****

|  |  |
| --- | --- |
| **论文类别** |  |
| **论文题目** | 231100599多种方法进行的外卖垃圾对环境影响的评估 |

**摘要**

**关键词：**：

**目录**

1. **背景**
   1. **目前研究状况**
   2. **问题重述**
2. **假设**
   1. **假设**
   2. **符号说明**
3. **数据获取**
4. **模型建立**
   1. **问题一**
   2. **问题二**
   3. **问题三**
      1. **层次分析法**
      2. **线性回归法**
      3. **主成分回归法**
      4. **距离判别法**
      5. **贝叶斯判别法**
      6. **支持向量机**
5. **模型优化**
6. **结论**
   1. **敏感性分析**
   2. **优缺点分析**
   3. **结论**
   4. **问题四**
7. **参考文献**
8. **附录**
9. **背景**
   1. **目前研究状况**
   2. **问题重述**
10. **假设**
    1. **假设**
    2. **符号说明**
11. **数据获取**
12. **模型建立**
    1. **问题一**

在选取统计范围时，本文考虑的因素一方面是获取数据的难度。统计范围不应过大，这主要表现在这个区域的出入口不应过多。因为过大的区域难以确保选取数据的精准程度。较少的出入口可以保证在一段时间内尽可能多的拦截到所要调查的外卖配送人员和订外卖的用户。统计范围也应该具有一定的代表性和鲜明的特色。比如这是一个老年人为主的小区还是青年人为主的小区，或是一个公司。同时这个区域应该具有一定的外卖流量，因为流量过少难以确保获得的数据具有代表性，容易出现较大的随机误差。统计范围还应该考虑本文作者的熟悉程度，不应前往过于偏僻或者过于陌生的地方。综合以上因素考虑，本文将调查范围确定在了某公司。

调查得到的数据经过统计后如表所示。同时，考虑到本文选取的数据有一定误差，因此利用在线调查得到的数据进行了修正。本文将实地调查得到的数据和在线调查得到的数据赋予同等的权重，对数据进行处理，公式如下：

其中，*A*代表处理过后的这一选项的数据，代表在线调查获得的某一选项的数目，代表实地调查获得的某一选项的数目。代表在线调查获得的参与调查的总人数，代表实地调查获得的参与调查的总人数。这样处理完后，得到的数据如表所示。依据此，本文制得个个问题各个选项的统计图如图所示。

对于第七问“您每次定外卖大约有多少塑料餐盒”，第八问“您每次定外卖大约有多少塑料袋？”，第九问“您每次定外卖大约有多少纸餐盒？”这三个问题，本文进一步作了点估计和区间估计来对废弃物进行进一步定量的估计。

基于假设：每一个选项内使用餐盒或塑料袋的人数是均匀分布的，因此本文对于每一个选项中餐盒或者塑料袋的数量，本文均用选项上下限的中值来代表这个选项的每一个人使用的餐盒或者塑料袋的数目。对于没有上界的选项则用选项的下界来代表选这个选项每一个人的使用的餐盒或者塑料袋的数目。因此在第七，八两问中若选择选项1则对应的值是1，若选择选项2则对应的值是2.5，若选择选项3则对应的值是4，若选择选项4则对应的值是5。在第九问中若选择选项1则对应的值是2，若选择选项2则对应的值是5，若选择选项3则对应的值是8，若选择选项4则对应的值是10

每一个选项内使用餐盒或塑料袋的人数是均匀分布的。（作为假设，要写在假设里吗）

本文首先利用点估计求得每一个点外卖的人使用的餐盒和塑料袋的个数。点估计利用算术平均值公式，即

其中，代表算术平平均值的估计值，代表求得的这个问题的算术平均值，即点估计的结果。代表回答这份问卷的总人数，代表第i个人的选项对应的值。求得的结果第七问为2.26220794882991，第八问为1.67887576144059，第九问为2.14449014609801。然后本文求的点估计的方差，公式为：

其中表示访查的估计值，表示求得的方差，代表回答这份问卷的总人数，代表上述求得的算术平均值，代表第*i*个人的选项对应的值。求得的结果第七问为1.62761856535801，第八问为1.33475670803274，第九问为1.57460238780043。

