

全国高等医药院校药学类规划教材

QUANGUO GAODENG YIYAO YUANXIAO

YAOXUELEI GUIHUA JIAOCAI

药学数理 统计方法

(供药学和生物制药专业用)

YAOXUE SHULI
TONGJI FANGFA

主编 刘艳杰

中国医药科技出版社

全国高等医药院校药学类规划教材

药学数理统计方法

(供药学和生物制药专业用)

主 编 刘艳杰

副主编 项荣武 阎心丽

编 委 (按姓氏笔画排序)

王 贺 张晓萍 胡忠盛 姜希伟

党 丹 梁露花

中国医药科技出版社

内 容 提 要

本书是全国高等医药院校药学类规划教材之一,主要针对药学专业的学生学习和掌握数理统计方法并培养学生将数学方法应用到今后的药学研究中。全书分为上下两篇,上篇是统计理论部分,包括概率论初步,随机变量,数据的统计描述,样本及抽样分布,参数估计,参数假设检验,非参数假设检验,方差分析,相关分析与回归分析九章内容,介绍了本科药学专业需要的全部统计理论。下篇是统计实验部分,采用国内通用的 SPSS 统计软件设计了六个统计实验,以此提高学生的应用和实践能力,这也是本书的一大特色之一。

本书可作为高等医药院校药学类专业本科教材,也可作为自学考试辅导教材和医药研究工作者的参考书。

图书在版编目(CIP)数据

药学数理统计方法/刘艳杰主编. —北京:中国医药科技出版社,2013.9

全国高等医药院校药学类规划教材

ISBN 978 - 7 - 5067 - 6297 - 7

I. ①药… II. ①刘… III. ①数理统计 - 应用 - 药理学 - 医学院校 - 教材 IV. ①R311

中国版本图书馆 CIP 数据核字(2013)第 190383 号

美术编辑 陈君杞

出版 中国医药科技出版社

地址 北京市海淀区文慧园北路甲 22 号

邮编 100082

电话 发行:010 - 62227427 邮购:010 - 62236938

网址 www.cmstp.com

规格 787 × 1092mm¹/₁₆

印张 23

字数 457 千字

版次 2013 年 9 月第 1 版

印次 2013 年 9 月第 1 次印刷

印刷 北京印刷一厂

经销 全国各地新华书店

书号 ISBN 978 - 7 - 5067 - 6297 - 7

定价 48.00 元

本社图书如存在印装质量问题请与本社联系调换

全国高等医药院校药学类规划教材常务编委会

名誉主任委员 邵明立 林蕙青

主任委员 吴晓明 (中国药科大学)

副主任委员 (按姓氏笔画排序)

刘俊义 (北京大学药学院)

匡海学 (黑龙江中医药大学)

朱依谆 (复旦大学药学院)

朱家勇 (广东药学院)

毕开顺 (沈阳药科大学)

吴少祯 (中国医药科技出版社)

吴春福 (沈阳药科大学)

张志荣 (四川大学华西药学院)

姚文兵 (中国药科大学)

高思华 (北京中医药大学)

彭 成 (成都中医药大学)

委 员 (按姓氏笔画排序)

王应泉 (中国医药科技出版社)

田景振 (山东中医药大学)

李 高 (华中科技大学同济药学院)

李元建 (中南大学药学院)

李青山 (山西医科大学)

杨 波 (浙江大学药学院)

杨世民 (西安交通大学药学院)

陈思东 (广东药学院)

侯爱君 (复旦大学药学院)

姜红祥 (山东大学)

官 平 (沈阳药科大学)

祝晨蓀 (广州中医药大学)

柴逸峰 (第二军医大学药学院)

黄 园 (四川大学华西药学院)

韩立民 (江西中医学院)

秘 书 夏焕章 (沈阳药科大学)

徐晓媛 (中国药科大学)

王益玲 (广东药学院)

浩云涛 (中国医药科技出版社)

高鹏来 (中国医药科技出版社)

出版说明

全国高等医药院校药学类专业规划教材是目前国内体系最完整、专业覆盖最全面、作者队伍最权威的药学类教材。随着我国药学教育事业的快速发展,药学及相关专业办学规模和水平的不断扩大和提高,课程设置的不断更新,对药学类教材的质量提出了更高的要求。

