

# 深圳蔬菜安全风险评估与抽样检验方案设计

## 深圳蔬菜安全风险评估抽样检验方案设计

本文着眼于利用安全指数法和主成分分析法建立风险评估模型对各个蔬菜输入渠道进行风险评估，并且利用TOPSIS法进行更加全面客观的评价；然后用聚类分析的方法对不同蔬菜产地之间的蔬菜风险进行了探究；再用熵值法对安全指数法进行改进；最后根据科学的推断统计原理对深圳市抽检方案进行误差分析并且设计出了一套误差相对于原方案更小的抽检方案。

市场上主要销售的蔬菜种类分为：干燥菜、鲜菜、酱腌菜和其他菜；而蔬菜的风险主要来源于农药残留(比如克百威、甲拌磷等)。首先，我们对每种菜的农药残留的种类和检出值进行分析，根据每种农药残留检测出来的平均值以及食品药品监督管理局公布的安全值，算出各种农药残留的超标率以及在这种蔬菜中所占的比例，进而建立蔬菜风险评估模型，从而得到上述各种菜的风险大小。结果发现鲜菜和酱腌菜的风险最大，其他菜的风险最小。因此重点对鲜菜的输入渠道进行风险评估有重大意义。

然后对批发市场、农贸市场等蔬菜常规输入渠道以及网络购物等特殊输入方式分别进行分析，建立安全指数模型计算出每种农药残留在每个月每个输入渠道的安全指数，并且与主成分分析法相结合建立蔬菜输入渠道的风险评估模型，从而得出不同渠道的风险大小并且还揭示了不同农药残留之间的危害性关系。结果发现农贸市场的风险最大，生产基地的风险最小。为了检验上述模型是否客观合理，我们采用TOPSIS法对各蔬菜输入渠道风险进行评估，发现结果和前面的模型相似，这正说明了模型的合理性。由于影响蔬菜安全的因素不仅有农药残留，还有蔬菜来源地，因此，本文以各种农药残留为指标用方差分析来检验不同蔬菜产地之间的蔬菜质量是否有显著差异。以山东和广东两省为例进行方差分析，结果发现两省的蔬菜质量差距较大。这表明产地对蔬菜质量有比较大的影响。接着用聚类分析的方法探究出了不同省份的蔬菜风险大小并且还依据风险大小把这些省份分为3类。由于不同种类农药残留的溶解度和毒性各不相同，人们对蔬菜进行清洗会不同程度的减少蔬菜上的农药残留，这就使得每种农药残留的人体实际摄入量相对于清洗前减少很多，因此用熵值法来确定各种农药残留被清洗后的风险大小，使得安全指数法中的人体实际摄入量变小更加贴近实际从而改进安全指数法，使得模型更加合理。但是该模型并不能准确地描绘出经过清洗后农药残留的具体残留量。

最后针对于附件2（深圳市1-3月的农产品抽检计划），我们借助分层抽样的误差区间估计原理对1-3月的农产品抽检计划进行误差估计，发现抽样误差较大，究其原因是对于各个输入渠道的抽检数分配不合理。于是对总体方差进行区间估计并且在抽样总数不变的情况下，人为地减小极限误差，运用分层抽样的方法更加科学地对各渠道以及各个区的抽检数进行分配，进而设计出一套抽样误差更小的抽检方案。我们综合考虑人口、耕地面积等因素，再次利用主成分分析法得出不同区的分配比重，从而结果更加客观真实。

**关键词：**安全指数法，主成分分析法，聚类分析，熵值法，分层抽样

## 深圳蔬菜安全风险评估与抽样检验方案设计

This paper focuses on establishing a risk assessment model for various vegetables input channel for risk assessment by using the safety index method and principal component analysis (PCA), and then apply TOPSIS method to comprehensive objective evaluation; Then the risk of different vegetable origin was studied by clustering analysis. The safety index method is improved by entropy method. Finally, based on the scientific inference theory, the error analysis of the sampling scheme in ShenZhen is analyzed and a set of sampling whose error is smaller than the original scheme is designed.

The main types of vegetables sold in the market are: dried vegetables, fresh vegetables, pickled vegetables and other dishes; However, the risk of vegetables is mainly derived from pesticide residues (such as ke budweiser, methyl mixed phosphorus, etc.). First of all, each pesticide residue detection value is analyzed,.According to the average value for each pesticide residue detected and TVC released the food and drug safety supervision bureau,we work out all kinds of pesticide residues which and the proportion in the vegetable, and vegetable risk assessment model is established, and then obtain he size of the risk of various dishes. The results showed that the risks of fresh and pickled vegetables were the greatest, and the risk of other dishes was minimal. Therefore, it is of great significance to carry out risk assessment on the input channel of fresh food.

And then some conventional input channels such as the wholesale market, the farmer's market vegetables and network shopping special input mode are analyzed respectively, the safety index model is established to calculate each pesticide residues of each input channel safety index in each month , and vegetables input channels of the risk assessment model is established with principal component analysis method , and it is concluded that the risk of different channels size also reveals the different relations between the dangers of pesticide residues. The result shows that the risk of agricultural market is the largest and the production base is the least risk. To test the above model is objective and reasonable, we adopt TOPSIS method to evaluate risk of vegetables input channels, finding results are similar to the results of the previous model, so the rationality of the scale model is demonstrated. Because the factors affecting the safety of vegetables come not only from pesticide residues, but also from vegetable source, therefore, we use all kinds of pesticide residues as an index based on the analysis of variance to test whether there are significant differences between the different vegetable producing area on vegetable quality. Based on the analysis of variance in shandong and guangdong provinces, the results showed that the quality of vegetables in the two provinces was relatively different. This indicates that the locality has a great influence on the quality of vegetables. Then, the risk size of vegetables in different provinces was explored by clustering analysis, and these provinces were classified into three categories based on the risk. Due to the solubility and toxicity

of different pesticide residues, each is not identical. washing vegetables would reduce the amount of pesticide residues to different levels, which makes the human body take in more of each pesticide residues less than that before cleaning, so using entropy value method to determine the risk of various kinds of pesticide residues after cleaning, making the safety index method of the human body more close to the actual consumption and more practical to improve the safety index method, makes the model more reasonable. However, the model does not accurately describe the residual amount of residue after cleaning.

Finally in attachment 2 ( The agricultural product sampling plan of 1-3 month in ShenZhen), we use the error of the stratified sampling interval estimation principle to evaluate 1-3 months sampling plans for error estimation of agricultural products, we found that the sampling error is bigger, the reason is on the distribution of the number of sampling on each input channel is not reasonable. So based on the interval estimation of population variance and on the basis of the total number of sampling, we reduce the limit error artificially, using stratified sampling method which is more scientific to allocate the sampling numbers to various channels and each area, and then we design a set of sampling error smaller sampling scheme. Considering the factors such as population and arable land area, we use the principal component analysis method to obtain the allocation proportion of different regions, thus the results are more objective and real.

**Keywords:** security index method, principal component analysis, cluster analysis, entropy method, stratified sampling

# 目录

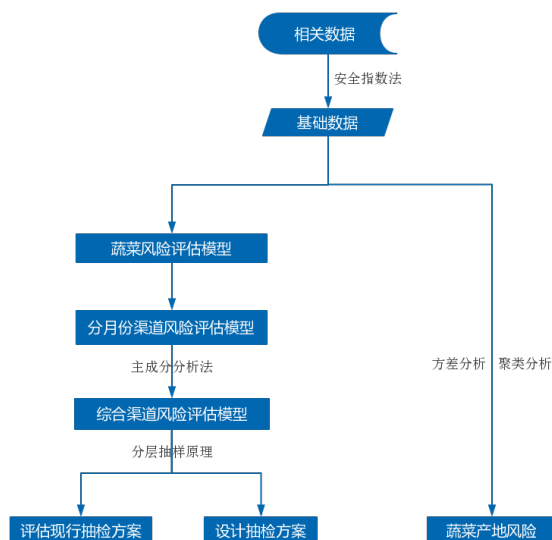
## 深圳蔬菜安全风险评估与抽样检验方案设计

1	建模思路	7
2	问题重述	7
3	问题分析	8
3.1	问题（1）的分析	8
3.2	问题（2）的分析	8
3.3	问题（3）的分析	8
3.4	问题（4）的分析	9
4	模型的假设及符号规定	9
4.1	模型的假设	9
4.2	符号说明	9
5	模型的准备	10
5.1	安全指数法	10
5.2	蔬菜风险评估模型	10
5.3	主成分分析法	11
5.3.1	定义	11
5.3.2	算法步骤	12
5.4	渠道风险评估模型	13
5.5	方差分析模型	13
5.6	聚类分析模型	14
5.6.1	离差平方和法（Ward法）	14
5.6.2	K-Means算法	15
5.7	TOPSIS法	15
5.7.1	基本原理	15
5.7.2	算法步骤	15
5.8	熵值法	16
5.8.1	基本原理	16
5.8.2	算法步骤	17
6	模型的建立与求解	18
6.1	深圳市蔬菜输入渠道的比例及推测	18
6.2	安全指数法的建立与求解	18
6.2.1	模型的建立	18

6.2.2	数据的处理以及模型的求解	18
6.3	蔬菜抽样误差模型的建立与求解	20
6.3.1	模型的建立	20
6.3.2	数据的处理	21
6.3.3	模型的求解	22
6.4	渠道风险评估模型的建立与求解	22
6.4.1	分月份渠道风险评估模型	22
6.4.2	综合渠道风险评估模型	24
6.5	第四问模型的建立与求解	27
6.5.1	数据的预处理	27
6.5.2	抽样误差的计算	28
6.5.3	计算抽样极限误差与置信区间	30
6.5.4	设计抽样方案	32
6.6	方差分析模型的建立与求解	35
6.7	聚类分析模型的建立与求解	37
6.7.1	模型的建立	37
6.7.2	数据的处理	37
6.7.3	模型的求解	37
6.8	致深圳市政府的一封信	37
7	模型的检验	38
8	模型的改进	40
8.1	熵值法	40
9	模型评价	42
9.1	优点	42
9.2	缺点	42
10	模型的总结与展望	43
10.1	总结	43
10.2	展望	43

# 1. 建模思路

## 深圳蔬菜安全风险评估与抽样检验方案设计



## 2 问题重述

食品安全是我们生活安全、消费安全、生命安全的重要前提和基础，它关乎到我们每个人的生活质量和生命健康，而蔬菜又是我们每个人日常生活中的必需品，蔬菜的消费量在整个食物支出中占有越来越高的比例，我国各种蔬菜的种植规模也在慢慢扩大。但是，由于不合理的添加物和化肥的使用，蔬菜中的不合格添加物等有害物质，不仅破坏了生态环境，更是严重威胁到了人们的生命健康。因此，对蔬菜安全风险进行有效评估以及设计合理的抽样检查方案具有重大意义。

现在蔬菜输入渠道主要包括农贸市场、批发市场、超市、配送中心等常规渠道，还包括网上购物、大单位、流动摊贩等直接输入方式。鉴于上述情况：

(1) 通过网上收集调查，得到深圳市各季节蔬菜消费种类以及不同输入渠道的比例。

(2) 建立忽略某些输入渠道的抽样而导致的食品安全风险的模型，然后基于题目所给附件，对深圳市的新鲜蔬菜风险进行评估。

(3) 根据前面所建立的模型以及所得到的结果，结合附件中深圳市蔬菜抽样检验方案，设计一套比较合理的蔬菜抽样检验方案，从而让监测工作更为有效而扎实，提高食品安全防护的效率。

(4) 在前面研究的基础上，写一份建议信给深圳市政府，谈谈你对于问题的认识 and 解决方法。

## 3 问题分析 深圳蔬菜安全风险评估与抽样检验方案设计

### 3.1 问题（1）的分析

对于问题（1），题目要求我们收集数据并且分析深圳市各季节新鲜蔬菜消费种类和各种蔬菜输入渠道的比例。由于网上相关数据较少，我们以深圳市食品药品监督管理局公布的抽检调查结果（补充附件）作为数据来源，可以间接得到深圳市新鲜蔬菜消费种类以及各种输入渠道（比如说农贸市场、批发市场）的比例。

### 3.2 问题（2）的分析

针对问题（2），题目要求我们建立模型来评估忽略某些输入渠道的抽样而导致的食品安全风险。对此，鉴于市场上消费的蔬菜种类繁多，我们按照附件中的分类方式主要把他们分为：干燥菜、鲜菜、酱腌菜、其他菜这四大类，然后建立一个比较完善的蔬菜输入渠道的安全风险评估模型。首先，上述每一种蔬菜都含有多种农药残留，并且各种被检测出来的农药残留的检出数在该种蔬菜总抽样数中的比例以及超标率不尽相同，为此，我们以每一种农药残留作为评价指标，并且以它们的被检测出来的比例作为它们各自的权重，从而建立蔬菜风险评估模型。然后，由于每一个渠道所输入各种蔬菜的比例、检出率各不相同，并且蔬菜中各种农药残留的检出率也不尽相同，各种农药残留在水中的溶解性以及自身的毒性也有较大差异，因此，为了进一步探究不同农药残留的危害性之间是否有一定的关系以及有怎样的关系，我们用安全指数法结合主成分分析法来建立渠道的风险评估模型。结果发现不同农药残留的风险之间确实有一定联系。然后，由于蔬菜质量也受来源地的影响，我们用方差分析的方法，以山东省和广东省为例对这两个省所产蔬菜中检测出的农药残留进行方差分析，发现两省之间的蔬菜质量差距较大，也进一步说明蔬菜产地也会对蔬菜质量进而对输入渠道的安全风险产生不可忽略的影响。最后，以不同省份为研究对象，不同农药残留每个月的安全指数之和为指标，用聚类分析的方法得出不同省份之间蔬菜风险的大小以及关系。结果发现广东和山东两省的蔬菜风险比较大，其他省份的蔬菜风险相对较小。

