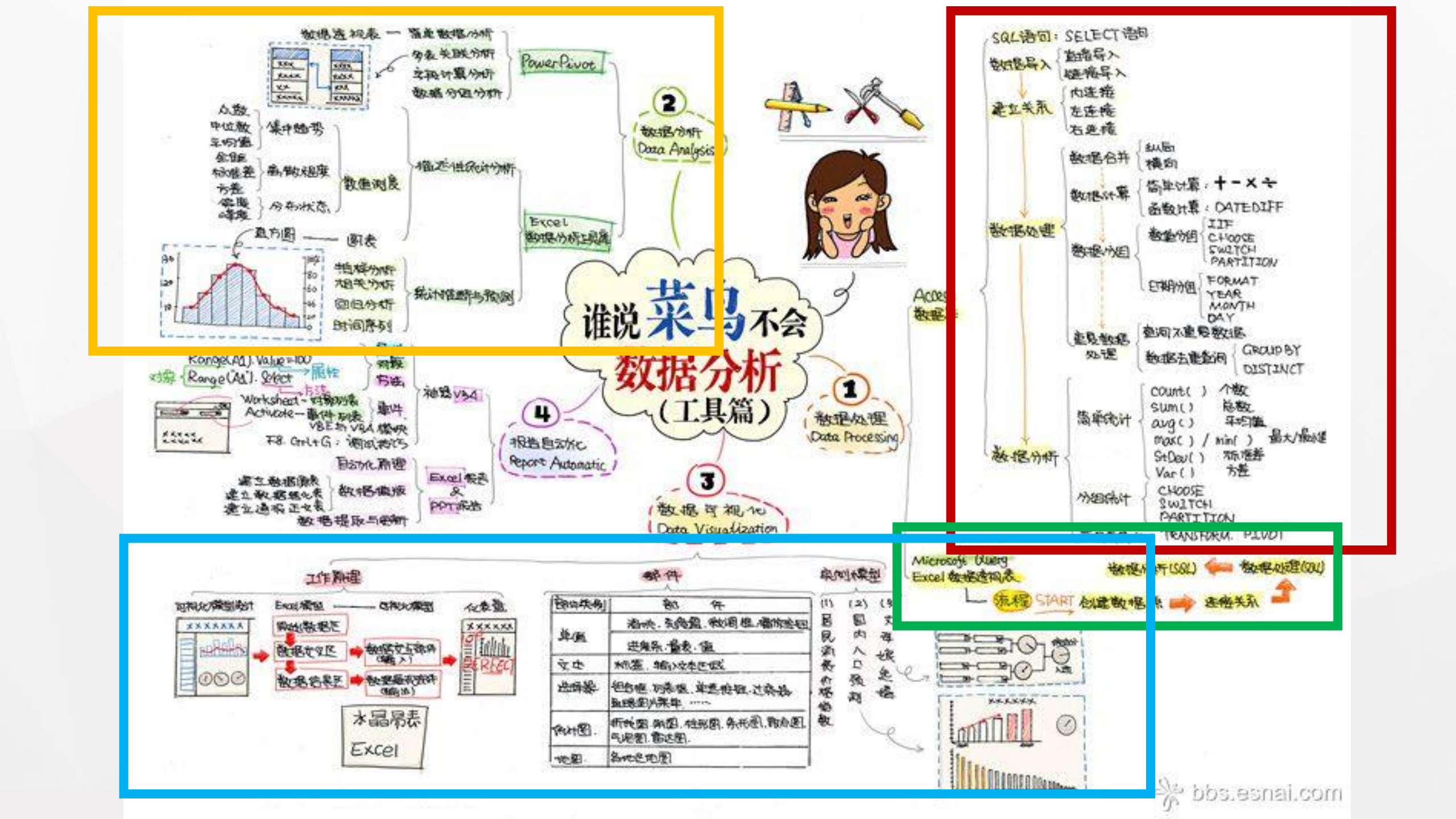
Excel也能搞定数据分析

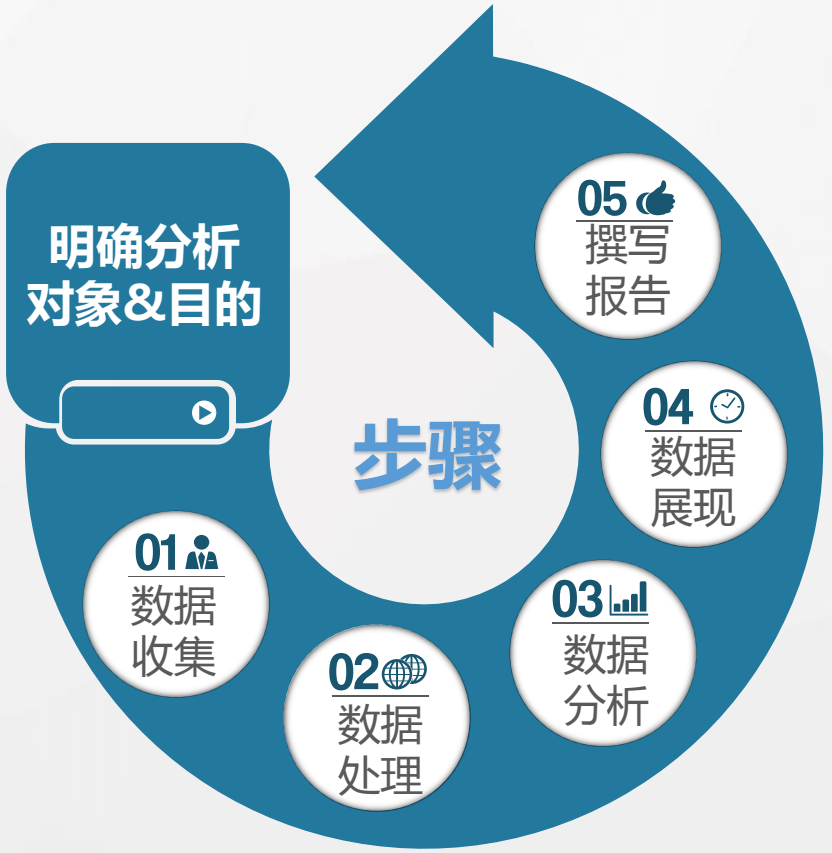

谁说菜鸟不会
数据分析
（工具篇）
1
数据处理
Data Processing
2
数据分析
Data Analysis
3
数据可视化
Data Visualization
4
报告自动化
Report Automatic
PowerPivot
Excel
数据分析工具库
描述性统计分析
统计推断与预测
数值测度
集中趋势
离散程度
分布状态
直方图
图表
VBA
Excel
&
PPT报告
SQL语句：SELECT语句
数据导入
建立关系
数据处理
数据分析
数据合并
数据计算
数据分组
重复数据处理
简单统计
分组统计
count( ) 个数
sum( ) 总数
avg( ) 平均值
max( ) / min( ) 最大/最小值
StDev( ) 标准差
Var( ) 方差
CHOOSE
SWITCH
PARTITION
FORMAT
YEAR
MONTH
DAY
GROUP BY
DISTINCT
TRANSFORM、PIVOT
Microsoft Query
Excel 数据透视表
流程
START
工作原理
水晶易表
Excel

何为数据分析

- 用适当的**统计方法**对收集来的大量**数据**进行**分析**
- 旨在把隐藏在看似杂乱无章的数据背后的信息**集中**和**提炼**出来，总结出研究对象的**内在规律**
- 比起SPSS、Python、R语言等数据分析工具，**EXCEL**是大家较为熟悉的办公软件。

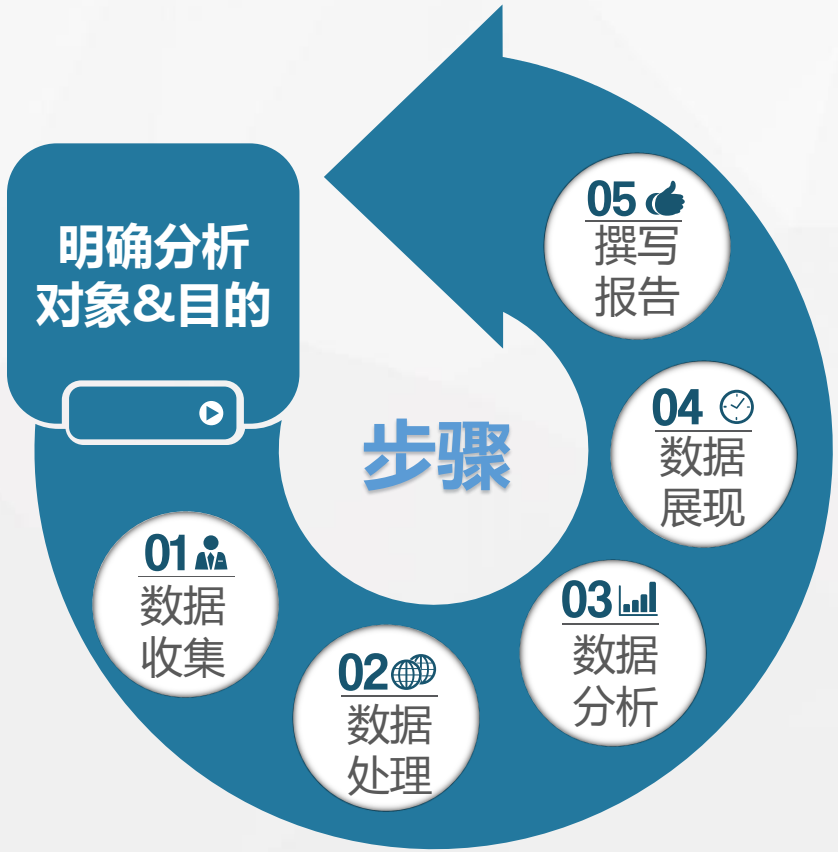
数据分析的背后是**数值之间的逻辑**

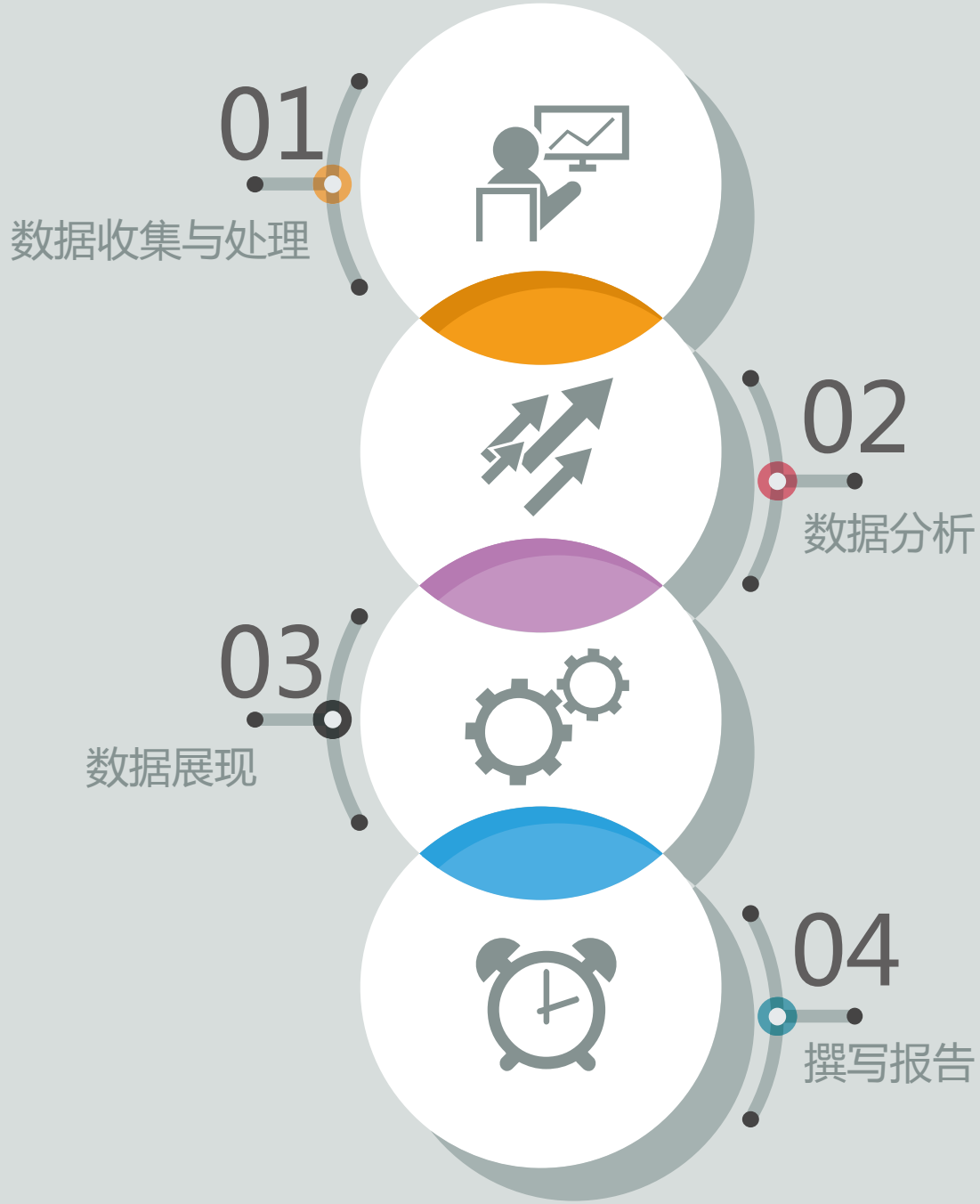
数值之间的逻辑背后是**业务的逻辑**



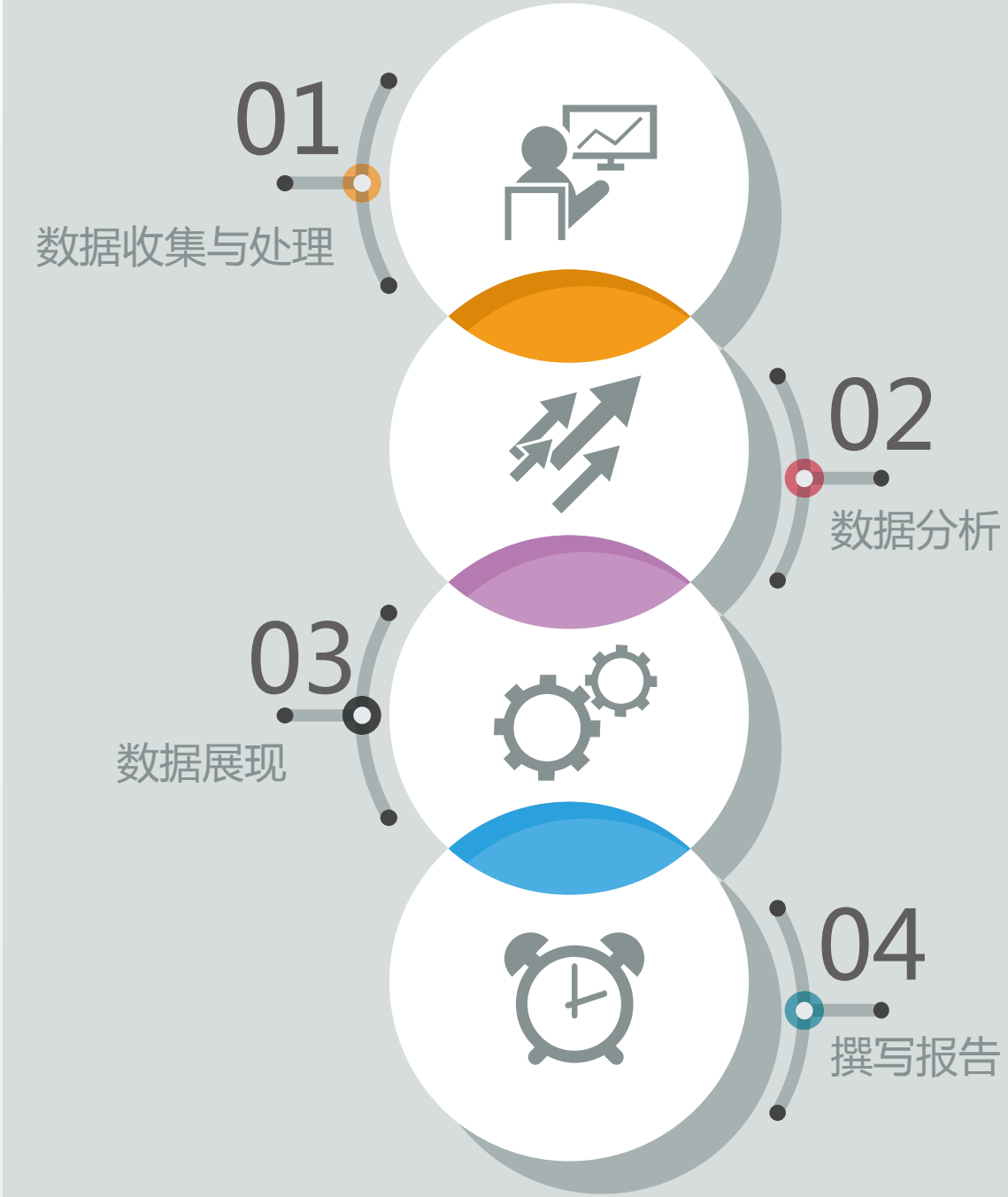
EXCEL能为我们做哪些数据分析工作？

目录

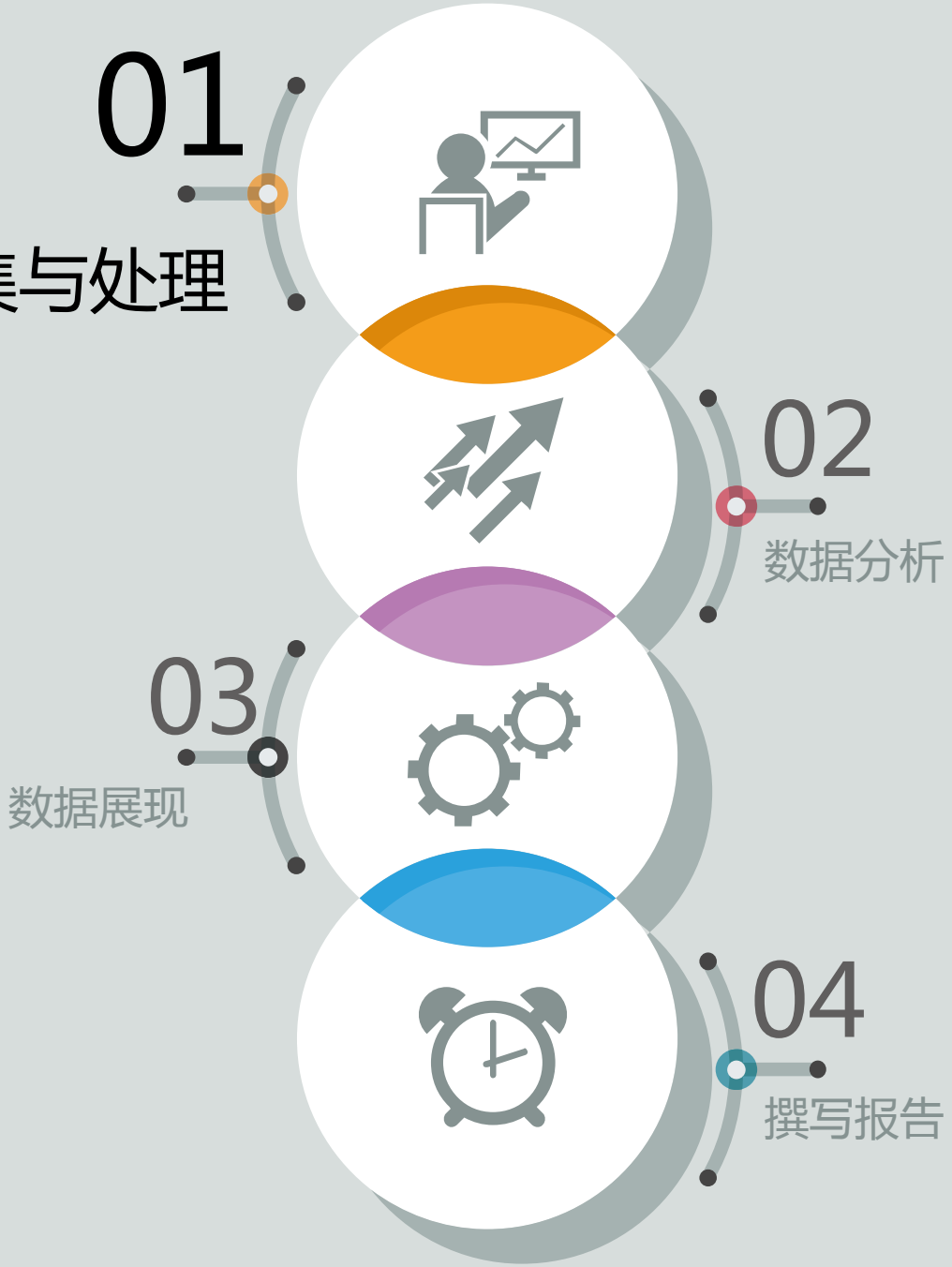

明确分析
对象&目的
步骤
01
数据
收集
02
数据
处理
03
数据
分析
04
数据
展现
05
撰写
报告


01
数据收集与处理
02
数据分析
03
数据展现
04
撰写报告

目录



目录



数据收集与处理

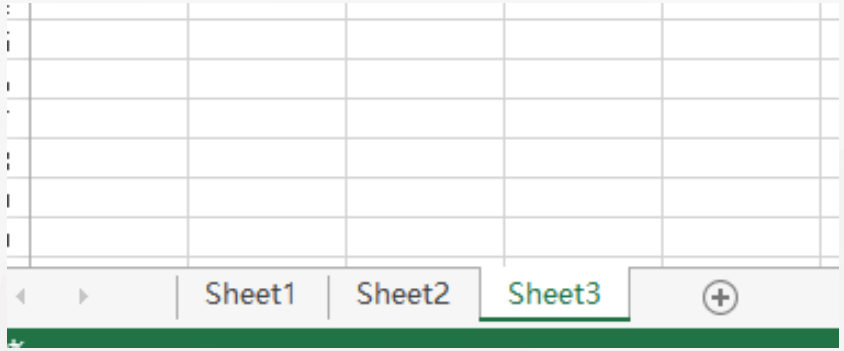
数据收集

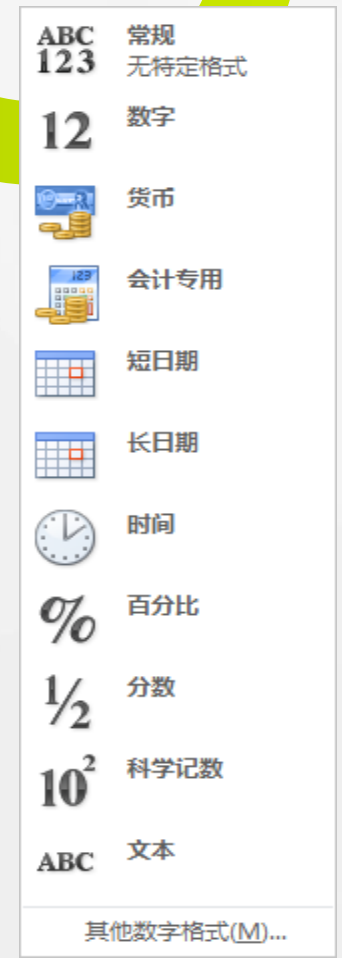
一般数据来源包括如下方式：

随着大数据时代的到来，我们每年所产生的数据成几何级数增长，数据的来源与形式多种多样。
从大量的、异构的数据中得到面向某一行业或某一问题的深入的分析，是大数据时代对我们每一个人的考验

