Workshop Desvendando o Metodo

Aplicação das relações da variável-r com modelos de mortalidade para a estimar a população de centenários no Brasil

Marília Nepomuceno

Contents

1. Brief introduction

Over the last decades, demographers have been given more and more attention to those people who live more than 100 years, the so-called centenarians.

There are two main debates regarding the population at very old ages:

a) Is there a limit of human lifespan?
b) How do mortality rates change at very old ages?

To answer these questions, it is crucial to have precise data about mortality and survival at very old ages.

Here, I will discuss the mortality models involved in this debate and an alternative approach to estimate the number of centenarians in the absence of reliable numbers.

2. Load required packages.

```
# Load packages to data manipulation and plots
library(tidyverse)
library(MortalityLaws)
```

3. Brazil in the international context

```r
# Read the data
Prop60= read.table("Data\\Proportion60.csv",header = TRUE, sep = ",")

# Look at the variable names and types
glimpse(Prop60)
```

```
## Rows: 74
## Columns: 6
## $ Year       <int> 1900, 1920, 1940, 1950, 1960, 1970, 1980, 1991, 2000, 20...
## $ Sex        <chr> "Men", "Men", "Men", "Men", "Men", "Men", "Men", "Men", ...
## $ Source     <chr> "Brazil-Census", "Brazil-Census", "Brazil-Census", "Braz...
## $ Prevalence <dbl> 71.57, 43.06, 35.89, 31.09, 21.10, NA, 10.36, 8.62, 15.9...
## $ e0         <dbl> 33.30, 33.80, 42.90, 45.30, 49.70, 54.60, 59.70, 63.13, ...
## $ e50        <dbl> NA, NA, 16.90, 20.58, 21.70, 22.36, 22.00, 24.20, 25.87,...
```

```r
# Create a vector with the country's order we want to show in the plot
country_levels <- c("Brazil-Census","Sweden","Japan","USA")

# Order the data
Prop60 <- Prop60 %>%
  mutate(Source = factor(Source, levels = country_levels))
```

```r
# Plot e0 vs. Proportion
Prop60 %>% filter(Source %in% c("Brazil-Census",
                                "Sweden",
                                "Japan","USA")) %>%
  filter(Sex == "Men" |  Sex == "Women") %>%
  ggplot(aes(x=Year, y=e0, colour=Source)) +
  geom_point(aes(colour = Source,shape = Source,size = Prevalence), alpha = 0.7)+
  theme_minimal()+
  scale_x_continuous(breaks = seq(1900, 2010, 10))+
  scale_y_continuous(breaks = seq(30, 90, 5))+
  scale_color_manual(name = "Country",
                     values = c("red",
                                "#147A5B",
                                "orange",
                                "#645EB2"),
                     labels = c("Brazil (Census)",
                                "Sweden",
                                "Japan","USA")) +
  scale_shape_manual(name = "Country",
                     values = c(16,15,17,18),
                     labels = c("Brazil (Census)",
                                "Sweden",
                                "Japan","USA"))+
  scale_size_area(name="Proportion (per 10,000)",breaks=c(1,5,10,15,20),max_size = 10)+
  geom_line(alpha = 0.5, size=1.05)+
  facet_grid(~Sex)+
  ylab("Life Expectancy at Birth")+
  theme(axis.text=element_text(size=9),axis.title=element_text(size=10))+
  ggtitle("Figure 1 - Life expectancy at birth and proportion of centenarians among the population aged
selected countries, men and women, 1900-2010")+
```

```
theme(plot.title = element_text(color="black", size=12),plot.caption = element_text(hjust=0, size=10))
theme(strip.text = element_text(size=10,lineheight=5.0, hjust=0))+
theme(plot.margin = unit(c(1,1,1.5,1), "cm"))+
guides(color = guide_legend(order = 1),
       size = guide_legend(order = 2),
       shape = guide_legend(order = 1))+
labs(caption = "Source: Author's calculations based on HMD(2020) and Brazilian Census Data.")
```