### 3.3 问题（3）的分析

针对问题（3），题目要求我们基于附件对深圳市的蔬菜安全风险进行评估。首先，对于蔬菜风险评估模型，根据我国不合格蔬菜详细信息列表（附件3），运用我们前面所构建的蔬菜风险评估模型，可以得出干燥菜，鲜菜等蔬菜种类的风险大小。结果发现鲜菜的风险高于其他三种蔬菜。然后，根据补充附件（深圳市食品药品监督管理局所公布的数据），我们以“月”为单位（由于数据有限）对深圳市的蔬菜安全风险进行评估。这里的蔬菜安全风险就是蔬菜输入渠道的安全风险。基于附件，我们可以算出每个月各个渠道、各种蔬菜的检出率，并且由前面的安全指数模型算出各种农药残留每个月在每个渠道的安全指数，进而得出各个渠道的安全风险大小。为了更加合理地检验和评价该模型，我们以每个渠道为研究对象，以每种农药残留在每个渠道中所有月份（2016.7—2017.1）的安全指数大小为指标，利用 $topsis$ 综合评价法对于各渠道风险进行再一次评估，发现结



果和渠道风险评估模型一致。在模型改进中，由于农药残留的溶解度和毒性均不同，在买回蔬菜后人们会对蔬菜进行清洗，这就使得安全指数模型中的每种农药残留值在买菜清洗后相对于买的时候减少很多，因此我们以每种农药残留为研究对象，以溶解度和毒性为指标，用熵值法来确定各个农药残留的清洗后的风险大小，从而改进安全指数法，使得模型更加合理。

## 深圳蔬菜安全风险评估与抽样检验方案设计

### 3.4 问题（4）的分析

针对问题（4），题目要求根据科学的统计推断原理来设计一套比较合理的深圳市新鲜蔬菜抽检方案。为了说明我们设计的方案是合理的，首先我们利用分层抽样的原理对深圳市现行的抽样方案进行误差估计。结果发现现行方案的抽样误差比较大。接着因为样本总体（抽样次数）有几千次之多，因此根据统计学原理可以认为抽样总体服从正态分布；然后用方差的区间估计方法对总体方差进行估计，再次利用分层抽样的原理算出抽样总数，最后以各渠道风险大小及各个区的人口密度为比例进行抽样数量的具体分配。在模型改进中，我们综合考虑人口密度、耕地面积等多方面因素利用主成分分析法算出各个区的比重从而更加合理地分配抽样数量。

## 4 模型的假设及符号规定

### 4.1 模型的假设

- (1)蔬菜的安全风险只来源于不合格添加物，不考虑其他因素比如土壤性质带来的影响；
- (2)在安全指数法中，人均体重为60kg；
- (3)忽略不同生产模式、反季节生产方式所带来的蔬菜质量差异；
- (4)检测结果准确无误；
- (5)各种不合格添加物所造成风险只与它们的超标率、检出率以及安全指数有关。

### 4.2 符号说明

- $a \sim k$  各种不合格添加物
- $A \sim I$  各种蔬菜输入渠道
- $q_{ij}$  第j种不合格添加物在第i种蔬菜中所占的比例
- $u_{ij}$  第j种不合格添加物在第i种蔬菜中检测出来的平均值
- $h_j$  第j种不合格添加物的标准值
- $If_{ij}$  第j种不合格添加物在第i种蔬菜中的超标率
- $P_{ik}$  第k种渠道第i种蔬菜所占的比例
- $W_{jk}$  第k种渠道的第j种不合格添加物的检出率

## 5 模型的准备

### 5.1 安全指数法

从相关文献<sup>[1]</sup>可以知道，由于不合格添加物的毒害作用与进入人体的量有关，所以评价蔬菜安全时以人体对蔬菜上不合格添加物的实际摄入量 and 人体日均安全摄入量指标更为合理，因此采用安全指数法来评估蔬菜中某种不合格添加物的风险，其计算公式为：

$$IFS_j = \frac{EDI_j \times Z_j}{ADI_j \times m}$$

其中， $EDI_j = R_j \times J \times K \times Q$  ( $R_j$ 为蔬菜中添加物j的残留水平，J为蔬菜的估计摄入量，K为蔬菜的可食用部分因子，Q为蔬菜的加工处理因子)。ADI<sub>j</sub>为每日允许摄入量，m为人体平均质量， $Z_j$ 为校正因子。在这个模型中，假设 $J=380g \times 人^{-1} \times d^{-1}$ ， $K=1$ ， $Q=1$ ， $m=60kg$ ， $Z_j=1$ 。不合格添加物的每日允许摄入量(ADI<sub>j</sub>)，如下表所示：

表 1: 不合格添加物的每日允许摄入量

不合格添加物	a	b	c	d	e
ADI <sub>j</sub> (单位: mg/kg)	0.001	0.004	0.030	0.003	0.010
f	h	i	l	g	k
0.070	0.005	0.003	0.030	0.002	0.0002

### 5.2 蔬菜风险评估模型

在某一种蔬菜中， $a \sim k$  (表2) 为在蔬菜检测到的各种不合格添加物， $u_{ij}$ 为某种不合格添加物在该蔬菜中所检测出来的数值的平均值， $h_j$ 为食品药品监督管理局所公布的该不合格添加物在蔬菜中的安全值，所以，某种不合格添加物的超标率为： $If_{ij} = \frac{u_{ij}-h_j}{h_j}$ ，由于每一种蔬菜中一般能够检测出多种不合格添加物，我们以每种不合格添加物的被检出数在该蔬菜被检总数中所占的比例作为它的权重，建立蔬菜风险评估模型：

$$r_i = \sum_j q_{ij} \times If_{ij}$$

表 2: 不合格添加物与对应符号

符号	a	b	c	d	e
不合格添加物	克百威	甲胺磷	乙酰甲胺磷	毒死蜱	甲拌磷
f	h	i	l	g	k
啉虫脒	氧乐果	甲基异柳磷	多菌灵	阿维菌素	氟虫脒

## 5.3 主成分分析法

### 5.3.1 定义

假定有 $n$ 个样本，每个样本共有 $p$ 个变量，构成一个 $n \times p$ 阶的数据矩阵。

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

当 $p$ 较大时，在 $p$ 维空间中考察问题比较麻烦。为了克服这一困难，就需要进行降维处理，即用较少的几个综合指标代替原来较多的变量指标，而且使这些较少的综合指标既能尽量多地反映原来较多变量指标所反映的信息，同时它们之间又是彼此独立的。

定义：记 $x_1, x_2, \dots, x_p$ 为原变量指标， $F_1, F_2, \dots, F_m (m \leq p)$ 为新变量指标。

$$\begin{cases} F_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ F_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \vdots \\ F_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \\ l_{i1}^2 + \dots + l_{ip}^2 = 1 \end{cases}$$

系数 $l_{ij}$ 的确定原则：

- (1)  $F_i$ 与 $F_j (i \neq j; i, j = 1, 2, \dots, m)$ 相互无关；
- (2)  $F_1$ 是 $x_1, x_2, \dots, x_p$ 的一切线性组合中方差最大者， $F_2$ 是与 $F_1$ 不相关的 $x_1, x_2, \dots, x_p$ 的所有线性组合中方差最大者；...； $F_m$ 是与 $F_1, F_2, \dots, F_{m-1}$ 都不相关的 $x_1, x_2, \dots, x_p$ 的所有线性组合中方差最大者。

则新变量指标 $F_1, F_2, \dots, F_m$ 分别称为原变量指标 $x_1, x_2, \dots, x_p$ 的第1,2,...,m主成分。

从以上的分析可以看出，主成分分析的实质就是确定原来变量 $x_j (j = 1, 2, \dots, p)$ 在诸成分 $F_i (i = 1, 2, \dots, m)$ 上的荷载 $l_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, p)$ 。

从数学上可以证明，它们分别是相关矩阵的 $m$ 个较大特征值所对应的特征向量。

### 5.3.2 算法步骤

设有 $n$ 个样本，每个样本观测 $p$ 个指标，将原始数据写成矩阵：

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2)$$

1.将原始数据标准化。为简单起见，不妨设上边矩阵已标准化。

2.建立变量的相关系数矩阵：

$$R = ((r_{ij})_{p \times p})$$

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

3.求 $R$ 的特征根( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ ) 及相应的单位特征向量：

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix}, \dots, a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix} \quad (3)$$

4.写出主成分：

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{pi}x_p, \quad i = 1, \dots, p$$

5.计算主成分贡献率和累计贡献率。

(1)贡献率：

$$C_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k}, (i = 1, 2, \dots, p)$$

(2)累计贡献率：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i = 1, 2, \dots, p)$$

一般取累计贡献率达85%至95%的特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 所对应的第1,2,...,m( $m \leq p$ )个主成分。

## 5.4 渠道风险评估模型

蔬菜的输入有A...I等多种渠道（见表3），输入到某种特定渠道的每种蔬菜，其自身固有的风险为 $r_i$ 。该渠道的风险除与 $r_i$ 有关外，还与每种蔬菜在该渠道中所占的比例 $P_{ik}$ 以及在检测过程中所检测出来的风险项的检出率 $W_{jk}$ 以及安全指数 $IFS_j$ 有关。综上，得出某种渠道的风险评估模型：

$$V_k = \sum_i (P_{ik} \times r_i \times \sum_j (IFS_j \times W_{jk}))$$

表 3: 渠道与对应符号

符号	A	B	C	D
进货渠道	农贸市场	批发市场	超市	配送中心
E	F	G	H	I
网络购物	流动摊贩	大单位（企业饭店）	超市	生产基地

## 5.5 方差分析模型

假设其它因素不变，只考虑一个因素A，且因素存在多个水平，在每个水平上做若干次实验，从实验结果推断是否该因素对指标有显著影响。这就是单因素的方差分析。

前提假设

设因素A的 $r$ 个水平为 $A_1, A_2, \dots, A_r$ ，每个水平下的指标服从正态分布 $N(u_1, \delta_2), N(u_2, \delta_2), \dots, N(u_r, \delta_2)$ 。因子对指标是否有影响取决于指标的正态分布是否一致，如果有影响，则正态分布应该存在差异，而正态分布由均值和方差决定，假设中方差相同，因此各个水平下的正态分布均值直接决定因素对指标是否有影响。

所以，问题可以转化为假设检验，设：

$$H_0 : u_1 = u_2 = u_3 = \dots = u_r$$

检验结果如果拒绝原假设，则认为因素A对指标有显著影响，否则认为无显著影响。

设：

$$u = (1/r) \sum_i u_i$$

$$i = u_i - u$$

则， $H_0$ 假设改写成，

$$H_0 : 1 = 2 = 3 = \dots = r = 0 \text{ 构造统计量,}$$

如果 $P_r$ 大于 $\alpha$ ，则接受 $H_0$ ，因素对指标无显著影响；

如果 $P_r$ 小于 $\alpha$ ，则拒绝 $H_0$ ，因素对指标有显著影响。

## 5.6 聚类分析模型

将一群物理对象或者抽象对象的划分成相似的对象类的过程。其中类簇是数据对象的集合，在类簇中所有的对象都彼此相似，而类簇与类簇之间的对象是彼此相异。

### 5.6.1 离差平方和法（Ward法）

该方法是Ward提出来的，所以又称Ward法。该方法的基本思想来自于方差分析，如果分类正确，同类样品的离差平方和应当较小，类与类的离差平方和较大。具体做法是先将n个样品各自成一类，然后每次缩小一类，没缩小一类，离差平方和就要增大，选择使方差增加最小的两类合并，直到所有样品归为一类为止。

设将n个样品分成k类， $G_1, G_2 \dots G_k$ ，用 $X_{it}$ 表示 $G_t$ 中的第I个样品， $n_t$ 表示 $G_t$ 中样品的个数， $\bar{X}_t$ 是 $G_t$ 的重心，则 $G_t$ 的样品离差平方和为：

$$S_t = \sum_{i=1}^{n_t} (X_{it} - \bar{X}_t)(X_{it} - \bar{X}_t)$$

如果 $G_p$ 和 $G_q$ 合并为新类 $G_r$ 类内离差平方和分别为

$$S_p = \sum_{i=1}^{n_p} (X_{ip} - \bar{X}_p)(X_{ip} - \bar{X}_p)$$

$$S_q = \sum_{i=1}^{n_q} (X_{iq} - \bar{X}_q)(X_{iq} - \bar{X}_q)$$

$$S_r = \sum_{i=1}^{n_r} (X_{ir} - \bar{X}_r)(X_{ir} - \bar{X}_r)$$

他们反映了各自类内样品的分散程度，如果 $G_p$ 和 $G_q$ 这两类相聚较近，则合并后所增加的离散平方和 $S_r - S_p - S_q$ 应较小；否则，应较大。于是定义 $G_p$ 和 $G_q$ 之间的平方距离应为：

$$D_{pq}^2 = S_r - S_p - S_q$$

其中 $G_r = G_p \cup G_q$ ，可以证明类间距离的递推公式为

$$D_{kr}^2 = \frac{n_k + n_p}{n_r + n_k} D_{kp}^2 + \frac{n_k + n_q}{n_r + n_k} D_{kq}^2 + \frac{n_k}{n_r + n_k} D_{pq}^2$$

这种系统聚类法称为离差平方和法或Ward方法。

### 5.6.2 K-Means算法

K-Means算法属于划分方法中的一种。K-Means算法的整个流程：首先从聚类对象中随机选出K个对象作为类簇的质心（当然了，初始参数的K代表聚类结果的类簇数），对剩余的每个对象，根据它们分别到这个K个质心的距离，将它们指定到最相似的簇，（因为K-Means是利用距离来量化相似度的，所以我们这里可以理解为是“将它们指定到离最近最近距离的质心所属类簇”）。然后重新计算质心位置。以上过程不断反复，直到准则函数收敛为止。通常采用平方误差准则。

