

## Survival Models: Week 11

## Graduation: ALT 2005-07

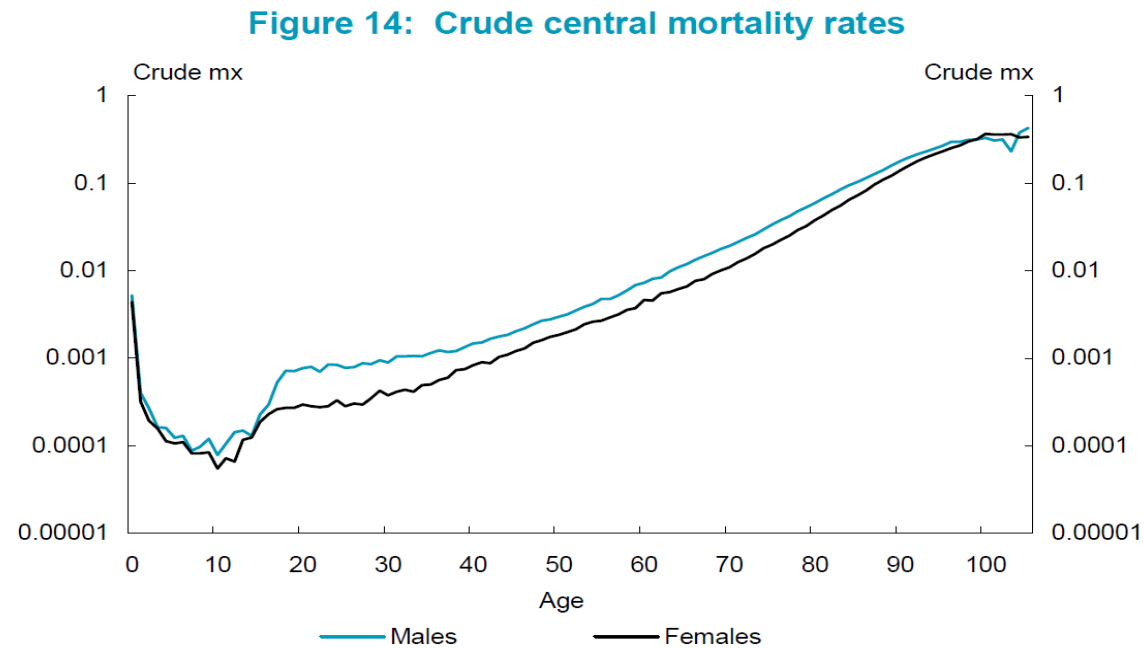
Please see pages 26-33 of ALT 2005-07 for further details.

- The first step in producing ALT 2005-07 was the calculation of exposed to risk and crude mortality rates. We will discuss this issue next week.
- The second step is the smoothing (or graduation) of the crude mortality rates.

“From a first principles perspective, however, there is no reason to suppose that these fluctuations are anything other than a reflection of the random nature of the underlying mortality distribution. Hence, when constructing a life table to represent the mortality experience of a population, it is customary to graduate the crude rates to obtain a curve that progresses smoothly with age.” (ALT 2005-07).

## Graduation: ALT 2005-07

Figure of crude rates taken from ALT 2005-07 page 31.



100 -  $\infty$

## Graduation: ALT 2005-07

Age 1-2  
↓

*main part*

- Manual graduation and cubic splines used to graduate the crude rates. We will only discuss the cubic spline graduation.
- Manual graduation was used at the extremes of age.
- For cubic splines the number and location of the knots needs to be determined.
- Tests were conducted to assess the appropriateness of the graduated rates.

## Graduation: ALT 2005-07

The knots were selected to minimize the following quantity:

$$\frac{\overset{\text{obs - expected}}{(A_x - E_x^c m_x)^2}}{\underset{\text{var of expected}}{E_x^c m'_x (1 - m'_x)}},$$

*don't use  $m_x$  in the bottom, or too complicated.*

where  $A_x$  is the observed number of deaths aged  $x$ ,  $E_x^c$  is the central exposed to risk aged  $x$ ,  $m_x$  is the graduated value of the central mortality rate at age  $x$ , produced by the cubic spline based on the particular knot selection, and  $m'_x$  is a preliminary value of  $m_x$ .

