

Medidas que juegan un rol determinante para la obtención del porcentaje de grasa corporal

Cynthia Marie Rivera Sánchez

6 de mayo de 2021

Resumen

Los conceptos de porcentaje de grasa corporal e índice de masa corporal son conocidos por la mayoría de las personas. Lo que pocos saben es que el porcentaje de grasa corporal suele ser más efectivo que el índice de masa corporal. Es por esto, que el propósito de este estudio fue hallar el mejor modelo de regresión lineal para el porcentaje de grasa corporal. La hipótesis establecía, que tanto el peso como la circunferencia del abdomen iban a formar parte de este modelo. Para probar esta hipótesis se realizaron diferentes tipos de gráficos, tablas y pruebas t. Se halló la correlación y asociación entre las variables, para luego crear modelos teniendo eso en consideración. Se desarrollaron diferentes modelos y utilizando las pruebas t se encontró que el mejor modelo de regresión lineal era el Modelo 1, el cual contenía las variables Peso, Abdomen, Antebrazo y Muñeca. Fue posible llegar a esta conclusión y aceptar la hipótesis propuesta, ya que todos los p-valores de este modelo eran significativamente menor que 0.05 lo cual nos permitía decir que tenían un efecto en el mismo.

1. Introducción

El porcentaje de grasa corporal es una medida la cual se utiliza para conocer la proporción de grasa que está presente en el cuerpo del ser humano [5]. El mismo es mucho más efectivo que el índice de masa corporal, pues a diferencia de este que solo ofrece la cantidad de grasa que hay en el cuerpo, el porcentaje de grasa corporal provee una idea de cuánto en realidad es grasa y cuánto es músculo [5]. Es decir, que un resultado de obesidad utilizando el índice de masa corporal, puede resultar ser sobrepeso cuando se evalúa con el porcentaje de grasa corporal, o viceversa [2]. Hoy día existen múltiples maneras de calcular esta medida desde la densitometría hidrostática hasta calculadoras que se pueden hallar en internet. Sin embargo, no todos estos métodos involucran los mismos elementos y hay otros factores que se deben tomar en consideración a la hora de obtener el porcentaje de grasa corporal, lo cual puede crear cierta discrepancia en los resultados que se obtienen.

El propósito de este análisis estadístico era determinar cuál sería el mejor modelo de regresión lineal utilizando los datos provistos. De acuerdo con la información buscada, la hipótesis planteada establecía que el mejor modelo iba tener tanto la variable de Peso como la de Abdomen. Para evaluarla, se estarán realizando diferentes pruebas con las cuales se espera que al final podamos aceptar o rechazar la hipótesis propuesta para así obtener una idea más clara de cuáles son las medidas que verdaderamente tienen un impacto al momento de calcular el porcentaje de grasa corporal.

2. Metodología

Para determinar el mejor modelo de regresión lineal para el porcentaje de grasa corporal se utilizaron los datos de *Fitting Percentage of Body Fat to Simple Body Measurements* suministrados por el Dr. A. Garth Fisher [4]. Estos, fueron evaluados utilizando diferentes técnicas estadísticas empleando el programa R [6] y RStudio [7], y los paquetes "broom" [1], "ggfortify" [9], "ggplot2" [3] y "leaps" [8]. Lo primero que se hizo fue evaluar a través de una tabla y una gráfica de regresión lineal las correlaciones que existían entre las variables, ya que estas tienen un impacto en el desarrollo de modelos de regresión lineal. Luego, se crearon diferentes modelos usando distintas estrategias. Algunos de estos fueron creados añadiendo variables, mientras que otros se crearon retirándolas. Una vez se obtuvieron todos los posibles modelos de regresión lineal, se procedió a comparar los mismos realizando gráficas para evaluar residuos, haciendo pruebas t, calculando el criterio de información Akaike (AIC) y el criterio de información Bayesiano (BIC), y obteniendo resúmenes de los modelos para encontrar el ideal. En resumen, este análisis estadístico está compuesto de tablas y gráficas de regresión lineal, utilizadas para evaluar los datos proporcionados y comparar los modelos de regresión lineal creados. Cabe señalar que no todas las gráficas ni tablas van estar en el cuerpo de este documento, sino que se añadirán en la sección de Appendices, pues no componen una parte integral de la conclusión llegada en este estudio, aunque fueron utilizadas para llegar a la misma.

