

Resolución Ejercicio 2: Telecomunicaciones (Análisis de supervivencia)

Felipe Neira Rojas & Angel Llanos Herrera

2024-09-27

Resolución

Importamos los datos de la empresa de telecomunicaciones sobre el tiempo de contrato que permanece activo en cada cliente (hasta su abandono)

```
library(readr)
Telecomunicaciones <- read_delim("C:/Users/angel/Desktop/Telco.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE)
```

Sabemos que, el periodo de interes del estudio es de 36 meses, por lo que, censuramos los clientes con tiempo de contrato mayor a 36.

```
# Cambiamos a variable indicadora
Telecomunicaciones$Churn <- ifelse(Telecomunicaciones$Churn == "No", 0, 1)

# Censurar los datos que exceden los 36 meses
Telecomunicaciones$tenure[Telecomunicaciones$tenure > 36] <- 36
Telecomunicaciones$Churn[Telecomunicaciones$tenure == 36] <- 0
# Censurado si no abandono en 36 meses
```

Visualizamos los datos censurados

```
# Cargar la librería necesaria
library(ggplot2)
# Convertir SeniorCitizen a factor
Telecomunicaciones$SeniorCitizen <- as.factor(Telecomunicaciones$SeniorCitizen)

# Seleccionar los primeros 100 registros
Telecomunicaciones_100obs <- head(Telecomunicaciones, 100)

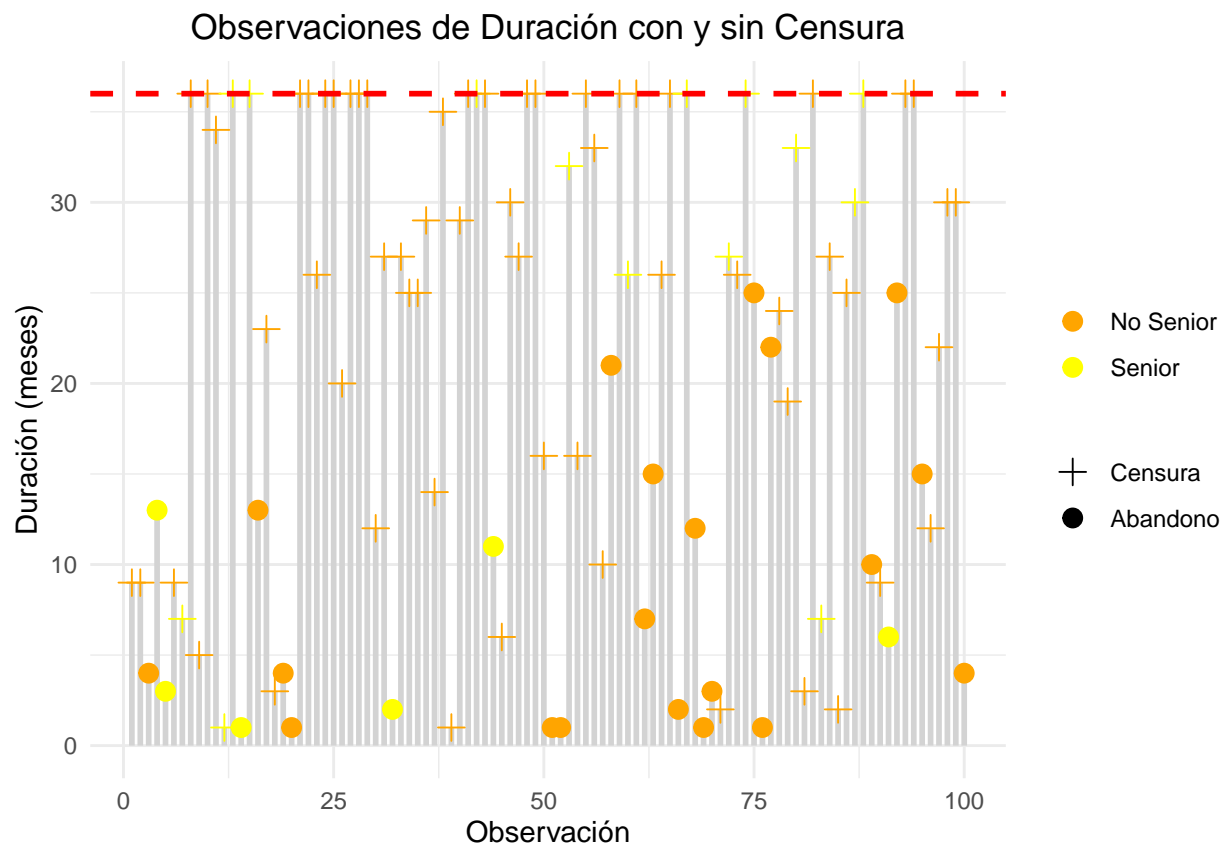
# Crear el gráfico
ggplot(Telecomunicaciones_100obs, aes(x = tenure,
  y = seq_along(tenure),
  color = SeniorCitizen)) +
  # Líneas desde 0 hasta la duración de cada observación
  geom_segment(aes(x = 0, xend = tenure,
    yend = seq_along(tenure)),
    linewidth = 1, color = "lightgray") +
```

```

# Puntos para cada observación, con diferentes formas según la censura
geom_point(aes(shape = factor(Churn)), size = 3) +
# Diferentes formas según la censura
scale_shape_manual(values = c(3, 19), labels = c("Censura", "Abandono")) +
# Asignar colores a SeniorCitizen
scale_color_manual(values = c("0" = "orange", "1" = "yellow"),
                  labels = c("No Senior", "Senior")) +

labs(
  title = "Observaciones de Duración con y sin Censura",
  x = "Duración (meses)",
  y = "Observación"
) +
theme_minimal() +
theme(legend.title = element_blank(),
      plot.title = element_text(hjust = 0.5)) +
coord_flip() +
geom_vline(xintercept = 36, linetype = "dashed", color = "red", size = 1)

