

Análisis de supervivencia de fallas del corazón

Alfie González, Santiago Battezzati & Elena Villalobos

Maestría en Ciencia de datos
Instituto Tecnológico Autónomo de México

27 de Mayo, 2021.

1. Introducción

La idea del trabajo es que muestren que saben usar los modelos que vimos en clase de manera apropiada para resolver un problema práctico. Les pido que consigan una base de datos, preferentemente de sus trabajos o en su defecto de repositorios de internet, que los motive. Una vez teniendo la base de datos se van a formular una serie de objetivos a resolver y en el trabajo van a plasmar que fue lo que hicieron para resolver esos objetivos.

Descripción del problema, contexto y objetivos a resolver

Qué son fallas del corazón, estadísticas mundiales y en México.

Las fallas del corazón son de las muertes más comunes

2. Base de datos

Descripción de la información: Describan cada una de sus variables, escalas de medición, y presenten un análisis exploratorio de datos

Este conjunto de datos contiene los registros médicos de 299 pacientes que tuvieron una falla en el corazón. Estos datos se colectaron durante el periodo de seguimiento y cada paciente tiene 13 características clínicas¹.

- **age**: Edad del paciente (años).
- **sex**: Mujer u hombre (binaria).
- **anaemia**: Disminución de glóbulos rojos o hemoglobina (booleana).
- **diabetes**: Si el paciente tiene diabetes (booleana).
- **smoking**: Tabaquismo, si el paciente fuma o no (booleana).

¹Base de datos obtenidas del Machine Learning Repository: <https://archive.ics.uci.edu/>

