

第五章 风险评估——聚类分析

目录

CONTENT

第一节 聚类分析基础

第二节 实战演练——风险评估分析程序

第一节

聚类分析基础



一、聚类分析的基本概念和常用距离

● (一) 聚类分析的基本概念

聚类分析(Cluster Analysis)是基于分析对象的特征,按照一定标准对分析对象进行分类的一种无监督式分析方法。它可以将分析对象划分为若干组,使得组内分析对象具有最高相似度,且组间分析对象差异较大。

聚类分析是将研究对象分为相对同质的群组的统计分析技术。是一种探索性的分析,在分类的过程中,人们不必事先给出一个分类的标准,聚类分析能够从样本数据出发,自动进行分类。

● (2) 聚类分析的常用距离

聚类分析要求组内差异极小,组间差异较大,而个体间差异程度用距离表示,聚类分析中常见的距离有欧氏距离、切贝谢夫距离、布洛克距离、明考斯基距离、夹角余弦距离等等。

二、聚类分析的方法

- 按分类对象不同，聚类分析可以分为样本聚类（Q型聚类）和变量聚类（R型聚类）。
- 按分析方法不同，聚类分析可以分为K-means聚类、层次聚类、两阶段聚类等。其中，最为主流的聚类分析方法为K-means聚类和层次聚类。

（一）K-means聚类

K-means聚类又称K-均值聚类、快速聚类，由波兰数学家Steinhaus于1956年提出。是一种“自上而下”的聚类方法，使用时需要人工预先指定样本中的聚类数目。

（二）层次聚类

层次聚类又称分层聚类、系统聚类。与K-means聚类不同，层次聚类是一种“自下而上”的聚类方法。它无需预先指定聚类数目，而是逐层合并至最后一层后，根据需要选择聚类数目。

三、大数据审计的聚类分析实现

(一) 了解被审计单位及其环境的聚类分析实现

1. 了解治理环境的聚类分析实现



注册会计师在承接审计业务后，需要对被审计单位的治理环境进行全面了解，以评判被审计单位是否存在道德风险。注册会计师对治理环境的了解应当是客观、全面的。以往注册会计师对被审计单位治理环境的了解主要基于定性分析和经验判断，这可能会导致由于判断不准确而增加的审计风险。聚类分析技术可以从定量分析层面帮助注册会计师更加深入地了解被审计单位，从而提升后续审计工作的效率和质量。



例如，董事会成员构成是公司治理的重要方面，基于董事会成员多样性形成的董事会断裂带内嵌于公司治理结构中。董事会成员的年龄、性别、职责、背景、任期等特征的排列组合无疑会成为内部成员构成发生变化的主要因素，这将对董事会群体的认知产生影响，进而影响企业行为。一般而言，较强的董事会断裂带表明董事会成员间制衡力度较强；而较弱的董事会断裂带容易形成相互包庇的治理环境。利用聚类分析技术可以将董事会成员划分至不同组别中，并进一步计算出被审计单位董事会断裂带强弱，从而为注册会计师了解被审计单位治理环境提供数据支持。

三、大数据审计的聚类分析实现

(一) 了解被审计单位及其环境的聚类分析实现

2. 选取可比公司的聚类分析实现

1

在风险导向审计模式下，注册会计师需要实施风险评估程序，以识别和评估财务报表层次和认定层次的重大错报风险。风险评估程序主要包括询问管理层和被审计单位内部其他人员、分析程序、观察和检查，其目的是了解被审计单位及其环境。

2

实务中，注册会计师为确定被审计单位财务报表中数据是否合理，通常会寻找一些合适的标杆，将被审计单位数据与标杆进行横向对比，根据两者间的差异来推断被审计单位是否存在潜在错报风险。标杆可以是与被审计单位所处行业、规模、财务状况、商业模式较为接近的一家或若干家企业。聚类分析技术可以用于寻找被审计单位的可比企业。

3

例如，审计人员可以通过设置行业、规模、财务状况、商业模式的代理指标作为聚类指标，将被审计单位和同行业上市公司进行聚类分析。由于聚类后同一组别内的样本非常相似，可以把和被审计单位处于同一组别的上市公司作为可比企业。如果进行一次聚类后发现被审计单位所处组别内样本过多或不符合寻找条件，则可以采用相同聚类算法进行多次聚类，直至找出最合适的标杆。

三、大数据审计的聚类分析实现

(一) 了解被审计单位及其环境的聚类分析实现

3. 识别特殊业务的聚类分析实现



聚类分析技术可以将对象划分至具有相似特征的组别，使得同一组别内的对象呈现出一定特性。注册会计师可以利用被审计单位会计信息系统中的脱敏数据进行聚类分析，使得具有相似特征的业务被划分至同一组别。



一般而言，同一企业内部相关交易或事项的变动具有规律性，如果个别业务数据单独被划分为一组，说明该业务与其他业务存在明显不同，需要特别关注，该业务也可以被标识为特殊业务（或孤立点）。这些特殊业务可能是问题业务或者有在未来可能成为问题业务的风险。进一步地对特殊业务进行深入考察，可以凝练出被审计单位的潜在问题，并对后续审计工作提供关键线索。

三、大数据审计的聚类分析实现

(二) 了解被审计单位内部控制的聚类分析实现

1. 风险预警的聚类分析实现

内部审计的一个重要目的是通过审计活动规范企业经营行为，增强企业合规性，降低企业面临的风险。企业日常经营过程中会产生大量数据，对这些数据进行归纳整理并深度挖掘，可以从中发现许多规律，这有助于企业预防可能发生的风险。从某种意义上看，聚类分析技术属于数据挖掘的一种，将其应用在事前监控，可以发现潜在问题，降低企业风险。考虑到聚类分析技术具有分离孤立点的特性，可以在实际工作中设置相关的控制阈值，当孤立点达到设定的控制阈值时，则可以根据智能决策规则或人工介入来形成审计预警方案。在问题爆发之前就采取有效的防范措施，可以极大减少企业损失。

三、大数据审计的聚类分析实现

(二)了解被审计单位内部控制的聚类分析实现

2.存货流转审计的聚类分析实现

企业的生产经营涉及存货的出库与入库，在这个过程中可能存在存货损耗。企业生产过程中出现正常的存货损耗是可以接受的，但也可能因为一些工作疏忽使得存货被藏匿在某些地方。逻辑上，被藏匿的存货存在时间和空间上的异常：一方面，被藏匿的存货在仓库台账中存在的时间比实际使用存货的存在时间长；另一方面，被藏匿的存货散落在各个生产环节，这些存货的出现次数高于实际使用存货的出现次数。基于上述特征，可以使用聚类分析技术对存货数据进行分组，计算特定规制内的库龄长且出现次数多的存货，确定其明细特征，从而评估存货流转是否存在问题。

三、大数据审计的聚类分析实现

(二)了解被审计单位内部控制的聚类分析实现

3.异常客户识别的聚类分析实现

一些特殊制造服务类企业（比如供电、供水、燃气公司）提供的产品覆盖面极广、数量庞大，产品的核算计费是根据计费设施的显示数据确定的。在这个过程中会经常发生少计、漏记的情况。利用聚类分析技术，可以帮助企业对业务数据进行分析，帮助企业识别异常客户。

例如，供水公司可以利用聚类分析挖掘重度用水用户中的疑点用户，并进一步调查重度用水用户的异常情况是否是水管漏水、偷水或其他情况导致的。

第二节

实战演练——风险评估分析程序

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/197024010126006026>