EXCEL这个分析工具是如何理解数据？

EXCEL认为数据存在在一张一张表单里sheet





字段：这些题名，刊名的属性表示的就是一个字段

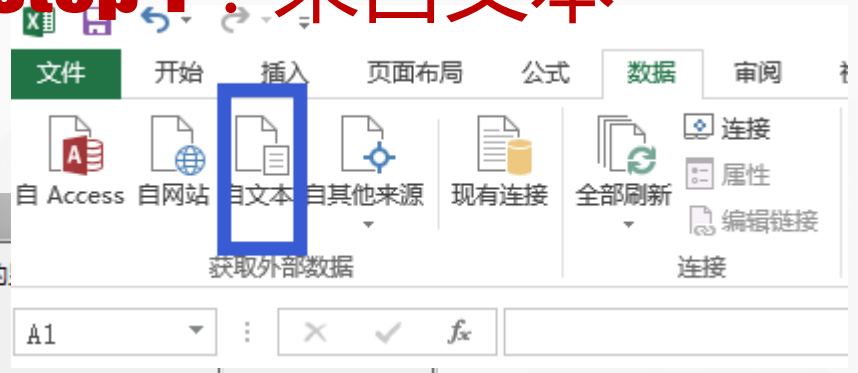
	A	B	C	D	E	
1	PT	AU	BA	BE	GP	AF
2	J	Lu, T; Liang, GZ; Peng, YL; Chen, T				Lu, Tingh; Liang, Guozheng; Peng, Yac
3	J	Tang, YS; Liang, GZ; Zhang, ZP; Han, J				Tang, Yusheng; Liang, Guozheng; Zhang
4	J	Li, J; Qiao, SR; Han, D; Li, M				Li Jun; Qiao Shengru; Han Dong; Li M
5	J	Yung, KL; Kong, J; Xu, Y				Yung, Kai-Leung; Kong, Jie; Xu, Yan
6	J	Zhang, GB; Fan, XD; Kong, J; Liu, YY				Zhang Guo-Bin; Fan Xiao-Dong; Kong Ji
7	J	Liu, YY; Yu, Y; Zhang, GB; Tang, MF				Liu, Yu-Yang; Yu, Yu; Zhang, Guo-Bin;
8	J	Li, XL; Li, JS; Hu, R; Kou, HC; Fu, HZ				Li Xiaoli; Li Jinshan; Hu Rui; Kou Ho
9	J	Luo, WZ; Shen, J; Li, QL; Fu, HZ				Luo Wenzhong; Shen Jun; Li Qinglin; F
10	J	Yang, J; Zhu, B; Mao, GW; Xu, YQ; Liu, JP				Yang Juan; Zhu Bing; Mao Gen-Wang; Xu
11	J	Rong, JH; Jiang, JS; Xie, YM				Rong, Jian Hua; Jiang, Jie Sheng; Xie
12	J	Wang, XJ; Feng, ZZ; Wang, FS; Yue, ZF				Wang Xinjun; Feng Zhenzhou; Wang Fush
13	J	An, WG; Li, WJ				An Weigang; Li Weiji
14	J	Xiao, FJ; Zhang, P; Liu, S; Zhu, JL				Xiao Fa-Jun; Zhang Peng; Liu Sheng; Z

记录:每一篇文章是一个记录包括题名，刊名，入藏号等

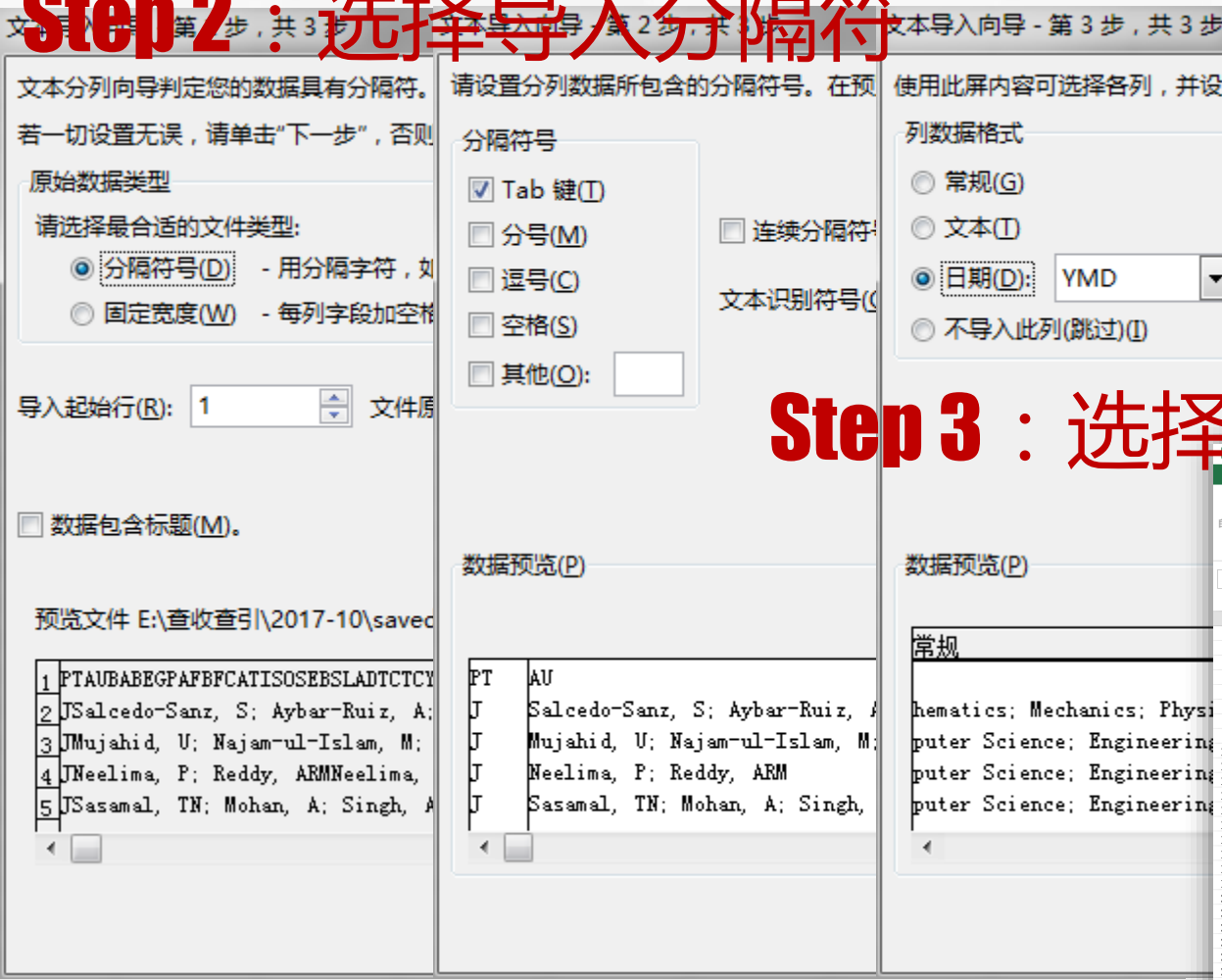
数据导入

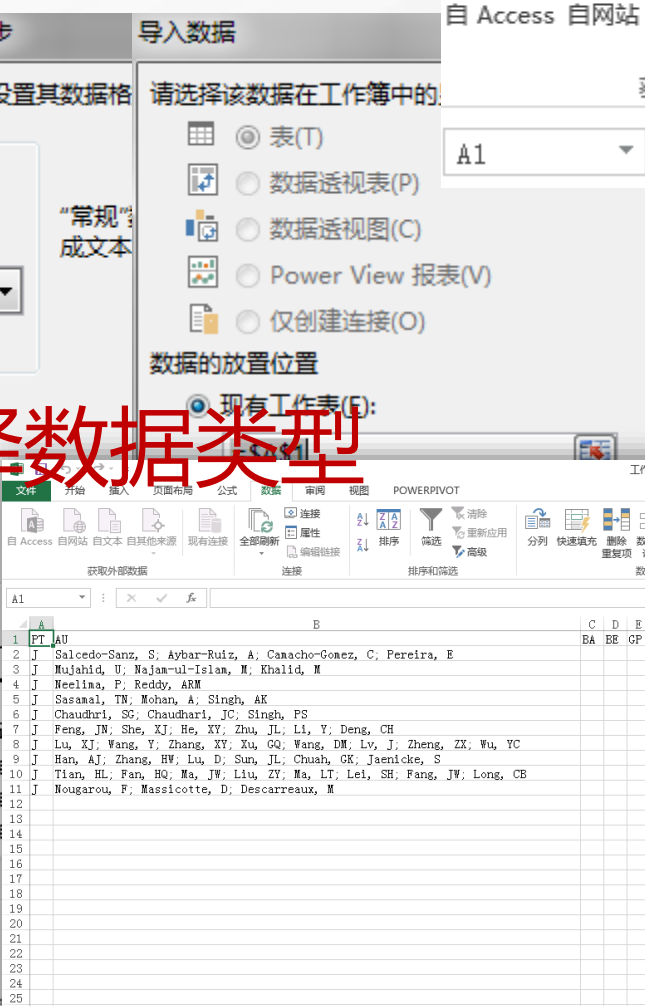
文本导入（举一个导入Web of Science数据的例子）

Step 1：来自文本



Step 2：选择导入分隔符



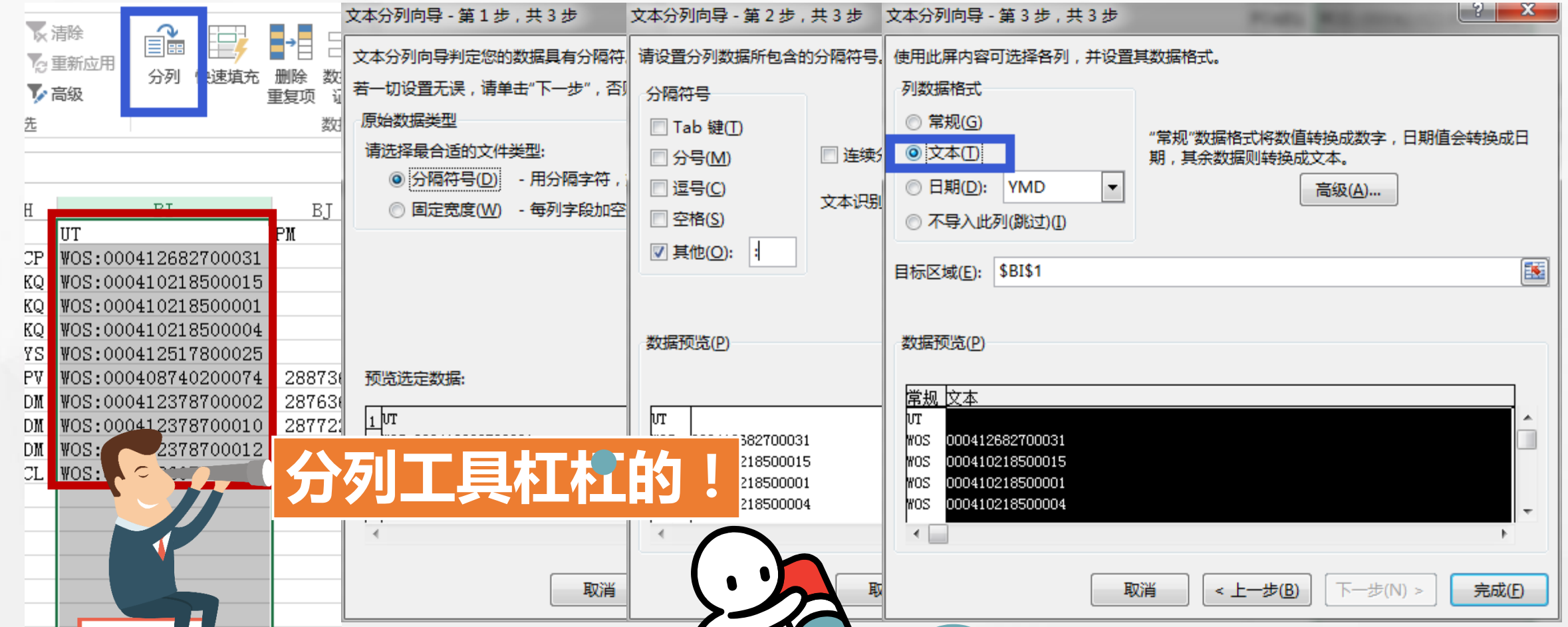


Step 3：选择数据类型

搞定！

数据导入

从数据库导入之后很多数据在数据库中设定了一些**标记**，不方便我们后期的处理与分析



分列工具杠杠的！

对于纯数字的字段一定要转化为文本格式
否则将作为数字保存

- A1是相对引用
- $A1绝对引用列是混合引用
- A$1绝对引用行是混合引用
- $A$1绝对引用行和列是绝对引用

F4变变变！

$在谁的前面就绝对引用谁

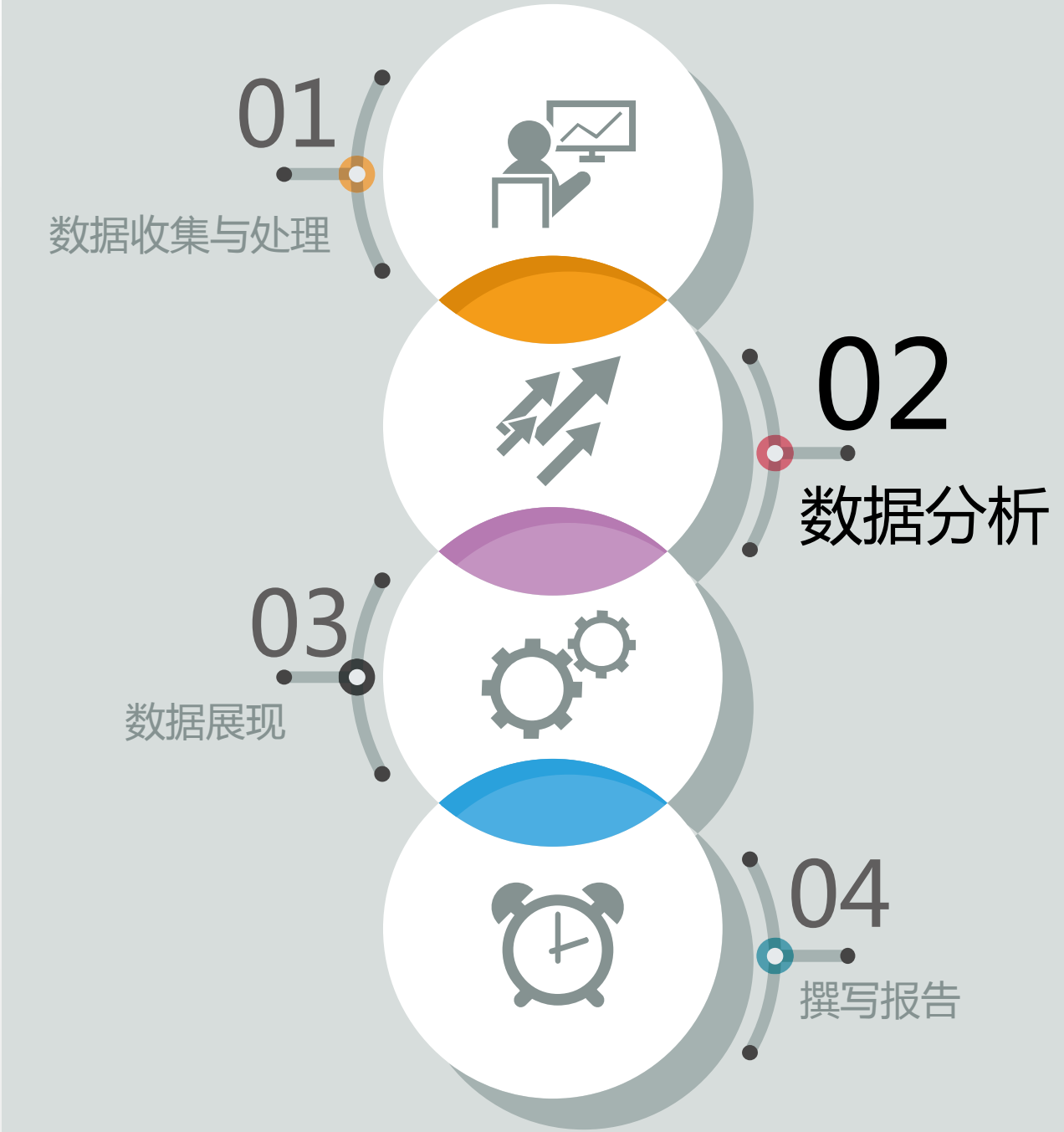
A1(相对引用)：在下拉拖动引用时,会变成引用A2,A3,A4...,右拉拖动时引用变成B1,C1,D1....