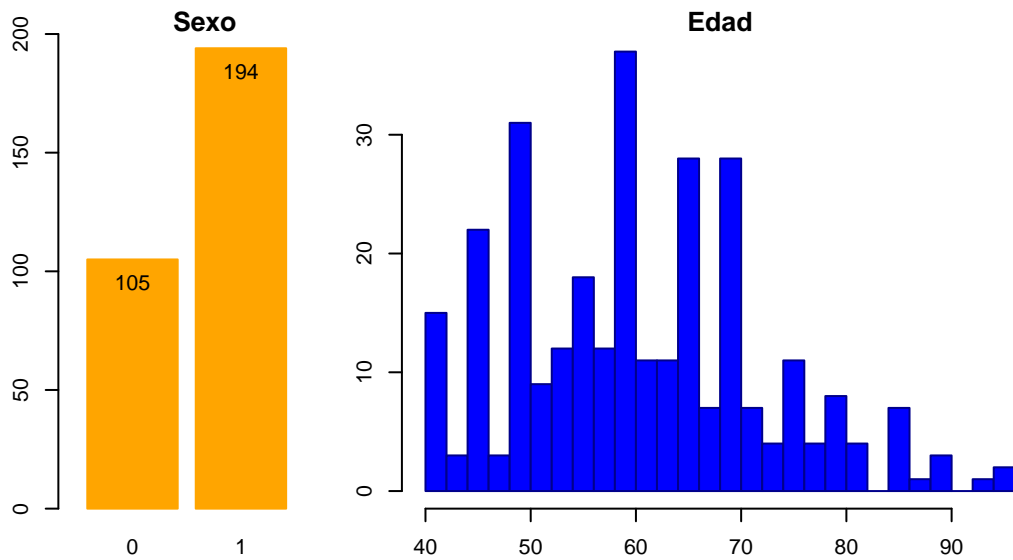


Figura 1: hola

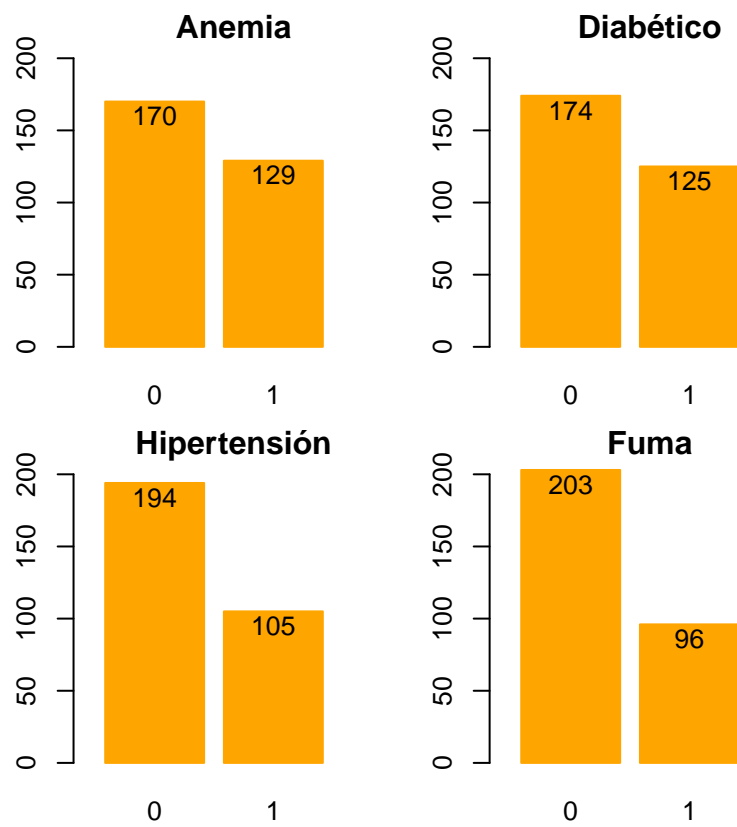


Figura 2: hola

- high-blood-pressure: Si el paciente tiene hipertensión (booleana).