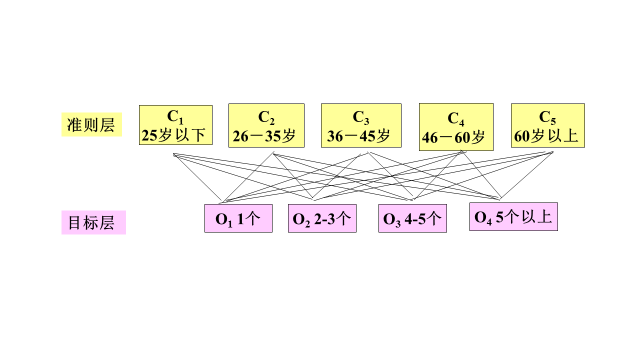
本文在点估计的基础上利用区间估计，公式如下：

其中，表示已有数据总体的待估参数，表示概率，表示置信下限，表示置信上限，表示可信度且满足。因此有

利用Matlab程序，本文将可信度设为0.95，求得平均值的置信区间第七问为[2.26143280891573,2.26298308874409]，方差[1.6271651117502, 1.62826182083831]。第八问平均值的置信区间[1.60296275457298, 1.60891124875577]，方差[1.30166457003991, 1.30590410543708]。第九问平均值的置信区间[1.67820983224883, 1.67954169063235]，方差[1.33436712094773, 1.33530931268636]。

* 1. **问题二**
  2. **问题三**
     1. **层次分析法**

为了在各个因子中进行选择，判断在某个特定性质的人中更可能产生多少垃圾，本文利用层次分析法来实现在复杂的不确定的问题中找到各项权重。本文将人的特质，如年龄，性别，收入，教育程度，所在公司等作为准则层。产生多少垃圾作为目标层，进行仅有一层但有多组的的层次分析。图中表示了使用多少餐盒对年龄的层次分析示意图。



首先，本文将某一个性质的人产生了多少垃圾的人数，即在某一准则的条件下选择某一目标选项的个数作为这一项的绝对权重，记为。则依据这个选项本文得到一个权向量(为目标层选项的个数)。将这一个性质的人选择产生垃圾相同的人数，互相作比的比值作为成对比较阵的权重，得到成对比较阵。因为其为一致阵，因此本文不需要进行一致性检验，其必然一致。利用特征根和特征向量公式

可得其特征向量，将各个准则的特征向量组合则可以得到同一问题下目标对准则的特征向量阵，及权向量阵。

* + 1. **线性回归法**

本文使用的第二种方法是线性回归法。可以将餐盒使用情况作为因变量，各个人的基本信息作为自变量。基于本文调查得到的样本，每一条数据均可以视为由自变量基本信息到因变量餐盒使用情况的映射。由于每一条数据自变量和因变量都是数值，可以根据已有的数据进行学习，通过线性回归得到自变量到各个因变量的函数关系。

本文记性别为自变量，年龄为自变量，教育程度为自变量，年收入为自变量。塑料盒使用情况为因变量，塑料袋使用情况为因变量，纸餐盒使用情况为因变量，餐具使用情况为因变量，餐食剩余情况为因变量，纸巾剩余情况为因变量。自变量和因变量的数值即为各选项的值。

本文利用回归公式

记自变量矩阵为，因变量矩阵为，系数矩阵为，采用最小二乘回归估计方法，公式如下：

进行对系数矩阵点估计的求解。通过Matlab程序计算得到的系数矩阵如表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 塑料盒 | 塑料袋 | 纸盒 | 餐具 | 餐食 | 纸巾 |
|  | 1.378606 | 1.18038 | 1.025224 | 1.001631 | 1.477334 | 1.539389 |
|  | 0.180875 | 0.055576 | 0.008579 | 0.07446 | -0.0514 | 0.065162 |
|  | 0.033322 | 0.034247 | 0.006007 | 0.057082 | 0.051542 | -0.01299 |
|  | 0.059538 | 0.01823 | -0.00986 | -0.03456 | -0.07101 | -0.00494 |
|  | 0.018077 | 0.002216 | 0.014431 | 0.021263 | 0.011921 | 0.005794 |

