

# Análisis de los datos y líneas futuras

**Los datos**

Dato	Explicación del dato
Fecha implante	Fecha en la que se ha insertado el aparato
Fecha último seguimiento	Fecha en la que se ha retirado el aparato
Fecha	Fecha en la que se ha tomado la medida (la fecha de la revisión)
Cámaras	Número de cámaras sobre las que actúa el aparato
STIMA, STIMVD, STIMVI	Porcentaje de estimulación de, respectivamente, la aurícula, el ventrículo derecho y el ventrículo izquierdo, acumulados desde la sesión anterior.
MSA, MSVD, MSVI	Anchura del impulso de salida de, respectivamente, la aurícula, el ventrículo derecho y el ventrículo izquierdo, en milisegundos.
VA, VD, VI	Respectivos valores de tensión a la salida de la aurícula, ventrículo derecho y ventrículo izquierdo medidos en el paciente.
Choque	Número de veces que ha saltado esta característica del aparato desde la revisión anterior.
ATP	Número de veces que ha saltado esta característica del aparato desde la revisión anterior.
Frecuencia mínima	Frecuencia programada a la que funciona el aparato.
R	Indica si el aparato tiene respuesta en frecuencia o no.
Episodios	Número de episodios que ha sufrido el paciente desde la revisión anterior.

**Problemas que nos  
hemos encontrado**

# Datos incompletos

Para poder incorporar una cita al modelo necesitamos saber el valor de todos los datos que vayamos a incorporar al mismo. Por esta razón, es importante que falten el menor número de datos posibles.

En el dataset original contamos con 524 datos.

De los 524 datos, no tenemos ninguno completo.

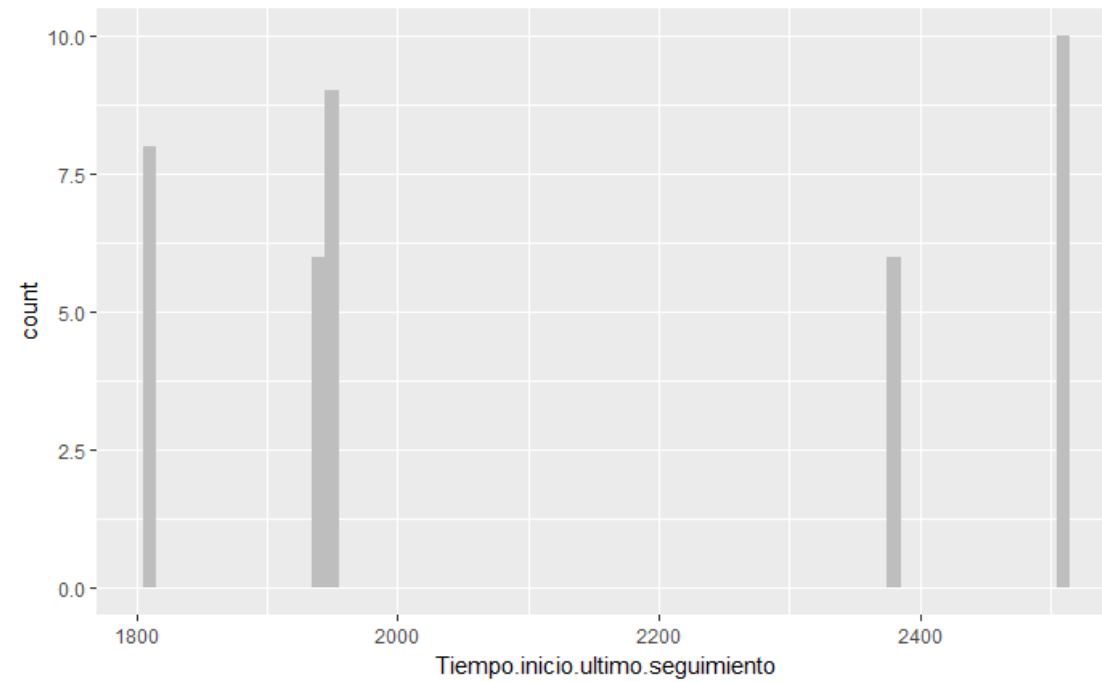
# ¿Qué datos faltan?

Datos	Valores conocidos	Valores desconocidos
Fecha de último seguimiento	<u>39</u> *	<u>485</u>
STIMA	455	69
STIMVI	<u>64</u>	<u>460</u>
Choque	440	84
ATP	440	84
MSA	455	69
VA	455	69
VI	<u>64</u>	<u>460</u>
MSVI	<u>64</u>	<u>460</u>

\* Las operaciones de modelado de machine learning se basan en la asunción de normalidad, para la cual la población con la que se trabaja debe tener un tamaño igual o superior a 50. Esta variable no la cumple.

