

# 課題2

## 準備

- 2011年をNAとするデータセットを用意する。
- 講義同様、対象年齢を20~60歳とする。また、解析時の目的変数は対数変換を施したものを使用する。

## 解析

- 女性の年齢別死亡率
  - 講義で履修した動的因子モデル（1因子/2因子）を用いて解析する。
- 女性の1因子モデル

```
num_params <- ncol(lmtxf)
# 動的因子モデル（1因子ローカルレベルモデル_女性死亡率対数）
mod_lmtxf1 <- SSMModel(lmtxf~SSMtrend(1,Q=1,"common")+SSMtrend(1,Q=0,"distinct"))

updatefn1 <- function(pars,mod_lmtxf1){
  mod_lmtxf1 <- SSMModel(lmtxf~SSMtrend(1,Q=1,"common")+SSMtrend(1,Q=0,"distinct"))
  mod_lmtxf1$Z[,1,] <- pars[1:num_params]#length(unique(mtx$Age))
  diag(mod_lmtxf1$H[,,]) <- exp(pars[num_params+1])#length(unique(mtx$Age))+1])
  mod_lmtxf1
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxf1 <- fitSSM(mod_lmtxf1,c(rep(0.1,num_params),-5),updatefn1,method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_time-strt_time))
```

```
## [1] "calculation time is:1.03243321577708"
```

```
conf_lmtxf1 <- predict(fit_lmtxf1$model, interval = "confidence", level=0.95)
pre_lmtxf1 <- predict(fit_lmtxf1$model, interval = "prediction", level=0.95)
```

- 1因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 `err_lmtxf` を算出する。

```
un_2011 <- 2011-1947+1
bet_2011 <- nrow(mtx_under100[mtx_under100$Year=="2011",])

obs_lmtxf <- lmtxf_nonNA

pred_lmtxf1 <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxf1[nm] <- data.frame(pre_lmtxf1[nm])[un_2011,1]
}

err_lmtxf <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxf[num_vec] <- pred_lmtxf1[num_vec] - obs_lmtxf[num_vec]
}
```

- 女性の2因子モデル

```
# 動的因子モデル (1 因子ローカルレベルモデル_女性死亡率対数)
mod_lmtxf_2facs <- SSMModel(lmtxf~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                           +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                           +SSMtrend(1, Q=0, "distinct"))

param2 <- 2*num_params

updatefn4 <- function(pars, mod_lmtxf_2facs) {
  mod_lmtxf_2facs <- SSMModel(lmtxf~SSMtrend(1, Q=1, "common", state_names="factor1_level")
                              +SSMtrend(1, Q=1, "common", state_names="factor2_level")
                              +SSMtrend(1, Q=0, "distinct"))

  mod_lmtxf_2facs$Z[, "factor1_level", ] <- pars[1:41]
  mod_lmtxf_2facs$Z[, "factor2_level", ] <- pars[42:82]
  diag(mod_lmtxf_2facs$H[, , ]) <- exp(pars[83])
  mod_lmtxf_2facs
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxf_2facs <- fitSSM(mod_lmtxf_2facs, c(fit_lmtxf1$opt$par[1:41], -20:20/200, -6), updatefn4, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_time-strt_time))
```

```
## [1] "calculation time is:39.0279831886292"
```

```
conf_lmtxf_2facs <- predict(fit_lmtxf_2facs$model, interval = "confidence", level=0.95)
pre_lmtxf_2facs <- predict(fit_lmtxf_2facs$model, interval = "prediction", level=0.95)
```

- 2因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 `err_lmtxf_2facs` を算出する。

```
obs_lmtxf_2facs <- lmtxf_nonNA

pred_lmtxf_2facs <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxf_2facs[nm] <- data.frame(pre_lmtxf_2facs[nm])[un_2011, 1]
}

err_lmtxf_2facs <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxf_2facs[num_vec] <- pred_lmtxf_2facs[num_vec] - obs_lmtxf_2facs[num_vec]
}
```

- 男性の1因子モデル

```
# 動的因子モデル (1 因子ローカルレベルモデル_男性死亡率対数)
mod_lmtxm1 <- SSMModel(lmtxm~SSMtrend(1, Q=1, "common")+SSMtrend(1, Q=0, "distinct"))

updatefn4 <- function(pars, mod_lmtxm1) {
  mod_lmtxm1 <- SSMModel(lmtxm~SSMtrend(1, Q=1, "common")+SSMtrend(1, Q=0, "distinct"))
  mod_lmtxm1$Z[, 1, ] <- pars[1:num_params]#length(unique(mtx$Age))
  diag(mod_lmtxm1$H[, ,]) <- exp(pars[num_params+1])#length(unique(mtx$Age))+1])
  mod_lmtxm1
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxm1 <- fitSSM(mod_lmtxm1, c(rep(0.1, num_params), -5), updatefn4, method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_time-strt_time))
```

```
## [1] "calculation time is:51.1458721160889"
```

```
conf_mtxm1 <- predict(fit_lmtxm1$model, interval = "confidence", level=0.95)
pre_mtxm1 <- predict(fit_lmtxm1$model, interval = "prediction", level=0.95)
```

- 1因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 `err_lmtxm` を算出する。

```
obs_lmtxm <- lmtxm_nonNA

pred_lmtxm1 <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxm1[nm] <- data.frame(pre_mtxm1[nm])[un_2011, 1]
}

err_lmtxm <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxm[num_vec] <- pred_lmtxm1[num_vec] - obs_lmtxm[num_vec]
}
```

- 男性の2因子モデル

```
# 動的因子モデル（1因子ローカルレベルモデル_女性死亡率対数）
mod_lmtxm_2facs <- SSMModel(lmtxm~SSMtrend(1,Q=1,"common",state_names="factor1_level")
                           +SSMtrend(1,Q=1,"common",state_names="factor2_level")
                           +SSMtrend(1,Q=0,"distinct"))

param2 <- 2*num_params

updatefn5 <- function(pars,mod_lmtxm_2facs){
  mod_lmtxm_2facs <- SSMModel(lmtxm~SSMtrend(1,Q=1,"common",state_names="factor1_level")
                              +SSMtrend(1,Q=1,"common",state_names="factor2_level")
                              +SSMtrend(1,Q=0,"distinct"))
  mod_lmtxm_2facs$Z[, "factor1_level",] <- pars[1:41]
  mod_lmtxm_2facs$Z[, "factor2_level",] <- pars[42:82]
  diag(mod_lmtxm_2facs$H[, ,]) <- exp(pars[83])
  mod_lmtxm_2facs
}

# 対数尤度最大化による未知パラメータの推定
strt_time <- Sys.time()
fit_lmtxm_2facs <- fitSSM(mod_lmtxm_2facs,c(fit_lmtxm1$opt$par[1:41],-20:20/200,-6),updatefn5,method="BFGS")
end_time <- Sys.time()
print(paste0("calculation time is:", end_time-strt_time))
```

```
## [1] "calculation time is:1.0575065334638"
```

```
conf_lmtxm_2facs <- predict(fit_lmtxm_2facs$model, interval = "confidence", level=0.95)
pre_lmtxm_2facs <- predict(fit_lmtxm_2facs$model, interval = "prediction", level=0.95)
```

- 2因子モデルで得られた2011年の予測値と、2011年の観測値との誤差 `err_lmtxm_2facs` を算出する。

```
obs_lmtxm_2facs <- lmtxm_nonNA

pred_lmtxm_2facs <- c(rep(0, bet_2011))
for (nm in 1:bet_2011) {
  pred_lmtxm_2facs[nm] <- data.frame(pre_lmtxm_2facs[nm])[un_2011,1]
}

err_lmtxm_2facs <- c(rep(0, bet_2011))
for (num_vec in 1:bet_2011) {
  err_lmtxm_2facs[num_vec] <- pred_lmtxm_2facs[num_vec] - obs_lmtxm_2facs[num_vec]
}
```

