

ICS 07.060

CCS A 45

HY

中华人民共和国海洋行业标准

HY/T XXX—202X

海洋经济统计数据集质量控制 技术规程

Code of practice for data set quality control of
ocean economy statistics

(报批稿)

202X-XX- XX发布

202X-XX-XX实施

中华人民共和国自然资源部 发布

目 次

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 一般规定.....	1
4.1 海洋经济统计数据集质量控制要求.....	1
4.2 海洋经济统计数据集质量控制流程.....	2
5 数据集质量控制内容与选择.....	2
5.1 质量控制对象.....	2
5.2 质量控制内容.....	3
5.3 质量控制内容的选择.....	3
6 齐全性检验.....	4
6.1 完整性检验.....	4
6.2 唯一性检验.....	4
7 规范性检验.....	4
7.1 一般规范性检验.....	4
7.2 指标规范性检验.....	5
8 逻辑关系检验.....	5
8.1 比较逻辑关系.....	5
8.2 相关逻辑关系.....	6
9 异常值检验.....	6
9.1 值域参考范围检验.....	6
9.2 统计异常值检验.....	7
10 数据集质量评价.....	8
10.1 数据集质量评价要求.....	8
10.2 数据集综合质量评价方法.....	10
11 数据集质量控制报告书编制.....	10
附 录 A（规范性） 狄克逊统计量.....	12
附 录 B（规范性） 狄克逊检验的临界值表.....	13
附 录 C（规范性） 数据集质量控制报告书范例.....	19
参 考 文 献.....	20

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。

本文件的某些内容仍可能涉及专利，本文件的发布机构不承担识别专利的责任。

本文件由中华人民共和国自然资源部提出。

本文件由全国海洋标准化技术委员会（SAC/TC 283）归口。

本文件起草单位：国家海洋信息中心。

本文件主要起草人：宋维玲、彭星、郑莉、付瑞全、周洪军、杨洋、张献丽、郭越。

海洋经济统计数据集质量控制技术规程

1 范围

本文件规定了海洋经济统计数据集质量控制基本要求、流程、内容和方法、质量评价方法以及质量控制报告书编制等内容。

本文件适用于常规海洋经济统计数据集的收集和汇总过程的数据质量控制。

2 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 4883-2008 数据的统计处理和解释 正态样本离群值的判断和处理

GB/T 20794-2021 海洋及相关产业分类

HY/T 094 沿海行政区域分类与代码

3 术语和定义

GB/T 20794-2021规定的以及下列术语和定义适用于本文件。

3.1

海洋经济 ocean economy

开发、利用和保护海洋的各类产业活动，以及与之相关联活动的总和。

[来源：GB/T 20794-2021，3.1]

3.2

海洋经济统计数据集 ocean economy statistical datasets

按照不同统计方式、渠道获取的各类海洋经济统计指标数据的集合。

4 一般规定

4.1 海洋经济统计数据集质量控制要求

4.1.1 全面性

海洋经济统计数据应报送及时，基本信息应全面完整，不存在漏报或重报的统计数据。

4.1.2 规范性

海洋经济统计数据应按规定的统计日期、名称、计量单位、格式等进行报送。

4.1.3 一致性

海洋经济统计数据与关联数据之间的逻辑关系应合理。

4.1.4 准确性

海洋经济统计数据应真实且误差控制在允许范围内，能反映客观实际情况。

4.1.5 可比性

海洋经济统计数据应连续、可比，在时间上和空间上的统计口径、计算方法规范统一。

4.2 海洋经济统计数据集质量控制流程

海洋经济统计数据集质量控制流程包括数据收集与处理、数据集质量控制、数据集质量评价、质量控制报告书编制等过程。应及时将质量控制报告反馈至数据报送单位，与数据报送单位沟通数据集质量检验情况，若未通过检验，则数据报送单位应及时进行核实补充。具体流程见图 1。

