

Universidad Autónoma del Estado de México
Licenciatura en Actuaría

Análisis de la Mortalidad en México mediante el Modelo Lee-Carter: Integración de Datos Históricos y Proyecciones

Autores:

**Fuentes Hinojosa Santiago de Jesus
Lopez Valerio Santiago**

Toluca, Estado de México a 26 de marzo de 2025

1. Objetivo

El objetivo principal de este trabajo es analizar la evolución de la mortalidad en la población mexicana a lo largo de más de un siglo, integrando datos históricos y recientes para aplicar el modelo Lee-Carter. Con ello, se busca identificar tendencias en las tasas de mortalidad y proyectar su comportamiento futuro, de esta manera se pretende facilitar la toma de decisiones informadas en ámbitos actuariales y de salud pública en México.

2. Introducción

En los estudios actuariales y demográficos, el análisis de la mortalidad juega un papel importante para la toma de decisiones en políticas públicas, planes de pensiones y seguros de vida. El presente proyecto tiene como objetivo analizar las tendencias de mortalidad en México, haciendo uso del modelo de Lee-Carter, uno de los métodos más reconocidos para modelar y pronosticar tasas de mortalidad a través del tiempo.

La relevancia de este estudio radica en la necesidad de entender cómo han evolucionado las tasas de mortalidad por edad y sexo a lo largo de más de un siglo, así como de contar con herramientas sólidas para proyectar estos patrones en el futuro. De esta forma, se pueden estimar expectativas de vida y riesgos actuariales con mayor precisión, contribuyendo a una mejor planificación y gestión de recursos en áreas como seguridad social, salud pública y seguros.

A lo largo del documento se abordarán los siguientes aspectos:

1- Revisión de datos y preparación de la base: Se describe el origen de los datos, su limpieza y la integración de distintas fuentes para conformar la base de datos final.

2- Análisis descriptivo: Se exploran las tasas de mortalidad por edad y año para hombres y mujeres, presentando gráficas que ilustran la evolución y los patrones más relevantes.

3- Aplicación del modelo Lee-Carter: Se explica la metodología del modelo y su ajuste a los datos para ambos sexos, mostrando los resultados y discutiendo su interpretación.

4- Conclusiones: Se presentan los hallazgos principales del análisis y se discuten sus implicaciones en ámbitos demográficos y actuariales, así como posibles líneas de investigación futura.

3. Revisión de Literatura

1. Mortality Forecasting in Latin America: An Assessment of the Lee-Carter Model

Mortality Forecasting in Latin America: An Assessment of the Lee-Carter Model

Janssen, F., & Debón, A. (2021). An age-specific evaluation of the Lee-Carter mortality forecasting model. *Scandinavian Actuarial Journal*, 2021(8), 654-679. <https://doi.org/10.1080/03461238.2021.192854>

Los autores evalúan la aplicación del modelo Lee-Carter para pronosticar la mortalidad en distintos países de América Latina. Su investigación destaca los siguientes aspectos relevantes para este proyecto:

1- Ventajas del modelo Lee-Carter: Se enfatiza cómo este modelo captura de forma efectiva la tendencia general de la mortalidad a través del tiempo, descomponiendo las tasas de mortalidad en factores que permiten aislar patrones de cambio a largo plazo.

2- Adaptabilidad a distintas poblaciones: El estudio muestra que el enfoque Lee-Carter puede ajustarse a diversas realidades demográficas, resultando útil para países con cambios acelerados en su estructura etaria y en sus patrones de mortalidad.

3- Importancia de la calidad de datos: Se subraya la necesidad de contar con series de mortalidad históricas consistentes y completas para mejorar la fiabilidad de las proyecciones, algo que este proyecto también busca asegurar al integrar múltiples fuentes de datos.

4- Proyecciones a mediano y largo plazo: El trabajo demuestra que el modelo Lee-Carter ofrece una forma relativamente sencilla y robusta de proyectar la evolución futura de la mortalidad, lo que lo convierte en una herramienta valiosa para la planificación de políticas públicas y la evaluación actuarial.

En conjunto, esta literatura respalda la pertinencia de emplear el modelo Lee-Carter en el presente proyecto, al ofrecer evidencia empírica de su eficacia y versatilidad para analizar y pronosticar tendencias de mortalidad en contextos demográficos complejos.

2. Latin American Mortality Database

Latin American Mortality Database

University of Wisconsin-Madison. (n.d.). Latin mortality project (2nd ed.). Center for Demography and Ecology. <https://www.ssc.wisc.edu/cdha/latinmortality2/>

Esta base de datos se utilizó para obtener las tablas de vida empleadas en la estimación del modelo Lee-Carter. En ella se encuentran las tablas de vida correspondientes a México clasificadas por sexo.

3. The Latin American Mortality Database (LAMBdA). Methodological Document Version II: August 2021

Al momento de acceder a las tablas de vida, podemos encontrar esta documentación la cual explica los métodos que siguen para estimar los datos que tienen el base de datos. El capítulo 3 es el que se centra en la parte de la estimación de las tablas de vida desde 1850-2020.

1- Fuentes de datos: se describen las principales fuentes de datos utilizados para la estimación de las tablas de vida, en los que se incluyen censos de población y estadísticas vitales de defunciones.

Estas fuentes fueron la clave para construir las estimaciones consistentes de la mortalidad, sin embargo, la disponibilidad de estos datos variaban a lo largo del tiempo y por países.

También relatan la complejidad a la que se enfrentaron, ya que una de las principales limitaciones en la construcción de tablas de vida fue la falta de datos precisos, especialmente antes de 1950. Esto debido a los censos de población y a las estadísticas de mortalidad para cada país.

2- Ajustes en las tasas de mortalidad: para estimar las tasas de mortalidad que vienen en las tablas de vida, las cuales son fundamentales para el modelo Lee-Carter, se utilizaron métodos como el modelo de ogiva generalizado de Coale y Demey (1967) y variantes de los métodos de Brass.

3- Mortalidad infantil: en el caso de la mortalidad infantil y juvenil, mencionan que hicieron énfasis especial en cómo estimar la mortalidad infantil y niños pequeños, por ejemplo tasas como 5Q0 y 1Q0 en aquellos países en donde los datos que se disponen son limitados.

4. Mortality forecasting at age 65 and above: an age-specific evaluation of the Lee-Carter model

Lee, R. D., & Carter, L. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659-671. <https://doi.org/10.1080/03461238.2021.1928542>

Mortality forecasting at age 65 and above: an age-specific evaluation of the Lee-Carter model

Este artículo analiza de manera profunda la precisión, el sesgo y la robustez del modelo Lee-Carter (LC) para la proyección de mortalidad a partir de los 65 años, evaluando su desempeño en cuatro países mediante un enfoque fuera de la muestra que se utiliza. El estudio muestra que el modelo LC es adecuado cuando la Tasa de Mejora de Mortalidad por Edad (ASRMI) es constante, pero señala que las ASRMI rara vez son constantes durante la práctica, lo que limita la precisión del modelo en la mayoría de los casos.

5. Mortality Modelling and forecasting: a review of methods

Mortality Modelling and forecasting: a review of methods

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. Actuarial Association of Australia. Retrieved from https://www.actuaries.org/CTTEES_TFM/Documents/Booth_Tickle.pdf

Este artículo proporciona una revisión exhaustiva de los métodos de modelado y pronóstico de mortalidad, destacando las técnicas más comunes que han sido utilizadas por los actuarios y demógrafos para pronosticar las tasas de mortalidad y la esperanza de vida a través del tiempo.

Dentro de los métodos de modelado de mortalidad, se exploran varias técnicas de modelado, una de esas técnicas es el modelo de Lee-Carter. Se discute la ventaja de este modelo por su simplicidad y su flexibilidad para capturar las tendencias de mejora de mortalidad a través del tiempo.

También se puede observar que se habla sobre las limitaciones de este modelo, este modelo no es capaz de captar los cambios abruptos de la mortalidad.

4. Análisis exploratorio de datos

A continuación, se presentarán las librerías de R necesarias para el desarrollo de este proyecto. Estas herramientas facilitarán la manipulación y el análisis de datos, así como la generación de resultados y visualizaciones de manera eficiente.

```
library(corrplot)
library(pacman)
library(demography)
library(ggplot2)
library(reshape2)
library(tidyverse)
library(dplyr)
library(skimr)
```

Después de haber cargado las librerías, procedimos a cargar la base de datos correspondiente a la población masculina, comprendida entre los años 1985 y 2005. Esta base de datos fue obtenida de la *Latin Mortality Database (Latin Mortality Database, n.d.)* recuperada de https://www.ssc.wisc.edu/cdha/latinmortality2/?page_id=28.

```
datos_lc5 <- read.table("Mex_mlt piv_5x1.txt", header = T, sep = "")
```

Para comenzar el análisis de nuestra base de datos, lo primero que debemos hacer es examinar su estructura, lo que nos permitirá saber las variables que contiene, su tipo y la forma en que se organizan.

```
str(datos_lc5)
```

```
## 'data.frame':    342 obs. of  10 variables:
##  $ ctry      : chr  "Mexico" "Mexico" "Mexico" "Mexico" ...
##  $ ctry.code : int   2310 2310 2310 2310 2310 2310 2310 2310 2310 2310 ...
##  $ year      : int   1895 1895 1895 1895 1895 1895 1895 1895 1895 1895 ...
##  $ age       : int    0  1  5 10 15 20 25 30 35 40 ...
##  $ mx        : num   0.30741 0.23314 0.0136 0.00742 0.00989 ...
##  $ qx        : num   0.2646 0.208 0.0657 0.0364 0.0483 ...
##  $ lx        : int  100000 73535 58242 54417 52435 49903 46862 43773 40780 37773 ...
##  $ Lx        : int   82268 253645 280694 267329 256098 242217 226896 211681 196684 1810 ...
##  $ Tx        : int  2872238 2789970 2536324 2255631 1988302 1732204 1489987 1263091 10 ...
##  $ ex        : num   28.7 37.9 43.5 41.5 37.9 ...
```

Podemos notar que tenemos 10 variables, de las cuales solo “ctry” es de tipo carácter, las demás son numéricas. Es indispensable explicar que significa cada variable:

-**mx**: tasa de mortalidad a edad x

-**qx**: probabilidad de muerte a edad x

- lx**: número de sobrevivientes a edad x
- Lx**: cantidad de años vividos entre las edades x y x+1
- Tx**: años totales vividos
- ex**: esperanza de vida a edad x

Luego, es fundamental describir nuestras variables: para las numéricas, usaremos `summary()`, la cual nos dará estadísticas básicas (mínimo, máximo, media, etc.). Estos pasos iniciales nos permiten asegurarnos de que los datos están en orden y listos para un mejor análisis.

```
summary(datos_lc5)
```

```
##      ctry      ctry.code      year      age
## Length:342      Min.   :2310      Min.   :1895      Min.   : 0.00
## Class :character 1st Qu.:2310      1st Qu.:1910      1st Qu.:15.00
## Mode  :character Median :2310      Median :1932      Median :40.00
##                      Mean  :2310      Mean   :1939      Mean   :40.32
##                      3rd Qu.:2310      3rd Qu.:1965      3rd Qu.:65.00
##                      Max.   :2310      Max.    :2005      Max.    :85.00
##      mx      qx      lx      Lx
## Min.   :0.00032 Min.   :0.00162 Min.   : 447 Min.   : 1072
## 1st Qu.:0.01065 1st Qu.:0.04892 1st Qu.: 27385 1st Qu.: 98900
## Median :0.02164 Median :0.09072 Median : 49999 Median :217112
## Mean   :0.07351 Mean   :0.21238 Mean   : 50492 Mean   :216807
## 3rd Qu.:0.09787 3rd Qu.:0.25204 3rd Qu.: 74606 3rd Qu.:300476
## Max.   :0.41755 Max.   :1.00000 Max.   :100000 Max.   :487300
##      Tx      ex
## Min.   : 1072 Min.   : 2.39
## 1st Qu.: 335332 1st Qu.:10.77
## Median :1275638 Median :25.78
## Mean   :1749231 Mean   :26.29
## 3rd Qu.:2750162 3rd Qu.:38.00
## Max.   :7154615 Max.   :72.10
```

Para este análisis, hemos optado por conservar únicamente las variables que consideramos de interés, con el fin de enfocar nuestro estudio en aquellos aspectos fundamentales de la base de datos. Al seleccionar cuidadosamente estas variables, buscamos asegurar que cada una aporte información relevante y contribuya de manera directa a la interpretación precisa de los resultados que obtendremos más adelante.

```
datos_lc2005 <- select(datos_lc5, year, age, mx, qx, lx, Lx, Tx, ex)
```

Nos hemos quedado únicamente con **year**, **age**, **mx**, **qx**, **lx**, **Lx**, **Tx**, **ex**.

Hasta este punto, hemos constatado que los datos disponibles abarcan el período comprendido entre 1895 y 2005. Con el objetivo de ampliar nuestro análisis temporal, procederemos a incorporar información adicional, concretamente, datos correspondientes al intervalo de 2005

a 2021. Esta ampliación se fundamenta en la información proporcionada en la fuente *Tablas de vida (UISP, n.d.)*, disponible en <https://uisp.insp.mx/wp/index.php/tablas-de-vida/>.

```
# 2006 -----
tablas2006general <- read.csv("2006.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df3 <- tablas2006general[-1, ]

# 2007 -----
tablas2007general <- read.csv("2007_1.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df4 <- tablas2007general[-1, ]

# 2008 -----
tablas2008general <- read.csv("2008.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df5 <- tablas2008general[-1, ]

# 2009 -----
tablas2009general <- read.csv("2009.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df6 <- tablas2009general[-1, ]

# 2010 -----
tablas2010general <- read.csv("2010.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df7 <- tablas2010general[-1, ]

# 2011 -----
tablas2011general <- read.csv("2011.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df8 <- tablas2011general[-1, ]

# 2012 -----
tablas2012general <- read.csv("2012.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df9 <- tablas2012general[-1, ]

# 2013 -----
tablas2013general <- read.csv("2013.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
```

```

df10 <- tablas2013general[-1, ]

# 2014 -----

tablas2014general <- read.csv("2014.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df11 <- tablas2014general[-1, ]

# 2015 -----

tablas2015general <- read.csv("2015.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df12 <- tablas2015general[-1, ]

# 2016 -----

tablas2016general <- read.csv("2016.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df13 <- tablas2016general[-1, ]

# 2017 -----

tablas2017general <- read.csv("2017.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df14 <- tablas2017general[-1, ]

# 2018 -----

tablas2018general <- read.csv("2018.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df15 <- tablas2018general[-1, ]

# 2019 -----

tablas2019general <- read.csv("2019.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df16 <- tablas2019general[-1, ]

# 2020 -----

tablas2020general <- read.csv("2020.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df17 <- tablas2020general[-1, ]

```



```
# 2021 -----

tablas2021general <- read.csv("2021.csv", sep="\t", header=TRUE, fileEncoding="UTF-16LE")
df18 <- tablas2021general[-1, ]
```

Procederemos ahora a seleccionar únicamente aquellas variables que coincidan con las presentes en el data frame correspondiente a los datos de 1895 a 2005.

```
datos_lc2006 <- select(df3, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2007 <- select(df4, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2008 <- select(df5, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2009 <- select(df6, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2010 <- select(df7, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2011 <- select(df8, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2012 <- select(df9, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2013 <- select(df10, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2014 <- select(df11, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2015 <- select(df12, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2016 <- select(df13, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2017 <- select(df14, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2018 <- select(df15, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2019 <- select(df16, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2020 <- select(df17, year, age, mx, qx, lx, Lx, Tx, ex )
datos_lc2021 <- select(df18, year, age, mx, qx, lx, Lx, Tx, ex )
```

Realizamos el merge con el objetivo de integrar las bases de datos correspondientes a periodos posteriores. Es importante señalar que esta fusión se llevará a cabo de manera progresiva, combinando los conjuntos de datos de forma secuencial y por pares.

```
merge1 <- merge(datos_lc2005, datos_lc2006, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge2 <- merge(merge1, datos_lc2007, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge3 <- merge(merge2, datos_lc2008, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge4 <- merge(merge3, datos_lc2009, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge5 <- merge(merge4, datos_lc2010, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge6 <- merge(merge5, datos_lc2011, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge7 <- merge(merge6, datos_lc2012, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge8 <- merge(merge7, datos_lc2013, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
merge9 <- merge(merge8, datos_lc2014, by=c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex"))
```

```

merge10 <- merge(merge9, datos_lc2015, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge11 <- merge(merge10, datos_lc2016, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge12 <- merge(merge11, datos_lc2017, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge13 <- merge(merge12, datos_lc2018, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge14 <- merge(merge13, datos_lc2019, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge15 <- merge(merge14, datos_lc2020, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",
merge16 <- merge(merge15, datos_lc2021, by=c("year","age", "mx", "qx", "lx", "Lx", "Tx",

```

Definimos la base final “base_final_hombres”, la cual ya integra los datos correspondientes al periodo comprendido entre 1895 y 2021.

```
base_final_hombres <- merge16
```

Durante el proceso de revisión y exploración de los datos, detectamos que algunos registros presentaban comas en lugares inesperados, lo que alteraba el formato numérico adecuado. Este inconveniente afectaba el procesamiento correcto de la información, pudiendo distorsionar nuestros cálculos y análisis posteriores.

```

# Verificacion de los datos
columnas_con_coma <- sapply(base_final_hombres, function(x) any(grepl(",",x)))
print(columnas_con_coma)

```

```

## year age mx qx lx Lx Tx ex
## FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE

```

```

# Eliminamos las comas
base_final_hombres$lx <- gsub(",", "", base_final_hombres$lx)

# Detectamos la presencia de mas comas en columnas Tx y ex
# Columna Tx
base_final_hombres$Tx <- gsub(",", "", base_final_hombres$Tx)

# Columna ex
base_final_hombres$Lx <- gsub(",", "", base_final_hombres$Lx)

```

Una vez concluido el proceso de limpieza de comas, procederemos a examinar detenidamente la estructura de la base de datos, verificando que cada uno de sus componentes se encuentre organizado según lo esperado. Posteriormente, convertiremos todas nuestras variables al tipo numérico

```
base_final_hombres$year <- as.numeric(base_final_hombres$year)
base_final_hombres$age <- as.numeric(base_final_hombres$age)
base_final_hombres$mx <- as.numeric(base_final_hombres$mx)
base_final_hombres$qx <- as.numeric(base_final_hombres$qx)
base_final_hombres$lx <- as.numeric(base_final_hombres$lx)
base_final_hombres$Lx <- as.numeric(base_final_hombres$Lx)
base_final_hombres$Tx <- as.numeric(base_final_hombres$Tx)
base_final_hombres$ex <- as.numeric(base_final_hombres$ex)
```

Procedamos a realizar un análisis de correlación lineal para investigar si se establecen relaciones significativas entre las variables de nuestro conjunto de datos.

```
cor(base_final_hombres[, c("year", "age", "mx", "qx", "lx", "Lx", "Tx", "ex")])
```

```
##           year           age           mx           qx           lx           Lx
## year  1.000000e+00 -5.232687e-20 -0.3927207 -0.2375079  0.6373840  0.6345414
## age  -5.232687e-20  1.000000e+00  0.3934822  0.6774502 -0.6918712 -0.4574899
## mx   -3.927207e-01  3.934822e-01  1.0000000  0.8202217 -0.5564214 -0.6771687
## qx   -2.375079e-01  6.774502e-01  0.8202217  1.0000000 -0.7457807 -0.6875409
## lx    6.373840e-01 -6.918712e-01 -0.5564214 -0.7457807  1.0000000  0.8236354
## Lx    6.345414e-01 -4.574899e-01 -0.6771687 -0.6875409  0.8236354  1.0000000
## Tx    4.667957e-01 -8.272127e-01 -0.4266558 -0.5874758  0.8411394  0.6327994
## ex    3.281810e-01 -9.104004e-01 -0.4947681 -0.6558508  0.8004225  0.6162034
##           Tx           ex
## year  0.4667957  0.3281810
## age  -0.8272127 -0.9104004
## mx   -0.4266558 -0.4947681
## qx   -0.5874758 -0.6558508
## lx    0.8411394  0.8004225
## Lx    0.6327994  0.6162034
## Tx    1.0000000  0.9749885
## ex    0.9749885  1.0000000
```

Podemos observar que existen relaciones signifativas. Por ejemplo entre mx y qx , lx y Lx , Tx y lx , Tx y Lx , Tx y ex , entre otros.

