

2.5.5 箱体图与茎叶图

除了以直方图、Q-Q 图（或 P-P 图）对变量进行描述外，SPSS 还常常借助箱体图或者茎叶图来描述数据形态。

1. 箱体图与茎叶图简介

(1) 箱体图 (Boxplots)

箱体图，也简称为箱图，即用一个箱状矩形来描述变量分布形态的图形。它以一个内部带有一条横线的矩形代表箱体，在箱体两端有伸出的一段触手，在部分图形触手之外的区域还有一些散点。距离箱体很远的被标记为“*”号，距离箱体较近的标记为“o”号。如图 2-68 所示。

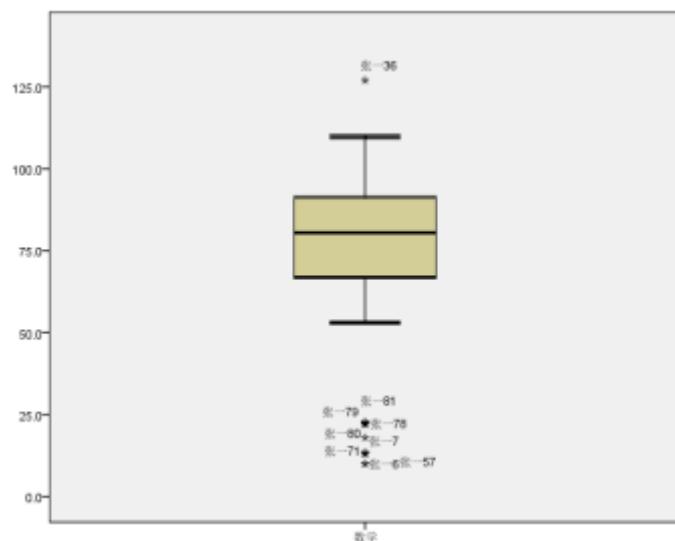


图 2-1 箱体图

在图 2-68 所示的箱体图中，矩形中部的横线实际上是整个数据序列的中位数的取值，箱体上下边缘是以中位数为基准上下延伸 25% 的个案后所构成的区域（即箱体部分包括了数据序列中 50% 的个案）。另外，再以上下边缘为基准向外延伸箱体的 1.5 倍作为其触手。

注意：在绘制箱体图时，触手的长度并不真正地延伸到箱体长度的 1.5 倍，而是以 1.5 倍区域内距离中位数最远的那个样本值作为触手的终点。

凡是观测值位于触手之外但 3 倍箱体之内的个案，被称为奇异值，以“o”作为标记；凡是观测值位于 3 倍箱体之外的个案，被称为极端值，以“*”作为标记。在图 2-68 中，“张一 81”、“张一 79”等人都是极端值，其名字前有“*”标记。

箱体图是对数据序列的四分位数的直观描述。其箱体部分即四分位的 Q3 和 Q1。因此，四分位间距即 Q3-Q1 的值，也就是箱体图中的箱体部分。

利用箱体图，能直观明了地识别数据序列中的异常值，利用箱线图还可以判断数据序列的集中或离散性水平，评价数据的偏态和尾重。

注意：利用箱体图，能够发现数据序列中的奇异值和极端值。在未来的数据分析中，不论相关性分析，还是差异显著性检验，极端值和奇异值都会对分析结果造成严重影响。因此，基于数据序列绘制出的箱体图在未来数据分析中作用很大。

(2) 茎叶图

茎叶图，也被称为枝叶图。它的思路是将序列中的数按位数进行比较，将数的大小基本不变或变化不大的位作为一个主干（茎），将变化大的位的数作为分枝（叶），列在主干的后面，这样就可以清楚地看到每个主干后面的几个数，每个数具体是多少。如图 2-69 所示。