## 解析結果の解釈

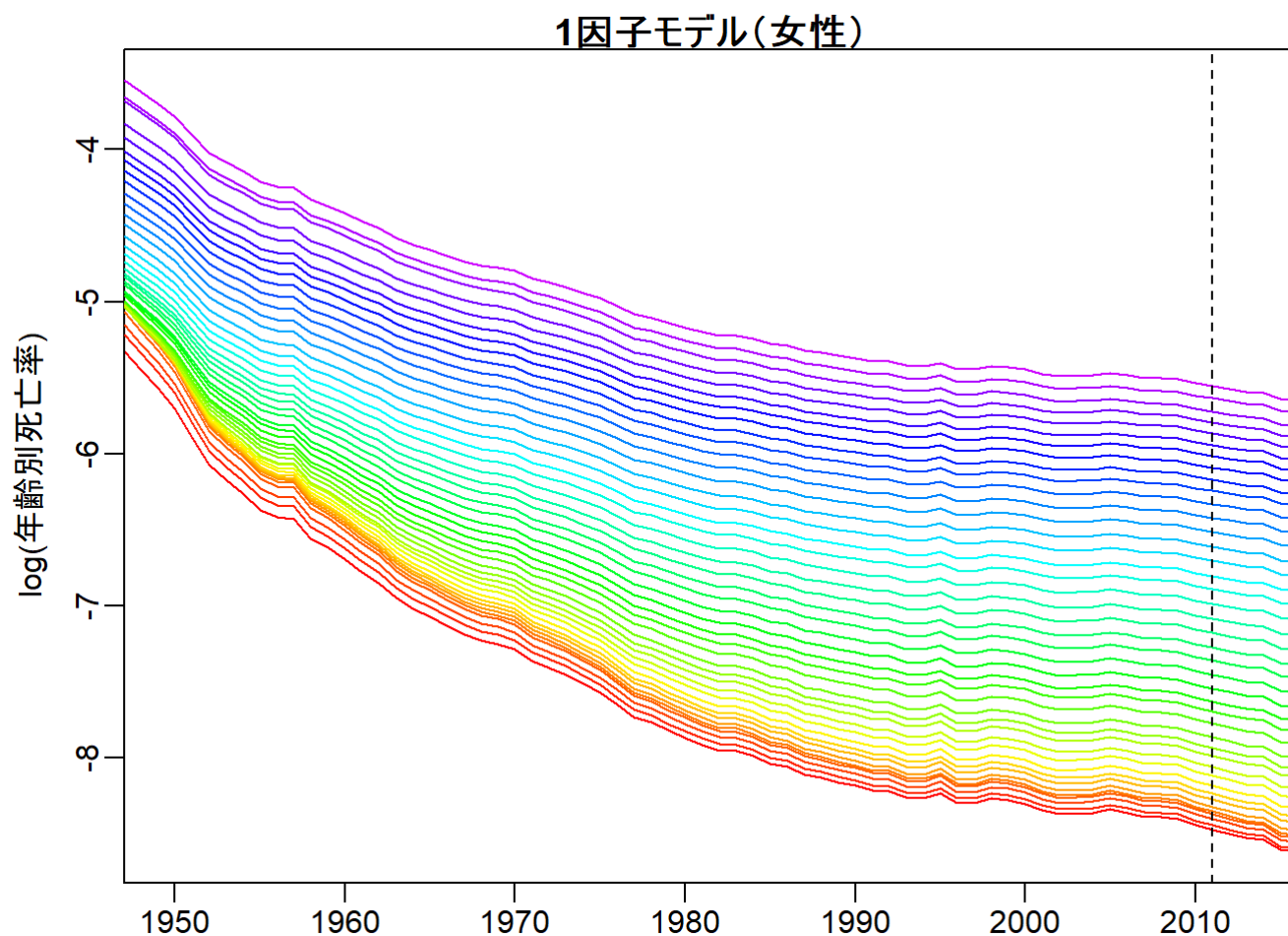
- 上記の解析で得た結果をもとに、2011年の20~60歳男女の年齢別死亡率について、予測値と観測値を比べてみる。
- まず、女性・男性それぞれにおけるモデル（1因子・2因子）の予測値を可視化する。

```

kfs_lm1xf <- KFS(fit_lm1xf1$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lm1xf$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log(年齢別死亡率)", main="1因子モデル (女性)")
abline(v=2011, lty=2)

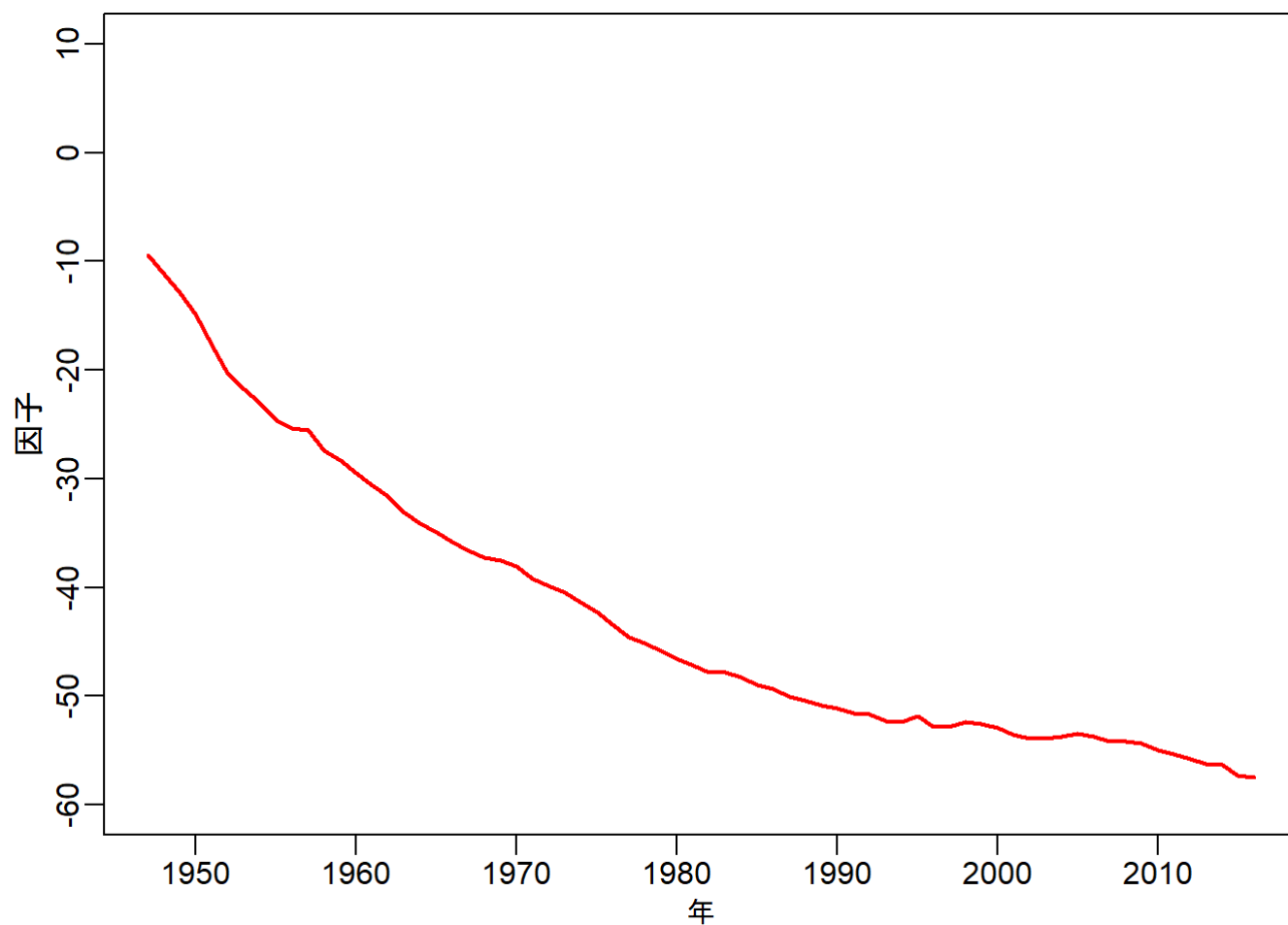
```



```

par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lm1xf$alphahat[, 1], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-60, 10))

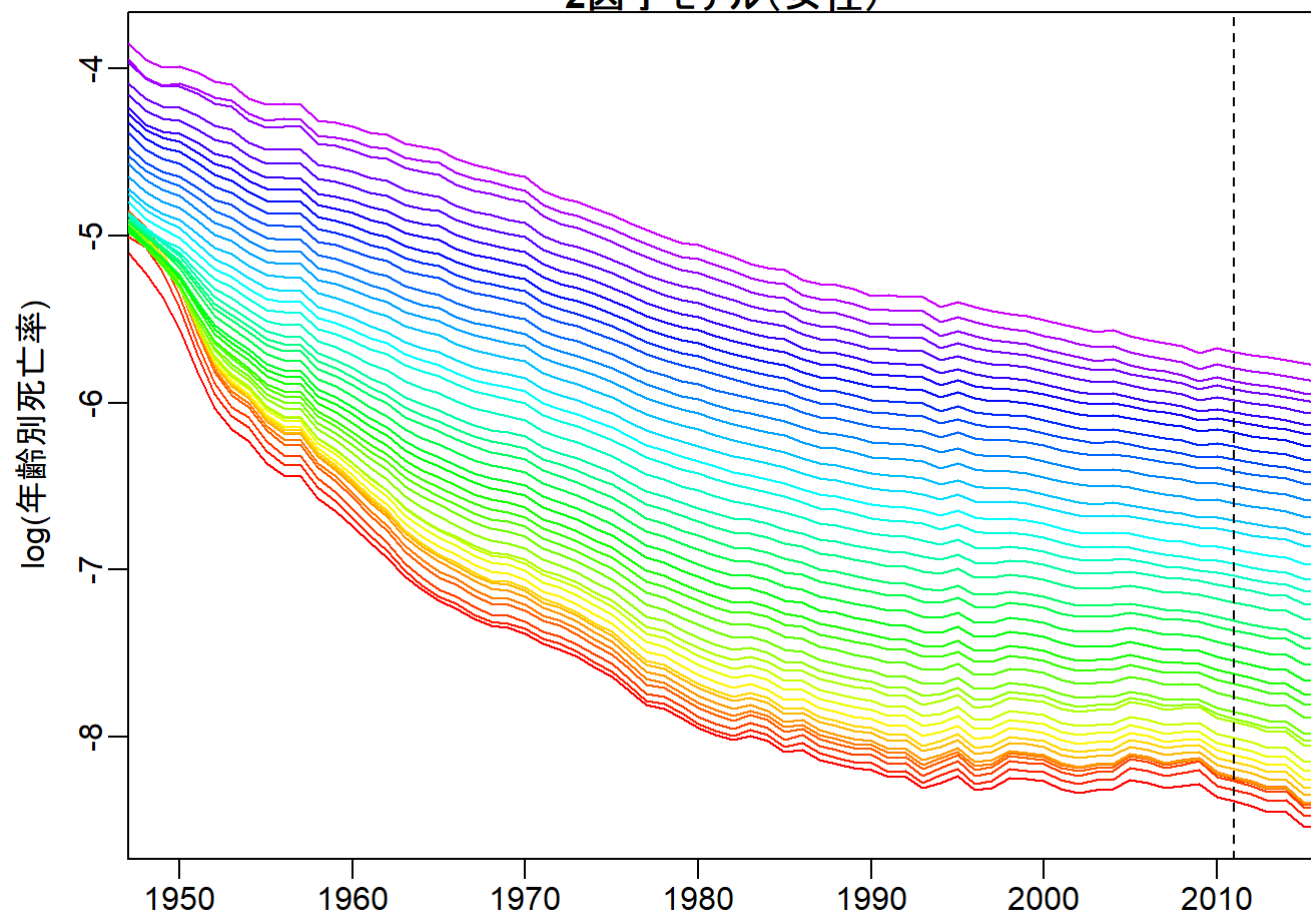
```



