TÉCNICAS MATEMÁTICAS PARA DIAGNOSIS MÉDICA

TRABAJO FIN DE GRADO

Curso 2021/2022

UNIVERSIDAD COMPLUTENSE
MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

GRADO EN MATEMÁTICAS

Alumna: Alexia Cazón López

Tutora: Ana Carpio Rodríguez

Madrid, 14 de Febrero de 2022

ÍNDICE

Índice de tablas 3

Ecuaciones y fórmulas 3

Resumen 4

Abstract 4

1. Introducción 5

1.1. Objetivos y plan de trabajo 5

2. Herramienta para la evaluación del riesgo de desarrollar cáncer de mama 6

2.1. Package BCRA 7

2.2. Análisis de los estudios de casos y controles 15

2.3. Modelo logístico 18

2.4. Adaptación del modelo logístico al formulario del NCI 20

2.5. Ecuación para predecir el riesgo absoluto de cancer de mama 23

3. Análogo covid-19 26

3.1. Estudio en 48 440 pacientes adultos sobre el covid-19 26

3.1. Muestras 26

3.2. Resultados e interpreatciones 27

3.3. Análisis estadístico 30

Conclusiones 33

Bibliografía 34

ÍNDICE DE TABLAS

Tabla 1: Función Recodificadora 11
Tabla 2: Ejemplo práctico paquete BCRA 13
Tabla 3: Resultados ejemplo práctico 14
Tabla 4: Tabla 2x2 inicial 16
Tabla 5: Tabla 2x2 17
Tabla 6: Tabla 1 2x2 para dos niveles de un factor de riesgo 17
Tabla 7: Tabla 2 2x2 para dos niveles de un factor de riesgo 18
Tabla 8: Tabla 2xk 18
Tabla 9: Parámetros estimados para el modelo del NCI 21
Tabla 10: Tabla de los riesgos relativos 21
Tabla 11: Riesgo específico basal 24
Tabla 12: Datos de los 48 440 sujetos 27
Tabla 13: Actividad física como factor de riesgo y sus OR 28
Tabla 14: La edad como factor de riesgo y sus OR 28
Tabla 15: El género como factor de riesgo y sus OR 29
Tabla 16: La raza como factor de riesgo y sus OR 29
Tabla 17: Ejemplo varón de 81 años 32

ECUACIONES Y FÓRMULAS

FÓRMULA 1: Número de muestras, función estimadora del riesgo relativo 12
FÓRMULA 2: Odd ratio 16
FÓRMULA 3: Tasa de incidencia 16
FÓRMULA 4: Riesgo atribuíble 17
FÓRMULA 5: Probabilidad logarítmica 19
FÓRMULA 6: Probabilidad de padecer una enfermedad 20
FÓRMULA 7: Proyección de la probabilidad de desarrollar cáncer de mama 23
FÓRMULA 8: Riesgo específico basal 23
FÓRMULA 9: Factor atribuíble 23
FÓRMULA 10: Probabilidad de sobrevivir a los factores de riesgo 24

RESUMEN

Se quiere conseguir un modelo matemático para la realización de formularios fiables para diagnosticar una enfermedad. Expondremos cómo la regresión estadística es una fuerte herramienta para el análisis y relación de diversas variables. Más allá, se aspira a presentar fórmulas y algoritmos matemáticos esenciales en el proceso de la diagnosis médica.

Los estudios de casos y controles sirven para obtener los odd ratios y sus riesgos absolutos y relativos correspondientes. Nos basaremos en las probabilidades que nos proporcionan los datos de estos estudios para llegar a un modelo logístico en términos de probabilidades logarítmicas.

Inicialmente, consideramos datos relacionados con el cáncer de mama. A partir de los métodos mencionados anteriormente podemos considerar igualmente datos COVID-19. Llegamos a formularios análogos consiguiendo relacionar ciertas circunstancias y causas con peores pronósticos. Finalmente, se pueden comprobar y testar datos de pacientes provenientes de bases de datos.

PALABRAS CLAVE: cáncer de mama, COVID-19, odd ratios, riesgo relativo y absoluto y formularios.

ABSTRACT

We want to achieve mathematical models to create reliable forms for medical diagnosis. We pretend to exhibit why regression analysis is a strong analytic tool to estimate and relate a set of variables. Beyond that, we aim to present essential mathematical formulas and algorithms for the medical diagnosis process.

Case and control studies are useful to obtain odd ratios and the corresponding absolute and relative risks. We base our report on the probabilities obtained from the data belonging to those studies to lead to a logistic model in terms of logarithms probabilities.

We start our study considering breast cancer data. Based on the previously mentioned methods, we can also tackle covid data. We obtain similar forms, reaching relations between some circumstances and causes with worst outcomes. From this, we can verify and test data patients from databases.

KEY WORDS: breast cancer, COVID-19, odd ratios, relative and absolute risks and forms.

1. INTRODUCCIÓN

En medicina, el diagnóstico es el proceso por el cual se identifica una enfermedad o condición de salud a partir de los signos, síntomas, historia clínica y examen físico del paciente. Cuando se quiere empezar a usar una nueva técnica de diagnosis se debe de llevar a cabo una validación de la misma, teniendo en cuenta la exactitud, fiabilidad, precisión y concordancia como cualidades imprescindibles. Esto se conseguirá mediante el seguimiento de los factores de riesgo de estas enfermedades y a través de los estudios de casos y controles. Para ello, se hace uso de los odds ratio que sirven para sobreestimar los riesgos relativos, medida que mejor estima el riesgo real y que nos permite deducir varias causas de una enfermedad.

El cáncer de mama es una patología que lleva siendo investigada muchos años, es por ello, que tenemos a nuestra disposición una gran cantidad de datos acerca de sus factores de riesgo y como consecuencia una amplia cantidad de resultados, hablando en términos de riesgos relativos y absolutos. Por tanto, constituye un buen modelo de referencia para la creación y diseño de formularios que podremos extrapolar para otras enfermedades.

Los modelos matemáticos, en particular los estadísticos que hemos introducido antes, van de la mano en el proceso de este juicio clínico. En esta memoria se demostrará cómo los cambios, que se producen en ciertas variables, van a influir sobre el valor que tome otra variable, por ejemplo, tener un peor pronóstico por COVID-19.

1.1. OBJETIVOS Y PLAN DE TRABAJO

Estudio de formularios fiables ya creados y publicados.

Presentación de la metodología, las fórmulas y los algoritmos que siguen estas técnicas de diagnosis.

Ejemplificar como funcionan, así como los resultados que ofrecen.

Análisis de otra enfermedad y sus factores de riesgo.

Exposición de los datos y modelos que nos permiten el estudio de esta nueva enfermedad.

Interpretación de los nuevos resultados y ejemplificaciones.

Diseño final de un formulario análogo al estudiado inicialmente.

