

团 体 标 准

T/CABC XXX—XXXX

商品源数据 数据质量核查实施规范

Product source data - specification for implementing data quality inspection

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

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 核查方案	2
4.1 核查流程	2
4.2 抽样方法	2
4.3 抽样数量	2
5 核查方法	2
5.1 技术审核	2
5.2 人工审核	3
5.3 其他方法	3
6 核查结果	3
6.1 结果分类	3
6.2 结果报告	5
附 录 A 抽样方案（规范性）	6
附 录 B 核查结果明细表（资料性）	7

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。

本文件由中国物品编码中心提出。

本文件由中国条码技术与应用协会归口。

本文件起草单位：XXXX。

本文件主要起草人：XXXX。

商品源数据 数据质量核查实施规范

1 范围

本文件规定了数据质量核查的术语和定义、核查方案、核查方法、核查结果等要求。
本文件适用于商品源数据的数据质量核查工作。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是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 2828.1-2012 计数抽样检验程序 第1部分：按接收质量限（AQL）检索的逐批检验抽样计划

GB/T 10111-2008 随机数的产生及其在产品质量抽样检验中的应用程序

T/CABC 2.1-2022 商品源数据 商品属性信息规范 第1部分：通用属性

T/CABC XXX-XXXX 商品源数据 数据质量控制实施规范

3 术语和定义

T/CABC XXX-XXXX界定的以及下列术语和定义适用于本文件。

3.1

数据质量核查 data quality inspection

对数据质量的现状与预期进行比较确认。

3.2

简单随机抽样 simple random sampling

从总体中抽取n个抽样单元构成样本，使n个抽样单元所有的可能组合都有相等被抽到概率的抽样。

[GB/T 10111-2008, 3.1.13]

3.3

分层简单随机抽样 stratified simple random sampling

每层都采用简单随机抽样的分层抽样。

注：如果从不同层抽出的个体/单位产品的比例与层在总体中的比例相等，则称为比例分配分层简单随机抽样。

[GB/T 10111-2008, 3.1.15]

3.4

单位数据项 data item

是组成商品数据的最小单元。例如每个属性值视为一个单位数据项，图片的每一个质量要求视为一个单位数据项。

3.5

样本数据集 sample data set

是样本数据中单位数据项的集合。

4 核查方案

4.1 核查流程

4.1.1 确定抽样方法，抽取样本数据。

4.1.2 确定样本数据的单位数据项数 M 。

4.1.3 核查样本数据中每个单位数据项，确定涉及的单位数据项数 m 。

注：涉及的单位数据项数是指该样本数据应涉及的所有单位数据项数，计算公式为：
 $m=M$ -冗余的单位数据项数+缺失的单位数据项数

4.1.4 出具核查结果。应符合本文件 6 核查结果。

4.2 抽样方法

应按照GB/T 10111—2008的方法，进行简单随机抽样或分层简单随机抽样。

4.3 抽样数量

4.3.1 规定每批数据应检查的样本数量，用 (N, n) 表示，式中 N 为批量， n 为抽取的样本量。其含义是从批量为 N 的批中抽取 n 个 GTIN 数据构成样本，并对样本全部检查。