Para el caso de las mujeres, se aplicó el mismo procedimiento. Con el fin de mantener la brevedad y evitar extendernos innecesariamente, a continuación se presenta únicamente el código utilizado junto con sus respectivos resultados.

```
# LIMPIEZA DE LA BASE DE DATOS
# Leemos los de intervalos de 5
datos_lc5_m <- read.table("Mex_flt piv_5x1_mujeres.txt", header = T, sep = "")

# Nos quedamos con las variables de nuestro interes
datos_lc2005_m <- select(datos_lc5_m, year, age, mxm, qxm, lxm, Lxm, Txm, exm)
```

NOTA:

Hasta aqui podemos ver que tenemos los datos desde 1895-2005

Continuamos con las demas tablas

```
datos_lc2006_m <- select(df3, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2007_m <- select(df4, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2008_m <- select(df5, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2009_m <- select(df6, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2010_m <- select(df7, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2011_m <- select(df8, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2012_m <- select(df9, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2013_m <- select(df10, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2014_m <- select(df11, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2015_m <- select(df12, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2016_m <- select(df13, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2017_m <- select(df14, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2018_m <- select(df15, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2019_m <- select(df16, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2020_m <- select(df17, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
datos_lc2021_m <- select(df18, year, age, mxm, qxm, lxm, Lxm, Txm, exm )
```

Realizamos el merge

Esto es con la finalidad de ir juntando las bases de datos de periodos posteriores

Debemos de ir haciendolo por pares

```
merge1m <- merge(datos_lc2005_m, datos_lc2006_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge2m <- merge(merge1m, datos_lc2007_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge3m <- merge(merge2m, datos_lc2008_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge4m <- merge(merge3m, datos_lc2009_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge5m <- merge(merge4m, datos_lc2010_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge6m <- merge(merge5m, datos_lc2011_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge7m <- merge(merge6m, datos_lc2012_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge8m <- merge(merge7m, datos_lc2013_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge9m <- merge(merge8m, datos_lc2014_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge10m <- merge(merge9m, datos_lc2015_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
merge11m <- merge(merge10m, datos_lc2016_m, by=c("year", "age", "mxm", "qx", "lx", "Lx", "Tx", "ex"))
```

```

merge12m <- merge(merge11m, datos_lc2017_m, by=c("year","age", "mxm", "qxm", "lxm", "Lxm", "Txm", "exm"))
merge13m <- merge(merge12m, datos_lc2018_m, by=c("year","age", "mxm", "qxm", "lxm", "Lxm", "Txm", "exm"))
merge14m <- merge(merge13m, datos_lc2019_m, by=c("year","age", "mxm", "qxm", "lxm", "Lxm", "Txm", "exm"))
merge15m <- merge(merge14m, datos_lc2020_m, by=c("year","age", "mxm", "qxm", "lxm", "Lxm", "Txm", "exm"))
merge16m <- merge(merge15m, datos_lc2021_m, by=c("year","age", "mxm", "qxm", "lxm", "Lxm", "Txm", "exm"))

##Nombramos base final hombres
base_final_mujeres <- merge16m

# Verificacion de los datos
columnas_con_coma <- sapply(base_final_mujeres, function(x) any(grepl(",",x)))
print(columnas_con_coma)

```

```

## year age mxm qxm lxm Lxm Txm exm
## FALSE FALSE FALSE FALSE TRUE TRUE TRUE FALSE

```

```

# Eliminamos las comas
base_final_mujeres$lxm <- gsub(",", "", base_final_mujeres$lxm)

# Detectamos la presencia de mas comas en columnas Tx y ex
# Columna Tx
base_final_mujeres$Txm <- gsub(",", "", base_final_mujeres$Txm)

# Columna ex
base_final_mujeres$Lxm <- gsub(",", "", base_final_mujeres$Lxm)

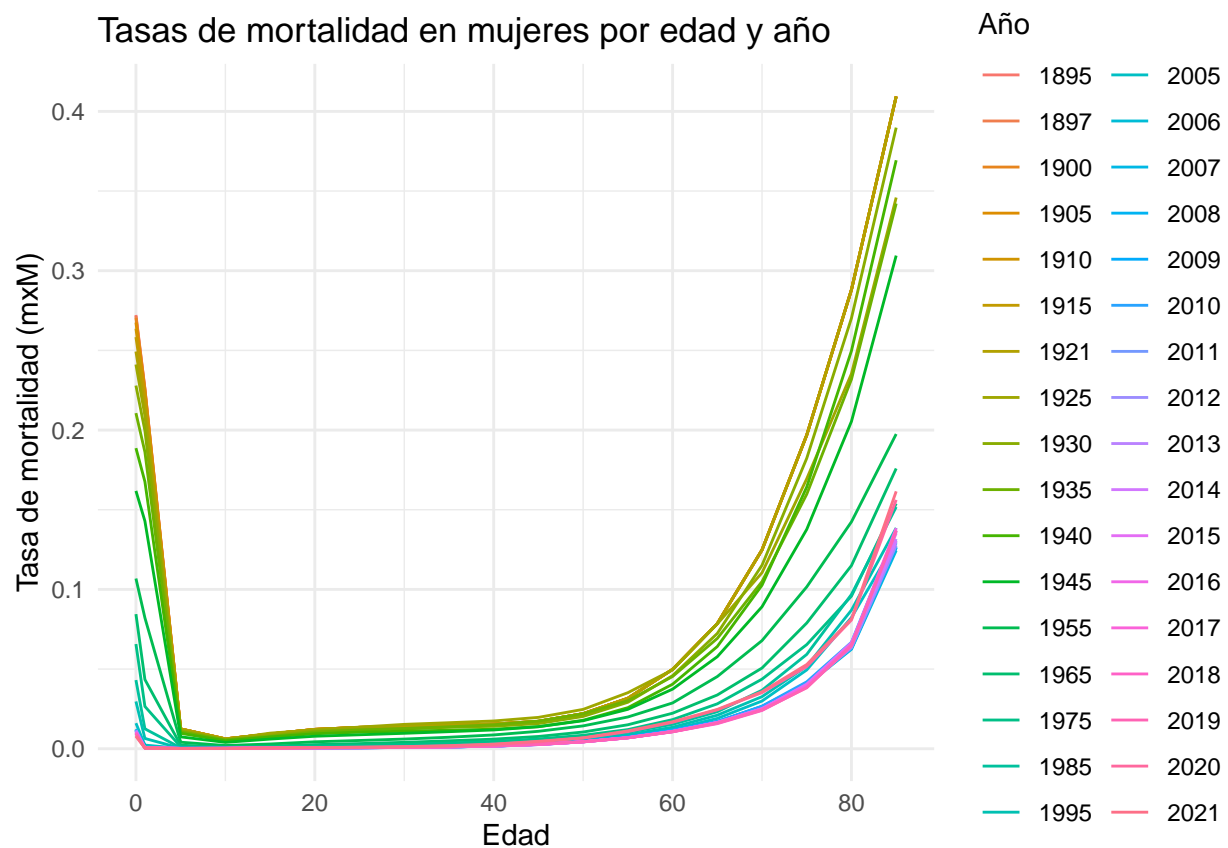
#Vamos a pasar las variables a numericas
base_final_mujeres$year <- as.numeric(base_final_mujeres$year)
base_final_mujeres$age <- as.numeric(base_final_mujeres$age)
base_final_mujeres$mxm <- as.numeric(base_final_mujeres$mxm)
base_final_mujeres$qxm <- as.numeric(base_final_mujeres$qxm)
base_final_mujeres$lxm <- as.numeric(base_final_mujeres$lxm)
base_final_mujeres$Lxm <- as.numeric(base_final_mujeres$Lxm)
base_final_mujeres$Txm <- as.numeric(base_final_mujeres$Txm)
base_final_mujeres$exm <- as.numeric(base_final_mujeres$exm)

```

Para culminar el punto 4, presentamos de forma gráfica las tasas de mortalidad en función del factor de edad y año, tanto para hombres como para mujeres. Con este análisis visual, buscamos identificar patrones y tendencias que reflejen la evolución de estas tasas a lo largo del tiempo.

```
# TASAS DE MORTALIDAD MUJERES -----
```