Figure 1 – Life expectancy at birth and proportion of centenarians among the population aged 60 or more: Brazil and selected countries, men and women, 1900–2010
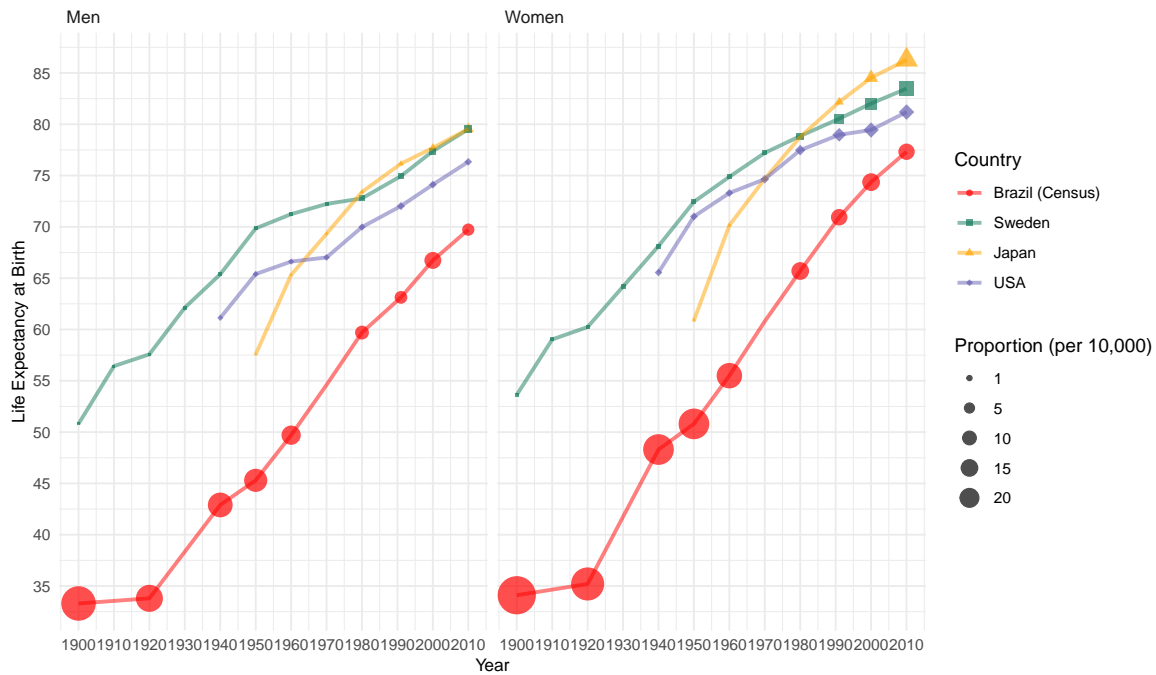
Source: Author's calculations based on HMD(2020) and Brazilian Census Data.

A comparison of the proportion of centenarians to the population aged 60 or more in Brazil and today's low mortality countries, as presented in the figure above helps to bring to light limitation on census data. Although the life expectancy at birth was considerably lower in Brazil than in the selected low-mortality countries, the proportion of centenarians was much higher, particularly in the earliest decades of the twentieth century. Over time, mortality levels decline, and the proportion of centenarians increased in Sweden, Japan and the US. The opposite trend happened in Brazil, probably as a result of improving census data quality.

To date, little is still known about the number of centenarians in developing countries. This prevents a more comprehensive discussion on the dynamic of this population beyond the limited set of lowest mortality countries.

To get a more feasible numbers of the centenarian population in Brazil, we use variable-r relations to estimate the centenarian population in Brazil from the combination of different mortality schedules.

4. The method

Variable-r relations can be of great value in the estimation of population measures when data are missing or are of bad quality. Also, they allow the re-establishment of stationary conditions in nonstable populations, which is especially useful for demographic measurement in the context of the demographic transition.

The connection is made by the set of age-specific growth rates, defined by the relationship between the sizes of the population at each age group in two points in time

$$e^{\int_0^t \bar{r}(x,t)dt} = \frac{{}_nN_x(t)}{{}_nN_x(0)}$$

where ${}_nN_x(0)$ and ${}_nN_x(t)$ are the populations at age group x and $x+n$, respectively, at time 0 and t, and $\bar{r}(x,t)$ is the mean age-specific growth rate over time intervals for the age groups x and $x+n$.

One of the applications of variable-r methods is to express the number of individuals at any given time t in terms of the number of people at another age at the same time t, age-specific population growth rates, and the probability of surviving between those ages.