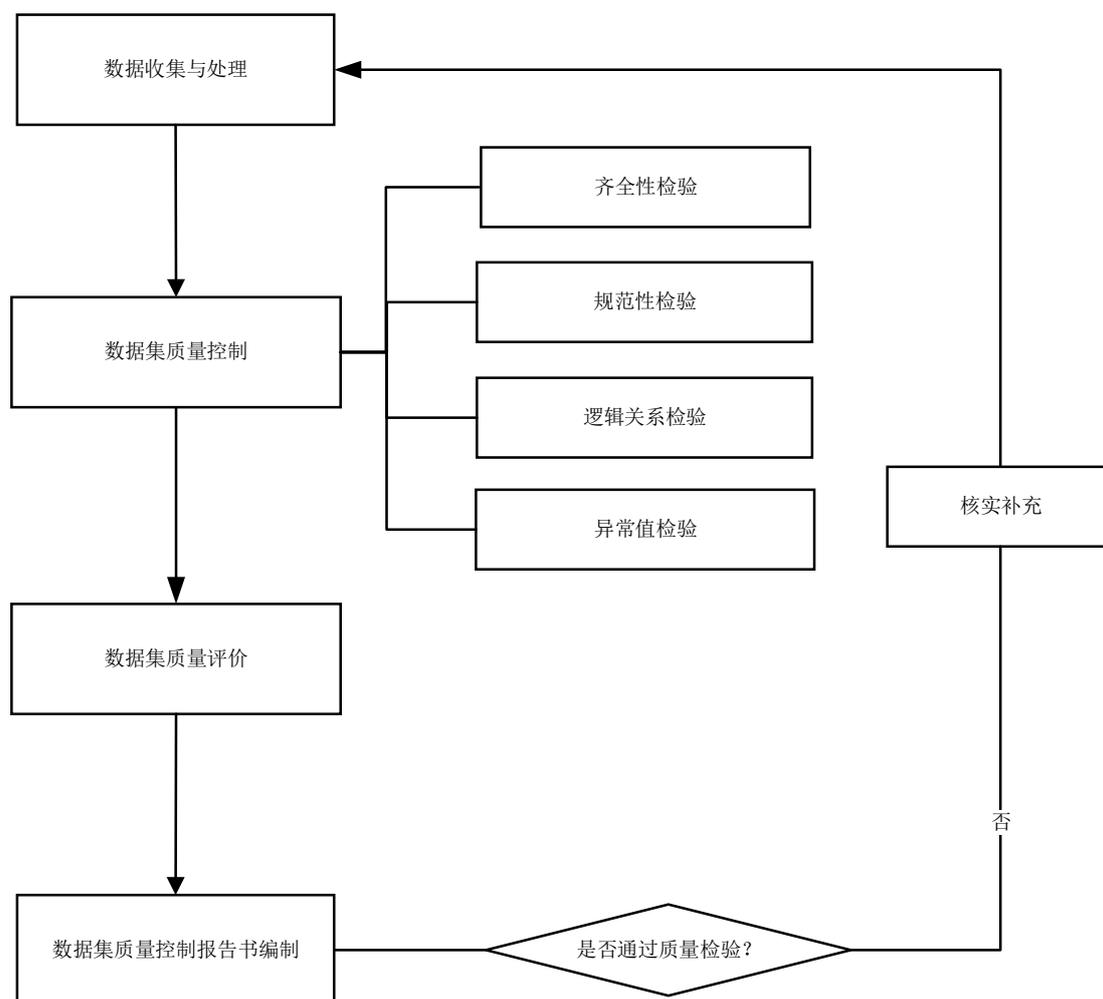


图1 海洋经济统计数据集质量控制工作流程图

5 数据集质量控制内容与选择

5.1 质量控制对象

收集海洋经济统计数据并进行归类汇总处理，形成各类海洋经济统计数据集。按照数据获取方式，海洋经济统计数据集主要包括海洋经济报表制度数据集、海洋经济调查数据集、

涉海企业直报数据集、海洋经济共享数据集等。

注 1：海洋经济报表制度数据集是指通过海洋经济统计调查制度、海洋生产总值核算制度等海洋经济报表制度获取的数据集。

注 2：海洋经济调查数据集是指通过重点调查、典型调查、抽样调查等调查方式获取的数据集。

注 3：涉海企业直报数据集是指通过涉海直报企业报送获取的数据集。

注 4：海洋经济共享数据集是指通过与涉海单位共享的方式获取的数据集。

5.2 质量控制内容

5.2.1 齐全性检验

按照海洋经济统计任务要求，检验数据集在规定时间内存在漏报、重报的情况。

5.2.2 规范性检验

按照报送数据规定的要求，检验海洋经济统计数据集基本信息的规范程度。

5.2.3 逻辑关系检验

检验海洋经济统计数据集的数据之间符合既定的比较逻辑或相关逻辑关系的情况。

5.2.4 异常值检验

通过统计学方法，检验海洋经济统计数据集存在异常值的情况，包括检验数据超出相应的参考值域范围或与正常观测数据有显著性差别等。

5.3 质量控制内容的选择

针对不同类型的海洋经济统计数据集，选择适宜的质量控制方法对数据的齐全性、规范性、逻辑关系、异常值等方面进行质量检验，见表 1。

表1 海洋经济统计数据集质量控制内容的选择

数据集	质量控制内容
海洋经济报表制度数据集	齐全性检验
	规范性检验
	逻辑关系检验
	异常值检验
海洋经济调查数据集	齐全性检验
	规范性检验
	逻辑关系检验
涉海企业直报数据集	齐全性检验
	规范性检验
	逻辑关系检验
	异常值检验
海洋经济共享数据集	齐全性检验
	规范性检验
	逻辑关系检验-比较逻辑关系检验

6 齐全性检验

6.1 完整性检验

6.1.1 数据漏报检验

检验海洋经济统计数据在规定时间内报送的情况。

6.1.2 数据基本信息漏报检验

检验收集的海洋经济统计数据存在基本信息空缺的情况。数据基本信息包括数据的地区名称、填报单位名称、数据日期、指标名称、计量单位、单位负责人、统计负责人、填表人、报出日期等。

