# Data Analytics in Life insurancec

Leonie Allgaier

## Table of contents

# 1 Reading in Death Rates from the Netherlands (1850-2020) and preparing the data

```
qxt <- read.table('Mx_1x1.txt', header = TRUE)
summary(qxt)
```

```
      Year          Age               Female              Male          
 Min.   :1850   Length:18870       Length:18870       Length:18870      
 1st Qu.:1892   Class :character   Class :character   Class :character  
 Median :1934   Mode  :character   Mode  :character   Mode  :character  
 Mean   :1934                                                           
 3rd Qu.:1977                                                           
 Max.   :2019                                                           
    Total          
 Length:18870      
 Class :character  
 Mode  :character  
```

One can see that our data set is not ready yet. The columns Age, Female, Male and Total are stored as characters which we would like to be real-valued numbers. Looking at some rows of the data we see:

```
tail(qxt)
```

```
      Year  Age   Female     Male    Total
18865 2019  105 0.660964 0.911892 0.690872
18866 2019  106 0.708861 1.385705 0.766349
18867 2019  107 0.779246 2.717839 0.871544
18868 2019  108 0.901401 5.768455 0.977354
18869 2019  109 1.347922        . 1.347922
18870 2019 110+ 5.814632        . 5.814632
```

Age contains "110+" which of course is not a number. In the other columns we sometimes have "." which of course is not interpretable as a number like this.

## 1.1 Fix the age:

```
qxt <- qxt %>%
  #cut the first 3 Characters of age and convert to integer
  mutate (Age = strtoi(substr(Age, 1, 3)))
tail(qxt)
```

```
      Year Age   Female     Male    Total
18865 2019 105 0.660964 0.911892 0.690872
18866 2019 106 0.708861 1.385705 0.766349
18867 2019 107 0.779246 2.717839 0.871544
18868 2019 108 0.901401 5.768455 0.977354
18869 2019 109 1.347922        . 1.347922
18870 2019 110 5.814632        . 5.814632
```

## 1.2 Fix Female, Male and Total

We simply replace "." by 0.

```
qxt <- qxt %>%
  mutate(Female = case_when(
    (substr(Female,1,1)=='.') ~ "0",
    TRUE ~  Female
    )) %>%
  mutate(Male = case_when(
    (substr(Male,1,1)=='.') ~ "0",
    TRUE ~ Male
    )) %>%
  mutate(Total = case_when(
    (substr(Total,1,1)=='.') ~ "0",
    TRUE ~ Total
    ))
```

```
qxt <- qxt %>%
  mutate(Female = as.double(Female))%>%
  mutate(Male = as.double(Male))%>%
  mutate(Total = as.double(Total))

sapply(qxt, class)
```

```
     Year       Age    Female      Male     Total
"integer" "integer" "numeric" "numeric" "numeric"
```

```r
summary(qxt)
```

```
      Year           Age            Female               Male         
 Min.   :1850   Min.   :  0   Min.   :0.000000   Min.   :0.000000  
 1st Qu.:1892   1st Qu.: 27   1st Qu.:0.001814   1st Qu.:0.002001  
 Median :1934   Median : 55   Median :0.009632   Median :0.010408  
 Mean   :1934   Mean   : 55   Mean   :0.107305   Mean   :0.109882  
 3rd Qu.:1977   3rd Qu.: 83   3rd Qu.:0.076138   3rd Qu.:0.082792  
 Max.   :2019   Max.   :110   Max.   :6.038035   Max.   :6.036331  
     Total         
 Min.   :0.000000  
 1st Qu.:0.002114  
 Median :0.010533  
 Mean   :0.113242  
 3rd Qu.:0.085005  
 Max.   :6.038035  
```

We now have our data ready to start working.

