

# 赶集网灰度发布平台可行性分析报告

题目：灰度发布平台设计与部署方案

姓 名\_\_\_\_\_王剑\_\_\_\_\_

部 门\_\_\_\_\_公共组\_\_\_\_\_

指导 员\_\_\_\_\_杜翔\_\_\_\_\_

指导 员\_\_\_\_\_蔡峰\_\_\_\_\_

职 称\_\_\_\_\_研发工程师\_\_\_\_\_

## 摘 要

依靠策略对用户进行分类，选择某群体进行试验。跟踪访问情况，采集指标数据，分析样本结果。

关键词：AB 测试；Bucket test 分流器设计；迭代开发；假设检验；

## ABSTRACT

A/B Testing is a way of conducting an experiment where you compare a control group to the performance of one or more test groups by randomly assigning each group a specific single-variable treatment.

Keywords: AB test; Bucket test shunt design; iterative development; hypothesis testing;

# 目 录

一、	概述.....	5
1.	课题背景.....	5
2.	国内外发展.....	5
二、	相关技术背景.....	6
1.	Haproxy 简介.....	6
2.	手机 Webapp 日志收集以及分析.....	6
3.	AB 测试.....	6
三、	分流器模型.....	7
1.	分流概述.....	7
2.	分流因子.....	7
3.	分流数学模型.....	7
4.	分流设计模式.....	9
5.	HAProxy 如何分流.....	10
四、	灰度实例开发及部署.....	12
1.	项目叠代.....	12
2.	多灰度部署冲突.....	12
五、	日志收集及分析.....	13
1.	分析指标.....	13
2.	AB 测试结果验证-假设检验.....	13

## 一、概述

### 1. 课题背景

为了真实验证一个算法模型的好坏，需要有一个系统能提供真实的流量来检验。

在用户使用系统时，由分流系统根据一定的策略来自动决定用户的分组号（TrackGroup id），保证自动抽取导入不同分组的流量具有可对比性，然后让不同分组的用户看到的不同模型提供的结果。用户在不同模型下的行为将被记录下来，这些行为数据通过数据分析形成一系列指标，而通过这些指标的比较，最后就形成了不同模型之间孰优孰劣的结论。只要分组的流量达到一定的程度，数据指标从统计意义上就具有可比性。<sup>[1]</sup>

### 2. 国内外发展

#### 2.1. Facebook

由于 Facebook 发布的产品比较多，经常出现互相影响的情况，一般均在做发布前评估让大家知道什么产品即将在什么时候推出去。

其发布工程的做法是阀门控制式的灰度发布，就是有所控制地选择发布的人群及其比例。灰度发布是控制发布的范围和速度。

通过数据监测来判断发布状态，才能得知某一阶段产品发布的质量，那种状况下才提高灰度发布的范围。

#### 2.2. 腾讯

QZone 是另外一个采用灰度发布的例子。大家都知道，QZone 在过去的一年中改进是巨大的，从以前慢悠悠的老爷爷变成了一个充满青春活力的小伙子。其中经历了大小无数次的发布，他们的发布也都是采用了灰度发布的策略，用户数据的升级并不是大面积的一次性升级，而是通过一个用户升级标志服务器，如果用户数据没有升级，后台会把此用户的数据逐步迁移到新版本上，然后将升级标志位置 1，升级过程中，用户仍然可以访问旧的数据，升级完成后的访问都将转发给新的版本。

QQ 的很多产品发布都采用灰度发布，有些是抽取部分 QQ 号段升级成新系统，然后根据用户反馈再大范围升级。我们的产品大部分也是采用灰度发布。

## 二、 相关技术背景

### 1. Haproxy 简介

HAProxy 提供高可用性、负载均衡以及基于 TCP 和 HTTP 应用的代理，支持虚拟主机，它是免费、快速并且可靠的一种解决方案。

HAProxy 特别适用于那些负载特大的 web 站点，这些站点通常又需要会话保持或七层处理。

HAProxy 多进程或多线程模型受内存限制、系统调度器限制以及无处不在的锁限制，很少能处理数千并发连接。事件驱动模型因为在有更好的资源和时间管理的用户端 (User-Space) 实现所有这些任务，所以没有这些问题。

