



Universitat Oberta
de Catalunya

Análisis multivariante de unos datos de ecotoxicología

Jenny Patricia Paredes Fierro

Máster Universitario en Bioinformática y Bioestadística UOC-UB
Área1_Estadística y Bioinformática

Director

Francesc Carmona Pontaque

Junio



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-
SinObraDerivada [3.0 España de Creative
Commons](#)

B) GNU Free Documentation License (GNU FDL)

Copyright © 2017 Jenny Patricia Paredes Fierro.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Análisis multivariante de unos datos de ecotoxicología</i>
Nombre del autor:	<i>Jenny Patricia Paredes Fierro</i>
Nombre del consultor/a:	Francesc Carmona Pontaque
Nombre del PRA:	
Fecha de entrega (mm/aaaa):	Mayo 2017
Titulación:	Máster Universitario en Bioinformática y Bioestadística UOC - UB
Área del Trabajo Final:	Área 1_Estadística y Bioinformática
Idioma del trabajo:	Castellano
Palabras clave	<i>Análisis Multivariante, ecotoxicología, PCA</i>

Resumen del Trabajo

En el presente trabajo se realiza el estudio de una zona poli contaminada del Camp de Tarragona expuesta a una polución procedente de cuatro incineradoras y a una adicional denominada zona de control que está climatológicamente aislada de todos los focos de polución, el estudio que se lleva a cabo consiste en un análisis multivariado de los datos de ecotoxicidad basadas en estudios previos realizados a muestras de agua, suelo, y sangre y tejido de algunos roedores. El análisis multivariante consistió en cinco métodos, análisis exploratorio, análisis de datos atípicos, análisis para reducir la dimensionalidad de la base de datos, análisis para establecer de manera diferencial el aporte relativo de cada foco de polución en cada una de las locaciones, análisis factorial, y finalmente un análisis de conglomerados, encontrándose como resultados el reconocimiento de la naturaleza de los datos, pudiéndose identificar datos atípicos, grupos homogéneos de variables de los datos, y finalmente la extracción de factores comunes de las variables de los datos de ecotoxicología. Al finalizar el análisis se ha identificado que la zona de control resulta ser la que presenta un mayor foco de polución de las cinco zonas, y también se ha identificado las variables específicas que aportan información toxicológica relevante en este estudio.

Abstract (in English, 250 words or less):

In this work, we present the study of a polycontaminated zone at Camp de Tarragona exposed to a pollution from four incinerators and an additional denominated control zone that is climatologically isolated from all the pollution focus, the study that is carried out consists of a multivariate analysis of ecotoxicity data based on previous studies done on samples of water, soil, and blood and tissue from some rodents. The multivariate analysis consisted of five methods, exploratory analysis, atypical data analysis, an analysis to reduce the dimensionality of the database, analysis to establish in a differential way the relative contribution of each source of pollution in each one of the locations, factor analysis, and finally an analysis of clusters, finding as results the recognition of the nature of the data, being able to identify atypical data, homogeneous groups of data variables, and finally the extraction of common factors from the variables of the ecotoxicology data. At the end of this analysis, we have been identified that the control zone present more pollution focus than the other four zones, and also we have identified the specific variables that provide relevant toxicological information in this study.

Índice

1. Introducción.....	5
1.1 Contexto y justificación del Trabajo.....	5
1.2 Objetivos del Trabajo.....	5
1.3 Enfoque y método seguido.....	6
1.4 Planificación del Trabajo	7
1.5 Breve sumario de productos obtenidos	9
1.6 Breve descripción de los otros capítulos de la memoria.....	9
2. Análisis de los datos.....	11
2.1 Análisis descriptivo	11
2.1.1 Estadística descriptiva.....	11
2.1. 2 Análisis Gráfico.....	22
2.2 Variables redundantes.....	27
2.3 Datos atípicos.....	30
2.3.1 Análisis gráfico para visualizar los datos extremos.....	30
2.3.2 Método para la detección de valores atípicos multivariantes.....	32
2.4 Análisis de componentes principales.....	35
2.5 Análisis factorial	50
2.6 Análisis de conglomerado	57
3. Conclusiones.....	67
4. Glosario	70
5. Bibliografía	71
6. Anexos	72
6.1 ANEXO 1.....	72
6.1.1 RSTUDIO.....	72

Lista de figuras

Figura 1: Planificación temporal de las tareas a realizar	9
Figura 2: Medias de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.....	12
Figura 3: Medias de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.	13
Figura 4: Medias de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.	13
Figura 5: Medias de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.	13
Figura 6: Medianas de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.	15
Figura 7: Medianas de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.	16
Figura 8: Medianas de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.	16
Figura 9: Medianas de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.	16
Figura 10: Desviaciones estándar de las variables del estudio en muestras de agua y suelo respecto a las locaciones.	18
Figura 11: Desviaciones estándar de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.....	18
Figura 12: Desviaciones estándar de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.	18
Figura 13: Desviaciones estándar de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.	19
Figura 14: Varianza de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.	20
Figura 15: Varianza de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones	21
Figura 16: Varianza de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.	21
Figura 17: Varianza de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.	21
Figura 18: Gráficas de dispersión de las medias de las variables del estudio de muestras de agua y suelo en cada una de las locaciones.	22
Figura 19: Gráficas de dispersión de las medias de las variables obtenidas de los roedores en cada una de las locaciones.....	22
Figura 20: Gráficas de dispersión de las medias de las variables cuyos datos son los resultados de los exámenes hematológicos de los roedores en cada una de las locaciones.....	23
Figura 21: Gráficas de dispersión de las medias de las variables cuyos datos son los resultados de los exámenes hematológicos de los roedores en cada una de las locaciones.....	23
Figura 22: Correlación de las variables sobre las medidas obtenidas de los roedores.	25
Figura 23: Correlación de las variables sobre los exámenes hematológicos de los roedores.....	26

Figura 24: Correlación de las variables sobre los exámenes hispatológicos de los roedores.....	27
Figura 25: Autovalores y autovectores correspondiente a las matriz S cuyos datos son las medidas obtenidas de los roedores.....	28
Figura 26: Autovalores y autovectores correspondientes a las matriz S cuyos datos son los resultados de los exámenes hematológicos.....	29
Figura 27: Autovalores y autovectores correspondiente a las matriz S cuyos datos son los resultados de los exámenes hispatológicos	29
Figura 28: Diagrama de cajas de las variables del estudio de las muestras del suelo y agua.....	30
Figura 29: Diagrama de cajas de las variables que son las medidas obtenidos de los roedores.....	31
Figura 30: Diagrama de cajas de las variables resultado de los exámenes hematológicos obtenidos de los roedores respecto a las locaciones.	32
Figura 31: Diagrama de cajas de las variables resultado de los exámenes hispatológicos de los roedores.....	32
Figura 32: Datos atípicos de las medidas obtenidas de los roedores	35
Figura 33: Gráfica de los outlier multivariados de las medidas obtenidas de los roedores.	35
Figura 34: Componentes principales de los datos del estudio de agua y suelo.	36
Figura 35: Sedimentación de los datos del estudio de las muestras del suelo y agua	36
Figura 36: Matriz de correlaciones estimada a partir de la solución factorial de los datos del estudio de las muestras de agua y suelo.	39
Figura 37: Componentes principales en dos dimensiones de los datos del estudio de las muestras de agua y suelo.	40
Figura 38: Componentes principales de las medidas obtenidas de los roedores	40
Figura 39: Sedimentación de las medidas obtenidas de los roedores	41
Figura 40: Matriz de correlaciones estimada a partir de la solución factorial de las medidas de los roedores.....	43
Figura 41: Componentes principales en dos dimensiones de las medidas de los roedores	43
Figura 42: Componentes principales de los exámenes hematológicos de los roedores.	44
Figura 43: Sedimentación de los exámenes hematológicos	44
Figura 44: Matriz de correlaciones estimada a partir de la solución factorial de los exámenes hematológicos	46
Figura 45: Componentes principales en dos dimensiones de los exámenes hematológicos	47
Figura 46: Componentes principales de los exámenes hispatológicos	47
Figura 47: Sedimentación de los exámenes hispatológicos	48
Figura 48: Matriz de correlaciones estimada a partir de la solución factorial de los exámenes hispatológicos.....	49
Figura 49: Componentes principales en dos dimensiones de los exámenes hispatológicos.....	50
Figura 50: Gráfico lineal del número de factores en el Análisis Factorial de las medidas de los roedores	51

Figura 51: Modelo factorial con tres factores sobre las medidas obtenidas de los roedores.....	52
Figura 52: Vectores y factores de los datos sobre las medidas obtenidas de los roedores	52
Figura 53: Representación de variables en factores para Análisis Factorial de los datos sobre las medidas de los roedores	53
Figura 54: Gráfico lineal del número de factores en el Análisis Factorial sobre as medidas de los roedores	53
Figura 55: Modelo factorial con cinco factores de las medidas de los roedores	54
Figura 56: Vectores y factores de los datos sobre los exámenes hematológicos de los roedores.....	55
Figura 57: Representación de variables en factores para Análisis Factorial de los exámenes hematológicos	55
Figura 58: Representación de variables en factores para Análisis Factorial de los exámenes hispatológicos	56
Figura 59: Modelo factorial con un factor de los exámenes hispatológicos.....	57
Figura 60: Dendograma por observaciones del estudio en muestras de agua y suelo.....	58
Figura 61: Observaciones por cluster de los datos del estudio de las muestras de agua y suelo	58
Figura 62: Formación de los cluster sobre los datos del examen de las muestras de agua y suelo	59
Figura 63: Dendograma por variables de los datos del estudio de las muestras de agua y suelo	59
Figura 64: Dendograma por observaciones de los datos sobre las medidas de los roedores.....	61
Figura 65: Observaciones por cluster de los datos sobre las medidas de los roedores	61
Figura 66: Formación de los cluster sobre los datos de las medidas de los roedores	61
Figura 67: Dendograma por variables de los datos sobre las medidas de los roedores	62
Figura 68: Dendograma por observaciones de los exámenes hematológicos ..	62
Figura 69: Observaciones por cluster de los datos de los exámenes hematológicos	63
Figura 70: Formación de los cluster sobre los datos de los exámenes hematológicos	63
Figura 71: Dendodrama por variables de los examenes hematólogicos	63
Figura 72: Dendograma por observaciones de los exámenes hispatológicos..	65
Figura 73: Observaciones por cluster de los datos de los exámenes hispatológicos.....	65
Figura 74: Formación de los cluster sobre los datos de los exámenes hispatológicos.....	65
Figura 75: Dendograma por variables de los exámenes hispatológicos	66

1. Introducción

1.1 Contexto y justificación del Trabajo

EL presente trabajo parte de un estudio de ecotoxicidad en una zona poli contaminada del Camp de Tarragona expuesta a la polución procedente de la Incineradora de Residuos Peligrosos de Constantí, de una incineradora municipal de residuos de tipo urbano, y del complejo petroquímico adyacente, así como del tráfico rodado de diversas carreteras y autovías.

Se ha planteado un estudio complejo, con cinco áreas de muestreo situadas cuatro de ellas en diferentes posiciones respecto de los focos de contaminación, en función de los vientos dominantes y del sentido general de la escorrentía, y una quinta geográfica y climatológicamente aislada de estos focos, que se ha utilizado como control. En cada una de estas zonas se han realizado estudios en muestras de agua y de suelo, así como también en muestras de sangre y tejidos de roedores capturados sobre el terreno, representando todo ello un número elevado de end-points diferentes.

En este contexto, es importante realizar un estudio ecotoxicidad en la zona afectada debido a que se trata de la aplicación de técnicas estadísticas a la solución de problemas reales como la ecotoxicología; siendo la ecotoxicología un área que durante mucho tiempo ha sido desconocida y que ha hecho su reaparición después de ser sujeto constante de preocupación sobre todo en los países industrializados. La finalidad del análisis multivariante es analizar simultáneamente conjuntos de datos multivariantes y así entender las relaciones entre las muchas variables, pudiéndose como resultado del trabajo distinguir entre los efectos observados en función de las distintas localizaciones, para así poder establecer de manera diferencial el aporte relativo de cada foco de polución en cada una de las localizaciones, además de establecer en qué proporción cada uno de ellos contribuye a aportar información toxicológica relevante.

1.2 Objetivos del Trabajo

Objetivos generales:

1. Distinguir entre los efectos observados en función de las distintas localizaciones, de manera que, aprovechando el uso de especies sésiles o de área de campeo restringida, podamos establecer de manera diferencial el aporte relativo de cada foco de polución en cada una de las localizaciones.

2. Establecer en qué proporción cada uno de los parámetros medidos contribuye a aportar información toxicológica relevante, o hasta qué punto algunos de ellos resultan redundantes o innecesarios, o en qué casos un determinado parámetro sencillo de obtener puede substituir (o predecir) la información dada por otros varios parámetros más complejos.

Objetivos específicos:

1. Realizar un análisis descriptivo de los datos de ecotoxicología
2. Establecer diferencias sobre los focos de polución en cada localización
3. Establecer que parámetros proporciona la mayor información toxicológica.
4. Encontrar grupos homogéneos de variables de la base de datos de ecotoxicología
5. Extraer factores comunes de las variables de la base de datos de ecotoxicología.
6. Sintetizar los resultados de manera apropiada

1.3 Enfoque y método seguido

Los principales objetivos de este trabajo es poder distinguir entre los efectos observados en función de las distintas localizaciones y poder establecer de manera diferencial el aporte relativo de cada foco de polución en cada una de las localizaciones, además de establecer en qué proporción cada uno de ellos contribuye a aportar información toxicológica relevante. Para llevar a cabo estos objetivos se debe realizar un análisis multivariante mediante los siguientes procesos:

- 1) Reconocimiento sobre la naturaleza multivariada de los datos ecotoxicológicos.

Realizar un análisis exploratorio de los datos que se basa en gráficos y estadísticos descriptivos más utilizados para poder describir y resumir las observaciones obtenidas.

- 2) Identificación de datos atípicos

Los métodos de detección de valores atípicos se pueden dividir en univariados y multivariados. Para el caso univariado se los puede identificar mediante una gráfica de cajas de cada variable o según el autor Daniel Peña [1] realizar una estandarización robusta de los datos , pero para el caso multivariado es mucho más complicado realizar una exploración para llegar a encontrar estos valores, debido al problema de la dimensión, por lo tanto se hace necesario conocer cuáles son los métodos

existentes que permiten detectar este tipo de outliers, de acuerdo al autor Iván Amón Uribe[2] hay 4 métodos utilizados, se aplicará en método de distancia de mahalanobis.

3) Análisis de componentes principales

Este análisis permitirá establecer que parámetros proporciona la mayor información toxicológica y poder establecer las diferencias de los focos de polución en cada locación.

4) Análisis Factorial

Aplicando este método se podrá encontrar grupos homogéneos de variables a partir del conjunto numeroso de variable

5) Análisis de Conglomerado

A partir del estudio se establecería factores comunes tanto de variables como de observaciones.

1.4 Planificación del Trabajo

Para la realización de este trabajo se ha planteado una serie de actividades basadas en métodos estadísticos que se describen en cinco actividades específicas.

Actividad 1: Análisis descriptivo de los datos

1. Estadística descriptiva

- Cálculo de la media, media, desviación estándar y varianza de cada variable respecto a las locaciones.

2. Análisis gráfico de los datos

- Diagrama de dispersión de cada variable respecto a las locaciones.

Actividad 2: Diferencia de los focos de polución en cada locación con PAC

1. Componentes principales

- Cálculo de las primeras componentes
- Gráfica de dos dimensiones de la componente PC1 y PC2

Actividad 3: Establecer que parámetros proporciona la mayor información toxicológica con PAC

1. Componentes principales

- Cálculo de las primeras componentes
- Gráfica de dos dimensiones de la componente PC1 y PC2

Actividad 4: Encontrar grupos homogéneos de variables de la base de datos de ecotoxicología

1. Análisis Factorial

- Obtener el número de factores para el análisis factorial
- Ejecutar el modelo adecuado dependiendo de los factores
- Visualizar la dependencia de cada variable de los dos factores más significativos.

Actividad 5: Extraer factores comunes de las variables de la base de datos de ecotoxicología

1. Análisis de Conglomerado

- Gráfica de los conglomerados por observaciones
- Gráfica de los conglomerados por variables
- Obtención de los datos para cada conglomerado
- Obtención de las variables en cada conglomerado

Actividad 6: Sintetizar los resultados de manera apropiada

1. De todos los resultados obtenidos escoger la relevancia que contribuye en el desarrollo del TFM

Fase Inicial		19 days	Wed 22/02/17	Mon 20/03/17
	Elección de Trabajo Final	5 days	Wed 22/02/17	Tue 28/02/17
	Plan de trabajo	11 days	Wed 01/03/17	Wed 15/03/17
Fase 1		15 days	Thu 16/03/17	Wed 05/04/17
	Analisis descriptivo de los datos	2 days	Thu 16/03/17	Fri 17/03/17
	Diferencia de los focos de polución en cada locación con PAC	7 days	Sun 19/03/17	Mon 27/03/17
	Establecer que parámetros proporciona la mayor información toxicológica con PAC	7 days	Tue 28/03/17	Wed 05/04/17
Fase 2		25 days	Thu 06/04/17	Wed 10/05/17
	Encontrar grupos homogéneos de variables de la base de datos de ecotoxicología	8 days	Thu 06/04/17	Mon 17/04/17
	Establecer factores comunes de las variables de las bases de datos	8 days	Tue 18/04/17	Thu 27/04/17
	Sintetizar los resultados de manera apropiada	9 days	Fri 28/04/17	Wed 10/05/17
Fase Final		30 days	Thu 11/05/17	Wed 21/06/17
	Memoria y presentación del trabajo final	10 days	Thu 11/05/17	Wed 24/05/17
	Preparación de la defensa	9 days	Thu 25/05/17	Tue 06/06/17
	Defensa Pública	11 days	Wed 07/06/17	Wed 21/06/17

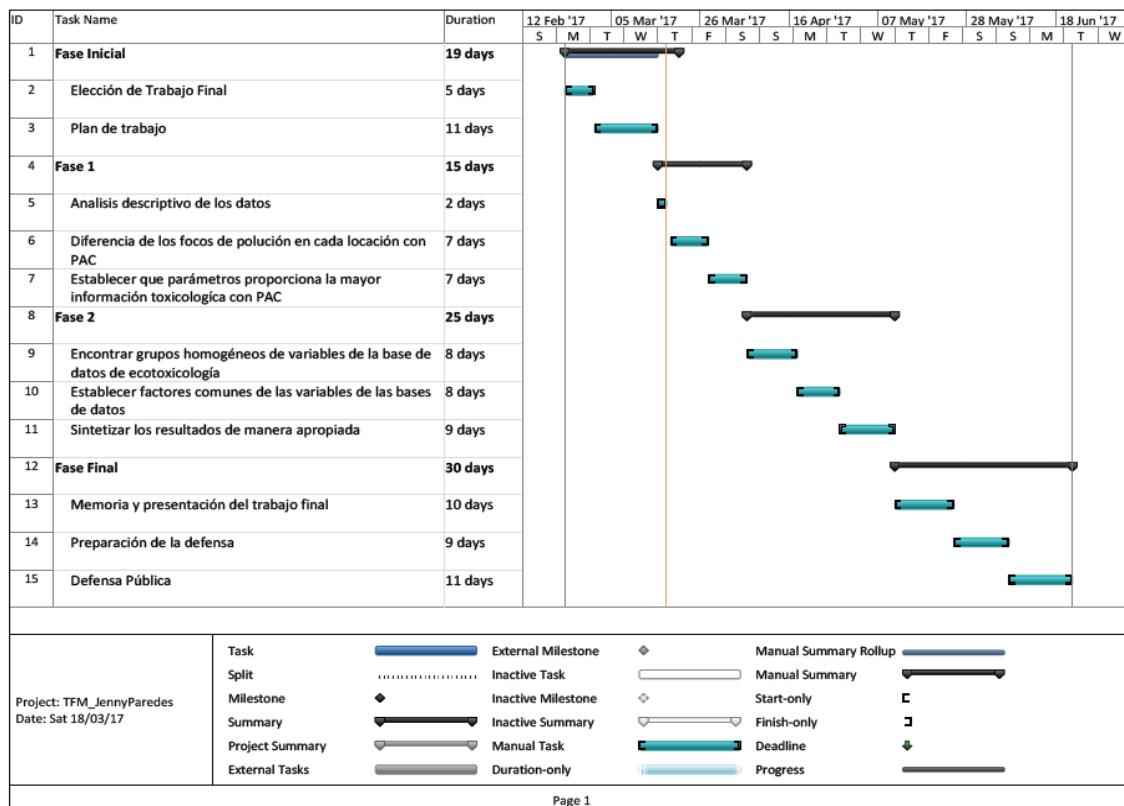


Figura 1: Planificación temporal de las tareas a realizar

1.5 Breve resumen de productos obtenidos

Los test aplicados a la tierra y agua tienen una contaminación media mayor en las zonas 4 (Zona Control) y 2 (Complejo Petroquímico Adyacente), en cambio en las muestras de sangre y tejidos de roedores capturados sobre el terreno nuestra una media mayor en los análisis respecto a las zonas 2(Complejo Petroquímico Adyacente) y la 0 (Incineradora de residuos peligrosos de Constantí).

En el análisis sobre las muestras de sangre de los animales las variables que tienen mayor aporte toxicológico son HCG, HCT, RBC, WBC, LYM y en el análisis sobre las medidas obtenida de los roedores las variables con mayor aporte toxicológico son BW, BCI, Taillenght, Handlenght y Footlenght.

1.6 Breve descripción de los otros capítulos de la memoria

En el apartado 2 se describe cada uno de los métodos y su aplicación en R. Los métodos descritos son:

- Análisis descriptivo de los datos
- Variables redundantes

- Datos Atípicos
- Análisis de Componentes Principales
- Análisis Factorial
- Análisis de Conglomerados

En los anexos se incluye la documentación necesaria utilizada para el análisis de los datos y los scripts de las con las funciones en R.

2. Análisis de los datos

2.1 Análisis descriptivo

2.1.1 Estadística descriptiva

Media: La media de la muestra se puede calcular a partir de las n mediciones en cada una de las p variables, de modo que, en general, habrá p medias significativas [3]

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad k = 1, 2, \dots, p$$

- Datos de los estudios en muestras de agua y suelo

```
##   AdultEnchytriaed JuvenilsEnchytriaed AdultPlatynotres AdultOppiatwelve
## 1              66          56.4           46            92
## 2              96         614.8           48            94
## 3              90         722.2           78            98
## 4              80         280.6           32            94
## 5              94        1076.2           84           100
##   AdultOppiattwentyeighth JuvenilsOppiatwentytwo Adultfolsomia
## 1                  72             1.8            5.0
## 2                  32             2.8            4.4
## 3                  90            12.6            5.8
## 4                  78            13.8            5.0
## 5                  84            12.2            8.0
##   Juvenilsfolsom Avoid.respectcontrol Nitrateveintiocho Seedfourteen
## 1      186.2                 14       2.653333  1.00
## 2      128.4                 26      1.030000  4.25
## 3      256.4                -40      1.573333  1.75
## 4      121.4                 46      1.916667  2.25
## 5      159.2                 8      3.116667  3.75
##   Freshplantfourteen Highplant clorosisplant willingplant
## 1      0.2900    86.6500          0          0
## 2      0.0925   71.0650          55         85
## 3      0.1125   66.5100          0          0
## 4      0.1675   77.7575          15         40
## 5      0.2325   84.2225          10          0
##   Malformationplant DaphniainmobilisationhuNAred
## 1              20            8.333333
## 2              85           21.666667
## 3              35           13.333333
## 4              40           15.000000
## 5              65           0.000000
##   Daphniainmobilisationfifty Daphniainmobilisationtwentyfive
## 1            21.666667            3.333333
## 2            21.666667           16.666667
## 3            13.333333           31.666667
## 4            15.000000           25.000000
## 5            3.333333           0.000000
##   SurvivalfishhuNAred Survivalfishfifty Survivalfishtwentyfive
## 1      80.00000          93.33333        90.00000
## 2     13.33333          96.66667        96.66667
## 3     26.66667          86.66667        100.00000
```

```

## 4          15.00000    90.00000   100.00000
## 5          100.00000   100.00000  100.00000
##  AbnormalityfishhuNAred Abnormalityfishfifty Abnormalityfish tweentyfive
## 1          27.40667    10.83333   17.99
## 2          96.33333    17.03667   10.74
## 3          100.00000    18.52000   10.00
## 4          100.00000    78.05667   30.00
## 5          0.00000     0.00000    0.00
##  HachinghuNAred Hachingfifty Hatchingtweentyfive AlgaehuNAred Algaefifty
## 1          53.703333   15.00000   49.47333  963333.3  2526666.7
## 2          12.500000   20.74000   24.44333  940000.0  940000.0
## 3          3.333333   15.50333   33.33333  846666.7  846666.7
## 4          6.666667   96.29667   80.00000  923333.3  923333.3
## 5          30.000000   20.00000   16.66667  2900000.0 3200000.0
##  Algaetwentyfive yielhuNAred Seed.germinationhuNAred seedgerminationfifty
## 1          3000000    840000.0    74          82
## 2          2713333    920000.0    76          82
## 3          3023333    763333.3    72          88
## 4          3340000    913333.3    74          80
## 5          3300000    2800000.0   86          86
##  seedgerminationtwentyfive rootlengthhuNAred rootlengthfifty
## 1          84          80.54      109.78
## 2          80          51.34      122.88
## 3          88          75.62      53.74
## 4          84          53.94      95.06
## 5          96          100.08     102.52
##  rootlengtweentyfive INAexgerminationhuNAred INAexgerminationfifty
## 1          109.92      27.16      40.76
## 2          110.22      17.86      45.82
## 3          94.50       24.72      21.64
## 4          98.02       17.74      34.54
## 5          104.40      39.18      39.96
##  INAexgermianationseventyfive
## 1          41.88
## 2          40.68
## 3          38.04
## 4          37.14
## 5          45.60

```

Figura 2: Medias de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.

