

影响城市居民身体健康的因素分析

摘要

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。基于以上信息，本研究旨在从多个角度深入分析影响城市居民身体健康的因素，以期改善城市居民健康状况提供有益的指导意见。本文通过建立模型，对约八千位居民的生活信息与健康状况进行了研究。我们将分别就饮食习惯的合理性、生活习惯和饮食习惯与人口特征之间的关系、慢性病与各项生活因素的关联性，以及针对不同人群提出的健康建议等问题展开深入研究和探讨。通过数据分析、数学建模以及相关理论探讨，我们将力求为城市居民的健康问题提供科学合理的解决方案，为未来的健康政策制定和健康促进提供有益的参考。

针对该问题，本文采用Python编程语言编写脚本，对附件A2中所提供的特定地区慢性病及其相关影响因素的流行病学数据进行了详尽的读取、细致的分析以及全面的总结。同时，本文以附件A3中所列的平衡膳食八大原则为基本框架展开逐步分析工作。通过编写运行Python脚本，我们能够高效地提取所需数据，并借助数据分析技术深入挖掘数据背后的信息。在此基础上，结合平衡膳食八大准则，我们系统地剖析慢性病与营养因素之间的关联，为四个问题提供全面且有深度的解答。

关键词：

一、问题重述

1.1 问题背景

在当今社会，城市化和现代化进程带来了社会经济的快速发展和居民生活水平的显著提升。然而，与此同时，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）在城市居民中逐渐崭露头角，成为影响健康的突出问题。这些慢性病不仅对个体健康产生严重威胁，也在一定程度上构成社会医疗资源的巨大负担。

心脑血管疾病、糖尿病、恶性肿瘤和慢性阻塞性肺病作为慢性病的代表，在城市居民中的患病率不断攀升，引发了广泛的关注。这些疾病通常具有长期的发展过程，而其对健康的影响也愈发显著。心脑血管疾病可能导致中风和心脏病发作，糖尿病会损害多个器官系统，恶性肿瘤具有较高的致死率，慢性阻塞性肺病则限制了呼吸系统的功能。这些疾病不仅影响个体的生活质量，还对医疗系统和社会经济造成了巨大的挑战。

近年来，城市居民的生活方式正在发生急剧变化，这导致了慢性病的患病率持续上升。这一变化在全球范围内都得到了广泛关注。人们的生活节奏加快，饮食结构发生变化，身体活动水平下降，都成为慢性病高发的因素。此外，现代化的职业结构和工作环境也在一定程度上影响了健康状况。这使得研究健康与生活方式、职业等因素之间的关系变得尤为重要。

在当前社会背景下，健康状况与多个因素之间的关系备受关注。年龄、饮食习惯、体育锻炼水平以及职业等因素，都被认为与健康紧密相连。年龄是慢性病发病率的重要因素，随着年龄增长，慢性病的风险也逐渐升高。此外，城市居民的饮食习惯与健康密切相关。现代社会的高盐、高糖、高脂的饮食结构，往往加剧了慢性病的发展。同时，缺乏适度的体育锻炼也使得居民更容易罹患慢性病。职业因素也是影响健康的重要因素之一，某些职业可能存在较高的工作压力和不良的工作环境，进而影响身体健康。

随着健康问题的凸显，社会各界开始广泛关注如何通过调整生活方式，提高居民的健康水平。合理膳食、适度体育锻炼以及健康生活习惯的养成，成为许多人关注的焦点。这些措施被认为是预防和控制慢性病的有效途径，有望为城市居民的健康带来积极的影响。

1.2 目标任务

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。

问题一：参考附件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。

问题二：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。

问题三：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。

问题四：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。

二、问题分析

尽管问题一、二、三、四各自侧重点不同，然而它们共同关注附件A2中居民的饮食习惯，以探究是否符合参考标准以及是否存在相关性等问题。首要任务在于整合、归纳和分析附件A2中的调查数据结果。由于附件A2所涵盖的表格数据庞大且呈现出明显的规律性，为此，本文着手编写并执行多个Python脚本，以实现附件A2数据的读取、归纳和综合总结。在此基础上，以解答问题一为目标，我们将附件A2中的数据与附件A3中《中国居民膳食指南》中所呈现的八项平衡膳食准则进行对比分析，从中提炼必要的图表与结论。

针对问题二，我们运用（方法）编写Python脚本，分析附件A2中生活习惯和饮食习惯与年龄、性别、婚姻状况、文化程度、职业等因素之间的相关程度。问题三与问题二相近，需要探究常见慢性病与吸烟、饮酒、饮食习惯、生活习惯、工作性质、运动等因素之间的关系及相关程度。分析方法与问题二类似，关注点在于不同的分析对象。

