

时间序列分析

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

时间序列基本概念

时间序列，英文名叫time series

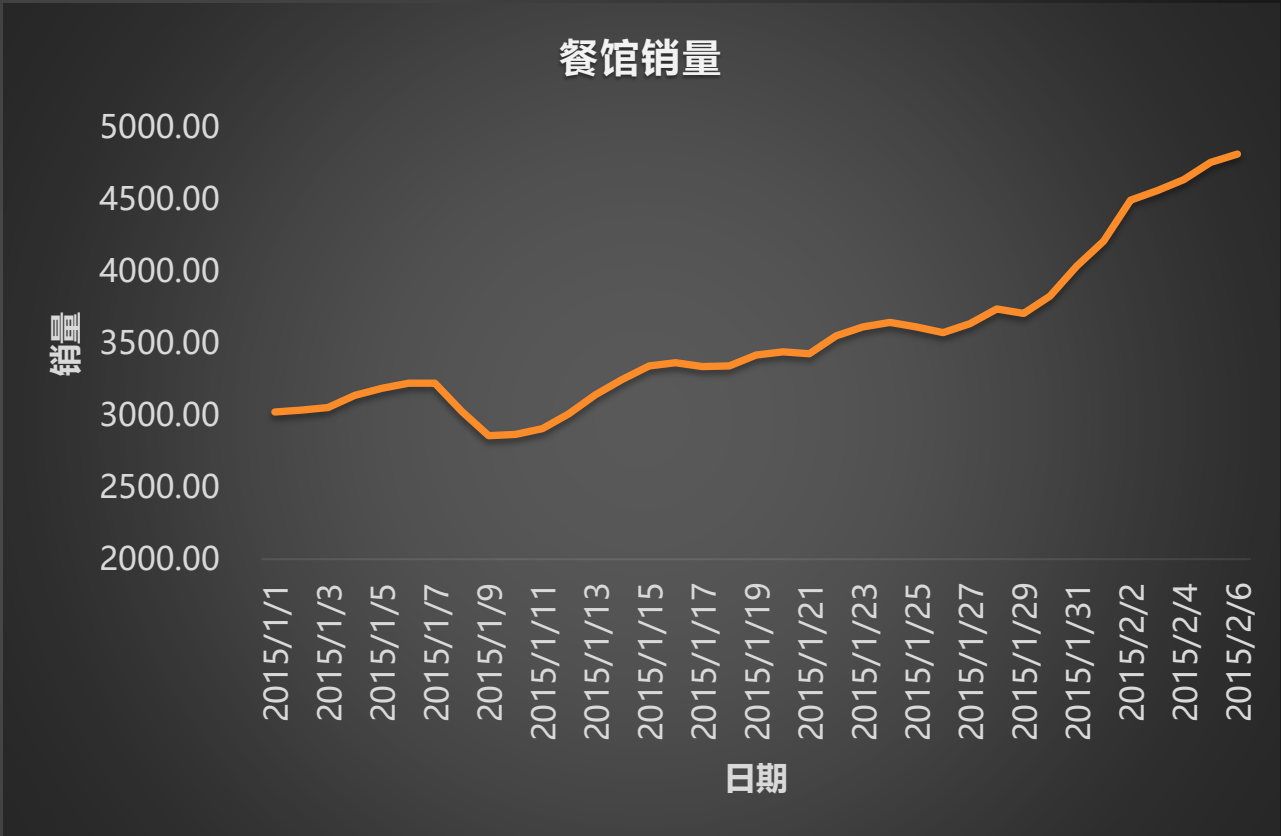
时间序列数据有两个要素：**时间**和**数值型指标**。

例如：

- 餐馆每天的销量数据。
- 2000-2013年人均GDP数据。
- 股票数据：某只股票每天的收盘价。
- 。

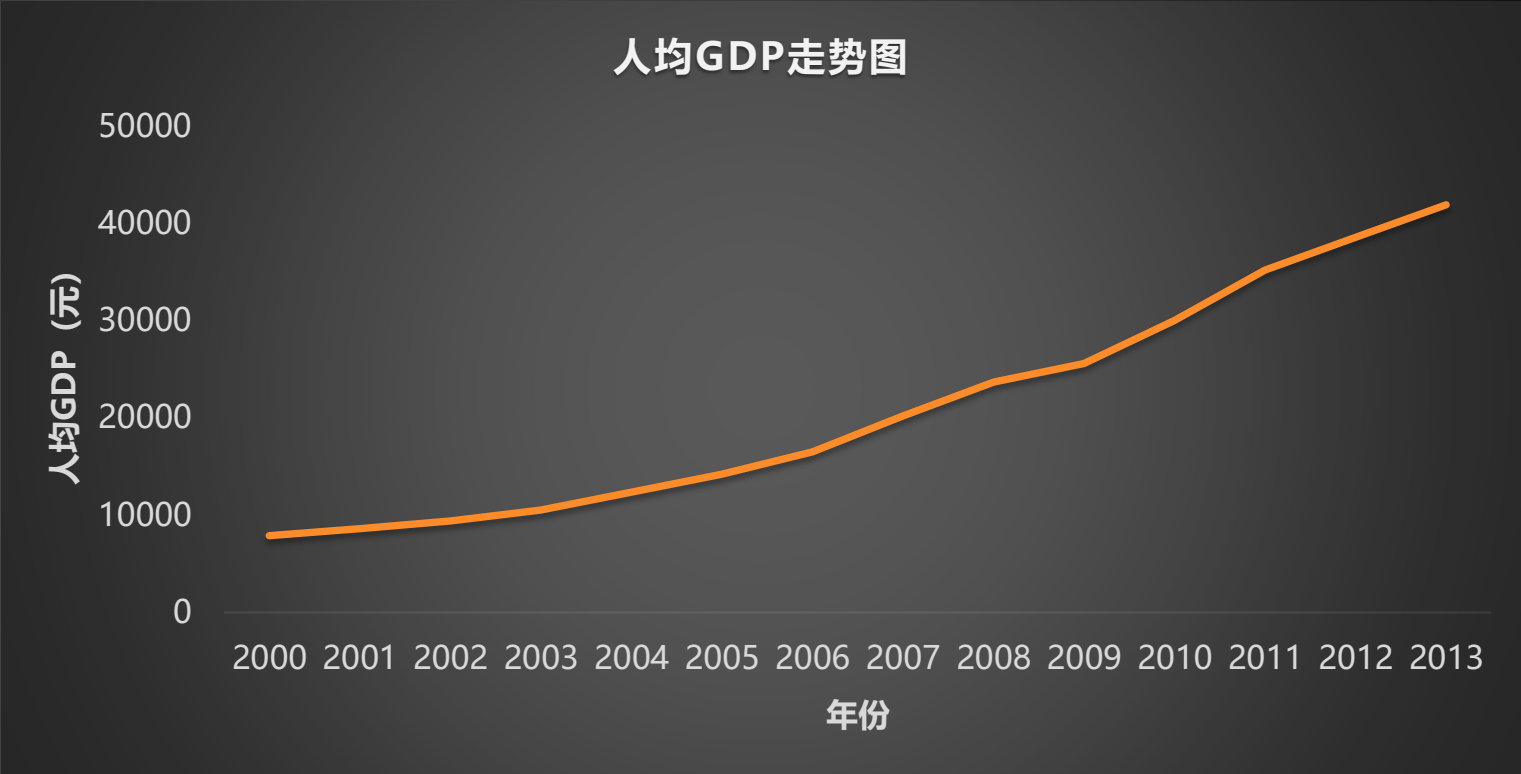
时间序列：餐馆销量数据

日期	销量
2015/1/1	3023.01
2015/1/2	3039.01
2015/1/3	3056.01
2015/1/4	3138.01
2015/1/5	3188.01
2015/1/6	3224.01
2015/1/7	3226.01
2015/1/8	3029.01
2015/1/9	2859.01
2015/1/10	2870.01
2015/1/11	2910.01
2015/1/12	3012.01
2015/1/13	3142.01
2015/1/14	3252.01
2015/1/15	3342.01



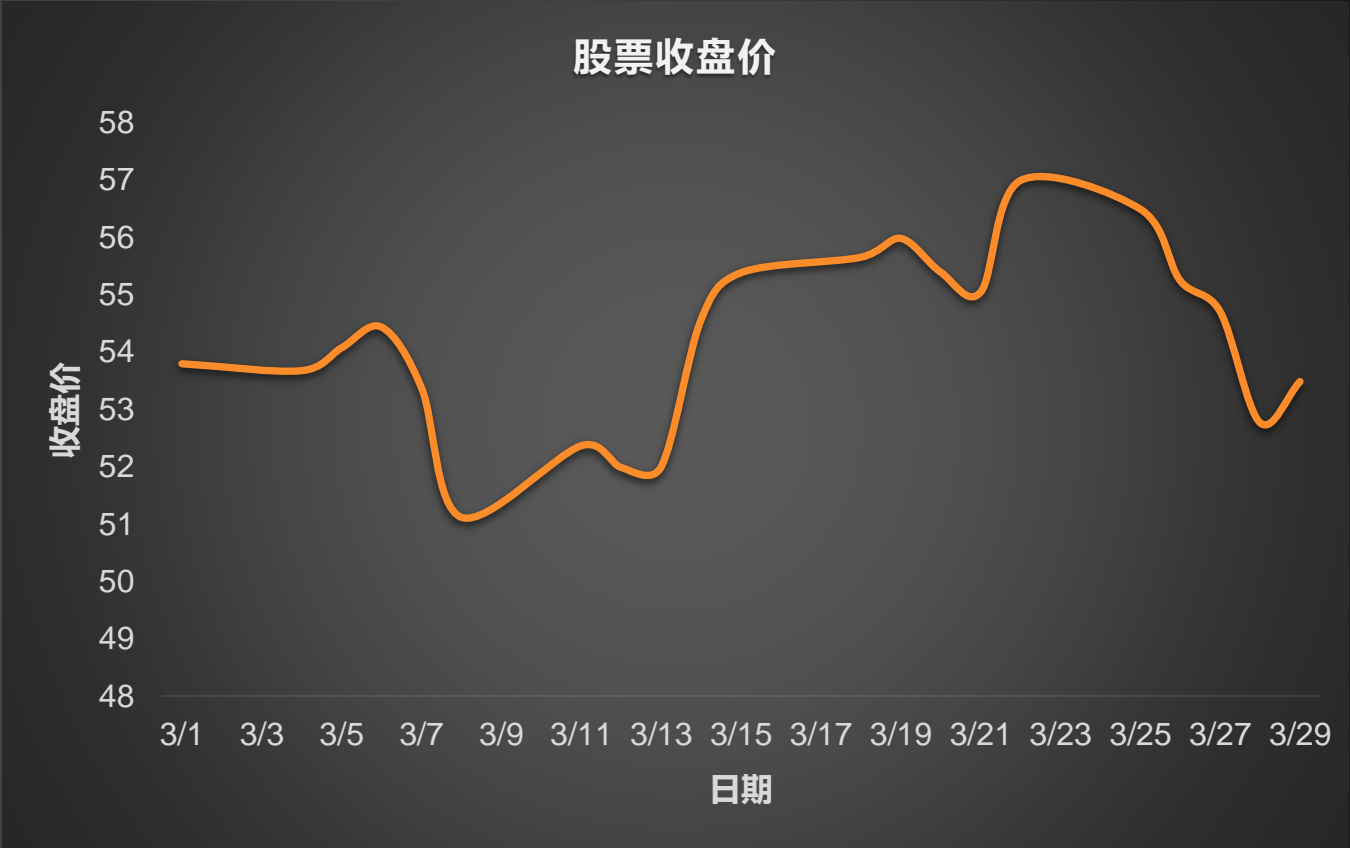
时间序列：人均GDP数据

年份	人均GDP (元)
2000	7857.7
2001	8621.7
2002	9398.1
2003	10542
2004	12335.6
2005	14185.4
2006	16499.7
2007	20169.5
2008	23707.7
2009	25607.5
2010	30015
2011	35197.8
2012	38549.5
2013	41907.6



