



分层随机抽样技术评价供水管网水质

童桢恭^{1,2}, 刘遂庆¹, 陶涛¹

(1 同济大学 污染控制与资源化研究国家重点实验室, 上海 200092 2 华东交通大学
土木学院, 江西 南昌 330013)

摘要: 目前, 自来水公司大多是依赖从输配水管网水质监测点检测出的各项水质参数数据来评判整个管网乃至用户的水质情况, 很少采用正式的统计样本原则进行样本采集。为此提出以最小方差和偏差的方法论, 采用分层随机抽样技术进行样本设计, 从而对管网水质进行较为真实的全面评价, 同时对影响管网水质的因素进行分析。最后以美国某城市管网为例说明分层技术在管网水质评价中的应用情况。

关键词: 分层随机抽样; 给水管网水质; 样本设计

中图分类号: TU991.2 **文献标识码:** B **文章编号:** 1000-4602(2005)03-0031-04

Stratified Random Sampling Technology for Water Quality Assessment in Distribution System

TONG Zhen-gong^{1,2}, LIU Su-qing¹, TAO Tao¹

(1 State Key Lab of Pollution Control and Resource Reuse Tongji University Shanghai
200092 China; 2 School of Civil Engineering East China Jiaotong University Nanchang
330013 China)

Abstract Presently water utilities mostly rely on the data collected from water quality monitoring station of water distribution system to estimate the quality of drinking water delivered to customers and rarely use the formal statistical samples protocol to collect these samples. Therefore, the minimum variance and deviation methodology was introduced, based on which, stratified random sampling technology was used for sample design, and thus making an overall assessment more truly on water quality in distribution system, and getting an insight into the factors affecting water quality. Finally, the application of stratified technology in assessment of water quality was explained by taking an American city's pipe network for example.

Key words stratified random sampling water quality in water distribution system; sample design

我国卫生部在 2001 年颁布的《生活饮用水水质卫生规范》中有明确的规定: 城市集中式供水管网的水质检验采样点数, 一般应按供水人口每两万人设一个采样点计算。供水人口超过 100 万时, 按上述比例计算出的采样点数可酌量减少。人口在

20 万以下时, 应酌量增加。在全部采样点中应有一定的点数选在水质易受污染的地点和管网系统陈旧部分等处, 具体采样点的选择应由供水单位与当地卫生监督机构根据本地区具体情况确定。

美国环境保护署 (USEPA) 在《安全饮用水法



(SDWA) 中对管网水质监测点也作出了规定: 对细菌性指标监测的采样点最少为 1 个/(1 000 人·月); 对于消毒副产物的监测, 每季度对每个水源的管网水至少要有 4 个采样点。

在采样点的选择上一般是根据地形结构确定。出于经济方面的原因以及人们的主观原因, 目前采样点很多是放在主要街道上的, 而且采样时间也经常是在早上, 这样的实测数据显然不能代表整个管网的实际情况, 即便检测数据中显示没有任何水质问题, 也有可能爆发水传播疾病。

1 分层随机抽样技术的应用

抽样调查是根据概率论的原理, 按照随机原则从调查对象总体中抽取一部分总体单位作为样本进行调查, 然后用样本指标以一定的精度和置信度去推断总体的相应指标, 即参数估计。如果总体由明显差异的几个部分组成, 为得到代表性较好的样本, 先按照差异情况将总体划分为若干子总体(层), 然后在各层进行随机抽样, 这就是分层随机抽样。其特点是经过分层将一个内部差异较大的总体划分为内部总体差异较小的多个小总体(层), 以达到提高抽样的经济效益和估计精度的目的。分层抽样是为了在不增加样本数的情况下提高抽样精度, 或者在一定精度要求下减少样本数的一种常用调查方法。该方法适用于总体情况复杂、各单位之间差异较大、单位较多的情况, 可以根据时间和空间的变化来进行分层研究。其主要流程如图 1 所示。

