

系列杀人犯罪侦查分析模型的创建与检验

曾 贇*

内容提要 当某一故意杀人犯罪发生时,侦查机关的第一要务是准确判定案件属性。其中确定案件是否属于系列杀人犯罪可谓重中之重,因为这既关系到案件能否得以及时、准确侦破,又关乎司法公正。为便于侦查机关准确判别某一故意杀人是否属于系列杀人犯罪,兹在明确界定系列杀人犯罪概念基础上,通过聚类分析、方差分析、二元逻辑回归分析等方法创建系列杀人犯罪分类模型和回归模型。依据系列杀人犯罪分类模型,侦查机关可初步基于犯罪地是否偏僻、被害人尸体是否在犯罪现场、凶犯除故意杀人是否还有抢劫、是否有犯罪前科等四个方面直观判定某一故意杀人是否为系列杀人犯罪;依据系列杀人犯罪回归模型,侦查机关可精准计算出某一故意杀人为系列杀人犯罪的概率值。分类模型的直观性和回归模型的精准性为侦查机关判别故意杀人犯罪案件属性提供了一个可具体测量的科学依据。

关键词 系列杀人 数据编码 聚类分析 方差分析 二元逻辑回归分析

十八届四中全会提出“以审判为中心”的刑事诉讼制度改革,从而为如何提高刑事司法公信力和司法公正指明了正确方向。2016年10月,最高人民法院、最高人民检察院、公安部、安全部、司法部联合下发《关于推进以审判为中心的刑事诉讼制度改革的意见》(以下简称《意见》),对以审判为中心的刑事诉讼制度改革作出全面部署。根据十八届四中全会精神和《意见》要求,保障无罪的人不受刑事追究和做到案件事实清楚及证据确实、充分被置于刑事诉讼活动的首要地位。那么侦查机关在侦办故意杀人案件中如何做到既能及时侦破案件,以防止系列杀手再次作案,又能准确侦破案件,以保障无罪的人不受刑事追究、防止冤假错案呢?通过对现有冤假错案的简单分析,我们可以发现一些冤假错案中的杀人真凶后来往往被证实为系列杀人凶犯。例如:呼格案的真凶为系列杀手赵志红、聂树斌案的真凶为系列杀手王书金、浙江“两张案”的真凶为系列杀手勾海峰。如果我们假设:对任一故意杀人案件,侦查机关根据系列杀人犯罪侦查分

* 广州大学法学院教授。本文系国家社会科学基金项目“刑事司法公信力评估指数研究”(项目批准号:14BFX059)的阶段性成果。

析模型能够及时、准确判定该案为系列杀人案件的概率值,那么一旦通过分析模型得出该案为系列杀人犯罪,则侦查机关就既能及时侦破案件,从而保证“命案必破”,又能准确办理案件,从而确保司法公正、防止冤假错案。以甘肃白银高承勇案为例,^①如果假设:在高犯第一次作案时侦查机关就能根据案件事实得出该案为系列杀人犯罪的结论,则高犯就不可能制造那么多起人间惨案。这是因为一旦我们建立起系列杀人犯罪侦查分析模型,侦查机关就可通过对故意杀人现有案件事实分析推知杀人凶犯再次实施犯罪的概率。假设根据系列杀人分析模型,得出甘肃白银案凶犯再次实施杀人犯罪的概率较高,那么根据系列杀人犯罪的显著特征,侦查机关就能及时划定凶犯再次作案的范围,并可快速锁定犯罪嫌疑人。以王书金案(聂树斌案)为例,^②如果侦查机关基于康某被害现场就能判定该案为系列杀人犯罪,那么司法机关就有理由相信该案很可能并非聂树斌所为,因为聂树斌的人生轨迹非常清晰,并不具有系列杀人犯罪的显著特征。若假设成立,则侦查机关一开始就不应将聂树斌列入该案的怀疑对象。

基于上述,对于任一已发生的故意杀人案件,准确分析和判别该案是否属于系列杀人犯罪可谓及时、准确侦破案件和确保司法公正、防止冤假错案的基础和前提。由是观之,科学创建系列杀人犯罪侦查分析模型无疑具有重要的实践意义。当然,本研究对于维护社会公共安全和防止类似人间悲剧重演来说亦具有重要的社会意义。为便于侦查机关及时、准确判别某一故意杀人是否属于系列杀人犯罪,兹在界定系列杀人犯罪概念基础上,试图采用聚类分析、方差分析、二元逻辑回归分析方法创建系列杀人分类模型和回归模型。首先,通过整群抽样、多阶抽样方法共收集单独个体多重杀人案例181件;其次,采用文本分析方法对所收集案例数据逐一进行编码,得到单独个体多重杀人犯罪数据编码表;再次,采用K均值聚类方法,创建系列杀人犯罪分类模型,采用交叉列表分析对分类模型的准确性进行检验;最后,采用二元logistics回归分析方法,创建系列杀人犯罪回归模型,同时采用交叉列表分析对回归模型的准确性进行检验。

一、模型数据的来源和编码

创建系列杀人犯罪侦查分析模型需要运用定量研究方法。如所周知,定量研究的前提是必须有可据以统计与分析的样本。当然,样本的选取依赖于概念界定的标准,研

① 犯罪人高承勇自1988年5月至2002年2月14年间实施强奸杀人11起,杀害11人。高犯作案手段极其残忍,其犯罪行为在甘肃白银一度引起公众恐慌。根据已有资料,高犯第一次犯罪时间是1988年5月16日,直至2004年8月5日,公安部门才组织专家对案件进行会诊,将白银、包头两地案件并案处理,确定为甘蒙“8·05”系列强奸杀人案。2016年8月,在距高犯首次作案28年后,侦查机关方将高犯抓获。2017年7月19日,白银市中级人民法院不公开审理了此案,检察机关指控被告人高承勇犯故意杀人罪、强奸罪、抢劫罪、侮辱尸体罪等4项罪,法院将择日宣判。

② 王书金一案中,被害人康某被司法机关认定为聂树斌所杀。1995年3月,聂树斌被石家庄中级人民法院判处死刑。在聂树斌被执行死刑21年后,2016年12月,最高人民法院再审改判聂树斌无罪。

究者对概念界定的标准不同,其所选取的样本也就各异,所得出的结论也就不尽相同。兹在界定系列杀人、多重杀人概念标准基础上,采用多阶抽样方法收集多重杀人案例 347 件,采用整群抽样方法抽取单独个体多重杀人案例 181 件。一般认为,多重杀人犯罪概念涵盖系列杀人和特大杀人两个不同子类。为科学创建系列杀人犯罪侦查分析模型,并据此得出判定某一故意杀人为系列杀人犯罪的概率计算公式,需要根据统计学的一般原理区分实验组和对照组。基于多重杀人概念界定标准,我们将系列杀人犯罪作为实验组,特大杀人犯罪作为对照组来分析和处理 181 件多重杀人案例数据。换言之,如果样本中没有对照组——特大杀人案例,则无法创建系列杀人犯罪侦查分析模型。基于此,通过文本分析方法对所选取的 181 件单独个体多重杀人案例数据逐一进行编码,得到单独个体多重杀人犯罪数据编码表。

(一) 系列杀人概念的界定

20 世纪 70 年代中期,美国联邦调查局研究员罗伯特·雷斯勒(Robert K. Ressler)首创“系列杀手”这一术语。自此,系列杀人犯罪便开始进入刑事法学者的研究视域。但是在如何界定系列杀人犯罪概念方面,学者之间仍有着各自不同的见解。其中最常被引用的是美国学者雷斯勒(Ressler)、伯吉斯(Burgess)和道格拉斯(Douglas)等关于系列杀人的定义。他们认为,所谓系列杀人是指在犯罪人实施杀人后的情感冷静期,连续实施三起或以上杀人事件,且每起杀人事件均发生在不同地点的暴力犯罪。^③然而这一定义中,有关犯罪地点、情感冷静期等概念要素的内涵却难以确定。质言之,在如何界定“犯罪地点”的个数和两起犯罪之间“情感冷静期”的时间间隔长度方面,学者相互之间难以达成一致意见。由于犯罪地点个数的计算标准难以确定,故有学者干脆提出将其从概念要素中予以剔除,例如美国弗吉尼亚大学法学院教授帕克·埃利奥特·迪茨(Dietz)即持此观点。^④对于“情感冷静期”时间间隔长度计算标准,多数学者采用模糊标准,例如美国西北大学教授詹姆斯·安拉·福克斯(Fox)和杰克·列文(Levin)认为,所谓系列杀人是指在几天、几周、几个月或几年时间内,由一人或几个犯罪人连续实施四次或以上杀人的暴力犯罪。^⑤

我国学者在界定系列杀人犯罪概念时主要考虑了以下要素:一是犯罪主体的同一性;二是杀人案件起数;三是犯罪时间的持续性;四是犯罪手段、模式的相似性。相较于国外论述,我国学者在界定系列杀人概念时增设了犯罪手段、模式要素,而略去犯罪地点要素。对于系列杀人犯罪起数的计算标准,我国学者有采用两起的计算标准,也有采

^③ R. K. Ressler, A. W. Burgess & J. E. Douglas, *Sexual Homicide: Patterns and Motives*, at 139 (Lexington Books, 1988).

