



# A/B测试驱动产品演进

陈冠诚

Testin VP & A/B测试首席顾问

01

## IBM Research

Research Staff Member

创新、大数据、云计算、机器学习、OpenPOWER、学术论文、发明、专利

02

## OneAPM

大数据首席架构师&技术总监

SaaS创业、产品、APM, Spark, Druid, 开源社区、机器学习、大数据

03

## Testin

Vice President

A/B测试、产品、增长黑客、数据分析、人工智能、创业、创新

# 目录

## CONTENTS

- 01 什么是A/B测试？
- 02 产品何时需要A/B测试？
- 03 如何做A/B测试？
- 04 关于Testin A/B测试

# 案例1：价值5,700万美金的A/B测试

## ORIGINAL

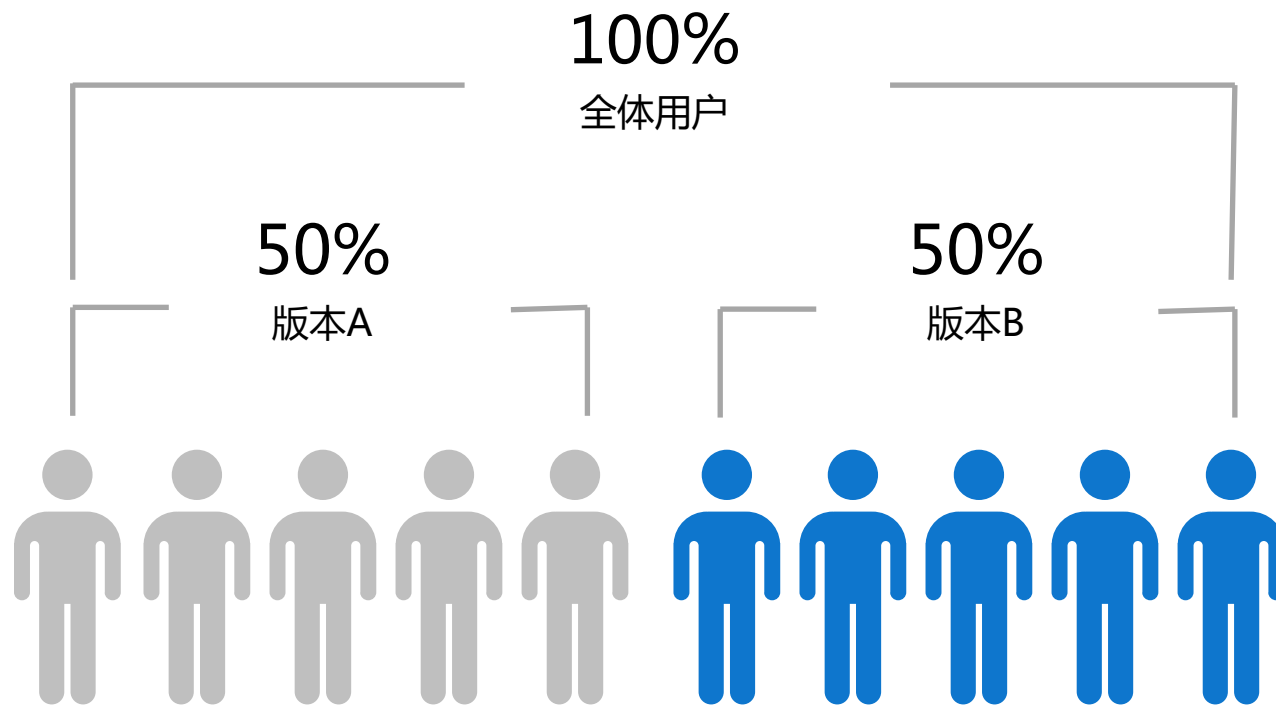


## VARIATION



+40.6%  
\$5,700万

A/B 测试是一种产品优化的方法，为同一个优化目标制定两个方案（比如两个页面），让一部分用户使用 A 方案，同时另一部分用户使用 B 方案，统计并对比不同方案的转化率、点击量、留存率等指标，以判断不同方案的优劣并进行决策。



