

附录四 Excel 在统计分析与数量方法中的应用

§1 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。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。

(1) 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。

(2) 操作简便

(3) 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。

(4) 数据库管理能力

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。

(5) 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。

(6) 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。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。

(7) 对象连接和嵌入功能

利用对象连接和嵌入功能，用户可将其它软件（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。

(8) 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。

§ 2 Excel 基本操作

2.1 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:

(1) 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。

(2) 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。

(3) 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。

2.2 数据的输入输出操作

(1) 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。

(2) 公式生成数据

Excel的数据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。

(3) 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:

i) 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。

ii) 选定区域之后，用鼠标右击该区域，选择“复制”，将区域内容复制到粘贴板之中。可以发现该区域已被虚线包围。

iii) 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。

i) 值复制

① 拖动鼠标选定待复制区域。

② 用鼠标右击选定区域，选择“复制”选项。

③ 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。

ii) 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 单元格中输入的是公式“=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。

2.3 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：

(1) 在“文件”菜单中选择“打开”子菜单。

(2) 在“打开文件”对话框中选择所要打开的文件类型及其所在的目录。

(3) 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：

(1) 编辑好文件后，在“文件”菜单中选择“另存为”子菜单。

(2) 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。

(3) 输入文件名之后,用鼠标单击“保存”按钮即可。

§3 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。

3.1 描述统计工具

(1) 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。

(2) 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。

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。

3.2 直方图工具

（1）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。

（2）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值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，并同时在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。

（3）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。

3.3 利用Excel绘制散点图

（1）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。

（2）操作步骤

数据如图1所示，可按如下步骤建立变量x-y，x-z的散点图。

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

图1 数据表

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2的图表。

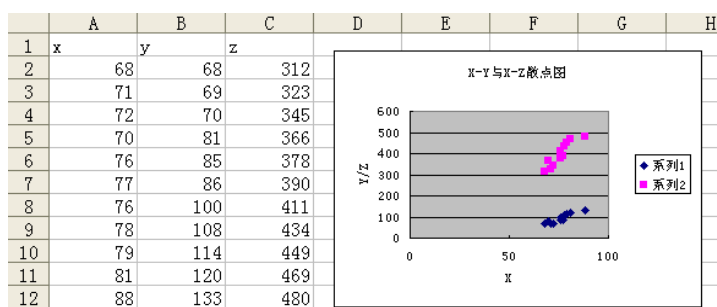


图2 散点图示意图

(3) 结果说明

如图2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。

3.4 数据透视表工具

(1) 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。

(2) 操作步骤

如图3所示，表中列出学生两门功课评定结果，可按如下步骤建立交叉频数表。

1. 选中图3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”。

	A	B	C
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3 数据表

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4将三个项目“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，然后单击“下一步”按钮。

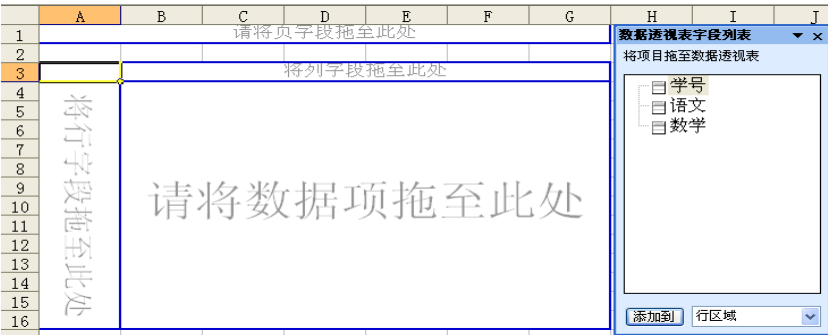


图4 透视表示意图

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5所示的数据透视表。

	A	B	C	D	E	F
1						
2						
3	计数项:学号	语文				
4	数学	差	良	优	中	总计
5	差	1		1		2
6	良	1			1	2
7	优				1	1
8	中	1	2		1	4
9	总计	3	2	1	3	9

图5 数据透视表

(3) 结果说明

如图5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。

3.5 排位与百分比工具

(1) 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。

(2) 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。

(3) 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)。

§4 Excel在推断统计中的应用

4.1 二项分布工具

(1) 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。

(2) 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6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6 二项分布框架表

2. 先在B7至F7单元格分别输入概率计算公式:

“=BINOMDIST(A7, 6, 0.3, 0)”, “=BINOMDIST(A7, 6, 0.3, 1)”

“=C7-B7”, “=1-C7”, “=1-D7”

