

# 落地页 A/B 测试 — 迷你备忘录

## 1. 总结

- a. 研究问题：验证新广告落地页是否在不恶化体验/成本的前提下显著提升核心转化率（CVR），并据此评估对单位经济学与业绩弹性的影响。
- b. 关键证据：
  - i. Treatment:  $n=145,310$ ,  $CVR=11.881\%$ ; Control:  $n=145,274$ ,  $CVR=12.039\%$ ;  $\Delta=-0.16$  个百分点 (pp)
  - ii.  $Z=-1.311$ ,  $p=0.1899$ ; 差值 95%CI 约  $[-0.39 \text{ pp}, +0.08 \text{ pp}]$
  - iii. 本次实验在 80% power、 $\alpha=0.05$  下的最小可检出效应 (MDE)  $=0.338\%$  pp。
- c. 主要结论：证据不足以支持“新页面提升转化”的假设，观测到的差异处于统计噪声区间。
- d. 决策与投研含义：本次实验  $\Delta CVR$  约  $-0.16\text{pp}$  ( $p=0.1899$ )，效果与噪声难以区分，因此不全量上线，保持旧版或小范围灰度观察；若想验证  $\leq 0.2\text{pp}$  的微小提升，需扩大样本/延长周期（当前实验的  $MDE=0.338\%\text{pp}$ ）。故不据此上调短期收入或利润预期，但可视为公司具备纪律化的实验文化与止损机制的正面信号；后续重点跟踪页面性能与投放 ROI、以及留存/复购等长期指标。

## 2. 设计与口径

- a. 分流与实验期：按用户 ID 1:1 随机分流 (Treatment / Control)，实验期为 [2017/01/02-2017/01/24，统一时区]；分流全程固定，不跨组、不复用。
- b. 样本/观测口径：仅纳入广告点击后成功到达落地页的有效访问；重复刷新/异常请求剔除，确保两组采集条件一致。数据来源为模拟公开数据集；指标定义与采集逻辑已固定于实验启动前。
- c. 主指标： $CVR = \text{支付成功的“独立用户数”} / \text{有效访问的“独立用户数”}$ ，用户级去重。
- d. 统计方法：双样本比例 Z 检验（双侧， $\alpha=0.05$ ），报告差值的 95%CI、Z 与 p；在当前样本量下给出  $MDE(80\% \text{ power}) = 0.338\% (\text{pp})$ 。
- e. 一致性与分层：按国家（如 CA/US/UK）及渠道/设备做分层复核。由于各组样本量明显低于总体，统计功效有限，加之存在多重比较风险，因此本次分组结果仅用于探索性趋势观察，不作显著性判断，也不做 Bonferroni 校正，需后续在目标样本

中进一步验证。

- f. 实验质量：本次实验未单独设置 A/A 检验或执行 SRM 检查，默认采样稳定，实验两组分流比例与预期接近；后续可补充相关检验以提升实验严谨性。

### 3. 结果

本次实验共纳入样本近 29 万人次（Treatment 组：145,310；Control 组：145,274）。主指标 CVR 在 Treatment 组为 11.881%，Control 组为 12.039%，绝对差值为 -0.158 个百分点（pp），方向为 Treatment 略低。采用双样本比例 Z 检验得出： $Z = -1.311$ ， $p = 0.1899$ ，差值的 95% 置信区间约为  $[-0.39pp, +0.08pp]$ ，说明两组间差异不具有统计显著性，无法拒绝原假设。结合当前样本规模估算，本实验在 80% 检验功效下的 MDE 约为 0.34pp，实际观察到的差异低于该阈值，位于统计噪声范围内。

进一步按国家（US/CA/GB）进行分层分析：美国组样本量为 203619，检验结果为  $Z = -1.505$ ， $p = 0.1323$ 。英国组样本量为 72,466， $Z = 0.475$ ， $p = 0.6349$ 。加拿大组样本量为 14,499， $Z = -1.297$ ， $p = 0.1947$ 。

所有国家的分层检验均未达到显著水平。由于分层样本量有限、检出功效不足，分层结果仅作探索性参考，不作为可采信的上线决策依据。未来若要把分层结果作为结论，需要预先设定分层维度，并在更大样本里复验。

### 4. 业务与财务映射

本次实验中，新页面的转化率比旧页面低约 0.16 个百分点，且结果在统计上不显著。即，如果把这种差异放大到业务层面，对整体成交额和收入的影响大概率非常有限。按照经营逻辑：成交额 = 流量 × 转化率 × 客单价，转化率变化直接决定了同样流量下能带来的订单量。假如日均有 100 万次访问，客单价 100 元，那么转化率下降 0.16%，大约意味着每天少了 1600 单、约 16 万元成交额。但由于置信区间跨零，这种变化可能是正的，也可能是负的，所以目前还不足以作为“收入会提升/下降”的依据。如果未来在更多流量和更长时间内验证出稳定的正向效果，才能真正体现为收入和利润的改善。

### 5. 风险与后续

本次实验的结果显示转化率差异较小，统计区间跨零，因此存在一定的不确定性。样本量虽然较大，但仍可能不足以识别细微效应，尤其在分国家后，样本量进一步

下降，导致检出能力有限，因此分层结果仅作趋势性参考。同时，本实验主要聚焦于转化率，未同步观测退款率、页面加载速度、留存等长期或守门指标，存在指标覆盖不全的风险。

后续可考虑延长实验周期或增加样本量，以提高对微弱差异的检出力；在设计阶段提前设定主要指标和关键分层，减少事后拆分带来的统计偏差；同时补充对退款、性能、留存等守门与长期指标的观测，确保对整体业务影响的评估更加全面。最后，在后续实验中引入 A/A 检验与样本比例匹配（SRM）检查，以进一步保障分流与数据采集的可靠性。

图表摘要

