

## Investigación

# Estudio estadístico sobre los factores de riesgo responsables de acontecimientos coronarios.

## Statistical study about the risk factors responsible of coronary events.

Alejandro Fernández-Jiménez, Álvaro Fernández Jiménez,  
Abilio José Fernández Vicente

Revista de Investigación



Volumen XIII, Número 1, pp. 04–24, ISSN 2174-0410  
Recepción: 16 Ago'22; Aceptación: 10 Ene'23

1 de abril de 2023

### Resumen

En este artículo realizamos un estudio estadístico para discernir la posible relación o no de varios factores con el riesgo de desarrollar enfermedad cardiovascular medido según el test REGICOR. Una vez decididos los factores de riesgo involucrados, utilizamos una regresión lineal para medir la relevancia de cada uno y dar una previsión del valor del riesgo cardiovascular de cada paciente. Además, con estos factores relevantes, con una regresión logística obtenemos una regla de clasificación para riesgo bajo-moderado y riesgo alto-muy alto.

**Palabras Clave:** Regresión lineal, regresión logística, REGICOR, factores de riesgo cardiovascular.

### Abstract

In this paper we perform a statistical study to understand the possible relationship between several factors and the risk to develop a cardiovascular disease according to the REGICOR test. Once we determine which ones are the risk factors involved, we use a linear regression to measure the relevance of each one of them and get an estimation of the cardiovascular risk of each patient. With the same factors, we perform a logistic regression to obtain a classification rule for low-moderate risk and high-very high risk.

**Keywords:** Linear regression, logistic regression, REGICOR, cardiovascular risk factors.

## 1. Introducción

**Riesgo cardiovascular.**

Las enfermedades cardiovasculares constituyen un grupo patológico de interés común por ser la principal causa de muerte en nuestro medio, el elevado grado de discapacidad y el alto gasto sanitario que generan. Estas son responsables de más del 30 % de las muertes en todo el mundo y su incidencia va en aumento [24].

La enfermedad cardiovascular isquémica (ECVI) se caracteriza por presentar una etiología multifactorial en la que intervienen factores de riesgo con diferente grado de penetración. En algunas ocasiones, dichos factores se presentan asociados, pudiéndose potenciar entre sí. Se entiende por factor de riesgo una característica biológica, hábito o enfermedad que permite identificar un grupo de personas con mayor probabilidad de presentar una determinada enfermedad a lo largo del tiempo en comparación con el resto de la población. Por un lado, se pueden definir los factores de riesgo endógenos o marcadores de riesgo; estos factores no son modificables y dependen bien de la dotación genética del individuo, bien de características ambientales. Por otro lado, se pueden definir los factores de riesgo exógenos que se caracterizan por ser, a priori, modificables. Éste es el caso del tabaquismo, la diabetes mellitus (DM), la hipertensión arterial (HTA) y la dislipemia [22].

**Tabla I.** Análisis de prevalencia de factores de riesgo en el grupo total de pacientes y en controles y en el grupo total sesgado por sexo.

	Pacientes n (%)	Controles n (%)	Odds ratio	IC 95%	p
<b>Total</b>					
Tabaquismo	77 (25,0)	51 (16,6)	1,7	1,12-2,53	0,008
HTA	179 (58,1)	138 (45,0)	1,64	1,29-3,05	< 0,05
DM	75 (24,4)	38 (12,4)	2,2	1,45-3,42	< 0,001
Dislipemia	79 (25,6)	57(18,6)	1,48	1,00-2,17	0,046
<b>Varones</b>					
Tabaquismo	63 (39,1)	44 (27,3)	1,75	1,09-2,80	0,020
HTA	85 (52,8)	58 (36,0)	1,70	1,25-3,07	0,003
DM	39 (24,2)	18 (11,2)	1,37	1,37-4,64	0,002
Dislipemia	51 (31,7)	27 (16,8)	2,28	1,34-3,80	0,002
<b>Mujeres</b>					
Tabaquismo	14 (9,5)	7 (7,8)	-	-	ns
HTA	94 (63,9)	80 (54,8)	-	-	ns
DM	36 (24,5)	20 (13,7)	1,96	1,07-3,59	0,027
Dislipemia	28 (19,0)	30 (20,5)	-	-	ns

DM: diabetes mellitus; HTA: hipertensión arterial, IC 95%: intervalo de confianza del 95%; ns: no significativo.

Figura 1. Tabla resumen de [22] que muestra la relación de distintos valores de riesgo cardiovascular en el grupo total de la población y sesgado según sexos.

### Objetivo del estudio.

La finalidad de este estudio es comparar y analizar la relevancia de los distintos parámetros clínicos con respecto al riesgo cardiovascular. Una vez satisfecho este objetivo utilizaremos esta información para obtener predicciones por medio de una regresión lineal y reglas de clasifica-

**Tabla II.** Análisis de factores de riesgo en el grupo de pacientes y de controles sesgado por sexo.

	Varón n (%)	Mujer n (%)	Odds ratio	IC 95%	p
<b>Pacientes</b>					
Tabaquismo	63 (39,1)	14 (9,5)	0,166	0,09-0,31	< 0,0001
HTA	85 (52,8)	94 (63,9)	1,58	1,00-2,50	0,047
DM	39 (24,2)	36 (24,5)	-	-	ns
Dislipemia	51 (31,7)	28 (19,0)	0,507	0,30-0,86	< 0,011
<b>Controles</b>					
Tabaquismo	44 (27,3)	7 (4,8)	0,137	0,06-0,31	< 0,001
HTA	58 (36,0)	80 (54,8)	2,31	1,45-3,67	< 0,001
DM	18 (11,2)	20 (13,7)	-	-	ns
Dislipemia	27 (16,8)	30 (20,5)	-	-	ns

DM: diabetes mellitus; HTA: hipertensión arterial, IC 95%: intervalo de confianza del 95%; ns: no significativo.

Figura 2. Tabla resumen de [22] que muestra la relación de distintos valores de riesgo cardiovascular en el grupo total de la población y sesgado según sexos.

ción a través de regresión logística.

Para obtener la predicción, durante el proceso queremos evaluar cada uno de los pasos implementados. Esto lo hacemos para garantizar que los cálculos realizados se encuentran apoyados bajo un marco teórico fiable.

Con este contexto extraemos dos rectas de regresión. En la primera incluimos todos los parámetros  $\beta_j$  susceptibles de ser factor de riesgo de evento cardiovascular. En la segunda, tras estudiar las conclusiones obtenidas, usamos solo los parámetros rechazados con contundencia para la hipótesis de contraste  $H_0: \beta_j = 0$  en la primera regresión. Con esta segunda recta de regresión conseguimos un análisis más depurado con el que se obtienen conclusiones mucho más precisas.

Sobre la regla de clasificación, nuestro objetivo es ser capaces de predecir si el paciente pertenece al grupo de riesgo bajo o moderado (riesgo de evento cardiovascular inferior al 10%), población 0, o al grupo de riesgo alto o muy alto (riesgo de evento cardiovascular no inferior al 10%), población 1.

### Recogida de datos.

Para la elaboración de este análisis se han recogido una serie de datos clínicos sobre varios pacientes de las localidades de Almarza, El Cubo, Espejo, Matute, Portelarból, San Andrés, Segoviela, Soria, y Tera; todas ellas pertenecientes a la provincia de Soria. Soria es una provincia del norte de España, ubicada en el extremo oriental de la comunidad autónoma de Castilla y León. Se trata de una de las regiones más despobladas de Europa, además de una de las más envejecidas, lo que hace que aumente la necesidad de investigar sobre potenciales factores de riesgo. Según la clasificación climática de Köppen, Soria está dentro de un clima oceánico de tipo CFB, es una región predominantemente rural y la dieta prevalente es la dieta mediterránea [29, 6].

La muestra total es de  $n = 219$  pacientes. Los datos fueron tomados por el Doctor Abilio José Fernández Vicente. Entre los datos que han sido recogidos se incluyen el nivel de colesterol (total, LDL y HDL), triglicéridos y glucosa en sangre, la tensión arterial (sistólica, diastólica y tensión diferencial), filtrado glomerular, presencia/ausencia de albuminuria, tabaquismo, edad y sexo del paciente. Además se ha realizado un seguimiento del historial clínico de los distintos pacientes con el fin de tener en cuenta las distintas particularidades de cada uno de ellos.

