分析流程  
  
数据源：  
data评价.xlsx  
  
算法配置：  
算法： 正态性校验  
变量： 自变量X:{您对当下针对口腔健康工作的评分}  
  
分析结果：  
正态性检验基于S-W检验或K-S检验得到结果： 变量分析项：您对当下针对口腔健康工作的评分样本N < 5000，采用S-W检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。

分析步骤  
  
1. 对数据进行Shapiro-Wilk（小数据样本，一般样本数5000以下）或者Kolmogorov–Smirnov（大数据样本，一般样本数5000以上）检验，查看其显著性。  
2. 若不呈现出显著性(P>0.05)，说明符合正态分布，反之说明不符合正态分布（PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布）。

详细结论

**输出结果1：总体描述结果**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量名 | 样本量 | 中位数 | 平均值 | 标准差 | 偏度 | 峰度 | S-W检验 | K-S检验 |
| 您对当下针对口腔健康工作的评分 | 336 | 4 | 3.911 | 1.015 | -0.44 | -0.657 | 0.835(0.000\*\*\*) | 0.239(0.000\*\*\*) |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | |

**图表说明：**

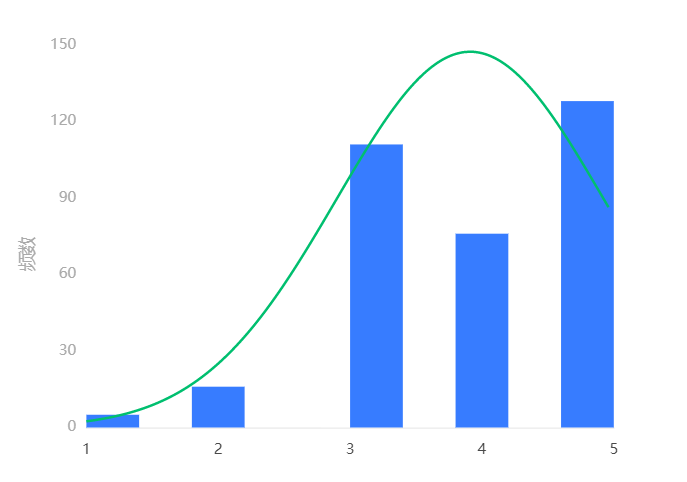
上表展示了您对当下针对口腔健康工作的评分描述性统计和正态性检验的结果，包括中位数、平均值等，用于检验数据的正态性。  
1. 通常正态分布的检验方法有两种，一种是Shapiro-Wilk检验，适用于小样本资料（样本量≤5000）；另一种是Kolmogorov–Smirnov检验，适用于大样本资料（样本量>5000）。  
2. 若呈现显著性(P<0.05)，则说明拒绝原假设（数据符合正态分布），该数据不满足正态分布，反之则说明该数据满足正态分布。  
PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布。

**智能分析**

分析项：您对当下针对口腔健康工作的评分样本N < 5000，采用S-W检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（-0.657）绝对值小于10并且偏度（-0.44）绝对值小于3，可以结合正态分布直方图、PP图或者QQ图进行进一步分析。)

