

# Tratamiento de la diabetes en adultos mayores

## ¿Más intenso es mejor?

CoderHouse Data Science

20 de Enero de 2023

Walter Gardella Sambeth

# Índice

<b>1. Resumen</b>	<b>2</b>
<b>2. Objetivos del modelo</b>	<b>2</b>
<b>3. Descripción de los datos</b>	<b>2</b>
<b>4. Análisis exploratorio</b>	<b>5</b>
<b>5. Conversión de variables</b>	<b>6</b>
<b>6. Análisis post-conversión</b>	<b>6</b>
6.1. Correlaciones entre variables	6
6.2. Análisis de frecuencia de combinaciones	7
6.3. Valores cero	8
6.4. Detección de outliers	8
<b>7. Cálculo de las tasas de mortalidad</b>	<b>9</b>
<b>8. Análisis de componentes principales</b>	<b>9</b>
<b>9. Selección de variables significativas</b>	<b>10</b>
<b>10. Aplicación de algoritmos de ML</b>	<b>11</b>

## 1. Resumen

Los consensos internacionales respecto del tratamiento de la Diabetes demuestran la conveniencia de enfoques intensificados y un abordaje temprano de los mismos.

Sin embargo, en los pacientes ancianos y con menor expectativa de vida los riesgos a corto plazo de los tratamientos pueden pesar más que el posible beneficio a largo plazo.

Como no es posible determinar la expectativa de vida de pacientes individuales, se trata de desarrollar y validar un modelo matemático que pronostique índices de mortalidad a 5 y 10 años para adultos mayores con diabetes, con o comorbilidades.

## 2. Objetivos del modelo

Este modelo busca encontrar qué variables socioeconómicas y de salud afectan más fuertemente esta tasa de mortalidad.

El principal beneficio recae sobre el sistema de salud, ya que reduciría los costos al evitar aplicar tratamientos más caros y complejos a pacientes que no van a obtener el beneficio esperado dada su reducida esperanza de vida.

Por otro lado, esos pacientes no estarían sometidos al riesgo adicional de aplicar esas alternativas terapéuticas, que en ningún caso son inocuas.

## 3. Descripción de los datos

Se trabaja con un dataset obtenido de la página [A Novel Dataset of Predictors of Mortality for Older Veterans Living with Type II Diabetes - Mendeley Data](#), que recoge información sobre veteranos de guerra de 65 o más años de edad al 01/01/2006.

Este dataset incluyó pacientes que cumplieron los siguientes criterios:

- Los pacientes tenían 65 años o más de edad al 1/1/2006.
- Ingresados al programa de atención de salud a Veteranos de las Fuerzas Armadas.
- Todos tenían diabetes diagnosticada clínicamente al momento de su inclusión en el ensayo.
- Al menos una visita al médico con controles registrados para tres parámetros típicos de la diabetes (Tensión arterial, BMI y Hemoglobina A1c) en los 24 meses anteriores al estudio.

Por otro lado, se aplicaron las siguientes consideraciones al recabar los datos:

- El sexo sólo se considera binario Hombre/Mujer.
- El estado civil sólo se considera Casado, Soltero (nunca casado o divorciado) o Viudo.
- Se definieron ocho grupos de prioridad para la inclusión en el estudio teniendo en cuenta niveles de discapacidad física y dificultades socioeconómicas.

A continuación se describen las variables que lo componen:

- SEX: 0=Mujer, 1=Hombre
- AGE: Edad en años
- RACE:
  - 1=Blanca
  - 2=Negra
  - 3=Otra
- MARRIED: Estado civil
  - MARRIED=Casado
  - SINGLE=Soltero (nunca casado o divorciado)
  - WIDOWED=viudo
- PRIORITY: Grupo de prioridad para la inclusión en el estudio.
  - 1: Discapacidad de 50% o mayor relacionada con el servicio.
  - 2: Discapacidad de 30 - 40% relacionada con el servicio.
  - 3: Discapacidad de 10 - 20% relacionada con el servicio.
  - 4: Discapacitados catastróficos - permanencia obligada en su hogar.
  - 5: Dificultades económicas.
  - 6: Discapacidad no relacionada con el servicio, otros.
  - 7: Sin discapacidad relacionada con el servicio ni dificultades económicas
  - 8: Sin discapacidad relacionada con el servicio ni dificultades económicas
- N\_IP: Número de días de internación.
- N\_OP: Cantidad de consultas en consultorios externos.
- DEATH\_5: Muerto dentro de los 5 años siguientes a la entrevista.
- DEATH\_10: Muerto dentro de los 10 años siguientes a la entrevista.
- FRAILTY: Índice de fragilidad. Indica la presencia o ausencia de treinta problemas de salud relacionados con la edad<sup>1</sup>.
- BMI: Índice de Masa Corporal (Peso / (altura en metros)<sup>2</sup>).
- SYSTOLIC: Presión arterial sistólica (máxima)
- DIASTOLIC: Presión arterial diastólica (mínima)

---

<sup>1</sup> Tomado de Orkaby AR, Nussbaum L, Ho Y, et al. The burden of frailty among U.S. veterans and its association With mortality, 2002–2012. J Gerontol A Biol Sci Med Sci 2018;74:1257–1264

- TRI: Triglicéridos en sangre.
- LDL: Colesterol LDL en sangre.
- HDL: Colesterol HDL en sangre.
- A1C: Concentración de Hemoglobina A1c en sangre. Indica el control a largo plazo de la diabetes
- MICROALB: Microalbuminuria en orina. Evalúa el estado de los riñones.
- SERUMALB: Albúmina en sangre.
- SERUMCRE: Creatinina en sangre.
- INSULIN: Prescripción de insulina: 0=no, 1=sí.
- ALPHA: Prescripción de inhibidor de la alfa-glucosidasa: 0=no, 1=sí.
- TZD: Prescripción de Tiazolidinedionas: 0=no, 1=sí.
- BIGUAN: Prescripción de Biguanidas: 0=no, 1=sí.
- SULF: Prescripción de Sulfonilureas: 0=no, 1=sí.
- OTHER\_MED: Prescripción de otros medicamentos contra la diabetes: 0=no, 1=sí.
- CHF: Enfermedad cardíaca congestiva: 0=no, 1=sí.
- ARRHYTHMIA: Arritmias cardíacas: 0=no, 1=sí.
- VALVULAR: Enfermedad valvular: 0=no, 1=sí.
- PHTN: Problemas en la circulación pulmonar: 0=no, 1=sí.
- PVD: Enfermedad vascular periférica: 0=no, 1=sí.
- HTN: Hipertensión arterial no complicada: 0=no, 1=sí.
- HTNCX: Hipertensión arterial complicada: 0=no, 1=sí.
- PARALYSIS: Parálisis: 0=no, 1=sí.
- NEUROOTHER: Otros problemas neurológicos: 0=no, 1=sí.
- PULMONARY: Enfermedad pulmonar crónica: 0=no, 1=sí.
- DMCX: Diabetes complicada: 0=no, 1=sí.
- HYPOTHYROID: Hipotiroidismo: 0=no, 1=sí.
- RENAL: Enfermedad renal: 0=no, 1=sí.
- LIVER: Enfermedad hepática: 0=no, 1=sí.
- PUD: Úlcera gástrica excluyendo casos con sangrado: 0=no, 1=sí.
- HIV: HIV/SIDA: 0=no, 1=sí.
- LYMPHOMA: Linfoma: 0=no, 1=sí.
- METS: Cáncer metastásico: 0=no, 1=sí.
- TUMOR: Tumor sólido sin metástasis: 0=no, 1=sí.
- RHEUMATIC: Artritis reumatoidea / Enfermedad en el colágeno vascular: 0=no, 1=sí.
- COAG: Coagulopatía: 0=no, 1=sí.
- OBESITY: Obesidad: 0=no, 1=sí.
- WEIGHTLOSS: Bajo peso: 0=no, 1=sí.
- ANEMIA: Anemia por deficiencias: 0=no, 1=sí.

