分析思路

第一步：确定分析的目标：寻找身高和体重的关系

目标中需要的变量：身高、体重

第二步：对数据本身的特质进行了解

身高、体重：判断是连续变量还是分类变量

分类变量：频率表格，众数。可视化：饼图、条形图

连续变量：频率表格、探索、描述等过程。可视化：直方图、箱体图

第三步：确定数据之间的相互影响和相互关系

1. 目标为连续变量，输入为连续变量

可视化：散点图

分析方法：相关分析

1. 目标是连续变量，输入为分类变量

可视化：堆积直方图、条形图、折线图

分析方法：T检验/方差分析

1. 目标是分类变量，输入为分类变量

可视化：分组饼图、网络图

分析方法：交叉表/透视表、卡方检验

第四步：数据分析/数据建模

1. 目标为连续变量，输入为连续变量

建模：回归、决策树、神经网络

回归：线性回归

1. 目标是连续变量，输入为分类变量

建模：logistic回归、决策树

1. 目标是分类变量，输入为分类变量

建模：决策树、分类