
SPSS—描述性统计分析—探索性分析

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/229.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS描述性统计分析菜单

除了可以计算基本的统计量之外，也可以给出一些简单的检验结果和图形，有助于用户进一步的分析数据。使得用户能够从大量的分析结果之中挖掘到所需要的统计信息。

适用范围

对资料的性质、分布特点等完全不清楚的时候

Analyze -> Descriptive Statistics -> Expore

数据源

ceramics.sav

因变量列表

用于选入待分析的变量

因子列表

用于选择分组变量，根据该变量取值不同，分组分析因变量列表中的变量

标注个案

选择标签变量

统计量

描述性

计算一般的描述性统计量，及指定的均数可信区间

M-估计量

描述集中趋势的统计量，用于稳健估计

界外值

分别输出5个极大值和极小值

百分位数

输出变量5%,10%,25%,50%,75%,90%,95%分位数

绘制

带校验的正态图

选择是否进行正态校验，且是否输出相应的Q-Q图

伸展与级别Levene检验

当选入分组变量时，该功能才被激活，主要用于比较各组之间的离散程度是否一致。在这里可以选择“未转换”，用于方差齐性检验

选项

输出结果

个案处理分析结果

包括观测量、缺失值等信息

描述性统计量

包括：均值、95%置信区间、方差、中位数、标准差、最大最小值、偏度和峰度等信息

集中趋势分布的3种较佳平稳测度

较佳测度之一：中位数等

中位数

与均值和众数大不相同，中位数是依赖于数据的主体部分而不是极值，因此它的值不是过分地受某几个观察值的影响

平稳估计量

如果对数据来源的总体做出某个假设(比如假定服从正态分布)，则会有更佳分布位置的估计量，这种估计量称为平稳或稳健测度的估计量

较佳测度之二：修正均值

由于均值深受极端值影响，因此可通过去掉一些远离主体数据的极端值，进而获得一个对于分布位置简单而平稳的估计量

5%修正均值

是通过去掉所有观察值中最大的5%和最小的5%的数据而获得

调整后的均值与中位数可更好的利用数据

较佳测度之三：M估计

将极端值计算在内，而赋予比靠近中央值较小的一个权重，这种方法可借助M估计或采用广义最大似然估计

M-estimators：平稳分布位置的最大似然估计量

Huber的M估计值

Tukey双权重估计值

Hampel重复递减M估计值

Andrew波形估计值

M-估计器

极值

这里用标注个案来标记极值

正态性检验

其中Premium变量对应的K-S检验P值和Shapiro-Wilk检验P值均为0.000，非常显著，应该拒绝原假设。所以，此变量的数据分布不是正态分布。

而Standard数据的分布不是显著的，可以认为是正态分布

在‘探索’里出现的Kolmogorov-Smirnov 检验，它的右上角有一个a的注释号。它将Kolmogorov-Smirnov 检验改进用于一般的正态性检验。

而在‘非参数检验’里出现的Kolmogorov-Smirnov 检验，是没有经过纠正或改进的。

该正态性检验只能做标准正态检验。

SPSS规定：当样本含量 $3 \leq n \leq 5000$ 时，结果以Shapiro—Wilk(W 检验)为难，当样本含量 $n > 5000$ 结果以Kolmogorm —Smimov(D检验)为准。

问题：

(1) 在实际应用中常出现检验结果与直方图、正态性概率图不一致，甚至几种假设检验方法结果完全不同的情况。

(2) Shapiro—Wilk 检验(W 检验)和经过Lilliefors 显著水平修正的Kolmogorov—Smirnov 检验(D 检验)是用一个综合指标(顺序统计量或D)来判定资料的正态性由于两种方法都是用一个指标反映资料的正态性，

所以当资料的正态峰和对称性两个特征有一个不满足正态性要求时，两种方法出现假阴性错误的机率均较大;而且两种方法的检验统计量都是进行大小排序后得到，所以易受异常值的影响。

(3) Kolmogorov—Smirnov

单一样本检验

是根据实际的累计频数分布和理论的累计频数分布的最大差异来检验资料的正态性，可对正态分布进行拟合优度检验。但它并非检验正态性的专用方法，因此它的检验效率是最低的，最容易受样本量和异常值等因素的影响。

方差齐性检验

假设检验：

H0: 两样本方差齐性(相等，或无显著性差异)

如上图，Sig > 0.2，并无显著差异。

正态Q-Q图

正态性检验可以通过直观的Q-Q图，进行人工验证。

Q-Q图是一种散点图,对应于正态分布的Q-Q图,就是由标准正态分布的分位数为横坐标,样本值为纵坐标的散点图.要利用QQ图鉴别样本数据是否近似于正态分布,只需看QQ图上的点是否近似地在一条直线附近,而且该直线的斜率为标准差,截距为均值.

如上图, batch=Standard Q-Q图上的点在一条直线附近,可以认为是正态分布,和正态性检验Lilliefors, Shapiro-Wilk得出的结果一致。

反趋势正态 Q-Q 图

如上图,反趋势正态概率Q-Q图以变量的观测值为X坐标,以变量的Z得分与期望值的偏差为Y坐标。

batch=Standard 图的观测点离期望值很集中,说明符合正态分布。

盒子图

Premiun中有部分异常数据，数据偏大。需要进行异常值检测。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发