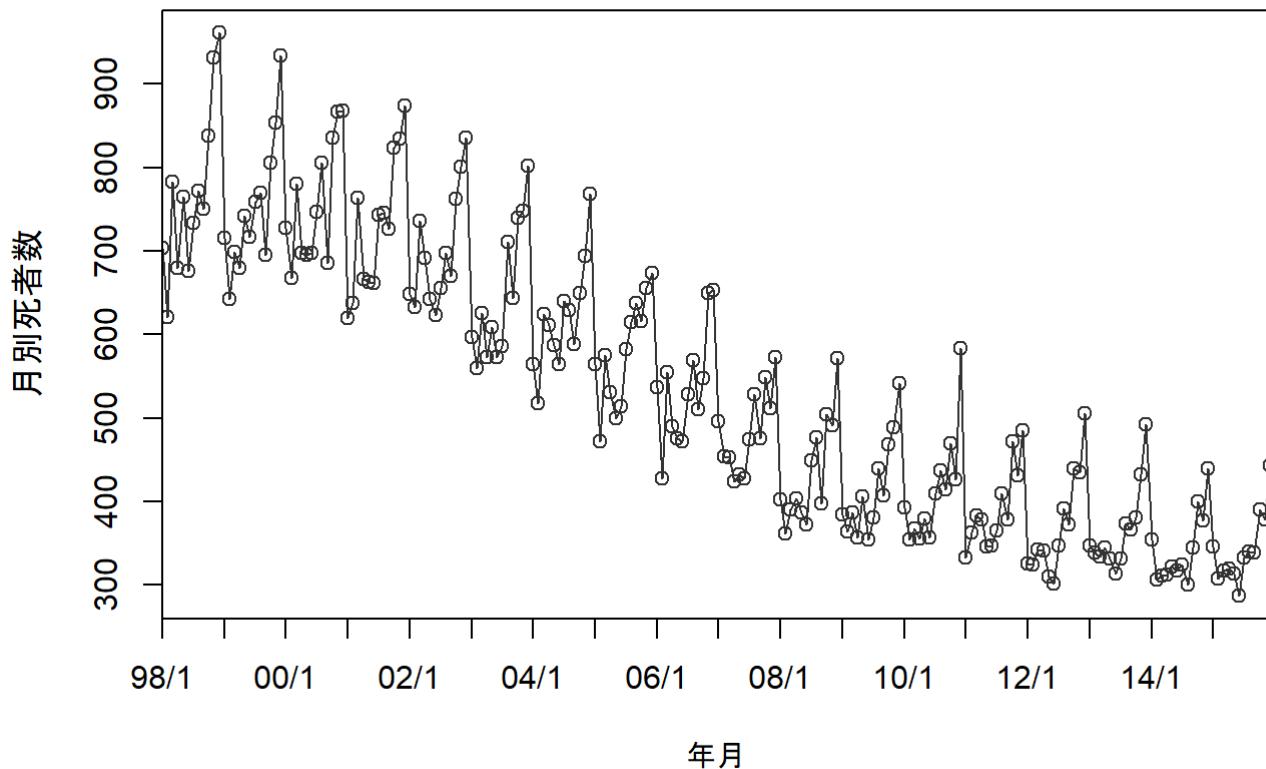


課題3

選択課題③ 解析例④の交通事故死者数の月次推移について観測値が正規分布に従う通常の構造時系列モデルを適用し、より良いモデルを追求する。

```
plot(tsuki, type="o", main="交通事故死者数の月次推移", xaxt="n", ylab="月別死者数", xlab="年月", col=
gray(0.2), xaxs="i")
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

交通事故死者数の月次推移



- モデル1:

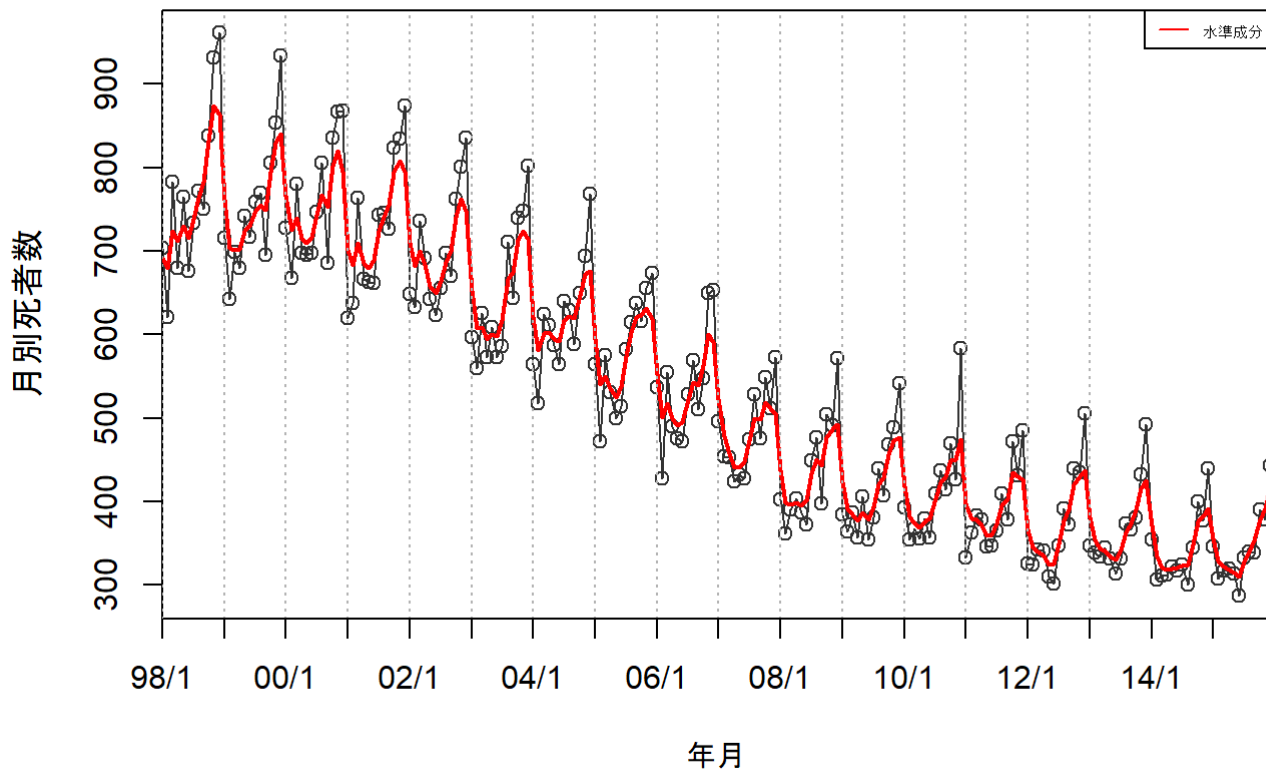
- まずはローカルレベルモデルで解析し、モデルの当てはまりを評価するためにAICを算出する。

```
# ローカルレベルモデル
mod_tri <- SSMmodel(tsuki ~ SSMtrend(1, Q=NA), H=NA)
fit_tri <- fitSSM(mod_tri, numeric(2), method = "SANN")
kfs_tri <- KFS(fit_tri$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri <- kfs_tri$logLik - sum(kfs_tri$Finf>0) * log(2*pi)/2
AIC_tri <- -2*logLik_tri + 2*2
print(paste("AIC_ローカルレベルモデル:", AIC_tri))
```

```
## [1] "AIC_ローカルレベルモデル: 2460.76641910979"
```

```
plot(tsuki, type="o", main="ローカルレベルモデル", xaxt="n", ylab="月別死者数", xlab="年月", col=gray(0.2), xaxs="i")
lines(kfs_tri$alphahat[, "level"], col=2, lwd=2)
legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

ローカルレベルモデル



• モデル2:

- データのプロットを見ると、1年単位の周期性のような変動傾向が見られるので、先ほどのローカルレベルモデルに季節変動を加えてみる。

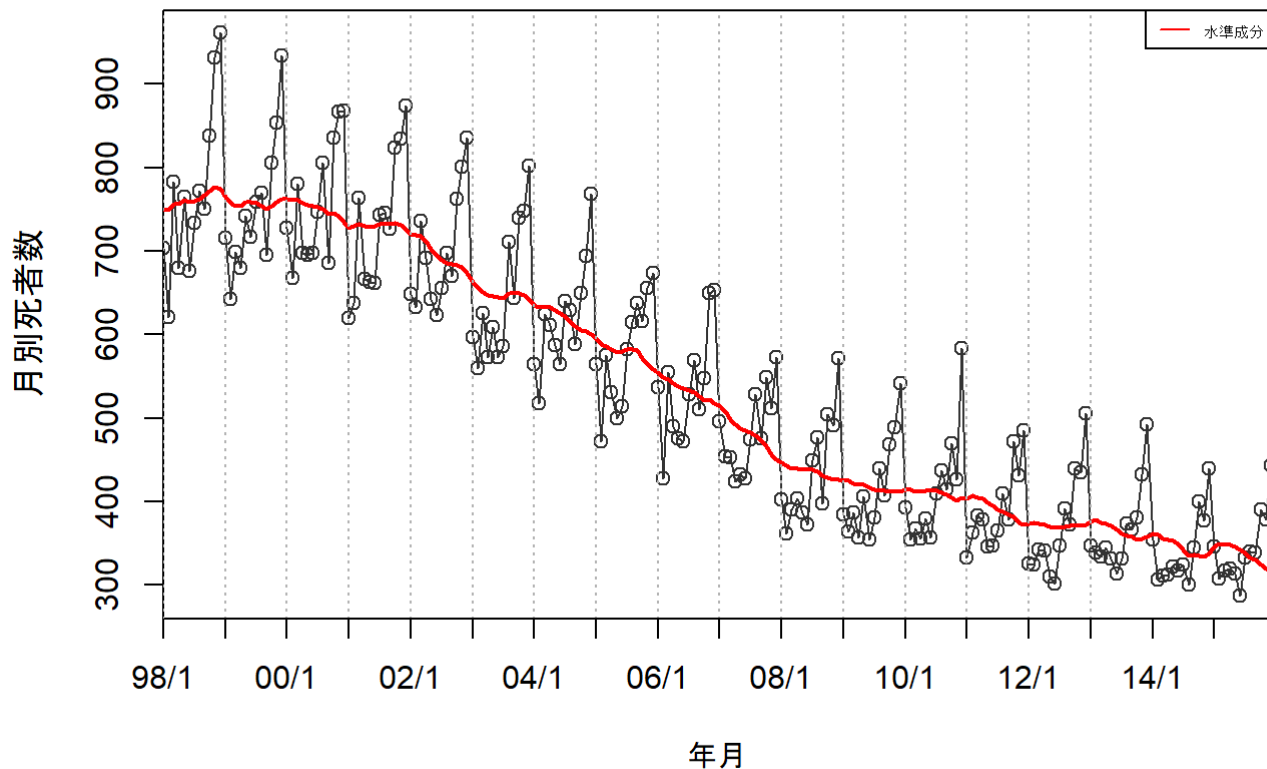
```
# ローカルレベルモデル+固定季節変動 (対数変換なし)
mod_tri1 <- SSMModel(tsuki ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=0), H=NA)
fit_tri1 <- fitSSM(mod_tri1, numeric(2), method = "SANN")
kfs_tri1 <- KFS(fit_tri1$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri1 <- kfs_tri1$logLik - sum(kfs_tri1$Finf>0) * log(2*pi)/2
AIC_tri1 <- -2*logLik_tri1 + 2*(2 + 12)
print(paste("AIC_ローカルレベルモデル+固定季節変動:", AIC_tri1))
```

```
## [1] "AIC_ローカルレベルモデル+固定季節変動: 2078.57392707883"
```

```
plot(tsuki, type="o", main="ローカルレベルモデル+固定季節変動", xaxt="n", ylab="月別死者数", xlab="年月", col=gray(0.2), xaxs="i")