K-Means的优点和不足：能处理大型数据集，结果簇相当紧凑，并且簇和簇之间明显分离。计算复杂性 $O(tkn)$ ，t:迭代次数、K：聚类数n:样本数；但是

(1)该算法必须事先给定类簇数和质点，簇数和质点的初始值设定往往会对聚类的算法影响较大。

(2)通常会在获得一个局部最优值时停止。

(3)并且只适合对数值型数据聚类。

(4)只适用于聚类结果为凸形的数据集，K-Means方法不适合发现非凸面形状类簇，或者大小差别很大的簇。

(5)对“噪音”和孤立点数据敏感，少量的该类数据对质点的计算会产生极大的影响。

## 5.7 TOPSIS法

### 5.7.1 基本原理

TOPSIS法是一种综合评价方法，它的基本思想是：

(1)将n个评价指标看成n条坐标轴，由此可以构造一个n维空间，则每个待评价的样本依照其各项指标的数据就对应n维空间中一个坐标点；

(2)针对各项指标从所有待评价样本中选出该指标的最优值（理想解即对应最优坐标点）和最差值（负理想解即对应最差坐标点），依次求出各个待评价样本的坐标点分别到最优坐标点和最差坐标点的距离 $d^*$ 和 $d^0$ ；

(3)构造评价参考值：

$$f = \frac{d^0}{d^0 + d^*}$$

则f值越大,代表评价结果越优越。

### 5.7.2 算法步骤

(1)构造决策矩阵，其中每一列是一个评价指标，每一行是一条待评价样本；为去掉量纲效应，做规范化处理得到 $B = (b_{ij})_{m \times n}$ ,其中：

$$b_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^m (a_{ij})^2}}$$

注：该规范化处理后，各待评价样本的同一评价指标评分值的平方和为1,这适合TOPSIS法中计算欧氏距离的场合。

(2)根据每个评价指标对评价结果的贡献程度的不同，确定相应的权重 $w = [w_1, \dots, w_n]$ .将 $B$ 的第 $j$ 列乘以其权重 $w_j$ , 得到加权规范化矩阵 $C = (c_{ij})_{m \times n}$ 。

(3)确定正理想解 $C^*$ 和负理想解 $C^0$ ：

$$C^* = [c_1^*, \dots, c_n^*]$$

$$C^0 = [c_1^0, \dots, c_n^0]$$

其中：

$$c_j^* = \begin{cases} \max_i c_{ij} & \text{若第}j\text{评价指标是正向, } j=1,2,\dots,n \\ \min_i c_{ij} & \text{若第}j\text{评价指标是负向标, } j=1,2,\dots,n \end{cases}$$

$$c_j^0 = \begin{cases} \min_i c_{ij} & \text{若第}j\text{评价指标是正向标, } j=1,2,\dots,n \\ \max_i c_{ij} & \text{若第}j\text{评价指标是负向标, } j=1,2,\dots,n \end{cases}$$

(4)计算每个待评价样本到正理想解和负理想解的距离 $d_i^*$ 和 $d_i^0$ ：

$$d_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, \quad i = 1, 2, \dots, m$$

$$d_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, \quad i = 1, 2, \dots, m$$

(5)计算每个待评价样本的评价参考值：

$$f_i = \frac{d_i^0}{d_i^0 + d_i^*}, \quad i = 1, 2, \dots, m$$

再将 $f_i$ 从大到小排列，得到各待评价样本的优劣顺序。

## 5.8 熵值法

### 5.8.1 基本原理

在信息论中，熵是对于不确定性的一种度量。信息量越大，不确定性就越小，熵也就越小；反之如果信息量越小，不确定性越大，熵也越大。根据熵的特性，可以通过计算熵值来判断一个事件



的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，指标的离散程度越大，该指标对综合评价的影响（权重）越大，它的熵值越小。

### 5.8.2 算法步骤

(1) 选取 $n$ 个对象与 $m$ 个指标，则 $x_{ij}$ 为第 $i$ 个对象的第 $j$ 个指标的评分( $i=1,2,\dots,n$ ;  $j=1,2,\dots,m$ );

(2) 指标的归一化处理：异质指标同质化由于各项指标的计量单位并不统一，因此在使用它们计算综合指标前，先要对它们进行归一化处理，即把指标的绝对值转化为其相对值，并令 $x_{ij} = |x_{ij}|$ ，从而解决各项不同质指标值的同质化问题。而且，由于正向指标和负向指标数值代表的含义不同(正向指标数值越大越好，负向指标数值越小越好)，因此，对于高低指标我们用不同的算法进行数据标准化处理。其具体方法如下：

正向指标：

$$x_{ij}^* = \frac{x_{ij} - \min\{x_{1j}, \dots, x_{nj}\}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}}$$

负项指标：

$$x_{ij}^* = \frac{\max\{x_{1j}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}}$$

则 $x_{ij}^*$ 为第 $i$ 个对象的第 $j$ 个指标的评分( $i=1,2,\dots,n$ ;  $j=1,2,\dots,m$ )。为了方便起见，归一化后的数据仍记为 $x_{ij}$ 。

(3) 计算第 $j$ 项指标中第 $i$ 个对象占该指标的比重：

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m$$

(4) 计算第 $j$ 项指标的熵值：

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), \quad j = 1, 2, \dots, m$$

其中， $k = \frac{1}{\ln(n)} > 0$ ,  $0 \leq e_j \leq 1$ 。

(5) 计算信息熵冗余度：

$$d_j = 1 - e_j, \quad j = 1, 2, \dots, m$$

(6) 计算各项指标的权值：

$$u_j^* = \frac{d_j}{\sum_{j=1}^m d_j}, \quad j = 1, 2, \dots, m$$

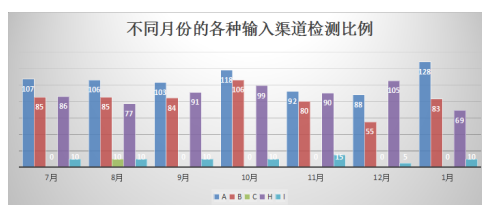
(7) 计算各对象的综合得分：

$$s_i = \sum_{j=1}^m u_j^* \times p_{ij}, \quad i = 1, 2, \dots, m$$

## 6 模型的建立与求解

### 6.1 深圳市蔬菜输入渠道的比例及推测

在得出各渠道每个月每种不合格添加物的安全指数后，基于补充附件（深圳市食品药品监督管理局公布的抽样检测结果），可以得出每个月各个蔬菜输入渠道的比例。如下图所示：



从上图可以看出，农贸市场、批发市场、小型超市这三个渠道在每个月被检测的比例远远高于其他渠道，这三种渠道比例能够占总渠道数的90%以上，因此我们推测这三个渠道的风险比较大；相比之下I渠道的比例很小，说明该渠道的风险明显低于其他渠道。为了验证这个猜想，我们建立渠道风险评估模型来评价每个渠道风险。

### 6.2 安全指数法的建立与求解

#### 6.2.1 模型的建立

评价蔬菜安全以人体对蔬菜不合格添加物的实际摄入量 and 人体日均安全摄入量指标更为合理，因此用安全指数法来评价蔬菜中某种不合格添加物的风险，其计算公式为：

$$IFS_j = \frac{EDI_j \times Z_j}{ADI_j \times m}$$

其中， $EDI_j = R_j \times J \times K \times Q$  ( $R_j$ 为蔬菜中添加物j的残留水平，这里取为每个渠道每个月该种不合格添加物检测出的最大值，该值可以从补充附件中查出；J为蔬菜的估计摄入量，K为蔬菜的可食用部分因子；Q为蔬菜的加工处理因子)。ADI<sub>j</sub>为每日允许摄入量(见表2)；m为人体平均质量；Z<sub>j</sub>为校正因子。在这个模型中，假设：J=380g×人<sup>-1</sup>×d<sup>-1</sup>，K=1，Q=1，m=60kg，Z<sub>j</sub>=1。

#### 6.2.2 数据的处理以及模型的求解

根据补充附件，找出每个渠道每个月该种不合格添加物检测出的最大值。结合表2经过计算可

以得到各渠道每个月各自不合格添加物的安全指数如下表所示:

农贸市场:

月份 \ 添加物	a	b	c	d	e	f	h	i	l	g	k
7月	8.23	0.22	1.07	6.65	0.08	0.19	0	0	0	0	0
8月	3.04	0	0	0	0	0	1.27	1.10	0	0	0
9月	2.22	0	0	9.14	0	0	0	0	0	0	0
10月	0	0	0	3.36	0	0.13	2.74	0	0	0	8.23
11月	0.89	0	0	0.38	0	0.16	0.22	0	0	0	3.80
12月	8.17	0	0	1.50	0	0	0	0	0	0	2.09
1月	1.77	0	0	0	0	0	0	0.04	0	0	
2月	0	0	0	0	0	0	0	0	0	0	0

批发市场:

月份 \ 添加物	a	b	c	d	e	f	h	i	l	g	k
7月	0	0	0	6.65	0	0	0	0	0	0	0
8月	0.82	0	0	0	0	0	0.38	0	0	0.38	0
9月	0	0	0	0	0	0	0	0	0	0	0
10月	0	0	0	0	0	0	2.98	0	0.02	0	0
11月	0	0	0	0	0	0.24	0.47	0	0	0	0
12月	0	0	0	0	0	0	0.63	0	0	0	6.65
1月	0	0	0	6.19	0	0	0.19	0	0	0	0
2月	0	0	0	0	0	0	0	0	0	0	0

小型超市:

月份 \ 添加物	a	b	c	d	e	f	h	i	l	g	k
7月	0	0	0	0	0	0	0	0	0	0	0
8月	0	0	0	0	0	0	0	0	0.02	0	0
9月	0	0	0	10.24	0	0.22	0.09	0	0	0	35.47
10月	1.33	0	0	3.36	0	0.23	0	0	0.11	0	0
11月	1.27	0	0	0	0	0	0	0.11	0	0	0
12月	0.63	0	0	4.39	0	0	0.2	0	0	0	10.13
1月	0	0	0	1.44	0	0.15	0.14	0	0	0	0
2月	0	0	0	0	0	0	0	0	0	0	0

当 $EDI_j < 1$ 时，说明该不合格添加物对人体的伤害不大，属于可以接受甚至忽略的范围内。当 $1 < EDI_j < 2$ 时，该不合格添加物对人体有一定的不可忽略的危害，当 $EDI_j > 2$ 时，说明该不合格添加物危害特别大。

从上表可知：在农贸市场渠道中，克百威、毒死蜱、氧乐果、氟虫腈的风险很高，甲胺磷的风险偏低，通过计算每个月所有不合格添加物安全指数 $EDI_j$ 的平均值，发现7、9、10、12月的风险偏高，其中7月最高。同理我们可以得到在批发市场中：毒死蜱、氧乐果、氟虫腈的风险很高，甲胺磷、乙酰甲胺磷、甲拌磷、甲基异柳磷的风险为0，其他不合格添加物的风险较低；7、10、12月的风险偏高，其中7月最高。在小型超市中：克百威、毒死蜱、氟虫腈的风险很高，甲胺磷、乙酰甲胺磷、甲拌磷、阿维菌素的风险为0，其他不合格添加物的风险较低；9、10、12月的风险偏高，其中9月最高。

综上，可以验证克百威、毒死蜱、氧乐果风险偏高，甲胺磷风险低的结论；另一方面，从月份来看，7月、9月的风险较高，这可能是由于蔬菜在这两个月大量上市，这也启示在设计抽样检测方案时要在7月和9月重点检查。

## 6.3 蔬菜风险评估模型的建立与求解

### 6.3.1 模型的建立

根据上面的蔬菜风险评估模型，在某一种蔬菜中， $a \sim k$ （表2）为在蔬菜中所检测到的各种不合格添加物， $u_{ij}$ 为某种不合格添加物在该蔬菜中所检测出的数值的平均值， $h_j$ 为食品药品监督管理局所公布的不合格添加物在蔬菜中的安全值，则某种不合格添加物的超标率为： $If_{ij} = \frac{u_{ij} - h_j}{h_j}$ ，由于每一种蔬菜中可能会检测出多种不合格添加物，故以每种不合格添加物被检出数在该蔬菜总检测数中所占的比例 $q_{ij}$ 为权重，建立蔬菜风险评估模型：

$$r_i = \sum_j q_{ij} \times If_{ij}$$

### 6.3.2 数据的处理

基于附件三(表：不合格蔬菜及其复制品)，由于附件三中的蔬菜大致分成干燥菜、鲜菜、酱腌菜、其他菜，且蔬菜的被检测出的不合格添加物类别、检出值、安全值及产地等信息均较详细，我们可以求出在每种蔬菜中各种不合格添加物的比例及超标率，结果(以鲜菜为例)如表4所示：

表 4: 鲜菜中主要不合格添加物情况

添加物	代森锰锌	啉虫脒	毒死蜱	二氧化硫	腐霉利
相关数据					
安全值	1.00	1.00	0.05	0.20	0.20
比例	0.017	0.032	0.055	0.077	0.023
平均值	5.80	4.05	0.76	4.59	3.15
超标率	4.80	3.05	14.13	21.93	14.72
添加物	氧乐果	甲拌磷	克百威	铅	山梨酸
相关数据					
安全值	0.02	0.01	0.02	0.10	0.50
比例	0.119	0.013	0.084	0.026	0.096
平均值	0.57	1.74	0.12	0.35	0.76
超标率	27.70	172.50	5.09	2.54	0.52