En el análisis estadístico descriptivo toxicológico se concluyó que en cada uno de los test aplicados a la tierra y agua tiene una mayor polución media en las zonas 4 (Zona Control) y 2 (Complejo Petroquímico Adyacente) y la zona 0(Incineradora de Residuos Peligrosos de Constantí) es el que tiene menor polución respecto al resto de locaciones.

- Datos de las medidas tomadas a los roedores

```

##      BW      BL      BCI      BCIApo     BCIMus     Liver     Kidney
## 0 16.11500 81.62500 1.012500 1.016667 0.9975000 5.300000 0.6760000
## 1 15.97200 82.17500 1.022000 1.007143 1.0333333 5.427500 0.8185000
## 2 14.96308 75.57692 1.046154 1.054545 1.0220000 5.642692 0.9738462
## 3 11.45075 76.22500 0.954750 0.925000 0.9761111 6.195750 0.8592500
## 4 21.82296 78.64815 1.072222 1.080000 1.0913636 5.315556 0.6718519

```

```

##      Spleen Taillenght Handlenght Footlenght
## 0 0.2475000 98.87400 27.00800 24.29000
## 1 0.3470000 92.73950 23.98700 24.95050
## 2 0.3207692 88.67880 27.35000 25.02731
## 3 0.3733333 77.32025 23.79825 22.14425
## 4 0.3444444 77.34630 22.45148 21.33222

```

Figura 3: Medias de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.

En el análisis estadístico descriptivo sobre las medidas recolectadas de los roedores se concluyó que en cada una de las medidas tiene una contaminación media mayor en las zonas 1 (Incineradora de Residuos Peligrosos de Constantí) y 2 (Complejo Petroquímico Adyacente).

- Datos de los exámenes hematológicos de los roedores

```

##      WBC      LYM      MID      GRA      RBC      HGB      HCT
## 0 4.235789 2.575789 0.2436842 1.418421 10.368421 164.3611 43.00737
## 1 4.710000 3.096250 0.1918750 1.420000 9.723125 156.2212 42.30687
## 2 4.568696 3.072174 0.2256522 1.272174 10.046957 160.5857 42.32304
## 3 5.888485 4.012121 0.2481818 1.627273 10.100606 149.5582 42.00333
## 4 5.473333 3.864762 0.2547619 1.354286 9.774762 148.5476 40.36429
##      MCV      MCH      MCHC Neutrophilis Lymphocytes Eosinophils
## 0 41.61474 15.91947 382.4763 14.000000 68.17647 2.823529
## 1 43.70000 16.10438 368.7331 9.315789 77.31579 2.631579
## 2 42.32217 16.03217 379.6391 11.750000 73.31250 3.812500
## 3 41.64879 14.81182 355.9285 13.081081 69.00000 3.000000
## 4 41.50952 15.20286 366.6767 8.740741 76.07407 2.296296
##      Monocytes Basophils Plaletes
## 0 13.823529 0.29411765 0.8823529
## 1 9.947368 0.05263158 0.7368421
## 2 10.125000 0.06250000 0.9375000
## 3 11.594595 0.00000000 0.7027027
## 4 11.666667 0.03703704 1.1851852

```

Figura 4: Medias de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.

En las muestras de sangre de los roedores capturados sobre el terreno se tiene una polución media mayor en las zonas 2(Complejo Petroquímico Adyacente) y la zona 0 (Incineradora de residuos peligrosos de Constantí).

- Datos de los exámenes hispatológicos de los roedores

```

##      Liverh Kidneyh Spleenh Lungh
## 0 125.00000 40.62500 29.41176 7.142857
## 1 20.00000 32.14286 61.53846 38.461538
## 2 28.00000 14.00000 144.11765 20.833333
## 3 66.666667 48.33333 40.00000 114.062500
## 4 3.846154 26.00000 37.50000 17.391304

```

Figura 5: Medias de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.

En las muestras de sangre de los roedores capturados sobre el terreno se tiene una polución media mayor en las zonas 3 (Tráfico rodado de diversas carreteras y autovías) y la zona 2 (Complejo Petroquímico Adyacente).

Mediana: Es el valor que divide al conjunto ordenado de datos, en dos subconjuntos con la misma cantidad de elementos. La mitad de los datos son menores que la mediana y la otra mitad son mayores [4]

- Ordena los números de menor a mayor
- **Numero par de datos:** La mediana es el dato que está en la posición del promedio entre los datos centrales
 $n/2 + (n+1)/2$.[4]
- **Número impar de datos:** La mediana es el dato que está en la posición $(n + 1)/2$.[4]
- **Datos de los estudios en muestras de agua y suelo**

```
##   AdultEnchytriaed JuvenilsEnchytriaed AdultPlatynotres AdultOppiattwelve
## 1              60                  52                 50                  90
## 2             100                 646                 40                 100
## 3              90                 828                 80                 100
## 4              80                 270                 40                  90
## 5              90                1109                 80                 100
##   AdultOppiattwentyeighth JuvenilsOppiatwenttwenty Adultfolsomia
## 1                  80                  1                   5
## 2                  20                  2                   4
## 3                  90                  12                  6
## 4                  80                  15                  4
## 5                  80                  10                  8
##   Juvenilsfolsom Avoid.respectcontrol Nitrateveintiocho Seedfourteen
## 1                 186                  20                 2.78                 1.0
## 2                 139                  40                 1.09                 4.5
## 3                 213                 -30                 1.53                 2.0
## 4                 127                  70                 1.91                 2.0
## 5                 152                  20                 3.13                 4.0
##   Freshplantfourteen Highplant clorosisplant willingplant
## 1                 0.20                86.890                  0                  0
## 2                 0.09                71.810                  50                 90
## 3                 0.11                66.975                  0                  0
## 4                 0.18                76.235                  20                  40
## 5                 0.25                83.685                  10                  0
##   Malformationplant DaphniainmobilisationhuNAred
## 1                  20                  10
## 2                  90                  20
## 3                  40                  15
## 4                  40                  15
## 5                  70                  0
##   Daphniainmobilisationfifty Daphniainmobilisationtwentyfive
## 1                  20                  0
## 2                  20                 15
## 3                  15                 30
```

```

## 4          15          25
## 5          5           0
##   SurvivalfishhuNAred Survivalfishfifty Survivalfishtwentyfive
## 1          90          100          100
## 2          15          100          100
## 3          30          90           100
## 4          15          90           100
## 5         100          100          100
##   AbnormalityfishhuNAred Abnormalityfishfifty Abnormalityfishtwentyfive
## 1          22.22        12.50        11.11
## 2          100.00       11.11        10.00
## 3          100.00       0.00         10.00
## 4          100.00       80.00        20.00
## 5          0.00         0.00         0.00
##   HatchinghuNAred Hachingfifty Hatchingtweentyfive AlgaehuNAred Algaefifty
## 1          11.11        20.00        50          990000      2600000
## 2          0.00         20.00        20          910000      910000
## 3          0.00         14.29        30          860000      860000
## 4          10.00        100.00       100         940000      940000
## 5          30.00        20.00        20          2900000     3200000
##   Algaetwentyfive yielhuNAred Seed.germinationhuNAred seedgerminationfifty
## 1          3000000     840000       80           80
## 2          2720000     920000       80           80
## 3          3020000     780000       80           90
## 4          3220000     930000       80           80
## 5          3300000     2810000      80           90
##   seedgerminationtwentyfive rootlengthhuNAred rootlenghfty
## 1          80           79.7        111.5
## 2          80           45.8        126.8
## 3          90           76.3        52.0
## 4          80           57.5        98.4
## 5          100          99.7        103.5
##   rootlengtweenyfive INAexgerminationhuNAred INAexgerminationfifty
## 1          116.2        27.4        45.7
## 2          113.6        16.1        45.0
## 3          99.1         27.4        21.3
## 4          106.5        17.3        33.5
## 5          106.8        38.8        41.9
##   INAexgermianationseventyfive
## 1          42.4
## 2          43.1
## 3          38.9
## 4          35.4
## 5          43.7

```

Figura 6: Medianas de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.

- Datos de las medidas obtenidas de los roedores

```

##      BW    BL    BCI BCIApo BCIMus Liver Kidney Spleen Taillenght Handlenght
## 0 14.54 81.5 1.015  1.00  1.010 5.500  0.695  0.215  101.930  26.395
## 1 15.76 78.0 1.005  1.00  1.035 5.635  0.730  0.225  94.905  25.000
## 2 13.07 74.0 1.040  1.00  1.030 5.520  0.770  0.280  89.230  26.915
## 3 11.38 76.0 0.950  0.95  0.970 6.010  0.860  0.350  79.330  24.245
## 4 13.34 78.0 1.020  1.00  1.045 5.490  0.730  0.350  75.640  23.900
##   Footlenght
## 0          23.485

```

```

## 1      23.910
## 2      23.910
## 3      21.170
## 4      21.110

```

Figura 7: Medianas de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.

- Datos de los exámenes hematológicos de los roedores

```

##      WBC    LYM    MID    GRA    RBC    HGB    HCT    MCV    MCH    MCHC
## 0  3.890  2.69  0.180  0.97 10.270 164.510 43.460 42.480 16.150 383.14
## 1  4.175  2.64  0.195  1.05  9.545 153.205 42.605 43.445 16.225 370.59
## 2  4.750  3.15  0.160  1.18 10.430 162.940 42.120 42.310 15.850 381.05
## 3  5.670  3.73  0.210  1.41  9.960 152.130 42.160 41.820 14.860 358.85
## 4  5.210  3.70  0.210  1.22  9.990 151.520 41.340 42.250 14.940 368.31
##   Neutrophilis Lymphocytes Eosinophils Monocytes Basophils Plaletes
## 0          12      70.0        2       11       0       0
## 1          7      79.0        1       9       0       0
## 2         11      73.5        2       9       0       1
## 3         11      73.0        2      11       0       0
## 4          7      76.0        2       9       0       1

```

Figura 8: Medianas de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.

- Datos de los exámenes hispatológicos de los roedores

```

##      Liverh Kidneyh Spleenh Lungh
## 0      100      0      0      0
## 1       0      0      0      0
## 2       0      0     100      0
## 3      50      0      0      0
## 4       0      0      0      0

```

Figura 9: Medianas de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.

Desviación Estándar: La raíz cuadra de la varianza

$$S_k = \sqrt{\frac{\sum_{j=1}^n (x_{jk} - \bar{x}_k)^2}{n}} \quad k = 1, 2, \dots, p$$

- Datos de los estudios en muestras de agua y suelo

```

##      AdultEnchytriaed JuvenilsEnchytriaed AdultPlatynotres AdultOppiattwelve
## 1      13.416408           21.43129      5.477226      8.366600
## 2      5.477226           173.28358     16.431677      8.944272
## 3      7.071068           210.41554      8.366600      4.472136
## 4      0.000000           68.71172     10.954451      5.477226
## 5      5.477226           129.25827      5.477226      0.000000
##      AdultOppiattwentyeighth JuvenilsOppiatwentyninety Adultfolsomia
## 1      10.954451           2.167948     0.7071068
## 2      21.679483           1.303840     0.5477226
## 3      0.000000           2.190890     1.3038405

```

## 4	10.954451	3.834058	1.4142136
## 5	5.477226	4.438468	0.7071068
## Juvenilsfolsom	Avoid.respectcontrol	Nitrateveintiocho	Seedfourteen
## 1	9.602083	63.08724	0.40513372
## 2	23.093289	39.74921	0.13076697
## 3	70.404545	24.49490	0.14011900
## 4	54.436201	45.05552	0.02081666
## 5	25.073891	50.69517	0.34019602
## Freshplantfourteen	Highplant	clorosisplant	willingplant
## 1	0.15588457	1.553963	0.00000
## 2	0.00500000	2.859528	34.15650
## 3	0.01500000	2.828510	0.00000
## 4	0.03201562	4.799468	10.00000
## 5	0.03500000	8.326613	11.54701
## Malformationplant	DaphniainmobilisationhuNAre		
## 1	0.00000		7.637626
## 2	19.14854		2.886751
## 3	10.00000		2.886751
## 4	0.00000		5.000000
## 5	19.14854		0.00000
## Daphniainmobilisationfifty	Daphniainmobilisationtwentfive		
## 1	2.886751		5.773503
## 2	2.886751		2.886751
## 3	2.886751		2.886751
## 4	5.000000		10.000000
## 5	2.886751		0.000000
## SurvivalfishhuNAre	Survivalfishfifty	Survivalfishtwentyfive	
## 1	26.457513	11.547005	17.320508
## 2	2.886751	5.773503	5.773503
## 3	5.773503	15.275252	0.000000
## 4	0.000000	10.000000	0.000000
## 5	0.000000	0.000000	0.000000
## AbnormalityfishhuNAre	Abnormalityfishfifty	Abnormalityfishtwentyfive	
## 1	30.334405	10.10363	22.24288
## 2	6.350853	11.24029	11.12847
## 3	0.000000	32.07758	10.00000
## 4	0.000000	10.55010	17.32051
## 5	0.000000	0.00000	0.00000
## HatchinghuNAre	Hatchingfifty	Hatchingtwentyfive	AlgaehuNAre
## 1	74.737166	13.228757	6.366360
## 2	21.650635	1.281718	7.696079
## 3	5.773503	6.199696	5.773503
## 4	5.773503	6.414361	34.641016
## 5	0.000000	0.000000	5.773503
			0.00
			0.00
## Algaetwentyfive	yielhuNAre	Seed.germinationhuNAre	seedgerminationfifty
## 1	0.00	50000.00	8.944272
## 2	20816.66	20000.00	5.477226
## 3	15275.25	37859.39	16.431677
## 4	336452.08	37859.39	13.416408
## 5	150000.00	17320.51	8.944272
## seedgerminationtwentfive	rootlengthhuNAre	rootlengthfifty	
## 1	5.477226	6.173573	12.018195
## 2	10.000000	8.618178	9.742022
## 3	8.366600	8.622471	9.842662
## 4	11.401754	10.488470	8.885831
## 5	5.477226	4.313004	12.273427
## rootlengtweenyfive	INAexgerminationhuNAre	INAexgerminationfifty	
## 1	13.960910	4.043884	7.252448
## 2	25.278588	3.941827	5.657031

```

## 3          9.716224      6.052851      4.664011
## 4          15.330590     2.250111      4.971720
## 5          12.510196     4.421199      3.636344
##  INAexgermianationseventyfive
## 1          4.651021
## 2          12.562524
## 3          6.536666
## 4          5.511170
## 5          6.276544

```

Figura 10: Desviaciones estándar de las variables del estudio en muestras de agua y suelo respecto a las locaciones.

- Datos de las medidas obtenidas de los roedores

```

##        BW       BL       BCI      BCIApo      BCIMus      Liver      Kidney
## 0  5.239348 10.108458 0.03985170 0.03892495 0.03370036 1.5898659 0.2147312
## 1  4.051375 11.871322 0.07585304 0.08287419 0.05750362 0.9314102 0.2825686
## 2  5.822177  9.994691 0.05933477 0.08201995 0.05115802 0.9898669 1.0463119
## 3  1.963502  6.145615 0.07228815 0.09574271 0.06446606 0.5746789 0.1831125
## 4 22.541927 10.040126 0.20872015 0.23874673 0.21350355 2.0382899 0.2302012
## Spleen Taillenght Handlenght Footlenght
## 0  0.1703518 12.636120   2.765560   3.078545
## 1  0.3495576  9.218399   4.177518   5.384912
## 2  0.1690899 14.757888   2.682813   3.391747
## 3  0.1427794  7.832010   4.969765   3.338165
## 4  0.1830581 13.160808   4.428479   1.888189

```

Figura 11: Desviaciones estándar de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones.

- Datos de los exámenes hematológicos de los roedores

```

##        WBC       LYM       MID       GRA       RBC       HGB       HCT
## 0  1.525831 0.9757921 0.1744901 1.0567311 0.875971 12.98072 3.160894
## 1  2.998077 1.9573311 0.1244304 1.2043421 1.156977 19.72742 4.389812
## 2  1.819384 1.2784471 0.1820046 0.6601168 1.293264 18.54788 4.726065
## 3  2.622432 1.8010219 0.1907625 0.9682984 1.234273 19.91400 5.029942
## 4  2.565072 2.1874543 0.2171087 0.5639377 1.337014 21.49690 4.253112
## MCV       MCH       MCHC Neutrophils Lymphocytes Eosinophils
## 0  2.983435 1.4406768 18.69476    7.778175 11.785772 2.480809
## 1  3.339581 1.2861776 16.24584    8.667004 9.672714 3.386419
## 2  2.949038 0.7902356 17.86734    5.039841 9.484856 3.544362
## 3  2.129888 0.9231449 19.24068   11.156409 17.605239 4.428443
## 4  2.550632 0.9381159 20.91993   4.670633 10.083978 1.957526
## Monocytes Basophils Plaletes
## 0   8.308394 1.2126781 1.3639260
## 1   5.431121 0.2294157 0.9334586
## 2   5.608030 0.2500000 1.2365948
## 3   6.985936 0.0000000 1.1270627
## 4   7.054731 0.1924501 1.1106837

```

Figura 12: Desviaciones estándar de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.

- Datos de los exámenes hispatológicos de los roedores

```

##      Liverh Kidneyh Spleenh Lungh
## 0 68.31301 80.03905 77.17436 26.72612
## 1 56.06119 57.53642 96.07689 76.79476
## 2 48.04512 30.68659 65.86528 58.82300
## 3 67.38123 64.97126 62.14555 150.92958
## 4 19.61161 50.24938 92.37212 65.03268

```

Figura 13: Desviaciones estándar de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.

Varianza: Es el promedio de las distancias al cuadrado que van de las observaciones a la media [4].

$$s_k^2 = \frac{\sum_{j=1}^n (x_{jk} - \bar{x}_k)^2}{n} \quad k = 1, 2, \dots, p$$

- Datos de los estudios en muestras de agua y suelo

```

##   AdultEnchytriaed JuvenilsEnchytriaed AdultPlatynotres AdultOppiatwelve
## 1           180             459.3          30                 70
## 2            30            30027.2         270                80
## 3            50            44274.7          70                20
## 4            0             4721.3         120                30
## 5            30            16707.7          30                 0
##   AdultOppiattwentyeighth JuvenilsOppiatwentytwo Adultfolsomia
## 1            120              4.7            0.5
## 2            470              1.7            0.3
## 3            0               4.8            1.7
## 4            120              14.7            2.0
## 5            30              19.7            0.5
##   Juvenilsfolsom Avoid.respectcontrol Nitrateveintiocho Seedfourteen
## 1            92.2            3980  0.1641333333  0.0000000
## 2            533.3            1580  0.0171000000  0.9166667
## 3            4956.8            600   0.0196333333  0.2500000
## 4            2963.3            2030  0.0004333333  0.2500000
## 5            628.7            2570  0.1157333333  0.2500000
##   Freshplantfourteen Highplant clorosisplant willingplant
## 1            0.024300  2.414800    0.0000            0.0000
## 2            0.000025  8.176900  1166.6667        366.6667
## 3            0.000225  8.000467    0.0000            0.0000
## 4            0.001025 23.034892  100.0000            0.0000
## 5            0.001225 69.332492  133.3333            0.0000
##   Malformationplant DaphniainmobilisationhuNAred
## 1            0.0000            58.333333
## 2            366.6667            8.333333
## 3            100.0000            8.333333
## 4            0.0000            25.000000
## 5            366.6667            0.000000
##   Daphniainmobilisationfifty Daphniainmobilisationtwentyfive
## 1            8.333333            33.333333
## 2            8.333333            8.333333
## 3            8.333333            8.333333
## 4            25.000000        100.000000

```

```

## 5          8.333333          0.000000
## SurvivalfishhuNAred Survivalfishfifty Survivalfishtwentyfive
## 1      700.00000      133.33333      300.00000
## 2      8.333333      33.33333      33.33333
## 3     33.333333     233.33333          0.00000
## 4      0.000000     100.00000          0.00000
## 5      0.000000      0.00000      0.00000
## AbnormalityfishhuNAred Abnormalityfishfifty Abnormalityfishtwentyfive
## 1      920.17613      102.0833      494.7457
## 2      40.33333      126.3440      123.8428
## 3      0.00000      1028.9712      100.0000
## 4      0.00000      111.3046      300.0000
## 5      0.00000      0.0000      0.0000
## HachinghuNAred Hachingfifty Hatchingtweentyfive AlgaehuNAred Algaefifty
## 1    5585.64403     175.00000     40.53053 3033333333 29633333333
## 2    468.75000      1.64280      59.22963 4900000000 49000000000
## 3    33.33333      38.43623      33.33333 1733333333 1733333333
## 4    33.33333      41.14403     1200.00000 1433333333 1433333333
## 5    0.00000      0.00000      33.33333      0      0
## Algaetwentyfive yielhuNAred Seed.germinationhuNAred seedgerminationfifty
## 1      0 2500000000          80          220
## 2    4333333333 4000000000          30          70
## 3    2333333333 1433333333          270          20
## 4  113200000000 1433333333          180          100
## 5  22500000000 3000000000          80          30
## seedgerminationtwentyp five rootlengthhuNAred rootlenghfifty
## 1      30      38.113      144.437
## 2      100      74.273      94.907
## 3      70      74.347      96.878
## 4      130      110.008      78.958
## 5      30      18.602      150.637
## rootlengtweentyfive INAexgerminationhuNAred INAexgerminationfifty
## 1    194.907      16.353      52.598
## 2    639.007      15.538      32.002
## 3    94.405      36.637      21.753
## 4    235.027      5.063      24.718
## 5    156.505      19.547      13.223
## INAexgermianationseventyfive
## 1      21.632
## 2      157.817
## 3      42.728
## 4      30.373
## 5      39.395

```

Figura 14: Varianza de las variables de los datos del estudio en muestras de agua y suelo respecto a las locaciones.

Como indica la figura 14 todas las variables tienen un alto grado de dispersión de los datos

- Datos de las medidas obtenidas de los roedores

```

##        BW       BL       BCI      BCIApo      BCIMus      Liver
## 0 27.45076 102.18092 0.001588158 0.001515152 0.001135714 2.5276737
## 1 16.41364 140.92829 0.005753684 0.006868132 0.003306667 0.8675250
## 2 33.89774  99.89385 0.003520615 0.006727273 0.002617143 0.9798365
## 3  3.85534  37.76859 0.005225577 0.009166667 0.004155873 0.3302558
## 4 508.13848 100.80413 0.043564103 0.057000000 0.045583766 4.1546256

```

```

##      Kidney      Spleen Taillenght Handlenght Footlenght
## 0 0.04610947 0.02901974 159.67152    7.648322   9.477442
## 1 0.07984500 0.12219053  84.97889   17.451654  28.997279
## 2 1.09476862 0.02859138 217.79524    7.197488  11.503948
## 3 0.03353019 0.02038596  61.34039   24.698569  11.143348
## 4 0.05299259 0.03351026 173.20687   19.611428   3.565256

```

Figura 15: Varianza de las variables de los datos sobre las medidas obtenidas de los roedores respecto a las locaciones

Como indica la figura 15 la variable BL, Taillenght, Handlenght y Footlenght tiene un alto grado de dispersión de los datos

- **Datos de los exámenes hematológicos de los roedores**

```

##      WBC       LYM       MID       GRA       RBC       HGB       HCT
## 0 2.328159 0.9521702 0.03044678 1.1166807 0.7673251 168.4990 9.991254
## 1 8.988467 3.8311450 0.01548292 1.4504400 1.3385962 389.1710 19.270450
## 2 3.310157 1.6344269 0.03312569 0.4357542 1.6725312 344.0239 22.335695
## 3 6.877151 3.2436797 0.03639034 0.9376017 1.5234309 396.5673 25.300317
## 4 6.579593 4.7849562 0.04713619 0.3180257 1.7876062 462.1168 18.088966
##      MCV       MCH       MCHC Neutrophils Lymphocytes Eosinophils
## 0 8.900882 2.0755497 349.4941    60.50000 138.9044 6.154412
## 1 11.152800 1.6542529 263.9272    75.11696 93.5614 11.467836
## 2 8.696827 0.6244723 319.2417    25.40000 89.9625 12.562500
## 3 4.536423 0.8521966 370.2039    124.46547 309.9444 19.611111
## 4 6.505725 0.8800614 437.6434    21.81481 101.6866 3.831909
##      Monocytes Basophils Plaletes
## 0 69.02941 1.47058824 1.860294
## 1 29.49708 0.05263158 0.871345
## 2 31.45000 0.06250000 1.529167
## 3 48.80330 0.00000000 1.270270
## 4 49.76923 0.03703704 1.233618

```

Figura 16: Varianza de las variables sobre los exámenes hematológicos de los roedores respecto a las locaciones.

Como indica la figura 16 la variable HGB, HCT, MCV, MCHC, Neutrophils, Lymphocytes, Monocytes tienen un alto grado de dispersión de los datos.

- **Datos de los exámenes hispatológicos de los roedores**

```

##      Liverh     Kidneyh     Spleenh      Lungh
## 0 4666.6667 6406.2500 5955.882    714.2857
## 1 3142.8571 3310.4396 9230.769   5897.4359
## 2 2308.3333 941.6667 4338.235   3460.1449
## 3 4540.2299 4221.2644 3862.069 22779.7379
## 4 384.6154 2525.0000 8532.609   4229.2490

```

Figura 17: Varianza de las variables sobre los exámenes hispatológicos de los roedores respecto a las locaciones.

Como indica la figura 17 todas las variables tienen un alto grado de dispersión de los datos.