lines(kfs_tri1$alphahat[, "level"], col=2, lwd=2)

legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

ローカルレベルモデル+固定季節変動



- モデル3:
 - 季節変動の程度がありそうなので、固定季節変動を可変季節変動に変更してみる。

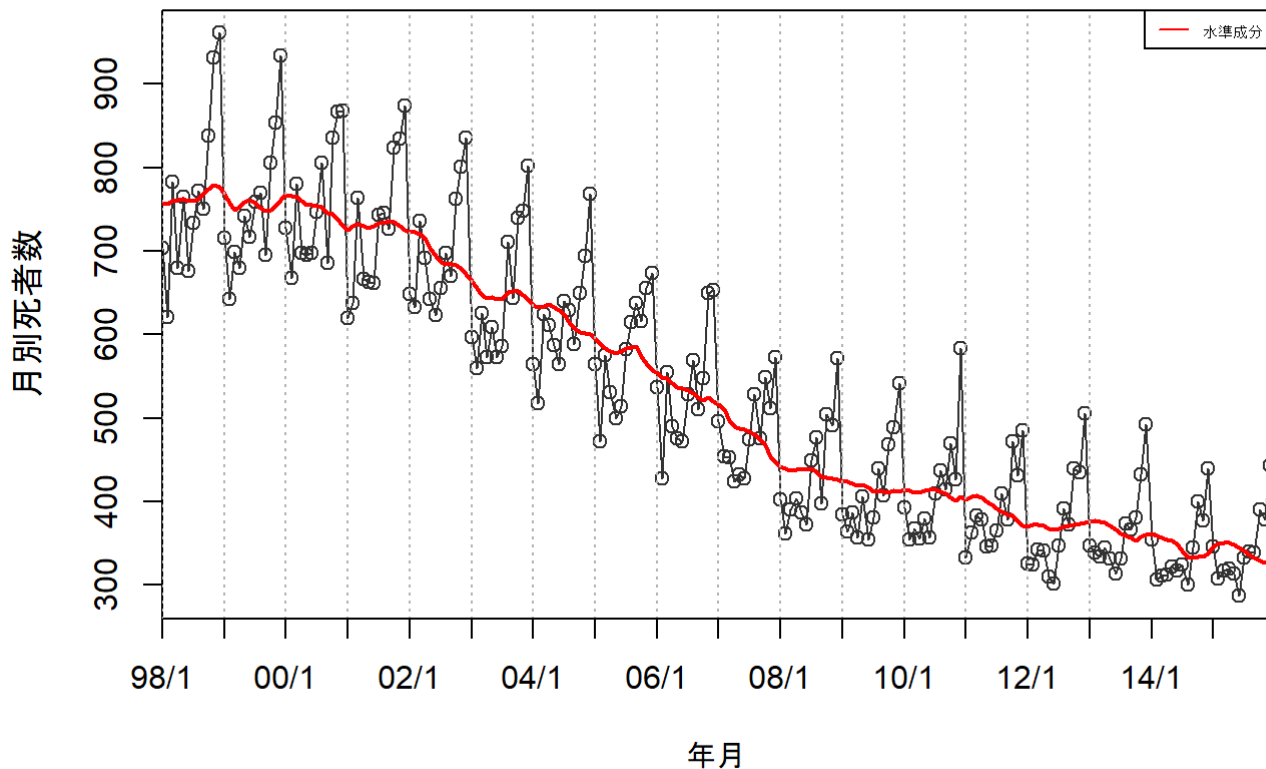
```
# ローカルレベルモデル+可変季節変動 (対数変換なし)
mod_tri2 <- SSMmodel(tsuki ~ SSMtrend(1, Q=NA) + SSMseasonal(12, Q=NA), H=NA)
fit_tri2 <- fitSSM(mod_tri2, numeric(3), method = "SANN")
kfs_tri2 <- KFS(fit_tri2$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri2 <- kfs_tri2$logLik - sum(kfs_tri2$Finf>0) * log(2*pi)/2
AIC_tri2 <- -2*logLik_tri2 + 2*(3 + 12)
print(paste("AIC_ローカルレベルモデル+可変季節変動:", AIC_tri2))
```