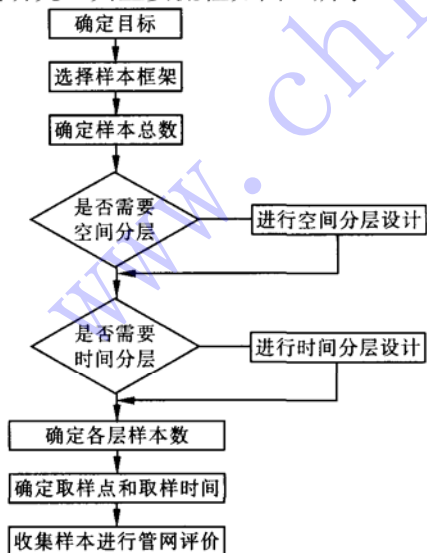


图 1 分层抽样方法流程图

Fig 1 Overview of sampling methodology

2 应用实例

2.1 背景

所引实例的基础资料来源于 EPANET 软件中的范例。实测数据的采集在康州中南水务局 (SC-CRWA) 所辖的供水服务区施行, 以 Cherry Hill/Bushy Plains 服务区为例说明。该区的供水服务面积大约为 5.2 km², 是整个 SCCRWA 的输配水管网系统中一个相对独立的区, 由位于管网系统南端的 Saltonstall 给水处理厂通过加压泵房向管网供水, 而在管网的北端, 有高位水池对管网的用水进行调节。

该区共有 40 根管段, 管径只有两种 (200 mm 和 300 mm), 管网全长为 10.97 km。在 1991 年 8 月 13 日到 15 日, 研究人员对该管网连续进行了 55 h 的余氯浓度采样检测, 共在 9 个不同采样点获得了 181 组有价值的原始数据。研究中样本设计的主要目的是评价管网中低余氯浓度的范围以及分析余氯衰减的主要原因。样本总数 $N = 55 \times 40 = 2\,200$ (相当于 40 根管子每小时模拟 1 次余氯浓度)。

2.2 样本总体设计

根据当地实际情况, 该服务区的管网水余氯浓度保持在 0.2 mg/L 以上为合格, 低于 0.2 mg/L 则意味着水质可能变坏。样本总体估计值可由比率 P 来定义。比率 P 是指氯浓度 < 0.2 mg/L 的比率:

$$P = \left(\sum_{i=1}^n Y_i \right) / n \quad (1)$$

式中 Y_i ——变量值, 如果所取样本值 ≥ 0.2 则取 0, 反之则取 1。

P 值一般可根据历史数据进行估计, 对于该服务区来说, 根据历史数据得出氯浓度 < 0.2 mg/L 的概率大约是 0.2 (即 20%)。根据置信水平, 利用公式 $n = \frac{t^2 P(1-P)}{\Delta^2 N + t^2 P(1-P)}$ (其中 t 为置信水平, 查表可得出; Δ 为边际误差) 就可推算出样本规模数, 结果见表 1。

表 1 $\alpha = 0.2$ 时所需最大样本数

Tab 1 Maximum number of samples required in 55 hours ($P = 0.2$)

置信度 / %	$d = P(0.2)$ 时的边际误差				
	0.01	0.02	0.03	0.04	0.05
80	1 204	511	260	154	101
85	1 322	602	315	189	125
90	1 461	728	396	240	161
95	1 620	904	521	327	221
97.5	1 726	1 049	634	408	280



2.3 分层数设计

由于管网水质随时空变化而变化, 故样本设计中考虑以时间和空间进行分层。空间分层主要是考虑以下 3 种分层: 管材、管径和离水厂的距离。因为这 3 个变量对余氯在管网中的衰减都有较大的影响。在时间方面主要考虑两个分层: 即用水高峰和非高峰时。空间变量分层结果见表 2。