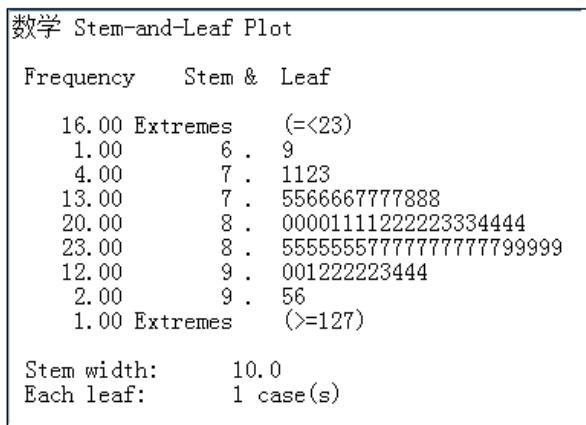


图 2-2 数学成绩的“茎叶图”

从图 2-69 可以看出, 最左边一列数据为“Frequency”, 即频数; 中间一列为“Stem”, 即茎(或主干), 最后一列为“Leaf”, 即叶子。由于茎叶图底部的“Stem width”取值为 10, 表示主干值应为标记值乘以 10。

以图 2-69 中的第 5 行为例“4.00 7. 1123”, 它表示频数为 4、主干为 7; 由于茎宽为 10, 所以实际主干值为 70, 即位于 70~74 分区域的个案有 4 个。叶子部分的值为 1123, 表示 4 个个案的叶子分别为 1、1、2、3, 即个案的原来观测值是 71、71、72 和 73。同理, 可以对第 6 行信息“13.00 7 .5566667777888”进行解释。

茎叶图的作用与直方图类似, 能够直观地看出数据在每个区段的频数, 还可以从整体上看出其分布形态。

2. 在 SPSS 中绘制箱体图与茎叶图的实践案例

(1) 案例要求

对于图 2-56 所示的原始数据文件 stuInfo.sav, 请在 SPSS 中绘制数学成绩的箱体图和茎叶图。

(2) 操作流程

首先, 在 SPSS 中打开 stuInfo.sav 文档, 使之处于“数据视图”状态。

其次, 利用菜单【分析】—【描述统计】—【探索】命令, 以便打开“探索”对话框。如图 2-65 所示。

第三, 在“探索”对话框中, 把变量“数学”从左侧列表移到右侧的“因变量列表”中;

第四, 在“探索”对话框中, 把变量“姓名”从左侧列表移到右侧的“标注个案”列表中;

第五, 在“探索”对话框中, 单击右侧的【绘制】按钮, 启动“探索: 图”对话框。如图 2-66 所示。

第六, 在“探索: 图”对话框中, 从左上角的“箱体”区域选中【按照因子分组】单选框; 从右上角的“描述性”区域中选中【茎叶图】复选框。表示绘制箱体图和茎叶图。

最后, 单击【继续】按钮返回到“探索”对话框中, 然后再单击【确定】按钮, 开始执行绘图过程。系统制作的图形将显示在新的输出界面中。

最终得到的箱体图和茎叶图如图 2-68 和如 2-69 所示。

(3) 补充说明

①在以“探索”模块绘制箱体图和茎叶图的过程中, 为了能在箱体图上清晰地标记出奇异值和极端值个案, 专门选择了“姓名”列作为标注个案。这样, 在箱体图中会显示出奇异值和极端值对应的个案的姓名。如果没有提供标注个案的变量, 则在箱体图的奇异值和极端值位置显示若干数码, 这些数码是个案在数据表中的自然序号。

②在“探索”对话框中, 人们可以把某个定类变量或定序变量添加到“因子列表”之中。比如, 若把“性别”添加到因子列表中, 则表示对待分析变量按照性别分组后再进行分析, 将会分别得到男生、女生的箱体图和茎叶图。

③在 SPSS 中, 还可以借助菜单【图形】—【旧对话框】—【箱图】/【直方图】/【散点图】命令来绘制各类统计图。