

TONGJI XUE SHIWU

统计学实务

顾晓安 朱建国 主编

立信会计出版社

C8
95

统计学实务

TONGJIXUE SHIWU

顾晓安 朱建国 主编

北方工业大学图书馆



00591869

立信会计出版社

SCG96/07

版权所有，剽窃翻印必究。

图书在版编目(CIP)数据

统计学实务/顾晓安,朱建国主编. —上海:立信会计出版社,2005.3

ISBN 7-5429-1422-7

I. 统… II. ①顾… ②朱… III. 统计学
IV. C8

中国版本图书馆 CIP 数据核字(2005)第 017933 号

出版发行	立信会计出版社
经 销	各地新华书店
电 话	(021)64695050×215 (021)64391885(传真) (021)64388409
地 址	上海市中山西路 2230 号
邮 编	200235
网 址	www.lixinaiph.com
E-mail	lxaph@sh163.net
E-mail	lxzbs@sh163.net (总编室)

印 刷	立信会计常熟市印刷联营厂
开 本	787×960 毫米 1/16
印 张	12
插 页	2
字 数	239 千字
版 次	2005 年 3 月第 1 版
印 次	2005 年 3 月第 1 次
印 数	3 000
书 号	ISBN 7-5429-1422-7/F·1285
定 价	19.00 元

如有印订差错 请与本社联系

前 言

以计算机与网络为代表的现代信息技术近年来取得了突飞猛进的发展,已经渗透到人类社会的每一个角落,成为现代社会不可或缺的基本元素。在统计工作中应用现代信息技术也已逐渐成为广大统计人员必备的基本职业素质,如何将现代信息技术融入统计学的教学过程去,是当前经济管理类专业教学改革中亟待研究的主要课题。而通过实验手段来提高学生应用现代信息技术解决实际统计问题的能力,不失为是一条行之有效的途径和有益的探索。

本书各项实验课题的设计,着重培养学生能够结合企业常见的统计业务,运用现代信息技术解决实际统计问题的能力。目前,市场上的大型统计软件很多,但是,由于其数据处理的原理和过程都是不公开的,这就给统计学的授课老师、学生和其他相关从业人士带来许多不便和疑惑。因此,本书在设计每一个实验课题时,尽可能选择通用性较好的软件,并对具体操作步骤详作介绍,使读者能对数据的统计处理过程有更全面透彻的了解和掌握。

本书既可以作为“社会经济统计学”或“统计学原理”等课题的课程实验教材,亦可以作为广大统计实际工作者学习使用现代信息技术进行实际统计工作的自学参考书。作为教材,本书紧密结合当前统计学教学的课程体系,既适合学生集中实验,更宜将各章实验内容灵活穿插进统计学理论教学中作为巩固和拓展之用。

本书是在上海理工大学商学院会计系与经济科学实验中心编著的内部教材基础上,根据几届本科生的试用情况,进一步充实实验内容,完善实验环境和步骤,精心编著而成。参加本书编著的教师都长期在高等院校从事统计实验和统计教学实践,具有丰富的教学经验,曾指导本科生的统计实验并指导多名研究生运用计算机统计的方法完成了部分科研项目的统计分析工作。

本书由顾晓安、朱建国担任主编,各章节具体分工如下:前言、第一、第二、第三、第四章由顾晓安编著,第五、第六、第七章由朱建国编著,第八、第九章由

顾晓安、朱建国和王国兴共同编著,并由顾晓安负责对全书初稿进行修改和总纂。

本书从计划编著到正式出版,自始至终都得到了立信会计出版社编辑余榕老师的热情帮助和大力支持,同时也得到了上海理工大学商学院的领导和教师们的支持和帮助,在此一并致谢。

由于编著者学识有限,加上时间比较仓促,书中不妥之处在所难免,恳请读者提出宝贵意见,以便今后修订(电子邮件:guxiaoan@126.com; zhjg@fait.org.cn)。

编著者

2005年2月

目 录

第一章 统计学实验概述	1
第二章 统计资料整理与分组	19
实验一	19
实验二	24
实验三	29
实验四	34
本章练习题及答案	42
第三章 平均指标	47
实验五	47
实验六	49
实验七	51
实验八	54
实验九	56
实验十	59
实验十一	62
本章练习题及答案	65
第四章 总量指标和相对指标	67
实验十二	67
实验十三	69
实验十四	72
实验十五	76
实验十六	80
本章练习题及答案	81