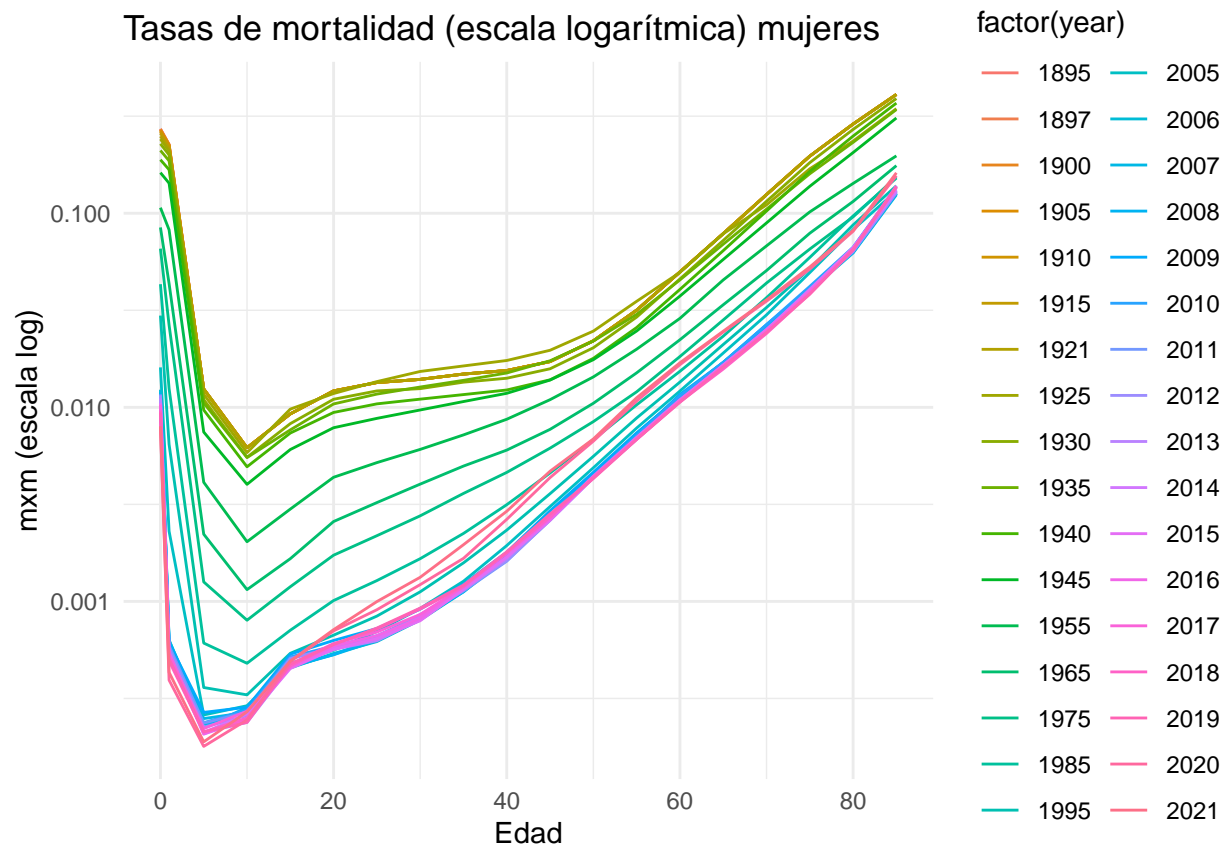
```
library(ggplot2)
ggplot(base_final_mujeres, aes(x = age, y = mxm, group = factor(year), color = factor(year))) +
  geom_line() +
  labs(
    title = "Tasas de mortalidad en mujeres por edad y año",
    x = "Edad",
    y = "Tasa de mortalidad (mxM)",
    color = "Año"
  ) +
  theme_minimal()
```



```
# Segunda grafica
```

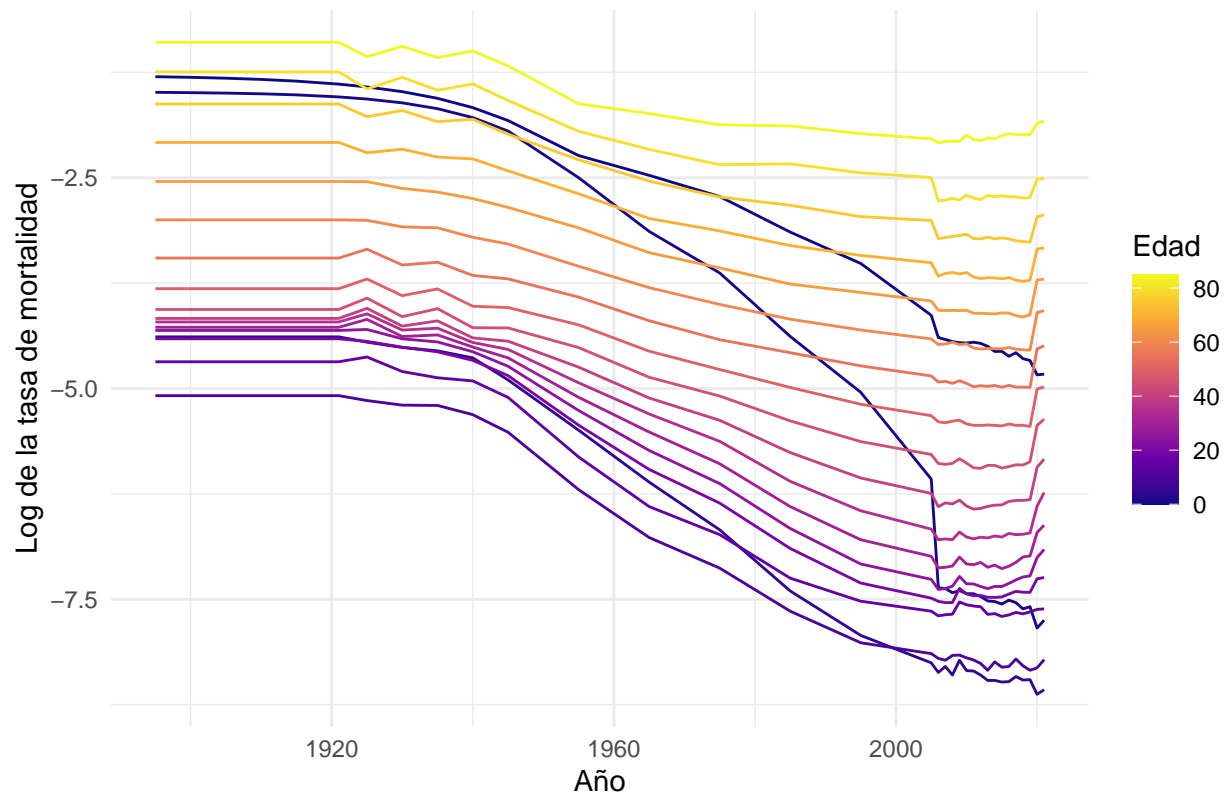
```
ggplot(base_final_mujeres, aes(x = age, y = mxm, group = year, color = factor(year))) +
  geom_line() +
  scale_y_log10() +
  labs(
    title = "Tasas de mortalidad (escala logarítmica) mujeres",
    x = "Edad",
    y = "mxm (escala log)"
  )
```

```
) +  
theme_minimal()
```



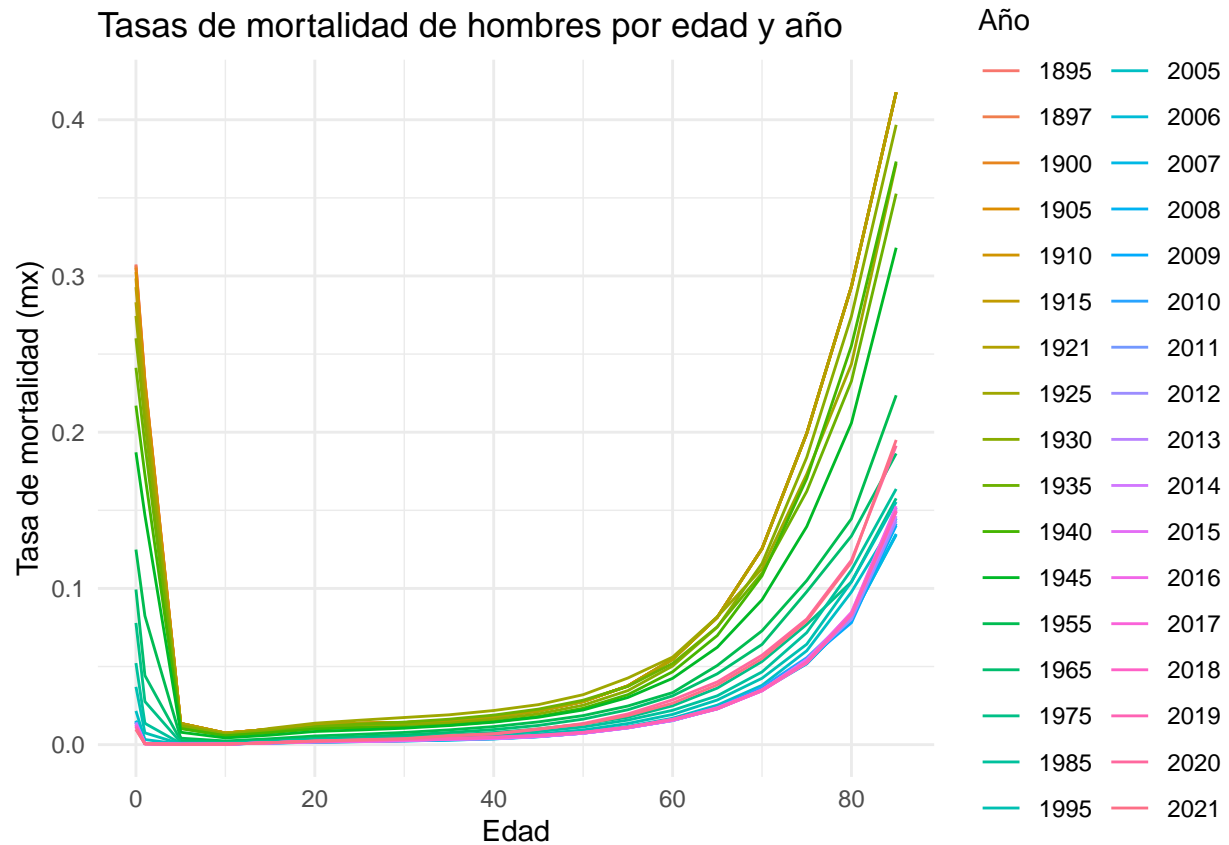
```
# Tercera gráfica  
ggplot(base_final_mujeres, aes(x = year,  
                                y = log(mxm),          # o bien scale_y_log10()  
                                group = age,  
                                color = age)) +  
  
  geom_line() +  
  # Escala continua de color (puedes usar otras paletas)  
  scale_color_viridis_c(option = "plasma") + # o "magma", "inferno", etc.  
  labs(  
    title = "Tasas de mortalidad de mujeres (log) por año y edad",  
    x = "Año",  
    y = "Log de la tasa de mortalidad",  
    color = "Edad"  
  ) +  
  theme_minimal()
```

