**37. 时间序列分析Ⅰ—平稳性及纯随机性检验**

**（一）基本概念**

**一、什么是时间序列？**

为了研究某一事件的规律，依据时间发生的顺序将事件在多个时刻的数值记录下来，就构成了一个时间序列。对时间序列进行观察、研究，找寻它变化发展的规律，预测它将来的发展趋势就是时间序列分析。

例如，国家或地区的年度财政收入，股票市场的每日波动，气象变化，工厂按小时观测的产量等等。

注：随温度、高度等变化而变化的离散序列，也可以看作时间序列。

**二、时间序列的特点**

（1）顺序性；

（2）随机性；

（3）前后时刻（不一定相邻）的依存性；

（4）整体呈趋势性和周期性。

**三、时间序列的分类**

按研究对象的数目：一元时间序列、多元时间序列；

按序列统计特性：平稳时间序列、非平稳时间序列；

按分布规律：高斯时间序列、非高斯时间序列。

**四、研究方法**

1. 平稳时间序列分析；

2. 非平稳时间序列分析（确定性分析、随机性分析）。

**五、其它**

任何时间序列经过合理的函数变换后都可以被认为是由下列三部分叠加而成：

（1）趋势项部分；

（2）周期项部分；

（3）随机项部分（随机信号、随机噪声）

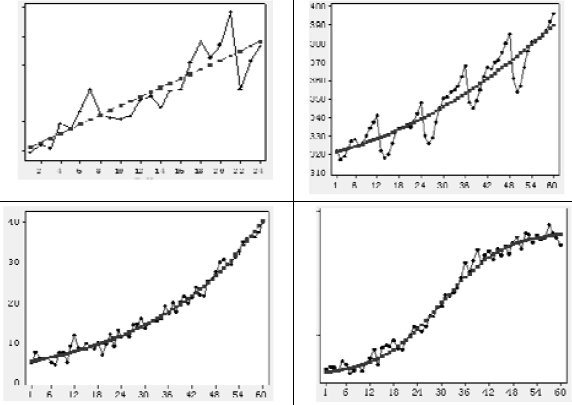


图1. 四种趋势：线性、二次、指数增长、S型

例如，手机销售的月记录按年增长（趋势项）；按季节周期波动（周期项）；随机信号和随机噪声。

时间序列分析的主要任务就是：上面三部分分解出来，是研究平稳随机过程的变化规律，建立特定的ARIMA 模型（要求大体平稳、可能含有周期但不能有规则性的线性指数等类型趋势项）。

**六、方法性工具**

1. 差分运算

（1）k步差分

间隔k期的观察值之差：Δk=*xt-xt-k*

（2）p阶差分

Δ*xt*=*xt*-*xt-*1称为一阶差分；

称为p阶差分；

SAS函数实现：diff*n*(*x* )

2. 延迟算子

延迟算子作用于时间序列，时间刻度减小1个单位（序列左移一位）： B*xt*=*xt-*1, ……, B*pxt*=*xt-p*.

SAS函数实现：lag*n*(*x*)

用延迟算子表示k步差分和p阶差分为：

Δk=*xt-xt-k*=(1-Bk) *xt*



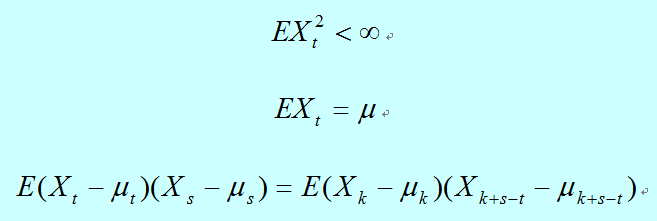
**（二）平稳时间序列**

**一、概念**

平稳时间序列按限制条件的严格程度，分为

严平稳时间序列：序列所有的统计性质都不会随着时间的推移而发生变化；

宽平稳时间序列：序列的主要性质近似稳定，即统计性质只要保证序列的二阶矩平稳，即对任意的时间*t，s，k*，序列*Xt*满足：



**二、平稳时间序列的统计性质**

（1）均值为常数；

（2）自协方差只依赖于时间跨度；

若定义自协方差函数为

γ(*t*,*s*) = E(*Xt*-*μt*)( *Xs*-*μs*)

