

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS ¶  
Maestria en Ciencia de Datos.

Metodos Estadisticos Multivariados  
Reporte Estadistico

MET.Rosa Isela Hernández Zamora

Alumnos: Jesus Emmanuel Ramos Davila  
Marco Antonio Obregon Flores  
Matricula: 1439401, 1723556

Fecha entrega: 03/28/2023

# Indice

## Introducción

El presente informe tiene como objetivo utilizar técnicas estadísticas multivariadas para analizar un conjunto de datos que contiene múltiples mediciones realizadas a un motor síncrono de imán permanente (PMSM). En específico, se utilizarán el análisis de componentes principales (PCA) y el análisis de factores para reducir la dimensionalidad del conjunto de datos y descubrir patrones y relaciones entre las variables. Estas técnicas se aplicarán en R Studio, utilizando los conocimientos adquiridos en el curso de Métodos Estadísticos Multivariados. El conjunto de datos contiene 13 variables distintas, incluyendo la temperatura del refrigerante del motor, la velocidad del motor, el par del motor y la corriente en el eje d del motor, entre otros. El análisis de componentes principales permitirá simplificar el conjunto de datos encontrando mezclas de variables que describen la mayor parte de la variación en los datos, mientras que el análisis de factores buscará reducir la dimensionalidad de los datos para explicar el máximo de información contenida en ellos.

El análisis del motor síncrono de imán permanente es importante para entender su funcionamiento y optimizar su rendimiento. El objetivo del análisis multivariado es identificar patrones y relaciones entre las variables, lo que puede llevar a descubrir factores importantes que afectan el rendimiento del motor. Al reducir la dimensionalidad de los datos mediante el uso de técnicas como el análisis de componentes principales y el análisis de factores, se puede obtener una mejor comprensión de la estructura subyacente de los datos y reducir la complejidad de la información. Esto puede permitir una mejor visualización de los patrones y relaciones, lo que puede conducir a una mejor identificación de los factores clave que afectan el rendimiento del motor. Además, la reducción de la dimensionalidad también puede ayudar a simplificar el análisis y mejorar la eficiencia del procesamiento de datos.

Los datos utilizados en este análisis son reales y fueron obtenidos de Kirgiz, W. (2021). Electric Motor Temperature. Los datos se pueden encontrar en <https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature> y contienen mediciones realizadas a un motor síncrono de imán permanente (PMSM). Es importante destacar que no se eliminaron valores atípicos o faltantes en los datos.

Además, se estandarizaron los datos antes de realizar el análisis, lo que significa que se convirtieron todas las variables a la misma escala para que tengan una media de cero y una desviación estándar de uno. Esto se hizo para que las variables se puedan comparar directamente entre sí y para evitar que una variable tenga más peso en el análisis solo porque tiene valores más grandes.

En resumen, este informe utilizará técnicas estadísticas avanzadas para analizar un conjunto de datos complejo y encontrar patrones y relaciones entre las variables del motor síncrono de imán permanente.

## Análisis descriptivo del conjunto de datos

Los registros corresponden a mediciones realizadas a un motor síncrono de imán permanente (PMSM), los cuales fueron muestreados a una frecuencia de 2 Hz. El conjunto de datos contiene múltiples sesiones de medición, las cuales se pueden distinguir por el identificador de perfil (`profile_id`) y tienen una duración variable de entre una y seis horas. En total, se registraron 185 horas de operación del motor.

El dataset utilizado en este análisis contiene un total de 1,330,816 mediciones realizadas al motor síncrono de imán permanente. Este es un conjunto de datos bastante grande que requiere técnicas estadísticas multivariadas avanzadas para su análisis y comprensión. La cantidad de mediciones en este conjunto de datos proporciona una gran cantidad de información sobre el comportamiento del motor, lo que puede ser útil para identificar patrones y relaciones complejas entre las variables y optimizar su rendimiento.

Cabe destacar que el motor es excitado por ciclos de conducción diseñados a mano, que establecen una velocidad de referencia y un par de referencia. Las corrientes y voltajes en coordenadas d/q son resultado de una estrategia de control estándar que intenta seguir la velocidad y el par de referencia, y las variables de velocidad y torque son las cantidades resultantes logradas por esa estrategia, derivadas de las corrientes y voltajes establecidos. La mayoría de los ciclos de conducción corresponden a caminatas aleatorias en el plano velocidad-par, con el fin de imitar ciclos de conducción del mundo real de manera más precisa que las excitaciones y rampas de subida y bajada constantes.

