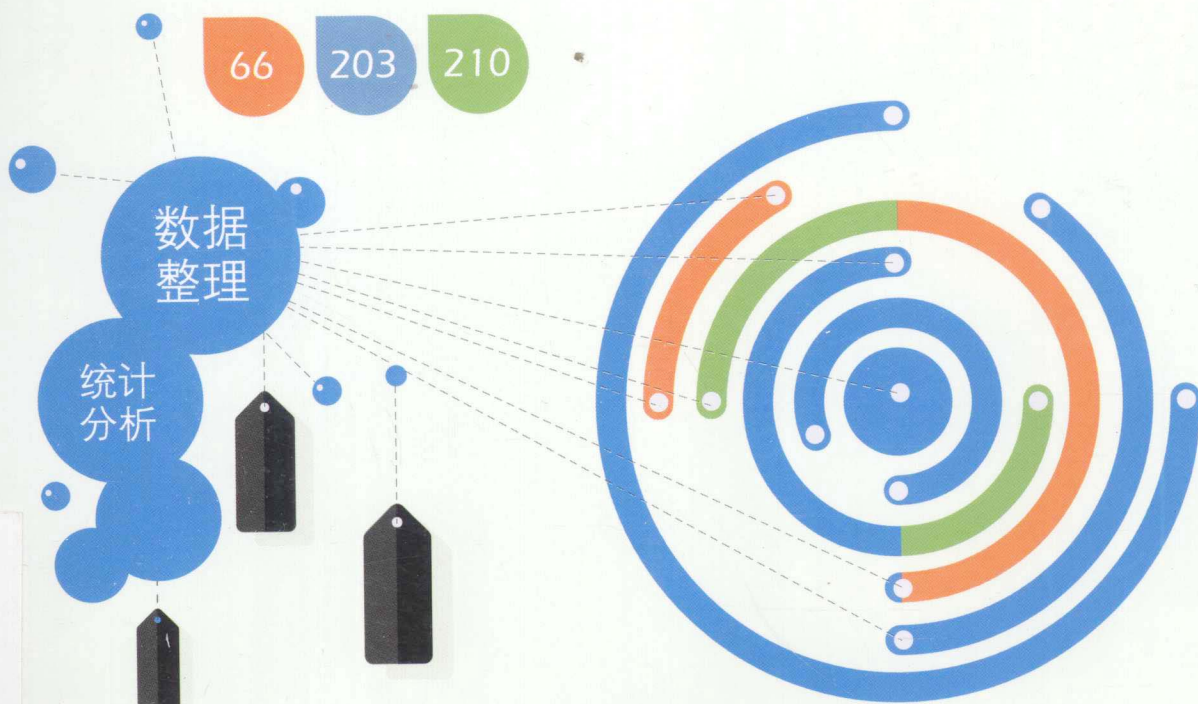


Excel 与 数据分析 (第3版)

宇传华 编著



十年沉淀，为你呈现

结合前两版的读者反馈，保留精华，补充变化，充分展示用Excel做统计的优势！



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

013056518

C819
13-3

Excel 与 数据分析 (第3版)

宇传华 编著



北航

C1663969

C819/13-3

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

812320310

内 容 简 介

本书由浅入深地介绍了 Excel 数据整理、统计分析、高级应用及电脑实验的方法。利用 Excel 函数,编制了近 100 个统计分析模板,如样本含量与检验功效的估计、Meta 分析、诊断实验评价、灰色模型预测、圆形分布分析等,这些模板由 Excel “数据分析”工具是无法实现的。此外,本书还利用 Excel 函数,制作了约 50 个统计学电脑实验,这些实验可帮助读者理解抽象的统计学概念,有利于统计学教学。本书配套的光盘操作简单,读者只需复制或键入数据,随即便可获得统计分析或电脑实验结果。

本书适用于和数据打交道的各界人士,可作为科研人员、管理人员、医务工作者、大中专院校师生的参考书或教材,尤其适用于学过或打算学习统计学的有关人员。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