第五章 抽样与估计	86
实验十七	86
实验十八	91
实验十九	93
实验二十	96
本章练习题及答案	98
 第六章 假设检验	101
实验二十一	103
实验二十二	104
实验二十三	107
实验二十四	110
实验二十五	114
实验二十六	117
本章练习题及答案	119
 第七章 统计指数	124
实验二十七	124
实验二十八	128
实验二十九	131
实验三十	135
实验三十一	137
本章练习题及答案	141
 第八章 动态数列	146
实验三十二	146
实验三十三	148
实验三十四	151
实验三十五	154
实验三十六	158
本章练习题及答案	163
 第九章 相关与回归分析	168
实验三十七	168

实验三十八·····	170
实验三十九·····	174
实验四十·····	177
本章练习题及答案·····	181

第一章 统计学实验概述

运用以计算机、网络等为代表的现代信息技术来完成实际统计工作和具体统计任务,在企业日常统计工作中已经是非常普遍的现象。现代信息技术作为一种高效的工具,在统计日常事务中有着非常广泛的应用前景。

统计学的教学目标就是要培养和训练学生能够应用有效的计算方法和计算工具来完成企业或其他经济主体的统计任务。就计算工具而言,过去主要是利用算盘和计算器,而现在则主要是利用现代信息技术。

众所周知,信息技术与教育相整合是当前教育改革和创新的重要趋势,也是经济管理类各专业的理论教学面向社会实际的重要发展方向。统计学是经济管理类各专业学生必须掌握的专业基础学科,为了使学生在毕业后能够很快地适应在企业各个部门运用现代信息技术进行统计业务的操作,在统计学的教学中引入相关业务的模拟实验教学有助于缩短学生毕业后进入岗位角色的时间,也有利于学生更好地掌握统计学理论和实务,提高学生应用计算机和现代网络信息技术的能力。

本书是为经济管理类各个专业(统计学专业除外)学习统计学原理或社会经济统计学原理课程设计的实验教材。亦可以作为统计实际工作者学习应用现代信息技术进行实际统计工作的参考书。

一、统计学计算机实验的目的

统计学计算机实验的目的可以归纳为以下三点:

- (1) 加深学生对统计理论和实务的理解。通过对实验过程的操作和实验结果的分析,使学生能够更好地掌握统计学的基本理论、基础知识、基本方法和基本技能。
- (2) 培养学生实事求是、理论联系实际和严谨认真的学习态度。
- (3) 培养学生动手实践的能力和自主学习的能力。

二、统计学计算机实验的时间安排

统计学计算机实验可以有两种时间安排方式:一种为集中式安排;另一种为分散式安排。两种时间安排方式各有利弊。

集中式时间安排是将统计学实验集中安排在统计学课程结束后的一段时间中作为课程设计项目进行,一般可以集中两周。这种安排便于教师集中辅导,时间利用效果较好。但是,学生往往会把统计学实验看成是独立于课程体系以外的一门实践课程,不能主动地将现代信息技术运用到统计实务操作的各个环节和阶段中,不利于对学生实际工作能力的培养。目前各校大多将统计学实验的内容作为统计学课程设计集中进行实验。

分散式时间安排是将统计实验教学插入统计学课程的各个章节中进行,将统计实验的内容作为课程作业或课外练习来完成,这种时间安排比较有利于实验设备的统筹安排,对于帮助学生理解计算机信息技术作为日常统计事务处理的有效工具比较有利,但是,不便于教师辅导。也有部分院校采用这种时间安排,在每门课程中划出部分课时用于实验指导。

三、实验环境和软件选择

统计学计算机实验的实验环境是指实验的软硬件配置和实验室环境配置。统计学计算机实验要求配备一定的实验用硬件设备及相关的系统软件和应用程序软件,为了使实验能够更好地体现实际操作情况,也可以进行适当的场景布置,使学生进行实验时有身临其境的感受。

