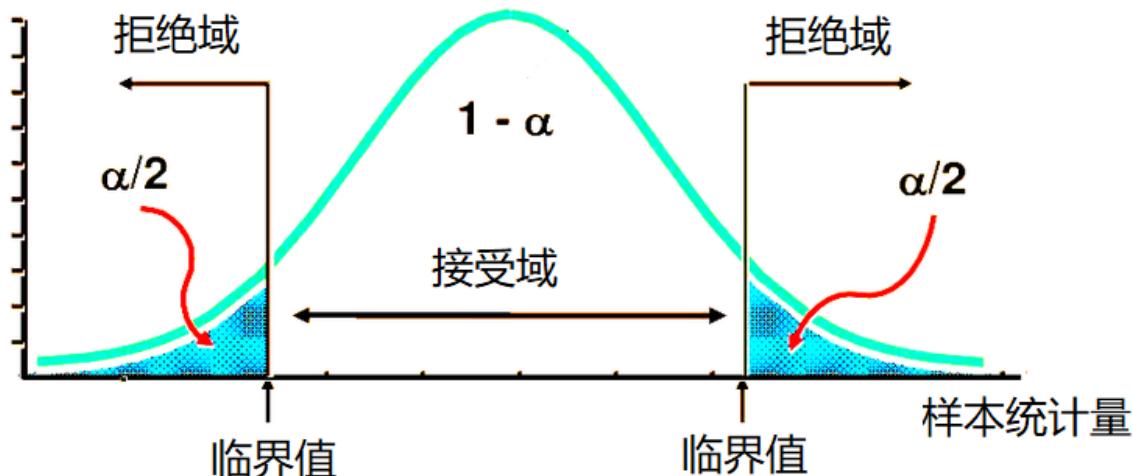


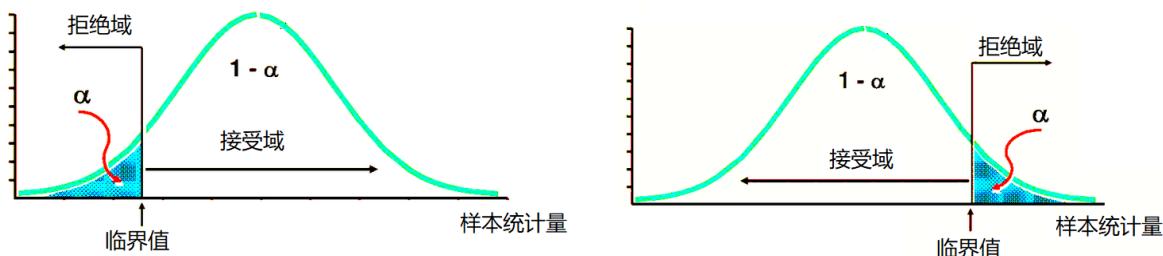
### 2.4.3 不同分布的拒绝域

对称型 (Z分布、t分布) :

双侧检验:

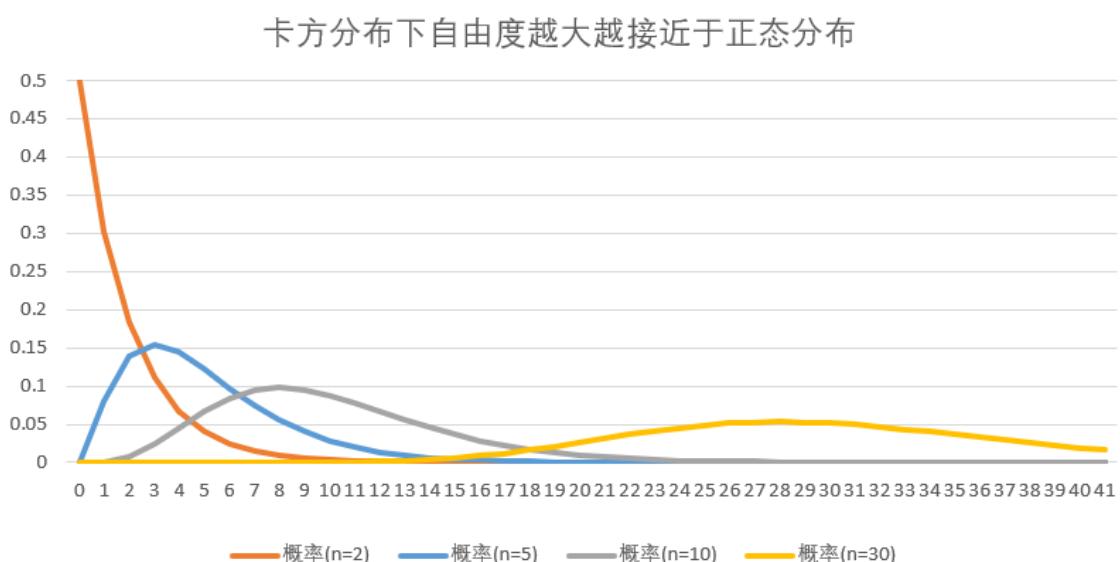


单侧检验:



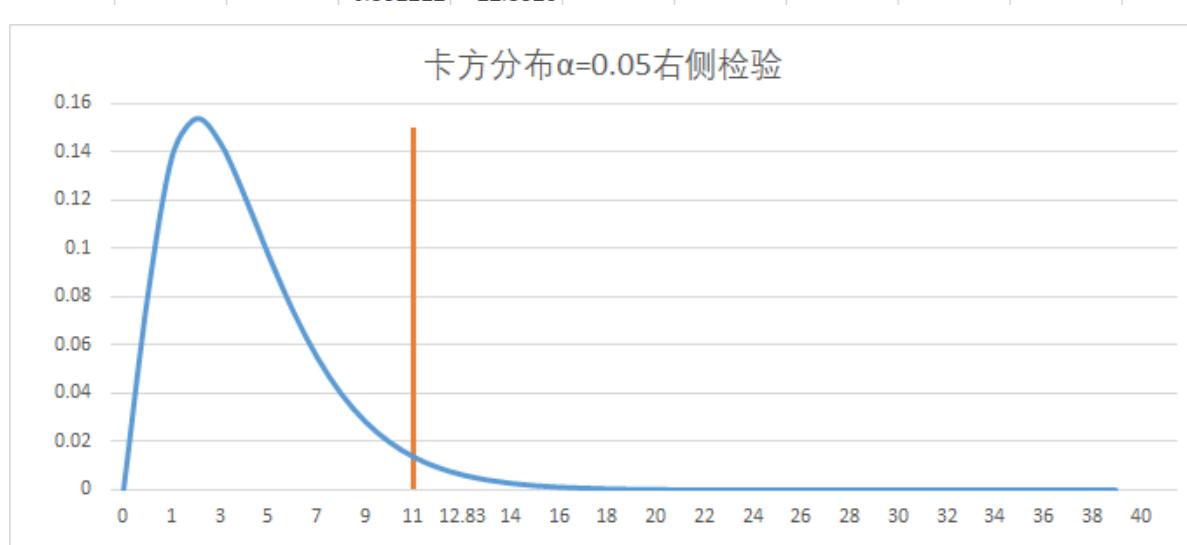
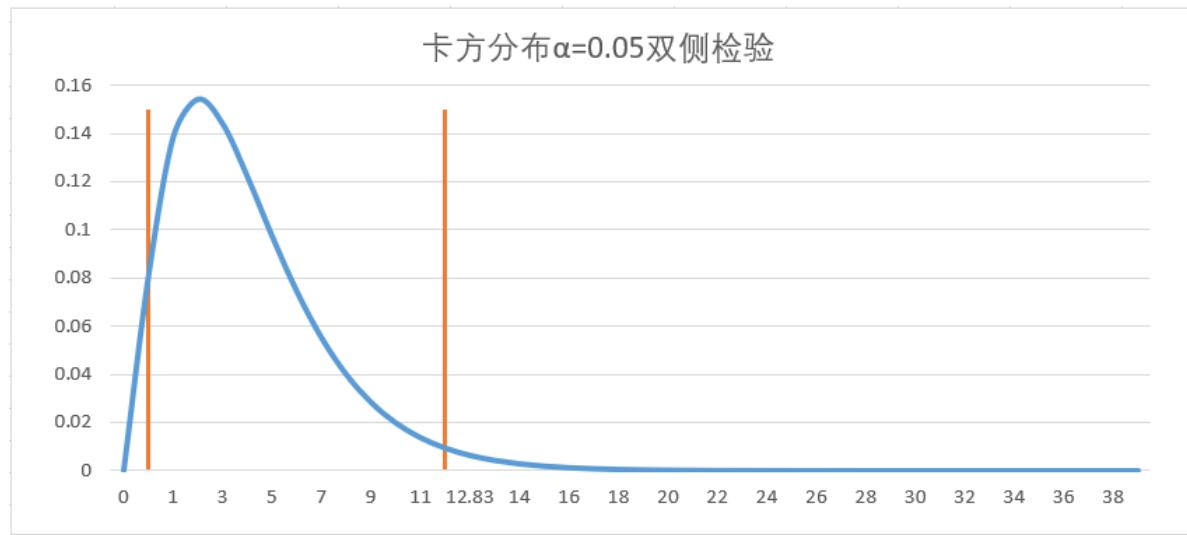
非对称型 (卡方分布、F分布) :

