Time Series (TS) Analysis
White Noise, Stationarity, and Variance Stabilization

Nagiza F. Samatova, samatova@csc.ncsu.edu
Professor, Department of Computer Science
North Carolina State University

Senior Scientist, Computer Science & Mathematics Division
Oak Ridge National Laboratory

Outline

- White Noise
  - ACF Analysis: Critical Value Lines
  - Portmanteau tests (Box.test() in R):
    - Box-Pierce test
    - Ljung-Box test
- Stabilizing the Variance
  - Box-Cox Transformation
- Stationarity
  - Stabilize the Mean with Differencing
  - Stabilize the Variance with Box-Cox Transformations

TS Parts: Systematic vs Non-systematic

TS Part	Definition	Detection	How to deal w/
Level	Average value of ts		
Trend	Long-term increase decrease in the data	lag.plot	De-trend via lag-1 differencing
Seasonality	Variations occurring during known periods of the year (monthly, quarterly, holidays)	lag.plot, Acf plots	De-seasonalize via lag-k differencing
Cycles	Other oscillating patterns about the trend (e.g., business or economic conditions)		
Auto-correlation	Correlation between neighboring points in ts	Acf, lag.plot	
Noise	Residuals after level, trend, seasonality, and cycles are removed	Normality tests	

Additive and Multiplicative TS Components

- A time series with additive components can be modeled as:

$$y_t = Level + Trend + Seasonality/Cycles + Noise$$

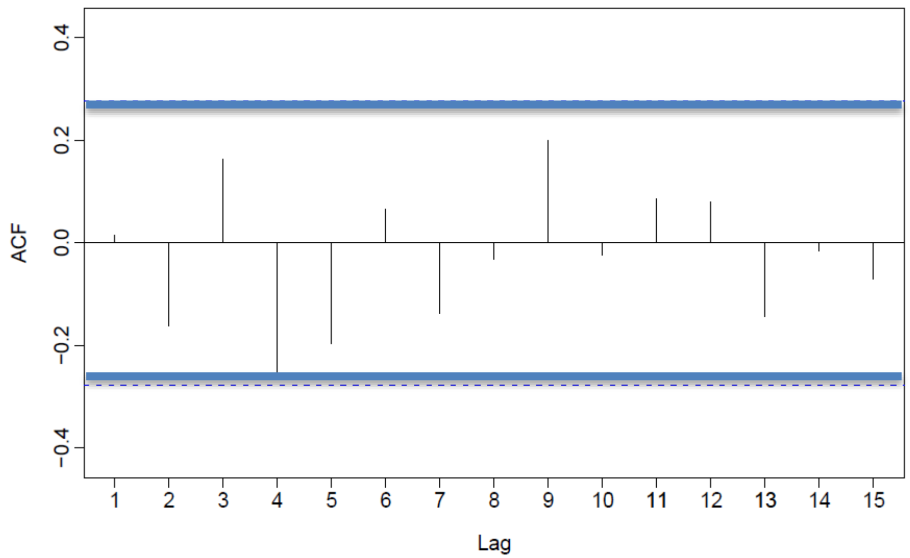
- A time series with multiplicative components is modeled as:

$$y_t = Level \times Trend \times Seasonality/Cycles \times Noise$$

- Forecasting methods attempt to isolate the systematic part and quantify the noise level.
  - The systematic part is used for generating point forecasts
  - The level of noise helps assess the uncertainty associated with the point forecasts

Residuals Must be White Noise

- White noise (WN) data is uncorrelated across time with zero mean and constant variance (independence is also required)
- WN: Uninteresting, with no predictable patterns
- Sample autocorrelations for white noise series are close to zero (i.e., uncorrelated data), or **between the critical value lines.**
- Better to perform **Portmanteau tests**



- Sampling distribution of e_k for white noise data is asymptotically $\mathrm{N}(0, \frac{1}{T})$.
- 95% of all e_k for white noise must lie within $\pm\frac{1.96}{\sqrt{T}}$
- If this is not the case, the residual series is probably not WN.
- Common to plot lines (**critical values**) at $\pm\frac{1.96}{\sqrt{T}}$ when plotting ACF.