6.2 唯一性检验

通过检验同一来源、同一时间区间、同一统计范围、同一频次、同一计量单位的统计数据的唯一性，确认数据是否存在重复报送的情况。

7 规范性检验

7.1 一般规范性检验

7.1.1 地区名称检验

检验海洋经济统计数据集地区名称的规范性。统计区域名称按照 HY/T 094，表 1-表 3。

示例：

不正确	正确
地区名称：广西省 地区名称：广西区 地区名称：广西 ...	地区名称：广西壮族自治区

7.1.2 填报单位名称检验

检验海洋经济统计数据集填报单位名称的规范性。填报单位名称应与其法人单位名称保持一致。

示例：

不正确	正确
填报单位：天津规划和自然资源局 填报单位：天津市规划局 填报单位：天津市自然资源局 ...	填报单位：天津市规划和自然资源局

7.1.3 数据日期检验

检验海洋经济统计数据集数据日期的规范性。按照统计数据反映的实际日期准确、完整填报，不应与报出日期混淆。

示例：

对于 2020 年报送的 2019 年数据：

不正确
数据日期：2020 年
...

正确
数据日期：2019 年

7.1.4 报出日期检验

检验海洋经济统计数据集报出日期的规范性。日期的表述形式为“年-月-日”，年份填满四位，月份填 01-12，日期填 01-31，月份、日期小于 10 的前位需补零。

示例：

不正确
报出日期：2020 年 2 月 9 日
...

正确
报出日期：2020 年 02 月 09 日

7.2 指标规范性检验

7.2.1 指标名称检验

检验海洋经济统计数据集指标名称的规范性。统计数据指标名称应写全称，避免出现缩写或英语简写等不规范的形式。

示例：

不正确
指标名称：海洋药物和生物制品增加值
指标名称：海洋生物医药业增加值
...

正确
指标名称：海洋药物和生物制品业增加值

7.2.2 指标计量单位检验

检验海洋经济统计数据集指标计量单位的规范性。统计数据计量单位应符合报送统一要求，避免因计量单位不一致而影响数据准确性。

示例：

按规定应填报千元：

不正确
计量单位：美元
计量单位：万元
...

正确
计量单位：千元

7.2.3 指标数据格式检验

检验海洋经济统计数据集指标数据格式的规范性。统计数据保留小数位数等格式应统一规范、符合要求。

示例：

按规定增速数据应保留一位小数：

不正确
数据格式：3%
...

正确
数据格式：3.3%

8 逻辑关系检验

8.1 比较逻辑关系

8.1.1 总量指标数据等于构成分项指标数据合计，或者总量指标数据大于、等于构成分项部分指标数据合计。

示例：计算海洋生产总值是否等于海洋第一产业增加值、海洋第二产业增加值与海洋第三产业增加值之和，检验数据间比较逻辑关系。

8.1.2 指标当期累计值等于当期值与上期累计值之和，或者指标当期累计值大于、等于上期累计值。

示例：计算涉海企业营业收入的当月累计值是否等于当月值与上月累计值之和，检验数据间比较逻辑关系。

8.1.3 低频指标数据等于对应高频指标数据合计。

示例：计算季度沿海港口货物吞吐量是否等于对应各月沿海港口货物吞吐量之和，检验数据间比较逻辑关系。

8.2 相关逻辑关系

8.2.1 指标数据之间存在较稳定的比例或比率关系。

示例：计算海洋盐业产值与产量的比例是否在合理市场价格 150 元/吨~300 元/吨范围内，检验数据间相关逻辑关系。

8.2.2 指标数据的变动趋势之间存在同向或反向一致性。

示例：计算验证海洋渔业产值增长率与海洋水产品产量增长率在方向上是否一致（在价格变动不大的情况下），检验数据间相关逻辑关系。

8.2.3 同一指标不同来源的统计数据之间符合逻辑一致性。

示例：计算验证国家旅游部门、地方旅游部门提供的数据合计在数量级、变动趋势上是否一致，检验数据间相关逻辑关系。

9 异常值检验

9.1 值域参考范围检验

9.1.1 指标极值检验

9.1.1.1 确定指标数据值域范围

海洋经济统计数据的比重、比例指标等宜采用指标极值检验。计算一定历史时期内比重、比例等指标数据的最小值和最大值，确定其值域范围。

9.1.1.2 检验数据是否超出值域范围

检验数据是否低于上述确定的最小值或高于最大值。

示例 1：某地区近 9 年近岸海域水质优良（一、二类）比例依次为 63%、63%、69%、66%、67%、71%、73%、68%、75%，此数据序列的最小值为 63%，最大值为 75%，通过判断报送的近岸海域水质优良（一、二类）比例数据是否大于 75%或小于 63%，检验数据质量。

示例 2：某涉海企业往年营业收入半年度数据占年度数据的比重最小值为 40%，最大值为 60%，通过判断涉海企业报送的半年度数据占全年度数据的比重是否大于 40%或小于 60%，检验数据质量。

