

MASTER UNIVERSITARIO

BIOESTADÍSTICA Y BIOINFOMÁTICA

*Análisis exploratorio multivariante para el estudio
de la distribución anormal de grasa, masa magra y
masa ósea en individuos infectados por VIH*

LONGINO JÁCOME PÉREZ

Asesor: Nuria Pérez Álvarez

Profesor responsable: Carles Ventura Royo

Enero de 2019

Índice general

Resumen	III
Abstract	IV
Índice de Figuras	V
Índice de Cuadros	VII
1. Introducción	1
Definición del problema	1
Justificación del problema	1
Objetivos	2
Objetivos generales y específicos	2
Tipo de estudio	2
Variables	2
2. Marco teórico	7
Densitometría con rayos X de doble energía	7
Índice de Masa Corporal	8
Osteoporosis	9
Lipodistrofia	10
Sarcopenia	11
Análisis estadístico	12
Machine Learning	12
Análisis de regresión	13
Análisis de clasificación	15
Análisis longitudinal	17
3. Metodología	19
Datos	19
Análisis metodológico	19
4. Resultados	23
Análisis preliminar	23
Análisis transversal	27
Masa ósea	27
Modelo de masa ósea general	36
Modelo de clasificación masa ósea general	44
Masa grasa	45
Modelo de masa grasa general	47
Modelo de clasificación	55
Masa magra	56

Modelo de regresión	58
Modelo de clasificación	65
Análisis longitudinal masa ósea	66
Modelo de Regresión lineal	67
5. Discusión	77
6. Conclusiones	81
A. Evaluación del plan de trabajo	83
B. Código R	87
C. Datos de NCBI/PUBMED	131
Bibliografía	133

Resumen

En la actualidad uno de los problemas graves de salud pública a nivel mundial es la infección por el Virus de Inmunodeficiencia Humana (VIH). El pronóstico de vida de los pacientes con VIH ha mejorado gracias a los tratamientos antirretrovirales existentes en la actualidad. Debido a ello se presentan otros problemas clínicos que ocasiona el VIH o el tratamiento antirretroviral mismo como osteoporosis, lipodistrofia y sarcopenia.

En este estudio el objetivo es evaluar las variables asociadas a la masa ósea, masa grasa y masa magra y su posible relación con la osteoporosis, lipodistrofia y sarcopenia en pacientes infectados por el VIH. Se utilizan técnicas de aprendizaje supervisado para obtener las relaciones entre las variables de interés en el estudio: técnicas de regresión y de clasificación. Se realiza un análisis longitudinal de datos desbalanceados a través de modelos lineales mixtos para modelar la relación entre los individuos y el tiempo de infección.

La base de datos consta de 1,475 individuos algunos de ellos con más de una dexa. Se consideran las variables densidad de masa ósea total, masa grasa total y masa magra total como variables de respuesta. Se caracteriza la presencia de osteoporosis a través de la variable T, los índices de masa grasa y de masa magra para lipodistrofia y la sarcopenia respectivamente.

El 65 % de la población de hombres presenta osteoporosis. En las mujeres, el 60 % presenta esta condición. El modelo de masa ósea para hombres explica el 33 % de los datos y el de mujeres el 37 %. Este modelo considera como covariantes las variables de masa grasa y masa magra. El 96.5 % de los hombres tiene la condición de lipodistrofia así como el 99.7 % de las mujeres. El modelo de regresión explica el 97.5 % de variabilidad tanto para hombres como para mujeres y considera como covariantes a las variables óseas y magras. El 11.8 % de los hombres tiene sarcopenia según el criterio tomado para definirla. El 0.6 % de las mujeres presenta esta condición. El modelo de regresión para el total de masa magra tiene una r^2 de 0.98 para hombres y 0.99 para mujeres. Para estos modelos las covariantes son las variables óseas y grasas.

El modelo de clasificación óseo, basado en el método kNN tiene una precisión del 60 % para hombres y 62 % para mujeres. El modelo correspondiente para lipodistrofia arroja una precisión del 96 % para hombres y 100 % para mujeres. En el caso de la masa magra se obtiene una precisión del 0.92 para hombres y 1 para mujeres. En cuanto al modelo longitudinal se observa como tendencia general la disminución de la masa ósea con el tiempo, sobre todo en aquellas observaciones con más de dos dexas, tanto para hombres como para mujeres.

Los resultados obtenidos en este trabajo en cuanto a los índices no difieren en gran medida de los obtenidos para otras poblaciones con VIH. Los porcentajes de individuos con osteoporosis son consistentes con los reportados por la literatura. Los valores de lipodistrofia obtenidos son muy altos. Los valores de sarcopenia resultan muy diferentes a los reportados por lo que deben ser revisados. A la luz de los resultados obtenidos se plantea la necesidad de implementar los análisis por edad, revisar puntos de corte y asociación de variables para eliminar problemas de colinealidad.

Palabras clave: Osteoporosis, lipodistrofia, sarcopenia, VIH, DEXA, Análisis multivariado, regresión lineal, modelo longitudinal

Abstract

Nowadays, one of the greatest problems in public health around the world is the infection by the human immunodeficiency virus (HIV). The life forecast for the patients with HIV has improved thanks to the existing antiretroviral treatments. But with this, other clinical problems appear due to the HIV or because of antiretroviral treatments itself, as osteoporosis, lipodystrophy and sarcopenia.

The objective of this study is to evaluate the associated variables to the bone mass, fat mass and lean mass, and its possible relation with osteoporosis, lipodystrophy, and sarcopenia in patients infected with HIV. Supervised learning techniques are used to obtain the relationship between the variables of interest in the study: regression and classification techniques. A longitudinal analysis of unbalanced data is done through mixed linear models in order to model the relationship between individuals and the time or infection.

The database includes 1,475 individuals some of which have more than one dexta. The database was divided in men and women in order to do the analysis separately. The variables density of total bone mass, total fat mass and total lean mass were considered as response variables. The criterion of the World Health Organization was taken to characterize the presence of osteoporosis through the T brand. The indexes of fat mass and lean mass were considered to evaluate the lipodystrophy and the sarcopenia.

The 65 % of men population presents osteoporosis. For women, 60 % presents such condition. The model of bone mass for men explains the 33 % of the data and for women the 37 %. The 96.5 % of men has the lipodystrophy condition and 99.7 % in women. The regression model explains the 97.5 % of the data for both men and women. The 11.8 % of men has sarcopenia according to the defined criteria. The 0.6 % of women presents this condition. The regression model for the total lean mass has an r^2 of 0.98 for men and 0.99 for women.

The bone classification model, based on the knn method, has an accuracy of 60 % for men and 62 % for women. The corresponding model for lipodystrophy shows an accuracy of 96 % for men and 100 % for women. In the case of lean mass, we got an accuracy of 0.92 for men and 1 for women. Regarding the longitudinal model, the decrease in bone mass with time is observed as a general trend, specially in those observations with more than two dexas, for both men and women.

The results obtained in this study in terms of the indices, do not differ greatly from those obtained for other populations with HIV. The percentages of individuals with osteoporosis are consistent with those reported by the literature. The lipodystrophy values obtained are very high, perhaps due to the cutoff points used. The sarcopenia values are very different from those reported, so they should be reviewed. Considering the result obtained it is necessary to perform analysis by age, review cutoff points and association of variables to eliminate collinearity problems.

Keywords: Osteoporosis, lipodistrofia, sarcopenia, VIH, DEXA, multivariate analysis, regression model, longitudinal model

Índice de figuras

2.1. Ejemplo de una dexa (Zerbi, 2017)	8
2.2. Clasificación del Machine Learning	13
2.3. Matriz de confusión	16
4.1. Análisis descriptivo de las variables del estudio Masculino	25
4.2. Análisis descriptivo de las variables del estudio Femenino	26
4.3. Valores promedio variables oseas T masculino	31
4.4. Valores promedio variables oseas T femenino	32
4.5. Masa ósea triangulo de Ward masculino	33
4.6. Masa ósea triangulo de Ward femenino	34
4.7. Distribucion de masa osea	35
4.8. Residuales del modelo lineal óseo reducido hombres	37
4.9. Distribución de la masa ósea total Hombres	38
4.10. Análisis de Residuales del modelo óseo hombres	40
4.11. Residuales del modelo lineal óseo reducido mujeres	42
4.12. Distribución de la masa ósea total mujeres	43
4.13. Análisis de Residuales del modelo óseo mujeres	44
4.14. Distribución masa grasa	48
4.15. Residuales de modelo lineal masa grasa hombres	50
4.16. Residuales de modelo lineal masa grasa mujeres	53
4.17. Análisis de residuales modelo masa grasa mujeres	54
4.18. Distribución masa magra	56
4.19. Residuales de modelo lineal reducido masa magra hombres	59
4.20. Análisis de residuales	60
4.21. Distribución de la masa ósea total mujeres	61
4.22. Residuales de modelo lineal masa magra mujeres	63
4.23. Análisis de residuos	64
4.24. Número de DEXAs registradas para un mismo paciente	66
4.25. Perfiles individuales del 20 % de las observaciones seleccionados al azar. Masa ósea	68
4.26. Regresión modelo longitudinal hombres	70
4.27. Análisis de residuales modelo longitudinal hombres	72
4.28. Regresión modelo longitudinal hombres	74
4.29. Análisis de residuales modelo longitudinal mujeres	76
A.1. Calendario	84
A.2. Hitos	85

Índice de cuadros

1.1. Variables del estudio	3
1.2. Parámetros generales evaluados en la DEXA	5
2.1. Valores t para definir osteoporosis	7
2.2. Criterios de la SEEDO	9
3.1. Principales indicadores del estudio	20
3.2. Variables de respuesta	20
3.3. Datos para estudio longitudinal	21
4.1. Resumen estadístico general de los datos del estudio por sexo	23
4.2. Resumen estadístico de los datos del estudio (Masculino)	24
4.3. Resumen estadístico de los datos del estudio (Femenino)	25
4.4. Pruebas de normalidad hombres	27
4.5. Pruebas de normalidad mujeres	27
4.6. Variables del estudio óseo	27
4.7. Matriz de correlaciones para variables óseas (masculino)	28
4.9. Matriz de correlaciones para variables óseas (femenino)	29
4.13. Correlaciones para la masa ósea total en hombres	36
4.14. Coeficientes del modelo lineal óseo para hombres	37
4.15. Test de Kolmogorov-Smirnov densidad masa ósea hombres	38
4.16. Modificación de Lillefors densidad masa ósea hombres	39
4.17. Correlaciones para la masa ósea total en hombres	39
4.18. Coeficientes del modelo lineal óseo para mujeres	41
4.19. Test de Kolmogorov-Smirnov densidad masa ósea mujeres	41
4.20. Modificación de Lillefors densidad masa ósea mujeres	41
4.21. Tabla de clasificación ósea para hombres método kNN (k=21, accuracy=0.60)	44
4.22. Tabla de clasificación ósea para mujeres método kNN (k=21, accuracy=0.62)	45
4.23. Variables consideradas para el análisis de masa grasa	46
4.24. Pruebas de normalidad para la variable grasa (TotalFg) en hombres . . .	46
4.25. Pruebas de normalidad para la variable grasa (TotalFg) en mujeres . . .	47
4.26. Correlaciones para la masa grasa total en hombres	47
4.27. Coeficientes del modelo lineal masa grasa para hombres	49
4.28. Correlaciones para la masa grasa total en mujeres	51
4.29. Coeficientes del modelo lineal masa grasa mujeres	52
4.30. Tabla de clasificación para hombres masa grasa método kNN (k=21, accuracy=0.96)	55
4.31. Tabla de clasificación para mujeres masa grasa método kNN (k=21) . . .	55
4.32. Variables independientes para el modelo de masa magra	57

4.33. Pruebas de normalidad para la variable magra (TotalLg) en hombres . . .	57
4.34. Pruebas de normalidad para la variable magra (TotalLg) en mujeres . . .	58
4.35. Correlaciones para la masa magra total en hombres	58
4.36. Coeficientes del modelo lineal graso en hombres	59
4.37. Correlaciones para la masa magra total en mujeres	61
4.38. Coeficientes del modelo lineal graso en mujeres	62
4.39. Tabla de clasificación para hombres magra método kNN (k=21, accuracy=0.917)	65
4.40. Tabla de clasificación para mujeres magra método kNN (k=21)	65
4.41. Número de DEXAs registradas para un mismo paciente	67
4.42. Coeficientes fijos del modelo longitudinal óseo para hombres	69
4.43. Efectos aleatorios del modelo longitudinal óseo para hombres	69
4.44. Coeficientes fijos del modelo longitudinal óseo para mujeres	73
4.45. Efectos aleatorios del modelo longitudinal óseo para mujeres	73
C.1. Datos bibliograficos de NCBI/PUBMED 2000-2018	132

Capítulo 1

Introducción

Definición del problema

Este trabajo busca definir las variables (o combinación de ellas) que mejor resuman la información de pacientes que presentan anormalidad en la distribución de grasa, masa magra y hueso.

Justificación del problema

Según la Organización Mundial de la Salud (OMS, 2018) el VIH es uno de los mayores problemas para la salud pública mundial. Desde su descubrimiento han fallecido más de 35 millones de personas. En 2016 un millón de personas fallecieron por causas relacionadas con este virus. Cada año se detectan 1.8 millones de nuevas infecciones.

Africa es la región más afectada con el 60 % de los casos registrados. El total mundial en 2016 se estimaba en cerca de los 37 millones de casos.

El virus de la inmunodeficiencia humana (VIH) ataca el sistema inmunitario y debilita los sistemas de defensa contra las infecciones y contra determinados tipos de cáncer. A medida que el virus destruye las células inmunitarias e impide el normal funcionamiento de la inmunidad, la persona infectada va cayendo gradualmente en una situación de inmunodeficiencia. La función inmunitaria se suele medir mediante el recuento de linfocitos CD4.

La inmunodeficiencia aumenta el riesgo de contraer numerosas infecciones, cánceres y enfermedades que las personas con un sistema inmunitario saludable pueden combatir.

La fase más avanzada de la infección por el VIH es el Síndrome de inmunodeficiencia adquirida o sida que, en función de la persona, puede tardar de 2 a 15 años en manifestarse. Las personas que padecen sida pueden contraer ciertos tipos de cáncer e infecciones o presentar otras manifestaciones clínicas de gravedad.

A medida que la infección va debilitando el sistema inmunitario, la persona puede presentar otros signos y síntomas, como inflamación de los ganglios linfáticos, pérdida de peso, fiebre, diarrea y tos. En ausencia de tratamiento pueden aparecer enfermedades

graves como tuberculosis, meningitis criptocócica, infecciones bacterianas graves o cánceres como linfomas o sarcoma de Kaposi, entre otros.

Los medicamentos contra el VIH están ayudando a las personas a vivir más, y la edad avanzada agrava el riesgo de enfermedades propias de la vejez, como osteoporosis, lipodistrofia y sarcopenia (Salud y Servicios Humanos de EU, 2017).

Objetivos

Objetivos generales y específicos

1. Caracterizar el estado de salud de los pacientes en cuanto a las variables contenidas en el DEXA
 - 1.1 Describir las características de los pacientes incluidos en el estudio por sexo
2. Identificar mediante métodos estadísticos las variables asociadas a la osteoporosis, lipodistrofia y sarcopenia en pacientes infectados por el VIH
 - 2.1 Analizar las variables que describen la osteoporosis
 - 2.2 Analizar las variables que describen la sarcopenia
 - 2.3 Analizar las variables que describen la lipodistrofia
3. Describir la evolución en el tiempo de osteoporosis en aquellos pacientes en los que se disponga de la información necesaria
 - 3.1 Describir la evolución en el tiempo de la densidad mineral ósea

Tipo de estudio

- Estudio transversal descriptivo de una serie de casos de pacientes con VIH
- Estudio longitudinal descriptivo de una serie de pacientes con VIH con más de una medida

Variables

La densitometría axial (DEXA) permite llevar a cabo un análisis de la composición corporal para cuantificar la cantidad de masa magra, grasa y densidad mineral ósea. Existen dos tipos de estudios densitométricos: la densitometría axial que se utiliza para evaluar la densidad de masa ósea (DMO) y la densitometría de cuerpo entero, que se emplea para analizar la composición corporal (Lorente Ramos *et al.*, 2012).

Las variables para este estudio se resumen en el cuadro 1.1. El archivo consta de 80 variables, divididas en las que caracterizan a los individuos estudiados (ID, género, fecha de nacimiento, peso, talla), variables relacionadas con la densidad ósea, la masa grasa y la masa magra. La columna contexto clasifica estas variables.

Cuadro 1.1: Variables del estudio

Variable	Descripcion	Contexto
ID	Identificador	
DateBirth	Fecha de nacimiento	
age_previous	Edad al estudio	
gender	Genero (H: hombre, D: Mujer)	
gender_num	Genero (1: hombre, 2: Mujer)	
DataAnalisi	Fecha del estudio	
IdCentre	ID Clinica	
TotalBMD	Densidad Masa Osea Total	óseo
Height	Talla (m)	
Weightg	Peso (g)	
RAFp	Right Arm Fat Percentage	Masa grasa
RAFg	Right Arm Fat Grams	Masa grasa
RALg	Right Arm Lean Grams	Masa magra
LAfP	Left Arm Fat Percentage	Masa grasa
LAfG	Left Arm Fat Grams	Masa grasa
LALg	Left Arm Lean Grams	Masa magra
BothAFp	Both Arm Fat Percentage	Masa grasa
BothAFg	Both Arm Fat Grams	Masa grasa
BothALg	Both Arm Lean Grams	Masa magra
RLFp	Right Leg Fat Percentage	Masa grasa
RLFg	Right Leg Fat Grams	Masa grasa
RLLg	Right Leg Lean Grams	Masa magra
LLFp	Left Leg Fat Percentage	Masa magra
LLFg	Left Leg Fat Grams	Masa grasa
LLLg	Left Leg Lean Grams	Masa magra
BothLFp	Both Leg Fat Percentage	Masa grasa
BothLFg	Both Leg Fat Grams	Masa grasa
BothLLg	Both Leg Lean Grams	Masa magra
TFp	Trunk Fat Percentage	Masa grasa
TFg	Trunk Fat Grams	Masa grasa
TLg	Trunk Lean Grams	Masa magra
TotalFp	Total Fat Percentage	Masa grasa
TotalFg	Total Fat Grams	Masa grasa
TotalLg	Total Lean Grams	Masa magra
L1BMD	Lumbar 1 Densidad masa osea	óseo
L1T	Lumbar 1 variable T	óseo
L1Z	Lumbar 1 Marca Z	óseo
L2BMD	Lumbar 2 Densidad masa osea	óseo
L2T	Lumbar 2 variable T	óseo
L2Z	Lumbar 2 Marca Z	óseo
L3BMD	Lumbar 3 Densidad masa osea	óseo
L3T	Lumbar 3 variable T	óseo
L3Z	Lumbar 3 Marca Z	óseo
L4BMD	Lumbar 4 Densidad masa osea	óseo
L4T	Lumbar 4 variable T	óseo

Variable	Descripcion	Contexto
L4Z	Lumbar 4 Marca Z	óseo
L1L4BMD	Lumbar 1-Lumbar 4 Densidad masa osea	óseo
L1L4T	Lumbar 1- Lumbar 4 variable T	óseo
L1L4Z	Lumbar 1-Lumbar 4 Marca Z	óseo
L2L4BMD	Lumbar 2-Lumbar 4 Densidad masa osea	óseo
L2L4T	Lumbar 2- Lumbar 4 variable T	óseo
L2L4Z	Lumbar 2-Lumbar 4 Marca Z	óseo
NeckFBMD	Cuello Femoral Densidad masa osea	óseo
NeckFT	Cuello Femoral variable T	óseo
NeckFZ	Cuello Femoral Marca Z	óseo
WardsBMD	Triangulo de Ward DMO	óseo
WardsT	Triangulo de Ward variable T	óseo
WardsZ	Triangulo de Ward Marca Z	óseo
TrochBMD	Trocanter DMO	óseo
TrochT	Trocanter variable T	óseo
TrochZ	Trocanter Marca Z	óseo
TotalFBMD	Total Femur bone mineral density	óseo
TotalFT	Total Femur variable T	óseo
TotalFZ	Total Femur Marca Z	óseo
coment_cetir	Comentario	
Weight	Peso (kg)	
BMI	Indice de Masa Corporal (kg/m ²)	
FMI	Indice de Masa Grasa (kg grasa /m ²)	Masa grasa
FFMI	Indice de Masa Magra (kg magra /m ²)	Masa magra
Apendicularleanmas	Masa magra apendicular /talla ²	Masa grasa
FMR	Porcentaje de Masa Grasa (% grasa tronco / % grasa piernas)	Masa grasa
FTrunkgFLegsg	Gramos de grasa tronco /grasa piernas	Masa grasa
Indexdistributionfat	Indice de distribucion de masa grasa (gramos grasa troncar /gramos de grasa extremidades)	Masa grasa
FtrunkpFlimbsp	% grasa tronco / % grasa extremidades	Masa grasa
FtrunkgFtotalg	Gramos de grasa tronco /grasa total	Masa grasa
FLegsgFtotalg	Gramos grasa en piernas / grasa total	Masa grasa
FlimbspFtotalg	Gramos grasa en extremidades / grasa total	Masa grasa
LLegFgBMI	Grasa pierna izquierda /BMI	Masa grasa
LLegFpBMI	Porcentaje de grasa pierna izquierda / BMI	Masa grasa
age	Edad	

Es importante resaltar 4 variables que resumen la información anterior (cuadro 1.2). Son los índices de masa corporal (BMI), el índice de masa grasa (FMI), el índice de masa magra (FFMI) y la densidad de masa ósea (BMD). Estos índices se calculan a partir de medidas resultantes de la DEXA (BMI, FMI, FFMI) o como la suma de algunas variables (BMD). Estos índices se podrían utilizar para tener una primera impresión del comportamiento general de los datos.

Cuadro 1.2: Parámetros generales evaluados en la DEXA

Variable	Definición	Inglés	Siglas
DMO	Densidad mineral ósea	Bone mineral density	BMD
IMC	Índice de masa corporal	Body mass index	BMI
IMG	Índice de Masa Grasa	Fat mass index	FMI
IMM	Índice de Masa Magra	Fat free mass index	FFMI

La densidad de masa ósea total (TotalBMD) es un promedio de las diferentes mediciones en las regiones del cuerpo. El índice de masa corporal (BMI) se calcula como el peso por metro al cuadrado ($\frac{kg}{m^2}$), el índice de masa grasa (FMI) son los kilogramos de grasa por metro al cuadrado ($\frac{kg \text{ masa grasa}}{m^2}$) y el índice de masa magra son los kilogramos de masa magra por metro cuadrado ($\frac{kg \text{ masa magra}}{m^2}$).

Capítulo 2

Marco teórico

Densitometría con rayos X de doble energía

La densitometría con rayos X de doble energía (DXA o DEXA) es una técnica utilizada originalmente para diagnosticar osteoporosis. En la práctica clínica las principales modalidades de DEXA son la densitometría ósea axial centrada en la columna lumbar y la cadera para cuantificar la densidad de masa ósea (DMO) y la imagen de cuerpo entero, para analizar la composición corporal total, evaluar pacientes con trastorno ponderal en enfermedades diagnósticas, retraso de crecimiento en pacientes pediátricos, lipodistrofia asociada a infección retroviral, seguimiento de artroplastias o establecer riesgo cardiovascular (Lorente Ramos *et al.*, 2012).

En la densitometría ósea se evalúan diversos parámetros (Cuadro 1.1). Lorente (2012) define el CMO como la cantidad de calcio en una región determinada. La DMO es la cantidad de mineral por unidad de área. Cuando este valor se compara con la base de datos de referencia se obtienen los valores para diagnóstico: las puntuaciones T y Z.

La **densidad mineral ósea** (DMO) se calcula dividiendo el CMO por unidad de superficie ($\frac{g}{cm^2}$). El t-score es un parámetro aceptado por la Organización Mundial de la Salud como criterio de diagnóstico de osteopenia y de osteoporosis en mujeres postmenopausicas y varones mayores de 50 años. Se refiere al número de desviaciones estándar del paciente estándar (sano de 20 años). Regularmente el t-score se obtiene de la medida de DMO de la media de L2-L4 (Figueras-Balsells, 2009). El z-score se utiliza para mujeres premenopáusicas, en varones menores a 50 años y en niños. En este caso, con valores menores a dos desviaciones el diagnóstico es de “densidad ósea baja para la edad” (N.J. Crabtree *et al.*, 2014).

Cuadro 2.1: Valores t para definir osteoporosis

Variable	t-score
Normal	< -1 d. e.
Osteopenia	-1 a -2.5 d. e.
Osteoporosis	> - 2.5 d. e.

La DEXA de cuerpo entero permite conocer la composición corporal además de la densidad mineral ósea en todo el cuerpo (figura 2.1). La distribución de tejidos del organismo incluye la grasa (masa grasa), partes blandas (masa magra) y hueso (DMO) en todo el cuerpo y por regiones. Estima la distribución de grasa corporal, la distribución de la grasa pélvica, la relación de grasa con distribución androide/genoide en la grasa pélvica y el índice de masa corporal (IMC). Lorente (2012) afirma que los valores de DMO de cuerpo entero sirven para estimar la mineralización pero no para diagnosticar osteopenia, para ello se debe utilizar la densitometría axial de la columna y la cadera.

Otra aplicación clínica de la DEXA es su uso en el estudio de lipodistrofias y lipoatrofias en pacientes con infección por VIH (Yang, Zhu & Paton, 2004)

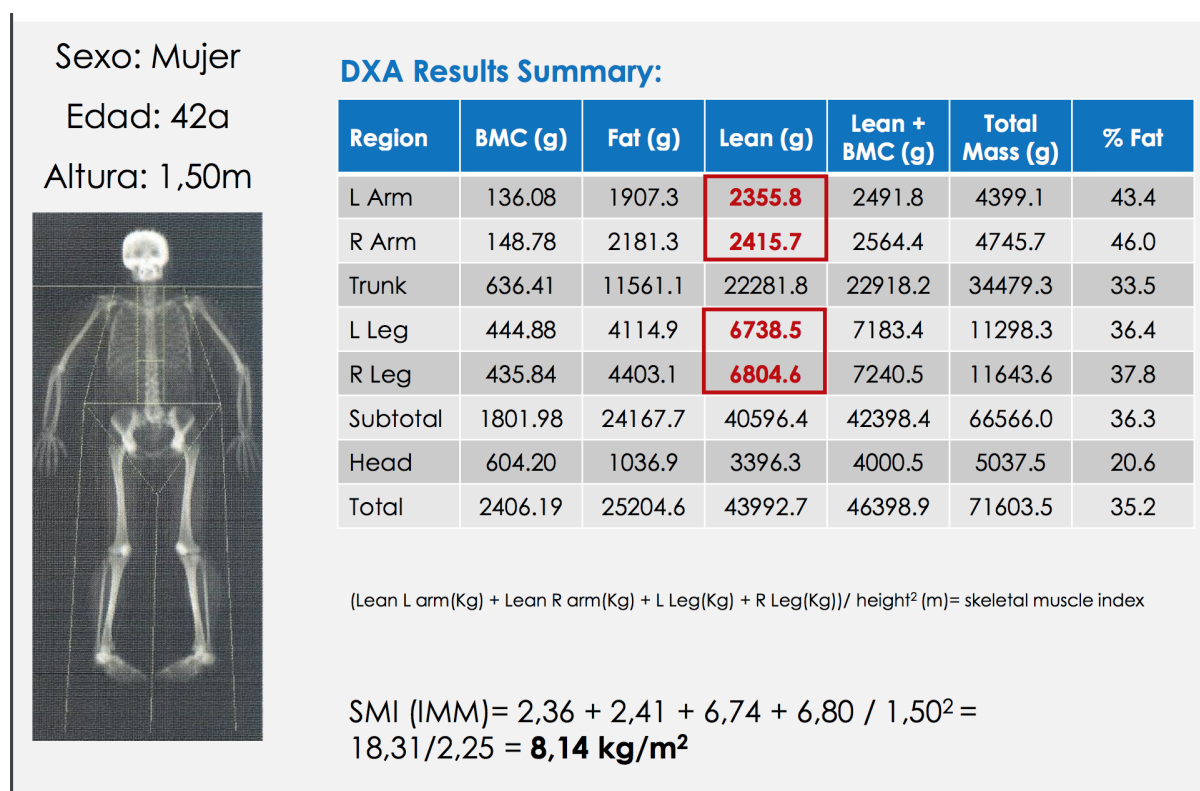


Figura 2.1: Ejemplo de una dexa (Zerbi, 2017)

Indice de Masa Corporal

El índice de masa corporal (IMC) mide la relación entre peso y talla $\frac{\text{peso(Kg)}}{\text{talla}^2(\text{m})}$. El IMC se correlaciona con el porcentaje de grasa corporal. En función del porcentaje de grasa corporal se considera que un individuo es normal con valores entre 12 al 20 % en hombres y del 20 al 30 % en mujeres adultas (Rubio, Salas-Salvadó, Barbany & Moreno, 2007).

Los criterios de la SEEDO (Sociedad Española para el Estudio de la Obesidad) para definir la obesidad en grados según el IMC en adultos se muestran en la tabla 2.2

Cuadro 2.2: Criterios de la SEEDO

Categoría	IMC
Peso insuficiente	< 18.5
Peso Normal	18.5 - 24.9
Sobrepeso	25.0 - 26.9
Preobesidad	27.0 - 29.9
Obesidad I	30.0 - 34.9
Obesidad II	35.0 - 39.9
Obesidad mórbida	40.0 - 49.9
Obesidad extrema	> 50

El IMC no es el mejor indicador de adiposidad en individuos con alto desarrollo muscular o en ancianos, sin embargo es el más utilizado y recomendado por diferentes sociedades médicas para uso clínico por su facilidad de uso y su capacidad de representar a la mayoría de la población.

En una persona sana se pueden definir cambios en la composición corporal en tres etapas según (González-Pérez, 2012):

- **infancia**, donde hay un mayor crecimiento de las extremidades inferiores en relación al tronco con gran mineralización del hueso. El contenido de agua en la composición corporal es hasta del 70 %.
- **adolescencia**, donde hay una aceleración del crecimiento en longitud y un aumento en la masa corporal total. Entre los 30-35 años se presenta el pico de densidad mineral ósea y a partir de este momento empieza a descender. En las mujeres este descenso se precipita con la aparición de la menopausia. El contenido de agua en la composición corporal representa el 60 %. En esta etapa la masa muscular alcanza su pico máximo entre los 25-35 años.
- **a partir de los 40 y 50 años**, se incrementa la masa magra en ambos sexos, produciéndose cambios en la distribución. El contenido de agua corporal puede llegar a un 50 %. Estas pérdidas de agua están en relación con las pérdidas de masa grasa. A partir de este momento el músculo empieza a declinar a un ritmo anual del 1-2 %. Este ritmo aumenta hasta en un 1.5 % entre los 50-60 años y un 3 % a partir de los 60 años. Una persona de 85 años tendría una reducción del 25 % con respecto a uno de 45 años.

Osteoporosis

La osteoporosis es una enfermedad que debilita los huesos y causa fracturas con facilidad. Aumenta el riesgo de fracturas de la cadera, la columna vertebral y la muñeca. Esta reducción del tejido óseo junto con la alteración de la microarquitectura del hueso conlleva una disminución de la resistencia ósea. Cuando se rompe el equilibrio del remodelado óseo y disminuye la mineralización ósea aparece la osteopenia que puede progresar a osteoporosis (Clay, Voss, Williams & Daume, 2008).

En la población general existen numerosos factores de riesgo para el desarrollo de osteoporosis. La osteoporosis es una enfermedad multifactorial relacionada con la edad y el sexo. Los adultos mayores de 50 años presentan mayor frecuencia de osteopenia u osteoporosis, que se puede considerar un nivel avanzado de la enfermedad. Las mujeres menopausicas es otro predisponente importante (Carrero & Ryan, 2009). Otros factores importantes son la raza, la toma prolongada de esteroides, escasa ingesta de calcio, consumo de tabaco y abuso de alcohol, sedentarismo y una dieta hiperproteica (Lambrou, Papadopoulos & Gammerman, 2013).

Los factores que pueden agravar el riesgo de osteoporosis y osteopenia en las personas con el VIH incluyen la infección propiamente dicha y algunos medicamentos empleados para tratarla. A pesar de que la causa de la reducción ósea no está definida con claridad, se cree que puede formar parte de la toxicidad mitocondrial de algunos antirretrovirales, aunque también se ha contemplado la posibilidad de la propia infección por VIH como causante de manera independiente al tratamiento (Clay *et al.*, 2008).

(Carrero & Ryan, 2009) reconoce la importancia de los tratamientos antirretrovirales como agentes etiológicos responsables de la pérdida de masa ósea, particularmente el tenofovir de la familia de inhibidores de la transcriptasa inversa. Cazanave (2008) afirma que la osteoporosis es 3.7 más frecuente en individuos con VIH, especialmente en hombres de mediana edad (43 años) que en individuos sanos. En el caso de las mujeres, reporta una edad promedio de 41 años.

Lipodistrofia

La lipodistrofia es una enfermedad caracterizada por cambios en la forma del cuerpo (cambios morfológicos) debido a la acumulación de grasa en ciertas áreas del cuerpo (lipohipertrofia), pérdida de grasa en ciertas áreas (lipoatrofia) o una combinación de ambos. Algunas veces van acompañados de cambios en los niveles de azúcar y grasa en la sangre (cambios metabólicos) (AIDS Information, 2002).

Se sabe que las lipodistrofias son un grupo heterogéneo de trastornos de etología diversa. Se clasifican como congénitas (de origen genético o metabólico) y adquiridas (por la infección del VIH y/o tratamientos antirretrovirales) (Vilar, 2003).

Al parecer, cualquier combinación de fármacos antirretrovirales puede ocasionar lipodistrofia, pero en particular se mencionan la terapia atiretroviral de gran actividad (TARGA) y los inhibidores de proteasa (IP). Los trastornos metabólicos asociados son la hiperinsulinemia, la hipertigliceridemia, hipercolesterolemia y la resistencia a la insulina (Lapphra, Vanprapar, Phongsamart, Chearskul & Chokephaibulkit, 2005). La lipohipertorfia se observa en el abdomen, las mamas y la región dorsocervical (giba de búfalo). La lipoatrofia se observa característicamente en la cara, las extremidades y las nalgas (Vilar, 2003). Estos cambios se corroboran mediante la absorciometría de rayos X de energía dual (DEXA). Se ha reportado que al menos una anormalidad física de las ya mencionadas se observa en el 50 % de los pacientes después de los 12-18 meses de tratamiento. Los signos metabólicos se han reportado desde un 28 % para la hipertrigliceridemia y hasta un 57 % para la hipercolesterolemia. No se conocen los mecanismos patogénéticos responsables de la lipodistrofia y de las manifestaciones metabólicas asociadas en pacientes con SIDA (Vilar, 2003).

Finkelstein (2015) estudia los factores de riesgo para lipodistrofia en pacientes con VIH y observa que el uso de antiretrovirales y ser mujer son los de mayor importancia. En el caso de lipoatrofia además añade la edad (mayores de 40) como factor adicional.

González (2012) afirma que el síndrome de lipodistrofia es un efecto secundario no deseable de la aparición de antirretrovirales debido al aumento en la supervivencia de los pacientes con VIH.

La grasa corporal se cuantifica a partir de la grasa total, la grasa en extremidades y la grasa en el tronco. También se suele determinar el porcentaje de grasa, ya sea en tronco y/o en las extremidades. Con estas mediciones se determina el **índice de masa grasa** o IMG mediante el cociente entre el porcentaje de grasa del tronco dividido por el porcentaje de grasa en las extremidades. Aunque no existe un punto de corte para definir lipodistrofia González (2012) la define con un $IMG \geq 1.5$ para hombres y ≥ 1.2 para mujeres.

Sarcopenia

La sarcopenia se refiere a la pérdida de masa y potencia muscular que se presenta durante el envejecimiento. Esta pérdida es universal, ocurre siempre con el paso de los años. A diferencia de la osteopenia en la sarcopenia no hay un nivel de pérdida a partir del cual se pueda considerar a una persona como sarcopénico (Rexach, 2006).

González (2012) sugiere que la sarcopenia es de origen multifactorial, siendo los factores más importantes la desnutrición, la disfunción neuromuscular y mitocondrial y los cambios hormonales.

La sarcopenia se ha identificado como una complicación neuromuscular asociada a la infección por VIH desde principios de los 80 (González-Pérez, 2012).

No queda claro el mecanismo por el cual los pacientes con infección por VIH tienen una mayor pérdida de masa muscular. Se desconoce si se trata de un efecto directo del virus o si es secundario al tratamiento antirretroviral o a otros factores (González-Pérez, 2012).

La masa muscular en el DEXA de cuerpo completo es equiparable a la variable **masa magra**, con el supuesto de que incluye no solo músculo, sino también tejidos no musculares no grasos (ligamentos, tendones, articulaciones, piel y vísceras).

González (2012) define dos formas de cuantificar la sarcopenia: **Sarcopenia ajustada a talla** o **sarcopenia cruda** y **sarcopenia ajustada a grasa** o **sarcopenia por residuos**.

Para cuantificar la masa muscular se mide la masa magra en las extremidades (masa muscular apendicular), en el tronco (masa muscular troncular) y la masa muscular total. Esta variable se expresa en gramos. Ahora, dividiendo la masa muscular en las extremidades por la talla del individuo al cuadrado se obtiene el índice de masa muscular (IMM en $\frac{kg}{m^2}$).

La sarcopenia cruda se define cuando el individuo presenta un IMM de dos desviaciones estándar por debajo de la masa esquelética máxima esperada en adultos de 35 años para cada sexo. En el caso de hombres este límite es de $7.26 \frac{kg}{m^2}$ y para mujeres es de $5.5 \frac{kg}{m^2}$.

La sarcopenia por residuos se define de acuerdo al modelo de regresión lineal que predice la masa muscular esquelética (en cm) y de la grasa corporal total (en kg). Cualquier valor por debajo del percentil 20 indica sarcopenia.

Análisis estadístico

El Análisis Exploratorio es un conjunto de procedimientos que tienen como objetivo proporcionar una visión general de las variables del estudio. Se apoya en un primer planteamiento descriptivo y se realiza sin ideas preconcebidas sobre el contenido de los datos con una mentalidad netamente exploratoria (Salgado, 1992).

Se constituye por un conjunto de procedimientos analíticos descriptivos y gráficos que muestran las particularidades de las variables sacando a la luz la estructura de los datos.

La aplicación del Análisis Exploratorio previo al análisis de los datos tiene como objetivos (Salgado, 1992):

1. Estudiar la tendencia, distribución y forma de cada uno de los indicadores incluidos en el modelo matemático.
2. Determinar la normalidad sobre el conjunto de datos.
3. Encontrar datos extremos, outliers o atípicos.
4. Obtener descripciones, comprobar supuestos y caracterizar diferencias entre subpoblaciones.
5. Ayuda a determinar si son adecuadas las técnicas estadísticas que se van a considerar en el análisis posterior.
6. Determina la necesidad de transformar los datos cuando no se observa una distribución normal, o bien, puede ayudar a decidir el uso de pruebas no paramétricas.

Machine Learning

Machine learning o aprendizaje automático es una disciplina de la inteligencia artificial. El objetivo es desarrollar sistemas que sean capaces de encontrar patrones complejos en grandes conjuntos de datos por si solos (Aldás & Uriel, 2017).

Los algoritmos de machine learning se clasifican en dos grupos: **Algoritmos supervisados**, que aplican lo que se ha aprendido con los datos históricos para sacar conclusiones sobre nuevos datos y **algoritmos no supervisados**, que extraen inferencias de conjuntos de datos (figura 2.2).

El aprendizaje supervisado es un conjunto de técnicas que permiten realizar predicciones futuras basadas en comportamientos o características analizadas en datos históricos, representada por medio de una función donde las entradas representan las características analizadas y la salida representa la variable que se va a predecir. Esta función es de tipo numérico en problemas de **regresión** y de tipo categórico en los problemas de **clasificación** (Zumel & Mount, 2014).