# 目录

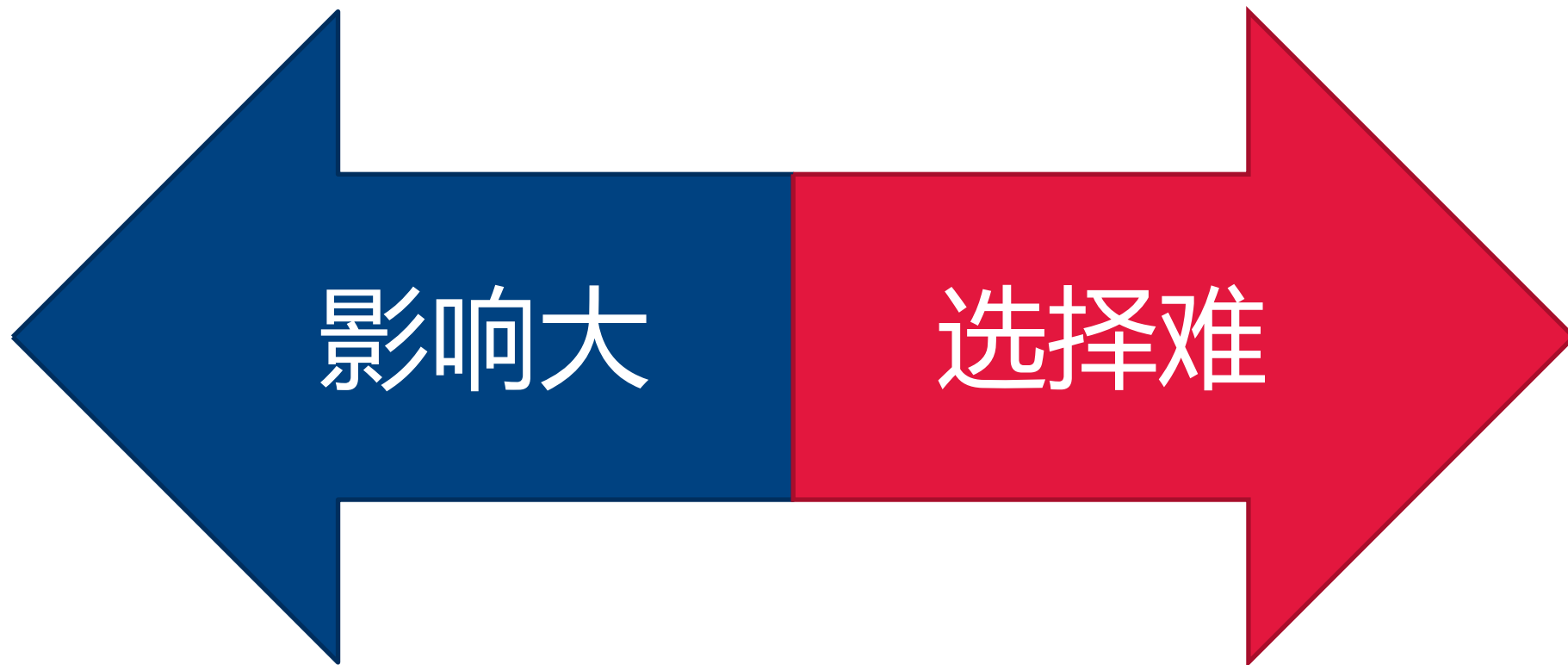
## CONTENTS

- 01 什么是A/B测试？
- 02 产品何时需要A/B测试？
- 03 如何做A/B测试？
