

**DETERMINACIÓN ESTADÍSTICA DE UMBRALES DE CALIDAD DE  
PARAMETROS FISICO-QUIMICOS EMPLEADOS EN LA DETERMINACIÓN  
DEL ESTADO ECOLÓGICO DE LOS RÍOS**

Concha DURÁN (Confederación Hidrográfica del Ebro)

Pilar GARGALLO (Universidad de Zaragoza)

Miriam PARDOS (Confederación Hidrográfica del Ebro)

Manuel SALVADOR (Universidad de Zaragoza)

# **INDICE**

## **1. OBJETIVO DEL ESTUDIO**

## **2. PLANTEAMIENTO DEL PROBLEMA**

## **3. DESCRIPCIÓN DE LOS DATOS**

### **3.1. Análisis estadístico de los parámetros biológicos**

### **3.2. Análisis estadístico de los parámetros físico-químicos**

## **4. METODOLOGIA ESTADISTICA**

## **5. DETERMINACIÓN DE LOS UMBRALES DE CALIDAD**

### **5.1. Umbrales univariantes**

### **5.2 Umbrales multivariantes**

## **6. CONCLUSIONES Y LINEAS DE INVESTIGACION FUTURAS**

## **1. OBJETIVO DEL ESTUDIO**

En este trabajo se realiza una determinación estadística de umbrales de calidad del agua a partir de parámetros físico-químicos determinados mediante un proceso de muestreo. Para ello se realiza un estudio de la relación de dichos parámetros con los parámetros biológicos del agua para, a partir de la misma, determinar los valores de los parámetros físico-químicos asociados a niveles de calidad biológica significativamente bajos. De esta forma se busca mejorar los procedimientos sistemáticos de determinación de la calidad del agua, basados habitualmente en el estudio de sus parámetros biológicos, complementándolos con la información proporcionada por sus parámetros físico-químicos que suelen ser más fáciles de determinar y que pueden ayudar a prevenir deterioros futuros del nivel de calidad del agua.

El estudio de la relación existente entre los parámetros físico-químicos y los biológicos se realiza mediante la utilización de modelos de regresión con variable dependiente cualitativa y, más concretamente, modelos de regresión logística-binaria dado que nuestro objetivo es determinar si existe o no un deterioro significativo de la calidad del agua. En dichos modelos las variables biológicas juegan el papel de variable dependiente y se utilizan para determinar si existe o no dicho deterioro, mientras que las variables físico-químicas juegan el papel de variables independientes. Utilizando los modelos estimados se determinan, para cada uno de los parámetros físico-químicos, los umbrales de calidad del agua para un nivel de riesgo dado. Así mismo, se realiza un estudio multivariante del problema que analiza las interacciones existentes entre los diversos parámetros con el fin de determinar aquéllos con más poder discriminante de la calidad, teniendo en cuenta las redundancias existentes entre ellos. Previamente a la estimación de dichos modelos se realiza un análisis exploratorio de los datos utilizados en el estudio con el fin de describir el grado de representatividad de los resultados obtenidos así como de reducir el impacto que puedan ejercer la presencia de observaciones atípicas en los mismos.

El informe se organiza de la siguiente manera:

- Sección 2: se describen las variables y los datos utilizados en el estudio
- Sección 3 se realiza un análisis exploratorio de los mismos
- Sección 4 se expone la metodología estadística utilizada para la determinación de los umbrales

- Sección 5 se muestran los resultados obtenidos al aplicar dicha metodología.
- Sección 6 concluye exponiendo los resultados más relevantes así como las líneas de investigaciones futuras que el equipo pretende abordar en trabajos posteriores.

## 2. PLANTEAMIENTO DEL PROBLEMA

Los datos corresponden a la medición, en distintos puntos de la cuenca del Ebro, de diversos parámetros biológicos y fisico-químicos de muestras tomadas en ríos. Las observaciones se realizaron en el periodo de tiempo comprendido entre los años 2002 y 2007, con una frecuencia quincenal-mensual, en el caso de las variables fisico-químicas, y anual en el caso de los indicadores de calidad biológicos, debido al mayor coste y grado de dificultad de la medición de éstos últimos. El número total de puntos muestreados fue igual a 355 dando lugar a un total de 6635 observaciones. El número de observaciones por punto de muestreo osciló entre 1 y 259 con un número medio igual a 19.

**Tabla 2.1: Frecuencia de observaciones por año**

Año	Frecuencia	Porcentaje
2002	1341	20,2
2003	999	15,1
2004	641	9,7
2005	1312	19,8
2006	1283	19,3
2007	1059	16,0
Total	6635	100,0

En la Tabla 2.1 se muestra el número de observaciones por año, apreciándose que durante los años 2003, 2004 y 2007 dicho número fue significativamente menor, especialmente en el año 2004 debido a razones presupuestarias.

Los indicadores biológicos considerados en el estudio fueron 3 de los utilizados por la Confederación Hidrográfica del Ebro (CHE) para la determinación del estado ecológico en ríos y aparecen listados en la Tabla 2.2. En dicha tabla se muestran, además, los umbrales utilizados por la CHE para determinar si el nivel de calidad del agua cumple los requisitos mínimos de acuerdo con lo indicado en la Directiva Marco del Agua.

**Tabla 2.2: Indicadores biológicos utilizados en el estudio**

<b>Indicador</b>	<b>Umbrales de calidad</b>
IBMWP (Macroinvertebrados)	Aguas contaminadas: IBMWP $\leq$ 60 Aguas no contaminadas: IBMWP $>$ 60
IPS (Diatomeas)	Calidad mala: IPS $<$ 17 Calidad buena: IPS $\geq$ 17
IVAM (Macrófitos)	Calidad mala: IVAM $<$ 4.5 Calidad buena: IVAM $\geq$ 4.5

Los indicadores físico-químicos considerados en el estudio fueron 21 y aparecen listados en la Tabla 2.3. Más concretamente se han utilizado 7 indicadores generales: temperatura del agua (en adelante, TEM\_AG), temperatura del aire (TEM\_AI), índice pH (PH), conductividad (COND20), oxígeno disuelto en concentración (O2) y en porcentaje de saturación (O2%) y sólidos en suspensión (MES), 1 de materia orgánica: demanda química de oxígeno (DQO\_D), 2 de acidificación: carbonatos (CO3CA) y bicarbonatos (CO32CA), 5 de salinidad: concentraciones de sodio (NA), calcio (CA), magnesio (MG), cloruros (CL) y sulfatos (SO4) y 6 indicadores de nutrientes: amonio (NH4), nitritos (NO2), nitratos (NO3), nitrógeno Kjeldahl (N\_KJED), fosfatos (PO4) y fósforo total (P\_TOT).

Nuestro objetivo se concreta en establecer umbrales para cada uno de estos indicadores que permitan predecir el índice de calidad del agua en un punto de muestreo con la mayor fiabilidad posible. Para tal fin se ha utilizado la metodología estadística que se describe en la sección 4. Previamente en la sección siguiente realizamos un estudio estadístico-descriptivo de los datos utilizados en el estudio.

**Tabla 2.3: Indicadores físico-químicos utilizados en el estudio**

<b>Tipo de indicador</b>	<b>Nombre</b>	<b>Significado</b>
Generales	TEM_AG	Temperatura del agua
	TEM_AI	Temperatura del aire
	PH	pH
	COND20	Conductividad a 20°C
	O2	Oxígeno disuelto (en concentración)
	O2%	Oxígeno disuelto (en % de saturación)
	MES	Sólidos en suspensión
Materia orgánica	DQO	Demanda Química de Oxígeno (al dicromato)
Acidificación	CO3CA	Carbonatos
	CO32CA	Bicarbonatos
Salinidad	NA	Sodio
	CA	Calcio
	MG	Magnesio
	CL	Cloruros
	SO4	Sulfatos
Nutrientes	NH4	Amonio Total
	NO2	Nitritos
	NO3	Nitratos
	N_KJED	Nitrógeno Kjeldahl
	PO4	Fosfatos
	P_TOT	Fósforo Total

### **3. DESCRIPCIÓN DE LOS DATOS**

En esta sección se lleva a cabo un análisis exploratorio de las variables consideradas en el estudio. Su finalidad es poner de manifiesto las características más relevantes de los datos prestando especial atención al grado de representatividad de la muestra así como al estudio de observaciones atípicas con el fin de disminuir su influencia en la determinación de los umbrales.

El estudio se lleva a cabo tanto desde un punto de vista gráfico como numérico. Dado que la información correspondiente a los datos físico-químicos es cuantitativa se analiza, en particular, la normalidad de dichas variables aplicando, si fuera preciso, una transformación de las mismas con el fin de aumentar su grado de normalidad. Ello es debido a que, a la hora de estimar los parámetros de los modelos utilizados en la sección 5 y hacer inferencias acerca de la población objeto de estudio, dichas estimaciones son más fiables y menos sensibles a la presencia de atípicos si se verifica la hipótesis de normalidad.

Así mismo, y con el fin de analizar la representatividad de la muestra, se ha realizado un estudio de los datos ausentes en cada variable, verificando la existencia o no de patrones sistemáticos en los mismos que puedan disminuir y/o sesgar significativamente el grado de representatividad y de acuracidad de los resultados.

Todos los cálculos presentados en el estudio se han realizado con el paquete estadístico SPSS 14.0.

#### **3.1. ANÁLISIS ESTADÍSTICO DE LOS PARÁMETROS BIOLÓGICOS**

Comenzamos realizando un estudio estadístico de los indicadores biológicos del estudio, a saber: IBMWP, IPS e IVAM (ver Tabla 2.2). Los resultados se presentan por separado para cada uno de ellos.

##### **3.1.1. Análisis estadístico del indicador IBMWP**

Se observa, en primer lugar, la existencia de un alto porcentaje de datos ausentes (66,5%) en la muestra analizada debido a que *dicho indicador sólo se ha medido a partir del año 2004* (ver Tabla 3.1.1.1).

**Tabla 3.1.1.1: Datos ausentes en el indicador IBMWP**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
2221	33,5%	4414	66,5%	6635	100,0%

En la Tabla 3.1.1.2 se muestran los valores de algunos estadísticos descriptivos de este indicador. Así mismo, en las Figuras 3.1.1.1 a 3.1.1.3 se muestran, respectivamente, el histograma de frecuencias, el diagrama de cajas y el diagrama de cuantiles (qqplot) normal. Finalmente, en la Figura 3.1.1.4 se muestra, en forma de diagrama de barras de error, la evolución anual de su valor medio mientras que en la Figura 3.1.1.5 se muestra su distribución geográfica.

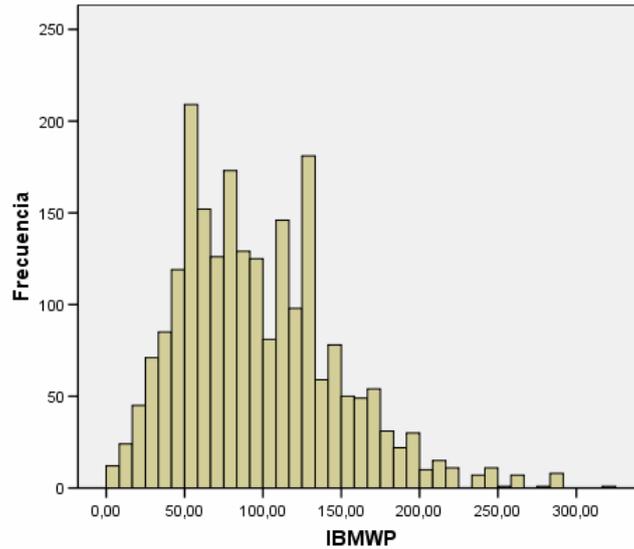
**Tabla 3.1.1.2: Análisis estadístico-descriptivo del indicador IBMWP**

		Estadístico	Error típ.
IBMWP	Media	96,9568	1,06653
	Intervalo de confianza para la media al 95%	Límite inferior Límite superior	
		94,8653 99,0483	
	Media recortada al 5%	94,3430	
	Mediana	87,0000	
	Varianza	2526,367	
	Desv. típ.	50,26298	
	Mínimo	7,00	
	Máximo	317,00	
	Rango	310,00	
	Amplitud intercuartil	72,00	
	Asimetría	,787	,052
	Curtosis	,715	,104

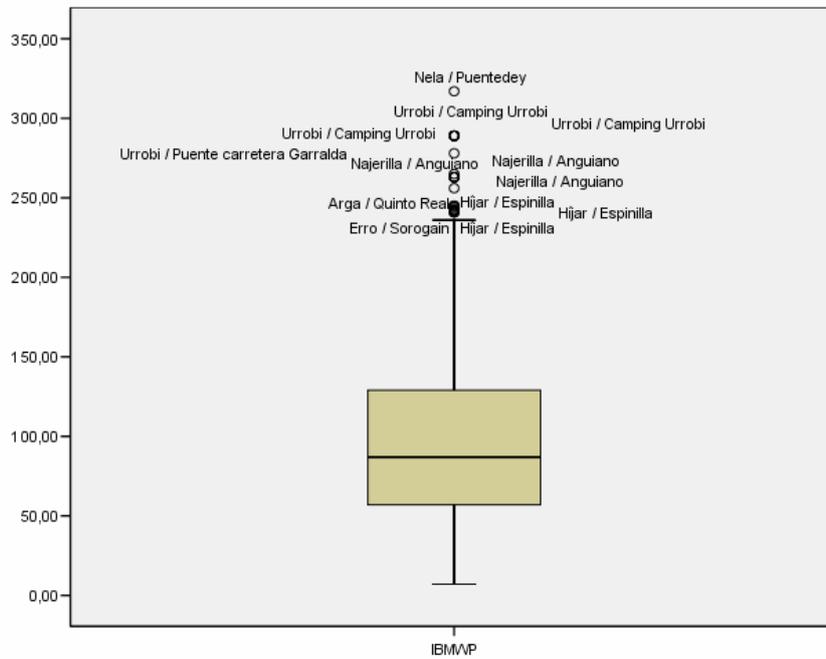
Se observa que, en general, *el nivel de calidad de las aguas de la cuenca ha sido bueno con un valor medio igual a 96,96 y un valor mediano igual a 87,00. Dicho nivel no ha sido, sin embargo, constante a lo largo del tiempo siendo peor en los años 2004 y 2005 y mejor en los años 2006 y 2007* (ver Figura 3.1.1.4). Existen, además, puntos de muestreo en los que el nivel de calidad de las aguas es muy bueno lo cual provoca la existencia de una asimetría positiva y una curtosis significativas (ver Tabla 3.1.1.2 y Figuras 3.1.1.1 a 3.1.1.3) y de algunos valores atípicos débiles

correspondientes a puntos de muestreo en los que la calidad del agua fue muy buena (ver Figura 3.1.1.2).

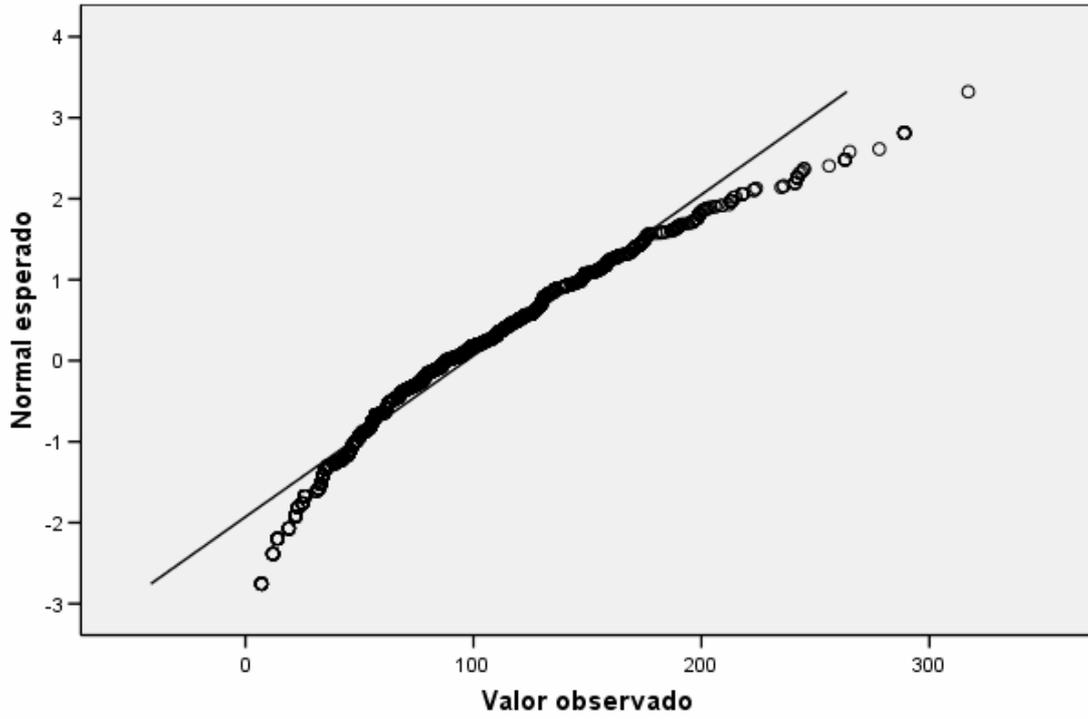
Estos hechos se aprecian claramente en la distribución geográfica del nivel de calidad del agua (ver Figura 3.1.1.5), *concentrándose la mayor parte de los puntos de muestreo con calidad muy buena en la zonas de montaña (norte y noroeste) de la cuenca.*



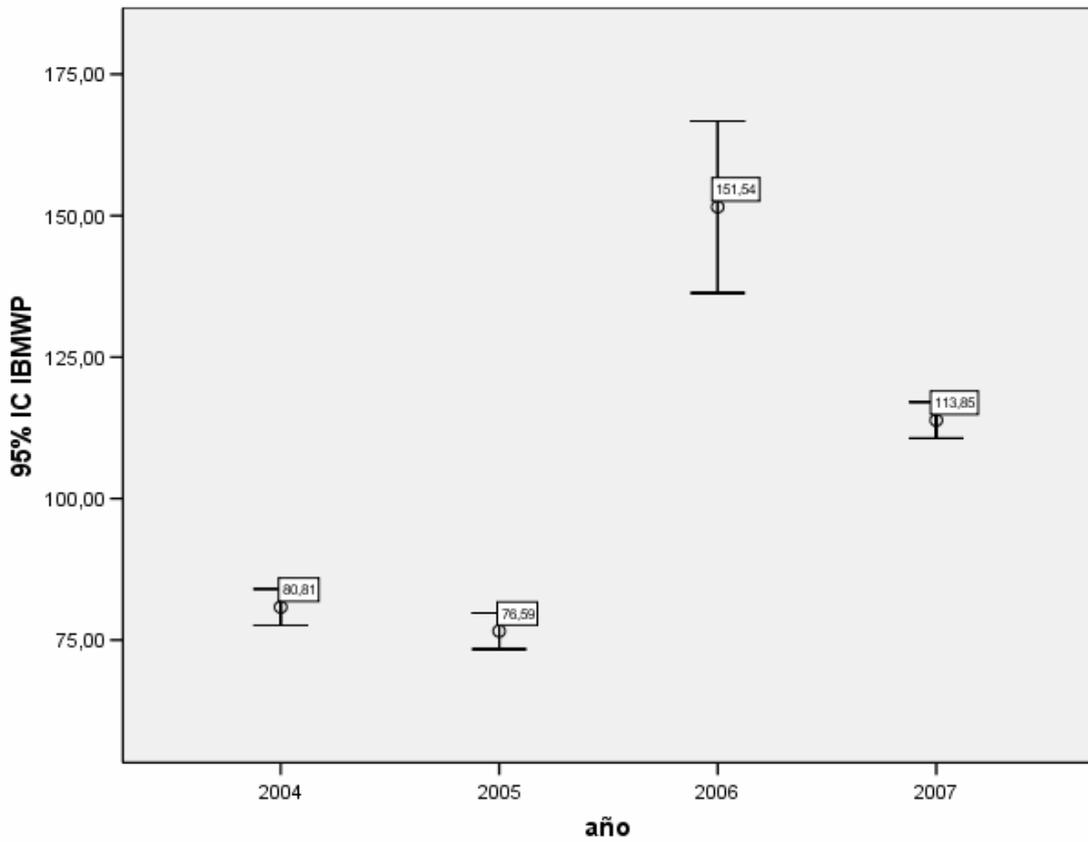
**Figura 3.1.1.1: Histograma de frecuencias del indicador IBMWP**



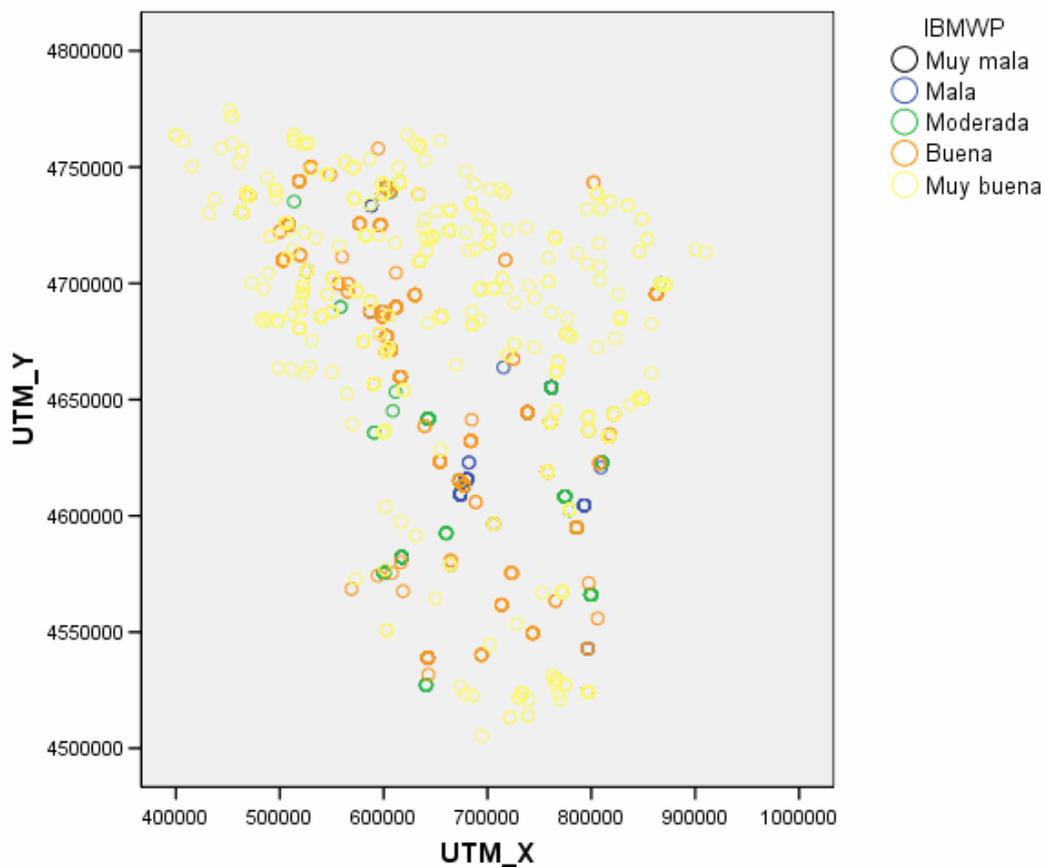
**Figura 3.1.1.2: Diagrama de cajas del indicador IBMWP**



**Figura 3.1.1.3: QQplot normal de indicador IBMWP**



**Figura 3.1.1.4: Evolución anual media del indicador IBMWP (encuadrados los valores medios del indicador en cada año)**



**Figura 3.1.1.5: Distribución geográfica de los niveles de calidad de las aguas medidos según el indicador IBMWP**  
 (Muy mala:  $IBMWP \leq 15$ ; mala:  $15 < IBMWP \leq 35$ ; moderada:  $35 < IBMWP \leq 60$ ; buena:  $60 < IBMWP \leq 100$ ; muy buena:  $IBMWP > 100$ )

### 3.1.2. Análisis estadístico del indicador IPS

Se observa, en primer lugar, *la existencia de un bajo porcentaje de datos ausentes (14,8%) la mayor parte de ellos debido a que dicho indicador no se observó en el año 2004* (ver Tabla 3.1.2.1).

**Tabla 3.1.2.1: Datos ausentes en el indicador IPS**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
5652	85,2%	983	14,8%	6635	100,0%

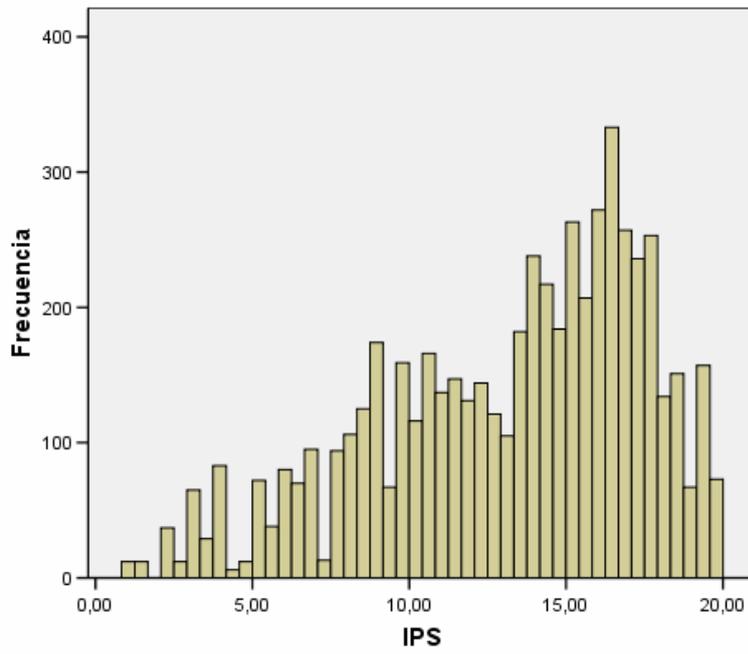
En la Tabla 3.1.2.2 se muestran los valores de algunos estadísticos descriptivos de este indicador. Así mismo, en las Figuras 3.1.2.1 a 3.1.2.3 se muestran su histograma de frecuencias, su diagrama de cajas y su diagrama qqplot normal. Finalmente, en la

Figura 3.1.2.4 se muestra, en forma de diagrama de barras de error, la evolución anual de su valor medio mientras que en la Figura 3.1.2.5 se representa su distribución geográfica.

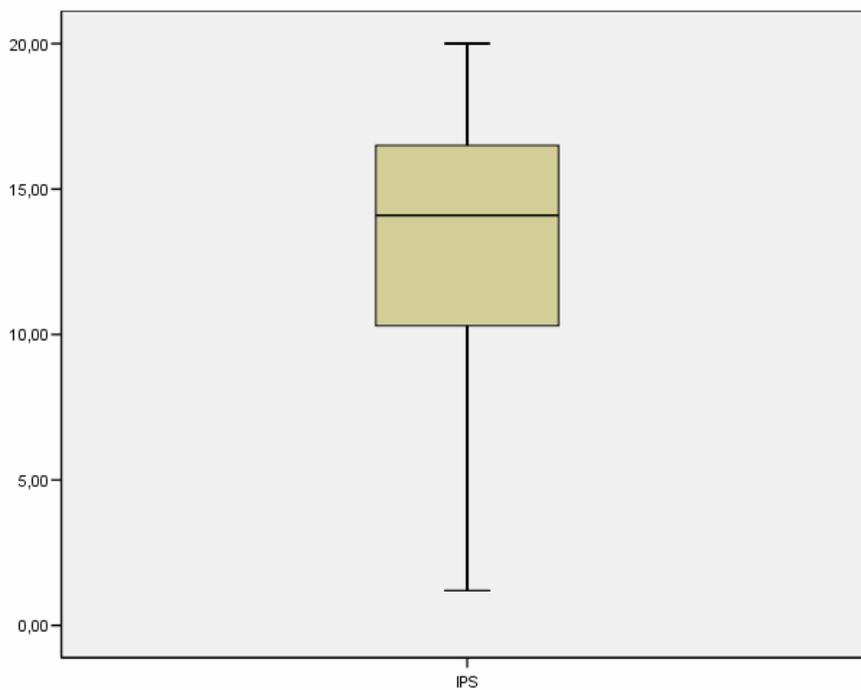
**Tabla 3.1.2.2: Análisis estadístico-descriptivo del indicador IPS**

		Estadístico	Error típ.	
IPS	Media	13,1759	,05675	
	Intervalo de confianza para la media al 95%	Límite inferior	13,0647	
		Límite superior	13,2872	
	Media recortada al 5%	13,3778		
	Mediana	14,1000		
	Varianza	18,204		
	Desv. típ.	4,26661		
	Mínimo	1,20		
	Máximo	20,00		
	Rango	18,80		
	Amplitud intercuartil	6,20		
	Asimetría	-,648	,033	
	Curtosis	-,378	,065	

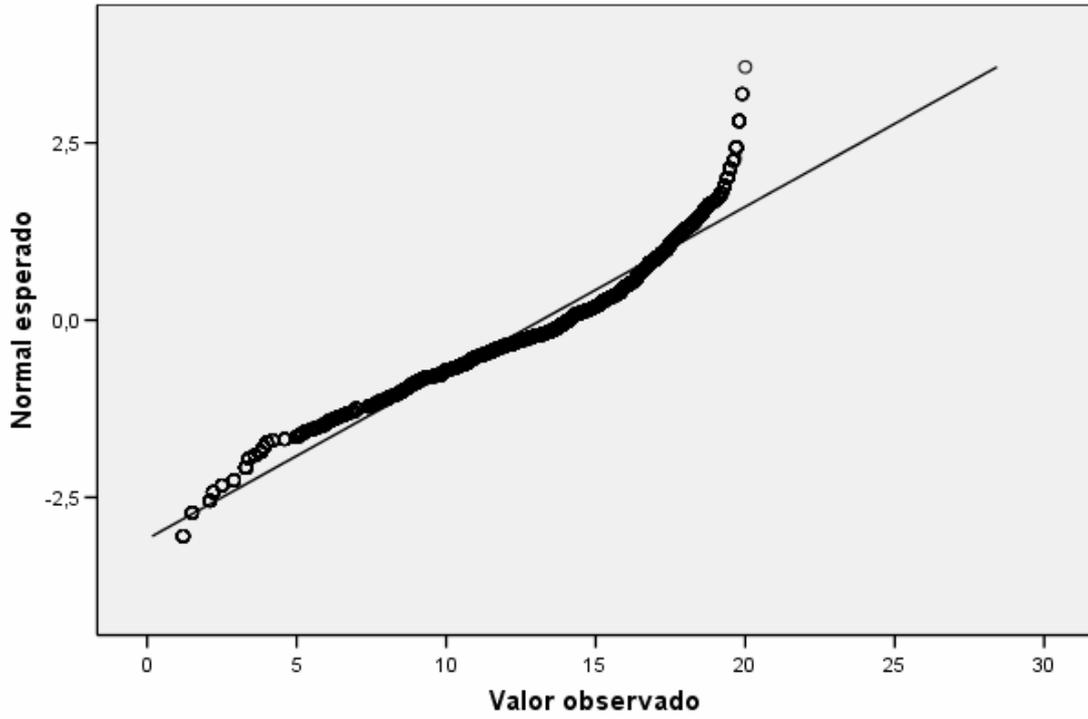
Se observa que, en general, *el nivel de calidad de las aguas de la cuenca ha sido bueno con un valor medio igual a 13,18 y un valor mediano igual a 14,10*. Dicho nivel no ha sido, sin embargo, constante a lo largo del tiempo siendo peor en los años 2005 y 2006 en los que el nivel de calidad fue moderado y mejor en los años 2002, 2003 y sobre todo 2007 (ver Figura 3.1.2.4). Existen, además, puntos de muestreo en los que el nivel de calidad de las aguas es muy bueno lo cual provoca la existencia de una asimetría negativa y una curtosis significativas (ver Tabla 3.1.2.2 y Figuras 3.1.2.1 a 3.1.2.3) no apreciándose, sin embargo, la existencia de puntos de muestreo con un comportamiento atípico. Si analizamos, finalmente, su distribución geográfica (ver Figura 3.1.2.5) se observa que *la mayor parte de los puntos de muestreo con niveles bajos de calidad del agua tienden a concentrarse a lo largo del cauce del río Ebro, siendo en las zonas de montaña en las que se sitúan la mayor concentración de puntos de muestreo con niveles altos-muy altos de calidad*.



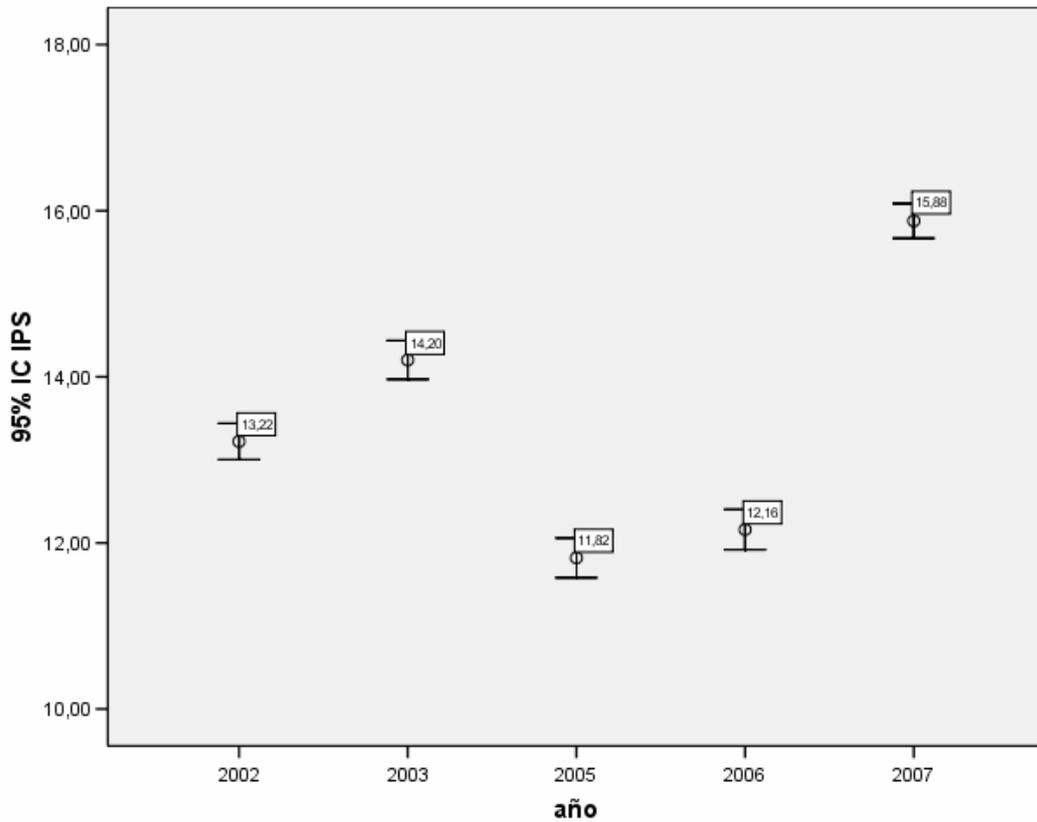
**Figura 3.1.2.1: Histograma de frecuencias del indicador IPS**



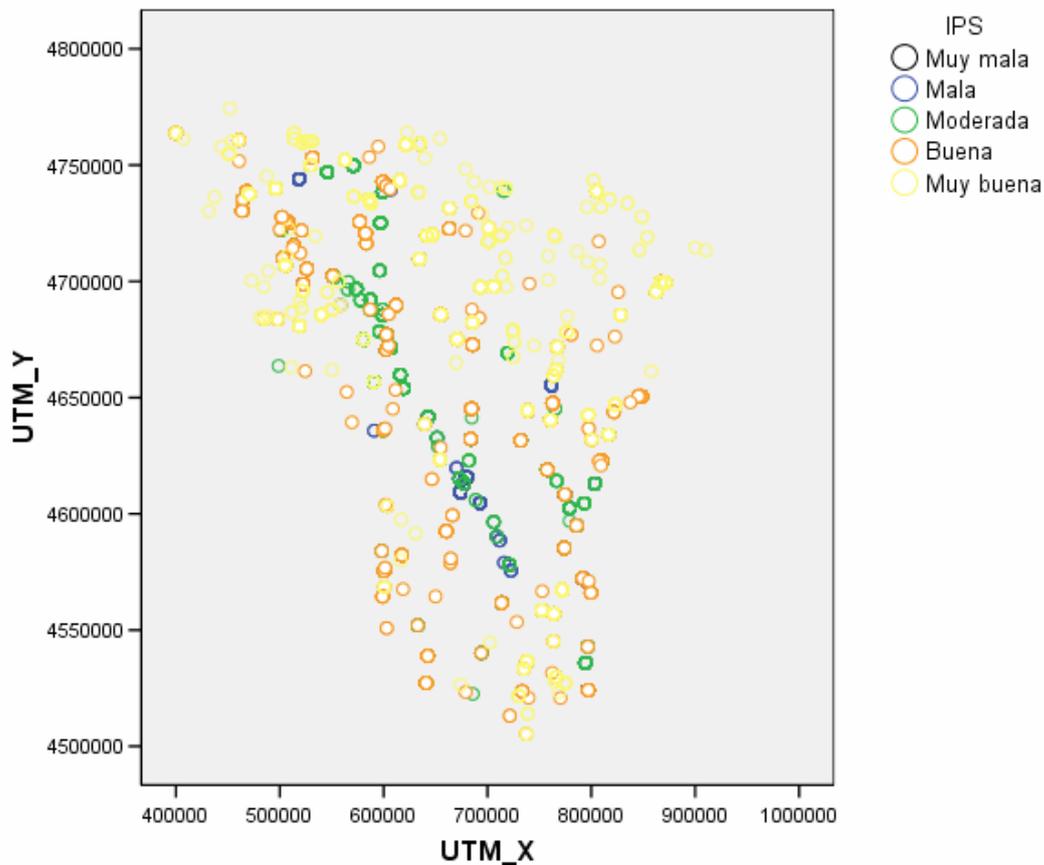
**Figura 3.1.2.2: Diagrama de cajas del indicador IPS**



**Figura 3.1.2.3: QQplot normal de indicador IPS**



**Figura 3.1.2.4: Evolución anual media del indicador IPS (encuadrados los valores medios del indicador en cada año)**



**Figura 3.1.2.5: Distribución geográfica de los niveles de calidad de las aguas medidos según el indicador IPS**  
 (Muy mala:  $IPS \leq 5$ ; mala:  $5 < IPS \leq 9$ ;  
 moderada:  $9 < IPS \leq 13$ ; buena:  $13 < IPS \leq 17$ ; muy buena:  $IPS > 17$ )

### 3.1.3. Análisis estadístico del indicador IVAM

Se observa, en primer lugar, la existencia de un *alto porcentaje de datos ausentes (86,00%) debido a que dicho indicador únicamente se observó en los años 2006 y 2007* (ver Tabla 3.1.3.1).

**Tabla 3.1.3.1: Datos ausentes en el indicador IVAM**

		Casos			
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
932	14,0%	5703	86,0%	6635	100,0%

En la Tabla 3.1.3.2 se muestran los valores de algunos estadísticos descriptivos de este indicador. Así mismo, en las Figuras 3.1.3.1 a 3.1.3.3 se muestran su histograma de frecuencias y sus diagramas de cajas y qqplot normal. Finalmente, en la Figura

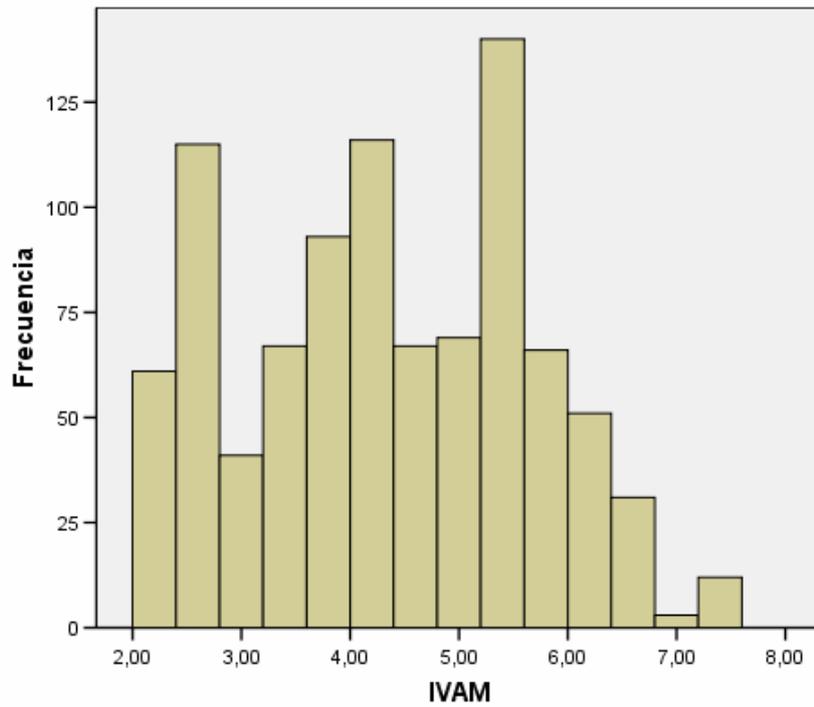
3.1.3.4 se muestra, en forma de diagrama de barras de error, la evolución anual de su valor medio mientras que en la Figura 3.1.3.5 se representa su distribución geográfica.

**Tabla 3.1.3.2: Análisis estadístico-descriptivo del indicador IVAM**

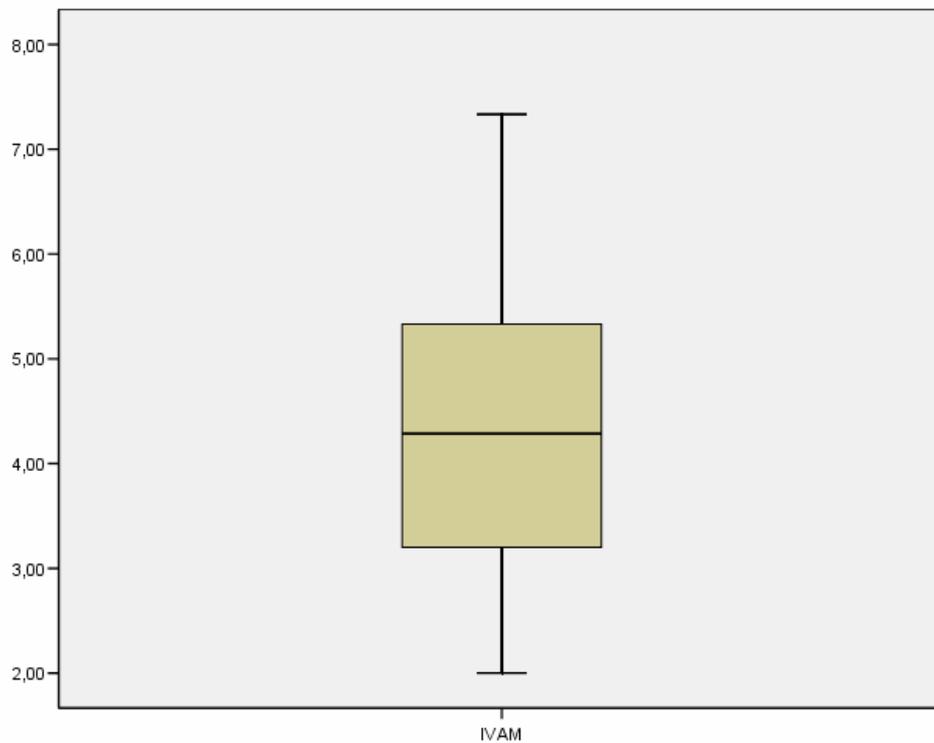
		Estadístico	Error típ.	
IVAM	Media	4,3098	,04299	
	Intervalo de confianza para la media al 95%	Límite inferior	4,2254	
		Límite superior	4,3942	
	Media recortada al 5%	4,3016		
	Mediana	4,2857		
	Varianza	1,722		
	Desv. típ.	1,31243		
	Mínimo	2,00		
	Máximo	7,33		
	Rango	5,33		
	Amplitud intercuartil	2,13		
	Asimetría	,040	,080	
	Curtosis	-,918	,160	

Se observa que, en general, *el nivel de calidad de las aguas de la cuenca ha sido moderado con un valor medio igual a 4,31 y un valor mediano igual a 4,29*. Dicho nivel no ha sido, sin embargo, constante a lo largo del tiempo siendo mejor en el año 2006, año en el que el nivel medio de la calidad del agua fue buena (ver Figura 3.1.3.4). Se aprecia, además, la existencia de una platicurtosis significativa (ver Tabla 3.1.3.2) debido a la existencia de varias modas en la distribución, algunas de ellas en la zona de valores bajos y otras en la de valores altos de la distribución (ver Figura 3.1.3.1) lo cual pone de manifiesto la existencia de comportamientos claramente diferenciados, respecto a este indicador, en grupos de puntos de muestreo a lo largo de la cuenca. No se aprecia sin embargo, la existencia de puntos de muestreo con un comportamiento atípico (ver Figura 3.1.3.2).

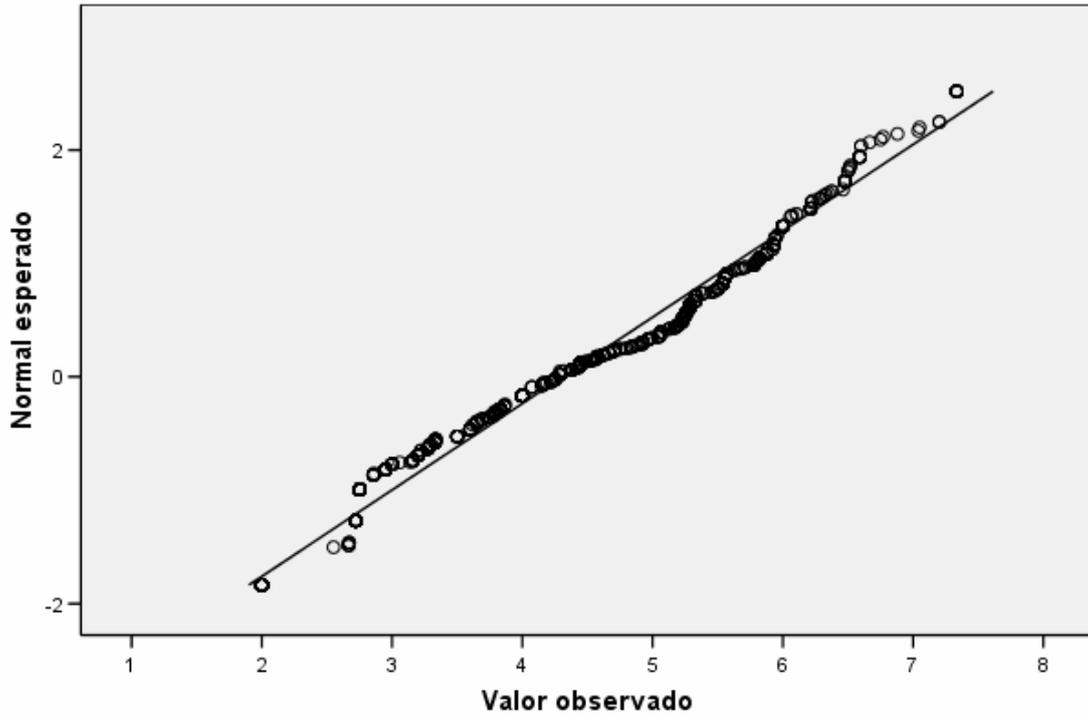
Si analizamos la distribución geográfica del indicador a lo largo de la cuenca (ver Figura 3.1.3.5) se observa que *las zonas con niveles de calidad más altos se sitúan en la zona Nordeste de la cuenca mientras que las zonas de peor calidad tiende a situarse en las zonas Suroeste y a lo largo del cauce del río Ebro*.



