

Universidad de Costa Rica
Escuela de Matemática
Departamento de Matemática y Ciencias Actuariales

Bitácora

Jimena Chacón Chavarría B81999
Maria Jimena Marchena Mendoza B74425
Jorge Andreas Sanabria Arce B97190

Estadística Actuarial I

Profesor: Maikol Solís

II-2023

Entrada 1
31 de agosto del 2023

Parte de planificación

Pregunta de Investigación

Definición de la idea

Investigar la existencia de una correlación entre diferentes factores (como el índice de masa corporal (IMC), sexo, consumo de tabaco, etc...) y el padecer de una enfermedad cardiovascular.

Conceptualización de la idea

Según la Real Academia Española (RAE), se tienen las siguientes definiciones:

- Investigar: *Realizar actividades intelectuales y experimentales de modo sistemático con el propósito de aumentar los conocimientos sobre una determinada materia.*
- Existencia: *Acto de existir. (existir: Dicho de una cosa: Ser real y verdadera.)*
- Correlación: *Medida de la tendencia de la evolución de dos variables.*
- Diferentes: *Diverso, distinto.*
- Factores: *Elemento o causa que actúan junto con otros.*
- Índice de masa corporal (IMC): Según la Organización Mundial de la Salud, *es un índice simple de peso para la talla que se usa comúnmente para clasificar el sobrepeso y la obesidad en adultos. Se define como el peso de una persona en kilogramos dividido por el cuadrado de su altura en metros (kg/m^2).*
- Sexo: *Condición orgánica, masculina o femenina, de los animales y las plantas.*
- Tabaquismo: *Adicción al consumo de tabaco.*
- Padecer: *Sentir física y corporalmente un daño, dolor, enfermedad, pena o castigo.*
- Cardiovascular: *Perteneciente o relativo al corazón o al aparato circulatorio.*

Identificación de tensiones

- **Complejidad de los factores:** Los factores a considerar pueden estar interrelacionados de maneras complejas.
- **Variables y factores desconocidos:** Pueden existir otras variables o factores no considerados inicialmente que podrían influir en los resultados.
- **Correlación no implica causalidad:** Si se encuentra una correlación entre algunos de estos factores y enfermedades cardiovasculares, no se puede asumir automáticamente que uno cause el otro. Puede haber factores desconocidos o no considerados que estén influyendo en ambas cosas.
- **Datos:** Dependiendo de la fuente y la calidad de los datos utilizados para la investigación, podría haber sesgos en la muestra que afecten los resultados. Así como los cambios en el tiempo, contexto cultural y geográfico, si se utilizan datos de un periodo y territorio específico, la generalización de los resultados a diferentes poblaciones podría ser un desafío.

Reformulación de la idea en modo pregunta

- ¿Cómo se puede ver afectada la salud cardiovascular?
- ¿Por qué es importante conocer los factores que se relacionan con padecer una enfermedad cardiovascular?
- ¿Como se puede determinar la relación entre distintos factores y las enfermedades cardiovasculares?
- ¿Qué factores se relacionan con la incidencia de enfermedades cardiovasculares?

Argumentación de las preguntas

¿Por qué es importante conocer los factores que se correlacionan con padecer una enfermedad cardiovascular?

Argumentos

Lógica: El conocimiento e identificación temprana de los factores que se correlacionan con padecer una enfermedad cardiovascular permite tomar medidas preventivas para reducir el riesgo de una enfermedad cardiovascular y sus complicaciones.

Ética: La salud es un tema delicado, pero fundamental. Por eso, se debe proporcionar información precisa sobre los factores de riesgo de enfermedades cardíacas.

Emocional: Al conocer los factores de riesgo, las personas pueden tomar decisiones informadas y tomar el control para cuidar debidamente su bienestar, lo cual les puede brindar más seguridad y apoyo.

Contraargumentos

Lógica: Inclusive sabiendo los factores que se correlacionan con padecer una enfermedad cardiovascular, en algunas ocasiones es inexplicable e inevitable el desarrollo de estos padecimientos, ya sea por temas genéticos así como factores externos que no están considerados en el análisis.

Ética: El conocimiento popular de los factores que se correlacionan con las enfermedades cardíacas pueden crear prejuicios o una estigmatización hacia algún grupo de personas.

Emocional: Se podría argumentar que conocer los factores que se correlacionan con padecer una enfermedad cardiovascular podría generar ansiedad en aquellas personas que se preocupan por su salud y de esta manera provocar un rechazo a la realidad de su salud.

