

某电商平台针对优化落地页设计的A/B测试

1. 项目介绍

某电商公司非常注重自己的落地页设计，希望通过改进设计来提高转化率。以往该公司全年转化率平均在13%左右，现在希望设计的新页面能够带来更高的转化率，希望新页面的转化率能有2%的提升，达到15%。在正式推出新页面之前，该公司希望通过AB测试在小范围的用户中进行测试，以确保新页面的效果能够达到预期目标。

2. 数据介绍

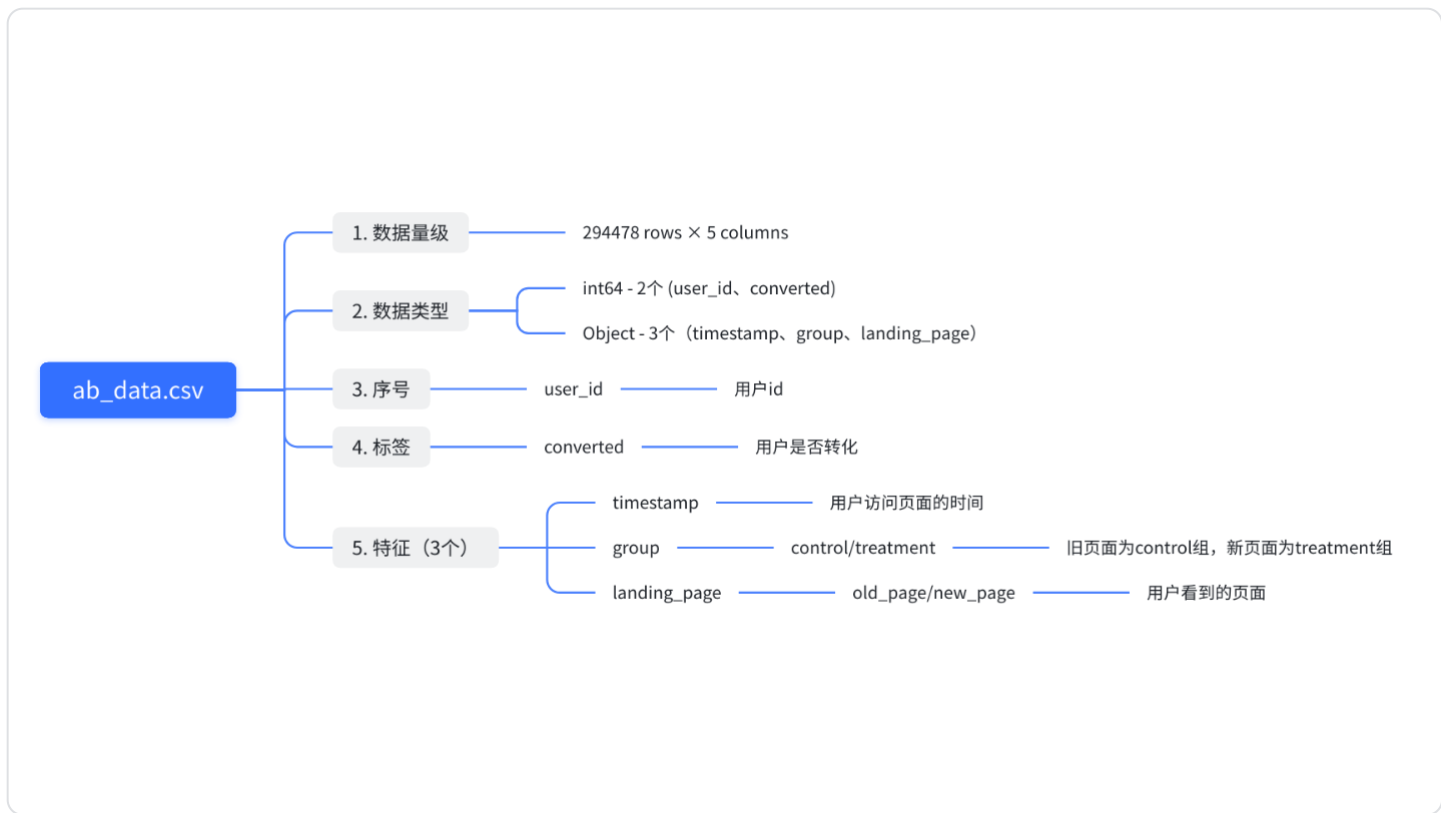
- 数据来源：



<https://www.kaggle.com/datasets/putdejdomthai/ecommerce-ab-testing-...>

Ecommerce AB Testing 2022 Dataset1

Kaggle is the world's largest data science community with powerful tools and resources to help you achieve your data science goals.



