在SPSS统计分析中你最容易忽略的一件事

作者：张华 赵一鸣 来源：临床流行病学和循证医学

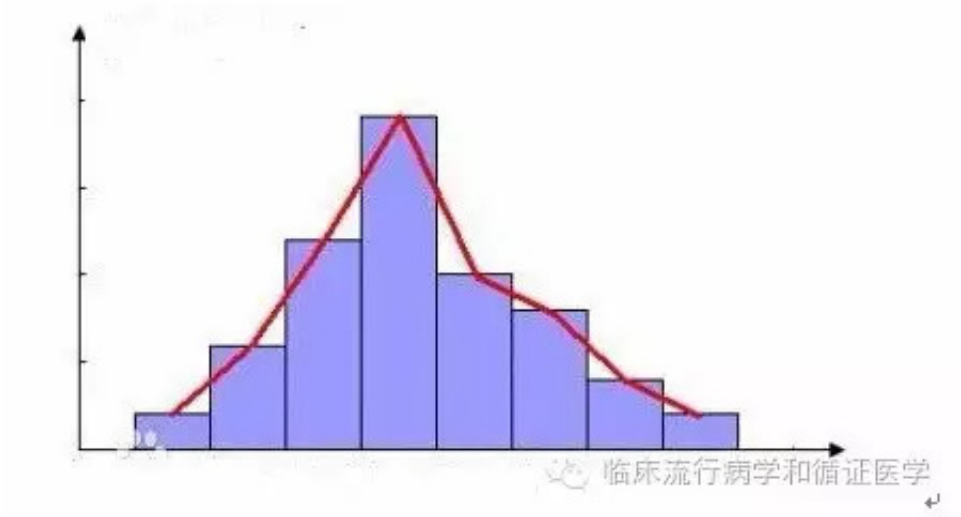
本文原地址：https://www.iikx.com/news/statistics/1930.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

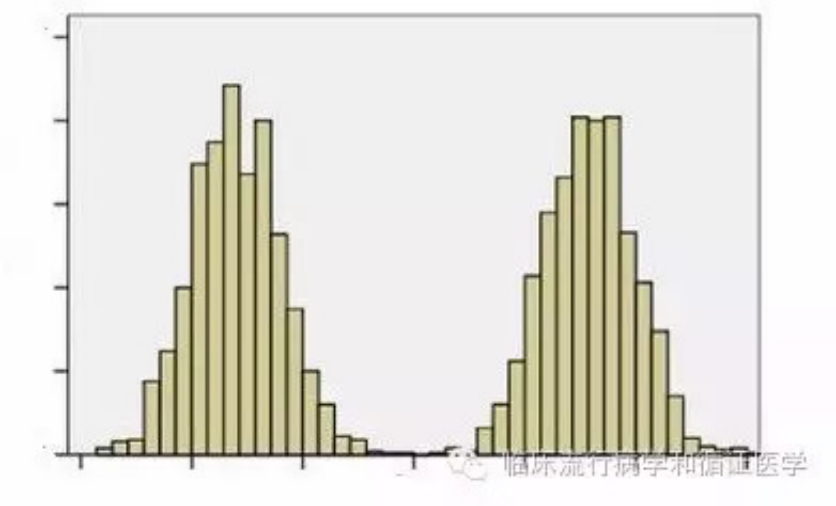
在SPSS统计分析时，我们一般是先对数据进行描述，包括连续性数据的集中趋势(计算均值、中位数和离散趋势(标准差、四分位数、极值)或分类变量的百分比等，再进行单因素统计检验和多因素分析。看似正确无误的流程，但却忽略了非常重要的工作，使用图形查看数据的分布情况。

单变量的数据描述首推直方图。直方图能看出什么呢?

首先从直方图里可以查看数据的分布，包括正态分布、均匀分布、逆高斯分布等，这关系到后续的数据描述和统计检验，如果是正态分布，当然可以均值、标准差进行描述，使用t检验、方差分析等方法进行检验，不符合正态分布只能使用中位数进行描述，使用非参数统计方法。有人会说，是否正态分布可以使用Kolmogorov-Smirnov检验，可我却不幸地告诉您，这个方法不是一个“好”方法，当样本量少于10时，即使不是正态分布，也检测不出来，而当样本量大于100时，即使是正态分布，也容易当成不是正态分布。到现在为止，图形仍然是判断是否正态分布的最好方法，这里的图形包括直方图、茎叶图、p-p图、Q-Q图等。