9.1.2 增长率极值检验

9.1.2.1 确定指标数据增长率值域范围

总量指标宜采用增长率极值检验。计算一定历史时期内海洋经济统计数据增长率的极大值和最小值，确定指标数据增长率的值域范围。

9.1.2.2 检验数据增长率是否超出值域范围

检验数据增长率是否低于上述确定的增长率最小值或高于增长率最大值。

示例：某省6年内海洋渔业增加值（单位：亿元）依次为441、425、460、380、326、336，计算得到5年内海洋渔业增加值增长率分别为-3.6%、8.2%、-17.4%、-14.2%、3.1%，增长率数据序列的最小值为-17.4%，最大值为8.2%。结合专家经验，综合确定海洋渔业增加值增速的值域范围为-15%至15%。通过判断报送的海洋渔业增加值的增长率是否大于15%或小于-15%，检验数据质量。

9.2 统计异常值检验

9.2.1 拉依达（3σ）检验

9.2.1.1 计算样本标准差和离差

对于总体服从或近似服从正态分布，且样本量 $n > 10$ 的海洋经济统计指标数据序列，宜采用拉依达检验。计算数据序列的均值 \bar{x} 、标准差 s 和离差 v_i ：

$$\bar{x} = \sum_{i=1}^n x_i \dots\dots\dots (1)$$

式中：

x_i ——第 i 个数据数值；

\bar{x} ——数据序列的算数平均值。

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \dots\dots\dots (2)$$

式中：

s ——数据序列的标准差。

$$v_i = x_i - \bar{x} \dots\dots\dots (3)$$

式中：

v_i ——第 i 个数据的离差。

9.2.1.2 确定异常值判定依据

选定3倍数据序列的标准差作为异常值判定依据。

9.2.1.3 判定异常值

若 $|v_i| > 3s$ ，判定 x_i 为异常值。若发现异常值，则将其剔除数据序列继续进行拉依达检验，如此循环，直至未发现异常值为止。

示例：现以涉海企业直报中营业收入数据为例。营业收入近似服从对数正态分布，使用拉依达检验法进行数据质量检验。某涉海企业某两年的月营业收入（单位：千元）分别为615、47、2337、2003、2702、2385、1980、1990、2166、3951、3299、4201、2835、2030、1019、1057、714、1928、2344、713、968、683、1885、1918。将其分别取对数：6.42、3.85、7.76、7.60、7.90、7.78、7.59、7.60、7.68、8.28、8.10、8.34、7.95、7.62、6.93、6.96、6.57、7.56、7.76、6.57、6.88、6.53、7.54、7.56。

计算涉海企业营业收入对数序列的均值和标准差，分别 $\bar{x} = 7.31$ ， $s = 0.92$ ；依次检验涉海企业营业收入对数序列每个样本的离差 v_i 是否大于3倍的标准差。结果发现检验数据3.85对应的离差 $|v_2| = |3.85 - 7.31| \Rightarrow 3.46 \times 3 = 10.38$ ，判定3.85为异常值，反推原始数据 $3.85^e = 47$ 为异常值，即营业收入47为异常值。对剔除异常值47后的营业收入数据序列继续进行拉依达检验，未发现异常值。因此，通过拉依达检验法判定某涉海企业营业收入47为异常值。

9.2.2 狄克逊 (Dixon) 检验

9.2.2.1 样本观测值排序

对于样本量 $3 \leq n \leq 100$ 的海洋经济统计指标数据序列, 宜采用狄克逊检验。将数据由小到大排列为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)}$ 。

9.2.2.2 计算狄克逊统计量

计算数据序列高端值和低端值狄克逊检验统计量, 分别为 D_n 、 D'_n (见附录 A)。

9.2.2.3 确定临界值

选定显著性水平 α , 对于单侧检验, 查临界值表得到临界值 $D_{1-\alpha}(n)$, 见附录 B.1, 临界值表按照 GB/T 4883-2008 中表 A.3、表 C.1; 对于双侧检验, 查临界值表得到临界值 $\hat{D}_{1-\alpha}(n)$, 见附录 B.2, 临界值表按照 GB/T 4883-2008 中表 A.3'、表 C.2。

根据指标属性选择单侧狄克逊检验或双侧狄克逊检验, 若该指标只有上限或下限, 宜采用单侧检验; 若该指标的取值范围在一个区间内, 宜采用双侧狄克逊检验。

9.2.2.4 判定异常值

对于单侧检验, 检验高端值, 若 $D_n > D_{1-\alpha}(n)$, 判定 $x_{(n)}$ 为异常值; 检验低端值, $D'_n > D_{1-\alpha}(n)$, 判定 $x_{(1)}$ 为异常值; 否则无异常值。

对于双侧检验, 若 $D_n > D_n'$, 且 $D_n > \hat{D}_{1-\alpha}(n)$, 判定 $x_{(n)}$ 为异常值; 若 $D'_n > D_n$, 且 $D'_n > \hat{D}_{1-\alpha}(n)$, 判定 $x_{(1)}$ 为异常值; 否则无异常值。