^④ Park Elliott Dietz, *Mass, Serial and Sensational Homicides*, 62 Bulletin of the New York Academy of Medicine 479 (1986).

^⑤ James Alan Fox and Jack Levin, *Multiple Homicide: Patterns of Serial and Mass Murder*, 23 Crime & Just. 410 (1998).

用“多起”这种模糊标准的。^⑥我国学者关于系列杀人概念界定存在的主要问题是:没有从统计学角度给予该概念一个可据计量的具体标准。基于国内外已有研究成果,我们认为,系列杀人犯罪概念宜界定为:同一犯罪人在二个以上犯罪事件中,杀害三个以上被害人,且多个杀人犯罪之间有情感冷静期的暴力犯罪。犯罪事件个数的计算以24小时为标准,在24小时之内无论发生多少起杀人犯罪,均为一个犯罪事件。两个犯罪事件之间“情感冷静期”的计算以间隔三天为标准。一般地,学者在研究系列杀人犯罪时往往将其视为多重杀人犯罪的子类。那么何谓多重杀人犯罪?所谓多重杀人是指多个犯罪事件或单个犯罪事件中故意或者有预谋地杀害三人以上的暴力犯罪。当然,多重杀人犯罪还包括另一子类——特大杀人。所谓特大杀人是指在单个犯罪事件中,有三个以上被害人被杀害,且一般仅有一个犯罪地点的犯罪。

(二)数据来源

为创建系列杀人犯罪分类与回归模型,首先,我们采用整群抽样与多阶抽样方法,从百度百科“刑事案件”名录中抽取自1980至2016年间发生的多重杀人案例名369件。^⑦其次,通过搜索百度网站和中国知网上有关多重杀人犯罪方面的相关报道和论文,在删除相同案例名后,共收集多重杀人案例名34件。经上述两种方法,共收集403件多重杀人案例名。对于所收集的403件多重杀人案例名,通过检索百度和中国知网中有关文献,逐一收集每件多重杀人案例的详细内容。案件内容收集完毕后,经初步整理,删除了56件内容无法查实的案例,保留了347件数据资料齐全的案例。在此基础上,从347件多重杀人案例中整群抽取由单独个体实施的多重杀人案例181件。根据系列杀人、特大杀人犯罪概念界定标准,181件单独个体多重杀人案例中有69件为系列杀人,有112件特大杀人,分别约占样本总量38.1%、61.9%。之所以选择由单独个体实施的多重杀人案例作为分析样本,主要基于以下两个方面的原因:

第一,自2000年始,由单独个体实施的多重杀人犯罪明显高于由共同犯罪人实施的多重杀人犯罪。据统计,在1990—1999年10年间,由共同犯罪人实施的多重杀人犯罪明显高于由单独个体实施的多重杀人犯罪,但自2000年始,由单独个体实施的多重杀人犯罪却明显高于前者。^⑧

第二,自2000年开始,由单独个体实施的多重杀人犯罪呈明显上升趋势。据统计,

^⑥ 我国有学者认为,所谓系列杀人案件,一般是指由同一个或同一伙犯罪分子实施的、手段相同或相近的二起以上的杀人案件。另有学者认为,所谓系列杀人案件,是指在一段时间内,连续发生的由一个或一伙人所作的多起杀人案件。参见吕云平:《系列杀人案件的成因、特点及侦查对策》,载《政法学刊》2005年第3期;姚丙育:《杀人案件的侦破与实例》,中国人民公安大学出版社2006年版,第115页。

^⑦ http://baike.baidu.com/link?url=14ArMHqBNQu1An2f3D-YcmDVSh_VIXoKrD-RIytSWzm-Zs3DPQDRLiscTBIjxLNLqUmLjkGN_oGDJI1TyAgsUVlc4bZzKROZoP1RWm_fWnCTgawHJS8MamI4-p_JfuqZP,最后访问时间:2017年4月3日。

^⑧ 根据描述统计,1990—1999年,由共同犯罪人实施的多重杀人案例共60件,单独个体实施的共33件;2000—2009年,由共同犯罪人实施的多重杀人案例共68件,单独个体实施的共83件。

虽然说自 2010 年始,多重杀人案件有所减少,但根据现有数据,由单独个体实施的多重杀人犯罪明显高于共同犯罪人实施的多重杀人犯罪这一趋势并未改变。就此而言,以单独个体实施的多重杀人犯罪案件为分析样本显然具有代表性。

(三) 数据编码

对所收集的 181 件单独个体实施的多重杀人案例材料,采用文本分析法对案例内容逐一进行编码。编码的依据主要来自于对案例研究的经验总结和对相关文献的梳理分析。一般认为,较早对系列杀人犯罪现场展开研究的当属美国联邦调查局国家暴力犯罪分析中心、行为科学研究处约翰·道格拉斯(John Douglas)、帕特里克·马尼拉(Patrick Mullany)等,其于 20 世纪 70 年代创设了有序犯罪现场(organized crime scenes)和无序犯罪现场两种类型。^⑨虽然他们关于故意杀人犯罪现场特征的规整有助于侦查机关分析案件的性质类型和凶犯的社会特征,但因为有序和无序的类型划分视角过于狭窄,且有关两类人格特征的划分范围亦过于狭小,故其适用范围较少。基此,我们主要基于对所收集案例的经验总结,通过开放式编码方法的采用,提取首犯时间等 34 个概念范畴。^⑩如表 1 所示,在 34 个概念范畴中,变量的类型包括四种:一是离散型变量,例如首犯时间等;二是连续型变量,例如第一次犯罪所处刑期等;三是二分类变量,例如性别等;四是多分类变量(名义变量),例如犯罪位置等。案例数据编码过程中,由于所编多分类变量的不同水平无法穷尽事物自身所有范围,因此需要研究者持续编码。换言之,我们需要根据新的案例内容持续增加多分类变量的不同水平,从而最终形成接近穷尽事物所有范围的编码表。例如:对于犯罪位置这一概念范畴,如果某个多重杀人案发地为矿井,则就需增加“矿井”这一编码。

采用开放式编码方法对案例数据进行初始编码是数据分析和理论发现的起点,因此初始编码的范围和内容也就较为宽泛。基此,经初始编码所得到的概念范畴,其抽象程度明显较低,故需要采用主轴编码方法提取抽象程度更高的范畴。通过对 34 个概念范畴进一步提炼,得到犯罪时间等 14 个抽象程度更高的主范畴。基于统计分析一般原理和数据分析处理初步结果,我们从中只提取 8 个主范畴作为创建分类模型与回归模型的因素指标,包括犯罪地点、犯罪现场、犯罪过程、犯罪行为、犯罪动机、被害人状况、

⑨ 据其研究,有序犯罪现场类型中杀人凶犯显著的人格特征是:精神变态者,显著的社会特征有六个方面:(1)智商高于平均值;(2)具有社会技能;(3)受过一定的教育;(4)出生于城市;(5)流动性较强;(6)性生活随意。无序杀人现场类型中杀人凶犯显著的人格特征是:将被害人非人化,显著的社会特征有五个方面:(1)低于平均智商;(2)高中阶段即退学;(3)缺乏社会能力;(4)没有稳定的家庭关系;(5)具有虐待情感和退缩型人格障碍。See Albert R. Roberts, Kristen M. Zgoba, Shahid M. Shahidullah, *Recidivism among four types of homicide offenders: An exploratory analysis of 336 homicide offenders in New Jersey*, 12 *Aggression and Violent Behavior* 497 (2007).