卡方分布:



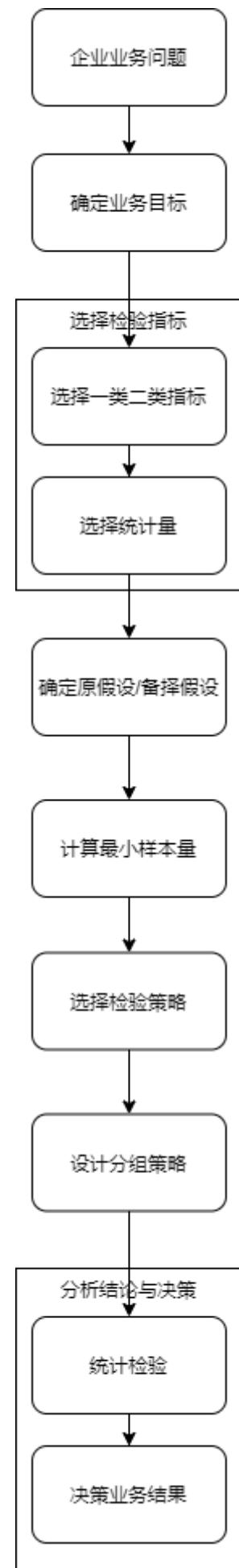
拒绝域:

(卡方分布在左侧的拒绝域特别小, 所以拒绝的区间的值也比较少), 所以卡方检验的拒绝域一般放在右侧。F分布同理



### 3 学习企业的ABTest的关键流程

#### 3.1 整体实验设计与分析流程



## 3.2 实验设计

### 3.2.0 确定业务目标

3.2.0 确定业务目标

1. 明确我们要提升的业务指标；  
2. 明确我们要改进的产品/策略。

如果明确这个部分，实验会显得比较精简。目标明确。

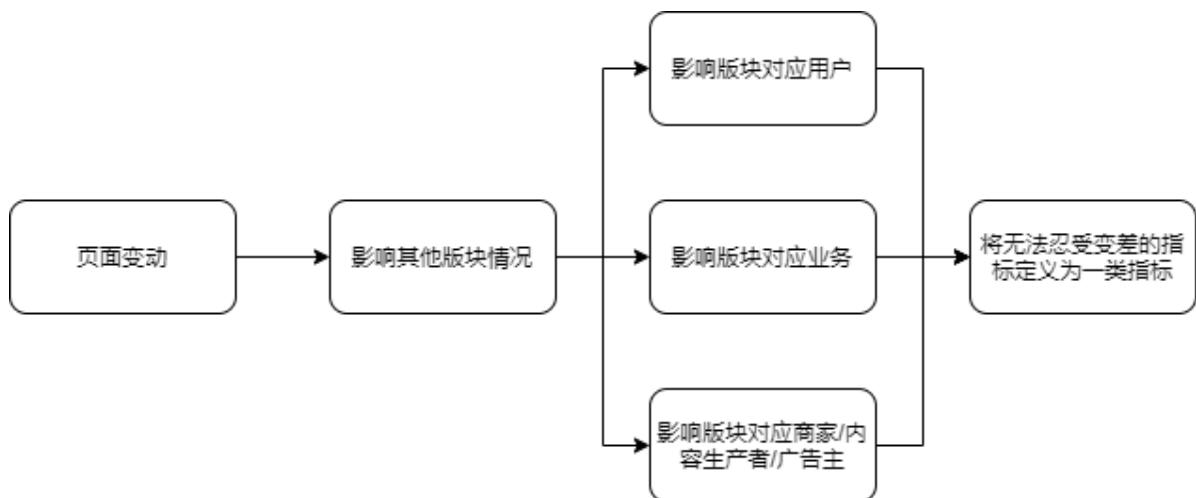
### 3.2.1 选择检验指标

#### 1. 选择一类指标

一类指标：不能容忍变差的指标；

二类指标：目标提升的指标。

如何确定一类指标？



场景举例：

业务场景	一类指标	二类指标
广告类业务		广告效果类指标：CTR、广告位曝光、广告位转化率、广告主消耗、客户满意度
产品功能改进业务	用户粘性指标：用户活跃时长、用户留存、用户点击深度、人均消费量	产品指标：产品CTR、跳出率、功能使用率、功能曝光率；用户粘性指标；
推荐算法类		效果类：转化率；品牌类：曝光量、传播量；内容类：内容消费率，消费时长；

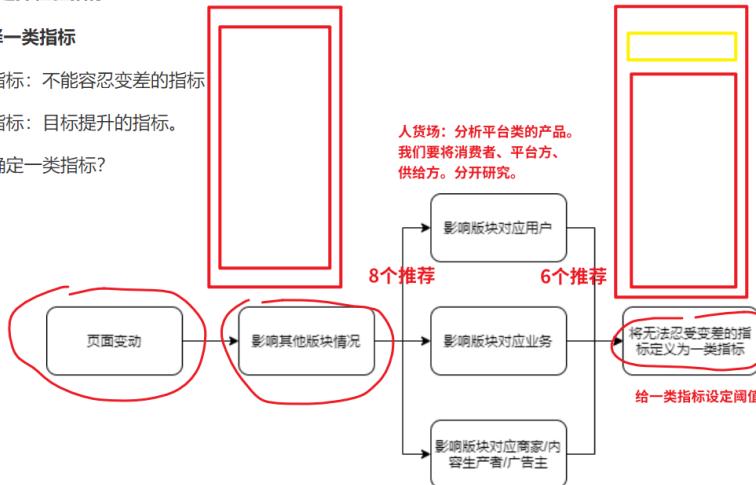
### 3.2.1 选择检验指标

#### 1. 选择一类指标

一类指标：不能容忍变差的指标

二类指标：目标提升的指标。

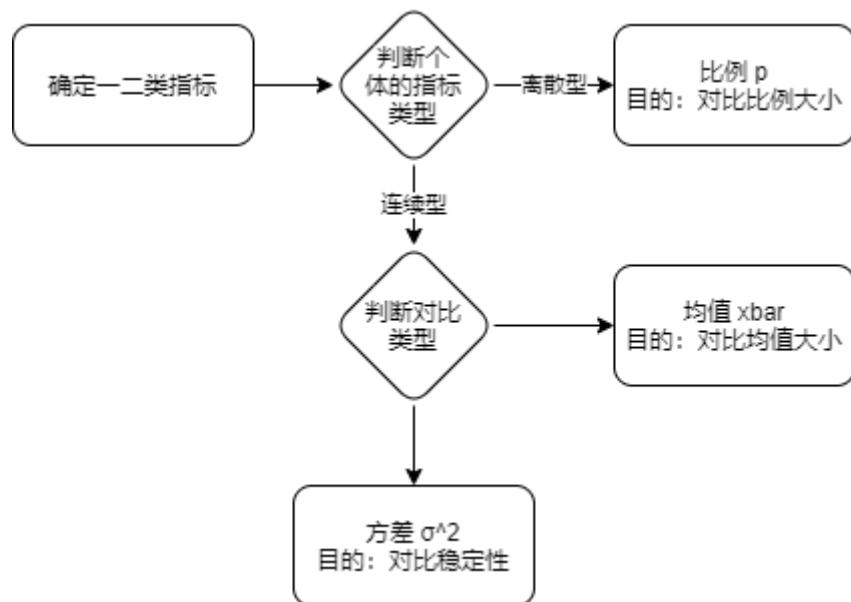
如何确定一类指标？



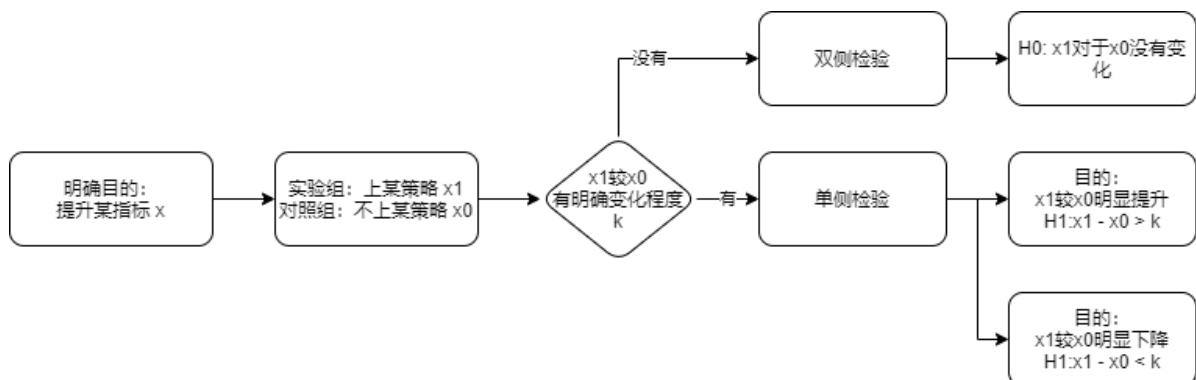
场景举例：

业务场景	一类指标	二类指标
广告类业务	用户粘性指标：用户活跃时长、 产品功能改进业务	广告效果类指标：CTR、广告位曝光、广告位转化率、广告主消耗、客户满意度 产品指标：产品CTR、跳出率、功能使用率、 功能曝光率；用户粘性指标； 效果类：转化率；品牌类：曝光量、传播量；内容类：内容消费率，消费时长；
推荐算法类	用户留存、用户点击深度、人均 消费量	