在涉及问题四时，我们需根据附件A2中居民的实际情况，对居民进行合理分类，据此为不同类别的人群提供有益于身体健康的膳食、运动等方面的恰当建议。解决这一问题可以采用多种分类标准，并从饮食习惯、生活习惯、工作性质、运动等多个角度出发，为不同类别的居民量身打造切实可行的健康指导方案。在这一过程中，将充分考虑个体差异，以为不同人群提供量身定制的健康建议。

三、模型假设

为了适当地对模型进行合理简化，本文给出如下假设或说明（部分）：

1. 根据第七次全国人口普查结果，平均每个家庭户的人口为2.62人。
2. 各种酒比例系数占比依据平均度数确定。
3. 根据世界卫生组织发布的《全球食品安全标准》和《全球膳食指南》，专家推荐的每人每天摄入油炸食品的量不超过25克。
4. 由于附件A2“油炸食品”一项数据没有明显相关结果可以得出，所以不做分析。

四、符号说明

五、模型的建立与求解

问卷调查表的结果为xlsx格式，本文编写Python脚本对数据进行预处理，包括异常值剔除、基本信息处理、生活习惯与饮食情况处理、健康指标处理等步骤。

BMI计算与异常值处理

根据体重和身高计算BMI指数。计算方法如下：

$$BMI = \text{体重} / \text{身高}^2$$

注意：这里体重单位是千克，身高单位是米。

由于一些样本的身高或体重数据填写存在明显错误，我们以BMI>72为条件，删除异常BMI值的样本，以确保数据真实可信。

基本信息处理

根据出生年计算年龄列，这里我们假设问卷年份为2023年。

根据问卷表中对应关系，处理性别、民族、婚姻状况、文化程度、职业等分类特征。

吸烟情况预处理

首先，根据年龄和开始吸烟年龄计算出烟龄。对于填写“记不清”的样本，我们假设其开始吸烟年龄为中国居民平均开始吸烟年龄19岁。其次，计算出平均每日吸烟支数CPD。最后，根据烟龄和CPD计算出吸烟指数。计算方法如下：

吸烟指数 = 烟龄 * CPD

饮酒情况处理

对于饮酒情况，我们计算每个样本的日均酒精摄入量。对不同种类的饮酒情况进行加权处理，取每种酒类的平均酒精含量。对应关系如下表：

酒的品类	酒精含量
高度白酒	45%
低度白酒	25%
啤酒	5%
黄酒、糯米酒	15%
葡萄酒	10%

饮食习惯处理

全谷物、鱼禽蛋瘦肉、奶制品、大豆制品

全谷物包括大米、小麦面粉和杂粮。将原始单位两换算成以克为单位。

鱼禽蛋瘦肉包括水产类、禽肉和蛋类。这里各种蛋类取平均值50克/颗。

奶制品包括鲜奶、奶粉和酸奶。这里假设一勺奶粉换算成50克鲜奶**。

大豆制品包括豆腐、豆腐丝、豆浆和干豆。对不同种类的大豆制品进行加权处理，取其平均大豆含量。对应关系如下表：

大豆制品	大豆含量
豆腐	22%
豆腐丝	67%
豆浆	6%
干豆	100%

含糖饮料

含糖饮料包括果汁饮料和其他饮料。对于这两类饮料，分别计算其平均含糖量，以克/杯为单位。对应关系如下表：

含糖饮料	含糖量
果汁饮料	53
其他饮料	43.5

食用油、盐

问卷表中，食用油和盐的食用量以家庭为单位，我们换算成以个人为单位。（以中国家庭平均人口为2.62人计算）

体育锻炼情况处理

计算每人每天体育锻炼时间。

健康指标处理

血压水平

根据收缩压和舒张压计算血压水平，根据血压值进行分类。分类依据如下表：

血压水平	收缩压	舒张压
单纯收缩期高血压	≥ 140	< 90
3级高血压	≥ 180	≥ 110
2级高血压	≥ 160	≥ 100
1级高血压	≥ 140	≥ 90
高血压	≥ 140	≥ 90
正常高值	≥ 100	≥ 80

糖代谢状态

根据空腹血糖值计算糖代谢状态，根据血糖值进行分类。分类依据如下表：

糖代谢状态	空腹血糖
糖尿病	≥ 7.0
空腹血糖受损	≥ 6.1

血脂水平

根据胆固醇、甘油三酯和高密度脂蛋白的值计算血脂水平，根据不同指标值进行分类。分类依据如下表：

血脂水平	胆固醇	甘油三酯	高密度脂蛋白
高胆固醇血症	>5.72	≤ 1.70	
高甘油三酯血症	≤ 5.72	>1.70	
混合型高脂血症	>5.72	>1.70	
低高密度脂蛋白血症			<0.90

尿酸水平

根据尿酸值计算尿酸水平，根据性别和尿酸值进行分类。

尿酸水平	尿酸值	性别
高尿酸血症	>420	男
高尿酸血症	>360	女

下面是对问题一、二、三和四的解答。

问题一

一、食物多样，合理搭配

- 1、坚持谷类为主的平衡膳食模式，每天的膳食应包括谷薯类、蔬菜水果、畜禽鱼蛋奶和豆类食物。
- 2、平均每天摄入 12 种以上食物，每周 25 种以上。做好荤素搭配、主副搭配，与第四版相比突出合理搭配的重要性。