表 2 空间变量分层

Tab 2 Spatial stratification variables

空间分层变量		样本数 N_h	备注
管径 / mm	200	20	共 40 条管段
	300	20	
离水厂距离 / km	≤ 1.5	6	共 40 条管段
	1.5~3.5	11	
	≥ 3.5	23	

对于时间分层, 根据其用水量变化曲线及时变化系数, 可得出在模拟的 55 h 中, 有 25 h 的时变化系数 > 1.0 故认为高峰时有 25 h 剩余的 30 h 为非高峰用水。

2.4 层间样本量设计

样本分配方法大体有 2 类, 即等数分配法 (平衡分配法) 和按比例分配法。比例分配使各层有统一的样本率, 在整体估计时常用, 因为它具有最小的标准误差。平衡分配就是各层中所取的样本数相同, 它主要用于比较各层变量因素对整体的影响。

两种分配方法在不同情况下的样本量如表 3 所示 (取 $P=0.2$, $\Delta=0.05$ 置信度为 85% 时的总体量为 125, 为便于计算, 取 120)。

表 3 平衡分配时每层的取样数

Tab 3 Number of samples per stratum for balanced allocation

设计次数	空间分层		时间分层	
			高峰时	非高峰时
1	管径 / mm	200	30	30
		300	30	30
2	管径 / mm	200	60 无时间分层	
		300	60 无时间分层	
3	离水厂距离 / km	≤ 1.50	20	20
		1.5~3.5	20	20
		≥ 3.5	20	20
4	离水厂距离 / km	≤ 1.50	40 无时间分层	
		1.5~3.5	40 无时间分层	
		≥ 3.5	40 无时间分层	
5	None	在所在管道中	60	60

在时间和空间分层基础上用平衡及比例分配方

法进行了不同组合, 共组合出 11 次的样本设计, 如表 4 所示。

表 4 比例分配时每层的取样数

Tab 4 Number of samples per stratum for proportionate allocation

设计次数	空间分层		时间分层	
			高峰时	非高峰时
6	管径 / mm	200	27	33
		300	27	33
7	管径 / mm	200	60 无时间分层	
		300	60 无时间分层	
8	离水厂距离 / km	≤ 1.50	8	10
		1.5~3.5	15	18
		≥ 3.5	31	38
9	离水厂距离 / km	≤ 1.50	18 无时间分层	
		1.5~3.5	33 无时间分层	
		≥ 3.5	69 无时间分层	
10	None	在所在管道中	54	66
11	None	在所在管道中	120	

2.5 结果与讨论

模拟结果如表 5 所示。每个样本设计的有效性可通过与标准误差比较来评价。

表 5 模拟结果分析

Tab 5 Simulation analysis results

样本设计	分配方法	空间分层	时间分层	模拟标准误差 / 10^{-2}	实际标准误差 / 10^{-2}	实际标准误差 $\div$ 模拟标准误差
1	平衡分配	管径	高峰 / 非高峰	4.41	3.48	1.04
2	平衡分配	管径	None	3.86	3.21	0.95
3	平衡分配	离水厂距离	高峰 / 非高峰	5.34	3.47	1.03
4	平衡分配	离水厂距离	None	3.29	2.76	0.82
5	平衡分配	None	高峰 / 非高峰	3.77	3.16	0.94
6	比例分配	管径	高峰 / 非高峰	4.48	3.33	0.99
7	比例分配	管径	None	3.7	3.03	0.90
8	比例分配	离水厂距离	高峰 / 非高峰	3.6	2.23	0.67
9	比例分配	离水厂距离	None	3.88	2.54	0.76
10	比例分配	None	高峰 / 非高峰	3.11	3.43	1.02
11	随机抽样	None	None	3.45	3.35	1.0