Tasas de mortalidad de mujeres (log) por año y edad



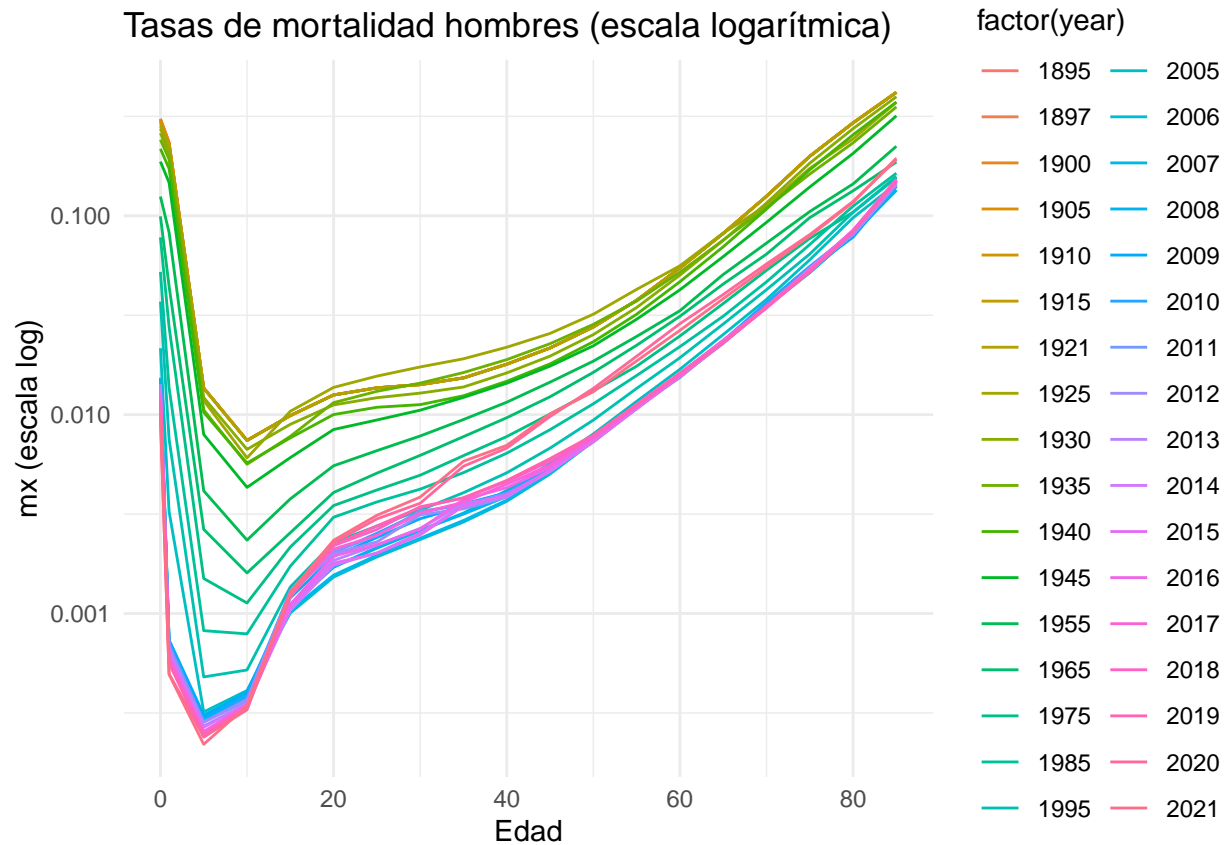
TASAS DE MORTALIDAD HOMBRES

```
library(ggplot2)
ggplot(base_final_hombres, aes(x = age, y = mx, group = factor(year), color = factor(year))) +
  geom_line() +
  labs(
    title = "Tasas de mortalidad de hombres por edad y año",
    x = "Edad",
    y = "Tasa de mortalidad (mx)",
    color = "Año"
  ) +
  theme_minimal()
```

Segunda grafica

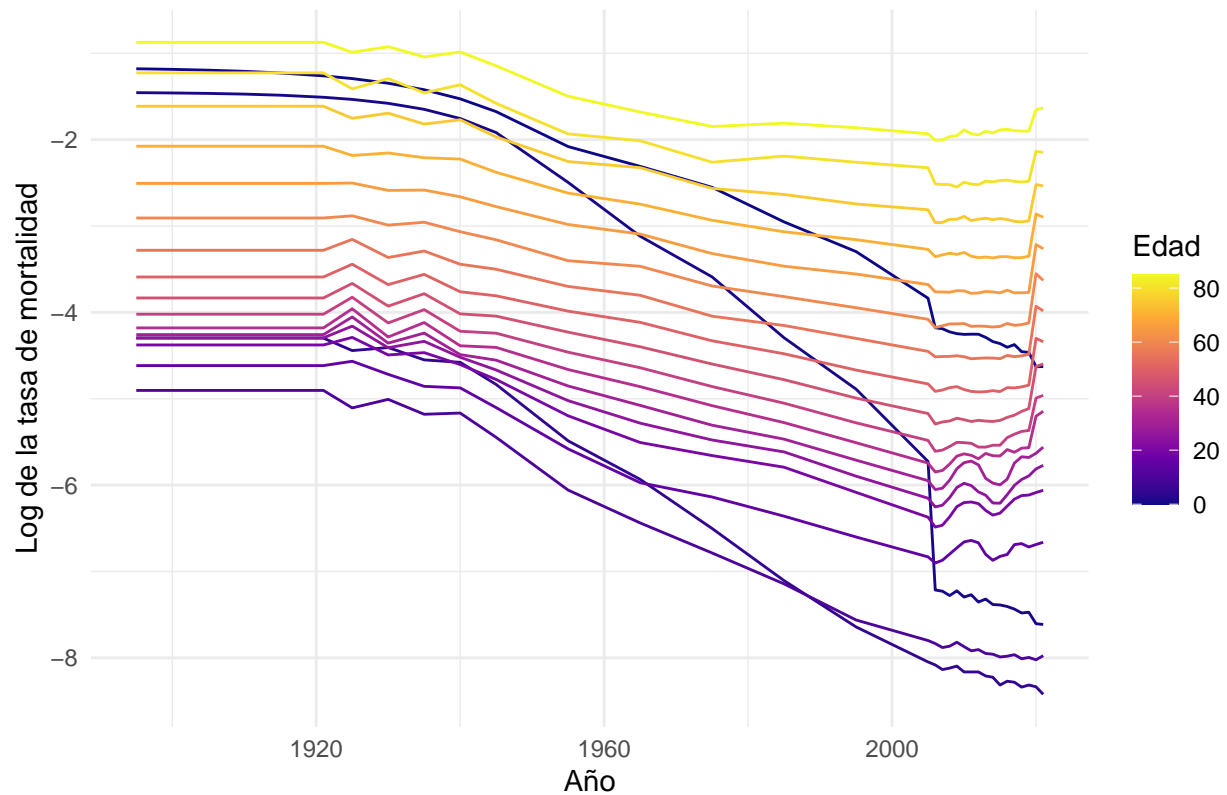
```
ggplot(base_final_hombres, aes(x = age, y = mx, group = year, color = factor(year))) +
  geom_line() +
  scale_y_log10() +
  labs(
    title = "Tasas de mortalidad hombres (escala logarítmica)",
    x = "Edad",
    y = "mx (escala log)"
  ) +
  theme_minimal()
```



```
# Tercera gráfica
ggplot(base_final_hombres, aes(x = year,
                              y = log(mx),          # o bien scale_y_log10()
                              group = age,
                              color = age)) +

  geom_line() +
  # Escala continua de color (puedes usar otras paletas)
  scale_color_viridis_c(option = "plasma") + # o "magma", "inferno", etc.
  labs(
    title = "Tasas de mortalidad de hombres (log) por año y edad",
    x = "Año",
    y = "Log de la tasa de mortalidad",
    color = "Edad"
  ) +
  theme_minimal()
```

Tasas de mortalidad de hombres (log) por año y edad



5. Metodología

A continuación, se describe la metodología empleada para el análisis de datos y la construcción del modelo de mortalidad:

1. Recopilación y fuente de datos

- Se reunieron registros históricos de mortalidad, específicamente de Latin Mortality Database (n.d.) y Tablas de vida (UISP, n.d.).
- Estos conjuntos de datos abarcan distintos periodos, desde finales del siglo XIX hasta inicios del siglo XXI, y contienen información relativa a tasas de mortalidad, número de sobrevivientes y años vividos, tanto para hombres como para mujeres.

2. Limpieza y preparación de la información

- Se revisó la estructura de los datos con funciones como `str()`, `head()` y `summary()` para identificar inconsistencias o errores de formato.
- Se corrigieron valores no válidos y se eliminaron o reemplazaron caracteres extra (por ejemplo, comas en campos numéricos). Además, se convirtieron las variables a tipos apropiados (numéricas, factores, etc.).

- Se unificaron las bases de datos provenientes de distintos periodos (por ejemplo, 1895-2005 y 2005-2013) mediante un proceso de **merge**, asegurando la correspondencia de las variables clave (edades, años y tasas de mortalidad).

3. Selección de variables relevantes

- Se identificaron las variables de mayor interés para el análisis actuarial, como m_x (tasa de mortalidad), l_x (sobrevivientes a la edad x), L_x (años vividos entre edades x y $x + 1$) y T_x (suma acumulada de años vividos).
- Esta depuración de variables facilitó la construcción de las matrices requeridas para el modelo.

4. Transformación de datos en formato demográfico

- Con ayuda de funciones del paquete **reshape2** (por ejemplo, **acast**), se convirtieron los data frames en matrices, donde las filas corresponden a edades y las columnas a años.
- Estas matrices (de tasas, sobrevivientes y años vividos) se transformaron en objetos de la clase **demogdata** mediante la función **demogdata** del paquete **demography**, especificando los vectores de edades, años y el tipo de dato (mortality).

5. Construcción del modelo (Lee-Carter)

- Se aplicó el modelo **Lee-Carter** utilizando la función **lca()** del paquete **demography**.
- El modelo se ajustó en dos configuraciones diferentes, aprovechando tanto la matriz de L_x como la de l_x para comparar y contrastar los resultados.
- Se emplearon los comandos **summary()** y **plot()** para evaluar la calidad del ajuste y visualizar las tendencias de la mortalidad a lo largo del tiempo.

6. Software y herramientas empleadas

- **Paquetes en R:**
 - **demography** para la creación de objetos demográficos y aplicación del modelo Lee-Carter.
 - **reshape2** (o **tidyr**) para la manipulación y reestructuración de datos.
 - **dplyr**, **ggplot2**, **readr** y otros paquetes auxiliares para la limpieza, transformación y visualización de la información.