这里BINOMDIST的语法为

BINOMDIST(number_s, trials, probability_s, cumulative)

其中 number_s 为试验成功的次数; trials 为独立试验的次数; probability_s 为每次试验中成功的概率; cumulative 为一逻辑值,用于确定函数的形式。如果 cumulative 为 TRUE, 函数 BINOMDIST 返回累积分布函数,即至多 number_s 次成功的概率; 如果为 FALSE, 返回概率密度函数,即 number_s 次成功的概率。

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7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	0.882351	1
8	1	0.302526	0.420175	0.117649	0.579825	0.882351
9	2	0.324135	0.74431	0.420175	0.25569	0.579825
10	3	0.18522	0.92953	0.74431	0.07047	0.25569
11	4	0.059535	0.989065	0.92953	0.010935	0.07047
12	5	0.010206	0.999271	0.989065	0.000729	0.010935
13	6	0.000729	1	0.999271	0	0.000729

图7 二项分布数据

4. 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。

5. 选择“柱状图”,然后单击“下一步”。

6. 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”。

7. 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。

4.2 其它分布的函数

(1) 函数CRITBINOM

1. 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。

2. 语法: CRITBINOM(trials, probability_s, alpha)

trials : 贝努利实验次数。

probability_s : 每次试验中成功的概率。

alpha : 临界概率。

3. 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。

(2) 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:

$$\text{HYPGEOMDIST}(1, 4, 8, 20) = 0.363261$$

(3) 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定时,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

(4) 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。

(5) 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分布函数。同样可以利用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的正态分布当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的正态分布当x=60时分布函数的数值。

(6) 函数NORMSDIST:

1. 说明: 函数NORMSDIST返回标准正态分布的分布函数。

2. 语法: NORMSDIST(z), 其中z为需要计算其分布的数值。

3. 举例: NORMSDIST(0)=0.5

(7) 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), 其中probability为正态分布的概率值。

3. 举例: NORMSINV(0.5)=0

(8) t分布函数TDIST

1. 说明: 函数TDIST返回t分布的概率。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, TDIST的计算公式为 $TDIST=P(X>x)$ 。如果tails=2, 函数TDIST返回双尾分布, TDIST的计算公式为 $TDIST=P(|X|>x)=P(X>x \text{ or } X<-x)$ 。

3. 举例: $TDIST(1.96, 60, 2)=0.054645$

4.3 随机抽样的工具

(1) 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。

(2) 操作步骤

如图8(a)所示有7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。

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B8单元格, 并在B1单元格输入标题“RANDOM”。
3. 选取单元格B2至B8, 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B8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8(b)所示, A2至A6单元格的样本即为随机抽取的5个样本。

	A
1	No
2	1
3	2
4	3
5	4
6	5
7	6
8	7

图8(a)

	A	B
1	No	random
2	4	0.313541
3	7	0.358698
4	6	0.636284
5	2	0.686976
6	5	0.785156
7	1	0.790632
8	3	0.951

图8(b)

(3) 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的进行随机数字模拟试验与随机游走模拟试验。

4.4 由样本推断总体

(1) 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。

(2) 操作步骤

1. 构造工作表。如图9(a)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。要显示公式，用鼠标选择“工具”菜单的“公式审核”子菜单的“公式审核模式”选项即可。

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	
9	计算结果	
10	抽样标准差	=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*抽样标准差
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9 (a)

	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	抽样标准差		0.889943818	30.9
11	自由度		9	32.5
12	t值		2.262157158	
13	置信区间半径		2.013192779	
14	置信区间上界		29.38680722	
15	置信区间下界		33.41319278	

图9 (b)

3. 输入样本数据，和用户指定的置信水平0.95, 如图9 (b) 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9 (b) 所示的计算结果。

(3) 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。

	A	B
1	总体均值的假设	
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	抽样标准差	=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)/抽样标准差
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H ₀ ", "接受H ₀ ")
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H ₀ ", "接受H ₀ ")
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10 假设检验公式设置

4.5 假设检验

(1) 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。

(2) 操作步骤

1. 构造工作表。如图10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A3:B4, A6:B8, A10:B11, A13:B15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。

3. 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11所示。

4. 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11中所示的计算结果。

(3) 结果说明：如图11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0.02030363。

	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	抽样标准差	1.75822638	30.9
11	计算Z值	-2.0475179	32.5
12	单侧检验		
13	单侧Z值	-1.6448536	
14	检验结果	拒绝 H_0	
15	单侧显著水平	0.02030363	
16	双侧检验		
17	双侧Z值	-1.959964	
18	检验结果	拒绝 H_0	
19	双侧显著水平	0.04060726	

