

文章编号: 1002-0268 (2001) 04-0079-06

公交线路客流时段分布估计新方法 ——首末站分层不等概率整群抽样法

杨新苗, 王 炜, 石小法, 尹红亮

(东南大学交通学院, 江苏 南京 210096)

摘要: 在对南京市公交线路客流调查数据研究的基础上, 提出一种公交线路客流时段分布估计的新方法, 即利用首末站分层不等概率整群抽样方法, 对线路客流进行估计, 并给出相应的方差估计公式。该方法在南京市公交企业中得到了实际的应用。

关键词: 线路客流时段分布估计; 客流预测; 公共交通; 运营调度优化

中图分类号: U492.4

文献标识码: A

Estimation of Temporal Patronage Distribution of Bus Lines by Stratified Clustering Sampling

YANG Xin-miao, WANG Wei, SHI Xiao-fa, YIN Hong-liang

(Transportation College Southeast University, Jiangsu Nanjing 210096 China)

Abstract: On studying the survey data from Nanjing, a new method based on the Stratified Clustering Sampling methodology has been raised in this paper, using sampling data from bus terminals to estimate temporal patronage distribution of bus lines, and an error function has been given. This method has got practical use in Nanjing public transit management.

Key words: Estimation of temporal patronage distribution; Patronage forecasting; Public transit; Transit operation optimization

0 公交线路客流时段分布估计

公交线路客流的时段分布是线路客流在全日各个时间段上的流量大小分布。常用的时间段单位是小时。

在公交线路调度管理人员制定线路运营计划的过程中, 需要掌握的两个重要信息是: 首先是线路客流在一天中总的流量大小; 其次是总流量在全天各个时段的分布。如此才能制定优化的运营计划。关于线路全日客流的总量大小, 一般可由线路每日运营收入资料推算获得。一般企业都有较为详细的线路运营收入资料, 客流总量的估算较为容易。

同线路客流总量相比, 线路客流时段分布数据的获得则较为困难。虽然在一些使用了公交 IC 卡系统

的城市, 可以借助 IC 卡系统获得线路时段分布信息, 但是要看到的是 IC 卡还没有得到广泛地普及和应用。对于大多数公交企业来讲, 目前获得线路客流时段分布数据的主要方式还是依靠人工调查。只有组织调查员, 到现场各个站点进行调查, 才能获得所需的时段分布数据。但是人工调查投入太大, 一般是每年或隔几年组织一次, 不可能经常进行。而对于线路客流时段分布信息的需求是每天都存在的。因此需要形成一种方法, 通过抽取少量关键站点, 由站点的客流时段分布规律, 推知线路客流时段分布的规律。

本文在对南京市公交线路客流调查数据研究的基础上, 提出了一种公交线路客流时段分布估计的新方法, 即利用首末站分层不等概率整群抽样方法, 对线路客流的时段分布进行估计, 并给出了相应的方差估

收稿日期: 2000-09-05

基金项目: 国家自然科学基金重点项目 (59838310) 及教育部优秀青年教师基金项目

作者简介: 杨新苗 (1974-), 男, 湖北武汉人, 东南大学博士研究生, 研究方向为交通运输规划与管理。

计公式。该方法是一种理论上高效，实用性强的新方法，并且在南京公交雅高巴士公司的调度管理中得到了实际的运用。

1 站点客流时段分布的特点^[1]

公交线路由一系列站点组成，线路客流的形成和变化，实际上是各个站点客流共同作用的结果。

公交站点的客流受到许多因素的影响，例如站点周边用地的性质、站点的换乘条件等。多因素的作用造成了站点客流变化的复杂性。为掌握站点客流的时段分布规律，需要进行大规模的实际调查和观测。

为配合南京市新一轮城市交通规划的进行，南京市公共交通总公司于 1997 年 5 月 23 日（星期五）零时起到二十四时止在南京市主城区 42 条公交线路上进行了全日随车客流的全样调查工作。调查总共出动调查人员 3 915 人，调查了 42 条线路，813 辆车，11 114 个车次。这 42 条线路基本上覆盖了南京市主城区内的主次干道，可以反映南京市主城区内的公交客流分布状况。这次随车客流调查为本研究提供了详实准确的数据。

线路客流时段分布估计关心的是客流在各个时段所占的比例。按以下公式进行标准化处理。

$$x_{ik} = X_{ik} / \sum_{k=1}^p X_{ik}$$

式中， X_{ik} 为某线路 i 站点第 k 个时段的上客量，单位：人次； x_{ik} 为处理后得到的上客量时段比例，单位：%， $x_{ik} \leq 1$ ，时段单位取为小时。

