天内，有23天时间实验组的总转化率GC小于对照组的GC，有13天时间实验组的净转化率NC小于对照组的NC，将结果带入到Sign Check网站进行计算。得到结果如下表：



#结论：这与我们的效应大小检验的结果一致！总转化率具有统计显著性，而净转化率没有统计显著性。

汇总

说明你是否使用了 Bonferroni 校正，并解释原因。若效应大小假设检验和符号检验之间存在任何差异，描述差异并说明你认为导致差异的原因是什么。

#Reason：没有使用Bonferroni校正，因为这两个评价指标之间并不是互相独立的而是具有紧密的关联性，若使用Bonferroni会显得太过保守。

建议

提供建议并简要说明你的理由。

A：可以开展这项实验，目前来看实验组与对照组数据的差异比较符合我们在开展实验前的假设，优达学城将改进整体的学生体验和提升导师为能够完成课程的学生提供帮助支持的能力。

后续试验（来自csdn大神）

对你会开展的后续试验进行概括说明，你的假设会是什么，你将测量哪些指标，你的转移单位将是什么，以及做出这些选择的理由。

如果想减少受挫并提前终止课程的学生数量，可以适当增加学生每节课完成后的正面反馈。例如，在学生完成一小节课后给出一些鼓励性提示。因此，我的假设为：

每完成一节课后会出现一条消息鼓励学员继续，最终提高学生的留存率。

度量选择：

1.不变度量

学生ID：此试验是在用户点击试学，并且登录id后，不会影响到登录用户数量，因此选择用户id做不变度量。

2.评估度量

留存率：试验可能会影响最终付费用户数量，留存率是个很好的评估指标。

3.转移单位

学生ID：此测试发生在用户登录后，id会被跟踪，用户id是合适的转移单位。

Yjin232

优达学城

2021年8月