**输出结果2：正态性检验直方图**

您对当下针对口腔健康工作的评分

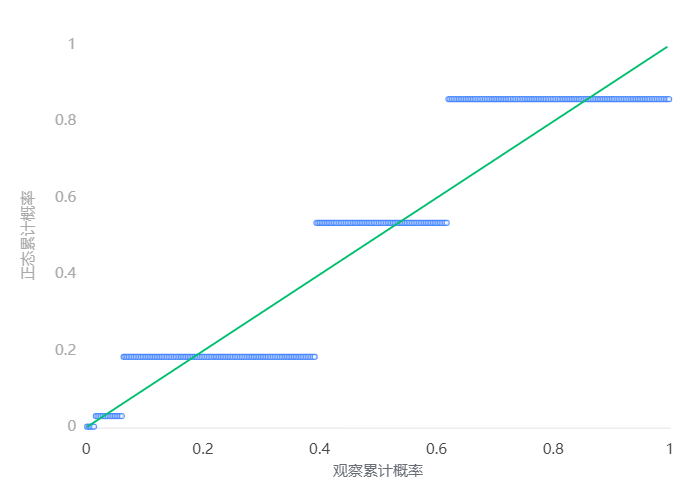


**图表说明：**

上图展示了您对当下针对口腔健康工作的评分数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

**输出结果3：正态性检验P-P图**

您对当下针对口腔健康工作的评分

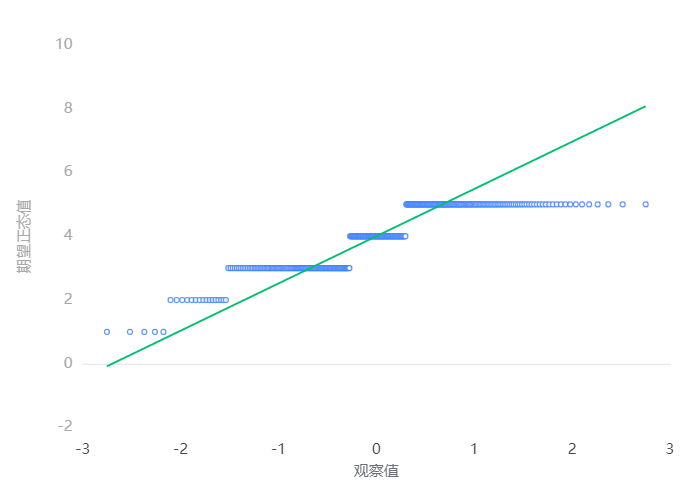


**图表说明：**

上图是您对当下针对口腔健康工作的评分计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

**输出结果4：正态性检验Q-Q图**

您对当下针对口腔健康工作的评分



**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

分析流程  
  
数据源：  
data评价.xlsx  
  
算法配置：  
算法： 正态性校验  
变量： 自变量X:{您是否期待针对口腔健康的小程序诞生}  
  
分析结果：  
正态性检验基于S-W检验或K-S检验得到结果： 变量分析项：您是否期待针对口腔健康的小程序诞生样本N < 5000，采用S-W检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。

分析步骤  
  
1. 对数据进行Shapiro-Wilk（小数据样本，一般样本数5000以下）或者Kolmogorov–Smirnov（大数据样本，一般样本数5000以上）检验，查看其显著性。  
2. 若不呈现出显著性(P>0.05)，说明符合正态分布，反之说明不符合正态分布（PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布）。

详细结论

**输出结果1：总体描述结果**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量名 | 样本量 | 中位数 | 平均值 | 标准差 | 偏度 | 峰度 | S-W检验 | K-S检验 |
| 您是否期待针对口腔健康的小程序诞生 | 336 | 5 | 4.476 | 0.894 | -1.894 | 3.503 | 0.636(0.000\*\*\*) | 0.397(0.000\*\*\*) |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | |

