

# R包—timeSeries

张家华  
1600022739

# TimeSeries

---

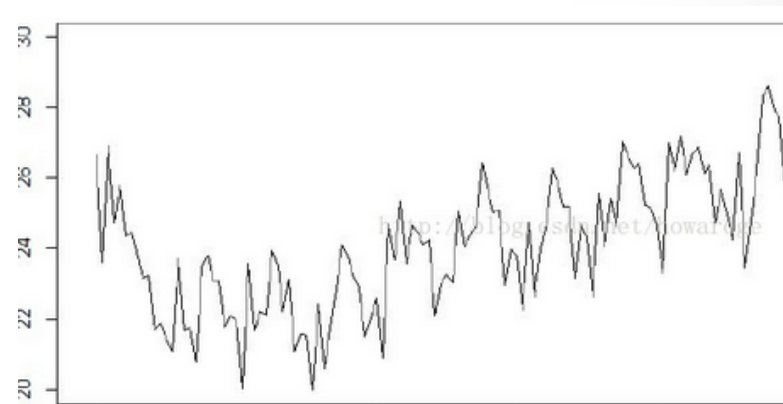
- R时间序列 ( time series ) 是一系列有序的数据。通常是等时间间隔的采样数据。如果不是等间隔，则一般会标注每个数据点的时间刻度。
- time series data mining 主要包括decompose ( 分析数据的各个成分，例如趋势，周期性 )，prediction ( 预测未来的值 ) 等。
- 本次主要讲针对**预测** ( forecast ) 问题。 即已知历史的数据，如何准确预测未来的数据。

# TimeSeries

Plot函数绘图：

一个季节性时间序列包含一个趋势部分，一个季节性部分和一个不规则部分。

可以看到这个时间序列在一定周期时间内存在的季节性变动：在何时达到峰值，何时达到谷底都会有清晰的显示。同样，这样的时间序列也可能是一个相加模型，随着时间推移，季节性波动时大致稳定的而不是依赖于时间序列水平，且对着时间的变化。



# TimeSeries

- 函数：使用了TimeSeries中**最复杂**的指数平滑法**Holt-Winters**
- Holt-Winters的思想是把数据分解成三个成分：平均水平（level），趋势（trend），周期性（seasonality）。
- Holt-Winters算法中提供了alpha、beta和gamma来分别对应当前点的水平、趋势部分和季节部分，参数的去执法范围都是0-1之间，并且参数接近0时，近期的观测值的影响权重就越小。

# TimeSeries

函数ARIMA : AutoRegressive Integrated Moving Average

ARIMA是两个算法的结合 : AR和MA。其公式如下 :

$$\bullet X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t - \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

