### 分析流程 数据源： data口腔健康调查表+评价.xlsx 算法配置： 算法： 卡方检验（自动选取最优求解器） 变量： 分组变量X:{您第一次去口腔医院是因为}；变量Y:{您在口腔上每年的支出约为} 分析结果： 卡方检验是分析两分类变量是否存在显著性差异：系统智能选择Yates校正卡方检验，显著性P值为0.019\*\*，水平上呈现显著性，拒绝原假设，因此您第一次去口腔医院是因为和您在口腔上每年的支出约为数据存在显著性差异。

### 分析步骤 1. 根据列联表的数据情况，选择适合的卡方检验（包括Pearson卡方检验、Yates校正卡方检验、Fisher精确检验），分析卡方检验是否呈现显著性（P<0.05，呈现显著性，拒绝原假设，则说明分类变量X与分类变量Y之间存在显著性差异）。 2. 若是Pearson卡方检验、Yates校正卡方检验呈现显著性，可接着根据效应指标对差异进行深入量化分析。

### 详细结论

**输出结果1：卡方检验结果**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 题目 | 名称 | 您第一次去口腔医院是因为 | | | | | | 合计 | X² | P |
| 亲戚建议 | 体检检查 | 其他医生建议 | 口腔疾病 | 朋友建议 | 没去过口腔医院 |
| 您在口腔上每年的支出约为 | 小于1000 | 9 | 57 | 5 | 126 | 6 | 62 | 265 | 41.855 | 0.019\*\* |
| 1000~3500 | 2 | 5 | 0 | 38 | 2 | 2 | 49 |
| 3500~5000 | 0 | 0 | 0 | 5 | 0 | 0 | 5 |
| 5000~7000 | 1 | 2 | 0 | 5 | 1 | 0 | 9 |
| 7000~9000 | 0 | 0 | 0 | 4 | 0 | 0 | 4 |
| 9000以上 | 1 | 1 | 0 | 0 | 0 | 2 | 4 |
| 合计 | | 13 | 65 | 5 | 178 | 9 | 66 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | | | |

**图表说明：**

上表展示了模型检验的结果，包括数据的频数、频数百分比、卡方值、显著性P值。若P<0.05，呈现显著性，拒绝原假设，则说明分类变量X与分类变量Y之间存在显著性差异。  
卡方检验方法的选取有以下规则：  
● 针对2×2列联表（R=2，C=2，即在列联表中行R、列C都只有两个分类水平，比如性别只有男、女两个分类水平）：可选择使用Pearson卡方检验、Yates校正卡方检验、fisher精确检验。  
1.所有的单元格理论数T≥5并且总样本量n≥40，用Pearson卡方进行检验。  
2.如果存在单元格理论数1 <=T<5，并且总样本量n≥40,用Yates校正卡方进行检验。  
3.如果存在单元格理论数T＜1或总样本量n＜40，则用Fisher精确检验。  
● 针对R×C列联表(R>2或C>2)：可选择使用Pearson卡方检验、Yates校正卡方检验。  
1.全部单元格理论数T>=1 且 1 <=T<5单元格的比例小于20% ，则使用Pearson卡方。  
2.若不能达到Pearson卡方检验的使用要求，则使用Yates校正卡方检验。

**智能分析：**

根据列联表的数据，系统智能选择Yates校正卡方检验，显著性P值为0.019\*\*，水平上呈现显著性，拒绝原假设，因此您第一次去口腔医院是因为和您在口腔上每年的支出约为数据存在显著性差异。

**输出结果2：交叉列联表热力图**



**图表说明：**

上图展示了热力图的形式展示了交叉列联表的值，主要通过颜色深浅去表示值的大小。

**输出结果3：效应量化分析**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名/分析项 | Phi | Crammer‘s V | 列联系数 | lambda |
| 您在口腔上每年的支出约为---您第一次去口腔医院是因为 | 0.353 | 0.158 | 0.333 | 0 |

**图表说明：**

上表展示了效应量化分析的结果，包括phi、Crammer's V、列联系数、lambda ，用于分析样本的相关程度。  
1. 当呈现出显著性差异（前提），结合分析效应量指标对差异性进行量化分析。  
2. 效应量化指标反映的是变量之间的相关程度。  
3. 根据交叉类型的不同，可以选用不同的效应量指标（交叉类型表示：交叉表横向格子数×纵向格子数）。  
4. phi系数：phi相关系数的大小，表示两样本之间的关联程度。当phi系数小于0.3时，表示相关较弱；当phi系数大于0.6时，表示相关较强（用于2×2交叉类型表）。  
5. Cramer's V：与phi系数作用相似，但Cramer's V系数的作用范围较广。当两个变量相互独立时，V=0，当数据中只有2个二分类变量时，Cramer's V系数的结果与phi相同（若m≠n，建议使用Cramer's V ）。  
6. 列联系数：简称C系数，用于3×3或4×4交叉表，但其受行列数的影响，随着R和C 的增大而增大。因此根据不同的行列和计算的列联系数不便于比较，除非两个列联表中行数和列数一致。  
7. lambda：用于反应自变量对因变量的预测效果，一般情况下，其值为1时表示自变量预测因变量效果较好，为0时表明自变量预测因变量较差（X或Y有定序数据时，建议使用lambda）。

**智能分析：**

效应量化分析的结果显示，分析项：您在口腔上每年的支出约为Cramer's V值为0.158，因此您第一次去口腔医院是因为和您在口腔上每年的支出约为的差异程度为弱程度差异。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 杨倩. 基于信息恒等式的属性数据分析[D]. 厦门大学, 2012.