4.3.2 确定接收质量限（AQL）：所使用的抽样方案及 AQL 应在合同中规定或由负责部门指定，可参考附录 A。

5 核查方法

5.1 技术审核

采用信息技术手段，按照《商品源数据 商品属性信息规范 第1部分：通用属性》和《商品源数据 数据质量控制实施规范》中的相关要求，对样本数据进行自动化校验。

对相互关联的属性建立逻辑关系，通过统计计算或估算进行技术判断。筛选可疑数据，标记错误数据。

属性值通过多个维度的规则对其进行约束。一个属性值可以被多个规则约束，一个规则也可以约束多个属性值。

5.1.1 数据规则

5.1.1.1 唯一

相同商品条码和包装外观的商品不应出现多条数据。

5.1.1.2 非空

属性值不应（或在满足某种条件时不应）出现空值。

5.1.1.3 无冗余

属性和属性值不应（或在满足某种条件时不应）出现多余或重复。

5.1.1.4 格式

属性值应符合规定格式。

5.1.1.5 值域

属性值应符合取值范围。

5.1.1.6 逻辑关系

属性值之间应具有某种逻辑依赖关系，如等于、大于、小于。

5.1.1.7 词法

属性值不应出现某个（类）字符、敏感词汇。

5.2 人工审核

依靠人力，按照《商品源数据 商品属性信息规范 第1部分：通用属性》和《商品源数据 数据质量控制实施规范》中的相关要求，对样本数据进行审核。

5.3 其他方法

可采用其他先进方法优化数据审核。

6 核查结果

6.1 结果分类

核查结果分为三类：商品图片（TP）、商品测量（CL）、商品属性（SX）。

6.1.1 商品图片结果

商品图片结果见表1。

表1 商品图片结果（缺陷编码表）

项目编号	项目	要求	缺陷编号	缺陷
TP-A	陈列图视图	必选正面、左面和顶面视图各1张，产品其他有内容面应拍摄图片；扁形商品必选的是正面和背面。	TP-A1	默认正面判断错误
			TP-A2	缺少必选图1张
			TP-A3	缺少必选图2张
			TP-A4	缺少其他有内容面图超过2张
			TP-A5	缺少其他有内容面图1张

TP-B	商品主图 俯角视图	除了扁平的特殊商品之外，应该拍摄3个不同的15°俯角3D视图，俯角拍摄时镜头中心对准产品正面中心点。所拍摄的角度误差不超过±5°。	TP-B1	缺少1张俯角视图
			TP-B2	缺少2张及以上俯角视图
			TP-B3	俯角角度大于15°+5°
			TP-B4	左右旋转角度误差超过±5°
TP-C	背景	纯白（RGB 255/255/255）	TP-C1	背景非纯白
TP-D	背景宽高比	1:1	TP-D1	背景宽高比非1:1
TP-E	商品图位置	在背景上居中	TP-E1	商品图在背景上未居中
TP-F	商品图边距	商品图最长边的顶点与背景边界的空白距离应占背景边长的2%~5%	TP-F1	商品图最长边的顶点与背景边界的空白距离占背景边长总长<2%
			TP-F2	商品图最长边的顶点与背景边界的空白距离占背景边长总长>5%
TP-G	商品轮廓	商品轮廓应完整、清晰	TP-G1	正面图商品轮廓不完整或不清晰
			TP-G2	非正面图商品轮廓不完整或不清晰
TP-H	色差	图片无偏色，与商品颜色一致	TP-H1	图片偏色，与商品颜色不一致
TP-I	对比度	对比度应整体平衡，高光均匀，不造成标签信息模糊	TP-I1	对比度严重不平衡
			TP-I2	高光不均匀，标签信息模糊
TP-J	反光和阴影	处理反光与阴影时，应注重细节和形状，不造成标签信息丢失	TP-J1	出现商品变形失真
			TP-J2	反光与阴影造成标签信息丢失
TP-K	多余内容	无指纹、污渍或水印等	TP-K1	出现与商品拍摄无关的物品（营销图片除外）
			TP-K2	有指纹、污渍或水印
TP-L	修图痕迹	图片缩放至100%时无明显修图痕迹	TP-L1	100%比例下有明显修图痕迹
TP-M	失真	压缩或放大时无失真	TP-M1	压缩或放大时失真
TP-N	像素	2400*2400~4800*4800	TP-N1	最低像素未达到2400*2400
			TP-N2	最高像素超过4800*4800
TP-O	分辨率	300dpi	TP-O1	分辨率非300dpi
TP-P	色彩模式	RGB	TP-P1	色彩模式非RGB
TP-Q	大小	≤7M	TP-Q1	大小>7M
TP-R	文件格式	文件格式符合要求	TP-R1	文件格式不符合要求
TP-S	命名	命名规范	TP-S1	命名不规范

6.1.2 商品测量结果

商品测量结果见表2。

表2 商品测量结果（缺陷编码表）

项目编号	项目	要求	缺陷编号	缺陷
CL-A	高宽深关系	高宽深关系正确	CL-A1	零售单元高、宽、深关系错误
			CL-A2	非零售单元高、宽、深关系错误

CL-B	数量级	尺寸、重量数量级与商品（包装）相符	CL-B1	尺寸数量级与商品（包装）不符
			CL-B2	重量数量级与商品（包装）不符
CL-C	数据完整	高宽深、重量数据完整	CL-C1	高宽深数据缺失
			CL-C2	重量数据缺失
CL-D	数值修约	按规则修约	CL-D1	未按规则修约。长度未精确到毫米，重量未精确到克。