Por último, para poder comparar, se ha obtenido el riesgo cardiovascular de cada paciente utilizando la escala REGICOR. La escala REGICOR estima el riesgo de acontecimiento coronario a los 10 años (angina o infarto de miocardio, con o sin síntomas, mortal o no) [20].

### **Estructura del artículo.**

Este trabajo se divide en dos partes, en las Secciones 2, 3 y 4 se realiza una regresión lineal para predecir el riesgo cardiovascular esperado del paciente, obteniendo una primera recta, una segunda más depurada y discutiendo la relevancia de los datos influyentes respectivamente. Más adelante, en la Sección 5 se obtiene una regla de clasificación para riesgo bajo-moderado o riesgo alto-muy alto utilizando una regresión logística. La regla obtenida presenta resultados muy satisfactorios en la identificación de pacientes con riesgo bajo-moderado pero no resulta tan fiable para riesgo alto-muy alto. Por último en la Sección 6 se exponen las principales conclusiones obtenidas y se proponen líneas de investigación con las que continuar este trabajo.

## **2. Primera recta de regresión lineal múltiple**

Esta sección se dedica a los primeros pasos para la obtención de una regla que estime el riesgo de accidentes cardiovasculares usando para ello algunos de los parámetros establecidos. Con este propósito, recurrimos a una recta de regresión. Una vez obtenida evaluaremos la fiabilidad que hemos conseguido con ella.

### **2.1. Discusión sobre la posible relación de los distintos parámetros con el riesgo cardiovascular**

En un primer análisis para confeccionar nuestra recta de regresión utilizaremos los parámetros  $\beta_1, \dots, \beta_{12}$ , todos ellos relacionados en la literatura científica con el riesgo cardiovascular.

Sospechamos que todos y cada uno de ellos son factores de riesgo cardiovascular por las razones que se enumeran a continuación.

- $\beta_1$ : Edad. La enfermedad cardiovascular aterosclerótica cursa con una frecuencia de aproximadamente 50% en personas de 30 años sin enfermedad conocida y su incidencia va en aumento, sobre todo en países industrializados, lo que está en relación con el envejecimiento [27]. El envejecimiento es un proceso fisiológico que conlleva una disminución progresiva de la capacidad elástica de las arterias [28]. Como consecuencia, esto implica un incremento de enfermedades cardiovasculares.
- $\beta_2$ : Colesterol total en sangre. Elevadas cantidades de colesterol favorece la aterosclerosis, un proceso inflamatorio crónico que se inicia desde la infancia y se desarrolla a lo largo de los años, de forma asintomática la mayor parte del tiempo. Se produce una retención, oxidación y modificación de lípidos en forma de estrías grasas en las paredes de las arterias que posteriormente evolucionan a placas fibrosas con engrosamiento de la pared de la arteria afectada, disminuyendo su diámetro interno o luz de manera crónica. Cuando estas placas se rompen causan trombosis y oclusión aguda parcial o total de la arteria afectada [16].

- $\beta_3$ : Colesterol HDL. Los niveles bajos de colesterol HDL se correlacionan con un riesgo elevado de desarrollar enfermedad aterosclerosa coronaria. La disminución de las HDL afecta al transporte reverso de colesterol, que es la vía metabólica responsable de la remoción del colesterol excedente de la células periféricas y su transporte hacia el hígado para reciclarlo o eliminarlo. Las HDL poseen además propiedades antiinflamatorias, antioxidativas, antiagregatorias, anticoagulantes y profibrinolíticas in vitro [25].
- $\beta_4$ : Colesterol LDL. Las lipoproteínas de baja densidad (LDL) son un agente causal de la enfermedad cardiovascular. Si las concentraciones séricas de colesterol LDL están elevadas de manera importante y persistente, éste logra penetrar las paredes de las arterias, se deposita y acumula entre las células, se liberan radicales libres de oxígeno, produciendo oxidación del LDL y liberando partículas proinflamatorias [28].
- $\beta_5$ : Triglicéridos. Existen evidencias de que las lipoproteínas ricas en triglicéridos, especialmente las de muy baja densidad (VLDL) son capaces de promover y desarrollar aterosclerosis [21].
- $\beta_6, \beta_7$ : Presión arterial sistólica, presión arterial diastólica. La hipertensión arterial, tanto sistólica como diastólica, produce daño endotelial aunque el mecanismo no es bien conocido; este hecho la convierte en factor de riesgo para la aterosclerosis [4]. Está demostrado que el aumento de 20 mmHg en la presión sistólica y de 10 mmHg en la presión diastólica por encima de los valores normales de 115/75 mmHg aumenta al doble el riesgo de muerte por enfermedad cardiovascular, independiente de otros factores de riesgo de ECV, para ambos sexos [19].
- $\beta_8$ : Nivel de glucosa en sangre. El control estricto de glucemia de forma prolongada en el tiempo puede reducir de forma significativa el riesgo de sufrir un evento cardiovascular mayor, concretamente en un 17%. Después de casi 10 años de seguimiento los cerca de 1800 pacientes con diabetes tipo 2, que habían sido asignados al azar a un control intensivo de la glucosa tenían entre 8 y 6 eventos cardiovasculares mayores menos por 1000 personas/año que los asignados a la terapia estándar, pero no se observó ninguna mejoría en la tasa de supervivencia global. La reducción de la mortalidad cardiovascular se situó en un 12% con terapia intensiva, si bien no presenta significación estadística con la terapia convencional [23].
- $\beta_9$ : Filtrado glomerular. Diversos estudios [1, 3, 11, 2] relacionan la insuficiencia renal crónica con el aumento de la morbimortalidad cardiovascular. El filtrado glomerular y la albuminuria son los principales indicadores de enfermedad renal crónica.
- $\beta_{10}$ : Albuminuria persistente. Se aplica lo mismo que lo anotado en el punto anterior. Tomaremos el valor 1 para pacientes cuyas niveles de albuminuria pertenezcan a la categoría A1, 2 para la categoría A2 y 3 para A3. Las categorías A1, A2 y A3 se refieren a una cantidad de entre 0 mg/g a 29 mg/g, 30 mg/g a 300 mg/g, y más de 300 mg/g respectivamente; donde, la medida que se está utilizando es miligramos de albumina por gramo de creatinina.
- $\beta_{11}$ : Diabetes. La diabetes mellitus aumenta de 2 a 5 veces el riesgo de enfermedad cardiovascular, siendo esta la primera causa de morbimortalidad en diabéticos [15]. Asignaremos el valor 1 a aquellos pacientes que sean considerados diabéticos y 0 al resto, sin hacer una discriminación más precisa.
- $\beta_{12}$ : Tabaquismo. El tabaquismo se ha considerado de forma clásica como uno de los principales factores de riesgo cardiovascular modificable [30]. El tabaquismo disminuye los niveles de HDL. En personas que fuman desde la adolescencia la mortalidad es 3 veces mayor que la de los no fumadores [13]. Asignaremos el valor 1 a aquellos pacientes que sean considerados fumadores habituales y 0 al resto, sin realizar una discriminación más

precisa sobre la cantidad de tabaco consumida o el número de años como fumadores habituales.

- $\beta_{13}$ : Sexo. La influencia del sexo en el riesgo cardiovascular está claramente demostrada dado que la incidencia de esta patología es mucho mayor en hombres menores de 50 años que en mujeres del mismo rango de edad, si bien, a partir de esta edad las incidencias tienden a igualarse [26]. No obstante, en este primer análisis no incluiremos este parámetro en nuestra recta de regresión, más adelante aportaremos explicaciones más elaboradas para justificar la decisión tomada.

Estos parámetros van a ser empleados para estimar el riesgo cardiovascular según el test REGICOR y con una regresión lineal.

Tras obtener la recta de regresión para estimar el riesgo de accidente cardiovascular del paciente, uno de los propósitos del análisis (quizás, incluso más interesante que la obtención de la propia recta de regresión) es decidir qué parámetros podrían no ser necesarios para la recta de regresión, es decir, que no aportan información relevante o que la información aportada ya es obtenida de manera implícita a través del resto de parámetros. Para este objetivo empleamos herramientas matemáticas para evaluar los datos aportados por la evidencia muestral. Como herramienta de apoyo, las conclusiones obtenidas de la evidencia muestral son analizadas desde el punto de vista médico para intentar decidir si los parámetros clasificados como no necesarios presentan un vínculo implícito con el resto de parámetros, si la relación de la variable "no necesaria" con la variable riesgo cardiovascular no es lineal o es inexistente o ambas circunstancias de forma simultánea.

