

箱线图

箱线图是一种数据的图形表示方法，它可以帮助我们了解**数据的位置、离散程度、观测值中的潜在离群点以及生成观测值的分布的对称性**。

在箱线图中，观测值是沿着纵轴排列的(当然如果为了美观画横向图也可以)。“箱子”的底部绘制在数据的下四分位数水平线上，顶部绘制在数据的上四分位数水平线上。数据的下（或上）四分位数是指其中四分之一的观测值小于（或大于）该值。箱子的宽度是任意的, 因为数据是纵向分布的(如前所述)，箱子中间有一条水平线，表示数据的中位数。中位数是排序后的数据中的中间值。

在箱子的顶部和底部绘制“胡须”。顶部的“胡须”连接到**位于上四分位数的 1.5 倍四分位距内的最大观测值**(注意这里连接到的是一定范围内的最大值)。四分位距是下四分位数和上四分位数之间的距离，即箱子的高度。底部的“胡须”则类似地绘制。超出“胡须”范围的观测值会被单独标出，例如用星号、小圆圈或短横线表示。

下图显示了来自指数分布（参数为 1）、标准正态分布和标准柯西分布的样本的箱线图。指数分布和柯西分布的样本中都有离群点（用“胡须”外的小圆圈表示）。中间的标准正态分布的箱线图显示数据相对于中位数相当对称，且不包含离群点。

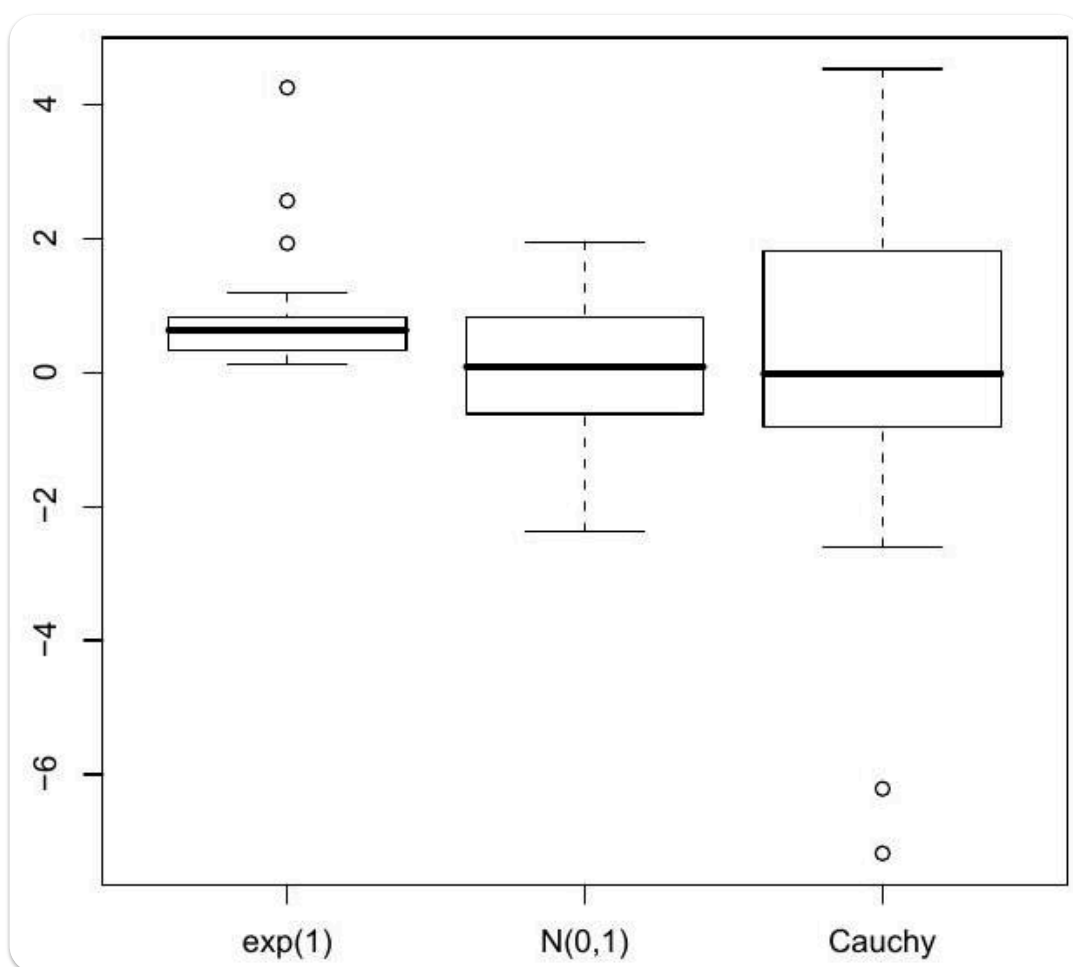


图: 来自标准指数分布 (左)、标准正态分布 (中) 和标准柯西分布 (右) 的样本 (样本量为 20) 的箱线图。