



## Trabajo Práctico Nro. 3

### Regresión Lineal

Presentado en la fecha: 28/09/2019

Hecho por: Andrada Alexander

Encina Guadalupe

Huarca Brian

# Contents

<b>Introducción</b>	<b>2</b>
<b>Objetivo</b>	<b>3</b>
<b>Regresión Lineal</b>	<b>4</b>
0.1 Datos Iniciales . . . . .	4
0.2 Elección del modelo . . . . .	6
0.2.1 Modelo 1 . . . . .	6
0.2.2 Modelo 2 . . . . .	9
0.2.3 Modelos con FORWARD, BACKWARD y STEPWISE . . . . .	10
0.3 Analisis de Colinealidad/Multicolinealidad . . . . .	12
Analisis de Factor de Inflación de Varianza (VIF) . . . . .	12
0.4 Distribución normal de residuos . . . . .	14
Distribución normal de los residuos . . . . .	14
<b>Conclusión</b>	<b>17</b>
<b>Anexo</b>	<b>18</b>

# Introducción

En el presente trabajo desarrollaremos el proceso llevado a cabo para la realización de la **regresión lineal** de un dataset que armamos de acuerdo a nuestras necesidades a partir de un original obtenido del sitio web *kaggle.com*. A lo largo del desarrollo, detallaremos los pasos que seguimos para la realización de la regresión lineal mediante el uso de la herramienta RStudio.

El conjunto en cuestión se trata de una representación de las variaciones en la temperatura de un motor sincrónico de imán permanente (PMSM).

Los motores de imán permanente son extensivamente usados en servomotores, accionamientos eléctricos para posicionamiento, robótica, máquinas herramienta, ascensores, etc. Incluso son utilizados para accionar submarinos propulsar trenes y son utilizados en vehículos eléctricos.

## Objetivo

Se trata de, a partir del conjunto elegido y armado, representar los datos en un gráfico a partir de la recta obtenida por el cálculo de regresión lineal, con el fin de poder realizar un análisis visual de los resultados y poder llegar a algún tipo de conclusión.

# Regresión Lineal

## 0.1 Datos Iniciales

La Figura 1 corresponde a los datos originales que obtuvimos de la página: Electric Motor Temperature | Kaggle

pmsm\_temperature\_data.csv (126.98 MB)

	# ambient	# coolant	# u_d	# u_q	# motor_s...	# torque	# i_d	# i_q	# pm	# stator_...	# stator_L...	# stator_...	profile_id
1	-0.75214297	-1.1184461	0.3279352	-1.2978575	-1.2224282	-0.2501821	1.0295724	-0.24586003	-2.522071	-1.8314217	-2.0661428	-2.0180326	4
2	-0.77126324	-1.1170286	0.3296648	-1.2976865	-1.2224293	-0.2491333	1.029509	-0.24583231	-2.5224178	-1.8309687	-2.0648587	-2.0176313	4
3	-0.78289163	-1.1166813	0.3327715	-1.3018217	-1.2224278	-0.24943107	1.0294477	-0.24581794	-2.5226731	-1.8304	-2.064073	-2.0173435	4
4	-0.78093535	-1.1167642	0.3336999	-1.301852	-1.2224301	-0.24863635	1.0328449	-0.2469548	-2.521639	-1.8303328	-2.0631368	-2.0176322	4
5	-0.7748426	-1.116775	0.3352061	-1.303118	-1.2224286	-0.24870083	1.0318071	-0.24660969	-2.5219002	-1.8304977	-2.0627947	-2.0181448	4
6	-0.7629362	-1.1169548	0.33490124	-1.3030168	-1.2224286	-0.248197	1.0318089	-0.24634062	-2.522203	-1.8319309	-2.0625494	-2.017884	4
7	-0.74922806	-1.1161705	0.33501354	-1.3020816	-1.2224296	-0.24791418	1.0304929	-0.24616154	-2.5225377	-1.8330117	-2.0621152	-2.0172427	4
8	-0.7384499	-1.1139084	0.3362563	-1.3051548	-1.2224321	-0.24832098	1.0301074	-0.24603489	-2.5228438	-1.8321822	-2.0619526	-2.0172133	4
9	-0.7309097	-1.1118276	0.3349053	-1.3037902	-1.2224315	-0.24778472	1.0298513	-0.24598087	-2.5228078	-1.8315759	-2.062443	-2.0177386	4
10	-0.72712964	-1.1094859	0.3359881	-1.3056333	-1.2224314	-0.24829444	1.0296358	-0.24588774	-2.5226767	-1.8314383	-2.062317	-2.0181001	4
11	-0.72371286	-1.1082836	0.33539978	-1.3045602	-1.2224283	-0.24791297	1.029509	-0.24583231	-2.5226302	-1.8314928	-2.0620575	-2.0176919	4
12	-0.71775365	-1.1085879	0.33443072	-1.3043374	-1.2224288	-0.24772124	1.0293863	-0.24580358	-2.5226388	-1.8318189	-2.0622487	-2.017435	4
13	-0.7048766	-1.1099857	0.33624125	-1.30582	-1.2224338	-0.24800174	1.0293179	-0.2457798	-2.5222142	-1.8320519	-2.0621185	-2.0175893	4

Figure 1: Datos Iniciales

En el conjunto de datos podemos observar 13 variables, las cuales detallaremos a continuación:

- *ambient* : temperatura del ambiente medida por un sensor térmico cercano al stator.
- *coolant* : Temperatura del refrigerante. El motor utiliza refrigeración líquida y la medición se realiza a la salida.
- *u\_dVoltage* : d-component.

- *u\_qVoltage* : q-component.
- *motor\_speedMotor* : velocidad.
- *torque* : fuerza del motor inducida por la corriente.
- *i\_dCurrent* : d-component.
- *i\_qCurrent* : q-component.
- *pmPermanent* : Temperatura de la superficie del imán, en representación de la temperatura del rotor. Se midió con una unidad de termografía infrarroja.
- *stator\_yokeStator* : Temperatura del yugo del estator medida por un sensor térmico.
- *stator\_toothStator* : Temperatura de los dientes del estator medida por un sensor térmico.
- *stator\_windingStator* : Temperatura del bobinado del estator medida por un sensor térmico.
- *profile\_idEach* : identificador de cada conjunto de mediciones.

A partir de este conjunto es que definiremos el mejor modelo para la regresión lineal, mediante el proceso que detallaremos en las siguientes páginas.

