

# 目 录

用 Excel 做数据分析—移动平均.....	2
用 Excel 做数据分析—抽样分析工具.....	4
实现 Excel 动态链接外部数据库.....	8
用 Excel 做数据分析—相关系数与协方差.....	18
用 Excel 做数据排序的常用方法与技巧.....	20
Excel 的数据分析—排位与百分比.....	24
用 Excel 做数据分析—描述统计.....	27
用 Excel 做数据分析—直方图.....	30
excel 数据分析.....	35
使用 Excel 做数据分析之回归分析方法.....	39
用单元格数据作为 Excel 工作簿名称.....	44

## 用 Excel 做数据分析—移动平均

某化工反应过程，每隔 2 分钟对系统测取一次压力数据。由于反应的特殊性，需要考察每 8 分钟的压力平均值，如果该压力平均值高于 15MPa，则认为自属于该平均值计算范围内的第一个压力数据出现时进入反应阶段，请使用 Excel 给出反应阶段时间的区间。

移动平均就是对一系列变化的数据按照指定的数据数量依次求取平均，并以此作为数据变化的趋势供分析人员参考。移动平均在生活中也不乏见，气象意义上的四季界定就是移动平均最好的应用。

注:本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘支持下加载“分析数据库”。加载成功后，可以在“工具”下拉菜单中看到“数据分析”选项。

### 操作步骤

1.打开原始数据表格，制作本实例的原始数据要求单列，请确认数据的类型。本实例为压力随时间变化成对数据，在数据分析时仅采用压力数据列。

	A	B	C
1	MC合成实验温度压力随温度变化表		
2	时间	系统压力/Mpa	
3	8:30	11.5	
4	8:32	12.5	
5	8:34	13.3	
6	8:36	15.2	
7	8:38	13	
8	8:40	12.1	
9	8:42	11.8	
10	8:44	10.1	
11	8:46	12.1	
12	8:48	12.6	
13	8:50	14.3	
14	8:52	13.4	
15	8:54	13.2	
16	8:56	10.3	



需要注意的是，因为平均值的求取需要一定的数据量，那么就要求原始数据量不少于求取平均值的个数，在 Excel 中规定数据量不少于 4。

2.选择“工具”-“数据分析”-“直方图”后，出现属性设置框，依次选择:

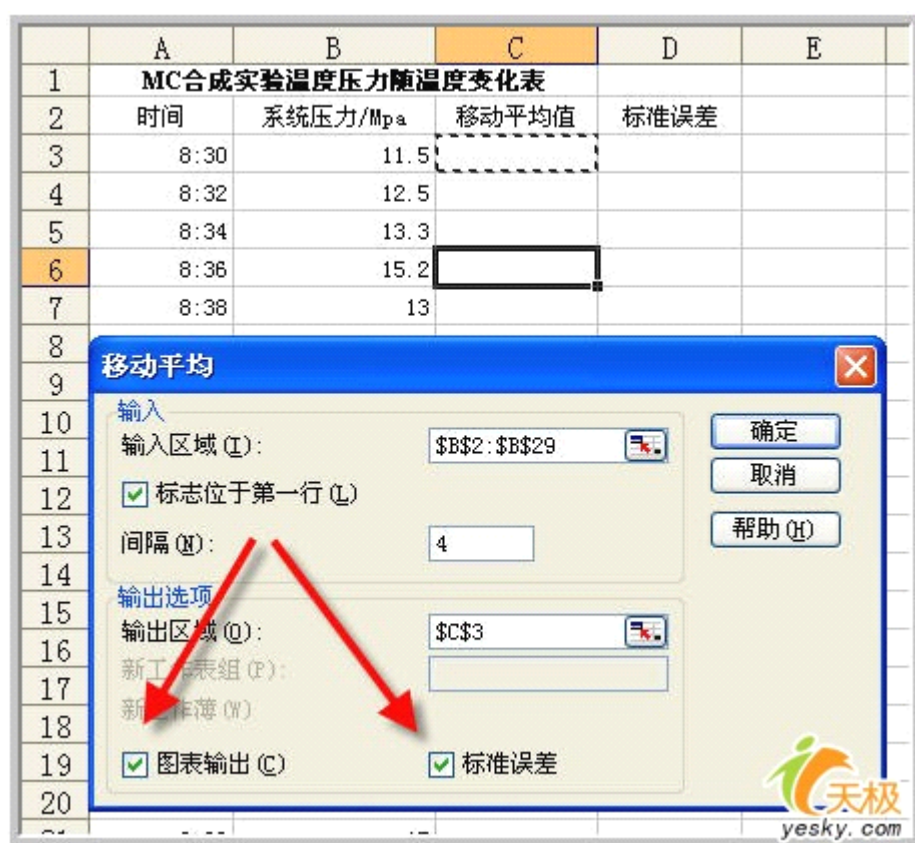
输入区域:原始数据区域;如果有数据标签可以选择“标志位于第一行”;

输出区域:移动平均数值显示区域;

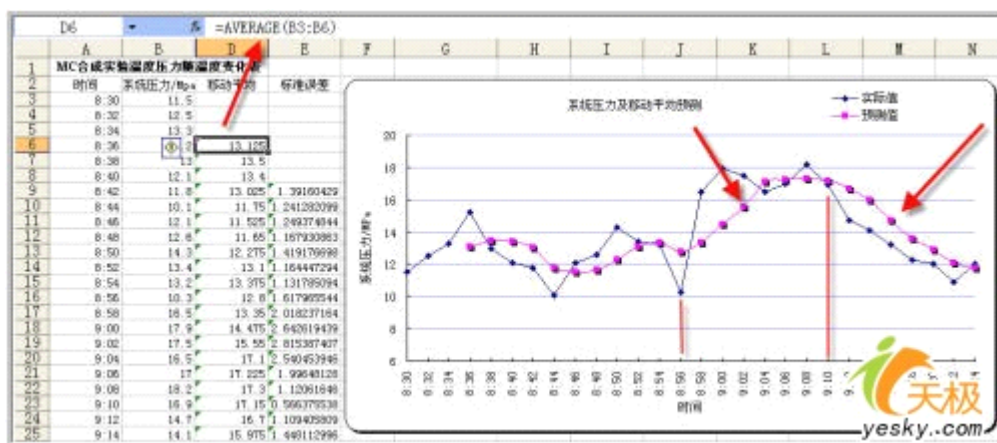
间隔:指定使用几组数据来得出平均值;

图表输出;原始数据和移动平均数值会以图表的形式来显示，以供比较;

标准误差:实际数据与预测数据(移动平均数据)的标准差,用以显示预测与实际值的差距。数字越小则表明预测情况越好。



3.输入完毕后，则可立即生成相应的数据和图表。



从生成的图表上可以看出很多信息。

根据要求，生成的移动平均数值在 9:02 时已经达到了 15.55MPa，也就是说，包含本次数据在内的四个数据前就已经达到了 15MPa，那么说明在 8 分钟前，也就是 8:56 时，系统进入反应阶段；采用同样的分析方法可以知道，反应阶段结束于 9:10，反应阶段时间区间为 8:56-9:10，共持续 14 分钟。

单击其中一个单元格“D6”，可以看出它是“B3-B6”的平均值，而单元格“E11”则是“SQRT(SUMXMY2(B6:B9,D6:D9)/4)”，它的意思是 B6-B9，D6-D9 对应数据的差的平方的平均值再取平方根，也就是数组的标准差。

## 用 Excel 做数据分析—抽样分析工具

省教育厅派专家组进行某校检查学生考试试卷，专家组拟对总体进行抽样调查，对学校某班的全体同学随机抽取 25 名作为调查样本。为了保证结果的非人为性，采用 Excel 帮助专家组做出抽查的结果。

抽样分析工具以数据源区域为总体，从而为其创建一个样本。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。如果确认数据源区域中的数据是周期性的，还可以对一个周期中特定时间段中的数值进行采样。也可以采用随机抽样，满足用户保证抽样的代表性的要求。

注：本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘的支持下加载“数据分析库”。加载成功后，可以在工具的下拉菜单中看到“数据分析”选项。

**操作步骤：**

1. 打开原始数据表格，制作本实例的原始数据无特殊要求，只要满足行或列中为同一属性数值即可。实例中显示的是学生学号。

	A	C	D	E
1	三维中学2006年第一学期期中考试化学成绩统计表			
2	学号	抽样结果		
3	0701			
4	0702			
5	0703			
6	0704			
7	0705			
8	0706			
9	0707			
10	0708			
11	0709			
12	0710			
13	0711			
14	0712			
15	0713			
16	0714			
17	0715			
18	0716			

2. 选择“工具”——“数据分析”——“抽样”后，出现对话框，依次选择：

	A	C	D	E	F
1	三维中学2006年第一学期期中考试化学成绩统计表				
2	学号	抽样结果			
3	0701				
4	0702				
5	0703				
6	0704				
7	0705				
8	<div> <div>抽样</div> <div> <div>输入</div> <div>           输入区域(I): <input type="text" value="\$A\$2:\$A\$58"/> </div> <div> <input checked="" type="checkbox"/> 标志(L)         </div> </div> <div> <div>抽样方法</div> <div> <input type="radio"/> 周期(E)           <div>间隔: <input type="text"/></div> </div> <div> <input checked="" type="radio"/> 随机(E)           <div>样本数: <input type="text" value="25"/></div> </div> </div> <div> <div>输出选项</div> <div> <input checked="" type="radio"/> 输出区域(O): <input type="text" value="\$C\$3:\$C\$3"/> </div> <div> <input type="radio"/> 新工作表组(E): <input type="text"/> </div> <div> <input type="radio"/> 新工作簿(W)           </div> </div> <div> <div>确定</div> <div>取消</div> <div>帮助(H)</div> </div> </div>				

输入区域：把原始总体数据放在此区域中，数据类型不限，数值型或者文本型均可；

抽样方法：有间隔和随机两种。间隔抽样需要输入周期间隔，输入区域中位于间隔点处的数值以及此后每一个间隔点处的数值将被复制到输出列中。当到达输入区域的末尾时，抽样将停止。(在本例题中没有采用)；随机抽样是指直接输入样本数，电脑自行进行抽样，不用受间隔的规律限制；

样本数：在此输入需要在输出列中显示需要抽取总体中数据的个数。每个数值是从输入区域中的随机位置上抽取出来的，请注意：任何数值都可以被多次抽取!所以抽样所得数据实际上会有可能小于所需数量。本文末尾给出了一种处理方法；

输出区域：在此输入对输出表左上角单元格的引用。所有数据均将写在该单元格下方的单列里。如果选择的是“周期”，则输出表中数值的个数等于输入区域中数值的个数除以“间隔”。如果选择的是“随机”，则输出表中数值的个数等于“样本数”；

3.然后单击确定就可以显示结果了(这是电脑自行随机抽样的结果)。

	A	C	D	E
1	三维中学2006年第一学期期中考试化学成绩统计表			
2	学号	抽样结果		
3	0701	0740		
4	0702	0707		
5	0703	0718		
6	0704	0751		
7	0705	0720		
8	0706	0704		
9	0707	0718		
10	0708	0701		
11	0709	0739		
12	0710	0743		
13	0711	0739		
14	0712	0702		
15	0713	0703		
16	0714	0721		
17	0715	0707		
18	0716	0744		
19	0717	0709		
20	0718	0708		
21	0719	0701		
22	0720	0755		
23	0721	0747		
24	0722	0747		
25	0723	0726		
26	0724	0710		
27	0725	0710		

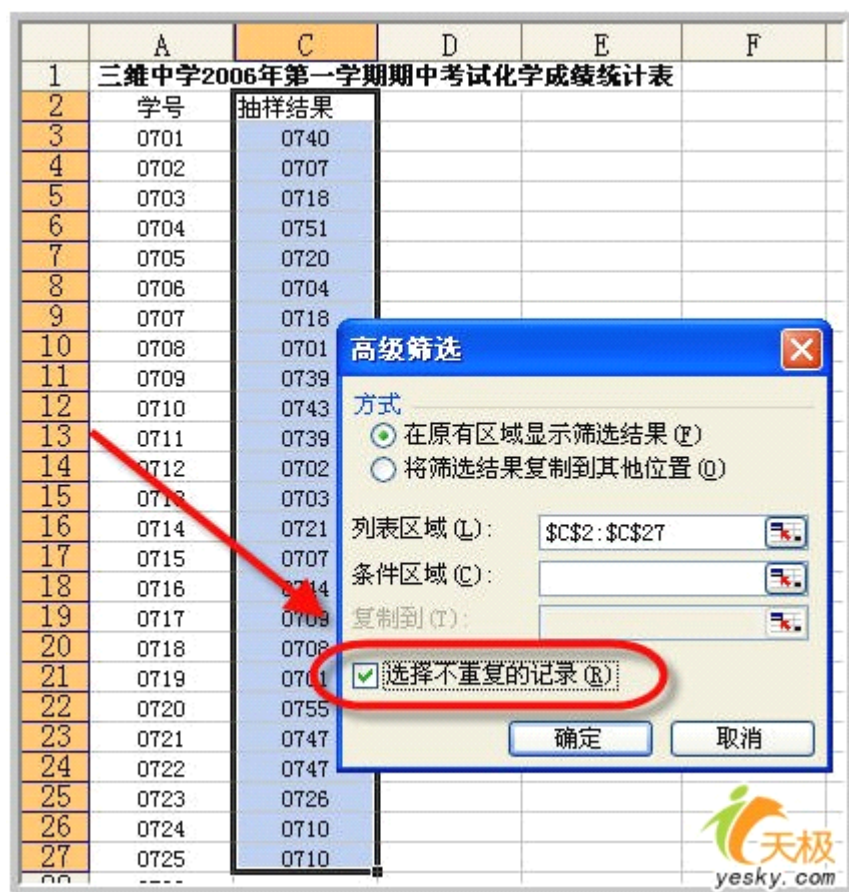
[原始数据文档在这里下载>>](#) [操作结果文档在这里下载>>](#)

需要说明的情况：



由于随机抽样时总体中的每个数据都可以被多次抽取,所以在样本中的数据一般都会有重复现象,解决此问题有待于程序的完善。可以使用“筛选”功能对所得数据进行筛选。

选中样本数据列,依次执行“数据”-“筛选”-“高级筛选”,如下图所示。



最后的样本结果如下图所示,请您根据经验适当调整在数据样本选取时的数量设置,以使最终所得样本数量不少于所需数量。

	A	C	D	E
1	三维中学2006年第一学期期中考试化学成绩统计表			
2	学号	抽样结果		
3	0701	0740		
4	0702	0707		
5	0703	0718		
6	0704	0751		
7	0705	0720		
8	0706	0704		
10	0708	0701		
11	0709	0739		
12	0710	0743		
14	0712	0702		
15	0713	0703		
16	0714	0721		
18	0716	0744		
19	0717	0709		
20	0718	0708		
22	0720	0755		
23	0721	0747		
25	0723	0726		
26	0724	0710		
28	0726			
29	0727			