⑩ 编码主要区分为三种类型:一是开放性编码,主要发生在编码的起始阶段,其目的在于提炼概念和范畴。二是主轴编码或关联性编码,其目的在于提炼抽象程度更高的概念和范畴,以确定现象及其属性的维度。三是选择性编码或重点编码,其目的在于系统处理诸范畴之间的关系。参见曾赞:《犯罪学中定性与定量研究》,载《山东警察学院学报》2015 年第 3 期。

犯罪人生平境遇、犯罪前科状况等。其理由主要有:一是作为实验组的系列杀人犯罪和作为对照组的特大杀人犯罪二者之间的所有变量因素需要基于相同的假设。由于初始编码范围广,而经提炼出的主范畴虽然抽象程度高,但范围依然较广,故研究者需要排除实验组和对照组不一样的假设条件。由于系列杀人凶犯一般未被当场抓获,而特大杀人凶犯则往往被及时抓获,故在创建系列杀人犯罪侦查分析回归模型时,我们需要将案发情形、犯罪人基本状况、犯罪人精神状况等因素指标予以排除。需要说明的是,虽然我们也还是保留了犯罪前科状况因素指标,但是在创建系列杀人犯罪回归模型(y_2)时也将予以删除。二是根据数据处理初步结果,有些数据缺失太多,无法通过均值补全,例如犯罪时间等。三是经初步统计分析,一些因素指标因为与其他因素指标之间高度相关,例如动机行为,故亦将其予以排除。

由于8个主范畴所涵盖的因素指标有些属于多分类变量、有些属于离散型变量,故为科学创建二元逻辑回归模型,需要将其转化为二分类变量。对于8个主范畴所涵盖因素指标的转换,我们主要基于经验概括与理论抽象,并通过重点编码方法的采用,将其依次转换为18个二值变量,由此得到18项测量指标。基于前述三种编码方法,最终形成“单独个体多重杀人犯罪数据编码表”(表1)。

表1:单独个体多重杀人犯罪数据编码表

概念范畴 (开放式编码)	主范畴 (主轴编码)	观测范畴 (二值编码)
1. 首犯时间	a. 犯罪时间	
2. 犯罪时点: ①凌晨(0-5); ②早上(5-11); ③中午(11-13); ④下午(13-17); ⑤傍晚(17-19); ⑥晚上(19-24)		
3. 犯罪季节: ①春天; ②夏天; ③秋天; ④冬天		
4. 犯罪位置: ①火车站、广场、城门口、银行门口、岗亭、街道、马路、村落等人群密集地区; ②被害人家(办公室、宿舍); ③犯罪人家(办公室、宿舍); ④菜地、玉米地、果园、防空洞等荒郊野外; ⑤街道、马路、院落、村落等偏僻处; ⑥招待所、茶馆、饭店、电影院、寺庙、敬老院等公共服务设施内; ⑦网吧、游戏厅、酒吧、红灯区等敏感地区内; ⑧出租车、地铁、电车、火车、船只、飞机等公共交通设施内	1. 犯罪地点	1. 犯罪地是否偏僻
5. 犯罪区域: ①特定地点; ②同一区域; ③区域不定	b. 危害结果	
6. 被杀害人数		
7. 财产损失		

概念范畴 (开放式编码)	主范畴 (主轴编码)	观测范畴 (二值编码)
8. 作案工具:①钝器;②锐器;③枪支;④爆炸物、易燃物;⑤毒物;⑥绳子等细长物;⑦交通工具	2. 犯罪现场	2. 是否未携带凶器作案 3. 尸体是否不在现场
9. 尸体处置:①尸体在犯罪现场;②埋尸;③藏尸;④分尸;⑤焚尸;⑥抛尸		
10. 犯罪手段:①绞杀;②扼杀;③钝器重击;④锐器刺戳;⑤割喉;⑥砍头;⑦电击;⑧毒杀;⑨枪击;⑩纵火;⑪爆炸;⑫锐器砍杀;⑬驾车冲撞	3. 犯罪过程	4. 是否采用绞杀、扼杀方法 5. 是否侮辱、摧残尸体
11. 犯罪标记:①折磨、拷打;②殴打;③拘禁、捆绑;④塞、堵(嘴)、蒙蔽(眼);⑤威胁被害人;⑥用刀戏谑刺划;⑦多次用力砍杀;⑧用力穿透目标;⑨切割生殖器;⑩损毁生殖器;⑪切挖腹部;⑫切挖胸腔;⑬切挖面部(眼睛);⑭奸尸;⑮食被害人身体部位;⑯喝被害人血;⑰切割乳头;⑱火烧阴毛;⑲切挖心肺;⑳解剖尸体;㉑侮辱、摧残尸体;㉒烹尸;㉓鸡奸;㉔扒光衣物		
12. 犯罪行为:①强奸;②抢劫;③盗窃;④放火;⑤投毒;⑥猥亵;⑦绑架;⑧爆炸;⑨盗窃、抢夺、抢劫枪支、弹药、爆炸物;⑩劫持;⑪故意伤害	4. 犯罪行为	6. 是否有强奸行为 7. 是否有抢劫行为
13. 犯罪动机:①权力与控制;②信仰与使命;③伪理性型;④家庭杀伐;⑤追求刺激;⑥追求利益;⑦心怀愤恨;⑧精神病人	5. 犯罪动机	8. 是否追求刺激 9. 是否追求利益 10. 是否心怀愤恨 11. 是否为权力控制
14. 动机行为:①变态杀人;②强奸杀人;③抢劫杀人;④套保杀人;⑤重罪杀人;⑥恐怖袭击;⑦报复杀人;⑧疯狂滥杀	c. 动机行为	
15. 被害人性别	6. 被害人状况	12. 被害人心智是否未达成熟 13. 被害人是否属于易遭受攻击群体 14. 犯罪人与被害人是否为熟人
16. 被害人生理特征:①智障者;②年老疾病者;③儿童;④少年		
17. 被害人社会特征:①妓女;②独行妇女;③其他独行者;④求职妇女;⑤其他求职者;⑥乘客;⑦旅客;⑧顾客;⑨雇员;⑩雇主;⑪警察;⑫家庭成员;⑬居民;⑭同村、同事、同学、生意合作者		
18. 被害人与犯罪人关系:①家庭成员(或亲戚)关系;②其他熟悉人关系;③陌生人关系		

概念范畴 (开放式编码)	主范畴 (主轴编码)	观测范畴 (二值编码)
19. 案发情形:①相互之间的争吵;②感情纠纷; ③家庭纠纷;④债权债务纠纷;	d. 案发情形	
20-25. 犯罪人出生日期、性别、民族、文化程度、 职业、婚姻	e. 犯罪人基本 状况	
26. 犯罪人早年生平境遇:①疾病;②父母一方 或双方去世;③父母离异;④被遗弃;⑤虐待;⑤ 身体缺陷;⑥孤独;⑦贫穷	7. 犯罪人生平 境遇	15. 个体早年是否 遭遇不幸
27. 犯罪人成年生平境遇:①遭遇解雇或解雇威 胁;②同所爱的人分离;③离异;④破产;⑥失业		16. 个体成年是否 遭遇不幸
28. 犯罪人情感支持状况		17. 情感是否长期 得不到安放
29. 精神病史	f. 犯罪人精神 状况	
30. 自杀史		
31. 犯罪前科次数	8. 犯罪人前科 状况	18. 是否有犯罪前 科
32. 第一次罪名		
33. 第一次刑期		
34. 第一次犯罪时间		

二、分类模型的创建及检验

在完成对系列杀人犯罪现场等数据的编码之后,需要首先考虑的问题是如何利用编码表创建一个能使侦查机关直观判定某一故意杀人是否为系列杀人的数学模型。我们认为,采用聚类分析方法创建分类模型是实现上述目的原初路径。之所以首先选择聚类分析方法创建分类模型,而不是首先采用回归分析方法创建回归模型,是因为案发当时如果嫌疑人未被及时抓获,那么侦查机关当然无法根据犯罪人相关资料来判定案件类型;即便嫌疑人被当场抓获,侦查机关亦难以判断犯罪人将来是否会再次实施杀人犯罪。就此而言,如要创建用于测量某一故意杀人是否为系列杀人犯罪的数学模型,就应以未知杀人案件属于何种类型为理论预设的前提。显然,针对未知事件类型的判别理当选择聚类分析法。兹根据表 1,在确立 8 个主范畴名下 18 项二值变量基础上,采用 K 均值聚类法,创建系列杀人犯罪分类模型。