全国高等医药院校药学类规划教材编写委员会在调查和总结上轮药学类规划教材质量和使用情况的基础上,经过审议和规划,组织中国药科大学、沈阳药科大学、广东药学院、北京大学药学院、复旦大学药学院、四川大学华西药学院、北京中医药大学、西安交通大学医学院、华中科技大学同济药学院、山东大学药学院、山西医科大学药学院、第二军医大学药学院、山东中医药大学、上海中医药大学和江西中医学院等数十所院校的教师共同进行药学类第三轮规划教材的编写修订工作。

药学类第三轮规划教材的编写修订,坚持紧扣药学类专业本科教育培养目标,参考执业药师资格准入标准,强调药学特色鲜明,体现现代医药科技水平,进一步提高教材水平和质量。同时,针对学生自学、复习、考试等需要,紧扣主干教材内容,新编了相应的学习指导与习题集等配套教材。

本套教材由中国医药科技出版社出版,供全国高等医药院校药学类及相关专业使用。其中包括理论课教材 82 种,实验课教材 38 种,配套教材 10 种,其中有 45 种入选普通高等教育“十一五”国家级规划教材。

全国高等医药院校药学类规划教材
编写委员会

2009 年 8 月 1 日

前 言

《药学数理统计方法》作为全国医药类院校的规划教材主要针对药学类各专业的学生。本书以培养数学抽象思维，启迪数学应用能力，结合药学专业特点融入统计软件的应用，提高学生计算和应用统计方法于药学研究领域的能力为宗旨，内容系统全面，例题典型实用，叙述力求深入浅出，简明易懂，注重学生学习兴趣的培养，特别是统计实验部分，更加增强了学生对统计应用的兴趣，从而提高了学生的数学和计算机应用能力。本书的主要特点是：

1. 作为国内首部针对药学类专业学生的数理统计教材，强调夯实基础，重点掌握统计方法，强化实际应用能力，体现学以致用。

2. 强化以计算机应用为基础的统计运算能力，本书采用国内应用广泛的统计软件SPSS设计的数学实验，通过数学实验课程使学生从另一个角度了解数学应用价值，提高其统计应用和计算能力。

3. 本书在知识点讲解和例题选择上紧密结合药学专业特点，以大量的药学应用实例说明统计学在药学领域的广泛应用，习题覆盖面广且类型全面，对学生掌握统计方法并应用于实际工作中有很大的启迪作用。

4. 配合大学生数学建模活动，引导学生掌握药学领域数学建模方法，促进数学与药学相互渗透和结合，提高学生的数学修养和创新实践能力。

本书可作为高等医药院校药学类专业本科教材，也可作为自学考试辅导教材和医药研究工作者的参考书。

本书编者来自沈阳药科大学多年从事药学基础数学教学的一线教师，他们将自己多年来的教学经验和体会凝聚在编写过程中，付出了大量的心血，在此对各位老师的努力表示衷心的感谢！同时，感谢中国医药科技出版社的大力支持和帮助！

由于时间紧迫，书中不免有一些错误和不妥之处恳请广大读者提出批评指正！

编者
2013年6月



上篇 统计理论

绪论	(2)
一、什么是统计	(2)
二、统计学的应用	(2)
三、统计软件简介	(4)
第一章 概率论初步	(7)
第一节 随机事件	(7)
一、样本空间与随机事件	(7)
二、事件之间的关系及其运算	(8)
第二节 随机事件的概率	(10)
一、概率的定义	(10)
二、概率的性质	(12)
三、古典概型	(13)
四、几何概型	(13)
第三节 条件概率与乘法公式	(14)
一、条件概率的引例	(14)
二、乘法定理	(16)
第四节 全概率公式和贝叶斯公式	(17)
第五节 事件的独立性	(19)
一、引言	(19)
二、贝努里 (Bernoulli) 试验	(21)
习题一	(22)
第二章 随机变量	(25)
第一节 随机变量及其分布	(25)