Table 1: Variables del conjunto de datos

Variable	Descripción
u_q	Tensión en el eje q del motor
coolant	Temperatura del refrigerante del motor
stator_winding	Temperatura del devanado del estator del motor
u_d	Tensión en el eje d del motor
stator_tooth	Temperatura del diente del estator del motor
motor_speed	Velocidad del motor
i_d	Corriente en el eje d del motor
i_q	Corriente en el eje q del motor
pm	Temperatura del imán permanente del motor
stator_yoke	Temperatura del yugo del estator del motor
ambient	Temperatura ambiente durante la medición
torque	Par del motor
profile_id	Identificador de la sesión de medición

## Análisis de Componentes Principales

El análisis de componentes principales (PCA) es un método estadístico que sirve para simplificar y resumir un conjunto de datos con muchas variables. En vez de usar todas las variables originales, el PCA encuentra mezclas de estas variables que capturan la mayor parte de la variación en los datos. Estas mezclas se llaman “componentes principales” y se usan para describir el conjunto de datos de una manera más sencilla y comprensible.

Piensa que tienes muchos datos distintos sobre un motor eléctrico, como la temperatura, el torque, la velocidad, etc. Puede ser complicado saber cómo todos estos datos se relacionan entre sí y qué información es la más relevante. Si el análisis de componentes principales descubre que la temperatura y el torque están muy relacionados, entonces puede juntarlos en un nuevo componente principal que abarque ambas variables a la vez. Así, se puede disminuir el número de variables y hacer que los datos sean más sencillos de entender.

## Análisis de Factores

El Análisis Factorial es, por tanto, una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos.

Para desarrollar el análisis de factores se realizaron pasos previos tales como estandarizar los datos , verificar si los datos cumplen la **normal multivariada**, revisar la **matriz de correlaciones** y realizar **supuestos e hipótesis**.

### Paso 1: Carga de Datos

```
## # A tibble: 1,330,816 x 12
##   u_q coolant stator_win~1 u_d stato~2 motor_~3 i_d i_q pm
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 -0.451 18.8 19.1 -0.350 18.3 2.87e-3 4.42e-3 3.28e-4 24.6
## 2 -0.326 18.8 19.1 -0.306 18.3 2.57e-4 6.06e-4 -7.85e-4 24.5
## 3 -0.441 18.8 19.1 -0.373 18.3 2.35e-3 1.29e-3 3.86e-4 24.5
## 4 -0.327 18.8 19.1 -0.316 18.3 6.10e-3 2.56e-5 2.05e-3 24.6
## 5 -0.471 18.9 19.1 -0.332 18.3 3.13e-3 -6.43e-2 3.72e-2 24.6
## 6 -0.539 18.9 19.1 0.00915 18.3 9.64e-3 -6.14e-1 3.37e-1 24.6
## 7 -0.653 18.9 19.1 0.239 18.3 1.34e-3 -1.01e+0 5.54e-1 24.6
## 8 -0.758 19.0 19.1 0.395 18.3 1.42e-3 -1.29e+0 7.06e-1 24.6
## 9 -0.727 19.0 19.1 0.547 18.3 5.77e-4 -1.49e+0 8.17e-1 24.6
## 10 -0.874 19.0 19.1 0.579 18.3 -1.25e-3 -1.63e+0 8.98e-1 24.6
## # ... with 1,330,806 more rows, 3 more variables: stator_yoke <dbl>,
## # ambient <dbl>, torque <dbl>, and abbreviated variable names
```

```
## # 1: stator_winding, 2: stator_tooth, 3: motor_speed
```