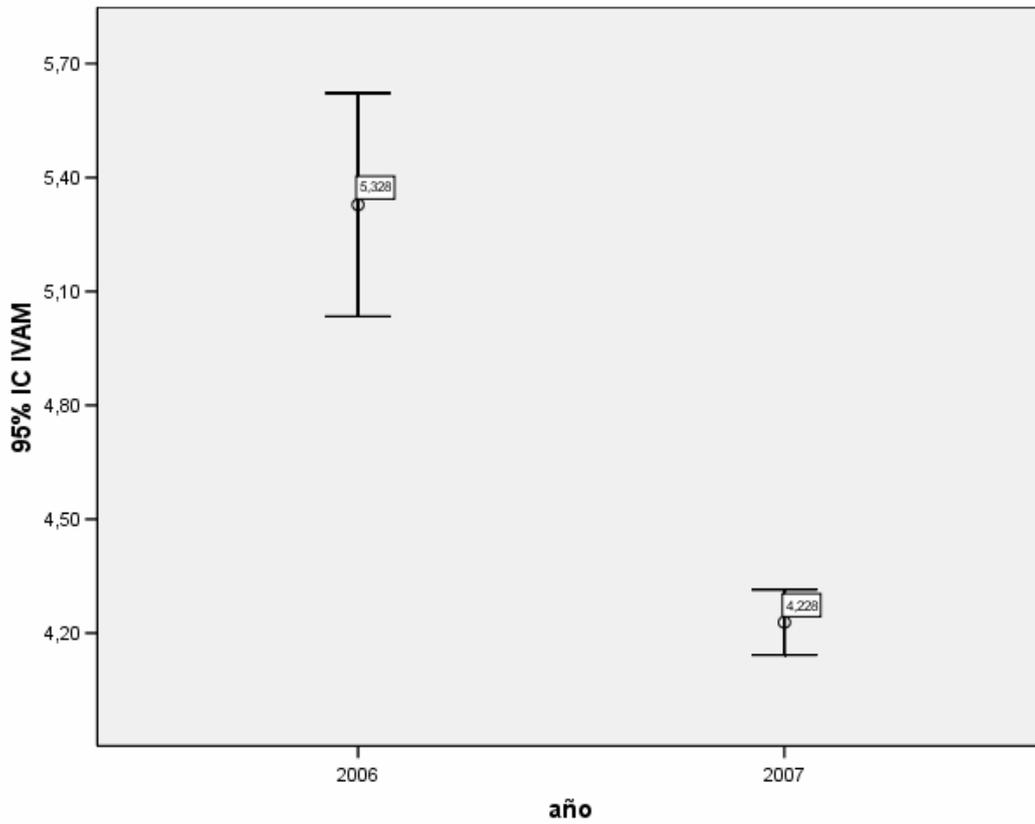
**Figura 3.1.3.1: Histograma de frecuencias del indicador IVAM**



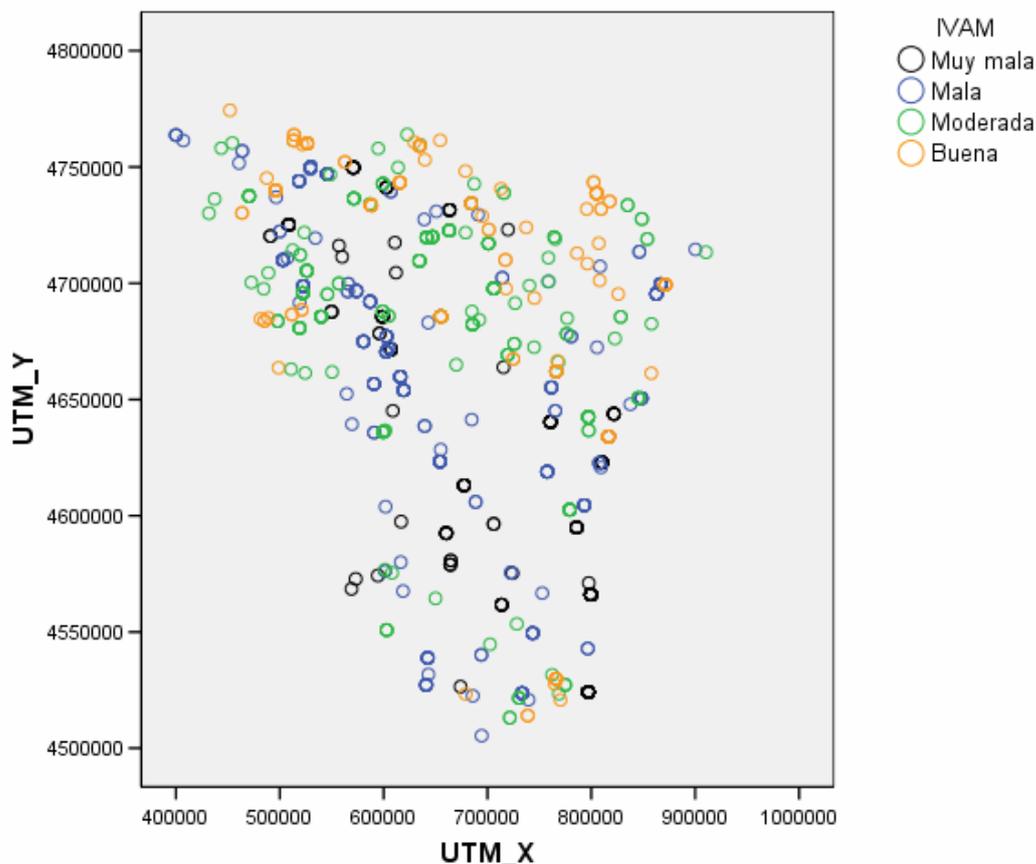
**Figura 3.1.3.2: Diagrama de cajas del indicador IVAM**



**Figura 3.1.3.3: QQplot normal de indicador IVAM**



**Figura 3.1.3.4: Evolución anual media del indicador IVAM (encuadrados los valores medios del indicador en cada año)**



**Figura 3.1.3.5: Distribución geográfica de los niveles de calidad de las aguas según el indicador IVAM**  
 (Muy mala:  $IVAM \leq 3,2$ ; mala:  $3,2 < IVAM \leq 4,5$ ; moderada:  $4,5 < IVAM \leq 5,7$ ; buena:  $5,7 < IVAM \leq 7,5$ ; muy buena:  $IVAM > 7,5$ )

### 3.1.4. Construcción del indicador de calidad conjunto

Dada la gran cantidad de datos ausentes existentes en los indicadores biológicos y con el fin de aprovechar al máximo la información disponible, se decidió construir un indicador de calidad conjunto que reflejara, con la mayor fiabilidad posible, la información contenida en los 3 indicadores biológicos anteriores. Dicho indicador se construyó de la siguiente manera:

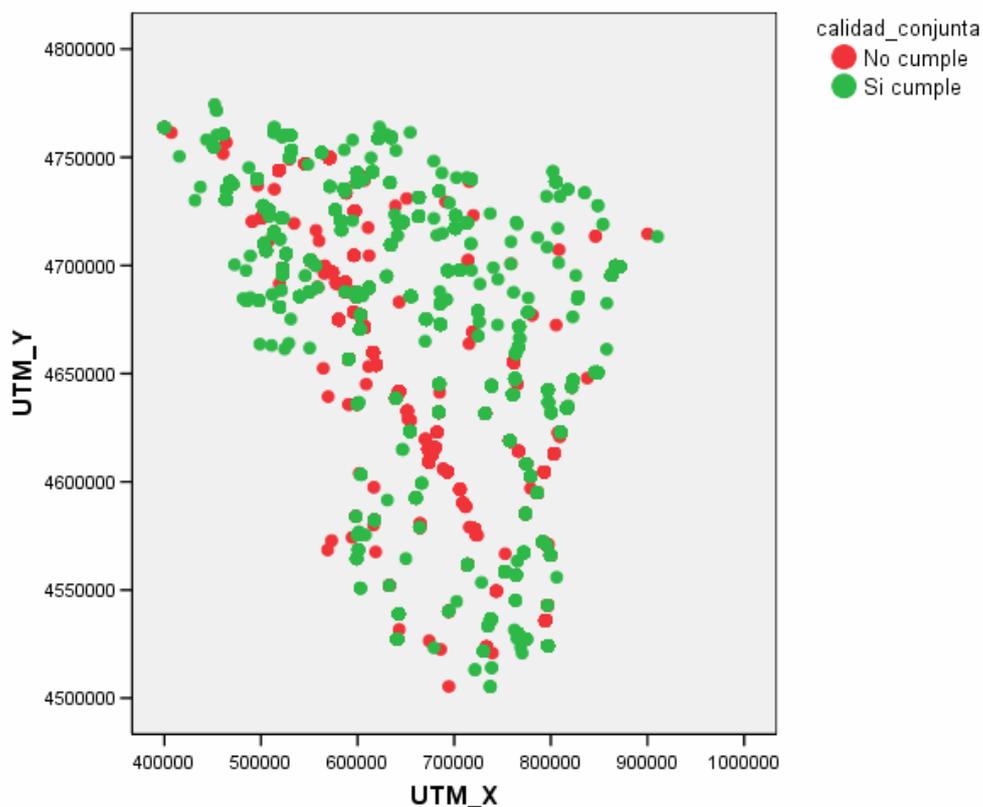
Diremos que un punto de muestreo cumple con los niveles de calidad mínimo exigibles desde un punto de vista biológico si verifica los requisitos listados en la columna umbrales de calidad de la Tabla 2.2 para todos los indicadores biológicos del estudio observados en dicho punto. En otro caso diremos que no cumple dichos requisitos, a no ser que no tenga ningún indicador observado, en cuyo caso el indicador de calidad conjunto se declarará ausente. Este indicador jugará el papel de variable dependiente en los modelos de regresión estimados en la sección 5 del trabajo.

En la Tabla 3.1.4.1 se presenta la distribución de frecuencias de dicho indicador. Se observa que *la mayor parte de los puntos de muestreo (53,10%) verifican los requisitos mínimos de calidad y que el porcentaje de datos ausentes (1,20%) se ha reducido considerablemente.*

**Tabla 3.1.4.1: Distribución de frecuencias del indicador de calidad conjunto**

		Frecuencia	Porcentaje	Porcentaje válido
Válidos	No cumple	3033	45,7	46,2
	Si cumple	3525	53,1	53,8
	Total	6558	98,8	100,0
Perdidos	Sistema	77	1,2	
Total		6635	100,0	

Si analizamos la distribución geográfica de dicho indicador (ver Figura 3.1.4.1) se puede apreciar que *los puntos de muestreo que no verifican los requisitos de calidad tienden a situarse a lo largo del cauce del río Ebro* resultado ya puesto de manifiesto en el análisis estadístico individualizado de los 3 parámetros biológicos utilizados en el estudio.



**Figura 3.1.4.1: Distribución geográfica del indicador de calidad conjunto**

## 3.2. ANÁLISIS ESTADÍSTICO DE LOS PARÁMETROS FÍSICO-QUÍMICOS

En esta sección se realiza un análisis estadístico-descriptivo de las variables físico-químicas del estudio listadas en la Tabla 2.3, prestando especial atención al análisis de la normalidad de las mismas y a la identificación de observaciones atípicas. Si es necesario, se aplicarán transformaciones para aumentar su grado de normalidad. Así mismo, se realiza un estudio detallado de los datos ausentes analizando la presencia de patrones significativos con respecto al nivel de calidad conjunta, que pueda distorsionar los procesos de estimación y validación de los modelos de regresión logística utilizados para la determinación de los umbrales de calidad. Los resultados se presentan por separado para cada uno de los 21 parámetros físico-químicos.

### 3.2.1. Análisis estadístico de los indicadores generales

En este apartado presentamos el análisis estadístico-descriptivo de los indicadores generales temperatura del agua (TEM\_AG), temperatura del aire (TEM\_AI), índice pH (PH), conductividad (COND20), oxígeno disuelto en concentración (O2) y en % de saturación (O2%) y sólidos en suspensión (MES).

#### 3.2.1.1 Análisis estadístico de la temperatura del agua

En las Tablas 3.2.1.1.1 a 3.2.1.1.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Más concretamente, en la Tabla 3.2.1.1.1 se analiza el porcentaje de datos ausentes mientras que en las Tablas 3.2.1.1.2 y 3.2.1.1.3 se analiza la existencia de diferencias sistemáticas entre los datos ausentes y los presentes en esta variable, tanto por niveles de calidad conjunta como por años. Este tipo de estudio se ha realizado en todos los demás indicadores físico-químicos y, por dicha razón, no se hará más referencia a este hecho a lo largo del estudio.

Se observa (ver Tabla 3.2.1.1.1) que hay un bajísimo porcentaje de datos ausentes (1,1%) y una levísima tendencia a concentrarse entre los datos que no cumplen los niveles de calidad (50% frente a un 46,2%) (ver Tabla 3.2.1.1.2). Si se hace el estudio por años (ver Tabla 3.2.1.1.3) se observa que los años con un porcentaje mayor de datos ausentes fueron 2002 y 2005. Dado, sin embargo, el escaso número de casos con datos ausentes, concluimos que, para esta variable, *no se aprecian la existencia de sesgos de importancia debido a la falta de representatividad de la muestra analizada.*

**Tabla 3.2.1.1.1: Porcentaje de datos ausentes de la variable TEM\_AG**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6561	98,9%	74	1,1%	6635	100,0%

**Tabla 3.2.1.1.2: Estudio de los datos ausentes de la variable TEM\_AG por nivel de calidad**

			miss_TEMAG		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3488	37	3525
		% de calidad_conjunta	99,0%	1,0%	100,0%
		% de miss_TEMAG	53,8%	50,0%	53,8%
		Residuos tipificados	,0	-,4	
	No cumple	Recuento	2996	37	3033
		% de calidad_conjunta	98,8%	1,2%	100,0%
		% de miss_TEMAG	46,2%	50,0%	46,2%
		Residuos tipificados	-,1	,5	
Total	Recuento	6484	74	6558	
	% de calidad_conjunta	98,9%	1,1%	100,0%	
	% de miss_TEMAG	100,0%	100,0%	100,0%	

**Tabla 3.2.1.1.3: Estudio de los datos ausentes de la variable TEM\_AG por años**

			miss_TEMAG		Total
			Presente	Ausente	
año	2002	Recuento	1307	34	1341
		% de año	97,5%	2,5%	100,0%
		% de miss_TEMAG	19,9%	45,9%	20,2%
		Residuos tipificados	-,5	4,9	
	2003	Recuento	992	7	999
		% de año	99,3%	,7%	100,0%
		% de miss_TEMAG	15,1%	9,5%	15,1%
		Residuos tipificados	,1	-1,2	
	2004	Recuento	637	4	641
		% de año	99,4%	,6%	100,0%
		% de miss_TEMAG	9,7%	5,4%	9,7%
		Residuos tipificados	,1	-1,2	
2005	Recuento	1298	14	1312	
	% de año	98,9%	1,1%	100,0%	
	% de miss_TEMAG	19,8%	18,9%	19,8%	
	Residuos tipificados	,0	-,2		
2006	Recuento	1275	8	1283	
	% de año	99,4%	,6%	100,0%	
	% de miss_TEMAG	19,4%	10,8%	19,3%	
	Residuos tipificados	,2	-1,7		
2007	Recuento	1052	7	1059	
	% de año	99,3%	,7%	100,0%	
	% de miss_TEMAG	16,0%	9,5%	16,0%	
	Residuos tipificados	,1	-1,4		
Total	Recuento	6561	74	6635	
	% de año	98,9%	1,1%	100,0%	
	% de miss_TEMAG	100,0%	100,0%	100,0%	

En las Tablas 3.2.1.1.4 y 3.2.1.1.5 así como en las Figuras 3.2.1.1.1 a 3.2.1.1.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. La temperatura media del agua en la cuenca es igual a 13,9° C y la mediana 13,5° C (ver Tabla 3.2.1.1.4). Los contrastes de normalidad rechazan la hipótesis de normalidad (ver Tablas 3.2.1.1.4 y 3.2.1.1.5 así como las Figuras 3.2.1.1.1 a 3.2.1.1.3) debido, fundamentalmente, al carácter platicúrtico de su distribución. Este carácter se debe, fundamentalmente, a la distribución uniforme de la temperatura del agua a lo largo de la cuenca *sin que se aprecien patrones de concentración geográfica dignos de mención* (ver Figura 3.2.1.1.4).

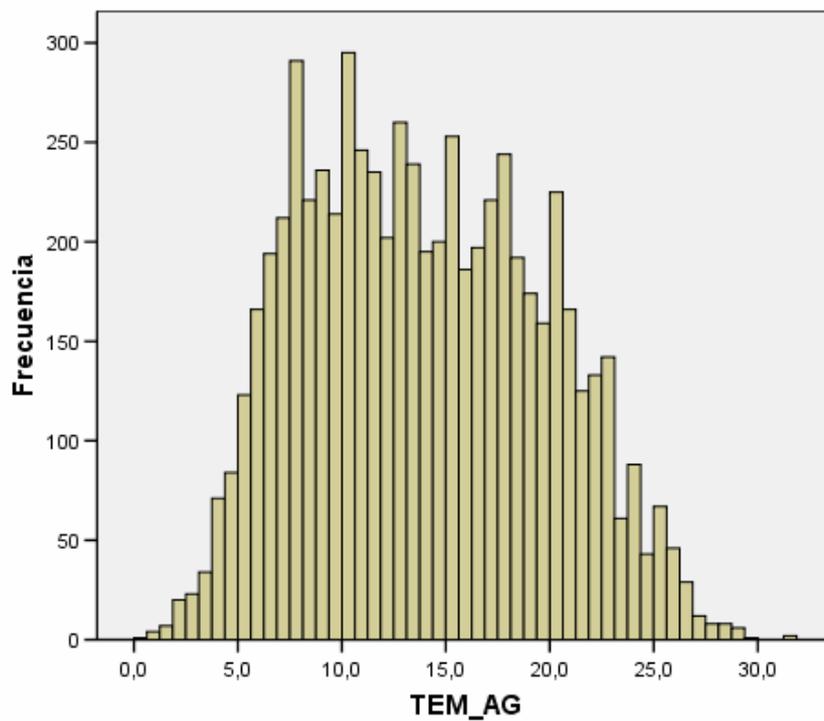
Analizando, sin embargo, su histograma (Figura 3.2.1.1.1) y su qqplot (Figura 3.2.1.1.3) no se aprecia la existencia de multimodalidad ni un alejamiento notable de dicha hipótesis además de no presentar observaciones atípicas (Figura 3.2.1.1.2) por lo que no aplicamos ningún tipo de transformación a esta variable.

**Tabla 3.2.1.1.4: Análisis estadístico-descriptivo de la variable TEM\_AG**

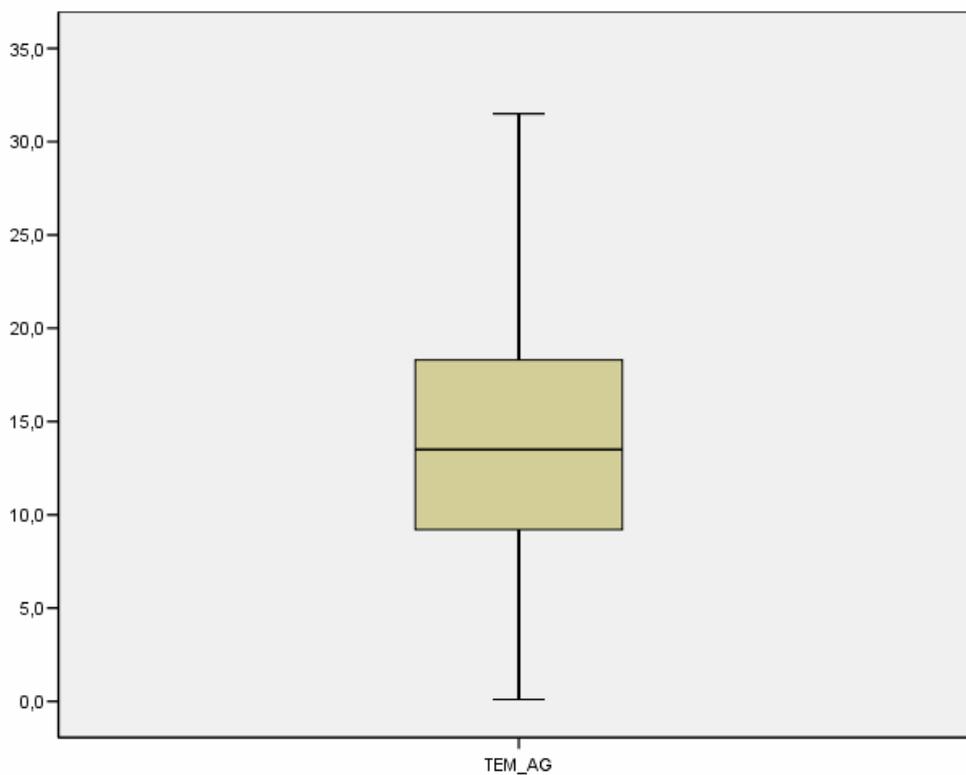
		Estadístico	Error típ.
TEM_AG	Media	13,900	,0707
	Intervalo de confianza para la media al 95%	Límite inferior 13,761	
		Límite superior 14,038	
	Media recortada al 5%	13,810	
	Mediana	13,500	
	Varianza	32,752	
	Desv. típ.	5,7230	
	Mínimo	,1	
	Máximo	31,5	
	Rango	31,4	
	Amplitud intercuartil	9,1	
	Asimetría	,207	,030
	Curtosis	-,787	,060

**Tabla 3.2.1.1.5: Pruebas de normalidad de la variable TEM\_AG**

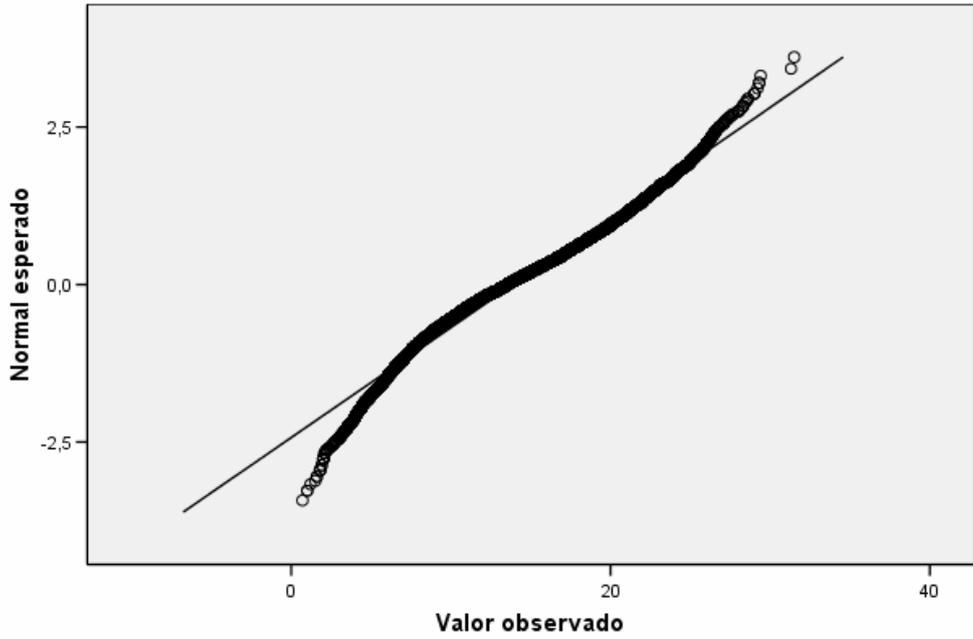
	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
TEM_AG	,056	6561	,000



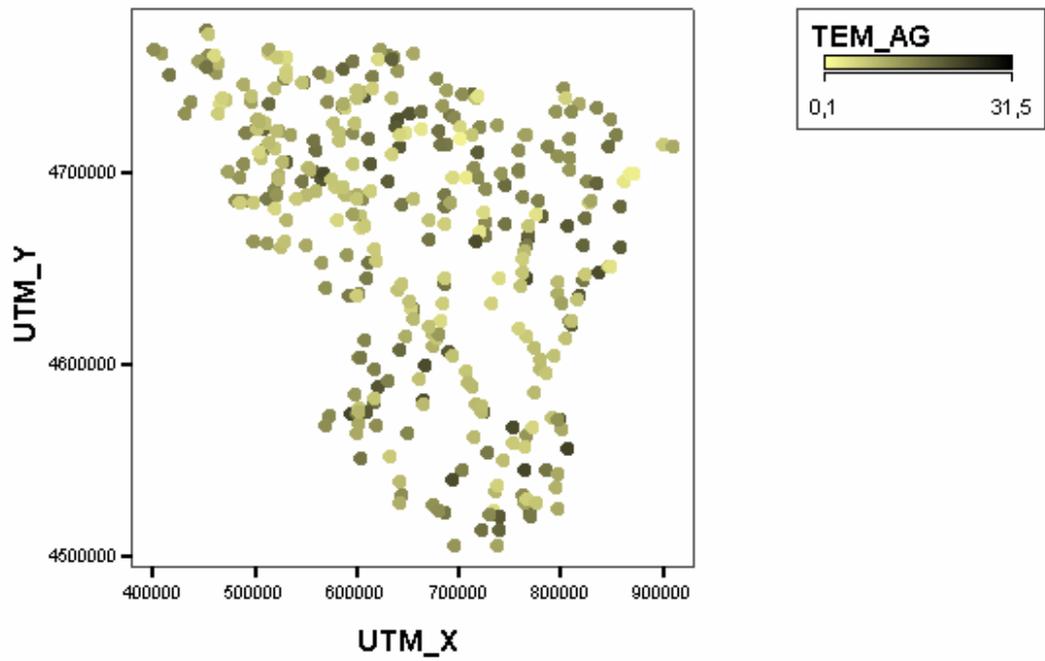
**Figura 3.2.1.1.1: Histograma de la variable TEM\_AG**



**Figura 3.2.1.1.2: Diagrama de cajas de la variable TEM\_AG**



**Figura 3.2.1.1.3: QQplot normal de la variable TEM\_AG**



**Figura 3.2.1.1.4: Distribución geográfica de los niveles de TEM\_AG**

### 3.2.1.2 Análisis estadístico de la temperatura del aire

En las Tablas 3.2.1.2.1 a 3.2.1.2.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa que hay un bajo porcentaje (6,4%) de datos de este tipo y una tendencia fuerte a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 72,6% de los datos que tienen valores ausentes en la variable TEM\_AI no cumplen los estándares de calidad frente a un 44,4% de los datos observados (ver Tabla 3.2.1.2.2). Si se hace el estudio por años (ver Tabla 3.2.1.2.3) se observa que los años con un porcentaje mayor de datos ausentes fueron 2004, 2005 y 2007.

**Tabla 3.2.1.2.1: Porcentaje de datos ausentes de la variable TEM\_AI**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6212	93,6%	423	6,4%	6635	100,0%

**Tabla 3.2.1.2.2: Estudio de los datos ausentes de la variable TEM\_AI por nivel de calidad**

			miss_TEMAI		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3409	116	3525
		% de calidad_conjunta	96,7%	3,3%	100,0%
		% de miss_TEMAI	55,6%	27,4%	53,8%
		Residuos tipificados	1,9	-7,4	
	No cumple	Recuento	2726	307	3033
		% de calidad_conjunta	89,9%	10,1%	100,0%
		% de miss_TEMAI	44,4%	72,6%	46,2%
		Residuos tipificados	-2,1	8,0	
Total	Recuento	6135	423	6558	
	% de calidad_conjunta	93,5%	6,5%	100,0%	
	% de miss_TEMAI	100,0%	100,0%	100,0%	

Dado, si embargo, el escaso número de casos con datos ausentes, concluimos que *no se aprecia una falta de representatividad de la muestra analizada que sea relevante.*

**Tabla 3.2.1.2.3: Estudio de los datos ausentes de la variable TEM\_AI por años**

			miss_TEMAI		Total
			Presente	Ausente	
año 2002	Recuento		1306	35	1341
	% de año		97,4%	2,6%	100,0%
	% de miss_TEMAI		21,0%	8,3%	20,2%
	Residuos tipificados		1,4	-5,5	
2003	Recuento		988	11	999
	% de año		98,9%	1,1%	100,0%
	% de miss_TEMAI		15,9%	2,6%	15,1%
	Residuos tipificados		1,7	-6,6	
2004	Recuento		573	68	641
	% de año		89,4%	10,6%	100,0%
	% de miss_TEMAI		9,2%	16,1%	9,7%
	Residuos tipificados		-1,1	4,2	
2005	Recuento		1164	148	1312
	% de año		88,7%	11,3%	100,0%
	% de miss_TEMAI		18,7%	35,0%	19,8%
	Residuos tipificados		-1,8	7,0	
2006	Recuento		1218	65	1283
	% de año		94,9%	5,1%	100,0%
	% de miss_TEMAI		19,6%	15,4%	19,3%
	Residuos tipificados		,5	-1,9	
2007	Recuento		963	96	1059
	% de año		90,9%	9,1%	100,0%
	% de miss_TEMAI		15,5%	22,7%	16,0%
	Residuos tipificados		-,9	3,5	
Total	Recuento		6212	423	6635
	% de año		93,6%	6,4%	100,0%
	% de miss_TEMAI		100,0%	100,0%	100,0%

En las Tablas 3.2.1.2.4 y 3.2.1.2.5 así como en las Figuras 3.2.1.2.1 a 3.2.1.2.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. La temperatura media del aire en la cuenca es igual a 16,36° C y la mediana 16,5° C (ver Tabla 3.2.1.2.4). Al igual que con la variable anterior, los contrastes de normalidad rechazan dicha hipótesis (ver Tablas 3.2.1.2.4 y 3.2.1.2.5) debido, nuevamente al carácter platicúrtico de su distribución. *La distribución de la temperatura del aire es uniforme en toda la cuenca, sin que se aprecien patrones geográficos dignos de interés* (ver Figura 3.2.1.2.4).

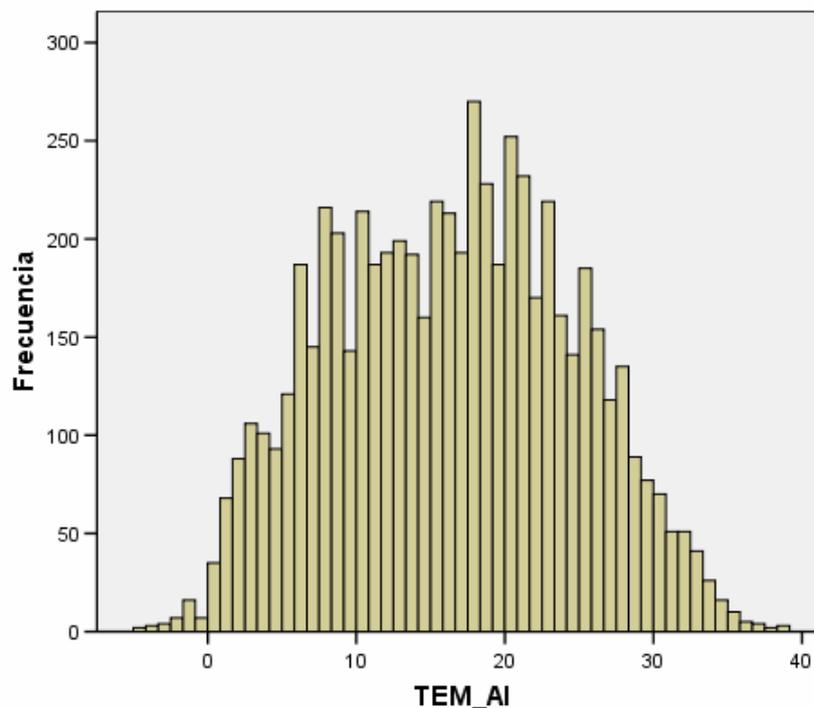
Analizando, sin embargo, su histograma (Figura 3.2.1.2.1) y su qqplot (Figura 3.2.1.2.3) no se aprecia la existencia de multimodalidad ni hay un alejamiento notable de dicha hipótesis así como no se manifiestan observaciones atípicas (Figura 3.2.1.2.2) por lo que no aplicamos ningún tipo de transformación a esta variable.

**Tabla 3.2.1.2.4: Análisis estadístico-descriptivo de la variable TEM\_AI**

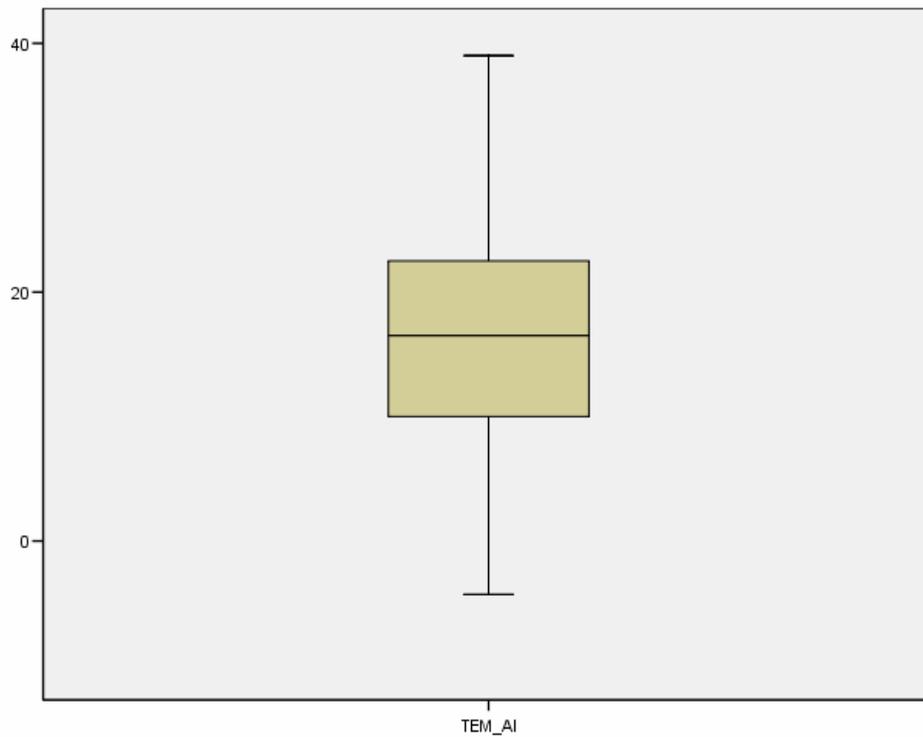
		Estadístico	Error típ.
TEM_AI	Media	16,36	,103
	Intervalo de confianza para la media al 95%	Límite inferior Límite superior	
		16,16 16,56	
	Media recortada al 5%	16,33	
	Mediana	16,50	
	Varianza	66,165	
	Desv. típ.	8,134	
	Mínimo	-4	
	Máximo	39	
	Rango	43	
	Amplitud intercuartil	13	
	Asimetría	,032	,031
	Curtosis	-,740	,062

**Tabla 3.2.1.2.5: Pruebas de normalidad de la variable TEM\_AI**

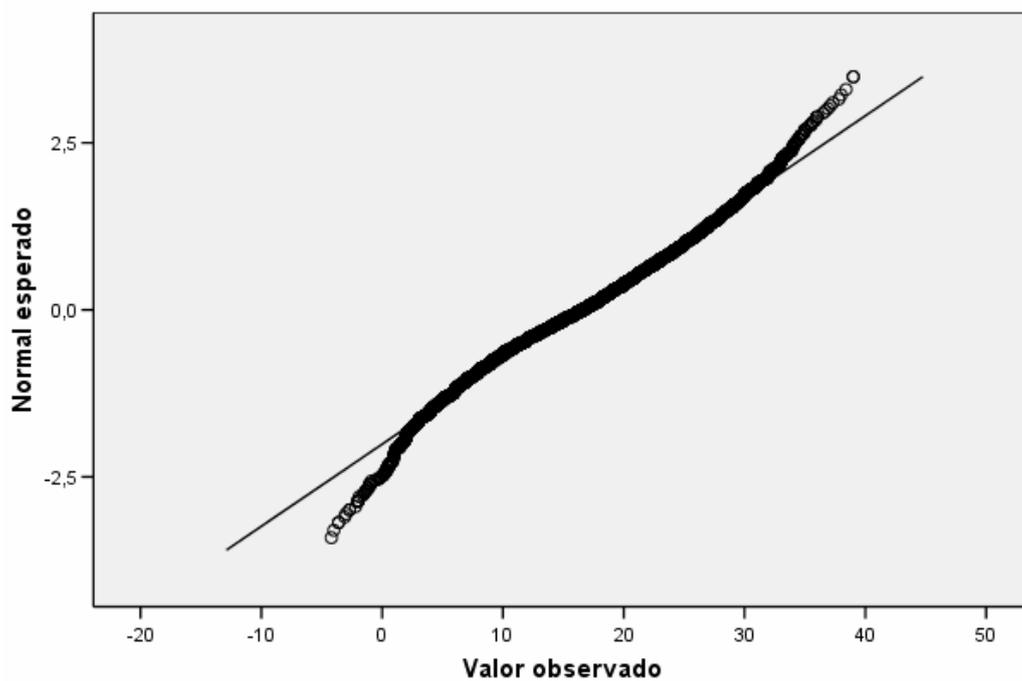
	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
TEM_AI	,043	6212	,000



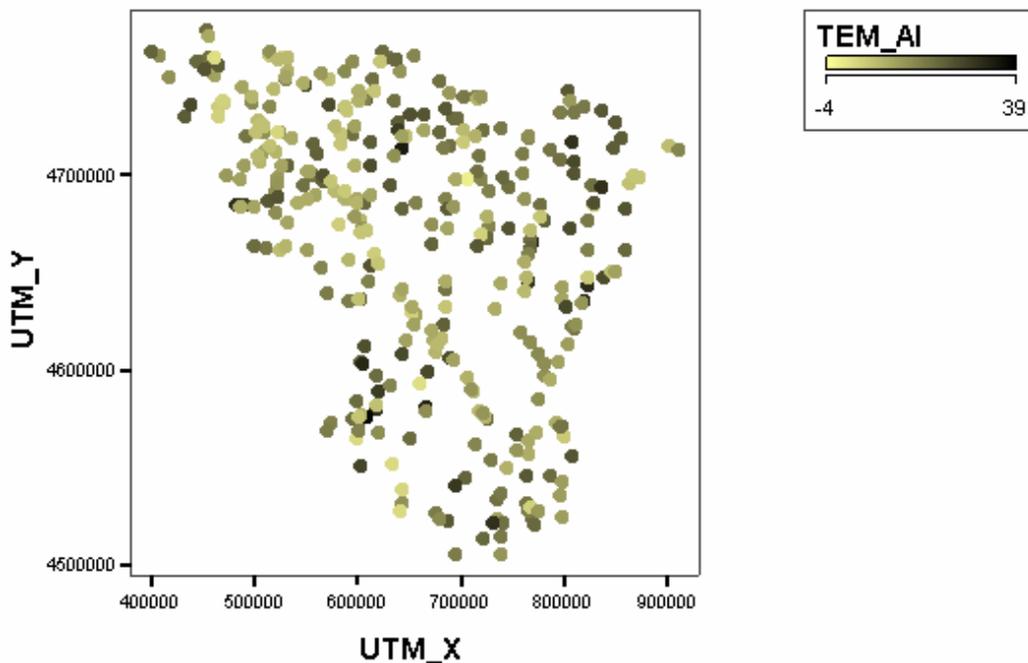
**Figura 3.2.1.2.1: Histograma de la variable TEM\_AI**



**Figura 3.2.1.2.2: Diagrama de cajas de la variable TEM\_AI**



**Figura 3.2.1.2.3: QQplot normal de la variable TEM\_AI**



**Figura 3.2.1.2.4: Distribución geográfica de los niveles de TEM\_AI**

### 3.2.1.3 Análisis estadístico del índice pH

En las Tablas 3.2.1.3.1 a 3.2.1.3.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.1.3.1) que hay un bajísimo porcentaje (1,1%) y un reparto equitativo (alrededor del 50%) entre los que cumplen y no cumplen los niveles de calidad (ver Tabla 3.2.1.3.2). Si se hace el estudio por años (ver Tabla 3.2.1.3.3) se observa que el año con un porcentaje mayor de datos ausentes fue 2002.

**Tabla 3.2.1.3.1: Porcentaje de datos ausentes de la variable PH**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6559	98,9%	76	1,1%	6635	100,0%

Dado el escaso número de casos con datos ausentes, concluimos que *no se aprecia una falta de representatividad de la muestra analizada que sea relevante.*

**Tabla 3.2.1.3.2: Estudio de los datos ausentes de la variable PH por nivel de calidad**

			miss_PH		Total
			Presente	Ausente	
calidad_ conjunta	Cumple	Recuento	3486	39	3525
		% de calidad_conjunta	98,9%	1,1%	100,0%
		% de miss_PH	53,8%	51,3%	53,8%
		Residuos tipificados	,0	-,3	
	No cumple	Recuento	2996	37	3033
		% de calidad_conjunta	98,8%	1,2%	100,0%
		% de miss_PH	46,2%	48,7%	46,2%
		Residuos tipificados	,0	,3	
Total		Recuento	6482	76	6558
		% de calidad_conjunta	98,8%	1,2%	100,0%
		% de miss_PH	100,0%	100,0%	100,0%

**Tabla 3.2.1.3.3: Estudio de los datos ausentes de la variable PH por años**

			miss_PH		Total
			Presente	Ausente	
año	2002	Recuento	1306	35	1341
		% de año	97,4%	2,6%	100,0%
		% de miss_PH	19,9%	46,1%	20,2%
		Residuos tipificados	-,5	5,0	
	2003	Recuento	992	7	999
		% de año	99,3%	,7%	100,0%
		% de miss_PH	15,1%	9,2%	15,1%
		Residuos tipificados	,1	-1,3	
	2004	Recuento	637	4	641
		% de año	99,4%	,6%	100,0%
		% de miss_PH	9,7%	5,3%	9,7%
		Residuos tipificados	,1	-1,2	
	2005	Recuento	1298	14	1312
		% de año	98,9%	1,1%	100,0%
		% de miss_PH	19,8%	18,4%	19,8%
		Residuos tipificados	,0	-,3	
	2006	Recuento	1274	9	1283
		% de año	99,3%	,7%	100,0%
		% de miss_PH	19,4%	11,8%	19,3%
		Residuos tipificados	,2	-1,5	
	2007	Recuento	1052	7	1059
		% de año	99,3%	,7%	100,0%
		% de miss_PH	16,0%	9,2%	16,0%
		Residuos tipificados	,2	-1,5	
Total		Recuento	6559	76	6635
		% de año	98,9%	1,1%	100,0%
		% de miss_PH	100,0%	100,0%	100,0%

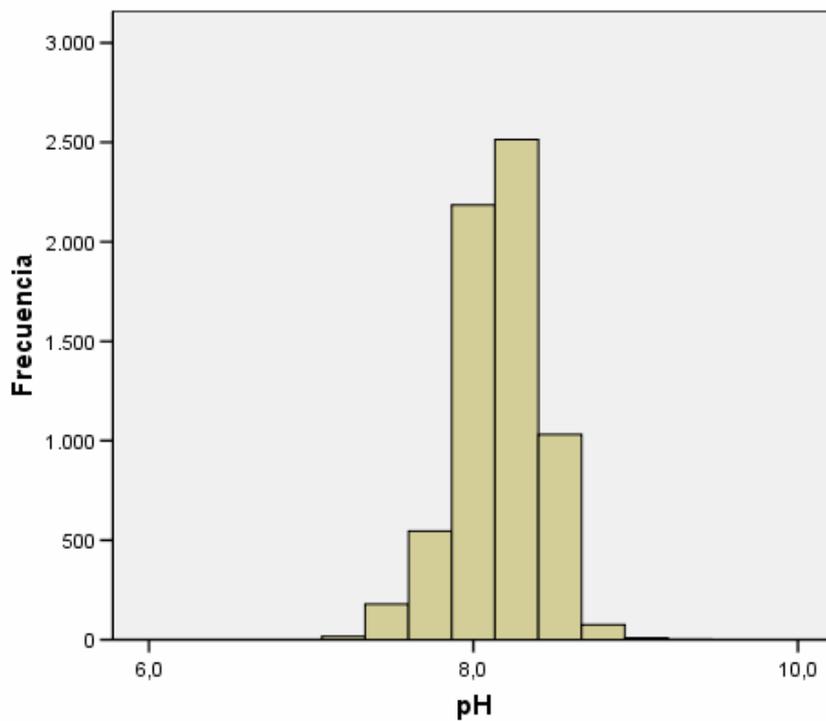
En las Tablas 3.2.1.3.4 y 3.2.1.3.5 así como en las Figuras 3.2.1.3.1 a 3.2.1.3.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El pH medio en las aguas de la cuenca es igual a 8,15 y la mediana 8,20 (ver Tabla 3.2.1.3.4). Se rechaza la hipótesis de normalidad (ver Tablas 3.2.1.3.4 y 3.2.1.3.5 y Figuras 3.2.1.3.1 a 3.2.1.3.3) debido a la excesiva leptocurtosis de la distribución. Dicha leptocurtosis se debe a la presencia de puntos de muestreo con valores del pH significativamente más altos y más bajos que el resto (ver Figura 3.2.1.3.2) apreciándose, además, una *tendencia de las zonas de montaña (norte de la cuenca) a tener un valor del pH más alto que el resto* (ver Figura 3.2.1.3.4). Dado que no se aprecia un problema de asimetría notable y que ninguna de las transformaciones utilizadas habitualmente para reducir la leptocurtosis tuvo excesivo éxito se optó por no transformar esta variable.

**Tabla 3.2.1.3.4: Análisis estadístico-descriptivo de la variable PH**

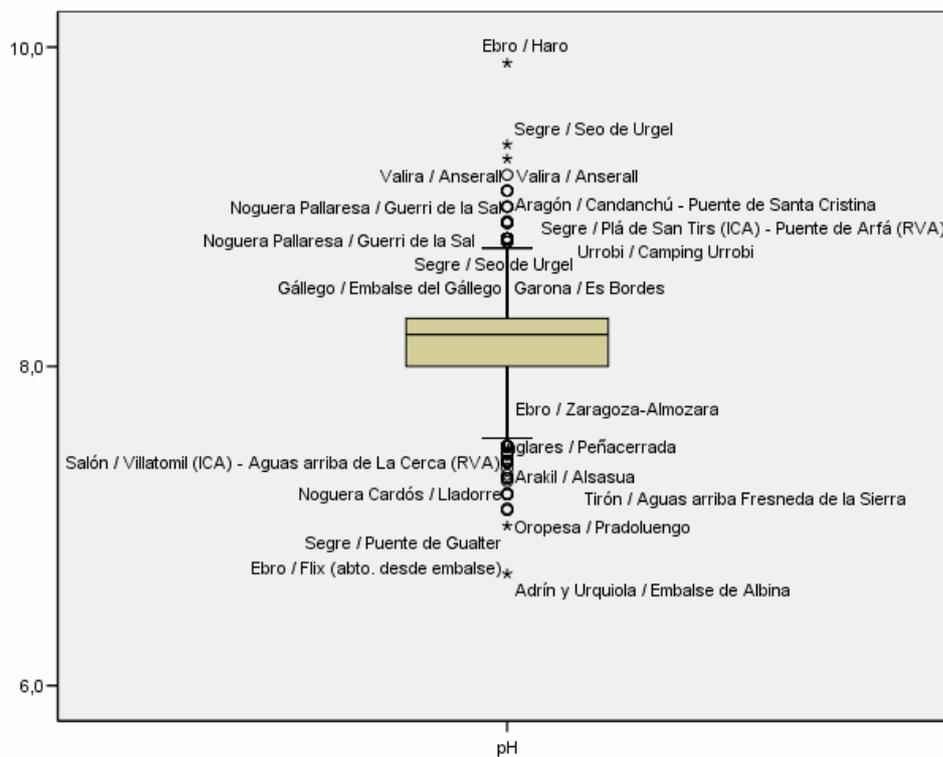
		Estadístico	Error típ.	
pH	Media	8,152	,0030	
	Intervalo de confianza para la media al 95%	Límite inferior	8,146	
		Límite superior	8,158	
	Media recortada al 5%	8,158		
	Mediana	8,200		
	Varianza	,060		
	Desv. típ.	,2450		
	Mínimo	6,7		
	Máximo	9,9		
	Rango	3,2		
	Amplitud intercuartil	,3		
	Asimetría	-,384	,030	
	Curtosis	1,500	,060	

**Tabla 3.2.1.3.5: Pruebas de normalidad de la variable PH**

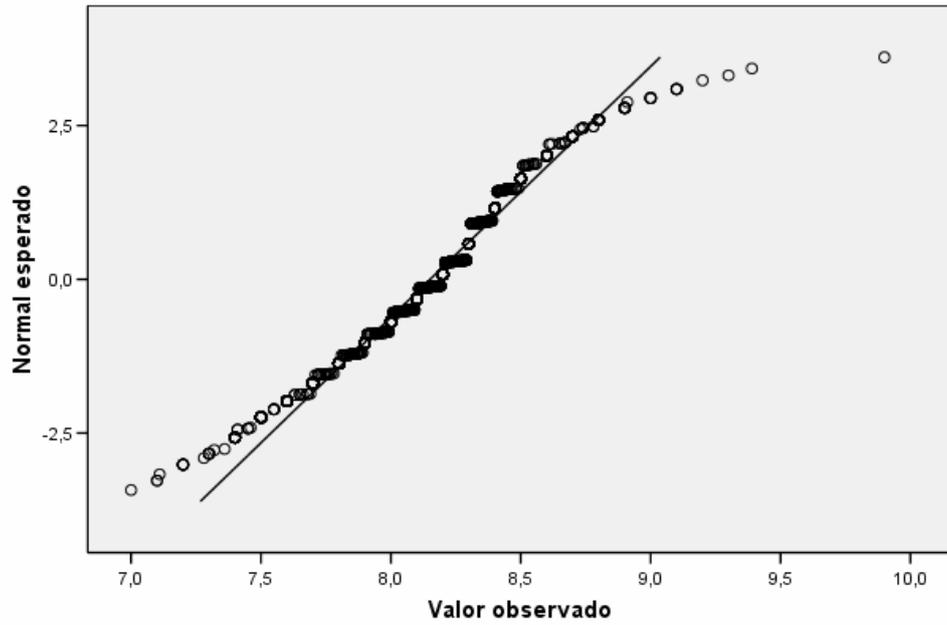
	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
pH	,118	6559	,000



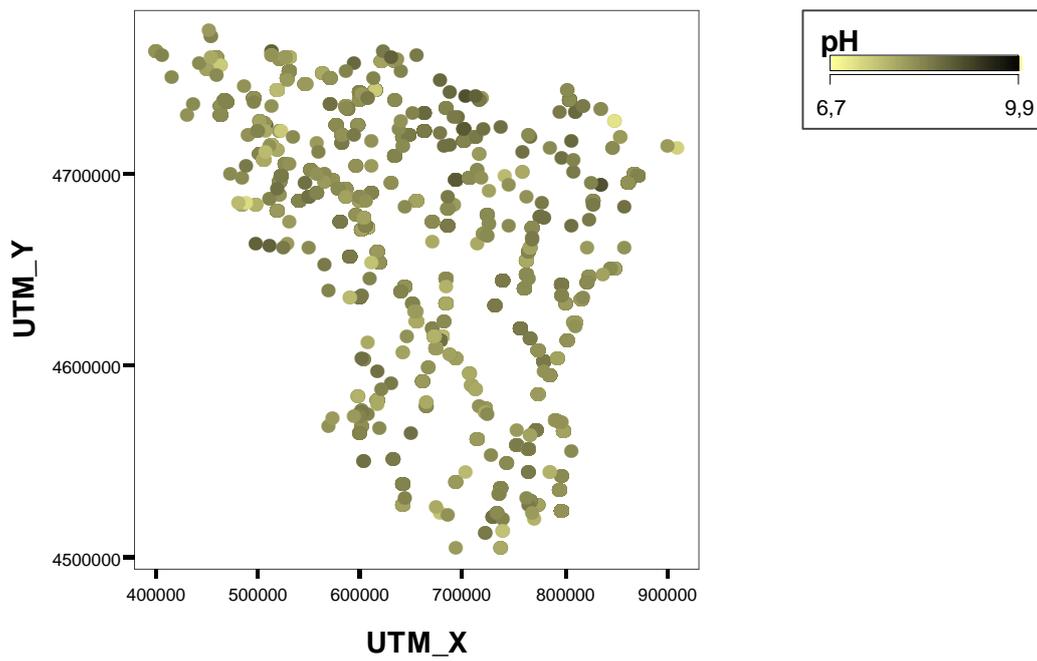
**Figura 3.2.1.3.1: Histograma de la variable PH**



**Figura 3.2.1.3.2: Diagrama de cajas de la variable PH**



**Figura 3.2.1.3.3: QQplot normal de la variable PH**



**Figura 3.2.1.3.4: Distribución geográfica de los niveles de pH**

### 3.2.1.4 Análisis estadístico de la conductividad a 20°C

En las Tablas 3.2.1.4.1 a 3.2.1.4.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a este indicador. Se observa (ver Tabla 3.2.1.4.1) que hay un bajísimo porcentaje (1,2%) y un reparto equitativo (alrededor del 50%) entre los que cumplen y no cumplen los niveles de calidad (ver Tabla 3.2.1.4.2). Si se hace el estudio por años (ver Tabla 3.2.1.4.3) se observa que el año con un porcentaje mayor de datos ausentes fue 2002.