Portmanteau Tests for White Noise

- ACF suffers from multiple testing problem
- **Core Idea**: consider a ***whole set*** of $\{e_k\}$ to test whether the set is significantly different from the zero set.
- **Portmanteau Tests:**
  - **Box-Pierce test :** $\boldsymbol{Q = T\sum_{k=1}^{h} e_k^2}$
  - **Ljung-Box test :** $\boldsymbol{Q^* = T(T+2)\sum_{k=1}^{h}(T-k)^{-1} e_k^2}$
    - $\boldsymbol{h}$ is the maximum lag considered ($h = 2m$, for seasonal data, or h=10, for non-seasonal data)
    - $\boldsymbol{Q}$ is **small** for WN
    - $\boldsymbol{Q}$ is **large** if some e_k values deviate from zero
    - $\boldsymbol{Q^*}$ is **small** for WN; better performance than $\boldsymbol{Q}$ for small samples
  - If data is WN, $\boldsymbol{Q^*}$ has $\boldsymbol{\chi^2}$ **distribution** with $(\boldsymbol{h-K})$ **degrees** of freedom where K = no. parameters in model.
  - When applied to raw data, set K = 0.
  - **If p-values are large, then the residuals are NOT distinguishable from a white noise series.**

Example: Portmanteau Tests

```
res <- residuals(naive(dj))

# lag=h and fitdf=K
> Box.test(res, lag=10, fitdf=0)
Box-Pierce test
X-squared = 14.0451, df = 10, p-value = 0.1709
> Box.test(res, lag=10, fitdf=0, type="Lj")
Box-Ljung test
X-squared = 14.4615, df = 10, p-value = 0.153
```

```
beer <- window(ausbeer,start=1992)
fc <- snaive(beer)
res <- residuals(fc)
Acf(res)
Box.test(res, lag=8, fitdf=0, type="Lj")
```

Box-Cox Transformation to Stabilize Variance

- If the data show different variation at different levels of the series, then a **Box-Cox transformation** can be useful to stabilize the variance:

$$y_t{}' = \begin{cases} \log_e(y_t),\ \lambda = 0 \\ \dfrac{y_t^\lambda - 1}{\lambda},\lambda \neq 0 \end{cases}$$

- **Automated Box-Cox transformations**:
    - Attempts to balance the seasonal fluctuations and random variation across the series.
    - Always check the results.
    - A low value of λ can give extremely large prediction intervals.

- A Box-Cox transformation followed by an additive ETS model is often better than an ETS model without transformation.
- A ***good value of λ is one that makes the size of the seasonal variation about the same across the whole series***, as this simplifies forecast model building:
    - $lambda = \ BoxCox.lambda(elec)$
    - $plot(BoxCox(elec, lambda))$

Explainable Variations

- Calendar variation
- Increasing population
- Inflation
- Strikes
- Changes in government
- Changes in law

Understand all possible sources of variation before modelling the time series.

Adjustments to Variations

- **Variations due to Month Length:**
  - **monthdays()** gives the number of days in each month or quarter
  - apply adjustments to month length:

$$
\begin{aligned}
y_t^* &= y_t \times \frac{\text{no. of days in an average month}}{\text{no. of days in month } t} \\
&= y_t \times \frac{365.25/12}{\text{no. of days in month } t}
\end{aligned}
$$

