

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.
- Cada motor arranca con diferentes grados de desgaste inicial y variación de fabricación que es desconocido para el usuario.
- Este desgaste y variación se considera normal, es decir, no se considera una condición de falla.
- Hay tres configuraciones operativas que tienen un efecto sustancial en el rendimiento del motor.

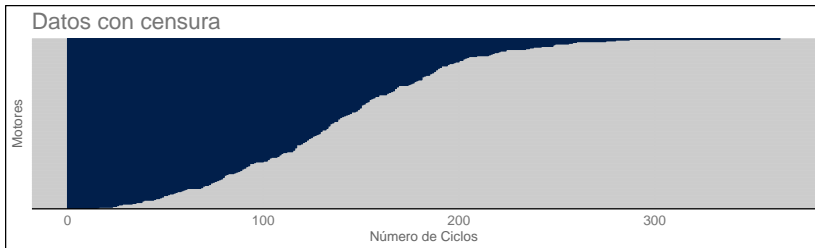
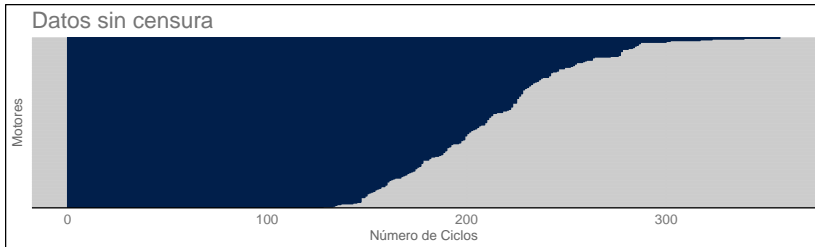
Objetivos

Análisis de Datos

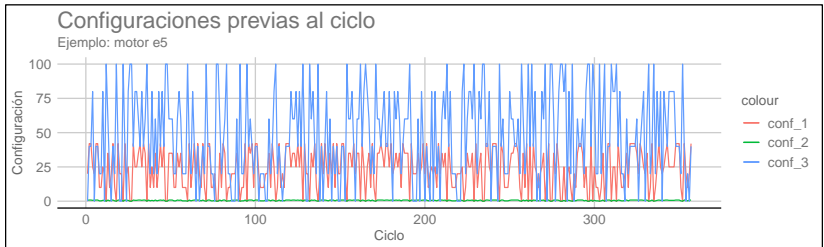
Análisis exploratorio de los datos

- 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

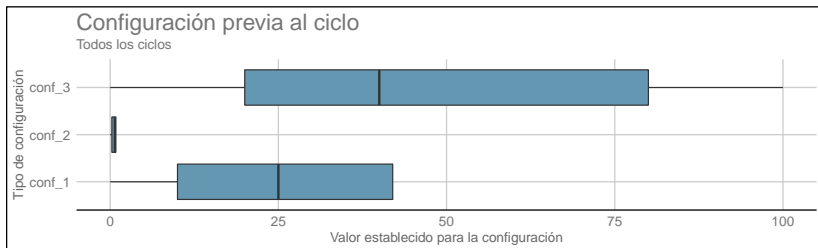


Análisis exploratorio de los datos

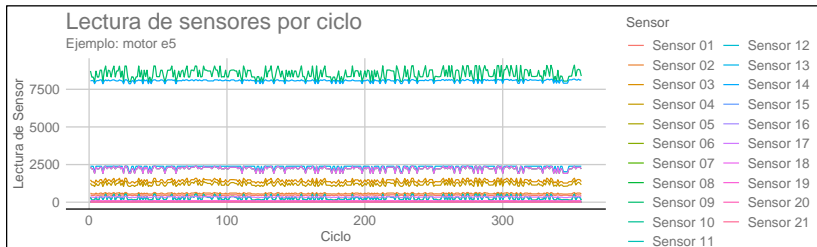
Cada configuración tiene mediciones en la siguiente escala:

- 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

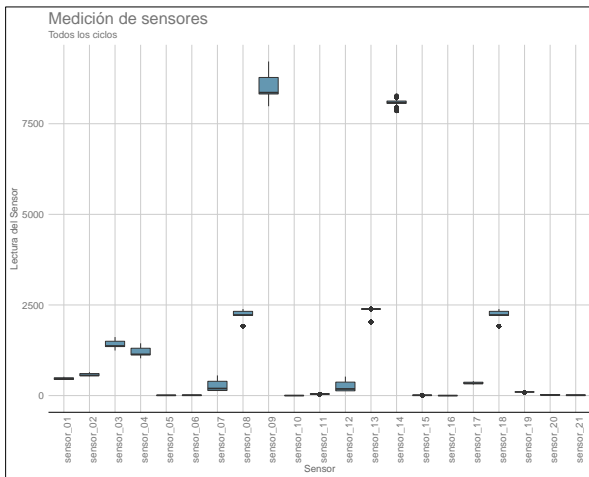


Análisis exploratorio de los datos

En general cada sensor tiene mediciones en la siguiente escala:

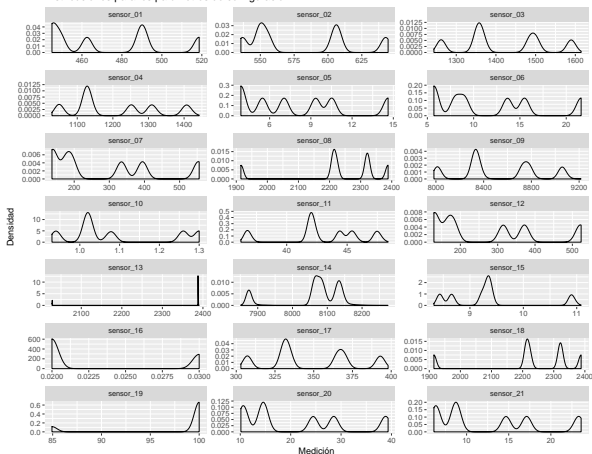
- Sensor 16: Cuenta con dos valores 0.02 y 0.03
- Sensor 10: De 0.93 a 1.3
- Sensores 5, 6 y 15: De 3.91 a 21.61
- Sensores 11, 20 y 21: De 6.124 a 48.39
- Sensor 19: Cuenta con sólo dos valores 84.93 y 100
- Sensores 1, 2, 7 y 12: De 129.2 a 644.4
- Sensor 17: Con valores enteros entre 303 y 398
- Sensores 3 y 4: De 1029 a 1615
- Sensores 8 y 13: De 1915 a 2391
- Sensor 18: Con valores enteros entre 1915 y 2388
- Sensores 9 y 14: De 7852 a 9217

Análisis exploratorio de los datos

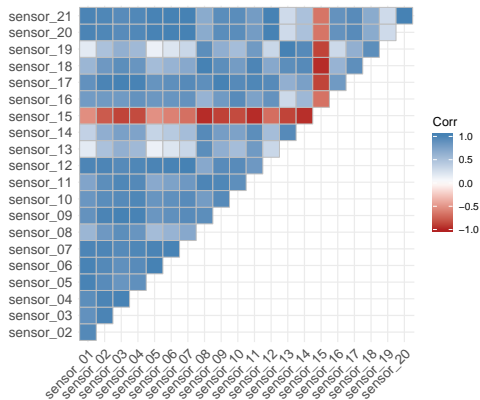


Análisis exploratorio de los datos

Distribuciones para los parámetros de configuración



Análisis exploratorio de los datos

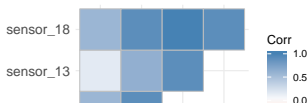


Análisis exploratorio de los datos

De la gráfica, anterior se decide dejar solamente una variable de aquellos pares que tengan una correlación arriba 0.90. De esta forma las variables finales son:

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

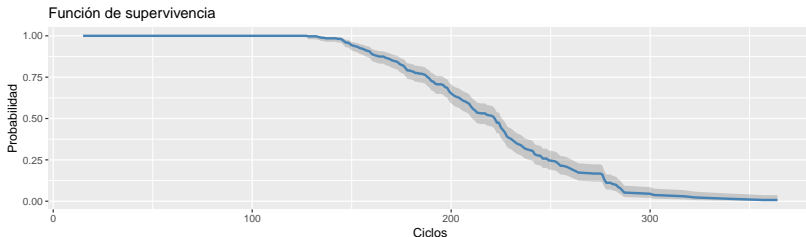
Quedando así la siguiente matriz de correlaciones:



Análisis inferencial

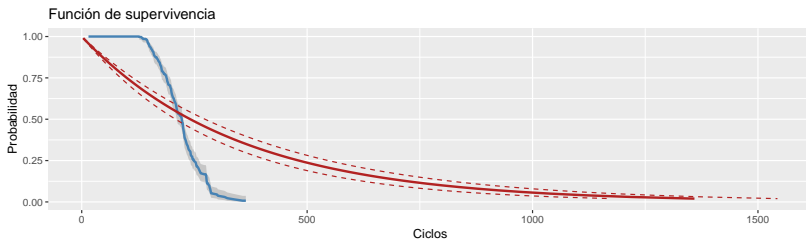
Estimador Kaplan Meier

- 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:



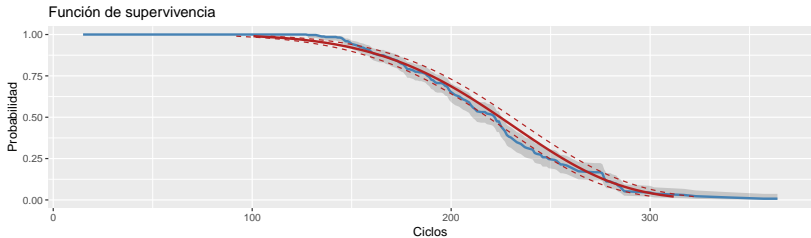
Comparación con modelos paramétricos

Comparación con el modelo exponencial:



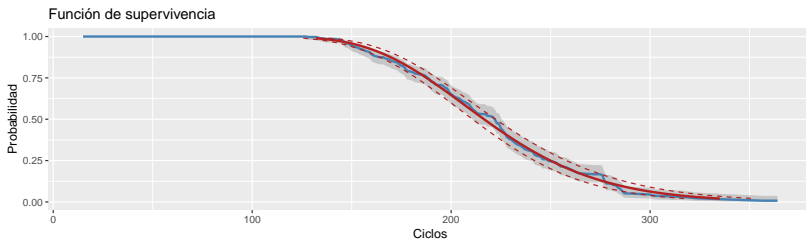
Comparación con modelos paramétricos

Comparación con el modelo Weibull:



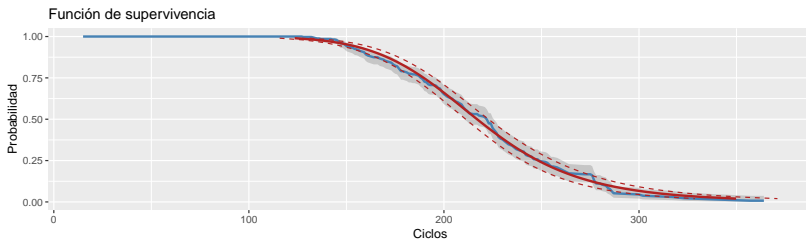
Comparación con modelos paramétricos

Comparación con el modelo lognormal:



Comparación con modelos paramétricos

Comparación con el modelo loglogístico:



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 1: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	5.305	0.044	119.378	0.000
Configuración 1	-0.003	0.003	-0.780	0.435
Configuración 2	0.275	0.178	1.549	0.121
Configuración 2	0.000	0.001	-0.886	0.376
Log(scale)	-2.123	0.053	-39.808	0.000

Modelo loglogístico

Cuadro 2: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	6.435	27.282	0.236	0.814
Sensor 01	0.005	0.059	0.088	0.930
Sensor 03	-0.009	0.002	-4.276	0.000
Sensor 08	0.067	0.036	1.853	0.064
Sensor 13	-0.005	0.023	-0.215	0.830
Sensor 18	-0.057	0.023	-2.452	0.014
Log(scale)	-2.123	0.054	-39.378	0.000

Modelo lognormal

Cuadro 3: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	5.297	0.043	122.392	0.000
Configuración 1	-0.003	0.003	-0.935	0.350
Configuración 2	0.301	0.179	1.685	0.092
Configuración 2	0.000	0.001	-0.530	0.596
Log(scale)	-1.573	0.046	-33.836	0.000

Modelo lognormal

Cuadro 4: Resultados del modelo de regresión

Coeficientes	Valor	Error estandar	Estadístico Z	p-value
Intercepto	7.552	25.476	0.296	0.767
Sensor 01	0.003	0.055	0.057	0.954
Sensor 03	-0.009	0.002	-4.471	0.000
Sensor 08	0.081	0.035	2.297	0.022
Sensor 13	-0.006	0.022	-0.280	0.779
Sensor 18	-0.070	0.023	-3.027	0.002
Log(scale)	-1.582	0.047	-33.582	0.000

Conclusiones

Conclusiones

Aquí van las conclusiones

Bibliografía

Bibliografía

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