2. HERRAMIENTA PARA LA EVALUACIÓN DEL RIESGO DE DESARROLLAR CÁNCER DE MAMA

El NCI (National Cancer Institute) desarrolla un formulario para estimar la probabilidad de desarrollar cáncer de mama a mujeres. Los profesionales a través de esta herramienta evalúan el riesgo de desarrollar este cáncer durante los 5 próximos años de vida y el riesgo de desarrollarlo de por vida (NCI, 2017).

Consiste en una calculadora interactiva que se basa en un modelo estadístico conocido como modelo de Gail, que lleva el nombre del Dr. Mitchell Gail, investigador del NCI. Este modelo ha sido probado y validado en grandes poblaciones de mujeres y se ha demostrado que proporciona estimaciones precisas del riesgo de padecer cáncer de mama. Además, estudios independientes han comprobado que es falible en el sentido de que el número de mujeres que predice que pueden tener este cáncer se acerca mucho al número real al que el tiempo demuestra en estas mujeres (Gail, et al., 2014).

El formulario está dirigido a mujeres blancas, negras, hispanas y asiáticas e isleñas del Pacífico en los Estados Unidos. Los resultados pueden ser inexactos en mujeres nativas de Alaska debido a una cantidad de datos limitados.

Por otro lado, la herramienta no puede estimar con precisión el riesgo de cáncer de mama para:

- Mujeres portadoras de una mutación productora de cáncer de mama en BRCA1 o BRCA2.
- Mujeres con antecedentes de cáncer de mama invasivo o in situ (carcinoma lobulillar in situ o carcinoma ductal in situ).
- Mujeres con antecedentes médicos de cáncer de mama de tipo DCIS o LCIS.
- Mujeres con antecedentes de síndromes raros que causan cáncer de mama como el síndrome de Li-Fraumeni.
- Mujeres que recibieron radiación para el tratamiento del linfoma de Hodgkin tienen un riesgo de cáncer de mama superior al promedio.
- Mujeres que recibieron Tratamiento con radiación al tórax.

Los resultados se basan en las respuestas a las siguientes preguntas:

1. ¿La mujer tiene antecedentes médicos de cáncer de mama o de carcinoma ductal in situ (DCIS) o carcinoma lobulillar in situ (LCIS) o ha recibido radioterapia previa en el tórax para el tratamiento del linfoma de Hodgkin?

2. ¿Tiene la mujer una mutación en el gen BRCA1 o BRCA2, o un diagnóstico de un síndrome genético que puede estar asociado con un riesgo elevado de cáncer de mama?
3. ¿Qué edad tiene la paciente?
4. ¿Cuál es la raza / etnia de la paciente?
    4.1. ¿Cuál es la subraza / etnia o lugar de nacimiento?
5. ¿La paciente ha tenido alguna vez una biopsia de mama con un diagnóstico benigno (no canceroso)?
    5.1. ¿Cuántas biopsias de mama con diagnóstico benigno ha tenido la paciente?
    5.2. ¿La paciente ha tenido alguna vez una biopsia de mama con hiperplasia atípica?
6. ¿Cuál era la edad de la mujer en el momento de su primer período menstrual?
7. ¿Qué edad tenía la mujer cuando dio a luz a su primer hijo?
8. ¿Cuántos de los parientes de primer grado de la mujer (madre, hermanas, hijas) han tenido cáncer de mama?

Como se puede observar, este formulario tiene muy en cuenta a quién se dirige y, además, proporciona distintas precisiones dependiendo de cada grupo étnico e intervalo de edad.

Para obtener los resultados, el NCI se apoya en programas estadísticos como el Macro SAS y el paquete R que proyectan el riesgo absoluto de cáncer de mama invasivo para mujeres, teniendo como información las respuestas a las preguntas anteriores. A continuación, se especifica en que consiste el último de ellos. Pero antes definamos algunos conceptos.

¿Qué es el riesgo absoluto? Cuando hablamos de riesgo absoluto nos referimos a la probabilidad de diagnosticarle a una mujer, que no tenga cáncer de mama, este tumor en un intervalo de tiempo definido $(T1, T2]$ dependiendo de unos factores de riesgo. A diferencia, el riesgo relativo se refiere a la probabilidad que tiene una mujer, de cierta edad y con ciertos factores de riesgo de manifestar este tumor, en comparación con la de una mujer de esa edad sin factores de riesgo.

2.1. PACKAGE BCRA

R es un software para computación estadística y gráfica. El "Package BCRA" es un paquete R que proyecta el riesgo absoluto de desarrollar cáncer de mama invasivo (Gail & Zhang, 2019).

La función principal de este programa es la función estimadora del riesgo absoluto, que se basa en el modelo estadístico de Gail que desarrollaremos en el siguiente apartado. Los parámetros necesarios para esta función aparecen especificados más abajo y necesitan ser recodificados con la función recodificadora.

Las constantes que influyen en el cálculo son: Riesgo atribuible, Coeficiente de regresión, Incidencias del cáncer de mama y Mortalidad del cáncer de mama. Estos factores son necesarios para la lista de constantes requeridas para calcular el riesgo absoluto y la función del riesgo relativo. Con los factores de riesgo mencionados y las edades del intervalo de proyección, la función estimadora del riesgo absoluto devolverá las correspondientes proyecciones de riesgo absoluto. Este paquete también define funciones que buscan errores en la entrada de datos que no especificaremos (BCRA, 2018).

Se usa el SEER (NCI Surveillance, Epidemiology, and End Results Program) para obtener datos necesarios para el cálculo.

Función estimadora del riesgo absoluto.

Argumentos de entrada:

1. Datos:
    - ID de la mujer: entero positivo.
    - T1, edad inicial: de 20 a 90 años.
    - T2, edad de proyección: de 20 a 90 años.
    - N_Biop, número de biopsias realizadas: entero positivo (suponemos el 0 entero).
    - HypPlas, biopsias que mostraron hiperplasia atípica: (0=no, 1=si, 99=NS/NC).
    - AgeMen, edad de la menarquía (primera menstruación): menor o igual a la edad inicial, 99=NS/NC.
    - Age1st, edad del primer hijo: mayor o igual a la edad de menarquía y menor o igual a la edad inicial.
    - N_Rels, número de familiares de primer grado con historial de cáncer de mama: entero positivo o 99=NS/NC.
    - Raza:
        1. Blanca.
        2. Africanas Americanas.
        3. Hispanas Americanas nacidas en EE. UU.
        4. Otras nativas de América de raza desconocida.
        5. Hispanas Americanas nacidas fuera.
        6. Chinas Americanas.
        7. Japonesas Americanas.
        8. Filipinas Americanas.

9. Hawaianas Americanas.
10. Otras isleñas del Pacífico.
11. Otras asiáticas.

2. Fila indicadora: 1 si los riesgos relativos están en el formato original y 0 si han sido recodificados a los valores 0, 1, 2 o 3.
3. Cálculo indicador: 0 para calcular el riesgo absoluto y 1 para calcular el riesgo absoluto promedio en mujeres blancas no hispanas y otras mujeres nativas de América.