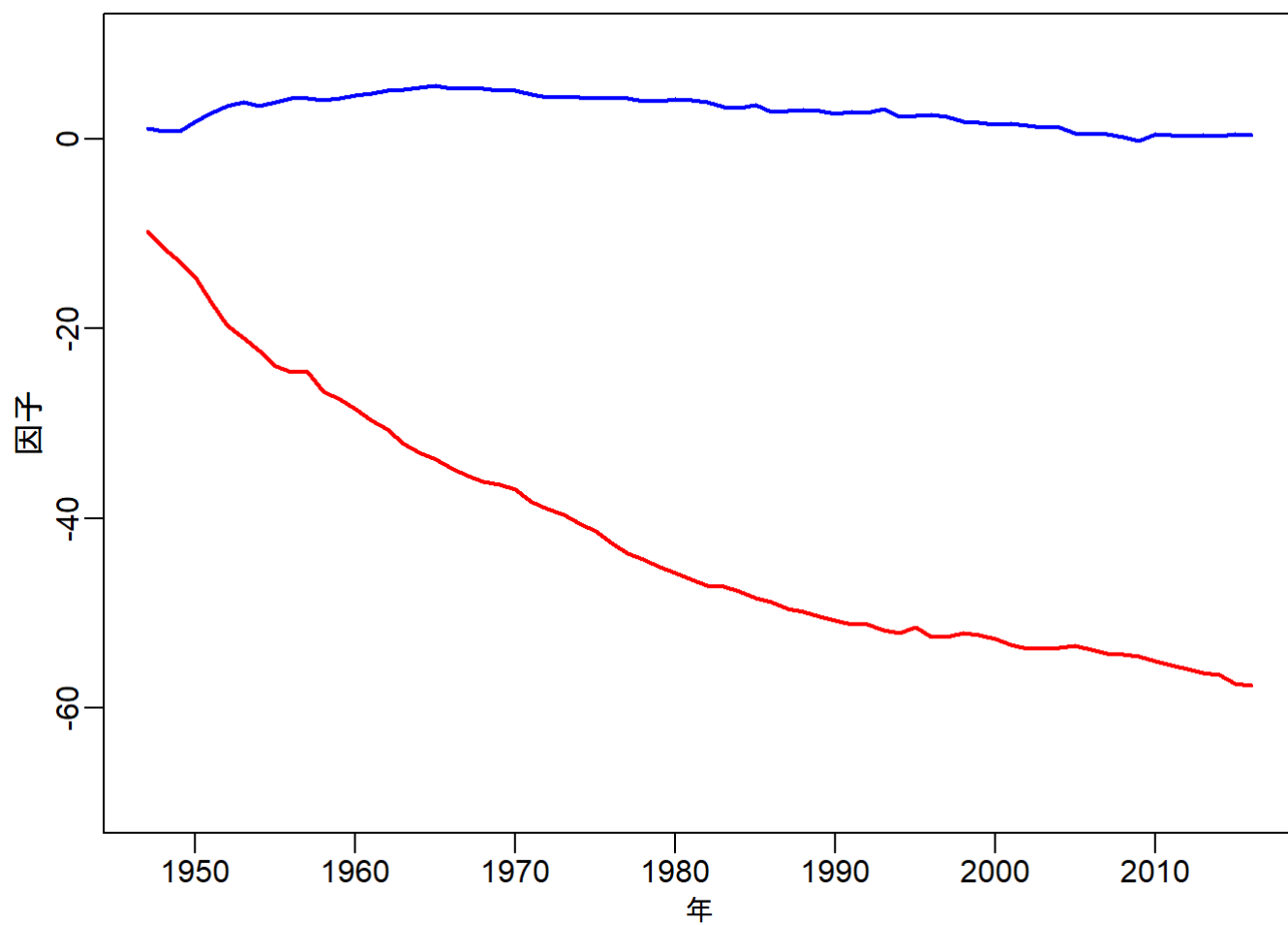
```
kfs_lmxf_2facs <- KFS(fit_lmxf_2facs$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmxf_2facs$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log
(年齢別死亡率)", main="2因子モデル (女性)")
abline(v=2011, lty=2)
```

