

偏度Skewness和峰度Kurtosis如何影响分布

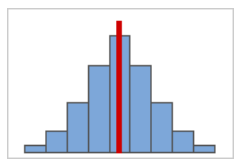
使用偏度和峰度可帮助初步了解数据。

三阶中心矩

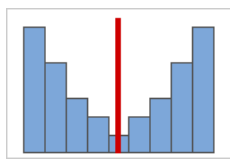
四阶中心矩

偏度 Skewness

偏度指的是数据不对称的程度。偏度值为 0、正数或负数可指示有关数据形状的信息。



图A

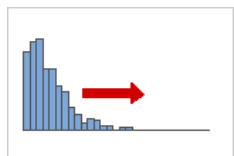


图B

对称或非偏斜分布

在数据变得更加对称时，其偏度值接近于零。如图A显示，根据定义，正态分布的数据显示出相对较小的偏度。通过沿正态数据直方图中间绘制一条线，可看到两侧互相构成镜像。

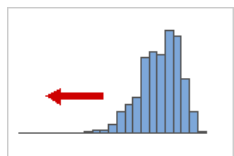
只是无偏并不表示具有正态性。图B显示了一个分布，此分布的两侧仍然彼此构成镜像，但数据远非正态分布。



正或向右偏斜分布

对正偏斜或向右偏斜的数据这样命名是因为分布的尾部指向右侧，且因为其偏度值将大于0（即为正）。

工资数据往往以这种方式偏斜：一个公司中许多员工得到的工资相对较少，而越来越少的人得到非常高的工资。



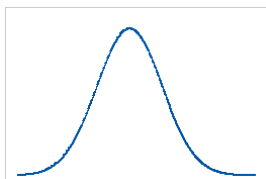
负或向左偏斜分布

左倾斜或负倾斜数据之所以如此命名，是因为分布点的尾部指向左侧，且产生负偏度值。

失效率数据通常向左偏斜。考虑灯泡：极少量的灯泡会立即烧坏，大多数灯泡会持续相当长的时间。

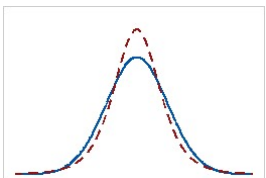
峰度Kurtosis

峰度表示分布的波峰和尾部与正态分布的区别。使用峰度可帮助您初步了解有关数据分布的一般特征。



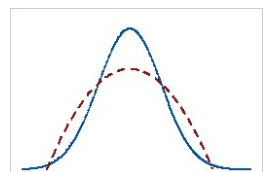
完全服从正态分布的数据的峰度值为 0。
正态分布数据建立峰度的基线。显著偏离 0 的样本峰度可指示该数据不是正态分布。

基线：峰度值 0。



正峰度

具有正峰度值的分配表明，相比于正态分布，该分布有更重的尾部 和更陡的峰值。
例如，遵循 T 分布的数据具有正峰度值。实线表示正态分布，虚线表示具有正峰度值的分布。



负峰度

具有负峰度值的分配表明，相比于正态分布，该分布有更轻的尾部和更平坦的峰值。
例如，遵循 Beta 分布（第一个和第二个形状参数等于2）的数据具有负峰度值。实线表示正态分布，虚线表示具有负峰度值的分布。