**图表说明：**

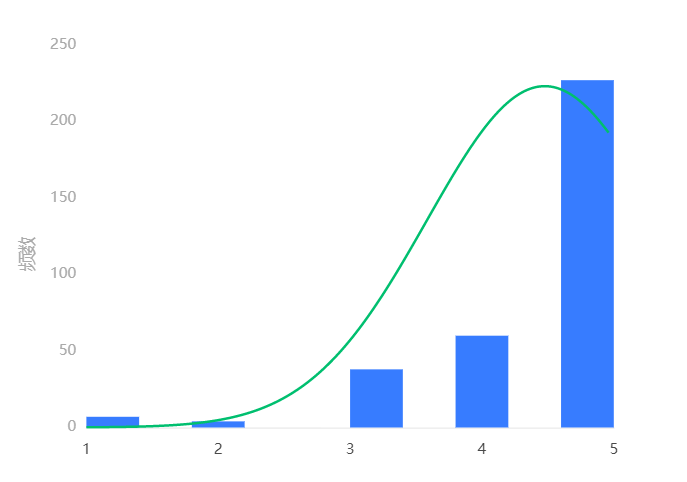
上表展示了您是否期待针对口腔健康的小程序诞生描述性统计和正态性检验的结果，包括中位数、平均值等，用于检验数据的正态性。  
1. 通常正态分布的检验方法有两种，一种是Shapiro-Wilk检验，适用于小样本资料（样本量≤5000）；另一种是Kolmogorov–Smirnov检验，适用于大样本资料（样本量>5000）。  
2. 若呈现显著性(P<0.05)，则说明拒绝原假设（数据符合正态分布），该数据不满足正态分布，反之则说明该数据满足正态分布。  
PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布。

**智能分析**

分析项：您是否期待针对口腔健康的小程序诞生样本N < 5000，采用S-W检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（3.503）绝对值小于10并且偏度（-1.894）绝对值小于3，可以结合正态分布直方图、PP图或者QQ图进行进一步分析。)

**输出结果2：正态性检验直方图**

您是否期待针对口腔健康的小程序诞生

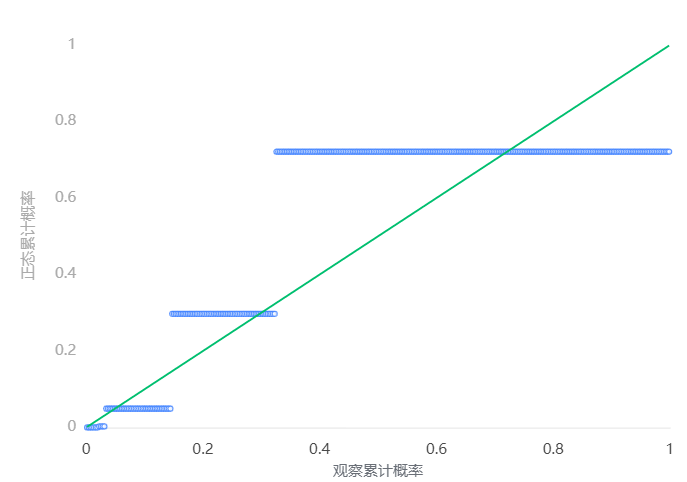


**图表说明：**

上图展示了您是否期待针对口腔健康的小程序诞生数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

**输出结果3：正态性检验P-P图**

您是否期待针对口腔健康的小程序诞生

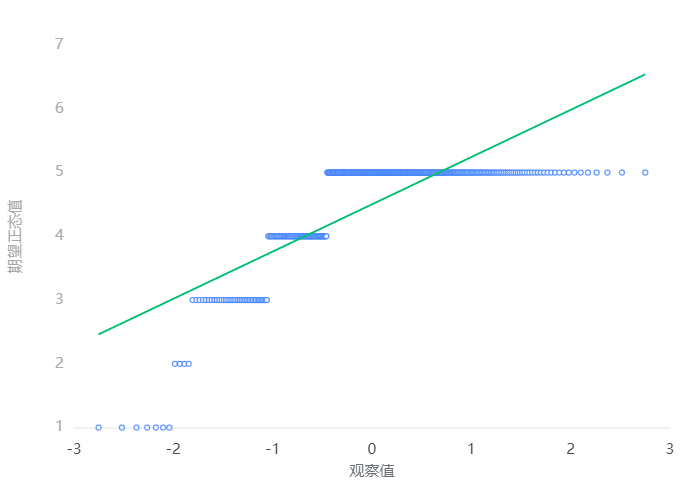


**图表说明：**

上图是您是否期待针对口腔健康的小程序诞生计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

**输出结果4：正态性检验Q-Q图**

您是否期待针对口腔健康的小程序诞生



**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

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。

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:{您在口腔上每年的支出}；变量Y:{您的受教育水平}  
  
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0.799，水平上不呈现显著性，不能拒绝原假设，因此您在口腔上每年的支出和您的受教育水平不存在显著性差异。

