



UNIVERSIDAD  
DE GRANADA

## Memoria final

**Análisis de la Efectividad de un Fármaco Osteoanabólico**

Pablo Gálvez Ortigosa  
Eduardo Rodríguez Cao  
Mario Megías Mateo  
José Luis Mera Cardoso

## Índice

<b>1. Resumen</b>	<b>3</b>
<b>2. Introducción</b>	<b>3</b>
<b>3. Materiales y métodos</b>	<b>4</b>
3.1. Materiales . . . . .	4
3.2. Métodos . . . . .	5
<b>4. Resultados</b>	<b>6</b>
<b>5. Discusión</b>	<b>11</b>
5.1. Efecto del peso . . . . .	11
5.2. Efecto de la dosis . . . . .	11
5.3. Influencia del tiempo . . . . .	12
5.4. Interacción dosis $\times$ tiempo . . . . .	12
<b>6. Conclusión</b>	<b>12</b>

### 1. Resumen

En este proyecto analizaremos la efectividad de un fármaco osteoanabólico. Para ello, compararemos distintos parámetros óseos medidos en 134 ratas durante un ensayo clínico aleatorizado, en el que cada rata es sometida a distintas dosis del tratamiento durante diferentes periodos de tiempo.

Nuestro primer objetivo será demostrar que el fármaco es eficaz, analizando la influencia de la dosis en la calidad ósea. En segundo lugar, queremos comprobar que tanto el peso de la rata a la llegada al ensayo, como al comienzo del tratamiento; así como la duración del tratamiento también influyen, y en este último caso determinar con qué fuerza lo hace. Intentaremos finalmente ajustar un rango de valores mínimo y máximo para la dosis administrada en función del tiempo que garantice una calidad ósea relevante.

Para satisfacer estos objetivos, se aplicarán distintas técnicas de estadística multivariante: ACP, análisis factorial, análisis clúster y regresión logística.

Los resultados finales muestran que la dosis administrada y el peso de la rata influyen en la calidad ósea. También calculamos que la fuerza con la que influye la duración es muy elevada (ODDS-Ratio  $\approx 11,44$ ). Finalmente, determinamos que dosis entre 80 y 400 mg/día durante 21 días, aseguran una calidad ósea relevante.

### 2. Introducción

El tratamiento de enfermedades óseas como la osteoporosis es un área de interés en la investigación biomédica, dado el gran impacto que estas patologías tienen en la calidad de vida y los costos asociados a su manejo. En este contexto, los fármacos osteoanabólicos representan una línea de investigación prometedora ya que buscan mejorar la calidad ósea estimulando la formación de hueso nuevo. Sin embargo, el desarrollo de estos tratamientos requiere validar rigurosamente su eficacia y determinar los parámetros óptimos para su administración.

Este trabajo se centra en el análisis de un fármaco osteoanabólico a través de un ensayo clínico aleatorizado realizado en ratas. Durante este ensayo, se evaluaron diversas variables óseas en 134 ratas expuestas a diferentes dosis del fármaco y durante diferentes periodos de tiempo. Adicionalmente, se recopilieron datos sobre el peso inicial de las ratas y el peso antes de empezar el tratamiento, variables que podrían influir en la calidad ósea. La riqueza de estos datos nos permite abordar el problema desde una perspectiva estadística multivariante, empleando herramientas avanzadas para identificar patrones y relaciones significativas.

La metodología aplicada incluye el análisis de componentes principales (ACP), análisis factorial, análisis clúster y regresión logística. Estas técnicas nos permitirán, por un lado, confirmar la eficacia del fármaco y, por otro, explorar cómo factores como la dosis, la duración del tratamiento y el peso de las ratas afectan la calidad ósea. Asimismo, se buscará ajustar un rango de dosis óptimo para garantizar resultados clínicos relevantes.

En este informe, se describirán los pasos seguidos para abordar el problema, incluyendo la descripción del conjunto de datos, la aplicación de técnicas estadísticas y los resultados

obtenidos. Finalmente, se discutirán las implicaciones de los hallazgos y se propondrán recomendaciones que podrían ser adoptadas por la empresa.

El objetivo principal de este trabajo es validar la efectividad del fármaco osteoanabólico y proponer directrices basadas en datos para su administración óptima, aportando así un conocimiento valioso para el área de investigación y desarrollo en tratamientos óseos.