3. 基本流程

3.1 确定实验目标及衡量指标

在本项目中实验目标是通过AB测试确定新落地页是否可以提升2%的转化率。衡量指标即为**页面转化率**

3.2 设计实验方案

设计实验的具体方案，包括实验变量（即页面的具体改动点）、实验时间、实验样本量、划分实验组和对照组等。还需要确定实验数据的收集方法和分析方式。

【注】一般分析方式包括描述性统计、假设检验、置信区间估计、回归分析等等。在进行AB测试的时候，需要分析师根据实验的目标和具体情况选择合适的分析方法，并结合实际情况进行解读，提出合理的建议和决策。

3.3 执行实验并收集数据

按照实验方案进行实验执行，同时记录实验数据。

3.4 数据分析和结果评估

对实验数据进行分析 and 评估，包括对实验结果的显著性检验、效果评估等。最后给出结论和建议。

4. 设计实验

4.1 提出假设

设计AB test实验的第一步通常是提出假设。假设是对于某个特定变化我们所期望的结果，也是后续实验的基础，我们需要在后续实验中通过数据验证这个假设是否成立。如果验证成立，我们可以将这个变化推广到全部用户。如果验证不成立，则需要继续优化这个假设或者放弃这个修改方案，以寻找更好的变化。

在这个实验中，我们希望新页面可以提升2%的转化率，原则上我们应选择单尾检验，准确的说，**应该选择右侧单尾检验，因为我们的假设是新页面的转化率要大于旧页面的转化率。但是，在本实验中，我们并不能确定新页面的性能一定比当前的页面更好。所以，这里选择双尾检验。**

【补充说明】单尾检验和双尾检验是假设检验中的两种常见形式。它们的区别在于对假设的方向性不同：

- 在单尾检验中，研究者针对假设提出了明确的方向性，例如“新设计的落地页转化率比原设计高”、“将按钮颜色从灰色改为红色将导致更多的点击”……因此只检验这个方向是否具有统计显著性，被称为单侧检验。
- 而在双尾检验中，研究者不确定假设所在的方向，例如“新设计的落地页转化率与原设计是否存在差异”、“改变页面布局会影响用户的满意度”……需要同时检验两个方向的统计显著性，被称为双侧检验。

在AB测试中，一般应根据实验目的、数据分布情况和统计分析方法等因素来确定单尾检验或双尾检验。例如，如果实验的目的是验证新设计的落地页转化率是否明显高于旧设计，就可以选择单尾检验。但如果不确定两种设计是否存在差异，则应选择双尾检验。

简单来说，可以这样辨别单尾/双尾检验：

- 如果备择假设 H_1 中是“ \neq ”，则是双尾检验；
- 如果备择假设 H_1 中是“ $>$ ”或者“ \geq ”，则是右侧单尾检验；
- 如果备择假设 H_1 中是“ $<$ ”或者“ \leq ”，则是左侧单尾检验。

此外，需要特别注意的是：**想支持的结论通常放在备择假设上。**所以在本案例中，原假设和备择假设分别是：

- 原假设 H_0 ： $P = P_0$
- 备择假设 H_1 ： $P \neq P_0$

其中， P_0 代表的是旧版落地页的转化率， P 代表的是新版落地页的转化率。

4.2 确定实验分组

在此次AB测试中，我们分为实验组和对照组两组：

- 对照组（control组）：这一组用户将看到旧版落地页
- 实验组（treatment组）：这一组用户将看到新版落地页

为了后续计算每一组的转化效率，需要记录每一位参与实验的用户的购买行为，也就是说无论用户看到的是新版落地页还是旧版落地页，都需要记录这位用户最终是否购买了产品。可通过在网站上添加