## 0.2 Elección del modelo

Para elegir el mejor modelo, realizamos la búsqueda de 5 maneras diferentes.

### 0.2.1 Modelo 1

El primer modelo está hecho de manera intuitiva, ya que podemos deducir que para conocer la temperatura del rotor consideramos las variables *speedMotor* y *coolant*.

```
> #Metodo 1
> #veo la correlacion de las variables con el metodo de Pearson para variables vistas al ojo
> mtcars.cor = cor(DF, method = "pearson")
> round(mtcars.cor, digits = 3)
```

	ambient	coolant	u_d	u_q	motor_speed	torque	i_d	i_q	pm	stator_yoke	stator_tooth
ambient	1.000	0.434	0.195	0.087	0.078	-0.262	0.006	-0.261	0.501	0.452	0.397
coolant	0.434	1.000	0.179	0.028	-0.033	-0.190	0.108	-0.186	0.431	0.874	0.689
u_d	0.195	0.179	1.000	-0.027	-0.234	-0.821	0.359	-0.797	-0.083	0.041	-0.066
u_q	0.087	0.028	-0.027	1.000	0.717	-0.037	-0.182	-0.026	0.101	0.106	0.149
motor_speed	0.078	-0.033	-0.234	0.717	1.000	0.025	-0.723	0.006	0.332	0.183	0.334
torque	-0.262	-0.190	-0.821	-0.037	0.025	1.000	-0.239	0.997	-0.073	-0.092	-0.011
i_d	0.006	0.108	0.359	-0.182	-0.723	-0.239	1.000	-0.204	-0.299	-0.180	-0.388
i_q	-0.261	-0.186	-0.797	-0.026	0.006	0.997	-0.204	1.000	-0.086	-0.099	-0.025
pm	0.501	0.431	-0.083	0.101	0.332	-0.073	-0.299	-0.086	1.000	0.695	0.768
stator_yoke	0.452	0.874	0.041	0.106	0.183	-0.092	-0.180	-0.099	0.695	1.000	0.950
stator_tooth	0.397	0.689	-0.066	0.149	0.334	-0.011	-0.388	-0.025	0.768	0.950	1.000
stator_winding	0.302	0.509	-0.151	0.125	0.393	0.081	-0.540	0.061	0.730	0.845	0.966
profile_id	0.385	0.500	0.301	-0.122	-0.166	-0.257	0.142	-0.256	0.157	0.398	0.281

```

stator_winding profile_id
ambient          0.302    0.385
coolant          0.509    0.500
u_d             -0.151    0.301
u_q             0.125   -0.122
motor_speed      0.393   -0.166
torque           0.081   -0.257
i_d             -0.540    0.142
i_q             0.061   -0.256
pm              0.730    0.157
stator_yoke      0.845    0.398
stator_tooth     0.966    0.281
stator_winding   1.000    0.182
profile_id       0.182    1.000
> |
```

Figure 2: Correlación de variables

Mediante el método de Pearson obtuvimos la correlación de las variables (Figure 2). Esto puede observarse más claramente en las Figuras 3 y 4, donde graficamos las matrices de correlación, histograma y dispersión.