The set of age-specific growth rates provides a growth correction by adjusting for any differences in the size of the two birth cohorts that come from variations in the number of births, and in cumulative differences of mortality and migration rates.

In the words of Preston et al 2001: "...all the pertinent history is contained in the age-specific growth rate function."

$$_nN_x(t) = {}_nN_y(t)e^{-\int_y^x \bar{r}(a,t)da}{}_{x-y}p_y(t)$$

where ${}_nN_x(t)$ is the population at age group x and $x+n$ at time t, ${}_nN_y(t)$ is the population aged y and $y+n$ at time t, which $x > y$, $\bar{r}(a,t)$ is the mean age-specific growth rate over time intervals, and ${}_{x-y}p_y(t)$ is the conditional probability of surviving from ages y to x at time t.

To solve the equation below and get ${}_nN_x(t)$, we need:

a) ${}_nN_y(t)$
b) $\bar{r}(a,t)$
c) ${}_{x-y}p_y(t)$

Note: Demographic rates pertain to discrete time intervals rather than to a point in time. Therefore, we adapt the equations above to discrete time intervals.

5. Application - Female 2000

5.1 Lower age limit

```
# define the age y and population at age group y and y+n in 2000
# y =  50

N50 <- 6506394
```

5.2 Age specific growth rates

```r
# Data to calculate growth rates

age.groups <- c('50-59', '60-69','70-79',' 80-89','90-99')

pop.1991 <- c(4863177,3388974,1731974,585157,74967)

pop.2000 <- c(6506394,4389501,2511989,943316,152286)

pop.2010 <- c(9679284,6084830,3547194,1507073,278400)

# Age-specific growth rates

t_1991.2000  <-  8.92 # difference in years between the 1991 and the 2000 Census
r_1991.2000  <-  log(pop.2000/pop.1991)/t_1991.2000
r_1991.2000
```

```
## [1] 0.03263380 0.02900093 0.04168308 0.05353376 0.07945205
```

```r
t_2000.2010  <-  10 # difference in years between the 2000 and the 2010 Census
r_2000.2010  <-  log(pop.2010/pop.2000)/t_2000.2010
r_2000.2010
```

```
## [1] 0.03972025 0.03265832 0.03450820 0.04685233 0.06032986
```

```r
# Hypothesis:
# The growth rate at age group 100-119 is equal to the growth rate at age group 90-99

r_1991.2000 <-  c(r_1991.2000,r_1991.2000[length(r_1991.2000)])
r_1991.2000
```

```
## [1] 0.03263380 0.02900093 0.04168308 0.05353376 0.07945205 0.07945205
```

```r
r_2000.2010 <-  c(r_2000.2010,r_2000.2010[length(r_2000.2010)])
r_2000.2010
```

```
## [1] 0.03972025 0.03265832 0.03450820 0.04685233 0.06032986 0.06032986
```

Convert the $\int_y^x \bar{r}(a,t)da$ for exact ages into equations for the age intervals in which data are normally found:

$$\int_{50}^{100} \bar{r}(a)da \approx 5 *_{10} r_{50} + 10 * \sum_{a=60,10}^{100-10} {}_{10}r_a + 5 *_{10} r_{100}$$

```r
# Sum age-specific growth rates

d = c(5,10,10,10,10,5)

S_1991.2000  <-  exp(-sum(r_1991.2000*d))
S_1991.2000
```

```
## [1] 0.07448723
```

```r
S_2000.2010  <-  exp(-sum(r_2000.2010*d))
S_2000.2010
```

```
## [1] 0.1060613
```

```r
# We wanted to make estimates separately for each census year,
# rather than for intecensal period.
# Therefore, we used to intercensal growth rates that are centered
# on each census date to calculate the mean growth rates

# Assumption: the growth rate changes linearly during the time interval

S_2000  <-  (S_1991.2000*S_2000.2010)^0.5
S_2000
```

```
## [1] 0.08888316
```

5.3 Conditional probability of surviving

5.3.1 The Debate on the Mortality Plateau

First, let's talk a bit about **the Debate on the Mortality Plateau**

There is widespread agreement that mortality rates increase exponentially between mid adulthood through ages 80 to 90. The Gompertz lar states that the force of mortality increases exponentially over adult ages, and it is probably the most important of the mortality laws. This assumption has proven to be an accurate and useful representation of the mortality pattern among diverse populations throughout the last centuries.

