

# 空间迁移分析中类型数据的logit 线性模型

F. 布劳威尔 P. 尼贾坎普

现代空间分析的一个重要进展,是已将重点放在非聚合行为抉择模型上,在交通运输研究、家庭市场研究和区位分析中已设计了多种离散抉择模型。这些领域中最流行的方法之一是众所周知的logit模型,在确定类型反应变量与连续的或类型的说明性变量之间的因果关系时,logit线性回归分析是一个广泛应用的判断步骤,在许多学科中,都可发现它的经验应用,如地理学、生物学、医学、社会学等,在迁移与运输分析中也可见到。

然而,在许多文章中,都忽视了logit模型的数量经济统计学特性。本文在对异方差性和条件均值给以特别注意的条件下,阐述了扰动项之不同类型分布函数的结果。在假定逻辑斯谛和极值分布函数时,就可得到logit和补充重对数数据变换。在这个假设与经验条件一致的情况下,logit数据变换就能解决异方差性和抉择概率(限于0至1的区间内)预测等问题。

乘法模型,在许多说明性变量以倍增法相互关联的学科中,都广泛应用以幂函数为基础的回归模型,当观察值服从多项分布时,

乘法模型是类型数据的对数线性模型的一部分,地理学、生物学、经济学、医学及社会学中可发现各种应用。

有类型数据的logit线性模型,在前一部分,比较了一些服从(对数)正态、逻辑斯谛或韦巴尔概率密度函数的不同类型的模型。现在,在二歧或多生分类方法中,要特别注意具有类型观察值的logit模型的经验应用(略)。

logit模型的应用,现在将上面论述的模型运用于一个荷兰调查样本,是关于2600户居民的生活质量与居住选择的(也可参见海塔和戈迪金的文章),通过这些模型的估计与检验,以更深刻地了解今后5年内居民迁居的可能性。表(1A)到(1E)给出了各种变量 $x$ 、 $y$ 、 $z$ 的观察值之列联表:

$x$  = 今后五年内从所在生活区移出的意愿: 1 = 大概不会; -1 = 预计会。

$y$  = 目前生活区与选择区的一致性: 1 = 一致; -1 = 不同。

$z_1$  = 户主年龄: 1 = 30岁以下; 0 = 30岁到60岁; -1 = 60岁以上。

$z_2$  = 生活于非城市区域的选择: 1 = 选

表1A 变量 $x$ 、 $y$ 和 $z_1$ 的列联表

	$z_1 = 1$		$z_1 = 0$		$z_1 = -1$		合计
	$y = 1$	$y = -1$	$y = 1$	$y = -1$	$y = 1$	$y = -1$	
$x = 1$	210	81	932	307	515	82	2130
$x = -1$	37	108	89	170	19	36	459
合 计	247	192	1021	477	534	118	2589

表 1 B 变量 $x$ 、 $y$ 和 $z_2$ 的列联表

	$z_2 = 1$		$z_2 = 0$		$z_2 = -1$		合计
	$y = 1$	$y = -1$	$y = 1$	$y = -1$	$y = 1$	$y = -1$	
$x = 1$	604	233	499	140	354	67	1897
$x = -1$	58	159	45	95	20	38	415
合 计	662	392	544	235	374	105	2312

表 1 C 变量 $x$ 、 $y$ 和 $z_3$ 的列联表

	$z_3 = 1$		$z_3 = 0$		$z_3 = -1$		合计
	$y = 1$	$y = -1$	$y = 1$	$y = -1$	$y = 1$	$y = -1$	
$x = 1$	867	259	609	147	170	65	2117
$x = -1$	79	165	44	101	22	45	456
合 计	946	424	653	248	192	110	2573

表 1 D 变量 $x$ 、 $y$ 和 $z_4$ 的列联表

	$z_4 = 1$		$z_4 = 0$		$z_4 = -1$		合计
	$y = 1$	$y = -1$	$y = 1$	$y = -1$	$y = 1$	$y = -1$	
$x = 1$	199	102	331	128	1014	213	1987
$x = -1$	21	75	33	102	74	106	411
总 计	220	177	364	230	1088	319	2398

表 1 E 变量 $x$ 、 $y$ 和 $z_5$ 的列联表

	$z_5 = 1$		$z_5 = 0$		$z_5 = -1$		合计
	$y = 1$	$y = -1$	$y = 1$	$y = -1$	$y = 1$	$y = -1$	
$x = 1$	1057	325	477	122	120	23	2124
$x = -1$	94	219	37	73	9	19	454
合 计	1154	544	514	195	129	42	2578

