## 数据类赛题解决方案

### 场景

我现在是一名数据分析与挖掘竞赛讲师，主要讲解现在的主流数据挖掘竞赛并提供解题方案，你需要根据我的需求给我提供对应的解决方案。

### 分析

我将给你一道题目，请你分析题目和考点，目前数据挖掘类赛题可能不仅仅涉及到数据挖掘的内容，有的适合也会考察到最优化的内容，例如线性规划、动态规划等等，你要做的是：

1. 分析每一个问题是否包含数据分析与挖掘的内容，并判断所需要建立的模型是统计分析类的模型还是最优化的模型；
2. 如果问题中包含数据分析与挖掘的内容，分析适用于当前场景下的数据分析与处理方法，要具有针对性而非通用性；
3. 目前数据挖掘类问题主要包含三个类别：

(1)时间序列分析，单变量或多变量的时间序列分析，核心在于挖掘序列本身存在的规律；

(2)对照实验：常使用控制变量的方法，即一个存在某特点，另一个不存在某特点的二元、多元分类或回归场景，被控制变量的部分通常会使用特定标识符或空值表示该部分不存在

(3)多元统计分析：和对照实验的唯一差别是没有观察组和对照组，属于比较常规的预测问题。

分析每个问题归属于哪一类别

针对每个问题，对于可能出现的不同情况，推荐2到4种解题方案，其详细解题思路及步骤用逻辑图markdown代码格式给出，语法可被xmind识别，其中不包含公式和英文缩写。逻辑图整体呈轴对称，左右(上下)结构，根节点在左侧，美观一些，色系统一，内容详细，使用适合论文风格的浅色系，直线箭头，不要有多余模块。

根据你提供的方向，给我一个可以完整求解该题目中所有数据挖掘类问题的解决思路，该比赛更多的考察数学与统计学上的严谨性与前后逻辑性，每题使用模型的要求：

1. 必须是最匹配当前问题场景的一个模型，尽可能使用基础统计学模型，保证强解释性
2. 当求解到问题二及后续的问题，必须要与前面已经完成的解决方案思路相匹配，逻辑逐步递进，模型逐步稳健，使得问题的解决方案与解决方案之间看起来更有整体性
3. 如果没有相对应的真实的量化结果，不要引入模型比较
4. 符合该比赛的考察标准，在选择方案时增加该板块的权重

* 这部分内容将作为论文的问题分析部分给出，该部分只涉及每个问题的思路，不涉及任何结论与总结，不涉及模型的建立，给出的内容不要包含我提示词的内容，直接用几个段落表示出来，保证学术性、论文性与数学性

修正：上传一篇优秀论文的思路部分的截图，末尾加上参考这个图片中的写作风格，根据它的风格，编写现在这道题目的思路，符合以上要求。

进一步丰富每段字数，使其每问不低于300字，同时提高描述的学术性，专业性，数学性，减少AI风格，语言干练，其文本风格应以完全可以用于论文之中，不要出现任何我前面提示词中出现的对你要求的内容，这会使得论文性变弱，以段落文本的格式给出。 这一部分不包含总结和结论

### 模型建立

基于你刚刚给出的最合适的解题思路，请根据赛题及附件内容，帮我生成第X问详细的解决方案，必须要和当前场景完全契合，而非通用解决方案。该内容应该至少百分之八十是数据分析与挖掘的内容，剩下的才是模型建立与分析，数据分析与挖掘的内容一定要详细，将题目场景和分析过程紧密结合起来，切忌随意套用模型，这一部分内容一定要全面，包括必要的分析，建模过程，结果展示，结果分析等小节，不包括应用推广、符号说明、模型假设等内容，小节的标题请自拟为与主题相关的。

上述回答的不错，对每一节的内容进行扩充，使其文本篇幅在原先的X倍以上，同时进一步增强论文的学术性、专业性、数学性，减少口语化。

换GPT：根据附件内容，按照以下思路，编写相应代码

### 模型检验

据上述各小问的求解结果，针对每一问的模型给出详细的分析，如果是优化类模型，包括灵敏度分析，描述性分析；如果是统计类模型，包括其解释性分析，稳健性分析；结论背后发现的隐藏的规则分析，以及其实际应用意义等等，标题自拟，符合论文规范即可。同样保证你给出内容的论文性、专业性、数学性，使其满足一名优秀的数据分析与挖掘竞赛讲师的标准。

### 摘要

根据你的解决方案，编写这道题的摘要部分，800-1000字

### 参考文献

复制摘要部分到scispace