时间序列分析实验报告书

班级：统计 2001 姓名：张逸敏　 实验日期：2023.4.11

实验一　时间序列的预处理

1 实验目的

掌握EVIEWS 软件的数据录入，平稳性检验，纯随机性检验。

2 实验条件

PC 机，R 语言

3 实验原理

根据平稳时间序列均值、方差为常数的性质，平稳序列的时序图应该显示出该序列始终在一个常数值附近随机波动，而且波动的范围有界的特点，如果序列的时序图表现出该序列有明显的趋势性或者周期性，那它通常不是平稳序列。根据这个性质，对于很多非平稳序列，可以通过查看其时序图将识别出来。

根据$Baelett$定理，可以构造统计量来检验序列的随机性。

一、　假设条件

原假设：延迟期数小于或者等于m期的序列值之间相互独立。

备择假设：延迟期数小于或者等于m期的序列值之间有相关性。

二、　检验统计量

$$Q_{BP}=n\sum_{k=1}^{m}\hat{\rho}_k^2$$

$$Q_{LB}=n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right)$$

式中，n为序列观察期数；m为指定延迟期数。

由于Q_{BP}统计量比较适用于大样本场合，在小样本场合不太适用，所以本案例中使用Q_{LB}统计量。

Box 和 Ljung 证明Q_{LB}统计量近似服从自由度为m的卡方分布。

当Q_{LB}统计量大于$\chi_{1-\alpha}^2(m)$分位点或者该统计量的p值小于α时，则可以以$1-\alpha$的置信水平拒绝原假设，认为该序列为非白噪声序列；否则，不能拒绝原假设，认为该序列为纯随机序列。

4 实验过程与结果

4.1 实验案例表述

对 1969 年 1 月至 1973 年 9 月在芝加哥海德公园内每 28 天发生的抢包案件数进行平稳性和纯随机性检验，原始数据截图如下：

6. 1969 年 1 月至 1973 年 9 月在芝加哥海德公园内每 28 天发生的抢包案件数见表 2—10（行数据）。

表 2—10

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 $\{x_t\}$ 的平稳性及纯随机性。

（2）对该序列进行函数运算：

$$y_t = x_t - x_{t-1}$$

并判断序列 $\{y_t\}$ 的平稳性及纯随机性。

4.2 实验过程与代码

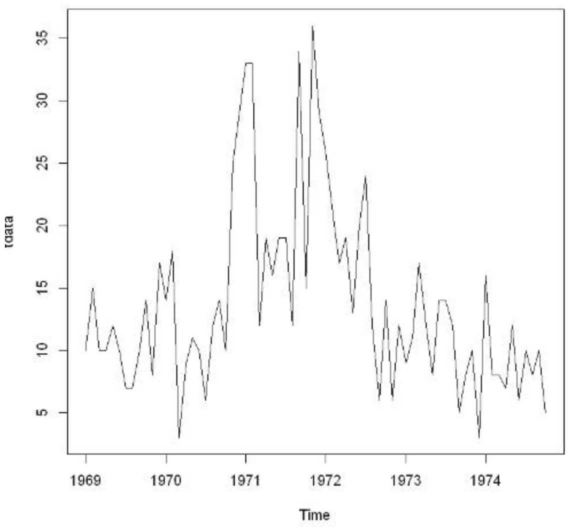
4.2.1 判断$\{x_t\}$的平稳性

首先读入数据，创建时间序列对象。

```r
# 导入数据，创建时间序列对象
data = c(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)
tdata = ts(data, start=c(1969,1), frequency = 12)
```