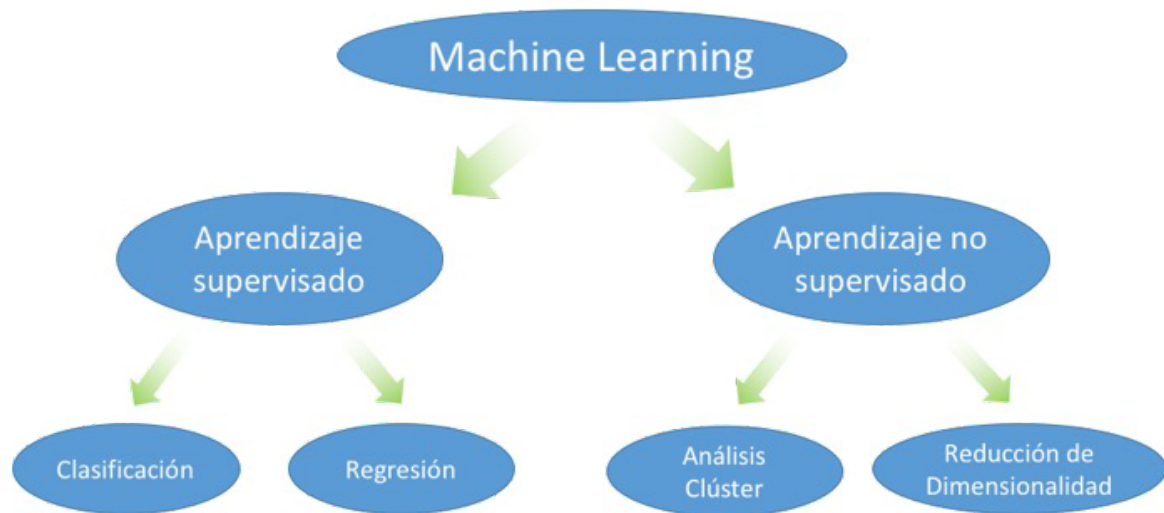


Figura 2.2: Clasificación del Machine Learning

Análisis de regresión

El análisis de regresión se usa para modelar la relación entre una variable continua llamada variable de respuesta o variable dependiente y una o más variables continuas llamadas variables explicativas o independientes.

El modelo estadístico general se define como:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_{kx}k_i + e_i$$

para $i = 1, 2, \dots, n$

Considerada esta ecuación como un modelo estadístico general, se tiene que cualquier observación es la suma de dos partes o componentes: una parte fija o determinista, $f(x)$, y una parte aleatoria desconocida, $g(x)$.

$$y = f(x) + g(x)$$

o planteado en términos del modelaje estadístico:

$$\text{Respuesta} = \text{modelo} + \text{incertidumbre}$$

Como cualquier modelo estadístico existe una hipótesis nula que se ha de probar. En el caso del modelo lineal la hipótesis es:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Si $H_0 : \beta_i = 0$, entonces no hay relación lineal entre las variables explicativas y de respuesta.

En caso contrario, si $H_1 : \beta_i \neq 0$, se infiere que hay una relación lineal entre ambas variables.

Hipótesis del modelo de regresión lineal

H1: *Linealidad en los parámetros:* Debe existir linealidad en los parámetros (independientemente de la linealidad en las variables).

H2: *Especificación correcta:* Supone que las variables explicativas del modelo son aquellas variables relevantes que explican a la variable de respuesta. Y que además están todas. No existe ninguna variable independiente que no explique nada de la dependiente. Si es así, el modelo está bien especificado.

H3: *Grados de libertad positivos:* Los grados de libertad (gl) de un modelo se definen como la diferencia entre el número de datos (n) y el número de variables explicativas (k). Como mínimo se debe disponer de tantos datos como parámetros a estimar. $gl = n - k > 0$.

H4: *Parámetros constantes:* Supone que los parámetros no cambian con el tiempo (con datos temporales) o con el individuo (en datos transversales).

H5: *Independencia lineal entre las variables explicativas:* Cada variable explicativa contiene información adicional sobre la variable de respuesta que no está contenida en otras. En caso de información repetida las variables explicativas serían linealmente dependientes de otras.

H6: *Normalidad:*

H7: *Homocedasticidad:* La varianza de los errores no puede cambiar con el tiempo (o con el individuo) de lo contrario se habla de heteroscedasticidad.

Diversos autores afirman que los modelos de una sola variable dependiente (análisis de regresión, análisis de la varianza y análisis de la covarianza) son robustos bajo la violación de los principios de normalidad y homocedasticidad ya mencionados, excepto cuando las muestras son muy pequeñas y desiguales. Los métodos del análisis multivariante son lo suficientemente robustos como para ser insensibles a desviaciones de los supuestos estadísticos, siempre que se trabaje con muestras grandes ($N > 30$) (Closas, Arriola, Kuc, Amarilla & Jovanovich, 2013).

Las pruebas de hipótesis empleadas para analizar la normalidad de los datos son el test de Kolmogorov-Smirnov y modificación de Lillefors. El test de Kolmogorov-Smirnov permite definir si una muestra procede de una población con una determinada distribución, es decir no está limitado a la distribución normal. Esta prueba asume que se conoce la media y varianza poblacional, lo que en la mayoría de los casos no es posible. Entonces el test toma los valores de la muestra para el cálculo. Sin embargo su uso es muy extendido. Para solventar este problema se desarrolló una modificación de esta prueba conocida como test lillefors, que asume que la media y la varianza son desconocidas. Es también una alternativa al test de Shapiro-Wilk cuando el número de observaciones es mayor de 50. En ambas pruebas se considera como hipótesis nula que los datos sí proceden de una distribución normal y como hipótesis alternativa que no lo hacen.

El estadístico de Jarque-Bera cuantifica que tanto se desvían los coeficientes de asimetría y curtosis de los esperados en una distribución normal. No requiere estimaciones de los parámetros que caracterizan la normal.

Cuando se tienen modelos con un gran número de variables explicativas puede ocurrir que dichas variables sean redundantes o que inclusive estén correlacionadas entre si. El tener variables correlacionadas en el modelo hace que sea inestable de tal forma que los parámetros del modelo se vuelven imprecisos y sus signos pueden incluso llegar a

ser contrarios. Durante el análisis exploratorio se pueden detectar estas variables con un análisis simple de correlaciones o con un análisis gráfico de dispersión (Salgado, 1992).

Una forma de reducir la dimensionalidad de los datos es mediante el análisis de componentes principales que en términos prácticos resumen la variabilidad de las variables originales al resumirla en vectores ortogonales.

En este trabajo se utiliza el método de selección de variables stepwise selection. Este método se encarga de abordar el problema de la construcción o selección del modelo. Si se incluyen cada vez más variables en un modelo de regresión puede ser que el ajuste mejore pero también aumenta la varianza y por tanto en la función de regresión estimada se produce un sobreajuste. Por otro lado, si se incluyen menos variables de las necesarias en el modelo, las varianzas se reducen pero los sesgos aumentarían obteniéndose una mala descripción de los datos. El objetivo de este método de selección de variables es buscar un modelo tal que se ajuste bien a los datos y que a la vez sea posible buscar un equilibrio entre bondad de ajuste y sencillez (Aachen, Batagelj & Ferligoj, 2006).

El p-valor para cada término comprueba la hipótesis nula de que el coeficiente es igual a cero (no tiene efecto). Un p-valor bajo (< 0.05) indica que puedes rechazar la hipótesis nula.

Análisis de clasificación

Para el análisis de clasificación se utiliza kNN vecinos más próximos (k-nearest neighbors) como método de clasificación supervisada. Pertenecce al grupo de algoritmos de aprendizaje supervisado no paramétrico y se utiliza para el reconocimiento de patrones y extracción de datos. El clasificador kNN se utiliza frecuentemente como base para clasificadores más complejos como las redes neurales, por su robustez y versatilidad (Bresler & Gil, 2011).

Para calcular la similitud, el método se basa en una función de distancia. La más común es la distancia euclídeana para datos continuos:

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}$$

El método incluye un conjunto de entrenamiento con clasificadores conocidos para determinar los parámetros del clasificador, un valor de k, que es el número de elementos para calcular las distancias que se van a tomar y un conjunto de validación para estimar el error de generalización que debe ser pequeño para evitar sobrevalorar la capacidad predictiva del modelo obtenido (sobre entrenamiento).

El resultado del modelo de clasificación se plasma en una matriz de confusión que ayuda a medir el cómo se comporta el modelo cuando se aplica a datos nuevos (Torgo, 2017). La estructura es la siguiente:

- **VP** es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- **VN** es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 2.3: Matriz de confusión

- **FN** es la cantidad de positivos que fueron clasificados incorrectamente como negativos.
- **FP** es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

Existen diferentes metodologías para determinar el valor de k . Se puede calcular la curva ROC y elegir el punto más alto como valor de k (Aachen *et al.*, 2006). Un método más sencillo es hacer varios ensayos del modelo de clasificación con diferentes valores de k y elegir el que tenga la eficiencia más alta (accuracy).

$$Eficiencia = \frac{VP + VN}{VP + VN + FN + FP}$$

La tasa de error del modelo se calcula como:

$$Tasa\ de\ error = \frac{FP + FN}{VP + VN + FN + FP}$$

Supuestos del método kNN (Gironés, Casas, Minguillón & Caihuelas, 2017)

1. Estandarización: Si las variables independientes del conjunto de entrenamiento están en diferentes escalas es necesario estandarizar antes de los cálculos de la distancia. Lo más común es normalizar:

$$x_s = \frac{x_i - \bar{x}}{\sigma}$$

2. Outliers: Los valores bajos de k (número de vecinos más cercanos) es sensible a los datos extremos, mientras que altos valores de k son más resilientes, pero implican más cálculos.

Existen otros métodos de clasificación (k-medias, clustering) pero el método kNN es el más sencillo e intuitivo (Burger, 2018).

Análisis longitudinal

Los estudios longitudinales son aquellos en que los individuos son observados a través del tiempo. La característica específica del estudio longitudinal es que tanto la variable de respuesta como el conjunto de covariables son repetidamente medidas a lo largo del tiempo. También se le conoce como estudio de medidas repetidas cuando las observaciones tienen varias lecturas. (Arnau & Bono, 2008).

Los estudios longitudinales ofrecen la oportunidad de estudiar tanto patrones individuales de cambio sobre el tiempo y sobre las condiciones.

En los estudios longitudinales hay covariables dependientes del tiempo y covariables que no lo son. Si Y es función de una variable X que cambia en el tiempo (témpero-dependiente) y otra Z que es estable en el tiempo (témpero-independiente) la ecuación que describe este modelo es:

$$Y_{it} = \beta_0 + \beta_1 x_{it} + z_i + \epsilon_{it}$$

donde el subíndice i hace referencia al individuo, el t al momento del tiempo y ϵ es un término de error (Z no cambia al ser estable y por eso tiene un solo subíndice) (Delgado-Rodríguez & Llorca, 2004).

La respuesta Y_{it} es observada en t unidades de tiempo para i medidas repetidas de unidades, individuos o sujetos. Los intervalos de tiempo se denominan *puntos de tiempo* y puede variar desde minutos o años. El conjunto de respuesta de las unidades de observación se denomina *perfil de respuesta* o *perfiles individuales* de cada unidad. Arnau (2008), quien define estos términos enfatiza que con estas condiciones un estudio longitudinal es sinónimo de diseño de medidas repetidas, estudios de panel o estudios de cohortes.

Los estudios longitudinales (Arnau & Bono, 2008) tienen varias dificultades. En primer lugar, hay dependencia entre las medidas repetidas de la misma unidad observacional y frecuentemente los datos son no balanceados o incompletos. Los datos no balanceados se refieren a que no hay el mismo número de observaciones entre los individuos.

Se ha sugerido que los estudios longitudinales pueden ser abordados desde cuatro ángulos estadísticos: 1) El análisis por separado de cada punto o intervalo de tiempo, 2) el análisis univariante de la varianza, 3) el análisis multivariante de las variables de contraste temporal y 4) los métodos basados en modelos mixtos (Arnau & Bono, 2008).

El primer método es equivalente a un estudio trasversal ya que no tiene ninguna implicación sobre el efecto de tiempo. Los puntos dos y tres se refieren al análisis de varianza univariante (ANOVA) y al análisis de varianza multivariante (MANOVA). Estas técnicas no consideran el problema de los datos desbalanceados por un lado y la posible correlación entre las medidas repetidas (Arnau & Bono, 2008). Los modelos lineales mixtos consideran la estructura de las covarianzas entre las medidas repetidas y además no es sensible al desbalanceo (Juárez de Galíndez & Balzarini, 2003).

Llorenç, (2011) propone el uso de modelos lineales mixtos, en particular el modelo de interceptos y pendientes aleatorias para el estudio de datos longitudinales. Este es un modelo con un solo factor aleatorio (los individuos denotados por a_i) que pertenecen a una población heterogénea y que presentan una variabilidad inherente (σ_A^2). La variabilidad de estas sub-replicas está condicionada a las variables explicativas. Cada nivel del factor

aleatorio se analiza en determinadas unidades de tiempo, aunque habitualmente están desestructuradas. El modelo asume que la variable objetivo puede describirse a través de un modelo lineal poblacional (efectos fijos) y un patrón individualizado para los efectos aleatorios. El modelo planteado por este autor es el siguiente (usando la notación anterior):

$$Y_{it} = \beta_0 + (\beta_1 + b_i)x_{it} + z_i + \epsilon_{it}$$

donde b_i se refiere al factor aleatorio y $b_i \sim N(0, \sigma_b^2)$, $z_i \sim N(0, \sigma_z^2)$ y $\epsilon_{it} \sim N(0, \sigma^2)$.

Delgado, (2004) destaca que los modelos de efectos aleatorios permiten que cada individuo tenga parámetros de regresión únicos y que los coeficientes del modelo se transforman en una odds ratio referida al individuo, que se asume constante en toda la población.

Capítulo 3

Metodología

Datos

Para este trabajo se cuenta con una base de datos con 4,319 DEXAS. Cada DEXA consta de 87 variables, entre las propias de la densitometría y alguna generales (sexo, edad, talla, peso). Las DEXAS pertenecen a 1,475 pacientes, medidos en diferentes puntos de tiempo aunque los pacientes tienen un número desigual de DEXAS. Dos archivos complementarios consisten en la definición de variables (cuadro 1.1) y el número de DEXA por paciente.

Análisis metodológico

Se utilizó R para el análisis de datos, RStudio para la integración del informe y Rmarkdown para la elaboración del documento. Dado que el archivo de datos es grande se utilizaron sentencias sql para la elaboración de filtros.

Del archivo original se eliminaron 11 variables que o estaban repetidas o no son relevantes en el estudio. Se dividió el estudio en transversal y en longitudinal. Para el estudio transversal se eliminaron todas las observaciones de un mismo individuo dejando solo la última para tener al final una observación por individuo. Se eliminaron 20 registros repetidos (mismo día de toma de información) y se eliminaron 49 registros con algún dato faltante, quedando el estudio transversal con 76 variables y 1,426 observaciones. Para simplificar el análisis se agregó una columna con el factor sexo, ya que el objetivo era realizar el estudio para hombres y mujeres por separado. También se agregó el estado óseo definido por la literatura a partir de la variable T del triangulo de ward, el estado graso a partir del índice de masa grasa y el estado magro, a partir del índice de masa magra.

Para el análisis exploratorio de los datos en esta primera fase se consideró analizar los indicadores que proporciona la muestra como aquellos relevantes para definir el estado de los individuos. Estos indicadores fueron:

Cuadro 3.1: Principales indicadores del estudio

Indice	Definición
TotalBMD	Densidad de masa ósea total
BMI	Indice de Masa Corporal (kg/m ²)
FMI	Indice de Masa Grasa (kg grasa /m ²)
FFMI	Indice de Masa Magra (kg magra /m ²)

Se realizó un análisis numérico básico y gráfico de estas variables para determinar su comportamiento e identificar posibles relaciones, a través de una análisis de correlaciones.

Se determinó que las variables de respuesta o dependientes son las que tienen que ver con la masa total de los componentes a estudiar (cuadro 3.1)

Cuadro 3.2: Variables de respuesta

Indice	Definición
TotalBMD	Densidad de masa ósea total
TotalFg	Masa grasa total
TotalLg	Masa magra total

Se hizo un análisis de normalidad para determinar cuánto difiere la distribución de los datos observados con respecto a lo esperado si estos proceden de una distribución normal. Se utilizaron representaciones gráficas de los datos mediante histogramas y curva de una distribución normal con la misma media y desviación estándar que muestran los datos. Se realizaron pruebas de hipótesis para analizar la normalidad de los datos. Se consideró como hipótesis nula que los datos sí proceden de una distribución normal y como hipótesis alternativa que no lo hacen. El p-value de estas pruebas indica la probabilidad de obtener una distribución como la observada si los datos proceden realmente de una población con una distribución normal. Se utilizaron los test de Kolmogorov-Smirnov y la modificación de Lillefors. También se aplicó el test de normalidad de Jarque-Bera que cuantifica que tanto se desvían los coeficientes de asimetría y curtosis de los esperados en una distribución normal.

Se elaboró un primer análisis de correlación y de regresión para la masa ósea total, considerando como covariables el resto de las variables óseas, además de la edad, el peso y la talla. Se analizaron los índices T proporcionados por la dexta para determinar el que se utilizaría para clasificar a los individuos según su estado óseo. Se determinó el porcentaje de individuos hombres y mujeres que presentan alguna de las condiciones óseas definidas: normal, osteopenia y osteoporosis.

Se realizó un análisis de regresión y un análisis predictivo de la regresión para la evaluación del modelo, considerando la masa ósea total como la variable de respuesta y las variables grasa y magra como covariantes. No se consideraron índices, porcentajes o proporciones para evitar colinealidad en las variables independientes. El conjunto de datos del modelo predictivo se dividió en entrenamiento (80 %) y prueba al (20 %).

Con el fin de simplificar los modelos se aplicó una reducción de variables por el método *Stepwise Selection*. En modelos con un gran número de variables explicativas puede ocurrir

que dichas variables sean redundantes o que estas variables estén correlacionadas entre sí. Introducir variables correlacionadas en un modelo lo vuelve inestable.

Con el modelo reducido se obtuvieron los coeficientes del mismo y se probó el modelo con los datos de prueba, obteniéndose gráficos de datos observados y predichos. Se realizó un análisis de residuales para evaluar los supuestos del modelo.

Estos pasos se hicieron por separado para hombres y mujeres. Los mismos pasos se siguieron para el modelo de masa grasa, considerando como covariantes las variables ósea y magra y para el modelo de masa magra, que en este caso considera las variables ósea y grasa como covariantes.

Para el análisis de clasificación se utilizó el método de kNN vecinos más próximos (k-nearest neighbors) como método de clasificación supervisada. Este método estima el valor de la función de densidad de probabilidad de que un elemento x pertenezca a la clase C a partir de la información proporcionada por el conjunto de datos de entrenamiento. Se dividió la base de datos en un conjunto de entrenamiento (80 %) y uno de validación (20 %). Se tomaron las variables magras y grasas para el análisis óseo, las variables magras y óseas para el análisis graso y las variables óseas y grasas para el análisis magro. Para determinar el valor k se tomaron los valores de k (1, 5, 10, 21) y se reportó el que tuvo un mayor valor de exactitud (accuracy).

En cuanto al modelo longitudinal se reprocesó la base de datos para filtrar ahora solo los individuos con más de una observación. Se eliminaron aquellas observaciones con datos faltantes. En la base de datos se tiene la fecha de la toma de dexa para cada observación y el ID como identificador del individuo. Como la fecha de toma de muestra no es continua ni el número de tomas constante, se determinó la primera fecha de toma de datos y la última, asignándole a cada observación un cuatrimestre de toma (por ejemplo para la toma de 18/03/2015 se asignó 2015.1, para 2/09/2013 se asignó 2013.3) A cada cuatrimestre se le asignó un número consecutivo empezando en 1999.4 como 1 y terminando en 2016.4 teniendo 69 unidades de tiempo. A cada una de las observaciones se le asignó el número consecutivo de toma de dexa quedando la tabla de datos con una columna extra denominada tiempo y que se relaciona con la fecha de toma de la dexa.

Cuadro 3.3: Datos para estudio longitudinal

ID	tiempo	DateAnalysis
1417	50	2012-03-28
1417	56	2013-07-13
1417	61	2014-10-27
1417	68	2016-09-14
2814	13	2002-10-14
2814	37	2008-12-10
2814	43	2010-06-10
2814	46	2011-02-22
2814	50	2012-02-27
5974	54	2013-02-04
5974	59	2014-06-19

Se dividieron los datos para hombres y mujeres. Se tomó el 10 % de los datos de hombres y 20 % de los de mujeres al azar para realizar los perfiles de los datos en el tiempo.

Para el modelo de regresión se consideró que la densidad de masa ósea total es la variable de respuesta y las variables grasa y magra las covariantes. Se consideró el ajuste por medidas repetidas no balanceadas (no todos los individuos tienen el mismo número de dexas) y se añadió el tiempo como una variable más.

En este modelo el tiempo es un factor aleatorio y las covariantes son efectos fijos (no cambian en función del tiempo).

El modelo utilizado es el siguiente:

```
TotalBMD ~ tiempo + Age + Height + Weight + RAFg + RALg +  
            AFg + LALg + RLFG + RLLg + LLFG + LLLg + TFG + TLg,  
random = ~ tiempo | ID
```

Se dividió la base en entrenamiento (80 %) y test (20 %) para probar el modelo.

Se obtuvo el modelo de regresión y la matriz de correlaciones del modelo. Se realizó el gráfico del modelo ajustado.

Capítulo 4

Resultados

Análisis preliminar

Se depuró la base de datos eliminando las variables que no se consideraron importantes ya sea por estar duplicadas (como el sexo) o por no ser útil (como la edad, ya que también se cuenta con la fecha de nacimiento). De las 80 variables del conjunto original quedan 76. El archivo consta de 4,319 observaciones, que corresponde a sujetos de estudio. Hay varios sujetos que tienen más de un dexta. Para la primera parte del trabajo, que es un estudio transversal, se consideró eliminar todos los estudios previos de los sujetos con más de una dexta, dejando solamente la última lectura. Esto reduce el número de observaciones a 1,495. Se detectaron 20 registros duplicados (individuos con dos registros en la misma fecha de estudio) que se eliminaron.

Se agregó una columna factor con el sexo, quedando la base de datos para el análisis transversal en 1,475 observaciones y 77 variables.

Se observan registros sin datos en alguna o algunas variables. Existen diferentes métodos de imputación para rellenar estos huecos en la base de datos, pero como no se conocía la naturaleza de los datos ni la importancia de las variables se optó por eliminar aquellos registros con datos faltantes. Después de eliminar las observaciones con algún dato faltante, la base de datos quedó con 1,426 observaciones (19 observaciones tenían algún dato faltante, apenas el 3 % de los datos).

La base de datos se dividió en hombres y mujeres, por ser este uno de los objetivos, quedando la base de datos de hombres en 1,082 observaciones y la de mujeres en 344 observaciones, ambas con 77 variables (cuadro 4.1).

Cuadro 4.1: Resumen estadístico general de los datos del estudio por sexo

Sexo	n	TotalBMD	BMI	FMI	FFMI
Masculino	1,082	1.175	24.17	5.263	17.88
Femenino	344	1.098	23.05	7.56	14.52

La base de datos cuenta con 1,426 individuos, de los cuales 1,082 son hombres y 344 son mujeres. Aunque se sabe que son individuos infectados por el VIH no se tiene ninguna medida de su grado de infección ni si están bajo algún tratamiento o terapia farmacológica. Esto es un problema porque los resultados obtenidos no reflejarán ningún resultado que caracterice a esta población.

La edad promedio de los hombres es de 46.17 años con un mínimo de 17.39 años y un máximo de 81.95. Las mujeres presentan una edad promedio es de 47.41 con un mínimo de 22.15 y un máximo de 80.73. Se puede observar que el rango de edades es amplio y no hay manera de determinar si la edad alta de hombres y mujeres se deba a que están recién infectados o que algún tratamiento haya prolongado su vida.

Para el análisis preliminar de los datos se consideraron las variables densidad masa ósea total (TotalBMD), el índice de masa corporal (BMI), índice de masa grasa (FMI) y el índice de masa magra (FFMI), así como la edad, la talla y el peso (cuadros 4.2 y 4.3).

Se puede observar que el promedio de masa ósea total es semejante hombres (1.175 ± 0.1) y mujeres (1.098 ± 0.11), lo que sugiere que no hay diferencias en este nivel. Pruebas de t muestran que no hay diferencias significativas para ninguno de estos indicadores.

Prueba t-student para masa ósea por sexo.

```
Null Hypothesis:          difference in means = 0

Alternative Hypothesis:   True difference in means is not equal to 0

Estimated Parameter(s):  mean of x = 1.175062
                        mean of y = 1.097776

Test Statistic:          t = 11.55734

P-value:                  1.007941e-27
```

Las gráficas de estos datos para hombres y mujeres muestran una tendencia a la normalidad, así como una correlación alta entre el índice de masa grasa y el índice de masa corporal (figuras 4.2 y 4.3).

Cuadro 4.2: Resumen estadístico de los datos del estudio
(Masculino)

	Media	D. E.	Minimo	Maximo
n	1,082			
Edad	46.17	10.83	17.39	81.95
Height (m)	1.734	0.07	1.47	1.93
Weightg (kg)	72.72	11.01	41.5	120.5
TotalBMD	1.175	0.1	0.87	1.9
BMI	24.17	3.22	14.19	37.96
FMI	5.263	2.49	0.06	15.28
FFMI	17.88	1.88	12.71	27.85

Cuadro 4.3: Resumen estadístico de los datos del estudio
(Femenino)

	Media	D. E.	Minimo	Maximo
n	344			
Edad	47.41	9.56	22.15	80.73
Height (m)	1.605	0.07	1.4	1.82
Weightg (kg)	59.31	11.63	34.6	99.4
TotalBMD	1.098	0.11	0.72	1.39
BMI	23.05	4.45	14.04	40.81
FMI	7.56	3.59	0.78	19.71
FFMI	14.52	1.67	10.71	21.87

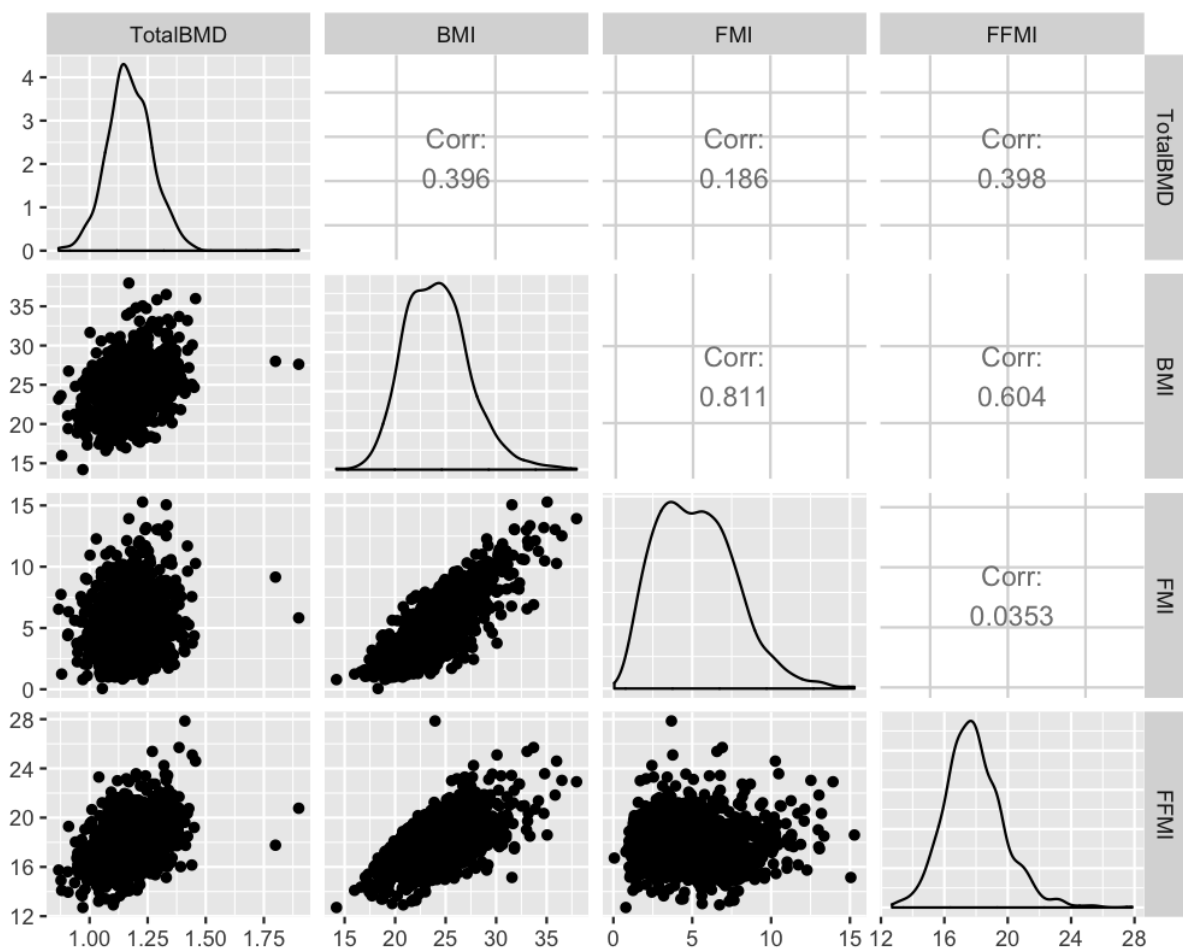


Figura 4.1: Análisis descriptivo de las variables del estudio Masculino

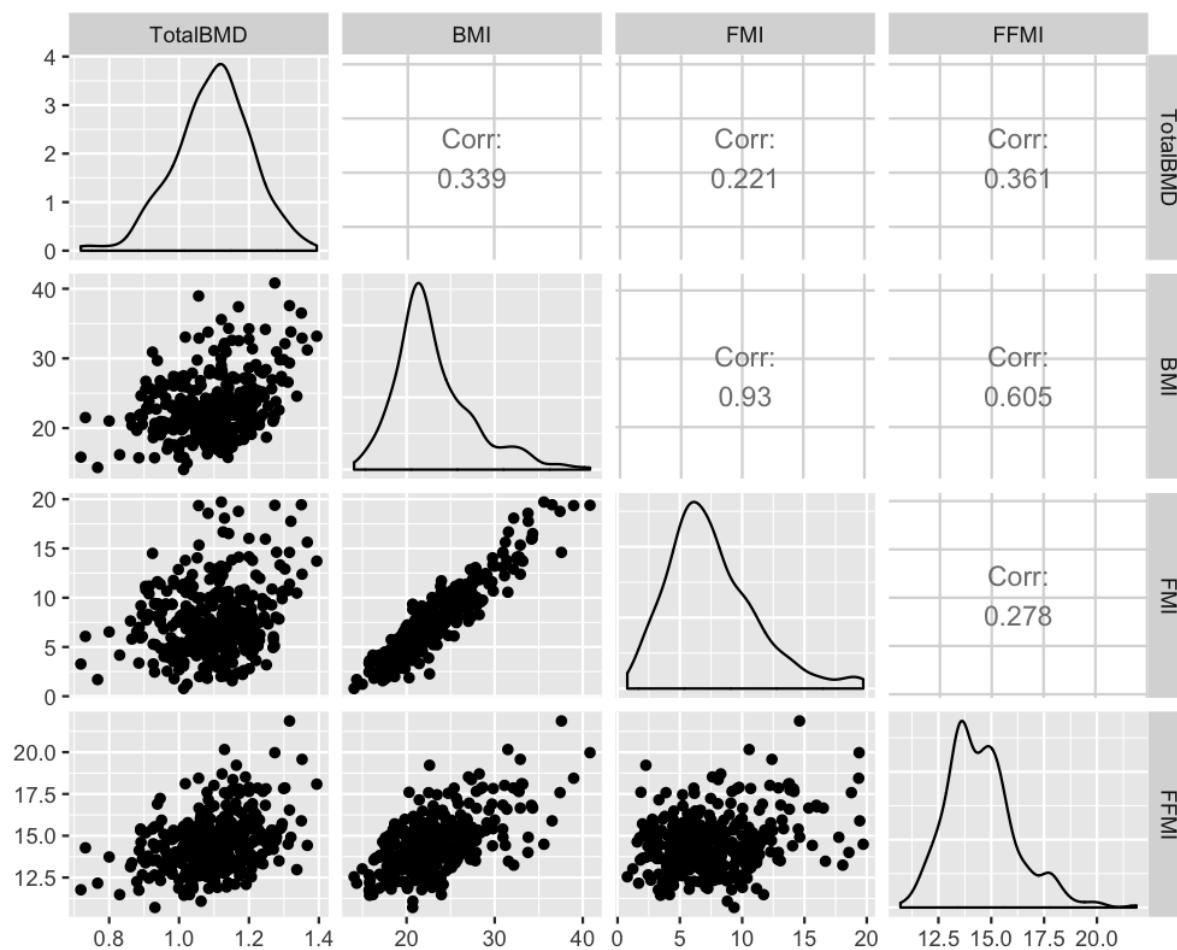


Figura 4.2: Análisis descriptivo de las variables del estudio Femenino

Cuadro 4.4: Pruebas de normalidad hombres

	TotalBMD	BMI	TotalFg	TotalLg
Kolmogorov-Smirnov	0.1907	0.1818	0.057	0.0442
Lillefors	0.0077	0.0067	0	0.0001
Jarque-Bera	0	0	0	0

Cuadro 4.5: Pruebas de normalidad mujeres

	TotalBMD	BMI	TotalFg	TotalLg
Kolmogorov-Smirnov	0.6900	0.00007	0.0157	0.1319
Lillefors	0.2491	0	0	0.0023
Jarque-Bera	0.0414	0	0	0

Hay razones para suponer que las variables índice de masa ósea para hombres y mujeres se distribuyen normalmente. El resto, masa grasa y masa magra parecen no tener este comportamiento, con excepción de la masa magra para mujeres.

El índice de masa corporal es el parámetro de la Organización Mundial de la Salud para el diagnóstico de sobrepeso y obesidad. Se considera que está asociado a la adiposidad total (índice de masa grasa) y tiene una fuerte correlación epidemiológica con la morbilidad asociada a la obesidad (Alvero-Cruz *et al.*, 2010).

La masa corporal total (BMI) se compone de la masa grasa (FMI) y la masa libre de grasa (FFMI) ya que todos estos índices están ajustados a la talla del individuo. No estaba en los objetivos originales del trabajo, pero creo que sería interesante ahondar en ello. La correlación en los hombres del BMI con el FMI es de 0.811 y con el FFMI es de 0.604. En el caso de las mujeres la correlación entre BMI y FMI es de 0.93 lo que puede sugerir que el índice de masa corporal está más influenciado por la grasa en las mujeres que en los hombres. En el caso de la masa magra no hay diferencia.

Análisis transversal

Masa ósea

Para el análisis de masa ósea se consideraron las variables relacionadas (cuadro 4.6). Se hizo un análisis descriptivo de las variables y se obtuvieron las correlaciones de las mismas.

Cuadro 4.6: Variables del estudio óseo

Variable	Descripción
TotalBMD	Densidad Masa ósea Total
L1BMD	Lumbar 1 Densidad masa ósea
L2BMD	Lumbar 2 Densidad masa ósea

Variable	Descripcion
L3BMD	Lumbar 3 Densidad masa ósea
L4BMD	Lumbar 4 Densidad masa ósea
L1L4BMD	Lumbar 1-Lumbar 4 Densidad masa ósea
L2L4BMD	Lumbar 2-Lumbar 4 Densidad masa ósea
NeckFBMD	Cuello Femoral Densidad masa ósea
WardsBMD	Triangulo de Ward DMO
TrochBMD	Trocanter DMO
TotalFBMD	Total Femur bone mineral density
BMI	Indice de Masa Corporal (kg/m^2)

Cuadro 4.7: Matriz de correlaciones para variables óseas (masculino)

	TotalBMD	L1BMD	L2BMD	L3BMD	L4BMD	L1L4BMD	L2L4BMD
TotalBMD	1	0.671	0.658	0.635	0.641	0.695	0.681
L1BMD	0.671	1	0.856	0.815	0.753	0.894	0.849
L2BMD	0.658	0.856	1	0.872	0.802	0.925	0.922
L3BMD	0.635	0.815	0.872	1	0.869	0.945	0.956
L4BMD	0.641	0.753	0.802	0.869	1	0.927	0.95
L1L4BMD	0.695	0.894	0.925	0.945	0.927	1	0.987
L2L4BMD	0.681	0.849	0.922	0.956	0.95	0.987	1
NeckFBMD	0.694	0.537	0.535	0.536	0.535	0.578	0.572
WardsBMD	0.64	0.508	0.509	0.515	0.497	0.543	0.541
TrochBMD	0.724	0.53	0.528	0.526	0.537	0.57	0.563
TotalFBMD	0.751	0.569	0.557	0.552	0.563	0.603	0.593
BMI	0.396	0.207	0.154	0.108	0.162	0.167	0.145

	NeckFBMD	WardsBMD	TrochBMD	TotalFBMD	BMI
TotalBMD	0.694	0.64	0.724	0.751	0.396
L1BMD	0.537	0.508	0.53	0.569	0.207
L2BMD	0.535	0.509	0.528	0.557	0.154
L3BMD	0.536	0.515	0.526	0.552	0.108
L4BMD	0.535	0.497	0.537	0.563	0.162
L1L4BMD	0.578	0.543	0.57	0.603	0.167
L2L4BMD	0.572	0.541	0.563	0.593	0.145
NeckFBMD	1	0.913	0.791	0.867	0.221
WardsBMD	0.913	1	0.736	0.825	0.104
TrochBMD	0.791	0.736	1	0.901	0.317
TotalFBMD	0.867	0.825	0.901	1	0.302
BMI	0.221	0.104	0.317	0.302	1

Cuadro 4.9: Matriz de correlaciones para variables óseas (femenino)

	TotalBMD	L1BMD	L2BMD	L3BMD	L4BMD	L1L4BMD	L2L4BMD
TotalBMD	1	0.751	0.743	0.765	0.722	0.784	0.776
L1BMD	0.751	1	0.906	0.883	0.799	0.933	0.896
L2BMD	0.743	0.906	1	0.925	0.838	0.958	0.951
L3BMD	0.765	0.883	0.925	1	0.895	0.975	0.978
L4BMD	0.722	0.799	0.838	0.895	1	0.934	0.95
L1L4BMD	0.784	0.933	0.958	0.975	0.934	1	0.995
L2L4BMD	0.776	0.896	0.951	0.978	0.95	0.995	1
NeckFBMD	0.763	0.626	0.614	0.659	0.61	0.662	0.657
WardsBMD	0.762	0.64	0.641	0.68	0.635	0.684	0.681
TrochBMD	0.736	0.607	0.566	0.602	0.578	0.621	0.61
TotalFBMD	0.796	0.659	0.632	0.669	0.63	0.682	0.673
BMI	0.339	0.2	0.097	0.1	0.087	0.124	0.102

	NeckFBMD	WardsBMD	TrochBMD	TotalFBMD	BMI
TotalBMD	0.763	0.762	0.736	0.796	0.339
L1BMD	0.626	0.64	0.607	0.659	0.2
L2BMD	0.614	0.641	0.566	0.632	0.097
L3BMD	0.659	0.68	0.602	0.669	0.1
L4BMD	0.61	0.635	0.578	0.63	0.087
L1L4BMD	0.662	0.684	0.621	0.682	0.124
L2L4BMD	0.657	0.681	0.61	0.673	0.102
NeckFBMD	1	0.933	0.834	0.92	0.262
WardsBMD	0.933	1	0.817	0.896	0.145
TrochBMD	0.834	0.817	1	0.946	0.372
TotalFBMD	0.92	0.896	0.946	1	0.361
BMI	0.262	0.145	0.372	0.361	1

Para el análisis de regresión en hombres, se consideró la densidad de masa ósea total como variable dependiente o de respuesta. El modelo con todas las variables explica el 70 % de la variabilidad. Para simplificar el modelo se utilizó el método stepwise. Así se eliminaron 7 variables. Se realizó un análisis de residuos del modelo reducido encontrándose que los residuales se distribuyen normalmente. También se observaron valores extremos.

