

《Excel 在统计中的应用》

作者：林飞

(经作者同意，首次在 <http://www.riskage.com> 发布)

第一节	中文 Excel 概述	3
一、	中文 Excel 简介	3
二、	Excel 的安装和启动	5
三、	Excel 工作界面简介	6
第二节	Excel 基本操作	8
一、	Excel 操作方法概述	8
二、	文件基本操作	9
三、	数据的输入输出操作	9
四、	数据的移动操作	11
五、	数据的删除操作	12
六、	与其它软件交换数据的方法	12
第三节	Excel 在描述统计中的应用	12
一、	描述统计工具	13
二、	直方图工具	14
三、	利用 Excel 绘制散点图	15
四、	数据透视表工具	17
五、	排位与百分比工具	18
第四节	Excel 在推断统计中的应用	19
一、	二项分布工具	19
二、	其它分布的函数	21
三、	随机抽样的工具	24
四、	由样本推断总体	25
五、	假设检验	27
六、	双样本等均值假设检验	29
七、	正态性的 X^2 检验	32
八、	列联表分析	34
九、	单因素方差分析	36
十、	线性回归分析	38
十一、	相关系数分析工具	40
十二、	协方差分析工具	41
十三、	自回归模型的识别与估计	41
十四、	季节变动时间序列的分解分析	45
附表 1	Excel 统计函数一览表	49
附表 2	Excel 数据分析工具一览表	56

Excel 在统计中的应用

第一节 中文 Excel 概述

一、中文 Excel 简介

Microsoft Excel 是美国微软公司开发的 Windows 环境下的电子表格系统，它是目前应用最为广泛的办公室表格处理软件之一。自 Excel 诞生以来 Excel 历经了 Excel 5.0、Excel 95、Excel 97 和 Excel 2000 等不同版本。随着版本的不断提高，Excel 软件的强大的数据处理功能和操作的简易性逐渐走入了一个新的境界，整个系统的智能化程度也不断提高，它甚至可以在某些方面判断用户的下一步操作，使用户操作大为简化。这些特性，已使 Excel 成为现代办公软件重要的组成部分。

Excel 具有强有力的数据库管理功能、丰富的宏命令和函数、强有力的决策支持工具，它具有以下主要特点。

(一) 分析能力

Excel 除了可以做一些一般的计算工作外，还有 400 多个函数，用来做统计、财务、数学、字符串等操作以及各种工程上的分析与计算。Excel 还专门提供了一组现成的数据分析工具，称为“分析工具库”，这些分析工具为建立复杂的统计或计量分析工作带来极大的方便。

(二) 操作简便

当需要将工作表上某个范围内的数据移到工作表上的另一个位置时，只需用鼠标键，选定要移动的资料，将该范围资料拖动至所需的位置，松开鼠标即可。如果要将公式或数据复制到临近的单元格内，可以拖动“填充柄”，公式或数据就会被复制到目标单元格中。

此外，在使用 Excel 时，可以单击鼠标右键，屏幕上将出现相应的“快捷菜单”它将帮助用户尽快地寻找到所需要的常用命令。

Excel 内有许多工具按钮，每一个按钮代表一个命令。例如，要建立一个新的工作簿文件，就可直接按下工具栏第一个按钮，而不必先选择“文件”菜单，然后再选择其中的“打开”命令。在 Excel 97 中，系统共有 14 组常用的工具栏，用户可以自由选择加入或隐藏这些工具栏。

(三) 图表能力

在 Excel 中，系统大约有 100 多种不同格式的图表可供选用，用户只要做几个简单的按键动作，就可以制作精美的图表。通过图表指南一步步的引导，可使用不同的选项，得到所需的结果，满意的话就继续，不满意则后

退一步，重新修改选项，直到最后出现完美的图表。

(四) 数据库管理能力

对于一个公司，每天都会产生许多新的业务数据，例如，销售数据、库存的变化、人事变动的数据资料等。这些数据必须加以处理，才能知道每段时间的销售金额、某个时候的存货量、要发多少薪水给每个员工等。要对这些数据进行有效的处理，就离不开数据库系统。所谓数据库系统，就是一组有组织的信息。例如，将每位职员简历写在一张卡片上，卡片放在盒子内，盒子内的数据通常组成列和行。再如，每种产品的产地、规格、单位、单价、数量组成一列，每一行都包含同一属性的数据，每列都包含同一品种的数据，即每列都有产地、规格、单位、单价、数量。

管理数据库可用专门的数据库管理软件，如 FoxPro ,Access ,Clipper , Sybase 等。在 Excel 中提供了类似的数据库管理功能，保存在工作表内的数据都是按照相应的行和列存储的，这种数据结构再加上 Excel 提供的有关处理数据库的命令和函数，使得 Excel 具备了能组织和管理大量数据的能力。

(五) 宏语言功能

利用 Excel 中的宏语言功能，用户可以将经常要执行的操作的全过程记录下来，并将此过程用一简单的组合按键或工具按钮保存起来。这样，在下次操作中，只需按下所定义的宏功能的相应按键或工具按钮即可，而不必重复整个过程。例如，可以定义一个打开最后编辑文件且可以自动执行的宏，以后当用户打开 Excel 后，将自动打开上一次编辑的工作簿。

在 Excel 中，高级用户可使用 Visual Basic 语言，进行宏命令的开发。利用宏命令，用户可以将 Excel 的下拉菜单和对话框更改或将图形按钮的说明更换，使它们更适合于用户的工作习惯和特殊要求。

(六) 样式功能

在 Excel 中，用户可以利用各种文字格式化的工具和制图工具，制作出美观的报表。Excel 工作表里的资料，在打印以前可将其放大或缩小进行观察，用户可以对要打印的文件作微调。

用户可将要打印出的格式制作好，并存储成样本，以后可以读取此样本文件，就可依据样本文件的格式打印出美观的报表。Excel 的专业文书处理程序具有样式工具。所谓样式，就是将一些格式化的组合用一个名称来表示，以后要使用这些格式化的组合时，只要使用此名称即可，因此可大幅度地节省报表格式化的时间。

(七) 对象连接和嵌入功能

利用对象连接和嵌入功能，用户可将其他软件（例如，画笔）制作的图形插入到 Excel 的工作表中。当需要更改图案时，只要在图案上双击鼠标键，制作该图案的程序就会自动打开，图案将出现在该图形编辑软件内，修改、编辑后的图形也会在 Excel 内显示出来。也可以将一个声音文件或动画文件嵌入到 Excel 工作表中，使工作表变成一幅声形并茂的报表。

（八）连接和合并功能

通常，每个工作在一张工作表上执行即可，早期的工作表软件都只能在一张工作表上执行。但有时需要同时用到多张工作表，例如，公司内每个分公司每月都会有会计报表，要将各分公司区的资料汇总起来，就需要用到连接和合并功能。Excel 很容易将工作表连接起来，并进行汇总工作。Excel 内一个工作簿可以存放许多工作表、图形等，每个工作簿文件最多可以由 255 张工作表组成。

二、Excel 的安装和启动

以下以 Excel 97 为例，说明 Excel 的安装过程。如果没作特别说明，本章介绍其它内容都是针对 Excel 97 软件的。

Excel 97 作为 Office 97 系统中的一个应用程序，它的工作平台为 Windows 95 或其以上的版本。对中文 Excel 而言，最好安装在中文 Windows 系统平台之上。

（一）安装 Excel 97

1. 在 CD-ROM 中放入 Office 97 光盘。

2. 单击 Windows 的“开始”按钮并选择“运行”命令。在弹出的“运行”对话框，“打开”框中寻找光盘上的安装程序“SETUP.EXE”，找到后单击“确定”按钮即可。（如果光盘具有自动执行的功能，则可省略以上步骤，直接点取屏幕上的“开始安装”按钮。）

3. 开始安装后系统将要求用户选择安装模式：“典型”、“自定义”、“快速”。用户可根据需要选择安装模式，一般可以选择“典型”安装方式，那么安装软件将装入最常用的 Office 组件，如果用户想定制所要安装的组件，那么可以选择“自定义”安装方式，而“快速”安装方式是一种最节约磁盘空间的安装方式，它仅装入 Office 运行中不能缺少的组件。

在剩下安装过程中则用户按屏幕提示选择。

（二）Excel 的启动与退出。

启动 Excel 的常用方法是：单击任务栏上的“开始”按钮，此时屏幕上出现一个弹出式菜单，将鼠标指向“程序”项后，屏幕出现另一个弹出菜单，单击“Microsoft Excel”，就可以启动 Excel 系统，此时屏幕上出现如

图附-1 所示的 Excel 主工作画面。若安装 Excel 时，生成了“快捷工具栏”，则双击其中的 Excel 按钮也可立即启动 Excel。

退出 Excel 常用的方法是，用鼠标单击标题栏的□按钮或进入“文件”菜单栏，单击“退出”选项。

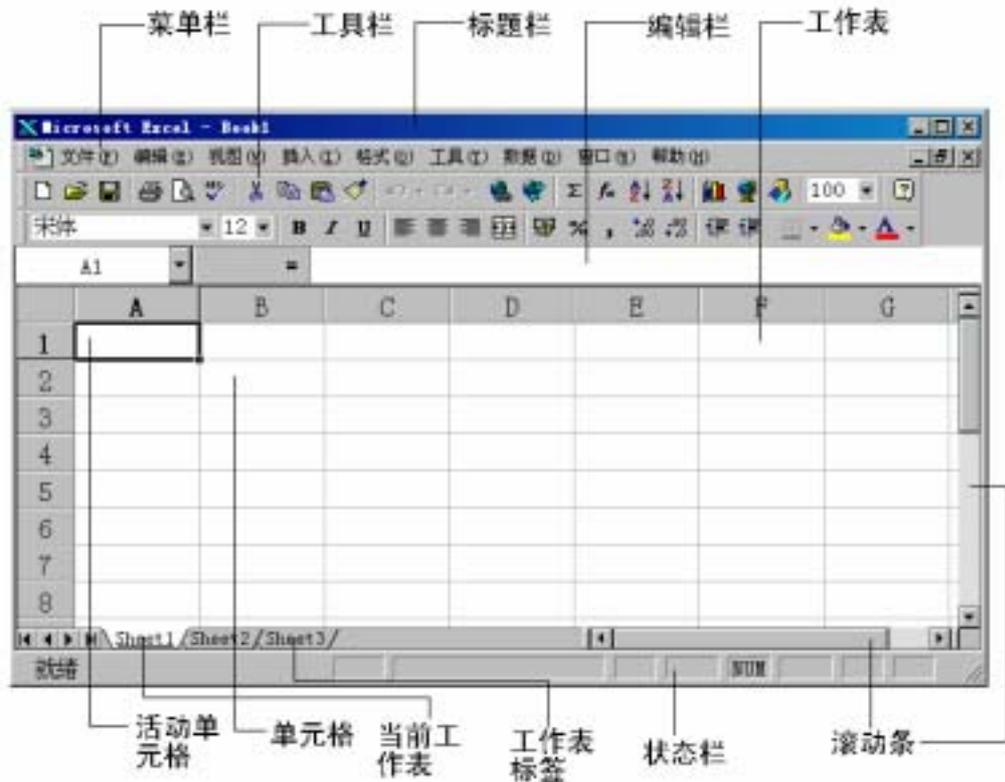


图 附-1 Excel 的工作界面

三、Excel 工作界面简介

按图附-1 从上到下的顺序 Excel 工作界面包含如下几项内容：“标题”栏、“菜单”栏、“工具”栏、“编辑”栏、工作表、工作表标签、滚动条、和“状态”栏。下面分别介绍它们的作用。

(一)“标题”栏

“标题”栏告诉用户正在运行的程序名称和正在打开的文件的名称。如图附-1 所示，标题栏显示“Microsoft Excel -Book1”表示此窗口的应用程序为 Microsoft Excel，在 Excel 中打开的当前文件的文件名为 Book1.xls。

(二)“菜单”栏

“菜单”栏按功能把 Excel 命令分成不同的菜单组，它们分别是“文件”、“编辑”、“视图”、“插入”、“格式”、“工具”、“表格”、

“帮助”。当菜单项被选中时，引出一个下拉式菜单，可以从中选取相应的子菜单。

另外，在屏幕的不同地方单击鼠标右键时，“快捷菜单”将出现在鼠标指针处。选取“快捷菜单”中的命令同从菜单栏的菜单上选取相应命令的效果是一样的，但选取速度明显增快。

(三)“工具”栏

Excel 可显示几种工具栏，这些工具可控制简化用户的操作。“工具”栏中的按钮都是菜单中常用命令的副本，当鼠标指向某一按钮后，稍等片刻在按钮右下方会显示该按钮命令的含意。用户可以配置“工具”栏的内容，通过“视图”菜单中的“工具”栏子菜单来选择显示不同类型的“工具”或全部显示出来。下面介绍出现在 Excel 开始屏幕中的两种“工具”栏。

1.“常用”工具栏

“常用”工具栏中为用户准备了访问 Excel 最常用命令的快捷按钮，如“新建文件”按钮，“打开文件”按钮，“保存文件”按钮等。

2.“格式”工具栏

“格式”工具栏专门放那些和文本外观有关的命令，如字体、字号、对齐方式及其他选项。

(四)“编辑”栏

“编辑”栏给用户提供活动单元格的信息。在“编辑”栏中用户可以输入和编辑公式，“编辑”栏位于图附-1 中第 5 行。

“编辑”栏由“名字”栏和“公式”栏组成。位于“编辑”栏左侧的“名字”栏中显示的是活动单元格的坐标，也可在“名字”栏中直接输入一个或一块单元格的地址进行单元格的快速选定；位于“编辑”栏右侧的“公式”栏可用于编辑活动单元格的内容，它包含三个按钮和一个编辑区。当向活动单元格输入数据时，公式栏中便出现三个按钮，三个按钮从左至右分别是：“ ”(取消)按钮、“ ”(确认)按钮和“=”(公式指南)按钮。

通常 Excel 在工作区中显示“编辑”栏。在“视图”菜单中的“编辑栏”命令是一个开关命令，它可以用于隐藏或显示“编辑”栏。

(五)工作表

工作簿窗口包含了 16 张独立的工作表 (sheet)。开始时，窗口中显示第一张工作表“Sheet1”，该表为当前工作表。当前工作表只有一张，用户可通过点击工作表下方的标签击活其他工作表为当前工作表。

工作表是一个由行和列组成的表格。行号和列号分别用字母和数字区别。行由上自下范围 1~65536，列号则由左到右采用字母编号 A~IV。因此

每张表为 256 列 65536 行，若从 Excel 导入的数据超过以上范围，则会被 Excel 自动截去。每一个行、列坐标所指定的位置称之为单元格。在单元格中用户可以键入符号、数值、公式以及其他内容。

(六) 工作表标签

工作表标签通常用“Sheet1”，“Sheet2”等名称来表示，用户也可以通过用鼠标右击标签名，选择弹出菜单中“重命名”命令来修改标签名。Excel 一般同时显示工作表队列中的前 3 个标签。利用标签队列左边的一组标签滚动按钮可显示队列中的后续工作表的标签。工作簿窗口中的工作表称之为当前工作表，当前工作表的标签为白色，其他为灰色。

(七)“滚动”栏

当工作表很大时，如何在窗口中查看表中的全部内容呢？可以使用工作簿窗口右边及下边的滚动栏，使窗口在整张表上移动查看，也可以通过修改常用“工具”栏中“显示比例框”的参数来扩大整个工作表的显示范围。