3. Resultados y análisis descriptivo

3.1. Tablas

AIC vs. BIC		
	AIC	BIC
Modelo 1	1423.471	1444.647
Modelo 2	1429.558	1454.264
Modelo 3	1420.225	1455.52

Cuadro 1: Tabla que compara el AIC y el BIC entre tres de los modelos de regresión lineal para determinar cuál es el mejor de ellos. El Modelo 1 contiene las variables Peso, Abdomen, Muñeca y Antebrazo. El Modelo 2 contiene las variables Edad, Cuello, Abdomen, Antebrazo y Muñeca. Finalmente, el Modelo 3 contiene las mismas variables que el Modelo 2, más Peso, Cadera y Muslo

Resumen del Modelo 1 de regresión lineal				
Residuos				
Mínimo	Primer Cuartil	Mediana	Tercer Cuartil	Máximo
-9.8002	-2.8728	-0.1545	2.8980	8.3845
Coeficientes				
	Estimado	Error Estándar	t-val	Pr(> t)
Intercepto	-31.29679	6.70886	-4.665	5.06e-06 ***
Abdomen	0.92137	0.05192	17.747	< 2e-16 ***
Peso	-0.12557	0.02292	-5.479	1.05e-07 ***
Muñeca	-1.39177	0.40991	-3.395	0.000799 ***
Antebrazo	0.44638	0.16822	2.654	0.008480 **
Códigos significantes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Error estándar residual: 4.021 con 247 grados de libertad				
R^2 múltiple: 0.7351, R^2 ajustado: 0.7308				
Prueba-F: 171.4 con 4 y 247 GL, p-valor: < 2.2e-16				

Cuadro 2: Tabla que provee un resumen del modelo de regresión lineal que luego del análisis de los datos se catalogó como el ideal. Este contiene las variables Peso, Abdomen, Muñeca y Antebrazo.

3.2. Gráficas

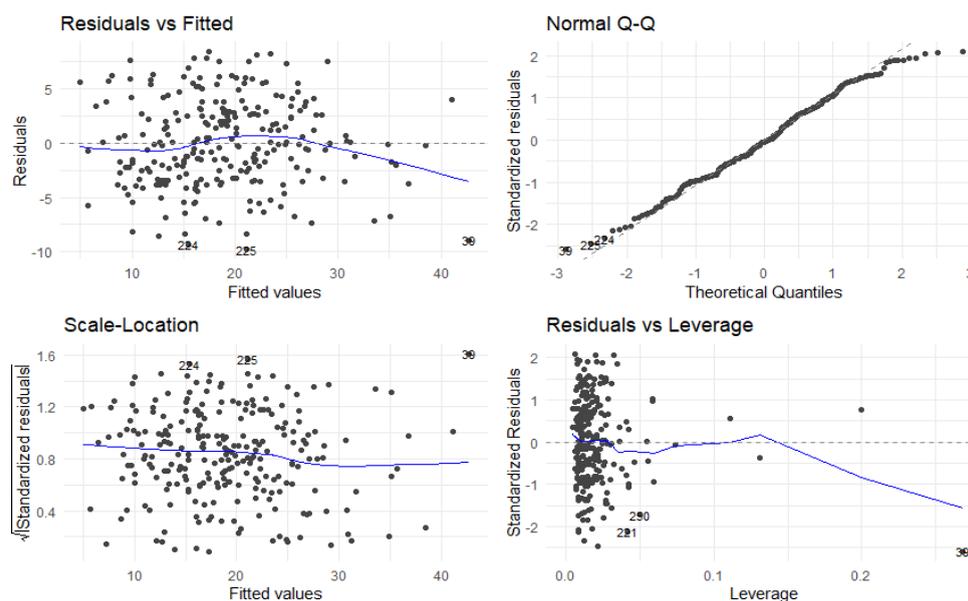


Figura 1: **Resumen de los residuos del Modelo 1.** Gráficas que despliegan la información sobre los residuos del Modelo 1, el cual contiene las variables Peso, Abdomen, Muñeca y Antebrazo. Estos no demuestran que haya ningún tipo de problema en el ajuste. No hay ninguna recta que presente una curva significativa para establecer algún problema, ni hay una desviación sistemática de línea por lo que se puede suponer normalidad. Sin embargo, sí se puede destacar que hay valores extremos que podrían estar influenciando los datos

4. Discusión e implicaciones

Dado que el problema presentado nos pedía encontrar el mejor modelo de regresión lineal para el porcentaje de grasa corporal, lo primero que se hizo fue evaluar la correlación y asociación entre las variables. Para esto se utilizó una tabla y una gráfica, las cuales se encuentran en los Apéndices A y B de este análisis estadístico (Cuadros 3 y 4, y Figura 2). En ellas se puede observar como hay variables que están altamente relacionadas, como por ejemplo Abdomen y Cadera que tienen una correlación de 0.874 o Muslo y Peso que tienen una correlación de 0.869. El saber esto es de suma importancia ya que dos variables que tengan una correlación alta nos podrían estar proveyendo la misma información, lo cual haría de nuestro modelo uno más complejo y no necesariamente certero.