则可由二元函数简化为一元函数γ(*t*-*s*)，得延迟k自协方差函数：

γ(*k*)= γ(*t*,*t*+*k*)

由此易知平稳时间序列必具有常数方差：

D(Xt)= E(*Xt*-*μt*)2=γ(*t*,*t*)= γ(0)

时间序列自相关函数：



延迟k自相关函数：



基本性质：

（1）ρ(0)=1;

（2）ρ(-k)= ρ(k);

（3）自相关阵为对称负定阵；

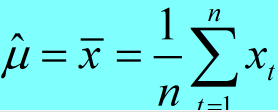
（4）非唯一性。

注意：协方差函数和相关函数——度量两个不同事件（Xt，Yt）彼此之间的相互影响的程度。

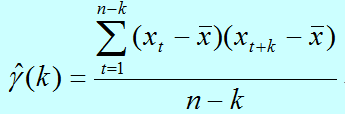
自协方差函数和自相关函数——度量用一事件（Xt）在两个不同时期之间的相互影响的程度。

**三、样本估计值**

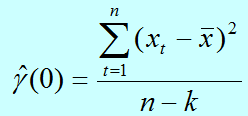
总体均值的估计值：



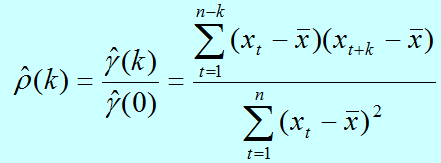
延迟k自协方差函数的估计值：



总体方差的估计值：



延迟k自相关函数的估计值：



**四、平稳性检验**

（1）时序图检验

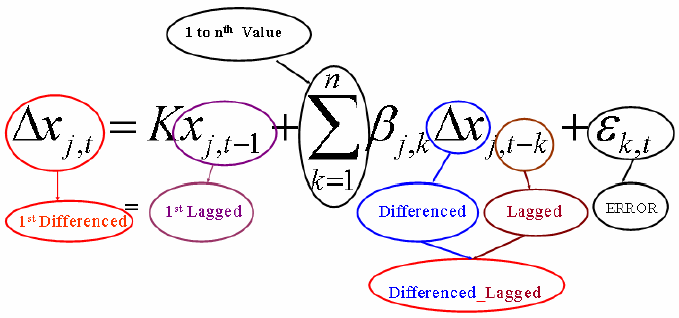
若无明显的趋势性和周期性，则平稳；

（2）自相关图检验

零均值平稳序列的自相关函数要么截尾要么拖尾；若时间序列零均值化后出现缓慢衰减或周期性衰减，则说明存在趋势性和周期性（非平稳）；

（3）单位根检验就是通过检验时间序列自回归特征方程的特征根是在单位圆内（平稳）还是在单位圆及单位圆外（非平稳）。通常用ADF检验法。

Dickey和Fuller (1979)利用如下的广义自回归模型



其中，Δ*xj,t*表示*x*的一阶差分；*xj,t-*1表示延迟一期；Δ*xj,t-k*表示延迟k期再一阶差分；*εk,t*表示扰动项。

上述回归模型生成的*xj,t-*1的t值正好对应ADF统计量，做假设检验：H0: 非平稳；H1：平稳。t值在1%, 5%, 10% 置信水平的临界值分别为：-3.524233, -2.902358, -2.588587. 以此判断序列是否平稳。

注：若Xt不平稳，可以依次对Xt做一阶、二阶…差分，直到序列平稳。

**例1**. 平稳性检验——ADF检验的SAS实现。

代码：

**data** simulation;

do i=**1** to **100**;

x=rannor(**1234**);

output;

end;