### 2. 手机 Webapp 日志收集以及分析

日志通过 1px\*1px 像素手机，而 pv 日志 utmgjeval 字段记载版本灰度值<sup>[3]</sup>。

### 3. AB 测试

AB 测试就是一种灰度发布方式，让一部用户继续用 A，一部分用户开始用 B，如果用户对 B 没有什么反对意见，那么逐步扩大范围，把所有用户都迁移到 B 上面来。

灰度发布可以保证整体系统的稳定，在初始灰度的时候就可以发现、调整问题，以保证其影响度缩短测试周期。减少测试干扰，提高测试准确性和价值。减少用户伤害，降低测试影响页面优化及改版方案评估决策。思考后台算法，引擎及业务逻辑优化结果。

### 三、 分流器模型

#### 1. 分流概述

通过分流因子确认分流实例，依靠 Haproxy 代理打点记录及分流，最后完成分流任务。

#### 2. 分流因子

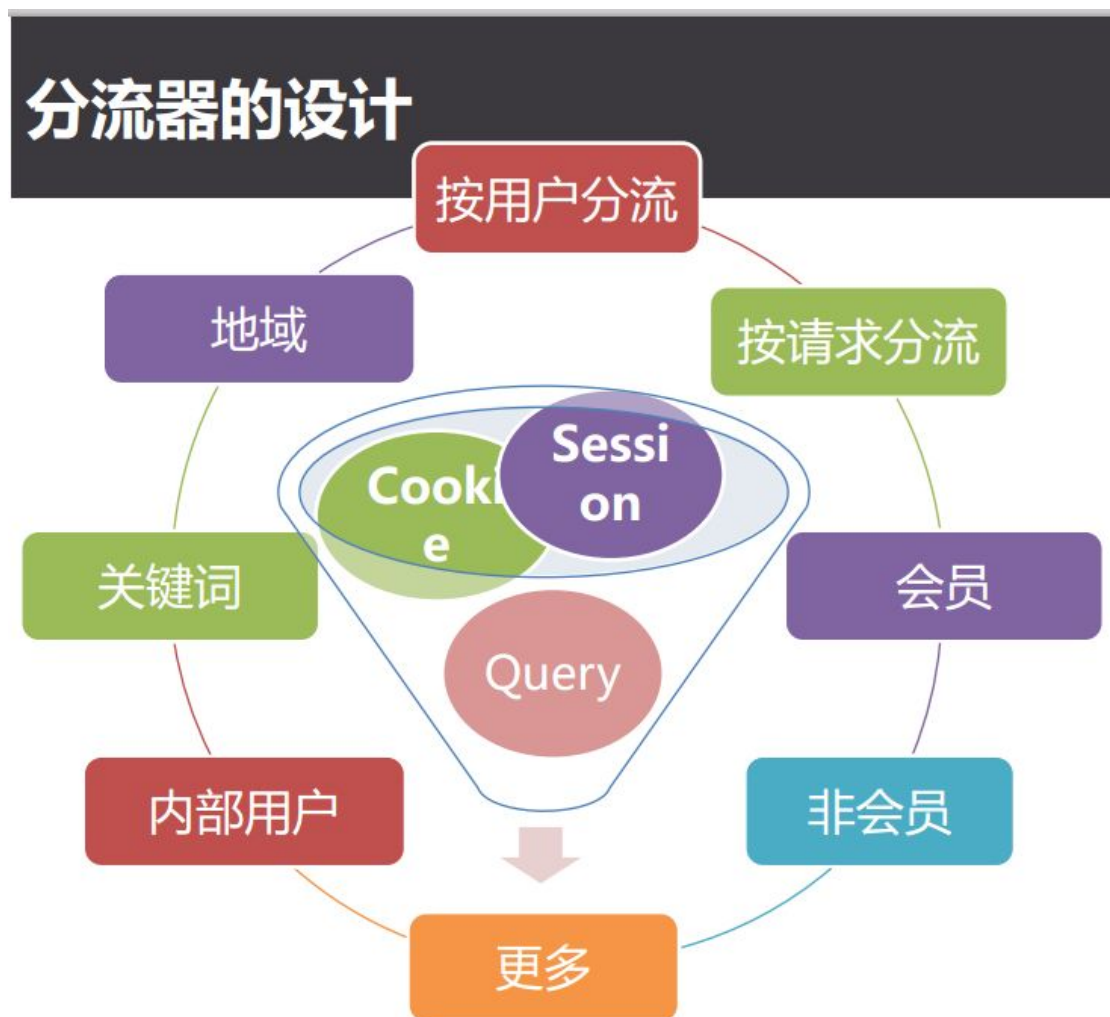


图 3.1 分流器的设计

#### 3. 分流数学模型

有四种常见流量分配函数，`user_id mods`，`cookies mods`，`cookie-day mods`，以及 `random`。因为要把流量无偏地分割给不同的实验，所以就有了流量分配函数。有四种流量分配函数，`user_id mods`，`cookies mods`，`cookie-day mods`，以及 `random`。Random 就是随机分配。

