

文章编号: 1000-050X(2000)04-0348-52

GIS 数字产品质量抽样检验方案探讨^{*}

刘大杰¹ 刘 春¹

(1 同济大学测量与国土信息工程系, 上海市四平路 1239 号 200092)

摘要 针对 GIS 数字产品生产过程, 采用“二级检查, 一级验收”的检查原则, 初步探讨了对数字产品实行抽样检验的基本原理和抽样方案, 以 OC 曲线分析了传统百分比抽样方法的不足, 并提出了对 GIS 数字产品采用挑选型一次极限质量方案作为生产单位的二级检查抽样方案, 而采用挑选型一次检后平均不合格品率上限方案作为验收单位的一级验收抽样方案。

关键词 抽样检验; 数字产品; GIS

分类号 P207; P208 文献标识码 A

对 GIS 来说, 空间数据是基础, 非空间数据是内涵, 是地理单元的纵深描述。不难理解, 数据质量的优劣将直接影响 GIS 应用分析结果的可靠程度和应用目标的实现。GIS 数据的质量问题是伴随着数据的采集、处理和应用的过程而产生并表现出来的。

目前, 在 GIS 数据质量方面已经取得了一些研究成果, 其中对数据质量的基本特点、GIS 数据误差的主要来源、数据误差的累积过程、几何误差的检测和表达以及属性误差的来源、估算等进行了一系列的探讨^{1,2,3}。随着数字图形产品和 GIS 产业的兴起, 将会有更多的数字产品进入测绘市场和 GIS 应用领域并为经济建设服务。对其产品质量进行科学的检验和评价已迫在眉睫, 只有这样, 才可以确保越来越多的数字产品符合 GIS 应用的质量标准。

抽样检验可以对产品的质量管理提出可靠的信息, 是质量控制的基本手段。因此, 探讨 GIS 产品的抽样检验方法对 GIS 的质量控制也是十分重要的。测绘数字产品是 GIS 产品的关键组成部分, 为此, 本文以测绘数字产品为例, 探讨 GIS 产品的抽样检验方法。首先简要地说明现有测绘数字产品的质量标准中抽样检验方法的要点, 分析其抽样检验方案的接收概率, 然后对抽样检验方案提出改进意见。

1 测绘数字产品质量标准中抽样检验的要点

为了保证测绘数字产品的质量, 我国有关部门在调查和试验分析的基础上, 既考虑我国目前 GIS 产品发展的现状, 又考虑质量控制的一般原则性要求, 制订了一些测绘数字产品的质量标准和规定^{4,5}。这些标准和规定的制订, 对我国 GIS 产业的健康发展有着重要作用。

有关部门所制订的测绘数字产品的质量标准和规定, 对测绘数字产品要求采用抽样检验的方法来检验产品的质量, 其要点是:

1) 采用定“缺陷”的方法来量化产品检验中发现的各类问题。

2) 规定采用回放图件与实地或与数字化地形原图目视对照检查和上机检查。

3) 实行二级监督检查制和一级验收制。二级监督检查是小组的过程检查和生产单位的检查; 一级验收是用户的抽样验收检查。小组的过程控制检查要求 100% 全面检查, 实现过程控制; 生产单位的抽样检查一般要求不小于批量的 30%; 用户验收的抽样检验一般不小于批量的 10%。

4) 在评定样本中各个单位产品(幅)的质量品级后, 根据样本质量作出“经检验, 判定批合格或不合格”的结论。

可以看到, 以上生产单位和用户的抽样检验

都属于计数抽样检验, 所采用的抽样检验方法是一种简单抽样法, 或称为百分比抽样法。

测绘数字产品的质量标准的制订对保证测绘数字产品的质量可起到促进作用, 且简单抽样检验的方案具有简单方便、易于实现的优点。

2 测绘数字产品抽样检验方法的接收概率

对一批产品进行抽样检验时, 其检验结果可能被接收, 也可能被拒绝。通常称可能被接收的概率为接收概率。一批产品的接收概率与不合格品率 p 密切相关, 故常将接收概率记为 $L(p)$ 。对于批量为 N 的被检验产品, 一次抽样检验方案用 (n, c) 表示, 其中, n 表示抽样的样本大小, c 表示合格判定数。当检验后查出不合格品数为 d , 则判定规则是: 当 $d \leq c$ 时接收此批产品; 当 $d > c$ 时则拒收此批产品。

