

# AB测试中样本量的确定

2021年5月13日 8:57

在数据分析过程中，一个完整的闭环是从数据中得到洞察，根据洞察得到某种假设，通过实验检验这一假设。

## 一、大数据中心和中心极限定理

A/B 测试样本量的选取基于大数定律和中心极限定理。通俗地讲：

- 1、**大数定律**：当试验条件不变时，随机试验重复多次以后，随机事件的频率近似等于随机事件的概率。
- 2、**中心极限定理**：对独立同分布且有相同期望和方差的n个随机变量，当样本量很大时，随机变量近似服从标准正态分布N(0,1)。
- 3、**AB测试实际上是在做一个假设检验**，AB测试主要有两种检验方式（不同的样本量，不同的检验方式）
  - Z检验，即检验实验组与对照组服从分布的均值是否相等
  - 卡方检验，即检验实验组是否服从理论分布（将对照组看成理论分布）

4、根据大数定律和中心极限定理，当样本量较大（大于30）时，可以通过**Z检验（也称U检验，是为了检验在零假设情况下测试数据是否可以接近正态分布的一种统计测试）**来检验测试组和对照组两个样本均值差异的显著性。

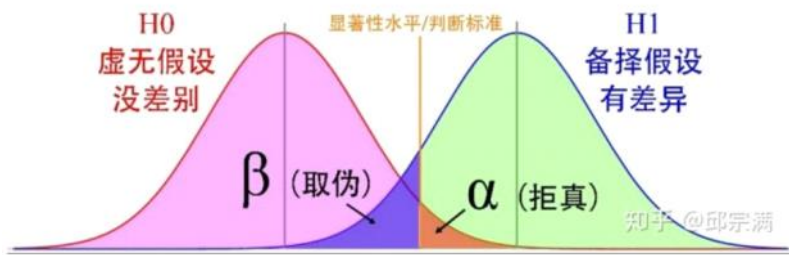
$$Z = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\mu_1$ 、 $\mu_2$ 是样本均值； $\sigma_1$ 、 $\sigma_2$ 是样本标准偏差； $n_1$ 、 $n_2$ 是样本容量

注：样本量小于30时，可进行**t检验**

在假设检验中有以下两类错误，如下图：

- **弃真**：原假设为真，而我们拒绝原假设，这种错误称为**一型错误**，其犯一类错误的概率记做显著性水平 $\alpha$
- **取伪**：原假设为假，而我们接受原假设，这种错误称为**二型错误**，其犯二类错误的概率记做 $\beta$



关于假设检验中的两类错误：

### 1、犯第I类错误的概率alpha与置信水平1-alpha

通常，将犯第I类错误的概率alpha(0.05)称为显著性，把没有犯一类错误的1-alpha(0.95)称为置信水平，即有1-alpha的概率正确接受了H0。

一般，alpha取值为0.05或更小的数值，即容忍犯第I类错误的概率最大为alpha。

### 2、犯第II类错误的概率beta与统计功效power=1-beta

通常，将犯第II类错误的概率称为beta；将1-beta称为统计功效，即正确拒绝H0的概率。

一般，beta取10%~20%，则统计功效的取值为80%~90%。

统计功效power的计算公式如下：

**Power for Comparing the Means of Two Normally Distributed Samples Using a Significance Level  $\alpha$**

To test the hypothesis  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 \neq \mu_2$  for the specific alternative  $|\mu_1 - \mu_2| = \Delta$ , with significance level  $\alpha$ ,

$$\text{Power} = \Phi \left( -Z_{1-\alpha/2} + \frac{\Delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \right)$$

where  $(\mu_1, \sigma_1^2)$ ,  $(\mu_2, \sigma_2^2)$  are the means and variances of the two respective groups and  $n_1, n_2$  are the sample sizes of the two groups.

$\Phi$  是标准正态分布的概率累积函数 (CDF)，有一个近似计算公式

$$\Phi(x) \approx \frac{1}{2} \left\{ 1 + \text{sign}(x) \left[ 1 - e^{-\frac{2}{\pi} \arctan^2 \left( \frac{x}{\sqrt{1-x^2}} \right)} \right] \right\}$$

$\Phi$ 表示z值对应的累计概率，即正态分布中的面积， $\Delta$ 是两组样本均值之差， $\sigma$ 为各组样本标准差， $n$ 为样本数。

在一般AB实验中，我们假设**AB两组是同质的，且样本量是一致的**，所以有 $\sigma_1 = \sigma_2$ ， $n_1 = n_2$ ，将上面的power公式进行转换最后可以得到如下关于样本量公式：

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

需要注意的是不同检验满足的分布是不一样的，对应的power公式也是不一样，而最小样本量公式也是从power来的，所以不同检验方式对应的最小样本量公式也是不一样的。本篇是以**满足正态分布的Z检验为例**。

### 3、统计显著性p-value

当 $p\text{-value} < \alpha$ 时，即原假设成立的概率小于预设的显著性水平，可拒绝原假设。p-value只说明两个样本有没有显著性差异，并不说明差异的大小

参考教程：

- AB测试原理及样本量计算 <<https://www.cnblogs.com/cqmcoding/p/13802023.html>>
- AB实验样本量确定 <[https://blog.csdn.net/qq\\_36153312/article/details/103934411](https://blog.csdn.net/qq_36153312/article/details/103934411)>
- 最小样本量计算 <<https://blog.csdn.net/junhongzhang/article/details/103220888>>
- 一文详解假设检验、两类错误和p值 <[https://blog.csdn.net/weixin\\_43823338/article/details/108050115](https://blog.csdn.net/weixin_43823338/article/details/108050115)>
- ABTest Guide <<https://abtestguide.com/abtestsize/>>
- Power and Sample Size <<http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality>>