The figure below shows the fit of the Gompertz model to the male population of Switzerland in 2010. The graph shows the assumption of exponential increase of mortality with age to be a good representation of the human mortality pattern.

```r
# This data object created in SSE.R
DX <- c(217, 17, 9, 8, 10, 7, 5, 4, 4, 11, 10, 6, 6, 8, 13, 12, 10,
        25, 32, 37, 49, 45, 48, 33, 39, 48, 48, 58, 42, 54, 47, 59, 58,
        64, 54, 66, 68, 74, 93, 74, 89, 88, 105, 112, 112, 105, 113,
        135, 157, 165, 183, 185, 226, 262, 226, 259, 301, 333, 315, 342,
        345, 352, 361, 404, 430, 467, 479, 530, 586, 636, 652, 690, 670,
        750, 760, 866, 884, 950, 985, 1054, 1074, 976, 900, 920, 953,
        983, 964, 898, 822, 712, 651, 613, 450, 395, 253, 195, 140, 87,
        73, 39, 33, 17, 9, 5, 2, 0, 0, 0, 0, 0, 0)
EX <- c(39980.5, 40331, 40832.5, 41767.5, 42090.5, 41878.5, 42626.5,
        43947, 45087.5, 45245.5, 44681, 44235.5, 43495, 42805.5, 42704,
        42827, 42852.5, 42653.5, 42908, 42928, 42327, 41416.5, 41457,
        41487.5, 41828, 43270.5, 44962, 46922, 49138.5, 51123, 53192,
        55358, 57517.5, 59765.5, 62033.5, 63767, 63794, 62494, 60825,
        59561.5, 58440.5, 57074.5, 56005.5, 55159.5, 53630, 51941, 50460.5,
        49869.5, 49207.5, 49311.5, 49792.5, 49721.5, 49760.5, 49505.5,
        48749.5, 47447.5, 45978.5, 44169, 41469, 38244, 36041, 34893,
        33470.5, 32602.5, 32242, 31326.5, 29867, 28764, 27864, 27003.5,
        25864.5, 24411.5, 23303.5, 22292.5, 21465.5, 20294.5, 19246.5,
        18193, 17190.5, 16204, 13758.095, 11652.475, 10355.17, 9281.875,
