時系列解析（1）

東京大学

北川源四郎

講義の概要

学部横断型教育プログラム「数理・データサイエンス教育プログラム」

開講科目名：	数理手法VII（時系列解析）
時限：	水曜５限
単位数：	２
学年：	B3/B4
担当教員：	北川源四郎
教室：	工６号館工６１号講義室

第４の科学： データサイエンス

	Human inspiration dependent	Cyber-enabled
演　繹 （モデル駆動型）	**理論科学**	**（第3の科学） 計算科学**
帰　納 （データ駆動型）	**実験科学**	**（第4の科学） データサイエンス**

教員プロフィール： 北川源四郎

現職：　東京大学 数理・情報教育研究センター 特任教授
明治大学 先端数理科学インスティテュート 客員教授
統計数理研究所 名誉教授，総合研究大学院大学 名誉教授

略歴：　東京大学理学系研究科数学専攻博士課程中退（1974），理学博士
統計数理研究所研究員，助教授，教授，所長（2002－11）
情報・システム研究機構長（2011－17）
この間、タルサ大学客員助教授（1980-81)、合衆国商務省センサス局研究員（1981-82）、総合研究大学院大学助教授・教授（1988-2011)、東京大学経済学研究科助教授（1988-91)、日本銀行金融研究所客員研究員(1996-98)、日本学術会議会員(2011-17)

研究分野：時系列解析（非定常モデリング，粒子フィルタ）
統計的モデリング（情報量規準GIC, EIC，ベイズモデリング）
船舶のオートパイロット、経済時系列の季節調整、地震波の自動処理、信号抽出、強風予測（JR列車安全運行システム）などの応用研究

所属学会： ASA (American Statistical Association: Fellow), ISI (International Statistical Institute), IASC (International Association for Statistical Computing), 日本統計学会, 日本数学会, 応用統計学会, 人工知能学会, 計算機統計学会，日本船舶海洋工学会, 日本地震学会, 計測自動制御学会, 応用経済時系列研究会, 日本金融・証券計量・工学学会

過去の学会活動：　日本統計学会会長，統計関連学会連合理事長, ISI Councilor, IASC Councilor

主な著書：情報量統計学(共立出版1983), 時系列解析プログラミング(岩波書店1993), 時系列解析の実際 I, II（朝倉書店1994)，時系列解析の方法（朝倉書店1998), 情報量規準（朝倉出版2004)，時系列解析入門（岩波書店、2005)，時系列解析（岩波書店2017 分担執筆）
Akaike Information Criterion Statistics（D.Reidel 1986)
Smoothness Prior Analysis of Time Series (Springer, 1996)
Practice of Time Series Analysis (Springer 1998)
Information Criteria and Statistical Modeling (Springer 2008)
Introduction to Time Series Modeling (Chapman & Hall 2010)

講義の目的

　時間とともに変動する現象を記録したデータが**時系列**である。時系列に基づき、複雑な現象を理解し、予測、制御や意思決定を行うための方法が**時系列解析**である。

　この講義では、時系列のモデリングのための前処理や特徴の可視化、統計的モデリングの方法、線形・定常時系列モデル、状態空間モデルおよび非線形・非ガウス型モデルについて、実際の問題への応用含めつつモデリングの方法を中心に解説し、**現実の問題に対応して適切なモデリングができるようになることを目標とする。**

講義予定

1	4/11	時系列の前処理と可視化	イントロダクション，時系列の前処理
2	4/18		共分散関数,スペクトルとピリオドグラム
3	4/25	モデリング	統計的モデリング・情報量規準AIC
4	5/2		最尤法、最小二乗法、ベイズモデル
5	5/9	定常時系列モデル	ARMAモデルによる時系列の解析
6	5/16		ARモデルの推定・応用
7	5/23		局所定常ARモデル
	5/30		
8	6/6	状態空間モデル	状態空間モデルによる時系列の解析
9	6/13		ARMAモデルの推定,トレンドの推定
10	6/20		季節調整モデル
11	6/27		ボラティリティ、時変係数ARモデル
12	7/4	非線形・非ガウス型モデル	非線形・非ガウス型状態空間モデル
13	7/11		粒子フィルタ
	7/18		

教科書・参考書

教科書

北川源四郎:「時系列解析入門」岩波書店(2005)

参考書

岩波データサイエンス刊行委員会編,　「時系列解析：状態空間モデル・因果解析・ビジネス応用」岩波データサイエンス６,　岩波書店

関連文献

1. 赤池弘次,　中川東一郎:「ダイナミックシステムの統計的解析と制御」〔新訂版〕サイエンス社 (2000)
2. 北川源四郎:「時系列解析プログラミング」岩波書店 (1993)
3. 赤池弘次,　北川源四郎編:「時系列解析の実際」1, 2 朝倉書店 (1994, 1995)
4. 尾崎統,　北川源四郎編:「時系列解析の方法」朝倉書店 (1998)
5. G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*, Springer (1996)
6. G. Kitagawa, *Introduction to Time Series Modeling*, CRC-Chapman and Hall (2010)

Rパッケージについて

Rのインストール
https://cran.r-project.org
http://cran.ism.ac.jp （統計数理研究所のミラーサイト）

本講義関連時系列解析用のRパッケージ

- TSSS （Time Series analysis with State Space model）
  http://jasp.ism.ac.jp/ism/TSSS/
  参考文献２（時系列解析プログラミング）に掲載されたプログラムを基に統計数理研究所が開発したR 関数群。教科書（時系列解析入門）のほとんどの例はTSSSで再現できる。（並列化版もあり）
- Timsac（TIMSAC for R package）
  http://jasp.ism.ac.jp/ism/timsac/
  時系列解析統合プログラムパッケージTIMSACを基に統計数理研究所が開発したR関数群

その他の時系列解析Rパッケージ

- dlm（An R package for dynamic linear model）
- KFAS （Exponential family state space models in R）

講義資料

数理・情報教育研究センターHP（関連教材）
www.mi.u-tokyo.ac.jp/teaching_material.html

- PowerPoint資料
- Rコード
- 講義で使う時系列データ

時系列解析教材

学部横断型教育プログラム「数理・データサイエンス教育プログラム」
数理手法VII（時系列解析）
北川源四郎

講義資料

- 第１回（4月12日）　ppt資料　Rコード

講義で使用する時系列データ

- 一括ファイル　時系列データ.zip
- hakusan_new.csv
- sunspot_new.csv
- maxtemp.csv
- blsfood_new.csv
- whard_new.csv
- mye1f_new.csv
- nikkei225_new.csv
- haibara_new.csv
- Lynx_new.csv
- rainfall_new.csv

April 10, 2018

講義で使う時系列データ

数理・情報教育研究センターHP（関連教材）
www.mi.u-tokyo.ac.jp/teaching_material.html

● 時系列データ.zip	下記のcsvファイルを含むzipファイル
• hakusan_new.csv	船舶の方向角速度，横揺れ，縦揺れ，エンジン，回転数（秒）
• sunspot_new.csv	太陽黒点数（年，月）
• maxtemp.csv	東京の最高気温(日，月）
• blsfood_new.csv	アメリカの食品産業に従事する労働者人口（月）
• whard_new.csv	ある商品の卸売高(月，日）
• mye1f_new.csv	地震波 の東西成分（1/100秒）
• nikkei225_new.csv	日経２２５株価 (日）
• haibara_new.csv	地下水位と気圧（分、時間）
• Lynx-new.csv	Canadian Lynx捕獲数（年）
• rainfall_new.csv	降雨データ（東京，日）

統計データのいろいろ

1. 独立な観測データ

   調査データ，治験データ

2. 時系列データ

   株価データ，気象データ

3. 空間（平面）データ

   GIS(Geographic Information System

4. 時空間データ

   気象データ，人口流動データ

時系列とは

<u>時系列: Time Series</u>

時間とともに不規則に変動する現象の記録

- 世の中の興味あるデータの多くは時系列
- 時系列は時刻をパラメータとする確率変数の実現値

Y_n: 確率過程　　　y_n ：時系列

時系列グラフの利用(可視化)

グラフ → データの特徴 問題点 → 分析法

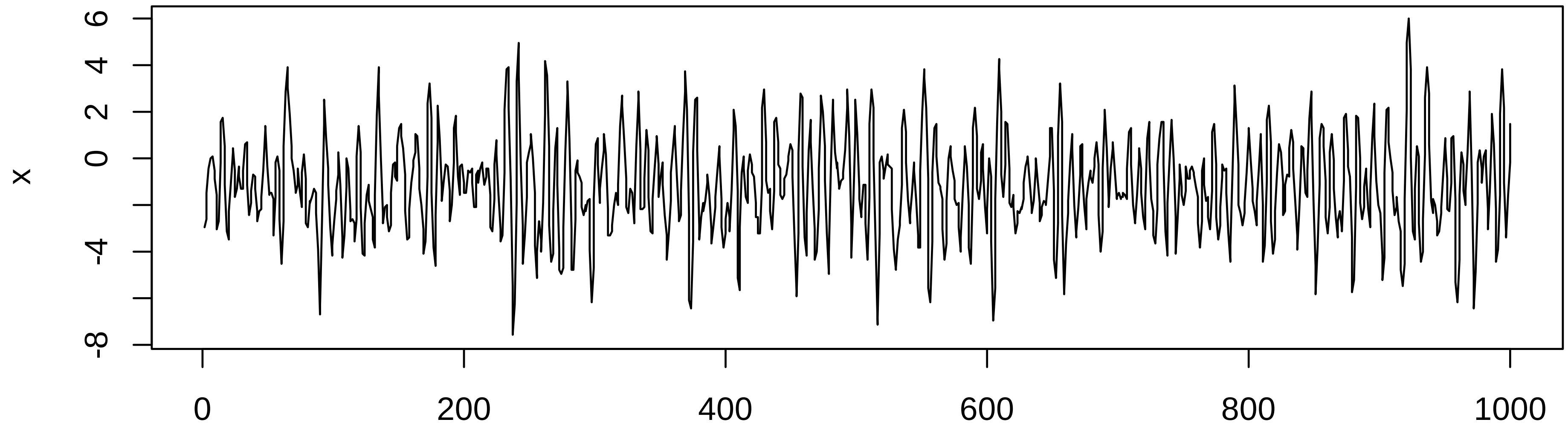
データプロット（直接表示）

時間相関を可視化（相関関数）

周期性を可視化（スペクトル、ピリオドグラム）

船舶の方向角データ

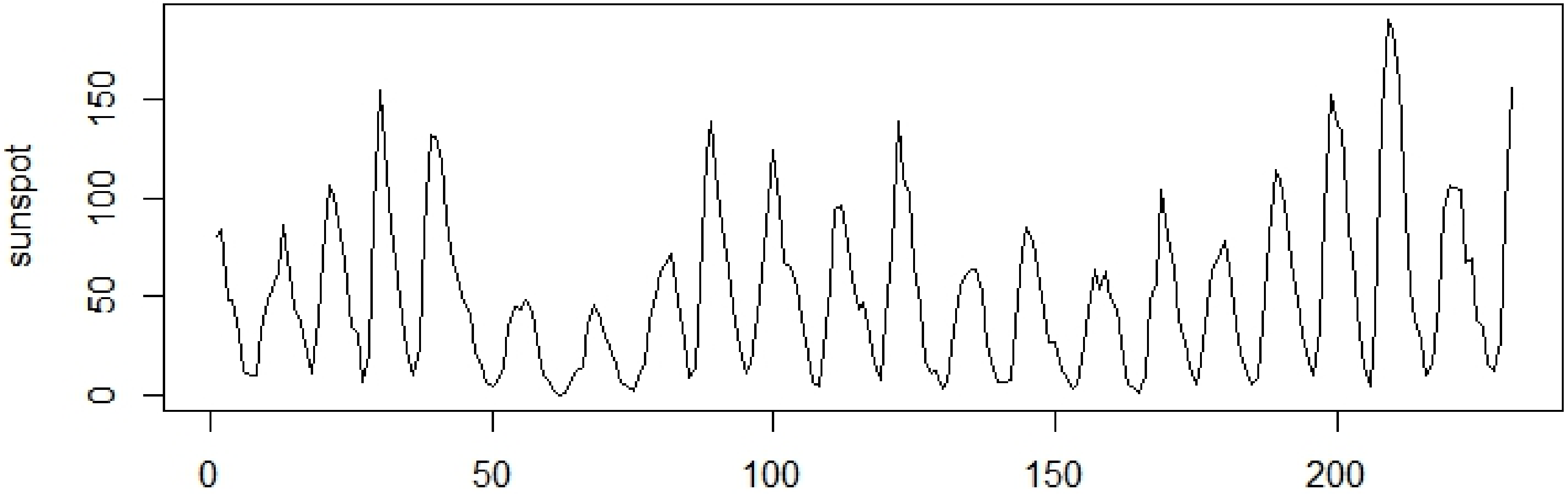
```
hakusan <- read.csv("hakusan_new.csv")
x <- as.ts(hakusan[,1])
plot(x)
```