图11 假设检验结果

	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12 数据表

4.6 双样本等均值假设检验

(1) 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12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？

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13（a）所示对话框。

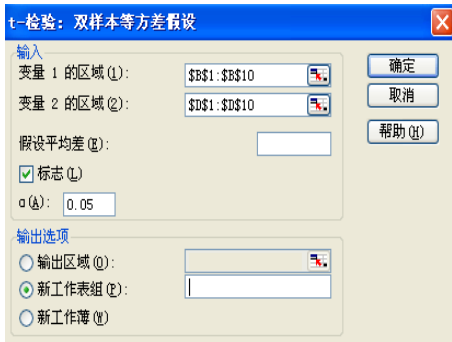


图13（a）

	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	36.33333333	31.55555556
5	方差	17.5	20.02777778
6	观测值	9	9
7	合并方差	18.76388889	
8	假设平均差	0	
9	df	16	
10	t Stat	2.339757116	
11	P(T<=t) 单尾	0.016294542	
12	t 单尾临界	1.745883669	
13	P(T<=t) 双尾	0.032589084	
14	t 双尾临界	2.119905285	

图13（b）

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，所以选择“标志”选项，再填写显著水平 α 为0.05，然后点击“确定”按钮。则可以得到图13（a）所示的结果。

（3）结果分析

如图13（b）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等于样本总个数减2，t统计量是两个样本差值减去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。可以看出t统计量均大于两个临界值，所以，在5%显著水平下，拒绝两个总体均值相等的假设，即两种装配方法所耗时间有显著的不同。

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。

4.7 正态性的 χ^2 检验

（1）简介

χ^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14所示，试用 χ^2 检验判断该样本是否来自一正态总体。

	A	B	C
1	正态性的卡方检验		
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14 分组数据

(2) 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15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15 区间概率计算公式

3. 计算预期频数值。如图16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	F
6	预期频数值	卡方值
7	=D7*样本个数	=(C7-E7)^2/E7
8	=D8*样本个数	=(C8-E8)^2/E8
9	=D9*样本个数	=(C9-E9)^2/E9
10	=D10*样本个数	=(C10-E10)^2/E10
11	=D11*样本个数	=(C11-E11)^2/E11
12	=SUM(E7:E11)	=SUM(F7:F11)
13		卡方统计量

图16 计算公式

4. 计算 χ^2 统计量。如图16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 χ^2 值, 在F12计算 χ^2 平方和, 这项就是最后计算出的 χ^2 统计量。在F13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。

	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, “拒绝总体为正态分布的假设”, “接受总体为正态分布的假设”)

图17 假设检验

5. 如图17所示, 分别在A14到B20单元格输入自由度、 χ^2 概率值、置信水平、临界值、 χ^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-正态分布参数个数-1=5-2-1=2。选定A14: 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。

(3) 结果分析: 如图18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	E	F
1	正态性的卡方检验					
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	预期频数值	卡方值
7		150	15	0.080756659	16.15133185	0.08207156
8	150	160	54	0.263821599	52.76431983	0.02893822
9	160	170	78	0.381168624	76.23372477	0.0409232
10	170	180	42	0.219453826	43.89076521	0.08145206
11	180		11	0.054799292	10.95985834	0.00014702
12		累积值	200	1	200	0.23353206
13						卡方统计量
14	自由度	2				
15	卡方概率值	0.889793352				
16						
17	置信水平	0.01				
18	临界值	9.210340372				
19	卡方检验结果	接受总体为正态分布的假设				

图18 卡方检验结果

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。

4.8 列联表分析

(1) 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是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19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
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	=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:B12)	=SUM(C10:C12)	=SUM(D10:D12)	=SUM(E10:E12)	=SUM(F10:F12)

图19 列联表

(2) 操作步骤

1. 建立期望频数表。如图19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。

2. 计算 χ^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加实际频数范围“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1E-07,如图20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20 列联表相关数据

3. 建立 χ^2 统计表。如图21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 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	F
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	=SUM(B20:E22)				
25						
26	置信水平	0.01				
27	临界值	=CHIINV(B26,6)				
28	检验结果	=IF(卡方统计量>临界值,“拒绝两种属性不相关的假设”,“接收两种属性不相关的假设”)				

图21 假设检验的公式

4. 进行假设检验。如图21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=(第一类属性的分类数-1)*(第二类属性的分类数-1)=(3-1)*(4-1)=6。