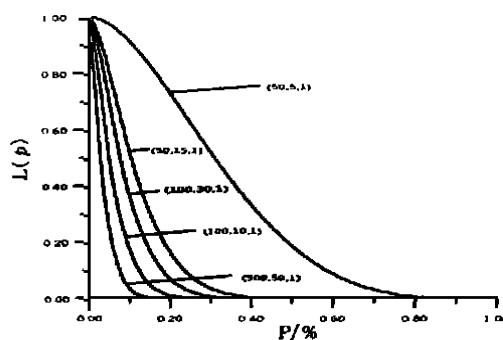
因批量为 N 的产品中不合格品数为 $D = N \cdot p$, 而抽取的 n 个样本中的不合格品数是一个服从超几何分布的随机度量, 其概率为:

$$P(d) = \frac{C_N^d C_{N-N}^{n-d}}{C_N^n} \quad (1)$$

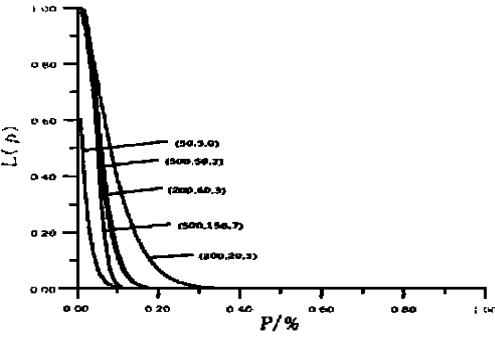
当 N 较大, $(n/N) \leq 0.10$ 时, 式(1)可以按照二项分布计算。所以按一次抽样方案 (n, c) 检验认为这批产品为合格的接收概率为:

$$L(p) = p(d \leq c) = \sum_{d=0}^c b(d, n, p) = \sum_{d=0}^c C_N^d p^d (1-p)^{n-d} \quad (2)$$

接收概率 $L(p)$ 是不合格品率 p 的递减函



(a) 单百分比



(b) 双百分比

图 1 抽样方案的 OC 曲线

Fig. 1 The OC Curve of Sampling Inspection

使用中才能体现出它的优越性, 而当前的 GIS 产品不属于这种产品, 故暂不宜采用。标准型抽样方案需要根据生产单位与使用单位分别提出的产

数, 通常将 $L(p)$ 与 p 之间的函数关系在一个直角坐标系中表示成抽样检验特性曲线(简称为 OC 曲线)。不同的抽样检验方案对应着不同的 OC 曲线。 OC 曲线是建立和选择抽样检验方案的一种依据。

在对测绘数字产品的抽样检验中, 规定生产单位抽样检验不少于批量的 30%, 用户单位抽样检验不少于批量 10% 的百分比抽样法。在实用上可以采用给定 c 后对样本数取 $n = N \cdot 10\%$ (或 $N \cdot 30\%$) 的百分比抽样, 也可以采用同时对判定数也取 $c = n \cdot k\%$ 的所谓双百分比抽样方案。

图 1(a) 是当批量 N 为 50、100、200、500, 分别抽样 30% 和 10%, 且取判定数 $c = 1$ 时的 OC 曲线; 图 1(b) 是取 $c = n \cdot 5\%$ 时的 OC 曲线。

由图 1 可以看到, 单百分比和双百分比都存在“大批量严, 小批量宽”的缺点^[6,7]。然而对于测绘数字产品, 由于文献[4]中规定以“幅”为单位产品, 一“幅”的内容相当多。在实用上, 批量 N 一般不会很大, 因此, 一般来说, 在采用百分比抽样方法时, “大批量严”的问题不会很严重。在批量 N 较小时, 适当选取判定值 c 对“小批量宽”的问题也会得到一些改善。但从 GIS 数字产品的现代化质量管理来看, 有必要选择更科学的抽样检验方案。

