9种常见数据分析方法

数据分析是从数据中提取有价值信息的过程，过程中需要对数据进行各种处理和归类，只有掌握了正确的数据分类方法和数据处理模式，才能起到事半功倍的效果，以下是数据分析员必备的9种数据分析思维模式：

**1. 分类**

分类是一种基本的数据分析方式，数据根据其特点，可将数据对象划分为不同的部分和类型，再进一步分析，能够进一步挖掘事物的本质。

**2. 回归**

回归是一种运用广泛的统计分析方法，可以通过规定因变量和自变量来确定变量之间的因果关系，建立回归模型，并根据实测数据来求解模型的各参数，然后评价回归模型是否能够很好的拟合实测数据，如果能够很好的拟合，则可以根据自变量作进一步预测。

**3. 聚类**

聚类是根据数据的内在性质将数据分成一些聚合类，每一聚合类中的元素尽可能具有相同的特性，不同聚合类之间的特性差别尽可能大的一种分类方式，其与分类分析不同，所划分的类是未知的，因此，聚类分析也称为无指导或无监督的学习。

数据聚类是对于静态数据分析的一门技术，在许多领域受到广泛应用，包括机器学习，数据挖掘，模式识别，图像分析以及生物信息。

**4. 相似匹配**

相似匹配是通过一定的方法，来计算两个数据的相似程度，相似程度通常会用一个是百分比来衡量。相似匹配算法被用在很多不同的计算场景，如数据清洗、用户输入纠错、推荐统计、剽窃检测系统、自动评分系统、网页搜索和DNA序列匹配等领域。

**5. 频繁项集**

频繁项集是指事例中频繁出现的项的集合，如啤酒和尿不湿，Apriori算法是一种挖掘关联规则的频繁项集算法，其核心思想是通过候选集生成和情节的向下封闭检测两个阶段来挖掘频繁项集，目前已被广泛的应用在商业、网络安全等领域。

**6. 统计描述**

统计描述是根据数据的特点，用一定的统计指标和指标体系，表明数据所反馈的信息，是对数据分析的基础处理工作，主要方法包括：平均指标和变异指标的计算、资料分布形态的图形表现等。

**7. 链接预测**

链接预测是一种预测数据之间本应存有的关系的一种方法，链接预测可分为基于节点属性的预测和基于网络结构的预测，基于节点之间属性的链接预测包括分析节点资审的属性和节点之间属性的关系等信息，利用节点信息知识集和节点相似度等方法得到节点之间隐藏的关系。与基于节点属性的链接预测相比，网络结构数据更容易获得。复杂网络领域一个主要的观点表明，网络中的个体的特质没有个体间的关系重要。因此基于网络结构的链接预测受到越来越多的关注。

**8. 数据压缩**

数据压缩是指在不丢失有用信息的前提下，缩减数据量以减少存储空间，提高其传输、存储和处理效率，或按照一定的算法对数据进行重新组织，减少数据的冗余和存储的空间的一种技术方法。数据压缩分为有损压缩和无损压缩。

**9. 因果分析**

因果分析法是利用事物发展变化的因果关系来进行预测的方法，运用因果分析法进行市场预测，主要是采用回归分析方法，除此之外，计算经济模型和投人产出分析等方法也较为常用。

以上是数据分析员应熟练掌握的9种数据分析思维方法，数据分析员应根据实际情况合理运用不同的方法，才能够快速精确的挖掘出有价值的信息！