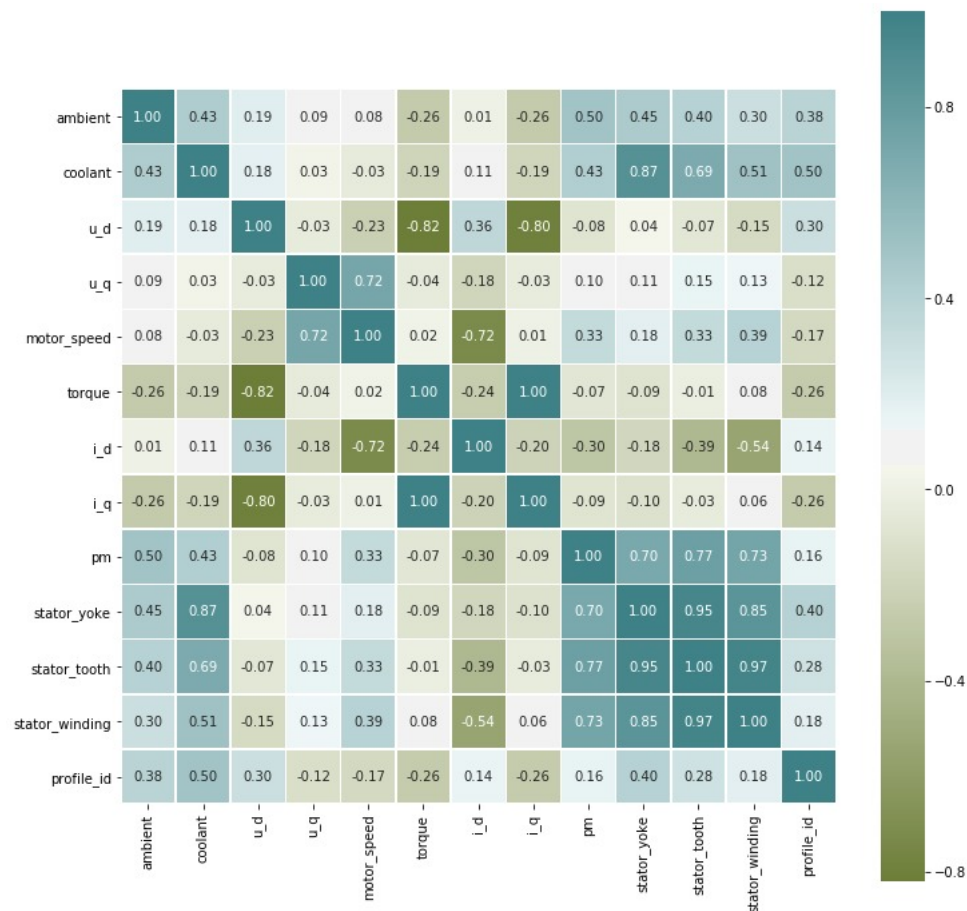


Figure 3: Matriz de correlación

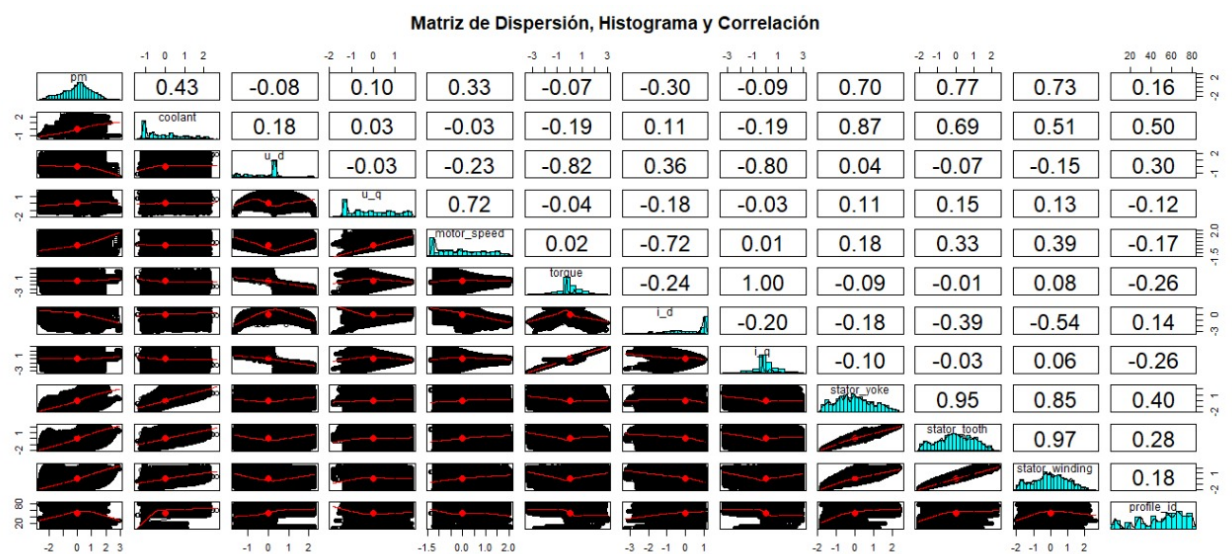


Figure 4: Matriz de dispersión, histograma y correlación de los datos originales



De igual manera, podemos observar en la Figura 5 la matriz resultante de la obtención del modelo 1, que sólo cuenta con las 3 variables mencionadas anteriormente.

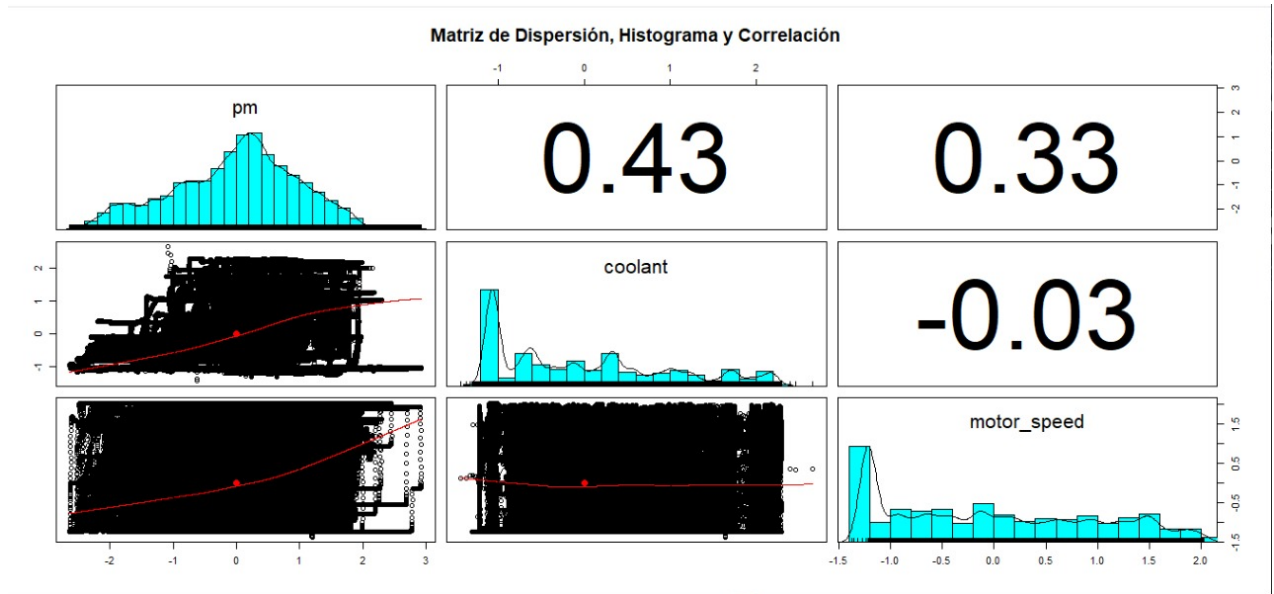


Figure 5: Matriz de dispersión, histograma y correlación del modelo 1

Luego mediante el método SUMMARY podemos observar los resultados del modelo (Figure 5), que nos arrojan un  $p$  -  $value$  menor al nivel de significancia, lo cual nos permite rechazar la hipótesis nula, pero un  $R$  -  $squared$  ajustado de 0.3058. Como este último es muy bajo, podemos continuar buscando un modelo mejor.

```
> summary(RM1)
call:
lm(formula = pm ~ coolant + motor_speed)

Residuals:
    Min       1Q   Median       3Q      Max
-2.7532  -0.5487   0.0033   0.5466   3.6493

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0042824  0.0008304  -5.157 2.51e-07 ***
coolant      0.4391478  0.0008289  529.812 < 2e-16 ***
motor_speed  0.3452340  0.0008299  416.013 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8296 on 998067 degrees of freedom
Multiple R-squared:  0.3058,    Adjusted R-squared:  0.3058
F-statistic: 2.198e+05 on 2 and 998067 DF,  p-value: < 2.2e-16

> |
```

Figure 6: Summary del modelo 1

## 0.2.2 Modelo 2

El segundo modelo lo obtenemos a partir de la observación del gráfico de Pearson para observar la correlación entre las variables (Figura 3). A partir de esta observación podemos ver que las variables correspondientes al estator (*stator\_yokeStator*, *stator\_toothStator*, *stator\_winding*) nos arrojan un porcentaje mayor a un 70%. En la Figura 7 se observa mejor lo antedicho. La elección continúa siendo muy arbitraria, no obstante, si realizamos un summary (Figura 8) podemos ver que esta vez nos arroja un  $r - squared$  ajustado de 0.7077.