从表中可以看出，氧乐果、克百威、二氧化硫残留量在鲜菜中所占的比例和超标率均明显比其他不合格添加物大，说明这几种添加物是比较常见的添加物，使用较频繁，也是蔬菜风险的主要来源（与之前安全指数法的结论相符）；而山梨酸和代森锰锌所占的比例和超标率均较小，说明这些不合格添加物不是蔬菜风险的主要影响因素，相对于其它不合格添加物的影响偏小。

### 6.3.3 模型的求解

在求出了每种不合格添加物的比例和超标率之后，以每种不合格添加物被检测数在该蔬菜被检测总数中所占的比例作为它的权重，运用蔬菜风险评估模型，可以得出每种蔬菜的风险指数如下表所示：

表 5: 每种蔬菜的风险指数

	鲜菜	干燥菜	酱腌菜	其他菜
风险	38.5939	14.3624	35.9788	1.2462

由上表可知，鲜菜和酱腌菜的风险明显高于干燥菜和其他菜，并且它们的风险大小相近：这可能与这两种菜需求量大而导致种植规模大、种植种类多、监管难以到位有关；干燥菜的风险适中，从数值上说只有鲜菜和酱腌菜的风险的一半不到；其他菜的风险最低并且远远低于前三种蔬菜：这可能与其它菜的抽样量过小或者是种植规模小有关。

## 6.4 渠道风险评估模型的建立与求解

### 6.4.1 分月份渠道风险评估模型

#### 模型的建立

蔬菜的输入有A...I等多种渠道（见表3），输入到某种特定渠道的每种蔬菜，其自身固有的风险为 $r_i$ 。该渠道的风险除与 $r_i$ 有关外，还与每种蔬菜在该渠道中所占的比例 $P_{ik}$ 以及在检测过程中所检

测出来的风险项的检出率 $W_{jk}$ 以及安全指数 $IFS_j$ 有关。综上，得出某种渠道的风险评估模型：

$$V_k = \sum_i (P_{ik} \times r_i \times \sum_j (IFS_j \times W_{jk}))$$

### 模型的求解

在题目中，由于只需要考虑新鲜蔬菜这一种蔬菜，所以我们把该蔬菜在某一渠道中所占的比例设为1。由上面可知，新鲜蔬菜的风险 $r = 38.5939$ 。在每种渠道中，每种不合格添加物的检出率如下表所示（以2017年1月为例）：

表 6: 不合格添加物检出率

	a	i	k	d	h	f
A	0.016	0.008	0	0	0	0
B	0	0	0.012	0.024	0.012	0
H	0	0	0.015	0.029	0.015	0.015

结合前面的每种不合格添加物的安全指数，可以得出，2016年7月-2017年1月各月各渠道的风险大小，如下表所示：

表 7: 每个月各渠道的风险大小

	7月	8月	9月	10月	11月	12月	1月
A	1.31	0.72	1.52	3.24	1.85	1.95	0.35
B	0.89	0.21	0.90	2.19	0.16	1.11	1.06
C	0	0.30	0	0	0	0	0
H	0	0.0025	6.17	0.58	0.60	2.41	0.75
I	0	0	0	0	0	0	0

基于上表，以“月”为单位，对各渠道的风险进行比较，如下图所示：

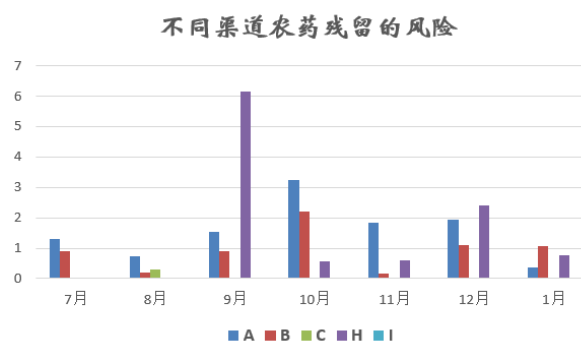


图 1

由上图可以得出，在2016年7月、10月、11月、12月，农贸市场的风险明显高于其他渠道，这说明农贸市场一直是鲜菜的主要供应渠道，它的数量多以及输入蔬菜的比例大导致了该渠道监管比较困难、风险较大。2016年9月小型超市的比例很好，远远高于其他渠道，这说明在九月份可能由超市输入的鲜菜相对于其它渠道多很多。从月份上来看，9月至12月的风险明显高于其它月份，这与蔬菜一般大量在夏秋季节上市有关。

#### 6.4.2 综合渠道风险评估模型

##### 模型的建立

由于上一模型只能分月份评价每个渠道的风险大小，不能给出每个渠道总的风险大小，又考虑到每个渠道的风险受克百威等十数种不合格添加物的影响，而每种不合格添加物对该渠道影响的大小又不尽相同，且不合格添加物之间不仅存在风险上的差异，而且也可能存在某种相关性，因此我们进一步用主成分分析法对不合格添加物种类进行降维处理。通过找出主成分并计算每个渠道的综合得分，建立综合渠道风险评估模型：

$$V'_k = \sum_i C_i F_{ik}$$

其中， $V'_k$ 为第K种渠道的风险； $C_i$ 为第i个主成分的贡献率； $F_{ik}$ 为第K种渠道第i个主成分的得分。

同时，为了衡量每种不合格添加物对人体的危害大小，我们将仍然采用所有不合格添加物，在2016年07月-2017年01月每种输入渠道中安全指数之和，来衡量不合格添加物的风险，具体如下问矩阵。

##### 模型的求解

具体做法如下：

(1)将原始数据（所有不合格添加物在每种输入渠道中的安全指数之和）写成矩阵：

$$\begin{bmatrix} 29.0194 & 20.2217 & \cdots & 26.1566 \\ 0.8233 & 0 & \cdots & 21.85 \\ \vdots & \vdots & \ddots & \vdots \\ 3.6353 & 0 & \cdots & 45.5987 \end{bmatrix} \quad (4)$$

由  $Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ ,  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, p$  将原始数据标准化。

其中， $\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$ ,  $s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}$

得到



$$Z = \begin{bmatrix} 2.1302 & 2 & 2 & 1.5342 & 2 & 2.0539 & 1.7752 & 2.0582 & 0.813 & 0 & 1.3979 \\ 0.0604 & 0 & 0 & 1.9356 & 0 & 0.8247 & 2.0518 & 0 & 0.3872 & 2 & 1.1677 \\ 0.0883 & 0 & 0 & 0 & 0 & 0 & 0.4425 & 0 & 0 & 0 & 0 \\ 0.2668 & 0 & 0 & 2.2598 & 0 & 2.0325 & 0 & 0.2019 & 2.2866 & 0 & 2.437 \end{bmatrix} \quad (5)$$

(2)得出各指标（不合格添加物的安全指数）之间的相关系数矩阵：

$$R = \begin{bmatrix} 1.0000 & 0.9958 & \cdots & 0.1716 \\ 0.9958 & 1.0000 & \cdots & 0.0982 \\ \vdots & \vdots & \ddots & \vdots \\ 0.1716 & 0.0982 & \cdots & 1.0000 \end{bmatrix} \quad (6)$$

从上面可以看出，有些不合格添加物比如说克百威和甲胺磷，克百威和乙酰甲胺磷之间的相关性比较大，高达99.58%；有些不合格添加物比如说甲胺磷和阿维菌素之间的相关性较小甚至负相关，这说明不合格添加物之间的相关性不尽相同，这是我们以后将继续研究的课题，以进一步探索不同不合格添加物的各种组合对于蔬菜风险有什么影响。

(3)求出相关系数矩阵R的特征根为：

$$\lambda = \begin{bmatrix} 5.8158 \\ 3.3054 \\ 1.8788 \\ \vdots \\ 0 \end{bmatrix} \quad (7)$$

即相应的单位特征向量：

$$a_1 = \begin{bmatrix} 0.4064 \\ 0.3996 \\ 0.3996 \\ 0.1256 \\ 0.3996 \\ 0.3126 \\ 0.1607 \\ 0.4073 \\ 0.0908 \\ -0.1418 \\ 0.1494 \end{bmatrix} \quad a_2 = \begin{bmatrix} -0.1029 \\ -0.1466 \\ -0.1466 \\ 0.4379 \\ -0.1466 \\ 0.3609 \\ -0.2412 \\ -0.0986 \\ 0.5255 \\ -0.0360 \\ 0.5075 \end{bmatrix} \quad a_3 = \begin{bmatrix} -0.0474 \\ -0.0165 \\ -0.0165 \\ 0.3822 \\ -0.0165 \\ 0.0249 \\ 0.5916 \\ -0.0408 \\ -0.1445 \\ 0.6839 \\ 0.0999 \end{bmatrix} \quad (8)$$

从上面分析可以得出前三个主成分为：

$$F_1 = 0.4064x_1 + 0.3996x_2 + 0.3996x_3 + 0.1256x_4 + 0.3996x_5 + 0.3126x_6 + 0.1607x_7 + 0.4073x_8 + 0.0908x_9 - 0.1418x_{10} + 0.1494x_{11}$$

$$F_2 = -0.1029x_1 - 0.1466x_2 - 0.1466x_3 + 0.4379x_4 - 0.1466x_5 + 0.3609x_6 - 0.2412x_7 - 0.0986x_8 + 0.5255x_9 - 0.0360x_{10} + 0.5075x_{11}$$

$$F_3 = -0.0474x_1 - 0.0165x_2 - 0.0165x_3 + 0.3822x_4 - 0.0165x_5 + 0.0249x_6 + 0.5916x_7 - 0.0408x_8 - 0.1445x_9 + 0.6839x_{10} + 0.0999x_{11}$$

(4)计算得出各个主成分的贡献率为：

第一个主成分的贡献率为52.8710%；第二个主成分的贡献率为30.0488%；第三个主成分的贡献率为17.0802%；前三个主成分的累计贡献率为100%。

如果按照90%以上的信息，选取主成分，则选取前三个主成分。第一主成分 $F_1$ 包含的信息量最大为52.8710%，它的主要代表变量为克百威、甲胺磷、乙酰甲胺磷、甲拌磷、甲基异柳磷；其权重系数分别是0.4064、0.3996、0.3996、0.3996、0.4073。第二主成分 $F_2$ 包含的信息量次之为30.0488%，它的主要代表变量为毒死蜱、啶虫脒、多菌灵、氟虫腈，其权重系数分别是0.4379、0.3609、0.5255、0.5075。第三主成分 $F_3$ 包含的信息量为17.0802%，它的主要代表变量为阿维菌素、氧乐果，其权重系数分别是0.5916、0.6839。这些代表变量反映了各自对新因子作用的大小，这也说明克百威、甲胺磷、乙酰甲胺磷、甲拌磷、甲基异柳磷之间存在着比较明显的关联，同理毒死蜱、啶虫脒、多菌灵、氟虫腈以及阿维菌素、氧乐果这两组变量组内也存在较强的相关性。

(5)根据前三个主成分得分及对应贡献率，代入 $V'_k = \sum_i C_i F_{ik}$ ，得到4个渠道的得分如下：

农贸市场：17.0999； 批发市场：11.7872； 大型超市：0.3309； 小型超市：18.6903。

因为在该模型中，我们所用的指标为安全指数，也就是说数值越大风险越大。所以农贸市场、批发市场、小型超市的风险远大于大型超市。

## 6.5 第四问模型的建立与求解

总体来讲，由于深圳市各个区的具体情况各不相同，比如说各个蔬菜的输入渠道比例和数量以及人口都可能有较大差异，因此设计抽样方案是比较复杂的。为此我们先基于附件二，根据科学的推断统计原理来得出深圳市现行抽样方案的抽样误差。

由于附件二中只给出了1,2,3月的抽检计划，因此我们分别对这3个月的抽样方案进行误差估计。采用分层抽样的方法进行误差估计，由于在分层抽样中，总体差异程度越大，抽样平均误差也大；在其他条件相同的情况下，样本容量越大，抽样平均误差越小，反之则越大，各层之间的差异的越大，层内的差异越小抽样越合理。基于上述理论，由于深圳市各区之间的蔬菜风险是否有较大差异未知，但是通过第三问的求解，我们发现不同蔬菜输入渠道之间的风险差异较大，因此我们以蔬菜输入渠道为根据进行分层，进行抽样误差的估计。

### 6.5.1 数据的预处理

基于补充附件统计出每个月每个渠道被检测出来的不合格添加物次数及安全指数，基于附件二可以得出每个月与各个输入渠道的抽样数量，如下表所示：

表 8

1月 渠道	相关数据	安全指数					抽检数量
		0.04	1.77	0.31	0.53	0	
A		0.04	1.77	0.31	0.53	0	155
B		6.19	6.65	0	6.19	0.19	337
H		0.14	10.13	3.93	1.44	0.15	130

2月 渠道	相关数据	安全指数						抽检数量
		1.08	0.91	1.65	1.54	4.43	23.12	
A		1.08	0.91	1.65	1.54	4.43	23.12	3.80
B		5.81	1.33	0.33	0	0	0	
H		2.36	6.84	0.14	0.13	18.05	6.97	2.79

2月 渠道	相关数据	安全指数					抽检数量
		0.89	0.53	4.20	0.21	1.14	
A		0.89	0.53	4.20	0.21	1.14	119
B		0	0	0	0	0	243
H		0.02	0	0	0	0	100

3月 渠道	相关数据	安全指数				抽检数量
A		0.08	0.44	2.64	0.41	116
B		0	0	0	0	248
H		3.27	1.12	0.14	0.13	87

从表中我们可以看出，从抽样数量上看，批发市场渠道的抽样数量明显高于其他渠道达到两倍之多；从被检测的添加物次数来看，农贸市场,小型超市的次数有明显多于批发市场，这就意味着这种抽样方法可能存在比较大的问题。