- FLUIDSLYTES: Desórdenes electrolíticos o de fluidos: 0=no, 1=sí.
- BLOODLOSS: Anemia por pérdida de sangre: 0=no, 1=sí.
- ALCOHOL: Alcoholismo: 0=no, 1=sí.
- DRUGS: Drogadicción: 0=no, 1=sí.
- PSYCHOSES: Psicosis: 0=no, 1=sí.
- DEPRESSION: Depresión clínica: 0=no, 1=sí.
- SEVERE\_DEP: Depresión severa: 0=no, 1=sí.
- CAD: Enfermedad arterial coronaria: 0=no, 1=sí.
- AMI: Infarto agudo de miocardio: 0=no, 1=sí.
- RETINOPATHY: Retinopatía: 0=no, 1=sí.
- HYPERG: Hiperglucemia: 0=no, 1=sí.
- AMPUTATION: Amputación de miembro inferior: 0=no, 1=sí.
- FEET: Infecciones por pie diabético: 0=no, 1=sí.
- SMOKER: Tabaquismo: 0=no, 1=sí.
- RETSCREEN: Bajo controles por posible retinopatía: 0=no, 1=sí.
- PCI: Cateterismo coronario: 0=no, 1=sí.
- ABI: Índice tobillo - brazo<sup>2</sup>: 0=no, 1=sí.
- ESLD: Etapa final de cirrosis o hígado graso: 0=no, 1=sí.
- CABG: Cirugía de by-pass coronario: 0=no, 1=sí.
- BP\_RX: Prescripción de medicamentos anti hipertensivos: 0=no, 1=sí.

El dataset se compone de 275190 filas y 70 columnas.

Las variables se distribuyen de la siguiente forma:

- Variables numéricas: 14
- Variables categóricas con valores múltiples: 3
- Variables categóricas binarias: 53

## 4. Análisis exploratorio

En el análisis exploratorio se estudiaron las distribuciones de las variables numéricas y categóricas, así como la presencia de nulos, valores duplicados e incongruentes.

Al analizar la distribución de las variables numéricas, se descubrió que presentaban gran dispersión en los valores por lo que en principio parecía existir una gran cantidad de outliers. Al verificar los valores se encontró que en realidad estos valores son plausibles desde el punto de vista médico y tendrían significancia como indicadores de comorbilidades, así que se los dejó incluidos para revisarlos en una etapa posterior.

---

<sup>2</sup> [Ver información en Mayo Clinic](#)

Se encontró que las variables `MICROALB` y `SERUMALB` tienen una elevada proporción de valores faltantes. Esto se explica en el primer caso porque la albuminuria es un hallazgo patológico indicador de daño renal, y en el segundo caso porque no es un análisis de rutina, sino que el médico tiene que tener motivos para solicitarlo.

El dataset en general presentó un 1,69% de registros duplicados, que fueron eliminados.

Respecto de la variable Sexo, el dataset está muy desbalanceado, con únicamente un 1,14% de mujeres registradas.

Se encontró que las variables `OBESITY` y `WEIGHTLOSS` (obesidad y bajo peso, respectivamente) tenían valores incoherentes, ya que en muchos casos no se relacionaban con los valores de BMI registrados (que son los marcadores clínicos de obesidad o bajo peso) e incluso en varios casos se presentaban ambos marcadores en la misma fila. Debido a todo esto se decidió eliminar estas columnas del dataset.

## 5. Conversión de variables

Considerando la dispersión encontrada en las variables numéricas y que los valores nulos en las variables `MICROALB` y `SERUMALB` no podían ser excluidos del análisis porque tienen significancia médica, se realizó una conversión de variables que permitió aprovechar para el análisis toda la información disponible.

- Las variables numéricas se agruparon en función de su significación médica en valores enteros (desde 1 a 2, 3, 4 o 5 según la variable).
- Todas las variables binarias se convirtieron a valores distintos de 0 (1=Falso, 2=Verdadero).
- Todos los valores nulos se convirtieron a 0.