### 3. Materiales y métodos

#### 3.1. Materiales

El conjunto de datos proporcionado tiene las siguientes variables:

1. Grupo de ratas (todo el grupo recibe la misma dosis durante el mismo tiempo)
2. Identificador único de cada rata
3. Estudio al que pertenece el grupo de ratas
4. Dosis suministrada a la rata
5. Peso de la rata en el momento de la medición
6. Separación trabecular (Tb.Sp)
7. Número de trabéculas (Tb.N)
8. Contenido mineral óseo (BMC) en mg
9. Densidad mineral ósea (BMD) en  $\text{mg}/\text{cm}^2$
10. Longitud de la tibia en mm
11. Superficie proyectada del área ósea (GPSurface) en  $\text{mm}^2$
12. Volumen proyectada del área ósea (GPVolume) en  $\text{mm}^3$

Cada número de ID tiene asociados tres registros, ya que se muestran las mediciones al comienzo del ensayo, al comienzo del tratamiento y al final del tratamiento (donde se miden todas las variables óseas). Para crear una verdadera base de datos se unifican estos tres registros en una única fila, conservando los parámetros óseos, los distintos pesos y los distintos tiempos. Es decir, se añaden las siguientes variables:

1. Tiempo intermedio: número de días a los que comienza el tratamiento
2. Duración: número de días que dura el tratamiento
3. Peso de la rata cuando llega al ensayo
4. Peso de la rata al comienzo del tratamiento
5. Peso final al terminar el tratamiento y al medir las variables óseas

Obtenemos así 134 registros con IDs únicos. A partir de un primer análisis exploratorio univariante básico, mostramos una tabla que incluye las principales medidas de posición, dispersión y forma más importantes de las variables cuantitativas de este conjunto de datos:

	BMC	BMD	Dosis	Duracion	GPSurface	GPVolume	Longitud	PesoFin	PesoIni	PesoInter	Tb.N	Tb.Sp	TiempoInter
Mean	0.28	142.76	154.33	16.12	40.37	5.60	38.70	302.82	105.68	175.16	3.13	0.33	18.24
Std.Dev	0.07	16.69	277.24	3.23	4.83	1.15	3.02	64.84	22.25	72.34	0.60	0.08	6.46
Min	0.15	105.62	0.00	14.00	28.96	3.35	32.45	163.00	63.00	77.00	1.67	0.21	14.00
Q1	0.24	134.18	0.00	14.00	36.76	4.76	37.31	266.95	76.05	138.75	2.71	0.28	14.00
Median	0.27	143.05	8.00	14.00	40.48	5.56	38.74	302.85	113.25	161.75	3.14	0.32	14.00
Q3	0.32	151.99	100.00	21.00	43.38	6.16	41.16	335.10	121.30	184.50	3.56	0.37	28.00
Max	0.50	186.94	1000.00	21.00	54.68	10.32	45.52	480.60	140.80	397.80	4.62	0.60	28.00
MAD	0.06	13.16	11.86	0.00	5.03	1.04	3.02	50.33	16.09	34.10	0.63	0.07	0.00
IQR	0.08	17.69	100.00	7.00	6.54	1.40	3.83	67.82	45.07	45.18	0.85	0.09	14.00
CV	0.25	0.12	1.80	0.20	0.12	0.20	0.08	0.21	0.21	0.41	0.19	0.23	0.35
Skewness	0.43	0.05	2.03	0.85	0.35	0.90	-0.17	0.42	-0.68	1.55	-0.13	1.17	0.85
SE.Skewness	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Kurtosis	0.49	0.06	3.17	-1.29	0.15	1.71	-0.54	0.36	-1.03	2.23	-0.39	1.47	-1.29

así como información básica de las variables cualitativas:

Grupo	Freq	% Total
20_0_Grupo	8	5.970149
20_1_Grupo	11	8.208955
20_10_Grupo	8	5.970149
20_11_Grupo	8	5.970149
20_2_Grupo	11	8.208955
20_4_Grupo	8	5.970149
20_5_Grupo	8	5.970149
20_6_Grupo	12	8.955224
20_7_Grupo	8	5.970149
20_8_Grupo	8	5.970149
20_9_Grupo	8	5.970149
23_1_Grupo	9	6.716418
23_2_Grupo	9	6.716418
23_3_Grupo	9	6.716418
23_Control	9	6.716418
<NA>	0	0.000000
Total	134	100.000000