## 2因子モデル(女性)

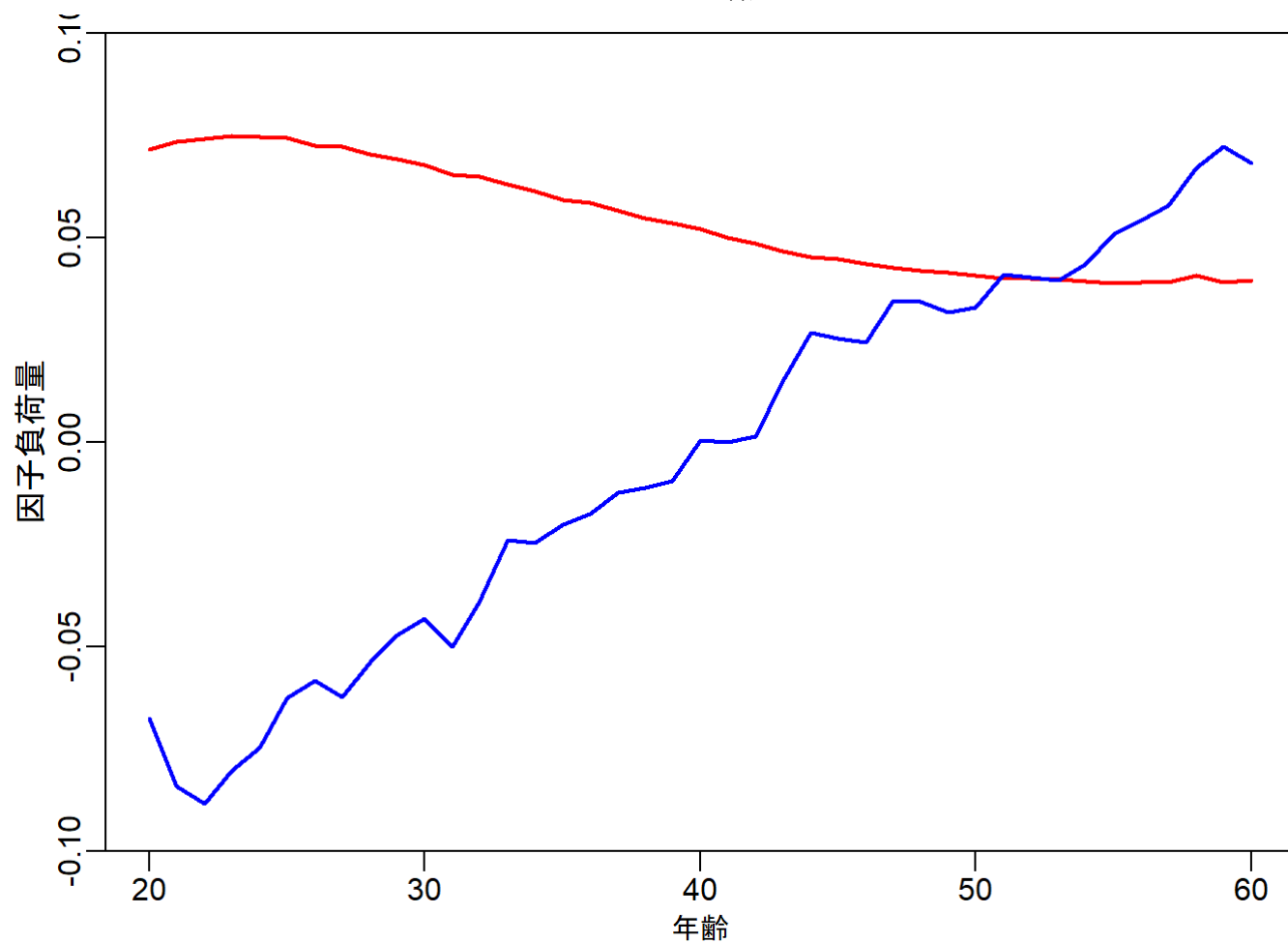


```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxf_2facs$alphahat[, "factor1_level"], type="l", col=2, lwd=2, xlab="年", ylab="
因子", ylim=c(-70, 10))
lines(1947:2016, kfs_lmtxf_2facs$alphahat[, "factor2_level"], type="l", col=4, lwd=2)
```



```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(20:60, fit_lmtxf_2facs$opt$par[1:41], type="l", col=2, lwd=2, xlab="年齢", ylab="因子負荷量", ylim=c(-0.1, 0.1), yaxs="i")
lines(20:60, fit_lmtxf_2facs$opt$par[42:82], col=4, lwd=2)
```





```
kfs_lmtxm <- KFS(fit_lmtxm1$mod)
```

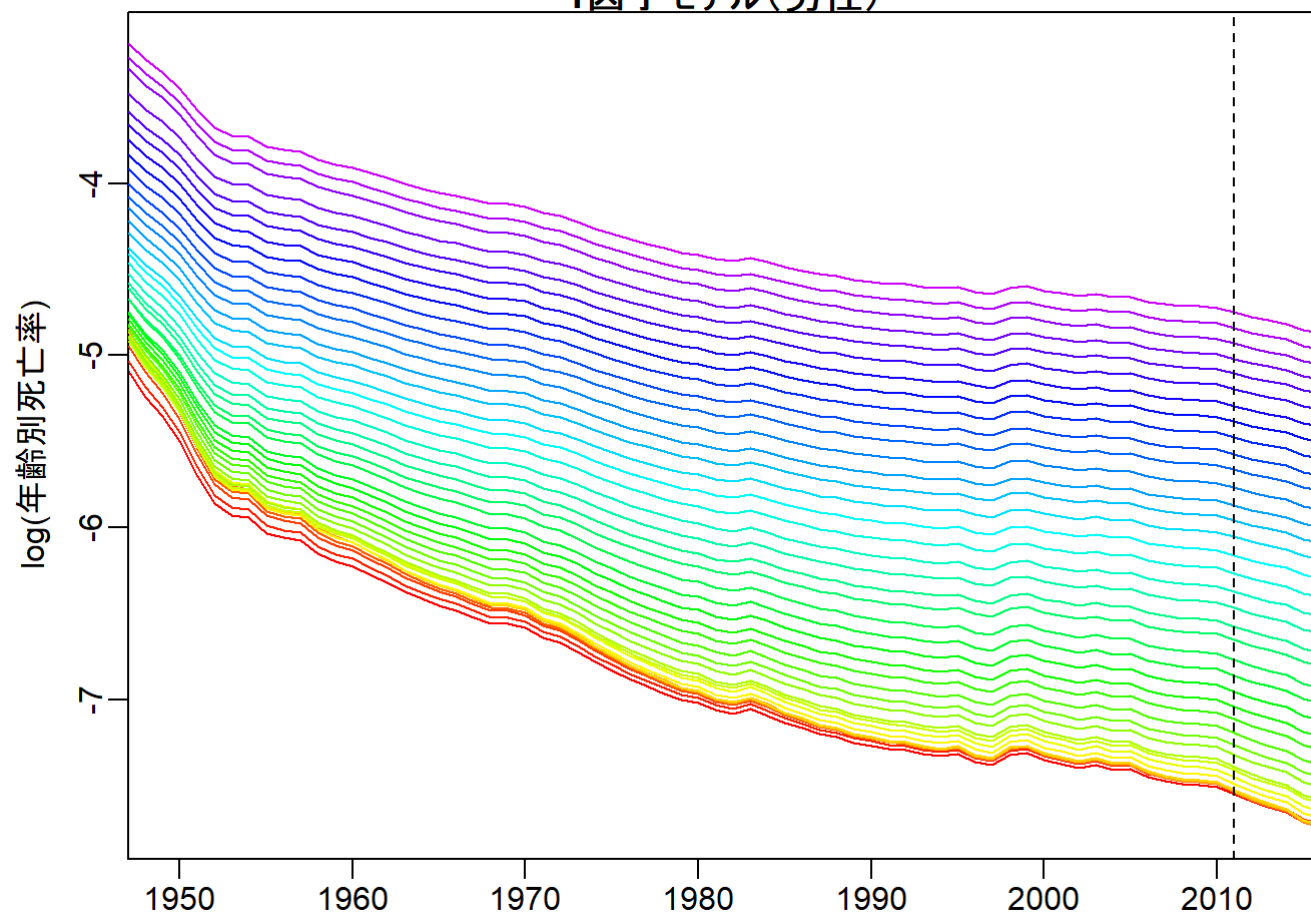
```
par(mar=c(2, 3, 1, 1))
```

```
par(mgp=c(2, .5, 0))
```