**A$1(混合引用)：**当你下拉复制时想保证引用的只是A1单元格时,A1就要加$符号,成A$1,这样在下拉时能保证对A列第一行的相对引用(即保持行号在引用时不产生变动)

**$A1(混合引用)：**当你右拉复制时想保证引用的只是A1单元格时,A1就要加$符号,成$A1,这样在右拉时能保证对A列第一行的相对引用(即保持列标在引用时不产生变动)

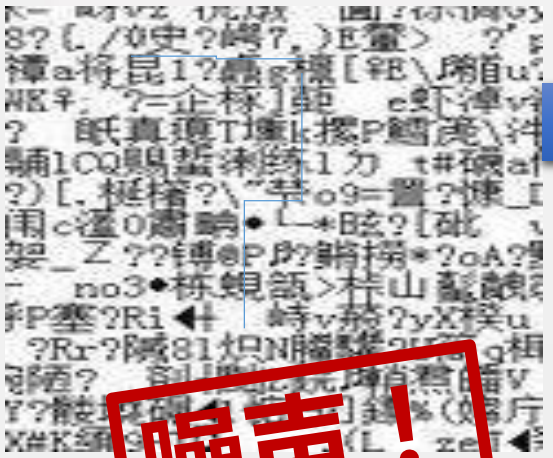
**$A$1(绝对引用)：**当你在下拉和右拉复制时想保证引用的只是A1单元格时,A1就要加$符号,成$A$1,这样在下拉和右拉时能保证对A列第一行的绝对引用(即保持行号列标在引用时不产生变动)

目录



数据分析的一般步骤

原始数据



噪声！

数据清洗

整齐的数据

月份	地区	品名	数量	零售价	销售额	品名	数量	零售价	销售额
一月	贵阳	料理机	31	264.4	￥8,196.4	榨汁机	22	216.2	￥4,756.40
一月	洛阳	料理机	36	226.9	￥8,168.4	榨汁机	23	224.5	￥5,163.50
一月	内蒙	料理机	45	213.8	￥9,621.0	榨汁机	13	213.6	￥2,776.80
一月	西安	料理机	40	260.2	￥10,408.0	榨汁机	10	205.5	￥2,055.00
二月	贵阳	料理机	35	246.8	￥8,638.0	榨汁机	43	216.3	￥9,301.80
二月	洛阳	料理机	27	267.9	￥7,233.3	榨汁机	35	221.5	￥7,752.50
二月	内蒙	料理机	23	208.5	￥4,795.5	榨汁机	23	213.1	￥4,901.30
二月	西安	料理机	30	260.8	￥7,824.0	榨汁机	12	209.6	￥2,515.20
三月	贵阳	料理机	32	265.2	￥8,486.4	榨汁机	34	216.1	￥7,347.40
三月	洛阳	料理机	47	236.5	￥11,115.5	榨汁机	41	223.5	￥9,163.50
三月	内蒙	料理机	41	215.8	￥8,847.8	榨汁机	40	213.2	￥8,528.00
三月	西安	料理机	25	260.1	￥6,502.5	榨汁机	25	209.8	￥5,245.00
四月	贵阳	料理机	63	232.1	￥14,622.3	榨汁机	15	212.3	￥3,184.50
四月	洛阳	料理机	52	212.1	￥11,029.2	榨汁机	30	221.5	￥6,645.00
四月	内蒙	料理机	37	219.7	￥8,128.9	榨汁机	16	214.1	￥3,425.60
四月	西安	料理机	50	269.2	￥13,460.0	榨汁机	12	209.5	￥2,514.00
五月	贵阳	料理机	54	211.2	￥11,404.8	榨汁机	24	216.9	￥5,205.60
五月	洛阳	料理机	57	221.1	￥12,602.7	榨汁机	25	217.5	￥5,437.50
五月	内蒙	料理机	47	219.6	￥10,321.2	榨汁机	22	214.9	￥4,727.80
五月	西安	料理机	46	260.1	￥11,964.6	榨汁机	15	208.5	￥3,127.50
六月	贵阳	料理机	23	263.9	￥6,069.7	榨汁机	13	216.2	￥2,810.60
六月	洛阳	料理机	43	263.9	￥11,347.7	榨汁机	30	221.5	￥6,645.00
六月	内蒙	料理机	15	263.9	￥3,958.5	榨汁机	32	214.1	￥6,851.20
六月	西安	料理机	35	263.9	￥9,236.5	榨汁机	15	208.5	￥3,127.50

整齐而完整！

数据加工

想要的数据

能够用于分析

数据清洗

导入的数据常常存在如下问题：

数据重复

缺失数据

逻辑错误

01 删除数据重复

进一步提高数据的纯度
最为常见的需求
常用的4种方法

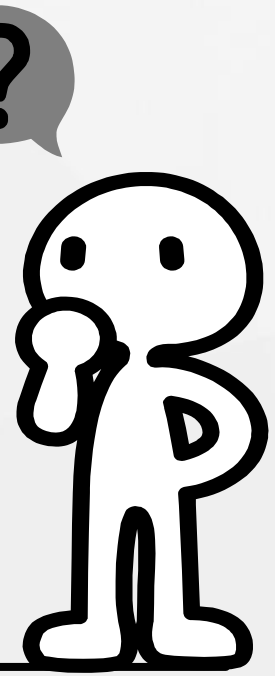
02 检查逻辑错误

在开始分析之前让分析工具在数据逻辑上进行初始筛选，提高后续工作的效率

03 补充缺失数据

提高数据的完整度

例如：
小A给你提供的是我们学校女性教师的发文情况
小B给你提供的是我校青年教师的发文情况
小C给你提供是的我校计算机学院发文情况
当我们把这些数据导入一个表中进行分析的时候，首先就是要根据工号等信息去除数据中的重复

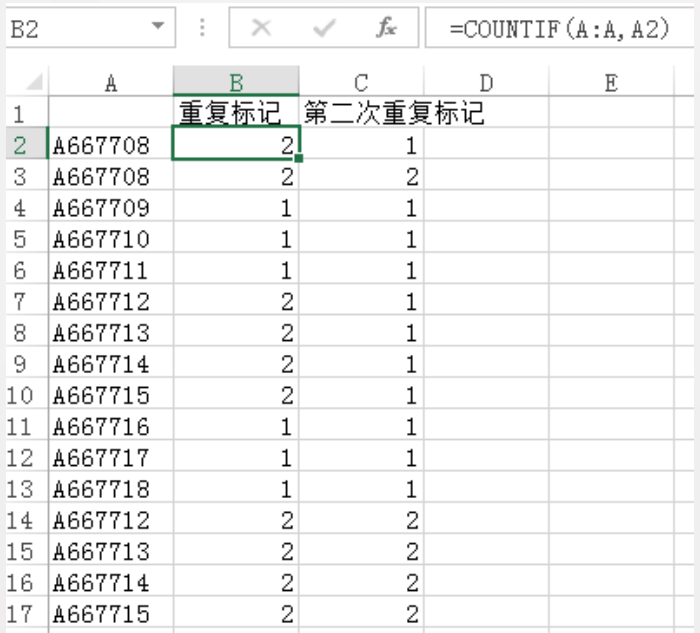


数据重复 方法一：用计数函数COUNTIF识别重复数据

COUNTIF(range,criteria):对区域中满足单个指定条件的单元格进行计数

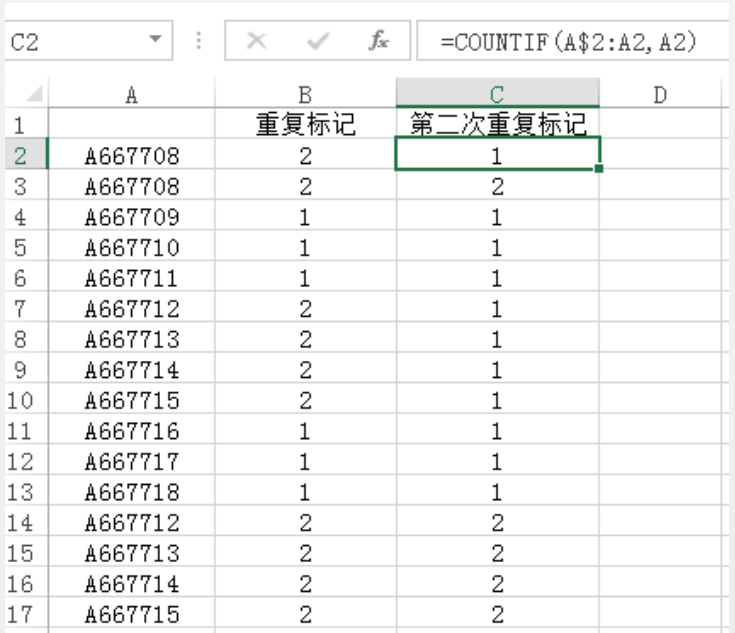
要计数的单元格范围

计算条件，其格式可以为数字、表达式或文本；
例如，条件可以表示为32、“32”、“>32”或“apples”



B2 =COUNTIF(A:A,A2)

	A	B	C	D	E
1		重复标记	第二次重复标记		
2	A667708	2	1		
3	A667708	2	2		
4	A667709	1	1		
5	A667710	1	1		
6	A667711	1	1		
7	A667712	2	1		
8	A667713	2	1		
9	A667714	2	1		
10	A667715	2	1		
11	A667716	1	1		
12	A667717	1	1		
13	A667718	1	1		
14	A667712	2	2		
15	A667713	2	2		
16	A667714	2	2		
17	A667715	2	2		



C2 =COUNTIF(A$2:A2,A2)

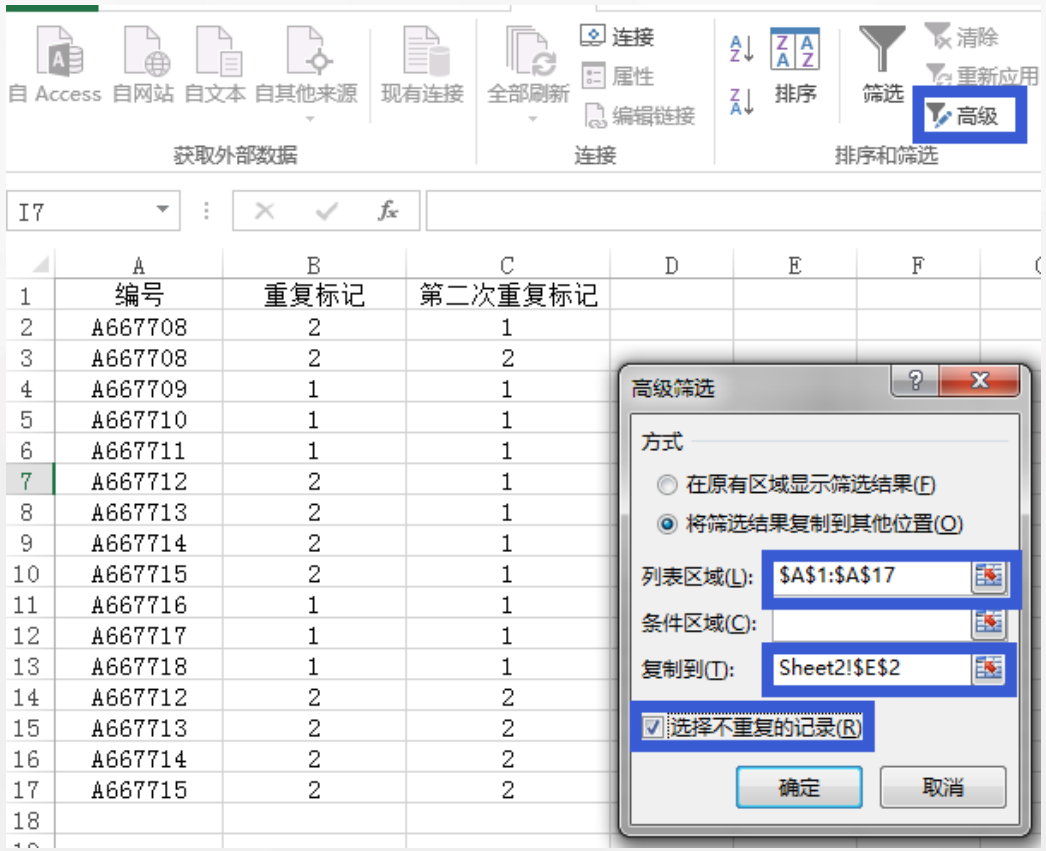
	A	B	C	D
1		重复标记	第二次重复标记	
2	A667708	2	1	
3	A667708	2	2	
4	A667709	1	1	
5	A667710	1	1	
6	A667711	1	1	
7	A667712	2	1	
8	A667713	2	1	
9	A667714	2	1	
10	A667715	2	1	
11	A667716	1	1	
12	A667717	1	1	
13	A667718	1	1	
14	A667712	2	2	
15	A667713	2	2	
16	A667714	2	2	
17	A667715	2	2	

重复标记：表示重复的次数

第二次重复标记：在该列表中第几次出现

用函数的计算方法比较灵活，可以根据函数计算的结果进行筛选

数据重复 方法二：用菜单操作来筛选重复数据

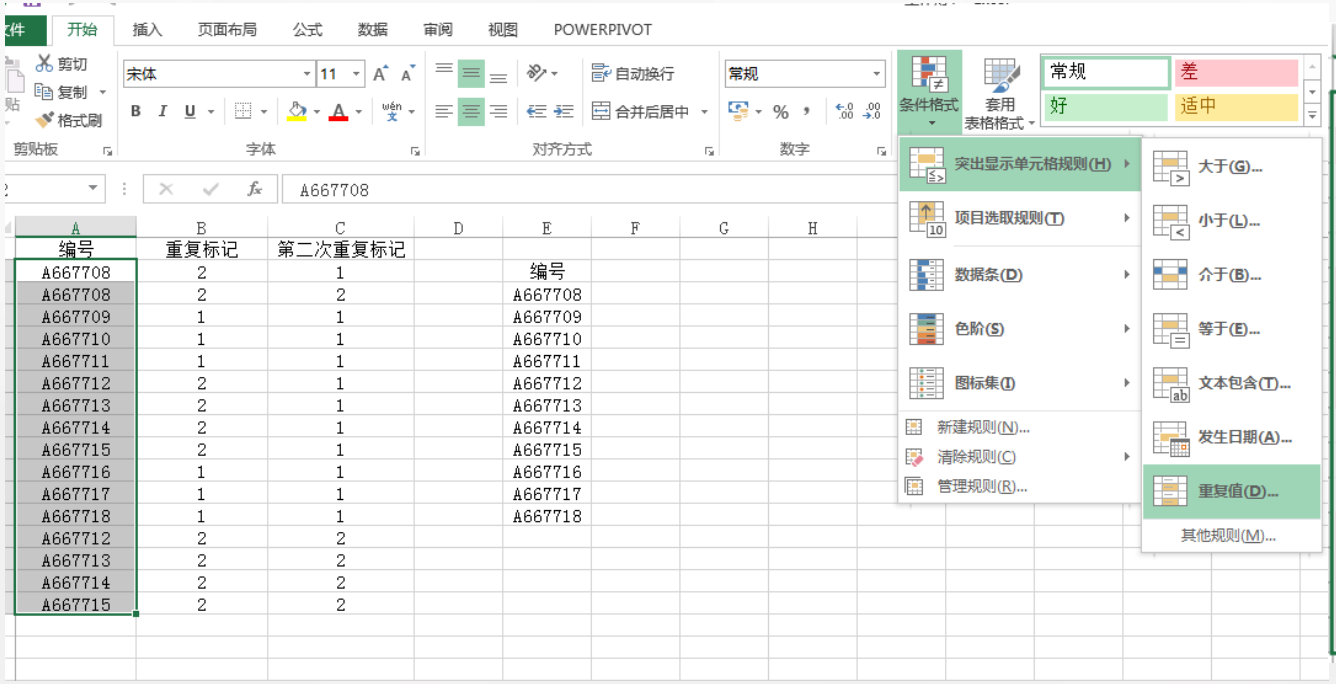


	A	B	C
1	编号	重复标记	第二次重复标记
2	A667708	2	1
3	A667708	2	2
4	A667709	1	1
5	A667710	1	1
6	A667711	1	1
7	A667712	2	1
8	A667713	2	1
9	A667714	2	1
10	A667715	2	1
11	A667716	1	1
12	A667717	1	1
13	A667718	1	1
14	A667712	2	2
15	A667713	2	2
16	A667714	2	2
17	A667715	2	2
18			

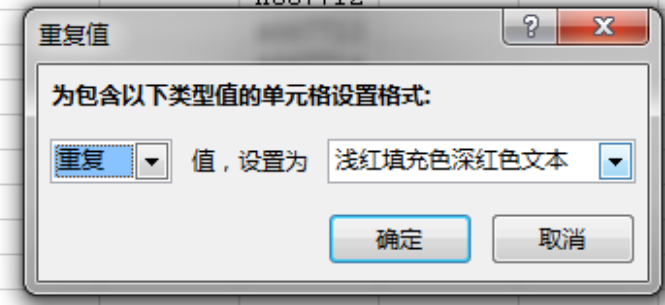
A	B	C	D	E
编号	重复标记	第二次重复标记		
A667708	2	1		编号
A667708	2	2		A667708
A667709	1	1		A667709
A667710	1	1		A667710
A667711	1	1		A667711
A667712	2	1		A667712
A667713	2	1		A667713
A667714	2	1		A667714
A667715	2	1		A667715
A667716	1	1		A667716
A667717	1	1		A667717
A667718	1	1		A667718
A667712	2	2		
A667713	2	2		
A667714	2	2		
A667715	2	2		

这种方法需要对筛选功能有一定的理解，注意一定要勾选**选择不重复记录**

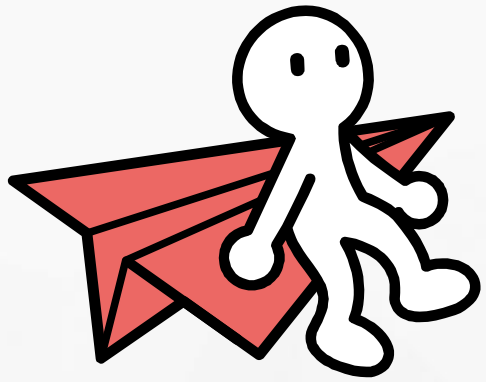
数据重复 方法三：用条件格式标识重复数据

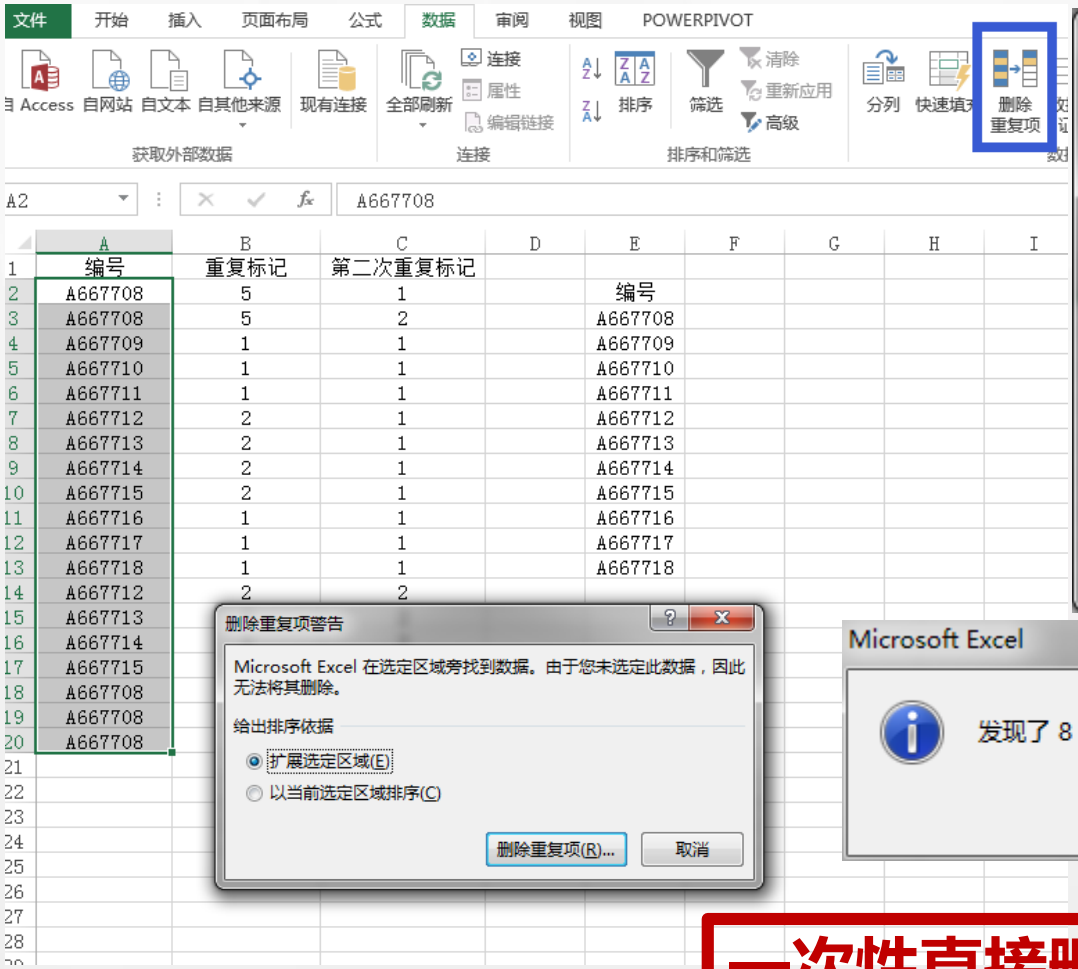


编号	重复标记	第二次重复标记		编号
A667708	2	1		
A667708	2	2		A667708
A667709	1	1		A667709
A667710	1	1		A667710
A667711	1	1		A667711
A667712	2	1		A667712
A667713	2	1		A667713
A667714	2	1		A667714
A667715	2	1		A667715
A667716	1	1		A667716
A667717	1	1		A667717
A667718	1	1		A667718
A667712	2	2		
A667713	2	2		
A667714	2	2		
A667715	2	2		

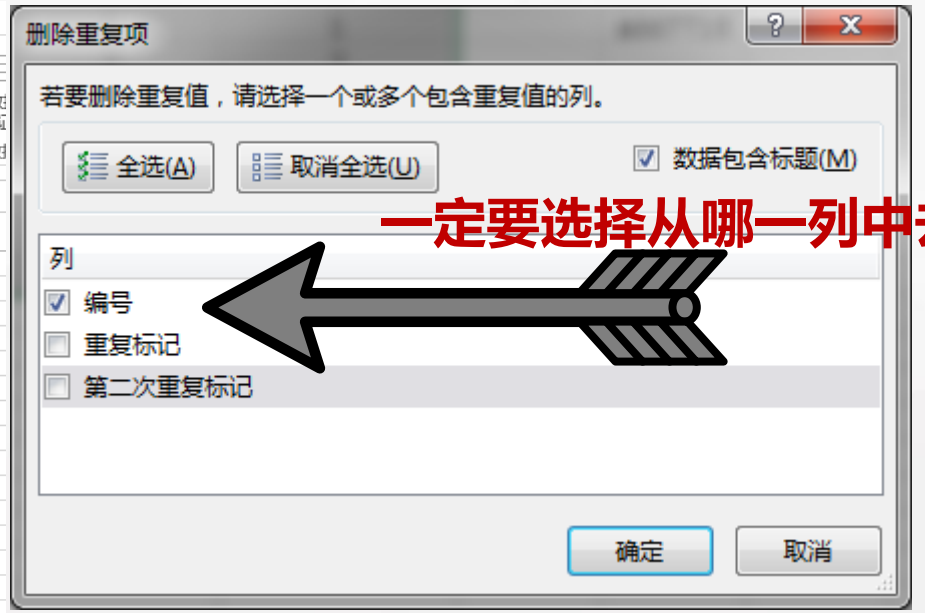


数据重复 方法四：用删除重复数据

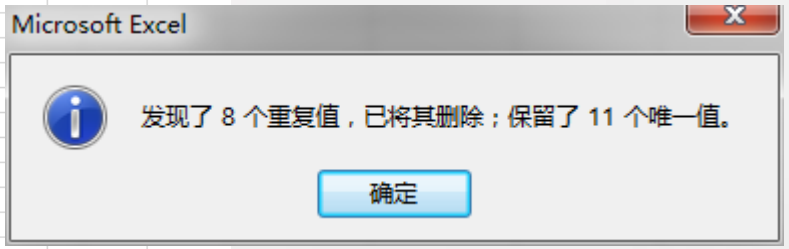




	A	B	C	D	E	F	G	H	I
1	编号	重复标记	第二次重复标记						
2	A667708	5	1		编号				
3	A667708	5	2		A667708				
4	A667709	1	1		A667709				
5	A667710	1	1		A667710				
6	A667711	1	1		A667711				
7	A667712	2	1		A667712				
8	A667713	2	1		A667713				
9	A667714	2	1		A667714				
10	A667715	2	1		A667715				
11	A667716	1	1		A667716				
12	A667717	1	1		A667717				
13	A667718	1	1		A667718				
14	A667712	2	2						
15	A667713								
16	A667714								
17	A667715								
18	A667708								
19	A667708								
20	A667708								
21									
22									
23									
24									
25									
26									
27									
28									