(3) 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1E-07，表明如果总体的两类属性是不相关的，那么得到以上观察的样本的概率是0.0000001。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。

4.9 单因素方差分析

(1) 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态分布的，总体方差是相等的，并且随机样本是独立的。这种工具使用于完全随机化试验的结果分析。例子如图22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22 单因素方差分析数据表

(2) 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23所示方式填写对话框。然后单击“确定”按钮。

方差分析：单因素方差分析

输入

输入区域 (I):

分组方式: ☒ 列 (C) ☐ 行 (R)

☒ 标志位于第一行 (L)

α (A):

输出选项

☐ 输出区域 (Q):

☒ 新工作表组 (P):

☐ 新工作簿 (W)

确定 取消 帮助 (H)

图23 单因素方差分析参数设置

(3) 结果分析

	A	B	C	D	E	F	G
1	方差分析：单因素方差分析						
2							
3	SUMMARY						
4	组	观测数	求和	平均	方差		
5	计划1	8	314	39.25	19.64286		
6	计划2	8	353	44.125	45.83929		
7	计划3	8	315	39.375	9.982143		
8	计划4	8	342	42.75	38.78571		
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	47.91667	1.677608	0.194419	2.946685
14	组内	799.75	28	28.5625			
15							
16	总计	943.5	31				

图24 单因素方差分析结果

按照如上的操作步骤即可得到图24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

立的话，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。

4.10 线性回归分析

(1) 简介

线性回归分析通过对一组观察值使用“最小二乘法”拟合直线，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25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25 线性回归数据

(2) 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；线性拟合图，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。

(3) 结果分析

按照如上的操作步骤即可得到图26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、判定系数 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 $Y=0.08309 \cdot X+2.77457$ 。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。

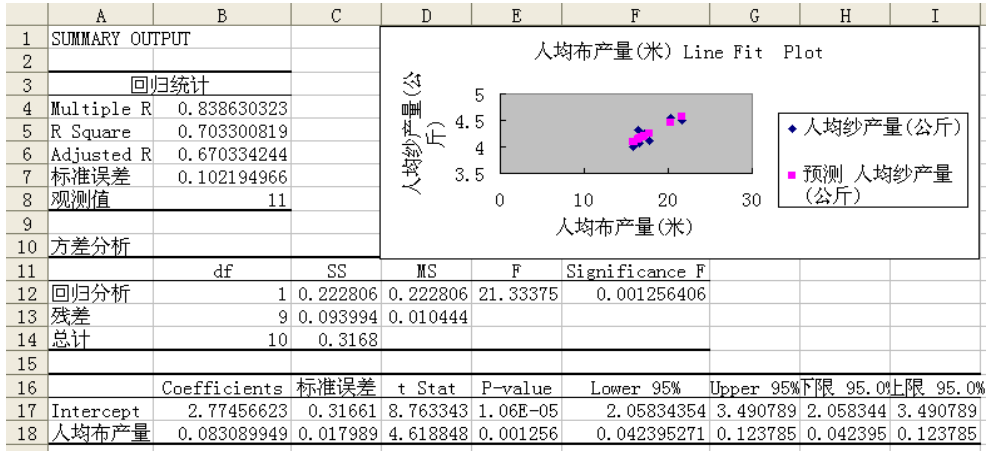


图26 回归分析计算结果

4.11 相关系数分析工具

(1) 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；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；还是两个集合中的数据互不相关（相关系数为零）。

(2) 操作步骤：采用图1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27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27 相关系数计算结果

(3) 结果说明：以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。

4.12 协方差分析工具

协方差分析的操作步骤与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。

4.13 自回归模型的识别与估计

(1) 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28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。

	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28 上证指数数据及零值化处理

(2) 操作步骤

1. 数据的零均值化处理。如图28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，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29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29 自相关函数计算公式