时间序列：股票数据

日期	收盘价
2019/3/29	53.49
2019/3/28	52.77
2019/3/27	54.71
2019/3/26	55.25
2019/3/25	56.5
2019/3/22	57
2019/3/21	55.05
2019/3/20	55.4
2019/3/19	55.99
2019/3/18	55.66
2019/3/15	55.39

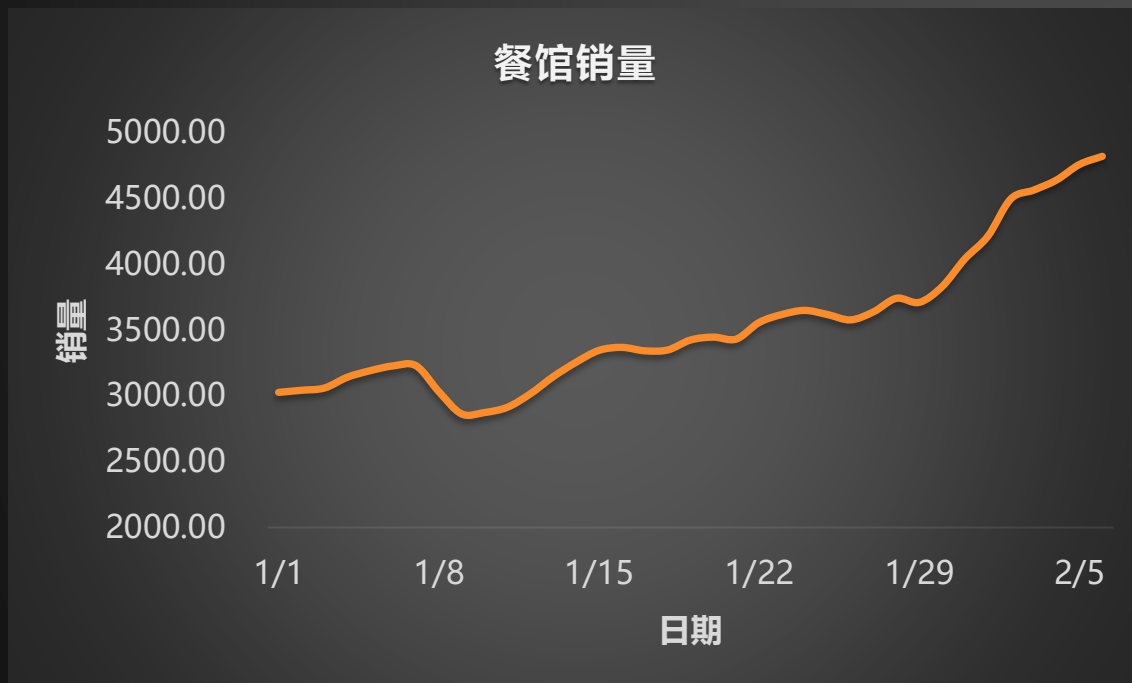


描述时间序列的折线图叫作序列图。

时间序列的分类

- 时间序列按照是否存在趋势分为**平稳序列**和**非平稳序列**。

非平稳序列

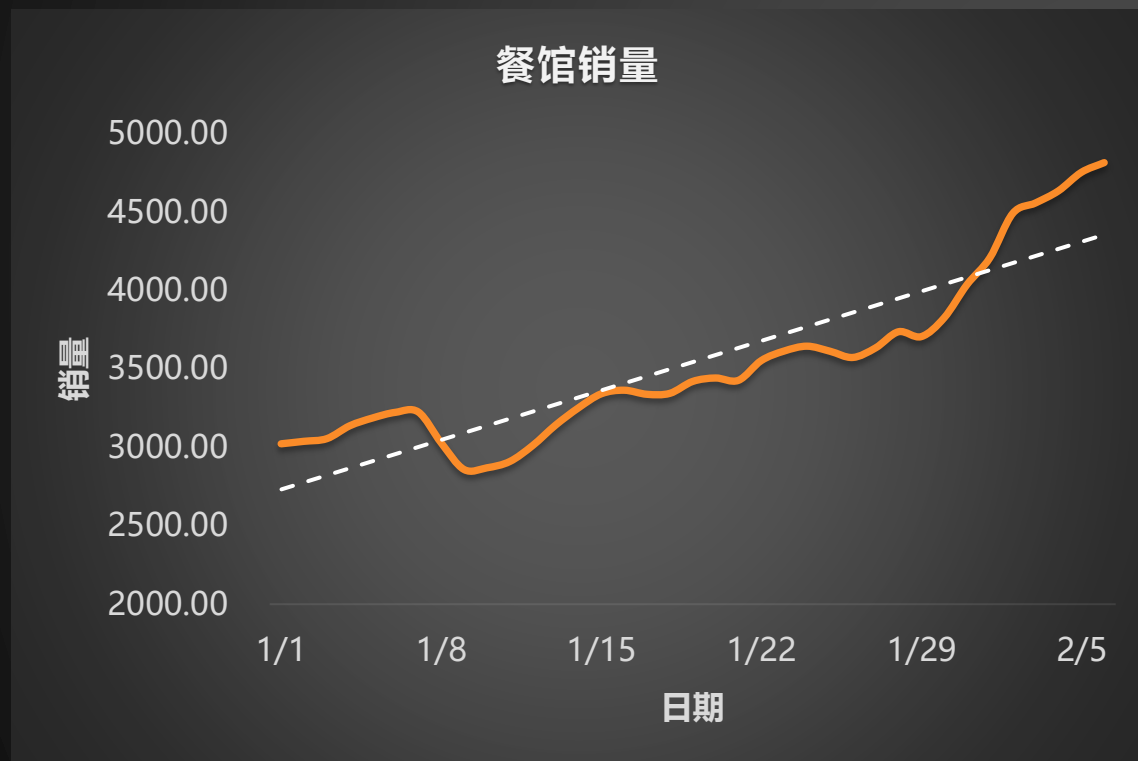


平稳序列

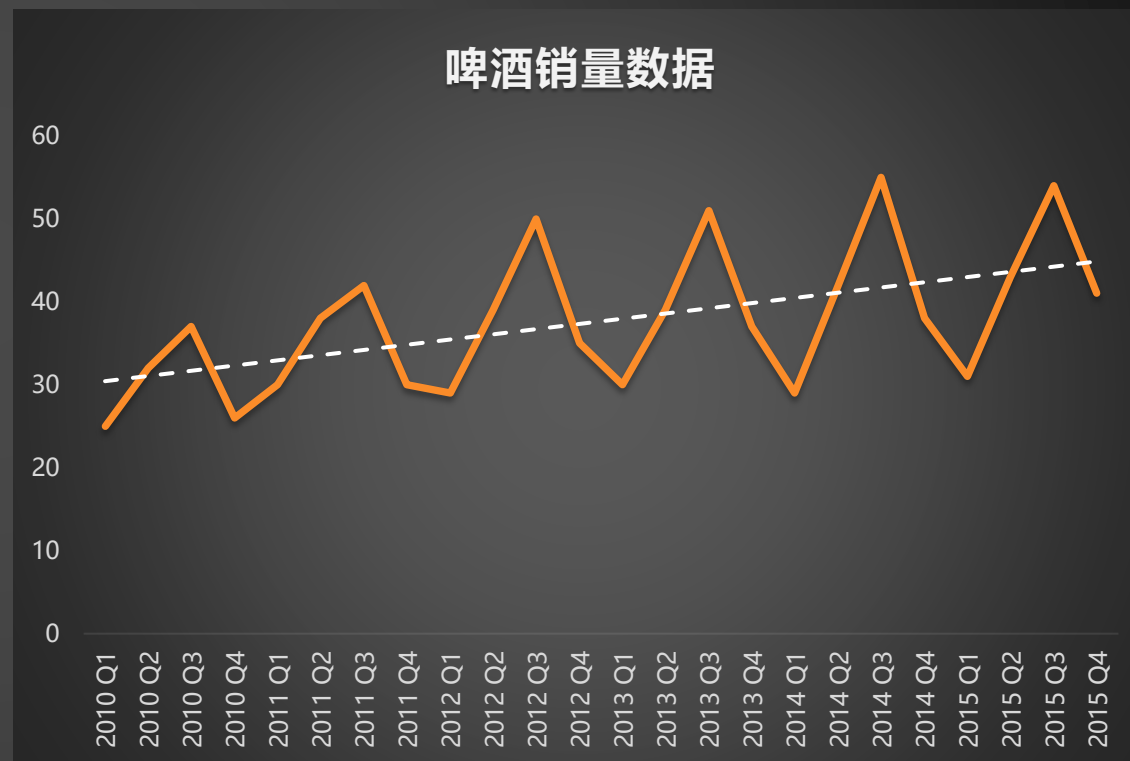


时间序列中的成分

- 时间序列中的成分主要有四种：趋势成分、季节成分、周期成分和随机成分。



趋势成分



趋势+季节性

时间序列中的成分



随机成分

时间序列中的成分

- **趋势**：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，称为长期趋势。
- **季节性**：时间序列在一年内重复出现的周期性波动。
- **周期性**：也称循环波动，是时间序列中呈现出来的围绕长期趋势的一种波浪形或震荡式变动，变动周期多在一年以上，常见于经济学数据中。
- **随机性**：时间序列中去除趋势、周期性和季节性之后的偶然性波动。