二、吃动平衡，合理运动，适当减肥

- 1、各年龄段人群都应天天进行身体活动，保持健康体重。健康体重测算方法： $18.5 \leq \text{BMI}$ (公斤体重/身高米的平方) < 24

- BMI判断标准数据来源：附件3中国居民膳食指南平衡膳食八条准则【1】

数据分析：根据BMI指数的分布来看，大约有55%的达到了正常标准，38%的人存在超重现象，7%的人身体消瘦可能会带来营养不良，通过查阅资料我们发现BMI指数大于40已经属于三类肥胖，针对三类肥胖人群的身体健康，往往需要药物或者减肥手术的介入，饮食方面也与正常人的标准不同，因此我们考虑把这类人群单列出来，先分析剩余人群的BMI分布的特征。

下图是刨除三类肥胖人群后的总体BMI分析，可以看出正常，超重，身体较轻人群占比无明显变化，从分布图中我们可以看到具有较好的对称性，因此我们可以选择重抽样方法去计算BMI的均值，计算后我们发现，该地区人群BMI的均值有95%的可能性落在[23.06, 23.21]这个区间，这与我们对饼状图的观测是相符的，从而可以知道这个群体总体是健康的，对大多数人来说只要坚持以下准则就可以保证身体健康。

- 2、食不过量，保持能量平衡。减少久坐时间，每小时动一动。

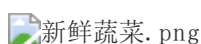
- 3、坚持日常身体活动，每周至少进行 5 天中等强度身体活动，累计 150min 以上；主动身体活动最好每天 6000 步。鼓励适当进行高强度有氧运动，加强抗阻运动 2~3 天/周，大约80%的群体运动量不足，对中老年人来说长期的运动不足将带来一系列危害。一些主要的影响包括肌肉减弱和骨密度降低，心血管疾病，代谢紊乱，心理健康问题，运动能力下降。因此，对于中老年人来说，定期参与适度的运动活动非常重要，可以选择合适的运动方式，如散步、跳舞、游泳、瑜伽等。但在开始运动前，最好咨询医生的建议，并根据自身情况制定适宜的运动计划。



针对三类肥胖群体，专家建议通过减少膳食热量、增加体力活动、改变行为习惯等生活方式干预，将体重控制到正常范围；生活方式干预同时配合减重药物治疗；重度肥胖患者（ $\text{BMI} \geq 35.0 \text{ kg/m}^2$ 或 $\text{BMI} \geq 32.5 \text{ kg/m}^2$ 合并 2 型糖尿病），可考虑手术减重。【2】

三、健康饮食，多吃蔬果、奶类、大豆，控制全谷物摄入

- 1、蔬菜水果、全谷物和奶制品是平衡膳食的重要组成部分。
- 2、保证每天摄入不少于 300g 的新鲜蔬菜，深色蔬菜应占 1/2。



- 数据判断标准来源：【1】

从饼状图中可以看出，新鲜蔬菜的摄入量不足与充足占比基本上是一致的，营养学方面知识告诉我们对摄入量不足的人群来说可能会带来营养不均衡，免疫力下降，消化，心血管，皮肤以及睡眠方向的问题。同时我们也需注意到，该地区人均蔬菜摄入量为284.95g/天，而根据中国统计年鉴2022发布的信息全国对蔬菜的人均日消费量为290.96g克，考虑到烹饪过程中蔬菜的损耗，我们有理由相信该地区居民的人均日蔬菜摄入量是高于全国整体水平的【3】，只要坚持每天摄入不少于 300g 的新鲜蔬菜，其中深色蔬菜如菠菜、西兰花、胡萝卜等应保证摄入占比不少于一半即可。

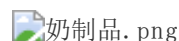
3、保证每天摄入 200~350g 的新鲜水果，果汁不能代替鲜果。



- 数据判断标准来源：【1】

从饼状图中可以看出大约有74%的人群属于水果摄入量不足，6%的人群水果摄入量过多，只有20%的人群水果摄入量属于正常的，这可能是由当地居民的消费观念和消费水平引起的。在一项GBD2015的研究中与水果摄入不足相关的疾病包括缺血性心脏病、缺血性脑卒中、出血性脑卒中、糖尿病、食管癌、喉癌、唇癌、口腔癌、鼻咽癌、气管、支气管和肺癌以及其他咽喉癌等慢性非传染性疾病【5】，因此水果摄入不足会带来很大的风险，但是同时水果也不能摄入过多，水果中富含大量的果糖，研究表明食用果糖过多会带来高血压的风险【6】，为了判断这样的饮食结构是否会导致血压的升高，分别计算正常人群与摄入过多人群的平均舒张压以及收缩压，分别为(74.32810750279955, 112.65845464725643)和(73.49658314350798, 111.49886104783599)，可知基本上没有差别，同时都处于正常血压范围，因此不会引起相关风险，针对水果摄入人群结构我们只需关心摄入不足的人群。