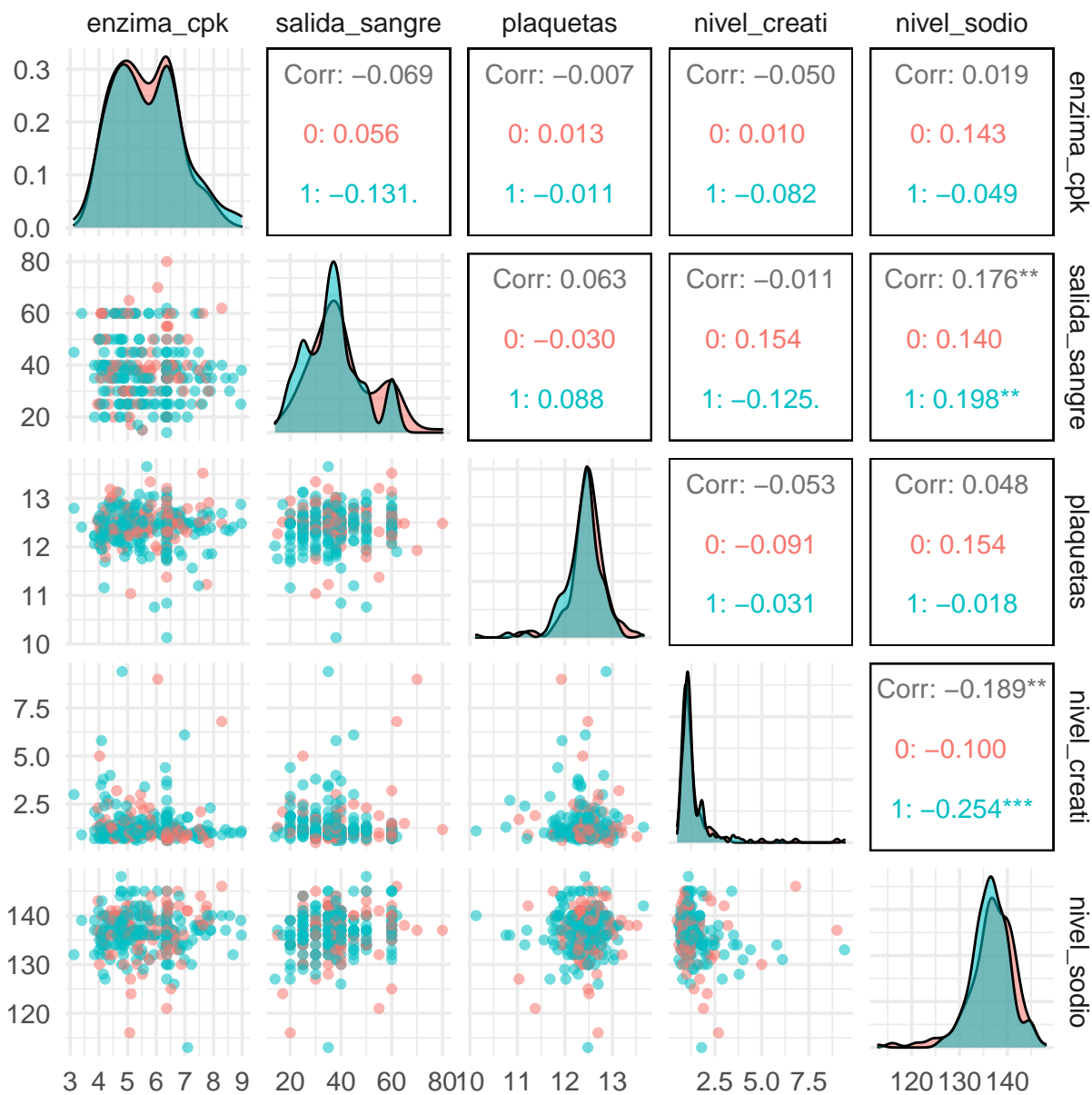


Figura 3: hola

- **ejection-fraction**: Fracción de eyección, porcentaje de sangre que sale del corazón en cada contracción (porcentaje).
- **creatinine-phosphokinase**: Nivel de la enzima CPK en sangre (mcg/L).
- **platelets**: Plaquetas en la sangre (kiloplaquetas/ml).
- **serum-creatinine**: Nivel de creatinina sérica en sangre (mg/dl).
- **serum-sodium**: Nivel de sodio sérico en sanre (mEq/L).