为进一步研究站点分布之间，以及站点客流分布与线路客流分布之间的相似性，引入站点分布相似系数。站点 i 与 j 之间的相似系数 p_{ij} 定义为

$$p_{ij} = 1 - d_{ij}$$

式中， $d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$ 为站点 i 与站点 j 之间的曼哈顿距离，其含义是站点 i 与站点 j 时段分布曲线绝对误差的和。 k 的取值为有车辆运营的时间段。

通过相似系数公式可以计算得到站点之间的相似矩阵，但研究更关心的是各站点分布与线路客流分布之间的相似性。所以取站点分布相似系数的默认值为该站点客流分布与线路客流分布之间的相似系数。

定义线路的站点分布相似系数平均值为该线路所有站点分布相似系数之和的平均值。表 1 为南京市区公交线路的站点分布相似系数平均值，数据来源于南京市 1997 年的公交随车客流调查。42 条市区线路中的 25 条，即 60% 线路的站点分布相似系数平均值在

0.7 以上。即这 60% 线路的站点客流时段分布曲线同线路客流分布曲线的平均绝对误差之和不超过 30%。

站点分布相似系数平均值（南京市 1997） 表 1

站点分布相似系数均值	线路条数	所占比例（%）
0~0.4	6	14.29
0.4~0.6	6	14.29
0.6~0.7	5	11.90
0.7~0.8	22	52.38
0.8~0.9	3	7.14

在一条线路的所有站点之中，客流量最大的站点往往是线路的首站和末站。表 2 是南京市随车客流调查得到的结果。61.9% 线路的首末站客流之和要占到所在线路客流的 20% 到 40%。南京市区客流中的 27.52% 发生在线路的首末站。

首末站客流比例（南京市 1997） 表 2

首末站客流占线路客流比例	线路条数	比例（%）
10%~20%	5	11.90
20%~30%	15	35.71
30%~40%	11	26.19
40%~50%	7	16.67
>50%	4	9.52

同其它站点相比，首末站的分布曲线同线路分布曲线之间的相似性更强。首末站分布相似系数描述的是两者之间的相似性，其计算同站点分布相似系数。

以首末站分布相似系数（ y ）为因变量，首末站客流之和占线路客流的比例（ x_1 ）以及线路的站点分布相似系数均值（ x_2 ）为自变量进行多元线性回归分析得到以下公式。

$$y = 0.492x_1 + 0.496x_2 + 0.333 \tag{1}$$

回归方程的相关系数 R 为 0.605。表 3 是对首末站分布回归方程的 F 检验。回归方程平方和为 $\sum (y_i - \bar{y})^2$ ，残差平方和为 $\sum (y_i - \hat{y}_i)^2$ ，其中 y_i 为实际观测值， $\hat{y}_i$ 为回归方程的预测值， $\bar{y}$ 为观测值的均值。 F 检验的值为 29.880，其显著水平为 0.000。因此可用该回归方程对首末站分布的相似系数进行评估。

首末站分布回归方程的 F 检验 表 3

	平方和	自由度	F 检验	显著水平
回归方程	0.254	2	29.880	0.000
残差	0.166	39		
总偏差	0.420	41		

表 4 是对首末站分布回归方程的 T 检验。各自变量的显著水平都为 0.000。因此，回归方程的所有变量都可用来估算首末站分布的相似性。

从以上分析，可得出以下结论：

- (1) 同一线路站点之间的客流时段分布相似；
- (2) 站点客流分布同线路的整体客流分布相似；
- (3) 首末站点客流分布与线路的客流分布很相似，首末站是线路的关键站点。

首末站客流一般要占到线路的 30% 左右。其客流分布同线路客流分布很相似。将首末站客流分布相似系数与首末站客流比例、站点分布相似系数均值进行回归分析的结果表明：如果首末站客流所占比例越大，那么首末站客流分布曲线同线路的曲线则越相似；而且站点之间的客流分布越相似，那么首末站同线路之间的相似性越强。

首末站分布回归方程的 T 检验 表 4

	值	标准差	T 检验	显著水平
常数项	0.254	0.065	5.148	0.000
首末站客流比例 (x_1)	0.166	0.082	6.023	0.000
站点分布相似系数均值 (x_2)	0.420	0.065	7.647	0.000

