

正态性检验算法

目录

SPSSAU 数据格式..... 1
 计算公式..... 2
 1. KolmogorovSmirnov 检验 2
 2. ShapiroWilk 检验 2
 3. JarqueBera 检验..... 2
 参考文献..... 2

相关系数研究两两变量之间的相关关系情况，在 SPSSAU 中支持：

- ✓ 批量对‘标题’进行分析；
- ✓ 支持区分不同组别 X 时进行分析；
- ✓ 提供 KolmogorovSmirnov、ShapiroWilk 和 JarqueBera 检验。



如果分析时有分组项 X，则放入对应‘分组项’框中即可。

SPSSAU 数据格式

X	Title1	Title2	Title3
Male	7.859206	7.699458	8.171911
Male	8.27663	9.567024	3.670163
Male	2.098827	3.968077	5.659362
Male	6.024826	0.623412	4.920464
Male	3.102632	4.343761	5.748209
Female	2.963518	7.800297	3.122878
Female	8.232136	0.099224	0.112001
Female	1.798314	8.374911	9.011049
Female	2.194865	0.566183	9.472328
Female	2.442311	7.202591	6.13348
Female	5.312099	1.245504	7.203394
Female	4.655058	5.38763	4.949512

比如研究 3 个 title 的正态分布情况，并且按 X 区分（即区分不同性别时的正态分布情况）。

计算公式

1. KolmogorovSmirnov 检验

$$D = \max|F_n(x) - F(x)|$$

其中：

$F_n(x)$ 是样本的经验累积分布函数（ECDF），基于正态分布

$F(x)$ 是理论分布的累积分布函数 CDF，基于正态分布

2. ShapiroWilk 检验

$$W = \frac{(\sum_{i=1}^n a_i X_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

其中：

X_i : 第 i 个值（排序后的样本值）

\bar{X} : 样本均值

a_i : 基于正态分布的顺序统计量的期望值，该常数是通过查表得到。

SPSSAU 借助 Python 中 scipy 包进行 ShapiroWilk 检验，可参考：

<https://docs.scipy.org/doc/scipy-0.19.0/reference/generated/scipy.stats.shapiro.html>

3. JarqueBera 检验

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right)$$

其中：

S : 样本的偏度（skewness）

$$S = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

K : 样本的峰度（kurtosis）

$$K = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3$$

s : 样本标准差

n : 样本大小

参考文献

[1] The SPSSAU project (2024). SPSSAU. (Version 24.0) [Online Application Software]. Retrieved from <https://www.spssau.com>.

【2】 SciPy: Scientific Computing Tools for Python (Version 1.9.3). Available at: <https://scipy.org>.

【3】 周俊,马世澎. SPSSAU 科研数据分析方法与应用.第 1 版[M]. 电子工业出版社,2024.