3 抽样检验方案的选择

产品的抽样检验方案有 3 种类型: 标准型、调整型和挑选型。调整型抽样检验方案主要适用于连续生产和验收的单位产品, 且必须在长期连续

品质量水平 p_0 和 p_1 来确定抽样检验方案, 而当前生产单位和用户单位对 p_0 和 p_1 的选取仍缺少经验和依据。

挑选型抽样方案的特点是对判为接收的批直接通过, 对判为拒收的批必须经过 100% 的检查, 将其中不合格品挑出来, 换成合格的(或修复为合格品)然后再接收。挑选型记数一次抽样方案有“一次极限质量抽样”和“一次检后平均不合格品率上限抽样”两种。下面分别讨论这两种方案与国家标准 GB/T 13546 规定的挑选计数一次抽样方案及实施程序的关系。

3.1 计数挑选一次极限质量抽样

对于一批产品, 其批量为 N 。假定此种产品的平均不合格品率为 \bar{p} , 如果采用一次抽样方案 (n, c) , 当 N 较大, $(n/N) \leq 0.10$, $\bar{p} \leq 0.10$ 时, 接受概率可按式(2)计算。按照均值的概念, 挑选型一次抽样方案的平均检验件数为:

$$I = n + (N - n) [1 - \sum_{d=0}^c b(d, n, \bar{p})] \quad (3)$$

如果规定一个极限不合格品率为 p_t , 这个极限不合格品率的接受概率为:

$$L(p_t) = \sum_{d=0}^c b(d, n, \bar{p}) \quad (4)$$

这里, $L(p_t)$ 可看作用户的风险, 所以挑选型一次极限质量抽样的准则就是给定用户风险 $L(p_t)$ 和极限不合格品率 p_t , 对任意给定的 N 和 \bar{p} , 求 c 与 n 的值, 并使得给出的 I 值最小。表 1 是挑选型一次极限质量抽样方法对应不同极限质量、不同批量大小所采用的抽样方案。

表 1 挑选型一次极限质量抽样方案表

Tab. 1 Sampling Schemes for One Time Limiting Quality Sampling Inspection of Count Selection

极限质量 $p_t/ \%$	批量大小 N	抽样方案 (n, c)
1.25	1~150	全检
	150~280	(122, 0)
	1~90	全检
2.00	91~150	(80, 0)
	151~280	(89, 0)
	1~50	全检
3.15	51~90	(48, 0)
	91~150	(55, 0)
	151~280	(63, 0)
5.00	1~50	全检
	51~90	(33, 0)
	91~150	(37, 0)
8.00	151~280	(42, 0)
	1~25	全检
	26~50	(22, 0)
	51~90	(25, 0)
	91~150	(26, 0)
	151~280	(27, 0)

由表 1 可以看到:

- 1) 当批量 N 很小时 ($N < 25 \sim 50$), 要求全部检查。
- 2) 当取极限质量为 $p_t = 5\%$, $N = 91 \sim 150$ 时, $(n, c) = (37, 0)$, 或 $p_t = 3.15$, $N = 151 \sim 280$ 时, $(n, c) = (63, 0)$, 抽样量 (n/N) 约为 30%。
- 3) 当 N 更大时, $(n/N) < 30\%$ 。

3.2 计数挑选型一次检后平均不合格品率上限抽样

设有一产品总体, 不合格品率为 p , 对批量为 N 的产品采用一次抽样方案 (n, c) , 用 AOQ 表示检后平均不合格品率, 有:

$$AOQ = \frac{N - n}{N} p \cdot L(p) + 0 \cdot [1 - L(p)] = \frac{N - n}{N} p \cdot L(p) \quad (5)$$