- 定常性

太陽黒点数データ

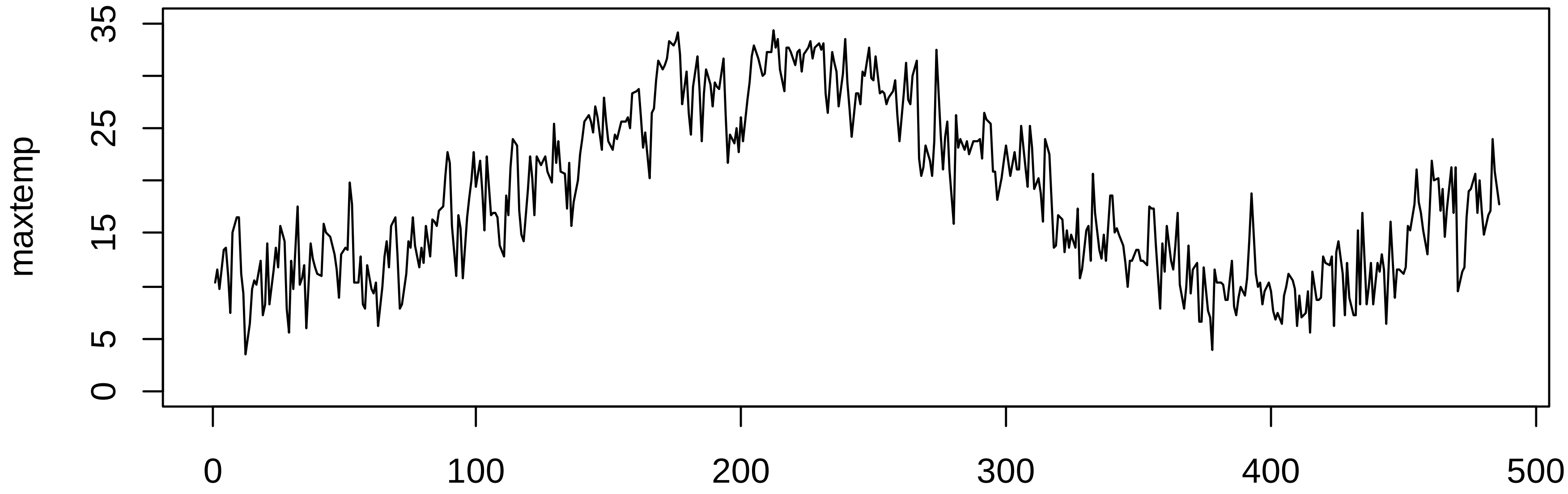
```
sunspot <- read.csv("sunspot_new.csv")
x <- as.ts(sunspot)
plot(x)
```



- 正値，上下非対称
- 擬似周期性
- 前後非対称性

日最高気温データ（東京）

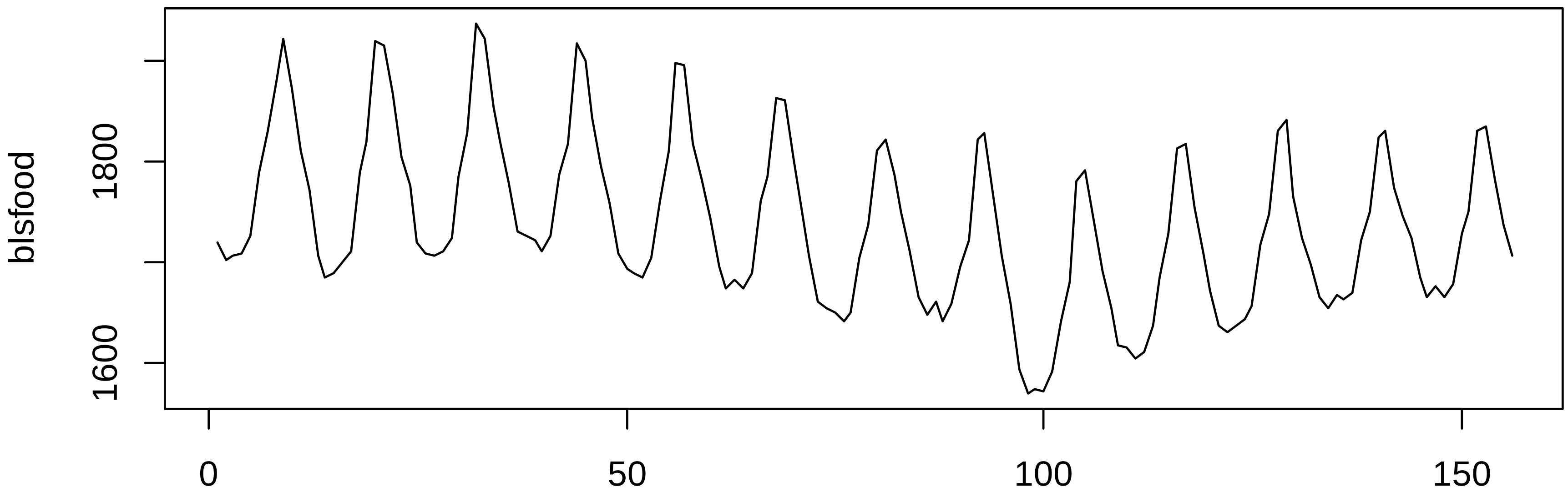
```
maxtemp <- read.csv("maxtemp.csv")
x <- as.ts(maxtemp)
plot(x,ylim=c(0,35))
```



- トレンド（長期の周期性）
- トレンド周りほぼ定常

米国食品産業従事者数データ

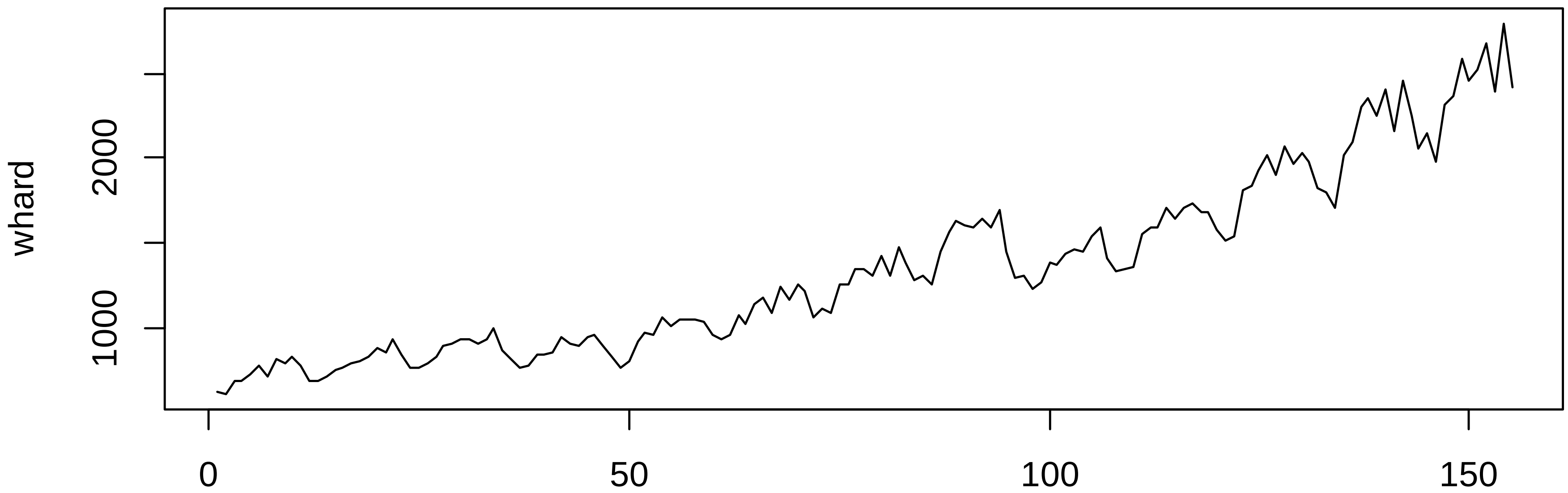
```
blsfood <- read.csv("blsfood_new.csv")
x <- as.ts(blsfood)
plot(x)
```



- 年周期性
- トレンド

卸売り高データ

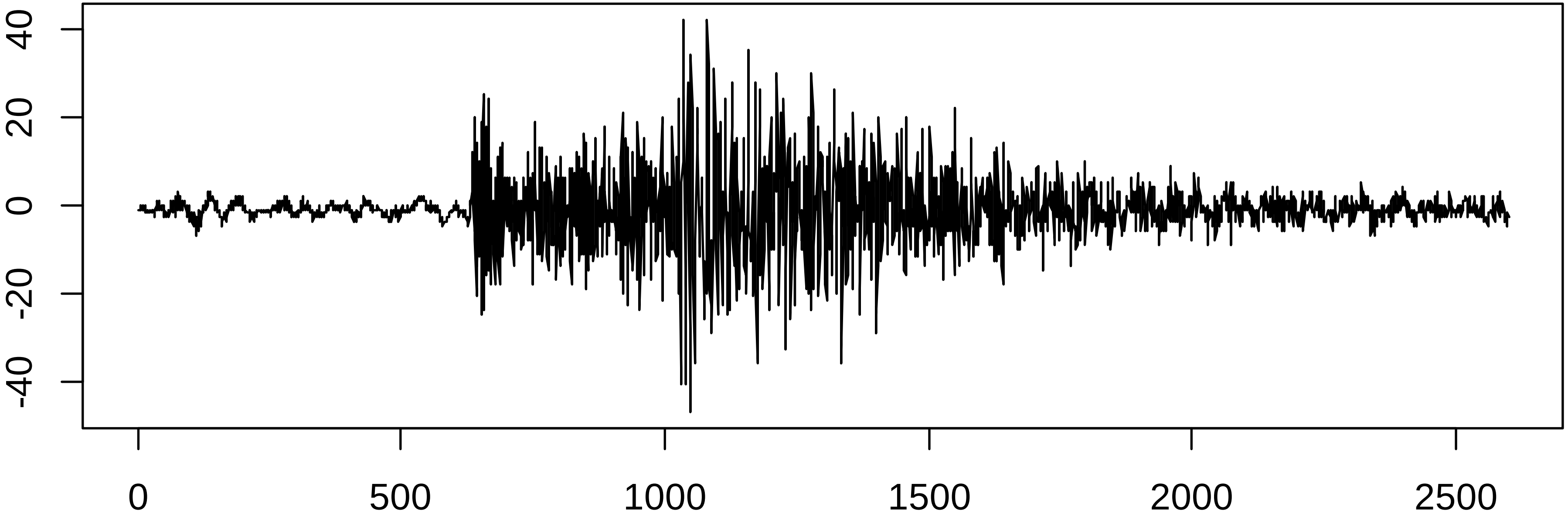
```
whard <- read.csv("whard_new.csv")
x <- as.ts(whard)
plot(x)
```



