

文章编号: 1005-0523(2010)06-0113-06

高速公路交通事故类平均系统聚类法实证分析

查伟雄, 李 敏, 付印平

(华东交通大学 交通运输与经济研究所, 江西 南昌 330013)

摘要: 道路的交通事故诱发的原因多种多样, 而同一段路上的交通事故的表现形式也很多。在选取事故形态表现形式和伤亡状况指标的基础上, 采用系统聚类法中的类平均法对某段高速公路的事故状况进行分析, 把具有某种共性的事故表现形式划归为一类, 并对结果进行判别分析来检验聚类的效果, 结果表明类别的划分突出特性, 反应交通事故的主要事故形态, 很大程度上反应事故原因, 有利于相关部门采取相应的措施。

关 键 词: 道路交通事故; 类平均法; 聚类分析; 事故形态

中图分类号: U491.31

文献标识码: A

道路交通事故统计分析是一项基础性工作, 道路交通事故统计指标是道路系统采集交通事故信息的依据^[1]。目前, 道路交通安全研究已趋于从交通事故成因、机理入手, 寻找道路交通事故发生时的人员、车辆、道路环境等方面的诱导因素, 以及交通事故与影响因素间的相互关系, 所以道路交通事故的调查和统计分析是交通事故研究的一个很重要的方面^[2]。将多元统计的方法用于交通中也是科学而必要的, 而聚类的方法在交通的应用日益广泛, 吴彪^[3]用 Ward 聚类法对道路交通事故进行多元统计分析, 黎莹^[4]用聚类分析对北京各区之间公共汽车交通状况进行分析, 但是类平均法用于交通事故形态分析相对较少, 且没有实质性的建议, 本文在道路交通事故统计指标的基础上, 选用系统聚类分析中的类平均法对沪昆高速公路(昌樟段)交通事故进行定量分析, 着重对结果进行系统性的分析并提出相关建议。

1 道路交通事故

高速公路具有通行能力高、舒适、运输成本低、经济效益好等特点, 现已成为我国综合交通运输系统的重要组成部分。然而, 高速公路运输在带来高效、快捷、方便的同时, 也带来了诸如交通事故增加等负面影响。近十几年来, 我国高速公路事故起数、死亡人数、受伤人数连续保持高位增长, 概括地说, 无论是从纵向的时间发展来看, 还是横向的和其他等级道路相比, 高速公路的安全形势都是不容乐观的。而且可以预见的是高速公路的安全形势会随着经济和社会的发展, 以及汽车保有量和高速公路通车里程的增加, 会变得越来越严峻^[5]。交通安全是高速公路发展的生命, 高速公路的交通安全管理不仅关系到高速公路的正常运行也影响到人们的生命、财产、生产、生活和工作。由于高速公路具有车流量大、行车速度高的特点, 目前在高速公路管理过程中还存在着一定的缺陷, 没有针对性。

2 聚类的适应性分析

反映同一事物特点的变量有很多, 我们往往根据所研究的问题选择部分变量对事物的某一方面进行研究。国内外通常采用交通事故次数、死亡人数、受伤人数以及直接经济损失这 4 项指标用于统计分析道路交通事故状况, 这 4 项指标既是认识交通事故的起点, 又是构造其他交通事故统计指标的基础, 基本涵盖了道路交通事故所造成各种损害的主要方面, 表征的交通事故情况直观、明确, 而且这 4 项指标相对比

收稿日期: 2010-09-01

作者简介: 查伟雄(1963—), 男, 教授, 研究方向为交通运输经济。

较容易获得,本文也用这 4 项指标进行分析。道路交通事故的形态主要有:正面相撞、侧面相撞、尾随相撞、对向刮擦、同向刮擦、刮撞行人、撞静止车辆、撞固定物、翻车、碾压、坠车和其他。

分类是人类认识世界的基本方法,聚类分析是对样品或变量进行分类的一种多元统计方法,目的在于将相似的事物归类,聚类是一种通过观察的学习方法,将某个对象集划分为若干组,这样就可以用上面选择的指标(事故次数,死亡人数,受伤人数,直接经济损失)通过聚类的方法,可以找出上述 11 种事故形态关于事故损害程度的分类,再来就分类结果提出政策建议。