若发现异常值, 则将其剔除数据序列继续进行狄克逊检验, 如此循环, 直至未发现异常值为止。

示例 1: 单侧狄克逊检验。以涉海企业直报中营业收入数据为例。使用单侧狄克逊检验进行数据质量检验。某涉海企业某 11 个月的月营业收入 (单位: 千元) 分别为 241、194、127、122、110、77、221、287、131、352、907。将其从小到大排序: 77、110、122、127、131、194、221、241、287、352、907。

检验高端值数据 907, $D_{11} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(2)}) = (x_{(11)} - x_{(9)}) / (x_{(11)} - x_{(2)}) = (907 - 287) / (907 - 110) = 0.78$; 检验低端值数据 77, $D'_{11} = (x_{(3)} - x_{(1)}) / (x_{(n-1)} - x_{(1)}) = (x_{(3)} - x_{(1)}) / (x_{(10)} - x_{(1)}) = (122 - 77) / (352 - 77) = 0.16$ 。选定显著性水平 $\alpha = 0.05$, 样本量 $n = 11$, 查单侧临界值表 $D_{1-\alpha}(n) = D_{0.95}(11) = 0.575$; $D_{11} > D_{0.95}(11)$, 判定 907 为异常值; $D'_{11} < D_{0.95}(11)$, 判定 77 为非异常值。

剔除异常值 907, 对剩余含有 10 个样本的营业收入数据序列继续进行单侧狄克逊检验。检验高端值数据 352, $D_{10} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(2)}) = (x_{(10)} - x_{(9)}) / (x_{(10)} - x_{(2)}) = (352 - 287) / (352 - 110) = 0.27$; 检验低端值数据 77, $D'_{10} = (x_{(2)} - x_{(1)}) / (x_{(n-1)} - x_{(1)}) = (x_{(2)} - x_{(1)}) / (x_{(9)} - x_{(1)}) = (110 - 77) / (287 - 77) = 0.16$; 查单侧临界值表 $D_{1-\alpha}(n) = D_{0.95}(10) = 0.477$; $D_{10} < D_{0.95}(10)$, 判定 352 为非异常值; $D'_{10} < D_{0.95}(10)$, 判定 77 为非异常值。因此, 判定营业收入 907 为异常值。

示例 2: 双侧狄克逊检验。仍以示例 1 中涉海企业直报中营业收入数据为例。使用双侧狄克逊检验进行数据质量检验。

选定显著性水平 $\alpha = 0.05$, 样本量 $n = 11$, 查临界值表 $\hat{D}_{1-\alpha}(n) = \hat{D}_{0.95}(11) = 0.619$; 由于 $D_{11} > D'_{11}$, 且 $D_{11} > \hat{D}_{0.95}(11)$, 判定 907 为异常值。

剔除异常值 907, 对剩余含有 10 个样本的营业收入数据序列继续进行双侧狄克逊检验。查临界值表 $\hat{D}_{1-\alpha}(n) = \hat{D}_{0.95}(10) = 0.530$; $D_{10} > D'_{10}$, $D_{10} < \hat{D}_{0.95}(10)$, 判定无异常值。因此, 判定营业收入 907 为异常值。

10 数据集质量评价

10.1 数据集质量评价要求

10.1.1 核心指标与非核心指标的确定

对海洋经济统计数据集中的指标，基于业务需要和使用频率来选取核心指标，宜选取数据源稳定、使用频率高的指标作为核心指标，数量不宜过多。

10.1.2 数据集质量判定规则

依据海洋经济统计数据集中每类数据集的各项质量检验结果，制定下列数据集质量判定规则（见表2）。

- a) 数据集齐全性检验的数据质量判定规则为：
 - 1) 所有指标数据均未遗漏或重复，则为“无明显缺陷”；
 - 2) 非核心指标数据遗漏或重复，则为“轻度缺陷”；
 - 3) 核心指标数据遗漏或重复，则为“重度缺陷”；
 - 4) 指标数据均未报送，则为“极重缺陷”。
- b) 数据集规范性检验的数据质量判定规则为：
 - 1) 所有指标数据均规范，则为“无明显缺陷”；
 - 2) 非核心指标数据不规范，则为“轻度缺陷”；
 - 3) 核心指标数据不规范，则为“重度缺陷”；
 - 4) 所有指标数据均不规范，则为“极重缺陷”。
- c) 数据集逻辑关系检验的数据质量判定规则为：
 - 1) 所有指标均满足逻辑关系，或遇特殊情况，数据报送单位可对不符合相关逻辑关系的指标数据进行解释说明，若解释符合实际情况，则为“无明显缺陷”；
 - 2) 非核心指标不满足逻辑关系，则为“轻度缺陷”；
 - 3) 核心指标不满足逻辑关系，则为“重度缺陷”；
 - 4) 所有指标数据均不满足逻辑关系，则为“极重缺陷”；
- d) 数据集异常值检验的数据质量判定规则为：
 - 1) 无法检验或检验无异常值，或遇特殊情况，数据报送单位可对异常数据进行解释说明，若解释符合实际情况，则为“无明显缺陷”；
 - 2) 非核心指标存在异常值，则为“轻度缺陷”；
 - 3) 核心指标存在异常值，则为“重度缺陷”；
 - 4) 所有指标数据均存在异常值，则为“极重缺陷”。

