# Excel 数据分析与可视化

## 本章概要

第

3

音

Excel 提供了强大的数据分析与可视化功能,可以利用数据分析工具来求解 复杂问题,并且利用图表进行数据可视化。本章主要学习数据分析的基础知识以 及常用的数据分析方法,以解决实际问题,并实现数据可视化。

## 学习目标

通过本章的学习,要求达到以下目标。

- (1) 掌握单变量模拟运算表和双变量模拟运算表两种模拟运算表。
- (2) 掌握单变量求解。
- (3) 掌握利用数据分析工具进行数据分析和预测。
- (4) 掌握利用图表进行数据可视化的方法。

## ◆ 3.1 数据分析

Excel 中提供了多种数据分析工具,可以方便、快速地解决较复杂的数据分析问题。例如,希望分析目标结果受一个或两个变量如何影响可以使用模拟运算表;已知目标结果,希望求解变量的值可以使用单变量求解;对历史数据进行预测可以使用移动平均或者指数平滑;希望分析两个变量之间的关系可以使用回归分析。本节将主要介绍 Excel 的数据分析高级应用。

## 3.1.1 模拟运算表

目标单元格是包含一个或多个单元格引用(变量)的公式或函数,模拟运算表 可以分析计算当公式中的一个或两个变量变化时,目标单元格值的变化。

模拟运算表有两种类型:单变量模拟运算表和双变量模拟运算表。单变量模 拟运算表中,用户可以对一个变量赋予不同的值,从而查看目标结果如何变化。 双变量模拟运算表中,用户可以对两个变量赋予不同的值,从而查看目标结果如 何变化。

例 3-1: 打开"投资计划表.xlsx",用模拟运算表制作一张投资计划表,当每月 可投资金额占月收入百分比变化时(15%,20%,25%,30%),计算投资的收益情况,计算结果为货币格式并保留两位小数,如图 3-1 所示。

	A	В	С	D	E
1	张先生投资计划表				
2	月收入	¥30,000.00			
3	每月可投资金额占比	20%			
4	利率(年)	3.0%			
5	期限(年)	1			
6					
7		投资比例改	变		
8		15%	20%	25%	30%
9	¥72,998.30	¥54,748.72	¥72,998.30	¥91,247.87	¥109,497.44

图 3-1 单变量模拟运算表样张

## 解答:

(1) 在 B8:E8 中创建模拟运算行参数单元格区域:在 B8 中输入 15%,C8 中输入 20%,自动填充到 E8。

(2) 创建目标公式: 在 A9 中输入公式: =FV(B4/12,B5 \* 12,B2 \* B3)。FV(rate, nper,pmt,[pv],[type])是财务函数,主要作用是在基于固定利率及等额分期付款方式下, 计算某项投资的未来值。rate 是各期利率(一个月为一期)。nper 是投资总期数。pmt 是 各期所投资的金额,在整个投资期间保持不变。pv 是一系列投资的当前值的累积和,即从 该项投资开始计算时已经入账的款项,如果省略 pv,则表示 pv 为 0。type 用以指定各期的 付款时间是在期初(取值 1) 还是期末(取值 0),如果省略 type,则表示 type 为 0。

(3)进行模拟运算:选择模拟运算区域 A8:E9,单击"数据"→"模拟分析"→"模拟运算 表",在"输入引用行的单元格"中指定单元格\$B\$3(投资占比),如图 3-2 所示。

	A	В	С	D	E	F	G	Н	I	J
1	张先生投资计划表									
2	月收入	¥30,000.00								
3	每月可投资金额占比	20%				模拟运	算表		-2	×
4	利率(年)	3.0%				输入引用行的前元格(R): \$B\$3 篇			E.C.	
5	期限(年)	1						55		
6						-163/(3	1/13/94342/6	na( <u>c</u> ).		<u>1986</u>
7		投资比例改	变			<b>施</b> 完 取消				
8		15%	20%	25%	30%	明定 取消				
9	¥72,998.30								1	

图 3-2 单变量模拟运算设置

(4) 选中 A9:E9, 右击, 选择"设置单元格格式"命令, 在"货币"中设置货币符号¥、小数 位数 2。

例 3-2: 打开"投资计划表.xlsx",用模拟变量表制作一张投资计划表,计算不同投资占比(15%,20%,25%,30%)和不同期限(1年,2年,3年,4年,5年)的投资收益情况,计算结果为货币格式并保留两位小数,如图 3-3 所示。

	А	В	С	D	E
1	张先生投资计划	表			
2	月收入	¥30,000.00			
3	每月可投资金额占比	20%			
4	利率(年)	3.0%			
5	期限(年)	1			
6					
12		投资比例和期降	艮改变		
13	¥72,998.30	15%	20%	25%	30%
14	1	¥54,748.72	¥72,998.30	¥91,247.87	¥109,497.44
15	2	¥111,162.68	¥148,216.91	¥185,271.13	¥222,325.36
16	3	¥169,292.52	¥225,723.36	¥282,154.20	¥338,585.04
17	4	¥229,190.44	¥305,587.25	¥381,984.06	¥458,380.88
18	5	¥290,910.21	¥387,880.28	¥484,850.34	¥581,820.41

图 3-3 双变量模拟运算表样张

(1) 在 B13:E13 和 A14:A18 分别创建模拟运算行参数和列参数单元格区域。

(2) 创建目标公式:在 A13 中输入公式:=FV(B4/12,B5\*12,B2\*B3)。

模拟运算表	? ×
输入引用行的单元格( <u>R</u> ):	\$B\$3
输入引用列的单元格( <u>C</u> ):	\$B\$5
确定	取消
图 3-4 双变量樽	

(3)进行模拟运算:选择模拟运算区域 A13:E18,单 击"数据"→"模拟分析"→"模拟运算表",在"输入引用行 的单元格"中指定单元格\$B\$3(投资占比),"输入引用列的 单元格"文本框指定单元格\$B\$5(期限),如图 3-4 所示。