Una vez se analizaron las correlaciones y asociaciones de los datos, se prosiguió a encontrar el modelo de regresión lineal ideal para estos. Se encontraron tres modelos que parecían funcionar. El Modelo 1 contenía las variables Peso, Abdomen, Muñeca y Antebrazo; el Modelo 2 contenía las variables Edad, Cuello, Abdomen, Antebrazo y Muñeca; y el Modelo 3, contenía las variables Edad, Peso, Cuello, Abdomen, Cadera, Muslo, Antebrazo y Muñeca. A estos modelos de regresión lineal se le hicieron varias pruebas para poder escoger el mejor de ellos. Primero, se comparó el Modelo 1 con el Modelo 3, donde se rechazó la hipótesis nula con un p-valor de 0.02785 la cual planteaba que el Modelo 1 proveía el mejor modelo de regresión lineal. Igualmente, se comparó el Modelo 2 con el Modelo 3 donde nuevamente se rechazó la hipótesis nula con un p-valor de 0.001982 la cual planteaba que el Modelo 2 proveía el mejor modelo de regresión lineal. Sin embargo, a pesar de que el Modelo 3 parecía ser el ideal, se decidió utilizar otros métodos para comparar el Modelo 1 con el Modelo 2 ya que estos no eran anidados, y de una vez comprobar si definitivamente el Modelo 3 era el mejor.

Se calcularon tanto el AIC como el BIC de los tres modelos los cuales están desplegados en el Cuadro 1. Dejándonos llevar por el AIC podemos observar que el menor lo tiene el Modelo 3, mientras que dejándonos llevar por el BIC el Modelo 1 tiene el menor, lo cual se debe a la penalización que tiene cada criterio. Una vez ya hecho este análisis, se dejó de considerar el Modelo 2 y se procedió a obtener el resumen de los modelos restantes para poder tomar una decisión (como quiera el resumen de este modelo y su gráfica de residuos se encuentran en los Apéndices B y C, Cuadro 7 y Figura 3).

Luego de realizar los resúmenes de los modelos se observó que el Modelo 3 contenía variables que en realidad aportaban poco o nada al mismo, como el fue el caso de la variable Cadera la cual tenía un p-valor de 0.14727 o la variable Edad que su p-valor fue 0.03876 lo cual está relativamente cerca de 0.05 (otros datos relacionados a este resumen se pueden observar en el Apéndice A, Cuadro 4, al igual que su gráfica de residuos en el Apéndice B, Figura 4). Cuando se compara el resumen del Modelo 3 con el resumen del Modelo 1 (Cuadro 2), se observa que los p-valores de las variables del Modelo 1 son más pequeños, por lo que tienen un mayor impacto en el modelo. Por otra parte, al evaluar sus residuos (Figura 1) no se encontró evidencia que permitiese decir que existiera algún problema en el ajuste. Además, el Modelo 1 es menos complejo que el Modelo 3, lo cual es algo que es importante tomar en consideración. Por estas razones, fue que se escogió el Modelo 1 como mejor modelo de regresión lineal.

5. Conclusión y recomendaciones

Luego de estudiar los resultados obtenidos por este análisis estadístico, la hipótesis puede ser aceptada. Según los hallazgos de este estudio, el mejor modelo de regresión lineal contiene las variables Peso, Abdomen, Muñeca y Antebrazo. A pesar de que las variables Peso y Abdomen tenían una correlación alta, basado en la información que se buscó estas no necesariamente proveen la misma información y es la razón por la que se encuentran ambas en el modelo. Por otra parte, el hecho de que se hubiesen considerados modelos con otras variables como la Edad muestra explícitamente que hay componentes que siempre van a tener un rol en el cálculo de esta medida. Es por esta razón que los expertos toman en consideración el género, la edad y la altura para crear unos rangos y lograr que el cálculo del porcentaje de grasa corporal sea uno más certero.

Sin embargo, como ya se mencionó anteriormente, existen múltiples maneras de obtener esta medida, por lo que a pesar de que el modelo presentado en este análisis estadístico funciona, no necesariamente significa que sea el único. Por otro lado, se pueden haber presentado posibles fuentes de error ya sea en la transcripción de los datos, lo cual pudo haber afectado los resultados o el hecho de que se tuviesen valores extremos los cuales podían estar influenciando los mismos. Es por esto que se recomienda que en algún futuro se realice esta prueba nuevamente. También, se recomienda incluir datos de mujeres para evaluar como el modelo se comporta, es decir, para averiguar si las medidas que se encontraron que tenían un rol importante cuando se calcula el porcentaje de grasa corporal en hombres aún lo siguen teniendo cuando se cambia el género o si otras medidas tendrían un mayor impacto.

Referencias

- [1] David Robinson, Alex Hayes and Simon Couch (2021). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.6. <https://CRAN.R-project.org/package=broom>
- [2] Goonasegaran AR, Nabila FN, Shuhada NS. Comparison of the effectiveness of body mass index and body fat percentage in defining body composition. Singapore Med J. 2012 Jun;53(6):403-8. PMID: 22711041.
- [3] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- [4] Johnson, R (1996) Fitting Percentage of Body Fat to Simple Body Measurements. Journal of Statistics Education, vol 4 No. 1. <http://www.amstat.org/publications/jse/v4n1/datasets.johnson.html> .
- [5] LBCCD (2019). "Porcentaje de grasa corporal: ¿Qué es y cómo calcularlo?", <https://www.sport.es/laborsadelcorredor/porcentaje-grasa-corporal-calcularlo/>
- [6] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [7] RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- [8] Thomas Lumley based on Fortran code by Alan Miller (2020). leaps: Regression Subset Selection. R package version 3.1. <https://CRAN.R-project.org/package=leaps>
- [9] Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages."The R Journal 8.2 (2016): 478-489.