(一) 聚类分析变量的确定

如表1所示,8个主范畴所涵盖的因素指标可以用作聚类分析的变量。兹基于对181件多重杀人案件的经验分析,就8个主范畴名下18项二值变量的确立作如下说明。

1. 犯罪地点。^① 根据表1,犯罪地点因素属于无序多分类变量。基于经验判断,其可转换为“犯罪地是否偏僻”二值变量。犯罪地为荒郊野外,街道、院落等的偏僻处,其均可被定义为犯罪地偏僻。犯罪地为犯罪人家(办公室、宿舍)的,也可被定义为偏僻场所,因为被害人一旦受蒙骗来到犯罪人家(办公室、宿舍)内,就难以摆脱犯罪人的控制。当然,其余诸项则可被定义为非偏僻场所。根据描述统计,181件多重杀人案例中有44件犯罪发生地为偏僻场所,约占样本总数24.3%。

2. 犯罪现场。根据表1,犯罪现场因素包括犯罪工具和尸体处置等2项无序多分类变量。基于案件事实,其可分别转换为“是否未携带凶器作案”、“尸体是否不在犯罪现场”等2项二值变量。对“是否未携带凶器作案”因素指标,凡凶犯未携带钝器、锐器、枪支、爆炸物、易燃物、毒物、交通工具等作案,重新编码为“是”,余则为“否”。需要说明的是,我们将使用随身携带的布料、皮鞭等细长物品作案视为未携带凶器作案。对“尸体是否不在犯罪现场”因素指标,凡凶犯有埋尸、藏尸、分尸、焚尸、抛尸、买卖尸体等行为的,重新编码为“是”,余则为“否”。即便被害人尸体确实是在犯罪现场被发现的,但如果杀人凶犯有埋尸、藏尸、分尸、焚尸等行为,从而导致被害人尸体不易被人们发现的,亦被视为被害人尸体不在犯罪现场。根据描述统计,181件多重杀人案例中凶犯未携带凶器作案的共23件,尸体不在犯罪现场的共45件,分别约占样本总量12.7%、24.9%。

3. 犯罪过程。根据表1,犯罪过程因素包括犯罪手段、犯罪标记等2项无序多分类变量。基于案件事实,其可分别转换为“是否采用绞杀、扼杀方法”“是否侮辱、毁损尸体”等2项二值变量。对于前者,凡凶犯采用绞杀、扼杀方法作案的,重新编码为“是”,采用诸如钝器重击、锐器刺戳、割喉、枪击、纵火、爆炸等方法作案的,则编码为“否”。对于后者,凡凶犯具有⑨-⑳等行为的,重新编码为“是”,余则为“否”。根据描述统计,181件多重杀人案例中凶犯采用绞杀、扼杀方法作案的共29件,^②有侮辱、毁损尸体行为的共28件,分别约占样本总量16.0%、15.5%。

4. 犯罪行为。根据表1,基于统计分析,犯罪行为因素可转换为“是否有强奸行为”、“是否有抢劫行为”等2项二值变量。根据描述统计,181件多重杀人案例中仅有故意杀人行为的共62件,约占样本总量34.3%;除了故意杀人行为之外,既有抢劫行为,又有强奸行为的共21件,约占样本总量11.6%;除了故意杀人行为之外,还具有强奸行为、抢劫行为的分别为32件、56件,约占样本总量17.7%、30.9%。

① 对于犯罪地点的研究,我国有学者采用犯罪制图和地理信息技术来绘制犯罪地图,以精准描述犯罪空间分布、组合、数量、密度等方面的特征。参见单勇:《犯罪地图的公开》,载《国家检察官学院学报》2016年第5期。

② 由于有些案件中的杀人凶犯既携带凶器,同时又采用绞杀、扼杀方法杀害被害人,因此采用绞杀、扼杀方法的案件数与未携带凶器作案的案件数不尽相同。

5. 犯罪动机。根据表1,犯罪动机因素属于无序多分类变量,其可转换为“是否为追求刺激型”、“是否为追求利益型”、“是否为心怀愤恨型”、“是否为权力和控制型”等4项二值变量。根据描述统计,181件多重杀人案例中犯罪动机为追求刺激型、追求利益型、心怀愤恨型、权力和控制型的分别有30件、40件、98件和13件,约占样本总量16.6%、22.1%、54.1%和7.2%。^⑬

6. 被害人状况。根据表1,被害人状况因素包括被害人生理—心理特征、被害人社会特征、被害人与犯罪之间关系等3项多分类变量。基于案件事实,其可分别转换为被害人心智是否未达成熟、被害人是否属于易遭受攻击人员、被害人与犯罪人是否为熟人等3项二值变量。对于被害人生理—心理特征因素的二值编码,凡被害人为智障者、年老疾病者、儿童、少年等,编码为“心智未达成熟”,余则为“否”;对于被害人社会特征因素的二值编码,凡被害人为妓女、独行妇女、其他独行者等,编码为“易遭受攻击人员”,余则为“否”。根据描述统计,181件多重杀人案例中被害人心智未达成熟、被害人为易遭受攻击人员、被害人与犯罪人为熟人的分别有49件、24件、78件,约占样本总量27.1%、13.3%和43.1%。

7. 犯罪人生平境遇。根据表1,犯罪人生平境遇因素包括个体早年生平境遇、成年生平境遇、情感支持状况等3项无序多分类变量。对于个体早年生平境遇因素的二值编码,凡个体早年时期遭遇疾病、父母一方或双方去世、父母离异、遗弃、虐待、身体缺陷、孤独、贫穷等不幸事件者,重新编码为“早年遭遇不幸”,余则为“否”;对于个体成年生平境遇因素的二值编码,凡个体成年时期遭遇解雇或解雇威胁、同所爱的人分离、离异、破产等不幸事件的,重新编码为“成年遭遇不幸”,余则为“否”;对于个体情感支持状况的二值编码,可直接将其转为“个体情感是否长期得不到安放”。根据描述统计,181件多重杀人案例中个体早年遭遇不幸事件、成年遭遇不幸事件、情感长期得不到安放的分别有31件、49件和50件,约占样本总量17.1%、27.1%和27.6%。

8. 犯罪人前科状况。根据表1,犯罪前科状况因素可转换为“是否有犯罪前科”二值变量。需要特别说明的是,凡犯罪人曾受过强制戒毒、行政拘留等行政处罚的,其亦被视为有犯罪前科记录。根据描述统计,181件多重杀人案例中犯罪人有前科记录的共57件,约占样本总量31.5%。

(二) 分类模型的创建

如果假设181件单独个体多重杀人案例相互之间存在着程度不同的相似性(亲疏关系),那么依据可以度量案例之间相似性的统计量,就可以把相似程度较大的案例聚

^⑬ 98件“心怀愤恨型”多重杀人犯罪案件中,系列杀人仅1件,余则为特大杀人案件。对于“心怀愤恨型”特大杀人犯罪,我国有学者将其归入独狼恐怖主义犯罪序列中,并将“纠正个人冤屈”作为独狼恐怖主义犯罪的动机目的之一。我国另有学者在论述恐怖主义的个体罪因时,将仇恨和报复作为恐怖主义的动机因素之一。参见曾赞:《论独狼恐怖主义犯罪的构成要素》,载《政法论坛》2016年第5期;兰迪:《恐怖主义的个体罪因》,载《国家检察官学院学报》2016年第9期。

合为一类,把另外一些相似程度较大的案例聚合为另一类。显然,案例之间的这种“亲疏程度”是重要的,因为它直接影响最终聚类结果。本研究中我们采用K均值聚类方法对181件多重杀人案例进行分类。K均值聚类方法采用欧式距离作为相似性测度的统计量,认为两个案例之间的距离越近,相互之间的相似度就越大。在样本容量较大时,研究者需要选择一批凝聚点或给出一个初始分类,让案例按照最短距离原则向凝聚点凝聚。向凝聚点凝聚的过程就是一个不断修改凝聚点或进行迭代的过程,这一过程将一直持续至分类比较合理或迭代稳定为止。