**Tabla 3.2.1.4.1: Porcentaje de datos ausentes de la variable COND20**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6557	98,8%	78	1,2%	6635	100,0%

**Tabla 3.2.1.4.2: Estudio de los datos ausentes de la variable COND20 por nivel de calidad**

			miss_COND20		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3482	43	3525
		% de calidad_conjunta	98,8%	1,2%	100,0%
		% de miss_COND20	53,7%	55,1%	53,8%
		Residuos tipificados	,0	,2	
	No cumple	Recuento	2998	35	3033
		% de calidad_conjunta	98,8%	1,2%	100,0%
		% de miss_COND20	46,3%	44,9%	46,2%
		Residuos tipificados	,0	-,2	
Total	Recuento	6480	78	6558	
	% de calidad_conjunta	98,8%	1,2%	100,0%	
	% de miss_COND20	100,0%	100,0%	100,0%	

Dado el escaso número de casos con datos ausentes, concluimos que *no se aprecia una falta de representatividad de la muestra analizada que sea relevante.*

**Tabla 3.2.1.4.3: Estudio de los datos ausentes de la variable COND20 por años**

			miss_COND20		Total
			Presente	Ausente	
año	2002	Recuento	1303	38	1341
		% de año	97,2%	2,8%	100,0%
		% de miss_COND20	19,9%	48,7%	20,2%
		Residuos tipificados	-,6	5,6	
	2003	Recuento	989	10	999
		% de año	99,0%	1,0%	100,0%
		% de miss_COND20	15,1%	12,8%	15,1%
		Residuos tipificados	,1	-,5	
	2004	Recuento	637	4	641
		% de año	99,4%	,6%	100,0%
		% de miss_COND20	9,7%	5,1%	9,7%
		Residuos tipificados	,1	-1,3	
	2005	Recuento	1300	12	1312
		% de año	99,1%	,9%	100,0%
		% de miss_COND20	19,8%	15,4%	19,8%
		Residuos tipificados	,1	-,9	
	2006	Recuento	1276	7	1283
		% de año	99,5%	,5%	100,0%
		% de miss_COND20	19,5%	9,0%	19,3%
		Residuos tipificados	,2	-2,1	
	2007	Recuento	1052	7	1059
		% de año	99,3%	,7%	100,0%
		% de miss_COND20	16,0%	9,0%	16,0%
		Residuos tipificados	,2	-1,5	
Total		Recuento	6557	78	6635
		% de año	98,8%	1,2%	100,0%
		% de miss_COND20	100,0%	100,0%	100,0%

En las Tablas 3.2.1.4.4 y 3.2.1.4.5 así como en las Figuras 3.2.1.4.1 a 3.2.1.4.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de conductividad en las aguas de la cuenca es igual a 801,05  $\mu\text{S}/\text{cm}$  con un nivel mediano igual a 570  $\mu\text{S}/\text{cm}$  (ver Tabla 3.2.1.4.4). La distribución de esta variable presenta una asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.1.4.4) reflejando la existencia de puntos de muestreo con un alto nivel de conductividad (ver Figuras 3.2.1.4.1 y 3.2.1.4.2). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.1.4.5 y Figura 3.2.1.4.3). ***Realizando una transformación logarítmica (variable LCOND20) se aumenta significativamente el grado de normalidad de la variable, tal y como lo demuestran los resultados expuestos en las Figuras 3.2.1.4.4 a 3.2.1.4.6, por lo que en el resto del estudio será dicha variable la considerada como variable independiente.***

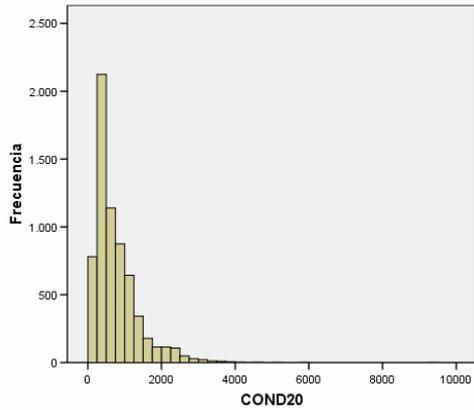
Analizando la distribución geográfica de LCOND20 (ver Figura 3.2.1.4.7) se observa que *los puntos de muestreo con un alto nivel de conductividad tienden a concentrarse a lo largo del cauce del río Ebro y en las zonas bajas de la cuenca.*

**Tabla 3.2.1.4.4: Análisis estadístico-descriptivo de las variables COND20 y LCOND20**

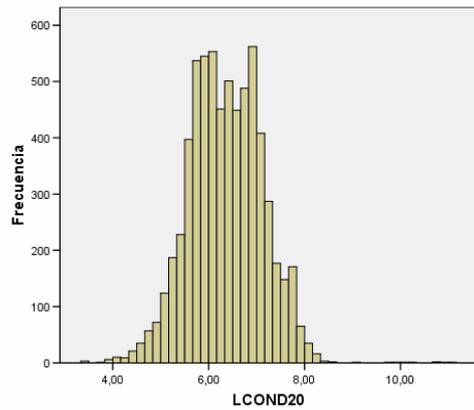
		Estadístico	Error típ.	
COND20	Media	801,05	18,832	
	Intervalo de confianza para la media al 95%	Límite inferior	764,13	
		Límite superior	837,96	
	Media recortada al 5%	690,48		
	Mediana	570,00		
	Varianza	2325308,4		
	Desv. típ.	1524,896		
	Mínimo	30		
	Máximo	62900		
	Rango	62870		
	Amplitud intercuartil	660		
	Asimetría	27,029	,030	
	Curtosis	916,776	,060	
	LCOND20	Media	6,3597	,00942
Intervalo de confianza para la media al 95%		Límite inferior	6,3412	
		Límite superior	6,3781	
Media recortada al 5%		6,3601		
Mediana		6,3456		
Varianza		,581		
Desv. típ.		,76239		
Mínimo		3,40		
Máximo		11,05		
Rango		7,65		
Amplitud intercuartil		1,08		
Asimetría		,115	,030	
Curtosis		,724	,060	

**Tabla 3.2.1.4.5: Pruebas de normalidad de las variables COND20 y LCOND20**

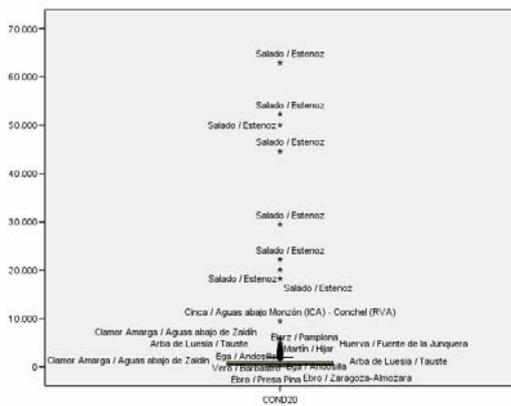
	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
COND20	,315	6557	,000
LCOND20	,031	6557	,000



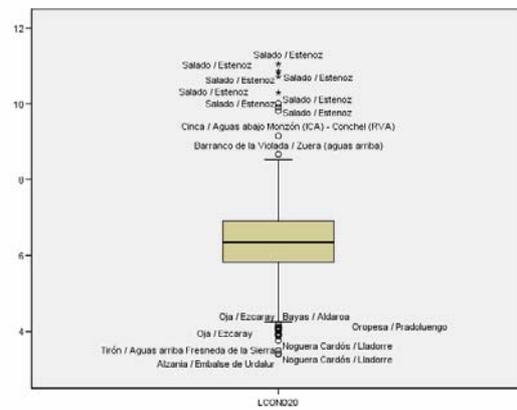
**Figura 3.2.1.4.1: Histograma de la variable COND20**



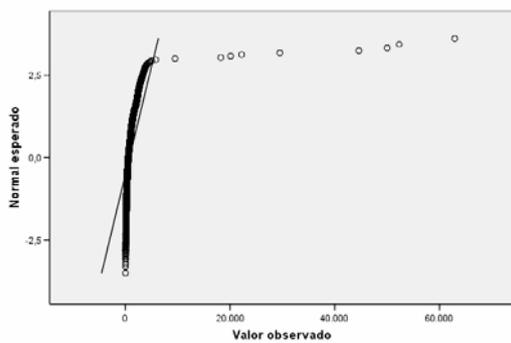
**Figura 3.2.1.4.4 : Histograma de la variable LCOND20**



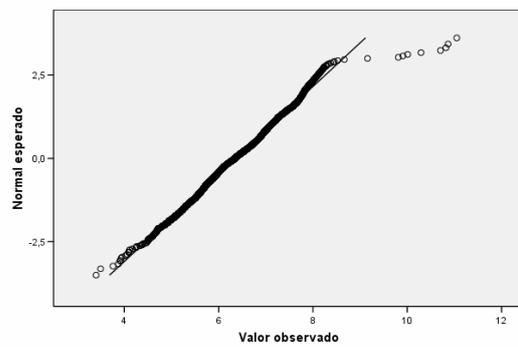
**Figura 3.2.1.4.2: Diagrama de cajas de la variable COND20**



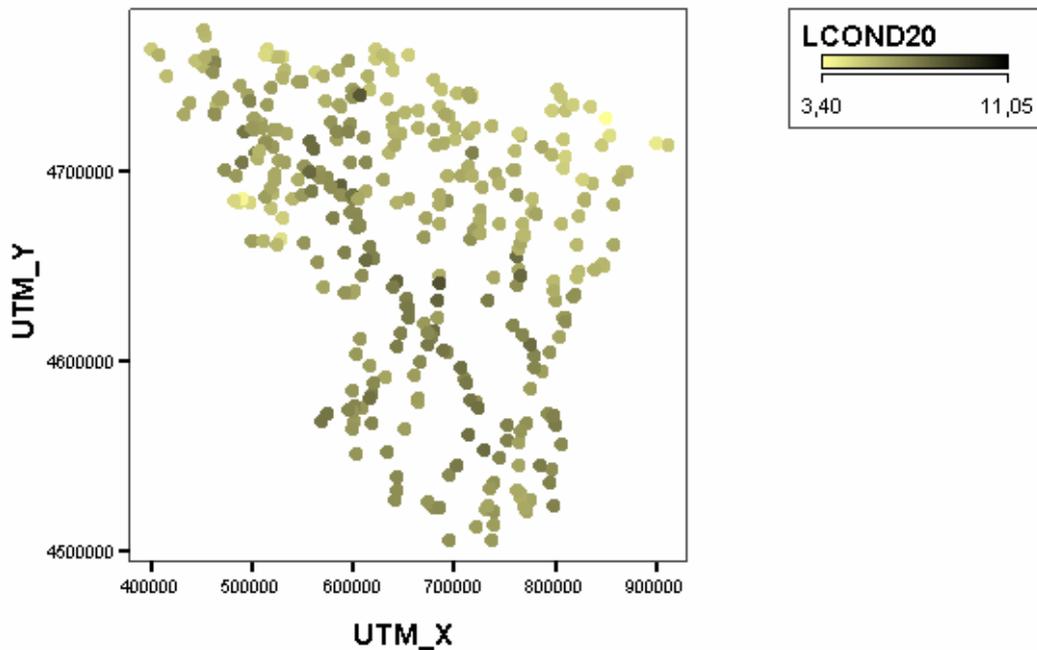
**Figura 3.2.1.4.5 : Diagrama de cajas de la variable LCOND20**



**Figura 3.2.1.4.3: QQplot normal de la variable COND20**



**Figura 3.2.1.4.6: QQplot normal de la variable LCOND20**



**Figura 3.2.1.4.7: Distribución geográfica de los niveles de LCOND20**

### 3.2.1.5 Análisis estadístico del oxígeno disuelto en concentración

En las Tablas 3.2.1.5.1 a 3.2.1.5.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.1.5.1) que hay un bajo porcentaje (6,5%) existiendo una fuerte tendencia a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 72,9% de los datos que tienen valores ausentes en la variable O2 no cumplen los estándares de calidad frente a un 44,4% de los datos observados (ver Tabla 3.2.1.5.2). Si se hace el estudio por años (ver Tabla 3.2.1.5.3) se observa que los años con un porcentaje mayor de datos ausentes fueron 2004, 2005 y 2007.

**Tabla 3.2.1.5.1: Porcentaje de datos ausentes de la variable O2**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6204	93,5%	431	6,5%	6635	100,0%

Sin embargo, dado el escaso número de casos con datos ausentes, concluimos que *no se aprecia una falta de representatividad de la muestra analizada que sea relevante*.

**Tabla 3.2.1.5.2: Estudio de los datos ausentes de la variable O2 por nivel de calidad**

			miss_O2		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3408	117	3525
		% de calidad_conjunta	96,7%	3,3%	100,0%
		% de miss_O2	55,6%	27,1%	53,8%
		Residuos tipificados	2,0	-7,5	
	No cumple	Recuento	2719	314	3033
		% de calidad_conjunta	89,6%	10,4%	100,0%
		% de miss_O2	44,4%	72,9%	46,2%
		Residuos tipificados	-2,2	8,1	
Total	Recuento	6127	431	6558	
	% de calidad_conjunta	93,4%	6,6%	100,0%	
	% de miss_O2	100,0%	100,0%	100,0%	

**Tabla 3.2.1.5.3: Estudio de los datos ausentes de la variable O2 por años**

			miss_O2		Total
			Presente	Ausente	
año 2002	Recuento	1304	37	1341	
	% de año	97,2%	2,8%	100,0%	
	% de miss_O2	21,0%	8,6%	20,2%	
	Residuos tipificados	1,4	-5,4		
2003	Recuento	991	8	999	
	% de año	99,2%	,8%	100,0%	
	% de miss_O2	16,0%	1,9%	15,1%	
	Residuos tipificados	1,9	-7,1		
2004	Recuento	577	64	641	
	% de año	90,0%	10,0%	100,0%	
	% de miss_O2	9,3%	14,8%	9,7%	
	Residuos tipificados	-,9	3,5		
2005	Recuento	1167	145	1312	
	% de año	88,9%	11,1%	100,0%	
	% de miss_O2	18,8%	33,6%	19,8%	
	Residuos tipificados	-1,7	6,5		
2006	Recuento	1217	66	1283	
	% de año	94,9%	5,1%	100,0%	
	% de miss_O2	19,6%	15,3%	19,3%	
	Residuos tipificados	,5	-1,9		
2007	Recuento	948	111	1059	
	% de año	89,5%	10,5%	100,0%	
	% de miss_O2	15,3%	25,8%	16,0%	
	Residuos tipificados	-1,3	5,1		
Total	Recuento	6204	431	6635	
	% de año	93,5%	6,5%	100,0%	
	% de miss_O2	100,0%	100,0%	100,0%	

En las Tablas 3.2.1.5.4 y 3.2.1.5.5 así como en las Figuras 3.2.1.5.1 a 3.2.1.5.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de concentración de oxígeno disuelto en las aguas de la cuenca es igual a 9,72 mg/L O<sub>2</sub> y el nivel mediano 9,80 mg/L O<sub>2</sub> (ver Tabla 3.2.1.5.4). La hipótesis de normalidad se rechaza debido, fundamentalmente, a la existencia de un elevado grado de leptocurtosis (ver Tablas 3.2.1.5.4 y 3.2.1.5.5 así como las Figuras 3.2.1.5.1 a 3.2.1.5.3). Ello es debido a la existencia de puntos de muestreo con valores significativamente más altos y más bajos de esta variable (ver Figura 3.2.1.5.2). Si se analiza la distribución geográfica de esta variable (ver Figura 3.2.1.5.4) se observa que *en la parte alta de la cuenca tienden a concentrarse los puntos de muestreo con mayores niveles de oxígeno, dándose la tendencia contraria en los puntos de muestreo de la parte baja.*

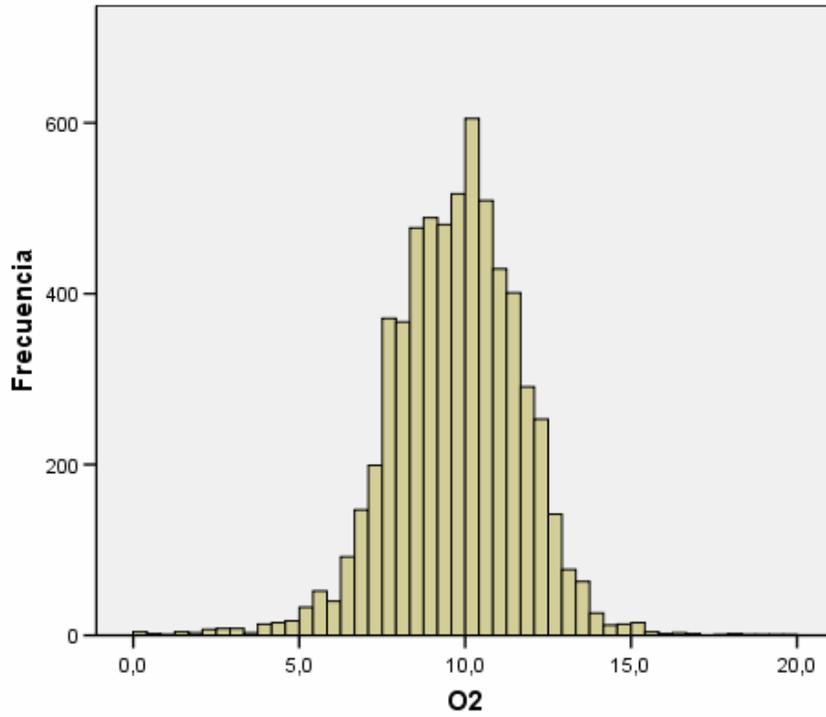
**Tabla 3.2.1.5.4: Análisis estadístico-descriptivo de la variable O<sub>2</sub>**

		Estadístico	Error típ.	
O <sub>2</sub>	Media	9,719	,0247	
	Intervalo de confianza para la media al 95%	Límite inferior	9,670	
		Límite superior	9,767	
	Media recortada al 5%	9,747		
	Mediana	9,800		
	Varianza	3,792		
	Desv. típ.	1,9473		
	Mínimo	,3		
	Máximo	19,7		
	Rango	19,4		
	Amplitud intercuartil	2,5		
	Asimetría	-,266	,031	
	Curtosis	1,582	,062	

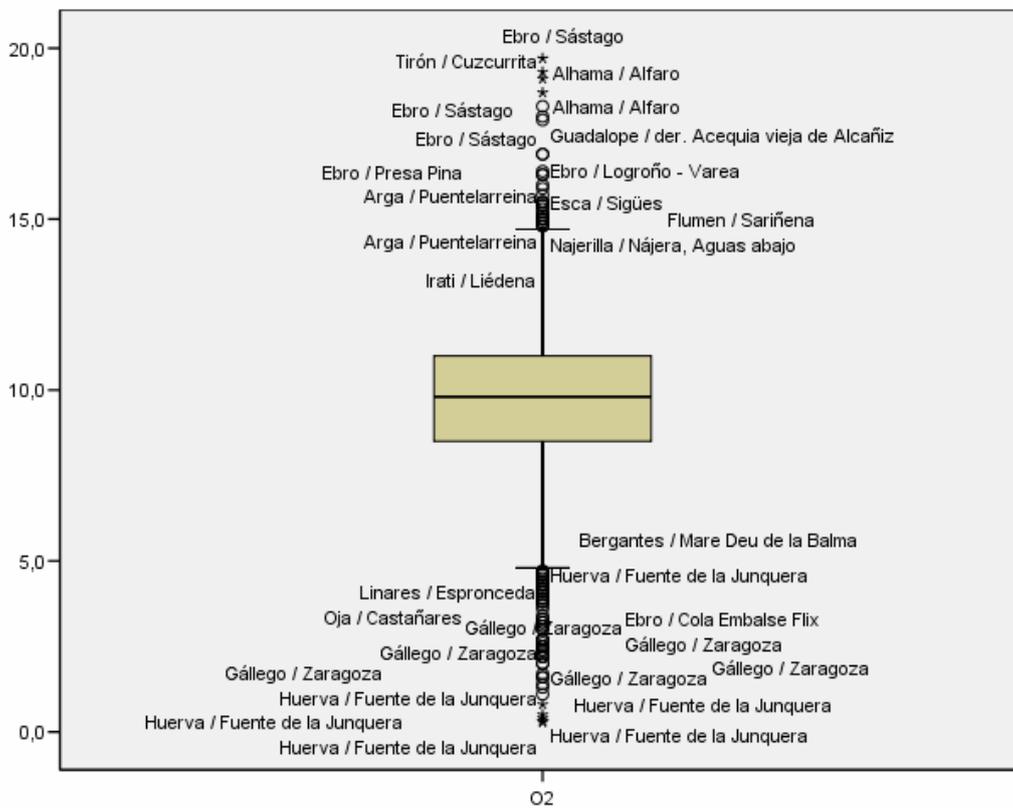
**Tabla 3.2.1.5.5: Pruebas de normalidad de la variable O<sub>2</sub>**

	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
O <sub>2</sub>	,027	6204	,000

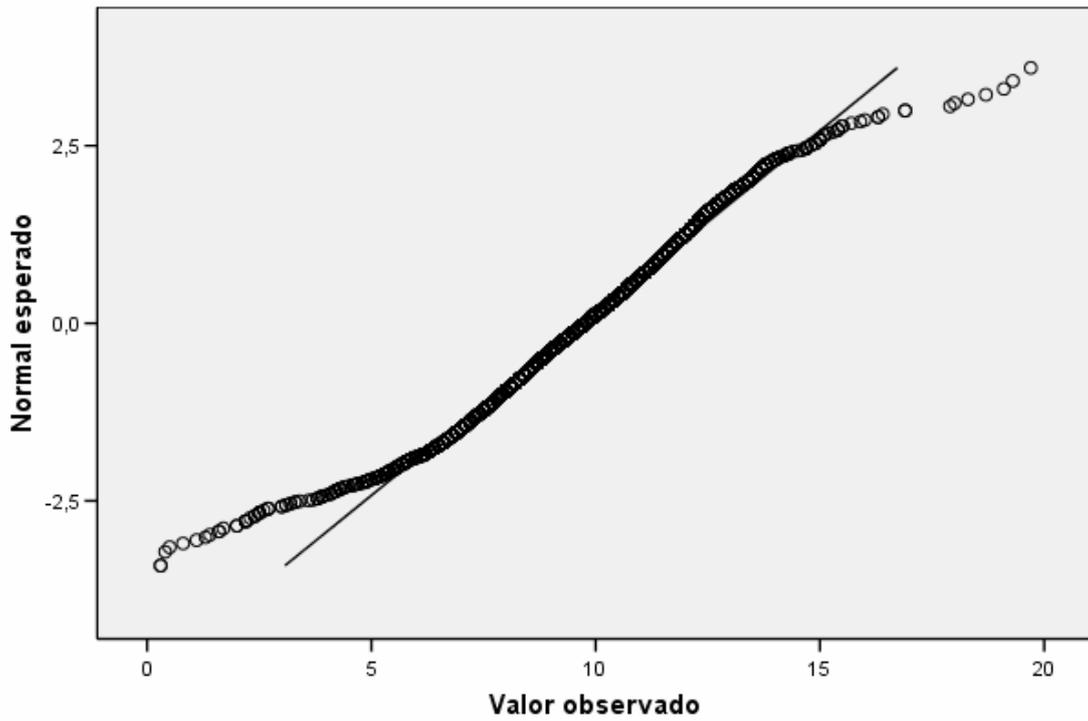
Dado que no se aprecia un problema de asimetría notable y que ninguna de las transformaciones utilizadas habitualmente para reducir la leptocurtosis tuvo excesivo éxito se optó por no transformar esta variable.



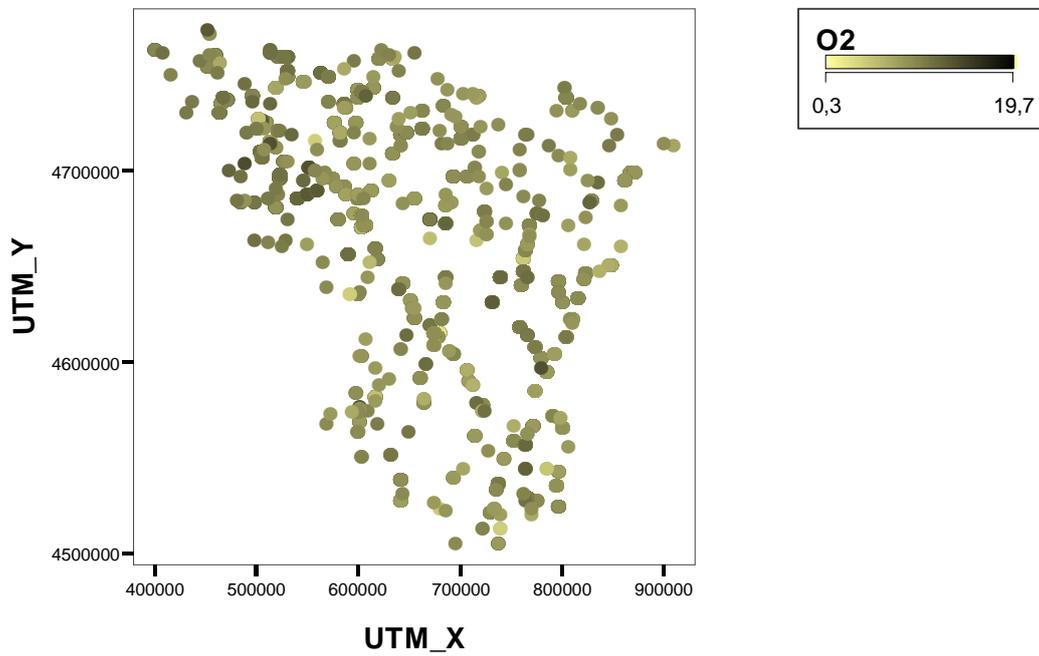
**Figura 3.2.1.5.1: Histograma de la variable O2**



**Figura 3.2.1.5.2: Diagrama de cajas de la variable O2**



**Figura 3.2.1.5.3: QQplot normal de la variable O2**



**Figura 3.2.1.5.4: Distribución geográfica de los niveles de O2**

### 3.2.1.6 Análisis estadístico del oxígeno disuelto en % de saturación

En las Tablas 3.2.1.6.1 a 3.2.1.6.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.1.6.1) que hay un bajo porcentaje (6,5%) y una tendencia fuerte a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 73,8% de los datos que tienen valores ausentes en la variable O2% no cumplen los estándares de calidad frente a un 44,3% de los datos observados (ver Tabla 3.2.1.6.2). Si se hace el estudio por años (ver Tabla 3.2.1.6.3) se observa que los años con un porcentaje mayor de datos ausentes fueron 2004, 2005 y 2007.

**Tabla 3.2.1.6.1: Porcentaje de datos ausentes de la variable O2%**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
6204	93,5%	431	6,5%	6635	100,0%

**Tabla 3.2.1.6.2: Estudio de los datos ausentes de la variable O2% por nivel de calidad**

			miss_O2A		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3413	112	3525
		% de calidad_conjunta	96,8%	3,2%	100,0%
		% de miss_O2A	55,7%	26,2%	53,8%
		Residuos tipificados	2,1	-7,8	
	No cumple	Recuento	2717	316	3033
		% de calidad_conjunta	89,6%	10,4%	100,0%
		% de miss_O2A	44,3%	73,8%	46,2%
		Residuos tipificados	-2,2	8,4	
Total	Recuento	6130	428	6558	
	% de calidad_conjunta	93,5%	6,5%	100,0%	
	% de miss_O2A	100,0%	100,0%	100,0%	

Sin embargo, dado el escaso número de casos con datos ausentes, concluimos que *no se aprecia una falta de representatividad de la muestra analizada que sea relevante.*

**Tabla 3.2.1.6.3: Estudio de los datos ausentes de la variable O2% por años**

			miss_O2A		Total
			Presente	Ausente	
año 2002	Recuento		1305	36	1341
	% de año		97,3%	2,7%	100,0%
	% de miss_O2A		21,0%	8,4%	20,2%
	Residuos tipificados		1,4	-5,4	
2003	Recuento		991	8	999
	% de año		99,2%	,8%	100,0%
	% de miss_O2A		16,0%	1,9%	15,1%
	Residuos tipificados		1,8	-7,0	
2004	Recuento		577	64	641
	% de año		90,0%	10,0%	100,0%
	% de miss_O2A		9,3%	15,0%	9,7%
	Residuos tipificados		-,9	3,5	
2005	Recuento		1164	148	1312
	% de año		88,7%	11,3%	100,0%
	% de miss_O2A		18,8%	34,6%	19,8%
	Residuos tipificados		-1,8	6,9	
2006	Recuento		1217	66	1283
	% de año		94,9%	5,1%	100,0%
	% de miss_O2A		19,6%	15,4%	19,3%
	Residuos tipificados		,5	-1,8	
2007	Recuento		953	106	1059
	% de año		90,0%	10,0%	100,0%
	% de miss_O2A		15,4%	24,8%	16,0%
	Residuos tipificados		-1,2	4,6	
Total	Recuento		6207	428	6635
	% de año		93,5%	6,5%	100,0%
	% de miss_O2A		100,0%	100,0%	100,0%

En las Tablas 3.2.1.6.4 y 3.2.1.6.5 así como en las Figuras 3.2.1.6.1 a 3.2.1.6.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de concentración de oxígeno disuelto en las aguas de la cuenca en % de saturación es igual a 94,33% de saturación y el nivel mediano 94,80% de saturación (ver Tabla 3.2.1.6.4). La hipótesis de normalidad se rechaza debido, fundamentalmente, a la existencia de un elevado grado de leptocurtosis (ver Tablas 3.2.1.6.4 y 3.2.1.6.5 así como las Figuras 3.2.1.6.1 a 3.2.1.6.3). Ello es debido a la existencia de puntos de muestreo con valores significativamente más altos y más bajos de esta variable (ver Figura 3.2.1.6.2). Si se analiza la distribución geográfica de esta variable (ver Figura 3.2.1.6.4) se observa que *en la parte alta de la cuenca tienden a concentrarse los puntos de muestreo con mayores niveles de oxígeno, dándose la tendencia contraria en los puntos de muestreo de la parte baja.*

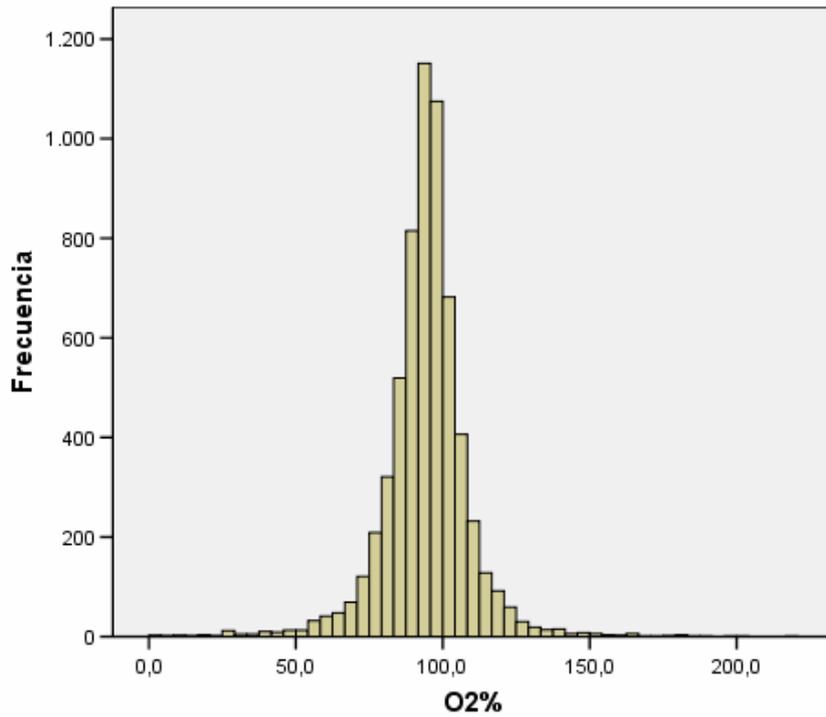
**Tabla 3.2.1.6.4: Análisis estadístico-descriptivo de la variable O2%**

		Estadístico	Error típ.
O2%	Media	94,331	,1892
	Intervalo de confianza para la media al 95%	Límite inferior Límite superior	
		93,961 94,702	
	Media recortada al 5%	94,469	
	Mediana	94,800	
	Varianza	222,161	
	Desv. típ.	14,9051	
	Mínimo	2,7	
	Máximo	219,0	
	Rango	216,3	
	Amplitud intercuartil	12,8	
	Asimetría	-,076	,031
	Curtosis	8,011	,062

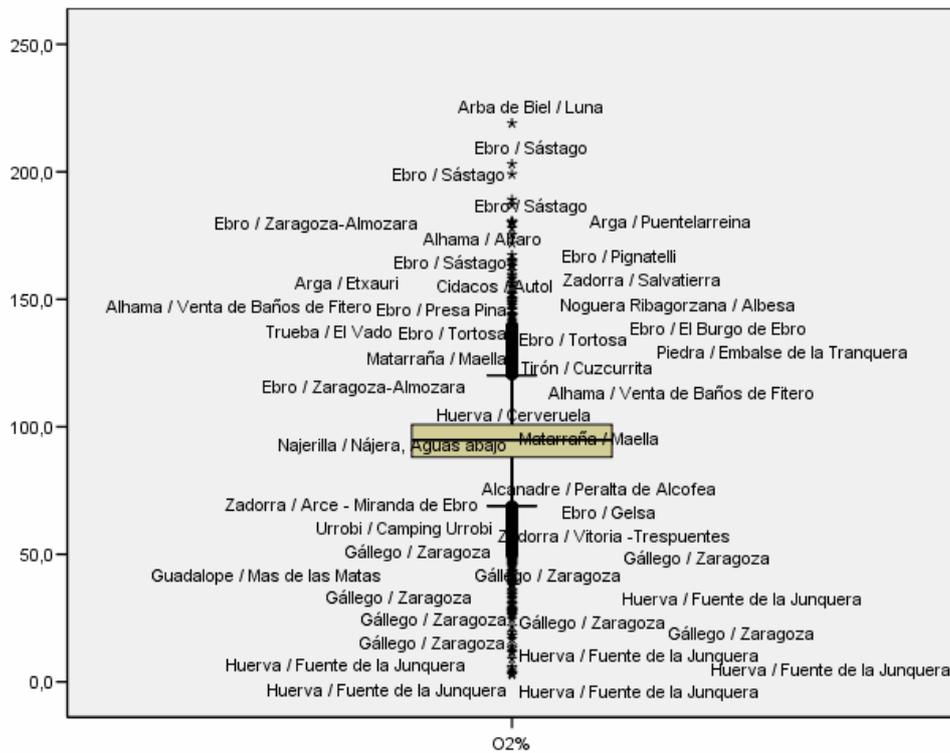
**Tabla 3.2.1.6.5: Pruebas de normalidad de la variable O2%**

	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
O2%	,095	6207	,000

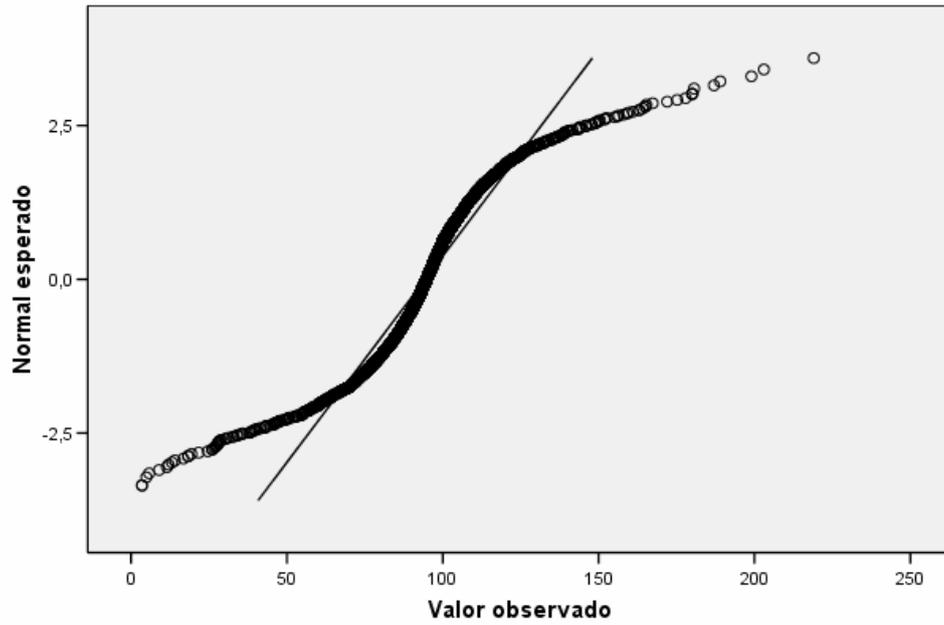
Dado que no se aprecia un problema de asimetría notable y que ninguna de las transformaciones utilizadas habitualmente para reducir la leptocurtosis tuvo excesivo éxito se optó por no transformar esta variable.



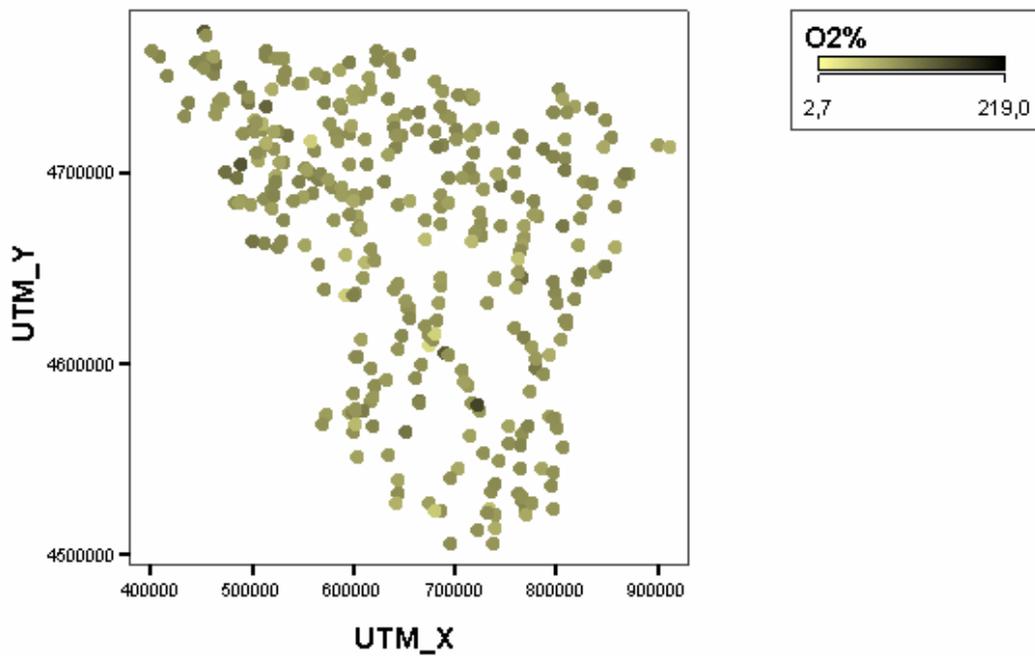
**Figura 3.2.1.6.1: Histograma de la variable O2%**



**Figura 3.2.1.6.2: Diagrama de cajas de la variable O2%**



**Figura 3.2.1.6.3: QQplot normal de la variable O2%**



**Figura 3.2.1.6.4: Distribución geográfica de los niveles de O2%**

### 3.2.1.7 Análisis estadístico de los sólidos en suspensión

En las Tablas 3.2.1.7.1 a 3.2.1.7.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.1.7.1) que, aunque el porcentaje de datos ausentes no es muy alto (12,2%), existe una tendencia significativa a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 63,6% de los datos que tienen valores ausentes en la variable MES no cumplen los estándares de calidad frente a un 43,8% de los datos observados (ver Tabla 3.2.1.7.2). Si se hace el estudio por años (ver Tabla 3.2.1.7.3) se observa que el año 2003 es el que menos datos ausentes presentó y 2007 el que más.

Existe, por lo tanto, una *tendencia significativa de los datos presentes a cumplir los requisitos de calidad que podría sesgar los resultados que se exponen en la sección 5 sobrevalorando los niveles de calidad de los puntos de muestreo para un valor dado de la variable sólidos en suspensión, por lo que los resultados obtenidos respecto a esta variable deberían tomarse con precaución.*

**Tabla 3.2.1.7.1: Porcentaje de datos ausentes de la variable MES**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
5826	87,8%	809	12,2%	6635	100,0%

**Tabla 3.2.1.7.2: Estudio de los datos ausentes de la variable MES por nivel de calidad**

			miss_MES		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	3232	293	3525
		% de calidad_conjunta	91,7%	8,3%	100,0%
		% de miss_MES	56,2%	36,4%	53,8%
		Residuos tipificados	2,5	-6,7	
	No cumple	Recuento	2520	513	3033
		% de calidad_conjunta	83,1%	16,9%	100,0%
		% de miss_MES	43,8%	63,6%	46,2%
		Residuos tipificados	-2,7	7,3	
Total	Recuento	5752	806	6558	
	% de calidad_conjunta	87,7%	12,3%	100,0%	
	% de miss_MES	100,0%	100,0%	100,0%	

**Tabla 3.2.1.7.3: Estudio de los datos ausentes de la variable MES por años**

			miss_MES		Total
			Presente	Ausente	
año	2002	Recuento	1197	144	1341
		% de año	89,3%	10,7%	100,0%
		% de miss_MES	20,5%	17,8%	20,2%
		Residuos tipificados	,6	-1,5	
	2003	Recuento	982	17	999
		% de año	98,3%	1,7%	100,0%
		% de miss_MES	16,9%	2,1%	15,1%
		Residuos tipificados	3,5	-9,5	
	2004	Recuento	555	86	641
		% de año	86,6%	13,4%	100,0%
		% de miss_MES	9,5%	10,6%	9,7%
		Residuos tipificados	-,3	,9	
	2005	Recuento	1107	205	1312
		% de año	84,4%	15,6%	100,0%
		% de miss_MES	19,0%	25,3%	19,8%
		Residuos tipificados	-1,3	3,6	
	2006	Recuento	1126	157	1283
		% de año	87,8%	12,2%	100,0%
		% de miss_MES	19,3%	19,4%	19,3%
		Residuos tipificados	,0	,0	
	2007	Recuento	859	200	1059
		% de año	81,1%	18,9%	100,0%
		% de miss_MES	14,7%	24,7%	16,0%
		Residuos tipificados	-2,3	6,2	
Total		Recuento	5826	809	6635
		% de año	87,8%	12,2%	100,0%
		% de miss_MES	100,0%	100,0%	100,0%

En las Tablas 3.2.1.7.4 y 3.2.1.7.5 así como en las Figuras 3.2.1.7.1 a 3.2.1.7.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de sólidos en suspensión en las aguas de la cuenca es igual a 31,6 mg/L con un nivel mediano igual a 8,0 mg/L (ver Tabla 3.2.1.7.4). La distribución de esta variable presenta una asimetría positiva y una leptocurtosis muy significativas (ver Tabla 3.2.1.7.4) reflejando la existencia de puntos de muestreo con un alto nivel de sólidos en suspensión (ver Figuras 3.2.1.7.1 y 3.2.1.7.2). Por todo ello se rechaza claramente la hipótesis de normalidad de esta variable (ver Tabla 3.2.1.7.5 y Figura 3.2.1.7.3). Dicho problema se soluciona en parte si se realiza **una transformación logarítmica (variable LMES)** la cual **aumenta el grado de normalidad** de la misma, tal y como lo demuestran los resultados expuestos en las Figuras 3.2.1.7.4 a 3.2.1.7.6. Por

dicha razón, en el estudio realizado en la sección 5 *la determinación de umbrales se realizará a partir de la variable transformada LMES.*

Finalmente, en la Figura 3.2.1.7.7 se puede observar como *los puntos de muestreo con un alto nivel de sólidos en suspensión tienden a concentrarse en los alrededores del cauce del río Ebro.*

**Tabla 3.2.1.7.4: Análisis estadístico-descriptivo de las variables MES y LMES**

		Estadístico	Error típ.	
MES	Media	31,60	2,469	
	Intervalo de confianza para la media al 95%	Límite inferior	26,76	
		Límite superior	36,44	
	Media recortada al 5%	15,63		
	Mediana	8,00		
	Varianza	35513,720		
	Desv. típ.	188,451		
	Mínimo	1		
	Máximo	7848		
	Rango	7847		
	Amplitud intercuartil	20		
	Asimetría	28,700	,032	
	Curtosis	998,375	,064	
	LMES	Media	2,2301	,01752
Intervalo de confianza para la media al 95%		Límite inferior	2,1958	
		Límite superior	2,2645	
Media recortada al 5%		2,1809		
Mediana		2,0794		
Varianza		1,789		
Desv. típ.		1,33753		
Mínimo		,00		
Máximo		8,97		
Rango		8,97		
Amplitud intercuartil		2,04		
Asimetría		,577	,032	
Curtosis		,399	,064	

**Tabla 3.2.1.7.5: Pruebas de normalidad de las variables MES y LMES**

	Kolmogorov-Smirnov		
	Estadístico	gl	Sig.
MES	,436	5826	,000
LMES	,082	5826	,000

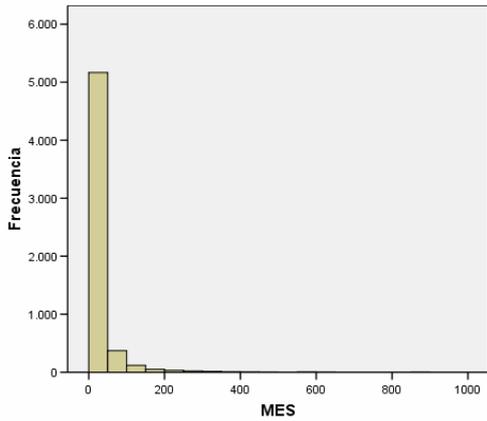


Figura 3.2.1.7.1: Histograma de la variable MES

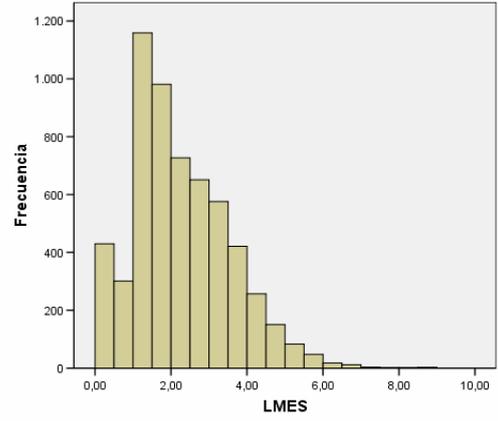


Figura 3.2.1.7.4 : Histograma de la variable LMES

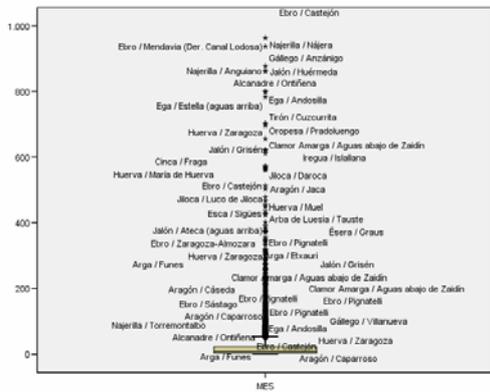


Figura 3.2.1.7.2: Diagrama de cajas de la variable MES

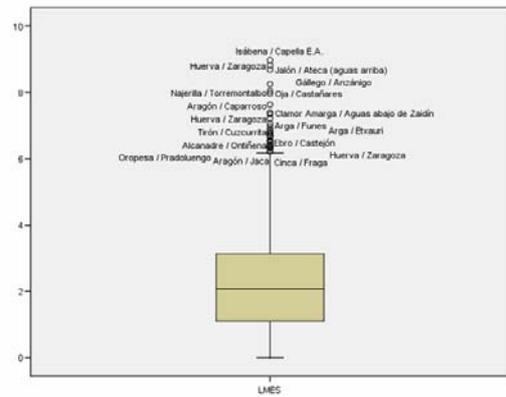


Figura 3.2.1.7.5: Diagrama de cajas de la variable LMES

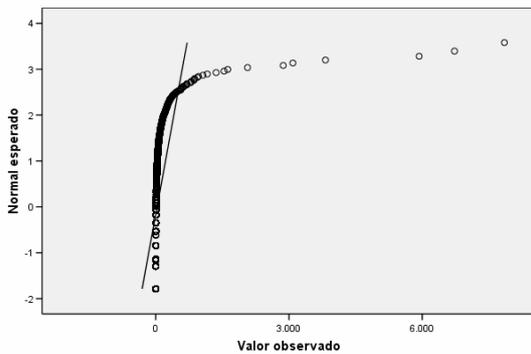


Figura 3.2.1.7.3: QQplot normal de la variable MES

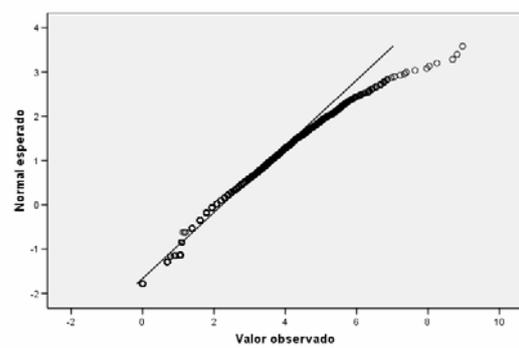
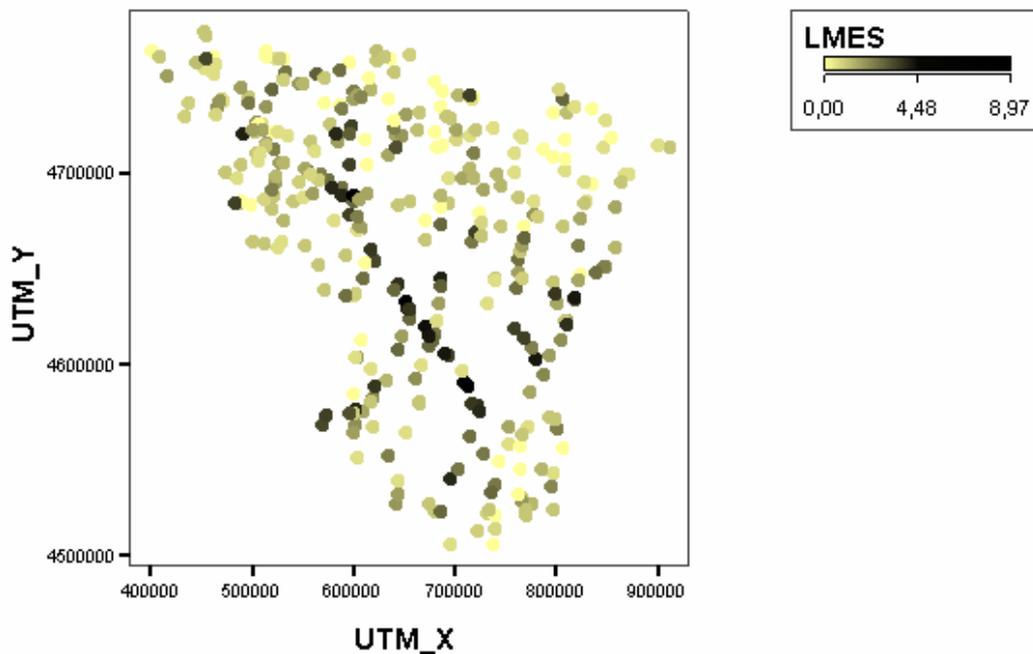


Figura 3.2.1.7.6: QQplot normal de la variable LMES



**Figura 3.2.1.7.7: Distribución geográfica de los niveles de LMES**

### **3.2.2. Análisis estadístico de los indicadores de materia orgánica**

En este apartado presentamos el análisis estadístico-descriptivo del indicador de materia orgánica, DQO, que mide la demanda química de oxígeno en mg/L O<sub>2</sub>. (ver Tabla 2.3).

#### **3.2.2.1 Análisis estadístico de la demanda química de oxígeno**

En las Tablas 3.2.2.1.1 a 3.2.2.1.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa que hay un alto porcentaje (52,7%) así como la existencia de una tendencia significativa a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 54,5% de los datos que tienen valores ausentes en la variable DQO no cumplen los estándares de calidad frente a un 37,2% de los datos observados (ver Tabla 3.2.2.1.2). Si se hace el estudio por años (ver Tabla 3.2.2.1.3) se observa cómo el porcentaje de datos ausentes aumenta considerablemente partir del año 2005, no existiendo en el año 2007 ningún dato observado.

Todo ello puede influir en los resultados obtenidos en este estudio debido a la *poca representatividad de los datos presentes claramente sesgados hacia valores altos*

de los niveles de calidad por lo que los umbrales obtenidos deben ser tomados con precaución.

**Tabla 3.2.2.1.1: Porcentaje de datos ausentes de la variable DQO**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
3137	47,3%	3498	52,7%	6635	100,0%

**Tabla 3.2.2.1.2: Estudio de los datos ausentes de la variable DQO por nivel de calidad**

			miss_DQO		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	1964	1561	3525
		% de calidad_conjunta	55,7%	44,3%	100,0%
		% de miss_DQO	62,8%	45,5%	53,8%
		Residuos tipificados	6,9	-6,6	
	No cumple	Recuento	1161	1872	3033
		% de calidad_conjunta	38,3%	61,7%	100,0%
		% de miss_DQO	37,2%	54,5%	46,2%
		Residuos tipificados	-7,5	7,1	
Total	Recuento	3125	3433	6558	
	% de calidad_conjunta	47,7%	52,3%	100,0%	
	% de miss_DQO	100,0%	100,0%	100,0%	

**Tabla 3.2.2.1.3: Estudio de los datos ausentes de la variable DQO por años**

			miss_DQO		Total
			Presente	Ausente	
año	2002	Recuento	1189	152	1341
		% de año	88,7%	11,3%	100,0%
		% de miss_DQO	37,9%	4,3%	20,2%
		Residuos tipificados	22,0	-20,9	
	2003	Recuento	985	14	999
		% de año	98,6%	1,4%	100,0%
		% de miss_DQO	31,4%	,4%	15,1%
		Residuos tipificados	23,6	-22,3	
	2004	Recuento	557	84	641
		% de año	86,9%	13,1%	100,0%
		% de miss_DQO	17,8%	2,4%	9,7%
		Residuos tipificados	14,6	-13,8	
	2005	Recuento	405	907	1312
		% de año	30,9%	69,1%	100,0%
		% de miss_DQO	12,9%	25,9%	19,8%
		Residuos tipificados	-8,6	8,2	
2006	Recuento	1	1282	1283	
	% de año	,1%	99,9%	100,0%	
	% de miss_DQO	,0%	36,6%	19,3%	
	Residuos tipificados	-24,6	23,3		
2007	Recuento	0	1059	1059	
	% de año	,0%	100,0%	100,0%	
	% de miss_DQO	,0%	30,3%	16,0%	
	Residuos tipificados	-22,4	21,2		
Total	Recuento	3137	3498	6635	
	% de año	47,3%	52,7%	100,0%	
	% de miss_DQO	100,0%	100,0%	100,0%	

En las Tablas 3.2.2.1.4 y 3.2.2.1.5 así como en las Figuras 3.2.2.1.1 a 3.2.2.1.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de dicromato en las aguas de la cuenca es igual a 9,90 mg/L con un nivel mediano igual a 8,60 mg/L (ver Tabla 3.2.1.4.4). La distribución de esta variable presenta una clara asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.2.1.4) reflejando la existencia de puntos de muestreo con un alto nivel de dicromato (ver Figuras 3.2.2.1.1 y 3.2.2.1.2). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.2.1.5 y Figura 3.2.2.1.3) debiendo realizarse una *transformación logarítmica (variable LDQO) para aumentar significativamente su grado de normalidad* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.2.1.4 a 3.2.2.1.6.

