



学一点数据分析-简单方法

第三节

二、简单实用的数据分析方法

1□对比分析法

对比分析法是指将两个或两个以上的数据进行比较，分析它们的差异，从而揭示这些数据所代表的事物发展变化情况和规律性。

对比分析法分为横向比较和纵向比较两种。横向比较是在同一时间条件下对不同总体指标的比较，这种比较方式也被称为静态比较；纵向比较是在同一总体条件下对不同时期指标数值的比较，也被称为动态比较。之前提及的“环比”和“同比”都属于纵向比较，是一种动态比较。

2□分组分析法

分组分析法指为了深入总体的内部，根据数据分析对象的特征，按照一定的标准（指标），把数据分析对象划分为不同的部分和类型来进行研究，以揭示其内在的联系和规律性。

分组分析法包括等距分组和不等距分组两种。在各单位数据变动较为均匀的情况下可以采用等距分组，反之则用不等距分组更佳。

二、简单实用的数据分析方法

3□ 结构分析法

指被分析总体内的各部分与总体之间进行对比的分析方法，即总体内各部分占总体的比例，属于相对指标。一般某部分的比例越大，说明其重要程度越高，对总体的影响越大。

4□ 平均分析法

运用计算平均数的方法来反映总体在一定时间、地点条件下某一数量特征的一般水平。平均指标可用于同一现象在不同地区、不同部门或单位间的对比，还可用于同一现象在不同时间内的对比。

5□ 综合评价分析法

综合评价分析法是一种多变量评价分析，运用多个指标对多个参评单位进行评价的方法。它的基本思路是将多个指标转化为一个能够反映综合情况的指标来进行分析评价。

二、简单实用的数据分析方法

6□相关分析法

相关分析法即通过相关表或相关图等方式来判断两个变量是否存在不确定的统计关系的数据分析方法。在数据分析中，将变量间确定性的统计关系称为函数关系，不确定性的统计关系称为相关关系。

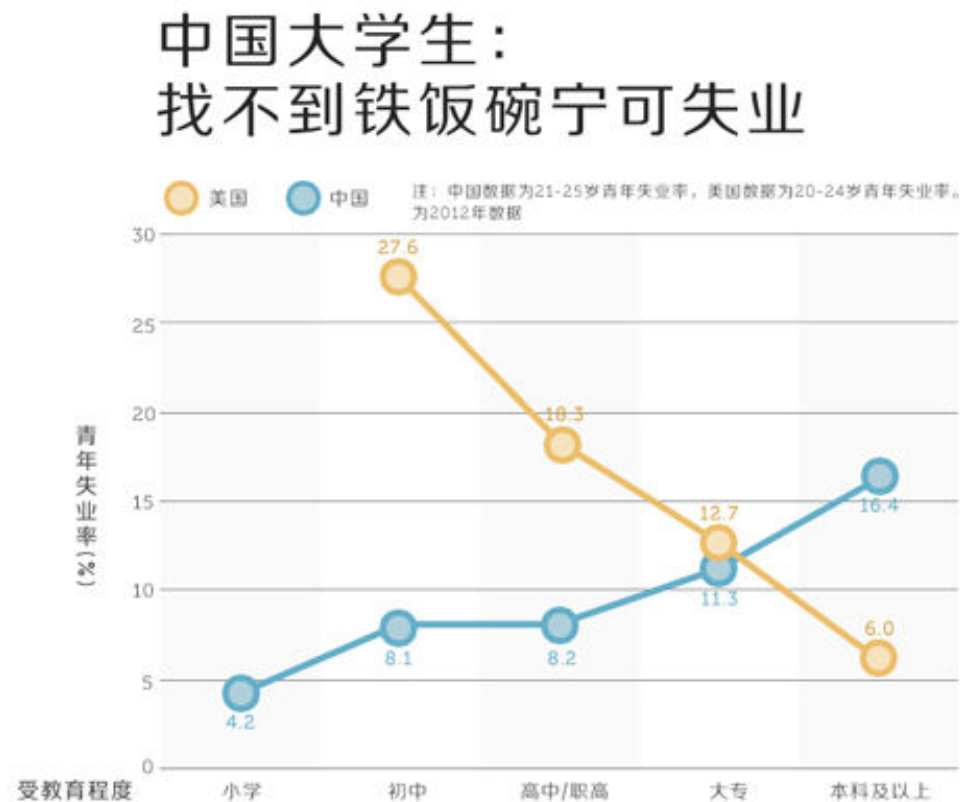
我们常见的相关关系多为两个变量之间的相关关系，可称为单相关（或称一元相关），与之相对，三个或三个以上变量之间的相关关系则属于复相关（或称多元相关）。当一个变量的值增加或减少时，另一个变量的值也随之增加或减少，这样的相关关系称为“正相关”，反之则是“负相关”。当变量之间的相关关系的散点图中的点接近一条直线时，称为“线性相关”；如果变量之间的相关关系的散点图中的点接近于一条曲线时，则称为“非线性相关”。

相关分析法中的相关表是指将变量的若干变量值从小到大按顺序排列，并将另一变量的值与之对应排列而形成的统计表。相关图又称“散点图”，指用直角坐标系的横轴表示一个变量，纵轴表示另一个变量，将两个变量间相对应的变量值用坐标点的形式描绘出来，用以表明相关点分布的图形。

案例：网易数独栏目《青年就业怪现象：没好工作宁愿待业》

2013年7月15日，网易“数读”栏目报道《青年就业怪现象：没好工作宁愿待业》中就运用了相关分析法来说明中美两国青年失业率和受教育程度之间的关系。在中、美两国，青年的失业率和受教育程度均存在线性相关关系，但有趣的是，通过相关图可见，中国青年的失业率和受教育程度之间存在的是正相关关系，而美国青年的失业率和受教育程度之间存在的却是负相关关系。也就是说，美国青年受教育程度越高，其失业率越低，而中国青年受教育程度越高，失业率不降反升。

值得一提的是，编辑用“中国大学生：找不到铁饭碗宁可失业”这一标题来对上述“怪现象”给予解释，这个判断略显主观武断，因为与配图一起刊出的报道中也提及“彼得森经济研究所认为造成中国这一怪异现象的原因主要在于教育结构与劳动力需求的错位以及青年人扭曲的就业预期”。可见，中国出现这种奇怪的“正相关”关系并不只是青年自身的原因。



Bbc 's classification

不同类型数据故事的“类别模型”

测量

最精简的新闻故事；计数与求和：“去年，全国各地的地方议会总共在采购回形针上花了X万亿英镑。”但通常这样一个笼统的数字很难让人明白到底是花多了还是花少了。因此，你得把数字放进特定的语境中——比如，可以运用：

比例

“去年，全国的地方议会在回形针上的支出占到全部文具预算的三分之二。”

国内比较 “地方议会在回形针上的支出多过为空巢老人送餐到家服务上的支出。

”海外比较 “去年，议会在回形针上的支出是国家海外救援预算的两倍。”

随时间变化

“四年来，议会在回形针上的开销增长了两倍。”

“排名表”

分类分析法 “紫党政务委员会用于购买纸夹的开销比黄党的多出50%。”或者你可以用数字把各个因素联系在一起：

关联比较法 “接受过文具用品公司捐款的那些政务委员会用在纸夹的开销更大，平均每一英镑的捐款，开销增长100英镑。

— 马丁·罗森鲍姆(Martin Rosenbaum), BBC

Exercise and Introduction : google trend data analysis example

How to get to the same scale?

How to visualize?

How to analyze?

Use python to do statistic analysis