将基于重点编码和二值编码方法取得的18项测量指标作为聚类分析的变量,将181件多重杀人案例序号作为个案依据,采用K均值聚类方法,可得初始聚类中心。经第一次迭代后,二个类的中心点分别偏移了1.770和1.237。经第七次迭代后,二个类的中心点偏移均小于系统判定标准(0.02),聚类结束,得到多重杀人犯罪最终聚类中心表(表2)。根据表2,第一类多重杀人犯罪的显著特征有:一是犯罪地偏僻;二是被害人尸体不在犯罪现场;三是杀人凶犯在杀害被害人后还有抢劫行为;四是杀人凶犯有犯罪前科。除此而外,第一类还具有以下非典型特征:一是一些杀人凶犯在作案时未携带凶器;二是杀人凶犯具有权力和控制动机欲望。第二类多重杀人犯罪的显著特征有:一是杀人凶犯的犯罪动机表现为心怀愤恨型;二是犯罪人与被害人是熟人。基于概念分析与经验观察,第一类多重杀人犯罪为系列杀人,第二类多重杀人犯罪为特大杀人。据此,根据多重杀人犯罪最终聚类中心表,得到系列杀人犯罪分类模型。依据系列杀人犯罪分类模型,侦查机关可以初步从犯罪地是否偏僻、被害人尸体是否不在犯罪现场、凶犯是否有抢劫行为、是否有犯罪前科等4个方面直观判别某一故意杀人是否属于系列杀人犯罪。当然,为能全面分析案件属性,侦查机关需根据18项二值变量,通过系列杀人犯罪分类模型来判别某一故意杀人案件是否属于系列杀人犯罪。

表2:多重杀人犯罪最终聚类中心表

	聚类			聚类	
	1	2		1	2
1. 犯罪地是否偏僻	1	0	10. 是否为权力与控制型	0.13	0.02
2. 是否未携带凶器作案	0.28	0.03	11. 是否为心怀愤恨型	0	1
3. 尸体是否不在犯罪现场	1	0	12. 被害人是否未达成熟	0	0
4. 是否采用绞杀、扼杀方法	0	0	13. 被害人是否属于易遭受攻击群体	0	0
5. 是否侮辱、摧残尸体	0	0	14. 犯罪人与被害人是否熟人	0	1
6. 是否有强奸行为	0	0	15. 个体早年是否遭遇不幸	0	0
7. 是否有抢劫行为	1	0	16. 个体成年是否遭遇不幸	0	0
8. 是否为追求刺激	0	0	17. 个体情感是否长期得不到安放	0	0
9. 是否为追求利益	0	0	18. 是否有犯罪前科	1	0

(三) 分类模型的检验

根据系列杀人犯罪分类模型的分类结果,181 件多重杀人案例中有 72 件被判定为系列杀人,109 件被判定为特大杀人犯罪。依据系列杀人和特大杀人犯罪概念,181 件多重杀人案例中有 69 件为系列杀人,112 件为特大杀人犯罪。为便于区分,将基于分类模型所得分类称为统计分类,将基于概念界定所作分类称为概念分类。为检验系列杀人犯罪分类模型的准确性,采用交叉列表分析法进行检验。通过交叉列表,得到概念分类与统计分类交叉列表(表 3)。根据表 3,统计分类中被确定为系列杀人犯罪的 72 件案例中有 66 件被概念分类定义为系列杀人犯罪。质言之,被分类模型预测为系列杀人犯罪的 72 件案例中有 66 件被准确预测,因此系列杀人犯罪分类模型的预测准确率为 95.7%。统计分类中被确定为特大杀人的 109 件案例中有 3 件被概念分类定义为系列杀人犯罪。质言之,被分类模型预测为特大杀人的 109 件案例中有 3 件案例被预测错误,因此系列杀人犯罪分类模型的预测错误率为 4.3%。

表 3:概念分类与统计分类交叉列表

			概念分类		合计
			0	1	
统计 分类	1.00	计数	6	66	72
		聚类 中的 %	8.3%	91.7%	100.0%
		是否系列杀人 中的 %	5.4%	95.7%	39.8%
	0.00	计数	106	3	109
		聚类 中的 %	97.2%	2.8%	100.0%
		是否系列杀人 中的 %	94.6%	4.3%	60.2%
合计		计数	112	69	181
		聚类 中的 %	61.9%	38.1%	100.0%
		是否系列杀人 中的 %	100.0%	100.0%	100.0%

根据系列杀人犯罪分类模型,若以聂树斌案中康某被害犯罪事实和甘肃白银案中高成勇所犯第一起犯罪事实作为文本分析对象,那么聂树斌案(王书金案)和甘肃白银案都被测定为系列杀人犯罪。假设在聂树斌案侦查、起诉或审理三个阶段中的任何一个阶段,尤其是侦查阶段,司法工作者能够依据系列杀人犯罪分类模型将该案判定为系列杀人犯罪,那么他们就有充分理由怀疑聂树斌案处理结论的正确性,因为根据聂树斌的人生轨迹,其不具有系列杀人犯罪的显著特征。假如根据高成勇所犯第一起犯罪事实,侦查机关就能确定该案为系列杀人犯罪,那么他们就可根据系列杀人犯罪特征来确定案件侦查的范围和划定杀人凶犯再次实施犯罪的重点区域。果真如此,高犯很可能不会在距其首次作案 28 年后才被侦查机关抓获。可以想象,如果侦查关能够依据统计分析模型来科学判定某一杀人案件的类型,那么类似的人间悲剧很可能会得以避免。

三、回归模型的创建及检验

根据前述,依据系列杀人犯罪分类模型,侦查机关可以直观判定某一杀人案件的类型属性,亦可据此测量某一故意杀人是否为系列杀人犯罪。但分类模型的不足之处亦显而易见:其一,据以创建模型的测量指标有 18 项之多,而其对因变量的影响却未经检验;其二,聚类结果往往会受凝聚点选择好坏的影响,从而使得系列杀人分类结果不够稳定;其三,模型的拟合效果无从检验。一言以蔽之,系列杀人犯罪分类模型最大的不足之处是稳定性不够。解决这一问题最好的办法是:基于聚类分析模型的分类结果,利用回归分析方法,创建回归模型,以准确计算出某一杀人案件为系列杀人犯罪的概率值。所谓回归分析实际上就是基于已知事件建立数学测量模型(概率模型),然后将其推至未知事件。然而由于未知事件不可能完全与已知事件完全相同,故回归模型测量结果的准确性常受质疑。但如果基于聚类分析结果,我们就可以克服回归分析法的这一不足,因为用于创建回归模型的测量值指标是通过聚类分析得到,而分类模型又是基于未知事件得到的。鉴于此,首先,利用聚类分析结果创建方差分析模型,以拣选出创建回归模型的测量指标。其次,采用二元逻辑回归方法,创建二元逻辑回归模型;再次,根据回归方程,计算出 181 件多重杀人为系列杀人犯罪的概率值;最后,通过交叉列表分析,检验回归模型的准确性。

(一) 回归模型测量指标的提取:基于方差分析

在创建二元逻辑回归模型之前,我们采用方差分析方法来提取对因变量具有显著影响的测量指标。依据系列杀人犯罪分类模型所得分类结果,将是否为系列杀人犯罪作为因变量,8 个主范畴名下的各二值变量作为自变量,利用 spss18.0 统计软件,采用方差分析方法,创建 8 个方差分析模型。在此基础上,从中分别提取 $p < 0.05$ 的变量作为创建系列杀人犯罪回归模型的测量指标。

1. 犯罪地点。通过对统计分类的描述统计,72 件系列杀人案例中有 13 件发生在犯罪人家(办公室、宿舍),有 10 件发生在菜地、玉米地、果园、防空洞等荒郊野外,有 16 件发生在街道、马路、院落、村落等地的偏僻处,即:犯罪发生在偏僻场所的案件共有 39 件,约占系列杀人犯罪样本总量 54.2%。通过方差分析,创建犯罪地点方差分析模型。模型检验为: $F = 84.247, p = 0.000$,说明模型具有统计学意义。根据方差分析表,犯罪地是否偏僻(x_1)对因变量具有显著影响。基此,系列杀人犯罪地点的一个显著特征是:多数案件发生在较为偏僻的场所。