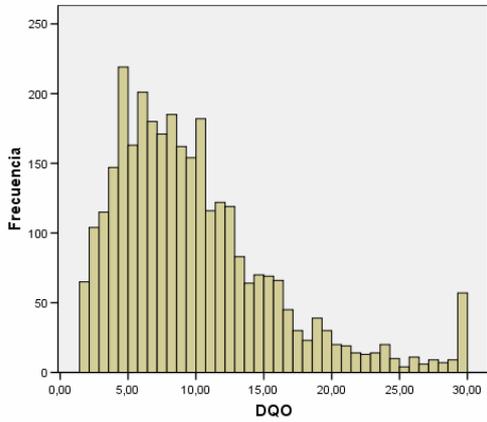
Por dichas razones, *los umbrales de calidad de esta variable se determinarán a partir de la variable transformada LDQO.*

**Tabla 3.2.2.1.4: Análisis estadístico-descriptivo de las variables DQO y LDQO**

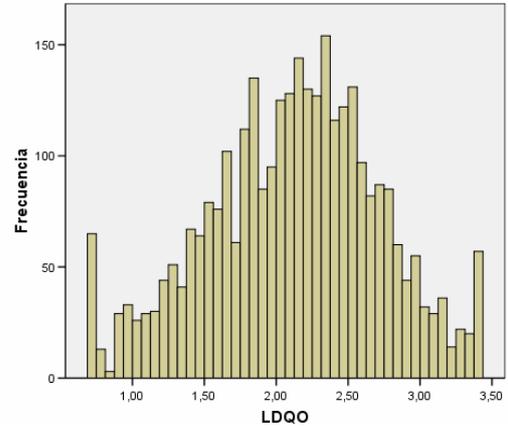
			Estadístico	Error típ.	
DQO	Media		9,9001	,10622	
	Intervalo de confianza para la media al 95%	Límite inferior	9,6919		
		Límite superior	10,1084		
	Media recortada al 5%		9,3712		
	Mediana		8,6000		
	Varianza		35,393		
	Desv. típ.		5,94918		
	Mínimo		2,00		
	Máximo		30,00		
	Rango		28,00		
	Amplitud intercuartil		6,90		
	Asimetría		1,304		,044
	Curtosis		1,781		,087
	LDQO	Media			2,1194
Intervalo de confianza para la media al 95%		Límite inferior	2,0983		
		Límite superior	2,1405		
Media recortada al 5%			2,1258		
Mediana			2,1518		
Varianza			,363		
Desv. típ.			,60218		
Mínimo			,69		
Máximo			3,40		
Rango			2,71		
Amplitud intercuartil			,80		
Asimetría			-,188	,044	
Curtosis			-,299	,087	

**Tabla 3.2.2.1.5: Pruebas de normalidad de las variables DQO y LDQO**

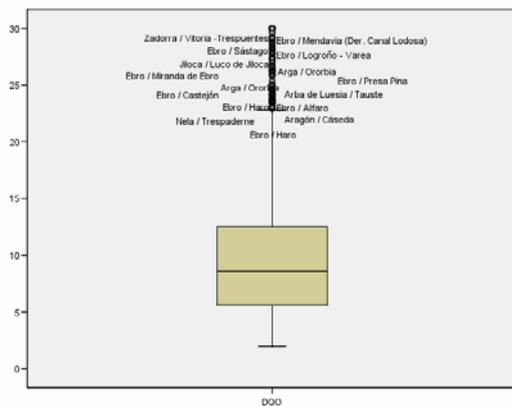
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
DQO	,102	3137	,000	,896	3137	,000
LDQO	,034	3137	,000	,991	3137	,000



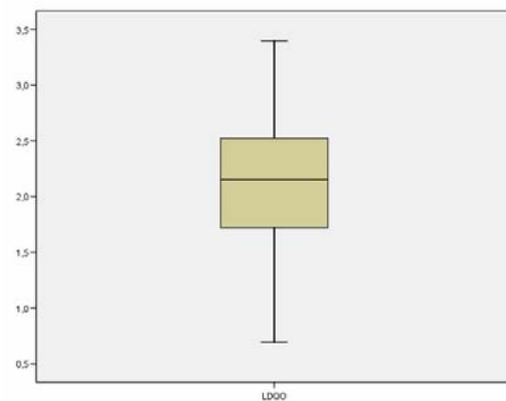
**Figura 3.2.2.1.1: Histograma de la variable DQO**



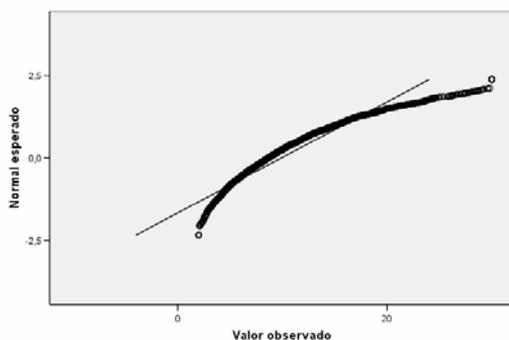
**Figura 3.2.2.1.4 : Histograma de la variable LDQO**



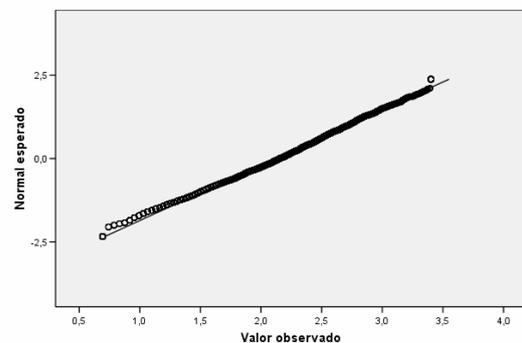
**Figura 3.2.2.1.2: Diagrama de cajas de la variable DQO**



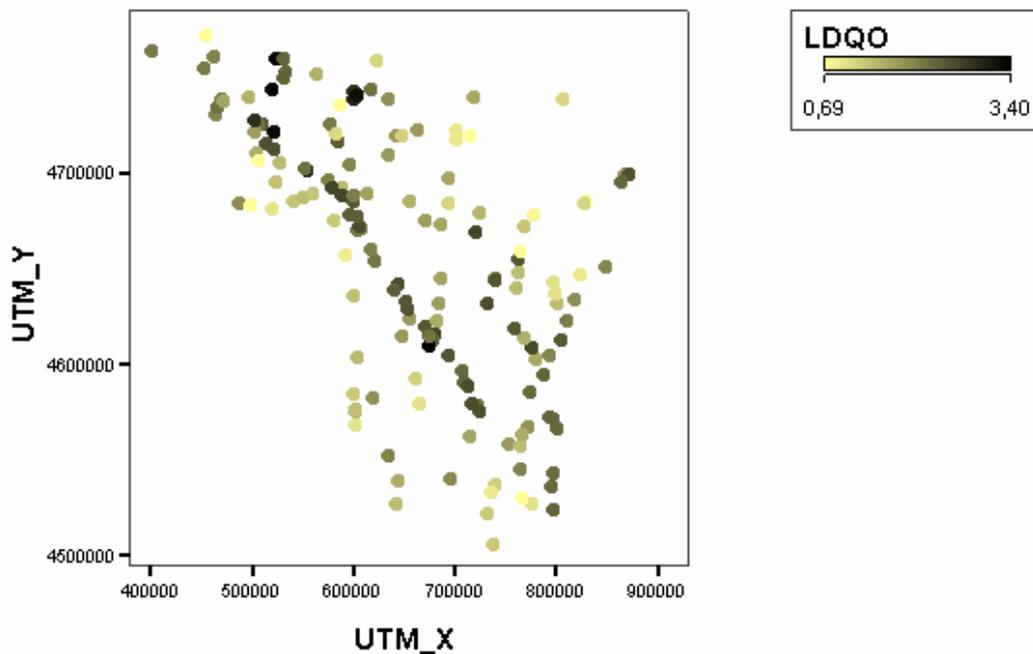
**Figura 3.2.2.1.5 : Diagrama de cajas de la variable LDQO**



**Figura 3.2.2.1.3: QQplot normal de la variable DQO**



**Figura 3.2.2.1.6: QQplot normal de la variable LDQO**



**Figura 3.2.2.1.7: Distribución geográfica de los niveles de LDQO**

En la Figura 3.2.2.1.7 se observa que *los puntos de muestreo con un alto nivel de DQO tienden a concentrarse a lo largo del cauce del río Ebro y en sus zonas cercanas, sobre todo en las zonas altas del cauce.*

### **3.2.3. Análisis estadístico de los indicadores de acidificación**

En este apartado presentamos el análisis estadístico-descriptivo de los indicadores de acidificación carbonatos (CO3CA) y bicarbonatos (CO32CA).

#### **3.2.3.1 Análisis estadístico de los carbonatos**

En las Tablas 3.2.3.1.1 a 3.2.3.1.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.3.1.1) que hay un altísimo porcentaje (88,4%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55.2% de los datos que tienen valores ausentes en la variable CO3CA cumplen los estándares de calidad frente a un 42.6% de los datos observados (ver Tabla 3.2.3.1.2). *Existe, por tanto, una sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán*

*tomarse con precaución.* Si se hace el estudio por años (ver Tabla 3.2.3.1.3) se observa como el porcentaje de datos ausentes ha disminuido considerablemente en el año 2007.

**Tabla 3.2.3.1.1: Porcentaje de datos ausentes de la variable CO3CA**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
771	11,6%	5864	88,4%	6635	100,0%

**Tabla 3.2.3.1.2: Estudio de los datos ausentes de la variable CO3CA por nivel de calidad**

			miss_C03CA		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	316	3209	3525
		% de calidad_conjunta	9,0%	91,0%	100,0%
		% de miss_C03CA	42,6%	55,2%	53,8%
		Residuos tipificados	-4,1	1,5	
	No cumple	Recuento	426	2607	3033
		% de calidad_conjunta	14,0%	86,0%	100,0%
		% de miss_C03CA	57,4%	44,8%	46,2%
		Residuos tipificados	4,5	-1,6	
Total		Recuento	742	5816	6558
		% de calidad_conjunta	11,3%	88,7%	100,0%
		% de miss_C03CA	100,0%	100,0%	100,0%

**Tabla 3.2.3.1.3: Estudio de los datos ausentes de la variable CO3CA por años**

			miss_C03CA		Total
			Presente	Ausente	
año 2002	Recuento	4	1337	1341	
	% de año	,3%	99,7%	100,0%	
	% de miss_C03CA	,5%	22,8%	20,2%	
	Residuos tipificados	-12,2	4,4		
2003	Recuento	0	999	999	
	% de año	,0%	100,0%	100,0%	
	% de miss_C03CA	,0%	17,0%	15,1%	
	Residuos tipificados	-10,8	3,9		
2004	Recuento	0	641	641	
	% de año	,0%	100,0%	100,0%	
	% de miss_C03CA	,0%	10,9%	9,7%	
	Residuos tipificados	-8,6	3,1		
2005	Recuento	118	1194	1312	
	% de año	9,0%	91,0%	100,0%	
	% de miss_C03CA	15,3%	20,4%	19,8%	
	Residuos tipificados	-2,8	1,0		
2006	Recuento	191	1092	1283	
	% de año	14,9%	85,1%	100,0%	
	% de miss_C03CA	24,8%	18,6%	19,3%	
	Residuos tipificados	3,4	-1,2		
2007	Recuento	458	601	1059	
	% de año	43,2%	56,8%	100,0%	
	% de miss_C03CA	59,4%	10,2%	16,0%	
	Residuos tipificados	30,2	-10,9		
Total		Recuento	771	5864	6635
		% de año	11,6%	88,4%	100,0%
		% de miss_C03CA	100,0%	100,0%	100,0%

En las Tablas 3.2.3.1.4 y 3.2.3.1.5 así como en las Figuras 3.2.3.1.1 a 3.2.3.1.8 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. Se observa, en primer lugar, (ver Figura 3.2.2.1.1) **la existencia de una distribución de tipo mixto por censura dado que, para el 64,9% de los datos observados únicamente se sabe que el valor de la variable  $CO_3CA \leq 5 \text{ mg/L } CO_3CA$ .** Si eliminamos estos casos el nivel medio de carbonatos en las aguas de la cuenca es igual a 13,05 mg/L  $CO_3CA$ , con un nivel mediano igual a 11,5 mg/L  $CO_3CA$  (ver Tabla 3.2.3.1.4). La distribución de los datos observados con precisión presenta, a su vez, una asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.3.1.4) reflejando la **existencia de puntos de muestreo con un alto nivel de carbonatos** (ver Figuras 3.2.3.1.2 y 3.2.3.1.3). **Dichos puntos tienden a concentrarse en las zonas montañosas del norte de la cuenca** (ver Figura 3.2.3.1.8).

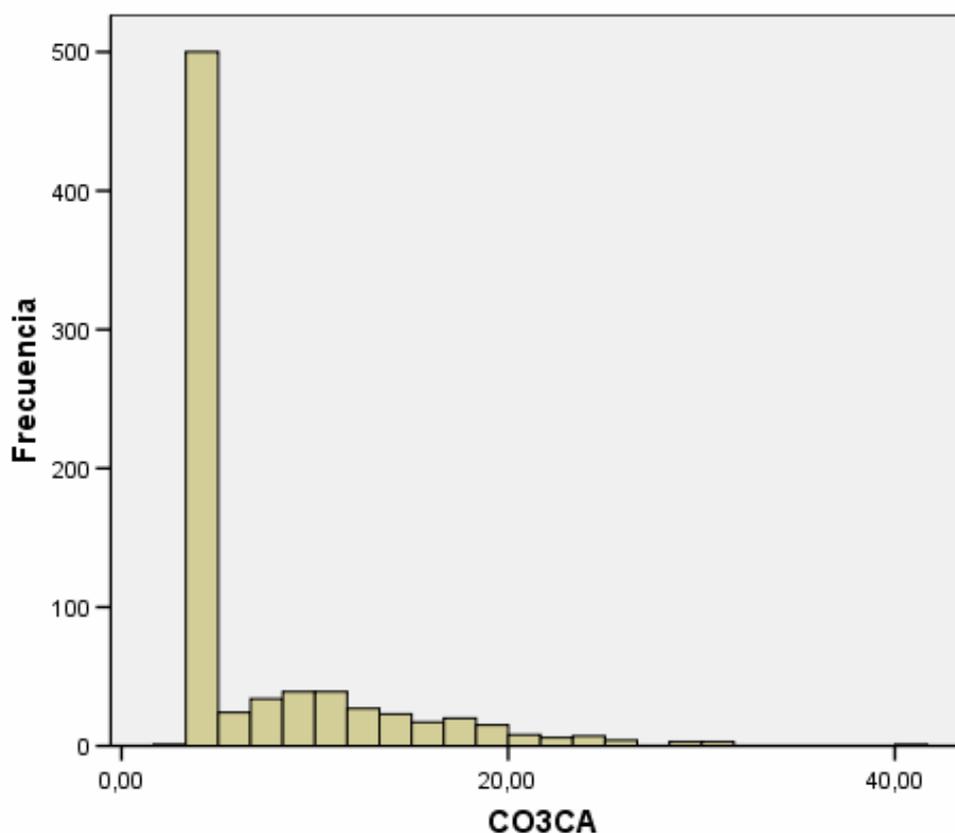
Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.3.1.5 y Figura 3.2.3.1.4) **siendo necesaria una transformación logarítmica (variable  $LCO_3CA$ ) para conseguir su normalidad**, tal y como lo demuestran los resultados expuestos en las Figuras 3.2.3.1.5 a 3.2.3.1.7.

**Tabla 3.2.3.1.4: Análisis estadístico-descriptivo de las variables  $CO_3CA$  y  $LCO_3CA$  si  $CO_3CA > 5 \text{ mg/L } CO_3CA$**

			Estadístico	Error típ.
$CO_3CA$	Media		13,0528	,35865
	Intervalo de confianza para la media al 95%	Límite inferior	12,3467	
		Límite superior	13,7589	
	Media recortada al 5%		12,6296	
	Mediana		11,5000	
	Varianza		34,858	
	Desv. típ.		5,90410	
	Mínimo		3,00	
	Máximo		41,50	
	Rango		38,50	
	Amplitud intercuartil		7,90	
	Asimetría		1,189	,148
	Curtosis		1,955	,295
	$LCO_3CA$	Media		2,4745
Intervalo de confianza para la media al 95%		Límite inferior	2,4225	
		Límite superior	2,5265	
Media recortada al 5%			2,4724	
Mediana			2,4423	
Varianza			,189	
Desv. típ.			,43485	
Mínimo			1,10	
Máximo			3,73	
Rango			2,63	
Amplitud intercuartil			,65	
Asimetría			,066	,148
Curtosis			-,337	,295

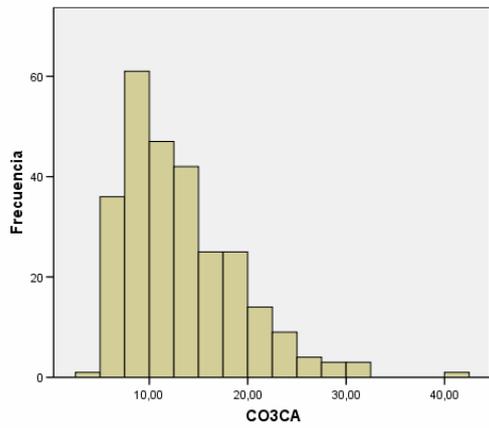
**Tabla 3.2.3.1.5: Pruebas de normalidad de las variables CO<sub>3</sub>CA y LCO<sub>3</sub>CA si CO<sub>3</sub>CA > 5 mg/L CO<sub>3</sub>CA**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CO <sub>3</sub> CA	,111	271	,000	,916	271	,000
LCO <sub>3</sub> CA	,052	271	,069	,990	271	,048

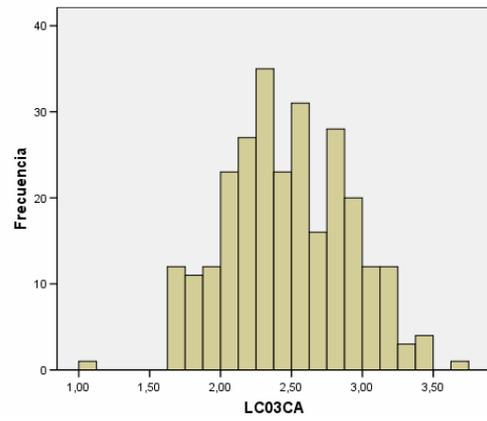


**Figura 3.2.3.1.1: Histograma de la variable CO<sub>3</sub>CA para todos los casos**

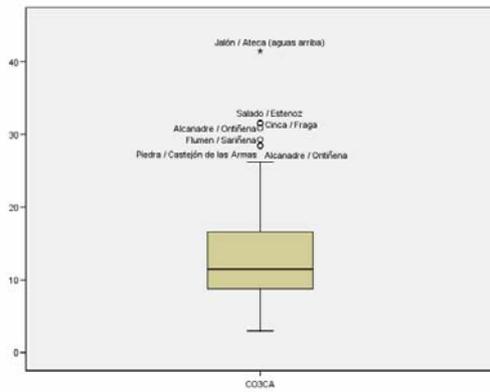
Debido a los resultados obtenidos *utilizaremos, para la determinación de los umbrales de esta variable, la variable transformada logarítmicamente, LCO<sub>3</sub>CA así como una variable indicador de censura ICO<sub>3</sub>CA que vale 1 si el valor de CO<sub>3</sub>CA está censurado (es decir, es menor o igual que 5 mg/L CO<sub>3</sub>CA) y 0 en caso contrario.*



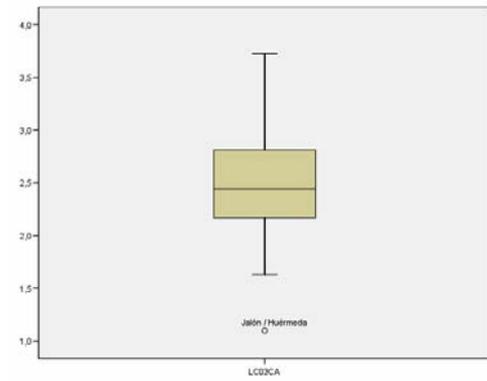
**Figura 3.2.3.1.2: Histograma de la variable CO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



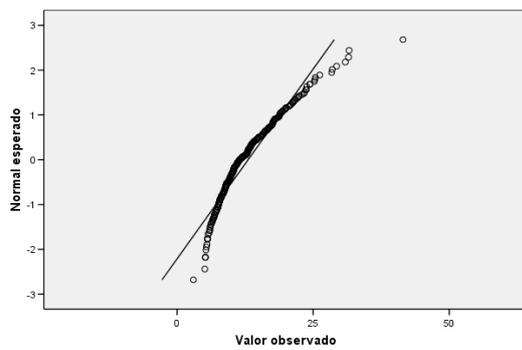
**Figura 3.2.3.1.5: Histograma de la variable LCO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



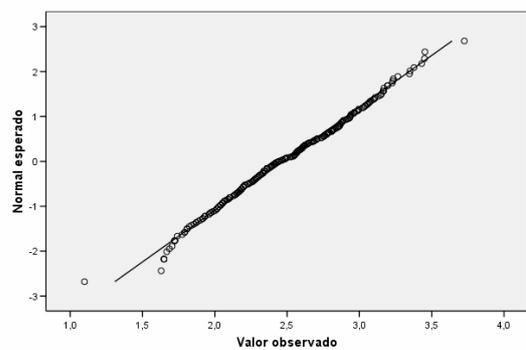
**Figura 3.2.3.1.3: Diagrama de cajas de la variable CO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



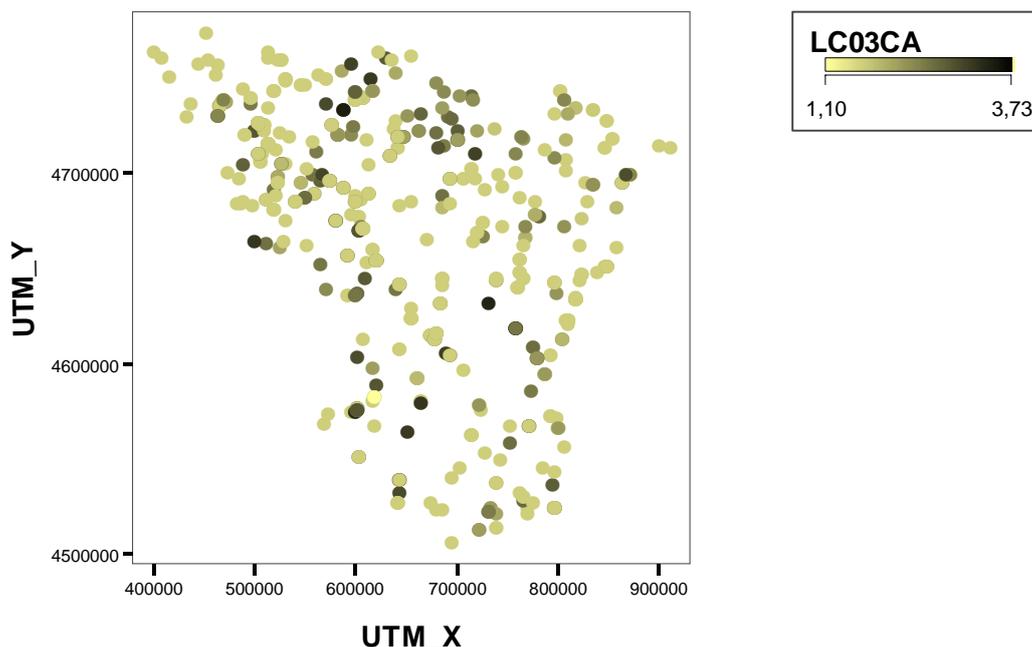
**Figura 3.2.3.1.6 : Diagrama de cajas de la variable LCO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



**Figura 3.2.3.1.4: QQplot normal de la variable CO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



**Figura 3.2.3.1.7: QQplot normal de la variable LCO3CA si CO3CA>5 mg/L CO<sub>3</sub>CA**



**Figura 3.2.3.1.8: Distribución geográfica de los niveles de LCO3CA**

### 3.2.3.2 Análisis estadístico de los bicarbonatos

En las Tablas 3.2.3.2.1 a 3.2.3.2.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.3.2.1) que hay un altísimo porcentaje (87,2%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55,1 % de los datos que tienen valores ausentes en la variable CO3CA cumplen los estándares de calidad frente a un 44,4% de los datos observados (ver Tabla 3.2.3.2.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución*. Si se hace el estudio por años (ver Tabla 3.2.3.2.3) se observa como el porcentaje de datos ausentes ha disminuido considerablemente en el año 2007.

**Tabla 3.2.3.2.1: Porcentaje de datos ausentes de la variable CO32CA**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
846	12,8%	5789	87,2%	6635	100,0%

**Tabla 3.2.3.2.2: Estudio de los datos ausentes de la variable CO32CA por nivel de calidad**

			miss_C032CA		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	363	3162	3525
		% de calidad_conjunta	10,3%	89,7%	100,0%
		% de miss_C032CA	44,4%	55,1%	53,8%
		Residuos tipificados	-3,6	1,4	
	No cumple	Recuento	454	2579	3033
		% de calidad_conjunta	15,0%	85,0%	100,0%
		% de miss_C032CA	55,6%	44,9%	46,2%
		Residuos tipificados	3,9	-1,5	
Total	Recuento	817	5741	6558	
	% de calidad_conjunta	12,5%	87,5%	100,0%	
	% de miss_C032CA	100,0%	100,0%	100,0%	

**Tabla 3.2.3.2.3: Estudio de los datos ausentes de la variable CO32CA por años**

			miss_C032CA		Total
			Presente	Ausente	
año	2002	Recuento	79	1262	1341
		% de año	5,9%	94,1%	100,0%
		% de miss_C032CA	9,3%	21,8%	20,2%
		Residuos tipificados	-7,0	2,7	
	2003	Recuento	0	999	999
		% de año	,0%	100,0%	100,0%
		% de miss_C032CA	,0%	17,3%	15,1%
		Residuos tipificados	-11,3	4,3	
	2004	Recuento	0	641	641
		% de año	,0%	100,0%	100,0%
		% de miss_C032CA	,0%	11,1%	9,7%
		Residuos tipificados	-9,0	3,5	
	2005	Recuento	118	1194	1312
		% de año	9,0%	91,0%	100,0%
		% de miss_C032CA	13,9%	20,6%	19,8%
		Residuos tipificados	-3,8	1,5	
2006	Recuento	191	1092	1283	
	% de año	14,9%	85,1%	100,0%	
	% de miss_C032CA	22,6%	18,9%	19,3%	
	Residuos tipificados	2,1	-,8		
2007	Recuento	458	601	1059	
	% de año	43,2%	56,8%	100,0%	
	% de miss_C032CA	54,1%	10,4%	16,0%	
	Residuos tipificados	27,8	-10,6		
Total	Recuento	846	5789	6635	
	% de año	12,8%	87,2%	100,0%	
	% de miss_C032CA	100,0%	100,0%	100,0%	

En las Tablas 3.2.3.2.4 y 3.2.3.2.5 así como en las Figuras 3.2.3.2.1 a 3.2.3.2.4 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de bicarbonatos en las aguas de la cuenca es igual a 190,82 mg/L CO<sub>3</sub>CA y el nivel mediano 191,00 mg/L CO<sub>3</sub>CA (ver Tabla 3.2.3.2.4). A pesar de que el test de Kolmogorov-Smirnov rechaza la hipótesis de normalidad para esta variable (ver Tabla 3.2.3.2.5) debido a la existencia de una leptocurtosis significativa (ver Tabla 3.2.3.2.4) el análisis gráfico realizado en las Figuras 3.2.3.2.1 a 3.2.3.2.3 pone de manifiesto que el grado de falta de normalidad no es alto. Los atípicos son, en su mayor parte, débiles y son debidos, en su inmensa mayoría, a la presencia de puntos de muestreo con valores altos de este indicador. Se observa (ver Figura 3.2.3.2.4), además, una *tendencia a tomar valores altos de este indicador a lo largo del cauce del río Ebro y en zonas de la cuenca cercanas al mismo*.

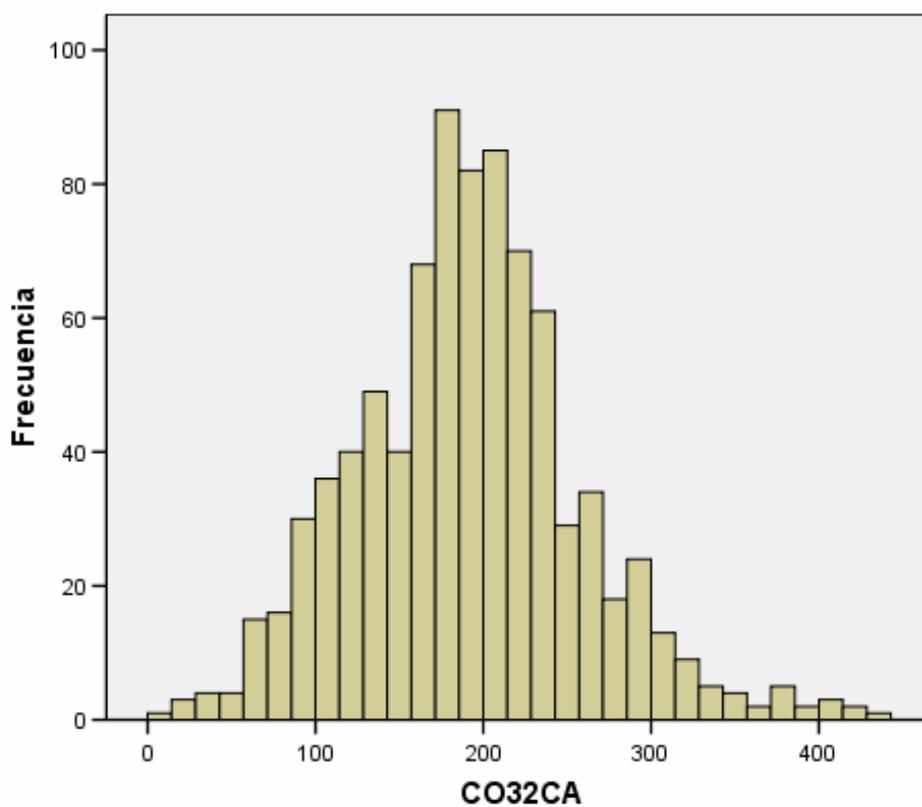
**Tabla 3.2.3.2.4: Análisis estadístico-descriptivo de la variable CO32CA**

			Estadístico	Error típ.
CO32CA	Media		190,82	2,303
	Intervalo de confianza para la media al 95%	Límite inferior	186,30	
		Límite superior	195,34	
	Media recortada al 5%		189,23	
	Mediana		191,00	
	Varianza		4486,577	
	Desv. típ.		66,982	
	Mínimo		14	
	Máximo		429	
	Rango		414	
	Amplitud intercuartil		82	
	Asimetría		,366	,084
	Curtosis		,649	,168

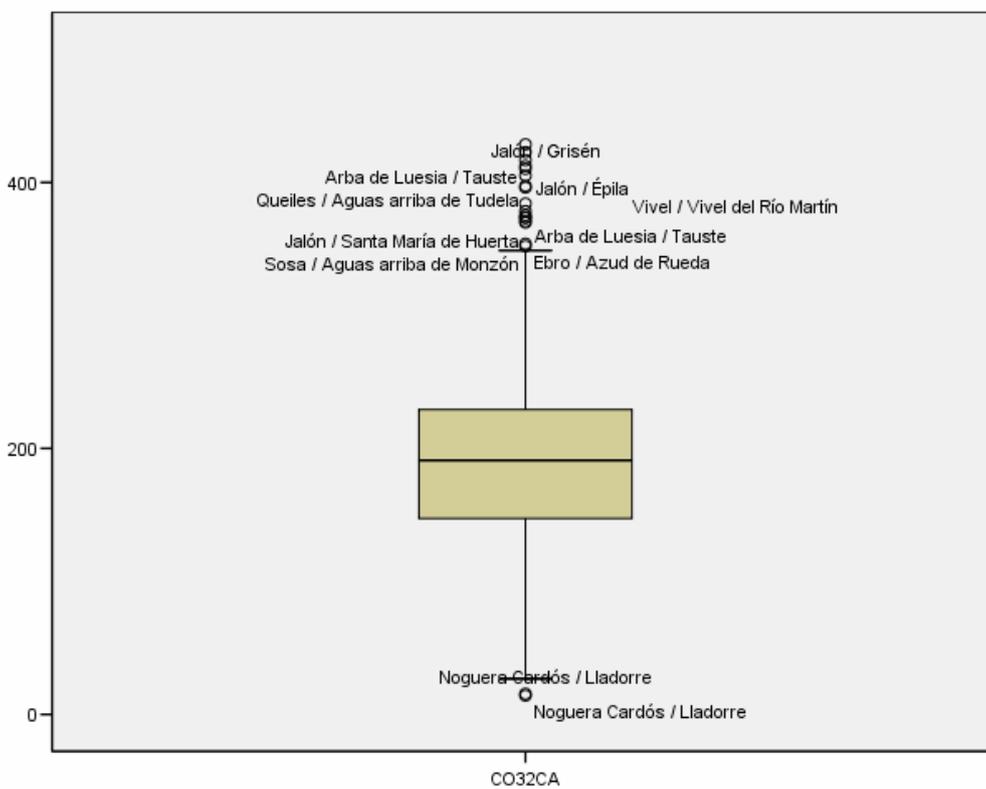
**Tabla 3.2.3.2.5: Pruebas de normalidad de la variable CO32CA**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CO32CA	,047	846	,000	,988	846	,000

Por todas estas razones no se realiza ningún tipo de transformación de esta variable utilizándose sus valores originales para la determinación de sus umbrales.



**Figura 3.2.3.2.1: Histograma de la variable CO32CA**



**Figura 3.2.3.2.2: Diagrama de cajas de la variable CO32CA**

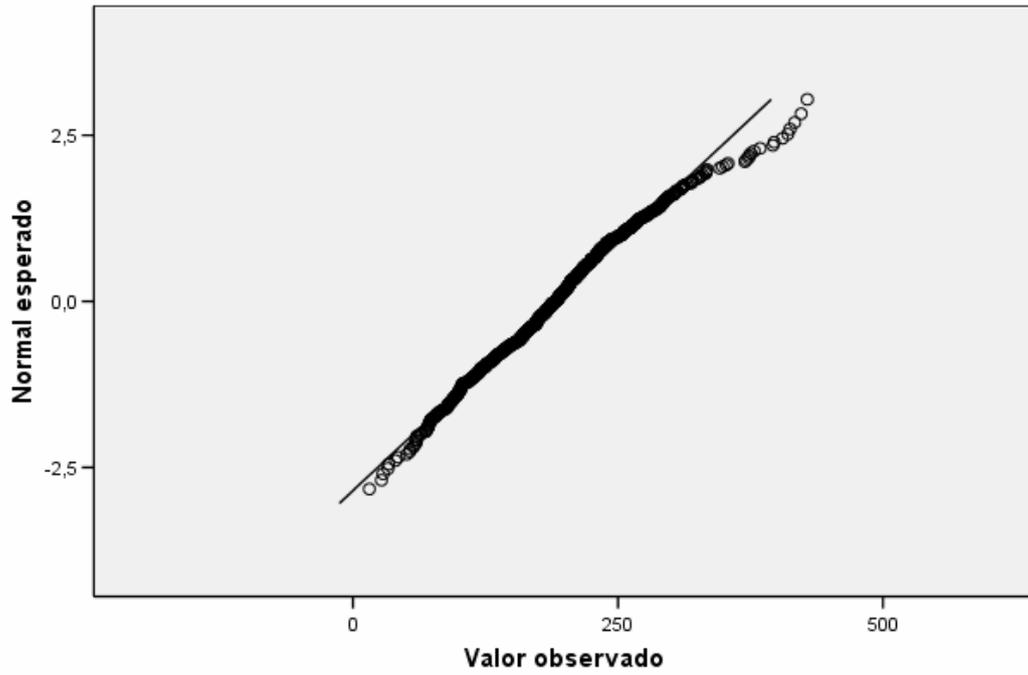


Figura 3.2.3.2.3: QQplot normal de la variable CO32CA

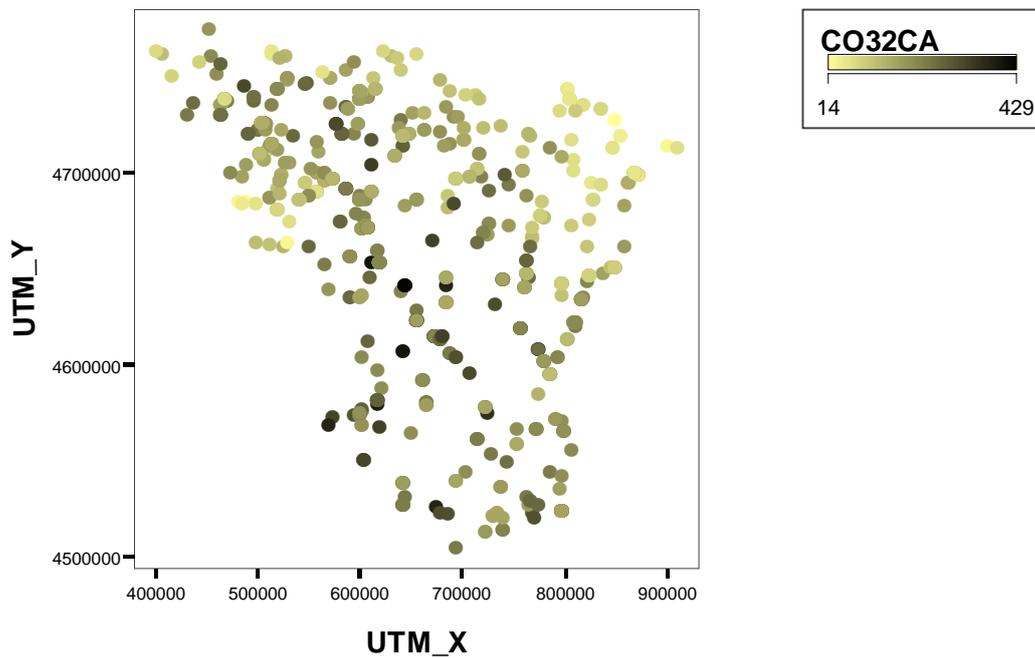


Figura 3.2.3.2.4: Distribución geográfica de los niveles de CO32CA

### 3.2.4. Análisis estadístico de los indicadores de salinidad

En este apartado presentamos el análisis estadístico-descriptivo de los indicadores de salinidad del sodio (NA), calcio (CA), magnesio (MG), cloruros (CL) y sulfatos (SO4).

#### 3.2.4.1 Análisis estadístico de los niveles de sodio

En las Tablas 3.2.4.1.1 a 3.2.4.1.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.4.1.1) que hay un alto porcentaje (79,1%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55,3% de los datos que tienen valores ausentes en la variable NA cumplen los estándares de calidad frente a un 47,7% de los datos observados (ver Tabla 3.2.4.1.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución*. Si se hace el estudio por años (ver Tabla 3.2.4.1.3) no se observan diferencias excesivamente importantes, salvo una disminución fuerte del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.4.1.1: Porcentaje de datos ausentes de la variable NA**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1386	20,9%	5249	79,1%	6635	100,0%

**Tabla 3.2.4.1.2: Estudio de los datos ausentes de la variable NA por nivel de calidad**

			miss_NA		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	646	2879	3525
		% de calidad_conjunta	18,3%	81,7%	100,0%
		% de miss_NA	47,7%	55,3%	53,8%
		Residuos tipificados	-3,1	1,6	
	No cumple	Recuento	709	2324	3033
		% de calidad_conjunta	23,4%	76,6%	100,0%
		% de miss_NA	52,3%	44,7%	46,2%
		Residuos tipificados	3,3	-1,7	
Total		Recuento	1355	5203	6558
		% de calidad_conjunta	20,7%	79,3%	100,0%
		% de miss_NA	100,0%	100,0%	100,0%

**Tabla 3.2.4.1.3: Estudio de los datos ausentes de la variable NA por años**

			miss_NA		Total
			Presente	Ausente	
año	2002	Recuento	279	1062	1341
		% de año	20,8%	79,2%	100,0%
		% de miss_NA	20,1%	20,2%	20,2%
		Residuos tipificados	-,1	,0	
	2003	Recuento	147	852	999
		% de año	14,7%	85,3%	100,0%
		% de miss_NA	10,6%	16,2%	15,1%
		Residuos tipificados	-4,3	2,2	
	2004	Recuento	101	540	641
		% de año	15,8%	84,2%	100,0%
		% de miss_NA	7,3%	10,3%	9,7%
		Residuos tipificados	-2,8	1,5	
	2005	Recuento	206	1106	1312
		% de año	15,7%	84,3%	100,0%
		% de miss_NA	14,9%	21,1%	19,8%
		Residuos tipificados	-4,1	2,1	
	2006	Recuento	194	1089	1283
		% de año	15,1%	84,9%	100,0%
		% de miss_NA	14,0%	20,7%	19,3%
		Residuos tipificados	-4,5	2,3	
	2007	Recuento	459	600	1059
		% de año	43,3%	56,7%	100,0%
		% de miss_NA	33,1%	11,4%	16,0%
		Residuos tipificados	16,0	-8,2	
Total		Recuento	1386	5249	6635
		% de año	20,9%	79,1%	100,0%
		% de miss_NA	100,0%	100,0%	100,0%

En las Tablas 3.2.4.1.4 y 3.2.4.1.5 así como en las Figuras 3.2.4.1.1 a 3.2.4.1.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de sodio en las aguas de la cuenca es igual a 88,44 mg/L Na con un nivel mediano igual a 34,15 mg/L Na (ver Tabla 3.2.4.1.4). La distribución de esta variable presenta, además, una fuerte asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.4.1.4) reflejando la *existencia de puntos de muestreo con un alto nivel de sodio* (ver Figuras 3.2.4.1.1 y 3.2.4.1.2). *Dichos puntos tienden a concentrarse a lo largo del cauce del río Ebro y en las zonas altas de la cuenca* (ver Figura 3.2.4.1.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.4.1.5 y Figura 3.2.4.1.3) siendo necesario realizar una *transformación logarítmica (variable LNA) para aumentar, de forma significativa, su grado de*

*normalidad* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.4.1.4 a 3.2.4.1.6.

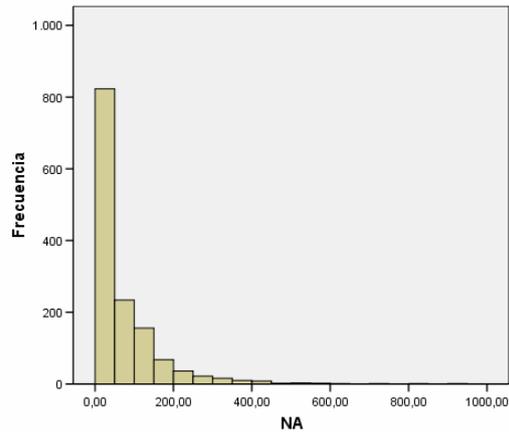
**Tabla 3.2.4.1.4: Análisis estadístico-descriptivo de las variables NA y LNA**

			Estadístico	Error típ.
NA	Media		88,4411	14,11415
	Intervalo de confianza para la media al 95%	Límite inferior	60,7537	
		Límite superior	116,1286	
	Media recortada al 5%		56,1508	
	Mediana		34,1500	
	Varianza		276104,092	
	Desv. típ.		525,45608	
	Mínimo		1,00	
	Máximo		15070,00	
	Rango		15069,00	
	Amplitud intercuartil		86,93	
	Asimetría		25,518	,066
	Curtosis		677,532	,131
	LNA	Media		3,4057
Intervalo de confianza para la media al 95%		Límite inferior	3,3299	
		Límite superior	3,4815	
Media recortada al 5%			3,4176	
Mediana			3,5308	
Varianza			2,069	
Desv. típ.			1,43832	
Mínimo			,00	
Máximo			9,62	
Rango			9,62	
Amplitud intercuartil			2,37	
Asimetría			-,106	,066
Curtosis			-,430	,131

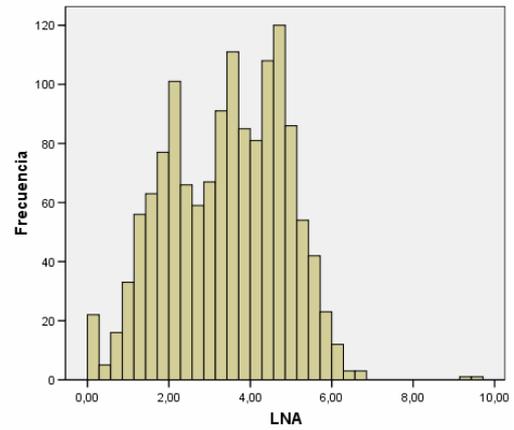
**Tabla 3.2.4.1.5: Pruebas de normalidad de las variables NA y LNA**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
NA	,434	1386	,000	,072	1386	,000
LNA	,061	1386	,000	,981	1386	,000

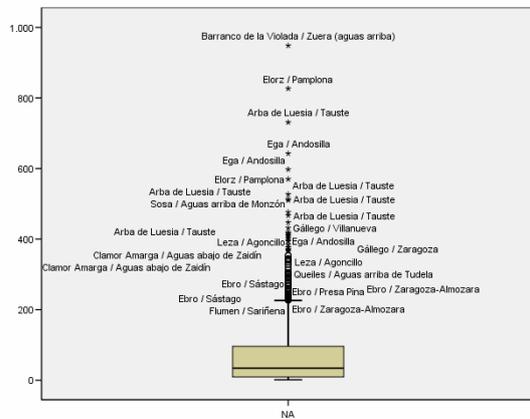
Por estas razones *la determinación de umbrales se llevará a cabo a partir de la variable LNA.*



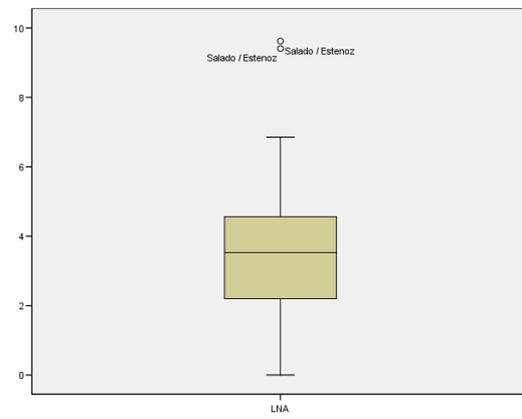
**Figura 3.2.4.1.1: Histograma de la variable NA**



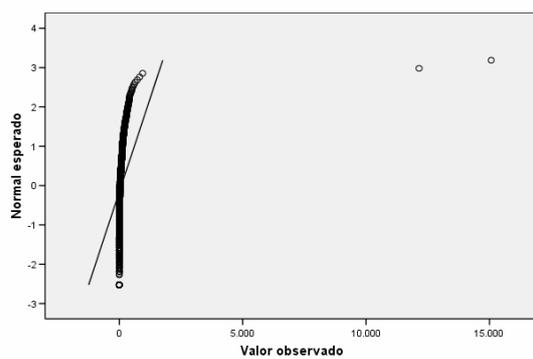
**Figura 3.2.4.1.4 : Histograma de la variable LNA**



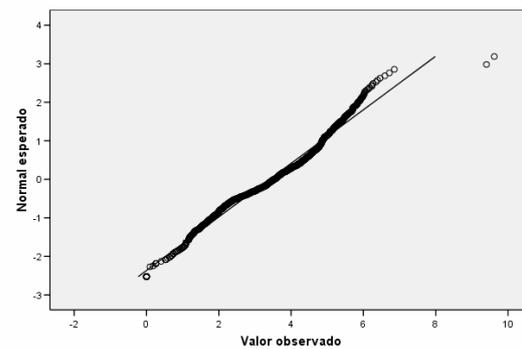
**Figura 3.2.4.1.2: Diagrama de cajas de la variable NA**



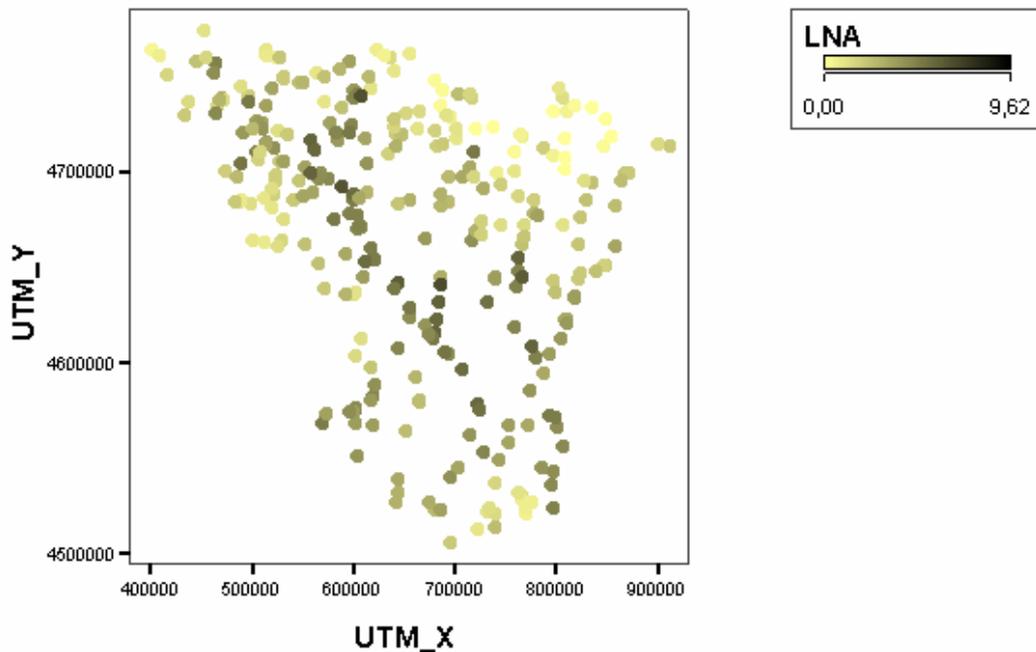
**Figura 3.2.4.1.5: Diagrama de cajas de la variable LNA**



**Figura 3.2.4.1.3: QQplot normal de la variable NA**



**Figura 3.2.4.1.6: QQplot normal de la variable LNA**



**Figura 3.2.4.1.7: Distribución geográfica de los niveles de LNA**

**3.2.4.2. Análisis estadístico del nivel de calcio**

En las Tablas 3.2.4.2.1 a 3.2.4.2.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.4.2.1) que hay un alto porcentaje (84%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55.4% de los datos que tienen valores ausentes en la variable CA cumplen los estándares de calidad frente a un 45.1% de los datos observados (ver Tabla 3.2.4.2.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución*. Si se hace el estudio por años (ver Tabla 3.2.4.2.3) no se observan diferencias excesivamente importantes, salvo una leve tendencia a aumentar el porcentaje de datos ausentes a partir del año 2003.