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

差分

非平稳序列



差分

平稳序列



差分

对于非平稳序列，可以通过差分将其转化为平稳序列。

差分简单来说，就是将时间序列中连续的两项做差（后一项减去前一项），得到的差值构成一个新的序列。用数学语言表达为：

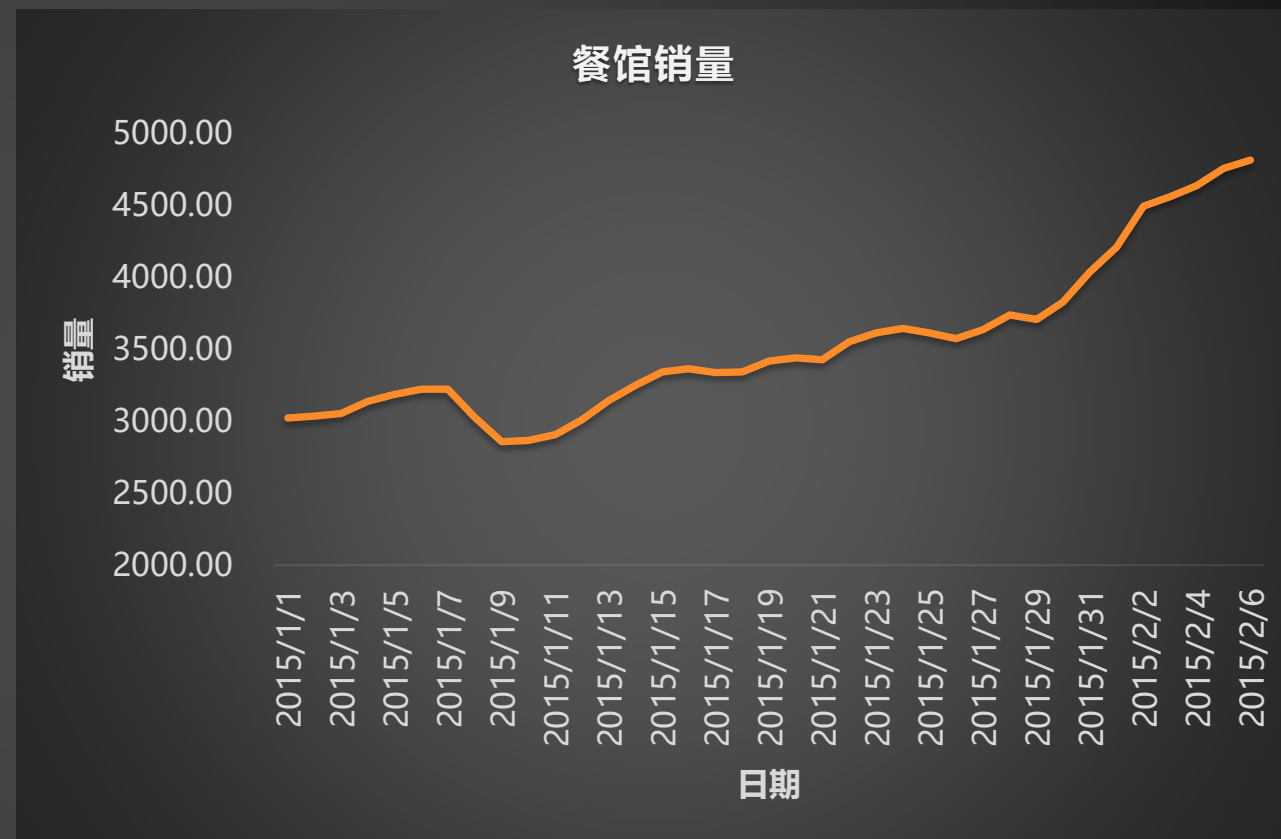
$$c_t = y_t - y_{t-1}$$

- 做一次差分就叫一阶差分，记差分阶数 d 为1。
- 做两次差分也叫二阶差分，记差分阶数 d 为2。
- 三次、四次等同理。

通常，对于一个非平稳序列来说，做一次差分就能够去除其中的线性趋势，从而变成一个平稳序列。

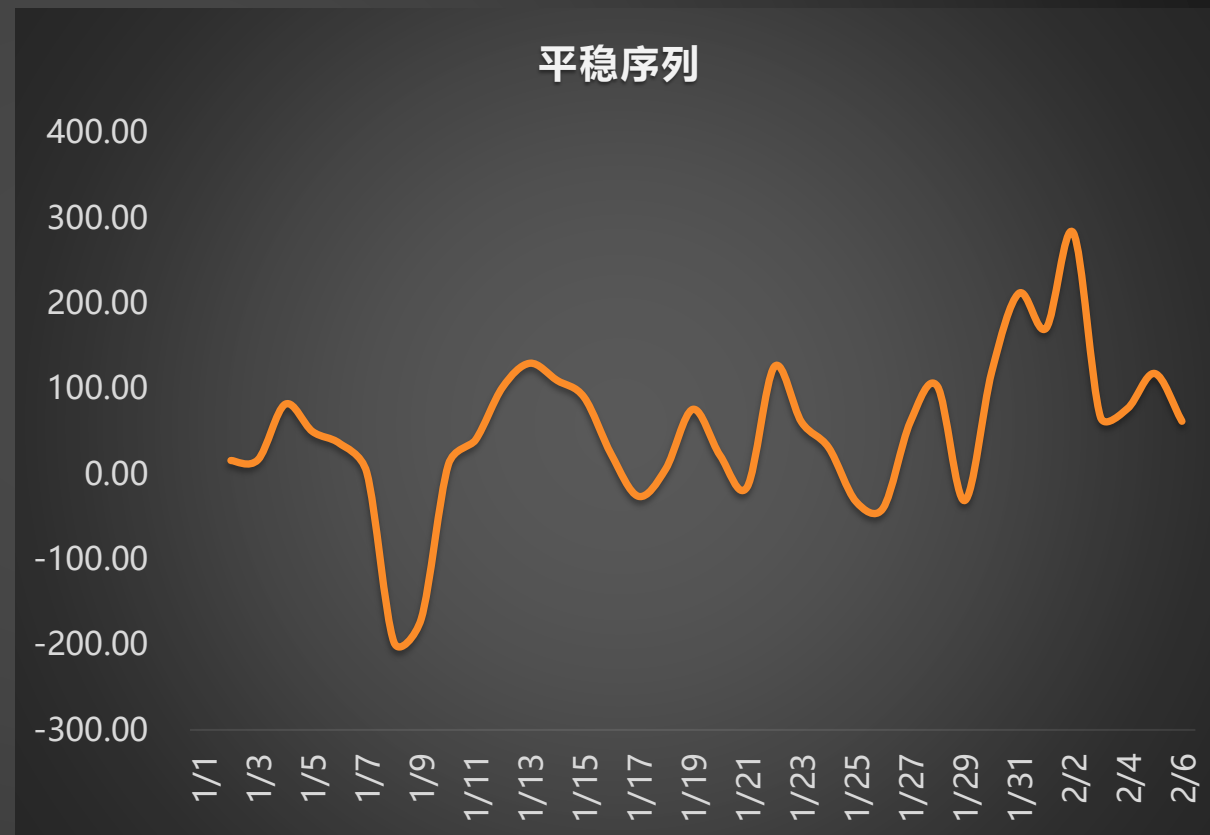
差分

日期	销量
2015/1/1	3023.01
2015/1/2	3039.01
2015/1/3	3056.01
2015/1/4	3138.01
2015/1/5	3188.01
2015/1/6	3224.01
2015/1/7	3226.01
2015/1/8	3029.01
2015/1/9	2859.01
2015/1/10	2870.01
2015/1/11	2910.01
2015/1/12	3012.01
2015/1/13	3142.01
2015/1/14	3252.01
2015/1/15	3342.01



差分

日期	销量	一阶差分
2015/1/1	3023.01	
2015/1/2	3039.01	16.00
2015/1/3	3056.01	17.00
2015/1/4	3138.01	82.00
2015/1/5	3188.01	50.00
2015/1/6	3224.01	36.00
2015/1/7	3226.01	2.00
2015/1/8	3029.01	-197.00
2015/1/9	2859.01	-170.00
2015/1/10	2870.01	11.00
2015/1/11	2910.01	40.00
2015/1/12	3012.01	102.00
2015/1/13	3142.01	130.00
2015/1/14	3252.01	110.00
2015/1/15	3342.01	90.00



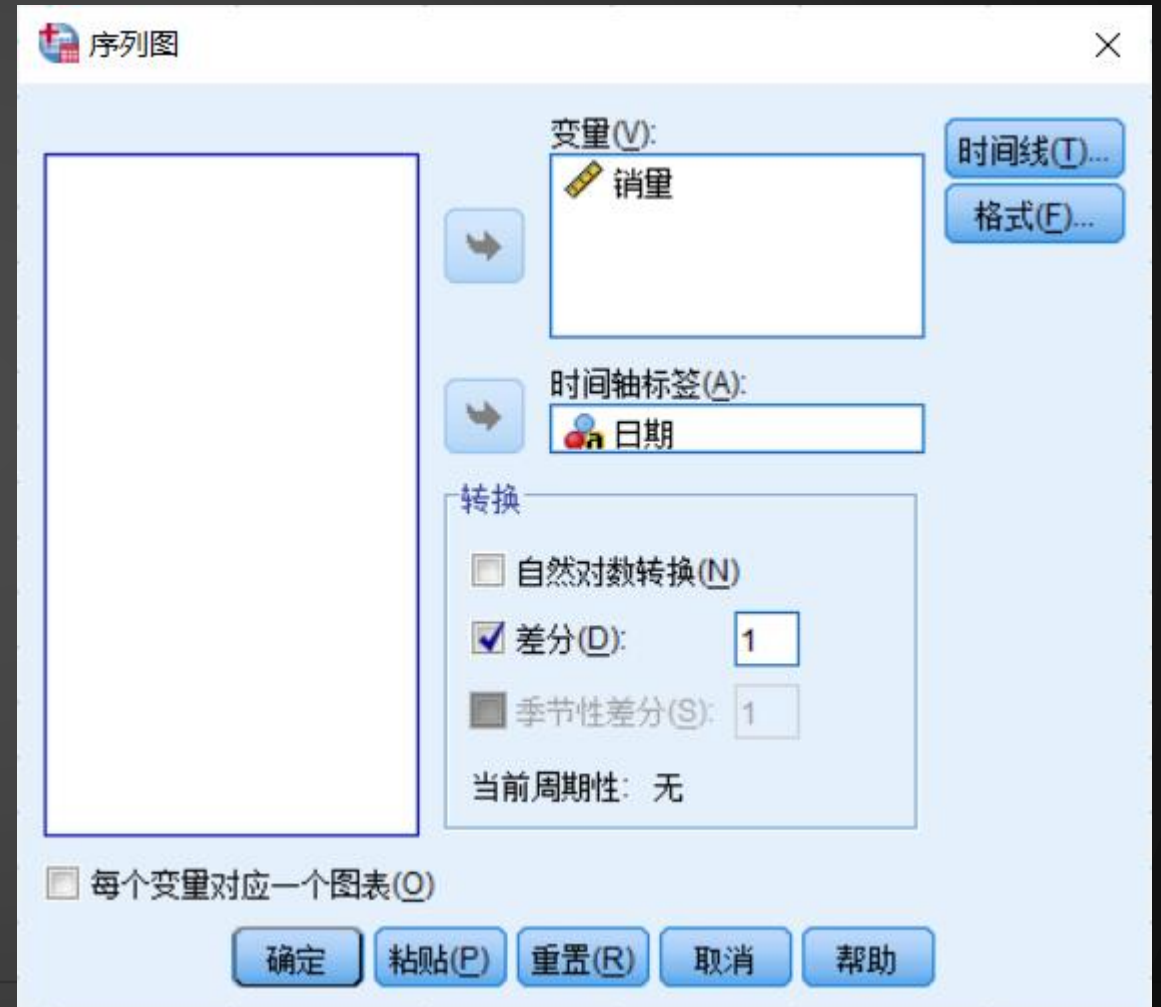
SPSS序列图及差分

序列图：

SPSS菜单选择【分析】 - 【预测】 - 【序列图】

差分：

SPSS菜单选择【分析】 - 【预测】 - 【序列图】，
勾选【差分】



增长率分析

增长率：也称增长速度，是时间序列中报告期观察值与基期观察值之比减1后的结果，用%表示

- 环比增长率：指报告期观察值与前一时期观察值之比减1，说明现象逐期增长变化的程度。

$$G_i = \frac{y_i - y_{i-1}}{y_{i-1}}, i = 1, 2, \dots, n$$

- 定基增长率：报告期观察值与某一固定时期观察值之比减1，说明现象在整个观察期内总的增长变化程度。

$$G_i = \frac{y_i - y_0}{y_0}, i = 1, 2, \dots, n$$

- 平均增长率：也称平均增长速度，是时间序列中逐期环比值的几何平均数减1后的结果。

$$\bar{G} = \sqrt[n]{\left(\frac{Y_1}{Y_0}\right)\left(\frac{Y_2}{Y_1}\right)\cdots\left(\frac{Y_n}{Y_{n-1}}\right)} - 1 = \sqrt[n]{\frac{Y_n}{Y_0}} - 1$$