2.1. 2 Análisis Gráfico

- Datos de los estudios en muestras de agua y suelo

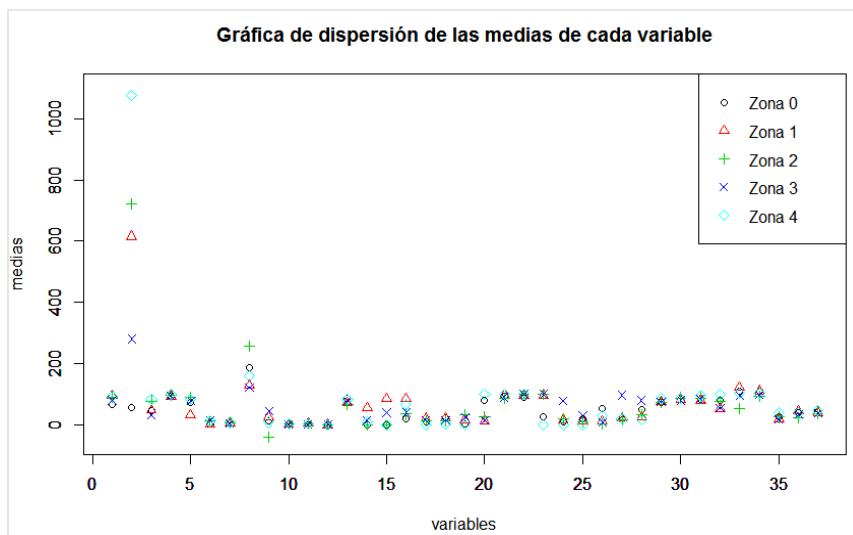


Figura 18: Gráficas de dispersión de las medias de las variables del estudio de muestras de agua y suelo en cada una de las locaciones.

En el análisis estadístico descriptivo y gráfico de los datos de ecotoxicología se concluyó que en cada uno de los test aplicados a la tierra y agua tiene una contaminación media mayor en las zonas 4 (Zona Control) y 2 (Complejo Petroquímico Adyacente) y la zona 0(Incineradora de Residuos Peligrosos de Constantí) es el que tiene menor polución respecto al resto de locaciones.

- Datos de las medidas obtenidas de los roedores

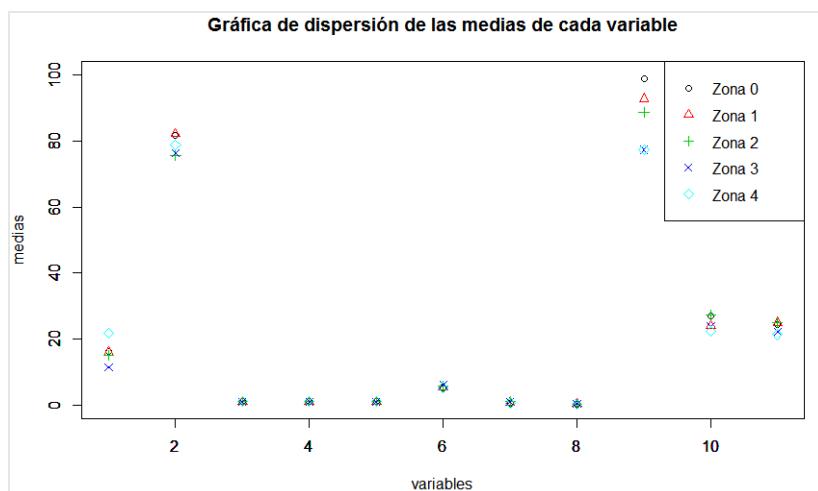


Figura 19: Gráficas de dispersión de las medias de las variables obtenidas de los roedores en cada una de las locaciones.

En el análisis estadístico descriptivo y gráfico de los datos de sobre las medidas recolectadas de los roedores se concluyó que en cada una de las medidas tiene una contaminación media mayor en las zonas 1 (Incineradora de Residuos Peligrosos de Constantí) y 2 (Complejo Petroquímico Adyacente).

- Datos de los exámenes hematológicos de los roedores

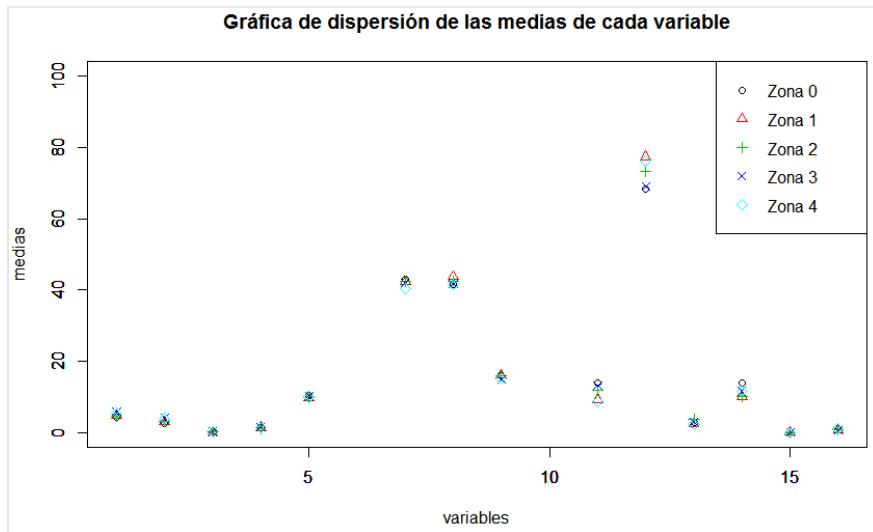


Figura 20: Gráficas de dispersión de las medias de las variables cuyos datos son los resultados de los exámenes hematológicos de los roedores en cada una de las locaciones.

En las muestras de sangre de los roedores capturados sobre el terreno se tiene una polución media mayor en las zonas 2(Complejo Petroquímico Adyacente) y la zona 0 (Incineradora de residuos peligrosos de Constantí).

- Datos de los exámenes hispatológicos de los roedores

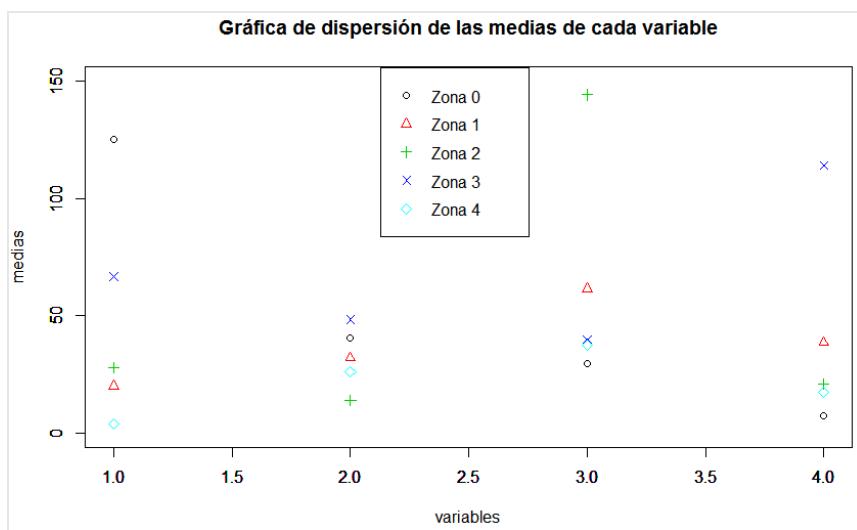


Figura 21: Gráficas de dispersión de las medias de las variables cuyos datos son los resultados de los exámenes hematológicos de los roedores en cada una de las locaciones.

En las muestras de sangre de los roedores capturados sobre el terreno se tiene una polución media mayor en las zonas 3 (Tráfico rodado de diversas carreteras y autovías) y la zona 2 (Complejo Petroquímico Adyacente).

- Matriz de Correlación

Permite medir el grado de relación que existe entre las variables, los valores están en el rango de -1 y 1.

$$\mathbf{R} = \mathbf{D}^{-1} \mathbf{S} \mathbf{D}^{-1}$$

Donde D es la matriz en cuya diagonal contiene las desviaciones típicas de las variables que contiene y S es la matriz de varianzas y covarianzas.

Se realizó un análisis de correlación de los datos para identificar la relación existente entre las variables de estudio.

- Datos de las medidas obtenidas de los roedores

	BW	BL	BCI	Liver	Kidney
## BW	1.0000000	0.40197245	0.81855662	-6.650018e-01	-0.247099639
## BL	0.40197245	1.0000000	-0.03070502	-3.250855e-01	-0.185477409
## BCI	0.81855662	-0.03070502	1.0000000	-5.966356e-01	-0.160010363
## Liver	-0.66500182	-0.32508545	-0.59663556	1.000000e+00	0.172903569
## Kidney	-0.24709964	-0.18547741	-0.16001036	1.729036e-01	1.000000000
## Spleen	-0.26638751	-0.14263491	-0.23729350	3.560354e-01	0.083009790
## Taillenght	0.07777150	0.09506007	0.17360274	-1.936162e-01	-0.007118844
## Handlenght	-0.05985199	-0.27026967	0.07580786	1.465242e-05	0.050925470
## Footlenght	-0.05963313	-0.19407492	0.13115160	-1.326385e-01	-0.021703242
	Spleen	Taillenght	Handlenght	Footlenght	
## BW	-0.26638751	0.077771498	-5.985199e-02	-0.05963313	
## BL	-0.14263491	0.095060070	-2.702697e-01	-0.19407492	
## BCI	-0.23729350	0.173602738	7.580786e-02	0.13115160	
## Liver	0.35603535	-0.193616210	1.465242e-05	-0.13263851	
## Kidney	0.08300979	-0.007118844	5.092547e-02	-0.02170324	
## Spleen	1.00000000	-0.185378064	-7.180818e-02	-0.21065550	
## Taillenght	-0.18537806	1.000000000	3.876435e-01	0.51373191	
## Handlenght	-0.07180818	0.387643501	1.000000e+00	0.36491131	
## Footlenght	-0.21065550	0.513731915	3.649113e-01	1.00000000	

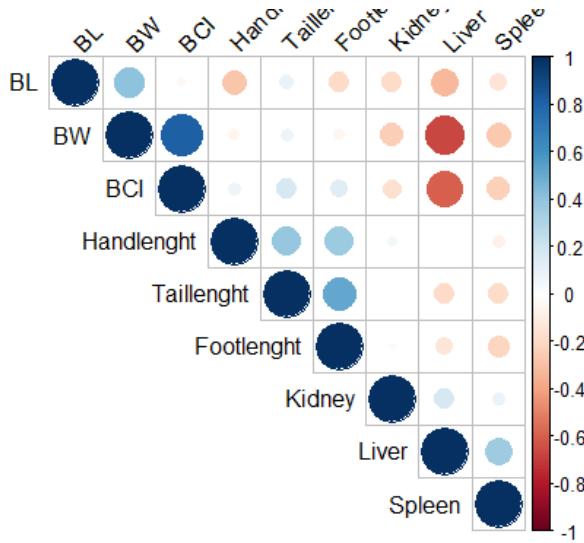


Figura 22: Correlación de las variables sobre las medidas obtenidas de los roedores.

La Figura 22 muestra que las variables BW – BCI y Taillenght - Footlenght presentan una relación positiva, mientras que una relación negativa existe entre las variables Liver – BW y Liver – BCI.

- Datos de los exámenes hematológicos de los roedores

	LYM	MID	GRA	RBC
## LYM	1.000000000	0.20329876	0.511294874	0.33116583
## MID	0.203298756	1.000000000	0.145228045	0.10276636
## GRA	0.511294874	0.14522804	1.000000000	0.19499041
## RBC	0.331165825	0.10276636	0.194990408	1.00000000
## HGB	0.317154741	0.14111658	0.194313030	0.80322692
## HCT	0.386021455	0.14724765	0.235513668	0.83972167
## MCV	-0.004470736	0.05372302	-0.003135731	-0.52675074
## MCH	-0.001469873	0.07072079	0.010866991	-0.24429670
## MCHC	-0.004410670	0.03412556	0.022293519	0.26849166
## Neutrophilis	-0.155544046	0.17681024	0.220645671	0.06936051
## Lymphocytes	0.197584062	-0.21078239	-0.215848475	-0.02649833
## Eosinophils	0.023093322	0.14704310	0.082458921	0.03223711
## Monocytes	-0.107772102	0.04179829	-0.025035003	-0.04835558
## Basophils	-0.146218176	0.02128678	0.401137160	-0.02609621
## Plaletes	-0.208500093	0.10275432	0.212175992	-0.08632331
##	HGB	HCT	MCV	MCH
## LYM	0.317154741	0.386021455	-0.004470736	-0.001469873
## MID	0.141116576	0.147247646	0.053723016	0.070720786
## GRA	0.194313030	0.235513668	-0.003135731	0.010866991
## RBC	0.803226918	0.839721673	-0.526750741	-0.244296700
## HGB	1.000000000	0.887321415	-0.093064739	0.377916265
## HCT	0.887321415	1.000000000	0.012999416	0.139439456
## MCV	-0.093064739	0.012999416	1.000000000	0.672791249
## MCH	0.377916265	0.139439456	0.672791249	1.000000000
## MCHC	0.594833446	0.160296664	-0.242325608	0.552737699
## Neutrophilis	0.033528166	0.006424095	-0.124104562	-0.052485938
## Lymphocytes	0.017037016	0.061183237	0.149242258	0.066674106
## Eosinophils	0.004414079	-0.009266907	-0.081708431	-0.039665403
## Monocytes	-0.044319205	-0.062706515	-0.005570299	0.004135340
## Basophils	-0.073507515	-0.172874665	-0.218936845	-0.073453870

```

## Plaletes      -0.173526014 -0.222997207 -0.186133362 -0.150374571
##             MCHC  Neutrophilis Lymphocytes Eosinophils
## LYM       -0.00441067 -1.555440e-01  0.19758406  0.023093322
## MID        0.03412556  1.768102e-01 -0.21078239  0.147043096
## GRA        0.02229352  2.206457e-01 -0.21584847  0.082458921
## RBC        0.26849166  6.936051e-02 -0.02649833  0.032237106
## HGB        0.59483345  3.352817e-02  0.01703702  0.004414079
## HCT        0.16029666  6.424095e-03  0.06118324 -0.009266907
## MCV       -0.24232561 -1.241046e-01  0.14924226 -0.081708431
## MCH        0.55273770 -5.248594e-02  0.06667411 -0.039665403
## MCHC       1.00000000  6.515056e-02 -0.07978150  0.041020894
## Neutrophilis 0.06515056  1.000000e+00 -0.77322999 -0.097615864
## Lymphocytes -0.07978150 -7.732300e-01  1.00000000 -0.196677453
## Eosinophils  0.04102089 -9.761586e-02 -0.19667745  1.000000000
## Monocytes    0.01424389  5.439309e-06 -0.54749124 -0.046065569
## Basophils     0.17564502  1.775854e-01 -0.30460123  0.134195272
## Plaletes     0.01946768  4.059790e-01 -0.41684561  0.031104139
##             Monocytes Basophils Plaletes
## LYM       -1.077721e-01 -0.14621818 -0.20850009
## MID        4.179829e-02  0.02128678  0.10275432
## GRA       -2.503500e-02  0.40113716  0.21217599
## RBC       -4.835558e-02 -0.02609621 -0.08632331
## HGB       -4.431921e-02 -0.07350752 -0.17352601
## HCT       -6.270651e-02 -0.17287466 -0.22299721
## MCV       -5.570299e-03 -0.21893685 -0.18613336
## MCH        4.135340e-03 -0.07345387 -0.15037457
## MCHC       1.424389e-02  0.17564502  0.01946768
## Neutrophilis 5.439309e-06  0.17758540  0.40597896
## Lymphocytes -5.474912e-01 -0.30460123 -0.41684561
## Eosinophils -4.606557e-02  0.13419527  0.03110414
## Monocytes   1.0000000e+00  0.08855369 -0.02587215
## Basophils    8.855369e-02  1.00000000  0.38020532
## Plaletes    -2.587215e-02  0.38020532  1.00000000

```

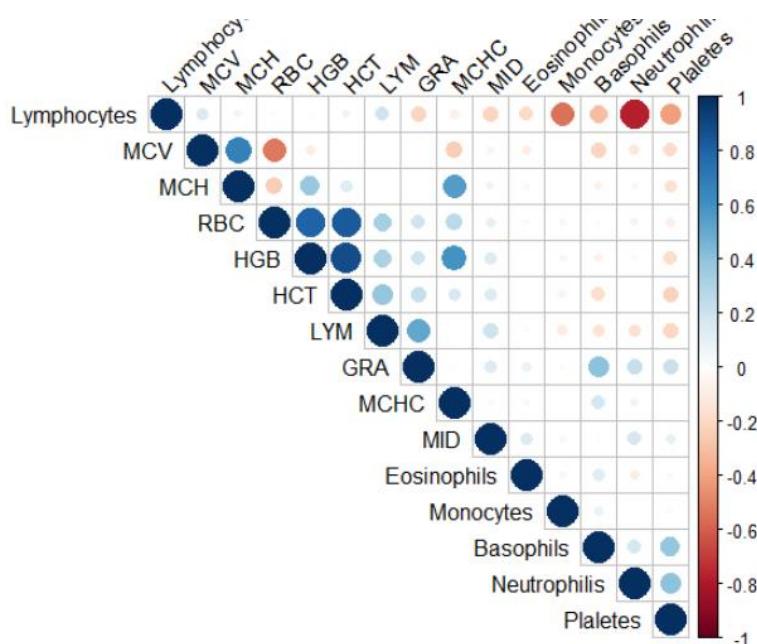


Figura 23: Correlación de las variables sobre los exámenes hematológicos de los roedores.

Como indica la figura 19 las variables que están correlacionadas son HGB – RBC, HGB – HCT, HCT – RBC, MCH – MCV, Neutrophils – Lymphocytes.

- Datos de los exámenes hispatológicos de los roedores

	Liverh	Kidneyh	Spleenh	Lungh
## Liverh	1.0000000	0.04292798	-0.11981702	0.06022104
## Kidneyh	0.04292798	1.0000000	-0.02646874	0.09005911
## Spleenh	-0.11981702	-0.02646874	1.0000000	0.02497212
## Lungh	0.06022104	0.09005911	0.02497212	1.0000000

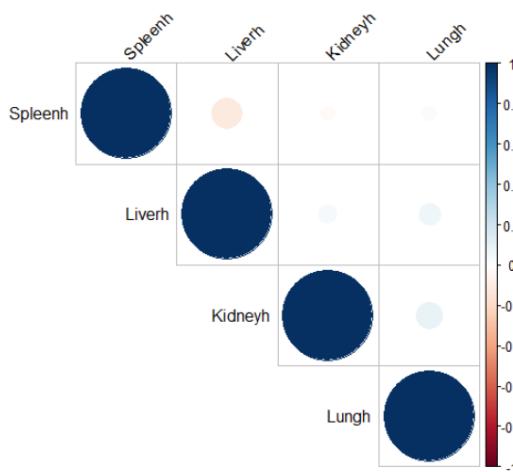


Figura 24: Correlación de las variables sobre los exámenes hispatológicos de los roedores

Como se observa en la figura 24 no hay una relación fuerte entre las variables respecto a los exámenes hispatológicos aplicado a los roedores.

2.2 Variables redundantes

Se quiere disminuir el tamaño de la matriz de datos para ello obtenemos la matriz S y calculamos sus autovalores y autovectores asociados, si se obtiene valores nulos en los autovalores indica que existen variables redundantes y dichas variables se debe excluir del estudio ya que es una combinación lineal de demás variables.

$$x_{i1} = \bar{x}_1 - \frac{w_2}{w_1}(x_{i2} - \bar{x}_2) - \dots - \frac{w_p}{w_1}(x_{ip} - \bar{x}_p).$$

- Datos de las medidas obtenidas de los roedores

\$values
[1] 2.124856e+02 1.552204e+02 6.475198e+01 1.415528e+01 9.124558e+00
[6] 8.910192e-01 2.419754e-01 3.485075e-02 2.053212e-03
##
\$vectors

```

##          [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.222862739 0.792997662 0.5611904006 0.029968289 0.003044736
## [2,] 0.165349349 0.529611090 -0.8047382757 -0.199442237 -0.067438866
## [3,] 0.002484231 0.004975681 0.0100626248 0.002700414 -0.001578297
## [4,] -0.030366347 -0.059124223 -0.0356756413 0.004635720 0.063312227
## [5,] -0.002428420 -0.010469030 -0.0009366854 0.002128733 0.014052719
## [6,] -0.003397031 -0.003113660 -0.0021273173 0.002978621 0.011375006
## [7,] 0.945656282 -0.255846545 -0.0239015252 0.162124131 0.115724577
## [8,] 0.109256226 -0.111831603 0.1650469555 -0.948035385 0.222459046
## [9,] 0.125870337 -0.095201428 0.0909420183 -0.185005904 -0.963444199
##          [,6]      [,7]      [,8]      [,9]
## [1,] -0.073848895 0.009828580 -0.0012904495 -0.009640923
## [2,] -0.012856383 0.006677374 -0.0013670164 0.005571732
## [3,] 0.008198800 0.001525301 -0.0059646634 0.999876446
## [4,] -0.994240182 0.006057393 0.0405540995 0.009201384
## [5,] 0.007800806 0.999806912 0.0021618785 -0.001492274
## [6,] -0.039687275 0.002252626 -0.9991119329 -0.005582856
## [7,] -0.004724738 -0.002340113 -0.0003779118 -0.001050938
## [8,] 0.007099491 -0.001910361 -0.0009625635 0.001474557
## [9,] -0.063965898 0.013846732 -0.0092658644 -0.001327231

```

Figura 25: Autovalores y autovectores correspondiente a las matriz S cuyos datos son las medidas obtenidas de los roedores

Como no tenemos ningún autovalor nulo se concluye que todas las variables son importantes para el estudio.

- Datos de los exámenes hematológicos de los roedores

```

## $values
## [1] 6.914988e+02 2.158002e+02 1.755767e+02 5.431632e+01 1.569499e+01
## [6] 8.426276e+00 7.726393e+00 1.495599e+00 5.212992e-01 1.520535e-01
## [11] 5.402086e-02 3.537942e-02 8.292341e-03 1.817713e-03 1.407364e-03
##
## $vectors
##          [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.0150671573 0.0371504907 -0.054572699 -0.0097151014 -0.101751787
## [2,] 0.0103333546 0.0433398261 -0.030152314 0.0027729616 -0.055375839
## [3,] 0.0006504278 -0.0013109247 -0.003020079 -0.0004545230 -0.006537554
## [4,] 0.0041257361 -0.0048463602 -0.021419016 -0.0119689772 -0.039686336
## [5,] 0.0261477088 0.0244606004 -0.051218035 -0.0011731930 -0.030943930
## [6,] 0.6391545054 0.3833026282 -0.604188687 0.0383273050 0.002499656
## [7,] 0.0887687857 0.1378967261 -0.229943184 0.0124420161 -0.011214343
## [8,] -0.0213031465 0.0352972494 -0.012971299 0.0193029198 0.124580571
## [9,] 0.0242911136 0.0007188799 0.018223498 0.0058331948 0.047501451
## [10,] 0.7616415601 -0.2991537190 0.560522184 -0.0182999074 0.009519345
## [11,] 0.0255926497 -0.4406820435 -0.303789555 -0.6020774052 0.290379143
## [12,] -0.0272118537 0.7103165655 0.399649371 -0.1918375937 0.253054514
## [13,] 0.0039230045 -0.0243027910 -0.003972294 0.0571093099 -0.834707207
## [14,] -0.0017296840 -0.1963642642 -0.091478081 0.7704863427 0.345416837
## [15,] 0.0015852541 -0.0127619251 0.003606359 0.0004717035 -0.023748256
## [16,] -0.0029264685 -0.0371415997 -0.003664878 -0.0343223295 -0.032186306
##          [,6]      [,7]      [,8]      [,9]      [,10]
## [1,] 0.186824055 0.76812342 0.048679938 -0.136987789 -0.095209815
## [2,] 0.150519377 0.52607492 -0.194919273 0.485765454 0.258772553
## [3,] -0.003539892 0.01435985 0.003241226 0.032100695 0.012167541
## [4,] 0.039862929 0.22751410 0.240298812 -0.655884583 -0.367687413
## [5,] 0.207386580 -0.06256042 -0.007152038 -0.017231269 0.039140608

```

```

## [6,] -0.023516846 -0.06677649  0.017568380  0.033013004 -0.023242592
## [7,]  0.006428455 -0.03107026  0.015410438 -0.115870602  0.091659034
## [8,] -0.870706214  0.23406540  0.059648800  0.018854288  0.053354438
## [9,] -0.325895286  0.08970979  0.011112847  0.023229028 -0.023041196
## [10,] -0.005598521  0.04150035 -0.008956853 -0.012333088  0.007402612
## [11,] -0.001108269  0.02761519 -0.280544707 -0.007346646 -0.140385051
## [12,]  0.037441570  0.00382312 -0.205903088 -0.019173506 -0.150744756
## [13,] -0.161699545 -0.07038101 -0.278704848  0.005019554 -0.157500675
## [14,]  0.050439459  0.03675493 -0.212824793 -0.006843382 -0.154192330
## [15,]  0.025498961  0.01284053  0.190126510 -0.361476615  0.801490731
## [16,]  0.046075745 -0.01100998  0.785686608  0.408695868 -0.203279921
##           [,11]          [,12]          [,13]          [,14]          [,15]
## [1,] -0.0654420249  0.274979940 -0.009127375 -0.003862759  0.009569441
## [2,]  0.1048338163 -0.303821899  0.004896253  0.007806873 -0.012541962
## [3,] -0.1879322671  0.842192747 -0.049541861  0.013914192  0.022645662
## [4,]  0.0131777862 -0.266789979  0.028214169 -0.001703171 -0.003748064
## [5,]  0.1466316817  0.082294287  0.837664273 -0.448094521  0.124525943
## [6,] -0.2593643374 -0.059795574 -0.011438155  0.001686991 -0.010529415
## [7,]  0.9079019929  0.193882707 -0.157158593  0.100997741  0.010117228
## [8,]  0.0392809400  0.020188308  0.341376902  0.191868353 -0.085204641
## [9,]  0.0355982318 -0.017984004 -0.385961119 -0.798703768  0.304594597
## [10,] 0.1071850455  0.025552552  0.021092509  0.033193437 -0.008339285
## [11,] 0.0045342738 -0.006215016  0.028695293  0.141887401  0.381827291
## [12,] 0.0030225513 -0.001434487  0.027851056  0.141315811  0.382739552
## [13,] -0.0007442788 -0.013185264  0.029606297  0.140780194  0.380658925
## [14,]  0.0053559973 -0.003754738  0.027781663  0.141431230  0.382462527
## [15,] -0.1386288434 -0.042892205 -0.018687409  0.112767328  0.393849713
## [16,]  0.0510150007 -0.014903269  0.024196102  0.142309662  0.380550953

```

Figura 26: Autovalores y autovectores correspondientes a las matriz S cuyos datos son los resultados de los exámenes hematológicos

En la figura 26 no se registra autovalores nulos por lo tanto todas las variables son consideradas en el estudio.