点估计有一个不足之处是它无法表示所得到的系数的准确度。因此本文利用了区间估计再次进行最小二乘回归。公式如下：

其中，表示已有数据总体的待估参数，表示概率，表示置信下限，表示置信上限，表示可信度且满足。因此有

利用Matlab程序，本文将可信度设为0.95，求得回归系数上下限如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 塑料盒下限 | 塑料盒上限 | 塑料袋下限 | 塑料袋上限 | 纸盒下限 | 纸盒上限 |
|  | 1.132305 | 1.624907 | 0.960902 | 1.399858 | 0.910536 | 1.139912 |
|  | 0.099211 | 0.262539 | -0.01719 | 0.128347 | -0.02945 | 0.046606 |
|  | -0.00504 | 0.071684 | 6.27E-05 | 0.068431 | -0.01186 | 0.02387 |
|  | 0.011431 | 0.107645 | -0.02464 | 0.061098 | -0.03226 | 0.012545 |
|  | -0.01256 | 0.048717 | -0.02509 | 0.02952 | 0.000163 | 0.028698 |
|  | 餐具下限 | 餐具上限 | 餐食下限 | 餐食上限 | 纸巾下限 | 纸巾上限 |
|  | 0.84703 | 1.156233 | 1.301779 | 1.652889 | 1.358263 | 1.720516 |
|  | 0.0232 | 0.12572 | -0.10961 | 0.006809 | 0.005108 | 0.125217 |
|  | 0.033003 | 0.081162 | 0.024199 | 0.078885 | -0.0412 | 0.015225 |
|  | -0.06476 | -0.00436 | -0.1053 | -0.03672 | -0.04032 | 0.030438 |
|  | 0.00203 | 0.040496 | -0.00992 | 0.03376 | -0.01674 | 0.028327 |

作出的残差图如图。当在检查相关系数时，本文发现相关系数如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 塑料盒 | 塑料袋 | 纸盒 | 餐具 | 餐食 | 纸巾 |
| 0.027248 | 0.007408 | 0.0043 | 0.032196 | 0.026184 | 0.004944 |

由于相关系数差距过大，因此本文尝试基于主成分回归进行分析。

* + 1. **主成分回归法**

本文依然可以像第二种方法一样将餐盒使用情况作为因变量，各个人的基本信息作为自变量。本文试图通过降维，将较多的原始数据信息和变量减少为较少的数据信息和较少变量，而新的变量可以大体上反映原变量的信息。

本文仍然记自变量矩阵为，因变量矩阵为。原变量为，新变量为，记样本的个数为，每个样本的变量数目为，则数据矩阵为：

由于数据的量纲差异较大，范围差异较大，本文需要对数据进行标准化处理。本文采用的是方差标准化法使得标准化后的数据方差为1，同时进行中心平移使得平均数为0。公式如下：

其中表示第行第列标准化后的数据，表示第行第列标准化前的数据，表示共有几列，表示共有几行。

其次，本文建立自变量的相关系数阵，计算公式和求得的相关系数阵如下：



|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 0.106472 | -0.02209 | -0.18079 |
| 0.106472 | 1 | 0.131604 | -0.16217 |
| -0.02209 | 0.131604 | 1 | 0.029231 |
| -0.18079 | -0.16217 | 0.029231 | 1 |

然后，本文求矩阵的特征根（对都满足）和特征向量，以确定与相关矩阵的个较大的特征值所对应的特征向量相等的原变量在各个主成分新变量上的荷载。是特征向量的第个值。公式如下：

其中表示各个特征向量，表示各个特征根，求得的特征根矩阵如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 0.773125 | 0.840018 | 1.079256 | 1.307601 |

特征向量矩阵如下：（每列为一个特征向量）

|  |  |  |  |
| --- | --- | --- | --- |
| -0.17622 | 0.755247 | -0.32618 | -0.54051 |
| -0.61883 | -0.38611 | 0.381381 | -0.5679 |
| 0.423931 | 0.349138 | 0.822693 | -0.14684 |
| -0.6374 | 0.398274 | 0.267078 | 0.603136 |