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 | 您在口腔上每年的支出约为 | | | | | | 合计 | X² | P |
| 小于1000 | 1000~3500 | 3500~5000 | 5000~7000 | 7000~9000 | 9000以上 |
| 您的受教育水平是 | 小学及以下 | 49 | 8 | 1 | 1 | 0 | 0 | 59 | 23.377 | 0.799 |
| 初中 | 162 | 33 | 2 | 7 | 4 | 4 | 212 |
| 高中 | 35 | 4 | 1 | 1 | 0 | 0 | 41 |
| 大专 | 9 | 1 | 0 | 0 | 0 | 0 | 10 |
| 本科 | 1 | 2 | 0 | 0 | 0 | 0 | 3 |
| 硕士研究生 | 4 | 0 | 0 | 0 | 0 | 0 | 4 |
| 博士研究生及以上 | 5 | 1 | 1 | 0 | 0 | 0 | 7 |
| 合计 | | 265 | 49 | 5 | 9 | 4 | 4 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | | | |

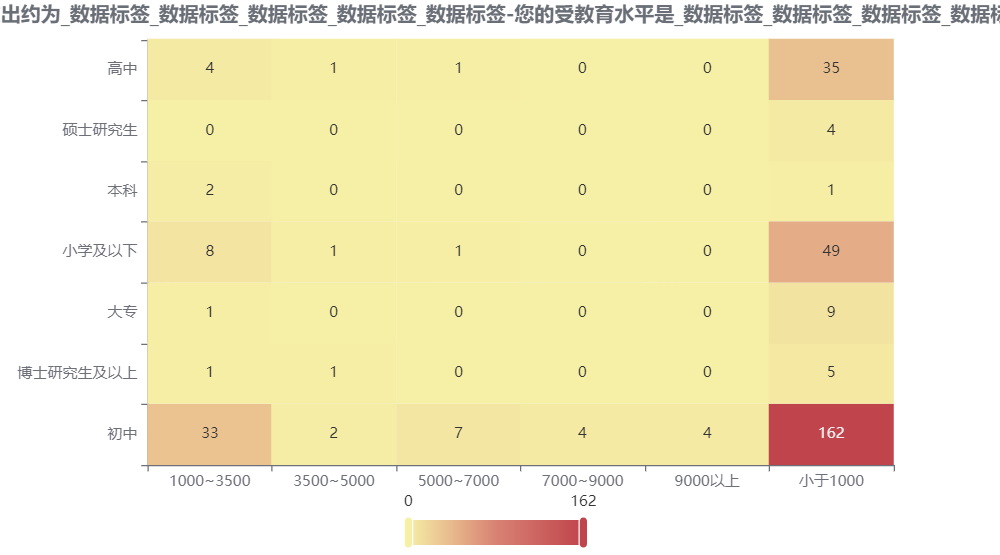
**图表说明：**

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0.799，水平上不呈现显著性，不能拒绝原假设，因此您在口腔上每年的支出和您的受教育水平不存在显著性差异。

**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 |
| 您的受教育水平---您在口腔上每年的支出 | 0.264 | 0.118 | 0.255 | 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：您的受教育水平Cramer's V值为0.118，因此您在口腔上每年的支出和您的受教育水平的差异程度为弱程度差异。