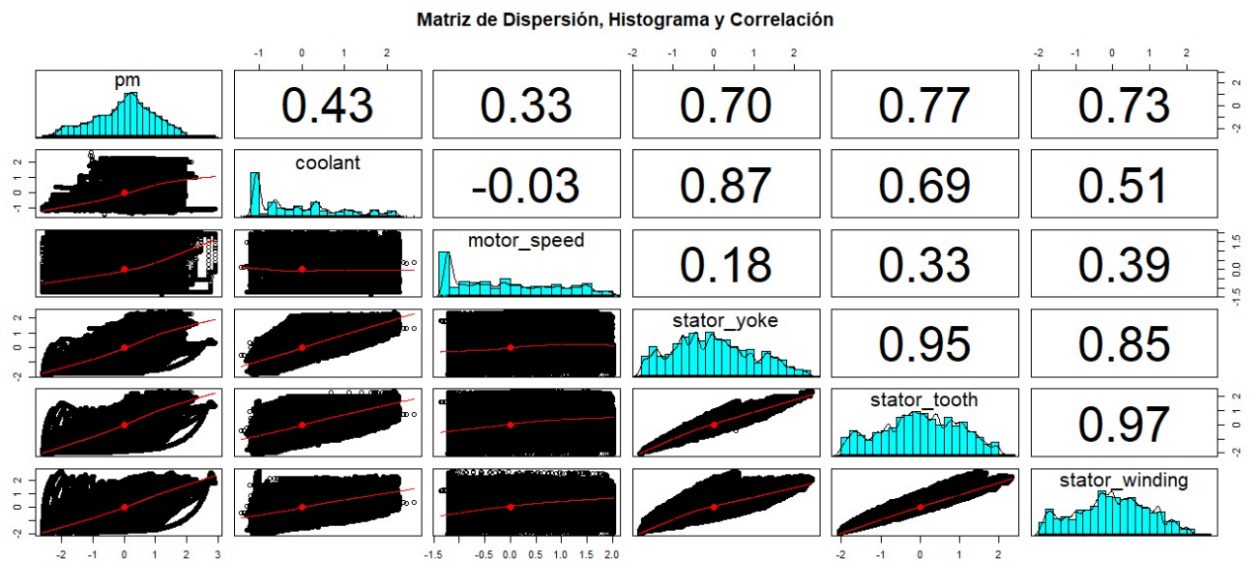


Figure 7: Matriz de dispersión del modelo 2

```
> RM2 = lm(pm~coolant+motor_speed+stator_yoke+stator_tooth+stator_winding)
> summary(RM2)

Call:
lm(formula = pm ~ coolant + motor_speed + stator_yoke + stator_tooth +
    stator_winding)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3467 -0.3368 -0.0069  0.3155  2.8389

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0013497  0.0005388   -2.505  0.0122 *
coolant      -0.0273215  0.0033814  -8.080 6.48e-16 ***
motor_speed  -0.0722185  0.0007214 -100.107 < 2e-16 ***
stator_yoke  -2.4429218  0.0103192 -236.736 < 2e-16 ***
stator_tooth  5.8183062  0.0120278  483.738 < 2e-16 ***
stator_winding -2.7851251  0.0050982 -546.294 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5383 on 998064 degrees of freedom
Multiple R-squared:  0.7077, Adjusted R-squared:  0.7077
F-statistic: 4.834e+05 on 5 and 998064 DF, p-value: < 2.2e-16

> |
```

Figure 8: Summary del modelo 2

### 0.2.3 Modelos con FORWARD, BACKWARD y STEPWISE

Para el modelo 3 utilizamos el método **forward**, que parte de un modelo vacío para ir agregando variables. En la figura 9 podemos observar que al hacer un summary nos arroja un  $r - squared$  ajustado de 0.774.

```
> summary(RLM.Fordward3)

Call:
lm(formula = pm ~ stator_tooth + ambient + coolant + stator_winding +
    u_q + stator_yoke + motor_speed + i_d + profile_id + u_d +
    torque + i_q, data = DF)

Residuals:
    Min       1Q   Median       3Q      Max
-2.74464 -0.30069 -0.00271  0.27996  2.30426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.210e-01  1.431e-03   84.511 < 2e-16 ***
stator_tooth  4.525e+00  1.186e-02  381.655 < 2e-16 ***
ambient       2.289e-01  5.825e-04  392.974 < 2e-16 ***
coolant      -2.439e-01  3.045e-03  -80.105 < 2e-16 ***
stator_winding -2.266e+00  5.873e-03 -385.917 < 2e-16 ***
u_q          -3.465e-01  1.060e-03 -326.880 < 2e-16 ***
stator_yoke  -1.547e+00  9.410e-03 -164.398 < 2e-16 ***
motor_speed   3.280e-01  1.734e-03  189.177 < 2e-16 ***
i_d           1.768e-01  1.409e-03  125.514 < 2e-16 ***
profile_id    -2.406e-03  2.665e-05  -90.276 < 2e-16 ***
u_d           -2.329e-02  1.076e-03  -21.653 < 2e-16 ***
torque        5.401e-02  7.711e-03   7.004 2.49e-12 ***
i_q           -3.451e-02  7.234e-03   -4.771 1.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 998057 degrees of freedom
Multiple R-squared:  0.774,    Adjusted R-squared:  0.774
F-statistic: 2.849e+05 on 12 and 998057 DF, p-value: < 2.2e-16
```

Figure 9: Summary del modelo 3

Para los modelos 4 y 5 aplicamos los métodos **backward**, que parte del conjunto completo de variables y elimina las menos importantes, y **stepwise**, que combina los dos métodos anteriores. En las Figuras 10 y 11 podemos observar que ambos modelos nos arrojan valores similares.

```

> summary(RLM.Backward3)

Call:
lm(formula = pm ~ ambient + coolant + u_d + u_q + motor_speed +
    torque + i_d + i_q + stator_yoke + stator_tooth + stator_winding +
    profile_id, data = DF)

Residuals:
    Min       1Q   Median       3Q      Max
-2.74464 -0.30069 -0.00271  0.27996  2.30426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.210e-01  1.431e-03   84.511 < 2e-16 ***
ambient      2.289e-01  5.825e-04  392.974 < 2e-16 ***
coolant      -2.439e-01  3.045e-03  -80.105 < 2e-16 ***
u_d          -2.329e-02  1.076e-03  -21.653 < 2e-16 ***
u_q          -3.465e-01  1.060e-03 -326.880 < 2e-16 ***
motor_speed   3.280e-01  1.734e-03  189.177 < 2e-16 ***
torque        5.401e-02  7.711e-03    7.004 2.49e-12 ***
i_d           1.768e-01  1.409e-03  125.514 < 2e-16 ***
i_q          -3.451e-02  7.234e-03   -4.771 1.84e-06 ***
stator_yoke   -1.547e+00  9.410e-03 -164.398 < 2e-16 ***
stator_tooth  4.525e+00  1.186e-02  381.655 < 2e-16 ***
stator_winding -2.266e+00  5.873e-03 -385.917 < 2e-16 ***
profile_id    -2.406e-03  2.665e-05  -90.276 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 998057 degrees of freedom
Multiple R-squared:  0.774,    Adjusted R-squared:  0.774
F-statistic: 2.849e+05 on 12 and 998057 DF, p-value: < 2.2e-16

> |