- Datos de los exámenes hispatológicos de los roedores

```

## $values
## [1] 13000.102  8016.664  4409.414  3500.991
##
## $vectors
##           [,1]          [,2]          [,3]          [,4]
## [1,] -0.05271487 -0.20962992  0.96839211  0.12447156
## [2,] -0.06483178 -0.04228270  0.11461685 -0.99038982
## [3,] -0.03965037  0.97654528  0.21112120 -0.01466324
## [4,] -0.99571373 -0.02503582 -0.06713828  0.05847930

```

Figura 27: Autovalores y autovectores correspondiente a las matriz S cuyos datos son los resultados de los exámenes hispatológicos

En la figura 27 también se tiene que no hay autovalores nulos, en conclusión, se indica que todas las variables son indispensables y se requieren en el estudio.

2.3 Datos atípicos

La detección de valores atípicos es una tarea muy importante en el análisis de datos. Los valores extremos describen el comportamiento de los datos anormales, es decir, datos que se desvían de la variabilidad natural de los datos. A menudo, los valores extremos son de interés primordial, por tanto, es importante identificarlos antes del modelado y análisis[5]

Los métodos estadísticos para la detección de valores atípicos multivariantes a menudo indican las observaciones que se encuentran relativamente lejos del centro de la distribución de datos. Se pueden implementar varias medidas de distancia para tal tarea [5]

La distancia de Mahalanobis es un criterio muy conocido que depende de los parámetros estimados de la distribución multivariada. Ésta describe la distancia entre cada punto de datos y el centro de masa. Cuando un punto se encuentra en el centro de masa, la distancia de Mahalanobis es cero y cuando un punto de datos se encuentra distante del centro de masa, la distancia es mayor a cero. Por lo tanto, los puntos de datos que se encuentran lejos del centro de masa se consideran valores atípicos [6]

Se puede también utilizar gráficos para detectar valores outliers a simple vista, pero corre el riesgo de cometer un error al momento de registrar el valor.

2.3.1 Análisis gráfico para visualizar los datos extremos

- **Datos de los estudios en muestras de agua y suelo**

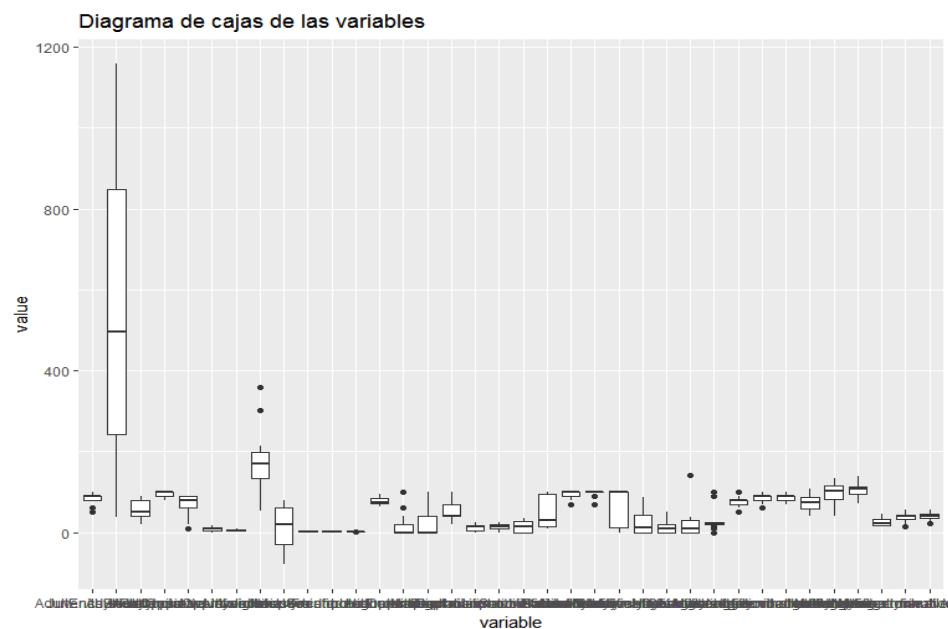


Figura 28: Diagrama de cajas de las variables del estudio de las muestras del suelo y agua

En la figura 28 podemos ver que hay un porcentaje mínimo de variables en el que se visualiza datos atípicos.

- **Datos de las medidas obtenidas de los roedores**

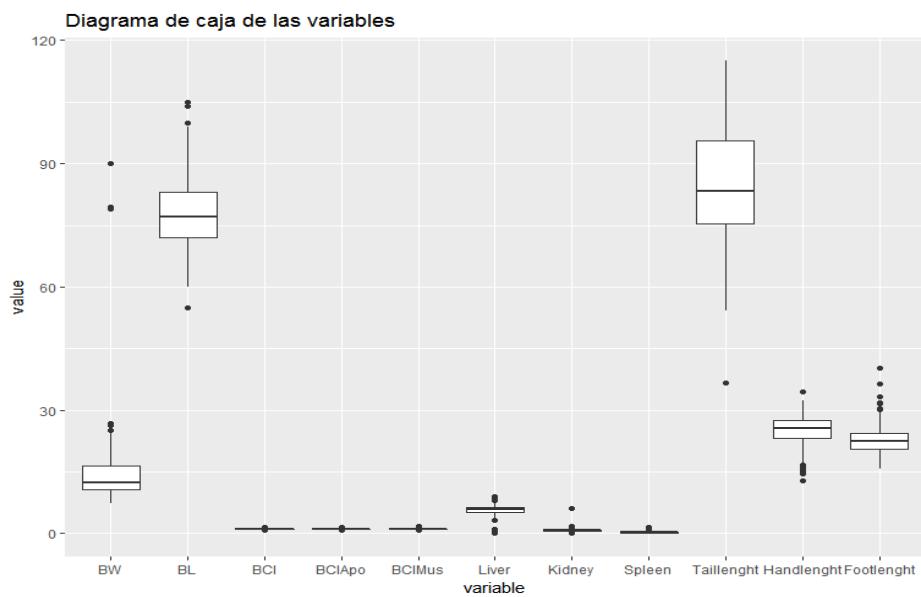


Figura 29: Diagrama de cajas de las variables que son las medidas obtenidos de los roedores.

En la figura 29 se muestra como 5 variables presentan datos extremos, tendremos que comprobar si efectivamente el dato atípico univariante es multivariante.

- **Datos de los exámenes hematológicos**

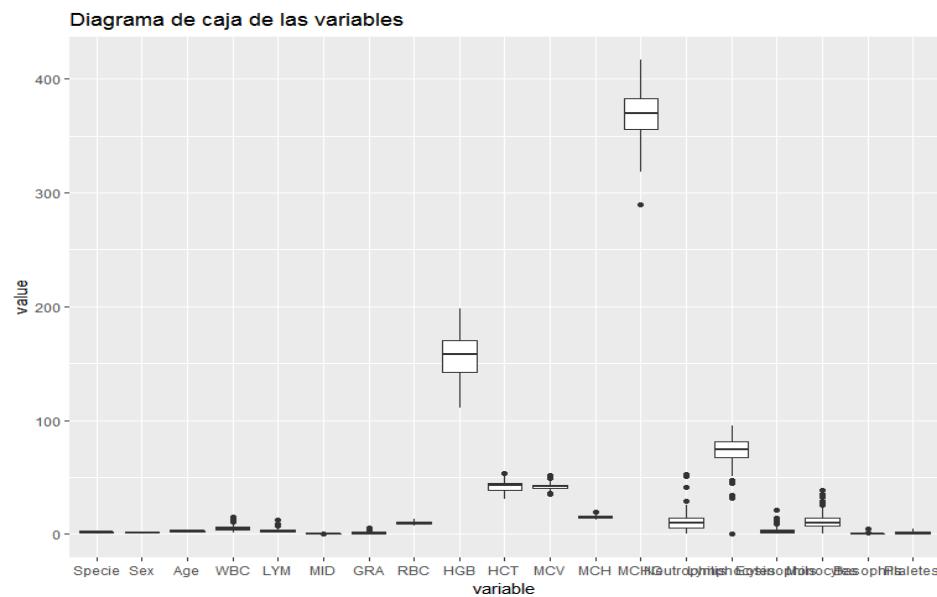


Figura 30: Diagrama de cajas de las variables resultado de los exámenes hematológicos obtenidos de los roedores respecto a las locaciones.

Como se muestra en la figura 30 hay un mínimo porcentaje de variables con datos extremos.

- **Datos de los exámenes hispatológicos**

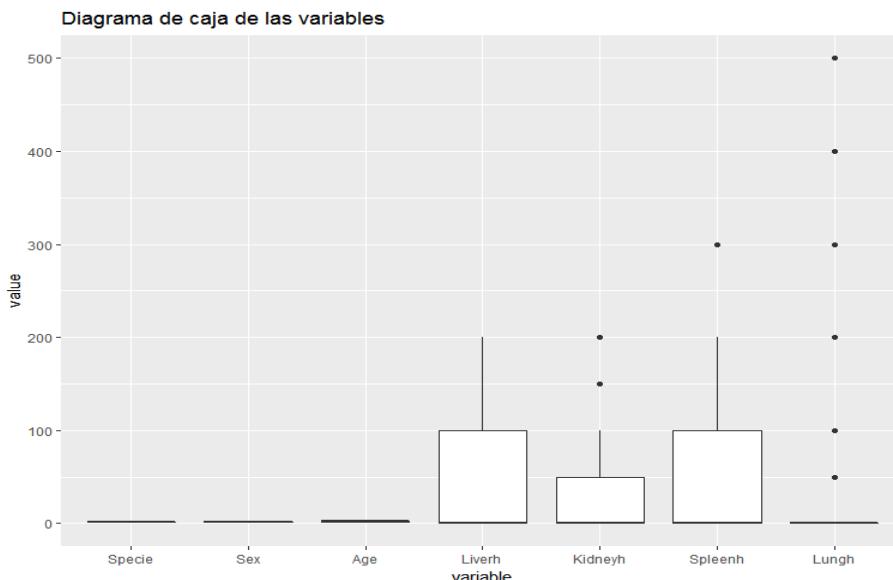


Figura 31: Diagrama de cajas de las variables resultado de los exámenes hispatológicos de los roedores.

En la figura 31 concuerdan en que existen datos extremos para cada variable, sin embargo, tendremos que probar si son datos atípicos.

2.3.2 Método para la detección de valores atípicos multivariantes

Para cerciorarnos de la existencia de datos atípicos aplicaremos el método de la distancia de mahalanobis.

- **Datos de las medidas obtenidas de los roedores**

```
result$outlier

##      Observation Mahalanobis Distance Outlier
## C2          C2            37534.372   TRUE
## C4          C4            24983.839   TRUE
## C3          C3            24302.871   TRUE
## I29         I29           1140.931    TRUE
## I216        I216          249.824     TRUE
## I22         I22           206.510     TRUE
## C23         C23           192.878    TRUE
## I25         I25           185.931     TRUE
## I016        I016          164.051     TRUE
## I16         I16           145.794     TRUE
## I011        I011          134.698     TRUE
```

## I115	I115	110.098	TRUE
## I013	I013	109.878	TRUE
## C18	C18	106.484	TRUE
## I112	I112	105.579	TRUE
## C20	C20	101.684	TRUE
## I24	I24	90.568	TRUE
## I217	I217	87.328	TRUE
## I34	I34	85.360	TRUE
## I015	I015	81.421	TRUE
## I21	I21	67.485	TRUE
## I14	I14	59.312	TRUE
## I012	I012	48.243	TRUE
## C6	C6	47.319	TRUE
## I020	I020	45.168	TRUE
## I113	I113	44.125	TRUE
## C24	C24	40.026	TRUE
## I23	I23	38.729	TRUE
## I11	I11	36.035	TRUE
## I09	I09	30.666	TRUE
## I01	I01	30.457	TRUE
## I13	I13	29.679	TRUE
## I337	I337	28.727	TRUE
## I116	I116	25.226	TRUE
## C17	C17	24.168	TRUE
## I18	I18	24.061	TRUE
## C22	C22	23.089	TRUE
## I39	I39	22.559	TRUE
## I117	I117	19.921	TRUE
## I311	I311	18.648	FALSE
## I019	I019	18.514	FALSE
## C14	C14	18.135	FALSE
## I312	I312	17.920	FALSE
## C19	C19	17.823	FALSE
## I316	I316	17.399	FALSE
## I319	I319	17.280	FALSE
## C12	C12	17.183	FALSE
## I110	I110	16.254	FALSE
## C15	C15	14.771	FALSE
## I17	I17	13.436	FALSE
## I317	I317	13.226	FALSE
## I212	I212	13.066	FALSE
## I318	I318	12.558	FALSE
## I226	I226	11.778	FALSE
## I221	I221	11.735	FALSE
## I340	I340	11.119	FALSE
## I334	I334	10.903	FALSE
## I010	I010	10.681	FALSE
## I12	I12	10.489	FALSE
## I19	I19	10.451	FALSE
## I213	I213	9.725	FALSE
## I35	I35	9.725	FALSE
## I017	I017	9.605	FALSE
## I111	I111	9.237	FALSE
## I03	I03	8.946	FALSE
## I330	I330	8.687	FALSE
## I331	I331	8.687	FALSE
## I322	I322	8.334	FALSE
## I114	I114	8.280	FALSE
## I315	I315	7.987	FALSE

## I321	I321	7.934	FALSE
## I28	I28	7.855	FALSE
## I327	I327	7.675	FALSE
## I325	I325	7.562	FALSE
## I118	I118	7.521	FALSE
## C25	C25	7.195	FALSE
## C5	C5	7.145	FALSE
## C8	C8	7.134	FALSE
## I05	I05	7.132	FALSE
## I329	I329	6.857	FALSE
## I218	I218	6.777	FALSE
## I02	I02	6.680	FALSE
## C21	C21	6.556	FALSE
## C9	C9	6.478	FALSE
## C16	C16	6.416	FALSE
## I328	I328	6.345	FALSE
## I38	I38	6.083	FALSE
## C26	C26	6.081	FALSE
## C27	C27	6.081	FALSE
## I014	I014	5.854	FALSE
## I119	I119	5.744	FALSE
## I018	I018	5.554	FALSE
## I26	I26	5.548	FALSE
## I27	I27	5.466	FALSE
## I214	I214	5.211	FALSE
## I332	I332	5.209	FALSE
## I224	I224	5.168	FALSE
## I335	I335	5.136	FALSE
## I225	I225	4.815	FALSE
## I313	I313	4.720	FALSE
## I333	I333	4.685	FALSE
## I338	I338	4.643	FALSE
## I336	I336	4.347	FALSE
## I04	I04	4.292	FALSE
## I215	I215	3.987	FALSE
## I324	I324	3.783	FALSE
## I222	I222	3.763	FALSE
## I314	I314	3.559	FALSE
## C13	C13	3.556	FALSE
## I120	I120	3.549	FALSE
## I06	I06	3.486	FALSE
## I320	I320	3.376	FALSE
## I210	I210	3.264	FALSE
## I310	I310	3.217	FALSE
## C1	C1	3.100	FALSE
## I33	I33	3.088	FALSE
## C11	C11	2.915	FALSE
## I339	I339	2.867	FALSE
## I223	I223	2.573	FALSE
## C7	C7	2.517	FALSE
## I32	I32	2.223	FALSE
## I36	I36	2.102	FALSE
## I37	I37	2.001	FALSE
## I15	I15	1.977	FALSE
## I08	I08	1.922	FALSE
## I326	I326	1.695	FALSE
## I220	I220	1.667	FALSE
## C10	C10	1.372	FALSE
## I31	I31	1.237	FALSE

## I07	I07	0.744	FALSE
## I211	I211	0.683	FALSE

Figura 32: Datos atípicos de las medidas obtenidas de los roedores

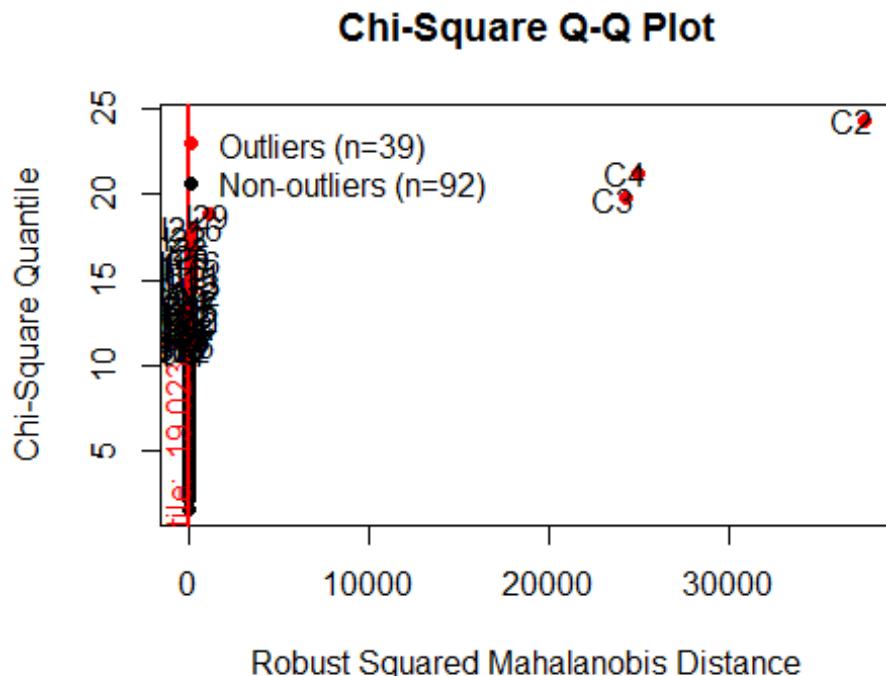


Figura 33: Gráfica de los outlier multivariados de las medidas obtenidas de los roedores.

Al realizar el análisis para averiguar si existen datos extremos, se concluye que se han encontrado 39 datos atípicos como lo indica en la figura 33, los mismos que serán excluidos de la base de datos.

Como resultado se obtuvo que los datos del estudio de muestras de agua y suelo, los datos de los exámenes hematológicos e hispatológicos de los roedores no hay presencia de datos atípicos.

2.4 Análisis de componentes principales

El Análisis de Componentes Principales (ACP) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. [7]

Las fases para el análisis de componentes principales [7]

- Análisis de la matriz de correlaciones
- Selección de los factores
- Análisis de la matriz factorial

- Interpretación de los factores
- Cálculo de las puntuaciones factoriales

Se calculan mediante $X_{ij} = a_{i1} \cdot Z_{1j} + \dots + a_{ik} \cdot Z_{kj} = \sum_{s=1}^k a_{is} \cdot Z_{sj}$

- **Análisis de los datos del estudio de las muestras de agua y suelo**

Para comenzar, nos interesa explorar la cantidad de componentes a extraer. Para ello, solicitamos tantos componentes como variables involucradas en el análisis.

```
##          eigenvalue percentage of variance cumulative percentage of variance
##  comp 1    12.031161           29.344295            29.34429
##  comp 2     6.154294           15.010472            44.35477
##  comp 3     6.066892           14.797297            59.15206
##  comp 4     3.973569           9.691632            68.84370
##  comp 5     2.464918           6.011995            74.85569
```

Figura 34: Componentes principales de los datos del estudio de agua y suelo.

Lo que interesa observar son los autovalores de los componentes extraídos, según los autovalores nos darán una medida de tolerancia para poder decidir con cuanta cantidad de componentes es recomendable quedarnos. Autovalores iguales o mayores a 1 indican que el componente logra explicar más varianza que una variable por sí sola.[8]

Por lo tanto, optaremos por las tres componentes que explican un total de 59,15% de la variabilidad.

Otro elemento que aporta a la delimitación de cantidad de factores/componentes a elegir, es el gráfico de sedimentación.

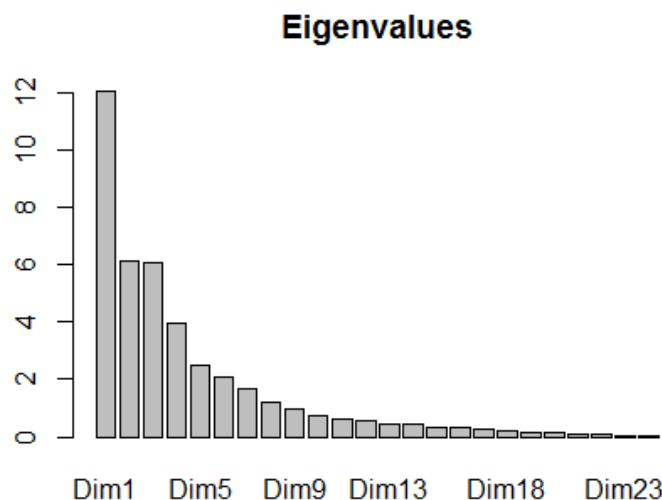


Figura 35: Sedimentación de los datos del estudio de las muestras del suelo y agua

El gráfico de sedimentación muestra la capacidad explicativa que va aportando cada componente a medida que se van incorporando al modelo

```

## $Dim.1
## $Dim.1$quanti
##                               correlation      p.value
## SurvivalfishhuNAred          0.8939288 1.757258e-09
## Algae fifty                  0.8749748 1.057064e-08
## Algae huNAred                0.8534569 5.879029e-08
## Nitrateveintiocho            0.8437688 1.167617e-07
## yielhuNAred                  0.8266714 3.526901e-07
## INAexgerminationhuNAred     0.7945016 2.109880e-06
## rootlengthhhuNAred           0.7909171 2.525660e-06
## Adultfolsomia                0.6957768 1.125547e-04
## AdultPlatynotres             0.6825272 1.707137e-04
## seedgerminationtwentyfive    0.5779226 2.481198e-03
## AdultOppiatwentyeighth       0.5224080 7.384993e-03
## Freshplantfourteen            0.4985424 1.119432e-02
## JuvenilsEnchytriaed           0.4966310 1.155925e-02
## Highplant                      0.4758915 1.619075e-02
## INAexgermianationseventyfive 0.4453221 2.569308e-02
## Algaetwentyfive               0.3835194 5.842152e-02
## Seed.germinationhuNAred       0.2891373 1.609676e-01
## JuvenilsOppiatwentytwo        0.2696421 1.924083e-01
## Avoid.respectcontrol           0.3352920 1.013261e-01
## Hachingfifty                  0.3887110 5.481389e-02
## Hatchingtweentyfive            0.4278500 3.288092e-02
## clorosisplant                 0.4707702 1.754226e-02
## Abnormalityfishtweentyfive    0.4756221 1.625966e-02
## Abnormalityfishfifty            0.5513854 4.276697e-03
## Daphniainmobilisationtwentyfive 0.6441359 5.111077e-04
## Daphniainmobilisationfifty     0.6818085 1.745101e-04
## willingplant                   0.6827172 1.697223e-04
## DaphniainmobilisationhuNAred   0.8456173 1.028044e-07
## AbnormalityfishhuNAred          0.8709759 1.487558e-08
##
## $Dim.1$quali
##              R2      p.value
## Zone 0.7455715 9.611866e-06
##
## $Dim.1$category
##      Estimate      p.value
## 5  5.314888 8.021151e-06
## 4 -2.511899 7.529092e-02
## 2 -3.134387 2.336674e-02
##
## $Dim.2
## $Dim.2$quanti
##                               correlation      p.value
## Malformationplant            0.8946145 1.636630e-09
## Seedfourteen                  0.8883582 3.077616e-09
## clorosisplant                 0.7636282 8.949776e-06
## AdultEnchytriaed              0.7130792 6.315571e-05
## willingplant                   0.6113361 1.167451e-03
## JuvenilsEnchytriaed            0.5201765 7.687466e-03
## INAexgerminationfifty          0.4713409 1.738724e-02

```

```

## Seed.germinationhuNAred          0.3504656 8.587394e-02
## rootlenghfty                  0.3433864 9.284526e-02
## yielhuNAred                   0.3281466 1.092794e-01
## AlgaehuNAred                  0.3097016 1.319252e-01
## AdultOppiattwelve             0.2861859 1.654782e-01
## Survivalfishfifty              0.2841638 1.686194e-01
## Survivalfishtwentyfive         0.2831900 1.701469e-01
## HatchinghuNAred                -0.2665797 1.977046e-01
## Nitrateveintiocho              -0.2824344 1.713387e-01
## Freshplantfourteen             -0.3155334 1.244274e-01
## Abnormalityfishfifty           -0.3672213 7.095043e-02
## Algaetwentyfive               -0.3928855 5.204076e-02
## Abnormalityfishtweentyfive     -0.4868441 1.358539e-02
## AdultOppiattwentyeighth        -0.5701185 2.925710e-03
## Hatchingtweentyfive            -0.6247023 8.430836e-04
##
## $Dim.2$quali
##           R2      p.value
## Zone 0.7046576 4.063203e-05
##
## $Dim.2$category
##   Estimate      p.value
## 2  3.303487 0.000280434
## 5  1.534859 0.132388079
## 4 -1.691013 0.095472229
## 1 -2.122369 0.032921000
##
##
## $Dim.3
## $Dim.3$quanti
##                               correlation      p.value
## JuvenilsOppiatwenttwenty       0.6946438 1.167346e-04
## Daphniainmobilisationtwentyfive 0.6566949 3.630572e-04
## Survivalfishtwentyfive         0.5658939 3.193391e-03
## AdultOppiattwentyeighth        0.4763760 1.606743e-02
## AdultEnchytriaed               0.4448540 2.586750e-02
## seedgerminationfifty          0.4444620 2.601434e-02
## JuvenilsEnchytriaed            0.4399261 2.776259e-02
## Juvenilsfolsom                 0.4395658 2.790540e-02
## AdultPlatynotres               0.3996410 4.778624e-02
## AdultOppiattwelve               0.3945415 5.097150e-02
## AbnormalityfishhuNAred          0.3714291 6.753510e-02
## Adultfolsomia                  0.3319320 1.050105e-01
## seedgerminationtwentyfive      0.3009530 1.437774e-01
## SurvivalfishhuNAred              -0.2859669 1.658164e-01
## Survivalfishfifty                -0.3320697 1.048576e-01
## Freshplantfourteen              -0.3813314 5.999593e-02
## rootlengtweentyfive             -0.4006815 4.715588e-02
## Avoid.respectcontrol            -0.4202234 3.648354e-02
## Algaefty                         -0.4278730 3.287049e-02
## Daphniainmobilisationfifty      -0.4556755 2.207090e-02
## HatchinghuNAred                  -0.5787963 2.435222e-03
## Highplant                         -0.5849838 2.129946e-03
## INAexgerminationfifty            -0.6572703 3.572800e-04
## rootlenghfty                      -0.8269727 3.462478e-07
##
## $Dim.3$quali
##           R2      p.value
## Zone 0.7624099 4.94308e-06

```

```

## 
## $Dim.3$category
##   Estimate      p.value
## 3  3.586366 3.706231e-05
## 2 -1.526345 1.317417e-01
## 1 -2.771237 3.419799e-03

```

Figura 36: Matriz de correlaciones estimada a partir de la solución factorial de los datos del estudio de las muestras de agua y suelo.