(4) 选中 A13 和 B14:E18,右击,选择"设置单元格格 式"命令,在"货币"中设置货币符号¥、小数位数2。

## 3.1.2 单变量求解

目标单元格是包含一个或多个单元格引用(变量)的公式或函数,如果已知目标单元格 的预期结果值,单变量求解可以计算出公式中某个变量的合适取值。例如,单变量求解经常 用于一元方程的求解,已知方程 y=f(x)和 y 的值,求解 x 是多少。

**例 3-3**:打开"单变量求解.xlsx",在"单变量求解 1"工作表的 B2 单元格中给出方程 6x<sup>4</sup>+2x<sup>3</sup>-3x<sup>2</sup>-3x=2的一个解,如图 3-5 所示。

#### 解答:

(1) 创建目标公式。在 B1 中输入公式: =6 \* B2^4+2 \* B2^3-3 \* B2^2-3 \* B2。

(2)进行单变量求解。选择 B1,单击"数据"→"模拟分析"→"单变量求解","目标单元格"中指定单元格 B1(目标公式),"目标值"为 2(目标结果),"可变单元格"指定单元格\$B\$2
(变量),如图 3-6 所示。单击"确定"按钮,在 B2 单元格中显示方程的一个解。注意:一元多次方程的解不唯一,单变量求解只能返回一个解。



单变量求解	? ×			
目标单元格(E):	B1			
目标值( <u>V</u> ):	2			
可变单元格( <u>C</u> ):	\$B\$2			
确定	取消			
图 3-6 单变量求解设置 1				

例 3-4:2019 年中国人口为 14 亿,如果控制人口在 2025 年不超过 15 亿,人口的年增长率应该控制为多少?打开"单变量求解.xlsx",在"单变量求解 2"工作表中计算人口年增长率,如图 3-7 所示。

## 解答:

(1) 根据题目要求,在 B2、B4 单元格中分别输入 2019 年人口和增长年数。

(2) 创建目标公式:在 B5 中输入公式:=B2\*(1+B3)^B4。

(3)进行单变量求解:选择 B5,单击"数据"→"模拟分析"→"单变量求解","目标单元格"中指定单元格 B5(目标公式),"目标值"为 15(目标结果),"可变单元格"指定单元格 \$B\$3(变量),如图 3-8 所示。单击"确定"按钮,在 B3 单元格中显示结果,并设置单元格格式 为百分比,保留两位小数。

	A	В	С
1	人口控制		
2	2019年人口	14	亿
3	人口年增长率	1.16%	
4	增长年数	6	年
5	2025年预计人口	14. 99998434	亿

单变量求解		? ×
目标单元格(E):	B5	
目标值( <u>V</u> ):	15	
可变单元格( <u>C</u> ):	\$B\$3	<b>5</b>
确定		取消

图 3-7 单变量求解人口年增长率样张

图 3-8 单变量求解设置 2

## 3.1.3 时间序列预测分析

预测(Forecast)是指用科学的方法预计、推断事物发展的必然性或可能性,即根据过去和现在预计未来。预测分析是一种常见的数据分析方法,可以根据历史数据(包括过去的和现在的),建立预测模型,应用预测分析工具,对预测对象的未来结果或趋势进行预测,从而减少对未来事物认识的不确定性,帮助人们制定决策。

## 1. 时间序列

时间序列是按事件发生的先后顺序排列起来的一组观察值或记录值。最早的时间序列 分析可以追溯到 7000 年前的古埃及。古埃及人把尼罗河涨落的情况逐天记录下来,构成一 个时间序列。通过对时间序列长期的观察,他们发现了尼罗河涨落的规律,从而提前预测尼 罗河的涨落,安排农业生产,由此古埃及的农业迅速发展。中国的二十四节气是古代劳动人 民通过观察太阳周年运动而形成的时间序列,图 3-9 展示了 2003—2008 年全国平均地表气 温的二十四节气时间序列图。



图 3-9 2003-2008 年全国平均地表气温的二十四节气时间序列图

时间序列中的时间可以是年份、季度、月份或其他任何时间形式。时间序列含有不同的成分,如趋势(如图 3-10 所示)、季节性(如图 3-11 所示)、周期性(如图 3-12 所示)和随机性,如表 3-1 所示。



2/1 7/1 12/1 5/1 10/1 3/1 8/1 1/1 6/1 11/1 4/1 9/1 2/1 7/1 12/1 5/1 10/1 3/1 图 3-10 纸黄金价格走势图



图 3-11 某商品利润图

表 3-1 时间序列成分及特点

成分	特点	示 例
趋势性	时间序列在长时期内呈现出来的某种持续上升或持 续下降的变动 时间序列中的趋势可以是线性和非线性	纸黄金价格、产品销售、国内生产 总值、股价
季节性	时间序列在短期内(一年内)重复出现的周期波动 旅游旺季、旅游淡季;销售旺季、销售淡季	气温、空气质量、降水量、旅客人数、季节性产品销售(冷饮、羽绒服等)、传染病传播

-			
•			
	а,	а.	a -

		续表
成分	特点	示 例
周期性	时间序列中呈现出来的围绕长期趋势的一种波浪形 或振荡式变动 一般周期在 2~15 年,循环的幅度和周期都不规则	太阳黑子、传染病流行
随机性/不 规则性	偶然性因素对时间序列产生影响,致使时间序列呈现 出某种随机波动,在时间序列中无法预计	股票市场中突然出现的利好或利 空消息使股价产生的波动



### 2. 时间序列预测分析方法

一般认为,事物的过去趋势会延伸到未来,实际数据的时间序列可以反映变量在一定时 期内的发展变化趋势与规律,因此可以从时间序列中找出变量变化的特征、趋势以及发展规 律,从而对变量的未来变化进行有效的预测。例如,企业记录了某商品第一个月,第二个 月,……,第 N 个月的销售量,利用时间序列分析方法,可以对未来各月的销售量进行预测。