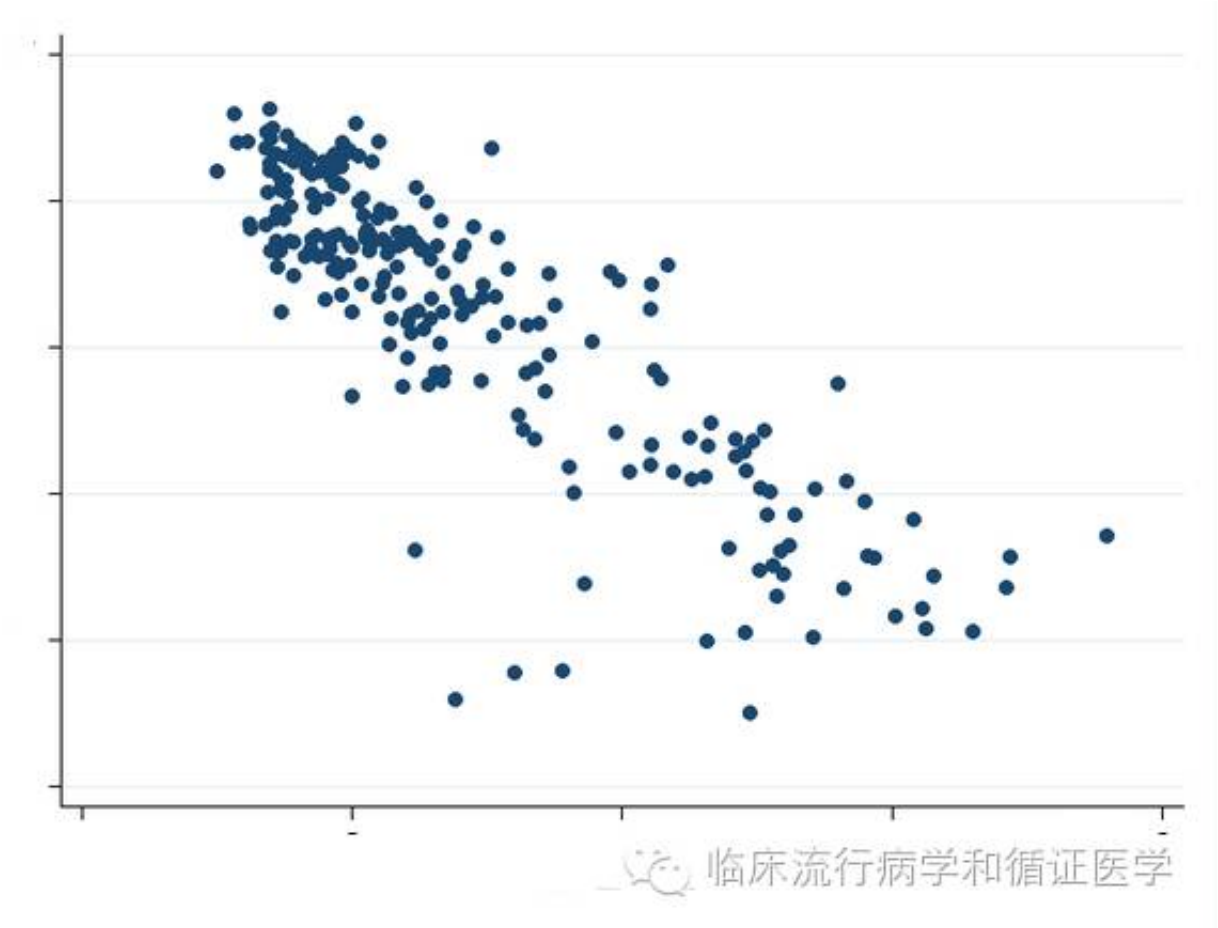


其次从分布里可以看出是单峰分布还是多峰分布。一般情况下一个变量是单峰分布，如果你遇到下面的双峰分布，你就要查看数据是否可以成亚组。如果下图是原发性高血压的晨起血压分布，那可能要恭喜你，也许这是个重大发现，接下来你仔细核查病人情况，也许你可能把病人分成不同的亚型，如对某种药的敏感和不敏感型。