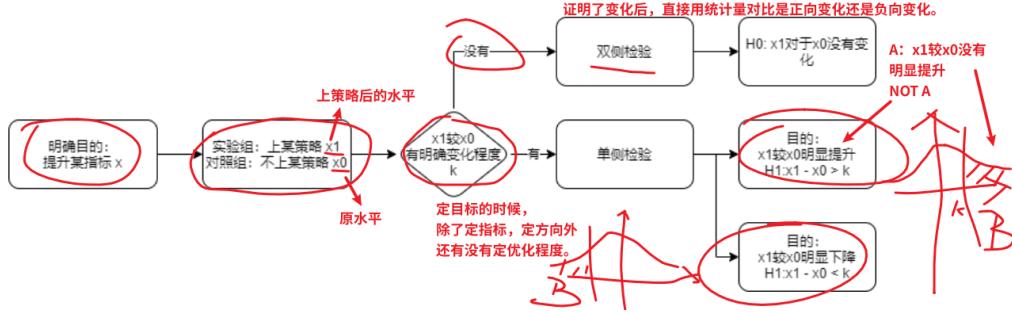
#### 2. 选择统计量



### 3.2.2 确定原假设与备择假设



### 3.2.2 确定原假设与备择假设



### 3.2.3 两类统计错误的防范

		实际情况	
		$H_0$ 正确	$H_0$ 错误
实验结论	拒绝 $H_0$	I类错误( $\alpha$ 错误)	正确
	接受 $H_0$	正确	II类错误( $\beta$ 错误)

- $\alpha + \beta$ 不一定等于1。
- 在样本容量确定的情况下， $\alpha$ 与 $\beta$ 不能同时增加或减少。
- 统计检验效力 ( $1-\beta$ ) 当 $H_0$ 为假时,得出拒绝 $H_0$ 的正确结论的概率,被称做检验的效力

- I类错误防范：

- 小概率 $\alpha$ 设置小些(避免小概率的触发)
- 增加样本量 (使异常数据的影响降低)

- II类错误防范：

- 调大 $\alpha$  (增加小概率的触发) 但是接受I类错误的代价远比II类错误的代价要大, 所以不予使用
- II类错误概率只能在实验结束后才能计算发生二类错误的概率, 这是一个事后值。所以在事前设计我们一般不考虑这个问题。默认二类错误的概率为20%。

### 3.2.4 样本量计算

统计学上根据统计量抽样分布和边际误差确定样本量

样本量计算工具：<http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equivalence>

业务层面是以一类错误临界值二类错误临界值计算

类型	名称	公式	备注
统计学	估计总体均值时的样本容量	$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$	
	估计总体比例时的样本容量	$n = \frac{(z_{\alpha/2})^2 \cdot \pi(1-\pi)}{E^2}$	

其中，区间估计算式：

$$\bar{x} \pm Z_{\alpha/2} * \sqrt{\sigma^2/n}$$

### 3.2.4 样本量计算

统计学上根据统计量抽样分布和边际误差确定样本量

样本量计算工具: <http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality>

业务层面是以一类错误临界值二类错误临界值计算

类型	名称	公式	备注
统计学	估计总体均值时的样本容量	$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$	NORM.inv
	估计总体比例时的样本容量	$n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2}$	

其中，区间估计算式：  
 $\bar{x} \pm Z_{\alpha/2} * \sqrt{\sigma^2/n}$  边际误差

估计均值之差时的样本量(双侧)	$n_A = \kappa n_B$ and $n_B = \left(1 + \frac{1}{\kappa}\right) \left(\frac{\sigma(z_{1-\alpha/2} + z_{1-\beta})}{\mu_A - \mu_B}\right)^2$	$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2})$ , $z = \frac{\mu_A - \mu_B}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$
估计比例之差时的样本量(双侧)	$n_A = \kappa n_B$ and $n_B = \left(\frac{p_A(1-p_A)}{\kappa} + p_B(1-p_B)\right) \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{p_A - p_B}\right)^2$	$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2})$ , $z = \frac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$
估计均值之差时的样本量(单侧)	$n_A = (\sigma_A^2 + \sigma_B^2/\kappa) \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\mu_A - \mu_B}\right)^2$	$1 - \beta = \Phi\left(\frac{ \mu_A - \mu_B  \sqrt{n_A}}{\sqrt{\sigma_A^2 + \sigma_B^2/\kappa}} - z_{1-\alpha}\right)$
估计比例之差时的样本量(单侧)	$n_A = \kappa n_B$ and $n_B = \left(\frac{p_A(1-p_A)}{\kappa} + p_B(1-p_B)\right) \left(\frac{z_{1-\alpha} + z_{1-\beta}}{p_A - p_B}\right)^2$	$1 - \beta = \Phi\left(\frac{ p_A - p_B }{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}} - z_{1-\alpha}\right)$

样本量计算工具: <http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality>

业务层面是以一类错误临界值二类错误临界值计算

类型	名称	公式	备注
统计学	估计总体均值时的样本容量	$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$	NORM.inv
	估计总体比例时的样本容量	$n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2}$	

其中，区间估计算式：  
 $\bar{x} \pm Z_{\alpha/2} * \sqrt{\sigma^2/n}$  边际误差

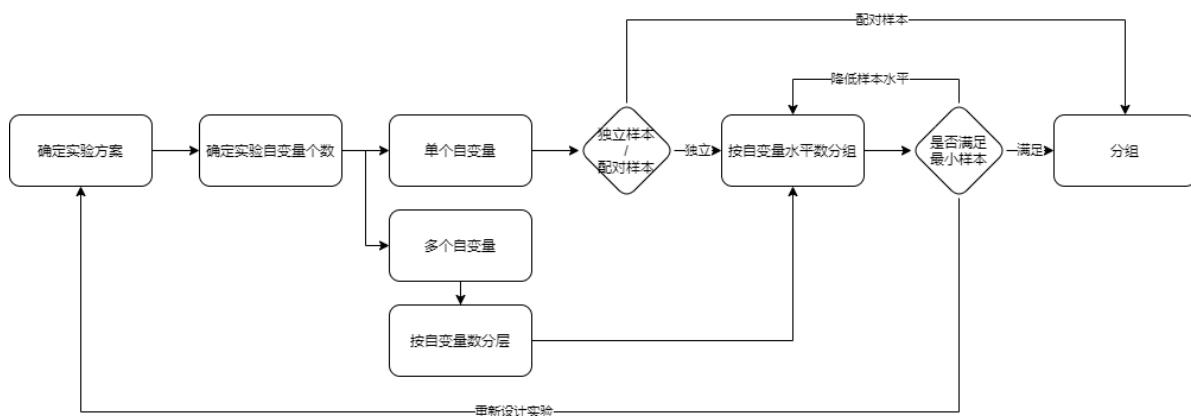
k: a组样本量与b组样本量之比。

估计均值之差时的样本量(双侧)	$n_A = \kappa n_B$ and $n_B = \left(1 + \frac{1}{\kappa}\right) \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\mu_A - \mu_B}\right)^2$	$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2})$ , $z = \frac{\mu_A - \mu_B}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$
估计比例之差时的样本量(双侧)	$n_A = \kappa n_B$ and $n_B = \left(\frac{p_A(1-p_A)}{\kappa} + p_B(1-p_B)\right) \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{p_A - p_B}\right)^2$	$1 - \beta = \Phi(z - z_{1-\alpha/2}) + \Phi(-z - z_{1-\alpha/2})$ , $z = \sqrt{\frac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}}$
估计均值之差时的样本量(单侧)	$n_A = (\sigma_A^2 + \sigma_B^2/\kappa) \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\mu_A - \mu_B}\right)^2$	$1 - \beta = \Phi\left(\frac{ \mu_A - \mu_B  \sqrt{n_A}}{\sqrt{\sigma_A^2 + \sigma_B^2/\kappa}} - z_{1-\alpha}\right)$
估计比例之差时的样本量(单侧)	$n_A = \kappa n_B$ and $n_B = \left(\frac{p_A(1-p_A)}{\kappa} + p_B(1-p_B)\right) \left(\frac{z_{1-\alpha} + z_{1-\beta}}{p_A - p_B}\right)^2$	$1 - \beta = \Phi\left(\frac{ p_A - p_B }{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}} - z_{1-\alpha}\right)$

