


应用统计优秀案例丛书

首届应用统计专业硕士 优秀案例选

 周勇 主编

 中国统计出版社
China Statistics Press

应用统计优秀案例丛书

首届应用统计专业硕士 优秀案例选



周勇 主编



中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

首届应用统计专业硕士优秀案例选 / 周勇主编. ——
北京: 中国统计出版社, 2016.4
(应用统计优秀案例丛书)
ISBN 978-7-5037-7771-4

I. ①首… II. ①周… III. ①统计学—案例 IV.
①C8

中国版本图书馆 CIP 数据核字(2016)第 081265 号

首届应用统计专业硕士优秀案例选

作 者/周 勇

责任编辑/姜 洋

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/http://www.zgtjcs.com/

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787mm×1092mm 1/16

字 数/600 千字

印 张/24.5

版 别/2016 年 5 月第 1 版

版 次/2016 年 5 月第 1 次印刷

定 价/60.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在
世界任何地区以任何文字翻印、拷贝、仿制或转载。
如有印装差错,由本社发行部调换。

《应用统计优秀案例丛书》

编委会

主 编：周 勇

副主编：李金昌 纪 宏

编 委：（按姓氏笔画为序）

王兆军 王学钦 王德辉 石 磊 朱建平

刘 杨 刘妍岩 刘 洪 孙六全 杜金柱

李 元 吴耀华 张鸣芳 张宝学 陈 敏

陈耀辉 林华珍 林金官 郑 明 房祥忠

赵彦云 郭建华 唐年胜 曹景林 崔恒建

程维虎 傅德印 童行伟 濮晓龙

编委会秘书：常 宁 阮 敬

《应用统计优秀案例丛书》序

统计学是关于收集、整理、分析和解释统计数据的科学,属于一门认识方法论性质的科学,其目的是探索数据内在的数量规律性,以达到对客观事物的科学认识。“由部分推及全体”的思想贯穿于统计学的始终。由于无论在科学研究还是社会生活中都有着大量的数据产生,所以统计学在自然科学、工程技术、生物医学、经济金融、社会科学、人文和管理科学等许多学科领域都有着广泛的应用,并且推动着这些学科的发展。

统计学科既是一门具有严谨科学理论与方法体系的独立学科,同时也是与数学、工业工程、生物医学、经济学、金融学、管理科学和行为科学等多学科交叉的一门应用学科,并同这些学科相结合而发展出许多相关的统计学二级学科和学科分支。例如生物统计学、生物计量学、环境计量学、精算学和金融计量学等。

统计学科和工业工程、计算机科学、社会科学的交叉,极大地推动了社会经济、科学与管理的发展。经济金融与社会科学的现代化越来越显现出与统计科学相互交叉、渗透而发展的特色。由此可见,统计学正成为整个科学门类中重要的基础及各类应用交叉学科的核心。统计学是我国高等教育中较新的理学门类下的一级学科。

在大数据时代,统计学遇到前所未有的机遇,同时也面临极大的挑战。首先是对统计学传统的理论和方法造成极大的困扰,传统统计学的方法和理论需要发展和创新,因此也是统计学发展的机遇。对统计学的学习和教学、研究都带来新的局面。

对于应用统计教学和人才培养而言,具体和实际的案例起着重要和特殊的作用,许多拔尖的统计分析人才和经营高手还在学校学习期间,就有过把所学到的统计知识成功和有效地应用到实践之中的案例,并由此获得了应用所学统计知识的快乐和兴趣。

到目前为止,国内有关统计学的应用案例还很少,对于统计学的教学和学

习仍然存在瓶颈,同时提高人们统计学的学习兴趣也不足,因此中国统计出版社出版这套《应用统计优秀案例丛书》,以期提高人们兴趣来学习统计学,并更加了解统计学的应用背景,加大统计学在学和用的有机结合的力度。对我国统计学的教育事业、人才培养发挥显著的作用。

《应用统计优秀案例丛书》选材都是来自每两年一届的全国应用统计专业硕士案例大赛获奖的优秀案例,经过编委会的精心挑选,组织修改而完成。这套书宗旨是面向大学统计专业、经济金融专业、社会科学专业和工业工程专业等的高年级学生、研究生以及青年教师,针对一些重要的统计应用领域,作较宽泛的介绍,既注重统计学的基础知识,又反映统计学具有宽广的应用领域。

这套书得到全国大部分统计学院的长期大力支持,编辑人员也为其付出了艰辛的工作。这套书凝聚了广大作者和编委会的诸多努力,仍然有许多不足,但我们相信它必将获得全国应用统计工作者、广大读者喜爱。我们真诚地希望大家关心和支持它的发展。使它越办越好,为统计教育事业做出进一步贡献。

周 勇

2015年12月

前 言

统计学是一门关于数据的科学和艺术。随着科学技术的发展,领域理论日益拓展和细分化,数据分析技术与工具也日益完善和先进。与此同时,统计学也得到了前所未有的分支学科领域扩容和深层次的延伸,理工农医文,没有一个领域用不到统计学。为适应我国现代统计事业发展对应用统计专门人才的迫切需要,完善应用统计人才培养体系,创新应用统计人才培养模式,提高应用统计人才培养质量,为我国经济建设培养实用型人才,2011年国务院学位办特设置应用统计硕士专业学位。