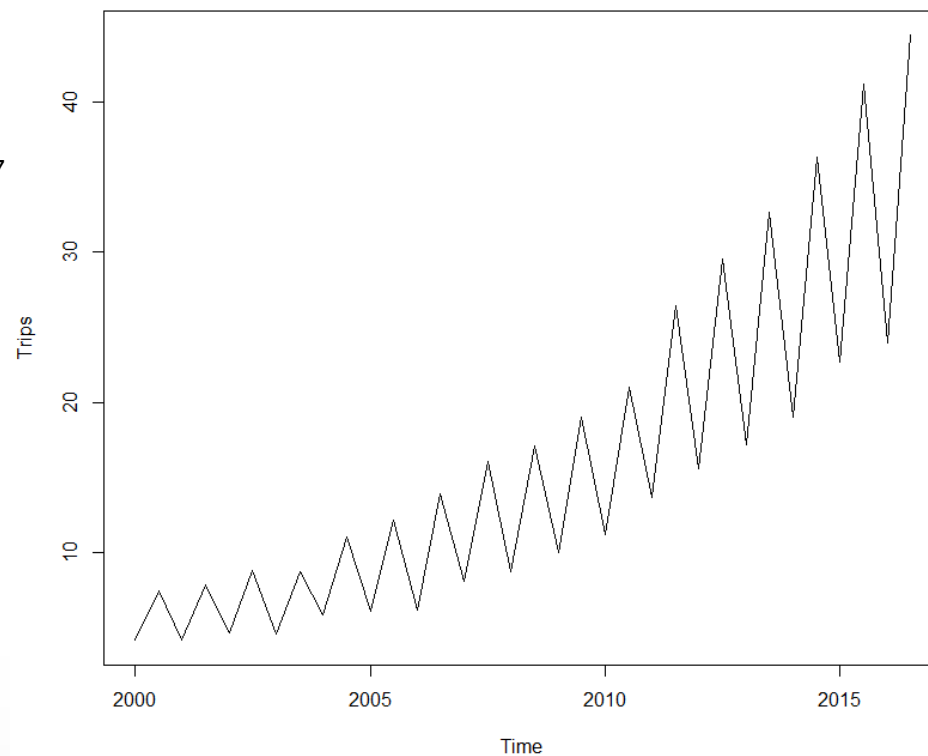
用到了差分, 即Integrated。例如一阶差分是把原数列每一项减去前一项的值。二阶差分是一阶差分基础上再来一次差分。

# 案例

## 导入数据 “2000-2015全国旅游景区人次表”

```
4.16 7.44 4.22 7.84 4.96 8.78 4.56 8.7 5.89 11.02 6.11 12.12  
6.19 13.94 6.65 16.1 8.73 17.12 9.99 19.02 11.21 21.03 14.67  
26.41 15.59 29.57 18.21 32.62 19.99 36.3 22.64 41.2 23.98  
44.4
```

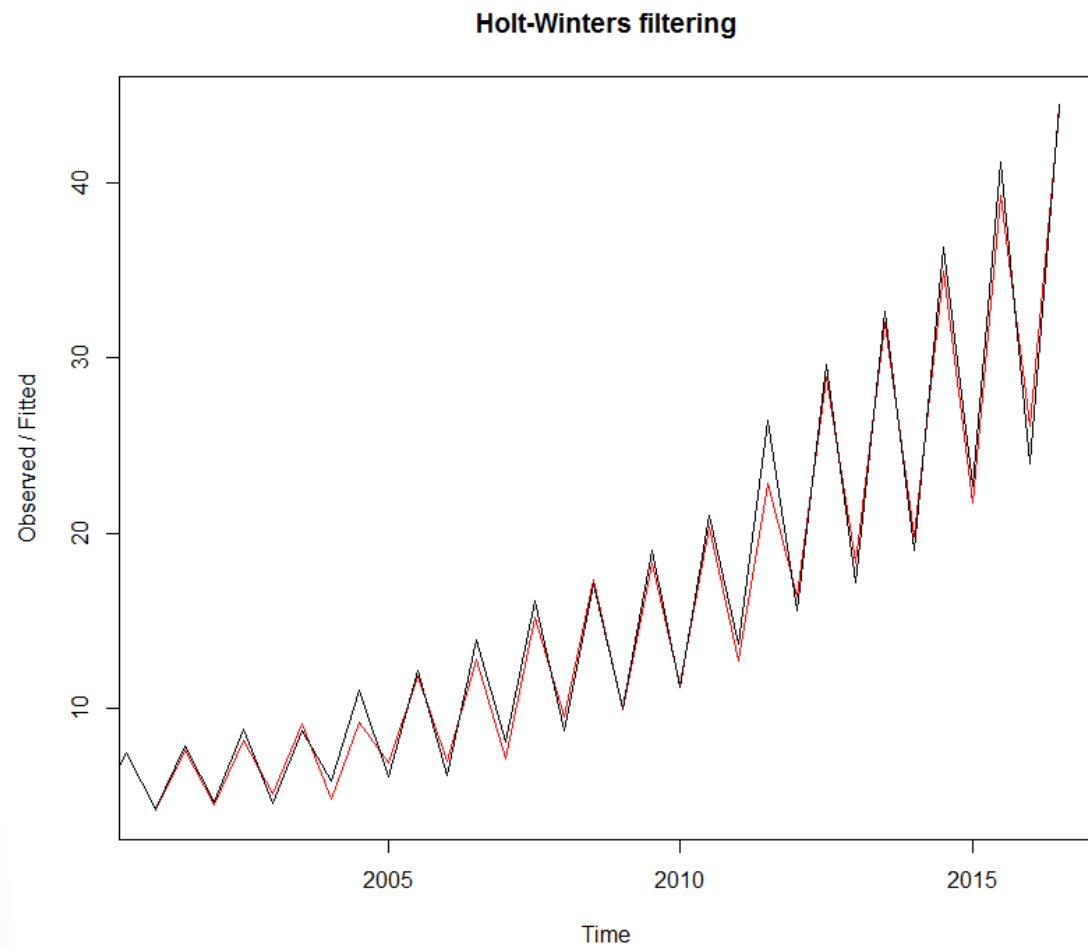
```
data<-read.csv("旅游景区.csv")  
data1 <- ts(data, frequency=2, start=c(2000,1))  
plot.ts(data)
```



