



**Universidad de Valladolid**



**ESCUELA DE INGENIERÍAS  
INDUSTRIALES**

**UNIVERSIDAD DE VALLADOLID**

**ESCUELA DE INGENIERIAS INDUSTRIALES**

**Grado en Ingeniería de Organización Industrial**

# **Detección de fallos en motores mediante técnicas estadísticas robustas**

**Autor:**

**Losada Torres, Sergio**

**Tutores:**

**García Escudero, Luis Ángel.  
Fernández Temprano, Miguel.**

**Departamento de Estadística  
e Investigación Operativa**

**Valladolid, Marzo 2016.**



## INDICE DE CONTENIDOS

RESUMEN	5
ABSTRACT	5
<b>CAPÍTULO 1: INTRODUCCIÓN Y OBJETIVOS</b>	<b>7</b>
1.1 INTRODUCCIÓN Y OBJETIVOS	9
1.1.1 Antecedentes	9
1.1.2 Motivación	10
1.1.3 Objetivos	11
1.1.4 Estructura del proyecto	11
1.2 PROGRAMAS UTILIZADOS	13
1.2.1 Lenguaje R	13
1.2.2 Statgraphics	14
<b>CAPÍTULO 2: MOTORES DE INDUCCIÓN</b>	<b>15</b>
2.1 INTRODUCCIÓN	17
2.2 ASPECTOS CONSTRUCTIVOS: Tipos de motores	17
2.2.1. Jaula de ardilla	18
2.2.2 Bobinado	19
2.3 PRINCIPIO DE FUNCIONAMIENTO	20
2.4 TIPOS DE ARRANQUE DE LOS MOTORES DE ROTOR DE JAULA DE ARDILLA	20
2.4.1 Arranque directo	21
2.4.2 Arranque por autotransformador	21
2.4.3 Conmutación estrella triángulo	22
2.5 CLASIFICACIÓN DE LOS MOTORES ASÍNCRONOS COMERCIALES	23
2.4.1 Clasificación según la NEMA	23
2.4.2 Clasificación según la norma CEI 60034-12	24
2.6 FALLOS EN LOS MOTORES DE INDUCCIÓN	25
2.6.1 Fallos en el estator	25
2.6.2 Fallos en el rotor	28
<b>CAPÍTULO 3: ESTADÍSTICA ROBUSTA</b>	<b>31</b>
3.1 DEFINICIÓN	33
3.2 ¿QUÉ ES UN OUTLIER U OBSERVACIÓN ATÍPICA?	33
3.3 ¿QUÉ DEBE CONSEGUIR UN PROCEDIMIENTO ROBUSTO?	34
3.4 NECESIDAD DE LOS PROCEDIMIENTOS ROBUSTOS	35

3.5 PRINCIPALES MÉTODOS DE ANÁLISIS ROBUSTO	36
3.5.1 Análisis multivariante	36
3.5.2 Regresión múltiple	38
3.6 DETECCIÓN CLÁSICA FRENTE A DETECCIÓN ROBUSTA	39
<b>CAPÍTULO 4: EXPERIMENTACIÓN</b>	<b>41</b>
4.1 DESCRIPCIÓN DE LA EXPERIMENTACIÓN	43
4.2 DESCRIPCIÓN DEL BANCO DE ENSAYOS	45
4.2.1 Motor	45
4.2.2 Variadores de Frecuencia	46
4.2.3 Freno	48
4.2.4 Sensores de velocidad	49
4.2.5 Transductores	50
4.2.6 Sistema de adquisición de datos	52
4.3 CARACTERÍSTICAS MEDIDAS	54
4.3.1 Relativas a la carga del motor	54
4.3.2 Características en el dominio de la frecuencia	55
4.3.3 Características en el dominio del tiempo	56
<b>CAPÍTULO 5: TÉCNICAS UTILIZADAS</b>	<b>59</b>
5.1 INTRODUCCIÓN	61
5.2 CONTROL DE CALIDAD	61
5.3 ANALISIS DISCRIMINANTE	62
5.4 DISTANCIA DE MAHALANOBIS	63
5.5 MINIMUM COVARIANCE DETERMINANT ESTIMATOR(MCD)	64
5.6 $T^2$ DE HOTELLING	65
<b>CAPÍTULO 6: ANÁLISIS DE LOS DATOS</b>	<b>67</b>
6.1 ANALISIS DESCRIPTIVO: Análisis de correlaciones	69
6.2 DESCRIPCIÓN Y CARACTERÍSTICAS DE LOS ENSAYOS.	73
6.3 ANÁLISIS MULTIVARIANTE	81
6.3.1 AB35	82
6.3.2 AB50	83
6.3.3 AB65	83
6.3.4 RED50	84
6.3.5 SM35	84
6.3.6 SM50	85

6.3.7 SM65	85
<b>CAPÍTULO 7: CONTROL ESTADÍSTICO DE CALIDAD USANDO EL MCD</b>	<b>91</b>
7.1 INTRODUCCIÓN	93
7.2 ANALISIS	93
7.2.1 Introducción	93
7.2.2 Fase I	96
7.2.3 Fase II	100
7.3 EFICIENCIA DEL MÉTODO EN EL CASO ROBUSTO	109
7.4 COMPARATIVA FRENTE A MÉTODOS CLÁSICOS	110
7.4.1 introducción	110
7.4.2 Detección R234	110
7.4.3 Detección R2-R3-R4	113
<b>CAPÍTULO 8: ESTUDIO ECONÓMICO</b>	<b>117</b>
8.1 INTRODUCCIÓN	119
8.2 ALCANCE DEL PROYECTO	119
8.3 GESTIÓN DEL TIEMPO	120
8.4 GESTIÓN DE LOS COSTES	125
<b>CAPÍTULO 9: CONCLUSIONES</b>	<b>127</b>
<b>CAPÍTULO 10: ANEXOS</b>	<b>131</b>
10.1 INTRODUCCIÓN	133
10.2 CÓDIGOS	133
10.3 GRÁFICOS DE CONTROL ROBUSTO DE CADA ENSAYO	140
10.3.1 AB50	140
10.3.2 AB65	140
10.3.3 SM35	141
10.3.4 SM50	141
<b>BIBLIOGRAFÍA</b>	<b>143</b>
BIBLIOGRAFIA	145
PAGINAS WEB	146



## **RESUMEN**

En el presente trabajo se ha estudiado el problema del diagnóstico de la rotura de una barra del rotor de un motor de inducción del tipo jaula de ardilla mediante técnicas estadísticas robustas recientes. Se consideran motores reales que pueden estar conectados a un variador de velocidad de dos fabricantes diferentes o conectados directamente a la red eléctrica que han sido sometidos a diferentes grados de rotura en una de sus barras. Estos motores se prueban bajo diferentes condiciones de uso y se busca distinguir entre los diferentes estados de rotura de la barra a través de una serie de características de los motores que pueden ser medidas con el motor en funcionamiento. Se verá a lo largo del trabajo como las técnicas robustas consideradas suponen un avance significativo frente a las técnicas habituales, lo que puede redundar en mejoras importantes a la hora de detectar fallos en motores en situaciones industriales reales.

## **PALABRAS CLAVE**

Estadística robusta, Minimum Covariance Determinant (MCD), Motores de inducción, Rotura de barras, Detección temprana de errores.

## **ABSTRACT**

In this paper it has been studied the problem of the diagnosis of a broken bar in a squirrel cage induction motor using some recent robust statistical methods. We consider real induction motors, connected to a variable speed drive from two different manufacturers or to the electrical network, subjected to different degrees of failure in one of its bars. These motors are tested under different conditions with the aim of distinguish between the different breakage statements in of the bars analyzing some motors features while the motor is working. As long as this paper goes on we will see the significant improvement of using these robust methods against the classical ones, which could mean important improvements if we want to detect failures in motors working in real industrial situations.

## **KEYWORDS**

Robust statistics, Minimum Covariance Determinant (MCD), Induction motor, Broken bar, early recognition errors.



***CAPÍTULO 1:***  
***INTRODUCCIÓN Y***  
***OBJETIVOS***



## **1.1 INTRODUCCIÓN Y OBJETIVOS**

### **1.1.1 Antecedentes**

Comenzaremos analizando la situación inicial en la que surge la realización del presente trabajo. El tema central del mismo es diagnosticar fallos por avería en motores eléctricos de inducción. Como sabemos, el aumento de la productividad de la maquinaria es un objetivo genérico de la mayoría de las empresas que buscan con ello la obtención de mejores resultados y el aumento de su competitividad. Los fallos en un motor implican parada en el dispositivo en el cual se está realizando un trabajo, y lo que es peor, si se trata de un trabajo en cadena, no solo puede afectar a la máquina estropeada y los operarios que la manipulen sino al resto de máquinas del sistema productivo y operarios lo que puede derivar en un problema grave tanto de plazos como económico para la empresa en cuestión. Por tanto, en este contexto surge el estudio que trata de encontrar vías de mejora en los métodos de mantenimiento de manera que se consiga prever el fallo antes de que se produzca para lograr con ello una disminución de los efectos propios de un fallo en el motor correspondiente.

En la actualidad la mayoría de las empresas se encuentran realizando una mejora de sus sistemas de mantenimiento, los cuales se han quedado obsoletos con la modernización e informatización de la maquinaria moderna. Las tareas de mantenimiento en fábricas con máquinas que funcionan con esta clase de motores pueden llegar a suponer un coste muy alto en relación al valor que pueden llegar a aportar, ya que si se dispone de un gran número de motores puede ser necesario un número elevado de operarios o de motores de repuesto. Además dichos operarios se pueden ver en la situación de tener que tomar decisiones rápidas que a veces pueden ser erróneas o perjudiciales para los intereses de su empresa.

### 1.1.2 Motivación

Vamos a proceder a la definición de necesidades a partir de las cuales surge el siguiente proyecto. Anteriormente ya hemos comentado la relación que existe entre el mantenimiento predictivo de una máquina y las causas y consecuencias de una posible avería de la misma, es decir, de la productividad de la misma. Centrándonos en los problemas ocasionados por una avería inesperada de una máquina, debemos trabajar para anticiparnos a los fallos que se puedan producir para de esta manera poder solucionarlos antes de que se produzcan haciendo que el tiempo de parada sea el mínimo posible y si se diera el caso de trabajar sobre la misma en un horario fuera del que este activo todo el sistema productivo para así ahorrar costes y desperdiciar horas de trabajo y desgaste del resto de las máquinas implicadas en la cadena.

Para llevar a cabo un método de trabajo como el que proponemos es necesario dotar a los operarios que trabajan en dicho sistema productivo de un método sencillo con el cual puedan determinar en qué momento actuar y cómo. Mediante un sistema informático de seguimiento simple de las máquinas se puede determinar cuándo es necesario emprender acciones de mantenimiento sobre la misma, de manera que no sea necesario una consulta a mandos superiores que siempre derivara en retrasos y más tiempo de parada de la máquina o sistema.

En cuanto a la parte de desarrollo del trabajo, no es necesaria una monitorización de todas las máquinas de la fábrica, ya que se puede orientar a los denominados cuellos de botella ya que son los puntos clave para aumentar la eficiencia de una fábrica. Por eso todo trabajo, como el nuestro, que apueste por la disminución de los tiempos de parada de máquina son necesarios y se va a invertir en ellos ya que sus resultados van a ser altamente beneficiosos.

Estos cambios suponen una serie de cambios en los métodos de trabajo de los talleres y fábricas, sobre todo en la mentalidad de los operarios, así como la necesidad de un operario especializado que sea capaz de interpretar los datos obtenidos en los sistemas de información instalados. Como responsables de dichos cambios es responsabilidad nuestra la de presentar a los operarios de forma clara y concisa la información sobre los nuevos métodos utilizados, así como de un seguimiento inicial para su buena implantación.

### 1.1.3 Objetivos

En base a lo descrito anteriormente se establece como objetivo principal de este TFG la de obtener información fiable sobre la salud de las barras de un motor de inducción del tipo jaula de ardilla conectado a un variador determinado y a una frecuencia de trabajo conocida, de manera que podamos clasificarlos en totalmente sanos o dañados-rotos, para lo cual utilizaremos una base de datos proporcionada por el departamento de Ingeniería Eléctrica, cuya experimentación será descrita con detalle en el Capítulo 4 de este trabajo.

Si bien, para poder llevar a cabo este objetivo se establecen a la par una serie de objetivos secundarios que no solo ayudan a la consecución del primaria, sino que pueden ser interesantes individualmente. Dichos objetivos son los siguientes:

- Realizar una comparativa entre los métodos estadísticos tradicionales frente a los métodos robustos, probando la mejora que suponen estos últimos respecto a los primeros.
- Realizar un método experimental de detección de fallos en motores de inducción eléctricos.
- Comprobar si es posible una detección temprana de fallos de los motores atendiendo solo a su frecuencia de trabajo sin tener en cuenta el variador o en función del variador sin tener en cuenta la frecuencia de trabajo.

Para la consecución de dichos objetivos se aplicaran métodos de análisis de datos multivariantes y métodos robustos que serán explicados en el apartado 5 de esta memoria.

### 1.1.4 Estructura del proyecto

La memoria de este proyecto se divide en diez capítulos de la siguiente forma:

- **CAPITULO 1:** Introducción y objetivos

En este apartado se explica el propósito principal del TFG, basado en una serie de antecedentes y objetivos específicos, así como una breve explicación de las herramientas que se van a usar para el desarrollo del mismo.

- **CAPITULO 2:** Motores de Inducción

En este apartado se presenta de manera breve las características básicas propias de los motores de inducción, principio de funcionamiento, tipos, etc...

- **CAPITULO 3:** Estadística Robusta

Se trata de exponer un poco el concepto de este tipo de métodos estadísticos y en que se basan, así como la explicación breve de los métodos más usados.

- **CAPITULO 4:** Características de la experimentación

En este capítulo se habla sobre la experimentación llevada a cabo en los laboratorios, explicando el banco de pruebas y sus elementos, y las variables y datos sobre los que se van a trabajar.

- **CAPITULO 5:** Métodos Utilizados

Se explica de manera detallada cada una de las técnicas que se van a utilizar para la realización del TFG.

- **CAPITULO 6:** Análisis de los datos

Se realizara un análisis de las características de la base de datos, sus variables y las particularidades de cada uno de los ensayos. Una vez se hayan analizando las variables resumen se realizara una selección de las más significativas para utilizar durante el TFG.

- **CAPITULO 7:** Control estadístico de calidad usando el MCD

En este capítulo se explica la parte central del trabajo, la detección temprana de fallos (técnica de control de calidad) y se complementa con ejemplos de varios ensayos, así como se analiza la eficiencia de los métodos clásicos respecto a los robustos.

- **CAPITULO 8:** Estudio económico

En este estudio se hace referencia al alcance, el tiempo y los costes estimados del proyecto.

- **CAPITULO 9:** Conclusiones

Exposiciones de los resultados finales y conclusión sobre los mismos.

- **ANEXOS**

- **BIBLIOGRAFÍA**

## 1.2 PROGRAMAS UTILIZADOS

### 1.2.1 Lenguaje R

R es un lenguaje y entorno de programación para análisis estadístico y gráfico. Se trata de un proyecto de GNU, sistema operativo de software libre, que es muy similar al lenguaje S, muy conocida en entornos estadísticos, que fue desarrollado en los Laboratorios Bell. El software libre proporciona una amplia variedad de técnicas estadísticas (modelado lineal y no lineal, series en el tiempo, agrupaciones...) y técnicas gráficas.

El funcionamiento de R es relativamente sencillo. El usuario dispone de una consola central desde la cual se introduce el código de programación deseado y a partir de esa consola van surgiendo ventanas en función de los distintos análisis que vayamos realizando como se puede ver en la figura 1.1. Además de realizar la escritura del código, desde esta consola se pueden cargar los paquetes estadísticos de R que se deseen en caso de que los que vienen cargados por defecto no nos sean suficientes para el análisis que queramos realizar. Al tratarse de un software libre existe la página web <https://www.r-project.org> en la cual puede descargarse el programa junto a manuales en varios idiomas, como los usados para este TFG [GAR08] así como distintos paquetes y su funcionamiento.

```

R Console
Archivo Editar Misc Paquetes Ventanas Ayuda

Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900081-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribución.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'scitation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> X2 <- read.table("C:/Users/Sergio/Desktop/2/Cuatri/TFG/Datos/AB/AB50.csv", sep = ";", dec = ".", header=TRUE)
>
> library(robustbase)
>
> # Aun quedan algunos outliers 8 38 45 54 66 71 76 78
> a <- covMod(X2[1:95,11:14])
> outliers <- (1:95)[1-e$mod.wt==1]
>
>
> X2 <- X2[1:213, -1]
> anno <- rep("green", 213)
> anno[96:213] <- "red"
>
> anno.num <- rep("1", 213)
> anno.num[96:213] <- "2"
>
>
> X2 <- cbind(X2, anno)
> names(X2)[11:14] <- c("B5", "B51", "B55", "B75")
>
> x11()
> pairs(X2[11:14], pch=anno.num, col=anno)
>
> x11()

```

Figura 1.1 Captura de la consola de trabajo de R Project.

## 1.2.2 Statgraphics

Statgraphics es una potente herramienta de análisis de datos que combina una amplia gama de procedimientos analíticos con extraordinarios gráficos interactivos para proporcionar un entorno integrado de análisis que puede ser aplicado en cada una de las fases de un proyecto. Se trata de un software muy intuitivo y que es bastante conocido ya que se trabaja sobre él en distintas asignaturas propias de la titulación. No obstante para aprender a usar ciertas herramientas necesarias para nuestro análisis como el análisis de multivariantes nos hemos ayudado de la guía que se puede encontrar para nuestra versión (STATGRAPHICS Centurion XVI.I) en la página web <http://www.statgraphics.net>. En la figura 1.2 podemos ver cómo funciona la interfaz principal del programa así como sus posibilidades de trabajo, las cuales hemos aprendido a utilizar usando [PER13].

	Motor	Alimentación	Estado	Estado2	Estado3	Estado4	Frecuencia	VelocidadM	Tension	Corriente	Des.
1	M11	AB	R1	R1	R12	R123	35	1042	152,3	1,56	0,00
2	M11	AB	R1	R1	R12	R123	35	1034	151	1,59	0,01
3	M11	AB	R1	R1	R12	R123	35	1029	152,5	1,62	0,01
4	M11	AB	R1	R1	R12	R123	35	1024	152	1,67	0,02
5	M11	AB	R1	R1	R12	R123	35	1021	153	1,7	0,02
6	M11	AB	R1	R1	R12	R123	35	1010	153,0	1,73	0,02
7	M11	AB	R1	R1	R12	R123	35	1014	152,5	1,79	0,03
8	M11	AB	R1	R1	R12	R123	35	1009	155,3	1,84	0,03
9	M11	AB	R1	R1	R12	R123	35	1007	154,2	1,88	0,03
10	M11	AB	R1	R1	R12	R123	35	1005	154	1,90	0,04
11	M11	AB	R1	R1	R12	R123	35	1001	155	1,98	0,04
12	M11	AB	R1	R1	R12	R123	35	996	154	2,04	0,05
13	M11	AB	R1	R1	R12	R123	35	991	154,7	2,13	0,05
14	M11	AB	R1	R1	R12	R123	35	987	156	2,17	0,05
15	M11	AB	R1	R1	R12	R123	35	985	154	2,23	0,06
16	M11	AB	R1	R1	R12	R123	35	985	154	2,23	0,06
17	M11	AB	R1	R1	R12	R123	35	1042	153	1,57	0,00
18	M11	AB	R1	R1	R12	R123	35	995	154	2,1	0,05
19	M11	AB	R1	R1	R12	R123	35	980	156,1	2,37	0,06
20	M11	AB	R1	R1	R12	R123	35	976	155,1	2,4	0,07
21	M11	AB	R1	R1	R12	R123	35	1042	152,9	1,57	0,00
22	M11	AB	R1	R1	R12	R123	35	1026	154,3	1,66	0,02
23	M11	AB	R1	R1	R12	R123	35	1016	154,7	1,76	0,03
24	M11	AB	R1	R1	R12	R123	35	1000	154,7	2	0,04

Figura 1.2 Captura del espacio de trabajo de Statgraphics.

***CAPÍTULO 2:***  
***MOTORES DE***  
***INDUCCIÓN***



## 2.1 INTRODUCCIÓN

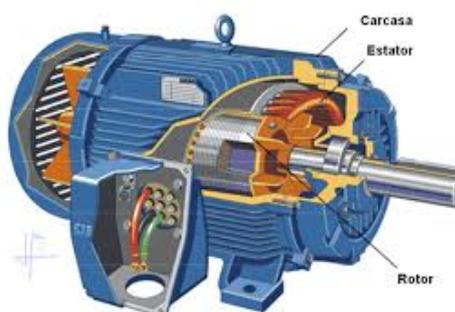
Los motores de inducción son los más ampliamente utilizados en la industria debidos a su simplicidad, bajo coste y facilidad de operación.

Debido a su gran uso, en torno al 85% de los motores que se usan en la industria [DUQ05], los motores de inducción con rotor de jaula de ardilla son un elemento muy importante en las líneas de producción de muchas empresas. Esta importancia es tal, que debido a que un fallo en un motor en una cadena de producción puede ocasionar un coste económico no esperado así como retrasos en la programación de dicha línea de producción.

El estudio de métodos de reducción de costes, tiempos o aumento de eficiencia en líneas de producción es objeto de muchos campos de investigación actuales en el ámbito de organización industrial, de ahí que el análisis de los motores que mueven las máquinas y sus posibles fallos sean de gran interés para dichas investigaciones.

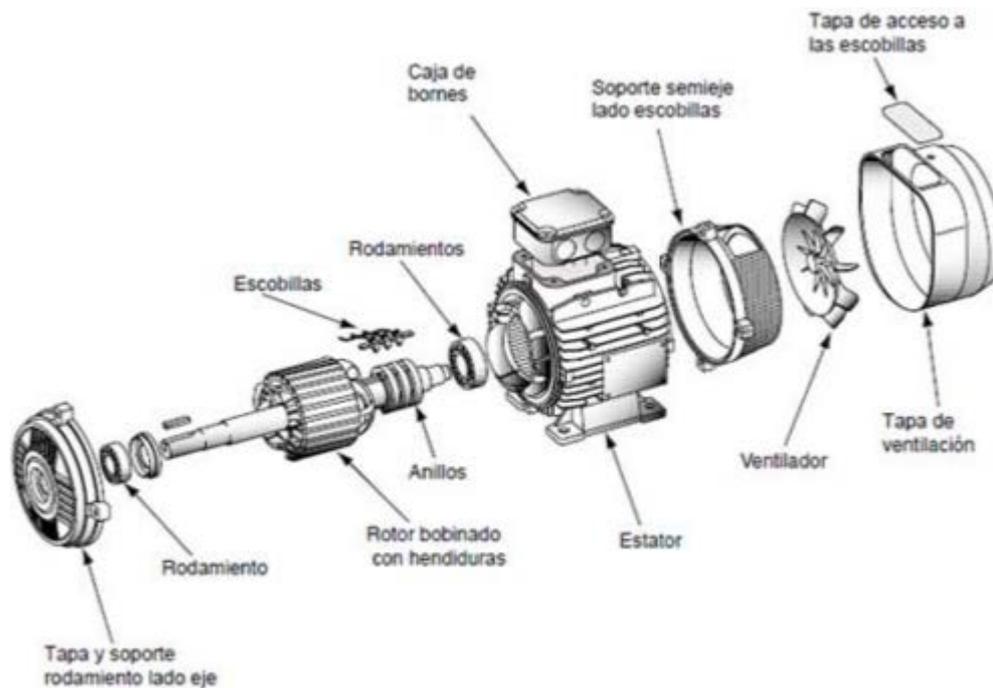
## 2.2 ASPECTOS CONSTRUCTIVOS: Tipos de motores

El motor asíncrono o de inducción al igual que cualquier otro dispositivo de conversión electromecánica de la energía de tipo rotativo, está formado principalmente por un estator y un rotor (ver figura 2.1). En el estator se coloca normalmente el inductor, alimentado por una red mono o trifásica, mientras que el rotor es el inducido. El estator está formado por un apilamiento de chapas de acero al silicio que disponen de unas ranuras en su periferia interior en las que se sitúa el devanado trifásico distribuido. El rotor está constituido por un conjunto de chapas apiladas, las cuales forman un cilindro giratorio, el cual tiene unas ranuras en la circunferencia exterior, donde se coloca el devanado. Es el rotor el que como veremos en el siguiente subapartado marca la diferencia de qué tipo de maquina asíncrona tenemos. Entre ambos elementos se encuentra lo que se conoce como entrehierro, que no es más que el aire que separa ambos elementos.



**Figura 2.1** Sección parcial de un motor de inducción de jaula de ardilla

Las máquinas asíncronas además de disponer de un estator y un rotor, están dotadas de otros elementos mecánicos necesarios para su funcionamiento, como pueden ser tapas, rodamientos, ventiladores, etc... (ver figura 2.2). De estos elementos mecánicos hay dos que poseen una importancia mayor que el resto que son la carcasa y el ventilador. La carcasa es la parte externa del motor, se encarga de proteger al estator y al rotor, y su material y forma pueden variar en función del uso que se quiera dar al motor correspondiente. En cuanto al ventilador, tiene la función de refrigerar de una manera más brusca que como lo hace la configuración de la carcasa.



**Figura 2.2** Esquema de los elementos de un motor de inducción

Los rotores tiene dos formas constructivas básicas que dan lugar a dos tipos de motores de inducción diferentes [DUQ05] que son: Motores de rotor de Jaula de Ardilla y motores de rotor bobinado.

### 2.2.1. Jaula de ardilla

El motor de corriente alterna trifásico de jaula de ardilla es el motor eléctrico industrial por excelencia ya que es fuerte, robusto y sencillo por lo que se usa en un gran número de máquinas sin la necesidad de realizar un mantenimiento exhaustivo. En este tipo el rotor se encuentra formado por una serie de barras en las cuales se colocan una serie de conductores de cobre puestos en cortocircuito por dos anillos laterales (ver figura 2.3).



**Figura 2.3** Rotor de motor de jaula de ardilla

Será esta clase de motores los que usaremos para nuestro estudio.

### 2.2.2 Bobinado

El motor de corriente alterna trifásico bobinado es menos usado que el de jaula de ardilla debido principalmente a que es más caro y necesita de un mantenimiento mucho más exhaustivo que los de jaula de ardilla. Como podemos ver en la figura 2.4, esta clase de rotores están formados por un arrollamiento trifásico en el que las tres fases se conectan por un lado en estrella y por otro lado se envían a uno anillos aislados entre si. Esta disposición hace posible que se pueda modificar desde el exterior para variar las características básicas del mismo.



**Figura 2.4** Bobinado de un motor

## 2.3 PRINCIPIO DE FUNCIONAMIENTO

El flujo magnético producido por la corriente alterna del estator genera un campo magnético rotativo que corta a los conductores del rotor obteniéndose sobre ellos, según la ley de inducción de electromagnética de Faraday, fuerza electromotriz inducida, con lo que circulará por el rotor una determinada intensidad. Según la ley de Lenz, las corrientes inducidas en el rotor generan fuerzas electromotrices y pares que tienen a arrastrar el rotor en el mismo sentido que gira el campo [FRA08]

Para que esto ocurra tiene que existir un movimiento relativo entre el campo y los conductores inducidos, por lo tanto, el motor nunca va a alcanzar la misma velocidad que el campo (velocidad síncrona), porque si lo hiciera se anularían las fuerzas electromotrices. Dependiendo del par resistente al que se vea sometido el rotor su velocidad se incrementara de una manera u otra. La figura 2.5 muestra las diferentes fuerzas electromagnéticas que actúan sobre el rotor de un motor de jaula de ardilla.

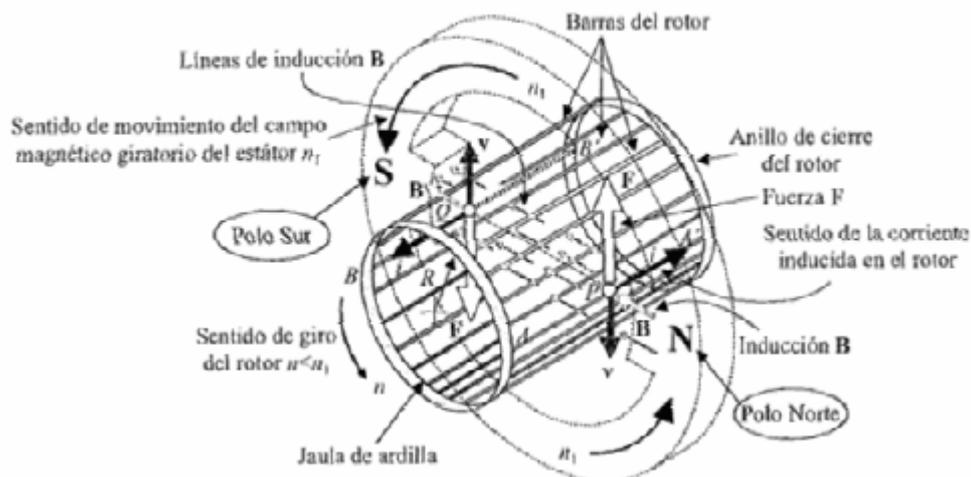


Figura 2.5 Campos magnéticos que actúan sobre el rotor [FRA08]

## 2.4 TIPOS DE ARRANQUE DE LOS MOTORES DE ROTOR DE JAULA DE ARDILLA

Se denomina arranque el proceso de puesta en marcha de una máquina eléctrica. Las dos características clave para este proceso son el par y la corriente de arranque. Para que un motor entre en funcionamiento es necesario generar un par de arranque sea superior al par resistente de la carga, y este proceso de arranque va acompañado de un consumo elevado de corriente, tanto que se asemeja a un cortocircuito, de ahí que dichas

corrientes de arranque se encuentran reguladas por las diferentes normas de cada país. Existen 3 maneras diferentes de arranque [FRA08] para estos motores, que son: Arranque directo, arranque por autotransformador y conmutación estrella-triángulo.

### 2.4.1 Arranque directo

Se emplea únicamente en motores de pequeña potencia. El esquema de conexiones se puede ver en la figura 2.6. Este método se aplica a maquinas con potencia inferior a 5kW cuando se trata de instalaciones conectadas a la red urbana. En grandes fabricas con gran potencia instalada, puede llegarse a arranques directos con motores de hasta 100 CV

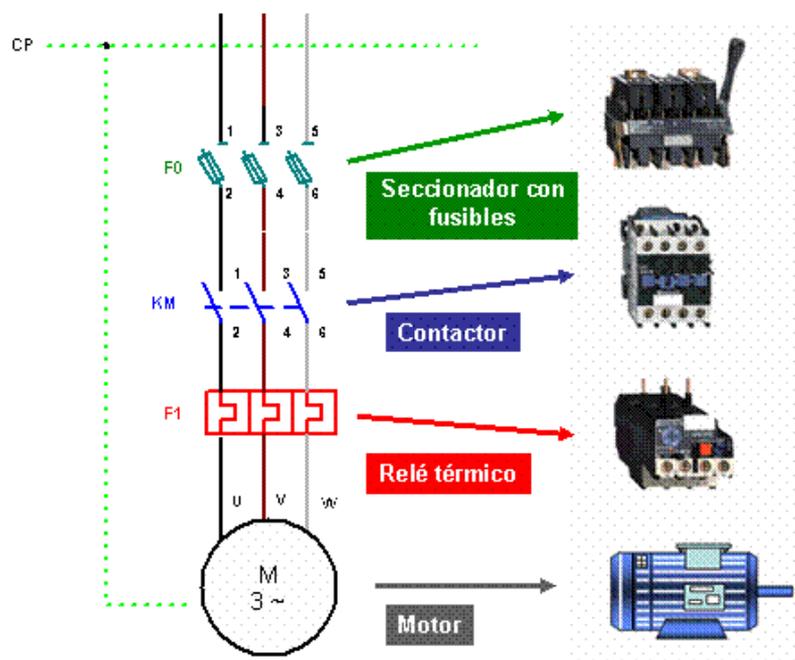


Figura 2.6 Esquema de funcionamiento de un arranque directo

### 2.4.2 Arranque por autotransformador

Consiste en intercalar un autotransformador entre la red y el motor (figura 2.7) de forma que la tensión aplicada en el arranque sea solo una fracción asignada. En este caso el fundamento viene por la relación directa entre el par y la tensión, de modo que si fraccionamos la tensión estaremos haciendo lo propio con el par. En cuanto a la corriente su valor esta disminuido por la impedancia de cortocircuito.

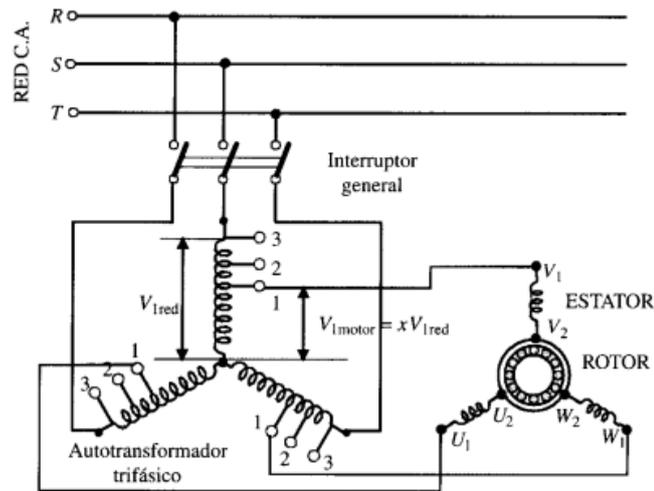


Figura 2.7 Esquema de arranque por autotransformador

### 2.4.3 Conmutación estrella triángulo

Este método solamente se puede utilizar en aquellos motores que estén preparados para funcionar en triángulo con la tensión de red. La máquina se conecta en estrella en el momento del arranque y se pasa después a triángulo para su funcionamiento (ver figura 2.8). En este caso debido a la relación entre las corrientes en triángulo y estrella, se consigue que durante el arranque se use una tercera parte de la que va a absorber el motor en su funcionamiento y lo mismo ocurre con el par.

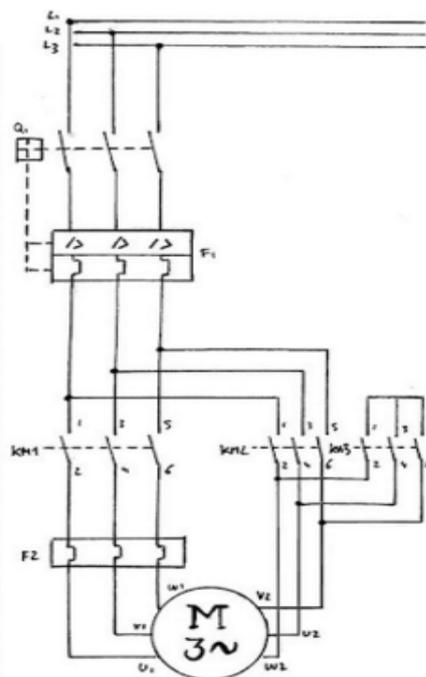


Figura 2.8 Esquema de arranque por conmutación

## 2.5 CLASIFICACIÓN DE LOS MOTORES ASÍNCRONOS COMERCIALES

Existen dos clasificaciones de los motores de inducción comerciales, que son la realizada por la NEMA (National Electrical Manufacturers Association) y otra la que establece la norma CEI-60034-12. Ambas los clasifican atendiendo al par de arranque y corriente necesarios para su funcionamiento, aunque la primera aborda alguna característica a mayores.