```

Figure 10: Summary del modelo 4

```

> summary(RLM.Stepwise3)

Call:
lm(formula = pm ~ stator_tooth + ambient + coolant + stator_winding +
    u_q + stator_yoke + motor_speed + i_d + profile_id + u_d +
    torque + i_q, data = DF)

Residuals:
    Min       1Q   Median       3Q      Max
-2.74464 -0.30069 -0.00271  0.27996  2.30426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.210e-01  1.431e-03   84.511 < 2e-16 ***
stator_tooth  4.525e+00  1.186e-02  381.655 < 2e-16 ***
ambient      2.289e-01  5.825e-04  392.974 < 2e-16 ***
coolant      -2.439e-01  3.045e-03  -80.105 < 2e-16 ***
stator_winding -2.266e+00  5.873e-03 -385.917 < 2e-16 ***
u_q          -3.465e-01  1.060e-03 -326.880 < 2e-16 ***
stator_yoke   -1.547e+00  9.410e-03 -164.398 < 2e-16 ***
motor_speed   3.280e-01  1.734e-03  189.177 < 2e-16 ***
i_d           1.768e-01  1.409e-03  125.514 < 2e-16 ***
profile_id    -2.406e-03  2.665e-05  -90.276 < 2e-16 ***
u_d          -2.329e-02  1.076e-03  -21.653 < 2e-16 ***
torque        5.401e-02  7.711e-03    7.004 2.49e-12 ***
i_q          -3.451e-02  7.234e-03   -4.771 1.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 998057 degrees of freedom
Multiple R-squared:  0.774,    Adjusted R-squared:  0.774
F-statistic: 2.849e+05 on 12 and 998057 DF, p-value: < 2.2e-16

> |

```

Figure 11: Summary del modelo 5

Para continuar con el trabajo, seleccionamos el modelo 5 realizado con stepwise, y en la sección a continuación explicaremos algunas mejoras que realizamos al mismo.

### 0.3 Analisis de Colinealidad/Multicolinealidad

Con el fin de someter a pruebas al mejor modelo elegido, se busco que no se cumpla con el supuesto de Colinealidad o Multicolinealidad (una variable predictora depende de una variables predictora distinta). Este estudio nos permite determinar si dentro de nuestro modelo hay modelos que expliquen otros modelos. ¿Qué problemas ocasiona?, Ya no pueden interpretarse los coeficientes. Si aumentamos  $X_1$  entonces aumentamos  $X_2$ , "No se pueden aislar efectos". Lo que pueden traer problematicas como por ejemplo que los coeficientes tienen un signo incorrecto o diferente al esperado.

¿Cómo detectar si estamos en presencia de Colinealidad o Multicolinealidad? Para este estudio se va a considerar, en primera instancia, como referencia visual:

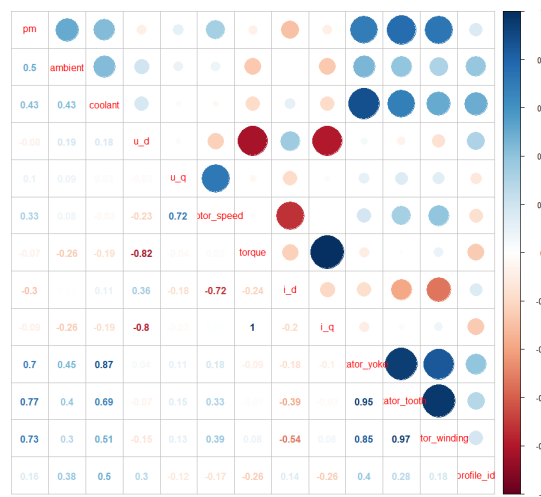


Figure 12: Matriz de Correlacion de Variables

Como segunda medida se a analizar el VIF partiendo de la matriz de correlación entre variables independientes y tomando los siguientes criterio:

- Si hay correlaciones mayores o iguales a 0.9, entonces indican problemas serios de multicolinealidad
- Correlaciones entre 0.5 y 0.9, entonces son problemas moderados o pueden indicar problemas

Partiendo de este modelo se puede observar como medida principal que el R-Cuadrado ajustado nos representa un 0.774, medida de significancia que mejor explica. Se va a

```
> summary(RLM.Stepwise3)

Call:
lm(formula = pm ~ stator_tooth + ambient + coolant + stator_winding +
    u_q + stator_yoke + motor_speed + i_d + profile_id + u_d +
    torque + i_q, data = DF)

Residuals:
    Min       1Q   Median       3Q      Max
-2.74464 -0.30069 -0.00271  0.27996  2.30426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.210e-01  1.431e-03   84.511 < 2e-16 ***
stator_tooth  4.525e+00  1.186e-02  381.655 < 2e-16 ***
ambient      2.289e-01  5.825e-04  392.974 < 2e-16 ***
coolant     -2.439e-01  3.045e-03  -80.105 < 2e-16 ***
stator_winding -2.266e+00  5.873e-03 -385.917 < 2e-16 ***
u_q         -3.465e-01  1.060e-03 -326.880 < 2e-16 ***
stator_yoke  -1.547e+00  9.410e-03 -164.398 < 2e-16 ***
motor_speed   3.280e-01  1.734e-03  189.177 < 2e-16 ***
i_d          1.768e-01  1.409e-03  125.514 < 2e-16 ***
profile_id   -2.406e-03  2.665e-05  -90.276 < 2e-16 ***
u_d         -2.329e-02  1.076e-03  -21.653 < 2e-16 ***
torque       5.401e-02  7.711e-03   7.004 2.49e-12 ***
i_q         -3.451e-02  7.234e-03   -4.771 1.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 998057 degrees of freedom
Multiple R-squared:  0.774,    Adjusted R-squared:  0.774
F-statistic: 2.849e+05 on 12 and 998057 DF,  p-value: < 2.2e-16

> |
```