```
## [1] "AIC_ローカルレベルモデル+可変季節変動: 2043.09870233152"
```

```
plot(tsuki, type="o", main="ローカルレベルモデル+可変季節変動", xaxt="n", ylab="月別死者数", xlab="年月", col=gray(0.2), xaxs="i")
lines(kfs_tri2$alphahat[, "level"], col=2, lwd=2)

legend("topright", "水準成分", lwd=1, col=2, cex = 0.5)
abline(v=1+0:15*12, lty=3, col=gray(0.7))
axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1", "06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

ローカルレベルモデル+可変季節変動



- モデル4:
 - モデル3を平滑化トレンドモデルにしてみる。

```
# 平滑化トレンドモデル+可変季節変動 (対数変換なし)
mod_tri3 <- SSMmodel(tsuki ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit_tri3 <- fitSSM(mod_tri3, numeric(3), method = "SANN")
kfs_tri3 <- KFS(fit_tri3$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri3 <- kfs_tri3$logLik - sum(kfs_tri3$Finf > 0) * log(2*pi)/2
AIC_tri3 <- -2*logLik_tri3 + 2*(3 + 13)
print(paste("AIC_平滑化トレンドモデル+可変季節変動:", AIC_tri3))
```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動: 2032.90509293277"
```

- ここまでで最も良いAICを記録したのは、平滑化トレンドモデル+可変季節変動。
- 続いて、上記モデルで目的変数に対数変換を施してみる。

```
# 平滑化トレンドモデル+可変季節変動（対数変換あり）
mod_tri4 <- SSMModel(log(tsuki) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA ), H=NA)
fit_tri4 <- fitSSM(mod_tri4, numeric(3), method = "SANN")
kfs_tri4 <- KFS(fit_tri4$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri4 <- kfs_tri4$logLik - sum(kfs_tri4$Finf > 0) * log(2*pi)/2
AIC_tri4 <- -2*logLik_tri4 + 2*( 3 + 13 ) + 2*sum(log(tsuki))
print(paste("AIC_平滑化トレンドモデル+可変季節変動（対数変換あり） :", AIC_tri4))
```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動（対数変換あり） : 2174.93509205638"
```

- AICが悪化してしまった。
- もう少し違うパターンからより良いモデルを探索してみることにする。
- 観測値のプロットに立ち戻ると、およそ1年単位での変動周期が、毎年上下動している。
- そこで、次のモデルを定義する。
- モデル6:
 - 平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり）

```
# 平滑化トレンドモデル+可変季節変動（対数変換あり）
lag_tri6 <- c(tsuki/stats::lag(tsuki, k=-12))
first <- rep(1, length=12)
lag_tri6 <- ts(append(first, lag_tri6))
mod_tri6 <- SSMModel(log(tsuki) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA ) + SSMregre
ssion(log(tsuki) ~ lag_tri6 , Q = NA), H = NA)
fit_tri6 <- fitSSM(mod_tri6, numeric(4), method = "SANN")
kfs_tri6 <- KFS(fit_tri6$model, smoothing=c("state", "mean", "disturbance"))
logLik_tri6 <- kfs_tri6$logLik - sum(kfs_tri6$Finf > 0) * log(2*pi)/2
AIC_tri6 <- -2*logLik_tri6 + 2*( 4 + 12 ) + 2*sum(log(tsuki))
print(paste("AIC_平滑化トレンドモデル+可変季節変動+前年同期比 :", AIC_tri6))
```

```
## [1] "AIC_平滑化トレンドモデル+可変季節変動+前年同期比 : 2018.27897295501"
```

- AICが若干向上した。
- 次に、「良いモデル」の追求について、「モデルの当てはまり」の観点から最良を示したモデル6が長期予測の観点からはどうなのかを確かめてみる。
- 今回は、直近4年分を欠損値として予測対象にする。

```
tsukiNA <- tsuki
tsukiNA[157:216] <- NA

mod_tri_pre1 <- SSMModel(log(tsukiNA) ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA ) + SS
Mregression(log(tsukiNA) ~ lag_tri6 , Q = NA), H = NA)
fit_tri_pre1 <- fitSSM(mod_tri_pre1, numeric(4), method = "SANN")
conf_tri_pre1 <- predict(fit_tri_pre1$model, interval="confidence", level=0.95)
pre_tri_pre1 <- predict(fit_tri_pre1$model, interval="prediction", level=0.95)
```