统计学计算机实验所需要配备的硬件设备一般包括能够运行各种统计学计算机软件 and 各类常见的通用应用软件的计算机,能接入国际互联网的上网设备等。

统计学计算机实验要求的软件包括操作系统软件和必要的应用软件。

(一) 操作系统软件

操作系统是最底层的系统软件,它是对硬件系统功能的扩充,是支撑应用软件运行的软件平台。统计学计算机实验要求的操作系统一般是 Windows 98 或 Windows 2000 或 Windows XP 等比较流行的操作系统软件。

(二) 统计学计算机软件

1. SPSS 软件

统计学计算机实验所使用的应用软件比较广泛,可供选择的余地很大。目前,市场上流行的有两大类:一类是与管理信息系统整合的大型软件系统“企业资源计划”(简称 ERP),在 ERP 中,统计业务处理作为企业管理信息系统的一个部分融入到企业管理的各个领域;另一类是以日常统计业务处理为核心的独立统计软件,这类软件有很多种类,有比较通用的,也有专用于某一行业或领域的专业统计软件,这些软件主要是为专业统计工作者服务的,例如,比较有名的统计软件 SPSS 系列包含了 SPSS 的基本操作功能、图形功能和统计分析功能,图形功能包括常见的统计图形和二维、三维交互图形,统计分析功能包括样本描述与数据准备、参数估计、假设检验、

非参数检验、方差分析、回归分析、相关分析、因子分析、聚类分析、判别分析和可靠性分析等,内容丰富、涉及面广且操作繁杂,一般仅供统计专业人员和进行大型专业统计时使用。

SPSS 是 Statistical Package for the Social Sciences(社会科学统计软件包)英文名称的首字母缩写。SPSS 公司于 2000 年正式将英文全称更改为 Statistical Product and Service Solutions,意为“统计产品与服务解决方案”。

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制,1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地拓展了它的应用范围,使 SPSS 能够应用于自然科学、技术科学、社会科学的各个领域。世界上许多有影响的报刊、杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、使用方便、功能齐全等方面特点给予了高度的评价与称赞,是目前世界上应用最广泛的专业统计软件。在国际学术界有条不成文的规定,在国际学术交流中,凡是用 SPSS 软件完成的计算和统计分析,可以不必说明算法,由此可见其影响之大和信誉之高。

SPSS 现在的最新版本为 11.03,大小约为 200M。一共由十个模块组成,其中 SPSS Base 为基本模块,其余九个模块为 Advanced Models、Regression Models、Tables、Trends、Categories、Conjoint、Exact Tests、Missing Value Analysis 和 Maps,分别用于完成某一方面的统计分析功能,它们均需要挂接在 Base 上运行。除此之外,SPSS 11 完全版还包括 SPSS Smart Viewer 和 SPSS Report Writer 两个软件,虽然它们并未整合进来,但功能上完全是 SPSS 的辅助软件。

SPSS 最突出的特点就是操作界面极为友好,它使用 Windows 的窗口方式展示各种管理和数据分析方法的功能,使用对话框展示各种功能选择项。SPSS 采用类似 EXCEL 表格的方式输入管理数据,数据接口较为通用,能方便的从其他数据库中读入数据。其统计过程包括了常用的、较为成熟的统计方法。