4、吃各种各样的奶制品，与第四版相比最高摄入量由原来的 300g 提高到500g。



- 数据判断标准：【1】

奶制品：从饼状图中可以看出，该地居民奶制品摄入结构极其不均衡，95%的人群奶制品摄入不足，奶类摄入不足是我国居民钙摄入不足的主要因素【7】，钙摄入不足会带来一系列的骨质疾病，但同时我们应该看到调查人群的年龄结构是中老年，对于该部分人来说钙的摄入超量会带来骨折，心脏病问题，对于该部分人群，我们的建议是奶制品摄入适量就好，可以通过每天增加一杯牛奶，使每天的钙摄入量达到800mg，不仅增加蛋白质，钙及维生素B1，B2，B6和维生素A，维生素K的供应，同时对心血管健康也是有益无害。

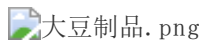
5、经常吃全谷物、大豆制品，适量吃坚果。与第四版相比突出全谷物在饮食中的地位。



- 数据判断标准：【8】

全谷物：谷物是膳食的重要基础组成部分，也是膳食纤维、B族维生素、矿物质等营养素的重要食物来源【9】。从饼状图中可以看出，该地区群体的全谷物摄入饮食结构存在不均衡现象，有88%的人群全谷物摄入是超标的，2.7%的人群摄入量不足，仅有9.6%的人全谷物摄入是正常范围的，根据调查人群的主要年龄结构，出现这种现象可能有两方面的原因，第一个是长期以来的饮食习惯根深蒂固，第二个是增加全谷物的摄入有助于降低全因死亡风险，降低心血管疾病，II型糖尿病，结直肠癌发病风险，维持正常体重，延缓体重增长【7】，但同时我们应该注意到也有少部分的全谷物摄入量是严重超

标的，这部分人群应该减少全谷物的摄入否则将给消化系统带来极大的负担，容易导致肥胖，从而引发一系列慢性病。

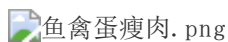


- 数据判断标准：【10】

大豆制品：豆类具有极高的营养价值，是植物性蛋白中唯一的优质来源，富含赖氨酸。以黄豆为基数，分别计算其他豆制品的营养指数，则得到如下图所示的分布图和饼状图，即53%的人群大豆摄入不达标，26%的人群大豆摄入超标，21%的人群大豆摄入处于正常水平，适量大豆的摄入增加有助于降低心血管发生风险，并且对女性人群而言可降低乳腺癌的发生【7】，但是摄入过多会对甲状腺带来极大负担，因此我们建议大豆摄入量不应低于正常水平，在正常水平基础上同时不应超过其5克，这是比较健康的饮食标准。

四、适量增加鱼禽肉蛋摄入

1、鱼禽、蛋类和瘦肉摄入要适量，平均每天 120~200g。



数据判断标准：【1】

鱼禽、蛋类和瘦肉均属于动物性食物，富含优质蛋白质、脂类、脂溶性纤维素、B族维生素和矿物质等，是平衡膳食的重要组成部分。从饼状图中可以看出该地70%居民存在鱼禽肉蛋摄入量不足的现象，11%摄入量超标，只有19%处于安全健康摄入量，饮食结构存在极大的失衡，考虑到调查人群的年轻主体为中老年人以及2018至2021年市场上相关产品的价格指数涨幅（见下图）【3】，推测可能是由于过去的生活水平习惯以及当前的市场环境导致的。增加鱼禽肉蛋的摄入量有助于降低全因死亡的风险，降低脑卒中转移的风险，特别的可以降低中老年人痴呆及认知功能障碍发病的风险【7】，因此根据膳食宝塔，我们建议

1、每周最好吃水产品 2 次，每天吃 1 个鸡蛋不弃蛋黄。

2、优先选择水产品，少吃肥肉、烟熏和腌制的肉制品。

3、水产品相对于畜肉脂肪含量更低，富含的不饱和脂肪酸更利于保护心血管系统。与第四版相比，水产品的摄入量更明确。同时，因鸡蛋的高氨基酸评分和高质量蛋白，而被着重强调。