Figure 13: Resumen del modelo

Aplicar VIF y a tratar de conservar la mayor significancia posible teniendo en cuenta los criterios mencionados

```
> vif(modelo)
      ambient      coolant      u_d      u_q  motor_speed      torque      i_d
      1.491004      41.500802      5.132772      5.029429      13.425392      263.800083      8.824513
      i_q  stator_yoke  stator_tooth  stator_winding  profile_id
      232.135474      395.315274      625.634149      153.127542      1.541900

> |
```

Figure 14: VIF1

El r-cuadrado inicial parte de un: 0.774

Me pide quitar: statorwinding, statoryoke, statortooth, iq, torque, coolant, motor-speed. SIEMPRE ELIMINO LA QUE MAYOR ESCALA TENGA

Orden de Eliminacion

STATORTOOTH: 625.634149 TORQUE: 262.824467 STATORYoke: 73.524228

Culminando la eliminacion de variables obtenemos el siguiente VIF final:

```
> vif(modelo1)
      ambient      coolant      u_d      u_q  motor_speed      i_d      i_q
      1.383457      2.348006      3.703333      4.844883      10.735217      6.263040      3.725995
      stator_winding  profile_id
      2.705360      1.529578

> |
```

Figure 15: VIF2



Conclusion del mejor modelo: r-cuadrado: 0.7014

```
> modeloi = lm(pm~ambient+coolant+u_d+u_q+motor_speed+i_d+i_q+stator_winding+profile_id)
> summary(modeloi)#r-cuadrado: 0.7014

Call:
lm(formula = pm ~ ambient + coolant + u_d + u_q + motor_speed +
    i_d + i_q + stator_winding + profile_id)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3966 -0.3374  0.0061  0.3358  2.5111

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.501e-01  1.639e-03   91.59  <2e-16 ***
ambient      3.155e-01  6.450e-04  489.22  <2e-16 ***
coolant     -9.212e-02  8.325e-04 -110.66  <2e-16 ***
u_d         -9.753e-02  1.050e-03  -92.86  <2e-16 ***
u_q         -4.182e-01  1.196e-03 -349.69  <2e-16 ***
motor_speed  6.841e-01  1.782e-03  383.84  <2e-16 ***
i_d          5.810e-01  1.364e-03  425.88  <2e-16 ***
i_q         -6.029e-02  1.053e-03  -57.23  <2e-16 ***
stator_winding 7.778e-01  8.972e-04  866.88  <2e-16 ***
profile_id   -2.978e-03  3.051e-05  -97.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5441 on 998060 degrees of freedom
Multiple R-squared:  0.7014,    Adjusted R-squared:  0.7014
F-statistic: 2.605e+05 on 9 and 998060 DF,  p-value: < 2.2e-16

> vif(modeloi)
          ambient          coolant           u_d           u_q      motor_speed           i_d           i_q
1.383457      2.348006      3.703333      4.844883      10.735217      6.263040      3.725995
stator_winding      profile_id
2.705360          1.529578
```

Figure 16: Resumen del modelo

## 0.4 Distribución normal de residuos

Tomando en cuenta la información que devuelve la función `summary` luego de seleccionar con el metodo de selección de variable Stepwise y El mejor modelo (despues de eliminar variables predictoras). Como puede observarse el error estandar residual es pequeño, este, se identifica como una medida absoluta ante la falta de ajuste del modelo. Cuanto menor sea el mismo, mejor se ajusta al modelo. A continuación se muestra su distribución normal:

Distribución normal para el modelo mejorado:

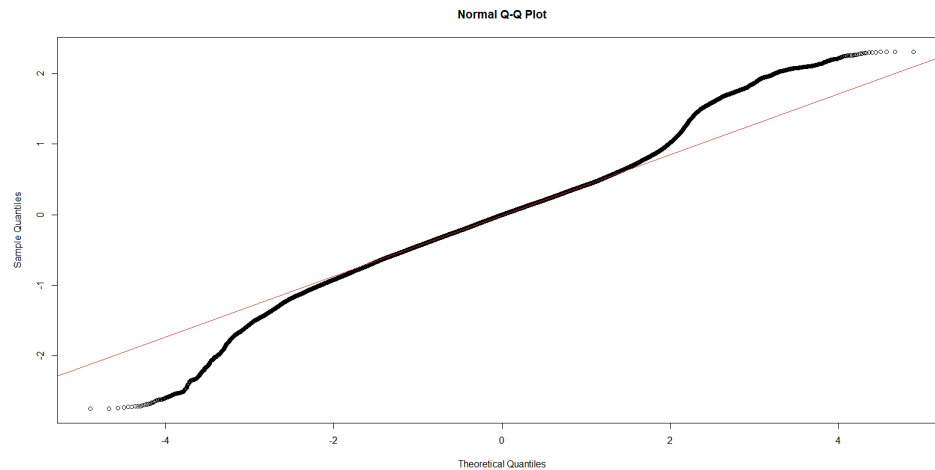


Figure 17: Distribución normal de los residuos

```
> summary(RLM.Stepwise3)

Call:
lm(formula = pm ~ stator_tooth + ambient + coolant + stator_winding +
    u_q + stator_yoke + motor_speed + i_d + profile_id + u_d +
    torque + i_q, data = DF)

Residuals:
    Min       1Q   Median       3Q      Max
-2.74464 -0.30069 -0.00271  0.27996  2.30426

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.210e-01  1.431e-03   84.511 < 2e-16 ***
stator_tooth  4.525e+00  1.186e-02  381.655 < 2e-16 ***
ambient      2.289e-01  5.825e-04  392.974 < 2e-16 ***
coolant     -2.439e-01  3.045e-03  -80.105 < 2e-16 ***
stator_winding -2.266e+00  5.873e-03 -385.917 < 2e-16 ***
u_q         -3.465e-01  1.060e-03 -326.880 < 2e-16 ***
stator_yoke  -1.547e+00  9.410e-03 -164.398 < 2e-16 ***
motor_speed   3.280e-01  1.734e-03  189.177 < 2e-16 ***
i_d          1.768e-01  1.409e-03  125.514 < 2e-16 ***
profile_id   -2.406e-03  2.665e-05  -90.276 < 2e-16 ***
u_d         -2.329e-02  1.076e-03  -21.653 < 2e-16 ***
torque       5.401e-02  7.711e-03    7.004 2.49e-12 ***
i_q         -3.451e-02  7.234e-03   -4.771 1.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4733 on 998057 degrees of freedom
Multiple R-squared:  0.774,    Adjusted R-squared:  0.774
F-statistic: 2.849e+05 on 12 and 998057 DF, p-value: < 2.2e-16

> |
```