### 2.4.1 Clasificación según la NEMA

- **Clase A:** Motor de jaula de ardilla normal o estándar, fabricado para un uso a velocidad constante. Tiene grandes áreas de ranuras para una buena disipación del calor. Tiene una alta densidad de corriente de arranque y su par de arranque es relativamente alto entre 1.5 y 1.75 del nominal. Posee un factor de potencia elevado.
- **Clase B:** Se les conoce como motores de propósito general, Su comportamiento deslizamiento-par es muy parecido al de clase A. Tienen un par de arranque menor que los de clase, pero siguen manteniendo una corriente de arranque alto.
- **Clase C:** Esta clase de motores tienen un rotor conocido como de doble jaula. Tienen un mayor par de arranque que A y B, entre 2 y 2.5 veces el par nominal. Sin embargo su corriente de arranque es menor. En condiciones de arranque frecuente tiende a sobrecalentarse por lo que se suele limitar su uso a condiciones en las que al arranque es difícil.
- **Clase D:** También conocidos como de alto par y de alta resistencia. El par de arranque de estos motores se acerca a 3 veces el nominal y una corriente de arranque que varía entre 3 y 8 veces la nominal. Al igual que los de clase C no están aconsejados para múltiples arranques debido a una mala disipación del calor. Su uso se limita a condiciones de alto par necesario con aplicaciones de carga repentina.
- **Clase F:** Conocidos como motores de doble jaula y bajo par. Su diseño está concebido para bajas corrientes de arranque. Son motores con una alta resistencia de rotor y se fabrican para su funcionamiento directo en línea. La idea original era que sustituyeran a los de clase B.

Clase NEMA	Par de arranque (veces el nominal)	Corriente de Arranque	Regulación de Velocidad (%)	Nombre de clase de motor
A	1.5-1.75	5-7	2-4	Normal
B	1.4-1.6	4.5-5	3-5	Propósito general
C	2-2.5	3.5-5	4-5	De doble jaula alto par
D	2.5-3.0	3-8	5-8,8-13	De alto par alta resistencia
F	1.25	2-4	>5	De doble jaula, bajo par y baja corriente de arranque.

Tabla 2.1 Clasificación motores comerciales según NEMA

#### 2.4.2 Clasificación según la norma CEI 60034-12

- **Categoría N:** El par y la corriente de arranque son normales, y el deslizamiento es bajo. En esta categoría se incluyen la mayoría de los motores comerciales.
- **Categoría H:** Tienen un par de arranque alto, una corriente de arranque normal y un deslizamiento bajo. Esta clase de motores se usa para requerimientos de par de arranque elevado.
- **Categoría D:** Los motores de esta categoría tienen un par de arranque alto, una corriente de arranque normal y poseen un deslizamiento mayor del 5%.

## 2.6 FALLOS EN LOS MOTORES DE INDUCCIÓN

A pesar de la robustez de que disponen y su fiabilidad, estos motores no se encuentran exentos de sufrir fallos durante su utilización en la industria. La aparición de estos fallos ocurre principalmente por la sollicitación de un esfuerzo mayor que para el que está diseñado a trabajar. La detección de estos fallos puede realizarse en línea, que consiste en realizar el análisis del motor dentro de su cadena de producción, y fuera de línea, que consiste en sacar el motor de su lugar de trabajo en la línea para llevar a cabo su análisis y reparación en caso de que fuese necesario[DUQ05]. Para clasificar los fallos, se hará en relación a las partes que afectan del motor.

### 2.6.1 Fallos en el estator

En el estator la mayor parte de los fallos vienen por sobreesfuerzos producidos en los bobinados. Dichos esfuerzos se pueden clasificar en función de su naturaleza.

#### 2.6.1.1 Esfuerzos térmicos

Los motores poseen una capa aislante preparada para trabajar hasta una temperatura límite, a partir de la cual dicha capa empieza a comportarse de manera diferente a lo esperado. La temperatura límite a la que puede responder el aislamiento viene dada por la denominada clase térmica del aislamiento, tal y como podemos ver en la tabla 2.2.

Clase térmica	Temperatura
Y	90°C
A	105°C
E	120°C
B	130°C
F	155°C
H	180°C
200	200°C
220	220°C
250	250°C

**Tabla 2.2** Clasificación de resistencia térmica

Cuando un motor trabaja de manera anormal, se producen incrementos de temperatura en las distintas partes del mismo. En el momento en el que la temperatura del motor en funcionamiento supera la característica de la clase térmica utilizada para el aislamiento, comienza un proceso de envejecimiento del mismo, lo que deriva en una pérdida de sus propiedades dieléctricas, con el consiguiente riesgo de que se produzca un cortocircuito.

### **2.6.1.2 Esfuerzos eléctricos**

Existen una serie de problemas eléctricos que pueden producir fallos en el estator los cuales son:

- **Efecto corona:** El efecto corona se presenta cuando el potencial de un conductor en el aire se eleva hasta valores tales que sobrepasan la rigidez dieléctrica del aire que rodea al conductor. El efecto corona se manifiesta por luminiscencias o penachos azulados que aparecen alrededor del conductor, más o menos concentrados en las irregularidades de su superficie.

- **Fenómeno tracking:** El fenómeno llamado tracking es la formación de caminos conductores en la zona de la superficie de un aislante eléctrico. Este fenómeno se ve agravado por la contaminación y la humedad. El tracking se escucha como un sonido de zumbido y chisporroteos intermitentes, con pausas y caídas y crecimientos en intensidad. La intensidad puede aumentar hasta llegar al punto de combustión súbita (flashover). Después de la combustión súbita todo este sonido se convierte en silencio.

- **Transitorios:**

Los transitorios de tensión pueden traer como consecuencia la aparición de sobretensiones con efectos de negativos. Las causas pueden estar debidas a:

- Interrupción de la corriente
- Desconexión de condensadores
- Descargas eléctricas
- Cortocircuitos
- Reencendidos repetidos

### **2.6.1.3 Esfuerzos mecánicos**

- **Roces entre estator y rotor**

Pueden producirse por diversas causas, entre las que destacan:

- Fallo de rodamientos
- Eje doblado
- Desalineamiento entre el rotor y el estator
- Campo magnético desequilibrado

**- Introducción de objetos extraños**

Este problema suele causar daños tanto en el estator como en el rotor, consiste en la introducción fortuita de objetos extraños que pueden ser tanto del exterior como partes propias del motor como tornillos, arandelas, etc., a través principalmente del circuito de ventilación, en el entrehierro.

Los elementos que pueden resultar dañados pueden ser el aislamiento, las bobinas, las barras o anillos de cortocircuito o las chapas del núcleo magnético.

**- Movimiento relativo de las bobinas**

Debido a las altas intensidades que se originan en las máquinas eléctricas de media y alta tensión, en los arranques o conmutaciones, producen grandes fuerzas electromagnéticas sobre las bobinas. Estos esfuerzos dan lugar a que el aislamiento pueda sufrir microfisuras o rupturas importantes. Estos fallos estructurales pueden acabar con la sujeción de la bobina floja lo que permite un movimiento relativo entre la bobina y su posición de equilibrio, lo que deriva en un incorrecto funcionamiento del motor o en el agrietamiento del aislante.

**- Fatiga producida por vibraciones**

Las vibraciones son en mayor o menor medida un elemento común en todas las máquinas eléctricas. Estas vibraciones pueden producir la separación y rotura de las láminas de mica que forman parte del sistema aislante. El motor de inducción puede no solo verse influenciado por las vibraciones de su propia máquina, sino por la producida por máquinas adyacentes y una mala sujeción de las bobinas propias, que puede acabar en el deterioro de las mismas.

**2.6.1.4 Esfuerzos de carácter medioambiental**

Los esfuerzos ambientales a los que se encuentran sometidos los motores pueden ser muy variados y dependen básicamente del tipo de aplicación. Estos son los más comunes:

**- Condensación**

Es muy común que se produzcan dentro de los aislamientos una condensación debida a la humedad acumulada. Esto produce derivaciones a masa y la pérdida de características aislantes. La mejor solución para evitar este problema no siempre es factible ya que es necesario que se hayan producido grandes periodos de parada para poder secar el interior de la máquina, o en caso de motores que no sufran grandes paradas, mantener por encima de la temperatura ambiente los devanados en el momento que la máquina este parada.

#### - **Funcionamiento en altitud elevada**

Los motores que trabajan por encima de 1000m de altitud debido al enrarecimiento del aire pueden presentar problemas de refrigeración. Este problema hace necesario que se disminuya la potencia nominal de la máquina para poder evitar averías. El porcentaje de potencia disminuido suele estar indicado por los fabricantes de acuerdo con la altitud y temperatura.

#### - **Contaminación**

Las partículas de suciedad junto con los restos de la condensación del aceite y sus vapores propagados a través de los conductos de ventilación, pueden llegar a formar con el tiempo una capa de suciedad que impide la evacuación del calor, es decir, una reducción de su capacidad para disipar el calor que afecta a la vida útil de los aislamientos y en la formación de caminos alternativos de la corriente hacia la masa.

### **2.6.2 Fallos en el rotor**

Según la naturaleza de los fallos.

#### **2.6.2.1 Esfuerzos electromagnéticos**

##### - **Efecto electromagnético**

Las fuerzas electromagnéticas son proporcionales al cuadrado de la corriente y son unidireccionales, tendiendo a desplazar las barras radialmente y produciendo una vibración. Esta vibración provoca la deflexión de las barras sobre todo en los arranques, que son los momentos de mayor corriente utilizada.

##### - **Campos magnéticos desequilibrados**

En un motor ideal las fuerzas magnéticas son equilibradas en direcciones opuestas debido a que el rotor se encuentra perfectamente centrado en el entrehierro, haciendo que no se produzca deflexión en el motor. Sin embargo, en un motor real como son todos los de trabajo, el rotor presenta cierta excentricidad con respecto al estator causado por el propio peso, desgaste de los rodamientos o alineación del motor entre ambos.

Una acción prolongada de estas resultantes puede llegar a ocasionar que se doble el eje y que el rotor choque con el estator. Este problema se hace más importante cuanto mayor corriente circule por los devanados.

### 2.6.2.2 Esfuerzos térmicos

#### - Sobrecargas térmicas

Un motor puede tener sobrecargas térmicas en distintas situaciones. Dichas sobrecargas pueden dar origen a fallos en el rotor y sobre el conjunto del motor. Las causas más comunes de los fallos provocados por las sobrecargas térmicas son:

- Elevado número de arranques consecutivos que pueden provocar un aumento de la temperatura sobre las barras y anillos del cortocircuito.
- Bloqueos del rotor debidos a carga elevada
- Barras rotas
- Fallo en la aceleración del motor
- Insuficiente ventilación
- Desequilibrio de la tensión de alimentación

#### - Desequilibrio térmico

Estos fallos pueden deberse a causas de diseño o a condiciones de trabajo por encima de las especificadas por el fabricante. Las causas más comunes son:

- Frecuentes arranques
- Desigual transferencia de calor entre las barras del rotor y el núcleo
- Pérdidas de fijación entre el núcleo y el eje debido a expansiones térmicas
- Puntos calientes en la superficie del rotor
- Desigual ventilación en zonas diferentes del rotor

#### - Puntos calientes y pérdidas

Hay muchas variables durante el proceso de diseño, fabricación o reparación que pueden ser causa de pérdidas impredecibles y de formación de puntos calientes como son:

- Falta de simetría en la apilación de las láminas
- Barras desigualmente sujetas
- Chispas en las laminaciones en ranuras o superficies del rotor
- Diseño inadecuado de la laminación
- Acortamiento irregular de las barras del rotor
- Uniones deficientes entre las barras y los anillos

#### - Chispeos en el rotor

En los arranques se pueden llegar a originar caídas de tensión que pueden exceder en más de seis veces su valor en régimen de funcionamiento. Se produce una vibración en las barras debido a las fuerzas magnéticas y que pueden producir desprendimientos de partículas de hierro de las barras que serán las que originen los chispeos. Dichos chispeos pueden llegar a producir la rotura de las barras.

### **2.6.2.3 Esfuerzos dinámicos**

#### - Esfuerzos cíclicos

Si el rotor es sometido a esfuerzos de naturaleza cíclica pueden llegar a producirse fallos por fatiga. Son varias las causas que pueden originar estos fallos como: Desalineamiento con el eje accionado, ejes doblados, fluctuaciones de la carga, correas de transmisión sobretensadas, holguras de rodamientos, etc.

#### - Fuerzas centrífugas

Someter al rotor a velocidades por encima de la velocidad máxima para la que este diseñado a trabajar, es una de las principales razones de los fallos dinámicos que se producen en los motores. Para evitar estas situaciones los motores cuentan con mecanismos para que no se sobrepase dicha velocidad y así evitar desplazamientos del eje o de la jaula respecto del núcleo del rotor.

#### - Solicitaciones ambientales

Elementos externos que entren dentro del motor pueden producir que el sistema de ventilación del rotor funcione mal produciendo calentamientos

### **2.6.2.4 Esfuerzos mecánicos**

- Porosidad de la jaula
- Holguras en las laminaciones del paquete
- Rupturas o agrietamientos localizados
- Incorrecta sujeción
- Fallo de rodamientos
- Desalineamientos

***CAPÍTULO 3:***  
***ESTADÍSTICA***  
***ROBUSTA***



### 3.1 DEFINICIÓN

Pequeñas desviaciones respecto a las suposiciones que hacen los modelos estadísticos clásicos (normalidad, independencia y homogeneidad) pueden provocar desviaciones importantes en los resultados obtenidos. Para hacer frente a estas pequeñas desviaciones surgen las técnicas de Estadística Robusta, las cuales no solo solucionan el problema, sino que lo hacen sin perder la eficiencia de los métodos clásicos. Cuando se produce un error importante como un dato mal copiado o una medición o valor anómalo del proceso que estamos estudiando, dicho error suele aparecer como outlier y la aparición de outliers es siempre peligrosa para los supuestos asumidos en los métodos estadísticos clásicos. [HUB04]

### 3.2 ¿QUÉ ES UN OUTLIER U OBSERVACIÓN ATÍPICA?

Se denomina outlier o observación atípica a un valor que es sospechoso de no pertenecer al conjunto de datos de los que proceden o que pueden ser producto de algún suceso extraño. Este suceso extraño puede provenir de cosas como un error de medición, de un simple error de paso de datos al ordenador o medio donde se acumulen las observaciones o simplemente que existe alguna observación cuyos valores son muy diferentes de los que toman la mayoría de las observaciones.

A veces estas observaciones atípicas se pueden ver a simple vista gráficamente como en la figura 3.1, pero la mayoría de las veces (especialmente en problemas en dimensiones altas) es necesario aplicar una serie de criterios para su detección.

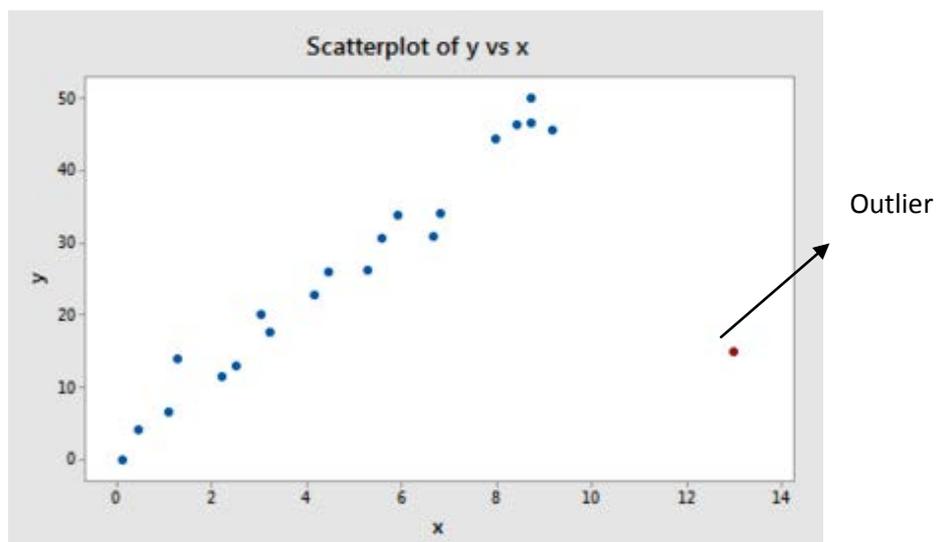


Figura 3.1 Ejemplo gráfico de outlier

Uno de los métodos más utilizados cuando hablamos de estadística univariante (variables tratadas de forma individual) se basa en el uso de cuartiles.

Si tenemos un conjunto de datos y lo ordenaremos de menor a mayor, el cuartil 1,  $Q_1$ , es el valor tal que desde ese valor hacia su izquierda se encuentran la primera cuarta parte de los valores de este conjunto de datos.

El cuartil 3,  $Q_3$ , es el valor tal que desde ese valor hacia su izquierda se encuentran las tres cuartas partes de los valores de este conjunto de datos. Y así sucesivamente.

Para detectar valores outliers moderados, tendríamos:

$$\begin{aligned}\text{LímInf} &= Q_1 - 1.5 (Q_3 - Q_1) \\ \text{LímSup} &= Q_3 + 1.5 (Q_3 - Q_1)\end{aligned}$$

Los outliers serán los valores que se encuentren o bien por encima del límite superior o bien por debajo del límite inferior.

Si quisiéramos detectar los valores outliers extremos deberíamos usar:

$$\begin{aligned}\text{LímInf} &= Q_1 - 3 (Q_3 - Q_1) \\ \text{LímSup} &= Q_3 + 3 (Q_3 - Q_1)\end{aligned}$$

Mientras en análisis univariantes resulta relativamente sencilla la detección de outliers, no ocurre lo mismo en los casos multivariantes. En casos de esta naturaleza, suelen ser los de mayor interés de estudio, pueden existir outliers multivariantes que pueden no tomar valores atípicos en ninguna de sus coordenadas por lo que son necesarias técnicas de mayor potencia que las que se usan para casos univariantes.

### **3.3 ¿QUÉ DEBE CONSEGUIR UN PROCEDIMIENTO ROBUSTO?**

Existen dos objetivos fundamentales [HUB04] en el uso de procedimientos estadísticos robustos, que son:

- Estimación robusta: Encontrar un ajuste robusto que sea lo más parecido posible al que obtendríamos sin la existencia de dichos outliers.
- Detección de outlier: Encontrar todas las observaciones que son outliers ya que suelen ser observaciones que pueden resultar interesantes una vez explicado porque son diferentes al resto de observaciones.

Por tanto para resolver estos dos problemas propios de la existencia de outliers, un procedimiento robusto debe satisfacer las siguientes características:

- Buena eficiencia: Debe de tener una buena eficiencia para el modelo asumido, es decir, que explique una gran parte de los sucesos que se producen.
- Robustez: Debe de ser lo suficientemente robusta para que una pequeña desviación en cualquiera de las suposiciones admitidas, apenas influya en el valor esperado en el modelo.
- Resistir errores de bulto o gross-errors : Visto desde el punto de vista de que en el caso de que existan desviaciones más grandes, que el modelo no sea totalmente diferente a lo que cabe esperar, es decir, de alguna manera que amortigüe los efectos de desviaciones grandes.

### **3.4 NECESIDAD DE LOS PROCEDIMIENTOS ROBUSTOS**

Como hemos dicho anteriormente, la existencia de outliers en un experimento hace que los valores o resultados obtenidos sean diferentes de los que esperábamos obtener, o si hablamos de un problema de clasificación, como el que tratamos en este TFG, podemos llegar a una clasificación diferente de la realidad, lo cual para problemas de Control de Calidad puede resultar en una situación de difícil arreglo.

Para ilustrar el problema de los outliers y la necesidad de los métodos robustos vamos a exponer un pequeño ejemplo de cómo afecta la no detección de outliers en un análisis multivariante de datos.

Un método clásico para detectar observaciones atípicas es la distancia de Mahalanobis, que explicaremos más adelante en el Capítulo 5, que usa como estimadores la media aritmética y la matriz de covarianza clásica. Como podemos ver en la gráfica 3.2, mediante esta distancia se crea un límite por el que por encima del mismo, se considera observación atípica y se descartaría. El problema viene debido a que mediante el uso de estos estimadores se puede producir el efecto de enmascaramiento, y es que exista un grupo de outliers que se encuentren juntos y que juntos afecten la determinación de la distancia de Mahalanobis ocultándose, lo que falsearía los resultados obtenidos en nuestra análisis.

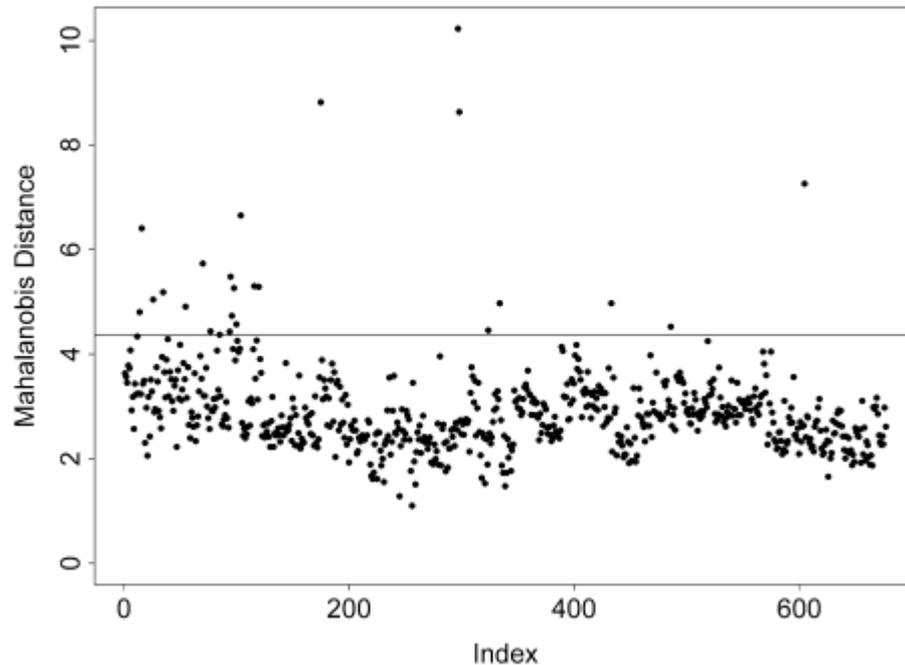


Figura 3.2 Representación de distancia de Mahalanobis no robusta [HUB08]

## 3.5 PRINCIPALES MÉTODOS DE ANÁLISIS ROBUSTO

### 3.5.1 Análisis multivariante

- Minimum Covariance Determinant Estimator (MCD): Este método se explicara con detenimiento en el capítulo 5, ya que es la técnica que utilizaremos para llevar a cabo nuestro análisis.

- Minimum Volume Ellipsoid Estimator (MVE): Se trata de un método alternativo (aunque con una filosofía parecida) al MCD [JEN07] que consiste en encontrar el elipsoide de volumen mínimo que cubre un grupo de observaciones que sea superior a un número  $h$  de puntos siendo  $h$  siempre un número mayor de la mitad de las observaciones de la muestra . Una vez que se ha detectado dicho elipsoide, se usan como estimadores robustos de localización el centro geométrico del elipsoide y como matriz de covarianza la matriz de covarianza definida por el elipsoide en sí mismo. Este método es usado en paquetes de programas altamente usados en la industria como el SAS.

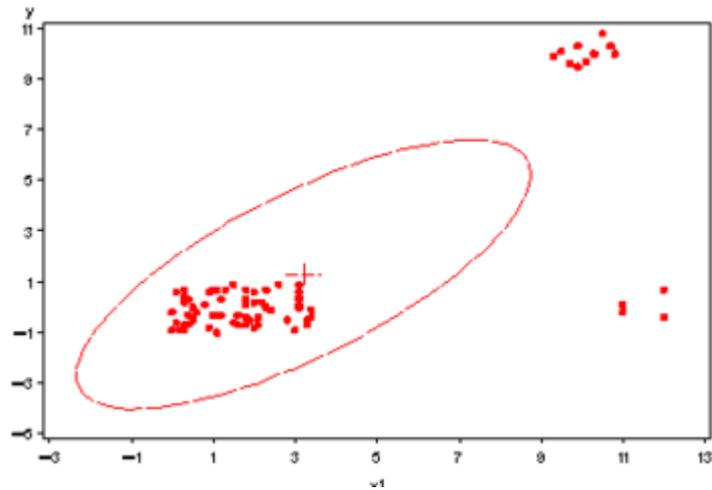


Figura 3.3 Ejemplo de MVE

- Stahel-Donoho estimator: Este fue uno de los primeros métodos de estimación robusta, su aparición data en torno a 1985. Se trata de un método en el que sus estimadores son una media y covarianza muestrales ponderadas, donde el peso de cada observación viene dado por lo que se conoce como outlyingness. Posee unas buenas propiedades robustas pero su principal problema radica en su programación informática ya que su alto número de operaciones a realizar hacen que su uso solo pueda realizarse con ordenadores especializados.

Para medir el buen o mal comportamiento de robustez de los procedimientos estadísticos que hemos descrito se usan dos herramientas: la “función de influencia” y el “punto de ruptura” que describimos brevemente a continuación:

- Función de influencia: Sirven para medir lo máximo que pueden verse afectados a nivel infinitesimal si existe alguna contaminación en algún punto.
- Punto de ruptura: Es la máxima cantidad de contaminación que puede resistir un estimador antes de dejar de ser totalmente operativo.

### 3.5.2 Regresión múltiple

El modelo de regresión múltiple asume que una variable respuesta  $Y$  puede ser explicada por una combinación de variables  $X$ .

$$Y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_p x_{pn} + e_n$$

Los métodos clásicos basados en mínimos cuadrados utilizados para calcular los distintos valores de  $\beta$  [HUB08] y de la varianza de la muestra son muy sensibles a la existencia de outliers de regresión, que son como podemos ver en la figura, aquellos puntos que no obedecen la distribución lineal formada por la mayoría de los puntos de la muestra.

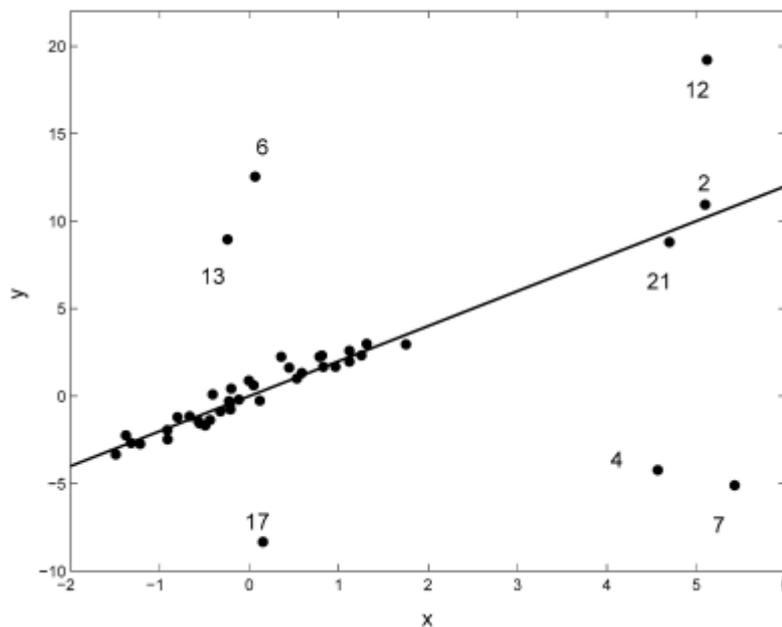


Figura 3.4 Diferentes outliers en una regresión [HUB08]

Existen como podemos ver en la figura diferentes tipos de puntos en una muestra de regresión:

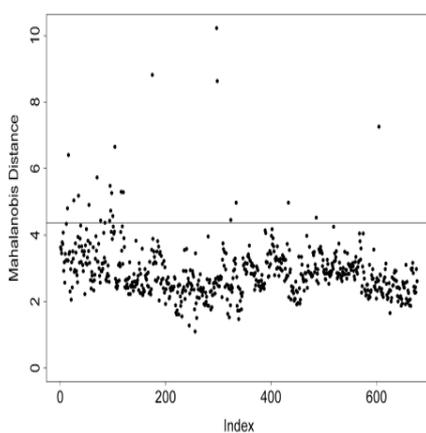
- Puntos regulares: Son los puntos de la muestra que se sitúan dentro de los patrones de la regresión.
- Puntos de influencia buenos: Se trata de puntos que mientras su valor correspondiente en la variable dependiente si seguiría el patrón de regresión correspondiente, se sitúa fuera de sitio en su valor de variable independiente( $X$ ). (Puntos 2 y 21 en la figura 3.4)
- Puntos de influencia malos: A diferencia de los anteriores, son puntos que ni siguen el patrón en la variable dependiente( $Y$ ), ni se sitúan en valores propios de la variable independiente( $X$ )(Puntos 4,7 y 12 en la figura 3.4)

- Outliers verticales: Son puntos que no son muy atípicos en la X, pero se encuentran bastante desviados del espacio correspondiente para Y en la recta de regresión. (Puntos 6,13 y 17 en la figura 3.4).

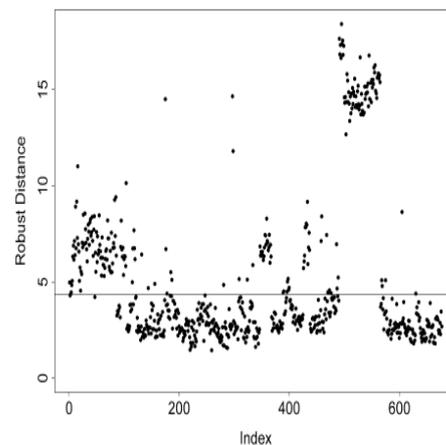
En casos en los que se disponga de pocas observaciones se puede llevar a cabo una inspección visual de los outliers y leverage points que aparecen en la muestra, pero cuando se trata de un número elevado de observaciones así como de una necesidad de ser más preciso es necesario buscar otra solución. Para llevar a cabo dicha detección se usa el método denominado Least Trimmed Squares (LTS). Este método fue propuesto por Rousseeuw en 1984 y su idea es la de coger como estimador el valor que minimiza la suma de los cuadrados de los residuos de un subgrupo con  $h$  observaciones de la muestra, o lo que es lo mismo, el estimador LTS es aquel valor de mínimos cuadrados que encuadra a los valores del conjunto con  $h$  observaciones.

### **3.6 DETECCIÓN CLÁSICA FRENTE A DETECCIÓN ROBUSTA**

Para ver con más claridad la diferencia que existe en la detección de outliers entre los métodos robustos y los métodos clásicos, en las gráficas inferiores podemos ver el número de puntos que serían considerados como atípicos si utilizamos la distancia de Mahalanobis, es decir, un método clásico, (figura 3.5a), mientras que en la otra imagen (figura 3.5b) podemos ver los que son considerados atípicos si utilizamos un método robusto. Salta a la vista como el número de observaciones raras es mucho más elevado utilizando los métodos robustos por lo que las conclusiones obtenidas de la manera clásica aparte de ser diferentes de las robustas, serán menos fiables.



**Figura 3.5a)** Distancia de Mahalanobis no robusta



**Figura 3.5b)** Distancia de Mahalanobis robusta



# ***CAPÍTULO 4:*** ***EXPERIMENTACIÓN***



## 4.1 DESCRIPCIÓN DE LA EXPERIMENTACIÓN

El proyecto que estamos tratando está basado en los datos provenientes de un experimento realizado por varios profesores del área de Ingeniería Eléctrica de la escuela de Ingenieros Industriales de la Universidad de Valladolid. Dicho experimento consiste en la reproducción de varios ensayos sobre un motor de inducción (como los explicados en el apartado 2) al cual se le va haciendo de manera progresiva roturas de diferentes profundidades en una barra del motor. Para ello se realizó mediante un taladro un agujero en la barra y dependiendo de la profundidad del mismo se definieron 4 estados diferentes de rotura:

- R1: En este estado se agrupan los motores que se encuentran completamente sanos.
- R2: En este estado se agrupan los motores con fallo incipiente. Se abre el motor, se saca el rotor y se realiza un agujero poco profundo (5,7mm) en la barra.
- R3: En este estado se agrupan los motores con un fallo ligeramente superior al anterior estado R2. Se aumento la profundidad del primer agujero a 12,1 mm.
- R4: En este estado se agrupan los motores con la barra totalmente perforada (18mm), que es el estado de mayor severidad de fallo.

Más concretamente para nuestro trabajo se utilizo una base de datos correspondiente a 1678 ensayos, con el motor conectado a la red y a dos variadores de frecuencia (explicados más adelante) con diferentes frecuencias de trabajo. La distribución de cada uno de los 1678 ensayos se puede ver en la tabla 4.1.

Conexión	Frecuencia	R1	R2	R3	R4
AB	35	138	42	40	45
AB	50	108	46	40	50
AB	65	106	41	49	45
RED	50	110	40	42	47
SM	35	101	42	42	45
SM	50	99	43	42	45
SM	65	102	82	46	45

**Tabla 4.1** Distribución de los datos en estados

Para llevar a cabo nuestro experimento estos datos serán usados de varias formas para llegar a lo que podemos denominar como 7 experimentos individuales y 5 grupales. Los ensayos individuales son aquellos que pertenecen a un variador determinado y a una frecuencia dada, mientras que los denominados grupales serán por pertenecer a un variador sin importar la frecuencia o viceversa, es decir, tendríamos los siguientes experimentos:

- Individuales:

1. AB35: Datos pertenecientes al variador AB y que trabajan a una frecuencia de 35 Hz.
2. AB50: Datos pertenecientes al variador AB y que trabajan a una frecuencia de 50 Hz.
3. AB65: Datos pertenecientes al variador AB y que trabajan a una frecuencia de 65 Hz.
4. RED50: Datos pertenecientes a motores conectados a la red y que trabajan a una frecuencia de 50 Hz.
5. SM35: Datos pertenecientes al variador SM y que trabajan a una frecuencia de 35 Hz.
6. SM50: Datos pertenecientes al variador AB y que trabajan a una frecuencia de 50 Hz.
7. SM65: Datos pertenecientes al variador AB y que trabajan a una frecuencia de 65 Hz.

- Grupales:

8. AB: Datos pertenecientes al variador AB, sin importar a qué frecuencia trabajan.
9. SM: Datos pertenecientes al variador AB, sin importar a qué frecuencia trabajan.
10. F35: Datos pertenecientes a motores que trabajan a 35 Hz, sin importar a que variador se encuentran conectados.
11. F50: Datos pertenecientes a motores que trabajan a 50 Hz, sin importar a que variador se encuentran conectados.
12. F65: Datos pertenecientes a motores que trabajan a 65 Hz, sin importar a que variador se encuentran conectados.

## 4.2 DESCRIPCIÓN DEL BANCO DE ENSAYOS

En este y posteriores apartados vamos a proceder a dar una descripción de los elementos que forman parte del banco de ensayos que se utilizó para la toma de datos por el departamento de Ingeniería Eléctrica.