五、少盐少油，控制糖分，正常饮酒

1、培养清淡饮食习惯，少吃高盐、高糖和油炸食品。

数据判断标准：



2、成人每天摄入烹调油 25~30g，控制添加糖在 25g 以下。



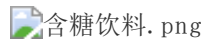
数据标准来源：【1】

食用油：食用油在食物的烹饪过程中有着功不可没的作用，既可以做到使菜肴增香又富含丰富的营养价值，然而食用油的摄入量得符合一定的标准。从饼状图中可以看出，调查人群中79%属于摄入超标，15%是摄入不足，只有7%的人是处于正常水平的摄入量。高脂肪酸的摄入将导致肥胖的发生，分别计算摄入量大于80以及处于正常水平两类人群的平均BMI值，分别为23.65以及23.64，在进行齐方差异性检验后得到p值为0.98，据此可进行t检验，得到p值为0.98，据此判断两边的BMI水平并无明显区别，且均处于正常水平。但这并不意味着摄入过多食用油没有风险，根据调查研究显示植物油中富含的反式脂肪摄入过多将会导致心血管方面的疾病【7】，因此对该地区人群，我们的建议是尽量吃清淡一点。

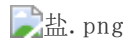
3、不喝或少喝含糖饮料，特殊人群不应饮酒。

数据判断标准来源：【11】

糖类：这里谈论的糖类主要为游离糖，即额外摄入糖，由于表格中所给单位为杯因此我们需要进行一定的计算，查阅资料可以知道，果汁类饮品的含糖中位数为10.6g/100ml，其它饮料的中位数为8.7g/100ml，若考虑一杯为500ml，则如下图所示该地居民额外糖类摄入量，可以看出75%处于正常摄入水平，只有25%是超标的，同时分布上具有较为严重的厚尾现象，对于糖类摄入过多的人群有调查报告显示，每多摄入250ml的含糖饮料将会使得成年II型糖尿病的风险增加18%【7】，这一部分人可能需要药物方面的协助，对于其他人群每天应尽量少喝饮料，最多不应该超过三杯。



4、与第四版相比，成人每日食盐和酒精摄入量控制更严格。食用盐<5g，无论男性女性摄入酒精<15g。

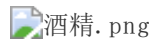


数据标准：【1】

盐：盐具有杀菌，调味道，促进新陈代谢的作用。从该地区的饼状结构图可以看出该地区群体摄入结构占比存在小程度上的失衡，但是由于超量中84%的人群只是轻微超量，推测这样的超量原因可能是由于当地烹饪习惯或者所处地理位置决定的，对于这部分群体应该因地制宜地改变标准，可认为他们盐摄入量是正常的，对于不属于这部分的超量群体应该告诫他们尽量少食盐。

数据标准：【1】

咨询身边喝酒的人得知市面上的常见白酒为38度、42度、45度和52度和55度。因此考虑用38度白酒作为低度酒代表，52度酒作为高度白酒代表，啤酒的酒精度数为4度，葡萄酒的酒精度数为13度，黄酒的酒精度数一般在12度，黄酒的酿造因地区而异因此可能会对数据分析结果带来极大的误差，经过数据处理后结果如下，可以发现该地区饮食指标在酒精这一项上是良好的。



六、规律进餐，足量饮水

此条为新加入内容，意在提醒大家规律进餐、适度饮食，不暴饮暴食、不偏食挑食、不过度节食。也督促大家不要口渴时才喝水，要少量多次、足量饮水。推荐喝白水或茶水，少喝或不喝含糖饮料，不用饮料代替白水。

从图中可以看出大约18%的人每周至少有一天不吃早餐，计算不吃早餐人群的血糖浓度，

0	1	2	3	4	5	6	7
5.12	4.94	4.93	5	5.10	5.12	5.31	5.04


从表格图中平均血糖的变化，我们发现了一个有趣的结果即不吃早餐血糖反而升高了，查阅资料后发现可能是由于糖尿病因素引起的，分别计算各部分糖尿病患者占比。


0	1	2	3	4	5	6	7
3.2%	1.86%	1.50%	1.13%	2.40%	1.17%	2.86%	1.54%


据此可以得出结论绝大多数不吃早餐的人血糖升高或者与糖尿病无关，或者接受了饮食控制方面的措施，计算糖尿病患者中各部分接受饮食控制的占比有

0	1	2	3	4	5	6	7
74%	0	80%	50%	67%	67%	0%	33%

简单计算Pearson相关系数后为 **-0.1306628581951439**，具有弱相关性，因此饮食控制这个原因猜测小概率是正确的。

不吃早餐.png

不吃中餐.png

不吃晚餐.png

七、会烹会选，会看标签

此条为新加入内容，意在提醒我们传承中国传统饮食文化，并享受食物的天然美味和烹饪带来的乐趣。也强调食品标签的重要性，应学会阅读食品标签，合理选择预包装食品。

八、公筷分餐，杜绝浪费

1、选择新鲜卫生的食物，不食用野生动物。食物制备生熟分开，熟食二次加热要热透。按需备餐，提倡光盘行动。

2、新冠肺炎疫情的发生提示我们讲究卫生，坚持公筷、公勺和分餐、份餐，避免食源性疾病的发生和传播。

问题二

针对问题一提出的标准，采用对不吃早餐，酒精，体育锻炼等进行了分类，二分类标准以不吃早餐为例，分成了健康与不健康，多分类标准以新鲜水果为例，分成了不足适量与超量三个档次，针对在此基础上又针对年龄，性别等进行更小的细分，以不吃早餐-年龄表格为例，数据分析为列联表分析，检验条件如下

