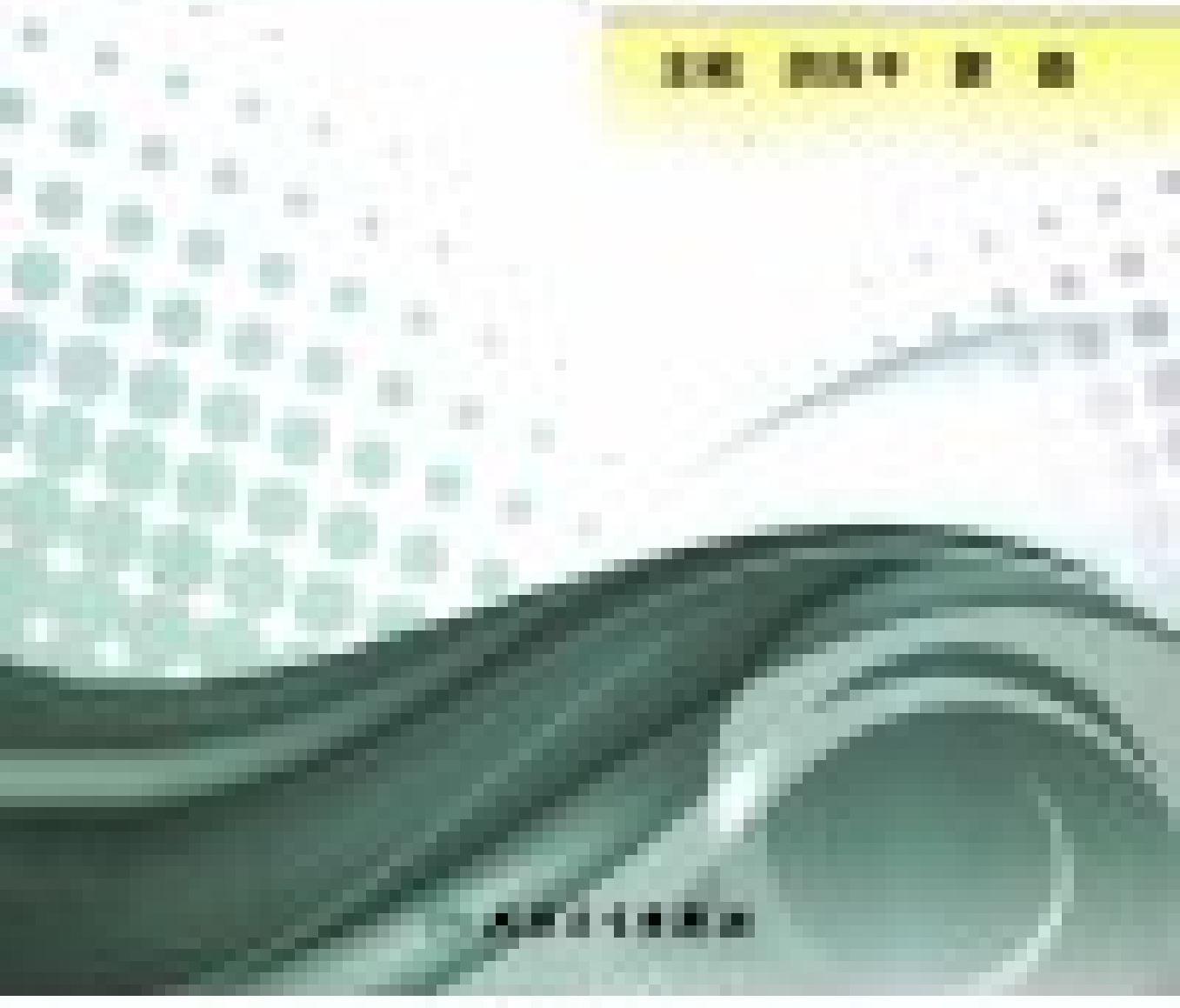
医学遗传统计分析与SAS应用

主编　胡良平　郭　晋

人民卫生出版社

医学统计分析与SAS应用



医学遗传统计分析与SAS应用

主　编　胡良平　郭　晋

编　委（以姓氏笔画为序）

王　琪　军事医学科学院
毛　玮　军事医学科学院
伍亚舟　第三军医大学
刘惠刚　首都医科大学
关　雪　军事医学科学院
李　霞　哈尔滨医科大学
李子建　济南军区疾病预防控制中心
李长平　天津医科大学
李伍举　军事医学科学院
张瑞杰　哈尔滨医科大学
易　东　第三军医大学
周诗国　军事医学科学院
胡良平　军事医学科学院
胡纯严　军事医学科学院
柳伟伟　军事医学科学院
贾元杰　军事医学科学院
高　辉　军事医学科学院
郭　晋　军事医学科学院
陶丽新　军事医学科学院
葛　毅　中国人民解放军后勤指挥学院
鲍晓蕾　军事医学科学院

人民卫生出版社

图书在版编目（CIP）数据

医学遗传统计分析与SAS应用/胡良平等主编. —北京：人民卫生出版社，2011.2
ISBN 978-7-117-13904-5

Ⅰ.①医… Ⅱ.①胡… Ⅲ.①医学遗传学-医学统计-统计分析-应用软件，SAS Ⅳ.①R394-31

中国版本图书馆CIP数据核字（2010）第258374号

医学遗传统计分析与SAS应用

主　　编：胡良平　郭　晋
出版发行：人民卫生出版社（中继线010-59780011）
地　　址：北京市朝阳区潘家园南里19号
邮　　编：100021
E-mail：pmph@pmph.com
购书热线：010-67605754　010-65264830
　　010-59787586　010-59787592
印　　刷：三河市富华印刷包装有限公司
经　　销：新华书店
开　　本：787×1092　1/16　**印张：**17　**插页：**4
字　　数：413千字
版　　次：2011年2月第1版　2011年2月第1版第1次印刷
标准书号：ISBN 978-7-117-13904-5/R·13905