Los elementos que se han utilizado son: Motores de inducción, variador de frecuencia, freno, sensores de velocidad y transductores. Además se contó con un sistema de adquisición de datos para medir los valores de la corriente conectado mediante USB a un ordenador portátil. Dichos datos fueron transferidos al programa Matlab para realizar los cálculos.



**Figura 4.1** Vista del Banco de ensayos

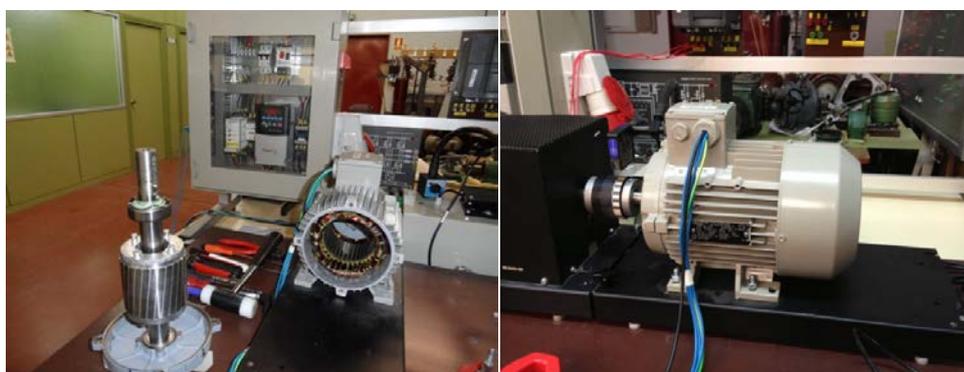
### 4.2.1 Motor

El motor que se ha utilizado para este experimento es el que se puede ver en la figura 4.2 perteneciente a la marca Siemens y cuyas características de funcionamiento podemos ver en la tabla:

Características técnicas del motor Siemens	
Potencia nominal	1,1 Kw
Tensión nominal	3x230/400 V
Conexión del estator	Estrella
Frecuencia nominal 50 Hz	50 Hz
Corriente nominal	2,6 A
Número de pares de polos	2

Número de barras del rotor	28
Velocidad nominal	1415 rpm
Denominación	3 ~ Mot 1LA7090-51
Número de serie	UD 0701/70749029-51
Eficiencia	EFF2
Índice de protección	IP55
Tamaño	90S IM B3
Clase térmica	ThcL F
Factor de Servicio	1,1

**Tabla 4.2** Características técnicas del motor Siemens



**Figura 4.2** Diferentes vistas del motor utilizado en los experimentos

#### 4.2.2 Variadores de Frecuencia

Los variadores de frecuencia, como su propio nombre indica, permiten controlar la velocidad de un motor de inducción mediante la variación de la frecuencia de la corriente de alimentación.

Para este trabajo se han usado dos variadores de frecuencia, uno de la marca Allen-Bradley (que representaremos como AB en nuestros ensayos) y otro de la marca Siemens (SM en nuestros ensayos) que se pueden ver en las figuras 4.3 y 4.4, respectivamente, y cuyas características de uso vienen recogidas en las tablas análogas.

Para que estos variadores entren en funcionamiento basta con conectarlos al banco de ensayos y llevar desde ellos la alimentación del motor para poder realizar las lecturas. En nuestro experimento hemos realizado lecturas a 35, 50 y 65 Hertzios para cada variador.

Características técnicas del variador PowerFlex 40 de Allen-Bradley	
Potencia nominal	2,2 kW
Tensión y frecuencia de alimentación	3x380 - 480 V (48-63 Hz)
Tensión de salida	0 - 460 V
Corriente nominal	7,5 A
Frecuencia de salida	0 - 400 Hz
Corriente máx. de salida en continuo	6 A
Sobrecarga temporal (60 seg)	9 A
Denominación	22B-D6P0N104
Número de serie	BD6P0NOW8500204
Índice de protección	IP20

**Tabla 4.3** Características técnicas del variador AB



**Figura 4.3** Vista del variador AB utilizado en la experimentación

Características Técnicas del variador MicroMaster 420 de Siemens	
Potencia nominal	4 kW
Tensión y frecuencia de alimentación	3x380 - 480 V 47-63 Hz
Tensión de salida	0 - Tensión de Entrada V
Frecuencia de salida	0 - 650 Hz
Corriente máx. de salida en continuo	10,2
Denominación	6SE6420-2UD24-0BA1
Número de serie	XAW902-006579
Índice de protección	IP20

**Tabla 4.4** Características técnicas del variador SM



**Figura 4.4** Imagen del variador SM utilizado en la experimentación

### 4.2.3 Freno

El freno es de polvo magnético y va acompañado de una unidad de control, todo de la marca Lucas-Nülle (figura 4.5). Dicha unidad de control nos permite fijar el valor del par resistente que ejerce el freno. Además, también incorpora unos indicadores de medida donde se muestra la velocidad de giro y el par resistente ejercido por el freno.

<b>Características del freno de polvo magnético Lucas- Nülle</b>	
Potencia nominal	1 kW
Tensión y frecuencia de alimentación	230 V 50 Hz
Par resistente	0 – 25 Nm
Velocidad de giro	0 – 6000 rpm
Denominación	SE 2662-5R

**Tabla 4.5** Características del freno magnético

<b>Características de la unidad de control Lucas- Nülle</b>	
Tensión y frecuencia de alimentación	230 V 50 Hz
Denominación	SO 3213-7S

**Tabla 4.6** Características de la unidad de control para el freno



Figura 4.5 Freno de polvo magnético y unidad de control de Lucas-Nulle

#### 4.2.4 Sensores de velocidad

El sensor de velocidad se encarga de medir la velocidad de giro del motor. El utilizado para esta experimentación figura 4.6 es de tipo digital y mide la velocidad mediante el contraste blanco-negro producido por una banda de esos mismos colores instalada en el eje del rotor del motor de inducción. El sensor es un S60W de la marca Datalogic (tabla 4.6).

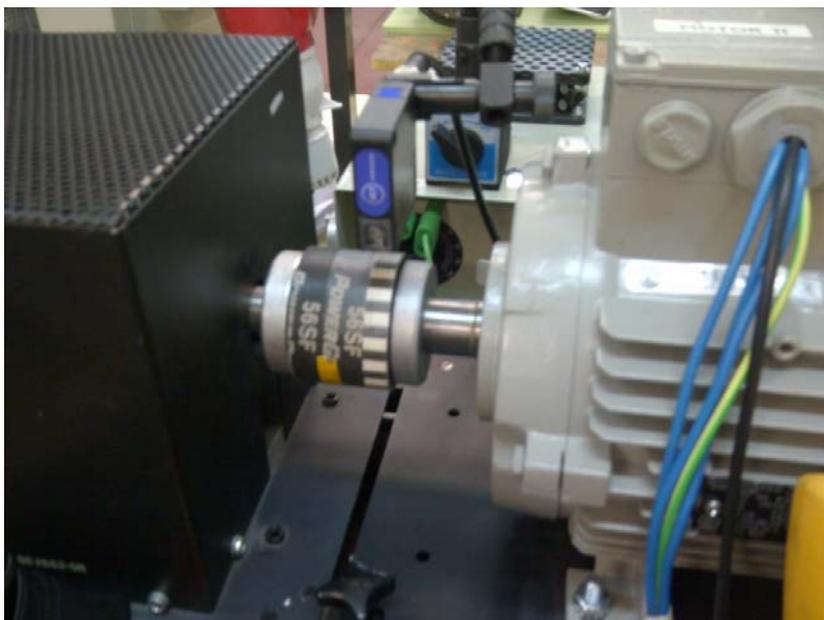


Figura 4.6 Sensor S60W

Características Técnicas S60W Datalogic	
Tensión de alimentación	10 – 30 Vdc
Rizado	2 Vpp max.
Consumo de corriente	30 mA
Outputs	PNP o NPN 30 Vdc.
Corriente de salida	100 mA

Voltaje de saturación de salida	$\leq 2V$
Tiempo de respuesta	100 $\mu s$
Frecuencia de conmutación	5 kHz
Indicadores	LED´s (amarillo, verde y rojo).
Ajustes	Botón de ajuste
Memoria de datos	EEPROM
Temperatura de operación	-10 – 55 °C
Distancia de funcionamiento	19 $\pm$ 2 mm

**Tabla 4.7** Especificaciones técnicas del sensor de contraste

#### 4.2.5 Transductores

Un transductor es un dispositivo que suministra una representación manipulable de la magnitud que tiene asociada. Dicha. En nuestro banco de ensayo los transductores han sido anclados a una placa de diseño y construcción propia en el mismo laboratorio. En nuestro experimento hemos contado con dos clases diferentes de transductores: Transductores de intensidad y de tensión.

##### 4.2.5.1 Transductores de intensidad

Su función es la de convertir la corriente alterna en tensión alterna proporcional a la intensidad. Para realizar la medida de de intensidad se han utilizado dos transductores de corriente Honeywell (figura 4.7) y otro Lem (figura 4.8)

Características técnicas del transductor de intensidad Honeywell	
Corriente primaria nominal	36 A
Tensión primaria(AC y DC)	10-500 V
Corriente secundaria nominal	25mA
Tensión de alimentación	15 Vdc
Consumo de corriente	10 ( $\pm$ 15V)mA
Ratio de conversión	2500:1000
Precisión a la corriente primaria nominal	$\pm$ 0.5%
Tiempo de respuesta a un escalón del 90% de la corriente primaria nominal	1 $\mu s$
Empaquetamiento	PCB

**Tabla 4.8** Características técnicas del sensor de contraste Honeywell



Figura 4.7 Transductor de corriente de Honeywell

Características técnicas del transductor de Corriente Lem(LA 25-NP)	
Corriente primaria nominal	25mA
Rango de medida	$\pm 36$ A
Corriente secundaria nominal	25mA
Tensión de alimentación	$\pm 15$ V
Consumo de corriente	10 mA
Precisión	$\pm 0.5\%$
Ancho de banda	150kHz
Tiempo de respuesta a un escalón del 90% de la corriente primaria nominal	1 $\mu$ s
Empaquetamiento	PCB

Tabla 4.9 Características técnicas del transductor de Corriente

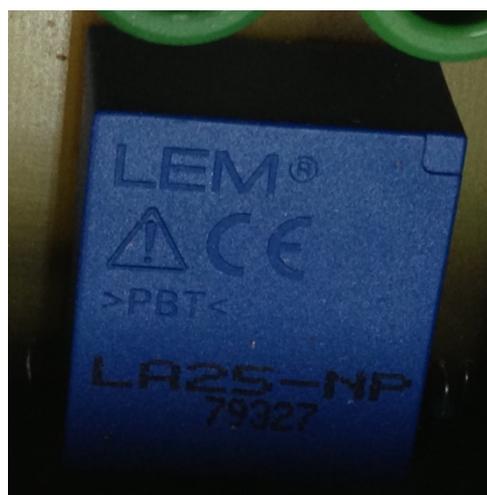


Figura 4.8 Transductor de corriente LEM

#### 4.2.5.2 Transductores de tensión

Para realizar la medida de tensión se empleó un transductor de Efecto Hall de LEM (Figura 4.9). Su función es la de transformar una tensión de 380V en una más pequeña 5 o 10 V.

Características técnicas del transductor de tensión LEM	
Corriente primaria nominal	10mA
Tensión primaria(AC y DC)	10-500 V
Corriente secundaria nominal	25mA
Tensión de alimentación	$\pm 12-15$ V
Consumo de corriente	10 ( $\pm 15$ )mA
Ratio de conversión	2500:1000
Precisión a la corriente primaria nominal	$\pm 0.9\%$
Tiempo de respuesta a un escalón del 90% de la corriente primaria nominal	40 $\mu$ s
Empaquetamiento	PCB

Tabla 4.10 Características técnicas del transductor de tensión

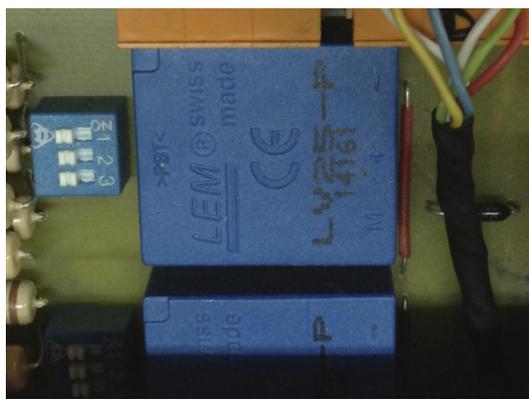


Figura 4.9 Transductor de corriente de LEM

#### 4.2.6 Sistema de adquisición de datos

Para transferir los datos obtenidos en cada una de las repeticiones del experimento con sus distintas variantes desde los transductores y sensores al ordenador se usó una plataforma de adquisición de datos de National Instruments (figura 4.10), conectado mediante un USB a un ordenador portátil, que consiste en un sistema modular formado por:

- Chasis compacto: NI cDAQ-9174. Este chasis permite la conexión de hasta 4 módulos de E/S.
- Módulo de adquisición: NI 9215.

Características técnicas del chasis NI cDAQ-9174	
Denominación	781157-01
Número de serie	14E8524
Alimentación	9-30 V
Potencia máxima consumida	15 W
Número de contadores/temporizadores	4
Resolución	32 bits
Interfaz de Bus	USB 2.0
Numero de ranuras	4

Tabla 4.11 Características técnicas del chasis

Características técnicas del módulo de adquisición NI 9215	
Denominación	779011-01
Número de serie	199269A-1L
Tipo de Medida	Voltaje
Tipo de aislamiento	Tierra
Canales diferenciales	4
Resolución	16 bits
Rango de Voltaje	-10V a 10V
Precisión máxima	3%

Tabla 4.12 Características técnicas del módulo de adquisición



Figura 4.10 Ordenador para la adquisición de datos y su sistema

## 4.3 CARACTERÍSTICAS MEDIDAS

Para nuestro experimento, como hemos dicho anteriormente, hemos utilizado una base de datos con 1678 observaciones de las cuales se recogieron 27 para ser utilizadas o en nuestro caso ser consideradas como variables. Existen también una serie de características que son propias de los motores de inducción y su carga de las que solo tendremos en cuenta el deslizamiento como explicaremos más adelante. Sin embargo, existen otras que se inicialmente consideramos que pueden dar una explicación del estado del motor, como son algunas relativas al dominio de la frecuencia del espectro de corriente, las amplitudes de ciertos armónicos, y otras variables se obtuvieron de aplicar varios estadísticos a la señal de la corriente estatórica que se encuentran relacionadas con el dominio del tiempo.

### 4.3.1 Relativas a la carga del motor

Este conjunto de variables, exceptuando el deslizamiento, solo son consideradas para conocer el estado de trabajo del motor durante los ensayos llevados a cabo, si bien algunas nos pueden ayudar a visualizar otras de las consideraremos importantes para la detección de fallos. Aunque este grupo de variables no entrara en ninguno de los análisis que haremos más adelante si conviene una pequeña descripción de las mismas.

- **Velocidad del Motor (VelocidadM):** Se trata de la velocidad a la que gira el rotor en cada ensayo medida en revoluciones por minuto. Para la misma como hemos descrito con anterioridad usaremos los sensores de velocidad colocados en el banco de pruebas.
- **Tensión:** Valor eficaz de la tensión consumida por el motor durante el ensayo y determinada a través de los transductores de tensión.
- **Corriente:** Valor eficaz de la corriente que consume el rotor mientras se lleva a cabo el ensayo.
- **Deslizamiento:** Es la única característica relativa a la carga del motor que usaremos de manera relativa para nuestro análisis. Hemos considerado eliminar las observaciones cuyo deslizamiento fuera menor que 0.01 ya que según recomendaciones del Departamento de Ingeniería Eléctrica estas observaciones pueden alterar resultados por considerarse poco fiables. En total hemos eliminado para nuestro análisis 204 observaciones.

El deslizamiento ( $s$ ) como tal es una característica que mide el valor relativo de la velocidad del motor. Su fórmula propia es la siguiente:

$$s = \frac{(n_1 - n)}{n_1}$$

Siendo  $n_1$  la velocidad angular de sincronismo y  $n$  la velocidad angular del rotor.

#### 4.3.2 Características en el dominio de la frecuencia

Debido a las propiedades de nuestro experimento la onda de corriente es propia del dominio temporal, por lo que será necesario transformarla en otra en el dominio de la frecuencia para así poder calcular las características en el dominio de la frecuencia. Para llevar a cabo este proceso hemos utilizado la Transformada rápida de Fourier (FFT) ya que requiere menos operaciones que la transformada simple de Fourier (FT).

Como podemos ver en la tabla 4.12 existe una frecuencia para cada banda lateral tal y como podemos ver en [DUQ05]. Para este experimento se midieron cuatro bandas alrededor del primer armónico, dos mediciones por cada lado, dos alrededor del quinto y otras dos alrededor del séptimo.

Banda lateral	Frecuencia
1er armónico, banda superior 1	$F_{bs} = (1+2s)f_1$
1er armónico, banda inferior 1	$F_{bi} = (1-2s)f_1$
1er armónico, banda superior 2	$F_{bs2} = (1+4s)f_1$
1er armónico, banda inferior 2	$F_{bi2} = (1-4s)f_1$
5º armónico, banda superior	$F_{b5s} = (1+2s)5f_1$
5º armónico, banda inferior	$F_{b5i} = (1-2s)5f_1$
7º armónico, banda superior	$F_{b7s} = (1+2s)7f_1$
7º armónico, banda inferior	$F_{b7i} = (1-2s)7f_1$

**Tabla 4.13** Diferentes armónicos utilizados

Siendo  $s$  el valor del desplazamiento y  $f_1$  el valor de la frecuencia.

### 4.3.3 Características en el dominio del tiempo

Para obtener las características en el dominio del tiempo lo que se ha hecho ha sido aplicar una serie de estadísticos a la onda temporal de la corriente estatística. Se define estadístico como una medida cuantitativa, derivada de un conjunto de datos de una muestra, con le objetivo de estimar características de una población. Para nuestro estudio se aplicaron los siguientes estadísticos:

• **Momento:** Un una distribución bidimensional de variables (x,y), se define el momento respecto al par (c,d) de orden r y s como:

$$m(c,d) = [ \sum_{i=1}^n \sum_{j=1}^m (X_i - c)^r (Y_i - d)^s n_{ij} ] / n$$

Donde:

$n_{ij}$  es la frecuencia del par  $(X_i, Y_i)$

$n$  es el total de observaciones de la muestra

$r$  y  $s$  son números naturales

$c$  y  $d$  son números reales

Los momentos que más se suelen utilizar debido a que se considera que son los que más información aportan son los calculados respecto el origen,  $c$  y  $d$  son 0, o cuando  $c$  y  $d$  se encuentran formados por las medias.

En nuestro experimento se han calculado hasta el momento de orden seis, aunque el primer momento fue calculado respecto el origen:

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i$$

El resto de momentos hasta el de orden seis fueron calculados respecto a la media de acuerdo a la formula:

$$m_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^n$$

• **Skewness:** El coeficiente de Skewness determina el grado de asimetría que posee una muestra respecto de su media. Para el caso de funciones simétricas el valor del coeficiente es cero. En nuestro hemos usado el coeficiente de Skewness de Pearson cuya fórmula viene dada por:

$$S_k = \frac{\sum \left( \frac{X_i - \bar{X}}{s} \right)^3}{n}$$

Siendo  $s$  la desviación típica y  $\bar{X}$  la media.

Podemos ver que si nos fijamos vemos la clara relación con el momento 3 respecto a la media.

Este indicador indica si la cola más larga de la distribución se encuentra desviada hacia la derecha, centrada o desviada hacia la izquierda de la distribución. Si se encuentra hacia la izquierda, el coeficiente será negativo y se dirá que la distribución se encuentra sesgada hacia a la izquierda, y viceversa.

• **Kurtosis:** El Coeficiente de Kurtosis analiza el grado de concentración que presentan los valores alrededor de la zona central de la distribución. Su fórmula más frecuente viene dada por:

$$\mathbf{Kurt} = \frac{\sum \left( \frac{X_i - \bar{X}}{s} \right)^4}{n}$$

Dependiendo de los valores que tome el coeficiente se considerara a la muestra:

- Mesocúrtica: Kurt=3, Este sería el valor propia de una distribución normal.
- Leptocúrtica: Kurt >3, La distribución es más apuntada que la normal y las colas son más pesadas.
- Platicúrtica: Kurt <3, La distribución tiene colas menos pesadas y es menos apuntada que la normal.

• **Cumulante:**

1<sup>er</sup> cumulante:  $m_1$

2<sup>o</sup> cumulante:  $m_2 - m_1$

3<sup>er</sup> cumulante:  $m_3 - 3m_1m_2 + 2m_1^3$

4<sup>o</sup> cumulante:  $m_4 + m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$

• **Promedio de valores absolutos (Absolut Mean):** Se calcula la media de los valores absolutos.

$$\overline{|X|} = \frac{1}{n} \sum_{i=1}^n |X_i|$$

- **Pico Absoluto ( $X_p$ ):** Es el valor máximo que toma la corriente estatística en cada ensayo.

$$X_p = \max |X|$$

- **Valor cuadrático medio( $X_r$ ):**

$$X_r = \left( \frac{1}{n} \sum_{i=1}^n \sqrt{X_i^2} \right)$$

- **Valor eficaz:**

$$X_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

- **Factor Cresta:** Se define como el cociente entre el pico absoluto y el valor eficaz de la corriente. Para una señal sinusoidal el factor de cresta es igual a  $\sqrt{2}$ . Este valor es usado para detectar picos absolutos con respecto al valor eficaz.

$$FC = \frac{X_p}{X_{rms}}$$

- **Factor Forma:** Se define como el valor eficaz de la corriente y la media de los valores absolutos.

$$FF = \frac{X_{rms}}{|\bar{X}|}$$

***CAPÍTULO 5:***  
***TÉCNICAS***  
***UTILIZADAS***



## **5.1 INTRODUCCIÓN**

En este apartado se van a describir cada una de las técnicas que se han utilizado para llevar a cabo la realización del trabajo.

Para realizar el análisis se llevo a cabo un procedimiento que basado en diversos análisis descriptivos para conocer las relaciones que existen entre las variables. Para poder decidir sobre que variables de las que disponemos trabajar llevamos una análisis discriminante lineal del que obtuvimos 4 variables que son las que más aparecen con más poder discriminante en cada uno de los ensayos.

Finalmente el método de clasificación elegido se basa en realizar un análisis de las observaciones sanas y descartar aquellas que son atípicas mediante un método robusto como el MCD, y posteriormente realizar una clasificación aplicando lo obtenido mediante el MCD para la distancia de Mahalanobis.

## **5.2 CONTROL DE CALIDAD**

Aunque no es una técnica como tal, sí se trata de un conjunto de técnicas entre las que se encuentran algunas de las que usaremos en este TFG.

El concepto de control estadístico de procesos está basado en la comparación de lo que está pasando en la actualidad (Fase II) con lo que ha pasado anteriormente (Fase I). En la fase I propia del control de calidad nos centramos en analizar el comportamiento y los límites de control que consideramos están dentro de nuestros umbrales deseados. Una vez que se han determinado los límites, se cogen dichas observaciones y se someten a los límites de control. Lo normal es que la mayoría de las observaciones se encuentren dentro de los límites. Las observaciones que han sido monitorizadas fuera de los límites son de nuevo examinadas para ver si pertenecen al mismo grupo o modelo que las que han sido situados dentro de los límites. Si tras una serie de análisis se determina que dichas observaciones deben ser descartadas, se repite el proceso de forma que se obtengan unos nuevos límites válidos. La fase II se considera a la monitorización en tiempo real de lo que se esté estudiando usando los límites finales obtenidos durante la fase I, de forma que todo aquello que ocurra fuera de dichos límites será descartado por tener un comportamiento que no corresponde con lo que hemos considerado válido.

### 5.3 ANALISIS DISCRIMINANTE

Debido a la gran cantidad de variables que existen en el estudio y la imposibilidad de realizar un análisis con la totalidad de las mismas sin problemas de sobreajuste, se decidió utilizar esta técnica para determinar cuáles son las variables que más aportan a la discriminación entre los distintos grupos del análisis para quedarnos con ellas y descartar las que no aportan mucha información relevante. Existen tres tipos de metodologías para la selección de variables en análisis discriminante: paso a paso (stepwise), hacia adelante (Forward selection), hacia atrás (Backward selection) y selección paso a paso (Stepwise selection).

- **Forward:** Se parte de un modelo que no tiene ninguna variable y se van añadiendo paso a paso aquella con mayor poder discriminante de las posibles hasta que el porcentaje de explicación apenas varíe al introducir una variable mas.
- **Backward:** Se parte de un modelo con todas las variables y se van quitando paso a paso las de menor poder discriminante hasta que el porcentaje de explicación disminuya bruscamente.
- **Stepwise selection:** Es uno de los métodos más empleados. Se trata de una mezcla entre los dos métodos anteriores ya que comienza con el método forward pero las variables pueden salir del modelo si se consideran redundantes para la explicación del modelo.

Para decidir que variable es la que mejor discrimina en cada caso existen una serie de criterios entre los que destacan:

- **F de Snedecor:** Se compara para cada variable las desviaciones de las medias de cada uno de los grupos a la media total, entre las desviaciones a la media dentro de cada grupo. Si F es grande para cada variable, entonces las medias de cada grupo están muy separadas y la variable discrimina bien. Si F es pequeña, la variable discriminará poco, ya que habrá poca homogeneidad en los grupos y éstos estarán muy próximos.

$$F = \frac{n - k - p - 1}{k - 1} \left( \frac{1}{\lambda} - 1 \right)$$

n: número de observaciones

k: número de grupos

p: número de variables

$\lambda$ : Lambda de Wilks

- **$\lambda$  de Wilks:** Cuando se considera a las variables de modo individual, la  $\lambda$  es igual al cociente entre la suma de cuadrados dentro de los grupos (S) y la suma de cuadrados total (T). Cuanto más pequeño es su valor, mayor poder discriminante, por lo que se trata de buscar la variable que haga menor el coeficiente:

$$\lambda = \frac{S}{T}.$$

## 5.4 DISTANCIA DE MAHALANOBIS

Se define la distancia de Mahalanobis [PEÑ07] entre un punto y un vector de medias por

$$MD(x_i) = \sqrt{(x_i - \bar{x})' S^{-1} (x_i - \bar{x})}$$

Donde  $X_i$  es el vector que contiene los valores de la observación,  $\bar{X}$  se trata del vector de medias correspondiente y  $S^{-1}$  se trata de la inversa de la matriz de covarianzas de la muestra de datos. La matriz de covarianza (S) de una muestra de datos, se trata de una matriz cuadrada en cuya diagonal se encuentran las varianzas de las distintas variables, mientras que los elementos que se encuentran fuera de la diagonal contienen las covarianzas posibles entre todos los pares posibles de variables

$$S = \begin{pmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{pp} & \cdots & s_p^2 \end{pmatrix}$$

Si tenemos en cuenta la definición de  $s_{ij}$  en el caso de que las variables sean independientes el valor de su correspondiente  $s_{ij}$  sería próximo a cero para los

casos en los que  $i$  es distinto de  $j$ , y si además las varianzas, que ocupan los términos diagonales en la matriz  $S$ , son todas parecidas entre sí, el uso de la distancia de Mahalanobis sería similar a usar la distancia euclídea. La ventaja de utilizar la distancia de Mahalanobis respecto a otras medidas es que nos permite adaptarnos a una estructura de dependencia y variabilidad, cosa que es fácil ocurra a medida que se aumenta el número de variables que toman partido.

Aunque es frecuente referirse también como distancia de Mahalanobis al valor de la misma elevada al cuadrado, nosotros para nuestro estudio tomaremos la definición explicada al comienzo de este apartado.

En nuestro análisis usaremos la distancia de Mahalanobis como el valor que distinguirá entre si una observación forma parte del estado  $R1$  o si forma parte del estado  $R234$ . Haremos dos usos de la distancia de Mahalanobis, uno en el que los estimadores (media y matriz de covarianzas) son no robustos y otro, el principal, en el que habremos obtenido estimadores robustos gracias a la aplicación de la técnica del MCD.

## **5.5 MINIMUM COVARIANCE DETERMINANT ESTIMATOR(MCD)**

El Minimum Covariance Determinant Estimator (MCD) es un método de estimación robusto que fue propuesto por Rousseeuw con la intención de reemplazar al MVE (minimum volume Ellipsoid Estimator), ver 3.5, debido a que se consideraba que este último era demasiado lento.

El objetivo principal del MCD es encontrar el conjunto de  $h$  observaciones, perteneciente del total de  $n$  observaciones, cuya matriz de covarianzas tenga el mínimo determinante [ROU99], siendo  $h$  generalmente la mitad de  $n$ . Al igual que el MVE el MCD no es operativo para cualquier número de observaciones. El número de observaciones para el que se considera que un método de estimación deja de ser operativo se considera punto de ruptura, y en el caso de el MCD este punto viene dado por el valor entero del cociente  $(n+p+1)/2$  donde  $n$  es el número de observaciones de la muestra inicial y  $p$  es el número de variables estudiadas [JEN07]. Se suele considerar este valor de  $h$  porque así se consigue el máximo punto de ruptura próximo al 50%.

Los estimadores obtenidos por el MCD nos ayudan a detectar los outliers de una manera más eficiente porque un determinante pequeño implica que nos encontramos con una submuestra que tiene poca dispersión, lo que hace más complicado que en dicha submuestra haya entrado algún outlier. Sin embargo, el problema principal de estos estimadores es que a medida que aumenta el valor de  $n$  y  $p$ , aumenta exponencialmente el número de posibles submuestras de tamaño  $h$  que analizar, lo que traducido en términos computacionales significa más operaciones y por ende más tiempo y ordenadores más complejos.

A pesar de saberse que el MCD era un método más eficiente que el MVE su traducción a un algoritmo informático era más compleja que para el MVE, por lo que como solución a dicho problema se creó un el algoritmo denominado FAST-MCD [HUB08], basado en los denominados Concentration steps (C-steps), cuya rapidez y eficiencia eran mayores que el MVE. El C-step lo que hace es monitorizar las diferentes distancias de Mahalanobis para  $n$  observaciones y ordenarlas, para luego construir un subconjunto de tamaño  $h$  con las observaciones más cercanas. Nuevos centros y matrices de covarianzas se obtienen en estos C-steps considerando la media muestral y la covarianza muestral de esas  $h$  observaciones más cercanas. Una vez se construye dicho subconjunto obtiene la media y covarianza del mismo [HUB08]. Este proceso puede repetirse hasta o bien el determinante de la matriz de covarianza sea 0 o que se igual al obtenido en el paso anterior.

Un breve resumen de cómo trabaja el FAST-MCD en su conjunto es la siguiente:

- 1) Para construir los primeros subconjuntos utilizamos un número fijado de subconjuntos aleatorios de  $p+1$  variables. De dichos subconjunto obtenemos la medias y la matrices de covarianzas.
- 2) Una vez tenemos los distintos subconjuntos iniciales llevamos a cabo pocas iteraciones del método C-step y seleccionamos un número pequeño de los mejores subconjuntos de tamaño  $h$ , cuyos determinantes de la matriz de covarianzas sean los más pequeños.
- 3) Para finalizar cogemos los subconjuntos seleccionados en el paso 2 y realizamos iteraciones del método C-step hasta que llegamos a un punto en el que apenas hay mejora entre iteraciones sucesivas. Cuando llegamos a este punto significa que hemos llegado a la convergencia del método y hemos obtenidos nuestros dos estimadores robustos, media y matriz de covarianzas.

## **5.6 T<sup>2</sup> DE HOTELLING**

Siendo  $X$  un vector aleatorio  $N_p(\mu, V)$ , una normal  $p$ -variante de media  $\mu$  y matriz de covarianzas  $V$ , la variable  $(x-\mu)'V^{-1}(x-\mu)$  se trata de una distribución chi-cuadrado con  $p$  grados de libertad.

Cuando la media poblacional ( $\mu$ ) y la matriz de varianzas-covarianzas ( $V$ ) son desconocidas, son sustituidas por un estimador apropiado. Los estimadores clásicos son la media muestral y la matriz de varianzas-covarianzas muestral que vienen dados por:

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^n \mathbf{x}_i}{n}$$

$$S_0 = \frac{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'}{n - 1}$$

Podemos apreciar que dicho estadístico es igual al cuadrado de la distancia de mahalanobis que vimos en el apartado anterior. El problema de este estadístico surge cuando existen múltiples outliers en la muestra así como cuando ocurren pequeños cambios en el valor del vector de medias.

Esto se debe a que su punto de ruptura tiene el valor de  $1/n$  por tanto a medida que aumente el tamaño de la muestra el valor se acerca a cero, lo que haría el estadístico  $T_i^2$  inefectivo.

A raíz de este problema surgen las alternativas robustas para los estimadores  $\mu$  y  $\Sigma$  que son los métodos MVE y MCD ya explicados con anterioridad. En estos casos el estimador  $T^2$  quedaría definido:

$$T_{MVE,i}^2 = (\mathbf{x}_i - \overline{\mathbf{x}_{MVE}})' \mathbf{S}_{MVE}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}_{MVE}})$$

$$T_{MCD,i}^2 = (\mathbf{x}_i - \overline{\mathbf{x}_{MCD}})' \mathbf{S}_{MCD}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}_{MCD}})$$

***CAPÍTULO 6:***  
***ANÁLISIS DE LOS***  
***DATOS***



## **6.1 ANALISIS DESCRIPTIVO: Análisis de correlaciones**

Para poder aplicar el algoritmo MCD a cada uno de los ensayos anteriores es necesario un análisis previo de las variables que componen nuestro estudio, tanto de su comportamiento individual, como de las relaciones que existen entre las mismas, con el fin de que la información que saquemos sea la mayor posible a la par que la más precisa, debido a que un número elevado de características o una alta correlación entre algunas, puede dar lugar a una información falseada y a problemas numéricos a la hora de obtener los estimadores robustos.

Existen dos causas principales por las que es de vital importancia el análisis de la correlación entre cada una de las variables de nuestro estudio. La primera es que para llevar a cabo el algoritmo MCD hace falta la matriz inversa de la covarianza, por lo que si en nuestro análisis introducimos variables con una alta correlación vamos a tener problemas numéricos para obtener dicha matriz y la segunda es el problema de sobreinformación debido a la introducción de dos variables cuya relación es lineal en el estudio. Como para las relaciones entre variables si siguen un patrón no importa a que subconjunto pertenezcan, introducimos en statgraphics las 1700 observaciones con las 24 características sobre las que realizaremos el ensayo, y llevamos a cabo un análisis multivariante. Cabe destacar que como aparece en el capítulo 4 sabemos de antemano que hay variables que están linealmente relacionadas entre sí como son los Momentos 1 y 2 con los Cumulantes 1 y 2 por lo que tomamos la decisión de eliminar los Momentos 1 y 2 antes de introducir los datos en el software Statgraphics.

Para tomar la decisión de eliminar o no variables, hemos llevado a cabo un análisis para ver qué relaciones existen entre las variables que se han medido para la realización del trabajo. Una vez que se obtuvo la matriz de correlaciones (figura 6.1) se observó el coeficiente de correlación entre cada una de las variables y se determinó que para no tener problemas numéricos en el transcurso del experimento, íbamos a considerar relación lineal aquellos coeficientes de correlación que sean mayores de 0.95.