Estudio	Freq	% Total
1	56	41.79104
2	42	31.34328
3	36	26.86567
<NA>	0	0.00000
Total	134	100.00000

### 3.2. Métodos

En primer lugar, se ha realizado un análisis exploratorio univariante preliminar de los datos:

- Se han detectado valores perdidos cuantitativos, menos de un 5% en cada variable, y con dos casos destacables. En uno, hay dos ratas sin variables óseas, de modo que se han eliminado del dataset. En otro, hay ratas con algunos valores perdidos, pero no todos, de modo que se han sustituido por la media.

- Se ha realizado un análisis descriptivo numérico clásico, dando las principales medidas de posición, dispersión y forma para poder entender mejor los datos.
- En las variables con valores extremos, estos se han identificado con gráficos boxplot y como tenemos muy pocos datos en el dataset que pueden ser significativos (suponemos que no son errores de medición), hemos tomado la decisión de no sustituirlos por la media y dejarlos como están.

En segundo lugar, se han aplicado diferentes técnicas estadísticas de Análisis Multivariante, comprobando los supuestos necesarios para cada una de ellas:

- Correlación: A nivel poblacional se ha justificado utilizando el contraste de esfericidad de Bartlett, que permite comprobar si las correlaciones son distintas de 0 de modo significativo. A nivel muestral se ha justificado observando la matriz de correlaciones, la matriz de correlaciones policrítica y otras representaciones gráficas.
- Normalidad univariante: Se ha hecho una exploración gráfica de la normalidad de las distribuciones individuales mediante histogramas y gráficos qqplots. Esto podría dar una idea de la posible distribución normal de las variables unidimensionales, pero la conclusión definitiva se ha obtenido mediante un test de hipótesis, que en este caso ha sido el de Shapiro-Wilks.
- Normalidad multivariante: Se ha estudiado mediante tests de hipótesis como el de Royston y el de Henze-Zirkler. Hay que tener en cuenta que la normalidad multivariante puede verse afectada por la presencia de outliers multivariantes, por lo que se ha hecho un análisis de estos.
- Se han realizado un Análisis de Componentes Principales para reducir la dimensión por medio de variables observables
- Se ha realizado un Análisis Factorial para identificar variables latentes (no observables) que tengan alta correlación con un grupo de variables observables y correlación prácticamente nula con el resto
- Se ha ajustado un modelo de regresión logística para clasificar las ratas en clases de la variable respuesta de calidad ósea.
- Se ha realizado Análisis Clúster para confirmar que la agrupación de la variable respuesta utilizada en los modelos de clasificación es adecuada.

## 4. Resultados

Primeramente, hemos realizado un análisis exploratorio bivariante y trivariante. Las gráficas más interesantes que se han obtenido en el análisis bivariante han sido:

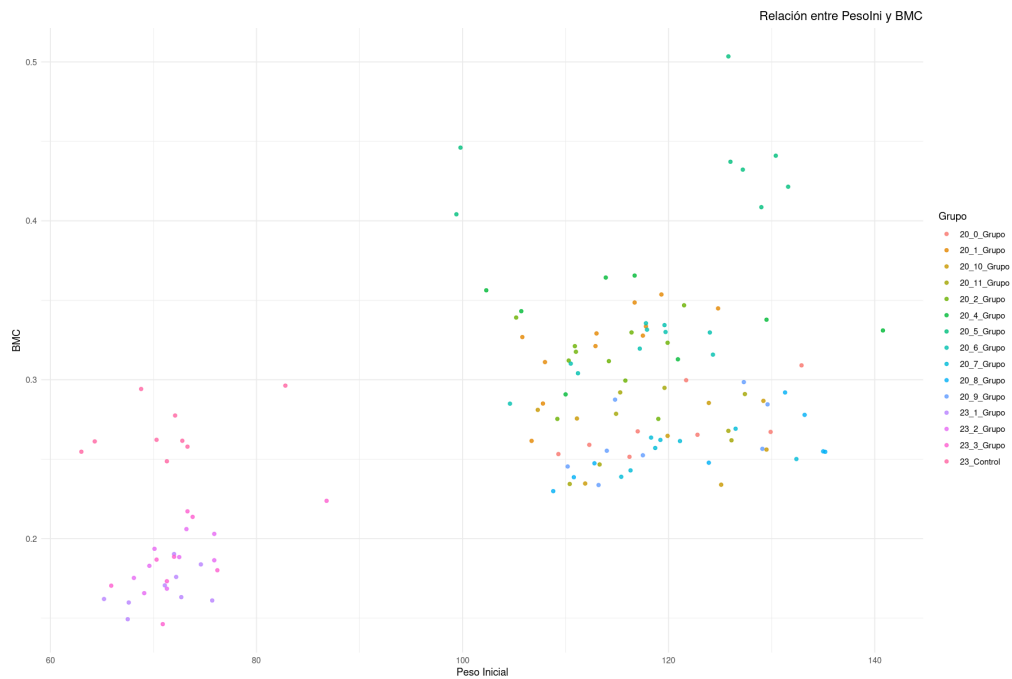


Figura 1: Relación entre las variables Peso Inicial y BMC

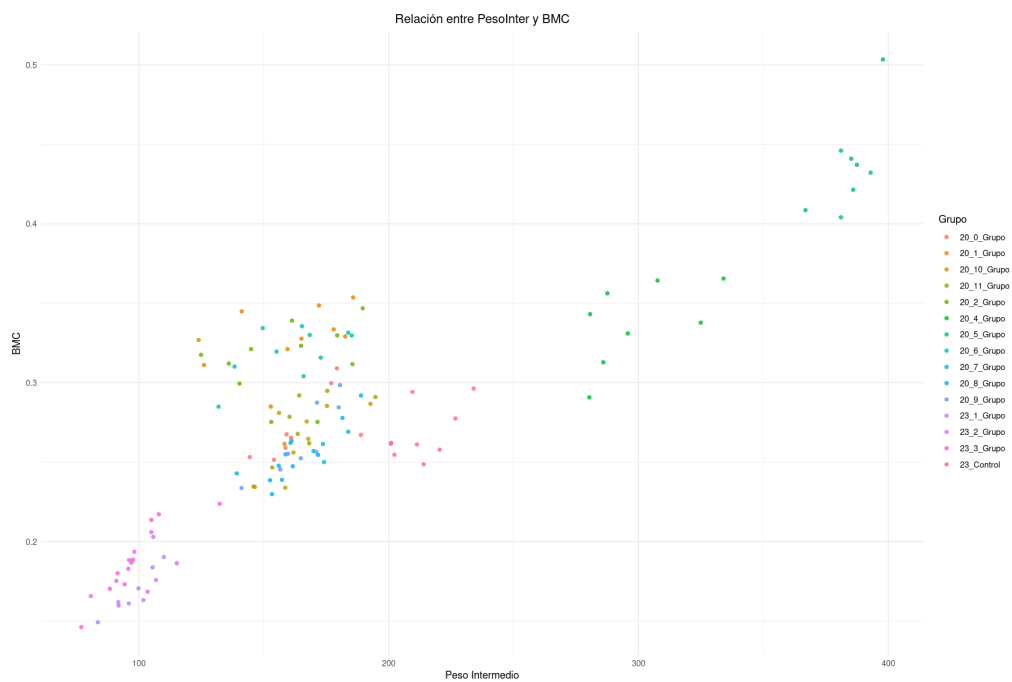


Figura 2: Relación entre las variables Peso Intermedio y BMC

donde en ambas gráficas se muestra una clara relación entre las variables. En cuanto al análisis trivariante se ha realizado una gráfica con la Dosis, Duración y BMC:

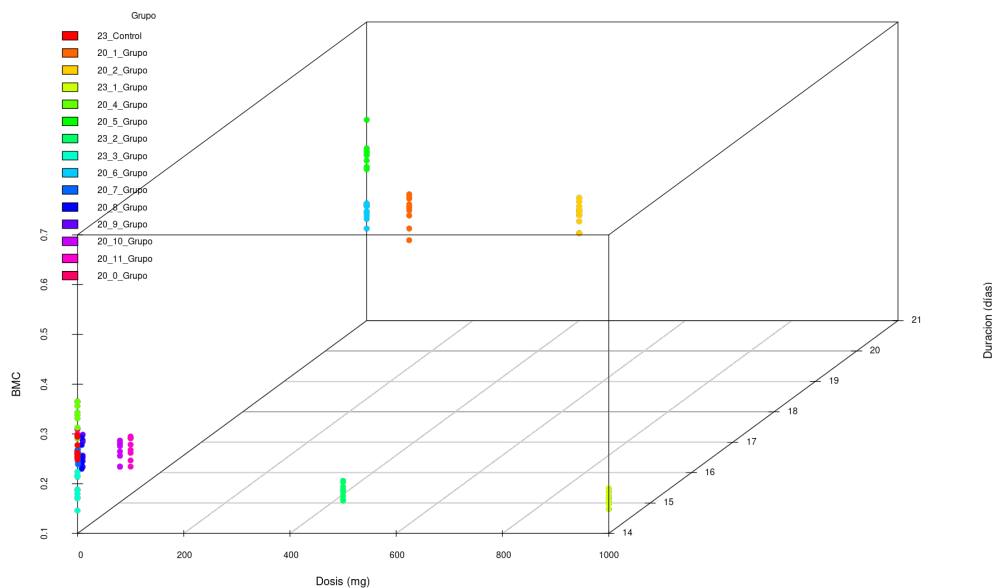


Figura 3: Relación entre las variables Dosis, Duración y BMC

El test de Bartlett ha dado evidencia para rechazar la hipótesis de independencia de los datos. De este modo, se ha concluido que existe correlación entre los datos y, por tanto, se puede plantear una reducción de la dimensión mediante un Análisis de Componentes Principales o un Análisis Factorial. En el Análisis Factorial se han obtenido dos factores latentes. El primero de ellos influye en 8 de las variables observadas, mientras que el segundo en solo 2.

Se ha realizado también, con objetivo de reducir dimensionalidad, un ACP. De las 13 variables de partida, se ha pasado a un conjunto de 3 variables observables, que son combinaciones lineales de las originales y acumulan casi un 80 % de la varianza explicada. Representando los datos utilizando la primera y la tercera componente principal, obtenemos:



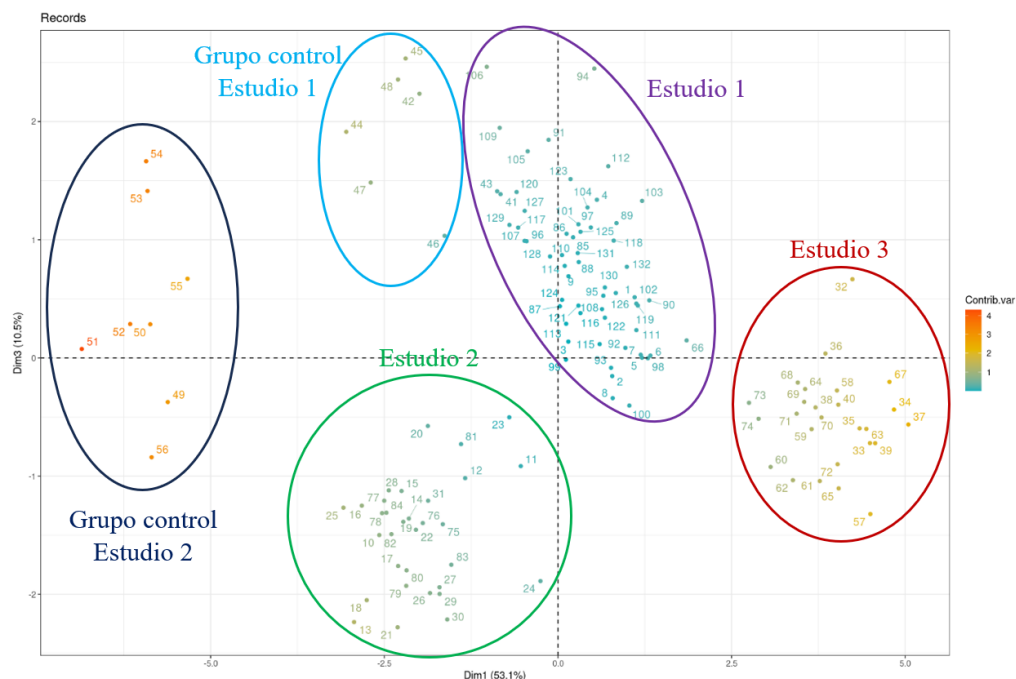


Figura 4: Datos representados utilizando la 1ª y 3ª componente principal. Se muestran con elipses las distintas agrupaciones.