余氯浓度采用 EPANET2.0 软件模拟, 模拟结果采用 Excel 进行统计分析。设管网各部分的水温 (取 20°C) 和 pH ($\text{pH}=7.0$) 值相同且恒定。由于受各种条件所限, 余氯在水体中的衰减系数 K_s 和管壁消耗系数 K_w 对于所有管道取相同的值, 即取 $K_s = -0.55/\text{d}$, $K_w = -0.15\text{ m/d}$ 。

范例中的管网具有水力模型, 因此可用软件模拟整个管网每条管段不同时刻的余氯浓度。采用公



式 $V = \sum_{h=1}^n W_h^2 \left[\frac{1-f_h}{n_h} S_h^2 \right]$ 计算实际标准误差 [其中 V 为标准误差的平方即方差, W_h 为各层的权重, f_h 为各层的抽样比, n_h 为样本数 (此中为 120), $S_h^2 = P(1-P)$]。

每个样本都可与简单随机样本 (即样本 11) 相比较评价它们之间的相关有效性, 比值 < 1 说明该样本设计比样本 11 (简单随机) 更有效, 它产生相对较低的标准误差。同理, 比值 > 1 则说明该样本设计比样本 11 效果更差。

理论上说, 比例分配 (样本 6~10) 的整体误差将比随机抽样 (样本 11) 的更低, 这点从表 5 也可看出。根据样本数的不同, 平衡分配中也有可能比简单随机 (样本 11) 更有效, 如样本 2、4、5。总体来说, 样本 8、9 (以离出厂水距离为空间分层) 最有效。

根据模拟分析结果可进行样本设计选择。如样本目标是最精确估计总体水质 (统计准确率最高) 则样本设计 8 或 9 为首选。然而, 如果样本目标是评价管径的影响, 那么样本设计 1、2、6、7 则为首选。

评价不同管材对水质的影响有助于水司制定管材维护计划。由于此例中管材信息不详, 所以无法进行管材对水质影响的分析, 但可通过分层技术估计低水质的影响范围。

3 评估低水质区域

主要以氯浓度作为衡量指标。样本要以最低的整体误差为最佳。模拟结果显示样本设计 8 和 9 的总体误差最小。样本 9 用离水厂距离作为空间分层且没有时间分层, 层中采用比例分配。

样本 9 的分析结果见表 6。总体氯浓度 < 0.2 mg/L 的 P 值为 31.1%。结果显示, 在距离水厂近的管段, 其水质恶化的可能性较低, 可能是停留时间

较短的缘故; 而远离水厂的管网水水质恶化的可能性就较大。样本 9 不仅显示了整个管网中低水质范围, 而且还给水司提供了很多有用的信息。

表 6 样本 9 结果

Tab 6 Results of sample 9

空间子层	离水厂距离	平均 $P\%$	标准误差 $/10^{-2}$
	整体	31.1	0.35
1	< 1.5 km	24.6	0.38
2	$1.5 \sim 3.5$ km	28.4	0.13
3	> 3.5 km	34.1	1.8

导致水质恶化的另一重要变量是水龄。在此没按水龄分层的原因是: 水龄是具有时间和空间概念的, 不能单用于严格的空间或时间分层中; 另外, 水龄也是个很难预测的变量, 它需要一个较好的水力模型。

4 结语

随机分层抽样技术对水司的运行维护提供了很好的工具。特别是对不同情况时的分析, 对某一给定管网系统, 选择不同的分层变量可满足不同需求。同时可借助历史数据或类似的系统知识结合到分层技术中。统计框架中提供了估计的标准误差和置信度水平, 这些对决策都是非常有用的。

参考文献:

- [1] Pratim Biswas A model for chlorine concentration decay in pipes[J]. Wat Res 1993 27(12): 1715- 1724
- [2] Rossman L A, Clark R M Modeling chlorine residuals in drinking water distribution systems[J]. Jour Envir Engng 1994 120(4): 803- 806

电话: (021) 65985869

E-mail zgong7212@ 163.net

收稿日期: 2004- 10- 20