### 6.5.2 抽样误差的计算

#### 一、分层抽样的原理

对于单层样本的抽样来说，简单随机抽样即是对总体单位不作任何分类或排队，完全按随机原则逐个地抽取样本单位的方式进行不重复抽样。抽样平均误差的计算公式如下：

(1)平均数的抽样平均误差

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)}$$

(2)成数的抽样平均误差

$$\mu_p = \sqrt{\frac{P(1-p)}{n} \left( \frac{N-n}{N-1} \right)}$$

当样本总数N较大时，上式可以写为：

$$\mu_p = \sqrt{\frac{P(1-p)}{n} \left( 1 - \frac{n}{N} \right)}$$

当总体为多层时，平均数抽样平均误差为：

$$\mu_{\bar{x}} = \sqrt{\frac{\bar{\sigma}_i^2}{n} \left( 1 - \frac{n}{N} \right)}$$

其中：

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i}$$

成数的抽样平均误差为：

$$\mu_p = \sqrt{\frac{\bar{P}_i(1-P_i)}{n} \left( 1 - \frac{n}{N} \right)}$$

其中：

$$\bar{P}_i(1-P_i) = \frac{\sum_{i=1}^k P_i(1-P_i)n_i}{\sum n_i}$$

## 二、分层抽样误差的求解

基于上述理论，这里的成数就是每个月每种输入渠道的蔬菜合格率，其计算公式为(抽样总数-不合格样品数)/抽样总数， $x_i$ 就是每个月每种输入渠道中所检测出来的不合格添加物的安全指数， $U_{ij}$ 为每个月每种渠道被检测出的不合格添加物的次数( $1 \leq i \leq 3, 1 \leq j \leq 4$ )由此可以得出：

1月份：

农贸市场(第一层)平均值：

$$x_1 = \frac{x_1 + x_2 + x_3 + x_4}{n_1} = \frac{0.0443 + 1.773 + 0.3103 + 0.5257}{155} = 0.0171$$

方差：

$$\sigma_1^2 = \frac{\sum((x_i - \bar{x})^2)}{n} = 3.4738$$

成数：

$$P_1 = \frac{n_1 - U_1}{n_1} = 0.9724$$

类似地，其他渠道各月的平均值、方差、成数分别如下表所示：

1月 渠道	相关数据	抽检数量	平均数	方差	成数
A		155	0.0171	3.4738	0.9742
B		337	0.057	121.8637	0.9881
H		130	0.1204	122.0188	0.9615

2月 渠道	相关数据	抽检数量	平均数	方差	成数
A		119	0.3655	579.7038	0.8992
B		243	0.0307	35.3596	0.9877
H		100	0.372968	420.60527	0.92

3月 渠道	相关数据	抽检数量	平均数	方差	成数
A		116	0.0308	7.2489	0.9655
B		248	0	0	1
H		87	0.0536	11.7458	0.954

从上表可以看出：A、H渠道的安全指数平均值相对于B渠道较高，并且方差也远远高于B渠道，为B渠道的20倍之多，而A、H渠道蔬菜的合格率(即成数)又略微低于B渠道，这说明A、H渠道应该是重点检查和监控的对象，应该要把更多的次数分配给A、H渠道。

在附件1中，由于每月抽检2600批次，文件中规定种植业的抽检比例至少50%，而由附件2深圳市1-3月抽检计划可知种植业类被检对象又包括蔬菜、水果和食用菌，所以我们根据附件2，统计出

每个月蔬菜和其他种植业产品(水果和食用菌)的抽检总数以及蔬菜的抽检数量在总抽检次数中的比例  $Q_i = \frac{n_i}{N}$ ，如下表所示：

月份 抽样数	1月	2月	3月
蔬菜抽样数量	559	465	624
其他抽样数量	537	374	526
蔬菜在总次数中的比例 $Q_i$	0.26646	0.2771	0.2713

从上表可以看出，每个月蔬菜的抽样次数略微高于水果和食用菌的总次数，其中2月和3月蔬菜在总抽检次数中的比例高于1月的。这提示我们在进行抽样方案设计分配抽样次数时要对蔬菜进行重点分配和检查。

接着开始计算平均数抽样平均误差和成数抽样平均误差，由公式可以得出每个月平均数和成数的抽样平均误差如下表所示：

月份 误差	1月	2月	3月
平均数平均误差	0.3327	0.6366	0.0819
成数平均误差	0.0204	0.0457	0.0052

从上面可以看出：3月的平均数抽样平均误差最小，仅为0.0819,2月的抽样平均误差最大，达到了0.6366；3月的成数平均误差最小，仅为0.0052,2月的最大，达到了0.0457.这说明不同月份之间的平均数平均误差和成数平均误差差距较大，而且2月的误差明显高于其他两月，这也说明了该抽样方案的不合理

### 6.5.3 计算抽样极限误差与置信区间

#### 一、极限误差与置信区间的定义

抽样极限误差即为指在进行抽样估计时，样本指标与总体指标之间可允许的最大误差范围。

抽样平均数极限误差：

$$\Delta_{\bar{x}} \geq |\bar{x} - \bar{X}|$$

抽样成数极限误差：

$$\Delta_p \geq |p - P|$$

$$p - \Delta_p \leq P \leq p + \Delta_p$$

基于理论上的要求，抽样极限误差需要用抽样平均误差为标准单位来衡量，抽样极限误差也可

以表示为抽样平均误差的若干倍，于是有计算公式：

$$\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}}$$

$$\Delta_p = t \cdot \mu_p$$

把极限误差相应除以 $\Delta_x$ 或 $\Delta_p$ ，得出相对的误差程度 $t$ 倍， $t$ 称为抽样误差的概率度。即：

$$t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}}$$

$$t = \frac{\Delta_p}{\mu_p}$$

一般地，在抽样平均误差为一定的条件下，当概率度 $t$ 的值越大，则抽样误差范围 $\Delta$ 越大，估计总体平均数或成数包含在相应的区间范围内的概率越大，从而抽样估计的可靠程度也就越高。反之亦然。

如果把可靠程度即概率用 $p$ 表示，概率度 $t$ 的大小决定 $p$ 的大小，即 $p$ 是 $t$ 的函数：

$$P = F(t)$$

(1)置信度 $F(t)$ ：指总体指标落在某个区间的概率保证程度。

为了计算的方便，在实际工作中，按不同的 $t$ 值和相应的概率编成正态分布概率表供查用。常用的有：

表 9: 概率度和置信度的函数关系

$t$	$F(t)$
1	0.6827
1.64	0.9000
1.96	0.9500
2	0.9545
3	0.9973

(2)置信区间：以一定的概率保证程度确定总体指标所在的区间。(分为置信下限和置信上限)

$$t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}} \Rightarrow t = \frac{|\bar{x} - \bar{X}|}{\mu_{\bar{x}}}$$

置信区间为：

$$\bar{X} - \Delta_{\bar{x}} \leq \bar{x} \leq \bar{X} + \Delta_{\bar{x}} \Rightarrow \bar{x} - \Delta_{\bar{x}} \leq \bar{X} \leq \bar{x} + \Delta_{\bar{x}}$$

同理：

$$p - \Delta_p \leq P \leq p + \Delta_p$$

## 二、极限误差的计算

基于上述理论，我们以95%的置信度来对总体指标进行区间估计，计得出安全指数和合格率(即成数)的极限误差。由于置信度为95%，所以由表（10）可得 $t=1.96$ 。然后由公式 $t = \frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}}$ 可以得出每个月的安全指数和合格率的极限误差如下表所示：

表 10: 安全指数和合格率的极限误差

月份 数据	1月	2月	3月
安全指数	0.6521	1.2477	0.1605
合格率	0.04	0.0896	0.0102

从上表我们可以看出，由于1-3月每个月的安全指数的平均值分别为0.06、0.1886、0.0171，而表中的安全指数极限误差明显高于平均值，这就使得置信区间左端出现负数(安全指数均为正数)，这就说明该抽样方法误差很大。而每个月的合格率分别为0.9791、0.9502、0.9823，如果按照95%的置信度去估计总体合格率，置信区间右端会出现大于1的数(合格率介于0-1之间)，如果要避免上述两种情况的发现，我们只能降低置信度，进而减小 $t$ 的值，这也充分说明该抽样方法的不准确。由公式 $\bar{x} - \Delta_{\bar{x}} \leq \bar{X} \leq \bar{x} + \Delta_{\bar{x}}$ 和 $p - \Delta_p \leq P \leq p + \Delta_p$ 得出每个月的安全指数和合格率的置信区间为：(去掉安全指数的负值和数值大于1的合格率)

表 11: 安全指数和合格率的置信区间

月份 数据	1月	2月	3月
安全指数	[0,0.7121]	[0,1.4363]	[0, 0.1776]
合格率	[0.9391,1]	[0.8606,1]	[0.9721,0.9925]

从上表可以看出，依照现行的抽样方法，有95%的把握说每个月蔬菜的安全指数(即风险大小)和合格率在上表所示的区间内。我们发现：3月的安全指数的置信区间上限明显低于1月和2月，其中2月最高；从蔬菜的合格率这一角度来看，2月的合格率区间跨度较大，说明蔬菜的质量和风险波动性高于其他月份。

### 6.5.4 设计抽样方案

鉴于上面对于附件中的抽样方案的评估结果，我们根据科学的统计推断原理来设计抽样方案。由上面的平均数抽样误差公式： $\Delta_{\bar{x}} = t \times \mu_{\bar{x}} = t \times \sqrt{\frac{\sigma_{\bar{x}}^2}{n}}$ 可以得到： $n = \frac{t^2 \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$  同理由成数的抽样误差可以得出： $n = \frac{t^2 P(1-p)}{\Delta_p^2}$ 。以安全指数为例，现在如果对每个月的安全指数的总体方差进行估计，然后自己确定一个比较理想的极限误差和置信度，就可以得出某个月的抽样次数；然后根据前面的各个蔬菜输入渠道的风险大小比例进行分配，最后以各个区的人口数作为衡量各个区面对的蔬菜风险大小的标准，把抽检次数再分配到各个区。所以首先应该对每个月安全指数的总体方差进行区间估计，从抽样理论可知：当总体方差未知、 $n$ 为大样本（即 $n \geq 30$ ）甚至当样本容量相当大时，即使



总体分布形式未知或总体为非正态分布，此时样本均值近似服从正态分布。所以问题转化为对单个正态分布的方差进行区间估计。

### 一、单个正态总体方差的区间估计

设总体  $X \sim N(\mu, \sigma^2)$ ,  $(X_1, X_2, \dots, X_n)$  为来自总体的一个样本，已给定置信度（水平）为  $1 - \alpha$ ，求  $\sigma^2$  的置信区间。

(1) 当  $\mu$  已知时，由于  $X \sim N(\mu, \sigma^2)$ ，因此， $\frac{X_i - \mu}{\sigma} \sim N(0, 1) (i = 1, 2, \dots, n)$  由  $\chi^2$  分布的定义知：

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n),$$

据  $\chi^2(n)$  分布上  $\alpha$  分位点的定义，有：

$$P\{\chi_{1-\frac{\alpha}{2}}^2(n) < \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} < \chi_{\frac{\alpha}{2}}^2(n)\} = 1 - \alpha$$

从而，

$$P\left\{\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)}\right\} = 1 - \alpha$$

故  $\sigma^2$  的置信度为  $1 - \alpha$  的置信区间为：

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)}\right)$$

(2) 当  $\mu$  未知时，据抽样分布有：

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

类似以上过程，得到：

$$P\left\{\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2}\right\} = 1 - \alpha$$

$\sigma^2$  的置信度为  $1 - \alpha$  的置信区间为：

$$\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2}\right)$$

$\sigma$ 的置信度为 $1 - \alpha$ 的置信区间为：

$$\left( \sqrt{\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}}} \right)$$

特别地，若置信系数为95%，则上述区间变为 $(0.615s^2, 1.905s^2)$  按照上述理论，我们选择2017年3月份的抽检方案作为样本，给出置信水平 $\sigma^2 = 0.95$ ，从而 $t=1.96$ 。由前面的每个月的抽检误差分析可得3月的抽检出来的不合格添加物的安全指数的平均方差：

$$s^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = (7.2489 \times 116 + 0 \times 248 + 11.7458 \times 87) / (116 + 248 + 87) = 4.1303$$

这里我们有95%的把握说总体的方差可以为 $\sigma_i^2 = s^2 \times 1.3 = 5.3694$ 。

## 二、抽样单位数目的计算

1平均数的计算公式：

$$\Delta_{\bar{x}} = t \times \mu_{\bar{x}} = t \times \sqrt{\frac{\sigma_{\frac{2}{x}}}{n}} \rightarrow n = \frac{t^2 \sigma_{\frac{2}{x}}}{\Delta_{\bar{x}}^2}$$

2.成数计算公式：

$$n = \frac{t^2 P(1 - P)}{\Delta_p^2}$$

总体的极限误差 $\Delta_{\bar{x}}$ 选取表11中最小的一个，即 $\Delta_{\bar{x}} = 0.1605$ 。每个月的抽样总数 $N=2600$ 不变。所以根据平均数计算公式可以得出 $n=612$ 。

## 三、抽样数的分配

### 1.分配到各渠道

根据前面的蔬菜输入渠道风险评估模型以及主成分分析法的得分，可以得出各个渠道的风险大小比例分别为：

农贸市场：17.0999/(17.0999+11.7872+0.3309+18.6903)=0.3569

同理可以得到批发市场、大型超市、小型超市的风险大小比例分别为：0.2460 0.0069 0.3901

按照风险比例越大分配抽样数量越大的原则，农贸市场、批发市场、大型超市、小型超市这4个输入渠道每个月的抽样数目分别为218.4410 150.5745 4.2271 238.7575。因为抽样数目为正整数，所以四舍五入后得到这4个输入渠道每个月的抽样数目分别为218 181 4 239。