### Paso 2: Estandarizar datos

```
##          u_q      coolant stator_winding      u_d stator_tooth motor_speed
## [1,] -1.27834410 -0.8311962  -1.43987472  0.4603935  -1.4450026  -1.2150662
## [2,] -1.29998946  0.3778941  -0.40021908  0.4642469  -0.2280093  -1.2150626
## [3,]  0.68739232  1.3867850   1.38311958 -1.1322555   1.5957661   1.4795303
## [4,]  0.08919665 -0.5129853   0.08196488 -1.1405594  -0.1277971  -0.2900407
## [5,]  0.93365153 -0.8302683  -1.35853668  0.4270964  -1.2968917  -0.1372255
## [6,] -0.94387184 -0.8088649  -0.87802889  0.2524234  -1.0398530  -1.0803338

##          i_d      i_q      pm stator_yoke      ambient      torque
## [1,]  1.0333025 -0.4371630 -0.09350299 -1.37623227  0.7425053 -0.3724405
## [2,]  1.0333128 -0.4371546  0.71427036 -0.04505362  0.9194281 -0.4450275
## [3,] -0.9582591  0.1539443  1.71937546  1.60801206  0.9222913  0.1856948
## [4,] -0.2317817  1.7182056 -0.41674479 -0.39157610 -0.7370514  1.6244578
## [5,]  1.0332965 -0.4371743 -1.20503791 -1.20390935 -0.3008730 -0.4487295
## [6,]  0.3984230  1.0283933 -1.35716354 -1.04509760 -0.8976195  0.9337019
```

### Paso 3: Revisar de cumplimiento de normal multivariada

Para este cumplimiento de normal multivarida creamos nuestras hipótesis

$$H_0 : \mu_1 = \mu_2 = \mu_3 \dots \mu_k$$

$$H_1 : \mu_1 \neq \mu_2 \dots \neq \mu_k$$

```
##          Test          HZ p value MVN
## 1 Henze-Zirkler 4.782402          0 NO
```

Para el cumplimiento de normal univariada creamos de igual manera nuestras hipótesis

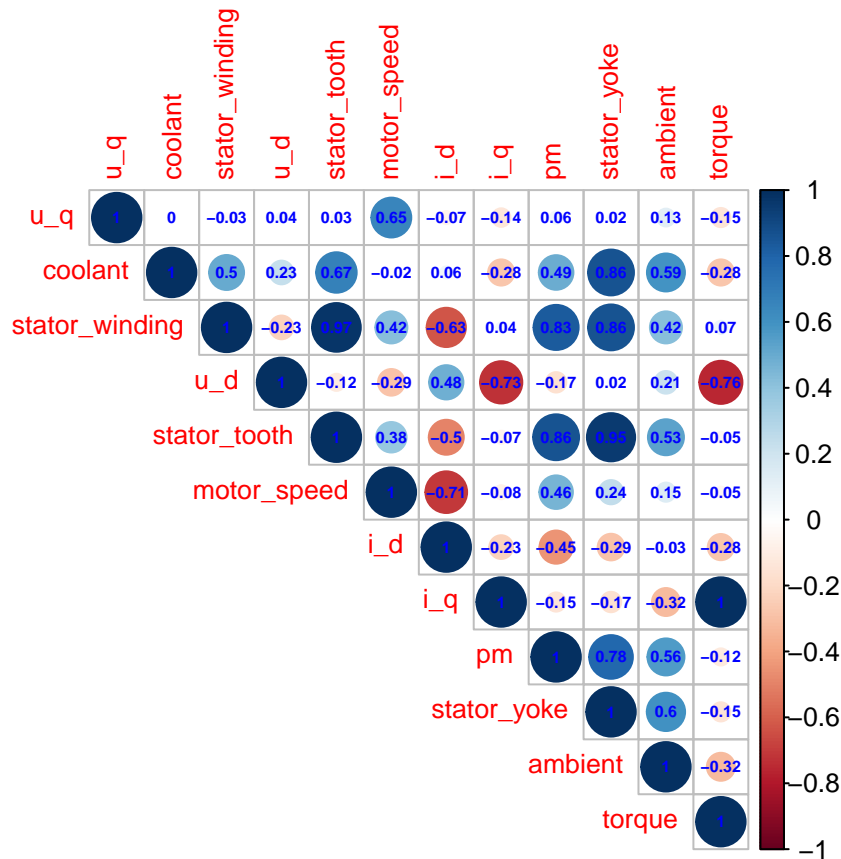
**H<sub>0</sub>** : los datos provienen de una distribución normal.

