

# AB测试：实验流程

1.实验流程

2.分配流量

  2.1正交试验

  2.2互斥试验

  2.3流量拆解

3.分流算法

4.最小样本量

  4.1比率类

  4.2均值类

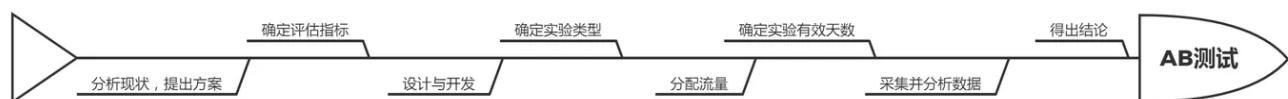
  4.3影响因子

5.实验有效天数

更新版本	时间	作者	内容
1.00	2025.06.30	胡鸿鑫	创建文档

**业务背景：**业务需要更多的实验来验证优化和得到结果，而更高效的实验能够满足业务需求；开发和验证一套更高效的实验方法是必要的

## 1.实验流程



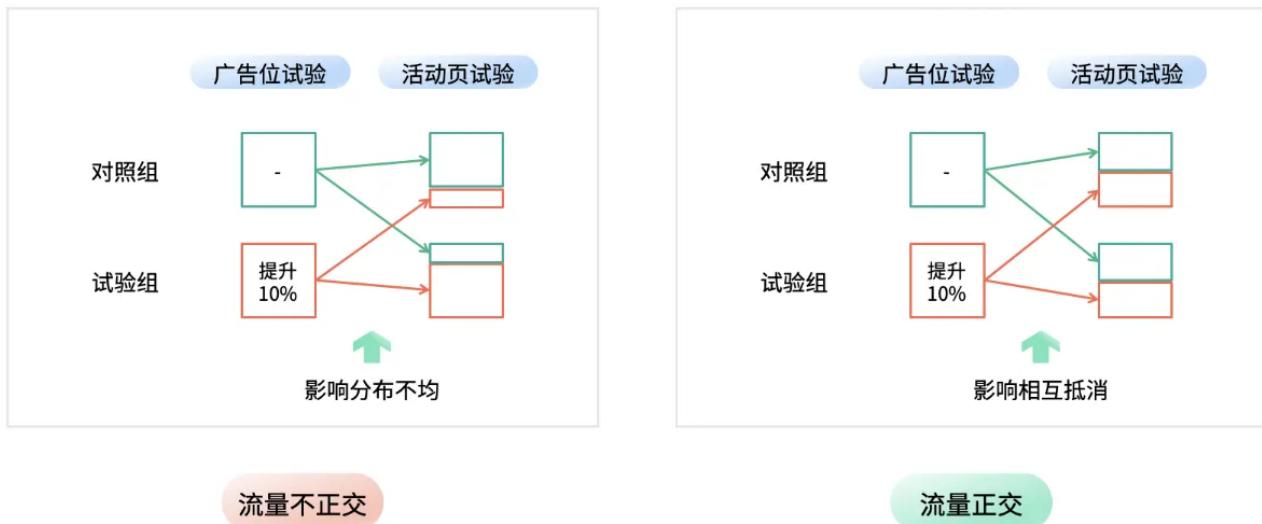
对于部分关键的点位，作出进一步的阐述

## 2.分配流量

由于流量有限，且使用全部流量进行实验对于部分产品来说影响较大，为了保证高效使用流量需要利用分层和分流的流量分配机制。

而根据不同的实验咋在共享流量的情况下，不同的实验之间是否会产生干扰，将实验类型分为正交实验和互斥实验；

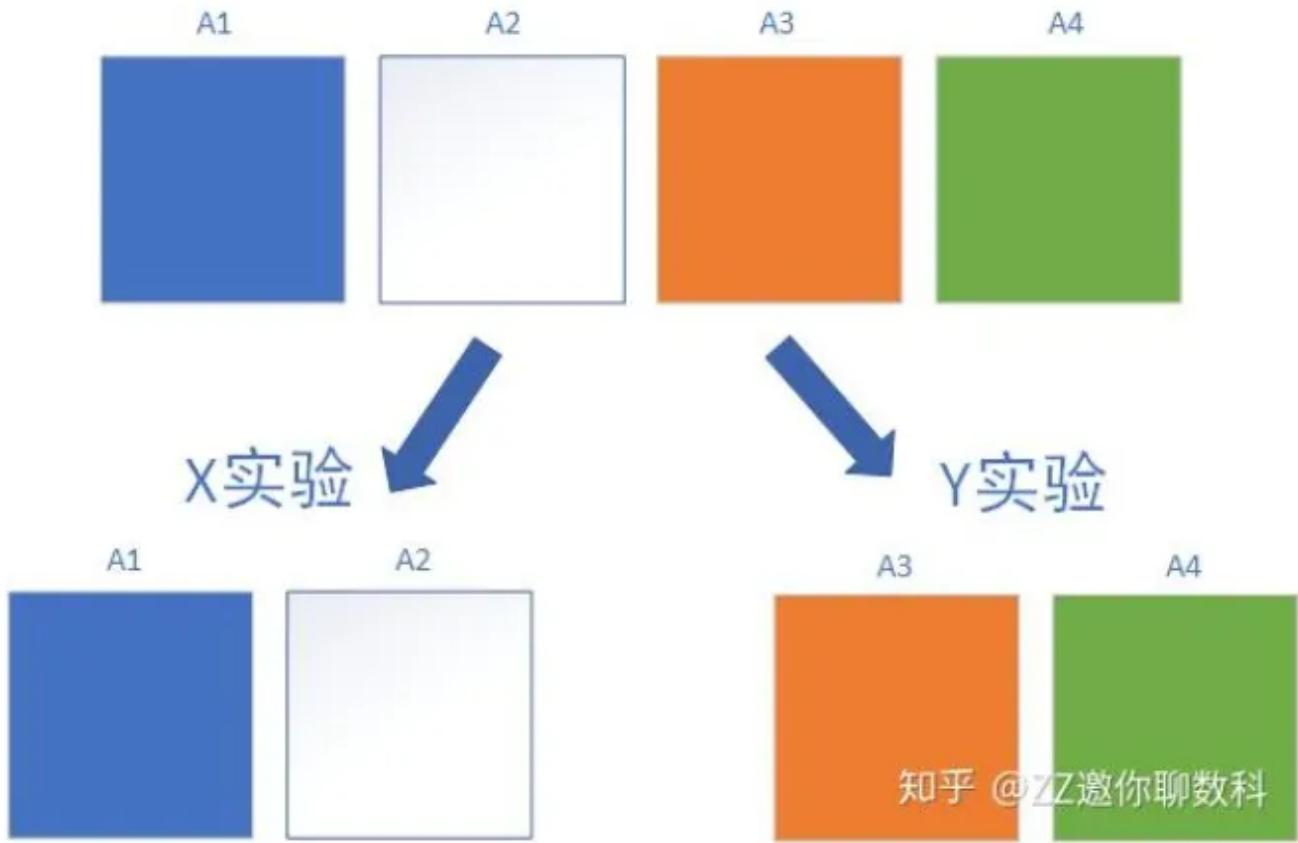
### 2.1 正交试验



正交实验中，影响的因素没有必然联系（需要证明独立性）；

正交的本质在于不同实验组对于实验影响是均等的，即影响是对单层实验组相同，核心思想还是控制变量

### 2.2 互斥试验



互斥实验在于用户只能进入单一实验，二者只能2选1；通过隔离区保证实验的可信度

### 2.3 流量拆解

实验往往是多组实验同时进行，即有正交。也有互斥；

将流量按照以下层级进行划分：流量→域→层

### 3. 分流算法

为了实现正交和互斥实验的随机均匀分流，使用Hash算法：

这里以Firebase的Remote config的AI回答为例，hash算法的核心在于确保 **用户体验的一致性** 和 **测试结果的准确性**；每个用户将会获得固定的哈希值，通过 murmurhash2算法将hash因子放大产生差异化，然后对100求余后+1，最终得到1~100数值

### 4. 最小样本量

业务背景：计算最小样本量能够满足统计学上面的显著差异，其中最小样本量的计算分为：均值类和比率类