- 年周期性
- 正値性
- トレンドと変動分散の連動

MYE1F（東西方向地震波）データ

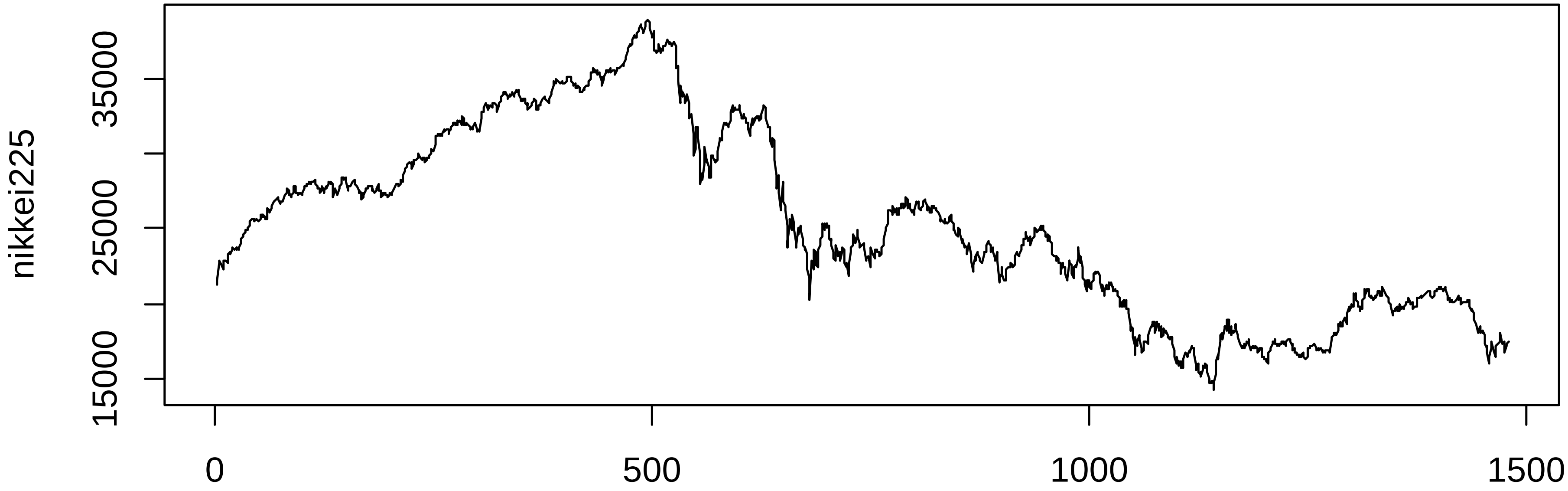
```
mye1f <- read.csv("mye1f_new.csv")
x <- as.ts(mye1f)
plot(x)
```



- トレンドなし
- 区分的定常性
- 分散非定常，共分散非定常

日経225平均株価データ

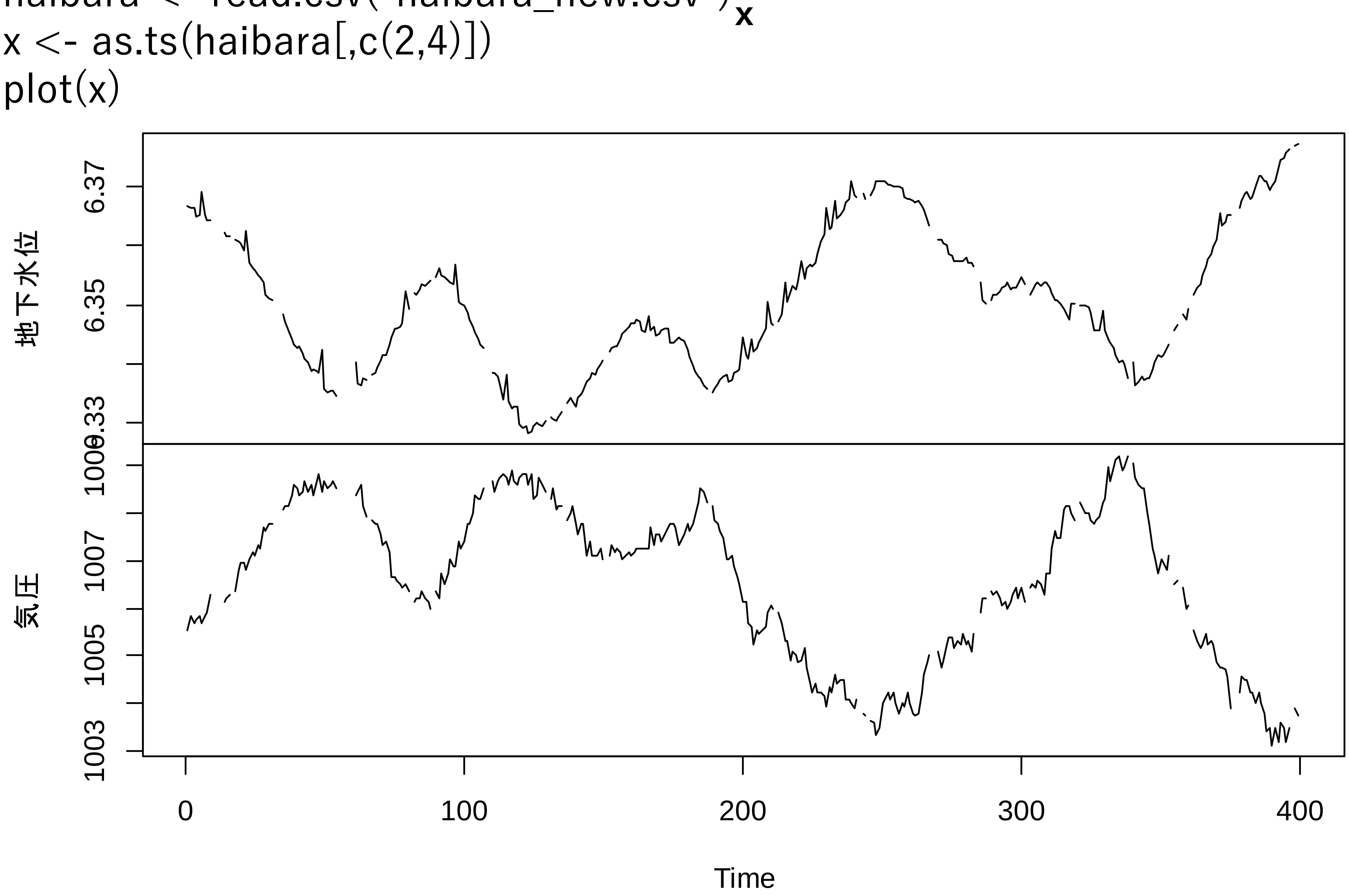
```
nikkei225 <- read.csv("Nikkei225_new.csv")
x <- as.ts(nikkei225)
plot(x)
```



- トレンド＋分散変動
- トレンド下降時の分散増加

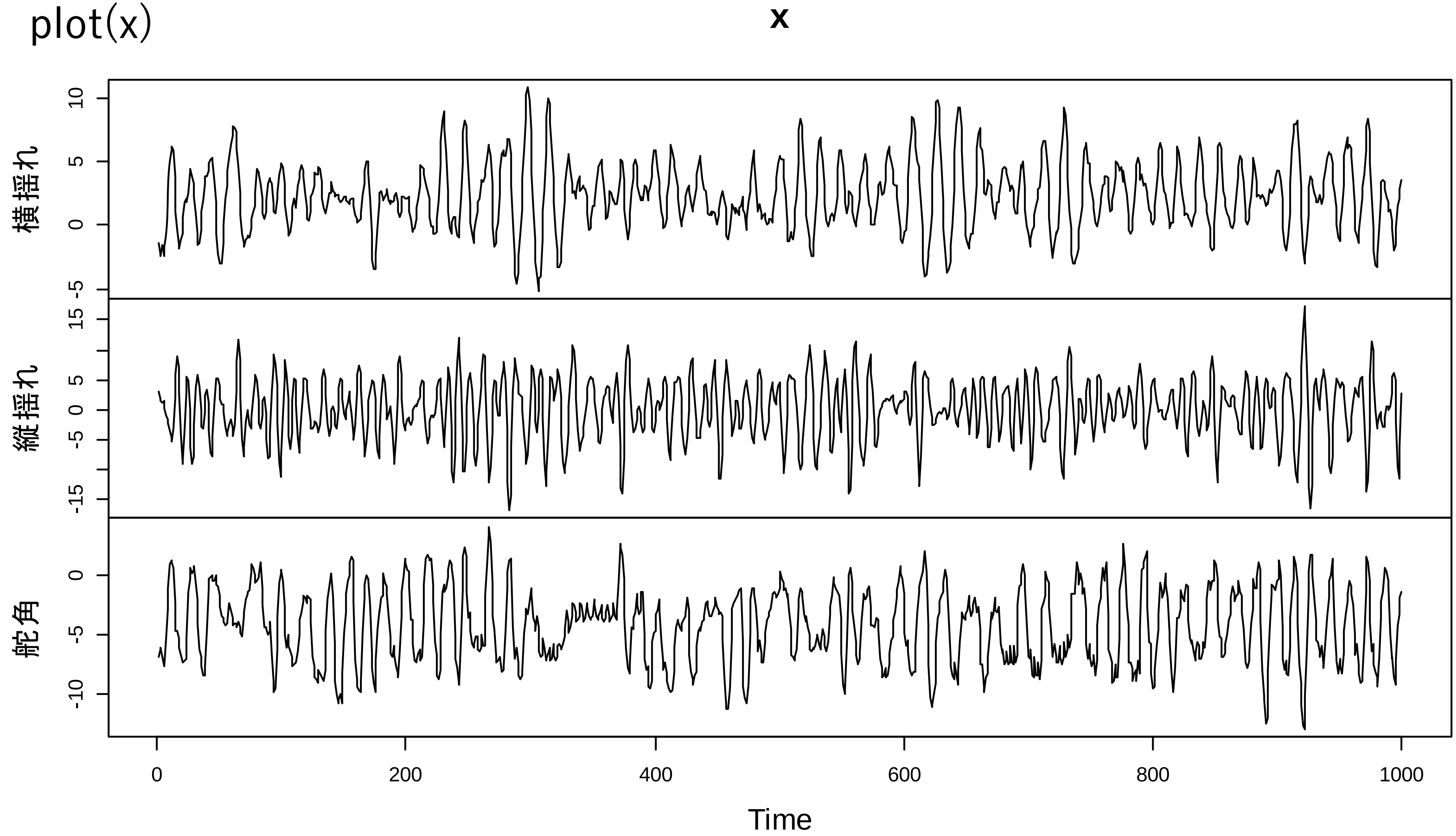
榛原地下水位・気圧データ

```
haibara <- read.csv("haibara_new.csv")
x <- as.ts(haibara[,c(2,4)])
plot(x)
```



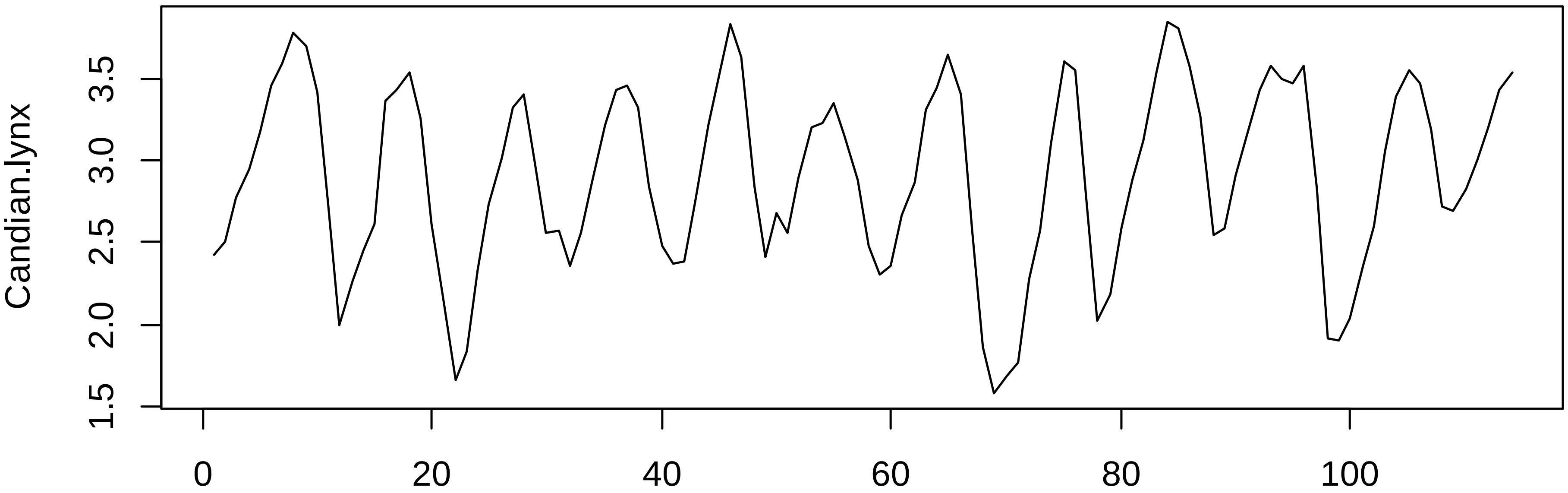
船舶データ（横揺れ、縦揺れ、舵角）

```
hakusan <- read.csv("hakusan_new.csv")
x <- as.ts(hakusan[,c(3,4,7)])
plot(x)
```