(八)“状态”栏

“状态”栏位于 Excel 窗口底部，它的左端是信息区，右端是键盘状态区。

在信息区中，显示的是 Excel 的当前工作状态。例如，当工作表准备接受命令或数据时，信息区显示“就绪”；当在“编辑”栏中键入新的内容时，信息区显示“输入”；当选取菜单命令或工具按钮时，信息区显示此命令或工具按钮用途的简要提示。

在键盘状态区中，显示的是若干按键的开关状态。例如，当按[Caps Lock] 键时，状态 栏中便显示“CAPS”。

与“编辑”栏相同，在“视图”菜单中的“状态”命令是一个开关命令，它可以用于隐藏或显示“状态”栏。

第二节 Excel 基本操作

一、Excel 操作方法概述

要完成任一项 Excel 操作一般都可以找到三种操作方法：鼠标操作、菜单操作和键盘命令操作。例如，想要将 A1 单元格的数据复制到 A2 单元格去，有如下几种操作方法：

(一) 鼠标操作法：先用鼠标选中 A1 单元格，然后缓慢移动鼠标到 A1 单元格的右下角，当鼠标的形状变为黑色实心“十”字形之后(以后称之为

“添充柄”)，拖动鼠标到 A2 单元格，然后放开鼠标，则 A1 的数据就复制到 A2 单元格了。

(二) 菜单操作法：先用鼠标选中 A1 单元格，选择“编辑”菜单中的“复制”命令，然后用鼠标选中 A2 单元格，再选择“编辑”菜单中的“粘贴”命令，数据就复制到 A2 单元格了。

(三) 键盘命令操作法：直接用鼠标选中 A2 单元格，从键盘输入“=A1”命令，则复制即告完成。

以上是 Excel 中很典型的三种操作方法。在实际使用过程中，应根据实际情况，尽量选择三种方法中最简洁的操作方法，以提高操作速度。

二、文件基本操作

(一) 新建文件：进入“文件”菜单栏，选择“新建”即可创建一个新的 Excel 文件。

(二) 打开文件：进入“文件”菜单栏，选择“打开”子菜单，可在 Excel 中打开一个已经存在的数据文件。它可以是 Excel 的数据文件，也可以是 Excel 兼容的其它软件的数据文件。可在不同窗口中同时打开多个数据文件，通过“窗口”菜单下方的不同选项，进行不同窗口的切换。

(三) 保存文件：进入“文件”菜单栏，选择“保存”命令，可保存当前数据文件。如果选择“另存为”，可将当前工作簿存为一个新的文件。保存文件的格式可以是 Excel 的数据文件，也可以是 Excel 兼容的其它软件的数据文件。

(四) 文件打印：进入“文件”菜单栏，选择“打印”，可打印当前的工作簿文件。打印之前，可以选择“文件”菜单栏的“页面设置”和“打印预览”选项，进行打印前的页面设置操作和打印效果的预先浏览。

三、数据的输入输出操作

(一) 数据的手动输入

建立一个新的 Excel 文件之后，便可进行数据的输入操作。Excel 中以单元格为单位进行数据的输入操作。一般用上下左右光标键，Tab 键或用鼠标选中某一单元格，然后输入数据。

Excel 中的数据按类型不同通常可分为四类：数值型，字符型，日期型，和逻辑型。Excel 根据输入数据的格式自动判断数据属于什么类型。如日期型的数据输入格式为“月/日/年”，“月-日-年”或“时：分：秒”。要输入逻辑型的数据，输入“true”(真)或“false”(假)即可。若数据由数字与小数点构成，Excel 自动将其识别为数字型，Excel 允许在数值型数据前加入货币符号，Excel 将其视为货币数值型，Excel 也允许数值型数据用科学记

数法表示，如 2×10^9 在 Excel 中可表示为 2E+9。除了以上三种格式以外的输入数据，Excel 将其视为字符型处理。

(二) 公式生成数据

Excel 的数据中也可由公式直接生成。例如：在当前工作表中 A1 和 B1 单元格中已输入了数值数据，欲将 A1 与 B1 单元格的数据相加的结果放入 C1 单元格中，可按如下步骤操作：用鼠标选定 C1 单元格，然后输入公式 “=A1+B1” 或输入 “=SUM(a1:b1)”，回车之后即可完成操作。C1 单元格此时存放实际上是一个数学公式 “A1+B1”，因此 C1 单元格的数值将随着 A1、B1 单元格的数值的改变而变化。Excel 提供了完整的算术运算符，如 + (加) - (减) * (乘) / (除) % (百分比) ^ (指数) 和丰富的函数，如 SUM(求和)、CORREL(求相关系数)、STDEV(求标准差)等，供用户对数据执行各种形式的计算操作，在 Excel 帮助文件中可以查到各类算术运算符和函数的完整使用说明。

(三) 复制生成数据

Excel 中的数据也可由复制生成。实际上，在生成的数据具有相同的规律性的时候，大部分的数据可以由复制生成。可以在不同单元格之间复制数据，也可以在不同工作表或不同工作簿之间复制数据，可以一次复制一个数据，也可同时复制一批数据，为数据输入带来了极大的方便。普通单元格的复制结果与公式单元格的复制结果相差较大，下面分别予以说明。

1. 普通单元格指的是非公式的单元格。普通单元格的复制，一般可以按如下步骤进行：

- (1) 拖动鼠标选定待复制的区域，选定之后该区域变为黑色。Excel 可以进行整行、整列或整个表格的选定操作，例如，如果要选定表格的第一列，可直接用鼠标单击列标 “A”，如果要选定表格的第一行，可直接用鼠标单击行标 “1”，如果要选定整个表格，可直接点击全选按钮，如图附-2 所示：

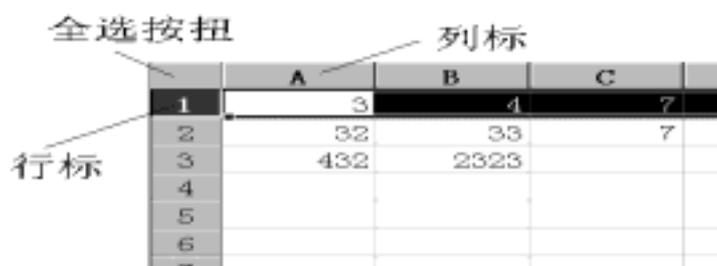


图 附-2

- (2) 选定完区域之后，用鼠标右击该区域，选择 “复制”，将区域

内容复制到粘贴版之中。可以发现该区域已被虚线包围。

(3) 用鼠标右击目标区域，选择“粘贴”，则单元格区域的复制即告完成。

2. 公式单元格的复制，一般可分为两种，一种是值复制，一种是公式复制。值复制指的是只复制公式的计算结果到目标区域，公式复制指的是仅复制公式本身到目标区域。下面对它们的操作步骤分别予以说明。

(1) 值复制： 拖动鼠标选定待复制区域。
用鼠标右击选定区域，选择“复制”选项。
用鼠标右击目标区域，再单击“选择性粘贴”子菜单。出现复制选项，选定“数值”选项，然后用鼠标单击“确定”按钮，则公式的值复制即告完成。

(2) 公式复制：

公式复制是 Excel 数据成批计算的重要操作方法，要熟练公式复制的操作首先要区分好两个概念：单元格的相对引用与绝对引用。Excel 中的公式中一般都会引用到别的单元格的数值，如果你希望当公式复制到别的区域之时，公式引用单元格不会随之相对变动，那么你必须要在公式中使用单元格的绝对引用。如果你希望当公式复制到别的区域之时，公式引用单元格也会随之相对变动，那么你必须要在公式中使用单元格的相对引用。在公式中如果直接输入单元格的地址，那么默认的是相对引用单元格，如果在单元格的地址之前加入“\$”符号那么意味着绝对引用单元格。例如，在当前工作表中 A1 和 B1 单元格中已输入了数值数据，用鼠标选定 C1 单元格，然后输入公式“=A1+B1”，此公式引用的便是两个相对的单元格 A1、B1，也就是说，如果将该公式复制到 C2 的单元格，公式所引用的单元格的地址将随着发生变化，公式将变为“=A2+B2”，如果将该公式复制到 F100 的单元格，那么公式将变为“=D100+E100”，这就是相对引用的结果，公式的内容随着公式的位置变化而相对变化。如果在 C1 单元格输入的是“=\$A\$1+\$B\$1”那么此公式引用的便是绝对的单元格，不论将公式复制到何处，公式的内容都不会发生变化。当然，绝对引用和相对引用亦可在同一公式之中混合交叉使用，例如，如果在 C1 单元中输入的是公式“=A\$1+B\$1”，那么意味着，公式的内容不会随着公式的垂直移动而变动，而是随着公式的水平移动而变动，如果将该公式复制到 F100 单元格，那么公式将变为，“=D\$1+E\$1”。可以作这样的归纳：公式中“\$”符号后面的单元格坐标不会随着公式的移动而变动，而不带“\$”符号后面的单元格坐标会随着公式的移动而变动。

在实际的使用中，如果能把单元格的相对引用与绝对引用灵活应用到 Excel 的公式之中，能为数据成批准确运算带来极大的方便。

四、数据的移动操作

数据的移动操作可按如下步骤进行：

- (一) 拖动鼠标选定待移动区域。
- (二) 用鼠标右击选定区域，选择“剪切”选项。
- (三) 用鼠标右击目标区域，选择“粘贴”，则单元格区域的移动即告完成。

与数据的复制操作不同，公式单元格的移动操作不存在值移动或公式移动的区别，也不存在绝对引用已相对引用的区别，移动操作将把公式单元格的公式内容原原本本移动到目标区域，不作任何改动。

五、数据的删除操作

数据的删除操作可按如下步骤进行：

- (一) 拖动鼠标选定待删除区域。
- (二) 用鼠标右击选定区域，选择“删除”，即可删除单元格区域的内容。

如果不小心删除了不该删除的区域，可以通过“编辑”菜单的“撤消”命令来恢复被删除的内容。“撤消”操作是 Excel 中较常用到的操作，如果不小心实施了错误的操作，那么可以通过“撤消”操作使工作表恢复原样。

六、与其它软件交换数据的方法

在 Excel 中可以打开其它类型的数据文件，如 FOXPRO 系列的 DBF 数据库文件，文本文件，lotus1-2-3 的数据文件等。具体操作方法如下：

- (一) 在“文件”菜单中选择“打开”子菜单。
- (二) 在“打开文件”对话框中选择所要打开的文件的类型及其所在的目录。

(三) 用鼠标双击该文件名，并按 Excel 提示步骤操作即可打开该文件。

Excel 文件同样也可存为其它类型的数据文件，具体操作方法如下：

- (一) 编辑好文件后，在“文件”菜单中选择“另存为”子菜单。
- (二) 在“另存为”对话框中选择所要打开文件的类型及其所在的目录。
- (三) 输入文件名之后，用鼠标单击“保存”按钮即可。

第二节介绍了 Excel 的一些比较主要的基本操作方法，在 Excel 中还有许多其它的基本操作，如表格显示格式控制，打印格式控制，Excel 帮助的使用等等，在应用 Excel 进行统计数据分析之前熟练这些操作是非常有必要的。第三节开始介绍 Excel 在统计数据分析中应用。

第三节 Excel 在描述统计中的应用

在使用 Excel 进行数据分析时，要经常使用到 Excel 中一些函数和数据分析工具。其中，函数是 Excel 预定义的内置公式。它可以接受被称为参数的特定数值，按函数的内置语法结构进行特定计算，最后返回一定的函数运算结果。例如，SUM 函数对单元格或单元格区域执行相加运算，PMT 函数在给定的利率、贷款期限和本金数额基础上计算偿还额。函数的语法以函数名称开始，后面是左圆括号、以逗号隔开的参数和右圆括号。参数可以是数字、文本、形如 TRUE 或 FALSE 的逻辑值、数组、形如 #N/A 的错误值，或单元格引用。给定的参数必须能产生有效的值。参数也可以是常量、公式或其它函数。

Excel 还提供了一组数据分析工具，称为“分析工具库”，在建立复杂的统计分析时，使用现成的数据分析工具，可以节省很多时间。只需为每一个分析工具提供必要的数据和参数，该工具就会使用适宜的统计或数学函数，在输出表格中显示相应的结果。其中的一些工具在生成输出表格时还能同时产生图表。如果要浏览已有的分析工具，可以单击“工具”菜单中的“数据分析”命令。如果“数据分析”命令没有出现在“工具”菜单上，则必须运行“安装”程序来加载“分析工具库”。安装完毕之后，必须通过“工具”菜单中的“加载宏”命令，在“加载宏”对话框中选择并启动它。

一、描述统计工具

(一) 简介：此分析工具用于生成对输入区域中数据的单变量分析，提供数据趋中性和易变性等有关信息。

(二) 操作步骤：

1. 用鼠标点击工作表中待分析数据的任一单元格。
2. 选择“工具”菜单的“数据分析”子菜单。
3. 用鼠标双击数据分析工具中的“描述统计”选项。
4. 出现“描述统计”对话框，对话框内各选项的含义如下：

输入区域：在此输入待分析数据区域的单元格范围。一般情况下 Excel 会自动根据当前单元格确定待分析数据区域。

分组方式：如果需要指出输入区域中的数据是按行还是按列排列，则单击“行”或“列”。

标志位于第一行/列：如果输入区域的第一行中包含标志项(变量名)，则选中“标志位于第一行”复选框；如果输入区域的第一列中包含标志项，则选中“标志位于第一列”复选框；如果输入区域没有标志项，则不选任何复选

框，Excel 将在输出表中生成适宜的数据标志。

均值置信度：若需要输出由样本均值推断总体均值的置信区间，则选中此复选框，然后在右侧的编辑框中，输入所要使用的置信度。例如，置信度 95%可计算出的总体样本均值置信区间为 10，则表示：在 5%的显著水平下总体均值的置信区间为 $(\bar{X} - 10, \bar{X} + 10)$ 。

第 K 个最大/小值：如果需要在输出表的某一行中包含每个区域的数据的第 k 个最大/小值，则选中此复选框。然后在右侧的编辑框中，输入 k 的数值。

输出区域：在此框中可填写输出结果表左上角单元格地址，用于控制输出结果的存放位置。整个输出结果分为两列，左边一列包含统计标志项，右边一列包含统计值。根据所选择的“分组方式”选项的不同，Excel 将为输入表中的每一行或每一列生成一个两列的统计表。

新工作表：单击此选项，可在当前工作簿中插入新工作表，并由新工作表的 A1 单元格开始存放计算结果。如果需要给新工作表命名，则在右侧编辑框中键入名称。

新工作簿：单击此选项，可创建一新工作簿，并在新工作簿的新工作表中存放计算结果。

汇总统计：指定输出表中生成下列统计结果，则选中此复选框。这些统计结果有：平均值、标准误差、中值、众数、标准偏差、方差、峰值、偏斜度、极差（全距）最小值、最大值、总和、样本个数。

5. 填写完“描述统计”对话框之后，按“确定”按钮即可。

(三) 结果说明：描述统计工具可生成以下统计指标，按从上到下的顺序其中包括样本的平均值 (\bar{X}) ，标准误差 (S/\sqrt{n}) ，组中值 (Medium)，众数 (Mode)，样本标准差 (S)，样本方差 (S^2) ，峰度值，偏度值，极差 (Max-Min)，最小值 (Min)，最大值 (Max)，样本总和，样本个数 (n) 和一定显著水平下总体均值的置信区间。