# **Algunos ejemplos**

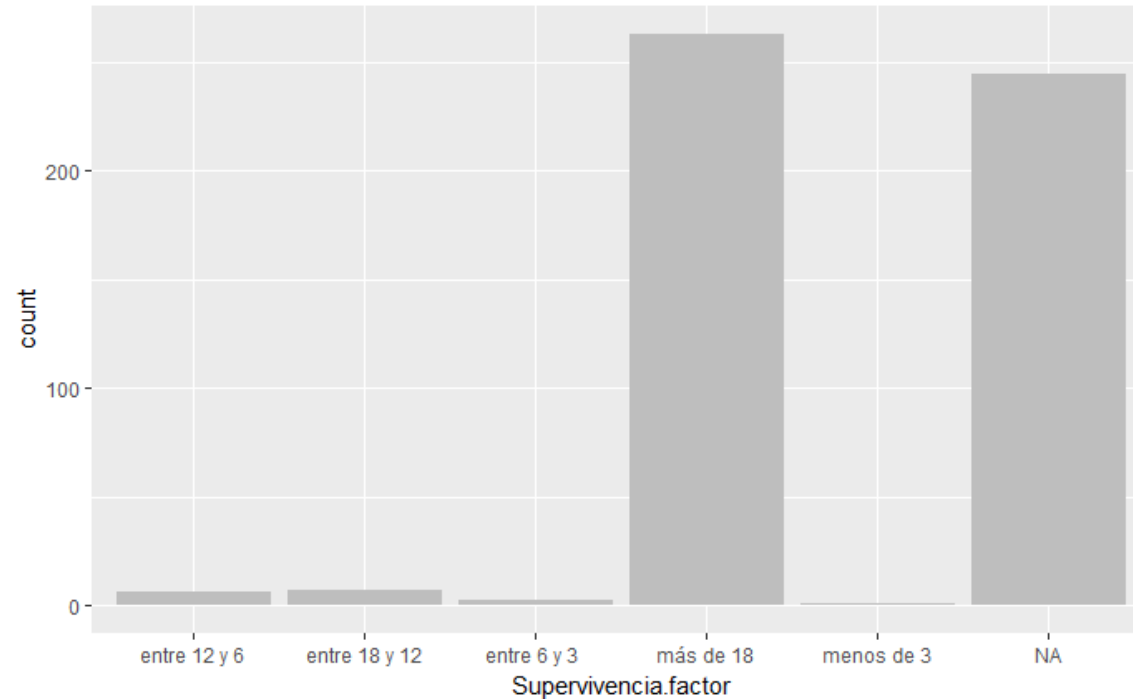
# Tiempo entre inicio y último seguimiento



En la gráfica podemos apreciar la distribución de los tiempos de vida que conocemos de distintos aparatos. La distribución no se asemeja a una normal.