时间序列预测分析是一种常见的预测分析方法,根据已有的历史数据对未来进行预测, 时间序列中的数据点越多,所产生的预测就越准确。

预测分析技术广泛应用于社会的方方面面,为金融、电信、医疗、零售业、制造业等众多 领域的决策者提供决策支持。例如,在金融领域,人们可以使用预测分析技术预测金融市场 趋势;零售业可以使用预测分析技术来预测库存数量,为优化管理库存提供帮助;在制造业 领域,企业可以使用预测分析技术预测生产需求,从而控制成本、增加利润。

常用的时间序列预测的主要方法有:移动平均法、指数平滑预测法、季节变动预测法、 自回归移动平均等。在 Excel 数据分析工具中集成了移动平均和指数平滑两种分析工具, 可以对数据进行时间序列预测分析。

## 3.1.4 移动平均

移动平均法是一种常用的时间序列预测方法,主要是利用一组最近的实际数据值来预

测未来的数据值,经常用于企业的需求量、销售量、销售额等预测。当产品需求或者销售量 既不快速增长也不快速下降,且不存在季节性因素时,移动平均法能有效地消除预测中的随 机波动。例如,移动平均法可以对企业未来的销售量进行预测,根据预测销量合理地安排生 产、管理库存、制定营销策略,指导企业进行科学决策。

移动平均的计算公式如下:

$$P_{t} = \frac{S_{t-1} + S_{t-2} + \dots + S_{t-n}}{r}$$

其中, $P_t$ 是对未来一期的预测值(即移动平均值);n是移动平均的时期个数; $S_{t-1}$ 是前一期的实际值, $S_{t-2}$ 是前两期的实际值,以此类推, $S_{t-n}$ 是前n期的实际值。

例 3-5: 打开"移动平均.xlsx",该文件中存放了某商家在本年度的 1~12 月份的销售额,请在 C、D、E 三列中计算在不同的移动平均时期个数下,该商家下一年度各月份的销售额预测值,如图 3-13 所示。

	А	В	C	D	Е
1	月份	销售额(万元)	移动平均(间隔2)	移动平均(间隔3)	移动平均(间隔4)
2	1月	179	#N/A	#N/A	#N/A
3	2月	162	171	#N/A	#N/A
4	3月	177	170	173	#N/A
5	4月	173	175	171	173
6	5月	167	170	172	170
7	6月	186	177	175	176
8	7月	193	190	182	180
9	8月	188	191	189	184
10	9月	178	183	186	186
11	10月	182	180	183	185
12	11月	159	171	173	177
13	12月	167	163	169	172

#### 图 3-13 移动平均样张

#### 解答:

(1)本年度1~12月份商品的销售额变化平稳,没有快速下降和增长,因此可以应用移动平均进行销售额预测。

(2)单击"数据"→"数据分析",在"数据分析"对话框中选择分析工具"移动平均",如图 3-14 所示。

数据分析	? 🗙
分析工具( <u>A</u> )	确定
方差分析:无重复双因素分析 相关系数 物方差 描述统计 指数平滑 F-检验双样本方差 傅利叶分析 百方图 <u>影动平均</u> 随机数发生器 ▼	·哨在 取消 帮助(出)
图 2.14 数据公坛工具	

(3)在"移动平均"对话框中,"输入区域"指定本年度销售额单元格区域\$B\$2:\$B\$13, "间隔"为2(即移动平均的时期个数*n*),"输出区域"指定存放结果的单元格区域\$C\$2:\$C\$13, 如图 3-15 所示,单击"确定"按钮,查看计算结果。

移动平均		? ×
输入		
输入区域(]):	\$B\$2:\$B\$13	确定
□标志位于第一行(L)	رعقتن	取消
间隔( <u>N</u> ):	2	帮助( <u>H</u> )
输出选项		
输出区域( <u>O</u> ):	\$C\$2:\$C\$13	
新工作表组(P):		
新工作薄(W)		
图表输出(C)	■ 标准误差	

图 3-15 移动平均设置

注意:n=2时,1月份无法计算移动平均值,显示错误信息#N/A。

(4) 重复步骤(3)计算 n=3 时的移动平均值。

注意: n=3 时,1 月份和 2 月份无法计算移动平均值,显示错误信息 # N/A。

(5) 重复步骤(3)计算 n=4 时的移动平均值。

注意: n=4 时,1 月份、2 月份和3 月份无法计算移动平均值,显示错误信息 # N/A。

通过例 3-5 可以看到,不同的移动平均时期个数对预测结果有影响,因此选择合适的移 动平均时期个数至关重要。为了选择合适的移动平均时期个数,可以对历史数据设置不同 的移动平均时期个数,和目前已知的数据进行对比,从而得出合适的移动平均时期个数。

## 3.1.5 指数平滑

#### 1. 指数平滑概述

指数平滑由布朗(Robert G.Brown)提出,他认为时间序列的态势具有稳定性或规则性,所以时间序列可被合理地顺势推延;最近的过去态势,在某种程度上会持续到未来,并且近期数据比远期数据更重要。

指数平滑法是在移动平均法基础上发展起来的。移动平均法用一组最近的实际数据值 来预测未来,但是并不考虑较远期的数据,在实际应用中不能客观地反映预测结果和数据趋 势的变化;而指数平滑法并不舍弃过去的数据,对不同的历史数据给予不同的权重值,即离 预测期较近的历史数据的权重值较大,离预测期较远的历史数据的权重值较小。指数平滑 法通过计算指数平滑值,配合一定的时间序列预测模型对未来进行预测,其原理是任一期的 指数平滑值都是本期实际观察值与前一期指数平滑值的加权平均。指数平滑法常用于中短 期的企业生产,经济发展等趋势预测。