        8314.4, 7685.01, 6897.155, 5820.03, 4728.975, 3801.81, 3072.335,
        2346.85, 1723.14, 1258.645, 871.085, 558.4, 357, 243, 152, 80.5,
        53.5, 25.5, 15.5, 5.5, 1.5, 1.5, 1, 0.5, 0, 0, 0)
ages        <- 0:110
mx.obs      <- DX / EX
x           <- 40:100       #age range to fit the model
m           <- length(x)
Dx_sub      <- DX[ages%in%x]#filter deaths in the age range to fit the model
```

```r
Ex_sub      <- EX[ages%in%x]#filter exposures in the age range to fit the model
mx.obs_sub  <- mx.obs[ages%in%x]#filter death rates in the age range to fit the model

fit.gompertz <- MortalityLaws::MortalityLaw(x = x, Dx = Dx_sub, Ex = Ex_sub,
        law = "gompertz", opt.method = "poissonL") #poisson distribution for death counts

my.cols.obs  <- c("grey20")
my.cols.mod  <- c("#d7191c")
par(mai=c(1,1.2,.1,.1))
plot(x,mx.obs_sub,log="y",type="n",axes=FALSE,xlab="",ylab="")
mtext("Age",side=1,line=2.5,cex=1.5)
mtext("Mortality (log scale)",side=2,line=3.5,cex=1.5)
abline(h=c(0.002,0.01,0.05,0.5),lty=3,col="grey70")
abline(v=seq(40,100,10),lty=3,col="grey70")
axis(1);axis(2,las=2,at=c(0.002,0.01,0.05,0.5));box()
points(x,mx.obs_sub,pch=5,cex=1.5,col=my.cols.obs,lwd=2)
lines(x,fit.gompertz$fitted.values,col=my.cols.mod,lwd=2.5)
legend("bottomright",c("Observed","Gompertz"),col=c(my.cols.obs,my.cols.mod),
        lwd=2.5,bty="n",cex=1.5,pch = c(5,NA),lty=c(NA,1))
```
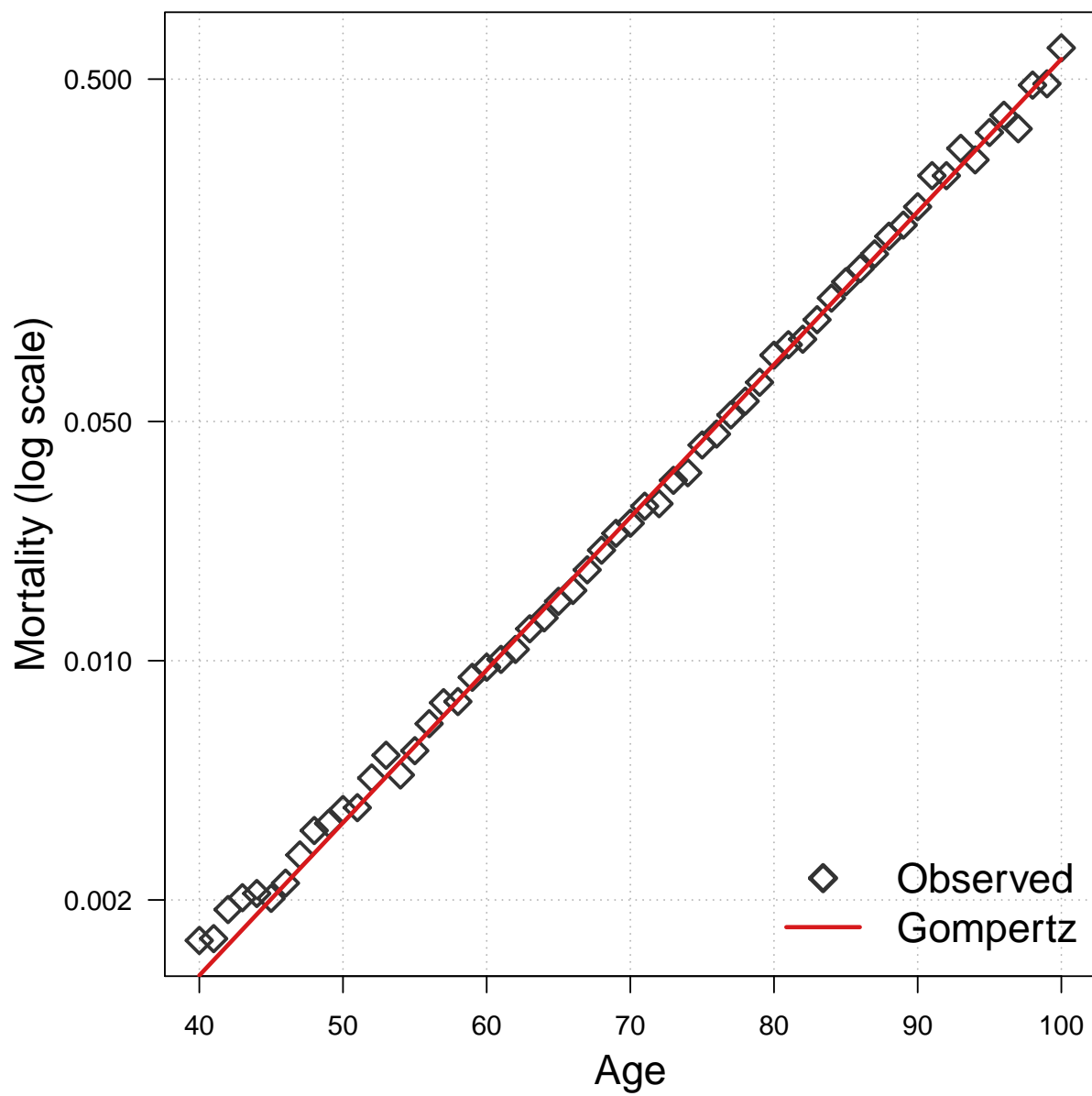
Figure 1: Observed and fitted death rates (in log-scale) using a log-linear (Gompertz) model for males aged 40 to 100 in Switzerland in the year 2010 (data: HMD).

