|  |  |
| --- | --- |
| ICS |  |
| CCS | 点击此处添加CCS号 |

|  |
| --- |
| 43 |

湖南省地方标准

DB 43/T XXXX—XXXX

机车车辆计量结果数字化处理规范

Specification for digitization processing of rolling stock metrology results

（本草案完成时间：2024年10月23日）

XXXX - XX - XX发布

XXXX - XX - XX实施

湖南省市场监督管理局  发布

目次

[前言 II](#_Toc180664023)

[1 范围 1](#_Toc180664024)

[2 规范性引用文件 1](#_Toc180664025)

[3 术语和定义 1](#_Toc180664026)

[4 数字化处理流程 1](#_Toc180664027)

[5 数据采集 1](#_Toc180664028)

[6 数据预处理 2](#_Toc180664029)

[7 数据存储 2](#_Toc180664030)

[8 数据分析 2](#_Toc180664031)

[9 数据可视化 2](#_Toc180664032)

[附录A（资料性） 机车车辆计量结果数字化处理流程 1](#_Toc180664033)

1. 前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由湖南省市场监督管理局提出并归口。

本文件起草单位：湖南省计量检测研究院、湖南科技大学、中车株洲电力机车有限公司、中南大学、湖南省质量和标准化研究院。

本文件主要起草人：

机车车辆计量结果数字化处理规范

* 1. 范围

本文件确立了机车车辆计量结果数字化处理流程，规定了数据采集、数据预处理、数据存储、数据分析、数据可视化等要求。

本文件适用于机车车辆计量结果数字化处理。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271.1 信息技术 词汇 第1部分：基本术语

DA/T 31 纸质档案数字化规范

JJF 1001 通用计量术语及定义

DB43/T XXXX 机车车辆计量数据管理规范

* 1. 术语和定义

GB/T 5271.1、DA/T 31、JJF 1001界定的以及下列术语和定义适用于本文件。

数字化 digitization

用计算机技术将模拟信号转换为数字信号的处理过程。

[来源：DA/T 31-2017，3.1]

数据 date

信息的可再解释的形式化表示,以适用于通信﹑解释或处理。

1. 可以通过人工或自动手段处理数据。

[来源：GB/T 5271.1-2000, 01.01.02]

* 1. 数字化处理流程

机车车辆计量结果数字化处理流程见图A.1。

* 1. 数据采集

数据采集前，应确认测量仪器处于校准状态，并记录相应的校准结果。

按安装流程安装测量仪器，进行多次测量确保数据稳定性，并应记录测量时间、地点以及测量参数。

应对采集的数据进行初步检查，检查数据是否存在缺失值、异常值，并记录数据检查结果。

* 1. 数据预处理

异常值宜采用删除、替换、离群值检测等方法处理，其中离群值检测宜采用箱线图、聚类等统计方法处理，数据单位应采用国际基本单位或其组合。

重复值宜采用删除、合并、标记等方法处理，其中合并可采用相似度匹配、聚类等方法处理。

缺失值宜采用填充缺失值、删除缺失值或进行插值处理，插值可采用线性插值、多项式插值、样条插值、最近邻插值等方法进行处理。

不一致值宜采用规范化、转换、匹配等方法处理，其中规范化可采用大小写转换、数据类型转换等方法处理。

格式不一致宜采用规范化、转换、格式化等方法处理，其中规范化可采用格式化字符串、正则表达式等方法处理。

应进行数据完整性检查。

* 1. 数据存储

数据存储应由数据库选择与匹配、数据库连接、数据导入三部分构成，数据库选择与匹配应符合DB43/T XXXX（机车车辆计量数据管理规范）。

数据库连接应提供常用的数据开放互连接口，并应对数据库连接的稳定性进行测试。

应将数据预处理得到的数据导入数据库中。

* 1. 数据分析

应对数据进行筛选，选择需要分析的数据。

连续信号和随机信号宜使用统计特征分析，连续信号还宜使用频域特征、时域特征等信号处理方法。

多源高维的数据挖掘中宜采用下列方法：

1. 假设检验：用于检验数据之间是否有显著差异，判断是否拒绝某个假设；
2. 回归分析：用于分析自变量和因变量之间的关系，建立回归模型，预测因变量的变化趋势；
3. 聚类分析：将数据分成若干类别，同一类别内的数据相似度较高，不同类别之间的数据相似度较低；
4. 主成分分析：将多个变量合并成少数几个新变量，保留原始变量的大部分信息，以减少维度；
5. 时间序列分析：用于分析时间序列数据，找出趋势、季节性和周期性等规律，预测未来的变化趋势；
6. 关联规则挖掘：从数据中发现频繁出现的模式和关联关系；
7. 决策树分析：基于样本数据建立决策树模型，通过选择属性值，逐步筛选出目标属性预测值；
8. 神经网络分析：用人工神经网络模拟人脑处理信息的过程，进行数据建模和预测等操作。
   1. 数据可视化

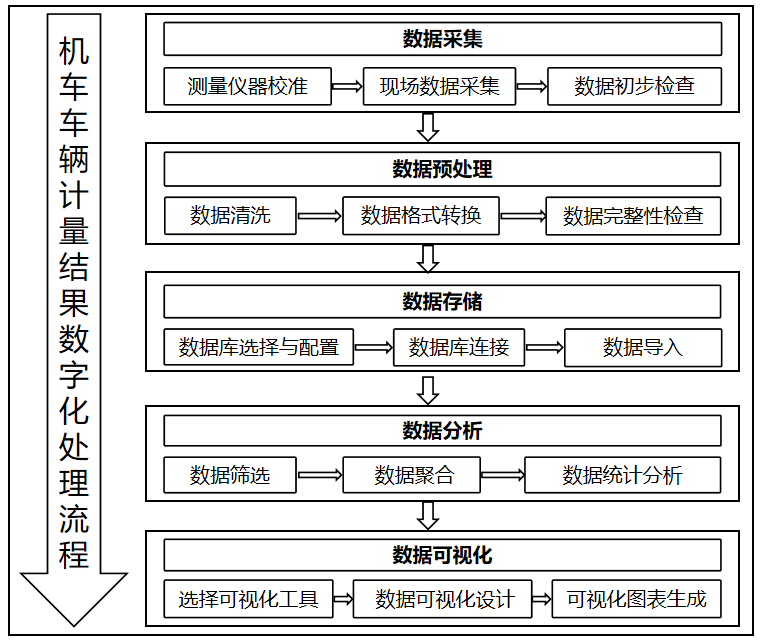
机车车辆计量结果数据可视化应根据数据特点和分析需求，合理选择可视化工具。

应选择直观的、易于理解的方式，将数据蕴含的信息进行可视化展示，宜使用的数据图表包括但不限于折线图、柱状图、饼状图等，图表应包含必要的说明注释。

应能生成可视化图表文件并保存。

2. （资料性）  
   机车车辆计量结果数字化处理流程

图A.1给出了机车车辆计量结果数字化处理流程。



* 1. 机车车辆计量结果数字化处理流程