```



En este gráfico podemos identificar visualmente una gran cantidad de censura de clientes, en dónde no se conoce el tiempo exacto en meses donde estos abandonaron la empresa.

La censura corresponde principalmente a las personas que no han abandonado la compañía luego de haber transcurrido los 36 meses, los cuales corresponden a los símbolos + que se encuentran en el final del evento (línea roja). Por otra parte, los símbolos que se encuentran dentro del periodo de estudio corresponde a personas que no se obtuvo más información de ella, considerándose como datos censurados.

Creamos la Tasa de abandono acumuladas para cada categoría de edad (Senior/No Senior)

```

# Cargar librerías necesarias
library(survival)
library(survminer)

# Crear un objeto de supervivencia
surv_obj <- Surv(time = Telecomunicaciones$tenure,
                 event = Telecomunicaciones$Churn)

# Ajustar el modelo de supervivencia por categoría de edad
fit <- survfit(surv_obj ~ SeniorCitizen,
              data = Telecomunicaciones)

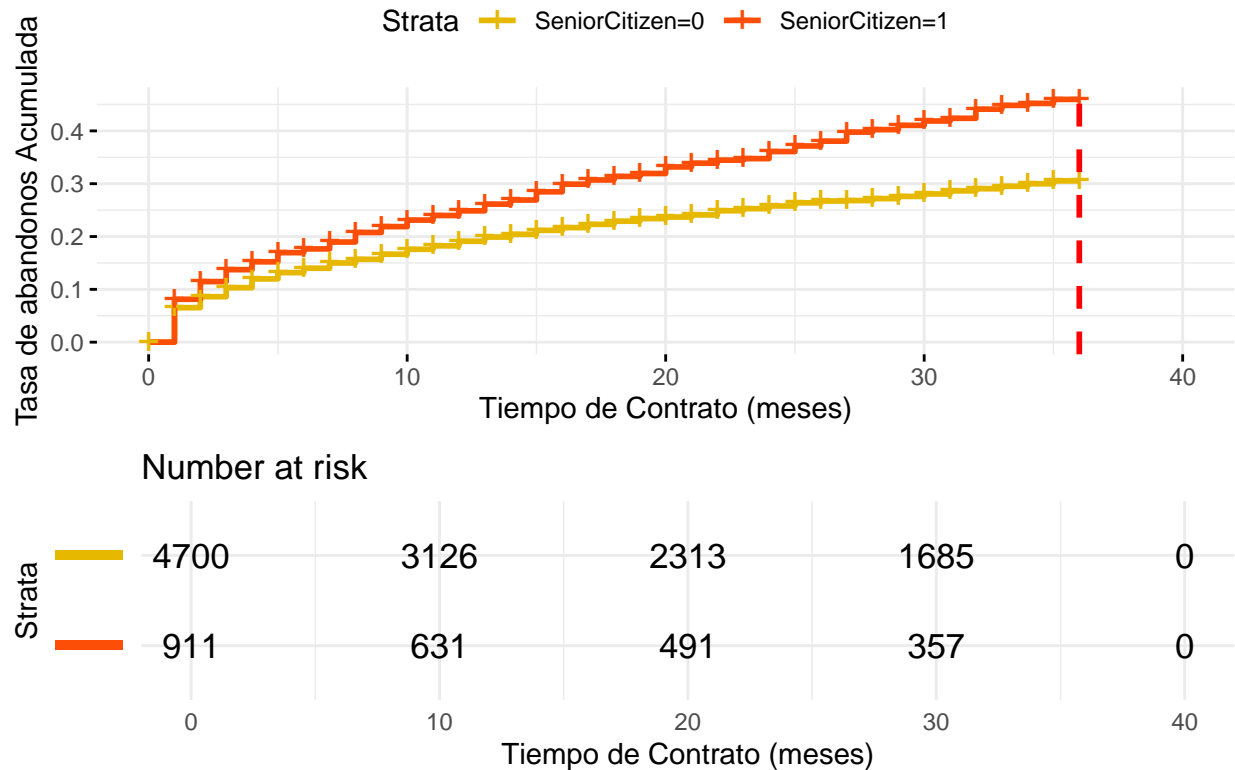
ggsurv <- ggsurvplot(fit,
                    data = Telecomunicaciones,
                    fun = "event",
                    conf.int = FALSE,
                    palette = c("#E7B800", "#FC4E07"),
                    ggtheme = theme_minimal(),
                    title = "Tasa de abandonos acumulada por Senior/No Senior",
                    xlab = "Tiempo de Contrato (meses)",
                    ylab = "Tasa de abandonos Acumulada",
                    risk.table = TRUE,
                    risk.table.height = 0.4,
                    # Ajustar la altura de la tabla de riesgo
                    tables.y.text = FALSE)

ggsurv$plot <- ggsurv$plot +
  geom_vline(xintercept = 36, linetype = "dashed", color = "red", size = 1) +
  theme(plot.title = element_text(hjust = 0.5))

print(ggsurv)