一定要选择从哪一列中去重复



	A	B	C
	编号	重复标记	第二次重复标记
	A667708	1	1
	A667709	1	1
	A667710	1	1
	A667711	1	1
	A667712	1	1
	A667713	1	1
	A667714	1	1
	A667715	1	1
	A667716	1	1
	A667717	1	1
	A667718	1	1

一次性直接删除不给你复查的机会

数据清洗

导入的数据常常存在如下问题：

数据重复

缺失数据

逻辑错误

01 删除数据重复

进一步提高数据的纯度

最为常见的需求

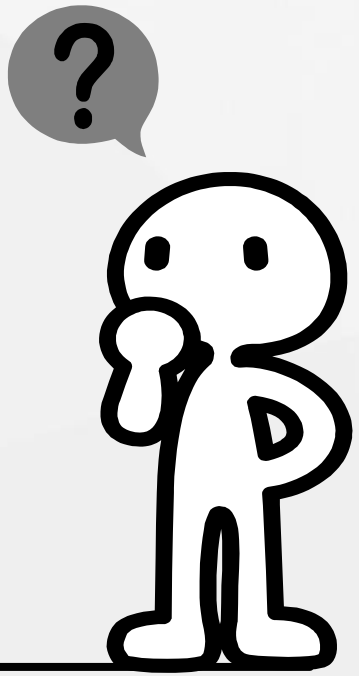
常用的4中方法

02 检查逻辑错误

在开始分析之间让分析工具在数据逻辑上进行初始筛选，提高后续工作的效率

补充缺失数据 03

提高数据的完整度

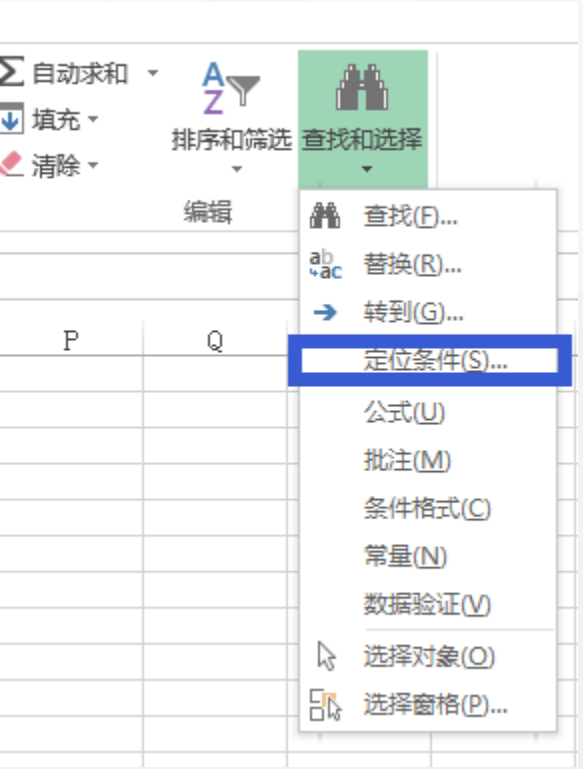


例如：数据里面总有一些空缺字段，
怎么办？还得分析！
找办法把这些数据补上！

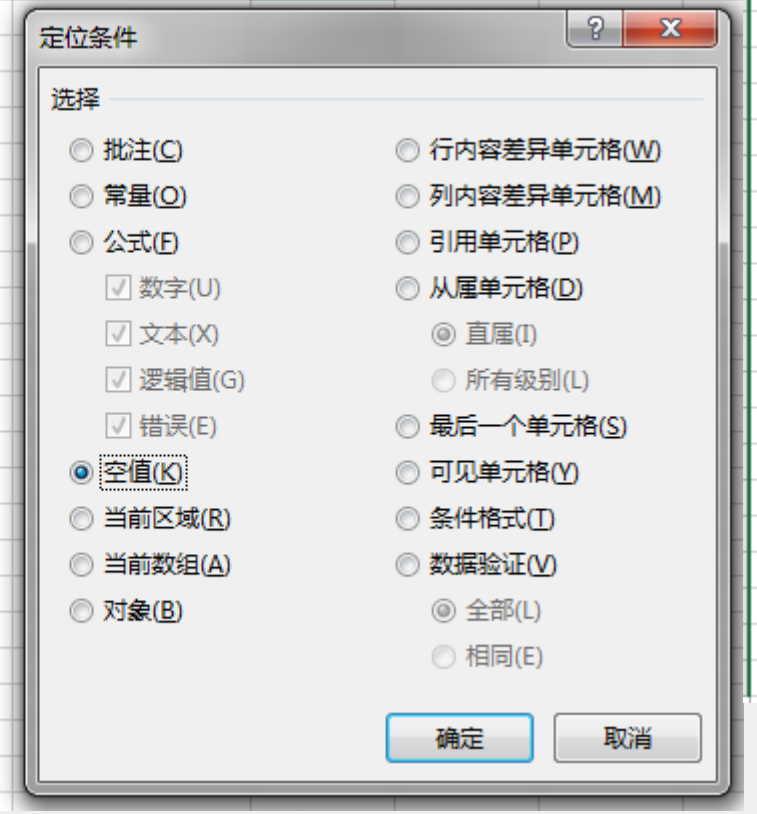
缺失数据

- 如何定位

快捷键：Ctrl+G弹出定位对话框



编号	篇数	份数
A667708	5	1
A667708	5	2
A667709	1	1
A667710	1	1
A667711	1	1
A667712	2	1
A667713	2	1
A667714	2	1
A667715	2	1
A667716	1	1
A667717	1	1
A667718	1	1
A667712	2	2
A667713	2	2
A667714	2	2
A667715	2	2
A667708		
A667708	5	4
A667708	5	5



编号	篇数	份数
A667708	5	1
A667708	5	2
A667709	1	1
A667710	1	1
A667711	1	1
A667712	2	1
A667713	2	1
A667714	2	1
A667715	2	1
A667716	1	1
A667717	1	1
A667718	1	1
A667712	2	2
A667713	2	2
A667714	2	2
A667715	2	2
A667708		
A667708	5	4
A667708	5	5

一次性选中空缺数据

缺失数据

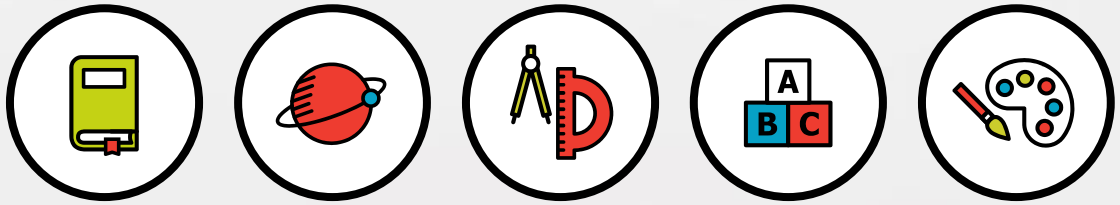
- 定位输入些什么呢？
  - 采用**四种策略**处理缺失数据

1、用一个样本的**统计量**的值代替**缺失值。均值、方差、计数等**

2、用一个**统计模型计算出来的值**去代替**缺失值**。常用的有回归模型、判别模型。

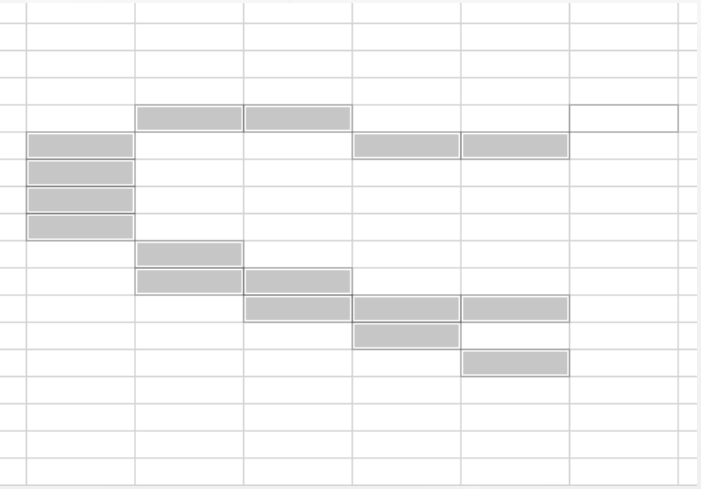
3、将有缺失的记录**删除**，不过可能会导致样本量的减少。

4、将有缺失值的**个案保留**，仅在相应的分析中做必要的删除。当调查的样本量较大，缺失值的数量不多，而变量之间不存在高度相关性时，可采用该方法。

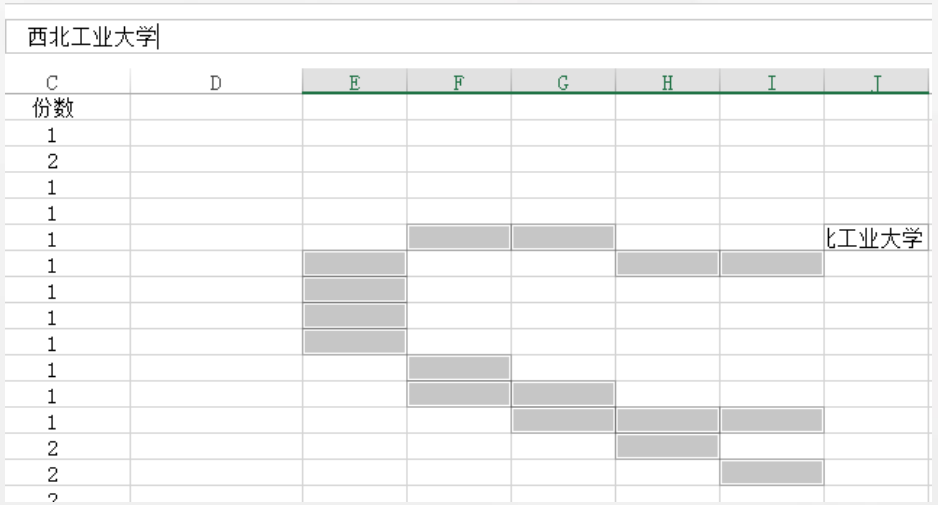


缺失数据

- 一次性填入多个相同数据　快捷键 “Ctrl+Enter” 键
- 例如：上一页我们提到了用某个量的均值来代替缺失数据，那么我们可以采用这个方法一次性补全缺失数据



Ctrl键



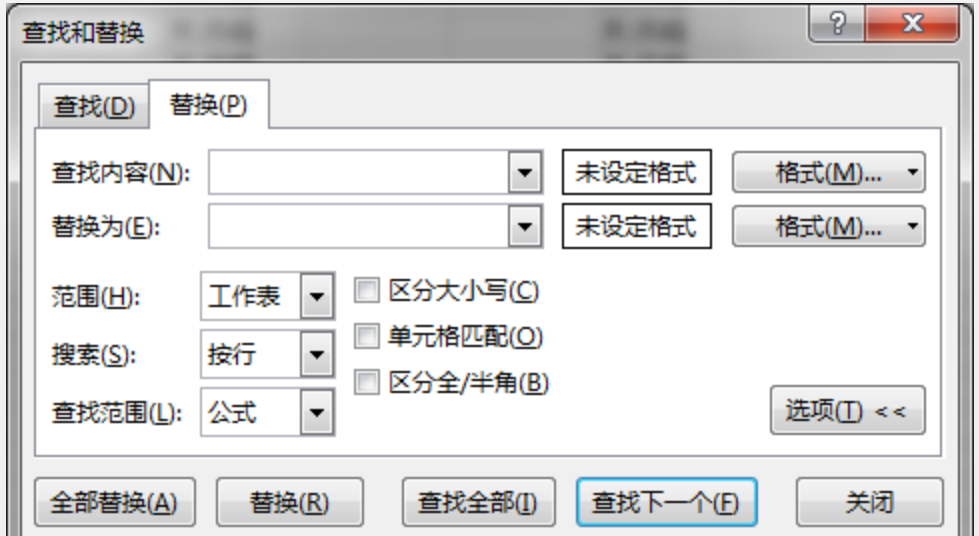
Ctrl+Enter键

缺失数据

- 如果不用鼠标键盘操作也可以尝试采用最为常用的方法
- 查找与替换

Ctrl+F
Ctrl+H
Ctrl+G

可以配合使用的快捷键



采用**通配符**模糊查找

搜索目标	搜索关键词的写法
以a开头的字符串	a*
以b结尾的字符串	*b
包含a的字符串	*a*
a排在第二位的字符串	?a*

数据清洗

导入的数据常常存在如下问题：

数据重复

缺失数据

逻辑错误

01 删除数据重复

进一步提高数据的纯度

最为常见的需求

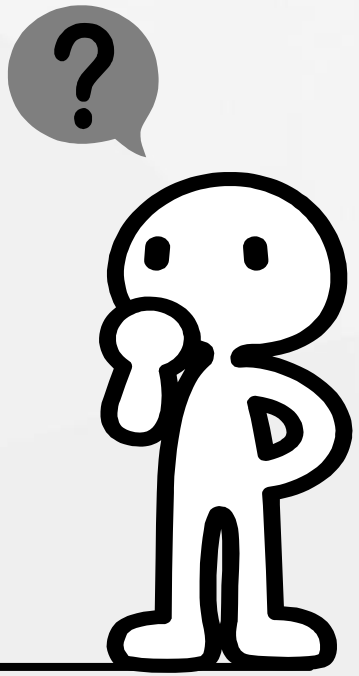
常用的4中方法

02 检查逻辑错误

在开始分析之间让分析工具在数据逻辑上进行初始筛选，提高后续工作的效率

补充缺失数据 03

提高数据的完整度



检查逻辑错误

- 利用IF函数检查错误

IF(logical_test,value_if_true,value_if_false):执行真假值判断，根据逻辑计算

表示计算结果为TURE或FALSE的表达式

为TRUE时返回的值

为FALSE时返回的值

注意：

1、条件表达式是用比较运算符(>,=,<)建立的式子，无比较就无判断

2、两个值若是**数值数据**可直接书写，若是**文本数据**则要用双引号标记

3、参数里所有用到的标点符号都是英文状态下的标点符号

4、IF函数可进行嵌套，最多可以有七层

检查逻辑错误

- 利用IF函数检查错误

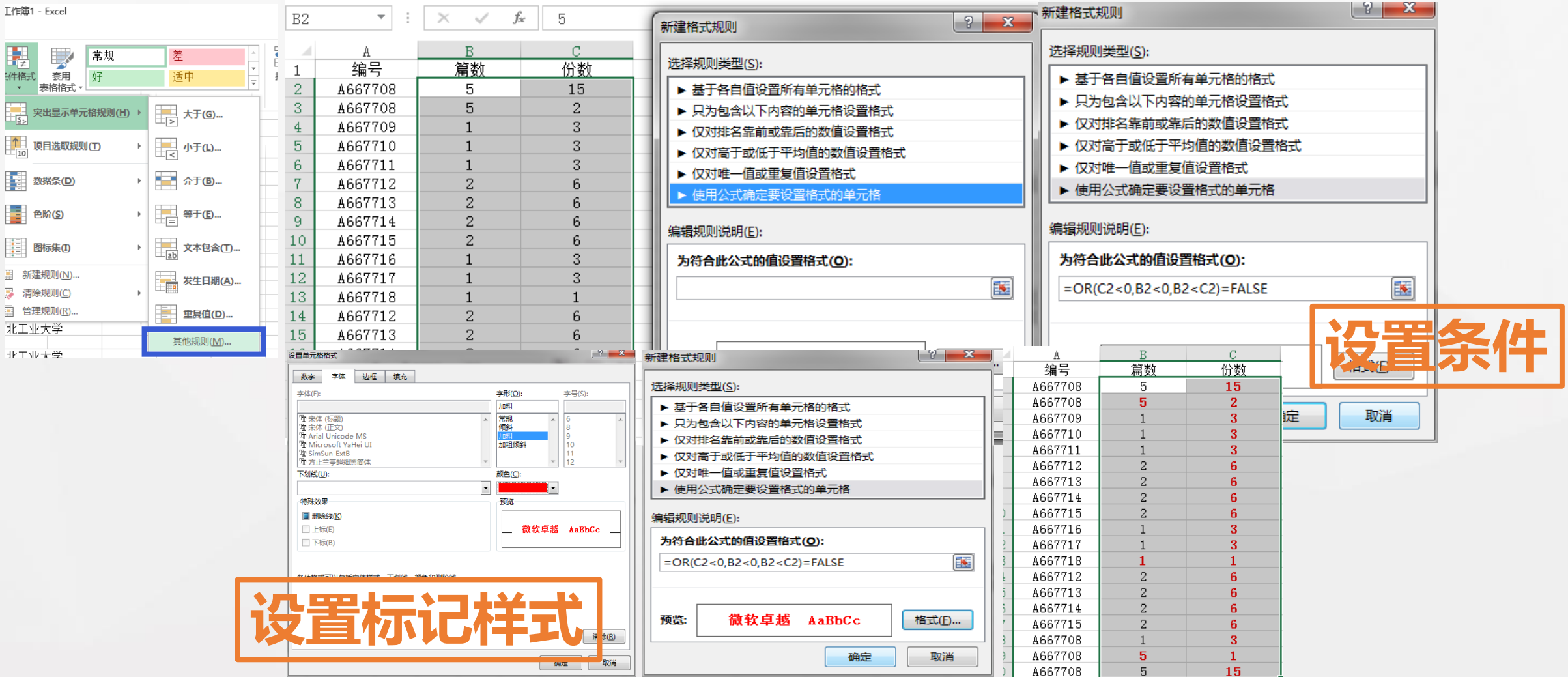
IF(logical_test,value_if_true,value_if_false):执行真假值判断，根据逻辑计算

D2 =IF(C2>B2,"正确","错误")

	A	B	C	D
1	编号	篇数	份数	
2	A667708	5	15	正确
3	A667708	5	2	错误
4	A667709	1	3	正确
5	A667710	1	3	正确
6	A667711	1	3	正确
7	A667712	2	6	正确
8	A667713	2	6	正确
9	A667714	2	6	正确
10	A667715	2	6	正确
11	A667716	1	3	正确
12	A667717	1	3	正确
13	A667718	1	1	错误
14	A667712	2	6	正确
15	A667713	2	6	正确
16	A667714	2	6	正确
17	A667715	2	6	正确
18	A667708	1	3	正确
19	A667708	5	1	错误
20	A667708	5	15	正确

检查逻辑错误

- 利用条件格式标记错误



检查逻辑错误

Excel本身能够对错误的信息进行标识，我们可以通过数据中出现错误符号判断是哪里出了问题！

错误符号	错误原因
####	数值或公式太长，单元格容纳不下
#DIV/0!	0为除数
#N/A	函数或公式中没有可用的数值
#NAME?	在公式中使用了excel不能识别的文本
#NULL!	使用了不正确的区域运算符或引用的单元格区域的交集为空
#NUM!	公式或函数中某些数字有问题
#REF!	单元格引用无效
#VALUE!	在公式中使用了错误的数据类型