**Tabla 3.2.4.2.1: Porcentaje de datos ausentes de la variable CA**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1061	16,0%	5574	84,0%	6635	100,0%

**Tabla 3.2.4.2.2: Estudio de los datos ausentes de la variable CA por nivel de calidad**

			miss_ca		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	471	3054	3525
		% de calidad_conjunta	13,4%	86,6%	100,0%
		% de miss_ca	45,1%	55,4%	53,8%
		Residuos tipificados	-3,8	1,7	
	No cumple	Recuento	574	2459	3033
		% de calidad_conjunta	18,9%	81,1%	100,0%
		% de miss_ca	54,9%	44,6%	46,2%
		Residuos tipificados	4,1	-1,8	
Total	Recuento	1045	5513	6558	
	% de calidad_conjunta	15,9%	84,1%	100,0%	
	% de miss_ca	100,0%	100,0%	100,0%	

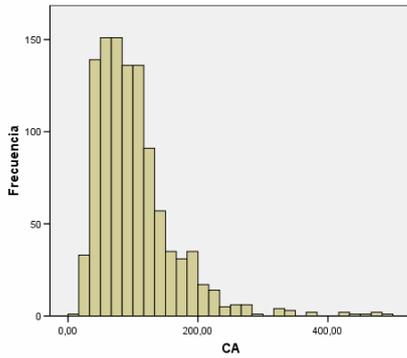
**Tabla 3.2.4.2.3: Estudio de los datos ausentes de la variable CA por año**

			miss_ca		Total
			Presente	Ausente	
año 2002	Recuento	270	1071	1341	
	% de año	20,1%	79,9%	100,0%	
	% de miss_ca	25,4%	19,2%	20,2%	
	Residuos tipificados	3,8	-1,7		
2003	Recuento	141	858	999	
	% de año	14,1%	85,9%	100,0%	
	% de miss_ca	13,3%	15,4%	15,1%	
	Residuos tipificados	-1,5	,6		
2004	Recuento	96	545	641	
	% de año	15,0%	85,0%	100,0%	
	% de miss_ca	9,0%	9,8%	9,7%	
	Residuos tipificados	-,6	,3		
2005	Recuento	194	1118	1312	
	% de año	14,8%	85,2%	100,0%	
	% de miss_ca	18,3%	20,1%	19,8%	
	Residuos tipificados	-1,1	,5		
2006	Recuento	176	1107	1283	
	% de año	13,7%	86,3%	100,0%	
	% de miss_ca	16,6%	19,9%	19,3%	
	Residuos tipificados	-2,0	,9		
2007	Recuento	184	875	1059	
	% de año	17,4%	82,6%	100,0%	
	% de miss_ca	17,3%	15,7%	16,0%	
	Residuos tipificados	1,1	-,5		
Total	Recuento	1061	5574	6635	
	% de año	16,0%	84,0%	100,0%	
	% de miss_ca	100,0%	100,0%	100,0%	

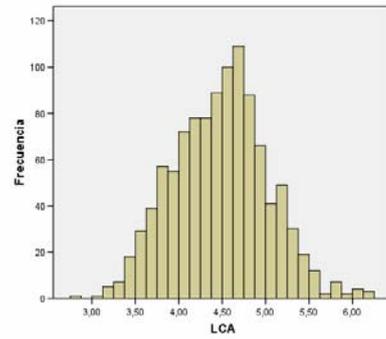
En las Tablas 3.2.4.2.4 y 3.2.4.2.5 así como en las Figuras 3.2.4.2.1 a 3.2.4.2.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de calcio en las aguas de la cuenca es igual a 102,52 mg/L Ca con un nivel mediano igual a 90,10 mg/L Ca (ver Tabla 3.2.4.2.4). La distribución de esta variable presenta, sin embargo, una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.4.2.4) reflejando la *existencia de puntos de muestreo con un alto nivel de calcio* (ver Figuras 3.2.4.2.1 y 3.2.4.2.2) *los cuales tienden a concentrarse a lo largo del cauce del río Ebro y en las zonas bajas de la cuenca* (ver Figura 3.2.4.2.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.4.2.5 y Figura 3.2.4.2.3) siendo *necesario realizar una transformación logarítmica (variable LCA)* para conseguir que se verifique dicha hipótesis tal y como lo demuestran los resultados expuestos en las Tabla 3.2.4.2.5 y 3.2.4.2.6 y en las Figuras 3.2.4.2.4 a 3.2.4.2.6.

**Tabla 3.2.4.2.4: Análisis estadístico-descriptivo de las variables CA y LCA**

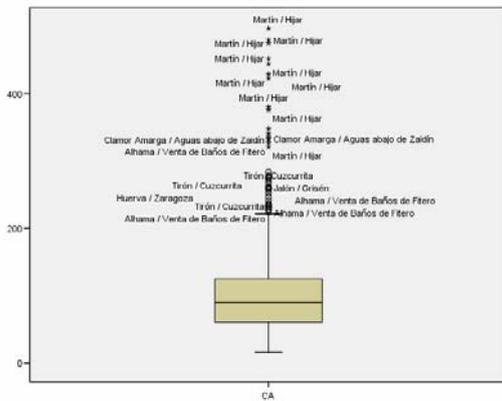
			Estadístico	Error típ.	
CA	Media		102,5187	1,90299	
	Intervalo de confianza para la media al 95%	Límite inferior	98,7846		
		Límite superior	106,2527		
	Media recortada al 5%		96,2872		
	Mediana		90,1000		
	Varianza		3842,260		
	Desv. típ.		61,98596		
	Mínimo		16,20		
	Máximo		497,00		
	Rango		480,80		
	Amplitud intercuartil		64,65		
	Asimetría		2,177		,075
	Curtosis		7,918		,150
	LCA	Media			4,4800
Intervalo de confianza para la media al 95%		Límite inferior	4,4473		
		Límite superior	4,5128		
Media recortada al 5%			4,4754		
Mediana			4,5009		
Varianza			,295		
Desv. típ.			,54317		
Mínimo			2,79		
Máximo			6,21		
Rango			3,42		
Amplitud intercuartil			,73		
Asimetría			,101	,075	
Curtosis			-,051	,150	



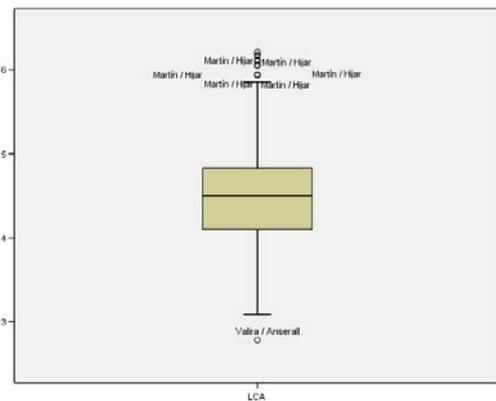
**Figura 3.2.4.2.1: Histograma de la variable CA**



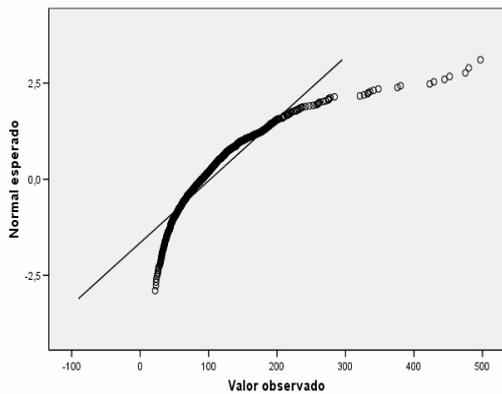
**Figura 3.2.4.2.4: Histograma de la variable LCA**



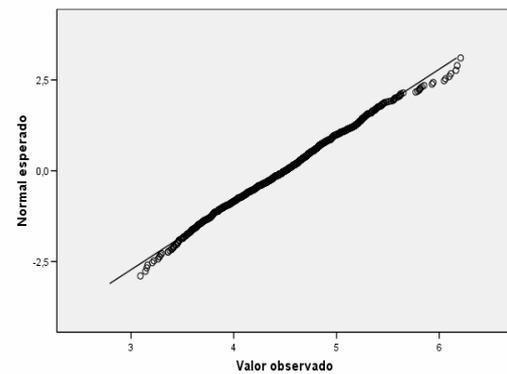
**Figura 3.2.4.2.2: Diagrama de cajas de la variable CA**



**Figura 3.2.4.2.5: Diagrama de cajas de la variable LCA**



**Figura 3.2.4.2.3: QQplot normal de la variable CA**



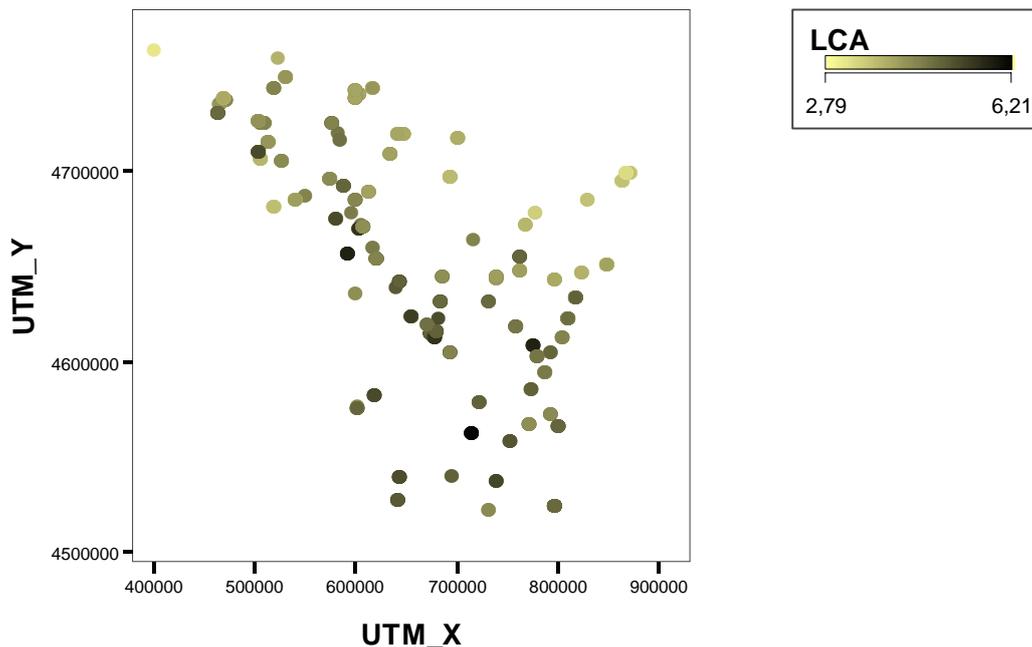
**Figura 3.2.4.2.6: QQplot normal de la variable LCA**

**Tabla 3.2.4.2.5: Pruebas de normalidad de la variable CA**

Kolmogorov-Smirnov			Shapiro-Wilk		
Estadístico	gl	Sig.	Estadístico	gl	Sig.
,116	1061	,000	,833	1061	,000

**Tabla 3.2.4.2.6: Pruebas de normalidad de la variable LCA**

Kolmogorov-Smirnov			Shapiro-Wilk		
Estadístico	gl	Sig.	Estadístico	gl	Sig.
,020	1061	,200	,997	1061	,037



**Figura 3.2.4.2.7: Distribución geográfica de los niveles de la variable LCA**

Por estas razones *la determinación de los umbrales de esta variable se llevará a cabo a partir de la variable LCA.*

### 3.2.4.3 Análisis estadístico del nivel de magnesio

En las Tablas 3.2.4.3.1 a 3.2.4.3.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.4.3.1) que hay un alto porcentaje (79,1%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55,3% de los datos que tienen valores ausentes en la variable MG cumplen los estándares de calidad frente a un 47,7% de los datos observados (ver Tabla 3.2.4.3.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución.* Si se hace el estudio por años (ver Tabla 3.2.4.3.3) no se observan diferencias excesivamente importantes, salvo la disminución notable del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.4.3.1: Porcentaje de datos ausentes de la variable MG**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1386	20,9%	5249	79,1%	6635	100,0%

**Tabla 3.2.4.3.2: Estudio de los datos ausentes de la variable MG por nivel de calidad**

			miss_MG		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	646	2879	3525
		% de calidad_conjunta	18,3%	81,7%	100,0%
		% de miss_MG	47,7%	55,3%	53,8%
		Residuos tipificados	-3,1	1,6	
	No cumple	Recuento	709	2324	3033
		% de calidad_conjunta	23,4%	76,6%	100,0%
		% de miss_MG	52,3%	44,7%	46,2%
		Residuos tipificados	3,3	-1,7	
Total	Recuento	1355	5203	6558	
	% de calidad_conjunta	20,7%	79,3%	100,0%	
	% de miss_MG	100,0%	100,0%	100,0%	

**Tabla 3.2.4.3.3: Estudio de los datos ausentes de la variable MG por años**

			miss_MG		Total
			Presente	Ausente	
año 2002	Recuento	279	1062	1341	
	% de año	20,8%	79,2%	100,0%	
	% de miss_MG	20,1%	20,2%	20,2%	
	Residuos tipificados	-,1	,0		
2003	Recuento	147	852	999	
	% de año	14,7%	85,3%	100,0%	
	% de miss_MG	10,6%	16,2%	15,1%	
	Residuos tipificados	-4,3	2,2		
2004	Recuento	101	540	641	
	% de año	15,8%	84,2%	100,0%	
	% de miss_MG	7,3%	10,3%	9,7%	
	Residuos tipificados	-2,8	1,5		
2005	Recuento	206	1106	1312	
	% de año	15,7%	84,3%	100,0%	
	% de miss_MG	14,9%	21,1%	19,8%	
	Residuos tipificados	-4,1	2,1		
2006	Recuento	194	1089	1283	
	% de año	15,1%	84,9%	100,0%	
	% de miss_MG	14,0%	20,7%	19,3%	
	Residuos tipificados	-4,5	2,3		
2007	Recuento	459	600	1059	
	% de año	43,3%	56,7%	100,0%	
	% de miss_MG	33,1%	11,4%	16,0%	
	Residuos tipificados	16,0	-8,2		
Total	Recuento	1386	5249	6635	
	% de año	20,9%	79,1%	100,0%	
	% de miss_MG	100,0%	100,0%	100,0%	

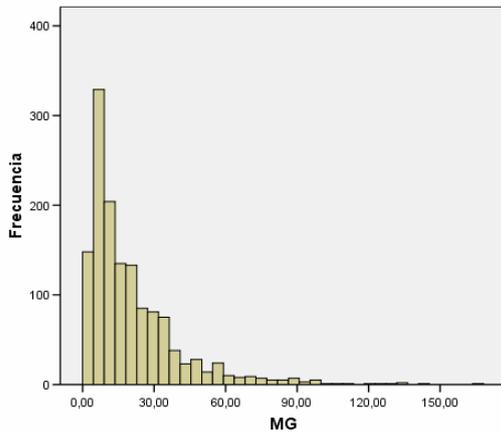
En las Tablas 3.2.4.3.4 y 3.2.4.3.5 así como en las Figuras 3.2.4.3.1 a 3.2.4.3.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de magnesio en las aguas de la cuenca es igual a 20,58 mg/L Mg con un nivel mediano igual a 14,20 mg/L Mg (ver Tabla 3.2.4.3.4). La distribución de esta variable presenta, sin embargo, una fuerte asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.4.3.4) reflejando la *existencia de puntos de muestreo con un alto nivel de magnesio* (ver Figuras 3.2.4.3.1 y 3.2.4.3.2). *Dichos puntos los cuales tienden a concentrarse en la zona suroeste de la cuenca* (ver Figura 3.2.4.3.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.4.3.5 y Figura 3.2.4.3.3) siendo *necesario realizar una transformación logarítmica (variable LMG)* para aumentar significativamente su grado de normalidad tal y como lo demuestran los resultados expuestos en las Figuras 3.2.4.3.4 a 3.2.4.3.6.

**Tabla 3.2.4.3.4: Análisis estadístico-descriptivo de las variables MG y LMG**

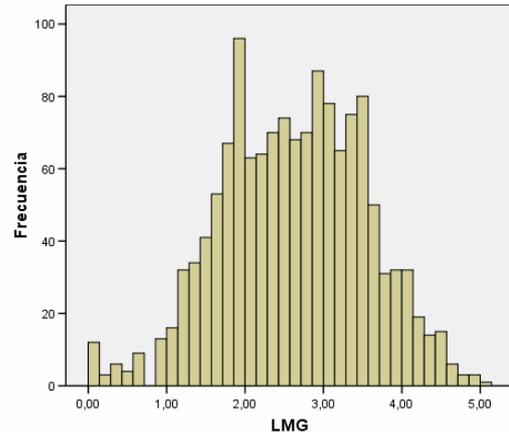
			Estadístico	Error típ.	
MG	Media		20,5850	,53671	
	Intervalo de confianza para la media al 95%	Límite inferior	19,5321		
		Límite superior	21,6378		
	Media recortada al 5%		18,1050		
	Mediana		14,2000		
	Varianza		399,245		
	Desv. típ.		19,98112		
	Mínimo		1,00		
	Máximo		167,00		
	Rango		166,00		
	Amplitud intercuartil		20,63		
	Asimetría		2,343		,066
	Curtosis		7,810		,131
	LMG	Media			2,6268
Intervalo de confianza para la media al 95%		Límite inferior	2,5783		
		Límite superior	2,6752		
Media recortada al 5%			2,6354		
Mediana			2,6532		
Varianza			,845		
Desv. típ.			,91921		
Mínimo			,00		
Máximo			5,12		
Rango			5,12		
Amplitud intercuartil			1,37		
Asimetría			-,136	,066	
Curtosis			-,230	,131	

**Tabla 3.2.4.3.5: Pruebas de normalidad de las variables MG y LMG**

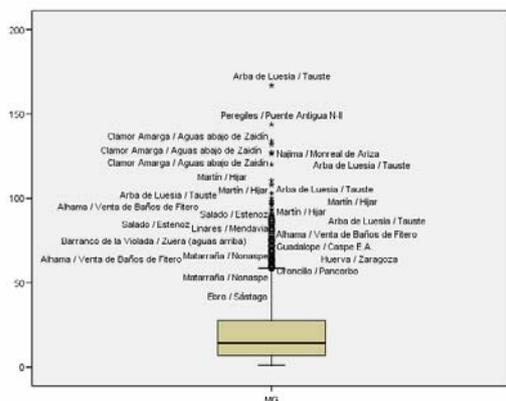
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
MG	,163	1386	,000	,773	1386	,000
LMG	,035	1386	,000	,994	1386	,000



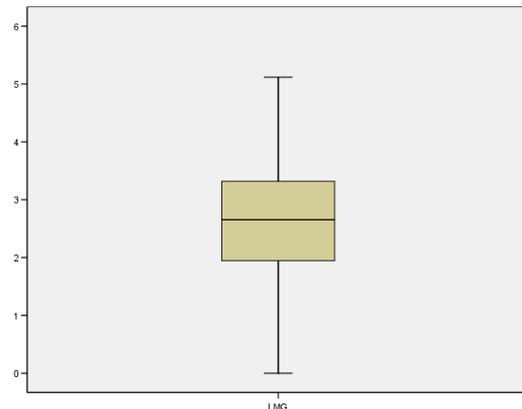
**Figura 3.2.4.3.1: Histograma de la variable MG**



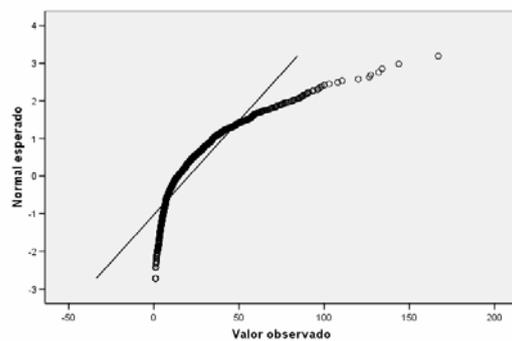
**Figura 3.2.4.3.4 : Histograma de la variable LMG**



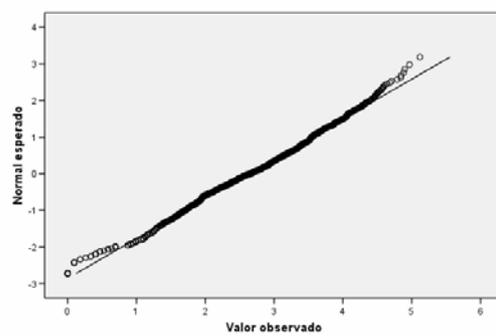
**Figura 3.2.4.3.2: Diagrama de cajas de la variable MG**



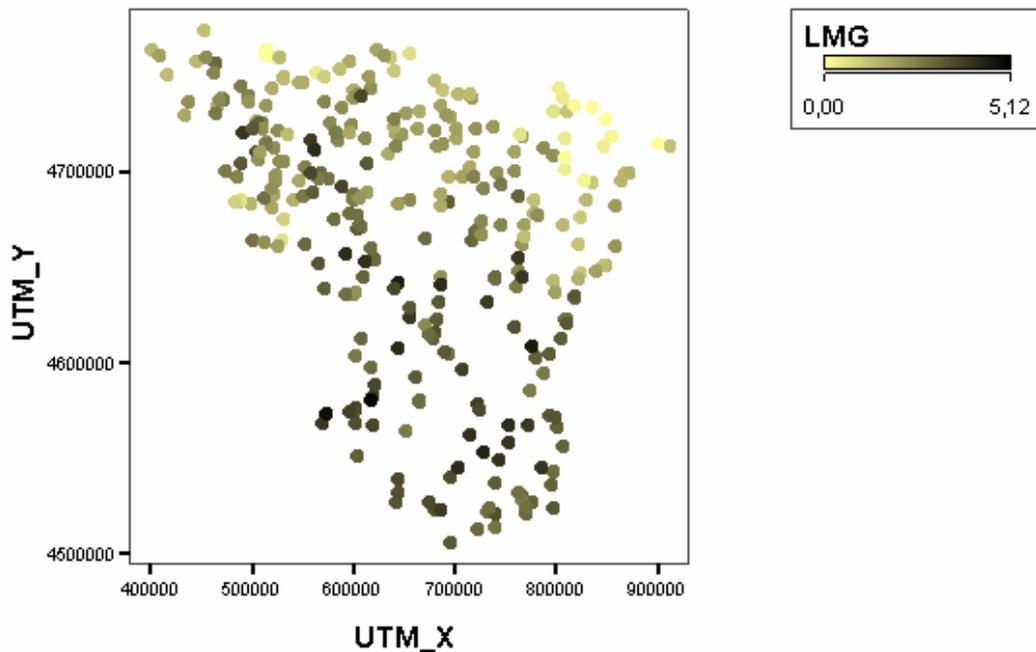
**Figura 3.2.4.3.5 : Diagrama de cajas de la variable LMG**



**Figura 3.2.4.3.3: QQplot normal de la variable MG**



**Figura 3.2.4.3.6: QQplot normal de la variable LMG**



**Figura 3.2.4.3.7: Distribución geográfica de los niveles de LMG**

Por estas razones *la determinación de los umbrales de esta variable se llevará a cabo a partir de la variable LMG.*

### 3.2.4.4. Análisis estadístico de los cloruros

En las Tablas 3.2.4.4.1 a 3.2.4.4.3 e muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.4.4.1) que hay un alto porcentaje (63.8%) así como la existencia de una tendencia significativa a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 47.7% de los datos que tienen valores ausentes en la variable CL no cumplen los estándares de calidad frente a un 43.7% de los datos observados (ver Tabla 3.2.4.4.2).

**Tabla 3.2.4.4.1: Porcentaje de datos ausentes de la variable CL**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
2405	36,2%	4230	63,8%	6635	100,0%

**Tabla 3.2.4.4.2: Estudio de los datos ausentes de la variable CL por nivel de calidad**

			miss_cl		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	1330	2195	3525
		% de calidad_conjunta	37,7%	62,3%	100,0%
		% de miss_cl	56,3%	52,3%	53,8%
		Residuos tipificados	1,7	-1,3	
	No cumple	Recuento	1033	2000	3033
		% de calidad_conjunta	34,1%	65,9%	100,0%
		% de miss_cl	43,7%	47,7%	46,2%
Total	Recuento	2363	4195	6558	
	% de calidad_conjunta	36,0%	64,0%	100,0%	
	% de miss_cl	100,0%	100,0%	100,0%	

Si se analiza el patrón de datos ausentes por años (ver Tabla 3.2.4.4.3) se observa que la tendencia a observar el valor de esta variable ha tendido a crecer, sobre todo en los últimos años alcanzando en el año 2007 un porcentaje de observación (52%) sensiblemente superior a los del resto de los años. *Existe, por tanto, una sobre-representación de los datos que sí cumplen los niveles de calidad mínimos así como de las observaciones del año 2007 con lo que los resultados expuestos respecto a esta variable deberán tomarse con cierta precaución.*

**Tabla 3.2.4.4.3: Estudio de los datos ausentes de la variable CL por años**

			miss_cl		Total
			Presente	Ausente	
año	2002	Recuento	482	859	1341
		% de año	35,9%	64,1%	100,0%
		% de miss_cl	20,0%	20,3%	20,2%
		Residuos tipificados	-,2	,1	
2003	Recuento	369	630	999	
	% de año	36,9%	63,1%	100,0%	
	% de miss_cl	15,3%	14,9%	15,1%	
	Residuos tipificados	,4	-,3		
2004	Recuento	159	482	641	
	% de año	24,8%	75,2%	100,0%	
	% de miss_cl	6,6%	11,4%	9,7%	
	Residuos tipificados	-4,8	3,6		
2005	Recuento	441	871	1312	
	% de año	33,6%	66,4%	100,0%	
	% de miss_cl	18,3%	20,6%	19,8%	
	Residuos tipificados	-1,6	1,2		
2006	Recuento	403	880	1283	
	% de año	31,4%	68,6%	100,0%	
	% de miss_cl	16,8%	20,8%	19,3%	
	Residuos tipificados	-2,9	2,2		
2007	Recuento	551	508	1059	
	% de año	52,0%	48,0%	100,0%	
	% de miss_cl	22,9%	12,0%	16,0%	
	Residuos tipificados	8,5	-6,4		
Total	Recuento	2405	4230	6635	
	% de año	36,2%	63,8%	100,0%	
	% de miss_cl	100,0%	100,0%	100,0%	

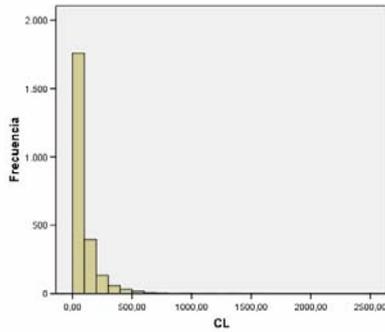
En las Tablas 3.2.4.4.4 y 3.2.4.4.5 así como en las Figuras 3.2.4.4.1 a 3.2.4.4.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de cloruros en las aguas de la cuenca es igual a 104,91 mg/L Cl con un nivel mediano igual a 28,8 mg/L Cl (ver Tabla 3.2.4.4.4). La distribución de esta variable presenta una fuerte asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.4.4.4) reflejando la *existencia de puntos de muestreo con un alto nivel de cloruros* (ver Figuras 3.2.4.4.1 y 3.2.4.4.2). *Dichos puntos tienden a concentrarse a largo del cauce del río Ebro y en las zonas cercanas al mismo* (ver Figura 3.2.4.4.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.4.4.5 y Figura 3.2.4.4.3) haciéndose *necesaria la realización de una transformación logarítmica (variable LCL) para aumentar su grado de normalidad* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.4.4.4 a 3.2.4.4.6.

**Tabla 3.2.4.4.4: Análisis estadístico-descriptivo de las variables CL y LCL**

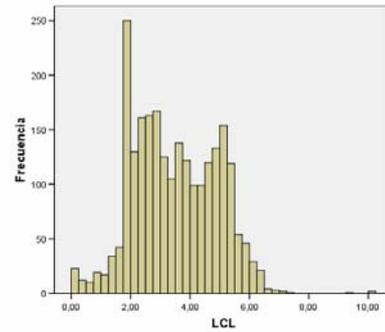
			Estadístico	Error típ.
CL	Media		104,9147	16,14085
	Intervalo de confianza para la media al 95%	Límite inferior	73,2633	
		Límite superior	136,5661	
	Media recortada al 5%		61,8244	
	Mediana		28,8000	
	Varianza		626567,869	
	Desv. típ.		791,56040	
	Mínimo		1,00	
	Máximo		27470,00	
	Rango		27469,00	
	Amplitud intercuartil		99,00	
	Asimetría		30,298	,050
	Curtosis		967,609	,100
	LCL	Media		3,4848
Intervalo de confianza para la media al 95%		Límite inferior	3,4285	
		Límite superior	3,5410	
Media recortada al 5%			3,4815	
Mediana			3,3604	
Varianza			1,979	
Desv. típ.			1,40682	
Mínimo			,00	
Máximo			10,22	
Rango			10,22	
Amplitud intercuartil			2,36	
Asimetría			,170	,050
Curtosis			-,312	,100

**Tabla 3.2.4.4.5: Pruebas de normalidad de las variables CL y LCL**

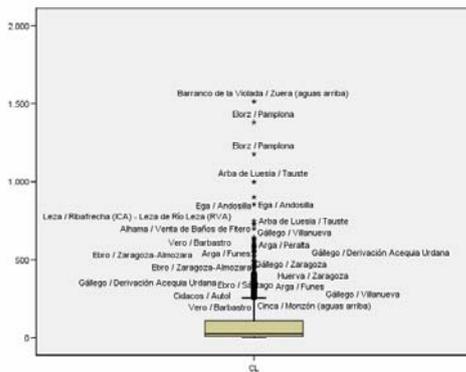
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
CL	,448	2405	,000	,053	2405	,000
LCL	,066	2405	,000	,977	2405	,000



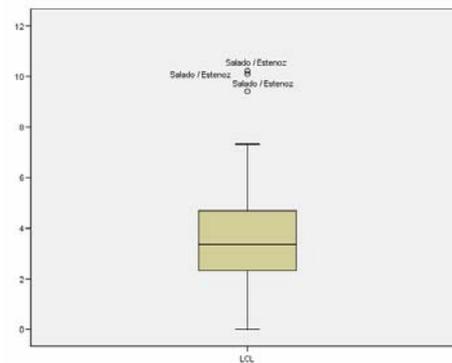
**Figura 3.2.4.4.1: Histograma de la variable CL**



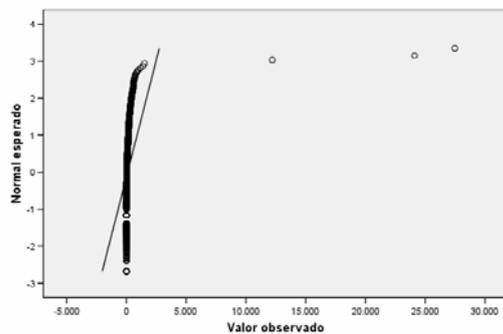
**Figura 3.2.4.4.4: Histograma de la variable LCL**



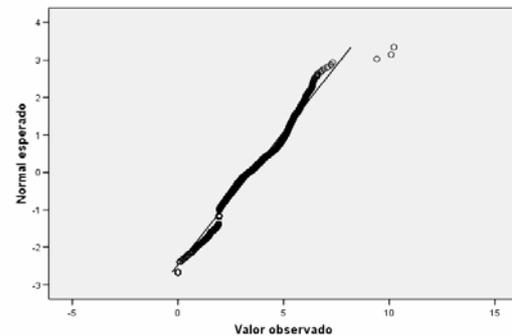
**Figura 3.2.4.4.2: Diagrama de cajas de la variable CL**



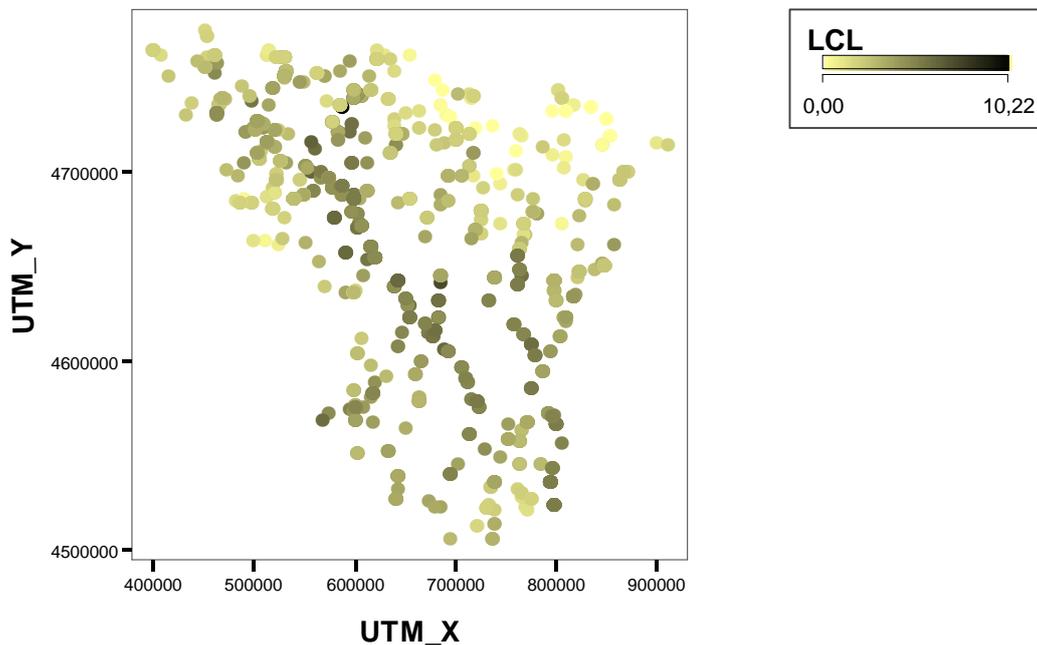
**Figura 3.2.4.4.5: Diagrama de cajas de la variable LCL**



**Figura 3.2.4.4.3: QQplot normal de la variable CL**



**Figura 3.2.4.4.6: QQplot normal de la variable LCL**



**Figura 3.2.4.4.7: Distribución geográfica de los niveles de LCL**

Por estas razones *la determinación de los umbrales de esta variable se llevará a cabo a partir de la variable LCL.*

### 3.2.4.5 Análisis estadístico de los sulfatos

En las Tablas 3.2.4.5.1 a 3.2.4.5.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.4.5.1) que hay un alto porcentaje (72,9%) aunque el reparto es bastante equitativo dado que un 54,4% de los datos que tienen valores ausentes en la variable SO<sub>4</sub> cumplen los estándares de calidad frente a un 51,9% de los datos observados (ver Tabla 3.2.4.5.2). Si se hace el estudio por años (ver Tabla 3.2.4.5.3) no se observan diferencias excesivamente importantes, salvo una disminución notable del porcentaje de datos ausentes en el año 2007. Por todo ello se concluye que *no se aprecia la existencia de una falta de representatividad de la muestra analizada en el estudio para la determinación de los umbrales de esta variable.*

**Tabla 3.2.4.5.1: Porcentaje de datos ausentes de la variable SO4**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1797	27,1%	4838	72,9%	6635	100,0%

**Tabla 3.2.4.5.2: Estudio de los datos ausentes de la variable SO4 por nivel de calidad**

			miss_SO4		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	915	2610	3525
		% de calidad_conjunta	26,0%	74,0%	100,0%
		% de miss_SO4	51,9%	54,4%	53,8%
		Residuos tipificados	-1,0	,6	
	No cumple	Recuento	847	2186	3033
		% de calidad_conjunta	27,9%	72,1%	100,0%
		% de miss_SO4	48,1%	45,6%	46,2%
		Residuos tipificados	1,1	-,7	
Total	Recuento	1762	4796	6558	
	% de calidad_conjunta	26,9%	73,1%	100,0%	
	% de miss_SO4	100,0%	100,0%	100,0%	

**Tabla 3.2.4.5.3: Estudio de los datos ausentes de la variable SO4 por años**

			miss_SO4		Total
			Presente	Ausente	
año	2002	Recuento	357	984	1341
		% de año	26,6%	73,4%	100,0%
		% de miss_SO4	19,9%	20,3%	20,2%
		Residuos tipificados	-,3	,2	
	2003	Recuento	227	772	999
		% de año	22,7%	77,3%	100,0%
		% de miss_SO4	12,6%	16,0%	15,1%
		Residuos tipificados	-2,6	1,6	
	2004	Recuento	126	515	641
		% de año	19,7%	80,3%	100,0%
		% de miss_SO4	7,0%	10,6%	9,7%
		Residuos tipificados	-3,6	2,2	
	2005	Recuento	319	993	1312
		% de año	24,3%	75,7%	100,0%
		% de miss_SO4	17,8%	20,5%	19,8%
		Residuos tipificados	-1,9	1,2	
2006	Recuento	273	1010	1283	
	% de año	21,3%	78,7%	100,0%	
	% de miss_SO4	15,2%	20,9%	19,3%	
	Residuos tipificados	-4,0	2,4		
2007	Recuento	495	564	1059	
	% de año	46,7%	53,3%	100,0%	
	% de miss_SO4	27,5%	11,7%	16,0%	
	Residuos tipificados	12,3	-7,5		
Total	Recuento	1797	4838	6635	
	% de año	27,1%	72,9%	100,0%	
	% de miss_SO4	100,0%	100,0%	100,0%	

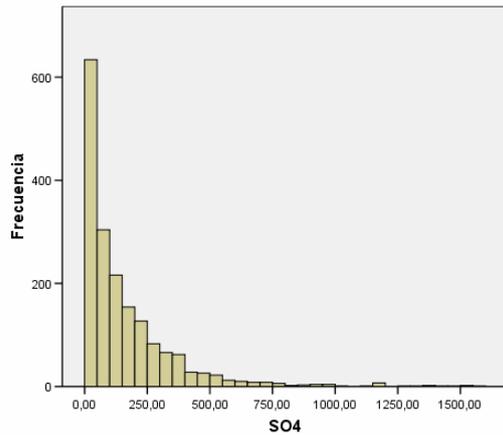
En las Tablas 3.2.4.5.4 y 3.2.4.5.5 así como en las Figuras 3.2.4.5.1 a 3.2.4.5.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de sulfatos en las aguas de la cuenca es igual a 158,94 mg/L SO<sub>4</sub> con un nivel mediano igual a 92 mg/L SO<sub>4</sub> (ver Tabla 3.2.4.5.4). La distribución de esta variable presenta, una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.4.5.4) reflejando la *existencia de puntos de muestreo con un alto nivel de sulfatos* (ver Figuras 3.2.4.5.1 y 3.2.4.5.2). *Estos puntos tienden a concentrarse a lo largo del cauce y en las zonas bajas de la cuenca* (ver Figura 3.2.4.5.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.4.5.5 y Figura 3.2.4.5.3) siendo necesario realizar una transformación logarítmica (variable LSO4) con el fin de aumentar significativamente su grado de normalidad tal y como lo demuestran los resultados expuestos en las Figuras 3.2.4.5.4 a 3.2.4.5.6.

**Tabla 3.2.4.5.4: Análisis estadístico-descriptivo de las variables SO<sub>4</sub> y LSO<sub>4</sub>**

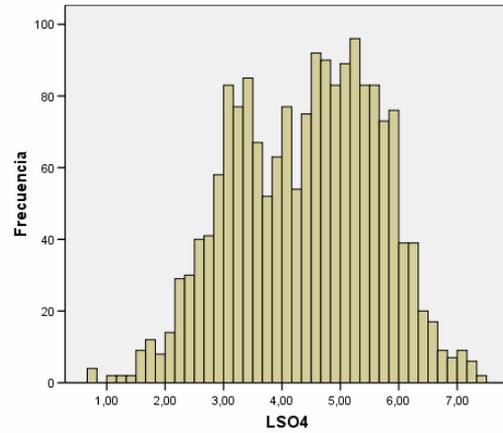
			Estadístico	Error típ.	
SO <sub>4</sub>	Media		158,9359	4,67278	
	Intervalo de confianza para la media al 95%	Límite inferior	149,7712		
		Límite superior	168,1006		
	Media recortada al 5%		131,6773		
	Mediana		92,0000		
	Varianza		39237,346		
	Desv. típ.		198,08419		
	Mínimo		2,00		
	Máximo		1584,00		
	Rango		1582,00		
	Amplitud intercuartil		181,95		
	Asimetría		2,948		,058
	Curtosis		12,336		,115
	LSO <sub>4</sub>	Media			4,4045
Intervalo de confianza para la media al 95%		Límite inferior	4,3475		
		Límite superior	4,4614		
Media recortada al 5%			4,4179		
Mediana			4,5218		
Varianza			1,514		
Desv. típ.			1,23056		
Mínimo			,69		
Máximo			7,37		
Rango			6,67		
Amplitud intercuartil			1,96		
Asimetría			-,179	,058	
Curtosis			-,611	,115	

**Tabla 3.2.4.5.5: Pruebas de normalidad de las variables SO4 y LSO4**

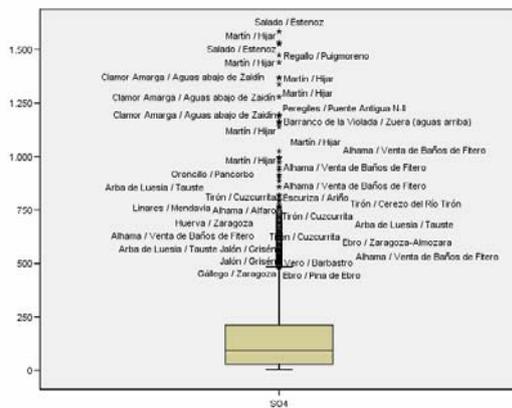
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
SO4	,214	1797	,000	,701	1797	,000
LSO4	,049	1797	,000	,988	1797	,000



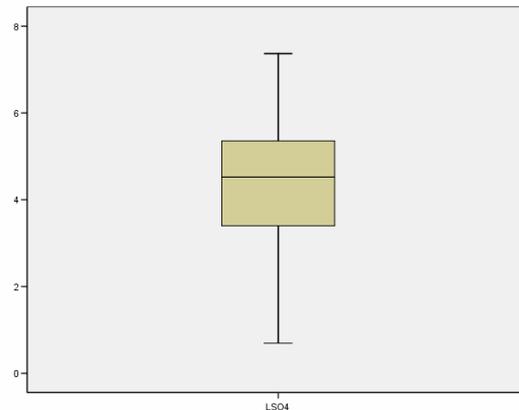
**Figura 3.2.4.5.1: Histograma de la variable SO4**



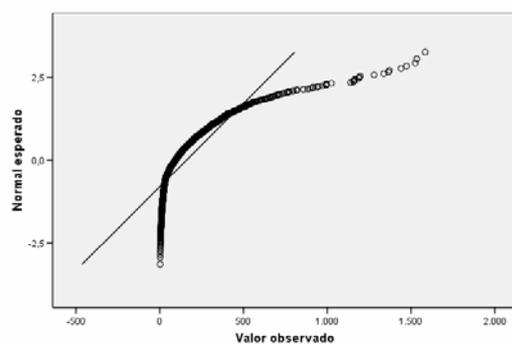
**Figura 3.2.4.5.4: Histograma de la variable LSO4**



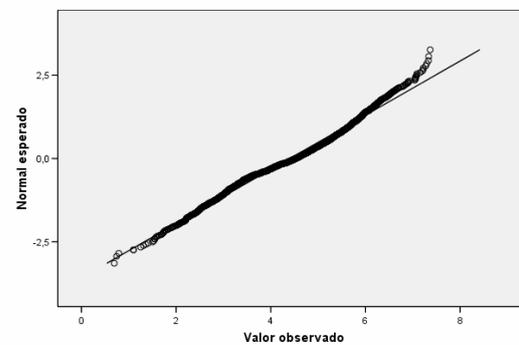
**Figura 3.2.4.5.2: Diagrama de cajas de la variable SO4**



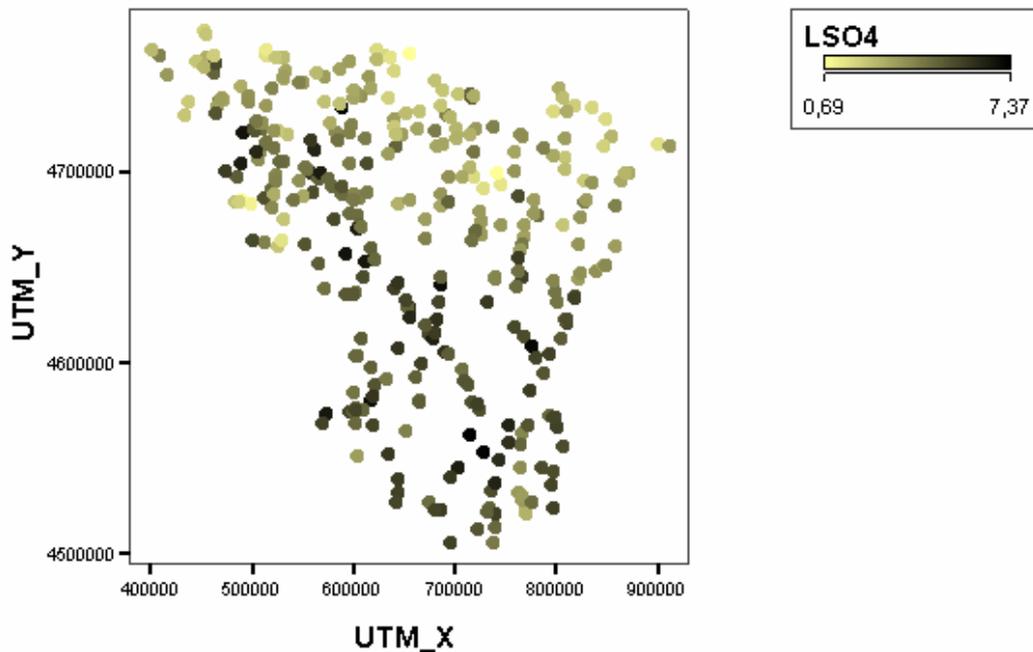
**Figura 3.2.4.5.5: Diagrama de cajas de la variable LSO4**



**Figura 3.2.4.5.3: QQplot normal de la variable SO4**



**Figura 3.2.4.5.6: QQplot normal de la variable LSO4**



**Figura 3.2.4.5.7: Distribución geográfica de los niveles de LSO4**

Por estas razones *la determinación de los umbrales de esta variable se llevará a cabo a partir de la variable LSO4.*

### **3.2.5. Análisis estadístico de los indicadores de nutrientes**

En este apartado presentamos el análisis estadístico-descriptivo de los indicadores de nutrientes NH<sub>4</sub>, NO<sub>2</sub>, NO<sub>3</sub>, N\_KJED, PO<sub>4</sub> y P\_TOT cuyo significado puede encontrarse en la Tabla 2.3.

#### **3.2.5.1 Análisis estadístico de los niveles de amonio**

En las Tablas 3.2.5.1.1 a 3.2.5.1.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.1.1) que hay un alto porcentaje (46,9%) así como la existencia de una tendencia significativa a concentrarse entre los datos que no cumplen los niveles de calidad conjunta dado que un 48,4% de los datos que tienen valores ausentes en la variable NH<sub>4</sub> no cumplen los estándares de calidad frente a un 44,3% de los datos observados (ver Tabla 3.2.5.1.2). Si se hace el estudio por años (ver Tabla 3.2.5.1.3) no se observan diferencias excesivamente importantes, siendo 2004 el año con mayor porcentaje de datos ausentes y 2007 el año con menos. Existe, por tanto, una *tendencia de los puntos de muestreo*

*que cumplen los niveles de calidad a estar sobre-representados lo cual disminuye el grado de representatividad de la muestra analizada y, en consecuencia, los resultados expuestos deberán tomarse con precaución.*

**Tabla 3.2.5.1.1: Porcentaje de datos ausentes de la variable NH4**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
3522	53,1%	3113	46,9%	6635	100,0%

**Tabla 3.2.5.1.2: Estudio de los datos ausentes de la variable NH4 por nivel de calidad**

			miss_NH4		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	1928	1597	3525
		% de calidad_conjunta	54,7%	45,3%	100,0%
		% de miss_NH4	55,7%	51,6%	53,8%
		Residuos tipificados	1,5	-1,6	
	No cumple	Recuento	1535	1498	3033
		% de calidad_conjunta	50,6%	49,4%	100,0%
		% de miss_NH4	44,3%	48,4%	46,2%
		Residuos tipificados	-1,7	1,8	
Total		Recuento	3463	3095	6558
		% de calidad_conjunta	52,8%	47,2%	100,0%
		% de miss_NH4	100,0%	100,0%	100,0%

**Tabla 3.2.5.1.3: Estudio de los datos ausentes de la variable NH4 por años**

			miss_NH4		Total
			Presente	Ausente	
año 2002	Recuento		580	761	1341
	% de año		43,3%	56,7%	100,0%
	% de miss_NH4		16,5%	24,4%	20,2%
	Residuos tipificados		-4,9	5,3	
2003	Recuento		510	489	999
	% de año		51,1%	48,9%	100,0%
	% de miss_NH4		14,5%	15,7%	15,1%
	Residuos tipificados		-,9	,9	
2004	Recuento		255	386	641
	% de año		39,8%	60,2%	100,0%
	% de miss_NH4		7,2%	12,4%	9,7%
	Residuos tipificados		-4,6	4,9	
2005	Recuento		701	611	1312
	% de año		53,4%	46,6%	100,0%
	% de miss_NH4		19,9%	19,6%	19,8%
	Residuos tipificados		,2	-,2	
2006	Recuento		741	542	1283
	% de año		57,8%	42,2%	100,0%
	% de miss_NH4		21,0%	17,4%	19,3%
	Residuos tipificados		2,3	-2,4	
2007	Recuento		735	324	1059
	% de año		69,4%	30,6%	100,0%
	% de miss_NH4		20,9%	10,4%	16,0%
	Residuos tipificados		7,3	-7,8	
Total		Recuento	3522	3113	6635
		% de año	53,1%	46,9%	100,0%
		% de miss_NH4	100,0%	100,0%	100,0%

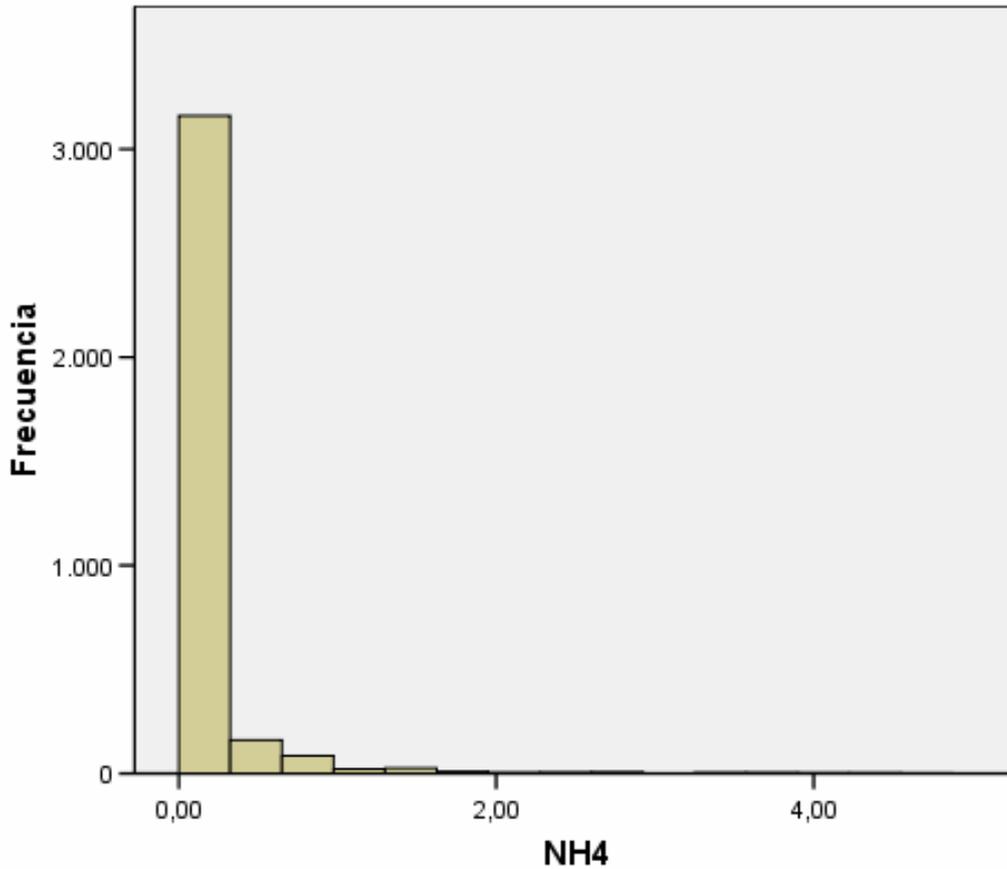
En las Tablas 3.2.5.1.4 y 3.2.5.1.5 así como en las Figuras 3.2.5.1.1 a 3.2.5.1.8 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. La distribución de esta variable tiene un claro *carácter mixto* (ver Figura 3.2.5.1.1) *debido a la presencia de un 81,7% de casos con información censurada en los que el nivel de amonio fue menor que 0,13 mg/L NH<sub>4</sub>*. Para el resto de los casos el nivel medio de amonio en las aguas de la cuenca es igual a 0,96 mg/L NH<sub>4</sub> con un nivel mediano igual a 0,38 mg/L NH<sub>4</sub> (ver Tabla 3.2.5.4.4). La distribución de la variable en los casos no censurados presenta, además, una asimetría positiva y una leptocurtosis significativas (ver Tabla 3.2.5.1.4) reflejando la *existencia de puntos de muestreo con un alto nivel de amonio* (ver Figuras 3.2.5.1.2 y 3.2.5.1.3). *Dichos puntos tienden a concentrarse en la zona alta y media del cauce del río Ebro* (ver Figura 3.2.5.1.8). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.1.5 y Figura 3.2.5.1.4) siendo *necesario realizar una transformación logarítmica (variable LN<sub>NH4</sub>) para aumentar su grado de normalidad* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.1.5 a 3.2.5.1.7.