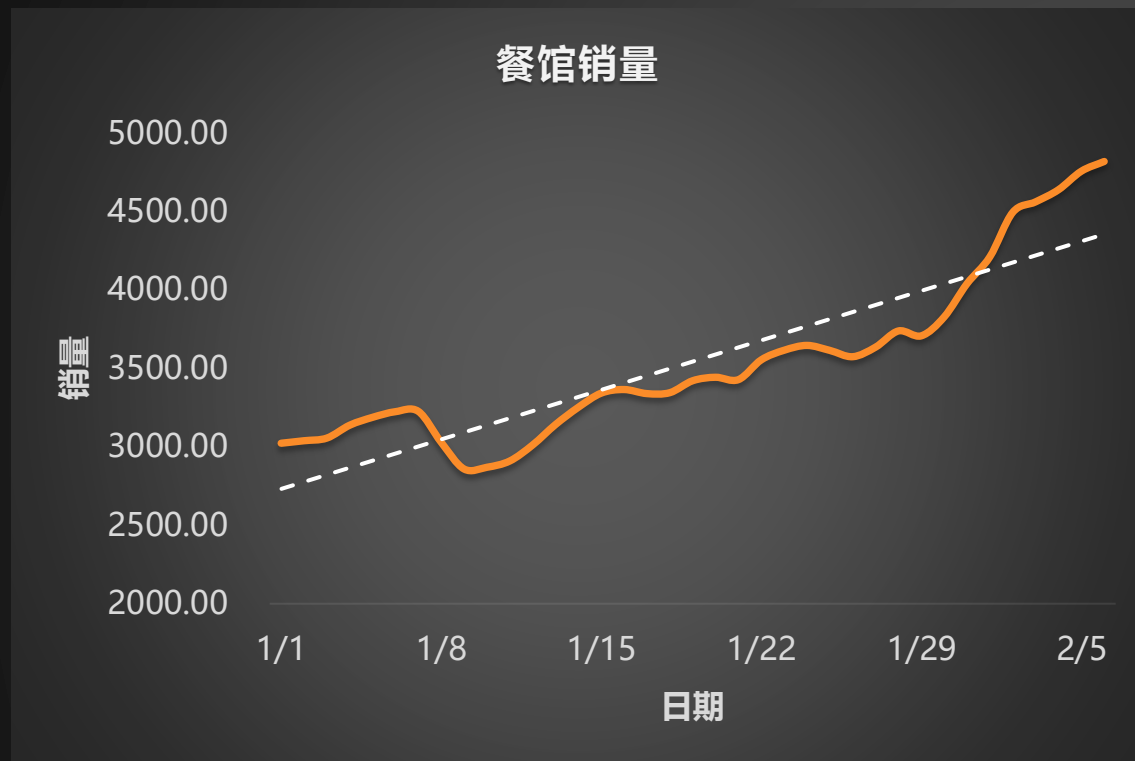
本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