Valor de salida:

Vector que te devuelve los riesgos absolutos si el cálculo indicador vale 1 o los riesgos relativos si el indicador vale 0.

Riesgo atribuible y Coeficiente de regresión:

1. Wh.Gail: blancas.
2. AA.CARE: Africanas Americanas.
3. HU.Gail: Hispano Americanas nacidas en EE. UU.
4. NA.Gail: nativas de América.
5. HF.Gail: Hispanas Americanas nacidas fuera de EE. UU.
6. Asian.AABCS: Asiáticas Americanas.

Incidencias del cáncer de mama y mortalidad del cáncer de mama:

1. Wh.1983_87: blancas, SEER 1983-1987
2. AA.1994_98: Africanas Americanas, SEER 1994-1998.
3. HU.1995_04: Hispanas Americanas (nacidas en EE. UU.), SEER 1995-2004.
4. NA.1983_87: Nativas de América y otras razas no conocidas, SEER 1983-1987.
5. HF.1995_04: Hispanas Americanas (nacidas fuera de EE. UU.), SEER 1995-2004.
6. Ch.1998_02: China-American, SEER 1998-2002.
7. Ja.1998_02: Japonesas Americanas, SEER 1998-2002.
8. Fi.1998_0: Filipinas Americanas SEER, 1998-2002.
9. Hw.1998_02: Hawaianas, SEER 1998-2002.
10. oP.1998_02: otras isleñas del Pacífico, SEER 1998-2002.
11. oA.1998_02: otras asiáticas, SEER 1998-2002.
12. Wh_Avg.1992_96: blancas, SEER 1992-1996.

Lista de constantes requeridas para el riesgo absoluto.

Argumentos de entrada:

1. Incidencias del cáncer de mama.
2. Mortalidad del cáncer de mama.
3. Coeficiente de regresión derivado del modelo de Gail.
4. Riesgo atribuible.

Valor de salida:

Archivo de texto con los datos convenientes.

Función recodificadora.

TABLA 1: FUNCIÓN RECODIFICADORA

Covarianza	Valor inicial	Valor recodificado
Número de biopsias (NB_Cat)	• 0 o 99 • 1 • 2,3, …, 99	• 0 • 1 • 2
Edad de la menarquía (AM_Cat)	• 14 o mas • 12, 13 • 11 o menos	• 0 • 1 • 2
Edad primer hijo (AF_Cat)	• 19 0 menos • 20-24 • 25-29 • 30 o mas	• 0 • 1 • 2 • 3
Familiares de primer grado con BrCa (NR_Cat)	• 0 o 99 • 1 • 2 o mas y no 99	• 0 • 1 • 2
Número de hiperplasias en biopsias (NB_Cat)	• 0 y alguna biopsia • ≥1 y alguna biopsia	• 0 • 1
Edad de la menarquía en hispano americanas nacidadas en EEUU (AM_Cat)	• Cualquier edad	• 0
Edad primer hijo africo-americanas (AF_Cat)	• 19 o mas joven o 99 • 20-29 • 30 o mas	• 0 • 1 • 2
Familiares de primer grado con BrCa en hispanas americanas (NR_Cat)	• 0 o 99 • 1 o 2 • 3 o mas	• 0 • 1 • 2

Función estimadora del riesgo relativo.

Argumentos de entrada:

1. Datos introducidos en la función principal.
2. Fila indicadora de si los datos están recodificados o no.

Valor de salida:

1. RR_Star1: riesgo relativo para mujeres de edad < 50.
2. RR_Star2: riesgo relativo para mujeres de edad >= 50.
3. Nº muestras: Número de muestras de riesgo, denotado por $Nº\ MR$. Hay 3 niveles para NB_Cat, 3 para AM_Cat, 4 para AF_Cat y 3 para NR_Cat. Por tanto, el total sería:

FÓRMULA 1: NÚMERO DE MUESTRAS, FUNCIÓN ESTIMADORA DEL RIESGO RELATIVO

$$Nº\ MR = NB\_Cat * 3 * 3 * 4 + AM\_Cat * 3 * 4 + AF\_Cat * 3 + NR\_Cat * 1 + 1$$

Ahora que ya hemos desarrollado el programa veamos un ejemplo práctico de como funcionaría el paquete para los siguientes datos:

TABLA 2: EJEMPLO PRÁCTICO PAQUETE BCRA

ID	T1	T2	N_Biop	HypPlas	AgeMen	Age1st	N_Rels	Race
1	45	53	99	99	10	20	1	0
2	45	53	99	1	10	20	1	1
3	45	53	99	0	10	20	1	2
4	45	53	0	99	10	20	1	3
5	45	53	1	99	10	20	1	4
6	45	53	1	99	14	19	1	5
7	45	53	99	99	99	19	1	6
8	45	53	1	1	14	19	1	7
9	45	53	99	1	14	99	1	8
10	45	53	1	0	14	19	1	9
11	45	53	99	0	99	99	1	10
12	45	53	0	0	14	19	1	11
13	45	53	0	99	10	20	1	12
14	45	53	0	1	10	20	1	0
15	45	53	0	0	10	20	1	1
16	45	53	1	0	10	20	1	2
17	35	40	4	99	11	25	0	3
18	35	40	4	99	11	98	0	4
19	35	40	4	99	11	10	0	5
20	35	40	4	99	36	25	0	6
21	27	90	99	99	13	22	0	7
22	27	90	99	99	13	22	99	8
23	18	26	99	99	13	22	99	9
24	27	26	99	99	13	22	99	10
25	85	91	99	99	13	22	99	11
26	85	90	99	99	13	22	99	12

Los resultados para estos valores fueron los siguientes (NA: no atribuye riesgo por razones de falta de precisión):

TABLA 3: RESULTADOS EJEMPLO PRÁCTICO

ID 1	NA
ID 2	NA
ID 3	NA
ID 4	2.108
ID 5	4.441
ID 6	3.976
ID 7	1.249
ID 8	5.775
ID 9	NA
ID 10	3.906
ID 11	NA
ID 12	NA
ID 13	NA
ID 14	NA
ID 15	NA
ID 16	2.689
ID 17	0.678
ID 18	1.022
ID 19	NA
ID 20	NA
ID 21	8.827
ID 22	6.767
ID 23	NA
ID 24	NA
ID 25	NA
ID 26	NA

2.2. ANÁLISIS DE LOS ESTUDIOS DE CASOS Y CONTROLES

Para poder comprender el algoritmo que sigue este formulario debemos introducir los estudios de casos y controles. Un estudio de casos y controles es una investigación sobre la proporción en la que las personas seleccionadas, que han pasado una enfermedad específica (los casos) o que no tienen esa enfermedad (los controles), han estado expuestas a los posibles factores de riesgo, con el fin de evaluar la hipótesis de que uno o más de estos es una causa de la enfermedad (Breslow NE, 1980).