**Tabla 3.2.5.1.4: Análisis estadístico-descriptivo de las variables NH<sub>4</sub> y LN<sub>NH4</sub> si NH<sub>4</sub>>0,13 mg/L NH<sub>4</sub>**

			Estadístico	Error típ.	
NH <sub>4</sub>	Media		,9559	,08052	
	Intervalo de confianza para la media al 95%	Límite inferior	,7978		
		Límite superior	1,1140		
	Media recortada al 5%		,5991		
	Mediana		,3800		
	Varianza		4,188		
	Desv. típ.		2,04652		
	Mínimo		,14		
	Máximo		24,25		
	Rango		24,11		
	Amplitud intercuartil		,56		
	Asimetría		5,982		,096
	Curtosis		46,543		,192
	LN <sub>NH4</sub>	Media			-,7749
Intervalo de confianza para la media al 95%		Límite inferior	-,8532		
		Límite superior	-,6966		
Media recortada al 5%			-,8617		
Mediana			-,9676		
Varianza			1,026		
Desv. típ.			1,01305		
Mínimo			-1,97		
Máximo			3,19		
Rango			5,15		
Amplitud intercuartil			1,34		
Asimetría			1,150	,096	
Curtosis			1,181	,192	

**Tabla 3.2.5.1.5: Pruebas de normalidad de las variables NH4 y LNH4 si NH4>0,13 mg/L NH4**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
NH4	,345	646	,000	,385	646	,000
LNH4	,120	646	,000	,902	646	,000



**Figura 3.2.5.1.1: Histograma de la variable NH4 para todos los casos**

*Por todas estas razones la determinación de los umbrales de la variable NH4 se realizará a partir de la variable LNH4 así como de una variable indicador INH4 que toma el valor 1 si  $NH4 \leq 0,13$  mg/L NH4 (es decir si el dato está censurado) y 0 en caso contrario.*

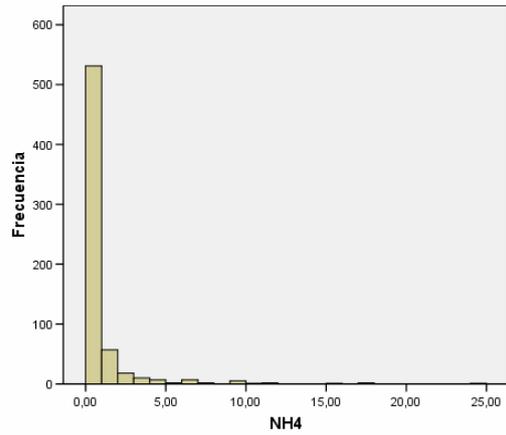


Figura 3.2.5.1.2: Histograma de la variable NH4 si  $NH_4 > 0,13$  mg/L  $NH_4$

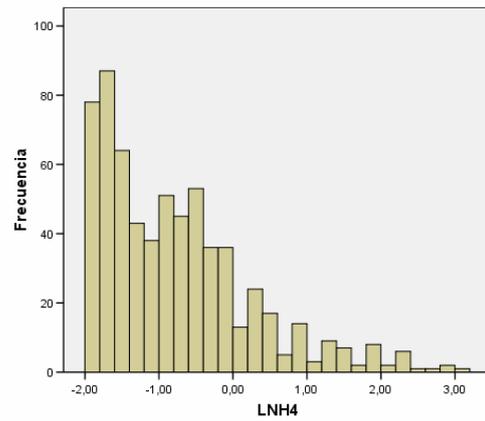


Figura 3.2.5.1.5 : Histograma de la variable LNH4 si  $NH_4 > 0,13$  mg/L  $NH_4$

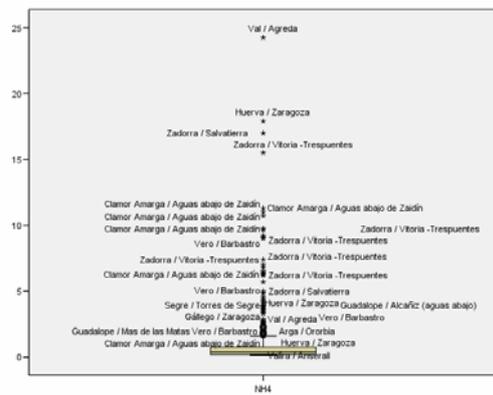


Figura 3.2.5.1.3: Diagrama de cajas de la variable NH4 si  $NH_4 > 0,13$  mg/L  $NH_4$

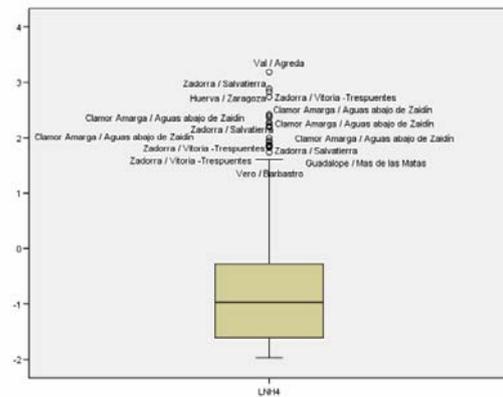


Figura 3.2.5.1.6 : Diagrama de cajas de la variable LNH4 si  $NH_4 > 0,13$  mg/L  $NH_4$

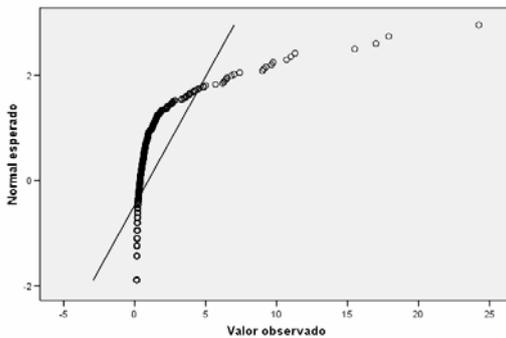


Figura 3.2.5.1.4: QQplot normal de la variable NH4 si  $NH_4 > 0,13$  mg/L  $NH_4$

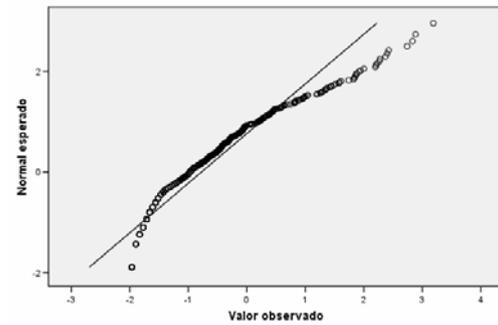
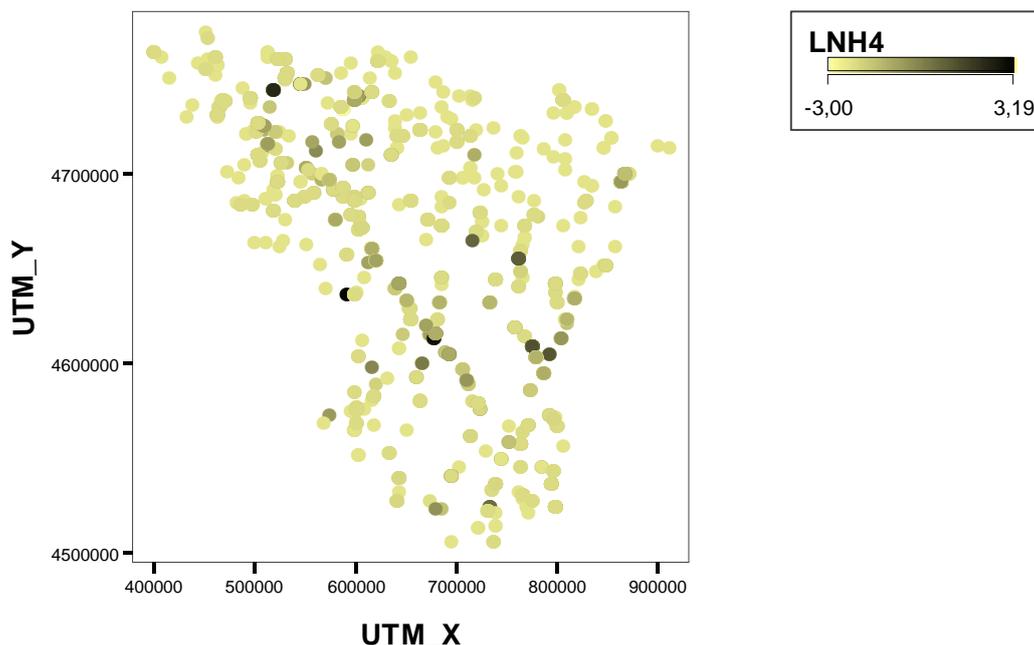


Figura 3.2.5.1.7: QQplot normal de la variable LNH4 si  $NH_4 > 0,13$  mg/L  $NH_4$



**Figura 3.2.5.1.8: Distribución geográfica de los niveles de LNH4**

### 3.2.5.2. Análisis estadístico de los nitritos

En las Tablas 3.2.5.2.1 a 3.2.5.2.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.2.1) que hay un alto porcentaje (72,2%) así como la existencia de una tendencia a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 54,7% de los datos que tienen valores ausentes en la variable NO<sub>2</sub> cumplen los estándares de calidad frente a un 51,4% de los datos observados (ver Tabla 3.2.5.2.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución*. Si se hace el estudio por años (ver Tabla 3.2.5.2.3) no se observan diferencias excesivamente importantes, salvo una disminución del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.5.2.1: Porcentaje de datos ausentes de la variable NO<sub>2</sub>**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1842	27,8%	4793	72,2%	6635	100,0%

**Tabla 3.2.5.2.2: Estudio de los datos ausentes de la variable NO2 por nivel de calidad**

			miss_NO2		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	925	2600	3525
		% de calidad_conjunta	26,2%	73,8%	100,0%
		% de miss_NO2	51,4%	54,7%	53,8%
		Residuos tipificados	-1,4	,9	
	No cumple	Recuento	876	2157	3033
		% de calidad_conjunta	28,9%	71,1%	100,0%
		% de miss_NO2	48,6%	45,3%	46,2%
		Residuos tipificados	1,5	-,9	
Total	Recuento	1801	4757	6558	
	% de calidad_conjunta	27,5%	72,5%	100,0%	
	% de miss_NO2	100,0%	100,0%	100,0%	

**Tabla 3.2.5.2.3: Estudio de los datos ausentes de la variable NO2 por año**

			miss_NO2		Total
			Presente	Ausente	
año	2002	Recuento	176	1165	1341
		% de año	13,1%	86,9%	100,0%
		% de miss_NO2	9,6%	24,3%	20,2%
		Residuos tipificados	-10,2	6,3	
	2003	Recuento	161	838	999
		% de año	16,1%	83,9%	100,0%
		% de miss_NO2	8,7%	17,5%	15,1%
		Residuos tipificados	-7,0	4,3	
	2004	Recuento	97	544	641
		% de año	15,1%	84,9%	100,0%
		% de miss_NO2	5,3%	11,3%	9,7%
		Residuos tipificados	-6,1	3,8	
	2005	Recuento	336	976	1312
		% de año	25,6%	74,4%	100,0%
		% de miss_NO2	18,2%	20,4%	19,8%
		Residuos tipificados	-1,5	,9	
2006	Recuento	460	823	1283	
	% de año	35,9%	64,1%	100,0%	
	% de miss_NO2	25,0%	17,2%	19,3%	
	Residuos tipificados	5,5	-3,4		
2007	Recuento	612	447	1059	
	% de año	57,8%	42,2%	100,0%	
	% de miss_NO2	33,2%	9,3%	16,0%	
	Residuos tipificados	18,5	-11,5		
Total	Recuento	1842	4793	6635	
	% de año	27,8%	72,2%	100,0%	
	% de miss_NO2	100,0%	100,0%	100,0%	

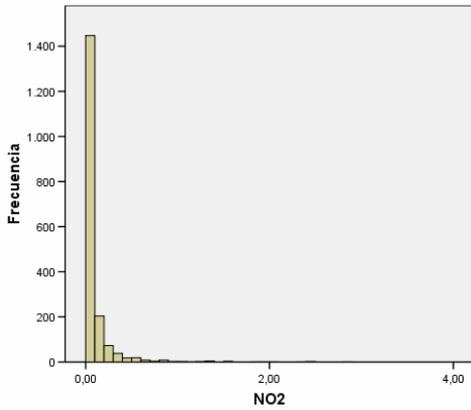
En las Tablas 3.2.5.2.4 y 3.2.5.2.5 así como en las Figuras 3.2.5.2.1 a 3.2.5.2.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de nitritos en las aguas de la cuenca es igual a 0,11 mg/L NO<sub>2</sub> con un nivel mediano igual a 0,05 mg/L NO<sub>2</sub> (ver Tabla 3.2.5.2.4). La distribución de esta variable presenta una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.5.2.4) reflejando la existencia de puntos de muestreo con un alto nivel de nitritos (ver Figuras 3.2.5.2.1 y 3.2.5.2.2). Dichos puntos tienden a concentrarse a lo largo del cauce del río Ebro y en la zona suroeste de la cuenca (ver Figura 3.2.5.2.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.2.5 y Figura 3.2.5.2.3) *siendo necesario realizar una transformación logarítmica (variable LNO2) con el fin de aumentar significativamente el grado de normalidad de la variable* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.2.4 a 3.2.5.2.6.

**Tabla 3.2.5.2.4: Análisis estadístico-descriptivo de las variables NO2 y LNO2**

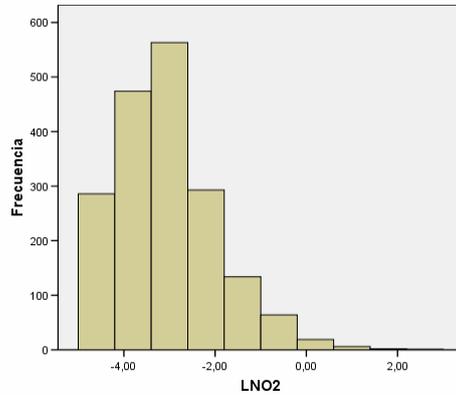
			Estadístico	Error típ.
NO2	Media		,1093	,00906
	Intervalo de confianza para la media al 95%	Límite inferior	,0915	
		Límite superior	,1271	
	Media recortada al 5%		,0645	
	Mediana		,0495	
	Varianza		,151	
	Desv. típ.		,38901	
	Mínimo		,01	
	Máximo		12,20	
	Rango		12,19	
	Amplitud intercuartil		,06	
	Asimetría		19,876	,057
	Curtosis		544,787	,114
	LNO2	Media		-3,0803
Intervalo de confianza para la media al 95%		Límite inferior	-3,1313	
		Límite superior	-3,0293	
Media recortada al 5%			-3,1490	
Mediana			-3,0058	
Varianza			1,246	
Desv. típ.			1,11606	
Mínimo			-4,69	
Máximo			2,50	
Rango			7,19	
Amplitud intercuartil			1,42	
Asimetría			,772	,057
Curtosis			,835	,114

**Tabla 3.2.5.2.5: Pruebas de normalidad de la variable NO2 y LNO2**

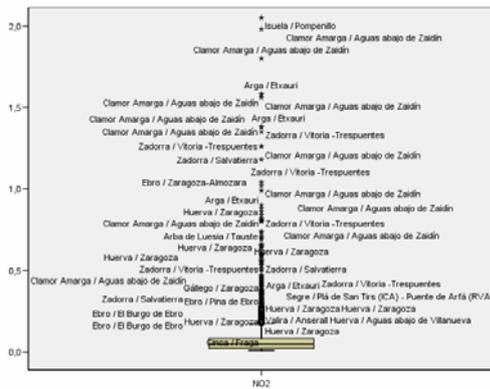
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
NO2	,399	1842	,000	,181	1842	,000
LNO2	,104	1842	,000	,942	1842	,000



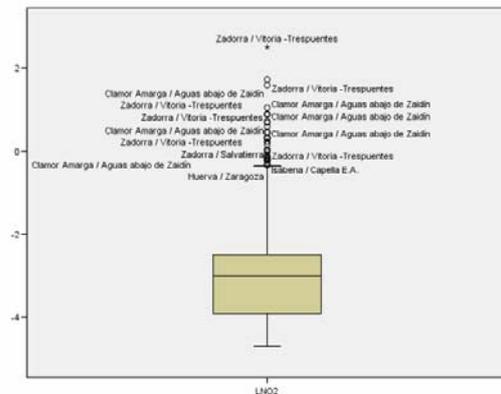
**Figura 3.2.5.2.1: Histograma de la variable NO2**



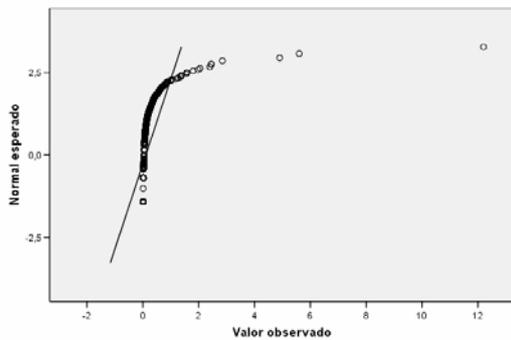
**Figura 3.2.5.2.4: Histograma de la variable LNO2**



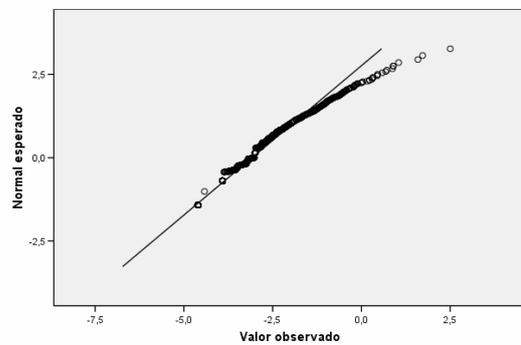
**Figura 3.2.5.2.2: Diagrama de cajas de la variable NO2**



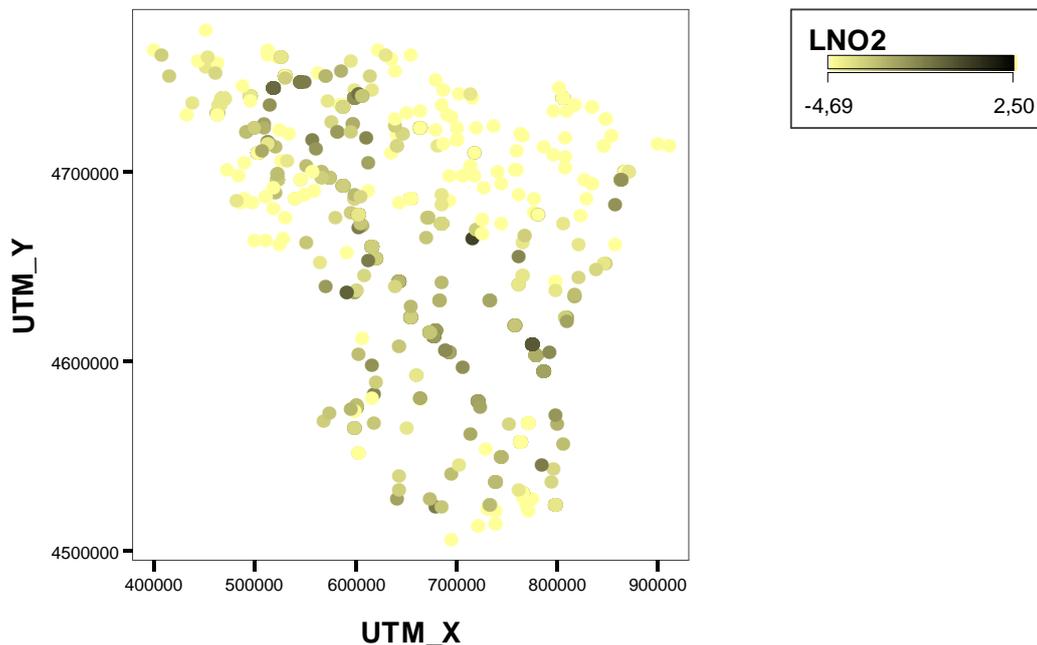
**Figura 3.2.5.2.5: Diagrama de cajas de la variable LNO2**



**Figura 3.2.5.2.3: QQplot normal de la variable NO2**



**Figura 3.2.5.2.6: QQplot normal de la variable LNO2**



**Figura 3.2.5.2.7: Distribución geográfica de los niveles de la variable LNO2**

Por estas razones la *determinación de los umbrales de la variable NO2 se realizará a partir de la variable LNO2.*

### 3.2.5.3. Análisis estadístico de los nitratos

En las Tablas 3.2.5.3.1 a 3.2.5.3.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.3.1) que hay un alto porcentaje (56,2%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 56% de los datos que tienen valores ausentes en la variable NO3 cumplen los estándares de calidad frente a un 50,9% de los datos observados (ver Tabla 3.2.5.3.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución.* Si se hace el estudio por años (ver Tabla 3.2.5.3.3) no se observan diferencias excesivamente importantes, salvo una disminución importante del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.5.3.1: Porcentaje de datos ausentes de la variable NO3**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
2906	43,8%	3729	56,2%	6635	100,0%

**Tabla 3.2.5.3.2: Estudio de los datos ausentes de la variable NO3 por nivel de calidad**

			miss_NO3		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	1449	2076	3525
		% de calidad_conjunta	41,1%	58,9%	100,0%
		% de miss_NO3	50,9%	56,0%	53,8%
		Residuos tipificados	-2,1	1,8	
	No cumple	Recuento	1400	1633	3033
		% de calidad_conjunta	46,2%	53,8%	100,0%
		% de miss_NO3	49,1%	44,0%	46,2%
		Residuos tipificados	2,3	-2,0	
Total	Recuento	2849	3709	6558	
	% de calidad_conjunta	43,4%	56,6%	100,0%	
	% de miss_NO3	100,0%	100,0%	100,0%	

**Tabla 3.2.5.3.3: Estudio de los datos ausentes de la variable NO3 por año**

			miss_NO3		Total
			Presente	Ausente	
año	2002	Recuento	481	860	1341
		% de año	35,9%	64,1%	100,0%
		% de miss_NO3	16,6%	23,1%	20,2%
		Residuos tipificados	-4,4	3,9	
	2003	Recuento	369	630	999
		% de año	36,9%	63,1%	100,0%
		% de miss_NO3	12,7%	16,9%	15,1%
		Residuos tipificados	-3,3	2,9	
	2004	Recuento	160	481	641
		% de año	25,0%	75,0%	100,0%
		% de miss_NO3	5,5%	12,9%	9,7%
		Residuos tipificados	-7,2	6,4	
	2005	Recuento	592	720	1312
		% de año	45,1%	54,9%	100,0%
		% de miss_NO3	20,4%	19,3%	19,8%
		Residuos tipificados	,7	-,6	
2006	Recuento	637	646	1283	
	% de año	49,6%	50,4%	100,0%	
	% de miss_NO3	21,9%	17,3%	19,3%	
	Residuos tipificados	3,2	-2,8		
2007	Recuento	667	392	1059	
	% de año	63,0%	37,0%	100,0%	
	% de miss_NO3	23,0%	10,5%	16,0%	
	Residuos tipificados	9,4	-8,3		
Total	Recuento	2906	3729	6635	
	% de año	43,8%	56,2%	100,0%	
	% de miss_NO3	100,0%	100,0%	100,0%	

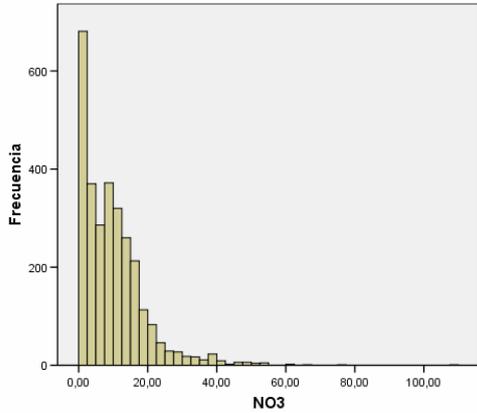
En las Tablas 3.2.5.3.4 y 3.2.5.3.5 así como en las Figuras 3.2.5.3.1 a 3.2.5.3.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de nitratos en las aguas de la cuenca es igual a 9,85 mg/L NO<sub>3</sub> con un nivel mediano igual a 8,30 mg/L NO<sub>3</sub> (ver Tabla 3.2.5.3.4). La distribución de esta variable presenta una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.5.3.4) reflejando la existencia de puntos de muestreo con un alto nivel de nitratos (ver Figuras 3.2.5.3.1 y 3.2.5.3.2). Dichos valores tienden a concentrarse a lo largo del cauce del río Ebro y en las zonas alta y suroeste de la cuenca (ver Figura 3.2.5.3.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.3.5 y Figura 3.2.5.3.3) siendo *necesario realizar una transformación logarítmica (variable LNO3) con el fin de aumentar significativamente el grado de normalidad de la variable* tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.3.4 a 3.2.5.3.6.

**Tabla 3.2.5.3.4: Análisis estadístico-descriptivo de las variables NO<sub>3</sub> y LNO<sub>3</sub>**

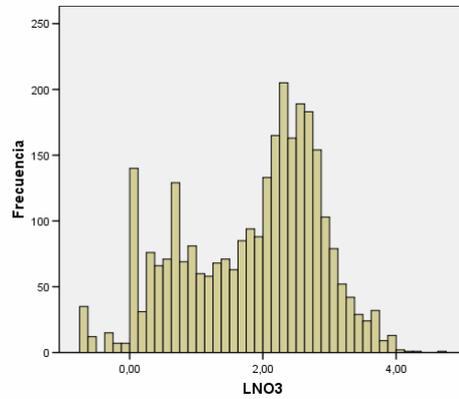
			Estadístico	Error típ.
NO <sub>3</sub>	Media		9,8535	,16713
	Intervalo de confianza para la media al 95%	Límite inferior	9,5258	
		Límite superior	10,1812	
	Media recortada al 5%		8,8456	
	Mediana		8,3000	
	Varianza		81,176	
	Desv. típ.		9,00980	
	Mínimo		,50	
	Máximo		109,00	
	Rango		108,50	
	Amplitud intercuartil		11,10	
	Asimetría		2,211	,045
	Curtosis		9,908	,091
	LNO <sub>3</sub>	Media		1,8437
Intervalo de confianza para la media al 95%		Límite inferior	1,8060	
		Límite superior	1,8813	
Media recortada al 5%			1,8641	
Mediana			2,1163	
Varianza			1,071	
Desv. típ.			1,03484	
Mínimo			-,69	
Máximo			4,69	
Rango			5,38	
Amplitud intercuartil			1,63	
Asimetría			-,434	,045
Curtosis			-,658	,091

**Tabla 3.2.5.3.5: Pruebas de normalidad de la variable NO3 y LNO3**

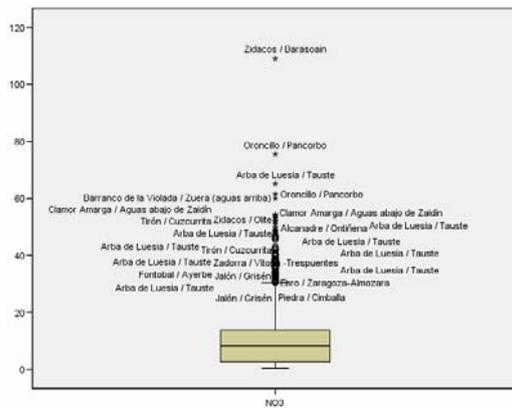
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
NO3	,150	2906	,000	,823	2906	,000
LNO3	,106	2906	,000	,962	2906	,000



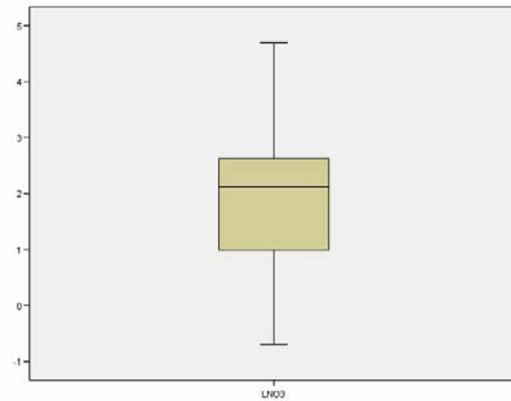
**Figura 3.2.5.3.1: Histograma de la variable NO3**



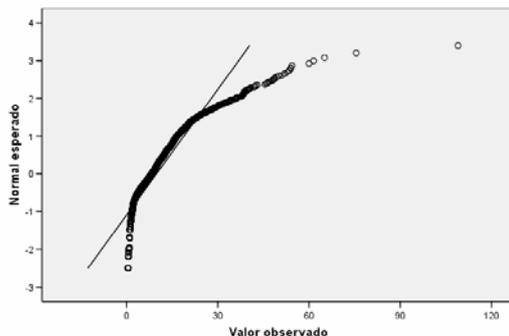
**Figura 3.2.5.3.4: Histograma de la variable LNO3**



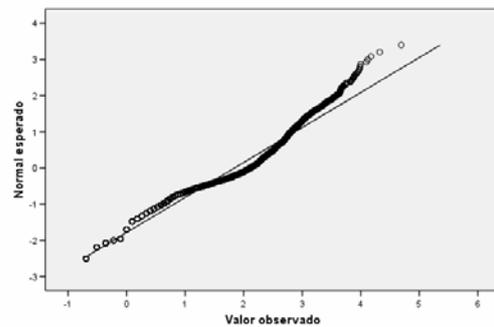
**Figura 3.2.5.3.2: Diagrama de cajas de la variable NO3**



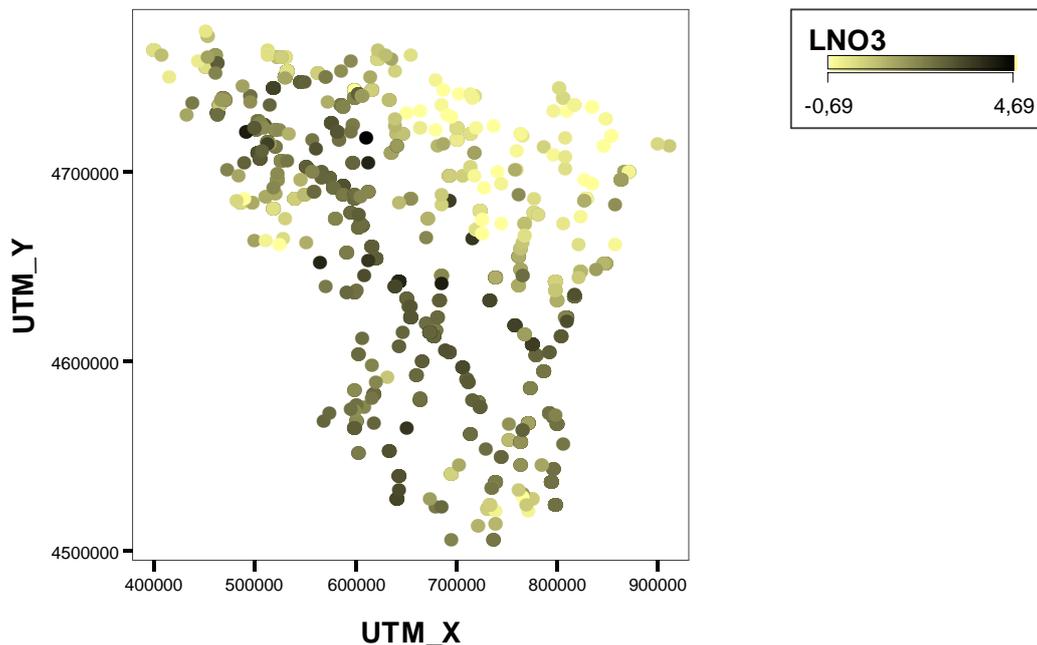
**Figura 3.2.5.3.5: Diagrama de cajas de la variable LNO3**



**Figura 3.2.5.3.3: QQplot normal de la variable NO3**



**Figura 3.2.5.3.6: QQplot normal de la variable LNO3**



**Figura 3.2.5.3.7: Distribución geográfica de los niveles de la variable LNO3**

Por todas estas razones *la determinación de los umbrales de la variable NO3 se realizará a partir de la variable LNO3.*

### **3.2.5.4 Análisis estadístico de los niveles de nitrógeno Kjeldahl**

En las Tablas 3.2.5.4.1 a 3.2.5.4.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.4.1) que hay un alto porcentaje (76,8%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 56,8% de los datos que tienen valores ausentes en la variable N\_KJED cumplen los estándares de calidad frente a un 43,3% de los datos observados (ver Tabla 3.2.5.4.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución.* Si se hace el estudio por años (ver Tabla 3.2.5.4.3) no se observan diferencias excesivamente importantes, salvo una disminución notable del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.5.4.1: Porcentaje de datos ausentes de la variable N\_KJED**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1542	23,2%	5093	76,8%	6635	100,0%

**Tabla 3.2.5.4.2: Estudio de los datos ausentes de la variable N\_KJED por nivel de calidad**

			miss_NKJED		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	649	2876	3525
		% de calidad_conjunta	18,4%	81,6%	100,0%
		% de miss_NKJED	43,3%	56,8%	53,8%
		Residuos tipificados	-5,5	3,0	
	No cumple	Recuento	850	2183	3033
		% de calidad_conjunta	28,0%	72,0%	100,0%
		% de miss_NKJED	56,7%	43,2%	46,2%
		Residuos tipificados	6,0	-3,2	
Total	Recuento	1499	5059	6558	
	% de calidad_conjunta	22,9%	77,1%	100,0%	
	% de miss_NKJED	100,0%	100,0%	100,0%	

**Tabla 3.2.5.4.3: Estudio de los datos ausentes de la variable N\_KJED por años**

			miss_NKJED		Total
			Presente	Ausente	
año	2002	Recuento	116	1225	1341
		% de año	8,7%	91,3%	100,0%
		% de miss_NKJED	7,5%	24,1%	20,2%
		Residuos tipificados	-11,1	6,1	
	2003	Recuento	104	895	999
		% de año	10,4%	89,6%	100,0%
		% de miss_NKJED	6,7%	17,6%	15,1%
		Residuos tipificados	-8,4	4,6	
	2004	Recuento	37	604	641
		% de año	5,8%	94,2%	100,0%
		% de miss_NKJED	2,4%	11,9%	9,7%
		Residuos tipificados	-9,2	5,0	
2005	Recuento	322	990	1312	
	% de año	24,5%	75,5%	100,0%	
	% de miss_NKJED	20,9%	19,4%	19,8%	
	Residuos tipificados	1,0	-5		
2006	Recuento	410	873	1283	
	% de año	32,0%	68,0%	100,0%	
	% de miss_NKJED	26,6%	17,1%	19,3%	
	Residuos tipificados	6,5	-3,6		
2007	Recuento	553	506	1059	
	% de año	52,2%	47,8%	100,0%	
	% de miss_NKJED	35,9%	9,9%	16,0%	
	Residuos tipificados	19,6	-10,8		
Total	Recuento	1542	5093	6635	
	% de año	23,2%	76,8%	100,0%	
	% de miss_NKJED	100,0%	100,0%	100,0%	

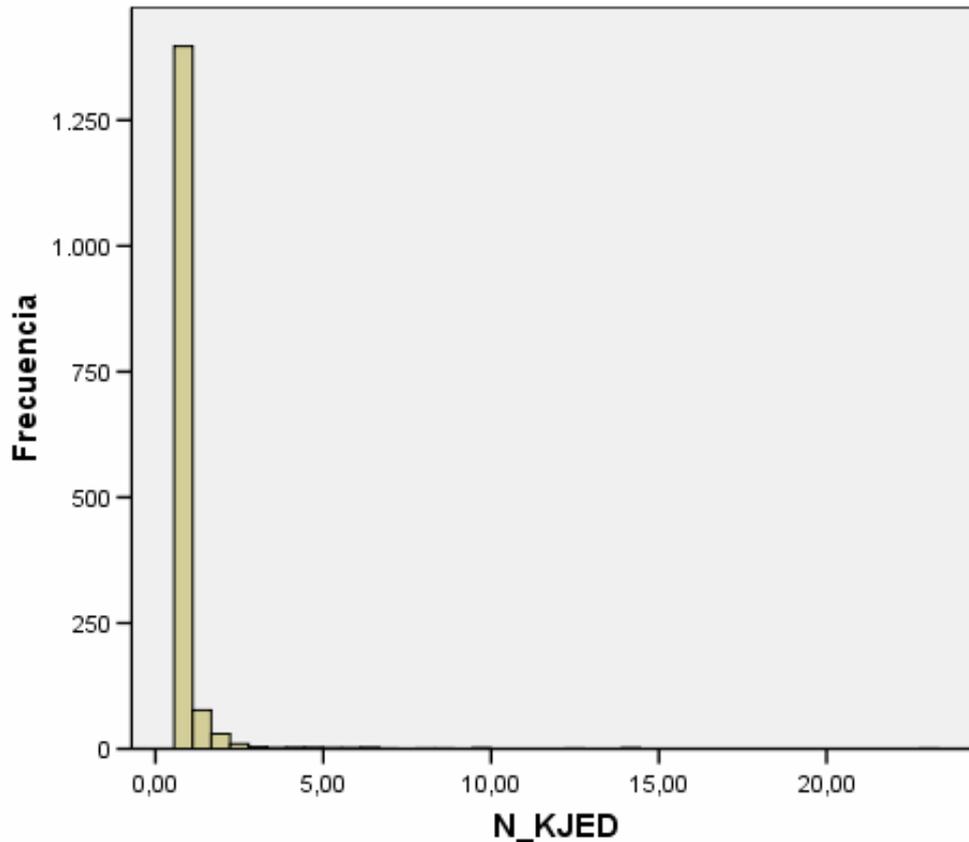
En las Tablas 3.2.5.4.4 y 3.2.5.4.5 así como en las Figuras 3.2.5.4.1 a 3.2.5.4.8 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. Se observa que esta variable tiene un claro *carácter mixto* (ver Figura 3.2.5.4.1) debido a que para el **89,5% de los datos observados su valor está censurado sabiéndose únicamente que  $N_{KJED} \leq 1$  mg/L N**. El nivel medio de nitrógeno Kjeldahl para el resto de los casos en las aguas de la cuenca es igual a 2,46 mg/L N con un nivel mediano igual a 1,50 mg/L N (ver Tabla 3.2.5.4.4). La distribución de esta variable para estos casos presenta, además, una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.5.4.4) reflejando la *existencia de puntos de muestreo con un alto nivel de nitrógeno Kjeldahl* (ver Figuras 3.2.5.4.2 y 3.2.5.4.3), *no observándose patrones geográficos claros en su distribución* (ver Figura 3.2.5.4.8). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.4.5 y Figura 3.2.5.4.4) siendo necesario realizar una *doble transformación logarítmica (variable  $LLN\_KJED$ ) para aumentar significativamente su grado de normalidad*, tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.4.5 a 3.2.5.4.7.

**Tabla 3.2.5.4.4: Análisis estadístico-descriptivo de las variables  $N\_KJED$  y  $LLN\_KJED$  si  $N\_KJED > 1$  mg/L N**

			Estadístico	Error típ.
N_KJED	Media		2,4580	,21891
	Intervalo de confianza para la media al 95%	Límite inferior	2,0257	
		Límite superior	2,8903	
	Media recortada al 5%		1,9770	
	Mediana		1,5000	
	Varianza		7,764	
	Desv. típ.		2,78632	
	Mínimo		1,05	
	Máximo		23,10	
	Rango		22,05	
	Amplitud intercuartil		1,00	
	Asimetría		4,221	,191
	Curtosis		22,536	,379
	LLNKJED	Media		-,8609
Intervalo de confianza para la media al 95%		Límite inferior	-1,0013	
		Límite superior	-,7204	
Media recortada al 5%			-,8710	
Mediana			-,9027	
Varianza			,819	
Desv. típ.			,90509	
Mínimo			-3,02	
Máximo			1,14	
Rango			4,16	
Amplitud intercuartil			1,46	
Asimetría			,103	,191
Curtosis			-,614	,379

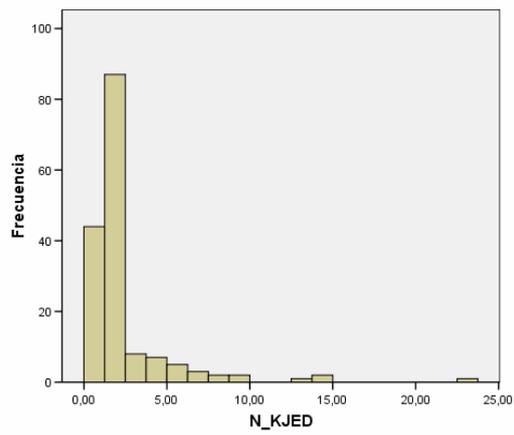
**Tabla 3.2.5.4.5: Pruebas de normalidad de las variables N\_KJED y LLN\_KJED si N\_KJED > 1 mg/L N**

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
N_KJED	,317	162	,000	,492	162	,000
LLNKJED	,095	162	,001	,975	162	,005

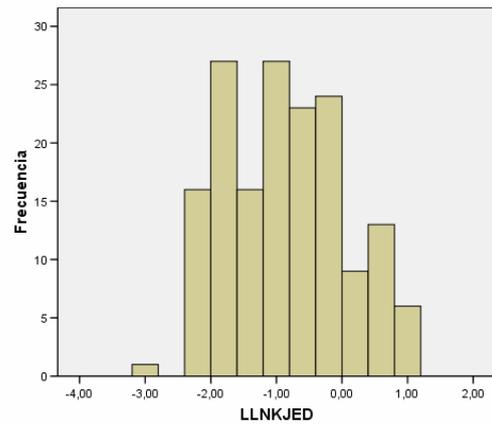


**Figura 3.2.5.4.1: Histograma de la variable N\_KJED para todos los casos**

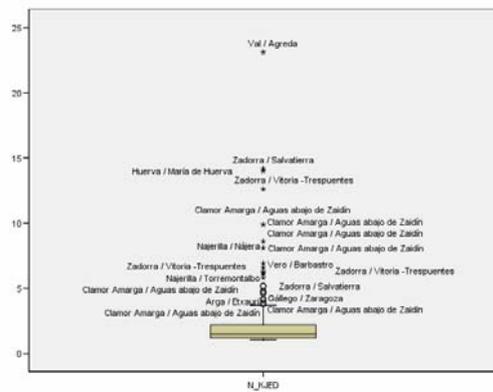
*Por todas estas razones la determinación de los umbrales de la variable N\_KJED se realizará a partir de la variable LLN\_KJED así como de una variable indicador INKJED que toma el valor 1 si  $N\_KJED \leq 1$  mg/L N (es decir si el dato está censurado) y 0 en caso contrario.*



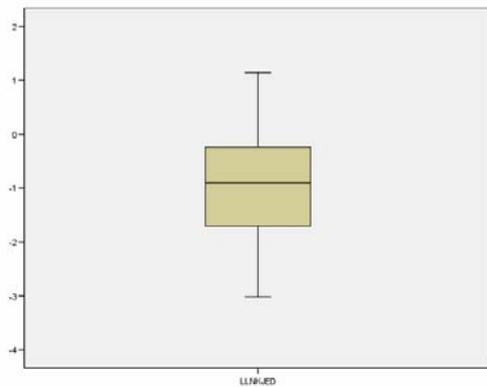
**Figura 3.2.5.4.2: Histograma de la variable N\_KJED si N\_KJED >1 mg/L N**



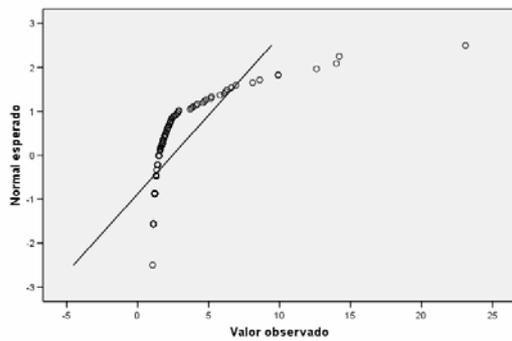
**Figura 3.2.5.4.5 : Histograma de la variable LLN\_KJED si N\_KJED >1 mg/L N**



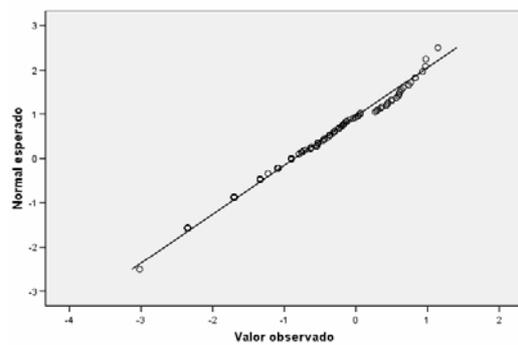
**Figura 3.2.5.4.3: Diagrama de cajas de la variable N\_KJED si N\_KJED >1 mg/L N**



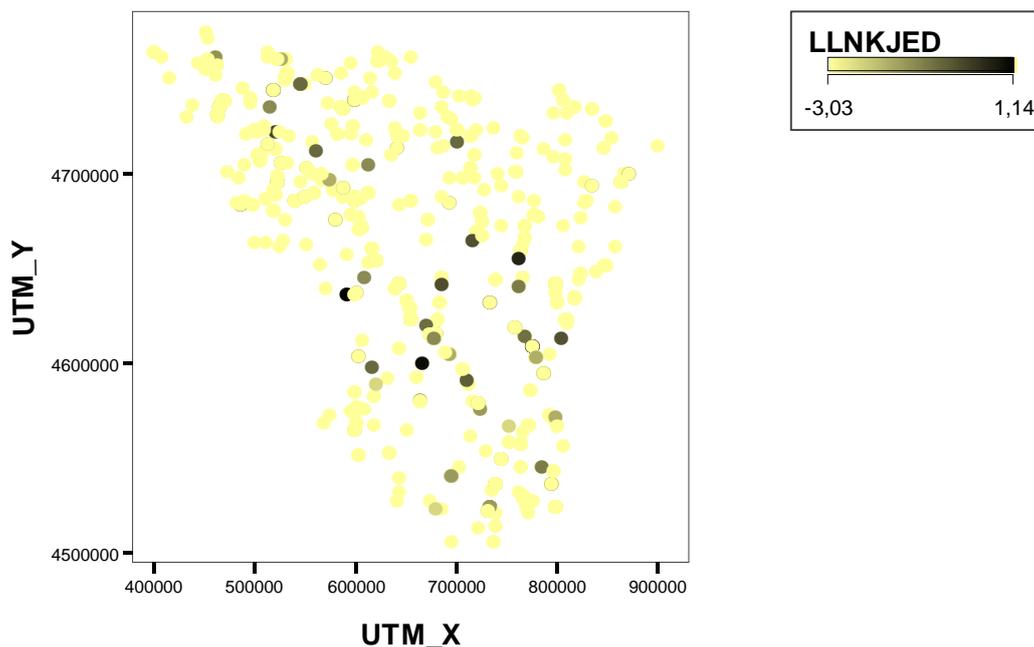
**Figura 3.2.5.4.6 : Diagrama de cajas de la variable LLN\_KJED si N\_KJED >1 mg/L N**



**Figura 3.2.5.4.4: QQplot normal de la variable N\_KJED si N\_KJED >1 mg/L N**



**Figura 3.2.5.4.7: QQplot normal de la variable LLN\_KJED si N\_KJED >1 mg/L N**



**Figura 3.2.5.4.8: Distribución geográfica de los niveles de LLN\_KJED**

### 3.2.5.5 Análisis estadístico de los fosfatos

En las Tablas 3.2.5.5.1 a 3.2.5.5.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.5.1) que hay un alto porcentaje (57,1%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55,9% de los datos que tienen valores ausentes en la variable PO4 cumplen los estándares de calidad frente a un 50,8% de los datos observados (ver Tabla 3.2.5.5.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución*. Si se hace el estudio por años (ver Tabla 3.2.5.5.3) no se observan diferencias excesivamente importantes, salvo una notable disminución del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.5.5.1: Porcentaje de datos ausentes de la variable PO4**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
2846	42,9%	3789	57,1%	6635	100,0%

**Tabla 3.2.5.5.2: Estudio de los datos ausentes de la variable PO4 por nivel de calidad**

			miss_PO4		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	1417	2108	3525
		% de calidad_conjunta	40,2%	59,8%	100,0%
		% de miss_PO4	50,8%	55,9%	53,8%
		Residuos tipificados	-2,1	1,8	
	No cumple	Recuento	1372	1661	3033
		% de calidad_conjunta	45,2%	54,8%	100,0%
		% de miss_PO4	49,2%	44,1%	46,2%
		Residuos tipificados	2,3	-2,0	
Total	Recuento	2789	3769	6558	
	% de calidad_conjunta	42,5%	57,5%	100,0%	
	% de miss_PO4	100,0%	100,0%	100,0%	

**Tabla 3.2.5.5.3: Estudio de los datos ausentes de la variable PO4 por años**

			miss_PO4		Total
			Presente	Ausente	
año	2002	Recuento	432	909	1341
		% de año	32,2%	67,8%	100,0%
		% de miss_PO4	15,2%	24,0%	20,2%
		Residuos tipificados	-6,0	5,2	
	2003	Recuento	371	628	999
		% de año	37,1%	62,9%	100,0%
		% de miss_PO4	13,0%	16,6%	15,1%
		Residuos tipificados	-2,8	2,4	
	2004	Recuento	161	480	641
		% de año	25,1%	74,9%	100,0%
		% de miss_PO4	5,7%	12,7%	9,7%
		Residuos tipificados	-6,9	6,0	
	2005	Recuento	579	733	1312
		% de año	44,1%	55,9%	100,0%
		% de miss_PO4	20,3%	19,3%	19,8%
		Residuos tipificados	,7	-,6	
	2006	Recuento	637	646	1283
		% de año	49,6%	50,4%	100,0%
		% de miss_PO4	22,4%	17,0%	19,3%
		Residuos tipificados	3,7	-3,2	
	2007	Recuento	666	393	1059
		% de año	62,9%	37,1%	100,0%
		% de miss_PO4	23,4%	10,4%	16,0%
		Residuos tipificados	9,9	-8,6	
	Total	Recuento	2846	3789	6635
		% de año	42,9%	57,1%	100,0%
		% de miss_PO4	100,0%	100,0%	100,0%

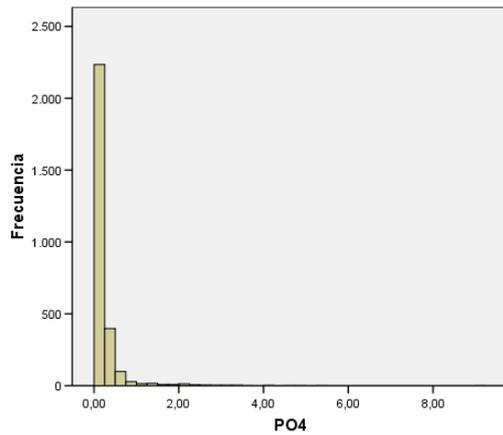
En las Tablas 3.2.5.5.4 y 3.2.5.5.5 así como en las Figuras 3.2.5.5.1 a 3.2.5.5.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de fosfatos en las aguas de la cuenca es igual a 0,225 mg/L PO<sub>4</sub> con un nivel mediano igual a 0,10 mg/L PO<sub>4</sub> (ver Tabla 3.2.5.5.4). La distribución de esta variable presenta una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.5.5.4) reflejando la *existencia de puntos de muestreo con un alto nivel de fosfatos* (ver Figuras 3.2.5.5.1 y 3.2.5.5.2). *Dichos puntos tienden a concentrarse a lo largo del cauce del río Ebro y zonas cercanas a la misma así como en las zonas altas de la cuenca* (ver Figura 3.2.5.5.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.5.5 y Figura 3.2.5.5.3) siendo necesario realizar una *transformación logarítmica (variable LPO4) que aumenta significativamente el grado de normalidad* de la variable tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.5.4 a 3.2.5.5.6.

