



普通高等教育“十二五”规划教材

常用统计软件

在生命科学中的应用



主编 郭春华



科学出版社

普通高等教育“十二五”规划教材

常用统计软件在生命科学中的应用

主 编 郭春华

副主编 施晓丽 张红平 徐宁迎

科学出版社

北 京

内 容 简 介

本教材将Excel、SPSS和SAS三个统计软件中与生物统计相关的内容汇集于一体,结合生命科学研究中的常见案例,着重介绍这三个软件的具体操作步骤和结果解释,旨在培养学生解决生物统计问题的实际动手能力。全书共二十章,内容包括三篇。第一篇主要介绍Excel在数据的整理、概率分布、 t 检验、方差分析、 χ^2 检验、相关与回归分析、规划求解与饲料配方中的应用;第二篇主要介绍SPSS在 t 检验、方差分析、 χ^2 检验、相关分析和回归分析中的应用;第三篇主要介绍SAS在 t 检验、方差分析、 χ^2 检验、相关分析、回归分析、主成分分析、聚类分析和重复测量数据分析中的应用。

本教材简单易懂,内容丰富,结构合理,以大量生命科学中经典案例作为例题,每章后面附有相应的练习题。本教材内容涵盖了生命科学中的常见问题,可作为动物科学、动物医学、农学、食品科学、水产养殖、草业科学、生物技术等专业本科生和研究生的教材、参考资料,也可供从事生命科学研究的专业人员学习和参考。

图书在版编目(CIP)数据

常用统计软件在生命科学中的应用 / 郭春华主编. —北京:科学出版社, 2011

ISBN 978-7-03-030107-9

I. 常… II. 郭… III. 统计分析—应用软件—应用—生命科学
IV. Q1-0

中国版本图书馆CIP数据核字(2011)第013395号

责任编辑:相 凌/ 责任校对:朱光兰

责任印制:张克忠/ 封面设计:陈四雄

科学出版社出版

北京东黄城根北街16号

邮政编码100717

<http://www.sciencep.com>

明辉印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2011年2月第 一 版 开本:(787×1092) 1/16

2011年2月第一次印刷 印张:18

印数:1—4 000 字数:410 000

定价:42.00元

(如有印装质量问题,我社负责调换)

本书编委会

主 编 郭春华 西南民族大学

副主编 施晓丽 贵州大学

张红平 四川农业大学

徐宁迎 浙江大学

编 者 (以下按姓氏拼音字母排序)

蔡惠芬 贵州大学

陈炼红 西南民族大学

戴国俊 扬州大学

高鹏飞 山西农业大学

何 俊 湖南农业大学

胡建宏 西北农林科技大学

黄艳玲 西南民族大学

孙桂荣 河南农业大学

王起山 上海交通大学

王雅春 中国农业大学

向 钊 西南大学

曾 涛 四川大学

张 豪 华南农业大学

审 稿 潘玉春 上海交通大学

前 言

“生物统计学”通常是动物科学、动物医学、农学、食品科学、水产养殖、草业科学、生物技术等专业学生的必修课，这是一门既重要又实用的专业基础课，而对于大多数学生来说又是一门比较难学的课程。尽管统计分析是一件十分枯燥而又繁琐的事情，但电脑的普及和统计分析软件的不断完善和智能化，又使数据统计分析变得十分容易，当然前提是至少掌握了一款统计分析软件的使用方法。由于“生物统计学”的课程设置在大多数高校都在 50 学时左右，该课程的主要任务是介绍生物统计和试验设计的基本概念、基本公式和基本方法，没有时间介绍统计软件在生物统计中的应用。因此，学生在学完该课程或类似课程之后，往往不知如何应用统计分析软件来对试验结果进行统计分析，即使在进入硕士或博士阶段的学习也是如此。

