**A/B Test**

1. **原理：**通过控制变量法，来验证新事物（产品新功能、运营/算法新策略等）对用户的影响范围大小，用来判断是否应该上线该功能和策略。所以A/B Test 从本质上来说是一个**对控制变量实验结果的假设检验。**

* **随机化：**AB实验把Treatment 到 Effect之间的所有confounder通过随机化实现了控制变量，只剩下Treatment作为唯一因素
* **假设检验：**假设检验是研究如何根据抽样后获得的样本来检查抽样前所作假设是否合理，它首先对实验组和对照组的关系提出了某种假设，然后计算这两组数据的差异和确定该差异是否存在统计上的显著性，最后根据上述结果对假设做出判断。
  + **中心极限定理**：在样本量足够大的时候，可以认为样本的均值近似服从正态分布。
  + 假设检验的理论核心是**证伪**，原假设H0是统计者想要拒绝的假设。拒绝的方式是我们看原假设成立的时候，H0（两组样本数据一样）是否是一个小概率事件。

**定义：AB testing (controlled experiments)** is a type of experiment with two variants, A and B, which are the control and variation in a controlled experiment.

* 一般Control是A——existing features，Treatment是B——new features
* 会对部分用户展开

1. **场景：**
   1. **产品迭代：**比如界面优化、功能增加、流程增加，这些 都可以使用实验测试产品改版是否成功
   2. **策略优化：**无论是运营策略还是算法策略，都可以通过AB实验的方式验证策略是否达到预期目标
   3. 问：什么时候不能用AB实验

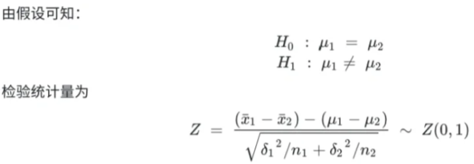
影响用户体验的实验：比如对用户价格歧视

1. **实验步骤/设计：**
   1. **确定实验指标：**一个核心（如总消费时长），多个观测（如曝光次数、播放次数、人均播放次数、人均消费时长），多个红线（不能下降：用户留存，收入）。
   2. **确定提升目标：**考虑成本和重点程度。
   3. **计算实验样本量：**根据核心指标方差，提升目标和置信度，计算满足统计功效需要的最小样本量。



* 1. **确定试验周期：**
     1. 每天可用样本量 = 每天全流量可用样本量 \* 每组可用流量比例
     2. 需要天数 = 最小样本量/每天可用样本量
     3. 考虑周期效应，对需要天数向上取到最近的7的倍数。
  2. **数据统计：**统计实验周期内多天累计数据，以每个设备/用户为最小样本
  3. **得出科学结论：**给出明确的置信区间和显著性结论。

1. 假设检验和显著性
   1. 原假设H0，检验统计量Z



* 1. **置信区间**：

根据样本对总体参数进行估计的区间范围。如果在使用interval estimation的全部区间中，有95%的interval包含population parameter，5%不包含population parameter。则interval estimation是在95% confidence level下构建的。

显著性水平为\alpha=0.05，则95%CI

如果CI包括H0，则不能拒绝原假设，结论不显著。

反之，拒绝原假设，结论显著。

* 1. P value检验

P value是样本所提供的证据对原假设H0的支持程度。P越小代表拒绝原假设的理由越充分。·

1. **辛普森悖论（Simpson's Paradox）**是指当数据被分组分析时，每个子组内的趋势与整体趋势相反或不同的现象。这种悖论揭示了在分组数据中可能出现的误导性结论。主要矛盾在于两组中样本分布不均匀导致。
   1. 实际案例：大学录取率

假设我们有两个系（A系和B系）以及它们的录取数据（看图）

如果我们看整体男女总的录取率：男性总录取率 = 25%，女性总录取率 = 46%

此时我们得出结论：女性的录取率高于男性的录取率。

但是，如果我们分系别A系的男性录取率和B系的男性录取率都比女性要高，这种差异就是辛普森悖论的一个典型案例：整体趋势与分组趋势相反或不同。

二、如何在实际工作中避免辛普森悖论

上述案例中如何避免辛普森悖论（看图）

**步骤一：分组分析（考虑混杂变量）**

在进行数据分析时，识别并控制潜在的混杂变量。比如案例中我们注意到申请者在A系和B系的分布不同。大部分女性申请者集中在录取率较高的B系，而大部分男性申请者集中在录取率较低的A系。