**H<sub>1</sub>** : los datos provienen de otra distribución.

```
##          Test          Variable Statistic    p value Normality
## 1 Anderson-Darling      u_q          12.0822 <0.001      NO
## 2 Anderson-Darling    coolant          37.8195 <0.001      NO
## 3 Anderson-Darling stator_winding      3.6821 <0.001      NO
## 4 Anderson-Darling      u_d          13.2541 <0.001      NO
## 5 Anderson-Darling stator_tooth       4.5108 <0.001      NO
## 6 Anderson-Darling motor_speed        11.8267 <0.001      NO
## 7 Anderson-Darling      i_d          17.9452 <0.001      NO
## 8 Anderson-Darling      i_q           8.2897 <0.001      NO
## 9 Anderson-Darling      pm           2.6402 <0.001      NO
## 10 Anderson-Darling stator_yoke        6.2023 <0.001      NO
## 11 Anderson-Darling    ambient         3.2206 <0.001      NO
## 12 Anderson-Darling    torque         8.5511 <0.001      NO
```

**Observaciones:** Se observa que no se cumplió con la prueba de normal multivariada dado su *p*-valor es **0** menor a alfa **0.05**, se rechaza **H<sub>0</sub>** los datos **no provienen de una normal multivariada**, con respecto a las pruebas de **normalidad univariada** se observa que ninguna variable cumplió con normalidad dados sus *p*-valores cercanos al cero y menores a alfa **0.05** por lo tanto los datos siguen otro tipo de distribución.

#### Paso 4: Grafica de correlaciones



**Observaciones:** Se observa fuertes correlaciones tanto positivas como negativas, Las correlaciones mas notables mostradas en la grafica son:

Relacion	Coficiente
coolant & stator_tooth	0.67
stator_winding & stator_tooth	0.97
stator_winding & i_d	-0.63
u_q & motor_speed	0.62
u_d & i_q	-0.73
motor_speed & i_d	-0.71
stator_tooth & pm	0.86
stator_winding & pm	0.83
coolant & stator_yoke	0.86
stator_winding & stator_yoke	0.86
stator_tooth & stator_yoke	0.95
pm & stator_yoke	0.78
coolant & ambient	0.59
stator_tooth & stator_yoke	0.95
pm & ambient	0.56
u_d & torque	-0.76
i_q & torque	1

**Observaciones:** Se observa una cantidad de fuertes correlaciones arriba de 0.70, tanto negativas como positivas. Una de las correlaciones más notorias es una correlación perfecta entre la variable **i\_q** y **torque** las cual es de 1.

### Paso 5: Prueba de esfericidad

Para esta prueba se usara la prueba de esfericidad de bartlett la cual sirve para identificar si la correlación entre pares de variables es cero o no.

Definimos nuestras hipótesis

Ho: La correlación entre cada par de variables es cero H1: La correlación entre cada par de variable diferente de cero

```
##
## Attaching package: 'psych'

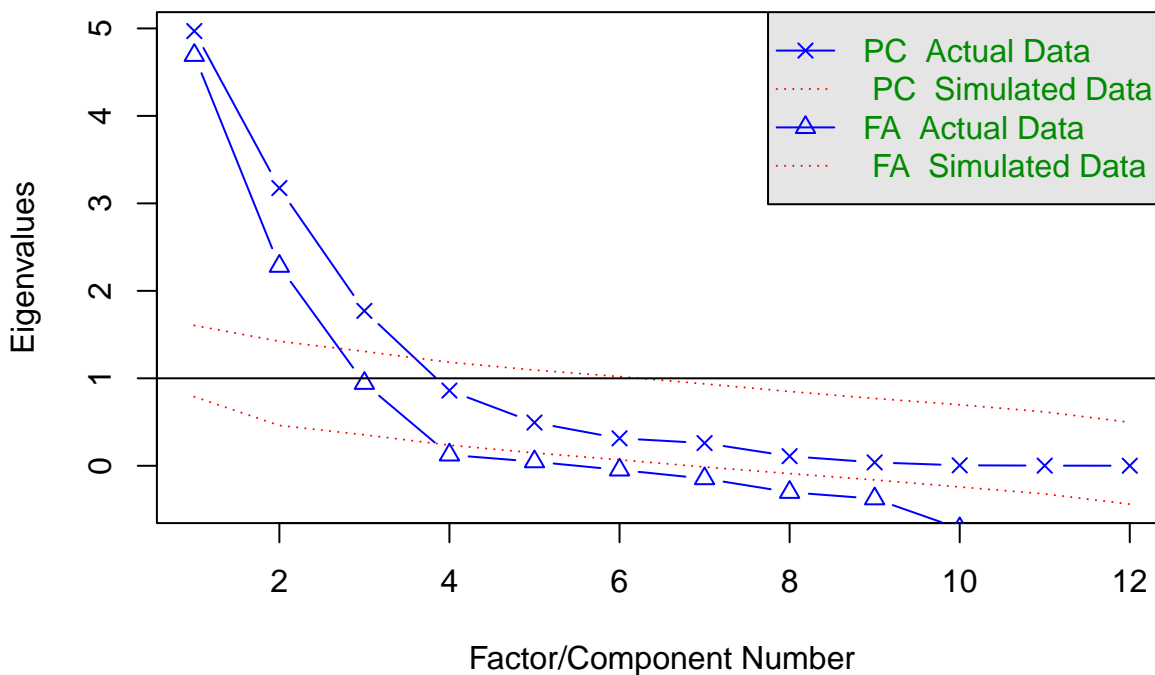
## The following objects are masked from 'package:ggplot2':
##
##   %+%, alpha
##
## $chisq
## [1] 2300.478
##
## $p.value
## [1] 0
##
## $df
## [1] 66
```