相应的追踪代码来实现：

- 0：代表用户在测试期间没有购买产品
- 1：代表用户在测试期间购买了产品

这样，后续就可以很容易地计算出每个组的均值，从而得到新旧两版落地页的转化率。

4.3 计算实验样本量、实验周期

1. 实验样本量的确定

根据大数定律和中心极限定理，样本量越大我们的估计就会越精确，但同时也意味着成本越高，所以需要知道实验所需的最小样本量是多少，在成本可控范围里，选择合适的样本量即可。

每一个实验组所需样本量计算公式如下：

$$N = \frac{\sigma^2}{\delta^2} (Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2$$

在这个公式当中， α 为犯第一类错误的概率， β 为犯第二类错误的概率， σ 代表的是样本数据的标准差， δ 代表的是预期实验组和对照组两组数据的差值。一般情况下，我们会设置：

- 显著性水平： $\alpha=0.05$ ，即在拒绝原假设之前，我们有95%的把握新版落地页的转化率比旧版落地页要高
- 统计功效（ $1-\beta$ ）： $\beta=0.2$ ，即表示测试检测特定效果的能力，如果该特定效果存在的话。在此案例中就是，如果新版落地页真的比旧版转换率要高，该测试有80%的概率能检测出这个状况。

当衡量指标为比率类指标时，标准差计算公式为：

$$\sigma^2 = P_A(1 - P_A) + P_B(1 - P_B)$$

其中， P_A 和 P_B 分别是对照组和实验组的观测值。在此案例中 $P_A=13\%$ ， $P_B=15\%$

因此，每个组所需的最小样本量为：

$$\begin{aligned} N &= \frac{\sigma^2}{\delta^2} (Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 \\ &= (0.13 * (1 - 0.13) + 0.15 * (1 - 0.15)) / (0.15 - 0.13)^2 * (1.96 + 0.84)^2 \\ &= 4716 \end{aligned}$$

2. 实验周期的确定

根据上面最小样本量的计算，我们知道此次AB测试至少需要9440个用户参与测试，假如该落地页以往每天的平均浏览量为1000，则实验周期至少需要的天数为：

试验周期=9440/1000=9.4≈10（天）

4.4 上线AB测试并收集数据

目前市面上大家熟知的大公司基本上都在做AB测试，比如：百度、阿里、腾讯、字节跳动、京东、滴滴、携程、美团等：

- 百度：百度统计可视化A/B测试：
<https://baijiahao.baidu.com/sid=1735048932168957887&wfr=spider&for=pc>
- 字节跳动：巨量引擎「AB实验工具」：<https://zhuanlan.zhihu.com/p/508366232>
- 阿里：淘宝推出的 A/B 测试平台，主要针对淘宝商家进行 A/B 测试
- 腾讯：腾讯云AB实验平台：<https://abtest.qq.com/>

大公司做AB测试的主要特点就是。一般都是自研系统。而对于规模不是那么大的企业，也有做AB测试的需求，比如互金、运动、在线教育、SaaS的都有做A/B测试，但是这部分企业一般是使用第三方的A/B测试工具。目前市面上的第三方AB测试工具主要有：

- 神测数据：<https://www.sensorsdata.cn/features/ABTesting.html>
- ABtester：提供web网站A/B测试<http://www.abtester.cn/>
- 热云数据：支持App、web、H5的A/B测试<https://www.appadhoc.com/>

4.5 数据分析及假设检验

某电商平台针对优化落地页设计的A/B测试

1. 确定实验样本量

```
In [1]: # 使用Python完成最小样本量计算
import numpy as np
import pandas as pd
import scipy.stats as stats
import statsmodels.stats.api as sms
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: # 根据我们的预期比率计算效果量
effect_size = sms.proportion_effectsize(0.13, 0.15)

In [3]: effect_size

Out[3]: -0.0576728617308947
```