En el análisis de regresión para mujeres se procedió de la misma forma, encontrando que el modelo con todas las variables explica el 76 % de la variabilidad. El método stepwise eliminó 8 variables. El análisis de residuales es mejor en este caso. No hay valores extremos.

En ambos casos se llevó a cabo un análisis predictivo de la regresión para la evaluación del modelo. El conjunto de datos se dividió en entrenamiento (80 %) y prueba al (20 %). La R^2 de la regresión es del 75 % en hombres y 76 % en mujeres.

Los modelos obtenidos son:

Hombres:

$$\text{TotalBMD} = 0.204 + 0.132*\text{Height} + 0.106*\text{L1BMD} + 0.08*\text{L2BMD} + 0.064*\text{L4BMD} + 0.059*\text{WardsBMD} + 0.117*\text{TrochBMD} + 0.179*\text{TotalFBMD} + 0.006*\text{BMI}$$

Coefficientes	Valor	Desviación
(Intercept)	0.204	0.0448
Height	0.132	0.0249
L1BMD	0.106	0.0219
L2BMD	0.08	0.0225
L4BMD	0.064	0.0178
WardsBMD	0.059	0.0187
TrochBMD	0.117	0.0312
TotalFBMD	0.179	0.0341
BMI	0.006	0.0006

Mujeres:

$$\text{TotalBMD} = 0.265 + 0.113*\text{Height} - 0.0005*\text{Age} + 0.072*\text{L1BMD} + 0.218*\text{L2L4BMD} + 0.117*\text{WardsBMD} + 0.186*\text{TotalFBMD} + 0.005*\text{BMI}$$

Coefficientes	Valor	Desviación
(Intercept)	0.265	0.080
Height	0.113	0.046
Age	-0.0005	0.0003
L1BMD	0.072	0.042
L2L4BMD	0.218	0.041
WardsBMD	0.117	0.046
TotalFBMD	0.186	0.054
BMI	0.005	0.0008

Aunque estos modelos no son lo que se busca, dado que el total de masa ósea está en función de sus componentes, sirve para determinar, de forma exploratoria, que en hombres el peso puede ser un factor determinante en el cambio de masa ósea. En los hombres la densidad mineral ósea total disminuye en -0.16 con respecto al peso manteniendo los otros factores constantes. En el caso de las mujeres la relación es inversa. La densidad mineral ósea aumenta en 0.11 con respecto al peso. Pero además en las mujeres la densidad mineral ósea disminuye en -0.0005 al cumplir un año más, manteniendo los otros factores constantes.

La literatura indica que la variable T es el mejor criterio para definir osteope-

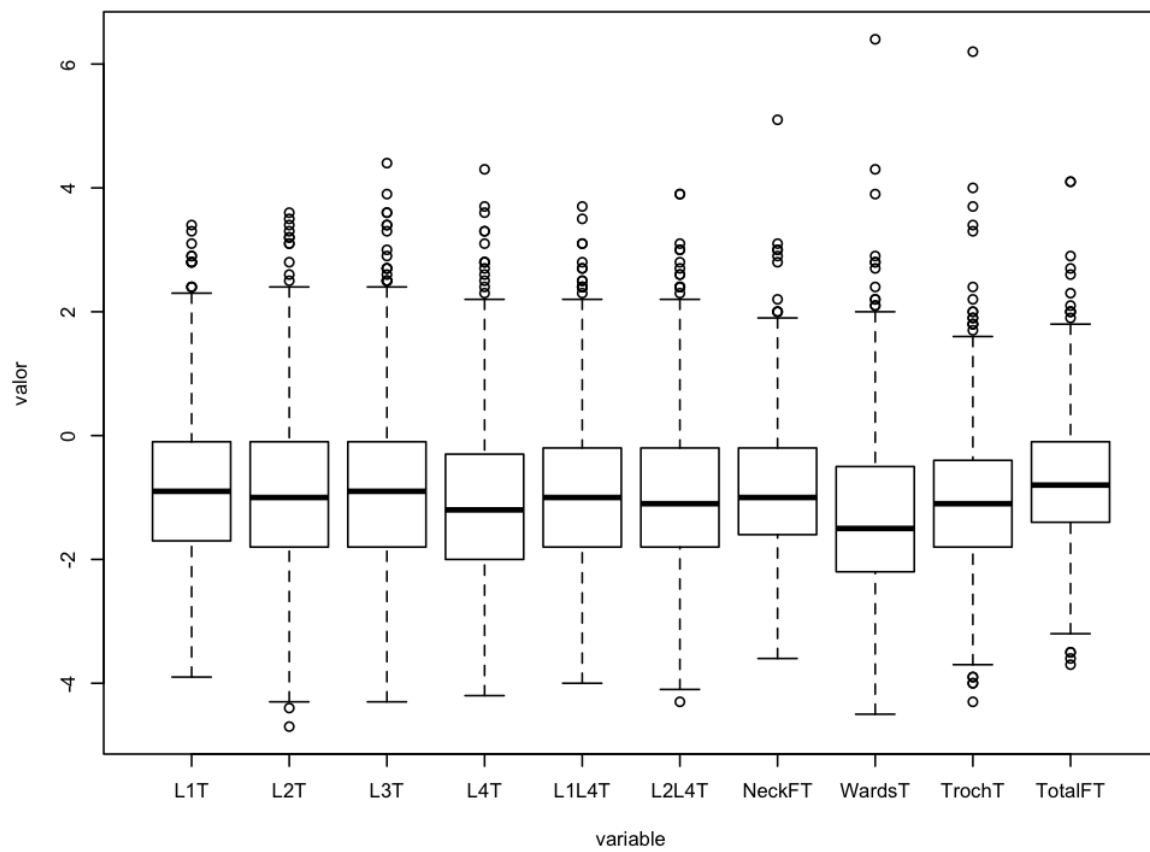


Figura 4.3: Valores promedio variables oseas T masculino

nia/osteoporosis o condición normal. Esta variable T se basa en el triángulo de Ward (zona situada entre los tres haces trabeculares de la cabeza del fémur) e indica que un valor menor a -2.5 es diagnóstico de osteoporosis, entre -1 y -2.5 es osteopenia y valores mayores a -1 es normal. La dexta tiene otros valores T calculados a partir de otras estructuras. No se encontraron diferencias significativas entre estas medidas (figura 4.3 para hombres y 4.4 para mujeres), por lo que se podría utilizar cualquiera de ellas, tanto para hombres como para mujeres.

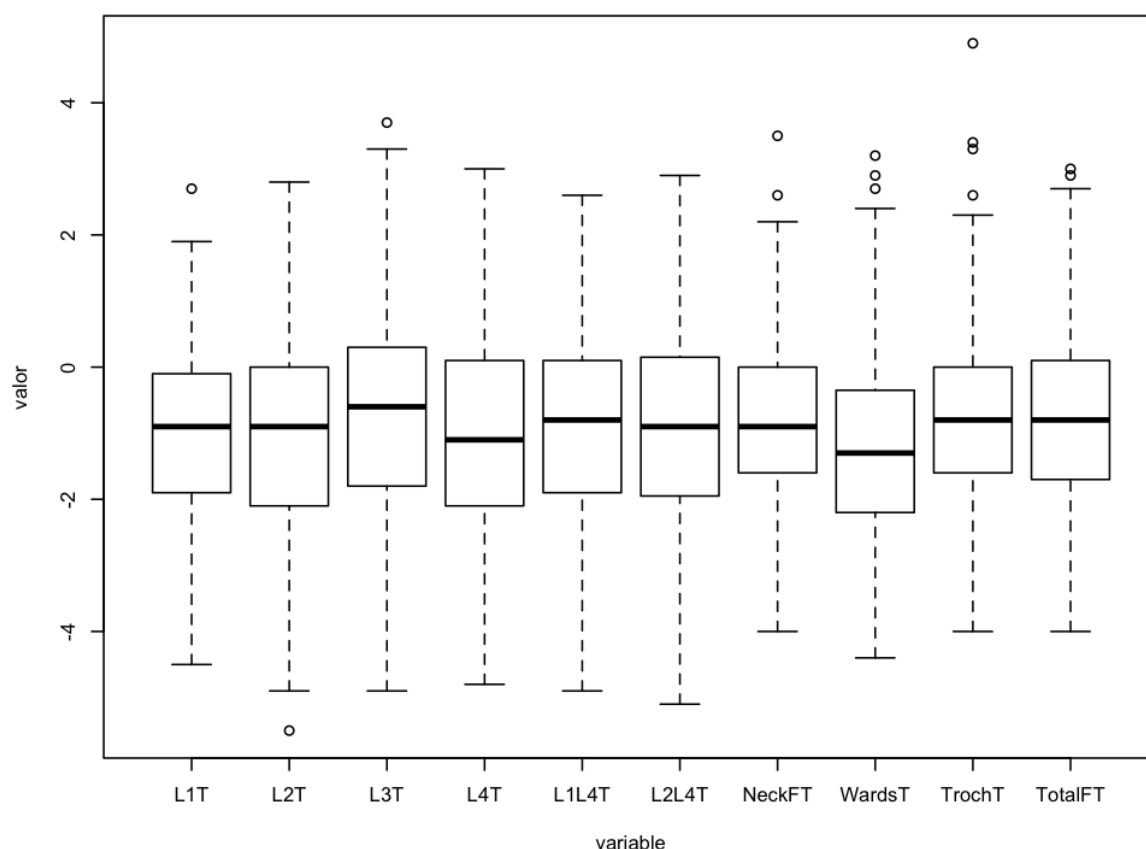


Figura 4.4: Valores promedio variables oseas T femenino

Se agregó el estado óseo según los criterios de la OMS y en un análisis de densidad de datos se observó que el triángulo de ward es el índice con menos sobreposición de datos, por lo que esta medida es la más eficiente (figura 4.5 para hombres y 4.6 para mujeres).

En la población se encontró que el 35 % de los individuos presentan la condición normal, 46 % presentan osteopenia y 19 % presentan osteoporosis en los hombres. En las mujeres los porcentajes son: 40 % normal, 41 % osteopenia y 19 % osteoporosis (figura 4.7).

Aparentemente no hay grandes diferencias en los porcentajes de osteopenia y osteoporosis entre hombres y mujeres.

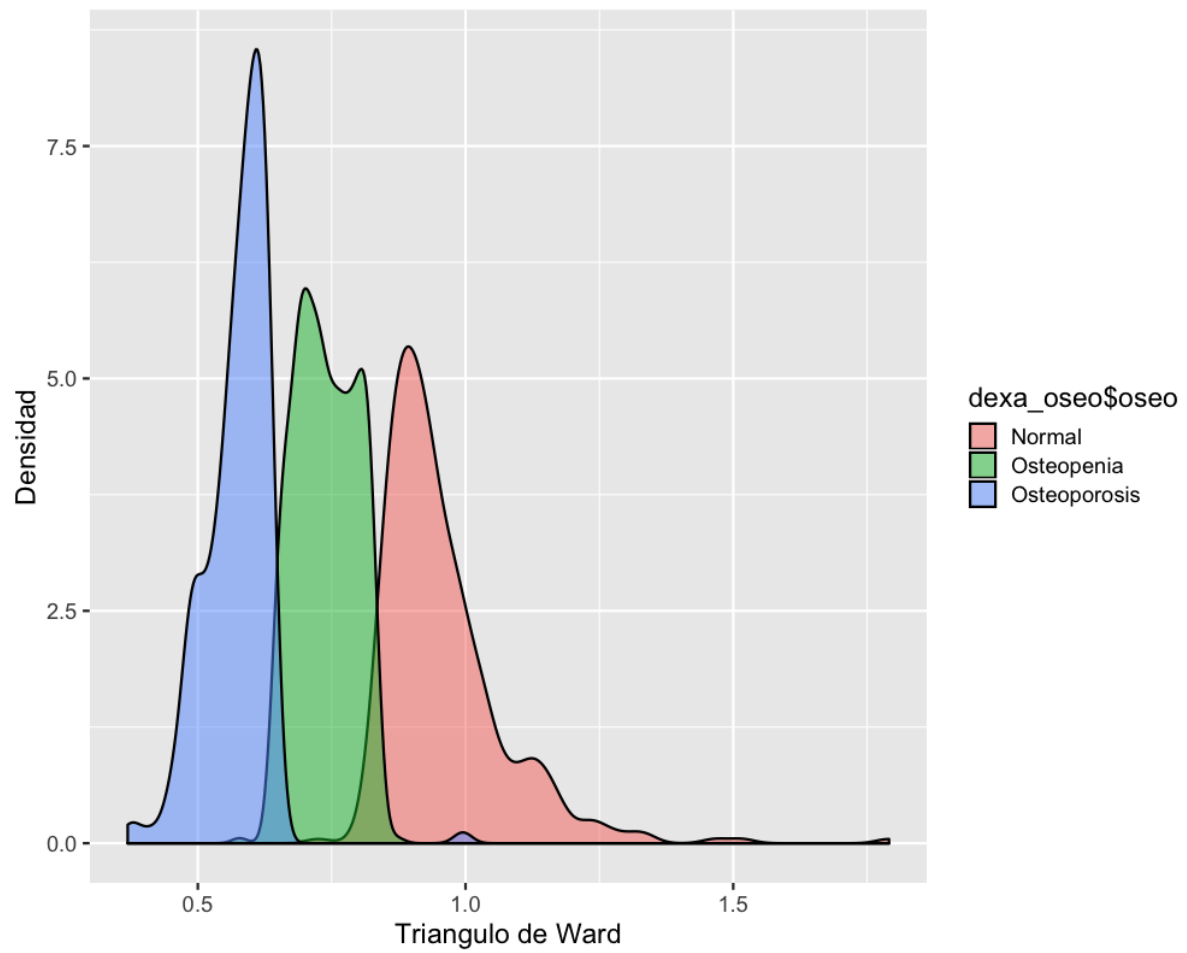


Figura 4.5: Masa ósea triangulo de Ward masculino

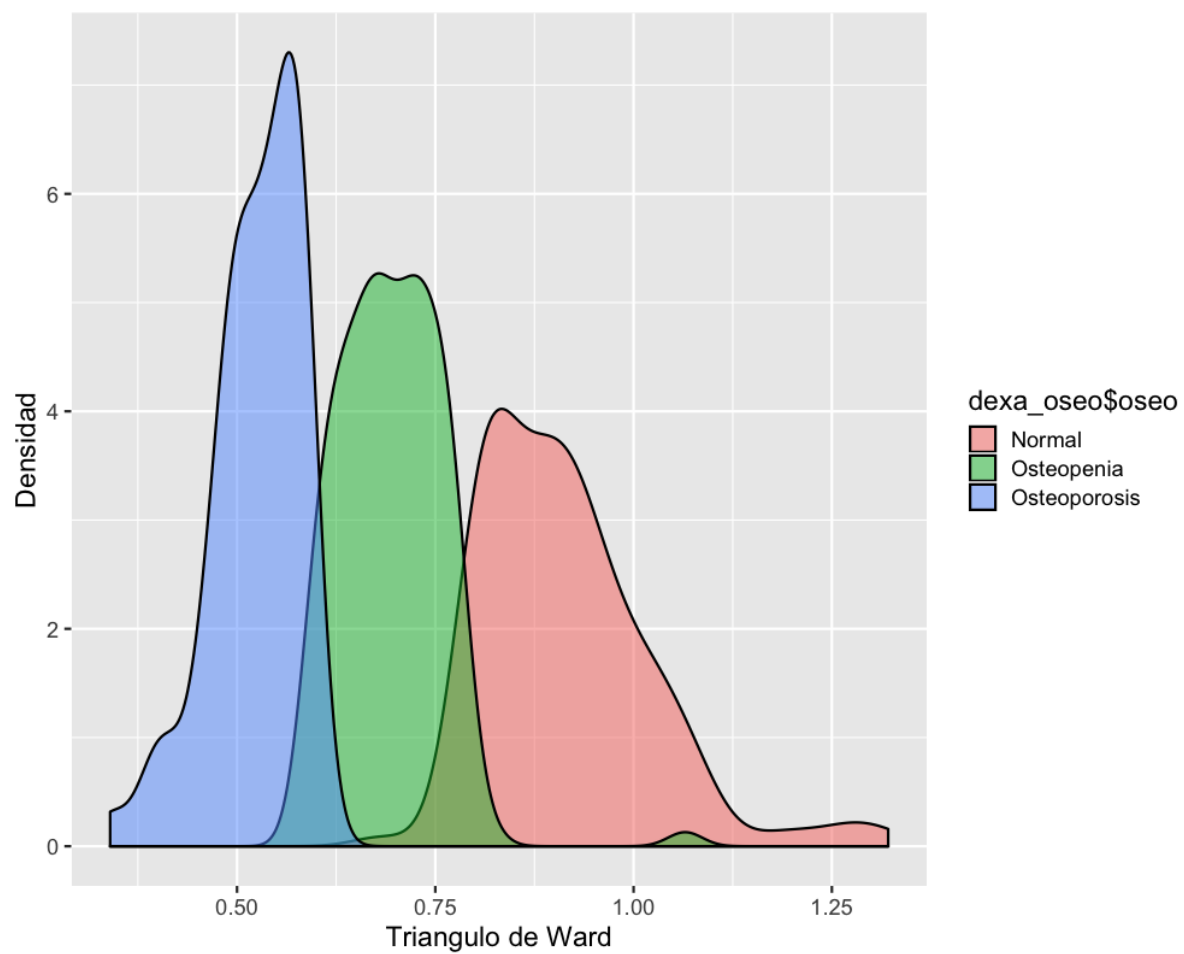


Figura 4.6: Masa ósea triangulo de Ward femenino

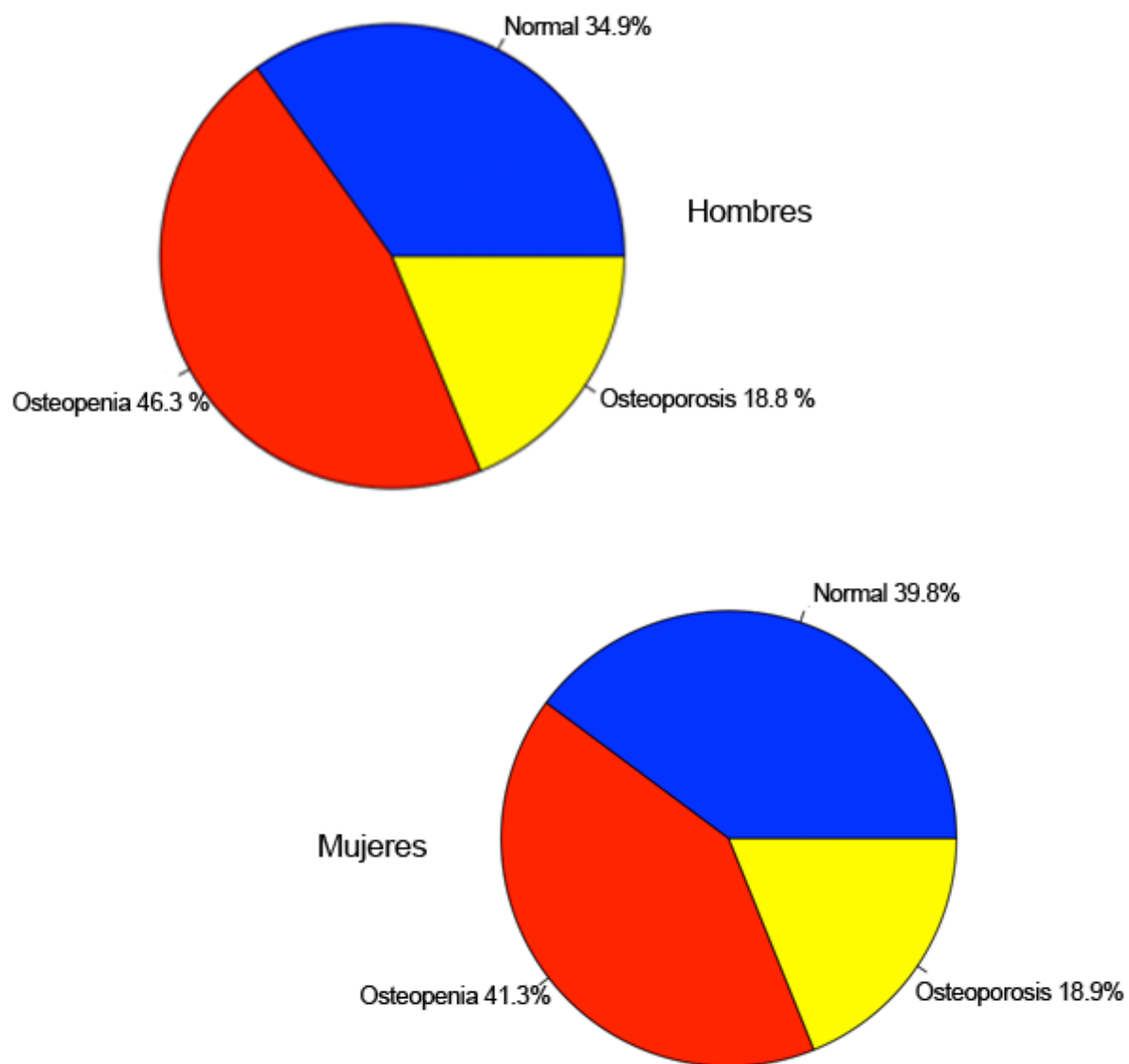


Figura 4.7: Distribucion de masa osea

Modelo de masa ósea general

Hombres

Este modelo considera las variables de masa grasa y masa magra para determinar si existe alguna relación con la masa ósea. Al igual que en el apartado anterior se considera la densidad de masa ósea total como la variable de respuesta y en este caso las variables grasa y magra como covariantes sin considerar los índices, marcas T y marcas Z.

Cuadro 4.13: Correlaciones para la masa ósea total en hombres

Weight+TotalBMD: 0.47	RALg+TotalBMD: 0.46
LALg+TotalBMD: 0.45	RLLg+TotalBMD: 0.46
LLLg+TotalBMD: 0.47	TLg+TotalBMD: 0.37
TotalBMD+Weight: 0.47	TotalBMD+RALg: 0.46
TotalBMD+LALg: 0.45	TotalBMD+RLLg: 0.46
TotalBMD+LLLg: 0.47	TotalBMD+TLg: 0.37

Se hace un primer ensayo con todas las variables y un análisis de correlaciones indica que la masa ósea total (TotalBMD) está correlacionada con otras variables con valores entre 0.37 y 0.47 (cuadro 4.11). El modelo lineal con todas las variables grasas y magras explica el 33 % de la variabilidad. Al hacer una reducción de variables por el método Stepwise Selection el modelo final presente la misma R^2 pero con menos variables.

El modelo reducido para explicar la masa ósea total contempla las variables grasas (grasa en el brazo derecho y en el tronco), masa magra (en el brazo izquierdo, en las piernas y en el tronco), además de la edad, el peso y la talla.

$$\text{TotalBMD} \sim \text{Height} + \text{Weight} + \text{Age} + \text{RALg} + \text{LAFg} + \text{RLFg} + \text{LLLg} + \text{TFg} + \text{TLg}$$

Los coeficientes de este modelo se muestran en el cuadro 4.14. El intercepto tiene un valor alto. Este es el valor que el modelo predice cuando todas las variables independientes valen cero. Sin embargo es cercano al valor máximo de la masa ósea total, lo cual es un indicador que el modelo no es tan bueno. La mayoría de los coeficientes presentar valores bajos lo que indica que no presentan un papel importante en la predicción del modelo, lo que no es una sorpresa, dados los valores bajos encontrados en el análisis de correlación.

La interpretación de estas covariables sería: la densidad mineral ósea disminuye en -0.002 por cada año de edad manteniendo los otros factores constantes. Aumenta en 0.007 por cada aumento de peso (un kilo por las unidades de peso) manteniendo el resto de factores constante. El más significativo es la disminución de la masa ósea en -0.24 por el aumento en talla (en metros) manteniendo los otros factores constantes, sería una relación interesante a investigar.

El signo de los coeficientes brinda información de su importancia en la masa ósea total. Las covariables con signo positivo indican que su aumento es positivo en la masa ósea

mientras que los que son negativos indican lo contrario. Hay 2 covariables positivas y 7 covariables negativas.

Cuadro 4.14: Coeficientes del modelo lineal óseo para hombres

Coeficientes	Valor	Desviación	Coeficientes	Valor	Desviación
(Intercepto)	1.2128	8.346e-02	LAFg	-6.03e-05	1.544e-05
Height	-2.45e-01	5.573e-02	RLFg	-3.96e-07	2.600e-05
Weight	7.02e-03	1.224e-03	LLFg	-8.90e-06	2.594e-05
Age	-1.60e-03	3.120e-04	TFg	-5.49e-07	1.765e-06
RALg	2.00e-05	8.045e-06	TLg	-1.71e-06	1.810e-06

El análisis de residuales muestra que tan bien se ajusta el modelo. Este gráfico (4.8) se realizó a partir de los datos de prueba con el modelo obtenido con los datos de entrenamiento.

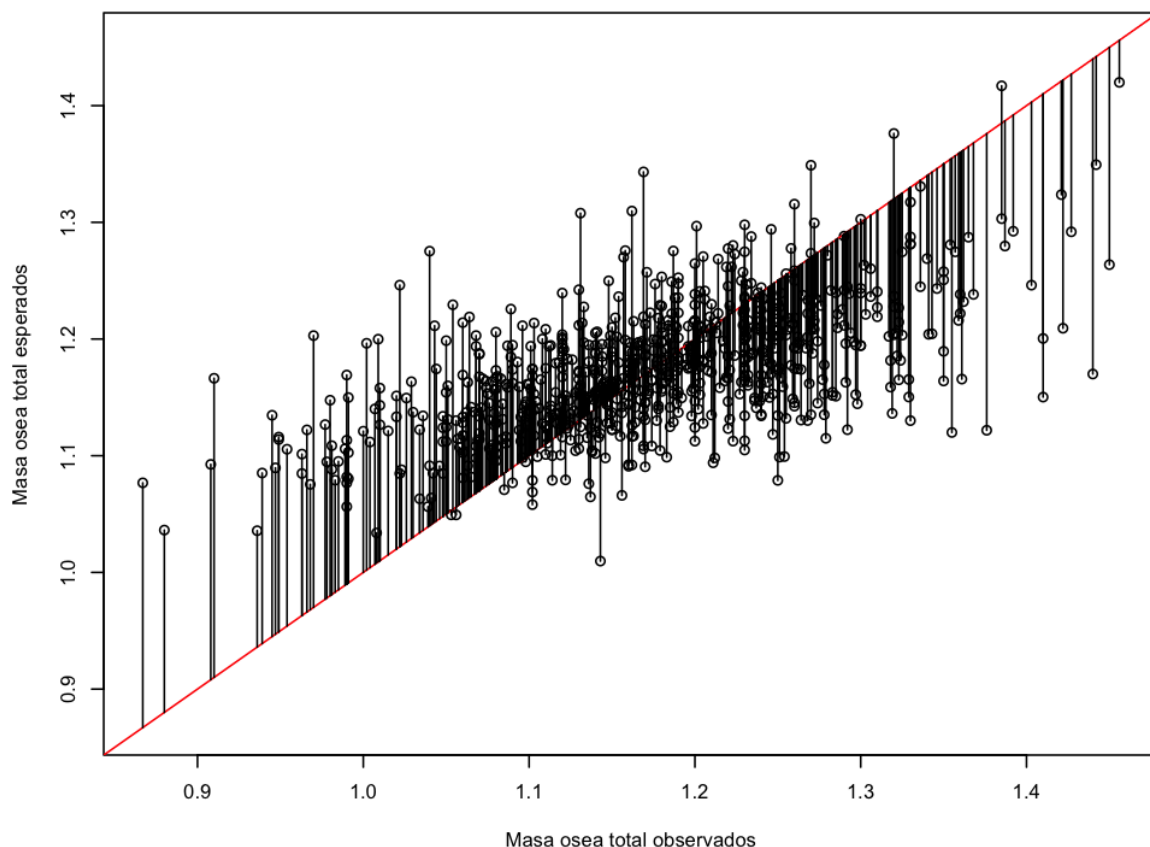


Figura 4.8: Residuales del modelo lineal óseo reducido hombres

La distribución de la variable de respuesta (TotalBMD) se verifica mediante la representación gráfica de los datos en un histograma y superponiendo la curva de una distribución

normal con la misma media y desviación estándar de los datos.

Gráficamente se observa que la variable se ajusta a una distriución normal, aunque la prueba de Kolmogorov-Smirnov rechaza la hipótesis nula de normalidad. La prueba ajustada de Lilliefors (que no supone los valores de media y desviación de los datos) permiten aceptar la hipótesis de normalidad.

En el análisis de residuos la gráfica qq-plot permite observar 3 datos outliers, que son los que pueden estar causando problemas en el ajuste. La densidad de los residuos muestran una distribución sin tendencia con distribución uniforme.

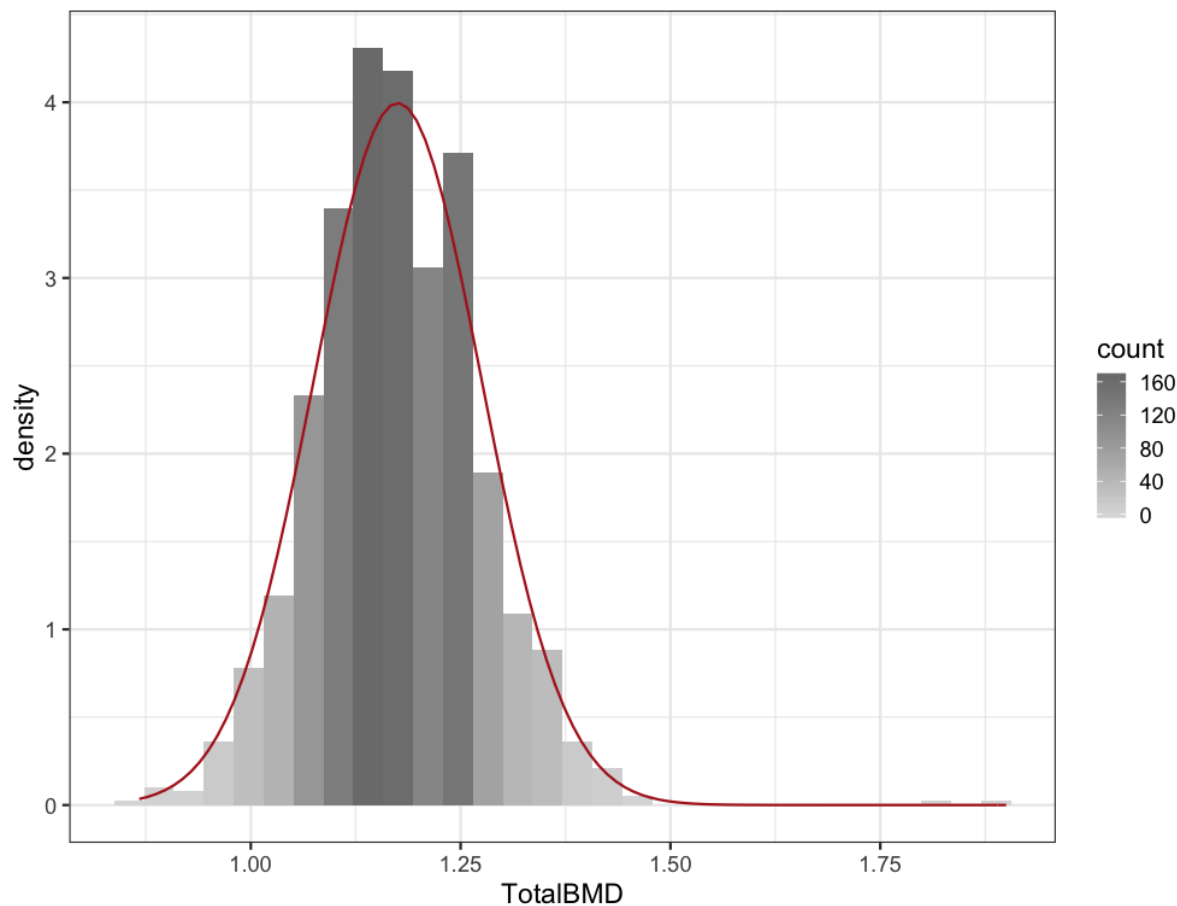


Figura 4.9: Distribución de la masa ósea total Hombres

Cuadro 4.15: Test de Kolmogorov-Smirnov densidad masa ósea hombres

Alternative Hypothesis:	two-sided
Test Name:	One-sample Kolmogorov-Smirnov test
Data:	dexa_oseoh\$TotalBMD
Test Statistic:	D = 0.03295043

P-value: 0.1906638

Cuadro 4.16: Modificación de Lillefors densidad masa ósea hombres

Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Data:	dexa_oseoh\$TotalBMD
Test Statistic:	D = 0.03295043
P-value:	0.007698875

Mujeres

Al igual que en el caso de los hombres para este análisis se considera como variable dependiente la densidad de masa ósea total (TotalBMD) y las variables de masa grasa y masa magra como independientes, sin considerar los índices, marcas T y marcas Z.

Se obtuvo la matriz de correlaciones y se destacan aquellas que involucran a la variable de respuesta.

En este caso se destaca la correlación negativa que existe entre la edad y la masa ósea total. Las correlaciones van de -0.34 a 0.45 (Cuadro 4.17).

Cuadro 4.17: Correlaciones para la masa ósea total en hombres

Weight+TotalBMD: 0.43	Age+TotalBMD: -0.34
RAIg+TotalBMD: 0.45	LALg+TotalBMD: 0.43
RLLg+TotalBMD: 0.43	LLLg+TotalBMD: 0.44
TLg+TotalBMD: 0.34	TotalBMD+Weight: 0.43
TotalBMD+Age: -0.34	TotalBMD+RAIg: 0.45
TotalBMD+LALg: 0.43	TotalBMD+RLLg: 0.43
TotalBMD+LLLg: 0.44	TotalBMD+TLg: 0.34

El modelo lineal con todas las variables grasas y magras explica el 37 % de la variabilidad. Se realizó una reducción de variables por el método Stepwise selection donde hay reducción de variables y ahora la R^2 es de 0.37

El modelo reducido contempla la talla, el peso y la edad, además de las variables de masa grasa (grasa en los brazos, en el tronco, en la pierna izquierda), de masa magra (en los brazos, en la pierna izquierda y en el tronco).

$$\text{TotalBMD} \sim \text{Height} + \text{Weight} + \text{Age} + \text{RAFg} + \text{LALg} + \text{RLFg} + \text{RLLg} + \text{Tfg} + \text{TLg}$$

En el cuadro 4.18 se muestran los coeficientes del modelo. En este caso el intercepto tiene un valor cercano al valor máximo de masa ósea. El valor de la masa ósea total va de 0.719 a 1.394 por tanto, si todas las variables independientes valieran cero el modelo predeciría un valor de masa ósea de 1.119. La talla es la única covariable que tiene alguna significancia en el modelo. La masa ósea disminuye en -0.22 por el aumento en la talla.

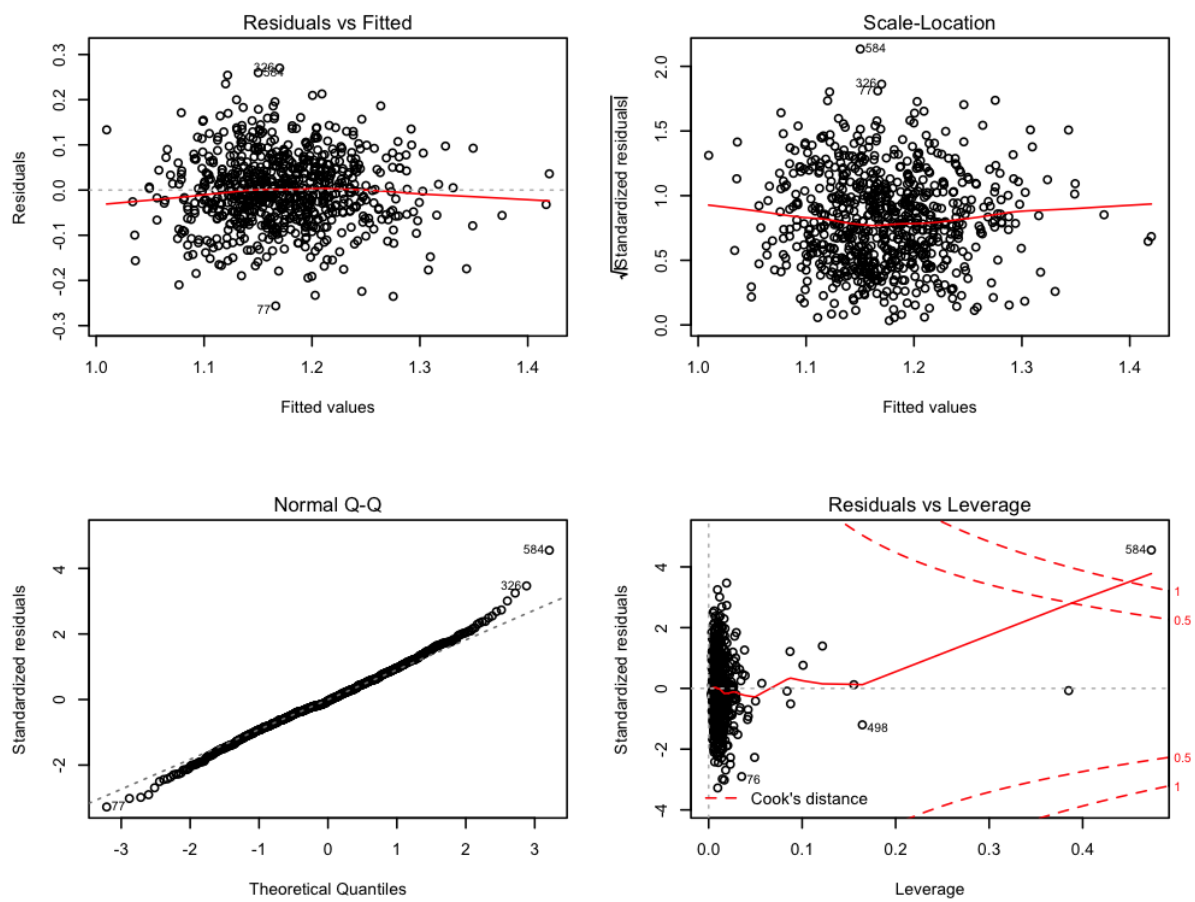


Figura 4.10: Análisis de Residuales del modelo óseo hombres

Cuadro 4.18: Coeficientes del modelo lineal óseo para mujeres

	Valor	Desviación
(Intercept)	1.119203	0.160
Height	-0.217605	0.113
Weight	0.035117	0.006
Age	-0.003770	0.0006
RAFg	-0.000038	0.00008
LAFg	-0.000052	0.00008
LALg	-0.000048	0.00006
RLFg	-0.000061	0.00003
RLLg	-0.000069	0.00003
TFg	-0.000030	0.000007
TLg	-0.000025	0.000006

En la figura 4.11 se observa la distribución de los residuales del modelo lineal. Aunque hay datos que no se ajustan los residuales tienen una distribución homogénea aunque la mayoría se alejan de la línea de referencia.

La variable de respuesta (TotalBMD) para mujeres no se distribuye de forma normal. En el gráfico 4.12 se pueden observar picos y las pruebas de hipótesis rechazan la hipótesis de normalidad.