21世纪,随着信息化、市场化的加深,统计学让数据“大有可观”,应用统计学更是让数据变“活”、变“亲切”、变得更有价值。正是由于统计在社会经济生活中的深入应用,具有较强应用能力并能够解决重要实际问题的统计学高级应用人才的社会需求日益旺盛。为培养应用统计专业学位研究生发现问题、研究问题、解决问题和评价问题的能力以适应社会需求,并探讨适合我国应用统计硕士的教学方法,提高我国高校统计学教师在应用统计方法与案例教学方面的水平,全国应用统计专业学位研究生教育指导委员会举办了“2014全国应用统计专业学位研究生案例大赛”。

大赛组委会于2013年12月发出比赛通知后,众多高校老师给予了大力支持,学生们也踊跃报名,积极参与。案例大赛采用小组制,每个小组由一位指导教师和三名学生组成。2014年7月,经过大赛评审专家的专业评审,大赛共决出了10个一等奖,20个二等奖,20个优秀奖,为了确保比赛的专业和公允,大赛评审专家库由具备应用统计专业硕士学位授予权单位的教师组成,从专家库中随机抽取专家对参赛作品进行评审。

参赛作品涵盖了经济、金融、生物医药、工业、环境、教育、电子商务等方方面面,讨论的问题都是目前社会和生活中的热点话题,所用数据也都来自实际,更有很多案例所用的是通过亲自调查得到的一手数据;在分析问题和建模的过程中,思路开阔,统计方法多种多样,有比较经典的多元统计分析、时间序

列预测等方法,也有BP神经网络、决策树、包络分析模型等较为前沿的方法,解决问题时运用的统计软件的种类也非常多,有matlab、R、SPSS、SAS、Eviews、Python等。比赛的过程和结果都充分达到了大赛的目的与预期。

为了扩大全国应用统计专业硕士案例大赛的影响力,实现举办大赛的初衷以及长期的良性发展,大赛将每两年举办一次,每次选取部分优秀案例出版案例集,作为各校案例教学的参考资料,以期在应用统计专业硕士的教学过程中,把统计知识体系渗透到案例教学,使学生在案例学习中掌握具体统计分析技术,扩展学生的眼界和思路,激发学生的学习兴趣,培养学生理论联系实际,用所学的统计方法进行统计建模和分析,解决实际问题的能力,达到更好的教学效果。

本次出版的案例是从获得一、二等奖的案例中精心挑选出来的,由于这是应用统计专业硕士案例大赛优秀案例集的第一次出版,尝试中难免会出现一些问题,如案例格式,案例编选的适用性等,这些都需要大家在实际教学和使用时发现问题,反馈问题,给我们提出宝贵的意见和建议,也希望有更多的学生和高校参与到大赛中来,以推动我国应用统计专业硕士的能力培养,推动未来应用统计人才的快速成长,推动我国统计事业的健康快速发展。

本书在编辑过程中得到全国统计界同仁的大力支持,首先要感谢首都经济贸易大学提供了大量人力和物力支持,以使此书得以顺利出版。其次感谢以下专家在本书出版前期做的大量工作,他们是陈敏、赵彦云、刘洪、顾六宝、王孟欣、郑明、朱仲义、朱建平、吴耀华、于丹、房祥忠、童行伟、张宝学、崔恒建、阮敬、毛炳寰、黄性芳、李昂、白欢朋、李钰、陈庆焕、汪红霞,正是这些工作使本书能早日与大家见面。再次,感谢编委会的认真工作,委员们多次开会讨论,不辞辛苦,最终形成本书的样稿。最后,感谢全国应用统计专业硕士的学子们,是他们积极参与案例大赛,付出艰辛的努力才使本书成为“有源之水和有米之炊”,使这本书具有新颖性和可读性。一直关心统计教育事业的各行各业广大读者,是你们的支持使得统计学的发展更上了一层楼,在此诚挚致谢!

李金昌 纪宏

2016年2月

目 录

基于中国海南省橡胶树台风灾害易损性的统计分析	郭菲菲	李昕泽	唐思汝	1	
某石油公司成品油调价策略分析	李 超	矫雨驰	冯晓明	17	
广州市二手房房价影响因素的相关研究	李 苏	刘志翔	闫炳硕	31	
打车软件用户满意度研究报告					
——以湖北省武汉市出租车乘客为例	刘 芳	许逸翰	郑文芳	70	
行业股票收益率联动性分析					
——以中国地产行业为例	赵丽娟	刘聪聪	赵 蕾	93	
陕西省西安市 PM _{2.5} 相关因素分析及预测	郑国义	牛彩云	田 学	114	
比特币的统计分析	靳爱慧	潘明阳	梁程远	138	
我国商业银行利率风险 VaR 度量					
——基于 SHIBOR 的实证研究	沈 华	李 黎	柯 婧	154	
函数型数据分析方法在蛱蝶分类识别中的应用	詹 静	郭文雯	王玲玲	171	
环境约束下东北老工业基地绿色全要素生产率的估算					
与分析	曹 政	杨 琪	刘 捷	193	
基于投入产出模型的主导产业选择方法及其应用研究					
——以上海市为例	俞 萍	刘 亚	丁 昊	215	
分层广义线性模型与车险索赔强度预测			王 莹	234	
基于 ARIMA 和 BP 神经网络集成模型的天然气价格预测	蒋晚秋		王 莹	258	
贫困山区小学生心理焦虑影响因素统计调查研究	张 菁	单 韦	吴季翔	278	
余额宝能走多久					
——以江西省南昌市高校数据为例	姜雯昱	李彭忠	吴 杉	301	
河北省老龄人口生活质量分析	麻东露	张世芳	刘宁婷	324	
基于环境保护视角下吉林省低碳经济综合评价及监测					
研究	侍术凯	安 佳	杨 苹	346	
梨温高速公路交通事故致因的统计分析研究					
.....	胡 辉	王峰娟	唐 蓉	周芷茗	362

