

- 一.数据预处理
- 二.描述性统计与绘图
- ▼ 三.基本数据分析
 - 1. 回归（包括OLS、WLS、GLS以及相对来说较为常见的逻辑回归）
 - 2. 多元统计相关
 - 3. 机器学习相关
- 四.可提供进阶纠错判断
- 五.综合后的自动化数据分析
- 六.结果整合

一.数据预处理

该部分主要分为两大功能：

1. 处理异常值和缺失值，修整数据长度和宽度，化为规整的，数据类型恰如其分的矩阵形式
2. 根据数据类型的占比和用户的功能选择，给数据分定“标签”，便于后续处理。

二.描述性统计与绘图

注：根据数据类型和数据预处理后输出的“标签”不同，可能会选择不同的图标进行展示

基本包括如下内容

1. 基本数字特征的展示，诸如均值，中位数，方差，上下四分位数，即一个summary
2. 基础的图标展示，比如针对一维数据的带着核密度曲线的直方图，针对多维数据的多维散点图
3. 略微进阶的图表，例如协方差阵的热力图等
4. 未完待续.....

三.基本数据分析

1. 回归（包括OLS、WLS、GLS以及相对来说较为常见的逻辑回归）

2. 多元统计相关

3. 机器学习相关

四.可提供进阶纠错判断

该项功能旨在提供给用户相应的统计学参考，比如

1. 会自动根据回归结果判断数据是否存在明显的异常值（强影响点），是否存在异方差、多重共线性、内生性、自相关
2. 会根据多元统计相关的统计方法给出数据的具体特征和可能存在的有效处理改进方法，例如如果数据存在明显的相关性，即主成分后第一主成分的方差贡献率奇高无比，那么可能代表数据亟待降维。
3. 会自动生成基于**基本机器学习方法**的分析结果并且提供可能的改进方案。

五.综合后的自动化数据分析

该功能旨在模拟用户针对初步分析后的数据进行改良（无论是修饰数据还是改良模型），并自动生成初步改良后的结果**视图**或**图表**

六.结果整合

改功能会通过外部程序将上述输出的所有结果，**包括**

1.字符串及文本

2.表格

3.图片

整合成pdf亦或是word进行输出，亦或是根据**用户选择**，生成原始文件夹并输出自动判断的**综述文档**