El objetivo principal de un estudio de casos y controles es proporcionar una estimación válida y precisa de al menos una relación de causa-efecto hipotética. La principal ventaja de estos estudios es que puede evaluar a la vez muchas hipótesis causales, hayan sido evaluadas antes o sean nuevas. Entre otras ventajas destacamos la eficiencia, la aplicación a enfermedades tanto poco habituales como comunes y la evaluación de causas raramente expuestas. Es por ello, que se escoge este tipo de investigación epidemiológica para conseguir información acerca de las causas de ciertos cánceres, en particular para esta calculadora interactiva del cáncer de mama.

Para adquirir una buena comprensión de como estos estudios permiten la estimación del riesgo relativo se necesita una descripción de como se toman muestras de los casos y los controles en la población. En el caso de los controles, tienen que representar una muestra aleatoria de sujetos que no tienen la enfermedad, aunque tengan riesgo de contraerla. Además, la muestra tiene que estar estratificada, en edad y sexo, por ejemplo, y que esté igual de distribuida que los casos. Los casos se identifican normalmente a través de un registro de cáncer u otro sistema que cubra una población bien definida.

Supongamos que p es la proporción de individuos expuestos a un factor de riesgo al principio del estudio. Sea $P_1 = P_1(t)$ la probabilidad de que una persona expuesta en un estrato desarrolle la enfermedad durante un período de estudio de longitud t, y sea $P_0 = P_0(t)$ la cantidad análoga para los no expuestos. Sean $Q = 1 - P$ y $q = 1 - p$. Entonces, las proporciones esperadas de individuos de cada una de las cuatro categorías resultantes pueden representarse así:

TABLA 4: TABLA 2X2 INICIAL

	Expuestos	No expuestos	Total
Enfermos	pP_1	qP_0	$pP_1 + qP_0$
No enfermos	pQ_1	qQ_0	$pQ_1 + qQ_0$
Total	p	q	1

A partir de los datos de esta tabla o cualquiera similar de dimensiones 2 x 2, queremos estimar el riesgo relativo, que denotaremos por ψ. Para ello, necesitamos saber cuando podemos excluir la hipótesis de que el riesgo relativo es igual a la unidad, y determinar un rango de valores que sean coherentes con los datos observados. Presentamos las siguientes definiciones:

Definición. Sean P_1 y P_0 las probabilidades de contraer una enfermedad. Se define el término odd ratio, para el que no hay una traducción española comúnmente aceptada y que denotaremos por OR, como la posibilidad de que una condición de salud o enfermedad se presente en un grupo de población frente al riesgo que ocurra en otro. Se expresa como el cociente entre el odd ratio en el grupo con el factor de riesgo y el odd ratio en el grupo sin el factor.

FÓRMULA 2: ODD RATIO

$$\psi = OR = \frac{P_1/P_0}{Q_1/Q_0} = \frac{P_1 Q_0}{P_0 Q_1}$$

Concluimos que la incidencia de la enfermedad se puede estimar directamente a partir de un estudio de casos y controles y usaremos los odds ratio como una aproximación de los riesgos relativos.

Definición. La tasa de incidencia en un tiempo t se define como la tasa de crecimiento de $P(t)$ a través de la siguiente ecuación:

FÓRMULA 3: TASA DE INCIDENCIA

$$\lambda(t) = \frac{1}{1 - P(t)} * \frac{dP(t)}{dt}$$

Los estudios de casos y controles también estiman el incremento relativo asociado a una exposición, lo que nos lleva a definir lo siguiente:

Definición. Sea $r = \frac{\lambda_1}{\lambda_0}$ y p la proporción de personas en la población expuestas a un factor de riesgo, se define el riesgo atribuible de la población, AR, como:

FÓRMULA 4: RIESGO ATRIBUÍBLE

$$AR = \frac{p(r-1)}{pr + (1-p)}$$

Éste representa la proporción de casos que ocurren en la población total.

Una manera más fácil de visualizar las tablas 2x2 sería representarlas de la siguiente manera:

TABLA 5: TABLA 2X2

	Expuestos	No expuestos	Total
Enfermos	a	b	n_1
No enfermos	c	d	m_0
Total	m_1	m_0	N

En este caso los odds ratios vendrían representados por $\psi = \frac{ad}{bc}$. Estamos interesados en extender las relaciones y definiciones anteriores a tablas con mas niveles de exposición a un factor de riesgo. Esto se hace agrupando los datos en tablas 2x2. Por ejemplo, sea la exposición E con dos niveles de exposición C+ y C-, sus riesgos relativos serían $\psi_1 = \frac{a_1 d_1}{b_1 c_1}$ y $\psi_2 = \frac{a_2 d_2}{b_2 c_2}$ y sus tablas 2x2 correspondientes:

TABLA 6: TABLA 1 2X2 PARA DOS NIVELES DE UN FACTOR DE RIESGO

Nivel de factor C+	Exposición a E	No hay exposición a E	Total
Casos	a_1	b_1	n_{11}
Controles	c_1	d_1	n_{01}
Total	m_{11}	m_{01}	N_1

TABLA 7: TABLA 2 2X2 PARA DOS NIVELES DE UN FACTOR DE RIESGO

Nivel de factor C-	Exposición a E	No hay exposición a E	Total
Casos	a_2	b_2	n_{12}
Controles	c_2	d_2	n_{02}
Total	m_{12}	m_{02}	N_2

Supongamos que existen $K > 2$ niveles de exposición, los sujetos a los que estudiamos se pueden clasificar en una tabla del siguiente tipo:

TABLA 8: TABLA 2XK

Niveles de exposición	Casos	Controles	Total
1	a_1	c_1	m_1
2	a_2	c_2	m_2
…	…	…	…
k	a_k	c_k	m_k
Total	n_1	n_0	N

La manera de proceder en el análisis de datos de este tipo de tablas es elegir un nivel de exposición, digamos el nivel 1, como base para comparar con cada uno de los otros niveles, utilizando los métodos ya vistos para las tablas 2x2. De esta forma se obtienen riesgos relativos $r_1, r_2, \dots, r_k$ para cada nivel de exposición.

2.3. MODELO LOGÍSTICO

Debido a que deseamos basar nuestro análisis en el mayor número total de sujetos, utilizamos una regresión logística no condicionada que explicaremos con detalle en este apartado. (Breslow NE, 1980)

Los epidemiólogos especializados en cáncer recopilan datos sobre una serie de variables que influyen en el riesgo de contraer una enfermedad o de padecer una condición de salud. Cada combinación de diferentes niveles de estas variables define una categoría a la que se le va a hacer una estimación de la probabilidad de desarrollar la enfermedad o condición de salud. A menudo, nos enfrentamos al problema de tener que hacer estimaciones que utilizan información de otras categorías debido al limitado número de casos disponibles dentro de un periodo de tiempo y el gran número de categorías generadas combinando factores.