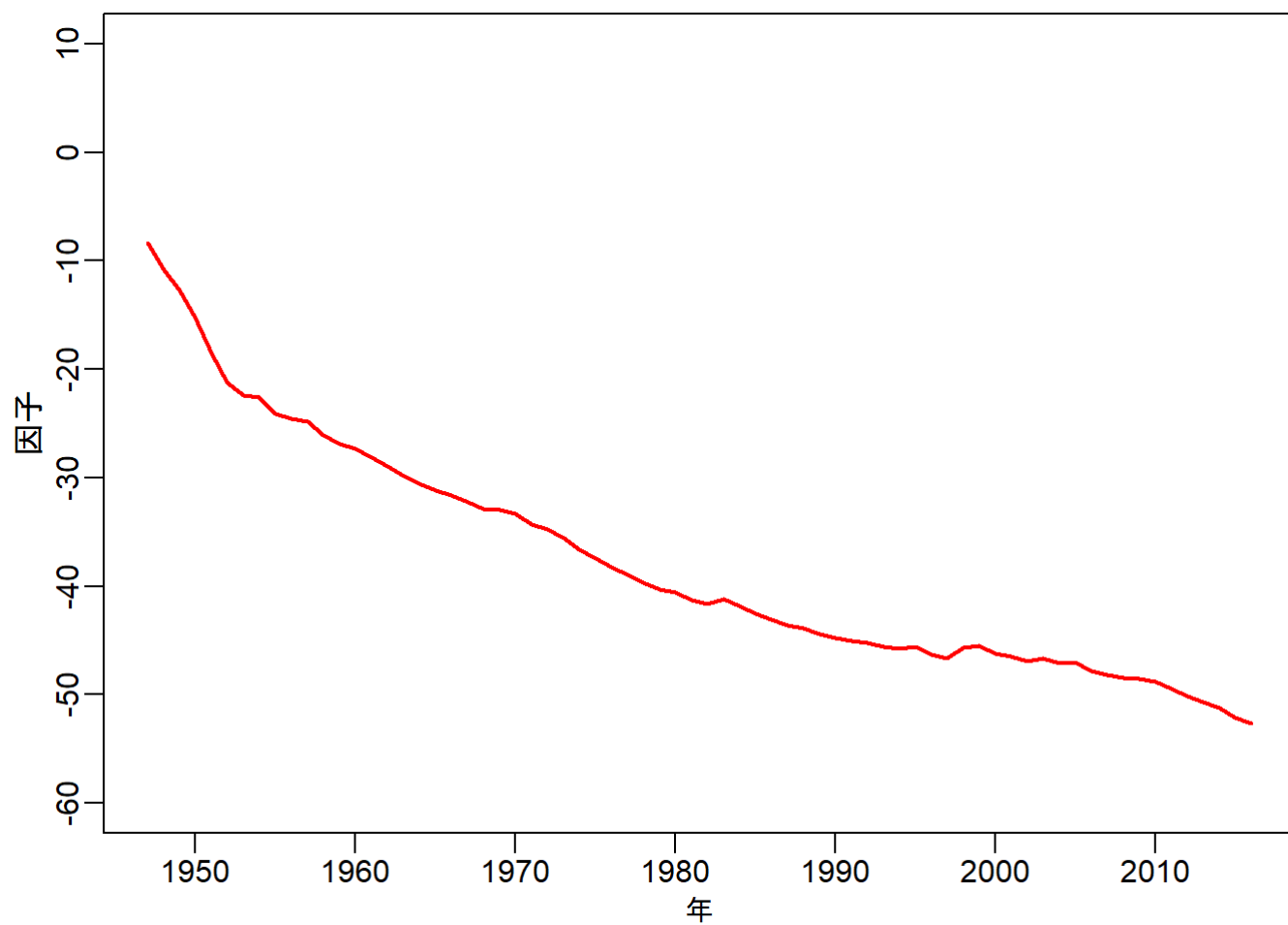
```
matplot(1947:2016, kfs_lmtxm$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log(年齢別死亡率)", main="1因子モデル (男性)")
```

```
abline(v=2011, lty=2)
```

## 1因子モデル(男性)



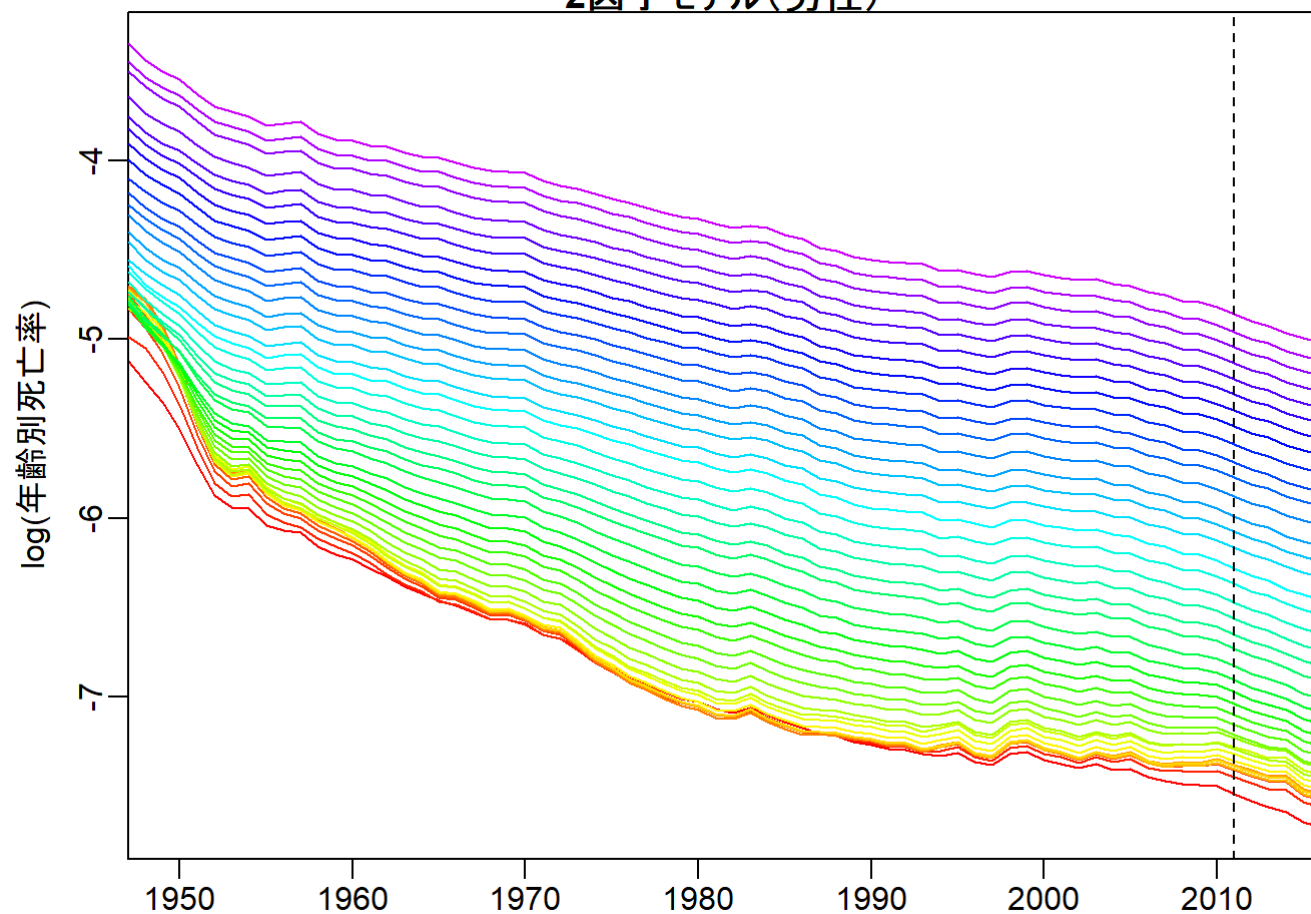
```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxm$alphahat[, 1], type="l", col=2, lwd=2, xlab="年", ylab="因子", ylim=c(-60, 10))
```