Appendices

A. Tablas

Resumen de las correlaciones entre las variables							
	Grasa Corporal	Edad	Peso	Altura	Cuello	Pecho	Abdomen
Grasa Corporal	1.000	0.289	0.613	-0.025	0.491	0.703	0.814
Edad	0.289	1.000	-0.013	-0.245	0.114	0.176	0.230
Peso	0.613	-0.013	1.000	0.487	0.831	0.894	0.888
Altura	-0.025	-0.245	0.487	1.000	0.321	0.227	0.190
Cuello	0.491	0.114	0.831	0.321	1.000	0.785	0.754
Pecho	0.703	0.176	0.894	0.227	0.785	1.000	0.916
Abdomen	0.814	0.230	0.888	0.190	0.754	0.916	1.000
Cadera	0.626	-0.050	0.941	0.372	0.735	0.829	0.874
Muslo	0.561	-0.200	0.869	0.339	0.696	0.730	0.767
Rodilla	0.508	0.018	0.853	0.501	0.672	0.719	0.737
Tobillo	0.267	-0.105	0.614	0.393	0.478	0.483	0.453
Bíceps	0.493	-0.041	0.800	0.319	0.731	0.728	0.685
Antebrazo	0.363	-0.085	0.630	0.322	0.624	0.580	0.503
Muñeca	0.348	0.214	0.730	0.398	0.745	0.660	0.620

Cuadro 3: Tabla que provee un resumen de las correlaciones existentes entre las variables. Entre estas se pueden observar varias correlaciones fuertes que hay que tomar en cuenta a la hora de realizar el modelos de regresión lineal

Continuación del resumen de las correlaciones entre las variables							
	Cadera	Muslo	Rodilla	Tobillo	Bíceps	Antebrazo	Muñeca
Grasa Corporal	0.626	0.561	0.508	0.267	0.493	0.363	0.348
Edad	-0.050	-0.200	0.018	-0.105	-0.041	-0.085	0.214
Peso	0.941	0.869	0.853	0.614	0.800	0.630	0.730
Altura	0.372	0.339	0.501	0.393	0.319	0.322	0.398
Cuello	0.735	0.696	0.672	0.478	0.731	0.624	0.745
Pecho	0.829	0.730	0.719	0.483	0.728	0.580	0.660
Abdomen	0.874	0.767	0.737	0.453	0.685	0.503	0.620
Cadera	1.000	0.896	0.823	0.558	0.739	0.545	0.630
Muslo	0.896	1.000	0.799	0.540	0.761	0.567	0.559
Rodilla	0.823	0.799	1.000	0.612	0.679	0.556	0.665
Tobillo	0.558	0.540	0.612	1.000	0.485	0.419	0.566
Bíceps	0.739	0.761	0.679	0.485	1.000	0.678	0.632
Antebrazo	0.545	0.567	0.556	0.419	0.678	1.000	0.586
Muñeca	0.630	0.559	0.665	0.566	0.632	0.586	1.000

Cuadro 4: Tabla que provee una continuación del resumen presentado en el Cuadro 3 de las correlaciones existentes entre las variables. Entre estas se pueden observar varias correlaciones fuertes que hay que tomar en cuenta a la hora de realizar el modelos de regresión lineal

Resumen de posibles modelos							
	Intercepto	Edad	Peso	Altura	Cuello	Pecho	Abdomen
1	Cierto	Falso	Falso	Falso	Falso	Falso	Cierto
2	Cierto	Falso	Cierto	Falso	Falso	Falso	Cierto
3	Cierto	Falso	Cierto	Falso	Falso	Falso	Cierto
4	Cierto	Falso	Cierto	Falso	Falso	Falso	Cierto
5	Cierto	Falso	Cierto	Falso	Cierto	Falso	Cierto
6	Cierto	Cierto	Cierto	Falso	Falso	Falso	Cierto
7	Cierto	Cierto	Cierto	Falso	Cierto	Falso	Cierto
8	Cierto	Cierto	Cierto	Falso	Cierto	Falso	Cierto

Cuadro 5: Tabla que provee un resumen de los posibles modelos de regresión lineal que pueden ser desarrollados con los datos provistos

Continuación del resumen de posibles modelos							
	Cadera	Muslo	Rodilla	Tobillo	Bíceps	Antebrazo	Muñeca
1	Falso	Falso	Falso	Falso	Falso	Falso	Falso
2	Falso	Falso	Falso	Falso	Falso	Falso	Falso
3	Falso	Falso	Falso	Falso	Falso	Falso	Cierto
4	Falso	Falso	Falso	Falso	Falso	Cierto	Cierto
5	Falso	Falso	Falso	Falso	Falso	Cierto	Cierto
6	Falso	Cierto	Falso	Falso	Falso	Cierto	Cierto
7	Falso	Cierto	Falso	Falso	Falso	Cierto	Cierto
8	Cierto	Cierto	Falso	Falso	Falso	Cierto	Cierto