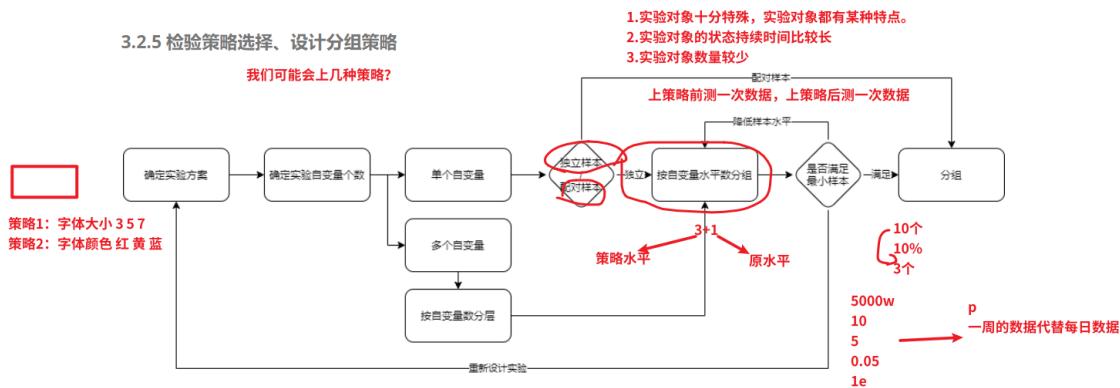
提高/降低  
的目标:

策略不同，导致我们一类和二类的统计量都可能服从不同的总体分布。

### 3.2.5 检验策略选择、设计分组策略



### 3.2.5 检验策略选择、设计分组策略



### 3.2.6 当企业没有AB测试的条件的时候, 如何解决问题?

- 1.没有系统
- 2.用户量不够
- 3.时间成本高

### 3.2.6 当企业没有AB测试的条件的时候, 如何解决问题?

- 1.没有系统      没有灰度发布的系统。
- 2.用户量不够      如果是比例的, 可以提高测试周期。      针对小众用户。
- 3.时间成本高      1.时间跨度过长。  
                      2.类似于转化周期过长。

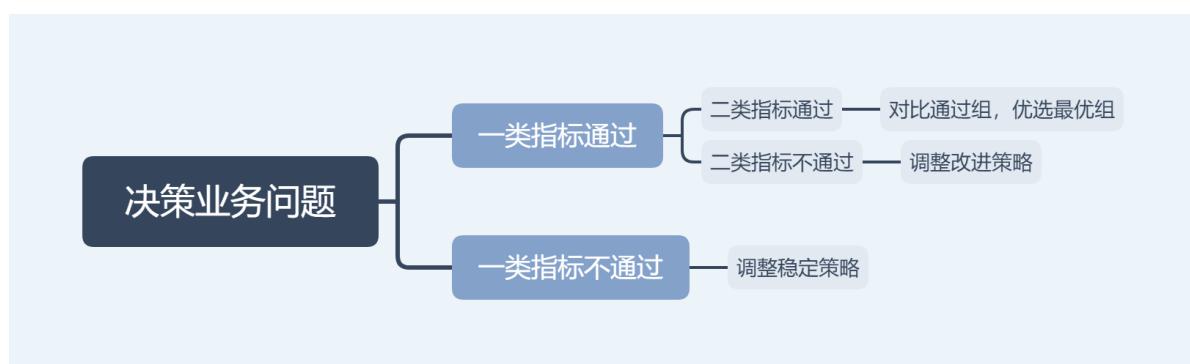
## 3.3 实验结论分析

### 3.3.1 决策统计检验

做实验决策可以通过统计量 及 统计量的P值来实现。

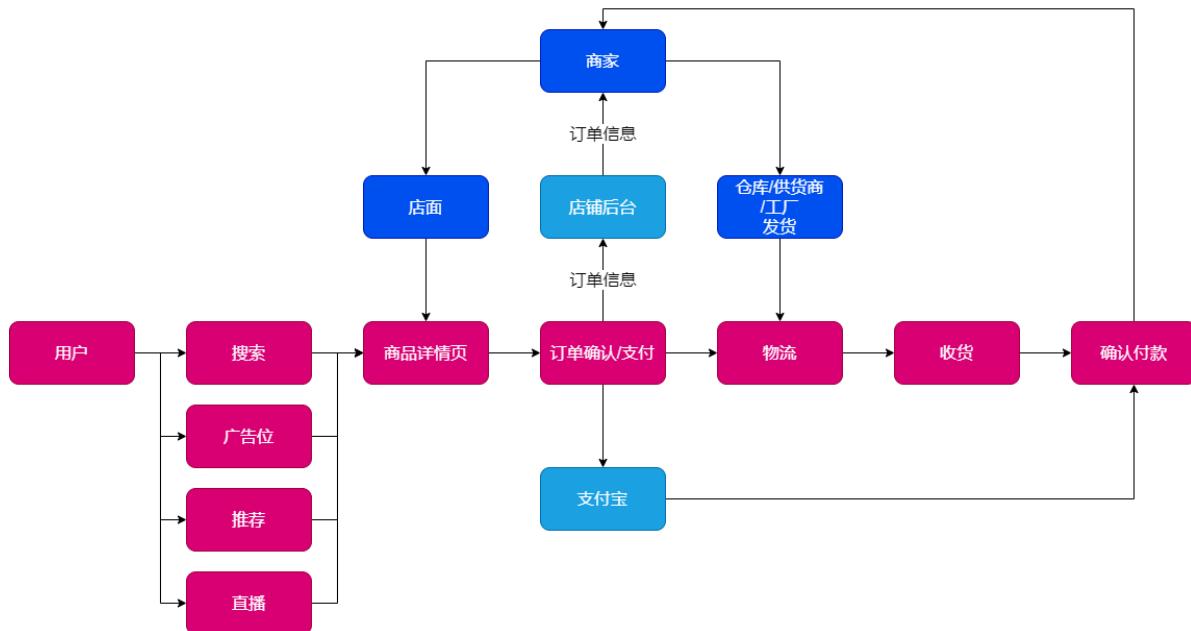
同时也可以通过样本量分布和显著性水平来确定拒绝域和接受域, 从而拒绝或者接受结果。见假设检验部分。

