**## 箱线图的正确绘制**

**中文网络上有很多误传的教程，认为箱线图上蓝色的两根线分别是**

**>上界 = Q1 - 1.5IQR**

**>下界 = Q3 + 1.5IQR**

**>**

**其实并不是。**

**请看这张图，注意两个黄色箭头的部分（下面的黄色箭头太小了，所以又用了一个绿色箭头指示它）**

**```python**

**#用python捏一个较为夸张的数据**

**import pandas as pd**

**data = pd.DataFrame([ 54, 65, 72, 76, 104, 432, 457])**

**data.boxplot()**

**```**

**![捏一个较为夸张的数据](https://img-blog.csdnimg.cn/3a5a85c9c7744eb0ae2a0b5493cec8e7.png)**

**如果按上述的公式计算，两个黄色箭头的长度应当是一样的，也就是1.5IQR，但显然不是这样。**

**实际上，箱线图里是没有“上界”、“下界”这种东西的。**

**这两根横线的实际意义是剔除了异常值之后，剩下的数据集里，最大和最小的数据所代表的位置。**

**Q1 - 1.5IQR 和 Q3 + 1.5IQR 这两个公式是用于判断异常值的。**

**也就是，取数据集中的一个值为x**

**$$**

**x = \begin{cases}**

**异常值 &\text{if } x>Q3+1.5\*IQR \space or \space x<Q1 - 1.5\*IQR\\**

**正常值 &\text{else}**

**\end{cases}（\alpha）**

**$$**

**## 画箱线图的正确步骤总结**

**1.计算数据集整体的四分位数，画箱子部分。**

**2.用公式($\alpha$)选出异常值之后，以剩下的最大值最小值为箱线图的两根横线。**

**3.最后再将异常值以圆点的形式画在箱线图上**