式中, $L(p)$ 是不合格品率为 p 时的样本接收概率。当 N 较大, $(n/N) \leq 0.10$, $p \leq 0.10$ 时, AOQ 可用泊松分布形式表示为:

$$AOQ = \frac{N - n}{N} p \cdot L(p) = \frac{N - n}{N} p \cdot \sum_{d=0}^c \frac{(np)^d}{d!} e^{-np} = \left(\frac{1}{n} - \frac{1}{N} \right) (np) \cdot \sum_{d=0}^c \frac{(np)^d}{d!} e^{-np} \quad (6)$$

对于既定方案 (n, c) 而言, 由式(5)可知, AOQ 是 p 的函数, 令 $(AOQ)'_p = 0$, 得:

$$\sum_{d=0}^c \frac{(np)^d}{d!} e^{-np} = \frac{(np)^{c+1}}{c!} e^{-np} \quad (7)$$

若记 $x = np$, 则有:

$$\sum_{d=0}^c \frac{x^d}{d!} e^{-x} = \frac{x^{c+1}}{c!} e^{-x} \quad (8)$$

记 $y = x \cdot \sum_{d=0}^c \frac{x^d}{d!} e^{-x}$, 则 AOQ 的上限 p_L 为:

$$p_L = \left(\frac{1}{n} - \frac{1}{N} \right) y \quad (9)$$

由上面的公式可得:

$$y = \frac{x^{c+2}}{c!} e^{-x} \quad (10)$$

$$n = \frac{N \cdot y}{y + N \cdot p_L} \quad (11)$$

所以, 挑选型一次检后平均不合格品率上限抽样的准则就是给定检后平均不合格品率上限 p_L 和 N , 任意固定值 c , 由式(8)确定 x , 式(10)确定 y , 再由式(11)确定 n , 得出的多对 (n, c) 中, 寻找同样使得给出的 I 值最小的那一对抽样方案。这里, I 值的确定同样由给定的平均不合格品率 \bar{p} 根据式(3)计算。表 2 是挑选型一次平均不合格品率上限抽样方法对应不同检后平均不合格品率

上限、不同批量大小所采用的抽样方案。

由表 2 可以看到:

1) 批量 N 很小时 ($N < 16 \sim 25$), 要求全部检查。

2) 当检后平均不合格品率上限为 $p_L = 2\%$, $N = 91 \sim 150$ 时, $(n, c) = (15, 0)$; 当 $p_L = 5\%$, $N = 51 \sim 90$ 时, $(n, c) = (7, 0)$, 抽样量 (n/N) 约为 10%。

3) 当 N 更大时, $(n/N) < 10\%$ 。

表 2 挑选型一次平均检后不合格品率上限抽样方案表

Tab. 2 Sampling Schemes for One Time Afterinspection

Percent Defective Upper Limit of Count Selection

检后平均不合格品率上限 $p_L / \%$	批量大小 N	抽样方案 (n, c)
0.2	1 ~ 90	全检
	91 ~ 150	(83, 0)
	151 ~ 280	(111, 0)
0.5	1 ~ 50	全检
	51 ~ 90	(41, 0)
	91 ~ 150	(50, 0)
0.8	151 ~ 280	(58, 0)
	1 ~ 25	全检
	26 ~ 50	(24, 0)
2.0	51 ~ 90	(31, 0)
	91 ~ 150	(35, 0)
	151 ~ 280	(40, 0)
5.0	1 ~ 16	全检
	17 ~ 25	(11, 0)
	26 ~ 50	(14, 0)
5.0	51 ~ 90	(15, 0)
	91 ~ 150	(16, 0)
	151 ~ 280	(17, 0)
5.0	1 ~ 16	全检
	17 ~ 25	(6, 0)
	26 ~ 50	(6, 0)
5.0	51 ~ 90	(7, 0)
	91 ~ 150	(7, 0)
	151 ~ 280	(7, 0)

3.3 GIS 数字产品抽样检验方案的选择

以上结果说明, 当批量 N 不太小也不太多时, 采用极限质量抽样方案的抽检量接近 30%, 采用检后平均不合格品率上限方案的抽检量约为 10%, 但它们都可以克服“批量大严, 批量小宽”的缺点。因此, 建议选取挑选型极限质量抽样方案作为生产单位的二级检查抽样方案, 极限质量可取 3% ~ 5%; 选取挑选型检后不合格品率上限抽样方案作为用户单位的验收抽样方案, 检后不合格品率上限可取 2% ~ 5%。