年龄	健康	不健康
----	----	-----

年龄	健康	不健康
20-30	76	10
30-40	607	219
40-50	2126	624
50-60	1917	432
60-70	808	102
70-80	668	42
80-90	183	7
90-100	3	0
100-110	2	1
110-120	0	0
120-130	2	0

- 1，当样本数总体超过且所有单元格T大于等于5时，用普通的卡方检验，若所得的p值约等于0.05，改用fisher检验
- 2，当样本数大于大于40但是存在单元格大于等于1小于5时，进行连续性校正
- 3，当样本数小于40或者存在单元格小于1时，采用fisher检验

卡方\连续校正卡方\fisher p值	年龄	性别	婚姻状况	文化程度	职业
不吃早餐	<0.001 (显著相关)	<0.001 (显著相关)	0.4338 (无关)	<0.001 (显著相关)	<0.001 (显著相关)
含糖饮料	<0.001(显著相关)	<0.001(显著相关)	<0.001(显著相关)	<0.001(显著相关)	<0.001(显著相关)
油炸面食（废弃）					
吸烟指数	<0.001(显著相关)	<0.001 (显著相关)	<0.001 (显著相关)	<0.001 (显著相关)	<0.001 (显著相关)
新鲜蔬菜	<0.001(显著相关)	0.394 (无关)	0.001 (相关)	0.056 (无关)	<0.001 (显著相关)
新鲜水果	<0.001(显著相关)	<0.001(显著相关)	0.018 (几乎无关)	<0.001 (显著相关)	0.002 (相关)

卡方\连续矫正卡方\fisher p值	年龄	性别	婚姻状况	文化程度	职业
体育锻炼	<0.001 (显著相关)	0.001 (相关)	<0.001 (显著相关)	<0.001 (显著相关)	<0.001 (显著相关)

表征解释：

饮食习惯：不吃早餐除与婚姻状况无统计学明显相关，与其他均有统计学上的明显相关；新鲜蔬菜除与文化程度和性别无明显相关，其他均具有明显相关；新鲜水果除与婚姻状况在统计学意义上几乎无关，与职业具有较弱的相关性，其他均具有明显相关。

生活习惯：吸烟指数与所有列变量都具有统计学上的明显相关性；体育锻炼除与性别具有相对较弱的统计学意义上的相关性，其他都是强相关性。

深度分析：由于上面的因素是对行变量进行分类得到的，产生的只是一对因子型变量，没有充分利用上每个样本各个行变量的具体数据，因此得出的结论只能在宏观上说明哪些之间具有显著的相关，而不能说明具体的哪些分组之间是相同的，哪些是不同的，现在考虑对每个具有显著相关的列行变量分组各自进行比较，查看各组之间的差异。分析如下（思维导图）

[分析思维表.pdf](#)

最终得到的结果如下

1，年龄段与各分量，分析待定

[年龄段.xlsx](#)

新鲜蔬菜：90-100；新鲜水果：80-90；体育锻炼：60-70；吸烟指数：20—30，30-40，40-50，90-100；酒精：无法判断，由于部分分组数目过少，组内酒精含量相同无标准差；含糖饮料：120-130；不吃早餐：60-70

2，性别与各分量

[性别.xlsx](#)

除体育锻炼男女之间无明显统计学差异，其他都有明显差异

3，文化程度与各分量

[文化程度.xlsx](#)

新鲜蔬菜：各组无明显差异；新鲜水果：小学；体育锻炼：高中/中专，大专/大本；吸烟指数：大本/大专；酒精：各组无明显差异；含糖饮料：文盲；不吃早餐：初中

4，婚姻与各分量

[婚姻.xlsx](#)


新鲜蔬菜：未婚；新鲜水果：丧偶；体育锻炼：未婚；吸烟指数：丧偶；酒精：未婚，已婚；含糖饮料：丧偶，未婚；不吃早餐：丧偶，未婚，已婚

5，职业与各分量


新鲜蔬菜指标上学生与其他职业具有最多的统计学差异，新鲜水果指标上是工人与农民，体育锻炼上是学生和离退休人员，吸烟指数：医务人员，学生，家庭妇女，酒精：家庭妇女，军人，学生；含糖饮料：学生，农民；不吃早餐：离退休人员。

[职业.xlsx](#)


不吃早餐

不吃早餐-年龄.png


[不吃早餐-年龄.xlsx](#)

不吃早餐-性别.png


[不吃早餐-性别.xlsx](#)

不吃早餐-婚姻状况.png

[不吃早餐-婚姻状况.xlsx](#)


不吃早餐-文化程度.png


[不吃早餐-文化程度.xlsx](#)

