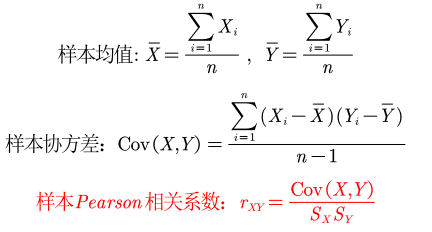
模型建立与求解

Pearson相关系数模型

存在两组数据为样本数据，

Pearson 相关系数用来确定两组数据集合是否在一条直线上，用于衡量两组变量间的线性关系。当两个变量都是正态连续变量，而且两者之间呈线性关系时，经常选用Pearson相关系数刻画二者的相关程度。具体计算公式如下



其中Sx是随机变量X的标准差，Sy是随机变量Y的标准差，并且rxy∈[-1,1]，当r=1时表明X与Y之间存在绝对线性正向关系，r=-1表明X与Y存在负向线性关系，r=0表明X与Y之间没有线性关系，但并不意味着两者没有非线性关系。



正态性检验

（1）JB检验（大样本 n> 30）

H0：{xi}服从正态分布 H1:{xi}不服从正态分布

检验统计量：

S：{xi}的偏度  k2：二阶中心矩 k3：三阶中心矩

K: {xi}的峰度 K =  k4:四阶中心矩

若P<,则在1-的置信水平上拒绝原假设。

若P>,则在1-的置信水平上接受原假设。



（2）Shapiro-wilk检验 （小样本 2<n<51）

H0：{xi}服从正态分布 H1:{xi}不服从正态分布

检验统计量：

为次序统计量

ai为在样本容量为n时特有的值。

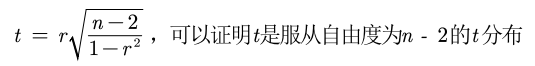
若P<,则在1-的置信水平上拒绝原假设。

若P> ,则在1-的置信水平上接受原假设。

Pearson模型的求解

建立原假设与备择假设 H0: r=0, H1: r≠0 r：皮尔逊相关系数

构造t统计量