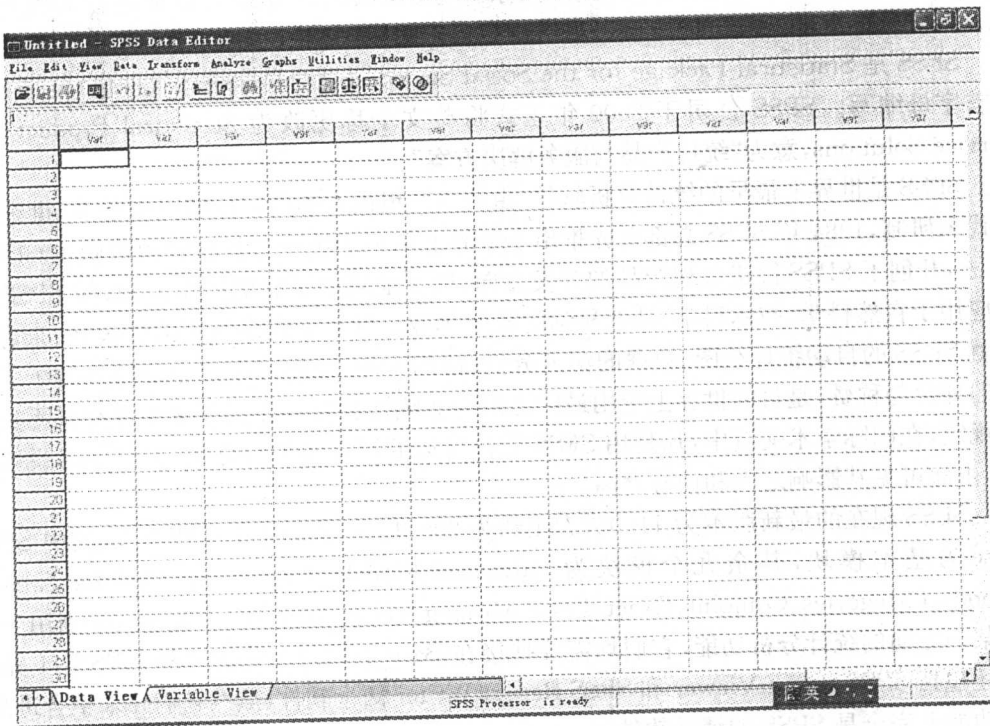
SPSS 是世界上最早采用图形菜单驱动界面的统计软件,它将几乎所有的功能都以统一、规范的界面展现出来,用户只要掌握一定的 Windows 操作技能,粗通统计分析原理,就可以使用该软件为特定的科研工作服务,它是统计人员的首选统计软件。

SPSS 软件目前只吸收较为成熟的统计方法,而对于最新的统计方法,SPSS 公司的做法是为之发展一些专门软件,如针对树结构模型的 Answer Tree,针对神经网络技术的 Neural Connection、专门用于数据挖掘的 Clementine 等,而不是直接纳入 SPSS。此外,其输出结果虽然漂亮,但不能为 WORD 等常用文字处理软件直接打开,只能采用拷贝、粘贴的方式加以交互调用。这些都可以说是 SPSS 软件的不足之处。

以下是 SPSS 的操作界面,如图表 1-1 所示:

图表 1-1

SPSS 的操作界面



2. SAS 软件

另一种常用的统计学计算机软件 SAS 系统的全称为 Statistics Analysis System, 最早由北卡罗来纳大学的两位生物统计学研究生编制, 并于 1976 年成立了 SAS 软件研究所, 正式推出了 SAS 软件。SAS 是用于决策支持的大型集成信息系统, 但该软件系统最早的功能限于统计分析, 至今, 统计分析功能也仍是它的重要组成部分和核心功能。

在数据处理和统计分析领域, SAS 系统被誉为国际上的标准软件系统, 并在 1996~1997 年度被评选为建立数据库的首选产品, 它堪称统计软件界的巨无霸。例如, 在以苛刻、严格著称于世的美国 FDA 新药审批程序中, 新药试验结果的统计分析规定只能用 SAS 进行, 其他软件的计算结果一律无效! 哪怕只是简单的平均数和标准差计算也不行! 由此可见 SAS 的权威地位。

SAS 现在的版本为 9.0 版, 大小约为 1G, 是一个组合软件系统, 它由多个功能模块组合而成, 其基本部分是 BASE SAS 模块。BASE SAS 模块是 SAS 系统的核心, 承担着主要的数据管理任务, 并管理用户使用环境, 进行用户语言的处理, 调用其他

SAS 模块和产品。各模块的安装及更新都可通过其安装程序非常方便地进行。SAS 系统具有灵活的功能扩展接口和强大的功能模块,在 BASE SAS 的基础上,还可以增加如下不同的模块而增加不同的功能:SAS/STAT(统计分析模块)、SAS/GRAPH(绘图模块)、SAS/QC(质量控制模块)、SAS/ETS(经济计量学和时间序列分析模块)、SAS/OR(运筹学模块)、SAS/IML(交互式矩阵程序设计语言模块)、SAS/FSP(快速数据处理的交互式菜单系统模块)、SAS/AF(交互式全屏幕软件应用系统模块),等等。