在抽样检验中, 根据合格判定数来判断检验

批是否合格可能出现两种错误判断: ①将合格批判断为不合格批, 这种错误将使生产方蒙受损失; ②将不合格批判断为合格批, 这将使用户蒙受损失。通常将产生错误①的概率记为 α , 称为第一类风险或生产方风险; 而将产生错误②的概率记为 β , 称为第二类风险或使用方风险。

为了控制这两种风险, 在抽样检验时, 往往规定两个不合格品率 p_0 和 p_1 ($p_0 < p_1$), 一般称 p_1 为极限质量。当一批产品的批不合格品率 $p < p_0$ 时, 认为这批产品质量较好, 应该以大概率接收, 或以小概率拒收, 通常将拒收概率控制在 α 之内; 而当批不合格品率 $p > p_1$ 时, 就认为这批产品的质量较差, 应当以小概率接收, 以大概率拒收。容易理解, α, β 与 p_0, p_1 有一定的关系, 并与抽样检验方案中的样本大小 n 和判定数 c 有关。采用抽样检验时, 两类错误是不可避免的, 或者说两种风险都存在。为了使双方的风险控制在一定的范围内, 更应该合理地确定 p_0, p_1 和 α, β 以及相应的抽样检验方案。

挑选型计数抽样模型依据的是控制用户方风险, 同时使抽样的总件数最小, 也即抽样的总费用最小。所以挑选型计数抽样方法适合于检查和验收检验, 它一方面充分考虑了用户的风险, 同时也在最小的抽样费用下严格控制了抽样误差。关于这方面的问题笔者将另文作进一步讨论。

4 结语

对 GIS 而言, 空间数据的采集往往是按批量有序进行的, 这样就可以把空间数据的采集看作为一个数字产品的生产过程, 对生产过程中数字产品质量的抽样检验就可以选用符合国标的一系列抽样方法。本文初步探讨 GIS 数字图形产品抽样检验所涉及的基本原理与方法, 并说明对数字产品采用传统的百分比抽样所引起的不足。然后提出了在生产单位的二级检查中采用挑选型极限质量抽样方案, 而在用户验收时采用挑选型检后不合格品率上限的抽样方案的改进建议。这两种方案既与目前采用的百分比抽样方案在一定情况下一致, 也简便易行。由于 GIS 数字图形产品涉及的内容和质量特性都比较复杂, 既有计数质量特性又有计量质量特性, 所以, 抽样检验模型仍有待进一步深入研究, 使之满足不同的抽样水平, 适合不同复杂程度的数字图形产品, 在使得抽样费用最小的情况下保证抽样的可靠性。

参 考 文 献

- 1 陈俊, 宫鹏. 实用地理信息系统. 北京: 科学出版社, 1998
- 2 阎正. 城市地理信息系统标准化指南. 北京: 科学出版社, 1998
- 3 Shi W Z. Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems. ITC Publication, Enschede, 1994
- 4 国家测绘局. 数字地形图产品质量标准(讨论稿). 1998
- 5 李佐清.“数字地形图产品质量标准”编写说明. 国家测绘局, 1998

- 6 于善奇. 抽样检验与质量控制. 北京: 北京大学出版社, 1991
- 7 马瑞, 王志成. 抽样检查计数. 哈尔滨: 哈尔滨工业大学出版社, 1993
- 8 刘大杰. GIS 空间数据的精度分析与质量控制. 上海: 上海科技文献出版社, 1999
- 9 马毅林, 于振凡, 于善奇, 等. 产品质量抽样检验. 北京: 中国标准出版社, 1997

刘大杰, 男, 60岁, 教授, 博士生导师。现从事测量数据处理和GIS空间数据质量研究。近期代表成果:《GIS空间数据的精度分析与质量控制》。

E-mail: Djliu@ceiliang.tongji.edu.cn

Study on Sampling Inspection Schemes to Digital Products in GIS

LIU Dajie¹ LIU Chun¹

(1 Dept. of Surveying and Geo-informatics, Tongji University, 1239 Siping Road, Shanghai, China 200092)

Abstract Spatial data is the basis of the GIS and non-spatial data is intention and the description of geographic cell. The quality of data affects greatly on the result of application analysis and the realization of the practical target in GIS. The quality of the GIS data must be taken into account in the procedures of data collection, processing and operation.