### 3.3.2 决策业务问题



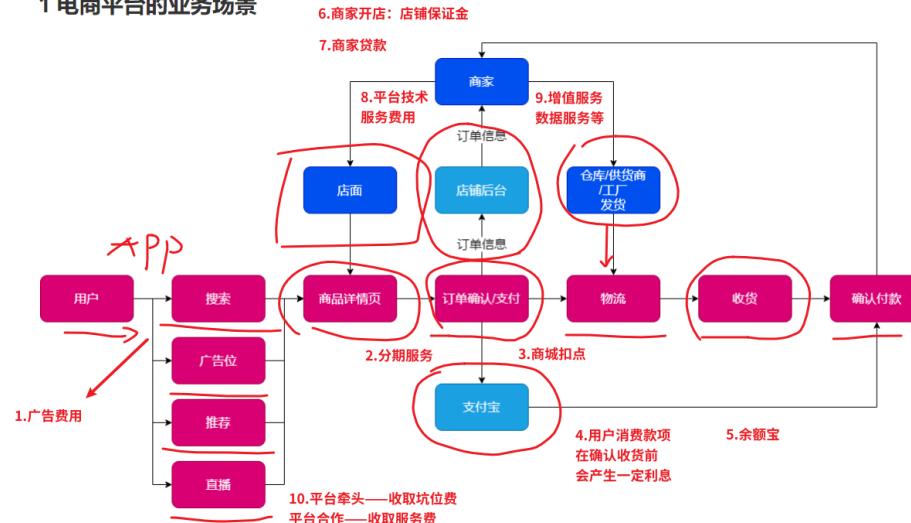
## 第三节 电商平台中小企业流量扶持问题

### 1 电商平台的业务场景



### ④ 第三节 电商平台中小企业流量扶持问题

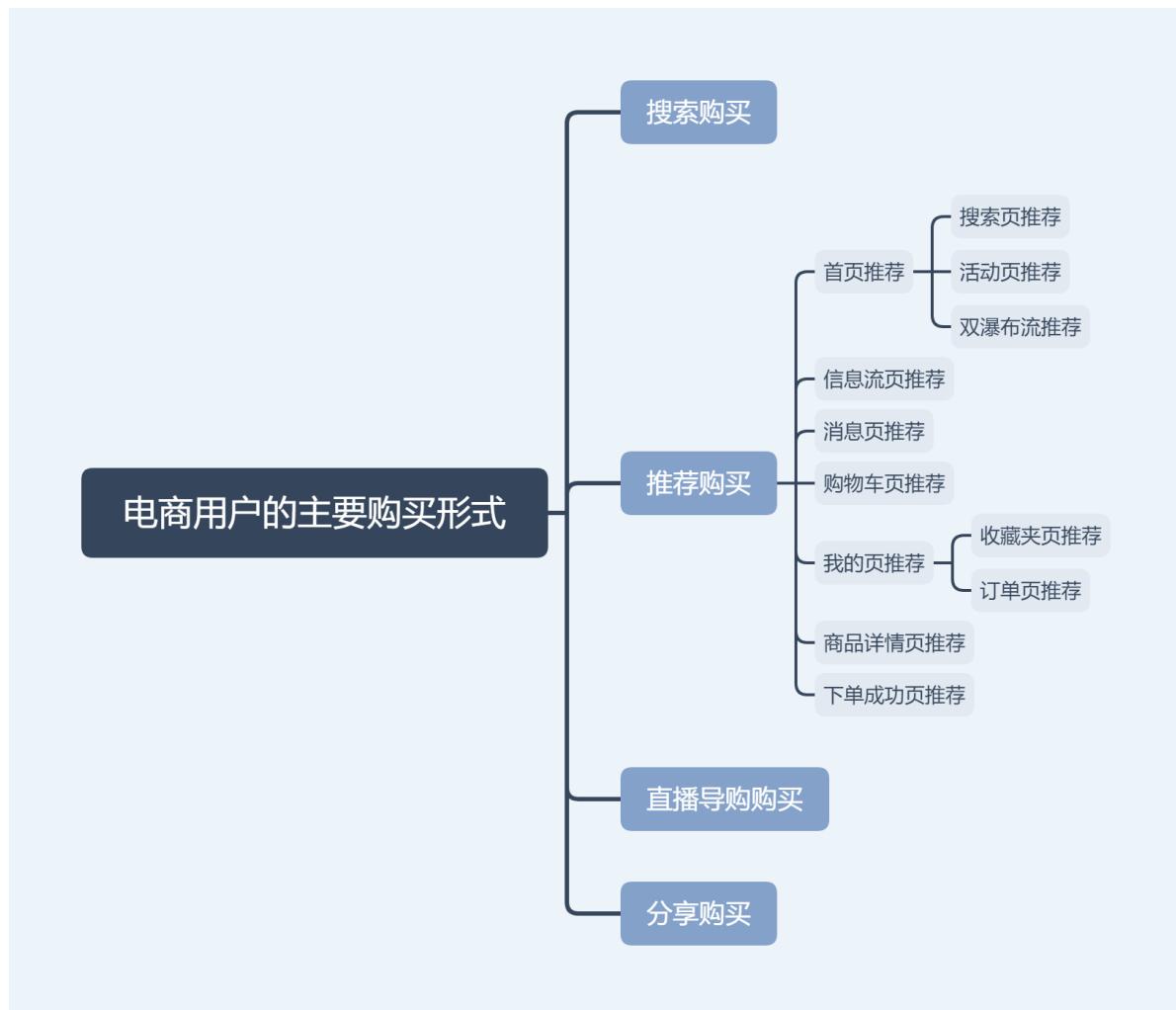
#### 1 电商平台的业务场景



### 2 拆解中小店铺流量扶持问题

#### 2.1 中小店铺面临的流量问题

- 用户在搜索页面更倾向于选择高等级店铺



搜索购买有一个特点：当用户搜索一件商品时，他很容易去对比搜索结果，根据多种情况选择最合适他的。这时，“高等级店铺”、“多付款人数”、“低价格”、“高评价”等等容易成为用户选择的因素。



- 流量较大的广告位，费用较高

而例如首页这样引流量十分大的页面，其广告费比较贵，中小店铺不太能负担得起这部分开销。所以购买这部分广告位的都是大店铺。

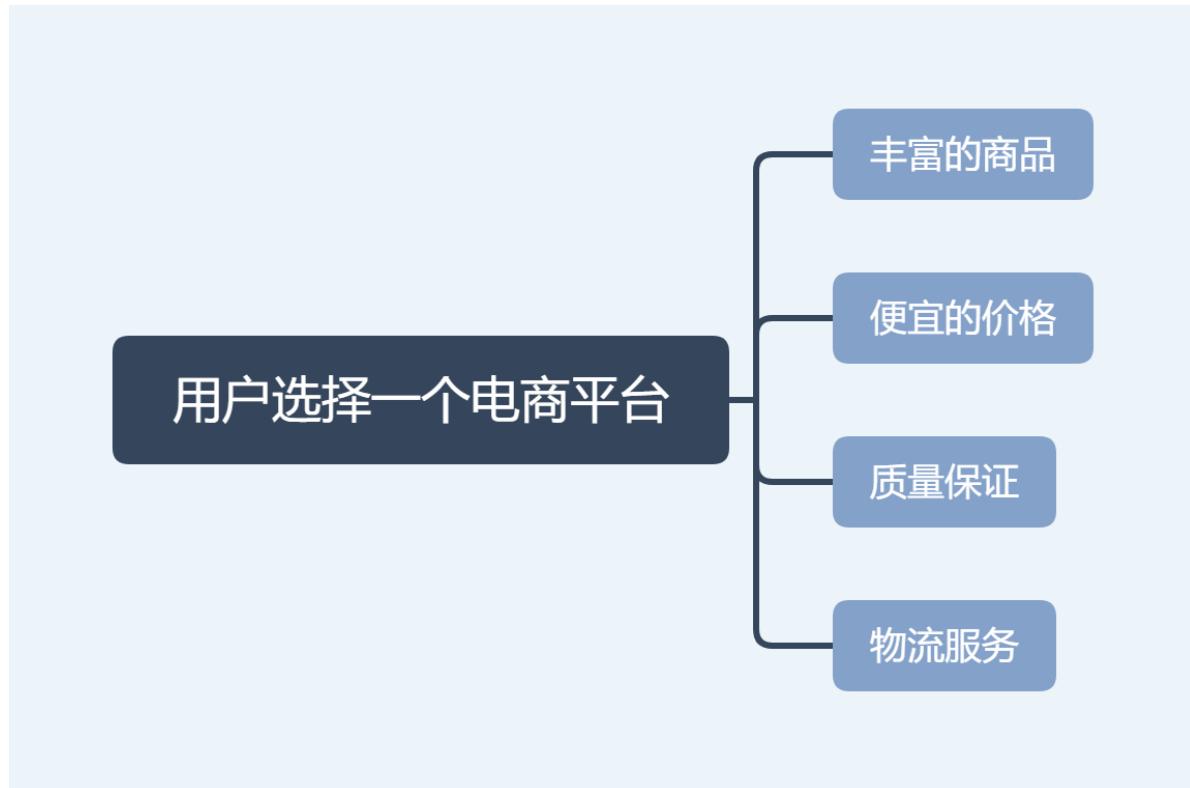
综合以上两个，就形成了大店铺流量越来越多，中小店铺流量越来越少的局面。

- 流量较大的广告位，费用较高

而例如首页这样引流量十分大的页面，其广告费比较贵，中小店铺不太能负担得起这部分开销。所以购买这部分广告位的都是大店铺。

综合以上两个，就形成了大店铺流量越来越多，中小店铺流量越来越少的局面。

## 2.2 中小店铺对于平台的价值



### 2.2 中小店铺对于平台的价值

国内常见的电商平台：  
丰富的商品是依靠丰富的垂类店铺实现的。

#### 大店铺：

- 1.更有优势的价格。
- 2.一般来说会有更好的商品质量。
- 3.能提供更好的服务。能有更多的人力预算，能够与更优质的物流公司合作。
- 4.大店铺能够有更多的仓储预算，能够售卖更多品类商品。

每个店铺展示资源、仓储资源都是有限的。

### 用户选择一个电商平台

#### 长尾理论：

一个企业的80%的利润是由20%的热门商品提供的。（28法则）  
2%的热门商品提供33%的利润，8%的中热门商品提供33%的利润。  
剩下的33%都是由90%的商品提供的。（长尾理论）