利用统计软件对大量试验数据进行统计分析是每一位科技工作者必须掌握的一项技能。常用的统计分析软件有 Excel、SPSS 和 SAS 等，它们各有优缺点，在实际中经常将这三个软件配合使用，能非常方便地解决生命科学研究中的统计问题。时下市场上有关这三个统计软件的书籍名目繁多，但是每本书都包括很多经济、管理、社会科学方面的内容，而没有一本专门针对生命科学有关专业学生的将这三个软件汇编在一起的书籍。本书的最大特点就是 Excel、SPSS 和 SAS 三个统计软件中与生物统计相关的内容汇编在一起，对生命科学相关专业的学生具有更强的针对性、实用性和可操作性。

本书的编写力求以生命科学教学、科研和生产中常见问题为主线，针对其所面临的实际问题，利用 Excel、SPSS 和 SAS 统计软件的强大功能，使广大学生能够快速而准确地掌握这三个软件的使用，解决统计分析中所面临的问题。

本书的内容包括三篇。第一篇主要介绍 Excel 在数据的整理、显著性检验、相关与回归分析、通径分析和饲料配方中的应用；第二篇主要介绍 SPSS 在显著性检验、相关与回归分析、通径分析中的应用；第三篇主要介绍 SAS 在显著性检验、相关与回归分析、通径分析、聚类分析、主成分分析和重复测量数据分析中的应用。

本书的写作特点是：改变一般软件使用说明的写作模式，按“简介—实例—操作步骤—结果解释”的顺序来安排结构，书中的例题和练习题大多选自作者长期在教学、科研中的试验案例，使学生能够在理解统计分析方法的基础上，能熟练地使用这三个软件，读懂软件的输出结果，并对输出结果做出合理的解释，最终能利用书中所述方法和步骤解决实际问题。

本书第一篇（Excel，第一章至第七章）由黄艳玲、郭春华、陈炼红、蔡惠芬、胡建宏和王起山编写，第二篇（SPSS，第八章至第十三章）由施晓丽、向钊、孙桂荣、曾涛、高鹏飞编写，第三篇（SAS，第十四章至第二十章）由张红平、戴国俊、王雅春、何俊、张豪、徐宁迎、郭春华编写，其他参加资料整理的有柏雪、刘港彪。本书由上海交通大学潘玉春教授审稿。在此一并表示感谢！

由于水平有限，书中难免有不当之处，敬请读者批评指正。

编 者

2010 年 9 月 6 日

目 录

前言

第一篇 Excel 在生命科学中的应用

第一章 Excel 2007 应用基础	2
第一节 认识 Excel 2007 工作界面	2
一、启动 Excel	2
二、Excel 2007 工作界面概况	2
第二节 数据的输入与编辑	3
一、数据的输入	3
二、数据的填充	6
三、数据的移动、复制和行列转置	7
第三节 公式与函数的使用	9
一、公式的使用	9
二、函数的使用	9
三、单元格引用	10
第四节 数据管理与分析	11
一、数据排序	11
二、数据筛选	11
三、分类汇总	13
四、数据透视表	14
第五节 Excel 绘图基础	14
一、认识图表	14
二、创建图表	15
三、编辑图表	15
第六节 Excel 统计分析功能概述	19
一、查找和了解统计函数	19
二、在 Excel 中加载数据分析模块	19
三、统计分析工具	20
习题	22
第二章 常见概率分布	23
第一节 常用理论分布	23
一、二项分布	23
二、泊松分布	24



三、正态分布	25
第二节 常用统计分布	27
一、 t 分布	27
二、 F 分布	28
三、 χ^2 分布	29
习题	30
第三章 t 检验	32
第一节 样本平均数与总体平均数的差异显著性检验	32
一、简介	32
二、实例	32
三、操作步骤	32
四、结果解释	33
第二节 非配对设计两样本平均数差异显著性检验	33
一、简介	33
二、实例	33
三、操作步骤	34
四、结果解释	35
第三节 配对设计两样本平均数差异显著性检验	35
一、简介	35
二、实例	35
三、操作步骤	36
四、结果解释	37
习题	37
第四章 方差分析	38
第一节 单因素试验资料的方差分析	38
一、简介	38
二、实例	38
三、操作步骤	39
四、结果解释	40
第二节 二因素试验资料的方差分析	41
一、二因素无重复观察值试验资料的方差分析	41
二、二因素有重复观察值试验资料的方差分析	42
习题	44
第五章 χ^2 检验	46
第一节 适合性检验	46
一、简介	46
二、实例	46
三、操作步骤	47
四、结果解释	48