基于中国海南省橡胶树台风灾害易损性的统计分析

郭菲菲 李昕泽 唐思汝

(指导教师:童行伟)

北京师范大学

一、问题描述

(一)选题背景及研究意义

天然橡胶因其特殊性能被广泛运用于国防、工业、交通、医药卫生领域和日常生活等方面,是四大基础工业原料之一,其重要性可以与石油、煤炭、钢铁相提并论,是国家的重要战略物资。随着经济的飞速发展,我国逐步成为了世界最大的橡胶消费国^[1]。海南省是我国最大的天然橡胶生产基地,海南种植面积达到 752 万亩,年产干胶 37.23 万吨,面积和产量均占全国一半以上。海南从事橡胶产业人员已达 200 多万人,占海南省 1/4 人口,其中胶农有 75 万多人。

自然灾害是如今人类面临的全球性的重大问题之一,尤其是热带气旋灾害更是以发生频率高、影响最严重著称。而台风又是热带气旋中产生的危害性最严重、破坏性最大的一种天气系统。台风持续的时间虽然并不长,但由于其强大的破坏力,故一旦登陆将会对陆地上大面积的作物造成危害。与此同时,由于海南省特殊的地理位置,在热带季风气候影响下,海南省成为了我国受热带气旋影响最严重的地区之一。已收集西北太平洋 1949—2009 年影响海南的热带气旋就有 358 场,强烈的热带气旋带来了严重的人员伤亡和经济损失,特别是对海南的橡胶林造成了严重的损害。例如 2005 年 18 号台风“达维”,是近十年对海南橡胶树造成损害最大的台风,统计海南农垦区受风害 3 级以上的橡胶树,其中已开割橡胶树受害率达 51%,未开割树受害率达 33.9%,总损失金额多达 24 亿元。^[2]

面对台风灾害给橡胶树带来的损失,运用统计方法明确致灾因子强度与损失之间的定量关系,对于识别风险、应对风险以及相关保险产品的设计均有一定的意义。实际保险理赔中,当一场台风影响过后,需要保险公司职员去到农场实地核算损失,时间耗费较长,可能耽误理赔,从而影响胶农的再生产。针对台风灾害对于海南省橡胶林造成的严重损失,对于承灾体进行合理的易损性分析,从而可以根据致灾因子强度定量判断损失情况,对于应对风险以及相关保险产品的设计均有一定的意义。之后的进一步研究则可以基于致灾因子强度和损失情况的定量关系计算保险赔付概率分布、期望损失等指标,最终完成保险指数构建,从而使得保险公司的及时理赔以保证胶农的及时投入再生产,这样对于保险公司和胶农而言是双赢的事情。

(二)研究框架

本次研究主要针对海南省并且基于农场尺度数据进行的。海南省是我国最南端的省份,在 1949—2011 年间,对海南周围 200 公里范围内海域产生影响的热带气旋共 366 场,其中台风及以上强度的共 230 场;登陆海南的热带气旋共 144 场,其中台风及以上强度的 89 场。^[3]

按照风险模型中“致灾因子—承灾体—易损性”的结构,致灾因子方面有历史实测风速以及

通过风场模拟得到的模拟风速;承灾体方面,将海南橡胶林农场的空间分布信息与收集到的各农场损失数据进行匹配,建立风速与损失率对应的数据集;然后通过运用合适的函数对致灾因子强度与损失率数据进行拟合,得到易损性曲线,并研究其拟合效果;得到易损性曲线之后,按照《中华人民共和国气象行业标准》^[4]按照风力等级的划分对不同风速下的损失进行离散化,得到每一个风力等级下对应的预测损失率,这样可以方便实际应用;最后,估计出损失率的方差,得到每一个风力等级下损失率的可能范围,以便于进一步识别风险和 Related 保险产品的设计。

二、数据

(一)致灾因子数据

致灾因子数据分为实测风速和模拟风速。实测风速即历史台风气象站点观测数据,观测数据又分为人工站观测数据和自动站观测数据。取各站点在每场台风过程中所观测极大风速的最大值作为该站点的风速指标。