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:{您对口腔健康了解吗}；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0.000\*\*\*，水平上呈现显著性，拒绝原假设，因此您对口腔健康了解吗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 | 您对口腔健康了解吗 | | | | 合计 | X² | P |
| 一般了解 | 有点了解 | 非常了解 | 完全了解 |
| 您在口腔上每年的支出约为 | 小于1000 | 99 | 133 | 32 | 1 | 265 | 42.242 | 0.000\*\*\* |
| 1000~3500 | 15 | 23 | 11 | 0 | 49 |
| 3500~5000 | 2 | 2 | 0 | 1 | 5 |
| 5000~7000 | 3 | 4 | 1 | 1 | 9 |
| 7000~9000 | 2 | 1 | 1 | 0 | 4 |
| 9000以上 | 0 | 4 | 0 | 0 | 4 |
| 合计 | | 121 | 167 | 45 | 3 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | |

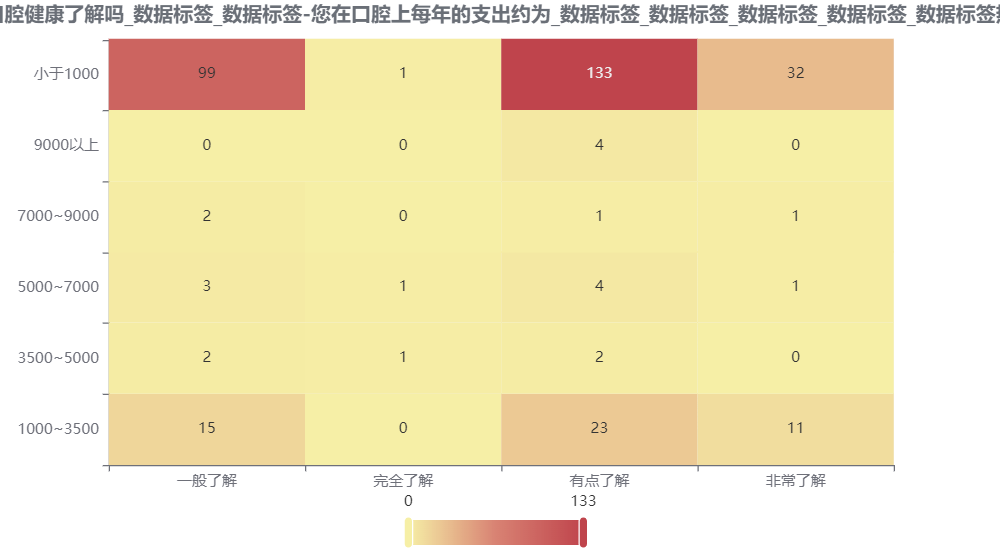
**图表说明：**

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0.000\*\*\*，水平上呈现显著性，拒绝原假设，因此您对口腔健康了解吗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---您对口腔健康了解吗 | 0.355 | 0.205 | 0.334 | 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0.205，因此您对口腔健康了解吗和您在口腔上每年的支出约为的差异程度为中等程度差异。

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:{您对口腔健康了解吗}；变量Y:{您第一次去口腔医院是因为}  
  
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0.000\*\*\*，水平上呈现显著性，拒绝原假设，因此您对口腔健康了解吗和您第一次去口腔医院是因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 | 您对口腔健康了解吗 | | | | 合计 | X² | P |
| 一般了解 | 有点了解 | 非常了解 | 完全了解 |
| 您第一次去口腔医院是因为 | 亲戚建议 | 6 | 6 | 1 | 0 | 13 | 42.186 | 0.000\*\*\* |
| 体检检查 | 22 | 26 | 16 | 1 | 65 |
| 其他医生建议 | 4 | 1 | 0 | 0 | 5 |
| 口腔疾病 | 53 | 98 | 26 | 1 | 178 |
| 朋友建议 | 7 | 1 | 0 | 1 | 9 |
| 没去过口腔医院 | 29 | 35 | 2 | 0 | 66 |
| 合计 | | 121 | 167 | 45 | 3 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | |

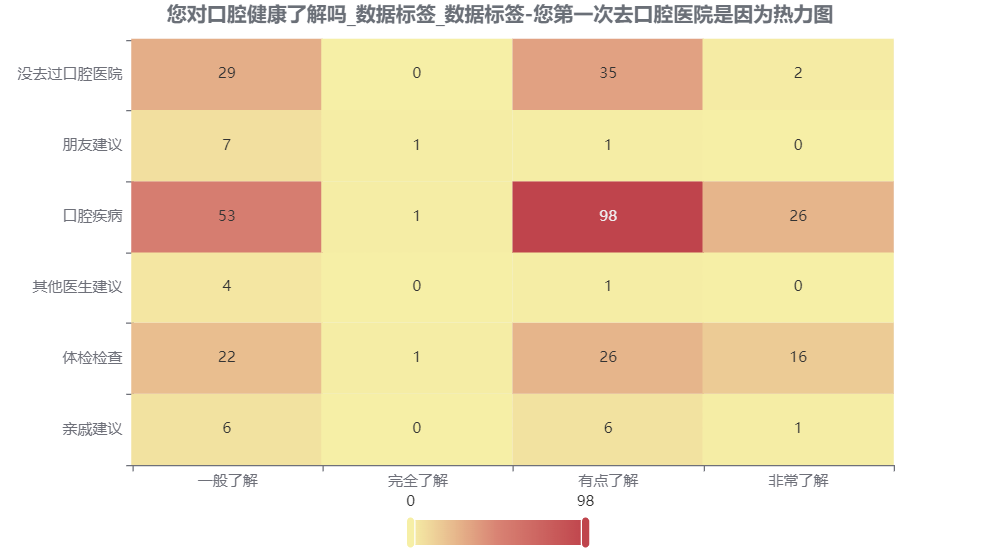
**图表说明：**

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0.000\*\*\*，水平上呈现显著性，拒绝原假设，因此您对口腔健康了解吗和您第一次去口腔医院是因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 |
| 您第一次去口腔医院是因为---您对口腔健康了解吗 | 0.354 | 0.205 | 0.334 | 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：您第一次去口腔医院是因为Cramer's V值为0.205，因此您对口腔健康了解吗和您第一次去口腔医院是因为的差异程度为中等程度差异。

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:{您是否患过口腔问题}；变量Y:{间隔多久您会进行口腔检查包括洗牙}  
  
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，因此您是否患过口腔问题和间隔多久您会进行口腔检查包括洗牙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**

**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。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 题目 | 名称 | 您是否患过口腔问题 | | | 合计 | X² | P | 合计 | X² | P |
| 不记得了 | 否 | 是 |
| 间隔多久您会进行口腔检查包括洗牙 | 小于一个月 | 1 | 2 | 8 | 11 | 18.358 | 0.019\*\* | 59 | 23.377 | 0.799 |
| 1~3个月 | 0 | 3 | 7 | 10 | 212 |
| 半年 | 11 | 41 | 83 | 135 | 41 |
| 一年 | 2 | 22 | 97 | 121 | 10 |
| 从未 | 0 | 13 | 46 | 59 | 3 |
| 合计 | | 14 | 81 | 241 | 336 | 4 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | 7 |
| 合计 | | 265 | 49 | 5 | 9 | 4 | 4 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | | | |

● 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，因此您是否患过口腔问题和间隔多久您会进行口腔检查包括洗牙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 |
| 间隔多久您会进行口腔检查包括洗牙---您是否患过口腔问题 | 0.234 | 0.165 | 0.228 | 0.07 |

**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：间隔多久您会进行口腔检查包括洗牙Cramer's V值为0.165，因此您是否患过口腔问题和间隔多久您会进行口腔检查包括洗牙的差异程度为弱程度差异。

分析流程  
  
数据源：  
data口腔健康调查表+评价.xlsx  
  
算法配置：  
算法： 卡方检验分析  
变量： 变量X:{您的受教育水平}；变量Y:{间隔多久您会进行口腔检查包括洗牙}  
  
分析结果：  
卡方检验用于检验两分组变量是否存在显著性差异： 基于您的受教育水平和间隔多久您会进行口腔检查包括洗牙，显著性P值为0.440，水平上不呈现显著性，接受原假设，因此对于您的受教育水平和间隔多久您会进行口腔检查包括洗牙不存在显著性差异。