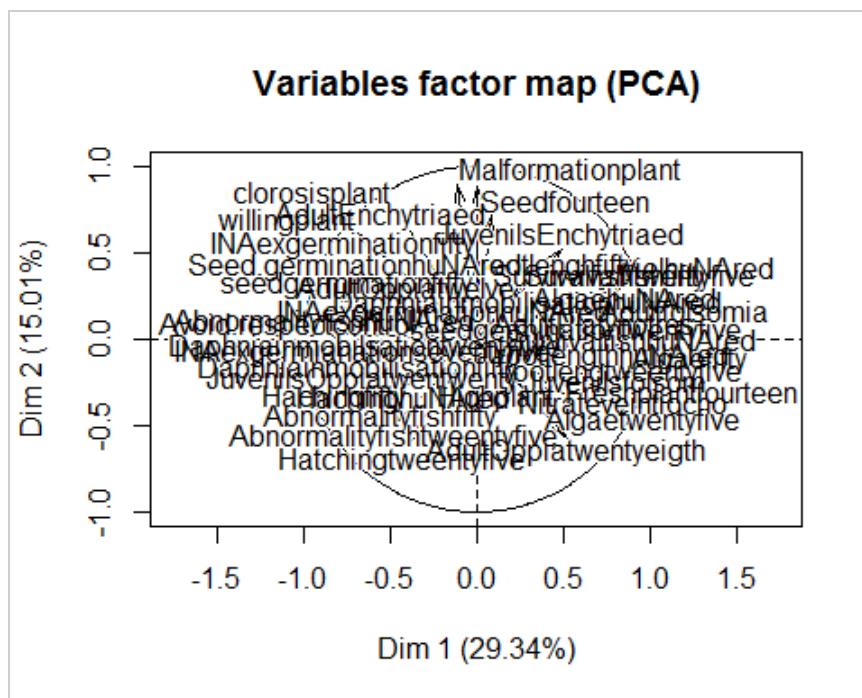
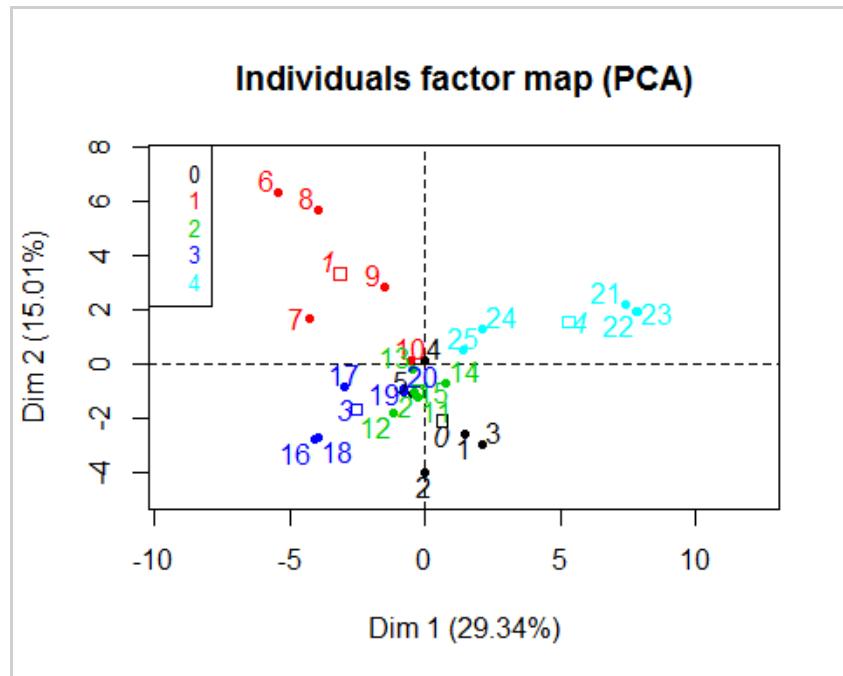


Figura 37: Componentes principales en dos dimensiones de los datos del estudio de las muestras de agua y suelo.

En cuanto a la interpretación de los componentes, observamos que PC1 tiene la mayor correlación positiva con las variables SurvivalfishhuNAred, Algaefifty, AlgaehuNAred, Nitrateveintiocho, yielhuNAred, INAexgerminationhuNAred, rootlengthhuNAred y una mayor correlación negativa con las variables DaphniainmobilisationhuNAred y AbnormalityfishhuNAred siendo los test que se realizaron del agua los que prevalecen en esta primera componente.

PC2 tiene una correlación positiva muy alta con las variables Malformationplant, Seedfourteen, Clorosisplant, AdulEnchytriaed y en esta componente predomina el test aplicado a la tierra.

También se puede identificar que la Zona 2 y la Zona 4 están relacionadas, y luego de analizar la estadística descriptiva se concluye que hay mayor polución en la zona 4 (Zona control) y en la zona 2 (Complejo Petroquímico Adyacente) y una menor contaminación en la zona 0 (Incineradora de Residuos Peligrosos de Constantí)

- Análisis de datos sobre las medidas recolectadas de los roedores

Se empezó obteniendo las componentes principales que se requiera

```
##          eigenvalue percentage of variance cumulative percentage of variance
## comp 1    3.7666611           34.242373                34.24237
## comp 2    2.1345989           19.405444                53.64782
## comp 3    1.4124455           12.840413                66.48823
## comp 4    1.1166159           10.151054                76.63928
## comp 5    0.7034334            6.394849                83.03413
```

Figura 38: Componentes principales de las medidas obtenidas de los roedores

Como resultado podemos ver los valores propios correspondientes a cada componente, la proporción de varianza total explicada y el porcentaje acumulado, por ello optaremos por las tres componentes que explican un total de 66.49% de la variabilidad.

Otro elemento que aporta a la delimitación de cantidad de factores/componentes a elegir, es el gráfico de sedimentación.

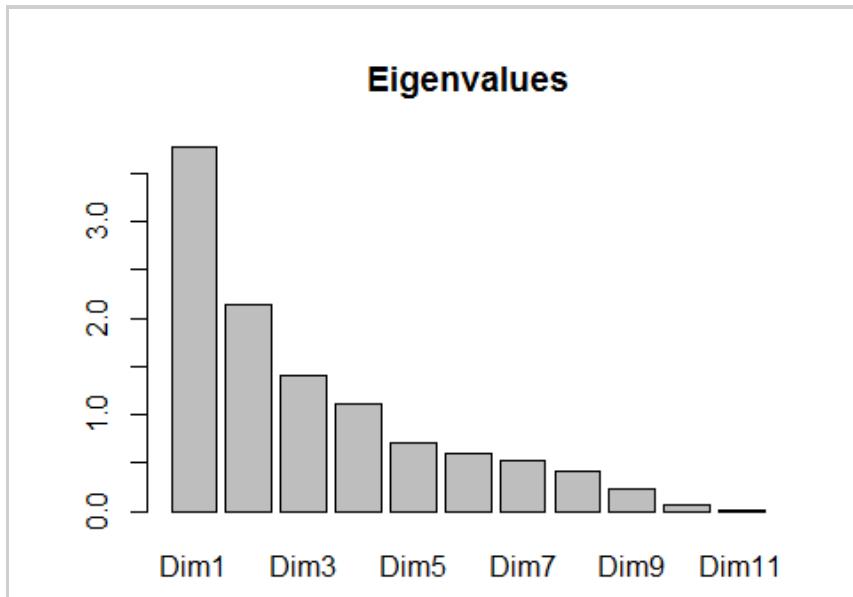


Figura 39: Sedimentación de las medidas obtenidas de los roedores

En la figura de sedimentación muestra la capacidad explicativa que va aportando cada componente a medida que se van incorporando al modelo

```

## $Dim.1
## $Dim.1$quanti
##           correlation      p.value
## BCI          0.8779809 1.527874e-30
## BW           0.8577548 9.544171e-28
## BCIApo       0.7506873 6.957540e-18
## Taillenght   0.5271042 6.741126e-08
## BCIMus        0.4110198 4.704386e-05
## Footlenght   0.3506748 6.101200e-04
## BL            0.2399313 2.124320e-02
## Spleen         -0.4319712 1.711022e-05
## Kidney         -0.6199431 4.412396e-11
## Liver          -0.6933053 1.873252e-14
##
## $Dim.1$quali
##           R2      p.value
## Specie 0.18924088 1.468653e-05
## Age     0.09405878 1.232966e-02
## Zone    0.12804106 1.694840e-02
## Sex     0.01848139 1.963120e-01
##
## $Dim.1$category
##           Estimate      p.value
## 1    1.0234399 1.468653e-05
## 3    0.8376747 3.056250e-03
## 1    0.6770489 9.434727e-02
## 4    0.2850816 1.737263e-01
## 2    0.2644059 1.963120e-01
## 1    -0.2644059 1.963120e-01
## 2    -0.3416220 4.227853e-02
## 3    -1.1257450 8.497288e-04
## 2    -1.0234399 1.468653e-05

```

```

## 
## 
## $Dim.2
## $Dim.2$quanti
##           correlation      p.value
## Footlenght  0.7399027 3.592148e-17
## Handlenght  0.7170387 9.047652e-16
## Taillenght  0.6502346 2.311165e-12
## Kidney      0.2581150 1.298608e-02
## BCIApo     -0.1777173 9.011000e-02
## BW          -0.3490502 6.493492e-04
## BL          -0.6277633 2.124735e-11
##
## $Dim.2$quali
##           R2      p.value
## Zone       0.30223601 2.249027e-06
## Specie    0.05743421 2.139762e-02
## Age        0.07142841 3.696451e-02
## Sex        0.01964182 1.827052e-01
##
## $Dim.2$category
##           Estimate      p.value
## 2  0.8893151 0.0004946657
## 1  0.7391371 0.0114135336
## 0  0.7572452 0.0142963359
## 1  0.4244436 0.0213976185
## 2  0.2051987 0.1827052161
## 1  -0.2051987 0.1827052161
## 2  -0.4244436 0.0213976185
## 3  -0.6910859 0.0099060190
## 4  -1.2122391 0.0013768646
##
##
## $Dim.3
## $Dim.3$quanti
##           correlation      p.value
## BCIMus     0.6895836 2.934667e-14
## Spleen     0.6174564 5.543176e-11
## BCI        0.4109051 4.729639e-05
## Liver      0.1564261 1.364728e-01
## Footlenght -0.1505976 1.518813e-01
## Handlenght -0.1556034 1.385729e-01
## Kidney     -0.1670429 1.114934e-01
## Taillenght -0.2016597 5.389759e-02
## BL          -0.4920098 6.326121e-07
##
## $Dim.3$quali
##           R2      p.value
## Specie    0.18316896 2.085148e-05
## Zone      0.14089375 9.639632e-03
## Age       0.05833105 6.894013e-02
##
## $Dim.3$category
##           Estimate      p.value
## 2  0.6165785 2.085148e-05
## 4  0.7164121 1.056503e-02
## 3  0.3251598 2.755693e-02
## 2  0.4388882 1.421689e-01
## 0  -0.4243399 1.379233e-01

```

```

## 1 -0.6433617 5.009094e-02
## 2 -0.2966882 3.369944e-02
## 1 -0.6165785 2.085148e-05

```

Figura 40: Matriz de correlaciones estimada a partir de la solución factorial de las medidas de los roedores

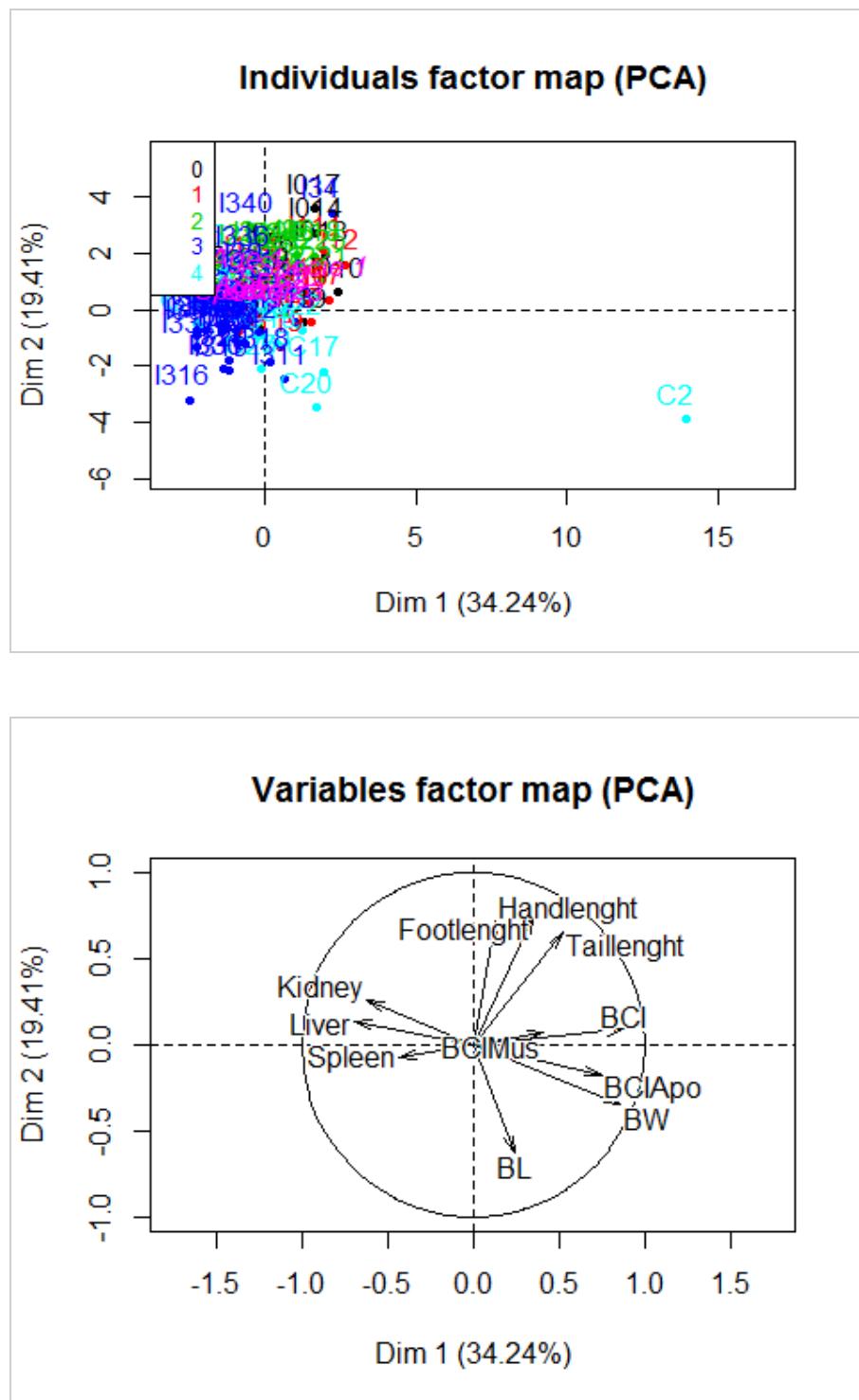


Figura 41: Componentes principales en dos dimensiones de las medidas de los roedores

En cuanto a la interpretación de los componentes, observamos que PC1 tiene la mayor correlación positiva con las variables BW, BCI, BCIApo, con lo cual sus valores positivos podrían asimilarse que los roedores han tenido unos niveles altos de peso.

PC2 tiene una correlación positiva muy alta con la variable Taillenght, Handlenght y Footlenght, con lo cual, destaca las medidas de longitud del roedor.

- Análisis de los exámenes hematológicos de los roedores

Se empezó extrayendo las componentes para identificar cuantas componentes se requiere.

	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	3.672746	22.954665	22.95467
## comp 2	2.517858	15.736610	38.69127
## comp 3	2.060990	12.881185	51.57246
## comp 4	1.807496	11.296850	62.86931
## comp 5	1.248461	7.802884	70.67219

Figura 42: Componentes principales de los exámenes hematológicos de los roedores.

Lo que interesa es prestar atención a los autovalores de los componentes extraídos, estos valores nos ayudaran a decidir con cuanta cantidad de componentes es recomendable quedarnos. Autovalores iguales o mayores a 1 indican que el componente logra explicar más varianza que una variable por sí sola, por lo tanto, optaremos por las tres componentes que explican un total de 51.57% de la variabilidad.

Otro elemento que aporta a la delimitación de cantidad de factores/componentes a elegir, es el gráfico de sedimentación.

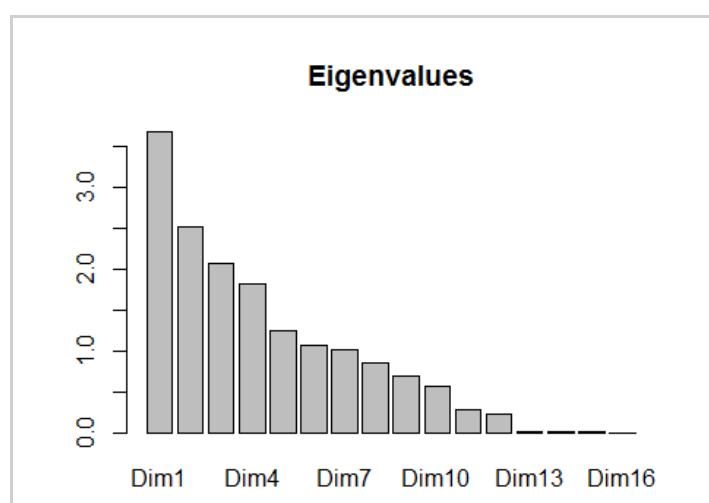


Figura 43: Sedimentación de los exámenes hematológicos

En la figura de sedimentación muestra la capacidad explicativa que va aportando cada componente a medida que se van incorporando al modelo.

```

## $Dim.1
## $Dim.1$quanti
##           correlation      p.value
## HGB          0.8355359 7.050739e-36
## HCT          0.8140215 1.049750e-32
## RBC          0.7994201 8.889515e-31
## WBC          0.7464869 6.176590e-25
## LYM          0.7037559 3.483553e-21
## GRA          0.5632177 1.700850e-12
## MCHC         0.3541960 2.881283e-05
## MID          0.3229241 1.499978e-04
## MCH          0.1406896 1.062587e-01
## Age          0.1261859 1.478123e-01
## Plaletes     -0.1349450 1.214626e-01
## MCV          -0.1498395 8.516846e-02
##
##
## $Dim.2
## $Dim.2$quanti
##           correlation      p.value
## Plaletes     0.6564931 9.588029e-18
## Neutrophilis 0.6337074 2.688462e-16
## Basophils    0.6051818 1.204934e-14
## GRA          0.4397480 1.184262e-07
## Monocytes    0.2383567 5.729832e-03
## MID          0.2175834 1.187539e-02
## Eosinophils   0.1764485 4.218755e-02
## WBC          0.1350790 1.210897e-01
## Specie       0.1260567 1.482312e-01
## HGB          -0.1964870 2.340532e-02
## Sex          -0.2084093 1.607089e-02
## HCT          -0.2253278 9.115193e-03
## MCH          -0.4250374 3.406659e-07
## MCV          -0.4743492 8.031005e-09
## Lymphocytes   -0.6853614 9.078693e-20
##
## $Dim.2$quali
##           R2      p.value
## Zone 0.0564011 0.1122187
##
## $Dim.2$category
##           Estimate      p.value
## 3  0.4286593 0.07866864
## 1 -0.6468167 0.03115854
##
##
## $Dim.3
## $Dim.3$quanti
##           correlation      p.value

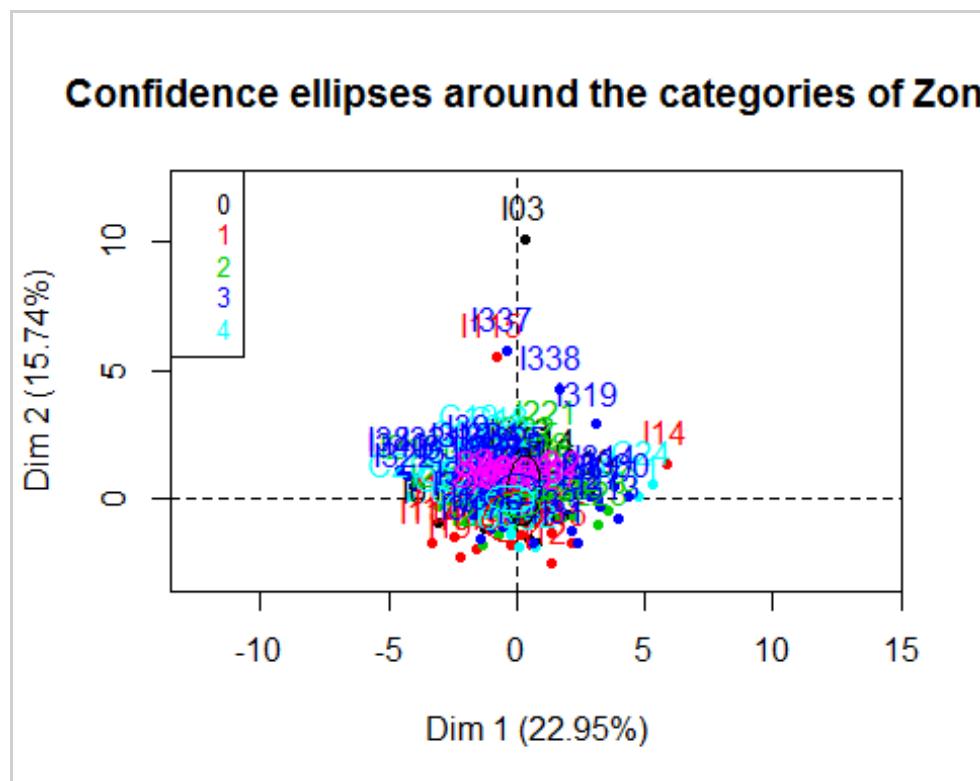
```

```

## MCHC          0.6364115 1.836496e-16
## HGB           0.4990102 9.750104e-10
## MCH           0.4664405 1.525557e-08
## Neutrophilis 0.2781103 1.190113e-03
## HCT            0.2568328 2.843143e-03
## RBC            0.2356954 6.312291e-03
## Monocytes     0.2129757 1.384435e-02
## Specie        -0.1735875 4.569398e-02
## Lymphocytes   -0.3240943 1.414601e-04
## GRA            -0.3830646 5.343239e-06
## WBC            -0.5738820 5.165633e-13
## LYM            -0.5831165 1.776854e-13
##
## $Dim.3$quali
##             R2      p.value
## Zone 0.2512266 1.551948e-07
##
## $Dim.3$category
##             Estimate      p.value
## 0  1.0768737 2.463714e-05
## 2  0.5080150 9.888137e-03
## 4 -0.6693254 3.243761e-02
## 3 -0.8581456 1.194270e-04

```

Figura 44: Matriz de correlaciones estimada a partir de la solución factorial de los exámenes hematológicos



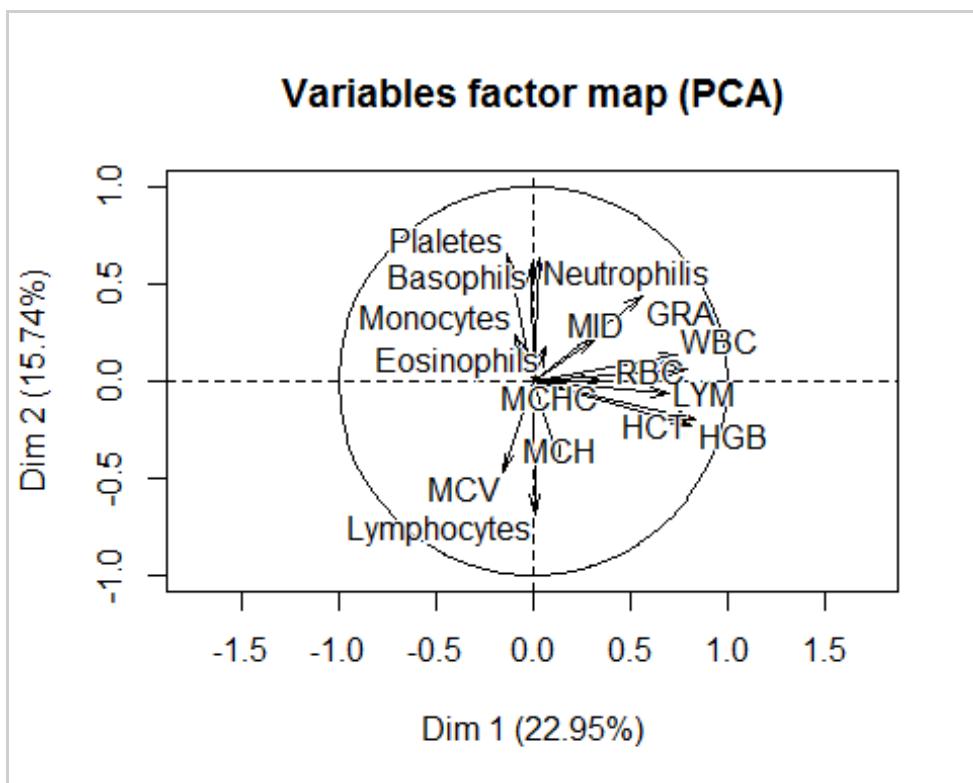


Figura 45: Componentes principales en dos dimensiones de los exámenes hematológicos

En cuanto a la interpretación de los componentes, observamos que PC1 tiene la mayor correlación positiva con las variables HCG, HCT, RBC, WBC y LYM, con lo cual sus valores positivos podrían asimilarse que han tenido un alto contenido de glóbulos rojos.