What about the age pattern of mortality at ages above 90?

There is no consensus regarding the mortality rate trajectory at the most advanced ages. Some studies conclude that the exponential growth of mortality with age is followed by a period of deceleration, with slower rates of mortality increase at the oldest ages leading to a plateau of human mortality. Models including Kannisto, Beard and Perks produce a mortality plateau. Others conclude that the mortality deceleration is an artifact of data with lower quality, and that mortality continues to grow exponentially through the highest ages.

```r
# load GBRENW 2010 female data
source("Data/HMD.R")
ages                    <- 0:110
year            <- 2010
cou             <- "GBRTENW"
names(Dx)       <- names(Ex) <- ages
mx.obs          <- Dx / Ex
## focus on old age mortality
x.start         <- 80
x               <- x.start:tail(ages,1)
m               <- length(x)
Dx_sub          <- Dx[ages%in%x]
Ex_sub          <- Ex[ages%in%x]
mx.obs_sub      <- mx.obs[ages%in%x]
mx.lt_sub       <- mx.lt[ages%in%x]


## select four models to fit to the data
models <- c("gompertz","ggompertz","kannisto","weibull")
fit.gompertz  <- MortalityLaw(x = x, Dx = Dx_sub, Ex = Ex_sub,
                              law = models[1], opt.method = "poissonL")
fit.ggompertz <- MortalityLaw(x = x, Dx = Dx_sub, Ex = Ex_sub,
                              law = models[2], opt.method = "poissonL")
fit.kannisto  <- MortalityLaw(x = x, Dx = Dx_sub, Ex = Ex_sub,
                              law = models[3], opt.method = "poissonL")
fit.weibull   <- MortalityLaw(x = x, Dx = Dx_sub, Ex = Ex_sub,