También se ha realizado un ACP separando las muestras por estudio, así como un análisis clúster:

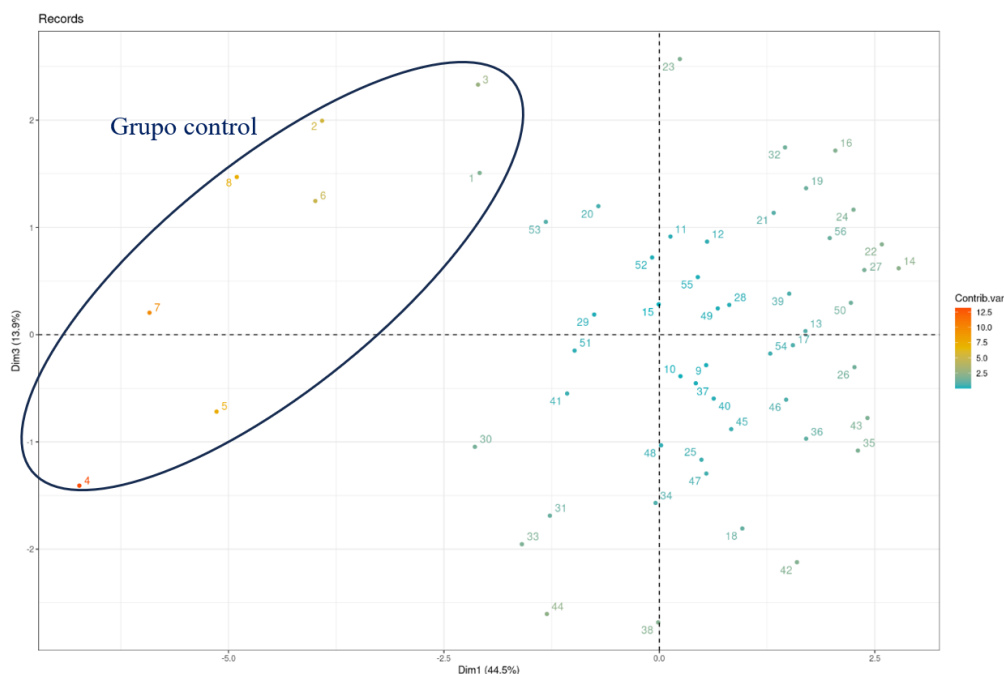


Figura 5: Datos representados utilizando la 1ª y 2ª componente principal (ACP restringido al estudio 1). Se muestran con elipses las distintas agrupaciones.