Excel 与数据分析 / 宇传华编著. —3 版. —北京: 电子工业出版社, 2013.7
ISBN 978-7-121-20739-6

I. ①E… II. ①宇… III. ①表处理软件 IV. ①TP391.13

中国版本图书馆 CIP 数据核字 (2013) 第 133525 号

责任编辑: 付 睿

印 刷: 北京丰源印刷厂

装 订: 三河市鹏成印业有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 36 字数: 806 千字

印 次: 2013 年 7 月第 1 次印刷

印 数: 4000 册 定价: 79.00 元 (含光盘 1 张)

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

著者说明

本书主要章节由宇传华主笔完成，以下人员在本书（包括前面的两个版本）及 Excel 程序的编写等方面做出了较大贡献。

人员名单

武汉大学

华中科技大学

北京法马苏提克医药有限公司

首都医科大学

广西中医药大学

腾讯公司

深圳市卫生人口和计生委员会

广州军区武汉总医院

深圳市宝安区妇幼保健院

上海市普陀区疾病预防控制中心

上海市嘉定区疾病预防控制中心

广东省检验检疫局

中山大学

王震坤 杨子娟 李 鹏 连 肖

姚 聪 孙 敏 季 洁 李 妍

夏 欣 魏娜娜 郝玉伟

王家春

方龙

祝慧萍

李京

林嵩艺

董国营

文芸

钟文明

蒋丽丽

马飞飞

颜杰

张晋昕

前 言

本书第 1 版是 2002 年出版的《Excel 与数据分析》，第 2 版是 2009 年出版的《Excel 统计分析与电脑实验》，前面版本中的很多内容，特别是统计分析程序，至今仍非常简单实用，第 3 版继承了上述版本的所有优点。这 10 年来，每当读者以各种方式对本书发出赞许之声，或者以各种方式告之本书已售罄的消息时，我就感觉到自己责任的重大，有一种强烈的努力编写好本书的欲望。

第 3 版是以 Excel 2010 为基础编写的。从 Excel 2003 过渡到 Excel 2010 的开始几天，操作起来有些不习惯，甚至有点觉得它过于花哨与烦琐。但经过一段时间的适应后，发觉 Excel 2010 改进真的很大，从“功能选项卡”到“命令组”，再到每一个“命令按钮”，层次更加清晰，选项内容也更加丰富；将“文件”选项卡独立于 Excel 工作表之外，更方便了对整个文件的操作。Excel 更稳定了，Excel 函数增多了，兼容性更强了，函数计算精确度更高了，Excel 图形可独立保护与撤销保护等新功能都提示用户从使用 Excel 2003 过渡到使用 Excel 2010 的正确性与迫切性。

第 3 版所配光盘中的每一个统计学分析模板与电脑实验程序均在 Excel 2010 中运行正常。从数量上看，自编的 Excel 程序（或界面）第 1 版为 66 个，第 2 版是 203 个，现在的第 3 版是 210 个。绝大多数程序只需键入数据，结果便自动在同一工作表内产生，具有即改即可见的效果。

选用 Excel 做统计分析的理由

与标准统计学软件 SAS、SPSS 等相比较，Excel 的优势在于：①强大的数据自动填充功能，这使得在 Excel 工作表中输入数据变得相当简单；②方便的数据汇总与数据透视分析功能，可快速获得整个工作表，甚至整个工作簿数据的探索性统计分析结果；③灵活的单元格绝对引用与相对引用功能，使得大量类似的 Excel 公式输入变得简单，很多情况下只要键入一个公式，通过拖放方式便可完成该单元格周围其他单元格公式的生成；④完美的图表内置格式，使得统计图形既制作方便，又美观好看；⑤Excel 的“数据分析”工具可完成大多数常用的统计分析，对于较复杂的统计学方法可简单利用 Excel 内置的函数自编程完成；⑥正规渠道购买的电脑内大多数预装有微软 Office 办公软件，且为中文界面，使用它进行统计学分析无版权问题。