第二节 独立性检验	48
一、简介	48
二、实例	48
三、操作步骤	49
四、结果解释	50
习题	50
第六章 相关与回归分析	51
第一节 相关分析	51
一、简介	51
二、实例	51
三、操作步骤	51
四、结果解释	52
第二节 回归分析	52
一、一元线性回归	53
二、多元线性回归	56
三、曲线回归	58
第三节 通径分析	59
一、简介	59
二、实例	59
三、操作步骤	59
四、结果解释	63
习题	64
第七章 规划求解与饲料配方	65
第一节 利用规划求解优化饲料配方	65
一、利用线性规划求解饲料配方的基本原理	65
二、实例	66
三、操作步骤	66
第二节 试差法求解饲料配方	69
一、简介	69
二、实例	70
三、操作步骤	70
习题	71

第二篇 SPSS 在生命科学中的应用

第八章 SPSS 简介	74
第一节 SPSS 简介	74
一、SPSS 的发展和版本	74
二、PASW statistics 18.0 的安装、启动和退出	75
三、SPSS 主要界面	76



四、SPSS 的帮助系统	78
第二节 建立数据文件与读取数据	79
一、建立新数据文件	79
二、读取数据	85
第三节 数据管理	87
一、数据整理	87
二、数据转换	91
习题	93
第九章 t 检验	95
第一节 样本平均数与总体平均数的差异显著性检验	95
一、简介	95
二、实例	95
三、操作步骤	95
四、结果解释	97
第二节 两样本平均数差异显著性检验	97
一、两非配对样本平均数差异显著性检验	97
二、两配对样本平均数差异显著性检验	100
习题	102
第十章 方差分析	103
第一节 单因素方差分析	103
一、简介	103
二、实例	103
三、操作步骤	103
四、结果解释	106
第二节 二因素方差分析	109
一、二因素无重复方差分析	109
二、二因素有重复交叉分组方差分析	112
第三节 正交试验分析	115
一、简介	115
二、实例	115
三、操作步骤	116
四、结果解释	116
第四节 裂区试验分析	117
一、简介	117
二、实例	117
三、操作步骤	117
四、结果解释	119
第五节 协方差分析过程	120
一、简介	120

二、实例·····	120
三、操作步骤·····	121
四、结果解释·····	123
习题·····	124
第十一章 χ^2 检验·····	127
第一节 适合性检验·····	127
一、简介·····	127
二、实例·····	127
三、操作步骤·····	127
四、结果解释·····	129
第二节 独立性检验·····	130
一、简介·····	130
二、实例·····	130
三、操作步骤·····	130
四、结果解释·····	131
习题·····	133
第十二章 相关分析·····	134
第一节 简单相关分析·····	134
一、实例·····	134
二、操作步骤·····	134
三、结果解释·····	135
第二节 偏相关分析·····	135
一、实例·····	135
二、操作步骤·····	135
三、结果解释·····	136
习题·····	136
第十三章 回归分析·····	138
第一节 线性回归分析·····	138
一、一元线性回归·····	138
二、多元线性回归分析·····	141
第二节 曲线估计·····	145
一、简介·····	145
二、实例·····	145
三、操作步骤·····	145
四、结果解释·····	146
第三节 非线性回归分析·····	147
一、简介·····	147
二、实例·····	147
三、操作步骤·····	147