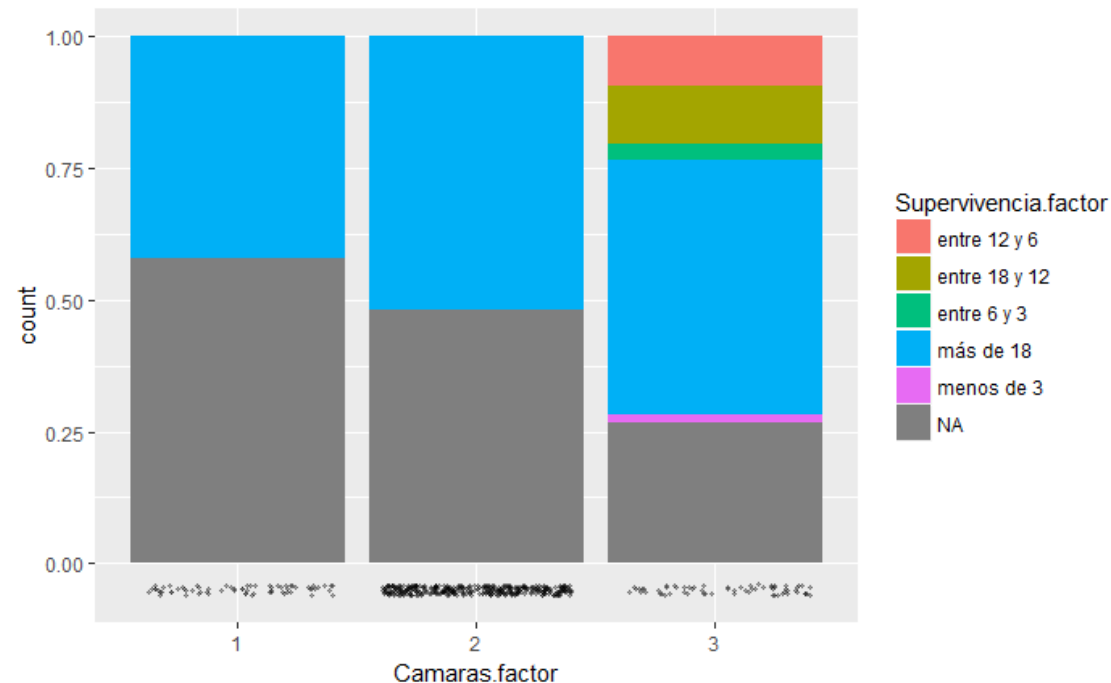


# Supervivencia



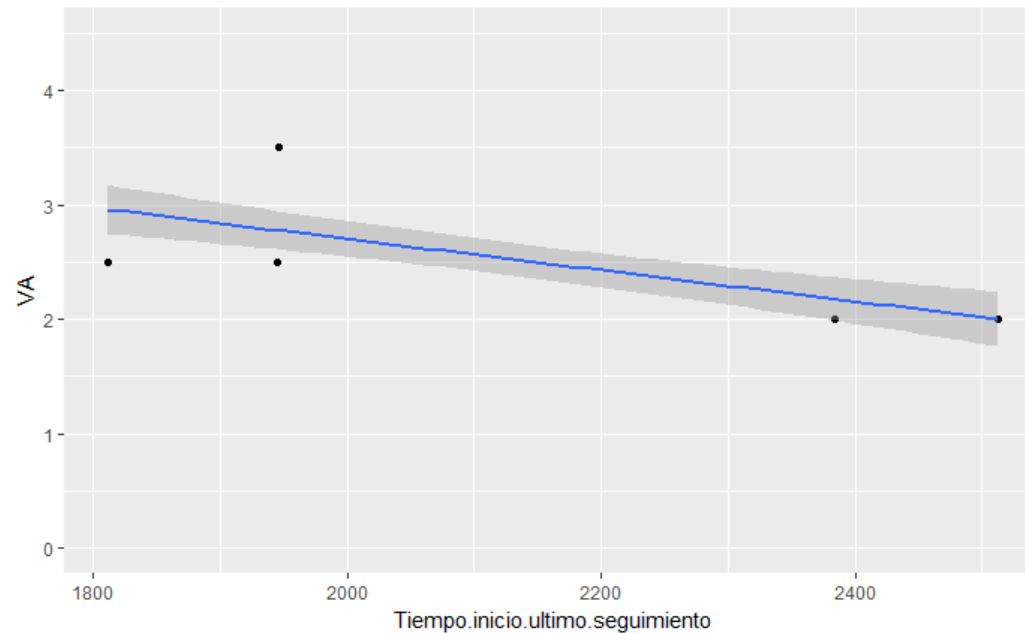
En la gráfica podemos observar la distribución de la variable Supervivencia, así como la cantidad de datos que desconocemos (los marcados como NA). Esta cantidad es demasiado grande como para no tenerlos en cuenta y asumir que la población presenta una distribución como la representada en la gráfica.

# Supervivencia en función del número de cámaras



Aquí podemos apreciar la relación entre el número de cámaras y la supervivencia del aparato. Los valores grises son los desconocidos. No será la única variable con la que solo nos encontremos diversidad de datos únicamente para las 3 cámaras.

# Relación entre el tiempo entre inicio y último seguimiento y la VA



En la gráfica se aprecia la relación entre el tiempo entre inicio y último seguimiento y VA, para todos los valores conocidos de ambos. Podríamos asumir que hay cierta relación, pero con tan pocos datos no es estadísticamente correcto.

**Solución**

A veces, es posible recuperar los datos aunque no los tengamos. Este es el caso de datos como Episodios, que si no se han registrado es porque no han sucedido, o STIMVI, que si no existe es porque no hay estimulación VI y por lo tanto podemos asumir que es 0.

DATO	Valor asignado a los nulos	Por qué
STIMA	0	Si el valor es nulo es porque al paciente no se le ha programado esa estimulación.
STIMVI	0	Si el valor es nulo es porque al paciente no se le ha programado esa estimulación.
Choque	0	Si el valor es nulo es porque al paciente no se le ha programado esa característica.
ATP	0	Si el valor es nulo es porque al paciente no se le ha programado esa característica.
MSA	?	Es una medida del paciente, y que no se mida no significa que no tuviera un valor determinado en ese momento.
VA	?	Es una medida del paciente, y que no se mida no significa que no tuviera un valor determinado en ese momento.
MSVI	?	Es una medida del paciente, y que no se mida no significa que no tuviera un valor determinado en ese momento.
VI	?	Es una medida del paciente, y que no se mida no significa que no tuviera un valor determinado en ese momento.

# La fecha de retirada

La fecha de retirada es un caso especial, pues es el dato “clave” para nuestro sistema. Queremos prever cuándo va a agotar su vida útil el aparato, pero si no tenemos ese dato en nuestro dataset de entrenamiento no podemos sustituirlo con un 0.