#### 丰富的商品

个性化商品（订制品、代购品、农副产品）

#### 便宜的价格

中小店铺能够做到一些差异化竞争。  
服装。

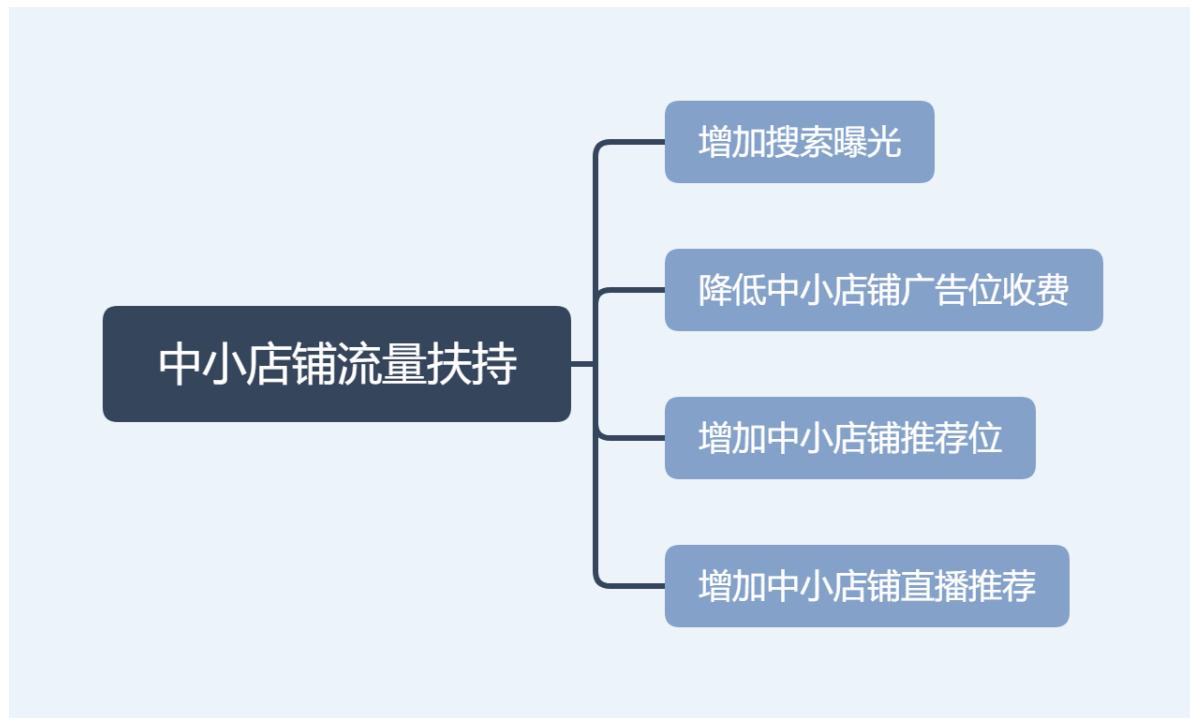
#### 质量保证

#### 物流服务

高用户粘性、高消费用户、  
高利润  
长尾商品

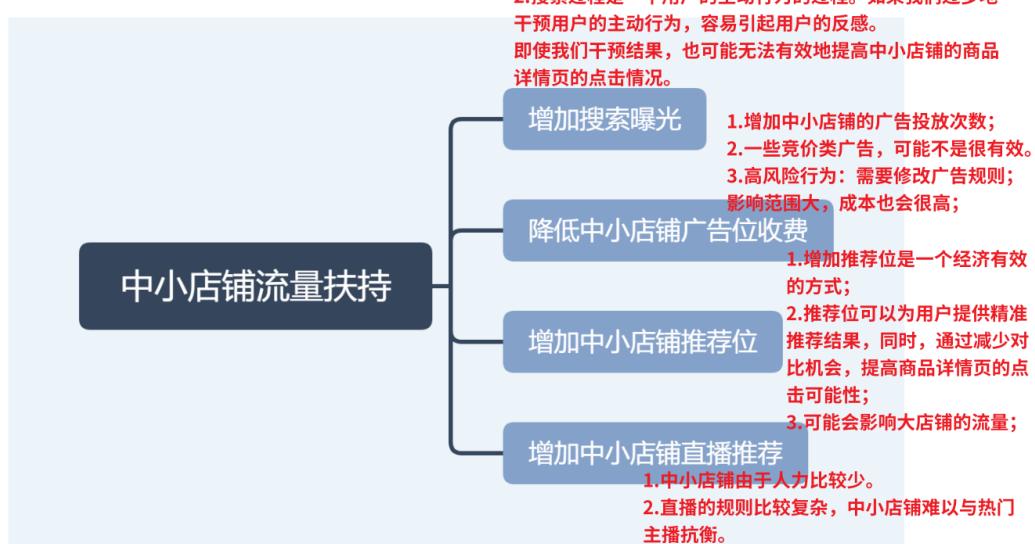
### 3 设计解决中小店铺流量扶持问题的流程

#### 3.1 业务思路

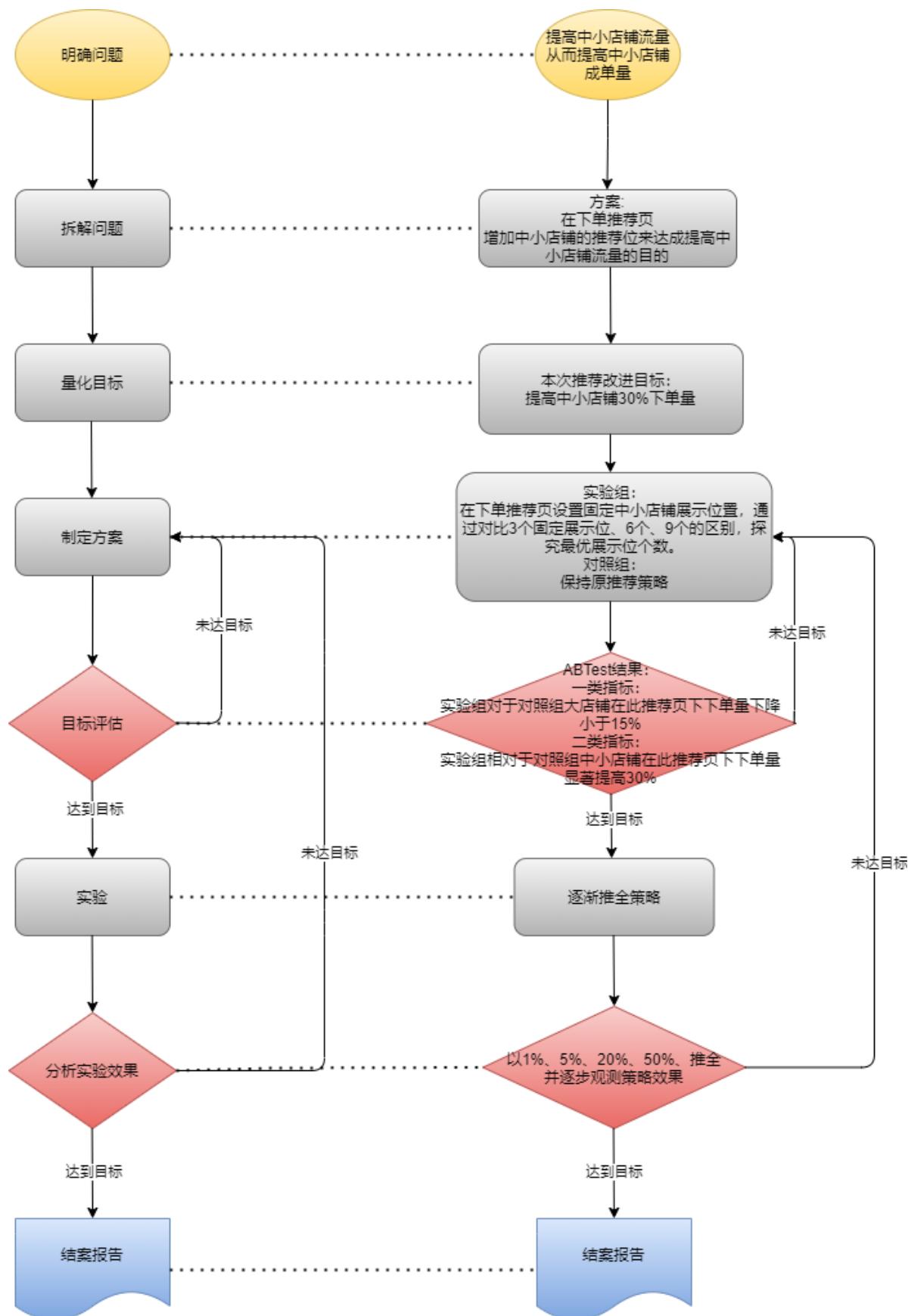


3.1设计解决中小店铺流量扶持问题的流程

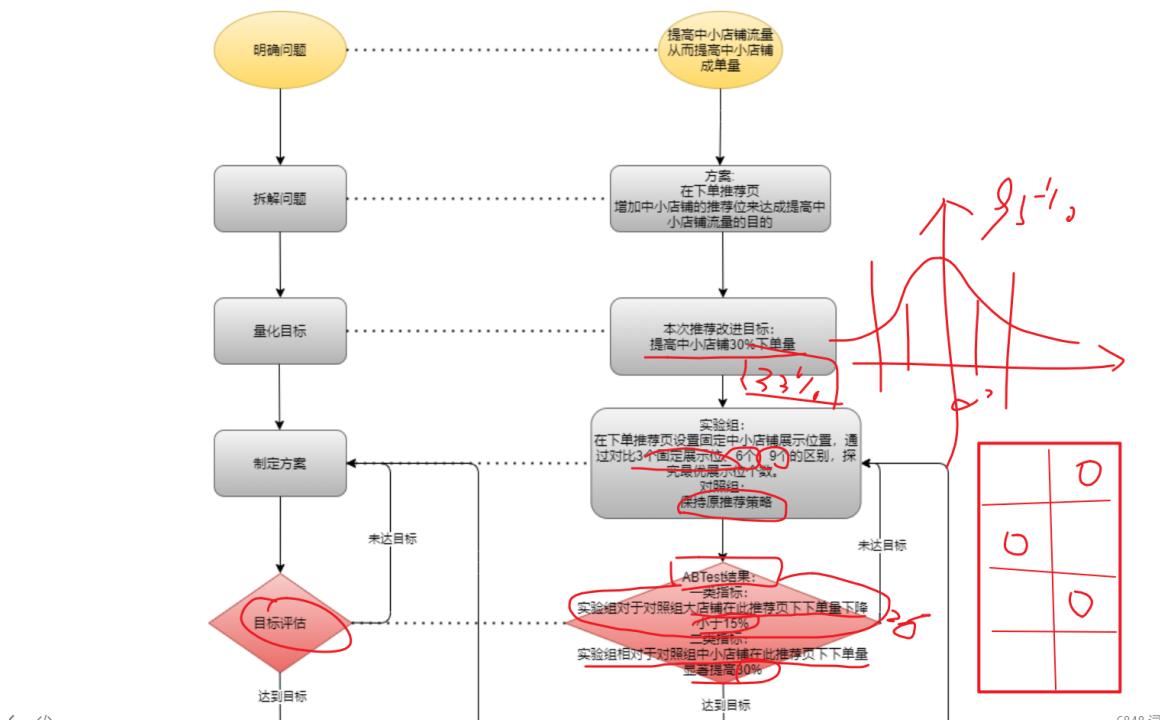
##### 3.1 业务思路



#### 3.2 解决问题流程



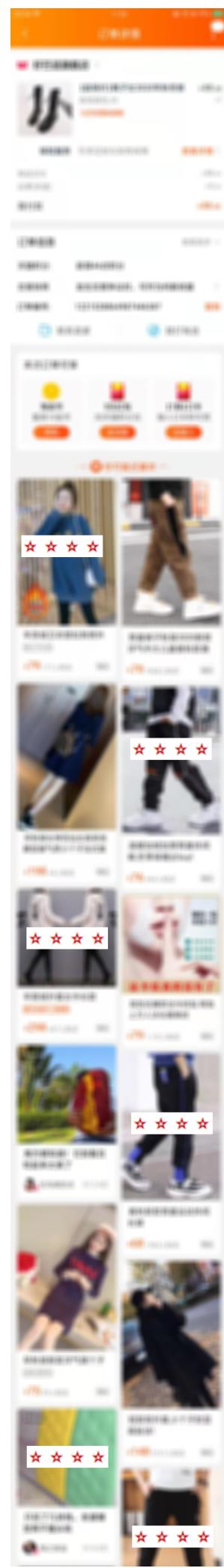
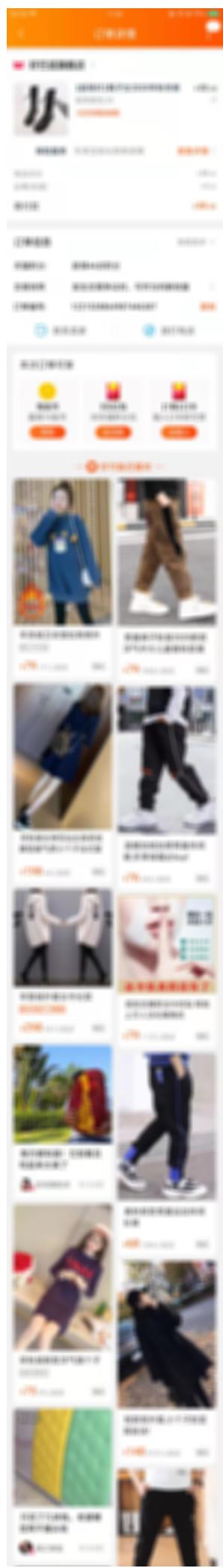
### 3.2 解决问题流程



## 第四节 利用ABTest寻求电商流量分配的最优解

实验方案:

在下单推荐页设置固定中小店铺展示位置, 通过对比3个固定展示位、6个、9个的区别, 探究最优展示位个数。





## 1 选择实验主要指标

### 1.1 二类指标的确定

目标：提升中小店铺在下单推荐页的下单量30%

由于总量无法作为假设检验统计量，通过总下单量 = 人数 \* 人均下单量，将统计量指标定为人均中小店铺下单量。此时就能确定二类指标：

二类指标：C类店铺人均下单量上升 30%



### 1 选择实验主要指标

#### 1.1 二类指标的确定

2sigma 33%

目标：提升中小店铺在下单推荐页的下单量30%

**总量** 由于总量无法作为假设检验统计量，通过总下单量 = 人数 \* 人均下单量，将统计量指标定为人均中小店铺下单量。此时就能确定二类指标：

二类指标：C类店铺人均下单量上升 30%

### 1.2 一类指标的确定

在下单推荐页将一部分流量固定分配给中小店铺，那大店铺的流量肯定会受到损失。从前期的分析来看，近一月的大店铺的流量仅有3%来源与下单推荐页，贡献成单量仅占2%。所以即使页面完全展示中小店铺结果，大店铺的整体流量影响能控制在3%内，大店铺的成单影响能控制在2%内。

一类指标：A类店铺 **人均下单量** 下降在两个标准差以内

二类指标：**C类店铺人均下单量上升 30%**

### 1.2 一类指标的确定

在下单推荐页将一部分流量固定分配给中小店铺，那大店铺的流量肯定会受到损失。从前期的分析来看，近一月的大店铺的流量仅有3%来源与下单推荐页，贡献成单量仅占2%。所以即使页面完全展示中小店铺结果，大店铺的整体流量影响能控制在3%内，大店铺的成单影响能控制在2%内。

一类指标：A类店铺 **人均下单量** 下降在两个标准差以内

15%

## 2 设计原假设与备择假设

一类指标：H1： 对照组人均下单量 - 实验组人均下单量 < 2 \* 实验前人均下单量标准差

H0： 对照组人均下单量 - 实验组人均下单量 >= 2 \* 实验前人均下单量标准差

二类指标：H1：实验组人均下单量 - 对照组人均下单量 > 30%对照组人均下单量