10.1.3 数据集质量标识规则

数据集各项质量检验结果的质量标识规则为：

- a) “无明显缺陷”赋值为“1”，表示一级质量；
- b) “轻度缺陷”赋值为“2”，表示二级质量；
- c) “重度缺陷”赋值为“3”，表示三级质量；
- d) “极重缺陷”赋值为“4”，表示四级质量。

表2 数据集各项质量检验结果的质量判定与质量分级

数据集质量控制内容	无明显缺陷	轻度缺陷	重度缺陷	极重缺陷
质量标识符赋值	1	2	3	4
齐全性检验 q_1	所有指标数据均未遗漏或重复	非核心指标数据遗漏或重复	核心指标数据遗漏或重复	所有指标数据均未报送

表 2 数据集各项质量检验结果的质量判定与质量分级 (续)

规范性检验 q_2	所有指标数据均规范	非核心指标数据不规范	核心指标数据不规范	所有指标数据均不规范
逻辑关系检验 q_3	所有指标均满足逻辑关系或有合理的解释说明	非核心指标不满足逻辑关系	核心指标不满足逻辑关系	所有指标数据均不满足逻辑关系
异常值检验 q_4	无法检验或检验无异常值或有合理的解释说明	非核心指标存在异常值	核心指标存在异常值	所有指标数据均存在异常值

10.2 数据集综合质量评价方法

10.2.1 数据集质量标识

根据 10.1.2 中数据集质量判定规则,对海洋经济统计数据集中每类数据集各项质量检验结果进行质量判定,再根据 10.1.3 中数据集质量标识规则,对数据集各项质量检验结果进行质量标识符赋值。

10.2.2 数据集综合质量评价

10.2.2.1 计算数据集的综合质量标识符数值

取数据集齐全性、规范性、逻辑关系、异常值检验的质量标识符数值平均值,作为数据集的综合质量标识符数值。

对于数据集 A ,其综合质量标识符数值 Q 的计算公式如下:

$$Q = \frac{q_1 + q_2 + q_3 + q_4}{4} \dots\dots\dots (4)$$

式中:

- q_1 ——数据集 A 的齐全性检验质量标识符数值;
- q_2 ——数据集 A 的规范性检验质量标识符数值;
- q_3 ——数据集 A 的逻辑关系检验质量标识符数值;
- q_4 ——数据集 A 的异常值检验质量标识符数值;
- Q ——数据集 A 的综合质量标识符数值。

10.2.2.2 数据集综合质量评价规则

根据上述 Q 值的取值范围对数据集的综合质量进行评价,评价规则为:

- a) 若 $1 \leq Q < 1.75$,则数据集质量级别为“优”;
- b) 若 $1.75 \leq Q < 2.5$,则数据集质量级别为“良”;
- c) 若 $2.5 \leq Q < 3.25$,则数据集质量级别为“中”;
- d) 若 $3.25 \leq Q \leq 4$,则数据集质量级别为“差”。

11 数据集质量控制报告书编制

根据海洋经济统计数据集质量控制及质量评价过程与结果,编制数据集质量控制报告书(见附录 C),并将报告反馈给数据报送单位。

数据集质量控制报告由正文和数据集质量评价附表构成。正文是质量控制过程、方法和结果的全面记录和描述,包括数据集质量控制与评价的组织、数据集概况、质量控制与评价依据、质量检验内容与方法、质量检验与评价步骤、质量评价结果、质量控制结论与建议等。

其中，建议宜针对质量检验问题进行表述。数据集质量评价附表是数据集综合质量评价结果的描述和反映。

附 录 A
(规范性)
狄克逊统计量

表A.1 狄克逊统计量

样本量	高端值检验统计量	低端值检验统计量
$n: 3-7$	$D_n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$	$D_n' = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}$
$n: 8-10$	$D_n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(2)}}$	$D_n' = \frac{x_{(2)} - x_{(1)}}{x_{(n-1)} - x_{(1)}}$
$n: 11-13$	$D_n = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(2)}}$	$D_n' = \frac{x_{(3)} - x_{(1)}}{x_{(n-1)} - x_{(1)}}$
$n: 14-100$	$D_n = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(3)}}$	$D_n' = \frac{x_{(3)} - x_{(1)}}{x_{(n-2)} - x_{(1)}}$

附 录 B
(规范性)
狄克逊检验的临界值表

表B.1~表B.2分别给出了单侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表和双侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表。