With the development of digital products and industrialization of GIS, more and more digital products appear on the surveying market and become the basis of applications. But for economical reasons a producer or purchaser of large database cannot perform manual check for the complete amount of data. Nevertheless, he or she should carry out some type of quality control specification and evaluation to the quality of digital products. The inspection can be done by drawing random samples, checking the objects within each sample very carefully and concluding from the result on the complete dataset using statistical method. Very often sampling methods are used for randomly checking purchased geodata, but very seldom these checks are based on solid statistical theory. So with the development of the standard of quality management and quality guaranty, the requirement of scientific inspection to digital products should be update. At present, there are about 20 national standards published on sampling inspection, which play a fine directorial role in practice. The theory of quality of control and inspection can't be used for GIS production perfectly unless the concept, range of the application, method and procedure of these standards have further been studied on making a good basis for the development of the industrialization in GIS.

The sampling inspection can provide reliable information for the quality management of products. As a basic means of quality control, it is also very important to study the sampling inspection for quality management of GIS products. For the digital surveying product it is a key component in digital surveying products, adopting a principle of “check-accept for the first rank, inspection for the second rank”. This paper briefly discusses the rationale of the sampling inspection method and the sampling inspection schemes to the digital product. The OC curve is also drawn to explain the deficiency of the percent sampling procedures for inspection. Meanwhile, the method of one time

(下转第 361 页)

Estimation of the Unit Weight Variance in Nonlinear Model Adjustment

WANG Xinzhou¹

(1 Science and Technology Office, WT USM, 129 Luoyu Road, Wuhan, China 430079)

Abstract In parameter estimation many models are nonlinear ones. The classical method dealing with these nonlinear models is linear approximation using the approximate value of parameter. Due to the fact that the difference of non-linearity between different nonlinear models, some nonlinear model can be linear approximation and the others can be not has been understand, the theory of parameter estimation for nonlinear model has been studied in many papers. But none studies the estimation of the unit weight variance in parameter estimation for nonlinear model. In parameter estimation for nonlinear model, how to estimate the unit weight variance? This paper specially studies the problem. At first this paper gives the expanded formula expectations, and variance of residuals. Then according to the theoretical relationship of residuals and their expectation and their variance, the formula of the unit weight variance in parameter estimation for nonlinear model is presented. At last we give an example to explain how to use the formula.

Key words nonlinear model; unit weight variance; residual

WANG Xinzhou, male, 46, professor, Ph. D supervisor. He is concentrated on the research and education in the theory and application of surveying data processing. His cardinal achievements include in the field the theory and method of deformation measurements the theory and algorithm of estimation of stochastic the parameter estimation theory of nonlinear. His published papers are more than 40.

E-mail: xzwang@wtusm.edu.cn

(上接第 352 页)

limiting quality of count selection is raised as the inspection scheme for production departments while the method of one time after-inspection average percent defective upper limit of count selection is proposed for acceptance departments. These two schemes are in agreement with existing percent sampling inspection schemes in certain circumstances, and have the advantage that can be easily conducted. More contents and quality characteristics have been involved in the digital product in GIS, and the model of sampling inspections still need a further research. The result of the research can be used in different sampling levels and suited for variant and complex digital products, and ensure the reliability of sampling while make the fee of sampling minimum.

Key words sampling inspection; digital products; GIS

LIUDajie, male, 60, professor, Ph. D supervisor. His chief research aspects involve geodata analysis and research of spatial data quality in GIS. His recently typical achievement is “Accuracy Analysis and Quality Control of Spatial Data in GIS”.

E-mail: Djliu@celiang.tongji.edu.cn