**Tabla 3.2.5.5.4: Análisis estadístico-descriptivo de las variables PO4 y LPO4**

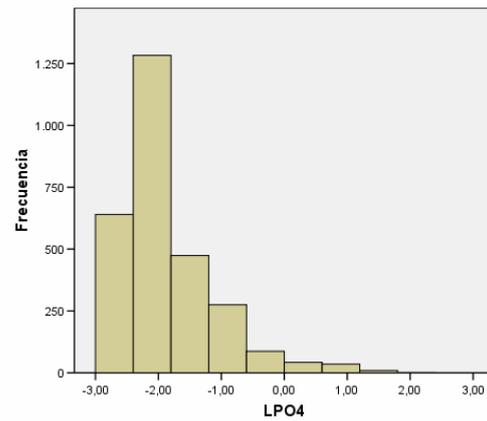
			Estadístico	Error típ.
PO4	Media		,2250	,00787
	Intervalo de confianza para la media al 95%	Límite inferior	,2095	
		Límite superior	,2404	
	Media recortada al 5%		,1586	
	Mediana		,1000	
	Varianza		,176	
	Desv. típ.		,42004	
	Mínimo		,05	
	Máximo		9,15	
	Rango		9,10	
	Amplitud intercuartil		,12	
	Asimetría		8,152	,046
	Curtosis		106,704	,092
	LPO4	Media		-1,9847
Intervalo de confianza para la media al 95%		Límite inferior	-2,0151	
		Límite superior	-1,9543	
Media recortada al 5%			-2,0548	
Mediana			-2,3026	
Varianza			,684	
Desv. típ.			,82711	
Mínimo			-3,00	
Máximo			2,21	
Rango			5,21	
Amplitud intercuartil			,79	
Asimetría			1,221	,046
Curtosis			1,919	,092

**Tabla 3.2.5.5.5: Pruebas de normalidad de las variables PO4 y LPO4**

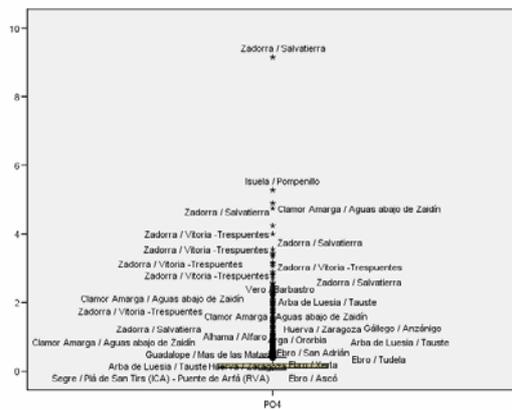
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PO4	,339	2846	,000	,361	2846	,000
LPO4	,200	2846	,000	,884	2846	,000



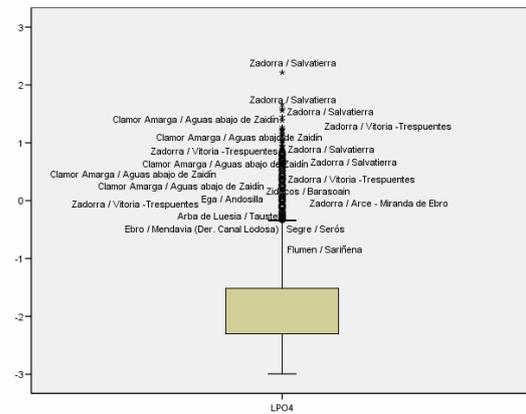
**Figura 3.2.5.5.1: Histograma de la variable PO4**



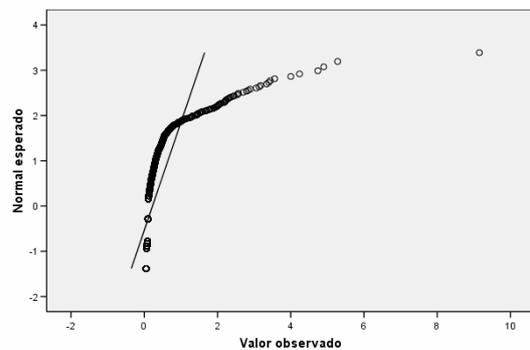
**Figura 3.2.5.5.4 : Histograma de la variable LPO4**



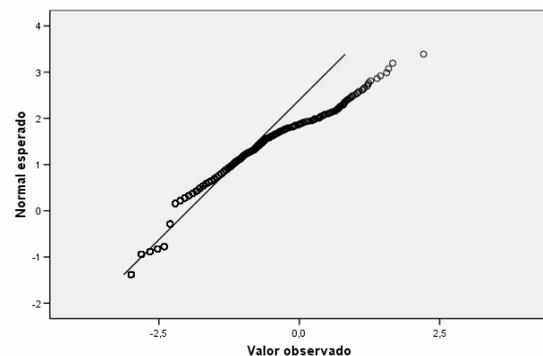
**Figura 3.2.5.5.2: Diagrama de cajas de la variable PO4**



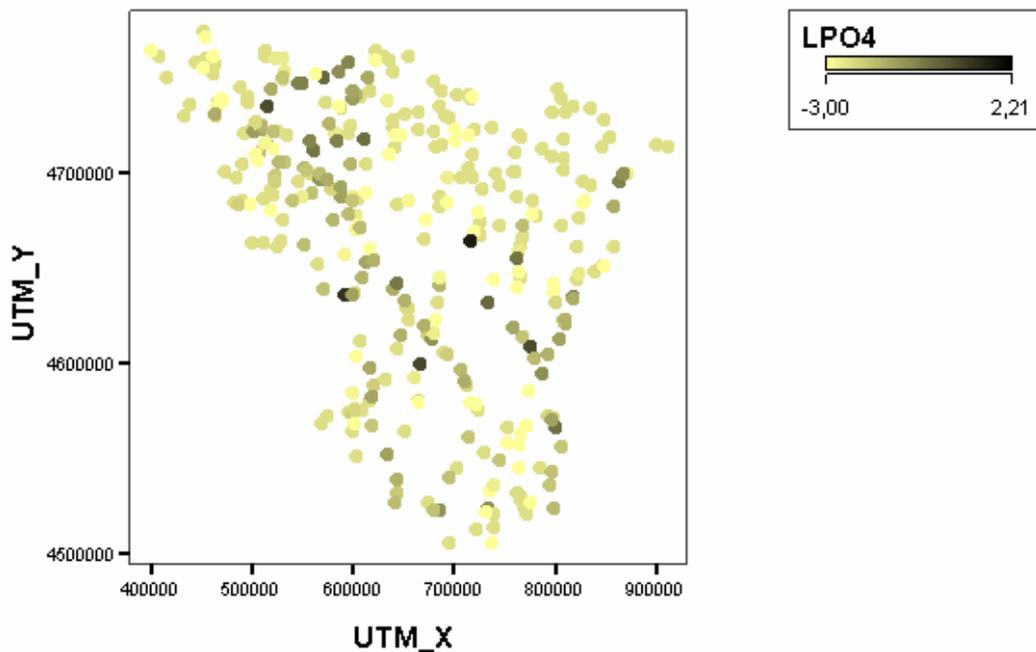
**Figura 3.2.5.5.5 : Diagrama de cajas de la variable LPO4**



**Figura 3.2.5.5.3: QQplot normal de la variable PO4**



**Figura 3.2.5.5.6: QQplot normal de la variable LPO4**



**Figura 3.2.5.5.7: Distribución geográfica de los niveles de LPO4**

Por estas razones, *la determinación de los umbrales de la variable PO4 se realizará a partir de la variable LPO4.*

### **3.2.5.6 Análisis estadístico de nivel de fósforo**

En las Tablas 3.2.5.6.1 a 3.2.5.6.3 se muestran los resultados obtenidos al estudiar los datos ausentes respecto a esta variable. Se observa (ver Tabla 3.2.5.6.1) que hay un alto porcentaje (70,3%) así como la existencia de una tendencia significativa a concentrarse entre los datos que sí cumplen los niveles de calidad conjunta dado que un 55,3% de los datos que tienen valores ausentes en la variable P\_TOT cumplen los estándares de calidad frente a un 50,2% de los datos observados (ver Tabla 3.2.5.6.2). Existe, por tanto, una *sobre-representación de los datos que no cumplen los niveles de calidad mínimos con lo que los resultados expuestos respecto a esta variable deberán tomarse con precaución.* Si se hace el estudio por años (ver Tabla 3.2.5.6.3) no se observan diferencias excesivamente importantes, salvo una disminución fuerte del porcentaje de datos ausentes en el año 2007.

**Tabla 3.2.5.6.1: Porcentaje de datos ausentes de la variable P\_TOT**

Casos					
Válidos		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1973	29,7%	4662	70,3%	6635	100,0%

**Tabla 3.2.5.6.2: Estudio de los datos ausentes de la variable P\_TOT por nivel de calidad**

			miss_PTOT		Total
			Presente	Ausente	
calidad_conjunta	Cumple	Recuento	969	2556	3525
		% de calidad_conjunta	27,5%	72,5%	100,0%
		% de miss_PTOT	50,2%	55,3%	53,8%
		Residuos tipificados	-2,2	1,4	
	No cumple	Recuento	963	2070	3033
		% de calidad_conjunta	31,8%	68,2%	100,0%
		% de miss_PTOT	49,8%	44,7%	46,2%
		Residuos tipificados	2,3	-1,5	
Total	Recuento	1932	4626	6558	
	% de calidad_conjunta	29,5%	70,5%	100,0%	
	% de miss_PTOT	100,0%	100,0%	100,0%	

**Tabla 3.2.5.6.3: Estudio de los datos ausentes de la variable P\_TOT por años**

			miss_PTOT		Total
			Presente	Ausente	
año 2002	Recuento	215	1126	1341	
	% de año	16,0%	84,0%	100,0%	
	% de miss_PTOT	10,9%	24,2%	20,2%	
	Residuos tipificados	-9,2	6,0		
2003	Recuento	174	825	999	
	% de año	17,4%	82,6%	100,0%	
	% de miss_PTOT	8,8%	17,7%	15,1%	
	Residuos tipificados	-7,1	4,6		
2004	Recuento	133	508	641	
	% de año	20,7%	79,3%	100,0%	
	% de miss_PTOT	6,7%	10,9%	9,7%	
	Residuos tipificados	-4,2	2,7		
2005	Recuento	356	956	1312	
	% de año	27,1%	72,9%	100,0%	
	% de miss_PTOT	18,0%	20,5%	19,8%	
	Residuos tipificados	-1,7	1,1		
2006	Recuento	477	806	1283	
	% de año	37,2%	62,8%	100,0%	
	% de miss_PTOT	24,2%	17,3%	19,3%	
	Residuos tipificados	4,9	-3,2		
2007	Recuento	618	441	1059	
	% de año	58,4%	41,6%	100,0%	
	% de miss_PTOT	31,3%	9,5%	16,0%	
	Residuos tipificados	17,1	-11,1		
Total	Recuento	1973	4662	6635	
	% de año	29,7%	70,3%	100,0%	
	% de miss_PTOT	100,0%	100,0%	100,0%	

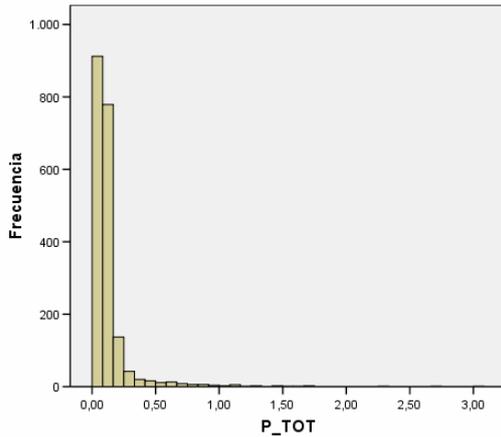
En las Tablas 3.2.5.6.4 y 3.2.5.6.5 así como en las Figuras 3.2.5.6.1 a 3.2.5.6.7 se muestran los resultados obtenidos al hacer un análisis estadístico exploratorio de esta variable. El nivel medio de fósforo total en las aguas de la cuenca es igual a 0,12 mg/L P con un nivel mediano igual a 0,09 mg/L P (ver Tabla 3.2.5.6.4). La distribución de esta variable presenta una asimetría positiva y leptocurtosis significativas (ver Tabla 3.2.5.6.4) reflejando la *existencia de puntos de muestreo con un alto nivel de fósforo* (ver Figuras 3.2.5.6.1 y 3.2.5.6.2). *Dichos puntos tienden a concentrarse a lo largo del cauce del río Ebro así como en las zonas altas de la cuenca* (ver Figura 3.2.5.6.7). Por todo ello se rechaza la hipótesis de normalidad de esta variable (ver Tabla 3.2.5.6.5 y Figura 3.2.5.6.3) siendo necesario realizar una *transformación logarítmica (variable LP\_TOT) con el fin de aumentar significativamente el grado de normalidad* de la variable tal y como lo demuestran los resultados expuestos en las Figuras 3.2.5.6.4 a 3.2.5.6.6.

**Tabla 3.2.5.6.4: Análisis estadístico-descriptivo de las variables P\_TOT y LP\_TOT**

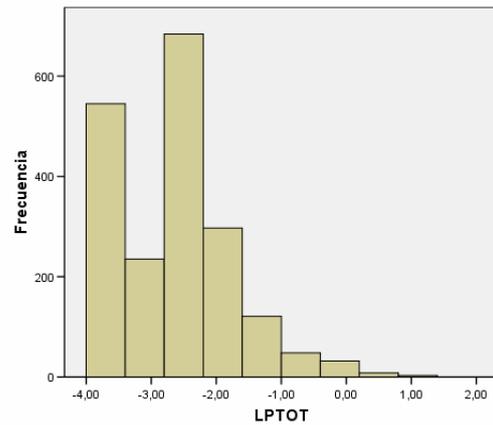
			Estadístico	Error típ.
P_TOT	Media		,1200	,00429
	Intervalo de confianza para la media al 95%	Límite inferior	,1116	
		Límite superior	,1284	
	Media recortada al 5%		,0888	
	Mediana		,0900	
	Varianza		,036	
	Desv. típ.		,19077	
	Mínimo		,03	
	Máximo		3,06	
	Rango		3,02	
	Amplitud intercuartil		,08	
	Asimetría		6,868	,055
	Curtosis		68,777	,110
	LPTOT	Media		-2,5453
Intervalo de confianza para la media al 95%		Límite inferior	-2,5805	
		Límite superior	-2,5101	
Media recortada al 5%			-2,6167	
Mediana			-2,4079	
Varianza			,635	
Desv. típ.			,79692	
Mínimo			-3,41	
Máximo			1,12	
Rango			4,53	
Amplitud intercuartil			1,25	
Asimetría			1,000	,055
Curtosis			1,342	,110

**Tabla 3.2.5.6.5: Pruebas de normalidad de las variables P\_TOT y LP\_TOT**

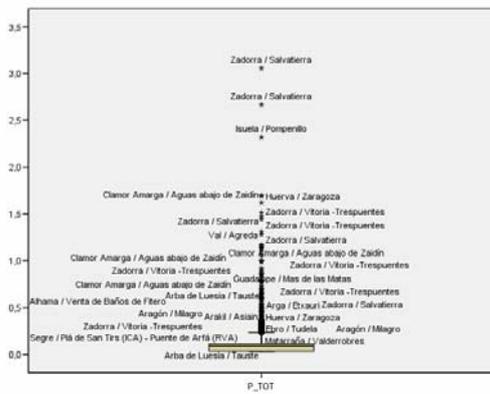
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
P_TOT	,324	1973	,000	,410	1973	,000
LPTOT	,139	1973	,000	,884	1973	,000



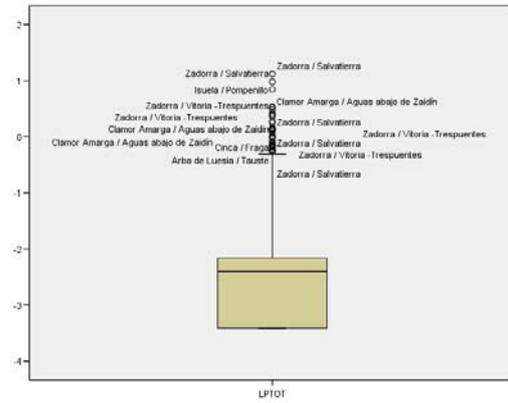
**Figura 3.2.5.6.1: Histograma de la variable P\_TOT**



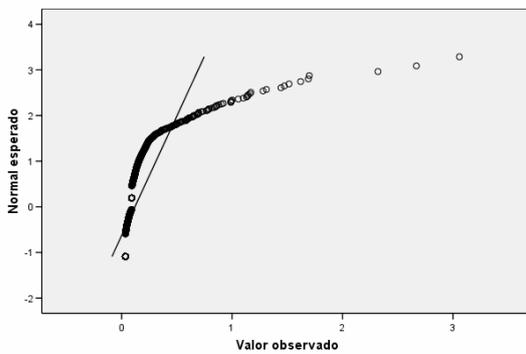
**Figura 3.2.5.6.4 : Histograma de la variable LP\_TOT**



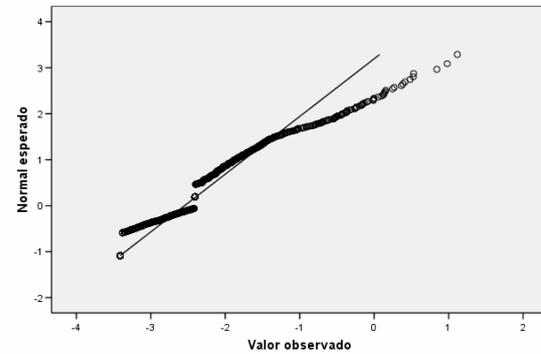
**Figura 3.2.5.6.2: Diagrama de cajas de la variable P\_TOT**



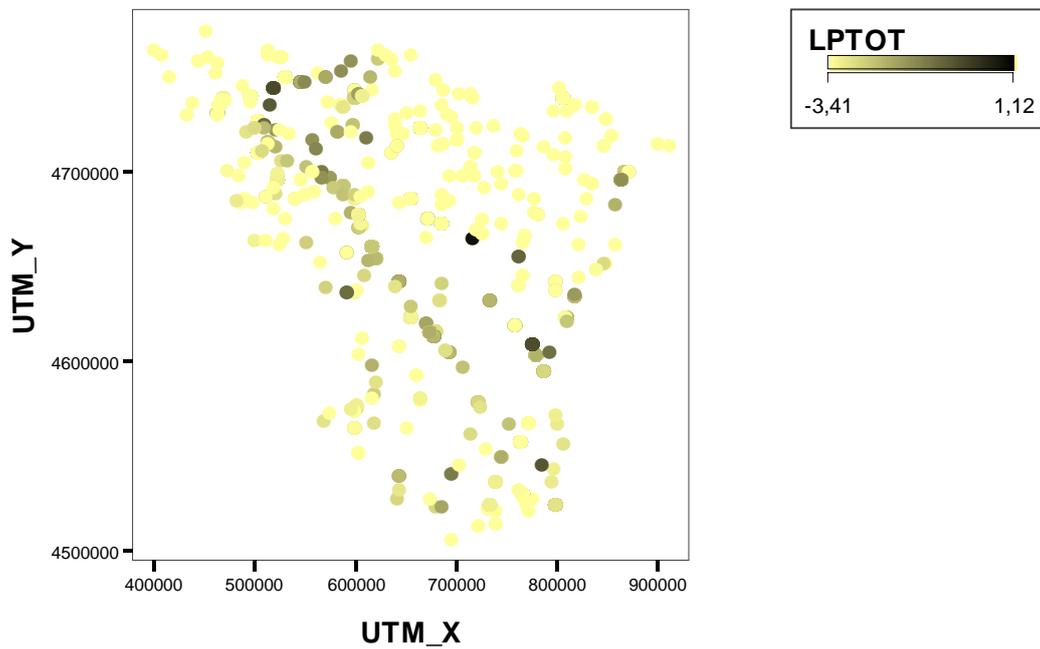
**Figura 3.2.5.6.5 : Diagrama de cajas de la variable LP\_TOT**



**Figura 3.2.5.6.3: QQplot normal de la variable P\_TOT**



**Figura 3.2.5.6.6: QQplot normal de la variable LP\_TOT**



**Figura 3.2.5.6.7: Distribución geográfica de los niveles de LP\_TOT**

Por estas razones la *determinación de los umbrales de la variable P\_TOT se realizará a partir de la variable LP\_TOT.*

#### 4. METODOLOGIA ESTADISTICA

Una vez realizado el análisis estadístico-descriptivo de las variables utilizadas en el estudio, en esta sección se describe, brevemente, la metodología estadística utilizada para la determinación de los umbrales de calidad a partir de la información proporcionada por el conjunto de puntos de muestreo situados a lo largo de la cuenca.

Para exponerla introducimos la siguiente notación. Sea  $Y =$  nivel de calidad conjunta codificada de forma binaria como:

$$Y = \begin{cases} 0 & \text{si la calidad del agua verifica los mínimos exigidos} \\ 1 & \text{si la calidad del agua no verifica los mínimos exigidos} \end{cases}$$

Sea  $X$  la variable físico-química cuyos umbrales de calidad se quieren establecer. Supondremos, además, que es necesario realizar una transformación funcional  $g$  tal que  $g(X)$  es aproximadamente normal.

La determinación de los umbrales de calidad se basa en la estimación de un modelo de regresión logística de la forma:

$$P(Y=1|X) = \frac{e^{\alpha+\beta g(X)}}{1 + e^{\alpha+\beta g(X)}} \quad (4.1)$$

donde  $P(Y=1|X)$  es la probabilidad de que si el valor del parámetro físico-químico es  $X$ , el agua de la zona considerada no verifique los estándares de calidad mínimo exigidos. Dicha probabilidad mide, por tanto, el riesgo de que dicha zona presente niveles de contaminación altos en el futuro.

Observar, además, que si  $X_0$  es tal que  $g(X_0) = 0$  se verifica que  $P(Y=1|X_0) = \frac{e^{\alpha}}{1 + e^{\alpha}}$  por lo que el parámetro  $\alpha$  mide el riesgo de que una zona con un valor del indicador  $X = X_0$ , no verifique los estándares de calidad mínimos. Así mismo, el parámetro físico-químico es relevante para medir los niveles de calidad del agua si  $\beta \neq 0$  de forma que, al ser la función  $g(X)$  creciente en  $X$  en todos los casos analizados, se tiene que si  $\beta > 0$  (respectivamente, si  $\beta < 0$ ) valores altos de  $X$  (respectivamente valores bajos de  $X$ ) vendrán asociados a riesgos altos de contaminación.

Fijado un nivel de riesgo  $p$  con  $0 < p < 1$  el umbral de calidad para el parámetro físico-químico  $X$  vendrá dado por la expresión:

$$\text{Umbral}_p = g^{-1} \left( \frac{\log\left(\frac{p}{1-p}\right) - \alpha}{\beta} \right) \quad (4.2)$$

donde  $g^{-1}$  denota a la función inversa de  $g$ . De esta forma se verifica que:

$$\text{a) Si } \beta > 0 \text{ y } X \geq \text{Umbral}_p \quad (4.3)$$

o

$$\text{b) Si } \beta < 0 \text{ y } X \leq \text{Umbral}_p \quad (4.4)$$

el agua tiene un riesgo del 100p% de estar contaminada en el futuro. Si  $\beta = 0$ ,  $X$  no es relevante para predecir los niveles de contaminación del agua y no debería ser utilizado para ello.

Así mismo, se lleva a cabo una validación intramuestral de los umbrales propuestos. Dicha validación utiliza los datos de la muestra analizada y los clasifica como muestras de agua contaminada o no contaminada de acuerdo a las reglas (4.3) o (4.4) según proceda, para un valor fijo de  $p$ . Utilizando dicho procedimiento de clasificación se calculan las siguientes cantidades:

$$\% \text{ Aciertos cumplen} = 100 * \frac{n_{\text{bien clasificados, cumplen}}}{n_{\text{cumplen}}} \quad (4.5)$$

$$\% \text{ Aciertos no cumplen} = 100 * \frac{n_{\text{bien clasificados, no cumplen}}}{n_{\text{no cumplen}}} \quad (4.6)$$

$$\% \text{ Aciertos totales} = 100 * \frac{n_{\text{bien clasificados}}}{n_{\text{total}}} \quad (4.7)$$

donde  $n_{\text{bien clasificados, cumplen}}$  y  $n_{\text{bien clasificados, no cumplen}}$  son el número de muestras bien clasificadas entre las muestras que cumplen y no cumplen los requisitos de calidad, respectivamente;  $n_{\text{cumplen}}$  y  $n_{\text{no cumplen}}$  son, respectivamente, el número de muestras que cumplen y no cumplen dichos requisitos de calidad;  $n_{\text{bien clasificados}} = n_{\text{bien clasificados, cumplen}} + n_{\text{bien clasificados, no cumplen}}$  es el número de muestras bien clasificadas y  $n_{\text{total}} = n_{\text{cumplen}} + n_{\text{no cumplen}}$  es el número total de muestras.

Estos indicadores nos proporcionan una evaluación, en términos prácticos, de la fiabilidad del indicador  $X$  a la hora de pronosticar el riesgo que corre una zona de no

verificar los indicadores de calidad mínimos exigidos para sus aguas. Así, el **%Aciertos cumplen** y **%Aciertos no cumplen** evalúan la fiabilidad del umbral dentro de los casos que cumplen y que no cumplen con los requisitos de calidad, respectivamente. Por su parte el **%Aciertos totales** evalúa la fiabilidad global del umbral en el total de la muestra.

Finalmente, se ha realizado una determinación multivariante de umbrales de calidad, para los distintos tipos de indicadores listados en la Tabla 2.3. Para ello se han estimado modelos de regresión logística de la forma:

$$P(Y=1|X_1, \dots, X_k) = \frac{e^{\alpha + \sum_{i=1}^k \beta_i g_i(X_i)}}{1 + e^{\alpha + \sum_{i=1}^k \beta_i g_i(X_i)}} \quad (4.8)$$

donde  $X_1, \dots, X_p$  denotan los parámetros físico-químicos que son relevantes para determinar la calidad del agua y  $g_1, \dots, g_p$  son las transformaciones funcionales utilizadas para aumentar su grado de normalidad.

Para un riesgo dado  $0 < p < 1$  se determinaría que la calidad del agua no cumple los estándares de calidad mínimos si:

$$\alpha + \beta_1 g_1(X_1) + \dots + \beta_k g_k(X_k) \geq \log\left(\frac{p}{1-p}\right) \quad (4.9)$$

Esta forma de actuar tiene la ventaja de que tiene en cuenta las interacciones existentes entre los parámetros físico-químicos y el inconveniente de ser más inestable al estar estimada con menor acuracidad. La selección de los parámetros físico-químicos  $X_1, \dots, X_k$  se ha llevado a cabo mediante procesos de selección de variables paso a paso lo cual nos permite seleccionar aquéllos con un mayor poder discriminante de los niveles de calidad del agua.

## 5. DETERMINACIÓN DE LOS UMBRALES DE CALIDAD

En esta sección se presentan los resultados obtenidos al aplicar la metodología descrita en la sección 4 a cada una de las variables físico-químicas consideradas en el estudio. Los resultados se presentan al igual que en la sección 3.2 agrupando las variables por bloques según el tipo de indicador (ver Tabla 2.3). Finalmente, se estima para cada tipo de indicador, un modelo multivariante que estima el riesgo de no cumplir con los estándares de calidad utilizando toda la información disponible.

### 5.1. UMBRALES UNIVARIANTES

#### 5.1.1. Umbrales de calidad para los indicadores generales

En este apartado presentamos los umbrales de calidad para todos los indicadores generales TEM\_AG, TEM\_AI, PH, COND20, O2, O2% y MES cuyo significado puede verse en la Tabla 2.3.

##### 5.1.1.1. Umbrales de calidad para la temperatura del agua

En la Tabla 5.1.1.1.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  de acuerdo con el análisis exploratorio realizado en la sección 3.2.1.1.

**Tabla 5.1.1.1.1: Estimación de los parámetros del modelo (4.1) para la variable TEM\_AG**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	0,05	0,00	122,24	1	0,00	1,05
$\alpha$	-0,84	0,07	155,86	1	0,00	0,43

Se observa que el coeficiente  $\beta$  es significativamente positivo. Por lo tanto, *altos valores de la temperatura de agua estarán asociados a un alto de riesgo de contaminación cifrándose en un 5% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento de 1° C de la temperatura del agua.*

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \frac{\log\left(\frac{p}{1-p}\right) + 0,84}{0,05} \quad (5.1)$$

de forma que se detectará riesgo alto de contaminación del agua si la temperatura de ésta es mayor o igual que dicho umbral. En la Tabla 5.1.1.1.2 se muestra el valor de la expresión (5.1) para una red de posibles valores de riesgo.

**Tabla 5.1.1.1.2: Umbrales de calidad para la variable TEM\_AG para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%
Umbral (° C)	17,05	25,32	34,33
Aciertos Cumplen(%)	73,80%	98,90%	100,00%
Aciertos No Cumplen(%)	37,50%	3,10%	0,00%
Aciertos Totales(%)	57,00%	54,60%	53,80%

Así, por ejemplo, si la temperatura del agua de una muestra es superior a 34,33°C se corre un riesgo superior al 70% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo, oscilando su fiabilidad global entre un 53,8% y un 57% las cuales son ligeramente superiores a la fiabilidad (53,8%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen y que se obtiene para niveles de riesgo superiores al 70% (ver Tabla 5.1.1.1.2). Todo ello refleja el *bajo poder discriminante de esta variable a la hora de determinar el nivel de calidad del agua, motivado por la distribución uniforme de la temperatura del agua a lo largo de la cuenca* (ver sección 3.2.1.1).

### 5.1.1.2. Umbrales de calidad para temperatura del aire

En la Tabla 5.1.1.2.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  de acuerdo a los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.1.2.

**Tabla 5.1.1.2.1: Estimación de los parámetros del modelo (4.1) para la variable TEM\_AI**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	0,01	0,00	5,84	1	0,02	1,01
$\alpha$	-0,35	0,06	36,34	1	0,00	0,71

Se observa que el coeficiente  $\beta$  es significativamente positivo a un nivel de significación del 5% pero no del 1% poniendo de manifiesto el débil poder discriminante de esta variable. En general, *altos valores de la temperatura del aire estarán asociados a un alto de riesgo de contaminación cifrándose en un 1% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento de 1°C la temperatura del aire.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \frac{\log\left(\frac{p}{1-p}\right) + 0.35}{0,01} \quad (5.2)$$

de forma que se detectará riesgo alto de contaminación si la temperatura del aire es mayor o igual que dicho umbral. En la Tabla 5.1.1.2.2 se muestra el valor de (5.2) para una red de posibles valores de riesgo

**Tabla 5.1.1.2.2: Umbrales de calidad para la variable TEM\_AI para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	42,00%	44,00%	46,00%	48,00%	50,00%
<b>Umbral (° C)</b>	3,38	14,07	24,65	35,17	45,66
<b>Aciertos Cumplen(%)</b>	5,40%	41,60%	83,00%	99,90%	100,00%
<b>Aciertos No Cumplen(%)</b>	94,50%	61,10%	17,80%	0,50%	0,00%
<b>Aciertos Totales(%)</b>	45,00%	50,30%	54,10%	55,70%	55,60%

Así, por ejemplo, si la temperatura del aire de una muestra es superior a 35,17°C se corre un riesgo superior al 48% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen para niveles de riesgo superiores al 46% pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen, oscilando su fiabilidad global entre un 54,1% y un 55,6% las cuales son ligeramente inferiores a la fiabilidad (55,6%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen y que se obtiene para niveles de riesgo superiores al 50% (ver Tabla 5.1.1.2.2). Todo ello refleja el *bajo poder discriminante de esta variable a la hora de determinar el nivel de calidad debido a la distribución uniforme de la misma a lo largo de la cuenca* (ver sección 3.2.1.2).

### 5.1.1.3. Umbrales de calidad para el índice pH

En la Tabla 5.1.1.3.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  de acuerdo al análisis exploratorio realizado en la sección 3.2.1.3.

**Tabla 5.1.1.3.1: Estimación de los parámetros del modelo (4.1) para la variable PH**

Parámetro	Estimación	Error Estándar	Wald	GL	Pvalor	Exp(B)
$\beta$	-1,28	0,11	144,87	1	0,00	0,28
$\alpha$	10,25	0,86	140,60	1	0,00	28341,12

Se observa que el parámetro  $\beta$  es significativamente negativo. Por lo tanto, *altos valores del pH estarán asociados a un bajo riesgo de contaminación cifrándose en un 128% la disminución del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento en una unidad del valor del pH.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \frac{\log\left(\frac{p}{1-p}\right) - 10,25}{-1,28} \quad (5.3)$$

de forma que se detectará riesgo alto de contaminación del agua si el pH es menor o igual que dicho umbral. En la Tabla 5.1.1.3.2 se muestra el valor de (5.3) para una red de posibles valores de riesgo

**Tabla 5.1.1.3.2: Umbrales de calidad para la variable PH para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%
<b>Umbral</b>	8,03	7,71	7,37	6,94
<b>Aciertos Cumplen(%)</b>	76,00%	95,40%	99,60%	100,00%
<b>Aciertos No Cumplen(%)</b>	36,90%	7,70%	0,10%	0,00%
<b>Aciertos Totales(%)</b>	57,90%	54,90%	53,60%	53,80%

Así, por ejemplo, si el pH de una muestra es inferior a 6,31 se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen

sobre todo en valores altos del nivel de riesgo, oscilando su fiabilidad global entre un 53,6% y un 57,9% que no mejoran a la fiabilidad (53,8%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen y que se obtiene para niveles de riesgo superiores al 80% (ver Tabla 5.1.1.3.2). Todo ello refleja el *bajo poder discriminante de esta variable a la hora de determinar el nivel de calidad del agua*.

#### 5.1.1.4. Umbrales de calidad para la conductividad a 20°C

En la Tabla 5.1.1.4.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  como resultado del análisis exploratorio llevado a cabo en la sección 3.2.1.4.

**Tabla 5.1.1.4.1: Estimación de los parámetros del modelo (4.1) para la variable COND20**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	1,78	0,05	1287,33	1	0,00	5,94
$\alpha$	-11,53	0,32	1303,51	1	0,00	0,00

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de conductividad estarán asociados a un alto riesgo de contaminación cifrándose en un 1,78% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento de la conductividad del 1%*.

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 11,53}{1,78}} \quad (5.4)$$

de forma que se detectarán riesgos alto de contaminación si el nivel de conductividad es mayor o igual que dicho umbral. En la Tabla 5.1.1.4.2 se muestra el valor de (5.4) para una red de posibles valores de riesgo

**Tabla 5.1.1.4.2: Umbrales de calidad para la variable COND20 para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral ( $\mu\text{S/cm}$ )	646,28	811,43	1039,78	1407,07	2218,05	3373,63	8519,87
Aciertos Cumplen(%)	76,50%	83,60%	91,40%	97,20%	99,40%	100,00%	100,00%
Aciertos No Cumplen(%)	68,20%	55,90%	39,10%	21,50%	7,90%	1,20%	0,30%
Aciertos Totales(%)	72,70%	70,80%	67,20%	62,20%	57,10%	54,30%	53,90%

Así, por ejemplo, si el nivel de conductividad de una muestra es superior a 2218,05  $\mu\text{S/cm}$  se corre un riesgo superior al 90% de que las aguas estén contaminadas

en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 90%), oscilando su fiabilidad global entre un 53,9% y un 72,7%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 80%) significativamente más altos que la fiabilidad (53,8%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen (ver Tabla 3.1.4.1). Todo ello refleja un *poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.1.4.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

#### 5.1.1.5. Umbrales de calidad para el nivel de oxígeno disuelto en concentración

En la Tabla 5.1.1.5.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  de acuerdo con los resultados del análisis exploratorio llevado a cabo en la sección 3.2.1.5.

**Tabla 5.1.1.5.1: Estimación de los parámetros del modelo (4.1) para la variable O2**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	-0,17	0,01	149,68	1	0,00	0,84
$\alpha$	1,42	0,14	107,93	1	0,00	4,14

Se observa que el parámetro  $\beta$  es significativamente negativo. Por lo tanto, *altos niveles de O2 estarán asociados a un bajo riesgo de contaminación cifrándose en un 17% el decremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento en una unidad del contenido en O2.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \frac{1,42 - \log\left(\frac{p}{1-p}\right)}{0,17} \quad (5.5)$$

de forma que se detectará riesgo alto de contaminación si el nivel de O2 es menor o igual que dicho umbral. En la Tabla 5.1.1.5.2 se muestra el valor de (5.5) para una red de posibles valores de riesgo.

**Tabla 5.1.1.5.2: Umbrales de calidad para la variable O<sub>2</sub> para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	50,00%	60,00%	70,00%	80,00%
<b>Umbral (mg/L O<sub>2</sub>)</b>	8,36	5,97	3,37	0,20
<b>Aciertos Cumplen(%)</b>	83,70%	99,10%	99,90%	100,00%
<b>Aciertos No Cumplen(%)</b>	29,90%	5,30%	1,30%	0,00%
<b>Aciertos Totales(%)</b>	59,80%	57,4%	56,20%	55,60%

Así, por ejemplo, si el nivel de O<sub>2</sub> de una muestra es inferior a 0,20 mg/L O<sub>2</sub> se corre un riesgo superior al 80% de que las aguas estén contaminadas en el futuro. En la Tabla 5.1.1.5.3 se presenta el riesgo estimado correspondiente al umbral de calidad de la variable O<sub>2</sub> utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE).

**Tabla 5.1.1.5.3: Nivel de riesgo para el umbral de calidad de la variable O<sub>2</sub> utilizado por la CHE**

<b>Riesgo (100p%)</b>	63,90%
<b>Umbral (mg/L O<sub>2</sub>)</b>	5,00
<b>Aciertos Cumplen(%)</b>	99,80%
<b>Aciertos No Cumplen(%)</b>	2,80%
<b>Aciertos Totales(%)</b>	56,70%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo, oscilando su fiabilidad global entre un 55,6% y un 59,8% las cuales son ligeramente superiores a la fiabilidad (55,6%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen y que se obtiene para niveles de riesgo superiores al 80% (ver Tabla 5.1.1.5.3). Todo ello refleja el *bajo poder discriminante de esta variable a la hora de determinar el nivel de calidad del agua*.

### 5.1.1.6. Umbrales de calidad para el nivel de oxígeno disuelto en porcentaje de saturación

En la Tabla 5.1.1.6.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  como resultado del análisis exploratorio llevado a cabo en la sección 3.2.1.6.

**Tabla 5.1.1.6.1: Estimación de los parámetros del modelo (4.1) para la variable O2%**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	-0,02	0,00	78,86	1	0,00	0,98
$\alpha$	1,31	0,17	56,12	1	0,00	3,70

Se observa que el parámetro  $\beta$  es significativamente negativo. Por lo tanto, altos niveles de O2% estarán asociados a un bajo riesgo de contaminación cifrándose en un 2% el decremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del 1% en el contenido de oxígeno en % saturación.

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \frac{1,31 - \log\left(\frac{p}{1-p}\right)}{0,02} \quad (5.6)$$

de forma que se detectará riesgo alto de contaminación si el nivel de O2% es menor o igual que dicho umbral. En la Tabla 5.1.1.6.2 se muestra el valor de (5.6) para una red de posibles valores de riesgo

**Tabla 5.1.1.6.2: Umbrales de calidad para la variable O2% para distintos niveles de riesgo**

Riesgo (100%)	50,00%	60,00%	70,00%	75,00%
Umbral (%)	80,17	55,34	28,29	12,90
Aciertos Cumplen(%)	93,40%	99,60%	99,90%	100,00%
Aciertos No Cumplen(%)	16,00%	2,90%	0,20%	0,00%
Aciertos Totales(%)	59,10%	56,80%	56,00%	55,70%

Así, por ejemplo, si el nivel de O2% de una muestra es inferior al 12,9% se corre un riesgo superior al 75% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un

comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo, oscilando su fiabilidad global entre un 55,7% y un 59,1% las cuales son ligeramente superiores a la fiabilidad (55,7%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 75% (ver Tabla 5.1.1.6.2). Todo ello refleja el *bajo poder discriminante de esta variable a la hora de determinar el nivel de calidad del agua.*

#### 5.1.1.7. Umbrales de calidad para los sólidos en suspensión

En la Tabla 5.1.1.7.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.1.7.

**Tabla 5.1.1.7.1: Estimación de los parámetros del modelo (4.1) para la variable MES**

Parámetro	Estimación	Error Estándar	Wald	Gl	Pvalor	Exp(B)
$\beta$	0,51	0,02	500,13	1	0,00	1,67
$\alpha$	-1,40	0,06	567,60	1	0,00	0,25

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de sólidos en suspensión estarán asociados a un alto riesgo de contaminación cifrándose en un 0,51% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en sólidos en suspensión del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 1,40}{0,51}} \quad (5.7)$$

de forma que se detectará riesgo alto de contaminación si el nivel de sólidos en suspensión es mayor o igual que dicho umbral. En la Tabla 5.1.1.7.2 se muestra el valor de (5.7) para una red de posibles valores de riesgo

**Tabla 5.1.1.7.2: Umbrales de calidad para la variable MES para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
<b>Umbral (mg/L)</b>	15,66	34,67	82,44	237,19	1163,07	5033,45	128065,54
<b>Aciertos Cumplen(%)</b>	78,80%	90,90%	96,70%	98,90%	99,90%	100,00%	100,00%
<b>Aciertos No Cumplen(%)</b>	49,10%	26,50%	9,70%	2,20%	0,30%	0,10%	0,00%
<b>Aciertos Totales(%)</b>	65,80%	62,70%	58,60%	56,60%	56,30%	56,20%	56,20%

Así, por ejemplo, si el nivel de sólidos en suspensión de una muestra es superior a 82,44 mg/L se corre un riesgo superior al 70% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 90%), oscilando su fiabilidad global entre un 56,2% y un 65,8% siendo para niveles de riesgo inferiores al 70% medianamente más altos que la fiabilidad (56,2%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.1.7.2). Todo ello refleja un *poder discriminante medio de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.1.7.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

Conviene recordar, sin embargo, (ver sección 3.2.1.7) que *existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

## 5.1.2. Umbrales de calidad para los indicadores de materia orgánica

En este apartado presentamos los umbrales para el indicador de materia orgánica DQO\_D cuyo significado aparece en la Tabla 2.3.

### 5.1.2.1. Umbrales de calidad para la demanda química de oxígeno

En la Tabla 5.1.2.1.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  como consecuencia de los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.2.1.

**Tabla 5.1.2.1.1: Estimación de los parámetros del modelo (4.1) para la variable DQO**

Parámetro	Estimación	Error Estándar	Wald	GL	pvalor	Exp(B)
$\beta$	1,42	0,08	347,65	1	0,00	4,13
$\alpha$	-3,62	0,17	429,65	1	0,00	0,03

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de dicromato estarán asociados a un alto riesgo de contaminación cifrándose en un 1,42% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en dicromato del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 3,62}{1,42}} \quad (5.8)$$

de forma que se detectarán riesgos alto de contaminación si el nivel de dicromato es mayor o igual que dicho umbral. En la Tabla 5.1.2.1.2 se muestra el valor de (5.8) para una red de posibles valores de riesgo.

**Tabla 5.1.2.1.2: Umbrales de calidad para la variable DQO para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%
Umbral (mg/L O <sub>2</sub> )	12,85	17,11	23,37	34,19
Aciertos Cumplen(%)	86,30%	93,50%	97,30%	100,00%
Aciertos No Cumplen(%)	39,40%	17,80%	7,70%	0,00%
Aciertos Totales(%)	68,90%	65,40%	64,00%	62,80%

Así, por ejemplo, si el nivel de calcio de una muestra es superior a 23,37 mg/L O<sub>2</sub> se corre un riesgo superior al 70% de que las aguas estén contaminadas en el futuro. En la Tabla 5.1.1.5.3 se presenta el riesgo correspondiente al umbral de calidad de la

variable DQO utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE), estimándose en un 55,45% el nivel de riesgo de que el nivel de calidad del agua no sea aceptable si se utiliza dicho umbral.

**Tabla 5.1.2.1.3: Nivel de riesgo para el umbral de calidad de la variable DQO utilizado por la CHE**

<b>Riesgo (100p%)</b>	55,45%
<b>Umbral (mg/L O<sub>2</sub>)</b>	15,00
<b>Aciertos Cumplen(%)</b>	90,30%
<b>Aciertos No Cumplen(%)</b>	27,90%
<b>Aciertos Totales(%)</b>	67,10%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 70%), oscilando su fiabilidad global entre un 62,8% y un 68,9% siendo para niveles de riesgo inferiores al 70% medianamente más altos que la fiabilidad (62,8%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 80% (ver Tabla 5.1.2.1.2). Todo ello refleja un *poder discriminante medio de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.2.1.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

Conviene recordar, sin embargo, (ver sección 3.2.2.1) que *existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.3. Umbrales de calidad para los indicadores de acidificación

En este apartado presentamos los umbrales para todos los indicadores de acidificación CO3CA y CO32CA cuyo significado aparece detallado en la Tabla 2.3.

#### 5.1.3.1. Umbrales de calidad para los carbonatos

En la Tabla 5.1.3.1.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = (I(X>5)*\ln(X), I(X\leq 5))$  donde  $I(X\leq 5)$  e  $I(X>5)$  denotan las variables indicadoras de  $X\leq 5$  y  $X>5$ , respectivamente, de acuerdo con el carácter mixto de dicha variable puesto de manifiesto en los resultados del análisis exploratorio llevado a cabo en la sección 3.2.3.1.

**Tabla 5.1.3.1.1: Estimación de los parámetros del modelo (4.1) para la variable CO3CA**

Parámetro	Estimación	Error Estándar	Wald	GI	Pvalor	Exp(B)
$\beta$	1,19	0,31	14,95	1	0,00	3,29
$\beta_{CO3CA\leq 5}$	1,56	0,31	25,46	1	0,00	4,78
$\alpha$	-2,99	0,77	14,99	1	0,00	0,05

Se observa que tanto el parámetro  $\beta$  como el coeficiente de la variable indicadora de censura  $CO3CA\leq 5$  son significativamente positivos. Por lo tanto, *altos niveles de carbonatos estarán asociados a un alto riesgo de contaminación cifrándose en un 1,19% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en carbonatos del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 2,99}{1,19}} \quad (5.9)$$

de forma que, se detectará un riesgo alto de contaminación si el nivel de carbonatos es mayor o igual que dicho umbral. En la Tabla 5.1.3.1.2 se muestra el valor de (5.9) para una red de posibles valores de riesgo

**Tabla 5.1.3.1.2: Umbrales de calidad para la variable CO<sub>3</sub>CA para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	50,00%	60,00%	70,00%	80,00%	90,00%
<b>Umbral (mg/L CO<sub>3</sub>CA)</b>	12,28	17,27	25,03	39,36	77,79
<b>Aciertos Cumplen(%)</b>	83,86%	94,30%	99,37%	100,00%	100,00%
<b>Aciertos No Cumplen(%)</b>	16,20%	9,39%	1,88%	0,23%	0,00%
<b>Aciertos Totales(%)</b>	45,01%	45,55%	43,40%	42,72%	42,59%

Así, por ejemplo, si el nivel de calcio de una muestra es superior a 25,05 mg/L CO<sub>3</sub>CA se corre un riesgo superior al 70% de que las aguas estén contaminadas en el futuro. La fórmula anterior es válida siempre y cuando el valor del umbral sea mayor que 5. En la Tabla 5.1.3.1.3 se muestra el riesgo estimado así como los diversos porcentajes de acierto si se toma como umbral 5. Se observa que no es muy conveniente utilizar dicho umbral dado el bajo nivel de riesgo que comporta lo cual llevaría a un bajo % de aciertos (57,59%) entre los datos que cumplen con los niveles de calidad y a una visión excesivamente negativa del nivel de calidad de las aguas de la cuenca.

**Tabla 5.1.3.1.3: Riesgo estimado si se toma como umbral 5 mg/L CO<sub>3</sub>CA**

<b>Riesgo (100p%)</b>	19,37%
<b>Aciertos Cumplen (%)</b>	57,59%
<b>Aciertos No Cumplen(%)</b>	30,00%
<b>Aciertos Totales(%)</b>	41,78%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 70%), oscilando su fiabilidad global entre un 42,59% y un 45,01% claramente inferiores a la regla de la mayoría que, en este caso, clasificaría todos los casos como que no cumplen con una fiabilidad del  $100\% - 42,59\% = 57,41\%$  (ver Tabla 5.1.3.1.2). Este *mal comportamiento del modelo se debe, muy probablemente, al alto porcentaje de censura (64,9%) existente en la muestra.*

Conviene recordar, además, (ver sección 3.2.3.1) que *existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra*

*no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.3.2. Umbrales de calidad para los bicarbonatos

En la Tabla 5.1.3.2.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = X$  de acuerdo a los resultados del análisis exploratorio llevado a cabo en la sección 3.2.3.2.

**Tabla 5.1.3.2.1: Estimación de los parámetros del modelo (4.1) para la variable CO32CA**

Parámetros	Estimación	Error Estándar	Wald	Gl	pvalor	Exp(B)
$\beta$	0,01	0,00	95,88	1	0,00	1,01
$\alpha$	-2,29	0,26	75,30	1	0,00	0,10

Se observa que el coeficiente  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de bicarbonatos estarán asociados a un alto riesgo de contaminación cifrándose en un 1% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento en una unidad del contenido en bicarbonatos.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \frac{\log\left(\frac{p}{1-p}\right) + 2.29}{0.01} \quad (5.10)$$

de forma que, se detectará riesgo alto de contaminación si el nivel de bicarbonatos es mayor o igual que dicho umbral. En la Tabla 5.1.3.2.2 se muestra el valor de (5.10) para una red de posibles valores de riesgo

**Tabla 5.1.3.2.2: Umbrales de calidad para la variable CO32CA para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L CO <sub>3</sub> CA)	170,13	200,29	233,15	273,24	333,55	389,13	511,90
Aciertos Cumplen(%)	55,90%	74,10%	85,70%	93,40%	99,70%	100,00%	100,00%
Aciertos No Cumplen(%)	79,50%	56,40%	27,50%	13,00%	3,70%	1,50%	0,00%
Aciertos Totales(%)	69,00%	64,30%	53,40%	48,70%	46,40%	45,30%	44,40%

Así, por ejemplo, si el nivel de calcio de una muestra es superior a 333,55 mg/L CO<sub>3</sub>CA se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de

fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 80%), oscilando su fiabilidad global entre un 44,40% y un 69%. Dicha ***fiabilidad es de tipo medio para niveles bajos de riesgo (menos de 60%)*** que la regla de la mayoría que, en este caso, clasificaría todos los casos como que no cumplen con una fiabilidad del  $100\% - 44,4\% = 55,60\%$  (ver Tabla 5.1.3.2.2).

Conviene recordar, sin embargo, que ***existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.***

#### 5.1.4. Umbrales de calidad para los indicadores de salinidad

En este apartado presentamos los umbrales para todos los indicadores de salinidad NA, CA, MG, CL y SO4 cuyo significado puede verse en la Tabla 2.3.

##### 5.1.4.1. Umbrales de calidad para los niveles de sodio

En la Tabla 5.1.4.1.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  como resultado del análisis exploratorio llevado a cabo en la sección 3.2.4.1.

**Tabla 5.1.4.1.1: Estimación de los parámetros del modelo (4.1) para la variable NA**

Parámetro	Estimación	Error Estándar	Wald	Gl	pvalor	Exp(B)
$\beta$	0,94	0,05	296,78	1	0,00	2,55
$\alpha$	-3,09	0,20	247,89	1	0,00	0,05

Se observa que el coeficiente  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de sodio estarán asociados a un alto riesgo de contaminación cifrándose en un 0,94% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en sodio del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 3.09}{0.94}} \quad (5.11)$$

de forma que se detectará riesgo alto de contaminación si el nivel de sodio es mayor o igual que dicho umbral. En la Tabla 5.1.4.1.2 se muestra el valor de (5.11) para una red de posibles valores de riesgo

**Tabla 5.1.4.1.2: Umbrales de calidad para la variable NA para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L Na)	26,99	41,61	66,68	118,53	281,65	625,26	3640,79
Aciertos Cumplen(%)	68,60%	79,10%	85,00%	94,00%	99,70%	100,00%	100,00%
Aciertos No Cumplen(%)	79,70%	65,60%	53,30%	32,90%	7,60%	0,80%	0,30%
Aciertos Totales(%)	74,40%	72,00%	68,40%	62,00%	51,50%	48,10%	47,80%

Así, por ejemplo, si el nivel de sodio de una muestra es superior a 281,65 mg/L Na se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de

fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 90%), oscilando su fiabilidad global entre un 47,80% y un 74,40%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 80%) significativamente más altos que la fiabilidad (100%-47,8% = 52,2%, ver Tabla 5.1.4.1.2) de la regla de la mayoría que clasificaría todos los datos de la muestra como que no cumplen. Todo ello refleja un *poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.4.1.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, (ver sección 3.2.4.1) que existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### 5.1.4.2. Umbrales de calidad para los niveles de calcio

En la Tabla 5.1.4.2.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1) , en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.4.2.