- 04 关于Testin A/B测试



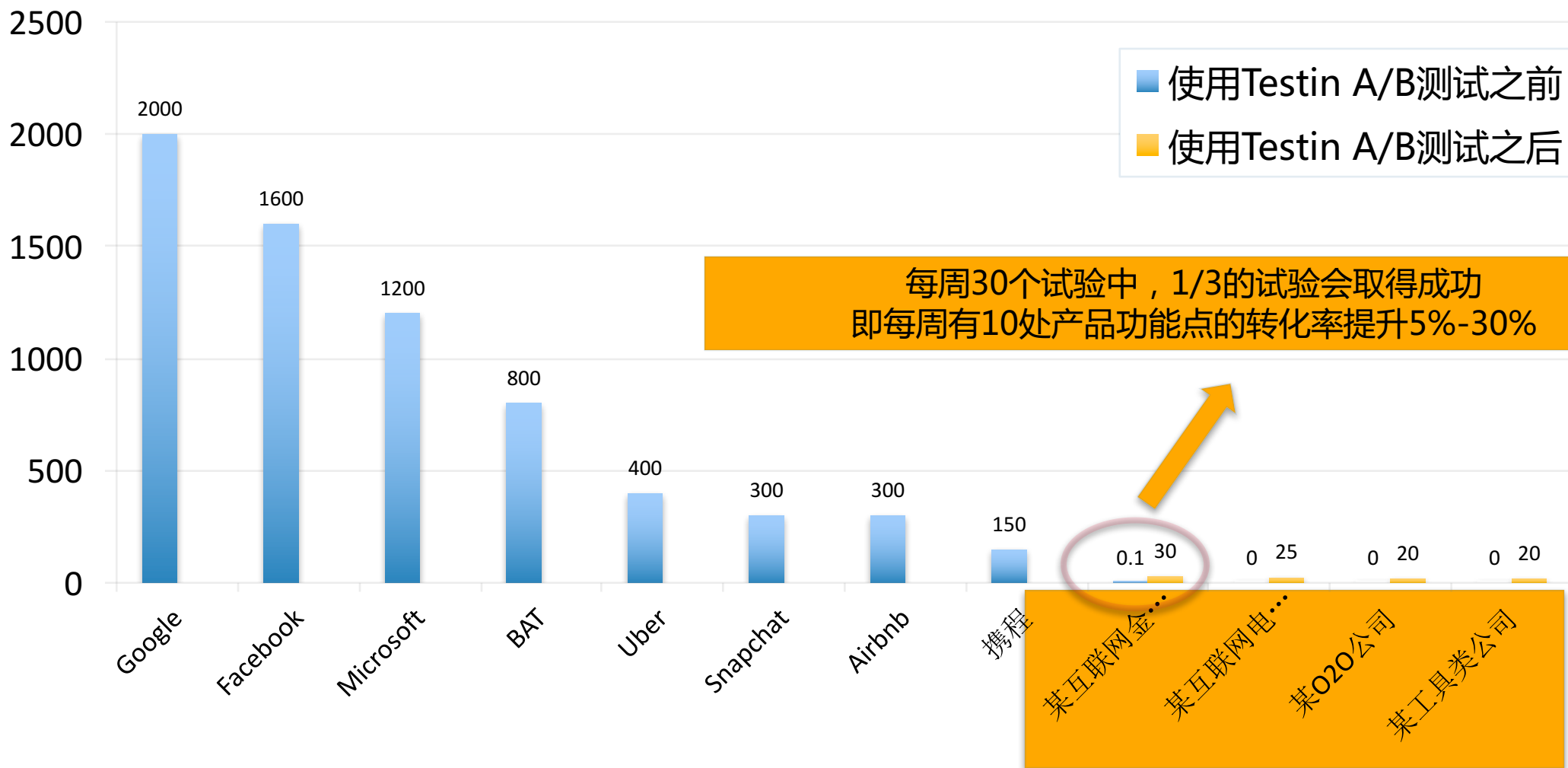
要不要上微信支付??

To be or not to be?