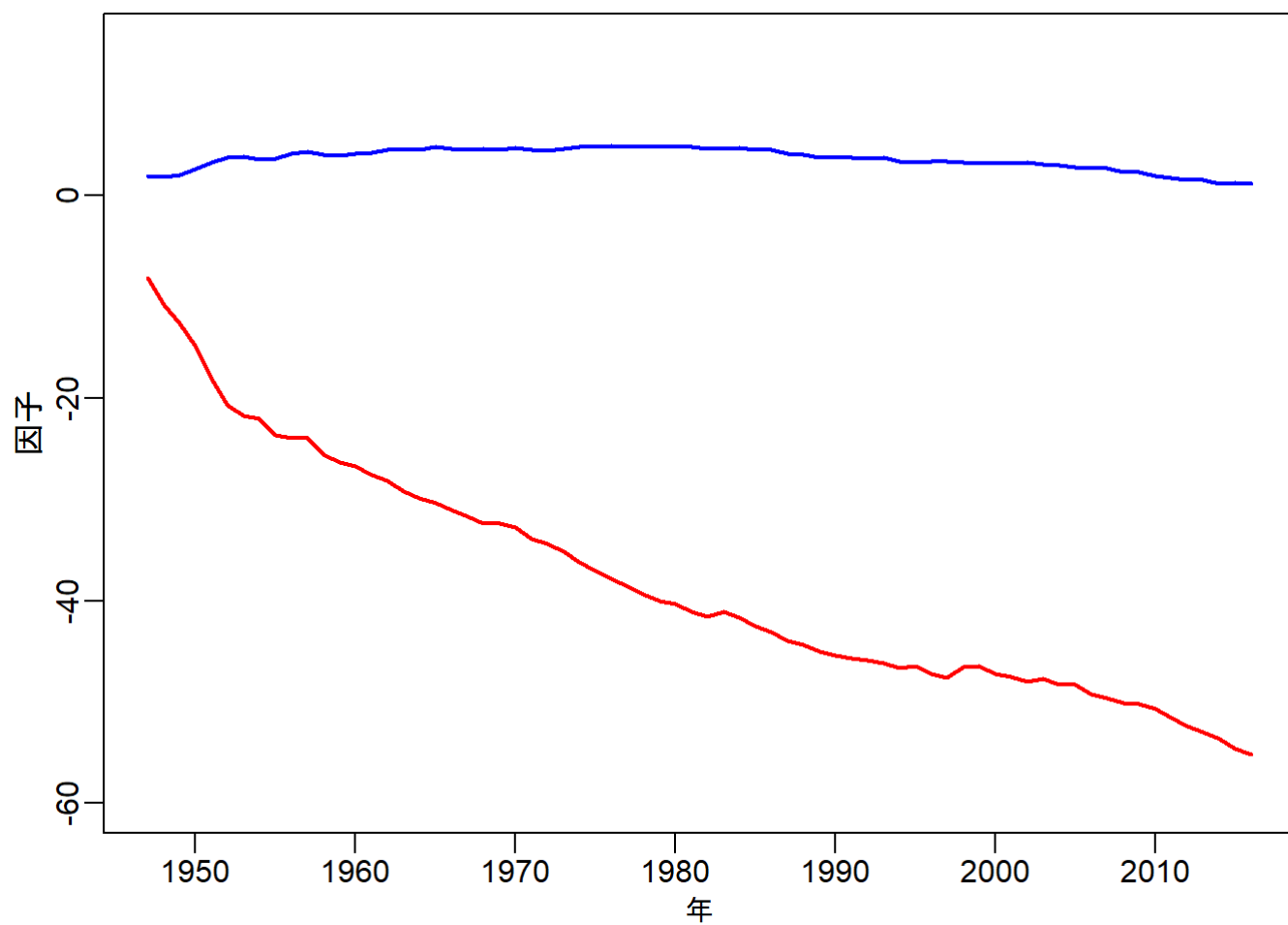
```
kfs_lmtxm_2facs <- KFS(fit_lmtxm_2facs$mod)

par(mar=c(2, 3, 1, 1))
par(mgp=c(2, .5, 0))
matplot(1947:2016, kfs_lmtxm_2facs$muhat, type="l", lty=1, col=rainbow(50)[1:41], xaxs="i", ylab="log
(年齢別死亡率)", main="2因子モデル (男性)")
abline(v=2011, lty=2)
```

## 2因子モデル(男性)



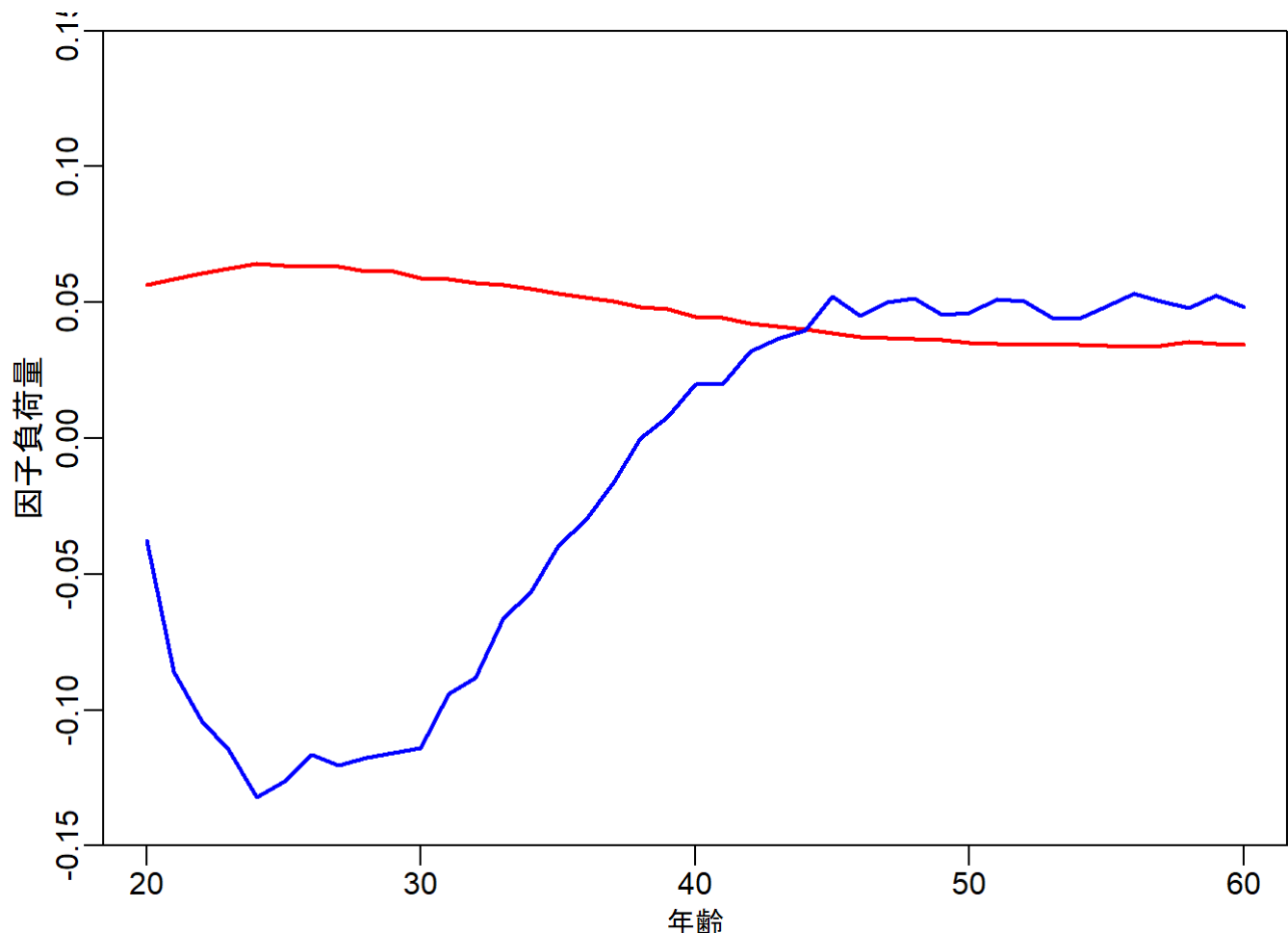
```
par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(1947:2016, kfs_lmtxm_2facs$alphahat[, "factor1_level"], type="l", col=2, lwd=2, xlab="年", ylab="
因子", ylim=c(-60, 15))
lines(1947:2016, kfs_lmtxm_2facs$alphahat[, "factor2_level"], type="l", col=4, lwd=2)
```



```

par(mar=c(2.5, 2.5, .5, .5))
par(mgp=c(1.5, .5, 0))
plot(20:60, fit_lmtxm_2facs$opt$par[1:41], type="l", col=2, lwd=2, xlab="年齢", ylab="因子負荷量", ylim=c(-0.15, 0.15), yaxs="i")
lines(20:60, fit_lmtxm_2facs$opt$par[42:82], col=4, lwd=2)