根据平滑次数不同,指数平滑法分为一次指数平滑法、二次指数平滑法和三次指数平滑 法等。一次指数平滑法适用于水平型历史数据的预测,二次指数平滑法适用于线性趋势变 化的历史数据的预测,三次指数平滑法适用于时间序列呈二次曲线趋势变化的预测。三种 方法基本思想一致,本节只介绍一次指数平滑法和二次指数平滑法。

## 2. 一次指数平滑

一次指数平滑预测应用于时间数列无明显的趋势变化,一次指数平滑值计算公式为:

 $S_{t}^{1} = \alpha y_{t} + (1 - \alpha) S_{t-1}^{1}$ 

根据指数平滑值进行预测,一次指数平滑预测公式为:

$$y_{t+1}^1 = \alpha y_t + (1 - \alpha) y_t^1$$

变量和参数说明如下。

- y<sub>t</sub>: 第 t 期的实际值。
- $S_t^1$ : 第 t 期的一次指数平滑值。
- S<sup>1</sup><sub>i-1</sub>: 第t-1期的一次指数平滑值。初始值S<sup>1</sup><sub>0</sub>的设定需要从时间序列的项数来考虑:当数据较多的时候,初始值的影响可以逐步平滑而降低到最小,此时可以用第一个数据作为初始值。数据较少时,初始值的影响较大,可以取最初几个实际值的平均值作为初始值。
- $y_{t+1}^1$ : 第 t+1 期的预测值。
- $y_t^1$ : 第 t 期的预测值,即一次指数平滑值  $S_t^1$ 。
- α: 平滑系数(0≤α≤1),是第 t 期实际值和预测值的比例分配。用指数平滑法时,确定一个合适的平滑系数非常重要,不同的平滑系数对预测结果产生不同的影响。当时间序列有较大随机波动时,应该为近期数据赋予更大的权重,因此平滑系数应设置较大;反之,当时间序列比较平稳时,平滑系数应设置较小。在实际预测中,可以选取不同的平滑系数进行预测并比较预测结果,然后根据结果选择更符合实际的平滑系数值。一般情况下,α 的取值可以参考以下经验:当时间序列呈稳定的水平趋势时,α 应取较小值,如 0.1~0.3;当时间序列具有明显的上升或下降趋势时,α 应取较大值,如 0.3~0.7。

在一次指数平滑模型中,第t+1期的预测值可以根据第t期的实际值和第t期的预测 值计算得到。例如,某种产品销售量的平滑系数为 0.4,2017 年实际销售量为 50 万件,2017 年的预测值为 52 万件,那么 2018 年的预测销售量: 50 万件×0.4+52 万件×(1-0.4) = 51.2 万件。

例 3-6: 打开文件"销量预测.xlsx",在"一次指数平滑"工作表中,根据表中 2008—2020 年商品的销量数据预测 2021 年商品的销量,如图 3-16 所示。

	A	В	С	D	Е	
1	销售	量数据表	平滑系数			
2	年份	销售量(万个)	0.1	0.3	0.5	
3	2008年	100	100	100	100	
4	2009年	113	101	104	107	
5	2010年	134	105	113	120	
6	2011年	156	110	126	138	
7	2012年	178	117	141	158	
8	2013年	210	126	162	184	
9	2014年	227	136	182	206	
10	2015年	237	146	198	221	
11	2016年	255	157	215	238	
12	2017年	260	167	229	249	
13	2018年	273	178	242	261	
14	2019年	289	189	256	275	
15	2020年	290	199	266	282	
16	2021年		208	273	286	

图 3-16 一次指数平滑预测销量样张

## 解答:

(1)为了分析平滑系数 α 不同取值的特点,分别取 α=0.1,α=0.3,α=0.5 计算一次指数平滑值。单击"数据"→"数据分析",在"数据分析"对话框中选择分析工具"指数平滑",如图 3-17 所示。

数据分析	? ×
分析工具(A) 方差分析:单因素方差分析 方差分析:可重复双因素分析 方差分析:无重复双因素分析 相关系数 协方差 描述统计	确定 取消 帮助( <u>H</u> )
指数平滑 F-检验双样本方差 傅利叶分析 直方图  ▼	

图 3-17 数据分析工具——指数平滑

(2) α=0.1 时,计算一次指数平滑值:在"指数平滑"对话框中,"输入区域"指定销售量 单元格区域\$B\$3:\$B\$15,"阻尼系数"为0.9(即1-α),"输出区域"指定存放结果单元格区域 \$C\$2,如图 3-18 所示设置。单击"确定"按钮查看结果。注意,因为返回的第一个指数平滑 值没有初始值,所以显示错误信息。将C2 单元格中的内容改为"0.1",将C14 单元格公式复 制到C15,得到2020年的一次平滑指数值,设置所有的指数平滑值为整数。

指数平滑		? ×
输入		确定
输入区域([):	\$B\$3:\$B\$15	PIDAL
阻尼系数(D):	0.9	取消
□ 标志(L)		帮助( <u>H</u> )
输出选项		
输出区域( <u>O</u> ):	\$C\$2	
新工作表组(P):		
新工作薄(W)		
图表输出( <u>C</u> )	□ 标准误差	

#### 图 3-18 指数平滑设置

(3) 重复步骤(2),分别计算 α=0.3 和 0.5 时的一次指数平滑值。

(4)预测未来两年的销售量:当 α=0.1 时,在 C16 中输入公式:=ROUND(\$B\$15 \* C2+C15 \* (1-C2),0),得到 2021 年预测销量为 208 万个;将公式分别复制到 D17 和 E17 中,得到当 α=0.3 时,2021 年销量为 273 万个;α=0.5 时,2021 年销量为 286 万个。

根据例 3-6 可知,指数平滑法对实际序列具有平滑作用,如图 3-19 所示,平滑系数越小,平滑作用越强,但对实际数据的变动反应较迟缓。当时间序列的变动出现线性趋势时, 用一次指数平滑法来进行预测将存在着明显的滞后偏差。因此,需要对一次指数平滑进行 修正。