分析步骤  
  
1. 分析卡方检验是否呈现显著性(P<0.05)。  
2. 若呈现显著性，具体根据类别的差异百分比进行描述。  
3. 若呈现显著性，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：卡方检验分析结果**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 题目 | 时间 | 您的受教育水平 | | | | | | | 总计 | X² | 校正X² | P |
| 小学及以下 | 初中 | 高中 | 大专 | 本科 | 硕士研究生 | 博士研究生及以上 |
| 间隔多久您会进行口腔检查包括洗牙 | 从未 | 7 | 42 | 6 | 3 | 0 | 0 | 1 | 59 | 24.389 | 24.389 | 0.44 |
| 一年 | 27 | 69 | 16 | 3 | 3 | 1 | 2 | 121 |
| 半年 | 20 | 88 | 17 | 4 | 0 | 2 | 4 | 135 |
| 1~3个月小于一个月 | 1 | 7 | 2 | 0 | 0 | 0 | 0 | 10 |
| 小于一个月 | 4 | 6 | 0 | 0 | 0 | 1 | 0 | 11 |
| 合计 | | 59 | 212 | 41 | 10 | 3 | 4 | 7 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | | | | | |

**图表说明：**

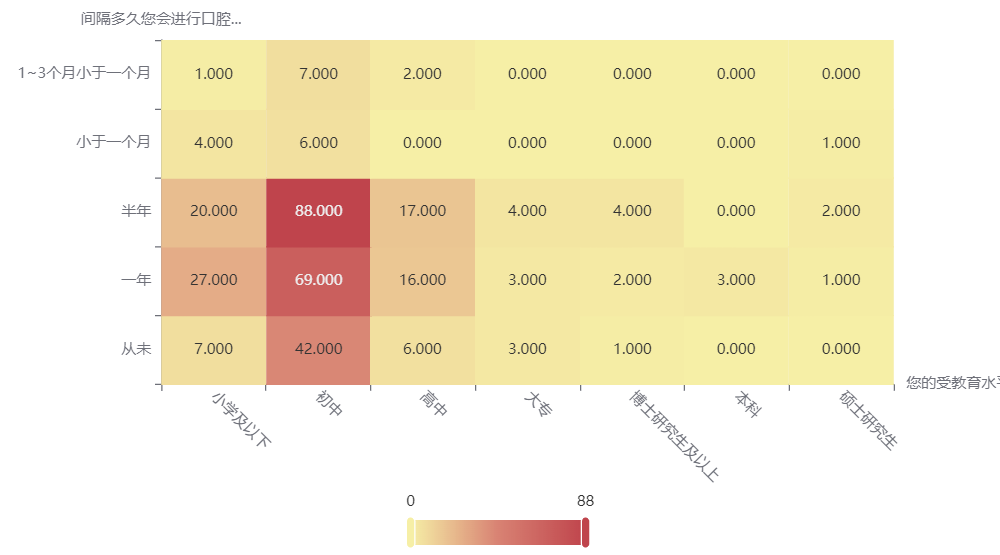
上表展示了模型检验的结果，包括数据的频数、频数百分比、卡方值、显著性P值。  
1. 分析模型是否呈现出显著性(P<0.05)。  
2. 若呈现显著性，拒绝原假设，则说明各样本之间存在显著性差异。具体根据类别的差异百分比进行描述。反之数据不存在显著性差异。

**图表分析：**

卡方检验分析的结果显示，基于您的受教育水平和间隔多久您会进行口腔检查包括洗牙，显著性P值为0.440，水平上不呈现显著性，接受原假设，因此对于您的受教育水平和间隔多久您会进行口腔检查包括洗牙不存在显著性差异。