筛选后的无  
重复数据

  
 yesky.com

## 实现 Excel 动态链接外部数据库

我们有时需要在 Excel 中调取其他数据库的数据，并且希望其他数据库数据改变时，Excel 中调取的数据也随之动态改变。下面介绍在 Excel 中通过“新建数据库查询”(Microsoft Query)的方法来实现动态链接数据库。

您在 Excel 中第一次使用“新建数据库查询”查询数据时，如果系统未安装 Microsoft Query，系统会提示您安装。

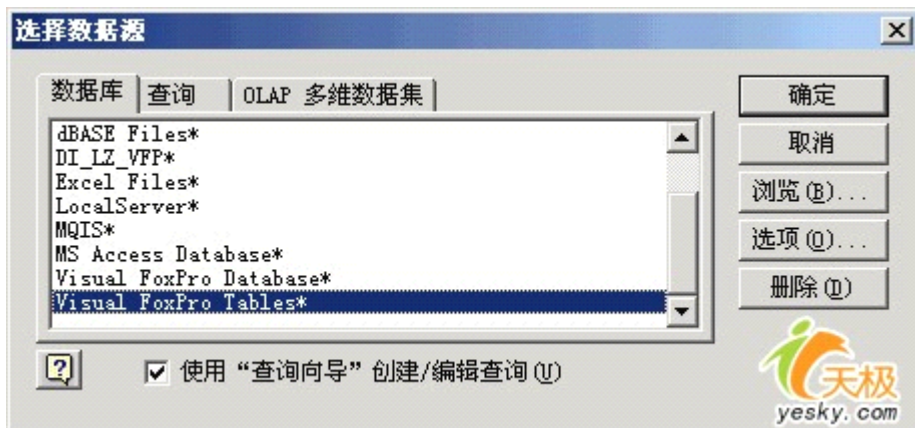
### 一、Excel 链接 vfp 数据库

vfp 是常用的数据库，先以此来介绍。假设从 px.dbf 表中取出相应学校的数据，已知学校的后 5 位代码，数据所在字段为 x16，数据值为 03 行和 06 行相加，放入 Excel 单个单元格中。

步骤如下：

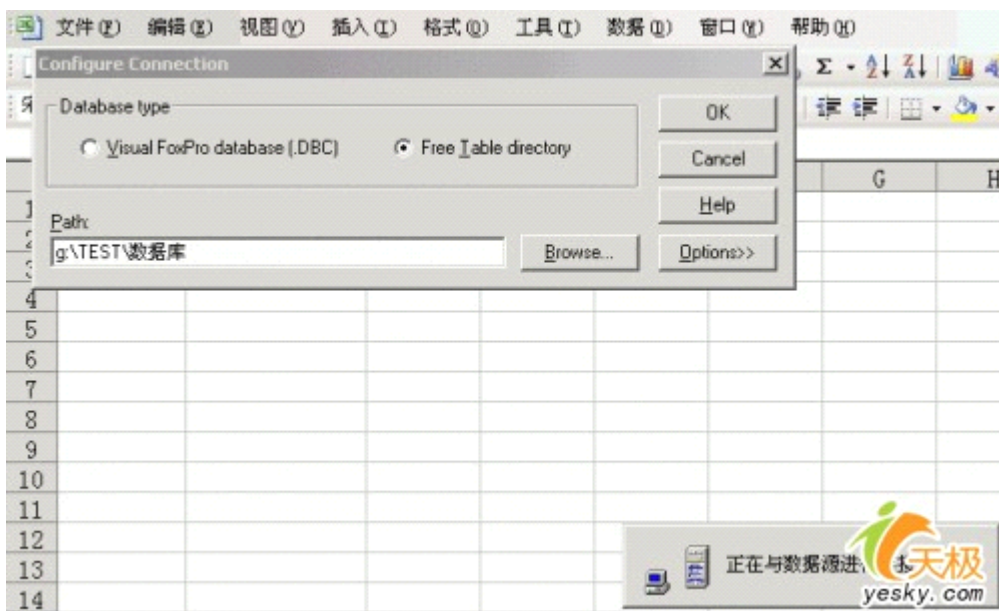
1、点击菜单“数据”—“导入外部数据”—“新建数据库查询”，出现“选择数据源”对话框，选择“Visual Foxpro Tables\*”，按“确定”。如果您对 SQL 语句不是很熟悉的话，请使用默认的“查询向导”创建。





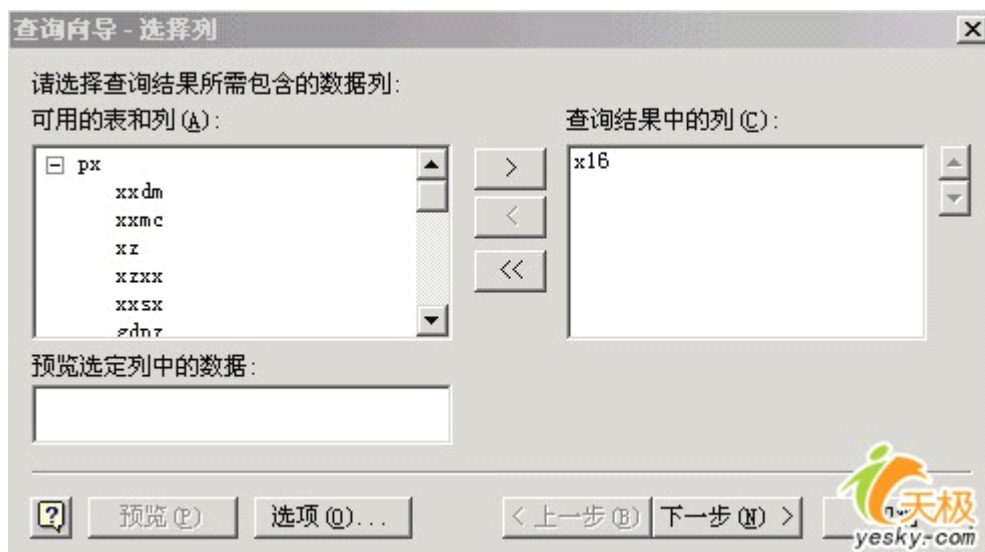
(图 1)

2、在“configure connection”对话框中，按“Browse”按钮选择表 px.dbf 所在的路径，按“ok”键。



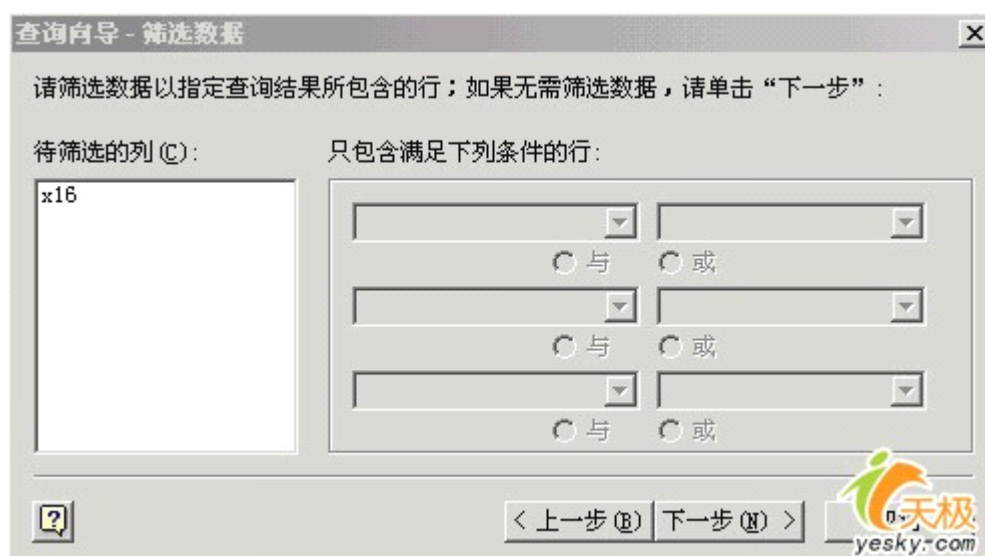
(图 2)

3、在“查询向导—选择列”对话框中，选择查询结果中的列，要选取的数据在“x16”字段，所以将可用表“px”中的“x16”列移至“查询结果中的列”，按“下一步”。



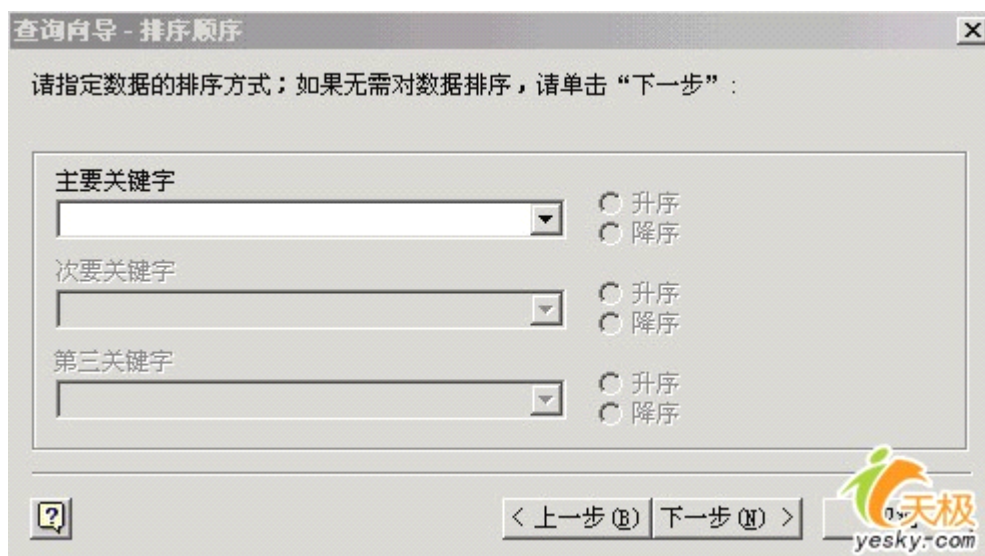
(图 3)

4、在“查询向导—筛选数据”对话框中，因为此例中设置的条件在字段学校代码“xxdm”和行号“bh”，非字段“x16”，所以点击“下一步”，暂时不筛选数据。



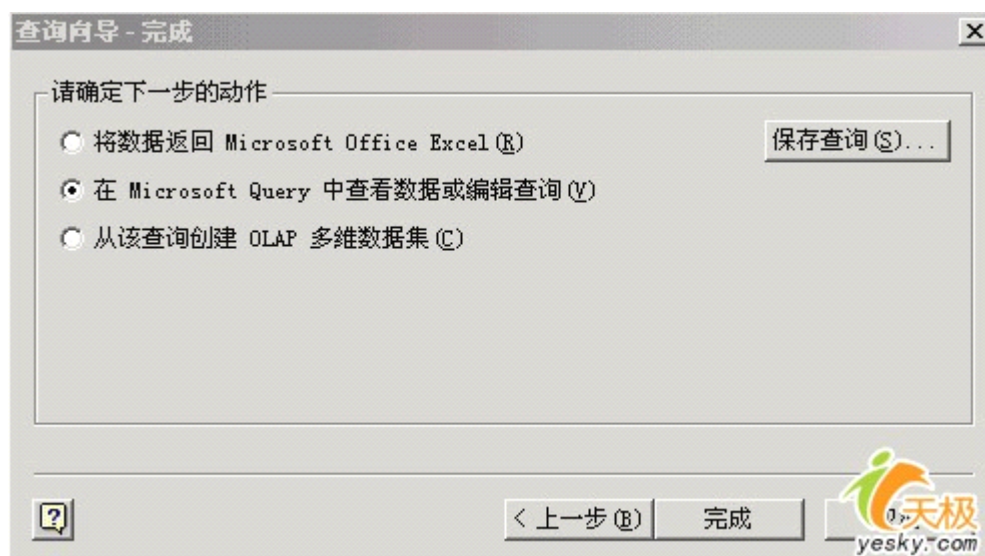
(图 4)

5、在“查询向导—排序顺序”对话框中，因为此例中所取的是单个数据，所在点击“下一步”。



(图 5)

6、出现“查询向导—完成”对话框，因还要对数据进行筛选，所以选择“在 Microsoft Query 中查看数据或编辑查询”，点击“完成”。

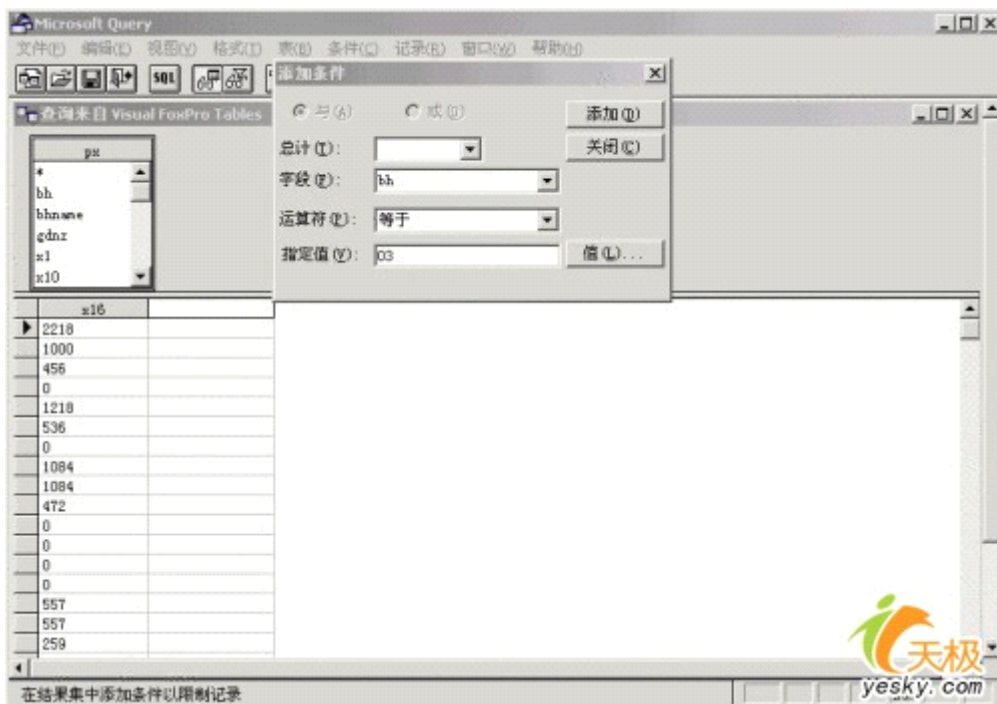


(图 6)