Cuadro 6: Tabla que provee una continuación del resumen de los posibles modelos de regresión lineal que pueden ser desarrollados con los datos provistos

Resumen del Modelo 2 de regresión lineal				
Residuos				
Mínimo	Primer Cuartil	Mediana	Tercer Cuartil	Máximo
-11.6196	-2.8049	-0.1558	2.6924	10.1805
Coeficientes				
	Estimado	Error Estándar	t-val	Pr(> t)
Intercepto	-6.78745	5.38359	-1.261	0.208588
Edad	0.08671	0.02214	3.916	0.000117 ***
Cuello	-0.55595	0.19919	-2.791	0.005666 **
Abdomen	0.72146	0.03745	19.263	1.2e-16 ***
Antebrazo	0.47392	0.17345	2.732	0.006745 **
Muñeca	-2.05184	0.43832	-4.681	4.72e-06 ***
Códigos significantes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Error estándar residual: 4.062 con 246 grados de libertad				
R^2 múltiple: 0.7308, R^2 ajustado: 0.7253				
Prueba-F: 133.5 con 5 y 246 GL, p-valor: < 2.2e-16				

Cuadro 7: Tabla que provee un resumen de uno de los modelos de regresión lineal creados. Este contiene las variables Edad, Cuello, Abdomen, Antebrazo y Muñeca.

Resumen del Modelo 3 de regresión lineal				
Residuos				
Mínimo	Primer Cuartil	Mediana	Tercer Cuartil	Máximo
-10.0574	-2.7411	-0.1912	2.6929	9.4977
Coeficientes				
	Estimado	Error Estándar	t-val	Pr(> t)
Intercepto	-20.06213	10.84654	-1.850	0.06558 .
Edad	0.05922	0.02850	2.078	0.03876 *
Peso	-0.08414	0.03695	-2.277	0.02366 *
Cuello	-0.43189	0.20799	-2.077	0.03889 *
Abdomen	0.87721	0.06661	13.170	1.2e-16 ***
Cadera	-0.18641	0.12821	-1.454	0.14727
Muslo	0.28644	0.11949	2.397	0.01727 *
Antebrazo	0.48255	0.17251	2.797	0.00557 **
Muñeca	-1.40487	0.47167	-2.978	0.00319 **
Códigos significantes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Error estándar residual: 3.965 con 243 grados de libertad				
R^2 múltiple: 0.7467, R^2 ajustado: 0.7383				
Prueba-F: 89.53 con 8 y 243 GL, p-valor: < 2.2e-16				

Cuadro 8: Tabla que provee un resumen de uno de los modelos de regresión lineal creados. Este contiene las variables Edad, Peso, Cuello, Abdomen, Cadera, Muslo, Antebrazo y Muñeca.