Canadian Lynx捕獲数（対数値）

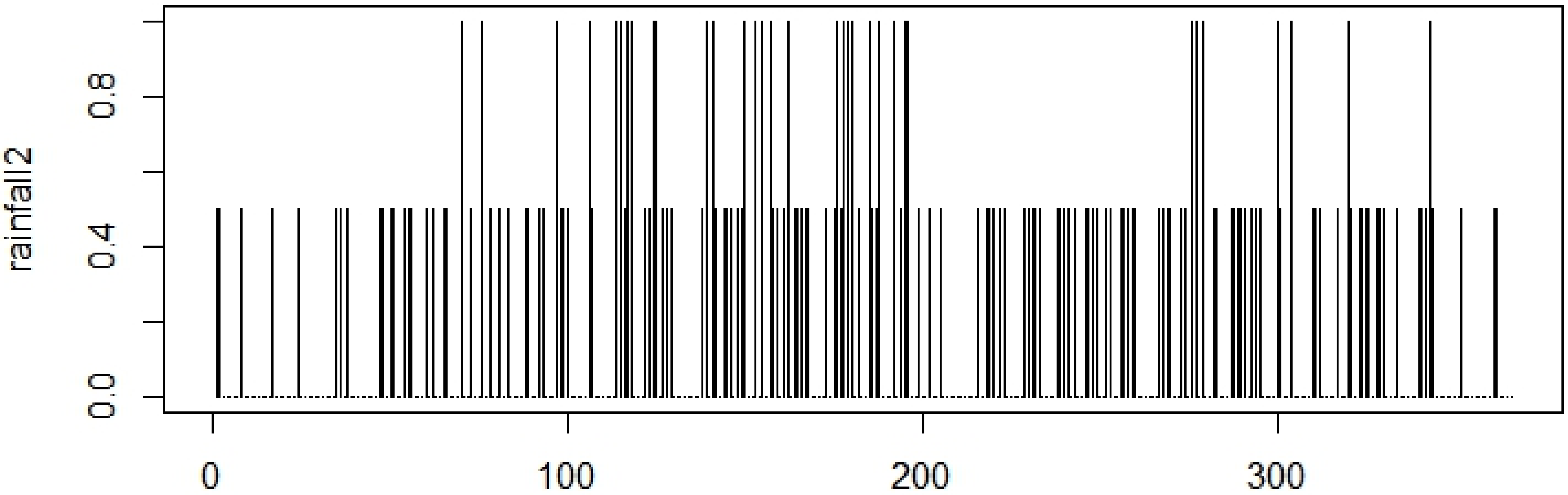
```
lynx <- read.csv("Lynx-new.csv")
x <- as.ts(lynx)
plot(x)
```



- 平均定常
- 擬似周期的
- 前後非対称

東京降雨データ（日単位）

```
rainfall <- as.ts(read.csv("rainfall_new.csv",header=TRUE))
rainfall2 <- rainfall[,4]/2
plot(rainfall2,type="h")
```



- 離散値
- 平均値非定常

時系列分類の観点

時系列は様々な観点から分類できる

- 連続時間時系列と離散時間時系列

  離散：等間隔時系列と 不等間隔時系列

- 一変量時系列と多変量時系列

- 定常時系列と非定常時系列

  定常： 性質が一定で時間的に変化しないもの

  非定常：性質が時間と共に変化するもの

- 線形時系列と非線形時系列

- ガウス型時系列と非ガウス型時系列

（本講義における）時系列解析

- 可視化（Visualization）

  時系列を図示したり，基本的な記述統計量を用いて時系列の特徴や解析結果を簡潔に表現する．

- モデル解析（Modeling）

  時系列モデルを推定し，時系列の特徴を捉えること

- 予測（Prediction）

  現在までに得られた情報から今後の変動を推測する

- 制御（Control）・意思決定

  操作可能な変数を適当に変化させ目的とする制御変数が望ましい変動をするように制御すること．

統計的モデリング

対象に関するあらゆる取得可能な情報（対象に関する理論，経験的知識，観測データ）およびモデリングの目的

事前情報とデータの持つ情報の統合 ➡ ベイズモデリング

時系列の前処理

時系列 ⇒ 前処理 ⇒ モデリング・分析

- 変数変換
- 階差（差分）
- 前期比・前年比
- 移動平均・移動メディアン

✓ モデリングを容易にするため
✓ 副作用もある

変数変換

目的：・線形，定常，正規性などを仮定したモデリングを容易にするため
・最適化を容易にするため

例

- 対数変換
- ロジット変換
- Box-Cox変換

変数変換

$$X \sim f(x) \quad \underset{x = h(y)^{-1}}{\overset{y = h(x)}{\rightleftarrows}} \quad Y \sim g(y)$$

$$g(y) = f(h^{-1}(y)) \left| \frac{dh^{-1}}{dy} \right|$$

$$f(x) = g(h(x)) \left| \frac{dh}{dx} \right|$$

変数変換

例　$y = h(x) = \log x$　（対数変換）

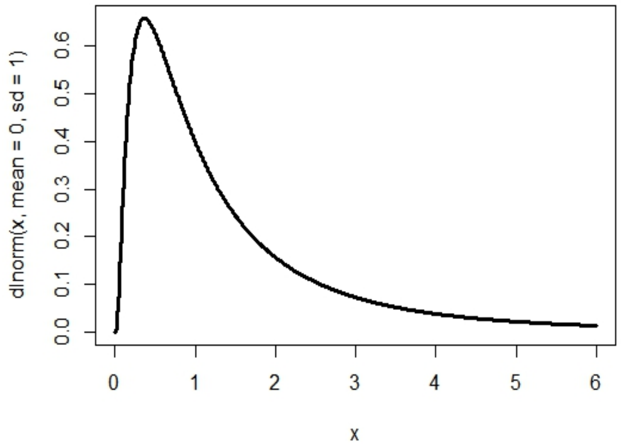
$$g(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$　（正規分布）

$$\Rightarrow \quad \frac{dh}{dx} = \frac{1}{x}$$

（対数正規分布）

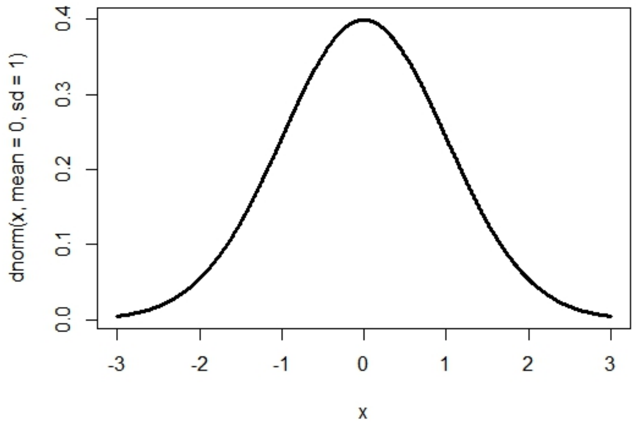
$$f(x) = g(h(x))\left|\frac{dh}{dx}\right| = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \frac{1}{x}$$

対数変換



$$y = \log x$$

$$x = e^y$$

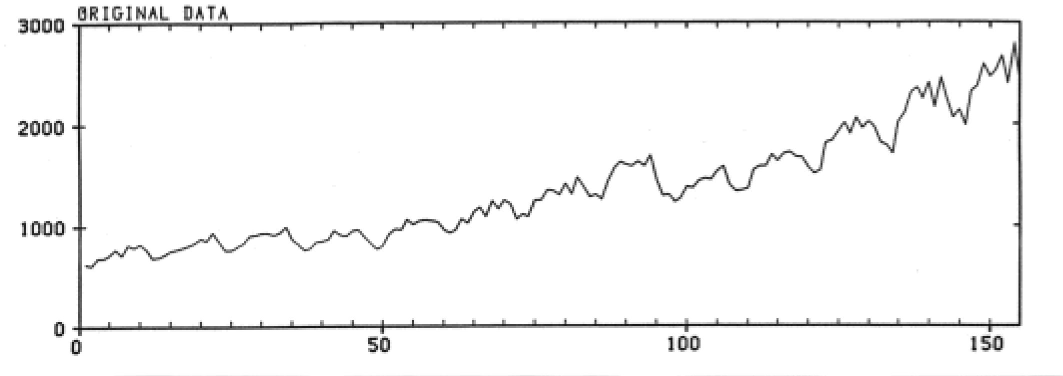


$0 < x < \infty$

$-\infty < y < \infty$

対数正規分布 → 正規分布

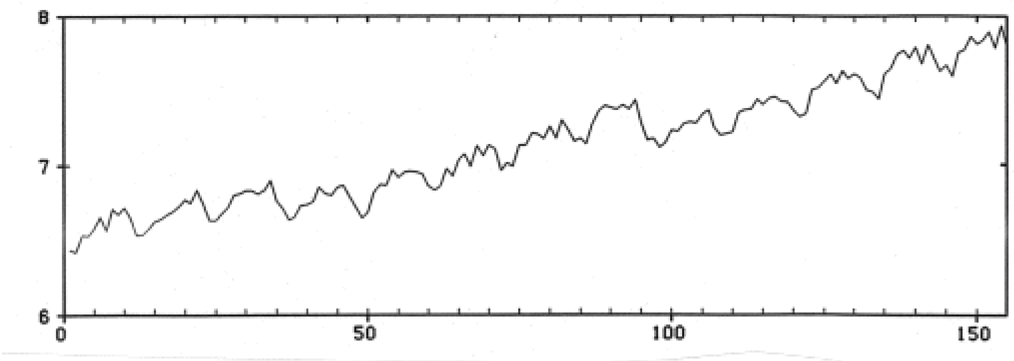
対数変換



- 分散一様化
- トレンドは消えない
- 変数の無制約化

$$(0,\infty) \to (-\infty,\infty)$$

$$y = \log x$$



Box-Cox変換

$$y_\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

$y_\lambda = x^\lambda$のとき

$\lambda = 2$	$y = x^2$
$\lambda = 1$	$y = x$
$\lambda = 0.5$	$y = \sqrt{x}$
$\lambda = 0$	$y = \log x$
$\lambda = -1$	$y = 1/x$