Como consecuencia de lo observado en la matriz de correlaciones se llegó a la conclusión de eliminar las siguientes variables: Cumulante 2, Cumulante 4, Momento 4 y Absolute mean.

	VelocidadC	BandInf	BandSup	BandInf2	BandSup2	Band5Inf	Band5Sup
VelocidadC		0,217	0,2029	0,1224	0,065	0,1763	0,1287
BandInf	0,217		0,8788	0,3739	0,4535	0,7414	0,3393
BandSup	0,2029	0,8788		0,3425	0,3962	0,6733	0,4041
BandInf2	0,1224	0,3739	0,3425		0,7332	0,4322	0,3978
BandSup2	0,065	0,4535	0,3962	0,7332		0,4493	0,4455
Band5Inf	0,1763	0,7414	0,6733	0,4322	0,4493		0,3882
Band5Sup	0,1287	0,3393	0,4041	0,3978	0,4455	0,3882	
Band7Inf	0,4271	0,4196	0,374	0,3574	0,3444	0,4439	0,4938
Band7Sup	0,0962	0,2903	0,3729	0,4047	0,3983	0,3735	0,7421
Momento1	-0,1386	-0,0886	-0,0796	-0,0977	-0,1284	-0,0632	-0,0391
Momento2	-0,3666	-0,0341	-0,0964	0,1858	0,2187	0,0838	0,1361
Momento3	0,0736	0,0605	0,0062	0,2125	0,2433	0,041	0,1592
Momento4	-0,3103	-0,0196	-0,0875	0,1825	0,2153	0,0799	0,1239
Cumulante1	-0,1386	-0,0886	-0,0796	-0,0977	-0,1284	-0,0632	-0,0391
Cumulante2	-0,3666	-0,0341	-0,0964	0,1858	0,2188	0,0838	0,1361
Cumulante3	0,2227	0,1222	0,0798	0,1942	0,2417	0,0518	0,1144
Cumulante4	0,3199	0,0236	0,0924	-0,1802	-0,2154	-0,0836	-0,1236
Cumulante6	0,4381	0,0447	0,0741	-0,1883	-0,2066	-0,1799	-0,0855
Skewness	0,1361	0,0497	0,0093	0,184	0,1897	0,0374	0,189
Kurtosis	0,508	0,0949	0,1276	-0,1308	-0,1756	-0,1383	-0,1199
AbsoluteMean	-0,3985	-0,0439	-0,1005	0,1825	0,2147	0,0817	0,1408
PicoAbsoluto	-0,2851	-0,03	-0,0989	0,2551	0,2542	0,0477	0,1478
xr	-0,4041	-0,048	-0,1031	0,1755	0,2098	0,0784	0,1439
FactorCresta	0,355	0,0248	0,0164	0,0416	-0,0237	-0,1028	-0,0323
FactorForma	0,4377	0,1412	0,1427	0,0416	-0,0295	0,0212	-0,1657

	Band7Inf	Band7Sup	Momento 1	Momento 2	Momento 3	Momento 4	Cumulante1
VelocidadC	0,4271	0,0962	-0,1386	-0,3666	0,0736	-0,3103	-0,1386
BandInf	0,4196	0,2903	-0,0886	-0,0341	0,0605	-0,0196	-0,0886
BandSup	0,374	0,3729	-0,0796	-0,0964	0,0062	-0,0875	-0,0796
BandInf2	0,3574	0,4047	-0,0977	0,1858	0,2125	0,1825	-0,0977
BandSup2	0,3444	0,3983	-0,1284	0,2187	0,2433	0,2153	-0,1284
Band5Inf	0,4439	0,3735	-0,0632	0,0838	0,041	0,0799	-0,0632
Band5Sup	0,4938	0,7421	-0,0391	0,1361	0,1592	0,1239	-0,0391
Band7Inf		0,4944	-0,1147	0,0098	0,0546	0,0216	-0,1147
Band7Sup	0,4944		-0,0282	0,1416	0,2365	0,1175	-0,0282
Momento1	-0,1147	-0,0282		0,0501	0,1334	0,0429	1
Momento2	0,0098	0,1416	0,0501		0,1525	0,9825	0,0501
Momento3	0,0546	0,2365	0,1334	0,1525		0,1413	0,1334
Momento4	0,0216	0,1175	0,0429	0,9825	0,1413		0,0429
Cumulante1	-0,1147	-0,0282	1	0,0501	0,1334	0,0429	
Cumulante2	0,0098	0,1416	0,05	1	0,1524	0,9825	0,05
Cumulante3	0,1139	0,165	-0,6527	-0,1048	0,6165	-0,1045	-0,6527
Cumulante4	-0,0181	-0,117	-0,0352	-0,9809	-0,1271	-0,9989	-0,0352
Cumulante6	0,0474	-0,1139	0,0488	-0,5861	0,1085	-0,511	0,0488
Skewness	0,0363	0,2586	0,1156	-0,001	0,9175	-0,0022	0,1156
Kurtosis	0,1015	-0,1403	0,0496	-0,6219	0,0822	-0,558	0,0496
AbsoluteMean	-0,0019	0,1545	0,0524	0,9942	0,154	0,9576	0,0524
PicoAbsoluto	0,019	0,1676	0,0439	0,9093	0,2535	0,889	0,0439
xr	-0,0071	0,157	0,0511	0,9927	0,15	0,9559	0,0511
FactorCresta	0,029	-0,0373	-0,0399	-0,5006	0,1199	-0,4351	-0,0399

	Cumulante2	Cumulante3	Cumulante4	Cumulante6	Skewness	Kurtosis	AbsMean
VelocidadC	-0,3666	0,2227	0,3199	0,4381	0,1361	0,508	-0,3985
BandInf	-0,0341	0,1222	0,0236	0,0447	0,0497	0,0949	-0,0439
BandSup	-0,0964	0,0798	0,0924	0,0741	0,0093	0,1276	-0,1005
BandInf2	0,1858	0,1942	-0,1802	-0,1883	0,184	-0,1308	0,1825
BandSup2	0,2188	0,2417	-0,2154	-0,2066	0,1897	-0,1756	0,2147
Band5Inf	0,0838	0,0518	-0,0836	-0,1799	0,0374	-0,1383	0,0817
Band5Sup	0,1361	0,1144	-0,1236	-0,0855	0,189	-0,1199	0,1408
Band7Inf	0,0098	0,1139	-0,0181	0,0474	0,0363	0,1015	-0,0019
Band7Sup	0,1416	0,165	-0,117	-0,1139	0,2586	-0,1403	0,1545
Momento1	0,05	-0,6527	-0,0352	0,0488	0,1156	0,0496	0,0524
Momento2	1	-0,1048	-0,9809	-0,5861	-0,001	-0,6219	0,9942
Momento3	0,1524	0,6165	-0,1271	0,1085	0,9175	0,0822	0,154
Momento4	0,9825	-0,1045	-0,9989	-0,511	-0,0022	-0,558	0,9576
Cumulante1	0,05	-0,6527	-0,0352	0,0488	0,1156	0,0496	0,0524
Cumulante2		-0,1047	-0,9809	-0,5861	-0,0011	-0,6219	0,9942
Cumulante3	-0,1047		0,1084	0,1424	0,5986	0,1263	-0,1036
Cumulante4	-0,9809	0,1084		0,535	0,014	0,5868	-0,9559
Cumulante6	-0,5861	0,1424	0,535		0,2053	0,9345	-0,6205
Skewness	-0,0011	0,5986	0,014	0,2053		0,1456	-0,0009
Kurtosis	-0,6219	0,1263	0,5868	0,9345	0,1456		-0,6552
AbsoluteMean	0,9942	-0,1036	-0,9559	-0,6205	-0,0009	-0,6552	
PicoAbsoluto	0,9093	-0,0066	-0,8786	-0,3781	0,1122	-0,4268	0,9067
xr	0,9927	-0,1048	-0,9548	-0,6188	-0,0001	-0,6645	0,9994
FactorCresta	-0,5006	0,2107	0,4476	0,736	0,223	0,6668	-0,5303
FactorForma	-0,4543	0,0608	0,4386	0,3613	-0,0461	0,6526	-0,4836

	xr	FactorCresta	FactorForma
VelocidadC	-0,4041	0,355	0,4377
BandInf	-0,048	0,0248	0,1412
BandSup	-0,1031	0,0164	0,1427
BandInf2	0,1755	0,0416	0,0416
BandSup2	0,2098	-0,0237	-0,0295
Band5Inf	0,0784	-0,1028	0,0212
Band5Sup	0,1439	-0,0323	-0,1657
Band7Inf	-0,0071	0,029	0,1608
Band7Sup	0,157	-0,0373	-0,1661
Momento1	0,0511	-0,0399	0,0048
Momento2	0,9927	-0,5006	-0,4543
Momento3	0,15	0,1199	-0,0064
Momento4	0,9559	-0,4351	-0,4208
Cumulante1	0,0511	-0,0399	0,0048
Cumulante2	0,9927	-0,5006	-0,4543
Cumulante3	-0,1048	0,2107	0,0608
Cumulante4	-0,9548	0,4476	0,4386
Cumulante6	-0,6188	0,736	0,3613
Skewness	-0,0001	0,223	-0,0461
Kurtosis	-0,6645	0,6668	0,6526
AbsoluteMean	0,9994	-0,5303	-0,4836
PicoAbsoluto	0,9042	-0,1341	-0,3767
xr		-0,5301	-0,513
FactorCresta	-0,5301		0,2845
FactorForma	-0,513	0,2845	

Tabla 6.1. Matriz de Correlaciones

Una vez hemos eliminado dichas variables obtenemos los siguientes resúmenes estadísticos para las variables no eliminadas, que se muestran en la tabla 6.1 obtenida con Statgraphics.

**Resumen Estadístico**

	<b>VelocidadC</b>	<b>BandInf</b>	<b>BandSup</b>	<b>BandInf2</b>	<b>BandSup2</b>	<b>Band5Inf</b>
<b>Recuento</b>	1679	1679	1679	1679	1679	1679
<b>Promedio</b>	1450,21	-62,3998	-61,5327	-69,7237	-71,1299	-75,3469
<b>Desviación Estándar</b>	337,778	8,40505	9,68522	7,26788	7,35293	7,61499
<b>Coefficiente de Variación</b>	23,2917%	-13,4697%	-15,74%	-10,4238%	-10,3373%	-10,1066%
<b>Mínimo</b>	965,786	-85,2774	-86,2632	-102,926	-106,247	-112,214
<b>Máximo</b>	1941,43	-6,1816	0	-34,2789	-33,2811	-46,0962
<b>Rango</b>	975,643	79,0958	86,2632	68,6471	72,9661	66,1177
<b>Sesgo Estandarizado</b>	-0,262859	8,10508	20,4176	8,42702	9,68566	-0,0329879
<b>Curtosis Estandarizada</b>	-10,5222	11,2641	44,421	16,0545	25,7382	1,93917

	<b>Band5Sup</b>	<b>Band7Inf</b>	<b>Band7Sup</b>	<b>Momento3</b>	<b>Cumulante1</b>	<b>Cumulante3</b>
<b>Recuento</b>	1679	1679	1679	1679	1679	1679
<b>Promedio</b>	-81,3199	-83,7533	-85,9506	0,0414716	0,00742177	-0,0470482
<b>Desviación Estándar</b>	8,36454	6,93721	7,89983	0,0996121	0,00804941	0,133048
<b>Coefficiente de Variación</b>	-10,286%	-8,2829%	-9,19114%	240,194%	108,457%	-282,79%
<b>Mínimo</b>	-112,7	-116,278	-116,395	-0,0872723	-0,0238234	-0,503179
<b>Máximo</b>	-38,8166	-49,6337	-34,7662	0,356531	0,032662	0,544873
<b>Rango</b>	73,8832	66,6439	81,6287	0,443804	0,0564854	1,04805
<b>Sesgo Estandarizado</b>	19,6501	6,21228	23,7967	27,5181	-1,77833	5,80739
<b>Curtosis Estandarizada</b>	30,6863	18,7945	44,4674	12,8395	-0,263838	9,48428

	<b>Cumulante6</b>	<b>Skewness</b>	<b>Kurtosis</b>	<b>PicoAbsoluto</b>	<b>xr</b>	<b>FactorCresta</b>
<b>Recuento</b>	1679	1679	1679	1679	1679	1679
<b>Promedio</b>	2,59687	0,00519751	1,52308	3,09339	1,5949	1,59457
<b>Desviación Estándar</b>	0,108923	0,0123637	0,0243382	0,442866	0,266839	0,118903
<b>Coefficiente de Variación</b>	4,19438%	237,877%	1,59796%	14,3165%	16,7307%	7,45675%
<b>Mínimo</b>	2,32552	-0,0105344	1,45573	2,13696	0,934932	1,38981
<b>Máximo</b>	3,09394	0,0442569	1,61511	5,47942	2,26119	2,94878
<b>Rango</b>	0,768413	0,0547913	0,159372	3,34245	1,32626	1,55897
<b>Sesgo Estandarizado</b>	23,7718	21,4533	9,24157	2,51555	-8,69175	65,8998
<b>Curtosis Estandarizada</b>	27,5474	5,87852	0,896615	8,00794	-1,54003	284,039

	<i>FactorForma</i>
Recuento	1679
Promedio	1,1153
Desviación Estándar	0,00714637
Coefficiente de Variación	0,640757%
Mínimo	1,10084
Máximo	1,14815
Rango	0,0473134
Sesgo Estandarizado	32,362
Curtosis Estandarizada	38,5527

**Tabla 6.1.** Resumen estadístico de las variables no eliminadas

## **6.2 DESCRIPCIÓN Y CARACTERÍSTICAS DE LOS ENSAYOS.**

Como hemos explicado inicialmente, lo que buscamos con este proyecto entre otras cosas, es mediante la capacidad discriminante de un número reducido de variables, realizar una clasificación adecuada para motores similares a los propios del experimento. Para nuestro estudio buscamos una clasificación de forma que distingamos entre motores completamente sanos (R1) respecto a motores que tienen alguna deficiencia (R2R3R4), por lo tanto de ahora en adelante representaremos las clasificaciones distinguiendo dos estados que serán R1-R234.

Debido a que existen un número muy elevado de variables, solo se estudiara un número reducido que se seleccionara mediante procesos de análisis discriminante. Tras llevar a cabo dichos análisis sobre los 7 ensayos individuales diferentes (ver Capítulo 4) y debido a los resultados obtenidos que se podrán ver en el siguiente capítulo, las variables elegidas son BandSup, Band5Sup, Band5Inf, Band7Sup (ver Capítulo 4 para explicación).

Antes de evaluar el poder discriminante que poseen dichas variables seleccionadas, se sugirió por parte del Departamento de Ingeniería Eléctrica el enfrentar una variable de carga del motor como el desplazamiento, contra dichas variables para ver tanto el poder de discriminación como la sospecha de que para valores muy pequeños del desplazamiento se obtienen valores falseados de las variables. En las gráficas 6.2, 6.3, 6.4 y 6.5 se puede apreciar como para las 4 variables a partir de valores inferiores a un desplazamiento del 0.01% se produce un solapamiento claro de los valores que hace que se pierda el poder discriminante de las variables para los distintos estados.

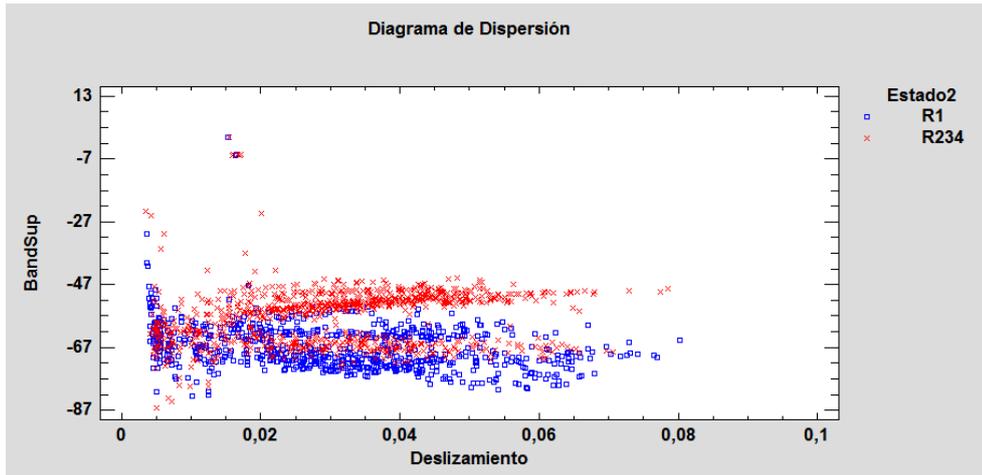


Figura 6.2 Distribución de la variable BandSup

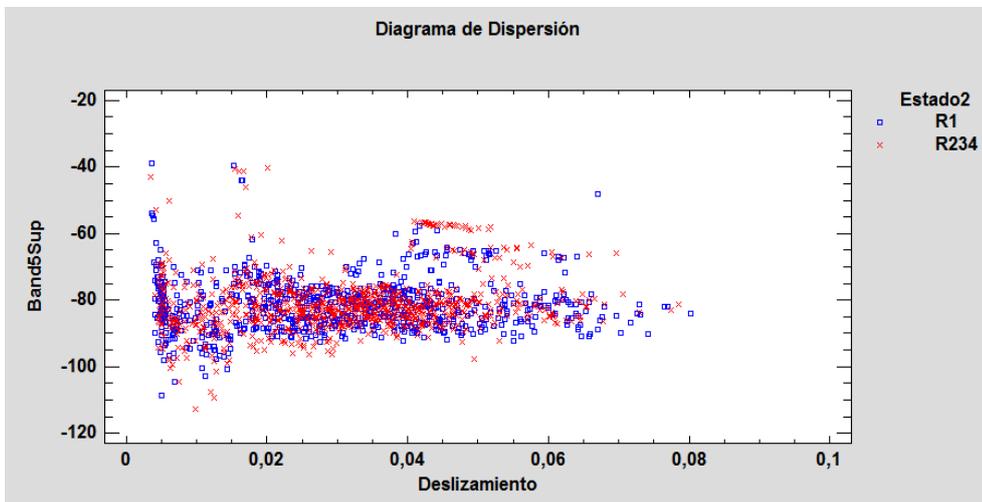


Figura 6.3 Distribución de la variable Band5Sup

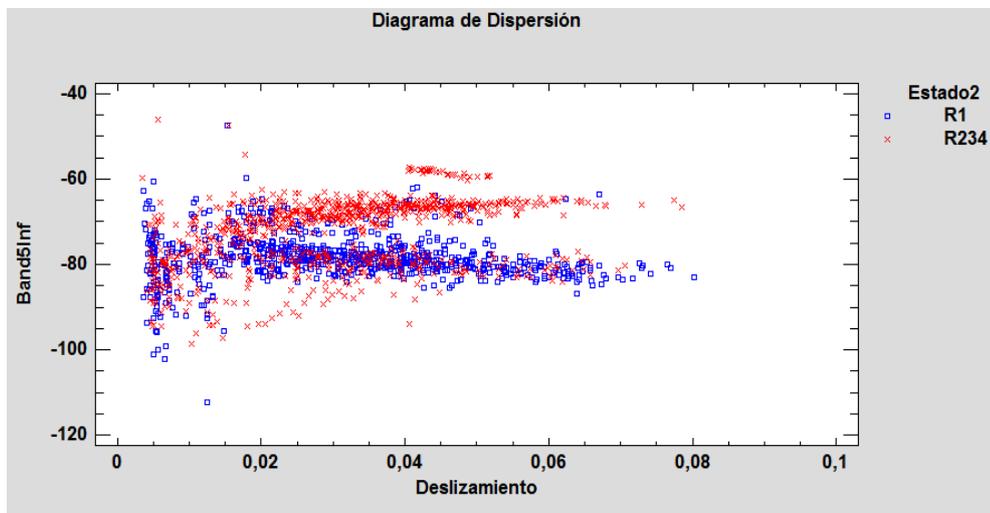


Figura 6.4 Distribución de la variable Band5Inf

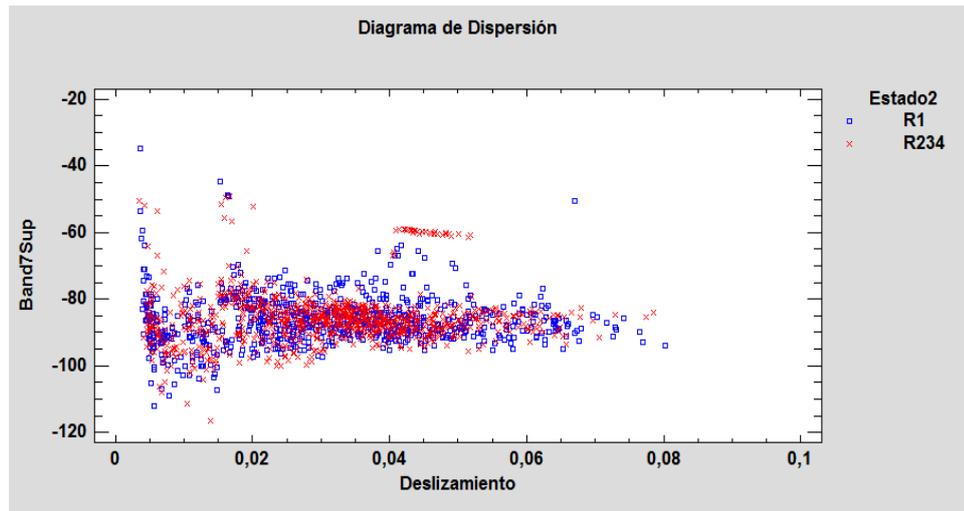


Figura 6.5 Distribución de la variable Band7Sup

Una vez que hemos descartado las observaciones cuyo deslizamiento está por debajo del 1% decidimos estudiar si se producían solapamientos entre los valores de los estados que queríamos diferenciar. Esto se hizo debido a que se esperan comportamientos diferentes entre los distintos estados y realizar un análisis previo de como se distribuyen para cada variable seleccionada las observaciones generales nos ayudara a establecer una base para las conclusiones que saquemos después de aplicar los distintos procedimientos.

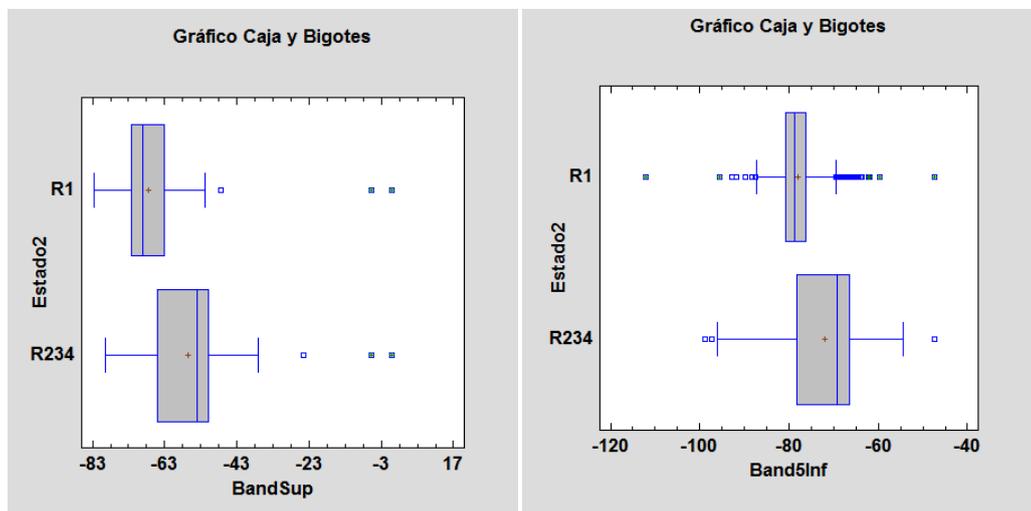


Figura 6.6 Box-Plot para la variable BandSup

Figura 6.7 Box-Plot para la variable Band5Inf

Analizando los gráficos superiores (6.6 y 6.7) que corresponden a la distribución de las observaciones para las variables BandSup y Band5Inf podemos determinar como existen un número de observaciones cuyos comportamientos son muy parecidos aunque se encuentren en diferentes estados. Una vez que acabemos el análisis y como veremos más adelante,

sabremos que este solapamiento se debe a las observaciones R2 que tiene comportamientos similares a las de R1 en bastantes ocasiones. No obstante nos sirve para ver sin necesidad de aplicar ningún algoritmo que existe una diferencia clara de comportamiento en función del estado en el que se encuentre la observaciones, lo que apoya el propósito de este trabajo.

No obstante no debemos olvidar que estos gráficos son respecto al conjunto de las observaciones sin reparar en que variador ni frecuencia pertenecen por lo que lo normal es que se produzca solapamiento como ocurre en las figuras 6.8 y 6.9 que representan las variables Band5Sup y Band7Sup.

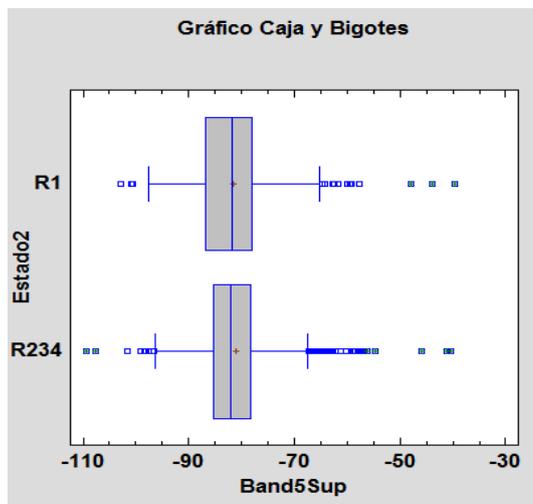


Figura 6.8 Box-Plot para la variable Band5Sup

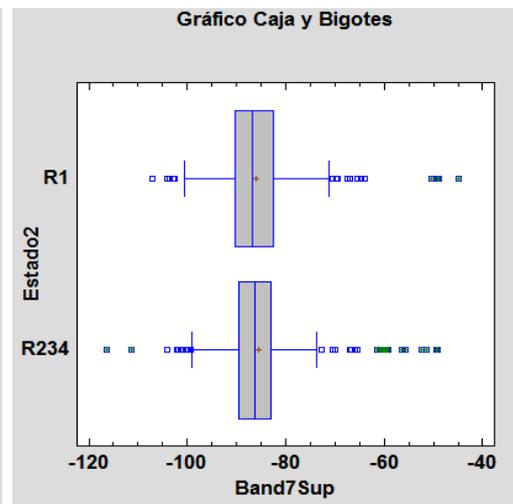


Figura 6.9 Box-Plot para la variable Band7Sup

Debido a esto y como ya se comentó con anterioridad nosotros llevaremos a cabo estudios más concretos en los que será más fácil visualizar los comportamientos diferentes por niveles, tal y como se puede ver en los gráficos 6.10 y 6.11 que muestran las mismas variables que los anteriores pero para el ensayo AB35.

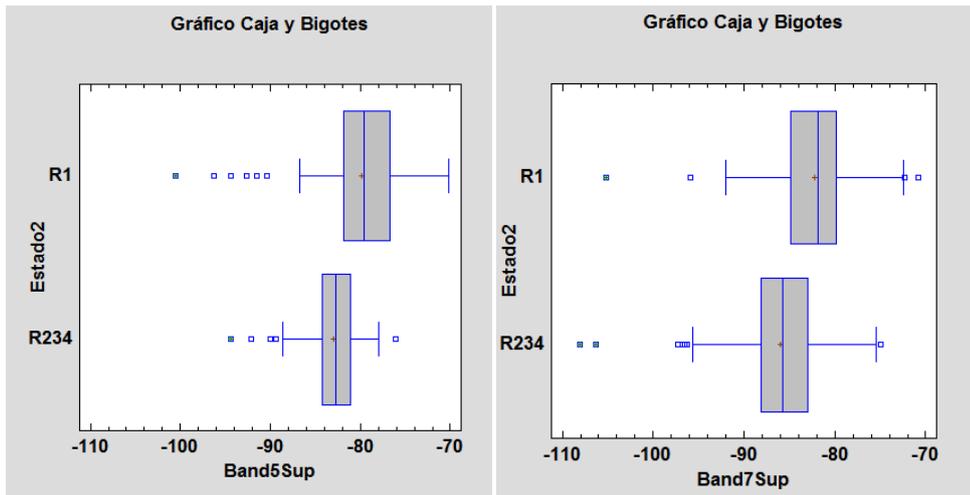


Figura 6.10 Box-Plot de AB35 de la variable Band5Sup Figura 6.11 Box-Plot de AB35 de la variable Band5Sup

Tal y como mencionamos con anterioridad la diferencia entre como se comportan para esas variables es mucho más apreciable a medida que hacemos un ensayo más específico.

No obstante y dado la manera en la que vamos a trabajar para obtener la clasificación de motores, conviene que veamos una representación gráfica por pares de variables de forma que veamos como representar variables juntas ayuda al algoritmo a obtener una mejor clasificación. A continuación mostraremos varios ejemplos de cómo se distribuyen las observaciones para algunos de los ensayos y atendiendo a dos clasificaciones diferentes que serán R1-R234 y R1-R2-R3-R4.

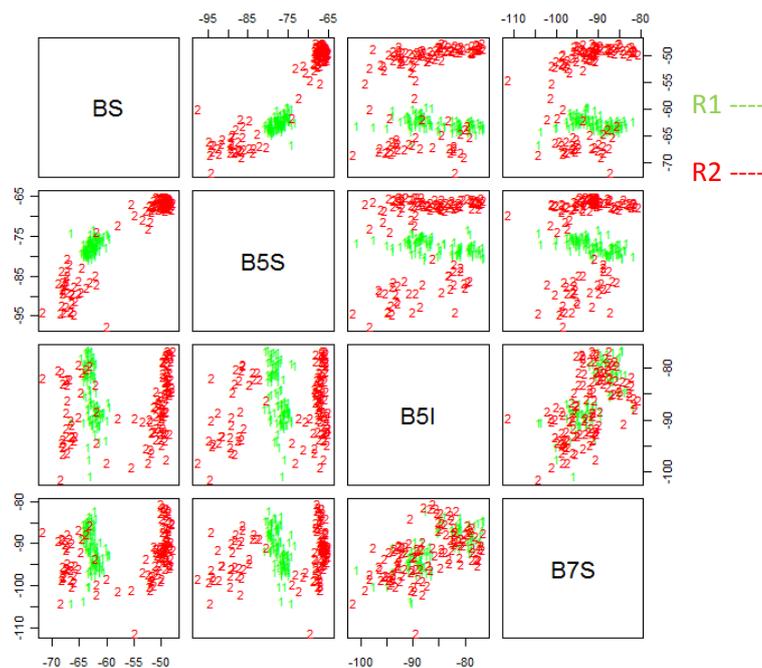
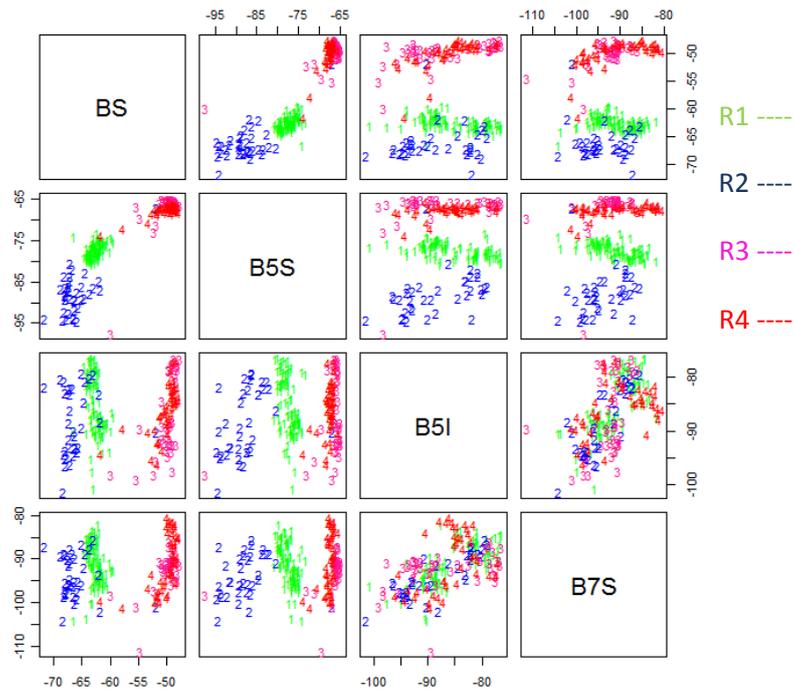


Figura 6.12 Representación pareada de las variables para el ensayo AB65 (R1-R234)



**Figura 6.13** Representación pareada de las variables para el ensayo AB65 (R1-R2-R3-R4)

Las gráficas 6.12 y 6.13 muestran la representación de pares para el ensayo AB65 distinguiendo la gráfica 6.12 entre los estados R1-R234 y la 6.13 entre R1-R2-R3-R4. Observando con detenimiento la discriminación R1-R234 como la formación de grupos es muy clara, lo que hace que nuestro método que busca grupos tenga alta eficiencia, cosa que podremos ver posteriormente en el Capítulo 8. Surge la duda acerca de las observaciones que se encuentran clasificadas como R234 que se encuentran distribuidas en dos grupos colocados a ambos lados de las R1, pero mirando la gráfica 6.13 vemos como por un lado se colocan las observaciones en estado R2 y en el otro las pertenecientes a las R3 y R4. Otra de las cosas que nos llama la atención es la proximidad existente entre las observaciones en estados R1-R2 entre sí, y que ocurre algo muy parecido entre las R3-R4. Más adelante veremos que este problema se traduce en la clasificación poco eficiente de las observaciones en estado R2 que se encuentran muy próximas en varios casos a R1 y nosotros queremos discriminar entre R1 y R234.

En cuanto a si es correcto usar estas cuatro variables, ya hemos comentado que se aprecia con claridad la formación de grupos, excepto en las representaciones que está involucrada Band7Sup cuya claridad es más reducida. No hay que olvidar como veremos más adelante que la elección de que variables usar para llevar a cabo el estudio se hizo cogiendo aquellas cuatro que más se repetían o eran más significantes en cada uno de los ensayos realizados. En las gráficas 6.14 y 6.15 veremos como si tiene un poder mayor de discriminación.