6.1.3 商品属性结果

商品属性结果见表3。

表3 商品属性结果（缺陷编码表）

项目编号	项目	要求	缺陷编号	缺陷
SX-A	必填项	必填项不应漏填	SX-A-(1)	必填单位数据项漏填 1 次
SX-B	必填项	必填项不应多余或重复	SX-B-(1)	必填单位数据项多余或重复填写 1 次
SX-C	必填项	必填项属性值应符合包装标签描述	SX-C-(1)	必填单位数据项不符合包装标签描述 1 次
SX-D	必填项	必填项单位应正确	SX-D-(1)	必填单位数据项单位错误 1 次
SX-E	条件必填项	条件必填项不应漏填	SX-E-(1)	条件必填单位数据项漏填 1 次
SX-F	条件必填项	条件必填项不应多余或重复	SX-F-(1)	条件必填单位数据项多余或重复填写 1 次
SX-G	条件必填项	条件必填项属性值应符合包装标签描述	SX-G-(1)	条件必填单位数据项不符合包装标签描述 1 次
SX-H	条件必填项	条件必填项单位应正确	SX-H-(1)	条件必填单位数据项错误 1 次
SX-I	选填项	选填项不应多余或重复	SX-I-(1)	选填单位数据项多余或重复填写 1 次
SX-J	选填项	选填项属性值应符合包装标签描述	SX-J-(1)	选填单位数据项不符合包装标签描述 1 次
SX-K	选填项	选填项单位应正确	SX-K-(1)	选填单位数据项错误 1 次

6.2 结果报告

按照结果分类与缺陷编码表，出具核查结果明细表，可参考附录B。

附录 A 抽样方案 (规范性)

A.1 抽样方案

表 A.1 抽样方案

批量	样本量 字码	样本量	接收质量限 (AQL)				
			1.5	2.5	4.0	6.5	10
			Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
2~8	A	2	↓	↓	↓	0 1	↓
9~15	B	3	↓	↓	0 1	↑	↓
16~25	C	5	↓	0 1	↑	↓	1 2
26~50	D	8	0 1	↑	↓	1 2	2 3
51~90	E	13	↑	↓	1 2	2 3	3 4
91~150	F	20	↓	1 2	2 3	3 4	5 6
151~280	G	32	1 2	2 3	3 4	5 6	7 8
281~500	H	50	2 3	3 4	5 6	7 8	10 11
501~1 200	J	80	3 4	5 6	7 8	10 11	14 15
1 201~3 200	K	125	5 6	7 8	10 11	14 15	21 22
3 201~10 000	L	200	7 8	10 11	14 15	21 22	↑
10 001~35 000	M	315	10 11	14 15	21 22	↑	↑
35 001~150 000	N	500	14 15	21 22	↑	↑	↑
150 001~500 000	P	800	21 22	↑	↑	↑	↑
500 001及以上	Q	1 250	↑	↑	↑	↑	↑

↓——使用箭头下面的第一个抽样方案。如果样本量等于或超过批量，则执行100%检验。

↑——使用箭头上面的第一个抽样方案。

Ac——接收数。

Re——拒收数。

A.1.1 检索方法

根据批量，在抽样方案表（表A.1）中由该批量所在行向右，在样本量栏内读出样本量 n ，再以样本量所在行和指定的接收质量限所在列相交处，读出接收数Ac和拒收数Re。若在相交处是箭头，则沿着箭头方向读出箭头所指的第一个接收数Ac和拒收数Re，然后由此接收数和拒收数所在行向左，在样本量栏内读出相应的样本量 n 。

附 录 B 核查结果明细表
(资料性)

B.1 核查结果明细表

表 B.1 核查结果明细表 (例)

批次名称/编号: XXXX-XXXX

送审单位	XXX 公司	核查单位	河北分中心	N/AQL/n	2000/4.0/125		
序号	GTIN	样本单位数据项数 M	涉及单位数据项数 m	核查方法	核查类别	缺陷编号	缺陷描述
1	6901234567892	60	63	人工	图片 TP/测量 CL/属性 SX	TP-C1/TP-D1/SX-D-(3)	背景非纯白/背景宽高比非 1:1/条件必填项漏填累计(3)次
2	6901234567908	60	60	人工	图片 TP/测量 CL/属性 SX	/	
.....
125	6901234567991	60	60			/	