然而,由于海南岛内气象人工站点只有 18 个,而损失数据和农场有 72 个,无法形成风速—损失率的一一对应关系,而自动站虽然分布广泛,但是由于建站较晚,所以监测到的数据有限,无法用来进行统计分析。因此需要对每个农场分配一个适合站点,以其所测风速作为该农场的风速指标进行估算。这样就会出现不同农场对应同一风速的情况,从而产生系统误差,导致不能很好的估计风速与损失率之间的关系。因此,提高易损性模型精度的一个重要修正就是风速数据修正,实现一个农场对应一个数据。因此就需要采用风场模型得出的模拟风速进行接下来的研究。此研究采用的模拟风速是减灾与应急管理研究院的同学采取风场模型得出的结果,精确度还有待提高。并且台风风速分为极大风速和最大风速,极大风速是指某个时段内出现的最大瞬时风速(瞬时风速是指 3s 的平均风速)值;最大风速是指在某个时段内出现的最大 10min 平均风速值。由于 Usbesk 等人研究发现损失与极大风速相关性最大^[5],因此下文的研究采用的是极大风速。

(二)承灾体数据

对于本研究,承灾体即为种植在海南省农场中的橡胶树,已收集 72 个农场及其下属 2548 个生产队的空间信息。在 1970 年到 2009 年农场橡胶种植信息方面已经收集的主要指标包括:橡胶种植面积,当年定植株数,橡胶种植总株数,已开割橡胶树面积,干胶总产量,年亩产干胶,单株年产干胶,以及由以上数据衍生出的未开割橡胶树面积、株数。

(三)橡胶树损失数据

已收集 9 场台风损失数据,其中 0016 号“悟空”、0308 号“天鹅”、0312 号“科罗旺”及 0518 号“达维”的损失数据分布较广,且损失程度各不相同,具有良好的代表性和可计算性,因此将其作为本次案例分析使用的数据^①。主要指标包括:农场名,农场中心经纬度,开割树损失金额,开割树总株数,开割树受害数,开割树受害率,开割树倒伏数,开割树半倒数,开割树断主干 2 米以下数,开割树断主干 2 米以上数,开割树全部主枝折断数;未开割树损失金额,未开割树总株数,未开割树受害数,未开割树受害率,未开割树倒伏数,未开割树半倒数,未开割树断主干 2 米以下数,未开割树断主干 2 米以上数,未开割树全部主枝折断数;总损失金额,总株数,总损失数,总损失率。根据《海南省农垦 50 年统计资料汇编》^[6]将橡胶树风灾受损类型如表 1 所示。

① 本书中涉及的案例数据、程序命令等资料,读者可登陆中国统计出版社网站(网址: <http://www.zgtjcb.com>)自行下载。

表 1 橡胶树风灾受损类型

损失类型	损失情况
倒伏	指橡胶树主干倾斜超过 45 度,即橡胶树主干倾斜后与地平线(面)夹角小于 45 度
半倒	指橡胶树主干倾斜 30 度至 45 度,即橡胶树主干倾斜后与地平线(面)夹角大于或等于 45 度且小于或等于 60 度。
断主干 2 米以下	五级风害断主干:指橡胶树主干 2 米以下折断,即橡胶树主干折断的断截面(部位)离地面高度小于 2 米(含 2 米)
断主干 2 米以上	四级风害断主干:指橡胶树主干 2 米以上折断,即橡胶树主干折断的断截面(部位)离地面高度大于 2 米(不含 2 米)
全部主枝折断	指橡胶树的主枝全部折断

三、指标选取

热带气旋带来的致灾因子主要有大风、降雨以及风暴潮,基于要研究的对象为海南省橡胶树的损失率,则选取风速。并且台风风速分为极大风速和最大风速,极大风速是指某个时段内出现的最大瞬时风速(瞬时风速是指 3s 的平均风速)值;最大风速是指在某个时段内出现的最大 10min 平均风速值。对于风速由于要考虑可以导致灾害发生的极值风速,因此采用极大风速。损失数据首先橡胶树分为开割和未开割两种,每一种橡胶树的倒伏情况又分为五种分别有对应的损失数据。根据分析历史数据中的极大风速和损失率可以来预测不同风速等级下损失的状况。

四、统计分析

(一)模型建立

根据风场模型得出的模拟极大风速和对应的损失率数据画出散点图,分析风速与损失率之间的相关性。通过计算 pearson 相关系数,得出开割橡胶树模拟极大风速与损失率的相关系数为 0.525,未开割橡胶树模拟极大风速与损失率之间的相关系数为 0.534,且 p 值均小于 0.001,说明相关性是显著的。可以看出极大风速与损失率之间是存在显著的相关性的,很有研究的必要和意义。并且未开割树相关性更高一些,可能是因为幼树扎根浅,质量轻,幼嫩易损伤等。

经过实际数据的过程发现,单纯的只采用参数回归或非参数回归去研究似乎结果总是不够理想。而事实上,在研究回归问题的时候可以综合运用参数回归和非参数回归,于是就想到采用一种广义回归模型^{[7],[8],[9]}。即 $Y = g(X\beta) + \epsilon$,其中 g 为连接函数,可以采用核光滑法或者样条光滑法进行估计^{[7][10]}。此处基于实际问题中损失率大于 0 小于 1,并且是单调递增的,可