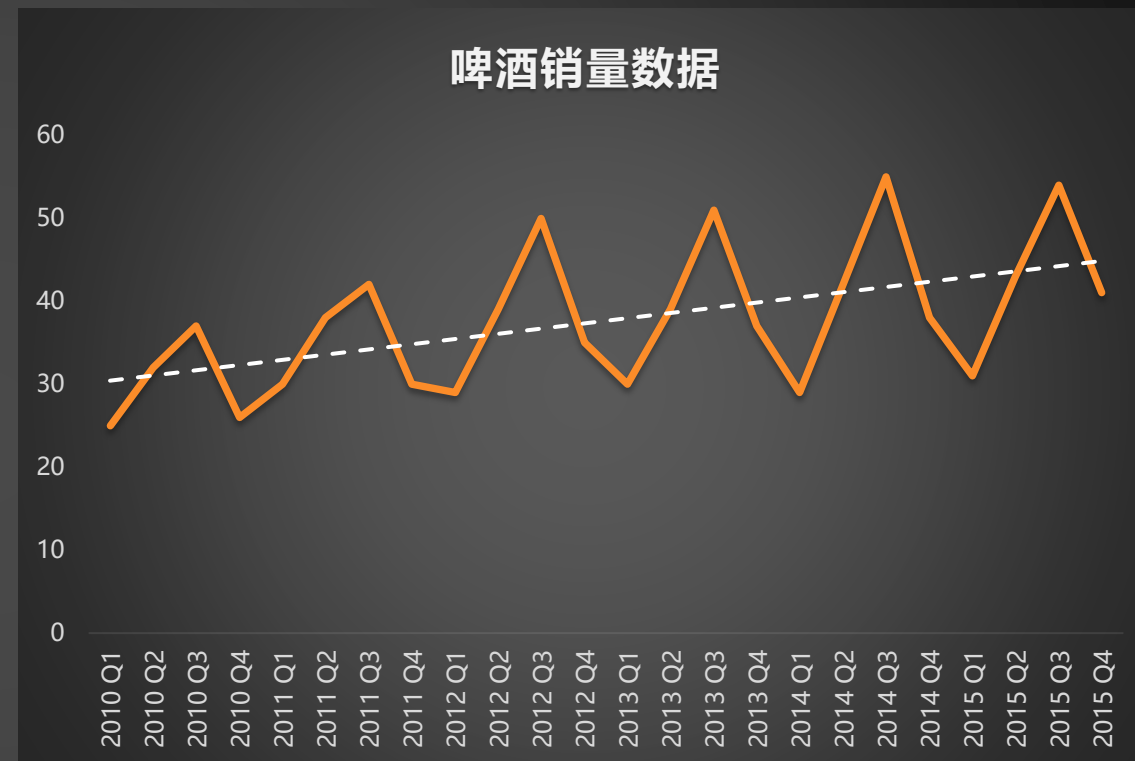
时间序列预测的程序

1. 确定时间序列的成分
2. 选择预测方法
3. 预测方法的评估

1、确定时间序列的成分



趋势

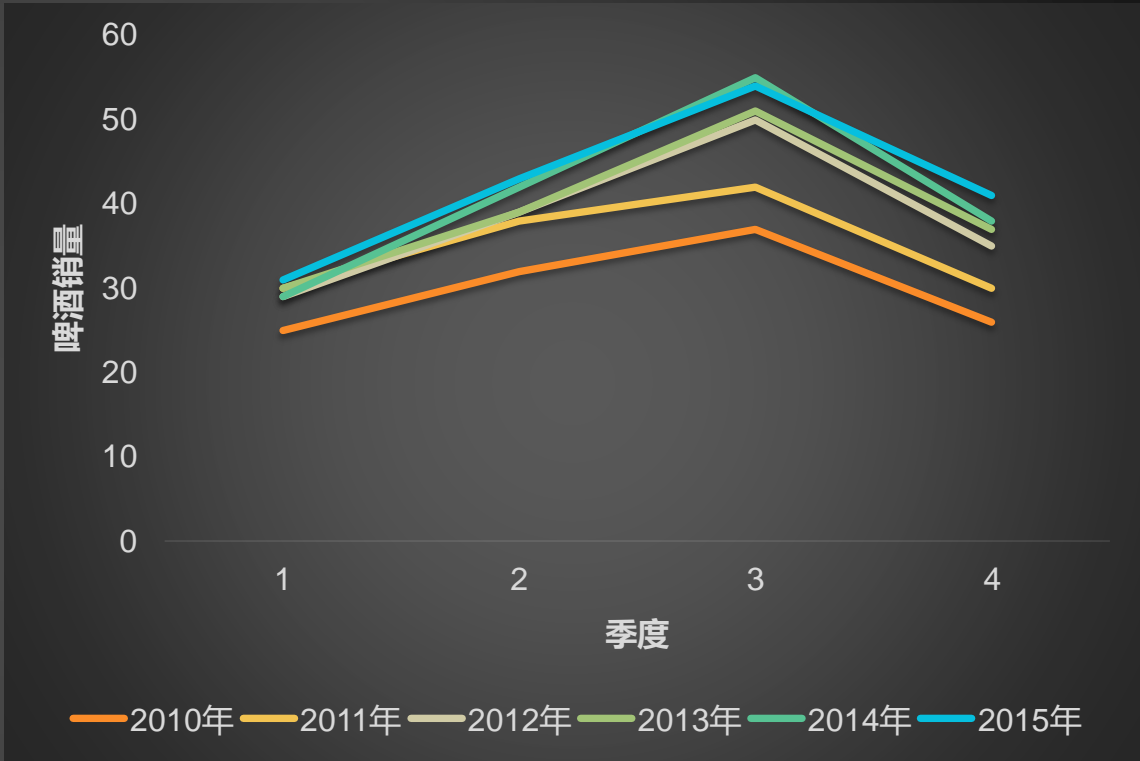


趋势+季节性

确定季节成分是否存在

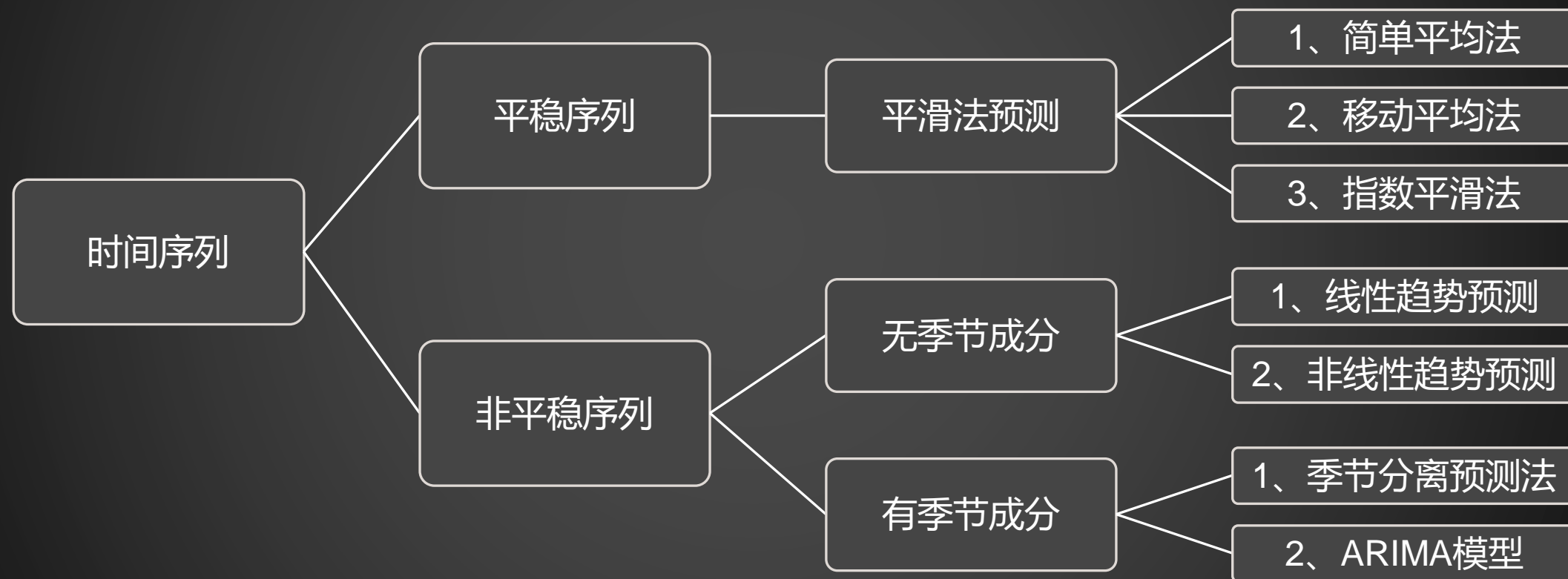
下面是一家啤酒生产企业2010~2015年各季度的啤酒销售量数据。

年份	季度			
	1	2	3	4
2010	25	32	37	26
2011	30	38	42	30
2012	29	39	50	35
2013	30	39	51	37
2014	29	42	55	38
2015	31	43	54	41



年度折叠时间序列图

2、选择预测方法



3、预测方法的评估

评估预测方法一般是误差分析，常见的误差有平均绝对误差、最大绝对误差、均方根误差等。

1、平均绝对误差，Mean Absolute Error, MAE

$$MAE = \frac{\sum_{i=1}^n |Y_i - F_i|}{n}$$

2、最大绝对误差，Max Absolute Error, MaxAE

3、均方根误差，Root Mean Squared Error, RMSE

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - F_i)^2}{n}}$$

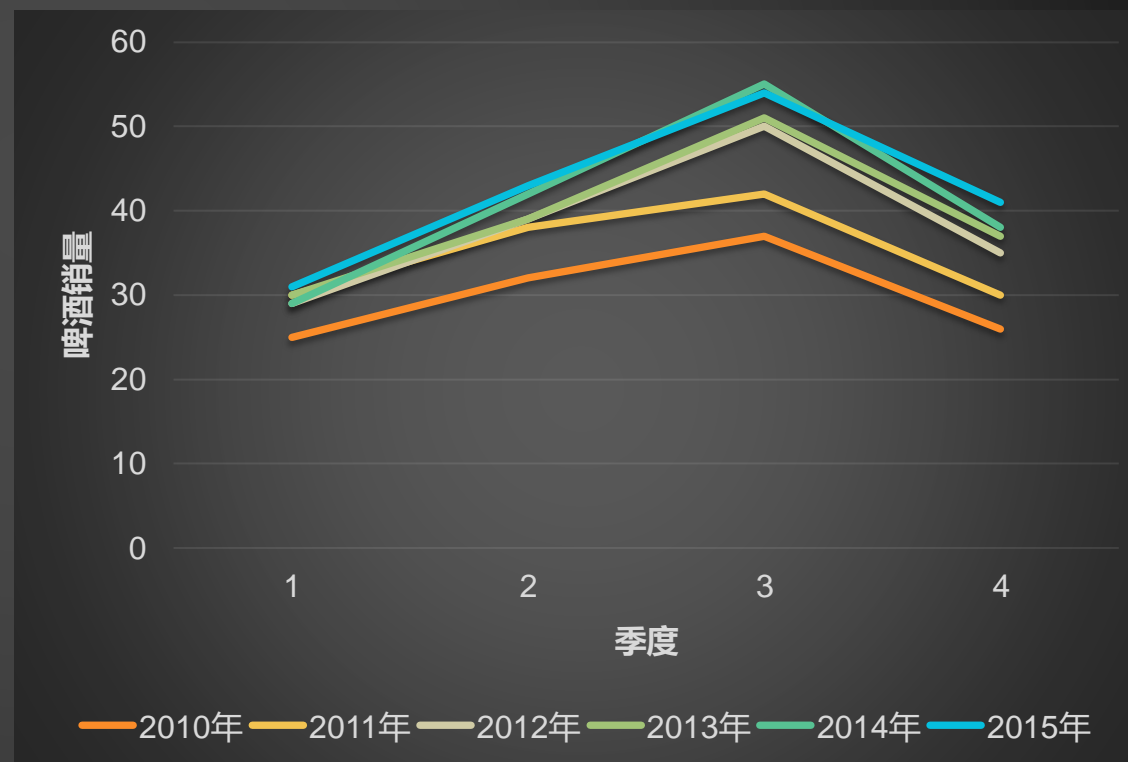
除了误差（越小越好）之外，还可以通过 R^2 及平稳 R^2 （越大越好）

还有，信息准则如BIC（贝叶斯信息准则），AIC（赤池信息准则）等，越小越好。

绘制年度折叠时间序列图

- 操作演示：用Excel绘制年度折叠时间序列图。
- 下面是一家啤酒生产企业2010~2015年各季度的啤酒销售量数据。

年份	季度			
	1	2	3	4
2010	25	32	37	26
2011	30	38	42	30
2012	29	39	50	35
2013	30	39	51	37
2014	29	42	55	38
2015	31	43	54	41

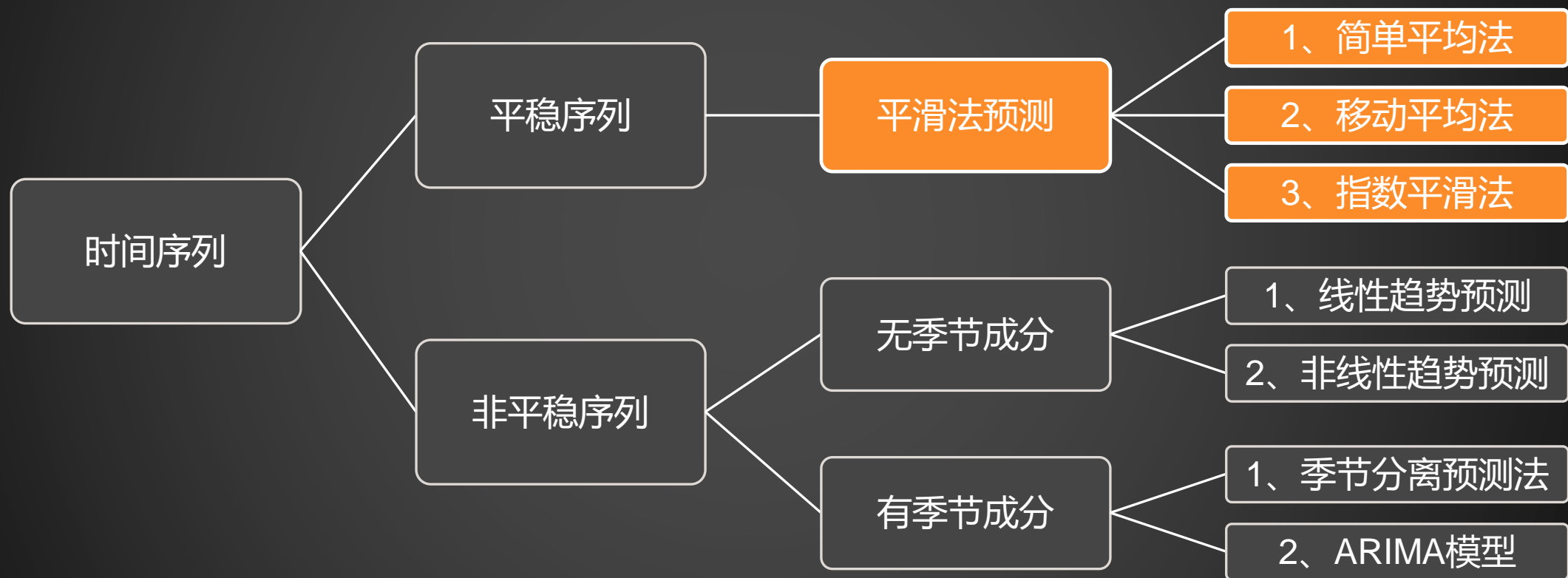


年度折叠时间序列图

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

时间序列预测方法



平滑法预测

使用条件：平稳序列

1、**简单平均法**：根据已有的 t 期观测值通过简单平均来预测下一期的数值。

2、**简单移动平均法**：将最近的 k 期数据加以平均，作为下一期的预测值。

例如 $k = 3$ （移动间隔），则 t 期的移动平均值为：

$$F_{t+1} = \frac{y_{t-2} + y_{t-1} + y_t}{3}$$

3、**指数平滑法**：通过对过去的观测值加权平均进行预测，有一次指数平滑法、二次指数平滑法等。

一次指数平滑法是将一段时期的预测值与观测值的线性组合作为 $t+1$ 期的预测值，计算公式为：

$$F_{t+1} = \alpha y_t + (1 - \alpha)F_t$$

其中， y_t 为 t 期的实际观测值， F_t 为 t 期的预测值， α 为平滑系数（ $0 < \alpha < 1$ ）。

简单移动平均法、指数平滑法实现方法：Excel公式计算、Excel数据分析工具库。

平滑法预测案例

右表为根据居民消费价格指数数据：

1、分别取移动间隔 $k=3$ 和 $k=5$ ，用简单移动平均法预测历史各年份和2014年的居民消费价格指数，计算出预测误差，并将原序列和预测后的序列绘制成图形进行比较。