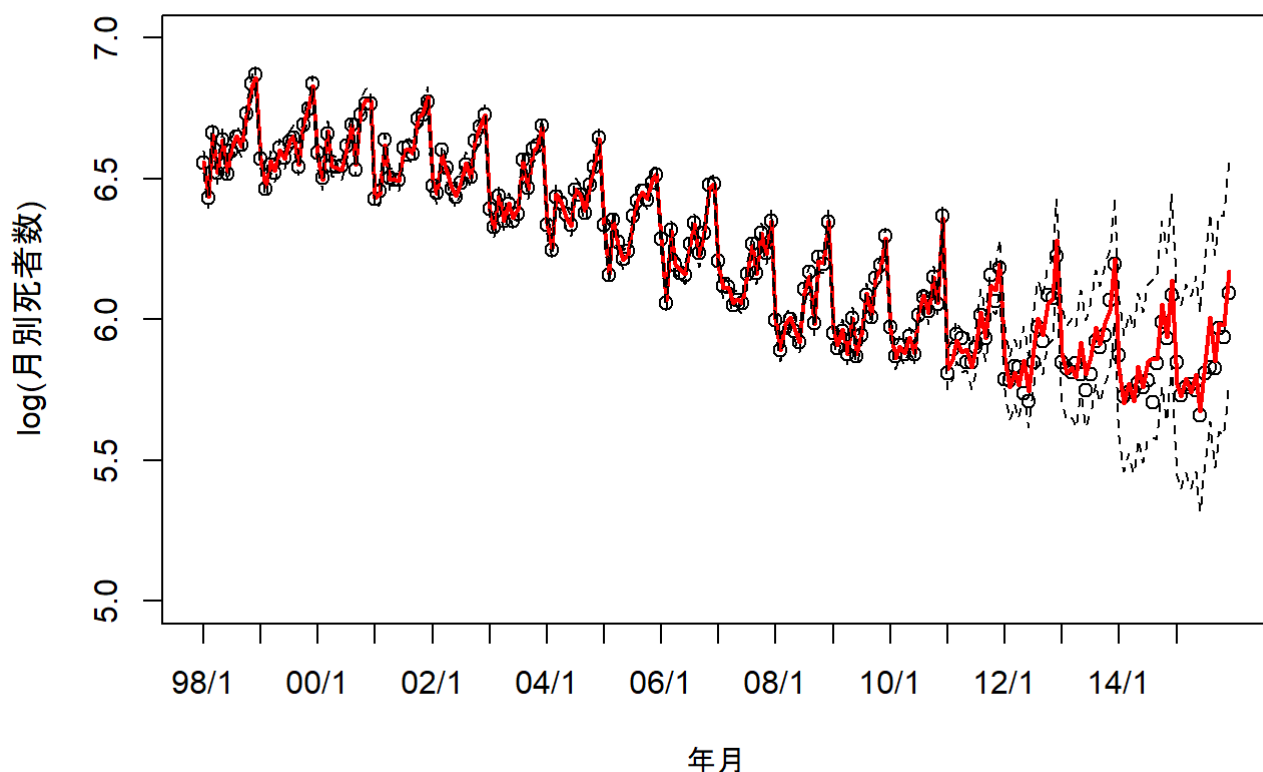
- 予測結果をプロットしてみる。

```
plot(log(tsuki), type="b", xaxt="n", main="(最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比", ylab="log(月別死者数)", xlab="年月", lty=3, ylim=c(5, 7))

lines(pre_tri_pre1[, 1], lwd=2, col="red")
lines(pre_tri_pre1[, 2], lty=2)
lines(pre_tri_pre1[, 3], lty=2)

axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

(最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比



- 比較のため、モデル4でも予測してみる。