### 2.分配到各个区

综合考虑耕地面积（公顷）、人口密度（人/平方米）、社会消费品零售总额（亿元）、超市数量、学校数量等因素，以上述因素为评价指标，各区为待评价对象，构造矩阵（见附录六），进行主成分分析，得到的抽检指标分配如下表：

表 12: 利用主成分分析解决抽检指标分配问题

地区 渠道	南山区	福田区	宝安区	罗湖区	龙华区
A	18	36	40	24	55
B	15	30	34	20	46
C	0	1	1	0	1
H	19	39	44	27	61

地区 渠道	盐田区	龙岗区	坪山区	光明新区	大鹏新区
A	3	28	3	9	1
B	3	23	3	8	1
C	0	1	0	0	0
H	3	31	4	10	1

对于蔬菜风险比较大的月份(比如说7、8、9月),可以适当地增加抽样数量。所以结合月份和渠道风险以及各个区具体情况来分配抽样数量,这将是后续继续研究的内容。

## 6.6 方差分析模型的建立与求解

我们发现不合格添加物的种类和在蔬菜中的含量会影响到蔬菜质量,所以以我国蔬菜大省山东省和广东省为研究对象,以苯甲酸等不合格添加物为指标,通过方差分析对各种不合格添加物的检出值进行分析来看这两个省之间蔬菜风险是否有显著差异,从而可以分析不同的产地对蔬菜质量的影响是否显著。从附件3可以得到山东省和广东省的主要集中不合格添加物的被检测值的平均值如下表所示:

地区 添加物	苯甲酸	毒死蜱	环己基氨基磺酸钠(甜蜜素)	甲拌磷
山东	2312	0.495	1693.529412	3.404
广东	2239.25	0	2041.33	0

地区 添加物	克百威	糖精钠	氧乐果	总砷
山东	0.068	0.421	0.199	0.94
广东	0	633.84	0.129	2.3

通过图表我们可以更加直观地看出:

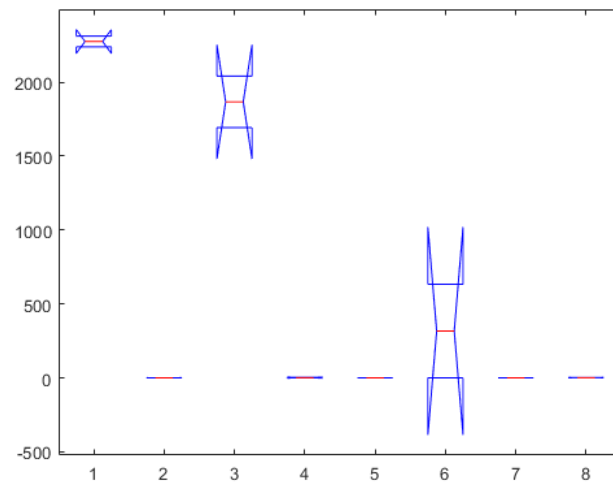


图 2: 方差分析1

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	12550962	7	1792994.6	54.39	4.03864e-06
Error	263745.6	8	32968.2		
Total	12814707.5	15			

图 3: 方差分析2

$H_0$ :各个蔬菜产地之间没有显著差异

从上图可以看出，检验值4.03864e-06明显小于0.05，拒绝原假设，说明不同省份之间的蔬菜质量有明显差异，从而蔬菜安全也受到产地的影响。弄清楚蔬菜产地是如何影响蔬菜质量的将是我们后续研究的内容。

## 6.7 聚类分析模型的建立与求解

### 6.7.1 模型的建立

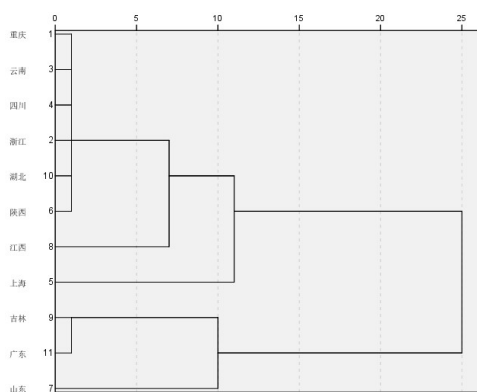
为了得到不同省份之间蔬菜风险的大小及其关系，我们以不同省份为研究对象，以每个省份在每个月的安全指数之和为指标用Ward法和K-Means方法分别进行聚类分析。从而得出各个省的蔬菜风险大小以及关系。

### 6.7.2 数据的处理

基于附件3，我们可以得出聚类分析的原始数据为：附录2-各省份聚类分析相关数据

### 6.7.3 模型的求解

Ward法求解结果如下图所示：



从上图可以看出：广东、山东、吉林三省聚成一类，说明这几个省份的蔬菜风险比较大，这可能是因为广东省进口蔬菜多而杂、山东省种植面积大而导致监管难度大所致。其它省份的蔬菜风险相对较小，尤其是重庆、陕西等省份，这是由于这些省份蔬菜种植面积小并且监管比较有效所致。K-Means方法求解结果如下：

根据蔬菜风险可以把产地分为3类：第一类：重庆、四川、江西；第二类：吉林、广东；第三类：其它省份。对比ward方法，我们可以发现对于产地的划分大致和Ward相同，但是山东没有被单独划成一类。综合上面两种结果，我们可以知道：在中东部的一些省份（比如说山东、广东），由于蔬菜种植规模大而导致蔬菜风险比较大；相比之下，西部省份的蔬菜风险就偏小。

## 6.8 致深圳市政府的一封信

尊敬的深圳市政府领导：

随着社会经济的快速发展，农业现代化进程的加快，蔬菜质量极易受到农药残留等各方面因素的影响，蔬菜质量安全问题层出不穷。而蔬菜质量安全不仅关系到人民群众的身体健康和生命安

全,还关系到社会的和谐稳定和经济的平稳发展,所以政府是否能够对蔬菜质量安全进行合理有效的风险评估和设计抽样方案显得至为关键。通过对深圳市新鲜蔬菜输入渠道风险的分析,我们根据建模和研究的结果给出如下建议:

第一,针对于蔬菜这个总体来说,市场上的主要销售的蔬菜种类大致可以分为鲜菜、干燥菜、酱腌菜、其他菜。通过研究发现鲜菜的安全风险是最大的。氧乐果、克百威、二氧化硫等农药残留量在鲜菜中所占的比例和超标率均明显比其他农药残留大,说明这几种农药是比较常见的农药,且这几种农药均为高毒农药,是蔬菜风险的主要来源,所以政府应当在检测过程中要对其重点监管,防止农药残留超标蔬菜进入人民群众的家庭;还应加大对农药产品的研究,使其代替高毒农药,从而减少高毒农药的使用;经分析还发现酱腌菜的风险也明显高于干燥菜和其他菜,所以应对鲜菜和酱腌菜的监管给予重视。

第二,针对于鲜菜以及各个输入渠道来说,基于对批发市场、农贸市场、超市等蔬菜常规输入渠道以及网络购物、流动摊贩等特殊输入方式的风险分别进行研究分析。农贸市场、批发市场、超市这三渠道尤其是农贸市场这一渠道在每个月被检测农药残留超标的比例远远高于其他渠道,这说明农贸市场一直是鲜菜的主要供应渠道,它的种类多以及输入蔬菜的比例大导致了该渠道监管比较困难、风险较大,因此政府在给各个渠道安排抽检次数时,应给这3个渠道分配较多抽检次数,对鲜菜抽检数量的增加、加大对农贸市场进出口监管力度、增加抽检人员等,使该渠道的风险尽可能降低。再分析发现在7月、9月、10月、12月的蔬菜风险大于其它月,这说明这几个月为蔬菜的上市季节,蔬菜种类及数量相比于其它月份明显增多,所以应该重点在这几个月(夏秋季节)加大抽检力度。在农贸市场这一渠道中,氧乐果、克百威、二氧化硫等农药残留的风险很高,甲胺磷农药残留的风险偏低,这也就验证了上面氧乐果、克百威、二氧化硫等农药残留风险偏高、甲胺磷农药残留风险低的结论。从而更加坚定地加大对上述农药残留的检测和监管力度。

第三,通过对蔬菜产地的研究发现:广东、山东、吉林三省聚成一类,说明这几个省份的蔬菜风险比较大。这可能是因为广东省进口蔬菜多而杂、山东省种植面积大而导致监管难度大所致。其它省份的蔬菜风险相对较小,尤其是重庆、陕西等省份,这是由于这些省份蔬菜种植面积小并且监管比较有效所致。所以在对蔬菜进行抽检时,应该重点对广东、山东等省份的蔬菜进行检测。

第四,针对于蔬菜的抽样方案设计来说,我们首先经过建模发现原方案的抽样误差较大。所以根据前面评估出的各个蔬菜输入渠道的风险大小以及深圳市各个区的人口密度,应该主要对农贸市场和批发市场进行抽样检查;对福田区、罗湖区、龙华新区等人口比较集中的地方进行重点抽样以降低风险。当通过科学的方法分析出各个渠道的风险后,通过利用科学的推断统计原理设计出一套合理的误差小的抽样方案,相信蔬菜风险会越来越低,菜篮子工程会做得越来越好!

## 7 模型的检验

为了进一步评价前面的渠道风险模型,我们用TOPSIS综合评价法来评估每种渠道的风险大小。首先构造决策矩阵(输入矩阵),其中行为每种不合格添加物,列为不同蔬菜输入渠道,以每种渠道中所得出的每种不合格添加物的安全指数的最大值为指标,进行综合评价,决策矩阵如下所

示：

$$\begin{bmatrix} 29.0194 & 20.2217 & \cdots & 26.1566 \\ 0.8233 & 0 & \cdots & 21.85 \\ \vdots & \vdots & \ddots & \vdots \\ 3.6353 & 0 & \cdots & 45.5987 \end{bmatrix} \quad (9)$$

(1)为去掉量纲效应，做规范化处理得到 $B = (b_{ij})_{m \times n}$  其中：

$$b_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}^2}}, \quad i = 1, \dots, m, j = 1, \dots, n$$

$$B = \begin{bmatrix} 0.84 & 1.00 & 1.00 & 0.27 & 1.00 & 0.42 & 0.42 & 0.91 & 0.23 & 0 & 0.28 \\ 0.02 & 0 & 0 & 0.34 & 0 & 0.17 & 0.48 & 0 & 0.11 & 1.00 & 0.23 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0.10 & 0 & 0 & 0 & 0 \\ 0.10 & 0 & 0 & 0.39 & 0 & 0.41 & 0 & 0.09 & 0.66 & 0 & 0.49 \end{bmatrix} \quad (10)$$

(2)根据每个评价指标对评价结果的贡献程度的不同，指定不同的权重： $w = [w_1, \dots, w_n]$ ，这里假设各个指标的贡献率相同，即 $w_j=1$ 。将 $B$ 的第 $j$ 列乘以其权重 $w_j$ ，得到加权规范矩阵 $C = (c_{ij})_{m \times n} = B$ 。

(3) 确定正理想解 $C^*$ 和负理想解 $C^0$ ，把每种不合格添加物安全指数的最大值作为正理想解，把最小值作为负理想解，可以得到：

$$C^* = [0.8367 \quad 1 \quad 1 \quad 0.3944 \quad 1 \quad 0.4182 \quad 0.4806 \quad 0.9107 \quad 0.6558 \quad 1 \quad 0.4871]$$

$$C^0 = [0.1048 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

(4)计算每个待评价样本到正理想解和负理想解的距离 $d_i^*$ 和 $d_i^0$ ：

$$d_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, \quad i = 1, \dots, m$$

$$d_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, \quad i = 1, \dots, m$$

(5)计算每个待评价样本的评价参考值：

$$f_i = \frac{d_i^0}{d_i^0 + d_i^*}, \quad i = 1, \dots, m$$

再将 $f_i$ 从大到小排列，得到各渠道的风险的优劣结果如下所示：0.6683 0.3511 0.0389 0.3010  
从结果可以看出，数值越大说明离最小值（风险最小，也即最优）距离越远，从而风险越大。

因此我们可以知道农贸市场的风险最大，而且是批发市场和小型超市的两倍左右。批发市场和小型超市风险接近，大型超市风险最小。这也与前面蔬菜输入渠道模型和主成分分析法得出的结果大致相同，这也就验证了我们前面的渠道风险评估模型是比较合理和准确的。

## 8 模型的改进

### 8.1 熵值法

在安全指数法模型中，我们把每个月每种不合格添加物检出的最大值作为每一种不合格添加物的人均实际摄入量。但是消费者把菜买回家以后，会进行各种形式的清洗，这就会使得蔬菜上的不合格添加物量减小，从而每一种不合格添加物的人均实际摄入量也会相应的减小。由于清洗后蔬菜上被去掉的不合格添加物量受到清洗方式、不合格添加物的溶解度等因素的影响，并且不合格添加物对人体的危害程度还与它本身的毒性有关，因此，我们想研究不同不合格添加物被清洗后的不合格添加物量与清洗前的关系。鉴于不合格添加物的溶解度与被清洗后的残留量之间没有明确的关系，我们以溶解度和毒性作为衡量不合格添加物被清洗后的风险大小的指标，用熵值法得出每种不合格添加物的比重（假设室温20度，用清水清洗），各种不合格添加物的溶解度( $mg/l$ )及毒性如下表所示：

表 13: 不合格添加物的溶解度及毒性

添加物 性质	a	b	c	d	e	f	h	i	g	k
溶解度	0.07	1	10	10	0.01	0.42	3.9	0.02	0.02	0.00019
毒性	3	3	1	1	3	1	3	1	3	1