SAS 是由大型机系统发展而来的,其核心操作方式就是程序驱动,经过多年的发展,现在已成为一套完整的计算机语言,其用户界面也充分体现了这一特点:它采用 MDI(多文档界面),用户在 PGM 视窗中输入程序,分析结果以文本的形式在 OUTPUT 视窗中输出。使用程序方式,用户可以完成所有需要做的工作,包括统计分析、预测、建模和模拟抽样等。但是,这使得初学者在使用 SAS 时必须学习 SAS 语言,入门比较困难。SAS 的 Windows 版本根据不同的用户群开发了几种图形操作界面,这些图形操作界面各有特点,使用时非常方便。

由于 SAS 系统是从大型机上的系统发展而来,在设计上也完全是针对专业用户进行的,因此其操作至今仍以编程为主,人机对话界面不太友好,并且在编程操作时需要用户最好对所使用的统计方法有较清楚的了解,非统计专业人员掌握起来较为困难。而且, SAS 极为高昂的价格和只租不卖的销售策略使得实力不足的个人和机构只能对其望而却步。

(三) 电子表格软件

企业管理人员在日常管理工作中,经常会碰到各种各样的统计问题,要编制各类表格,进行各种计算,电子表格软件强大的表格格式化功能、计算和函数功能、图表制作功能等,是最适合非统计专业的经济管理人员使用的应用统计软件。当前最流行的电子表格软件是微软公司的 Office 系列办公自动化软件的组件之一 Excel。

Microsoft Excel 具有功能强大、技术先进、使用方便的特点,广泛应用于会计、财务、金融、营销、贸易、统计、行政等领域。是各类经济管理人员最有用的工具。Excel 可以用来制作电子表格、完成许多复杂的数据运算,进行数据的分析和预测,并且具有强大的制作图表的功能等。Excel 作为一种电子表格软件,它不仅有强大的数据处理能力,而且它可以通过查询功能去调用 Access、SQL SERVER、ORACLE、DB2 等数据库软件建立的大型数据库的内容。用户可以在工作表中对这些数据进行筛选、排序、查询、编辑、进行数据分析和统计处理、打印统计报表等,也可以对从数据库获得的数据进行进一步的加工、分析和综合。此外,Excel 的 VBA 功能强大,如果与工作表结合应用,可以更好地完成日常统计工作提供帮助。

对于非统计专业的学生进行统计学实验使用 Microsoft Excel 是比较理智的。首先,

Excel 是比较通用的软件,作为实验教学可以用比较小的代价取得,操作上也比较简便。其次,Excel 不仅能够方便的处理表格和进行图形分析,其更强大的功能体现在对数据的自动处理和计算,尤其是 Excel 的统计函数的应用,可以为缺少专业统计学背景的经济管理类各专业学生完成专业领域的统计任务提供很大的方便。最后,Excel 与 Word 和 Access 等的紧密结合,可以为学生提交实验报告等带来方便。

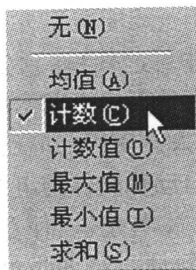
将 Excel 运用到统计领域一般有四种方法:直接统计法、函数法、透视表法和使用数据分析加载宏。

1. 直接统计法

直接统计法就是利用 Excel 的自动计算功能,在工作表上直接进行快速的统计,统计结果显示在状态栏上。Excel 具有自动计算功能,可以进行“求和、计数、均值、最大值、最小值”等计算。选定单元格后,Excel 将在状态栏(就是 Excel 工作表窗口底部的水平区域;如果未显示,可单击“视图”菜单中的“状态栏”命令)中显示单元格区域的总和,Excel 启动时自动计算的缺省设置为“求和”。当我们需要计数或计算均值时,右击状态栏,从显示的快捷菜单中单击“计数”或“均值”(如图表 1-2 所示)。这样,当我们选择了一个数据区域系统会自动显示选定区域内单元格的数目或这些数据的均值、最大值、最小值等。如图表 1-2 所示:

图表 1-2

Excel 自动计算功能菜单



【例 1】 某企业机修车间 10 名工人的年龄分别为 23, 32, 36, 40, 43, 48, 51, 21, 26, 32。计算其平均年龄。直接统计结果如图表 1-3 所示。