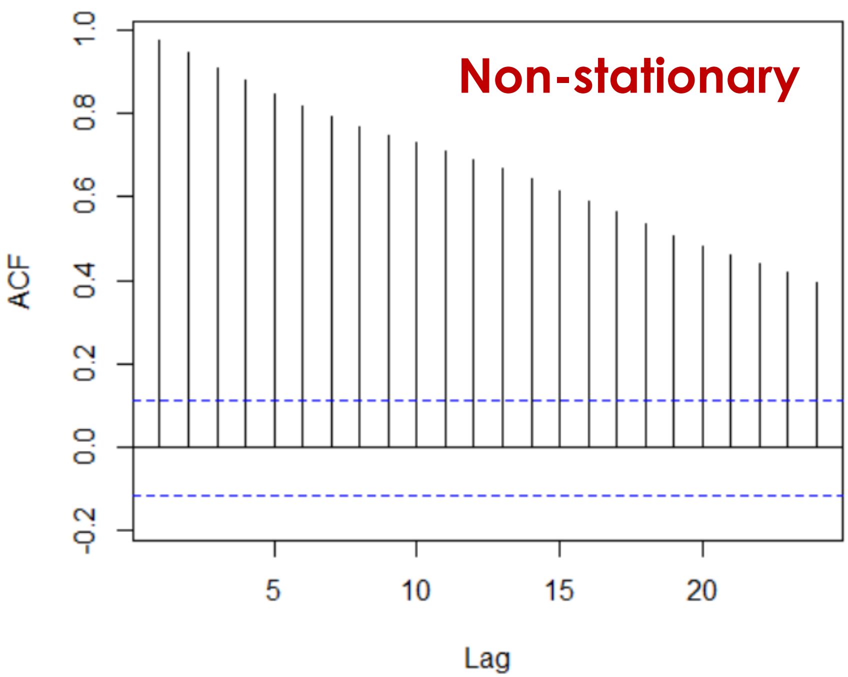
Stationarity

- **Stationary time series**: for all s, the distribution of $(y_t, y_{t+1}, \dots, y_{t+s})$ does not depend on t.
- Stationary series has the following properties:
  - Roughly horizontal
  - Constant variance
  - No patterns predictable in the long-term
- To identify **non-stationary** time series:
  - Plot the time series
  - The ACF of stationary ts drops to zero quickly
  - The ACF of non-stationary data decreases slowly

- **Dealing with non-stationarity**
  - Stabilize the **mean** with **differencing**
  - Stabilize the **variance** with **Box-Cox transformations**

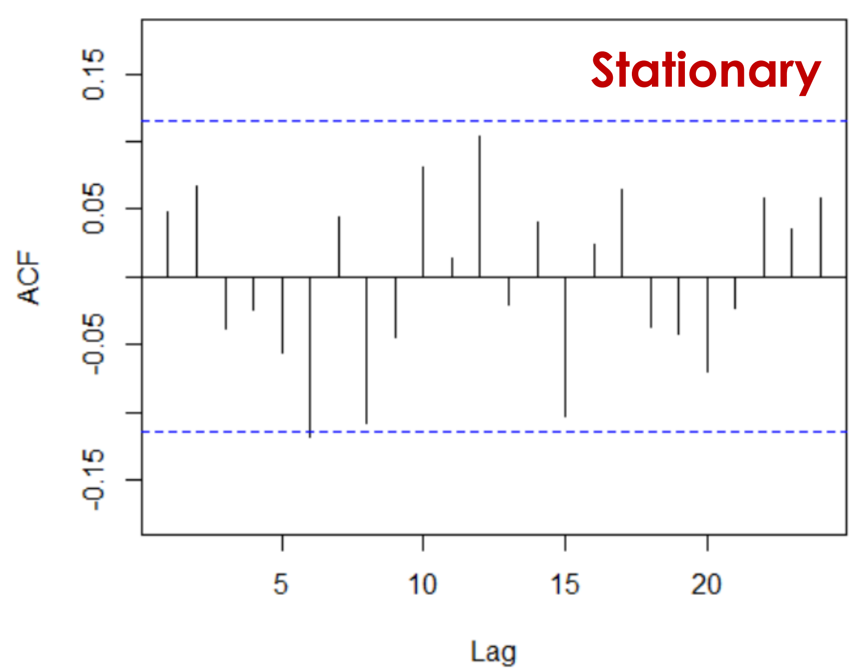
ACF of Stationary vs. Non-Stationary

`Acf(dj)`



- Slowly decreasing
- r_1 is large and positive

`Acf(diff(dj))`



- ACF drops to zero quickly

Unit Root Test: Order of Differencing

- To determine the required **order of differencing**:
  - **Augmented Dickey Fuller test**: null hypothesis is that the data are **non-stationary and non-seasonal**.
  - **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test**: null hypothesis is that the data are **stationary and non-seasonal**.