7、进入“Microsoft Query”，看到查询结果 x16 列出了所有行的数据。

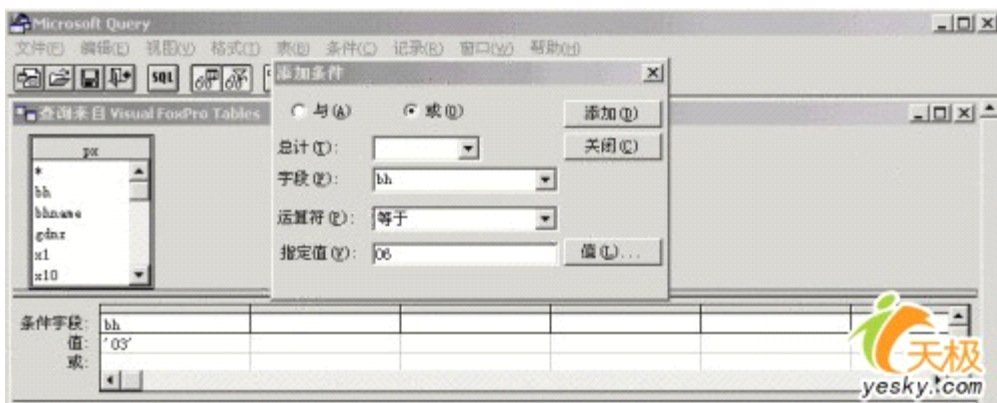
接下来要通过设置条件来筛选所需的数据。

点击菜单“条件”—“增加条件(A)...”，字段“bh”,运算符“等于”，指定值为“03”，按“添加”。



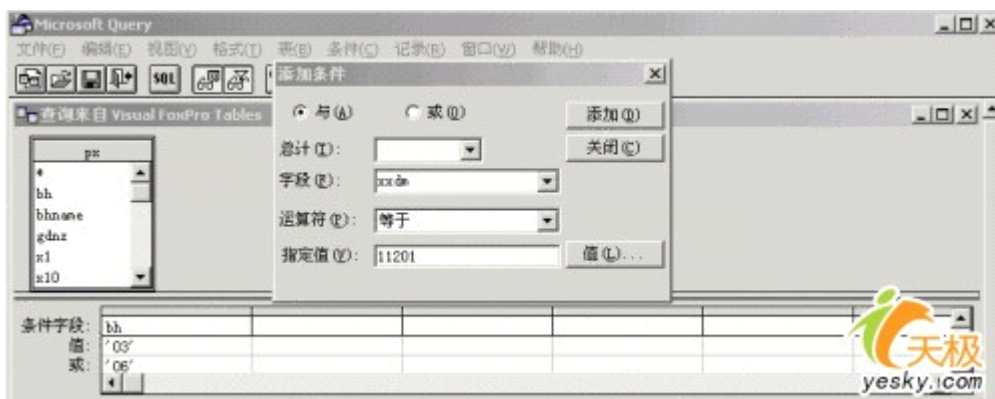
(图 7)

继续“添加条件”对话框，选择“或”，字段“bh”,运算符“等于”，指定值为“06”，按“添加”。



(图 8)

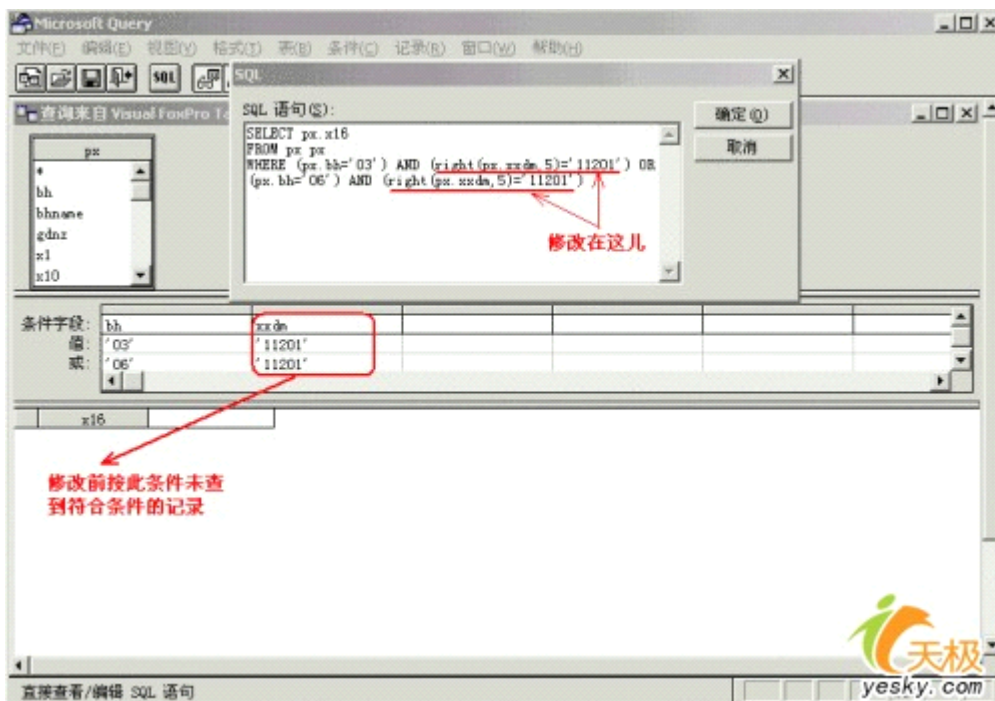
继续“添加条件”对话框，选择“与”，字段“xxdm”,运算符“等于”，指定值为“11201”，按“添加”后按“关闭”。



(图 9)

但是看到符合条件 x16 为空，这是因为学校代码的后五位为“11201”。

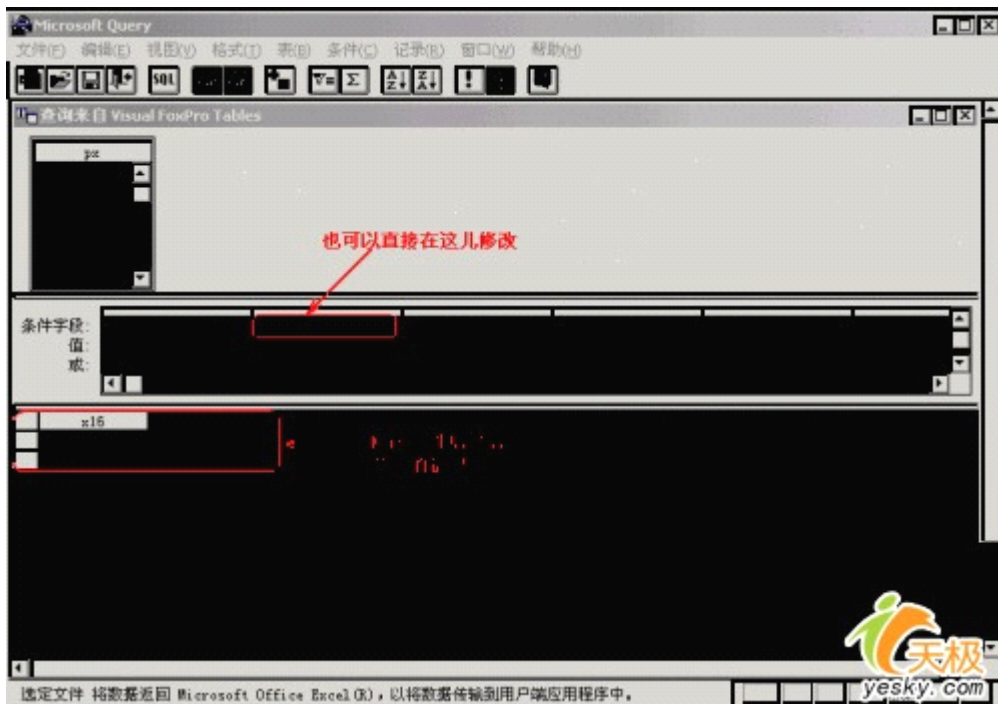
可以通过“显示 SQL”来直接修改条件语句,将 SQL 语句中 px.xxdm="11201"改成 right(px.xxdm,5)="11201",按“确定”。



(图 10)

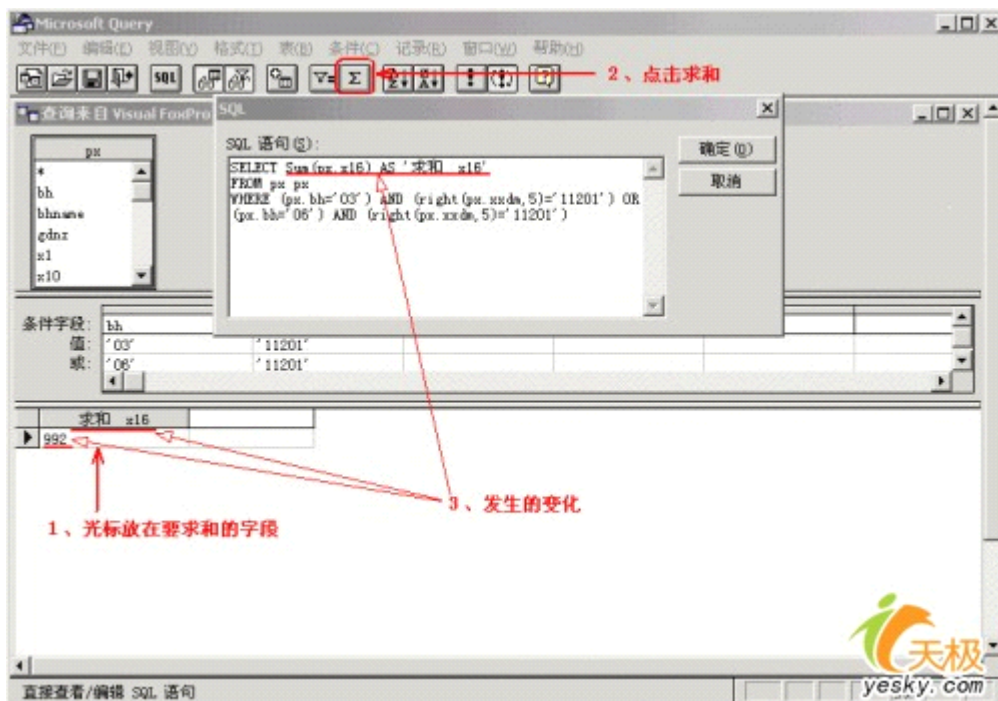
也可以在“条件字段”直接修改。





(图 11)

光标放在 x16 字段，点击“循环总计”，将查询出的两个值相加，当然也可以直接在 SQL 语句中修改。



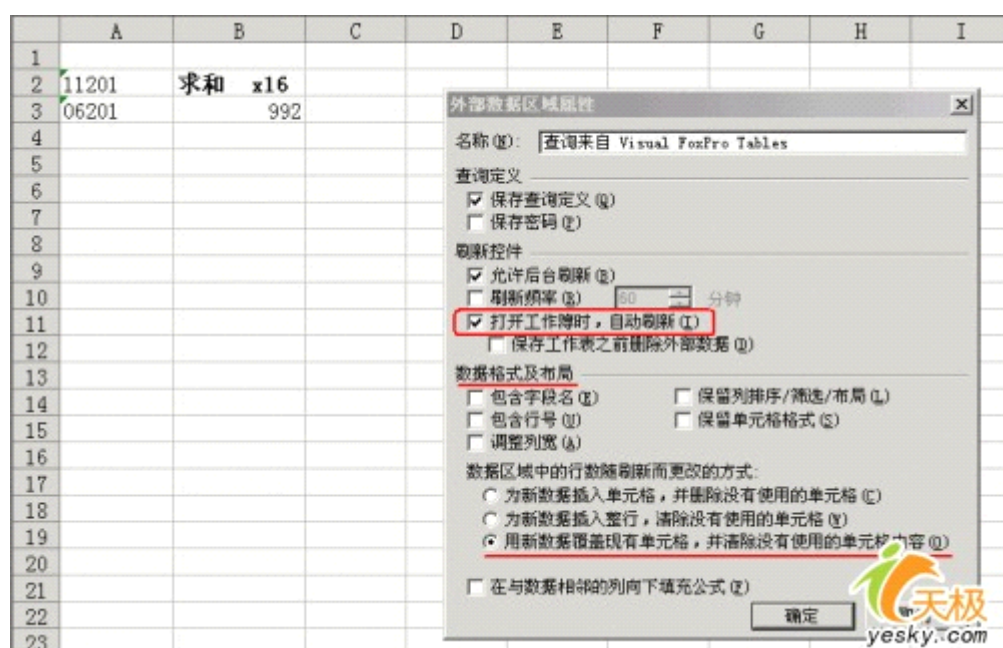
(图 12)

8、关闭“Microsoft Query”，弹出“导入数据”对话框，选择“数据的放置位置”，选择“\$B\$2”，按“确定”。



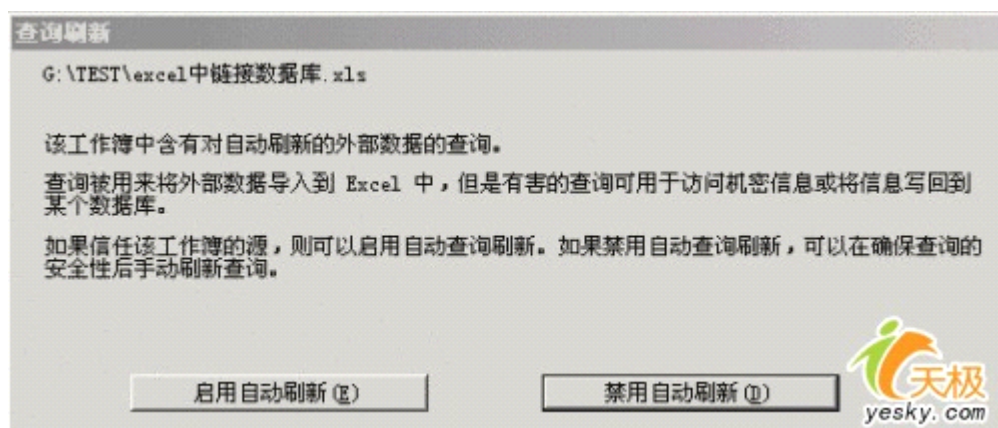
(图 13)

9、但我们发现显示了标题行，真正的数据却显示到了\$B\$3，这时我们需要修改\$B\$2单元格的“数据区域属性”。点击\$B\$2单元格，按右键菜单的“数据区域属性”，弹出“外部数据区域属性”窗口，将“数据格式与布局”中“包含字段名”等前面的对勾去掉，将“数据区域中的行数随刷新而更改的方式”改成“用新数据覆盖现有单元格，并清除没有使用的单元格内容”，按“确定”。但还是有标题，没有关系，点击\$B\$2单元格，按右键菜单的“刷新数据”即可。



(图 14)