B. Gráficas

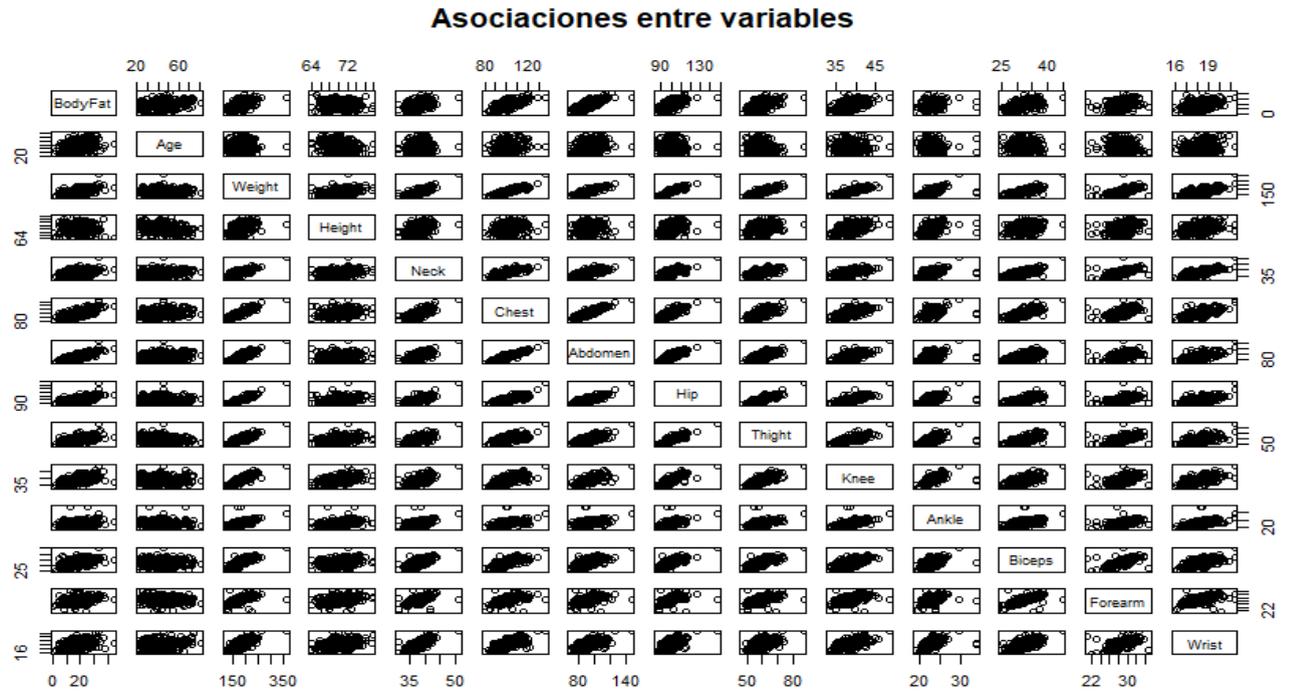


Figura 2: Gráficas que despliegan las diferentes asociaciones que existen entre las variables. Estas en conjunto con la información de los Cuadros 3 y 4 ayudan a tomar una decisión importante con respecto al modelo de regresión lineal

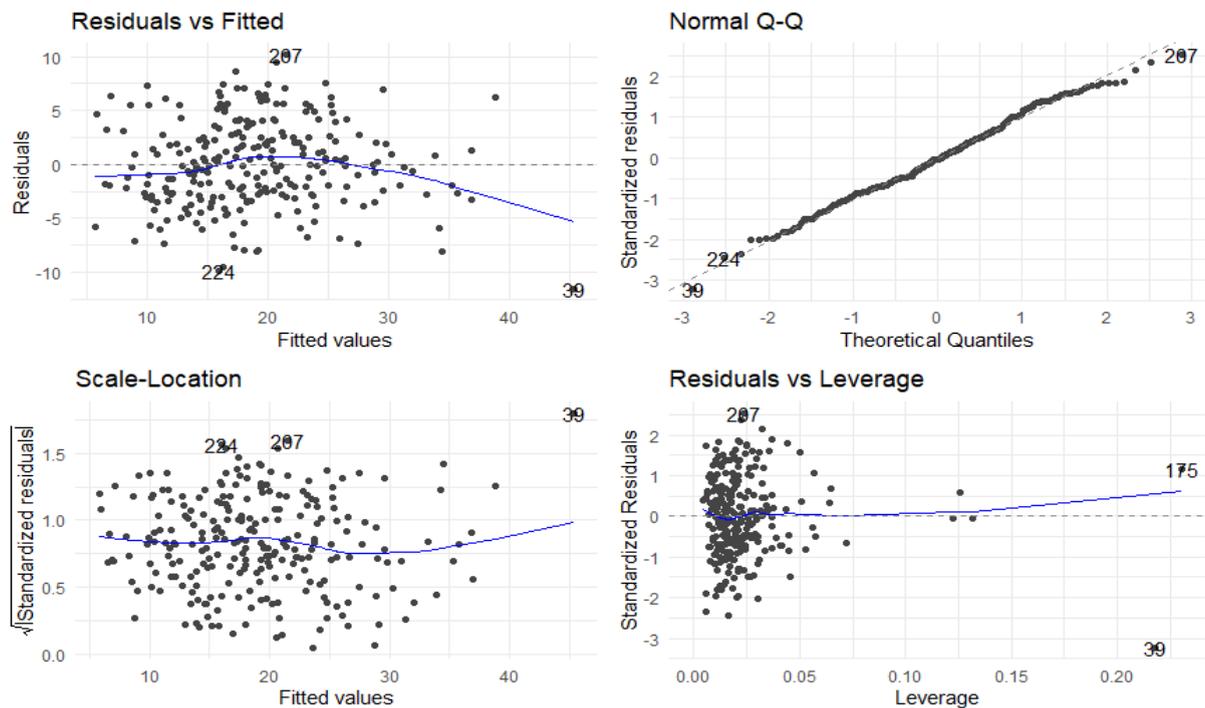


Figura 3: **Resumen de los residuos del Modelo 2.** Gráficas que despliegan la información sobre los residuos del Modelo 2, el cual contiene las variables Edad, Cuello, Abdomen, Antebrazo y Muñeca. Estos no demuestran que haya ningún tipo de problema en el ajuste. No hay ninguna recta que presente una curva significativa para establecer algún problema, ni hay una desviación sistemática de línea por lo que se puede suponer normalidad. Sin embargo, sí se puede destacar que hay valores extremos que podrían estar influenciando los datos