Por este motivo en este primer análisis no hemos incluido la variable sexo. Ya existe mucha evidencia que sugiere la relación entre el sexo del paciente y el riesgo que este tiene de padecer un accidente cardiovascular. Por contra, nuestros datos para algunos parámetros varían mucho según el sexo del paciente. Por ejemplo, solo el 8,80 % de las mujeres en nuestra muestra son fumadoras en contraposición con el elevado 20,21 % en el caso de los hombres. En nuestros datos, solo el 11,20 % de las mujeres son diabéticas en contraposición con el 17,02 % de los hombres.

Los resultados vistos evidencian una relación entre las variables sexo y ser fumador habitual, y las variables sexo y padecer diabetes, especialmente fuerte en el primer caso. Al elaborar nuestro estudio, y decidir qué variables son relevantes, si incluyéramos el parámetro sexo podríamos encontrarnos en una situación indeseada. Si rechazamos el parámetro "ser fumador habitual" sería difícil decidir si el motivo del rechazo es debido a que ser fumador no afecta al riesgo cardiovascular o si, por el contrario, esta variable se encuentra expresada de forma implícita en la variable "sexo del paciente". Tenemos mucha evidencia de que la variable "sexo del paciente" es factor de riesgo [5, 10, 22, 17] y no necesitamos incluir esta variable en el primer análisis para decidir si es relevante para la recta de regresión, tenemos evidencia de que sí lo es.

## 2.2. Coeficientes de la recta de regresión

La estimación de los parámetros obtenida viene dada por la Tabla 1.

Lo que estamos haciendo es estimar el riesgo de acontecimiento coronario a los 10 años (angina o infarto de miocardio, con o sin síntomas, mortal o no) en tanto por 1,  $\hat{y}$ , creando para ello la siguiente regla

	Parámetro	Estimación
$\hat{\beta}_0$	Ordenada en el origen	-0,102876
$\hat{\beta}_1$	Edad	0,000620
$\hat{\beta}_2$	Colesterol total	0,0006419
$\hat{\beta}_3$	HDL	-0,0010525
$\hat{\beta}_4$	LDL	0,0003768
$\hat{\beta}_5$	Triglicéridos	$-2,7749 \cdot 10^{-5}$
$\hat{\beta}_6$	Presión Arterial Sistólica	0,0004922
$\hat{\beta}_7$	Presión Arterial Diastólica	$4,4558 \cdot 10^{-5}$
$\hat{\beta}_8$	Glucosa	$4,2925 \cdot 10^{-5}$
$\hat{\beta}_9$	Filtrado Glomerular	0,0146038
$\hat{\beta}_{10}$	Albuminuria	0,0004476
$\hat{\beta}_{11}$	Diabetes	0,0227179
$\hat{\beta}_{12}$	Tabaquismo	0,0201834

Tabla 1. Resumen de la información dada por la recta de regresión estimada.

$$\hat{y} = -0,102876 + 0,000620 \cdot \text{edad} + 0,0006419 \cdot \text{col} \\ - 0,0010525 \cdot \text{HDL} + 0,0003768 \cdot \text{LDL} - 2,7749 \cdot 10^{-5} \cdot \text{Trig} \\ + 0,0004922 \cdot \text{PAS} + 4,4558 \cdot 10^{-5} \cdot \text{PAD} + 4,2925 \cdot 10^{-5} \cdot \text{Glu} \\ + 0,0146038 \cdot \text{Fil Glo} + 0,0004476 \cdot \text{Cat Alb} \\ (+0,0227179 \text{ si diabetico}) \\ (+0,0201834 \text{ si fumador habitual}).$$

De esta forma, para la recta de regresión calculada, la estimación para  $\sigma^2$  es

$$\hat{\sigma}^2 = s_R^2 = \frac{1}{n-k-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \frac{1}{n-k-1} \sum_{i=1}^n e_i^2 = \frac{\text{RSS}}{n-k-1} = 0,000348595,$$

donde  $n = 219$  se corresponde con el número de datos,  $k = 12$  con el número de parámetros que se tienen en cuenta para el análisis y  $e_i := \hat{y}_i - y_i$  se corresponde con el residuo.

Las sumas de cuadrados son,

- (total)  $TSS := \sum_{i=1}^n (y_i - \bar{y})^2 = 0,22266849$ ,
- (residual)  $RSS := \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0,07181063$ ,
- (explicada por modelo)  $MSS := \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0,15085786$ .

Podemos definir el coeficiente de determinación como

$$R^2 := \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS} \leq 1.$$

Este parámetro resulta útil, pues, si  $R^2$  toma un valor próximo a 1, el modelo explica bien la realidad. En cambio, si toma un valor cercano a 0, no explica bien la realidad. En nuestro caso,  $R^2 = 0,6775$ . El p-valor del contraste global de la regresión,  $H_0: \beta_1 = \dots = \beta_{12} = 0$  es  $3,893 \cdot 10^{-44}$  (el estadístico es 36,06 y usamos la distribución  $F$  de Fisher-Snedecor  $F_{k;n-k-1}$  para contrastar, con  $n = 219$  y  $k = 12$ ), un valor muy bajo. Es razonable aceptar que la realidad se explica bien por esta recta de regresión.

Por último, para cada parámetro vamos a dar sus intervalos de confianza para  $\alpha = 5\%$  (ver Tabla 2) y a estudiar la hipótesis de contraste,  $H_0: \beta_j = 0$  (ver Tabla 3).

Intervalos de confianza con $\alpha = 5\%$			
	Parámetro	Extremo inferior	Extremo superior
$\beta_0$	Ordenada en el origen	-0,141498	-0,064254
$\beta_1$	Edad	0,000415	0,000826
$\beta_2$	Colesterol total	0,000221	0,001063
$\beta_3$	HDL	-0,001478	-0,000627
$\beta_4$	LDL	-0,000800	$4,6049 \cdot 10^{-5}$
$\beta_5$	Triglicéridos	-0,000112	$5,6259 \cdot 10^{-5}$
$\beta_6$	Presión Arterial Sistólica	0,000325	0,000659
$\beta_7$	Presión Arterial Diastólica	-0,000244	0,000333
$\beta_8$	Glucosa	-0,000129	0,000215
$\beta_9$	Filtrado Glomerular	-0,006714	0,035922
$\beta_{10}$	Albuminuria	-0,012435	0,013330
$\beta_{11}$	Diabetes	0,012548	0,032888
$\beta_{12}$	Tabaquismo	0,012551	0,027816

Tabla 2. Con una probabilidad del 95% cada uno de los parámetros  $\beta_j, j = 0, \dots, 12$  mencionados en esta tabla se encuentran en dicho rango de valores.

De la tabla 3 se deduce que no hay suficiente evidencia estadística para afirmar que las variables "Triglicéridos", "Presión Arterial Diastólica", "Glucosa" y "Albuminuria" aportan significativamente a la recta de regresión, esto podría deberse a que las variables no están relacionadas con el riesgo de evento cardiovascular o porque son descritas de manera implícita como combinación lineal de variables que sí aceptamos.

En un intento de buscar una respuesta a esta pregunta, cabría la posibilidad de que la variable "Glucosa" haya sido alterada artificialmente por la presencia de diabéticos (13,7% de nuestra muestra), los cuales se encuentran bajo medicación y presentan valores sobre el parámetro "Glucosa" que no reflejan la realidad del paciente. Sus valores "Glucosa" podrían haber sido radicalmente distintos si se hubieran tomado en otro momento. Sería interesante repetir el estudio sin incluir a los diabéticos y analizar tras ello los resultados obtenidos.

Con las variables "Filtrado Glomerular" y "LDL" podríamos tener nuestras dudas sobre si lo correcto es rechazar la hipótesis de contraste o aceptarla ya que sus valores se encuentran en un punto intermedio. No obstante, no existe ningún problema con rechazar ambas.