PC2 tiene una correlación positiva muy alta con las variables Plaletes, Neutrophils, Basophils, lo contrario del PC1 destaca el análisis de los glóbulos blancos.

- Análisis de los exámenes hispatológicos

Iniciaremos explorando la cantidad de componentes a extraer y para ello, solicitamos tantos componentes como variables involucradas en el análisis

```
##          eigenvalue percentage of variance cumulative percentage of variance
##  comp 1   1.1737729      29.34432                29.34432
##  comp 2   1.0541891      26.35473                55.69905
##  comp 3   0.9525303      23.81326                79.51231
##  comp 4   0.8195078      20.48769                100.00000
```

Figura 46: Componentes principales de los exámenes hispatológicos

Como podemos ver con tres componentes podemos explicar un 79,51% de la variabilidad.

Otro elemento que aporta a la delimitación de cantidad de factores/componentes a elegir, es el gráfico de sedimentación.

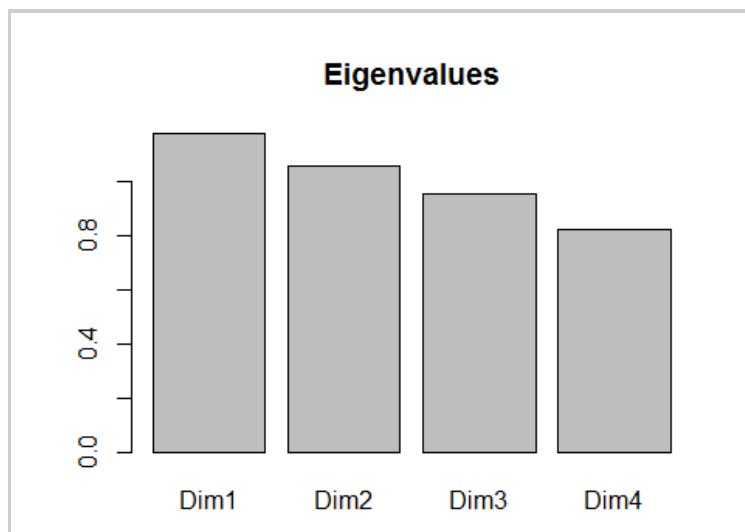


Figura 47: Sedimentación de los exámenes hispatológicos

En la figura de sedimentación muestra la capacidad explicativa que va aportando cada componente a medida que se van incorporando al modelo.

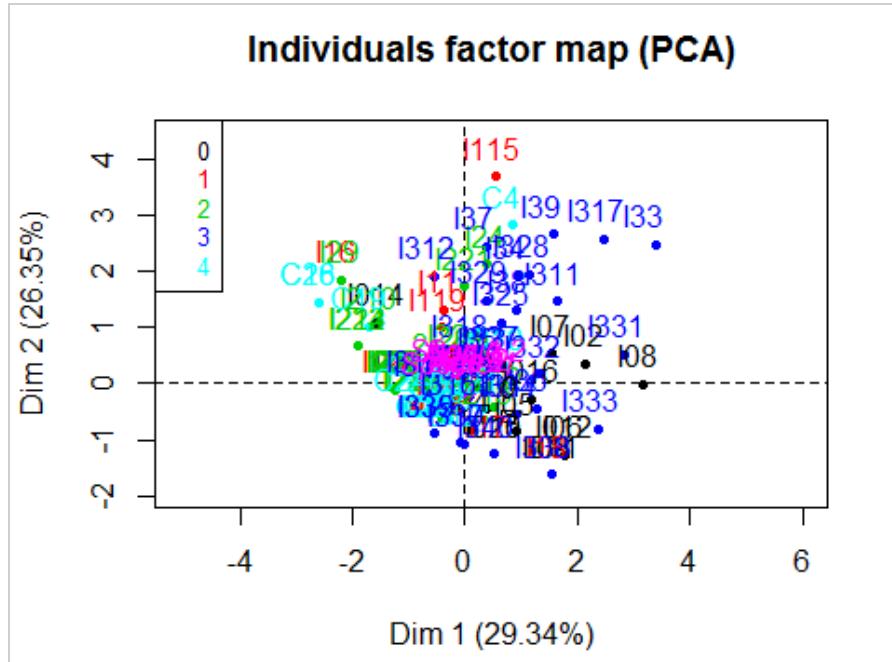
```
## $Dim.1
## $Dim.1$quanti
##           correlation      p.value
## Liverh    0.6877842 5.991118e-20
## Kidneyh   0.4586329 2.829338e-08
## Lungh     0.4267253 3.025328e-07
## Spleenh   -0.5552361 4.037351e-12
##
## $Dim.1$quali
##           R2      p.value
## Zone 0.3138014 7.189959e-10
##
## $Dim.1$category
##           Estimate      p.value
## 3  0.6111995 3.833186e-05
## 0  0.7990455 5.214865e-04
## 4 -0.4208267 1.421614e-02
## 2 -0.7787068 1.164409e-05
##
##
## $Dim.2
## $Dim.2$quanti
##           correlation      p.value
## Lungh    0.6807906 1.967251e-19
## Spleenh   0.5981396 2.908609e-14
## Kidneyh  0.4272806 2.909041e-07
```

```

## Liverh   -0.2244407 9.399654e-03
##
## $Dim.2$quali
##               R2      p.value
## Zone 0.08953591 0.01662857
##
## $Dim.2$category
##     Estimate      p.value
## 3  0.3612071 0.024354597
## 4 -0.1756259 0.192648589
## 0 -0.5151648 0.006739281
##
## $Dim.3
## $Dim.3$quanti
##     correlation      p.value
## Liverh    0.4356472 1.597928e-07
## Lungh     0.4000957 1.829320e-06
## Spleenh    0.2363269 6.169516e-03
## Kidneyh   -0.7394693 2.863801e-24
##
## $Dim.3$quali
##               R2      p.value
## Zone 0.05382522 0.1288313
##
## $Dim.3$category
##     Estimate      p.value
## 4 -0.3763558 0.02092184

```

Figura 48: Matriz de correlaciones estimada a partir de la solución factorial de los exámenes hispatológicos



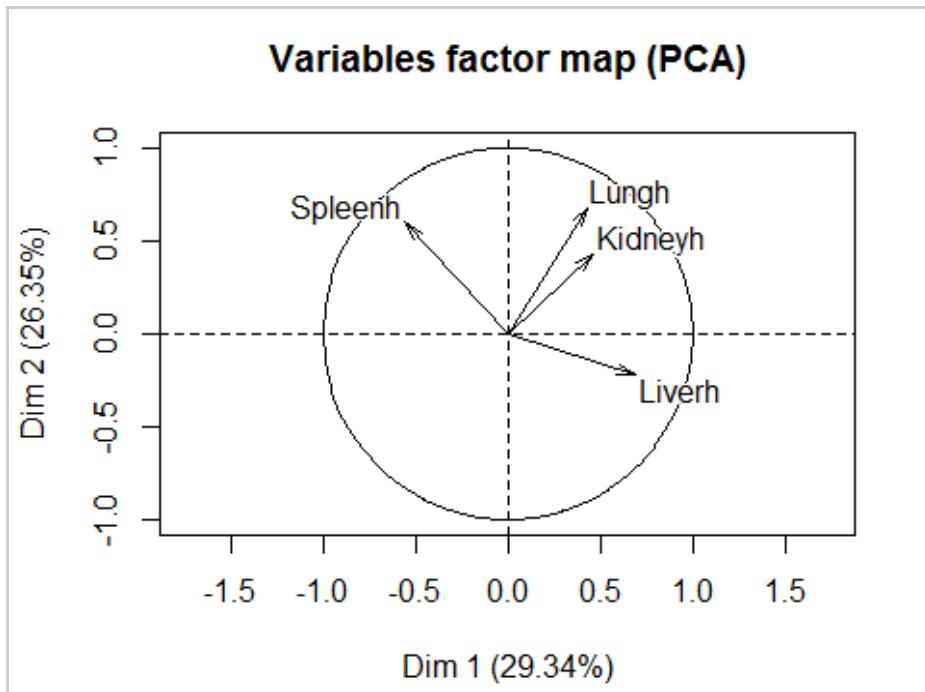


Figura 49: Componentes principales en dos dimensiones de los exámenes hispatológicos.

En cuanto a la interpretación de los componentes, observamos que PC1 tiene la mayor correlación positiva con la variable Liverh.

PC2 tiene una correlación positiva con las variables Lungh y Spleenhh.

2.5 Análisis factorial

El análisis factorial está relacionado con los componentes principales, pero existen ciertas diferencias. En primer lugar, los componentes principales se construyen para explicar las varianzas, mientras que los factores se construyen para explicar las covarianzas o correlaciones entre las variables. En segundo lugar, componentes principales es una herramienta descriptiva, mientras que el análisis factorial presupone un modelo estadístico formal de generación de la muestra dada.[1]

El modelo del Análisis Factorial viene dado habitualmente por las ecuaciones:

$$\begin{aligned}
 X_1 &= a_{11} F_1 + a_{12} F_2 + \dots + a_{1k} F_k + u_1 \\
 X_2 &= a_{21} F_1 + a_{22} F_2 + \dots + a_{2k} F_k + u_2 \\
 &\dots \\
 X_p &= a_{p1} F_1 + a_{p2} F_2 + \dots + a_{pk} F_k + u_p
 \end{aligned}$$

Donde, (F_1, F_2, \dots, F_k) ($k < p$) son los Factores Comunes , (u_1, u_2, \dots, u_p) los Factores únicos o específicos , y los Coeficientes (a_{ij}) { $i=1, \dots, p$; $j=1, \dots, k$ } las Cargas factoriales [9]

- Análisis de las medidas de los roedores

Comenzaremos con reducir al mínimo de factores de las variables correspondientes a los datos.

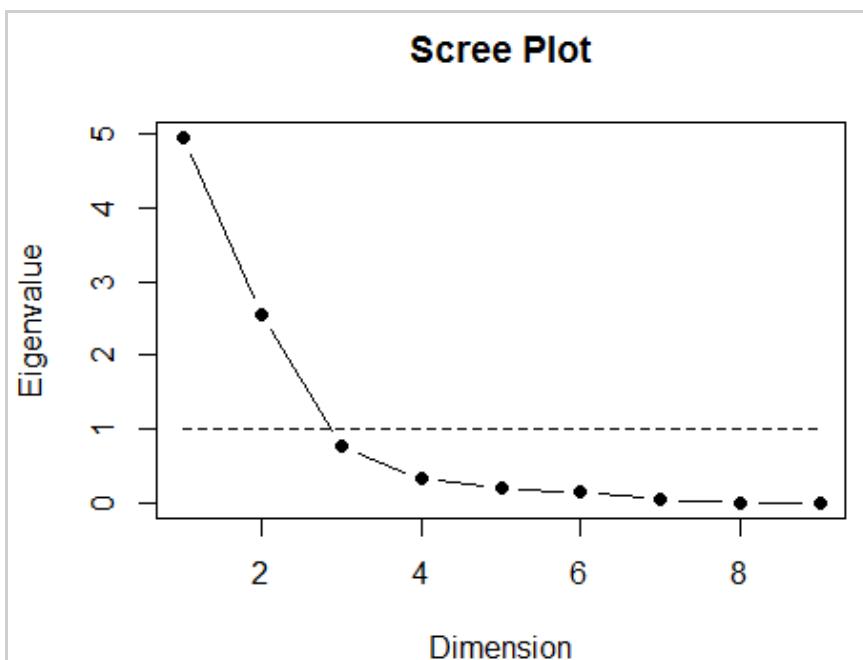


Figura 50: Gráfico lineal del número de factores en el Análisis Factorial de las medidas de los roedores

Al observar la figura 50 se determina que la solución con tres factores es la más apropiada para el análisis.

Ejecutamos el modelo con el número de factores ideal

```
modelo1  
##  
## Call:  
## factanal(x = Tabla1[, 5:13], factors = 3, scores = "regression",  
## rotation = "varimax")  
##  
## Uniquenesses:  
##          BW          BL         BCI        Liver       Kidney      Spleen  
##          0.162       0.005       0.078       0.654       0.668       0.751  
## Taillenght Hand lenght Foot lenght  
##          0.072       0.655       0.480  
##  
## Loadings:  
##           Factor1 Factor2 Factor3  
## BW          0.872          0.271
```

```

## BL          0.181  -0.314   0.930
## BCI         0.910   0.249  -0.179
## Liver      -0.542  -0.112  -0.198
## Kidney     -0.547           -0.176
## Spleen     -0.217  -0.243  -0.379
## Taillength 0.226   0.911   0.217
## Handlength 0.553   0.553  -0.195
## Footlength 0.114   0.712
##
##                  Factor1 Factor2 Factor3
## SS loadings     2.327   1.877   1.269
## Proportion Var  0.259   0.209   0.141
## Cumulative Var 0.259   0.467   0.608
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 17.32 on 12 degrees of freedom.
## The p-value is 0.138

```

Figura 51: Modelo factorial con tres factores sobre las medidas obtenidas de los roedores

Respecto a la varianza explicada hay tres factores extraídos que explican el 60,8% de la variabilidad. Se observa que el factor 1 explica un 25,4% de la varianza, el factor 2 explica un 20,9% y el factor 3 explica 14,1% de la varianza.

Mostramos los vectores y factores característicos

```

##             Factor1    Factor2    Factor3
## I02 -0.22045316  0.20486677  1.208187601
## I03  0.07038436 -0.06126991 -0.002001254
## I04 -0.42714482  1.18959570 -0.826605460
## I05 -0.74382436  0.31599227 -0.839256414
## I06 -0.02853293  1.10650990 -0.636815926
## I07  0.06047017 -0.36891051 -0.399834425

```

Figura 52: Vectores y factores de los datos sobre las medidas obtenidas de los roedores

En la siguiente gráfica se visualiza la dependencia de cada variable de los dos factores más significativos.

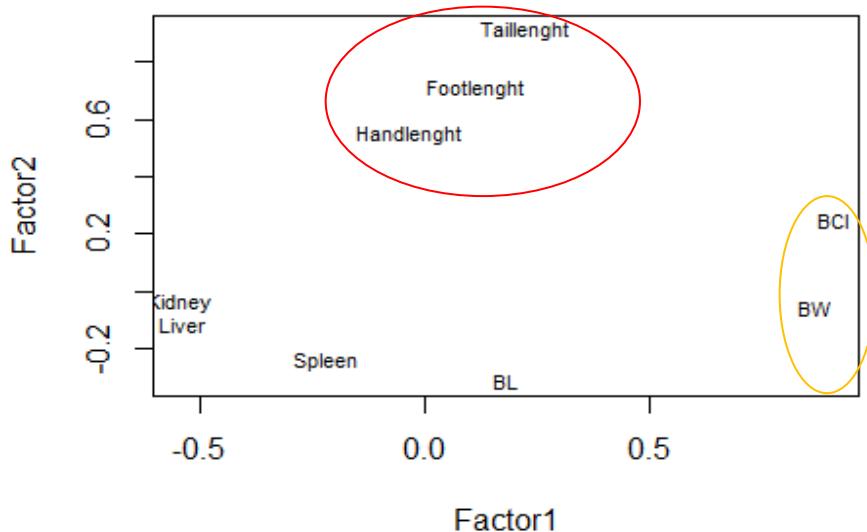


Figura 53: Representación de variables en factores para Análisis Factorial de los datos sobre las medidas de los roedores

En la figura 53 se puede ver que las variables BW y BCI tienen saturación alta en el factor 1, las variables Taillenght, Footlength y Handlength tiene saturación alta en el factor 2 y la variable BL tiene saturación alta en el factor 3, también lo corroboramos revisando los valores en la figura 51.

- **Análisis de los datos obtenidos sobre los análisis hematológicos realizado a los roedores**

Comenzaremos con reducir al mínimo de factores de las variables correspondientes a los datos.

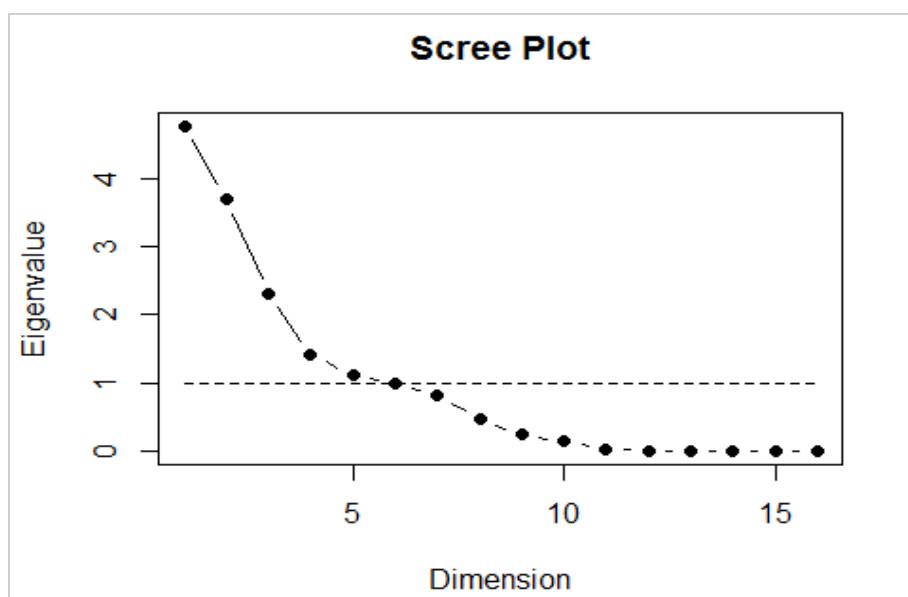


Figura 54: Gráfico lineal del número de factores en el Análisis Factorial sobre as medidas de los roedores

Al observar la figura 54 se obtiene que la solución con cinco factores resulta la más apropiada y por lo tanto se ejecutar el modelo con el número de factores ideal.

```

modelo2

##
## Call:
## factanal(x = tabla2.[, 5:20], factors = 5, scores = "regression",
## rotation = "varimax")
##
## Uniquenesses:
##          WBC        LYM        MID        GRA        RBC
## 0.005    0.089    0.862    0.354    0.005
##          HGB        HCT        MCV        MCH        MCHC
## 0.005    0.005    0.005    0.005    0.005
## Neutrophilis  Lymphocytes  Eosinophils  Monocytes  Basophils
## 0.398    0.005    0.954    0.696    0.811
##          Plaletes
## 0.768
##
## Loadings:
##          Factor1 Factor2 Factor3 Factor4 Factor5
## WBC      0.158   0.983
## LYM      0.169   0.915  -0.206
## MID      0.277   0.208
## GRA      0.771   0.215
## RBC      0.883   0.175   -0.426
## HGB      0.882   0.178
## HCT      0.960   0.233   0.109
## MCV     -0.126   -0.143   0.961  -0.185
## MCH      0.809   0.578
## MCHC     0.222   0.969
## Neutrophilis      0.772
## Lymphocytes      -0.997
## Eosinophils       0.194
## Monocytes         0.544
## Basophils        -0.147   0.311  -0.176   0.180
## Plaletes         -0.169   0.424  -0.150
##
##          Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings  2.674   2.609   2.375   1.851   1.527
## Proportion Var 0.167   0.163   0.148   0.116   0.095
## Cumulative Var 0.167   0.330   0.479   0.594   0.690
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 1729.12 on 50 degrees of freedom.
## The p-value is 0

```

Figura 55: Modelo factorial con cinco factores de las medidas de los roedores

Respecto a la varianza explicada hay cinco factores extraídos que explican el 69,0% de la variabilidad. Se observa que el factor 1 y 2 explican un valor casi similar 16,7 % y 16.3% respectivamente, el factor 3 explica 14.8%, el factor 4 explica 11,6% y el factor 5 explica 9,5% de la varianza.

También presentaremos los vectores y factores característicos

	Factor1	Factor2	Factor3	Factor4	Factor5
## I02	1.3051511	-1.0952622	-0.2564927	-1.76338347	-0.27951837
## I03	-1.1062919	0.5753207	2.4556234	-2.04499725	1.79744638
## I05	1.1162844	-0.5393243	-1.7489305	-0.04062691	-1.05715453
## I07	0.5766890	-0.9072280	-0.2279419	0.35104380	0.48422358
## I08	0.6741668	-0.6569392	1.8518054	-0.78856597	0.32330457
## I09	0.6773456	-1.0349521	-0.5708333	-0.54579470	-0.03017594

Figura 56: Vectores y factores de los datos sobre los exámenes hematológicos de los roedores

Mostraremos como depende cada variable de los dos factores más significativos.

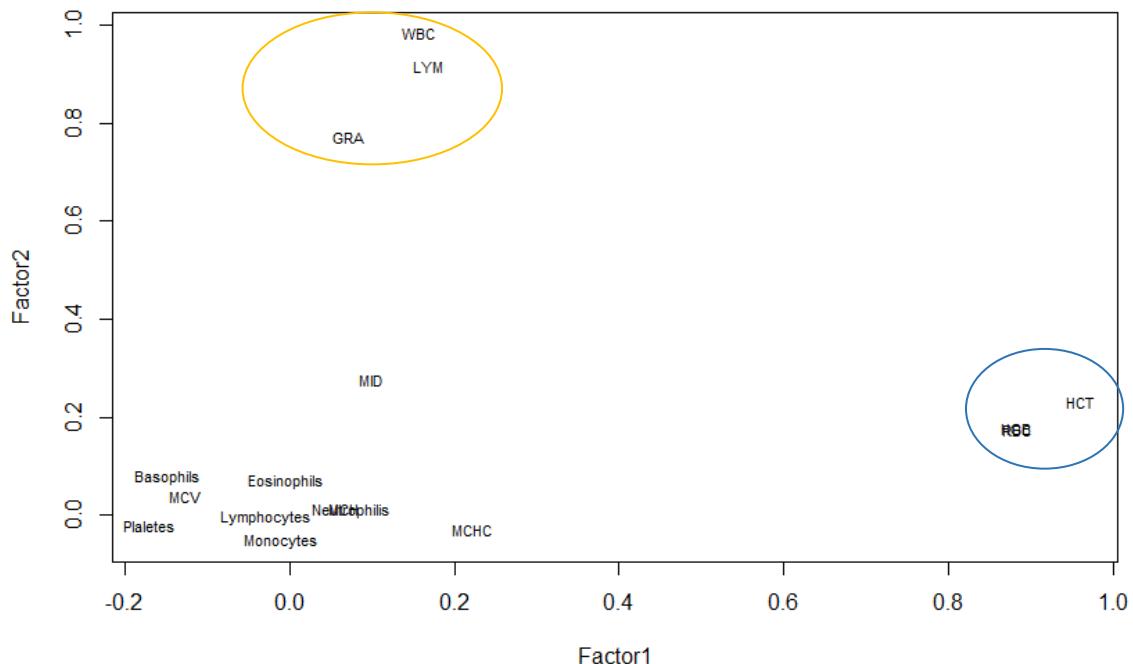


Figura 57: Representación de variables en factores para Análisis Factorial de los exámenes hematológicos

En la figura 57 y en base a la figura 56 se puede ver que las variables HCT, HGB Y RBC tienen saturación alta en el factor 1, las variables WBC, LYM y GRA tienen una saturación alta en el factor 2, la variable Lymphocytes tiene una saturación alta en el factor 3, la variable MCV y MCH tienen una saturación alta en el factor 4 y la variable MCHC tiene una saturación alta en el factor 5.