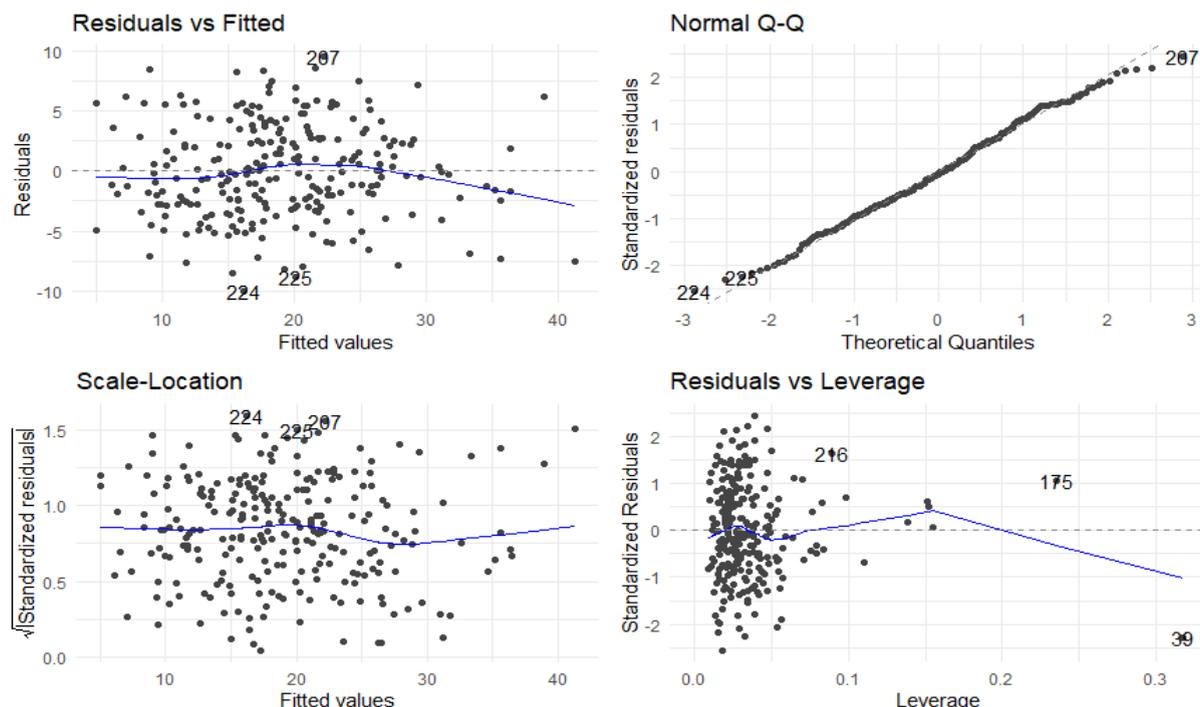


Figura 4: **Resumen de los residuos del Modelo 3.** Gráficas que despliegan la información sobre los residuos del Modelo 2, el cual contiene las variables Edad, Peso, Cuello, Abdomen, Cadera, Muslo, Antebrazo y Muñeca. Estos no demuestran que haya ningún tipo de problema en el ajuste. No hay ninguna recta que presente una curva significativa para establecer algún problema, ni hay una desviación sistemática de línea por lo que se puede suponer normalidad. Sin embargo, sí se puede destacar que hay valores extremos que podrían estar influenciando los datos