```
mod_tri_pre2 <- SSMModel(tsukiNA ~ SSMtrend(2, Q=list(0, NA)) + SSMseasonal(12, Q=NA), H=NA)
fit_tri_pre2 <- fitSSM(mod_tri_pre2, numeric(3), method = "SANN")
conf_tri_pre2 <- predict(fit_tri_pre2$model, interval="confidence", level=0.95)
pre_tri_pre2 <- predict(fit_tri_pre2$model, interval="prediction", level=0.95)
```

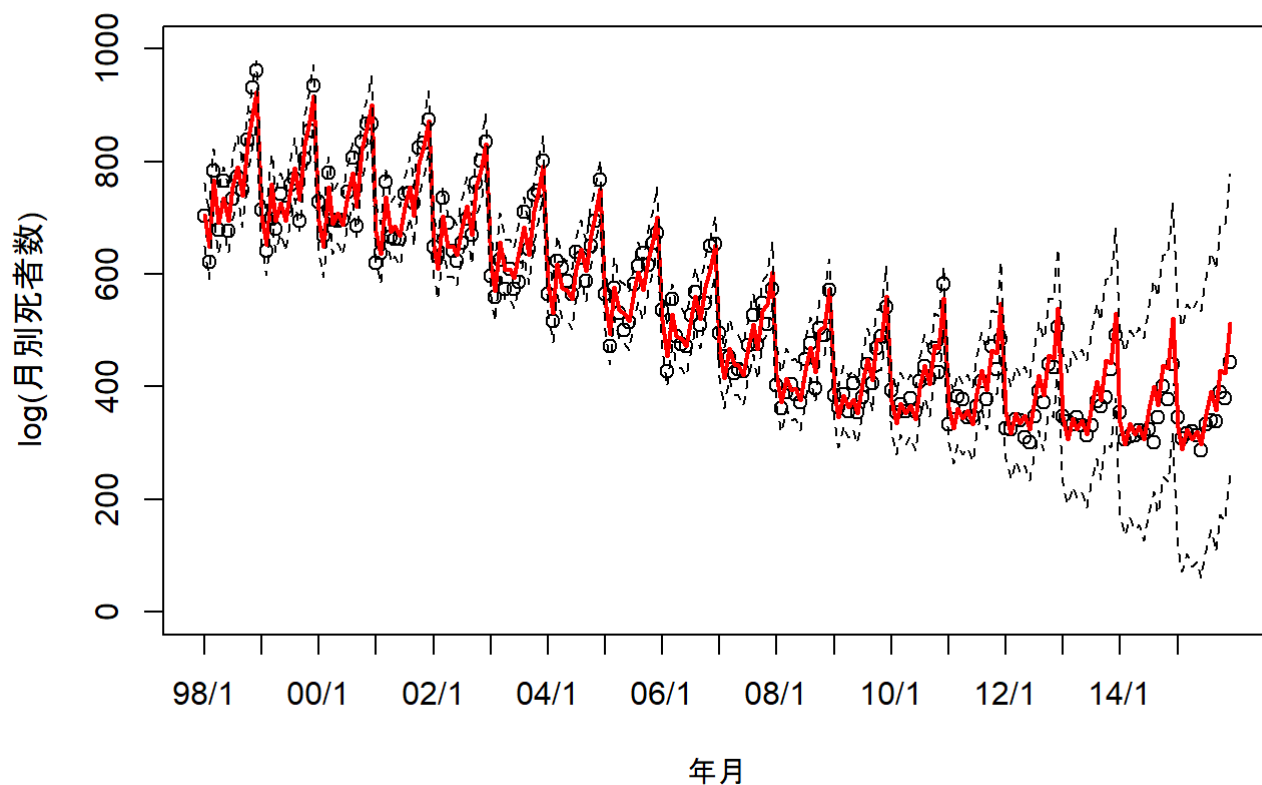
- 予測結果をプロットしてみる。

```
plot(tsuki, type="b", xaxt="n", main="(最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比", ylab="log(月別死者数)", xlab="年月", lty=3, ylim=c(0, 1000))

lines(pre_tri_pre2[, 1], lwd=2, col="red")
lines(pre_tri_pre2[, 2], lty=2)
lines(pre_tri_pre2[, 3], lty=2)

axis(side=1, at=1+0:17*12, labels=c("98/1", "99/1", "00/1", "01/1", "02/1", "03/1", "04/1", "05/1",
"06/1", "07/1", "08/1", "09/1", "10/1", "11/1", "12/1", "13/1", "14/1", "15/1"))
```

(最新3年分予測) 平滑化トレンドモデル+変動季節変動+前年同期比



- モデル6（平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり））では、各月の不定期的な変動要素にも対応しつつ予測していることが伺える一方、モデル4（平滑化トレンドモデル+可変季節変動（対数変換なし））では、一定の季節変動のもとでの予測であることが伺える。
- よって、今回トライした中では、モデルの当てはまり、予測の両面でベターなモデルは モデル6：平滑化トレンドモデル+可変季節変動+前年同期比（対数変換あり）である。