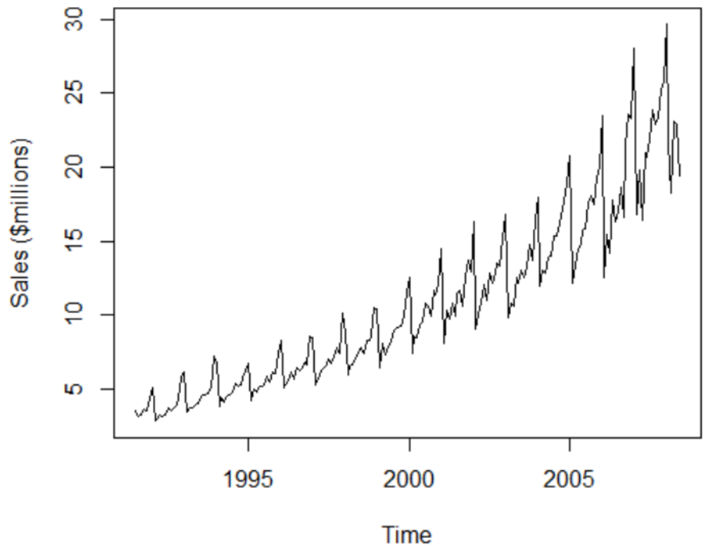
```
ndiffs(x)
nsdiffs(x)
```

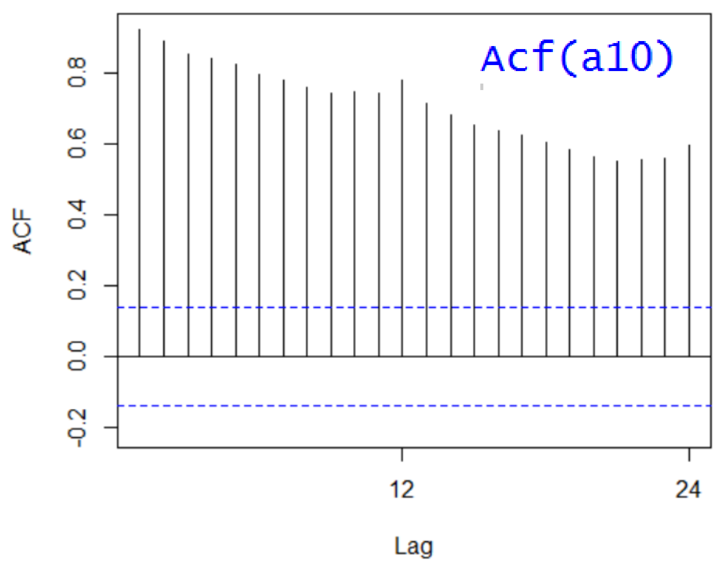
Automated differencing

```
ns <- nsdiffs(x)
if(ns > 0)
  xstar <- diff(x,lag=frequency(x),
                differences=ns)
else
  xstar <- x
nd <- ndiffs(xstar)
if(nd > 0)
  xstar <- diff(xstar,differences=nd)
```