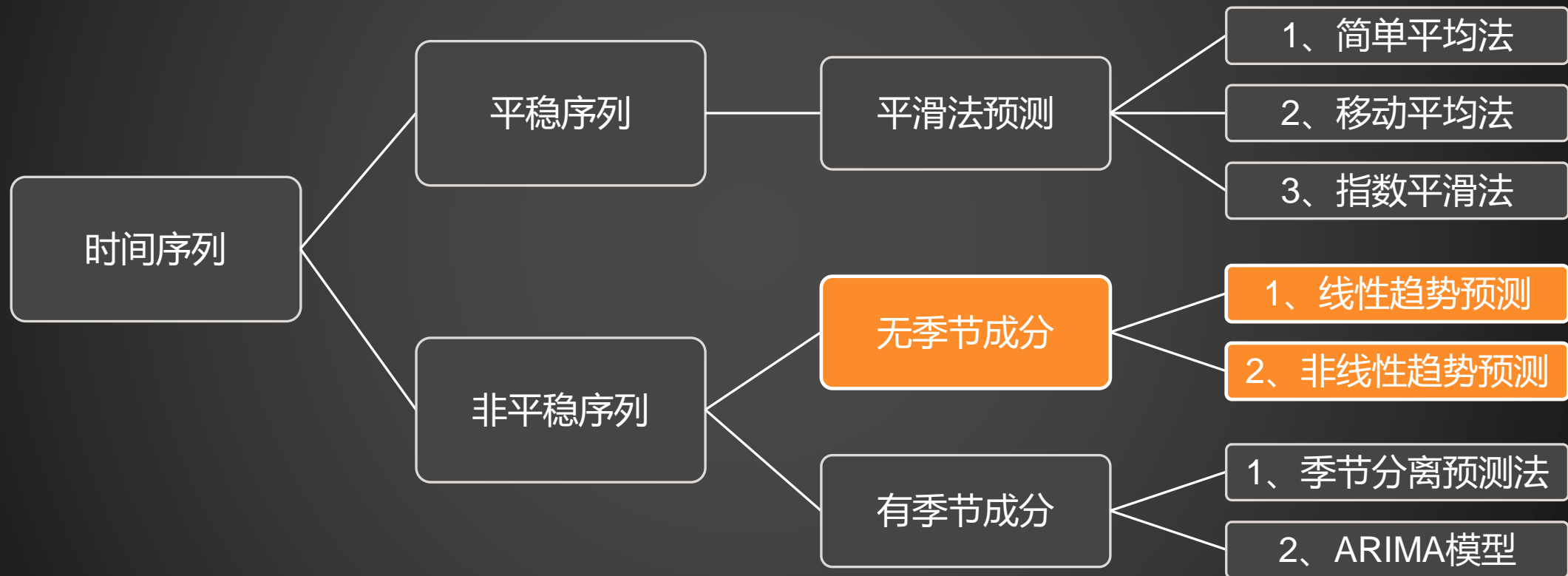
2、选择平滑系数 $\alpha = 0.3$ 和 $\alpha = 0.5$ ，用指数平滑法预测历史各年份和2014年的居民消费价格指数，计算出预测误差，并将原序列和预测后的序列绘制成图形进行比较。

年份	居民消费价格指数（上面=100）
2000	100.4
2001	100.7
2002	99.2
2003	101.2
2004	103.9
2005	101.8
2006	101.5
2007	104.8
2008	105.9
2009	99.3
2010	103.3
2011	105.4
2012	102.6
2013	102.6
2014	

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

时间序列预测方法



线性趋势预测

根据给出的啤酒产量数据，利用回归分析计算出各期的预测值和预测误差，预测2014年的啤酒产量，并将原序列和各期的预测值序列绘制成图形进行比较

操作步骤：

- 1、给时间编号作为x，啤酒产量作为y，做线性回归
- 2、利用回归方程进行预测，并给出预测误差
- 3、将原始数据和预测值绘制成折线图

年份	啤酒产量（万千升）
2000	2231.3
2001	2288.9
2002	2402.7
2003	2540.5
2004	2948.6
2005	3126.1
2006	3543.6
2007	3954.1
2008	4156.9
2009	4162.2
2010	4490.2
2011	4834.5
2012	4778.6
2013	5061.5
2014	

非线性趋势预测

根据下表中的煤炭占能源消费总量的比重数据，拟合适当的趋势曲线，计算出各期的预测值和预测误差，预测2014年的煤炭占能源消费总量的比重，并将原序列和各期的预测值序列绘制成图形进行比较。

操作步骤：

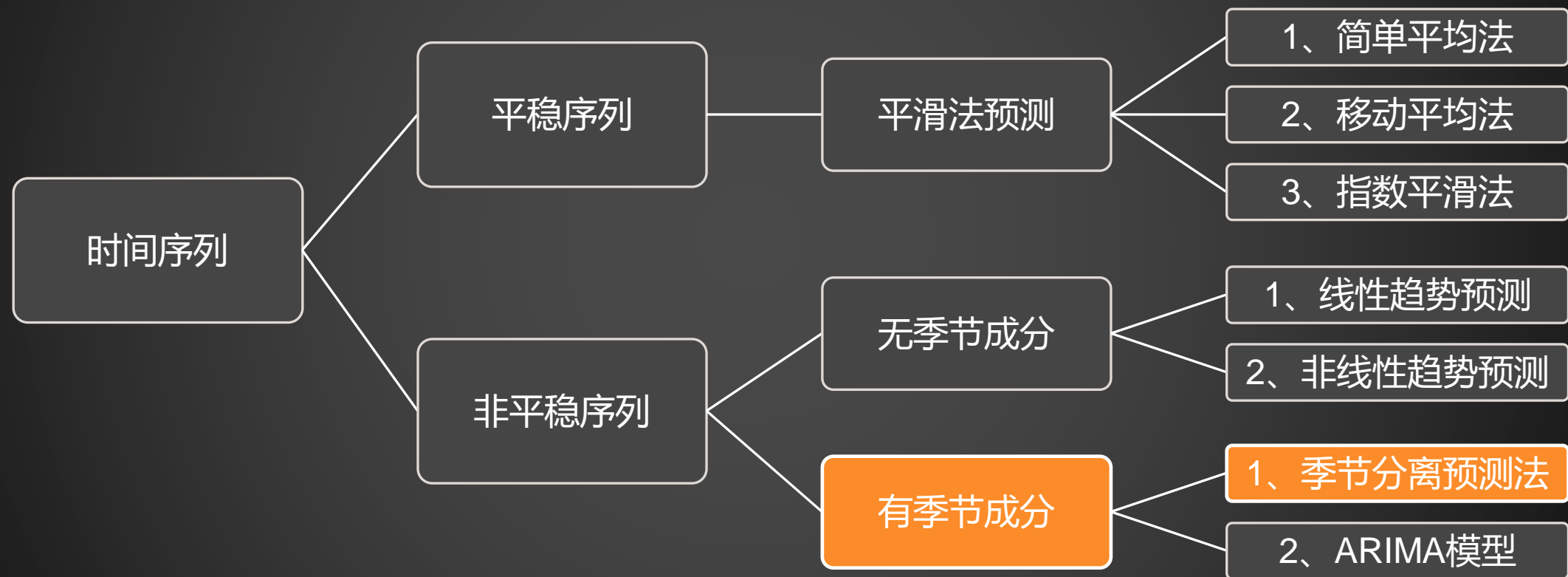
- 1、绘制出原始数据的散点图，判断大致趋势
- 2、利用Excel中的【趋势线选项】进行预测

年份	煤炭占能源消费总量的比重
2000	69.2
2001	68.3
2002	68
2003	69.8
2004	69.5
2005	70.8
2006	71.1
2007	71.1
2008	70.3
2009	70.4
2010	68
2011	68.4
2012	66.6
2013	66
2014	

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

时间序列预测方法



季节分离预测法

下面是一家啤酒生产企业2010~2015年各季度的啤酒销售量数据，预测2016年各个季度的啤酒销量数据。

年份	季度			
	1	2	3	4
2010	25	32	37	26
2011	30	38	42	30
2012	29	39	50	35
2013	30	39	51	37
2014	29	42	55	38
2015	31	43	54	41

思路：将季节成分分离，得到季节分离后的序列，利用线性回归进行预测，再将季节成分乘回去，得到最终预测值。

季节分离预测法

操作步骤：

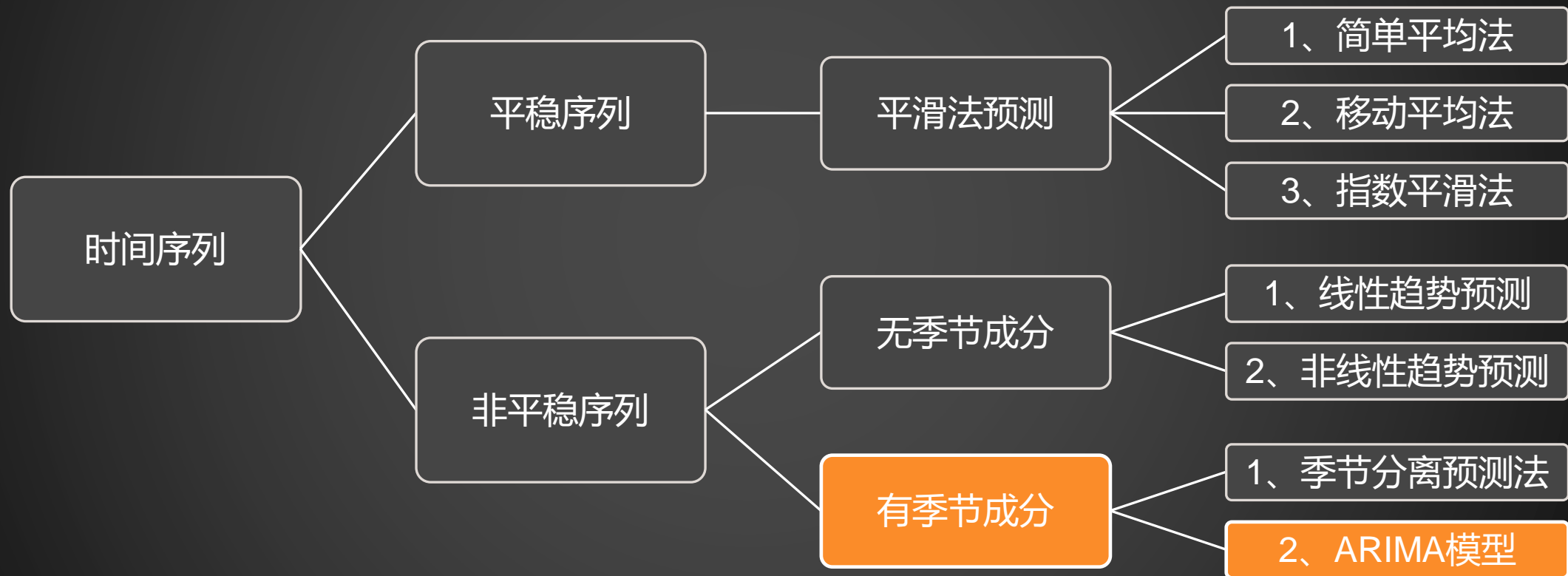
- 1、将原始序列 (Y) 变成列的形式，并对时间进行编号
- 2、对原始序列先做一个4项移动平均，再做一个2项移动平均，得到中心化移动平均值 (CMA)
- 3、求出季节比率 (Y/CMA)，按照季度求出季节比率的均值，得到季节指数 (S)
- 4、分离季节成分，即 Y/S ，得到季节分离后的序列 (线性趋势)
- 5、利用线性回归进行预测，得到回归预测值
- 6、再将季节指数乘回去，即回归预测值*季节指数，得到最终预测值