二. 直方图工具

(一) 简介：直方图工具，用于在给定工作表中数据单元格区域和接收区间的情况下，计算数据的个别和累积频率，可以统计有限集中某个数

值元素的出现次数。例如，在一个有 50 名学生的班级里，可以通过直方图确定考试成绩的分布情况，它会给出考分出现在指定成绩区间的学生个数，而用户必须把存放分段区间的单元地址范围填写在直方图工具对话框中的“接收区域”框中。

(二) 操作步骤：

1. 用鼠标点击表中待分析数据的任一单元格。
2. 选择“工具”菜单的“数据分析”子菜单。
3. 用鼠标双击数据分析工具中的“直方图”选项。
4. 出现“直方图”对话框，对话框内主要选项的含义如下：

输入区域：在此输入待分析数据区域的单元格范围。

接收区域(可选)：在此输入接收区域的单元格范围，该区域应包含一组可选的用来计算频数的边界值。这些值应当按升序排列。只要存在的话，Excel 将统计在各个相邻边界直之间的数据出现的次数。如果省略此处的接收区域，Excel 将在数据组的最小值和最大值之间创建一组平滑分布的接收区间。

标志：如果输入区域的第一行或第一列中包含标志项，则选中此复选框；如果输入区域没有标志项，则清除此该复选框，Excel 将在输出表中生成适宜的数据标志。

输出区域：在此输入结果输出表的左上角单元格的地址。如果输出表将覆盖已有的数据，Excel 会自动确定输出区域的大小并显示信息。

柏拉图：选中此复选框，可以在输出表中同时显示按降序排列频率数据。如果此复选框被清除，Excel 将只按升序来排列数据。

累积百分比：选中此复选框，可以在输出结果中添加一系列累积百分比数值，并同时直方图表中添加累积百分比折线。如果清除此选项，则会省略以上结果。

图表输出：选中此复选框，可以在输出表中同时生成一个嵌入式直方图表。

5. 按需要填写完“直方图”对话框之后，按“确定”按钮即可。

(三) 结果说明：完整的结果通常包括三列和一个频率分布图，第一列是数值的区间范围，第二列是数值分布的频数，第三列是频数分布的累积百分比。

三、利用 Excel 绘制散点图

(一) 简介：散点图是观察两个变量之间关系程度最为直观的工具之一，

利用 Excel 的图表向导，可以非常方便的创建并且改进一个散点图，也可以在一个图表中同时显示两个以上变量之间的散点图。

(二) 操作步骤：数据如图附-3 所示，

	A	B	C 编辑
1	x	y	z
2	68	68	312
3	71	69	323
4	72	70	345
5	70	81	366
6	76	85	378
7	77	86	390
8	76	100	411
9	78	108	434
10	79	114	449
11	81	120	469
12	88	133	480

图 附-3

可按如下步骤建立变量 x-y, x-z 的散点图：

1. 拖动鼠标选定数值区域 A2:C12，不包括数据上面的标志项。
2. 选择“插入”菜单的“图表”子菜单，进入图表向导。
3. 选择“图表类型”为“散点图”，然后单击“下一步”。
4. 确定用于制作图表的数据区。Excel 将自动把你前面所选定的数据区的地址放入图表数据区的内。
5. 在此例之中，需要建立两个系列的散点图，一个是 x-y 系列的散点图，一个是 x-z 系列的散点图，因此，必须单击“系列”标签，确认系列 1 的“X 值”方框与“数值方框”分别输入了 x, y 数值的范围，在系列 2 的“X 值”方框与“数值方框”分别输入了 x, z 数值的范围。在此例中，这些都是 Excel 已经默认的范围，所以，可忽略第 5 步，直接单击“下一步”即可。
6. 填写图表标题为“X-Y 与 X-Z 散点图”，X 轴坐标名称为“X”与 Y 轴坐标名称“Y/Z”，单击“下一步”。
7. 选择图表输出的位置，然后单击“完成”按钮即生成图附-4 的图表。

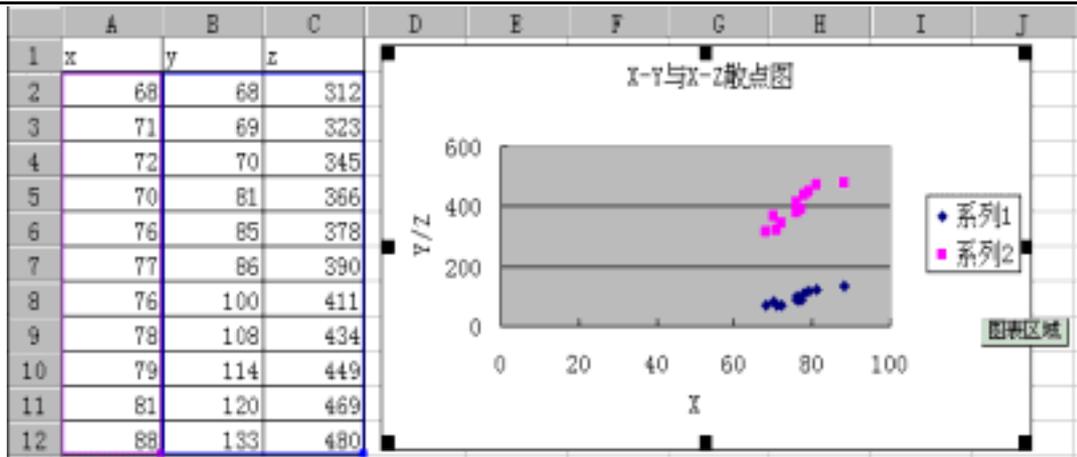


图 附-4

(三) 结果说明：如图附-4 所示，Excel 中可同时生成两个序列的散点图，并分为两种颜色显示。通过散点图可观察出两个变量的关系，为变量之间的建立模型作准备。

四、数据透视表工具

(一) 简介：数据透视表是 Excel 中强有力的数据列表分析工具。它不仅可以用来作单变量数据的次数分布或总和的分析，还可以用来作双变量数据的交叉频数分析、总和分析和其它统计量的分析。

(二) 操作步骤：如图附-5 所示，表中列出学生两门功课评定结果，

	A	B	C	D	E	F
1	学号	语文	数学			
2	1001	优	差			
3	1002	良	中			
4	1003	中	中			
5	1004	差	中			
6	1005	差	差			
7	1006	中	良			
8	1007	中	优			
9	1008	差	良			
10	1009	良	中			

图 附-5

可按如下步骤建立交叉频数表：

1. 选中图附-5 中表格中有数据的任一单元格，然后选择“数据”菜单的“数据透视表”子菜单，进入数据透视表向导。
2. 选择“Microsoft Excel 数据清单或数据库”为数据源。单击“下一步”。
3. 选择待分析的数据的区域，一般情况下 Excel 会自动根据当前单元格确定待分析数据区域，因此你只要直接单击“下一步”按钮即可。
4. 确定数据透视表的结构，在此例中，要建立的是一个交叉频数表，

分别按语文和数学的成绩对学生的人数进行交叉频数分析,因此可按图附-6 将三个按钮“学号”、“语文”、“数学”分别拖放到表格的指定部位,并且双击“求和项:学号”,将其改为记数项,结果如图附-6 所示,然后单击“下一步”按钮。



图 附-6

5. 选择数据透视表的显示位置之后,单击“完成按钮”,可出现如图附-7 所示的数据透视表。

	A	B	C	D	E	F
1	计数项:学号	语文				
2	数学	差	良	优	中	总计
3	差	2		2	2	6
4	良	1	2		2	5
5	优			1	2	3
6	中	2	2		1	5
7	总计	5	4	3	7	19

图 附-7

(三) 结果说明:如图附-7 的结果所示,数据透视表可以作为一个交叉频数分析工具。完成数据透视表之后,可按需要修改数据表的显示格式。例如,如果想要把表格中的频数替换成为百分比数。可以用鼠标右击频数的任一单元格,选择“字段”子菜单,单击“选项”按钮,将“数据显示方式”替换成为“占总和的百分比”,然后单击“确定”按钮即可。按同样方式,可将数据透视表修改成为其它不同样式。

五、排位与百分比工具

(一) 简介:此分析工具可以产生一个数据列表,在其中罗列给定数据集

中各个数值的大小次序排位和相应的百分比排位。用来分析数据集中各数值间的相互位置关系。

(二) 操作步骤：

1. 用鼠标点击表中待分析数据的任一单元格。
2. 选择“工具”菜单的“数据分析”子菜单。
3. 用鼠标双击数据分析工具中的“排位与百分比”选项。
4. 填写完“排位与百分比”对话框，单击“确定”按钮即可。

(三) 结果说明：输出的结果可分为四列，第一列“点”是数值原来的存放位置，第二列是相应的数值，第三列是数值的排序号，第四列是数值的百分比排位，它的计算方法是：小于该数值的数值个数/(数值总个数-1)。

第四节 Excel 在推断统计中的应用

一、二项分布工具

(一) 简介：在 Excel 中想要计算二项分布的概率分布、累积概率，需要利用 Excel 的工作表函数 BINOMDIST。函数 BINOMDIST 适用于固定次数的独立实验，实验的结果只包含成功或失败二种情况，且每次实验成功的概率固定不变。例如，已知次品概率的情况下，函数 BINOMDIST 可以计算 10 个产品中发现 2 个次品的概率。以下例子说明如何在 Excel 中计算二项分布的概率，以及如何建立二项分布图表。

(二) 操作步骤：例子如下所示，一个推销员打了六个电话，推销成功的概率是 0.3，那么可以按以下步骤建立推销成功次数的概率分布图表。

1. 如图附-8 所示，先在 Excel 之下建立好概率分布表格的框架。

	A	B	C	D	E	F
1	二项分布概率分布表					
2	试验总次数	6				
3	每次成功概率	0.3				
4						
5		概率				
6	成功次数(k)	P(Y=k)	P(Y<=k)	P(Y<k)	P(Y>k)	P(Y>=k)
7	0					
8	1					
9	2					
10	3					
11	4					
12	5					
13	6					

图 附-8

2. 如图附-9 所示, 先在 B7 至 F7 单元格分别输入概率计算公式。

	A	B	C	D	E	F
1	二项分布概率分布表					
2	试验总次数	6				
3	每次成功概率	0.3				
4						
5		概率				
6	成功次数(k)	P(Y=k)	P(Y<=k)	P(Y<k)	P(Y>k)	P(Y>=k)
7	0	=BINOMDIST(A7,6,0.3,0)	=BINOMDIST(A7,6,0.3,1)	=C7-B7	=1-C7	=1-D7
8	1					
9	2					
10	3					
11	4					
12	5					
13	6					

图 附-9

3. 公式的拷贝。选取 B7 至 F7 单元格, 拖动“填充柄”至 F13 单元格即可完成公式的拷贝操作。结果图附-10 所示。

	A	B	C	D	E	F
1	二项分布概率分布表					
2	试验总次数	6				
3	每次成功概率	0.3				
4						
5	概率					
6	成功次数(k)	P(Y=k)	P(Y<=k)	P(Y<k)	P(Y>k)	P(Y>=k)
7	0	0.117649	0.117649	0	0.88235	1
8	1	0.302526	0.420175	0.11765	0.57983	0.88235
9	2	0.324135	0.74431	0.42018	0.25569	0.57983
10	3	0.18522	0.92953	0.74431	0.07047	0.25569
11	4	0.059535	0.989065	0.92953	0.01094	0.07047
12	5	0.010206	0.999271	0.98906	0.00073	0.01094
13	6	0.000729	1	0.99927	0	0.00073

图 附-10

- 下面开始创建二项分布图表。选取 B7 至 B13 单元格，选取“插入”菜单的“图表”子菜单。
- 选择“柱状图”，然后单击“下一步”。
- 单击“系列”标签，单击“分类(X)轴标志”框，并用鼠标选取 A7 至 A13 单元格为图表 X 轴的轴标，然后单击“下一步”。
- 分别键入图表名称“二项分布图”，X 轴名称“成功次数”，Y 轴名称“成功概率”，单击“完成”按钮即可生成二项分布图表。

(三) 结果说明: 如图附-10 所示，利用 Excel 的 BINOMDIST 的函数可以计算出二项分布的概率以及累积概率。BINOMDIST 函数可以带四个参数，各参数的含义分别是：实验成功的次数，实验的总次数，每次实验中成功的概率，是否计算累积概率。第四个参数是一个逻辑值，如果为 TRUE，函数 BINOMDIST 返回累积分布函数，如果为 FALSE，返回概率密度函数。

二、其它分布的函数

(一) 函数 CRITBINOM：

- 说明：函数 CRITBINOM 可称为 BINOMDIST 的逆向函数，它返回使累积二项式分布概率 $P(X \leq x)$ 大于等于临界概率值的最小值。
- 语法：CRITBINOM(trials, probability_s, alpha)
 Trials：贝努利实验次数。
 Probability_s：一次试验中成功的概率。
 Alpha：临界概率。
- 举例：CRITBINOM(6, 0.5, 0.75) 等于 4，表明如果每次试验成功的概率为 0.5，那么 6 次试验中成功的次数小于等于 4 的概率恰好超过或等于 0.75。

(二) 函数 HYPGEOMDIST :

1. 说明：函数 HYPGEOMDIST 返回超几何分布。给定样本容量、总体容量和样本总体中成功的次数，函数 HYPGEOMDIST 返回样本取得给定成功次数的概率。使用函数 HYPGEOMDIST 可以解决有限总体的问题，其中每个观察值或者为成功或者为失败，且给定样本区间的所有子集有相等的发生概率。
2. 语法：HYPGEOMDIST(sample_s, number_sample, population_s, number_population)
Sample_s： 样本中成功的次数。
Number_sample： 样本容量。
Population_s： 样本总体中成功的次数。
Number_population： 样本总体的容量。
3. 举例：容器里有 20 块巧克力，8 块是焦糖的，其余 12 块是果仁的。如果从中随机选出 4 块，下面函数计算式计算出只有一块是焦糖巧克力的概率：HYPGEOMDIST(1, 4, 8, 20)= 0.363261。

(三) 函数 NEGBINOMDIST :

1. 说明：函数 NEGBINOMDIST 返回负二项式分布。当每次试验成功概率固时，函数 NEGBINOMDIST 返回在到达指定次数成功之前，出现 n 次失败的概率。此函数与二项式分布相似，只是它的成功次数固定，试验总数为变量。与二项分布类似的是，试验次数被假设为自变量。
2. 语法：NEGBINOMDIST(number_f, number_s, probability_s)
Number_f： 失败次数。
Number_s： 成功的临界次数。
Probability_s： 成功的概率。
3. 举例：例如，如果要找出 5 个反应敏捷的人，且已知具有这种特征的候选人的概率为 0.3。以下公式将计算出在找到 5 个合格候选人之前，需要面试 10 个候选人的概率：
NEGBINOMDIST(10, 5, 0.3)= 0.06871

(四) 函数 POISSON:

1. 说明：函数 POISSON 返回泊松分布。泊松分布通常用于预测一段时间内事件发生指定次数的概率，比如一分钟内通过收费站的轿车的数量为 n 的概率。
2. 语法：POISSON(x, mean, cumulative)

X： 事件数。

Mean： 期望值。

Cumulative： 为一逻辑值，确定所返回的概率分布形式。如果 cumulative 为 TRUE，函数 POISSON 返回累积分布函数，即，随机事件发生的次数在 0 和 x 之间（包含 0 和 1）；如果为 FALSE，则返回概率密度函数，即，随机事件发生的次数恰好为 x。