```

Tasa de abandonos acumulada por Senior/No Senior



Observaciones:

En la tasa de fallos acumulada podemos visualizar gráficamente la proporción de personas que abandonaron la compañía en el periodo t , desde los 0 a 36 meses. Se categorizó según edades, en donde visualmente podemos notar una mayor proporción de abandono de las personas mayores (Senior), con respecto a las personas menores (no senior), notando esta tendencia desde el segundo mes. Por otra parte, las personas más jóvenes en promedio permanecen más tiempo en la compañía.

A modo de análisis, se recomiendan plantear estrategias que incentiven a las personas de mayor edad a permanecer en la compañía.

Graficamos las curvas de Kaplan-meier categorizadas por grupo etario.

```
library(survival)
library(survminer)

# Objeto de supervivencia
surv_obj <- Surv(time = Telecomunicaciones$tenure,
                  event = Telecomunicaciones$Churn)

# Ajustar
fit <- survfit(surv_obj ~ SeniorCitizen,
               data = Telecomunicaciones)

ggsurv <- ggsurvplot(fit,
                    data = Telecomunicaciones,
                    conf.int = FALSE,
```

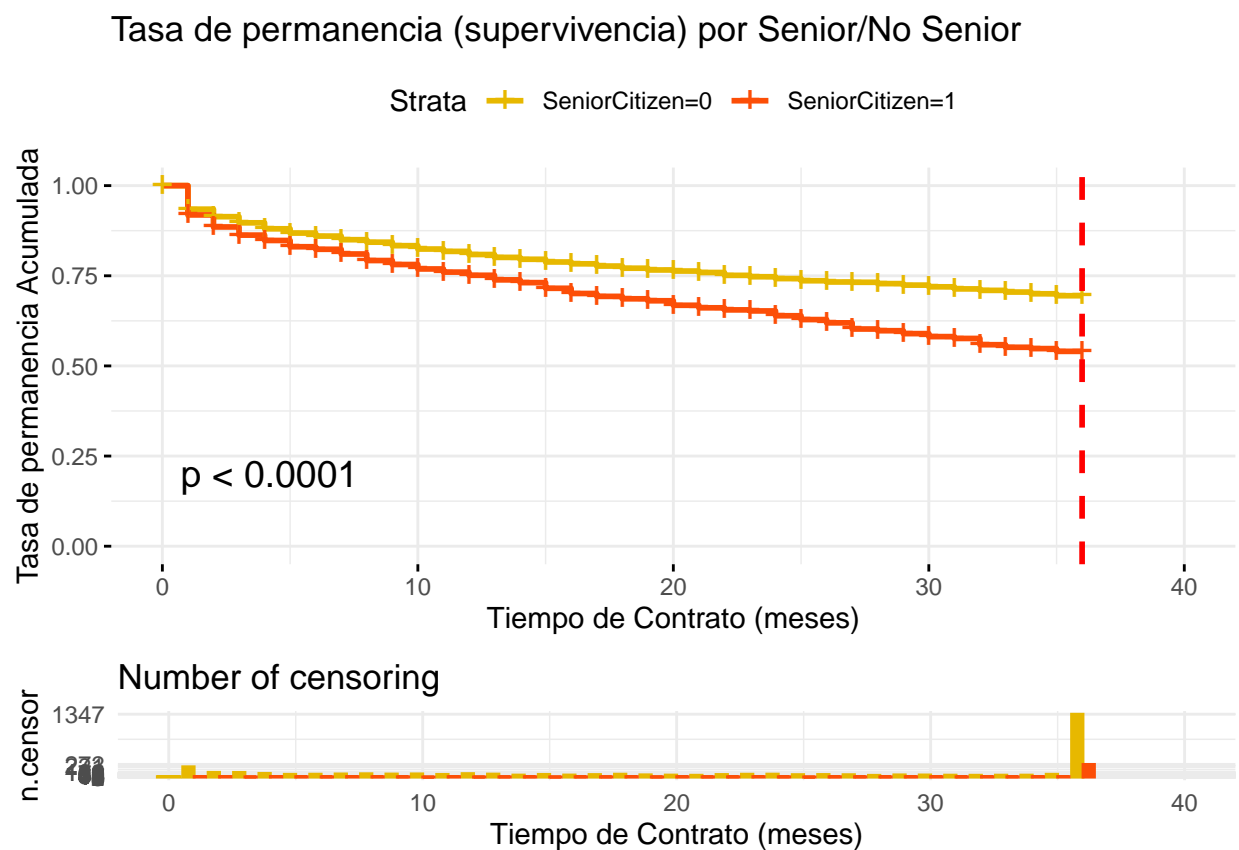
```

ggtheme = theme_minimal(),
palette = c("#E7B800", "#FC4E07"),
ncensor.plot = TRUE,
pval = TRUE,
title = "Tasa de permanencia (supervivencia) por Senior/No Senior",
xlab = "Tiempo de Contrato (meses)",
ylab = "Tasa de permanencia Acumulada")

ggsurv$plot <- ggsurv$plot +
  geom_vline(xintercept = 36, linetype = "dashed", color = "red", size = 1)

print(ggsurv)