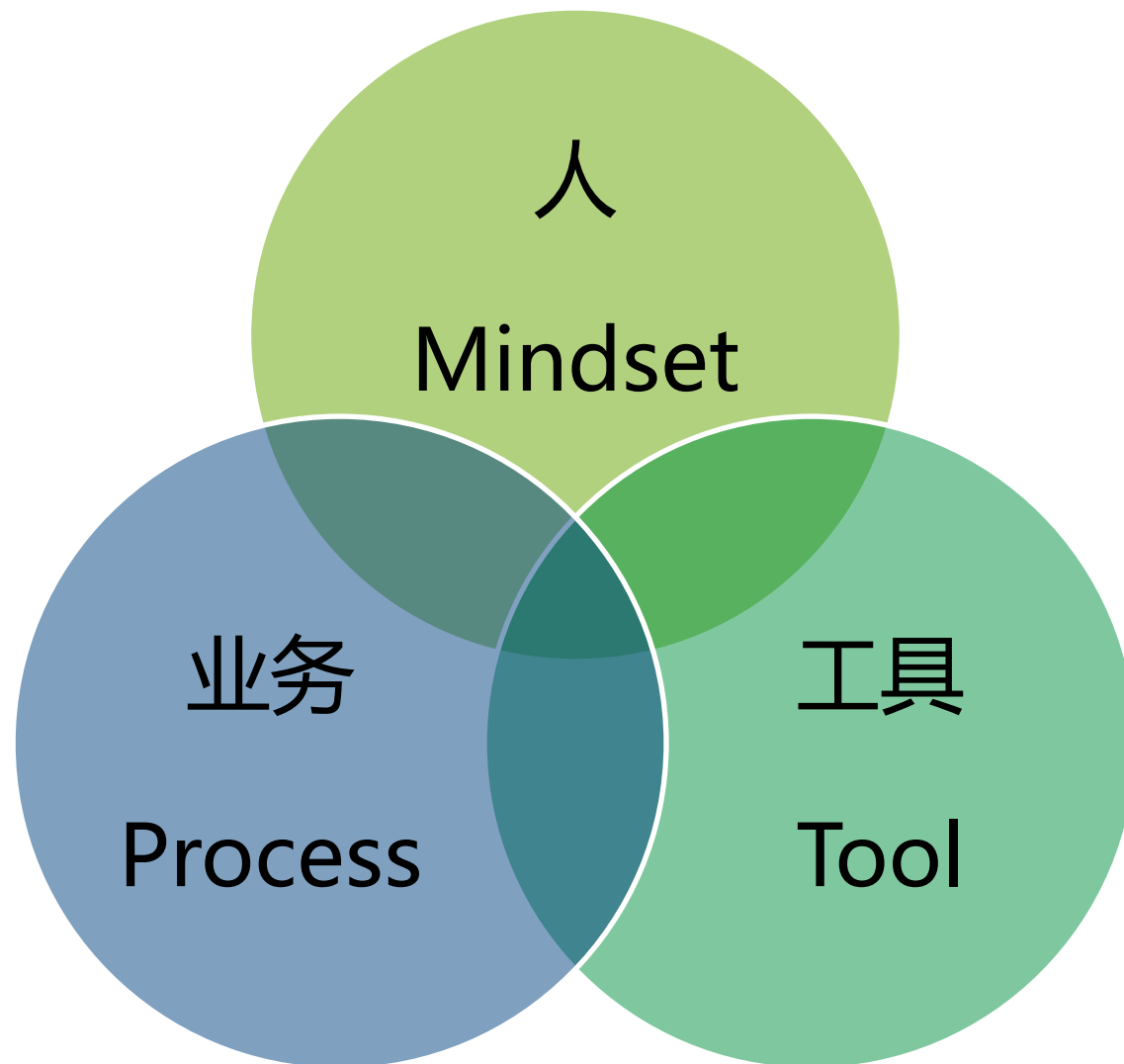
# 典型公司的A/B测试频率（每周）

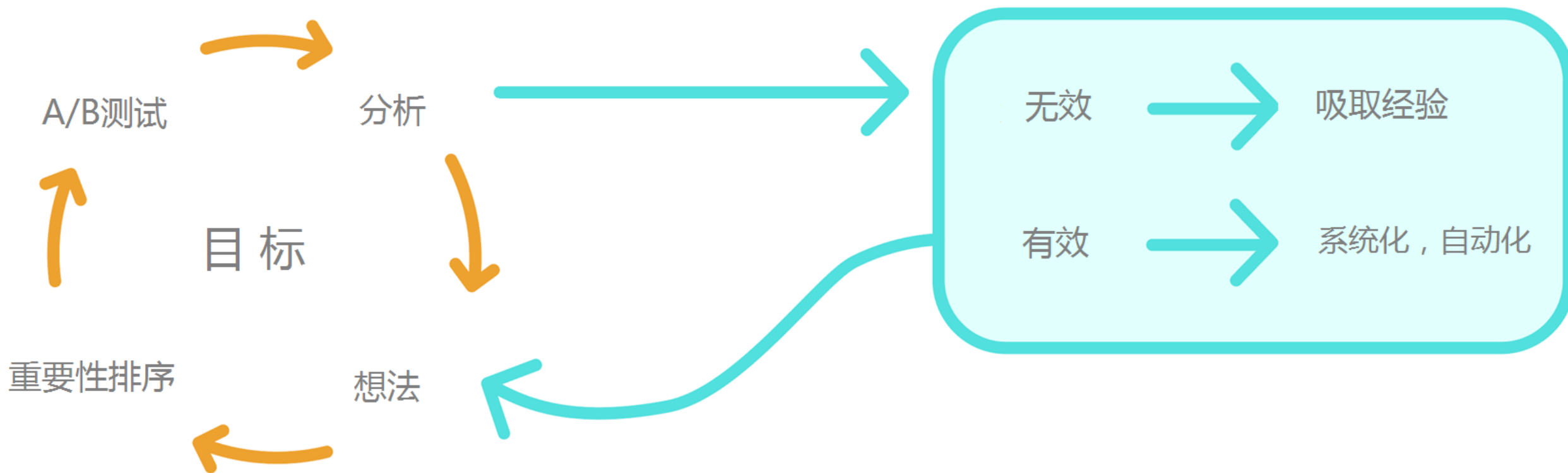


# 目录

## CONTENTS

- 01 ● 什么是A/B测试？
- 02 ● 产品何时需要A/B测试？
- 03 ● 如何做A/B测试？
- 04 ● 关于Testin A/B测试





	I 起步阶段 0-1 Test/周	II 成长阶段 1-3 Test/周	III 成熟阶段 3-10 Test/周	IV 大规模应用 > 10 Test/周
组织架构的特点	开始尝试做A/B测试	开始在主要产品功能改版流程中引入A/B测试	A/B测试成为产品迭代流程的一部分，有系统的A/B测试指标评价体系	有适用于多产品线的标准化A/B测试流程，包括系统化的指标体系、实验平均体系、OEC改进体系
技术架构的特点	没有成型的A/B测试实验平台，用最简单的分流方式和数据分析能力进行实验	有成型的A/B测试系统，有做A/A测试和统计功效分析	开始需要可视化A/B测试，后端A/B测试等高级功能，以便满足多样的A/B测试需求	A/B测试系统成为一个公共平台，能适应多种场景、多业务需求
实验评价体系的特点	有一个最基本的指标评估原则，例如转化率	从单一指标演进为多维度的指标体系，系统跟踪实验对产品多方面的影响	经过大量实验，总结并形成一个系统的OEC评价方法，由多项业务指标加权得到	OEC稳定，半年或者一年才变动一次，并被用于指导组织的整体业绩