## 6. Análisis post-conversión

### 6.1. Correlaciones entre variables

Utilizando el paquete `pandas_profiling` se analizaron las correlaciones entre las variables convertidas y se encontró que los siguientes grupos presentaban alta correlación entre sí:

- `AGE/RATE_5/RATE_10`: La mortalidad se correlaciona muy fuertemente con el grupo etario, ya que los mayores de 89 quedan todos juntos en el mismo grupo.
- `SERUMCRE/RENAL`: Los valores anormalmente altos de creatinina sérica son uno de los síntomas relacionados con enfermedades renales.

- DEATH\_5/DEATH\_10: Dado que indican si el paciente falleció dentro de los 5 o 10 años del estudio, claramente se correlacionan.
- HTN/BP\_RX: La detección de hipertensión arterial se correlaciona con la prescripción de medicamentos para tratarla.
- HTNCX/RENAL: Una de las complicaciones de la hipertensión arterial es el daño renal.
- DMCX/RETINOPATHY: Una de las complicaciones de la diabetes es la retinopatía.
- LIVER/ESLD: La variable ESLD indica la etapa final de cirrosis o hígado graso, ambas enfermedades hepáticas.
- DEPRESSION/SEVERE\_DEP: Todos los casos de depresión severa se indican también como depresión, por lo que los valores se correlacionan.
- HDL/TRI: **Éste es el único caso que a priori no resulta claramente explicable.**

Se tomó la decisión de que, si en pasos futuros se consideraba eliminar alguna de estas variables de alta correlación, se elegiría retener las siguientes:

- HDL/TRI: Se eligió TRI porque tiene menos nulos.
- SERUMCRE/RENAL: Se eligió RENAL
- HTN/BP\_RX: Se eligió BP\_RX
- HTNCX/RENAL: Se eligió RENAL para no sumar correlación.
- DMCX/RETINOPATHY: Se eligió DCMX.
- LIVER/ESLD: Se emplea LIVER
- DEPRESSION/SEVERE\_DEP: Se eligió DEPRESSION

Como se puede observar, en los casos en que la correlación se debió a variables que indican grados de afectación o una clase general y otra particular, se mantuvo la de menor grado o mayor generalidad.

## 6.2. Análisis de frecuencia de combinaciones

Se realizó un análisis de la frecuencia de combinaciones de valores para algunos pares de variables que podían tener relación médica:

- SYSTOLIC - BMI
- DIASTOLIC - BMI
- TRI - BMI
- HDL - BMI
- LDL - BMI
- FRAILTY - BMI
- MICROALB - N\_OP (Número de visitas a consultorios externos)
- FRAILTY - N\_OP
- BMI - N\_OP



El análisis demostró que hay combinaciones de valores que son más frecuentes que otras, pero no mostró ninguna asociación clara.

### 6.3. Valores cero

El análisis de `pandas_profiling` reporta que las siguientes variables presentan gran número de valores cero:

- `N_IP` (número de internaciones): Esto resulta razonable teniendo en cuenta que la gran mayoría de los enfermos de diabetes, a pesar de que tengan alguna complicación, no suelen requerir internación.
- `FRAILTY`: El valor cero indica que el paciente no tiene una salud frágil; considerando que esto es razonable considerar que aproximadamente un tercio de los pacientes tengan una salud robusta a pesar de padecer una enfermedad crónica. En todo caso sería necesario tener a disposición la metodología de cálculo del índice de fragilidad para poder efectuar un análisis similar al realizado con la obesidad y el bajo peso.

### 6.4. Detección de outliers

Ya teniendo las variables convertidas, se realizó la evaluación de la cantidad de outliers por separado para las correlaciones a 5 y 10 años.