数据加工

清洗之后可以对数据进行的加工

数据抽取【提取数据中的部分元素、合并一些数据】

数据计算【AVERAGE,SUM,MAX,MIN,Date,If】

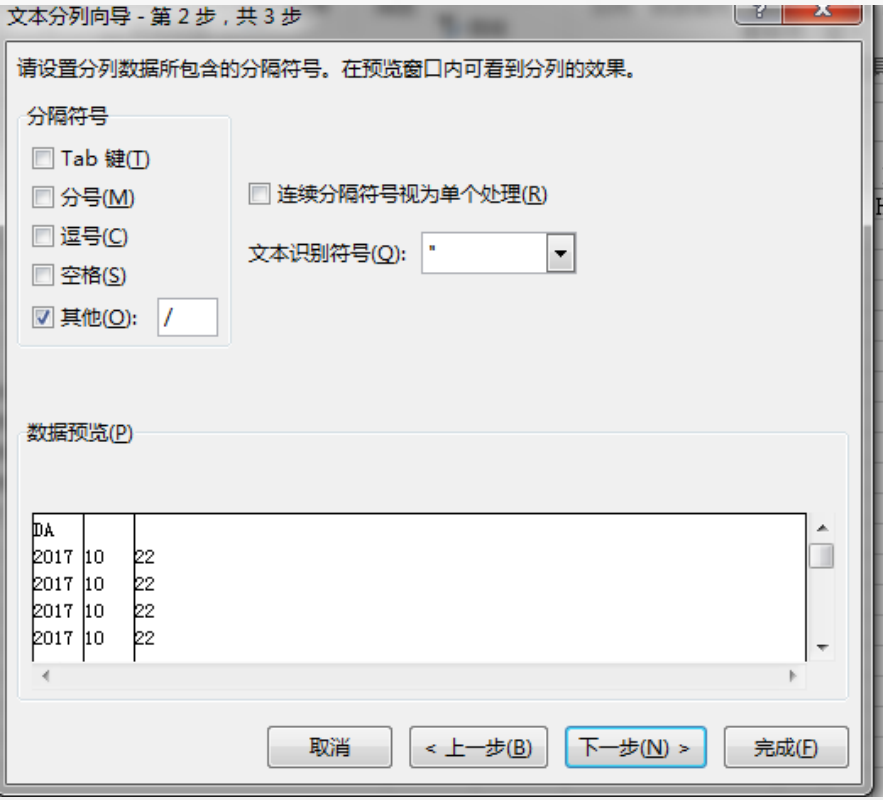
数据分组【VLOOKUP函数，采用近似匹配，SEARCH函数】

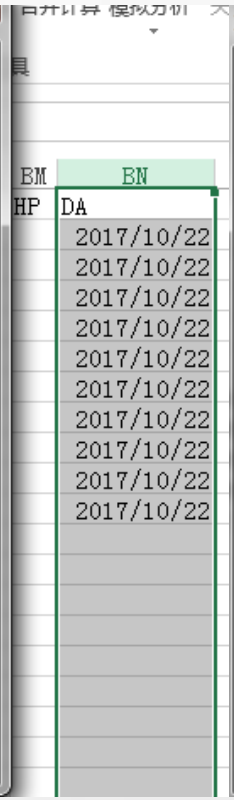
数据转换【数据行列变换】

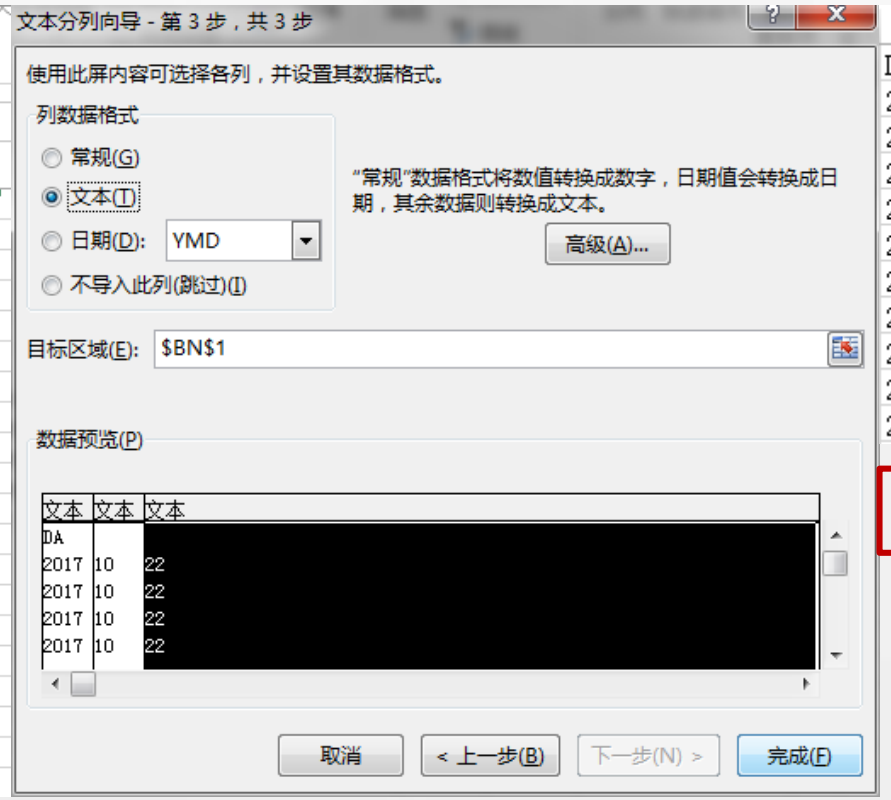
数据抽样【RAND函数，RAND()】

数据抽取

- 分列法抽取数据







BN	BO	BP
DA		
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22
2017	10	22

分列抽取法提取年月日

数据抽取

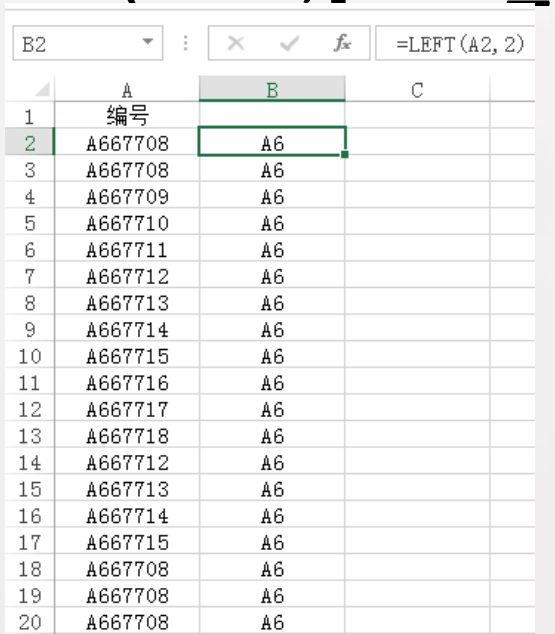
- 利用函数抽取数据

当数据中有特定的分隔符时，采用分列法非常方便。

但是有时候，我们需要提取特定的几个字符

✓ LEFT(TEXT,[num_chars]):得到字符串左部指定个数的字符

✓ RIGHT(TEXT,[num_chars]):得到字符串右部指定个数的字符



	A	B	C
1	编号		
2	A667708	A6	
3	A667708	A6	
4	A667709	A6	
5	A667710	A6	
6	A667711	A6	
7	A667712	A6	
8	A667713	A6	
9	A667714	A6	
10	A667715	A6	
11	A667716	A6	
12	A667717	A6	
13	A667718	A6	
14	A667712	A6	
15	A667713	A6	
16	A667714	A6	
17	A667715	A6	
18	A667708	A6	
19	A667708	A6	
20	A667708	A6	

数据抽取

- 字段合并
  - CONCATENATE函数
  - “&”（逻辑与）运算符

D2 =CONCATENATE(A2,"开具检索证明",B2,"篇",C2,"份")

	A	B	C	D	E
1	编号	篇数	份数		
2	A667708	5	15	A667708开具检索证明5篇15份	
3	A667708	5	2	A667708开具检索证明5篇2份	
4	A667709	1	3	A667709开具检索证明1篇3份	
5	A667710	1	3	A667710开具检索证明1篇3份	
6	A667711	1	3	A667711开具检索证明1篇3份	
7	A667712	2	6	A667712开具检索证明2篇6份	
8	A667713	2	6	A667713开具检索证明2篇6份	
9	A667714	2	6	A667714开具检索证明2篇6份	

20 =A20&"开具检索证明"&B20&"共"&C20&"份"

	A	B	C	D
	编号	篇数	份数	
	A667708	5	15	A667708开具检索证明5共15份
	A667708	5	2	A667708开具检索证明5共2份
	A667709	1	3	A667709开具检索证明1共3份
	A667710	1	3	A667710开具检索证明1共3份
	A667711	1	3	A667711开具检索证明1共3份
	A667712	2	6	A667712开具检索证明2共6份
	A667713	2	6	A667713开具检索证明2共6份
	A667714	2	6	A667714开具检索证明2共6份
	A667715	2	6	A667715开具检索证明2共6份

数据抽取

- 字段匹配
- VLOOFUP匹配函数：在表格的首列查找指定的数据，并返回指定数据所在行中指定列处的单元格内容

VLOOKUP(lookup_value,table_array,col_index_num,range_lookup)

在表格或区域的第一列中查找的
其参数可以是值或引用

包含数据的单元格区域
可以使用绝对区域或区域名称引用，table_array第一列的值是由loofup_value搜索的值

希望返回的匹配值的序列号，
其数值为1时，返回table_array第一列中的值

1近似匹配（可缺省）
0精细匹配

VLOOKUP(lookup_value,table_array,col_index_num,range_lookup)

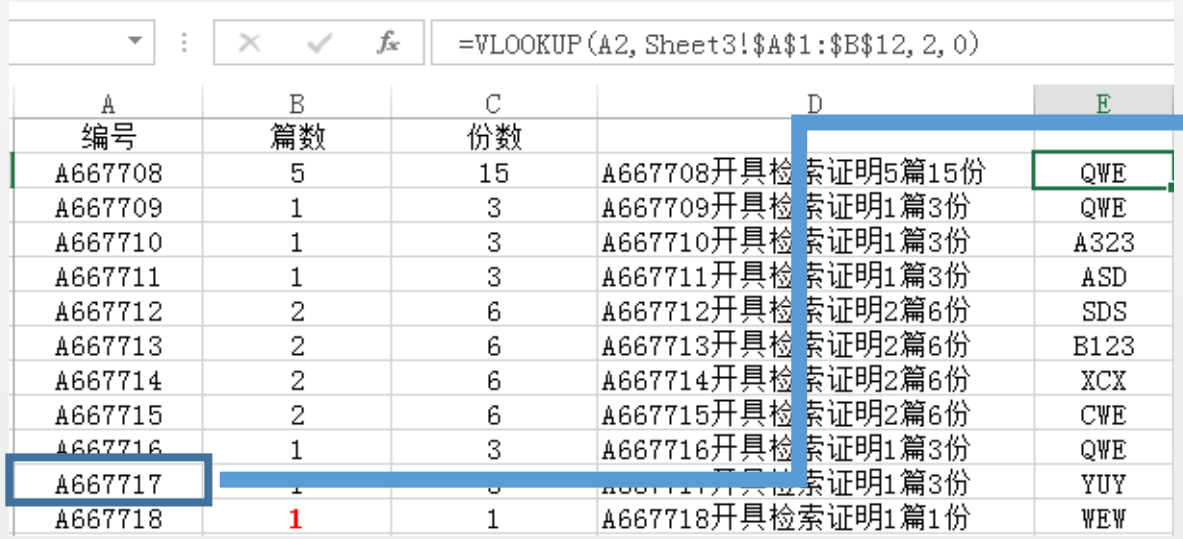
要在表格或区域的第一列中查找的值，其参数可以是值或引用

包含数据的单元格区域，可以使用绝对区域或区域名称引用，table_array 第一列的值是由loofup_value搜索的值

希望返回的匹配值的序列号，其数值为1时，返回table_array第一列中的值

1近似匹配（可忽略）
0精细匹配

=VLOOKUP(A2,Sheet3!$A$1:$B$12,2,0)



	A	B	C	D	E
	编号	篇数	份数		
	A667708	5	15	A667708开具检索证明5篇15份	QWE
	A667709	1	3	A667709开具检索证明1篇3份	QWE
	A667710	1	3	A667710开具检索证明1篇3份	A323
	A667711	1	3	A667711开具检索证明1篇3份	ASD
	A667712	2	6	A667712开具检索证明2篇6份	SDS
	A667713	2	6	A667713开具检索证明2篇6份	B123
	A667714	2	6	A667714开具检索证明2篇6份	XCX
	A667715	2	6	A667715开具检索证明2篇6份	CWE
	A667716	1	3	A667716开具检索证明1篇3份	QWE
	A667717	1	3	A667717开具检索证明1篇3份	YUY
	A667718	1	1	A667718开具检索证明1篇1份	WEW

	A	B
	编号	序列
	A667717	YUY
	A667714	XCX
	A667718	WEW
	A667712	SDS
	A667708	QWE
	A667709	QWE
	A667716	QWE
	A667715	CWE
	A667713	B123
	A667711	ASD
	A667710	A323

数据加工

清洗之后可以对数据进行进一步的加工

数据抽取【**提取数据中的部分元素、合并一些数据**】

数据计算【AVERAGE,SUM,MAX,MIN,Date,If】

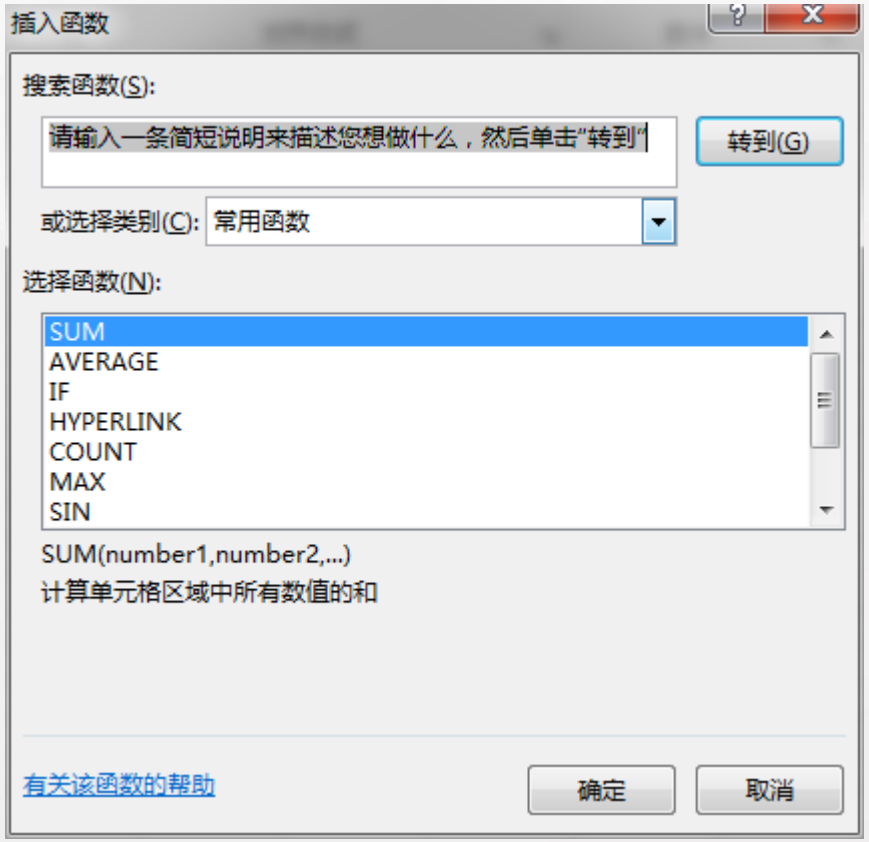
数据分组【VLOOKUP函数，采用近似匹配，SEARCH函数】

数据转换【数据行列变换】

数据抽样【RAND函数，RAND()】

数据计算

- 函数计算



在状态栏看结果

数据计算

- 函数计算
  - 关于时间的操作

日期加减法

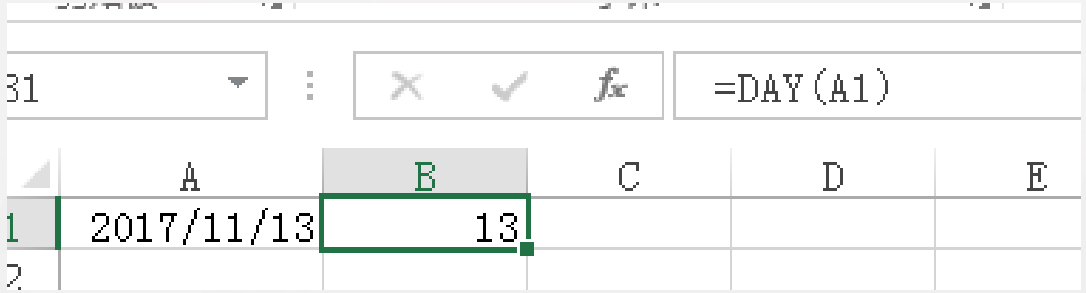
显示	公式	快捷键
2017-11-15	=TODAY()	CTRL+;
11:31		Ctrl+shift+;
2017-11-15 11:31	=NOW()	“Ctrl+;”(分号)，再按空格键，接着按 “Ctrl+shift+;”

DATE(year,month,day):返回表示特定日期

YEAR(serial_number):返回某日期对应的年份

MONTH(serial_number):返回以序列号表示的日期中的月份。用整数1-12表示

DAY(serial_number):返回以序列号表示的日期的天数，用整数1-31表示



报表常常要求时效性，因此对于时间的操作在计算中十分重要！

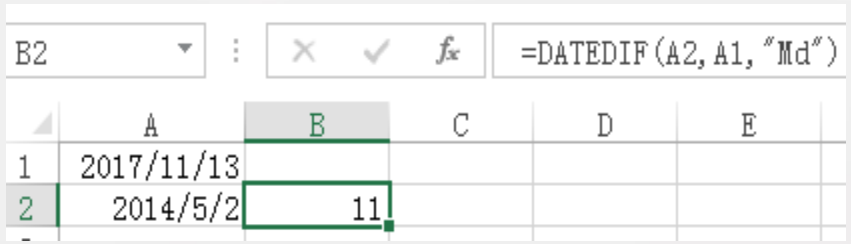
数据计算

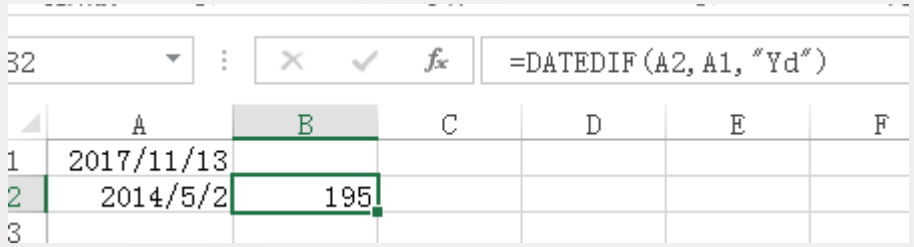
- 函数计算
  - 关于时间的操作

DATEDIF(start_date,end_data,unit):返回两个日期之间的年/月/日间隔数

时间段内的起始天数　**时间段内的结束天数**　**Unit有Y/M/D/MD/YM/YD六种形式**

Y时间段中的整年数，M时间段中的整月数
MD时间段中的天数的差，忽略月和年
YM时间段中的月数的差，忽略日期中的日和年
YD时间段中的天数的差，忽略年





数据加工

清洗之后可以对数据进行进一步的加工

数据抽取【提取数据中的部分元素、合并一些数据】

数据计算【AVERAGE,SUM,MAX,MIN,Date,If】

数据分组【VLOOKUP函数，采用近似匹配，SEARCH函数】

数据转换【数据行列变换】

数据抽样【RAND函数，RAND()】

数据分组

- 采用VLOOKUP函数来实现
- **为什么要分组?**比如我们要已经获得我校职工一年发表SCI论文数的统计表，我想想要看看大家各个区间各有多少人？

姓名	SCI论文篇数
张无忌	2
赵敏	3
周芷若	1
王大锤	4
张三丰	2
刘翔	3
王二狗	12
贤二	23
朱子琪	3
刘若星	2
张宇州	16
夏天	77

B2 =VLOOKUP(A2,$D$2:$F$5,3)

	A	B	C	D	E	F
1	SCI论文篇数	奖励	姓名	阈值	分组	备注
2	2	无	张无忌	0	0-5	无
3	3	无	赵敏	5	1-5篇	院级奖励
4	1	无	周芷若	20	5-20篇	校级奖励
5	4	无	王大锤	20	20篇及以上	省部级奖励
6	2	无	张三丰			
7	3	无	刘翔			
8	12	院级奖励	王二狗			
9	23	省部级奖励	贤二			
10	3	无	朱子琪			
11	2	无	刘若星			
12	16	院级奖励	张宇州			
13	77	省部级奖励	夏天			
14						