2 线路客流时段分布的特点^[1]

同站点分布相似系数相类似，首先也将线路客流时段分布曲线进行标准化处理。定义线路分布相似系数如下。

线路 i 与线路 j 之间的分布相似系数 q_{ij} 定义为

$$q_{ij} = 1 - d_{ij}$$

式中， $d_{ij} = \sum_{k=1}^p |y_{ik} - y_{jk}|$ 为线路 i 与线路 j 时段分布曲线之间的曼哈顿距离。 k 的取值为有车辆运营的时间段。表 5 计算的是各条线路与全市客流分布之间的相似系数。

线路分布相似系数（南京市 1997） 表 5

线路分布相似系数	线路条数	所占比例 (%)
< 0.5	5	11.90
0.5~0.6	2	4.76
0.6~0.7	5	11.90
0.7~0.8	2	4.76
0.8~0.9	20	47.62
0.9~1.0	8	19.05

将近 70% 线路的客流分布同全市客流分布的相似性在 0.8 以上。即 10 条线路中有 7 条线路的分布很相似，而且它们的相似性在 0.8 以上。南京市的平均线路分布相似系数 0.762。不同城市会有不同的结果，但其趋势应该是大部分线路的客流分布是相似

的，只有少数线路的客流时段分布存在显著不同。

影响线路分布之间相似性的因素有：

(1) 线路的运营区域

在同一条路上，走向大致相同的线路，其客流时段分布应该是趋于相似的。在我国大城市中，习惯在城市主干道上布设若干条走向相同的线路，若无较强外部因素作用，这些线路的客流分布应该是相近的。

(2) 线路的运营规模线路的运营规模即配车数，对线路客流的分布有一定影响。在南京的调查发现，运营规模相近的线路，其线路客流分布相似性较强，这一点对于大运量的线路更为明显。

(3) 线路的运营性质

运营性质对线路的客流时段分布有较大影响，例如旅游线路的峰值分布同常规线路就有较大区别。

3 首末站分层不等概率整群抽样法²

如果以乘客为基本抽样单元，难以设计一个以公交乘客为基础单位的抽样框。因为客流的发生和离去都是集中在站点，考虑以站点为单位进行整群抽样。

整群抽样 (Cluster Sampling) 是一种基本的抽样方法。在这里，将各个站点视为初级单元，每位乘客是次级单元。站点就是需要研究的群。

整群抽样群划分的原则是：尽可能地使群与群之间的差异小，而且群内的差异则越大越好。这样保证每个群都有较好的代表性。由前知站点分布之间的相似性强，符合群划分的原则。

在大多数情况下，群的大小 M_i 是不相等的。此时需对群进行不等概率抽样。通常采用的不等概率抽样方法为多项抽样方法。

多项抽样方法的定义如下：设 $Z_1, Z_2, \dots, Z_n$

是一组概率， $\sum_{i=1}^N Z_i = 1$ ，按照这组概率对总体中的 N 个单元进行放回抽样，每次抽到第 i 个单元的概率为 Z_i ，独立地进行这样的抽样 n 次。特别当每个单元具有一个说明其大小或规模的度量时，例如单位的职工人数等，则可取

$$Z_i = \frac{M_i}{M_0}$$

式中， $M_0 = \sum_{i=1}^N M_i$ 是总体中所有单元的“大小”之和。这时的多项抽样由于每个单元在每次抽样中的入样概率与单元大小成比例，故称为与大小成比例的概率抽样 (Sampling with probability proportional to size)，即为 PPS 抽样。

实施 PPS 抽样的主要方法是 Lahiri 方法: 令 $M^* = \max_{1 \leq i \leq N} \{M_i\}$, 每次抽取一个 $[1, N]$ 范围内的随机数 i 及 $[1, M^*]$ 范围内的随机数 m , 若 $M_i \geq m$, 则第 i 个单元入样; 否则, 重抽 (i, m) 。此时第 i 个单元的入样概率与 M_i 成正比, 从而 $Z_i = M_i / M_0$ 。