接下来绘制时序图。

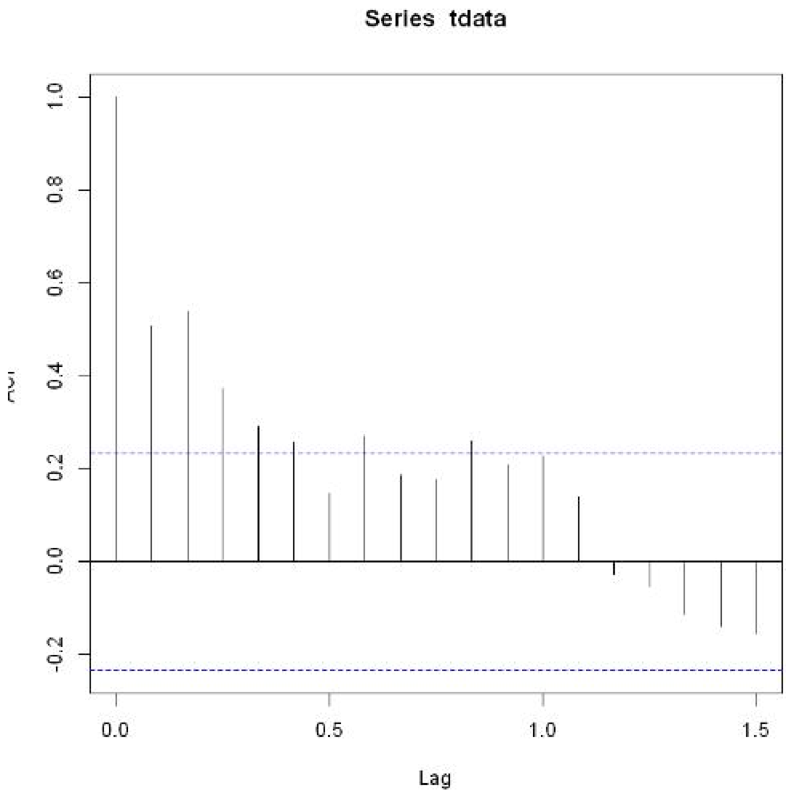
```r
# 绘制时序图
plot(tdata)
```



观察时序图，没有发现明显的趋势或者周期，基本上可以视为平稳序列。

为了稳妥起见，还需要进行自相关图检验。

```
# 自相关图检验
acf(tdata)
```



自相关图显示大多数自相关系数在 2 倍标准差以外，说明序列具有短期相关性，是平稳序列。

4.2.2 判断$\{x_t\}$的纯随机性

利用 LB 统计量进行纯随机性检验，结果如下。

```
Box.test(tdata, type="Ljung-Box", lag=6)
Box.test(tdata, type="Ljung-Box", lag=12)
```

```
	Box-Ljung test

data:  tdata
X-squared = 64.016, df = 6, p-value = 6.85e-12


	Box-Ljung test

data:  tdata
X-squared = 88.975, df = 12, p-value = 7.794e-14
```

延迟 6 期的 LB 统计量和延迟 12 期的 LB 统计量的 p 值均<0.05，表示延迟期数

小于等于 6、12 的序列值之间存在相关性，时间序列不具有纯随机性，值得继续进行建模。

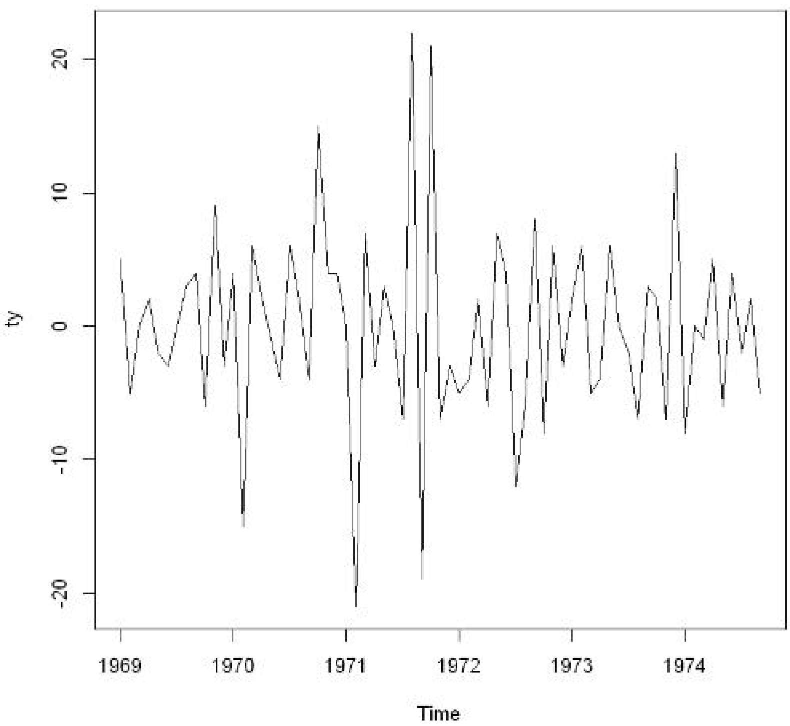
4.2.3 判断$\{y_t = x_t - x_{t-1}\}$的平稳性

进行差分运算，构造 y 数组。

```
n = length(data)
y = data[2:n] - data[1:n - 1]
ty = ts(y, start = c(1969,1), frequency = 12)
```