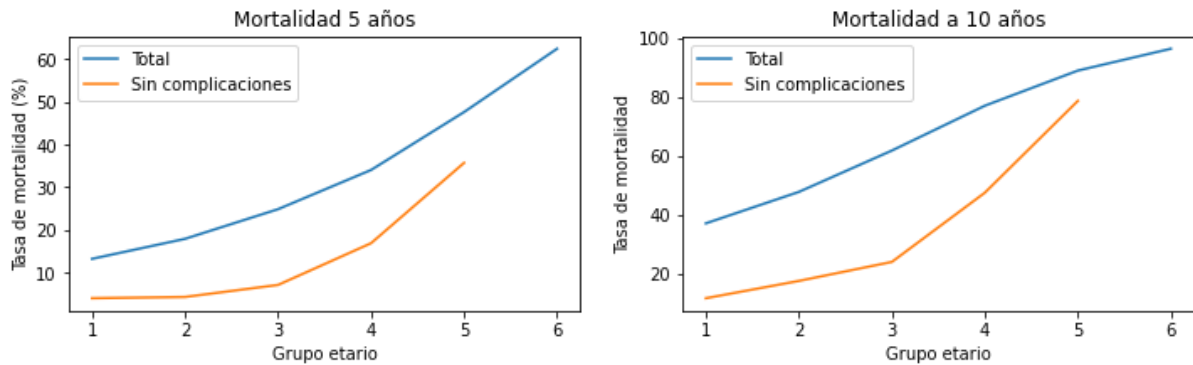
En ambos casos se trabajó haciendo 10 rondas de entrenamiento y selección de outliers con los métodos **Isolation Forest**, **Minimum Covariance Determinant**, **Local Outlier Factor** y **One-Class SVM**. En cada ronda se registró el error medio absoluto de la regresión con los datos crudos y los datos sin los outliers detectados por cada método. Al final se revisó si algún método era consistentemente mejor para eliminar los outliers.

Los resultados fueron interesantes en el sentido de que el método más eficaz para detectar outliers en la correlación para predicción a 5 años (**OneClassSVM**) no fue el mismo que para la predicción a 10 años (**IsolationForest**). Además, aunque la cantidad de outliers detectados fue prácticamente igual (2678 vs 2679) la reducción en el MAE obtenida terminó siendo muy diferente para ambas correlaciones (>3.5% para la predicción a 5 años contra <0,03% para la predicción a 10 años).

De todas maneras se eliminaron los outliers respectivos de ambas correlaciones.

## 7. Cálculo de las tasas de mortalidad

Una vez eliminados los outliers, se calculó la tasa de mortalidad para cada grupo etario en general y únicamente para los pacientes sin complicaciones:



Claramente se puede apreciar que la tasa de mortalidad para los pacientes sin complicaciones es menor que para la población en general.

## 8. Análisis de componentes principales

Considerando que la cantidad de variables hasta el momento resulta muy elevada, se realizó un PCA para ver si era posible reducir la complejidad del modelo. Para el análisis se escalieron las variables usando **StandardScaler** y se evaluó por separado si se afectaba el número de componentes manteniendo las variables de alta correlación detectadas anteriormente o quitándolas de acuerdo a lo indicado en el punto 6.1.

Los resultados fueron los siguientes:

	Componentes necesarios para explicar...	
	60% de la varianza	80% de la varianza
Con variables de alta correlación	31	46
Sin variables de alta correlación	29	45

Se concluye que en este caso el PCA fue insuficiente para poder reducir satisfactoriamente la complejidad del modelo.

## 9. Selección de variables significativas

Se realizó la selección de variables significativas por separado para las correlaciones de predicción a 5 y 10 años. Para el proceso se compararon los resultados de dos métodos:

- Por un lado se utilizó el método Stepwise, con un p-valor de 0.05 como límite para considerar significativa a cada variable.
- Por el otro se utilizó el método **BorutaPy**. Este método utiliza un **RandomForestClassifier** cuyos parámetros se convierten en los hiperparámetros del método. Como **BorutaPy** requiere que la variable objetivo tenga valores discretos, se la escaló multiplicando cada valor por 100000 a fin de retener 5 cifras significativas.