本书特色

本书从实例出发, 易学易懂。所配光盘内包括有大量实例数据与程序, 很多情况下, 只需在编好的 Excel 程序中键入有关数据或参数, 随即便可获得计算结果。

所配光盘程序均为开放代码, 读者通过书中有关提示, 结合光盘文件可以自编新的程序, 解决更复杂的统计学分析难题。部分 Excel 程序采用了工作表保护, 目的是防止 Excel 工作表内重要数据或计算公式发生不经意的改动。如果需要去除保护, 只需选择“审核→撤销工作表保护”命令, 输入密码“0”即可。

本书光盘程序是在 Windows 7 操作系统下, 以 Excel 2010 为基础, 屏幕分辨率为 1366×768 像素下制作完成的。

本书内容

本书共分为数据整理、统计分析、高级应用和电脑实验 4 篇。在数据整理篇中, 较为全面系统地介绍了数据的编辑、查询、建立子数据清单、数据透视表与透视图分析等方法。在统计分析篇中, 利用 Excel 图表向导功能制作了多种常见统计图形; 利用 Excel 内置函数编制了大量统计分析模板, 如卡方检验、正态分布等优度拟合检验、Fisher 精确概率法、非参数检验、Ridit 分析、简单相关与回归分析、多重相关与回归分析、加权 logistic 回归分析等。在高级应用篇中编制了样本含量与检验功效的估计、寿命表、诊断试验评价的样本含量估计、ROC 曲线分析、Meta 分析、TOPSIS 法、层次分析法等统计模块。在电脑实验篇中, 利用“开发工具→插入→ActiveX 控件→滚动条”制作滚动条, 结合 Excel 内置函数和内置图表, 制作了大量电脑实验, 如正态分布、二项分布、Poisson 分布等十余个概率分布电脑实验; z 分布、 t 分布、 F 分布、卡方分布等抽样分布电脑实验、中心极限定理、随机现象、假设检验、置信区间等共计近 50 个电脑实验程序。

第 3 版还新增加了灰色预测模型(第 12.6 节)、圆形分布分析(第 12.8 节)、Bland-Altman 法(第 16.6 节)等统计学方法与 Excel 实现程序。

对于“实用型”用户, 可直接利用有关程序进行统计学分析; 对于“高级”用户, 可根据书中提示, 结合光盘的相应文件, 对相应工作表略加修改, 完成更复杂的计算。

本书编者

本书第 1、2、3 版的所有编写人员已在“著者说明”页列出。

在第 3 版的修改编写过程中, 第 1~5 章以及附录 A、B、C、D、E 由王震坤负责, 第

6~10 章由杨子娟负责, 第 11~16 章由李鹏负责, 第 17~20 章由连肖和姚聪负责。此外, 王震坤在编写第 12.6 节, 以及联络其他编者等方面做出了较大贡献。

限于作者水平有限, 时间仓促, 书中错谬之处在所难免, 恳请各位同行、专家和广大读者批评指正。

宇传华 (E-mail: yuchua@163.com)

武汉大学

2013 年 6 月

目 录

数据整理篇

第 1 章 Excel 基本知识	2
1.1 Excel 的安装	3
1.2 Excel 工作界面简介	3
1.3 鼠标与键盘的操作	9
1.4 单元格区域与整个工作表的选取	12
1.5 工作表重命名与复制副本	13
1.6 工作表的格式化	14
1.7 工作表的打印	19
第 2 章 公式与函数	21
2.1 公式的建立	22
2.2 函数的结构	23
2.3 公式中的运算符	24
2.4 运算符的优先顺序	26
2.5 单元格的混合引用	26
2.5.1 相对引用	27
2.5.2 绝对引用	27
2.5.3 混合引用	27
2.5.4 A1 引用样式	28
2.5.5 应用实例	28
2.6 编辑公式常用的快捷键	30
2.7 编辑公式时常见的错误信息	30
2.8 公式保护与撤销保护的技巧	33
2.9 常用数学函数应用举例	34
2.9.1 常用数学函数	35
2.9.2 数组运算	39
2.9.3 矩阵的运算与求解线性方程组	40