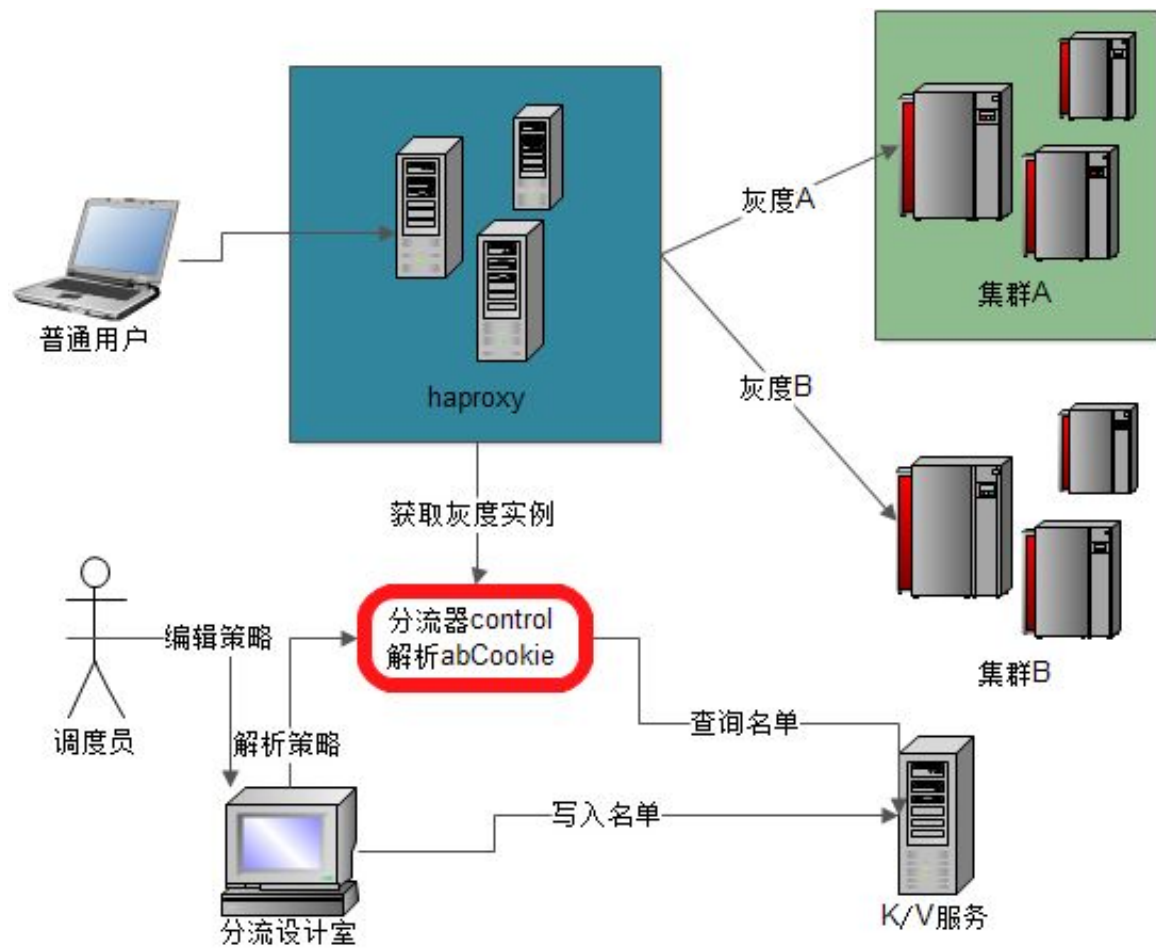
#### **a. cookies mods**

很简单，这里单说 `cookies mods`，因为剩下的两种其实与之类似。常用的一种做法是把用户 `cookie` 转换为一个数字，对其取模，Experiment 就可以指定取哪个模的流量作为训练样本。可以把这个模的表达式为  $f(\text{cookie}) \% 1000$ ，意为把流量分割为 1000 块。更合理的表达式为  $f(\text{cookie}, \text{layer}) \% 1000$ ，这个式子与之前的式子不同之处在于把 `layer` 也作为变量引入进来，这样可以保证 `layer` 之间的实验是独立的。考虑取第一个式子的情况，如果 `layer1` 的实验 `E11`，取 `mod=1` 的流量，`layer2` 的实验 `E21` 也取 `mod=1` 的流量，则这两个实验是相关的。所以如果要想保证 `layer` 间实验的独立性，使用第二个式子即可保证同一个访问在不同层的模不会完全一样。

要决定一个访问是否能用于一个实验，除了分配函数，还有另一个因素：`condition`。这个因素引入了比“模”这种随机标识更为丰富的访问本身的特性，如访问来自的国别、语言、浏览器等等，以便于定制更为有针对性的实验。如果还需要更为细致的访问定义，则可由 `trigger` 这个东西来完成，但因为这方面的特征过于复杂，所以它是定义在实验内部的，并不在流量分发阶段完成判断。



#### 4. 分流设计模式



做赶集必须的!

