**A题 影响城市居民身体健康的因素分析**

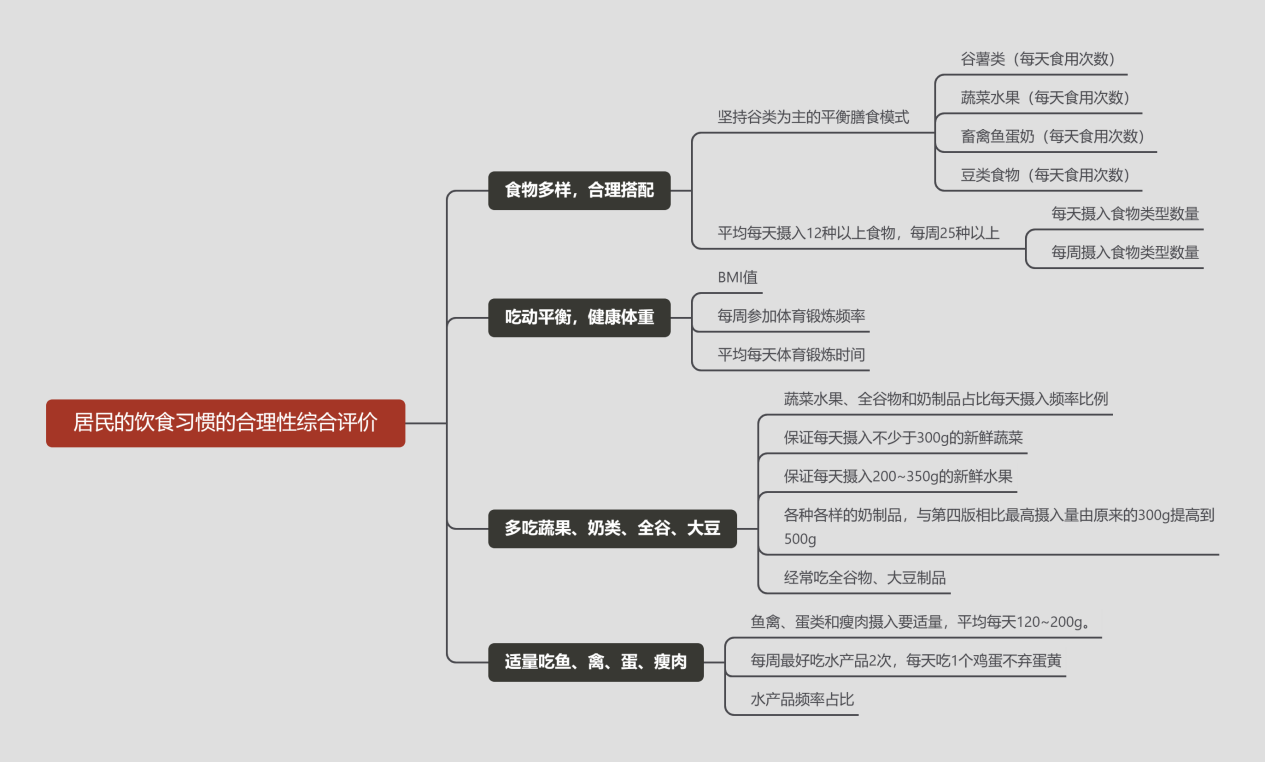
以心脑血管疾病、糖尿病、恶性肿瘤以及慢性阻塞性肺病为代表的慢性非传染性疾病（以下简称慢性病）已经成为影响我国居民身体健康的重要问题。随着人们生活方式的改变，慢性病的患病率持续攀升。众所周知，健康状况与年龄、饮食习惯、身体活动情况、职业等都有密切的关系。如何通过合理地安排膳食、适量的身体运动、践行健康的生活方式，从而达到促进身体健康的目的，这是全社会普遍关注的问题。附件A1是某市卫生健康研究部门对部分居民所做的“慢性非传染性疾病及其相关影响因素流行病学”调查问卷表，附件A2是相应的调查数据结果，附件A3是中国营养学会最新修订的《中国居民膳食指南》中为平衡居民膳食提出的八条准则。

请你们团队研究解决下面问题：

**问题1 参考附件A3，分析附件A2中居民的饮食习惯的合理性，并说明存在的主要问题。**

搭建一个指标体系，用于衡量居民的饮食习惯的合理性，根据附件A3整理出指标体系，然后对每一项指标绘制图表进行描述性分析，说明居民的饮食习惯与附件3《中国居民膳食指南》的差距。

这里的难度是整理数据，分析难度并不大



**问题2 分析居民的生活习惯和饮食习惯是否与年龄、性别、婚姻状况、文化程度、职业等因素相关。**

解法1：相关性分析，首先可以整理出生活习惯指标与饮食习惯指标的相关变量，然后逐个对年龄、性别、婚姻状况、文化程度、职业等因素进行关联分析，然后对前面关联分析后的结果进行整合，获取其相关系数的均值，进而确定整体是否与以上因素存在相关关系，以及个体上，哪些变量相关性低，或者不呈现相关性。

解法2：逻辑回归，首先可以整理出生活习惯指标与饮食习惯指标的相关变量，这些变量作为X，然后以此对年龄、性别、婚姻状况、文化程度、职业等人口因素做为Y，例如以性别作为Y举例，先分析其F检验是否呈现显著性，如过存在显著性，那么说明整体上是存在影响关系的，接着查看每一项的标准化回归系数，核对个体上的显著性关系；

解法3：机器学习+模型解释（shap模型），同方法2，先核对指标，然后通过机器学习建模分类或回归模型，把模型输入shap模型之类，这样可以从非线性的角度确定各个指标对人口因素（Y）的影响

问题3 根据附件A2中的数据，深入分析常见慢性病（如高血压、糖尿病等）与吸烟、饮酒、饮食习惯、生活习惯、工作性质、运动等因素的关系以及相关程度。

这道题同问题2一样，唯一的不同是把Y变了，这里的Y为（0：没有患病，1：有高血压或糖尿病），然后再对这些变量进行整理即可，建议问题2可以采用解法3，然后套入与问题2一样的解法，这样问题3解题难度就降低了，想要炫技的话可以采用不同的机器学习来对比。

问题4 依据附件A2中居民的具体情况，对居民进行合理分类，并针对各类人群提出有利于身体健康的膳食、运动等方面的合理建议。

这道题的关键核心是分类的方向，从题意来看很多类型的分类，例如对于有无患病（高血压或糖尿病）、亦或根据人口特征分类，例如少年、青年、中年、老年，或者肥胖群体，再或者饮食习惯等等，所以事实上这道题有很多的做法，但是万变不离其中，再做完分类后，针对各类人群提出有利于身体健康的膳食、运动等方面的合理建议这个做法是都是一样的分析步骤，这个分析可以直接复制第一问的分析，只是这个时候是按人群来划分了