## **output/csv/第一问/**

### **1. cirrhosis\_describe.csv**

* **内容**：肝硬化（cirrhosis）数据集的描述性统计表。
* **包含**：每个变量的计数（count）、均值（mean）、标准差（std）、最小值（min）、25%分位数、50%分位数（中位数）、75%分位数、最大值（max）等。
* **作用**：帮助快速了解肝硬化数据的分布、集中趋势和离散程度。

### **2. heart\_describe.csv**

* **内容**：心脏病（heart）数据集的描述性统计表。
* **结构同上**，用于了解心脏病数据的基本分布特征。

### **3. stroke\_describe.csv**

* **内容**：中风（stroke）数据集的描述性统计表。
* **结构同上**，用于了解中风数据的基本分布特征。

## **output/plt/第一问/**

### **相关性热力图（基于 Spearman 相关系数）**

#### 1. cirrhosis\_corr.png

* **内容**：肝硬化数据集中所有数值型变量之间的 Spearman 相关系数热力图。
* **作用**：展示变量之间的单调相关性，颜色越深表示相关性越强，便于发现潜在影响因子。

#### 2. heart\_corr.png

* **内容**：心脏病数据集的 Spearman 相关系数热力图。
* **作用同上**。

#### 3. stroke\_corr.png

* **内容**：中风数据集的 Spearman 相关系数热力图。
* **作用同上**。

### **患病率分布条形图**

#### 4. cirrhosis\_rate.png

* **内容**：肝硬化数据集按 Stage（分期）分组的患病人数条形图。
* **作用**：展示不同分期的样本数量分布。

#### 5. heart\_rate.png

* **内容**：心脏病数据集按 HeartDisease 分组的患病/未患病人数条形图。
* **作用**：直观显示患病与未患病的比例。

#### 6. stroke\_rate.png

* **内容**：中风数据集按 stroke 分组的患病/未患病人数条形图。
* **作用同上**。

### **年龄分组患病率条形图**

#### 7. cirrhosis\_Age\_rate.png

* **内容**：肝硬化数据集按年龄分段（如0-30, 30-40, ...）统计各年龄段的患病率（Stage均值）条形图。
* **作用**：揭示不同年龄段的患病风险。

#### 8. heart\_Age\_rate.png

* **内容**：心脏病数据集按年龄分段统计各年龄段的患病率（HeartDisease均值）条形图。
* **作用同上**。

#### 9. stroke\_age\_rate.png

* **内容**：中风数据集按年龄分段统计各年龄段的患病率（stroke均值）条形图。
* **作用同上**。

### **箱线图（Boxplot）**

#### 10. cirrhosis\_XXX\_boxplot.png

* **内容**：肝硬化数据集，按 Stage 分组，对每个数值型变量（XXX）绘制的箱线图。
* **作用**：展示不同分期下各变量的分布、离群值和中位数，便于对比组间差异。

#### 11. heart\_XXX\_boxplot.png

* **内容**：心脏病数据集，按 HeartDisease 分组，对每个数值型变量绘制的箱线图。
* **作用同上**。

#### 12. stroke\_XXX\_boxplot.png

* **内容**：中风数据集，按 stroke 分组，对每个数值型变量绘制的箱线图。
* **作用同上**。

> 其中 XXX 代表具体的数值型变量名（如年龄、血压、胆固醇等），每个变量都会生成一个对应的箱线图文件。