Los conjuntos de variables devueltos por cada método fueron los siguientes:

a) Método Stepwise, predicción a 5 años:

A1C	AGE	BIGUAN	BMI	CAD
CHF	DIASTOLIC	FRAILTY	HIV	INSULIN
LDL	LYMPHOMA	MARRIED	MICROALB	NEUROOTHER
PARALYSIS	PHTN	PRIORITY	PULMONARY	RACE
SERUMALB	SEX	SMOKER	SULF	SYSTOLIC
TUMOR	TZD			

b) Método Stepwise, predicción a 10 años:

A1C	AGE	BIGUAN	BMI	DEPRESSION
DIASTOLIC	INSULIN	MARRIED	PHTN	PRIORITY
RACE	SEX	SMOKER	TZD	VALVULAR

c) Método BorutaPy, predicción a 5 años:

AGE	ARRHYTHMIA	BIGUAN	BMI	CHF
DIASTOLIC	FRAILTY	MARRIED	PRIORITY	SMOKER
TRI				

d) Método BorutaPy, predicción a 10 años:

A1C	AGE	ARRHYTHMIA	BIGUAN	BMI
CHF	DIASTOLIC	FLUIDSLYTES	FRAILTY	HYPOTHYROID
LDL	MARRIED	PRIORITY	PVD	RACE
RENAL	SMOKER	TRI	TUMOR	VALVULAR

Los resultados muestran que cada método seleccionó un número diferente de variables con cada conjunto de datos. Se decidió que luego de optimizar los

hiperparámetros de las correlaciones se decidirá qué conjunto de variables aplicar a cada caso.

## 10. Aplicación de algoritmos de ML

Para calcular las regresiones se decidió emplear los siguientes algoritmos:

- Stochastic Gradient Descent (**SGDRegressor**)
- Support Vector Machines (**LinearSVR**)

La ventaja de utilizar estos algoritmos respecto de otros como los basados en RandomForest es que devuelven el coeficiente que aplica a cada variable en la regresión lineal, por lo que es posible comparar estos para pacientes con comorbilidades o sin comorbilidades a fin de obtener las conclusiones que buscamos.

Se calcularon training sets del 33% de los datos por separado para la correlación a 5 años y a 10 años, y se los mantuvo a partir de este momento hasta el fin del trabajo.

Se aplicaron ambos algoritmos (sin optimizar) a los cuatro conjuntos de variables obtenidos anteriormente, y se obtuvieron las siguientes métricas:

Correl.	Método	Variables	$R^2$	$AR^2$	MAE	RMSE
5 años	SGDRegressor	Stepwise	0.951469	0.951462	0.015293	0.020558
		BorutaPy	0.949764	0.949761	0.013602	0.020917
	LinearSVR	Stepwise	0.911165	0.911152	0.018904	0.027815
		BorutaPy	0.897135	0.897128	0.026700	0.029931

Correl.	Método	Variables	$R^2$	$AR^2$	MAE	RMSE
10 años	SGDRegressor	Stepwise	0.992612	0.992612	0.011434	0.013265
		BorutaPy	0.992706	0.992705	0.011608	0.013181
	LinearSVR	Stepwise	0.992554	0.992554	0.011930	0.013317
		BorutaPy	0.983758	0.983757	0.016019	0.019669

Luego se corrió **GridSearchCV** para cada algoritmo y conjunto de datos, con lo que se encontraron valores optimizados para los hiperparámetros de cada algoritmo en cada contexto.

Al correr las regresiones con hiperparámetros optimizados, las métricas obtenidas fueron las siguientes:

Correl.	Método	Variables	R <sup>2</sup>	AR <sup>2</sup>	MAE	RMSE
5 años	SGDRegressor	Stepwise	0.951453	0.951446	0.014766	0.020562
		BorutaPy	0.951368	0.951366	0.014808	0.020580
	LinearSVR	Stepwise	0.951408	0.951400	0.014654	0.020571
		BorutaPy	0.951368	0.951365	0.014796	0.020580

Correl.	Método	Variables	R <sup>2</sup>	AR <sup>2</sup>	MAE	RMSE
10 años	SGDRegressor	Stepwise	0.994032	0.994031	0.010467	0.011923
		BorutaPy	0.994025	0.994024	0.010476	0.011930
	LinearSVR	Stepwise	0.994032	0.994031	0.010455	0.011923
		BorutaPy	0.994025	0.994024	0.010466	0.011930

Los resultados muestran que ambos algoritmos alcanzan métricas muy similares; sin embargo se observó que **SGDRegressor** fue mucho más lento que **LinearSVR**, así que finalmente se decidió mantener este último y usar para cada correlación el conjunto de variables más parsimonioso (para la correlación a 5 años el devuelto por BorutaPy y para la correlación a 10 años el devuelto por Stepwise).