不吃早餐-职业.png

[不吃早餐-职业.xlsx](#)


酒精

酒精-年龄.png


酒精-性别.png

酒精-婚姻状况.png


酒精-文化程度.png

酒精-职业.png


含糖饮料

含糖饮料-年龄.png


[含糖饮料-年龄.xlsx](#)

含糖饮料-性别.png


[含糖饮料-性别.xlsx](#)

含糖饮料-婚姻状况.png

[含糖饮料-婚姻状况.xlsx](#)

含糖饮料-文化程度.png

[含糖饮料-文化程度.xlsx](#)


含糖饮料-职业. png


[含糖饮料-职业. xlsx](#)


油炸面食

油炸面食-年龄. png


油炸面食-性别. png

油炸面食-婚姻状况. png


油炸面食-文化程度. png

油炸面食-职业. png

吸烟指数

吸烟-年龄. png

[吸烟指数-年龄. xlsx](#)

吸烟-性别. png

[吸烟指数-性别. xlsx](#)

吸烟-婚姻状况. png

[吸烟指数-婚姻状况. xlsx](#)


吸烟-文化程度. png

[吸烟指数-文化程度. xlsx](#)


吸烟-职业. png

[吸烟指数-职业. xlsx](#)


新鲜蔬菜

新鲜蔬菜-年龄. png


[新鲜蔬菜-年龄. xlsx](#)

新鲜蔬菜-性别. png


[新鲜蔬菜-性别. xlsx](#)

新鲜蔬菜-婚姻状况. png

[新鲜蔬菜-婚姻状况. xlsx](#)

新鲜蔬菜-文化程度. png

[新鲜蔬菜-文化程度. xlsx](#)

新鲜蔬菜-职业. png

[新鲜蔬菜-职业.xlsx](#)


新鲜水果

新鲜水果-年龄.png


[新鲜水果-年龄.xlsx](#)

新鲜水果-性别.png


[新鲜水果-性别.xlsx](#)

新鲜水果-婚姻状况.png

[新鲜水果-婚姻状况.xlsx](#)


新鲜水果-文化程度.png

[新鲜水果-文化程度.xlsx](#)


新鲜水果-职业.png

[新鲜水果-职业.xlsx](#)


体育锻炼

体育锻炼-年龄.png


[体育锻炼-年龄.xlsx](#)

体育锻炼-性别.png


[体育锻炼-性别.xlsx](#)

体育锻炼-婚姻状况.png

[体育锻炼-婚姻状况.xlsx](#)

体育锻炼-文化程度.png

[体育锻炼-文化程度.xlsx](#)

体育锻炼-职业.png

[体育锻炼-职业.xlsx](#)

数据文件

[R.csv](#)

问题三

[R.csv](#)

多元logistic混合回归（SPSS分析）

流程图：

[未命名文件.png](#)

针对慢性病与哪些因素有关以及各自的相关水平，我们考虑采用多元混合logistics回归

尿酸水平：

首先我们做霍斯默-莱梅肖检验以判断这组数据能够用二元logistics回归模型，测得p值为0.24大于0.05说明是可以的。接着利用前向逐步回归的方法筛选出显著性较强的变量，分别是体育锻炼和吸烟指数，针对这两个显著性变量分别计算它们的OR值，其中体育锻炼OR值为1.001说明增强体育锻炼有助于维持正常尿酸，而吸烟指数的OR值为1说明相对高尿酸人群而言它的增加，不会带来倍数的明显变化，针对无关变量采取向后回归的方法，分别计算在去除变量后显著性，结果如下

p值		含糖饮料	新鲜蔬菜	酒精	新鲜水果	不吃早餐	
		.932	.758	.658	.603	.353	

糖代谢水平：

去除157个未知血糖样本，对剩余的样本我们首先做二分类只分为正常与不正常，按照流程后分析结果如下

不相关变量：新鲜蔬菜，新鲜水果，酒精，不吃早餐，对应的p值分别为0.329，0.293，0.512，0.154；

相关变量：吸烟档次，对应的OR值分别是0.302，0.258，0.383，相对1的绝对值为0.698，0.742，0.617，处于吸烟第二档次对于维持正常血糖标准最困难；含糖饮料，OR值为0.997，相对1为0.003，说明少喝含糖饮料有助于维持血糖正常水平；体育锻炼，OR值为1.001，相对1为0.001，体育锻炼基本不影响血糖正常水平。

异常糖代谢：

首先做平行线检验，p值不显著，因此不满足有序回归的条件，选择无序多分类回归，以糖尿病患者为参考组，似然比检验后结果如下

指标	新鲜蔬菜	新鲜水果	酒精	含糖饮料	体育锻炼	不吃早餐	吸烟指数	常值截距
p值	.247	.961	.814	.776	.966	.808	.002	.922

由此可见只有吸烟指数指标具有统计学意义，查询相应OR值为0.999，说明空腹血糖受损人群相对于糖尿病而言，吸烟指数每增加一个单位，发生风险（减少）幅度为0.999倍。