λ の決定法？

ロジット変換

x:　確率, 割合

$$\boxed{y = \log\left(\frac{x}{1-x}\right)}$$

$$(0,1) \quad \leftrightarrow \quad (-\infty, \infty)$$

$$e^y = \frac{x}{1-x} \quad \Rightarrow \quad (1-x)e^y = x$$

$$\Rightarrow e^y = x(1+e^y)$$

逆変換

$$x = \frac{e^y}{1+e^y}$$

より一般の変換

$$y = \log\left(\frac{x-a}{b-x}\right)$$

$$(a,b) \quad \leftrightarrow \quad (-\infty,\infty)$$

$$e^y = \frac{x-a}{b-x} \quad \Rightarrow \quad (b-x)e^y = (x-a)$$

$$\Rightarrow \quad a + be^y = x(1+e^y)$$

$$\Rightarrow \quad x = \frac{a+be^y}{1+e^y}$$

逆変換

$$x = \frac{a+be^y}{1+e^y}$$

階差 (差分)

y_t　時系列

- 1階階差　$\Delta y_t = y_t - y_{t-1}$

直線 $y_t = a + bt$ のとき $\Delta y_t = y_t - y_{t-1} = b$

- 2階階差　$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$

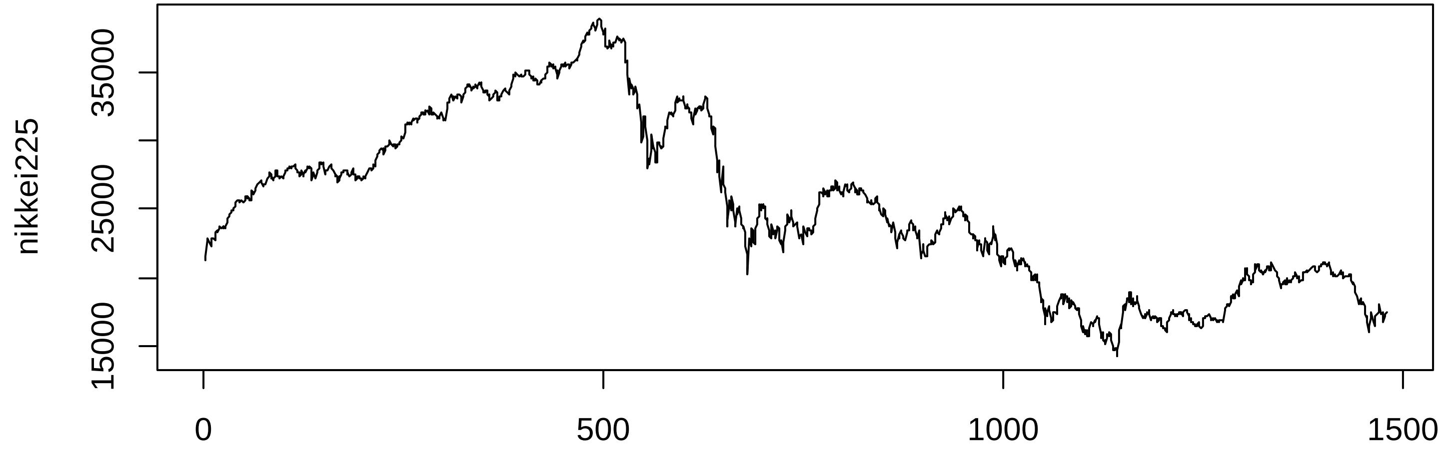
$y_t = a + bt + ct^2$ のとき

$$\Delta y_t = y_t - y_{t-1} = (b-c) + 2ct$$

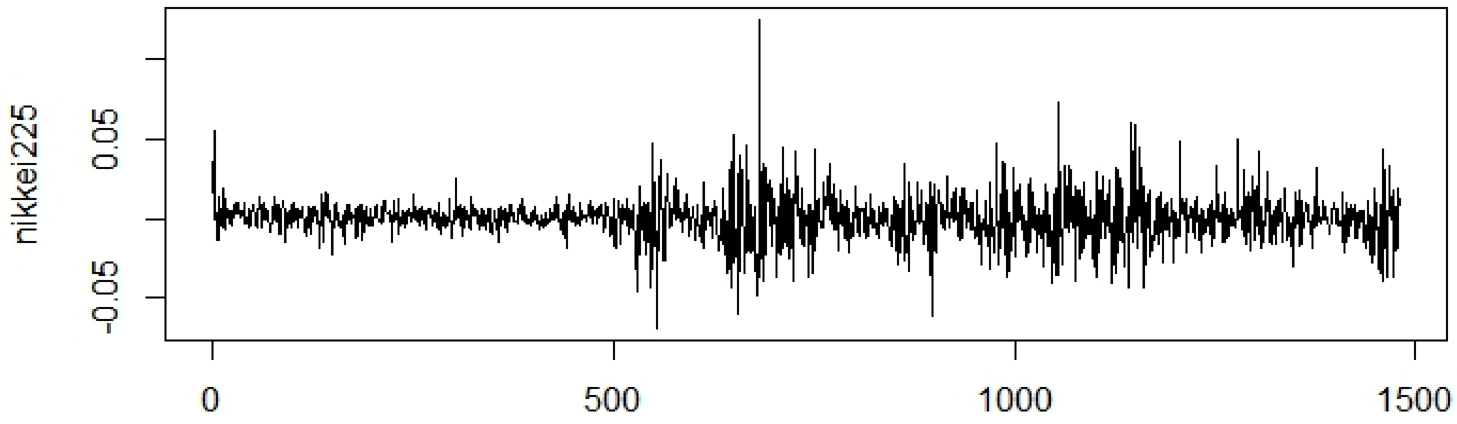
$$\Delta^2 y_t = 2c$$

（対数）差分

```
nikkei225 <- read.csv("Nikkei225_new.csv")
x <- as.ts(nikkei225)
plot(x)
```




```
nikkei225 <- read.csv("nikkei225_new2.csv")
x <- as.ts(nikkei225)
y <- log(x)
z <- diff(y)
plot(z)
```



前期比

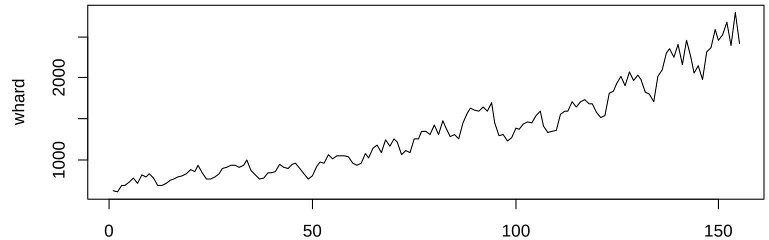
前期比　　$x_t = y_t / y_{t-1}$

$$y_t = T_t$$
$$T_t = (1+\alpha)T_{t-1}$$

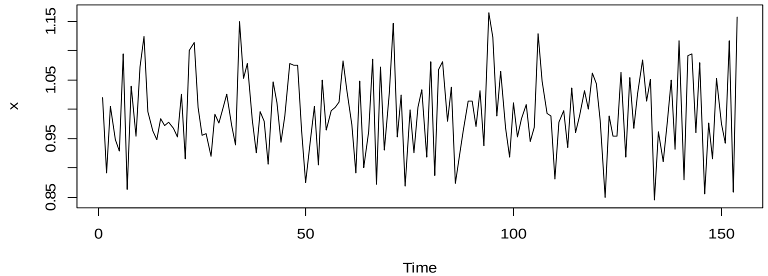
$$x_t = \frac{T_t}{T_{t-1}} = \frac{(1+\alpha)T_t}{T_{t-1}} = 1+\alpha$$

実際は $y_t = T_t w_t$ または $T_t + w_t$

$$x_t = \frac{y_t}{y_{t-1}} = \frac{T_t \cdot w_t}{T_{t-1} \cdot w_{t-1}}$$
$$= \frac{(1+\alpha)T_{t-1} \cdot w_t}{T_{t-1} \cdot w_{t-1}} = (1+\alpha)\frac{w_t}{w_{t-1}}$$



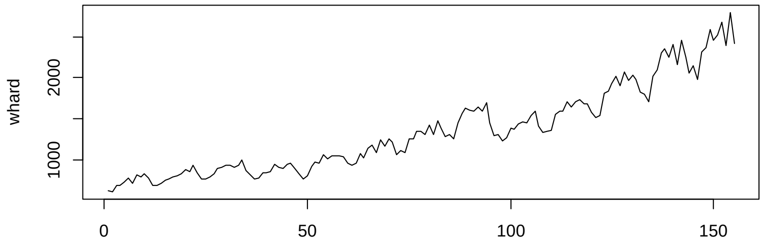

```
x <- as.ts(whard)
plot(x/lag(x))
```



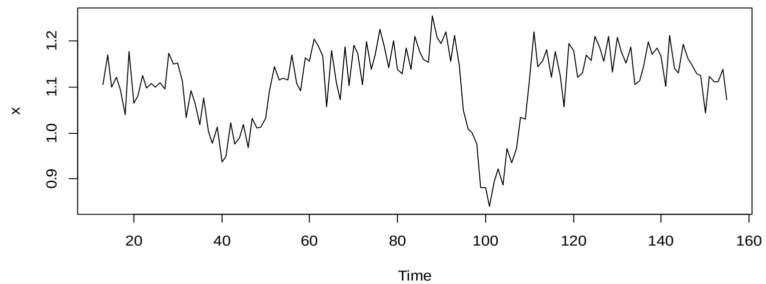
前年同期比

前年同期比　　$x_t = y_t / y_{t-p}$

$y_t = S_t \qquad S_t \cong S_{t-p}$ ⇒ $x_t = \dfrac{y_t}{y_{t-p}} = \dfrac{S_t}{S_{t-p}} \cong 1$



plot(x/lag(x,k=-12))



注意

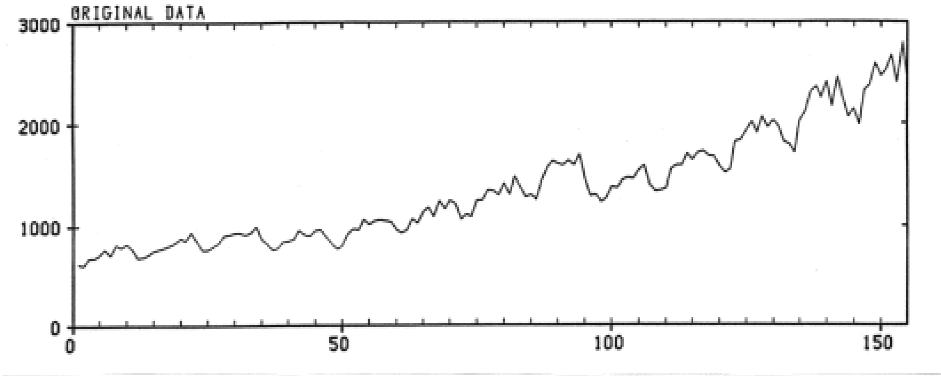
$$y_t = a + bt + \varepsilon_t$$

⇒

$$\Delta y_t = b + \varepsilon_t - \varepsilon_{t-1}$$

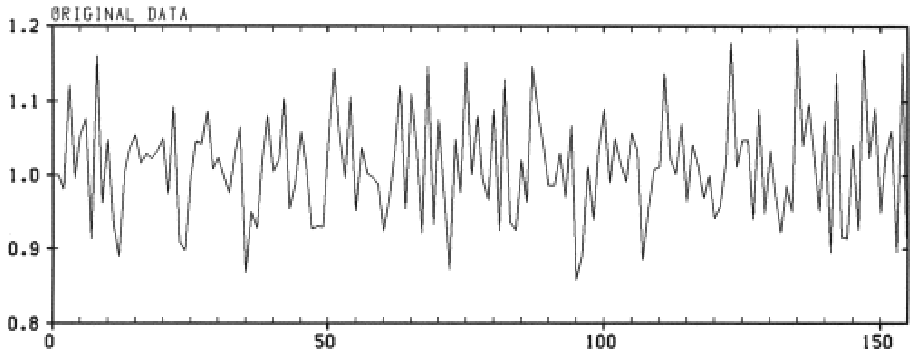
$$\Delta^2 y_t = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$$