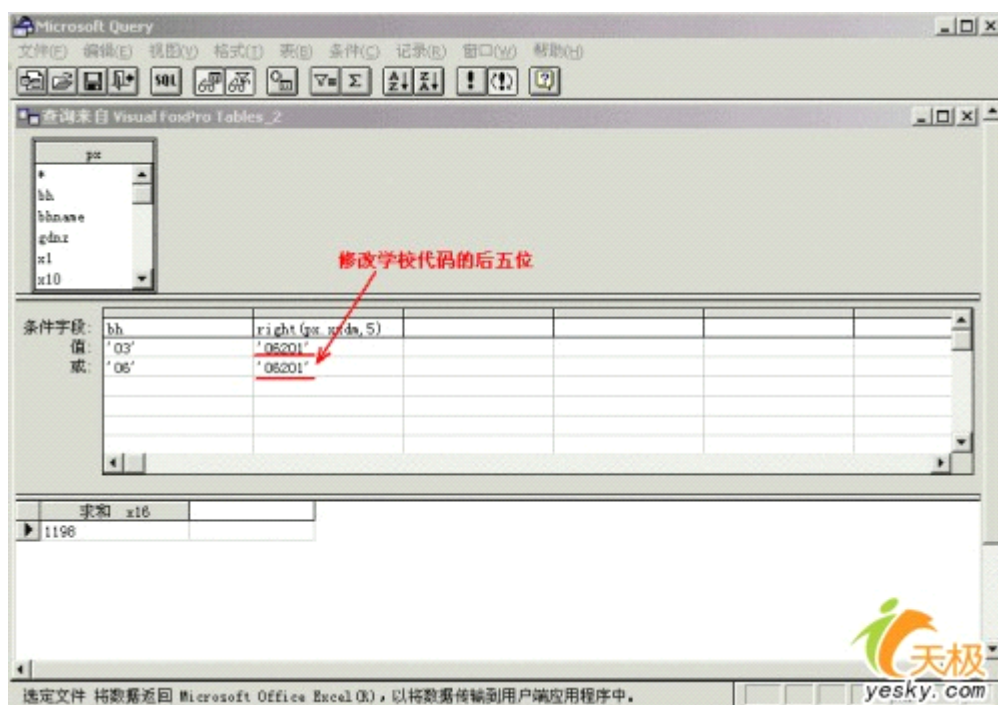
10、如果 px.dbf 数据库有所改变，那么如何在 Excel 中刷新，可以点击链接数据所在单元格右键菜单中的“刷新”手工刷新;也可以在“数据区域属性”—“刷新控件”中选中“打开工作簿时，自动刷新”(见图 14)，从而在每次打开工作簿时，在“查询刷新”的提示框中按“启动自动刷新”按钮刷新(见图 15)。



(图 15)

11、如果需要编辑查询，则点击\$B\$2 单元格，按右键菜单的“查询编辑”，如果出现“查询向导”无法编辑此查询的提示，只要按“确定”即可进入“Microsoft Query”。

12、\$B\$3 单元格查询条件与\$B\$2 相差的只是学校代码，可以将\$B\$2 单元格复制到\$B\$3 单位格，再用\$B\$3 的“编辑查询”来修改，修改一下学校代码的值即可。



(图 16)

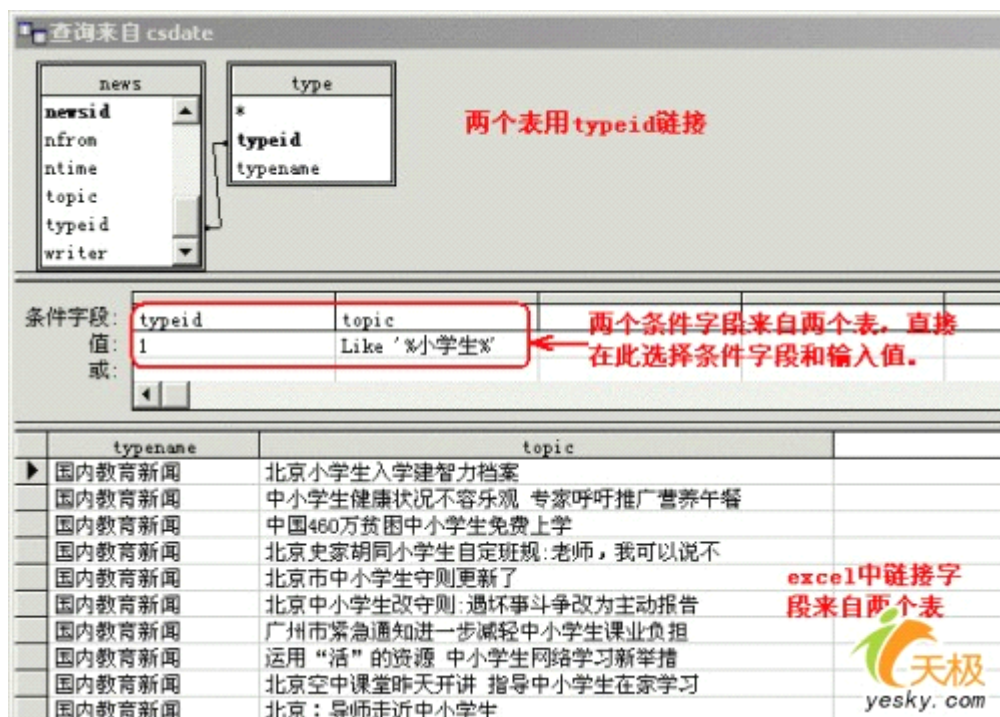
## 二、Excel 链接 SQL 数据库

链接 SQL 数据库与链接 vpf 大致是相同的。这里介绍的是连接两表的查询。

点击菜单“数据”—“导入外部数据”—“新建数据库查询”，出现“选择数据源”对话框，可直接选取所需链接的 SQL 数据库名。

如果 SQL 数据库有密码的话，将提示输入密码，按“确定”。如果 SQL SERVER 服务管理器未开启的话，将出现错误提示。

如果不使用“查询向导”的话，可直接进入“Microsoft Query”。添加表 news 和表 type，news 表中的 typeid 和 type 表的 typeid 建立关联。在“条件字段”中选择字段名，并在“值”中输入条件值，这里输入的条件是 type 表的 typeid 字段值为 1 且 news 表的 topic 中含有“小学生”。点击选择要输出的字段名，这里选择的是 type 表中的 typename 字段和 news 表的 topic 字段。



(图 17)

当然，如果您对 SQL 语句熟悉，也可以直接输入 SQL 语句。

接下来的步骤请参考一、中的 8、至 10、。



## 用 Excel 做数据分析—相关系数与协方差

化学合成实验中经常需要考察压力随温度的变化情况。某次实验在两个不同的反应器中进行同一条件下实验得到两组温度与压力相关数据，试分析它们与温度的关联关系，并对在不同反应器内进行同一条件下反应的可靠性给出依据。

相关系数是描述两个测量值变量之间的离散程度的指标。用于判断两个测量值变量的变化是否相关，即，一个变量的较大值是否与另一个变量的较大值相关联(正相关)；或者一个变量的较小值是否与另一个变量的较大值相关联(负相关)；还是两个变量中的值互不关联(相关系数近似于零)。设(X,Y)为二元随机变量，那么：

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \sqrt{DY}}$$

为随机变量 X 与 Y 的相关系数。p 是度量随机变量 X 与 Y 之间线性相关密切程度的数字特征。

注:本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘中加载“分析数据库”。加载成功后，可以在“工具”下拉菜单中看到“数据分析”选项。

### 操作步骤

1. 打开原始数据表格，制作本实例的原始数据需要满足两组或两组以上的数据，结果将给出其中任意两项的相关系数。

	A	B	C
1	MC合成实验温度压力随温度变化表		
2	温度(℃)	压力A(MPa)	压力B(MPa)
3	70	0	0
4	75	0.1	0.1
5	80	0.15	0.15
6	85	0.2	0.21
7	90	0.23	0.25
8	95	0.28	0.28
9	100	0.31	0.33
10	105	0.34	0.38
11	110	0.4	0.45
12	115	0.45	0.5
13	120	0.52	
14	125	0.61	
15	130	0.71	



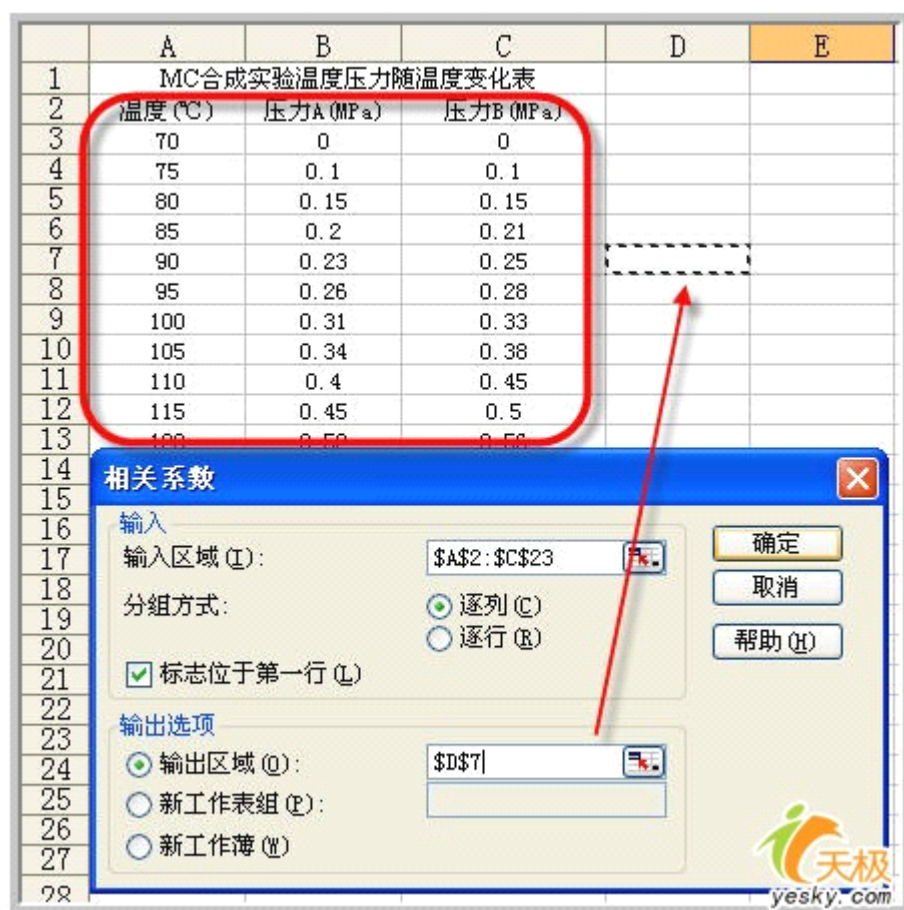
2. 选择“工具”-“数据分析”-“描述统计”后，出现属性设置框，依次选择：



输入区域:选择数据区域, 注意需要满足至少两组数据。如果有数据标志, 注意同时勾选下方“标志位于第一行”;

分组方式:指示输入区域中的数据是按行还是按列考虑, 请根据原数据格式选择;

输出区域可以选择本表、新工作表组或是新工作簿;



3.点击“确定”即可看到生成的报表。

可以看到, 在相应区域生成了一个 3×3 的矩阵, 数据项目的交叉处就是其相关系数。显然, 数据与本身是完全相关的, 相关系数在对角线上显示为 1; 两组数据间在矩阵上有两个位置, 它们是相同的, 故右上侧重复部分不显示数据。左下侧相应位置分别是温度与压力 A、B 和两组压力数据间的相关系数。

	A	B	C	D	E	F	G
1	MC合成实验温度压力随温度变化表						
2	温度(℃)	压力A(MPa)	压力B(MPa)				
3	70	0	0				
4	75	0.1	0.1				
5	80	0.15	0.15				
6	85	0.2	0.21				
7	90	0.23	0.25				
8	95	0.28	0.28				
9	100	0.31	0.33				
10							
11	温度与压力A的相关系数						
12	温度与压力B的相关系数						
13							
14							
15	130	0.71	0.76				
16	135	0.82	0.9				
17	140	0.95	1.04				
18	145	1.12	1.24				
19	150	1.25	1.46				
20	155	1.41	1.7				
21							

从数据统计结论可以看出，温度与压力 A、B 的相关性分别达到了 0.95 和 0.94，这说明它们呈现良好的正相关性，而两组压力数据间的相关性达到了 0.998，这说明在不同反应器内的相同条件下反应一致性很好，可以忽略因为更换反应器造成的系统误差。

协方差的统计与相关系数的活的方法相似，统计结果同样返回一个输出表和一个矩阵，分别表示每对测量值变量之间的相关系数和协方差。不同之处在于相关系数的取值在 -1 和 +1 之间，而协方差没有限定的取值范围。相关系数和协方差都是描述两个变量离散程度的指标。

## 用 Excel 做数据排序的常用方法与技巧

在用 Excel 制作相关的数据表格时，我们可以利用其强大的排序功能，浏览、查询、统计相关的数字。下面，我们以图 1 所示的“员工基本情况登记表”为例，来全面体验一番 Excel 的排序功能。

	A	B	C	D	E	F	G	H	I	J	K	L
1	序号	姓名	部门	性别	民族	籍贯	出生年月	学历	工龄	职称	工资	备注
2	6	毕志坚	助理室	男	汉	安徽和县	1942年3月	大专	46	经济师	1520	
3	68	王辉荣	二队	男	汉	马鞍山	1944年11月	大专	43	经济师	1520	
4	53	周亦	财务科	男	汉	安徽南陵	1946年9月	大专	40	经济师	1520	
5	19	钟晓然	运输科	男	汉	安徽庐江	1946年9月	初中	39	助理会计师	1160	
6	27	胡楚君	备件科	男	汉	安徽无为	1945年10月	中专	39	助理经济师	1200	
7	61	周长州	加油站	男	汉	四川重庆	1944年6月	大学	39	高级工程师	1680	
8	62	孙景芸	一队	男	汉	安徽滁州	1946年1月	中专	39	政工师	1400	
9	78	董乐然	三队	男	汉	安徽枞阳	1947年11月	中专	39	助理经济师	1160	
10	94	史李敏	六队	男	回	辽宁本溪	1943年2月	大学	38	高级工	1400	
11	58	牛霞强	加油站	男	汉	安徽宣城	1948年1月	大学	37	助理工	1160	
12	90	陈善工	五队	男	汉	安徽肥东	1943年3月	大学	37	高级工	1400	

### 一、快速排序

如果我们对员工资料按某列属性(如“工龄”由长到短)进行排列,可以这样操作:选中“工龄”列任意一个单元格(如 I3),然后按一下“常用”工具栏上的“降序排序”按钮即可(参见图 1)。

小提示:①如果按“常用”工具栏上的“升序排序”按钮,则将“工龄”由短到长进行排序。②如果排序的对象是中文字符,则按“汉语拼音”顺序排序。③如果排序的对象是西文字符,则按“西文字母”顺序排序。

## 二、多条件排序

如果我们需要按“学历、工龄、职称”对数据进行排序,可以这样操作:选中数据表格中任意一个单元格,执行“数据→排序”命令,打开“排序”对话框(图 2),将“主要关键字、次要关键字、第三关键字”分别设置为“学历、工龄、职称”,并设置好排序方式(“升序”或“降序”),再按下“确定”按钮就行了。



## 三、按笔划排序

对“姓名”进行排序时,国人喜欢按“姓氏笔划”来进行:选中姓名列任意一个单元格,执行“数据→排序”命令,打开“排序”对话框(参见图 2),单击其中的“选项”按钮,打开“排序选项”对话框(图 3),选中其中的“笔划排序”选项,确定返回到“排序”对话框,再按下“确定”按钮即可。