**Observaciones:** Dado que el  $p\_valor$  es menor a alfa **0.05** , se rechaza Ho por lo tanto las correlaciones son diferente de 0.

### Paso 6: Determinar numero de factores

Para determinar el número de factores, procederemos a realizar un Análisis de Componentes Principales (PCA), el cual nos sugerirá el número de factores a considerar.

### Parallel Analysis Scree Plots



```
## Parallel analysis suggests that the number of factors = 3 and the number of components = 3
```

**Observaciones:** Se puede observar que el numero factores optimo esta entre 3 y 4, Procedemos a obtener un resumen del análisis PCA para revisar cuanta varianza explicada es la que se tiene cuando se toman 3

o 4 componentes.

```
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation      2.2293 1.7819 1.3309 0.92737 0.7031 0.56018 0.50943
## Proportion of Variance 0.4142 0.2646 0.1476 0.07167 0.0412 0.02615 0.02163
## Cumulative Proportion 0.4142 0.6787 0.8263 0.89801 0.9392 0.96536 0.98698
##          PC8      PC9      PC10      PC11      PC12
## Standard deviation      0.33131 0.19658 0.07310 0.04296 0.02463
## Proportion of Variance 0.00915 0.00322 0.00045 0.00015 0.00005
## Cumulative Proportion 0.99613 0.99935 0.99980 0.99995 1.00000
```

**Observaciones:** Se puede observar que al elegir 3 factores obtenemos 82% de la varianza explicada, la cual es un buen porcentaje, Procedemos a usar el algoritmo ahora rotando los ejes usando el metodo de 'varimax'.

```
## Principal Components Analysis
## Call: principal(r = R, nfactors = 3, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          RC1      RC2      RC3      h2      u2 com
## u_q          -0.09  0.23  0.77 0.65 0.349 1.2
## coolant       0.78  0.31 -0.22 0.76 0.244 1.5
## stator_winding 0.90 -0.24  0.21 0.92 0.084 1.3
## u_d          -0.04  0.88 -0.20 0.81 0.193 1.1
## stator_tooth   0.97 -0.09  0.16 0.97 0.034 1.1
## motor_speed    0.24 -0.10  0.95 0.97 0.035 1.1
## i_d          -0.37  0.50 -0.59 0.74 0.265 2.7
## i_q          -0.13 -0.92 -0.14 0.89 0.112 1.1
## pm           0.85 -0.04  0.28 0.81 0.192 1.2
## stator_yoke    0.98  0.07  0.01 0.96 0.044 1.0
## ambient       0.66  0.35  0.03 0.55 0.448 1.5
## torque       -0.11 -0.94 -0.11 0.91 0.086 1.1
##
##          RC1      RC2      RC3
## SS loadings      4.70 3.10 2.11
## Proportion Var    0.39 0.26 0.18
## Cumulative Var    0.39 0.65 0.83
## Proportion Explained 0.47 0.31 0.21
## Cumulative Proportion 0.47 0.79 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.07
##
## Fit based upon off diagonal values = 0.98
```