Dichas estimaciones se llevan a cabo en términos de un modelo, que relaciona el riesgo de padecer la enfermedad con varias combinaciones de factores de riesgo a través de un método matemático. Este modelo nos permite predecir el riesgo incluso para categorías en las que se dispone de escasa información. Se conoce como modelo lineal de regresión.

Definición. Sea P el riesgo de contraer la enfermedad, entonces la probabilidad logarítmica de este riesgo viene definida por:

FÓRMULA 5: PROBABILIDAD LOGARÍTMICA

$$y = logit(P) = \log\left(\frac{P}{1-P}\right)$$

Expresando P en términos de y nos quedaría: $P = \frac{\exp(y)}{1+\exp(y)}$

Sea $\beta = \log OR$, siendo OR los odds ratios definidos anteriormente. Si expresamos como $P(x)$ la probabilidad de padecer una condición de salud, siendo x una variable de regresión que vale 0 o 1 si hay exposición o no a un cierto factor de riesgo y $r(x) = \frac{P(x)Q_0}{Q(x)P_0}$ el riesgo relativo, entonces $logit\big(P(x)\big) = \alpha + \beta x$ con $\alpha = logit(P_0)$.

Podemos generalizar el modelo de manera que relacione una variable, que se denotará por $y = 1$ si contrae la enfermedad o por $y = 0$ si no la contrae, y una serie de K variables de regresión $x = (x_1, x_2, \dots, x_k)$ que valdrán 0 o 1 si están expuestas o no al factor de riesgo durante un periodo de proyección, a través de la ecuación:

$$pr(y = 1|x) = \frac{\exp(\alpha + \sum \beta_k x_k)}{1 + \exp(\alpha + \sum \beta_k x_k)}$$

O equivalentemente:

$$logit\big(pr(y=1|x)\big) = \alpha + \sum_{k=1}^{K} \beta_k x_k$$

Generalizando el modelo llegamos a que la probabilidad de que una persona dentro de una clasificación de niveles de factores de riesgo tenga la enfermedad viene dada por:

FÓRMULA 6: PROBABILIDAD DE PADECER UNA ENFERMEDAD

$$logit\big(P_i(x)\big) = \alpha_i + \sum_{k=1}^{K} \beta_k x_k$$

2.4. ADAPTACIÓN DEL MODELO LOGÍSTICO AL FORMULARIO DEL NCI

Estamos interesados en adaptar este modelo de regresión logística al formulario creado por el NCI para predecir los riesgos relativos de cada sujeto (Breslow NE, 1980).

Sea P_i la probabilidad de tener cáncer de mama de un sujeto en un estrato i con k factores de riesgo, la ecuación por la que se obtendrá los riesgos relativos asociados viene dada por:

$$\begin{aligned} logit(P_i) = \alpha_i + \sum_{k=1}^{7} \beta_k x_k = {} & 0.74948 + 0.09401\,(AgeMen) + 0.52926(NBiop) \\ & + 0.21863(Age1st) + 0.95830(NumRel) + 0.01081(AgeCat) \\ & - 0.28804(NBiop * AgeCat) - 0.19081(\text{Age1st X NumRel}) \end{aligned}$$

Los parámetros estimados para esta ecuación, que son los logaritmos de los odds ratio, vienen recogidos en la tabla 9. Derivan del modelo de regresión logístico no condicional y se obtienen los riesgos relativos correspondientes a través de la exponenciación de estos. Por ejemplo, para el número de biopsias será $exp(0.52926) = 0.09401$.

TABLA 9: PARÁMETROS ESTIMADOS PARA EL MODELO DEL NCI

	NBiop	AgeMen	Age1st	NumRel	AgeCat	NBiop x AgeCat	Age1st x NumRel
Parámetro	0.52926	0.09401	0.21863	0.95830	0.01081	-0.28804	-0.19081
Covarianza	0.956	0.001	-0.005	-0.015	0.201	-0.956	0.012
		0.161	0.000	-0.010	0.023	0.002	0.005
			0.128	0.174	-0.032	0.002	-0.106
				1.907	-0.010	0.020	-0.908
					0.418	-0.280	0.002
						1.244	-0.015
							0.534

A continuación, presentamos la tabla final con los riesgos relativos asociados. Dado que la categoría de menor exposición a los factores se utiliza como referencia para la comparación con otros grupos, usamos $r_1 = 1.000$. Para el cálculo del riesgo relativo de un sujeto en comparación con otro de la misma edad pero sin ningún factor de riesgo se localizan los riesgos relativos asociados para AgeMen, NBiops y la combinación de Age1st con NumRel y multiplicando los 3. Por ejemplo, para una mujer de 54 años con edad de menarquía a los 15, una biopsia previa, sin hijos y con un familiar de primer grado con cáncer de mama es $1.000 * 1.273 * 2.756 = 3.508$.

TABLA 10: TABLA DE LOS RIESGOS RELATIVOS

Factor de riesgo		Riesgo relativo	Nº de casos 2852	Nº de controles 3146
AgeMen	≥14	1.000	790	926
	12-13	1.099	1554	1735
	<12	1.207	508	485

NBiop	0		1.000	635	794
Age<50	1		1.698	113	93
	≥2		2.882	66	24
NBiop	0		1.000	1551	1817
Age≥50	1		1.273	312	300
	≥2		1.620	175	118
Age1st	<20	0	1.000	167	285
		1	2.607	44	40
		≥2	6.798	8	0
	20-24	0	1.244	708	1042
		1	2.681	208	123
		≥2	5.775	25	5
	25-29	0	1.548	986	1106
		1	2.756	247	178
		≥2	4.907	46	20
	≥30	0	1.927	307	291
		1	2.834	87	50
		≥2	4.169	19	6

2.5. ECUACIÓN PARA PREDECIR EL RIESGO ABSOLUTO DE CANCER DE MAMA

La probabilidad de que una mujer de edad a, con un factor de riesgo dependiente de la edad $r(t)$, desarrolle cáncer de mama en los próximos τ años viene dado por (Gail, et al., 2014):

FÓRMULA 7:PROYECCIÓN DE LA PROBABILIDAD DE DESARROLLAR CÁNCER DE MAMA

$$\Pr\{a,\tau,r(t)\} = \int_a^{a+\tau} h_1(t)r(t)\exp\left\{-\int_a^{\tau} h_1(t)r(u)du\right\}\{S_2(t)/S_2(a)\}dt$$

donde $h_1(t)$ es el riesgo específico basal de desarrollar cáncer de mama. Este se obtiene de las tasas de riesgo específico con una edad promedio, denotado por h_1^*, y usando que

FÓRMULA 8:RIESGO ESPECÍFICO BASAL

$$h_1(t) = h_1^*(t)F(t) \text{ y}$$

FÓRMULA 9: FACTOR ATRIBUÍBLE

$$F(t) = 1 - \text{AR} \text{ para la edad } t.$$

El factor $F(t)$ utilizado en el modelo fue 0,5229 para mujeres menores de 50 años y 0,5264 para mujeres de 50 años o más.