2. 犯罪现场。通过对是否未携带凶器作案、尸体是否不在犯罪现场等 2 项测量指标的描述统计,72 件系列杀人案例中未携带凶器作案共 20 件,尸体不在犯罪现场共 40 件,分别约占系列杀人犯罪样本总量 27.8%、55.6%。通过多因素方差分析,创建犯罪现场方差分析模型。模型检验为: $F = 34.548, P = 0.000$,说明所用模型具有统计学意

义。根据方差分析表,是否未携带凶器作案(x_2 , $F=9.376$, $P=0.003$),尸体是否不在犯罪现场(x_3 , $F=27.657$, $p=0.000$)等2项自变量对因变量有显著影响。基此,系列杀人犯罪现场的显著特征有:一是凶犯一般未携带凶器作案;二是尸体一般不在案发现场。

3. 犯罪过程。通过对是否采用绞杀、扼杀方法,是否侮辱、毁损尸体等2项测量指标的描述统计,72件系列杀人案例中采用绞杀、扼杀方法作案的共26件,有侮辱、损毁尸体行为的共29件,分别约占系列杀人犯罪样本总量36.1%、40.3%。通过多因素方差分析,创建犯罪过程方差分析模型。模型检验为: $F=41.114$, $P=0.000$,说明所用模型具有统计学意义。根据方差分析表,是否采用绞杀、扼杀方法(x_4 , $F=15.295$, $P=0.000$),是否侮辱、摧残尸体(x_5 , $F=27.368$, $P=0.000$)等2项自变量对因变量具有显著影响。基于此,系列杀人犯罪过程的显著特征有:一是杀人凶犯一般采用扼杀、绞杀的方法作案;二是凶犯在完成杀人犯罪行为后往往还有侮辱、摧残尸体行为。

4. 犯罪行为。通过描述统计,72件系列杀人案例中杀人凶犯除有故意杀人行外,还有强奸行为的共31件,有抢劫行为的共55件,同时有强奸、抢劫两个行为的共21件,分别约占系列杀人犯罪样本总量43.1%、76.4%和29.2%,此处将同时具有强奸、抢劫两个行为的案件分别重复计入强奸、抢劫行为总量中。基于方差分析,创建犯罪行为方差分析模型。模型检验为: $F=243.719$, $P=0.000$,说明模型完全具有统计学意义。根据方差分析表,是否有强奸行为(x_6 , $F=91.443$, $P=0.000$)、是否有抢劫行为(x_7 , $F=119.315$, $P=0.000$)等2项自变量对因变量均具有显著影响。基于此,系列杀人犯罪行为的一个显著特征是:杀人凶犯除具有故意杀人行之外,还具有强奸、抢劫等犯罪行为。

5. 犯罪动机。通过描述统计,72件系列杀人案例中犯罪动机为追求刺激型30件、追求利益型37件、心怀愤恨型1件、权力和控制型4件,分别约占系列杀人样本总量41.7%、51.4%、1.4%和5.6%。系列杀人案例中仅有1件系列杀人凶犯出于报复动机。基于方差分析方法,创建犯罪动机多因素方差分析模型。模型检验为: $F=689.239$, $P=0.000$,说明模型具有统计学意义。根据方差分析表,犯罪动机是否为追求刺激型(x_8 , $F=913.799$, $P=0.000$)、是否为追求利益型(x_9 , $F=788.853$, $P=0.000$)、是否为权力和控制型(x_{10} , $F=332.067$, $P=0.000$)等3项自变量对因变量均具有显著影响。基于此,系列杀人犯罪动机的一个显著特征是:杀人凶犯一般出于追求刺激、利益等而实施系列杀人犯罪。

6. 被害人状况。根据描述统计,72件系列杀人案例中被害人心智未达成熟、被害人为易遭受攻击人员、被害人与犯罪人为熟人的分别有19件、22件和8件,约占系列杀人犯罪样本总量26.4%、30.6%和11.1%。基于方差分析方法,创建被害人状况多因素方差分析模型。模型检验为: $F=17.670$, $P=0.000$,说明模型具有统计学意义。根据方差分析表,被害人是否属于易遭受攻击群体(x_{11} , $F=8.854$, $P=0.003$)对因变量具有显著影响,其余2项因素对因变量不具有显著影响。基此,系列杀人犯罪被害人状况的一个显著特征是:被害人多为易遭受攻击人员。

7. 犯罪人生平境遇。根据描述统计,72件系列杀人案例中个体早年遭遇不幸、成年

时期遭遇不幸和情感长期得不到安放的分别为 14 件、17 件和 22 件,约占系列杀人样本总量 19.4%、23.6% 和 30.6%。基于多因素方差分析方法,创建犯罪人生平境遇方差分析模型。模型检验为: $F=0.769, P=0.614$,说明模型完全不具有统计学意义。根据方差分析表,个体早年是否遭遇不幸($P=0.982$)、个体成年是否遭遇不幸($P=0.186$)和个体情感是否长期得不到安放($P=0.264$)等 3 项因素对因变量均不具有显著影响。

8. 犯罪人前科状况。根据描述统计,72 件系列杀人案例中有犯罪前科的案例共 39 件,约占系列杀人样本总量 54.2%。基于方差分析,创建方差分析模型。模型检验为: $F=33.443, P=0.000$,说明模型具有统计学意义。根据方差分析表,是否有犯罪前科(x_{12})对因变量具有显著影响。基此,系列杀人犯罪人前科状况的一个显著特征是:系列杀人凶犯多数有犯罪前科记录。

综合上述,通过方差分析,从 8 个主范畴方差分析模型中初步提取了 12 项因素作为系列杀人犯罪回归模型的测量指标,具体包括:(1)犯罪地是否偏僻(x_1);(2)是否未携带凶器作案(x_2);(3)尸体是否不在犯罪现场(x_3);(4)是否采用绞杀、扼杀方法(x_4);(5)是否侮辱、摧残尸体(x_5);(6)是否有强奸行为(x_6);(7)是否有抢劫行为(x_7);(8)是否为追求刺激型(x_8);(9)是否为追求利益型(x_9);(10)是否为权力和控制型(x_{10});(11)被害人是否属于易遭受攻击群体(x_{11});(12)是否有犯罪前科(x_{12})。

(二) 系列杀人犯罪回归模型的创建:基于二元逻辑回归分析

在拣选出系列杀人犯罪回归模型测量指标后,需要通过多重共线性诊断来判断各自变量之间的显著相关性。通过多重共线性诊断,那些可以通过其他自变量来解释的因素将被从模型中剔除。在此基础上,通过二元逻辑回归分析方法创建系列杀人犯罪回归模型。

1. 多重共线性诊断

在创建二元逻辑回归模型前,需要通过多重共线性诊断来检验各自变量之间相关性,因为二元逻辑回归的参数估计要求各自变量之间相互独立,即一个自变量不能用其他自变量来解释。也就是说,多重共线性问题会掩盖检验的显著性,从而导致模型不稳定。据以判定多重共线性的统计量主要有方差膨胀因子(VIF)、条件指数(CI)、方差比例等。一般地,当 $VIF > 10$ 时被认为存在严重的多重共线性,也有人主张用 $VIF > 5$ 作为判定严重的多重共线性的标准。当 $CI > 10$ 时,被认为存在弱多重共线性;当 $CI > 30$ 时,被认为存在中等多重共线性;当 $CI > 100$ 时,被认为存在严重的多重共线性。对于方差比例判定标准而言,一般认为方差比例的值越大,则多重共线性越强。依据系列杀人概念所得的概念分类结果,将是否为系列杀人作为因变量,同时将基于方差分析所得的 12 项测量指标作为自变量,通过多元回归分析方法可得多元回归模型系数表和共线性诊断表(表 4)。根据表 4,我们发现:其一, x_8 的方差膨胀因子值最大($VIF=5.284$);其二, CI 最大值=9.248,接近 10;其三, x_8 、 x_9 的方差比例最大,分别为 0.83、0.78。基