## 函数运用

我们采用R中提供的**HoltWinters**算法进行预测，实现和结果如下：

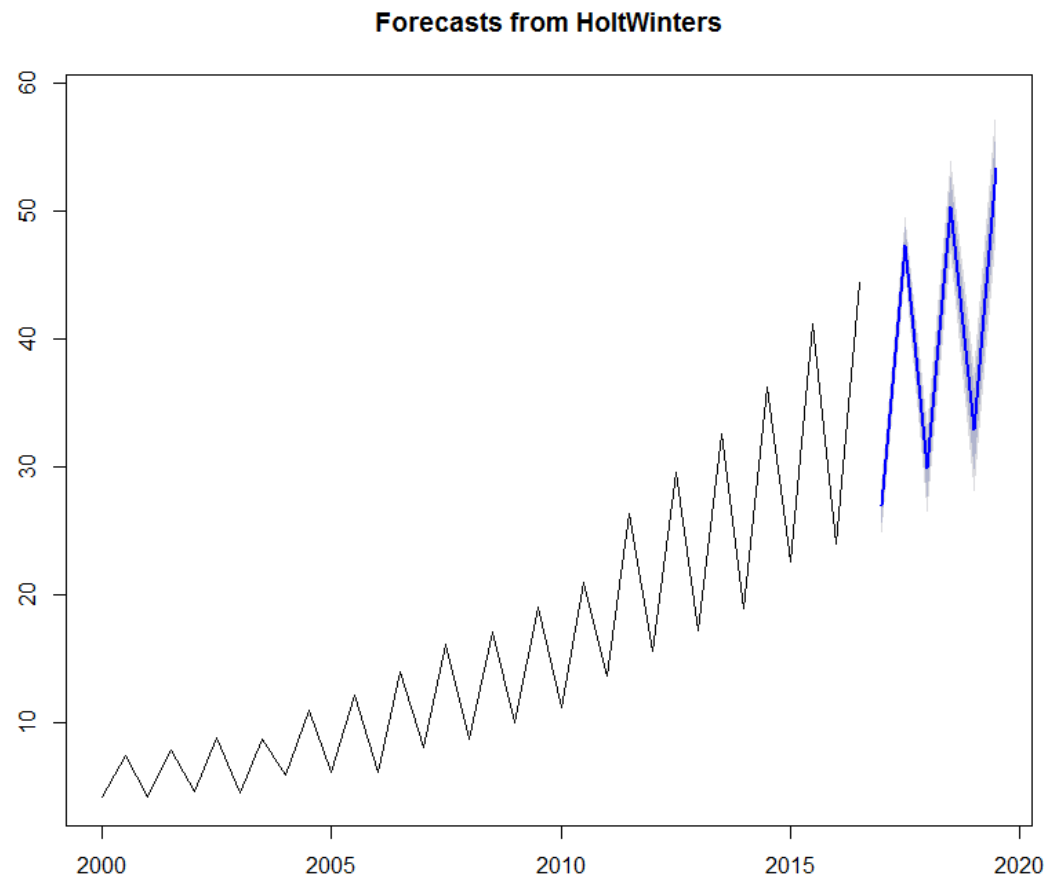
```
data1forecast <- HoltWinters(data1)  
plot(data1forecast)
```