Cuadro 4.19: Test de Kolmogorov-Smirnov densidad masa ósea mujeres

Alternative Hypothesis:	two-sided
Test Name:	One-sample Kolmogorov-Smirnov test
Data:	dexa_oseom\$TotalBMD
Test Statistic:	D = 0.03842409
P-value:	0.6900688

Cuadro 4.20: Modificación de Lilliefors densidad masa ósea mujeres

Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Data:	dexa_oseom\$TotalBMD
Test Statistic:	D = 0.03842409
P-value:	0.2491346

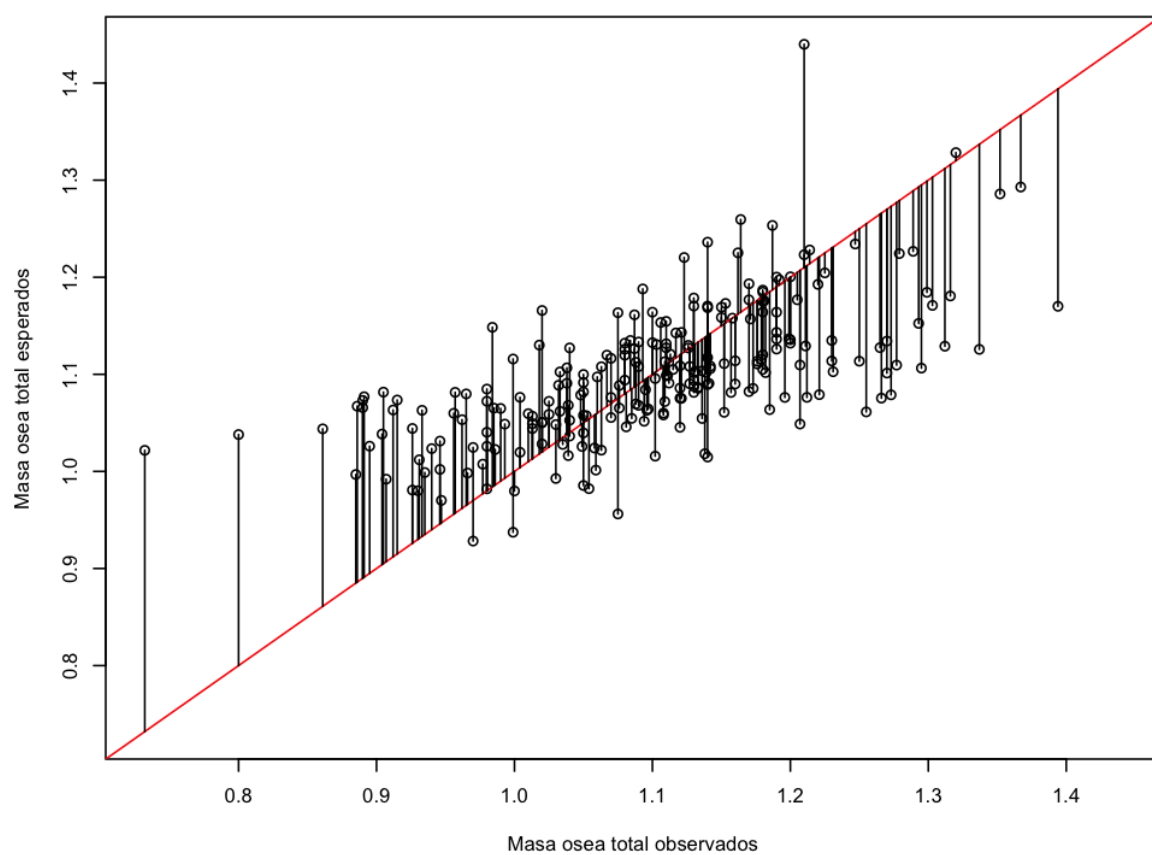


Figura 4.11: Residuales del modelo lineal óseo reducido mujeres

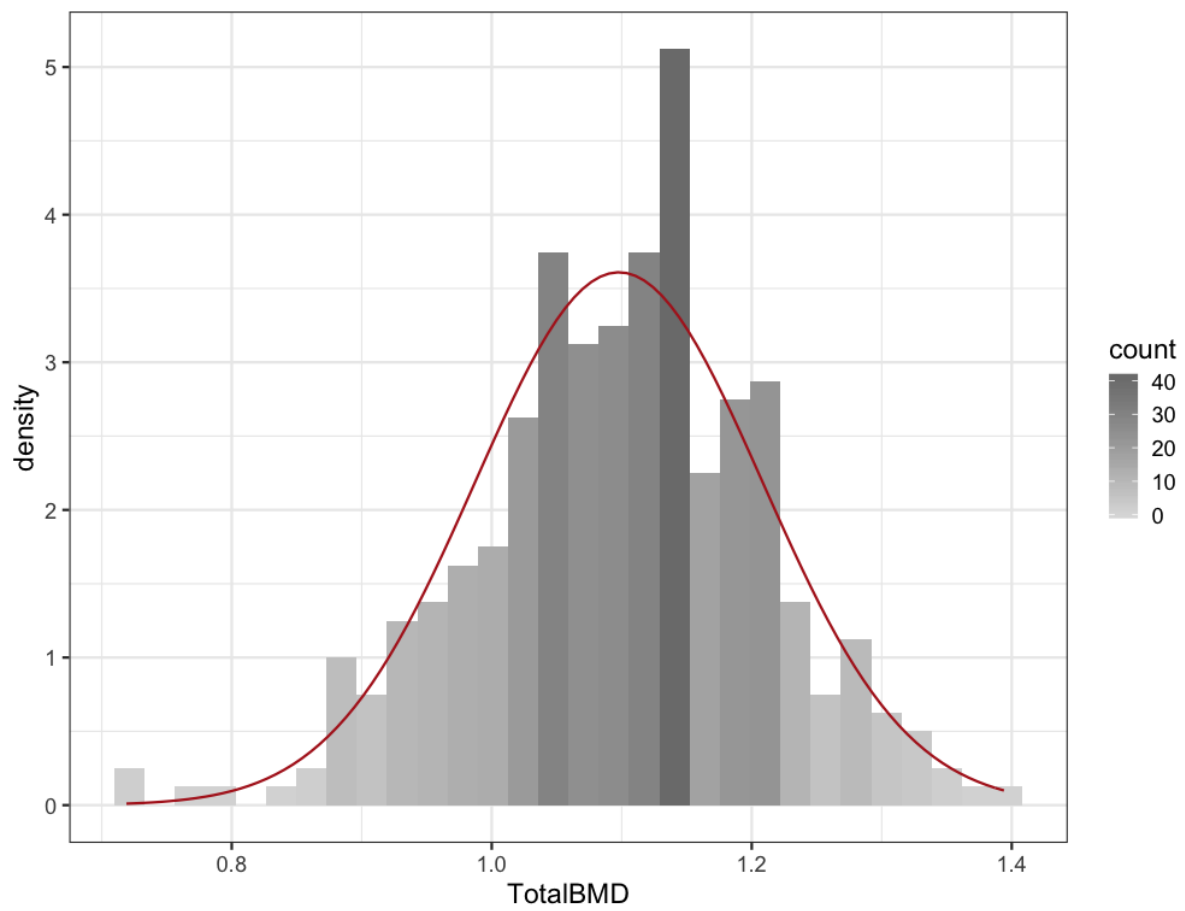


Figura 4.12: Distribución de la masa ósea total mujeres

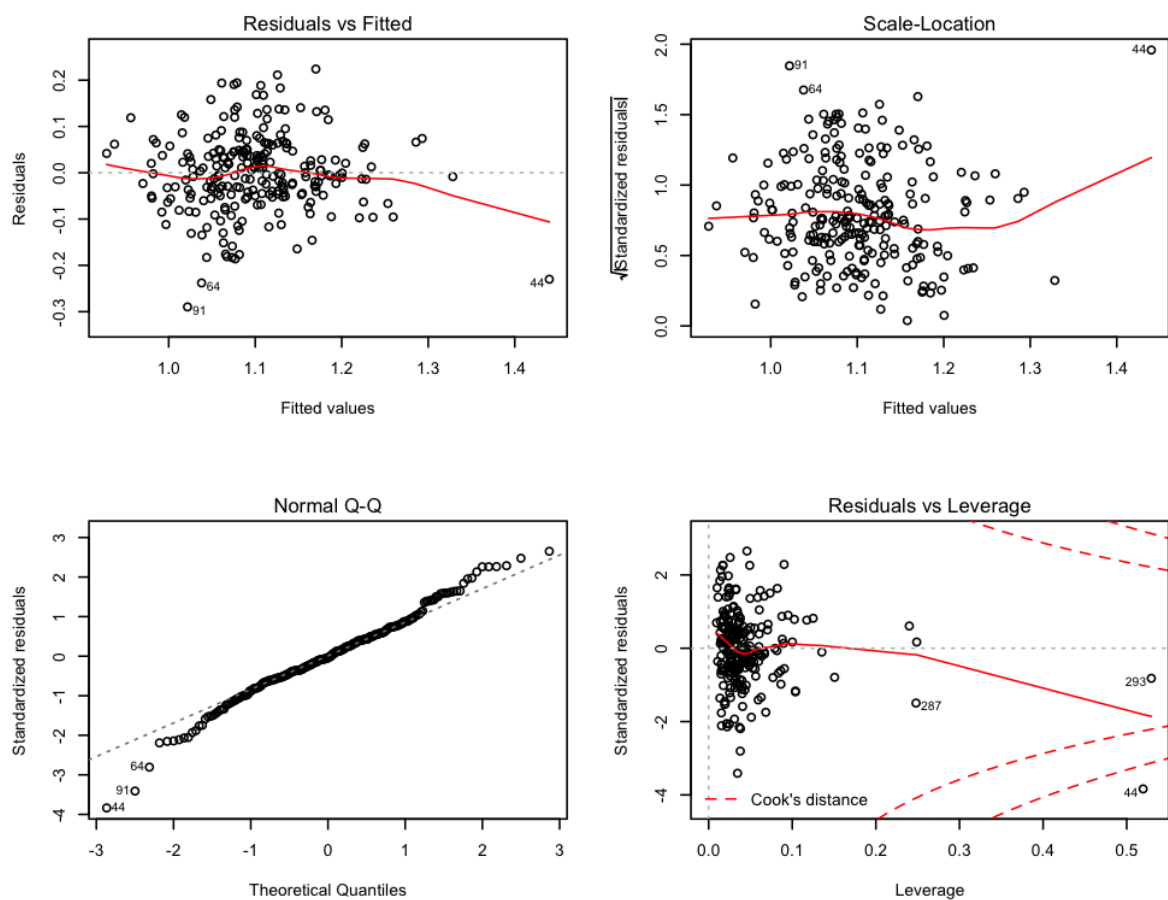


Figura 4.13: Análisis de Residuales del modelo óseo mujeres

Modelo de clasificación masa ósea general

En cuanto a los modelos de clasificación, se utilizó el método de kNN vecinos más próximos. Se obtuvieron los valores de exactitud (accuracy), obteniendo 0.60 para hombres y de 0.62 para mujeres (cuadros 4.21 y 4.22).

Cuadro 4.21: Tabla de clasificación ósea para hombres
método kNN ($k=21$, $\text{accuracy}=0.60$)

Estado óseo	Normal	Osteopenia	Osteoporosis	Total renglón
Normal	39	35	1	75
	0.520	0.467	0.013	0.346
	0.736	0.235	0.067	
	0.180	0.161	0.005	
Osteopenia	12	84	5	101
	0.119	0.832	0.050	0.465
	0.226	0.564	0.333	

Estado óseo	Normal	Osteopenia	Osteoporosis	Total renglón
	0.055	0.387	0.023	
Osteoporosis	2	30	9	41
	0.049	0.732	0.220	0.189
	0.038	0.201	0.600	
	0.009	0.138	0.041	
Total columna	53	149	15	217
	0.244	0.687	0.069	

Cuadro 4.22: Tabla de clasificación ósea para mujeres
método kNN (k=21, accuracy=0.62)

Estado óseo	Normal	Osteopenia	Osteoporosis	Total renglón
Normal	17	12	1	30
	0.567	0.400	0.033	0.435
	0.739	0.300	0.167	
	0.246	0.174	0.014	
Osteopenia	6	21	0	27
	0.222	0.778	0.000	0.391
	0.261	0.525	0.000	
	0.087	0.304	0.000	
Osteoporosis	0	7	5	12
	0.000	0.583	0.417	0.174
	0.000	0.175	0.833	
	0.000	0.101	0.072	
Total columna	23	40	6	69
	0.333	0.580	0.087	

Para hombres se obtiene una precisión de predicción de 60 % y un error de predicción de 40 % para el conjunto de validación.

Para mujeres los valores son una predicción del 62 % y un error del 38 % en el conjunto de validación.

Masa grasa

Este modelo considera las variables de masa ósea y masa magra para determinar si existe alguna relación con la masa grasa. Se considera la masa grasa total (TotalFg) como

la variable de respuesta y en este caso las variables óseas y magras como covariantes, sin considerar índices. Se complementa el análisis con las variables de edad, peso y talla.

Cuadro 4.23: Variables consideradas para el análisis de masa grasa

Variable	Descripción	Contexto
Height	Talla (m)	
Weight	Peso (kg)	
age	Edad	
RALg	Right Arm Lean Grams	Masa magra
LALg	Left Arm Lean Grams	Masa magra
RLLg	Right Leg Lean Grams	Masa magra
LLLg	Left Leg Lean Grams	Masa magra
TLg	Trunk Lean Grams	Masa magra
L1BMD	Lumbar 1 Densidad masa ósea	óseo
L2BMD	Lumbar 2 Densidad masa ósea	óseo
L3BMD	Lumbar 3 Densidad masa ósea	óseo
L4BMD	Lumbar 4 Densidad masa ósea	óseo
L1L4BMD	Lumbar 1-Lumbar 4 Densidad masa ósea	óseo
L2L4BMD	Lumbar 2-Lumbar 4 Densidad masa ósea	óseo
NeckFBMD	Cuello Femoral Densidad masa ósea	óseo
WardsBMD	Triangulo de Ward DMO	óseo
TrochBMD	Trocanter DMO	óseo
TotalFBMD	Total Femur bone mineral density	óseo

La variable de respuesta (TotalFg) se distribuye normalmente según las pruebas de hipótesis mostradas para hombres (cuadro 4.24) y para mujeres (cuadro 4.25)

Cuadro 4.24: Pruebas de normalidad para la variable grasa (TotalFg) en hombres

Test Name:	One-sample Kolmogorov-Smirnov test
Test Statistic:	D = 0.05201889
P-value:	0.005727021
Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Test Statistic:	D = 0.05201889
P-value:	2.897577e-07
Test Name:	Jarque Bera Test

Test Statistic:	X-squared = 60.5732
Test Statistic Parameter:	df = 2
P-value:	7.027712e-14

Cuadro 4.25: Pruebas de normalidad para la variable grasa (TotalFg) en mujeres

Test Name:	**One-sample Kolmogorov-Smirnov test
Test Statistic:	D = 0.08390724
P-value:	0.01575393
Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Test Statistic:	D = 0.08390724
P-value:	4.247766e-06
Test Name:	Jarque Bera Test
Test Statistic:	X-squared = 84.73846
Test Statistic Parameter:	df = 2
P-value:	0

El índice de masa grasa (IMG) se utiliza para definir la lipodistrofia. En este trabajo se definió con el criterio de $IMG \geq 1.5$ para hombres y $IMG \geq 1.2$ para mujeres. De esta forma, el 96.5 % de hombres y el 99.7 % de las mujeres presentan lipodistrofia (figura 4.14).

Modelo de masa grasa general

Hombres

El modelo de masa grasa inicial considera todas las variables, tanto óseas como magras para calcular el modelo. Al hacer un análisis de correlaciones se observa que ninguna de ellas es mayor a 0.3 lo que indica de entrada que la correlación entre variables es baja.

Cuadro 4.26: Correlaciones para la masa grasa total en hombres

TotalFg+Height: 0.11	TotalFg+Weight: 0.76	TotalFg+Age: 0.2
TotalFg+RALg: 0.07	TotalFg+LALg: 0.1	TotalFg+RLLg: 0.19
TotalFg+LLLg: 0.21	TotalFg+TLg: 0.02	TotalFg+L1BMD: 0.11
TotalFg+L2BMD: 0.07	TotalFg+L3BMD: 0.02	TotalFg+L4BMD: 0.07
TotalFg+L1L4BMD: 0.07	TotalFg+L2L4BMD: 0.05	TotalFg+NeckFBMD: 0.08
TotalFg+WardsBMD: -0.02	TotalFg+TrochBMD: 0.13	TotalFg+TotalFBMD: 0.12

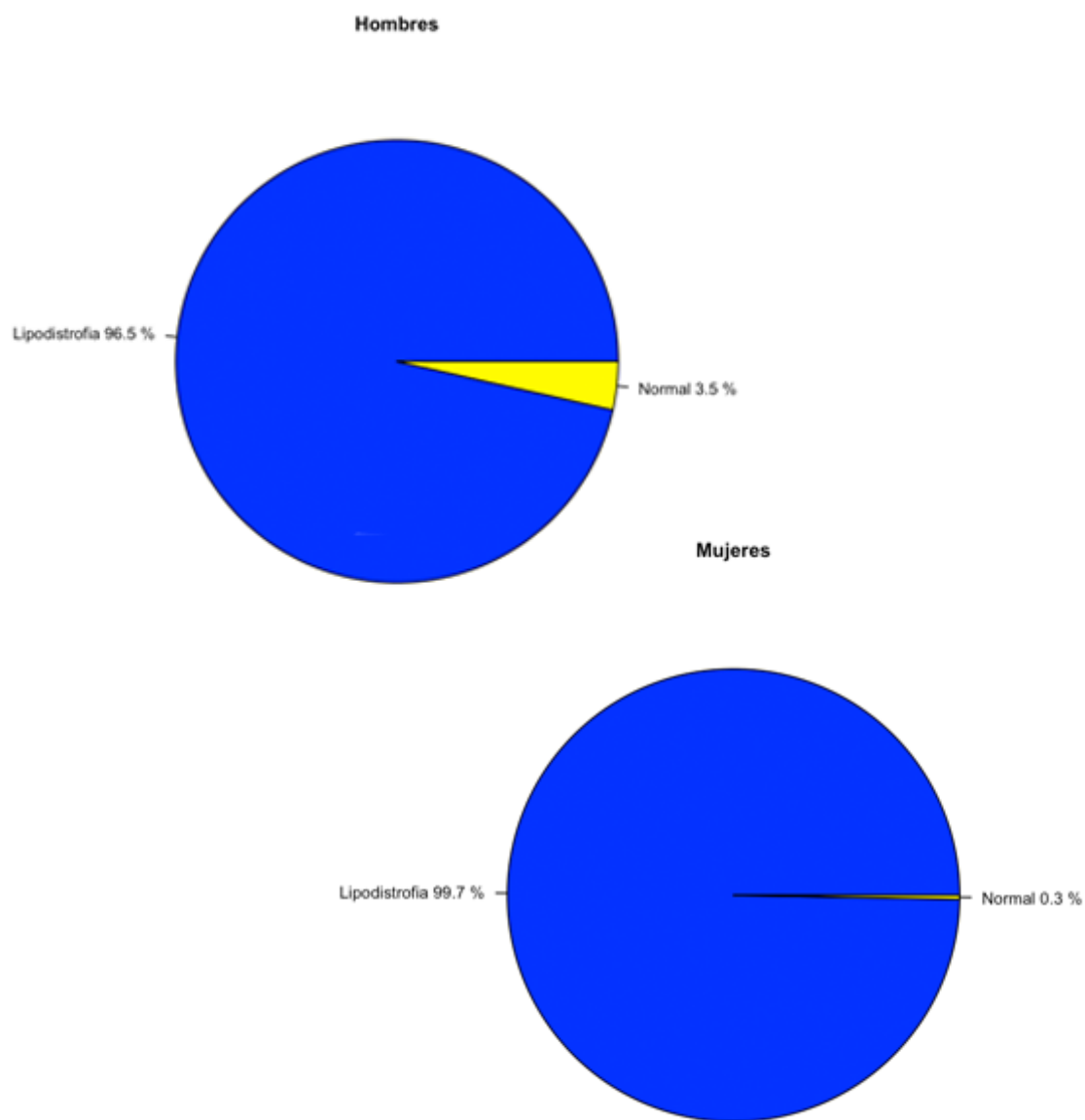


Figura 4.14: Distribución masa grasa

Para el modelo de regresión donde se utilizan todas las covariantes del estudio se explica el 97.48 % de la variabilidad (18 variables). Una reducción por el método Stepwise (9 variables) explica el 97.49 % de la variabilidad.

El modelo reducido de masa grasa total contempla las variables de masa magra (masa magra en ambos brazos, masa magra en ambas piernas y masa magra en el tronco) y solo una variable de masa ósea, la masa ósea entre las lumbares 2 y 4, además de la talla, el peso y la edad.

$$\text{TotalFg} \sim \text{Height} + \text{Weight} + \text{Age} + \text{RALg} + \text{LALg} + \text{RLLg} + \text{LLLg} + \text{TLg} + \text{L2L4BMD}$$

Los valores de la variable de respuesta van de 184 a 44,160 (gramos totales de grasa). El intercepto tiene un valor menor al primer cuartil (casi dos kilos de grasa cuando todas los coeficientes valen 0). El valor negativo en la talla indica que por cada metro de talla, el valor de masa grasa disminuye en -4.230 kg de masa grasa, cuando todos los coeficientes valen cero. La relación con el peso es que por cada kilo de aumento en peso del individuo, la masa grasa aumenta en aproximadamente un kilo. En la edad, la masa grasa disminuye en apenas 6 gramos por cada año de edad del paciente, cuando el resto de los coeficientes vale 0. El último punto a destacar es la masa ósea en las lumbares 2 y 4. La masa grasa disminuye un kilo en función de la masa ósea en este punto. En este análisis se destaca el signo de los coeficientes. En todos, excepto el peso, hay disminución de masa grasa al aumentar estos coeficientes.

Los coeficientes de este modelo se muestran en el cuadro 4.27.

Cuadro 4.27: Coeficientes del modelo lineal masa grasa para hombres

	Valor	Desviación
(Intercept)	1979.573	1078.0
Height	-4230.013	684.1
Weight	985.408	5.26
Age	-5.963	3.67
RALg	-0.697	0.134
LALg	-1.846	0.153
RLLg	-0.921	0.093
LLLg	-1.441	0.096
TLg	-0.781	0.016
L2L4BMD	-1080.4887	248.0

El modelo reducido de masa grasa que tiene 9 variables en comparación con las 18 del modelo original tiene un buen ajuste, según la gráfica de residuales de los datos observados y esperados (figura 4.15). Salvo un pequeño grupo de datos que no se ajustan, la mayoría están sobre la curva hipotética de máximo ajuste.

El análisis de residuales muestra que el ajuste es bueno, aunque en la curva qq se observan valores extremos y una distribución con dos picos.

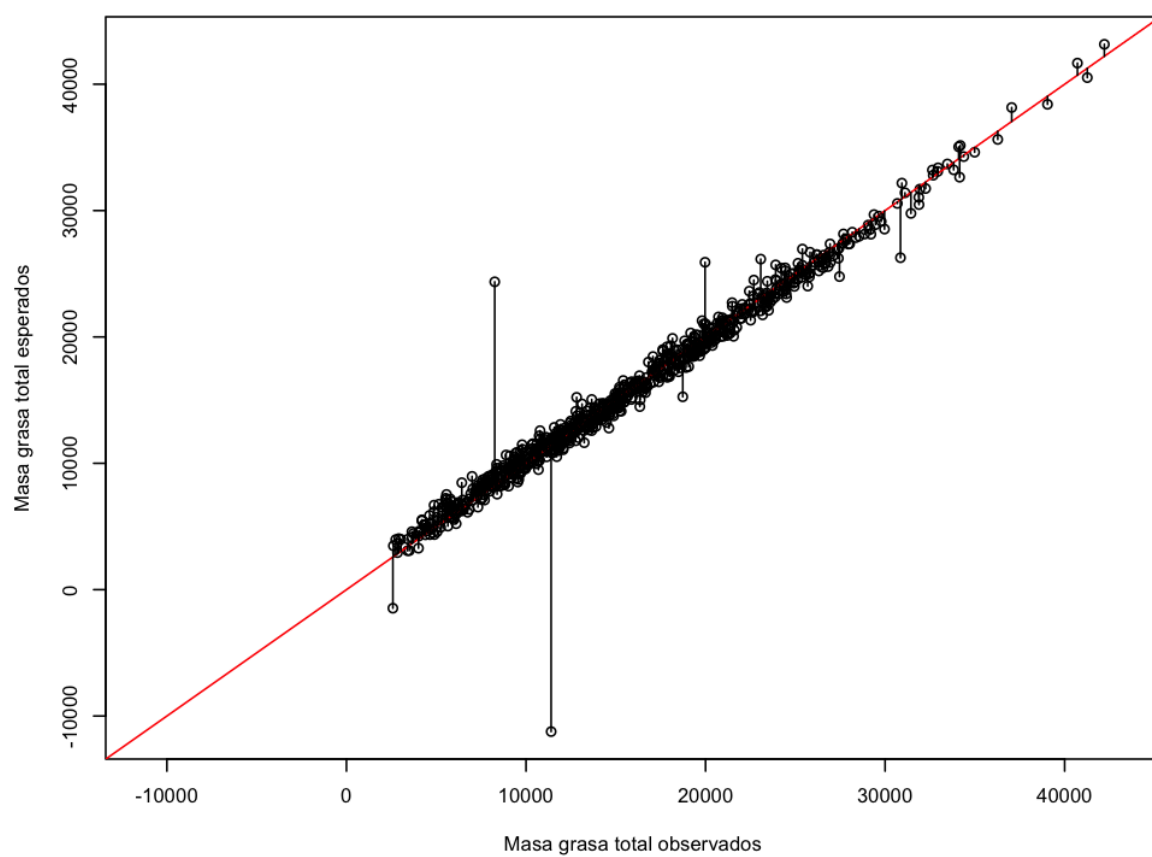
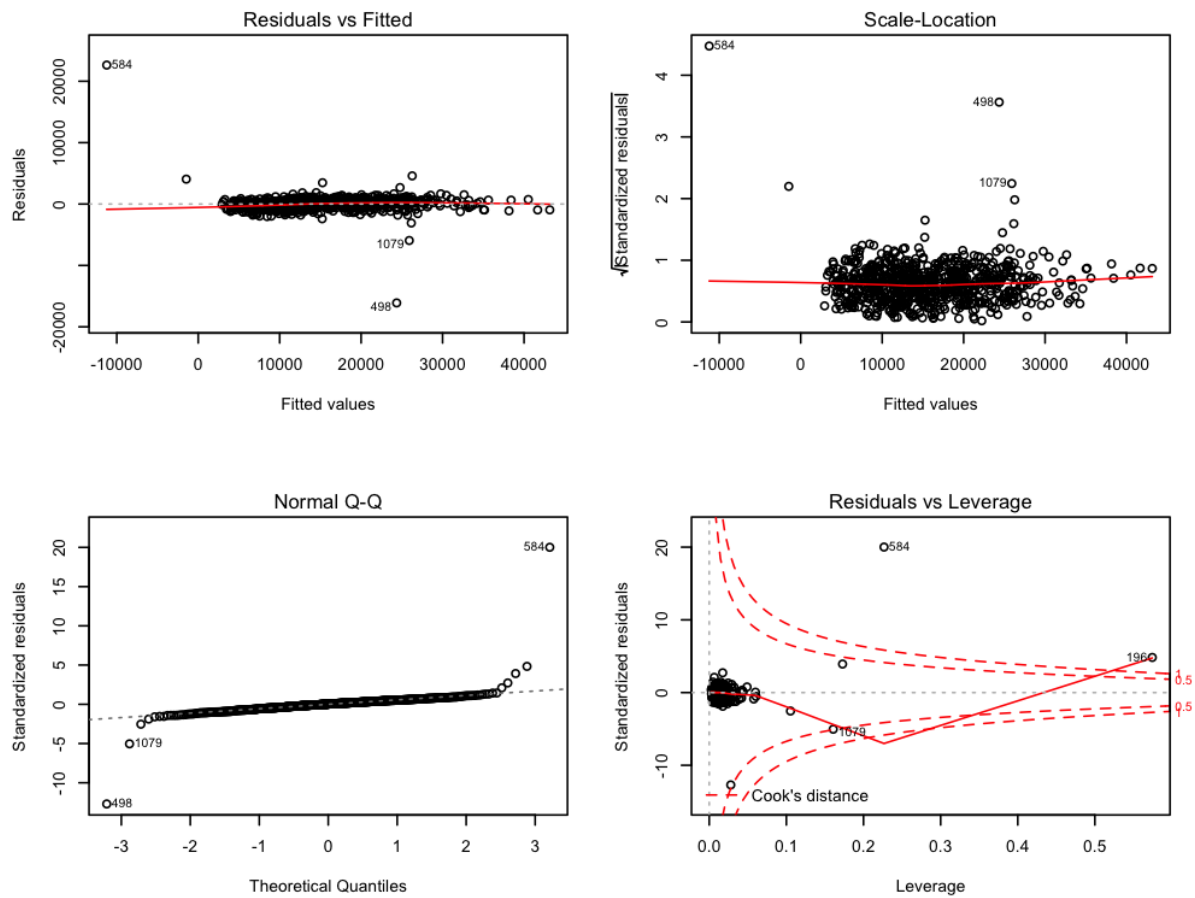


Figura 4.15: Residuales de modelo lineal masa grasa hombres



Mujeres

El análisis de correlaciones de las variables óseas y magras con la variable de respuesta muestra una correlación alta con el peso y en menor grado con la masa magra en ambos brazos.

Cuadro 4.28: Correlaciones para la masa grasa total en mujeres

TotalFg+Weight: 0.88
TotalFg+RALg: 0.32
TotalFg+LALg: 0.38

Para el modelo de regresión donde se utilizan todas las covariantes del estudio se explica el 98.89 % de la variabilidad (18 variables). Una reducción por el método Stepwise (10 variables) explica el 97.5 % de la variabilidad.

$$\text{TotalFg} \sim \text{Height} + \text{Weight} + \text{Age} + \\ \text{RALg} + \text{LALg} + \text{RLLg} + \text{LLLg} + \text{TLg} + \\ \text{L1L4BMD} + \text{TotalFBMD}$$

Los coeficientes de este modelo se muestran en el cuadro 4.29

Cuadro 4.29: Coeficientes del modelo lineal masa grasa mujeres

	Valor	Desviación
(Intercept)	5456.62	1639.0
Height	-4981.36	1055.0
Weight	987.93	6.59
Age	-9.77	6.07
RALg	-0.91	0.465
LALg	-1.94	0.467
RLLg	-1.34	0.194
LLLg	-0.78	0.188
TLg	-0.87	0.029
L1L4BMD	-926.28	442.0
TotalFBMD	-1276.20	535.4

El valor del intercepto es dos veces el valor mínimo de la variable de respuesta. Solo el peso tiene un coeficiente positivo, que se interpreta como que por cada kilo de peso, el total de masa grasa en las mujeres aumenta un kilo también (muy parecido a lo que pasa en los hombres). El resto de las variables del modelo reducido (masa magra en ambos brazos, en ambas piernas y en el tronco) presentan valores negativos, lo que indica que un aumento en alguna de estas variables cuando el resto vale cero disminuye la masa grasa total. Lo mismo ocurre con los coeficientes de variables óseas (en las lumbares 1 y 4 y con el total de densidad ósea del fémur).

El análisis de residuales de los datos observados y esperados muestra que el modelo se ajusta bien, aunque hay solo dos datos que muestran diferencias significativas.

El análisis de residuales (figura 4.17) muestra que el ajuste del modelo es bueno, aunque la curva qq presenta valores extremos.

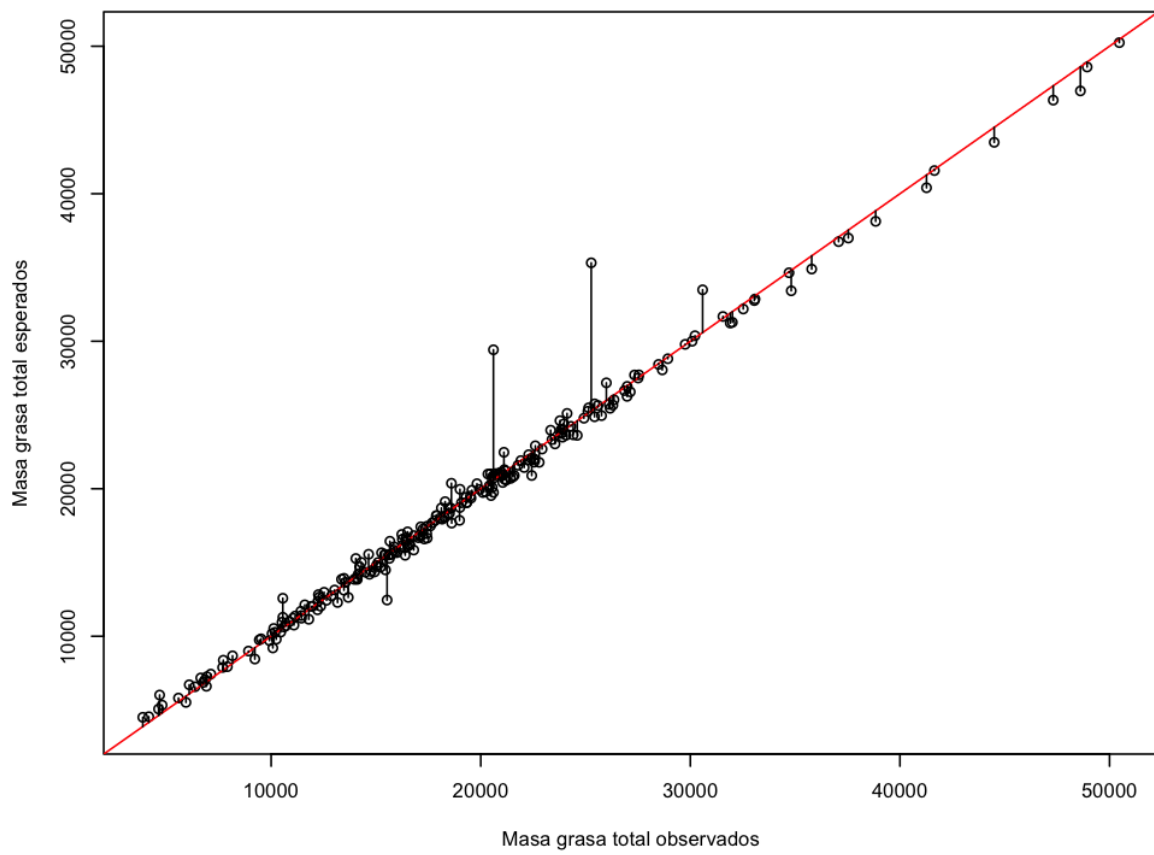


Figura 4.16: Residuales de modelo lineal masa grasa mujeres

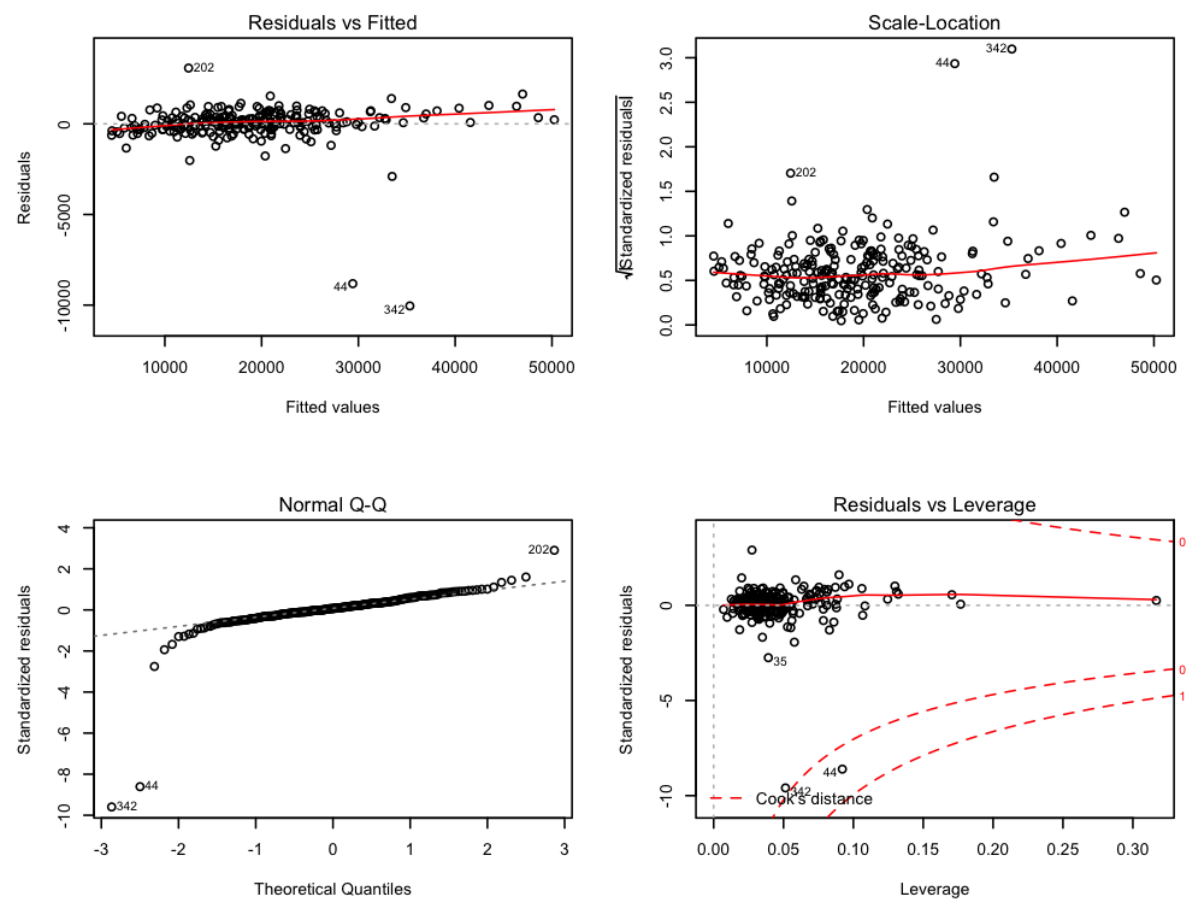


Figura 4.17: Análisis de residuales modelo masa grasa mujeres

Modelo de clasificación

El modelo de clasificación utilizado es el método de kNN vecinos más próximos. Para los hombres se obtuvo una predicción del 96 % y un error de predicción del 4 % para el conjunto de validación. En el caso de las mujeres no hubo falsos positivos ni falsos negativos quizá por el tamaño de la muestra y el porcentaje de lipodistrofia del conjunto de entrenamiento.

Hombres

Cuadro 4.30: Tabla de clasificación para hombres masa grasa método kNN (k=21, accuracy=0.96)

Estado graso	Lipodistrofia	Normal	Total renglón
Lipodistrofia	207	2	209
	0.990	0.010	0.963
	0.972	0.500	
	0.954	0.009	
Normal	6	2	8
	0.750	0.250	0.037
	0.028	0.500	
	0.028	0.009	
Total columna	213	4	217
	0.982	0.018	

Mujeres

Cuadro 4.31: Tabla de clasificación para mujeres masa grasa método kNN (k=21)

Estado graso	Lipodistrofia	Total renglón
Lipodistrofia	68	68
	0.986	
Normal	1	1
	0.014	
Total columna	69	69

Masa magra

Para el análisis de masa magra se consideró como variable de respuesta el total de masa magra (TotalLg) y como variables independientes las óseas y grasas (cuadro 4.32). El análisis de complemento con el peso, la tala y la edad.