图 3.4.1 分流器 usecase

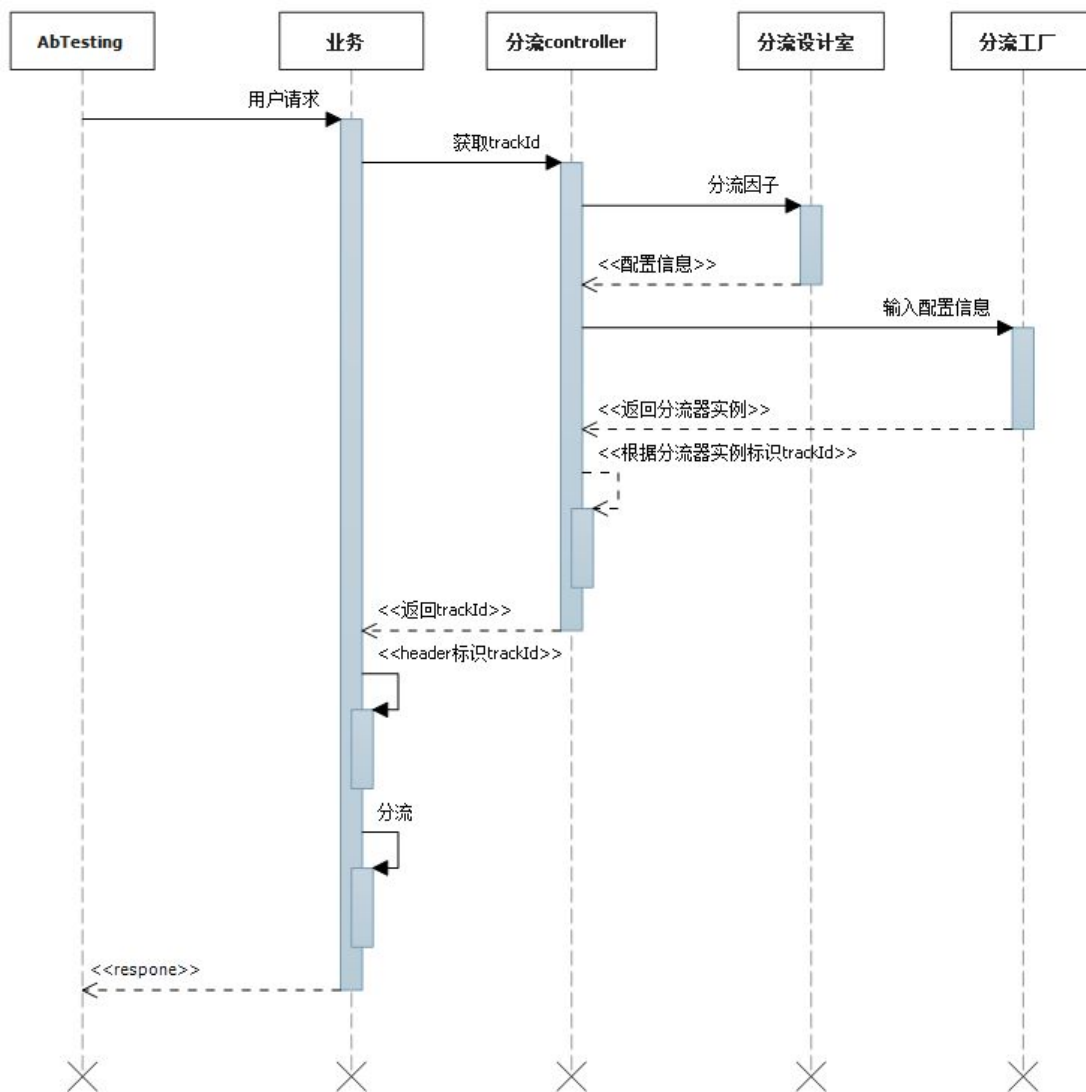


图 3.4.2 分流器的时序图

## 5. HAProxy 如何分流

通过算法将一部分流量（或人）导入后端指定的几台机器上面。这个的难点在于：

如何保证用户一旦进入新版本后不会再跳回老版本？

如何控制流量和后端指定 app 数量之间的比率？

首先确认了不能让 haproxy 种 cookie<sup>[2]</sup>，核心配置是放在 frontend 下，通过 `hdr_reg(trackId)` 区分，把流量分成了 N 等分，按百分比来渐变的。后端根据不同的机器数量进行百分比的切换。。这样就能保证灰度发布过程中不至于出现前端大部分的流量集中访问到后端较少的机器上的极端情况。按照灰度发布的要求算出配置文件，然后自动生成配置文件，验证后 reload 即可。

```
acl cookie_gray_new hdr_reg(cookie) -i ha_gray_cookie=(1|2|3|4|5)(.*?)
acl cookie_gray_old hdr_reg(cookie) -i ha_gray_cookie=(6|7|8|9|0)(.*?)
use_backend app_1 if cookie_gray_new
use_backend app_2 if cookie_gray_old
```