表B.1 单侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表

n	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.005$
3	0.885	0.941	0.988	0.994
4	0.679	0.765	0.889	0.92
5	0.557	0.642	0.782	0.823
6	0.484	0.562	0.698	0.744
7	0.434	0.507	0.637	0.68
8	0.479	0.554	0.681	0.723
9	0.441	0.512	0.635	0.676
10	0.41	0.477	0.597	0.638
11	0.517	0.575	0.674	0.707
12	0.49	0.546	0.642	0.675
13	0.467	0.521	0.617	0.649
14	0.491	0.546	0.640	0.672
15	0.470	0.524	0.618	0.649
16	0.453	0.505	0.597	0.629
17	0.437	0.489	0.580	0.611
18	0.424	0.475	0.564	0.595
19	0.412	0.462	0.550	0.580
20	0.401	0.450	0.538	0.568
21	0.391	0.440	0.526	0.556
22	0.382	0.431	0.516	0.545
23	0.374	0.422	0.507	0.536
24	0.367	0.413	0.497	0.526
25	0.360	0.406	0.489	0.519
26	0.353	0.399	0.482	0.510
27	0.347	0.393	0.474	0.503
28	0.341	0.387	0.468	0.496
29	0.337	0.381	0.462	0.489
30	0.332	0.376	0.456	0.484
31	0.327	0.371	0.450	0.478
32	0.323	0.367	0.445	0.473
33	0.319	0.362	0.441	0.468
34	0.315	0.358	0.436	0.463
35	0.311	0.354	0.432	0.458
36	0.308	0.350	0.427	0.454
37	0.305	0.347	0.423	0.450

表B.1 单侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表 (续)

n	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.005$
38	0.301	0.343	0.419	0.446
39	0.298	0.340	0.416	0.442
40	0.296	0.337	0.413	0.439
41	0.293	0.334	0.409	0.435
42	0.290	0.331	0.406	0.432
43	0.288	0.328	0.403	0.429
44	0.285	0.326	0.400	0.425
45	0.283	0.323	0.397	0.423
46	0.281	0.321	0.394	0.420
47	0.279	0.318	0.391	0.417
48	0.277	0.316	0.389	0.414
49	0.275	0.314	0.386	0.412
50	0.273	0.312	0.384	0.409
51	0.271	0.310	0.382	0.407
52	0.269	0.308	0.379	0.405
53	0.267	0.306	0.377	0.402
54	0.265	0.304	0.375	0.400
55	0.264	0.302	0.373	0.398
56	0.262	0.300	0.371	0.396
57	0.261	0.298	0.369	0.394
58	0.259	0.297	0.367	0.392
59	0.258	0.295	0.366	0.391
60	0.256	0.294	0.363	0.388
61	0.255	0.292	0.362	0.387
62	0.253	0.291	0.361	0.385
63	0.252	0.289	0.359	0.383
64	0.251	0.288	0.357	0.382
65	0.250	0.287	0.355	0.380
66	0.249	0.285	0.354	0.379
67	0.247	0.284	0.353	0.377
68	0.246	0.283	0.351	0.376
69	0.245	0.282	0.350	0.374
70	0.244	0.280	0.348	0.372
71	0.243	0.279	0.347	0.371
72	0.242	0.278	0.346	0.370
73	0.241	0.277	0.344	0.368
74	0.240	0.276	0.343	0.368
75	0.239	0.275	0.342	0.366
76	0.238	0.274	0.341	0.365
77	0.237	0.273	0.340	0.364
78	0.236	0.272	0.338	0.363
79	0.235	0.271	0.337	0.361

表B.1 单侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表 (续)

n	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.005$
80	0.234	0.270	0.336	0.360
81	0.233	0.269	0.335	0.359
82	0.232	0.268	0.334	0.358
83	0.232	0.267	0.333	0.356
84	0.231	0.266	0.332	0.356
85	0.230	0.265	0.331	0.355
86	0.229	0.264	0.330	0.353
87	0.228	0.263	0.329	0.352
88	0.228	0.262	0.328	0.352
89	0.227	0.262	0.327	0.351
90	0.226	0.261	0.326	0.350
91	0.225	0.260	0.325	0.349
92	0.225	0.259	0.324	0.348
93	0.224	0.259	0.323	0.347
94	0.223	0.258	0.323	0.346
95	0.223	0.257	0.322	0.345
96	0.222	0.256	0.321	0.344
97	0.221	0.255	0.320	0.344
98	0.221	0.255	0.320	0.343
99	0.220	0.254	0.319	0.341
100	0.219	0.254	0.318	0.341

