

# AB测试：实验结果分析

## 1.假设检验

### 1.1基本概念

#### 1.1.1原假设H0与备择假设H1

#### 1.1.2检验统计量

## 2.实验结果分析

### 2.1P值

### 2.2两类错误和统计功效

#### 2.2.1两类错误

### 2.3统计功效

### 2.4置信区间

### 2.5实验结果分析体系

| 更新版本 | 时间         | 作者  | 内容   |
|------|------------|-----|------|
| 1.00 | 2025.07.01 | 胡鸿鑫 | 创建文档 |
|      |            |     |      |
|      |            |     |      |

在完成了实验后，我们得到了关键的实验数据，但是对于波动的数据，肉眼观察并不能带来更好的数据反馈，需要有更加科学的方法来检验数据

## 1.假设检验

定义：假设检验是先对总体的参数提出某种假设，然后利用样本数据判断假设是否成立的过程

### 1.1基本概念

#### 1.1.1原假设H0与备择假设H1

我们常把没有把握不能轻易肯定的命题作为备择假设  $H_1$ ，而把没有充分理由不能轻易否定的命题作为零假设  $H_0$ ，或者说我们将希望通过实验结果推翻的假设记为零假设。

原假设和备择假设是一个完备事件组，而且相互对立。在一项假设检验中，原假设和备择假设必有一个成立，而且只有一个成立。

### 1.1.2 检验统计量

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}, \text{ 其中 } n \text{ 为实验参与人数, } p \text{ 为成功人数除以参与人数的百分比, 其中 } Z$$

的临界值为1.96

- Z绝对值大于临界值，即拒绝零假设，差异在统计上是显著的
- 可能只是随机波动造成的

## 2. 实验结果分析

### 2.1 P值

在实际进行AB测试的时候，需要判断两个版本之间是否存在差异，p值告诉我们两个版本的实验结果之间存在显著差异的概率

### 2.2 两类错误和统计功效

#### 2.2.1 两类错误

- 第 I 类错误（弃真错误）：原假设为真时拒绝原假设；第 I 类错误的概率记为  $\alpha$ ，也好就是我们前文提到的显著性水平。
- 第 II 类错误（取伪错误）：原假设为假时未拒绝原假设。第 II 类错误的概率记为  $\beta$ 。

### 2.3 统计功效

实际进行AB测试时，统计功效就是，当两个不同版本之间存在显著差异时，实验能正确做出存在差异判断的概率。

### 2.4 置信区间

$$(\hat{p}_2 - \hat{p}_1) \pm Z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{N_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{N_2}}$$

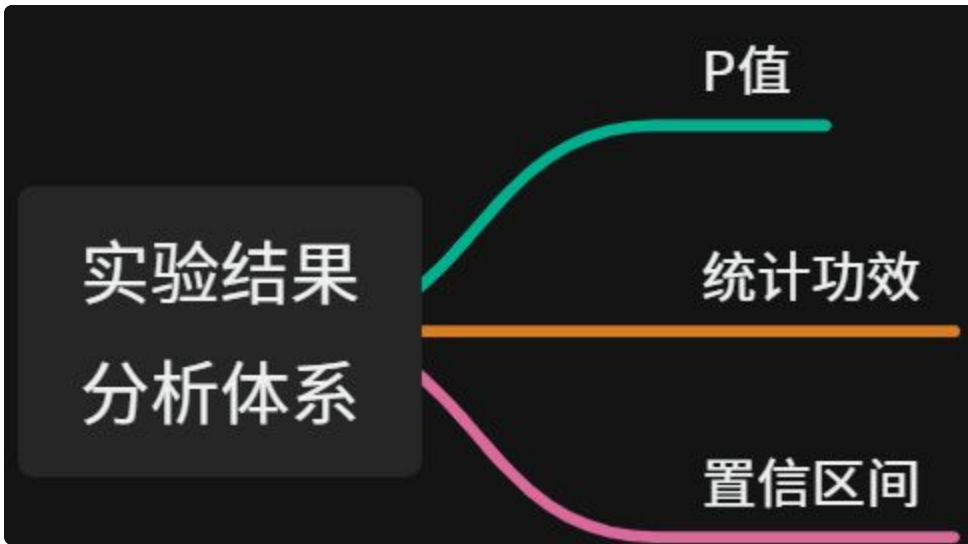
使用样本推断总体，在统计上主要由参数估计与假设检验；

其中，参数估计是用样本统计量去估计总体参数，分为点估计和区间估计，区间估计由样本统计量加减估计误差得到，这个区间称为 **置信区间**；

由样本量所构造的总体参数的估计区间称为**置信区间**。

在实际进行AB测试时，置信区间可以辅助确定版本间是否有存在显著差异的可能性：如果置信区间上下限的值同为正或负，认为存在有显著差异的可能性；如果同时有负值和正值，那么则认为不存在有显著差异的可能性。

2.5实验结果分析体系



通常，我们认为：AB测试实验数据在95%的置信水平区间内，P值小于0.05，功效大于80%的情况下，实验结果是可信赖的。但实际情况可能有所差异，部分数据存在缺陷，需要进行具体的业务分析来判断