

# 数据挖掘-全表统计

- 概述
- 输入/输出
- 参数设置
- 示例

## 概述

全表统计是对观测数据进行不同的统计分析，可统计信息：最小值、最大值、平均值、标准差、方差、总和、行数、唯一值、缺失值、偏度、峰度、中位数、下四分位、上四分位、众数等指标；还可使用箱线图和直方图尽可能简单全面表达数据所蕴含的数值范围、分布等信息。

## 输入/输出

|    |                |
|----|----------------|
| 输入 | 一个输入端口，用于接收数据集 |
| 输出 | 没有输出端口         |

## 参数设置

| 参数名称    | 说明  | 备注 |
|---------|---|----|
| 选择列     | 选择需要统计的特征列（列数<=10）  | 必填 |
| 可选统计    | 选择需要的指标进行统计分析。可选的指标：偏度、峰度、总和、众数、方差、标准差、缺失值。 <div><span>ⓘ</span> 使用全表统计节点默认统计的指标：行数、最大值、最小值、平均值、下四分位、中位数、上四分位、唯一值。</div> |    |
| 连续数据分桶数 | 设置连续数据分桶数，分桶数为>=2的整数，统计结果中，数值变量会按照这个分桶数来分箱展示  | 必填 |
| 显示异常值   | 设置是否在箱线图中显示异常值。默认不勾选。   |    |

## 示例

使用“鸢尾花数据”，选择4个特征进行全表统计，查看分析效果如图：