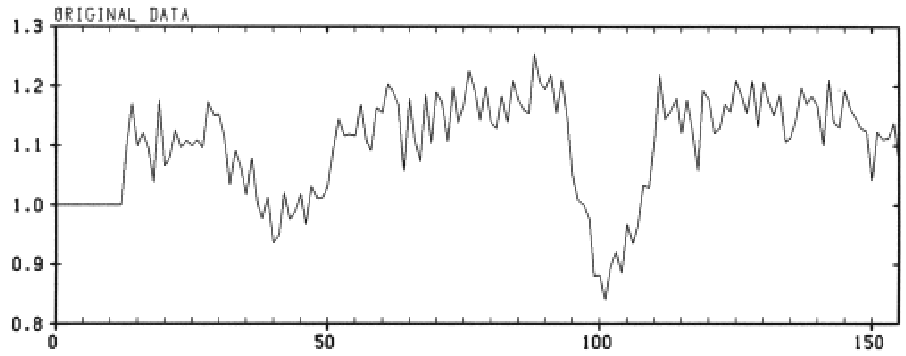
前期比，前年同期比



前期比

前年同期比



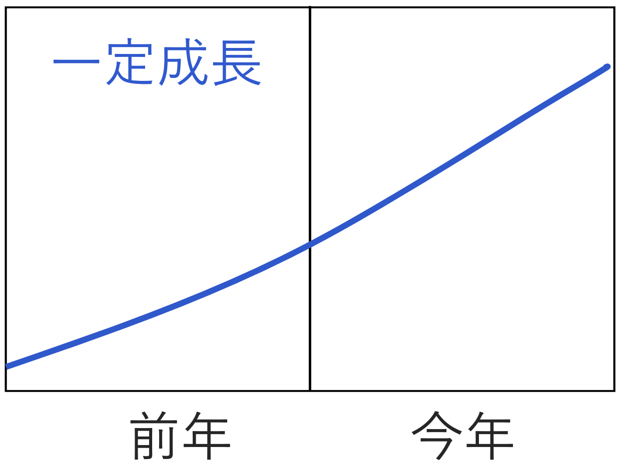


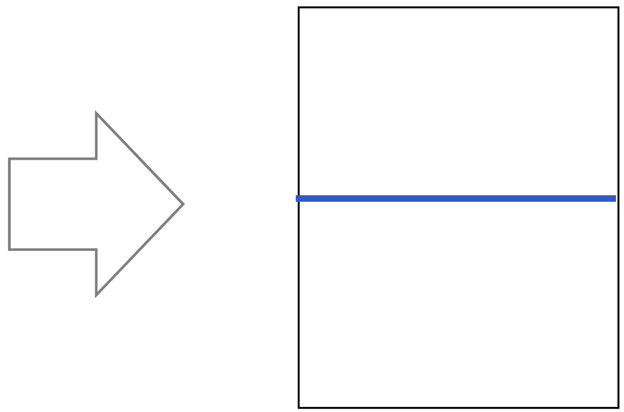
前年同期比の問題点

$$y_t = T_t \cdot S_t$$

T_t トレンド　　S_t 季節成分

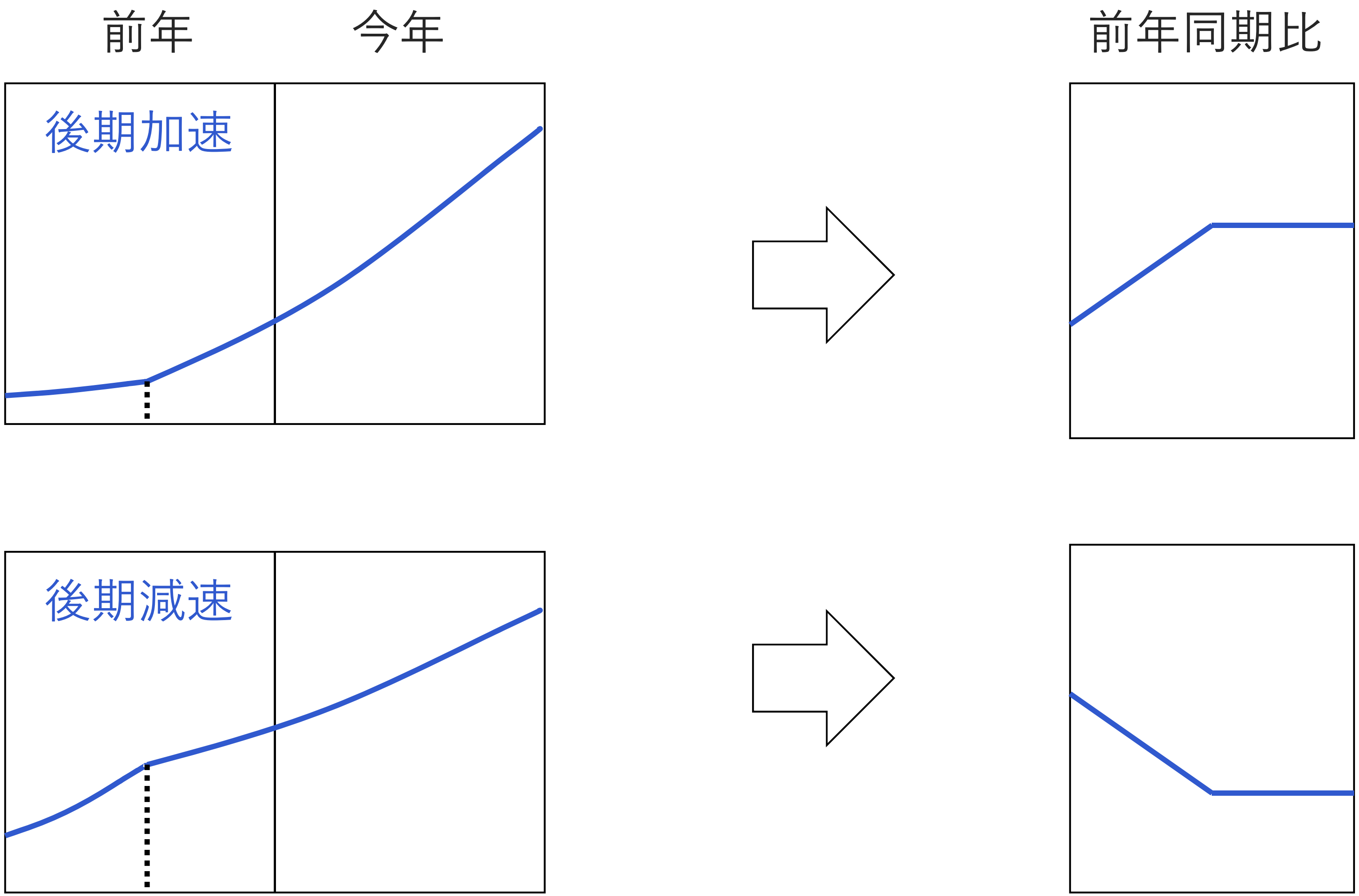
$$\frac{y_t}{y_{t-12}} = \frac{T_t \cdot S_t}{T_{t-12} \cdot S_{t-12}} \cong \frac{T_t}{T_{t-12}}$$



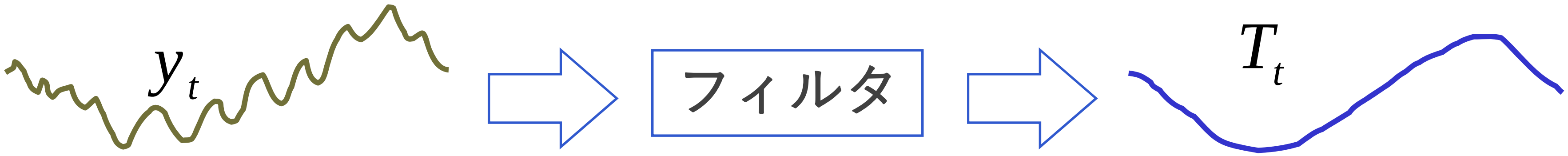


前年同期比

前年同期比の問題点



移動平均フィルタ

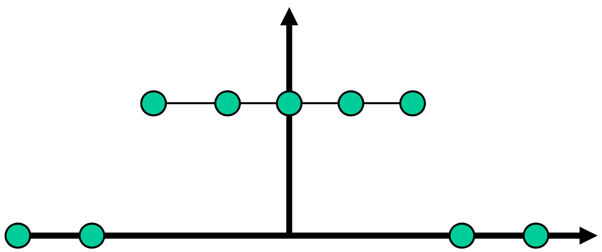


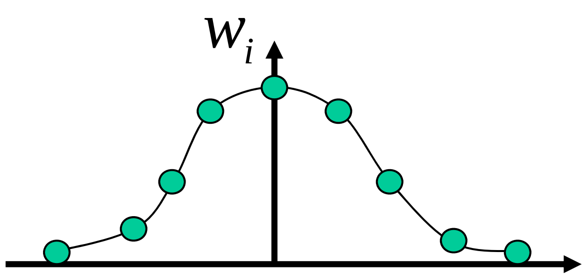
$$t_n = \frac{1}{2k+1}\left(y_{t-k} + \cdots + y_t + \cdots + y_{t+k}\right)$$

$$t_n = w_{-k} y_{t-k} + \cdots + w_0 y_t + \cdots + w_k y_{t+k}$$

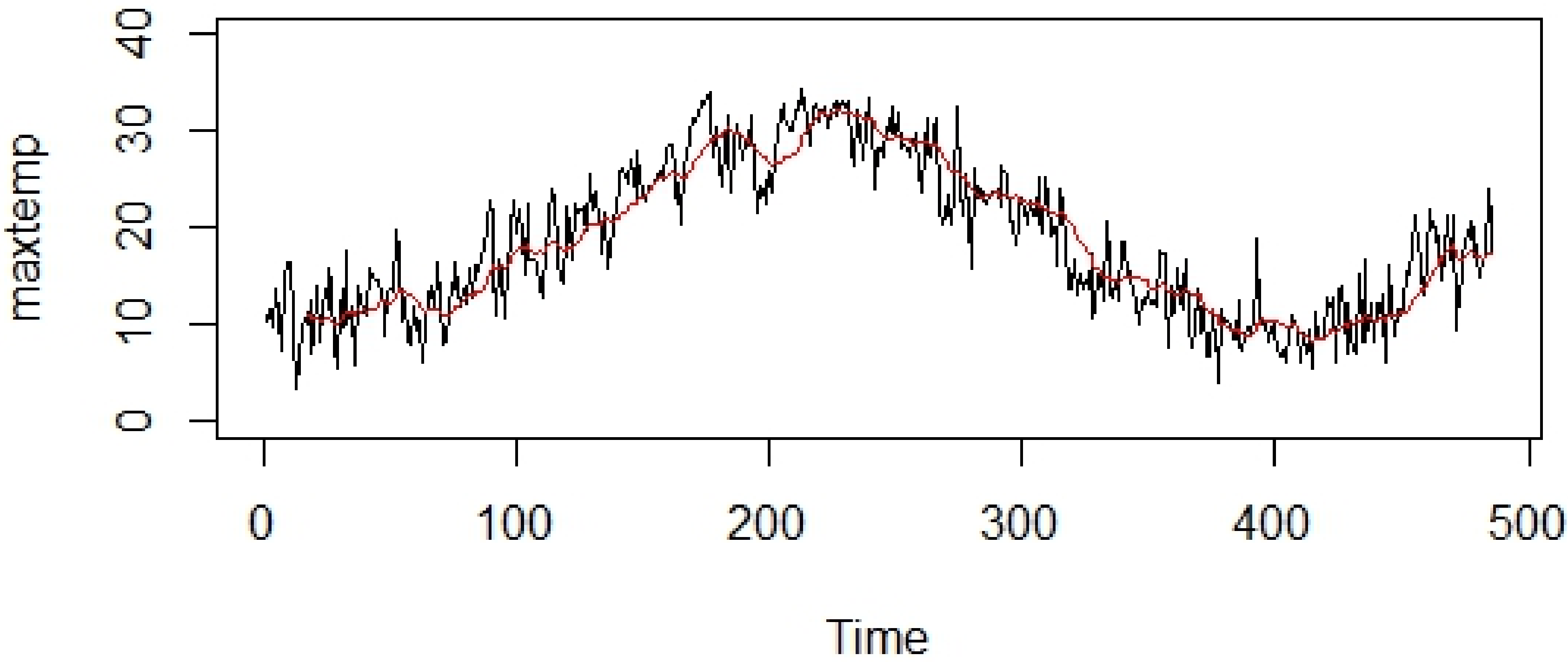
重みつき移動平均　　w_i　重み係数

$2k+1$：項数





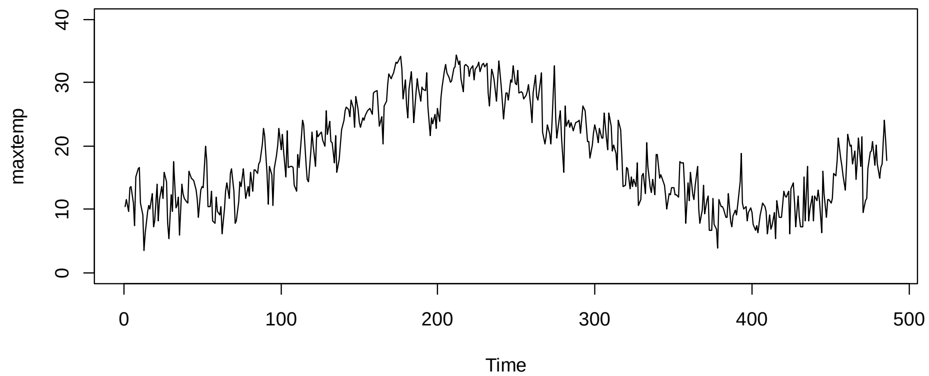
```
plot(maxtemp,ylim=c(0,40))
x <- SMA(maxtemp,17)
lines(x,col=2,lwd=1)
```