3. 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30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φ1i至φ8i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: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。)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30,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要求解的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	0.039	-0.01
3	2	0.611510228		0.819	1	0.819	0.612	0.452	0.327	0.164	0.039
4	3	0.451726387		0.612	0.819	1	0.819	0.612	0.452	0.327	0.164
5	4	0.327219978		0.452	0.612	0.819	1	0.819	0.612	0.452	0.327
6	5	0.164215603		0.327	0.452	0.612	0.819	1	0.819	0.612	0.452
7	6	0.038753484		0.164	0.327	0.452	0.612	0.819	1	0.819	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	0.819	1
10											
11		pac		phi	phi2	phi3	phi4	phi5	phi6	phi7	phi8
12		0.819067942		0.81906794	0.9665959	0.970905957	0.970658198	0.96642008	0.971606597	0.9695316	0.994145739
13		-0.180131808			-0.18013	-0.203440087	-0.201318242	-0.1977986	-0.20228701	-0.184046	-0.209162739
14		0.024149372				0.024149372	0.013923671	8.3184E-05	7.78075E-05	-0.004311	-0.022366733
15		0.010548888					0.010548888	0.19010611	0.194742612	0.1745534	0.149172412
16		-0.218146596						-0.2181466	-0.24110288	-0.212368	-0.168038894
17		0.023803467							0.023803467	-0.059919	-0.122154966
18		0.08413202								0.084132	0.371806711
19		-0.297066716									-0.297066716

图30 偏相关系数的计算

图31 回归参数设置

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 $(-1.96/\sqrt{20}, 1.96/\sqrt{20})$,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31

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。

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	回归统计								
4	Multiple R	0.933296876							
5	R Square	0.871043059							
6	Adjusted R Square	0.815487504							
7	标准误差	13.02367183							
8	观测值	19							
9									
10	方差分析								
11		df	SS	MS	F	Significance F			
12	回归分析	1	20622.17	20622.17	121.5815	3.63018E-09			
13	残差	18	3053.089	169.616					
14	总计	19	23675.26						
15									
16		Coefficients	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	1.062842172	0.096391	11.0264	1.95E-09	0.860332854	1.265351	0.86033285	1.26535149

图32 模型结果

(3) 结果分析：按以上操作步骤，可得到图32所示AR(1)模型结果。因此，零均值化模型的估计结果是 $Z = 1.06284 * Z(-1)$ ，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)。

	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	调节季节指数	消除季节指数	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	1.9445351	1.860272838	1.820272369	163.7117637	3
5	1996.4	105	168.75	163.25	0.6431853	0.651123701	0.637122931	164.8033605	4
6	1997.1	113	178	173.375	0.6517664	0.68532461	0.670588435	168.5087217	5
7	1997.2	159	184.25	181.125	0.8778468	0.891178814	0.872016265	182.3360486	6
8	1997.3	335	193	188.625	1.7760106		1.820272369	184.0383921	7
9	1997.4	130	201.5	197.25	0.6590621		0.637122931	204.0422558	8
10	1998.1	148	210.25	205.875	0.7188828		0.670588435	220.7016887	9
11	1998.2	193	216.5	213.375	0.9045108		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								
15	1991.2								
16	1991.3								
17	1991.4								
18									

图33 旅游人口数相关数据

4.14 季节变动时间序列的分解分析

(1) 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33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。

(2) 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/D4”，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。
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/G2”，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33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34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35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34 回归参数设置

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	回归统计								
4	Multiple R	0.940402088							
5	R Square	0.884356088							
6	Adjusted R Square	0.871506764							
7	标准误差	12.23969631							
8	观测值	11							
9									
10	方差分析								
11		df	SS	MS	F	Significance F			
12	回归分析	1	10310.7	10310.7	68.82511	1.65377E-05			
13	残差	9	1348.291	149.8102					
14	总计	10	11658.99						
15									
16		Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
17	Intercept	120.0409961	7.915042	15.16619	1.03E-07	102.1359275	137.9461	102.13593	137.946065
18	时间	9.68161373	1.167009	8.29609	1.65E-05	7.041655393	12.32157	7.0416554	12.3215721

图35 回归分析计算结果

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36所示。

(3) 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。

	A	B	C	D	E	F	G	H	I
14	1999.1	164.89902					0.670588435		13
15	1991.2	222.87305					0.872016265		14
16	1991.3	482.85492					1.820272369		15
17	1991.4	175.17492					0.637122931		16

图 36 预测结果

§ 5 线性规划

利用 Excel 提供的规划求解法可以解运筹学的问题，诸如线性规划、指派问题、运输问题、食谱问题、机器分配问题、人事安排等等，只要对生产、制造、分配、投资、财务、工程等求最大利润、最小损失等问题，就可使用规划求解法找到答案。

5.1 最佳产品组合（求最大值）

1. 问题描述及模型

有一家玩具工厂每天生产两种玩具（珈菲猫和忍者龟）。制造一个珈菲猫可赚 30 元，制造一个忍者龟可赚 20 元。制造珈菲猫需要 2 小时机工与 1 小时手工；忍者龟需要 1 小时机工与 1 小时手工。不论机工与手工都有时间上的限制，同时产品制造数量亦有限制，条件如下：机工不能超过 100h，手工不能超过 80h，珈菲猫生产量不能超过 45 个，求该工厂每周最佳生产量与最大利润为多少。