## Graduation: ALT 2005-07

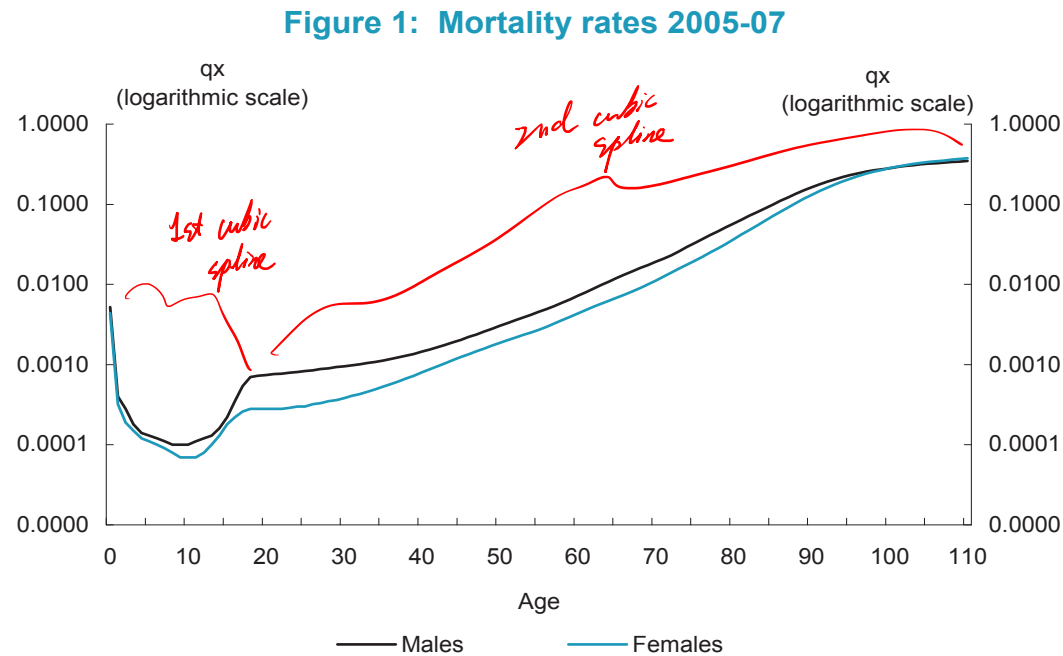
- The initial knot placement was based on a visual inspection of the data.
- More knots were placed where the crude rates were changing the fastest.
- The final knot locations were:

Males: 4 16 19 21 22 32 57 60 73 75 78 89 94

Females: 4 11 17 20 22 35 36 57 58 65 78 91 105

## Graduation: ALT 2005-07

Figure of graduated rates taken from ALT 2005-07 page 3.