将r(相关系数)与n(样本量)带入统计量并计算得到检验p值，如下图所示



\*\*表示在α=99%的置信水平上拒绝原假设

\*表示在α=95%的置信水平上拒绝原假设

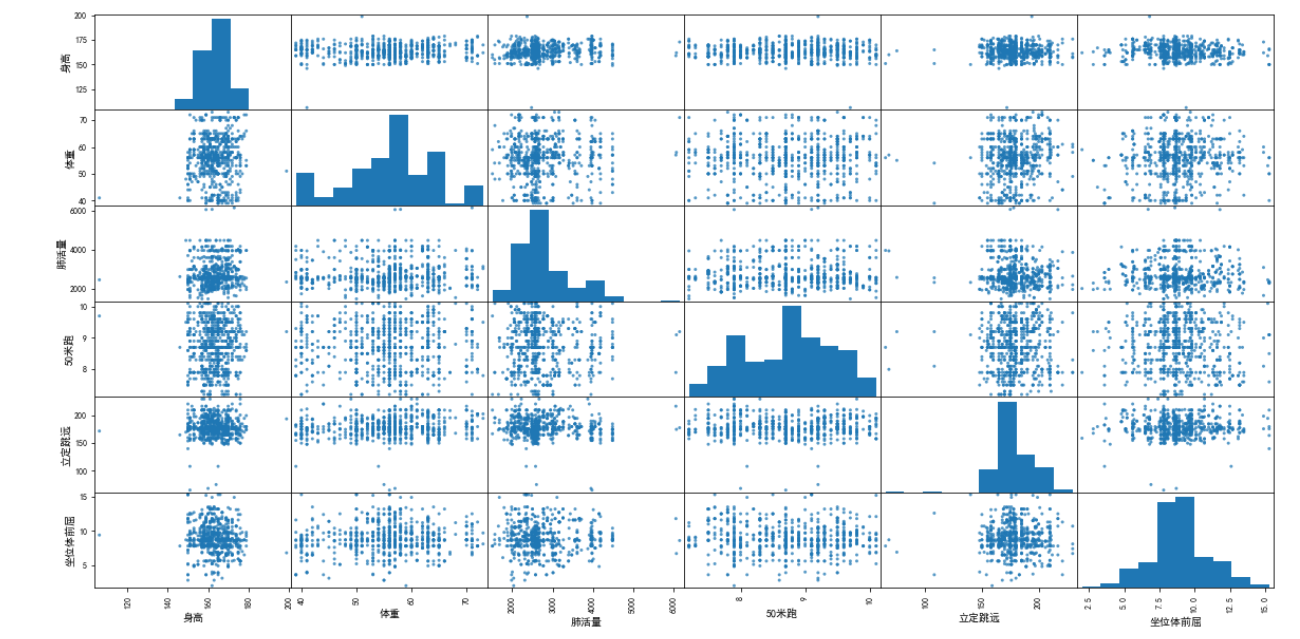
若P< ,则在1-的置信水平上拒绝原假设。

若P>  ,则在1-的置信水平上接受原假设。

Spearman模型的建立

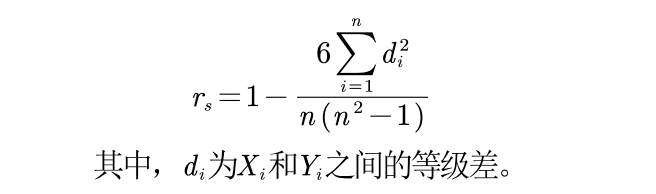
1. 描述性统计

求出每个指标的最大值最小值均值标准差，画出箱线图，分析数据特征，做一个简单的描述性统计，画出散点图



1. Sprearman相关系数

根据上述分析，此数据明显不符合正态性假定，也没有明显的线性关系，因此选择spearman相关系数分析指标两两之间的关系。对于样本容量为*n*的样本，*n*个原始数据被转换成等级数据，Spearman相关系数rs为

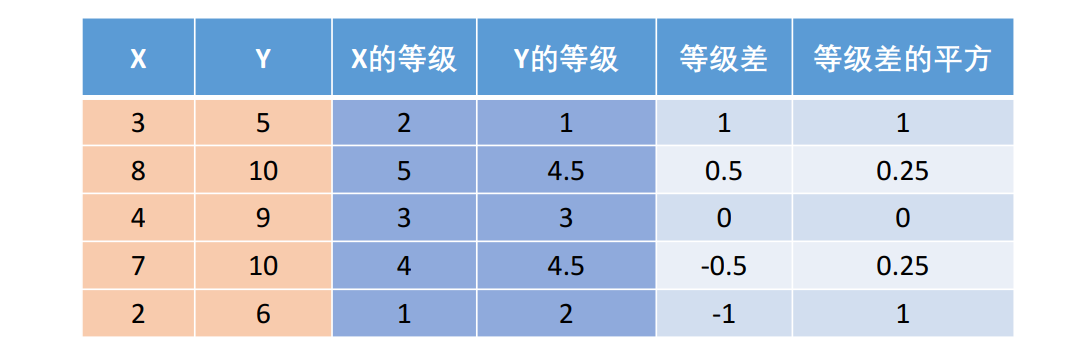




模型的求解

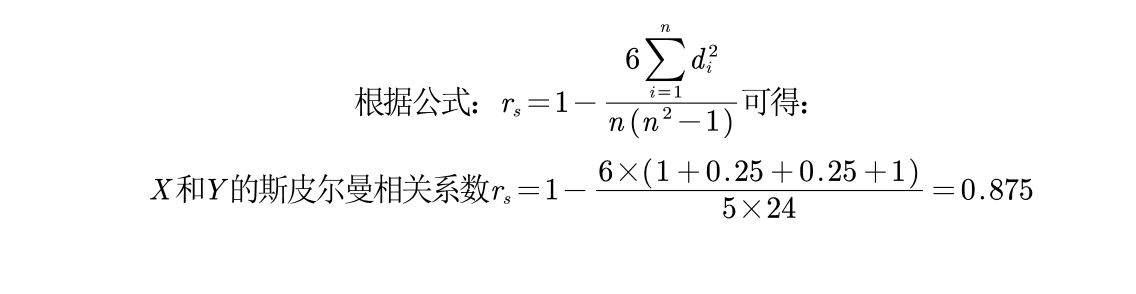
1. 划分等级

将两个指标从小到大划分不同等级，具体等级如下表所示



1. 相关系数检验

将求出的等级差代入计算公式，得到X与Y的spearman相关系数



斯皮尔曼相关系数显著性检验

H0:= 0 H1: 0 ：斯皮尔曼相关系数

检验统计量： 小样本（n<=30）：直接查临界值表

大样本：

若P< ,则在1-的置信水平上拒绝原假设。

若P>  ,则在1-的置信水平上接受原假设。