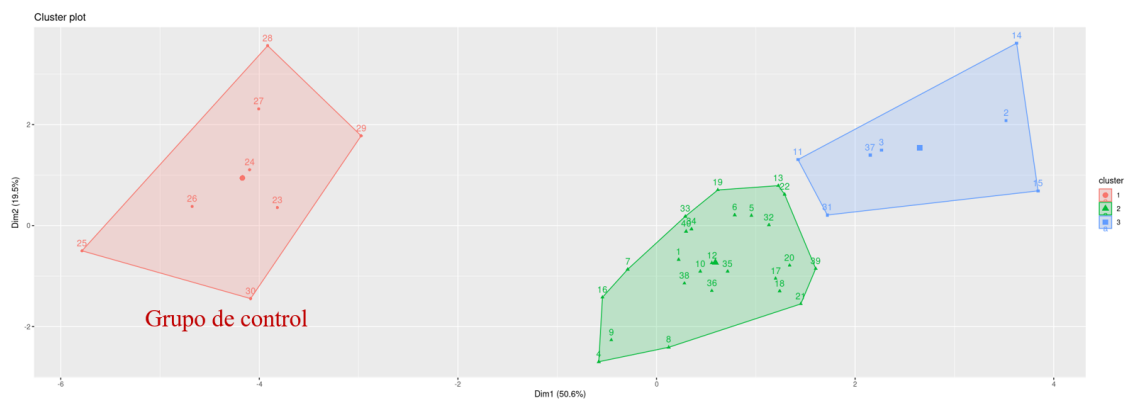


Figura 6: Análisis con 3 clústers restringido al estudio 2

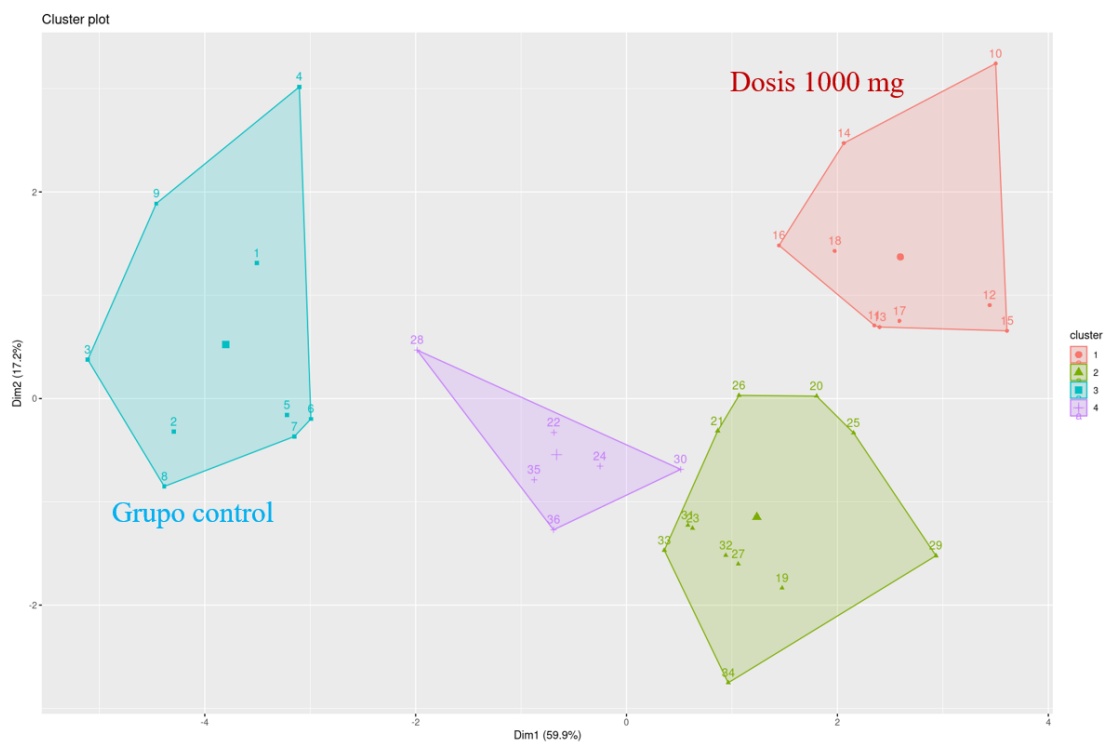


Figura 7: Análisis clúster restringido al estudio 3

En cuarto lugar, mediante los tests correspondientes se ha concluido que no hay normalidad univariante para la mayoría de las variables. Por tanto, se ha decidido no realizar un Análisis de Discriminante. En su lugar, ajustaremos varios modelos de regresión logística.

Para ello, se ha definido una variable categórica (Relevante / No relevante) a partir de un cierto valor en la variable BMC. Cuando se utilizan como variables predictoras Duración, Peso Inicial, Peso Intermedio y Dosis se obtienen unas ODDS-Ratio de 11,44 para la Duración y de 17,19 para el Peso Intermedio. Este modelo obtiene menos de un 4 % de error. Por otra parte, si se utiliza únicamente la Dosis como variable predictora, se

obtienen unas ODDS-Ratio de 1,77.

## 5. Discusión

### 5.1. Efecto del peso

Tal y como se observa en la Figura 1, se aprecia una ligera distinción en la cantidad de minerales en el hueso (BMC) cuando la rata comienza el estudio pesando más o menos. En general, comprobamos que a mayor peso inicial la rata obtiene al final de el estudio un mayor BMC. Asimismo, las ratas del estudio 3 aparecen claramente diferenciadas del resto, ya que son las que menor peso inicial tenían al comienzo.

De forma similar, en la Figura 2, apreciamos que cuando el peso intermedio (peso al inicio del tratamiento) es mayor, también se obtiene un mayor BMC (la relación es casi lineal), y en general, todas las variables óseas mejoran.