利用贡献率公式和累计贡献率公式





求得直到第三主成份累计贡献率为0.806719，大于0.8，因此取前三个特征值作为主成分。设主成分为：

其中为

|  |  |  |  |
| --- | --- | --- | --- |
| -0.54051 | -0.5679 | -0.14684 | 0.603136 |
| -0.32618 | 0.381381 | 0.822693 | 0.267078 |
| 0.755247 | -0.38611 | 0.349138 | 0.398274 |

利用上述最小二乘回归，点估计和区间估计方法，由回归公式

得到系数矩阵如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 塑料盒 | 塑料袋 | 纸盒 | 餐具 | 餐食 | 纸巾 |
|  | -0.20454 | -0.16981 | -0.1595 | -0.22205 | -0.17519 | -0.17574 |
|  | 0.268683 | 0.244276 | 0.354158 | 0.269743 | 0.315164 | 0.291636 |
|  | 0.400675 | 0.292137 | 0.467573 | 0.369267 | 0.323962 | 0.476837 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 塑料盒下限 | 塑料盒上限 | 塑料袋下限 | 塑料袋上限 | 纸盒下限 | 纸盒上限 |
|  | -0.25638 | -0.1527 | -0.22186 | -0.11776 | -0.21677 | -0.10224 |
|  | 0.220762 | 0.316604 | 0.19616 | 0.292392 | 0.301224 | 0.407092 |
|  | 0.350522 | 0.450828 | 0.241779 | 0.342494 | 0.412173 | 0.522973 |
|  | 餐具下限 | 餐具上限 | 餐食下限 | 餐食上限 | 纸巾下限 | 纸巾上限 |
|  | -0.27623 | -0.16787 | -0.23183 | -0.11854 | -0.2317 | -0.11978 |
|  | 0.219655 | 0.31983 | 0.262799 | 0.367528 | 0.239903 | 0.34337 |
|  | 0.316847 | 0.421688 | 0.269158 | 0.378765 | 0.422693 | 0.53098 |

最后进行逆标准化变换后得到的用原变量表示的回归方程

得到系数矩阵：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 塑料盒下限 | 塑料盒上限 | 塑料袋下限 | 塑料袋上限 | 纸盒下限 | 纸盒上限 |
|  | 1.992579 | 1.992579 | 1.423933 | 1.423933 | 1.069573 | 1.069573 |
|  | 0.448787 | 0.433149 | 0.285051 | 0.271199 | 0.20587 | 0.19792 |
|  | 0.059731 | 0.02112 | 0.059945 | 0.025746 | 0.022947 | 0.003317 |
|  | 0.275291 | 0.354778 | 0.197871 | 0.268274 | 0.157076 | 0.197486 |
|  | 0.022146 | 0.086706 | 0.006613 | 0.063796 | 0.026417 | 0.059239 |
|  | 餐具下限 | 餐具上限 | 餐食下限 | 餐食上限 | 纸巾下限 | 纸巾上限 |
|  | 1.231911 | 1.231911 | 1.35436 | 1.35436 | 1.607607 | 1.607607 |
|  | 0.270195 | 0.259909 | 0.234369 | 0.222196 | 0.360714 | 0.348438 |
|  | 0.047083 | 0.021686 | 0.057646 | 0.02759 | 0.027531 | -0.00278 |
|  | 0.168294 | 0.220576 | 0.197587 | 0.259462 | 0.222029 | 0.284422 |
|  | 0.005789 | 0.048253 | 0.013488 | 0.063743 | 0.033965 | 0.084641 |

* + 1. **距离判别法**

本文尝试依据已有的数据，随机抽样作为学习样本，作为测试集，将庞大的样本在数目上高度归并，找出各个样本之间公有的信息和特征，从而得出类似的废弃物情况时人群的共同特征从而确定关系。