3 昌樟高速数据聚类分析

聚类分析就是一种分类方法,它将一批样品、变量(或指标),按照它们在性质上相似、疏远程度进行科学分类。对于同一样本,采用不同的系统聚类方法的结果大体一致,而类平均法是聚类效果较好、应用比较广泛的一种聚类方法,因此本文采用类平均法进行聚类。具体步骤如下。

3.1 统计数据的特征矩阵^[6]

聚类分析所研究的内容包括两个方面,一是对样品进行分类,设 n 个样品,每个样品均用 p 个指标的观测向量 $\mathbf{x}_i (i=1,2,\cdots,n)$ 来表征,要根据 $\mathbf{x}_i$ 间某种相似性度量,将这 n 个样品进行分类。另一方面是对变量进行分类,即对所考察的 p 个指标 $X=(X_1,X_2,\cdots,X_p)'$,根据 n 个观测值 $\mathbf{x}_i=\begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{pi} \end{bmatrix}' (i=1,2,\cdots,n)$ 及某些相似性原则将这 p 个变量 $X_1,X_2,\cdots,X_p$ 进行分类。根据上述可以得出原始的观测数据:

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

(1)

2003 到 2008 年整合的原始数据如表 1。

表 1 昌樟事故形态及其伤亡状况分布统计

序号	事故形态	事故次数/起	死亡人数/人	受伤人数/人	直接经济损失/元
1	侧面相撞	51	14	54	861 120
2	正面相撞	15	6	19	808 010
3	尾随相撞	225	59	291	9 614 128
4	对向刮擦	1	0	2	500
5	同向刮擦	15	4	10	28 004
6	刮撞行人	11	8	4	16 300
7	撞静止车辆	27	22	51	964 000
8	撞固定物	152	26	177	7 982 476
9	翻车	97	23	126	2 201 300
10	碾压	1	1	0	4 000
11	坠车	7	14	78	454 896
12	其他	7	3	4	11 250
13	全线合计	609	180	816	22 945 984

注:来源于 2003—2008 年沪昆高速公路(昌樟段)全线交通安全相关数据的采集

3.2 数据的标准化

数据的标准化是对变量的数值和量纲进行的一种数据变换。首先对每个变量进行中心化变换,即 $x'_{ij}=x_{ij}-\bar{x}_j (i=1,2,3,\cdots,n;j=1,2,3,\cdots,p)$,然后用该变量的中心化的结果进行标准化。即有

$$x^*_{ij}=\frac{x'_{ij}}{S_j}=\frac{x_{ij}-\bar{x}_j}{S_j}$$

(2)

$$S_j = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2; \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$
$$(i=1, 2, 3, \dots, n; j=1, 2, 3, \dots, p)$$

(3)

其中: x_{ij}^* 为标准化数据, 经过标准化变换处理后, 每个变量即数据矩阵中每列数据的平均值为 0, 方差为 1, 且也不再具有量纲, 同样也便于不同变量之间的比较。

数据化标准处理^[7]结果见表 2。

表 2 昌樟事故形态及其伤亡状况标准化矩阵

序号	事故形态	事故次数	死亡人数	受伤人数	直接经济损失
1	侧面相撞	0.004	−0.061	−0.156	−0.319
2	正面相撞	−0.500	−0.546	−0.546	−0.335
3	尾随相撞	2.440	2.670	2.490	2.340
4	对向刮擦	−0.696	−0.910	−0.736	−0.580
5	同向刮擦	−0.500	−0.670	−0.647	−0.571
6	刮撞行人	−0.556	−0.425	−0.714	−0.575
7	撞静止车辆	−0.332	0.425	−0.190	−0.287
8	撞固定物	1.420	−0.667	1.220	1.840
9	翻车	−0.647	−0.485	−0.647	0.088
10	碾压	−0.696	−0.849	−0.758	−0.579
11	坠车	−0.612	−0.061	0.110	−0.442
12	其他	−0.612	−0.728	−0.714	−0.576

3.3 建立距离矩阵

常用的距离算法很多, 本文采用最常用的欧式距离来进行计算各样本之间的距离, 得距离矩阵 D 为

$$D = \left(d_{ij} \right)_{n \times n}, d_{ij} = \left[\sum_{k=1}^p \left(x_{ik} - x_{jk} \right)^2 \right]^{\frac{1}{2}}, i, j = 1, 2, \dots, n$$

(4)