四、结果解释	149
第四节 通径分析	150
一、简介	150
二、实例	151
三、操作步骤及结果解释	151
习题	156

第三篇 SAS 在生命科学中的应用

第十四章 SAS 概述	158
第一节 SAS 系统介绍	158
第二节 SAS 系统的启动与关闭	159
一、启动	159
二、退出	160
第三节 SAS 程序的结构	161
一、SAS 程序的结构	161
二、程序的输入、修改调试和运行	162
第四节 SAS 语言与数据管理	162
一、SAS 系统对数据的管理	162
二、数据的三种输入方式	163
习题	167
第十五章 t 检验	168
第一节 样本平均数与总体平均数差异显著性检验	168
一、简介	168
二、实例与程序	168
三、结果	169
第二节 配对设计两样本平均数差异显著性检验	169
一、简介	169
二、实例与程序	170
三、结果	171
第三节 非配对设计两样本平均数差异显著性检验	171
一、简介	171
二、实例与程序	171
三、结果	173
四、特别提示	175
习题	176
第十六章 方差分析	177
第一节 单因素试验资料的方差分析	177
一、简介	177
二、ANOVA 和 GLM 过程的应用	177

第二节 多因素方差试验资料的分析	182
一、二因素交叉分组无重复观测值试验资料的方差分析	182
二、二因素交叉分组有重复观测值试验资料的方差分析	184
三、二因素系统分组资料的方差分析	186
第三节 正交试验设计资料的方差分析	189
一、无重复观测值的正交试验设计资料的方差分析	189
二、有重复观测值的正交试验设计资料的方差分析	191
第四节 拉丁方试验设计资料的方差分析	192
一、实例及 SAS 程序	192
二、输出结果及解释	193
第五节 F 检验时误差效应的确定	193
第六节 协方差分析	194
一、简介	194
二、实例及 SAS 程序	194
三、输出结果及解释	194
习题	196
第十七章 χ^2 检验	200
第一节 适合性检验	200
一、实例与程序	200
二、输出结果及解释	201
第二节 独立性检验	202
一、实例与程序	202
二、输出结果及解释	203
三、特别提示	205
习题	205
第十八章 相关分析	206
第一节 简单相关分析	206
一、简介	206
二、实例与程序	206
三、结果输出及解释	206
第二节 偏相关分析	207
一、简介	207
二、实例与程序	207
三、结果输出及解释	208
习题	210
第十九章 回归分析	212
第一节 一元线性回归分析	212
一、一元线性回归	212
二、可化为线性回归的曲线回归	214



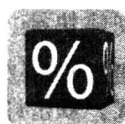
第二节 多重线性回归	215
一、多重回归分析	215
二、利用 PROC GLM 进行多重回归分析	221
第三节 通径分析	223
第四节 曲线回归	223
一、多项式回归	223
二、分段回归	224
第五节 非线性模型回归	227
一、简介	227
二、实例及 SAS 程序	227
三、SAS 输出结果及解释	228
第六节 多元线性回归	229
一、简介	229
二、SAS 程序	229
三、变量的筛选	231
习题	232
第二十章 SAS 高级应用	234
第一节 主成分分析	234
一、基本知识	234
二、应用举例	236
第二节 聚类分析	239
一、基本知识	239
二、聚类分析的 SAS 过程应用	240
三、系统聚类过程 CLUSTER	241
四、逐步聚类过程 FASTCLUS	244
五、变量聚类过程 VARCLUS	248
第三节 重复测量数据分析	256
一、单因素重复测量数据的方差分析	257
二、两因素重复测量数据的方差分析	260
习题	263
附录	265
附录 1 Excel 常用统计函数和统计分析工具	265
附录 2 SPSS 主要菜单英汉对照	267
附录 3 SAS 常用命令和过程	270
参考文献	272

第一篇



在生命科学中的应用

Excel ZAI SHENG MING KE XUE
ZHONG DE YING YONG



第一章 Excel 2007 应用基础

第一节 认识 Excel 2007 工作界面

Excel 2007 相比早期版本 Excel 的工作界面有很大的不同，全新的界面使得操作变得更加方便快捷。本节主要介绍 Excel 2007 的启动和工作界面，为更好地认识和使用 Excel 2007 打好基础。

一、启动 Excel

启动 Excel 常用的有两种方法：

(1) 一般程序方式：依次单击【开始→程序→Microsoft Office】在其子菜单中选择【Microsoft Office Excel 2007】，即可新建一个 Excel 文件，进入 Excel 工作界面。