说明：这里是季度数据，采用4项移动平均，如果是月度数据，则采用12项移动平均。

本章主要内容

1. 时间序列基本概念
2. 差分、增长率分析
3. 时间序列预测的程序
4. 平滑法预测方法
5. 趋势预测方法
6. 季节分离预测法
7. ARIMA模型

时间序列预测方法



ARIMA模型是什么？

- ARIMA是Autoregressive Integrated Moving Average的缩写，ARIMA模型翻译为：自回归移动平均模型。
- ARIMA可以拆分为AR+I+MA三部分，其中AR代表一个模型，MA代表一个模型，I代表差分。
- 由于ARIMA模型是基于平稳序列的，对于非平稳序列，通过差分将其转化为平稳序列。
- 差分前面已经讲过，下面分别介绍这两个模型：MA模型和AR模型。

MA模型

MA是英文moving average的缩写，即移动平均，所以MA模型叫作移动平均模型。

移动平均是指将时间序列表示为一系列不相关的随机变量的线性组合，这一系列不相关的随机变量我们称之为**白噪声**。

用数学语言表达，就是： $y_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t$

其中， y_t 表示第t期的观测值， θ_i 表示加权系数， $\{e_{t-q}\}$ 表示白噪声序列， e_t 表示误差项。

用通俗的话来说，现在的你是由过去不断地犯错所造成的。

$MA(q)$ 表示q阶移动平均。

当q取1时，则为一阶移动平均，用数学语言表达为： $y_t = \theta_1 e_{t-1} + e_t$

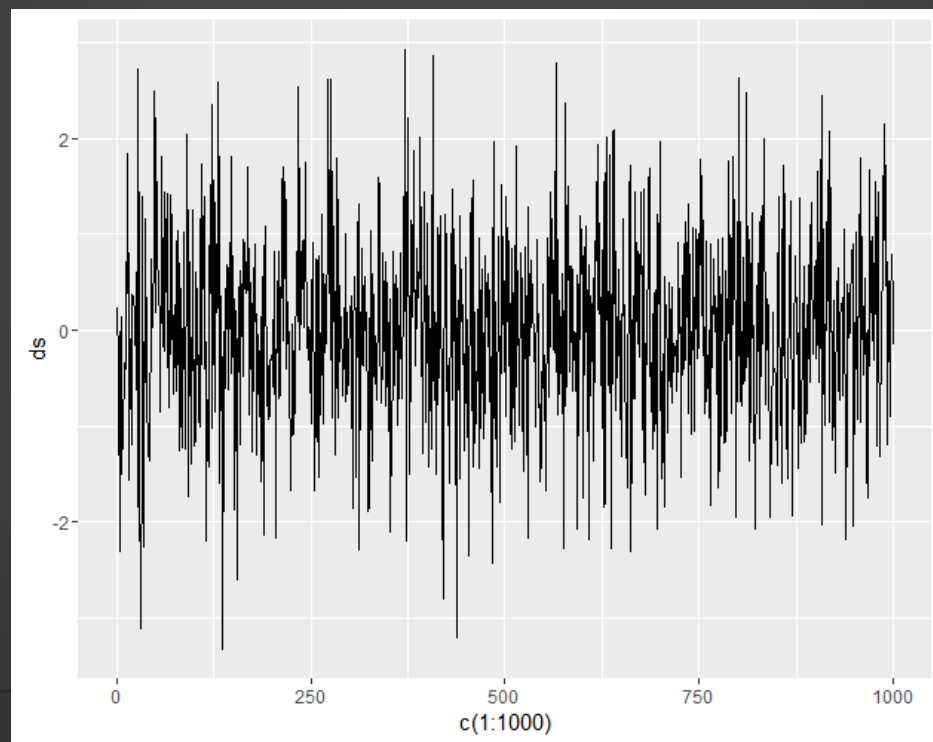
当q取2时，则为二阶移动平均，用数学语言表达为： $y_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t$

一般情况下，三阶移动平均即可搞定问题。

问题：如何确定阶数q？

白噪声

白噪声, white noise, 由白噪声构成的时间序列有如下特点: 完全随机, 无趋势性, 无相关性。对于一个一般的时间序列来说, 去除趋势成分、季节成分等后, 剩下的随机成分可以看成是一个白噪声序列。满足高斯分布 (正态分布) 的白噪声过程称为高斯白噪声过程。

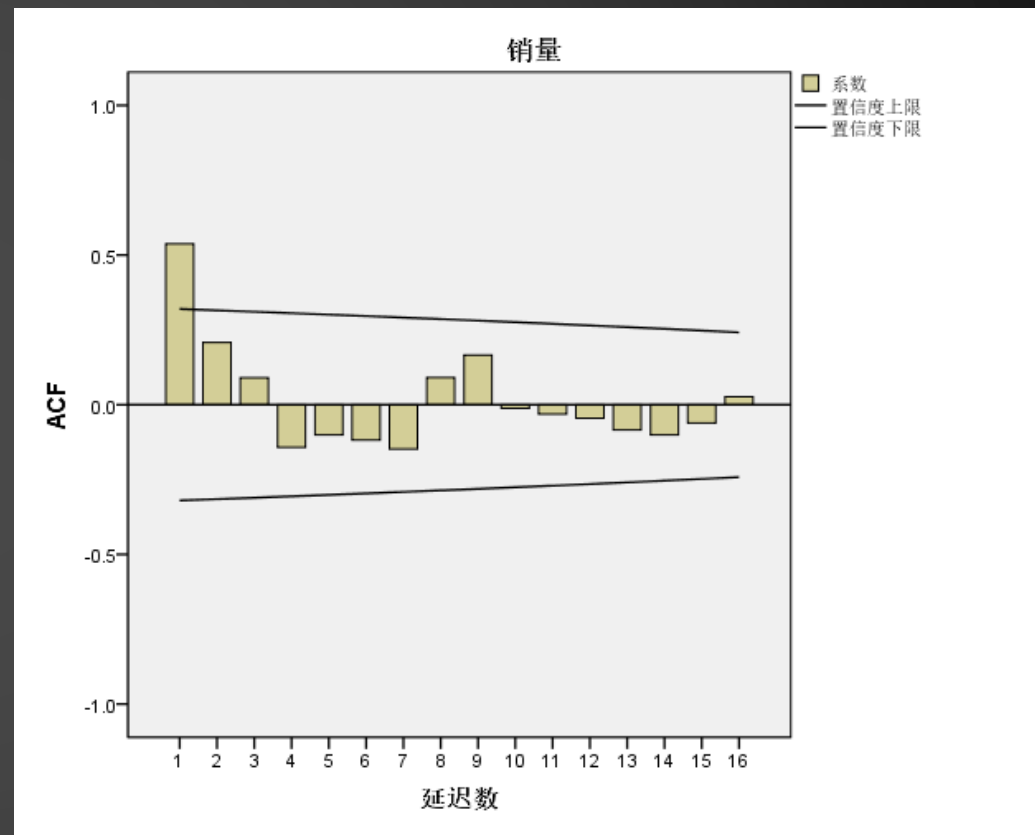


ACF图

MA模型中的阶数 q 由ACF图确定。那什么是ACF?

ACF, autocorrelation function, 自相关函数, 反映了时间序列中不同项之间的相关性, 但是, 对于 $x(t-k)$ 与 $x(t)$, 会受到中间 $k-1$ 个随机变量的干扰。

ACF对应的图像是ACF图, 叫作自相关图, 由SPSS可以得到, 在ACF图中, q 阶后不断减小到零, 这种形态称为截尾。



AR模型

AR是英文autoregressive的缩写，即自回归，所以AR模型叫作自回归模型。

自回归，顾名思义，就是用自身的历史数据去预测自身。用数学语言表达，就是：

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

其中， y_i 表示第*i*期的观测值， ϕ_i 表示加权系数， e_t 表示误差项。

用通俗的话来说，就是现在的你，是由过去的你不断创造的。

$AR(p)$ 表示

阶自回归。

当

取1时，则为一阶自回归，用数学语言表达为： $y_t = \phi_1 y_{t-1} + e_t$

当

取2时，则为二阶自回归，用数学语言表达为： $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$

一般情况下，三阶自回归即可搞定问题。

问题：如何确定阶数

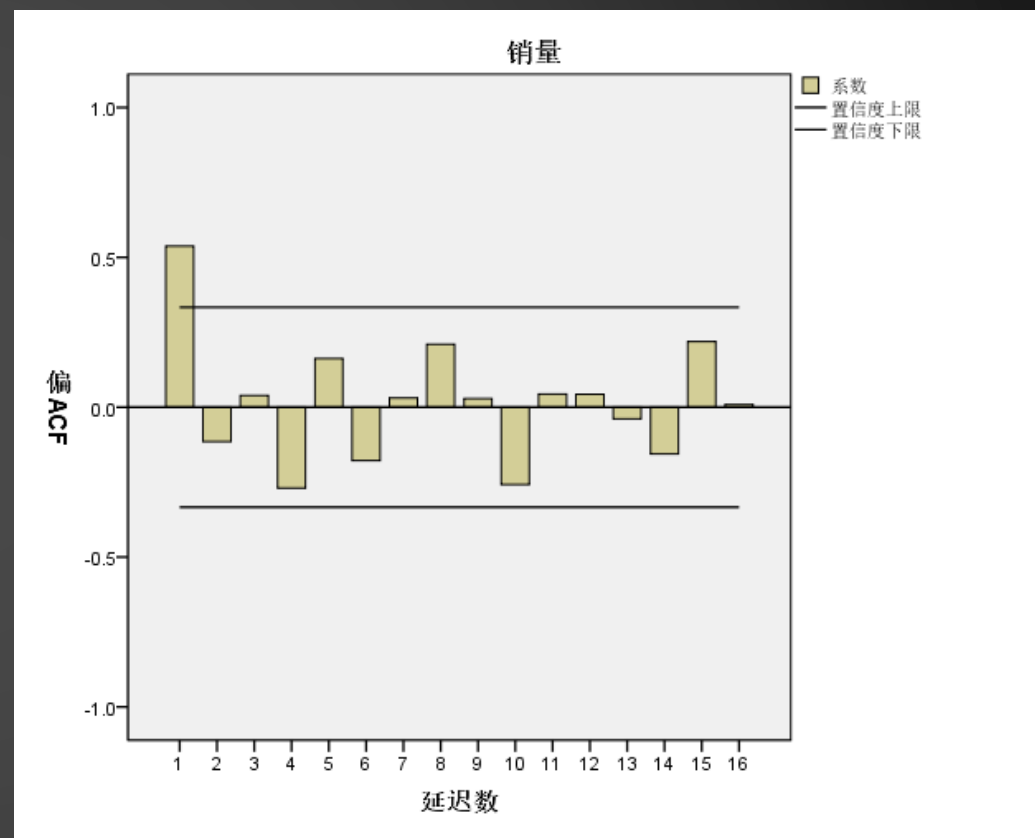
?