3. 举例：POISSON(2, 5, FALSE)=0.084224 表明，若某一收费站每分钟通过的轿车平均数量为 5 辆，那么某一分钟通过 2 辆的概率为 0.084224。

(五) 正态分布函数 NORMDIST：

1. 说明：正态分布在模拟现实世界过程和描述随机样本平均值的不确定度时有广泛的用途。函数 NORMDIST 返回给定平均值和标准偏差的正态分布的累积函数。同样可以用类似“七”中的方法，利用 NORMDIST 函数建立正态分布密度函数图，这里不再赘述。

2. 语法：NORMDIST(x, mean, standard_dev, cumulative)

X： 为需要计算其分布的数值。

Mean： 分布的算术平均值。

Standard_dev： 分布的标准偏差。

Cumulative： 为一逻辑值，指明函数的形式。如果 cumulative 为 TRUE，函数 NORMDIST 返回累积分布函数；如果为 FALSE，返回概率密度函数。

3. 举例：例如，公式 NORMDIST(6, 5, 2, 0) 返回平均值为 5、标准差为 2 的正态函数当 X=6 时概率密度函数的数值，公式 NORMDIST(60, 50, 4, 1) 返回平均值为 50、标准差为 4 的正态分布函数当 X=60 时累积分布函数的数值。

(六) 函数 NORMSDIST：

1. 说明：函数 NORMSDIST 返回标准正态分布的累积函数。

2. 语法：NORMSDIST(z)

Z 为需要计算其分布的数值。

3. 举例：NORMSDIST(0)=0.5

(七) 函数 NORMSINV：

1. 说明：函数 NORMSINV 返回标准正态分布累积函数的逆函数。

2. 语法：NORMSINV(probability)

Probability：正态分布的概率值。

3. 举例：NORMSINV(0.5)=0

(八) t 分布函数 TDIST:

1. 说明：函数 TDIST 返回 student 的 t 分布数值。T 分布用于小样本数据集合的假设检验。使用此函数可以代替 t 分布的临界值表。

2. 语法：TDIST(x, degrees_freedom, tails)

X：为需要计算分布的数字。

Degrees_freedom：为表示自由度的整数。

Tails：指明返回的分布函数是单尾分布还是双尾分布。如果 tails = 1，函数 TDIST 返回单尾分布。如果 tails = 2，函数 TDIST 返回双尾分布。

3. 举例：TDIST(1.96, 60, 2)=0.054645

三、随机抽样的工具

(一) 简介：Excel 中的 Rand() 函数可以返回大于等于 0 小于 1 的均匀分布随机数，Rand() 不带任何参数运行，每次计算时时都将返回一个新的数值。RAND() 函数可以被用来作为不重复抽样调查的工具。

(二) 操作步骤：如图附-11 所示有 10 个象征性的样本数据，欲从中随机抽取 5 个数据可按如下步骤操作：

	A
1	No
2	1
3	2
4	3
5	4
6	5
7	6
8	7
9	8
10	9
11	10

图 附-11

1. 选择 B2 单元格，输入公式 “=RAND()” 并回车。
2. 拖动 B2 单元格右下角的填充柄至 B11 单元格，并在 B1 单元格输入标题 “RANDOM”。
3. 选取单元格 B2 至 B11，右击选中的区域选择 “复制”，再次右击选

中的区域，选择“选择性粘贴”，单击选项“数值”后，点击“确定”按钮。

4. 选取单元格 A2 至 B11 单元格，选择“数据”菜单项下的排序子菜单。
5. 选取“RANDOM”为主要关键字，然后点击“确定”按钮。排序结果如图附-12 所示，A2 至 A6 单元格的样本即为随机抽取的 5 个样本。

	A	B
1	No	random
2	10	0.216688
3	3	0.234034
4	1	0.302342
5	8	0.437267
6	9	0.610631
7	7	0.64232
8	4	0.656722
9	2	0.68924
10	5	0.882674
11	6	0.953918

图 附-12

(三) 结果说明：

1. 以上进行的是不重复随机抽样，可以用类似的方法，利用 Excel 的 RANDBETWEEN(TOP, BOTTOM) 函数实现总体的重复随机抽样。RANDBETWEEN(TOP, BOTTOM) 函数可随机返回介于 TOP 与 BOTTOM 之间的整数，抽取此整数对编号的样本可作为总体的重复随机抽样的结果。
2. RAND() 函数返回的是 0 与 1 之间均匀的随机数，利用 Excel 数据分析工具中的随机数发生器，可以生成用户指定类型分布的随机数。例如 0-1 正态分布的随机数，指定参数的泊松分布的随机数等。
3. 利用 Excel 易于产生各类型随机数的特性，可以用类似的方法方便地进行随机数字模拟试验与随机游走模拟试验。

四、由样本推断总体

- (一) 简介：利用 Excel 的几个函数，如求平均函数 AVERAGE、标准差函数 STDEV、T 分布函数 TINV 等的组合使用可以构造出一个专门用于实现

样本推断总体的 Excel 工作表。以下例子先计算样本的平均数和标准差，然后在一定置信水平上估计总体均值的区间范围。

(二) 操作步骤：

1. 构造工作表。如图附-13 所示，首先在各个单元格输入以下内容，其中左边是变量名，右边是相应的计算公式。
2. 为表格右边的公式计算结果定义左边的变量名。选定 A4: B6, A8: B8 和 A10: B15 单元格(先选择第一部分，再按住 CTRL 键选取另外两个部分)，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“最左列”选项，然后点击“确定”按钮即可。

	A	B
1		以样本均值推断总体均值的置信区间
2		
3	样本统计量	
4	样本个数	=COUNT(样本数据)
5	样本均值	=AVERAGE(样本数据)
6	样本标准差	=STDEV(样本数据)
7	用户输入	
8	置信水平	0.95
9	计算结果	
10	抽样标准误	= '样本标准差' /SQRT('样本个数')
11	自由度	= '样本个数' -1
12	t值	=TINV(1-'置信水平', '自由度')
13	置信区间半径	= 't值' * '抽样标准误'
14	置信区间上界	= '样本均值' - '置信区间半径'
15	置信区间下界	= '样本均值' + '置信区间半径'

图 附-13

3. 输入样本数据，和用户指定的置信水平 0.95, 如图附-13 所示。
4. 为样本数据命名。选定 D1: D11 单元格，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“首行”选项，然后点击“确定”按钮，得到图附-14 所示的计算结果。

	A	B	C	D
1	以样本均值推断总体均值的置信区间			样本数据
2				28.5
3	样本统计量			26.4
4	样本个数	10		33.5
5	样本均值	31.4		34.3
6	样本标准差	2.814249456		35.9
7	用户输入			29.6
8	置信水平	0.95		31.3
9	计算结果			31.1
10	抽样标准误	0.889943818		30.9
11	自由度	9		32.5
12	t值	2.262158887		
13	置信区间半径	2.013194318		
14	置信区间上界	29.38680568		
15	置信区间下界	33.41319432		

图 附-14

(三) 结果说明：以上例子说明如何交叉组合使用 Excel 的公式和函数，以构造出一个能实现样本推断总体有关计算的 Excel 工作表。实际上，在用 Excel 进行数据统计处理之时，许多统计功能可以使用和上例类似的方法，通过组合使用 Excel 的各类统计函数和公式加以实现的。

五、假设检验

(一) 简介：假设检验是统计推断中的重要内容。以下例子利用 Excel 的正态分布函数 NORMSDIST、判断函数 IF 等，构造一张能够实现在总体方差已知情况下进行总体均值假设检验的 Excel 工作表。

(二) 操作步骤：

1. 构造工作表。如图附-15 所示，首先在各个单元格输入以下的内容，其中左边是变量名，右边是相应的计算公式。
2. 为表格右边的公式计算结果定义左边的变量名。选定 A3: B4, A6: B8, A10: A11, A13: A15 和 A17: B19 单元格，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“最左列”选项，然后点击“确定”按钮即可。

	A	B
1		总体均值的假设检验
2	样本统计量	
3	样本个数	=COUNT(样本数据)
4	样本均值	=AVERAGE(样本数据)
5	用户输入	
6	总体标准差	
7	总体均值假设值	
8	置信水平	
9	计算结果	
10	抽样标准误	= '总体标准差' /SQRT('样本个数')
11	计算Z值	= ('样本均值' - '总体均值假设值') / '抽样标准误'
12	单侧检验	
13	单侧Z值	=NORMSINV(1-'置信水平')
14	检验结果	=IF(ABS('计算Z值')>ABS('单侧Z值'),"拒绝Ho","接收Ho")
15	单侧显著水平	=1-NORMSDIST(ABS('计算Z值'))
16	双侧检验	
17	双侧Z值	=NORMSINV((1-'置信水平')/2)
18	检验结果	=IF(ABS('计算Z值')>ABS('双侧Z值'),"拒绝Ho","接收Ho")
19	双侧显著水平	=IF('计算Z值'>0, 2*(1-NORMSDIST('计算Z值')), 2*NORMSDIST('计算Z值'))

图 附-15

2. 输入样本数据，以及总体标准差、总体均值假设、置信水平数据。如图附-16 所示。
3. 为样本数据命名。选定 C1: C11 单元格，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“首行”选项，然后点击“确定”按钮，得到如图附-16 中所示的计算结果。

	A	B	C
1		总体均值的假设检验	样本数据
2	样本统计量		28.5
3	样本个数	10	26.4
4	样本均值	31.4	33.5
5	用户输入		34.3
6	总体标准差	5.56	35.9
7	总体均值假设值	35	29.6
8	置信水平	0.95	31.3
9	计算结果		31.1
10	抽样标准误	1.758226379	30.9
11	计算Z值	-2.047517909	32.5
12	单侧检验		
13	单侧Z值	-1.644853	
14	检验结果	拒绝Ho	
15	单侧显著水平	0.020303562	
16	双侧检验		
17	双侧Z值	-1.959961082	
18	检验结果	拒绝Ho	
19	双侧显著水平	0.040607125	

图 附-16

(三) 结果说明：如图附-16 所示，该例子的检验结果不论是单侧还是双侧均为拒绝 H_0 假设。所以，根据样本的计算结果，在 5% 的显著水平之下，拒绝总体均值为 35 的假设。同时由单侧显著水平的计算结果还可以看出，在总体均值是 35 的假设之下，样本均值小于等于 31.4 的概率仅为 0.020303562。

六、双样本等均值假设检验

(一) 简介：双样本等均值检验是在一定置信水平之下，在两个总体方差相等的假设之下，检验两个总体均值的差值等于指定平均差的假设是否成立的检验。我们可以直接使用在 Excel 数据分析中提供双样本等均值假设检验工具进行假设检验。以下通过一例说明双样本等均值假设检验的操作步骤。例子如下，某工厂为了比较两种装配方法的效率，分别组织了两组员工，每组 9 人，一组采用新的装配方法，另外一组采用旧的装配方法。18 个员工的设备装配时间图附-17 中表格所示。根据以下数据，是否有理由认为新的装配方法更节约时间？

	A	B	C	D
1	组别	旧方法装配时间	组别	新方法装配时间
2	1	32	2	35
3	1	37	2	31
4	1	35	2	29
5	1	38	2	25
6	1	41	2	34
7	1	44	2	40
8	1	35	2	27
9	1	31	2	32
10	1	34	2	31

图 附-17

(二) 操作步骤：以上例子可按如下步骤进行假设检验。

1. 选择“工具”菜单的“数据分析”子菜单，双击“t-检验：双样本等方差假设”选项，则弹出图附-18 所示对话框。

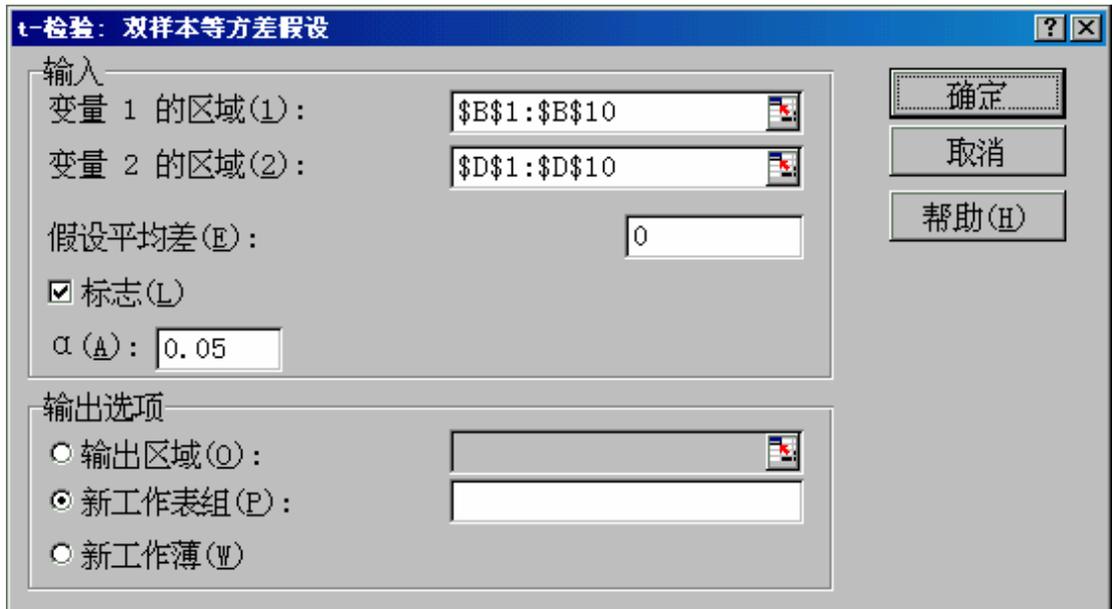


图 附-18

2. 分别填写变量 1 的区域：\$B\$1:\$B\$10，变量 2 的区域：\$D\$1:\$D\$10，由于我们进行的是等均值的检验，填写假设平均差为 0，由于数据的首行包括标志项选择标志选项，所以选择“标志”选项，再填写显著水平 为 0.05, 然后点击“确定”按钮。则可以得到图附-19 所示的结果。

(三) 结果分析：如图附-19 中所示，表中分别给出了两组装配时间的平均值、方差和样本个数。其中，合并方差是样本方差加权之后的平均值, Df 是假设检验的自由度它等于样本总个数减 2，t 统计量是两个样本差值减去

	A	B	C
1	t-检验：双样本等方差假设		
2			
3		旧方法装配时间	新方法装配时间
4	平均	35.22222222	31.55555556
5	方差	24.44444444	20.02777778
6	观测值		9
7	合并方差	22.23611111	
8	假设平均差		0
9	df		16
10	t Stat	1.649484617	
11	P(T<=t) 单尾	0.059269899	
12	t 单尾临界	1.745884219	
13	P(T<=t) 双尾	0.118539799	
14	t 双尾临界	2.119904821	

图 附-19

假设平均差之后再除于标准误差的结果，“P(T<=t)单尾”是单尾检验的显著水平，“t 单尾临界”是单尾检验 t 的临界值，“P(T<=t)双尾”是双尾检验的显著水平，“t 双尾临界”是双尾检验 t 的临界值。由下表的结果可以看出 t 统计量均小于两个临界值，所以，在 5%显著水平下，不能拒绝两个总体均值相等的假设，即两种装配方法所耗时间没有显著的不同。