2.10	创建自定义函数	43
2.11	获取函数帮助信息	44
第 3 章	数据清单的建立与利用	45
3.1	数字的单元格格式	46
3.2	数据的有效性设置	50
3.3	建立数据清单	54
3.4	数据“记录单”的作用	56
3.5	数据的自动填充	57
3.5.1	简单数据的自动填充	57
3.5.2	等比数列数据的填充	59
3.5.3	公式的自动填充	60
3.5.4	智能标记的应用	60
3.6	数据的分列	61
3.7	添加与编辑批注	63
3.8	特殊数学符号的键入	64
3.9	数据的导入与导出	66
第 4 章	数据的编辑与查询	68
4.1	窗口的冻结与拆分	69
4.2	数据的剪贴与删除	70
4.3	数据的查找与替换	71
4.4	行列转置与选择性粘贴的其他妙用	72
4.5	将连续型变量划分等级	75
4.6	年龄的计算与当前日期的生成	76
4.7	数据的跨工作表操作	76
4.8	多个工作表的合并计算	78
4.9	查询与新建子数据清单	81
4.9.1	排序	81
4.9.2	自动筛选	82
4.9.3	高级筛选	85
4.9.4	选择不重复的记录	86
4.9.5	人工筛选	86
4.10	两个类似文件的比较	88

第5章 分类汇总与透视分析	90
5.1 分类汇总	91
5.2 数据透视表	94
5.2.1 创建数据透视表	95
5.2.2 数据透视表的结构及其工具栏	96
5.2.3 由数据透视表获得均值标准差	98
5.2.4 产生列联表	100
5.2.5 创建子数据清单	106
5.2.6 设置报告格式与更新数据	106
5.2.7 在数据透视图表中创建公式	107
5.3 数据透视图	108
5.3.1 创建数据透视图	109
5.3.2 调整数据透视图	109
5.3.3 将数据透视图更改为静态图	111
5.4 数据透视的其他应用	113
5.4.1 多个工作表的数据透视	113
5.4.2 数据透视表网页	113

统计分析篇

第6章 统计图表	116
6.1 统计表的制作	117
6.2 统计图的结构及特点	118
6.3 Excel 标准图表类型简介	119
6.4 创建统计图的一般步骤	121
6.5 统计图的编辑与修饰	124
6.6 绘制统计图举例	127
6.6.1 分段条图	127
6.6.2 误差条图	129
6.6.3 百分条图	131
6.6.4 饼图	132
6.6.5 复合条饼图	133
6.6.6 线图与半对数线图	135

6.6.7	雷达图	136
6.6.8	气泡图	138
6.6.9	箱图	140
6.6.10	人口金字塔图	143
6.6.11	三角函数制作线图	146
6.6.12	曲面图	147
第7章	统计描述	149
7.1	数据分析工具的调用	150
7.2	统计学中的数据类型	151
7.2.1	数据类型的辨析	152
7.2.2	三类数据的计算	153
7.2.3	三类数据的转化	153
7.3	区间数据的统计学描述	154
7.3.1	区间数据的频数表与直方图	154
7.3.2	集中位置指标	160
7.3.3	离散程度指标	163
7.3.4	数据分析工具: 描述统计	169
7.4	名义数据的统计描述	170
7.4.1	COUNTIF 函数清点计数	171
7.4.2	率的标准化法	171
7.4.3	动态数列	173
第8章	区间数据的假设检验	176
8.1	假设检验的一般步骤	177
8.2	样本均数与总体均数的比较	178
8.2.1	已知样本均数和标准差	178
8.2.2	已知样本原始数据	179
8.3	配对资料的比较	180
8.4	两个样本均数的比较	181
8.4.1	两样本的总体方差齐性检验	181
8.4.2	总体方差齐同的 t 检验	183
8.4.3	总体方差不齐的 t 检验	185
8.4.4	大样本的 z 检验	186