设每周生产 x_1 个珈菲猫， x_2 个忍者龟时利润最大，建立如下线性规划模型

$$\begin{aligned}
 \max \quad & z = 30x_1 + 20x_2 \\
 \text{s. t.} \quad & 2x_1 + x_2 \leq 100 \\
 & x_1 + x_2 \leq 80 \\
 & x_1 \leq 45 \\
 & x_1, x_2 \geq 0
 \end{aligned}$$

建立一个如图 37 所示的工作表。

	A	B	C	D	E	F
4					珈菲猫	忍者龟
5				利润	30	20
6				最佳产量		
7						
8			最大用量	已用量		
9		机工时间	100	0	2	1
10		手工时间	80	0	1	1
11		珈菲猫数量	45	0	1	0
12						
16					0	0
17					0	

图 37 人机工时分配问题

表 1 单元格所含公式

单元格	公 式
D9	=E6*E9+F6*F9
D10	=E6*E10+F6*F10
D11	=E6*E11+F6*F11
E16	=E5*E6
F16	=F5*F6
E17	=SUM(E16, F16)

在图 37 有关单元格所含公式如表 1 所示。

2. 求最优解

求最优解的步骤如下：

1) 在图 37 中选取“工具”、“规划求解”，画面显示如图 38 所示。

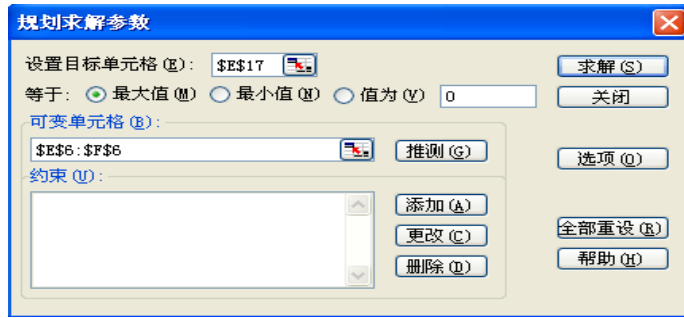


图 38 “规划求解参数”对话框

2) 在“设定目标单元格”框输入 E17，此处为我们的目标解。

3) 选定“最大值”选项。

4) 在“变量单元格”框输入 E6：F6，此为最佳产量。

5) 选取“添加”钮，画面显示如图 39 所示。

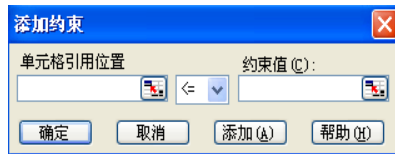


图 39 “添加约束条件”对话框

6) 在“单元格引用位置”框输入 D9:D11。

7) 选定运算符<=。

8) 在“约束值”框输入 C9:C11。

9) 选择“添加”。

10) 重复步骤 6)～8)，以同样方法输入限制条件：E6:F6>=0。

11) 选择“确定”，画面显示如图

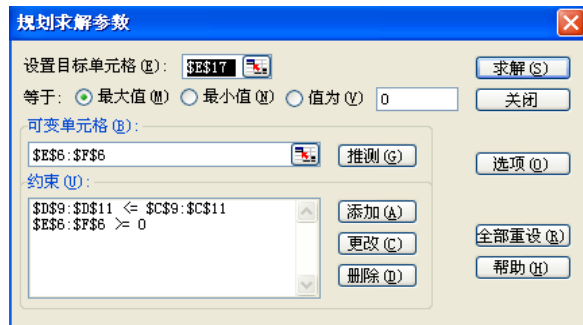


图 40 完成规划求解参数输入画面

12) 选择“求解”按钮，规划求解程序开始运算，解得结果如图 41 所示。

在图 41 中，单元格 E6, F6 显示珈菲猫和忍者龟的最佳产量分别为 20 个，60 个，单元格 E17 显示最大利润为 1800。

另外，在“报告”框中有三个选项，分别为“运算结果报告”、“敏感性报告”、“极限值报告”，表示有三种报表形式供选择参考。

	B	C	D	E	F	G	H	I	J	K
4				咖啡猫	忍者龟					
5			利润	30	20					
6			最佳产量	20	60					
7										
8		最大用量	已用量							
9	机工时间	100	100	2	1					
10	手工时间	80	80	1	1					
11	咖啡猫数量	45	20	1	0					
12										
13										
14										
15										
16				600	1200					
17				1800						