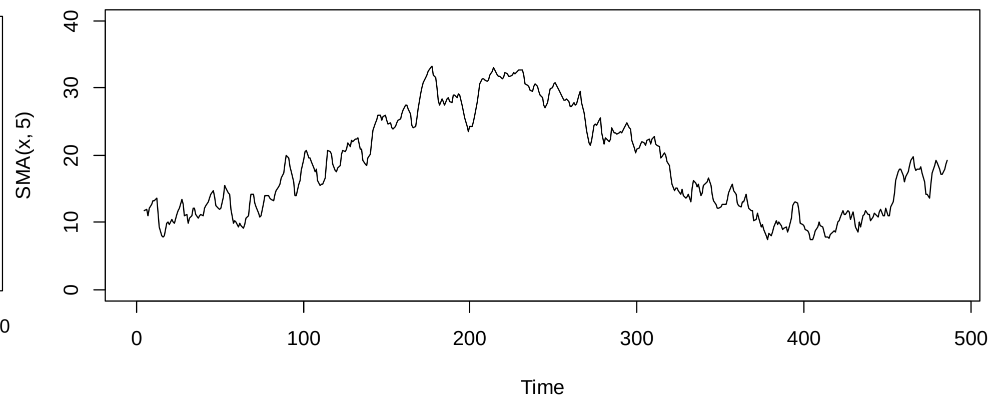
項数の影響

```
options(repos="http://cran.ism.ac.jp")
install.packages("TTR")
library(TTR)
```

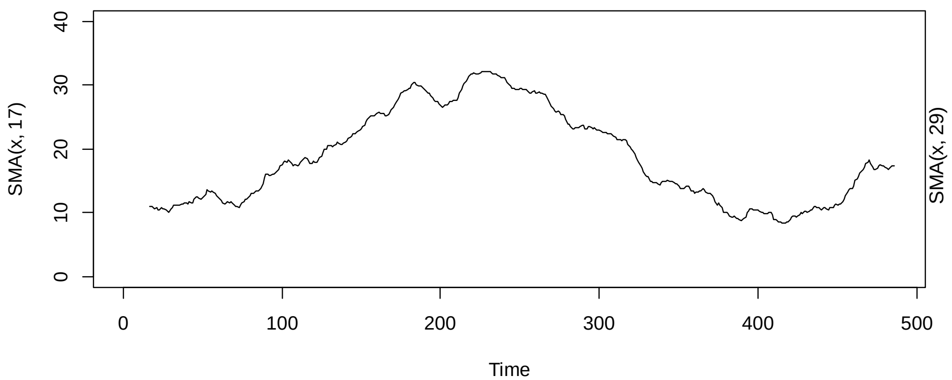
plot(x,ylim=c(0,40))



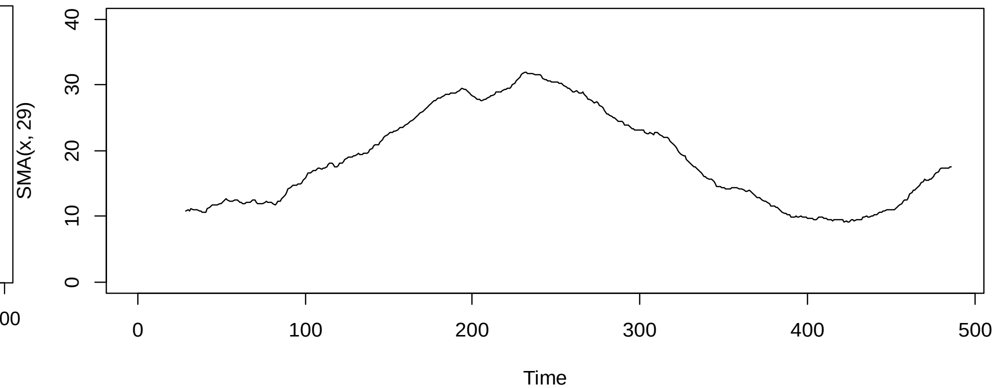
plot(SMA(x,5),ylim=c(0,40))



plot(SMA(x,17),ylim=c(0,40))



plot(SMA(x,29),ylim=c(0,40))



移動平均の性質

$$y_n = T_n + w_n, \qquad T_n = a + bn \text{ の場合}$$

$(2k+1)$項移動平均

$$
\begin{aligned}
t_n &= \tfrac{1}{2k+1}\left(y_{n-k} + \cdots + y_n + \cdots + y_{n+k}\right) \\
&= \tfrac{1}{2k+1}\left(T_{n-k} + \cdots + T_n + \cdots + T_{n+k}\right) + \tfrac{1}{2k+1}\left(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}\right) \\
&= T_k + \tfrac{1}{2k+1}\left(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}\right)
\end{aligned}
$$

$$E\left(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}\right) = E(w_{n-k}) + \cdots + E(w_n) + \cdots + E(w_{n+k}) = 0$$

$$
\begin{aligned}
&E\left(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}\right)^2 \\
&\quad = E(w_{n-k}^2 + \cdots + w_{n+k}^2) + 2E(w_{n-k}w_{n-k+1} + \cdots + w_{n+k-1}w_{n+k}) = (2k+1)\sigma^2
\end{aligned}
$$

$$E(w_n^2) = \sigma^2 \qquad E(w_n w_m) = 0 \quad (n \neq m)$$

$$E\left(\frac{w_{n-k} + \cdots + w_n + \cdots + w_{n+k}}{2k+1}\right)^2 = \frac{1}{(2k+1)^2} E\left(w_{n-k} + \cdots + w_n + \cdots + w_{n+k}\right)^2 = \frac{\sigma^2}{2k+1}$$

平均： 不変　　　分散： $\dfrac{1}{2k+1}$

移動メディアン・フィルタ

$$\boxed{t_n = \text{median}(t_{n\text{-}k}, \ldots, y_n, \ldots, y_{n+k})}$$

$y_{(n-k)}, \cdots, y_{(n+k)}$:　$y_{n-k}, \ldots, y_{n+k}$　を小さい順に並べたもの

$$\text{median}(y_1, \cdots, y_n) = \begin{cases} y_{\left(\frac{n+1}{2}\right)} & n：奇数 \\ \frac{1}{2}\left(y_{\left(\frac{n}{2}-1\right)} + y_{\left(\frac{n}{2}+1\right)} \right) & n：偶数 \end{cases}$$

長所：　異常値，急激な構造変化に対応できる

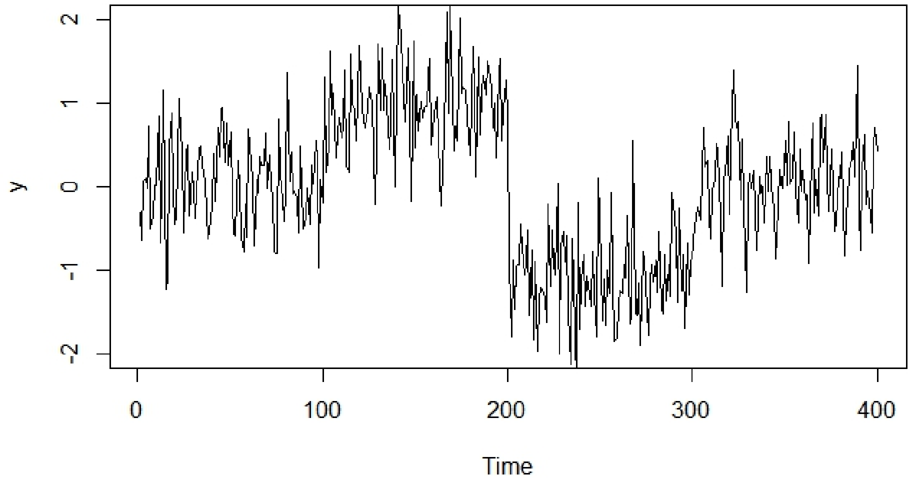
移動平均フィルタと移動メディアンフィルタ

```
# 移動平均フィルタ

plot(maxtemp,ylim=c(0,40))
y <- maxtemp
ndata <- length(maxtemp)
y[1:ndata] <- NA
kfilter <- 17
n0 <- kfilter+1
n1 <- ndata-kfilter
for(i in n0:n1){
   i0 <- i-kfilter
   i1 <- i+kfilter
   y[i] <- mean(maxtemp[i0:i1])
}
lines(y,col=2,ylim=c(0,40))
```

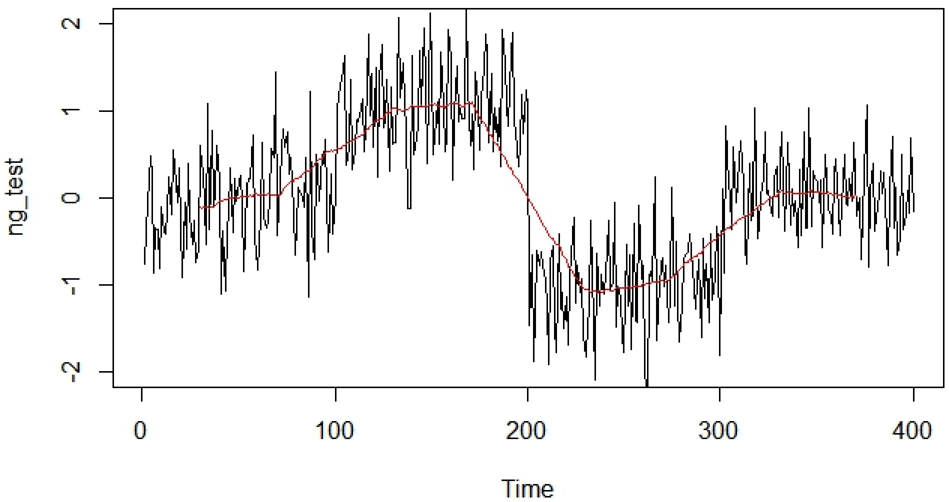
```
# 移動メディアンフィルタ

plot(maxtemp,ylim=c(0,40))
y <- maxtemp
ndata <- length(maxtemp)
y[1:ndata] <- NA
kfilter <- 17
n0 <- kfilter+1
n1 <- ndata-kfilter
for(i in n0:n1){
i0 <- i-kfilter
i1 <- i+kfilter
y[i] <- median(maxtemp[i0:i1])
}
lines(y,col=3,ylim=c(0,40),lwd=2)
```