小提示:如果需要按某行属性对数据进行排序,我们只要在上述“排序选项”对话框中选中“按行排序”选项即可。

#### 四、自定义排序

当我们对“职称”列进行排序时,无论是按“拼音”还是“笔划”,都不符合我们的要求。对于这个问题,我们可以通过自定义序列来进行排序:

先把相应的职称序列按需要排序的顺序输入到相应的单元格区域(如 N2 至 N18)中(图 4);执行“工具→选项”命令,打开“选项”对话框(图 5),切换到“自定义序列”标签下,在“从单元格中导入序列”右侧的方框中输入“\$N\$2:\$N\$18”(也可以用鼠标选择输入),然后单击“导入”按钮,将相应的序列导入到系统中,确定返回。

N
高级工程师
高级经济师
高级会计师
高级政工师
工程师
经济师
会计师
政工师
助理工程师
助理经济师
助理会计师
助理政工师
技术员
经济员
会计员
天极 其 yesky.com

小提示:序列导入后,原来 N2 至 N18 区域中输入的数据可以删除,导入的序列在其他 Excel 文档中均可直接使用。

选中“职称”列任意一个单元格,执行“数据→排序”命令,打开“排序”对话框,单击其中的“选项”按钮,打开“排序选项”对话框(参见图 3),按“自定义排序次序”选项右侧的下拉按钮,在随后弹出的下拉列表中,选中上述“导入”的序列,确定返回到“排序”对话框,再按下“确定”按钮即可。

## 五、用函数进行排序

有时,我们对某些数值列(如“工龄、工资”等)进行排序时,不希望打乱表格原有数据的顺序,而只需要得到一个排列名次。对于这个问题,我们可以用函数来实现(以“工龄”为例):在“工龄”右侧插入一个空白列(J 列),用于保存次序(图 6),然后选中 J2 单元格,输入公式:=RANK(I2,\$I\$2:\$I\$101),然后再次选中 J2 单元格,将鼠标移至该单元格右下角成“细十字线状”时(这种状态,我们通常称之为“填充柄”状态),按住左键向下拖拉至最后一条数据为止,次序即刻显示出来(图 6)。



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	序号	姓名	部门	性别	民族	籍贯	出生年月	学历	工龄	排序	职称	工资	备注
2	1	何岸洲	经理室	男	汉	河南永城	1953年5月	大学	36	12	高级工程师	1800	
3	2	刘继舟	经理室	男	汉	安徽当涂	1947年10月	大专	36	12	政工师	1400	
4	3	潘长川	工会	男	汉	安徽阜阳	1954年2月	大学	34	26	高级政工师	1680	
5	4	孙景芸	经理室	男	汉	安徽当涂	1963年12月	大专	26	56	工程师	1580	
6	5	梅诚雨	经理室	男	汉	皖安庆	1962年11月	大专	18	82	工程师	1580	
7	6	陆华忠	助理室	男	汉	安徽和县	1942年3月	大专	46	1	经济师	1520	
8	7	高尔艺	办公室	男	汉	安徽和县	1956年10月	大专	30	45	助理工	1300	
9	8	陈继伊	办公室	男	汉	浙江建德	1956年2月	大专	26	56			
10	9	陆金雅	办公室	男	汉	安徽无为	1954年6月	初中	34	26			
11	10	王耀芳	办公室	男	汉	安徽霍邱	1965年12月	大专	17	85			
12	11	潘永明	办公室	男	汉	安徽当涂	1966年10月	中专	17	88			

本文图片来自网络

小提示:若要升序排序,可在公式最后增加一个“非零”参数,如将上述公式改为:=RANK(I2,\$I\$2:\$I\$101,1)。

## 六、让序号不参与排序

当我们对数据表进行排序操作后,通常位于第一列的序号也被打乱了,如何不让这个“序号”列参与排序呢?我们在“序号”列右侧插入一个空白列(B列),将“序号”列与数据表隔开。用上述方法对右侧的数据区域进行排序时,“序号”列就不参与排序了。

小提示:插入的空列会影响表格的打印效果,我们可以将其隐藏起来:选中B列(即插入的空列),右击鼠标,再选择“隐藏”选项即可。

## Excel 的数据分析—排位与百分比

某班级期中考试进行后,按照要求仅公布成绩,但学生及家长要求知道排名。故欲公布成绩排名,学生可以通过成绩查询到自己的排名,并同时得到该成绩位于班级百分比排名(即该同学是排名位于前“X%”的学生)。

排序操作是 Excel 的基本操作,Excel“数据分析”中的“排位与百分比排位”可以使这个工作简化,直接输出报表。

注:本功能需要使用 Excel 扩展功能,如果您的 Excel 尚未安装数据分析,请依次选择“工具”-“加载宏”,在安装光盘中加载“分析数据库”。加载成功后,可以在“工具”下拉菜单中看到“数据分析”选项。

## 操作步骤

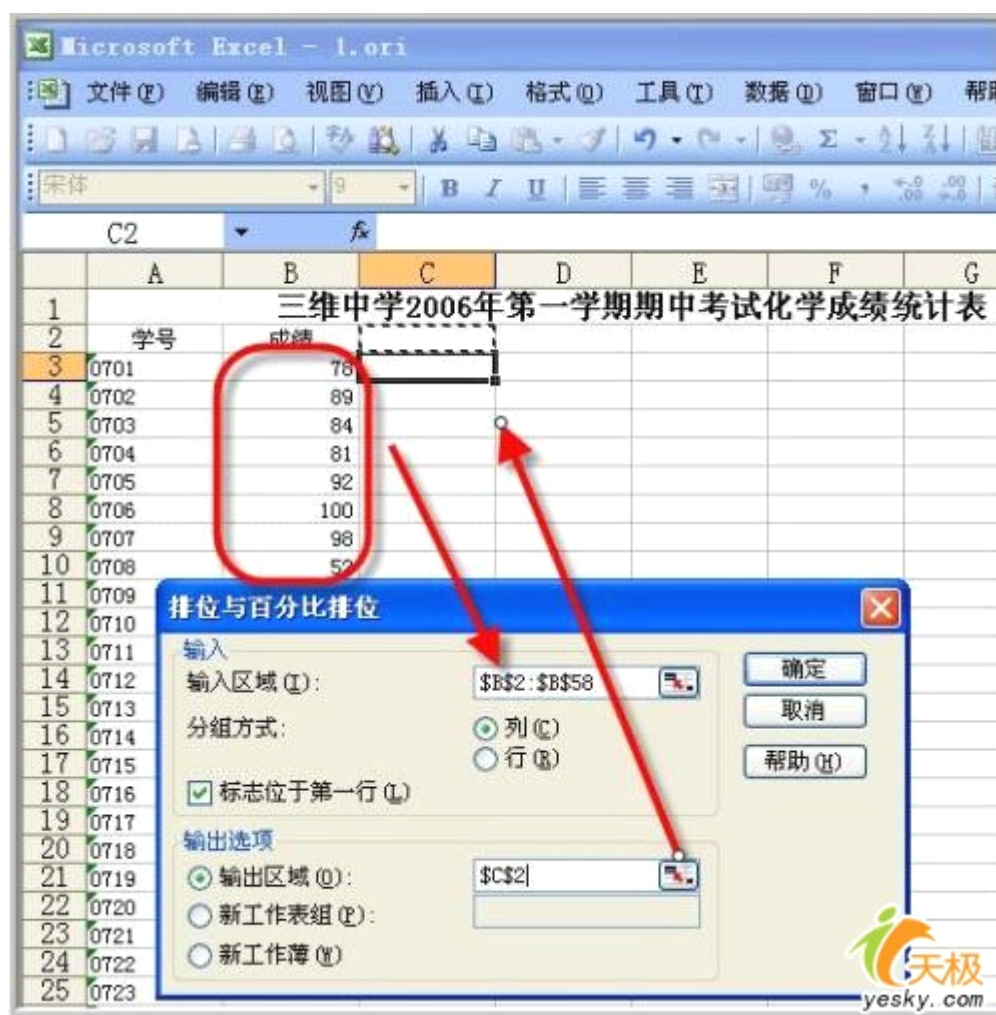
1. 打开原始数据表格，制作本实例的原始数据无特殊要求，只要满足行或列中为同一属性数值即可。

2. 选择“工具”-“数据分析”-“描述统计”后，出现属性设置框，依次选择；

输入区域:选择数据区域，如果有数据标志，注意同时勾选下方“标志位于第一行”；

分组方式:指示输入区域中的数据是按行还是按列考虑，请根据原数据格式选择；

输出区域可以选择本表、新工作表组或是新工作簿。



3. 点击“确定”即可看到生成的报表。

可以看到，此时生成一个四列的新表格，其中“点”是指排序后原数据的序数，在本实例中对应与学号，这也是很实用的一个序列；“成绩”即为排序后的数据系列；“排位”采取重复数据占用同一位置的统计方法；“百分比”是按照降序排列的，为了得到真正的“百分比排位”，还需要稍微作一下调整。

Microsoft Excel - 1.ori

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D) 窗口(W) 帮助(H)

宋体 9 B I U

H21

	A	B	C	D	E	F	G
1	三维中学2006年第一学期期中考试化学成绩统计表						
2	学号	成绩	点	成绩	排位	百分比	
3	0701	78	6	100	1	98.10%	
4	0702	89	30	100	1	98.10%	
5	0703	84	21	99	3	94.50%	
6	0704	81	49	99	3	94.50%	
7	0705	92	7	98	5	81.80%	
8	0706	100	12	98	5	81.80%	
9	0707	98	22	98	5	81.80%	
10	0708	52	29	98	5	81.80%	
11	0709	56	35	98	5	81.80%	
12	0710	86	45	98	5	81.80%	
13	0711	68	48	98	5	81.80%	
14	0712	98	46	95	12	80.00%	
15	0713	60	5	92	13	76.30%	
16	0714	70	36	92	13	76.30%	
17	0715	74	37	91	15	74.50%	
18	0716	89	2	89	16	67.20%	
19	0717	53	16	89	16	67.20%	
20	0718	15	20	89	16	67.20%	
21	0719	45	34	89	16	67.20%	
22	0720	89	33	88	20	65.40%	
23	0721	88	30	87	21	65.40%	

天极 aryesky.com

4.在“百分比”列的下一列输入“百分排名”，在第一个单元格中输入公式“=1-G3(对应于‘百分排名’)”，回车。选中该单元格，向下拖动直至填充完毕。这样就达到了显示百分比排名的目的。

完成的报表实例如下图所示。

H3								
	A	B	D	E	F	G	H	
1	三维中学2006年第一学期期中考试化学成绩统计表							
2	学号	成绩	点	成绩	排位	百分比	百分排名	
3	0701	78	6	100	1	98.10%	1.90%	
4	0702	89	30	100	1	98.10%	1.90%	
5	0703	84	21	99	3	94.50%	5.50%	
6	0704	81	49	99	3	94.50%	5.50%	
7	0705	92	7	98	5	81.80%	18.20%	
8	0706	100	12	98	5	81.80%	18.20%	
9	0707	98	22	98	5	81.80%	18.20%	
10	0708	52	29	98	5	81.80%	18.20%	
11	0709	56	35	98	5	81.80%	18.20%	
12	0710	86	45	98	5	81.80%	18.20%	
13	0711	68	48	98	5	81.80%	18.20%	
14	0712	98	46	95	12	80.00%	20.00%	
15	0713	60	5	92	13	76.30%	23.70%	
16	0714	70	36	92	13	76.30%	23.70%	
17	0715	74	37	91	15	74.50%	25.50%	
18	0716	89	2	89	16	67.20%	32.20%	
19	0717	53	16	89	16	67.20%	32.20%	

## 用 Excel 做数据分析—描述统计

某班级期中考试进行后，需要统计成绩的平均值、区间，以及给出班级内部学生成绩差异的量化标准，借此来作为解决班与班之间学生成绩的参差不齐的依据。要求得到标准差等统计数值。

样本数据分布区间、标准差等都是描述样本数据范围及波动大小的统计量，统计标准差需要得到样本均值，计算较为繁琐。这些都是描述样本数据的常用变量，使用 Excel 数据分析中的“描述统计”即可一次完成。

**注：**本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘加载“分析数据库”。加载成功后，可以在“工具”下拉菜单中看到“数据分析”选项。

### 操作步骤

1. 打开原始数据表格，制作本实例的原始数据无特殊要求，只要满足行或列中为同一属性数值即可。



Microsoft Excel - 1.1.xls

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D)

宋体 9 B I U

H15 fx

	A	B	C	D	E
1	三维中学2006年第一学期期中考试化学成绩统计表				
2	学号	成绩			
3	0701	78			
4	0702	89			
5	0703	84			
6	0704	81			
7	0705	92			
8	0706	100			
9	0707	98			
10	0708	52			
11	0709	56			
12	0710	86			
13	0711	68			
14	0712	98			
15	0713	60			
16	0714	70			
17	0715	74			
18	0716	89			
19	0717	53			

单列或多列  
均可

天极  
yesky.com

2. 选择“工具”-“数据分析”-“描述统计”后，出现属性设置框，依次选择：

输入区域：原始数据区域，可以选中多个行或列，注意选择相应的分组方式；

如果数据有标志，注意勾选“标志位于第一行”；如果输入区域没有标志项，该复选框将被清除，Excel 将在输出表中生成适宜的数据标志；

输出区域可以选择本表、新工作表或是新工作簿；

汇总统计：包括有平均值、标准误差(相对于平均值)、中值、众数、标准偏差、方差、峰值、偏斜度、极差、最小值、最大值、总和、总个数、最大值、最小值和置信度等相关项目。

其中：

中值：排序后位于中间的数据的值；

众数：出现次数最多的值；



峰值：衡量数据分布起伏变化的指标，以正态分布为基准，比其平缓时值为正，反之则为负；

偏斜度：衡量数据峰值偏移的指数，根据峰值在均值左侧或者右侧分别为正值或负值；

极差：最大值与最小值的差。

第 K 大(小)值：输出表的某一行中包含每个数据区域中的第 k 个最大(小)值。

平均数置信度：数值 95% 可用来计算在显著性水平为 5% 时的平均值置信度。



结果示例如下(本实例演示了双列数据的描述统计结果):

成绩		学习时间	
平均	78.64285714	平均	62.91428571
标准误差	2.408241878	标准误差	1.926593502
中位数	85	中位数	68
众数	98	众数	78.4
标准差	18.02163202	标准差	14.41730562
方差	324.7792208	方差	207.8587013