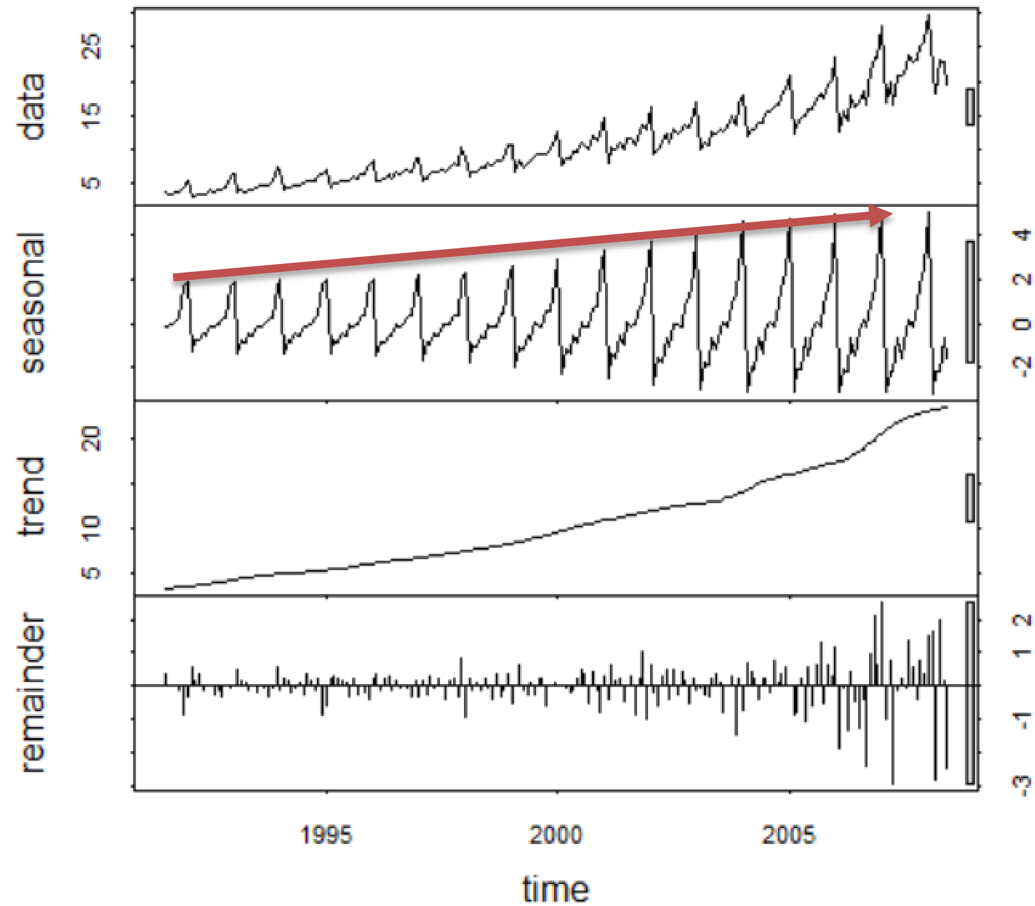
Example: Annual Antidiabetic Drug Sales

```
plot (a10, ylab="Sales ($millions)")
```