En el Apéndice A demostramos que las variables "Riesgo de evento cardiovascular" y Filtrado Glomerular son dependientes. Además, durante la discusión nos convencimos de que la relación entre ambas no era lineal, pues el riesgo incrementaba considerablemente al cambiar de G1 a G2 y volvía a disminuir significativamente al cambiar de G2 a G3a.

La variable "LDL" se encuentra altamente relacionada con las variables "Colesterol" y "HDL" ya que es la fracción más importante del colesterol "no HDL", por ello el valor del parámetro "LDL" puede estar relacionado de forma implícita con el de los parámetros "Colesterol" y "HDL", explicando esto un p-valor del 8,04%. Por ello, aunque este valor no sea muy elevado, por ser superior al 5% y estar relacionado con otros dos parámetros que sí vamos a incluir en la segunda recta de regresión, decidimos no incluir el "LDL" en el segundo análisis más depurado que realizaremos posteriormente.



Hipótesis de contraste $H_0: \beta_j = 0$		
	Parámetro	p-valor
$\beta_1$	Edad	0,000001 %
$\beta_2$	Colesterol total	0,298 %
$\beta_3$	HDL	0,00022 %
$\beta_4$	LDL	8,04 %
$\beta_5$	Triglicéridos	51,56 %
$\beta_6$	Presión Arterial Sistólica	0,000002 %
$\beta_7$	Presión Arterial Diastólica	76,09 %
$\beta_8$	Glucosa	62,32 %
$\beta_9$	Filtrado Glomerular	17,83 %
$\beta_{10}$	Albuminuria	94,54 %
$\beta_{11}$	Diabetes	0,00171 %
$\beta_{12}$	Tabaquismo	0,000045 %

Tabla 3. Realizamos hipótesis de contrastes para todos los parámetros (excepto ordenada en el origen, el hecho de que la ordenada en el origen pueda ser 0 o no, no es relevante ya que simplemente actúa como una corrección de la recta) y damos el p-valor. Cuanto más bajo sea el p-valor, con mayor contundencia podemos rechazarlo y cuanto más alto, existe menor evidencia estadística para poder rechazarla, siendo lo más natural en ese caso aceptar la hipótesis de contraste,  $\beta_j = 0$ .

### 2.3. Análisis de normalidad de los residuos.

Para varios de los resultados obtenidos durante esta Sección, hemos usado que los residuos se distribuían según una variable aleatoria normal de media 0. El objetivo de esta Subsección es comprobar que podemos suponer que esto es cierto. Suponemos que los residuos se distribuyen según una normal de media 0 y desviación típica 0,01811 (desviación típica obtenida si suponemos que los residuos tienen media 0).

Sea  $X \sim \mathcal{N}(0, 0,01811)$  queremos contrastar la hipótesis de rechazo

$H_0$ : Los datos recogidos sobre el colesterol en sangre es una muestra aleatoria de  $X$ .

Para contrastar esta hipótesis usaremos el test de Kolmogorov-Smirnov. Sean

$$F_X(t) = \mathbb{P}(\{X \leq t\}), \tag{1}$$

$$F_n(t) = \frac{1}{n} \# \{1 \leq j \leq n : x_j \leq t\}. \tag{2}$$

Se define el estadístico de Kolmogorov-Smirnov como

$$\begin{aligned} \delta_n &= \sup |F_n(t) - F_X(t)| \\ &= \max_{1 \leq j \leq n} \left[ \max \left\{ \left| F_X(x_j) - \frac{j-1}{n} \right|, \left| F_X(x_j) - \frac{j}{n} \right| \right\} \right]. \end{aligned} \tag{3}$$

Para nuestra muestra el estadístico de Kolmogorov-Smirnov toma el valor  $\delta_n = 0,0823548$  y  $\sqrt{n} \cdot \delta_n = \sqrt{219} \cdot 0,08235 = 1,2187 < 1,3581$  (siendo este último número el percentil  $\alpha = 5\%$  de la variable aleatoria  $Z$  de Kolmogorov-Smirnov). No rechazamos la hipótesis de contraste y por ende, podemos asumir que nuestros datos se distribuyen siguiendo una variable aleatoria normal. La Figura 3 pretende mostrar de manera visual la discusión que acabamos de tener, los residuos de nuestras predicciones se han distribuido de acuerdo a una variable aleatoria normal.

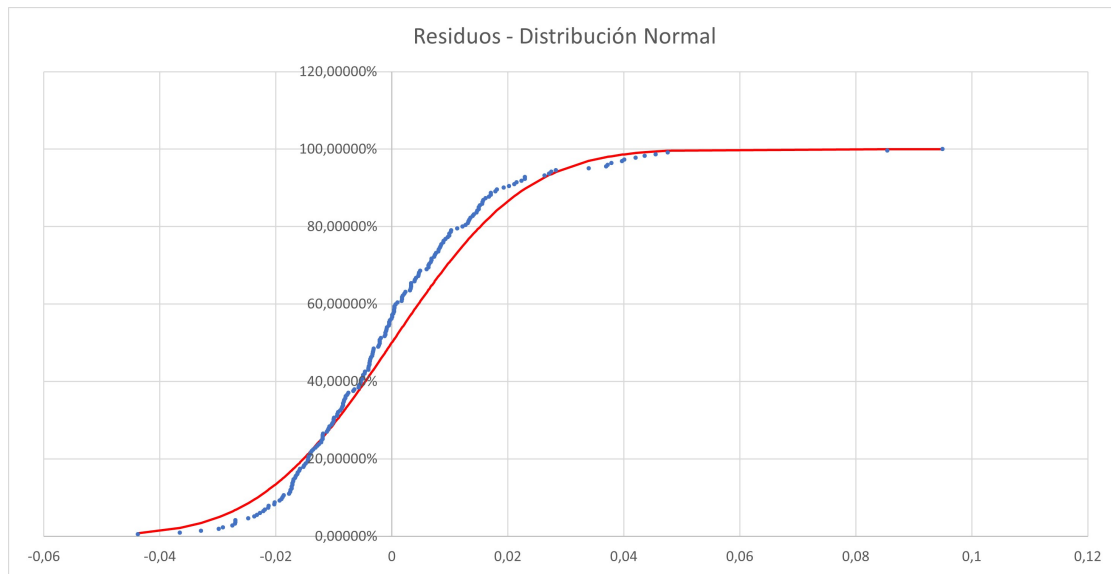


Figura 3. Comparación de las funciones  $F_n(t)$  (en azul) y  $F_X(t)$  (en rojo). Ambos trazos resultan parecidos sugiriendo que vienen motivados por una misma distribución y siendo esta hipótesis comprobada por el test de Kolmogorov-Smirnov.

### 3. Segunda recta de regresión, solo con los parámetros relevantes.

Tal y como apuntábamos anteriormente, en esta Sección vamos a repetir el análisis para obtener una recta de regresión más depurada. Con este propósito utilizamos solo los parámetros para los que hemos rechazado contundentemente la hipótesis de contraste  $H_0: \beta_j = 0$  en el primer análisis. Estos son "Edad", "Colesterol total", "HDL", "Presión Arterial Sistólica", "Diabetes" y "Tabaquismo". También podríamos incluir los parámetros "LDL" y "Filtrado Glomerular" pero ya hemos discutido por qué es mejor no hacerlo. Además, incorporamos el parámetro "Sexo", que no hemos tenido en cuenta en el anterior análisis porque podría habernos confundido sobre si el "Tabaquismo" (entre otros) era factor de riesgo o no, al estar ambos parámetros muy relacionados en nuestra muestra. No obstante, dado que existe gran cantidad de evidencia médica que afirma que el sexo del paciente y el riesgo cardiovascular están relacionados [5, 10, 22, 17], añadimos este parámetro en esta segunda regresión lineal más depurada.

La estimación de los parámetros que obtenemos en este caso viene dada por la Tabla 4.