## 2 Exercise 1.2 a)

We know that we can get the Force of Mortality from our data like this:

$$q_x(t) = 1 - p_x(t) \text{ and } p_x(t) = \exp(-\mu_x(t)) \Rightarrow \mu_x(t) = -\log(1 - q_x(t))$$

### 2.1 i) calculating mu and storing it in the right matrix

We can filter the years and age we want and then calculate mu as seen above:

```
########### Define here for which years/ages we want the model:
x_1 = 50#50
x_2 = 90#90
t_1 = 1980#1980
t_2 = 2006#2006
m = x_2 - x_1 + 1
n = t_2 - t_1 + 1
#################

mu_tx <- qxt %>%
  filter(Year >= t_1 & Year <= t_2)%>%
  filter(Age >= x_1 & Age <= x_2) %>%
  mutate(mu_total = -log(1-Total))%>%
  mutate(mu_female = -log(1-Female))%>%
  mutate(mu_male = -log(1-Male)) %>%
  subset(select = -c(Male, Total,Female))
head(mu_tx)
```

```
  Year Age    mu_total   mu_female     mu_male
1 1980  50 0.004086338 0.002854069 0.005325153
2 1980  51 0.005099983 0.003597463 0.006637983
3 1980  52 0.005205525 0.003435896 0.007035693
4 1980  53 0.005513170 0.003585420 0.007524236
5 1980  54 0.006176032 0.004029106 0.008446572
6 1980  55 0.006881624 0.004580474 0.009354618
```

We continue with only the values for the total population.

```
mu <- mu_tx %>%
  subset(select = -c(mu_female, mu_male)) %>%
  pivot_wider(names_from = Age,
              names_prefix = "age_",
              values_from = mu_total)
```

```
mu_xt = t(mu)
mu = mu[-1]
```

## 2.2 ii) Estimation of alpha, beta and kappa

We can estimate

$$A = \left(A_{t,x}\right)_{x\in\{x_1,\dots,x_m\},\ t\in\{t_1,\dots,t_n\}} = (\hat{\mu}_x(t))_{x\in\{x_1,\dots,x_m\},\ t\in\{t_1,\dots,t_n\}}$$

and

$$\hat{\alpha}_x = \bar{A}_x = \frac{1}{n}\sum_{t=t_1}^{t_n}\log(\hat{\mu}_x(t)) = \frac{1}{27}\sum_{t=1980}^{2006}\log(\hat{\mu}_x(t))$$

```
A_tx = mu
alpha = double(m)

for (x in 1:m) {
  alpha[x] = 1/n * sum(log(A_tx[,x]))
  A_tx[,x] = log(A_tx[,x]) - alpha[x]
}

A_xt = t(A_tx)
```

and then by applying svd on

$$(A_{x,t} - \hat{\alpha}_x)_{x,t}$$

we can estimate

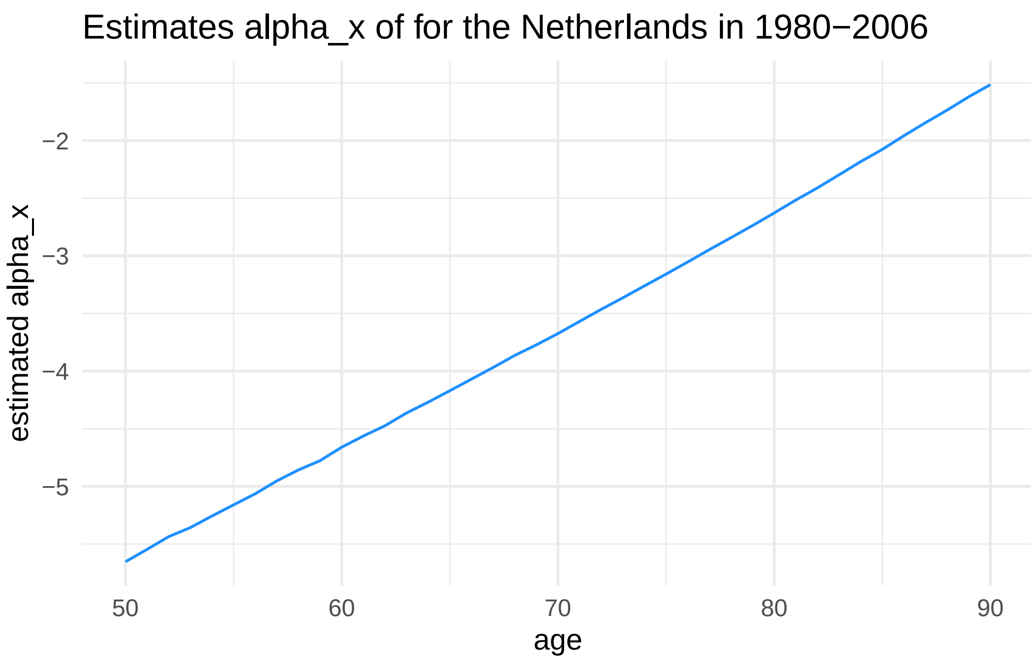
$$\hat{\beta} = c \cdot d_{1,1} \cdot u_1, \quad \hat{\kappa} = c^{-1} \cdot v_1 \text{ with } c \text{ s.t.: } \sum_x \hat{\beta}_x = 1$$

```
SVD = svd(A_xt)
d = SVD$d
u = SVD$u
v = SVD$v

beta = d[1] * u[,1]
c = 1/sum(beta)
beta = beta * c
kappa = 1/c * v[,1]
```