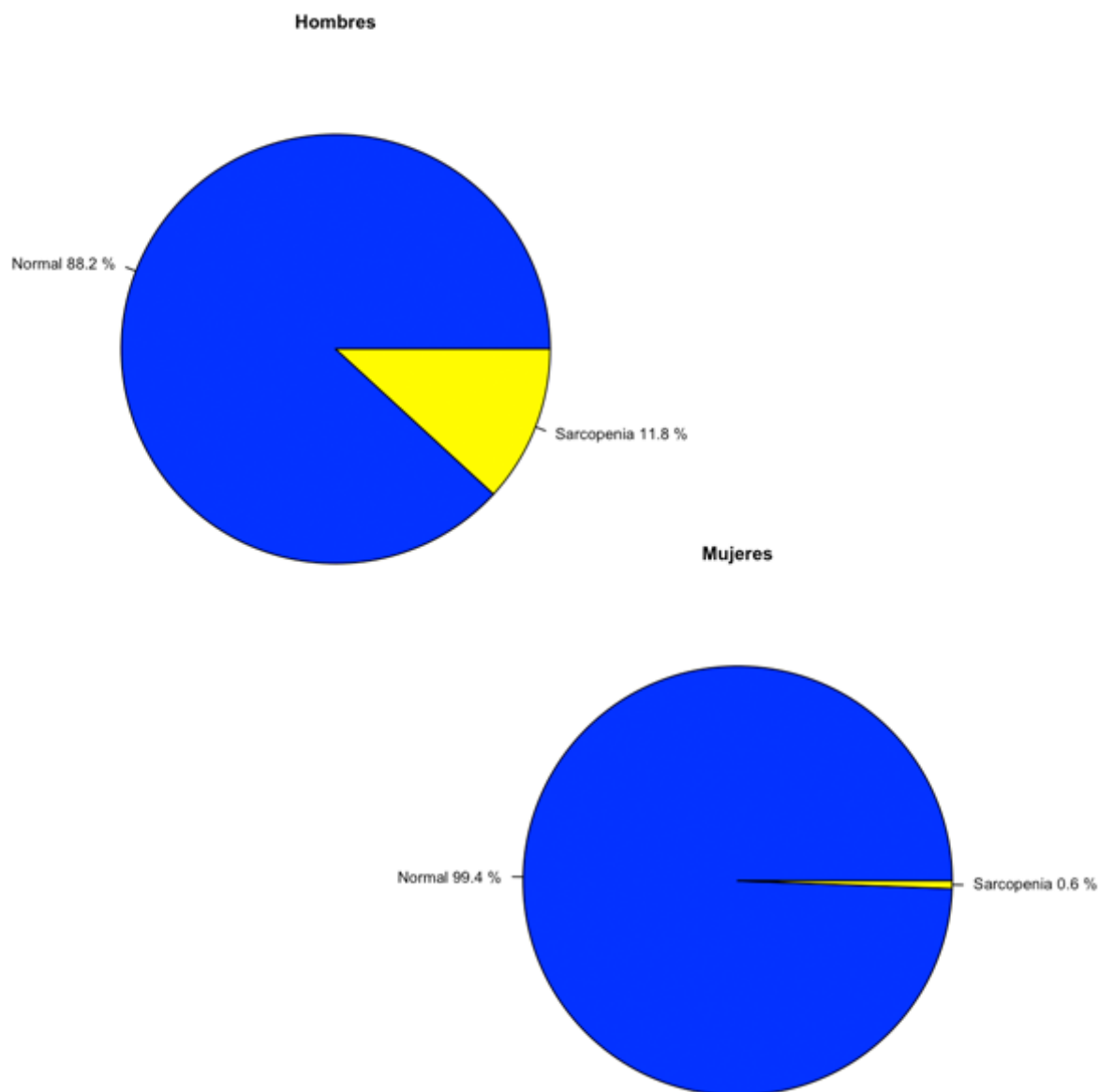


Figura 4.18: Distribución masa magra

El 11.8% de los hombres presentan la condición de sarcopenia tal como se calcula a partir del índice de masa magra (IMM). En este caso son aquellos individuos con un IMM (FFMI en inglés) menor a $7.26 \frac{kg}{m^2}$. En las mujeres el límite es de $5.5 \frac{kg}{m^2}$ y la presenta apenas el 0.6% (figura 4.18).

Cuadro 4.32: Variables independientes para el modelo de masa magra

Variable	Descripcion	Contexto
Height	Talla (m)	
Weight	Peso (kg)	
age	Edad	
RAFg	Right Arm Fat Grams	Masa grasa
LAfG	Left Arm Fat Grams	Masa grasa
RLFg	Right Leg Fat Grams	Masa grasa
LLFg	Left Leg Fat Grams	Masa grasa
TFg	Trunk Fat Grams	Masa grasa
L1BMD	Lumbar 1 Densidad masa ósea	óseo
L2BMD	Lumbar 2 Densidad masa ósea	óseo
L3BMD	Lumbar 3 Densidad masa ósea	óseo
L4BMD	Lumbar 4 Densidad masa ósea	óseo
L1L4BMD	Lumbar 1-Lumbar 4 Densidad masa ósea	óseo
L2L4BMD	Lumbar 2-Lumbar 4 Densidad masa ósea	óseo
NeckFBMD	Cuello Femoral Densidad masa ósea	óseo
WardsBMD	Triangulo de Ward DMO	óseo
TrochBMD	Trocanter DMO	óseo
TotalFBMD	Total Femur bone mineral density	óseo

La variable de respuesta (TotalLg) se distribuye normalmente en hombres y no se distribuye normalmente en mujeres, según las pruebas de hipótesis realizadas (cuadros 4.33 y 4.34).

Cuadro 4.33: Pruebas de normalidad para la variable magra (TotalLg) en hombres

Test Name:	One-sample Kolmogorov-Smirnov test
Test Statistic:	D = 0.04195949
P-value:	0.04429871
Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Test Statistic:	D = 0.04195949
P-value:	0.0001226718
Test Name:	Jarque Bera Test
Test Statistic:	X-squared = 192.5752
Test Statistic Parameter:	df = 2
P-value:	0

Cuadro 4.34: Pruebas de normalidad para la variable magra (TotalLg) en mujeres

Test Name:	One-sample Kolmogorov-Smirnov test
Test Statistic:	D = 0.06286123
P-value:	0.1318889
Test Name:	Lilliefors (Kolmogorov-Smirnov) normality test
Test Statistic:	D = 0.06286123
P-value:	0.002314665
Test Name:	Jarque Bera Test
Test Statistic:	X-squared = 75.05737
Test Statistic Parameter:	df = 2
P-value:	0

Modelo de regresión

Hombres

El análisis de correlaciones de las variables óseas y grasas con respecto a la variable de respuesta, total de masa magra (TotalLg) arroja que hay una correlación de casi 0.6 con la talla de los individuos hombres, de 0.7 con respecto al peso y una correlación significativa con otras variables óseas (cuadro 4.35).

Correlaciones:

Cuadro 4.35: Correlaciones para la masa magra total en hombres

TotalLg+Height: 0.59
TotalLg+Weight: 0.71
TotalLg+L1L4BMD: 0.31
TotalLg+NeckFBMD: 0.36
TotalLg+TrochBMD: 0.41
TotalLg+TotalFBMD: 0.38

El modelo de regresión de la masa magra contra las variables óseas y grasas (18 variables) explica el 98.11 de la variabilidad. La reducción de variables por el método stepwise selection arroja un total de 8 variables y explica el 98.1 % de la variabilidad. El modelo reducido incluye el peso, la talla, la masa grasa en ambos brazos, en ambas piernas, en el tronco y la densidad de masa ósea en la cuarta lumbar y en el cuello femoral.

$$\text{TotalLg} \sim \text{Height} + \text{Weight} + \text{RAFg} + \text{LAFg} + \text{RLFg} + \text{LLFg} + \text{TFg} + \text{NeckFBMD}$$

Los coeficientes del modelo se muestran en el cuadro 4.36. El valor mínimo de la masa magra (TotalLg) es de 33.45 kg y el máximo es de 88.9 kg. El modelo predice que cuando todos los coeficientes valen cero, el valor inicial es de 1.7 kg. También se puede observar que por cada metro de altura el individuo tiene 1.4 kg de masa magra menos. Por cada kilo

de peso, la masa magra aumenta en casi un kilo. Otro coeficiente significativo es la masa ósea en el cuello femoral. Por cada gramo en este coeficiente la masa magra disminuye en un kilo.

Cuadro 4.36: Coeficientes del modelo lineal graso en hombres

	Valor	Desviación
(Intercept)	1660.50	0.085
Height	-1421.46	0.057
Weight	977.25	5.67
RAFg	-0.83	0.515
LAFg	-1.42	0.51
RLFg	-1.26	0.27
LLFg	-0.81	0.27
TFg	-0.97	0.012
NeckFBMD	-1004.67	0.024

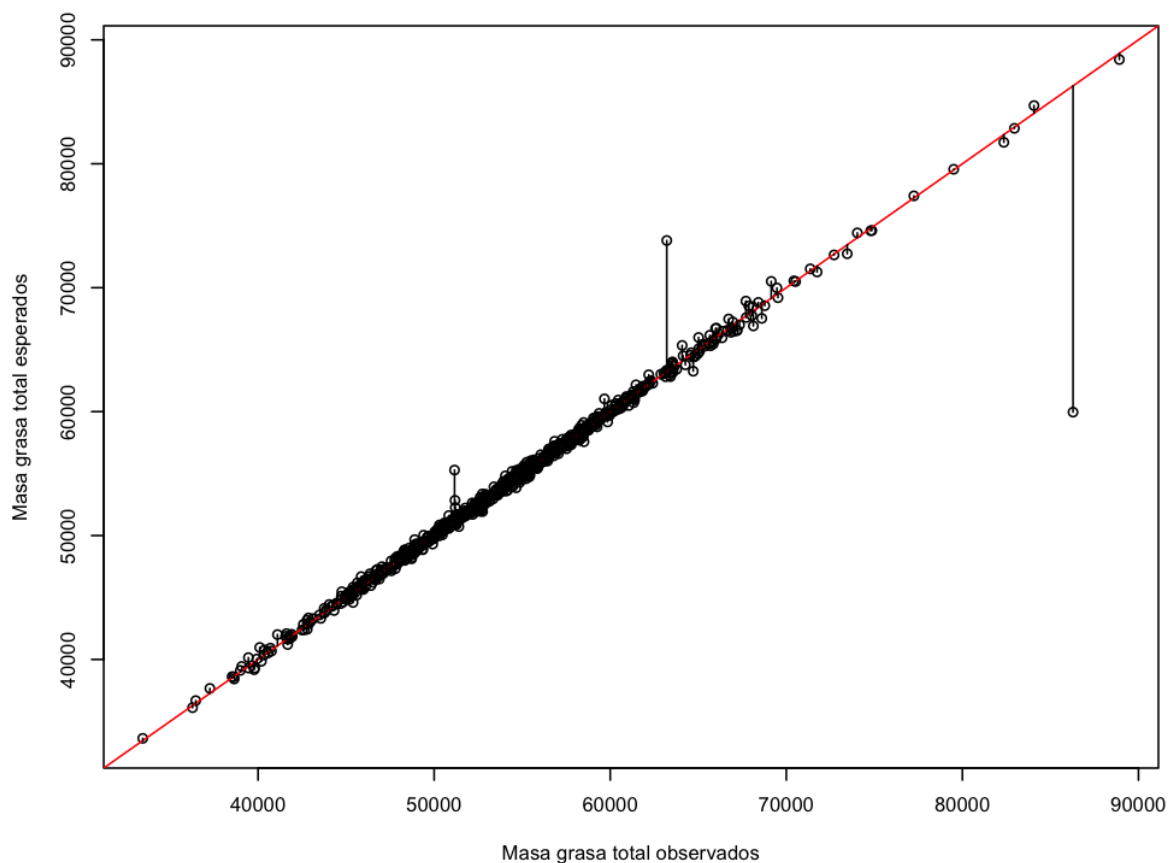


Figura 4.19: Residuales de modelo lineal reducido masa magra hombres

Se ajustó el modelo a un conjunto de datos de entrenamiento (80 %) y se probó con el 20 % restante. El gráfico de los residuales (valores observados contra esperados en el conjunto de prueba) se muestra en la figura 4.19

En este caso se observa un buen ajuste, ya que la mayoría de los datos están muy cercanos a la línea de referencia y solo 3 de ellos presentan valores dispersos.

El análisis de residuales muestra, en la gráfica 4.20, estos valores que son outliers. Sin embargo el ajuste se puede considerar bueno y los residuales distribuidos normalmente.

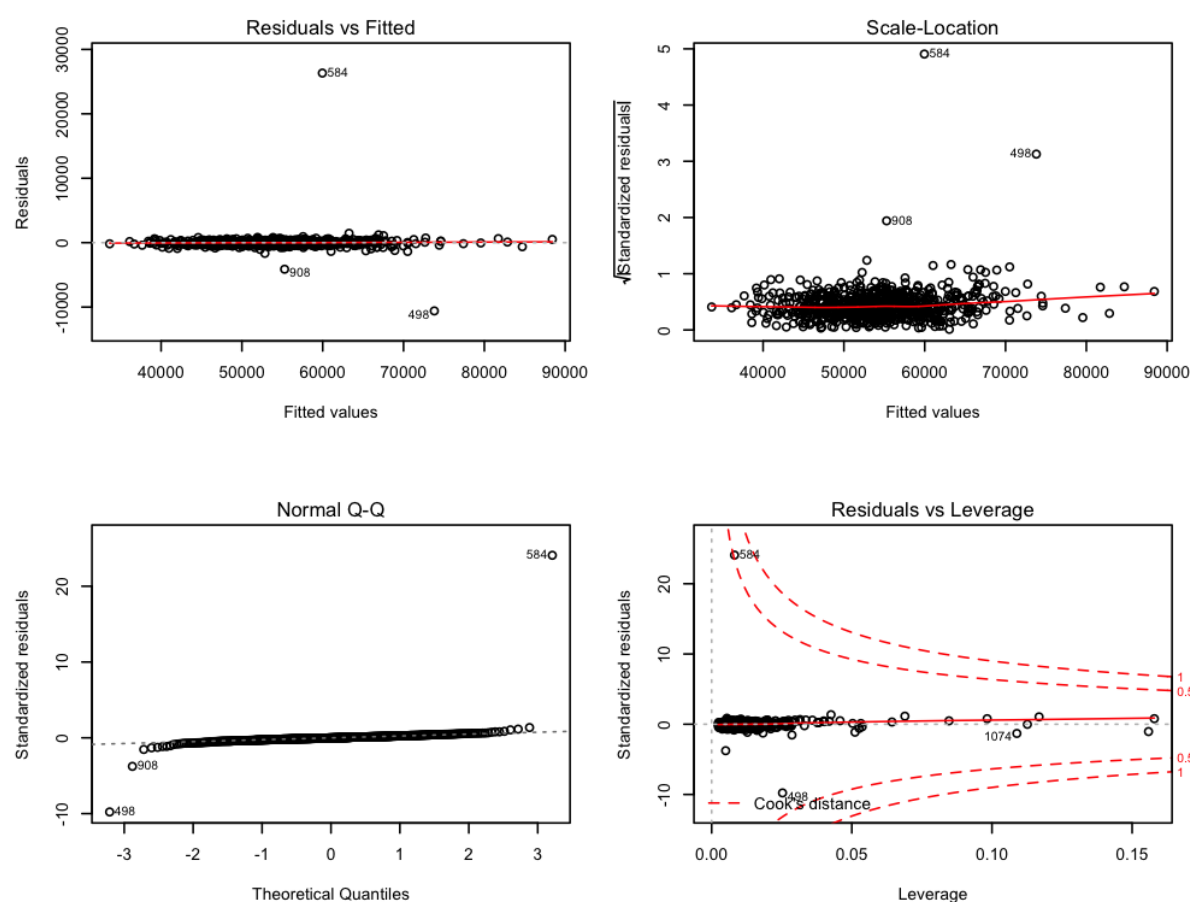


Figura 4.20: Análisis de residuales

Mujeres

En cuanto a las mujeres, las variables que tienen una correlación significativa con la variable de respuesta (TotalLg) son el peso, la talla, varias variables óseas (lumbares 1, 3, 4, 1-4, 2-4, cuello femoral, triángulo de ward, trocanter, total de masa ósea de femur). Destaca que no hay correlación mayor con 0.3 con alguna variable grasa (cuadro 4.37).

Cuadro 4.37: Correlaciones para la masa magra total en mujeres

TotalLg+Height: 0.58
TotalLg+Weight: 0.62
TotalLg+L1BMD: 0.31
TotalLg+L3BMD: 0.34
TotalLg+L4BMD: 0.35
TotalLg+L1L4BMD: 0.34
TotalLg+L2L4BMD: 0.34
TotalLg+NeckFBMD: 0.38
TotalLg+WardsBMD: 0.34
TotalLg+TrochBMD: 0.43
TotalLg+TotalFBMD: 0.41

Hay que recordar que la variable TotalLg para mujeres no se distribuye normalmente según las pruebas de hipótesis realizadas. El gráfico correspondiente (4.21) muestra que la distribución de esta variable tiene dos picos y está sesgada a la izquierda.

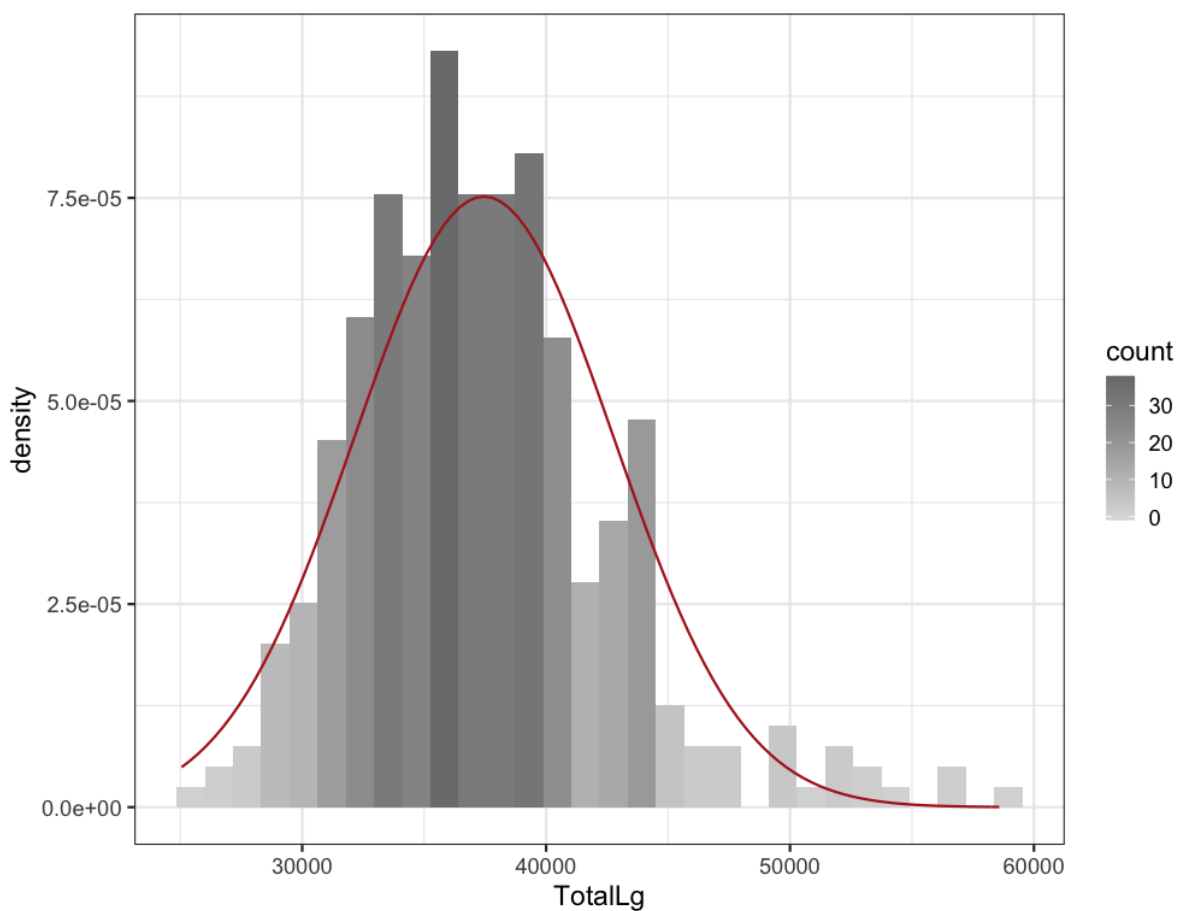


Figura 4.21: Distribución de la masa ósea total mujeres

El modelo lineal completo con 18 variables explica el 99.47 % de la variabilidad. Al hacer la reducción por el método stepwise se obtiene un modelo con 9 variables que explica el 99.48 % de la variabilidad. Este modelo incluye las variables peso y talla, además de la masa grasa en ambas piernas, ambos brazos, la masa grasa en el tronco y la masa ósea en las lumbares 1 y 4 y en el triángulo de ward. Los coeficientes de la regresión se muestran en el cuadro 4.38

$$\text{TotalLg} \sim \text{Height} + \text{Weight} + \text{RAFg} + \text{LAFg} + \text{RLFg} + \text{LLFg} + \text{TFg} + \text{L1L4BMD} + \text{WardsBMD}$$

Cuadro 4.38: Coeficientes del modelo lineal graso en mujeres

	Valor	Desviación
(Intercept)	3043.20	574.2
Height	-2213.24	420.4
Weight	993.10	5.58
RAFg	-1.02	0.224
LAFg	-0.88	0.228
RLFg	-1.16	0.086
LLFg	-0.92	0.088
TFg	-1.02	0.01
L1L4BMD	-956.92	176.2
WardsBMD	-637.67	174.9

Los valores de la variable TotalLg van de 25,056 a 58, 578 gramos de masa magra. El modelo propuesto tiene un intercepto de 3,043-2, que es el valor del modelo cuando todos los coeficientes valen cero. La interpretación del modelo dice que por cada metro de altura, la masa magra disminuye en 2.21 kg; por cada kilo de peso del individuo, aumenta la masa magra en un kilo aproximadamente. El resto de los coeficientes son negativos, por lo que su aumento disminuye la masa magra, cuando el resto de los coeficientes valen cero.

La gráfica de los residuales ajustados al conjunto de prueba de los datos muestra que el ajuste es muy bueno, con un par de datos outliers (figura 4.22).

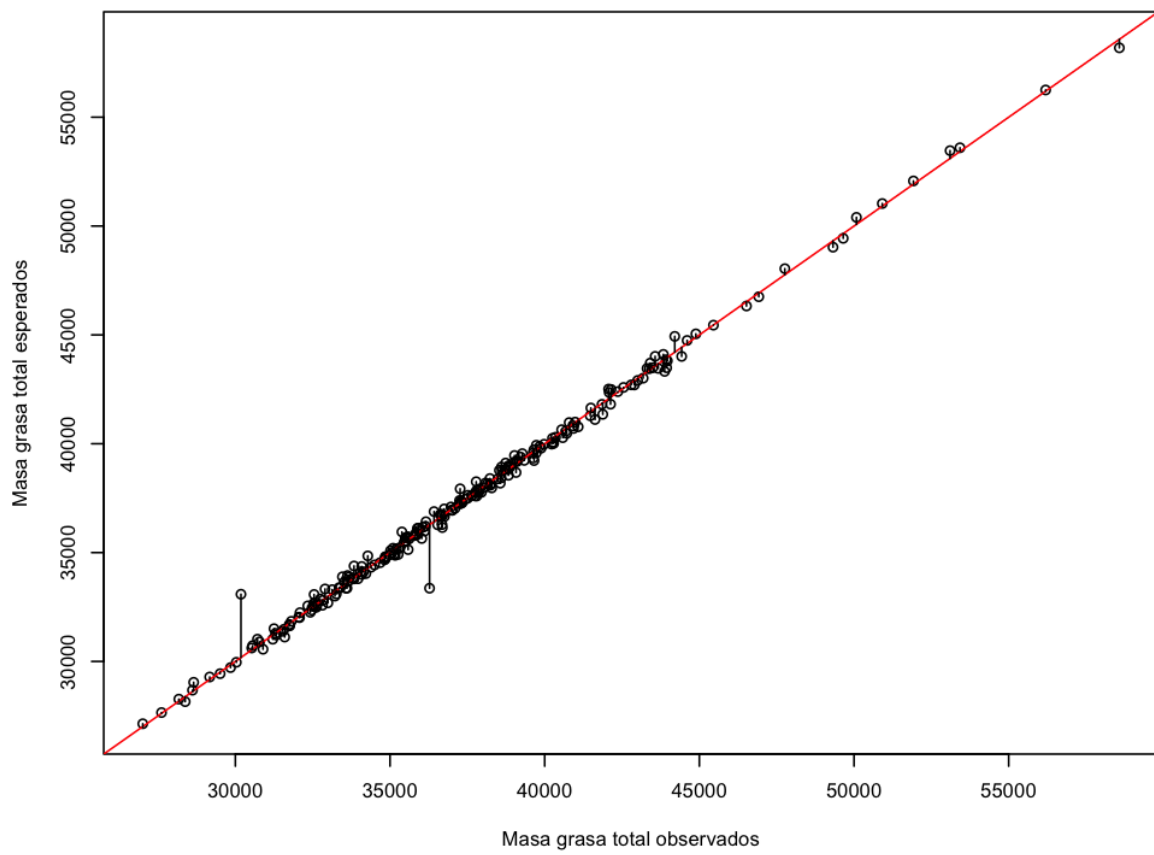


Figura 4.22: Residuales de modelo lineal masa magra mujeres

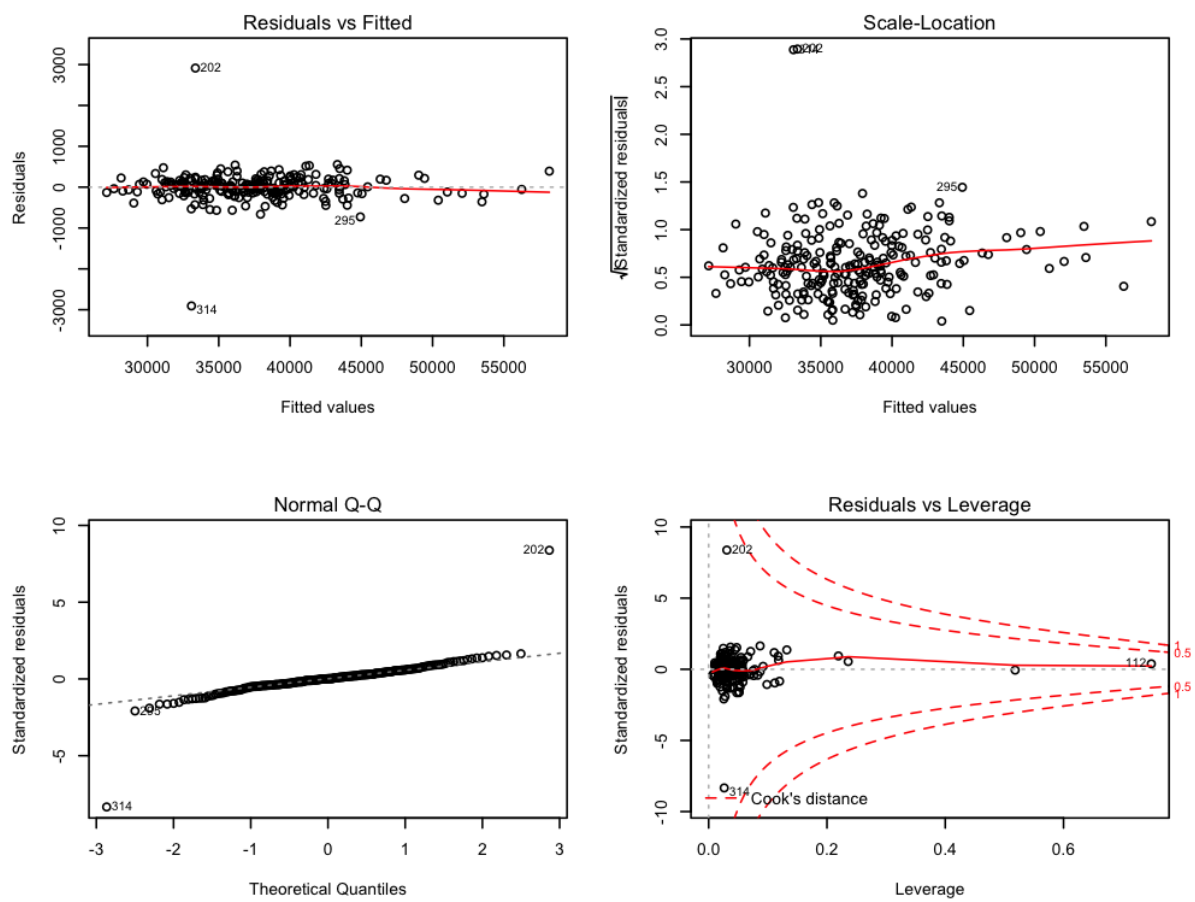


Figura 4.23: Análisis de residuos

Modelo de clasificación

El modelo de clasificación utilizado es el método de kNN vecinos más próximos. Para los hombres se obtuvo una predicción del 91.7 % y un error de predicción del 7.3 % para el conjunto de validación. En el caso de las mujeres no hubo falsos positivos ni falsos negativos quizá por el tamaño de la muestra y el porcentaje de sarcopenia del conjunto de entrenamiento (cuadros 4.39 y 4.40).

Cuadro 4.39: Tabla de clasificación para hombres magra
método kNN (k=21, accuracy=0.917)

Estado magro	Normal	Sarcopenia	Total renglón
Normal	194	4	198
	0.980	0.020	0.912
	0.933	0.444	
	0.894	0.018	
Sarcopenia	14	5	19
	0.737	0.263	0.088
	0.067	0.556	
	0.065	0.023	
Total columna	208	9	217
	0.959	0.041	

Cuadro 4.40: Tabla de clasificación para mujeres magra
método kNN (k=21)

Estado magro	Normal	Total renglón
Normal	67	67
	0.971	
Sarcopenia	2	2
	0.029	
Total columna	69	69

Análisis longitudinal masa ósea

Para el análisis longitudinal de los datos se consideró la masa ósea total (TotalBMD) como la variable de respuesta y las variables grasas, la edad, el peso y la talla como las variables independientes. Dos factores son importantes a considerar en este tipo de análisis: el tiempo, ya que los individuos presentan más de una toma de datos en diferentes tiempos y las medidas repetidas, ya que los individuos tienen más de una toma, pero no es constante. En la gráfica 4.24 se observa que hay muchos individuos con dos tomas de dexa (287) y hay inclusive uno que tiene 18 tomas.

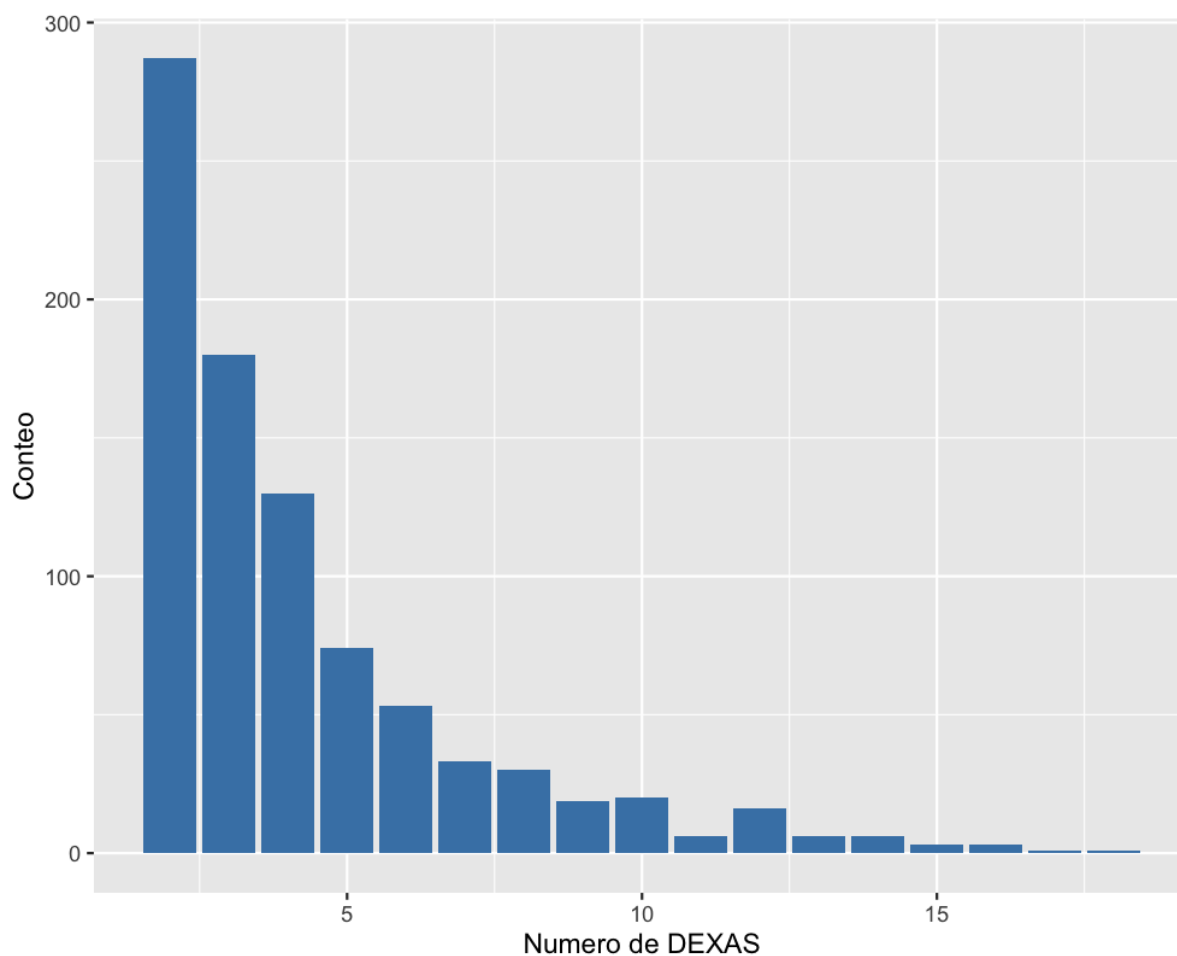


Figura 4.24: Número de DEXAs registradas para un mismo paciente

De los 4,319 registros de la base depurada (sin datos con NA), se eliminaron aquellos con solo una dexa, quedando la base con 3,545 observaciones correspondientes a 1,473 individuos. La base de datos dividida por sexo arroja 2,730 registros para hombres y 815 registros para mujeres. Se realizó un ajuste de datos para el tiempo, convirtiendolo en una variable donde hay unidades de tiempo, desde el tiempo de la primera toma hasta la última. En total se agruparon los datos en 70 unidades de tiempo uniforme.

Cuadro 4.41: Número de DEXAs registradas para un mismo paciente

frecuencia	cuantos
2	287
3	180
4	130
5	74
6	53
7	33
8	30
9	19
10	20
11	6
12	16
13	6
14	6
15	3
16	3
17	1
18	1

Para observar cómo afecta el tiempo a las observaciones de masa ósea total se tomaron al azar el 10 % de la muestra para hombres (273 datos) y el 20 % para mujeres (163 datos) y se realizaron los perfiles, esperando encontrar un patrón (figura 4.25).

Como tendencia general se observa que hay una disminución de la masa ósea con el tiempo, sobre todo en aquellas observaciones con más de dos dexas. El tiempo será entonces una variable a considerar en el análisis de regresión lineal. También se observan individuos que tienden a subir su masa ósea con el tiempo. Hay otros que inclusive bajan y luego suben su masa ósea. Esto sugiere que hay una variabilidad dentro de los individuos. Esta sería otra variable a considerar.

Modelo de Regresión lineal

Hombres

Para el modelo de regresión lineal de los datos longitudinales en masa ósea se tomó la densidad de masa ósea total como la variable de respuesta y tanto las variables magras y grasas son las covariantes.

También es necesario considerar la posible variabilidad dentro de las observaciones y que el tiempo es la variable de importancia a estudiar.

Este modelo determina el tiempo como un factor aleatorio para cada ID y las variables grasas y magras como efectos fijos (que no dependen del tiempo).

Se dividió la base de datos en entrenamiento (80 %) para ajustar el modelo y de prueba (20 %) para probarlo.

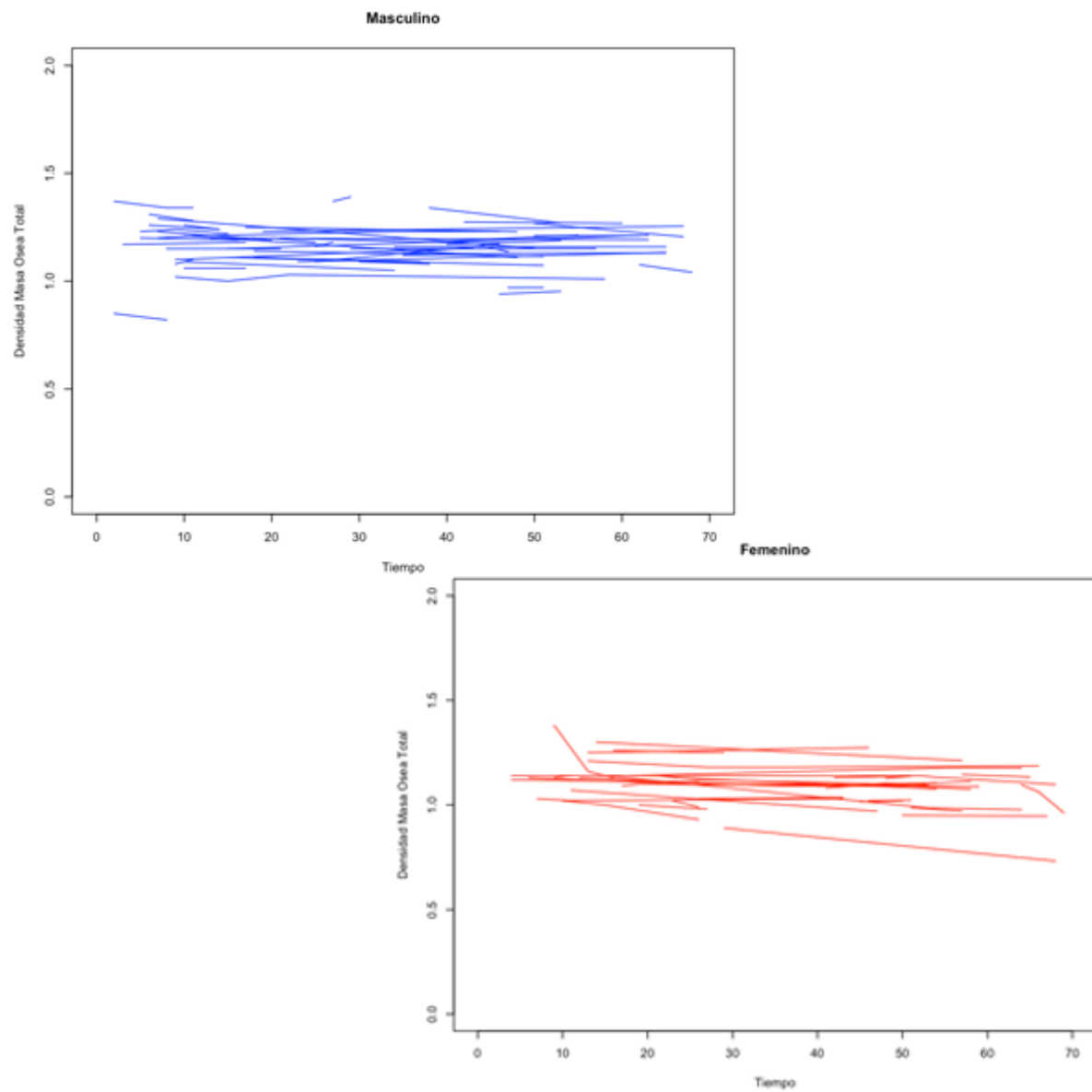


Figura 4.25: Perfiles individuales del 20 % de las observaciones seleccionados al azar. Masa ósea

El ajuste del modelo lineal muestra que hay coeficientes positivos (como el tiempo, el peso, la masa magra en ambos brazos y masa grasa en brazo izquierdo) y coeficientes negativos (la edad, la talla, la masa grasa en brazo derecho, ambas piernas y en el tronco, la masa magra en ambas piernas y en el tronco).

Se muestran los coeficientes de la regresión y la matriz de correlación obtenida.

Cuadro 4.42: Coeficientes fijos del modelo longitudinal óseo para hombres

	Valor	Desviación
(Intercept)	1.004	0.075
tiempo	0.0002	0.0001
Age	-0.001	0.0003
Height	-0.079	0.045
Weight	0.006	0.0009
RAFg	-0.00002	0.00002
RALg	0.000	0.000004
LAFg	0.000001	0.00002
LALg	0.000001	0.000006
RLFg	-0.000005	0.000008
RLLg	-0.000001	0.000005
LLFg	-0.000002	0.000008
LLLg	-0.000002	0.000005
TFg	-0.000003	0.000001
TLg	-0.000001	0.0000008

Cuadro 4.43: Efectos aleatorios del modelo longitudinal óseo para hombres

	Desviación	Correlación
(Intercept)	0.097	
tiempo	0.001	-0.65
Residual	0.038	

La parte de los resultados que se refiere a los efectos aleatorios muestra que la varianza residual es de $\sigma^2 = 0.038^2$ y la varianza de la constante es $\sigma^2 = 0.096^2$. La varianza explicada por el factor tiempo es muy pequeña $\sigma^2 = 0.001^2$. Para la parte de los efectos fijos del modelo la constante se estima en 1.004 y el valor de los diferentes factores es prácticamente 0, excepto para el peso (-0.079), que además es negativo.