定　　价：36.00元
打击盗版举报电话：010-59787491　E-mail：WQ@pmph.com
（凡属印装质量问题请与本社销售中心联系退换）

内/容/提/要

本书介绍了医学遗传数据的统计分析方法与SAS实现、简明遗传学的基本概念和原理以及医学遗传统计分析的计算原理。详细介绍了如何使用SAS/Genetics模块和其他相关模块实现统计计算、输出结果的解释，并结合具体问题、统计和专业知识，作出令人信服的专业结论。

对医学遗传资料，首先应根据分析目的、资料所具备的前提条件正确判断其数据结构，从而选择分析方法并准确地向SAS/Genetics中录入数据。本书第1章对遗传数据结构做了详细介绍；而后针对拟分析的遗传资料，由简单到复杂的顺序，介绍各种遗传统计分析方法及遗传数据分析的SAS实现：基因、基因型频率的测定、哈代-温伯格平衡定律的验证；连锁不平衡检验、单体型分析；一般人群病例-对照遗传数据分析；遗传结果比较与校正；家系数据分析、连锁分析、芯片数据的分析和物种遗传关系确定分析等。本书最后还提供了SAS/Genetics中全部过程的语法结构和用法简介以及与遗传有关的基本概念和基础知识，供读者参考。

本书能满足基础医学科研人员、生物试验室人员对遗传资料、遗传试验数据分析的需要，可作为高等院校遗传专业研究生工具书，也可作为医学院校、高等院校生物系本科生和研究生参考书，适用于遗传专业、生物医学专业和统计专业学者学习和借鉴。

前言

很多疾病与遗传有关，人们简称其为遗传病。前辈遗传了什么给他们的后代？遗传了细胞中染色体上的某些基因。若前辈属于某些遗传病的患者或其虽属非显性遗传病患者但却携带着遗传病的患病基因，则他们的后代患这些遗传病的概率就很大。经过多年的医学实践和探索，人类已经弄清一小部分遗传病的遗传规律，但还有很多遗传病的遗传规律根本不清楚。然而，随着全世界医学科技事业的蓬勃发展，特别是近年来蛋白质组学、代谢组学、基因组学和后基因组学的迅猛发展，人类对基因有了更全面深入的了解，越来越多的与疾病发生、发展、预后、康复和长寿有关的基因陆续地被发现。