Para asegurar la influencia de estos dos pesos en la calidad ósea final, realizamos un ACP y proyectamos los datos sobre la primera y la tercera componente, como se puede ver en la Figura 4. En la imagen se pueden apreciar agrupaciones de ratas con pesos muy similares: tanto el grupo de control del Estudio 2 (a la izquierda de la imagen) como el del Estudio 1 (en la parte superior de la imagen) se separan del resto de muestras notablemente. También se observan agrupaciones de las ratas del Estudio 3 (a la derecha de la imagen), del Estudio 2 (en la parte inferior) y del Estudio 1 (parte central).

### 5.2. Efecto de la dosis

El Análisis Factorial ha mostrado la existencia de dos factores latentes que son suficientes para explicar los datos. El primero de ellos correlaciona mucho con todas las variables óseas (salvo Tb.Sp y Tb.N), Peso Inicial, Peso Intermedio Peso Final y Dosis.

Para confirmar este efecto de la dosis se ha realizado un ACP y un análisis clúster. En este caso, para encontrar agrupaciones por dosis y no únicamente por pesos se aplicarán dichas técnicas a grupos de ratas con pesos similares, es decir, a ratas del mismo estudio.

En el ACP del estudio 1 (Figura 5) se observa una clara separación entre el grupo de control y el resto de ratas. En cuanto al análisis clúster del estudio 2 (Figura 6), también se muestra una clara diferenciación entre el grupo de control y el resto de ratas. Por último, en el análisis clúster del estudio 3 se observa una clara separación del grupo de control y también de los grupos de ratas que han recibido una dosis de 1000 mg/día.

También se ha ajustado un modelo de regresión logística utilizando el percentil 65 del BMC para definir una variable respuesta Calidad Ósea Relevante. Al ajustar una regresión logística con la dosis como variable predictora (eliminando las dosis de 500 y 1000, para las que no se obtienen buenos resultados, véase 3) se obtienen unas ODDS-Ratio de 1,77.

### 5.3. Influencia del tiempo

Utilizando la anterior variable respuesta, se ajusta ahora un modelo de regresión logística con predictores las variables Peso Intermedio, Peso Inicial, Dosis y Duración. Se obtiene que la Duración del tratamiento es significativa con bastante fuerza, obteniendo unas ODDS-Ratio de 11,44.

### 5.4. Interacción dosis $\times$ tiempo

Fijándonos en la Figura 3, comprobamos que dosis de más de 400 miligramos producen un empeoramiento de la mayoría de variables óseas. Por tanto, dichas dosis quedan descartadas.

En cuanto al resto de dosis, cuando la duración del tratamiento es de 14 días, se observa que dosis entre 1 y 80 mg/día producen resultados interesantes. Cuando la duración es de 21 días, los mejores resultados se obtienen con dosis entre 80 y 400 mg. Es en este último caso cuando las variables óseas mejoran más por lo que sería en principio la mejor opción.

## 6. Conclusión

Se ha conseguido comprobar de forma satisfactoria que tanto el peso de cada rata, al inicio del ensayo y al comienzo del tratamiento; así como la duración del tratamiento, influyen de forma significativa en la calidad ósea final. En cuanto a la duración también se ha podido medir la fuerza con la que ésta influye en la calidad ósea.

También se ha demostrado que ciertos grupos de control se separan perfectamente de las ratas a las que sí se les ha suministrado el fármaco, lo cual muestra su eficacia. Sin embargo, no se ha conseguido diferenciar el resto de ratas por dosis.

Por último, se ha encontrado un rango de valores mínimo y máximo para la dosis administrada tanto para 14 como para 21 días que garantizan una calidad ósea relevante.

Para mejorar los resultados, se ha intentado tratar los valores extremos de distintas formas (imputación con la media, KNN, ...), pero los resultados obtenidos han sido peores en todos los casos.

Se han encontrado muchas dificultades al trabajar con este conjunto de datos, ya que la cantidad de registros es muy escasa. Además, se utilizan muy pocas dosis y en la mayoría de casos, cada dosis solo se prueba con una única duración, complicando la comparación.

Con el fin de obtener mejores resultados, sería aconsejable realizar nuevos estudios con un mayor número de dosis y duraciones.