El modelo se ajustó a los datos de prueba y se obtuvo el gráfico de datos originales y observados, donde se detectan algunos datos que no se ajustan al modelo, aunque la mayoría se ajustan bien.

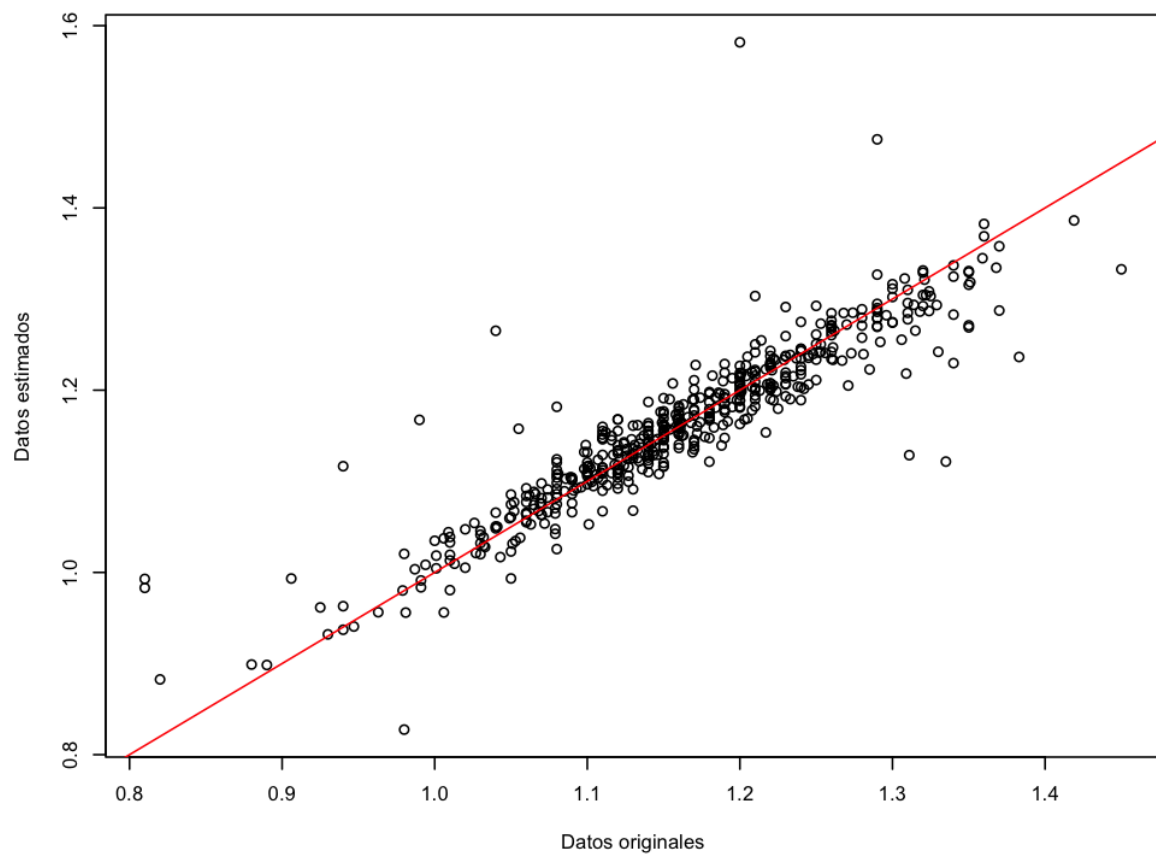


Figura 4.26: Regresión modelo longitudinal hombres

Correlation:

	(Intr)	tiempo	Age	Height	Weight	RAFg	RALg	LAfG	LALg	RLFg
tiempo	0.160									
Age	-0.342	-0.543								
Height	-0.933	-0.140	0.167							
Weight	0.010	0.113	-0.091	-0.150						
RAFg	0.014	-0.028	-0.025	0.010	-0.152					
RALg	-0.008	-0.061	0.053	0.022	-0.112	-0.381				
LAfG	-0.002	0.072	-0.003	0.003	0.034	-0.935	0.387			
LALg	-0.015	-0.135	0.077	0.032	-0.263	0.409	-0.472	-0.413		
RLFg	-0.012	-0.015	0.090	0.000	-0.173	-0.011	0.021	-0.014	0.037	
RLLg	0.053	0.040	0.014	-0.049	-0.330	-0.006	-0.021	0.060	-0.020	-0.181
LLFg	-0.001	-0.076	0.042	0.008	-0.063	0.003	0.009	-0.018	0.037	-0.891
LLLg	-0.002	-0.044	0.066	0.001	-0.286	0.062	-0.020	-0.017	0.003	0.309
TFg	0.008	-0.182	-0.085	0.132	-0.759	0.059	0.114	-0.043	0.221	0.051
TLg	0.017	0.018	-0.014	-0.019	-0.580	0.164	0.003	-0.116	0.129	0.144

	RLLg	LLFg	LLLg	TFg
LLFg	0.260			
LLLg	-0.645	-0.243		
TFg	0.243	0.030	0.154	
TLg	0.146	0.036	0.079	0.467

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-6.2909107294	-0.2628875461	-0.0006869513	0.2857343378	11.5444740652

Number of Observations: 2184

Number of Groups: 649

Análisis de residuales

Para saber si el modelo es adecuado se analizan los residuales normalizados del modelo estimado a partir de REML. Los residuos frente a los valores predichos para ver si hay homocedasticidad muestran cierta heterogeneidad de varianzas. El histograma de los residuos muestran normalidad. La gráfica qq-plot muestra una cierta asimetría en los datos y valores extremos (figura 4.27).

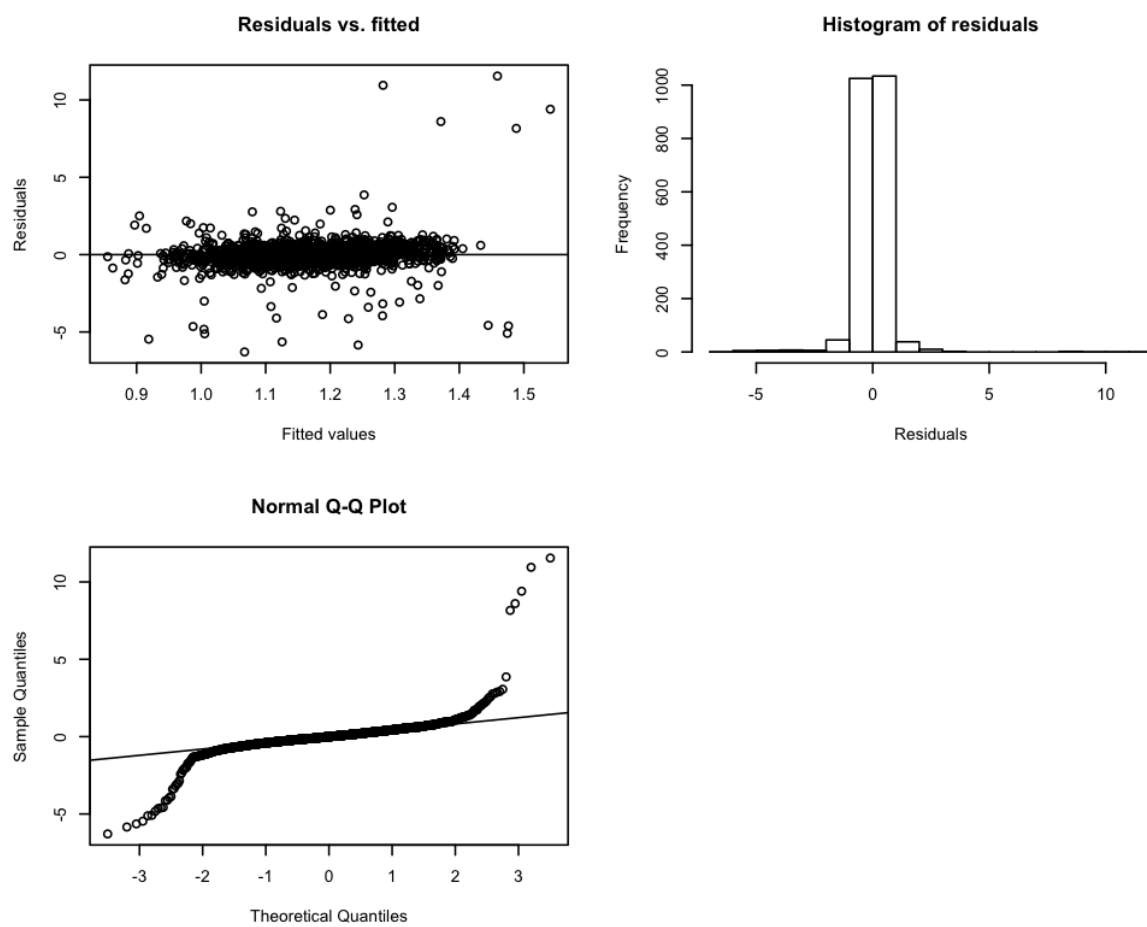


Figura 4.27: Análisis de residuales modelo longitudinal hombres

Mujeres

Para el modelo de regresión lineal de los datos longitudinales en masa ósea en mujeres se consideró de igual forma a la variable Total de masa ósea como variable de respuesta y las variables grasas y magras como covariantes.

Se dividió la base de datos en entrenamiento (80 %) para ajustar el modelo y de prueba (20 %) para probar el modelo.

El ajuste del modelo lineal muestra que hay coeficientes positivos (el peso y la talla, masa grasa en el brazo derecho, en ambas piernas y en el tronco, la masa magra en brazo y pierna izquierda y tronco), coeficientes negativos (la edad, el tiempo, masa grasa en el brazo izquierdo y masa magra en brazo derecho).

Se muestran los coeficientes de la regresión y la matriz de correlación obtenida.

Cuadro 4.44: Coeficientes fijos del modelo longitudinal óseo para mujeres

	Valor	Desviación
(Intercept)	1.093	0.142
tiempo	-0.0004	0.0002
Age	-0.003	0.0008
Height	0.025	0.089
Weight	0.0003	0.0008
RAFg	0.000007	0.00002
RALg	-0.0000007	0.00001
LAfG	-0.00002	0.00002
LALg	0.00002	0.000008
RLFg	0.0000008	0.000005
RLLg	-0.000004	0.000007
LLFg	0.000001	0.000006
LLLg	0.00001	0.000007
TFg	0.000004	0.000001
TLg	0.000001	0.000001

Cuadro 4.45: Efectos aleatorios del modelo longitudinal óseo para mujeres

	Desviación	Correlación
(Intercept)	0.10073044	
tiempo	0.00143005	-0.512
Residual	0.02949162	

La parte de los resultados que se refiere a los efectos aleatorios muestra que la varianza residual es de $\sigma^2 = 0.03^2$ y la varianza de la constante es $\sigma^2 = 0.1^2$. La varianza explicada por el factor tiempo es muy pequeña $\sigma^2 = 0.001^2$. Para la parte de los efectos fijos del modelo la constante se estima en 1.09 y el valor de los diferentes factores es prácticamente 0, excepto para el peso (0.025), que además es positivo.

El modelo se ajustó a los datos de prueba y se obtuvo el gráfico de datos originales y observados, donde se detectan algunos datos que no se ajustan al modelo, aunque la mayoría se ajustan bien.

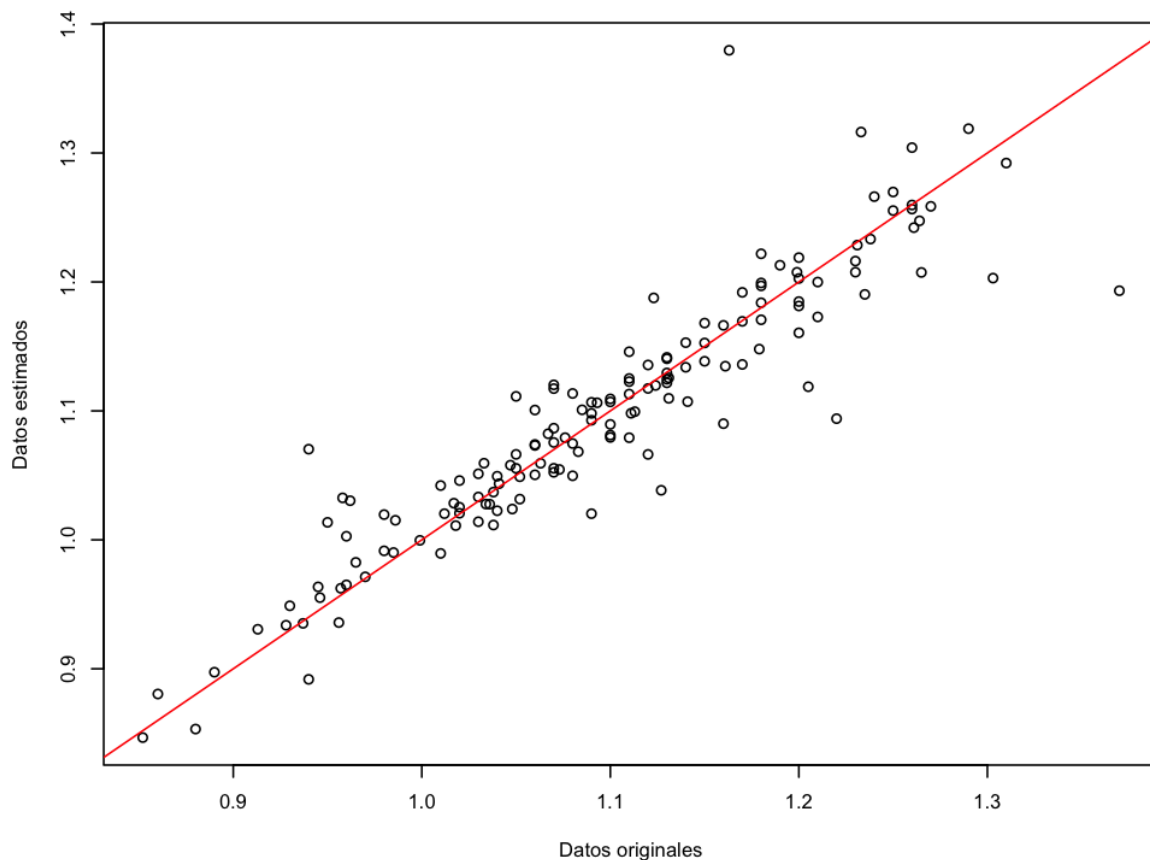


Figura 4.28: Regresión modelo longitudinal hombres

Correlation:

	(Intr)	tiempo	Age	Height	Weight	RAFg	RALg	LAFg	LALg	RLFg
tiempo	0.200									
Age	-0.311	-0.767								
Height	-0.950	-0.081	0.102							
Weight	0.019	0.040	-0.064	-0.057						
RAFg	0.050	0.042	-0.028	-0.047	-0.052					
RALg	-0.020	-0.067	0.046	-0.022	-0.262	-0.369				
LAFg	-0.043	0.001	0.001	0.058	-0.059	-0.929	0.327			
LALg	-0.005	-0.001	-0.006	0.003	-0.032	0.297	-0.388	-0.303		
RLFg	0.005	-0.017	0.045	-0.022	-0.161	0.047	0.056	-0.046	0.002	
RLLg	0.056	0.025	-0.008	-0.083	-0.173	-0.006	-0.064	0.021	-0.056	-0.066

	RLLg	LLFg	LLLg	TFg
LLFg	0.109			
LLLg	-0.622	-0.110		
TFg	0.097	-0.031	0.074	
TLg	-0.065	0.092	-0.072	0.051

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-5.39010708	-0.27379131	0.03010816	0.31454697	6.20205577

Number of Observations: 652

Number of Groups: 199

Análisis de residuales

Los residuos frente a los valores predichos muestran algo de heterocedasticidad. El histograma de los residuos muestran normalidad. La gráfica qq-plot muestra una cierta asimetría en los datos y algunos valores extremos (figura 4.29).

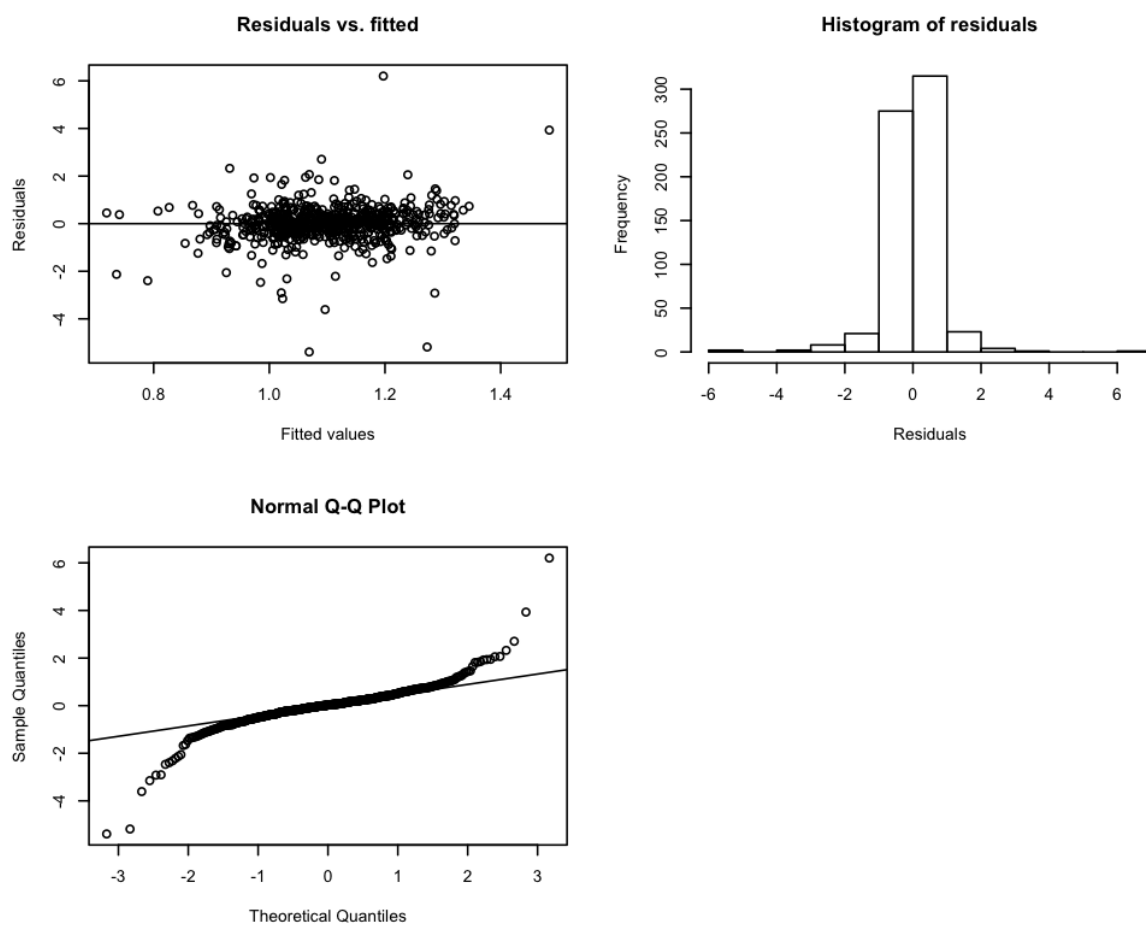


Figura 4.29: Análisis de residuales modelo longitudinal mujeres

Capítulo 5

Discusión

En este trabajo se analizaron los datos de un grupo de individuos con VIH. Se realizaron estudios transversales relativos a la masa ósea, masa grasa y masa magra de 1,426 individuos, 1,082 hombres y 344 mujeres. De igual forma se realizaron estudios longitudinales de aquellos individuos con más de una toma de datos correspondientes a 866 individuos (3,545 observaciones), 661 hombres (2,730 observaciones) y 205 mujeres (815 observaciones).

La edad promedio de los hombres es de 46 ± 10.83 años y en el caso de las mujeres de 47 ± 9.56 , lo que implica que es una población joven. El peso promedio de los hombres es de 72 ± 11.01 kg y el de las mujeres de 59 ± 11.63 kg lo que habla de una población sin sobrepeso.

El total de densidad de masa ósea (TotalBMD) es de 1.17 ± 0.1 en hombres y 1.09 ± 0.11 en mujeres. Lazcano (2009) reportó valores entre 0.97 y 1.14 para mujeres sanas (1.09 en promedio). En otros estudios (El-Maouche, Xu, Jr, Dobs & T Brown, 2011) se han reportado valores de TotalBMD para hombres en el orden de 1.02

El índice de masa corporal (BMI) promedio para hombres es de 24.17 ± 3.22 y para mujeres de 23.05 ± 4.45 , que caen en la categoría normal, según la Sociedad española para el estudio de la obesidad (Rubio *et al.*, 2007).

El índice de masa grasa (FMI) tiene en promedio un valor de 5.3 ± 2.49 para hombres y 7.5 ± 3.59 para mujeres que está dentro del promedio normal para hombres (3-6) y mujeres (5-9) (Kelly, Wilson & Heymsfield, 2009).

El índice de masa magra (FFMI) obtenido fue de 17.9 ± 1.88 kg/m² para hombres y 14.5 ± 1.67 para mujeres, siendo los valores normales de 7.26 para hombres y 5.5 para mujeres.

Se observa que los valores obtenidos en este estudio no difieren en gran medida que los de otras poblaciones, exceptuando la masa muscular.

Estos índices generales se han utilizado para diagnosticar el estado corporal de los individuos (Porbén & Espinosa, 2003) por su simplicidad de cálculo. Sin embargo, deben ser tomados con cautela para estudios más finos, dado que están influidos por el sexo, la raza y la edad lo que conduce a clasificaciones erróneas (Alvero-Cruz *et al.*, 2010). Los datos obtenidos en este estudio muestran una correlación alta entre el índice de masa corporal y el índice de masa grasa (0.8 para hombres y 0.9 para mujeres), lo que resulta lógico ya que la grasa es un componente importante del peso.

En el caso de los hombres se encontró que el peso es un factor importante en la relación con la masa ósea. En los hombres la densidad mineral ósea total disminuye en -0.16 con respecto al peso manteniendo los otros factores constantes. En las mujeres la relación es inversa. La densidad mineral ósea aumenta en 0.11 con respecto al peso. Pero además en las mujeres la densidad mineral ósea disminuye en -0.0005 al cumplir un año más, manteniendo los otros factores constantes.

Al hacer el estudio del estado óseo según los criterios de la OMS, se encontró que cualquier medida de variable T puede ser utilizada para caracterizar el estado óseo, pero el triángulo de Ward es el índice con menos sobreposición de datos. En los hombres el 46 % presenta osteopenia y el 19 % osteoporosis (65 % presentan una condición anormal del estado óseo). En las mujeres, 41 % presentan osteopenia y 19 % osteoporosis (60 % de condición ósea anormal). En un estudio en pacientes con VIH, Landazurri (2014) encontraron 50 % de condición ósea anormal para hombres y 28 % para mujeres, aunque estos valores están relacionados con el consumo de tratamientos antiretrovirales.

Diversos estudios han mostrado una correlación positiva y significativa entre la masa ósea con el peso, la masa muscular y la masa grasa, encontrando la mejor correlación entre masa ósea y masa grasa, resaltando el hecho de que las personas con sobre peso y grasa tienen menor probabilidad de sufrir osteoporosis (Oria, 2003).

Las investigaciones sobre la influencia del tejido magro y del tejido graso en la pérdida de masa ósea ha arrojado que los pacientes con osteoporosis tienen una cantidad menor de ambos tejidos. Para Zanchetta (2001) la masa magra es el principal determinante del contenido de masa ósea del esqueleto entero ya que cita estudios donde hay una constante en la relación contenido de masa ósea del esqueleto / masa magra. Esta relación es constante en niños, hombres y mujeres postmenopáusicas. En mujeres entre la menarca y la menopausia explica que los estrógenos determinan una acumulación mayor de calcio para una masa magra similar.

La osteoporosis también se asocia con la pérdida de la función muscular y la limitación de la movilidad, aunque no se sabe bien cómo se comporta esta asociación. Las muestras de biopsia de pacientes con osteoporosis muestran atrofia de las fibras musculares de tipo II. El grado de atrofia de las fibras es proporcional al grado de pérdida de la densidad mineral ósea (Rastogi, Corriere & Ferrucci, 2014).

Sayers (2010) afirma que la masa magra es uno de los determinantes más fuertes de la masa ósea a lo largo de la vida. Gregson (2013) sugiere que hay una interacción fuerte entre la masa grasa y la masa ósea, por la capacidad de la grasa de acumular hormonas y otros esteroideos, en particular los estrógenos, que benefician la formación de hueso.

Zerbi (2017) enfatiza el hecho de que las células musculares y las óseas tienen un origen común: las células-tronco mesenquimales. Afirma que son susceptibles a influencias similares genéticas y epigenéticas. Se desarrollan paralelamente durante la adolescencia y declinan paralelamente durante la vejez. Es tal la relación de estas células que acuña un término nuevo: la Osteosarcopenia, que integra los mecanismos de pérdida y factores de riesgo similares en los procesos de osteoporosis y sarcopenia.

En cuanto a la masa grasa, el 96.5 % de los hombres presenta la condición de lipodistrofia, según los puntos de corte definidos en la literatura. En las mujeres, el 99.7 % presenta esta condición. Otros estudios han arrojado porcentajes del orden del 28 % con lipodistrofia (Doradea & Rodríguez, 2017) y hasta el 42 % en pacientes con VIH (Valle Gutiérrez &

Gómez Rodrigo, 2001).

Los hombres presentan la condición de sarcopenia en el 11 % de los individuos y en el 0.6 % de las mujeres, según los puntos de corte definidos en el capítulo 2. En la literatura se reporta que la sarcopenia está fuertemente ligada a la edad. Nemerovsky (2016) reporta valores de hasta el 70 % en individuos mayores de 80 años. Baumgartner (1998) reporta valores de sarcopenia dependientes de la edad de 17 % y 24 % para hombres y mujeres menores de 70 años y hasta 57 % en hombres y 60 % en mujeres mayores de 80 años.

Se modeló la relación entre la densidad de masa ósea total y las medidas de masa grasa y masa magra. El modelo obtenido considera el peso, la talla, la masa grasa y la masa magra en diferentes partes del cuerpo, además de la edad para mujeres. Este modelo explica el 33 % de la variabilidad para hombres y 37 % para mujeres.

En el modelo para masa grasa la variable de respuesta fue el total de masa grasa (TotalFg) y las covariables la masa ósea y la masa magra. El modelo para hombres explica el 97.48 % de la variabilidad y el modelo para mujeres el 97.5 %. Se encontró una relación 1:1 entre la masa grasa y el peso tanto para hombres y mujeres. Por el signo de los coeficientes de las regresiones, existe una relación inversa entre las variables óseas y magras con respecto a la masa grasa. Este comportamiento ya había sido descrito por Rodríguez (2013) y lo atribuye a la colinealidad que existe entre el peso y la masa grasa.

Para el modelo de masa magra la variable de respuesta elegida fue el total de masa magra (TotalLg) y las covariables la masa ósea y la masa grasa. El modelo para hombres explica el 98 % de la variabilidad y el 99 % para mujeres. Se encontró una relación entre la masa magra y el peso de los individuos, siendo esta en relación positiva 1:1, es decir al aumentar o disminuir uno, el otro aumenta o disminuye en la misma proporción.

La pérdida de masa magra relacionada con la edad está típicamente compensada por la ganancia de masa grasa. Después de los 70 años, la masa magra y la masa grasa tienden a disminuir en forma paralela (Rastogi *et al.*, 2014).

Lo que queda claro con cualquiera de las regresiones mostradas es que introducir el peso en los análisis, como covariable, es un determinante de sesgo, por la colinealidad que presenta tanto con la masa grasa como con la magra (T. Rodríguez, 2013).

En cuanto al modelo de clasificación se eligió el algoritmo kNN (k-nearest neighbors), por ser simple y de fácil interpretación. Es necesario aclarar que también se ensayaron otros modelos (clasificador bayesiano, árboles de decisión y random forest) pero al final el método kNN dió la mayor precisión.

Para el modelo óseo la matriz de confusión obtenida para hombres dió un valor de precisión de 60 % (porcentaje clasificación exitosa) para hombres y de 62 % para mujeres.

En el modelo de masa grasa, los valores de precisión fueron del 96 % para hombres y ninguno para mujeres, debido quizá al tamaño de muestra del conjunto de entrenamiento (que es menor que número de datos con la condición de lipodistrofia).

En el modelo de masa magra la precisión fue del 92 % para hombres y sin valor para mujeres.

La elección del valor de k se puede hacer mediante el método de curvas ROC, pero para este trabajo se ensayaron varios valores y al final se eligió la que obtuvo la mayor precisión.

Los modelos de clasificación utilizados fueron aceptables en los tres casos donde se obtuvieron resultados. Es necesario recordar que la clasificación de facto utilizada proviene de los datos mismos (variable T, FMI y FFMI).

Para el modelo de masa ósea con datos longitudinales se filtraron los datos para aquellos individuos con más de una DEXA. El número de DEXAs por individuo no es constante, lo que lleva a utilizar modelos que consideren tanto la variable del tiempo como el hecho de que los individuos tienen medidas repetidas y desiguales. Arnau (2007) considera que el estudio de medidas repetidas desde la perspectiva temporal se están volviendo populares aunque requiere técnicas de análisis avanzados y recursos de cómputo.

Dado el análisis de residuales de los modelos propuestos (para hombres y mujeres) se podría aplicar una transformación, intentar averiguar si el incremento de la varianza para determinados valores es debido a alguna covariable, utilizar otro tipo de modelos como los modelos lineales generalizados mixtos con una distribución de errores diferente (gama, logarítmica) (Cayuela, 2010) o utilizar alguna técnica no paramétrica como los árboles de regresión o, simplemente, quedarse con estos modelos a pesar de sus limitaciones (Corra Morales & Salazar, 2016).

Algunas de las variables de respuesta no se comportan totalmente de forma normal. El teorema del límite central requiere que la población o poblaciones de las que proceden las muestras tengan una distribución normal, no que la tengan las muestras. Si las muestras se distribuyen de forma normal, se puede aceptar que así lo hacen las poblaciones de origen. En el caso de que las muestras no se distribuyan de forma normal pero se tenga certeza de que las poblaciones de origen sí lo hacen, entonces, los resultados obtenidos por los contrastes paramétricos sí son válidos. El teorema del límite central, permite reducir los requerimientos de normalidad cuando las muestras suficientemente grandes.

Capítulo 6

Conclusiones

Este trabajo presenta un estudio que permite en base a dos aproximaciones (predicción y clasificación) poder abordar problemas de modelaje, donde se tiene modelos aprendidos con datos de entrenamiento y que se aplican después a datos de validación. Con esta metodología se ha podido determinar un marco general de modelización que sigue el siguiente esquema:

$$\textit{Respuesta} = \textit{Modelo} + \textit{incertidumbre}$$

El modelo de regresión transversal para variables óseas, grasas y magras permite una reducción de dimensión, obteniendo un listado de aquellas variables que podrían influir en las variables de respuesta.

Los modelos óseos descritos en este trabajo tienen un r^2 bajo. La distribución ósea obtenida tiene porcentajes por debajo de lo reportado en la literatura.

La distribución grasa presenta porcentajes muy por debajo de lo reportado en la literatura.

La sarcopenia es una entidad de alta prevalencia en las personas mayores. Desde los 50 años de edad, la masa muscular disminuye a un ritmo de 12 % a 15 % por década. La literatura reporta que la sarcopenia está presente en un 20 % en adultos de 70 a 75 años y del 50 % en los mayores de 80 años en los varones y del 25 % al 40 % en las mujeres en los mismos rangos de edad. El resultado obtenido en este trabajo supera con mucho estos porcentajes, aunque no se determinó si es debido a los puntos de corte o a la presencia de la infección por VIH.

Al no existir mayor información como tiempo de infección, administración de tratamientos, ingesta de suplementos o comorbilidades (presencia de uno o más trastornos y el efecto de estos trastornos en la infección por VIH) es imposible hacer una caracterización óptima de los pacientes que permita hacer mejores inferencias, por lo que el estudio se queda solo al nivel descriptivo.

En el apartado de clasificación los patrones obtenidos con el algoritmo kNN permitieron realizar predicciones con un nivel de error menor a 10 % en la mayoría de los casos. Aun cuando el error pudiera considerarse bajo, el modelo no puede aplicarse cuando la incidencia de la condición de estudio presenta un porcentaje bajo.

Sin embargo, los resultados obtenidos son prometedores ya que muestran resultados significativos. Es importante el desarrollo de análisis futuros que por su planeación permitan caracterizar mejor las relaciones entre masa grasa, masa magra y masa ósea en el contexto de un grupo vulnerable que tiende a un envejecimiento progresivo.

Apéndice A

Evaluación del plan de trabajo

A continuación se hará una evaluación del plan de trabajo presentado en el PEC 1 entregada el 15 de octubre de 2018. En esa entrega se plantearon los siguientes objetivos:

Objetivos generales y específicos

1. Caracterizar el estado de salud de los pacientes en cuanto a las variables contenidas en la DEXA
 - 1.1 Describir las características de los pacientes incluidos en el estudio por sexo
2. Identificar mediante métodos estadísticos las variables asociadas a la osteopenia, lipodistrofia y sarcopenia en pacientes infectados por el VIH
 - 2.1 Analizar las variables que describen la osteopenia y osteoporosis
 - 2.2 Analizar las variables que describen la sarcopenia
 - 2.3 Analizar los variables que describen la lipodistrofia
3. Describir la evolución en el tiempo de osteopenia en aquellos pacientes en los que se disponga de la información necesaria
 - 3.1 Describir la evolución en el tiempo de la densidad mineral ósea

Aunque en el momento del PEC 1 se pensó que el trabajo era ambicioso, se cumplieron con los objetivos a nivel de análisis exploratorio. Definitivamente puede hacerse un análisis detallado de estos objetivos pero este trabajo es un buen inicio.

Enfoque y método a seguir

En cuanto al análisis estadístico se consideró utilizar algoritmos de aprendizaje supervisado, aunque no se puntualizó cuáles. Originalmente se pensó en emplear algoritmos de clasificación y regresión.

- **Clasificación.** Cada registro de la base de datos pertenece a una clase, la cual se indica mediante el valor de un atributo el cual llamamos la clase de la instancia. Este atributo puede tomar diferentes valores discretos. El objetivo será predecir la clase de nuevas instancias a partir del resto de atributos. Se ensayaron los algoritmos:
 - Método k vecinos más próximos
 - Método clasificador bayesiano
 - Método Árboles de decisión

- Método Ranforest

Al final se analizaron los datos con el método k vecinos más próximos, pero en el apéndice 2 están el resto de los análisis.

- **Regresión.** Describir una función real que asigne a cada instancia un valor real. La principal diferencia respecto a la clasificación es que el valor a predecir es numérico.

Se utilizó regresión lineal simple con todas las variables disponibles para cada variable de respuesta. Se utilizaron los métodos Stepwise y boruta para reducir variables. Finalmente se dejó solo el primero como método para reducción de variables. Se entrenaron los modelos con un conjunto de entrenamiento y se probó con un conjunto de prueba, que son los resultados que se muestran en el trabajo.

También se realizó una regresión longitudinal mediante modelos lineales mixtos para medidas repetidas no balanceadas, por ser el método que mejor se ajusta al tipo de datos.

Las tareas que se propusieron para desarrollar el trabajo en el PEC 1 fueron:

- Verificación y depuración de la base de datos
- Crear base de datos para hombres y para mujeres
- Descripción de los pacientes por género
- Definir puntos de corte para osteoporosis, osteopenia y normal
- Definir puntos de corte para lipodistofia
- Definir puntos de corte para sarcopenia
- Definir puntos de corte para el índice de masa corporal (IMC)
- Describir a los pacientes por género y estado de masa ósea
- Describir a los pacientes por género y estado de masa grasa
- Describir a los pacientes por género y estado de masa magra
- Definir las variables que mejor permitan clasificar los estados de masa osea, masa grasa y masa grasa
- Describir el efecto en el tiempo de las medias DEXA para osteopenia, osteoporosis, normal

Se cumplieron con todas ellas y con el calendario que se propuso para su cumplimiento:

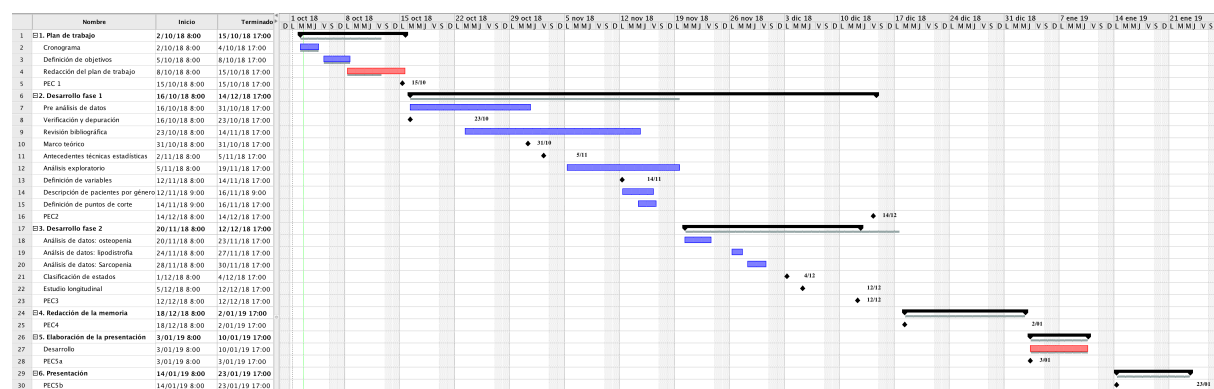


Figura A.1: Calendario

A pesar de que el tiempo era el único factor limitante reconocido en el plan de trabajo se cumplieron con todos los objetivos planteados. Al cumplirse en tiempo las tareas se cubrieron los objetivos del trabajo.



Figura A.2: Hitos

Se realizó un estudio al detalle para la condición de normal / osteoporosis, sin embargo la planeación arrojó que se podría hacer un análisis de lipodistrofia y sarcopenia, considerando que la preparación de la base de datos y puntos de corte afectaba las tres condiciones y que las técnicas estadísticas que se utilizaron eran las mismas en todos los casos.

De esta forma el resultado final es una primera aproximación al problema, que considero puede servir como base para trabajos más detallados y especializados. En lo personal, me parece que los datos longitudinales podrían ser interesantes para un análisis a fondo.