C. Código utilizado

```
#MATE5002 - Proyecto 2

#Primero, instalar todos los paquetes que van a utilizarse
install.packages("tidyverse")
install.packages("PASWR2")
install.packages("MASS")
install.packages("ggplot2")
install.packages("ggfortify")
install.packages("broom")
install.packages("leaps")

#Incluir las librerías que se van a utilizar
library(tidyr)
library(ggplot2)
library(ggfortify)
library(broom)
library(leaps)

#Agregar el archivo que va a utilizarse y evaluar el mismo
Body_Fat <- read.csv(file.choose())
class(Body_Fat)
dim(Body_Fat)
colnames(Body_Fat)
attach(Body_Fat)

#Evaluar posibles asociaciones entre las variables
pairs(Body_Fat[,-1], main = "Asociaciones entre variables")
```

```
round(cor(Body.Fat[,-1]), 3)
```

```
#Crear posibles modelos de regresión lineal añadiendo datos  
bodyfat.mod0 <- lm(BodyFat~1)  
bodyfat.mod0
```

```
#Utilizando la función step  
step(bodyfat.mod0, .~. + Age + Weight + Height + Neck + Chest + Abdomen + Hip + Thight +  
Knee + Ankle + Biceps + Forearm + Wrist)
```

```
#Manualmente  
add1(bodyfat.mod0, .~. + Age + Weight + Height + Neck + Chest + Abdomen + Hip + Thight +  
Knee + Ankle + Biceps + Forearm + Wrist)
```

```
bodyfat.mod6 j- lm(BodyFat ~ Abdomen)  
add1(bodyfat.mod6, .~. + Age + Weight + Height + Neck + Chest + Hip + Thight + Knee +  
Ankle + Biceps + Forearm + Wrist)
```

```
bodyfat.mod26 <- lm(BodyFat ~ Abdomen + Weight)  
add1(bodyfat.mod26, .~. + Age + Height + Neck + Chest + Hip + Thight + Knee + Ankle +  
Biceps + Forearm + Wrist)
```

```
bodyfat.mod2613 <- lm(BodyFat ~ Abdomen + Weight + Wrist)  
add1(bodyfat.mod2613, .~. + Age + Height + Neck + Chest + Hip + Thight + Knee + Ankle +  
Biceps + Forearm)
```

```
bodyfat.mod261213 <- lm(BodyFat ~ Abdomen + Weight + Wrist + Forearm)  
add1(bodyfat.mod261213, .~. + Age + Height + Neck + Chest + Hip + Thight + Knee + Ankle +  
Biceps)
```

```
#Crear posibles modelos de regresión lineal removiendo datos  
bodyfat.mod12345678910111213 <- lm(BodyFat ~ Age + Weight + Height + Neck + Chest +  
Abdomen + Hip + Thight + Knee + Ankle + Biceps + Forearm + Wrist)
```

```
#Utilizando la función step  
step(bodyfat.mod12345678910111213, .~. - Age - Weight - Height - Neck - Chest - Abdomen -  
Hip - Thight - Knee - Ankle - Biceps - Forearm - Wrist)
```

```
#Manualmente  
drop1(bodyfat.mod12345678910111213)
```

```
bodyfat.mod1234567810111213 <- lm(BodyFat ~ Age + Weight + Height + Neck +  
Chest + Abdomen + Hip + Thight + Ankle + Biceps + Forearm + Wrist)  
drop1(bodyfat.mod1234567810111213)
```

```
bodyfat.mod123467810111213 <- lm(BodyFat ~ Age + Weight + Height + Neck +  
Abdomen + Hip + Thight + Ankle + Biceps + Forearm + Wrist)  
drop1(bodyfat.mod123467810111213)
```

```
bodyfat.mod12467810111213 <- lm(BodyFat ~ Age + Weight + Neck + Abdomen + Hip +
```

```
Thight + Ankle + Biceps + Forearm + Wrist)
drop1(bodyfat.mod12467810111213)
```

```
bodyfat.mod124678111213 <- lm(BodyFat ~ Age + Weight + Neck + Abdomen + Hip +
Thight + Biceps + Forearm + Wrist)
drop1(bodyfat.mod124678111213)
```

```
bodyfat.mod1246781213 <- lm(BodyFat ~ Age + Weight + Neck + Abdomen + Hip +
Thight + Forearm + Wrist)
drop1(bodyfat.mod1246781213)
```

```
bodyfat.mod124681213 <- lm(BodyFat ~ Age + Weight + Neck + Abdomen + Thight +
Forearm + Wrist)
drop1(bodyfat.mod124681213)
```

```
bodyfat.mod12681213 <- lm(BodyFat ~ Age + Weight + Abdomen + Thight + Forearm +
Wrist)
drop1(bodyfat.mod12681213)
```

```
bodyfat.mod1261213 <- lm(BodyFat ~ Age + Weight + Abdomen + Forearm + Wrist)
drop1(bodyfat.mod1261213)
```

```
bodyfat.mod261213.drop <- lm(BodyFat ~ Weight + Abdomen + Forearm + Wrist)
drop1(bodyfat.mod261213.drop)
```

```
#Comparar los modelos
anova(bodyfat.mod261213, bodyfat.mod12345678910111213)
```

```
anova(bodyfat.mod261213.drop, bodyfat.mod1261213)
```

```
anova(bodyfat.mod261213.drop, bodyfat.mod12681213)
```

```
anova(bodyfat.mod261213.drop, bodyfat.mod124681213)
```

```
anova(bodyfat.mod261213.drop, bodyfat.mod1246781213)
```

```
anova(bodyfat.mod0, bodyfat.mod12345678910111213)
```

```
anova(bodyfat.mod0, bodyfat.mod261213)
```

```
#Evaluar modelos
tidy(bodyfat.mod261213)
```

```
tidy(bodyfat.mod12345678910111213)
```

```
bodyfat.mod1461213 |> lm(BodyFat ~ Age + Neck + Abdomen + Forearm + Wrist)
tidy(bodyfat.mod1461213)
```

```
tidy(bodyfat.mod1246781213)
```

```
#Ver el summary de lo que se piensa ser el modelo ideal  
summary(bodyfat.mod261213)
```

```
summary(bodyfat.mod1461213)
```

```
summary(bodyfat.mod1246781213)
```

```
#Hacer las gráficas de los residuos  
autoplot(bodyfat.mod261213, ncol = 2, label.size = 3) + theme_minimal()
```

```
autoplot(bodyfat.mod1461213) + theme_minimal()
```

```
autoplot(bodyfat.mod1246781213) + theme_minimal()
```

```
#Asegurándonos de haber seleccionado el modelo indicado  
bodyfat.leaps <- leaps::regsubsets(x = Body_Fat[,3:15], y = BodyFat)  
bodyfat.leapssum <- summary(bodyfat.leaps)  
bodyfat.leapssum$which
```

```
Evaluar los AIC  
AIC(bodyfat.mod261213)  
BIC(bodyfat.mod261213)
```

```
AIC(bodyfat.mod1461213)  
BIC(bodyfat.mod1461213)
```

```
AIC(bodyfat.mod1246781213)  
BIC(bodyfat.mod1246781213)
```

```
#Prueba anova final  
anova(bodyfat.mod261213, bodyfat.mod1246781213)
```

```
anova(bodyfat.mod1461213, bodyfat.mod1246781213)
```