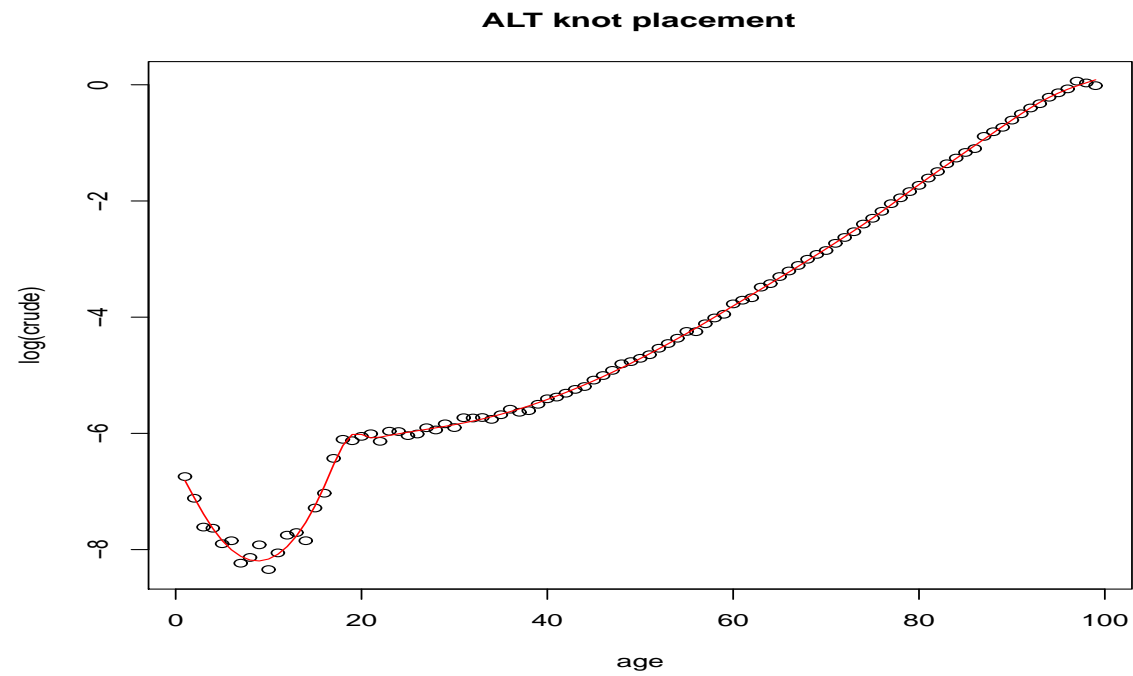


## Graduation ALT 2005-07 in R

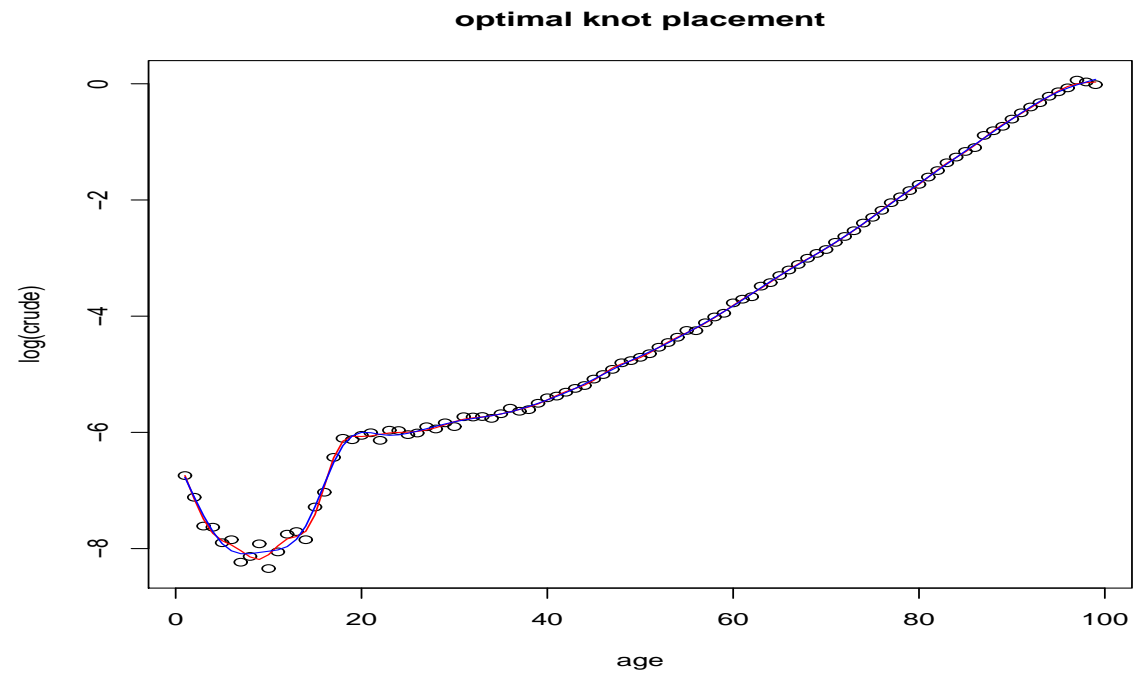
- Extracted the raw data for Males from Appendix C of ALT 2005-07
- Smoothed the data using natural cubic splines in R.
- Used specific knot placements and also AIC/BIC to choose an “optimal” number of knots.



## Graduation ALT 2005-07 in R



## Graduation ALT 2005-07 in R



## Graduation ALT 2005-07 in R

```
#  
library(splines)  
data<-read.table("crude.csv",header=T,sep=",")  
colnames(data)<-c("age","exposed","deaths","crude")  
attach(data)  
  
#graduating the number of deaths. need offset to allow  
#for different exposed to risk  
#using knots from ALT 2005-07  
fit<-glm(deaths~ns(age,knots=c(4,16,19,21,22,32,57,60,73,75,78,89,94))  
+offset(log(exposed)),family="poisson")  
values<-age  
temp<-data.frame("age"=values)  
fit<-predict.glm(fit,temp)  
plot(age,log(crude),main="ALT knot placement")  
lines(values,log(exp(fit)/exposed),col="red")  
  
#using AIC and BIC to find "optimal" number of knots  
resAIC<-resBIC<-rep(0,30)  
  
for(i in 1:50) {
```

$$\text{fit} = \log(\hat{\text{death}})$$

$$\log\left(\frac{\exp(\text{fit})}{\text{exposed}}\right) = \log\left(\frac{\text{death}}{\text{exposed}}\right)$$