注：在毒性一栏中，1表示低毒,3表示高毒。

(1)指标的归一化处理

由于溶解度数值越高越好，因此，它属于正项指标；毒性的数值越低越好，因此它属于负项指标，对于高低指标，我们用不同的算法进行数据标准化处理，其具体方法如下：

计算溶解度：

根据

$$x_{ij}^* = \frac{x_{ij} - \min\{x_{ij}, \dots, x_{nj}\}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{ij}, \dots, x_{nj}\}}$$

对溶解度这一正向指标进行标准化处理后可以得出

$$x_{ij}^* = [0.0070, 0.1000, 1.0000, 1.0000, 0.0010, 0.0420, 0.3900, 0.0020, 0.0020, 0.001]$$

同理，根据

$$x_{ij}^* = \frac{\max\{x_{ij}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{ij}, \dots, x_{nj}\}}$$



对毒性这一负向指标进行标准化处理后可得： $x_{ij}^* = [0.001, 0.001, 1, 1, 0.001, 1, 0.001, 1, 0.001, 0.001]$

由于熵值法中，属性的值不能出现0，所以我们将上述经标准化处理后的数据中的0 改为0.001，更改后的标准化数据为：

$$\begin{bmatrix} 0.0070 & 0.1000 & 1.0000 & 1.0000 & 0.0010 & 0.0420 & 0.3900 & 0.0020 & 0.0020 & 0.0010 \\ 0.0010 & 0.0010 & 1.0000 & 1.0000 & 0.0010 & 1.0000 & 0.0010 & 1.0000 & 0.0010 & 0.0010 \end{bmatrix} \quad (11)$$

(2)计算第j项指标下第i个对象占该指标的比重：

根据

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, \quad i = 1, \dots, n, j = 1, \dots, m$$

可以得出(单位 $10^{-4}$ )：

$$\begin{bmatrix} 28 & 393 & 3929 & 3929 & 3.93 & 165 & 1500 & 7.86 & 7.86 & 3.93 \\ 2.50 & 2.50 & 2500 & 2500 & 2.50 & 2496 & 2.50 & 2496 & 2.50 & 2.50 \end{bmatrix} \quad (12)$$

从上面可以看出，有些不合格添加物比如说乙酰甲胺磷、毒死蜱的溶解度很大，远远高于其他不合格添加物。有些不合格添加物比如是a、k的溶解度又很小，几乎不溶，这对于消费者买回后是否能够洗掉绝大部分添加物有很大影响。有些不合格添加物比如说克百威、甲胺磷、甲拌磷、氧乐果、阿维菌素的毒性较大而且溶解度又较低，所以要引起高度重视。

(3)由 $e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij}$ (其中 $k = 1/\ln(n) > 0$ )计算第j项指标的熵值：得到溶解度和毒性这两个指标的熵值为0.5429 0.6072。由 $d_j = 1 - e_j$ 计算信息熵冗余度得到两个指标的信息熵冗余度分别为 $d_1 = 0.4571$ ,  $d_2 = 0.3928$

(4)由 $u_j^* = \frac{d_j}{\sum_{j=1}^m d_j}$ 计算各项指标的权值得到：

$$w_1 = 0.4571/(0.3928 + 0.4571) = 0.5378$$

$$w_2 = 0.3928/(0.3928 + 0.4571) = 0.4622$$

由此可见，溶解度和毒性这两个指标的权重差不多，溶解度权重稍微大一点，这说明影响消费者洗菜后风险的主要因素是该不合格添加物的溶解度。

(5)根据 $s_i = \sum_{j=1}^m u_j^* \times p_{ij}$ 计算不合格添加物的综合得分如下：

$$s = [0.0016, 0.0212, 0.3267, 0.3267, 0.0003, 0.1242, 0.0825, 0.1158, 0.0005, 0.0003]$$

从最终结果可以看出，得分越高的不合格添加物(比如说乙酰甲胺磷、毒死蜱)的溶解度越大，毒性越小。也就是说清洗后风险越小。于是我们就用 $(1-s_i)$ 作为安全指数法中的校正因子 $Z_j$ ，这就使得经过清洗后风险小的不合格添加物(溶解度大的，毒性小的)的人体最大摄入量大幅度减少，与

实际更加符合。

## 9 模型评价

### 9.1 优点

#### 1.蔬菜风险评估模型

(1)对于影响蔬菜风险的主要因素(不合格添加物)进行了比较全面的分析,结果也比较符合现实情况。

(2)计算简单,其中所需的量容易求得。

#### 2.渠道风险评估模型(包括安全指数法、主成分分析法)

(1)安全指数法采用比值定义的方法,可以客观准确的评价每种不合格添加物的风险大小,其适用面比较广。

(2)主成分分析法不仅能够找出不同不合格添加物之间的危害性的关系,还能得出每个渠道的综合得分(即风险大小)。

#### 3.分层抽样模型

(1)能够基于附件有效地定性定量地评估抽样误差,为后面设计出更好的抽样方案做好准备。

(2)方案设计模型能够比较准确而合理地基于前面的抽样误差评估结果,借助于分层抽样的原理设计出抽样误差比原方案小、又能细化到每个月每个区各个蔬菜输入渠道的抽检方案。

#### 4.其它模型

(1)TOPSIS法可以较为全面地结合所有不合格添加物对各渠道风险进行评估,而且结果和现实很接近(农贸市场的风险明显大于其它渠道)。

(2)熵值法通过考虑到不合格添加物溶解度和毒性对安全指数法中的人体每日摄入量进行了修改,使得安全指数法更加与实际情况相吻合。

### 9.2 缺点

#### 1.蔬菜风险评估模型

(1)模型比较简单,没有将产地对于蔬菜安全的影响考虑进去。

(2)蔬菜种类过多,没有进行进一步的风险评估,只对鲜菜等四大类菜进行了粗略评估。

#### 2.渠道风险评估模型(包括安全指数法、主成分分析法)

(1)安全指数法中的人体实际摄入量无法获得,只能近似估计。

(2)在主成分分析法中,由于实际不合格添加物种类繁多,没有逐一考虑,只对一些主要的检出率比较大的进行分析,导致结果不准确。

#### 3.分层抽样模型

(1)在进行抽样误差估计时，对极个别的少见的合格添加物忽略不计。可能会导致抽样误差不准确。

(2)在进行抽样数量计算和设计分配方案时，总体方差估计可能不够准确，各个区之间的风险大小没有进行全面评估，而只是以人口密度作为各个区的蔬菜风险衡量标准，从而导致分配方案不够合理。

#### 4.其它模型

熵值法只能说改进安全指数法中的人体实际合格添加物摄入量，并不能准确反映蔬菜经过清洗后到底能洗去多少；而且合格添加物的被清洗掉的比例也与清洗方式和次数以及气温有关，熵值法并没有考虑这些因素的影响。

## 10 模型的总结与展望

### 10.1 总结

1.通过对蔬菜及蔬菜输入渠道风险的评估研究，我们不仅找出了不同蔬菜种类的风险大小，而且对于各输入渠道的风险进行了比较详细而准确的分析。发现结果和现实中的情况比较接近：鲜菜和酱腌菜的风险相对较大；农贸市场和批发市场的风险也比较大，这可能与这两个渠道、数量较多所占比重较大而导致监管难度增大所致。通过分析还发现你有之间也存在着一定的相关性。

2.通过对现行蔬菜抽样方案的抽样误差分析，发现现行抽样方案误差较大。接着根据科学的推断统计的原理设计出误差更小、更加合理的抽样方案。

### 10.2 展望

1.以后将继续研究蔬菜产地以及种植方式对蔬菜质量的影响，建立更加合理的蔬菜输入渠道评估模型。

2.以后将对深圳市各个区的蔬菜风险进行全面评估，即不仅仅考虑人口密度对蔬菜风险的影响，还会考虑自然环境等因素的影响，建立更加完善的各个区的蔬菜风险评估模型。

## 参考文献

- [1] 王冬群等. 慈溪市蔬菜合格添加物时空变化及质量安全风险评估[J]. 2009,21(6):609-613.
- [2] 李金昌. 分层整群抽样初探[J]. 1994,59(3).
- [3] 陈军. 分层抽样与整群抽样的结合使用[J]. 统计与决策. 2006,15.070.
- [4] 张超. 基于主成分分析法与平衡理论的食品安全抽检模型[J]. 2011.
- [5] 左方华等. 广西蔬菜合格添加物动态及质量安全风险评估[J]. 2011,39(1):208-210.

- [6] 温雅君.批发市场蔬菜不合格添加物调查及风险评估[J]. 2011.

附 录 一

鲜菜中的各种添加物的检出情况

添加物	二氧化硫残留量（SO2计）	6-苄基腺嘌呤
比例	0.003	0.019
标准值	≤0.1g/kg	不得检出
检出值	1.02g/kg	0.0765mg/kg
超标率	9.2	
添加物	涕灭威（涕灭威、涕灭威砒和涕灭威亚砒之和）	亚硫酸盐
比例	0.0032	0.0032
标准值	≤0.03mg/kg	≤0.2g/kg
检出值	0.51mg/kg	3.24g/kg
超标率		
添加物	苯甲酸	苯甲酸及其钠盐（苯甲酸计）
比例	0.00967	0.0064
标准值	≤1.0g/kg	≤1.0g/kg
检出值	3.3g/kg	2.1925g/kg
超标率	2.3	1.192
添加物	苯醚甲环唑	苯醚甲环唑
比例	0.0032	0.0032
标准值	≤0.5mg/kg	不得检出
检出值	1.1mg/kg	不得检出
超标率	1.2	
添加物	代森锰锌	滴滴涕
比例	0.0169	0.0032
标准值	≤1 mg/kg	≤0.2mg/kg
检出值	5.8mg/kg	1.0mg/kg
超标率	4.8	4
添加物	啶虫脒	毒死蜱
比例	0.0323	0.0548
标准值	≤1mg/kg	≤0.1mg/kg
检出值	4.046mg/kg	1.0476mg/kg
超标率	3.046	9.4700
添加物	毒死蜱,甲基异柳磷	毒死蜱,克百威
比例	0.0032	0.0032
标准值	≤0.1mg/kg—≤0.01mg/kg	≤0.1mg/kg—
检出值	0.343mg/kg—0.16mg/kg	0.22mg/kg—0.10mg/kg
超标率	2.43 45	1.2
添加物	多菌灵	二氧化硫
比例	0.0064	0.00967
标准值	≤0.5mg/kg	不得添加
检出值	1.3212mg/kg	0.02286g/kg
超标率	1.6424	

添加物	二氧化硫残留量	啉虫脒
比例	0.0032	0.0323
标准值	≤0.2mg/kg	≤1mg/kg
检出值	0.83mg/kg	4.046mg/kg
超标率	3.15	3.046
添加物	毒死蜱	毒死蜱,甲基异柳磷
比例	0.0774	0.0032
标准值	≤0.05mg/kg	≤0.1mg/kg—≤0.01mg/kg
检出值	0.75627mg/kg	0.343mg/kg—0.16mg/kg
超标率	14.1254	2.43
添加物	毒死蜱,硫丹	多菌灵
比例	0.0032	0.0064
标准值		≤0.05mg/kg—不得检出
检出值	0.15mg/kg—0.14mg/kg	1.3212mg/kg
超标率	2	1.6424
添加物	二氧化硫	二氧化硫残留量
比例	0.0129	0.0226
标准值	≤0.05 mg/kg	≤0.1g/kg
检出值	1.32002g/kg	0.721g/kg
超标率	25.4000	6.21
添加物	苯甲酸、山梨酸用量占其最大使用量比例之和	氟虫腈
比例	0.0032	0.0096
标准值	不超过1	≤0.02mg/kg
检出值	1.31	0.056mg/kg
超标率		0.12
添加物	腐霉利,硫丹	腐霉利,氯氟氰菊酯和高效氯氟氰菊酯
比例	0.0064	0.0032
标准值	≤0.2mg/kg—不得检出	≤0.2mg/kg—≤0.5mg/kg
检出值	2.71mg/kg	0.64mg/kg—1.25mg/kg
超标率	12.5500	2.2
添加物	镉	镉
比例	0.0064	0.0064
标准值	≤0.1mg/kg	≤0.2mg/kg
检出值	0.155mg/kg	0.35mg/kg5
超标率	0.55	0.75
添加物	铬	环己基氨基磺酸钠
比例	0.0064	0.0064
标准值	≤0.5mg/kg	≤1.0g/kg
检出值	0.7mg/kg75	1.5g/kg
超标率	0.4	0.5