PPS 抽样的估计, 可直接应用 Hansen-Hurwitz 估计量。记 Y_{ij} 为总体中第 i 个群中第 j 个次级单元的观测值 ($i=1, 2, \dots, N; j=1, 2, \dots, M_i$), 其中 M_i 是群的大小。 y_{ij} 为样本中第 i 个群中第 j 个次级单元的观测值 ($i=1, 2, \dots, n; j=1, 2, \dots, m_i$), 其中 m_i 是群的大小。抽样估计中用到的一些统计量如下。

$$\begin{aligned} Y_i &= \sum_{j=1}^{M_i} Y_{ij} & y_i &= \sum_{j=1}^{m_i} y_{ij} \\ \bar{Y}_i &= \sum_{j=1}^{M_i} Y_{ij} / M_i & \bar{y}_i &= \sum_{j=1}^{m_i} y_{ij} / m_i \\ Y &= \sum_{i=1}^N Y_i / N & \bar{y} &= \sum_{i=1}^n y_i / n \\ \bar{Y} &= \sum_{i=1}^N \sum_{j=1}^{M_i} Y_{ij} / M_0 & \bar{\bar{y}} &= \sum_{i=1}^n \bar{y}_i / n \end{aligned}$$

则总体总和 Y 的 Hansen-Hurwitz 估计为

$$Y_{HH} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{z_i} = \frac{M_0}{n} \sum_{i=1}^n \frac{y_i}{m_i} = M_0 \bar{\bar{y}} \quad (2)$$

Y_{HH} 是 Y 的无偏估计, 它的方差为

$$V(Y_{HH}) = \frac{1}{n} \sum_{i=1}^N Z_i \left(\frac{Y_i}{Z_i} - Y \right)^2 = \frac{M_0}{n} \sum_{i=1}^N M_i (Y_i - \bar{Y})^2 \quad (3)$$

它的一个无偏估计为

$$\begin{aligned} v(Y_{HH}) &= \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{y_i}{z_i} - Y_{HH} \right)^2 \\ &= \frac{M_0^2}{n(n-1)} \sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2 \end{aligned} \quad (4)$$

在实际的抽样中, 并没有严格按照 PPS 的抽样方法进行抽样, 而是直接抽取了各线路的首末站点, 这是一种基于实用的考虑。具体的原因如下:

- (1) 首末站整群抽样样本比例较高, 方差可控;
- (2) 首末站整群抽样操作性好。

我国公交线路调度采用的是首末站调度管理的方法。每条线路的首末站都配有相应的调度管理人员。在没有电子化乘客计数系统的条件下, 在首末站实行整群抽样, 是唯一可以实现, 而且能长期进行的抽样调查方法。南京公交雅高巴士公司已经使用这一方法

进行调查。其具体操作方法是在每辆车离开首末站的时候, 在记录车辆发车时间的同时, 统计车上的乘客人数。该方法将抽样调查结合到日常的运营调度管理中, 没有增加任何调查费用, 而且调查的精度和长期性都可以得到保证。

首末站客流量比例较大, 而且首末站往往是线路客流最大的站点。按照 PPS 的抽样规则, 最有可能抽到的站点应该是首末站点。所以可近似认为直接抽取首末站点的整群抽样方法符合 PPS 抽样的要求。在此假设下, 可运用 PPS 的方差估计公式对结果进行估计。

表 6 以南京市公交 6 路为例, 给出了以首末站整群抽样方法估计线路客流时段分布的一个算例。6 路的首末站客流占到线路客流的 26.1%。乘客是抽样的次级单元, 抽样到的每位乘客都有一个上车时间同其相对应。如果线路的运营时间是从 5:00—22:00, 那么以小时为单位划分, 便有 18 个时间单位, 即 18 个时间段指标。按此指标设计, 抽样到的每位乘客都有 18 个指标观测值。例如调查到某乘客的上车时间为 8:41, 那么该样本 8:00—9:00 时间段的指标值取为 1, 其它时间段指标值取为 0。求线路的客流时段分布实际上是求时间段指标的总和估计。

为计算各时间段指标的估计量在 95% 的置信区间下相对误差约为多少, 引入了双倍变异系数, 它相当于两倍的标准差同估计值的比值。同时也利用实际调查数据计算了实际的变异系数。18 个指标中有将近 50% (8 个) 的实际变异落在双倍变异系数范围之内。为综合比较两者之间的差别, 按照各个时间段客流量大小, 对两者进行加权平均, 得到双倍变异系数的均值为 23.41%, 实际变异系数的均值为 18.67%。双倍变异系数的均值表明, 在 95% 的置信度下, 估计值同实际值之间的平均相对误差不超过 23.41%。而由全样调查数据计算出来的结果为 18.67%, 这表明实际的变异落在估计的变异范围之内。