3. 二次指数平滑

一次指数平滑法通常适用于历史数据呈水平线变化的预测,不适用于历史数据呈线性

#### 数据智能——数据处理与管理实践



增长或减少变化的预测。二次指数平滑是对一次指数平滑的再平滑,它适用于具有线性趋势的时间序列,计算公式如下。

$$S_{t}^{1} = \alpha y_{t} + (1 - \alpha) S_{t-1}^{1}$$
  
$$S_{t}^{2} = \alpha S_{t}^{1} + (1 - \alpha) S_{t-1}^{2}$$

变量和参数说明如下。

- y<sub>t</sub>: 第 t 期的实际值。
- $S_t^1$ : 第 t 期的一次指数平滑值。
- $S_{t-1}^1$ : 第t-1期的一次指数平滑值。
- $S_t^2$ : 第 t 期的二次指数平滑值。
- α: 平滑系数。
- $S_{t-1}^2$ : 第t-1期的二次指数平滑值。

和一次指数平滑方法不同,二次指数平滑值并不用于直接预测,只是用来求出线性预测 模型的参数,从而建立预测的数学模型,然后运用数学模型计算预测值。二次指数平滑的数 学模型如下。

$$Y_{t+T} = a + bT$$
  

$$a = 2S_t^1 - S_t^2$$
  

$$b = \frac{\alpha}{1 - \alpha} (S_t^1 - S_t^2)$$

例 3-7: 打开文件"销量预测.xlsx",在"二次指数平滑"工作表中,根据表中 2008—2020 年商品的销量数据预测未来两年商品的销量,如图 3-20 所示。

#### 解答:

(1)由销售量数据表可以看出商品销售量呈线性增长趋势,因此应该对一次平滑指数 值进行修正,应用二次指数平滑法进行预测。设置 α=0.5。单击"数据"→"数据分析",在 "数据分析"对话框中选择分析工具"指数平滑"。

(2) 计算一次指数平滑值:在"指数平滑"对话框中,"输入区域"指定销售量单元格区域\$B\$3:\$B\$15,"阻尼系数"为 0.5,"输出区域"指定存放结果单元格区域\$C\$2,得到一次平滑指数值。将 C2 单元格中的内容改为"一次指数平滑",将 C14 单元格公式复制到 C15,得

	A	В	С	D	E	F	G
1	销售	量数据表	平滑系数	0.5		预测模型	副参数表
2	年份	销售量(万个)	一次指数平滑	二次指数平滑		а	293.037
3	2008年	100	100	100		Ь	10.5286
4	2009年	113	107	103			
5	2010年	134	120	112			
6	2011年	156	138	125			
7	2012年	178	158	142			
8	2013年	210	184	163			
9	2014年	227	206	184			
10	2015年	237	221	203			
11	2016年	255	238	220			
12	2017年	260	249	235			
13	2018年	273	261	248			
14	2019年	289	275	261			
15	2020年	290	283	272			
16	2021年	304					
17	2022年	314					

#### 图 3-20 二次指数平滑预测销量样张

到 2020 年的一次平滑指数值,设置所有的指数平滑值为整数。

(3) 计算二次指数平滑值:在"指数平滑"对话框中,"输入区域"指定销售量单元格区域\$C\$3:\$C\$15,"阻尼系数"为0.5,"输出区域"指定存放结果单元格区域\$D\$2,得到二次平滑指数值。将D2单元格中的内容改为"二次指数平滑",将D14单元格公式复制到D15,得到2020年的二次平滑指数值,设置所有的指数平滑值为整数。

(4) 计算预测模型参数: 在 G2 中输入公式=2 \* C15-D15, 计算参数 a。在 G3 中输入公式=D1/(1-D1) \* (C15-D15), 计算参数 b。

(5) 计算销售量预测值:在 B16 中输入公式=ROUND(G2+G3,0),得到 2021 年销售量预测值为 304。在 B17 中输入公式=ROUND(G2+G3 \* 2,0),得到 2022 年销售量预测值为 314。

#### 3.1.6 回归分析

## 1. 回归分析概述

回归分析是一种常见的预测分析方法。回归(Regression)一词是由英国著名生物学家 兼统计学家 Galton 在研究人类遗传问题时提出的。为了研究父代身高与子代身高的关系, Galton 收集了上千对父亲及其子女的身高数据。经过对数据的深入分析,发现了两者之间 存在着一定的关系:父母的身高增加时,孩子的身高也倾向于增加。但是还有一个有趣的 现象:父母高的孩子,平均身高没有他们父母的平均身高高;父母矮的孩子,平均身高却高 于他们父母的平均身高。这是因为大自然具有一种神奇的约束力,人类身高的分布相对稳 定而不产生两极分化,Galton 把子代身高会向平均身高靠近的趋势称为"回归"。

回归分析是为了寻找多个变量间相关关系的一种方法。它通过对变量数据的分析,去 寻找隐藏在数据背后的相关关系,并用数学模型来描述这种关系,以便对未来进行预测。

回归分析中有两类变量:因变量(也称为响应变量)和自变量(也称为回归变量)。因变量通常是实际问题中所关心的一类指标,用Y表示。自变量是影响因变量取值的一个或多个变量,用X表示,如图 3-21 所示。例如,在产品的销售中,用户满意度对产品的销售至关重要,而产品的质量、价格、售后服务都会对满意度产生影响。因此,可以建立用户满意度与

产品的质量、价格以及售后服务之间的回归模型,对用户的满意度进行预测,其中,用户满意 度是因变量,产品的质量、价格、售后服务都是自变量。



回归分析利用数学统计方法对数据进行处理,确 定因变量与自变量之间的关系,建立变量之间的函数 表达式,即回归方程,并将回归方程作为预测模型,根 据自变量的未来变化预测因变量。