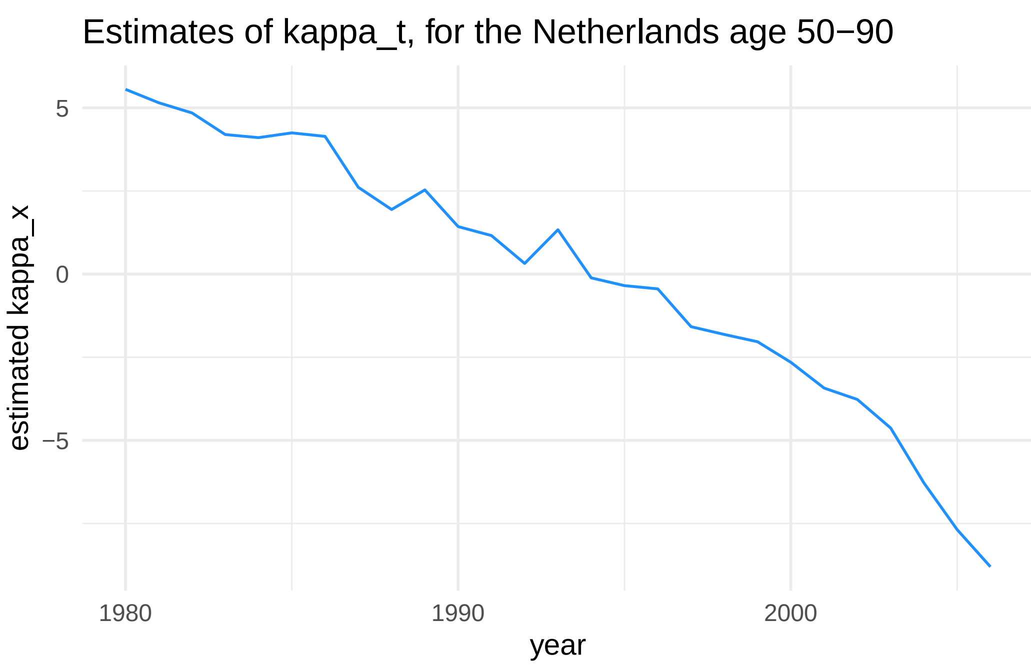
Creating the Plots:

```r
alpha_x = ggplot() +
  geom_line(aes(x_1:x_2, alpha), size=.5,col='dodgerblue')+
  labs(
    title = "Estimates alpha_x of for the Netherlands in 1980-2006",
    x = "age",
    y = "estimated alpha_x"
  )+
  theme_minimal()


beta_x = ggplot() +
  geom_line(aes(x_1:x_2, beta), size=.5, col='dodgerblue')+
  labs(
    title = "Estimates of beta_x, for the Netherlands in 1980-2006",
    x = "age",
    y = "estimated beta_x"
  )+
  theme_minimal()

kappa_t = ggplot() +
  geom_line(aes(t_1:t_2, kappa), size=.5, col='dodgerblue')+
  labs(
    title = "Estimates of kappa_t, for the Netherlands age 50-90",
    x = "year",
    y = "estimated kappa_x"
  )+
  theme_minimal()
```