Excel 中还提供了以下类似的假设检验的数据分析工具，它们的名称和作用如下：

1. “t-检验：双样本异方差假设”：此分析工具可以进行双样本 student t-检验，与双样本等方差假设检验不同，该检验是在两个数据集的方差不等的前提假设之下进行两总体均值差额的检验，故也称作异方差 t-检验。可以使用 t-检验来确定两个样本均值实际上是否相等。当进行分析的样本个数不同时，可使用此检验。如果某一样本组在某次处理前后都进行了检验，则应使用“成对检验”。
2. “t-检验：成对双样本均值分析”：此分析工具可以进行成对双样本学生氏 t-检验，用来确定样本均值是否不等。此 t-检验并不假设两个总体的方差是相等的。当样本中出现自然配对的观察值时，可以使用此成对检验，例如，对一个样本组进行了两次检验，抽取实验前的一次和实验后的一次。
3. “z - 检验：双样本均值分析”：此分析工具可以进行方差已知的双样本均值 z - 检验。此工具用于检验两个总体均值之间存在差异的假设。例如，可以使用此检验来确定两种汽车模型性能之间的差异情况。

七、正态性的 X^2 检验

(一) 简介： X^2 检验可以用来判断所观测的样本是否来自某一特定分布的总体，这种检验亦称为一致性检验。以下例子，已知某样本的相关统计量和分组频数分布如图附-20 所示，试图用 X^2 检验判断该样本是否来自一正态总体。

	A	B	C
1	正态性的 X^2 检验		
2			
3	样本个数	样本均值	样本标准差
4	200	164	10
5			
6	分组下界	分组上界	真实频数
7		150	15
8	150	160	54
9	160	170	78
10	170	180	42
11	180		11
12		累积值	200

图 附-20

(二) 操作步骤

1. 创建变量名。选定 A3:C4 单元格，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“首行”选项，然后点击“确定”按钮即可。
2. 计算预期正态概率值。如图附-21 表中所示，在 D6 单元格输入标志项，在 D7:D11 单元格输入公式，分别计算各组的预期正态概率值，在 D12 计算累积概率值。

	D
6	预期正态概率
7	=NORMDIST(B7, 样本均值, 样本标准差, 1)
8	=NORMDIST(B8, 样本均值, 样本标准差, 1)-NORMDIST(A8, 样本均值, 样本标准差, 1)
9	=NORMDIST(B9, 样本均值, 样本标准差, 1)-NORMDIST(A9, 样本均值, 样本标准差, 1)
10	=NORMDIST(B10, 样本均值, 样本标准差, 1)-NORMDIST(A10, 样本均值, 样本标准差, 1)
11	=1-NORMDIST(A11, 样本均值, 样本标准差, 1)
12	=SUM(D7:D11)

图 附-21

3. 计算预期频数值。如图附-22 所示，在 E6 单元格输入标志项，在 E7:E11 单元格输入公式，分别计算各组的预期频数，在 E12 计算累积频数值。

	E
6	预期频数值
7	=D7*样本个数
8	=D8*样本个数
9	=D9*样本个数
10	=D10*样本个数
11	=D11*样本个数
12	=SUM(E7:E11)

图 附-22

4. 计算 X^2 统计量。如图附-23 所示，在 F6 单元格输入标志项，在 F7:F11 分别输入计算公式，分别计算 X^2 值，在 E12 计算 X^2 平方和，这项就是最后计算出的 X^2 统计量。在 E13 单元格输入标志项“卡方统计量”，为以后的引用作准备。先选中 F12、F13 两个单元格，选择“插入”菜单的“名称”子菜单的“指定”选项，用鼠标点击“尾行”选项，然后点击“确定”按钮即可。

	F
1	
2	
3	
4	
5	
6	卡方值
7	=(C7-E7)^2/E7
8	=(C8-E8)^2/E8
9	=(C9-E9)^2/E9
10	=(C10-E10)^2/E10
11	=(C11-E11)^2/E11
12	=SUM(F7:F11)
13	卡方统计量

图 附-23

5. 如图附-24 所示，分别在 A14 到 B20 单元格输入自由度、 X^2 概率值、置信水平、临界值、 X^2 检验结果几项的标志值及计算公式。其中的自由度=区间分段数-正太分布参数个数-1=5-2-1=2。

	A	B
14	自由度	2
15	卡方概率值	=CHIDIST(卡方统计量,'自由度')
16		
17	置信水平	0.01
18	临界值	=CHIINV('置信水平','自由度')
19	卡方检验结果	=IF(卡方统计量>临界值,"拒绝总体为正太分布的假设","接受总体为正太分布的假设")

图 附-24

(三) 结果分析：如图附-25 所示，按照以上操作步骤可以得到表中的计算结果。

	A	B	C	D	F
1		正态性的x2检验			
2					
3	样本个数	样本均值	样本标准差		
4	200	164	10		
5					
6	分组下界	分组上界	真实频数	预期正态概率	卡方值
7		150	15	0.080756711	0.082073
8	150	160	54	0.263821592	0.0289383
9	160	170	78	0.381168632	0.0409231
10	170	180	42	0.219453775	0.0814512
11	180		11	0.054799289	0.000147
12		累积值	200	1	0.2335326
13					卡方统计量
14	自由度	2			
15	卡方概率值	0.889793097			
16					
17	置信水平	0.01			
18	临界值	9.210351036			
19	卡方检验结果	接受总体为正太分布的假设			

图 附-25

按同样的方法可以作总体泊松分布、总体超几何分布等其它分布的检验。此类统计应用也是由 Excel 各类公式和函数综合使用而实现的, 为了以后使用方便, 和上面的一些例子一样, 一般需要将整个表格的计算框架和标志项罗列好, 再保存成文件, 以后只要对数据项稍作修改即可很快得到计算结果。如果对 Excel 宏语言较为熟悉, 还可以将它编成一个宏语言程序, 加入 Excel 的工具栏, 这样以后使用起来更为方便。

八、列联表分析

(一) 简介：列联表分析经常用来判断同一个调查的对象两个特性之间是否存在明显相关关系。例如，房地产商常常设计列联表问卷，调查顾客的职业和顾客所选房子的类型有否有明显的相关关系。列联表分析同样也可以由 Excel 加以实现，下面用一个例子给予说明。如图附-26 所示，表中是某装修公司的调查报告，试用列联表分析方法分析在顾客的所在地区和所选房子的地板类型之间是否存在明显的相关关系。

	A	B	C	D	E	F
1	列联表分析					
2						
3	真实频数	地区1	地区2	地区3	地区4	行总数
4	木质地板	72	8	12	23	115
5	钢砖地板	26	10	16	33	85
6	大理石地板	7	10	14	19	50
7	列总数	105	28	42	75	250

图 附-26

(二) 操作步骤

1. 建立期望频数表。如图附-27 所示，先建立期望频数表的框架，然后在 B10 单元格输入公式“=B\$7*\$F4/\$F\$7”，再利用“填充柄”将公式复制到表格的其它单元格，最后利用 Excel 的求和函数 sum 计算行和与列和。

	A	B	C	D	E	F
9	期望频数	地区1	地区2	地区3	地区4	总数
10	木质地板	=B\$7*\$F4/\$F\$7	=C\$7*\$F4/\$F\$7	=D\$7*\$F4/\$F\$7	=E\$7*\$F4/\$F\$7	=SUM(B10:E10)
11	钢砖地板	=B\$7*\$F5/\$F\$7	=C\$7*\$F5/\$F\$7	=D\$7*\$F5/\$F\$7	=E\$7*\$F5/\$F\$7	=SUM(B11:E11)
12	大理石地板	=B\$7*\$F6/\$F\$7	=C\$7*\$F6/\$F\$7	=D\$7*\$F6/\$F\$7	=E\$7*\$F6/\$F\$7	=SUM(B12:E12)
13	总数	=SUM(B10:E12)	=SUM(C10:C12)	=SUM(D10:D12)	=SUM(E10:E12)	=SUM(B13:E13)

图 附-27

2. 计算 X^2 概率值。在 A15 单元格输入标志项“卡方概率值”，先点击 B15 单元格，从“插入”菜单中“函数”子菜单，选择“统计函数”中的“CHITEST”函数，单击“确定按钮”，然后在弹出的对话框中分别添入实际频数范围“B4:E6”和预期频数范围“B10:E12”。最后单击“确定”按钮即可得到计算结果 1.3E-07，如图附-28 所示。

	A	B	C	D	E	F
1	列联表分析					
2						
3	真实频数	地区1	地区2	地区3	地区4	行总数
4	木质地板	72	8	12	23	115
5	钢砖地板	26	10	16	33	85
6	大理石地板	7	10	14	19	50
7	列总数	105	28	42	75	250
8						
9	期望频数	地区1	地区2	地区3	地区4	总数
10	木质地板	48.3	12.88	19.32	34.5	115
11	钢砖地板	35.7	9.52	14.28	25.5	85
12	大理石地板	21	5.6	8.4	15	50
13	总数	105	28	42	75	250
14						
15	卡方概率值	1E-07				

图 附-28

3. 建立 χ^2 统计表。如图附-29 所示，先建立表格的框架，然后在 B20 单元格输入公式 “ $= (B4-B10)^2/B10$ ”，再利用填充柄将公式复制到表格的其它单元格。最后计算 χ^2 卡方统计量，分别在 A24 与 B24 单元输入标志项与计算公式。

	A	B	C	D	E
19		地区1	地区2	地区3	地区4
20	木质地板	$= (B4-B10)^2/B10$	$= (C4-C10)^2/C10$	$= (D4-D10)^2/D10$	$= (E4-E10)^2/E10$
21	钢砖地板	$= (B5-B11)^2/B11$	$= (C5-C11)^2/C11$	$= (D5-D11)^2/D11$	$= (E5-E11)^2/E11$
22	大理石地板	$= (B6-B12)^2/B12$	$= (C6-C12)^2/C12$	$= (D6-D12)^2/D12$	$= (E6-E12)^2/E12$
23					
24	卡方统计量	$= \text{SUM}(B20:E22)$			

图 附-29

4. 进行假设检验。如图附-30 所示，分别输入置信水平、临界值和假设检验的结果其中 CHIINV 函数的自由度= $(\text{第一类属性的分类数}-1) * (\text{第二类属性的分类数}-1) = (3-1) * (4-1) = 6$ 。

	A	B
26	置信水平	0.01
27	临界值	$= \text{CHIINV}(B26, 6)$
28	检验结果	$= \text{IF}(\text{卡方统计量} > \text{临界值}, \text{“拒绝两种属性不相关的假设”}, \text{“接受两种属性不相关的假设”})$

图 附-30

(三) 结果分析：以上的操作步骤完成整个列联表的分析。其中，B15 单元格的卡方概率值与 B24 单元格的卡方统计量是表格的两个重要计算结果。其中卡方概率值等于 $1.3E-07$ 表明：如果总体的两类属性，即所在地区和所选地板类型，是不相关的，那么得到以上观察的样本的概率是 0.00000013 。这个概率几乎接近于 0，所以可以认为总体的这两个属性是显著相关的。

九、单因素方差分析

(一) 简介：单因素方差分析可用于检验两个或两个以上的总体均值相等的假设是否成立。此方法是对双均值检验（如 t-检验）的扩充。检验假定总体是服从正太分布的，总体方差是相等的，并且随机样本是独立的。这种工具试用于完全随机化试验的结果分析。例子如图附-31 表中所示，一产品制造商雇佣销售人员向销售商打电话。制造商想比较四种不同电话频率计划的效率，他从销售人员中随机选出 32 名，将他们随机分配到 4 种计划中，在一段时期内记录他们的销售情况已经在表中列出，试问其中是否有一种计划会带来较高的销售水平。

	A	B	C	D
1	单因素方差分析			
2				
3	计划1	计划2	计划3	计划4
4	36	39	44	31
5	40	45	43	43
6	32	54	38	46
7	44	53	40	43
8	35	46	41	36
9	41	42	35	49
10	44	35	37	46
11	42	39	37	48

图 附-31

(二) 操作步骤

1. 选择“工具”菜单的“数据分析”子菜单，双击“方差分析：单因素方差分析”选项，弹出单因素方差分析对话框。
2. 按图附-32 所示方式填写对话框。然后单击“确定”按钮。

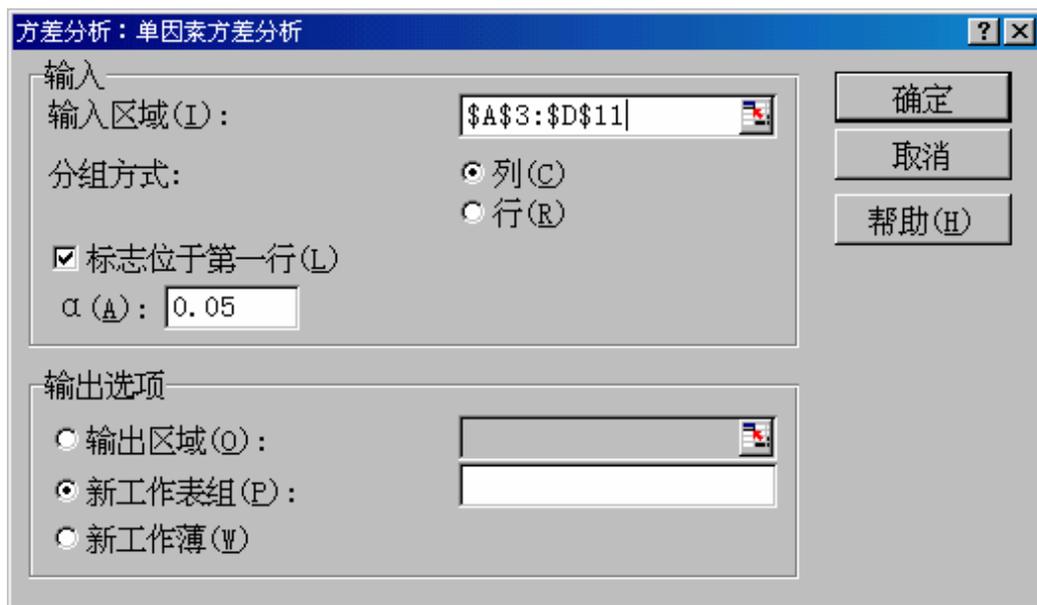


图 附-32

(三) 结果分析：按照如上的操作步骤即可得到图附-33 的计算结果。其中表格的第二部分则是方差分析的结果。SS 列分别给出了四个分组的组间方差、组内方差以及总方差，DF 列分别给出了对应方差的自由度，MS 列是平均值方差，由 SS 除以 DF 得到，它是总体方差的两个估计值。F 列是 F 统计量的计算结果，如果四个总体均值相等的假设成立的化，它应该服从 F 分布，即近似为 1，它是最终的计算结果，通过将

	A	B	C	D	E	F	G
1	方差分析：单因素方差分析						
2							
3	SUMMARY						
4	组	计数	求和	平均	方差		
5	计划1	8	314	39.25	19.6429		
6	计划2	8	353	44.125	45.8393		
7	计划3	8	315	39.375	9.98214		
8	计划4	8	342	42.75	38.7857		
9							
10							
11	方差分析						
12	差异源	SS	df	MS	F	P-value	F crit
13	组间	143.75	3	47.9167	1.67761	0.19442	2.94668
14	组内	799.75	28	28.5625			
15							
16	总计	943.5	31				

图 附-33

它与一定置信水平下的 F 临界值 F_{crit} 比较，可以判断均值相等的假设是否成立，在本例中， $1.67761 < 2.94668$ ，所以不能拒绝四个总体均值相等的假设。P-value 列，是单尾概率值，表明如果四个总体均值相等的假设成立的化，得到如上样本结果的概率是 19.442%，即得到以上样本并不是小概率事件，同样也得到不能拒绝四个总体均值相等的假设的结论。

