

# Resolución ejercicio: Telefonía (Análisis de supervivencia)

Felipe Neira Rojas & Angel Llanos Herrera

2024-09-27

## Resolución

Importamos los datos de la empresa de telefonía (datos simulados) sobre el tiempo de cada celular que cada celular funciona correctamente despues de ser reparado por cierto servicio técnico. (Y librerías necesarias)

```
# Importar las librerías necesarias
library(survival)
library(fitdistrplus) # Ajustar distribución
library(tidyr)
library(ggplot2)
library(survminer)
library(flexsurv) # EMV
library(readr)

datos_celulares <- read_delim("C:/Users/angel/Desktop/TELEFONIA.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE)
```

En el caso parametrico del analisis de supervivencia, es necesario saber la distribución de los datos, en este caso, el tiempo de durabilidad del celular despues de ser reparado por cierto servicio técnico.

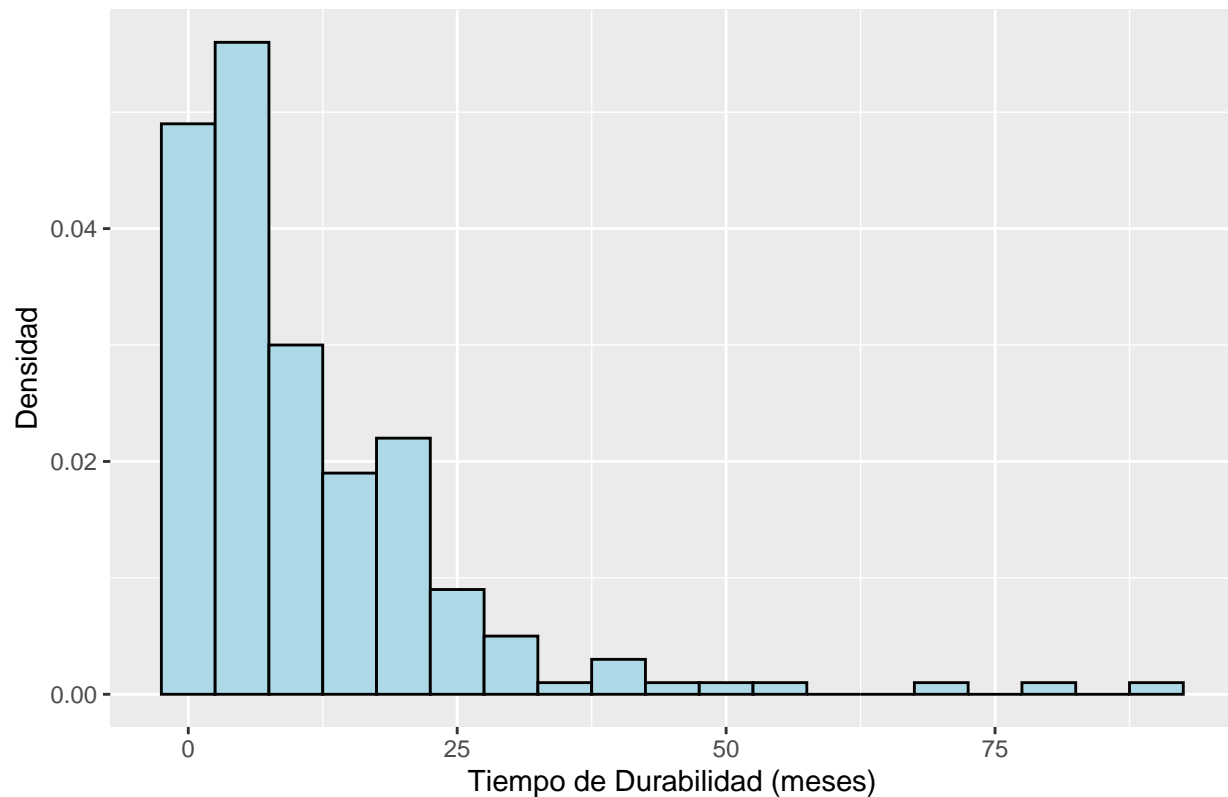
Suponiendo que  $X$  : Tiempo\_Durabilidad se acerca a una distribución Weibull, realizamos un Grafico de densidad y prueba de Kolmogorov-Smirlov para comprobar este supuesto.