En este caso, la regla que estamos dando (en forma de recta de regresión) para estimar el riesgo de evento cardiovascular es

$$\hat{y} = -0,080727 + 0,000497 \cdot \text{edad} + 0,000293 \cdot \text{col} \\ - 0,000672 \cdot \text{HDL} + 0,0003768 + 0,000506 \cdot \text{PAS} \\ (+0,023995 \text{ si diabetico}) \\ (+0,017748 \text{ si fumador habitual}) \\ (+0,007638 \text{ si el paciente es hombre}).$$

La estimación para  $\sigma^2$  es

$$\hat{\sigma}^2 = s_R^2 = \frac{1}{n-k-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \frac{1}{n-k-1} \sum_{i=1}^n e_i^2 = \frac{\text{RSS}}{n-k-1} = 0,00033854$$

y tenemos,  $R^2 = 0,6792$ , un valor muy similar al anterior. El p-valor del contraste global de la

	Parámetro	Estimación
$\hat{\beta}_0$	Ordenada en el origen	-0,080727
$\hat{\beta}_1$	Edad	0,000497
$\hat{\beta}_2$	Colesterol total	0,000293
$\hat{\beta}_3$	HDL	-0,000672
$\hat{\beta}_4$	Presión Arterial Sistólica	0,000506
$\hat{\beta}_5$	Diabetes	0,023995
$\hat{\beta}_6$	Tabaquismo	0,017748
$\hat{\beta}_7$	Sexo	0,007638

Tabla 4. Resumen de la información dada por la segunda recta de regresión estimada.

regresión  $H_0: \beta_1 = \dots = \beta_7 = 0$  es  $1,1127 \cdot 10^{-48}$  (el estadístico es 63,82 y usamos la distribución  $F_{k;n-k-1}$  para contrastar, con  $n = 219$  y  $k = 7$ ), un valor incluso más pequeño que el del análisis anterior, siendo muy razonable aceptar que la realidad se corresponde fielmente con las predicciones que damos con esta recta de regresión.

Por último, al igual que hicimos con la primera recta de regresión, para cada parámetro damos sus intervalos de confianza al 5 % (ver Tabla 5) y estudiamos las hipótesis de contraste  $H_0: \beta_j = 0$  (ver Tabla 6).

Intervalos de confianza con $\alpha = 5\%$			
	Parámetro	Extremo inferior	Extremo superior
$\beta_0$	Ordenada en el origen	-0,101623	-0,059831
$\beta_1$	Edad	0,000354	0,000639
$\beta_2$	Colesterol total	0,000223	0,000364
$\beta_3$	HDL	-0,000832	-0,000513
$\beta_4$	Presión Arterial Sistólica	0,000374	0,000637
$\beta_5$	Diabetes	0,016272	0,031717
$\beta_6$	Tabaquismo	0,010298	0,025199
$\beta_7$	Sexo	0,002399	0,012877

Tabla 5. Con una probabilidad del 95% cada uno de los parámetros  $\beta_j, j = 0, \dots, 7$  mencionados en esta tabla se encuentran en dicho rango de valores.

Los p-valores de las hipótesis de contraste  $H_0: \beta_j = 0$  son tan pequeños para todos los parámetros que podemos afirmar que todos ellos son relevantes y aportan información a la recta de regresión. Confirmando nuestras sospechas hechas en el primer análisis que apuntaban a que estos eran los parámetros relevantes. El hecho de que los p-valores sean mucho más pequeños que en el análisis anterior es también un buen síntoma que sugiere que hemos conseguido eliminar duplicidades, es decir, los parámetros que intervienen ahora no presentan tantas relaciones lineales entre ellos.

Con respecto a los intervalos de confianza para un valor  $\alpha = 5\%$ . Este segundo análisis ha mejorado las estimaciones de los parámetros. Consecuentemente, hemos calibrado el impacto que tienen factores de riesgo y cómo estos afectan de una manera u otra al riesgo de evento

Hipótesis de contraste $H_0: \beta_j = 0$		
	Parámetro	p-valor
$\beta_1$	Edad	$6,72 \cdot 10^{-9} \%$
$\beta_2$	Colesterol total	$2,75 \cdot 10^{-12} \%$
$\beta_3$	HDL	$1,05 \cdot 10^{-12} \%$
$\beta_4$	Presión Arterial Sistólica	$1,24 \cdot 10^{-10} \%$
$\beta_5$	Diabetes	$4,38 \cdot 10^{-7} \%$
$\beta_6$	Tabaquismo	$4,78 \cdot 10^{-4} \%$
$\beta_7$	Sexo	0,4466 %

Tabla 6. Realizamos hipótesis de contrastes para todos los parámetros (excepto ordenada en el origen, el hecho de que la ordenada en el origen pueda ser 0 o no, no es relevante ya que simplemente actúa como una corrección de la recta) y damos el p-valor, cuanto más bajo sea este, con mayor contundencia podemos rechazarlo y cuanto más alto, existe menor evidencia estadística para rechazarlo y por tanto lo natural será aceptar la hipótesis de contraste,  $\beta_j = 0$ .

cardiovascular ya que todos los intervalos de confianza nuevos tienen una longitud inferior a la que tenían en el primer análisis (depurar la recta eligiendo solo los valores relevantes nos ha permitido tener un conocimiento más preciso de la información de la que disponemos), destacando especialmente el parámetro "Colesterol total", la longitud de su intervalo de confianza es del orden de casi 10 veces más estrecho (una excepcional mejora en la estimación de la cuantificación de la influencia del colesterol como factor de riesgo).

Además, los intervalos de confianza del segundo análisis de los parámetros "Colesterol total", "Presión Arterial Sistólica" y "Diabetes" son contenidos completamente por los intervalos de confianza obtenidos para estos mismos parámetros en el primer análisis. Para los otros tres parámetros, los intervalos de confianza del segundo análisis están contenidos casi completamente en los respectivos intervalos de confianza del primero. Puede ser interesante estudiar si en nuestra muestra existe algún tipo de relación entre los parámetros "Edad", "HDL" o "Tabaquismo" con el parámetro que hemos añadido para este segundo análisis, la variable "Sexo", y que esta sea la razón por la que sus intervalos de confianza no estén contenidos completamente en los del análisis previo. Ya hemos visto que sí que existe esta entre el "Tabaquismo" y el "Sexo", el análisis para los otros dos es más complicado pero sugiere nuevas preguntas interesantes.

Resumimos las conclusiones más interesantes a las que hemos llegado con este segundo análisis depurado que cambia (esencialmente mejora) algunos de los resultados obtenidos antes y aporta nuevas observaciones importantes.

- Con una probabilidad del 95 %, envejecer 10 años conlleva el aumento de entre un 0,35 % y un 0,64 % del riesgo de evento cardiovascular.
- El aumento del colesterol es perjudicial: Un aumento de 10 mg/dL incrementa el riesgo de evento cardiovascular entre un 0,22 % y un 0,36 % con una probabilidad del 95 %, en cambio, el aumento del HDL es beneficioso para prevenir eventos cardiovasculares, un aumento de 10 mg/mL disminuye el riesgo de evento cardiovascular entre un 0,51 % y un 0,83 % con una probabilidad del 95 %.
- Con una probabilidad del 95 %, el incremento en 10 mmHg de la presión arterial sistólica conlleva el aumento de entre un 0,37 % y un 0,64 % del riesgo de evento cardiovascular.

- Los pacientes diabéticos presentan con una probabilidad del 95 % un aumento de entre un 1,63 % y un 3,17 % del riesgo de evento cardiovascular.
- Los fumadores habituales presentan con una probabilidad del 95 % un aumento de entre un 1,03 % y un 2,52 % del riesgo de evento cardiovascular.
- El sexo del paciente influye en el riesgo que este tiene de sufrir un evento cardiovascular, con un 95 % de probabilidad, un hombre tiene un riesgo de evento cardiovascular de entre un 0,24 % y un 1,29 % superior al que tiene una mujer.

#### 4. Estudio de datos influyentes.

En todo estudio de naturaleza bioestadística es habitual encontrar datos influyentes o “anómalos”. La razón de esto es que en la muestra solemos encontrar pacientes con condiciones o patologías especiales que requieren ser tratados de forma personalizada, susceptibles de alejarse demasiado de las estimaciones predichas por el modelo. Por ello, es importante detenerse en los datos anómalos para estudiarlos con mayor cuidado y de manera individualizada. Queremos decidir si estos pacientes requieren un tratamiento excepcional o si el modelo ha fallado para ellos.

