

Proyecto Final Análisis de Supervivencia

Análisis de Supervivencia para fallas de motor

Víctor Samayoa - 175750 Saúl Caballero - 133930 Delia
Del Águila - 167188

Introducción

Introducción

Se obtuvo la base de datos usada para la competencia de desafío de pronóstico en la Conferencia Internacional sobre pronóstico y gestión de la salud (PHM08). [1]

- La base consiste en múltiples series de tiempo multivariadas.
- Cada serie es de un motor diferente pero de un mismo tipo.
- Hay tres configuraciones operativas que tienen un efecto sustancial en el rendimiento del motor.
- Se cuentan con medidas de 21 sensores en cada ciclo.

Objetivos

Los objetivos son:

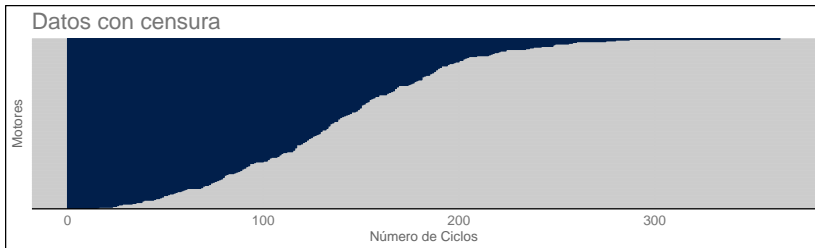
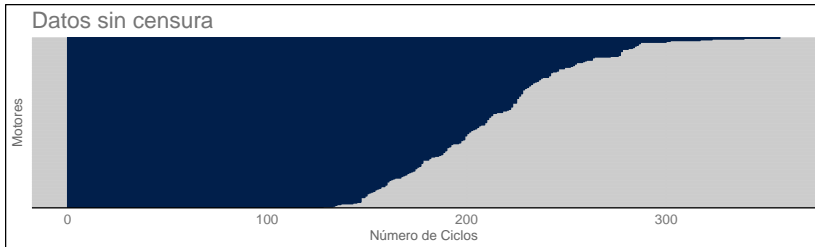
- Estimar la función de supervivencia para los motores donde el evento de falla será cuando el motor no pueda seguir funcionando y tenga que ser mandado a mantenimiento.
- Entender si las configuraciones y las lecturas de sensor ayudan a explicar la falla del motor.
- Construir un modelo predictivo que permita estimar el ciclo previo a la falla.

Análisis de Datos

Análisis exploratorio de los datos

- La base de datos cuenta con 75,738 registros correspondientes a 436 motores
- Cada registro corresponde a un ciclo de un motor en específico e incluye las configuraciones iniciales del ciclo así como las mediciones de los sensores.
- El 50 % de censura por la derecha.

Análisis exploratorio de los datos

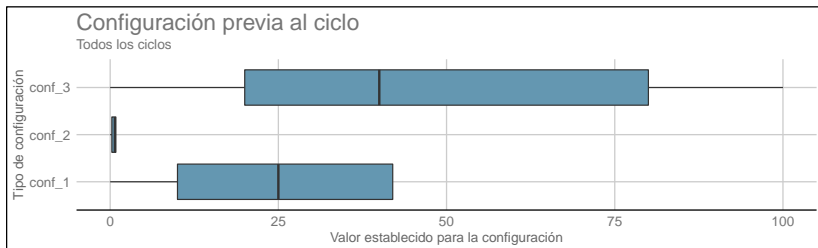


Análisis exploratorio de los datos

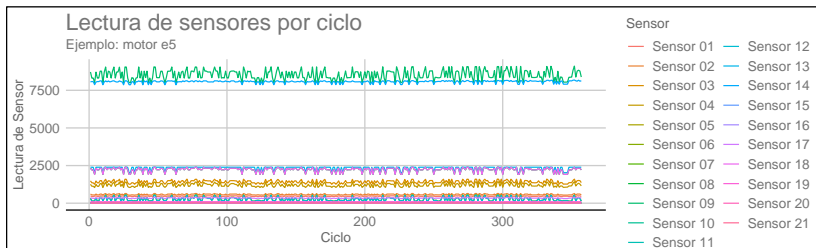
Cada configuración tiene mediciones en la siguiente escala:

- Configuración 1: De 0 a 42.1
- Configuración 2: De 0 a 0.842
- Configuración 3: Valores discretos de 0 a 100 con saltos de 20 puntos