Apéndice B

Código R

```
# Librerías utilizadas
library(haven)
library(readr)
library(readxl)
library(dplyr)
library(ggplot2)
library(GGally)
library(caret)
library(ellipse)
library(sqldf)
library(rpart)
library(rpart.plot)
library(Hmisc)
library(party)
library(pastecs)
library(pander)
library(MASS)
library(DAAG)
library(Boruta)
library(EnvStats)
library(ggpubr)
library(class)
library(gmodels)
library(e1071)
library(C50)
library(randomForest)
library(lubridate)
library(nlme)
library(nortest)
library(tseries)

# Importación de datos
setwd("~/Dropbox/TFM/Final 1.0/fuentes/")
dexa <- read_sav("Totes DEXES completes_09-01-18_English_selection.sav")
```

```

attach(dexa)
ID = toString(ID)

variables <- read_csv("variable names-Tabla 1.csv",
  locale = locale(encoding = "UTF8"))

frecuencia_dexa <- read_excel("summary_content_dataset.xlsx")

# Se eliminan las columnas que no se utilizarán.
dexa <- dexa[,-3] # age_previous
dexa <- dexa[,-6] # id_center
dexa <- dexa[,-3] # dato duplicado
dexa <- dexa[,-62] # comentarios no considerados
dexal <- dexa
dexal$sexo[dexal$gender_num == 1] <- "hombre"
dexal$sexo[dexal$gender_num == 2] <- "mujer"
dexal$sexo <- as.factor(dexal$sexo)

# Para análisis transversal.
# Se eliminan los ID repetidos y se deja solo el último análisis

dexa$fecha=as.character(dexa$DateAnalysis) # Solo para hacer el filtro
dexa <- dexa %>%
  group_by(ID) %>%
  top_n(n = 1, wt=fecha)

dexa <- dexa[,-77] # eliminamos la columna agregada

dexa$sexo[dexa$gender_num == 1] <- "hombre"
dexa$sexo[dexa$gender_num == 2] <- "mujer"
dexa$sexo <- as.factor(dexa$sexo)

# glimpse(dexa) # Estructura de la base de datos.

dexa<-dexa[!duplicated(dexa$ID),] # Se eliminan duplicados con la misma fecha de análisis
dexa <- na.omit(dexa) # Se eliminan todos los datos que tengan NA's
sapply(dexa, function(x) sum(is.na(x))) # La base de datos no contiene NA

dexa_h <- dexa[ which(dexa$gender_num==1),] # Base de datos de hombres
dexa_m <- dexa[ which(dexa$gender_num==2),] # Base de datos de mujeres

mydata <- aggregate(dexa[c("TotalBMD", "BMI", "FMI", "FFMI")],
  by=list(Sexo=dexa$gender_num), mean, na.rm=TRUE)
a<-dim(dexa_h)
b<-dim(dexa_m)
n<-c(a[1], b[1])
mydata <- data.frame(n,mydata)
Sexo <- c('Masculino', 'Femenino')
mydata <- data.frame(Sexo,mydata)

```

```

mydata <- mydata[,-3]
pandoc.table( mydata
               , caption = "Resumen estadístico general de los datos del estudio por s
               , justify = 'lrrrrr'
               , round = 3
               , decimal.mark = "."
               , big.mark = ","
               , emphasize.rownames = TRUE
               , split.cells = 5
               )

```

```

a<-dim(dexa_h)
estad <- stat.desc(dexa_h)
d <- c('n', 'Edad', 'Height (m)', 'Weightg (kg)', 'TotalBMD', 'BMI ', 'FMI', 'FFMI')
e <- c(a[1], estad[9,76], estad[9,6], estad[9,62], estad[9,5], estad[9,63],
      estad[9,64], estad[9,65])
f <- c('', round(estad[13,76],2), round(estad[13,6],2), round(estad[13,62],2),
      round(estad[13,5],2), round(estad[13,63],2), round(estad[13,64],2),
      round(estad[13,65],2))

g <- c('', round(estad[4,76],2), round(estad[4,6],2), round(estad[4,62],2),
      round(estad[4,5],2), round(estad[4,63],2), round(estad[4,64],2),
      round(estad[4,65],2))

h <- c('', round(estad[5,76],2), round(estad[5,6],2), round(estad[5,62],2),
      round(estad[5,5],2), round(estad[5,63],2), round(estad[5,64],2),
      round(estad[5,65],2))

```

```

mydata <- data.frame(d,e,f,g,h)
names(mydata) <- c("", "Media", "D. E.", "Minimo", "Maximo")

```

```

pandoc.table( mydata
               , caption = "Resumen estadístico de los datos del estudio\n(Masculino)"
               , justify = 'lrrrr'
               , round = 3
               , decimal.mark = "."
               , big.mark = ","
               , emphasize.rownames = TRUE
               , split.cells = 5
               )

```

```

a<-dim(dexa_m)
estad <- stat.desc(dexa_m)
d <- c('n', 'Edad', 'Height (m)', 'Weightg (kg)', 'TotalBMD', 'BMI ', 'FMI', 'FFMI')
e <- c(a[1], estad[9,76], estad[9,6], estad[9,62], estad[9,5], estad[9,63],
      estad[9,64], estad[9,65])
f <- c('', round(estad[13,76],2), round(estad[13,6],2), round(estad[13,62],2),
      round(estad[13,5],2), round(estad[13,63],2), round(estad[13,64],2),

```

```

round(estad[13,65],2))

g <- c('', round(estad[4,76],2), round(estad[4,6],2), round(estad[4,62],2),
round(estad[4,5],2), round(estad[4,63],2), round(estad[4,64],2),
round(estad[4,65],2))

h <- c('', round(estad[5,76],2), round(estad[5,6],2), round(estad[5,62],2),
round(estad[5,5],2), round(estad[5,63],2), round(estad[5,64],2),
round(estad[5,65],2))

mydata <- data.frame(d,e,f,g,h)
names(mydata) <- c("", "Media", "D. E.", "Minimo", "Maximo")

pandoc.table( mydata
, caption = "Resumen estadístico de los datos del estudio\n(Femenino)"
, justify = 'lrrrr'
, round = 3
, decimal.mark = "."
, big.mark = ","
, emphasize.rownames = TRUE
, split.cells = 5
)

## Pruebas de t para las variables
t.test(dexa_h$TotalBMD, dexa_m$TotalBMD)

t.test(dexa_h$FMI, dexa_m$FMI)

t.test(dexa_h$FFMI, dexa_m$FFMI)

dexa_h_variables <- sqldf("select TotalBMD, BMI, FMI, FFMI from dexa_h")

png("~/Dropbox/TFM/Final 1.0/figurasR/fig1.png",
width=1000,height=800,units="px",
pointsize=8,bg="white",res=150)

ggpairs(data=dexa_h_variables,columns=1:4)

dexa_m_variables <- sqldf("select TotalBMD, BMI, FMI, FFMI from dexa_m")

png("~/Dropbox/TFM/Final 1.0/figurasR/fig2.png",
width=1000,height=800,units="px",
pointsize=8,bg="white",res=150)

ggpairs(data=dexa_m_variables,columns=1:4)

#### TotalBMD
## Prueba de Kolmogorov-Smirnov
ks.test(x = dexa_h$TotalBMD,"pnorm", mean(dexa_h$TotalBMD), sd(dexa_h$TotalBMD))

```

```

ks.test(x = dexa_m$TotalBMD,"pnorm", mean(dexa_m$TotalBMD), sd(dexa_m$TotalBMD))

## Modificación de Lillefors
lillie.test(x = dexa_h$TotalBMD)
lillie.test(x = dexa_m$TotalBMD)

## Test de normalidad de Jarque-Bera
jarque.bera.test(x = dexa_h$TotalBMD)
jarque.bera.test(x = dexa_m$TotalBMD)

#### BMI
## Prueba de Kolmogorov-Smirnov
ks.test(x = dexa_h$BMI,"pnorm", mean(dexa_h$BMI), sd(dexa_h$BMI))
ks.test(x = dexa_m$BMI,"pnorm", mean(dexa_m$BMI), sd(dexa_m$BMI))

## Modificación de Lillefors
lillie.test(x = dexa_h$BMI)
lillie.test(x = dexa_m$BMI)

## Test de normalidad de Jarque-Bera
jarque.bera.test(x = dexa_h$BMI)
jarque.bera.test(x = dexa_m$BMI)

#### TotalFg
## Prueba de Kolmogorov-Smirnov
ks.test(x = dexa_h$TotalFg,"pnorm", mean(dexa_h$TotalFg), sd(dexa_h$TotalFg))
ks.test(x = dexa_m$TotalFg,"pnorm", mean(dexa_m$TotalFg), sd(dexa_m$TotalFg))

## Modificación de Lillefors
lillie.test(x = dexa_h$TotalFg)
lillie.test(x = dexa_m$TotalFg)

## Test de normalidad de Jarque-Bera
jarque.bera.test(x = dexa_h$TotalFg)
jarque.bera.test(x = dexa_m$TotalFg)

#### TotalLg
## Prueba de Kolmogorov-Smirnov
ks.test(x = dexa_h$TotalLg,"pnorm", mean(dexa_h$TotalLg), sd(dexa_h$TotalLg))
ks.test(x = dexa_m$TotalLg,"pnorm", mean(dexa_m$TotalLg), sd(dexa_m$TotalLg))

## Modificación de Lillefors
lillie.test(x = dexa_h$TotalLg)
lillie.test(x = dexa_m$TotalLg)

## Test de normalidad de Jarque-Bera
jarque.bera.test(x = dexa_h$TotalLg)
jarque.bera.test(x = dexa_m$TotalLg)

```

```

dexa_h_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                        L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                        TrochBMD, TotalFBMD, BMI from dexa_h")
correlaciones <- cor(dexa_h_oseo[,4:15])
pandoc.table( correlaciones
              , caption = "Matriz de correlaciones para variables oseas (masculino)"
              , justify = 'lrrrrrrrrrrrrr'
              , round = 3
              , decimal.mark = "."
              , big.mark = ","
              , emphasize.rownames = TRUE
              , split.cells = 5
              )

dexa_m_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                        L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                        TrochBMD, TotalFBMD, BMI from dexa_m")
correlaciones <- cor(dexa_m_oseo[,4:15])
pandoc.table( correlaciones
              , caption = "Matriz de correlaciones para variables oseas (femenino)"
              , justify = 'lrrrrrrrrrrrrr'
              , round = 3
              , decimal.mark = "."
              , big.mark = ","
              , emphasize.rownames = TRUE
              , split.cells = 5
              )

png("~/Dropbox/TFM/Final 1.0/figurasR/fig3.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

ggpairs(data=dexa_h_oseo,columns=4:15)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig4.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

ggpairs(data=dexa_m_oseo,columns=4:15)

dexa_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                        L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                        TrochBMD, TotalFBMD, BMI from dexa_h")
fit1 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit1)

#response <- dexa_oseo[, "TotalBMD"]
#boruta_output <- Boruta(response ~ ., data = dexa_oseo, doTrace=2)
#boruta_signif <- names(boruta_output$finalDecision[boruta_output$finalDecision %in%

```

```
#           c("Confirmed", "Tentative"]])
#boruta_output
```

```
dexa_oseo <- sqldf("select Height, Weightg, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_h")
fit2 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit2)
```

```
# Stepwise Regression
```

```
dexa_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_h")
fit <- lm(TotalBMD ~ ., data=dexa_oseo)
step <- stepAIC(fit, direction="both")
step$anova
```

```
dexa_oseo <- sqldf("select Height, TotalBMD, L1BMD, L2BMD, L4BMD,
                    WardsBMD, TrochBMD, TotalFBMD, BMI  from dexa_h")
fit3 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit3)
```

Comparar los tres modelos con ANOVA y elegir el mejor.

```
anova(fit1, fit2, fit3)
```

```
dexa_oseo$prediction <- predict(fit3,newdata=dexa_oseo)
summary(fit3)
coefficients(fit3)
anova(fit3)
layout(matrix(c(1,2,3,4),2,2))
#plot(fit3)
```

```
dexa_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_m")
fit1 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit1)
```

```
dexa_oseo <- sqldf("select Weightg, Age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_m")
fit2 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit2)
```

```
# Stepwise Regression
```

```
dexa_oseo <- sqldf("select Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_m")
fit <- lm(TotalBMD ~ ., data=dexa_oseo)
```

```

step <- stepAIC(fit, direction="both")
step$anova

dexa_oseo <- sqldf("select Height, age, TotalBMD, L1BMD, L2L4BMD,
                    WardsBMD, TotalFBMD, BMI from dexa_m")
fit3 <- lm(TotalBMD ~ ., data=dexa_oseo)
summary(fit3)

anova(fit1, fit2, fit3)

dexa_oseo$prediction <- predict(fit3,newdata=dexa_oseo)
summary(fit3)
coefficients(fit3)
anova(fit3)
layout(matrix(c(1,2,3,4),2,2))
#plot(fit3)

# Modelo predictivo
# Masculino

dexa_oseo <- sqldf("select Height, TotalBMD, L1BMD, L2BMD, L4BMD,
                    WardsBMD, TrochBMD, TotalFBMD, BMI from dexa_h")

# Datos de entrenamiento y validación
set.seed(123)
split.indices <- sample(nrow(dexa_oseo), nrow(dexa_oseo)*0.8, replace = F)
train <- dexa_oseo[split.indices, ]
test <- dexa_oseo[-split.indices, ]

fit <- lm(TotalBMD ~ ., data=train)
summary(fit)
anova(fit)
pred = predict(fit, test)
rss = sum((train$TotalBMD - pred)^2)
rss
#plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
#      sub="TotalBMD= Height + L1BMD + L2BMD + L4BMD + WardsBMD + TrochBMD + TotalFBMD",
#      cex.sub=0.5)

# Femenino

dexa_oseo <- sqldf("select Height, age, TotalBMD, L1BMD, L2L4BMD,
                    WardsBMD, TotalFBMD, BMI from dexa_m")

#Dividing data in train & test samples

set.seed(123)
split.indices <- sample(nrow(dexa_oseo), nrow(dexa_oseo)*0.8, replace = F)
train <- dexa_oseo[split.indices, ]

```

```

test <- dexa_oseo[-split.indices, ]

fit <- lm(TotalBMD ~ ., data=train)
summary(fit)
anova(fit)
pred = predict(fit, test)
rss = sum((train$TotalBMD - pred)^2)
rss
#plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
#      sub="TotalBMD= Height + Age + L1BMD + L2L4BMD + WardsBMD + TotalFBMD + BMI",
#      cex.sub=0.5)

# Análisis basado en marca T
# Masculino

png("~/Dropbox/TFM/Final 1.0/figurasR/fig5.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

dexa_T <- sqldf("select L1T, L2T, L3T, L4T, L1L4T, L2L4T,
                      NeckFT, WardsT, TrochT, TotalFT from dexa_h")
summary(dexa_T)

valor <- c(dexa_T$L1T, dexa_T$L2T, dexa_T$L3T,
           dexa_T$L4T, dexa_T$L1L4T, dexa_T$L2L4T, dexa_T$NeckFT,
           dexa_T$WardsT, dexa_T$TrochT, dexa_T$TotalFT)

variable <- rep(1:10, each = 1082)

variable <- factor(variable, labels = c("L1T", "L2T", "L3T", "L4T",
                                       "L1L4T", "L2L4T", "NeckFT", "WardsT",
                                       "TrochT", "TotalFT"))

variable <- gl(10,1082, labels = c("L1T", "L2T", "L3T", "L4T", "L1L4T",
                                   "L2L4T", "NeckFT", "WardsT",
                                   "TrochT", "TotalFT"))

plot(valor ~ variable)

p.aov <- aov(valor ~ variable)
summary(p.aov)
model.tables(p.aov, type = "mean")

# ggpairs(data=dexa_T,columns=1:10)

# Femenino

png("~/Dropbox/TFM/Final 1.0/figurasR/fig6.png",
    width=1000,height=800,units="px",

```

```

    pointsize=8,bg="white",res=150)

dexa_T <- sqldf("select L1T, L2T, L3T, L4T, L1L4T, L2L4T,
                    NeckFT, WardsT, TrochT, TotalFT from dexa_m")
summary(dexa_T)

valor <- c(dexa_T$L1T, dexa_T$L2T, dexa_T$L3T,
           dexa_T$L4T, dexa_T$L1L4T, dexa_T$L2L4T, dexa_T$NeckFT,
           dexa_T$WardsT, dexa_T$TrochT, dexa_T$TotalFT)

variable <- rep(1:10, each = 344)

variable <- factor(variable, labels = c("L1T", "L2T", "L3T", "L4T",
                                       "L1L4T", "L2L4T", "NeckFT", "WardsT",
                                       "TrochT", "TotalFT"))
variable <- gl(10,344, labels = c("L1T", "L2T", "L3T", "L4T", "L1L4T",
                                  "L2L4T", "NeckFT", "WardsT",
                                  "TrochT", "TotalFT"))

plot(valor ~ variable)

p.aov <- aov(valor ~ variable)
summary(p.aov)
model.tables(p.aov, type = "mean")

#ggpairs(data=dexa_T,columns=1:10)

dexa$oseo[dexa$WardsT > -1] <- "Normal"
dexa$oseo[dexa$WardsT <= -1 & dexa$WardsT > -2.5] <- "Osteopenia"
dexa$oseo[dexa$WardsT <= -2.5] <- "Osteoporosis"
dexa$oseo <- as.factor(dexa$oseo)

dexa_h <- dexa[ which(dexa$gender_num==1),]
dexa_m <- dexa[ which(dexa$gender_num==2),]
dexa_oseo <- sqldf("select oseo, Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI from dexa_h")

#dexa_h%>%
#   ggplot(aes(x=dexa_oseo$TotalBMD, fill=dexa_oseo$oseo))+
#   geom_density(alpha=0.5) + xlab("Masa Osea Total") + ylab ("Densidad")

png("~/Dropbox/TFM/Final 1.0/figurasR/fig7.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

dexa_h%>%
  ggplot(aes(x=dexa_oseo$WardsBMD, fill=dexa_oseo$oseo))+
  geom_density(alpha=0.5) + xlab("Triangulo de Ward") + ylab ("Densidad")

```

```

dexa_oseo <- sqldf("select oseo, Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_m")

png("~/Dropbox/TFM/Final 1.0/figurasR/fig8.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

dexa_m%>%
  ggplot(aes(x=dexa_oseo$WardsBMD, fill=dexa_oseo$oseo))+
  geom_density(alpha=0.5) + xlab("Triangulo de Ward") + ylab ("Densidad")

# Modelo de clasificación
# Masculino

dexa_oseo <- sqldf("select oseo, Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI  from dexa_h")

table(dexa_oseo$oseo)

round(prop.table(table(dexa_oseo$oseo)) * 100, digits = 1)
lbls <- paste(names(table(dexa_oseo$oseo)),
              round(prop.table(table(dexa_oseo$oseo)) * 100,
                    digits = 1), "%")

png("~/Dropbox/TFM/Final 1.0/figurasR/fig9.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

pie(table(dexa_oseo$oseo), labels = lbls, col = c("Blue", "Red", "Yellow"))

# Modelo de clasificación
# Masculino

dexa_oseo<-sqldf("select  oseo, TotalBMD,
                    Height, Weight, Age,
                    RAFg, RALg, LAFg, LALg,
                    RLFg, RLLg, LLFg, LLLg,
                    TFG, TLg
                    from dexa_h")

# Método k vecinos más próximos

set.seed(123)
split.indices <- sample(nrow(dexa_oseo), nrow(dexa_oseo)*0.8, replace = F)

train <- dexa_oseo[split.indices, -1]

```

```

test <- dexa_oseo[-split.indices, -1]

train_labels <- dexa_oseo[split.indices, 1]
test_labels <- dexa_oseo[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
#n <- as.data.frame(lapply(dexa_oseo[2:16], normalize))
n <- as.data.frame(lapply(dexa_oseo[2:15], normalize))

summary(n$TotalBMD)
n_train <- n[split.indices, ]
n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test,
                 cl = train_labels, k = 21)
CrossTable(x = test_labels, y = test_pred,
           prop.chisq = FALSE)

#accuracy_knn = (61+87+24)/(61+87+24+14+8+6+17)
accuracy_knn = (39+84+9)/(39+84+9+35+1+12+5+2+30)
accuracy_knn

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 1)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (44+55+15)/(44+55+15+29+2+31+15+4+22)
accuracy_knn

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 5)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (45+62+13)/(45+62+13+29+1+29+10+4+24)
accuracy_knn

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 10)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (42+70+14)/(42+70+14+33+24+7+1+26)
accuracy_knn

# Método clasificador bayesiano

classifier <- naiveBayes(train, train_labels)
eval_pred <- predict(classifier, test)
head(eval_pred)
CrossTable(eval_pred, test_labels, prop.chisq = FALSE, prop.t = FALSE,
           prop.r = FALSE, dnn = c('estimado', 'observado'))

```

```
accuracy = (38+59+23)/(38+59+23+20+1+34+17+3+22)
accuracy
```

```
# Método Arboles de decisión
```

```
tree <- C5.0(train, train_labels)
tree
summary(tree)
tree_pred <- predict(tree, test)
CrossTable(test_labels, tree_pred,
            prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,
            dnn = c('clase observada', 'clase estimada'))
accuracy = (44+71+15)/(44+71+15+30+1+23+7+3+23)
accuracy
```

```
boost10 <- C5.0(train, train_labels,
                trials = 10)
summary(boost10)
tree_pred10 <- predict(boost10, test)
CrossTable(test_labels, tree_pred10,
            prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,
            dnn = c('clase observada', 'clase estimada'))
accuracy = (54+57+15)/(54+57+15+20+1+31+13+3+23)
accuracy
```

```
boost20 <- C5.0(train, train_labels,
                trials = 20)
# summary(boost20)
tree_pred20 <- predict(boost20, test)
CrossTable(test_labels, tree_pred20,
            prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,
            dnn = c('clase observada', 'clase estimada'))
accuracy = (52+66+16)/(52+66+16+22+1+26+9+2+23)
accuracy
```

```
boost50 <- C5.0(train, train_labels,
                trials = 50)
# summary(boost50)
tree_pred50 <- predict(boost50, test)
CrossTable(test_labels, tree_pred50,
            prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,
            dnn = c('clase observada', 'clase estimada'))
accuracy = (55+64+14)/(55+64+14+19+1+29+8+2+25)
accuracy
```

```
##### Método Ranforest
```

```
res = randomForest(oseo ~ ., data=dexa_oseo)
```

```

res
#varImpPlot(res)
accuracy = (256+362+89)/(256+362+89+118+4+104+35+3+111)
accuracy

#### Femenino

png("~/Dropbox/TFM/Final 1.0/figurasR/fig10.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

dexa_oseo <- sqldf("select oseo, Height, Weightg, age, TotalBMD, L1BMD, L2BMD,
                    L3BMD, L4BMD, L1L4BMD, L2L4BMD, NeckFBMD, WardsBMD,
                    TrochBMD, TotalFBMD, BMI from dexa_m")

dexa_oseo<-sqldf("select oseo, TotalBMD,
                    Height, Weight, Age,
                    RAfg, RALg, LAfg, LALg,
                    RLfg, RLLg, LLfg, LLLg,
                    TFg, TLg
                    from dexa_m")

table(dexa_oseo$oseo)

round(prop.table(table(dexa_oseo$oseo)) * 100, digits = 1)
lbls <- paste(names(table(dexa_oseo$oseo)),
              round(prop.table(table(dexa_oseo$oseo)) * 100,
                    digits = 1), "%")
pie(table(dexa_oseo$oseo), labels = lbls, col = c("Blue", "Red", "Yellow"))

##### Método k vecinos más próximos
### mujeres

set.seed(123)
split.indices <- sample(nrow(dexa_oseo), nrow(dexa_oseo)*0.8, replace = F)

train <- dexa_oseo[split.indices, -1]
test <- dexa_oseo[-split.indices, -1]

train_labels <- dexa_oseo[split.indices, 1]
test_labels <- dexa_oseo[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
n <- as.data.frame(lapply(dexa_oseo[2:15], normalize))
n_train <- n[split.indices, ]

```

```

n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test,
                 cl = train_labels, k = 21)
CrossTable(x = test_labels, y = test_pred,
           prop.chisq = FALSE)

accuracy_knn = (17+21+5)/(17+21+5+12+1+6+7)
accuracy_knn

##### Método clasificador bayesiano

classifier <- naiveBayes(train, train_labels)
eval_pred <- predict(classifier, test)
head(eval_pred)
CrossTable(eval_pred, test_labels, prop.chisq = FALSE, prop.t = FALSE,
           prop.r = FALSE, dnn = c('estimado', 'observado'))
accuracy = (19+15+8)/(19+15+8+5+1+8+3+3+7)
accuracy

```

***** Sarcopenia *****

```

dexa$sarco[dexa$FFMI < 20] <- "Normal"
dexa$sarco[dexa$FFMI >= 20] <- "Sarcopenia"
dexa$sarco <- as.factor(dexa$sarco)
dexa_h <- dexa[ which(dexa$gender_num==1),]
dexa_m <- dexa[ which(dexa$gender_num==2),]

```

Masa magra Hombres

```

dexa_sarco_h <- sqldf("select sarco, Age,
                        Height, Weight,
                        RAf, LAf,
                        RLf, LLf,
                        TF,
                        L1BMD, L2BMD, L3BMD, L4BMD,
                        L1L4BMD, L2L4BMD,
                        NeckFBMD, WardsBMD, TrochBMD,
                        TotalFBMD,
                        TotalLg
                        from dexa_h")

```

```

table(dexa_sarco_h$sarco)
round(prop.table(table(dexa_sarco_h$sarco)) * 100, digits = 1)
lbls <- paste(names(table(dexa_sarco_h$sarco)), round(prop.table(table(dexa_sarco_h$sarco)) * 100, digits = 1))
png("~/Dropbox/TFM/Final 1.0/figurasR/fig40.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

```

```
pie(table(dexa_sarcoh$sarco), labels = lbls, main = "Hombres", col = c("Blue", "Yellow"))
```

```
dexa_sarcoh<-sqldf("select Age,
                        Height, Weight,
                        RAfg, LAfg,
                        RLfg, LLfg,
                        TFg,
                        L1BMD, L2BMD, L3BMD, L4BMD,
                        L1L4BMD, L2L4BMD,
                        NeckFBMD, WardsBMD, TrochBMD,
                        TotalFBMD,
                        TotalLg
                        from dexa_h")
```

```
cors <- cor(dexa_sarcoh)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0.3 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dexa_sarcoh)[cor.idx[,1]], "+",
                    colnames(dexa_sarcoh)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
j<- grep("TotalLg", cor.names)

i=1
for(i in i:j) {
  print(cor.names[j])
}
```

```
### Modelo de regresión
#### Masculino
```

```
dexa_sarco<-sqldf("select Age,
                        Height, Weight,
                        RAfg, LAfg,
                        RLfg, LLfg,
                        TFg,
                        L1BMD, L2BMD, L3BMD, L4BMD,
                        L1L4BMD, L2L4BMD,
                        NeckFBMD, WardsBMD, TrochBMD,
                        TotalFBMD,
                        TotalLg
                        from dexa_h")
```

```
fit <- lm(TotalLg ~ ., data=dexa_sarco)
summary(fit)
```

```
# Stepwise Regression
```

```
step <- stepAIC(fit, direction="both")
step$anova
```



```
## Modelo reducido
```

```
dexa_sarco<-sqldf("select TotalLg,  
                    Height, Weight,  
                    RAfg, LAfg,  
                    RLfg, LLfg,  
                    TFg,  
                    NeckFBMD  
                    from dexa_h")
```

```
fit2 <- lm(TotalLg ~ ., data=dexa_sarco)  
summary(fit2)
```

```
coeficientes <- coefficients(fit2)  
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)  
pandoc.table( coeficientes  
              , caption = "Coeficientes modelo magro hombres"  
              , justify = 'lrrrrrrrr'  
              , round = 3  
              , decimal.mark = "."  
              , big.mark = ",",  
              , emphasize.rownames = TRUE  
              , split.cells = 5  
              )
```

```
##
```

```
## -----  
## (Intercept)      Height    Weight    RAfg    LAfg    RLfg    LLfg    TFg    NeckFBMD  
## -----  
## 1660.50          -1421.46    977.25    -0.83    -1.42    -1.26    -0.81    -0.97    -1004.4  
## -----
```

```
##
```

```
## Table: Coeficientes modelo magro hombres
```

```
### Probar el modelo magro hombres
```

```
set.seed(123)  
N.train <- ceiling(0.7 * nrow(dexa_sarco))  
N.test <- nrow(dexa_sarco) - N.train  
trainset <- sample(seq_len(nrow(dexa_sarco)), N.train)  
testset <- setdiff(seq_len(nrow(dexa_sarco)), trainset)  
  
modelo <- lm(TotalLg ~ ., dexa_sarco, trainset)  
  
res <- residuals(modelo)  
pred <- modelo$fitted.values  
obs <- dexa_sarco[trainset, "TotalLg"]  
val.range <- range(pred, obs)
```

```
png("~/Dropbox/TFM/Final 1.0/figurasR/fig41.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)
```

```
plot(obs, pred,
     xlim = val.range, ylim = val.range,
     xlab = "Masa grasa total observados",
     ylab = "Masa grasa total esperados",
     main = "")
abline(0,1, col = "red")
segments(obs, pred, obs, pred + res)
```

```
png("~/Dropbox/TFM/Final 1.0/figurasR/fig42.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)
```

```
layout(matrix(c(1,2,3,4),2,2))
plot(modelo)
```

```
## Masa magra Mujeres
```

```
dexa_sarcom<-sqldf("select sarco, Age,
                      Height, Weight,
                      RAFe, LAFe,
                      RLFe, LLFe,
                      TFe,
                      L1BMD, L2BMD, L3BMD, L4BMD,
                      L1L4BMD, L2L4BMD,
                      NeckFBMD, WardsBMD, TrochBMD,
                      TotalFBMD,
                      TotalLg
                    from dexa_m")
```

```
table(dexa_sarcom$sarco)
round(prop.table(table(dexa_sarcom$sarco)) * 100, digits = 1)
lbls <- paste(names(table(dexa_sarcom$sarco)), round(prop.table(table(dexa_sarcom$sarco)) * 100, digits = 1))
png("~/Dropbox/TFM/Final 1.0/figurasR/fig43.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)
```

```
pie(table(dexa_sarcom$sarco), labels = lbls, main = "Mujeres", col = c("Blue", "Yellow"))
```

```
dexa_sarcom<-sqldf("select Age,
                      Height, Weight,
                      RAFe, LAFe,
```

```

        RLFg, LLFg,
        TFg,
        L1BMD, L2BMD, L3BMD, L4BMD,
        L1L4BMD, L2L4BMD,
        NeckFBMD, WardsBMD, TrochBMD,
        TotalFBMD,
        TotalLg
        from dexa_m")

cors <- cor(dexa_sarcom)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0.3 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dexa_sarcom)[cor.idx[,1]], "+",
                    colnames(dexa_sarcom)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
j<- grep("TotalLg", cor.names)

i=1
for(i in i:j) {
  print(cor.names[j])
}

## Analisis de normalidad magro mujeres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig56.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

ggplot(data = dexa_sarcom, aes(x = TotalLg)) +
  geom_histogram(aes(y = ..density.., fill = ..count..)) +
  scale_fill_gradient(low = "#DCDCDC", high = "#7C7C7C") +
  stat_function(fun = dnorm, colour = "firebrick",
               args = list(mean = mean(dexa_sarcom$TotalLg),
                           sd = sd(dexa_sarcom$TotalLg))) +
  ggtitle("") +
  theme_bw()

#### Modelo de regresión
#### Mujeres

dexa_sarco<-sqldf("select Age,
                        Height, Weight,
                        RAFg, LAFg,
                        RLFg, LLFg,
                        TFg,
                        L1BMD, L2BMD, L3BMD, L4BMD,

```

```

        L1L4BMD, L2L4BMD,
        NeckFBMD, WardsBMD, TrochBMD,
        TotalFBMD,
        TotalLg
        from dexa_m")

fit <- lm(TotalLg ~ ., data=dexa_sarco)
summary(fit)

# Stepwise Regression
step <- stepAIC(fit, direction="both")
step$anova

## Modelo reducido

dexa_sarco<-sqldf("select TotalLg,
                    Height, Weight,
                    RAFg, LAFg,
                    RLFg, LLFg,
                    TFG,
                    L1L4BMD, WardsBMD
                    from dexa_m")

fit2 <- lm(TotalLg ~ ., data=dexa_sarco)
summary(fit2)

coeficientes <- coefficients(fit2)
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)
coeficientes

### Probar el modelo magro mujeres

set.seed(123)
N.train <- ceiling(0.7 * nrow(dexa_sarco))
N.test <- nrow(dexa_sarco) - N.train
trainset <- sample(seq_len(nrow(dexa_sarco)), N.train)
testset <- setdiff(seq_len(nrow(dexa_sarco)), trainset)

modelo <- lm(TotalLg ~ ., dexa_sarco, trainset)

res <- residuals(modelo)
pred <- modelo$fitted.values
obs <- dexa_sarco[trainset, "TotalLg"]
val.range <- range(pred, obs)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig44.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

```

```

plot(obs, pred,
      xlim = val.range, ylim = val.range,
      xlab = "Masa grasa total observados",
      ylab = "Masa grasa total esperados",
      main = "")
abline(0,1, col = "red")
segments(obs, pred, obs, pred + res)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig45.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

layout(matrix(c(1,2,3,4),2,2))
plot(modelo)

## Masa magra
### Modelo de clasificación
#### Masculino
##### Método k vecinos más próximos

dexa_sarcoh<-sqldf("select sarco, Age,
                    Height, Weight,
                    RAFg, LAFg,
                    RLFg, LLFg,
                    TFG,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalLg
                    from dexa_h")

set.seed(123)
split.indices <- sample(nrow(dexa_sarcoh), nrow(dexa_sarcoh)*0.8, replace = F)

train <- dexa_sarcoh[split.indices, -1]
test <- dexa_sarcoh[-split.indices, -1]

train_labels <- dexa_sarcoh[split.indices, 1]
test_labels <- dexa_sarcoh[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
n <- as.data.frame(lapply(dexa_sarcoh[2:20], normalize))

```

```

n_train <- n[split.indices, ]
n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 1)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (194+5)/(194+5+4+14)
accuracy_knn

## Masa magra
### Modelo de clasificación
#### Femenino
##### Método k vecinos más próximos

dexa_sarcom<-sqldf("select sarco, Age,
                    Height, Weight,
                    RAfg, LAfg,
                    RLFg, LLFg,
                    TFg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalLg
                    from dexa_m")

set.seed(123)
split.indices <- sample(nrow(dexa_sarcom), nrow(dexa_sarcom)*0.8, replace = F)

train <- dexa_sarcom[split.indices, -1]
test <- dexa_sarcom[-split.indices, -1]

train_labels <- dexa_sarcom[split.indices, 1]
test_labels <- dexa_sarcom[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
n <- as.data.frame(lapply(dexa_sarcom[2:20], normalize))
n_train <- n[split.indices, ]
n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 1)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (67+2)/(67+2)
accuracy_knn

```

Lipodistrofia *****

```
dexa$lipo[dexa$FMI < 1.5 & dexa$sexo == "hombre"] <- "Normal"
dexa$lipo[dexa$FMI >= 1.5 & dexa$sexo == "hombre"] <- "Lipodistrofia"
dexa$lipo[dexa$FMI < 1.2 & dexa$sexo == "mujer"] <- "Normal"
dexa$lipo[dexa$FMI >= 1.2 & dexa$sexo == "mujer"] <- "Lipodistrofia"
dexa$lipo <- as.factor(dexa$lipo)
```

```
dexa_h <- dexa[ which(dexa$gender_num==1),]
dexa_m <- dexa[ which(dexa$gender_num==2),]
```

```
dexa_lipoh<-sqldf("select lipo, Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalFg from dexa_h")
```

```
table(dexa_lipoh$lipo)
round(prop.table(table(dexa_lipoh$lipo)) * 100, digits = 1)
lbls <- paste(names(table(dexa_lipoh$lipo)), round(prop.table(table(dexa_lipoh$lipo))
```

```
png("~/Dropbox/TFM/Final 1.0/figurasR/fig30.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)
```

```
pie(table(dexa_lipoh$lipo), labels = lbls, main = "Hombres", col = c("Blue", "Yellow"
```

```
dexa_lipoh<-sqldf("select Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalFg from dexa_h")
```

```
cors <- cor(dexa_lipoh)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dexa_lipoh)[cor.idx[,1]], "+",
                    colnames(dexa_lipoh)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
#print(cor.names)
```

```
# Y cuales son las correlaciones con la masa ósea total
```

```

j<- grep("TotalFg", cor.names)

i=1
for(i in i:j) {
  print(cor.names[j])
}
correlaciones <- cor(dexa_lipoh[,1:19])
pandoc.table( correlaciones
  , caption = "Matriz de correlaciones para variables magras (masculino)"
  , justify = 'lrrrrrrrrrrrrrrrrrrrrrr'
  , round = 3
  , decimal.mark = "."
  , big.mark = ","
  , emphasize.rownames = TRUE
  , split.cells = 5
)

dexa_lipom <- sqldf("select lipo, Height, Weight, age, RAFp, RAFg, LAFp, LAFg,
  BothAFp, BothAFg, RLFp, RLFg, LLFp, LLFg, BothLFp,
  BothLFg, TFp, TFG, TotalFp, TotalFg, BMI, FMI, FMR,
  FTrunkgFLegsg, Indexdistributionfat, FtrunkpFlimbsp,
  FtrunkgFtotalg, FLegsgFtotalg, FlimbspFtotalg, LLegFgBMI,
  LLegFpBMI from dexa_m")
correlaciones <- cor(dexa_lipom[,5:31])
pandoc.table( correlaciones
  , caption = "Matriz de correlaciones para variables magras (femenino)"
  , justify = 'lrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr'
  , round = 3
  , decimal.mark = "."
  , big.mark = ","
  , emphasize.rownames = TRUE
  , split.cells = 5
)

### Modelo de regresión
#### Masculino

dexa_lipo<-sqldf("select Height, Weight, Age,
  RALg, LALg,
  RLLg, LLLg,
  TLg,
  L1BMD, L2BMD, L3BMD, L4BMD,
  L1L4BMD, L2L4BMD,
  NeckFBMD, WardsBMD, TrochBMD,
  TotalFBMD,
  TotalFg from dexa_h")

fit <- lm(TotalFg ~ ., data=dexa_lipo)

```

```

summary(fit)

# Stepwise Regression
step <- stepAIC(fit, direction="both")
step$anova

## Modelo reducido

dexa_lipo<-sqldf("select TotalFg,
                Height, Weight, Age,
                RALg, LALg,
                RLLg, LLLg,
                TLg,
                L2L4BMD
                from dexa_h")

fit2 <- lm(TotalFg ~ ., data=dexa_lipo)
summary(fit2)

### Probar el modelo graso hombres

set.seed(123)
N.train <- ceiling(0.7 * nrow(dexa_lipo))
N.test <- nrow(dexa_lipo) - N.train
trainset <- sample(seq_len(nrow(dexa_lipo)), N.train)
testset <- setdiff(seq_len(nrow(dexa_lipo)), trainset)

modelo <- lm(TotalFg ~ ., dexa_lipo, trainset)

res <- residuals(modelo)
pred <- modelo$fitted.values
obs <- dexa_lipo[trainset, "TotalFg"]
val.range <- range(pred, obs)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig32.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

plot(obs, pred,
     xlim = val.range, ylim = val.range,
     xlab = "Masa grasa total observados",
     ylab = "Masa grasa total esperados",
     main = "")
abline(0,1, col = "red")
segments(obs, pred, obs, pred + res)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig33.png",
    width=1000,height=800,units="px",

```

```

    pointsize=8,bg="white",res=150)

layout(matrix(c(1,2,3,4),2,2))
plot(modelo)

### Analisis masa grasa mujeres

dexa_lipom<-sqldf("select lipo, Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalFg from dexa_m")

table(dexa_lipom$lipo)
round(prop.table(table(dexa_lipom$lipo)) * 100, digits = 1)
lbls <- paste(names(table(dexa_lipom$lipo)), round(prop.table(table(dexa_lipom$lipo))

png("~/Dropbox/TFM/Final 1.0/figurasR/fig34.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

pie(table(dexa_lipom$lipo), labels = lbls, main = "Mujeres", col = c("Blue", "Yellow")

dexa_lipom<-sqldf("select Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalFg from dexa_m")

cors <- cor(dexa_lipom)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0.3 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dexa_lipom)[cor.idx[,1]], "+",
                    colnames(dexa_lipom)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
#print(cor.names)

# Y cuales son las correlaciones con la masa ósea total
j<- grep("TotalFg", cor.names)

```

```

i=1
for(i in i:j) {
  print(cor.names[j])
}

```

```

### Modelo de regresión
#### Femenino

```

```

dexa_lipo<-sqldf("select Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,
                    L1L4BMD, L2L4BMD,
                    NeckFBMD, WardsBMD, TrochBMD,
                    TotalFBMD,
                    TotalFg from dexa_m")

```

```

fit <- lm(TotalFg ~ ., data=dexa_lipo)
summary(fit)

```

```

# Stepwise Regression

```

```

step <- stepAIC(fit, direction="both")
step$anova

```

```

## Modelo reducido

```

```

dexa_lipo<-sqldf("select TotalFg,
                    Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1L4BMD,
                    TotalFBMD
                    from dexa_m")

```

```

fit2 <- lm(TotalFg ~ ., data=dexa_lipo)
summary(fit2)

