

数据挖掘-假设检验

概述

通过特征变量与目标变量之间的偏差来检验数据之间的相关性或回归分析中的拟合结果。

- 概述
- 输入/输出
- 参数设置
- 示例

输入/输出

输入	一个输入端口，用于接收数据集
输出	一个输出端口，用于输出假设检验结果数据集

参数设置

参数名称	说明	备注
选择特征列	用于选择需要检验的字段列	必填
选择标签列	用于选择做为标签列的字段（必须为数值型）	

示例

使用“鸢尾花数据”，选择4个特征列和1个标签列，输出结果如下图：

已挑选列	P值	自由度	统计量
Petal_Length	0.0	84	271.8
Petal_Width	0.0	42	271.75
Sepal_Length	6.6659873176888595E-9	68	156.26666666666665
Sepal_Width	6.01603148220331E-5	44	89.54628704628703

鼠标点击右键查看分析结果：

查看分析结果

已挑选列	P值	自由度	统计量
Petal_Length	0	84	271.8
Petal_Width	0	42	271.75
Sepal_Length	6.6659873176888595e-9	68	156.26666666666665
Sepal_Width	0.0000601603148220331	44	89.54628704628703

×

以上图知：已挑选的特征列的P值都趋近于0，则说明已挑选的这4个特征都与标签列具有较高的相关性，该特征的P值越趋近于0则该特征与标签列字段越相关。