按相似方法可进行无重复双因素方差分析，有重复双因素方差分析。

十、线性回归分析

(一) 简介：线性回归分析通过对一组观察值使用“最小二乘法”直线拟合，用来分析单个因变量是如何受一个或几个自变量影响的。例子如图附-34 所示，表中是我国 1987 年至 1997 年的布匹人均产量和人均纱产量，试用线性回归分析的方法分析两组数据之间的关系。

	A	B	C
1	年份	人均布产量(米)	人均纱产量(公斤)
2	1987	15.96	4.03
3	1988	17.06	4.23
4	1989	16.92	4.26
5	1990	16.63	4.07
6	1991	15.79	4
7	1992	16.37	4.31
8	1993	17.23	4.26
9	1994	17.73	4.11
10	1995	21.59	4.5
11	1996	17.17	4.21
12	1997	20.23	4.55

图 附-34

(二) 操作步骤

1. 选择“工具”菜单的“数据分析”子菜单，双击“回归”选项，弹出回归分析对话框。其中主要选项的含义如下：Y 值输入区域，在此输入对因变量数据区域，该区域必须由单列数据组成；X 值输入区域，在此输入对自变量数据区域，Excel 将对此区域中的自变量从左到右按升序排列，自变量的个数最多为 16；置信度，如果需要在汇总输出表中包含附加的置信度信息，则选中此复选框，然后在右侧的编辑框中，输入所要使用的置信度，95%为默认值；常数为零，如果要强制回归线通过原点，则选中此复选框；输出区域，在此输入对输出表左上角单元格的引用。汇总输出表至少需要有七列的宽度，包含的内容有 anova 表、系数、y 估计值的标准误差、 r^2 值、观察值个数，以及系数的标准误差；新工作表，单击此选项，可在当前工作簿中插入新工作表，并由新工作表的 A1 单元格开始粘贴计算结果，如果需要给新工作表命名，则在右侧的编辑框中键入名称；新工作簿，单击此选项，可创建一新工作簿，并在新工作簿中的新工作表中粘贴计算结果；残差，如果需要以残差输出表的形式查看残差，则选中此复选框；标准残差，如果需要在残差输出表中包含标准残差，则选中此复选框；残差图，如果需要生成一张图表，绘制每个自变量及其残差，则选中此复选框；线形拟合图，如果需要为预测值和观察值生成一个图表，则选中此复选框；正态概率图，如果需要绘制正态概率图，则选中此复选框。

2. 按如下方式填写对话框：X 值输入区域为 \$B\$1:\$B\$12，Y 值输入区域为 \$C\$1:\$C\$12，并选择“标志”和“线性拟合图”两个复选框，然后单击“确定”按钮即可。

(三) 结果分析

按照如上的操作步骤即可得到图附-35 下表的计算结果。结果可以分为四个部分，第一部分是回归统计的结果包括多元相关系数、可决系数 R^2 、调整之后的相关系数、回归标准差以及样本个数。第二部分是方差分析的结果包括可解释的离差、残差、总离差和它们的自由度以及由此计算出的 F 统计量和相应的显著水平。第三部分是回归方程的截距和斜率的估计值以及它们的估计标准误差、t 统计量大小双边拖尾概率值、以及估计值的上下界。根据这部分的结果可知回归方程为

$Y=8.46433 \cdot X-18.288$ 。第四部分是样本散点图，其中蓝色的点是样本的真实散点图，红色的点是根据回归方程进行样本历史模拟的散点。如果

觉得散点图不够清晰可以用鼠标拖动图形的边界达到控制图形大小的目的。用相同的方法可以进行多元线性方程的参数估计，还可以在自变量中引入虚拟变量以增加方程的拟合程度。对于非线性的方程的参数

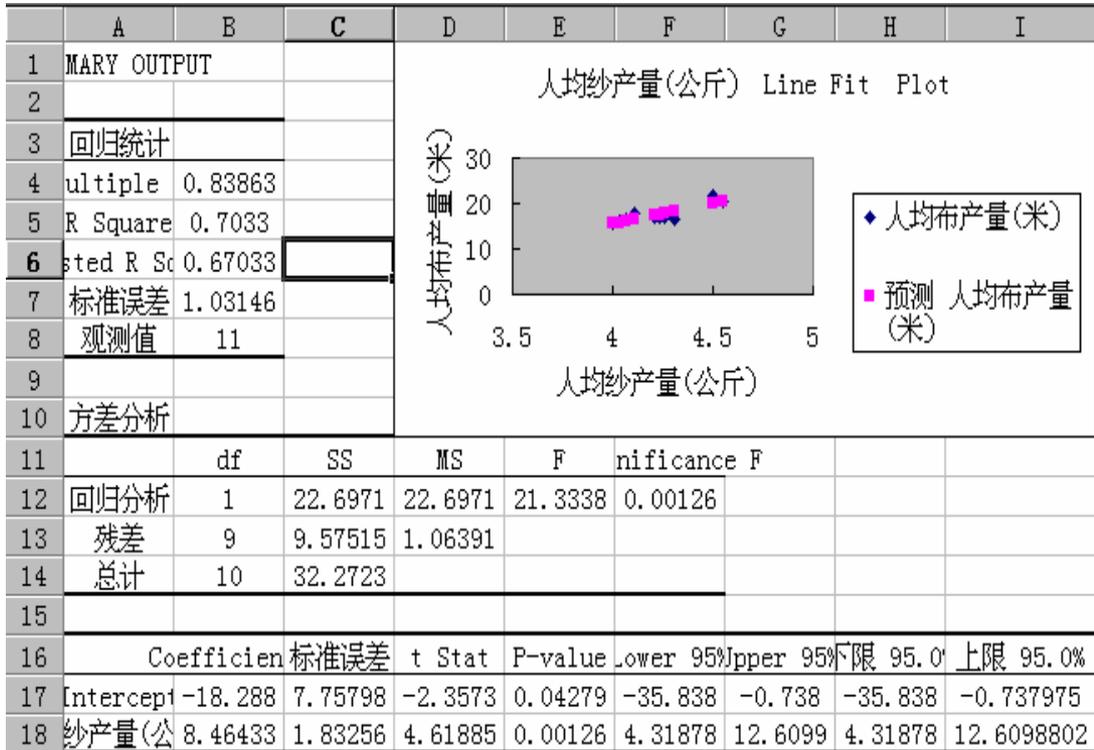


图 附-35

估计，可以在进行样本数据的线性化处理之后，再按以上步骤进行参数估计。

十一、相关系数分析工具

(一)简介：此分析工具可用于判断两组数据之间的关系。可以使用“相关系数”分析工具来确定两个区域中数据的变化是否相关，即，一个集合的较大数据是否与另一个集合的较大数据相对应(正相关)；或者一个集合的较小数据是否与另一个集合的较小数据相对应(负相关)；还是两个集合中的数据互不相关(相关系数为零)。

(二)操作步骤：采用图附-3表中的数据，可按如下步骤计算变量 x, y, z 之间的相关系数。

1. 用鼠标点击表中待分析数据的任一单元格。
2. 选择“工具”菜单的“数据分析”子菜单。
3. 用鼠标双击数据分析工具中的“相关系数”选项。
4. 填写完“相关系数”对话框，单击“确定”按钮即可得到各个变量的相关系数矩阵，结果如图附-36所示。

	A	B	C	D
1		x	y	z
2	x	1		
3	y	0.929167	1	
4	z	0.922982	0.984245	1

图 附-36

(三) 结果说明：以上下三角矩阵计算出三个变量 x, y, z 两两之间的相关系数，如变量 x, y 之间的相关系数为：0.929167，所以可以判断 x, y 之间存在着较高的正线性相关关系。

十二、协方差分析工具

协方差分析的操作步骤同“六”的相关系数分析较为相似，只须在第 3 步中将“相关系数”选项替换成为“协方差”选项即可。

十三、自回归模型的识别与估计

(一) 简介：时间序列分析已广泛的应用于科研，生产，社会生活的方方面面。它通过对时间序列作统计规律性的分析，构造出拟合时间序列的最佳模型。构造出的时间序列模型一方面浓缩了时间序列的信息，可简化对时间序列的表示，另一方面可以用来预测时间序列未来的可能取值，作为人们科学决策的重要依据。例子如图附-37 所示，表中是自 1999 年 4 月 1 日起的 20 个交易日内的上证指数的时间序列，试用自回归模型加以拟合。

(二) 操作步骤

1. 数据的零均值化处理。如图附-37 中所示，在 C1 中输入序列名“Z”，在 C2 中输入公式“=上证指数-AVERAGE(上证指数)”，然后在 C2 单元格中，拖动 Excel “填充柄”将公式复制到 C3 至 C22 单元格，即可生成上证指数的零均值化序列。
2. 计算自相关函数。在 E1 和 F1 单元格分别输入标志项 Lag 和 ac，

	A	B	C
1	日期	上证指数	Z
2	36251	1168.5	=上证指数-AVERAGE(上证指数)
3	36252	1184.4	=上证指数-AVERAGE(上证指数)
4	36255	1194.51	=上证指数-AVERAGE(上证指数)
5	36256	1199.6	=上证指数-AVERAGE(上证指数)
6	36257	1201.63	=上证指数-AVERAGE(上证指数)
7	36258	1202.47	=上证指数-AVERAGE(上证指数)
8	36259	1205.14	=上证指数-AVERAGE(上证指数)
9	36262	1195.34	=上证指数-AVERAGE(上证指数)
10	36263	1178.19	=上证指数-AVERAGE(上证指数)
11	36264	1181.52	=上证指数-AVERAGE(上证指数)
12	36265	1170.62	=上证指数-AVERAGE(上证指数)
13	36266	1166.24	=上证指数-AVERAGE(上证指数)
14	36269	1159.47	=上证指数-AVERAGE(上证指数)
15	36270	1171.6	=上证指数-AVERAGE(上证指数)
16	36271	1144.06	=上证指数-AVERAGE(上证指数)
17	36272	1137.16	=上证指数-AVERAGE(上证指数)
18	36273	1140	=上证指数-AVERAGE(上证指数)
19	36276	1112.79	=上证指数-AVERAGE(上证指数)
20	36277	1091.69	=上证指数-AVERAGE(上证指数)
21	36278	1091.09	=上证指数-AVERAGE(上证指数)

图 附-37

在 E2 到 E9 单元格中分别输入置后期数 1 至 8。在 F2 单元格输入计算自相关函数的公式 “=SUMPRODUCT(OFFSET(C\$2,0,0,20-E2), OFFSET(C3,0,0,20-E2))/VAR(\$C\$2:\$C\$21)/19”，然后利用“填充柄”将 F2 单元格公式复制到 F3:F9 单元格，结果如图附-38 所示。

	E	F
1	Lag	ac
2	1	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E2), OFFSET(C3,0,0,20-E2))/VAR(\$C\$2:\$C\$21)/19
3	2	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E3), OFFSET(C4,0,0,20-E3))/VAR(\$C\$2:\$C\$21)/19
4	3	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E4), OFFSET(C5,0,0,20-E4))/VAR(\$C\$2:\$C\$21)/19
5	4	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E5), OFFSET(C6,0,0,20-E5))/VAR(\$C\$2:\$C\$21)/19
6	5	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E6), OFFSET(C7,0,0,20-E6))/VAR(\$C\$2:\$C\$21)/19
7	6	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E7), OFFSET(C8,0,0,20-E7))/VAR(\$C\$2:\$C\$21)/19
8	7	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E8), OFFSET(C9,0,0,20-E8))/VAR(\$C\$2:\$C\$21)/19
9	8	=SUMPRODUCT(OFFSET(C\$2,0,0,20-E9), OFFSET(C10,0,0,20-E9))/VAR(\$C\$2:\$C\$21)/19

图 附-38

2. 计算偏自相关函数。计算偏自相关函数的步骤较为复杂，必须利用 Excel 的逆矩阵等函数求解 Yule-Walker 方程组，由于我们选择了置后期数为 8，为了求解偏自相关函数，我们必须求解 8 个 Yule-Walker 方程组。首先，利用自相关函数的计算结果，填写 H2:O9 范围内的对称矩阵如图附-39 中 H2:O9 单元格所示。其次，利用 Excel 数组公

式分别求解 8 个方程组的结果，结果分别放在 1i 至 8i 的八列之中，第一个方程组的结果放在 H12 中，第二个方程组的结果放在 I12:I13 中，第三个方程组的结果放在 J12:J14 中，以此类推。所输入的 8 个数组公式分别为：

“ MMULT(MINVERSE(OFFSET(H2,0,0,1,1)),OFFSET(F2,0,0,1)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,2,2)),OFFSET(F2,0,0,2)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,3,3)),OFFSET(F2,0,0,3)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,4,4)),OFFSET(F2,0,0,4)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,5,5)),OFFSET(F2,0,0,5)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,6,6)),OFFSET(F2,0,0,6)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,7,7)),OFFSET(F2,0,0,7)) ”,
“ MMULT(MINVERSE(OFFSET(H2,0,0,8,8)),OFFSET(F2,0,0,8)) ”。

(说明 1. 在 Excel 中输入数组公式时,先用鼠标选定所有需放置结果的单元格地址范围然后输入数组公式，例如

“ =MMULT(MINVERSE(OFFSET(H2,0,0,2,2)), OFFSET(F2,0,0,2)) ”，然后同时按下“ CTRL+SHIFT+回车 ”三个按键，完成数组公式的输入，公式会自动加上一对大括号，它由 Excel 自动添入。2. 以上数组公式中包含的各个函数的含义及其用法请参看附表 1。)最后，将每一个方程组的最后一个解，用值复制的方式复制到 pac 这一列，即可得到 8 个偏自相关系数。如图附-39，表中 H12:O19 单元格的 8 列分别给出了 8 个数组公式计算的结果，F12:F19 单元格的内容即是所求解的 8 个偏自相关系数。

	E	F	G	H	I	J	K	L	M	N	O
1	Lag	ac									
2	1	0.819067942		1	0.819	0.612	0.452	0.327	0.164	0.0388	-0.01
3	2	0.611510228		0.819	1	0.819	0.612	0.452	0.327	0.1642	0.039
4	3	0.451726387		0.612	0.819	1	0.819	0.612	0.452	0.3272	0.164
5	4	0.327219978		0.452	0.612	0.819	1	0.819	0.612	0.4517	0.327
6	5	0.164215603		0.327	0.452	0.612	0.819	1	0.819	0.6115	0.452
7	6	0.038753484		0.164	0.327	0.452	0.612	0.819	1	0.8191	0.612
8	7	-0.009794647		0.039	0.164	0.327	0.452	0.612	0.819	1	0.819
9	8	-0.116257009		-0.01	0.039	0.164	0.327	0.452	0.612	0.8191	1
10											
11		pac		ϕ_{1i}	ϕ_{2i}	ϕ_{3i}	ϕ_{4i}	ϕ_{5i}	ϕ_{6i}	ϕ_{7i}	ϕ_{8i}
12		0.819067942		0.819	0.967	0.971	0.972	0.967	0.972	0.9704	0.997
13		-0.180361799			-0.18	-0.205	-0.21	-0.2	-0.204	-0.184	-0.202
14		0.025905952				0.026	0.048	0.002	0.002	-0.015	-0.085
15		-0.022714997					-0.02	0.19	0.195	0.1951	0.256
16		-0.21914027						-0.22	-0.243	-0.226	-0.23
17		0.024732305							0.025	-0.058	-0.115
18		0.084685469								0.0847	0.386
19		-0.310257383									-0.31