## 11. Impacto de cada variable en la tasa de mortalidad

Una vez definidos los algoritmos y conjuntos de variables, se filtraron los conjuntos de datos para separar aquellos pacientes con y sin comorbilidades y se calcularon las cuatro regresiones a fin de obtener los coeficientes de cada variable para cada conjunto de datos.

Luego se calculó un factor R de riesgo definido como el valor absoluto del cociente entre el coeficiente de cada variable para los pacientes sin comorbilidades respecto del correspondiente para los pacientes con comorbilidades. El significado de este coeficiente es que cuando una persona sin comorbilidades tiene un valor mayor de esta variable, su tasa de mortalidad tiende a subir respecto del resto de la población sin comorbilidades.

Como los signos de los coeficientes no son todos positivos, y no se mantienen entre las dos correlaciones evaluadas al mismo tiempo, también se calculó un indicador “‡” que muestra si el cambio de coeficientes es perjudicial (pasa de negativo para los pacientes con comorbilidades a positivo para los pacientes sin ellas) o “♥” si es “beneficioso” (el cambio es a la inversa). En el primer caso, el coeficiente que matemáticamente tendía a bajar la tasa de mortalidad pasa a hacerla crecer, por lo que la variable sería un predictor más fuerte de la tasa de mortalidad final. En el

segundo caso, la variable tendería a disminuir el valor de la tasa de mortalidad calculada.

Los coeficientes encontrados son los siguientes:

*Correlación a 5 años*

<b>Variable</b>	<b>Relación R</b>	<b>Indicador</b>
Arritmia cardíaca	43.40	‡
Uso de Biguanidas	6.53	‡
IMC	2.93	‡
Enf. cardíaca congestiva	58.92	
Tabaquismo	7.27	
Triglicéridos en sangre	6.17	
Índice de fragilidad (1)	3.27	
Presión diastólica	2.09	
Edad	0.93	
Prioridad de inclusión	0.39	
Casado/a	1.68	♥

*Correlación a 10 años*

<b>Variable</b>	<b>Relación R</b>	<b>Indicador</b>
Problemas de circulación pulmonar	253.99	‡
Uso de insulina	63.96	‡
Uso de Tiazolidinedionas	48.76	‡
Uso de Biguanidas	40.33	‡
IMC	5.69	‡
Sexo	5.60	‡
Hemoglobina A1C	362.50	
Enfermedad valvular	92.36	
Tabaquismo	76.87	
Depresión clínica	59.26	
Prioridad de inclusión	1.72	
Raza	1.65	
Presión diastólica	1.55	
Edad	0.99	
Casado/a	0.39	

En ambos casos, las variables con valores cercanos a 1 tienen aproximadamente la misma influencia para los dos conjuntos de pacientes, y cuanto mayor el valor más atención debería dársele en un paciente sin comorbilidades. Entre todas las variables, se les debería prestar más atención a aquellas con indicador “‡” (aunque

en el caso de “Sexo” para la correlación a 10 años sólo refleja que la esperanza de vida para los varones es menor que para las mujeres.

## 12. Conclusiones

Llegados a este punto estamos preparados para poder responder las preguntas que nos habíamos formulado al inicio del trabajo.

Como se puede ver en el punto [Cálculo de las tasas de mortalidad](#), se puede concluir que hay una diferencia notoria entre la tasa de mortalidad a 5 y 10 años de los pacientes diabéticos ancianos con o sin comorbilidades, teniendo este último grupo una mejor sobrevida en ambos casos.

También hemos encontrado qué variables permiten explicar satisfactoriamente las tasas de mortalidad observadas y hemos encontrado que hay claras diferencias en el impacto que tienen las mismas en pacientes con comorbilidades y aquellos que no las presentan.

Este conocimiento permitiría realizar programas de seguimiento de estas variables en pacientes diabéticos añosos, concentrando esfuerzo y recursos en atender con prontitud aquellas comorbilidades que tienen el mayor impacto real sobre la esperanza de vida del paciente.