Figure 18: Resumen del Modelo



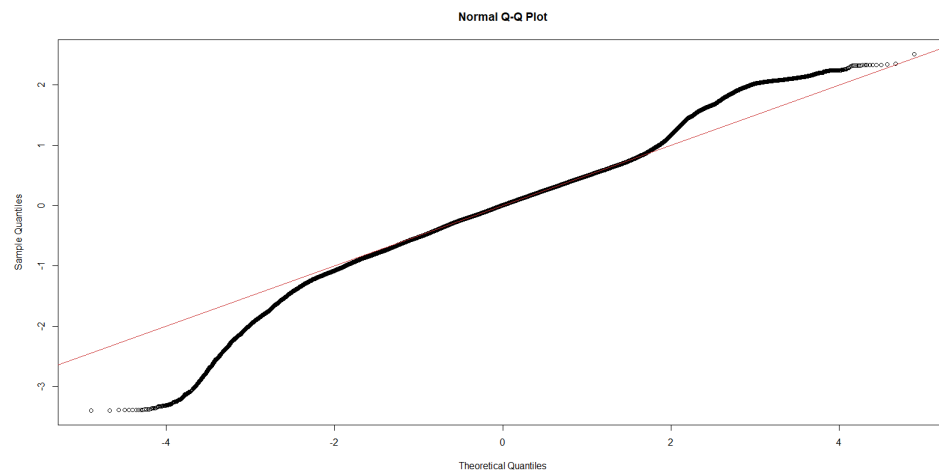


Figure 19: Distribución normal de los residuos

```

> model1 = lm(pm~ambient+coolant+u_d+u_q+motor_speed+i_d+i_q+stator_winding+profile_id)
> summary(model1)#r-cuadrado: 0.7014

Call:
lm(formula = pm ~ ambient + coolant + u_d + u_q + motor_speed +
    i_d + i_q + stator_winding + profile_id)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3966 -0.3374  0.0061  0.3358  2.5111

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.501e-01  1.639e-03   91.59  <2e-16 ***
ambient      3.155e-01  6.450e-04  489.22  <2e-16 ***
coolant      -9.212e-02  8.325e-04 -110.66  <2e-16 ***
u_d          -9.753e-02  1.050e-03  -92.86  <2e-16 ***
u_q          -4.182e-01  1.196e-03 -349.69  <2e-16 ***
motor_speed   6.841e-01  1.782e-03  383.84  <2e-16 ***
i_d           5.810e-01  1.364e-03  425.88  <2e-16 ***
i_q          -6.029e-02  1.053e-03  -57.23  <2e-16 ***
stator_winding 7.778e-01  8.972e-04  866.88  <2e-16 ***
profile_id   -2.978e-03  3.051e-05  -97.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5441 on 998060 degrees of freedom
Multiple R-squared:  0.7014,    Adjusted R-squared:  0.7014
F-statistic: 2.605e+05 on 9 and 998060 DF, p-value: < 2.2e-16

> vif(model1)
          ambient          coolant           u_d           u_q      motor_speed           i_d           i_q
stator_winding 1.383457      2.348006      3.703333      4.844883      10.735217      6.263040      3.725995
profile_id      2.705360      1.529578
> |

```

Figure 20: Resumen del Modelo

## Conclusión

Luego de desarrollar el análisis, y haciendo énfasis en el uso de herramientas estadísticas y funciones, concluimos que el Análisis de Regresión Lineal nos permite realizar un mejor estudio, desarrollo y optimización predictivo de la Regresión Multilineal.

Al mejorar y someter a pruebas el modelo de Regresión Multilineal se puede observar que La Temperatura del Rotor (PM) presenta una alta relación y correlación con respecto a sus variables predictoras. Esta relación estimó un  $r$ -cuadrado de 0.7014, indicando así una relación considerable para nuestro análisis .