图 3.5.1 HAProxy 代码<sup>[2]</sup>

## 四、 灰度实例开发及部署

### 1. 项目叠代

使用 git 管理版本，并行项目同步问题使用 rsync svn 方案管理。部署灰度版本使用阿波罗 git 上线

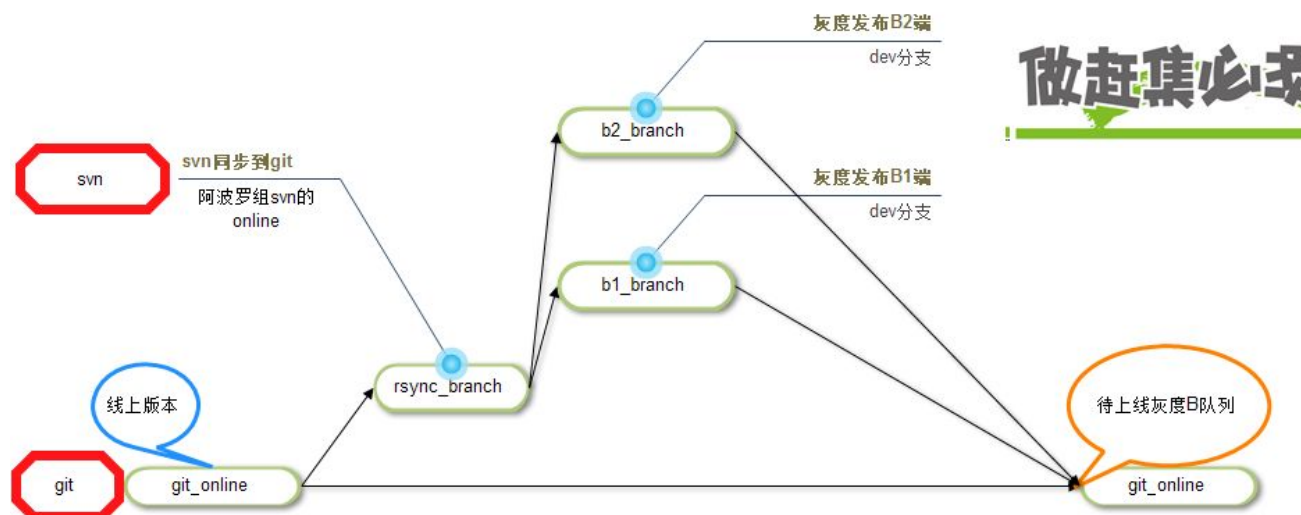


图 4.1.1 git 分支管理

### 2. 多灰度部署冲突

- 1, 按照域名硬切换不同灰度版本  
eg:r576.3g.ganji.com 或者 r678.3g.ganji.com
- 2, 按照 groupId 部署不同代码  
按照 trackGroupId 进行分组, nginx 指向不同 webroot。

```
-- wap_abtesting_group_576
|-- app
|-- plugin
|-- templates
|-- webroot
-- wap_abtesting_group_678
|-- app
|-- plugin
|-- templates
|-- webroot
-- wap_abtesting_group_857
|-- app
|-- plugin
|-- templates
|-- webroot
-- wap_abtesting_group_900