$H_0$ : Los datos de Tiempo\_Durabilidad siguen una distribución Weibull.

$H_1$ : Los datos de Tiempo\_Durabilidad no siguen una distribución Weibull.

```
# Densidad para Tiempo_Durabilidad
ggplot(datos_celulares, aes(x = Tiempo_Durabilidad)) +
  geom_histogram(aes(y = ..density..), binwidth = 5, fill = "lightblue", color = "black") +
  labs(title = "Densidad de Tiempo de Durabilidad de Celulares",
    x = "Tiempo de Durabilidad (meses)",
    y = "Densidad")+ theme(plot.title = element_text(hjust = 0.5))
```

## Densidad de Tiempo de Durabilidad de Celulares



```
datos_celulares$Tiempo_Durabilidad[datos_celulares$Tiempo_Durabilidad==0] <- 0.01

ajuste_weibull <- fitdist(datos_celulares$Tiempo_Durabilidad, "weibull")
forma <- ajuste_weibull$estimate["shape"]
escala <- ajuste_weibull$estimate["scale"]

# Prueba de Kolmogorov-Smirnov
ks.test(datos_celulares$Tiempo_Durabilidad, "pweibull", shape = forma, scale = escala)
```

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: datos_celulares$Tiempo_Durabilidad
## D = 0.089475, p-value = 0.08133
## alternative hypothesis: two-sided
```

Por lo tanto, con un nivel de significancia del 0.05, no se rechaza la hipótesis nula (Los datos de Tiempo\_Durabilidad siguen una distribución Weibull), podemos decir que hay suficiente evidencia estadística que apoya que los datos siguen una distribución Weibull.

Siguiendo con el ejercicio. Sabemos que, el periodo de interés del estudio es de 36 meses, por lo que, censuramos los datos con tiempo de durabilidad mayor a 36.

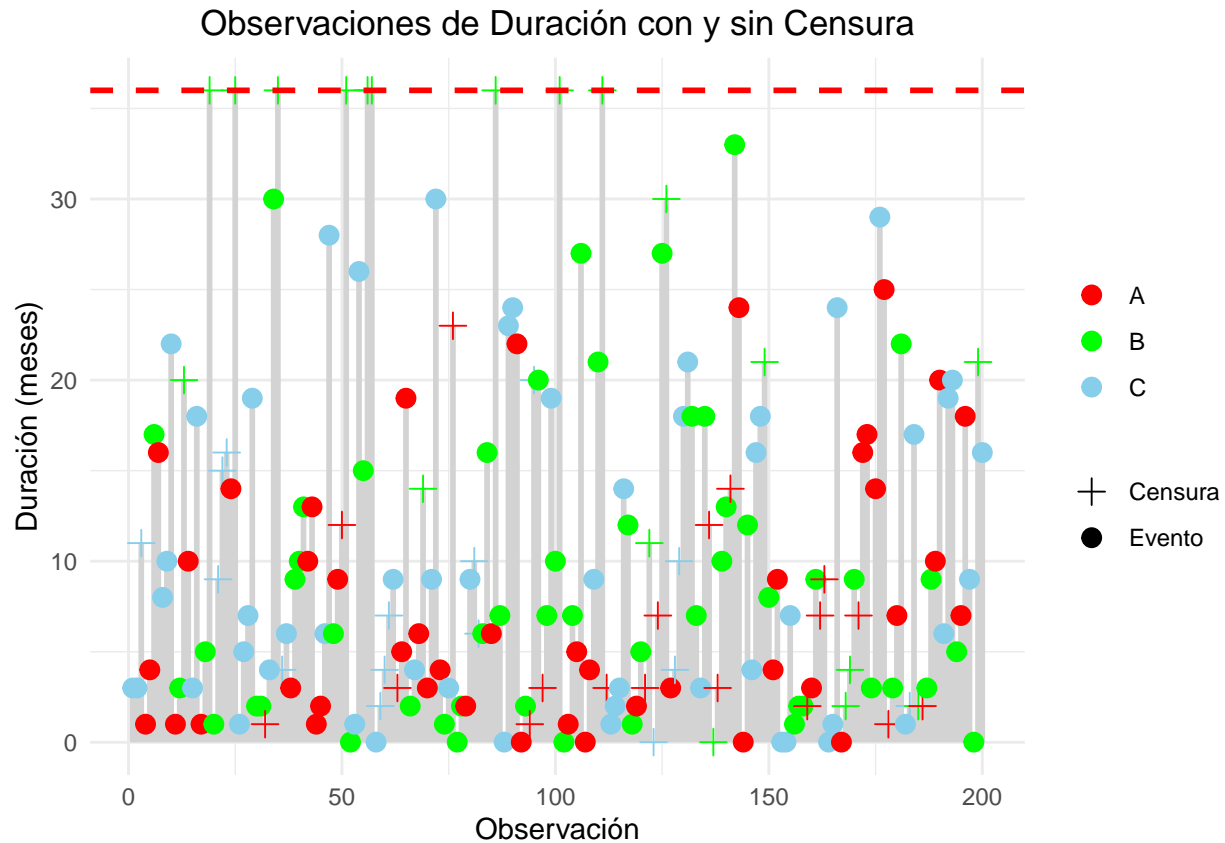
```
# Censurar los datos que exceden los 36 meses
datos_celulares$Tiempo_Durabilidad[datos_celulares$Tiempo_Durabilidad > 36] <- 36
```