在研究基因与疾病关系的过程中，一系列不可回避的问题被提出来了：如何进行遗传科研课题的研究设计、如何正确地收集医学遗传资料、如何识别遗传资料的数据结构、如何针对不同的遗传资料和分析目的选用相应的遗传资料分析方法、如何方便快捷地使用国际上公认的统计软件实现遗传数据的统计分析、如何结合遗传专业知识对计算结果作出合理的解释和给出科学的专业结论。本书正是在这样一个强大医学需求的背景下应运而生的。

本书分为以下 3 篇，第 1 篇医学遗传统计实例分析与 SAS 实现，包含 10 章内容，第 1 章医学遗传资料数据结构与分析方法选择，提纲挈领地呈现出遗传资料中四种类型的数据结构；第 2 章到第 9 章由浅入深地介绍了遗传资料的实例分析与 SAS 实现；第 10 章 SAS/Genetics 模块概述，较详细地介绍了此模块的功能、语法、过程和语句。第 2 篇简明遗传学基本概念与原理，包含 3 章内容，它们分别为第 11 章遗传学基本概念、第 12 章孟德尔遗传原理与遗传病和第 13 章群体遗传学的基本原理。第 3 篇医学遗传统计分析的计算原理，包含

7章内容，分别为第1篇实例分析中所涉及的各种遗传资料统计分析方法的计算原理。

笔者有幸于2007年招收了一位具有扎实生物学知识基础的硕士研究生，他的名字叫郭晋。在三年的研究生学习生涯中，他刻苦钻研，勇于拼搏，不仅在基础统计学、多因素试验设计、多元统计分析和SAS软件应用等方面取得了长足进步，而且，较为全面地掌握了医学遗传资料统计分析的理论和方法，并能巧妙地运用SAS/Genetics模块实现遗传资料的统计分析和结果解释。本书中绝大部分内容的初稿都是由他完成的。

在本书即将出版之际，笔者真挚地感谢为本书作出过很多贡献的来自哈尔滨医科大学的李霞和张瑞杰教授、第三军医大学的易东教授和军事医学科学院基础医学研究所的李伍举教授等；还要感谢所有为本书付出过辛勤劳动的人们，他们不仅直接和间接地参与了某些章节的编写工作，还认真地为全书作了审阅和校对工作；胡纯严为本书制作了方便用户调用SAS软件的SAS引导程序，即SASPAL软件。正是由于他们的积极参与、不懈努力和真心奉献，才使这部具有明显特色的遗传统计学专著得以问世！

由于笔者水平有限，书中难免会出现这样或那样的不妥，甚至错误之处，恳请广大读者不吝赐教，以便再版时修正。

胡良平

于北京军事医学科学院生物医学统计学咨询中心

2010年12月

目/录

第1篇 医学遗传统计实例分析与SAS实现

第 2 篇 简明遗传学基本概念与原理

第3篇 医学遗传统计分析的计算原理

第1篇

医学遗传统计实例分析与SAS实现

第1章

医学遗传资料数据结构与分析方法选择

在本书的开篇，我们详细地介绍一下常见的几种医学遗传资料的数据结构。数据结构对于认清数据的本质非常重要，同时，标准的数据结构也便于顺利地实现向标准统计软件（如SAS软件）中录入，从而便于后续的数据整理和分析。医学遗传数据大体分为一般人群实验数据、家系数据以及基因芯片数据等，在数据收集过程中，应尽可能地将数据整理或转化为本章所呈现的格式。

1.1 一般人群实验研究的数据结构

一般人群实验研究的数据，主要信息为个体的患病情况、位点基因型等，根据实验设计类型又可细分为下面几类。

1.1.1 位点基因型数据结构

这是一类最简单的数据结构，即将一般人群中个体的基本信息以及研究者所关心的基因或位点的基因型列入表中，成为数据库的形式，便于遗传资料信息的保存和提取。如表1-1所示。

表1-1 可疑影响因素及位点基因型状态的数据结构

受试者编号	性别(M/F)	年龄	A16G	Gln27Glu	T164I	A523C
1	M	5	A A	G G	T T	C C
2	M	6	A A	G G	T T	C C
3	M	6	A A	G G	T T	C C
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	M	44	A G	G G	T T	A A

对数据结构的分析：这种数据结构的录入格式不言自明，因为每行表达了一个受试者有关的全部信息，而每列代表一个变量。第1列为“编号”，仅表达数据所在行的位置；第2列