*# of knots*

```
fit<-glm(deaths~ns(age,df=i)+offset(log(exposed)),family="poisson")
resAIC[i]<-AIC(fit)
resBIC[i]<-AIC(fit,k=log(length(age)))
}
> sort.list(resAIC)[1]
[1] 36
> sort.list(resBIC)[1]
[1] 17
```

```
fit1<-glm(deaths~ns(age,df=36)+offset(log(exposed)),family="poisson")
fit2<-glm(deaths~ns(age,df=17)+offset(log(exposed)),family="poisson")
values<-age
temp<-data.frame("age"=values)
fit1<-predict.glm(fit1,temp)
fit2<-predict.glm(fit2,temp)
plot(age,log(crude),main="optimal knot placement")
lines(values,log(exp(fit1)/exposed),col="red")
lines(values,log(exp(fit2)/exposed),col="blue")
```

## Mortality Improvements

- ALT 2005-07 also provide improvement factors that can be applied to allow for improvements in mortality over time.
- Improvements are incorporated using the following relationship:

$$q_x(t) = q_x \left( 1 + \frac{I_x}{100} \right)^{(t-2006)},$$

where  $I_x$  is the improvement factor and  $q_x(t)$  is the mortality rate  $t$  years into the future.

- See page 35 of ALT 2005-07 for further details.

*Appendix E.*

## Examples: Other Countries

- The R package *MortalitySmooth* contains functions for smoothing mortality data.
- The package also contains selected data from the Human Mortality Database (HMD).
- The data in the HMD contains deaths, exposed to risk, and other information over many years for a number of different countries.
- Data of this form allows us to investigate the mortality experience of cohorts of individuals.
- On the following slides we will show plots of data from Denmark and Japan.

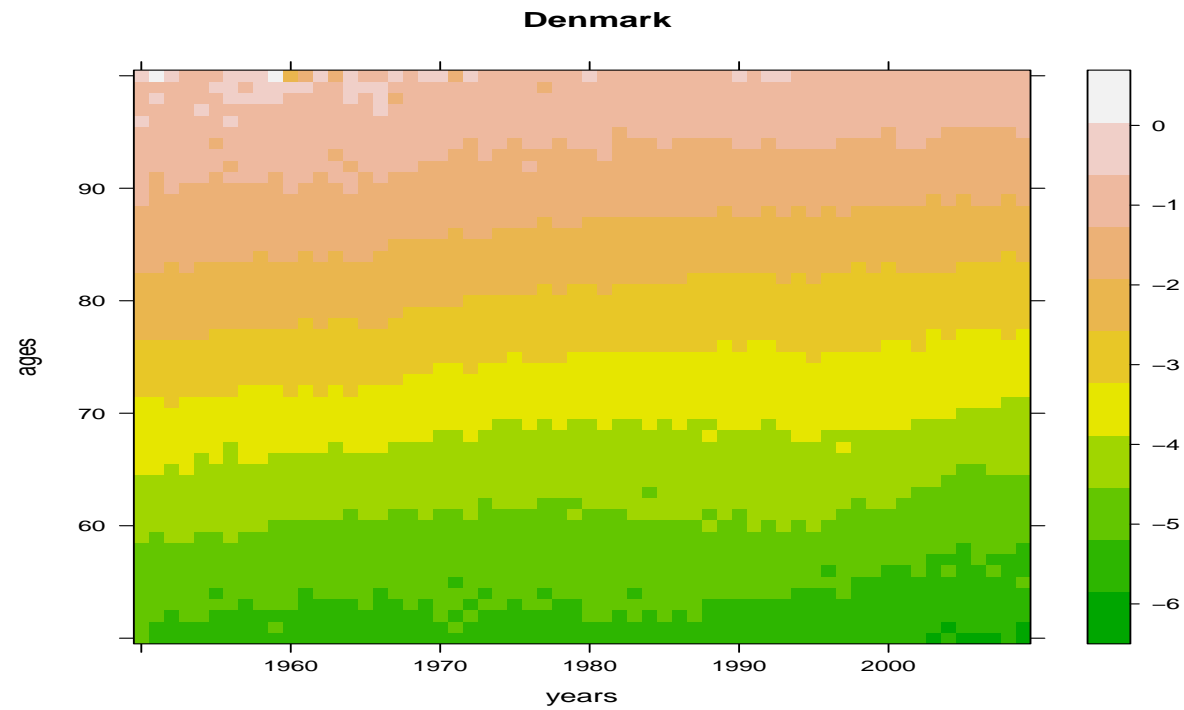
Age, time, country.