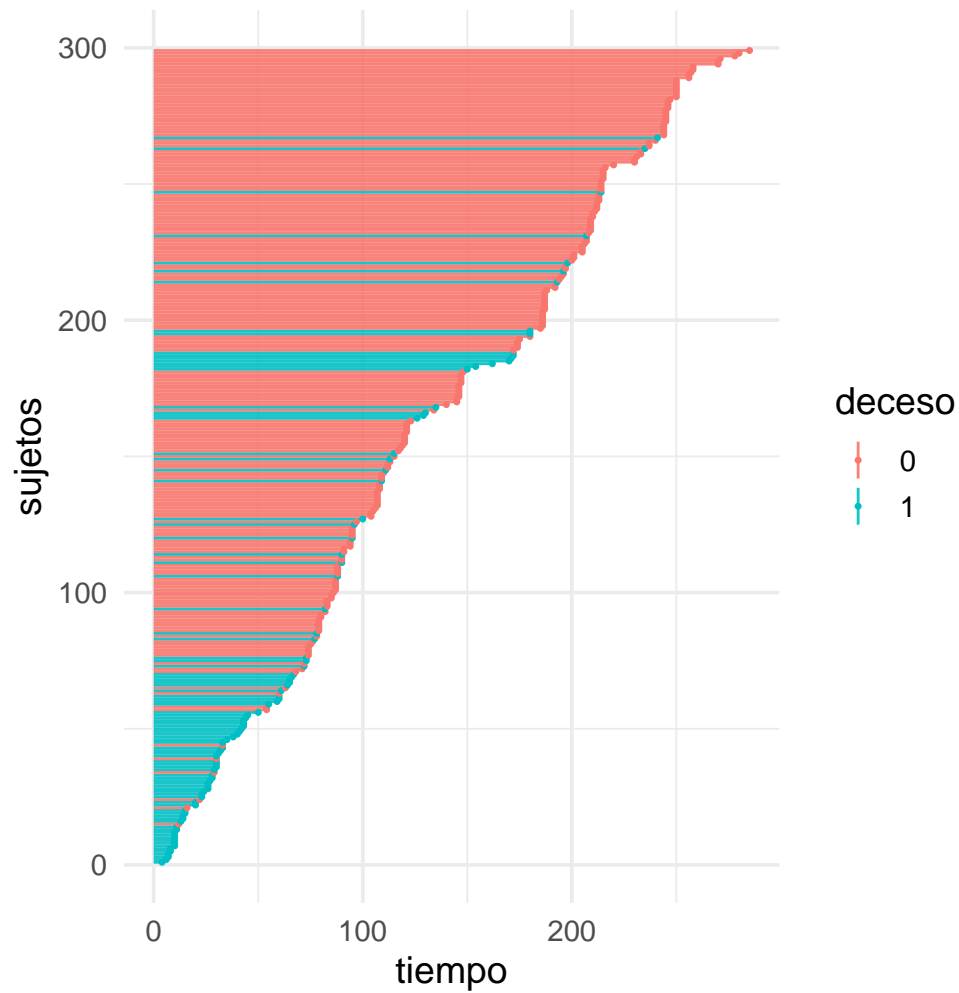


Figura 4: hola

- **time**: Periodo de seguimiento (días).
- **DEATH-EVENT**: Si el paciente falleció durante el período de seguimiento (booleana).

2.1. Recolección

105 mujeres, y 194 hombres. Todos los pacientes tiene disfunción sistólica ventricular.