# 目录

## CONTENTS

- 01 ● 什么是A/B测试？
- 02 ● 产品何时需要A/B测试？
- 03 ● 如何做A/B测试？
- 04 ● 关于Testin A/B测试



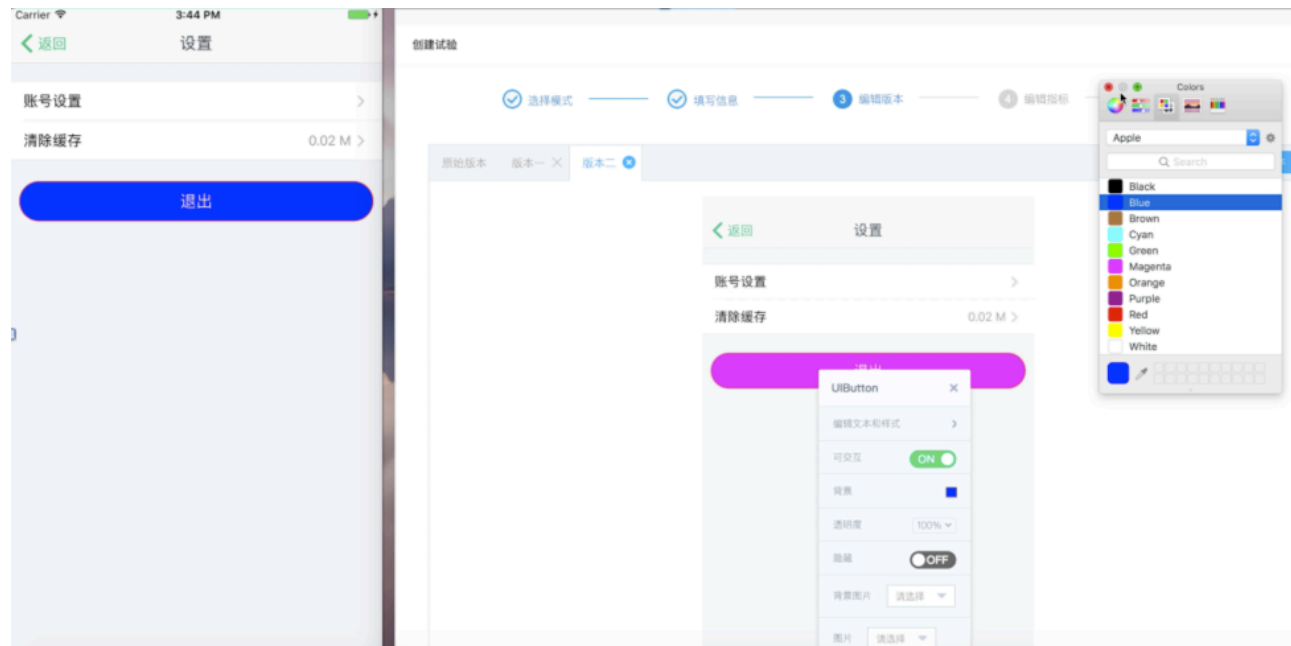
## • 优点

- » 不需要开发，所见即所得
- » 实时更新，无需应用市场审核



## • 缺点

- » 只适合小改版（控件的样式，文案等）
- » 不适合大改版（例如流程的改版）





应用概况

应用配置

流量趋势 ?

流量占用状态

流量走势图 ?

昨日全部流量

0

运行中的实验

0

历史累计流量

0

历史运行过的实验

0

默认层



实验列表

共计: 0个实验

创建实验

实验模式

实验层

实验状态

全部用户

按创建时间由后到先

空  
(0个实验)