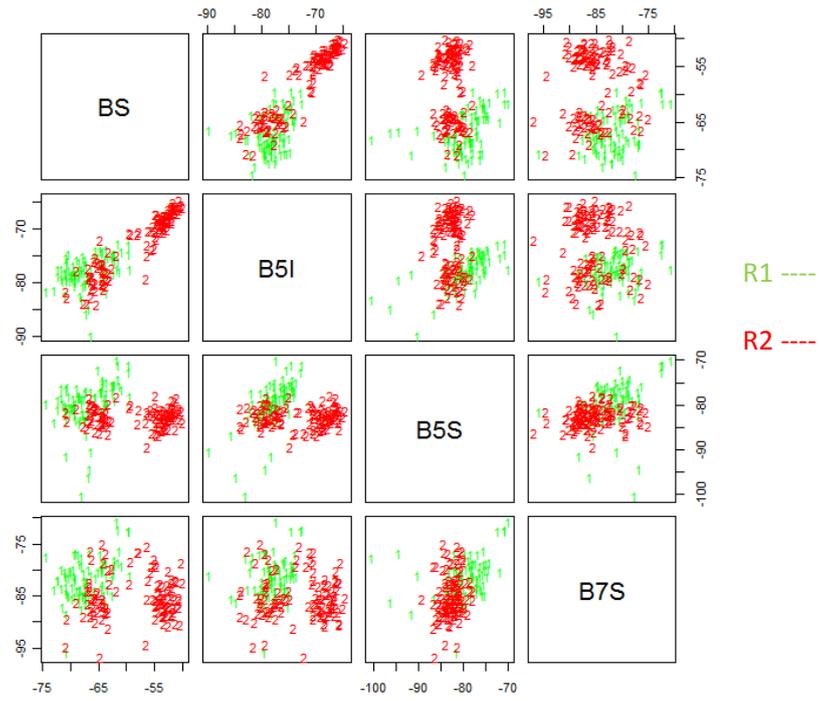


Figura 6.14 Representación pareada de las variables para el ensayo SM50 (R1-R234)

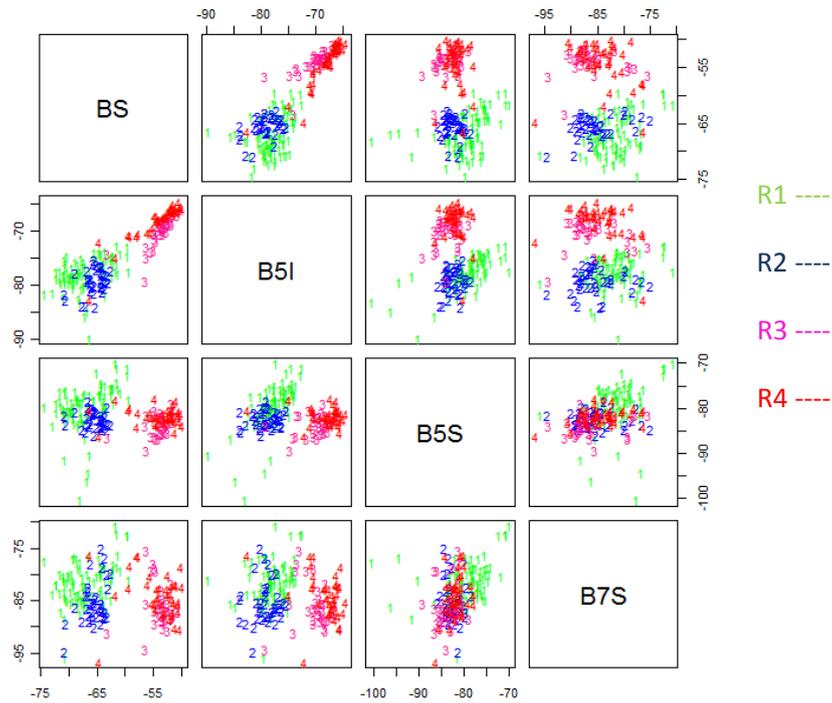


Figura 6.15 Representación pareada de las variables para el ensayo SM50 (R1-R2-R3-R4)

Las gráficas 6.14 y 6.15 muestran lo mismo que hemos comentado para las 6.12 y 6.13 pero en este caso el ensayo cogido es el SM50. Podemos observar que la claridad en la formación de grupos no es tan elevada como en el ensayo AB65 y eso se traducirá en la eficiencia del método para ese ensayo como veremos en el capítulo 8. Tal y como ocurre para AB65, las observaciones R1 y R2 se encuentran próximas entre sí, al igual que las R3 y R4, por lo que que esta distribución se dé en estas dos gráficas nos invita a pensar, posteriormente se verá con más detenimiento, la semejanza de comportamientos entre estos estados entre sí.

En relación a lo que comentábamos antes respecto a la variable Band7Sup, podemos ver como en este caso la diferencia entre unos estados y otros es más visible, mientras que otras han perdido claridad respecto al anterior. Este problema se repetirá con todas las variables, que en un ensayo muestren bastante representadas junto a otras, y que en otros apenas muestren diferencia clara entre los distintos comportamientos por estados.

Los problemas que hemos mostrado en estos dos ejemplos de ensayos individuales, se multiplican cuando realizamos un ensayo de los denominados grupales como el mostrado en las figuras 6.16 y 6.17. Estos problemas entre otras cosas vienen determinados por un mayor número de observaciones que da pie a más comportamientos diferentes o posibles grupos en los que fijarse.

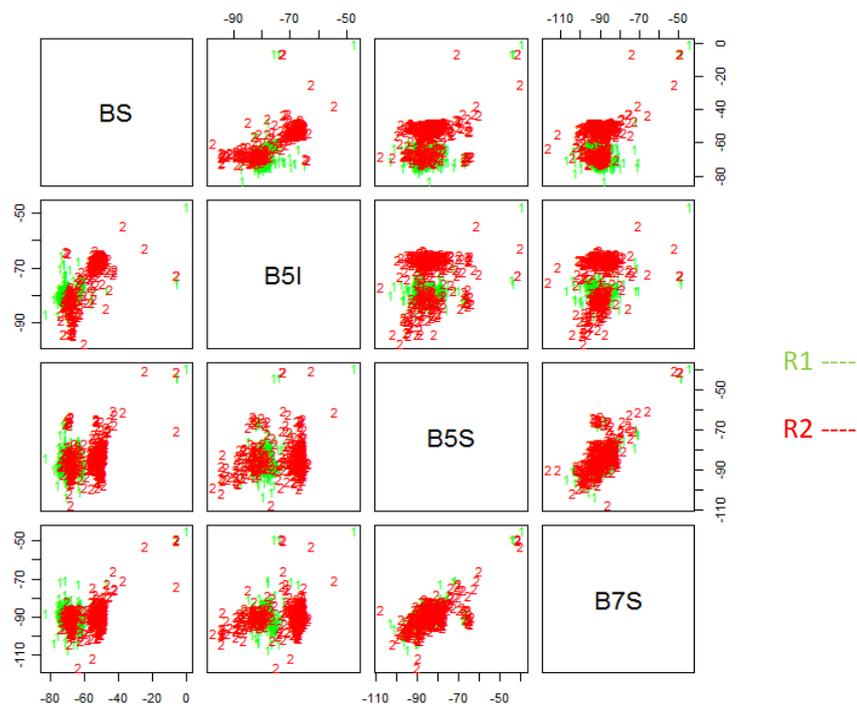


Figura 6.16 Representación pareada de las variables para el ensayo AB (R1-R234)

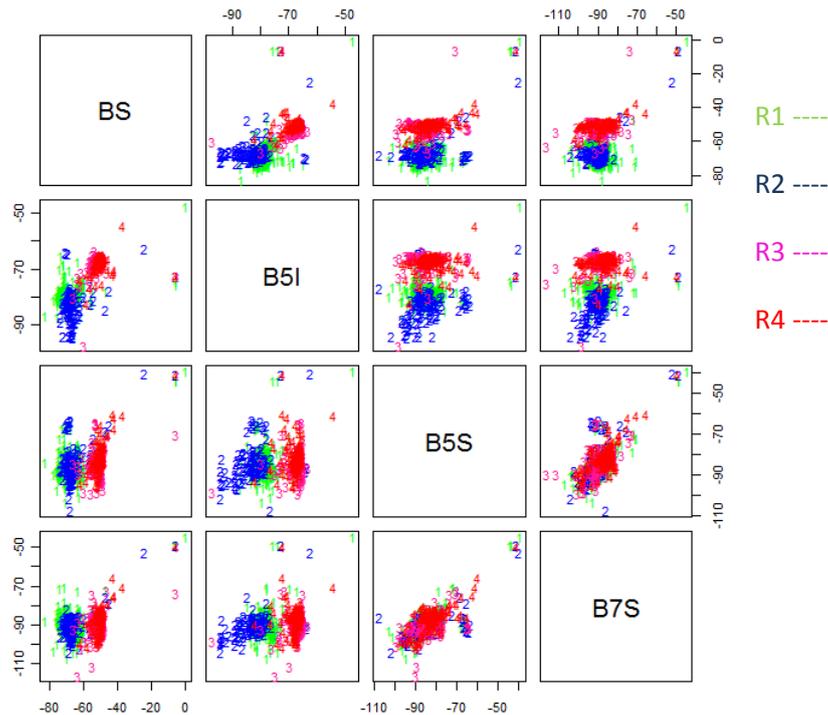


Figura 6.17 Representación pareada de las variables para el ensayo AB (R1-R2-R3-R4)

Si miramos con detenimiento las gráficas de pares para el ensayo AB vemos claramente en la gráfica 6.16 correspondiente a la discriminación que vamos a estudiar nosotros, R1-R234, que resulta casi imposible buscar grupos bien definidos sobre los que trabajar. Esto nos da la idea que más adelante será más clara aún de que no se puede aplicar nuestro método a ensayos grupales. En el Capítulo 8 veremos como el acierto en la detección correcta del estado correspondiente a cada observación es mucho menor que en los ensayos individuales. El caso de la discriminación R1-R2-R3-R4 si que parece observando las gráficas obtendríamos mejores resultados, pero no podemos decirlo con seguridad porque no es objetivo de nuestro trabajo y no lo estudiaremos.

### 6.3 ANÁLISIS MULTIVARIANTE

Como hemos dicho anteriormente lo que pretendemos es construir clasificadores que consigan discriminar la combinación de estados R1-R234, es decir, diferenciar entre una barra sana y una barra con cualquier tipo de rotura ya sea leve o grave o lo que es lo mismo, realizar un control de calidad para los motores.

Tal y como mencionamos con anterioridad un gran número de variables en nuestro análisis puede conllevar problemas de sobreajuste, lo que sumado a

qué buscamos un método que ayude a la reducción de costes, basar nuestro estudio en 20 variables no tiene mucho sentido.

El procedimiento a llevar será utilizar los métodos paso a paso en cada uno de los 7 experimentos individuales ya descritos para poder decidir que cuatro variables son las más significativas en conjunto, por lo que aplicaremos los métodos Backward y forward de acuerdo al estadístico F (ver Capítulo 5) que usando los parámetros que vienen por defecto en el software Statgraphics, es decir,  $F_{in}=F_{out}=4$ . Tal y como dijimos inicialmente, buscamos clasificar los motores en el estado R1 o en el R234, y es el criterio que se usó para elegir las 4 variables. No obstante, debido a los resultados obtenidos tras aplicar los procedimientos (Tablas 6.1-6.7) no nos fijaremos tanto en los coeficientes de las variables elegidas que se obtienen a la hora de aplicar los dos métodos sino que nos fijaremos en cuáles son los que más se repiten entre todos los análisis. En dichas tablas lo que veremos será los diferentes coeficientes estandarizados que se utilizarían para predecir a que nivel pertenecerían las nuevas observaciones si estuviéramos trabajando con análisis discriminante simple. Como ejemplo de lo que hablamos tomaremos los valores para AB35. Los valores que vemos en las tablas 6.1a son los coeficientes estandarizados de cada variable seleccionada para la primera función discriminante construida para distinguir entre los distintos valores que existen para el estado, es decir, distinguir entre R1 y R234. Dicha función sería:  $0,539641*BandInf+0,586222*Band5Inf-0,547456*Band7Inf-2,82147*Cumulante6+ 3,5644*Kurtosis$ .

### 6.3.1 AB35

En la tabla 6.1 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo AB35.

Backward		Forward	
BandInf	0,5396	BandInf	0,5396
Band5Inf	0,5862	Band5Inf	0,5862
Band7Inf	-0,5475	Band7Inf	-0,5475
Cumulante6	-2,8215	Cumulante6	-2,8215
Kurtosis	3,5644	Kurtosis	3,5644

**Tabla 6.1.** Variables y valores función discriminante

### 6.3.2AB50

En la tabla 6.2 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo AB50.

Backward		Forward	
FactorCresta	-0,7486	FactorCresta	-1,0439
AbsMean	0,7693	AbsMean	0,2702
BandSup	1,1922	BandSup	1,0814
Band5Sup	-0,4333	Band5Sup	-0,4957
Momento3	1,0857	Cumulante3	-0,2696
Cumulante3	-0,3121	Kurtosis	1,3137
Cumulante6	1,3933	BandSup2	0,2751
Skewness	-1,08764		

Tabla 6.2. Variables y valores función discriminante

### 6.3.3 AB65

En la tabla 6.3 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo AB65.

Backward		Forward	
AbsMean	-2,0422	BandSup	1,8258
BandSup	1,9239	BandInf	1,2809
BandInf	1,3611	BandSup2	0,3493
BandSup2	0,3461	Band5Inf	-2,8902
Band5Inf	-3,1253	Band7Inf	0,2968
Band7Inf	0,4001	Momento3	0,6911
Cumulante1	-0,7447	Skewness	0,5115
Cumulante3	-0,8057		
Cumulante6	0,602		
Kurtosis	-1,7676		

Tabla 6.3 Variables y valores función discriminante

### 6.3.4 RED50

En la tabla 6.4 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo RED50.

Backward		Forward	
Momento3	-0,434	Skewness	0,3103
Cumulante1	2,2826	Kurtosis	0,51
Cumulante3	2,3555	Band5Inf	-0,6929
Kurtosis	-0,5361	Band5Sup	-0,8262
xr	0,3737	Band7Sup	0,3925
Band5Inf	0,6716	BandSup	1,602
Band5Sup	0,9117		
Band7Sup	-0,4053		
BandSup	-1,7011		

**Tabla 6.4** Variables y valores función discriminante

### 6.3.5 SM35

En la tabla 6.5 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo SM35.

Backward		Forward	
Band7Inf	-0,2284	Band7Inf	-0,2133
BandInf2	0,4635	BandInf2	0,4291
Cumulante3	-0,3975	Cumulante3	-0,3581
Skewness	-0,4012	Skewness	-0,1946
Kurtosis	1,174	FactorForma	0,9158
FactorCresta	-1,0657	BandSup	1,0899
BandSup	1,2096	Band5Sup	-0,5705
Band5Sup	-0,6112	Band5Inf	0,3032
Band7Sup	-0,2915	Band7Sup	-0,3146
Kurtosis	-1,7676		

**Tabla 6.5** Variables y valores función discriminante

### 6.3.6 SM50

En la tabla 6.6 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo SM50.

Backward		Forward	
BandInf	-0,7032	BandInf	-0,538
BandSup2	-0,2587	Momento3	-1,183
Momento3	-1,2599	Skewness	2,7714
Skewness	2,9217	Kurtosis	1,5381
Kurtosis	1,5132	xr	1,883
xr	2,0943	FactorCresta	-1,6956
FactorCresta	-1,7371	FactorForma	-1,1661
FactorForma	-1,0686	BandSup	1,0689
BandSup	0,9081	Band7Sup	-0,3002
Band5Inf	0,4131		
Band7Sup	-0,2734		

Tabla 6.6 Variables y valores función discriminante

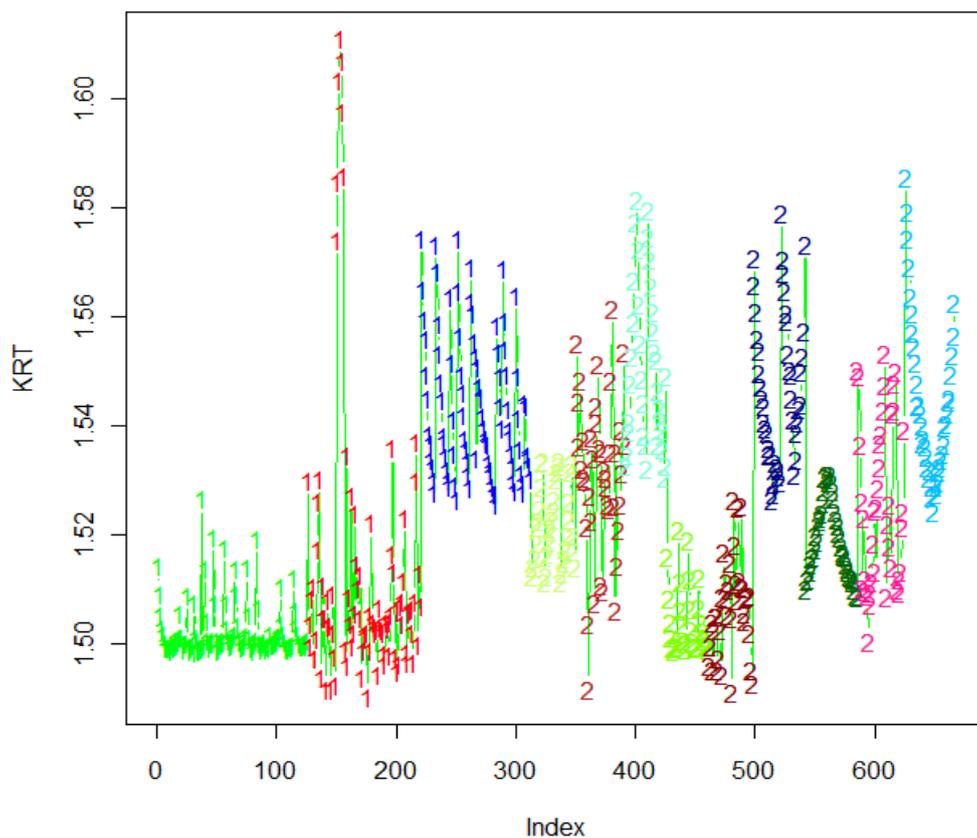
### 6.3.7 SM65

En la tabla 6.7 se muestran las variables y coeficientes de las funciones discriminantes obtenidas usando los métodos de selección Backward y forward para el ensayo SM65.

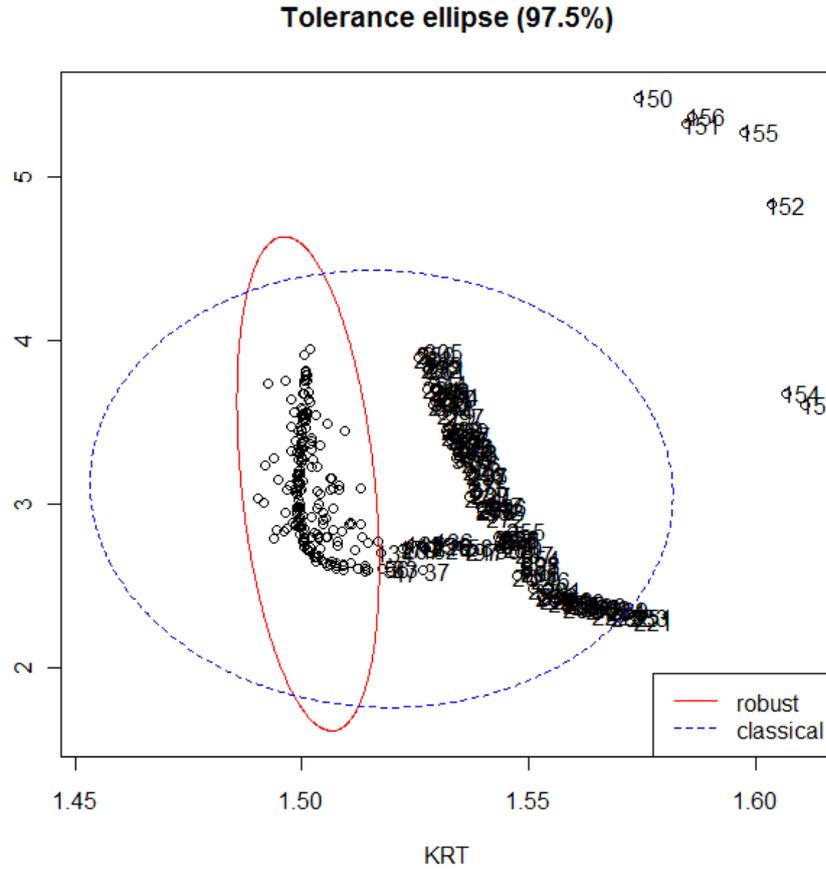
Backward		Forward	
Cumulante3	0,3363	Cumulante3	0,3363
Skewness	5,2477	Skewness	5,2477
Kurtosis	-0,9175	Kurtosis	-0,9175
Xr	2,1997	xr	2,1997
FactorCresta	-0,9176	FactorCresta	-0,9176
FactorForma	-1,7308	FactorForma	-1,7308
BandSup	-0,3635	BandSup	-0,3636

Tabla 6.7 Variables y valores función discriminante

Como podemos observar a lo largo de los resultados obtenidos para los 7 experimentos individuales elegir cuatro variables para realizar el análisis es complicado, exceptuando la variable BandSup que se repite prácticamente en todos. Las otras 3 variables que hemos elegido son una mezcla entre las veces que se repiten y sus coeficientes, de forma que finalmente las elegidas son: BandSup, Band5Sup, Band5Inf y Band7Sup. Hay que destacar que antes de la elección de estas variables se procedió a realizar una prueba con 5 variables para introducir una variable de las más repetidas y que no fuera en el dominio de la frecuencia como es la variable de Kurtosis pero como observaremos en las figuras 6.18, 6.19, 6.20 y 6.21 su comportamiento no aconseja su utilización.



**Figura 6.18** Evolución de Kurtosis para variador AB



**Figura 6.19** Elipse de tolerancia de Kurtosis para variador AB

Las figuras 6.18 y 6.19 muestran cómo evoluciona la variable Kurtosis para las tres frecuencias posibles en el variador AB. La curvatura que aparece en la figura 6.19 da pie a pensar que el comportamiento de esta variable atiende a un patrón lo que desaconseja su uso. Todos los métodos se basan implícitamente en la normal multivariante y esa curvatura no parece encajar con el tipo de realizaciones que tienen las distribuciones normales multivariantes.

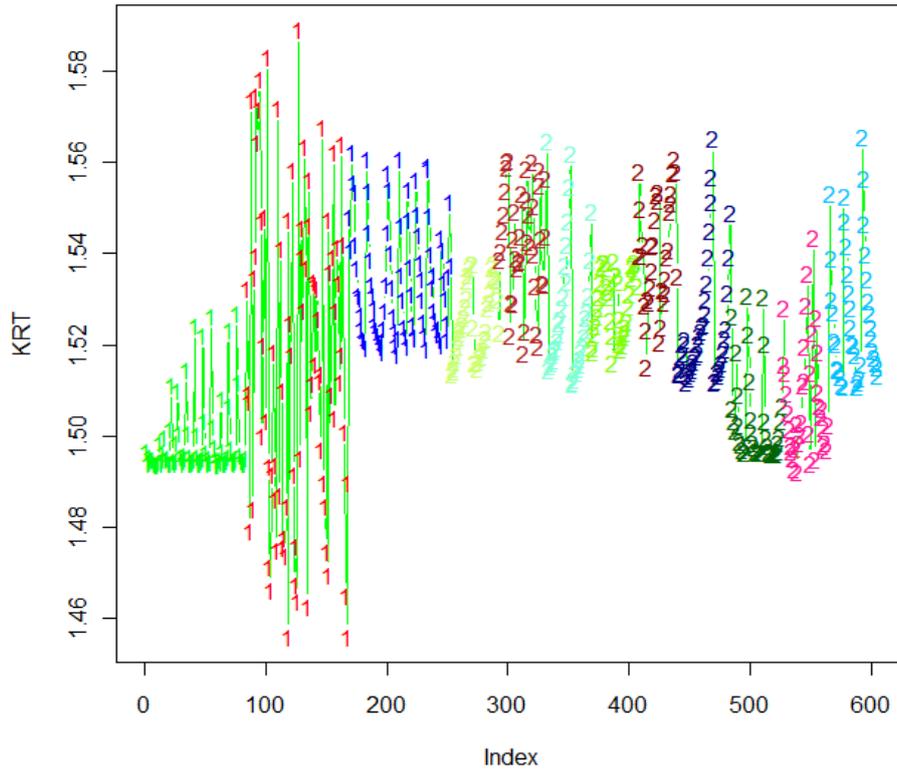


Figura 6.20 Evolución de Kurtosis para variador SM

**Tolerance ellipse (97.5%)**

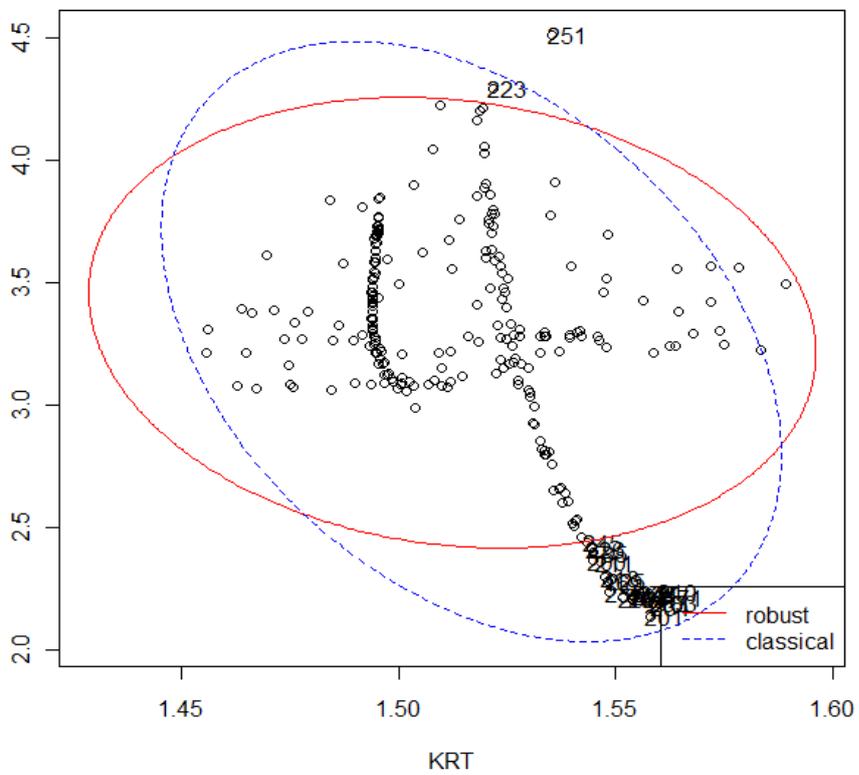


Figura 6.21 Elipse de tolerancia de Kurtosis para variador SM

Podemos apreciar que las figuras 6.20 y 6.21 muestran lo mismo que las anteriores pero para el variador SM, que se repiten los problemas que mencionamos con anterioridad. Existe curvatura y en la figura 6.20 podemos ver como hay un combinación de variador SM y frecuencia cuyo abanico de valores para Kurtosis es demasiado amplio como para poder utilizar esta variable a favor de una clasificación.



***CAPÍTULO 7:***  
***CONTROL***  
***ESTADÍSTICO DE***  
***CALIDAD USANDO EL***  
***MCD***



## **7.1 INTRODUCCIÓN**

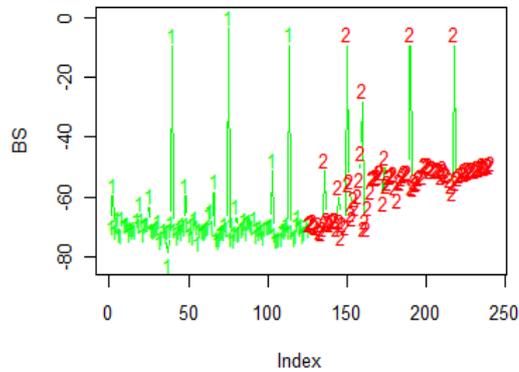
Una vez hemos utilizado los métodos clásicos y el software Statgraphics para llevar a cabo una primera clasificación de los estados de los motores, así como de las variables más significativas en conjunto, que serán las que usemos para nuestro estudio, procedemos a la parte principal del proyecto, que es aplicar las técnicas robustas explicadas en el Capítulo 5 mediante el uso del software R, a cada uno de los 12 posibles experimentos que explicamos con anterioridad. El programa utilizado para llevar a cabo el control de calidad en R se compone de varias partes ya que para llevar a cabo un análisis más profundo no solo es necesario obtener conclusiones de la parte final robusta, sino que gráficas que nos muestren la evolución de las variables a través de los estados o elipses de tolerancia para la detección de outliers pueden ser de gran utilidad para el procedimiento. Dicho programa completo se puede ver para el experimento AB35 en los anexos, siendo brevemente modificado para cada experimento en función del número de outliers, observaciones y distribución de los estados del motor en la muestra.

## **7.2 ANALISIS**

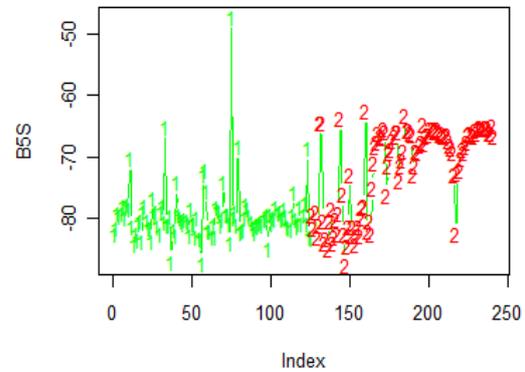
### **7.2.1 Introducción**

Inicialmente lo que realizamos es un análisis simple [GAR14] basado sobre todo en gráficas acerca de cómo evolucionan las variables que seleccionamos anteriormente en el análisis discriminante (ver figura 6.2) para ver si siguen un patrón en función de los niveles que descarte su uso por completo o simplemente para que nos ayude a fortalecer el poder discriminante que descubrimos anteriormente. Se representan en gráficas pareadas enfrentando una a una las cuatro variables para ver si hay grupos bien diferenciados entre los estados, y también se hace un análisis individual de cada una para ver si los valores varían en función del estado en el que se encuentre el motor. Como ejemplo podemos ver las figuras 7.1 y 7.2 que corresponden al análisis AB35. (De ahora en adelante se considerara que las observaciones pertenecientes al estado R1 se encuentran representadas con un 1 y de color verde, mientras que las pertenecientes a los estados R234 se encuentran representadas por un 2 y de color rojo.) La otra parte previa se trata de aplicar el método MCD para la detección de outliers, de forma que gráficamente quedan representadas por elipses de tolerancia (figura 7.3). Una vez que hemos realizado este pequeño análisis previo se llevaran a cabo dos fases pertenecientes al control de calidad, Fase I ver, por ejemplo [GAR14], donde eliminamos observaciones atípicas y realizamos un análisis de las observaciones R1 para establecer un criterio de clasificación, y Fase II ver, por ejemplo [GAR14], que es en la que se lleva a cabo un control de

calidad para cada observación que poseemos para determinar si se encuentran en estado R1 o R234 en función de lo obtenido en la fase I.

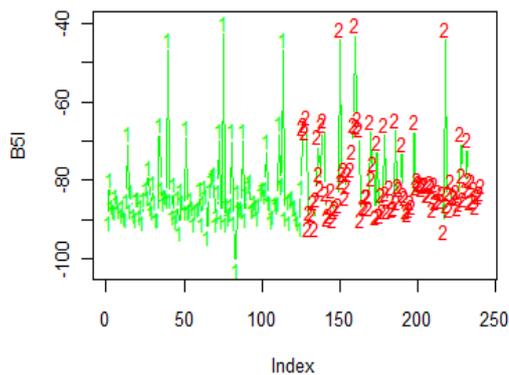


**Figura 7.1a** Evolución de BandSup para AB35

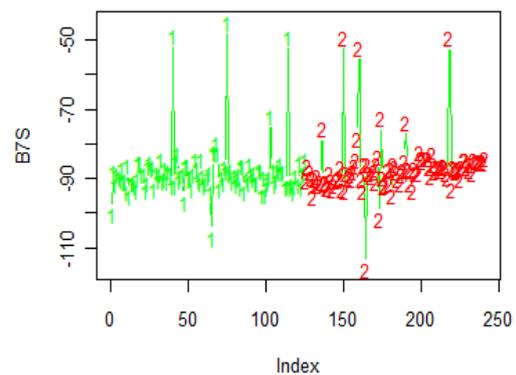


**Figura 7.1b** Evolución de Band5Sup para AB35

Como podemos ver en las gráficas a y b en este caso se aprecia visualmente como las observaciones que se encuentran en el estado R1, para las variables Banda Superior (BS) y Banda 5 Superior (B5S), tienen valores inferiores a la mayoría de las observaciones del estado R234. Posteriormente se averiguó que las observaciones representadas por el 2 que se encuentran a la misma altura que las de R1 se debían mayoritariamente a observaciones cuyo estado era el R2.



**Figura 7.1c** Evolución de BandSup para AB35



**Figura 7.1d** Evolución de Band5Sup para AB35

Sin embargo, en las figuras 7c y 7d se puede apreciar que para este experimento en concreto (AB35) no aportan apenas diferencias visuales las variables Banf5Inf y Band7Sup. No obstante, y como explicamos con anterioridad la elección de variables se hizo teniendo en cuenta los diferentes experimentos individuales, por lo que habrá otros experimentos en los que

estas dos variables sean muy significativas para poder hacer una buena clasificación. A pesar de este hecho no debemos olvidar que se trata de un experimento multivariable por lo que tal y como veremos en la figura 7.2 aunque una variable sola no muestre claras diferencias, si puede hacerlo, representada junto con otra de las elegidas.