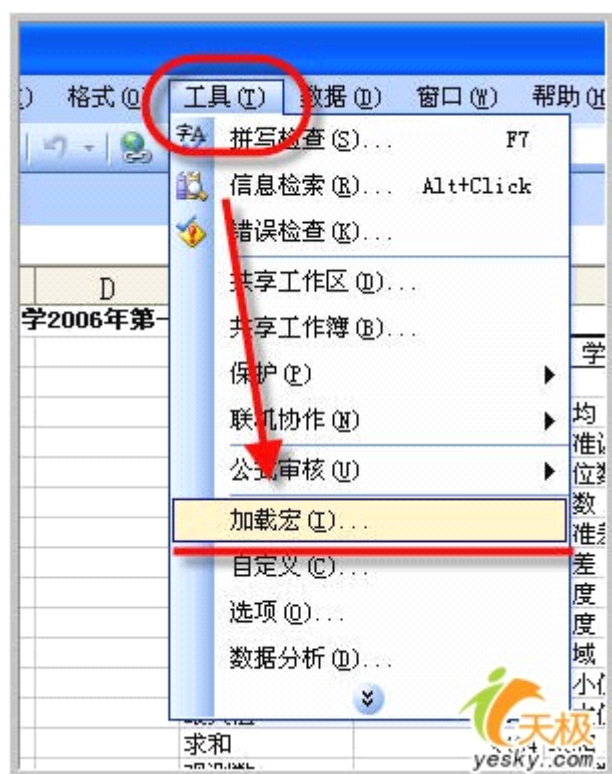
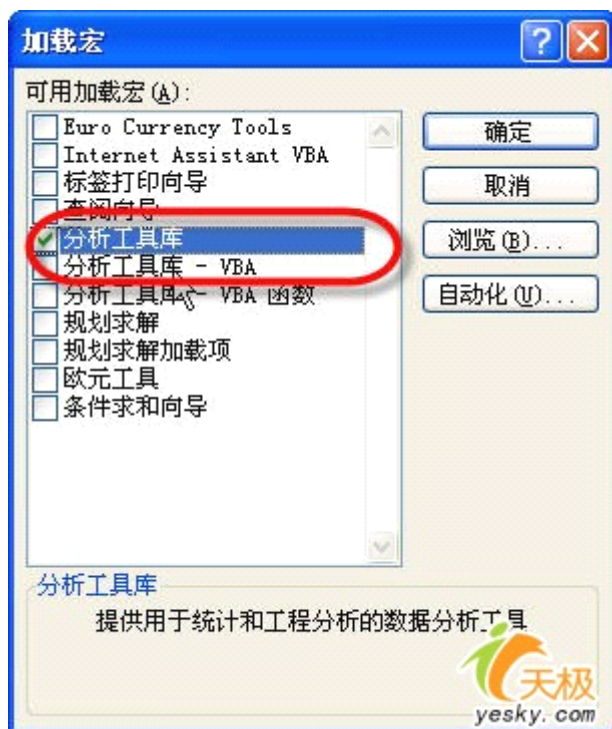
峰度	1.464424408	峰度	1.464424408
偏度	-1.130551511	偏度	-1.13055151
区域	85	区域	68
最小值	15	最小值	12
最大值	100	最大值	80
求和	4404	求和	3523.2
观测数	56	观测数	56
最大 (1)	100	最大 (1)	80
最小 (1)	15	最小 (1)	12
置信度 (95.0%)	4.826224539	置信度 (95.0%)	3.860979631

## 用 Excel 做数据分析—直方图

使用 Excel 自带的数据分析功能可以完成很多专业软件才有的数据统计、分析，这其中包括：直方图、相关系数、协方差、各种概率分布、抽样与动态模拟、总体均值判断，均值推断、线性、非线性回归、多元回归分析、时间序列等内容。下面将对以上功能逐一作使用介绍，方便各位普通读者和相关专业人员参考使用。

**注：**本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘中加载“分析数

据库”。加载成功后，可以在“工具”下拉菜单中看到“数据分析”选项。



### 实例 1

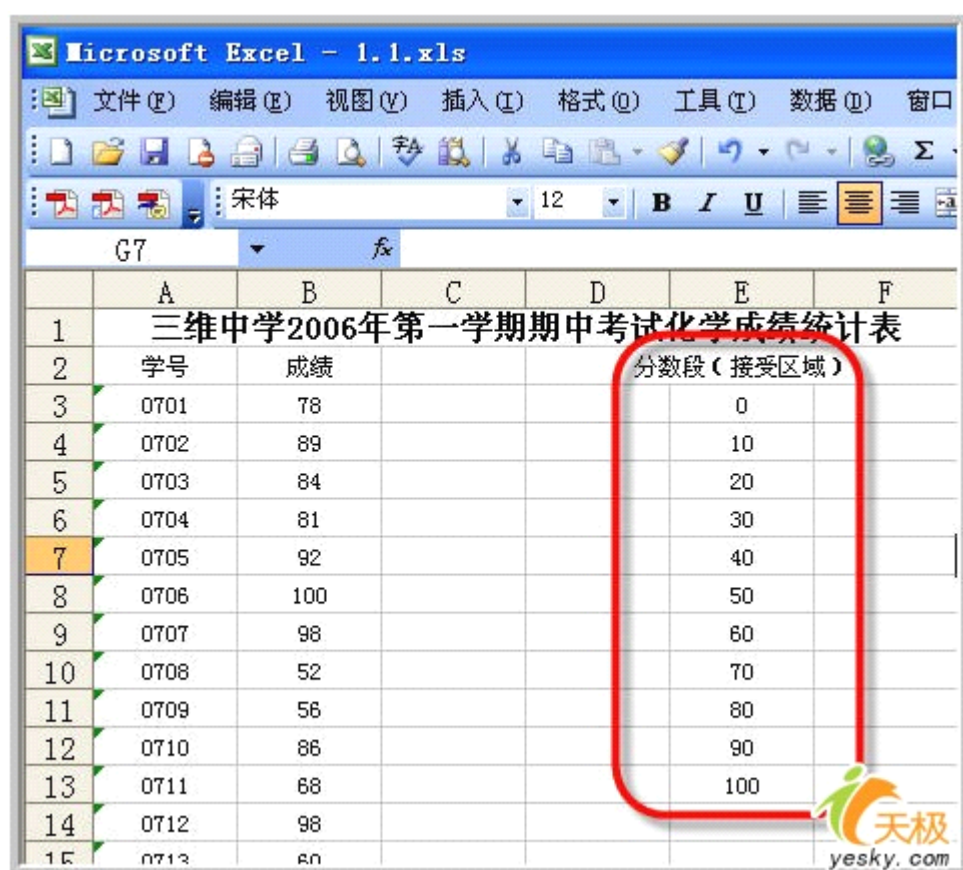
某班级期中考试进行后，需要统计各分数段人数，并给出频数分布和累计频数表的直方图以供分析。

以往手工分析的步骤是先将各分数段的人数分别统计出来制成一张新的表格，再以此表格为基础建立数据统计直方图。使用 Excel 中的“数据分析”功能可以直接完成此任务。

### 操作步骤

1.打开原始数据表格，制作本实例的原始数据要求单列，确认数据的范围。本实例为化学成绩，故数据范围确定为 0-100。

2.在右侧输入数据接受序列。所谓“数据接受序列”，就是分段统计的数据间隔，该区域包含一组可选的用来定义接收区域的边界值。这些值应当按升序排列。在本实例中，就是以多少分数段作为统计的单元。可采用拖动的方法生成，也可以按照需要自行设置。本实例采用 10 分一个分数统计单元。



	A	B	C	D	E	F
1	三维中学2006年第一学期期中考试化学成绩统计表					
2	学号	成绩			分数段 (接受区域)	
3	0701	78			0	
4	0702	89			10	
5	0703	84			20	
6	0704	81			30	
7	0705	92			40	
8	0706	100			50	
9	0707	98			60	
10	0708	52			70	
11	0709	56			80	
12	0710	86			90	
13	0711	68			100	
14	0712	98				
15	0713	60				

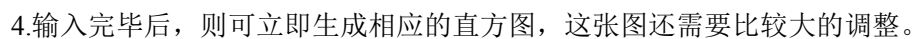
3.选择“工具”-“数据分析”-“直方图”后，出现属性设置框，依次选择：

输入区域：原始数据区域；

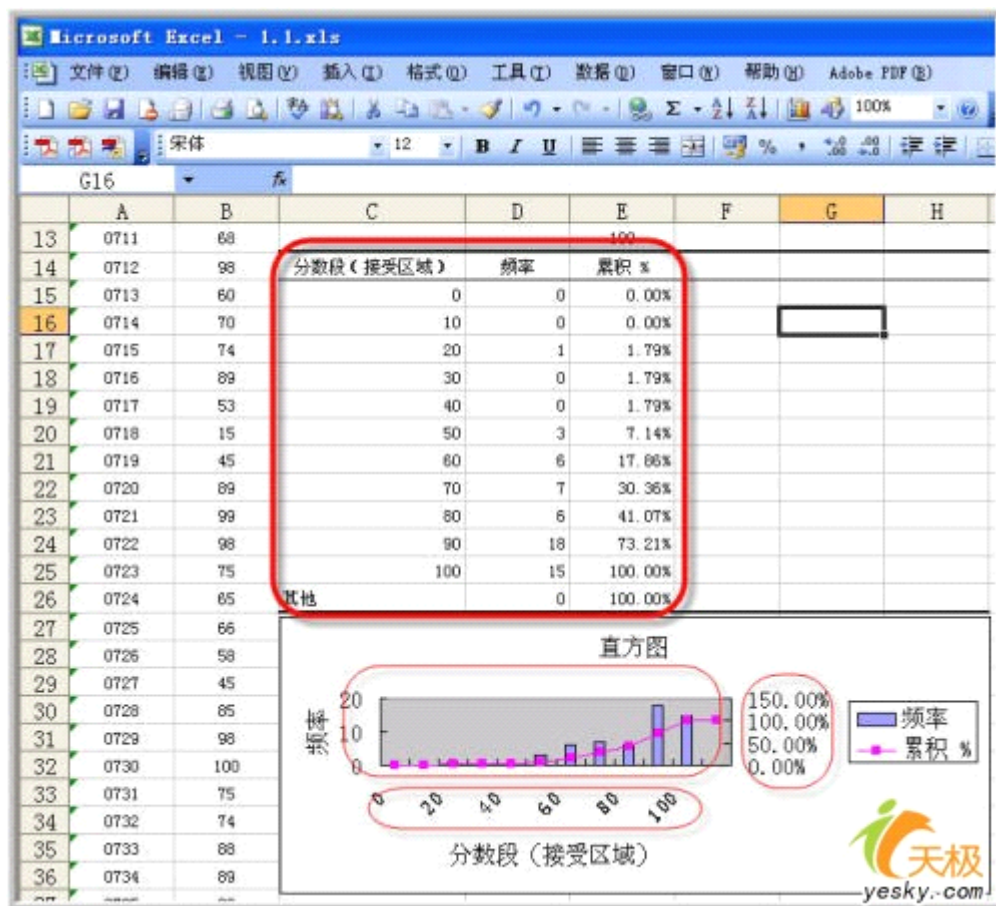
接受区域：数据接受序列；

如果选择“输出区域”，则新对象直接插入当前表格中；

若选择“累计百分率”，则会在直方图上叠加累计频率曲线;





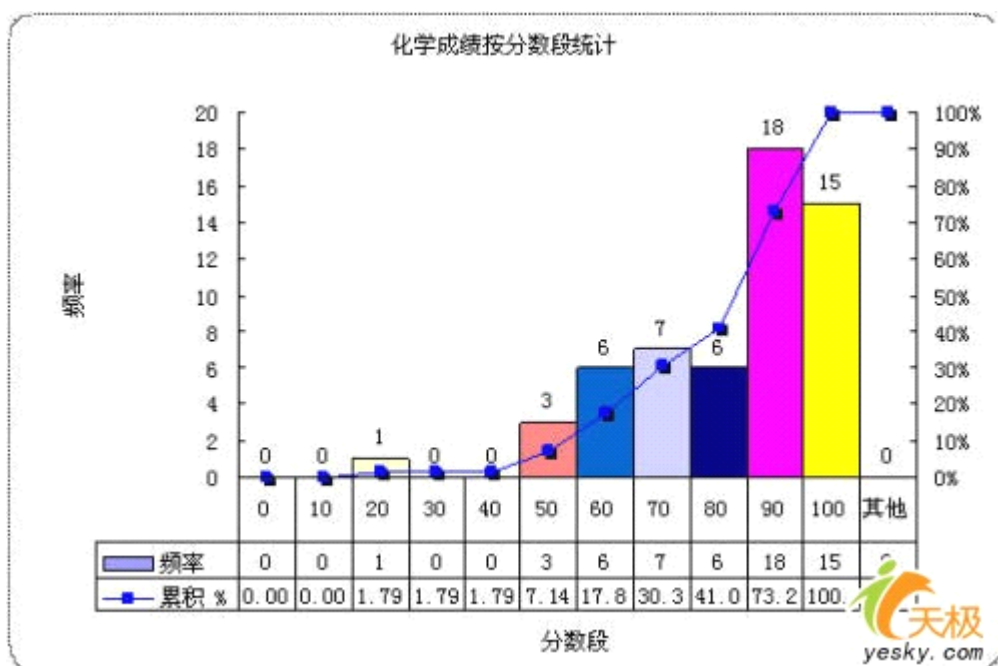


主要是：

纵横坐标的标题、柱型图的间隔以及各种数据的字体、字号等等。

为了达到柱型图之间无缝的紧密排列，需要将“数据系列格式”中的“选项”中“分类间距”调整为“0”。其余细节，请双击要调整的对象按照常规方法进行调整，这里不再赘述。

调整后的直方图参考如下



## excel 数据分析

### 公式

公式是 Excel 的核心功能之一，它使工作表具备“计算”能力，用户只需输入原始数据，进一步的计算用公式来实现，准确、快速又方便。公式由等号(=)、运算符(加(+)、减(-)、乘(\*)、除(/)、乘方(^)、括号(())、数等要素组成。

- ① 等号：这是公式的标志。若想对数值进行计算，必须在数值运算公式前加等号。
- ② 操作符：表示执行哪种运算。
- ③ 数：引用的单元格、函数或常数。

【案例三】用 EXCEL 公式计算《统计学原理》教材第 22 页表 2—4 中全地区从业总人数、全地区工业增加值总额、全地区上交税金总额（见统计学原理.XLS 的‘第二章第三节 统计表及其设计’工作表）。

#### 计算步骤：

①在 A10 单元格输入‘表 2—4’，在 B10 单元格输入‘某年某地区甲乙两地工业企业从业人数、工业增加值和税金’，将 B10 单元格到 D10 单元格合并，并居中和自动换行；

②在 A11 单元格输入‘按地区分组’，在 B11 单元格输入‘从业人数（万人）’，在 C11 单元格输入‘工业增加值（万元）’，在 D11 单元格输入‘上交税金（万元）’；

③在 A12 单元格输入‘（甲）’，在 B12 单元格输入‘（1）’，在 C12 单元格输入‘（2）’，在 D12 单元格输入‘（3）’；

④在 A13 单元格输入‘甲地’，在 B13 单元格输入‘11.47’，在 C13 单元格输入‘216622.42’，在 D13 单元格输入‘643.20’；在 A14 单元格输入‘乙地’，在 B14 单元格输入‘35.99’，在 C14 单元格输入‘368573.59’，在 D14 单元格输入‘9141.46’；