图 附-39

3. 模型的识别与估计。自相关函数序列呈现明显拖尾性，而偏自相关

函数序列在 $k > 1$ 之后，都在区间 $(-\frac{1.96}{\sqrt{20}}, \frac{1.96}{\sqrt{20}})$ ，即 $(-0.438, 0.438)$

之间，因此可以认为自相关函数在 $K > 1$ 之后截尾，因此我们选用 AR(1) 模型进行数据拟合。复制 C2: C20 的数据，将之以值复制的形式复制到 D3: D21 的单元格，并在 D1 中填入标志项“Z(-1)”。选择“工具”菜单的“数据分析”子菜单，双击“回归”选项，弹出回归分析对话框。按图附-40 所示的方式填写对话框。然后单击“确定”按钮，即可得到 AR(1) 模型的估计结果。

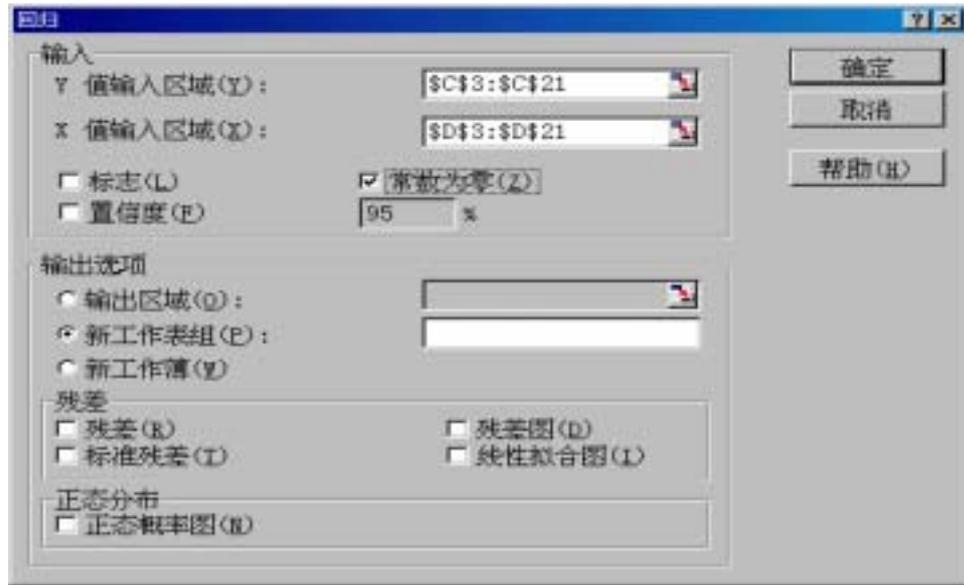


图 附-40

(三) 结果分析：按以上操作步骤，可得到图附-41 所示 AR(1)模型

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	回归统计								
4	Multiple R	0.93329							
5	R Square	0.87104							
6	Adjusted R Square	0.81548							
7	标准误差	13.0237							
8	观测值	19							
9									
10	方差分析								
11		df	SS	MS	F	Significance F			
12	回归分析	1	20621.4	20621.4	121.577	3.63113E-09			
13	残差	18	3053.09	169.616					
14	总计	19	23674.5						
15									
16		Coefficient	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
17	Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
18	X Variable 1	1.06284	0.09639	11.0264	1.9E-09	0.860332697	1.26535	0.86033	1.26535

图 附-41

结果。因此，零均值化模型的估计结果是 $\hat{Z} = 1.06284 * Z(-1)$ ，还原成上证指数，最终的时间序列模型是：上证指数估计值-上证指数的平均值 = 1.06284(上一天上证指数-上证指数平均值)。

十四、季节变动时间序列的分解分析

(一) 简介：分解分析法是分析时间序列常用的统计方法。季节时间序列是趋势变动(T)、季节变动(S)、随机变动(I)综合影响的结果，分解过程

要从原始序列中消除随机变动,然后分别识别出季节变动和趋势变动的变化模式。下面结合具体例子介绍在 Excel 中如何实现时间序列的分解分析。如图附-42 所示,表中 A1 至 B13 单元格是 1996 至 1998 年各季度某海滨城市旅游人口数(千人),试预测 1999 年各季度旅游人口数。

(二) 操作步骤：

1. 计算一次移动平均,消除随机波动。在 C3 单元格填入公式“=AVERAGE(B2:B5)”,然后用“填充柄”将公式复制到 C4:C11 单元格。
2. 中心化移动平均数。在 D4 单元格输入公式“=AVERAGE(C3:C4)”,然后用“填充柄”将公式复制到 D5:D11 单元格。
3. 计算各个季节指数。在 E4 单元格输入公式“=B4/E4”,然后用“填充柄”将公式复制到 E5:E11 单元格。
4. 计算平均季节指数。在 F4 单元格中输入公式“=AVERAGE(E4,E8)”,然后用“填充柄”将公式复制到 F5:F7 单元格。

	A	B	C	D	E	F	G	H	I
1	季度	旅游人数	一次移动平均	中心化移动平均	季节指数	平均季节指数	调整季节指数	消除季节变动	时间
2	1996.1	77					0.670588435	114.8245272	1
3	1996.2	115	148.75				0.872016265	131.8782741	2
4	1996.3	298	157.75	153.25	1.94454	1.860272838	1.820272369	163.7117637	3
5	1996.4	105	168.75	163.25	0.64319	0.651123701	0.637122931	164.8033606	4
6	1997.1	113	178	173.375	0.65177	0.68532461	0.670588435	168.5087217	5
7	1997.2	159	184.25	181.125	0.87785	0.891178814	0.872016265	182.3360496	6
8	1997.3	335	193	188.625	1.77601		1.820272369	184.0383921	7
9	1997.4	130	201.5	197.25	0.65906		0.637122931	204.0422558	8
10	1998.1	148	210.25	205.875	0.71888		0.670588435	220.7016887	9
11	1998.2	193	216.5	213.375	0.90451		0.872016265	221.326147	10
12	1998.3	370					1.820272369	203.2662838	11
13	1998.4	155					0.637122931	243.2811512	12
14	1999.1								13
15	1999.2								14
16	1999.3								15
17	1999.4								16

图 附-42

5. 计算调整后的季节指数。为了让季节指数的总平均为 1,必须对季节指数加以调整。在 G4 单元格中输入公式“=F4/AVERAGE(\$F\$4:\$F\$7)”,然后用“填充柄”将公式复制到 G5:G7 单元格。G4:G7 就是最终计算出的季节指数,按 G4:G7 给出的 4 个季度的季节指数,将季节指数填充到 G2:G13 的其它单元格。
6. 消除旅游人数序列中的季节变动。在 H2 单元格中输入“=B2/F2”,然后用“填充柄”将公式复制到 H3:H13 单元格。则 H 列就是消除季节变动之后的旅游人数时间序列。

7. 对消除季节变动的旅游人数进行回归分析。在 I 列填入时间序号 1 至 12，如图附-42 所示。选择“工具”菜单的“数据分析”子菜单，双击“回归”选项，弹出回归分析对话框。按图附-43 所示的方式填写对话框。然后单击“确定”按钮，即可得到剔除了季节波动的时间序列的线性趋势模型。估计结果如图附-44 所示，其中 B35 单元格是线性趋势模型的截距，B36 单元格是斜率。

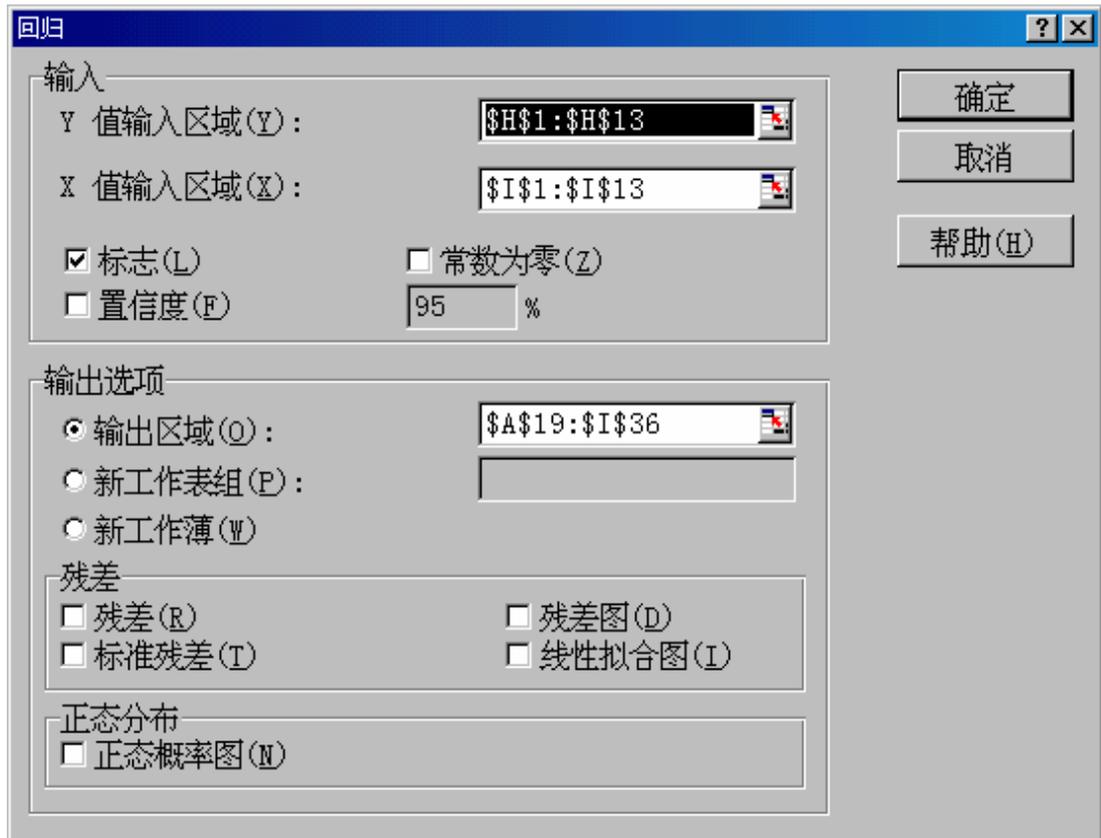


图 附-43

	A	B	C	D	E	F	G	H	I
19	SUMMARY OUTPUT								
20									
21	回归统计								
22	Multiple R	0.95448							
23	R Square	0.91103							
24	Adjusted R Square	0.90213							
25	标准误差	11.762							
26	观测值	12							
27									
28	方差分析								
29		df	SS	MS	F	Significance F			
30	回归分析	1	14166.4156	14166.4156	102.4	1.42572E-06			
31	残差	10	1383.44549	138.344549					
32	总计	11	15549.8611						
33									
34		Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
35	Intercept	118.864	7.2390081	16.4199564	1.5E-08	102.7346797	134.99	102.7347	134.994
36	时间	9.95318	0.98358751	10.1192649	1.4E-06	7.761612654	12.145	7.761613	12.1448

图 附-44

8. 预测。在 G14:G17 单元格中分别填入刚才计算出的四个调整后的季节指数，在 B14 单元格中输入公式 “=(\$B\$35+I14*\$B\$36)*G14”，然后利用“填充柄”将公式复制到 B15:B17 单元格，B14:B17 单元格中就是 1999 年各个季度旅游人数的预测值，如图附-45 所示。

	A	B	C	D	E	F	G	H	I
1	季度	旅游人数	一次移动平均	中心化移动平均	季节指数	平均季节指数	整季季节指数	消除季节变动	时间
14	1999.1	166.4773					0.67059		13
15	1999.2	225.1622					0.87202		14
16	1999.3	488.1278					1.82027		15
17	1999.4	177.1935					0.63712		16

图 附-45

(三) 结果分析：以上步骤完成了整个季节时间序列的分析和预测过程。使用了分解分析的方法，能将时间数列的各个影响因素都分解出来，由这种方法得到的预测模型和预测结果都比直接使用回归分析要更为可靠合理。参看以上分析步骤，用类似的方法还可以进行月份时间序列、双循环变动时间序列等的分解分析和预测。

附表 1 Excel 统计函数一览表

函数名称	函数功能
AVEDEV	返回一组数据与其均值的绝对偏差的平均值，用于评测这组数据的离散度。
AVERAGE	返回指定序列算术平均值。
AVERAGEA	计算参数清单中数值的算数平均值。不仅数字，而且文本和逻辑值（如 TRUE 和 FALSE）也将计算在内。
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。
BETAINV	返回 beta 分布累积函数的逆函数值。即，如果 $probability = BETADIST(x, \dots)$ ，则 $BETAINV(probability, \dots) = x$ 。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。
BINOMDIST	返回一元二项式分布的概率值。函数 BINOMDIST 适用于固定次数的独立实验，实验的结果只包含成功或失败二种情况，且成功的概率在实验期间固定不变。例如，函数 BINOMDIST 可以计算三个婴儿中两个是男孩的概率
CHIDIST	返回 X^2 分布的单尾概率。 X^2 分布与 X^2 检验相关。使用 X^2 检验可以比较观察值和期望值。例如，某项遗传学实验假设下一代植物将呈现出某一组颜色。使用此函数比较观测结果和期望值，可以确定初始假设是否有效。
CHIINV	返回 X^2 分布单尾概率的逆函数。如果 $probability = CHIDIST(x, ?)$ ，则 $CHIINV(probability, ?) = x$ 。使用此函数比较观测结果和期望值，可以确定初始假设是否有效。
CHITEST	返回独立性检验值。函数 CHITEST 返回 X^2 分布的统计值及相应的自由度。可以使用 X^2 检验确定假设值是否被实验所证实。

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。例如，如果通过邮购的方式订购产品，依照给定的置信度，可以确定最早及最晚到货的时间。
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。例如，可以检测某地的平均温度和空调使用情况之间的关系。
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。
COUNTA	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。
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。例如，可利用它来检验教育程度与收入档次之间的关系。
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。例如，使用函数 CRITBINOM 来决定最多允许出现多少个有缺陷的部件，才可以保证当整个产品在离开装配线时检验合格。
DEVSQ	返回数据点与各自样本均值偏差的平方和。
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，例如，在计算银行自动提款机支付一次现金所花费的时间时，可通过函数 EXPONDIST，确定这一过程最长持续一分钟的发生概率。
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。例如，分析进入高校的男生、女生的考试分数，确定女生分数的变化程度是否与男生不同。
FINV	返回 F 概率分布的逆函数值。如果 $p = \text{FDIST}(x, \dots)$ ，则 $\text{FINV}(p, \dots) = x$ 。在 F 检验中，可以使用 F 分布比较两个数据集的变化程度。例如，可以分析美国、加拿大的收入分布，判断两个国家是否有相似的收入变化程度。