H0：实验组人均下单量 - 对照组人均下单量 <= 30%对照组人均下单量

### 3 计算分组样本量

选择两个总体均值之差计算公式，从假设上看一类二类指标都为单侧检验，但由于我们现在还没有做抽样，所以我们不知道实验组的 $\sigma^2$ ，我们在这里假定AB组 $\sigma^2$ 相等，推导公式为：

$$n_A = kn_B \text{ and } n_B = (1 + \frac{1}{k})(\sigma \frac{z_{1-\alpha} + z_{1-\beta}}{\mu_A - \mu_B})^2$$

python实现代码如下

```
#引用包
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import pymysql
import warnings
warnings.filterwarnings('ignore')
plt.rcParams['font.family']='SimHei'
plt.rcParams['axes.unicode_minus']= False
```

```
#读取SQL数据
con=pymysql.connect(host='127.0.0.1',port=3306,user='root',passwd='123456',db='ABTest',use_unicode=True, charset="utf8")
data=pd.read_sql("select * from ABTest",con=con)
data.head()
```

$$n_A = kn_B \text{ and } n_B = (1 + \frac{1}{k})(\sigma \frac{z_{1-\alpha} + z_{1-\beta}}{\mu_A - \mu_B})^2$$

```
# 根据常规定默认值确定α ,β,K 值 α = 0.05, β = 0.2 ,K = 1 (组件样本均衡)
alpha = 0.05
beta = 0.2
k = 1
```

求  $z_{1-\alpha}$  和  $z_{1-\beta}$

```
z_alpha = stats.norm.ppf(1-alpha)
z_beta = stats.norm.ppf(1-beta)
print(z_alpha,z_beta)

1.959963984540054 0.8416212335729143
```

求  $\sigma$

```
df_C = data.loc[(data.组=='D') & (data.店铺类型=="C"),["riqi","下单量"]]
sigema = df_C["下单量"].std()
sigema
0.4560376923365469
```

求二类目标提升值：  $\mu_A - \mu_B$

```
mean = df_C.groupby("riqi")["下单量"].mean().mean()*0.3  
mean  
0.03106555555555555
```

```
yangbenlang = (1+1/k)*np.power(sigema*(z_alpha+z_beta)/mean,2) #随着mua-miub差值  
愈大所需要的样本量愈小  
yangbenlang  
  
3382.835657318574
```