```

## 五、 日志收集及分析

依靠 Javascript 植入或者 Gif 统计网站 Pv,再分析使用 Hadoop 的 mapreduce 进行数据分析。通过 trackID 分离样本容量。

### 1. 分析指标

- 用户访问深度变化
- 跳出比率
- 某一转换率指标，比如注册用户等。

### 2. AB 测试结果验证-假设检验

#### 2.1. Z 检验

A 版本转化率(control)和 B 版本的转化率 (treatment) 的分布的一个例子。

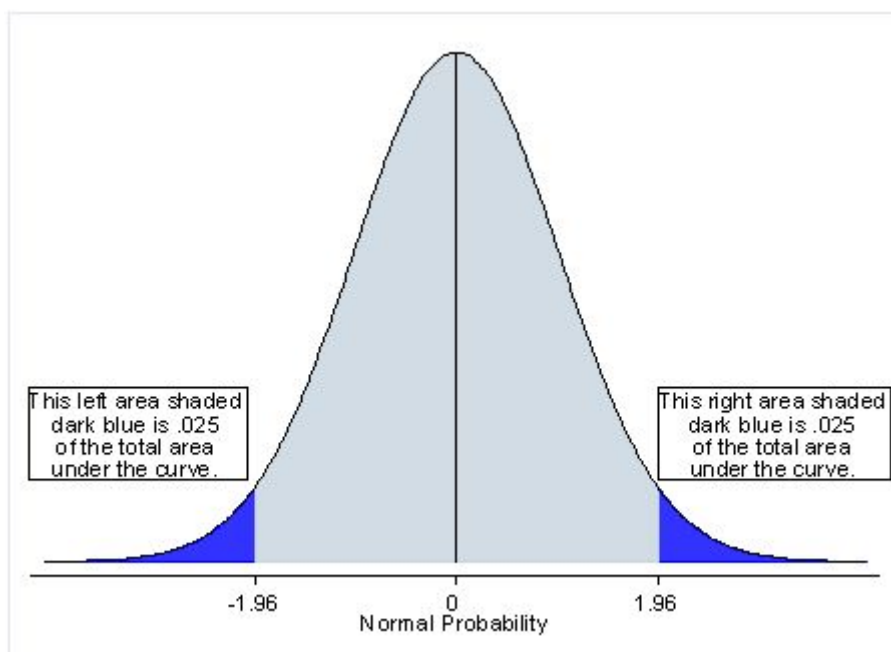


图 5.0.1 双尾检验

#### 2.2. 实验数据

A 版本			
6	9	10	5
7	7	3	4
4	9	3	6
9	5	9	4
4	7	8	4

8	7	8	8
4	5	9	8
9	6	8	3
6	8	10	3
7	6	5	4
8	8	8	5
5	2	9	8
5	4		

图 5.1.1 A 版本实验数据

B 版本			
10	8	10	6
9	8	10	10
3	9	4	8
5	7	5	7
3	7	3	9
6	9	7	9
5	5	7	9
3	9	4	8
2	10	5	1
7	6	6	7
6	3	8	8
8	8	6	8
5	10		