式中: d_{ij} 为样品 x_i 和 x_j 之间的欧氏距离。

利用 SPSS 软件得出的欧式距离矩阵见表 3。

表 3 昌樟事故形态及其伤亡状况的距离矩阵(Proximity Matrix)

个案	1:侧面相撞	2:正面相撞	3:尾随相撞	4:对向刮擦	5:同向刮擦	6:刮撞行人	7:撞静止车	8:撞固定物	9:翻车	10:碾压	11:坠车	12:其他
1:侧面相撞	0.000	0.642	27.403	1.615	0.926	0.822	0.350	9.068	1.522	1.541	0.465	1.201
2:正面相撞	0.642	0.000	35.293	0.267	0.081	0.103	1.100	12.979	3.981	0.235	0.692	0.132
3:尾随相撞	27.403	35.293	0.000	41.508	38.026	37.241	26.747	6.912	16.412	41.216	30.103	39.557
4:对向刮擦	1.615	0.267	41.508	0.000	0.105	0.256	2.298	16.613	6.107	0.004	1.466	0.041
5:同向刮擦	0.926	0.081	38.026	0.105	0.000	0.067	1.510	14.737	4.751	0.084	0.972	0.021
6:刮撞行人	0.822	0.103	37.241	0.256	0.067	0.000	1.129	14.636	4.564	0.202	0.834	0.095
7:撞静止车	0.350	1.100	26.747	2.298	1.510	1.129	0.000	9.617	1.802	2.164	0.428	1.765
8:撞固定物	9.068	12.979	6.912	16.613	14.737	14.636	9.617	0.000	4.021	16.507	11.071	15.622
9:翻车	1.522	3.981	16.412	6.107	4.751	4.564	1.802	4.021	0.000	6.002	2.449	5.348
10:碾压	1.541	0.235	41.216	0.004	0.084	0.202	2.164	16.507	6.002	0.000	1.404	0.024
11:坠车	0.465	0.692	30.103	1.466	0.972	0.834	0.428	11.071	2.449	1.404	0.000	1.144
12:其他	1.201	0.132	39.557	0.041	0.021	0.095	1.765	15.622	5.348	0.024	1.144	0.000

3.4 采用类平均距离法进行聚类

类平均法定义两类之间的距离平方为这两类元素两两之间距离平方的平均 D_{pq}^2 , 即

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{x_i \in G_p} \sum_{x_j \in G_q} d_{ij}^2$$

(5)

设 G_p 与 G_q 合并成新类 G_r , 则任一类 G_k 与 G_r 之间的距离为

$$\begin{aligned} D_{kr}^2 &= \frac{1}{n_k n_r} \sum_{x_i \in G_k} \sum_{x_j \in G_r} d_{ij}^2 = \frac{1}{n_k n_r} \left[\sum_{x_i \in G_k} \left(\sum_{x_j \in G_p} d_{ij}^2 + \sum_{x_j \in G_q} d_{ij}^2 \right) \right] = \\ &= \frac{1}{n_k n_r} \left(\sum_{x_i \in G_k} \sum_{x_j \in G_p} d_{ij}^2 + \sum_{x_i \in G_k} \sum_{x_j \in G_q} d_{ij}^2 \right) = \frac{1}{n_k n_r} \left(n_k n_p D^2(G_k, G_p) + n_k n_q D^2(G_k, G_q) \right) = \\ &= \frac{n_p}{n_r} D^2(G_k, G_p) + \frac{n_q}{n_r} D^2(G_k, G_q) = \frac{n_p}{n_r} D_{kp}^2 + \frac{n_q}{n_r} D_{kq}^2 \end{aligned}$$

(6)

即在已经建立好的欧式距离矩阵中找出最小的元素, 将其合并为新类, 然后按照上述类平均距离的定义公式进行计算, 得出新的距离矩阵, 再重复上述过程知道所有的元素都聚为一类为止。类平均法系统聚类的谱系图见图 1。