```

```

coeficientes <- coefficients(fit2)
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)
pandoc.table( coeficientes
               , caption = "Coeficientes del modelo lineal masa grasa mujeres"
               , justify = 'lrrrrrrrrrrr'
               , round = 3
               , decimal.mark = "."
               , big.mark = ",",

```

```

        , emphasize.rownames = TRUE
        , split.cells = 5
    )

### Probar el modelo graso mujeres

set.seed(123)
N.train <- ceiling(0.7 * nrow(dexa_lipo))
N.test <- nrow(dexa_lipo) - N.train
trainset <- sample(seq_len(nrow(dexa_lipo)), N.train)
testset <- setdiff(seq_len(nrow(dexa_lipo)), trainset)

modelo <- lm(TotalFg ~ ., dexa_lipo, trainset)

res <- residuals(modelo)
pred <- modelo$fitted.values
obs <- dexa_lipo[trainset, "TotalFg"]
val.range <- range(pred, obs)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig36.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

plot(obs, pred,
     xlim = val.range, ylim = val.range,
     xlab = "Masa grasa total observados",
     ylab = "Masa grasa total esperados",
     main = "")
abline(0,1, col = "red")
segments(obs, pred, obs, pred + res)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig35.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

layout(matrix(c(1,2,3,4),2,2))
plot(modelo)

### Modelo de clasificación
#### Masculino
##### Método k vecinos más próximos

dexa_lipoh<-sqldf("select lipo, Height, Weight, Age,
                    RALg, LALg,
                    RLLg, LLLg,
                    TLg,
                    L1BMD, L2BMD, L3BMD, L4BMD,

```

```

        L1L4BMD, L2L4BMD,
        NeckFBMD, WardsBMD, TrochBMD,
        TotalFBMD,
        TotalFg from dexa_h")

set.seed(123)
split.indices <- sample(nrow(dexa_lipoh), nrow(dexa_lipoh)*0.8, replace = F)

train <- dexa_lipoh[split.indices, -1]
test <- dexa_lipoh[-split.indices, -1]

train_labels <- dexa_lipoh[split.indices, 1]
test_labels <- dexa_lipoh[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
n <- as.data.frame(lapply(dexa_lipoh[2:20], normalize))
n_train <- n[split.indices, ]
n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 1)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (207+2)/(207+2+2+6)
accuracy_knn

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 5)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (209+1)/(209+1+7)
accuracy_knn

#### Femenino
##### Método k vecinos más próximos

dexa_lipom<-sqldf("select lipo, Height, Weight, Age,
        RALg, LALg,
        RLLg, LLLg,
        TLg,
        L1BMD, L2BMD, L3BMD, L4BMD,
        L1L4BMD, L2L4BMD,
        NeckFBMD, WardsBMD, TrochBMD,
        TotalFBMD,
        TotalFg from dexa_m")

set.seed(123)
split.indices <- sample(nrow(dexa_lipom), nrow(dexa_lipom)*0.8, replace = F)

```

```

train <- dextra_lipom[split.indices, -1]
test <- dextra_lipom[-split.indices, -1]

train_labels <- dextra_lipom[split.indices, 1]
test_labels <- dextra_lipom[-split.indices, 1]

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
n <- as.data.frame(lapply(dextra_lipom[2:20], normalize))
n_train <- n[split.indices, ]
n_test <- n[-split.indices, ]

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 1)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (68+1)/(68+1+5)
accuracy_knn

test_pred <- knn(train = n_train, test = n_test, cl = train_labels, k = 5)
CrossTable(x = test_labels, y = test_pred, prop.chisq=FALSE)
accuracy_knn = (209+1)/(209+1+7)
accuracy_knn

# Análisis de regresión Masa Ósea hombres

dextra_oseoh<-sqldf("select  TotalBMD,
                          Height, Weight, Age,
                          RAFg, RALg, LAFg, LALg,
                          RLFg, RLLg, LLFg, LLLg,
                          TFG, TLg
                          from dextra_h")

cors <- cor(dextra_oseoh)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0.3 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dextra_oseoh)[cor.idx[,1]], "+",
                    colnames(dextra_oseoh)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
#print(cor.names)

# Y cuales son las correlaciones con la masa ósea total
j<- grep("TotalBMD", cor.names)

i=1
for(i in 1:j) {
  print(cor.names[j])
}

```

```

set.seed(123)
N.train <- ceiling(0.7 * nrow(dexa_oseoh))
N.test <- nrow(dexa_oseoh) - N.train
trainset <- sample(seq_len(nrow(dexa_oseoh)), N.train)
testset <- setdiff(seq_len(nrow(dexa_oseoh)), trainset)

```

```

modelo <- lm(TotalBMD ~ ., dexa_oseoh, trainset)
model.summary <- summary(modelo)
print(model.summary)
res <- residuals(modelo)
pred <- modelo$fitted.values
obs <- dexa_oseoh[trainset, "TotalBMD"]

val.range <- range(pred, obs)

```

```

#plot(obs, pred,
#      xlim = val.range, ylim = val.range,
#      xlab = "Masa osea total observados",
#      ylab = "Masa osea total esperados",
#      main = "Residuales del modelo lineal para datos de entrenamiento")
#abline(0,1, col = "red")
#segments(obs, pred, obs, pred + res)

```

```

coeficientes <- coefficients(modelo)
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)
pandoc.table( coeficientes
               , caption = "Resumen estadístico general de los datos del estudio por s
               , justify = 'lrrrrrrrrrrrrrrr'
               , round = 3
               , decimal.mark = "."
               , big.mark = ", "
               , emphasize.rownames = TRUE
               , split.cells = 5
               )

```

```

coefficients(modelo)
model.summary$coefficients
var.matrix <- vcov(modelo)
#print(var.matrix)
std.err <- sqrt(diag(vcov(modelo)))
print(std.err)
scaled.var.matrix <- model.summary$cov.unscaled * model.summary$sigma^2
#print(scaled.var.matrix)

t.vals <- coef(modelo) / std.err
print(t.vals)
model.df <- df.residual(modelo)
rss <- sum(res^2)

```

```

mean.rss <- rss / model.df
res.std.error <- sqrt(mean.rss)
print(res.std.error)
r.squared <- cor(pred, obs)^2
print(r.squared)

ci <- confint(modelo)
apply(ci, 1, function(x) paste0("95% CI: [",
                                round(x[1], 2), ",", round(x[2], 2), "]"))

layout(matrix(c(1,2,3,4),2,2))
#plot(modelo)

step <- stepAIC(modelo, direction="both")
step$anova

modelo2 <- lm(TotalBMD ~ Height + Weight + Age + RALg + LAFg +
              RLFg + LLFg +
              TFg + TLg, dextra_oseoh, trainset)

model2.summary <- summary(modelo2)
print(model2.summary)
res2 <- residuals(modelo2)
pred2 <- modelo2$fitted.values
obs2 <- dextra_oseoh[trainset, "TotalBMD"]

val.range <- range(pred2, obs2)

coeficientes <- coefficients(modelo2)
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)
pandoc.table( coeficientes
              , caption = "Resumen estadístico general de los datos del estudio por s
              , justify = 'lrrrrrrrrrr'
              , round = 3
              , decimal.mark = "."
              , big.mark = ","
              , emphasize.rownames = TRUE
              , split.cells = 5
              )

png("~/Dropbox/TFM/Final 1.0/figurasR/fig11.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

plot(obs2, pred2,
     xlim = val.range, ylim = val.range,
     xlab = "Masa osea total observados",

```



```

    ylab = "Masa osea total esperados",
    main = "")
abline(0,1, col = "red")
segments(obs2, pred2, obs2, pred2 + res2)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig12.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

layout(matrix(c(1,2,3,4),2,2))
plot(modelo2)

png("~/Dropbox/TFM/Final 1.0/figurasR/fig13.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

ggplot(data = dexa_oseoh, aes(x = TotalBMD)) +
  geom_histogram(aes(y = ..density.., fill = ..count..)) +
  scale_fill_gradient(low = "#DCDCDC", high = "#7C7C7C") +
  stat_function(fun = dnorm, colour = "firebrick",
               args = list(mean = mean(dexa_oseoh$TotalBMD),
                           sd = sd(dexa_oseoh$TotalBMD))) +
  ggtitle("") +
  theme_bw()

shapiro.test(x = dexa_oseoh$TotalBMD)
ks.test(x = dexa_oseoh$TotalBMD,"pnorm", mean(dexa_oseoh$TotalBMD), sd(dexa_oseoh$Tot
lillie.test(x = dexa_oseoh$TotalBMD)

# Análisis de regresión Masa Osea Mujeres

dexa_oseom<-sqldf("select  TotalBMD,
                        Height, Weight, Age,
                        RAfg, RALg, LAfg, LALg,
                        RLFg, RLLg, LLFg, LLLg,
                        TFg, TLg
                        from dexa_m")

cors <- cor(dexa_oseom)
# Cuáles variables están altamente correlacionadas excluyendo autocorrelación
cor.idx <- which(abs(cors) >= 0.3 & cors != 1, arr.ind = TRUE)
cor.names <- paste0(colnames(dexa_oseom)[cor.idx[,1]], "+",
                    colnames(dexa_oseom)[cor.idx[,2]], ": ", round(cors[cor.idx], 2))
#print(cor.names)

```

```

# Y cuales son las correlaciones con la masa ósea total
j<- grep("TotalBMD", cor.names)

i=1
for(i in i:j) {
  print(cor.names[j])
}

set.seed(123)
N.train <- ceiling(0.7 * nrow(dexa_oseom))
N.test <- nrow(dexa_oseom) - N.train
trainset <- sample(seq_len(nrow(dexa_oseom)), N.train)
testset <- setdiff(seq_len(nrow(dexa_oseom)), trainset)

modelo <- lm(TotalBMD ~ ., dexa_oseom, trainset)
model.summary <- summary(modelo)
print(model.summary)
res <- residuals(modelo)
pred <- modelo$fitted.values
obs <- dexa_oseom[trainset, "TotalBMD"]

val.range <- range(pred, obs)

#plot(obs, pred,
#   xlim = val.range, ylim = val.range,
#   xlab = "Masa osea total observados",
#   ylab = "Masa osea total esperados",
#   main = "Residuales del modelo lineal para datos de entrenamiento")
#abline(0,1, col = "red")
#segments(obs, pred, obs, pred + res)

#coefficients(modelo)
model.summary$coefficients
var.matrix <- vcov(modelo)
#print(var.matrix)
std.err <- sqrt(diag(vcov(modelo)))
print(std.err)
scaled.var.matrix <- model.summary$cov.unscaled * model.summary$sigma^2
#print(scaled.var.matrix)

t.vals <- coef(modelo) / std.err
print(t.vals)
model.df <- df.residual(modelo)
rss <- sum(res^2)
mean.rss <- rss / model.df
res.std.error <- sqrt(mean.rss)
print(res.std.error)
r.squared <- cor(pred, obs)^2

```

```

print(r.squared)

ci <- confint(modelo)
apply(ci, 1, function(x) paste0("95% CI: [",
                                round(x[1], 2), ",", round(x[2], 2), "]"))

layout(matrix(c(1,2,3,4),2,2))
#plot(modelo)

step <- stepAIC(modelo, direction="both")
step$anova

modelo2 <- lm(TotalBMD ~ Height + Weight + Age + RAFg + LAFg +
              LALg + RLFg + RLLg + TFG + TLg, dexta_oseom, trainset)

model2.summary <- summary(modelo2)
print(model2.summary)
res2 <- residuals(modelo2)
pred2 <- modelo2$fitted.values
obs2 <- dexta_oseom[trainset, "TotalBMD"]
coefficients(modelo)
val.range <- range(pred2, obs2)

coeficientes <- coefficients(modelo2)
coeficientes <- format(coeficientes, digits = 2, nsmall = 1, scientific=F)
pandoc.table( coeficientes
              , caption = "Resumen estadístico general de los datos del estudio por s
              , justify = 'lrrrrrrrrrrr'
              , round = 3
              , decimal.mark = "."
              , big.mark = ","
              , emphasize.rownames = TRUE
              , split.cells = 5
              )

png("~/Dropbox/TFM/Final 1.0/figurasR/fig14.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

plot(obs2, pred2,
     xlim = val.range, ylim = val.range,
     xlab = "Masa osea total observados",
     ylab = "Masa osea total esperados",
     main = "")
abline(0,1, col = "red")
segments(obs2, pred2, obs2, pred2 + res2)

```

```

png("~/Dropbox/TFM/Final 1.0/figurasR/fig15.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

layout(matrix(c(1,2,3,4),2,2))
plot(modelo2)

## Analisis de normalidad

png("~/Dropbox/TFM/Final 1.0/figurasR/fig16.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

ggplot(data = dextra_oseom, aes(x = TotalBMD)) +
  geom_histogram(aes(y = ..density.., fill = ..count..)) +
  scale_fill_gradient(low = "#DCDCDC", high = "#7C7C7C") +
  stat_function(fun = dnorm, colour = "firebrick",
               args = list(mean = mean(dextra_oseom$TotalBMD),
                           sd = sd(dextra_oseom$TotalBMD))) +
  ggtitle("") +
  theme_bw()

shapiro.test(x = dextra_oseom$TotalBMD)
ks.test(x = dextra_oseom$TotalBMD,"pnorm", mean(dextra_oseom$TotalBMD), sd(dextra_oseom$TotalBMD))
lillie.test(x = dextra_oseom$TotalBMD)

##### Análisis Longitudinal DEXA

png("~/Dropbox/TFM/Final 1.0/figurasR/fig20.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

frecuencia_dexa <- frecuencia_dexa[-1476,] # eliminamos el renglón del total
colnames(frecuencia_dexa) <- c("ID", "frecuencia") # cambiamos los nombres de las columnas
frecuencias <- sqldf("select * from frecuencia_dexa where frecuencia > 1") # Solo aquellas con frecuencia mayor a 1

### Gráfico de conteo de dexas
ggplot(frecuencias, aes(x=frecuencia)) + geom_bar(fill="steelblue") +
  labs(x = "Numero de DEXAS", y = "Conteo")

```

```
##### Análisis Longitudinal asignación unidades de tiempo
```

```
dexal <- na.omit(dexal) # Se eliminan todos los datos que tengan NA's
```

```
datos_frecuencia <- sqldf("select frecuencia, count(*) as cuantos from frecuencias  
                           group by frecuencia")
```

```
pandoc.table( datos_frecuencia  
              , caption = "Número de DEXAS"  
              , justify = 'cc'  
              , round = 3  
              , decimal.mark = "."  
              , big.mark = ","  
              , emphasize.rownames = TRUE  
              , split.cells = 5  
              )
```

```
x <- ymd(dexal$DateAnalysis)
```

```
dexal$cuatrimestre <- quarter(x, with_year = TRUE)
```

```
summary(dexal$cuatrimestre)
```

```
a<-sqldf("select distinct(cuatrimestre) from dexal order by cuatrimestre")
```

```
i <- 1  
while (i < 70) {  
  a$tiempo[i]<-i  
  i = i+1  
}
```

```
base<-sqldf("select a.ID, b.tiempo, a.DateBirth, a.DateAnalysis, a.TotalBMD,  
  a.Height, a.Weightg, a.RAFp, a.RAFg, a.RALg, a.LAFp, a.LAFg, a.LALg, a.BothAFp,  
  a.BothAFg, a.BothALg, a.RLFp, a.RLFg, a.RLLg, a.LLFp, a.LLFg, a.LLLg, a.BothLFp,  
  a.BothLFg, a.BothLLg, a.TFp, a.TFg, a.TLg, a.TotalFp, a.TotalFg, a.TotalLg,  
  a.L1BMD, a.L1T, a.L1Z, a.L2BMD, a.L2T, a.L2Z, a.L3BMD, a.L3T, a.L3Z, a.L4BMD,  
  a.L4T, a.L4Z, a.L1L4BMD, a.L1L4T, a.L1L4Z, a.L2L4BMD, a.L2L4T, a.L2L4Z,  
  a.NeckFBMD, a.NeckFT, a.NeckFZ, a.WardsBMD, a.WardsT, a.WardsZ, a.TrochBMD,  
  a.TrochT, a.TrochZ, a.TotalFBMD, a.TotalFT, a.TotalFZ, a.Weight, a.BMI,  
  a.FMI, a.FFMI, a.Apendicularleanmas, a.FMR, a.FTrunkgFlegsg,  
  a.Indexdistributionfat, a.FtrunkpFlimbsp, a.FtrunkgFtotalg, a.FlegsgFtotalg,  
  a.FlimbsgFtotalg, a.LLegFgBMI, a.LLegFpBMI, a.Age, a.sexo,  
  a.cuatrimestre from dexal a, a b where a.cuatrimestre=b.cuatrimestre")
```

```
longitudinal <- sqldf("select * from base where (ID in (select ID from frecuencias))  
                      order by ID, DateAnalysis")
```

```
## Conteo
```

```

todos <- sqldf("select count(distinct(ID)) from longitudinal")
hombres <- sqldf("select count(distinct(ID)) from longitudinal where sexo='hombre'")
mujeres <- sqldf("select count(distinct(ID)) from longitudinal where sexo='mujer'")

testM <- sqldf("select * from longitudinal where sexo='hombre'")
testF <- sqldf("select * from longitudinal where sexo='mujer'")

## Tomaré el 20% de los datos para hacer los análisis
## Si tomo el 100% sale emplastado

set.seed(123)
split.indices <- sample(nrow(testM), nrow(testM)*0.10, replace = F)
muestra <- testM[split.indices, ]
testM <- muestra

set.seed(123)
split.indices <- sample(nrow(testF), nrow(testF)*0.20, replace = F)
muestra <- testF[split.indices, ]
testF <- muestra

##### Análisis Longitudinal Masa ósea hombres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig21.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, TotalBMD from testM")
plot(df$tiempo, df$TotalBMD, xlab = "Tiempo", ylab="Densidad Masa Osea Total", main
     type="n", xlim = c(0,70), ylim=c(0,2))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$TotalBMD[df$ID == i]), col = 4)
}

##### Análisis Longitudinal

df <- sqldf("select * from longitudinal where sexo='hombre'")

set.seed(123)
split.indices <- sample(nrow(df), nrow(df)*0.8, replace = F)
train <- df[split.indices, ]
test <- df[-split.indices, ]

df<- sqldf("select ID, tiempo, TotalBMD, Age, BMI, FMI, FFMI from df")

```

```
head(df)
m2 <- lme(TotalBMD ~ tiempo + Age+ BMI + FMI + FFMI, random = ~ tiempo | ID, data=tra
summary(m2)
```

```
##### Análisis Longitudinal
##### Analisis de residuales
```

```
res <- residuals(m2, type = "normalized")
fit <- fitted(m2)
```

```
#test 1
```

```
pred = predict(m2, test)
```

```
png("~/Dropbox/TFM/Final 1.0/figurasR/fig40reshom.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)
```

```
plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
     cex.sub=0.5)
abline(0,1, col = "red")
# test
```

```
##### Análisis Longitudinal
```

```
df <- sqldf("select * from longitudinal where sexo='mujer'")
```

```
set.seed(123)
split.indices <- sample(nrow(df), nrow(df)*0.8, replace = F)
train <- df[split.indices, ]
test <- df[-split.indices, ]
```

```
df<- sqldf("select ID, tiempo, TotalBMD, Age, BMI, FMI, FFMI from df")
head(df)
m2 <- lme(TotalBMD ~ tiempo + Age+ BMI + FMI + FFMI, random = ~ tiempo | ID, data=tra
summary(m2)
```

```
##### Análisis Longitudinal
##### Analisis de residuales
```

```
res <- residuals(m2)
```

```
fit <- fitted(m2)
```

```
#test 1
```

```

pred = predict(m2, test)
png("~/Dropbox/TFM/Final 1.0/figurasR/fig40resmuj.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
     cex.sub=0.5)
abline(0,1, col = "red")
# test

```

Análisis Longitudinal Masa ósea mujeres

```

png("~/Dropbox/TFM/Final 1.0/figurasR/fig22.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, TotalBMD from testF")
plot(df$tiempo, df$TotalBMD, xlab = "Tiempo", ylab="Densidad Masa Osea Total", main = "Densidad Masa Osea Total",
     type="n", xlim = c(0,70), ylim=c(0,2))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$TotalBMD[df$ID == i]), col = 2)
}

```

Análisis Longitudinal indice de masa corporal hombres

```

png("~/Dropbox/TFM/Final 1.0/figurasR/fig23.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, BMI from testM")
plot(df$tiempo, df$BMI, xlab = "Tiempo", ylab="Indice de Masa Corporal", main = "Indice de Masa Corporal",
     type="n", xlim = c(0,70), ylim=c(0,50))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$BMI[df$ID == i]), col = 4)
}

```

Análisis Longitudinal indice de masa corporal mujeres

```

png("~/Dropbox/TFM/Final 1.0/figurasR/fig24.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, BMI from testF")
plot(df$tiempo, df$BMI, xlab = "Tiempo", ylab="Indice de Masa Corporal", main = "Indice de Masa Corporal",
     type="n", xlim = c(0,70), ylim=c(0,50))
for(i in unique(df$ID)) {

```



```

    lines(lowess(df$tiempo[df$ID == i], df$BMI[df$ID == i]), col = 2)
}

##### Análisis Longitudinal masa grasa hombres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig25.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, FMI from testM")
plot(df$tiempo, df$FMI, xlab = "Tiempo", ylab="Indice de Masa Grasa", main = "Masculino",
     type="n", xlim = c(0,70), ylim=c(0,20))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$FMI[df$ID == i]), col = 4)
}

##### Análisis Longitudinal masa grasa mujeres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig26.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, FMI from testF")
plot(df$tiempo, df$FMI, xlab = "Tiempo", ylab="Indice de Masa Grasa", main = "Femenino",
     type="n", xlim = c(0,70), ylim=c(0,20))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$FMI[df$ID == i]), col = 2)
}

##### Análisis Longitudinal masa magra hombres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig27.png",
     width=1000,height=800,units="px",
     pointsize=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, FFMI from testM")
plot(df$tiempo, df$FFMI, xlab = "Tiempo", ylab="Indice de Masa Magra", main = "Masculino",
     type="n", xlim = c(0,70), ylim=c(0,30))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$FFMI[df$ID == i]), col = 4)
}

##### Análisis Longitudinal masa magra mujeres

png("~/Dropbox/TFM/Final 1.0/figurasR/fig28.png",

```

```

width=1000,height=800,units="px",
points=8,bg="white",res=150)

df<- sqldf("select ID, tiempo, FFMI from testF")
plot(df$tiempo, df$FFMI, xlab = "Tiempo", ylab="Indice de Masa Magra", main = "Feme
type="n", xlim = c(0,70), ylim=c(0,30))
for(i in unique(df$ID)) {
  lines(lowess(df$tiempo[df$ID == i], df$FFMI[df$ID == i]), col = 2)
}

##### Análisis Longitudinal
### Bases para regresión lineal
# Oseo

longitudinal_oseo<-sqldf("select ID, tiempo,
  Height, Weight, RAFg, RALg, LAFg, LALg,
  RLFg, RLLg, LLFg, LLLg,
  TFG, TLg,
  Age, sexo,
  TotalBMD
  from longitudinal")

##### Análisis Longitudinal óseo hombres

df <- sqldf("select * from longitudinal_oseo where sexo='hombre'")

df<- sqldf("select ID, tiempo, TotalBMD, Height, Weight, Age,
  RAFg, RALg, LAFg, LALg,
  RLFg, RLLg, LLFg, LLLg,
  TFG, TLg from df")

set.seed(123)
split.indices <- sample(nrow(df), nrow(df)*0.8, replace = F)
train <- df[split.indices, ]
test <- df[-split.indices, ]

head(df)
m2 <- lme(TotalBMD ~ tiempo + Age + Height + Weight + RAFg + RALg + LAFg + LALg + RLF
summary(m2)

# Análisis de residuales
res <- residuals(m2, type = "normalized")
fit <- fitted(m2)
png("~/Dropbox/TFM/Final 1.0/figurasR/fig40aresidualeshom.png",
width=1000,height=800,units="px",

```

```

        pointsize=8,bg="white",res=150)

par(mfrow = c(2, 2))
plot(res ~ fit, xlab = "Fitted values", ylab = "Residuals", main = "Residuals vs. fit")
abline(h = 0)
hist(res, main = "Histogram of residuals", xlab = "Residuals")
qqnorm(res)
qqline(res)

#test

pred = predict(m2, test)
rss = sum((train$TotalBMD - pred)^2)
rss
png("~/Dropbox/TFM/Final 1.0/figurasR/fig29a.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
     cex.sub=0.5)
abline(0,1, col = "red")
# test

##### Análisis Longitudinal óseo mujeres

df <- sqldf("select * from longitudinal_oseo where sexo='mujer'")

df<- sqldf("select ID, tiempo, TotalBMD, Height, Weight, Age,
              RAFg, RALg, LAFg, LALg,
              RLFg, RLLg, LLFg, LLLg,
              TFg, TLg from df")

set.seed(123)
split.indices <- sample(nrow(df), nrow(df)*0.8, replace = F)
train <- df[split.indices, ]
test <- df[-split.indices, ]

head(df)
m2 <- lme(TotalBMD ~ tiempo + Age + Height + Weight + RAFg + RALg + LAFg + LALg + RLFg + RLLg + LLFg + LLLg + TFg + TLg, data=df)
summary(m2)

```

```

# Análisis de residuales
res <- residuals(m2, type = "normalized")
fit <- fitted(m2)
png("~/Dropbox/TFM/Final 1.0/figurasR/fig40aresidualesmuj.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

par(mfrow = c(2, 2))
plot(res ~ fit, xlab = "Fitted values", ylab = "Residuals", main = "Residuals vs. fit")
abline(h = 0)
hist(res, main = "Histogram of residuals", xlab = "Residuals")
qqnorm(res)
qqline(res)

#test

pred = predict(m2, test)
rss = sum((train$TotalBMD - pred)^2)
rss
png("~/Dropbox/TFM/Final 1.0/figurasR/fig29b.png",
    width=1000,height=800,units="px",
    pointsize=8,bg="white",res=150)

plot(test$TotalBMD, pred, xlab="Datos originales", ylab="Datos estimados",
     cex.sub=0.5)

# test
abline(0,1, col = "red")

```

Apéndice C

Datos de NCBI/PUBMED

El siguiente código realiza una búsqueda en las bases de datos de NCBI/PUBMED según los criterios definidos. Este no debe ser ejecutado muchas veces ya que a mayor tráfico se corre el riesgo de ser baneado por demasiados accesos. La salida se muestra en el cuadro A.1.

Código modificado de (r-project, 2018).

```
library(rentrez)
library(pasteecs)
library(pander)

search_year <- function(year, term){
  query <- paste(term, "AND (", year, "[PDAT])")
  entrez_search(db="pubmed", term=query, retmax=0)$count
}

year <- 2000:2018
a <- "((((osteopenia OR osteoporosis) AND hiv)))"
b <- "((((sarcopenia) AND hiv)))"
c <- "((((lipodystrophy) OR (lipodistrofia)) AND hiv)))"
d <- "((((dexa) AND hiv)))"
busqueda <- data.frame(Fecha=year)
papers <- sapply(year, search_year, term=a, USE.NAMES=FALSE)
busqueda[,2] <- data.frame(papers)
papers <- sapply(year, search_year, term=b, USE.NAMES=FALSE)
busqueda[,3] <- data.frame(papers)
papers <- sapply(year, search_year, term=c, USE.NAMES=FALSE)
busqueda[,4] <- data.frame(papers)
papers <- sapply(year, search_year, term=d, USE.NAMES=FALSE)
busqueda[,5] <- data.frame(papers)
colnames(busqueda)<-c("Anio", "Osteoporosis", "Sarcopenia",
                     "Lipodistrofia", "dexa")
pandoc.table( busqueda
               , caption = "Datos bibliograficos de NCBI/PUBMED 2000-2018"
               , justify = 'ccccc'
               , round = 3
```

```

, decimal.mark = "."
, big.mark = ""
, emphasize.rownames = TRUE
, split.cells = 5
)

```

Cuadro C.1: Datos bibliograficos de NCBI/PUBMED
2000-2018

Anio	Osteoporosis	Sarcopenia	Lipodistrofia	dexa
2000	17	1	122	2
2001	33	1	170	6
2002	20	1	164	1
2003	36	1	215	11
2004	27	3	192	7
2005	35	0	204	10
2006	37	1	188	8
2007	33	1	189	8
2008	26	1	207	7
2009	43	2	167	10
2010	38	1	136	6
2011	66	4	143	11
2012	60	2	123	8
2013	75	6	110	8
2014	84	4	108	9
2015	65	5	84	6
2016	72	5	66	6
2017	65	5	49	5
2018	60	7	43	3

Bibliografía

Aachen, B., Batagelj, V. & Ferligoj, A. (2006). *Data science and clasification*. Springer-Verlag. Retrieved from <https://zdoc.site/studies-in-classification-data-analysis-and.html>

AIDS Information, C. for. (2002). Hoja informativa sobre la lipodistrofia. Retrieved from <http://www.centerforaids.org/pdfs/liposp.pdf>

Aldás, J. & Uriel, E. (2017). *Análisis multivariante aplicado con R*. Madrid: Paraninfo.

Alvero-Cruz, J.R., Álvarez Carnero, E., Fernández-García, J.C., Barrera Expósito, J., Carrillo de Albornoz Gil, M. & Sardinha, L.B. (2010). Validez de los índices de masa corporal y de masa grasa como indicadores de sobrepeso en adolescentes españoles: Estudio Escola. *Medicina Clínica*, 135, 8–14. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S002577531000148X>

Arnau, J. & Bono, R. (2008). Estudios longitudinales. Modelos de diseño y análisis. *Escritos de psicología*, 2, 32–41.

Arnau-Gras, J. (2007). Estudios longitudinales de medidas repetidas. Modelos de diseño y de análisis. *Avances en medición*, 9–26.

Baumgartner, R.N., Koehler, K.M., Gallagher, D., Romero, L., Heymsfield, S.B., Ross, R.R., Garry, P.J. & Lindeman, R.D. (1998). Epidemiology of sarcopenia among the elderly in New Mexico. *American Journal of Epidemiology*, 147, 755–763.

Bresler, R. & Gil, R. (2011). Clasificación kNN de documentos usando GPU. *Computación y Sistemas*, 15, 63–77.

Burger, S. (2018). *Introduction to Machine Learning with R*. Rigorous mathematical modeling. Boston: O'Reilly.

Carrero, A. & Ryan, P. (2009). Tenofovir y enfermedad ósea en pacientes infectados por el VIH. *Rev Esp Enferm Metab Oseas*, 18, 56–62.

Cayuela, L. (2010). Modelos lineales mixtos en R. EcoLab, Centro Andaluz de Medio Ambiente, Universidad de Granada.

Cazanave, C., Dupon, M., Lavignolle-Aurillac, V., Barthe, N., Lawson-Ayayi, S., Mehsen, N., Mercié, P., Morlat, P., Thiébaud, R., Dabis, F. & Groupe d'Epidémiologie Clinique du SIDA en Aquitaine. (2008). Reduced bone mineral density in HIV-infected patients: Prevalence and associated factors. *AIDS (London, England)*, 22, 395–402.

Clay, P.G., Voss, L.E., Williams, C. & Daume, E.C. (2008). Valid treatment options for osteoporosis and osteopenia in HIV-infected persons. *The Annals of Pharmacotherapy*, 42,

Closas, A., Arriola, E., Kuc, C., Amarilla, M. & Jovanovich, E. (2013). Análisis multivariante, conceptos y aplicaciones en psicología educativa y psicometría. *Enfoques*, XXV, 65–92.

Corra Morales, J.C. & Salazar, C. (2016). *Introducción a los modelos mixtos*. Medellín: Universidad Nacional de Colombia.

Crabtree, N.J., Arabi, A., Bachrach, L.K., Fewtrell, M., El-Hajj Fuleihan, G., Kecske-methy, H.H., Jaworski, M., Gordon, C.M. & International Society for Clinical Densitometry. (2014). Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: The revised 2013 ISCD Pediatric Official Positions. *Journal of Clinical Densitometry: The Official Journal of the International Society for Clinical Densitometry*, 17, 225–242.

Delgado-Rodríguez, M. & Llorca, J. (2004). Estudios longitudinales: Concepto y particularidades. *Revista española de salud pública*, 78. Retrieved from http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-57272004000200002

Doradea, M. & Rodríguez, K. (2017). *Prevalencia de lipodistrofia diagnosticada a través de pliegues cutáneos y circunferencias, en pacientes ambulatorios con VIH/SIDA, del hospital nacional “Angélica Vidal de Najarro”, San Bartolo, Ilopango; junio-julio 2017* (Licenciado en nutrición). Universidad del Salvador, San Salvador.

El-Maouche, D., Xu, X., Jr, J., Dobs, A. & T Brown, T. (2011). Prevalence of Low Bone Mineral Density in a Low-Income Inner-City Population. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, 26, 388–96.

Figueras-Balsells, M. (2009). *Análisis longitudinal de los factores asociados a la pérdida de densidad mineral ósea en pacientes infectados con VIH* (Master de Estadística e Investigación Operativa). Universidad Politécnica de Cataluña, Barcelona.

Finkelstein, J., Gala, P., Rochford, R., Glesby, M. & Mehta, S. (2015). HIV/AIDS and lipodystrophy: Implications for clinical management in resource-limited settings. *J. Int AIDS Soc.*, 18. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4297925/>

Gironés, J., Casas, J., Minguillón, J. & Caihuelas, R. (2017). *Minería de datos. Modelos y algoritmos*. barcelona: Editorial UOC.

González-Pérez, N. (2012). *Sarcopenia en pacientes VIH: Prevalencia y factores asociados* (Grado de doctor en medicina). Universidad Autónoma de Madrid, Madrid.

Gregson, C.L., Paggiosi, M.A., Crabtree, N., Steel, S.A., McCloskey, E., Duncan, E.L., Fan, B., Shepherd, J.A., Fraser, W.D., Smith, G.D. & Tobias, J.H. (2013). Analysis of Body Composition in Individuals With High Bone Mass Reveals a Marked Increase in Fat Mass in Women But Not Men. *The Journal of Clinical Endocrinology & Metabolism*, 98, 818–828. Retrieved from <https://academic.oup.com/jcem/article/98/2/818/2833372>

Juárez de Galíndez, M. & Balzarini, M. (2003). Modelación de la estructura de covarianza entre observaciones de un mismo individuo en modelos de curvas de crecimiento arbóreo. *Revista de la Sociedad Argentina de Estadística*, 7, 1–15.

Kelly, T.L., Wilson, K.E. & Heymsfield, S.B. (2009). Dual Energy X-Ray Absorptiometry Body Composition Reference Values from NHANES. *PLOS ONE*, 4, e7038. Retrieved

from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007038>

Lambrou, A., Papadopoulos, H. & Gammerman, A. (2013). Osteoporosis Risk Assessment with Well-Calibrated Probabilistic Outputs. In H. Papadopoulos, A.S. Andreou, L. Iliadis & I. Maglogiannis (Eds.), *Artificial Intelligence Applications and Innovations, Aiai 2013* (Vol. 412, pp. 432–441). Berlin: Springer-Verlag Berlin.

Landazurri, N. & Dávila, P. (2014). *Factores asociados a osteoporosis y osteopenia en pacientes con infección por VIH en la Unidad de Atención Integral de PVV del Hospital Enrique Garcés, Quito-Ecuador durante el período de enero de 2013 a junio de 2013*. (Médico Cirujano). Pontificia Universidad Católica del Ecuador, Quito, Ecuador.

Lapphra, K., Vanprapar, N., Phongsamart, W., Chearskul, P. & Chokeyhaibulkit, K. (2005). Dyslipidemia and Lipodystrophy in HIV-Infected Thai Children on Highly Active Antiretroviral Therapy (HAART). *J Med Assoc Thai*, 88, 956–966.

Lazcano-Ponce, E., Tamayo, J., Díaz, R., Burguete, A.I. & Salmerón, J. (2009). Tendencias de correlación para la densidad mineral ósea en mujeres mexicanas: Pruebas de predisposición familiar. *Salud Pública de México*, 51, s93–s99. Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0036-36342009000700012&lng=es&nrm=iso&tlng=en

Llorenc, B. (2011). Modelos lineales generalizados mixtos. Algunos casos prácticos. In *X Congreso galego de estadística e investigación de operaciones* (p. 6). Pontevedra.

Lorente Ramos, R., Azpeitia Armán, J., Arévalo Galeano, N., Muñoz Hernández, A., García Gómez, J. & Gredilla Molinero, J. (2012). Absorciometría con rayos X de doble energía. Fundamentos, metodología y aplicaciones clínicas. *Radiología*, 54, 410–423. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0033833811003729>

Nemerovsky, J. (2016). Sarcopenia. *Revista argentina de gerontología y geriatría*, 1, 28–33.

OMS. (2018). VIH/sida. *VIH/sida: Nota descriptiva*. Retrieved from <http://www.who.int/es/news-room/fact-sheets/detail/hiv-aids>

Oria, E. (2003). Factores preventivos y nutricionales de la osteoporosis. *An. Sist. Sanit. Navar.*, 26, 81–90.

Porbén, S. & Espinosa, A. (2003). Composición corporal. *Acta médica*, 11, 26–37. Retrieved from http://bvs.sld.cu/revistas/act/vol11_1_03/act05103.htm

r-project. (2018). Introduction: The NCBI, entrez and rentrez. Retrieved from https://cran.r-project.org/web/packages/rentrez/vignettes/rentrez_tutorial.html

Rastogi, R., Corriere, M. & Ferrucci, L. (2014). Age-related and disease-related muscle loss: The effect of diabetes, obesity, and other diseases. *Diabetes & endocrinology*, 2, 819–829.

Rexach, J.A.S. (2006). Consecuencias clínicas de la sarcopenia. *Nutr. Hosp.*, 5.

Rodríguez, T. (2013). *Masa ósea y masa grasa corporal en mujeres premenopáusicas sanas de Extremadura* (Tesis doctoral). Universidad de Extremadura, España.

Rubio, M., Salas-Salvadó, J., Barbany, M. & Moreno, B. (2007). Consenso SEEDO 2007 para la evaluación del sobrepeso y la obesidad y el establecimiento de criterios de

intervención terapéutica. *Rev. Esp. Obes.*, 7–48.

Salgado, I. (1992). *El análisis exploratorio de datos biológicos*. México: Marc ediciones.

Salud y Servicios Humanos de EU, D. de. (2017). La infección por el VIH y la lipodistrofia. *infoSIDA*. Retrieved from <https://infosida.nih.gov/understanding-hiv-aids/fact-sheets/22/61/la-infeccion-por-el-vih-y-la-lipodistrofia>

Sayers, A. & Tobias, J.H. (2010). Fat Mass Exerts a Greater Effect on Cortical Bone Mass in Girls than Boys. *The Journal of Clinical Endocrinology & Metabolism*, 95, 699–706. Retrieved from <https://academic.oup.com/jcem/article/95/2/699/2597010>

Torgo, L. (2017). *Data mining with R* (Segund.). Florida: Taylor & Francis Group.

Valle Gutiérrez, F.J. del & Gómez Rodrigo, F.J. (2001). Lipodistrofia y alteraciones metabólicas en pacientes infectados por el virus de la inmunodeficiencia humana tratados con inhibidores de la proteasa. *Clínica e Investigación en Arteriosclerosis*, 13, 254–259. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0214916801788092>

Vilar, A. (2003). Lipodistrofias: Bases moleculares y manifestaciones clínicas. *Endocrinol. Nutr.*, 133–144.

Yang, Y., Zhu, W.D.J. & Paton, N.I. (2004). Comparison of dual-energy X-ray absorptiometry machines for measuring fat distribution changes of HIV-associated lipodystrophy. *Antiviral Therapy*, 9, 771–778.

Zanchetta, J. & Talbot, J. (2001). *Osteoporosis: Fisiopatología, diagnóstico, prevención y tratamiento*. Buenos Aires: Editorial médica panamericana.

Zerbi, C. (2017). Osteosarcopenia: Valoración para la prevención de caídas y fracturas. Simposio, Córdoba, Argentina. Retrieved from <https://www.osteoporosis.org.ar/pdf/sio2017/sarcopenia.pdf>

Zumel, N. & Mount, J. (2014). *Practical data science with R*. New York: Manning Publications.