```

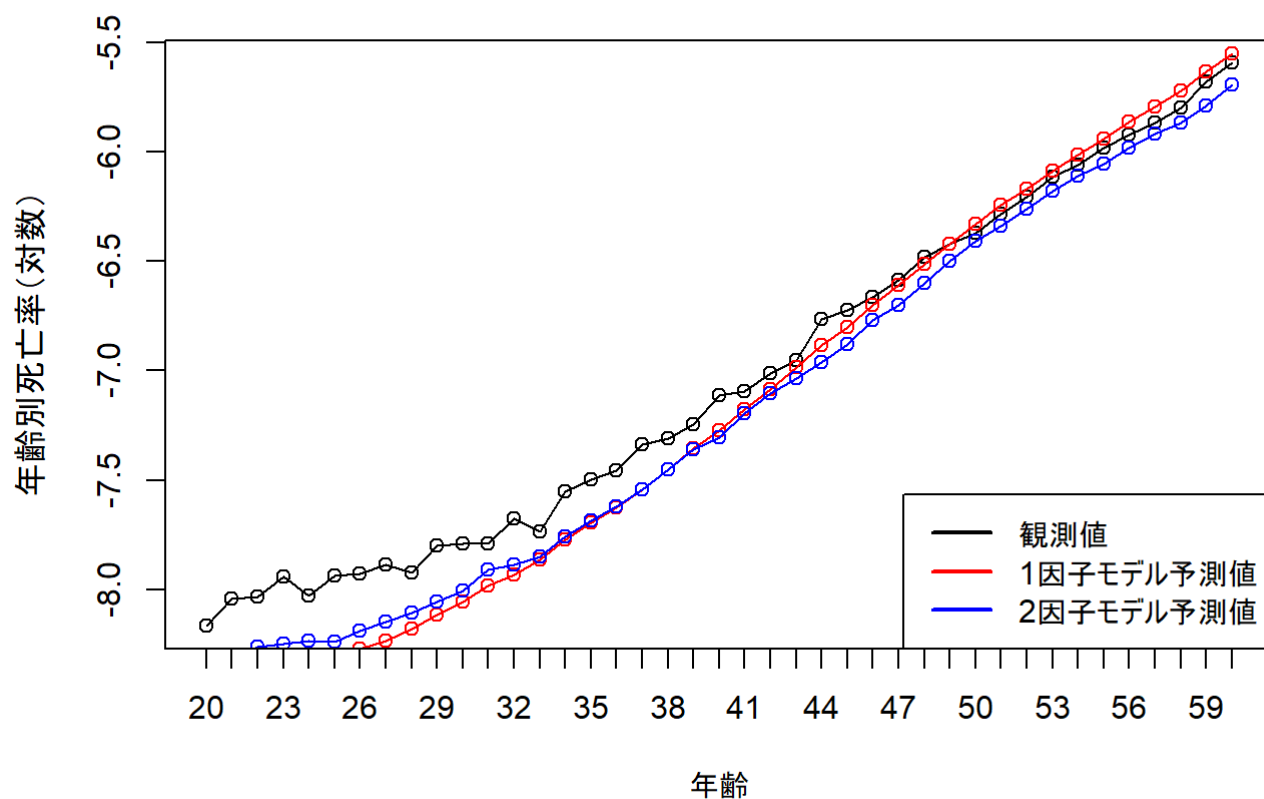


男性・2因子モデルの1950年前後における若い世代（20代前半）の死亡率が上がり、逆転現象が起きているが、これは第二次世界大戦による徴兵の影響を表現していると考えられる。- ちなみに、第1因子は時代、第2因子は年齢に対する因子を表している。

- 続いて、予測対象であった2011年の死亡率について、観測値とそれぞれのモデルでの予測値とを可視化し比較する。

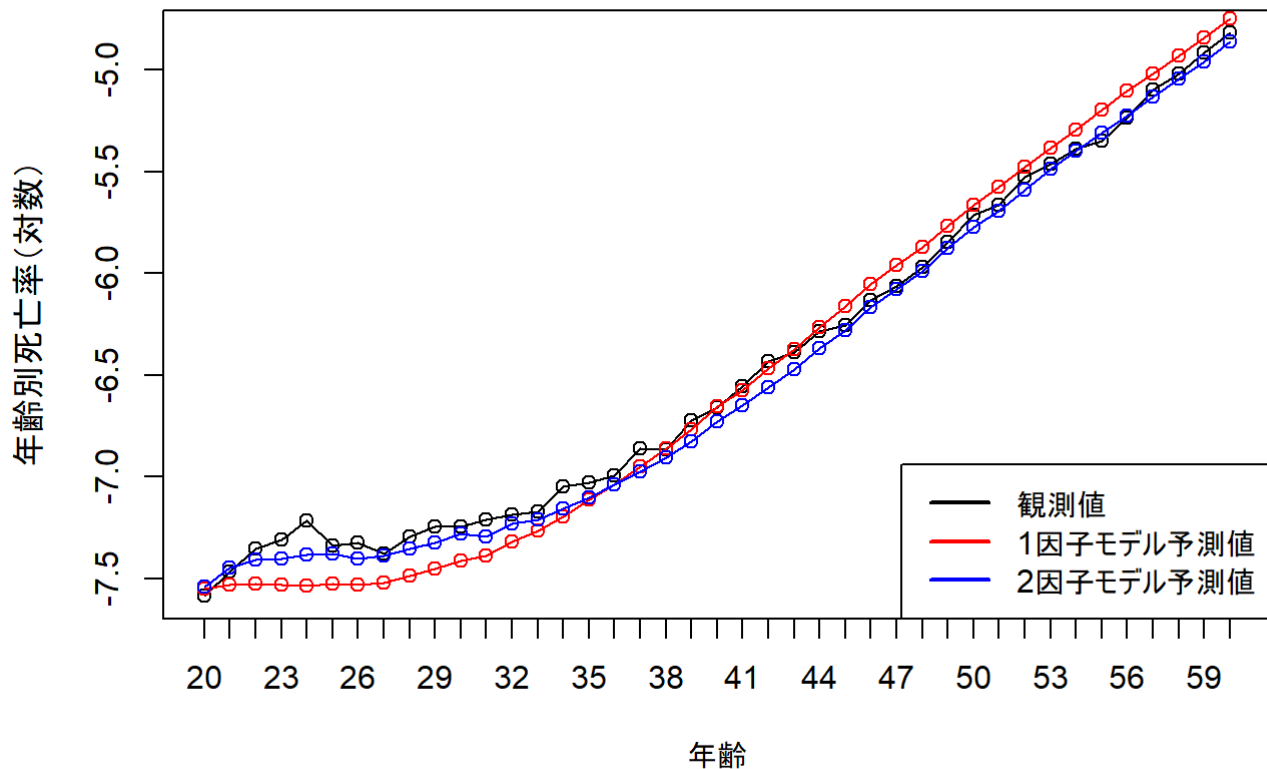
```
plot(obs_lmtxf, type="o", main="2011年の女性20-60歳死亡率（対数変換あり）", ylab="年齢別死亡率（対数）", xlab="年齢", xaxt="n")
lines(pred_lmtxf1, col="red", type="o")
lines(pred_lmtxf_2facs, col="blue", type="o")
legend("bottomright", c("観測値", "1因子モデル予測値", "2因子モデル予測値"), lwd=2, col=c(1, 2, 4))
axis(side=1, at=1:41, labels=c(thr))
```

## 2011年の女性20-60歳死亡率(対数変換あり)



```
plot(obs_lmtxm, type="o", main="2011年の男性20-60歳死亡率(対数変換あり)", ylab="年齢別死亡率(対数)", xlab="年齢", xaxt="n")
lines(pred_lmtxm1, col="red", type="o")
lines(pred_lmtxm_2facs, col="blue", type="o")
legend("bottomright", c("観測値", "1因子モデル予測値", "2因子モデル予測値"), lwd=2, col=c(1, 2, 4))
axis(side=1, at=1:41, labels=c(thr))
```