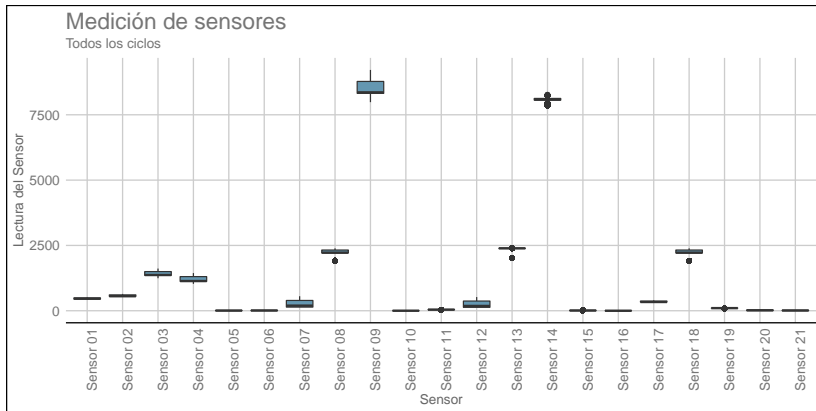
Análisis exploratorio de los datos



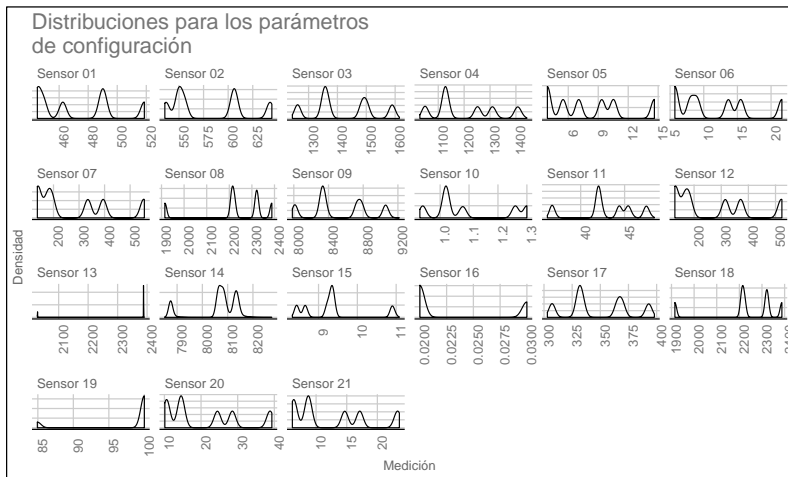
Análisis exploratorio de los datos



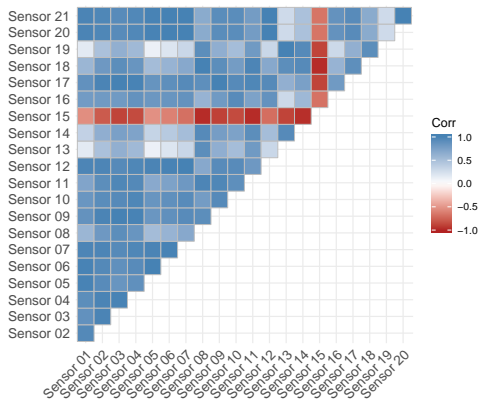
Análisis exploratorio de los datos



Análisis exploratorio de los datos



Análisis exploratorio de los datos

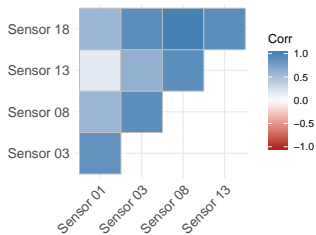


Análisis exploratorio de los datos

Se decide dejar solamente una variable de aquellos pares que tengan una correlación arriba 0.90:

- sensor 01
- sensor 03
- sensor 08
- sensor 13
- sensor 18

Análisis exploratorio de los datos



Análisis exploratorio de los datos

En general cada sensor tiene mediciones en la siguiente escala:

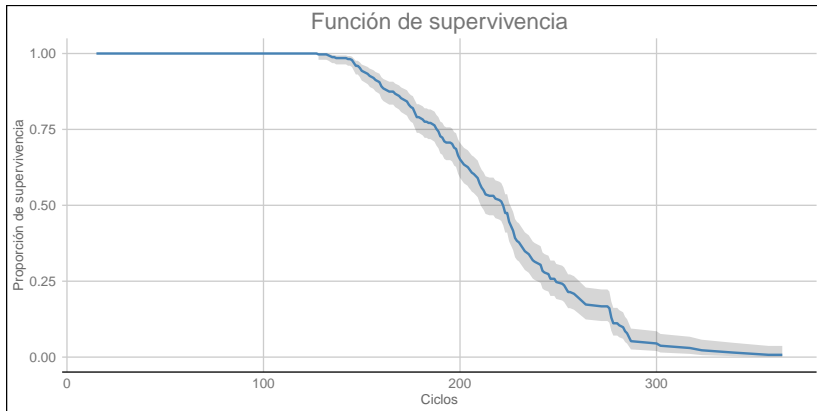
- Sensor 1: De 129.2 a 644.4
- Sensor 3: De 1029 a 1615
- Sensores 8 y 13: De 1915 a 2391
- Sensor 18: Con valores enteros entre 1915 y 2388

Análisis inferencial

Estimador Kaplan Meier

- Se procede a obtener el estimador Kaplan Meier para la función de supervivencia.
- La función de supervivencia con el estimador de Kaplan Meier se comporta de la siguiente forma:

Estimador Kaplan Meier



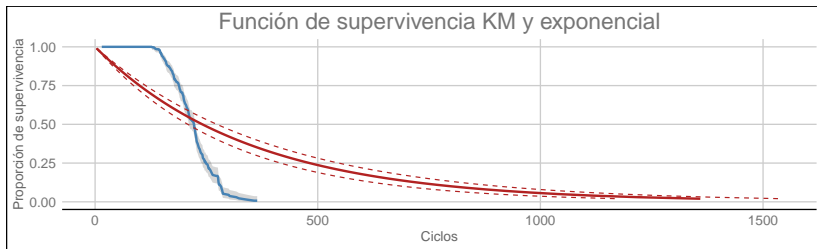
Estimador Kaplan Meier

Cuadro 1: Tabla de cuantiles Kaplan-Meier

Cuantil	Valor	Intervalo de confianza	
		Inferior	Superior
0.25	189	178	197
0.5	222	211	226
0.75	249	241	263

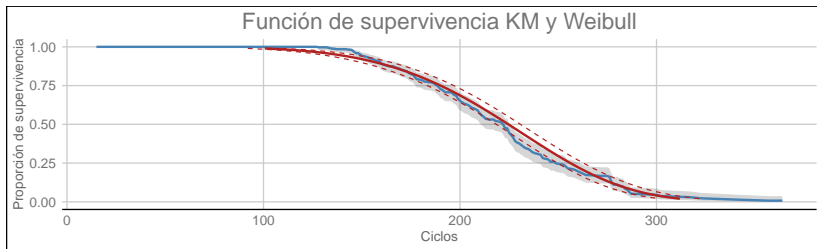
Comparación con modelos paramétricos

Comparación con el modelo exponencial:



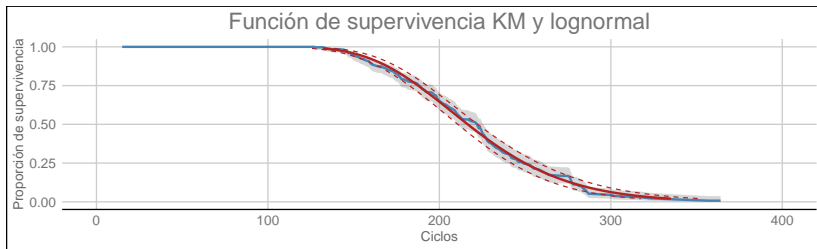
Comparación con modelos paramétricos

Comparación con el modelo Weibull:



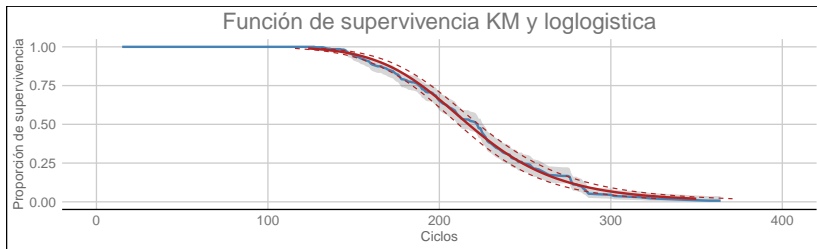
Comparación con modelos paramétricos

Comparación con el modelo lognormal:



Comparación con modelos paramétricos

Comparación con el modelo loglogístico:

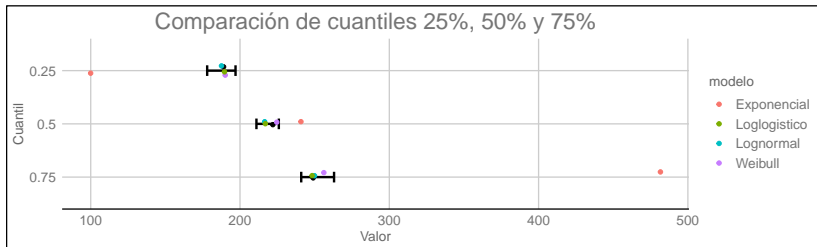


Cuantiles

Cuadro 2: Tabla de cuantiles

Modelo	Cuantiles		
	0.25	0.25	0.75
Exponencial	99.947	240.815	481.629
Weibull	190.106	224.611	256.169
Lognormal	187.604	216.525	249.905
Loglogistico	189.618	216.983	248.297

Cuantiles



Modelos de vida acelerada

- Con base en las gráficas anteriores, observamos que el estimador de la función de supervivencia se ajusta tanto a un modelo lognormal como loglogístico.
- Se procede a realizar ajustes de regresión utilizando como covariables la información asociada a la configuración inicial.
- Se procede a realizar ajustes de regresión utilizando como covariables la información asociada a la información de los sensores.

Modelo loglogístico

Cuadro 3: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-valor
Intercepto	-381.787	335.472	-1.138	0.25
Configuración 1	-0.039	0.039	-1.023	0.30
Configuración 2	0.597	0.668	0.893	0.37
Configuración 3	-0.007	0.005	-1.232	0.21
Sensor 01	0.852	0.731	1.165	0.24
Sensor 03	-0.010	0.003	-3.240	0.00
Sensor 08	-0.295	0.312	-0.945	0.34
Sensor 13	0.332	0.292	1.138	0.25

Modelo loglogístico

Cuadro 4: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	-57.363	21.934	-2.615	0.009
Configuración 3	-0.002	0.001	-2.046	0.041
Sensor 01	0.144	0.048	2.982	0.003
Sensor 03	-0.009	0.002	-4.096	0.000
Sensor 13	0.050	0.018	2.690	0.007
Sensor 18	-0.049	0.020	-2.507	0.012
Log(scale)	-2.127	0.054	-39.451	0.000

Modelo lognormal

Cuadro 5: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-valor
Intercepto	-379.195	343.086	-1.105	0.26
Configuración 1	-0.039	0.039	-0.997	0.31
Configuración 2	0.589	0.679	0.867	0.38
Configuración 3	-0.006	0.005	-1.198	0.23
Sensor 01	0.846	0.748	1.132	0.25
Sensor 03	-0.010	0.003	-3.384	0.00
Sensor 08	-0.280	0.319	-0.876	0.38
Sensor 13	0.330	0.298	1.105	0.26

Modelo lognormal

Cuadro 6: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	9.005	0.906	9.938	0.000
Sensor 03	-0.009	0.002	-4.477	0.000
Sensor 08	0.082	0.024	3.451	0.001
Sensor 13	-0.007	0.002	-4.118	0.000
Sensor 18	-0.070	0.023	-3.055	0.002
Log(scale)	-1.582	0.047	-33.582	0.000

Log-verosimilitud del modelo -1161.2106817

Interpretación

- Un aumento de una desviación estándar del sensor 3 resulta en una disminución del 0.89 % del tiempo de vida.
- Un aumento de una desviación estándar del sensor 8 resulta en un aumento del 8.54 % del tiempo de vida.
- Un aumento de una desviación estándar del sensor 13 resulta en una disminución del 0.69 % del tiempo de vida.
- Un aumento de una desviación estándar del sensor 18 resulta en una disminución del 6.76 % del tiempo de vida.

Conclusiones

Conclusiones

- Se encontró un buen ajuste de la función de supervivencia con Kaplan-Meier.
- Se encontró que los modelos paramétricos loglogístico o lognormal podían ajustar bien al estimador de Kaplan-Meier.
- Se generaron modelos con ambas distribuciones y se encontró que el mejor ajuste lo tenía el modelo lognormal.
- Se interpretaron los coeficientes y los sensores que mayor impacto tienen sobre la supervivencia son el 8 y el 18.
- Se generó el modelo predictivo con el modelo lognormal.

Bibliografía

Bibliografía

[1] <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>