由于本文调查得到的数据在自变量相同时对应的因变量却不同，但判别时需要各个样本不同，因此本文将相同自变量时的因变量取代数平均值，作为自变量时因变量的代表。对于塑料盒，塑料袋，纸盒使用数量这三个问题将选项的值转化为问题的数量，如有几个塑料盒，进行求数量平均值。对于餐具，餐食，纸巾这三个问题有剩余则赋0，无剩余则赋1.当不存在这样的自变量组合时因变量补0。由于取得的代数平均值并不为整数，因此本文将求得的代数平均值对更加接近的选项的值进行四舍五入，再次转化为原始选项的值。转化出来的选项如表所示。然后本文将原始自变量组合对应的因变量的值用转化后的值替代，用以接下来进一步的分析。如表所示。

利用马氏距离判别法，公式如下：



其中，均表示两个行向量，表示协方差矩阵，表示求得的数据的马氏距离。。

本文随机抽取的判别后的结果与最初的调查结果比对如下表。判别准确率如下表。由于各个问题的判别准确率均较低，不足以准确的反应各个变量之间的特征，因此本文考虑采用另外的方法。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 塑料盒 | 塑料袋 | 纸餐盒 | 餐具 | 餐食 | 纸巾 |
| 0.385042 | 0.411111 | 0.5528 | 0.7167 | 0.6 | 0.5583 |

* + 1. **贝叶斯判别法**

在以上的距离判别方法中，没有考虑到各个样本在总体中出现的频率多少，也没有考虑到误判带来的损失。而贝叶斯判别法在距离判别的基础上进行了修正，公式如下：



其中表示后验概率，表示先验概率，表示样本出现的频率，Σ表示总体的协方差矩阵。判别规则有后验概率最大原则和平均错判损失最小，判别规则如下：若满足：

C:\Users\amin\Desktop\1.png

则将判给，其中为总体，为总体的概率密度函数，为的先验概率，即样本发生时它属于某类的概率。为总体的个数。判别分析的解公式如下：





其中用表示将来自总体的样品错判到总体的条件概率，为这种错判所造成的损失。为待判样本的一个分划。是平均错判损失。使最小的一组解是贝叶斯判别分析的解。

利用MATLAB程序，本文依然随机抽样作为学习样本，作为测试集，进行贝叶斯判别求解。由于按照上一节所述方法转换过后的数据不存在选项4，因此选项4的条件概率为0。计算后发现除了纸盒的问题存在多种判别结果外，其余5个问题仅存在一种判别结果，不能从中分类归纳得出结论，因此本文将原始数据与判别结果相同的去掉，再次进行判别分析。

在剩余餐纸的情况当中，表中显示了判别结果为1，也就是有剩余餐纸扔掉的选项，与人群信息的关系。依据表中信息可以得出男性较女性更通常将剩余的餐纸扔掉。可以得出年龄低的人群更易于将餐纸扔掉而成为外卖垃圾。可以得出学历较低的人群更加会把餐纸扔掉。可以得出尚未工作和高收入人群更倾向于扔掉餐纸。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 性别 | 判别结果分类 | 总计 | 学历 | 判别结果分类 | 总计 |
| 男 | 102 | 434 | 高中及以下 | 31 | 61 |
| 女 | 24 | 519 | 大专 | 5 | 77 |
| 年龄 | 判别结果分类 | 总计 | 本科 | 55 | 463 |
| 25岁以下 | 88 | 128 | 研究生及以上 | 35 | 352 |
| 26－35岁 | 37 | 222 | 年收入 | 判别结果分类 | 总计 |
| 36－45岁 | 1 | 330 | 0-1万元 | 0 | 70 |
| 46－60岁 | 0 | 236 | 1万元-10万元 | 0 | 259 |
| 60岁以上 | 0 | 37 | 10万元到20万元 | 12 | 311 |
|  |  | | 20万元到35万元 | 21 | 129 |
| 35万元以上 | 18 | 100 |
| 还没工作 | 75 | 84 |