以取 $g = \frac{e^x}{1+e^x}$ 、 $g = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 。

通过对历史数据的拟合效果的对比同时结合数据实际背景意义,广义线性模型更好一些,并且由于 $g = \frac{e^x}{1+e^x}$ 、 $g = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 两种连接函数得出的拟合效果相差不多,故选用形式相对较为简单的连接函数 $g = \frac{e^x}{1+e^x}$ 。

综上,本研究最后确定采用的易损性函数模型为:

$$Y = g(X\beta) + \epsilon = \frac{e^{X\beta}}{1 + e^{X\beta}} + \epsilon, X = (1, x), \beta = (\beta_1, \beta_2)^T \quad (1)$$

其中 $x = (x_1, x_2, \dots, x_n)^T$, n 是样本个数, x 是风速, y 是损失率。

基于上一部分建立的模型对风速和损失数据进行处理得到不同类型损失的易损性曲线。为了方便实际应用,根据《中华人民共和国气象行业标准》按照风力等级的划分对不同风速下的损失进行离散化,得到每一个风力等级下对应的预测损失率。具体做法是采用每一个风力等级风速范围的中数所对应的的损失率预测值当做这个风力等级下的损失率。从而每一种橡胶树的每一种损失类型都得到从五级到大于等于十五级共 11 种损失率。

然而实际应用中只给出不同风力等级对应的损失率总是有误差存在,为了方便应用更希望给出一个损失率的置信区间。为此需要先得到估计参数的渐进正态性从而得到参数的方差,然后再计算得到损失率的方差,从而给出损失率的置信区间。

(二) 参数估计的渐进正态性

由于本研究采用的易损性曲线模型是 $g(X\beta) = \frac{e^{X\beta}}{1 + e^{X\beta}}$, 其中估计未知参数 β 采用的最小二乘法,即得出 β 的估计值为

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n [y_i - g(X_i\beta)]^2 \quad (2)$$

则

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma^*), \Sigma^* = A^{-1} \Sigma A^{-1'} \quad (3)$$

其中

$$A = \frac{1}{n} \sum_{i=1}^n \frac{\partial U_i(\beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n \frac{[\exp(-X_i\beta)]^2}{[1 + \exp(-X_i\beta)]^4} X_i^T X_i$$

$$\Sigma = \frac{1}{n} \sum_{i=1}^n U_i(\beta) U_i(\beta)^T = \frac{1}{n} \sum_{i=1}^n \frac{[\exp(-X_i\beta)]^2 [y_i - g(X_i)]^2}{[1 + \exp(-X_i\beta)]^4} X_i^T X_i$$

以上结论的具体证明过程如下:

由于 $\hat{\beta}$ 满足式(2),则对式(2)关于 β 求导等于 0,即

$$\sum_{i=1}^n [y_i - g(X_i\beta)] X^T f'(X_i\beta) = 0$$

为了方便下面的推导,将上式记作 $U_n(\beta) = \sum_{i=1}^n U_i(\beta) = 0$

由泰勒展式得 $0 = U_n(\hat{\beta}) = U_n(\beta) + U'_n(\beta)(\hat{\beta} - \beta)$

整理得

$$\hat{\beta} - \beta = -[U'_n(\beta)]^{-1} U_n(\beta) = -\left[\sum_{i=1}^n U'_i(\beta)\right]^{-1} \sum_{i=1}^n U_i(\beta)$$

$$= -\frac{1}{\sqrt{n}} \left[\frac{1}{n} \sum_{i=1}^n U'_i(\beta)\right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i(\beta)$$

由中心极限定理(CLT)得 $\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i(\beta) \xrightarrow{d} N(0, \text{Var}(U_i(\beta)))$

且
$$\frac{1}{n} \sum_{i=1}^n U'_i(\beta) \xrightarrow{n \rightarrow \infty} E(U'(\beta))$$

所以
$$\sqrt{n}(\hat{\beta} - \beta) = -[E(U'(\beta))]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i(\beta)$$

$$\xrightarrow{d} N(0, [E(U'(\beta))]^{-1} \text{Var}(U_i(\beta)) ([E(U'(\beta))]^{-1})^T)$$

记 $A = E(U'(\beta)), \Sigma = \text{Var}(U_i(\beta))$

则 A 和 Σ 估计值为 $\hat{A} = \frac{1}{n} \sum_{i=1}^n U'_i(\hat{\beta})$ 和 $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n U_i(\hat{\beta}) U_i^T(\hat{\beta})$ 。将函数 $g(X\beta) = \frac{e^{X\beta}}{1 + e^{X\beta}}$ 的具体形式代入计算即可得到要证的结论。

(三) 损失方差估计

根据历史样本数据可以估计出 β ，之后我们想求出 $g(X\beta) = \frac{e^{X\beta}}{1 + e^{X\beta}}$ 的方差，从而得到估计损失的不确定性，最后可以给出损失数据的区间估计。根据 δ 方法可以知道需要先求出 β 的渐进正态分布的方差^{[11][12]}。