回归分析一般有以下两种分类方法。

(1) 根据因变量和自变量的个数划分。

① 一元回归分析: 自变量只有一个。

② 多元回归分析:自变量有两个或两个以上。

(2) 根据因变量和自变量的相关关系划分。

① 线性回归分析: 自变量和因变量是线性关系。

② 非线性回归分析:自变量和因变量是非线性 关系。

回归分析的主要步骤如下。

(1) 根据预测目标,确定自变量和因变量。

明确预测的具体目标,也就确定了因变量。例如,预测目标是下一年度的销售量,那么 销售量 Y 就是因变量。寻找与预测目标的相关影响因素,即自变量,并从中选出主要的影 响因素,例如,影响销售量的主要因素有产品价格、产品质量、产品服务等。

(2) 建立回归预测模型。

依据自变量和因变量的历史数据进行计算,在此基础上建立回归分析方程,即回归分析 预测模型。

(3)进行相关分析,确定相关系数。

回归分析是对具有因果关系的自变量和因变量所进行的数理统计分析处理。只有自变量与因变量确实存在某种关系时,建立的回归方程才有意义。因此,作为自变量的因素与作为因变量的预测对象是否有关,相关程度如何,以及判断这种相关程度的可靠性多高,是进行回归分析必须要解决的问题。进行相关分析,一般要求出相关关系,以相关系数的大小来判断自变量和因变量的相关程度。

(4) 检验回归预测模型,计算预测误差。

回归预测模型是否可用于实际预测,取决于对回归预测模型的检验和对预测误差的计算。 回归方程只有通过各种检验,且预测误差较小,才能将回归方程作为预测模型进行预测。

(5) 计算并确定预测值。

利用回归预测模型计算预测值,并对预测值进行分析,确定最后的预测值,并计算预测 值的置信区间。

2. 回归分析模型

回归分析模型描述了自变量(X)与因变量(Y)之间的关系,可以表示为:

## Y = f(X) + e

其中, f(X)是自变量 X(一个或多个)的函数; e 是随机误差变量, 表示其他未知的因素 或随机因素对因变量 Y 产生的影响。本节只讨论自变量为一个的一元回归分析模型。

根据 *f*(*X*)不同,回归分析模型主要可以分为线性回归、对数回归、指数回归、幂次回归和多项式回归。

(1) 线性回归。

线性回归的回归线是线性的,是最常用的回归模型之一,如图 3-22 所示。线性回归使用最佳的拟合直线(也就是回归线)在因变量 Y 和自变量 X 之间建立一种关系,表示如下:

 $Y = b_0 + b_1 X$ 

参数说明如下。

b<sub>0</sub>、b<sub>1</sub>:回归系数。对于线性回归线来说,b<sub>0</sub>表示直线的截距,b<sub>1</sub>表示直线的斜率。



(2) 多项式回归。

对于一个回归方程,如果自变量的指数大于1,那么它就是多项式回归,表示如下。

 $Y = b_0 + b_1 X + b_2 X^2 + \dots + b_n X^n$ 

在多项式回归中,最佳拟合线不是直线,而是一条曲线,图 3-23 给出了一个数据点拟合的多项式回归方程。



<sup>(3)</sup> 其他回归模型。

除了线性回归和多项式回归,还可以建立其他的回归模型,如表 3-2 所示。

#### 表 3-2 回归模型

回归模型	模型表示
对数回归	$Y = b_0 + b_1 \ln(X)$
指数回归	$Y = b_0 \exp(b_1 X)$
幂次回归	$Y\!=\!b_{\scriptscriptstyle 0}X^{b_1}$

在 Excel 中,可以给图表添加趋势线进行回归分析,Excel 中提供了线性、多项式、对数、 指数、幂次等多种趋势线(即回归线)。

例 3-8:客服中心客户电话的接听数量与回访数量之间存在着一定关系,打开文件"客服中心接听与回访数据.xlsx",通过回归分析,生成接听量与回访量之间的回归方程,并根据接听量对回访量进行预测。

## 解答:

① 选中接听量和回访量两列数据,插入散点图,如图 3-24 所示,从图中可以看出,回访量和接听量之间呈明显的线性关系。

② 添加趋势线(即回归方程):选择"图表工具"→"设计"→"添加图表元素"→"趋势 线"→"其他趋势线选项",在"设置趋势线格式"面板中选择"线性"回归分析,并选中"显示公 式""显示 R 平方值"复选框,如图 3-25 所示。R 平方值反映了回归线的估计值与对应的实 际数据之间的拟合程度,可以作为模型拟合度优劣的度量,用于评价模型的可靠性,取值范 围为 0~1。R 平方值越大,表明数据的拟合程度越高,因变量和自变量之间的相关性越大。



设置趋势线格式 ▼ × <sup>趋势线选项</sup> ▼						
۵ 🗘 🗈						
▲ 趋势线选项						
○ 指数(凶)						
● 线性(L)						
○ 对数(O)						
√ ○ /D\	顺序([	2) 2	÷			
○ 幂W)						
✓ ○ 移动平 均(M)	周期(E	2)	A V			
趋势线名称						
● 自动( <u>A</u> )	线性 (	回访量)				
○ 自定义(C)						
趋势预测						
向前( <u>F</u> )	0.0	周期				
向后(B)	0.0		周期			
设置截距(S)	C	0.0				
✓ 显示公式(E)						
✓ 显示 R 平方值( <u>R</u> )	)					

图 3-25 设置趋势线格式

③利用回归方程进行预测:步骤②为散点图中添加的线性趋势线如图 3-26 所示,回访 量与接听量间的线性回归方程为 y=0.8966x-1.0161,x 表示接听量,y 表示回访量。R<sup>2</sup> 为 0.9748,表示数据拟合程度达到 97.48%,拟合程度高,说明接听量与回访量之间存在着线 性关系,可以通过回归方程根据接听量预测回访量。例如,如果客服人员接听量为 50 个,那 么预测回访量为 44 个。如果回访量低于 44,说明回访量没有达标。