- Análisis de los exámenes hispatológicos

Empezaremos graficando el modelo factorial

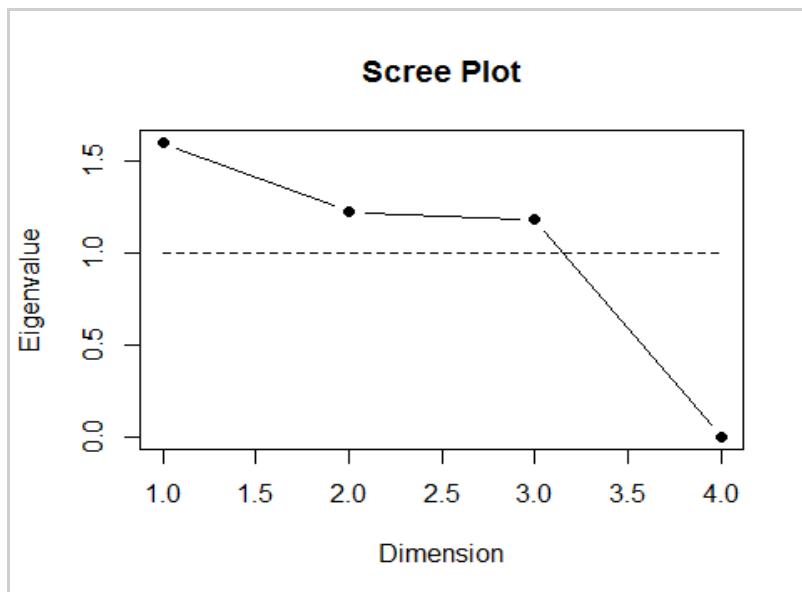


Figura 58: Representación de variables en factores para Análisis Factorial de los exámenes hispatológicos

Al observar la figura 58 se obtiene que la solución con un factor resulta la más apropiada.

```
modelo3
##
## Call:
## factanal(x = tabla3.[, 5:8], factors = 1, scores = "regression",
## rotation = "varimax")
##
## Uniquenesses:
## Liverh Kidneyh Spleenh Lungh
## 0.652 0.992 0.962 0.990
##
## Loadings:
##             Factor1
## Liverh    0.590
## Kidneyh
## Spleenh -0.196
## Lungh
##
##             Factor1
## SS loadings   0.405
## Proportion Var 0.101
##
## Test of the hypothesis that 1 factor is sufficient.
```

```
## The chi square statistic is 0.74 on 2 degrees of freedom.  
## The p-value is 0.689
```

Figura 59: Modelo factorial con un factor de los exámenes hispatológicos

Respecto a la varianza explicada hay un factor extraído que explican el 10.1% de la variabilidad. El modelo no es bueno ya que la varianza explicada por un factor es muy baja.

2.6 Análisis de conglomerado

El análisis de agrupamiento es un término genérico para una amplia gama de métodos numéricos para examinar datos multivariados con el fin de descubrir o descubrir grupos o grupos de observaciones que son homogéneos y separados de otros grupos. [10]

Hay diferentes técnicas de agrupamiento.[10]

- **Algoritmos jerárquicos acumulativos:** En una clasificación jerárquica, los datos no se dividen en un número determinado de clases o clusters en un solo paso. En cambio, la clasificación consiste en una serie de particiones que pueden correr desde un solo "cluster" que contiene todos los individuos, a n agrupaciones que contienen un solo individuo [10]
- **Métodos tipo K-means:** Esté método trata de encontrar la mejor partición posible del conjunto de objetos para un número dado de clase k. Con este método a diferencia de la clasificación jerárquica es necesario establecer a priori el número de clases que debe tener la solución buscada. [11]
- **Métodos de máxima verosimilitud de clasificación.**

Nosotros aplicaremos el primer método el algoritmo jerárquico acumulativo

- Análisis de los datos del estudio de muestras de agua y suelo

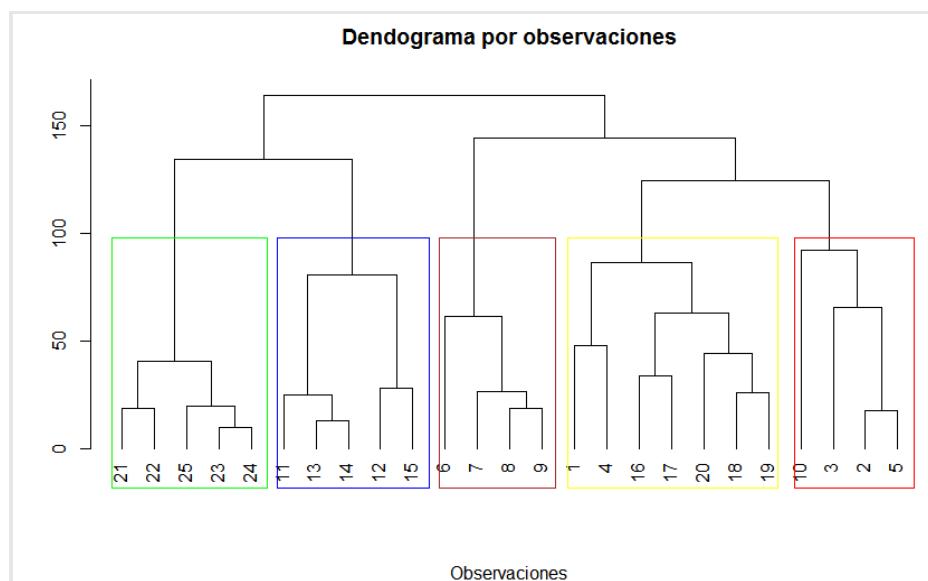


Figura 60: Dendrograma por observaciones del estudio en muestras de agua y suelo.

Realizando 5 conglomerados, uno contendría las observaciones 21-25, otro las observaciones 11-15, otro las observaciones 6-9, otro las observaciones 1,4,16-20 y el otro las observaciones restantes 2,3,5,10.

Las observaciones más similares y que tienen una distancia menor son 23,24, ,13,14 y las observaciones más distintas al resto es la 10 que es la última observación que se incorpora al grupo 5.

Estudiemos como se han agrupado las 25 observaciones en estudio directamente sobre los datos

```
table(cluster)
```

```
## cluster
## 1 2 3 4 5
## 7 4 4 5 5
```

Figura 61: Observaciones por cluster de los datos del estudio de las muestras de agua y suelo

El primer conglomerado contiene 7 observaciones, el segundo y tercero conglomerado están formados por 4 observaciones cada uno, el cuarto y quinto conglomerado incluye 5 observaciones cada uno.

Revisamos cómo se forman los clústeres directamente sobre los datos.

```
##      Zone AdultEnchytriaed JuvenilsEnchytriaed cluster
## 1      1                      80                  37       1
```

## 4	1	80	90	1
## 16	4	80	285	1
## 17	4	80	393	1
## 18	4	80	270	1
## 19	4	80	243	1

Figura 62: Formación de los cluster sobre los datos del examen de las muestras de agua y suelo

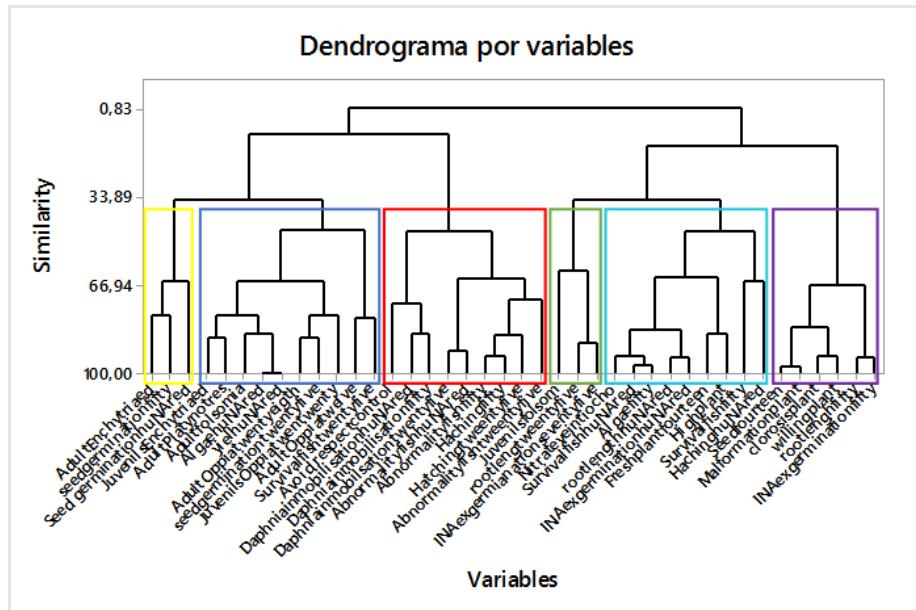


Figura 63: Dendrograma por variables de los datos del estudio de las muestras de agua y suelo

Realizando 6 conglomerados, uno contendría 3 variables, otro constaría de 11 variables, otro tendría 9 variables, otro estaría conformado de 3 variables, otro incluiría 9 variables y la ultima 6 variables.

Las variables más similares y que tienen una distancia menor son SurvivalfishhuNAred – Algaefifty, Seedfourteen – Malformationplant y las variables más distintas al resto es la Juvenilsfolsom, Survivalfishfifty, HachinghuNAred.

Primer conglomerado:

- AdultEnchytriaed
- Seedgerminationfifty
- SeedgerminationhuNAred

Segundo conglomerado:

- JuvenilsEnchytriaed
- AdultPlatynotres
- Adultfolsom
- AlgaehuNAred
- yielhuNAred
- AdultOppiattwelve

- Survivalfishtwentyfive
- AdultOppiatwentyeighth
- Seedgerminationtwentyfive
- JuvenilsOppiatwentytwo
- Algaetwentyfive

Tercer conglomerado:

- Avoidrespectcontrol
- DaphniainmobilisationhuNAred
- Daphniainmobilisationfifty
- Daphniainmobilisationtwentyfive
- AbnormalityfishhuNAred
- Abnormalityfishfifty
- Hachingfifty
- Hatchingtweentyfive
- Abnormalityfishabweentyfive

Cuarto Conglomerado:

- Juvenilsfolsom
- Rootlengtweentyfive
- INAexgermianationseventyfive

Quinto conglomerado:

- Nitrateveintiocho
- SurvivalfishhuNAred
- Algaefifty
- rootlengthhuNAred
- INAexgerminationhuNAred
- Freshplantfourteen
- Highplant
- Survivalfishfifty
- HachinghuNAred

Sexto conglomerado:

- Seedfourteen
- Malformationplant
- Clorosisplant
- Willingplant
- rootlenghfifty
- INAexgerminationfifty

- Análisis de los datos obtenidos sobre las medidas obtenidas de los animales

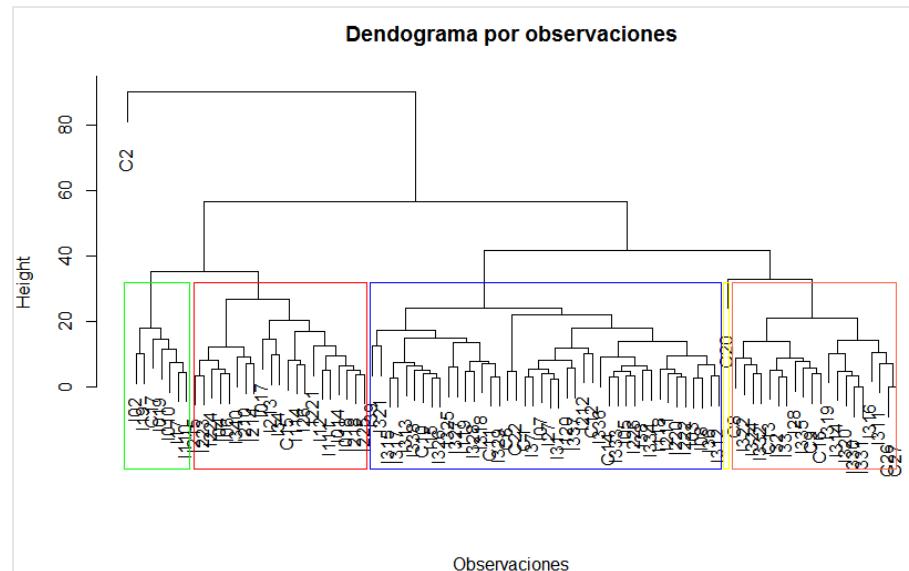


Figura 64: Dendrograma por observaciones de los datos sobre las medidas de los roedores

Podemos considerar que en general los datos son bastante homogéneos, ya que la mayoría de observaciones quedan a una distancia inferior a 20.

Estudiemos como se han agrupado las 92 observaciones en estudio directamente sobre los datos

```
## cluster1
##  1  2  3  4  5
##  7 42 21 21  1
```

Figura 65: Observaciones por cluster de los datos sobre las medidas de los roedores

Realizando 5 conglomerados, el primer conglomerado contendría 7 observaciones, el segundo tendría 42 observaciones, el tercero y cuarto conglomerado estarían conformados por 21observaciones cada uno y el quinto conglomerado con una sola observación.

Examinamos cómo se forman los clústeres directamente sobre los datos.

```
##          BL  BCI Liver Kidney Spleen cluster1
## I02      83.0 0.96  7.41   0.71   0.28      1
## I010     85.0 1.02  3.90   0.70   0.12      1
## I019     93.5 0.97  6.53   0.65   0.11      1
## I17      88.0 1.01  4.92   0.66   0.11      1
## I19      90.0 0.99  4.61   0.73   0.17      1
## I110     86.0 1.01  6.29   0.63   0.11      1
```

Figura 66: Formación de los cluster sobre los datos de las medidas de los roedores

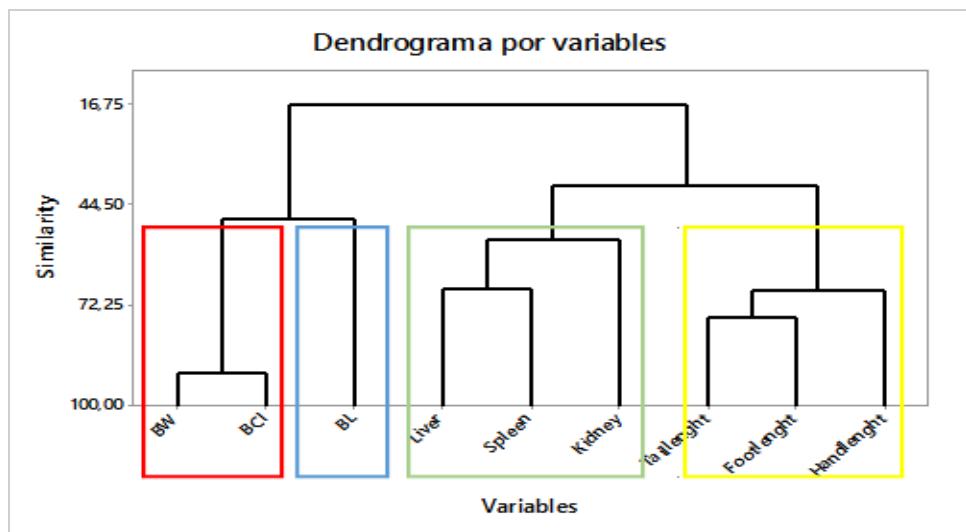


Figura 67: Dendrograma por variables de los datos sobre las medidas de los roedores

Realizando 4 conglomerados, un primer conglomerado contendría las variables BW y BCI, un segundo conglomerado abarcaría las variables Taillenght, Footlenght, Handlenght, un tercer conglomerado incluiría la variable BL y un cuarto conglomerado estaría conformado por las variables restantes Liver, Kidney, Spleen.

Las variables más similares y que tienen una distancia menor son BW y BCI y la variable más distinta al resto es BL.

- **Análisis de los datos obtenidos sobre los exámenes hematológicos realizado a los roedores**

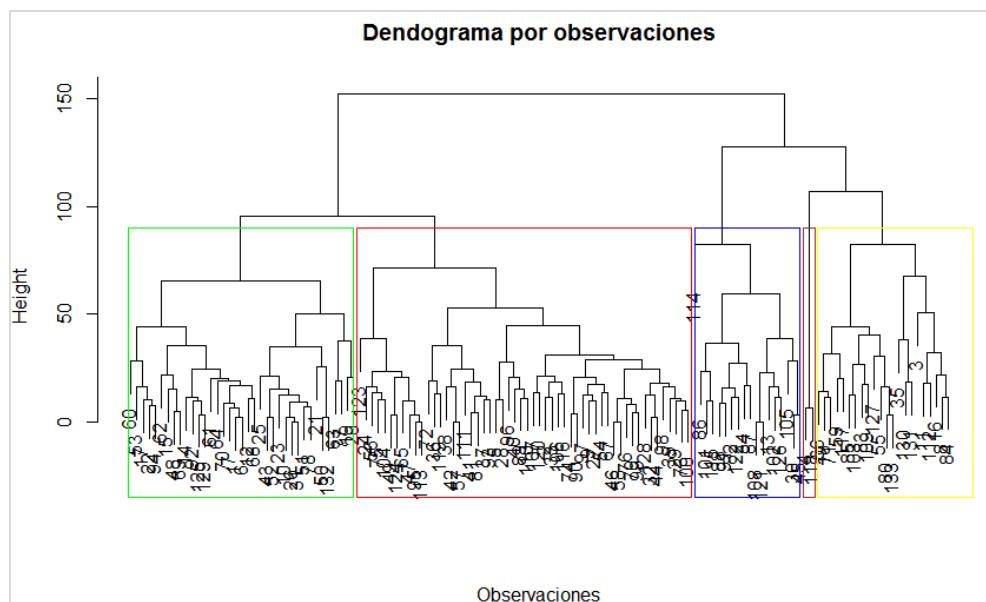


Figura 68: Dendrograma por observaciones de los exámenes hematológicos

Podemos considerar que en general los datos son bastante homogéneos, sobre todo en los dos primeros conglomerados ya que la mayoría de observaciones quedan a una distancia inferior a 30.

La observación más distinta al resto es la 114.

Revisamos como se han agrupado las 133 observaciones directamente sobre los datos

```
## cluster2
## 1 2 3 4 5
## 37 54 22 2 18
```

Figura 69: Observaciones por cluster de los datos de los exámenes hematológicos

Realizando 5 conglomerados, el primero contendría 37 observaciones, el segundo tendría 54 observaciones, el tercero incluiría 22 observaciones, el cuarto constaría de 2 observaciones y el quinto estaría conformado por 18 observaciones, ahora agruparemos las observaciones asociadas al clúster formado.

```
##    Age WBC  LYM  MID  GRA cluster2
## 1    3 1.70 0.82 0.01 0.87      1
## 7    2 3.34 2.48 0.08 0.79      1
## 10   3 2.44 1.53 0.13 0.78      1
## 12   3 3.61 2.20 0.17 1.23      1
## 14   2 5.18 3.69 0.64 0.85      1
```

Figura 70: Formación de los cluster sobre los datos de los exámenes hematológicos

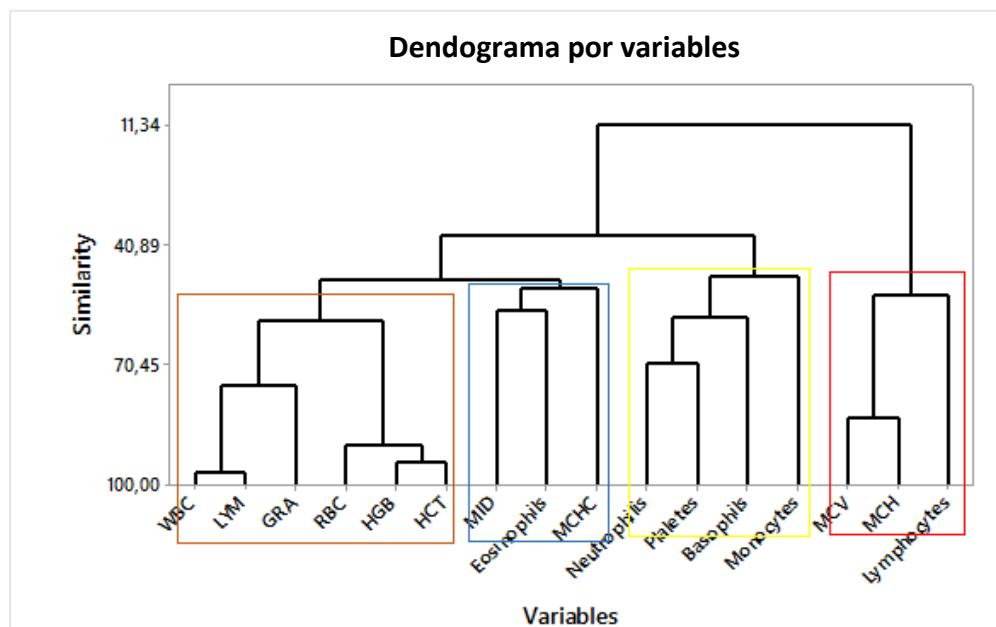


Figura 71: Dendograma por variables de los exámenes hematológicos

Las variables más similares y que tienen una distancia menor son WBC, LYM y HCB, HCT y las variables más distintas al resto son MCHC, Monocytes, Lymphocytes.

Las 16 variables están conformadas en 4 conglomerados.

Primer conglomerado (6 variables):

- WBC
- LYM
- GRA
- RBC
- HGB
- HCT

Segundo conglomerado (3 variables):

- MID
- Eosinophils
- MCHC

Tercer conglomerado (4 variables):

- Neutrophils
- Plaletes
- Basophils
- Monocytes

Cuarto conglomerado (3 variables):

- MCV
- MCH
- Lymphocytes.

- **Análisis de los datos obtenidos sobre los exámenes hispatológicos realizado a los animales**

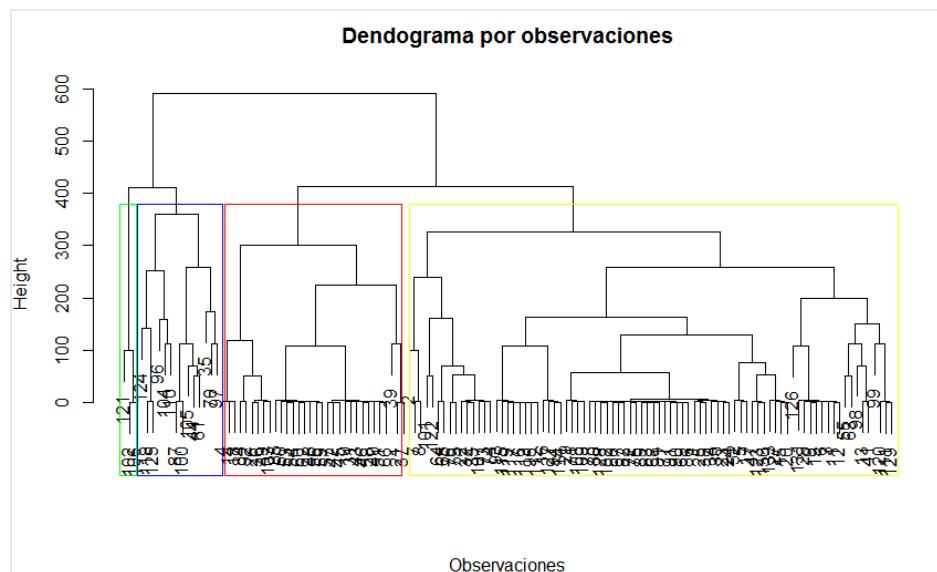


Figura 72: Dendrograma por observaciones de los exámenes hispatológicos

Podemos considerar que en general los datos son bastante homogéneos, sobre todo en los dos últimos conglomerados ya que la mayoría de observaciones quedan a una distancia inferior a 0.

Revisamos como se han agrupado las 133 observaciones directamente sobre los datos

```
## cluster3
## 1 2 3 4
## 84 32 14 3
```

Figura 73: Observaciones por cluster de los datos de los exámenes hispatológicos

Realizando 4 conglomerados, el primer conglomerado tendría 84 observaciones, el segundo 32, el tercero 14 y el cuarto 3 observaciones, lo siguiente será formar los clústeres directamente sobre los datos.

```
##   Animals Zone Specie Sex Age Liverh Kidneyh Spleenh Lungh cluster3
## 1     I01    0      2   2   3       0       0       0       0       1
## 2     I02    0      2   1   3     100     200       0       0       1
## 3     I03    0      2   1   3     200       0       0       0       1
## 5     I05    0      2   2   1     100      50       0       0       1
## 6     I06    0      2   2   2     200       0       0     NA       1
## 7     I07    0      2   2   2      NA     200       0       0       1
```

Figura 74: Formación de los cluster sobre los datos de los exámenes hispatológicos

Dendrograma por variables

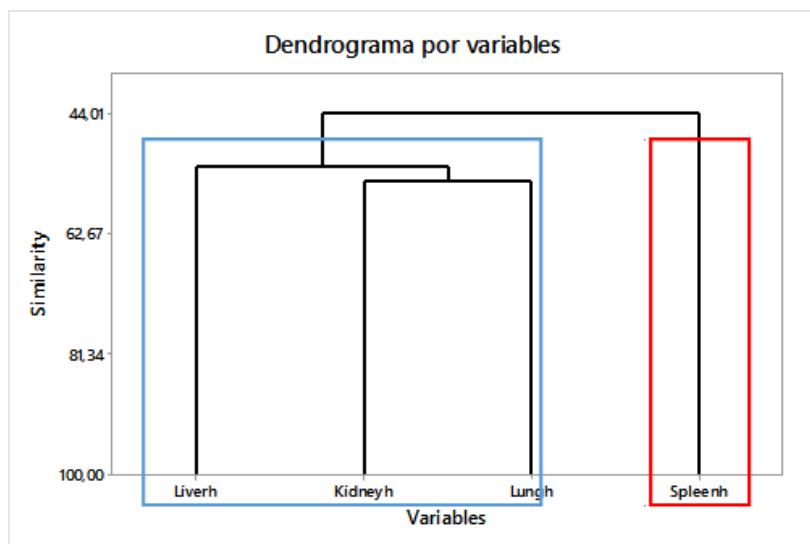


Figura 75: Dendrograma por variables de los exámenes hispatológicos

Las variables más similares y que tienen una distancia menor son Kidneyh, Lungh y la variable más distinta al resto es Spleenhh.

El dendrograma muestra que las 4 variables están conformadas en 2 conglomerados, el primer conglomerado con 3 variables, Liverh, Kidneyh, Lungh y el segundo conglomerado consta la variable Spleenhh

3. Conclusiones

Al realizar el análisis descriptivo se encontró que los test aplicados a las muestras de tierra y agua tienen una contaminación media mayor en las zonas 4 (Zona Control) y 2 (Complejo Petroquímico Adyacente), en cambio en las muestras de sangre y tejidos de roedores capturados sobre el terreno muestra una media mayor en los análisis respecto a las zonas 2(Complejo Petroquímico Adyacente) y la 0 (Incineradora de residuos peligrosos de Constantí).