```



Ya obtenida la función de supervivencia, podemos visualizar la probabilidad de que las personas se mantengan en la empresa en el tiempo t .

Podemos notar una mayor probabilidad de permanencia en las personas no mayores, resultado que se complementa con el gráfico de tasa de abandono, en donde los dos gráficos coinciden con que las personas mayores tienen mayor probabilidad de abandonar la compañía,

Buscamos la mediana de la durabilidad de cada grupo etario

```

# Calcular la media de permanencia para cada grupo
medias <- summary(fit)$table[, "rmean"]

# Mostrar las medias

```

```
print(medias)
```

```
## SeniorCitizen=0 SeniorCitizen=1  
##          28.47155          25.44849
```

Observaciones:

Logrando calcular la media de permanencia de personas mayores y no mayores en la compañía, podemos decir que en promedio, las personas mayores tienen a abandonar la compañía 3 meses antes que las personas no mayores, logrando obtener una diferencia entre las medias que se comparará mediante una prueba de Log-Rank.

Ahora, determinamos si hay diferencias significativas en la durabilidades entre servicios técnicos. Para esto aplicamos una Prueba de Log-Rank, donde las hipótesis son:

H_0 : No existe diferencia en la función de supervivencia entre los servicios técnicos.

H_1 : Existe al menos una diferencia en la función de supervivencia entre los servicios técnicos.

```
survdif(formula = surv_obj ~ SeniorCitizen,  
         data = Telecomunicaciones)
```

```
## Call:  
## survdif(formula = surv_obj ~ SeniorCitizen, data = Telecomunicaciones)  
##  
##           N Observed Expected (O-E)^2/E (O-E)^2/V  
## SeniorCitizen=0 4700      1141      1249       9.34      57.1  
## SeniorCitizen=1  911       360       252      46.28      57.1  
##  
## Chisq= 57.1 on 1 degrees of freedom, p= 4e-14
```

Con un nivel de significancia del 0.05 podemos rechazar la hipótesis nula, por lo que, hay suficiente evidencia estadística para concluir que existe al menos una diferencia significativa en la continuidad de las personas en la compañía según su grupo etario.

Conclusiones:

Como logramos realizar un estudio de confiabilidad con análisis de datos censurados de Kaplan-Meier, en donde logramos estimar la tasa de abandonos (fallos) y la función de permanencia (supervivencia) de los clientes en la compañía.

Ambos gráficos y funciones nos dieron como resultado una mayor tasa de abandonos de la compañía en personas mayores hasta el mes 36, teniendo las personas no mayores una probabilidad más alta de permanecer en la compañía.

En términos de la empresa, esto refleja una tendencia de abandono de las personas Senior, lo que motiva a tomar medidas correctivas y estrategias de marketing para lograr mantener durante más tiempo a este grupo etario dentro de la compañía. Sin embargo, es importante seguir considerando el grupo etario menor, aunque tenga una menor probabilidad de abandono, siempre será mejor minimizar esa probabilidad a lo largo del tiempo.