Non-stationary

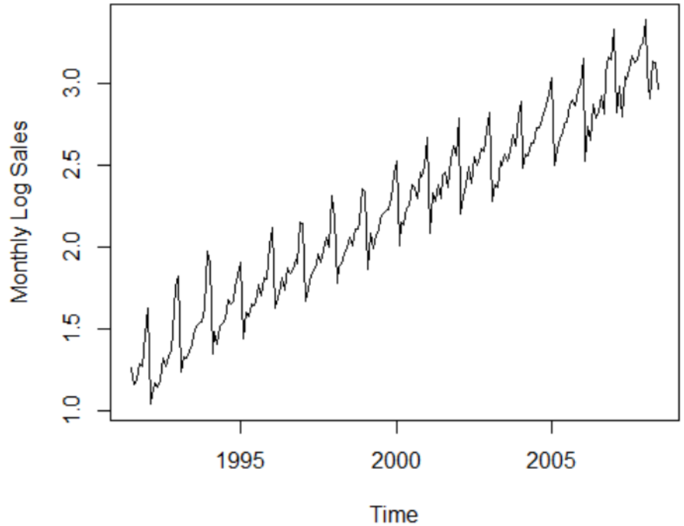



```
plot(stl(a10, s.window=12))
```

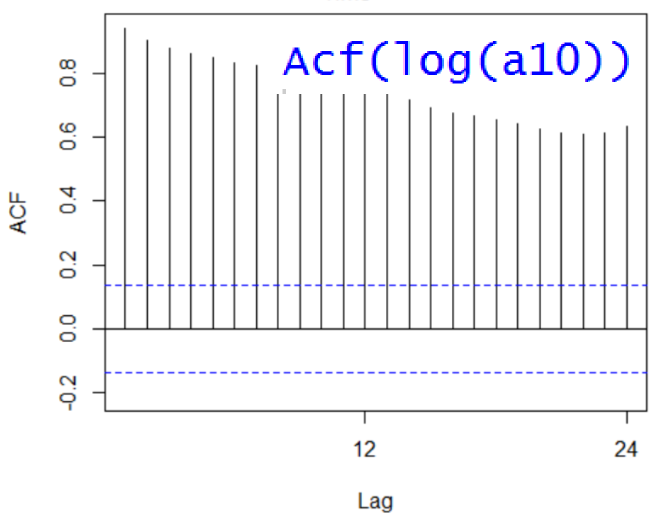
Ex: Non-stationary → Stationary: Step 1

Step 1: Stabilize the variance: Log Transformation

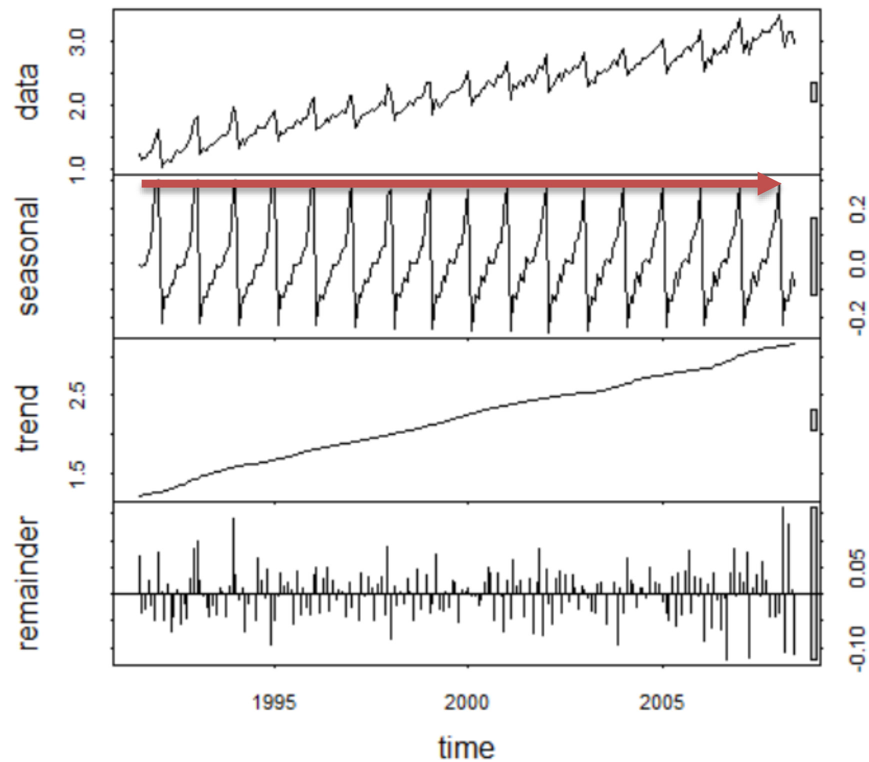
```
plot (log(a10), ylab="Monthly Log Sales")
```




```
Acf(log(a10))
```