一、随机变量	(25)
二、分布函数	(26)
第二节 离散型随机变量及其分布	(27)
一、离散型随机变量及其分布律	(27)
二、常见的离散型随机变量的概率分布	(27)
第三节 连续型随机变量及其分布	(32)
第四节 随机变量函数的分布	(37)
一、随机变量函数	(37)
二、离散型随机变量函数的分布	(37)
三、连续型随机变量函数的分布	(38)
第五节 多维随机变量及其分布	(40)
一、二维随机向量及其分布	(40)
二、边缘分布	(44)
三、条件分布	(46)
四、随机变量的独立性	(48)
五、两个随机变量的函数的分布	(49)
第六节 随机变量的数字特征	(50)
一、随机变量的数学期望	(51)
二、随机变量的方差	(56)
三、协方差与相关系数	(59)
四、矩、协方差矩阵	(62)
第七节 大数定律与中心极限定理	(63)
一、大数定律	(63)
二、中心极限定理	(66)
习题二	(68)

第三章 数据的统计描述

(72)

第一节 数据类型	(72)
一、变量与数据	(72)
二、数据类型	(72)
第二节 集中趋势的测度	(73)
一、数值平均数	(74)
二、位置平均数	(78)
第三节 离散程度的测度	(83)
一、离散程度的绝对指标	(83)
二、离散程度的相对指标	(86)
三、数据的标准化	(86)
四、是非标志标准差	(87)

五、原点矩与中心矩	(88)
第四节 分布偏态与峰度的测度	(89)
一、分布偏态的测度	(89)
二、分布峰度的测度	(90)
第五节 统计表	(91)
一、统计表的结构	(91)
二、统计表的分类	(91)
三、统计表设计规则	(92)
第六节 统计图	(93)
一、统计图的结构	(93)
二、统计图的种类	(94)
三、直方图	(94)
四、条形图	(96)
五、饼图	(98)
习题三	(99)
第四章 样本及抽样分布	(101)
第一节 数理统计基本概念	(102)
一、总体与个体	(102)
二、样本和简单随机样本	(103)
三、总体、样本、样本值的关系	(104)
第二节 随机抽样方法	(104)
第三节 分布函数与概率密度函数的近似解	(107)
一、经验分布函数	(107)
二、利用直方图求密度函数的近似解	(108)
第四节 统计量和抽样分布	(109)
一、统计量的概念	(109)
二、抽样分布	(111)
三、分位数与分位点	(115)
习题四	(116)
第五章 参数估计	(119)
第一节 参数估计的概念	(119)
第二节 点估计量的求法	(120)
一、数的矩估计法 (K. Pearson 提出)	(120)
二、极大似然估计法 (R. A. Fisher 提出)	(121)
第三节 估计量的评判标准	(124)
一、无偏性	(124)

二、有效性	(126)
三、相合性 (一致性)	(126)
第四节 区间估计	(126)
一、单个正态总体期望与方差的区间估计	(128)
二、两个正态总体的区间估计	(129)
习题五	(131)
第六章 参数假设检验	(133)
第一节 假设检验基本原理	(133)
一、小概率事件原理	(133)
二、假设检验的步骤	(135)
三、假设检验的两类错误	(135)
四、双侧检验与单侧检验	(136)
五、假设检验应注意的问题	(137)
第二节 单个正态总体的假设检验	(137)
一、单个正态总体均值检验	(137)
二、单个正态总体方差的 χ^2 检验	(140)
第三节 两个正态总体的假设检验	(141)
一、两个正态总体均值的假设检验	(141)
二、两个正态总体方差的假设检验	(145)
第四节 样本频率的假设检验	(146)
一、单个样本频率的假设检验	(147)
二、两个样本频率的假设显著性检验	(148)
习题六	(149)
第七章 非参数假设检验	(152)
第一节 拟合优度检验	(152)
第二节 列联表的 χ^2 检验	(155)
第三节 秩和检验	(158)
习题七	(163)
第八章 方差分析	(166)
第一节 单因素方差分析	(166)
一、单因素方差分析模型	(166)
二、方差分析的基本原理	(167)
三、方差分析的计算与分析	(169)
第二节 两两间多重比较的检验方法	(172)
一、T 方法	(172)

