数据的标准化

数据的标准化分为很多种，可以使用Excel，也可以使用MATLABSPSS实现，有的时候也可以用数据无量纲化，数据的归一化等等。

数据标准化，使用在评价问题中，几乎每一个评价问题都需要进行数据标准化，很多时候预测方法也会使用。

一、Min-max 标准化

Min-max标准化方法是对原始数据进行线性变换。设minA和maxA分别为属性A的最小值和最大值，将A的一个原始值x通过min-max标准化映射成在区间[0,1]中的值x'，其公式为：

新数据=（原数据-最小值）/（最大值-最小值）

MATLAB源代码：

clc,clear

a=[4 5 6 5 4 6 5 8 9 10 12 4 9];%原始数据

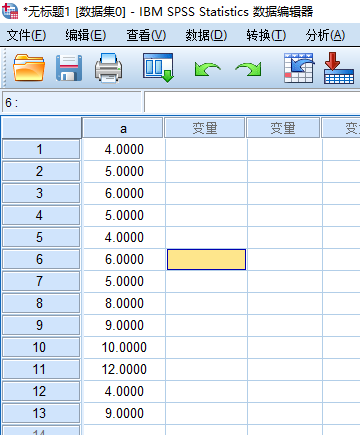
max=max(a);%求原始数据a的最大值

min=min(a);%求原始数据a的最小值

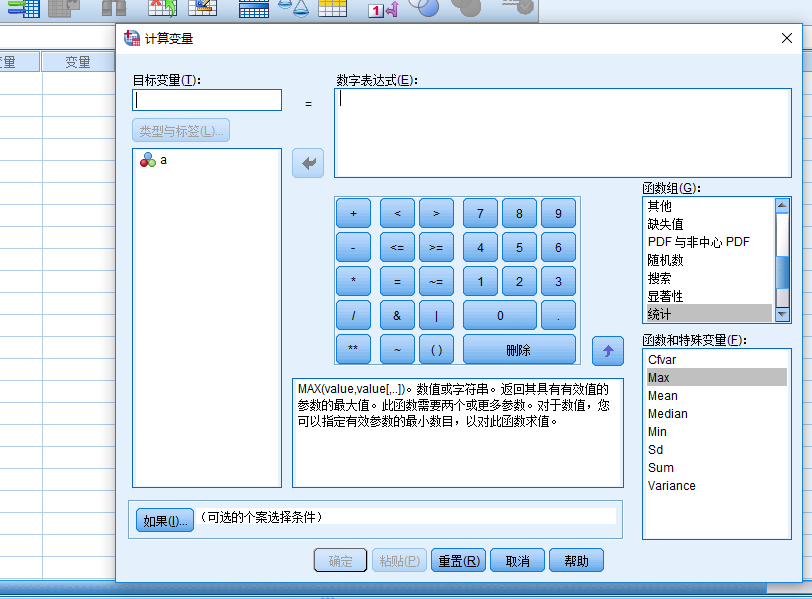
b=(a-min)/(max-min)%对数据进行Min-max标准化，标准化后的数据为b

SPSS图解：

1. 输入数据：



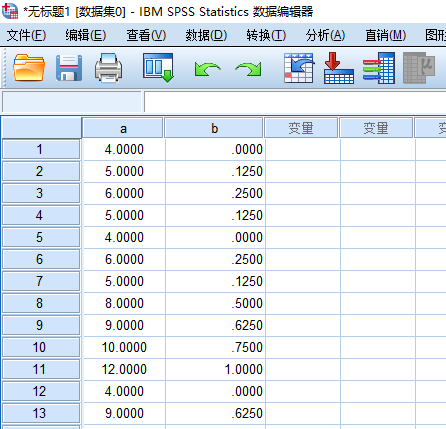
1. 依次点击“转换”、“计算变量”，弹出“计算变量”对话框，如图：



1. 在“目标变量”中输入标准化后数据的名称，在“数字表达式”中输入新数据=（原数据-最小值）/（最大值-最小值），本题输入为“(a - 4) / (12 - 4)”，如图：



4.单击“确定”即可（调整一下小数位数，默认两位小数位。）：



二、z-score 标准化

这种方法基于原始数据的均值（mean）和标准差（standard deviation）进行数据的标准化。将A的原始值x使用z-score标准化到x'。

z-score标准化方法适用于属性A的最大值和最小值未知的情况，或有超出取值范围的离群数据的情况。

新数据=（原数据-均值）/标准差

spss默认的标准化方法就是z-score标准化。

用Excel进行z-score标准化的方法：在Excel中没有现成的函数，需要自己分步计算，其实标准化的公式很简单。

　　步骤如下：

　　1.求出各变量（指标）的算术平均值（数学期望）xi和标准差si ；

　　2.进行标准化处理：

　　zij=（xij－xi）/si

　　其中：zij为标准化后的变量值；xij为实际变量值。

　　3.将逆指标前的正负号对调。

标准化后的变量值围绕0上下波动，大于0说明高于平均水平，小于0说明低于平均水平。

MATLAB源代码：

clc,clear

a=[4 5 6 5 4 6 5 8 9 10 12 4 9];%原始数据

s=var(a);%求原始数据a的标准差s

c=mean(a);%求原始数据a的均值c

b=(a-c)/s%b为标准化后的数据

三、Decimal scaling小数定标标准化

这种方法通过移动数据的小数点位置来进行标准化。小数点移动多少位取决于属性A的取值中的最大绝对值。将属性A的原始值x使用decimal scaling标准化到x'的计算方法是：

x'=x/(10^j)

其中，j是满足条件的最小整数。

例如 假定A的值由-986到917，A的最大绝对值为986，为使用小数定标标准化，我们用每个值除以1000（即，j=3），这样，-986被规范化为-0.986。

注意，标准化会对原始数据做出改变，因此需要保存所使用的标准化方法的参数，以便对后续的数据进行统一的标准化。

除了上面提到的数据标准化外还有对数Logistic模式、模糊量化模式等等：

对数Logistic模式：

模糊量化模式：



X为原数据，b为标准化后的数据

——————————————————————————————————