

核密度图

目录

数据格式..... 1

计算步骤..... 2

 1. 数据准备..... 2

 2. 选择核函数..... 2

 3. 选择带宽..... 2

 4. 计算核密度估计..... 2

 5. 绘制图..... 3

核密度图是一种可直观查看数据分布的图形工具，它通过平滑数据点来创建一个连续的概率密度函数，从而帮助我们更直观地理解数据的分布情况。在 SPSSAU 中支持：

- ✓ 批量放入‘标题’进行绘制；
- ✓ 支持放入分类项（比如性别），绘制不同类别情况下的核密度图；
- ✓ 支持选择不同的核函数；
- ✓ 系统自动设置带宽值 h ，同时也支持自定义带宽值 h 。



数据格式

X	data1	data2	data3
A	81.3282	75.93940353	5.239129333
A	28.9825	85.50764657	77.36784448
A	27.8162	53.62511055	58.55069696
A	66.5765	27.41458722	60.26764323
B	78.4136	15.56179381	46.08047493
B	45.5203	27.3542373	87.78188518
B	82.6031	80.5082716	21.16291234
B	53.1135	35.59697641	57.68042561

数据中包括 1 个 Category 以及分析数据（当然分析时可以不放入 Category）。

计算步骤

1. 数据准备

收集一组样本数据 x_1, x_2, \dots, x_n 。

2. 选择核函数

选择一个适当的核函数 $K(x)$ ，满足 $\int_{-\infty}^{\infty} K(x)dx = 1$ 。SPSSAU 提供的核函数包括均匀核（Uniform）、三角核（Triangular）、二次核（Epanechnikov）、四次核（Quartic）、高斯核（Gaussian）和余弦核（Cosine），默认是高斯核（Gaussian）。关于核函数的计算公式，如下：

核函数	公式
均匀核（Uniform）	$K(x) = \begin{cases} \frac{1}{2}, & x \in [-1, 1] \\ 0, & \text{Others} \end{cases}$
三角核（Triangular）	$K(x) = \begin{cases} 1 - x , & x \in [-1, 1] \\ 0, & \text{其它} \end{cases}$
二次核（Epanechnikov）	$K(x) = \begin{cases} \frac{3}{4}(1 - x^2), & x \in [-1, 1] \\ 0, & \text{Others} \end{cases}$
四次核（Quartic）	$K(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2, & x \in [-1, 1] \\ 0, & \text{Others} \end{cases}$
高斯核（Gaussian）[默认]	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in R$
余弦核（Cosine）	$K(x) = \begin{cases} \frac{\pi}{4} \cos \frac{\pi}{2} x, & x \in [-1, 1] \\ 0, & \text{Others} \end{cases}$

3. 选择带宽

选择一个合适的带宽 $h > 0$ ，控制核函数的平滑程度。带宽越小，估计越不平滑；带宽越大，估计越平滑。SPSSAU 默认使用‘Silverman 大拇指法则’法计算 h 值，其计算公式如下：

$$h = \left(\frac{3 \times n}{4} \right)^{-\frac{1}{5}} \times \min \left(s, \frac{IQR}{1.35} \right)$$

其中：

s 是样本标准差

IQR 是四分位距

n 是样本量

\min 为取二者的较小值

4. 计算核密度估计

对于任意 x ，核密度估计 $\hat{f}(x)$ ，定义为：

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

5. 绘制图

将 $\hat{f}(x)$ 对 x 作图,即得核密度图。