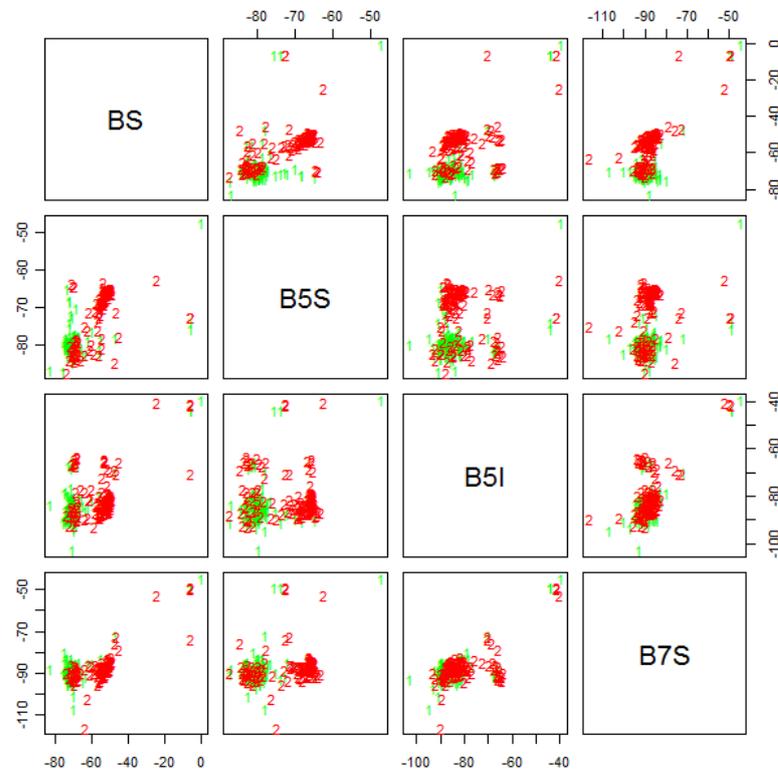


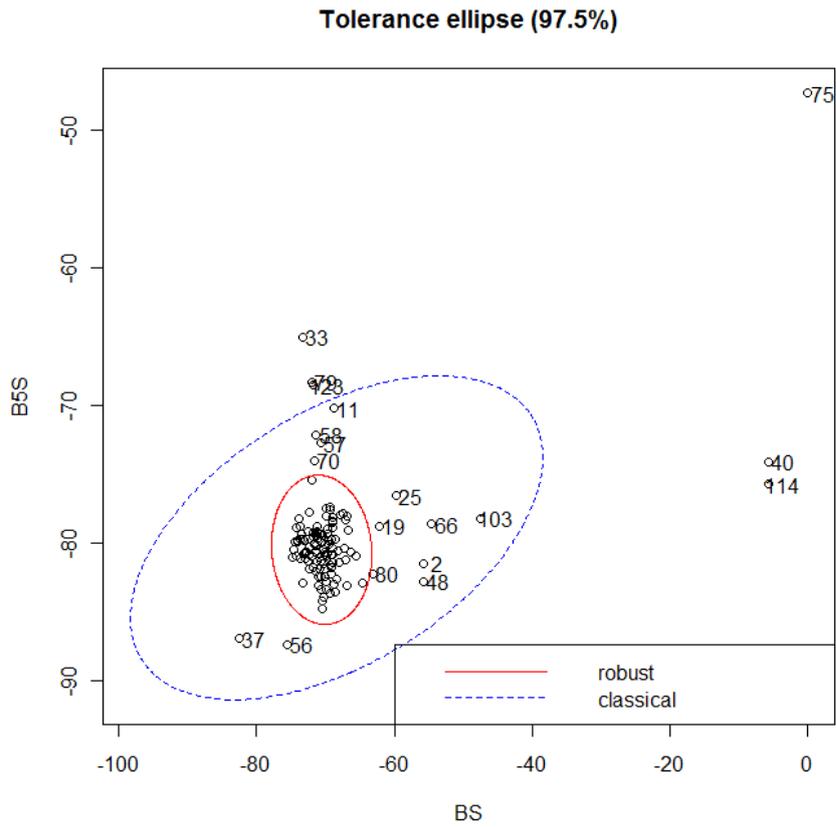
Figura 7.2 Representación Pareada para AB35

En la figura superior 7.2 se muestra una gráfica de pares de variables cuyo cometido principal es mostrar la aparición de pequeños grupos de observaciones que se comportan distinto. Como podemos ver en la misma existen en todas dos “nubes” de observaciones. El interés de estas gráficas radica en ver si estos grupos son de diferentes estados, lo que se aprecia muy bien si enfrentamos las variables B5S y B5I ya que hay un grupo en verde y otro en rojo bien diferenciados, ya que las observaciones en rojo que aparecen sobre la nube verde son mínimas, y como hemos dicho anteriormente, seguramente forman parte del estado R2 que en esa variable tiene un comportamiento parecido al R1.

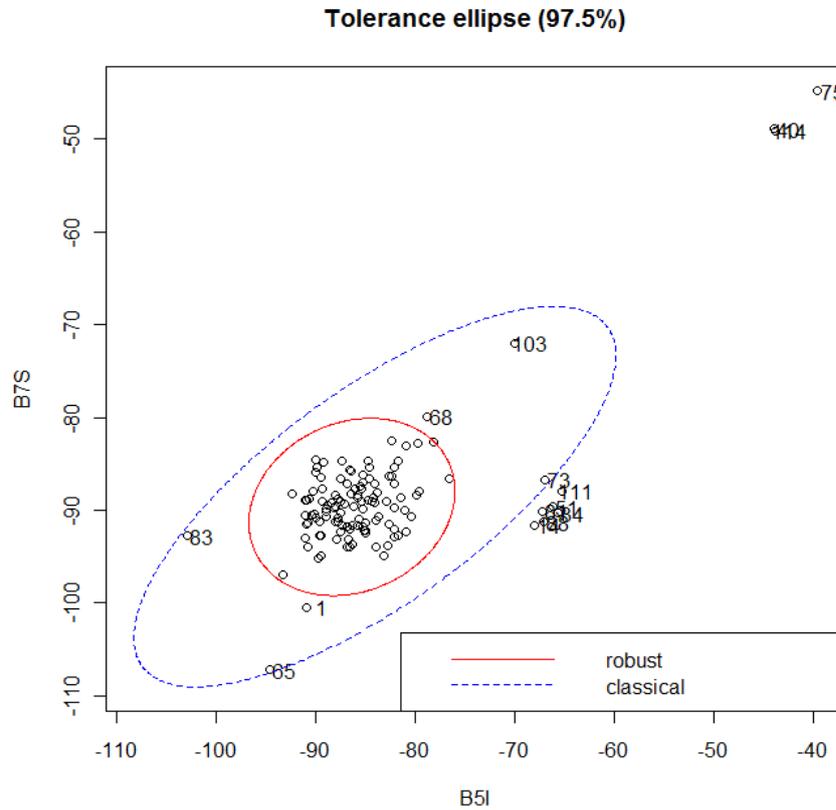
En cuanto a la detección de outliers, fue programada para su detección en esta fase de introducción aunque como su uso se hará en la fase I para descartar observaciones a la hora de establecer los criterios de explicación se explicara a continuación en la fase I.

### 7.2.2 Fase I

En esta fase lo que se quiere ilustrar es que el uso de  $T^2$  standard puede decirnos que un proceso está bajo control cuando en realidad no lo está debido a la existencia de un pequeño número de observaciones atípicas. Por otro lado el uso del  $T^2$  robusto nos muestra que el proceso no está bajo control y que debemos quitar las observaciones atípicas. Para la detección de estas observaciones atípicas usaremos el comando `covMcd` perteneciente al paquete contribuido *RobustBase* del software R. Dicho paquete puede descargarse de la página <http://cran.r-project.org>. Cómo funciona el método MCD ya ha sido explicado con anterioridad en el capítulo 5, pero con las figuras 7.3a y 7.3b que muestran las elipses de tolerancia se ve más clara la idea que quiere representar. Los puntos que se encuentran fuera de la línea discontinua azul, serían las observaciones del estado R1 que se considerarían outlier sin aplicar el método MCD, y las que se encuentran entre la elipse roja y la ya mencionada azul, son aquellos que se añadirían a la lista de outliers definida con métodos clásicos. Se aprecia claramente la gran diferencia de observaciones que serían consideradas como outliers si se usa un método o el otro. La no detección de estos outliers influye en la matriz de dispersiones en la fase I lo que hace que se muestren ocultos y hagan que haya observaciones que se muestren fuera de control en la fase II cuando no deberían de estarlo. Para la realización de estas elipses de tolerancia el propio software utiliza los resultados obtenidos después de aplicar el comando `covMCD`, no obstante si nosotros no hubiéramos utilizado dicho comando, el software tiene por defecto aplicarlo para la realización de estas elipses de tolerancia.

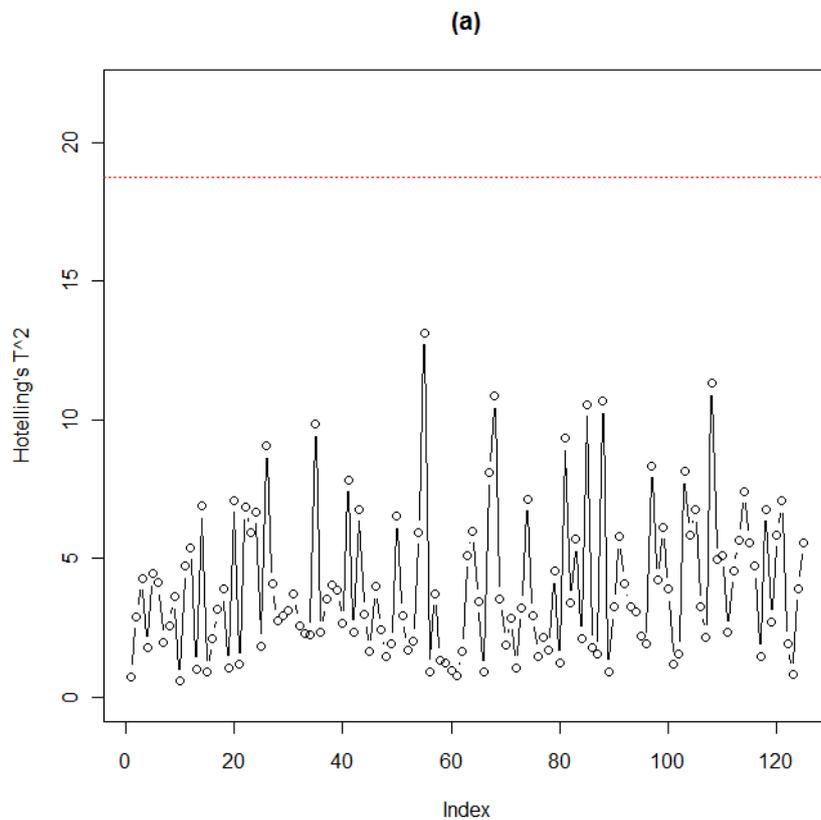


**Figura 7.3a** Elipse de tolerancia con BS y B5S para AB 35

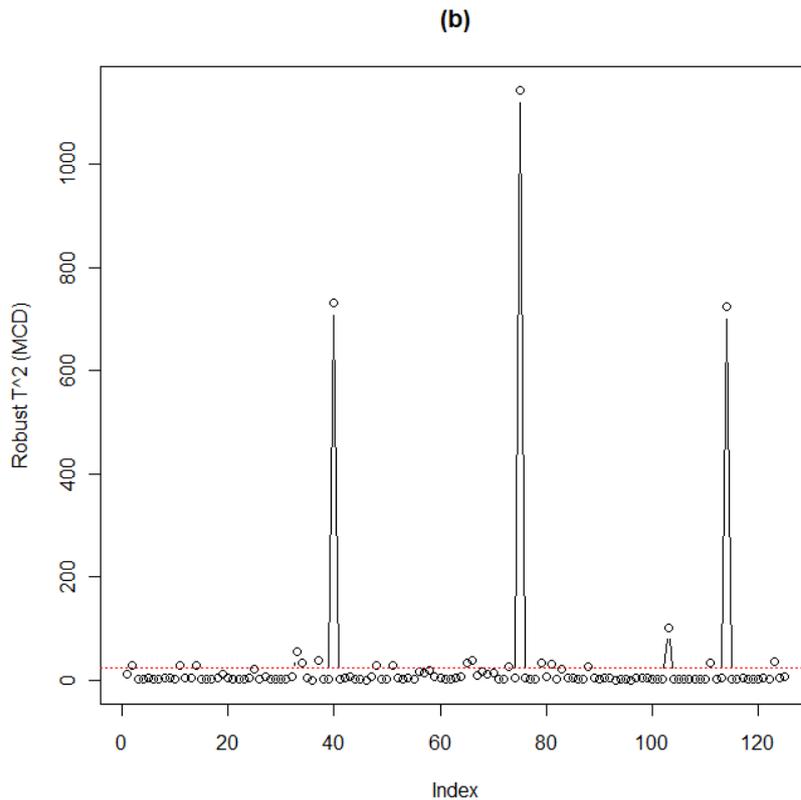


**Figura 7.3b** Elipse de tolerancia con B5I y B7S para AB 35

La importancia de la detección de outliers se antoja clave para una buena clasificación de los motores. El procedimiento que se lleva a cabo es una medición del estadístico  $T^2$  de Hotelling para las observaciones del motor que están clasificadas como estado R1 en la base de datos, tal y como explicamos en el capítulo 5, mediante la distancia de Mahalanobis. Una vez que se tienen los valores de cada observación, se realizan simulaciones de la normal multivariante standard (media igual a cero y varianza igual a la matriz identidad) para obtener por remuestreo los valores críticos por los que una observación es considerada finalmente atípica o no. Consta de dos partes, una en la que se monitoriza la distancia de mahalanobis clásica, es decir, con los estimadores no robustos, media muestral y matriz de covarianzas de las observaciones, y una segunda en la que se utilizan los estimadores robustos que se han obtenido mediante el método MCD aplicado con anterioridad. La diferencia entre aplicar un método y el otro, se puede apreciar de manera intuitiva en las figuras 7.4a y 7.4b. Vemos como se observa con claridad los valores que serían considerados outlier entre un método o no, así como la diferencia entre los valores sobre los que basar la clasificación, por tanto, podemos decir que un método no descubre outliers, lo que hace que después la varianza se presente inflada como se comento anteriormente.



**Figura 7.4a** Valores de  $T^2$  no robustos



**Figura 7.4b** Valores de  $T^2$  robustos

Tal y como vemos en la figura 7.4b, tras la aplicación del MCD a la muestra de motores que se encuentran en estado R1 para cada ensayo, se detectan observaciones que por ser consideradas outliers serán eliminadas. En la tabla 7.1 se muestra que porcentaje del total de las observaciones R1, son eliminadas en cada ensayo para tener una fase fiable (“todas las observaciones dentro de control”) para poder llevar a cabo la fase II de detección de fallos que veremos a continuación. Si miramos detenidamente la tabla 7.1 vemos como el porcentaje de observaciones eliminadas es mayor, generalmente, en los ensayos grupales, llegando casi al 50% de observaciones atípicas en el ensayo F65. Estos datos no hacen más que apoyar la idea que ya hemos comentado varias veces anteriormente, de aplicar este método a los ensayos grupales. Estos valores de observaciones consideradas outliers para los ensayos grupales son debidos a las características del MCD, ya que si algún ensayo individual de los incluidos en el grupal se comporta de una manera muy distinta a los otros, es probable que considere todas sus observaciones como atípicas debido a la distancia que se encuentran de las otras, que son las consideradas por el método como normales. Es por esto por lo que no podemos aplicar este método de una forma más general porque no tiene sentido establecer unos límites de control que son superados por una parte de motores sanos.

ENSAYO	Outliers(%)
AB35	20,97
AB50	20,00
AB65	5,43
AB	22,44
SM35	15,66
SM50	12,94
SM65	12,94
SM	19,76
RED50	28,41
F35	17,31
F50	16,42
F65	42,94

**Tabla 7.1** Porcentaje de Outliers detectados por el MCD en cada ensayo

Valores tan altos como el 42% para el ensayo grupal F65 son debidos a que hay un grupo que contiene aproximadamente la mitad de los datos, que es el que detecta el MCD, y que luego existen grupos muy diferentes de datos. En la fase I de estos experimentos grupales se mezclan observaciones muy heterogéneas debido a que las diversas características de funcionamiento (frecuencia y variador) que tienen los motores pertenecientes a dicho ensayo tienen comportamientos muy diferentes.

### 7.2.3 Fase II

En esta última fase es donde se encuentra la fase propia de clasificación y el control de calidad como tal de todas las observaciones propias del experimento. El procedimiento está basado en el valor del estadístico  $T^2$  correspondiente a todas las observaciones de los diferentes estados, no solo al R1 como anteriormente. Se mostraran los resultados tanto de manera robusta como no robusta con el fin de realizar una comparación que demuestre que es mejor la robusta. Para determinar que observaciones se encuentran dentro de los límites para ser clasificada como R1 es necesario un valor critico que marque el umbral. Dicho umbral límite como explicamos en el apartado 5.2 será construido acorde a las observaciones en estado R1

que han permanecido una vez hemos eliminado las consideradas outliers. El valor límite puede ser construido mediante la siguiente fórmula [LA10]:

$$V.C = \frac{(m-1)^2}{m} \beta_{\alpha; \frac{p}{2}; \frac{m-p-1}{2}}$$

Donde :

- m es el número de observaciones de una submuestra libre de outliers ( $m \leq n$ ).
- p es el número de características monitorizadas
- $\beta_{\alpha; u, v}$  es el percentil  $1-\alpha$  de una distribución  $\beta$  con parámetros u y v.

No obstante para nuestro procedimiento y debido a una mayor simplicidad utilizaremos la siguiente fórmula equivalente:

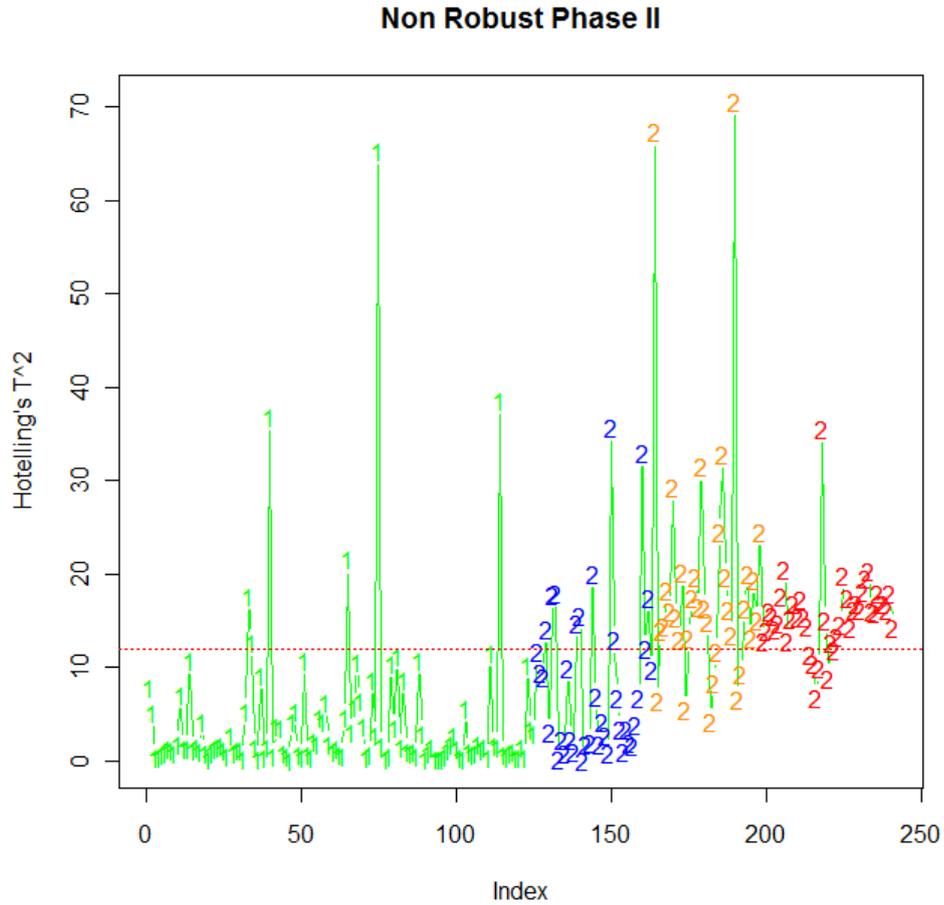
$$V.C = \frac{(p * (m + 1) * (m - 1))}{(m * (m - p))} * \chi_{0.975; p, m-p}^2$$

Donde :

- $\chi_{0.975}^2$  es el percentil  $1-\alpha=0.975$  de una distribución chi-cuadrado con parámetros u y v.

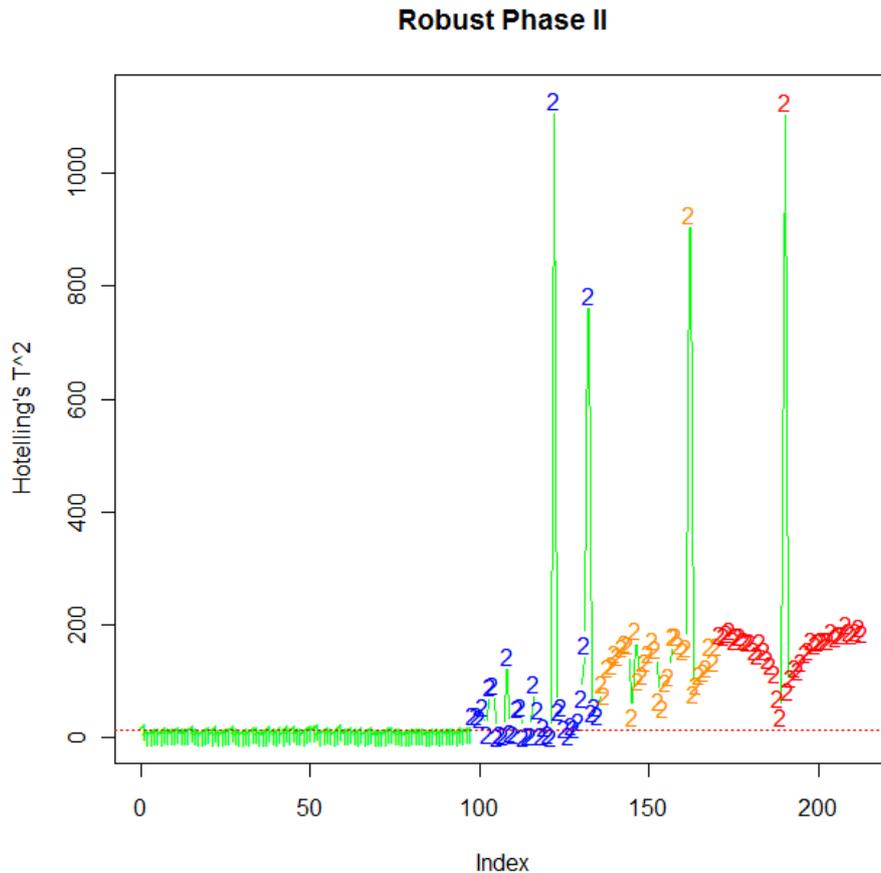
Por tanto, una vez obtenemos el límite mediante la expresión anterior lo que hace que una observación sea considerada como R1 o R234 es si se encuentra su valor de la  $T^2$  de Hotelling asociado, por debajo o por encima respectivamente de dicho valor crítico.

La idea de este procedimiento se puede ver en las gráficas 7.5a y 7.5b. En estas gráficas se representan los gráficos de control para los distintos estados asignándoles un color distinto a cada uno, verde para R1, azul para R2, naranja para R3 y rojo para R4. Estos colores se mantendrán para las diferentes gráficas mostradas hasta el final de este apartado



**Figura 7.5a** Gráfica de control no robusta para AB35

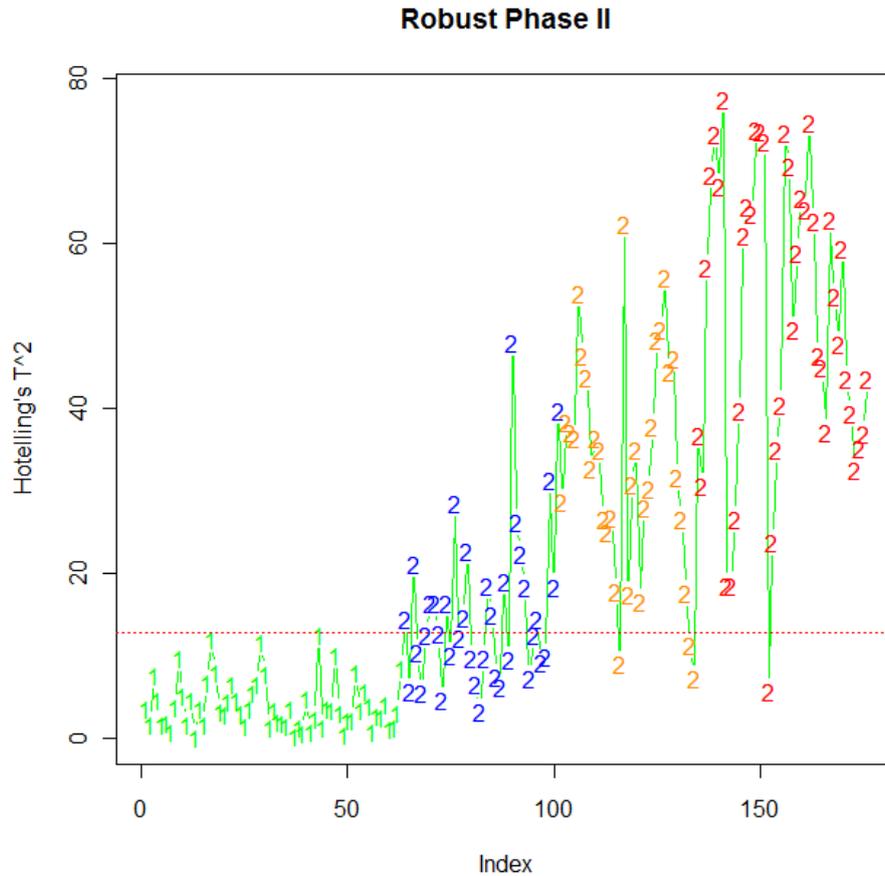
Tal y como hemos comentado con anterioridad en la gráfica 7.5a se puede observar como clasificaría nuestro método las distintas observaciones del experimento, sin eliminar las observaciones atípicas. Se aprecia con claridad como no es un método fiable ya que hay muchas observaciones que serían mal clasificadas. Al tratarse este de un ejemplo para explicar mejor el procedimiento, no se detendrá en contabilizar y hallar la eficiencia del método en este momento, sino que se hará más adelante.



**Figura 7.5b** Gráfica de control robusta para AB35

La gráfica superior 7.5b, muestra el mismo procedimiento que el explicado con anterioridad, pero utilizando el método robusto, es decir, con estimadores obtenidos por la aplicación del algoritmo MCD y tras la eliminación de aquellas observaciones atípicas del estado R1 para establecer el umbral límite con las técnicas standard, es decir, usando la formula V.C anterior. Si comparamos las observaciones representadas como unos en las figuras 7.5a y 7.5b vemos la importancia de la eliminación de outliers llevada a cabo en la fase robusta ya que las observaciones raras han desaparecido. Como hemos mencionado con anterioridad, no nos vamos a detener ahora a analizar la eficiencia de cuales son detectados bien y cuáles no, pero se aprecia con claridad la mejor eficiencia del método robusto frente al no robusto.

Como este apartado se considera el tema central de este TFG, se muestran a continuación las gráficas de control robustas y no robustas de algunos ensayos “individuales” y “grupales”. Para los “individuales” hemos elegido el ensayo RED 50 (figuras 7.6a y 7.6b) y el SM65 (figuras 7.7a y 7.7b), mientras que para los “grupales” se ha optado por el ensayo F50 (figuras 7.8a y 7.8b) y el AB (figuras 7.9<sup>a</sup> y 7.9b). La elección de estos ensayos se ha hecho con el fin de mostrar un rango de posibilidades lo más amplio posible, que nos permita llegar a conclusiones fiables de cara a los objetivos propuestos al inicio.



**Figura 7.6a** Gráfica de control robusta para RED50

El ensayo RED50 se puede considerar como estándar, en el sentido de resultados obtenidos, ya que detecta por encima de la mitad de las observaciones en estado R2 como malas y muestra una clara mejora entre el método no robusto y el no robusto. En cuanto a las observaciones en las que el daño del motor es más importante (R3 y R4) apenas hay 4 observaciones que no serían detectadas como malas.

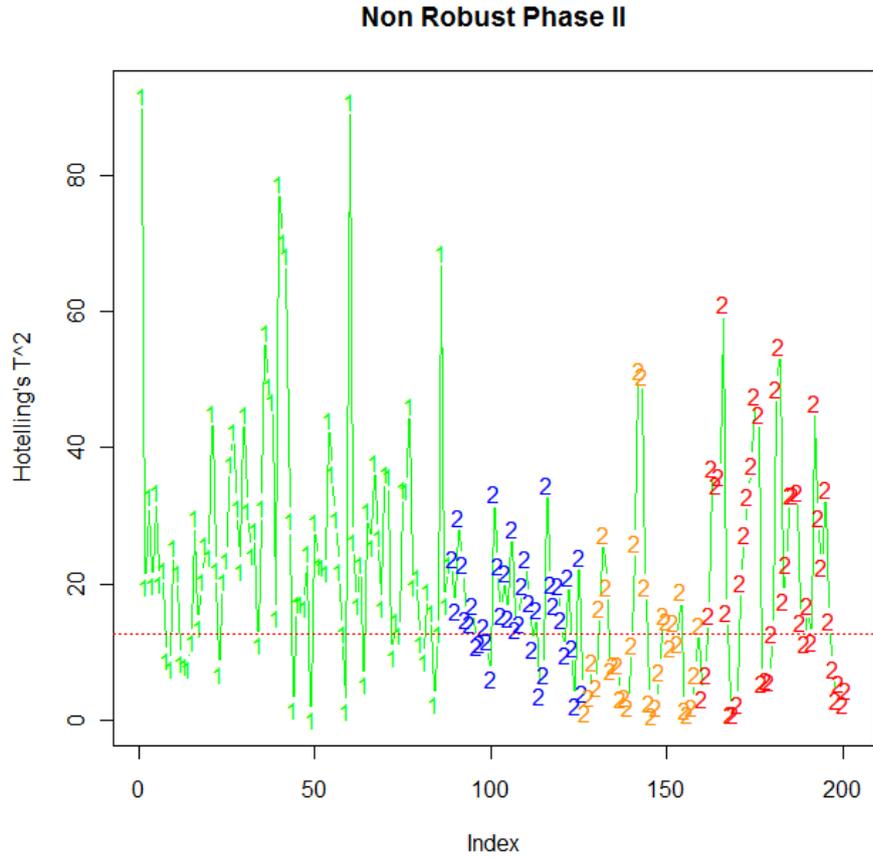


Figura 7.6b Gráfica de control no robusta para RED50

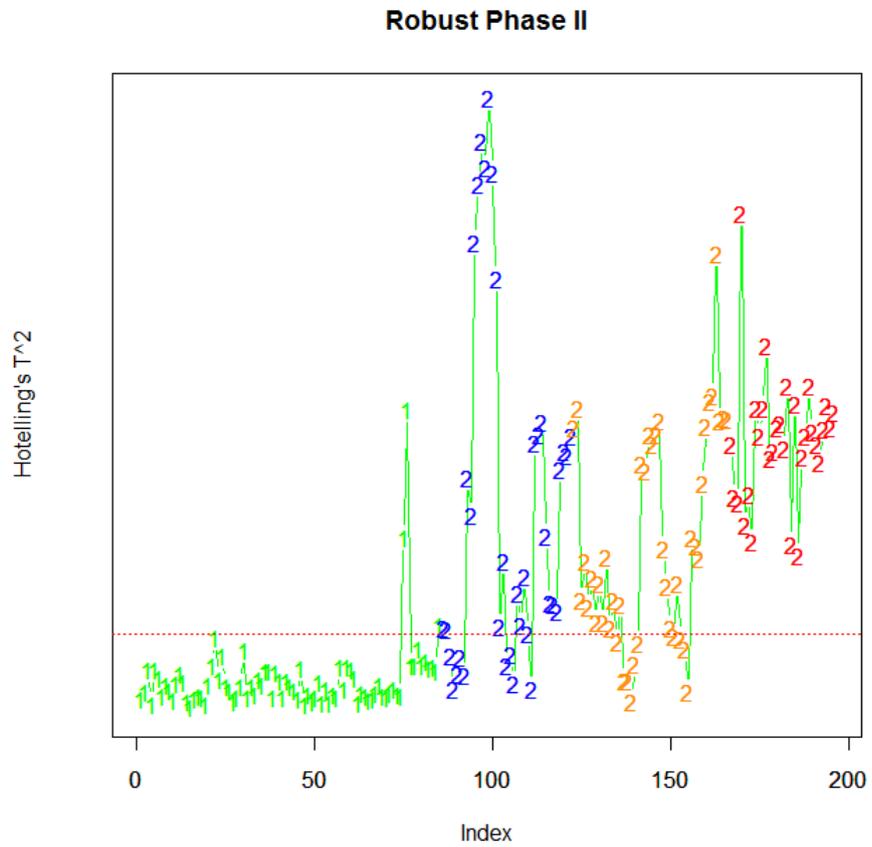
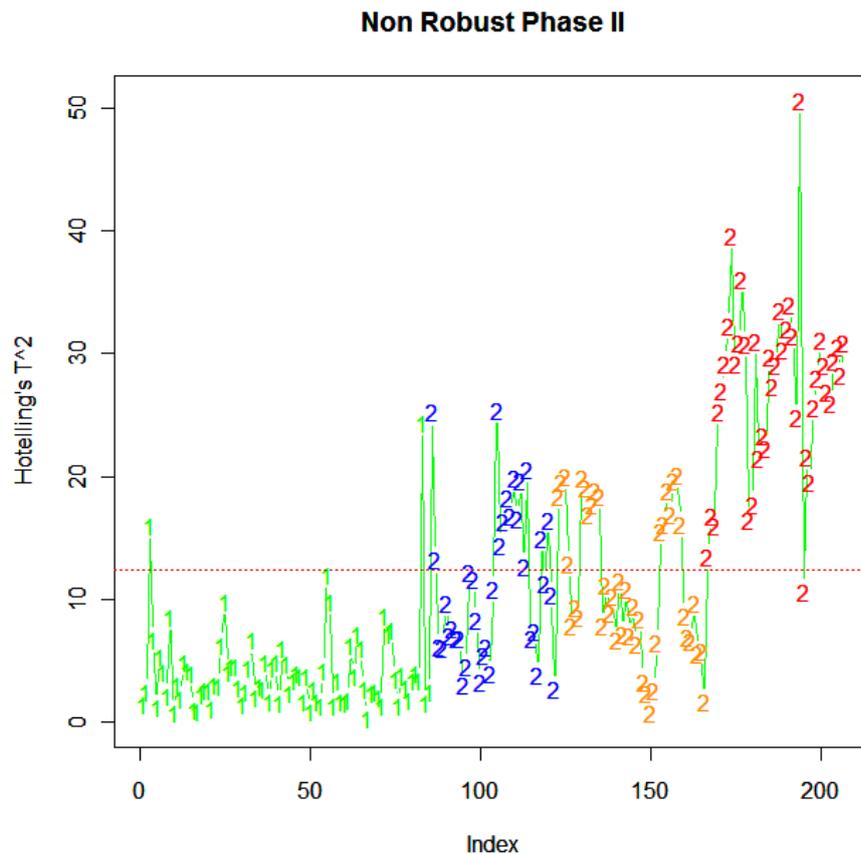


Figura 7.7a Gráfica de control robusta para SM65

El ensayo SM65 se ha considerado importante mostrarlo debido a que es el peor para detectar los R3. Como podemos ver en la figura 7.7a existe un número elevado de “2” ’s de color naranja, observaciones R3, que no son consideradas defectuosas, mientras que en el resto de ensayos no ocurre lo mismo. No obstante para que se pueda ver con claridad el porcentaje de bien detectados en función del estado se tratara el tema en el apartado 7.3. En cuanto a la figura 7.7b que muestra la gráfica de control para el ensayo no robusto, tiene una forma parecida a la robusta pero al igual que ocurre con el resto de ensayos, detecta un porcentaje menor de los motores defectuosos en el caso no robusto.



**Figura 7.7b** Gráfica de control no robusta para SM65

Aunque ya se ha comentado anteriormente la decisión de descartar los ensayos denominados “grupales” debido a los comportamientos tan diversos que tienen sus ensayos integrantes, mostramos en las figuras 7.8a, 7.8b, 7.9a y 7.9b los resultados obtenidos para F50 y AB para una mejor apreciación del problema. Podemos observar como el problema de las observaciones en estado R2 se hace más evidente que en los ensayos “individuales” ya que prácticamente ninguna es detectada como defectuosa.