Los valores h_1^* vienen dados en la siguiente tabla:

TABLA 11: RIESGO ESPECÍFICO BASAL

Grupo de edad	h_1^*
20-24	2,7
25-29	16,8
30-34	60,3
35-39	114,6
40-44	203,7
45-49	280,8
50-54	320,9
55-59	293,8
60-64	369,4
65-69	356,1
70-74	307,8
75-79	301,3

Por último, la probabilidad de sobrevivir a los factores de riesgo a la edad t es la siguiente:

FÓRMULA 10: PROBABILIDAD DE SOBREVIVIR A LOS FACTORES DE RIESGO

$$S_2(t) = \exp\left\{-\int_0^{\tau} h_2\,(u)du\right\}$$

El riesgo de mortalidad por causas que no sean el cáncer se denota por h_2 y sus valores para la población estadounidense están disponibles en la base de datos del SEER (SEER, 1995-2020).

Para obtener el riesgo relativo $r(t)$ se multiplican 3 riesgos relativos específicos de las categorías A (edad de menarquía), B (número de biopsias y edad de la mujer) y C (número de familiares de primer grado afectados y edad del primer nacimiento vivo). Estos riesgos relativos para mujeres blancas aparecen

recogidos en la Tabla 10. Para otros grupos étnicos podemos encontrarlos en las referencias (Rayna K. Matsuno, 2011) y (Matthew P. Banegas, 2017) que amplían el método de Gail a mujeres hispanas, asiáticas y africanas.

3. ANÁLOGO COVID-19

Tras la llegada de una pandemia a todos los rincones del mundo y gracias a los avances médicos y tecnológicos disponibles, tenemos en nuestras manos una infinidad de datos y estudios que nos permiten la aplicación de modelos matemáticos para la mejora de la salud de la población mediante la prevención y asociación de causas y efectos en todo lo relacionado con la enfermedad que causa el virus SARS-CoV-2.

Nuestro objetivo ahora es extrapolar los métodos matemáticos especificados al COVID-19, teniendo a disposición bases de datos que nos faciliten el número de casos dependiendo de la edad, etnia y otras condiciones relacionadas con la salud de los sujetos. Ya sabemos que los estudios de casos y controles nos proporcionan una estimación válida y precisa de al menos una relación de causa-efecto hipotética, por ello, nos apoyaremos en este tipo de investigación y junto al modelo de regresión logístico, nos permitirá asociar distintas categorías con distintos riesgos relativos. Podemos aproximarnos a nuestra meta gracias a investigaciones que nos proporcionan estos datos necesarios, como es un estudio sobre el sedentarismo y su asociación con padecer complicaciones severas cuando se contrae el Covid-19. (Robert Sallis, 2021)

3.1. ESTUDIO EN 48 440 PACIENTES ADULTOS SOBRE EL COVID-19

Los centros para el control y la prevención de enfermedades de EE. UU. han estudiado factores de riesgo del COVID- 19, incluyendo la edad avanzada, el sexo (masculino) y la presencia de condiciones que afectan a la salud como diabetes, sobrepeso, enfermedades cardiovasculares… Este estudio busca incluir la inactividad física como factor de riesgo entre los últimos mencionados. Para ello, usa una regresión logística dependiente de varias variables para evaluar si la inactividad está asociada a consecuencias severas al padecer COVID-19.

Entre sus objetivos también debemos de nombrar el de concienciar a la población general de los beneficios del ejercicio físico, no sólo para este virus, si no ante la exposición de otras enfermedades.

3.1. MUESTRAS

Las muestras de los casos y controles se toman del KSPC, Kaiser Permanent Southern California (KPSC, 2021), sobre personas mayores de edad, que tengan un seguimiento de por lo menos 6 meses y con al menos 3 visitas ambulatorias en las que se haya registrado su nivel de actividad física antes de tener la prueba positiva en COVID-19. Así, el número total de muestras que se toman son de 48440 pacientes

que cumplieron estos requisitos. Se distinguen 3 categorías de actividad física dependiendo del tiempo que se dedique a ello por semana: inactividad, entre 0 y 10 minutos; actividad moderada, entre 11 y 149 minutos; y actividad constante, a partir de 150 minutos.

TABLA 12: DATOS DE LOS 48 440 SUJETOS

Patient characteristics by exercise level					
	Consistently inactive (n=6984)	Some activity (n=38 338)	Consistently meeting PA guidelines (n=3118)	Total (n=48 440)	P value*
Age at index date					<0.0001
Mean (SD)	49.4 (16.88)	47.8 (16.95)	40.6 (15.72)	47.5 (16.97)	
Median (Q1, Q3)	49 (36.0, 60.0)	47 (34.0, 60.0)	38 (27.0, 52.0)	47 (33.0, 60.0)	
Age group, n (%)					<0.0001
<60 years	5176 (14.3)	28 492 (78.4)	2652 (7.3)	36 320	
60–69 years	973 (14.2)	5585 (81.3)	313 (4.6)	6871	
70–79 years	433 (12.9)	2803 (83.4)	126 (3.7)	3362	
80+ years	402 (21.3)	1458 (77.3)	27 (1.4)	1887	
Gender, n (%)					<0.0001
Female	4244 (14.2)	24 284 (81)	1464 (4.9)	29 992	
Male	2740 (14.9)	14 053 (76.2)	1654 (9)	18 447	
Unknown	0 (0)	1 (100)	0 (0)	1	
Race/ethnicity, n (%)					<0.0001
Asian	365 (13)	2228 (79.7)	204 (7.3)	2797	
Black	476 (13.6)	2857 (81.8)	160 (4.6)	3493	
Hispanic	4734 (15)	25 007 (79.5)	1729 (5.5)	31 470	
Native American/Alaskan	9 (10.2)	75 (85.2)	4 (4.5)	88	
Pacific Islander	37 (12)	254 (82.2)	18 (5.8)	309	
White	1148 (13)	6873 (77.6)	835 (9.4)	8856	
Other	215 (15.1)	1044 (73.2)	168 (11.8)	1427	
BMI					<0.0001
Mean (SD)	32.2 (7.39)	31.3 (7.06)	28.2 (5.45)	31.2 (7.07)	
Median (Q1, Q3)	31.4 (27.3, 36.2)	30.2 (26.4, 35.1)	27.4 (24.5, 30.9)	30.2 (26.3, 35.0)	
BMI group, n (%)					<0.0001
<25 kg/m^2	1010 (11.9)	6521 (77)	933 (11)	8464	
25–29 kg/m^2	1895 (12.5)	12 025 (79.4)	1216 (8)	15 136	
30–39 kg/m^2	3141 (16)	15 652 (79.7)	842 (4.3)	19 635	
≥40 kg/m^2	936 (18)	4134 (79.6)	126 (2.4)	5196	
Smoking, n (%)					<0.0001
Ever	1558 (15.5)	8008 (79.6)	492 (4.9)	10 058	
Never	4084 (13.7)	23 882 (80)	1886 (6.3)	29 852	
Utilisations, clinical characteristics and comorbidities, n (%)					
Emergency encounters	1019 (14.5)	5702 (81.4)	287 (4.1)	7008	<0.0001
Inpatient encounters	317 (16)	1618 (81.8)	43 (2.2)	1978	<0.0001
Ever had organ transplant	12 (8.5)	129 (91.5)	0 (0)	141	0.0005
Pregnant at index date	184 (12.5)	1224 (83.4)	59 (4)	1467	<0.0001
Cardiovascular disease	689 (16.5)	3410 (81.6)	82 (2)	4181	<0.0001
COPD	788 (14.5)	4449 (81.7)	210 (3.9)	5447	<0.0001
Renal disease	459 (17.3)	2149 (81)	46 (1.7)	2654	<0.0001
Cancer	108 (12)	768 (85.4)	23 (2.6)	899	<0.0001
Metastatic cancer	47 (16.4)	232 (80.8)	8 (2.8)	287	0.0326
Hypertension	1682 (15.6)	8827 (81.7)	297 (2.7)	10 806	<0.0001
Diabetes, n (%)					<0.0001
A1C<7%	1849 (13.8)	10 813 (80.7)	733 (5.5)	13 395	
7%≤A1C<8%	316 (14.8)	1758 (82.6)	55 (2.6)	2129	
A1C≥8%	500 (16)	2566 (82)	63 (2)	3129	