数据分组这里使用的是数据的模糊匹配，因此忽略的VLOOKUP的最后一个参数

数据加工

清洗之后可以对数据进行进一步的加工

数据抽取【**提取数据中的部分元素、合并一些数据**】

数据计算【AVERAGE,SUM,MAX,MIN,Date,If】

数据分组【VLOOKUP函数，采用近似匹配，SEARCH函数】

数据转换【数据行列变换】

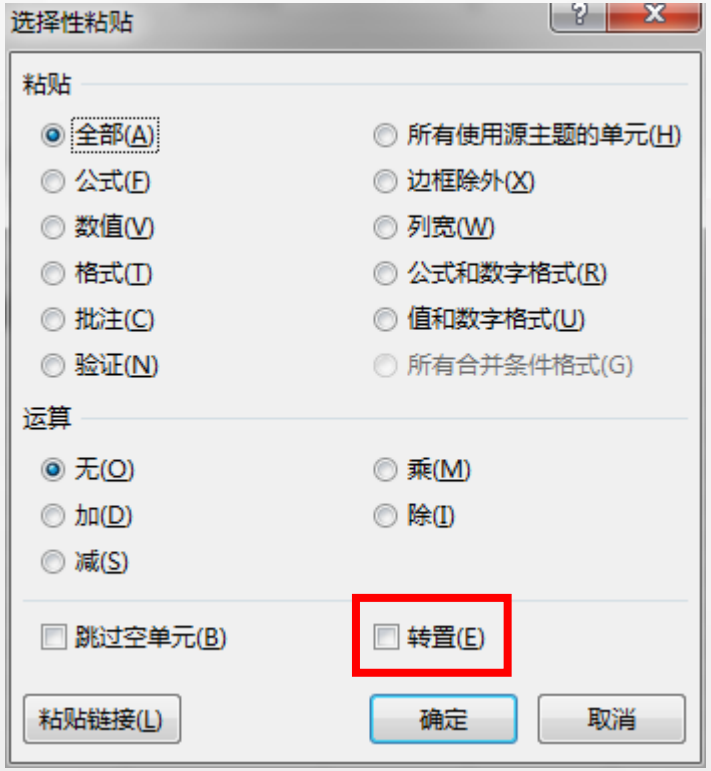
数据抽样【RAND函数，RAND()】

数据转换

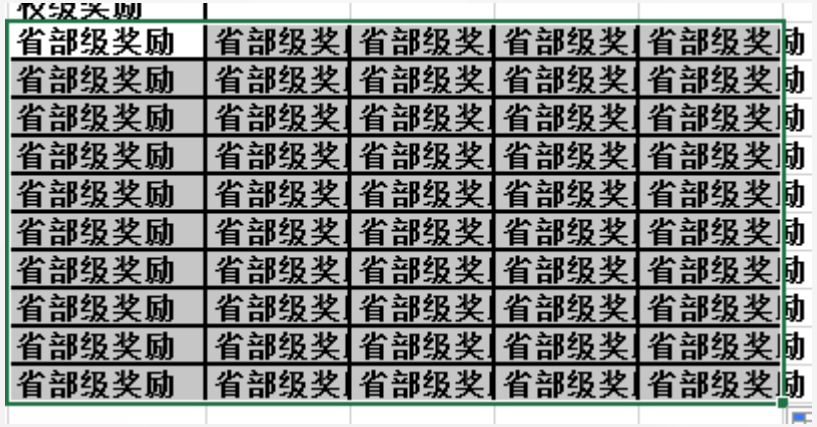
- 数据表的行列互换

1. 选择性粘贴

2. 快捷键Ctrl+Alt+V



3.不用复制的粘贴



数据加工

清洗之后可以对数据进行进一步的加工

数据抽取【Left,Right,CONCATENATE(文本1，文本2，...),VLOOKUP】

数据计算【AVERAGE,SUM,MAX,MIN,Date,If】

数据分组【VLOOKUP函数，采用近似匹配，SEARCH函数】

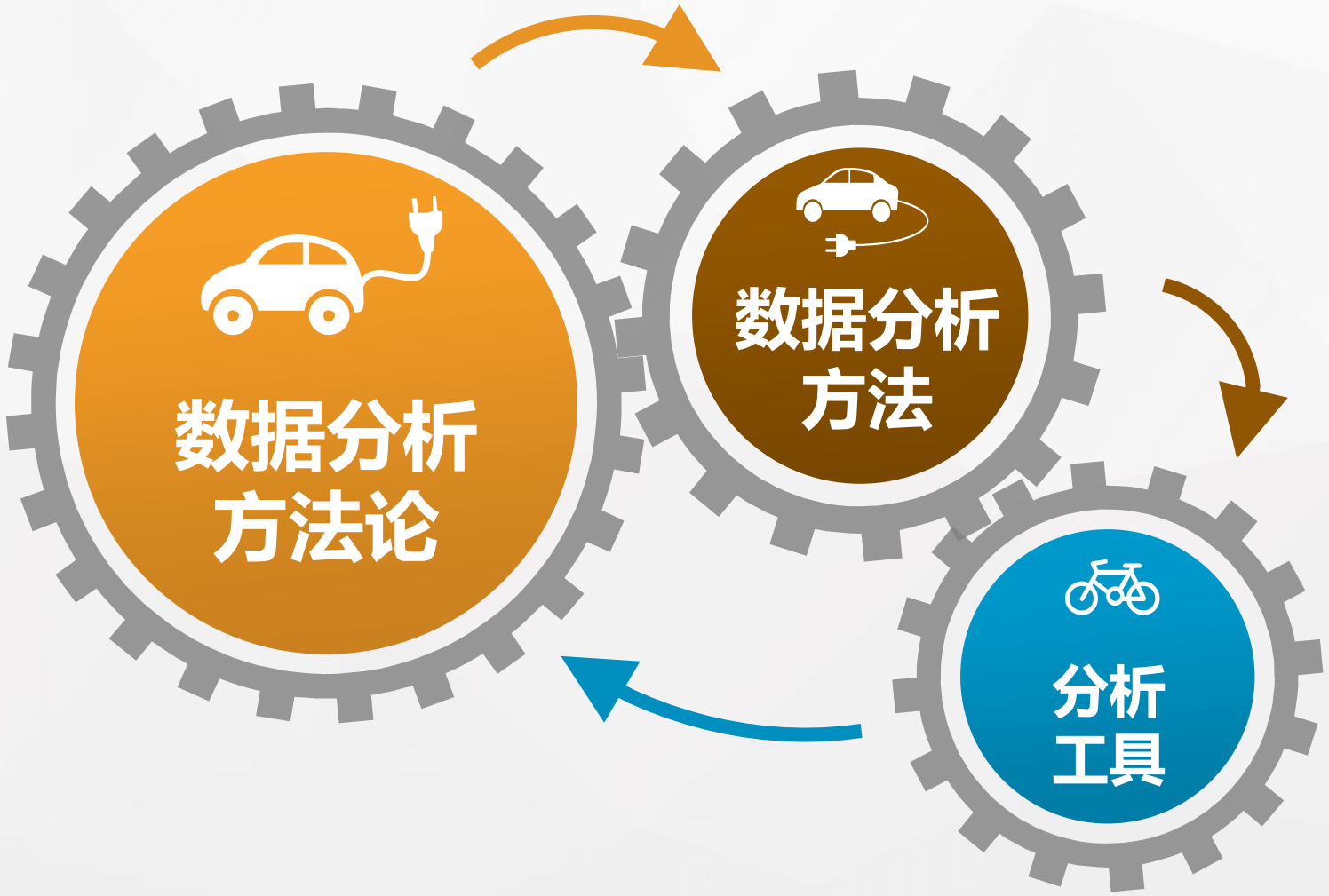
数据转换【数据行列变换】

数据抽样【RAND函数，RAND()】

数据抽样

- 利用RAND()函数
- 返回[0,1]的均匀分布随机数，而且每次计算工作表是都要返回一个新的数值
- 如果在编辑栏中输入“=Rand()”后，保持编辑状态，按<F9>,则生成的随机数将永远保存，不再返回新的数值。
- 如果要生成从a到b之间的随机实数=RAND()*(b-a)+a
- 举个例子

	A	B	C	D	E	F	G	H
	序号	SCI论文篇	奖励	姓名	编号		随机数	姓名
2	1	2	无	张无忌	A667708		1	张无忌
3	2	3	无	赵敏	A667709		7	王二狗
4	3	1	无	周芷若	A667710		10	刘若星
5	4	4	无	王大锤	A667711		8	贤二
6	5	2	无	张三丰	A667712		2	赵敏
7	6	3	无	刘翔	A667713		11	张宇州
8	7	12	院级奖励	王二狗	A667714		10	刘若星
9	8	23	省部级奖励	贤二	A667715		10	刘若星
10	9	3	无	朱子琪	A667716			
11	10	2	无	刘若星	A667717			
12	11	16	院级奖励	张宇州	A667718			
13	12	77	省部级奖励	夏天	A667719			


数据分析
方法论
数据分析
方法
分析
工具

数据分析方法论

从宏观的角度指导如何进行数据分析，比较像数据分析的前期规划，指导后期的数据分析工作如何开展

PEST分析法

政治环境

politics

对组织经营活动具有实际与潜在影响的政治力量和有关的法律、法规等因素

经济环境

economy

GDP的变化发展趋势、利率水平、通货膨胀程度及趋势、失业率、居民可支配收入水平、汇率水平、能源供给成本、市场机制的完善程度、市场需求状况

社会文化环境

society

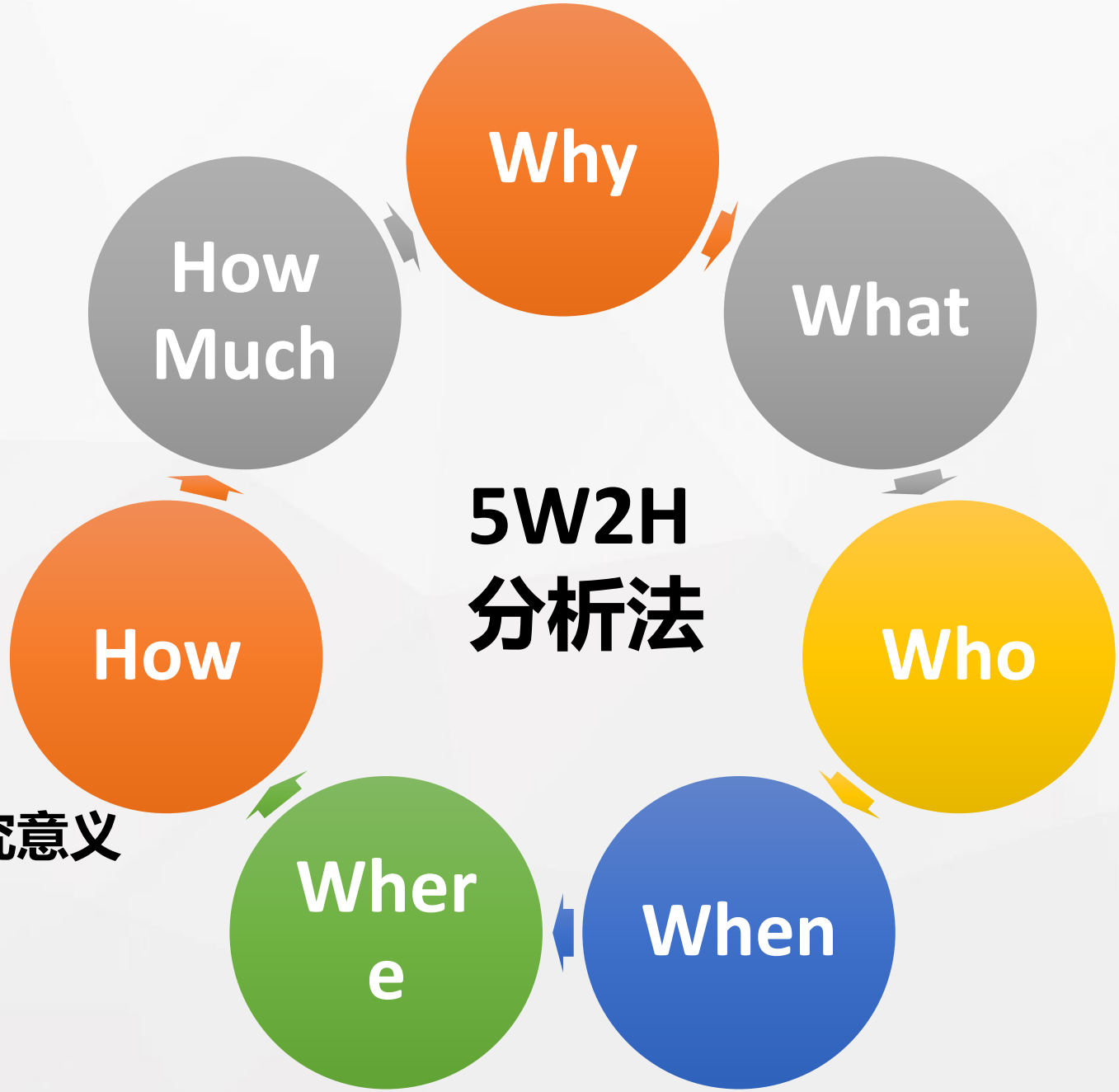
构成社会环境的要素包括人口规模、年龄结构、种族结构、收入分布、消费结构和水平、人口流动性等

技术环境

technology

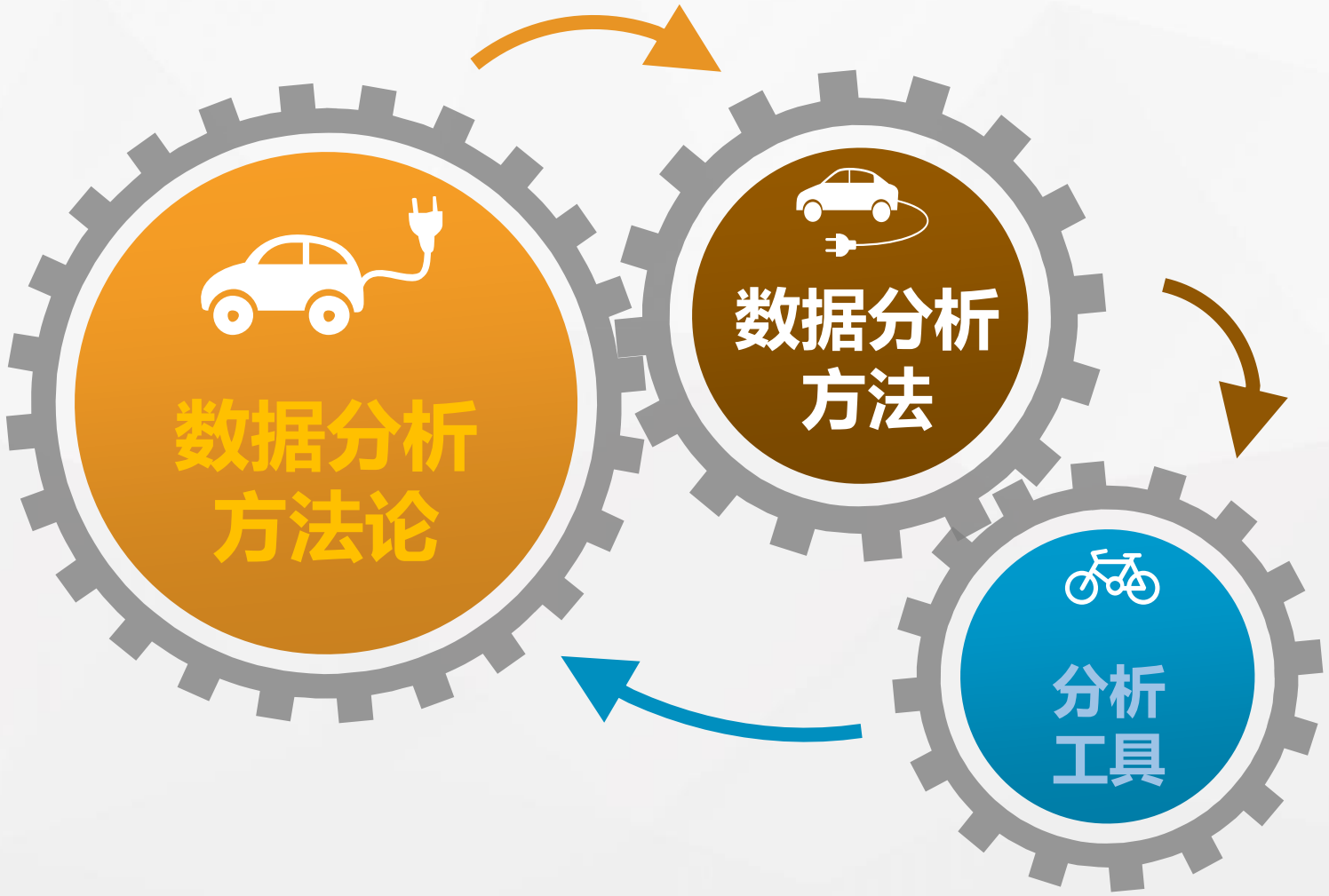
包括那些引起革命性变化的发明，与企业生产有关的新技术、新工艺、新材料的出现和发展趋势以及应用前景

5W2H分析法

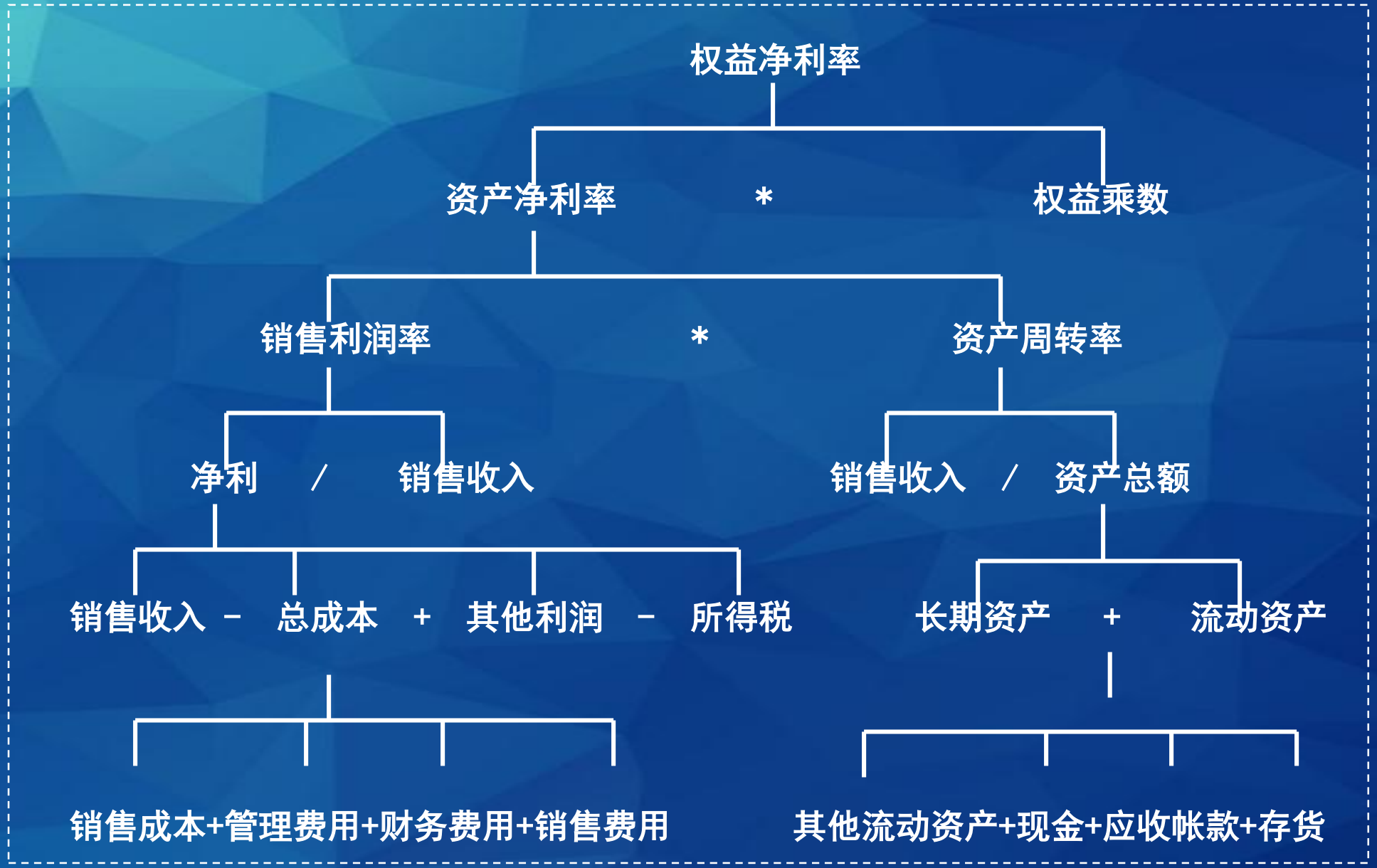


项目课题的研究背景与研究意义

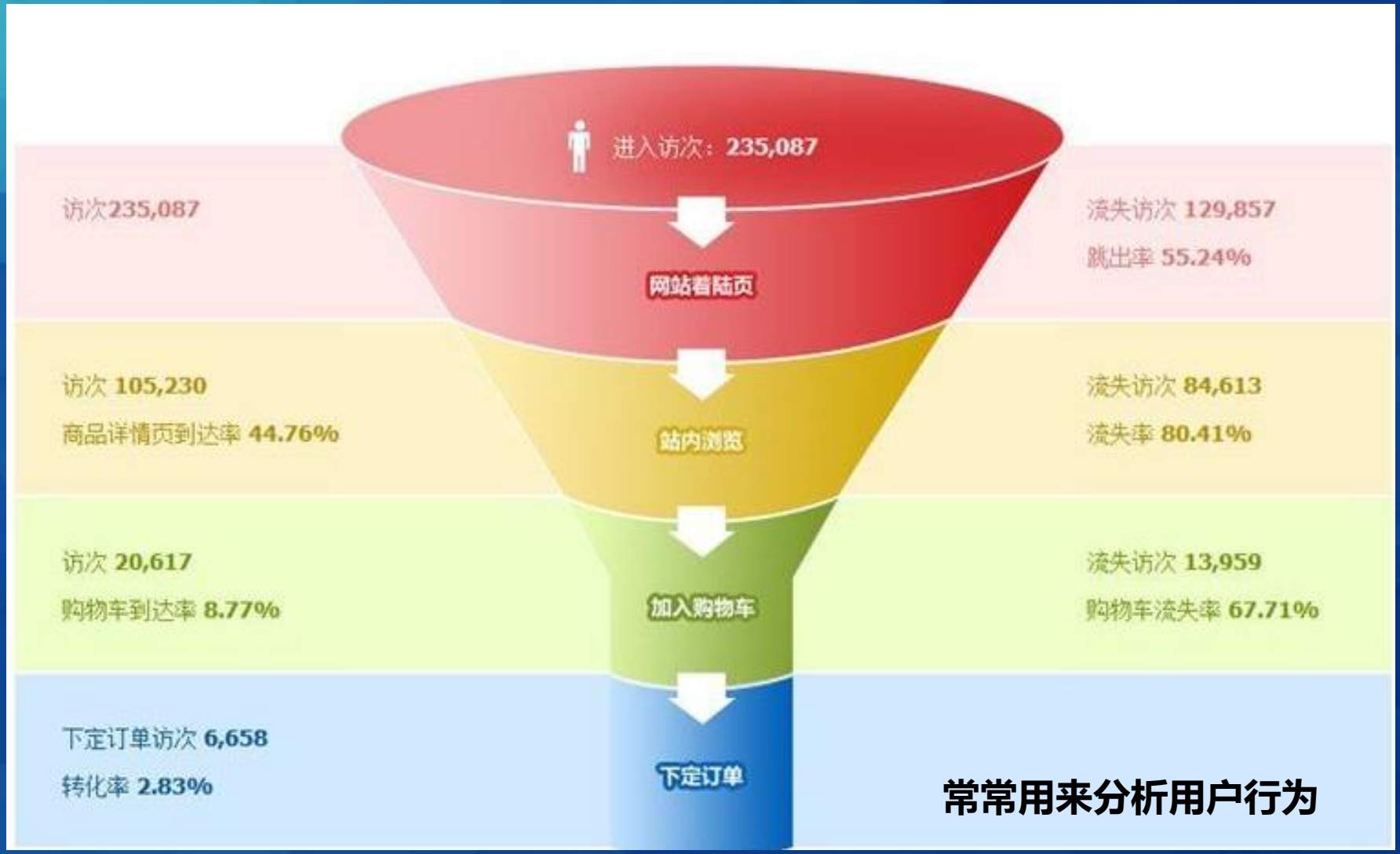
数据分析



杜邦分析法



- 对比分析法
- 分组分析法
- 结构分析法
- 平均分析法
- 交叉分析法
- 综合评价分析法
- 杜邦分析法
- 矩阵关联分析法
- **漏斗图分析法**



常常用来分析用户行为

数据分析工具

数据透视表：有机的综合了数据排序、筛选、分类汇总等数据处理分析功能；同时，也是解决函数公式速度瓶颈的有效手段之一

术语	内容
轴	数据透视表中的一个维度，例如行、列或页
数据源	创建数据透视表的数据表、数据库
字段	数据信息的种类，相当于数据表中的列
字段标题	描述字段内容的标志，可通过拖动字段标题对数据透视表进行透视分析
透视	通过改变一个或多个字段的位置来重新安排数据透视表
汇总函数	EXCEL用来计算表格中数据的值的函数，数值和文本的默认汇总函数分别是求和与计数
刷新	重新计算数据透视表，以反映目前数据源状态