**run**;

**data** timeseries;

set simulation;

x\_1st\_lag= lag1(x);

x\_1st\_diff= dif1(x);

x\_1st\_diff\_1st\_lag= dif1(lag1(x));

x\_1st\_diff\_2nd\_lag= dif1(lag2(x));

x\_1st\_diff\_3rd\_lag= dif1(lag3(x));

x\_1st\_diff\_4th\_lag= dif1(lag4(x));

x\_1st\_diff\_5th\_lag= dif1(lag5(x));

**run**;

**proc** **reg** data=timeseries;

model x\_1st\_diff = x\_1st\_lag

x\_1st\_diff\_1st\_lag

x\_1st\_diff\_2nd\_lag

x\_1st\_diff\_3rd\_lag

x\_1st\_diff\_4th\_lag

x\_1st\_diff\_5th\_lag;

**run**;

运行结果：

REG 过程

模型: MODEL1

因变量: x\_1st\_diff

|  |  |
| --- | --- |
| **读取的观测数** | 100 |
| **使用的观测数** | 94 |
| **具有缺失值的观测数** | 6 |

| **方差分析** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **源** | **自由度** | **平方和** | **均方** | **F 值** | **Pr > F** |
| **模型** | 6 | 111.38082 | 18.56347 | 15.25 | <.0001 |
| **误差** | 87 | 105.88424 | 1.21706 |  |  |
| **校正合计** | 93 | 217.26507 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **均方根误差** | 1.10320 | **R 方** | 0.5126 |
| **因变量均值** | 0.02507 | **调整 R 方** | 0.4790 |
| **变异系数** | 4399.76165 |  |  |

| **参数估计值** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **参数估计值** | **标准误差** | **t 值** | **Pr > |t|** |
| **Intercept** | **1** | -0.01634 | 0.11418 | -0.14 | 0.8866 |
| **x\_1st\_lag** | **1** | -0.70975 | 0.20949 | -3.39 | 0.0011 |
| **x\_1st\_diff\_1st\_lag** | **1** | -0.26217 | 0.19212 | -1.36 | 0.1759 |
| **x\_1st\_diff\_2nd\_lag** | **1** | -0.15780 | 0.17907 | -0.88 | 0.3806 |
| **x\_1st\_diff\_3rd\_lag** | **1** | -0.01973 | 0.16308 | -0.12 | 0.9040 |
| **x\_1st\_diff\_4th\_lag** | **1** | 0.07067 | 0.13938 | 0.51 | 0.6134 |
| **x\_1st\_diff\_5th\_lag** | **1** | 0.00340 | 0.10591 | 0.03 | 0.9745 |

x\_1st\_lag的t值 = -3.39 < t0.05=-2.902358, （或从P值 = 0.0011 < 0.05判断）故拒绝原假设H0，即序列平稳。