δ 方法：设 β 为是参数，并设 $\hat{\beta}$ 为 β 的渐进正态估计，即 $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma^*)$ ，又设 $g(X\beta)$ 是满足 $g'(X\beta) \neq 0$ ，且 $g'(X\beta)$ 为有限的函数，

则

$$\sqrt{n}[g(X\hat{\beta}) - g(X\beta)] \xrightarrow{d} N(0, (g'(X\beta))^T \Sigma^* g'(X\beta)) \quad (4)$$

以上结论的具体证明过程如下：

$$g(X\hat{\beta}) = g(X\beta) + (g'(X\beta) + R_n)(\hat{\beta} - \beta)$$

当 $\hat{\beta} \rightarrow \beta$ 时，则 $R_n \xrightarrow{p} 0$ ，从而得到 $g'(X\beta) + R_n \xrightarrow{p} g'(X\beta)$

$$\text{则 } \sqrt{n}[g(X\hat{\beta}) - g(X\beta)] = \sqrt{n}[(g'(X\beta) + R_n)(\hat{\beta} - \beta)] \xrightarrow{d} N(0, (g'(X\beta))^T \Sigma^* g'(X\beta))$$

因此 设 $X_0 = (1, x_0)$ 为历史风速数据，则

$$\sqrt{n}[g(X_0\hat{\beta}) - g(X_0\beta)] \xrightarrow{d} N(0, \left(\frac{X_0^T \exp(-X_0\beta)}{(1 + \exp(-X_0\beta))^2}\right)^T \Sigma^* \frac{X_0^T \exp(-X_0\beta)}{(1 + \exp(-X_0\beta))^2}) \quad (5)$$

从而估计出损失率 $g(X\beta)$ 的方差，进一步可以算出每一种风力等级下损失的一个范围，从而为灾害识别和保险产品设计及理赔提供依据。

五、案例分析及结果

通过上述研究，本次案例分析主要是基于以上方法对 0016 号“悟空”、0308 号“天鹅”、0312 号“科罗旺”及 0518 号“达维”四场典型台风的加总的模拟极大风速以及损失率的易损性分析，分别得出开割橡胶树和未开割橡胶树的五种损失类型的损失率的预测值以及方差的预测值，从而进一步得出每一级风力等级下的损失率范围的预测。具体结果如下列图表所示。