3.2. RESULTADOS E INTERPREATCIONES

Como mencionamos antes, queremos concluir que hay relación entre el sedentarismo y peores consecuencias al contraer el COVID-19. Para ello, necesitamos calcular los odd ratios en las distintas

categorías en las que dividimos los sujetos según su nivel de actividad física y compararlo con resultados de otros factores de riesgo ya conocidos.

TABLA 13: ACTIVIDAD FÍSICA COMO FACTOR DE RIESGO Y SUS OR

	OR en hospitalizaciones	OR en admisiones UCI	OR en muertes
Inactividad	2.26	1.73	2.49
Actividad Moderada	1.89	1.58	1.88
Actividad constante	1	1	1

La tabla 13 arroja resultados que sugieren que los pacientes constantemente inactivos comparados con aquellos en la categoría de alguna actividad tenían menores probabilidades de hospitalización y muerte, por lo que cualquier cantidad de actividad física puede tener un beneficio. Sin embargo, no ser constante con el ejercicio proporcionó probabilidades sustancialmente más altas de resultados adversos al COVID-19 que cumplir constantemente las pautas de actividad física.

TABLA 14: LA EDAD COMO FACTOR DE RIESGO Y SUS OR

Edad	OR en hospitalizaciones	OR en admisiones UCI	OR en muertes
<60	1	1	1
60-69	2.30	2.40	4.01
70-79	3.72	3.44	10.40
+80	6.13	3.52	27.31

La edad siempre ha sido un factor de riesgo conocido, como vemos en el COVID-19 no es menos. Los odd ratios muestran que la edad es un factor relevante, de hecho, se puede confirmar que tener una edad avanzada está asociado con más muertes al padecer el virus. Estos resultados son mucho más indicativos que los que nos dan en las categorías de actividad física y va a ser el que nos da los resultados más claros.

TABLA 15: EL GÉNERO COMO FACTOR DE RIESGO Y SUS OR

Género	OR en hospitalizaciones	OR en admisiones UCI	OR en muertes
Masculino	1.82	2.38	1.72
Femenino	1	1	1

El género, factor que no pudimos tener en cuenta en el estudio del cáncer de mama, si lo podemos tomar como factor de riesgo en este estudio pues tenemos evidencia de mayores probabilidades de hospitalización y muerte en hombres que en mujeres. Esto conlleva un nuevo estrato que antes no habíamos considerado y que arroja resultados en los odd ratios significativos y similares a los de la categoría de actividad física.

TABLA 16: LA RAZA COMO FACTOR DE RIESGO Y SUS OR

Raza	OR en hospitalizaciones	OR en admisiones UCI	OR en muertes
Asiáticos	2.04	2.29	1.30
Negros	1.33	1.25	1.18
Hispanos	1.22	1.31	1.08
Nativos de EE. UU.	1.16	1.42	1.69
Isleños del Pacífico	1.69	1.77	2.53
Blancos	1	1	1

Estos datos también nos sugieren que la etnia es un factor a tener en cuenta para validar la hipótesis de que es causa de peores consecuencias cuando se padece COVID-19.

Otros factores de riesgo conocidos son altos niveles de índice de masa corporal, fumar y condiciones de salud como diabetes, trasplantes de órganos, embarazo, enfermedades cardiovasculares, cáncer, cáncer con metástasis, hipertensión, enfermedades renales y enfermedad pulmonar obstructiva crónica.

Podemos concluir que hemos encontrado evidencias de causa-efecto relacionadas con la cantidad de ejercicio que realice un sujeto por semana y la gravedad de las consecuencias al padecer este virus, pero éstas pueden estar asociadas a factores de riesgo ya comprobados como índice de masa corporal elevado y diabetes. Además, se echa en falta la medida de intensidad de la actividad física y la relación de esto con la edad avanzada que muchas veces impide su realización.

3.3. ANÁLISIS ESTADÍSTICO

Para poder llegar a las conclusiones anteriores hay que comparar distintos datos de las muestras para lo que nos hará falta el modelo de regresión. Este modelo incluye la información sobre la correlación entre diferentes factores de riesgo y permite el control estadístico de los factores más confusos para obtener inferencias más confiables.

Como incluimos en la tabla se toma como valor de referencia la mujer, blanca, menor de 60 años que realiza actividad constante y necesitamos $p = 3 + 1 + 5 = 9$ variables (tres categorías para la edad, una para el sexo y cinco para la raza) para representar todas las características de los sujetos. Así, la tasa de padecer el virus sigue el modelo ya visto antes (Gail, et al., 2014):

$$\log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) = \alpha + x\beta$$

siendo $y \in \{0,1\}$ dependiendo de si padece o no el virus y $x \in \{0,1\}$ si tiene algunos de los 9 factores de riesgo o no. Como ya hemos mencionado antes α es el logaritmo del odd ratio del valor de referencia y β los logaritmos de los odd ratios asociados al resto de categorías.