线路进行了相应的计算。结果表明南京市近 70% 的线路的平均双倍变异系数不超过 30%, 它们的实际变异绝大多数落在估计变异范围之内。从满足调度优化工作的需要出发, 相对偏差不大于 30% 的时段分布估计是可用的。即在南京市, 70% 的线路可采用首末站整群抽样的方法估计线路客流分布。但还有 30% 的线路变异系数较大, 仅依靠首末站整群抽样是不够的。

首末站整群抽样算例 表 6

时段	线路客流 实际值 Y	线路客流 估计值 Y_{HH}	标准差 $\sqrt{v(Y_{HH})}$	双倍变异系数 $1.96 \frac{\sqrt{v(Y_{HH})}}{Y_{HH}}$	实际变异系数 $\frac{ Y - Y_{HH} }{Y}$
5:00—6:00	123	139	15.26	21.45%	13.36%
6:00—7:00	837	1113	119.33	21.02%	32.94%
7:00—8:00	1421	1661	52.91	6.25%	16.86%
8:00—9:00	982	1167	161.55	27.13%	18.84%
9:00—10:00	1023	1264	126.29	19.58%	23.60%
10:00—11:00	838	576	54.16	18.44%	31.31%
11:00—12:00	814	680	21.77	6.28%	16.48%
12:00—13:00	640	384	24.37	12.42%	39.93%
13:00—14:00	895	996	313.22	61.63%	11.30%
14:00—15:00	1102	1251	90.41	14.17%	13.48%
15:00—16:00	819	665	31.43	9.27%	18.84%
16:00—17:00	1080	961	44.33	9.04%	10.98%
17:00—18:00	998	939	303.17	63.31%	5.96%
18:00—19:00	887	872	127.48	28.64%	1.64%
19:00—20:00	436	502	67.23	26.26%	15.10%
20:00—21:00	401	274	36.43	26.06%	31.67%
21:00—22:00	260	163	13.89	16.72%	37.35%
22:00—23:00	90	39	26.94	134.16%	56.27%
总和	13646	13646	—	—	—

深入地调查发现，调度管理人员实际需要的并不是每条线路的客流时段分布变化曲线，而是按照类型划分的线路客流变化曲线。调度管理人员在制定计划时往往将客流变化规律类似、运营地域相邻、规模相近的线路合并在一起，根据这一类型线路的变化规律，制定相同的运营计划。所以线路客流时段分布估计更为准确的定义应该是：利用抽样调查手段，估计预测线路类型的客流时段分布。

调度管理人员制定计划时的线路归并过程实际上是一个总体的分层过程。考虑综合运用分层抽样与整群抽样方法，解决线路客流时段分布估计预测问题。

分层抽样将大小为 N 的总体分成 L 个不相重迭的子总体，大小分别为 $N_1, N_2, \dots, N_L$ (N_h 皆已知)，每个子总体称为层，从每层中独立进行抽样。

在首末站整群抽样研究的基础上，为解决所有线路客流时段分布的预测问题，特提出了首末站分层不等概率整群抽样法，它的具体步骤是：

- (1) 将全体线路划分为若干个大层；
依据线路客流变化规律、运营地域、运营规模等特点，采用定量计算与定性分析相结合的方法将全体线路划分为若干大层。总体分层有很多成熟的方法。
- (2) 在每个大层之中，将每条线路看作一个小层，在小层中进行不等概率整群抽样。
在每个小层中以 PPS 方法进行整群抽样。在实际的抽样框设计中，抽取的是各条线路的首末站。基于

同样的原因，可以认为这种抽样设计近似等同于 PPS 抽样，并可利用 PPS 方差估计公式来估计小层的方差。

- 采用分层不等概率整群抽样具有以下优点：
- (1) 可对层的参数进行估计并且考虑它们的精度
通过该方法可以估计每个大层（线路类型）的参数及其精度。
- (2) 使样本更具有代表性
因为分层抽样中每层都一定有单元被抽取，从而使样本的均匀性更好。

(3) 分层抽样可以提高估计量的精度
这是采用分层抽样方法的重要原因，分层扩大了样本数据，而且层间变差不进入到最后估计量的抽样误差中，这样估计的精度就可以有较大幅度的提高。

(4) 实施中的组织管理以及数据汇总都比较方便
小层的划分同车队编制是一致，而且小层内的整群抽样同首末站调度管理相结合，在组织管理上抽样都可以得到保障。抽样的数据可按照日常运营表格的统计途径进行汇总。