血脂水平：

去除149个未知血脂样本，对剩余的样本我们首先做二分类只分为正常与不正常，按照流程后分析结果如下

不相关变量：含糖饮料，新鲜蔬菜，新鲜水果，不吃早餐，对应p值分别为0.395，0.456，0.102，0.224

相关变量：酒精对应的OR值为1.135，相对1的绝对值为0.135，说明酒精每增加一个单位水平，将会增加1.135倍血脂异常风险；体育锻炼OR值为1.001，说明体育锻炼每增加一个单位基本不影响血脂异常风险；以吸烟重度为参考组，第一档次每增加一个单位想对于吸烟重度的每增加一个单位血脂异常发生档次变化倍数为0.307，第二档次则为0.358。

异常血脂：

首先做平行线检验，p值不显著，因此不满足有序回归的条件，选择无序多分类回归，以高胆固醇血症为参考组，似然比检验后结果如下

指标	新鲜蔬菜	新鲜水果	酒精	含糖饮料	体育锻炼	不吃早餐	吸烟指数指标
p值	.296	.956	.212	.133	<.001	.233	<.001

由此可见只有吸烟指数指标和体育锻炼具有统计学意义，低高密度脂蛋白血症，混合型高脂血症和高甘油三酯血症患者体育锻炼OR值为0.999，体育锻炼每增加一个单位，异常发生风险为参考组的0.999倍；低高密度脂蛋白血症，混合型高脂血症吸烟指数OR值为1.001，体育锻炼每增加一个单位，异常发生风险为参考组的1.001倍，高甘油三酯血症OR为1，相对参考值差异不大。

血压水平：

去除134个未知血压样本，对剩余的样本我们首先做二分类只分为正常与不正常，按照流程后分析结果如下

不相关变量：新鲜蔬菜，新鲜水果，酒精相应p值为.357,.230,.069；

相关变量：体育锻炼OR值为1.001；吸烟指数为1.000；含糖饮料为.996;不吃早餐.937（分析同上）

异常血压：

首先做平行线检验，p值不显著，因此不满足有序回归的条件，选择无序多分类回归，以3级高血压为参考组，似然比检验后结果如下

指标	新鲜蔬菜	新鲜水果	酒精	含糖饮料	体育锻炼	不吃早餐	吸烟指数指标
p值	.601	.573	.337	.085	.008	.740	.058

为减少误判概率，选择实接受吸烟指数指标具有统计学意义，其中单纯收缩期高血压，1级高血压（轻度），2级高血压（中度）吸烟指数指标改变引起风险的变化倍数与参考组没有差异；单纯收缩期高血压体育锻炼OR值为.995，体育锻炼每增加一个单位，异常发生风险为参考组的.995倍，1级高血压（轻度），2级高血压（中度）体育锻炼OR值为.999，体育锻炼每增加一个单位，异常发生风险为参考组的.999倍，

六、敏感度分析

七、模型的评价与推广

7.1 模型的优点

7.2 模型的缺点

7.3 模型的推广

八、参考文献

- [1] 附件3 中国居民膳食指南平衡膳食八条准则
- [2] 超重或肥胖人群体重管理专家共识及团体标准-中华健康管理学杂志 2018 年6月第 12 卷第3期 Chin J Health Manage, June 2018, Vol. 12, No. 3
- [3] [中国统计年鉴—2022 \(stats.gov.cn\)](http://stats.gov.cn)
- [4] 张嫣然, 宇传华, 胡迪等. 1990—2019年中国归因于水果摄入不足的食管癌死亡趋势[J]. 中华疾病控制杂志, 2023, 27(01):4-10. DOI:10.16462/j.cnki.zhjbkz.2023.01.002.
- [5] 罗丽莎, 张干深, 宇传华等. 1990年与2015年中国归因于水果摄入不足的疾病负担分析[J]. 中国卫生统计, 2018, 35(02):168-171+176.
- [6] 食用果糖过多增加患高血压风险[J]. 辽宁中医杂志, 2012, 39(02):330.
- [7] 2021中国膳食指南研究报告 中国营养学会
- [8] 2022中国膳食指南研究报告 中国营养学会
- [9] 中国营养学会. 中国居民膳食指南[M]. 人民卫生出版社, 2016.
- [10] 豆类食物换算表 国家食物与营养咨询委员会
- [11] 邓陶陶, 梁栋, 李湖中等. 我国市场常见饮料中糖含量调查[J]. 中国食物与营养, 2018, 24(04):5-8.
- [12] Kaur, Kulwinder & Singh, Surinder & Meena, Kailash & Saraya, Prabhnoor & Kaur, Kirandeep. (2020). Smoking index- A measure to quantify cumulative smoking exposure. Panacea Journal of Medical Sciences. 10. 71-74. 10.18231/j.pjms.2020.018.