2.2. Análisis exploratorio

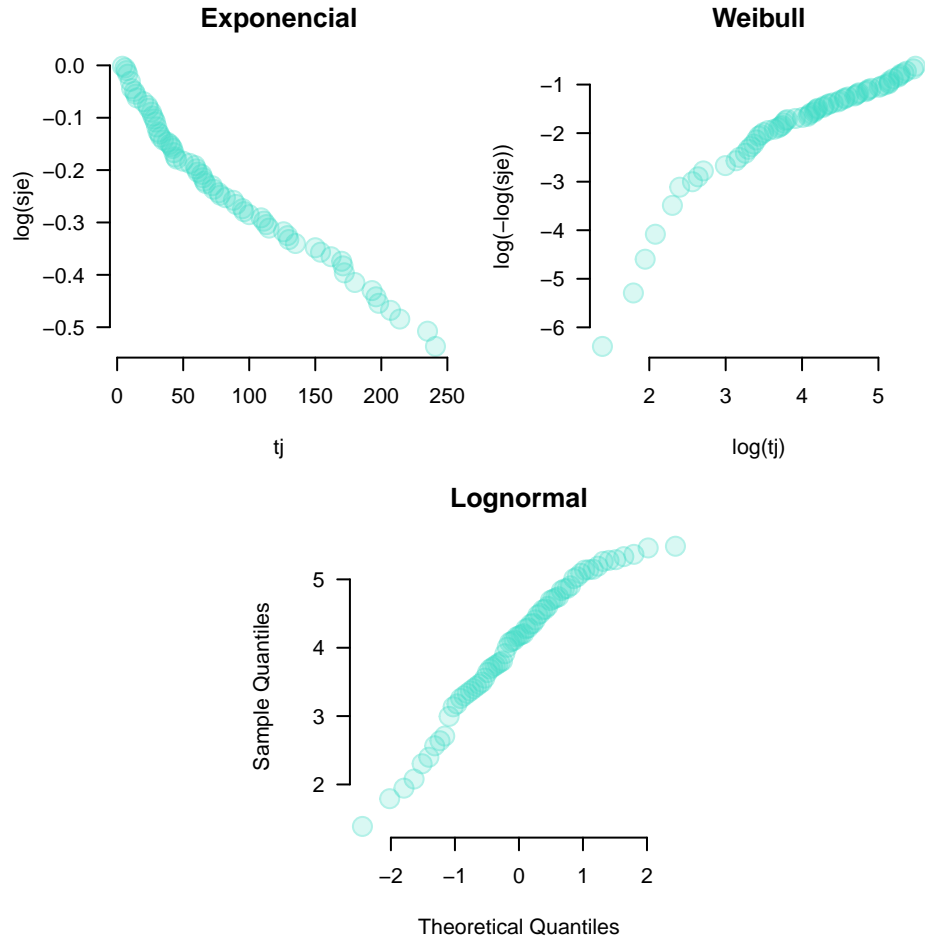


Figura 5: hola

3. Modelado e implementación

3.1. Verificación de supuestos

describan con detalle el modelo, con todas sus especificaciones, que usarán para resolver sus objetivos. Corran su modelo en R, Python o el paquete que mejor les acomode.

4. Resultados

Interpretación de resultados: presenten un resumen de sus estimadores (puntuales y por intervalo) e interpreten en el contexto del problema. Planteen pruebas de hipótesis y tomen decisiones. Hagan uso de sus resultados para responder a los objetivos planteados e incluyan predicciones.