图 5.1.2 B 版本实验数据

版本	A	B
平均值	6.34	6.72
标准差	2.162859	2.373665
方差	4.677959	5.634286

图 5.1.3 A 版本实验数据

Z	P 值	差异度
>=2.58	<=0.01	非常显著
>=1.96	<=0.05	显著
<1.96	>0.05	不显著

图 5.1.4 Z 值置信区间

2.3. 计算公式

$$z = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}}$$

图 5.1.5 Z 检验

类似 Excel 中的 ZTEST 函数<sup>[4]</sup>，其中：avg(x) 平均数；U0 原始均值；N 样本容量；Sigma 如图 5.1.6。

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

图 5.1.6 标准方差

类似 Excel 中的 STDEVP 函数<sup>[5]</sup>。

#### 2.4. 计算步骤

- 提出假设（原假设 H0，没有差异，那么备择假设 H1 就是有差异）
- 规定显著性水平，如 0.05。
- 计算 Z 测试值。

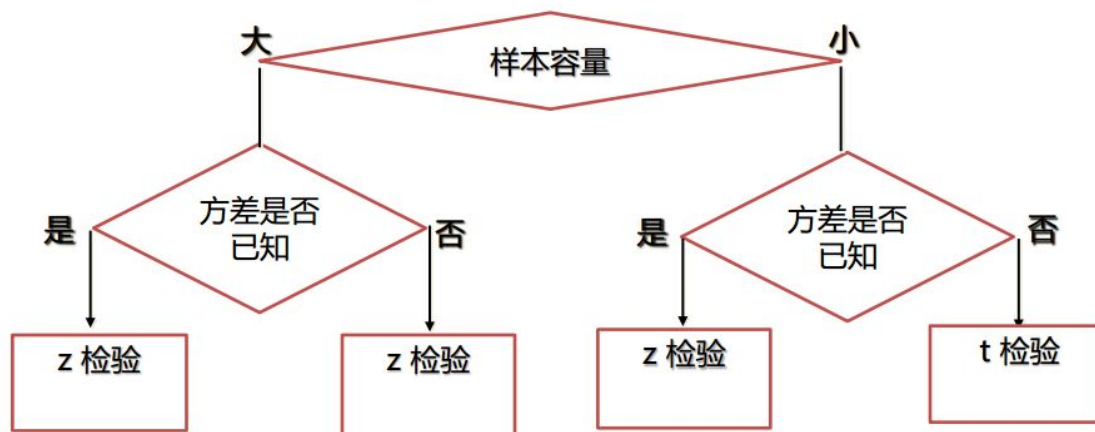
#### 2.5. 结论

比较计算所得 Z 值与理论 Z 值，推断收生的概率，采用双尾比较结果如图 5.1.7

$$|-0.84| < 1.96$$

图 5.1.7 Z 值比较结论

#### 2.6. 计算方法扩展



**Z检验、T检验、G检验、秩和检验、卡方检验等**

图 5.1.8 假设检验方法

## 参考文献

[1]OASIS. Universal description , discovery and integration[EB/OL]. <http://www.uddi.org>, 2001-11-14.

[2]Answered. haproxy cookie can't roundrobin

<http://stackoverflow.com/questions/17267405/haproxy-cookie-cant-roundrobin/17397935#17397935>

[3]GanjiMobile.[www.ganji.com](http://www.ganji.com). 移动互联网日志统计格式 [EB/OL].

<http://wiki.corp.ganji.com/%E5%85%AC%E5%85%B1%E7%BB%84/%E7%A7%BB%E5%8A%A8%E4%BA%92%E8%81%94%E7%BD%91/%E6%97%A5%E5%BF%97%E7%BB%9F%E8%AE%A1>

[4]Microsoft.标准方差及 Ztest 函数[EB/OL].

<http://office.microsoft.com/en-us/excel-help/stdevp-HP005209281.aspx>

[5]Microsoft.标准方差及 Ztest 函数[EB/OL].

<http://office.microsoft.com/en-us/excel-help/stdevp-HP005209281.aspx>