## 3.1.7 习题与实践

#### 1. 填空题

- (1) 模拟运算表最多能解决\_\_\_\_\_个变量对于计算结果的影响问题。
- (2) 按因变量和自变量之间关系,回归分析可以分为线性回归和
- (3)时间序列含有的成分包括趋势、\_\_\_\_\_、\_\_\_\_和随机性。
- (4) 当历史数据呈线性上升趋势时,应该使用 指数平滑进行预测。
- (5) 是根据目标结果来倒推生成该结果的输入值,属于一种"逆"运算。

#### 2. 选择题

- (1) 以下数据分析工具中,通常用于预测分析的为()。 B. 百分比排位 C. 相关系数 A. 方差 D. 移动平均 (2) 以下可用来解决曲线拟合问题的是( )。 A. 移动平均 B. 指数平滑 C. 回归分析 D. 单变量求解 (3) 单变量求解就是求解具有()个变量的方程。 A. 1 B. 2 C. 3 D. 任意 (4) 关于模拟运算表,以下叙述正确的是()。 A. 在引用列变量的单变量模拟运算表中,计算公式一定位于变量列的右上角单元 B. 在引用行变量的单变量模拟运算表中,计算公式一定位于变量列的左下角单元 C. 在双变量模拟运算表中,计算公式一定位于模拟运算表左上角行变量和列变量 交叉单元格
  - D. 模拟运算表可以删除其中某个格的值
- (5)进行单变量求解的时候,需要设置的参数包括目标单元格、可变单元格和()。

A. 目标值 C. 输入列引用的单元格 B. 输入行引用的单元格 D. 结果单元格

## ◆ 3.2 数据可视化

数据可视化是关于数据视觉表现形式的科学技术研究,它利用图形、图像处理、计算机 视觉及用户界面,通过表达、建模以及对数据立体、表面、属性、动画的显示,加以可视化解 释,以便于人们更好地发现和利用数据的价值。

Excel 具有很强的图表处理功能,可以很方便地将工作表中的有关数据制作成专业化的图表。图表是一个图形对象,通过它可以将一组枯燥的数据更形象地展示出来,让他人看得更直观、更清晰。因此,对于数据分析的结果呈现,大多数人都会选择以图表来反映数据。

## 3.2.1 可视化基础

要使用图表来展示数据,就必须要创建图表,这是数据可视化的第一步。但是在创建图 表之前,首先要明确以下四点内容。

#### 1. 图表类型

图表类型一定要与分析需求一致。只有使用的图表类型与需求分析一致,才能更好地 反映数据的分析结果,因此有必要了解数据与图表之间的各种关系,如表 3-3 所示。

数据关系	对应的图	图表类型	说明
比较关系	柱形图		柱形图是使用最频繁的图表类型,用于显示一段时间内的数据变化或显示各项数据之间的比较情况。由于柱形图可以通过数量来表现数据之间的差异,因此被广泛地应用于时间序列数据和频率分布数据的分析
	条刑	的图	条形图也是用于显示各项数据之间的比较情况,但它弱化了时间的变化, 偏重于比较数量大小
拉执子系	折线图		折线图是以折线的方式展示某一时间段的相关类别数据的变化趋势,强 调时间性和变动率,适用于显示与分析在相等时间段内的数据趋势
旭労大尔	面积图		面积图主要是以面积的大小来显示数据随时间而变化的趋势,也可表示 所有数据的总值趋势
	饼图		饼图一般用于展示总和为 100%的各项数据的占比关系,该图表类型只能对一列数据进行比较分析
占比关系 环形图		影图	要对包含多列的目标数据进行占比分析可以使用系统提供的圆环图来详 细说明数据的比例关系。它由一个或者多个同心的圆环组成,每个圆环 表示一个数据系列,并划分为多个环形段,每个环形段的长度代表一个数 据值在相应数据系列中所占的比例。此外,在表格中从上到下的数据记 录顺序,在圆环图中对应从内到外的圆环
	雷达图		在对同一对象的多个指标进行描述和分析时,可选用该类型的图表,使阅读者能同时对多个指标的状况和发展趋势一目了然
其他关系	XY 散 点图	散点图	散点图将沿横坐标(X轴)方向显示的一组数值数据和沿纵坐标轴(Y轴) 方向显示的另一组数值数据合并到单一数据点,并按不均匀的间隔或簇 显示出来,常用于比较成对的数据,或显示独立的数据点之间的关系

表 3-3 不同关系对应的图表类型

续表

79

数据关系	对应的图	图表类型	说 明
其他关系	XY 散 点图	气泡图	气泡图是散点图的变体,因此,其要求的数据排列方式与散点图一样,即 确定一行或一列表示 X 轴数值,在其相邻的一行或一列表示相应的 Y 轴 数值
	股价图		股价图主要用于展示股票价格的波动情况,若要在工作表中使用股价图, 其数据的组织方式非常重要,必须严格按照每种图表类型要求的顺序来 排列
	旭日图		旭日图非常适合显示分层数据,并将层次结构的每个级别均通过一个环 或圆形表示,最内层的圆表示层次结构的顶级(不含任何分层数据的旭日 图与圆环图类似)。若具有多个级别类别的旭日图,则强调外环与内环的 关系
	树状图		树状图是一种直观和易读的图表,所以特别适合展示数据的比例和数据 的层次关系。如分析一段时期内什么商品销量最大、哪种产品赚钱最 多等
	箱型图		箱型图不仅能很好地展示和分析出数据分布区域和情况,而且能直观地 展示出一批数据的"四分值"、平均值以及离散值
	瀑布图		瀑布图是由麦肯锡顾问公司所独创的图表类型,因为形似瀑布流水而称 为瀑布图(Waterfall Plot)。此种图表采用绝对值与相对值结合的方式, 适用于表达数个特定数值之间的数量变化关系