## 2011年の男性20-60歳死亡率(対数変換あり)



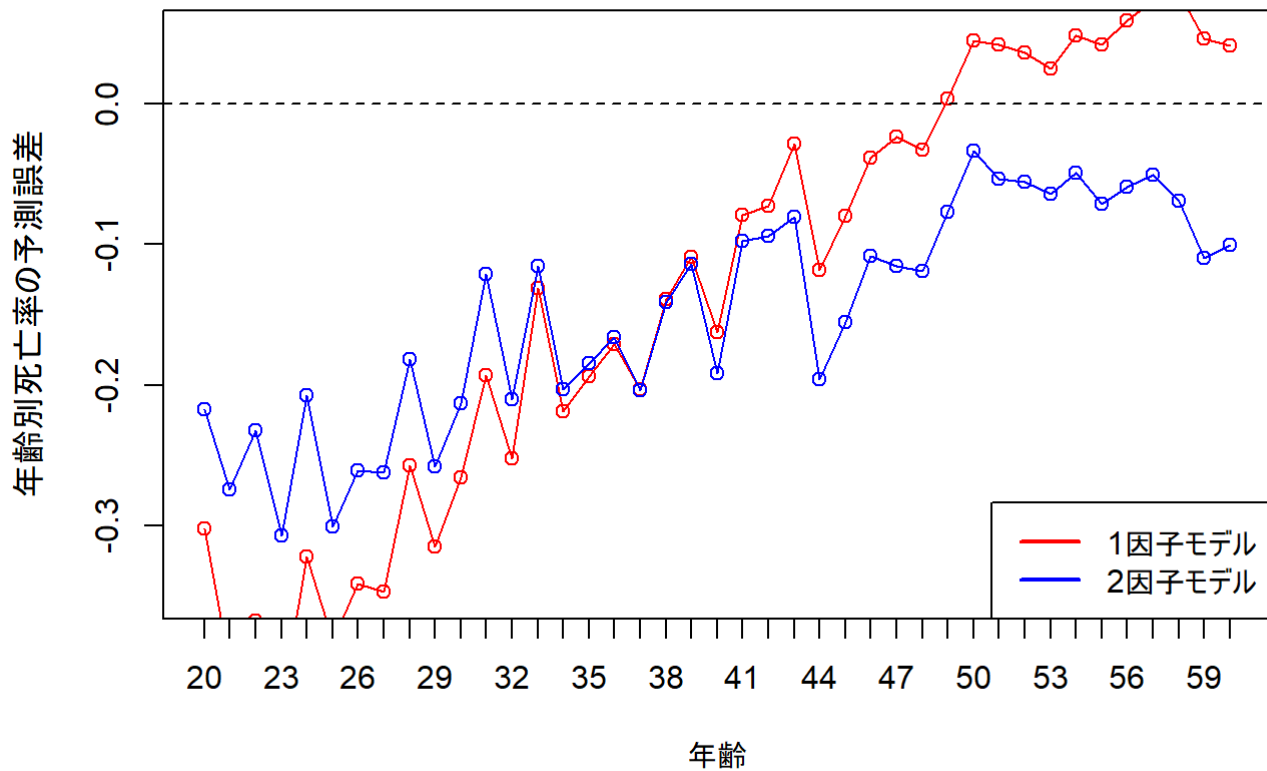
両者とも、若い世代では1因子モデルの方が若干観測値と近い値を予測できているように見える一方、年配の世代は2因子モデルの方が観測値と近い値を予測しているように見える。

- 次に、先ほど算出した、観測値と予測値の誤差について可視化する。

```
plot(err_lmtxf, type="o", main="2011年20-60歳死亡率予測誤差（女性）", ylab="年齢別死亡率の予測誤差", xlab="年齢", xaxt="n", col="red", ylim=c(-0.35, 0.05))
lines(err_lmtxf_2facs, col="blue", type="o")
legend("bottomright", c("1因子モデル", "2因子モデル"), lwd=2, col=c(2, 4))
axis(side=1, at=1:41, labels=c(thr))
abline(h=0, lty=2)
```

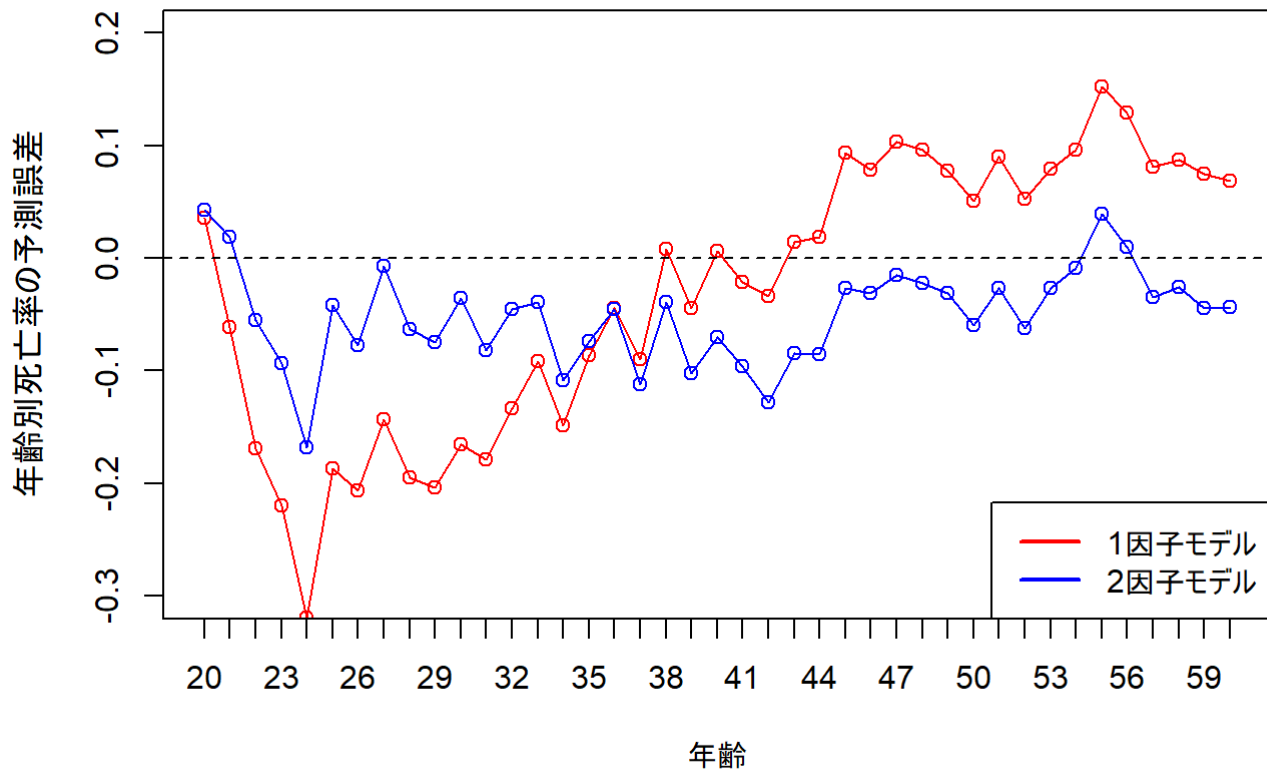


## 2011年20-60歳死亡率予測誤差(女性)



```
plot(err_lmtxm, type="o", main="2011年20-60歳死亡率予測誤差 (男性)", ylab="年齢別死亡率の予測誤差", xlab="年齢", xaxt="n", col="red", ylim=c(-0.3, 0.2))
lines(err_lmtxm_2facs, col="blue", type="o")
legend("bottomright", c("1因子モデル", "2因子モデル"), lwd=2, col=c(2, 4))
axis(side=1, at=1:41, labels=c(thr))
abline(h=0, lty=2)
```

## 2011年20-60歳死亡率予測誤差(男性)



両者とも、若い世代では1因子モデルの方が若干予測誤差が小さい一方、年配の世代は2因子モデルの方が若干予測誤差が小さい。 - これは、20代の若い世代間での死亡率にあまり差がなく、戦争や災害といった時代のイベントによる変動の影響を受けやすい一方、高齢になってくると、そういったイベントだけでなく、年齢による死亡率の変動も比較的大きくなり、年齢の影響を受けやすくなるためと考えられる。