⑤在 A15 单元格输入‘全地区’，在 B15 单元格输入‘=B13+B14’，然后敲回车键，则 B15 单元格显示的结果为 47.46；

⑥然后用鼠标左键点击一次 B15 单元格，再次将鼠标光标移动到该单元格的右下角，这时鼠标变成十字星，利用该填充功能将 C15 和 D15 单元格分别填充为‘=C13+C14’和‘=D13+D14’；

⑦将 B13 到 D15 单元格格式全部选为数值，小数位保留 2 位，然后敲回车键，则在 B15 单元格显示的结果为 47.46，C15 单元格显示的结果为 585196.01，D15 单元格显示的结果为 9784.66；

⑧给各单元格绘制边框（见统计学原理.XLS 的‘第三节 统计表及其设计’工作表）。计算过程完成。

【案例四】用 EXCEL 公式计算《统计学原理》教材第 22 页表 2—5 中各地区和全地区每人平均工业增加值、每人平均上交税金（见统计学原理.XLS 的‘第二章第三节 统计表及其设计’工作表）。

计算步骤：

①将 A10 到 D15 单元格全部选中，然后点击复制，再用鼠标点击 A18 单元格，点击粘贴；将输入‘某年某地区甲乙两地工业企业从业人数、工业增加值和税金’的单元格内容改写为‘某年某地区甲乙两地工业企业人均增加值和税金’，再将 B18 单元格到 D18 单元格的合并撤消，在 C18 单元格内输入‘单位：元’；

②将 B19 到 B23 单元格删除，并将右边单元格左移；

③在新的 B20 单元格内输入‘（1）’，在 B21 单元格内输入‘=C13/B13’，然后敲回车键，再利用填充功能将 B21 单元格内的公式填充到 B22 和 B23 单元格中；

④在 C20 单元格内输入‘（2）’，在 C21 单元格内输入‘=D13/B13’，然后敲回车键，再利用填充功能将 C21 单元格内的公式填充到 C22 和 C23 单元格中；

⑤设定各单元格数值格式和绘制边框（见统计学原理.XLS 的‘第三节 统计表及其设计’工作表）。计算过程完成。

【案例五】用 EXCEL 公式计算《统计学原理》教材第 40 页表 3—4 中销售收入各组的企业比重、销售收入的上限、下限、组中值、组距（见统计学原理.XLS 的‘第三章第三节 次数分布’工作表）。

计算步骤：

①用鼠标右键点击统计学原理.XLS‘第二章第三节统计表及其设计’工作表标签，将它复制为统计学原理.XLS‘第二章第三节统计表及其设计（2）’工作表，再将该工作表标签改写成‘第三章第三节次数分布’工作表。

②将 A1 单元格改写成‘表 3—4’，将 B1 单元格改写成‘某地区商业企业销售收入统计表’；将 A2 单元格改写成‘销售收入（万元）’，将 B2 单元格改写成‘企业数（个）’；将第 3 行删除；在 A3 单元格内输入 80~90，在 A4 单元格内输入 90~100。在第 5 行处插入一行，然后，在 A5 单元格内输入 100~110，在 A6 单元格内输入 110~120。

③在 B3 到 B7 单元格内分别输入 2、4、16、6、=B3+B4+B5+B6，然后敲回车键，再利用填充功能将 B7 单元格内的公式填充到 C7 单元格中。再将 C4 单元格内的公式填充到 C5 单元格中。这时，销售收入各组的企业比重数就计算出来了。

④在 D2 单元格内输入‘下限’，在 E2 单元格内输入‘上限’，在 F2 单元格内输入‘组中值’，在 G2 单元格内输入‘组距’，然后在 F3 单元格内输入‘=(D3+E3)/2’，在 G3 单元格内输入‘=E3-D3’，敲回车键，再利用填充功能将 F3 和 G3 单元格内的公式填充到 F4 到 G6 单元格内。

⑤在 D3 到 E4 单元格内依次输入 80、90、90、100，然后将 D3 到 E4 单元格选中，再利用填充功能将 D3 到 E4 内有规律的数字遵循其规律填充到 D5 到 E6 单元格内。这时，销售收入的上限、下限、组中值、组距就计算出来了。

⑥设定各单元格数值格式和绘制边框（见统计学原理.XLS 的‘第三章第三节次数分布’工作表）。计算过程完成。

五、绘制曲线图

### (一)绘制直方图

【案例六】用 EXCEL 图表功能绘制《统计学原理》教材第 43 页表 3—4 中销售收入各组的直方图（见统计学原理.XLS 的‘第三章第三节次数分布’工作表）。

绘制步骤：

①在 A1 到 B13 单元格内创建表 5—1（1），在 A2 到 B12 单元格内输入相关数据。

②点击命令条中的图表（H）图标，然后选择柱形图，按着操作步骤，就可以绘制出表 3—4 中销售收入各组的直方图（见统计学原理.XLS 的‘第三章第三节次数分布’工作表）。

## 第二章 EXCEL 函数

### 第一节 平均指标类函数

#### 一、算术平均数

【案例七】用 EXCEL 函数计算《统计学原理》教材第 64 页（例 5-1）工人平均日产量（见统计学原理.XLS 的‘第五章平均指标和变异指标’工作表）。

计算步骤：

①在 A1 到 B13 单元格内创建表 5—1（1），在 A2 到 B12 单元格内输入相关数据，在 A13 单元格内输入‘平均’。

②点击 B13 单元格，然后再点击命令条中的函数（F）图标，选择函数分类中的统计，在函数名中选‘AVERAGE’（见图 2—1—1），点击确定，出现图 2—1—2。

③点击图 2—1—2 确定按钮，计算完成。

#### 二、调和平均数

【案例八】用 EXCEL 函数计算《统计学原理》教材第 69 页（例 5-4）三批产品的平均价格（见统计学原理.XLS 的‘第五章平均指标和变异指标’工作表）。

计算步骤：

①在 A16 到 DB21 单元格内创建表 5—3，在 A22 单元格内输入‘平均’。

②点击 B22 单元格，然后再点击命令条中的函数（F）图标，选择函数分类中的数学与三角函数，在函数名中选‘SUM’，在 Number1 对话框内输入 C18: C20，点击确定。然后将 B22 单元格内的计算公式改写成‘=SUM（C18: C20）/SUM（D18: D20）’。计算过程结束。

### 第二节 变异指标类函数

#### 一、标准差

【案例九】用 EXCEL 函数计算《统计学原理》教材第 82 页（例 5-13）工人日产量零件数的标准差（见统计学原理.XLS 的‘第五章平均指标和变异指标’工作表）。

计算步骤：

①在 A25 到 A38 单元格内创建表 5—15，在 A39 单元格内输入‘标准差’。

②点击 A40 单元格，然后再点击命令条中的函数（F）图标，选择函数分类中的统计函数，在函数名中选‘STDEVP’，在 Number1 对话框内输入 A29: A38，点击确定。计算过程结束。

也可以采用 EXCEL 数据分析工具计算。

计算步骤：

①在 A25 到 A38 单元格内创建表 5—15，在 B39 单元格内输入‘标准差’。

②点击 A41 单元格，然后再点击命令条中的工具（F）图标，选择数据分析（D）（见图 2—2—1），在数据分析的分析工具（A）对话框内将‘描述统计’中选，点击确定（见图 2—2—2）。

③在描述统计的输入区域对话框内输入\$A\$28:\$A\$38, 分组方式选择‘逐列’, 将标志位于第一行选中。在输出选项的输出区域对话框内输入\$A\$41, 再将汇总统计选中, 点击确定。

④在 B40 单元格内输入‘=B47\*SQRT((COUNT(A29:A38)-1)/COUNT(A29:A38))’, 敲回车键。计算过程结束 (见图 2—2—3)。

### 第三节 平均发展速度类函数

#### 一、乘幂函数

power (底,指数)

【案例十】用 EXCEL 乘幂函数计算《统计学原理》教材第 99 页 (例 6-6) 1985—2004 年间平均发展速度 (见统计学原理.XLS 的‘第六章动态数列’工作表)。

计算步骤:

①在 A1 到 B4 单元格内创建表 6—1 (1), 在 C2 单元格内输入‘平均发展速度’。

②在 C3 单元格, 输入‘=POWER(B4/B3,1/(A4-A3))’, 然后, 敲回车键。计算过程结束。

#### 二、几何平均函数

GEOMEAN (数值区域, 数值区域, ...)

【案例十一】用 EXCEL 几何平均数函数计算《统计学原理》教材第 100 页 (例 6-7) 1999—2004 年的平均发展速度 (见统计学原理.XLS 的‘第六章动态数列’工作表)。

计算步骤:

①在 A7 到 G11 单元格内创建表 6—7 (1), 在 A11 单元格内输入‘平均发展速度’。

②在 B11 单元格输入‘=GEOMEAN(C10:G10)’, 然后, 敲回车键。计算过程结束。

### 第四节 最小平方类函数

#### 一、截距函数

INTERCEPT (因变量数据单元格区域, 自变量数据单元格区域)

【案例十一】用 EXCEL 截距函数计算《统计学原理》教材第 107 页 (例 6-12) 1991—2000 年间某地区粮食产量的直线趋势方程常数项 a (见统计学原理.XLS 的‘第六章动态数列’工作表)。

计算步骤:

①在 A14 到 C25 单元格内创建表 6—14 (1), 在 A27 单元格内输入‘直线趋势方程常数项’。

②在 B27 单元格输入‘=INTERCEPT(C16:C25,B16:B25)’, 然后, 敲回车键。计算过程结束。

#### 二、斜率函数

SLOPE (因变量数据单元格区域, 自变量数据单元格区域)

【案例十二】用 EXCEL 截距函数计算《统计学原理》教材第 107 页 (例 6-12) 1991—2000 年间某地区粮食产量的直线趋势方程系数 b (见统计学原理.XLS 的‘第六章动态数列’工作表)。

计算步骤:

①在 A14 到 C25 单元格内创建表 6—14 (1), 在 A28 单元格内输入‘直线趋势方程系数’。

②在 B28 单元格输入‘=SLOPE(C16:C25,B16:B25)’, 然后, 敲回车键。计算过程结束。

## 第三章 EXCEL 相关与回归分析

### 第一节 EXCEL 相关分析

#### 一、绘制散点图



【案例十三】用 EXCEL 图表功能绘制《统计学原理》教材第 158 页散点图（见统计学原理.XLS 的‘第十章相关与回归分析’工作表）。

绘制步骤：

①在 A1 到 B12 单元格内创建表 9—1。

②点击插入命令条中的图表图标，然后选择散点图，点击下一步，按照图表向导—4 步骤就可以绘制出散点（见统计学原理.XLS 的‘第十章相关与回归分析’工作表）。

## 二、计算相关系数

【案例十四】用 EXCEL 数据分析功能计算《统计学原理》教材第 160 页（例 9-1）生产总值（亿元）与社会商品零售总额（亿元）的相关系数（见统计学原理.XLS 的‘见统计学原理.XLS 的‘第十章相关与回归分析’工作表）。

计算步骤：

①在 A19 到 C27 单元格内创建表 9—4。

②点击工具命令条中的数据分析，在数据分析对话框内选择相关系数，点击确定。在相关系数对话框内的输入区域对话框输入\$B\$20: \$C\$27，在分组方式中选择‘逐行’，将标志位于第一行选中，在输出区域对话框内输入\$A\$28，点击确定（见统计学原理.XLS 的‘第十章相关与回归分析’工作表），计算过程结束。

## 第二节 EXCEL 回归分析

### 一、建立一元线性回归方程

【案例十五】用 EXCEL 数据分析功能建立《统计学原理》教材第 164（例 9-2）一元线性回归方程（见统计学原理.XLS 的‘第五章平均指标和变异指标’工作表）。

建立一元线性回归方程步骤：

①在 A33 到 A41 单元格内创建表 9—7（1），在 A42 单元格内输入‘建立一元线性回归方程’。

②点击 A43 单元格，然后再点击工具命令条中的数据分析，在数据分析对话框内选择回归，点击确定。在回归对话框内的 Y 值输入区域对话框输入\$B\$34: \$B\$41，X 值输入区域对话框输入\$A\$34: \$A\$41，将标志选中，在输出选项对话框输出区域内输入\$A\$42（见图 3—2—1），点击确定（见图 3—2—2）。

③建立的财政收入 y 关于增加值 x 一元线性回归方程：

建立一元线性回归方程过程结束（见统计学原理.XLS 的‘第十章相关与回归分析’工作表）。

## 使用 Excel 做数据分析之回归分析方法

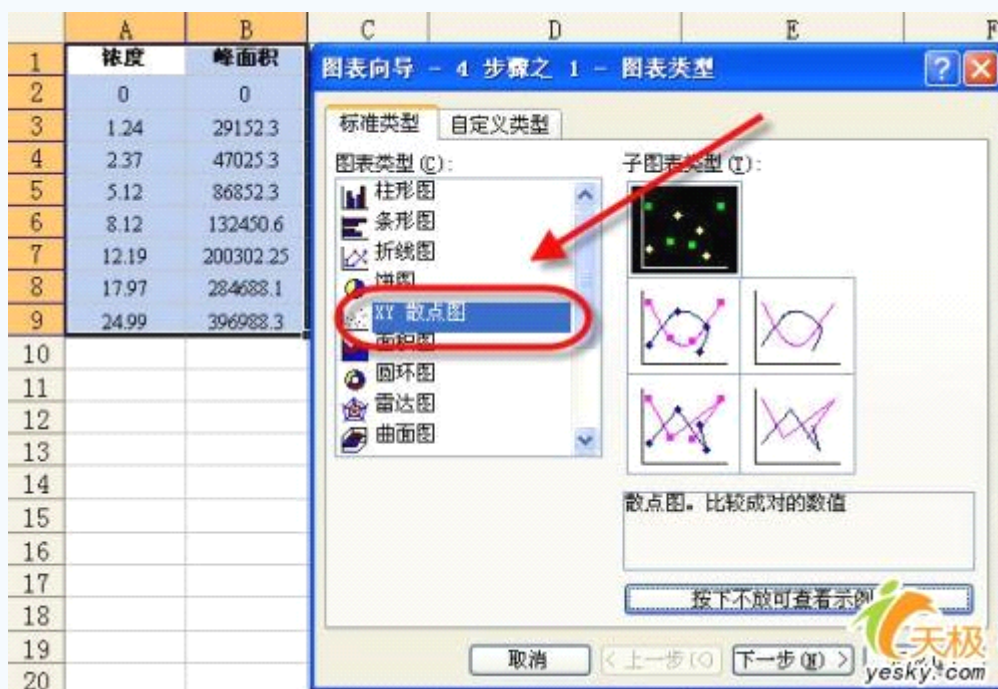
在数据分析中，对于成对成组数据的拟合是经常遇到的，涉及到的任务有线性描述，趋势预测和残差分析等等。很多专业读者遇见此类问题时往往寻求专业软件，比如在化工中经常用到的 Origin 和数学中常见的 MATLAB 等等。它们虽很专业，但其实使用 Excel 就完全够用了。我们已经知道在 Excel 自带的数据库中已有线性拟合工具，但是它还稍显单薄，今天我们来尝试使用较为专业的拟合工具来对此类数据进行处理。