8.5	多样本方差 Bartlett 齐性检验	187
8.6	单因素方差分析	188
8.6.1	单因素方差分析工具	188
8.6.2	单因素方差分析图示	190
8.6.3	无原始数据的单因素方差分析	191
8.6.4	单因素方差分析的多重比较	191
8.7	双因素方差分析	193
8.7.1	无重复双因素方差分析	193
8.7.2	有重复双因素方差分析	195
第 9 章	名义与有序数据的假设检验	198
9.1	卡方检验	199
9.1.1	四格表资料的 χ^2 检验	199
9.1.2	四格表配对资料的 χ^2 检验	203
9.1.3	四格表资料的 Fisher 精确概率检验	205
9.1.4	行×列表资料的 χ^2 检验	208
9.2	频数分布的拟合优度检验	210
9.2.1	正态分布	210
9.2.2	二项分布	213
9.2.3	Poisson 分布	215
9.2.4	均匀分布	216
9.3	非参数检验	217
9.3.1	配对资料 Wilcoxon 符号秩检验	218
9.3.2	两独立样本非参数检验	221
9.3.3	多样本 Kruskal-Wallis 秩和检验	225
9.3.4	多个相关样本的 Friedman 检验	228
9.4	Ridit 分析	230
第 10 章	相关与回归	235
10.1	简单相关与回归	236
10.1.1	Excel 函数实现简单相关与回归	238
10.1.2	数据分析工具实现简单相关与回归	243
10.1.3	添加趋势线实现直线回归分析	247
10.2	曲线回归	248

10.2.1	对数曲线拟合	249
10.2.2	乘幂曲线拟合	251
10.2.3	指数曲线拟合	252
10.2.4	多项式曲线拟合	253
10.3	非线性回归	255
10.3.1	对数曲线拟合	255
10.3.2	乘幂曲线拟合	256
10.3.3	指数曲线拟合	257
10.3.4	多项式曲线拟合	257
10.4	多重相关与回归	258
10.4.1	多重相关与多重协方差	258
10.4.2	多重回归分析	260
10.4.3	用函数实现多重回归与相关	261
10.5	加权 logistic 回归	263

高级应用篇

第 11 章	变量求解与方案优选	268
11.1	单变量求解	269
11.1.1	目标搜索	269
11.1.2	求解非线性方程	270
11.2	规划求解	271
11.3	方案管理器	274
11.4	常见的方案优选方法	277
11.4.1	确定型决策	277
11.4.2	非确定型决策	278
11.4.3	风险型决策	281
11.5	综合评价方法	282
11.5.1	TOPSIS 法	282
11.5.2	层次分析法	285
第 12 章	预测分析	288
12.1	自相关与自回归	289
12.1.1	线性时序趋势	289

12.1.2	DW 统计量	290
12.1.3	自相关	291
12.1.4	自回归	292
12.2	移动平均法	294
12.3	指数平滑法	295
12.3.1	数据分析工具——指数平滑	296
12.3.2	指数平滑的初始值	297
12.3.3	求解最佳平滑常数	298
12.4	季节预测法	299
12.4.1	哑变量多重回归	300
12.4.2	AR(4) 模型	303
12.4.3	季节指数法	304
12.5	Markov 预测法	306
12.6	灰色预测模型	310
12.6.1	GM(1,1)模型的预测与评价方法	310
12.6.2	灰色模型预测实例	312
12.7	傅利叶分析	314
12.8	圆形分布分析	316
12.8.1	角度变换	316
12.8.2	平均角及其假设检验	317
12.8.3	圆形分布分析实例	318
12.8.4	两个或多个样本平均角的比较	320
第 13 章	寿命表	324
13.1	现时寿命表	325
13.2	去死因寿命表	330
13.3	健康期望寿命表	334
第 14 章	随机化分组与样本含量估计	337
14.1	随机化分组	338
14.1.1	完全随机设计	338
14.1.2	配对设计	339
14.1.3	配伍组设计	340
14.1.4	三个随机化分组程序介绍	342