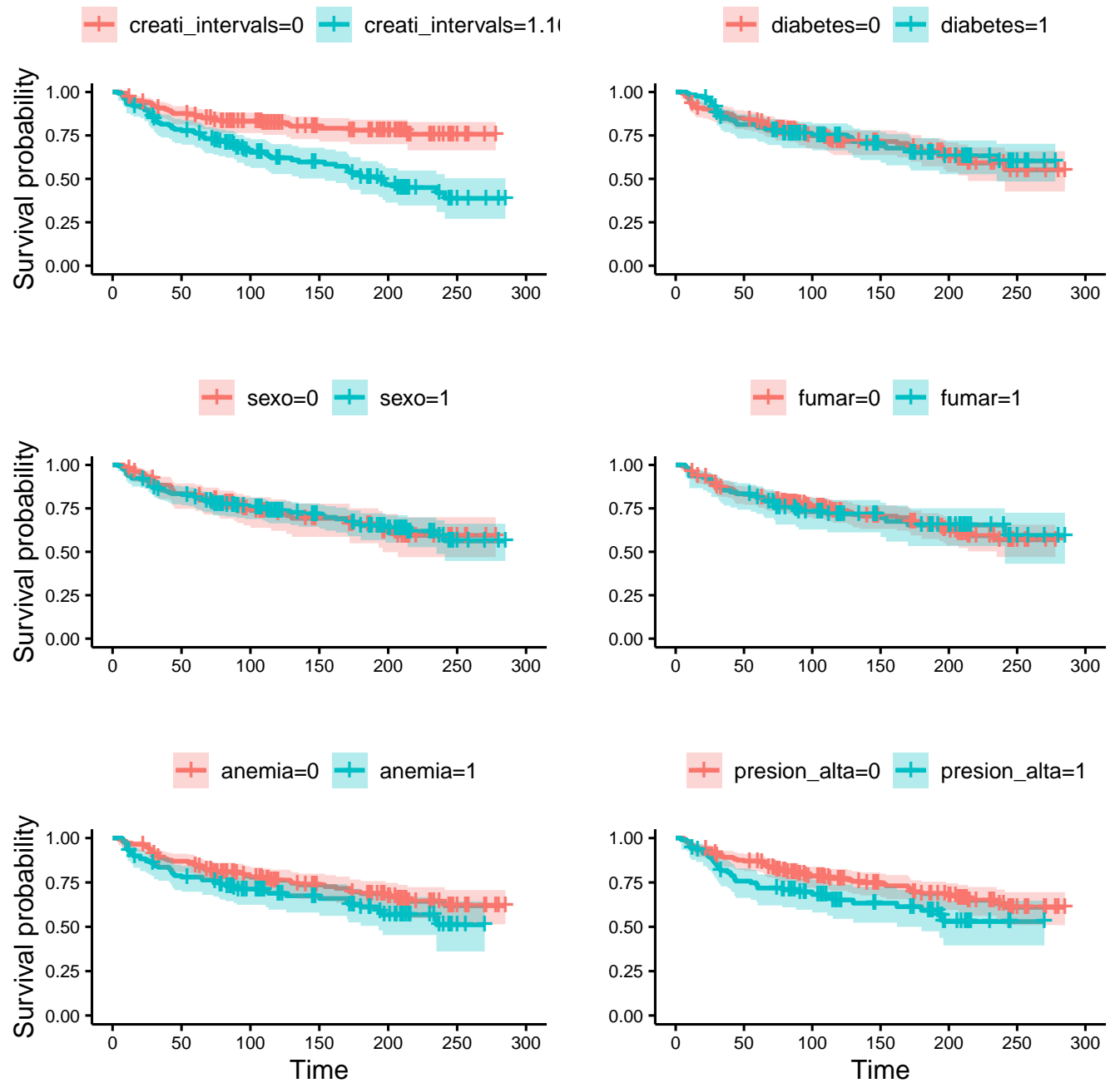


Figura 6: hola

5. Discusión

Referencias

Incluyan una lista de las fuentes que consultaron para hacer su trabajo, desde páginas de internet, libros, revistas o apuntes de clase.

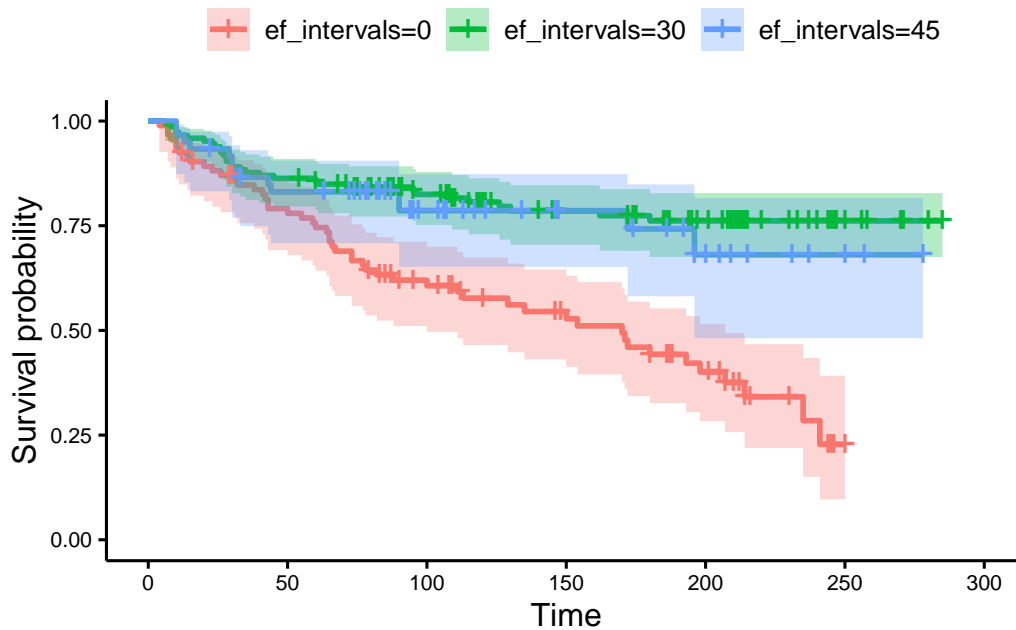


Figura 7: hola

Apéndice

Apéndice. Incluyan si quieren, todo el código utilizado. Por favor no incluyen código dentro de ninguna de las secciones anteriores. NOTA: Las gráficas que consideren útiles las pueden incluir en cualquiera de las secciones de la i-iv con comentarios para que el lector vea lo que ustedes quieren que vean. Las gráficas que no sean indispensables las pueden mandar al apéndice.

4) Preparen una presentación de 15 minutos más o menos, el formato es libre. Todos los integrantes tienen que hablar y la calificación de la presentación será individual, mientras que la calificación del trabajo será por equipo. Se penalizará a aquellos equipos que se tarden más del tiempo asignado originalmente en su presentación.