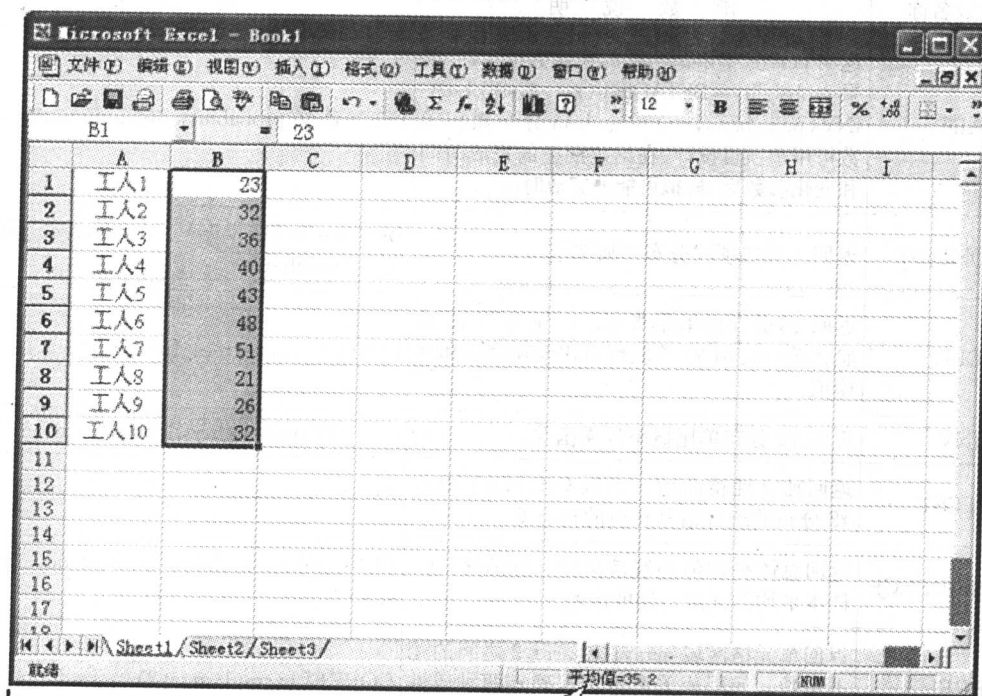
2. 函数法

如果统计时涉及的单元格区域很大,有上百条记录,直接统计时需要滚动屏幕才能选定单元格区域,这时最好使用函数法。

函数法是使用 Excel 内置函数完成统计任务的方法,用于统计的函数有很多,各种统计函数的名称、说明和语法形式如图表 1-4 所示:

图表 1-3

自动计算平均值



平均值=35.2

图表 1-4

统计函数汇总

函数名称	函数说明	语法形式
AVEDEV	返回一组数据与其均值的绝对偏差的平均值,即离散度。	AVEDEV(number1,number2,...)
AVERAGE	返回参数算术平均值。	AVERAGE(number1,number2,...)
AVERAGEA	计算参数清单中数值的平均值(算数平均值)。不仅数字,而且文本和逻辑值(如 TRUE 和 FALSE)也将计算在内。	AVERAGEA(value1,value2,...)
BETADIST	返回 Beta 分布累积函数的函数值。Beta 分布累积函数通常用于研究样本集合中某些事物的发生和变化情况。	BETADIST(x,alpha,beta,A,B)

(续表)