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。使用此函数可以完成相关系数的假设检验
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。如果 $y = \text{FISHER}(x)$ ，则 $\text{FISHERINV}(y) = x$ 。
FORECAST	根据给定的数据计算或预测未来值。此预测值为基于一系已知的 x 值推导出的 y 值。以数组或数据区域的形式给定 x 值和 y 值后，返回基于 x 的线性回归预测值。使用此函数可以对未来销售额、库存需求或消费趋势进行预测。
FREQUENCY	以一系列垂直数组返回某个区域中数据的频率分布。例如，使用函数 FREQUENCY 可以计算在给定的值集和接收区间内，每个区间内的数据数目。由于函数 FREQUENCY 返回一个数组，必须以数组公式的形式输入。
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。例如，给定公立和私立学校的测试成绩，可以检验各学校间的差别程度。
GAMMADIST	返回 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。
GAMMAINV	返回 分布的累积函数的逆函数。如果 $P = \text{GAMMADIST}(x, \dots)$ ，则 $\text{GAMMAINV}(p, \dots) = x$ 。使用此函数可以研究出现分布偏斜的变量。
GAMMALN	返回伽玛函数的自然对数， $\ln(\Gamma(x))$ 。
GEOMEAN	返回正数数组或数据区域的几何平均值。例如，可以使用函数 GEOMEAN 计算可变复利的平均增长率。
GROWTH	根据给定的数据预测指数增长值。根据已知的 x 值和 y 值，函数 GROWTH 返回一组新的 x 值对应的 y 值。可以使用 GROWTH 工作表函数来拟合满足给定 x 值和 y 值的指数曲线。
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。

HYPGEOMDIST	返回超几何分布。给定样本容量、样本总体容量和样本总体中成功的次数，函数 HYPGEOMDIST 返回样本取得给定成功次数的概率。使用函数 HYPGEOMDIST 可以解决有限总体的问题，其中每个观察值或者为成功或者为失败，且给定样本区间的所有子集有相等的发生概率。
INTERCEPT	利用已知的 x 值与 y 值计算直线与 y 轴的截距。截距为穿过 known_x 担和 known_y 担数据点的线性回归线与 y 轴交点。当已知自变量为零时，利用截距可以决定因变量的值。例如，当所有的数据点都是在室温或更高的温度下取得的，可以用函数 INTERCEPT 预测在 0 时金属的电阻。
KURT	返回数据集的峰值。峰值反映与正态分布相比某一分布的尖锐度或平坦度。正峰值表示相对尖锐的分布。负峰值表示相对平坦的分布。
LARGE	返回数据集里第 k 个最大值。使用此函数可以根据相对标准来选择数值。例如，可以使用函数 LARGE 得到第一名，第二名，或第三名的得分。
LINEST	乘法计算对已知数据进行最佳直线拟合，并返回描述此直线的数组。因为此函数返回数值数组，故必须以数组公式的形式输入。
LOGEST	在回归分析中，计算最符合观测数据组的指数回归拟合曲线，并返回描述该曲线的数组。由于这个函数返回一个数组，必须以数组公式输入。
LOGINV	返回 x 的对数正态分布累积函数的逆函数，此处的 $\ln(x)$ 是含有 mean (平均数) 与 standard-dev (标准差) 参数的正态分布。如果 $p = \text{LOGNORMDIST}(x, \dots)$ 那么 $\text{LOGINV}(p, \dots) = x$ 。使用对数正态分布可以分析经过对数变换的数据。
LOGNORMDIST	返回 x 的对数正态分布的累积函数，其中 $\ln(x)$ 是服从参数为 mean 和 standard_dev 的正态分布。使用此函数可以分析经过对数变换的数据。
MAX	返回数据集中的最大数值。
MAXA	返回参数清单中的最大数值。文本值和逻辑值 (如 TRUE 和 FALSE) 也作为数字来计算。函数 MAXA 与

	函数 MINA 相似。
MEDIAN	返回给定数值集合的中位数。中位数是在一组数据中居于中间的数，换句话说，在这组数据中，有一半的数据比它大，有一半的数据比它小。
MIN	返回给定参数表中的最小值。
MINA	返回参数清单中的最小数值。文本值和逻辑值（如 TRUE 和 FALSE）也作为数字来计算。
MODE	返回在某一数组或数据区域中出现频率最多的数值。跟 MEDIAN 一样，MODE 也是一个位置测量函数。
NEGBINOMDIST	返回负二项式分布。当成功概率为常数 probability_s 时，函数 NEGBINOMDIST 返回在到达 number_s 次成功之前，出现 number_f 次失败的概率。此函数与二项式分布相似，只是它的成功次数固定，试验总数为变量。与二项分布类似的是，试验次数被假设为自变量。
NORMDIST	返回给定平均值和标准偏差的正态分布的累积函数。此函数在统计方面应用范围广泛（包括假设检验）
NORMINV	返回给定平均值和标准偏差的正态分布的累积函数的逆函数。
NORMSDIST	返回标准正态分布的累积函数，该分布的平均值为 0，标准偏差为 1。可以使用该函数代替标准正态曲线面积表。
NORMSINV	返回标准正态分布累积函数的逆函数。该分布的平均值为 0，标准偏差为 1。
PEARSON	返回 Pearson（皮尔生）乘积矩相关系数，r，这是一个范围在 -1.0 到 1.0 之间（包括 -1.0 和 1.0 在内）的无量纲指数，反映了两个数据集合之间的线性相关程度。
PERCENTILE	返回数值区域的 K 百分比数值点。可以使用此函数来建立接受阈值。例如，可以确定得分排名在 90 个百分点以上的检测候选人。
PERCENTRANK	返回特定数值在一个数据集中的百分比排位。此函数可用于查看特定数据在数据集中所处的位置。例如，可以使用函数 PERCENTRANK 计算某个特定的能力测

	试得分在所有的能力测试得分中的位置。
PERMUT	返回从给定数目的对象集合中选取的若干对象的排列数。排列可以为有内部顺序的对象或为事件的任意集合或子集。排列与组合不同，组合的内部顺序无意义。此函数可用于彩票计算中的概率。
POISSON	返回泊松分布。泊松分布通常用于预测一段时间内事件发生的次数，比如一分钟内通过收费站的轿车的数量。
PROB	返回一概率事件组中落在指定区域内的事件所对应的概率之和。如果没有给出 upper_limit，则返回 x_range 内值等于 lower_limit 的概率。
QUARTILE	返回数据集的四分位数。四分位数通常用于在销售额和测量值数据集中对总体进行分组。例如，可以使用函数 QUARTILE 求得总体中前 25% 的收入值。
RANK	返回一个数值在一组数值中的排位。数值的排位是与数据清单中其它数值的相对大小（如果数据清单已经排过序了，则数值的排位就是它当前的位置）。
RSQ	返回根据 known_y's 和 known_x's 中数据点计算得出的 Pearson 乘积矩相关系数的平方。详细内容，则参阅函数 REARSON。R 平方值可以解释为 y 方差与 x 方差的比例。
SKEW	返回分布的偏斜度。偏斜度反映以平均值为中心的分布的不对称程度。正偏斜度表示不对称边的分布更趋向正值。负偏斜度表示不对称边的分布更趋向负值。
SLOPE	返回根据 known_y's 和 known_x's 中的数据点拟合的线性回归直线的斜率。斜率为直线上任意两点的重直距离与水平距离的比值，也就是回归直线的变化率。
SMALL	返回数据集中第 k 个最小值。使用此函数可以返回数据集中特定位置上的数值。
STANDARDIZE	返回以 mean 为平均值，以 standard-dev 为标准偏差的分布的正态化数值。
STDEV	估算样本的标准偏差。标准偏差反映相对于平均值（mean）的离散程度。
STDEVA	估算基于给定样本的标准偏差。标准偏差反映数值相

	对于平均值 (mean) 的离散程度。文本值和逻辑值 (如 TRUE 或 FALSE) 也将计算在内。
STDEVP	返回以参数形式给出的整个样本总体的标准偏差。标准偏差反映相对于平均值 (mean) 的离散程度。
STDEVPA	计算样本总体的标准偏差。标准偏差反映数值相对于平均值 (mean) 的离散程度。
STEYX	返回通过线性回归法计算 y 预测值时所产生的标准误差。标准误差用来度量根据单个 x 变量计算出的 y 预测值的误差量。
TDIST	返回学生氏- t 分布。T 分布用于小样本数据集合的假设检验。使用此函数可以代替 t 分布的临界值表。
TINV	返回指定自由度的学生氏- t 分布的逆函数。
TREND	返回一条线性回归拟合线的一组纵坐标值 (y 值)。即找到适合给定的数组 known_y's 和 known_x's 的直线 (用最小二乘法)，并返回指定数组 new_x's 值在直线上对应的 y 值。
TRIMMEAN	返回数据集的内部平均值。函数 TRIMMEAN 先从数据集的头部和尾部除去一定百分比的数据点，然后再求平均值。当希望在分析中剔除一部分数据的计算时，可以使用此函数。
TTEST	返回与学生氏- t 检验相关的概率。可以使用函数 TTEST 判断两个样本是否可能来自两个具有相同均值的总体。
VAR	估算样本方差。
VARA	估算基于给定样本的方差。不仅数字，文本值和逻辑值 (如 TRUE 和 FALSE) 也将计算在内。
VARP	计算样本总体的方差。
VARPA	计算样本总体的方差。不仅数字，文本值和逻辑值 (如 TRUE 和 FALSE) 也将计算在内。
WEIBULL	返回韦伯分布。使用此函数可以进行可靠性分析，比如计算设备的平均故障时间。
ZTEST	返回 z 检验的双尾 P 值。Z 检验根据数据集或数组生成 x 的标准得分，并返回正态分布的双尾概率。

	可以使用此函数返回从某总体中抽取特定观测值的似然估计。
--	-----------------------------

附表 2 Excel 数据分析工具一览表

“ F - 检验：双样本方差分析 ” 分析工具	此分析工具可以进行双样本 F - 检验，用来比较两个样本总体的方差。例如，可以对参加游泳比赛的两个队的时间记分进行 F- 检验，查看二者的样本方差是否不同。
“ t - 检验：成对双样本均值分析 ” 分析工具	此分析工具及其公式可以进行成对双样本学生氏 t - 检验，用来确定样本均值是否不等。此 t - 检验并不假设两个总体的方差是相等的。当样本中出现自然配对的观察值时，可以使用此成对检验，例如对一个样本组进行了两次检验，抽取实验前的一次和实验后的一次。
“ t - 检验：双样本等方差假设 ” 分析工具	此分析工具可以进行双样本学生氏 t - 检验。此 t- 检验先假设两个数据集的平均值相等，故也称作齐次方差 t - 检验。可以使用 t - 检验来确定两个样本均值实际上是否相等。
“ t - 检验：双样本异方差假设 ” 分析工具	此分析工具及其公式可以进行双样本学生氏 t - 检验。此 t - 检验先假设两个数据集的方差不等，故也称作异方差 t - 检验。可以使用 t - 检验来确定两个样本均值实际上是否相等。当进行分析的样本组不同时，可使用此检验。如果某一样本组在某次处理前后都进行了检验，则应使用“成对检验”。
“ z - 检验：双样本均值分析 ” 分析工具	此分析工具可以进行方差已知的双样本均值 z - 检验。此工具用于检验两个总体均值之间存在差异的假设。例如，可以使用此检验来确定两种汽车模型性能之间的差异情况。
“ 抽样分析 ” 分析工具	此分析工具以输入区域为总体构造总体的一个样本。当总体太大而不能进行处理或绘制时，可以选用具有代表性的样本。如果确认输入区域中的数据

	<p>是周期性的，还可以对一个周期中特定时间段中的数值进行采样。例如，如果输入区域包含季度销售量数据，以四为周期进行取样，将在输出区域中生成某个季度的样本。</p>
“傅立叶分析”分析工具	<p>此分析工具可以解决线性系统问题，并能通过快速傅立叶变换（FFT）分析周期性的数据。此工具也支持逆变换，即通过对变换后的数据的逆变换返回初始数据。</p>
“回归分析”分析工具	<p>此工具通过对一组观察值使用“最小二乘法”直线拟合，进行线性回归分析。本工具可用来分析单个因变量是如何受一个或几个自变量影响的。例如，观察某个运动员的运动成绩与一系列统计因素的关系，如年龄、身高和体重等。在操作时，可以基于一组已知的体能统计数据，并辅以适当加权，对尚未进行过测试的运动员的表现作出预测。</p>
“描述统计”分析工具	<p>此分析工具用于生成对输入区域中数据的单变值分析，提供有关数据趋中性和易变性的信息。</p>
“排位和百分比排位”分析工具	<p>此分析工具可以产生一个数据列表，在其中罗列给定数据集中各个数值的大小次序排位和相应的百分比排位。用来分析数据集中各数值间的相互位置关系。</p>
“随机数发生器”分析工具	<p>此分析工具可以按照用户选定的分布类型，在工作表的特定区域中生成一系列独立随机数字。可以通过概率分布来表示主体的总体特征。例如，可以使用正态分布来表示人体身高的总体特征，或者使用双值输出的伯努利分布来表示掷币实验结果的总体特征。</p>
“相关系数”分析工具	<p>此分析工具及其公式可用于判断两组数据集（可以使用不同的度量单位）之间的关系。可以使用“相关系数”分析工具来确定两个区域中数据的变化是否相关，即，一个集合的较大数据是否与另一个集合的较大数据相对应（正相关）；或者一个集合的较小数据是否与另一个集合的较小数据相对应（负相关）；还是两个集合中的数据互不相关（相关性为零）。</p>

<p>“ 协方差 ” 分析工具</p>	<p>此分析工具及其公式用于返回各数据点的一对均值偏差之间的乘积的平均值。协方差是测量两组数据相关性的量度。可以使用协方差工具来确定两个区域中数据的变化是否相关，即，一个集合的较大数据是否与另一个集合的较大数据相对应(正协方差)；或者一个集合的较小数据是否与另一个集合的较小数据相对应(负协方差)；还是两个集合中的数据互不相关(协方差为零)。</p>
<p>“ 移动平均 ” 分析工具</p>	<p>此分析工具及其公式可以基于特定的过去某段时期中变量的均值，对未来值进行预测。移动平均值提供了由所有历史数据的简单的平均值所代表的趋势信息。使用此工具可以预测销售量、库存或其它趋势。</p>
<p>“ 直方图 ” 分析工具</p>	<p>在给定工作表中数据单元格区域和接收区间的情况下，计算数据的个别和累积频率，用于统计有限集中某个数值元素的出现次数。例如，在一个有 20 名学生的班级里，可以确定以字母打分(如 A、B-等)所得分数的分布情况。直方图表会给出字母得分的边界，以及在最低边界与当前边界之间某一得分出现的次数。出现频率最多的某个得分即为数据组中的众数。</p>
<p>“ 指数平滑 ” 分析工具</p>	<p>此分析工具及其公式基于前期预测值导出相应的新预测值，并修正前期预测值的误差。此工具将使用平滑常数 a，其大小决定了本次预测对前期预测误差的修正程度。</p>
<p>“ Anova :单因素方差分析 ” 分析工具</p>	<p>此分析工具通过简单的方差分析(anova)，对两个以上样本均值进行相等性假设检验(抽样取自具有相同均值的样本空间)。此方法是对双均值检验(如 t-检验)的扩充。</p>
<p>“ Anova :可重复双因素分析 ” 分析工具</p>	<p>此分析工具是对单因素 anova 分析的扩展，即每一组数据包含不止一个样本。</p>
<p>“ Anova :无重复双因素分析 ” 分析工具</p>	<p>此分析工具通过双因素 anova 分析(但每组数据只包含一个样本)，对两个以上样本均值进行相等性假设检验(抽样取自具有相同均值的样本空间)。此方法是对双均值检验(如 t-检验)的扩充。</p>

参考书目：

- [1] [美]MICHAEL R. MIDDLETON, 《使用 Excel 进行数据分析》，中国水利水电出版社，1997 年。
- [2] 王美今, 《经济预测》，厦门大学出版社, 1997 年。
- [3] 何晓群, 《现代统计分析方法与应用》，中国人民大学出版社，1998 年。