求  $\sigma$

```
df_A = data.loc[(data.组=='D') & (data.店铺类型=="A"), ["riqi", "下单量"]]  
sigema = df_A["下单量"].std()  
sigema  
  
1.90306885529098
```

求一类目标降低值:  $\mu_A - \mu_B$

```
mean = df_A.groupby("riqi")["下单量"].mean().std()*2  
mean  
  
0.19578164968724276
```

```
yangbenlang = (1+1/k)*np.power(sigema*(z_alpha+z_beta)/mean,2) #随着mua-miub差值  
愈大所需要的样本量愈小  
yangbenlang  
  
1483.2102112443183
```

日活: 1000w

平均每组0.1%流量。

## 4 检验策略选择

由于我们明确了一类指标的下降阈值和二类指标的上升目标，所以都使用单侧检验。

## 5 设计分组策略

A组：下单推荐页前12个推荐，9个C类店铺商品

B组：下单推荐页前12个推荐，6个C类店铺商品

C组：下单推荐页前12个推荐，3个C类店铺商品

D组：不干预（对照组）

## 6 实验结论分析

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import pymysql
import warnings
warnings.filterwarnings('ignore')
plt.rcParams['font.family']='simHei'
plt.rcParams['axes.unicode_minus']= False
```

```
con=pymysql.connect(host='127.0.0.1',port=3306,user='root',passwd='123456',db='ABTest',use_unicode=True, charset="utf8")
data=pd.read_sql("select * from ABTest",con=con)
data.head()
```

### 一类指标假设检验

```
# 抽取A类店铺一天数据
df = data.loc[(data.riqi=='2020-05-14') & (data.店铺类型=="A"), ["组", "下单量"]]
df.head()
```

```
df.groupby("组")["下单量"].count() #查看每组样本量
```

```
df = df[df.组 != "C"] #C组样本悬殊故C组不参与检验
df.组.unique()
```

```
df.groupby(["组"])["下单量"].mean()
```

```
#统计对照组(D组)的标准差
df_std = data.loc[data.店铺类型=="A", ["组", "riqi", "下单量"]]
std = df_std[df_std.组=="D"].groupby("riqi")["下单量"].mean().std()
std
```

```
#为让A类店铺无明显感知利润下降，我们设置A类店铺下降阈值为2个标准差
muzhicha = std *2
muzhicha

0.19578164968724276
```

## 求AD组

```
alpha =0.05
# H0: μD - μA >=0.195    H1: μD - μA <0.195
```

计算统计量  $\bar{x}_D - \bar{x}_A$

```
dif_AD = df.下单量[df.组=="D"].mean() - df.下单量[df.组=="A"].mean()
dif_AD

0.6861399544246107
```

$$\frac{s_A^2}{n_A} + \frac{s_D^2}{n_D}$$

```
varsum_AD = df.下单量[df.组=="A"].var()/df.下单量[df.组=="A"].count() + df.下单量
[df.组=="D"].var()/df.下单量[df.组=="D"].count()
varsum_AD

0.0004817503857263059
```

$$\bar{x}_D - \bar{x}_A \sim N(0.195, varsum_A D)$$

## 计算dif\_AD 在相应的分布的概率p

```
p_A = stats.norm.cdf(dif_AD, loc=muzhicha, scale=np.sqrt(varsum_AD)) # norm.dist
p_A
```

## 求BD组

```
dif_BD = df.下单量[df.组=="D"].mean() - df.下单量[df.组=="B"].mean()
varsum_BD = df.下单量[df.组=="D"].var()/df.下单量[df.组=="D"].count() + df.下单量
[df.组=="B"].var()/df.下单量[df.组=="B"].count()
# dif ~ N(0.1, varsum)
# 计算dif 在相应的分布的概率p
p_B = stats.norm.cdf(dif_BD, loc=muzhicha, scale=np.sqrt(varsum_BD))
p_B
```

```

if (p_A < alpha) & ( p_B < alpha):
    if dif_A <dif_B:
        print("A策略对A类店铺影响小")
    else:
        print("B策略对A类店铺影响小")
elif p_A < alpha:
    print("A策略对A类店铺影响小与阈值" + str(muzhicha))
elif p_B < alpha:
    print("B策略对A类店铺影响小与阈值" + str(muzhicha))
else:
    print("A.B策略对A类店铺影响超过阈值" + str(muzhicha))

```

B策略对A类店铺影响小与阈值0.19578164968724276

## 二类指标假设检验

```

df_C = data.loc[(data.riqi=='2020-05-14') & (data.店铺类型=="C"),["组","下单量"]]
df_C.head()

```

```
df_C.groupby(["组"])["下单量"].count()
```

```

#去除C组
df_C = df_C[df_C.组 != "C"]
df_C.groupby(["组"])["下单量"].mean()

```

## 计算提升值 $\mu_A - \mu_B$

```

tisheng = data[(data.组=='D') & (data.店铺类型=="C")].groupby("riqi")["下单量"].mean().mean()*0.3
tisheng
0.03106555555555555

```

## ABTest封装函数

```

def abtest(df: pd.DataFrame, alpha=0.05, group_col: str = None, value_col: str = None):
    """
    :param df: 被分析DataFrame对象
    :param alpha: 临界值
    :param group_col: 组列的名字, 默认为df的第一列
    :param value_col: 值列的名字, 默认为df的第2列
    :return:best_group_name, pdf
            best_group_name:最优组
            pdf:最优组与其他组的差异性
    ...
    # 列名
    if not group_col:
        group_col = df.columns[0]
    if not value_col:
        value_col = df.columns[1]
    
```

```

# 寻找最优组与最优质值
best_group_name = df.groupby(group_col)[value_col].mean().sort_values(ascending=False).index.tolist()[0]
best_group_values = df[df[group_col] == best_group_name][value_col] # 最优组的values
# 去除最优组的组名
group_names = list(set(df[group_col].unique().tolist()) - set(best_group_name))
# 初始化返回数据
pdf = pd.DataFrame(columns=[group_col, 'mean', 'pvalue', 'ptype'])
# 计算差异性
for group_name in group_names:
    group_values = df[df[group_col] == group_name][value_col]

    dif = best_group_values.mean() - group_values.mean()
    var = best_group_values.var()/best_group_values.count() + group_values.var()/group_values.count()

    pvalue = 1-stats.norm.cdf(dif, loc=tisheng, scale=np.sqrt(var))

    if pvalue >= alpha:
        ptype = "无显著差异"
    else:
        ptype = "有显著差异"
    # 添加数据
    pdf.loc[pdf.shape[0]] = {group_col: group_name, 'mean': group_values.mean(), 'pvalue': pvalue, 'ptype': ptype}

return best_group_name, best_group_values.mean(), pdf

```

abtest(df\_C)

	组	mean	pvalue	ptype
0	A	0.0064	0.0	有显著差异
1	D	0.1318	0.0	有显著差异)

## 7 实验结论与后续决策

一类指标评估：B策略的一类指标下降幅度在两个标准差内，符合要求；

二类指标评估：B策略的二类指标最优，同时上升幅度显著大于30%，符合要求；

决策：考虑推全B策略。

为了避免我们的取的那天数据是一个异常值，可以多用几天的数据来做检验。

## 第五节 总结与拓展

### 1 如何做一个好的ABTest

- 1、确定对照组和实验组，最好是做**单变量**的实验，一次只改变一个变量。
- 2、分流时尽量**排除混杂因素**，一般情况下采用随机分流即可。如果随机分流无法保证样本分布于总体分布一致。建议采用手动的分层随机分流。
- 3、检查流量是否达到**最小样本量要求**，达不到要求则没法进行后续的分析，实验结果不可信。
- 4、准确收集**用户行为数据**，这就要求埋点必须正确。

## 第五节 总结与拓展

### 1 如何做一个好的ABTest

- 1、确定对照组和实验组，最好是做**单变量**的实验，一次只改变一个变量。
- 2、分流时尽量**排除混杂因素**，一般情况下采用随机分流即可。如果随机分流无法保证样本分布于总体分布一致。建议采用手动的分层随机分流。
- 3、检查流量是否达到**最小样本量要求**，达不到要求则没法进行后续的分析，实验结果不可信。**如果可以扩展实验周期，我们可以将实验周期从一天扩展到一周或者n周。**
- 4、准确收集**用户行为数据**，这就要求埋点必须正确。

对比样本与总体的用户的属性分布  
最常用的属性：地域

↑  
**多水平实验，减少水平数**  
**如果可以扩展实验周期，我们可以将实验周期从一天扩展到一周或者n周。**

## 2 ABTest的局限

### 2.1 细小改变与重大改变的博弈

### 2.2 数据驱动与业务灵感的平衡

### 2 ABTest的局限

**控制变量法。一次只变化一个自变量。**

#### 2.1 细小改变与重大改变的博弈

#### 2.2 数据驱动与业务灵感的平衡