## 案例

可见Holt-Winters很好地拟合了季节峰值，  
为了预测未来期数的值，我们同样采用forecast函数包，  
以预测未来6年人数为例：

```
library("forecast")  
data2<-forecast.HoltWinters(data1forecast,h=6)  
plot.forecast(data2)
```



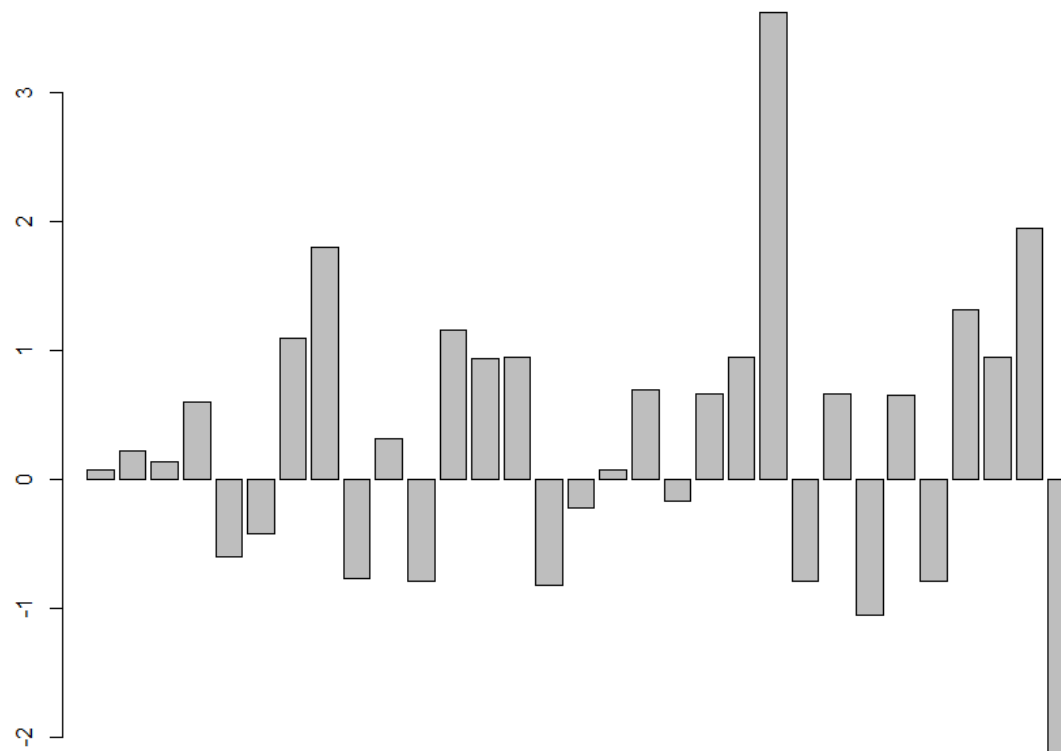


# 案例

接下来采用Ljung-Box 和画出直方图,  
检查预测误差的随机性

```
barplot(rev(data2$residuals),horiz=F,xlim=c(38,5),axes=T)
```

Data2\$residuals(残差图)