Esta metodología aseguró la coherencia de los datos y la aplicación adecuada del modelo Lee-Carter para el estudio de la mortalidad a lo largo de distintos periodos históricos. A continuación se explicará como se utilizó cada línea de código para el caso de la base de datos de hombres.

```
mx_matrix <- acast(base_final_hombres, age ~ year, value.var = "mx")
# Conviértela a matriz numérica si no lo es:
mx_matrix <- as.matrix(mx_matrix) # Asegúrate de que sea 'matrix' y no 'table'
mx_matrix <- apply(mx_matrix, 2, as.numeric)
base_final_hombres$mx <- as.numeric(base_final_hombres$mx)
```

Creamos una matriz correspondiente a la tasa de mortalidad. En esta matriz, las filas representan las edades y las columnas los años. Cada celda contiene el valor de la tasa de mortalidad (mx) correspondiente a esa combinación

```
Lx_matrix <- acast(base_final_hombres, age ~ year, value.var = "Lx")
Lx_matrix <- as.matrix(Lx_matrix)
Lx_matrix <- apply(Lx_matrix, 2, as.numeric)
str(Lx_matrix)
```

```
## num [1:19, 1:34] 82268 253645 280694 267329 256098 ...
## - attr(*, "dimnames")=List of 2
## ..$ : NULL
## ..$ : chr [1:34] "1895" "1897" "1900" "1905" ...
```

Ahora transformamos la información correspondiente a L_x en una matriz organizada por edad y año, y de asegurar que todos los datos sean numéricos. Para comprobar la transformación, se muestra la estructura de la matriz resultante.

```
lx_matrix <- acast(base_final_hombres, age ~ year, value.var = "lx")
lx_matrix <- as.matrix(lx_matrix)
lx_matrix <- apply(lx_matrix, 2, as.numeric)
str(lx_matrix)
```

```
## num [1:19, 1:34] 100000 73535 58242 54417 52435 ...
## - attr(*, "dimnames")=List of 2
## ..$ : NULL
## ..$ : chr [1:34] "1895" "1897" "1900" "1905" ...
```

Por ultimo, convertimos los datos de l_x en una matriz, estructurada de manera que cada fila representa una edad y cada columna un año. Se aplica la conversión a formato numérico, y luego se utiliza `str()` para revisar la estructura resultante de la matriz.

```
my_data <- demogdata(
  data = mx_matrix,          # Matriz de tasas (o defunciones)
  pop = Lx_matrix,
  ages = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85),
  years = c(1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985),
  type = "mortality", # Tipo de dato
  label = "Mis datos",
```

```

    name = "Hombres"
)

```

Las anteriores líneas de código crean un objeto demográfico llamado `my_data` utilizando la función `demogdata`. Básicamente, se establecen los elementos esenciales para el análisis, incluyendo la matriz de tasas de mortalidad, la población (o medida relacionada), y los vectores de edades y años. Además, se especifica el tipo de dato (en este caso, “mortality”) y se asignan etiquetas para identificar el conjunto de datos. Este objeto se utiliza posteriormente para realizar análisis demográficos basados en L_x .

```

# Con matriz lx
my_data2 <- demogdata(
  data = mx_matrix,          # Matriz de tasas (o defunciones)
  pop = lx_matrix,
  ages = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85),
  years = c(1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985),
  type = "mortality",      # Tipo de dato
  label = "Mis datos",
  name = "Hombres"
)

```

El siguiente bloque de código crea un objeto demográfico llamado `my_data2` utilizando la función `demogdata`. En este caso, se utiliza la matriz de tasas de mortalidad (`mx_matrix`) y la matriz `lx_matrix` para la población. Se establecen los vectores de edades y años, se define el tipo de dato (“mortality”) y se asignan etiquetas al conjunto de datos. Este objeto servirá para llevar a cabo análisis demográficos utilizando la información basada en la variable l_x .

```

# Realizamos el ajuste
fit_lc <- lca(my_data, series = "Hombres", adjust = "e0")
summary(fit_lc)

```

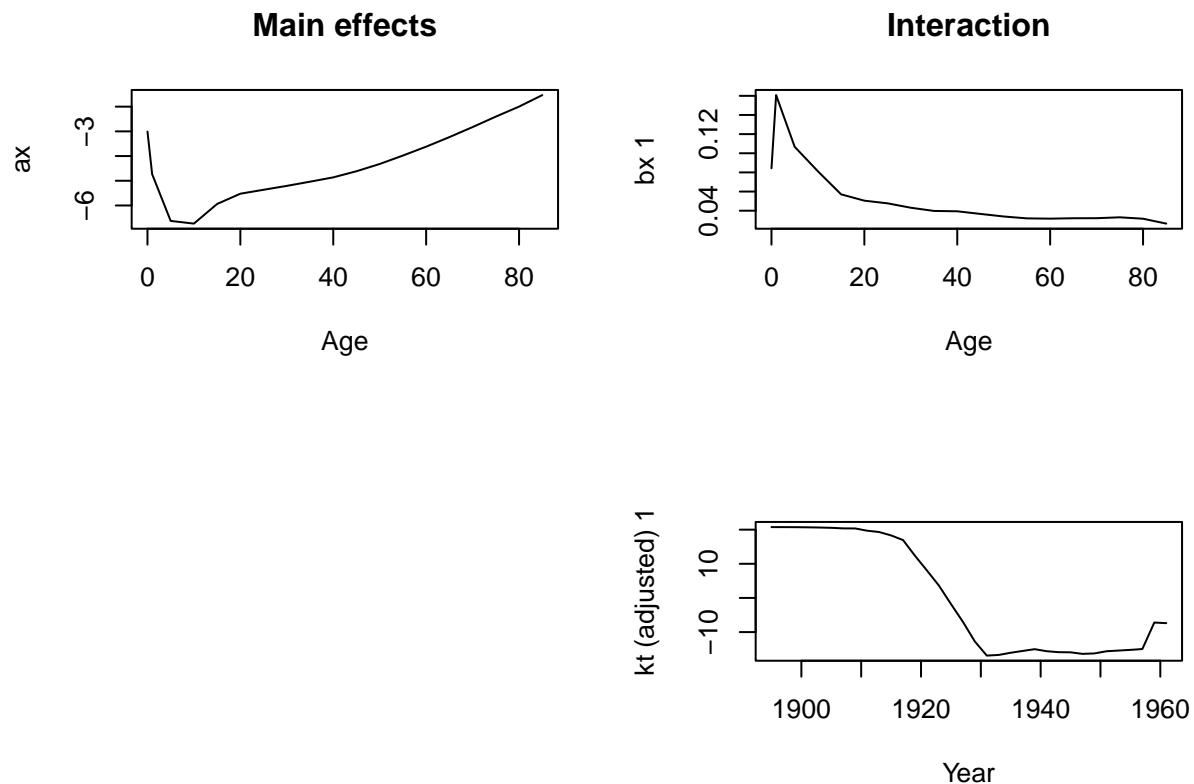
```

## Lee-Carter analysis
##
## Call: lca(data = my_data, series = "Hombres", adjust = "e0")
##
## Adjustment method: e0
## Region: Mis datos
## Years in fit: 1895 - 2021
## Ages in fit: 0 - 85
##
## Percentage variation explained: 97.8%
##
## ERROR MEASURES BASED ON MORTALITY RATES
##
## Averages across ages:
##      ME      MSE      MPE      MAPE

```

```
## -0.00056  0.00011  0.03198  0.12531
##
## Averages across years:
##      IE      ISE      IPE      IAPE
## -0.03518  0.00696  2.21575  9.81128
##
##
## ERROR MEASURES BASED ON LOG MORTALITY RATES
##
## Averages across ages:
##      ME      MSE      MPE      MAPE
## 0.00843  0.03643  0.00141  0.02939
##
## Averages across years:
##      IE      ISE      IPE      IAPE
## 0.66557  2.60205  0.07310  2.29206
```

```
plot(fit_lc)
```



En este bloque de código se aplica el modelo Lee-Carter para analizar las tasas de mortalidad en la serie de datos de hombres. Se realizan dos ajustes, cada uno utilizando un conjunto de datos distinto:

- **Con my_data:**
Se ajusta el modelo Lee-Carter a los datos contenidos en `my_data` y se obtienen tanto un resumen como una representación gráfica de los resultados.
- **Con my_data2:**
Se repite el proceso utilizando `my_data2`, lo que permite comparar cómo se comporta el modelo cuando se utiliza otra medida de población.

Esta comparación nos ayuda a evaluar el rendimiento y la consistencia del modelo bajo diferentes variables dentro de los datos.

El análisis para la base de datos de mujeres se realiza de forma análoga al que aplicamos para la base de datos de hombres. Dado que los procedimientos y métodos empleados son equivalentes, a continuación se presentan únicamente las líneas de código utilizadas para procesar y analizar los datos femeninos.

```
# Corremos el modelo
# Generamos las matrices

mxm_matrix <- acast(base_final_mujeres, age ~ year, value.var = "mxm")

# Conviértela a matriz numérica si no lo es:
mxm_matrix <- as.matrix(mxm_matrix) # Asegúrate de que sea 'matrix' y no 'table'
mxm_matrix <- apply(mxm_matrix, 2, as.numeric)
base_final_mujeres$mxm <- as.numeric(base_final_mujeres$mxm)

# 'apply' recorre cada columna y la convierte a numeric, devolviendo una matriz
str(mxm_matrix)
```

```
## num [1:19, 1:34] 0.27222 0.22613 0.01243 0.00619 0.00925 ...
## - attr(*, "dimnames")=List of 2
## ..$ : NULL
## ..$ : chr [1:34] "1895" "1897" "1900" "1905" ...
```

```
# Corremos las otras matrices
Lxm_matrix <- acast(base_final_mujeres, age ~ year, value.var = "Lxm")
Lxm_matrix <- as.matrix(Lxm_matrix)
Lxm_matrix <- apply(Lxm_matrix, 2, as.numeric)
str(Lxm_matrix)
```

```
## num [1:19, 1:34] 84510 263995 293708 281295 270752 ...
## - attr(*, "dimnames")=List of 2
## ..$ : NULL
## ..$ : chr [1:34] "1895" "1897" "1900" "1905" ...
```



```

lxm_matrix <- acast(base_final_mujeres, age ~ year, value.var = "lxm")
lxm_matrix <- as.matrix(lxm_matrix)
lxm_matrix <- apply(lxm_matrix, 2, as.numeric)
str(lxm_matrix)

```

```

##  num [1:19, 1:34] 100000 76169 60754 57096 55353 ...
##  - attr(*, "dimnames")=List of 2
##    ..$ : NULL
##    ..$ : chr [1:34] "1895" "1897" "1900" "1905" ...