En esta situación, tenemos dos posibles soluciones:

1. Recoger más datos.
2. Intentar inferir los datos que faltan.

# Solución: inferir los datos que faltan

Sabemos que el tiempo de vida del aparato tiene que depender de alguno de los datos con los que estamos trabajando. Realizamos un análisis de relaciones entre algunas variables y el tiempo de vida del aparato para encontrar estas dependencias.

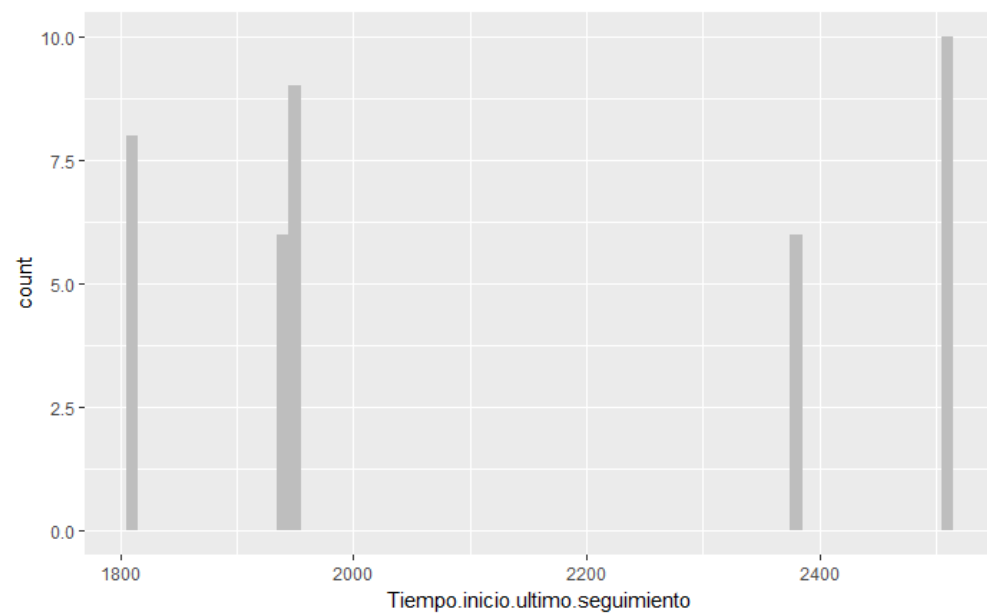
Encontramos dos relaciones con coeficientes de correlación altos:

1. El tiempo de vida que le queda al aparato y el tiempo que lleva funcionando
2. El tiempo de vida que le queda al aparato y el voltaje de la batería en determinado momento.