此, x_8 、 x_9 等 2 个自变量被认为同其他 9 个自变量之间存在多重共线性。

表 4: 多重共线性诊断表

模型	维数	特征值	条件指数	方差比例												
				(常量)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	1	6.035	1.000	.01	.01	.01	.01	.01	.01	.01	.00	.00	.00	.01	.01	
	2	1.471	2.025	.03	.00	.02	.00	.01	.01	.01	.02	.01	.07	.00	.00	.03
	3	1.036	2.413	.00	.00	.00	.00	.01	.00	.01	.00	.01	.01	.34	.00	.01
	4	.869	2.635	.00	.00	.13	.00	.10	.16	.00	.01	.01	.00	.00	.09	.00
	5	.666	3.009	.04	.01	.04	.07	.02	.01	.00	.00	.01	.07	.00	.05	.35
	6	.622	3.114	.11	.00	.25	.00	.00	.00	.02	.00	.00	.01	.00	.46	.03
	7	.549	3.314	.13	.13	.08	.16	.03	.02	.13	.04	.00	.00	.00	.02	.00
	8	.415	3.815	.49	.19	.04	.00	.16	.03	.02	.02	.01	.00	.00	.02	.13
	9	.383	3.967	.09	.00	.33	.08	.45	.01	.00	.00	.00	.00	.01	.22	.12
	10	.339	4.219	.05	.29	.06	.07	.00	.13	.44	.01	.00	.00	.05	.12	.02
	11	.298	4.501	.04	.04	.02	.14	.10	.05	.21	.16	.08	.04	.00	.00	.22
	12	.245	4.963	.01	.28	.00	.31	.12	.50	.15	.05	.04	.02	.00	.00	.05
	13	.071	9.248	.00	.04	.03	.14	.00	.06	.00	.69	.83	.78	.58	.01	.04

对于 x_8 、 x_9 等 2 个自变量同其余自变量之间的多重共线性问题,我们还可以通过多元回归分析方法来进行进一步检验。将 x_8 、 x_9 分别作为因变量,将其余 11 个变量作为自变量,创建犯罪动机是否为追求刺激型和追求利益型等 2 个多元回归模型。根据前者, x_8 同 x_1 ($P=0.001$)、 x_2 ($P=0.005$)、 x_3 ($P=0.000$)、 x_5 ($P=0.000$)、 x_6 ($P=0.026$)、 x_7 ($P=0.000$)、 x_9 ($P=0.000$)、 x_{10} ($P=0.000$)、 x_{12} ($P=0.000$) 等 9 个自变量之间存在显著相关性。质言之, x_8 可以通过其余 9 个自变量来解释。根据后者, x_9 同 x_3 ($P=0.000$)、 x_7 ($P=0.000$)、 x_8 ($P=0.000$)、 x_{10} ($P=0.000$)、 x_{12} ($P=0.028$) 等 5 个自变量之间存在显著相关性。质言之, x_9 可以通过其余 5 个变量来解释。基于前述,在创建二元逻辑回归分析模型时,将是否为追求刺激型(x_8)、是否为追求利益型(x_9)等 2 项因素指标予以剔除。在剔除 x_8 、 x_9 等 2 项自变量后,我们发现 10 个自变量中 VIF 最大值仅为 1.956,条件指数最大值为 4.55。据此,多重共线性问题得以消除。

2. 系列杀人犯罪回归模型的创建

将基于系列杀人概念所得的概念分类——是否为系列杀人作为因变量,将 x_1 、 x_2 、 x_3 、 x_4 、 x_5 、 x_6 、 x_7 、 x_{10} 、 x_{11} 、 x_{12} 等 10 项测量指标作为自变量,通过二元逻辑回归分析,采用逐步向前进入法(似然比),运用 spss18.0,创建系列杀人犯罪回归模型。所谓逐步向前进入法是指在每次引入一个新变量时对先前引入方程的老变量逐个进行检验,将对因变量不具有显著影响的变量从中剔除,直至没有剔除的变量方再引入新的变量。通过逐步向前回归,模型中仅保留了 4 项对因变量具有显著影响的自变量指标,具体包括:犯罪地是否偏僻(x_1 , $P=0.000$)、尸体是否不在犯罪现场(x_3 , $P=0.001$)、是否有抢劫行

为($x_7, P=0.005$)、是否有犯罪前科($x_{12}, P=0.000$)。根据回归模型结果(表5),可得系列杀人犯罪回归方程: $y_1 = -5.454 + 4.608 * x_1 + 3.669 * x_3 + 5.583 * x_7 + 2.686 * x_{12}$ 。经检验:模型的-2倍对数似然值为50.651,说明模型具有统计学意义;模型的Nagelkerke R^2 为0.884,说明模型的拟合效果极佳。^⑭

表5:系列杀人犯罪二元logistics回归模型(y_1)

		B	S. E.	Wals	df	Sig.	Exp (B)	EXP(B)的95%	
								下限	上限
步骤1 ^a	是否有抢劫行为 x_7	3.839	.499	59.260	1	.000	46.491	3.839	.499
	常量	-1.719	.249	47.610	1	.000	.179	-1.719	.249
步骤2 ^b	犯罪地是否偏僻 x_1	4.939	.804	37.770	1	.000	139.568	4.939	.804
	是否有抢劫行为 x_7	5.014	.736	46.383	1	.000	150.558	5.014	.736
	常量	-3.540	.586	36.537	1	.000	.029	-3.540	.586
步骤3 ^c	犯罪地是否偏僻 x_1	3.939	.956	16.982	1	.000	51.365	3.939	.956
	尸体是否不在犯罪现场 x_3	3.467	.940	13.611	1	.000	32.051	3.467	.940
	是否有抢劫行为 x_7	5.619	.928	36.637	1	.000	275.614	5.619	.928
	常量	-4.489	.820	29.991	1	.000	.011	-4.489	.820
步骤4 ^d	犯罪地是否偏僻 x_1	4.608	1.161	15.764	1	.000	100.263	4.608	1.161
	尸体是否不在犯罪现场 x_3	3.669	1.059	12.014	1	.001	39.218	3.669	1.059
	是否有抢劫行为 x_7	5.583	1.019	30.013	1	.000	265.820	5.583	1.019
	是否有犯罪前科 x_{12}	2.686	.961	7.811	1	.005	14.671	2.686	.961
	常量	-5.454	.997	29.937	1	.000	.004	-5.454	.997

当采用该模型来测量某一故意杀人是否为系列杀人犯罪时,上述模型(y_1)所面临的一个问题是:在不知道凶手是谁的情况下,是否有犯罪前科这一测量指标并不具有实际测量功能。鉴于此,将是否有犯罪前科(x_{12})这一指标予以剔除,保留 x_1 、 x_2 、 x_3 、 x_4 、 x_5 、 x_6 、 x_7 、 x_{10} 、 x_{11} 等9项测量指标。将前述9项测量指标作为自变量,再次创建二元逻辑回归模型(表6)。经逐步向前回归,模型中仅保留了4项对因变量具有显著影响的测量指标,包括:犯罪地是否偏僻($x_1, P=0.000$)、尸体是否不在犯罪现场($x_3, P=0.003$)、是否采用绞杀、扼杀方法($x_4, P=0.045$)、是否有抢劫行为($x_7, P=0.000$)。根据回归结果,系列杀人犯罪二元逻辑回归方程为: $y_2 = -4.683 + 3.974 * x_1 + 3.014 * x_3 + 2.762 * x_4 + 5.650 * x_7$ 。经检验:模型的-2倍对数似然值为56.193,说明模型具有统计学意义。Nagelkerke R^2 为0.869,说明模型的拟合效果极好;采用 Hosmer—Lemeshow 检验, $\text{sig}=0.958 > 0.05$,说明模型的拟合度极强。^⑮

⑭ Nagelkerke R^2 被称为伪决定系数,以有别于多元回归方程中决定系数这一称谓。Nagelkerke R^2 越接近1,模型的拟合效果就越强。

⑮ 它是一种 Logistic 模型拟合优度的检验方法,由 Hosmer 和 Lemeshow 在1989年提出。其原假设为模型能很好地拟合,因此当 $\text{sig.} > 0.05$ 时,接受原假设,即认为模型能够很好地拟合数据。

表 6: 系列杀人犯罪二元逻辑回归模型(y_2)