GSD\_IOS



相册



收藏



钱包



卡包



表情



设置



微信



通讯录



发现



我

## • 优点

- » 大小改版都适用
- » 灵活、可配置
- » 能结合后端业务逻辑

## • 缺点

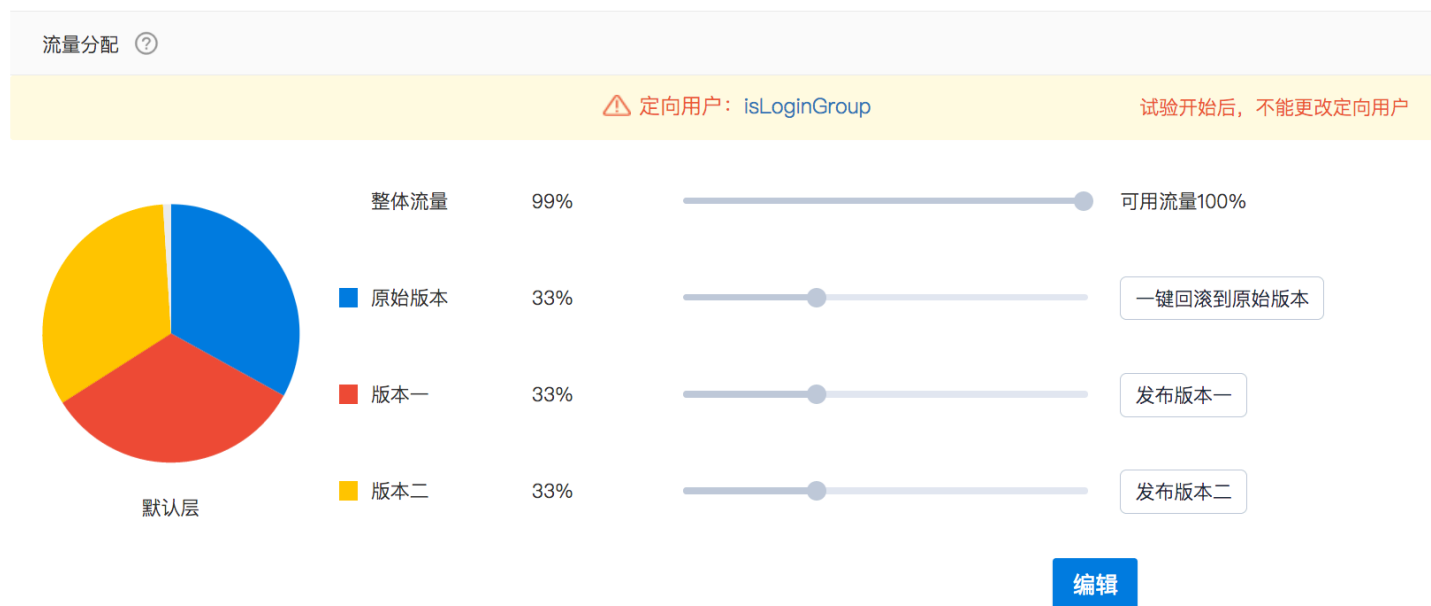
- » 需要开发配合
- » 需要通过应用市场审核



```
//获取int类型的试验变量themeColor的值
private static final int BLUE=1;
private static final int RED=2;
private static final int YELLOW=3;
int color=TestinApi.getIntegerFlag("themeColor",BLUE);

if(color==BLUE){
    //使用默认主题
}else if(color==RED){
    setTheme(R.style.Theme_ZhihuDailyPurify_Light_NoActionBar_Red);
}else if(color==YELLOW){
    setTheme(R.style.Theme_ZhihuDailyPurify_Light_NoActionBar_Yellow);
}
```

- 唯一性
  - » 通过精准且高效的Hash算法，确保单个用户每次登录应用时被分到的试验版本是唯一的
- 均匀性
  - » 分流的人群，各维度分配比例均匀
- 灵活性
  - » 用户可以随时在试验的进行过程中调节试验版本之间的流量分配比例
- 定向性
  - » 可以根据用户标签来实现精准定向分流
    - 设备标签 (SDK自动采集)
    - 用户标签 (需要上传)
- 分层分流
  - » 满足并行做大量A/B测试的需求