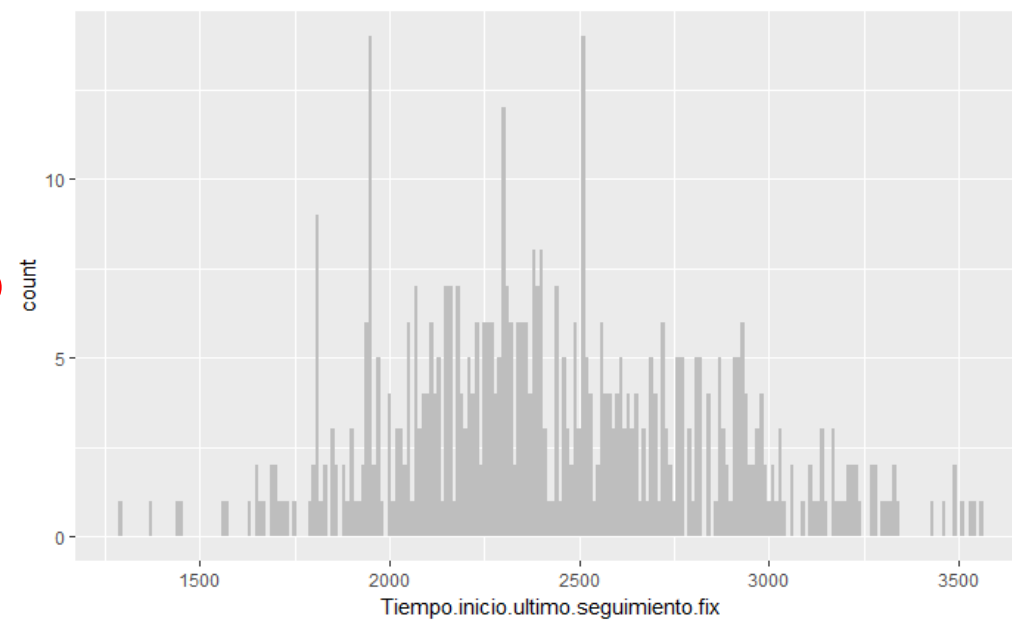
# **Resultados de la solución**

# Según el voltaje de la batería



Original

VS



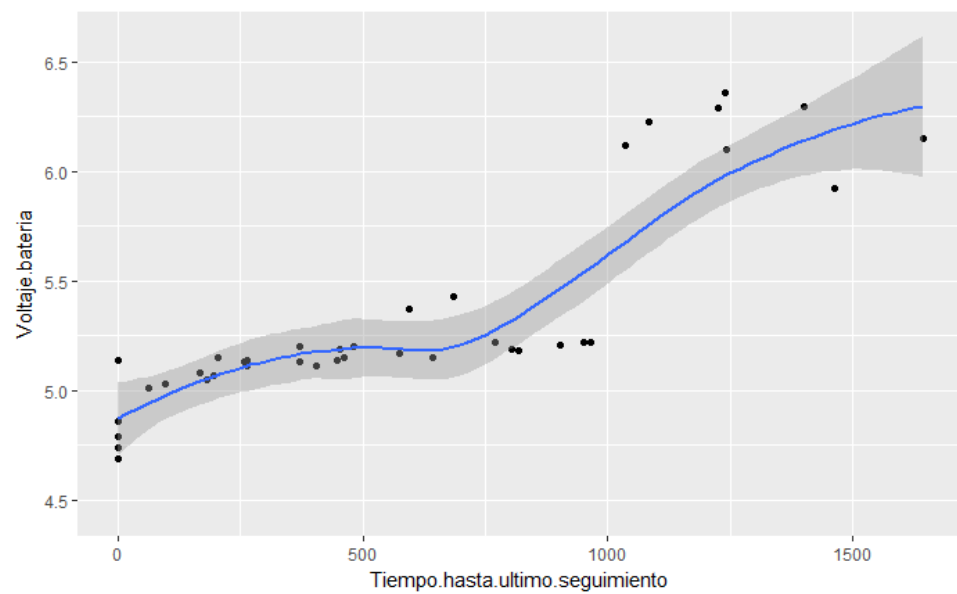
Corregido

## Diapositiva 18

---

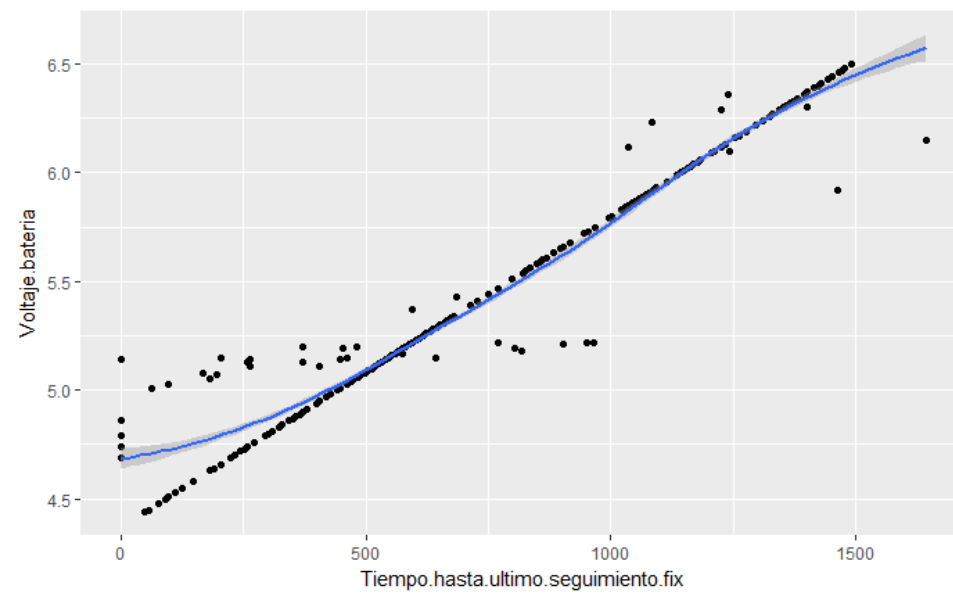
**EC1**

Elena Campillo; 11/04/2017