## Examples: Other Countries

```
library(MortalitySmooth)
data(HMDdata)
x<-50:100
y<-1950:2009
newdata1<-selectHMDdata(country="Denmark",data="Rates",sex="Females",ages=x,years=y)
newdata2<-selectHMDdata(country="Japan",data="Rates",sex="Females",ages=x,years=y)
plot(newdata1,cut=10,col.regions=terrain.colors(11),main="Denmark")
plot(newdata2,cut=10,col.regions=terrain.colors(11),main="Japan")
> newdata1[1:3,1:3]
```

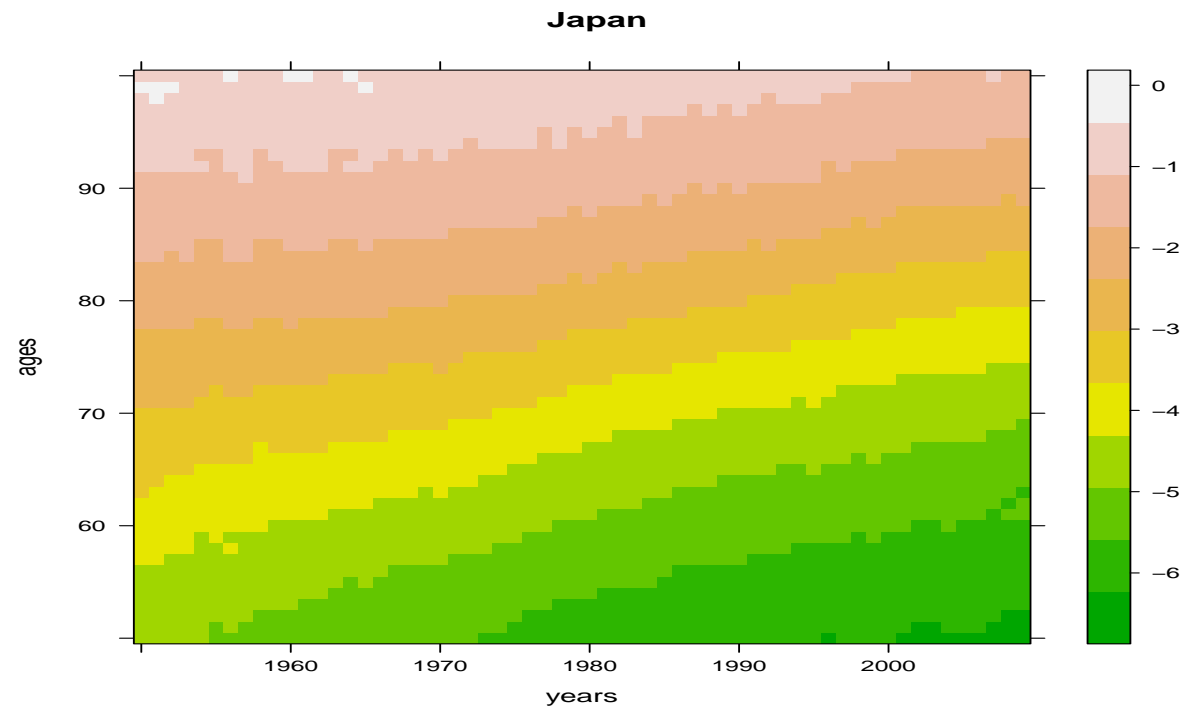
	1950	1951	1952
50	0.005733326	0.005169963	0.005274262
51	0.005738661	0.005425875	0.004788330
52	0.006253357	0.006952491	0.004620687

## Examples: Other Countries





## Examples: Other Countries



## Examples: Denmark 2009 v 1950

#Looking at two different years

```
Exposure<-selectHMDdata(country="Denmark",data="Exposures",sex="Females",ages=1:100,years=c(1950,2009))
```

```
Deaths<-selectHMDdata(country="Denmark",data="Deaths",sex="Females",ages=1:100,years=c(1950,2009))
```