En el análisis de componente principales respecto al estudio de las muestras de agua y tierra las variables que tienen mayor aporte toxicológico son SurvivalfishhuNAred, Algaefifty, AlgaehuNAred, Nitrateveintiocho, yielhuNAred, INAexgerminationhuNAred, rootlengthhuNAred,, Malformationplant, Seedfourteen, Clorosisplant, AdulEnchytriaed, DaphniainmobilisationhuNAred , AbnormalityfishhuNAred , también se hizo el análisis sobre las muestras de sangre de los animales y las variables que tienen mayor aporte toxicológico son HCG, HCT, RBC, WBC, LYM, por último el análisis sobre las medidas obtenidas de los roedores se concluyó que las variables con mayor aporte toxicológico son BW,BCI, Taillenght ,Handlenght y Footlenght.

Se realizó el estudio para identificar el mayor foco de polución respecto a las distintas locaciones, se logró identificar que la Zona 2 y la Zona 4 están relacionadas, y al analizar la estadística descriptiva se concluye que hay mayor polución en la zona 4 (Zona control) y en la zona 2(Complejo Petroquímico Adyacente) y una menor contaminación en la zona 0 (Incineradora de Residuos Peligrosos de Constantí)

Respecto al análisis factorial de los datos sobre las medidas de los roedores se encontró dos grupos homogéneos, el factor 1 cuyas variables son BW y BCI estas variables se refieren al peso e índice de condición corporal y factor dos que consta de las variables Taillenght, Footlenght y Handlenght que son las longitudes de la pata, mano y cola del roedor y el factor tres consta de la variable BL que es la longitud corporal, también los exámenes hematológicos revelaron 5 grupos homogéneos en el factor 1 las variables HCT, HGB Y RBC que son estudios de los glóbulos rojos, el factor 2 contiene las variables WBC, LYM y GRA, el factor 3 está conformado por la variable Lymphocytes, el factor 4 incluye la variable MCV, MCH y el factor 5 contiene la variable MCHC.

Al realizar el análisis de conglomerados respecto a los datos del estudio de las

muestras de agua y suelo se encontraron factores comunes de las observaciones y se han distribuido en 5 conglomerados y en cada uno de ellos se encuentran los datos referentes a cada zona, respecto al análisis por variables realizamos 6 conglomerados, el primer conglomerado contendría 3 variables, el segundo constaría de 11 variables, el tercero tendría 9 variables, el cuarto estaría conformado de 3 variables, el quinto incluiría 9 variables y el sexto 6 variables. Las variables más similares y que tienen una distancia menor son SurvivalfishhuNAred –Algaefifty, Seedfourteen – Malformationplant y las variables más distintas al resto son las variables Juvenilsfolsom, Survivalfishfifty, HachinghuNAred.

Respecto al análisis de conglomerados de los datos sobre las medidas obtenidas de los roedores se puede considerar que en general los datos son bastante homogéneos, ya que la mayoría de observaciones quedan a una distancia inferior a 20, y se distribuyen en 5 conglomerados, un conglomerado contendría 7 observaciones, otro conglomerado obtendría 42 observaciones, otro conglomerado incluirá 21 observaciones, otro conglomerado abarcaría 21 observaciones y otro conglomerado una sola observación, en lo referente a las variables, existen 4 conglomerados, uno contiene las variables BW y BCI, otro las variables Taillenght, Footlength, Handlength, otro la variable BL y el otro las variables restantes Liver, Kidney, Spleen al comparar con el análisis de factores obtenemos que los tres primeros factores contienen las mismas variables.

En el análisis hematológico se puede considerar que en general los datos son bastante homogéneos, sobre todo en los dos primeros conglomerados ya que la mayoría de observaciones quedan a una distancia inferior a 30. Se obtuvo 5 conglomerados, un primer conglomerado contendría 37 observaciones, un segundo conglomerado tendría 54 observaciones, un tercer conglomerado incluiría 22 observaciones, un cuarto conglomerado constaría de 2 observaciones y un quinto conglomerado estaría conformado por 18 observaciones, en lo referente al análisis de las variables aquellas que tienen más similares y que tienen una distancia menor son WBC , LYM y HCB , HCT y las variables más distinta al resto son MCHC, Monocytes, Lymphocytes, se consideró 4 conglomerados, en el primer conglomerado se incluyen a las 6 variables WBC, LYM, GRA, RBC, HGB, HCT, el segundo conglomerado está conformado por las 3 variables MID, Eosinophils, MCHC, el tercer conglomerado está conformado por las 4 variables Neutrophilis, Plaletes, Basophils, Monocytes y el cuarto conglomerado que consta de 3 variables que son MCV, MCH, Lymphocytes.

En el análisis de conglomerado sobre los exámenes hispatológicos se puede considerar 4 conglomerados, el primer conglomerado tendría 84 observaciones, el segundo 32, el tercero 14 y el cuarto 3 observaciones, el

estudio sobre las variables se formó 2 conglomerados una con tres variables Liverh, Kidneyh, Lungh y el segundo conglomerado consta la variable Spleenh.

La realización de este trabajo ha significado un reto para mí la aplicación de métodos estadísticos en datos de ecotoxicología, siendo la toxicología un área totalmente nueva y novedosa para mí. Al seguir una planificación previa iniciándose con un Plan de Trabajo me ha permitido orientarme en los objetivos del trabajo en sí. La metodología usada en el desarrollo de este trabajo ha permitido concluir los objetivos propuestos para la realización de este estudio.

Al concluir este trabajo se ha identificado algunas líneas de trabajo futuro, una de estas sería, poder identificar los niveles de toxicidad que representan los datos calculados de las muestras en agua, tierra y muestras de tejidos y sangre de roedores analizados. Una segunda línea de trabajo sería realizar un análisis de la influencia de los focos de polución en las poblaciones circundantes al área contaminada.

4. Glosario

R: sistema para computación estadística y creación de gráficos. Proporciona, entre otros, un lenguaje de programación, gráficos de alto nivel, interfaces para otros lenguajes y facilidades para la depuración de errores.

BW: Variable peso corporal

BL: Variable longitud corporal

BCI: Índice de la condición corporal

BCIApo: Índice de la condición corporal sólo Apodemus

BCIMus: Índice de condición corporal sólo Mus

HGB: Es la hemoglobina y es una proteína en los glóbulos rojos que transporta el oxígeno de los pulmones al resto del cuerpo.

RBC: Es el conteo de glóbulos rojos y por lo tanto es el número total de glóbulos rojos.

HCT: Es el hematocrito y mide el porcentaje del volumen de sangre ocupado por los glóbulos rojos.

WBC: Es el conteo de glóbulos blancos y es el número total de glóbulos blancos.

LYM: Linfocitos

MID: Monocitos

MCV: El volumen corpuscular medio (MCV) mide el volumen promedio de todos los glóbulos rojos

MCH: La hemoglobina corpuscular media mide la cantidad de hemoglobina en una célula promedio

MCHC: La concentración de hemoglobina corpuscular media mide la cantidad de concentración de hemoglobina en una célula promedio.

5. Bibliografía

- [1] D. Peña, *Análisis de datos multivariantes*. McGraw-Hill, 2002.
- [2] J. A. M. García y I. A. Uribe, «Técnicas para detección de outliers multivariantes», *Rev. En Telecomunicaciones E Informática*, vol. 3, n.º 5, pp. 11-25, 2013.
- [3] R. Johnson y D. Wichern, *Applied multivariate statistical analysis*, Sexta. Pearson Education, 2007.
- [4] A. D'Amelio y L. Marconi, *Medidas de tendencia central y de variabilidad*.
- [5] M. Quaglino y J. Merello, «Métodos multivariados en estudios de vulnerabilidad social en la provincia de santa fe», 2012. [En línea]. Disponible en: <http://biblioteca.ucm.es/tesis/bio/ucm-t26247.pdf>. [Accedido: 25-may-2017].
- [6] S. Matsumoto, Y. Kamei, A. Monden, y K. i Matsumoto, «Comparison of Outlier Detection Methods in Fault-proneness Models», en *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 2007, pp. 461-463.
- [7] M. Terrádez, *Análisis de componentes principales*.
- [8] R. Zamora y J. Esnaola, «Análisis Factorial y Análisis de componentes principales».
- [9] S. De la Fuente, «Análisis Factorial». 2011.
- [10] B. Everitt, *An R and S-PLUS® Companion to Multivariate Analysis*. Sprenger.
- [11] C. Guisande y A. Vaamonde, *Gráficos estadísticos y mapas con R*, EdicionesDíazdeSanto. 2012.

6. Anexos

6.1 ANEXO 1

6.1.1 RSTUDIO

En primer lugar, importamos nuestra base de datos en R. Para ello añadimos una línea al archivo.

- Datos de los test en muestras de agua y tierra

```
Ejer <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/Ejer.csv")
Ejer[,1] <- as.factor(Ejer[,1])
```

- Datos de las medidas de los roedores

```
tabla1 <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/tabla1.csv")
rownames(tabla1) <- tabla1$Animals
tabla1 <- tabla1[,-1]
tabla1[,1] <- as.factor(tabla1[,1])
tabla1[,2] <- as.factor(tabla1[,2])
tabla1[,3] <- as.factor(tabla1[,3])
tabla1[,4] <- as.factor(tabla1[,4])
```

- Datos de los exámenes hematológicos de los roedores

```
tabla2 <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/tabla2.csv")
rownames(tabla2) <- tabla2$Animals
tabla2 <- tabla2[,-1]
tabla2[,1] <- as.factor(tabla2[,1])
```

- Datos de los exámenes hepatológicos de los roedores

```
tabla3 <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/tabla3.csv")
rownames(tabla3) <- tabla3$Animals
tabla3 <- tabla3[,-1]
tabla3[,1] <- as.factor(tabla3[,1])
```

Análisis descriptivo

- Análisis de datos del estudio de las muestras de agua y suelo

```
(Xbe<-apply(Ejer[,2:42],MARGIN=2,FUN=tapply,INDEX=Zone,mean,na.rm=T))
apply(Ejer[,2:42],MARGIN=2,FUN=tapply,INDEX=Zone,median,na.rm=T)
apply(Ejer[,2:42],MARGIN=2,FUN=tapply,INDEX=Zone,sd,na.rm=T)
```

```
apply(Ejer[, 2:42], MARGIN=2, FUN=tapply, INDEX=Zone, var, na.rm=T)
```

- Análisis de las medidas de los roedores

```
(Xb1<-
apply(tabla1[, 5:15], MARGIN=2, FUN=tapply, INDEX=Zone, mean, na.rm=T))
apply(tabla1[, 5:15], MARGIN=2, FUN=tapply, INDEX=Zone, median, na.rm=T)
apply(tabla1[, 5:15], MARGIN=2, FUN=tapply, INDEX=Zone, sd, na.rm=T)
apply(tabla1[, 5:15], MARGIN=2, FUN=tapply, INDEX=Zone, var, na.rm=T)
```

- Análisis de los exámenes hematológicos de los roedores

```
(Xb2<-
apply(tabla2[, 5:20], MARGIN=2, FUN=tapply, INDEX=Zone, mean, na.rm=T))
apply(tabla2[, 5:20], MARGIN=2, FUN=tapply, INDEX=Zone, median, na.rm=T)
apply(tabla2[, 5:20], MARGIN=2, FUN=tapply, INDEX=Zone, sd, na.rm=T)
apply(tabla2[, 5:20], MARGIN=2, FUN=tapply, INDEX=Zone, var, na.rm=T)
```

- Análisis de los exámenes hispatológicos de los roedores

```
(Xb3<-apply(tabla3[, 5:8], MARGIN=2, FUN=tapply, INDEX=Zone, mean, na.rm=T))
apply(tabla3[, 5:8], MARGIN=2, FUN=tapply, INDEX=Zone, median, na.rm=T)
apply(tabla3[, 5:8], MARGIN=2, FUN=tapply, INDEX=Zone, sd, na.rm=T)
apply(tabla3[, 5:8], MARGIN=2, FUN=tapply, INDEX=Zone, var, na.rm=T)
```

Análisis Gráfico

- Análisis de datos de los exámenes de las muestras de suelo y agua

```
Ejer1<-Ejer[,-c(30:33)]
attach(Ejer1)

XE<-apply(Ejer1[, 2:38], MARGIN=2, FUN=tapply, INDEX=Zone, mean, na.rm=T)
plot(XE[1,], typ="p", col=1, pch=1, ylim=c(-50, 1100), xlab = "Media de las
variables")
par(new=TRUE)
plot(XE[2,], typ="p", col=2, pch=2, ylim=c(-50, 1100))
par(new=TRUE)
plot(XE[3,], typ="p", col=3, pch=3, ylim=c(-50, 1100))
par(new=TRUE)
plot(XE[4,], typ="p", col=4, pch=4, ylim=c(-50, 1100))
par(new=TRUE)
plot(XE[5,], typ="p", col=5, pch=5, ylim=c(-50, 1100))
legend("topright", c("Zona 0", "Zona 1", "Zona 2", "Zona 3", "Zona
4"), pch=1:5, col=1:5)
```

- Análisis de las medidas de los roedores

```
plot(Xb1[1,], typ="p", col=1, pch=1, ylim=c(0, 100), xlab = "Media de las
variables")
par(new=TRUE)
```

```

plot(Xb1[2,],typ="p",col=2,pch=2,ylim=c(0,100))
par(new=TRUE)
plot(Xb1[3,],typ="p",col=3,pch=3,ylim=c(0,100))
par(new=TRUE)
plot(Xb1[4,],typ="p",col=4,pch=4,ylim=c(0,100))
par(new=TRUE)
plot(Xb1[5,],typ="p",col=5,pch=5,ylim=c(0,100))
legend("topright",c("Zona 0", "Zona 1", "Zona 2", "Zona 3", "Zona
4"),pch=1:5,col=1:5)

```

- Análisis de los exámenes hematológicos de los roedores

```

plot(Xb2[1,],typ="p",col=1,pch=1,ylim=c(0,1100),xlab = "Media de las
variables")
par(new=TRUE)
plot(Xb2[2,],typ="p",col=2,pch=2,ylim=c(0,100))
par(new=TRUE)
plot(Xb2[3,],typ="p",col=3,pch=3,ylim=c(0,100))
par(new=TRUE)
plot(Xb2[4,],typ="p",col=4,pch=4,ylim=c(0,100))
par(new=TRUE)
plot(Xb2[5,],typ="p",col=5,pch=5,ylim=c(0,100))
legend("topright",c("Zona 0", "Zona 1", "Zona 2", "Zona 3", "Zona
4"),pch=1:5,col=1:5)

```

- Análisis de los exámenes hispatológico de los roedores

```

plot(Xb3[1,],typ="p",col=1,pch=1,ylim=c(0,150),xlab = "Media de las
variables")
par(new=TRUE)
plot(Xb3[2,],typ="p",col=2,pch=2,ylim=c(0,150))
par(new=TRUE)
plot(Xb3[3,],typ="p",col=3,pch=3,ylim=c(0,150))
par(new=TRUE)
plot(Xb3[4,],typ="p",col=4,pch=4,ylim=c(0,150))
par(new=TRUE)
plot(Xb3[5,],typ="p",col=5,pch=5,ylim=c(0,150))
legend("topright",c("Zona 0", "Zona 1", "Zona 2", "Zona 3", "Zona
4"),pch=1:5,col=1:5)

```

Gráfica de las correlaciones

- Análisis de datos del estudio de las muestras de agua y suelo

```

Ejer.<-na.omit(Ejer)
(cc <- cor(Ejer.[,2:42], use = "complete.obs"))

corrplot(cc, type = "upper", order = "hclust",
tl.col = "black", tl.srt = 45)

```

- Análisis de las medidas de los roedores

```

library(corrplot)
tabla1.<-na.omit(tabla1[,-c(8,9)])
(cc1 <- cor(tabla1.[,5:13], use = "complete.obs"))

corrplot(cc1, type = "upper", order = "hclust",
tl.col = "black", tl.srt = 45)

```

- Análisis de los exámenes hematológicos de los roedores

```

library(corrplot)
tabla2.<-na.omit(tabla2)
(cc2 <- cor(tabla2.[,5:20], use = "complete.obs"))

corrplot(cc2, type = "upper", order = "hclust",
tl.col = "black", tl.srt = 45)

```

- Análisis de los exámenes hispatológicos de los roedores

```

library(corrplot)
tabla3.<-na.omit(tabla3)
(cc3 <- cor(tabla3.[,5:8], use = "complete.obs"))

##          Liverh      Kidneyh     Spleenhh    Lungh
## Liverh  1.00000000  0.04292798 -0.11981702  0.06022104
## Kidneyh  0.04292798  1.00000000 -0.02646874  0.09005911
## Spleenhh -0.11981702 -0.02646874  1.00000000  0.02497212
## Lungh    0.06022104  0.09005911  0.02497212  1.00000000

```

Variables redundantes

- Análisis de las medidas de los roedores

```

tabla1.<-na.omit(tabla1[,-c(8,9)])
S1<-cov(tabla1.[,5:13])
eigen(S1)

```

- Análisis de los exámenes hematológicos de los roedores

```

tabla2.<-na.omit(tabla2)
S2<-cov(tabla2.[5:20])
eigen(S2)

```

- Análisis de los exámenes hispatológicos de los roedores

```

tabla3.<-na.omit(tabla3)
S3<-cov(tabla3.[,5:8])
eigen(S3)

```

Datos atípicos

- Análisis de datos del estudio de las muestras de agua y suelo

```
Ejerda<-Ejer[,-c(29:33)]  
library(ggplot2)  
library(reshape)  
ggplot(data = melt(Ejerda), aes(x = variable, y = value)) +  
geom_boxplot()
```

- Análisis de las medidas de los roedores

```
library(ggplot2)  
library(reshape)  
ggplot(data = melt(tabla1), aes(x = variable, y = value)) +  
geom_boxplot(aes())
```

```
tabla1.<-na.omit(tabla1[,-c(8,9)])  
TDA <- tabla1[,5:13]  
library(MVN)  
  
result <- mvOutlier(TDA, qqplot=T, alpha = 0.5, method = c("quan"),  
position = NULL, offset = 0.5)  
  
result$outlier
```

- Análisis de los exámenes hematológicos de los roedores

```
library(ggplot2)  
library(reshape)  
ggplot(data = melt(tabla2), aes(x = variable, y = value)) +  
geom_boxplot(aes())
```

- Análisis de los exámenes hispatológicos de los roedores

```
library(ggplot2)  
library(reshape)  
ggplot(data = melt(tabla3), aes(x = variable, y = value)) +  
geom_boxplot(aes())
```

Análisis de componentes principales

- Análisis de datos del estudio de las muestras de agua y suelo

```
Ejer <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/Ejer.csv")  
Ejer[,1] <- as.factor(Ejer[,1])  
dim(Ejer)  
  
## [1] 25 42
```

```

attach(Ejer)

library(FactoMineR)
res.pcaE <- PCA(Ejer,quali.sup = 1)
res.pcaE$eig[1:5, ]
barplot(res.pcaE$eig[,1], main = "Eigenvalues",
names.arg = paste("Dim", 1:nrow(res.pcaE$eig), sep = ""))
dimdesc(res.pcaE, proba = 0.2)
res.pcaE$var$coord
plot(res.pcaE, habillage = 1)

```

- Análisis de los exámenes hematológicos de los roedores

```

library(FactoMineR)
res.pca2 <- PCA(tabla2, quali.sup = 1:4)
res.pca2$eig[1:5, ]
barplot(res.pca2$eig[,1], main = "Eigenvalues",
names.arg = paste("Dim", 1:nrow(res.pca2$eig), sep = ""))
dimdesc(res.pca2, proba = 0.2)
res.pca2$var$coord
plot(res.pca2, habillage = 1)
plotellipses(res.pca2, keepvar = all)
plotellipses(res.pca2, keepvar = c(1))

```

- Análisis de los exámenes hispatológicos de los roedores

```

library(FactoMineR)
res.pca3 <- PCA(tabla3,quali.sup = 1:4)
res.pca3$eig[1:5, ]
barplot(res.pca3$eig[,1], main = "Eigenvalues",
names.arg = paste("Dim", 1:nrow(res.pca3$eig), sep = ""))
dimdesc(res.pca3, proba = 0.2)
res.pca3$var$coord
plot(res.pca3, habillage = 1)

```

Análisis Factorial

- Análisis de datos del estudio de las muestras de agua y suelo

```

Tabla1<-na.omit(tabla1[,-c(8:9)])
Tabla1<-Tabla1[-
c(1,9,11,12,13,15,16,20,21,23,24,26,28,32,33,35,36,37,41,42,43,44,45,4
9,56,57,68,69,70,72,83,84,86,88,89,90,97,102,130),]

library(psy)
modelo1 <- factanal(Tabla1[,5:13],
3,rotation="varimax",scores="regression")

modelo1

```

```

scree.plot(modelo1$correlation)
head(modelo1$scores)
load <- modelo1$loadings[,1:2]
plot(load,type="n")
text(load,labels=names(Tabla1[,5:13])),cex=.7)

```

- Análisis de los datos obtenidos sobre los análisis hematológicos realizado a los animales

```

library(psy)
tabla2.<-na.omit(tabla2)
modelo2 <- factanal(tabla2.[,5:20],
5,rotation="varimax",scores="regression")
modelo2

scree.plot(modelo2$correlation)
head(modelo2$scores)

load <- modelo2$loadings[,1:2]
plot(load,type="n")
text(load,labels=names(tabla2.[,5:20])),cex=.7)

```

- Análisis de los datos obtenidos sobre los análisis hispatológicos realizado a los animales

```

library(psy)
modelo3 <-
factanal(tabla3.[,5:8],1,rotation="varimax",score="regression")
scree.plot(modelo3$correlation)

modelo3

head(modelo3$scores)

load <- modelo3$loadings[,1]
plot(load,type="n")
text(load,labels=names(tabla3.[,5:8])),cex=.7)

```

Análisis de conglomerado

- Análisis de datos del estudio de las muestras de agua y suelo

```

X<-Ejer[,2:42]
Z=scale(X)
cj=hclust(dist(Z)^2)
plot(cj,main="Dendograma",hang=-1)
polygon(x=c(0.7,0.7,5.5,5.5),y=c(-25,90,90,-25),border="green")

```

```

polygon(x=c(5.8,5.8,10.5,10.5),y=c(-25,90,90,-25),border="blue")
polygon(x=c(10.8,10.8,14.4,14.4),y=c(-25,90,90,-25),border="brown")

polygon(x=c(14.8,14.8,21.3,21.3),y=c(-25,90,90,-25),border="yellow")
polygon(x=c(21.8,21.8,25.5,25.5),y=c(-25,98,98,-25),border="red")

cluster<-cutree(hclust(dist(Z)),method="complete"),h=10
table(cluster)

```

Obtener el total de clústeres y cuantas observaciones contiene cada cluster

```

cluster<-cutree(hclust(dist(Z)),method="complete"),h=10
table(cluster)

```

Vamos a estudiar cómo se forman los clústeres directamente sobre los datos.

```

grup<-cbind(data.frameTabla4[,1:3]),cluster)
library(reshape)
grup<-sort_df(grup,vars='cluster')
head(grup)

```

- Análisis de los datos respecto a las medidas de los roedores

```

Tabla1<-Tabla1[-
c(1,9,11,12,13,15,16,20,21,23,24,26,28,32,33,35,36,37,41,42,43,44,45,4
9,56,57,68,69,70,72,83,84,86,88,89,90,97,102,130),]

hc1<-hclust(dist(Tabla1[,1:13]),method = "euclidean",method="complete")
plot(hc1)

polygon(x=c(0.6,0.6,8.3,8.3),y=c(-25,32,32,-25),border="green")
polygon(x=c(8.8,8.8,29.3,29.3),y=c(-25,32,32,-25),border="red")
polygon(x=c(29.7,29.7,71.3,71.3),y=c(-25,32,32,-25),border="blue")
polygon(x=c(71.5,71.5,72.2,72.2),y=c(-25,32,32,-25),border="yellow")
polygon(x=c(72.6,72.6,92,92),y=c(-25,32,32,-25),border="tomato1")

cluster1<-cutree(hclust(dist(Tabla1[,1:13])),method="complete"),h=35
table(cluster1)

grup1<-cbind(data.frame(Tabla1[,6:10]),cluster1)
library(reshape)
grup1<-sort_df(grup1,vars='cluster1')
head(grup1)

```

- Análisis de los datos respecto a los exámenes hematológicos de los roedores

```

Tabla2 <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/tabla2.csv")
hc2<-hclust(dist(Tabla2[,2:21]),method = "euclidean",method="complete")
plot(hc2)

polygon(x=c(0.6,0.6,37,37),y=c(-35,90,90,-35),border="green")

```

```

polygon(x=c(37.5,37.5,91.5,91.5),y=c(-35,90,90,-35),border="red")
polygon(x=c(92,92,109,109),y=c(-35,90,90,-35),border="blue")
polygon(x=c(109.5,109.5,111.5,111.5),y=c(-35,90,90,-35),border="brown")
polygon(x=c(111.8,111.8,137,137),y=c(-35,90,90,-35),border="yellow")

cluster2<-cutree(hclust(distTabla2[,2:21]),method="complete"),h=90)
table(cluster2)

grup2<-cbind(data.frame(Tabla2[,5:9]),cluster2)
library(reshape)
grup2<-sort_df(grup2,vars='cluster2')
head(grup2)

```

- Análisis de los datos respecto a los exámenes hispatológicos de los roedores

```

Tabla3 <- read.csv2("D:/Master/Segundo Semestre/TFM/Datos/tabla3.csv")

hc3<-hclust(distTabla3[,2:9],method = "euclidean",method="complete")
plot(hc3)
polygon(x=c(0.3,0.3,3,3),y=c(-140,380,380,-140),border="green")
polygon(x=c(3.3,3.3,18,18),y=c(-140,380,380,-140),border="blue")
polygon(x=c(18.3,18.3,48.7,48.7),y=c(-140,380,380,-140),border="red")
polygon(x=c(50,50,134,134),y=c(-140,380,380,-140),border="yellow")

cluster3<-cutree(hclust(distTabla3[,2:9]),method="complete"),h=370)
table(cluster3)

grup3<-cbind(data.frame(Tabla3[,1:9]),cluster3)
library(reshape)
grup3<-sort_df(grup3,vars='cluster3')
head(grup3)

```