

数据统计分析案例

2022.02.29



1.1 对比分析

- 对比分析法是将两个或两个以上的数据进行比较，分析其中的差异，从而揭示这些事物代表的发展变化情况和规律性。
- 特点:非常直观地看出事物某方面的变化或差距，而且可以准确、量化地表示出变化的差距是多少。
- 对比分析法通常是把两个相互联系的指标数据进行比较，从数量上展示和说明研究对象规模的大小水平的高低、速度的快慢，以及各种关系是否协调。对比分析一般来说有以下几种对比方法:纵向对比、横向对比、标准对比、实际与计划对比。
- 见notebook案例：对比分析个品牌销量表现TOP10

同比、定比和环比分析

- 在数据分析中，有一个重要的分析方法，叫趋势分析法：
- 将两期或连续数期报告中某一指标进行对比，确定其增减变动的方向、数额和幅度，以确定该指标的变动趋势。趋势分析法中的指标，有同比分析、定比(定基比)分析和环比分析，以及同比增长率分析、定比(定基比)增长率分析和环比增长率分析。

同比、定比和环比概述

- 同比：本期数据与历史同期数据比较。例如：2020年2月份与2019年2月份相比较。
- 定比：本期数据与特定时期的数据比较。例如，2020年2月与2019年12月份相比较。
- 环比：本期数据与上期数据比较。例如，2020年2月份与2020年1月份相比较。
- 同比的好处是可以排除一部分季节因素；环比的好处是可以更直观地表明阶段性的变换，但是会受季节性因素影响；定比常用于财务数据分析。

下面来看一个生活中经常出现的场景：

- 同比:去年这个这时候**100**元钱能买**10**斤猪肉，今年这个时候只能买**5**斤。
- 定比: 年龄**50**岁是**25**岁的两倍。
- 环比: 这个月好像比上个月胖了。
- 下面简单介绍一下同比、定比和环比计算的公式：

- 1.同比:

- 同比=本期数据/上年同期数据
 - 同比增长率=【(本期数-同期数)/同期数】*100%
-

- 2.定比:

- 定比=本期数据/固定数据
- 定比增长率=【(本期数据-固定期数据)/固定期数据】*100%

- 3.环比:

- 环比增长率反应本期比上期增长了多少, 公式如下:
- 环比增长率=【(本期数-同期数)/上期数】*100%
- 环比发展速度是本期水平与上一期水平之比, 反应前后两期的发展变化情况,
- 环比发展速度=(本期数/上期数)*100%
- 环比增长速度=环比发展速度-1
- 见案例

贡献度分析（帕累托法则）

- 贡献度分析又称**80/20**法则、二八法则、帕累托法则、帕累托定律、最省力法则或不平衡原则。该法则是由意大利经济学家“帕累托”提出的。**80/20**法则认为:原因和结果、投入和产出、努力和报酬之间本来存在着无法解释的不平衡。例如, 一个公司**80%**的利润常常来自**20%**的产品, 那么使用贡献度分析就可以分析获利最高的**20%**的产品。
- 下面简单介绍一下 贡献率相关算法:

$$\text{累计贡献率(\%)} = (\text{累加销售收入} / \text{销售总收入}) \times 100\%$$

- 通过上述公式得出累计贡献率，当累计贡献率接近**80%**时(不定正好是**80%**)，然后找到该产品在图表中相应的位置并进行标注。

图书编号	
B4	0.307463
B5	0.445179
B3	0.51233
B8	0.572204
B22	0.631536
B9	0.680833
B16	0.725675
B6	0.764115
B1	0.796161
B13	0.817665
B25	0.836678
B2	0.854559
B15	0.871132
B10	0.88764
B18	0.900241
B17	0.91184
B12	0.923435

以上产品累计贡献率已经接近总销售收入的**80%**，所以这部分产品应作为主打产品，重点营销。

案例：产品贡献度分析：

下面分析淘宝电商全彩熙烈图书2018年上半年销售收入占比**80%**的产品。

使用公式计算出产品累计贡献率，可以看出，到编号**B13**，累计贡献度达到**0.817665**，（接近总销售收入的**80%**），其中共有**10**种产品，接下来在图表中进行标注：

差异化分析

- 任何事物都存在差异，如同上课听讲，有人津津有味，有人昏昏欲睡。
-

那么，通过差异化分析，比较不同事物之间在某个指标上存在的差异，根据差异定制不同的策略。对于产品而言，差异化分析是指企业在其提供给顾客的产品上，通过各种方法满足顾客的偏好，使顾客能够把它同其他竞争企业提供的同类产品有效地区别开来，从而使企业在市场竞争中占据有利的地位。比较常见的有性别差异、年龄差异。通过差异化分析比较不同性别之间在某个指标上存在的差异，通过分析结果对不同性别定制不同的方案。例如，分析不同性别的同学在学习成绩上的差异，了解男生和女生之间的这些差异，因材施教，定制不同的弥补弱项的方案。对于女生，可以有意识地培养她的思维能力;而对于男生，可以买些书籍，来增强他薄弱的方面。

年龄差异化分析，了解不同年龄的需求，投其所好，使企业的利润最大化。例如，网购、自媒体、汽车、旅游等行业，通过年龄差异化分析，找出不同年龄段用户群体的喜好，从而增加产品销量。



相关性分析

- 任何事物之间都存在一定的联系。例如，夏天温度的高低与空调的销量就存在相关性。当温度升高时，空调的销量也会相应提高。

相关性分析是指对多个具备相关关系的数据进行分析，从而衡量数据之间的相关程度或密切程度。相关性可以应用到所有数据的分析过程中。如果一组数据的改变引发另一组数据朝相同方向变化，那么这两组数据存在正相关性，例如，身高与体重，一般个子高的人体重会重一些，个子矮的人体重会轻一些；如果一组数据的改变引发另一组数据朝相反方向变化，那么这两组数据存在负相关性，例如，运动与体重。

时间序列分析

- 顾名思义，时间序列就是按照时间顺序排列的一组数据序列。时间序列分析就是找出数据变化发展的规律，从而预测未来的走势。

时间序列分析有以下几种表现形式：

长期趋势变化: 受某种因素的影响，数据依据时间变化，按某种规则稳步增长或下降。使用的分析方法有移动平均法、指数平滑法等。

季节性周期变化: 受季节更替等因素影响，数据依据固定周期规则性的变化。季节性周期变化，不局限于自然季节，还包括月、周等短期周期。例如，空调、羽绒服、冷饮的销售，双十一、双十二流量在一周之内 的波动等。采用的方法为季节指数。

循环变化: 指一种较长时间的上、下起伏周期性波动，一般循环时间为 2~ 15 年。

- 随机性变化: 由许多不确定因素引起的数据变化，在时间序列中无法预计。