(2) 快捷图标方式：如果桌面有快捷方式的，可以直接双击快捷方式打开 Excel 文件。如果没有的可以按照如下方法创建快捷方式：依次单击【开始→所有程序→Microsoft Office】进入子菜单右击【Microsoft Office Excel 2007】，选择【发送到→桌面快捷方式】，桌面上即可显示快捷方式。双击即可进入工作界面。

(3) 新建文档方式：在桌面空白处单击鼠标右键，选择【新建→Microsoft Office Excel 工作表】，即可创建 Excel 文件，进入 Excel 工作界面。

二、Excel 2007 工作界面概况

启动 Excel 后可以看到图 1-1 所示界面。

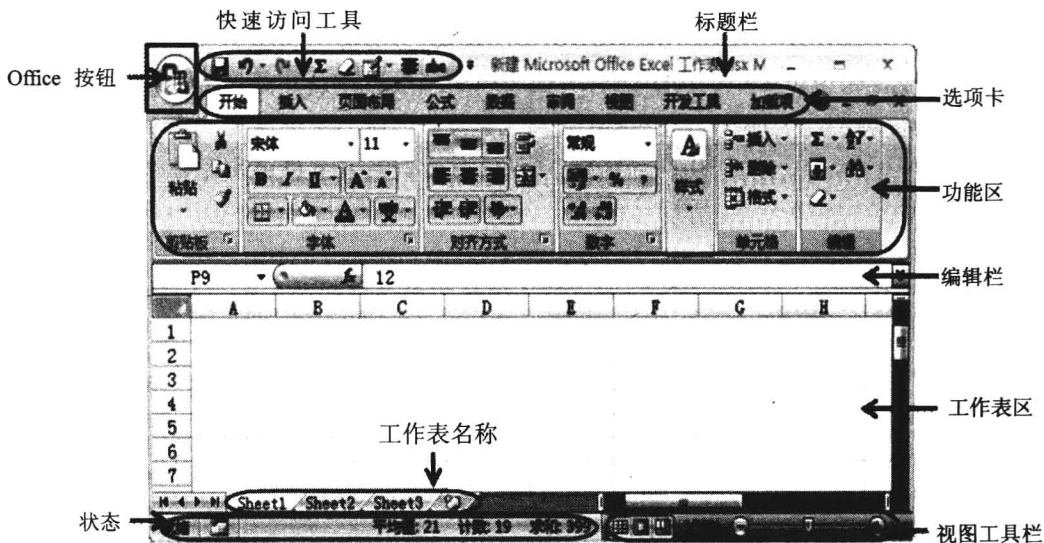


图 1-1 Excel 2007 工作界面

Office 按钮：位于整个窗口的左上方，单击此按钮，弹出下拉菜单，可以进行新建、打开、保存、打印等操作任务；右侧还会显示最近使用过的 Excel 文件。

快速访问工具栏：使用者可以根据自己的需要自定义该工具栏，如编辑文档时可进行快速保存、撤消等操作。

标题栏：中间部分显示的是 Excel 文档的名称，在右上角的三个按钮可以进行最小化、还原、最大化、关闭等操作。

选项卡：Excel 启动时一般的会显示【开始】、【插入】、【页面布局】、【公式】、【数据】、【审阅】和【视图】七个选项卡。【开发工具】和【加载项】需要另外设置才能载入，这将在本章第二节中进行介绍。

功能区：此区是多个逻辑组的集合，每个逻辑组里都有很多相应的命令，直接点击可以快速完成相应的操作。

编辑栏：编辑栏由两部分组成，左侧显示的是输入内容的单元格位置，右侧则为内容。当单元格要输入较长内容或有公式时在此编辑效果较好。如若要输入函数可点击按钮，弹出【插入函数】选项框，选择相应函数确定即可。

工作表区：该区是进行各项操作的窗口，由单元格、网格线、行号、列号、工作表标签等组成。Excel 默认工作簿由 Sheet1、Sheet2、Sheet3 三个工作表组成。点击 Sheet3 旁边的插入工作表图标可插入新的工作表。当工作表较多时可以通过工作表标签左侧的方向按钮进行移动查找。