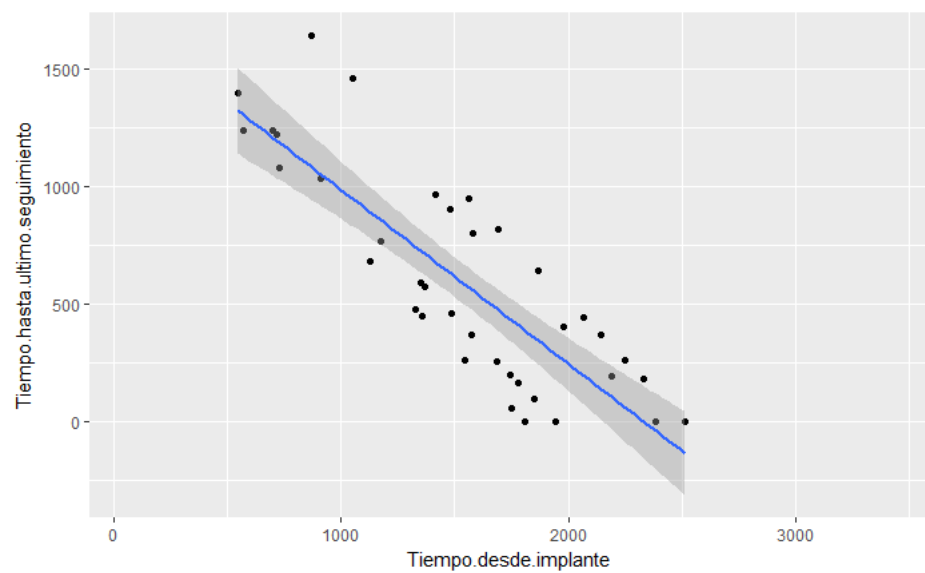


Original

VS

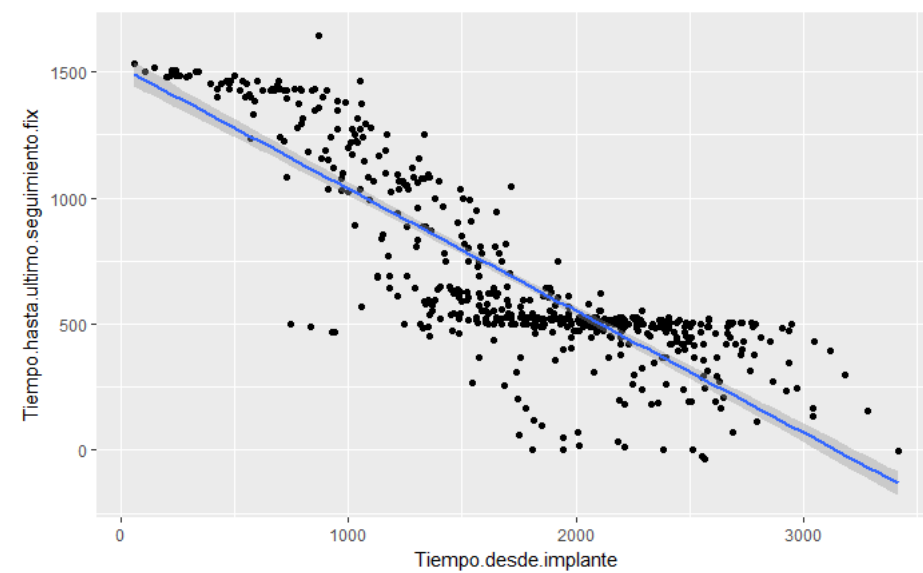


Corregido



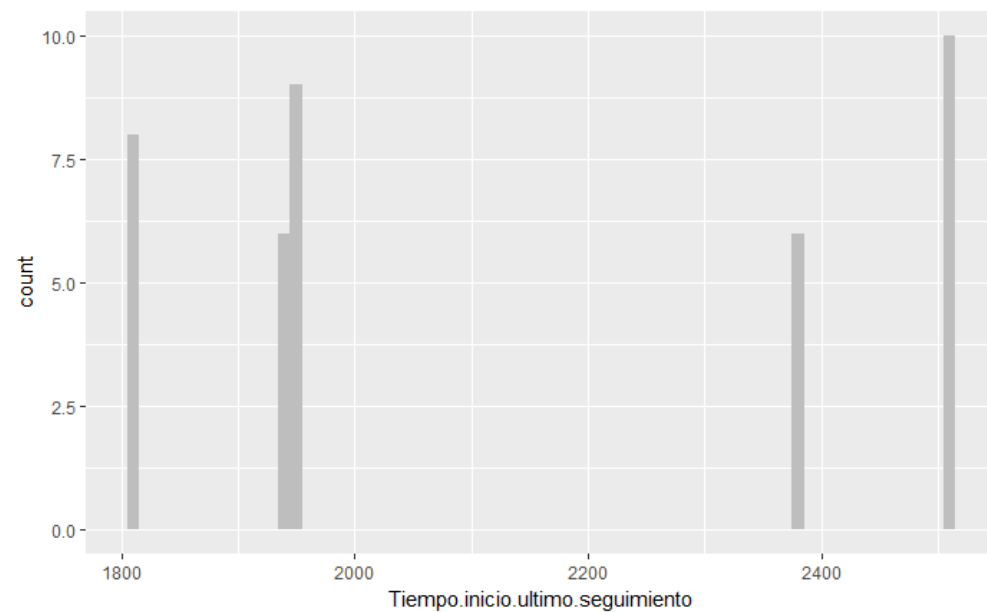
Original

VS



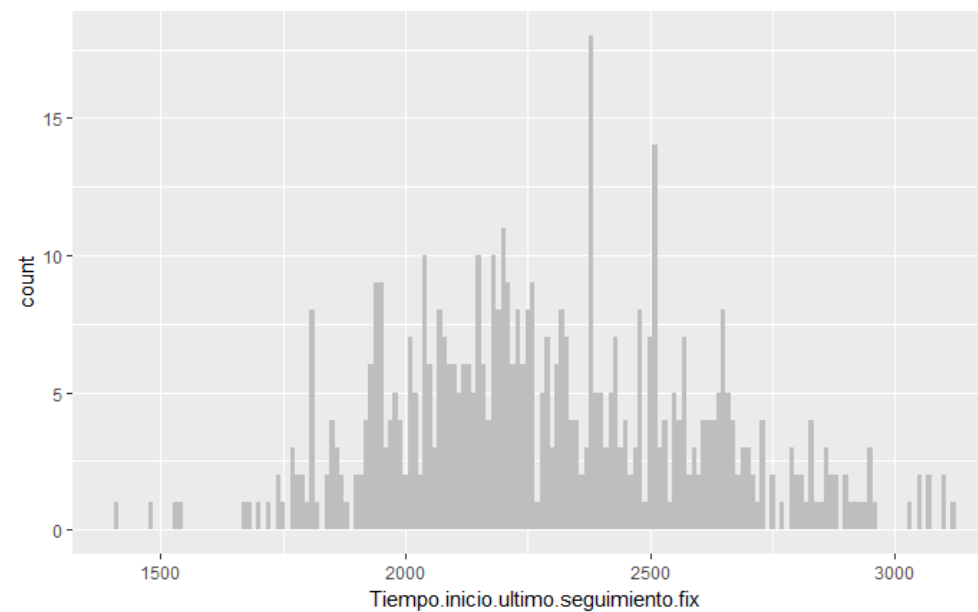
Corregido