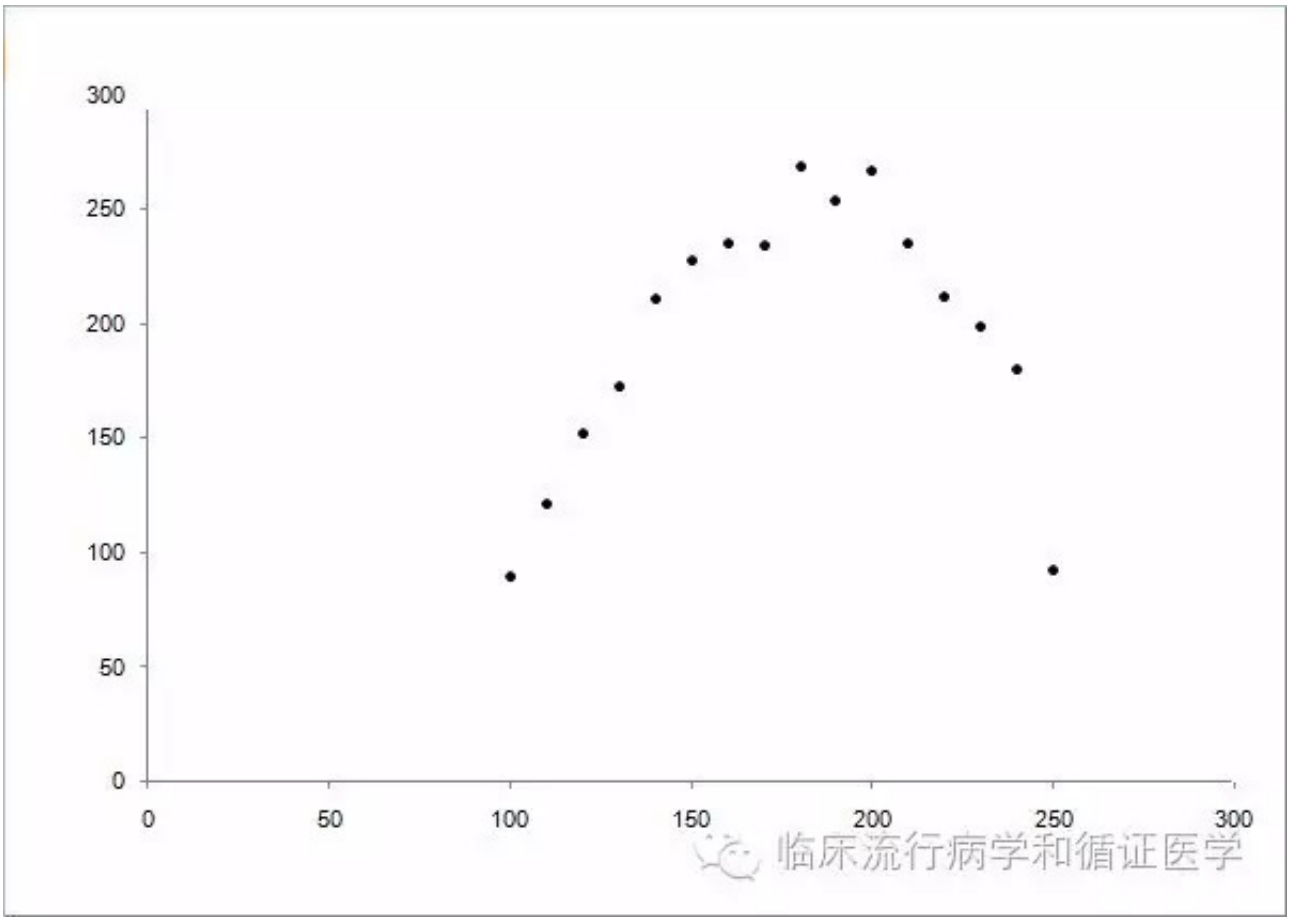


两个变量的图形描述首推散点图，从散点图看到的信息也有助于我们进一步的数据分析。

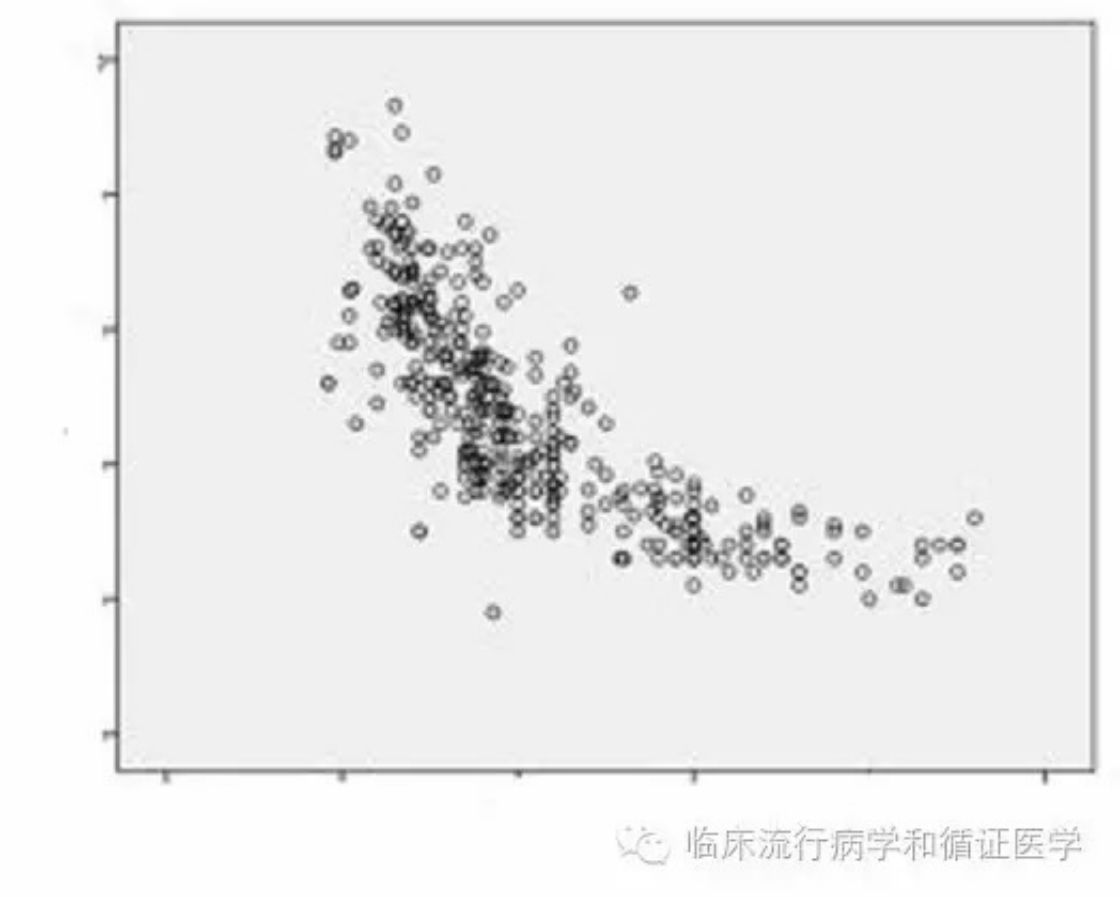
在做相关分析前，查看散点图是第一步，我们期望散点图是这样的。



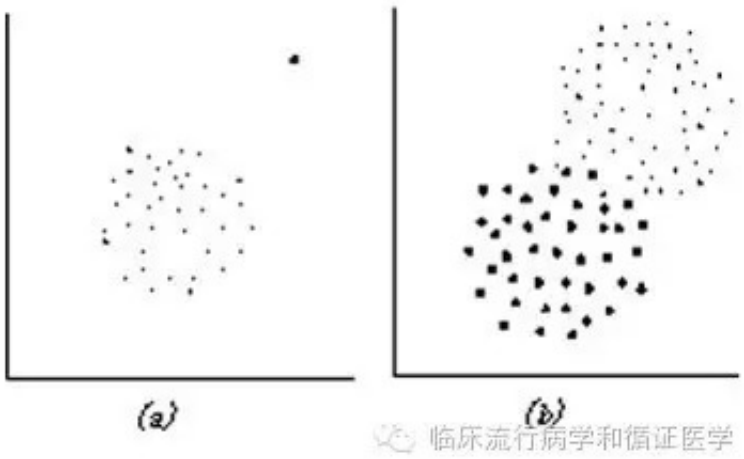
但是也可能是这样的，
有可能两个变量是抛线线函数关系，我们就需要将一个变量取对数后再进行相关分析。



也有可能是下图关系，
提示我们有可能是双曲线函数关系，我们需要将其中一个变量取倒数后进行相关分析。



当然也有可能是下面两种关系。如果你的图形是左侧形态，那就是对特殊点进行数据核查，是否存在错误;如果是右侧形态，
有可能数据可以分亚组，可能需要分层后
进行相关分析，如何分层需要进行数据核查，这也许成为你的新的发现哦。



总之，在SPSS统计分析中你最容易忽略的一件事：在进行数据分析前，要对变量进行图形描述，然后选择合适的数据分析方法，这一步不会在文章中体现，但却非常重要，是数据分析的良心活。另外很多重要的发现也是从数据描述中得到的，在数据分析前查看数据分布，也许会给你很大

惊喜和收获。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发