- 科学统计

- » 使用了科学的统计分析方法来对试验数据进行分析，并给出可靠的试验结果

- 区间估计

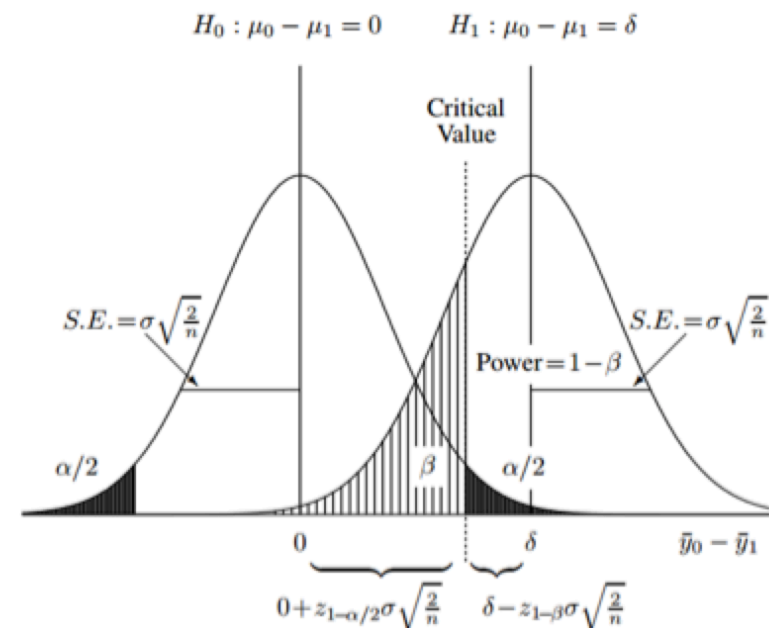
- » 使用区间估计，给出了95%置信区间，因此避免了点估计带来的决断风险

- 统计显著性判断

- » 通过p-value来判断不同试验版本之间差异显著性

- 精益分析

- » 对试验数据进行去噪音处理，去除噪音数据，以提高统计结果的质量





原始版本



新版本

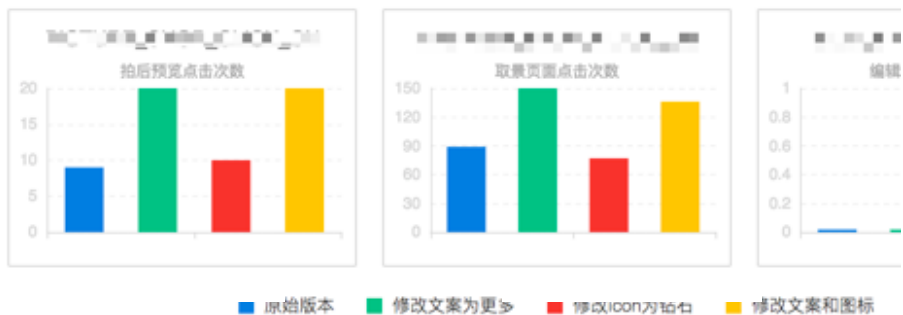
版本	UV	人员比例	签到按钮人均点击次数	点击量相比原始版本的增减幅	95%置信区间	p-value	Power
原始版本	27471	5%	0.69				
新版本	27661	5%	0.72	+4.17%	[+1.7%,+6.6%]	0.03	100%

- 试验目标：期望通过更改签到按钮的文案提高签到的人数，从而提高留存率
- 签到按钮文案试验，流量第一天两个版本各5%，第二天无错误后调节成50%原始版，50%新版本
- 新版本签到按钮平均点击次数高于原始版本
- 新版本的签到次数比原始版本签到次数提高**4.17%**
- 95%置信区间结果显示小范围人群的试验结果推广到全量用户之后，有95%概率获得1.7%至6.6%的提升；p-value小于0.05，显示新老版本有显著统计差异

## 试验概况

试验报告

均值 总值 转化人数 转化率



## 流量走势图

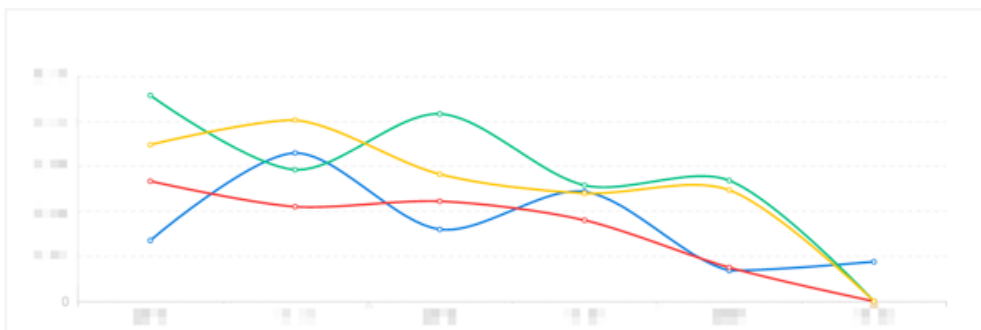
按时 按日

03/17 — 03/2



均值 总值 转化人数 转化率

按时 按日



## 统计分析

下载数据(CSV)

版本	UV	均值	变化度[95%置信区间]	p-value
原始版本	9198	90	0.00978474	
多	9281	150	+65.18% [8.15%,122.20%]	0.01 效果明显优于原始版本
石	9258	84	0.00907323	
标	9298	139	0.01494945	



A/B测试  
就找Testin!