表B.2 双侧狄克逊检验统计量的临界值 $D_{1-\alpha}^*(n)$ 表

n	$\alpha=0.05$	$\alpha=0.01$
3	0.970	0.994
4	0.829	0.926
5	0.710	0.821
6	0.628	0.740
7	0.569	0.680
8	0.608	0.717
9	0.564	0.672
10	0.530	0.635
11	0.619	0.709
12	0.583	0.660
13	0.557	0.638
14	0.587	0.669
15	0.565	0.646
16	0.547	0.629
17	0.527	0.614
18	0.513	0.602
19	0.500	0.582
20	0.488	0.570
21	0.479	0.560
22	0.469	0.548
23	0.460	0.537
24	0.449	0.522
25	0.441	0.518
26	0.436	0.509
27	0.427	0.504
28	0.420	0.497
29	0.415	0.489
30	0.409	0.480
31	0.403	0.473
32	0.399	0.468
33	0.395	0.463
34	0.390	0.460
35	0.388	0.458
36	0.438	0.442
37	0.380	0.450
38	0.377	0.447
39	0.375	0.442
40	0.370	0.438
41	0.367	0.433
42	0.364	0.432

表 B.2 双侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表 (续)

n	$\alpha=0.05$	$\alpha=0.01$
43	0.362	0.428
44	0.359	0.425
45	0.357	0.422
46	0.353	0.419
47	0.352	0.416
48	0.350	0.413
49	0.346	0.412
50	0.343	0.409
51	0.342	0.407
52	0.340	0.405
53	0.338	0.402
54	0.337	0.400
55	0.335	0.399
56	0.334	0.399
57	0.330	0.396
58	0.329	0.393
59	0.327	0.390
60	0.325	0.389
61	0.232	0.387
62	0.321	0.385
63	0.320	0.383
64	0.319	0.382
65	0.318	0.379
66	0.316	0.377
67	0.315	0.375
68	0.313	0.376
69	0.313	0.375
70	0.312	0.375
71	0.310	0.373
72	0.309	0.373
73	0.308	0.371
74	0.306	0.370
75	0.305	0.368
76	0.304	0.363
77	0.304	0.363
78	0.303	0.362
79	0.303	0.361
80	0.302	0.358
81	0.301	0.358
82	0.301	0.355

表 B.2 双侧狄克逊检验统计量的临界值 $D_{1-\alpha}(n)$ 表 (续)

n	$\alpha=0.05$	$\alpha=0.01$
83	0.301	0.355
84	0.298	0.353
85	0.297	0.351
86	0.297	0.351
87	0.296	0.349
88	0.295	0.349
89	0.294	0.347
90	0.293	0.347
91	0.291	0.344
92	0.290	0.344
93	0.289	0.343
94	0.289	0.343
95	0.288	0.343
96	0.288	0.342
97	0.286	0.340
98	0.285	0.340
99	0.285	0.339
100	0.284	0.339

附录 C
(规范性)
数据集质量控制报告书范例

图C.1给出了数据集质量控制报告书范例。

海洋经济统计数据集质量控制报告

一、基本概况

1. 数据集质量控制与评价的组织
相关机构与人员、时间、地点及形式等。
2. 数据集概况
任务来源、数据集范围与口径、数据量、数据格式等。
3. 质量控制与评价依据
列出依据文本，如报表制度、合同书、任务书等。

二、数据集质量控制与评价

1. 质量检验内容与方法
2. 质量检验与评价步骤
3. 质量评价结果

三、质量控制结论与建议

1. 结论
综合描述数据集质量检验与质量评级最终结论及存在的问题。
2. 建议
提出提高数据集质量的工作建议。

四、其他

附表：

数据集质量评价表

数据集	齐全性检验 质量标识符	规范性检验 质量标识符	逻辑关系检验 质量标识符	异常值检验 质量标识符	综合 质量标识符
A_1	q_{11}	q_{12}	q_{13}	q_{14}	Q_1
...
A_n	q_{n1}	q_{n2}	q_{n3}	q_{n4}	Q_n
A	q_1	q_2	q_3	q_4	Q
综合质量评价结果：优/良/中/差					

注：1) 数据集 A_1, A_2, \dots, A_n 是数据集 A 的子集；

2) $Q_1 = \frac{q_{11} + q_{12} + q_{13} + q_{14}}{4}$; $Q_n = \frac{q_{n1} + q_{n2} + q_{n3} + q_{n4}}{4}$; $Q = \frac{q_1 + q_2 + q_3 + q_4}{4}$ 。

图 C.1 数据集质量控制报告书范例图

参 考 文 献

- [1] 何广顺, 丁黎黎, 宋维玲等著. 海洋经济分析评估理论、方法与实践[M]. 北京: 海洋出版社, 2014.
 - [2] 郭越, 宋维玲, 董伟. 构建海洋统计数据质量监控体系的思考[J]. 海洋开发与管理, 2010, 27 (11): 4-7.
 - [3] 杨鑫, 刘文长等著. 质量控制过程中的统计技术[M]. 北京: 化学工业出版社, 2014.
 - [4] 成邦文, 董丽娅, 杨峻. 研究与开发机构统计数据质量与异常点的对数正态分布检验与识别[J]. 统计研究, 2000, (1): 42-45.
 - [5] 国家统计局, 国家统计质量保证框架, 2021年6月.
 - [6] 国土资源部信息中心, 国土资源数据库数据质量检查验收规范, 2008年12月.
-