# Anexo

Codigo para el Analisis de Obtencion del Mejor Modelo

```
1 library(readr)
2 library(readxl)
3 library(openxlsx)
4 library(car)
5 library(leaps) #Para lm y probar todas las combinaciones
6 library(xtable)#Para reportar tablas en LaTeX para informes
7 library(GGally)
8 library(corrplot)
9 library(psych)
10 library(foreign)
11 library(nortest)
12
13
14 #Dataframe Original sin Filtrar
15 DF = read.csv("C:/Users/brian/Documents/RegresionMultiple/pmsm_temperature_data.csv",
16             header = TRUE, sep = ",")
17
18 #Debug del dataset
19 View(DF)
20 names(DF)
21 options(scipen=6) #para evitar notacion cientifica
22
23 #Veo la Correlacion de las Variables con el metodo de Pearson
24 mtcars.cor = cor(DF, method = "pearson")
25 round(mtcars.cor, digits = 3)
26
27 #Veo correlaciones entre las variables (Grafico) y elijo---Veo todas las combinaciones
28 x11(15)
29 pairs.panels(x=DF, ellipses = F, lm=T, method = "pearson")
30
31 #Modelo 1. Se deduce a partir de variables tomadas arbitrariamente
32 RM1 = lm(pm~ coolant+motor_speed)
33 summary(RM1)
34
35 #DataFrame modelo1
36 DF1 = select(DF1, pm, coolant, motor_speed)
37 View(DF1)
38
39 #Grafico De Correlacion e Histograma para el PRIMER MODELO
40 x11(15)
```

```
40 chart.Correlation(DF1, histogram = T, pch = 19)
41 x11(15)
42 pairs.panels(DF1, pch=21,main="Matriz de Dispersin, Histograma y Correlacin")
43
44 #Modelo 2 -- viendo las correlaciones del diagrama de Pearson
45 DF2 = select(DF, pm, coolant, motor_speed, stator_yoke, stator_tooth, stator_winding)
46 View(DF2)
47
48 #Grafico de Correlacion del 2do Modelo
49 x11(15)
50 pairs.panels(DF2, pch=21,main="Matriz de Dispersin, Histograma y Correlacin")
51
52 attach(DF2)
53 RM2 = lm(pm~coolant+motor_speed+stator_yoke+stator_tooth+stator_winding)
54 summary(RM2)
55
56
57 #Redefino el DF poniendo como primer variable a la DEpendiente
58 DF3 = select(DF, pm, ambient, coolant, u_d, u_q, motor_speed, torque, i_d, i_q, stator_
  yoke, stator_tooth, stator_winding, profile_id)
59 View(DF3)
60 names(DF3)
61
62 #Grafico de Correlacion Teniendo en cuenta todas la Variables y apuntando a pm como
  variable dependiente
63 x11(15)
64 pairs.panels(DF3, pch=21,main="Matriz de Dispersin, Histograma y Correlacin")
65
66 #r-cuadrado del modelo general
67 attach(DF3)
68 RM3 = lm(pm~., data=DF3)
69 summary(RM3)
70 names(DF3)
71 #####INTERVALO DE CONFIANZA DEL GENERAL#####
72 confint(RM3)
73 #####INTERVALO DE CONFIANZA DEL GENERAL#####
74
75 ###Para todos los modelos#Metodos de Seleccin de Variables
76 #RLM Vacio
77 RLM.Vacio3=lm(formula = pm~1,DF3)
78 summary(RLM.Vacio3)
79
80 #RLM Completo
81 RLM.Completo3 = lm(formula=pm~.,DF3)
82 summary(RLM.Completo3)
83
84
85 #Regresion FordWard
86 RLM.Fordward3=step(RLM.Vacio3,
87   scope = list(lower=RLM.Vacio3, upper=RLM.Completo3),
88   direction = "forward")
89 summary(RLM.Fordward3)
90
91
92 #Regresion Backward
```

```

93 RLM.Backward3=step(RLM.Completo3,
94                   scope = list(lower=RLM.Vacio3, upper=RLM.Completo3),
95                   direction = "backward")
96 summary(RLM.Backward3)
97
98
99 #Regresion Stepwise
100 RLM.Stepwise3=step(RLM.Vacio3,
101                   scope = list(lower=RLM.Vacio3, upper=RLM.Completo3),
102                   direction = "both")
103 summary(RLM.Stepwise3)
104
105 #####INTERVALO DE CONFIANZA DEL GENERAL#####
106 confint(RLM.Stepwise3)
107 #####INTERVALO DE CONFIANZA DEL GENERAL#####
108
109 #####
110 #Propiedad 1: Estudio de Colinealidad/Multicolinealidad
111 modelo = lm(pm~., DF3)
112 summary(modelo)#r-cuadrado: 0.774
113 vif(modelo)
114
115 #Me pide quitar: stator_winding, stator_yoke, stator_tooth, i_q, torque, coolant, motor_
    speed,
116 #SIEMPRE ELIMINO LA QUE MAYOR ESCALA TENGA
117
118 #OPCION1--QUITAR STATOR_TOOTH: 625.634149
119 modelo1 = lm(pm~ambient+coolant+u_d+u_q+motor_speed+torque+i_d+i_q+stator_yoke+stator_
    winding+profile_id)
120 summary(modelo1)#r-cuadrado: 0.741
121 vif(modelo1)
122 #Conclusion:
123 #Me pide quitar: coolant, motor_speed, torque, i_q, stator_yoke, stator_winding
124
125
126 #OPCION2--QUITAR torque:262.824467
127 modelo1 = lm(pm~ambient+coolant+u_d+u_q+motor_speed+i_d+i_q+stator_yoke+stator_winding+
    profile_id)
128 summary(modelo1)#r-cuadrado: 0.741
129 vif(modelo1)
130 #Conclusion:
131 #Me pide quitar: coolant, stator_yoke, stator_winding
132
133
134 #OPCION3--QUITAR STATOR_yoke: 73.524228
135 modelo1 = lm(pm~ambient+coolant+u_d+u_q+motor_speed+i_d+i_q+stator_winding+profile_id)
136 summary(modelo1)#r-cuadrado: 0.7014
137 vif(modelo1)
138 #Conclusion:
139 #Me pide quitar: motor_speed: 10.735217 --- ME QUEDO CON 9NUEVE VARIABLES
140
141 #####INTERVALO DE CONFIANZa Mejor Modelo#####
142 confint(modelo1)
143 #####INTERVALO DE CONFIANZa Mejor Modelo#####
144

```

```

145 #####
146 #####Someter a prueba el mejor modelo#####
147
148 #Principio 2: Distribucion normal de residuos
149 #Estudio de Residuos---Normalidad: Tanto el analisis grafico como es test de hipotesis
    confirman la normalidad.
150 #Para modelo general
151 x11(15)
152 qqnorm(resid(RLM.Stepwise3))
153 qqline(resid(RLM.Stepwise3),col=c("brown3"))
154
155 #Estudio con Modelo Mejorado
156 x11(15)
157 qqnorm(resid(modelo1))
158 qqline(resid(modelo1),col=c("brown3"))
159
160 # Dispersion de los residuos
161 residuos.procesador <- rstandard(procesador_modelo)
162 plot(residuos.procesador)
163
164 #####COMO HAY RESIDUOS DISPERSION Y NO COHERENTES, Y UNA FORMA DE COLA
    DISTORSIONADA#####
165 #####Error
    #####
166 #####SE BUSCA APLICAR AJUSTAR Y CORREGIR EL MODELO
    #####
167 #####Error
    #####
168 #####Error
    #####
169
170 #Buscar Lambda:
171 #Y*=Y^Lambda---Para ello buscar el valor de Lambda
172 x11(15)
173 bc=boxcox(modelo1, lambda=seq(-5,5, length = 0.01))#Desde -5 hasta 5.. en 0.01 en 0.01
174 bc=boxcox(RLM.Stepwise3, lambda=seq(-5,5, length = 0.01))#Desde -5 hasta 5.. en 0.01 en
    0.01
175 lambda = bc$x[which.max(bc$y)]
176 lambda
177
178 DF3$pm = with(DF3, pm^lambda)
179
180 modelomejorado <- lm(pm ~ . , DF3)
181 summary(modelomejorado)
182
183 x11(15)
184 qqnorm(resid())
185 qqline(resid(modelomejorado))
186
187 #####Error #####
188 #####Error #####
189 #####Error #####
190
191
192 #Principio 3: de Homoscedasticidad/Estudio de Residuos

```

```
193
194 plot(modelo1)
195 plot(RLM.Stepwise3)
196
197 #Histograma de Residuos--- Modelo Mejorado
198 residuos1 = residuals(modelo1)
199 x11(15)
200 hist(residuos1)
201
202 #Histograma de Residuos--- 5toModelo
203 residuos2 = residuals(RLM.Stepwise3)
204 x11(15)
205 hist(residuos2)
206
207 residuos <- rstandard(modelo1)
208 mean(residuos) # acceptable
209 sum(residuos)
```