**五、纯随机性检验**

若序列值彼此之间没有任何相关性，即过去的行为对未来的发展没有丝毫影响，此时称为纯随机序列。

从统计分析的角度而言，纯随机序列是没有任何分析价值的序列。因此，为了确保平稳序列还值不值得分析，还需要对平稳序列进行纯随机性检验。

1. 纯随机序列（白噪声序列）

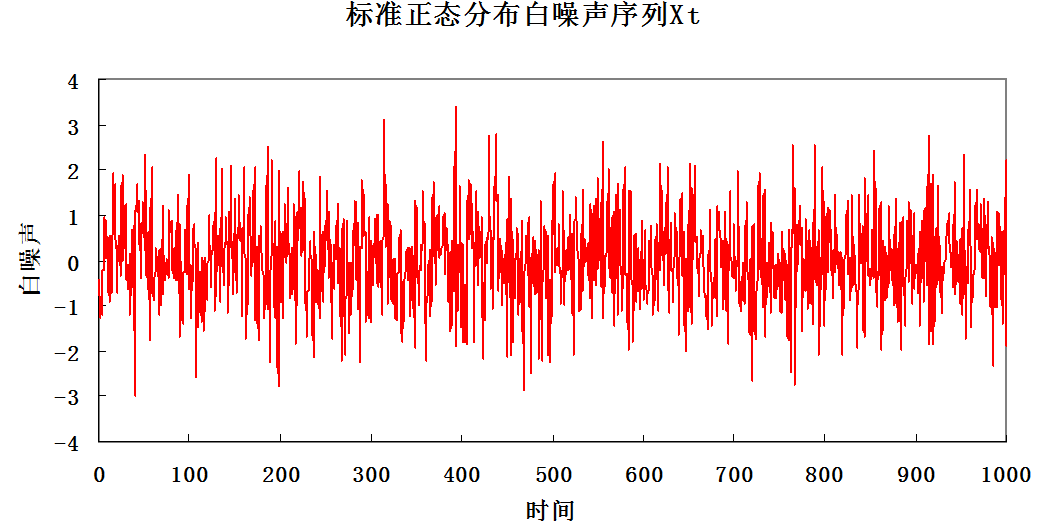
若对任取的时间t和s，时间序列Xt满足：

（1）E(Xt) = μ；（常数均值）

（2）r(t,s) =σ2， 若t=s；（方差齐性）

（3）r(t,s) =0， 若t≠s. （纯随机性）

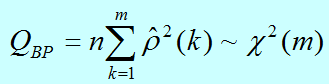
则称Xt为纯随机序列或白噪声序列（白光具有该特性），简记为Xt～WN(μ, σ2)。白噪声序列是最简单的平稳时间序列。随机生成的1000个服从标准正态分布的白噪声序列观察值：

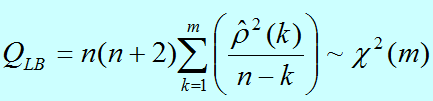


2. 纯随机性检验

Barlett证明：n个观察值的纯随机时间序列，延迟为k（≠0）的自相关函数ρ(k) 近似服从正态分布N(0,1/n).

由此可以构造QBP统计量（适合样本数n≥50）和QLB统计量（适合小样本）来检验序列的纯随机性：





再做假设检验：

H0: ρ(1)= ρ(2)=…=ρ(m)，即延迟≤m的序列之间相互独立；

H1: 至少有一个ρ(k)≠0，即延迟≤m的序列之间有相关性。

注：m一般取值为6、12。这是因为平稳序列通常具有短期相关性，只要序列时期足够长，自相关系数都会收敛于零。

**例2.** 数据如下表，时间间隔为天，起始时间自定义。

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **10** | **15** | **10** | **10** | **12** | **10** | **7** | **7** | **10** | **14** | **8** | **17** |
| **14** | **18** | **3** | **9** | **11** | **10** | **6** | **12** | **14** | **10** | **25** | **29** |
| **33** | **33** | **12** | **19** | **16** | **19** | **19** | **12** | **34** | **15** | **36** | **29** |
| **26** | **21** | **17** | **19** | **13** | **20** | **24** | **12** | **6** | **14** | **6** | **12** |
| **9** | **11** | **17** | **12** | **8** | **14** | **14** | **12** | **5** | **8** | **10** | **3** |
| **16** | **8** | **8** | **7** | **12** | **6** | **10** | **8** | **10** | **5** |  |  |

（1）判断该序列*xt*的平稳性及纯随机性；

（2）判断*xt*的一阶差分*yt*的平稳性及纯随机性。

代码：

**data** datas1;

input x\_t @@;

time=intnx('day',**'01jan2014'd**,\_n\_-**1**);

format time monyy.;

cards;

10 15 10 10 12 10 7 7 10 14 8 17

14 18 3 9 11 10 6 12 14 10 25 29

33 33 12 19 16 19 19 12 34 15 36 29

26 21 17 19 13 20 24 12 6 14 6 12

9 11 17 12 8 14 14 12 5 8 10 3

16 8 8 7 12 6 10 8 10 5

;

**run**;

**proc** **gplot** data = datas1;

plot x\_t\*time;

symbol i=join v=star cv=red ci=green;

**run**;

**proc** **arima** data = datas1;

identify var=x\_t nlag=**24**;