```
datos_celulares$Evento[datos_celulares$Tiempo_Durabilidad == 36] <- 0
# Censurado si no hay evento en 36 meses
```

Visualizamos los datos censurados

```
# Cargar la librería necesaria
library(ggplot2)

# Crear el gráfico
ggplot(datos_celulares, aes(x = Tiempo_Durabilidad,
                             y = seq_along(Tiempo_Durabilidad),
                             color = Servicio_Tecnico)) +
  # Líneas desde 0 hasta la duración de cada observación
  geom_segment(aes(x = 0, xend = Tiempo_Durabilidad,
                    yend = seq_along(Tiempo_Durabilidad)),
               linewidth = 1, color = "lightgray") +
  # Color de las líneas
  # Puntos para cada observación, con diferentes formas según la censura
  geom_point(aes(shape = factor(Evento)), size = 3) +
  # Diferentes formas según la censura
  scale_shape_manual(values = c(3, 19), labels = c("Censura", "Evento")) +
  # Cuadrado para censura, cruz para evento
  scale_color_manual(values = c("A" = "red", "B" = "green", "C" = "skyblue"),
                     # Asignar colores a cada servicio técnico
                     labels = c("A", "B", "C")) +
  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)
```



Podemos observar una gran cantidad de datos censurados, tanto en dispositivos que siguieron funcionando después de los 36 meses, como en datos que se perdió información. Estos datos censurados ocurren en los 3 tipos de servicio técnico.

Creamos la Tasa de fallos Acumuladas para cada servicio técnico.

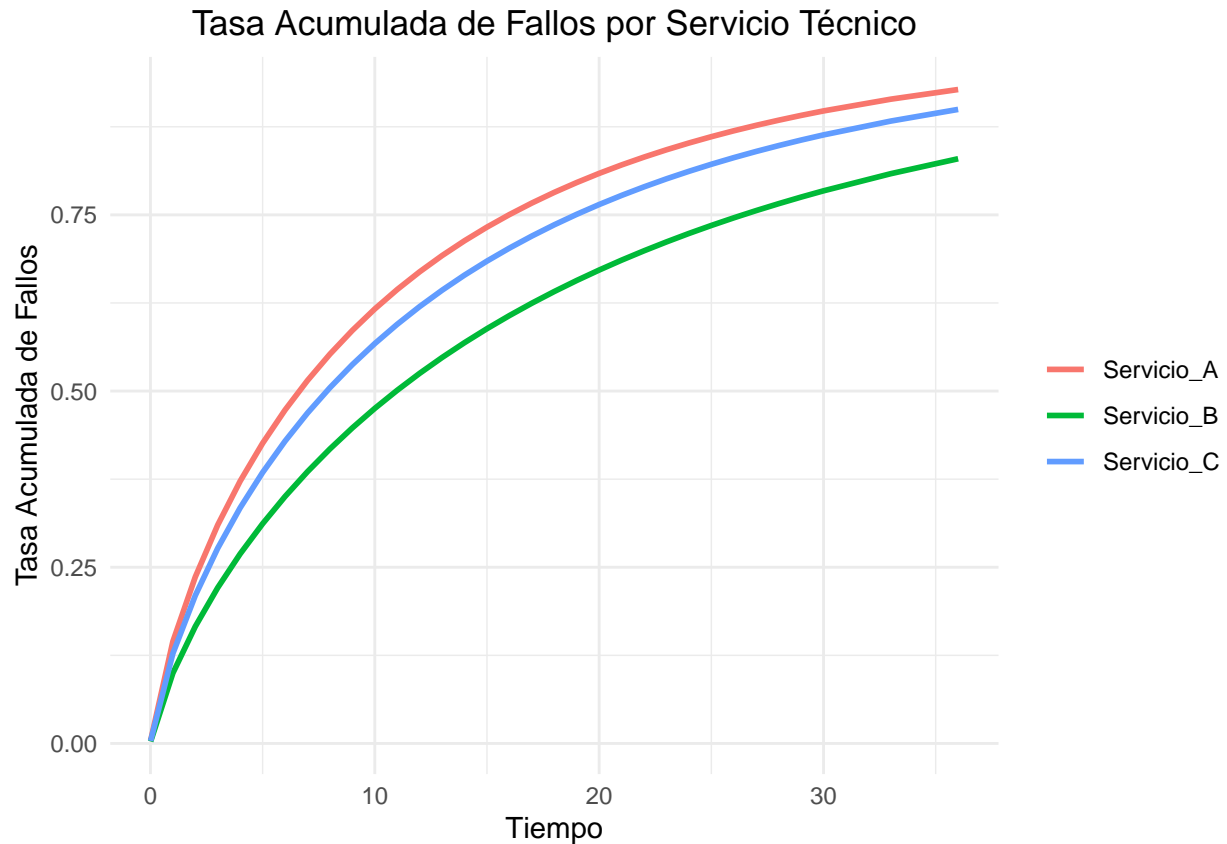
```
# Ajuste de la funcion de supervivencia con ajuste Weibull (EMV)
# Estratificado
flex <- flexsurvreg(Surv(datos_celulares$Tiempo_Durabilidad,
                        datos_celulares$Evento) ~ datos_celulares$Servicio_Tecnico,
                  data = datos_celulares, dist = "weibull") # Ajuste weibull
flexgg <- flex %>% summary(type = "survival") %>% data.frame

tasa_fallos_acumuladas <- data.frame(tiempo =
  flexgg$datos_celulares.Servicio_Tecnico.A.time,
  Servicio_A =
    (1- flexgg$datos_celulares.Servicio_Tecnico.A.est),
  Servicio_B =
    (1- flexgg$datos_celulares.Servicio_Tecnico.B.est),
  Servicio_C =
    (1- flexgg$datos_celulares.Servicio_Tecnico.C.est))

tasa_fallos_long <- tasa_fallos_acumuladas %>% pivot_longer(cols = -tiempo, names_to = "Servicio_Tecnico")

# Graficar
ggplot(tasa_fallos_long, aes(x = tiempo, y = Tasa_Fallos,
                             color = Servicio_Tecnico)) +
```