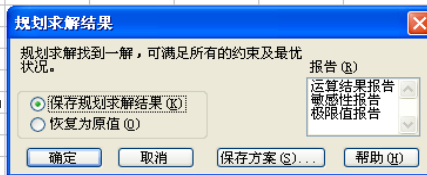


图 41 “规划求解结果”对话框

注意：在使用“规划求解”以前，一定先要经过安装，使用“工具”、“加载宏”，在“加载宏”框里，打开“规划求解”，选择“确定”。

3. 分析求解结果

在图 41 中，“报告”框有三种报表选项，可以只选 1 种、2 种或 3 种，我们打算选择 3 种，方法如下：

1) 按住 Ctrl 键，使用鼠标左键分别选定 3 种报告。

2) 选择“确定”，就会在画面底端的工作表标签中显示 3 个报表，分别为“运算结果报告 1”、“敏感性报告 1”、“极限值报告 1”。

3) 选取“运算结果报告 1”，画面就会显示“运算结果报告 1”。

4. 重设规划求解参数

当要重新开始算新的问题时，应重设规划求解参数，方法如下：

1) 选取“工具”、“规划求解...”。

2) 选取“全部重设”，再选取“确定”，画面会显示“规划求解参数”对话框，如图 38。

5. 保存模型

当选择“文件”、“保存”指令保存文件时，模型也会跟着一并保存，但是保存的模型为最后一次修改过的模型，若想要保存某些特定模型，可以采用“规划求解参数”对话框中的“选项”。选项里的“保存模型...”钮可保存模型，方法如下：

1) 先选定空白单元格范围准备存放模型，例如，选定范围 B20:B25。

2) 选取“工具”、“规划求解...”指令，显示“规划求解参数”对话框。

3) 选取“选项”钮，画面显示如图 42 所示的“规划求解选项”对话框。

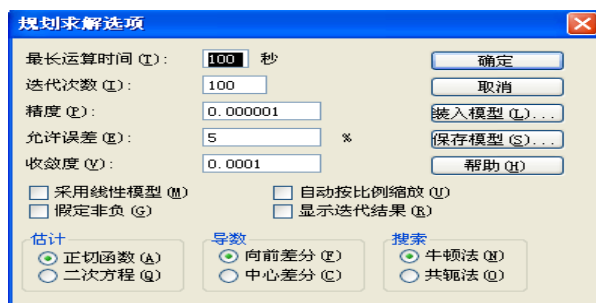


图 42 “规划求解选项”对话框

4) 选取“保存模型”钮。

5) 选取“确定”，现在单元格 B20:B25 已存有模型。

若未装入模型，先选定单元格 B20:B25，然后选取“选项”、“装入模型”。

提示：在同一个工作表中，可以保存若干个模型，只要在不同的空白区域保存不同的模型即可。因此当要执行某个模型，只需先选定该模型所在的单元格范围，接着使用“装入模型”钮装入即可。

5.2 营养调配（求最小值）

1. 问题描述及模型

艾美为了健康，打算每天由柳丁、木瓜、香蕉、胡萝卜 4 种水果获取维生素 A、B、C、D。假设柳丁每个 20 元，木瓜每个 30 元，香蕉每根 15 元，胡萝卜每根 25 元。每天她最少必需摄取 500 单位维生素 A，250 单位维生素 B，600 单位维生素 C，以及 300 单位维生素 D，各种水果所含维生素及价格明细如表 2 所示。试求花费最少，又能满足各种维生素摄取量的最佳解。

表 2 维生素含量、价格和需求表

水果 摄取量	柳丁	木瓜	香蕉	胡萝卜	最少需求
A	5	50	10	80	500
B	15	30	40	20	250
C	90	20	30	30	600
D	10	15	10	50	300
市价	20	30	15	25	

设柳丁、木瓜、香蕉、胡萝卜的最佳需求量分别为 $x_i(i=1,2,3,4)$ 个。建立如下线性规划模型。

$$\begin{aligned} \min \quad & z = 20x_1 + 30x_2 + 15x_3 + 25x_4 \\ \text{s. t.} \quad & 5x_1 + 50x_2 + 10x_3 + 80x_4 \geq 500 \\ & 15x_1 + 30x_2 + 40x_3 + 20x_4 \geq 250 \\ & 90x_1 + 20x_2 + 30x_3 + 30x_4 \geq 600 \\ & 10x_1 + 15x_2 + 10x_3 + 50x_4 \geq 300 \\ & x_i \geq 0 \text{ 且为整数, } i = 1,2,3,4 \end{aligned}$$