#### 2. 图表标题

图表标题是传达图表内容的第一手信息,一定要谨慎和仔细,如果图表标题设置得不合适,不仅不能很好地传递信息,而且容易让他人曲解。

#### 3. 图表大小

图表大小不合适会影响数据结果的分析,尤其对于数据多的图表,过小的图表会让数据 挤在一起,不易阅读。

## 4. 图表位置

图表的位置要根据分析目的来确定,一般情况下是浮在数据表中,如果单独放大查看图表,可以将其移动到图表工作表中。

由此可见,要创建一个图表至少要经过如上4个过程,下面以在"国内生产总值数据"工 作表中使用图表来分析 2016—2020 年国内生产总值和增长率数据为例。

例 3-9: 在"国内生产总值数据.xlsx"中记录了 2016—2020 年国内生产总值和增长率的 基本信息,现使用图表来分析国内生产总值和增长率数据。其结果如图 3-27 所示。

(1)选择 A2:C7 单元格区域,打开"插入"选项卡,在"图表"组中单击"对话框启动器" 按钮,在打开的"插入图表"对话框中打开"所有图表"选项卡,在其中选择需要的图表类型。 由于需要分析两组数据系列,所以采用不同的图表类型来区别这两组数据系列。国内生产 总值数据系列用簇状柱形图图表类型,增长速度数据系列用折线图图表类型,所以选用的图 表类型为"组合"。

(2)由于国内生产总值数据和增长速度数据两组数量单位不同,绘制到同一个图表中, 增长速度数据较小不能正常显示,为了方便查看和分析每个数据系列的数据,需要为增长速 度数据系列指定添加一个次坐标轴,如图 3-28 所示。







图 3-28 选择图表类型并添加次坐标轴

(3) 在返回的工作表中即可查看到创建的图表,为图表添加标题——"2016 年至 2020 年国内生产总值及其增长速度",调整图表大小、位置。

(4) 对图表进行美化。选择图表,打开"图表工具设计"选项卡,在"图表样式"组的列

表框中选择"样式 6"。右击图表,在弹出的菜单中选择"设置图表区域格式"命令,在弹出的 对话框中选择"图表选项"→"边框"→"圆角",如图 3-29 所示。

图表1 · : × ✓ fx								
AB	C	DE	F (	н	T	ТТ	设置图表区格式	* X
1			. ,				图表选项 ▼ 文本选项	
2 <u>年份</u> 国内生产总值(亿元)	比上年增长				X #	±1.0(T)	🖏 🗋 🖪	
3 2010年 羊 740,395.00 4 2017在 ¥ 832,036,00	6.80%				00 9 Da 4	5切(1)		
1         2011中         1         032,030.00           5         2018年         ¥         919,281.00	6. 70%				-8.8	之中(5)	4 沈辉	
6 2019年 ¥ 986, 515.00	6.00%					·1201还坝; 	A 12/1E	
7 2020年 ¥ 1,015,986.00	2.30%					D	○ 无线荼( <u>N</u> )	
8		0			1	重设以匹配样式(A)	<ul> <li>实线(<u>S</u>)</li> </ul>	
<b>9 2016</b> 年3	至2020年国内	生产总值及	<b>支</b> 其增长速	度	A	字体(E)	○ 渐变线(G)	
			con none.	~	di 3	更改图表类型(Y)	<ul> <li>自动(U)</li> </ul>	
12					協力	弓存为模板(S)	颜色( <u>C</u> )	<u>ð</u> •
13 ¥1,000,000.00			_	_	E i	先择教据(E)	透明度(T) ▋	1
14					<b>1 €</b>	运行 (V)	nder miter (1.6.0)	•
16 \$800,000.00					n -	=	SC/RC(VV)	•
17 ¥600,000.00							复合类型( <u>C</u> )	= -
18					면지	866(G)	短划线类型(D)	
19 ¥400,000.00					16.0	重于顶层(K) ▶	端点类型(A)	Ŧ
20 X200.000.00					96.8	≝于底层(K) ▶	联接类型(J)	-
21 4100,000,000					打	旨定宏(№)	箭头前端类型(B)	≓▼
22 ¥-					- 🕑 i	设置图表区域格式(E)	箭头前端大小(S)	= -
2016年	2017年			2020年	唱奏	改据透视图选项(Q)	箭头末端类型(E)	<b>≓</b> .▼
25	━━ 国内生产总值(						· 等处主进于小(N)	
26		0			8	· Z·		
<ul> <li>→ Sheet1 ⊕</li> </ul>					捕索	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	↓ ◎用(区)	
2014 Ban					項九			

#### 图 3-29 设置圆角

(5)选择图表,单击图表右上角的"+"按钮,在展开的菜单中将鼠标光标移动到"图例" 选项上,单击其右侧按钮,在弹出的子菜单中选择"顶部"选项即可将图例位置从底部显示变 为从顶部显示,如图 3-30 所示。



#### 图 3-30 改变图例位置

(6) 在图表中,所有的数据都通过 Excel 自动转换为相对大小的图形,如果需要查看指 定数据系列精确的数值,可以借助数据标签来辅助查看。在图表中选择国内生产总值数据 系列,单击图表右上角的"+"按钮,在展开的菜单中将鼠标光标移动到"数据标签"选项上, 单击其右侧按钮,在弹出的子菜单中选择"居中"选项即可,同样也可为增长率添加数据标 签,如图 3-31 所示。

美化图表的方法有很多种,可以设置图表的样式还有设置形状效果,更改布局样式及设 置艺术效果的图表标题等。但是需要注意以下两点。

(1) 多种色块不要过多。

当需要在图表的多个较大区域使用颜色时,这个颜色不要太抢眼,而且在一张图表中,