```
geom_line(size = 1) +
labs(title = "Tasa Acumulada de Fallos por Servicio Técnico",
     x = "Tiempo",
     y = "Tasa Acumulada de Fallos") +
theme_minimal() +
theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5))
```



Observaciones: - Se obtuvieron la tasa de fallos acumulada para cada servicio técnico, en donde al llegar a los 36 meses, generando una probabilidad de que al menos el 75% de los dispositivos arreglados en cualquier servicio técnico tenga un fallo.

- Respecto a las curvas de tasa de fallo acumulada, podemos notar que el mejor servicio lo proporciona el servicio técnico C, teniendo menor tasa de fallos entre el mes 3 y 36. Por otro lado, el servicio técnico C y A presentan tasa de fallos más elevadas, siendo este último el que posee la mayor tasa.

Graficamos las curvas de durabilidad (supervivencia) estratificadas por Servicios tecnicos.

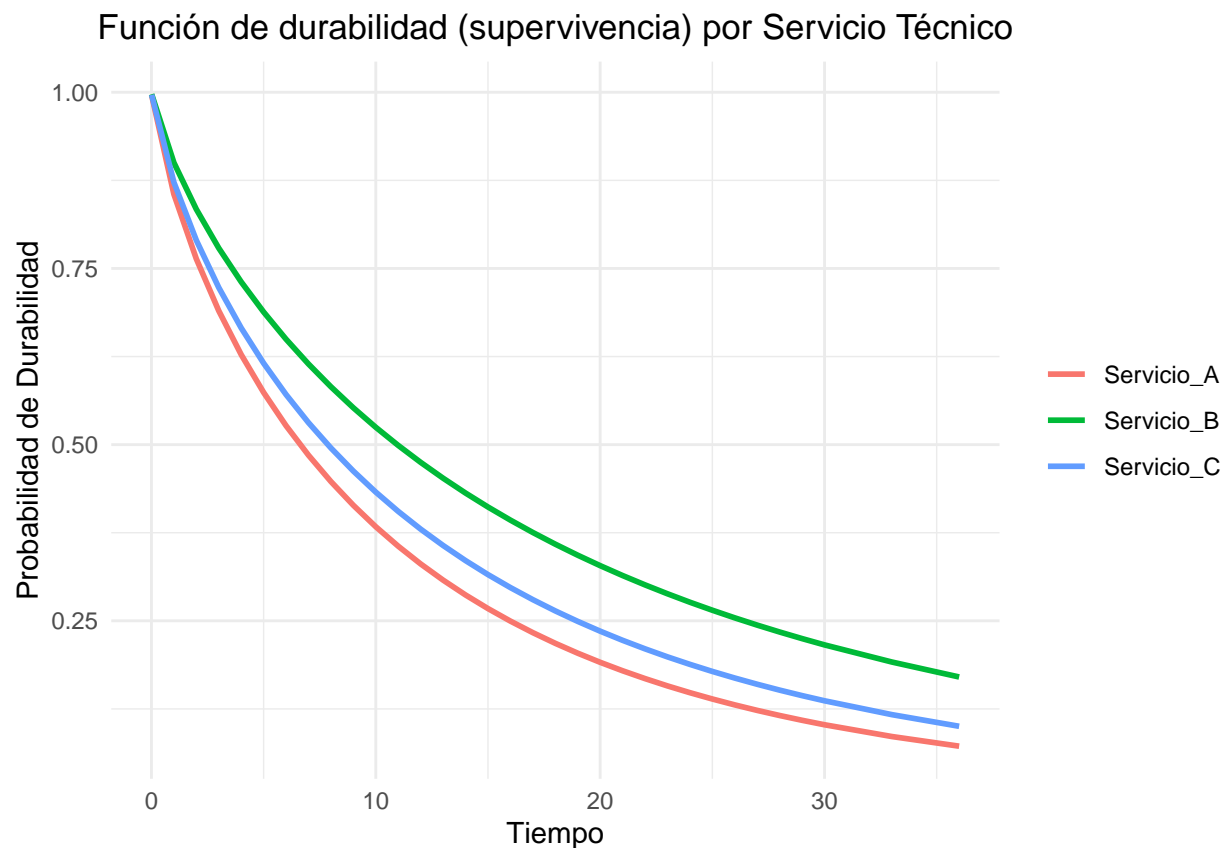
```
supervivencia <- data.frame(tiempo = flexgg$datos_celulares.Servicio_Tecnico.A.time,
                             Servicio_A =
                               flexgg$datos_celulares.Servicio_Tecnico.A.est,
                             Servicio_B =
                               flexgg$datos_celulares.Servicio_Tecnico.B.est,
                             Servicio_C =
                               flexgg$datos_celulares.Servicio_Tecnico.C.est)
```