二、S 方法	(173)
第三节 双因素方差分析	(174)
一、无重复试验的方差分析	(174)
二、双因素有重复试验的方差分析	(178)
第四节 多因素正交试验设计与方差分析	(181)
一、正交试验设计与正交表	(181)
二、正交试验结果的分析	(183)
三、多指标正交试验的分析方法	(190)
习题八	(192)
第九章 相关分析与回归分析	(196)
第一节 相关分析	(197)
一、散点图	(197)
二、样本相关系数	(199)
三、相关系数的显著性检验	(200)
四、等级相关分析	(202)
第二节 一元线性回归分析	(203)
一、一元线性回归模型	(204)
二、回归系数的最小二乘估计	(204)
三、回归方程的显著性检验	(206)
四、相关分析和回归分析时的注意事项	(210)
五、用回归方程进行预测和控制	(211)
第三节 一元拟线性回归分析	(213)
一、确定可能的函数形式	(214)
二、参数估计	(214)
三、曲线回归方程的比较	(215)
第四节 多元线性回归分析	(217)
一、多元线性回归方程的建立	(218)
二、多元线性回归方程的显著性检验	(220)
第五节 多项式回归	(222)
一、多项式回归分析的一般方法	(222)
二、一元二次多项式回归分析	(223)
第六节 Logistic 回归模型	(225)
一、模型结构	(226)
二、Logistic 回归模型的参数估计与假设检验	(228)
第七节 几种常见的 Logistic 回归模型	(229)
一、二项分类变量资料非条件 Logistic 回归	(229)
二、多分类结果变量的 Logistic 回归	(232)

三、配比设计条件的 Logistic 回归模型	(236)
习题九	(239)

下篇 统计实验

SPSS 18.0 软件系统介绍	(244)
实验 1 数据文件管理	(247)
一、试验目的与要求	(247)
二、试验原理	(247)
三、试验内容与步骤	(247)
四、课上练习	(254)
五、课后练习	(258)
实验 2 统计描述	(259)
一、频数分析 (Frequencies)	(260)
二、描述统计 (Descriptives)	(263)
三、探索分析 (Explore)	(266)
四、课上练习	(270)
实验 3 T 检验	(271)
一、独立样本 T 检验	(271)
二、配对样本 T 检验	(273)
三、单样本 T 检验	(275)
四、课上练习	(276)
五、课后练习	(276)
实验 4 列联表独立性检验	(277)
一、数据输入	(277)
二、课上练习	(283)
实验 5 方差分析	(283)
一、单因素方差分析	(283)
二、无重复观察值的两因素方差分析	(289)
三、有重复观察值的两因素方差分析	(293)
四、课上练习	(297)
实验 6 相关分析和回归分析	(298)
一、相关性统计分析的 SPSS 操作	(298)
二、一元线性回归分析	(301)
三、多元线性回归分析	(305)
四、曲线回归分析	(309)
五、课上练习	(312)
六、课后练习	(313)

习题参考答案	(314)
附录	(323)
附表 1 标准正态分布表	(323)
附表 2 标准正态分布双侧临界值表	(325)
附表 3 χ^2 分布表	(326)
附表 4 t 分布表	(328)
附表 5 F 分布表	(330)
附表 6 配对比较符号秩和检验用 T 界值表	(334)
附表 7 两总体比较秩和检验用 T 界值表	(335)
附表 8 三总体比较秩和检验用 H 界值表	(336)
附表 9 多重比较中的 q 表	(337)
附表 10 多重比较中的 S 表	(339)
附表 11 检验相关系数 $\rho=0$ 的临界值表	(341)
附表 12 等级相关系数的临界值表	(342)
附表 13 正交表	(343)

上篇

SHANG PIAN



统计理论

绪 论

南丁格尔曾说过“若想了解上帝在想什么，我们就必须学统计，因为统计学就是在量测他的旨意。”

这里的上帝就是客观世界。

统计学已经广泛应用于现实生活的各个领域之中，并在解决现代科学的那些最重要的和最多样化的课题中起着主导作用。现实生活中也离不开统计，每天晚上我们都要收听明天的天气预报，尤其注意明天下雨或下雪的概率；在投资股票时，你需要了解股票场价格的信息，了解每只股票的财务信息；在黄金周，你打算出门旅行时，一定要了解旅游的目的地价格、服务以及旅游人数等。作为从事统计工作的专业人士，更需要了解和掌握统计学的基本理论和方法。