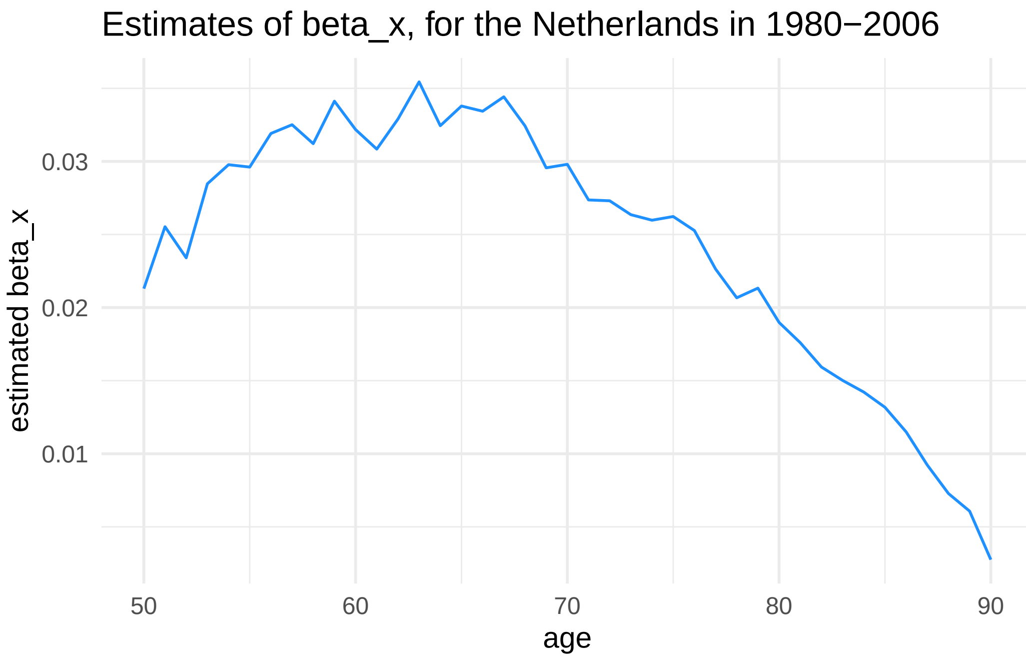
# 3 Exercise 1.2 b) Plotting and their interpretation:




Alpha is in the age interval we're looking at almost perfectly linear growing, only when looking very closely we can see some variation there. From this we can we derive that the base shape of the mortality of the population we're viewing is growing with age. (The older one is the higher ones mortality gets).



In Kappa we have more variation but see a general downwards trend over the years. In Kappa we see the change of mortality according to what year we have. Therefore we can say, that overall the mortality decreases over the years. This can for example be due to new medical breakthroughs and therefore higher chance of surviving various diseases in later years.




Beta gives us the sensitivity of age x to Kappa. The larger Beta is, the more impact the current year has on an age group. We have Beta increasing at first but then decreasing again with a maximum around 60-65. Therefore at this age the mortality is most sensitive to the current year. Before and after we have less impact of the current year on the mortality with close to zero impact approaching age 90.