PACF图

AR模型中的阶数 p 由PACF图确定。

PACF, partial autocorrelation function, 偏自相关函数, 剔除了中间 $k-1$ 个随机变量的干扰之后 $x(t-k)$ 对 $x(t)$ 影响的相关程度。

PACF对应的图像是PACF图, 偏自相关图, 由SPSS可以得到, 在PACF图中, p 阶后不断震荡, 这种形态称为拖尾。



ARIMA模型操作步骤

ARIMA模型就是将以上所讲的模型综合在一起，形成了一个强大的预测工具。

假定现在有一个非平稳序列需要预测，我们将它交给ARIMA模型。ARIMA模型拿到后，会这样处理：

1. 首先，进行差分，将其转化为一个平稳序列，确定参数 d 。
2. 绘制ACF图和PACF图，确定参数 q 和 p 的值。
3. 建模，SPSS选择ARIMA模型，输入参数，即ARIMA(p, d, q)
4. 最后，根据R方、残差、信息准则AIC/BIC等评估模型。

这就是ARIMA模型的基本步骤，接下来说说ARIMA模型的评估。

ARIMA模型的评估

对于ARIMA模型来说，有以下评估指标：

- R^2 及平稳的 R^2 ：类似于回归分析中的判定系数，越大越好；
- 残差：MAE（平均绝对误差）、MaxAE（最大绝对误差）、RMSE（均方根误差），当然是越小越好；
- 信息准则：BIC（贝叶斯信息准则）、ACI（赤池信息准则）等，越小越好。

说明：以上评估指标，SPSS会自动给出。

ARIMA模型：SPSS实现

主要步骤：

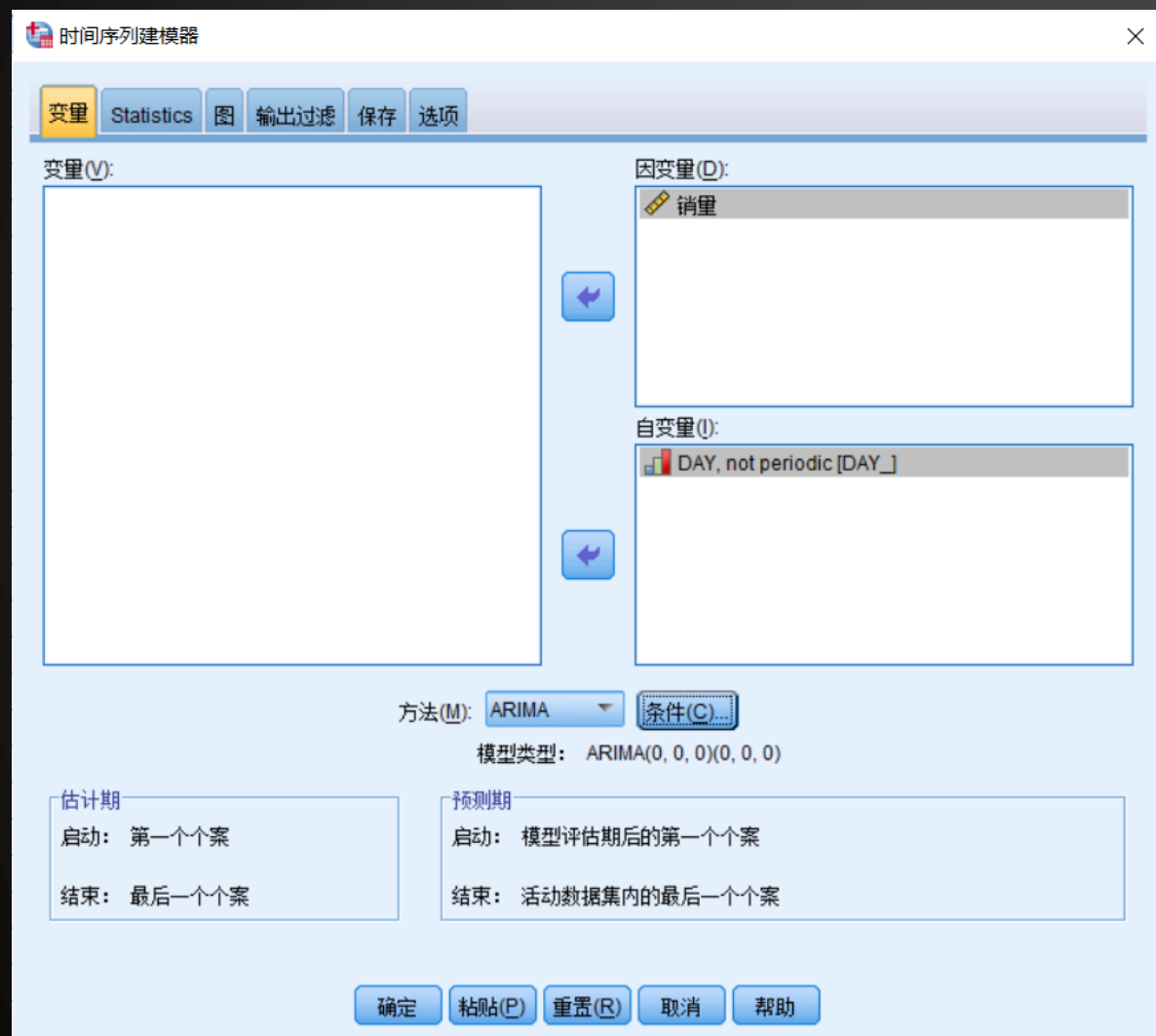
1. 读取数据，并在SPSS中定义日期：天、月份、季度或者年。
2. 差分： SPSS菜单【分析】 - 【预测】 - 【序列图】， 勾选【差分】
3. 确定阶数 q, p ： SPSS菜单【分析】 - 【预测】 - 【自相关】
4. 创建ARIMA模型： SPSS菜单【分析】 - 【预测】 - 【创建模型】， 在建模对话框中， 选择因变量、方法（ARIMA）、误差选项、图表、预测变量以及预测周期
5. 解释分析结果

ARIMA模型案例：餐馆销量预测

现有某餐厅近30多天的营业数据（销量），试预测该餐厅未来7天的销量。

日期	销量
2015/1/1	3023.01
2015/1/2	3039.01
2015/1/3	3056.01
2015/1/4	3138.01
2015/1/5	3188.01
2015/1/6	3224.01
2015/1/7	3226.01
2015/1/8	3029.01
2015/1/9	2859.01
2015/1/10	2870.01

ARIMA模型SPSS实现主要步骤截图



☒ 按模型显示拟合测量、Ljung-Box 统计和离群值的数量(U)

拟合测量

- ☒ 平稳的 R 方(Y) ☒ 平均绝对误差(E)
☒ R 方 ☐ 最大绝对误差百分比(B)
☒ 均方根误差(Q) ☒ 最大绝对误差(X)
☐ 平均绝对误差百分比(P) ☒ 标准化的 BIC(L)

比较模型的统计

- ☐ 拟合优度(G)
☐ 残差自相关函数 (ACF)
☐ 残差部分自相关函数(PACF)(U)

个别模型的统计

- ☒ 参数估计(M)
☒ 残差自相关函数 (ACF)
☒ 残差部分自相关函数(PACF)

☒ 显示预测值(S)

模型比较图

- ☐ 平稳的 R 方(Y) ☐ 最大绝对误差百分比(X)
☐ R 方 ☐ 最大绝对误差(B)
☐ 均方根误差(Q) ☐ 标准化的 BIC(N)
☐ 平均绝对误差百分比(P) ☐ 残差自相关函数(ACF)(U)
☐ 平均绝对误差(M) ☐ 残差部分自相关函数(PACF)

单个模型图

- ☒ 序列(E) ☒ 残差自相关函数 (ACF)
☒ 残差部分自相关函数(PACF)

每张图显示的内容

- ☒ 观察值(O)
☒ 预测值(S)
☒ 拟合值(I)
☒ 预测值的置信区间(V)
☒ 拟合值的置信区间(L)

保存变量

变量(V):

描述	保存	变量名的前缀
预测值	<input checked="" type="checkbox"/>	预测值
置信区间的下限	<input checked="" type="checkbox"/>	LCL
置信区间的上限	<input checked="" type="checkbox"/>	UCL
噪声残值	<input checked="" type="checkbox"/>	NResidual

变量

Statistics

图

输出过滤

保存

选项

预测期

- ☐ 模型评估期后的第一个个案到活动数据集内的最后一个个案(F)
- ☒ 模型评估期后的第一个个案到指定日期之间的个案(C)

日期(D):

日
44

用户缺失值

- ☒ 视为无效(I)
- ☐ 视为有效(V)

置信区间宽度(%) (W):

95

输出中的模型标识前缀(P):

模型

ACF 和 PACF 输出中的显示标签最大数(X):

24

带有季节性的ARIMA模型

对于带有季节性的非平稳序列，在使用ARIMA模型预测时，除了进行差分之外，还需要进行季节差分，才能将其转化为平稳序列。

在绘制ACF图和PACF图时，勾选【差分】、【季节性差分】，一般进行1阶差分

周期长度：

- 对于季度数据，周期长度为4；
- 对于月度数据，周期长度为12。

The screenshot shows the 'Time Series' dialog box in Minitab. The 'Variable (V):' field contains '啤酒销量'. The 'Transform' section has 'Natural Log Transform (N)' unchecked, 'Differencing (D):' checked with a value of 1, and 'Seasonal Differencing (S)' checked with a value of 1. The 'Current Periodicity' is set to 4. The 'Output' section has 'Autocorrelation (A)' and 'Partial Autocorrelation (I)' both checked. The 'Time Series' list on the left includes '时间编号', 'YEAR, not periodic [...]', and 'QUARTER, period 4...'. The 'Options (O)...' button is in the top right. At the bottom are buttons for '确定', '粘贴(P)', '重置(R)', '取消', and '帮助'.

变量(V): 啤酒销量

转换

- ☐ 自然对数转换(N)
- ☒ 差分(D): 1
- ☒ 季节性差分(S): 1

当前周期性: 4

输出

- ☒ 自相关(A)
- ☒ 偏自相关(I)

确定 粘贴(P) 重置(R) 取消 帮助

带有季节性的ARIMA模型

模型

转换函数

界外值

ARIMA 阶数

结构(S):

	非季节性	季节性
自回归(p)	2	2
差分(d)	1	1
移动平均值(q)	3	2

当前周期性:4

转换

- ☒ 无(N)
☐ 平方根(Q)
☐ 自然对数(I)

☒ 在模型中包括常数(I)

继续

取消

帮助

带有季节性的ARIMA模型：啤酒销量预测

下面是一家啤酒生产企业2010～2015年各季度的啤酒销售量数据，预测2016年各个季度的啤酒销量数据。

年份	季度			
	1	2	3	4
2010	25	32	37	26
2011	30	38	42	30
2012	29	39	50	35
2013	30	39	51	37
2014	29	42	55	38
2015	31	43	54	41