Still Non-stationary



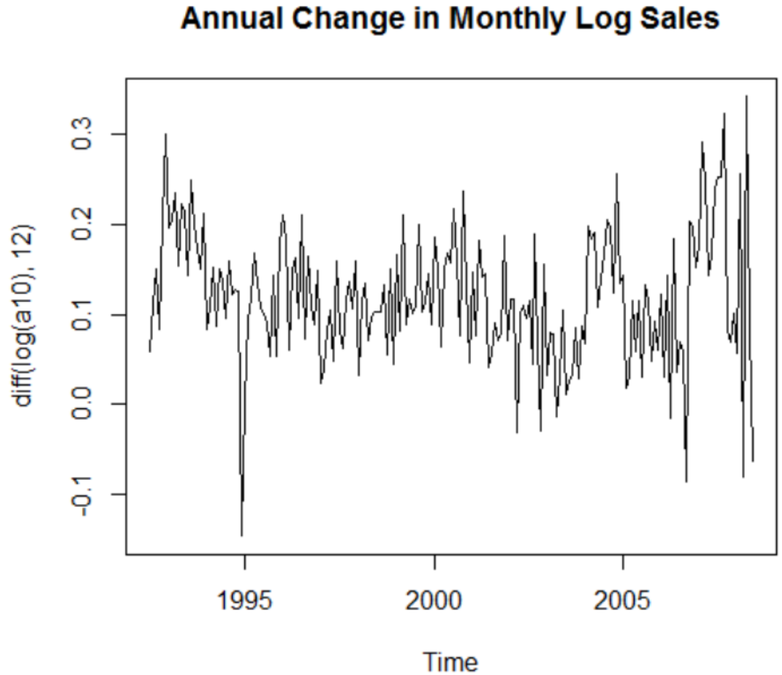

```
plot(stl(log(a10), s.window=12))
```

Ex: Non-stationary → Stationary: Step 2

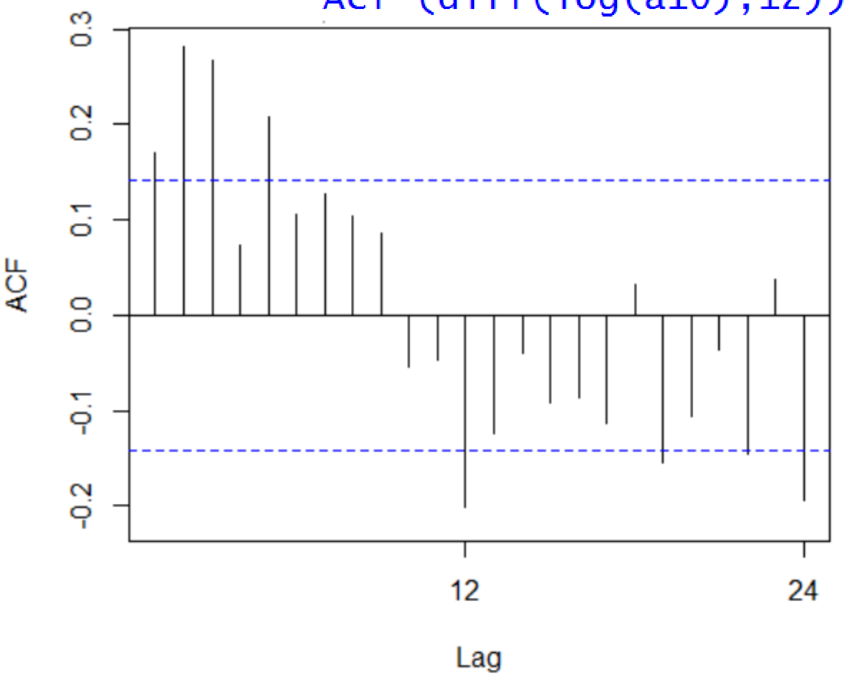
Step 2: **Remove trend and seasonality: Seasonal Difference**

Stationary

```
plot (diff(log(a10),12), main="Annual Change in Monthly Log sales")
```




```
Acf (diff(log(a10),12))
```



Acknowledgements

- Books
  - Free and online (otexts.com/fpp): Forecasting Principles & Practice by R. Hyndman, G. Athanasopoulos ← Excellent Book!!!
  - Practical Time Series Forecasting with R: A Hand-on Guide by Shmueli & Lichtendahl
- Packages
  - R: fpp (install.packages ("fpp", dependencies=TRUE))