建立一个如图 43 所示的工作表。

	A	B	C	D	E	F	G	H
3								
4					柳丁	木瓜	香蕉	胡萝卜
5				单价	20	30	15	25
6				数量	0	0	0	0
7								
8			最少需求	最佳结果				
9		维生素A	500	0	5	50	10	80
10		维生素B	250	0	15	30	40	20
11		维生素C	600	0	90	20	30	30
12		维生素D	300	0	10	15	10	50
13								
14								
15								
16					0	0	0	0
17		支出			0			

图 43 营养分配问题

在图 43 中，有关单元格所含公式如表 3 所示。

表3 单元格所含公式

单元格	公 式
D9	=E6*E9+F6*F9+G6*G9+H6*H9
D10	=E6*E10+F6*F10+G6*G10+H6*H10
D11	=E6*E11+F6*F11+G6*G11+H6*H11
D12	=E6*E12+F6*F12+G6*G12+H6*H12
E16	=E5*E6
F16	=F5*F6
G16	=G5*G6
H16	=H5*H6
E17	=SUM(E16:H16)

2. 求最佳组合解

在图 43 中，求最优解，步骤如下：

- 1) 选取“工具”、“规划求解…”。
- 2) 在“设定目标单元格”框输入 E17。
- 3) 选定“最小值”选项。
- 4) 在“变量单元格”框输入 E6:H6。
- 5) 选取“添加”，分别输入表 4 中所示的参数。

表4 输入约束条件参数

单元格引用地址	运算符号	限制值
D9:D12	>=	C9:C12
E6:H6	>=	0

- 6) 选择“确定”，完成输入。
- 7) 由“规划求解参数”框，选择“求解”。
- 8) 选择“确定”，计算结果如图 44 所示。

	A	B	C	D	E	F	G	H
3								
4					柳丁	木瓜	香蕉	胡葡萄
5				单价	20	30	15	25
6				数量	4.140127	0	1.815287	5.764331
7								
8			最少需求	最佳结果				
9		维生素A	500	500	5	50	10	80
10		维生素B	250	250	15	30	40	20
11		维生素C	600	600	90	20	30	30
12		维生素D	300	347.7707006	10	15	10	50
13								
14								
15								
16					82.80254777	0	27.22929936	144.1082803
17		支出			254.1401274			

图 44 最佳运算结果

图 44 算出每日支出最少为 254.14 元，每日吸收维生素 D 为 345.7707006 单位，其它维生素维持最低需求量。购买水果数量分别如单元格 E6:H6 所示。

3. 取得整数解

在图 44 中，发现数字解都为小数，可以把它们改为用整数显示，方法如下：

- 1) 按住 Ctrl 键，分别选定单元格 E6、G6、H6、D12、E16、G16、H16、E17。

- 2) 选取“格式”、“单元格...”、“数字”。
- 3) 在“分类”框中选定“自定义”的0格式。
- 4) 选择“确定”，发现画面显示结果如图 45 所示。

	B	C	D	E	F	G	H
3							
4				柳丁	木瓜	香蕉	胡葡萄
5			单价	20	30	15	25
6			数量	4	0	2	6
7							
8		最少需求	最佳结果				
9	维生素A	500	500	5	50	10	80
10	维生素B	250	250	15	30	40	20
11	维生素C	600	600	90	20	30	30
12	维生素D	300	348	10	15	10	50
13							
14							
15							
16				83	0	27	144
17	支出			254			

图 45 数字以整数显示

习题

1. 某企业 50 名工人日产量数据如图 46 所示。运用 Excel 进行数据分组并制作直方图。

	A	B	C	D	E
1	148	116	128	125	129
2	140	109	123	137	119
3	127	132	114	107	124
4	120	135	108	113	130
5	110	129	132	132	118
6	104	123	124	140	107
7	128	110	120	137	113
8	135	108	125	119	122
9	129	148	116	119	128
10	123	135	118	127	114

图 46 日产量数据

2. 某工厂制造 A, B 两种产品，制造 A 每吨需用煤 9 吨，电力 4 千瓦，3 个工作日；制造 B 每吨需用煤 5 吨，电力 5 千瓦，10 个工作日。已知制造产品 A 和 B 每吨分别获利 7000 元和 1200 元，现在该厂由于条件限制，只有煤 360 吨，电力 200 千瓦，工作日 300 个可以利用，问 A、B 两种产品各应生产多少吨才能获利最大？