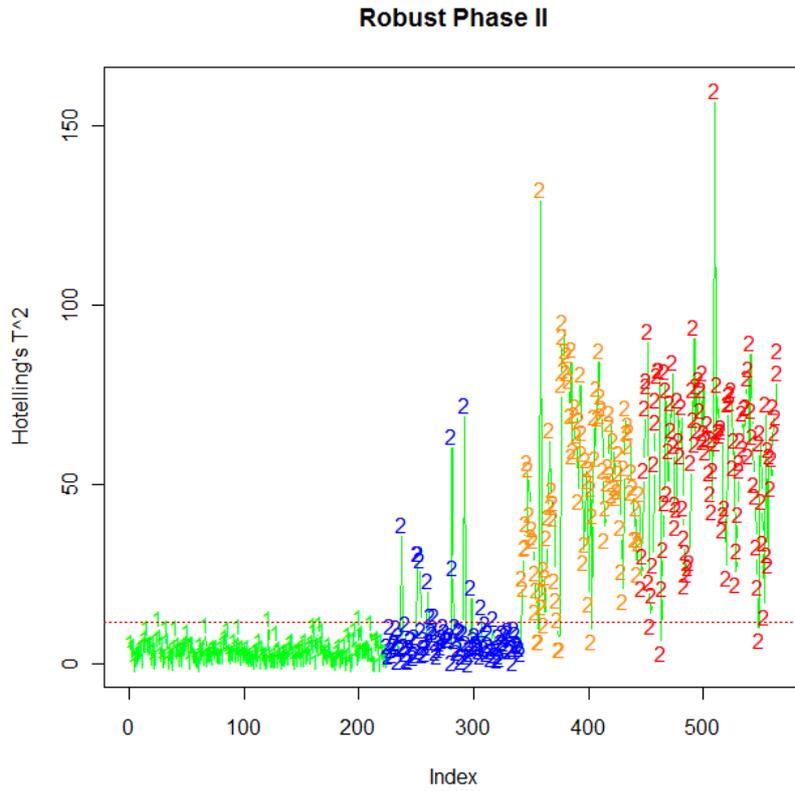


Figura 7.8a Gráfica de control robusta para F50

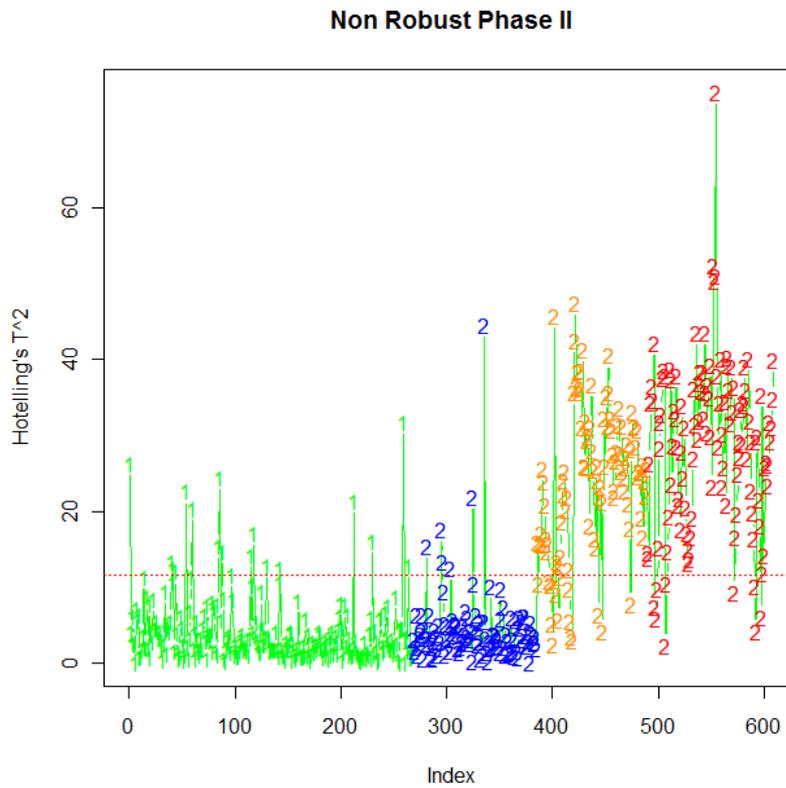


Figura 7.8b Gráfica de control no robusta para F50

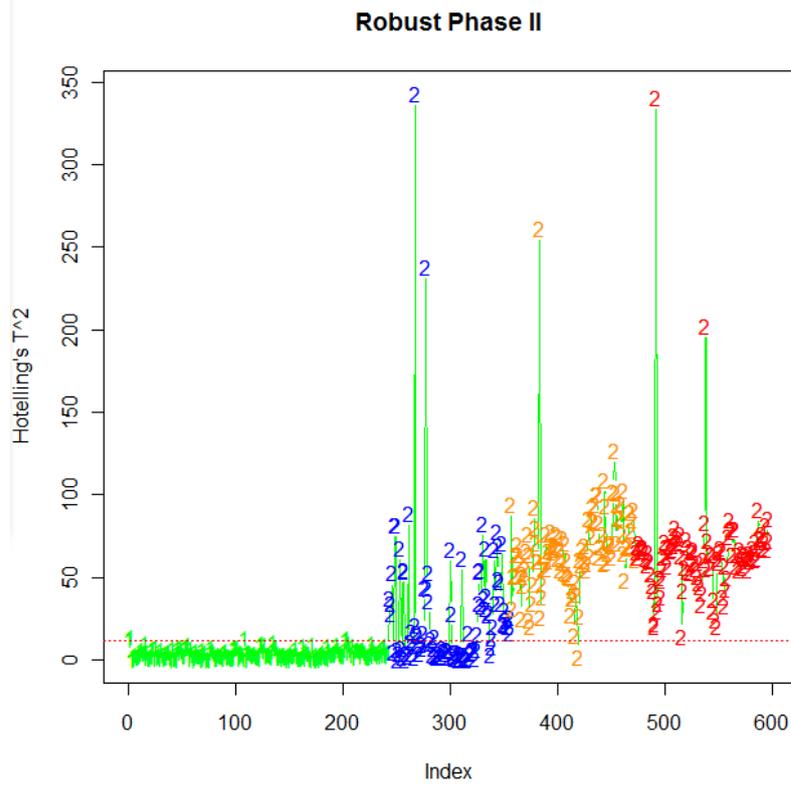


Figura 7.9a Gráfica de control robusta para AB

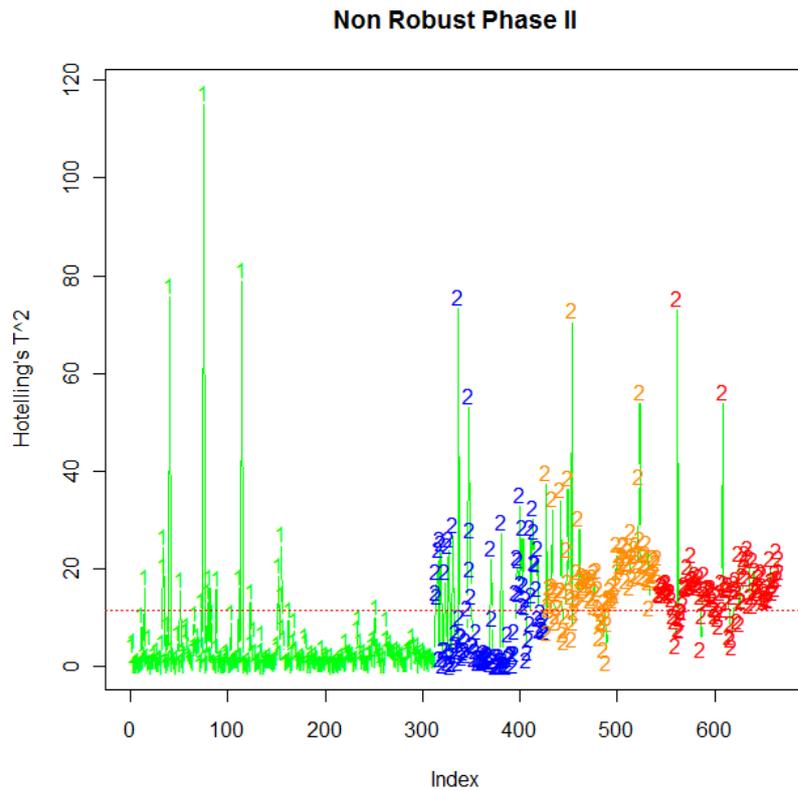


Figura 7.9b Gráfica de control no robusta para AB

### **7.3 EFICIENCIA DEL MÉTODO EN EL CASO ROBUSTO**

Una vez que hemos explicado el funcionamiento del método vamos a proceder a evaluar los resultados obtenidos en términos de eficiencia para los diferentes experimentos que hemos hecho. Lo que buscamos principalmente es detectar que un motor ha dejado de estar en el estado R1 con el fin de obtener una detección temprana del fallo y que no llegue a un estado R3 o R4, y con esa idea obtenemos dos clases de información como podemos ver en la tabla 7.1. Lo primero que vemos es qué porcentaje de motores sanos son bien clasificados por nuestro método en cada uno de los distintos experimentos. En relación a esta información podemos determinar que hemos obtenido un alto grado de eficiencia para nuestro método ya que el porcentaje de acierto en la clasificación es superior al 98% en todos los experimentos, ya se trate de un experimento de los denominados individual o de los de un grupo con una característica común (recordemos que puede ser la frecuencia de trabajo o la clase de variador). Sin embargo, de la segunda información obtenida no podemos decir que sea eficiente en todos los ensayos. En la tabla donde se ven los porcentajes de que observaciones han sido bien detectadas y cuáles no para los estados R1-R234 puede parecer que los experimentos de grupo tiene una eficiencia similar a los individuales, pero existen 3 principales problemas que hacen que debamos descartar el uso del método para el control de calidad de los experimentos grupales:

-1 El número de outliers detectados en las observaciones R1 para los ensayos AB, SM, F35, F50 y F65 como se puede ver en la tabla 7.1. Es muy elevado en porcentaje del total por lo que disponemos de un número de observaciones un poco bajo para poder fiarnos de la clasificación.

-2 Mientras que los ensayos individuales tienen problemas en la clasificación debido a las observaciones que se encuentran en estado R2, las observaciones grupales tienen problemas para R3 y R4 también.

-3 Como hemos explicado en el Capítulo 5, el fundamento del método MCD, se basa en recoger un porcentaje de observaciones muy concentradas, y a partir de ellas, decidir que observaciones son outliers o no. El problema de un ensayo grupal es que los R1 de los ensayos individuales que forman parte del grupal tienen comportamientos muy distintos, puede darse el caso de que sean considerados como atípicos y no ser tenidos en cuenta para a posteriori construir los estimadores robustos que se utilizaran cuando apliquemos la distancia de mahalanobis.

ENSAYO		R2-R3-R4	ENSAYO		R2-R3-R4
AB35	MAL	11,97	SM65	MAL	22,13
	BIEN	88,03		BIEN	77,87
AB50	MAL	23,08	SM	MAL	36,36
	BIEN	76,92		BIEN	63,64
AB65	MAL	1,65	RED50	MAL	20,35
	BIEN	98,35		BIEN	79,65
AB	MAL	16,34	F35	MAL	22,36
	BIEN	83,66		BIEN	77,64
SM35	MAL	23,97	F50	MAL	32,84
	BIEN	76,03		BIEN	67,16
SM50	MAL	30,91	F65	MAL	5,74
	BIEN	69,09		BIEN	94,26

Tabla7.1 Porcentaje de observaciones bien o mal clasificadas

## 7.4 COMPARATIVA FRENTE A MÉTODOS CLÁSICOS

### 7.4.1 introducción

Como dijimos al comienzo de este documento, uno de los objetivos de este TFG era probar la mayor eficacia de los métodos robustos frente a los no robustos a la hora de clasificar una observación o dato de una forma u otra con acierto. Debido al carácter de los datos en los que nos basamos y con la intención no solo de probar la mayor eficacia de los robustos sino de mostrar cómo serían los resultados si quisiéramos realizar otra clasificación que no fuera la de separar los estados R1-R234. No obstante y para no basarnos solo en lo obtenido por el software R, mostraremos los resultados obtenidos con otro software como statgraphics cuyos resultados entrarían dentro de los métodos no robustos y sin usar el algoritmo MCD.

### 7.4.2 Detección R234

Como esperábamos, los resultados obtenidos por métodos robustos superan ampliamente en eficiencia a los resultados obtenidos con el mismo software mediante un algoritmo no robusto o de los obtenidos por otro software cuya función de clasificación viene programada de serie.

Analizando los ensayos uno por uno, vemos como los resultados de clasificación entre el método robusto y el no robusto (figura 7.9a) son peores los obtenidos por el no robusto, variando entre una mejora que a veces es mínima, entorno al 5% de mejor detectados, a casos en los que nos vamos a una mejora de más del 50% de casos detectados correctamente por parte del robusto como es el caso del ensayo RED 50. La gran diferencia entre usar un

método u otro se aprecia a la hora de distinguir la clasificación R234 ya que para la clasificación del estado R1 apenas hay diferencias significativas. Puede llamarnos la atención que si que hay un par de ensayos en los que se clasifica mejor en el no robusto que en el robusto, si bien la diferencia es bastante pequeña en comparación con lo indicado anteriormente.

Otra cosa que podemos apreciar es que no es posible obtener ningún patrón en cuanto a la mejor o peor detección para el caso no robusto ya que por ejemplo para el variador AB el no robusto lo detecta con mejor eficiencia en relación a como detecta los anteriores a medida que aumenta la frecuencia, pero luego con el variador SM a medida que aumenta la frecuencia pierde eficiencia, y sin embargo si podemos interpretar que existe un patrón para el robusto, ya que ocurre para los variadores AB y SM que para la frecuencia intermedia se clasifica con peor eficiencia que para la primera, para volver a aumentar la eficiencia a ser la mayor, con la frecuencia más elevada.

Como excepción que confirma la regla aparece el ensayo AB65 en el que por muy poca diferencia el método robusto parece clasificar peor las observaciones que el no robusto.

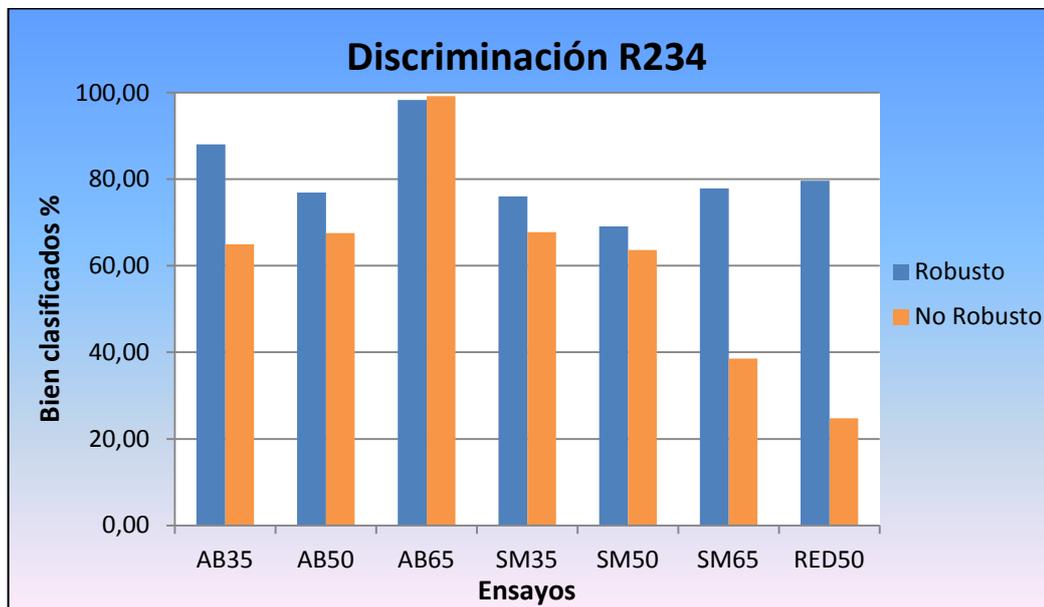


Figura 7.9a Comparativa para el estado R234 entre Robustos y No Robustos

Para finalizar la comparación entre los resultados robustos y no robustos para esta discriminación conviene fijarse en la figura 7.10 y la tabla 7.2 homóloga. En ellas se puede ver una comparativa de como decae la eficiencia entre detectar correctamente las observaciones que se encuentran en estado R234 si utilizamos el método no robusto en lugar del robusto. La diferencia entre un método y otro es bastante notable ya que pasamos de un 70% de motores defectuosos bien detectados en el ensayo de menor eficiencia por el método de robusto, a un peor ensayo de un 25% de motores defectuosos bien detectados si usamos el método no robusto. Estos resultados son importantes de cara a nuestro análisis ya que nosotros buscamos detectar cuanto antes y correctamente un posible fallo en un motor y si queremos hacerlo de una forma eficiente y que en algún momento pueda llegar a tener su aplicación real, no podemos fiarnos de un procedimiento como el no robusto a tenor de estos resultados.

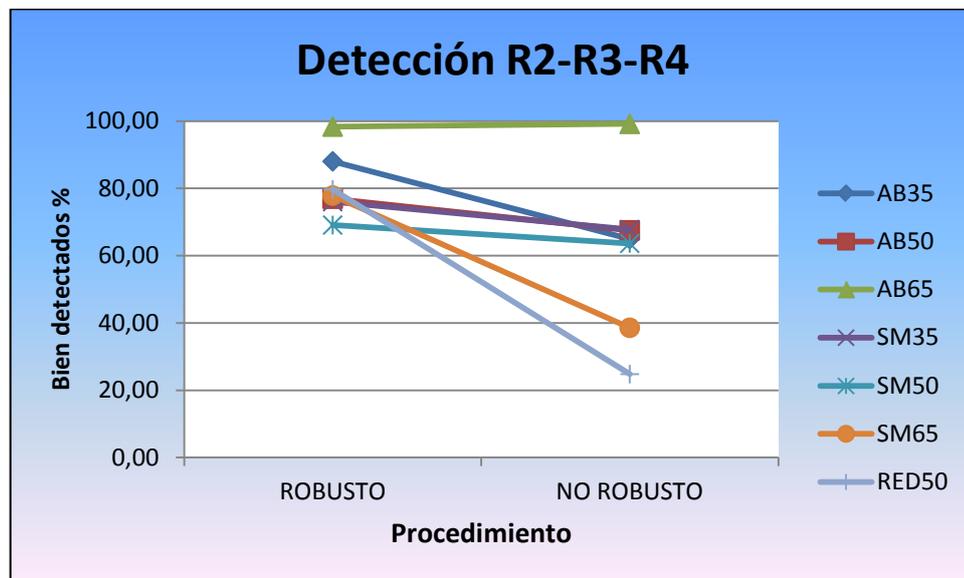


Figura 7.10 Comparativa entre procedimientos para la discriminación R234

ROBUSTO			NO ROBUSTO		
ENSAYO		R2-R3-R4	ENSAYO		R2-R3-R4
AB35	MAL	11,97	AB35	MAL	35,04
	BIEN	88,03		BIEN	64,96
AB50	MAL	23,08	AB50	MAL	32,48
	BIEN	76,92		BIEN	67,52
AB65	MAL	1,65	AB65	MAL	0,83
	BIEN	98,35		BIEN	99,17
AB	MAL	16,34	AB	MAL	33,24
	BIEN	83,66		BIEN	66,76
SM35	MAL	23,97	SM35	MAL	32,23
	BIEN	76,03		BIEN	67,77
SM50	MAL	30,91	SM50	MAL	36,36
	BIEN	69,09		BIEN	63,64
SM65	MAL	22,13	SM65	MAL	61,45
	BIEN	77,87		BIEN	38,55
SM	MAL	36,36	SM	MAL	50,14
	BIEN	63,64		BIEN	49,86
RED50	MAL	20,35	RED50	MAL	75,22
	BIEN	79,65		BIEN	24,78
F35	MAL	22,36	F35	MAL	48,95
	BIEN	77,64		BIEN	51,05
F50	MAL	32,84	F50	MAL	35,48
	BIEN	67,16		BIEN	64,52
F65	MAL	5,74	F65	MAL	23,46
	BIEN	94,26		BIEN	76,54

Tabla 7.2 Porcentaje de observaciones bien o mal detectadas para R234

### 7.4.3 Detección R2-R3-R4

Analizar de manera resumida los resultados obtenidos si hubiéramos deseado hacer esta discriminación entre nuestros motores sin ser el objeto principal de estudio atiende a dos motivos principales. El más importante es el de mostrar cual es el principal problema que tiene nuestro algoritmo ya sea robusto o no, de obtener mejores datos en eficiencia. Este motivo no es otro de la gran confusión que supone diferenciar los motores que se encuentran en estado R2 como R1 o R234. Como podemos ver en las figuras 7.11a, 7.11b y 7.11c, el porcentaje de detectados correctos para estado R2 es ampliamente inferior que los obtenidos para los estados R3 y R4. Exceptuando en dos ensayos, en el resto más de la mitad de los motores cuyo estado es R2 son clasificados de manera equivocada, sin embargo, los motores en estados R3 y R4 son clasificados correctamente en prácticamente la totalidad de los casos en los ensayos. Esto nos deja una clara conclusión a sacar para nuestro estudio y no es otra cosa que los motores que se encuentran en estado R1 o estado R2 se comportan de manera muy

parecida, independientemente de la frecuencia a la que trabajen o al variador al que se encuentren conectados.

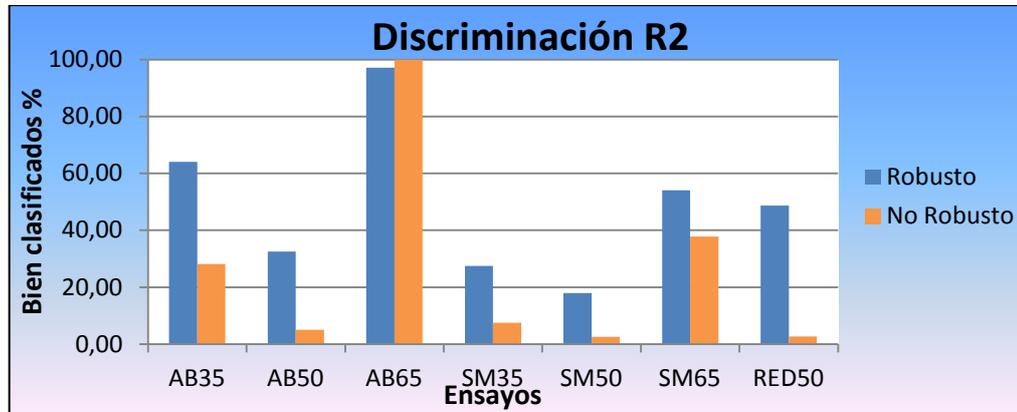


Figura 7.11a Comparativa para el estado R2 entre Robustos y No Robustos

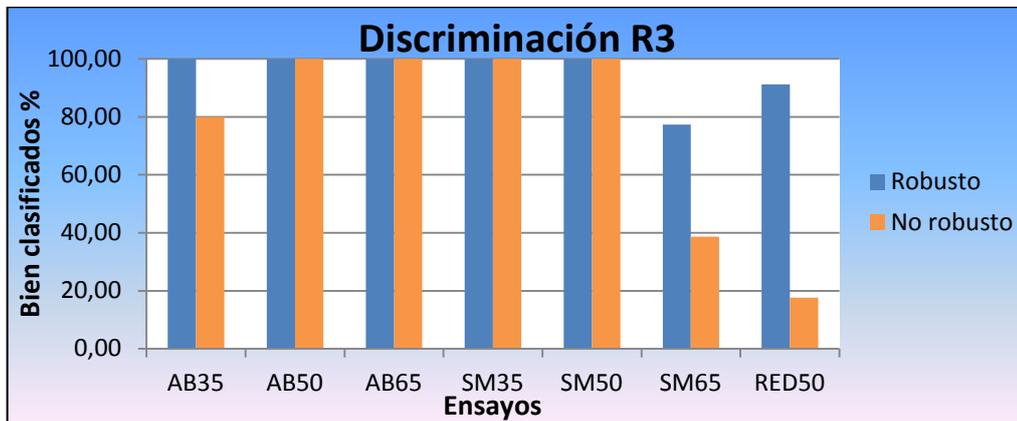


Figura 7.11b Comparativa para el estado R3 entre Robustos y No Robustos

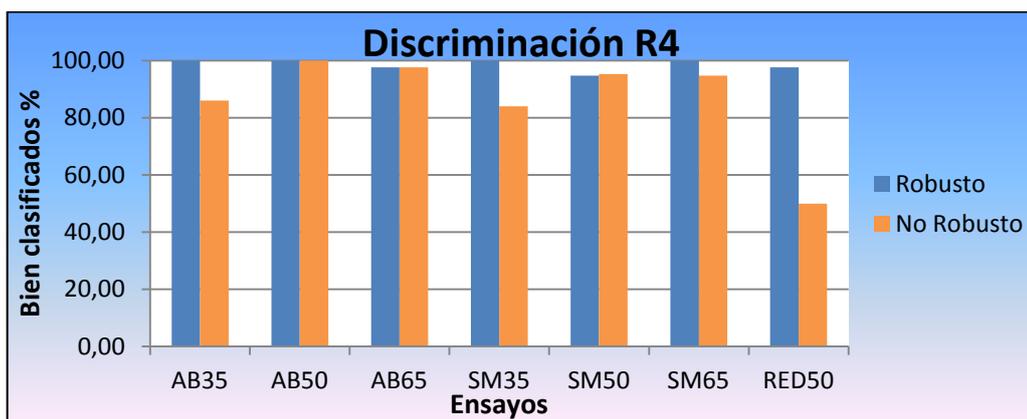


Figura 7.11c Comparativa para el estado R4 entre Robustos y No Robustos

En relación a la mejor detección de los métodos robustos y no robustos podemos ver los mismos resultados (figuras 7.12a, 7.12b y tabla 7.3) que

comentábamos en el apartado anterior y es la clara mejora de los primeros frente a los segundos. Observando las figuras y la tabla vemos dos cosas que conviene resaltar, cuanto más dañado se encuentra el motor menos diferencia existe entre usar un método u otro, y la otra es el claro patrón que sufren todos los ensayos menos el AB65, de tener un alto porcentaje para R1, sufrir un descenso considerable en R2, para ir aumentando progresivamente en R3 y R4 para alcanzar prácticamente los mismos niveles que en R1.

De cara a los objetivos planteados al inicio, no solo hemos demostrado en este apartado la mejora de aplicar métodos robustos frente a los no robustos, sino que por norma general el método robusto que nosotros usamos detecta los motores con fallos antes de que lo haga el no robusto, y no solo eso, sino que con nuestro método robusto rara vez un motor llega sin ser detectado como defectuoso el estado R4, rotura total de la barra, mientras que si observamos los resultados obtenidos por el método no robusto, hay ensayos en los que motores en estado R4 siguen siendo clasificados como válidos. De cara a una situación real, es una gran ventaja tanto económica como de ahorro de tiempo la que supone usar el método robusto, ya que no supone el mismo dinero ni se tarda lo mismo sustituir un motor roto por completo que en arreglar un pequeño desperfecto.

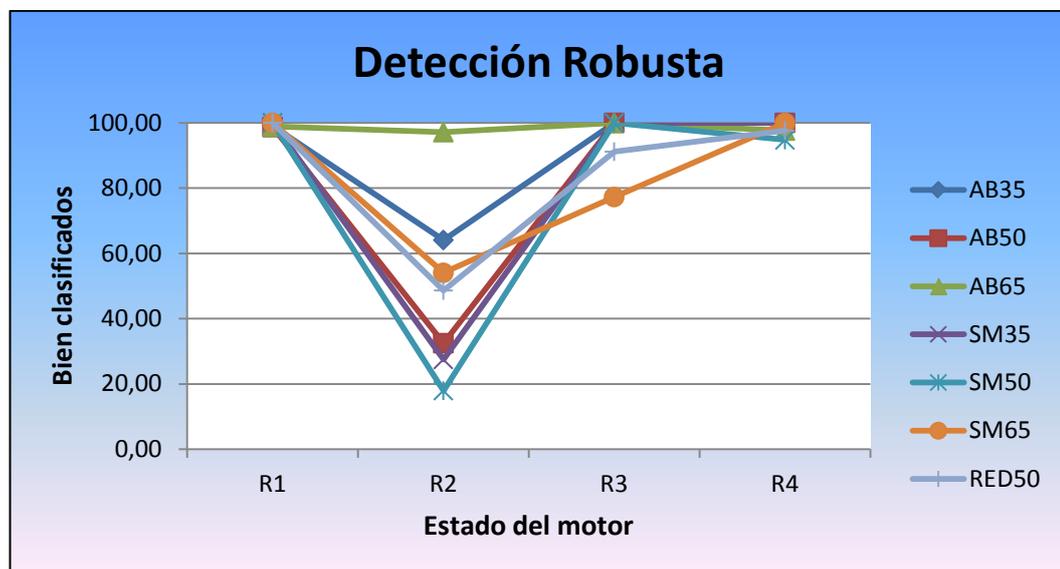


Figura 7.12a) Comparativa del método robusto para la discriminación R1-R2-R3-R4

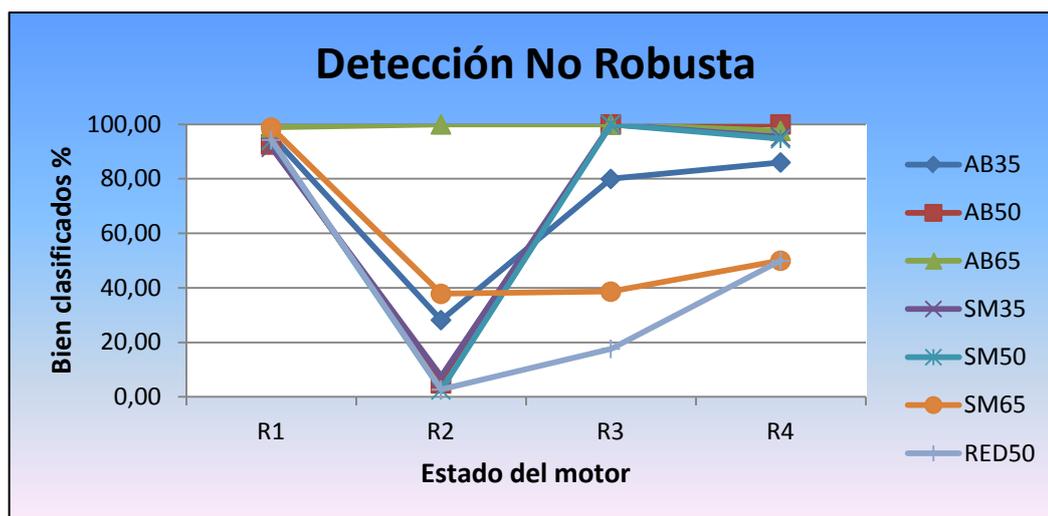


Figura 7.12b) Comparativa del método no robusto para la discriminación R1-R2-R3-R4

ROBUSTO					NO ROBUSTO				
ENSAYO		R2	R3	R4	ENSAYO		R2	R3	R4
AB35	MAL	35,90	0,00	0,00	AB35	MAL	71,79	20,00	13,95
	BIEN	64,10	100,00	100,00		BIEN	28,21	80,00	86,05
AB50	MAL	67,50	0,00	0,00	AB50	MAL	95,00	0,00	0,00
	BIEN	32,50	100,00	100,00		BIEN	5,00	100,00	100,00
AB65	MAL	2,86	0,00	2,38	AB65	MAL	0,00	0,00	2,38
	BIEN	97,14	100,00	97,62		BIEN	100,00	100,00	97,62
AB	MAL	50,00	0,86	0,00	AB	MAL	68,42	17,24	16,00
	BIEN	50,00	99,14	100,00		BIEN	31,58	82,76	84,00
SM35	MAL	72,50	0,00	0,00	SM35	MAL	92,50	0,00	4,65
	BIEN	27,50	100,00	100,00		BIEN	7,50	100,00	95,35
SM50	MAL	82,05	0,00	5,26	SM50	MAL	97,44	0,00	5,26
	BIEN	17,95	100,00	94,74		BIEN	2,56	100,00	94,74
SM65	MAL	45,95	22,73	0,00	SM65	MAL	62,16	61,36	50,00
	BIEN	54,05	77,27	100,00		BIEN	37,84	38,64	50,00
SM	MAL	88,89	11,30	5,74	SM	MAL	91,38	40,87	19,67
	BIEN	11,11	88,70	94,26		BIEN	8,62	59,13	80,33
RED50	MAL	51,35	8,82	2,38	RED50	MAL	97,30	82,35	50,00
	BIEN	48,65	91,18	97,62		BIEN	2,70	17,65	50,00
F35	MAL	66,67	1,37	0,00	F35	MAL	84,62	39,73	24,42
	BIEN	33,33	98,63	100,00		BIEN	15,38	60,27	75,58
F50	MAL	88,89	5,77	1,67	F50	MAL	88,89	8,65	6,67
	BIEN	11,11	94,23	98,33		BIEN	11,11	91,35	93,33
F65	MAL	18,06	0,00	1,19	F65	MAL	54,17	18,18	2,41
	BIEN	81,94	100,00	98,81		BIEN	45,83	81,82	97,59

Tabla 7.3 Porcentaje de observaciones bien o mal detectadas para los estados R2-R3-R4

***CAPÍTULO 8:***  
***ESTUDIO ECONÓMICO***



## 8.1 INTRODUCCIÓN

Para llevar a cabo un breve estudio económico del proyecto que estamos tratando, es necesario incluir las acciones requeridas para la planificación, supervisión y control de los costes durante el ciclo de vida del mismo. Hay que resaltar que en este caso el estudio ya ha sido realizado y vamos a llevar a cabo un análisis posterior al mismo. No obstante para no entrar en cuestiones complejas e innecesarias, se va a emplear la metodología clásica sobre gestión de proyectos [PMBOK5] como guión para establecer el coste final del presente trabajo, ya que este es el objetivo principal del capítulo. Para llevar a cabo la gestión de costes de un proyecto existen tres dimensiones importantes que debemos tener en cuenta, que son: alcance, tiempo y tipo de proyecto.

## 8.2 ALCANCE DEL PROYECTO

El alcance de un proyecto incluye todo el trabajo que hay que realizar para que el proyecto incluya todo lo requerido, es decir, que cumpla sus objetivos. Definir de manera clara el alcance de un proyecto favorece la realización exitosa del mismo así como nos ayuda a definir los límites del mismo, es decir, que se incluye y que no. Es imprescindible que todas las partes que toman parte en el proyecto conozcan el alcance del mismo ya que puede dar lugar a malos entendidos, así como si se produjese algún cambio en el mismo ha de ser comunicado a todos los interesados.

En nuestro caso el alcance del proyecto viene un poco delimitado por el departamento de estadística e investigación operativa, no obstante, y de cara a una mejor valoración para el punto de vista económico, se incluye un breve resumen de cada uno de los pasos propios de la metodología que estamos usando y de cómo se definieron para nuestro proyecto.

**1. Planificar la gestión del alcance:** Se trata de elaborar un plan en el que se explique cómo se va a definir, validar y controlar el alcance del proyecto.

Para este proyecto debido a su carácter el control y validación se iba produciendo según el alumno iba haciendo avances y el director y subdirector del proyecto los iban dando como correctos.

