目标客户的统计信息以离散型随机变量为主，为了研究其人口学特征与购买的关系，这里采用卡方独立性检验的方法。其中诸如年龄、工龄等特征为了方便处理使用划分区间的方法进一步做离散化处理。

卡方检验是一种不要求总体正态性的非参数检验，其中统计量刻画了两个分类变量的相关程度，统计量越大，表明观察值频数与期望值频数的差别越大，进而表明行列两个分类变量之间越有可能具有相关性。

本题中提出原假设：客户的身份信息与是否购买无关；备择假设：客户的身份信息与是否购买无关。事先指定犯错误的概率，则各变量卡方值如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 户籍情况（B） | 居住时间（B2） | 居住区域（B3） | 驾龄（B4） | 家庭人口数（B5） |
| 卡方值 | 2.195 | 20.573 | 16.097 | 12.346 | 3.752 |
| p值 | 0.334 | 0.015 | 0.007 | 0.194 | 0.586 |
|  | 婚姻家庭情况（B6） | 育儿数（B7） | 年龄（B8） | 最高学历（B9） | 工作年限（B10） |
| 卡方值 | 3.525 | 0.883 | 4.399 | 6.937 | 6.207 |
| p值 | 0.833 | 0.830 | 0.733 | 0.139 | 0.516 |
|  | 单位性质（B11） | 职位（B12） | 家庭年收入（B13） | 个人年收入（B14） | 可支配年收入（B15） |
| 卡方值 | 22.775 | 19.538 | 22.303 | 13.516 | 20.750 |
| p值 | 0.004 | 0.034 | 0.008 | 0.095 | 0.004 |
|  | 房贷支出占比（B16） | 车贷支出占比（B17） | |  |  |
| 卡方值 | 82.424 | 35.075 |  |  |  |
| p值 | 2.61E-16 | 1.45E-06 |  |  |  |

将下列结果汇总如下图所示，则由此初步分析知收支状况、借贷情况、居住地带、工作单位性质及职业等属性较明显地与客户是否愿意购买电动汽车相关。其中贷款状况强烈影响客户的购买意愿，偿还房贷支出占家庭总收入超10%的客户群体中仅有1%的客户有意愿购买电动汽车，而这一支出占比超20%的家庭购买意愿降低到0%，偿还车贷的群体也有着类似的结论；对于不同收入水平的家庭而言，整体呈现中等收入家庭（个人）的购买意愿更强，这一结论也是符合常识的；居住地带上，市中心的客户群体更愿意购买电动汽车，这一情况一定程度上与城市中相关设施建设落实相关；而在工作单位与个人职业属性上，整体呈现出从事个体工商业群体、自由职业的群体及公司职员相对更愿意购买电动汽车，这一情况一定程度上与其收入水平及工作内容及电动车的平均售价有关。而部分特征与客户是否有意愿购买的相关度不高，如客户的婚姻家庭信息、年龄与车龄信息等