数据分析容易陷入的误区：

1、分析的目标不明确，为了分析而分析

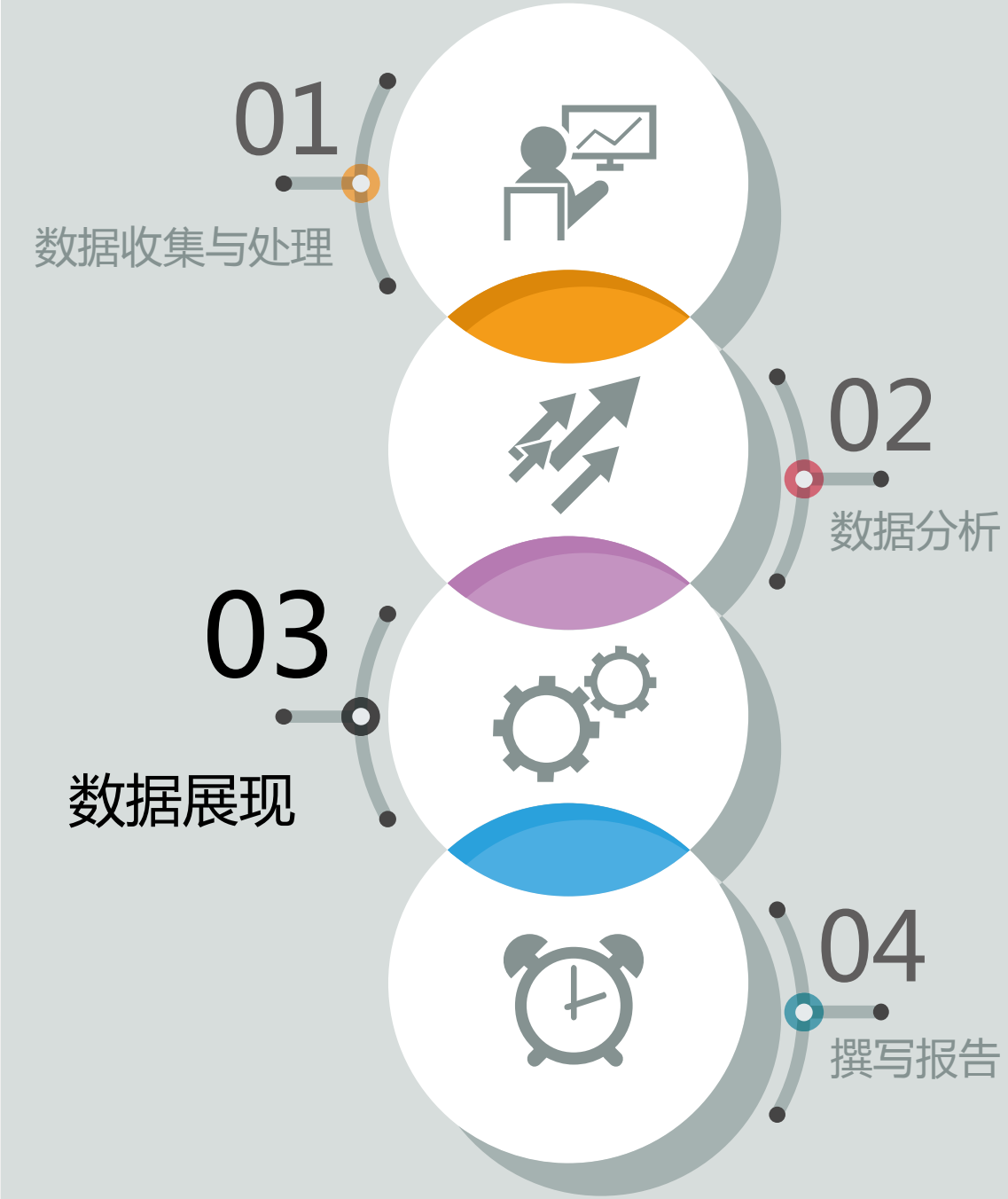
2、缺乏业务知识，分析结果偏离实际

3、一味的追求使用高级分析方法，热衷研究模型

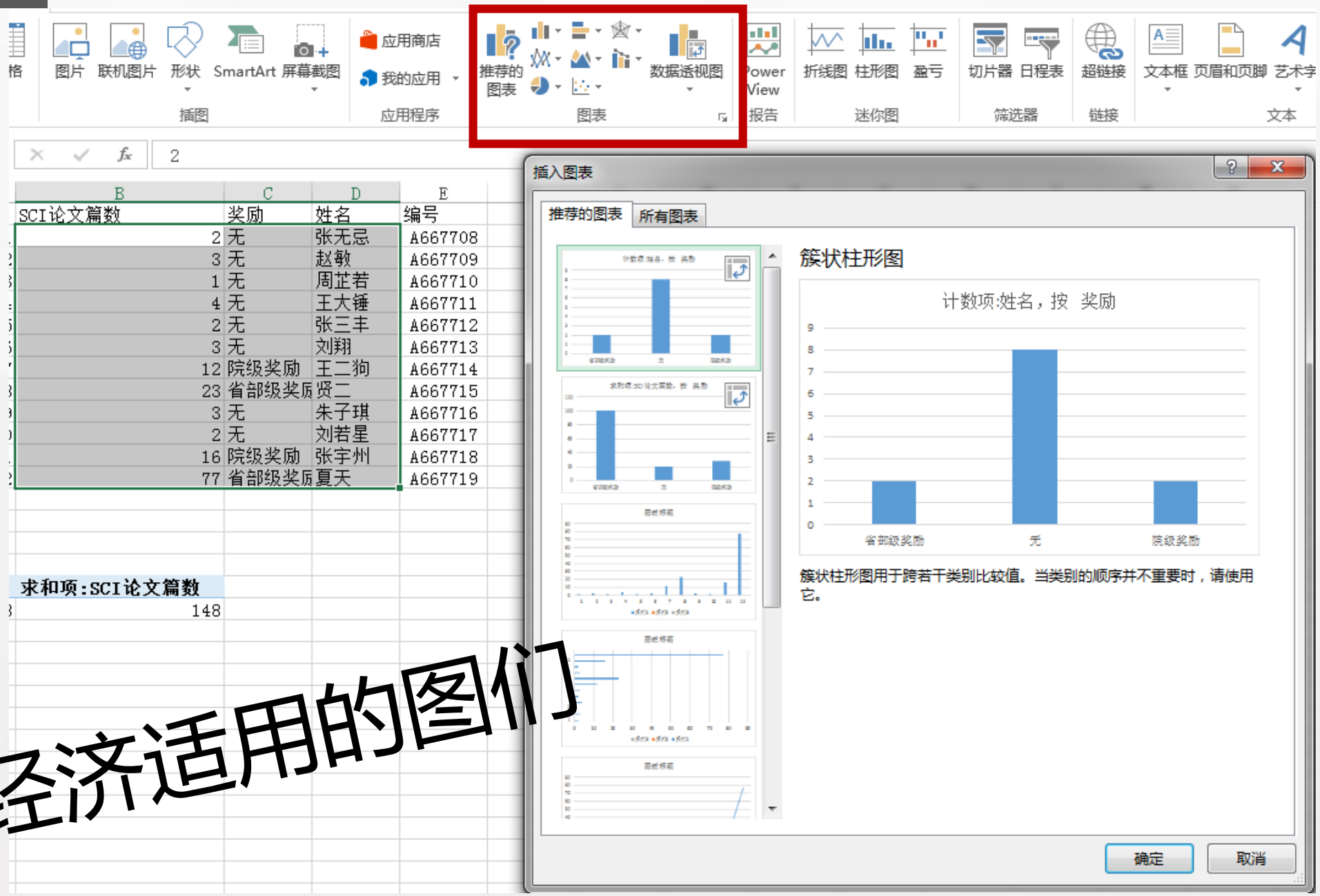
数据分析的背后是数值之间的逻辑
数值之间的逻辑背后是业务的逻辑

一切一切分析的基础是对业务的理解
所有的工具都是技巧

目录



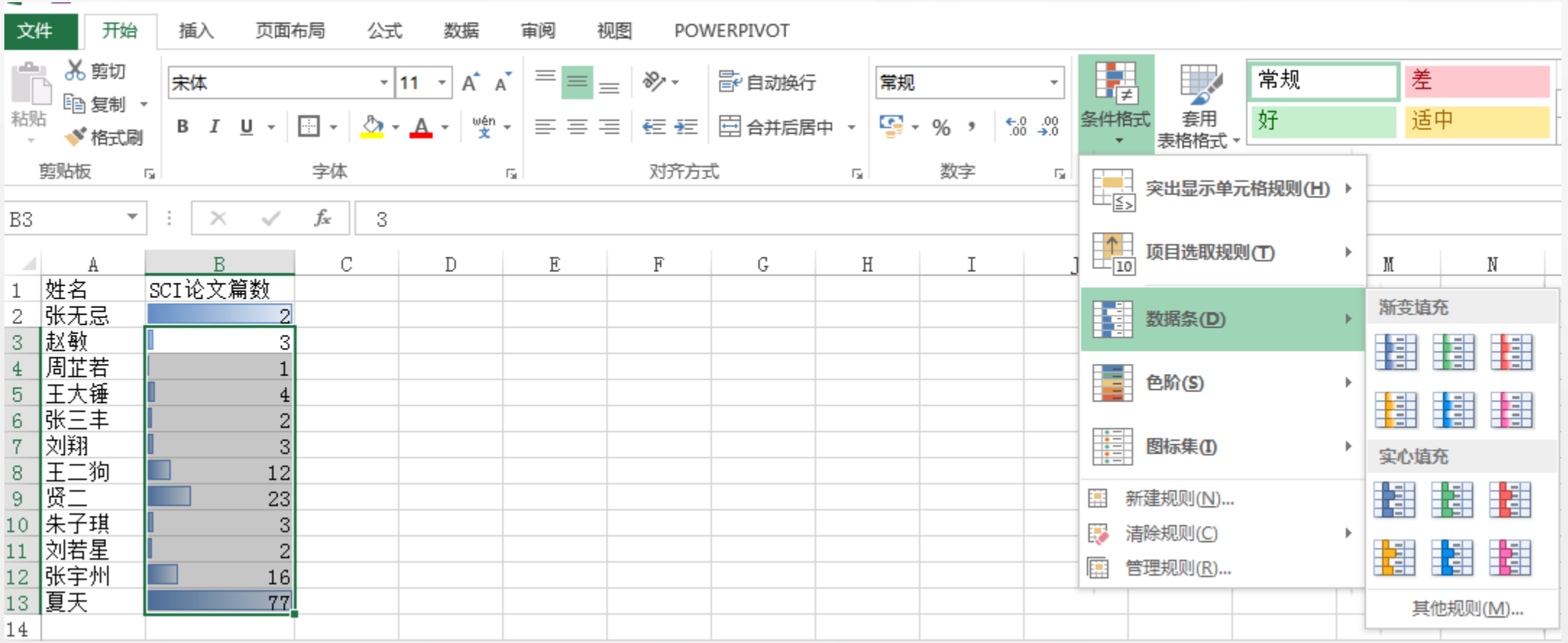
数据展现



经济适用的图们

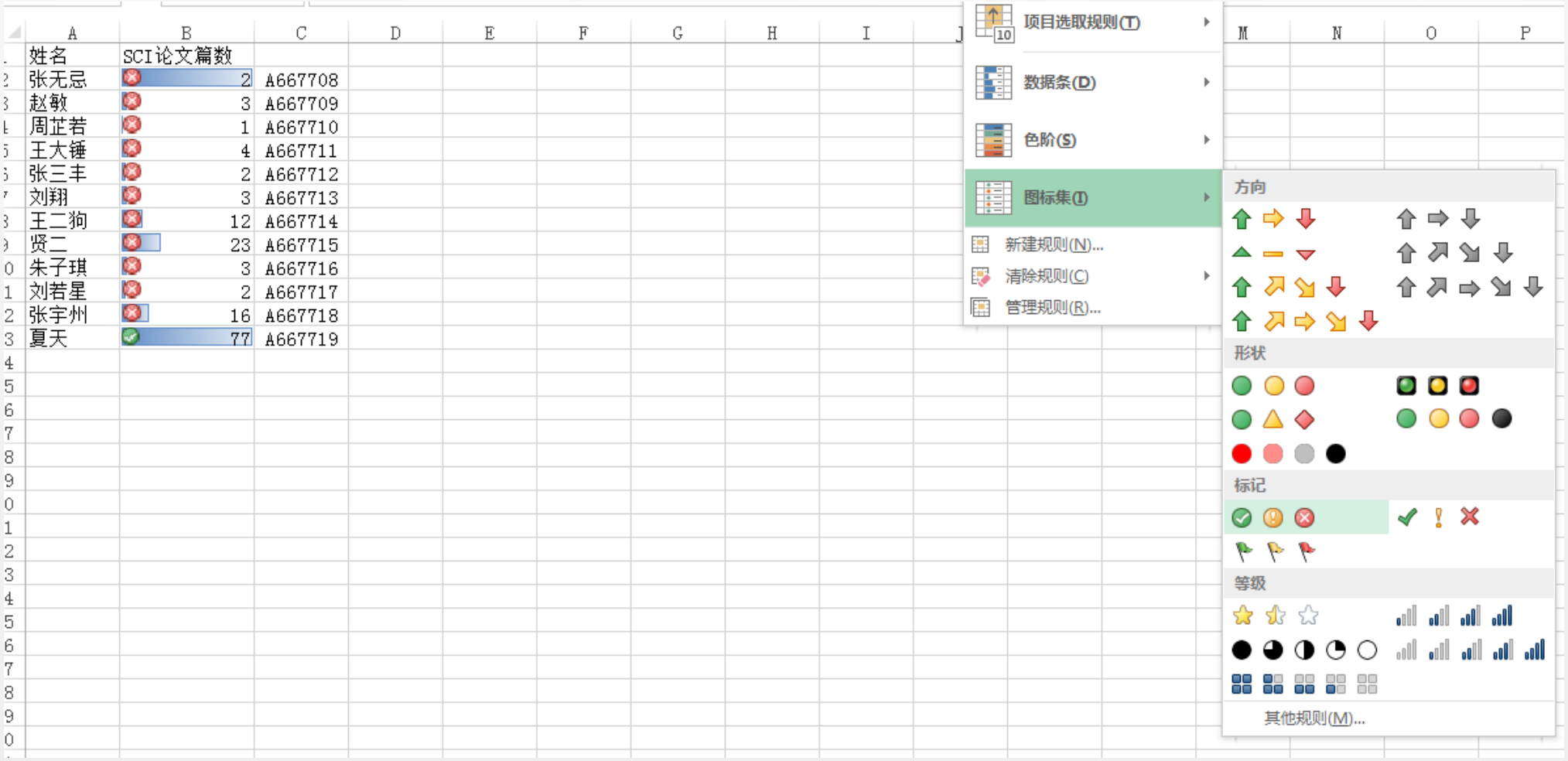
让图标更加有效更加专业的TIPS:

- 突出显示单元格-----筛选功能、条件格式
- 项目选取---条件格式
- 数据条



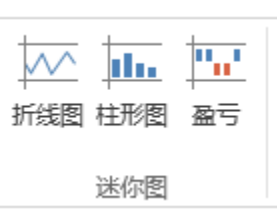
让图标更加有效更加专业的TIPS:

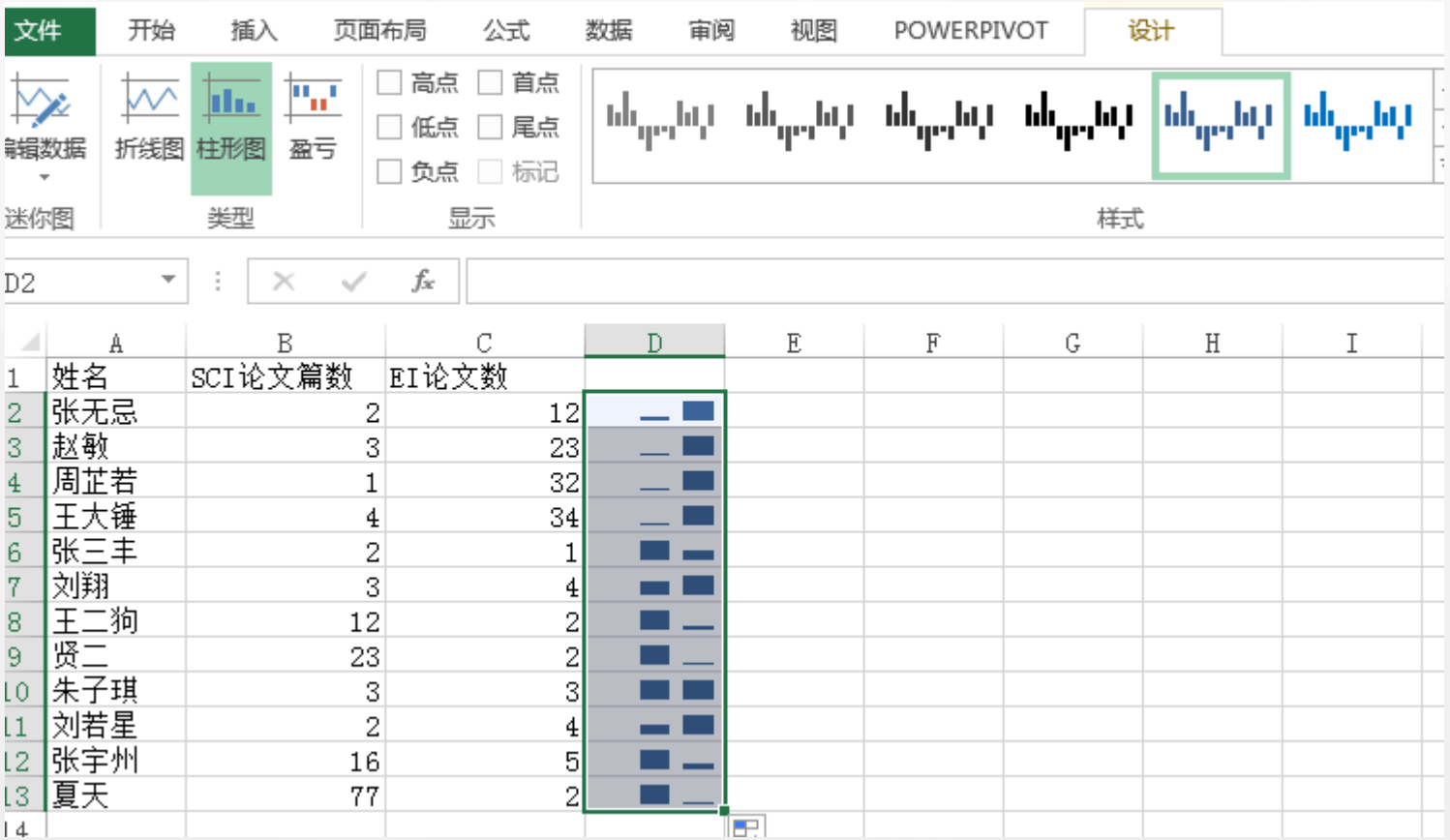
- 图标集



让图标更加有效更加专业的TIPS:

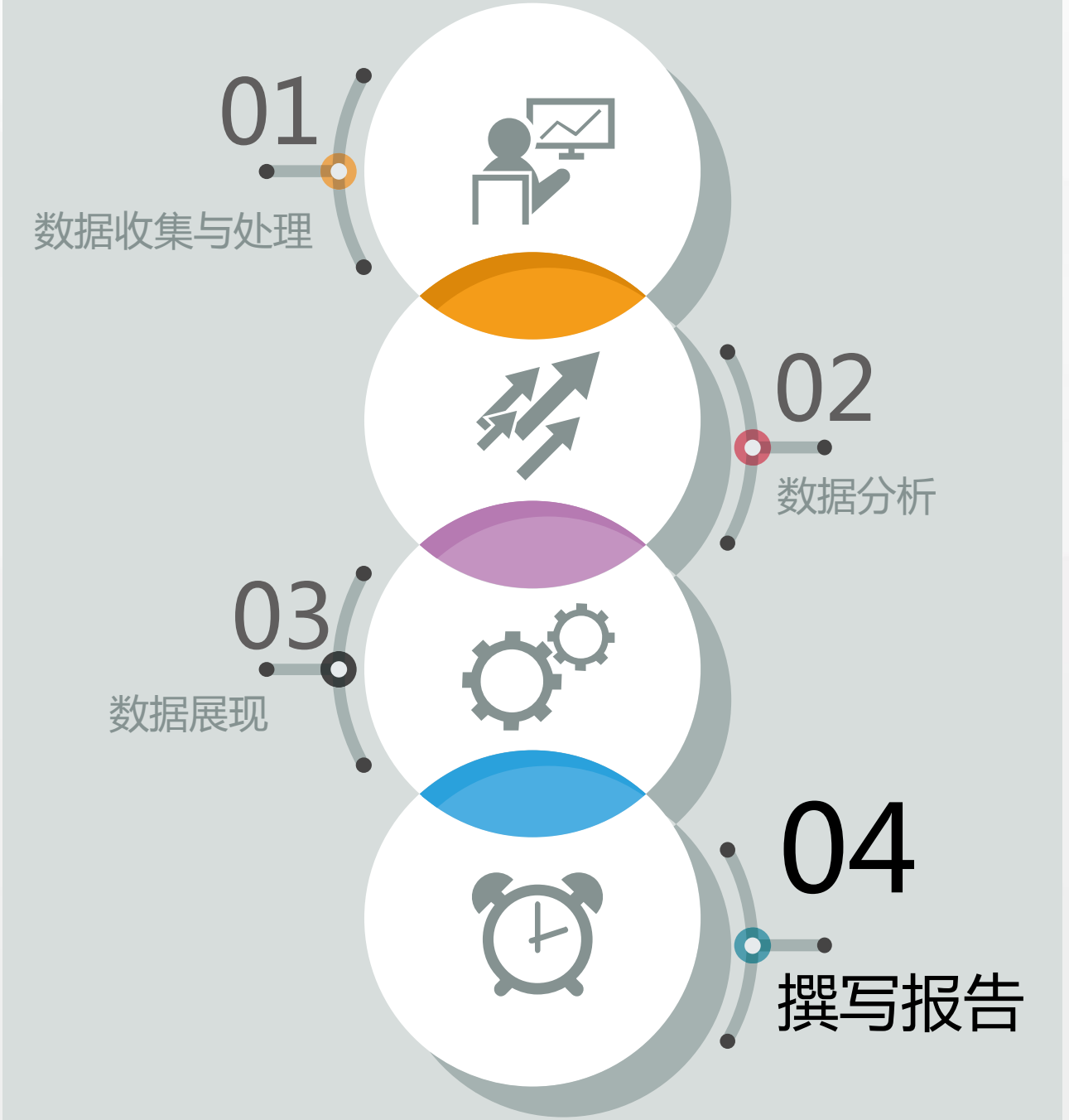
- 迷你图





	A	B	C	D	E	F	G	H	I
1	姓名	SCI论文篇数	EI论文数						
2	张无忌	2	12						
3	赵敏	3	23						
4	周芷若	1	32						
5	王大锤	4	34						
6	张三丰	2	1						
7	刘翔	3	4						
8	王二狗	12	2						
9	贤二	23	2						
10	朱子琪	3	3						
11	刘若星	2	4						
12	张宇州	16	5						
13	夏天	77	2						
14									

目录



数据分析报告的定义

运用数据来反映、研究和分析某项事物的现状、问题、原因、本质和规律，并得出结论，提出解决问题的办法的一种分析应用文体

01 规范性

术语规范，要与业内公认的术语一致
标准统一，前后一致

02 严谨性

数据真实完整
分析过程科学全面
分析结果可靠
内容实事求是

03 创新性

引入新的研究模型或者分析方法，用实际的结果来验证或改良模型

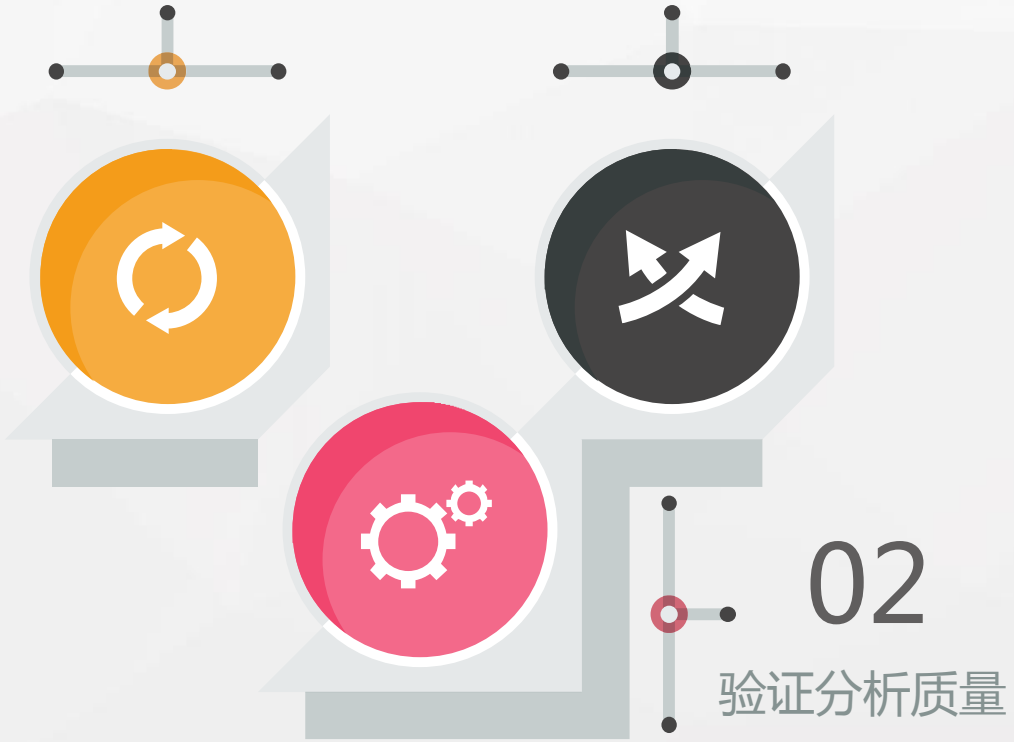
04 重要性

体现数据分析的重点，选取重点的关键指标

数据分析报告的作用

01
展示分析结果

03
提供决策依据



02
验证分析质量

数据分析报告的种类

专题分析报告	综合分析报告	日常数据通报
• 内容的单一性 • 分析的深入性	• 全面性 • 联系性	• 进度性 • 规范性 • 时效性
对某一问题的分析，不要求全面，但是要有一定深度	全面评价一个地区、单位、部门等发展情况	以定期的数据报表为依据

Office各软件制作报告的优劣对比

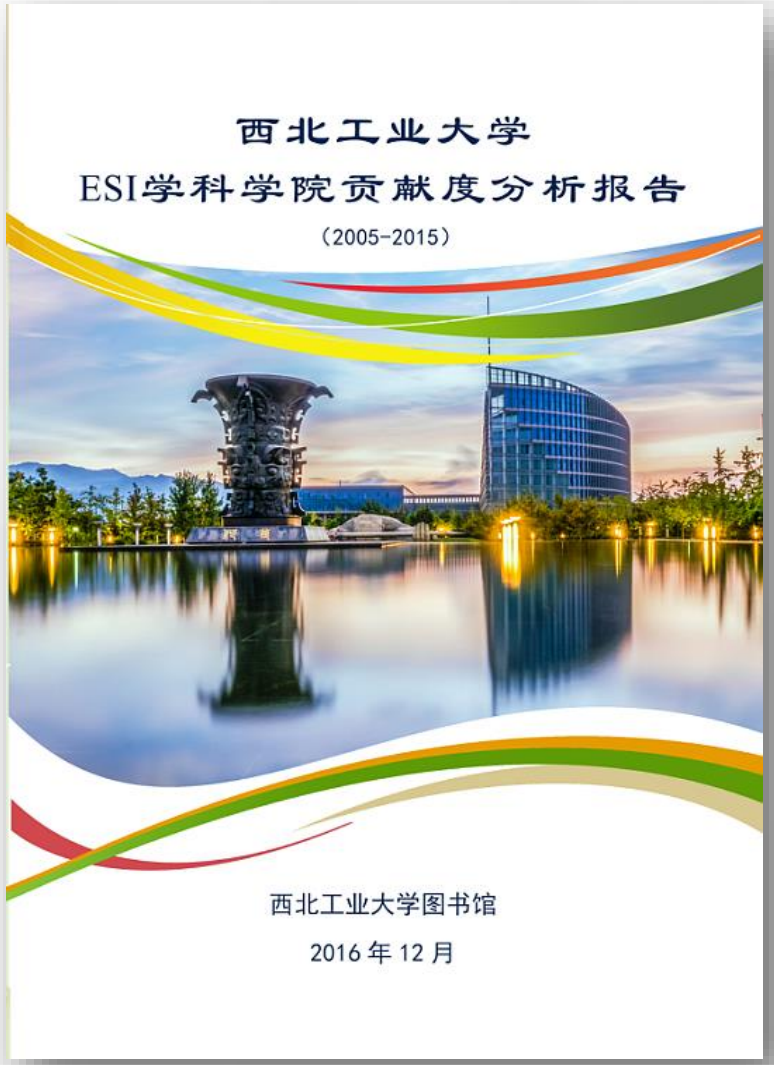
项目			
优势	• 易于排版 • 可打印装订成册	• 可含有动态图表 • 结果可实时更新 • 交互性更强	• 可加入丰富的元素 • 适合演示汇报 • 增强展示效果
劣势	• 缺乏交互性 • 不适合演示汇报	• 不适合演示汇报	• 不适合大幅文字
适用范围	• 综合分析报告 • 专题分析报告 • 日常数据通报	• 日常数据通报	• 综合分析报告 • 专题分析报告

数据分析报告

报告结构

- 标题页
- 目录
- 前言
- 正文
- 结论与建议
- 附录

数据分析报告示例



前 言

为了更好地促进我校“十三五”发展规划和“双一流”建设，分析我校学科发展状况，以及各学院对 ESI 学科的贡献情况，为学院和学校相关职能部门提供数据参考，图书馆特组织信息咨询与发展研究部和信息技术部的工作人员编制了本报告。

基本科学指标数据库（Essential Science Indicators，简称 ESI）是由世界著名的学术信息出版机构美国科技信息所(ISI)于 2001 年推出的衡量科学研究绩效、跟踪科学发展趋势的基本分析评价工具，ESI 已成为当今世界范围内普遍用以评价高校、学术机构、国家/地区国际学术水平及影响力的重要评价指标工具之一。

本报告统计了 2005—2015 年西北工业大学各学院在 SCIE/SSCI 引文数据库中的论文收录和被引情况，分析了论文的中科院分区、国内外合作、涉及 ESI 学科及学院分布等指标的分布情况，并根据 ESI 的学科分类，从论文数量和被引频次两个方面分析了我校已进入全球前 1%的 3 个学科的学院贡献度，最后对我校未来可能进入 ESI 全球前 1%的学科进行了预测分析。报告揭示了各学院在 SCIE/SSCI 上的学术产出、学术影响力和发展趋势，以及对 ESI 学科的贡献情况，供学院和学校相关职能部门参考。希望各学院和相关职能部门对报告的内容和框架提出宝贵意见和建议，以便我们对报告不断进行完善，进一步满足学校“双一流”建设的需要。

在本次数据统计分析过程中，我们发现，我校师生发表的国际学术论文中论文作者及其机构的署名不规范，给我们的统计工作带来了诸多困难，也影响到最终统计结果的准确率。因此，我们向广大师生提出以下建议：

1. 作者姓名的英文署名应使用全拼形式，且姓在前、名在后。
   例如：张大伟可写作 zhang dawei 或者 zhang da-wei
2. 作者机构的英文署名应包含二级单位即作者所在学院、部门或实验室的名称，且使用全拼形式，二级单位的名称写在学校名称之前。
   例如：航空学院：School of Aeronautics，Northwestern Polytechnical University。
3. 尽量提供通讯作者的邮箱地址。

由于学术水平和分析能力有限，本报告若有疏漏和不妥之处，敬请批评指正。

西北工业大学图书馆
2016 年 12 月

目 录

数据分析学习推荐

学习推荐

数据分析类论坛

① 中国统计网 www.itongji.cn

② EXCEL Home http://www.excelhome.net/

③ Excel技巧网 http://www.exceltip.net/forum.php

学习推荐

数据分析公众号

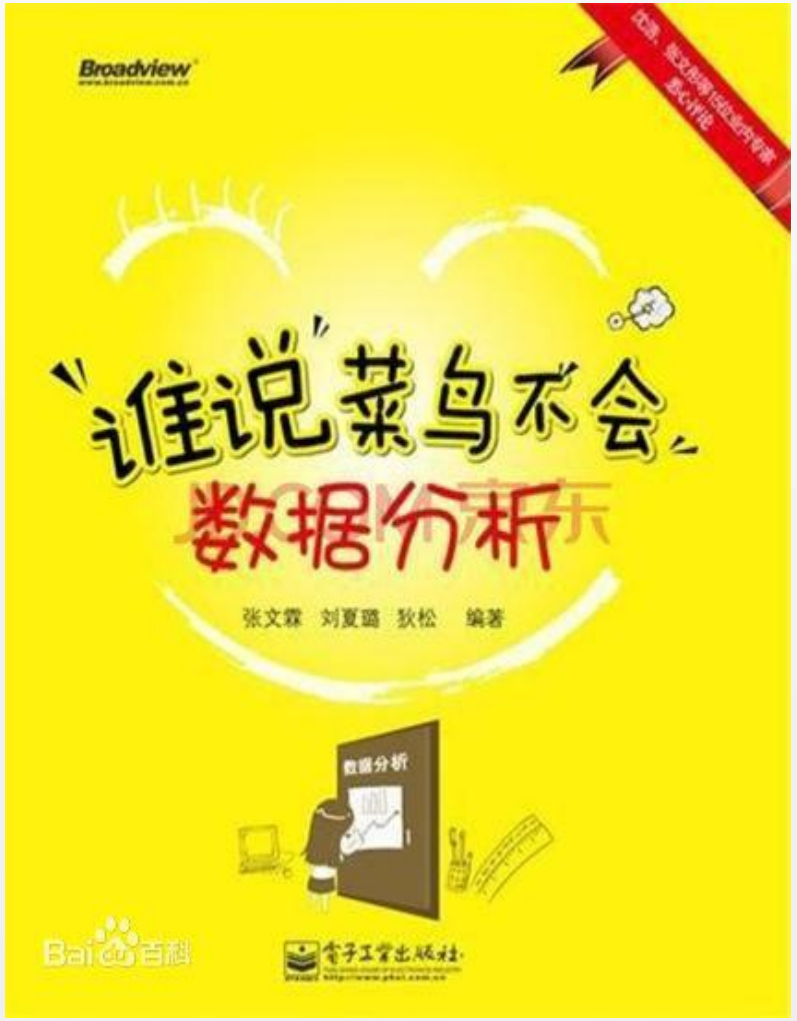
微信公众号：artofdata
沈浩老师 中国传媒大学教授
微信公众号中分享了大量数据处理、分析方法、工具心得。比较偏重理论学习

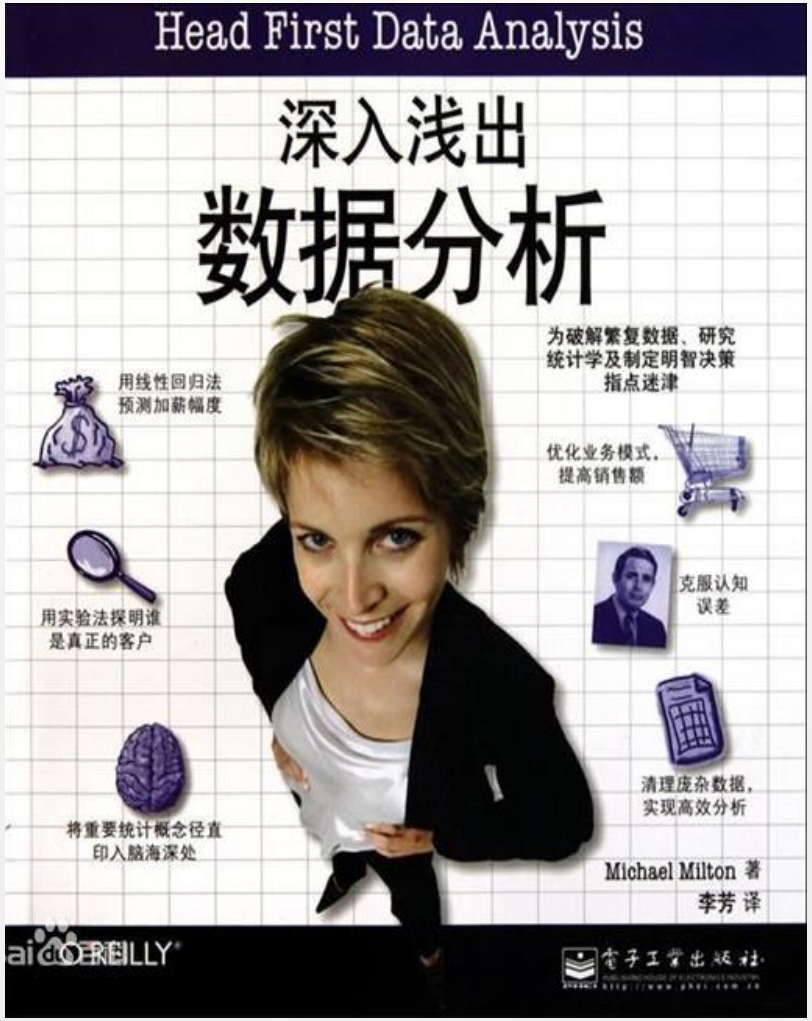
微信号: excelpx-tete
每天一篇excel原创教程，由浅入深，全面学习excel技巧、函数、图表和VBA编程。有excel问题也可以提问哦！

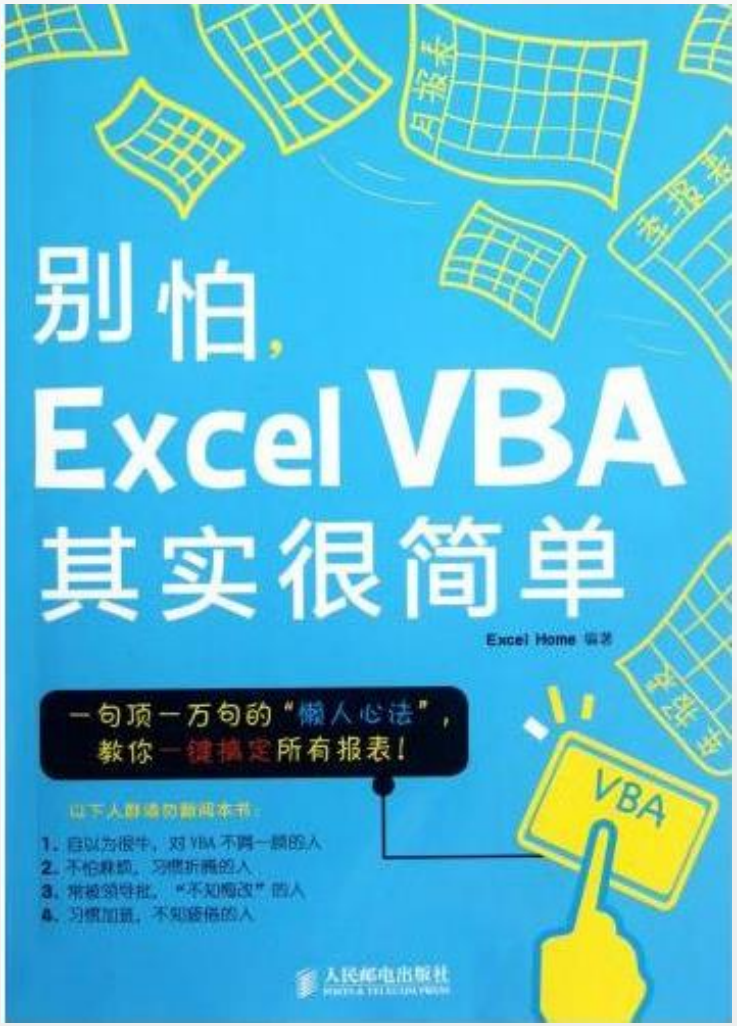
微信号：iexcelhome
ExcelHome每日分享excel操作技巧、excel函数公式、透视表、excel图表和VBA教程，助您轻松提高办公效率，搞定数据分析!

学习推荐

推荐几本书


Broadview
谁说菜鸟不会数据分析
张文霖 刘夏璐 狄松 编著
电子工业出版社


Head First Data Analysis
深入浅出
数据分析
为破解繁复数据、研究统计学及制定明智决策指点迷津
用线性回归法预测加薪幅度
优化业务模式，提高销售额
克服认知误差
用实验法探明谁是真正的客户
清理庞杂数据，实现高效分析
将重要统计概念径直印入脑海深处
Michael Milton 著
李芳 译
O'REILLY
电子工业出版社


别怕，Excel VBA 其实很简单
Excel Home 编著
一句顶一万句的“懒人心法”，教你一键搞定所有报表！
VBA
人民邮电出版社

谢谢
Q&A?

西北工业图书馆微信公众平台