(一) 开割橡胶树

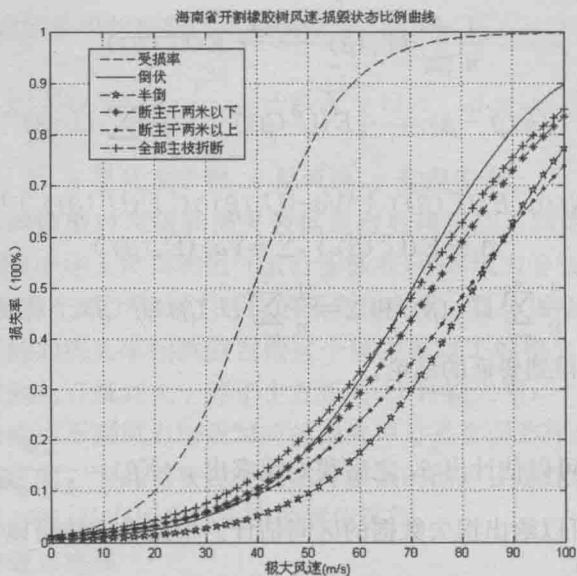


图 1 开割橡胶树模拟风速—损毁状态比例曲线

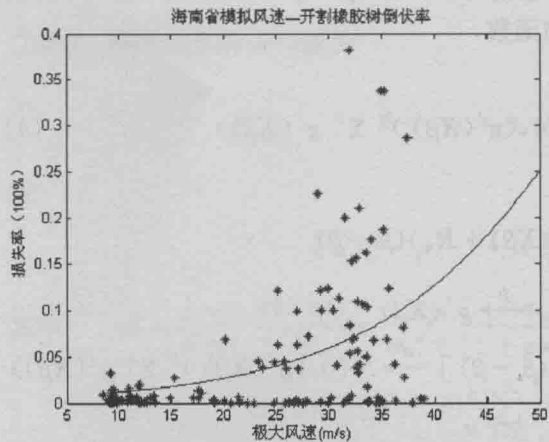


图 2 开割橡胶树模拟风速—倒伏率

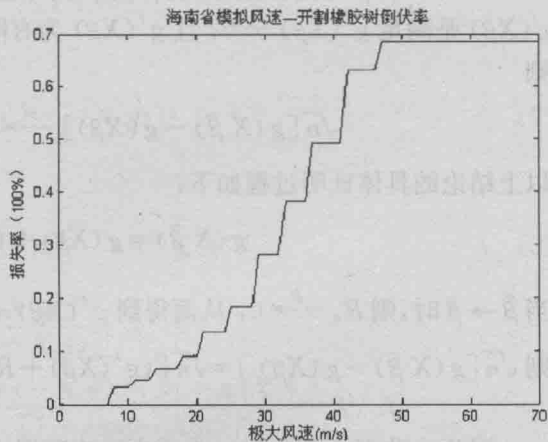


图 3 开割橡胶树模拟风速—倒伏率阶梯函数

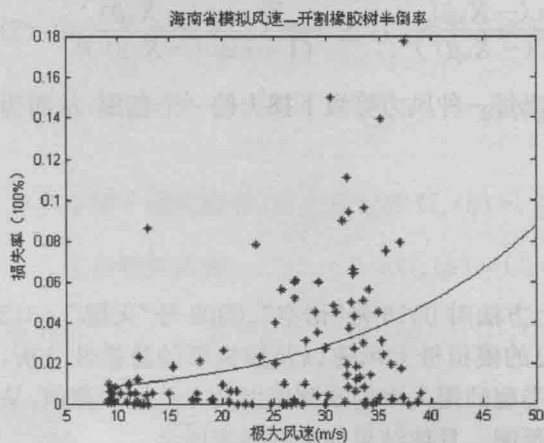


图 4 开割橡胶是模拟风速—半倒率

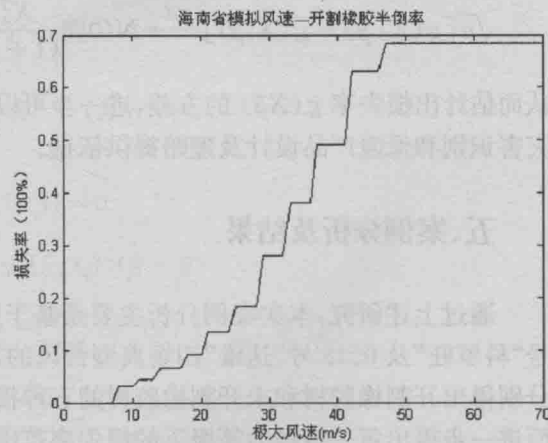


图 5 开割橡胶是模拟风速—半倒率阶梯函数

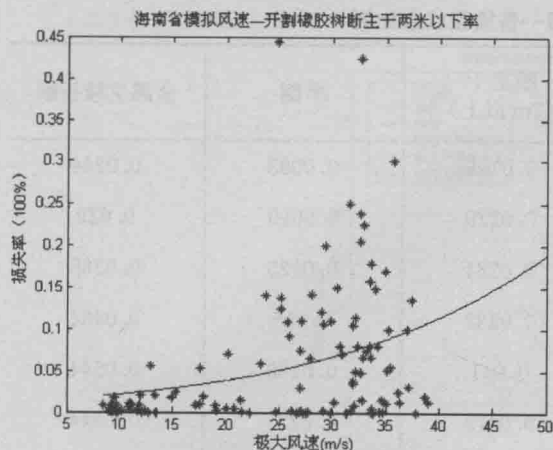


图6 开割橡胶是模拟风速—断主干
两米以下率

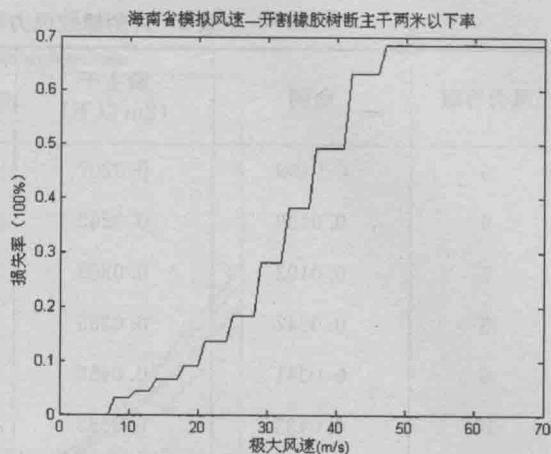


图7 开割橡胶是模拟风速—断主干
两米以下率阶梯函数

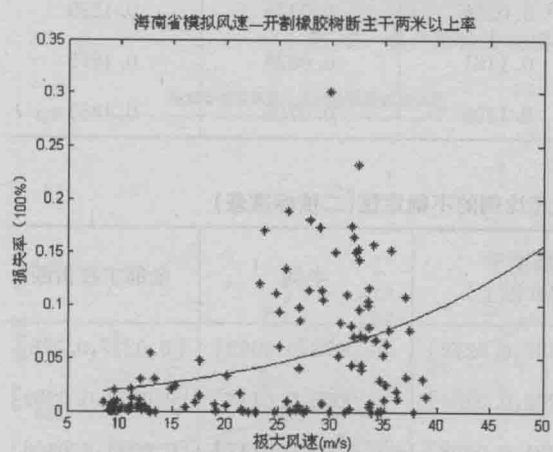


图8 开割橡胶是模拟风速—断主干
两米以上率

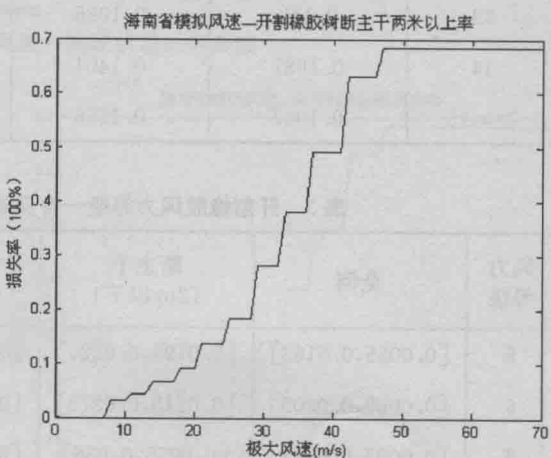


图9 开割橡胶是模拟风速—断主干
两米以上率阶梯函数

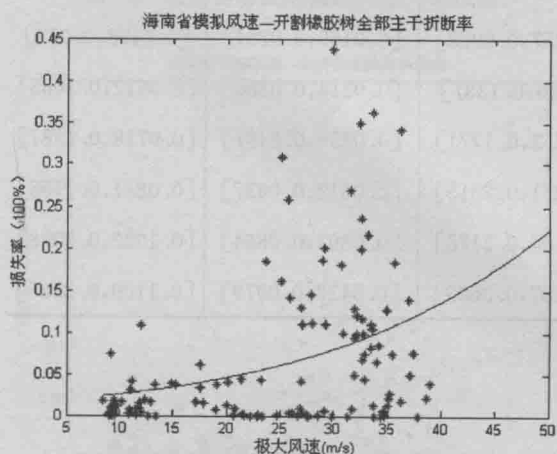


图10 开割橡胶是模拟风速—全部
主枝折断率

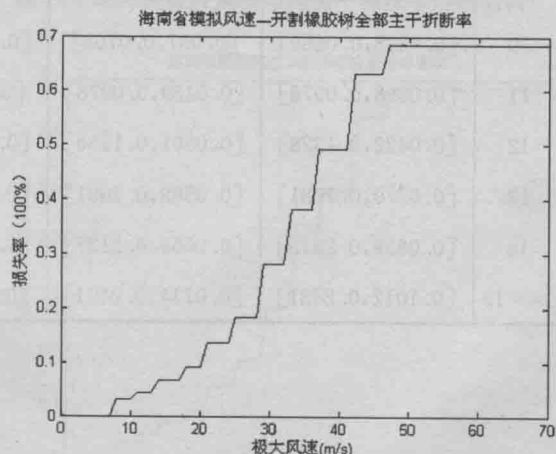


图11 开割橡胶是模拟风速—全部
主枝折断率阶梯函数

表 2 开割橡胶风力等级—各损毁状态比例

风力等级	全倒	断主干 (2m 以下)	断主干 (2m 以上)	半倒	全部主枝折断
5	0.0109	0.0207	0.0196	0.0083	0.0244
6	0.0139	0.0245	0.0229	0.0099	0.029
7	0.0193	0.0308	0.0284	0.0125	0.0365
8	0.0247	0.0365	0.0332	0.015	0.0434
9	0.0341	0.0456	0.041	0.0189	0.0544
10	0.0433	0.0539	0.0479	0.0225	0.0644
11	0.0643	0.0709	0.0619	0.03	0.0848
12	0.0875	0.0879	0.0759	0.0378	0.1053