添加物	甲氨基阿维菌素苯甲酸盐	甲胺磷	赭曲霉毒素A
比例	0.0032	0.0064	0.0032
标准值	$\leq 0.02\text{mg/kg}$	$\leq 0.05\text{mg/kg}$	$\leq 5.0 \mu\text{g/kg}$
检出值	0.05mg/kg	0.6145	14.3 $\mu\text{g/kg}$
超标率	1.5000	11.2900	1.8600
添加物	甲基异柳磷	酒精度	总汞g
比例	0.0064	0.0032	0.0026
标准值	$\leq 0.01\text{mg/kg}$	$3.8 \pm 1.0\%\text{vol}$	$\leq 0.01\text{mg/k}$
检出值	0.3195mg/kg	2.2%vol	0.156mg/kg
超标率	30.9500		14
添加物	乐果	硫丹	敌敌畏
比例	0.0064	0.035	0.003
标准值	$\leq 1\text{mg/kg}$	不得检出	$\leq 0.2\text{mg/kg}$
检出值	2.18mg/k	0.1024mg/kg	0.83mg/kg
超标率	1.18		3.15
添加物	氯氰菊酯和高效氯氰菊酯	灭蝇胺	苯甲酸及其钠盐（苯甲酸计）
比例	0.0032	0.0032	0.0032
标准值	$\leq 1\text{mg/kg}$	$\leq 0.5\text{mg/kg}$	$\leq 1\text{g/kg}$
检出值	1.32mg/kg	1.668mg/kg	1.4g/kg
超标率	0.32	2.3360	0.4
添加物	铅	铅	毒死蜱,硫丹g
比例	0.0258	0.0032	0.0032
标准值	$\leq 0.1\text{mg/kg}$		$\leq 0.05\text{mg/kg}$ —不得检出
检出值	0.35375mg/kg	0.47mg/kg	0.15mg/kg—0.14mg/k
超标率	2.5375	0.5667	2
添加物	山梨酸	水胺硫磷	毒死蜱
比例	0.096	0.0129	0.0548
标准值	$\leq 0.5\text{g/kg}$	不得检出	$\leq 0.1\text{mg/kg}$
检出值	0.760g/kg	0.535mg/kg	1.0476mg/kg
超标率	0.5200		9.470
添加物	辛硫磷	亚硫酸盐	二氧化硫
比例	0.035	0.0096	0.00967
标准值	$\leq 0.05\text{mg/kg}$	$\leq 0.02 \text{ g/kg}$	$\leq 0.5\text{mg/kg}$
检出值	0.127mg/kg	0.0687g/kg	0.02286g/kg
超标率	1.5400	2.4350	
添加物	氧乐果	乙二胺四乙酸二钠	腐霉利
比例	0.1193	0.0096	0.0226
标准值	$\leq 0.02\text{mg/kg}$	$\leq 0.25\text{g/kg}$	$\leq 0.2\text{mg/kg}$
检出值	0.5739mg/kg	3.01g/kg	3.145mg/kg
超标率	27.6950	11.0400	14.7250

添加物	荧光增白物质	镉	镉
比例	0.0064	0.0161	0.0032
标准值	不得检出	$\leq 0.05\text{mg/kg}$	$\leq 0.5\text{mg/kg}$
检出值	阳性	0.16mg/kg	2.1mg/kg
超标率		2.2	3.2
添加物	总砷	环己基氨基磺酸钠	甲拌磷
比例	0.0096	0.0032	0.0129
标准值	$\leq 0.5\text{mg/kg}$	$\leq 0.3\text{g/kg}$	$\leq 0.01\text{mg/kg}$
检出值	1.03mg/kg	2.54g/kg	1.735mg/k
超标率	1.06	7.4667	172.5000
添加物	6-苄基腺嘌呤	克百威	氯氟氰菊酯和高效氯氟氰菊酯
比例	0.0064	0.0838	0.0032
标准值	0.0 (mg/kg)	$\leq 0.02\text{mg/kg}$	$\leq 0.5\text{mg/kg}$
检出值	0.66mg/kg	0.1218mg/k	0.9mg/kg
超标率		5.0900	0.8000
添加物	大肠杆菌	柠檬黄,亮蓝	山梨酸
比例	0.0129	0.0032	0.0032
标准值	$\leq 30\text{MPN}/100\text{g}$	不得检出	不得添加
检出值	62880MPN/100g 0.0084g/kg	0.033g/kg	
超标率	2095	16	15.2000
添加物	毒死蜱	乙酰甲胺磷	二氧化硫残留量
比例	0.0774	0.0032	0.0903
标准值	$\leq 0.05\text{mg/kg}$	$\leq 1\text{mg/kg}$	$\leq 0.2\text{g/kg}$
检出值	0.75627mg/kg	2.72mg/kg	4.585g/kg
超标率	14.1254	1.72	21.9250
添加物	二氧化硫	毒死蜱,克百威	
比例	0.0129	0.0032	
标准值	$\leq 0.05\text{ mg/kg}$	$\leq 0.1\text{mg/kg}$	
检出值	1.32002g/kg	0.22mg/kg—0.10mg/kg	
超标率	25.40000	1.2	



## 附 录 二

各省份聚类分析相关数据

省份 \ 添加物		a	d	e	h	m	f	w	s
重庆		0	0	0	0.9779	0.741	0	12.2867	698.26
浙江		0	0	0	0	1.0916	0	0	11.2241
云南		0	1.7944	0	0	0	0	0	39.1219
四川		0	0	0	0	0.7091	0	55.944	640.7524
上海		5.32	2.4489	0	0.342	0.1077	3.1902	0	2.9495
陕西		0	0	0	0	1.8734	0	0	28.1743
山东		3.901	4.18	4.3117	2.767	11.7965	0	35.0656	2.85
江西		0	0	0	0	0.957	0	0	309.8267
吉林		0	0	0	0	3010.5	0	0	88.2098
湖北		0	0	0	0	10.8242	0	0	66.2738
广东		0	0	0	0.1634	1082.7	0	0	266.9356
地区 \ 添加物		o	u	p	y	t	q	z	v
重庆		2.0267	8.36	0	0	0	0	0	0
浙江		0	0	2.2167	0.1364	0.6295	0	0	0
云南		0	1.0311	0	0	0	0	0	0
四川		13.933	0.0504	0.9078	0.5354	0	0.1419	0	0
上海		0	0	3.6206	0	0.4053	0	0	0
陕西		0.6967	0	2.3011	0.6156	8.1143	0	7.7152	1.406
山东		0	2.1508	0.1794	0.1429	7.4733	0.7845	9.4424	0
江西		0	0	4.4333	0.6126	0	428.018	8.9242	0
吉林		0	0.1491	0	0	1.3553	0.0786	0	0
湖北		0	1.8747	0	0	1.1286	0	10.3636	0.798
广东		0	0	3.2553	0.0081	8.8312	0.1129	0	0

### 附 录 三

综合渠道风险评估模型-不合格添加物的安全指数

$R=$

$$\begin{bmatrix} 1.0000 & 0.9958 & 0.9958 & 0.1138 & 0.9958 & 0.6139 & 0.4091 & 0.9999 & 0.0487 & -0.3840 & 0.1716 \\ 0.9958 & 1.0000 & 1.0000 & 0.0679 & 1.0000 & 0.5507 & 0.4719 & 0.9955 & -0.0392 & -0.3333 & 0.0982 \\ 0.9958 & 1.0000 & 1.0000 & 0.0679 & 1.0000 & 0.5507 & 0.4719 & 0.9955 & -0.0392 & -0.3333 & 0.0982 \\ 0.1138 & 0.0679 & 0.0679 & 1.0000 & 0.0679 & 0.7686 & 0.1931 & 0.1255 & 0.7232 & 0.3355 & 0.9154 \\ 0.9958 & 1.0000 & 1.0000 & 0.0679 & 1.0000 & 0.5507 & 0.4719 & 0.9955 & -0.0392 & -0.3333 & 0.0982 \\ 0.6139 & 0.5507 & 0.5507 & 0.7686 & 0.5507 & 1.0000 & 0.0321 & 0.6209 & 0.7852 & -0.2687 & 0.8818 \\ 0.4091 & 0.4719 & 0.4719 & 0.1931 & 0.4719 & 0.0321 & 1.0000 & 0.4138 & -0.4947 & 0.6563 & -0.1540 \\ 0.9999 & 0.9955 & 0.9955 & 0.1255 & 0.9955 & 0.6209 & 0.4138 & 1.0000 & 0.0549 & -0.3767 & 0.1809 \\ 0.0487 & -0.0392 & -0.0392 & 0.7232 & -0.0392 & 0.7852 & -0.4947 & 0.0549 & 1.0000 & -0.3230 & 0.9334 \\ -0.3840 & -0.3333 & -0.3333 & 0.3355 & -0.3333 & -0.2687 & 0.6563 & -0.3767 & -0.3230 & 1.0000 & -0.0553 \\ 0.1716 & 0.0982 & 0.0982 & 0.9154 & 0.0982 & 0.8818 & -0.1540 & 0.1809 & 0.9334 & -0.0553 & 1.0000 \end{bmatrix}$$

(13)

### 附 录 四

综合渠道风险评估模型-所有不合格添加物在每种输入渠道中的安全指数之和

$$\begin{bmatrix} 29.0194 & 20.2217 & 1.0724 & 16.7833 & 0.076 & 0.6062 & 4.218 & 1.0978 & 0.0443 & 0 & 26.1566 \\ 0.8233 & 0 & 0 & 21.1744 & 0 & 0.2434 & 4.8753 & 0 & 0.0211 & 0.38 & 21.85 \\ 1.2033 & 0 & 0 & 0 & 0 & 0 & 1.0513 & 0 & 0 & 0 & 0 \\ 3.6353 & 0 & 0 & 24.7212 & 0 & 0.5999 & 0 & 0.1077 & 0.1246 & 0 & 45.5987 \end{bmatrix}$$

(14)

## 附录五

### 模型的检验-决策矩阵规范化处理

$$B = \begin{bmatrix} 0.8367 & 1.0000 & 1.0000 & 0.2678 & 1.0000 & 0.4182 & 0.4158 & 0.9107 & 0.2332 & 0 & 0.2794 \\ 0.0237 & 0 & 0 & 0.3378 & 0 & 0.1679 & 0.4806 & 0 & 0.1111 & 1.0000 & 0.2334 \\ 0.0347 & 0 & 0 & 0 & 0 & 0 & 0.1036 & 0 & 0 & 0 & 0 \\ 0.1048 & 0 & 0 & 0.3944 & 0 & 0.4139 & 0 & 0.0893 & 0.6558 & 0 & 0.4871 \end{bmatrix}$$

## 附录六

### 深圳市各区相关信息

$$\begin{bmatrix} 99 & 6961 & 737.11 & 2384 & 845 \\ 8 & 18317 & 1665.6 & 2925 & 477 \\ 772 & 7187 & 715.45 & 21717 & 2471 \\ 22 & 12389 & 1152.47 & 2383 & 342 \\ 248 & 8609 & 263.66 & 30122 & 5591 \\ 10 & 2964 & 71.09 & 527 & 60 \\ 627 & 5527 & 670.52 & 12474 & 1748 \\ 566 & 2132 & 70.90 & 166 & 110 \\ 1398 & 3417 & 110.40 & 3236 & 176 \\ 346 & 460 & 56.00 & 81 & 17 \end{bmatrix} \quad (15)$$

## 附录七

### TOPSIS法：

```
x=[29.0194 0.2217 1.0724 16.7833 0.076 0.6062 4.218 1.0978 0.0443 0 26.1566; 0.8233 0 0 21.1744 0
0.24341 4.8753 0 0.0211 0.38 21.85; 1.2033 0 0 0 0 0 1.0513 0 0 0 0; 3.6353 0 0 24.7212 0 0.5999 0
0.1077 0.1246 0 45.5987]
[m,n]=size(x);
y=[]
for i=1:n
    y(:,i)=x(:,i)/sum([x(:,i)]) end
```

```

for i=1:n
w(i)=max(y(:,i));
b(i)=min(y(:,i));
end
D1=zeros(m,1);
D2=zeros(m,1);
for i=1:m
for j=1:n
 $ma(i,j) = (y(i,j) - w(j))^2;$ 
 $mi(i,j) = (y(i,j) - b(j))^2;$ 
D1(i,1)=D1(i,1)+ma(i,j);
D2(i,1)=D2(i,1)+mi(i,j);
end
D1(i,1)=sqrt(D1(i,1));
D2(i,1)=sqrt(D2(i,1));
end
for i=1:m
c(i)=D2(i,1)/(D1(i,1)+D2(i,1));
end
for i=1:m
c(i)
end
0.6683
ans =
0.3511
ans =
0.0389
ans =
0.3010

```

熵值法：

```

function [s,w]=shang(x)
x=x';
[n,m] = size(x);

```

```

for i=1:n
for j=1:m
p(i,j)=x(i,j)/sum(x(:,j));
end
end
k=1/log(n);
for j=1:m
e(j)=-k*sum(p(:,j).*log(p(:,j))));
end
for j=1:m
e(j)
end
d=ones(1,m)-e;
w=d./sum(d);
 $s = w * p'$ ;

```

### 主成分分析法：

```

x=[29.0194 0.2217 1.0724 16.7833 0.076 0.6062 4.218 1.0978 0.0443 0 26.1566 0.8233 0 0 21.1744 0
0.24341 4.8753 0 0.0211 0.38 21.85 1.2033 0 0 0 0 0 1.0513 0 0 0 0 3.6353 0 0 24.7212 0 0.5999 0
0.1077 0.1246 0 45.5987 ];
stdr=std(x);
 $[n, m] = size(x)$ ;
sddata=x./stdr(ones(n,1),:)
 $[p, princ, egenvalue] = princomp(sddata)$ 
p3=p(:,1:3)
sc=princ(:,1:3)
egenvalue
per=100*egenvalue/sum(egenvalue)

```

### 安全指数法：

```

function f = IBS(x,y) f = 0.38 * x/(y * 60); end

```

## 总程序

```

#include < iostream >
#define Z1
#define K1
#define ADa10
#define m60
#define Ra0.001
#define Q1
#define J380

using namespace std;
void main() {
int size;
cout<< "请输入数据组数;
cin >> size;
double *r = new double[size];
double *ga = new double[size];
double *Ifa = new double[size];
double *VA = new double[size];
double *PA = new double[size];
double *WaA = new double[size];
double *result = new double[size];
double VAResult = 0;
cout <<" 请输入ga与Ifa的值";
for (int i = 0; i < size; i++) {
cin >> ga[i] >> Ifa[i];
}
for (int i = 0; i < size; i++) {
r[i] = ga[i] * Ifa[i];
}
double EDIa = Ra * J * K * Q;
double IBSa = EDIa * Z / (ADa * m);
cout << "请输入PA与WaA的值";
for (int i = 0; i < size; i++) {
cin >> PA[i] >> WaA[i];
}
}

```

```
}  
for (int i = 0; i < size; i++) {  
    result[i] = PA[i] * WaA[i];  
}  
for (int i = 0; i < size; i++){  
    VA[i] = PA[i] * r[i] * result[i];  
    VAResult += VA[i];  
}  
cout << VAResult;  
}
```