在进行第二次贝叶斯判别以后，发现对于第九问由于选项过于极重，去掉后样本量过小，因此无法进行分析。对于第十，第十一两问因为选项只有两个，去掉后选项划分，即需要判别的类别仅剩一类，因此无法进一步进行判别分析。对于有多个选项的第七、八个问题第七问获得了0.7766的回判准确率，第八问获得了0.9405的回判准确率。第二次贝叶斯判别结果如表所示。将两次判别的结果综合如表所示，第一个表是第七问选项为1，也就是经过转换后通常每次外卖使用1个塑料盒的人的特征。第二个表是第七问选项为3，也就是经过转换后通常每次外卖使用4-5个塑料盒的人的特征。第三个表是第八问选项为2，也就是经过转换后通常每次外卖使用2-3个塑料袋的人的特征。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 性别 | 判别结果分类 | 总计 | 比率 | 学历 | 判别结果分类 | 总计 | 比率 |
| 男 | 57 | 434 | 0.131336 | 高中及以下 | 3 | 61 | 0.04918 |
| 女 | 15 | 519 | 0.028902 | 大专 | 12 | 77 | 0.155844 |
| 年龄 | 判别结果分类 | 总计 | 比率 | 本科 | 15 | 463 | 0.032397 |
| 25岁以下 | 9 | 128 | 0.070313 | 研究生及以上 | 41 | 352 | 0.116477 |
| 26－35岁 | 22 | 222 | 0.099099 | 年收入 | 判别结果分类 | 总计 | 比率 |
| 36－45岁 | 27 | 330 | 0.081818 | 0-1万元 | 8 | 70 | 0.114286 |
| 46－60岁 | 8 | 236 | 0.033898 | 1万元-10万元 | 7 | 259 | 0.027027 |
| 60岁以上 | 6 | 37 | 0.162162 | 10万元到20万元 | 10 | 311 | 0.032154 |
|  | | | | 20万元到35万元 | 10 | 129 | 0.077519 |
| 35万元以上 | 25 | 100 | 0.25 |
| 还没工作 | 12 | 84 | 0.142857 |

分析如下：在使用塑料盒较少的选项中，男性占比远多于女性。在老年人中使用的餐盒较少。同时在学历较高的人中占比较高，使用的餐盒较少。在尚未工作或者年收入高的人群中使用的餐盒较少。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 性别 | 判别结果分类 | 总计 | 比率 | 学历 | 判别结果分类 | 总计 | 比率 |
| 男 | 7 | 434 | 0.016129 | 高中及以下 | 4 | 61 | 0.065574 |
| 女 | 15 | 519 | 0.028902 | 大专 | 11 | 77 | 0.142857 |
| 年龄 | 判别结果分类 | 总计 | 比率 | 本科 | 7 | 463 | 0.015119 |
| 25岁以下 | 9 | 128 | 0.070313 | 研究生及以上 | 0 | 352 | 0 |
| 26－35岁 | 1 | 222 | 0.004505 | 年收入 | 判别结果分类 | 总计 | 比率 |
| 36－45岁 | 2 | 330 | 0.006061 | 0-1万元 | 8 | 70 | 0.114286 |
| 46－60岁 | 3 | 236 | 0.012712 | 1万元-10万元 | 11 | 259 | 0.042471 |
| 60岁以上 | 7 | 37 | 0.189189 | 10万元到20万元 | 3 | 311 | 0.009646 |
|  | | | | 20万元到35万元 | 0 | 129 | 0 |
| 35万元以上 | 0 | 100 | 0 |
| 还没工作 | 0 | 84 | 0 |

在使用塑料盒较多的选项中，女性占比远高于男性，年龄大的人使用的较多，学历和收入较低的人占比较高，而学历高或者收入高的人的人则极少使用大量餐盒。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 性别 | 判别结果分类 | 总计 | 比率 | 学历 | 判别结果分类 | 总计 | 比率 |
| 男 | 56 | 434 | 0.129032 | 高中及以下 | 6 | 61 | 0.098361 |
| 女 | 103 | 519 | 0.198459 | 大专 | 34 | 77 | 0.441558 |
| 年龄 | 判别结果分类 | 总计 | 比率 | 本科 | 73 | 463 | 0.157667 |
| 25岁以下 | 11 | 128 | 0.085938 | 研究生及以上 | 46 | 352 | 0.130682 |
| 26－35岁 | 21 | 222 | 0.094595 | 年收入 | 判别结果分类 | 总计 | 比率 |
| 36－45岁 | 63 | 330 | 0.190909 | 0-1万元 | 16 | 70 | 0.228571 |
| 46－60岁 | 56 | 236 | 0.237288 | 1万元-10万元 | 34 | 259 | 0.131274 |
| 60岁以上 | 8 | 37 | 0.216216 | 10万元到20万元 | 52 | 311 | 0.167203 |
|  | | | | 20万元到35万元 | 45 | 129 | 0.348837 |
| 35万元以上 | 10 | 100 | 0.1 |
| 还没工作 | 2 | 84 | 0.02381 |