13	0.118	0.1085	0.0926	0.0475	0.1299
14	0.1687	0.1401	0.1181	0.0628	0.1675
≥15	0.1947	0.1556	0.1305	0.0706	0.1857

表 3 开割橡胶风力等级—各损毁状态比例的不确定性(二倍标准差)

风力等级	全倒	断主干 (2m 以下)	断主干 (2m 以上)	半倒	全部主枝折断
5	[0.0055,0.0163]	[0.0193,0.022,]	[0.0187,0.0212]	[0.0073,0.0093]	[0.0237,0.025]
6	[0.0069,0.0209]	[0.0218,0.0273]	[0.0202,0.0284]	[0.0085,0.0113]	[0.0272,0.0309]
7	[0.0095,0.0291]	[0.0255,0.036]	[0.0221,0.0408]	[0.0104,0.0147]	[0.0324,0.0406]
8	[0.012,0.0373]	[0.0286,0.0443]	[0.0235,0.0528]	[0.012,0.0179]	[0.037,0.0498]
9	[0.0164,0.0518]	[0.0332,0.058]	[0.0249,0.0731]	[0.0146,0.0232]	[0.0439,0.0649]
10	[0.0208,0.0659]	[0.037,0.0708]	[0.0257,0.0923]	[0.0169,0.0281]	[0.0498,0.079]
11	[0.0308,0.0978]	[0.0439,0.0978]	[0.026,0.1337]	[0.0214,0.0386]	[0.0612,0.1085]
12	[0.0422,0.1328]	[0.0501,0.1256]	[0.0252,0.1771]	[0.0259,0.0497]	[0.0718,0.1387]
13	[0.0579,0.1781]	[0.0569,0.1601]	[0.0231,0.2315]	[0.0312,0.0637]	[0.0841,0.1758]
14	[0.0859,0.2515]	[0.0665,0.2137]	[0.0184,0.3175]	[0.0393,0.0864]	[0.1022,0.2328]
≥15	[0.1012,0.2881]	[0.0711,0.2401]	[0.0157,0.3603]	[0.0432,0.0979]	[0.1109,0.2605]