注：本功能需要使用 Excel 扩展功能，如果您的 Excel 尚未安装数据分析，请依次选择“工具”-“加载宏”，在安装光盘支持下加载“分析数据库”。加载成功后，可以在“工具”下拉菜单中看到“数据分析”选项

**实例** 某溶液浓度正比对应于色谱仪器中的峰面积，现欲建立不同浓度下对应峰面积的标准曲线以供测试未知样品的实际浓度。已知 8 组对应数据，建立标准曲线，并且对此曲线进行评价，给出残差等分析数据。

这是一个很典型的线性拟合问题，手工计算就是采用最小二乘法求出拟合直线的待定参数，同时可以得出 R 的值，也就是相关系数的大小。在 Excel 中，可以采用先绘图再添加趋势线的方法完成前两步的要求。

选择成对的数据列，将它们使用“X、Y 散点图”制成散点图。



X、Y 散点图

在数据点上单击右键，选择“添加趋势线”-“线性”，并在选项标签中要求给出公式和相关系数等，可以得到拟合的直线。



给出公式和相关系数

由图中可知，拟合的直线是  $y=15620x+6606.1$ ， $R^2$  的值为 0.9994。

因为  $R^2 > 0.99$ ，所以这是一个线性特征非常明显的实验模型，即说明拟合直线能够以大于 99.99%地解释、涵盖了实测数据，具有很好的一般性，可以作为标准工作曲线用于其他未知浓度溶液的测量。

为了进一步使用更多的指标来描述这一个模型，我们使用数据分析中的“回归”工具来详细分析这组数据。

	A	B	C	D	E
1	浓度	峰面积		SUMMARY OUTPUT	
2	0	0			
3	1.24	29152.3		回归统计	
4	2.37	47025.3		Multiple R	0.999602689
5	5.12	86852.3		R Square	0.999205535
6	8.12	132450.6		Adjusted R Square	0.856348392
7	12.19	200302.25		标准误差	5901.98335
8	17.97	284688.1		观测值	8
9	24.99	396988.3			

回归

输入

Y 值输入区域 (Y):

X 值输入区域 (X):

☒ 标志 (L)
☒ 常数为零 (Z)

☐ 置信度 (C)
 %

输出选项

☒ 输出区域 (O):

☐ 新工作表组 (P):

☐ 新工作簿 (B):

残差

☒ 残差 (R)
☒ 残差图 (D)

☒ 标准残差 (I)
☒ 线性拟合图 (L)

正态分布

☒ 正态概率图 (N)

确定

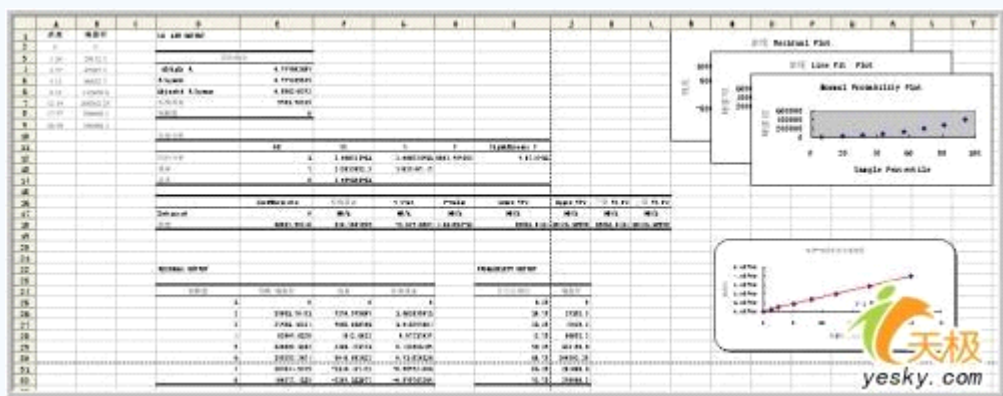
取消

帮助 (H)

## 回归工具

在选项卡中显然详细多了，注意选择 X、Y 对应的数据列。“常数为零”就是指明该模型是严格的正比例模型，本例确实是这样，因为在浓度为零时相应峰面积肯定为零。先前得出的回归方程虽然拟合程度相当高，但是在  $x=0$  时，仍然有对应的数值，这显然是一个可笑的结论。所以我们选择“常数为零”。

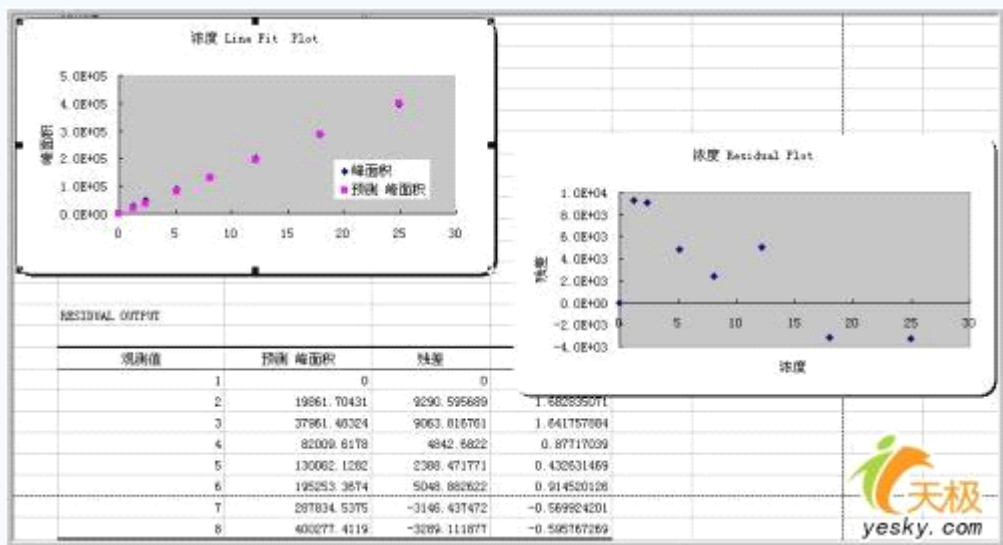
“回归”工具为我们提供了三张图，分别是残差图、线性拟合图和正态概率图。重点来看残差图和线性拟合图。



残差图、线性拟合图和正态概率图

在线性拟合图中可以看到，不但有根据要求生成的数据点，而且还有经过拟和处理的预测数据点，拟合直线的参数会在数据表格中详细显示。本实例旨在提供更多信息以起到抛砖引玉的作用，由于涉及到过多的专业术语，请各位读者根据实际，在具体使用中另行参考各项参数，此不再对更多细节作进一步解释。

残差图是有关于世纪之与预测值之间差距的图表，如果残差图中的散点在中州上下两侧零乱分布，那么拟合直线就是合理的，否则就需要重新处理。



图表

更多的信息在生成的表格中，详细的参数项目完全可以满足回归分析的各项要求。下图提供的是拟合直线的得回归分析中方差、标准差等各项信息。



D	E	F	
SUMMARY OUTPUT			
回归统计			
Multiple R	0.999602689		
R Square	0.999205535		
Adjusted R Square	0.856348392		
标准误差	5901.98335		
观测值	8		
方差分析			
	df	SS	
回归分析	1	3.06672E+11	
残差	7	24383385	
总计	8	3.069161	

## 用单元格数据作为 Excel 工作簿名称

1、启动 Excel，执行“工具→宏→Visual Basic 编辑器”命令，进入 VBA 编辑状态(如图 1)。

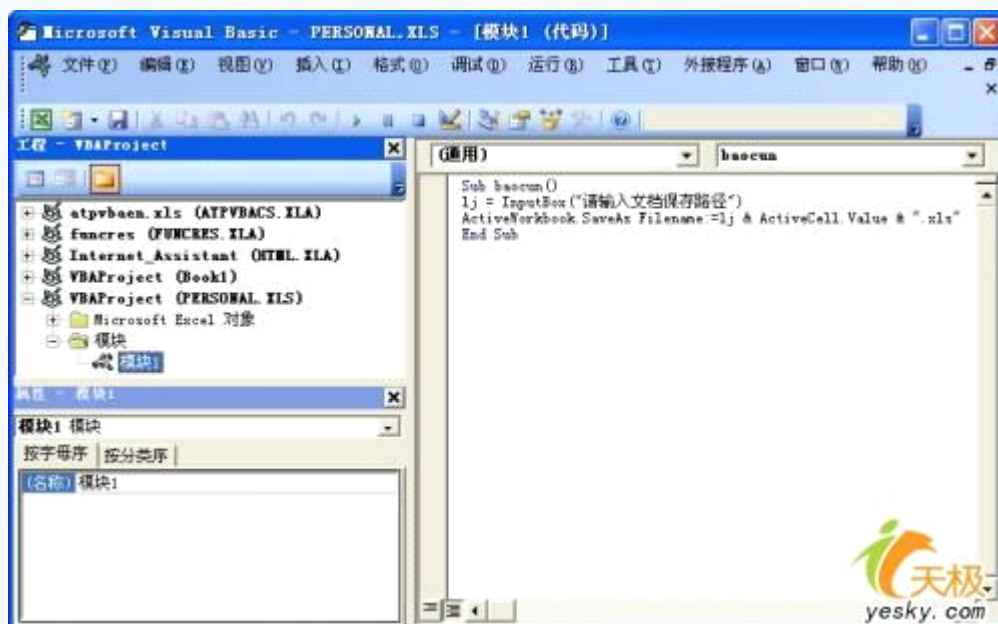


图 1 VBA 编辑

2、在左侧的“工程资源管理器”窗口中，选中“VBAproject (PERSONAL. XLS)” (个人宏工作簿)选项。

3、执行“插入→模块”命令，插入一个模块(模块 1)。

4、将下述代码输入到右侧的代码编辑窗口中：

```
Sub baocun()  
lj = InputBox("请输入文档保存路径")  
ActiveWorkbook.SaveAs Filename:=lj & ActiveCell.Value & ".xls"  
End Sub
```

退出 VBA 编辑状态。

5、以后要保存某个工作簿文档时，先选中作为名称的字符所在的单元格(参见图 2)，然后执行“工具→宏→宏”命令，打开“宏”对话框(如图 3)。

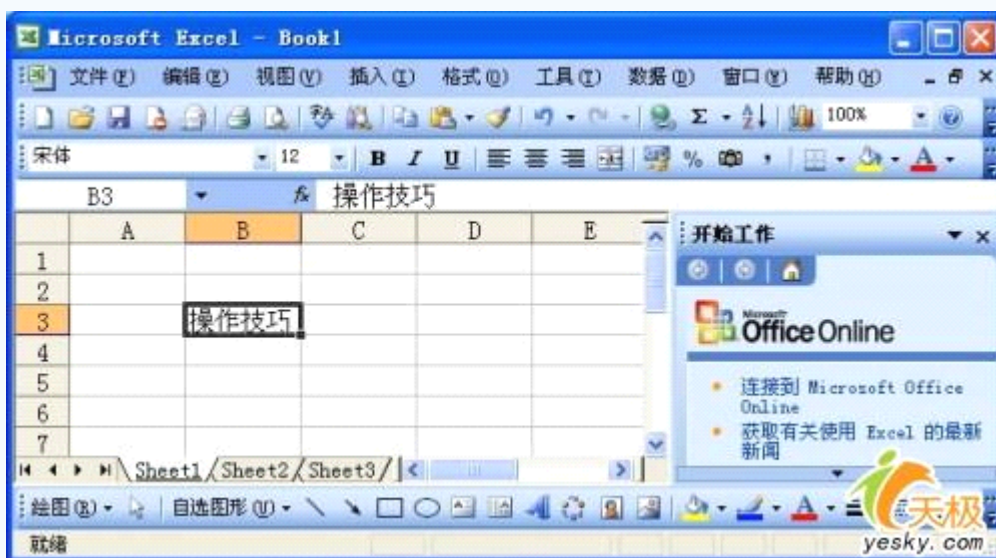


图 2 单元格

ER-BOTTOM-COLOR: #000000; BORDER-TOP-COLOR: #000000; BORDER-RIGHT-COLOR: #000000" alt= 教你用单元格数据作为 Excel 工作簿名称  
src="http://images.itdb.cn/News/2006/12/07/82118002352.jpg" border=1>

图 3 宏对话框

6、选中刚才编辑的宏 (PERSONAL.XLS!baocun)，单击“执行”按钮，系统弹出如图 4 所示的对话框。

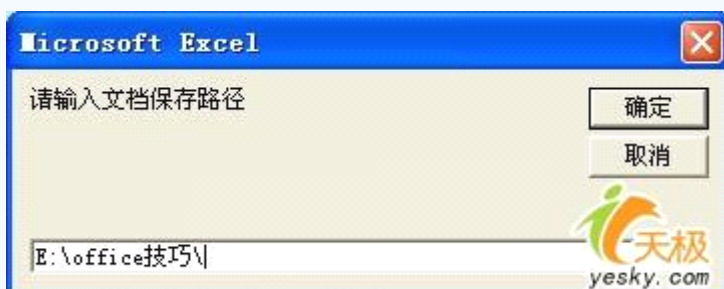


图 4 对话框

7、输入保存文档的路径(如“E:\office 技巧\”), 单击“确定”按钮。

文档保存成功(参见图 5)。

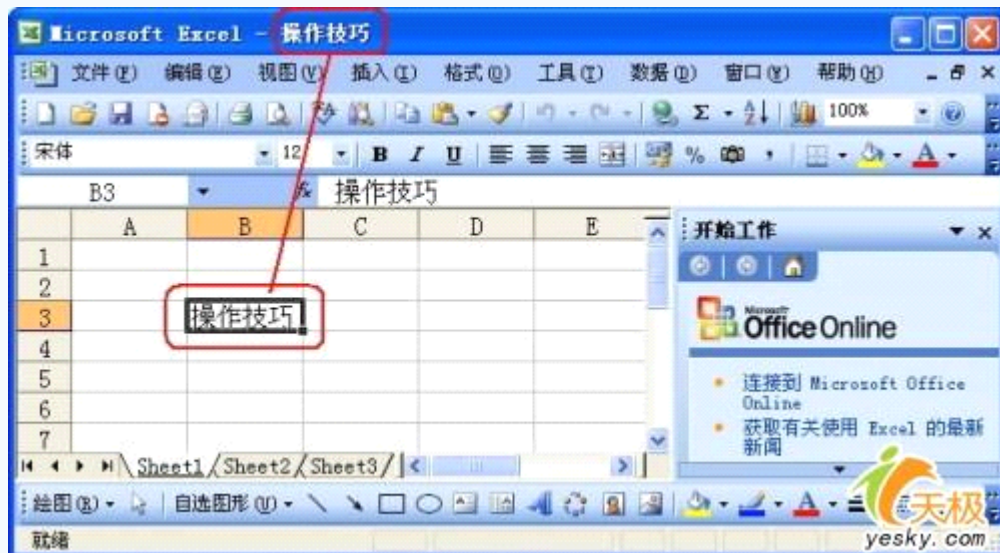


图 5 保存成功

注意：如果不需要保存路径，文档将被保存到“我的文档”文件夹中。