移動平均フィルタと移動メディアンフィルタ

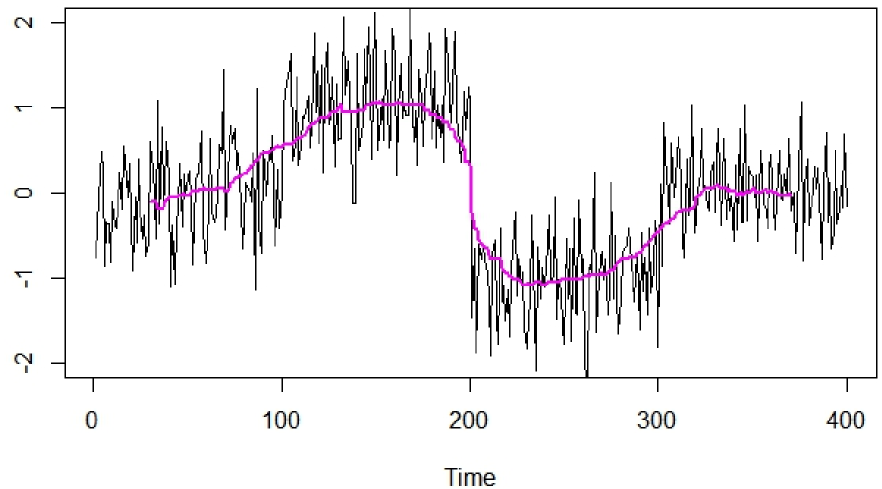
```
x <- rep(0,400)
x[101:200] <- 1
x[201:300] <- -1
y <- x + rnorm(400, mean=0, sd=0.5)
plot(y)
ng_test <-as.ts(y)
plot(ng_test)
```



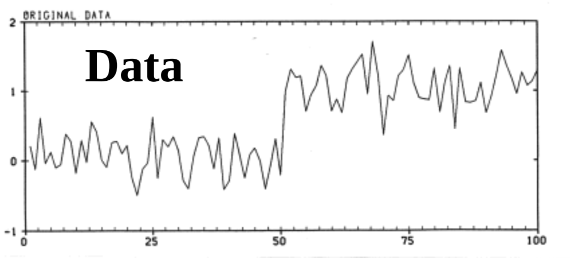
Moving-average(29)

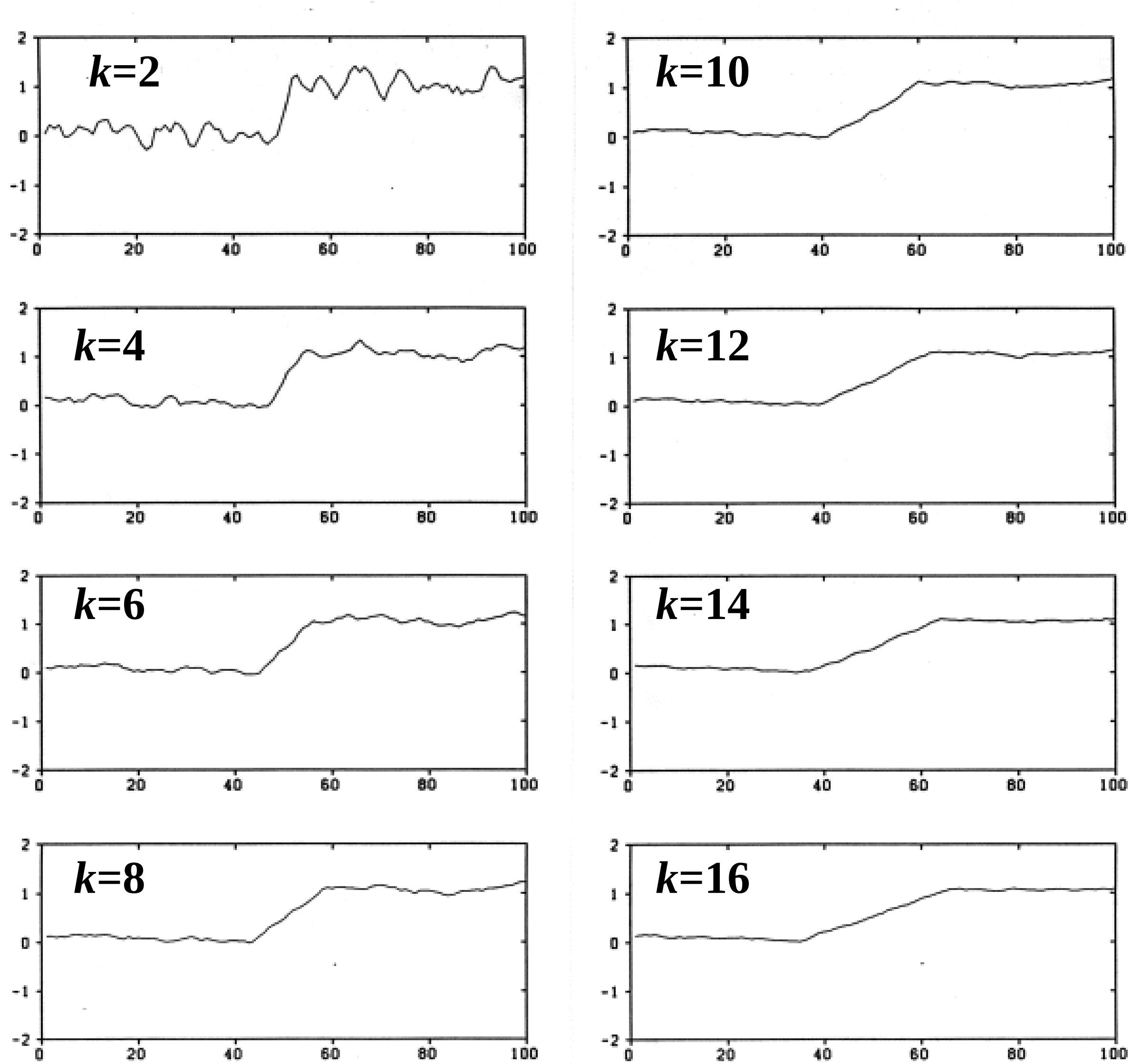


Moving-median(29)



移動平均フィルタ

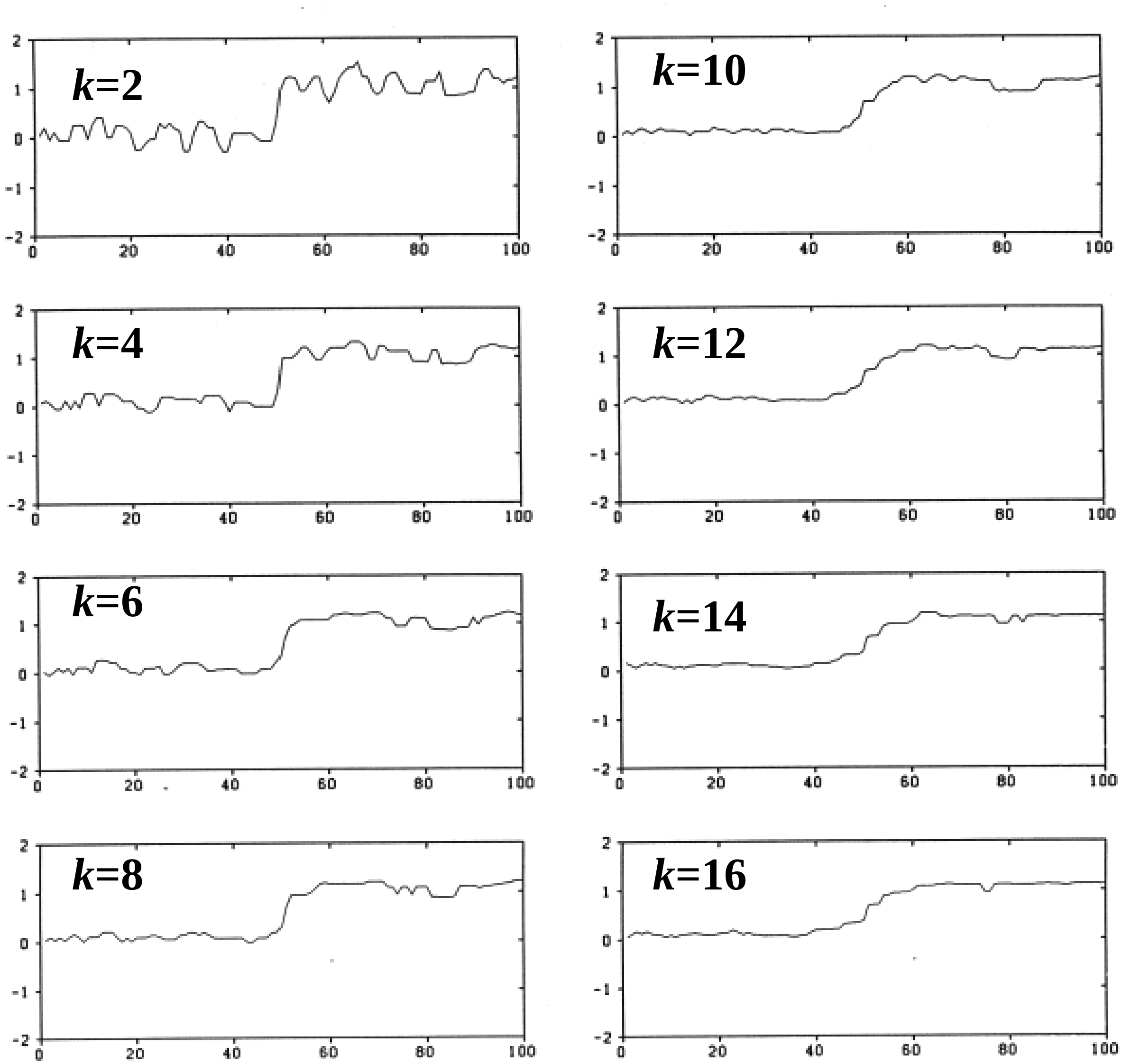




長所
- 滑らかな推定値

欠点
- 構造変化を正確に検出できない
- 異常値に敏感

移動メディアン

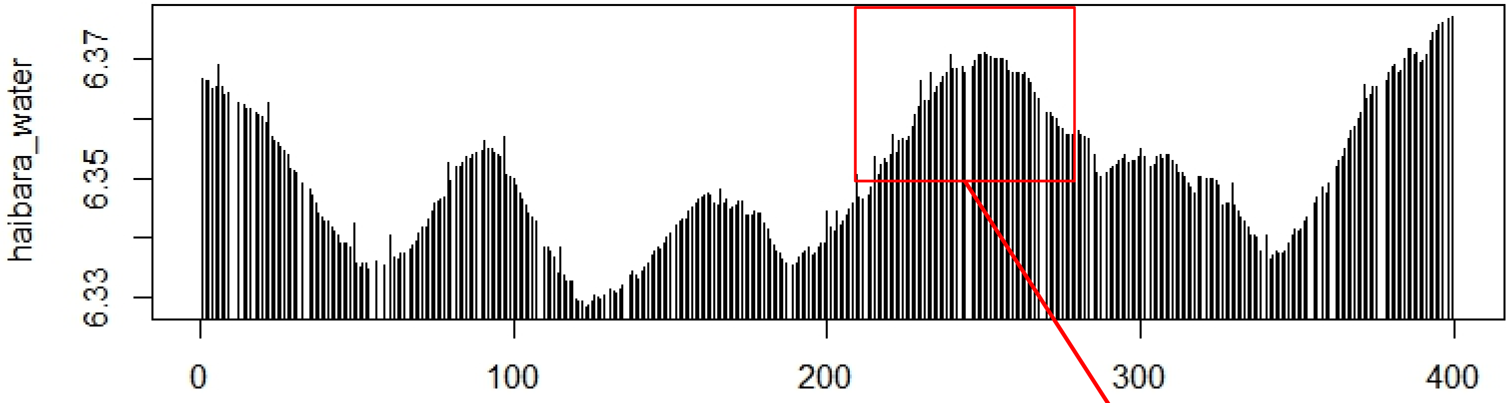


長所
- 構造変化を検出できる
- 異常値に頑健

欠点
- 変動が大きい

欠測値と異常値（外れ値）

```
# 榛原 地下水位データ
haibara <- as.ts(read.csv("haibara_new.csv"))
haibara_water <- haibara[,2]
plot(haibara_water,type="h")
```




```
plot(haibara_water[211:280],type="h",ylim=c(6.35,6.38))
```