**Tabla 5.1.4.2.1: Estimación de los parámetros del modelo (4.1) para la variable CA**

Parámetro	Estimación	Error Estándar	Wald	G1	pvalor	Exp(B)
$\beta$	1,98	0,15	164,08	1	0,00	7,27
$\alpha$	-8,64	0,69	156,49	1	0,00	0,00

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de calcio estarán asociados a un alto riesgo de contaminación cifrándose en un 1,98% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en calcio del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 8.64}{1.98}} \quad (5.12)$$

de forma que se detectará riesgo alto de contaminación si el nivel de calcio es mayor o igual que dicho umbral. En la Tabla 5.1.4.2.2 se muestra el valor de (5.12) para una red de posibles valores de riesgo

**Tabla 5.1.4.2.2: Umbrales de calidad para la variable CA para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L Ca)	77,95	95,62	119,48	156,78	235,96	343,89	790,33
Aciertos Cumplen(%)	64,30%	74,70%	83,90%	92,60%	98,70%	99,60%	100,00%
Aciertos No Cumplen(%)	80,20%	62,50%	38,70%	19,70%	4,70%	1,40%	0,00%
Aciertos Totales(%)	73,40%	68,00%	59,00%	52,50%	47,10%	45,60%	45,10%

Así, por ejemplo, si el nivel de calcio de una muestra es superior a 235,96 mg/L Ca se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 80%), oscilando su fiabilidad global entre un 45,10% y un 73,40%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 70%) significativamente más altos que la fiabilidad ( $100\% - 45,1\% = 54,9\%$ , ver Tabla 5.1.4.2.2) de la regla de la mayoría que clasificaría todos los datos de la muestra como que no cumplen. Todo ello refleja un *poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.4.2.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, (ver sección 3.2.4.2) que existe una sobre representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.4.3. Umbrales de calidad para los niveles de magnesio

En la Tabla 5.1.4.3.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en la sección 3.2.4.3.

**Tabla 5.1.4.3.1: Estimación de los parámetros del modelo (4.1) para la variable MG**

Parámetro	Estimación	Error Estándar	Wald	Gl	pvalor	Exp(B)
$\beta$	0,95	0,07	170,37	1	0,00	2,59
$\alpha$	-2,40	0,20	144,89	1	0,00	0,09

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de magnesio estarán asociados a un alto riesgo de contaminación cifrándose en un 0,95% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en magnesio del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 2,40}{0,95}} \quad (5.13)$$

de forma que se detectará riesgo alto de contaminación si el nivel de magnesio es mayor o igual que dicho umbral. En la Tabla 5.1.4.3.2 se muestra el valor de (5.13) para una red de posibles valores de riesgo.

**Tabla 5.1.4.3.2: Umbrales de calidad para la variable MG para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%
Umbral (mg/L Mg)	12,38	18,95	30,12	53,03	124,18	271,98
Aciertos Cumplen(%)	65,20%	75,90%	88,50%	95,80%	100,00%	100,00%
Aciertos No Cumplen(%)	72,60%	52,20%	29,90%	9,30%	0,80%	0,00%
Aciertos Totales(%)	69,10%	63,50%	57,90%	50,60%	48,10%	47,70%

Así, por ejemplo, si el nivel de magnesio de una muestra es superior a 124,18 mg/L Mg se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los

casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 80%), oscilando su fiabilidad global entre un 47,70% y un 69,10%, siendo, para niveles de riesgo inferiores al 60%. moderadamente más altas que la fiabilidad (100%-47,7% = 52,3%, ver Tabla 5.1.4.3.2) de la regla de la mayoría que clasificaría todos los datos de la muestra como que no cumplen. Todo ello refleja un *poder discriminante medio de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.4.3.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

Conviene recordar, sin embargo, que existe una *sobre-representación de datos que no cumplen los niveles de calidad mínimos (ver sección 3.2.4.3), lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### 5.1.4.4. Umbrales de calidad para los cloruros

En la Tabla 5.1.4.4.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.4.4.

**Tabla 5.1.4.4.1: Estimación de los parámetros del modelo (4.1) para la variable CL**

Parámetro	Estimación	Error Estándar	Wald	G1	pvalor	Exp(B)
$\beta$	1,21	0,05	622,32	1	0,00	3,36
$\alpha$	-4,57	0,18	624,96	1	0,00	0,01

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de cloruros estarán asociados a un alto riesgo de contaminación cifrándose en un 1,21% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en cloruros del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 4,57}{1,21}} \quad (5.14)$$

de forma que se detectará riesgo alto de contaminación si el nivel de cloruros es mayor o igual que dicho umbral. En la Tabla 5.1.4.4.2 se muestra el valor de (5.14) para una red de posibles valores de riesgo

**Tabla 5.1.4.4.2: Umbrales de calidad para la variable CL para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
<b>Umbral (mg/L Cl)</b>	43,57	60,89	87,70	136,84	267,26	495,23	1934,54
<b>Aciertos Cumplen(%)</b>	83,80%	88,90%	91,60%	94,70%	99,30%	99,80%	100,00%
<b>Aciertos No Cumplen(%)</b>	73,60%	64,10%	55,70%	38,80%	13,10%	3,20%	0,30%
<b>Aciertos Totales(%)</b>	79,30%	78,00%	75,90%	70,20%	61,60%	57,60%	56,40%

Así, por ejemplo, si el nivel de cloruros de una muestra es superior a 267,26 mg/L Cl se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 95%), oscilando su fiabilidad global entre un 56,4% y un 79,3%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 90%) significativamente más altos que la fiabilidad (56,4%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.4.4.2). Todo ello refleja *un poder discriminante alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.4.4.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

Conviene recordar, sin embargo, que existe una *sobre-representación de datos que cumplen los niveles de calidad mínimos (ver sección 3.2.4.4), lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.4.5. Umbrales de calidad para los sulfatos

En la Tabla 5.1.4.5.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en la sección 3.2.4.5.

**Tabla 5.1.4.5.1: Estimación de los parámetros del modelo (4.1) para la variable SO4**

Parámetro	Estimación	Error Estándar	Wald	GI	Pvalor	Exp(B)
$\beta$	1,00	0,05	344,60	1	0,00	2,72
$\alpha$	-4,53	0,25	332,25	1	0,00	0,01

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de sulfatos estarán asociados a un alto riesgo de contaminación cifrándose en un 1% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en sulfatos del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\log\left(\frac{p}{1-p}\right) + 4,53} \quad (5.15)$$

de forma que se detectará riesgo alto de contaminación si el nivel de sulfatos es mayor o igual que dicho umbral. En la Tabla 5.1.4.5.2 se muestra el valor de (5.15) para una red de posibles valores de riesgo

**Tabla 5.1.4.5.2: Umbrales de calidad para la variable SO4 para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%
Umbral (mg/L SO <sub>4</sub> )	91,81	137,61	213,90	366,34	823,12	1735,47
Aciertos Cumplen(%)	72,70%	82,20%	87,70%	95,20%	99,60%	100,00%
Aciertos No Cumplen(%)	74,40%	59,60%	37,30%	16,60%	3,00%	0,00%
Aciertos Totales(%)	73,50%	71,30%	63,50%	57,40%	53,10%	51,90%

Así, por ejemplo, si el nivel de sulfatos de una muestra es superior a 823,12 mg/L SO<sub>4</sub> se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al

80%), oscilando su fiabilidad global entre un 51,9% y un 73,5%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 70%) significativamente más altos que la fiabilidad (51,9%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 95% (ver Tabla 5.1.4.5.2). Todo ello refleja un *poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.4.5.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos (ver sección 3.2.4.5), lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.5. Umbrales de calidad para los indicadores de nutrientes

En este apartado presentamos los umbrales para todos los indicadores de nutrientes NH<sub>4</sub>, NO<sub>2</sub>, NO<sub>3</sub>, N\_KJED, PO<sub>4</sub> y P\_TOT cuyo significado puede verse en la Tabla 2.3.

#### 5.1.5.1. Umbrales de calidad para el nivel de amonio

En la Tabla 5.1.5.1.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$ , de acuerdo a los resultados obtenidos en la sección 3.2.5.1.

**Tabla 5.1.5.1.1: Estimación de los parámetros del modelo (4.1) para la variable NH<sub>4</sub>**

Parámetro	Estimación	Error Estándar	Wald	G1	pvalor	Exp(B)
$\beta$	1,02	0,07	190,20	1,00	0,00	2,77
$\alpha$	1,77	0,15	133,03	1,00	0,00	5,87

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de amonio estarán asociados a un alto de riesgo de contaminación cifrándose en un 1,02% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en amonio del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) - 1,77}{1,02}} \quad (5.16)$$

de forma que se detectará riesgo alto de contaminación si el nivel de amonio es mayor o igual que dicho umbral. En la Tabla 5.1.5.1.2 se muestra el valor de (5.16) para una red de posibles valores de riesgo

**Tabla 5.1.5.1.2: Umbrales de calidad para la variable NH<sub>4</sub> para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L NH <sub>4</sub> )	0,18	0,26	0,41	0,69	1,52	3,16	15,95
Aciertos Cumplen(%)	94,60%	96,94%	97,93%	99,20%	99,84%	99,90%	100,00%
Aciertos No Cumplen(%)	28,34%	21,95%	16,55%	10,75%	4,69%	2,48%	0,20%
Aciertos Totales(%)	65,23%	63,70%	61,85%	59,98%	57,67%	56,71%	55,80%

Así, por ejemplo, si el nivel de amonio de una muestra es superior a 1,52 mg/L NH<sub>4</sub> se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Observar que, a pesar del carácter mixto de esta variable (ver sección 3.2.5.1), *la*

*influencia de la censura no es significativa concordando los resultados obtenidos en el proceso de determinación de umbrales tanto si se tiene en cuenta como si no.* En la Tabla 5.1.5.1.3 se presenta el riesgo correspondiente al umbral de calidad de la variable NH4 utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE).

**Tabla 5.1.5.1.3: Nivel de riesgo para el umbral de calidad de la variable NH4 utilizado por la CHE**

<b>Riesgo (100p%)</b>	69,73%
<b>Umbral (mg(L NH<sub>4</sub>))</b>	0,40
<b>Aciertos Cumplen (%)</b>	97,90%
<b>Aciertos No Cumplen(%)</b>	16,5%
<b>Aciertos Totales(%)</b>	61,9%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 90%), oscilando su fiabilidad global entre un 55,67% y un 65,23%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 80%) significativamente más altos que la fiabilidad (55,67%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.5.1.2). Todo ello refleja un *poder discriminante medio de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.5.1.8) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre representación de datos que cumplen los niveles de calidad mínimos (ver sección 3.2.5.1), lo cual puede distorsionar al alza, las estimaciones de la probabilidad de que una muestra cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.5.2. Umbrales de calidad para los nitritos

En la Tabla 5.1.5.2.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados del análisis exploratorio llevado a cabo en la sección 3.2.5.2.

**Tabla 5.1.5.2.1: Estimación de los parámetros del modelo (4.1) para la variable NO2**

Parámetro	Estimación	Error Estándar	Wald	Gl	pvalor	Exp(B)
$\beta$	1,46	0,07	384,81	1	0,00	4,32
$\alpha$	4,51	0,24	346,43	1	0,00	90,92

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de nitritos estarán asociados a un alto riesgo de contaminación cifrándose en un 1,46% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en nitritos del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) - 4,51}{1,46}} \quad (5.17)$$

de forma que se detectará riesgo alto de contaminación si el nivel de nitritos es mayor o igual que dicho umbral. En la Tabla 5.1.5.2.2 se muestra el valor de (5.17) para una red de posibles valores de riesgo.

**Tabla 5.1.5.2.2: Umbrales de calidad para la variable NO2 para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L NO <sub>2</sub> )	0,05	0,06	0,08	0,12	0,21	0,34	1,06
Aciertos Cumplen(%)	71,90%	88,40%	92,20%	96,40%	98,60%	99,50%	100,00%
Aciertos No Cumplen(%)	75,20%	55,90%	44,30%	33,90%	19,20%	11,00%	0,00%
Aciertos Totales(%)	73,50%	72,60%	68,90%	66,00%	60,00%	56,40%	52,50%

Así, por ejemplo, si el nivel de calcio de una muestra es superior a 0,21 mg/L NO<sub>2</sub> se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. En la Tabla 5.1.5.2.3 se presenta el riesgo estimado correspondiente al umbral de calidad de la variable NO2 utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE).

**Tabla 5.1.5.2.3: Nivel de riesgo para el umbral de calidad de la variable NO<sub>2</sub> utilizado por la CHE**

<b>Riesgo (100p%)</b>	85,00%
<b>Umbral (mg/L NO<sub>2</sub>)</b>	0,15
<b>Aciertos Cumplen (%)</b>	97,80%
<b>Aciertos No Cumplen(%)</b>	24,90%
<b>Aciertos Totales(%)</b>	62,40%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 95%), oscilando su fiabilidad global entre un 52,5% y un 73,5%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 90%) significativamente más altos que la fiabilidad (52,5%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.5.2.2). Todo ello *refleja un poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.5.2.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre representación de datos que no cumplen los niveles de calidad mínimos (ver sección 3.2.5.2), lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

### 5.1.5.3. Umbrales de calidad para los nitratos

En la Tabla 5.1.5.3.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo a los resultados obtenidos en la sección 3.2.5.3.

**Tabla 5.1.5.3.1: Estimación de los parámetros del modelo (4.1) para la variable NO3**

Parámetro	Estimación	Error Estándar	Wald	Gl	Pvalor	Exp(B)
$\beta$	1,34	0,05	607,50	1	0,00	3,83
$\alpha$	-2,59	0,12	489,12	1	0,00	0,08

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de nitratos estarán asociados a un alto de riesgo de contaminación cifrándose en un 1,34% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en nitratos del 1%*.

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) + 2,59}{1,34}} \quad (5.18)$$

de forma que se detectará riesgo alto de contaminación si el nivel de nitratos es mayor o igual que dicho umbral. En la Tabla 5.1.5.3.2 se muestra el valor de (5.18) para una red de posibles valores de riesgo

**Tabla 5.1.5.3.2: Umbrales de calidad para la variable NO3 para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L NO <sub>3</sub> )	6,87	9,29	12,91	19,30	35,30	61,60	210,67
Aciertos Cumplen(%)	69,20%	77,50%	88,00%	95,20%	99,50%	100,00%	100,00%
Aciertos No Cumplen(%)	82,10%	69,10%	44,10%	17,30%	4,50%	0,20%	0,00%
Aciertos Totales(%)	75,60%	73,40%	66,40%	56,90%	52,80%	51,00%	50,90%

Así, por ejemplo, si el nivel de nitratos de una muestra es superior a 35,3 mg/L NO<sub>3</sub> se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. En la Tabla 5.1.5.3.3 se presenta el riesgo estimado correspondiente al umbral de calidad de la variable NO<sub>3</sub> utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE).

**Tabla 5.1.5.3.3: Nivel de riesgo para el umbral de calidad de la variable NO3 utilizado por la CHE**

<b>Riesgo (100p%)</b>	80,76%
<b>Umbral ((mg/L NO<sub>3</sub>))</b>	20,00
<b>Aciertos Cumplen (%)</b>	95,80%
<b>Aciertos No Cumplen(%)</b>	15,80%
<b>Aciertos Totales(%)</b>	56,50%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 80%), oscilando su fiabilidad global entre un 50,9% y un 75,6%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 70%) significativamente más altos que la fiabilidad (50,9%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.5.3.2). Todo ello refleja un *poder discriminante medio-alto de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.5.3.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### 5.1.5.4. Umbrales de calidad para el nivel de nitrógeno Kjeldahl

En la Tabla 5.1.5.4.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(\ln(X))$  de acuerdo a los resultados obtenidos en el análisis exploratorio llevado a cabo en la sección 3.2.5.4.

**Tabla 5.1.5.4.1: Estimación de los parámetros del modelo (4.1) para la variable N\_KJED**

<b>Parámetro</b>	<b>Estimación</b>	<b>Error Estándar</b>	<b>Wald</b>	<b>GI</b>	<b>Pvalor</b>	<b>Exp(B)</b>
<b><math>\beta</math></b>	0,50	0,09	28,88	1,00	0,00	1,65
<b><math>\alpha</math></b>	1,69	0,27	38,16	1,00	0,00	5,40

Se observa que el valor del parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de nitrógeno Kjeldahl estarán asociados a un alto de riesgo de contaminación cifrándose en un 0,50% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en log(nitrógeno Kjeldahl) del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = \exp \left[ \exp \left[ \frac{\log \left( \frac{p}{1-p} \right) - 1,69}{0,50} \right] \right] \quad (5.19)$$

de forma que se detectará riesgo alto de contaminación si el nivel de nitrógeno Kjeldahl es mayor o igual que dicho umbral. En la Tabla 5.1.5.4.2 se muestra el valor de (5.19) para una red de posibles valores de riesgo

**Tabla 5.1.5.4.2: Umbrales de calidad para la variable N\_KJED para distintos niveles de riesgo**

Riesgo (100p%)	60,00%	70,00%	80,00%	90,00%
Umbral (mg/L N)	1,08	1,20	1,73	16,08
Aciertos Cumplen(%)	94,92%	96,92%	98,15%	100,00%
Aciertos No Cumplen(%)	14,24%	11,06%	6,00%	0,10%
Aciertos Totales(%)	49,17%	48,23%	45,90%	43,40%

Así, por ejemplo, si el nivel de nitrógeno Kjeldahl de una muestra es superior a 1,2 mg/L N se corre un riesgo superior al 70% de que las aguas estén contaminadas en el futuro. Observar que, a pesar del carácter mixto de esta variable (ver sección 3.2.5.4), la influencia de la censura no es significativa concordando los resultados obtenidos en el proceso de determinación de umbrales tanto si se tiene en cuenta como si no. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 70%), oscilando su fiabilidad global entre un 43,4% y un 49,23%, valores claramente inferiores a 100%-43,3% = 56,7% (ver Tabla 5.1.5.4.2) que es la fiabilidad de la regla de la mayoría que clasificaría todos los datos de la muestra como que no cumplen. Todo ello refleja un

*poder discriminante muy bajo de esta variable a la hora de determinar el nivel de calidad del agua muy probablemente debido al alto porcentaje de censura (89,5% de los datos observados, ver sección 3.2.5.4)*

*Conviene recordar, además, que existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos (ver sección 3.2.5.4), lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### 5.1.5.5. Umbrales de calidad para los fosfatos

En la Tabla 5.1.5.5.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo con los resultados del análisis exploratorio llevado a cabo en la sección 3.2.5.5.

**Tabla 5.1.5.5.1: Estimación de los parámetros del modelo (4.1) para la variable PO4**

Parámetro	Estimación	Error Estándar	Wald	G1	Pvalor	Exp(B)
$\beta$	1,37	0,07	401,09	1	0,00	3,92
$\alpha$	2,75	0,15	339,84	1	0,00	15,59

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de fosfatos estarán asociados a un alto de riesgo de contaminación cifrándose en un 1,37 % el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en fosfatos del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) - 2,75}{1,37}} \quad (5.20)$$

de forma que se detectará riesgo alto de contaminación si el nivel de fosfatos es mayor o igual que dicho umbral. En la Tabla 5.1.5.5.2 se muestra el valor de (5.20) para una red de posibles valores de riesgo

**Tabla 5.1.5.5.2: Umbrales de calidad para la variable PO4 para distintos niveles de riesgo**

<b>Riesgo (100p%)</b>	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
<b>Umbral (mg/L PO<sub>4</sub>)</b>	0,13	0,18	0,25	0,37	0,67	1,16	3,87
<b>Aciertos Cumplen(%)</b>	82,10%	88,40%	91,60%	95,80%	98,90%	99,60%	100,00%
<b>Aciertos No Cumplen(%)</b>	59,10%	47,20%	35,50%	20,50%	8,50%	5,00%	0,40%
<b>Aciertos Totales(%)</b>	70,80%	68,10%	64,00%	58,80%	54,40%	53,10%	51,00%

Así, por ejemplo, si el nivel de fosfatos de una muestra es superior a 0,67 mg/L PO<sub>4</sub> se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. En la Tabla 5.1.5.5.3 se presenta el riesgo estimado correspondiente al umbral de calidad de la variable PO<sub>4</sub> utilizado hasta el momento por la Confederación Hidrográfica del Ebro (CHE).

**Tabla 5.1.5.5.3: Nivel de riesgo para el umbral de calidad de la variable PO4 utilizado por la CHE**

<b>Riesgo (100p%)</b>	75,07%
<b>Umbral (mg/L PO<sub>4</sub>)</b>	0,30
<b>Aciertos Cumplen (%)</b>	94,10%
<b>Aciertos No Cumplen(%)</b>	26,40%
<b>Aciertos Totales(%)</b>	60,80%

Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 80%), oscilando su fiabilidad global entre un 51% y un 70,8%, siendo para la mayor parte de los casos (niveles de riesgo inferiores al 70%) significativamente más altos que la fiabilidad (51%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.5.5.2). Todo ello refleja un *poder discriminante medio de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.5.5.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre-representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza,*

*las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### 5.1.5.6. Umbrales de calidad para el nivel de fósforo

En la Tabla 5.1.5.6.1 se muestran las estimaciones de los parámetros del modelo de regresión logística (4.1), en el que se ha tomado  $g(X) = \ln(X)$  de acuerdo con los resultados del análisis exploratorio llevado a cabo en la sección 3.2.5.6.

**Tabla 5.1.5.6.1: Estimación de los parámetros del modelo (4.1) para la variable P\_TOT**

Parámetro	Estimación	Error Estándar	Wald	Gl	Pvalor	Exp(B)
$\beta$	0,96	0,07	182,99	1	0,00	2,60
$\alpha$	2,44	0,19	165,89	1	0,00	11,45

Se observa que el parámetro  $\beta$  es significativamente positivo. Por lo tanto, *altos niveles de fósforo estarán asociados a un alto de riesgo de contaminación cifrándose en un 0,96% el incremento del riesgo de que el agua presente niveles de calidad no aceptables por cada incremento del contenido en fósforo total del 1%.*

Utilizando la expresión (4.2) el umbral para un riesgo  $0 < p < 1$  vendrá dado por:

$$\text{Umbral}_p = e^{\frac{\log\left(\frac{p}{1-p}\right) - 2,44}{0,96}} \quad (5.21)$$

de forma que se detectará riesgos alto de contaminación si el nivel de fósforo total es mayor o igual que dicho umbral. En la Tabla 5.1.5.6.2 se muestra el valor de (5.21) para una red de posibles valores de riesgo

**Tabla 5.1.5.6.2: Umbrales de calidad para la variable P\_TOT para distintos niveles de riesgo**

Riesgo (100p%)	50,00%	60,00%	70,00%	80,00%	90,00%	95,00%	99,00%
Umbral (mg/L P)	0,08	0,12	0,19	0,33	0,78	1,70	9,57
Aciertos Cumplen(%)	50,80%	89,60%	96,20%	98,90%	99,90%	100,00%	100,00%
Aciertos No Cumplen(%)	62,90%	39,10%	20,70%	9,40%	3,30%	0,30%	0,00%
Aciertos Totales(%)	56,80%	64,40%	58,50%	54,30%	51,80%	50,30%	50,20%

Así, por ejemplo, si el nivel de fósforo total de una muestra es superior a 0,78 mg/L P se corre un riesgo superior al 90% de que las aguas estén contaminadas en el futuro. Así mismo, y con el fin de evaluar empíricamente el comportamiento del umbral utilizando los datos de la muestra, se presentan los valores de los indicadores de fiabilidad dados por las expresiones (4.5), (4.6) y (4.7) de la sección anterior. Se

observa que el umbral presenta un comportamiento excelente en el %Aciertos Cumplen pero a costa de un comportamiento claramente deficiente a la hora de identificar los casos que no cumplen sobre todo en valores altos del nivel de riesgo (superiores al 70%), oscilando su fiabilidad global entre un 50,2% y un 64,4%, siendo para valores intermedios del nivel de riesgo (entre el 60 y el 70%) significativamente más altos que la fiabilidad (50,2%) de la regla de la mayoría que clasificaría todos los datos de la muestra como que cumplen, la cual se obtiene para niveles de riesgo superiores al 99% (ver Tabla 5.1.5.6.2). Todo ello refleja un *poder discriminante medio-bajo de esta variable a la hora de determinar el nivel de calidad del agua, ya puesto de manifiesto en su distribución geográfica (ver Figura 3.2.5.6.7) que tiende a ser similar a la del nivel de calidad (ver Figura 3.1.4.1).*

*Conviene recordar, sin embargo, que existe una sobre representación de datos que no cumplen los niveles de calidad mínimos, lo cual pueden distorsionar al alza, las estimaciones de la probabilidad de que una muestra no cumpla los estándares de calidad, por lo que estas estimaciones deberían tomarse con las debidas precauciones.*

#### **5.1.6 Resumen de los resultados obtenidos**

En la Tabla 5.1.6. se muestra un resumen de los resultados obtenidos junto con una estimación de los umbrales correspondientes a un riesgo empírico del 5% (%Aciertos que cumplen = 95%) entre las muestras que cumplen los requisitos de calidad. Se observa que, para dicho nivel de riesgo, la fiabilidad de la mayor parte de los indicadores es superior al 55%. Destacan, en este sentido los cloruros (69,8%), nitritos (68,2%), la conductividad a 20° C (65,4%), el nivel de amonio (65%) y el de demanda química de oxígeno (64,9%) cuya fiabilidad oscila entre el 65% y el 70% mostrando (para este nivel de riesgo) un mayor poder discriminante que el resto. Conviene hacer notar, sin embargo, que con la única excepción del nivel de conductividad, en todos los demás parámetro físico-químicos se aprecia la existencia de un porcentaje no despreciable de datos ausentes los cuales presentan patrones sistemáticos que podrían sesgar los resultados obtenidos, por lo que los comentarios anteriores deberían tomarse con las debidas precauciones.

Así mismo se muestran las fiabilidades correspondientes a los umbrales que ha venido utilizando la CHE para algunos de los indicadores anteriores. Dichos umbrales tienen un riesgo empírico en las muestras que cumplen que oscila entre un 2,1% y un 9,7% (% aciertos que cumplen 97,9% y 90,3%) y un nivel de fiabilidad de tipo medio

que oscila entre un 56,5% y un 67,1% siendo el umbral correspondiente a la demanda química de oxígeno y el más fiable de todos ellos.

**Tabla 5.1.6: Resumen de los resultados obtenidos**

Tipo Indicador	Parámetro	Relación	Umbral	%Aciertos Cumplen	Fiabilidad	Umbral CHE	%Aciertos Cumplen	Fiabilidad	Sesgo Missing y censura
Generales	TEM_AG	Directa	22,1	95,00%	57,10%				No
	TEM_AI	Directa	29,5	95,00%	55,30%				No
	PH	Inversa	7,71	95,00%	54,90%				No
	COND20	Directa	1180	95,00%	65,40%				No
	O2	Inversa	7,3	95,00%	59,30%				No
	O2%	Inversa	78	95,00%	58,80%				No
	MES	Directa	62,35	95,00%	59,30%				No
Materia Orgánica	DQO	Directa	19	95,00%	64,90%	15	90,30%	67,10%	Sobre-representación no cumplen
Acidificación	CO3CA	Directa	18,075	95,00%	62,13%				Sobre-representación no cumplen y censura significativa
	CO32CA	Directa	280,4	95,00%	49,00%				Sobre-representación no cumplen
Salinidad	NA	Directa	126	95,00%	60,70%				Sobre-representación no cumplen
	CA	Directa	184,8	95,00%	49,80%				Sobre-representación no cumplen
	MG	Directa	47,985	95,00%	51,50%				Sobre-representación no cumplen
	CL	Directa	141,34	95,00%	69,80%				Sobre-representación cumplen
	SO4	Directa	360,4	95,00%	57,70%				Sobre-representación no cumplen
Nutrientes	NH4	Directa	0,18	95,00%	65,00%	0,4	97,90%	61,90%	Sobre-representación cumplen y censura no significativa
	NO2	Directa	0,095	95,00%	68,20%	0,15	97,80%	62,40%	Sobre-representación no cumplen
	NO3	Directa	18,8	95,00%	57,70%	20	95,80%	56,50%	Sobre-representación no cumplen
	N_KJED	Directa	1,1	95,00%	48,70%				Sobre-representación no cumplen y censura significativa
	PO4	Directa	0,34	95,00%	59,40%	0,3	94,10%	60,80%	Sobre-representación no cumplen
	P_TOT	Directa	0,162	95,00%	60,10%				Sobre-representación no cumplen

## 5.2. UMBRALES MULTIVARIANTES

En esta sección se estiman modelos de regresión logística multivariantes de la forma (4.8) con el fin de analizar el signo y la potencia discriminante conjunta de los indicadores físico-químicos utilizados en el estudio, a la hora de predecir el nivel de calidad del agua. Dada la gran cantidad de datos ausentes existentes en la mayor parte de dichos indicadores, se ha optado por llevar a cabo el proceso de construcción de los modelos agrupando los indicadores físico-químicos de acuerdo al tipo listado en la Tabla 2.3. Así mismo, y con el fin de determinar aquéllos indicadores con mayor potencia discriminante, se ha llevado a cabo un proceso de selección de variables paso a paso tomando como niveles de significación de entrada y salida el 5%. Los resultados obtenidos se presentan a continuación.

### 5.2.1. Umbrales multivariantes para los indicadores generales

En este caso utilizamos como variables independientes del modelo TEM\_AG, TEM\_AI, PH, LCOND20, O2, O2% y LMES cuyo significado puede verse en la Tabla 2.3. La variable dependiente es el indicador de calidad conjunta definido en la sección 3.1.4. El número de casos utilizados en el proceso de selección fue 5678 que constituye un 85,6% de los datos de la muestra.

En las Tablas 5.2.1.1 y 5.2.1.2 se muestran los resultados obtenidos. Más concretamente, en la Tabla 5.2.1.1 se presentan las variables finalmente escogidas junto con el orden en el que fueron seleccionadas mientras que en la Tabla 5.2.1.2 se muestran los resultados obtenidos en el proceso de validación intramuestral de cada uno de los modelos estimados por el proceso de selección.

Las variables finalmente seleccionadas fueron (por este orden) LCOND20, LMES y pH y el modelo escogido viene dado por la expresión:

$$P(\text{No cumplir}) = -4,243 + 1,658 * \text{LCOND20} + 0,305 * \text{LMES} - 0,895 * \text{pH}$$

Este resultado concuerda con los obtenidos en el proceso de construcción de los umbrales univariantes de cada una de las variables independientes en cuanto al poder discriminante y al signo de la misma con respecto al nivel de calidad del agua. Más concretamente se observa que las variables con mayor poder discriminante son LCOND20, LMES y, en menor medida pH, las cuales tienen una relación directa, directa e inversa, respectivamente con el riesgo de no cumplir los niveles mínimos de calidad

del agua. El resto de las variables ejerce un poder discriminante poco significativo y, por dicha razón, no han sido seleccionadas.

**Tabla 5.2.1.1: Resultados del proceso de selección de variables**

		Variables en la ecuación					
		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 <sub>a</sub>	LCOND20	1,845	,054	1159,071	1	,000	6,330
	Constante	-12,049	,350	1188,145	1	,000	,000
Paso 2 <sub>b</sub>	LCOND20	1,693	,055	952,153	1	,000	5,433
	LMES	,313	,025	152,143	1	,000	1,368
	Constante	-11,786	,350	1131,437	1	,000	,000
Paso 3 <sub>c</sub>	LCOND20	1,658	,055	912,022	1	,000	5,248
	LMES	,305	,025	143,808	1	,000	1,356
	pH	-,895	,138	42,122	1	,000	,409
	Constante	-4,243	1,196	12,589	1	,000	,014

a. Variable(s) introducida(s) en el paso 1: LCOND20.

b. Variable(s) introducida(s) en el paso 2: LMES.

c. Variable(s) introducida(s) en el paso 3: pH.

Si analizamos los resultados expuestos en la Tabla 5.2.1.2 (calculados para un nivel de riesgo del 50%) se observa que, desde un punto de vista práctico, la principal potencia discriminante la ejerce la variable LCON20 (73,4% de aciertos totales) siendo la contribución de las otras dos variables más bien marginal al aumentar el porcentaje de aciertos en un 1,5%.

**Tabla 5.2.1.2: Resultados del proceso de validación intramuestral de los modelos estimados por el proceso de selección de variables**

**Tabla de clasificación<sup>a</sup>**

Observado			Pronosticado		
			calidad_conjunta		Porcentaje correcto
			Cumple	No cumple	
Paso 1	calidad_conjunta	Cumple	2520	677	78,8
		No cumple	834	1647	66,4
	Porcentaje global				73,4
Paso 2	calidad_conjunta	Cumple	2540	657	79,4
		No cumple	793	1688	68,0
	Porcentaje global				74,5
Paso 3	calidad_conjunta	Cumple	2539	658	79,4
		No cumple	770	1711	69,0
	Porcentaje global				74,9

a. El valor de corte es ,500

### 5.2.2. Umbrales multivariantes para los indicadores de acidificación

En este caso utilizamos como variables independientes del modelo LCO3CA y CO32CA cuyo significado puede verse en la Tabla 2.3. La variable dependiente es el indicador de calidad conjunta definido en la sección 3.1.4. El número de casos utilizados en el proceso de selección fue 742 que constituye únicamente un 11,2% de los datos de la muestra. Ello es debido al alto número de casos que tienen ausentes al menos uno de los dos indicadores por lo que los resultados que se exponen a continuación deben tomarse con precaución.

En la Tablas 5.2.2.1 se muestran los resultados obtenidos. Más concretamente, se presentan las variables finalmente escogidas junto con el orden en el que fueron seleccionadas.

**Tabla 5.2.2.1: Resultados del proceso de selección de variables**

**Variables en la ecuación**

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso	CO32CA	,013	,001	82,187	1	,000	1,013
1	Constante	-2,140	,276	60,148	1	,000	,118

a. Variable(s) introducida(s) en el paso 1: CO32CA.

Se observa que la variable finalmente seleccionada es el indicador CO32CA. Ello es debido al bajo poder discriminante mostrado por el indicador CO3CA, muy

probablemente, por el alto porcentaje de censura de la misma. El modelo estimado concuerda, en líneas generales, con el estimado en la sección 5.1.3.2 utilizando toda la información disponible y remitimos al lector a los comentarios hechos en dicha sección a la hora de analizar la validación intramuestral del mismo.

### **5.2.3. Umbrales multivariantes para los indicadores de salinidad**

En este caso utilizamos como variables independientes del modelo LNA, LCA, LMG, LCL y LSO4 cuyo significado puede verse en la Tabla 2.3. La variable dependiente es el indicador de calidad conjunta definido en la sección 3.1.4. El número de casos utilizados en el proceso de selección fue 1041 que constituye un 15,7% de los datos de la muestra. Ello es debido al alto número de casos que tienen ausentes al menos uno de los dos indicadores por lo que los resultados que se exponen a continuación deben tomarse con precaución.

En las Tablas 5.2.3.1 y 5.2.3.2 se muestran los resultados obtenidos. Más concretamente, en la Tabla 5.2.3.1 se presentan las variables finalmente escogidas junto con el orden en el que fueron seleccionadas mientras que en la Tabla 5.2.3.2 se muestran los resultados obtenidos en el proceso de validación intramuestral de cada uno de los modelos estimados por el proceso de selección.

Las variables seleccionadas fueron (por este orden) LNA, LMG, LCA, LCL y LSO4 el modelo escogido viene dado por la expresión:

$$P(\text{No cumplir}) = -7,587 + 0,914 * LCA + 1,16 * LCL - 1,316 * LMG + 0,585 * LSO4$$

no siendo elegida, finalmente, la variable LNA debido a que su efecto viene explicado a través de las variables seleccionadas, como se pone de manifiesto en los resultados mostrados. Observamos que, habiendo sido seleccionada en el primer paso del algoritmo, fue posteriormente eliminada al ser introducidas en el modelo el resto de las variables, puesto que deja de ser significativo su efecto.

**Tabla 5.2.3.1: Resultados del proceso de selección de variables**

**Variables en la ecuación**

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	LNA	1,085	,070	239,949	1	,000	2,958
	Constante	-3,642	,258	198,741	1	,000	,026
Paso 2	LMG	-,301	,124	5,873	1	,015	,740
	LNA	1,243	,098	162,130	1	,000	3,465
	Constante	-3,400	,275	153,207	1	,000	,033
Paso 3	LCA	1,455	,298	23,820	1	,000	4,286
	LMG	-,944	,183	26,520	1	,000	,389
	LNA	1,158	,101	131,231	1	,000	3,185
	Constante	-7,903	,985	64,356	1	,000	,000
Paso 4	LCA	1,375	,296	21,630	1	,000	3,956
	LCL	,609	,277	4,833	1	,028	1,838
	LMG	-,952	,182	27,293	1	,000	,386
	LNA	,616	,264	5,455	1	,020	1,852
	Constante	-7,964	,974	66,851	1	,000	,000
Paso 5	LCA	1,016	,346	8,607	1	,003	2,763
	LCL	,772	,291	7,018	1	,008	2,164
	LMG	-1,258	,240	27,502	1	,000	,284
	LNA	,407	,285	2,043	1	,153	1,502
	LSO4	,463	,231	4,018	1	,045	1,589
	Constante	-7,571	,994	57,959	1	,000	,001
Paso 6	LCA	,914	,337	7,371	1	,007	2,495
	LCL	1,160	,107	117,207	1	,000	3,190
	LMG	-1,316	,236	31,045	1	,000	,268
	LSO4	,585	,215	7,413	1	,006	1,794
	Constante	-7,587	,987	59,100	1	,000	,001

- a. Variable(s) introducida(s) en el paso 1: LNA.
- b. Variable(s) introducida(s) en el paso 2: LMG.
- c. Variable(s) introducida(s) en el paso 3: LCA.
- d. Variable(s) introducida(s) en el paso 4: LCL.
- e. Variable(s) introducida(s) en el paso 5: LSO4.

Los resultados obtenidos muestran la existencia de un alto nivel de colinealidad entre las variables que están fuertemente relacionadas entre sí tal y como lo muestra su matriz de correlaciones (ver Tabla 5.2.3.3). De hecho un análisis factorial reveló la existencia de un único factor relacionado directamente con todos los indicadores que explica el 84,98% de la variabilidad total.

Desde un punto de vista práctico y teniendo en cuenta tanto los resultados expuestos en la Tabla 5.2.3.2 (calculados para un nivel de riesgo del 50%) como los obtenidos para cada indicador por separado (secciones 5.1.4.1 a 5.1.4.5 y 5.1.6) se

observa que la principal potencia discriminante la ejerce la variable LCL siendo la contribución del resto de la variables más bien marginal.

**Tabla 5.2.3.2: Resultados del proceso de validación intramuestral de los modelos estimados por el proceso de selección de variables**

**Tabla de clasificación<sup>a</sup>**

Observado	Pronosticado				
	calidad conjunta		Porcentaje correcto		
	Cumple	No cumple			
Paso 1	calidad conjunta	Cumple	312	157	66,5
		No cumple	106	466	81,5
	Porcentaje global				74,7
Paso 2	calidad conjunta	Cumple	312	157	66,5
		No cumple	104	468	81,8
	Porcentaje global				74,9
Paso 3	calidad conjunta	Cumple	324	145	69,1
		No cumple	101	471	82,3
	Porcentaje global				76,4
Paso 4	calidad conjunta	Cumple	324	145	69,1
		No cumple	108	464	81,1
	Porcentaje global				75,7
Paso 5	calidad conjunta	Cumple	325	144	69,3
		No cumple	105	467	81,6
	Porcentaje global				76,1
Paso 6	calidad conjunta	Cumple	324	145	69,1
		No cumple	110	462	80,8
	Porcentaje global				75,5

a. El valor de corte es ,500

**Tabla 5.2.3.3: Matriz de correlaciones de los indicadores de salinidad**

**Matriz de correlaciones**

		LCA	LCL	LMG	LNA	LSO4
Correlación	LCA	1,000	,748	,859	,744	,900
	LCL	,748	1,000	,719	,977	,782
	LMG	,859	,719	1,000	,718	,891
	LNA	,744	,977	,718	1,000	,782
	LSO4	,900	,782	,891	,782	1,000
Sig. (Unilateral)	LCA		,000	,000	,000	,000
	LCL	,000		,000	,000	,000
	LMG	,000	,000		,000	,000
	LNA	,000	,000	,000		,000
	LSO4	,000	,000	,000	,000	

Se concluye, por lo tanto, que cualquiera de los indicadores de salinidad tienen una potencia discriminante significativa a la hora de predecir los niveles de calidad del agua, siendo la información que proporcionan muy redundante al estar fuertemente relacionados entre sí. Cualquiera de ellos puede ser seleccionado para llevar a cabo dicha labor de discriminación. Conviene hacer notar, finalmente, que estos resultados están basados en una muestra muy reducida (15,7% del total) debido al alto número de casos con datos ausentes en al menos uno de dichos indicadores, por lo que los resultados obtenidos deberían tomarse con precaución.

#### **5.2.4. Umbrales multivariantes para los indicadores de nutrientes**

En este caso utilizamos como variables independientes del modelo LNH<sub>4</sub>, LNO<sub>2</sub>, LNO<sub>3</sub>, LLN\_KJED, LPO<sub>4</sub> y LP\_TOT cuyo significado puede verse en la Tabla 2.3. La variable dependiente es el indicador de calidad conjunta definido en la sección 3.1.4. El número de casos utilizados en el proceso de selección fue 1031 que constituye un 15,5% de los datos de la muestra. Ello es debido al alto número de casos que tienen ausentes al menos uno de los dos indicadores por lo que los resultados que se exponen a continuación deben tomarse con precaución.

En las Tablas 5.2.4.1 y 5.2.4.2 se muestran los resultados obtenidos. Más concretamente, en la Tabla 5.2.4.1 se presentan las variables finalmente escogidas y el orden en el que fueron seleccionadas mientras que en la Tabla 5.2.4.2 se muestran los resultados obtenidos en el proceso de validación intramuestral de cada uno de los modelos estimados por el proceso de selección.

Las variables seleccionadas fueron (por este orden) LNO<sub>2</sub>, LNO<sub>3</sub> y LPO<sub>4</sub> y el modelo escogido viene dado por la expresión:

$$P(\text{No cumplir}) = 3,225 + 0,724 * LNO_2 + 0,547 * LNO_3 + 0,789 * LPO_4$$

no siendo seleccionadas, finalmente, el resto de las variables debido a su menor poder discriminante y a que su efecto viene explicado a través de las variables seleccionadas. Este hecho se pone de manifiesto si analizamos la matriz de correlaciones de las variables independientes (ver Tabla 5.2.4.3) en las que se aprecia que todas las correlaciones son significativas y moderadamente positivas. De hecho un Análisis Factorial puso de manifiesto la existencia de un factor común relacionado de forma directa con todas ellas explicando un 57,53% de la variación total.

**Tabla 5.2.3.1: Resultados del proceso de selección de variables**

**Variables en la ecuación**

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	LNO2	1,207	,084	208,803	1	,000	3,343
	Constante	4,333	,280	238,690	1	,000	76,179
Paso 2	LNO2	,943	,090	110,016	1	,000	2,566
	LNO3	,542	,089	37,444	1	,000	1,720
	Constante	2,426	,393	38,020	1	,000	11,313
Paso 3	LNO2	,724	,098	55,128	1	,000	2,063
	LNO3	,547	,090	36,629	1	,000	1,728
	LPO4	,789	,150	27,756	1	,000	2,201
	Constante	3,225	,447	52,079	1	,000	25,164

a. Variable(s) introducida(s) en el paso 1: LNO2.

b. Variable(s) introducida(s) en el paso 2: LNO3.

c. Variable(s) introducida(s) en el paso 3: LPO4.

Este hecho también se refleja en los resultados expuestos en la Tabla 5.2.4.2 (calculados para un nivel de riesgo del 50%) así como en los obtenidos en las secciones 5.1.5.1 a 5.1.5.6 y 5.1.6 en los que se observa que, desde un punto de vista práctico, la principal potencia discriminante la ejerce la variable LNO2 siendo la contribución del resto de las variables más bien marginal.

**Tabla 5.2.4.2: Resultados del proceso de validación intramuestral de los modelos estimados por el proceso de selección de variables**

**Tabla de clasificación<sup>a</sup>**

Observado			Pronosticado		
			calidad conjunta		Porcentaje correcto
			Cumple	No cumple	
Paso 1	calidad conjunta	Cumple	198	159	55,5
		No cumple	64	610	90,5
	Porcentaje global				78,4
Paso 2	calidad conjunta	Cumple	201	156	56,3
		No cumple	74	600	89,0
	Porcentaje global				77,7
Paso 3	calidad conjunta	Cumple	215	142	60,2
		No cumple	75	599	88,9
	Porcentaje global				79,0

a. El valor de corte es ,500

**Tabla 5.2.4.3: Matriz de correlaciones de los indicadores de nutrientes**

**Matriz de correlaciones**

		LNH4	LLNKJED	LNO2	LNO3	LPTOT	LPO4
Correlación	LNH4	1,000	,587	,581	,208	,582	,492
	LLNKJED	,587	1,000	,426	,094	,461	,411
	LNO2	,581	,426	1,000	,514	,592	,624
	LNO3	,208	,094	,514	1,000	,369	,383
	LPTOT	,582	,461	,592	,369	1,000	,809
	LPO4	,492	,411	,624	,383	,809	1,000
Sig. (Unilateral)	LNH4		,000	,000	,000	,000	,000
	LLNKJED	,000		,000	,000	,000	,000
	LNO2	,000	,000		,000	,000	,000
	LNO3	,000	,000	,000		,000	,000
	LPTOT	,000	,000	,000	,000		,000
	LPO4	,000	,000	,000	,000	,000	

Se concluye, por lo tanto, los indicadores de nutrientes tienen una potencia discriminante significativa a la hora de predecir los niveles de calidad del agua, siendo NO<sub>2</sub>, NO<sub>3</sub> y PO<sub>4</sub> los de mayor potencia. Sin embargo, la información que proporcionan estos 3 indicadores es bastante redundante al estar significativamente relacionados entre sí. Cualquiera de los 3 indicadores anteriores puede ser seleccionado, por tanto, para llevar a cabo dicha labor de discriminación. Conviene hacer notar, finalmente, que estos resultados están basados en una muestra muy reducida (15,5% del total) debido al alto número de casos con datos ausentes en al menos uno de dichos indicadores, por lo que los resultados obtenidos deberían tomarse con precaución.

## 6. CONCLUSIONES

En este trabajo se ha llevado a cabo una determinación estadística de los umbrales de calidad del agua en las diversas estaciones de los ríos de la cuenca del Ebro a partir de 21 parámetros físico-químicos que informan acerca de características generales del agua (temperatura, pH, conductividad), materia orgánica presente en la misma y sus niveles de acidificación, salinidad y nutrientes. El valor de dichos parámetros fue determinado mediante un proceso de muestreo llevado a cabo en las diferentes estaciones meteorológicas de muestreo localizadas a lo largo de la cuenca.

Los umbrales se han determinado mediante modelos de regresión logística-binaria en los que se ha tomado como variable dependiente un indicador de la calidad del agua construido a partir de los índices biológicos IBMWP, IPS e IVAM, habitualmente utilizados para medir dicha calidad, y como variables independientes cada uno de los indicadores físico-químicos considerados en el estudio. Así mismo, se ha realizado, para cada grupo de indicadores físico-químicos, un estudio multivariante del problema que analiza las interacciones existentes entre ellos, determinando cuáles son los que tienen más poder discriminante de la calidad y poniendo de manifiesto las redundancias existentes entre ellos.

A modo de resumen, en la Tabla 6.1. se muestran los umbrales obtenidos para los diferentes parámetros físico-químicos, según la metodología estadística utilizada en el trabajo (ver Sección 4) así como su comparación con los umbrales propuestos por la Confederación en el que caso de que los hubiera, y los posibles sesgos de la determinación realizada.

Los resultados obtenidos (ver Tabla 6.1) muestran que, para un riesgo empírico del 5% entre las muestras que cumplen, la **conductividad** a 20°C, los indicadores de salinidad (más concretamente, los **cloruros**); las concentraciones de **nitritos** y la **demanda química de oxígeno** son los parámetros físico-químicos con mayor fiabilidad (en torno a un 65%-70% de aciertos) a la hora de predecir la calidad del agua. Así mismo los carbonatos muestran una fiabilidad moderadamente alta (en torno al 60% de aciertos). Se aprecia, además, la existencia de redundancias en la información proporcionada por los indicadores de salinidad y, en menor medida por los indicadores de nutrientes, por lo que bastaría la utilización de un indicador de cada grupo para obtener una potencia discriminante similar a la del conjunto de todos ellos.

Conviene hacer notar, sin embargo, que, con la única excepción de los indicadores generales (temperaturas del agua y del aire, pH, conductividad a 20°C, sólidos en suspensión y nivel de oxígeno disuelto tanto en concentración como en % de saturación), existe un alto porcentaje de datos ausentes y, en algunos casos, además, censura, apreciándose una falta de representatividad en las muestras analizadas, por lo que los resultados anteriores deberían tomarse con precaución.

Sería interesante realizar un análisis multivariante del problema utilizando la información proporcionada por todos los indicadores considerados en el estudio, en lugar de hacerlo por grupos. Para ello sería necesario disponer de una muestra suficientemente amplia de puntos de muestreo que tengan observados simultáneamente el valor de los índices biológicos y de las variables fisico-químicas utilizadas en el análisis, algo que no ha sido posible en el presente estudio en el que para ningún caso de la muestra se disponía de dicha información. A partir de los datos observados se podrían utilizar herramientas más adaptadas al problema como, por ejemplo, árboles de decisión, que permiten la elaboración de esquemas secuenciales de decisión basados en la determinación de umbrales univariantes para cada indicador. Así mismo sería interesante replicar el estudio por tipologías de masas de agua definidas por el Cedex dado que los requisitos de calidad mínimos utilizados en este estudio no tienen por qué ser los mismos para cada una de estas tipologías.

Todo ello queda como un futuro objetivo que el equipo que ha elaborado este informe planea llevar a cabo en estudios posteriores del problema.

**Tabla 6.1: Resumen de los resultados obtenidos**

Tipo Indicador	Parámetro	Relación	Umbral (95% Riesgo empírico)	Fiabilidad	Umbral CHE	%Aciertos Cumplen	Fiabilidad	Sesgo Missing y censura
Generales	TEM_AG	Directa	22,1 °C	57,10%				No
	TEM_AI	Directa	29,5 °C	55,30%				No
	PH	Inversa	7,71	54,90%				No
	COND20	Directa	1180 µS/cm	65,40%				No
	O2	Inversa	7,3 mg/L O <sub>2</sub>	59,30%				No
	O2%	Inversa	78 mg/L O <sub>2</sub>	58,80%				No
	MES	Directa	62,35 mg/L	59,30%				No
Materia Orgánica	DQO	Directa	19 mg/L O <sub>2</sub>	64,90%	15	90,30%	67,10%	Sobre-representación no cumplen
Acidificación	CO3CA	Directa	18,075 mg/L CO <sub>3</sub> CA	62,13%				Sobre-representación no cumplen y censura significativa
	CO32CA	Directa	280,4 mg/L CO <sub>3</sub> CA	49,00%				Sobre-representación no cumplen
Salinidad	NA	Directa	126 mg/L Na	60,70%				Sobre-representación no cumplen
	CA	Directa	184,8 mg/L CA	49,80%				Sobre-representación no cumplen
	MG	Directa	47,985 mg/L MG	51,50%				Sobre-representación no cumplen
	CL	Directa	141,34 mg/L CL	69,80%				Sobre-representación cumplen
	SO4	Directa	360,4 mg/L SO <sub>4</sub>	57,70%				Sobre-representación no cumplen
Nutrientes	NH4	Directa	0,18 mg/L NH <sub>4</sub>	65,00%	0,4	97,90%	61,90%	Sobre-representación cumplen y censura no significativa
	NO2	Directa	0,095 mg/L NO <sub>2</sub>	68,20%	0,15	97,80%	62,40%	Sobre-representación no cumplen
	NO3	Directa	18,8 mg/L NO <sub>3</sub>	57,70%	20	95,80%	56,50%	Sobre-representación no cumplen
	N_KJED	Directa	1,1 mg/L N	48,70%				Sobre-representación no cumplen y censura significativa
	PO4	Directa	0,34 mg/L PO <sub>4</sub>	59,40%	0,3	94,10%	60,80%	Sobre-representación no cumplen
	P_TOT	Directa	0,162 mg/L P	60,10%				Sobre-representación no cumplen