```

supervivencia_long <- supervivencia %>% pivot_longer(cols = -tiempo, names_to = "Servicio_Tecnico", value_name = "Supervivencia")

# Graficar
ggplot(supervivencia_long, aes(x = tiempo, y = Supervivencia,
                              color = Servicio_Tecnico)) +
  geom_line(size = 1) +
  labs(title = "Función de durabilidad (supervivencia) por Servicio Técnico",
       x = "Tiempo",
       y = "Probabilidad de Durabilidad") +
  theme_minimal() +
  theme(legend.title = element_blank(), plot.title = element_text(hjust = 0.5))

```



Obtenida la función de supervivencia para cada servicio técnico podemos comprobar las observaciones pasadas, en dónde podemos visualizar una mayor probabilidad de supervivencia en los dispositivos del servicio técnico B. Por otra parte, los servicios técnicos A y C presentan una menor probabilidad de supervivencia, siendo los dispositivos del servicio técnico A los que tienen una menor probabilidad

Apoyandonos en que Tiempo\_Durabilidad sigue una distribución Weibull, calculamos la mediana como:

```

ajuste_weibull_A <- fitdist(
  datos_celulares$Tiempo_Durabilidad[datos_celulares$Servicio_Tecnico=="A"],
  "weibull")
forma_A <- ajuste_weibull_A$estimate["shape"]
escala_A <- ajuste_weibull_A$estimate["scale"]

```

```

mediana_A <- qweibull(0.5, shape = forma_A, scale = escala_A)

ajuste_weibull_B <- fitdist(
  datos_celulares$Tiempo_Durabilidad[datos_celulares$Servicio_Tecnico=="B"],
  "weibull")
forma_B <- ajuste_weibull_B$estimate["shape"]
escala_B <- ajuste_weibull_B$estimate["scale"]
mediana_B <- qweibull(0.5, shape = forma_B, scale = escala_B)

ajuste_weibull_C <- fitdist(
  datos_celulares$Tiempo_Durabilidad[datos_celulares$Servicio_Tecnico=="C"],
  "weibull")
forma_C <- ajuste_weibull_C$estimate["shape"]
escala_C <- ajuste_weibull_C$estimate["scale"]
mediana_C <- qweibull(0.5, shape = forma_C, scale = escala_C)

print("-----Medianas-----")

## [1] "-----Medianas-----"

paste("A = ", round(mediana_A,2), "B = ", round(mediana_B,2), "C = ", round(mediana_C,2))

## [1] "A = 4.73 B = 7.67 C = 6.1"

```

Observaciones: - Servicio Técnico A:

Los celulares enviados al servicio tecnico A presentaron la menor mediana de durabilidad, indicando que tienden a fallar más rápidamente, ya que al menos el 50% de los celulares reparados por este servicio tecnico funcionan bien durante 4 meses o más.

- Servicio Técnico B:

La mediana de durabilidad de un celular enviado al servicio tecnico B es la más alta. Esto nos dice que, en promedio, los celulares reparados por este servicio tienen una mejor durabilidad. Al menos el 50% de los celulares funcionaron bien durante 7 meses o más.

Esto no solo sugiere que tardan más en fallar, sino que existe una mayor consistencia en la calidad de las reparaciones realizadas por el servicio tecnico B.

-Servicio Técnico C:

La mediana de durabilidad de los dispositivos reparados en el servicio técnico C es de 6.1, representando que al menos el 50% de los celulares funcionaron 6 o más meses

Es importante considerar que esta mediana fue calculada considerando los datos censurados, donde se asumió que la durabilidad del celular fue al menos igual al tiempo de seguimiento. Por lo tanto, los meses representan el limite inferior para la durabilidad real de estos dispositivos.

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

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

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