状态栏：在状态栏的空白处常常会显示选定区域数据的平均值、计数、求和三个统计量。

视图工具栏：Excel 默认的页面是“普通视图”，通过该工具栏还可以切换到“分页预览”和“页面布局”模式。通过缩放工具可以将工作表区的内容放大或缩小（注：按住 Ctrl 键同时滚动鼠标也有此放大或缩小页面的功能）。

第二节 数据的输入与编辑

对数据进行分析，前提是必须先输入数据。数据的输入最基本的是手动输入，即直接在单元格中键入数据。当数据具有一定的规律性时则可以采用填充功能自动输入。本节将介绍几种数据的输入方法和基本的编辑操作。

一、数据的输入

（一）输入数值

在 Excel 中一般的数值可以直接输入，但当输入以 0 开头的数字（小数除外）时数字前面的 0 会自动被舍弃，因此必须通过特殊的方式来实现。

例如，输入以 0 开头的可运算的四位数值，可以按下述方法实现。

选中需要输入数据的单元格右击选择【设置单元格格式→分类→自定义】在类型中输入所需位数相同个数的“0”，如图 1-2 所示。如要输入“0001”，在【自定义】的【类型】框里输入 4 个 0，单击【确定】，然后在相应的单元格中直接输入“1”就行了。

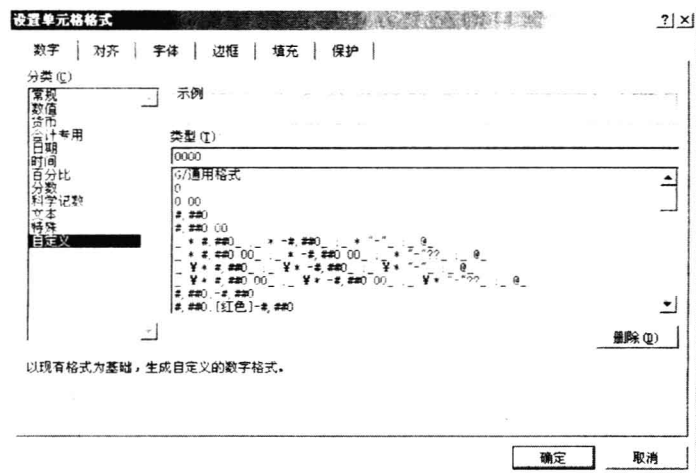


图 1-2 自定义数据类型

(二) 输入日期

输入日期时用户可以使用多种格式，一般使用斜杠“/”或者“-”来分隔年月日。年份常常用两位数表示。例如，要输入 2010-7-20 可以如图 1-3 所示输入。

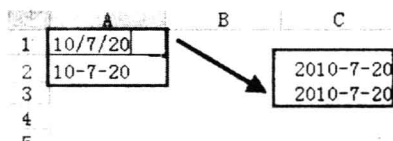


图 1-3 输入日期

(三) 输入文本

1. 输入以 0 开头的编号

方法一：选中需要输入数据的单元格右击选择【设置单元格格式—分类—文本】见图 1-4。

方法二：使用“'”符号实现，如图 1-5 所示。

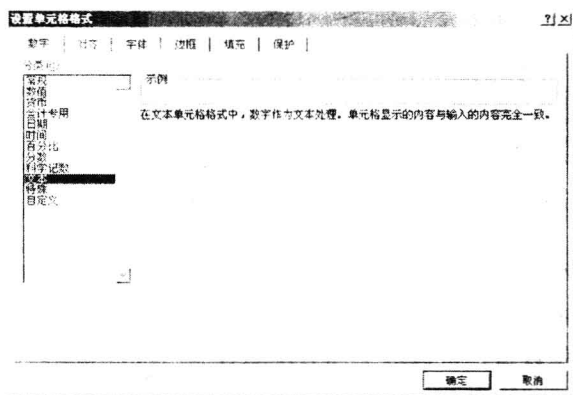


图 1-4 设置单元格格式图

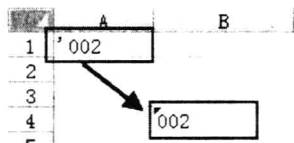


图 1-5 在需要输入的字符前面加“'”符号