**Observaciones:** Se observa una varianza acumulada del 83%, con respecto a los **residuales RSMR** se observa un valor muy bajo de **0.07** cercano a cero. Con respecto a las cargas elegidas estas muestran comunalidades (

$$h_2$$

) altas y la varianza no explicada

$$u_2$$

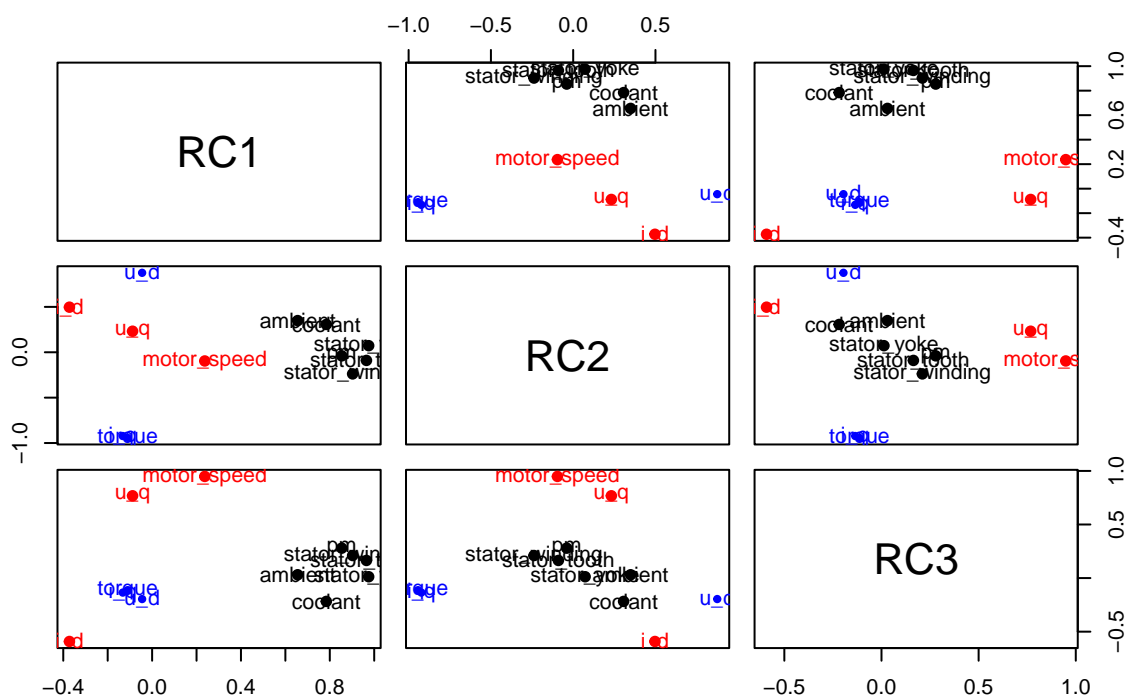
es muy baja. También observamos con el método de *varimax* de una manera muy clara los **variables dominantes para cada factor** los cuales son:

- **Factor 1** : stator\_winding, stator\_tooth, pm, stator\_yoke, coolant, ambient
- **Factor 2** : u\_d,i\_q,torque
- **Factor 3** : u\_q,motor\_speed,i\_d,

## Paso 7: Representación gráfica

Representacion grafica de cada uno de las variables.

### Principal Component Analysis



**Observaciones:** Se puede observar una agrupación muy notoria en las variables “stator\_winding, stator\_tooth, pm, stator\_yoke, coolant, ambient”, mientras que motor\_speed y u\_q están cercanas entre ellas, también se observa que las variable torque y u\_d están muy cercanas, la única variable que está muy alejada de los grupos antes mencionados es la variable i\_d.

### Conclusiones

Se concluye que aunque no se cumplieron los supuesto de normal multivariada dadas las pruebas de hipótesis, se obtuvo una varianza acumulada de 82% usando 3 factores con lo cual se redujo la dimensión de variables de 12 variables a solo 3, Por otra estos factores mostraron **comunalidades muy altas** y **varianza no explicada muy baja**, Con respecto a las variables dominantes de cada factor estas quedaron de la siguiente forma:

- **Factor 1 :** stator\_winding, stator\_tooth, pm, stator\_yoke, coolant, ambient
- **Factor 2 :** u\_d, i\_q, torque
- **Factor 3 :** u\_q, motor\_speed, i\_d,

## Conclusiones

## Referencias

Jolliffe, I. T. (2002). Principal component analysis. Springer.

Kirgiz, W. (2021). Electric Motor Temperature. Recuperado el 10 de enero de 2023, de <https://www.kaggle.com/datasets/wkirgsn/electric-motor-temperature>