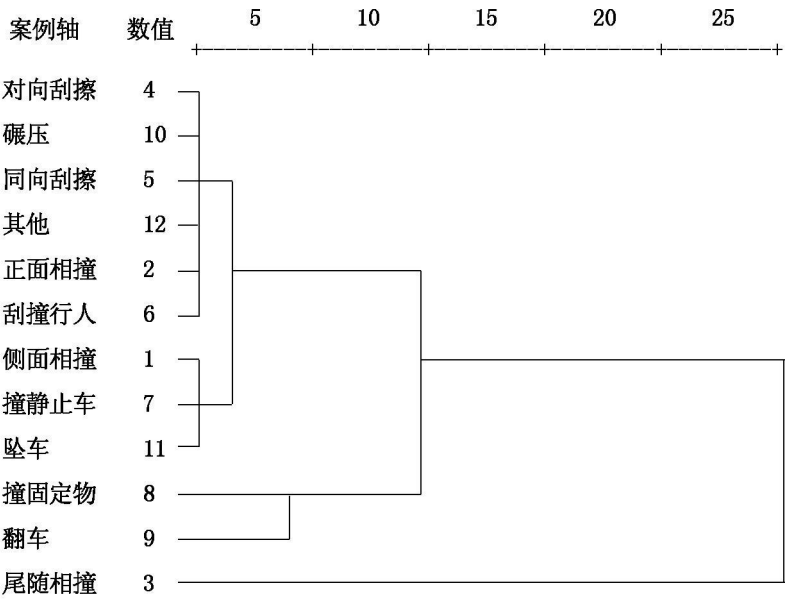


图 1 昌樟事故形态及其伤亡状况的系统聚类谱系图

从图 1 聚类分析中可以看到距离为 5 时事故形态很明显的被分为 3 种: 第 1 类是尾随相撞; 第 2 类是撞固定物、翻车; 第 3 类是对向刮擦、碾压、其他、同向刮擦、刮撞行人、正面相撞、侧面相撞、撞静止车、坠车计数。

结合数据分析可知第 1 类事故形态即尾随相撞是最多的, 其发生次数多、伤亡人数多并且带来的经济损失非常严重; 第 2 类即撞固定物和翻车的事故形态也很多, 它们导致的伤亡人数也较多, 所带来的经济损失也比较大; 第 3 类事故形态相对较少, 从选取的 4 项指标来看造成的各项损失都较小。

4 判别分析验证

判别分析是根据所研究个体的某些指标的观测值来推断该个体所属类型的一种统计方法, 根据已知观测对象的分类和若干表明观测对象特征的变量值, 建立判别函数和判别准则, 并使其错判率最小, 对于一个未知分类的样本, 将所测指标代入判别方程, 从而判断它来自哪个总体。当然, 这种准则在某种意义上是最优的, 如错判概率最小或错判损失最小等。

此实证分析中采取上述聚类结果作为原始类别, 采用贝叶斯判别法进行判别可得出判别类别以及其被判别为各项的概率。这样就对前面的聚类分析是否准确可信做了一个检验, 见表 4。

表 4 昌樟事故形态及其伤亡状况的判别分析后验概率

事故形态	原始类别	判别类别	判别为 1 的概率	判别为 2 的概率	判别为 3 的概率
侧面相撞	3	3	0	0.001 584 578 2	0.998 415 421 8
正面相撞	3	3	0	0.000 000 000 2	0.999 999 999 8
尾随相撞	1	1	0.999 999 884 2	0.000 000 115 8	0
对向刮擦	3	3	0	0	1
同向刮擦	3	3	0	0.000 000 002 3	0.999 999 997 7
刮撞行人	3	3	0	0	1
撞静止车辆	3	3	0	0	1
撞固定物	2	2	0.000 005 560 0	0.999 994 440 0	0
翻车	2	2	0.000 000 002 4	0.999 999 884 6	0.000 000 113 0
碾压	3	3	0	0	1
坠车	3	3	0	0.000 000 007 1	0.999 999 992 9
其他	3	3	0	0.000 000 000 1	0.999 999 999 9

由表 4 可以看出通过贝叶斯判别法进行判别可以得出原始聚类的类别和判别类别完全一致, 并且每项为判别类别的概率均在 99% 以上, 即每项的错判概率均在 1% 以下, 由此可见通过聚类得到的类别相当准确, 各种事故形态的错判概率几乎都为 0, 通过判别分析验证采取此聚类方法是准确可靠的, 用此结果来分析问题也是可信的。

5 结论与建议

高速公路由于没有行人、非机动车的干扰, 所以同普通道路相比, 其事故种类有很大不同。结合数据发现沪昆高速公路(昌樟段)事故形态中尾随相撞、撞固定物以及翻车事故起数、伤亡人数和直接经济损失都是很严重的, 这同我国高速公路交通事故特性是一致的^[8]。尾随相撞与超速行驶、疲劳驾驶密切相关, 而翻车和撞固定物与酒后驾驶密切相关。针对这些可能出现的原因, 相关部门应该建立更加高效的交通安全管理体制, 与高速公路运营部门联手合作, 要重点整治超速行驶, 疲劳驾驶和酒后驾驶的行为, 以进一步提高高速公路的安全水平, 具体建议如下:

- (1) 严禁超速, 途中设立用以测速的仪器, 对高速公路上行驶的车辆严格限速, 有的高速公路, 还可以根据当时天气和道路通行情况来设立可变限速交通标志, 增加道路的安全性。在事故多发点周边需要设定标志标线, 增加用以确定车距的标线, 方便驾驶员确定车距, 减少追尾事件的发生。
- (2) 禁止疲劳驾驶, 高速公路开车 4 h 必须把车开到服务区活动一下身体, 休息 20 min, 洗把脸, 喝点茶水, 以利于提神醒脑。交通管理部门应加大宣传力度, 同时可以考虑在车上使用一些先进的技术来检测驾驶员的连续行车时间, 超过一定的时间可以给予处罚。
- (3) 禁止酒后驾车, 可在高速公路入口处增设驾驶员是否酒后驾驶的检测, 同时通过贴宣传标语、播放视频等方式加强对驾驶人员的交通安全教育, 让驾驶人员意识到酒后驾驶的危害, 懂得珍爱生命, 以期减少交通事故。
- (4) 与管理部门携手建立更加高效的交通安全管理体制, 进一步规范道路标志, 使其更加人性化, 给予驾驶员更加良好的行车环境。对于不服从管理的人员在必要时可以采取法律手段进行强制管理。

6 结束语

多元统计中的系统聚类法在道路交通事故分析中起到很大作用, 本文通过对交通事故的形态进行类平均聚类分析, 发现高速公路事故形态中尾随相撞、撞固定物以及翻车事故的起数、伤亡人数和直接经济损失都是很严重的, 并针对这些形态可能出现的原因对相关部门做出实质性的建议。但聚类法不仅仅是可以分析主要的事故形态, 还可以对各大城市的交通状况进行聚类分析, 对复杂和宏观的数据进行聚类分

析,这种有理有据的定量分析,为组织管理者了解交通状况、制定政策和法规提供可靠而有力的证据,从而提高城市交通水平,为城市发展做出贡献。但是由于没有对记录的每项事故具体数据进行分析,上述的建议只是停留在常规原因的讨论,这方面有待深入研究。

参考文献:

- [1] 费雪良,胡江碧,刘小明.我国道路交通事故统计指标体系问题探讨[J].公路,2006(7):125-128.
- [2] 赵海存.道路交通事故数据统计分析系统研究[D].西安:长安大学,2005.
- [3] 吴彪,许洪国,张文会.基于 Ward 聚类法的道路交通事故多元统计分析[J].黑龙江工程学院学报:自然科学版,2009,23(3):4-6.
- [4] 黎莹.北京各区之间公共汽车交通状况聚类分析[J].商业文化,2007(4):225-226.
- [5] 钟连德,孙小端,陈永胜,等.中国高速公路事故特点及分布规律研究[J].交通论坛,2007,7(4):16-21.
- [6] 何晓群.多元统计分析[M].2版.北京:中国人民大学出版社,2004.
- [7] 卢纹岱,朱一力,沙捷,等.SPSS for windows 统计分析[M].2版.北京:电子工业出版社,2002.
- [8] 彭萌,王新泉,吴超,等.高速公路事故特征实例分析研究[J].工业安全与环保,2009,30(10):41-43.

Empirical Analysis of Highway Traffic Accident in Average Linkage Clustering Method

Zha Weixiong, Li Min, Fu Yingping

(Institute of Transportation and Economics, East China Jiaotong University, Nanchang 330013, China)

Abstract: Road traffic accidents are induced by different reasons. Forms of the road traffic accident at the same section are various. Based on choosing different accident forms and casualties indicators, clustering method is adopted to analyze accident status of a section. Some similar accident forms can be classified as a category. Results are analyzed to examine the effect of cluster. The results show that the category has its characteristics, and that forms of major traffic accidents reflect the reasons of the accident, which can help the related departments take corresponding measures.

Key words: road traffic accident; average linkage method; cluster analysis; accident forms

(责任编辑 刘棉玲)