```

Ajustamos el modelo

```

my_data_mujeres <- demogdata(
  data = mxm_matrix,          # Matriz de tasas (o defunciones)
  pop = Lxm_matrix,
  ages = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85),
  years = c(1895,1897,1900,1905,1910,1915,1921,1925,1930,1935,1940,1945,1955,1965,1975,1980,1985,1990,1995,2000,2005,2010,2015,2021),
  type = "mortality", # Tipo de dato
  label = "Mis datos",
  name = "Mujeres"
)
my_data_mujeres2 <- demogdata(
  data = mxm_matrix,          # Matriz de tasas (o defunciones)
  pop = lxm_matrix,
  ages = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85),
  years = c(1895,1897,1900,1905,1910,1915,1921,1925,1930,1935,1940,1945,1955,1965,1975,1980,1985,1990,1995,2000,2005,2010,2015,2021),
  type = "mortality", # Tipo de dato
  label = "Mis datos",
  name = "Mujeres"
)

```

Realizamos el ajuste

```

fit_lc_mujeres <- lca(my_data_mujeres, series = "Mujeres", adjust = "e0")
summary(fit_lc_mujeres)

```

```

## Lee-Carter analysis
##
## Call: lca(data = my_data_mujeres, series = "Mujeres", adjust = "e0")
##
## Adjustment method: e0
## Region: Mis datos
## Years in fit: 1895 - 2021
## Ages in fit: 0 - 85
##
## Percentage variation explained: 98.5%

```

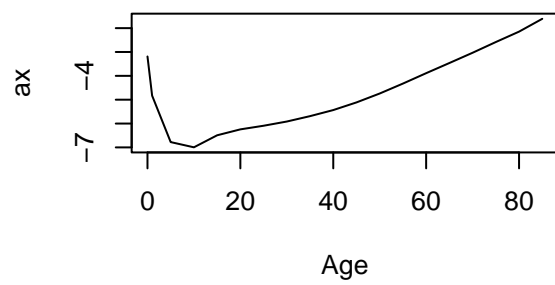
```

##
## ERROR MEASURES BASED ON MORTALITY RATES
##
## Averages across ages:
##      ME      MSE      MPE      MAPE
## -0.00054  0.00010  0.03130  0.11586
##
## Averages across years:
##      IE      ISE      IPE      IAPE
## -0.03506  0.00633  2.37309  9.37368
##
##
## ERROR MEASURES BASED ON LOG MORTALITY RATES
##
## Averages across ages:
##      ME      MSE      MPE      MAPE
## 0.01360  0.03026  0.00061  0.02705
##
## Averages across years:
##      IE      ISE      IPE      IAPE
## 1.11345  2.30286  0.01271  2.12253

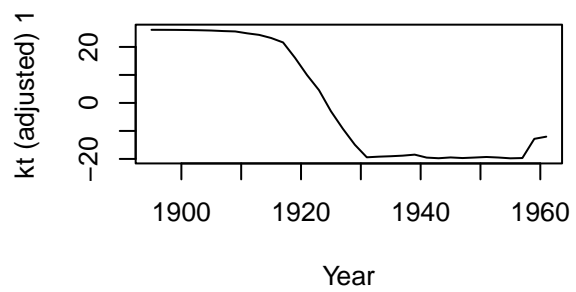
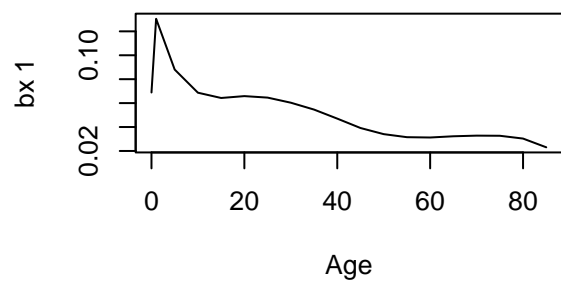
```

```
plot(fit_lc_mujeres)
```

Main effects



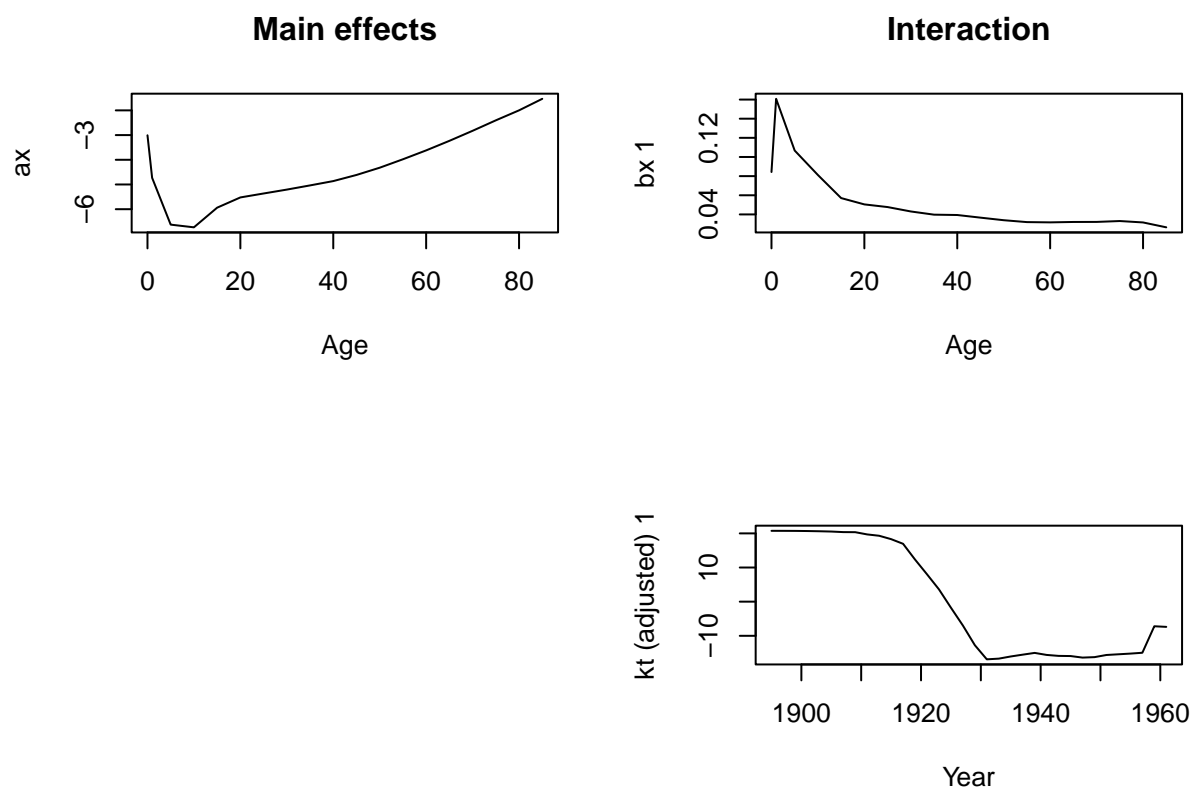
Interaction



Comparación de componentes del modelo Lee-Carter Hombres vs Mujeres

Hombres

```
plot(fit_lc)
```

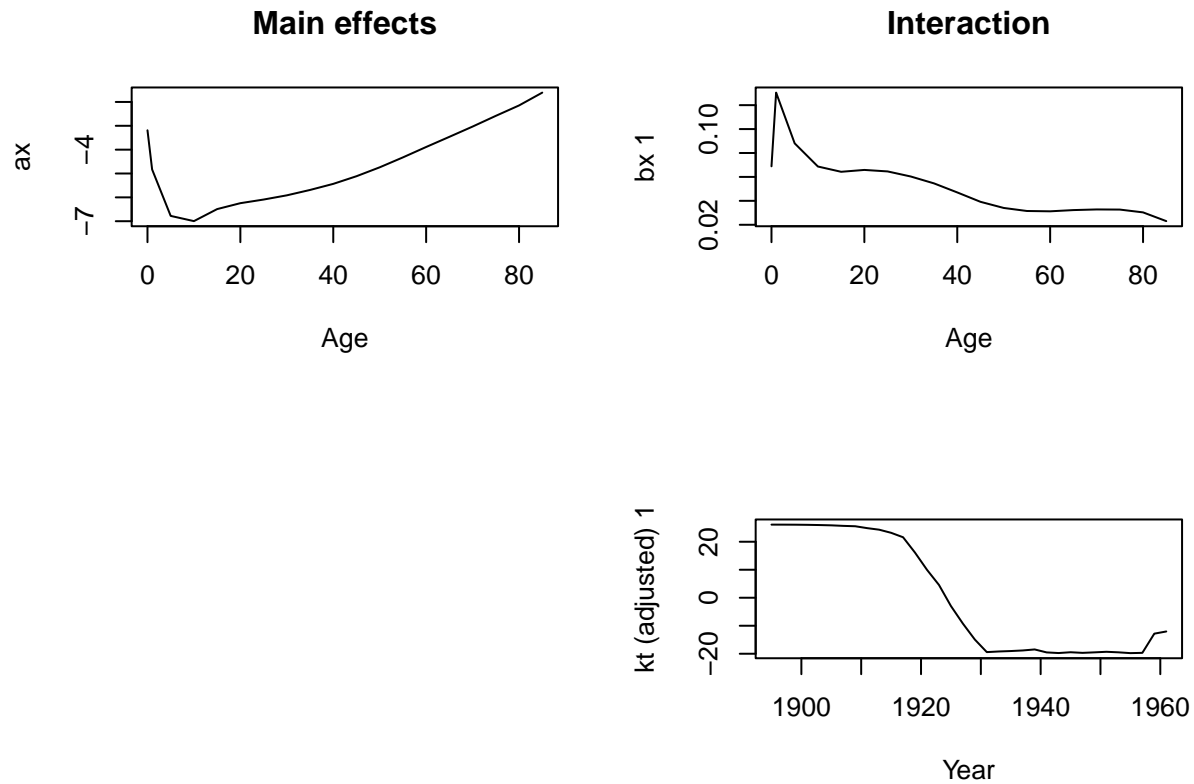


Interpretaciones

- α_x : Efectos principales. La tasa de mortalidad aumenta significativamente con la edad. La mortalidad por edad es más pronunciada conforme los hombres envejecen. Podemos observar un aumento más pronunciado en la mortalidad para los hombres de edad avanzada, en comparación con las mujeres, lo cual nos indica que las tasas de mortalidad en los hombres son más altas en edades más avanzadas.
- β_x : Interacción. La curva para los hombres indica que la mortalidad en los jóvenes ha disminuido, mientras que para las edades más avanzadas, hay un descenso en las tasas de mortalidad, lo que nos indica que son menos pronunciadas en los últimos años.
- κ_t : Tendencia temporal ajustada. Nos muestra una caída importante en las tasas de mortalidad desde a principios del siglo XX, esto nos indica una mejora en el sistema de salud pública y los avances médicos que ha tenido México.