```
library(survival)

# Crear objeto de supervivencia
surv_obj <- Surv(datos_celulares$Tiempo_Durabilidad, datos_celulares$Evento)

# Prueba de log-rank
log_rank_test <- survdiff(surv_obj ~ Servicio_Tecnico, data = datos_celulares)
log_rank_test
```

```
## Call:
## survdiff(formula = surv_obj ~ Servicio_Tecnico, data = datos_celulares)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Servicio_Tecnico=A 61      43      33.9      2.463      3.551
## Servicio_Tecnico=B 71      52      64.8      2.520      5.121
## Servicio_Tecnico=C 68      53      49.4      0.269      0.438
##
##  Chisq= 6   on 2 degrees of freedom, p= 0.05
```

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 las durabilidades de los celulares entre los servicios técnicos.

Ahora bien, si comparamos con base en los resultados de la mediana más alta (Servicio técnico B) con otro servicio técnico.

```
log_rank_result <- survdiff(surv_obj ~ Servicio_Tecnico,
                             data = datos_celulares,
                             subset = Servicio_Tecnico %in% c("B", "A"))

log_rank_result
```

```
## Call:
## survdiff(formula = surv_obj ~ Servicio_Tecnico, data = datos_celulares,
##          subset = Servicio_Tecnico %in% c("B", "A"))
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Servicio_Tecnico=A 61      43      34.3      2.19      3.87
## Servicio_Tecnico=B 71      52      60.7      1.24      3.87
##
##  Chisq= 3.9   on 1 degrees of freedom, p= 0.05
```

Podemos ver que con nivel de significancia del 0.05 se rechaza la hipótesis nula, concluyendo que la diferencia significativa de durabilidades de los celulares se encuentra entre el servicio técnico B y A.

Por último, con base en el gráfico de supervivencia era evidente esta diferencia, y por lo mismo, es válido concluir y comprobar la inexistencia de diferencias significativas entre la durabilidad del servicio técnico A y C ( $p > 0.2$ )

```
log_rank_result <- survdiff(surv_obj ~ Servicio_Tecnico,
                             data = datos_celulares,
                             subset = Servicio_Tecnico %in% c("A", "C"))

log_rank_result
```



```
## Call:
## survdiff(formula = surv_obj ~ Servicio_Tecnico, data = datos_celulares,
##           subset = Servicio_Tecnico %in% c("A", "C"))
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## Servicio_Tecnico=A 61      43    36.8      1.05      1.93
## Servicio_Tecnico=C 68      53    59.2      0.65      1.93
##
## Chisq= 1.9  on 1 degrees of freedom, p= 0.2
```

#### Conclusiones:

El análisis de supervivencia realizado en la durabilidad de los celulares reparados por 3 diferentes servicios reflejó diferencias significativas en la durabilidad de los dispositivos ( $p = 0.05$ ). Donde el servicio técnico B presentó la mayor mediana de durabilidad después de reparación (7 meses y 10 días), lo que indica que, en promedio, los celulares reparados por este servicio tuvieron una mayor vida útil (ver gráfico de supervivencia).

Considerando los datos censurados incluidos en el estudio, podemos concluir que, en promedio, los dispositivos reparados en el servicio técnico B tuvieron una vida útil superior a las demás. Sin embargo, es importante tener en cuenta que la presencia de datos censurados implica que el tiempo de vida real de algunos dispositivos podría haber sido incluso mayor al observado.