**2. Recopilar requisitos:** La finalidad de este apartado es la de determinar y documentar las necesidades de las personas implicadas en el proyecto de forma que se puedan cumplir los objetivos.

Al ser un proyecto académico la recopilación de requisitos no se considero la idea de realizarlo ya que se esperaba que las necesidades fueran evolucionando a medida que se avanzaba en el mismo.

**3. Definir el alcance:** Lo más importante de este apartado es definir los límites del mismo y una descripción detallada. Los límites para este proyecto quedaron determinados una vez se terminó con el análisis de la parte experimental y de las conclusiones que se obtenían de la misma.

**4. Crear la Estructura de Desglose del Trabajo (EDT):** Es el proceso de dividir el trabajo en partes más pequeñas y de mejor manejo. La EDT es una descomposición jerárquica del alcance total del proyecto a realizar para cumplir con los objetivos y generar los entregables. Lo que se consigue al descomponer el alcance en partes más pequeñas es una mejor estimación del coste así como la asignación de un responsable directo del mismo. Cada paquete tendrá asociadas una serie de actividades concretas a realizar de forma que quede todo incluido, y que no se realiza trabajo extra innecesario. Para nosotros esta parte fue fácil de realizar ya que el alumno se encargaba de la realización del trabajo, y el director y subdirector de la supervisión, validación y asesoramiento.

**5. Validar el alcance:** Formalizar la aceptación de los entregables que se hayan completado. A medida de que el alumno iba finalizando cada una de las partes, éstas iban siendo validadas por los directores.

**6. Controlar el alcance:** Se trata de llevar un control del estado del proyecto y de llevar a cabo la gestión de las posibles modificaciones que ocurran en el mismo. Esta labor fue llevada a cabo por parte de los directores durante todo el proyecto.

### **8.3 GESTIÓN DEL TIEMPO**

Gestionar el tiempo en un proyecto incluye todos los procesos requeridos para llevar a cabo la ejecución del proyecto en el plazo correspondiente. En este caso el plazo de ejecución de 30 semanas. Veamos los pasos a seguir para lograr una correcta gestión del tiempo siguiendo el mismo procedimiento que para el alcance, y una vez hayamos definido los pasos necesarios para el mismo, desarrollaremos de manera más detallada para nuestro proyecto.

**1. Planificar la gestión del cronograma:** En este primer paso se establecen las políticas, los procedimientos y la documentación para planificar, desarrollar, gestionar, ejecutar y controlar el cronograma. Para este proyecto no se trabajó de forma muy exhaustiva esta parte ya que al ser un proyecto académico la mayoría de documentos y procedimientos se encuentran estipulados por norma

**2. Definir las actividades:** identificar y documentar las acciones específicas que se deben realizar para lograr el éxito del proyecto. En la primera reunión se establecieron las diferentes actividades que eran necesarias para la consecución exitosa del proyecto.

**3. Secuenciar las actividades:** se trata de identificar las relaciones que hay entre las distintas actividades que forman parte del proyecto. Es un momento clave puesto que una buena secuenciación permite ahorrar tiempo y recursos, lo que se traduce en ahorro de costes.

**4. Estimar los recursos de las actividades:** estimar el tipo y la cantidad de materiales, recursos, equipos o suministros.

**5. Estimar la duración de las actividades:** Dar a cada actividad una cantidad de periodos de trabajo necesaria para que finalicen a tiempo.

**6. Desarrollar el cronograma:** analizar las secuencias, duraciones, requisitos de recursos y restricciones para crear el modelo de programación del proyecto.

**7. Controlar el cronograma:** Llevar un control sobre en qué estado se encuentran cada una de las actividades para actualizar el avance y ver si es necesaria la introducción de cambio.

Una vez hemos descrito de forma genérica los pasos a seguir, vamos a realizar el plan de gestión del tiempo para nuestro proyecto de manera resumida:

- Se establece de qué manera vamos a programar nuestro proyecto así como que método de los conocidos utilizaremos. Nosotros decidimos utilizar el diagrama de Gantt, que es un gráfico de barras en la que las actividades están definidas en el eje vertical y el tiempo en horizontal, de forma que a mayor duración mayor longitud horizontal de las barras.
- Debemos pensar con qué nivel de exactitud queremos hacer las estimaciones sobre la duración de las actividades, en nuestro caso el nivel de exactitud no es muy alto debido a que hay factores externos que pueden desviar el tiempo de las actividades. Por esto decidimos utilizar las semanas como medida del tiempo.

Para poder realizar el diagrama de Gantt es necesario definir las actividades y sus relaciones de procedencia. Normalmente en este paso se tiene en cuenta la EDT realizada pero como en nuestro caso es muy simple no se realizó. Las actividades que realizaremos son las siguientes:

- Realizar la toma de contacto con el tutor y cotutor del proyecto para ver en qué consiste el trabajo y cómo se va a llevar a cabo.
- Aprender cómo funciona el programa R, utilizado para la elaboración del proyecto.

- Mejorar los conocimientos sobre motores de inducción
- Realizar el algoritmo elegido en el programa R.
- Ejecutar la aplicación realizada en R con cada una de las posibilidades que nos ofrecen los datos.
- Llevar a cabo una reunión para el análisis de la primera ejecución de la aplicación.
- Buscar mejoras para la aplicación y ejecutarla de nuevo con las directrices estipuladas en la reunión.
- Analizar los resultados y establecer que vamos a usar para conseguir los objetivos.
- Realizar un informe sobre lo que se ha estudiado y sus implicaciones.
- Tener una reunión con tutor y cotutor para la validación de los estudios realizados
- Tener una reunión con el departamento de ingeniería eléctrica para la explicación de la experimentación llevada a cabo por ellos.
- Buscar información sobre toda la metodología utilizada
- Elaborar una memoria para el TFG
- Supervisar los resultados
- Preparar una presentación
- Realizar la presentación

Una vez tenemos todas las actividades definidas a priori se procede a realizar la tabla de relaciones de procedencia (tabla 8.1), cuya importancia es alta debido a que permite saber que actividades pueden o no realizarse simultáneamente y a qué tiempo.

Actividad	Descripción	Precedente
A	Realizar la toma de contacto	-
B	Aprender a utilizar el programa R	A
C	Mejorar conocimientos sobre motores	B
D	Informarse sobre la metodología utilizada	B
E	Realizar algoritmo en R	D
F	Ejecutar la aplicación realizada en R	E
G	Reunión para analizar resultados	F
H	Buscar mejoras y ejecutar la aplicación	G
I	Analizar resultados y descartar los no válidos	H
J	Reunión con tutor y co-tutor	I
K	Reunión departamento ingeniería eléctrica	J
L	Supervisar resultados	K
M	Elaborar una memoria	L
N	Revisar Memoria	M
N	Corrección Memoria	N
O	Preparar Presentación	N
P	Realizar Presentación	O

Tabla 8.1 Actividades y precedentes

En relación a las dependencias mostradas en la tabla, existen tres clases de dependencia en función del origen de la limitación:

- Dependencia obligatoria: Propias a la naturaleza del trabajo que se lleva a cabo.
- Dependencia discrecional: Definidas por el equipo de Dirección del Proyecto. Están basadas en experiencia previa.
- Dependencia externa: Independiente de las actividades del proyecto.

También existen dependencias en función de la relación que hay entre las diferentes actividades:

- Procedencia final a comienzo: La actividad que viene después solo puede empezar, una vez haya finalizado la anterior.
- Procedencia comienzo a comienzo: La actividad sucesora puede empezar una vez haya empezado la antecesora.
- Procedencia final a final: La actividad sucesora puede finalizar una vez haya finalizado su antecesora.
- Procedencia comienzo a final: la actividad sucesora no puede finalizar hasta que su sucesora haya comenzado.

Una vez hemos definido las actividades que forman parte del proyecto y sus relaciones de procedencia, es necesario estimar los recursos que debemos utilizar para cada una de ellas así como la duración que esperamos que tengan. Como recursos entendemos personas, maquinas o instalaciones que

debemos utilizar para que la actividad se lleva a cabo de forma satisfactoria. Por lo general este paso es complejo en la realización de un proyecto de mayores dimensiones, ya que en nuestro caso los recursos así como las unidades que disponemos de cada uno son prácticamente ilimitados debido a su naturaleza. Explicado esto procedemos a asignar recursos y duración estimada a cada una de las actividades de la tabla A continuación vamos a estimar los recursos necesarios para cada actividad.

Actividad	Recursos(personas)	Recursos(materiales)	Duración(sem)
A	Ing. Organización y tutores	-	1
B	Ing. Organización	Ordenador, R y Tutorial	2
C	Ing. Organización	Ordenador, Libros	1
D	Ing. Organización	Ordenador, Libros	2
E	Ing. Organización	Ordenador, R	3
F	Ing. Organización	Ordenador, R	3
G	Ing. Organización y tutores	-	1
H	Ing. Organización	-	3
I	Ing. Organización	-	4
J	Ing. Organización y tutores	-	1
K	Ing. Organización y Ing. Industrial	-	1
L	Tutores (Doc. Estadística)	-	1
M	Ing. Organización	-	6
N	Tutores (Doc. Estadística)	-	2
Ñ	Ing. Organización	-	1
O	Ing. Organización	-	2
P	Ing. Organización	-	1

**Tabla 8.2** Actividades y su duración

En relación a la duración de las actividades hay que resaltar que la duración no está referida a una dedicación completa sino que en el transcurso de las mismas se han realizado otras solapadas, como puede ser el caso de la realización de la memoria que se ha ido realizando a medida que se avanzaba en el proyecto. La duración de muchas de estas actividades ha surgido alguna modificación durante el proceso debido a varias causas, pero siempre han sido cambios esperados. Una vez tenemos tanto las actividades y sus relaciones como su duración podemos efectuar el diagrama de Gantt (figura 8.1).

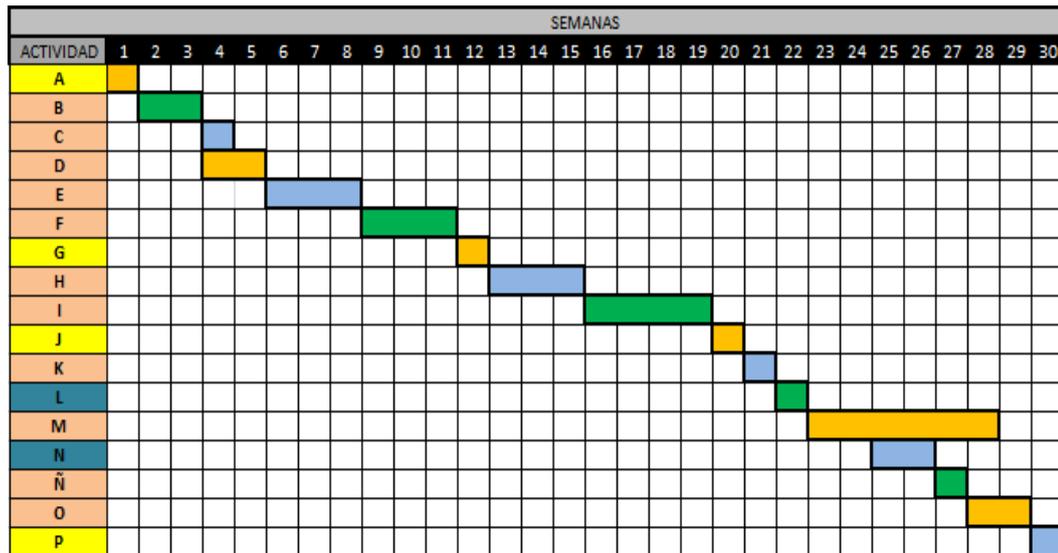


Figura 8.1 Diagrama de Gantt

## 8.4 GESTIÓN DE LOS COSTES

Para llevar a cabo una adecuada gestión de los costes de un proyecto debemos seguir los siguientes pasos:

1. Planificar los costes: En nuestro proyecto al tratarse de un proyecto académico no conlleva grandes inversiones por lo que las políticas y procedimientos para la gestión y ejecución de costes pasa desapercibida para un proyecto de estas características.
2. Estimar los costes: Aunque se trate de un proyecto de ámbito académico puede tener su interés en el ámbito industrial por lo que asignaremos a cada recurso de los utilizados independientemente del carácter de estos recursos. En la tabla 8.3 se indica el coste de cada recurso utilizado. Los sueldos brutos están calculados en base a una encuesta del Colegio de Ingenieros de Guipúzcoa teniendo en cuenta antigüedad y rama, y en torno al salario medio de un profesor doctor en una universidad pública de España.

Recurso	Coste Estimado
Ingeniero Organización	20 €/hora
Ingeniero Industrial	33 €/hora
Profesor Titular	25 €/hora
Ordenador	Propiedad de las partes del proyecto
Programas( R, Excel, Statgraphics)	Versiones educativas
Gastos varios(luz, papel, impresora...)	% del total

Tabla 8.3 Recursos utilizados y sus costes

3. Determinar el presupuesto: Una vez que hemos asignado un coste a los recursos que vamos a utilizar para el proyecto, hay que sumar los costes estimados para cada actividad. Es aconsejable destinar una partida siempre para imprevistos que puedan surgir durante la realización del proyecto, en nuestro, debido a la naturaleza del mismo no será necesario.

Recurso	Coste	Cantidad	Total
Ing. Organización	20 €/hora	760 horas	15200 €
Profesor titular	25 €/hora	50 horas	1500 €
Ing. Industrial	33 €/hora	3 horas	99 €
			16799 €
Gastos varios	5%	-	840 €
TOTAL			17639 €

**Tabla 8.4** Resumen de costes

4. Controlar los costes: monitorear el estado del proyecto para gestionar posibles cambios. La gestión de los costes se encarga principalmente del coste de los recursos necesarios para completar el proyecto. Al gestionar los costes no se debe tratar de ahorrar sin antes analizar las consecuencias finales de eliminar algunas actividades

# ***CAPÍTULO 9:***

## ***CONCLUSIONES***



Este trabajo está basado en la aplicación de métodos estadísticos robustos a un problema con un gran interés práctico tanto en la industria como en ingeniería como es el de clasificar los fallos que se dan en motores de inducción. Este problema posee una gran importancia ya que esta clase de motores son muy usados en la actualidad por lo que desarrollar un método que se encargue de la detección de fallos antes de que estos aparezcan y sin la necesidad de parar los motores puede suponer una gran reducción de costes.

Además en este trabajo también se ha querido realizar una comparación entre cómo se comportan los métodos robustos, que son relativamente novedosos, a un método no robusto. Los métodos robustos por su propia naturaleza deben suponer una clara mejora respecto a los métodos no robustos, si hay observaciones anómalas. Por lo tanto, este trabajo sirve para comparar la capacidad de detección de fallos entre unos otros después de utilizar los procedimientos estadísticos adecuados.

Mediante el uso de estos métodos se ha intentado detectar en distintos motores de inducción conectados a diferentes frecuencias y variadores, los distintos fallos que sufren en función de su gravedad. Dicho nivel de gravedad se ha medido asignando estados del R1 al R4 en siendo R1 el estado del motor sano y el R4 el estado de un motor completamente roto. Concretamente se ha buscado distinguir entre los motores que se encuentran completamente sanos (R1), respecto los que tienen algún fallo ya sea leve o muy grave (R234), descartando realizar un análisis sobre detectar como es la clase de fallo que tiene el motor, que sería clasificar entre R1-R2-R3-R4. Tras la realización del experimento con los métodos robustos se llegó a la conclusión de la gran eficiencia que se conseguía en la detección temprana de errores ya que en el 95% de los casos no hacía falta que un motor llegase al estado R4 para que fuese considerado como defectuoso, sino que era detectado entre los estados R2 y R3. Este dato supera claramente a los experimentos realizados con métodos no robustos cuyo eficiencia media se sitúa en torno al 75%, lo que supone una clara mejora que hace rentable la aplicación de dichos métodos.

Para terminar, se definió como objetivo comprobar si dicho método de detección temprana de fallos podía ser aplicado de manera más general para motores que trabajasen a la misma frecuencia sin tener en cuenta el variador al que se encuentran conectados o para motores conectados al mismo variador pero que trabajaran a frecuencia diferente. Tras la realización del experimento se determinó que debido a los comportamientos de estos motores, no era posible la aplicación de este método para ese análisis, ya que los comportamientos eran tan diversos que era imposible establecer un umbral de clasificación para la clasificación R1-R234.



***CAPÍTULO 10:***  
***ANEXOS***



## 10.1 INTRODUCCIÓN

En este apartado se incluirá el código que se ha utilizado para llevar a cabo el trabajo. Como se ha utilizado el mismo código para cada uno de los ensayos exceptuando los cambios propios del archivo correspondiente de datos y la variación del tamaño de la muestra, solo se incluirá el correspondiente al ensayo AB35 para que sirva de ejemplo. Para que sea mejor la comprensión se dividirá en las fases explicadas en el capítulo 7.

A continuación del código utilizado se mostraran las gráficas de control obtenidas para cada uno de los ensayos que no se mostraron en el Capítulo 7 de forma que se pueda ver la diferencia entre ellos a la hora de clasificar correctamente las observaciones.

Por último se adjuntan las tablas que contienen el porcentaje de aciertos en los distintos ensayos y para diferentes tipos de discriminación.

## 10.2 CÓDIGOS

*#Llamamos al archivo en el que tenemos almacenados los datos sobre los que vamos a trabajar#*

```
Datos<-read.table
("C:/Users/Sergio/Desktop/2ºCuatri/TFG/Datos/AB/AB35.csv",sep
p = ";",dec=".",header=TRUE)
```

*library(robustbase) #Activamos la libreria robustbase de R para poder calcular el MCD#*

*# Damos características de color (verde para R1, azul para R2, naranja R3 y rojo para R4) y asignamos un número (1 para R1 y 2 para R234) para distinguir de una manera más clara los estados en las gráficas que realizaremos en el ensayo correspondiente#*

```
sano <- rep("green",241)
sano[126:163] <- "blue"
sano[164:198] <- "darkorange"
sano[199:241] <- "red"
sano.num <- rep("1",241)
sano.num[126:241] <- "2"
```

```
datos <- cbind(datos,sano) #Asignamos las características
que creamos anteriormente al vector de datos X2, que
contiene los datos que vamos a trabajar#
```

```
names(datos)[11:14] <- c("BS","B5S","B5I","B7S") #Damos
nombre a las columnas#
```

```
#####
#####
# :Parte 1:Resúmenes gráficos iniciales para las variables
seleccionadas.#####
#####
#####
```

```
pairs(datos[11:14],pch=sano.num,col=sano) #Representación
de las gráficas de pares#
```

```
pairs(datos[1:125,11:14],pch=sano.num) #Representación
únicamente de las observaciones que se encuentran en estado
R1 pareada#
```

```
#Representación con los 4 colores de cómo evoluciona cada
variable de manera individual para todas las observaciones
que disponemos (R1-R4)#
```

```
par(mfrow=c(2,2))
```

```
plot(datos[,11],col=sano,pch=sano.num,type="b",ylab="BS")
```

```
plot(datos[,12],col=sano,pch=sano.num,type="b",ylab="B5S")
```

```
plot(datos[,13],col=sano,pch=sano.num,type="b",ylab="B5I")
```

```
plot(datos[,14],col=sano,pch=sano.num,type="b",ylab="B7S")
```

```
#Representación de elipses de tolerancia en las que
enfrentamos dos a dos las variables y obtenemos de una
manera gráfica una comparación entre la detección de
outliers clásica/robusta#
```

```
par(mfrow=c(1,2))
```

```
tolEllipsePlot(datos[1:125,11:12],classic=TRUE,xlab="BS",yla
b="B5S")
```

```
tolEllipsePlot(datos[1:125,13:14],classic=TRUE,xlab="B5I",yl
ab="B7S")
```

*#FIN DE LA PARTE I#*

```
#####
#####
PARTE 2: Se lleva a cabo una comparación de la fase I no
robusta y robusta. Buscamos demostrar que la existencia de
unos pocos outliers nos lleva a pensar que el proceso está
bajo control, es decir, sin outliers, y que podemos usar esa
estimación de medias y varianzas para la fase II, cuando no
es así. #####
#####
```

```
#####
#Fase I clásica
#####
```

*#Construcción de una función (número de observaciones m y dimensión p) que llamaremos critical.usual, a la cual llamaremos más adelante, y que nos proporciona el valor crítico. Dicho valor crítico se calcula por simulación (simulamos B muestras de la normal, calculamos las distancias de mahalanobis correspondientes y buscamos el percentil 95).#*

```
critical.usual <- function(m,p){
  B <- 1000
  sim <- (1:B)
  for (i in 1:B){
    X <- matrix(rnorm(m*p),ncol=p)
    sim[i] <- max(mahalanobis(X,apply(X,2,mean),cov(X)))
  }
  res <- sort(sim)[floor(B*0.95)]
  return(res)
}

#Seleccionamos las observaciones que contienen solo motores
sanos (R1)#

X <- datos[1:125,11:14]
```

```

#Llamamos a la función para obtener el valor crítico#

m <- dim(X)[1]

p <- dim(X)[2]

cri <- critical.usual(m,p)

#Monitorizamos los Ti2 de las observaciones R1 con su valor
crítico de forma gráfica para ver si queda dentro de los
límites de control#

par(mfrow=c(1,2))

X <- matrix(rnorm(m*p),ncol=p)

ti <- mahalanobis(X,apply(X,2,mean),cov(X))

plot(ti,type="b",ylim=c(0,max(cbind(ti,cri+3))),ylab="Hotelling's T^2",main="(a)")

abline(h=cri,col=2,lty=3)

#####
#Fase I Robusta
#####

#Construcción de una función (número de observaciones m y
dimensión p) que llamaremos critical.rob, a la cual
llamaremos más adelante, y que nos proporciona el valor
crítico. Dicho valor crítico se calcula por simulación
(simulamos B muestras de la normal, calculamos las
distancias de mahalanobis obtenidas por el MCD y buscamos el
percentil 95)#

Critical.rob <- function(m,p){

  B <- 1000

  sim <- (1:B)

  for (i in 1:B){

    X <- matrix(rnorm(m*p),ncol=p)

    a <- covMcd(X)

```

```

    sim[i] <- max(mahalanobis(X,a$center,a$cov))
  }

  res <- sort(sim)[floor(B*0.95)]

  return(res)
}

#Seleccionamos las observaciones que contienen solo motores
sanos (R1)#

X <- datos[1:125,11:14]

#Llamamos a la función para obtener el valor critico#

m <- dim(X)[1]

p <- dim(X)[2]

cri <- critical(m,p)

#Monitorizamos los valores de Ti2 robustos con su valor
crítico en un gráfico para ver si queda dentro de los
límites que denominamos de control y para comprobar si los
resultados obtenidos en la parte clásica son o no fiables#

a <- covMcd(X)

ti <- mahalanobis(X,a$center,a$cov)

plot(ti,type="b",ylim=c(0,max(cbind(ti,cri+3))),ylab="Robus
t T^2 (MCD)",main="(b)")

abline(h=cri,col=2,lty=3)

#FIN DE LA PARTE 2#

#####
#####
PARTE 3: Comparación de los resultados obtenidos de la fase
II robusta y no robusta#####
#####
#####
#####

#####
#Fase II Clásica
#####

```

```
#Asignación de los valores m y p que utilizaremos luego para
obtener el valor crítico, y creación del vector datos.tot
que contiene todas las observaciones, no eliminamos los
outliers detectados, del ensayo#
```

```
datos.0 <- datos[1:125,11:14]
datos.1 <- datos[-(1:125),11:14]
datos.tot <- datos[,11:14]
m <- dim(datos.0)[1]
p <- dim(datos.0)[2]
```

```
#Calculamos el valor crítico que vamos a usar para chequear
las nuevas observaciones#
```

```
cri <- (p*(m+1)*(m-1))/(m*(m-p))*qf(0.975,p,m-p)
```

```
#Obtención del valor de la  $T_i^2$  de Hotelling para todas las
observaciones mediante la distancia de mahalanobis con
estimadores no robustos, es decir, sin usar el MCD#
```

```
ti<- mahalanobis(datos.tot,apply(datos.0,2,mean),cov(datos.0))
```

```
# Monitorización de la distancia de Mahalanobis obtenida
antes para todas las observaciones junto con el valor
crítico obtenido en la función critical.usual #
```

```
plot(ti,type="b",ylim=c(0,max(cbind(ti,cri+3))),col=sano,pc
h=sano.num,ylab="Hotelling's T^2",main="Non Robust Phase
II")
abline(h=cri,col=2,lty=3)
```

```
#####
#Fase II Robusta
#####
```

```
#Aplicamos el MCD a las observaciones del motor sano (R1) y
obtenemos las atípicas#
```

```
a <- covMcd(X2[1:125,11:14])
outliers <- (1:125)[1-a$mcd.wt==1]
```

```
#Quitamos los outliers de las observaciones de los motores
sanos#
```

```
datos.0 <- datos[1:125,11:14]
```

```

datos.0 <- datos[-outliers,]

#Creamos los nuevos vectores de datos tras haber quitado
los outliers #

datos.1 <- datos[-(1:125),11:14]

datos.tot <- datos.1[-outliers,11:14]

sano <- sano[-outliers]

sano.num <- sano.num[-outliers]

#Construcción del valor critico del control de calidad
mediante la fórmula siguiente, explicada en el apartado
7.2.3#

m <- dim(datos.0)[1]

p <- dim(datos.0)[2]

cri <- (p*(m+1)*(m-1))/(m*(m-p))*qf(0.975,p,m-p)

# Obtención del valor de la  $T_i^2$  de Hotelling para todas las
observaciones mediante la distancia de mahalanobis con
estimadores robustos, es decir, usando los resultados
obtenidos tras la aplicación del MCD #

ti <- mahalanobis(Xtot,apply(X2.0,2,mean),cov(X2.0))

#Monitorización de la distancia de Mahalanobis obtenida
antes para todas las observaciones junto con el valor
crítico obtenido en la función critical.rob #

plot(ti,type="b",ylim=c(0,max(cbind(ti,cri+3))),col=sano,pc
h=sano.num,main="Robust Phase II",ylab="Hotelling's T^2")

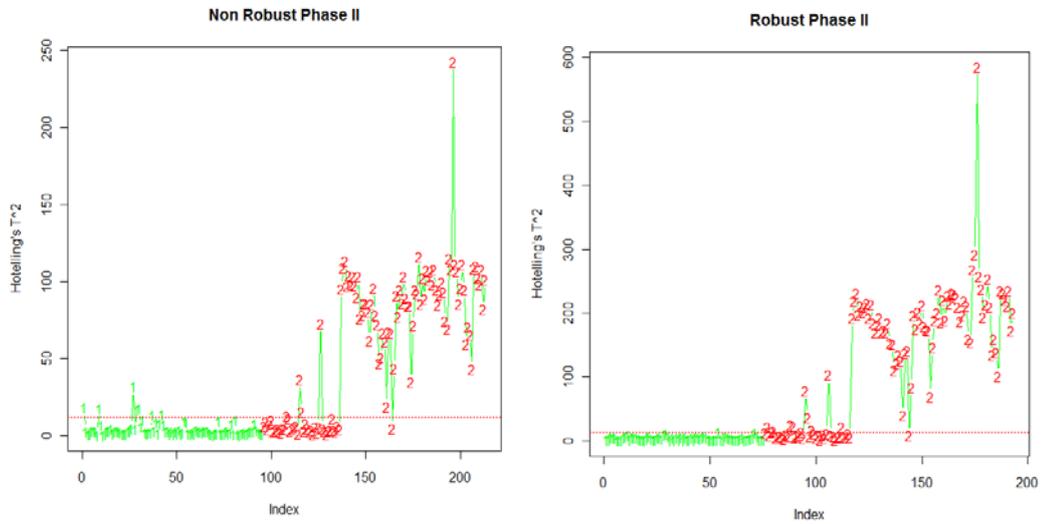
abline(h=cri,col=2,lty=3)

#FIN DEL PROGRAMA#

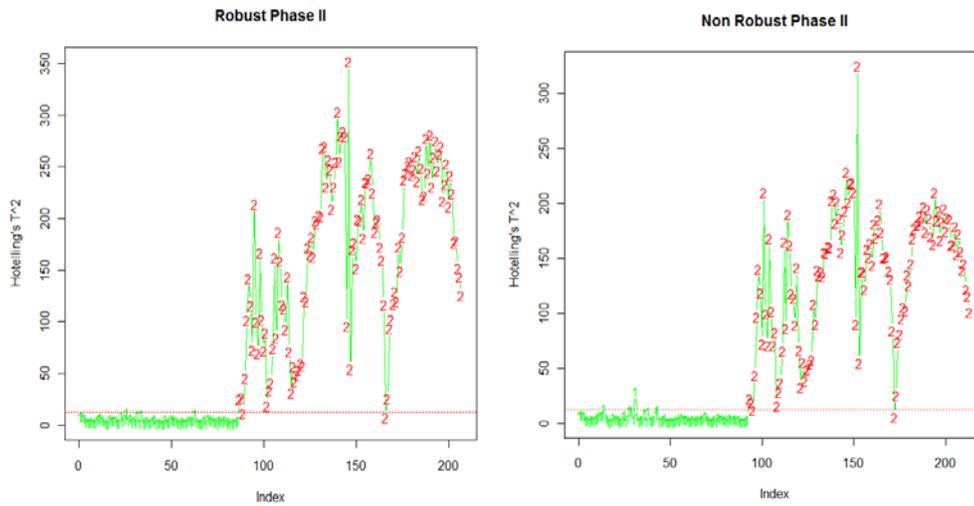
```

## 10.3 GRÁFICOS DE CONTROL ROBUSTO DE CADA ENSAYO

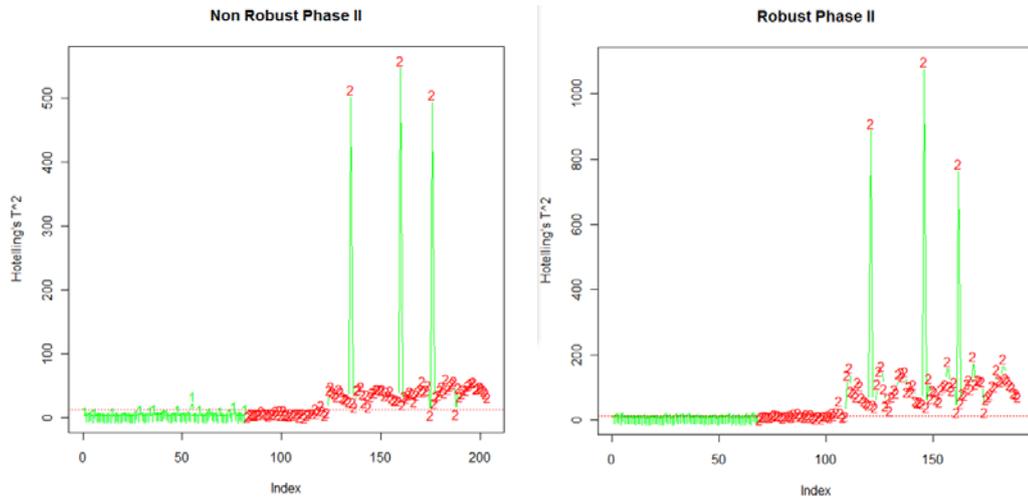
### 10.3.1 AB50



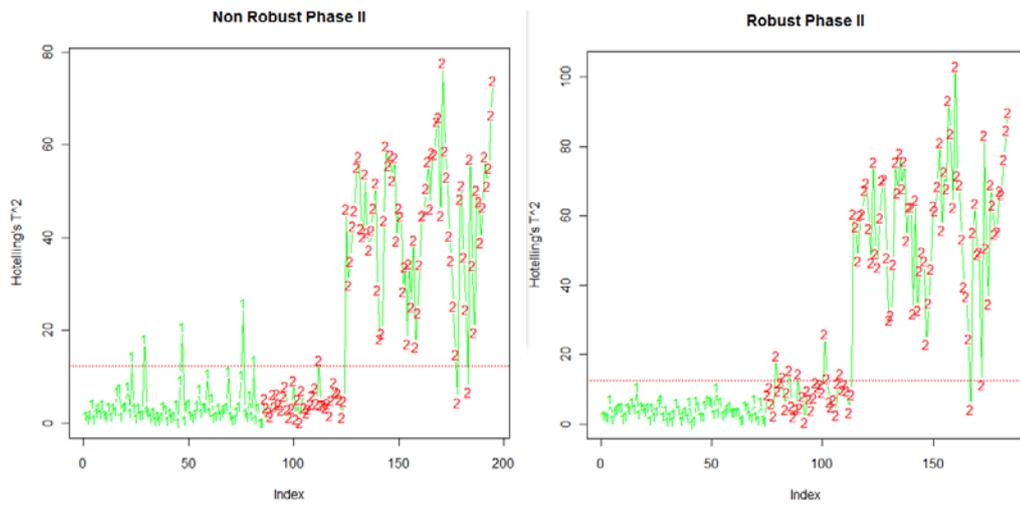
### 10.3.2 AB65



### 10.3.3 SM35



### 10.3.4 SM50





# ***BIBLIOGRAFÍA***



## BIBLIOGRAFIA

[DUQ05] Duque, O. y Pérez, M.(2005). “Motores de inducción-Técnicas de mantenimiento predictivo”. Madrid-Valladolid: @becedario.

[FRA08] Fraile J. (2008) “Máquinas eléctricas 6ª edición”. Madrid: Mc.Graw-Hill.

[GAR08] Garcia, A. “Estadística aplicada con R”. UNED

[GAR11] Garcia, L.A., Duque, O., Morinigo, D., Perez, M. (2011) “Robust condition monitoring for early detection of broken rotor bars in induction motors”. Expert systems with applications 38 (2011) 2653-2660

[GOR14] Gordo, D. (2014) “Detección de fallos en motores de inducción con técnicas de machine learning”, trabajo de fin de grado. Escuela de Ingenierías Industriales (Universidad de Valladolid)

[HUB08] Hubert, M., Rousseeuw, P., Van Aelst, S. (2008) “High-Breakdown Robust Multivariate Methods”. Statistical Science 2008, Vol.23, No 1, 92-119

[HUB04] Huber P. (2004) “Robust Statistics”. New Jersey: Wiley

JEN[07] Jensen, W. A., Birch, J.B., & Woodall, W. H.(2007). “High breakdown estimation methods for phase I multivariate control charts. Quality and Reliability Engineering International, 2007, 23(5), 615-629.

[PEÑ02] Peña, D. (2002) “Análisis de datos multivariante”. Madrid:Mc.Graw-Hill

[PER13] Pérez, C. (2013) “Análisis multivariante de datos. Aplicaciones con IBM SPSS, SAS y StatGraphics”. Madrid:Garceta

[PMBOK13] “Guía para la dirección de proyectos”. Project Management Institute, Inc. (2013)

[ROD13] Rodriguez, B. (2013). “Detección de fallos en motores de inducción utilizando histogramas”, proyecto fin de carrera. Escuela de Ingenierías Industriales (Universidad de Valladolid)

[ROU99] Rousseeuw, P.J, Van Driessen, K. “A fast algorithm for the minimum covariance determinant estimator” TECHNOMETRICS, VOL.41 No. 3

#### **PAGINAS WEB**

- a) <http://cran.r-project.org/>
- b) [http:// projecteuclid.org/](http://projecteuclid.org/)
- c) <http:// instrumentaciónycontrol.net>