Decimos que una observación es influyente si tiene mucho impacto en  $\hat{\beta}$  y en  $\hat{y}$ , a estas observaciones las llamaremos puntos críticos.

Consideremos  $\hat{Y}$  el vector con las estimaciones  $\hat{y}_i$ ,  $Y$  el vector con las observaciones  $y_i$  y  $X$  la matriz de datos, entonces, se satisface,

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y.$$

Definimos la matriz  $H := X(X^T X)^{-1} X^T$ , y tomamos la coordenada  $h_{ii}$  que definimos como leverage del dato  $i$ -ésimo. Pasamos ahora a definir la distancia de Cook como,

$$D_i := \frac{1}{k+1} \left( \frac{e_i}{s_R \sqrt{1-h_{ii}}} \right)^2 \frac{h_{ii}}{1-h_{ii}}.$$

Podemos usar el leverage,  $h_{ii}$  y la distancia de Cook para identificar puntos críticos [8, 9].

Las distancias de Cook de nuestra muestra son por lo general bastante pequeñas. Esta medida resalta por encima del resto solo a los pacientes 33 (distancia de Cook 0,1742) y 158 (distancia de Cook 0,1613), exceptuando estos dos datos, el resto de elementos de la muestra lleva asociada una distancia de Cook inferior a 0,075.

Para el leverage consideraremos anómalos aquellos datos que sean superiores a  $3 \frac{k+1}{n} = 0,17808$ . Tenemos 6 datos que satisfacen esto, los pacientes 15 ( $h_{ii} = 0,2112$ ), 29 ( $h_{ii} = 0,1869$ ), 79 ( $h_{ii} = 0,2290$ ), 115 ( $h_{ii} = 0,3521$ ), 159 ( $h_{ii} = 0,3813$ ) y 161 ( $h_{ii} = 0,2389$ ).

Los pacientes 33 y 158 que eran influyentes según la distancia de Cook son también los únicos dos pacientes cuyo riesgo de evento cardiovascular es superior al 15 %, 18 % y 19 % respectivamente. Como su riesgo de evento cardiovascular es tan elevado comparado con el del resto de pacientes es normal que estos dos puntos sean considerados influyentes, la coordenada  $y$  se encuentra lejos de la del resto de datos.

Pasamos al estudio personalizado de los datos influyentes según el “leverage” (ninguno coincide con los datos influyentes según la distancia de Cook)

- Paciente 15: Se trata de un paciente pluripatológico en tratamiento con múltiples fármacos, entre ellos *Acecumarol*. Este es un anticoagulante, un tratamiento preventivo para disminuir el riesgo de accidentes tromboembólicos. Por ello, es de esperar que algunos de

sus factores de riesgo hayan sido modificados por el tratamiento farmacológico mientras que otros factores han permanecido inalterados, creando un dato con parámetros poco comunes que puede ser leído de manera influyente por la regresión. Además el paciente era diabético, en tratamiento hipoglucemiante, lo que hace que el valor de las cifras de glucemia estén alteradas artificialmente.

- Paciente 29: Se trata de un paciente con hipertrigliciremia severa con valores de colesterol dentro de la normalidad, en este paciente habría que descartar una hipertrigliceremia familiar, en cuyo caso habría factores genéticos que podrían influir en su riesgo cardiovascular. Este paciente sobrepasa en gran manera los valores habituales de triglicéridos y es probable que por ello sea un dato influyente para la recta de regresión.
- Paciente 79: La distancia de Cook para este paciente es muy pequeña,  $1,85 \cdot 10^{-6}$ , así que según este criterio, el paciente 79 se encuentra lejos de ser un dato influyente. Su única anomalía es padecer diabetes mellitus tipo I, único paciente con esta disfunción de entre todo los analizados. Esta característica podría hacer que este dato sea influyente. No obstante, no somos capaces de decidir si el modelo ha fallado ya que el dato podría no ser influyente (la predicción y el dato "REGICOR" real fueron muy similares).
- Paciente 115: La tensión diferencial, considerada factor de riesgo por varios autores [18], de esta paciente es muy elevada. Los valores de "LDL" y "Triglicéridos" son demasiado elevados, mientras que los de "HDL" se encuentran por debajo de lo aconsejable. Además, el paciente es diabético y fumador habitual. Al tener tantos parámetros en rangos muy alejados de las cifras habituales es razonable pensar que este paciente deba ser considerado un dato influyente para nuestra recta de regresión. Su riesgo de evento cardiovascular debería ser estudiado de forma especial y personalizada.
- Paciente 159: Paciente de avanzada edad pluripatológico, polimedicado, bien controlado. Los datos de los distintos parámetros son muy satisfactorios para una persona de su edad (posiblemente como consecuencia de su medicación) que lo convertiría en un dato influyente por estar en rangos distintos a lo que suele ser habitual para pacientes de su edad.
- Paciente 161: Paciente diabético fumador de avanzada edad relativamente bien controlado por medicación lo que hace que sus parámetros puedan estar modificados artificialmente.

Es interesante observar que la gran mayoría de los datos influyentes son diabéticos (pacientes que no constituyen un porcentaje tan representativo de la población). Por ello, para realizar un análisis más depurado y fiable, sería aconsejable estudiar la población diabética y no diabética de forma separada.

Para nuestro estudio, al contar solo con 6 datos influyentes sobre un total de 219, retirarlos del análisis no cambiaría de forma significativa los resultados obtenidos.

## 5. Regla de clasificación

Queremos predecir si un paciente con unas determinadas características es susceptible de tener un riesgo de evento cardiovascular bajo o moderado (inferior al 10%), clasificado en la población 0, o de padecer un riesgo alto o muy alto (igual o superior al 10%), población 1.

En nuestra muestra contamos con 206 pacientes en la población 0 (riesgo bajo o moderado) y con 13 en la población 1 (riesgo alto o muy alto).

La regla de clasificación que hemos aplicado es la regresión logística, usando el comando "Solver" del programa Excel para resolver la combinación de  $\hat{\beta}_j$  que maximiza el estimador log VEROSIMILITUD.

Con esto, la regla obtenida consiste en calcular el valor

$$f(x_i) = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{Edad} + \hat{\beta}_2 \cdot \text{Colesterol Total} + \hat{\beta}_3 \cdot \text{HDL} + \hat{\beta}_4 \cdot \text{LDL} \\ + \hat{\beta}_5 \cdot \text{Triglicéridos} + \hat{\beta}_6 \cdot \text{PAS} + \hat{\beta}_7 \cdot \text{PAD} + \hat{\beta}_8 \cdot \text{Glúcidos} \\ + \hat{\beta}_9 \cdot \text{Filtrado Glomerular} + \hat{\beta}_{10} \cdot \text{Albuminuria} \\ (+\hat{\beta}_{11} \text{ si el paciente es diabetico}) \\ (+\hat{\beta}_{12} \text{ si el paciente es fumador}) \\ (+\hat{\beta}_{13} \text{ si el paciente es hombre}),$$

donde los valores de los coeficientes  $\hat{\beta}_j$  vienen dados en la Tabla 7.

Regresión Logística		
	Parámetro	Estimación para la regresión logística
$\hat{\beta}_0$	Ordenada en el origen	-800
$\hat{\beta}_1$	Edad	1,0139631
$\hat{\beta}_2$	Colesterol total	0,1004485
$\hat{\beta}_3$	HDL	1,0092167
$\hat{\beta}_4$	LDL	1,03056338
$\hat{\beta}_5$	Triglicéridos	1,0258
$\hat{\beta}_6$	Presión Arterial Sistólica	1,030683
$\hat{\beta}_7$	Presión Arterial Diastólica	1,0169
$\hat{\beta}_8$	Glucosa	1,0222
$\hat{\beta}_9$	Filtrado Glomerular	1,000156
$\hat{\beta}_{10}$	Albuminuria	1,0002
$\hat{\beta}_{11}$	Diabetes	1,0000833
$\hat{\beta}_{12}$	Tabaquismo	1,000083
$\hat{\beta}_{13}$	Sexo	1,00015

Tabla 7. Valores propuestos por la función "Solver" de Excel para la regla de clasificación de regresión logística.