## 案例

使用“Ljung-Box”进行检验  
预测误差的随机性

```
Box.test(data2$residuals,lag=10,type = "Ljung-Box")
```

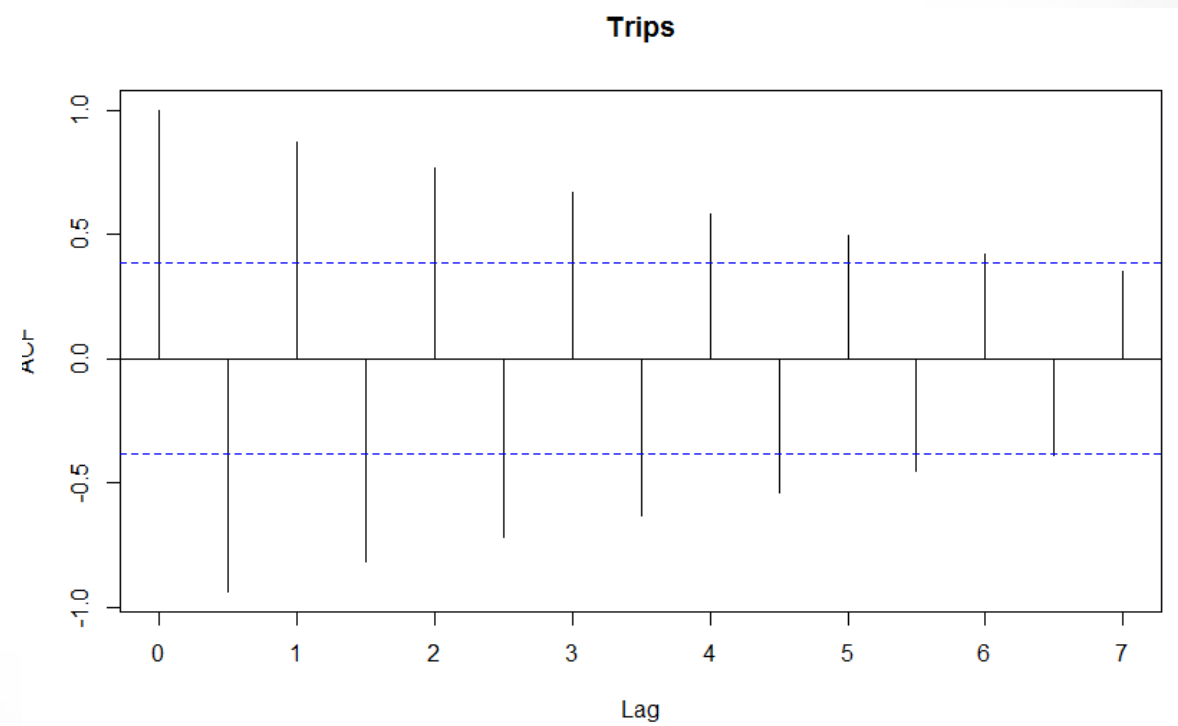
前提假设:误差结果是白噪音。  
结果:卡方较大,自由度为10,  
P值<0.05,有95%的把握拒绝  
原假设。

```
> Box.test(data2$residuals,lag=10,type = "Ljung-Box")  
  
Box-Ljung test  
  
data: data2$residuals  
X-squared = 25.848, df = 10, p-value = 0.003951
```

# 案例

ARIMA,  
对结果进行监测：

```
acf(data2$residuals)  
acf(diff(data1, lag=1))  
acf(diff(diff(data1, lag=7)))
```



# 问题

---

总结：在使用此函数时，样本数量越大，波动越大，越能体现出其价值。

Q1：使用**Holt-Winters**把上证指数10年历史数据分解成：平均水平、周期、趋势，并进行今后五年的季度指数预测，并进行检验。

Q2：对上述问题进行季度趋势预测，更改参数使得预测结果更加具有随机性。并使结果用diff图表现出来。

**R**

---

**THANKS**