**输出结果2：卡方交叉热力图**

您的受教育水平---间隔多久您会进行口腔检查包括洗牙热力图



**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 |
| 间隔多久您会进行口腔检查包括洗牙 | 0.269 | 0.135 | 0.26 | 0.05 |

**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：间隔多久您会进行口腔检查包括洗牙 Cramer’s V值为0.135，因此间隔多久您会进行口腔检查包括洗牙和您的受教育水平是的差异程度为弱程度差异。

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:{您在口腔上每年的支出}；变量Y:{您的受教育水平}  
  
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0.799，水平上不呈现显著性，不能拒绝原假设，因此您在口腔上每年的支出和您的受教育水平不存在显著性差异。

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 | 您在口腔上每年的支出约为 | | | | | | 合计 | X² | P |
| 小于1000 | 1000~3500 | 3500~5000 | 5000~7000 | 7000~9000 | 9000以上 |
| 您的受教育水平是 | 小学及以下 | 49 | 8 | 1 | 1 | 0 | 0 | 59 | 23.377 | 0.799 |
| 初中 | 162 | 33 | 2 | 7 | 4 | 4 | 212 |
| 高中 | 35 | 4 | 1 | 1 | 0 | 0 | 41 |
| 大专 | 9 | 1 | 0 | 0 | 0 | 0 | 10 |
| 本科 | 1 | 2 | 0 | 0 | 0 | 0 | 3 |
| 硕士研究生 | 4 | 0 | 0 | 0 | 0 | 0 | 4 |
| 博士研究生及以上 | 5 | 1 | 1 | 0 | 0 | 0 | 7 |
| 合计 | | 265 | 49 | 5 | 9 | 4 | 4 | 336 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | | | |

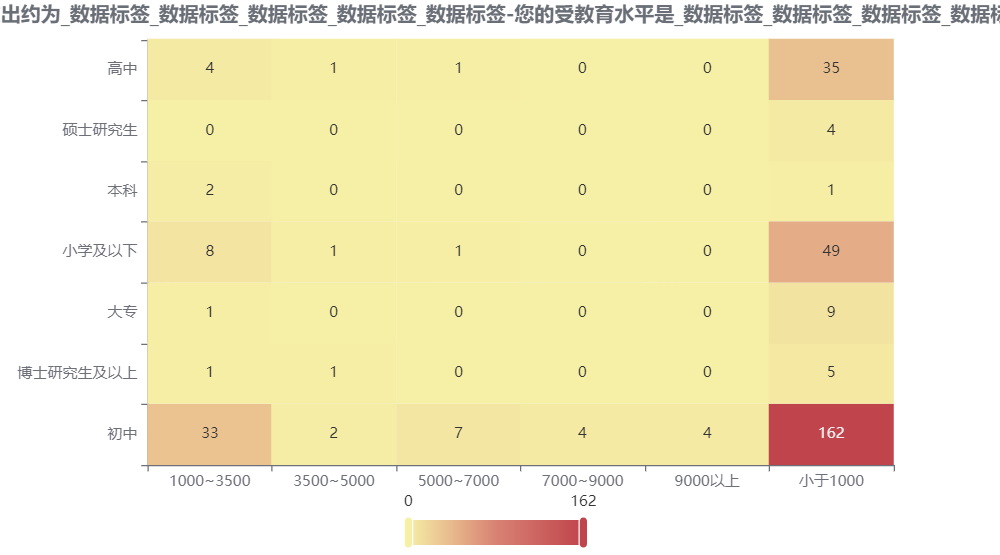
**图表说明：**

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0.799，水平上不呈现显著性，不能拒绝原假设，因此您在口腔上每年的支出和您的受教育水平不存在显著性差异。

**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 |
| 您的受教育水平---您在口腔上每年的支出 | 0.264 | 0.118 | 0.255 | 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：您的受教育水平Cramer's V值为0.118，因此您在口腔上每年的支出和您的受教育水平的差异程度为弱程度差异。