在使用塑料盒较多的选项中，女性占比远高于男性，年龄大的人使用的较多，学历和收入较低的人占比较高，而学历高或者收入高的人的人则较少使用大量塑料袋。

综上所述，综合各个结果可以发现，通常女性产生的外卖固体废弃物比男性多，年龄在青年人中产生的外卖固体废弃物较多，而年老或者未成年人的人则很少。学历和收入较低的人产生的外卖固体废弃物较多产生的频率更加高，因此学历和收入较低的人产生的外卖固体废弃物较多，在高学历高收入人群中产生的外卖固体废弃物则较少。

* + 1. **支持向量机**

支持向量机（SVM）是一种在解决非线性问题中有极大优势的算法，其将低维数据映射到高维空间中并找到最优超平面作为将给定数据分类的标准。

本文使用MATLAB和LIBSVM-3.22[Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm]进行了支持向量机的计算，其示意图[https://www.cnblogs.com/spoorer/p/6551220.html]如下：



本文从所有数据中抽取作为测试数据，作为训练数据训练出一个SVM模型并通过计算计算已知样本点的错判率和在训练集上的错误率而评估出模型的切合程度。

使用的MATLAB代码见附录。最后计算出模型在训练集上的准确度为，在测试集上的错误率 是。此处列举了模型认为分类正确概率最大的10组数据：



由于有时会出现前四题作答完全一样但是第五题选择内容不一样的数据，因此这时模型只能给出一个可能的选择；同样，这些数据也干扰了模型的学习，因而无法取得大于正确率的SVM模型，这也是本方法的局限所在。

最终观察各组数据，得出结论：训练模型显示，拥有大专及以上学历的26至45岁之间的女性在外卖上产生的垃圾比拥有高中学历或其它年龄段女性产生的更多；46至60岁拥有研究生及以上学历的男性订外卖时，单次产生的垃圾比其余男性单次产生的垃圾多。

1. **模型优化**
2. **结论**
   1. **敏感性分析**

* 1. **优缺点分析**
  2. **结论**
  3. **问题四**

1. **参考文献**
2. **附录**

load('data.txt');

data\_sorted = data(1:718, :);

test\_data = data(719:end, :);

[yy, ps] = mapstd(data\_sorted);

dd = mapstd(test\_data);

load('types.txt');

s = svmtrain(yy, types(1:718, :));

sv\_index = s.SupportVectorIndices;

al = s.Alpha;

bb = s.Bias;

mean\_and\_std\_trans = s.ScaleData;

check = svmclassify(s, yy);

err\_rate = 1 - sum(types == check) / length(types);

solution = svmclassify(s, dd);



|  |  |
| --- | --- |
| **官方网站** | www.dengfengbei.com |
| **微信公众号** | **Dengfengbeijingsai**  C:\Users\ZHANGCHAO\AppData\Local\Microsoft\Windows\INetCacheContent.Word\登峰杯微信号二维码.jpg |
| **官方QQ群** | （1）“登峰杯”学术作品学生QQ群 571526693  （2）“登峰杯”数学建模学生QQ群 571535826  （3）“登峰杯”机器人学生QQ群 571540979  （4）“登峰杯”结构设计学生QQ群 592858677  （5）“登峰杯”数据挖掘学生QQ群 144821810  （6）“登峰杯”艺术创意设计学生QQ群 318850726 |
| **官方邮箱** | dengfengbei@126.com |
| **联系电话** | 010-52909593，18310079788  （工作日9:00~12:00，13:00~17:00） |