                              law = models[4], opt.method = "poissonL")


## choose colors
my.cols.obs <- c("grey40","grey20")
#my.cols.mod <- c("#d7191c","chartreuse4","darkorange","dodgerblue3")
my.cols.mod <- c("#008E77","#458200","#A46400","#C23F86")
my.lty.mod  <- c("81","44","63","8222")
par(mai=c(1,1.2,.1,.1))
plot(x,mx.obs_sub,log="y",t="n",axes="F",xlab="",ylab="")
mtext("Age",side=1,line=2.5,cex=1.5)
mtext("Mortality (log scale)",side=2,line=3.5,cex=1.5)
abline(h=c(0.05,0.1,0.2,0.5,1),lty=3,col="grey70")
abline(v=seq(80,110,5),lty=3,col="grey70")
axis(1);axis(2,las=2);box()
points(x,mx.obs_sub,pch=5,cex=1.5,col=my.cols.obs[2],lwd=2)
points(x,mx.lt_sub,pch=1,cex=1.7,col=my.cols.obs[1],lwd=2)
legend("topleft",c("Observed","Life Table"),col=rev(my.cols.obs),
       pch=c(5,1),bty="n",cex=1.5,lwd=2,lty=NA,pt.cex=c(2,2.25))
matlines(x,cbind(fit.gompertz$fitted.values,
                 fit.weibull$fitted.values,
                 fit.ggompertz$fitted.values,
```

```
                fit.kannisto$fitted.values),
        col=my.cols.mod,
        lwd=2.5,
        lty=my.lty.mod)
legend("bottomright",c("Gompertz","Weibull","Gamma-Gompertz","Kannisto"),col=my.cols.mod,
       lwd=2.5,lty=my.lty.mod,bty="n",cex=1.2)
```
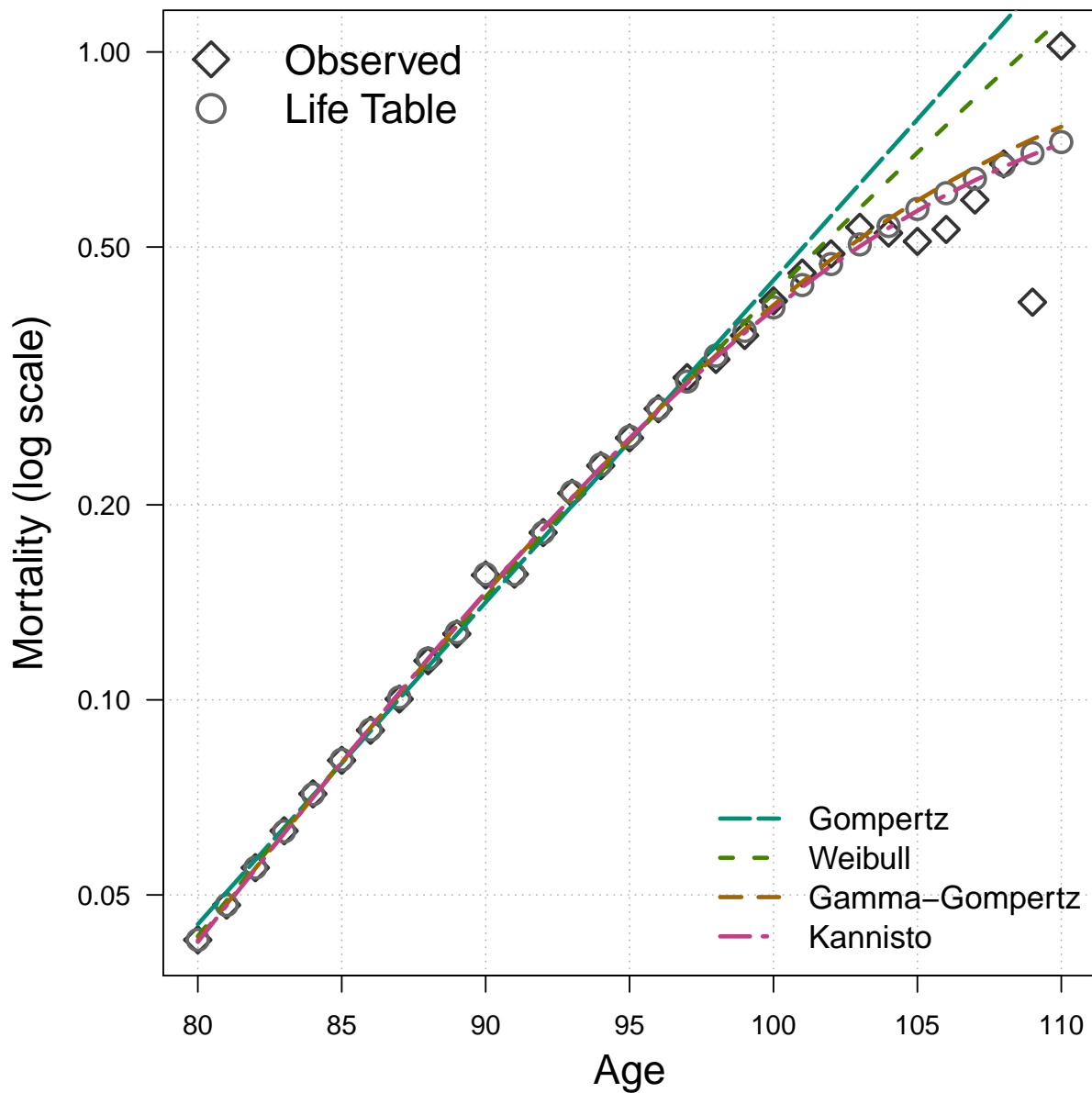
Figure 2: Observed and life table death rates (in log scale) for females in England and Wales in the year 2010, with fit of four different parametric mortality models. The Gompertz model is log linear, whereas the Kannisto and Gamma-Gompertz models deviate after age 95, tapering toward a plateau. The Weibull extrapolation is intermediate in this case.