# Según el voltaje de la batería y el tiempo que lleva operativo

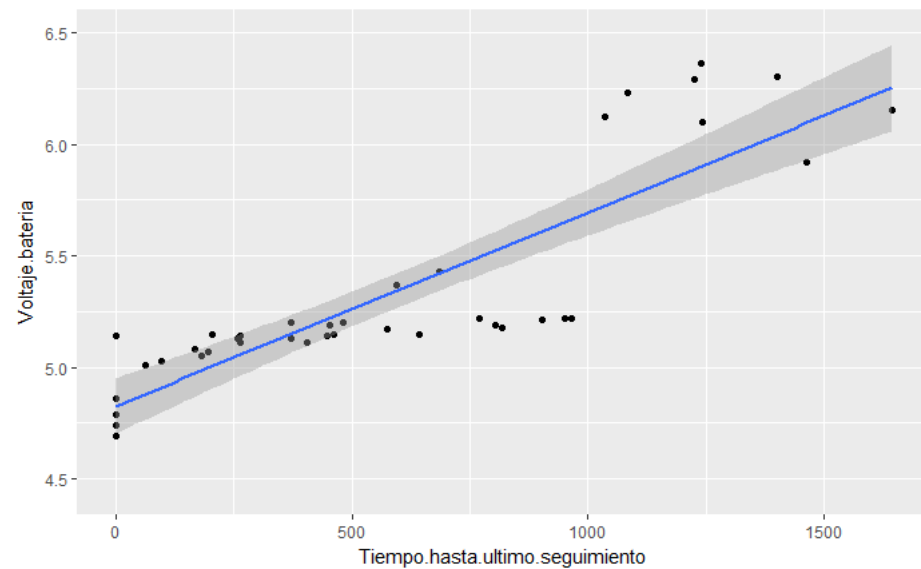


Original

VS

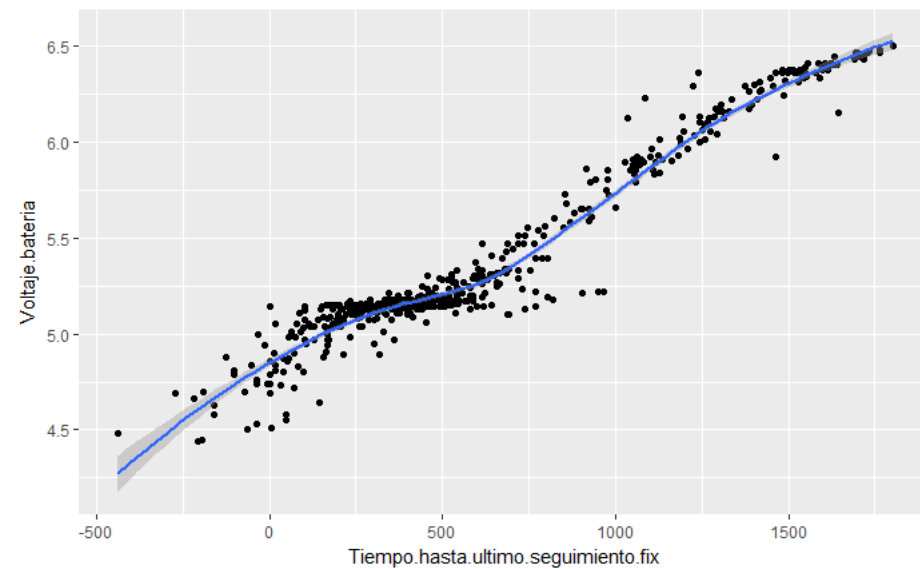


Corregido

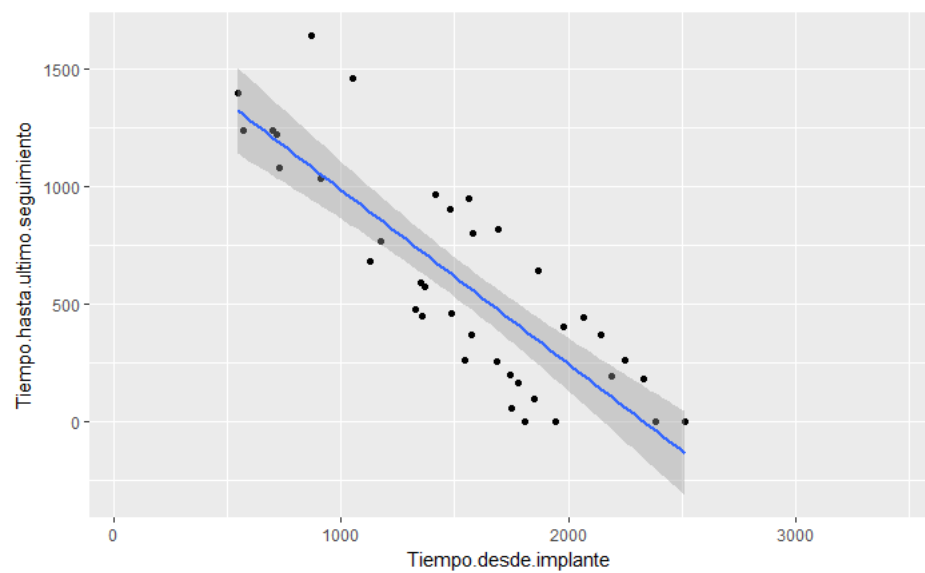


Original

VS

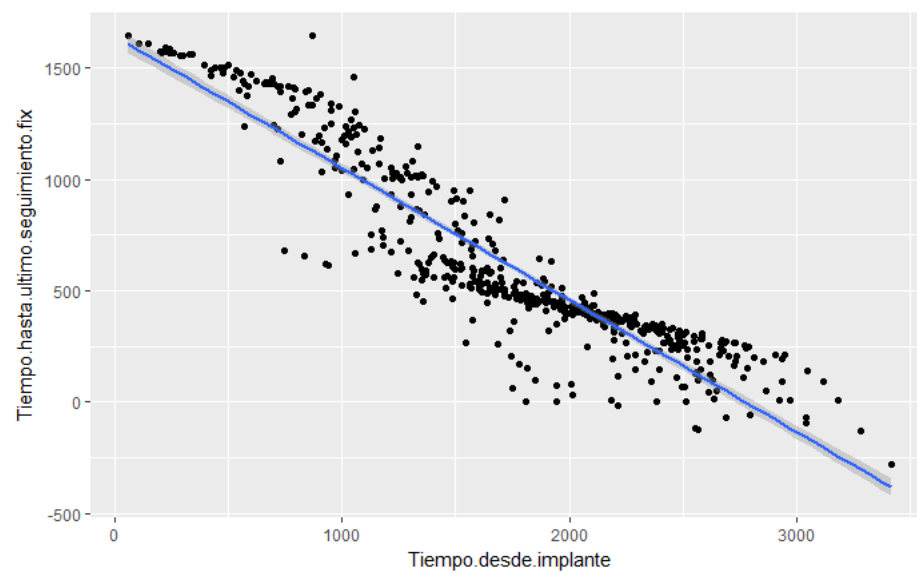


Corregido



Original

VS



Corregido



# **Ventajas de inferir los datos**

- Podemos trabajar con los datos para crear el modelo.
- Nos aseguramos de que se adecúan en cierta medida a la realidad.

# **Desventajas de inferir los datos**

- Los resultados no son estadísticamente probables.