Mujeres

```
plot(fit_lc_mujeres)
```



Interpretaciones:

- α_x : Efectos principales. La tasa de mortalidad aumenta significativamente con la edad, pero de manera menos pronunciada en comparación con la de los hombres. Las mujeres tienden a tener una tasa de mortalidad más baja.
- β_x : Interacción. La curva de mortalidad para las mujeres en uno de los periodos se suaviza, lo que nos indica que existe una tendencia más estable a través del tiempo.
- κ_x : Tendencia temporal ajustada. La tasa de mortalidad en el caso de las mujeres presenta una caída más moderada, esto nos sugiere que puede estar relacionada con factores sociales, acceso a la salud o incluso diferencias de género, entre otros factores.

Conclusiones

La diferencia en la esperanza de vida implica que las mujeres necesitarán más recursos a lo largo de su vida, ya que presentan mayor longevidad. Los sistemas de pensiones en México deberían de adaptarse a este hecho.

Sabemos que México enfrenta un envejecimiento poblacional, lo que conlleva a una mayor demanda de servicios de salud y pensiones. Es gracias a este modelo que podemos crear medidas para el cuidado de bienestar social.

6. Propuesta de Modelo

Descripción del modelo propuesto

El modelo Lee-Carter es una metodología estadística que fue desarrollada en 1992 por el demógrafo Raymond Lee y el economista Robert Carter, fue diseñada con el propósito de modelar y predecir las tasas de mortalidad de una población a lo largo del tiempo.

Este modelo se ha convertido en una herramienta esencial en la demografía, especialmente en el análisis y proyección de la mortalidad de una población a lo largo de los años. El objetivo principal es capturar y proyectar las tasas de mortalidad de una población para distintos grupos de años. Esto lo hace apoyándose de datos históricos de mortalidad y de población.

Es útil para entender tendencias de mortalidad a largo plazo, esto lo hace útil en áreas como pensiones, salud pública y estudios de longevidad.

Fórmula del Modelo Lee-Carter

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

Donde los parámetros son los siguientes:

- α_x : **Tendencia temporal** para el grupo de edad x . Este parámetro captura la evolución general de la mortalidad dentro de la población a lo largo del tiempo.
- β_x : **Tasa de mortalidad base** para el grupo de edad x en el período base. Representa la mortalidad para cada grupo de edad suponiendo que no hubiera cambios en el tiempo.
- κ_t : **Componente de interacción entre edad y tiempo**, describe cómo las tasas de mortalidad de cada grupo de edad cambian a través del tiempo.
- $\epsilon_{x,t}$: **Término de error** asociado a cada observación de las tasas de mortalidad, que captura las variaciones no explicadas por los otros parámetros que hay en el modelo.

Ventajas del modelo

- 1. Simplicidad y eficacia:** además de que es un modelo simple, el modelo Lee-Carter es eficaz para describir y proyectar las tasas de mortalidad.
- 2. Flexibilidad:** puede adaptarse a distintas poblaciones y datos de mortalidad, también se puede adaptar a distintas edades y períodos en el tiempo.
- 3. Proyecciones a largo plazo:** permite realizar proyecciones de mortalidad a largo plazo, lo que lo hace útil para planificar estrategias de salud pública.

Limitaciones del modelo

- 1. Linealidad del modelo:** el modelo asume que la mortalidad sigue una tendencia logarítmica a través del tiempo. Esto puede no ser completamente realista, ya que pueden haber eventos que alteren las tasas de mortalidad.

2. No captura las complejidades: como es un modelo simple, no logra capturar las complejidades más detalladas de las tasas de mortalidad, tales como los impactos de eventos demográficos significativos o factores sociales.

Justificación de la elección del modelo

Elegimos el modelo Lee-Carter debido a que es ideal para proyectar las tasas de mortalidad a largo plazo. Su capacidad para descomponer las tasas de mortalidad en tres componentes clave la tasa base por edad, la tendencia temporal y la interacción entre edad y tiempo, lo que lo convierte en una herramienta poderosa para la proyección precisa de la mortalidad en cualquier población. Además, su flexibilidad permite modelar diferentes escenarios y adaptarlo a diversas condiciones demográficas, lo que lo hace adecuado para México.

Decidimos aplicar este modelo a México debido a que proporciona una herramienta útil para anticipar las tendencias futuras de mortalidad en el país. Las proyecciones obtenidas permitirán planificar con mayor precisión los recursos necesarios en áreas críticas como salud, pensiones y bienestar social. En particular, el modelo tiene aplicaciones clave dentro del sistema de pensiones y la seguridad social, ya que facilita la proyección del número de personas que alcanzarán la edad de jubilación y la estimación de las tasas de mortalidad de ese grupo.

Además, creemos que con este trabajo estamos contribuyendo significativamente al área demográfica en México, ya que, al investigar el modelo Lee-Carter aplicado a datos históricos del país, hemos notado que existen pocos estudios que aborden este tema desde esta perspectiva. Con este análisis, esperamos aportar un valor adicional a la comprensión de la mortalidad en México y ofrecer una herramienta útil para futuras investigaciones y también en políticas públicas.

Supuestos del modelo Lee-Carter

- 1. Linealidad temporal:** las tasas de mortalidad siguen una tendencia logarítmica a través del tiempo.
- 2. Independencia de las tasas de mortalidad:** las tasas de mortalidad por edad son independientes entre sí.
- 3. Interacción entre edad y tiempo:** existe un término de interacción que modela cómo las tasas de mortalidad de cada grupo de edad van cambiando con el tiempo.
- 4. Homogeneidad dentro de los grupos de edad:** se asume que todos los individuos dentro de un grupo de edad tienen la misma tasa de mortalidad.
- 5. Ausencia de choques externos:** no se consideran eventos fuera de lo normal (como pandemias o guerras) que alteren considerablemente las tasas de mortalidad.
- 6. Normalidad del error:** el término de error ($\epsilon_{x,t}$) que captura las desviaciones en las tasas de mortalidad sigue una distribución normal.
- 7. Estabilidad de las proyecciones:** suponemos que las tendencias históricas de mortalidad continúan de forma estable en el futuro.

Proceso de estimación

La estimación del modelo se realiza a través de un análisis de componentes principales (PCA) o también de descomposición en valores singulares (SVD).

Lo primero que se hace es realizar el logaritmo de las tasas de mortalidad (m_x), ya que el modelo proyecta de esta manera la mortalidad. Después se realiza la descomposición de los componentes para estimar (α), (β) y (κ).

La optimización de estos parámetros obtenidos se realiza a través de un proceso de optimización, de manera que se minimizan la diferencia entre las tasas de mortalidad observadas y las predicciones generadas por el modelo.

Parámetros del modelo Lee-Carter

Parámetro α

- α_x : **Tendencia temporal** para el grupo de edad x . Este parámetro captura la evolución general de la mortalidad dentro de la población a lo largo del tiempo.

Estimación

$$\alpha_t = \frac{1}{X} \sum_{x=1}^X (\log(m_{x,t}) - \beta_x)$$

Muestra cómo las tasas de mortalidad para cada edad evolucionan de manera distinta en cada año. Captura la interacción específica entre la edad x y el año t .

Se puede calcular a partir de la descomposición de las tasas de mortalidad observadas, utilizando un proceso de optimización para reducir los errores en la predicción.

Parámetro β

- β_x : **Tasa de mortalidad base** para el grupo de edad x en el período base. Representa la mortalidad para cada grupo de edad suponiendo que no hubiera cambios en el tiempo.

Estimación

$$\beta_x = \frac{1}{T} \sum_{t=1}^T \log(m_{x,t})$$

Representa la tasa base de mortalidad para cada grupo de edad en el período base, que por lo general es el período inicial. Muestra cómo es la mortalidad para cada edad dentro de un período en específico. Se calcula a partir de la media de las tasas de mortalidad para cada grupo de edad en todos los años que se tengan en los datos.

Parámetro kappa

- κ_t : **Componente de interacción entre edad y tiempo**, describe cómo las tasas de mortalidad de cada grupo de edad cambian a través del tiempo.

Estimación

Este parámetro captura la manera en que las tasas de mortalidad para un grupo de edad específico cambian con el tiempo, es decir, cómo es que evoluciona la mortalidad para cada grupo en función de los años.

$$\kappa_{x,t} = \log(m_{x,t}) - \beta_x - \alpha_t$$

El conjunto de estas componentes hace el modelo Lee-Carter

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

De esta manera hemos visto la estimación de los parámetros que conforman este modelo.

Plan de simulaciones para el segundo parcial

El objetivo de las simulaciones será evaluar la capacidad predictiva del modelo Lee-Carter y compararlo con un modelo de mortalidad dinámico que refleje mejor la realidad de la mortalidad y que no tenga las limitaciones que tenga el modelo Lee-Carter.

Para el segundo parcial, además de realizar las simulaciones correspondientes, se pretende comparar el modelo Lee-Carter con modelos dinámicos que reflejen mejor la realidad. Modelos de Cohorte Dinámica o modelos de aprendizaje automático. También se pretende comparar las proyecciones realizadas con las proyecciones que realizan las instituciones como el INEGI o la CONAPO.