**run**;

**data** datas2;

set datas1;

y\_t = dif1(x\_t);

**run**;

**proc** **gplot** data = datas2;

plot y\_t\*time;

symbol i=join v=star cv=red ci=green;

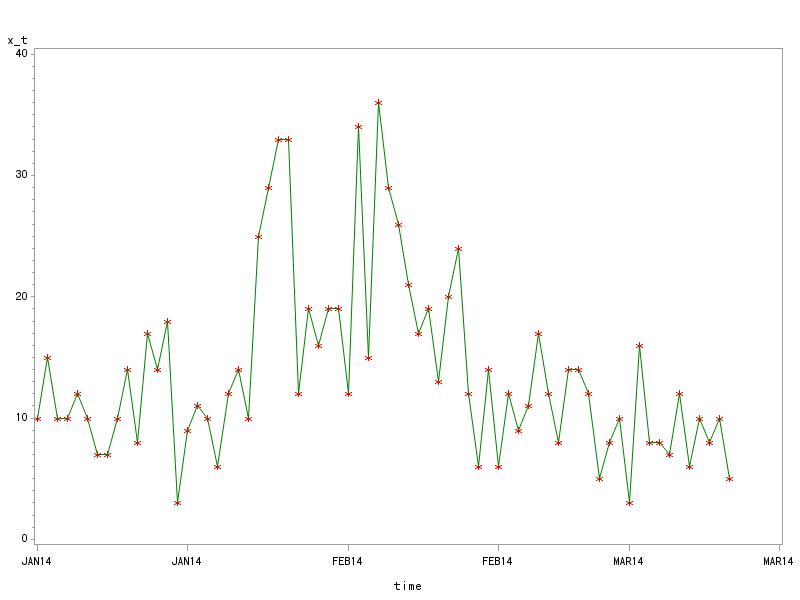
**run**;

**proc** **arima** data = datas2;

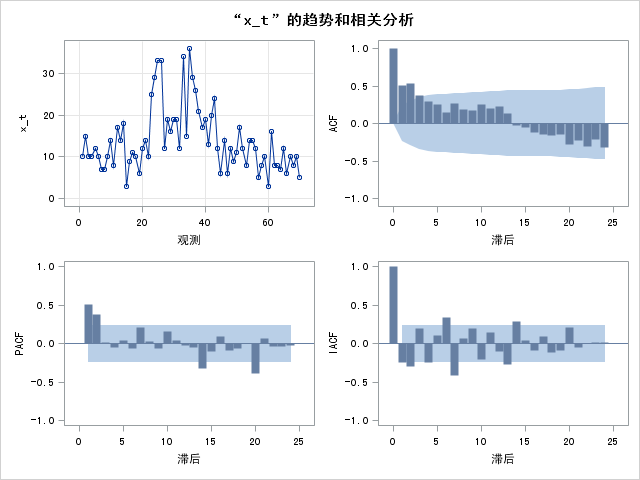
identify var=y\_t nlag=**24**;