De esta forma, si el valor de  $f(x_i)$  es menor que 0 clasificamos el dato  $x_i$  en la población 0, y en caso contrario en la población 1. Es cierto que la función "Solver" a veces puede ser imprecisa. No obstante, la regla de clasificación que hemos dado ha sido acertada para un 93,15 % de los datos (204 del total de 219), aunque al haber usado los mismos datos para comprobar la fiabilidad y crear la regla de clasificación este valor tan elevado de aciertos puede haber sido adulterado, deberíamos comprobar esta regla con datos nuevos. En caso de no poder conseguir datos nuevos, para evaluar la regla, para cada dato de los que disponemos podemos crear una regla de clasificación con esta metodología (regresión logística), hacer la predicción de ese dato y evaluar si la predicción ha acertado.

Nuestra regla ha clasificado de forma altamente precisa los datos de la población 0, solo 7 fallos de entre 206 pacientes de esta población, pero clasifica con poca precisión a los pacientes de la población 1, 8 fallos de entre 13 pacientes.

## 6. Conclusiones

En este estudio hemos realizado dos rectas de regresión lineal para mejorar los resultados proporcionados por esta herramienta. De esta manera se ha conseguido cuantificar de forma muy precisa el impacto de 7 de los 13 factores de riesgo de los cuales sospechábamos una potencial relación con el riesgo cardiovascular. De los 6 factores de riesgo iniciales restantes, uno de ellos, "LDL" está relacionado implícitamente de forma lineal con otros dos de los factores incluidos en el segundo análisis, "Colesterol total" y "HDL". Otro de los potenciales factores de riesgo descartados para la segunda recta de regresión, el "Filtrado glomerular" ha sido estudiado cualitativamente de forma independiente en el Apéndice A. Para futuros estudios se sugiere estudiar la posible independencia de los 4 potenciales factores de riesgo restantes ("Triglicéridos", "Presión Arterial Diastólica", "Glucosa", "Albuminaria"). También se sugiere corroborar los resultados aquí obtenidos con otros datos.

En este trabajo también hemos propuesto una regla de clasificación por medio de una regresión logística. Para un próximo estudio proponemos la creación de reglas óptimas de clasificación para estas dos poblaciones. Con este fin, como funciones para cada una de las poblaciones se propone: usar funciones tales que para los parámetros  $\beta_1 - \beta_{10}$  sean una normal de dimensión 10 con media y varianza muestral (intentando obtener el dato de la albuminaria y no solo la clasificación en A1, A2 o A3) y para los parámetros  $\beta_{11}, \beta_{12}, \beta_{13}$  Bernoullis tomando como  $p$  el siguiente valor

$$p = \frac{\text{casos valor 1 en la población } i}{\text{casos totales en la población } i}, \quad i = 1, 2$$

independientes cada una de ella del resto de parámetros que intervienen en la función de densidad. Tras esto se aconseja incluir costes de mala clasificación si se desea evitar clasificar pacientes con riesgo alto o muy alto como pacientes de riesgo bajo o moderado, mejorando su función como una herramienta de prevención de riesgo. También sería interesante usar algún test para cerciorarnos de que estas funciones de densidad reflejan la realidad.

### A. El filtrado glomerular es un factor de riesgo cardiovascular no lineal.

La Enfermedad Renal Crónica (ERC) tiene en nuestro medio una prevalencia aproximada del 15%, siendo mayor en varones, personas con edad avanzada y pacientes con enfermedad cardiovascular. Fue incluida como factor de riesgo independiente de Enfermedad Cardiovascular por el *Joint National Committee on Prevention Detection and Treatment of High Blood pressure* en 2003 [7]. Partiendo de estas premisas, parece interesante utilizar los marcadores analíticos de función renal para el diagnóstico precoz de patología cardiovascular ya que la marcada acumulación de factores de riesgo cardiovascular en estos nefrópatas podría hacer considerar a la enfermedad renal crónica como un trastorno cardiovascular (e.g. [12, 14]). En esta Sección discutiremos que existe una dependencia entre estos dos factores pero que esta no es lineal.

La metodología empleada es un contraste  $\chi^2$  de independencia (supondremos los valores de las probabilidades desconocidos y los estimaremos por cantidad de cada clase dividido por el tamaño de la muestra). Vamos a contrastar la siguiente hipótesis de rechazo.

$H_0$ : el filtrado glomerular y el riesgo cardiovascular se distribuyen según dos variables aleatorias independientes.

Por último, antes de la recopilación de datos, conviene explicar el significado de las distintas categorías en relación al filtrado glomerular, así como los criterios de clasificación.



Estadíos del filtrado glomerular estimado ( $mL/min/1,73 m^2$ ) en función de creatinina con fórmula CKD-EP<sup>1</sup>

- G1  $\geq 90\%$ : FG normal o elevado.
- G2 60 – 89%: FG ligeramente disminuido.
- G3a 45 – 59%: FG ligera-moderadamente disminuido.
- G3b 30 – 44%: FG moderada-gravemente disminuido.
- G4 15 – 29%: FG gravemente disminuido.
- G5<sup>2</sup> < 15%: Fallo renal.

Tabla de observados				
Fil. Glomerular/RC (REGICOR)	Bajo	Moderado	Alto	Muy Alto
G1	70	17	4	0
G2	53	39	6	2
G3a	10	7	1	0
G3b	3	5	0	0
G4	2	0	0	0

Tabla 8. Tabla de observados relación filtrado glomerular con riesgo cardiovascular (según test REGICOR).

Tabla de esperados				
Fil. Glomerular/RC (REGICOR)	Bajo	Moderado	Alto	Muy Alto
G1	57.34	28.26	4.57	0.83
G2	63.01	31.05	5.02	0.91
G3a	11.34	5.59	0.90	0.16
G3b	5.04	2.48	0.40	0.07
G4	1.26	0.62	0.10	0.02

Tabla 9. Tabla de esperados relación filtrado glomerular con riesgo cardiovascular (según test REGICOR).

El estadístico de Pearson toma el valor  $b = 19,00$ , lo cual implica que el  $p$ -valor es 8,84% (contrastamos con  $\chi^2_{12}$ ). Esto permite rechazar la hipótesis de contraste  $H_0$  con un nivel de confianza superior al 90%.

Realicemos un análisis cualitativo sobre la relación entre la filtración glomerular y el riesgo cardiovascular. Las diferencias más abultadas que encontramos entre observados y esperados se encuentran en las clases G1 y G2. Para G1 el riesgo observado parece más bajo que el esperado, mientras que para G2 los resultados parecen indicar lo contrario. En cambio, para las clases G3a, G3b y G4, los resultados observados concuerdan con los esperados, lo cuál sugiere que una vez alcanzado el estado G3a (o superior) la calidad del filtrado glomerular es independiente del riesgo cardiovascular.

En base a estos resultados podemos concluir que los pacientes en estadio G1 (filtrado glomerular normal o elevado) presentan en la práctica un RCV menor que el que cabría esperar si ambos fenómenos fueran independientes. El adecuado filtrado glomerular sería capaz de compensar, al menos parcialmente, el efecto nocivo de estos factores de riesgo.

<sup>1</sup>Individuos de raza negra: multiplicar por 1,159

<sup>2</sup>En nuestra muestra, ningún paciente ha alcanzado la categoría de fallo renal y en consecuencia no la tendremos en cuenta para nuestro estudio.

En el estadio G2 ya el filtrado glomerular se encuentra discretamente disminuido (60 – 89 %), la uremia aumenta levemente y sería responsable de un estado de inflamación crónica a nivel de toda la anatomía con el consiguiente aumento de RCV como se aprecia en nuestro estudio.

Cuando la insuficiencia renal ya está consolidada, es decir filtrado glomerular inferior al 60 % (estadios G3a a G5), la conjunción de los distintos factores de riesgo ya han lesionado de forma importante e irreversible tanto el sistema cardiovascular como la función renal y en consecuencia el deterioro renal no se ve reflejado en un aumento del RCV que ya es de por sí elevado.

Del análisis de estos resultados se desprende que la detección precoz de los pacientes con disminución ligera del filtrado glomerular (G2), previa a la insuficiencia renal crónica, permitiría implementar las actuaciones sobre los Factores de Riesgo Cardiovascular modificables y con ello disminuir de forma costo-efectiva la morbimortalidad cardiovascular así como la evolución hacia la enfermedad renal crónica.