Expresando la última fórmula en términos de la probabilidad de contraer el virus al padecer un factor de riesgo x nos queda:

$$P(y=1|x) = \frac{\exp(\alpha + x\beta)}{1 + \exp(\alpha + x\beta)}$$

Para estimar el riesgo relativo para variables específicas con distintos valores $x_i \in X$, tal que $x_1 = c_i, \dots, x_i = c_i$ con $c_i \in \{0,1\}$, $i \in \{0, \dots, 9\}$, usamos la fórmula general:

$$P(y=1|x_1 = c_i, \dots, x_i = c_i) = \frac{\sum_{x_i \in X} P(y=1|x_i)P(x_i)}{\sum_{x_i \in X} P(x_i)}$$

Recordemos nuestra fórmula para el cálculo del riesgo absoluto:

$$\Pr\{a,\tau,r(t)\} = \int_a^{a+\tau} h_1(t)r(t)\exp\left\{-\int_a^{\tau} h_1(t)r(u)du\right\}\{S_2(t)/S_2(a)\}dt$$

Podemos observar que sólo nos faltan las constantes $h_1^*, h_1, F(t), AR, S_2$ y h_2 , correspondientes al riesgo específico promedio y basal, el factor dependiente del riesgo atribuible, la probabilidad de sobrevivir a otros factores de riesgo y el riesgo de mortalidad por causas que no sean COVID-19, para el cálculo final y estos se pueden conseguir en una base de datos de pacientes con esta patología y en personas sanas.

Para complementar este estudio podemos realizar un formulario parecido al del capítulo del cáncer de mama ya que a través de las respuestas a ciertas preguntas podríamos obtener los odd ratios necesarios y a partir de estos devolver el riesgo relativo y absoluto del paciente de tener un peor pronóstico al padecer COVID-19.

El formulario estaría dirigido a hombres y mujeres mayores de edad que quieran evaluar el riesgo absoluto y relativo que tienen de mortalidad, hospitalización e ingreso en UCI tras recibir un resultado positivo al realizarse un test de COVID-19. Las preguntas podrían ser las siguientes:

1. Señale en que franja de actividad física se encuentra el paciente:
    - 1.1. Entre 0-10 minutos por semana
    - 1.2. Entre 10-149 minutos por semana
    - 1.3. Más de 150 minutos por semana
2. ¿Qué sexo tiene el paciente?
    - 2.1. Si es mujer, ¿está embarazada la paciente?
3. ¿Qué edad tiene el paciente?
4. ¿Cuál es la raza / etnia del paciente?
5. ¿Ha fumado el paciente alguna vez?
6. ¿Cuál es el índice de masa corporal del paciente?
7. ¿Ha estado el paciente ingresado?
8. ¿Ha estado el paciente ingresado de emergencia?
9. ¿Ha recibido el paciente un trasplante de órganos?
10. Indique cual de las siguientes enfermedades padece el paciente:
    - 10.1. Enfermedad Cardiovascular
    - 10.2. Enfermedad Renal
    - 10.3. Cáncer
    - 10.4. Cáncer con metástasis

10.5. Hipertensión

10.6. Diabetes

10.7. Enfermedad Pulmonar Obstructiva Crónica

Veamos un ejemplo de un paciente con las siguientes respuestas:

TABLA 17: EJEMPLO VARÓN DE 81 AÑOS

Preguntas	Respuestas
Pregunta 1	0-10 minutos por semana
Pregunta 2	Hombre
Pregunta 3	81
Pregunta 4	Blanco
Pregunta 5	No fuma
Pregunta 6	<25
Pregunta 7	No
Pregunta 8	No
Pregunta 9	No
Pregunta 10	Ninguna

En este caso el individuo es un hombre blanco de 81 años que no realiza actividad física y cuyas características clínicas no son relevantes, no padece otras enfermedades ni otras condiciones de salud que hayamos visto que influyen en un peor pronóstico. Su riesgo relativo en comparación con otro individuo de la misma edad pero sin ningún factor de riesgo es $2.26 * 6.13 * 1.82 * 1 = 25.21$ y podemos observar la gran diferencia en comparación con un hombre con las mismas condiciones de salud pero de 59 años, cuyo factor de riesgo es 4,11.

CONCLUSIONES

En esta memoria, tras un profundo estudio de un formulario que nos servía para una calculadora interactiva de cáncer de mama (NCI, 2017) hemos explicado como desarrollar otro para el COVID-19 (Robert Sallis, 2021). La presentación y comprensión de las fórmulas y algoritmos que se han necesitado para el primero nos aseguran la fiabilidad del segundo, basándonos en los odd ratios de los factores de riesgo que se han tenido en cuenta para el COVID-19, entre los que destacamos entre otros la actividad física. Además, hemos observado las relaciones entre las distintas variables y cómo varían en los resultados finales de los riegos relativos, interpretándolos y concluyendo que la actividad física es de los factores que más influyen en peores pronósticos de COVID-19. Finalmente se expone un diseño de un nuevo formulario de la mano de fórmulas esenciales como son las de los odd ratios, la probabilidad de contraer el virus y la de la proyección en un rango de tiempo de la probabilidad de padecerlo.

BIBLIOGRAFÍA

BCRA, P. (2018). *Package BCRA*. Obtenido de Breast Cancer Risk Assessment: https://dceg.cancer.gov/tools/risk-assessment/bcra/bcra-manual.pdf

Breslow NE, D. N. (1980). *Statistical Methods in Cancer Research Volume I: The Analysis of Case-Control Studies*. Lyon: IARC W. DAVIS .

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mutvihill, J. J. (2014). Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. *Journal of the National Cancer Institue*. Recuperado el 4 de Abril de 2020, de http://jnci.oxfordjournals.org/

Gail, M., & Zhang, F. (2019). *BCRA R Package*. Recuperado el 2020, de https://dceg.cancer.gov/tools/risk-assessment/bcra

KPSC. (2021). *Kaiser Permanente Thrive*. Obtenido de Kaiser Foundation Health Plan: https://thrive.kaiserpermanente.org/care-near-you/southern-california/

Matthew P. Banegas, E. M. (2017). Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women. *ournal of the National Cancer Institute*, Volume 109.

Mitchell H. Gail, J. P.-S. (2007). Projecting Individualized Absolute Invasive Breast Cancer Risk in African American Women. *Journal Cancer Institute*, 1782–1792. Recuperado el 2020, de https://academic.oup.com/jnci/article/99/23/1782/993533#app1

NCI, N. C. (2017). *Breast Cancer Risk Assessment Tool*. Obtenido de Breast Cancer Risk Assessment Tool: https://bcrisktool.cancer.gov/calculator.html

Rayna K. Matsuno, J. P. (2011). Projecting Individualized Absolute Invasive Breast Cancer Risk in Asian and Pacific Islander American Women. *Journal of the National Cancer Institute*, 951–961.

Robert Sallis, Y. D. (2021). *Physical inactivity is associated with a higher risk for severe COVID-19 outcomes: a study in 48 440 adult patients*. California, USA: Br J Sports Med.

SEER. (1995-2020). *Surveillance, Epidemiology, and End Results Program*. Recuperado el 2020, de https://seer.cancer.gov/explorer/application.html?site=55&data_type=1&graph_type=2&compareBy=sex&chk_sex_3=3&chk_sex_2=2&race=1&age_range=1&stage=101&rate_type=2&advopt_precision=1&advopt_display=2