一、什么是统计

统计学 (statistics) 是通过收集数据，分析数据和由数据得出结论的一组概念、原则和方法。统计分析数据的方法大体上可分为描述统计 (descriptive statistics) 和推断统计 (inferential statistics) 两大类。

描述统计是研究数据收集、处理和描述的统计学方法。其内容包括如何取得研究所需要的数据，如何用图表形式对数据进行处理，如何通过对数据的综合、概括与分析，得出所关心的数据特征。

推断统计则是研究如何利用样本数据来推断总体特征的统计学方法，主要方法是参数估计和假设检验。

二、统计学的应用

(一) 统计学在现实生活中的应用实例

说出哪些领域应用统计，这很困难，因为几乎所有的领域都应用统计；说出哪些领域不使用统计，同样也很困难，因为几乎找不到一个不用统计的领域。因此，统计是适用于所有学科领域的通用的数据分析方法，是一种通用的数据分析语言。这里我们举几个著名的统计实例。

应用案例一：从“女士品茶”中得到的统计实验设计

20 世纪 20 年代后期的一个夏日午后，一群风度翩翩的学者偕夫人及漂亮的女友，正在英国剑桥的户外餐桌旁，悠闲地品茶论道。席间，一位美丽的女士惊呼，午茶的调制顺序对味道有很大的影响。把茶加进牛奶里和把牛奶加进茶里，喝起来风味完全不同。出于对女性的尊重，那些学者们面带绅士的微笑，内心却不以为然，甚至是藐视，依据他们的科学头脑分析，茶和牛奶两种物质混合结果的化学成分不会因为调制顺序不

同而产生不同，怎么会喝起来不一样呢？文中暗表，这个命题的假设前提是不论调制顺序如何，牛奶和茶的比例是固定的或是基本不变的。正当众学者对美丽女士的说法嗤之以鼻时，有个身材瘦小，嘴上留着灰白胡子的绅士挺身而出，抓住了这个问题。此人便是在统计发展史上地位显赫、大名鼎鼎的费雪（Ronald Aylmer Fisher, 1890 ~ 1962），英国统计学家。费雪当时显得非常兴奋，好像发现了新大陆。“让我们来检验这个命题。”在众位学者的帮助下，他开始进行实验。他们设计并调制出很多杯不同的茶，有些先放茶水再加牛奶，有些先放牛奶再加茶水，然后按照既定的顺序一杯一杯拿给美丽女士品尝分辨，但她并不知道每杯茶的调法。费雪端给她第一杯茶时她品尝了一口，然后说出这杯茶是先放茶水后加的牛奶，还是先放牛奶后加的茶水。费雪记录上她的说法，再送上第二杯，……费雪设计了各种可能的实验方法，来测试美丽女士能否分辨出不同的茶。问题是，如果美丽女士只是哗众取宠而没有真本领能分辨出不同的茶，她还是有猜出的可能；如果给她两杯调制方法不同的茶，她可能一次全部猜错或全部猜对。如果美丽女士有真本事，确实能够分辨调制方法不同的茶，但她还是可能弄错，或是茶水和牛奶没有混合好，或茶水温度不够影响了味道，或她喝了很多以后感觉已经不太灵敏。这就是费雪提出来的实验设计思想，1935年，费雪完成了在科学实验理论和方法上具有划时代意义的一本书《实验设计》。在书的第二章，费雪就提到了剑桥午后的品茶和那位美丽的女士，在书里，费雪讨论了各种可能结果，描述了该准备多少茶，依照什么顺序拿给她，然后她回答的正确与否，计算出各种结果的概率。至于剑桥午后品茶的那位女士，据说她能分辨出每一杯茶，全部答对，看来，这位女士不仅仅是美丽。对常人来说这只是一个有趣的故事，而对科学家来说这却是一个理论的诞生。