**步骤二：计算加权平均录取率(分层分析)**

使用加权平均的方法来计算整体的录取率，计算各子组的加权平均录取率。尽管在A系和B系中男性的录取率均高于女性，但整体上女性的录取率高于男性。这主要是由于不同系别申请者分布不均匀且每个系的录取率差异较大造成的。

**步骤三：数据可视化**

通过数据可视化来辅助理解和避免辛普森悖论。例如，使用分组的柱状图或堆积柱状图来展示各子组的录取率及其在整体中的贡献。

user\_log表，记录用户下单时间和商品，表内字段包含user\_id、date、goods\_id；求每天的去重用户数以及每天人均下单次数。

答案是：

Select date, count(distinct (user\_id)), count(goods\_id)/count(distinct(user\_id)) as avg\_goods

From user\_log

group by date

**from – on – join – where – group by – having – select – distinct – order by – limit**

1. **GMV异动分析 -** 互联网商分/经分/数分面试高频Case
   1. **Gross Merchandise Volume总交易额**：在一个特定的时间范围内，一个电商平台上所有商品的销售总额。这个指标通常被用来衡量电商平台的交易活跃程度和市场规模。特定时间内生成的总订单金额 = 已付款订单金额 + 未付款订单金额 + 取消订单金额 = （实际销售额）+（拒收订单金额 + 退货订单金额）+ （取消订单金额）
   2. **答题逻辑：**
      1. 逻辑框架：简要分析主要维度，
      2. 结合实际业务场景，分析指标含义
      3. 可落地的业务建议：
   3. **Example：**

**某电商平台上月GMV下降20%，如何分析？**

* + 1. **确认GMV异动是真异常：**确定数据的真实性（是否是“真异常”），根据数据下滑的时间节点等
    2. **外因分析（**不要占太多篇幅，点到即止。面试官更关注如何结合业务场景拆解指标）
       - 1. **自然周期（节日/季节/周期/突发性事件）影响**：比如冬天西瓜GMV下降，这是季节导致的供给下降（冬天西瓜不产了）
         2. **竞对动作：**竞对产品是否上新功能、运营是否有新动作（拉新/促销/补贴）。（确认数据是否真异常）
         3. **政策影响**
    3. **GMV指标拆解**

先**拆解**，后定位**归因（现象数据+具体下钻（现象+数据）+计算权重）**，再对应业务场景给**建议（现象原因+对应建议）**，最终**落地**

**GMV总订单金额 = 订单量 \* 单均价 = 流量 \* 转化率 \* 单均价**

* + - * 1. **流量：横向拆分，定位原因**
      * **渠道来源：**APP/小程序/PC，还是搜索页/直播间/信息流广告/活动页。

**可能原因**：活动页流量位变少

**业务建议：**迭代流量分配政策

* + - * **用户类型：**新用户/老用户；普通/vip
      * **行业品类：**服饰/美妆/食品/...
        1. **转化率**：用户行为纵向拆分，定位原因：
      * **进店**：入口数量变少、UI视觉效果变差、运营活动导流不好

**业务建议**：产品功能改版、展示图片优化、运营活动重新上线

* + - * **加购**：库存不够、预售较多
      * **下单**：价格力稍逊于其他平台
      * **支付**：页面bug
        1. **单均价**：横向拆分，定位原因：
      * **用户类型**：老用户消费能力下降且占比提升，新用户增长不足且消费能力一般
      * **行业品类**
    1. **落地方法：**
       - 1. 分析后找到相关数据交叉检验。
         2. 最终归因要与业务方确认，避免分析脱离现实。并给出落地建议。
         3. 跟踪业务动态将解决方案落地：根据业务需求搭建指标体系，追踪落地效果；如果效果不好需要复盘重新拆解。