**run**;

运行结果：



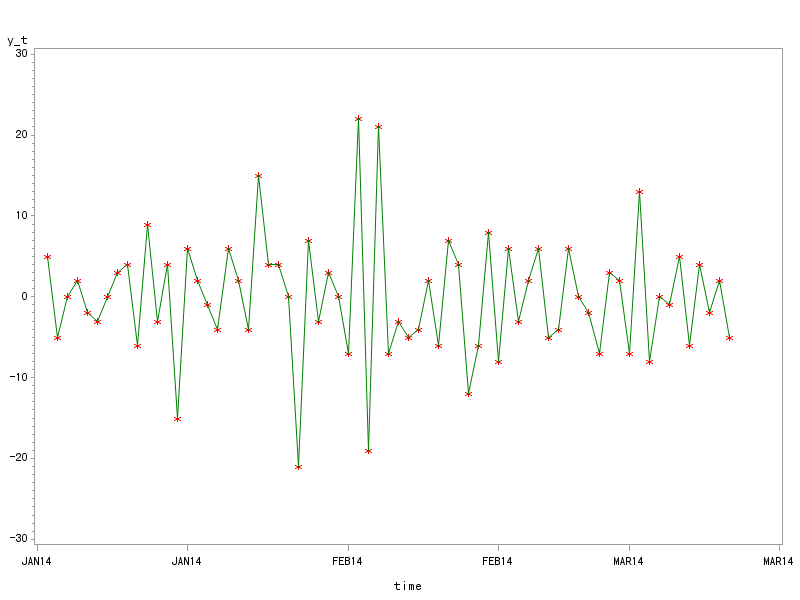
从时序图看，Xt有明显的周期性和递增递减趋势，故不平稳。



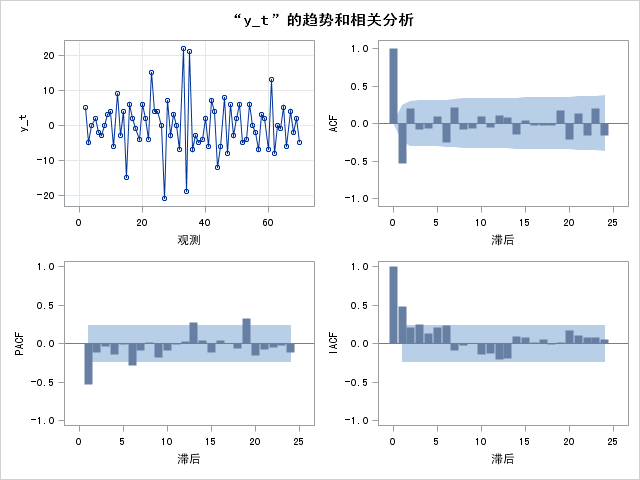
从ACF图看，Xt的自相关系数递减到零的速度相当缓慢，在很长的延迟时期里，自相关系数一直为正，而后又一直为负，故判断该序列非平稳。

| **白噪声的自相关检查** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **至滞后** | **卡方** | **自由度** | **Pr > 卡方** | **自相关** | | | | | |
| **6** | 64.02 | 6 | <.0001 | 0.506 | 0.539 | 0.374 | 0.291 | 0.258 | 0.148 |
| **12** | 88.98 | 12 | <.0001 | 0.270 | 0.186 | 0.178 | 0.258 | 0.207 | 0.226 |
| **18** | 96.32 | 18 | <.0001 | 0.138 | -0.027 | -0.053 | -0.112 | -0.139 | -0.155 |
| **24** | 137.26 | 24 | <.0001 | -0.145 | -0.284 | -0.229 | -0.306 | -0.211 | -0.313 |

延迟为6、12的检验P值均小于0.05，故拒绝原假设，认为Xt为非纯随机序列（非白噪声序列）。



Yt的时序图波动范围有界且没有明显的周期性、递增（递减）趋势，故可以初步判断该序列平稳。



从ACF自相关图看，延迟1阶后的样本自相关系数很快衰减到零附近，且1阶后的样本自相关系数均落在了两倍标准误的范围之内，且在零值附近波动，故可认为Yt平稳。

| **白噪声的自相关检查** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **至滞后** | **卡方** | **自由度** | **Pr > 卡方** | **自相关** | | | | | |
| **6** | 29.46 | 6 | <.0001 | -0.529 | 0.195 | -0.080 | -0.059 | 0.092 | -0.256 |
| **12** | 35.94 | 12 | 0.0003 | 0.216 | -0.075 | -0.070 | 0.101 | -0.048 | 0.104 |
| **18** | 38.61 | 18 | 0.0032 | 0.075 | -0.142 | 0.045 | -0.032 | -0.026 | -0.022 |
| **24** | 57.43 | 24 | 0.0001 | 0.173 | -0.214 | 0.129 | -0.158 | 0.195 | -0.165 |

延迟为6、12的检验P值均小于0.05，故拒绝原假设，认为Yt为非纯随机序列（非白噪声序列）。