表 2 参数估计与模型有效性

变 量	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$x^2$
$x, y, z_1$	-2.335 ° (0.160) 214.099	0.816 ° (0.160) 26.146	0.168 (0.160) 1.114	-0.059 (0.361) 0.027	0.512 ° (0.160) 10.314	0.157 (0.361) 0.189	0.021 (0.361) 0.003	0.093 (0.361) 0.066	86.1982
$x, y, z_2$	-2.849 ° (0.239) 141.689	0.782 ° (0.239) 10.678	0.160 (0.239) 0.447	0.495 (0.292) 2.869	0.565 ° (0.239) 5.576	-0.028 (0.292) 0.009	-0.129 (0.292) 0.194	-0.017 (0.292) 0.004	0.4988
$x, y, z_3$	-2.874 ° (0.241) 142.556	0.796 ° (0.241) 10.943	0.185 (0.241) 0.593	0.640 ° (0.275) 5.414	0.591 ° (0.241) 6.026	0.027 (0.275) 0.010	0.005 (0.275) 0.000	-0.010 (0.275) 0.001	7.1611
$x, y, z_4$	-2.920 ° (0.256) 129.861	0.759 ° (0.256) 8.774	0.064 (0.256) 0.062	-0.626 ° (0.284) 4.849	0.546 ° (0.256) 4.547	-0.194 (0.284) 0.466	-0.334 (0.284) 1.378	-0.063 (0.284) 0.049	5.4090
$x, y, z_5$	-3.101 ° (0.363) 73.102	0.814 ° (0.363) 5.039	0.200 (0.363) 0.305	1.124 ° (0.388) 8.412	0.557 (0.363) 2.357	0.017 (0.388) 0.002	-0.022 (0.388) 0.003	0.053 (0.388) 0.019	0.7894

显著水平为0.95

表 3 参数估计与简化模型的有效性

变 量	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$x^2$	自由度	临界值
$x, y, z_1$	-2.443 (0.132) 343.357	0.921 (0.118) 60.533	— — —	— — —	0.633 (0.083) 57.979	91.153	9	16.919
$x, y, z_2$	-2.596 (0.142) 332.062	0.769 (0.125) 37.867	— — —	— — —	0.576 (0.092) 39.252	21.328	9	16.919
$x, y, z_3$	-2.983 (0.150) 397.413	0.935 (0.122) 58.873	0.659 (0.094) 49.489	— — —	0.742 (0.088) 71.405	8.570	8	15.507
$x, y, z_4$	-3.138 (0.159) 389.492	0.918 (0.140) 43.288	-1.046 (0.096) 117.898	— — —	0.797 (0.096) 69.628	19.611	8	15.507
$x, y, z_5$	-2.635 (0.140) 352.884	0.992 (0.107) 85.850	1.093 (0.111) 97.417	— — —	— — —	93.306	9	16.919

择非城市区: 0 = 不太愿选择非城市区。

$z_3$  = 买房与租房的选择: 1 = 愿买: 0 = 愿租: -1 = 无选择。

$z_4$  = 生活于市区的选择: 1 = 选择城市区: 0 = 不太愿选择城市区: -1 = 绝不选择城市区。

$z_5$  = 对生活区有更多商业服务设施的选择: 1 = 是: 0 = 无所谓: -1 = 不。

表2列出了与表1A—1E有关的5个模型的饱和形式的参数估计, 括号内为标准误差。

表2给出了参数估计的统计显著性, 在1个自由度及显著水平为0.95时,  $\chi^2$ 临界值等于3.841。第一个变量(即迁出的意愿)的主效应在5个模型中都是显著的, 并且符号为正。此外, 生活区与选择区一致性的主效应在5个模型中均不显著, 虽然在4个模型中这两个变量的一阶相互作用效应是显著的。正如 $\chi^2$ 值所指示的那样, 其他相互作用都很弱。第一个模型中, 第一与第二变量相互作用效应的 $\chi^2$ 值为0.120, 大大低于显著水平为0.95时的临界值, 对模型3、4、5第三变量的主效应是显著的。凡是表示生活在城市区的选择、买房或租房的选择及在生活区是否建设更多商业网点的选择等变量, 对于这样一个logit线性模型都是有意义的。

由表2得到这样一个结论: 并非饱和logit模型的所有参数都是显著的。因而, 表3给出一个相应的减少了参数个数的修正模型, 即消除意义不大的参数。

与户主年龄有关的预计从现生活区迁出的模型在两种情况下均被否定, 因为按照 $\chi^2$ 检验统计量模型的有效性最小。当估计饱和模型中减数(显著)参数时, 只有变量 $x$ 、 $y$ 和 $z_3$ 的模型未被否定: 表3中只有一类相互作用参数, 即变量 $x$ 与 $y$ 的一阶相互作用参数。比较表2和表3, 可得出的主要结论是: 变量的关系水平很稳定。仅有一个显著模型是简化型的, 那就是第三个模型, 其中买房或租房的选择与变量 $x$ 及 $y$ 相关,  $x$ 、 $y$ 的意义前面已经谈过。

**结论** 本文论述了分类数据的回归模型logit和补余重对数变换是常用的数据变换。利用logit变换可解决异方差性问题和概率预报问题。必须作出的假设是扰动项服从逻辑斯谛或极值分布函数, 当此假设合理时, 对于空间迁移分析, logit模型是一个有用的工具。

谢炳庚、卜照义摘译自《Economic Geography》, V.60, No.2, 1984, 廖小军校