应用案例二：从《红楼梦》中给出作者的判断

众所周知，《红楼梦》一书共120回，一般认为前80回是曹雪芹所著，后40回为高鄂所续。长期以来红学界对这个问题一直有争议。1986年复旦大学李贤平教授带领他的学生用统计方法进行了研究，他们将120回看成是120个样本，然后确定与情节无关的虚词作为变量（所以要抛开情节，是因为在一般情况下，同一情节大家描述的都差不多，但由于个人写作特点和习惯不同，所用的虚词是不会一样的），让学生数出每一回里虚词出现的次数，作为数据，用统计分析中的聚类分析方法进行分类。聚类结果将120回分成两类：即前80回为一类，后40回为一类，很形象地证实了120回的《红楼梦》不是出自同一人的手笔。之后又与曹雪芹的其他著作进行类似分析，进一步证实前80回确实为曹雪芹所著，而后40回是否为高鄂写的呢？论证结果推翻了后40回为高鄂一个人所写。这个论证在经学界轰动很大，他们用统计分析方法支持了红学界的观点，使红学界为之赞叹。

（三）统计学在医药领域中的应用

医、药学是较早使用数理统计方法的领域之一。在医学研究中，为防治一种疾病，统计方法常被作为重要的研究工具，用来发现和验证导致这种疾病的种种因素。例如，应用统计方法证实肺癌与吸烟的关系。在药学研究中，通过临床试验，应用正交设计、交叉设计、回归分析、方差分析、列联表分析等统计方法，来确定一种药物对治疗某种疾病是否有效，以及比较几种药物或治疗方案的效力。下面列举几个在药学中应用

实例:

案例一:用正交试验法研究三七提取工艺,试验的考察指标为测定每个样品中的三七总皂苷量。在对溶媒和提取方法初步选择以后,选定乙醇浓度、提取方法和乙醇用量作为考察的3个因素,每个因素各取3个水平。数据进行方差分析,得最佳工艺为采用渗漉法、75%乙醇收集10倍量体积。通过科学的数据处理,使三七提取工艺规范化,数据化。

案例二:在一种片剂批的含量均匀度检查的初试中,随机地取出10片,逐片进行含量测定,得到10个数据,判断该厂这种片剂含量均匀度是否符合要求?

案例三:在临床前试验中,为确定一种新化合物的致癌性将其按不同剂量给两个动物组服用:一个是空白对照组,把空白用于对照组的65个动物,有6个引发癌症(9.2%);另一组服用新化合物,把一定剂量用于该组的60个动物,结果有9个引发癌变(15%),已知该新化合物不会降低致癌率,在单侧检验中,为确定两组动物的致癌率在 $\alpha=0.05$ 水平是否有差异?

三、统计软件简介

(一) SAS 统计软件

SAS 是英文 Statistical Analysis System 的缩写,翻译成汉语是统计分析系统,最初由美国北卡罗来纳州立大学两名研究生开始研制,1976年创立 SAS 公司。SAS 系统具有十分完备的数据访问、数据管理、数据分析功能。在国际上,SAS 被誉为数据统计分析的标准软件。SAS 系统是一个模块组合式结构的软件系统,共有三十多个功能模块。SAS 是用汇编语言编写而成的,通常使用 SAS 需要编写程序,比较适合统计专业人员使用,而对于非统计专业人员学习 SAS 比较困难。SAS 最新版为 9.0 版。网址: <http://www.sas.com/>。

(二) SPSS 统计软件

SPSS 是英文 Statistical Package for the Social Science 的缩写,翻译成汉语是社会统计学程序包,20 世纪 60 年代末由美国斯坦福大学的三位研究生研制,1975 年在芝加哥组建 SPSS 总部。SPSS 系统特点是操作比较方便,统计方法比较齐全,绘制图形、表格较有方便,输出结果比较直观。SPSS 是用 FORTRAN 语言编写而成。适合进行从事社会学调查中的数据分析处理。最新版为 13.0 版。网址: <http://www.spss.com/>。

(三) BMDP 统计软件

BMDP 是英文 Biomedical computer programs 的缩写,翻译成汉语是生物医学计算程序,美国加州大学于 1961 年研制,是世界上最早的统计分析软件。特点是统计方法齐全,功能强大。但 1991 年的 7.0 版后没有新的版本推出,使用不太普及,最后被 SPSS 公司收购。

(四) Stata 统计软件

Stata 统计软件由美国计算机资源中心 (Computer Resource Center) 1985 年研制。特点是采用命令操作,程序容量较小,统计分析方法较齐全,计算结果的输出形式简洁,绘出的图形精美。不足之处是数据的兼容性差,占内存空间较大,数据管理功能需要加