绘制时序图。

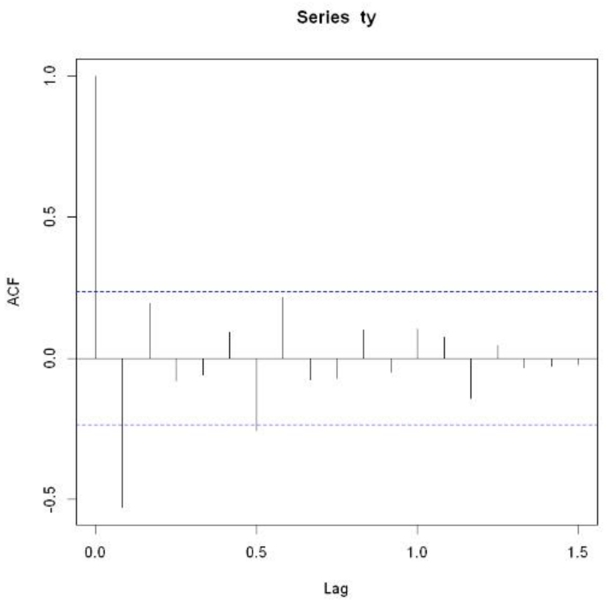
```
plot(ty)
```



观察时序图可以看到序列值在 0 附近随机波动，可以认为是平稳序列。

进行自相关图检验。

```
acf(ty)
```



自相关系数很快趋向于零，可以认为是平稳序列。

4.2.4 判断$\{y_t = x_t - x_{t-1}\}$的纯随机性

利用 LB 统计量进行纯随机性检验。

```
Box.test(ty, type="Ljung-Box", lag=6)
Box.test(ty, type="Ljung-Box", lag=12)
```

```
        Box-Ljung test

data:  ty
X-squared = 29.458, df = 6, p-value = 4.982e-05


        Box-Ljung test

data:  ty
X-squared = 35.943, df = 12, p-value = 0.0003308
```

延迟 6 期的 LB 检验和延迟 12 期的 LB 检验的 p 值均<0.05，表示延迟期数小于等于 6、12 的序列值之间存在相关性，时间序列不具有纯随机性，值得继续进行建模。

5 实验分析与总结

5.1 实验分析

（1）时序图显示芝加哥海德公园每 28 天的抢包案件数大多在 10 到 20 之间，波动是比较平稳的。

（2）自相关图检验，考察该样本序列的自相关图，可以发现延迟 3 阶以后，自相关系数都在两倍标准差范围左右，没有明显超出两倍标准差。可以看出，这是一个典型的短期相关的样本自相关图。由时序图和样本自相关图的性质，可以认为该序列平稳。

（3）纯随机性检验。利用 LB 统计量进行纯随机性检验，计算在 6 延迟阶数和 12 延迟阶数下的 LB 统计量的 p 值。在 6 和 12 延迟阶数下的 LB 统计量的 p 值都为 0（<0.001），所以有大于99.99%的置信水平认为芝加哥海德公园每 28 天的抢包案件数属于非白噪声序列。

总结（1）、（2）、（3）的结果可以发现，这个序列是一列平稳的、蕴含相关信息的序列，值得去分析和研究。

5.2 实验总结

（1）在实际应用中人们发现 Q 统计量在大样本场合检验效果很好，但在小样本场合就不太精确。为了弥补这一缺陷，Box 和 Ljung 又推导出 LB 统计量：

$$Q_{LB} = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right)$$

本案例中，由于样本数 n=70，较小，所以采用 LB 统计量。

（2）在纯随机性检验中，只检验了前 6 期和前 12 期延迟的 LB 统计量就直接判断该序列不具有纯随机性，而没有进行全部 69 期延迟检验。这是因为平稳序列通常具有短期相关性，如果序列值之间存在显著的相关关系，通常只存在于延迟时期比较短的序列值之间。如果短期之间不存在相关，长期延迟之间就更不会存在显著相关。如果序列具有短期相关性，那么就可以判断序列不具有纯随机性，值得进行下一步建模。如果考虑的延迟期数太长，那么可能会淹没该序列的短期相关性，因为平稳序列只要延迟时期足够长，自相关系数都会收敛于零。