14.2	样本含量的估计	343
14.2.1	Nomogram 图示法	345
14.2.2	假设检验的样本含量估计	346
14.2.3	置信区间的样本含量估计	353
14.3	检验功效的估计	357
14.4	流行病学研究的样本含量与检验功效	362
14.4.1	病例对照研究	362
14.4.2	队列研究	365
14.5	方差分析等方法的样本含量与检验功效	366
14.5.1	方差分析的检验功效	366
14.5.2	方差分析的样本含量	367
14.5.3	其他方法的样本含量与检验功效	368
第 15 章	高级科研统计图与 Meta 分析	369
15.1	统计地图	370
15.2	P-P 图与 Q-Q 图	373
15.2.1	创建绘图数据	373
15.2.2	绘制正态 P-P 图	374
15.2.3	绘制正态 Q-Q 图	375
15.3	生存曲线	375
15.3.1	Kaplan-Meier 法计算生存率及其标准误	376
15.3.2	绘制生存曲线	377
15.4	重复测量数据的曲线图	378
15.5	Meta 分析统计量的置信区间图	379
15.5.1	Meta 分析的计算公式	380
15.5.2	连续资料的 Meta 分析	383
15.5.3	离散资料的 Meta 分析	387
15.5.4	相关系数资料的 Meta 分析	391
15.5.5	P 值的 Meta 分析	392
第 16 章	诊断试验评价	394
16.1	常用统计指标	395
16.2	ROC 曲线原理及其绘制方法	399
16.2.1	有序分类资料 ROC 分析	400

16.2.2	连续资料 ROC 分析	403
16.2.3	值小为阳性的 ROC 分析	404
16.3	诊断试验评价的 Meta 分析	406
16.3.1	SROC 分析	406
16.3.2	ROC 面积的 Meta 分析	412
16.4	样本含量估计	415
16.4.1	灵敏度与特异度的置信区间估计	415
16.4.2	ROC 曲线下面积的置信区间估计	416
16.4.3	固定 FPR 的灵敏度置信区间估计	417
16.4.4	单个准确度指标的假设检验	418
16.4.5	两灵敏度或特异度的比较	420
16.4.6	ROC 曲线下面积的比较	421
16.5	Kappa 值估计	422
16.6	Bland-Altman 法	425

电脑实验篇

第 17 章	概率分布	428
17.1	正态分布	429
17.1.1	采用公式制作正态分布电脑实验	429
17.1.2	正态曲线下的面积	431
17.1.3	正态分布综合电脑实验	433
17.2	二项分布	435
17.2.1	概率质量函数	435
17.2.2	累计分布函数	436
17.2.3	二项分布电脑实验	436
17.2.4	样本率及率的标准误	438
17.3	Poisson 分布	438
17.3.1	概率质量函数与累计分布函数	439
17.3.2	Poisson 分布电脑实验	439
17.4	其他连续分布	440
17.4.1	对数正态分布	441
17.4.2	均匀分布	441
17.4.3	指数分布	442

17.4.4	Weibull 分布	442
17.4.5	Beta 分布	444
17.4.6	Gamma 分布	444
17.5	其他离散分布	445
17.5.1	均匀分布	445
17.5.2	超几何分布	446
第 18 章	随机抽样与抽样误差	448
18.1	随机数发生器与随机函数	449
18.1.1	随机数发生器	449
18.1.2	随机数函数	454
18.2	有放回抽样与无归还抽样	455
18.2.1	“抽样”分析工具	455
18.2.2	无归还抽样	456
18.3	随机抽样电脑实验	456
18.3.1	从正态总体中随机抽样	457
18.3.2	从非正态总体中随机抽样	461
第 19 章	假设检验的概率分布与统计用表	466
19.1	z 分布	467
19.1.1	标准正态分布曲线下面积	467
19.1.2	标准正态分布分位数	469
19.1.3	标准正态分布的其他应用	469
19.1.4	z 分布电脑实验	472
19.2	t 分布	473
19.2.1	t 分布的图形及其分布特征	474
19.2.2	t 分布曲线下面积与检验界值	476
19.2.3	t 分布电脑实验	477
19.3	F 分布	479
19.3.1	F 分布的图形及其分布特征	480
19.3.2	F 分布曲线下面积与检验界值	481
19.3.3	F 分布电脑实验	483
19.4	χ^2 分布	484
19.4.1	χ^2 分布的图形及其分布特征	484