#2009

```
deaths<-round(Deaths[,2])
```

```
exposed<-Exposure[,2]
```

```
age<-1:100
```

```
fit<-glm(deaths~ns(age,df=15)+offset(log(exposed)),family="poisson")
```

```
values<-age
```

```
temp<-data.frame("age"=values)
```

```
fit<-predict.glm(fit,temp)
```

```
plot(age,log(deaths/exposed),main="Denmark 2009 and 1950")
```

```
lines(values,log(exp(fit)/exposed),col="red")
```

#1950

```
deaths<-round(Deaths[,1])
```

```
exposed<-Exposure[,1]
```

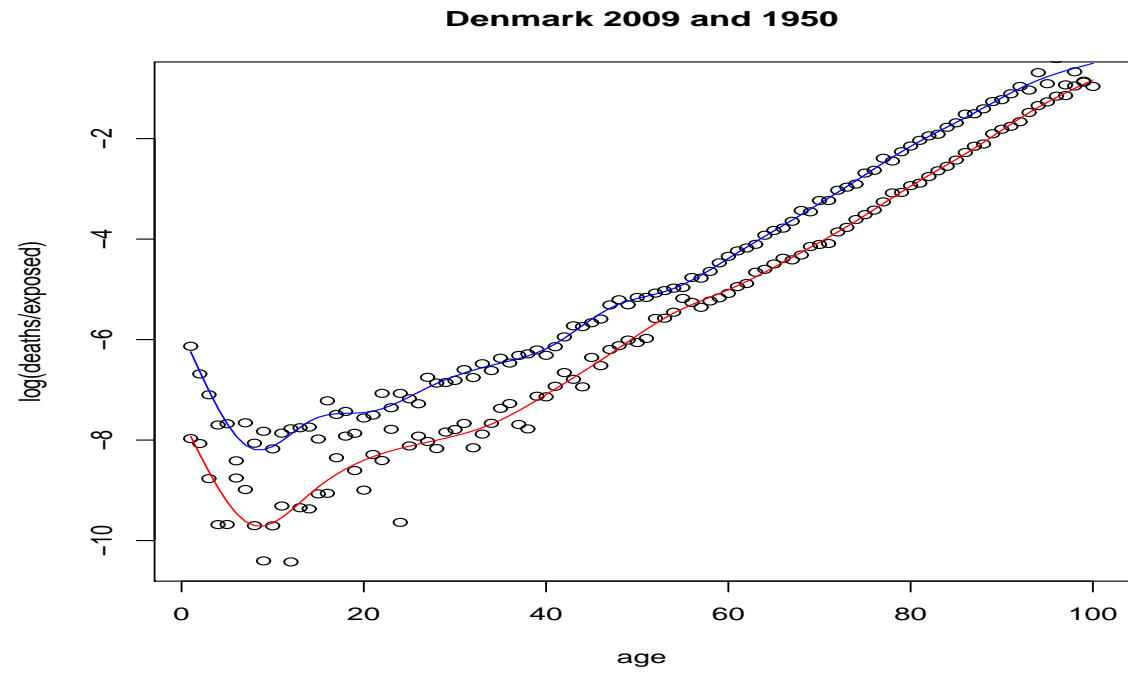
```
age<-1:100
```

```
fit<-glm(deaths~ns(age,df=15)+offset(log(exposed)),family="poisson")
```

```
values<-age
```

```
temp<-data.frame("age"=values)
fit<-predict.glm(fit,temp)
points(age,log(deaths/exposed),main="ALT knot placement")
lines(values,log(exp(fit)/exposed),col="blue")
```

## Examples: Denmark 2009 v 1950



## Examples: Denmark and Japan age 50 1950 to 2009

```
#Japan
```

```
Exposure<-selectHMDdata(country="Japan",data="Exposures",sex="Females",ages=50,years=1950:2009)
```

```
Deaths<-selectHMDdata(country="Japan",data="Deaths",sex="Females",ages=50,years=1950:2009)
```

```
deaths<-round(Deaths)
```

```
exposed<-Exposure
```

```
year<-1950:2009
```

```
fit<-glm(deaths~ns(year,df=5)+offset(log(exposed)),family="poisson")
```

```
values<-year
```

```
temp<-data.frame("year"=values)
```

```
fit<-predict.glm(fit,temp)
```

```
plot(year,log(deaths/exposed),main="Denmark and Japan (Red) aged 50")
```

```
lines(values,log(exp(fit)/exposed),col="red")
```

```
#Denmark
```

```
Exposure<-selectHMDdata(country="Denmark",data="Exposures",sex="Females",ages=50,years=1950:2009)
```

```
Deaths<-selectHMDdata(country="Denmark",data="Deaths",sex="Females",ages=50,years=1950:2009)
```

```
deaths<-round(Deaths)
```

```
exposed<-Exposure
```

```
year<-1950:2009
```

```
fit<-glm(deaths~ns(year,df=5)+offset(log(exposed)),family="poisson")
```