函数名称	函数说明	语法形式
BETAINV	返回 beta 分布累积函数的逆函数值。即,如果 $\text{probability}=\text{BETADIST}(x,\cdots)$, 则 $\text{BETAINV}(\text{probability},\cdots)=x$ 。beta 分布累积函数可用于项目设计,在给定期望的完成时间和变化参数后,模拟可能的完成时间。	$\text{BETAINV}(\text{probability},\alpha,\beta,A,B)$
BINOMDIST	返回一元二项式分布的概率值。	$\text{BINOMDIST}(\text{number_s},\text{trials},\text{probability_s},\text{cumulative})$
CHIDIST	返回 χ^2 分布的单尾概率。 χ^2 分布与 χ^2 检验相关。使用 χ^2 检验可以比较观察值和期望值。	$\text{CHIDIST}(x,\text{degrees_freedom})$
CHIINV	返回 χ^2 分布单尾概率的逆函数。	$\text{CHIINV}(\text{probability},\text{degrees_freedom})$
CHITEST	返回独立性检验值。函数 CHITEST 返回 χ^2 分布的统计值及相应的自由度。	$\text{CHITEST}(\text{actual_range},\text{expected_range})$
CONFIDENCE	返回总体平均值的置信区间。置信区间是样本平均值任意一侧的区域。	$\text{CONFIDENCE}(\alpha,\text{standard_dev},\text{size})$
CORREL	返回单元格区域 array1 和 array2 之间的相关系数。使用相关系数可以确定两种属性之间的关系。	$\text{CORREL}(\text{array1},\text{array2})$
COUNT	返回参数的个数。利用函数 COUNT 可以计算数组或单元格区域中数字项的个数。	$\text{COUNT}(\text{value1},\text{value2},\cdots)$
COUNTA	返回参数组中非空值的数目。利用函数 COUNTA 可以计算数组或单元格区域中数据项的个数。	$\text{COUNTA}(\text{value1},\text{value2},\cdots)$
COVAR	返回协方差,即每对数据点的偏差乘积的平均数,利用协方差可以决定两个数据集之间的关系。	$\text{COVAR}(\text{array1},\text{array2})$
CRITBINOM	返回使累积二项式分布大于等于临界值的最小值。此函数可以用于质量检验。	$\text{CRITBINOM}(\text{trials},\text{probability_s},\alpha)$
DEVSQ	返回数据点与各自样本均值偏差的平方和。	$\text{DEVSQ}(\text{number1},\text{number2},\cdots)$
EXPONDIST	返回指数分布。使用函数 EXPONDIST 可以建立事件之间的时间间隔模型。	$\text{EXPONDIST}(x,\lambda,\text{cumulative})$

(续表)

函数名称	函数说明	语法形式
FDIST	返回 F 概率分布。使用此函数可以确定两个数据系列是否存在变化程度上的不同。	FDIST(x, degrees_freedom1, degrees_freedom2)
FINV	返回 F 概率分布的逆函数值。	FINV(probability, degrees_freedom1, degrees_freedom2)
FISHER	返回点 x 的 Fisher 变换。该变换生成一个近似正态分布而非偏斜的函数。	FISHER(x)
FISHERINV	返回 Fisher 变换的逆函数值。使用此变换可以分析数据区域或数组之间的相关性。	FISHERINV(y)
FORECAST	根据给定的数据计算或预测未来值。	FORECAST(x, known_y's, known_x's)
FREQUENCY	以一系列垂直数组返回某个区域中数据的频率分布。	FREQUENCY(data_array, bins_array)
FTEST	返回 F 检验的结果。F 检验返回的是当数组 1 和数组 2 的方差无明显差异时的单尾概率。可以使用此函数来判断两个样本的方差是否不同。	FTEST(array1, array2)
GAMMADIST	返回伽玛分布。可以使用此函数来研究具有偏态分布的变量。伽玛分布通常用于排队分析。	GAMMADIST(x, alpha, beta, cumulative)
GAMMAINV	返回伽玛分布的累积函数的逆函数。	GAMMAINV(probability, alpha, beta)
GAMMALN	返回伽玛函数的自然对数, $\Gamma(x)$ 。	GAMMALN(x)
GEOMEAN	返回正数数组或数据区域的几何平均值。	GEOMEAN(number1, number2, ...)
GROWTH	根据给定的数据预测指数增长值。	GROWTH(known_y's, known_x's, new_x's, const)
HARMEAN	返回数据集合的调和平均值。调和平均值与倒数的算术平均值互为倒数。	HARMEAN(number1, number2, ...)
HYPGEOMDIST	返回超几何分布。	HYPGEOMDIST(sample_s, number_sample, population_s, number_population)
INTERCEPT	利用已知的 x 值与 y 值计算直线与 y 轴的截距。	INTERCEPT(known_y's, known_x's)
KURT	返回数据集的峰值。	KURT(number1, number2, ...)