## 4.1 比率类

首先需要明确以下数据：

FB提供公式：

$$N = \frac{[Z_{\alpha/2} \sqrt{2P(1-P)} + Z_{\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)}]^2}{(P_1 - P_2)^2}$$

知乎ZZ提供：[最小样本量计算](#)

其中k为 样本量之比

$$n_A = \kappa n_B \quad \text{and}$$

$$n_B = \left( \frac{p_A(1-p_A)}{\kappa} + p_B(1-p_B) \right) \left( \frac{z_{1-\alpha/2} + z_{1-\beta}}{p_A - p_B} \right)^2$$

这里使用：

7留为7%， 提升7留为9%， 即MDE（期望绝对提升）为2%；

- $\alpha$  (显著性水平) = 0.05 →  $z_{1-\alpha/2} = 1.96$
- $\beta$  (统计功效) = 0.8 →  $z_{1-\beta} = 0.84$

代入公式得到：2881人数，即验证1：1分流AB测试需要B组人数为2881人数（A组同样）

## 4.2 均值类

知乎zz提供：（同上）

$$n_A = \kappa n_B$$

$$n_B = \left( 1 + \frac{1}{\kappa} \right) \left( \sigma \frac{z_{1-\alpha/2} + z_{1-\beta}}{\mu_A - \mu_B} \right)^2$$

其中k同上，为分组比， $\sigma$ 为历史数据的标准差；也可以直接使用经验公式：15.86

$$n = 16 \frac{\sigma^2}{\delta^2}$$

### 4.3影响因子

- 统计功效&MDE：通常越小的MDE需要更大的用户量
- 多指标：为每个样本计算样本量，并使用所需的最大样本量
- 避免结果偷看：在没有达到计算出的样本量之前

## 5.实验有效天数

主要考虑以下两个因素：

- 实验需要多少天能达到流量的最小样本量
- 用户的行为周期和适应期