## 4 Exercise 1.2 c)

Estimating sigma

$$\hat{\sigma}_\epsilon^2 = \frac{1}{mn}\sum_{x=x_1}^{x_m}\sum_{t=t_1}^{t_n}\{\log(\mu_x(t)) - \alpha_x - \beta_x\kappa_t\}^2$$

$$= \frac{1}{41*27}\sum_{x=50}^{90}\sum_{t=1980}^{2006}\{\log(\mu_x(t)) - \alpha_x - \beta_x\kappa_t\}^2$$

```
res = mu

for (x in 1:m){
  for (t in 1:n){
    res[t,x] = (log(res[t,x]) - alpha[x] - beta[x]*kappa[t])^2
  }
}

sigma = 1/(m*n)*sum(res)

glue::glue('Estimate for sigma^2 is {sigma}.')
```

```
Estimate for sigma^2 is 0.000832367155000623.
```

# 5 Exercise 1.2 d)

Computing the percentage of variance explained by our model.

$$\frac{\sum_{x=x_1}^{x_m}\sum_{t=t_1}^{t_n}(\log(\hat{\mu}_x(t))-\alpha_x)^2}{\sum_{x=x_1}^{x_m}\sum_{t=t_1}^{t_n}(\log(\mu_x(t))-\alpha_x)^2}$$

```r
temp1 = mu
temp2 = mu

for (x in 1:m){
  for (t in 1:n){
    temp1[t,x] = (beta[x] * kappa[t])^2
    temp2[t,x] = (log(temp2[t,x]) - alpha[x])^2
  }
}

POV = sum(temp1)/sum(temp2)

glue::glue('Percentage of Variance explained by the model is {POV}.')
```

```
Percentage of Variance explained by the model is 0.925044439093751.
```

# 6 Exercise 1.2 e)

```
C = 1/(n - 1) * (kappa[n] - kappa[1])

glue::glue('Estimate for C is {C}')
```

```
Estimate for C is -0.552248996439625
```

```
sigma_kappa = 0

for (i in 2:n){
  sigma_kappa = sigma_kappa + (kappa[i] - kappa[i - 1] - C) ^ 2
}

sigma_kappa = sigma_kappa/(n - 1)

glue::glue('Estimate for sigma_kappa^2 is {sigma_kappa}.')
```

```
Estimate for sigma_kappa^2 is 0.395841834296157.
```

# 7 Exercise 1.2 f)

We already know alpha_60 and beta_60, and want a forecast of the mortality rate in 2040. So Kappa is to be predicted into the future. We do this by using

$$\kappa_{t_n+k} = \kappa_{t_n} + k \cdot C + \sum_{j=1}^{k} \epsilon_{t_n+j} \Rightarrow \hat{\kappa}_{t_n+k} = \hat{\kappa}_{t_n} + k \cdot \hat{C} = \hat{\kappa}_{2006} + 34 \cdot \hat{C}$$

```
kappa_2040 = kappa[27] + 34*C
kappa_2040
```

```
[1] -27.57756
```

$$\Rightarrow \hat{\mu}_{60}(2040) = \exp(\hat{\alpha}_{60} + \hat{\beta}_{60} \cdot \hat{\kappa}_{2040})$$

```
mu_60_2040 = alpha[11] + beta[11] * kappa_2040
```

To get the probability of a 60 year old person surviving until there over 75 we use the formula given on slide 33.

$${}_sp_x(t) = \exp\left(-\sum_{j=0}^{\lfloor s \rfloor - 1} \mu_{x+j}(t+j) - (s - \lfloor s \rfloor)\mu_{x+\lfloor s\ rfloor}(t + \lfloor s \rfloor)\right)$$

therefore

$$\Rightarrow_{15}\hat{p}_{60}(2040) = \exp\left(-\sum_{j=0}^{14} \hat{\mu}_{60+j}(2040+j)\right)$$

Predicting kappa as seen before and with that predicting the future force of mortality we get:

```
sum_mu = 0
kappa_tmp = kappa_2040

for (j in 0:14){
sum_mu = sum_mu + exp(alpha[11 + j] + beta[11 + j] * kappa_tmp)
kappa_tmp = kappa_tmp + C
}

prob = exp(- sum_mu)
```

```r
perc = prob * 100
glue::glue('Probability of a person aged 60 in 2040 to become
              strictly older than 75 is {perc}%.')
```

```
Probability of a person aged 60 in 2040 to become
strictly older than 75 is 88.7847173667374%.
```