5.3.2 Estimate probabilities of surviving in Brazil

As a way to address the current disagreement in the literature, we decided to offer estimates for the number of centenarians in Brazil from the application of three different mortality models:

  a) Gompertz
  b) Kannisto
  c) Weibull

Note: In the case of Brazil, instead of using the number of deaths and the exposures, we used mortality rates from the Life Tables calculated by the IBGE.

```r
#--------
# Brazil

ages <- c(0,1,seq(5,90,5))

x.start  <-  70 # beginning of age interval to estimate the parameters
x.end  <-   90 # open age interval to estimate the parameters
x <-  seq( x.start,x.end,by=5)

single.age  <-  seq(50,120,1)

#-------------
# Female 2000

women2000 <-  read.table("Data\\Mulheres_Censo_TabMort2000-90.txt", header=T)

Lx.f2000  <-  matrix(0,ncol=4,nrow=length(single.age))
colnames(Lx.f2000)  <-  c( 'age','Kannisto', 'Gompertz', 'Weibull')
Lx.f2000[,1] <-  single.age

Lx.f2000[,"Kannisto"] <-  convertFx(x=single.age,
                              data = predict(MortalityLaw(x = x,
                                                  mx = women2000$mx[ages%in%x],
                                                  law = 'kannisto'), x =single.age),
                              from = "mx", to = "Lx")
Lx.f2000[,"Gompertz"] <-  convertFx(x=single.age,
                              data = predict(MortalityLaw(x = x,
                                                  mx = women2000$mx[ages%in%x],
                                                  law = 'gompertz'), x =single.age),
                              from = "mx", to = "Lx")
Lx.f2000[,"Weibull"] <-   convertFx(x=single.age,
                              data = predict(MortalityLaw(x = x,
                                                  mx = women2000$mx[ages%in%x],
                                                  law = 'weibull'), x = single.age),
                              from = "mx", to = "Lx")
```

After having predicted the mortality according to the three different models, we will calculate the probability of surviving between age 50-59 and 100-109

$$_{100}p_{50} \approx \frac{_{10}L_{100}}{_{10}L_{50}}$$

Then,

$$_{110}p_{50} \approx \frac{_{10}L_{110}}{_{10}L_{50}}$$

```
#-------
# 50-59/100-109 and 50-59/110-119


x1 = seq(50,59,1)
x2 = seq(100,109,1)
x3 = seq(110,119,1)

# Women
Lx.SR.f2000 = matrix(0,ncol=3,nrow=2)

colnames(Lx.SR.f2000) = c('Kannisto', 'Gompertz', 'Weibull')
rownames(Lx.SR.f2000) = c('p50.59-100.109', 'p50.59-110.119')

Lx.SR.f2000["p50.59-100.109","Kannisto"] =
  sum(Lx.f2000[single.age%in%x2,"Kannisto"])/sum(Lx.f2000[single.age%in%x1,"Kannisto"])
Lx.SR.f2000["p50.59-110.119","Kannisto"] =
  sum(Lx.f2000[single.age%in%x3,"Kannisto"])/sum(Lx.f2000[single.age%in%x1,"Kannisto"])

Lx.SR.f2000["p50.59-100.109","Gompertz"] =
  sum(Lx.f2000[single.age%in%x2,"Gompertz"])/sum(Lx.f2000[single.age%in%x1,"Gompertz"])
Lx.SR.f2000["p50.59-110.119","Gompertz"] =
  sum(Lx.f2000[single.age%in%x3,"Gompertz"])/sum(Lx.f2000[single.age%in%x1,"Gompertz"])

Lx.SR.f2000["p50.59-100.109","Weibull"] =
  sum(Lx.f2000[single.age%in%x2,"Weibull"])/sum(Lx.f2000[single.age%in%x1,"Weibull"])
Lx.SR.f2000["p50.59-110.119","Weibull"] =
  sum(Lx.f2000[single.age%in%x3,"Weibull"])/sum(Lx.f2000[single.age%in%x1,"Weibull"])

Lx.SR.f2000
```

```
##                   Kannisto     Gompertz      Weibull
## p50.59-100.109 0.0015379331 8.413082e-04 1.393372e-03
## p50.59-110.119 0.0000051817 9.359693e-08 1.505751e-06
```

5.3.2 Results

The estimated number of female centenarians in Brazil is:

```
N100 = N50 * S_2000 *Lx.SR.f2000
N100
```

```
##                   Kannisto     Gompertz      Weibull
## p50.59-100.109 889.400298 486.53598143 805.7991866
## p50.59-110.119   2.996623   0.05412793   0.8707893
```

```
colSums(N100)
```

```
## Kannisto Gompertz  Weibull
## 892.3969 486.5901 806.6700
```

6 Exercise

Calculate the number of male centenarians in Brazil in 2000. Consider the same set of age-specific growth rates for women.

```
#-----------
# Male 2000

men2000  <-  read.table("Data\\Homens_Censo_TabMort2000-90.txt", header=T)

N50.male <-  6000922

#N100.male = ??
```