需要估计的量是大层各个时间段的客流量。对于某个确定的大层，设计如下的总和估计量

$$Y_j = \sum_{i=1}^n \frac{m_i}{m} Y_{HHj}$$

式中， Y_j 为第 j 个时间段的客流估计量，该大层总共有 n 个小层（ n 条线路）， m_i 为该大层中各小层的大

分层不等概率整群抽样算例

表 7

时段	标准差 $\sqrt{v(Y_j)}$	双倍变异 $\frac{1.96 \sqrt{v(Y_j)}}{Y_j}$	实际变异 $\frac{ Y-Y }{Y}$	时段	标准差 $\sqrt{v(Y_j)}$	双倍变异 $\frac{1.96 \sqrt{v(Y_j)}}{Y_j}$	实际变异 $\frac{ Y-Y }{Y}$
5:00—6:00	150.44	31.54%	25.33%	14:00—15:00	329.59	14.45%	19.43%
6:00—7:00	123.59	6.29%	38.43%	15:00—16:00	219.68	13.20%	11.27%
7:00—8:00	430.78	11.98%	26.88%	16:00—17:00	202.88	8.96%	7.04%
8:00—9:00	1082.26	35.29%	29.66%	17:00—18:00	541.12	25.48%	5.28%
9:00—10:00	207.25	10.04%	18.20%	18:00—19:00	413.82	27.83%	1.96%
10:00—11:00	399.57	25.54%	5.71%	19:00—20:00	356.40	27.62%	25.63%
11:00—12:00	229.92	18.03%	7.59%	20:00—21:00	92.31	16.44%	21.22%
12:00—13:00	138.86	12.05%	9.37%	21:00—22:00	38.75	15.90%	23.30%
13:00—14:00	406.17	20.25%	21.53%	22:00—23:00	29.94	30.92%	35.88%

小（线路客流量）， $\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$ ， Y_{HHi} 为第 i 小层的 Hansen—Hurwitz 估计量， Y_{HHi} 的计算见公式（2）。 Y_j 的方差估计为

$$v(Y_j) = \frac{m_i^2}{m^2} \sum_{i=1}^n v(Y_{HHi})$$

式中， $v(Y_{HHj})$ 的计算见公式（4）。

表 7、图 1 给出了以分层不等概率整群抽样方法估计线路客流时段分布的一个算例。算例的数据来自于南京市 1997 年的随车客流调查。该线路高层包括 3 条规模和客流都很相似的线路，分别是 2 路、6 路、17 路。在各条线路，即各小层中，分别抽取其首站和尾站。总的样本抽样比例达到了 30.50%。18 个指标中有将近 56%（10 个）的实际变异落在双倍变异系数范围之内。以各时段的客流量大小为权，计算得到双倍变异系数的均值为 18.48%，实际变异系数的均值为 17.00%。按照分层抽样，不论估计的变异系数还是实际的变异系数均比单条线路的首末站整群抽样的精度有所提高。本文的两个算例选取的线路样本量充足，而且代表性好，两者之间差别不显著，但对于较差的首末站整群抽样线路，分层不等概率整群抽样的效果是显著优越的。

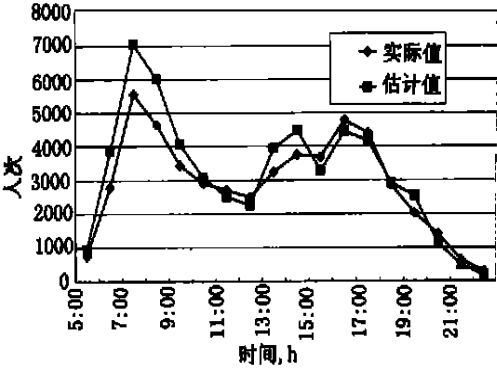


图 1 算例的估计值和实际值

4 结论

通过首末站分层不等概率整群抽样，抓住了线路的关键站点，可以准确地获得不同线路类型的时段分布曲线，是解决公交线路客流时段分布估计问题的一个新思路。在获得现状线路客流时段分布数据的基础上，可以结合模糊神经网络等预测方法，对未来的变化趋势进行预测。

参考文献:

[1] 南京市公共客运交通规划课题组. 南京市近期公共客运交通规划（研究报告）. 1998

[2] 冯士雍, 施锡铨. 抽样调查——理论、方法与实践. 上海科学技术出版社, 1996.