		B	S. E,	Wals	df	Sig.	Exp(B)	EXP(B)的 95%	
								下限	上限
步骤 1 ^a	是否有抢劫行为 x_7	3.839	.499	59.260	1	.000	46.491	17.492	123.564
	常量	-1.719	.249	47.610	1	.000	.179		
步骤 2 ^b	犯罪地是否偏僻 x_1	4.939	.804	37.770	1	.000	139.568	28.892	674.207
	是否有抢劫行为 x_7	5.014	.736	46.383	1	.000	150.558	35.562	637.410
	常量	-3.540	.586	36.537	1	.000	.029		
步骤 3 ^c	犯罪地是否偏僻 x_1	3.939	.956	16.982	1	.000	51.365	7.889	334.418
	尸体是否不在犯罪现场 x_3	3.467	.940	13.611	1	.000	32.051	5.080	202.223
	是否有抢劫行为 x_7	5.619	.928	36.637	1	.000	275.614	44.680	1700.170
	常量	-4.489	.820	29.991	1	.000	.011		
步骤 4 ^d	犯罪地是否偏僻 x_1	3.974	1.047	14.392	1	.000	53.176	6.826	414.259
	尸体是否不在犯罪现场 x_3	3.014	.998	9.122	1	.003	20.361	2.881	143.911
	是否采用绞杀、扼杀方法 x_4	2.762	1.378	4.014	1	.045	15.830	1.062	235.974
	是否有抢劫行为 x_7	5.650	.969	34.027	1	.000	284.244	42.585	1897.237
	常量	-4.683	.870	28.989	1	.000	.009		

(三) 系列杀人犯罪回归模型的运算

根据系列杀人犯罪二元逻辑回归方程(y_2), 可以计算出某一故意杀人犯罪为系列杀人的概率, 概率公式为: $P = ez / 1 + ez$ ($z = -4.683 + 3.974 * x_1 + 3.014 * x_3 + 2.762 * x_4 + 5.650 * x_7$)。以聂树斌案(王书金案)为例, 根据案发现场, 可作如下认定: 犯罪地偏僻(x_1); 被害人尸体被发现在犯罪现场(x_3); 凶犯采用布条绞杀方法致被害人死亡(x_4)。因为被害人所穿连衣裙不在案发现场(根据王书金的供述, 在当时他想要把连衣裙带回家给他的妻子穿, 后担心被侦查机关发现, 故又将连衣裙就地埋藏), 故可认定凶犯有抢劫行为(x_7);^⑩也可以认定为没有抢劫行为, 因为当时的侦办干警可能并未想到这一细节。根据计算, 如果认定凶犯除故意杀人外还有抢劫行为, 则该案被判定为系列杀人犯罪的概率值约为 0.999; 如果认定凶犯没有抢劫行为, 则被判定为系列杀人犯罪的概率约为 0.886。根据任一计算结果, 王书金案(聂树斌案)被判定为系列杀人案无疑。对甘肃白银高承勇第一起犯罪案件, 根据案发现场, 可作如下认定: 犯罪地(在被害人家)偏僻; 被害人尸体在犯

^⑩ 参见焦新波、张君:《十年前王书金案采访实录》, 载《民主与法制》2015 年第 7 期。

罪现场;采用锐器刺戳;现场有侵财迹象。^① 根据计算:该案为系列杀人的概率值约为 0.997。根据计算结果,甘肃白银案被判定为系列杀人犯罪。

(四) 系列杀人犯罪回归模型的检验

根据概率公式 $P = ez / (1 + ez)$ ($z = -4.683 + 3.974 * x_1 + 3.014 * x_3 + 2.762 * x_4 + 5.650 * x_7$), 分别计算出基于概念分类所得 69 件系列杀人案件的概率值。通过计算, 69 件系列杀人案件的概率值分布为: ①有 3 件 < 0.5 ; ②有 23 件 < 0.95 , 但 > 0.5 ; ③有 43 件 > 0.95 。一般认为, 概率值大于 0.5 时, 被认为某事很可能发生。基此, 根据系列杀人犯罪回归模型计算结果, 有 3 件预测错误, 有 66 件预测准确, 模型预测的准确率约为 95.7%。当然, 我们还可以采用交叉列表方法来检验模型的准确性。根据交叉列表分析, 181 件多重杀人案例中有 69 件为系列杀人、112 件为特大杀人。69 件系列杀人案件被模型准确预测为系列杀人犯罪的有 66 件, 错误预测为特大杀人犯罪的有 3 件, 模型预测准确率为 95.7%; 模型的总体预测准确率为 95%。112 件特大杀人案例被模型准确预测为特大杀人犯罪有 106 件, 错误预测为系列杀人犯罪有 6 件, 模型预测准确率为 94.6%。

结 语

综合全文, 在通过文本分析法得到多重杀人犯罪数据编码表的基础上, 采用 K 均值聚类方法创建系列杀人犯罪分类模型; 采用方差分析、多重共线性诊断、二元回归分析等方法创建系列杀人犯罪回归模型。依据系列杀人犯罪分类模型, 侦查机关可以简单而又直观判定某一故意杀人是否为系列杀人犯罪; 依据系列杀人犯罪回归模型, 侦查机关可以精准计算某一故意杀人为系列杀人犯罪的概率, 从而准确判定案件的属性。总的来看, 系列杀人犯罪侦查分析模型的创建是刑事诉讼实践活动中的一项重要发现, 它为提升我国刑事司法公信力和公正性提供了一个可具体测量的科学路径。然行文至此, 尚有以下两个问题需要说明和厘清:

第一, 系列杀人犯罪分类和回归模型赖以创建的基础数据既非官方资料, 也非第一手调查或访谈资料, 而是来自于媒体的纪实报道或中国知网上的相关论文, 那么据此创建的模型可信吗? 我们认为, 据此创建的模型十分可信, 因为模型所用的数据均是客观的。显然, 对于犯罪地点(x_1), 被害人尸体身在何处(x_3), 是否采用绞杀、扼杀方法(x_4), 杀人凶犯是否还具有抢劫行为(x_7), 犯罪人是否有前科(x_{12})等 5 项自变量数据资料, 我们相信媒体报道和相关论文不存在虚假的、错误的陈述。就此而言, 基于系列杀人犯罪分类模型与

^① 1988 年 5 月 26 日, 白银市公安局白银分局刑警王磊在工农路派出所接到报案后赶赴案发现场, 经勘验, 床上都是血, 被害人身上二十多处伤, 颈部被切开, 手上也有搏斗伤。现场有侵财迹象, 受害人屋里被翻得乱七八糟。根据办案民警回忆, “从尸检情况看, 没有发生性行为, 但有猥亵迹象”。参见《老刑警曝光白银命案细节: 没有受害人穿红衣服》, http://news.youth.cn/sh/201608/t20160831_8608673.htm, 最后访问时间: 2017 年 5 月 30 日。

回归模型,侦查机关只需要根据犯罪现场客观事实就可直接测量某一故意杀人是否为系列杀人犯罪,而无需高深的刑事侦查技术和复杂的实验设备。

第二,系列杀人犯罪分类模型与回归模型究竟孰优孰劣?纵观全文,二者间有相同点:一是两者对某一故意杀人是否为系列杀人的预测准确率均高达95.7%;二是分类模型中的核心解释变量与回归模型中的4项测量指标基本相同。两者间也有不同点:一是前者中的18项测量指标均来自经验判断,后者中的4项测量指标则是通过方差分析、多重共线性诊断、二元逻辑回归分析等统计方法的综合应用得到的,故后者对因变量的解释力明显优于前者。二是前者的拟合效果未经任何检验,后者则经似然比检验、Nagelkerke R^2 和 Hosmer—Lemeshow 等三种不同方法检验,故后者的可靠性明显优于前者。

Abstract: When an intentional homicide crime happens, the first task of investigation organs is to judge whether it is a serial murder, which is the most important, because it is not only related to whether investigation organs can solve the murder case correctly and promptly, but also related to judicial justice. In order to judge scientifically whether the murder is a serial case, on the basis of clearing the definition of serial murder, the classification model and regression model of the serial murder will be created through the cluster analysis, analysis of variance and dualistic logistic regression analysis. According to the classification model of serial murder, investigation organs can simply and directly judge whether the murder is a serial one from whether the crime place is isolated or not, whether the corpse is at the crime scene or not, whether there is a robbery or not, whether criminals have the criminal history or not. According to the regression model of serial murder, investigation organs can accurately calculate the probability of whether the murder is a serial one. The intuition of classification model and the accuracy of regression model provide a measurable and scientific foundation for investigation organs to judge the attribute of serial murder.

(责任编辑:李 游)