## Referencias

- [1] M. Al de Francisco, L. Aguilera y V. Fuster. Enfermedad cardiovascular, enfermedad renal y otras enfermedades crónicas. Es necesaria una intervención más temprana en la enfermedad renal crónica. *Nefrología* 29.1 (2009), págs. 6-9.
- [2] R. Alcanzar, L. Orte y A. Otero. Enfermedad renal crónica. *Nefrología* (2008), págs. 3-6.
- [3] R. Alcázar Arroyo, L. Orte, E. González Parra, J. Górriz, J. Navarro, A. Martín de Francisco, M. Egocheaga y F. Álvarez Guisasaola. Documento de consenso SEN-semFYC sobre la enfermedad renal crónica. *Nefrología* 28.3 (2008), págs. 273-282.
- [4] R. W. Alexander. Hypertension and the pathogenesis of atherosclerosis: oxidative stress and the mediation of arterial inflammatory response: a new perspective. *Hypertension* 25.2 (1995), págs. 155-161.
- [5] A. Arias Morales, R. García Hernández y M. Oliva Pérez. Riesgo cardiovascular global en pacientes ancianos hipertensos. *Revista Cubana de Medicina* 53.2 (2014), págs. 178-188.
- [6] A. Asesio Sánchez, A. Fernández y A. Morcillo. *Influencia de la Dieta Mediterránea en la percepción de la salud: Estudio en población adolescente del IES Antonio Machado de Soria y población adulta de la provincia de Soria*. Editorial Académica Española, 2018.
- [7] A. V. Chobanian, G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo Jr, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright Jr et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *hypertension* 42.6 (2003), págs. 1206-1252.
- [8] R. D. Cook. Detection of influential observation in linear regression. *Technometrics* 19.1 (1977), págs. 15-18.
- [9] R. D. Cook. Influential observations in linear regression. *Journal of the American Statistical Association* 74.365 (1979), págs. 169-174.
- [10] J. M. B. Díez, J. L. del Val García, J. T. Pelegrina, J. L. M. Martínez, R. M. Peñacoba, I. G. Tejón, E. M. R. Quintana, M. P. Sajkiewicz, A. A. Boronat, B. Á. Pérez et al. Epidemiología de las enfermedades cardiovasculares y factores de riesgo en atención primaria. *Revista española de cardiología* 58.4 (2005), págs. 367-373.
- [11] S. G. García, R. M. Barmudez, J. B. Sanjuan, A. C. Amenós, R. D. Piquet y A. L. M. de Fran. Recomendaciones sobre la utilización de ecuaciones para la estimación del filtrado glomerular en adultos. *Nefrología* 26.6 (2006), págs. 658-665.

- [12] M. Gorostidi, M. Sánchez-Martínez, L. M. Ruilope, A. Graciani, J. Juan, R. Santamaría, M. D. del Pino, P. Guallar-Castillón, F. de Álvaro, F. Rodríguez-Artalejo et al. Prevalencia de enfermedad renal crónica en España: impacto de la acumulación de factores de riesgo cardiovascular. *Nefrología* 38.6 (2018), págs. 606-615.
- [13] R. R. Gorrita Pérez, A. Gilvonio Cárdenas e Y. Hernández Martínez. Caracterización del hábito de fumar en un grupo de escolares adolescentes. *Revista Cubana de Pediatría* 84.3 (2012), págs. 256-264.
- [14] M. I. Gutiérrez Pérez y J. C. Rodríguez Sánchez. *Problemas nefrourológicos y genito-uritarios*.
- [15] S. M. Haffner, S. Lehto, T. Rönnemaa, K. Pyörälä y M. Laakso. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *New England journal of medicine* 339.4 (1998), págs. 229-234.
- [16] W. Insull Jr. The pathology of atherosclerosis: plaque development and plaque responses to medical treatment. *The American journal of medicine* 122.1 (2009), S3-S14.
- [17] J. P. S. de Lafuente, Y. Sáez, M. Vacas, M. Santos, J. D. Sagastagoitia, E. Molinero y J. A. Iriarte. Diferencias de sexo en los factores de riesgo cardiovascular en pacientes con enfermedad coronaria comprobada angiográficamente. *Clínica e investigación en arteriosclerosis* 21.4 (2009), págs. 173-178.
- [18] J. A. Linares, J. B. Simó, F. S. Sancho, S. R. Martínez, M. Y. Marco y M. M. Pérez. La presión diferencial como factor independiente de riesgo cardiovascular. *Atención Primaria* 36.1 (2005), págs. 19-24.
- [19] E. M. T. Lira. Impacto de la hipertensión arterial como factor de riesgo cardiovascular. *Revista Médica Clínica Las Condes* 26.2 (2015), págs. 156-163.
- [20] Á. C. Matía Cubillo, J. M. Soto Esteban y M. E. Martín Pascual. *Riesgo cardiovascular*.
- [21] J. Millán, A. Hernández-Mijares, J. F. Ascaso, M. Blasco, A. Brea, Á. Díaz, P. González-Santos, T. Mantilla, J. Pedro-Botet, X. Pintó et al. La auténtica dimensión del colesterol-no-HDL: colesterol aterogénico. *Clínica e Investigación en Arteriosclerosis* 28.6 (2016), págs. 265-270.
- [22] V. Moreno, A. García Raso, M. García Bueno, C. Sánchez-Sánchez, E. Meseguer, R. Mata y P. Llamas. Factores de riesgo vascular en pacientes con ictus isquémico. Distribución según edad, sexo y subtipo de ictus. *Rev Neurol* 46.10 (2008), págs. 593-8.
- [23] M. Morillas Bueno. ¿Hasta dónde bajar la glucemia para la prevención cardiovascular en la diabetes? *Cardiología hoy 2015* (2015), págs. 593-598.
- [24] M. Naghavi, H. Wang, R. Lozano, A. Davis, X. Liang, M. Zhou et al. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 385.9963 (2015), págs. 117-171.
- [25] Ó. Pérez-Méndez. Lipoproteínas de alta densidad (HDL). ¿Un objetivo terapéutico en la prevención de la aterosclerosis? *Archivos de cardiología de México* 74.1 (2004), págs. 53-67.
- [26] M. Ramirez Iñiguez de la Torre et al. "Determinación del Riesgo Cardiovascular en una población laboral aparentemente sana. Relación con Variables Sociodemográficas y Laborales". Tesis doct. Universitat de les Illes Balears, 2016.
- [27] G. A. Roth, M. H. Forouzanfar, A. E. Moran, R. Barber, G. Nguyen, V. L. Feigin, M. Naghavi, G. A. Mensah y C. J. Murray. Demographic and epidemiologic drivers of global cardiovascular mortality. *New England Journal of Medicine* 372.14 (2015), págs. 1333-1341.
- [28] D. Sarre-Álvarez, R. Cabrera-Jardines, F. Rodríguez-Weber y E. Díaz-Greene. Enfermedad cardiovascular aterosclerótica. Revisión de las escalas de riesgo y edad cardiovascular. *Medicina interna de México* 34.6 (2018), págs. 910-923.

- [29] Wikipedia. *Provincia de Soria* — *Wikipedia, The Free Encyclopedia*. <http://es.wikipedia.org/w/index.php?title=Provincia%20de%20Soria&oldid=145180302>. [Online; accessed 11-August-2022]. 2022.
- [30] S. Yusuf, S. Hawken, S. Ôunpuu, T. Dans, A. Avezum, F. Lanas, M. McQueen, A. Budaj, P. Pais, J. Varigos et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The lancet* 364.9438 (2004), págs. 937-952.

**Sobre el/los autor/es:**

*Nombre:* Alejandro Fernández-Jiménez

*Correo electrónico:* alejandro.fernandezjimenez@maths.ox.ac.uk

*Institución:* University of Oxford

*Orcid:* 0000-0002-4548-4374.

*Nombre:* Álvaro Fernández Jiménez

*Correo electrónico:* alvfj1404@gmail.com

*Institución:* Universidad Católica de Valencia

*Orcid:* 0000-0002-1580-9764.

*Nombre:* Abilio José Fernández Vicente

*Correo electrónico:* abiliofernandez@yahoo.es

*Institución:* Sanidad de Castilla y León (SaCyL)

*Orcid:* 0000-0002-9507-7937.