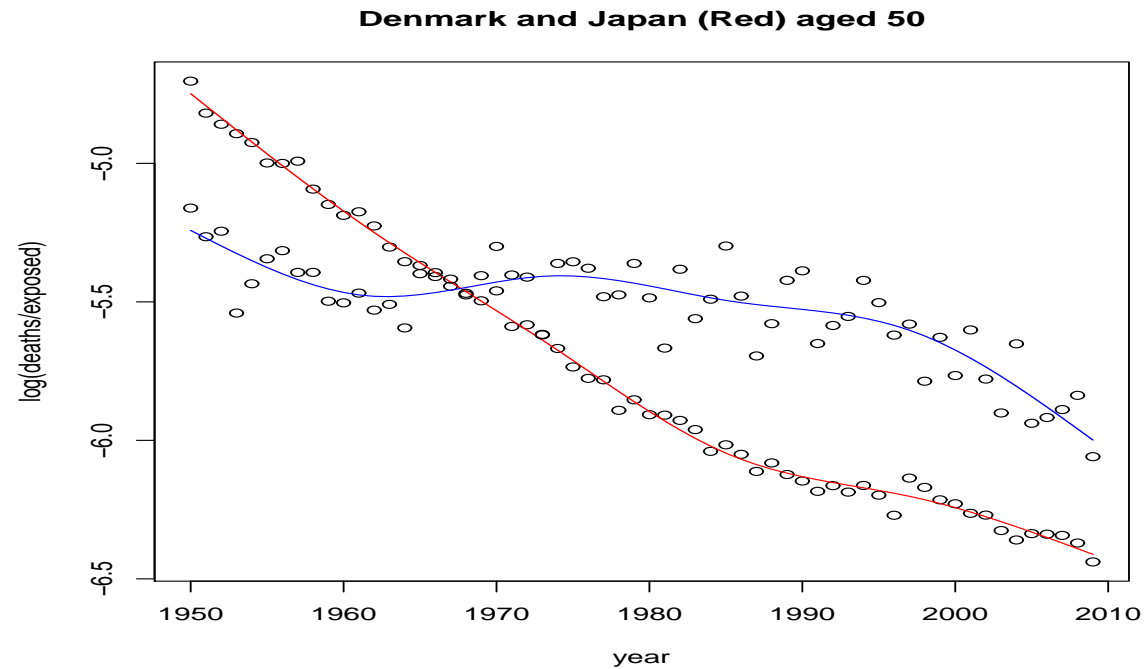
```
values<-year
```

```
temp<-data.frame("year"=values)
```

```
fit<-predict.glm(fit,temp)
```

```
points(year,log(deaths/exposed))  
lines(values,log(exp(fit)/exposed),col="blue")
```

## Examples: Denmark and Japan age 50 1950 to 2009



(won't be in exam)

## Smoothing in Two Dimensions

The package Mortality smooth has functions for smoothing. These functions will smooth the types of data we have seen above and will also smooth counts in two-dimensions. The functions will also choose the “optimal” values of the tuning parameters. See Camarda C (2012). “MortalitySmooth: An R Package for smoothing Poisson Counts with P-splines”, Journal of Statistical Software. Volume 50, for more details.

```
#Smoothing in two dimensions
#code taken from "Camarda C (2012). "MortalitySmooth: An R Package for
#smoothing Poisson Counts with P-splines", Journal of Statistical Software.
x <- 10:100
y <- 1930:2010
Y <- selectHMDdata("Sweden", "Deaths", "Females", ages = x, years = y)
E <- selectHMDdata("Sweden", "Exposures", "Females", ages = x, years = y)
fit2D <- Mort2Dsmooth(x = x, y = y, Z = Y, offset = log(E))
```



## Smoothing in Two Dimensions - Denmark