Conclusión

La realidad de la salud de las personas y asegurar que se comunique de una manera efectiva es esencial y beneficiosa para todos. Conocer y comprender los factores que pueden estar ligados a una enfermedad cardiovascular es necesario desde las perspectivas lógicas y éticas. Además, esta comprensión puede brindar un apoyo adicional a las personas, para que estas tomen medidas preventivas para un mayor bienestar y una mejor calidad de vida.

¿Cómo se puede determinar la relación entre distintos factores y las enfermedades cardiovasculares?

Argumentos

Lógica: Por medio de conceptos y métodos estadísticos, así como estudios científicos como base, es posible concluir una relación entre los distintos factores y las enfermedades cardiovasculares por medio de un análisis de los datos que permita identificar ciertos patrones, o bien, combinaciones de factores que aumenten la incidencia de enfermedades cardiovasculares.

Ética: Al utilizar datos recuperados de Estados Unidos, se asegura una base amplia y diversa para realizar un análisis completo de las correlaciones de los datos con respecto a las personas que padecen de enfermedades cardiovasculares.

Emocional: Es importante analizar la relación de variables tales como sexo y edad, entre otras, y la presencia de enfermedades cardiovasculares en distintos individuos dado que con esto se puede utilizar como base a estudios que ahonden más en el tema para ver cómo se correlacionan entre ellas.

Contraargumentos

Lógica: Es difícil determinar si una variable tiene realmente una correlación con las enfermedades cardiovasculares dado a diferencias de cada persona por temas genéticos u otros factores externos que impliquen una predisposición de padecer alguna enfermedad cardiovascular.

Ética: Se requiere una base de datos lo suficientemente amplia para no tener resultados incompletos así como conocimientos necesarios en estadística para poder hacer un análisis correcto de los resultados. Los modelos utilizados pueden tener errores o factores externos que no se contemplen por lo que arrojen resultados imprecisos.

Emocional: La falta de conocimiento estadístico de las personas utilizado en la investigación podrían provocar una omisión al tema o crear desconfianza.

Conclusión

Una determinación certera de la relación entre las enfermedades cardiovasculares y distintos factores es un proceso que combina todos los argumentos. La lógica científica es primordial dado a los métodos estadísticos que son posibles de utilizar para tener resultados precisos y congruentes a lo esperado. La ética en la investigación que se basa en un buen manejo de los datos y transparencia en los resultados, así como la necesidad de comunicar de manera efectiva los hallazgos para lograr un impacto real y beneficioso en el ámbito.

¿Qué factores se correlacionan con la incidencia de enfermedades cardiovasculares?

Argumentos

Lógica: Existen diversos factores, comprobados científicamente, vinculados con la incidencia de enfermedades cardiovasculares. Entre ellos están incluidos el tabaquismo, dietas poco saludables, hipertensión arterial, entre otros.

Ética: Una distorsión de la información debe ser evitada a todo costo para no generar desinformación a las personas. Además la información de los factores que se correlacionan con la incidencia de enfermedades cardiovasculares debe basarse en fuentes confiables y presentarse de una manera concisa.

Emocional: Las personas informadas pueden optar por decisiones y hábitos de vida más saludables teniendo en cuenta estos factores que se correlacionan con la incidencia de enfermedades cardiovasculares.

Contraargumentos

Lógica: Es posible que la predisposiciones genéticas tengan mayor "impacto" en los padecimientos cardiovasculares que los factores variables que se pueden presentar.

Ética: Se podrían presentar casos en los que las personas infieran incorrectamente en los datos y los resultados y, de esta manera, crear una divulgación falsa y preocupante.

Emocional: El entendimiento de estos factores puede ocasionar en las personas sentimientos negativos como por ejemplo un estrés o ansiedad adicional así como inclusive un sentimiento de culpa por elecciones de vida.

Conclusión

Se entiende que hay factores lógicos que se correlacionan con la incidencia de enfermedades cardiovasculares. Es decir, una persona que fuma o no hace ejercicio en general va a tener una mayor probabilidad de padecer de alguna enfermedad a largo plazo. Sin embargo, es fundamental asegurarse de brindar información actualizada y precisa al alcance de todos, realizado, primordialmente, con una buena intención de beneficiar de las personas.

¿Cómo se puede ver afectada la salud cardiovascular?

Argumentos

Lógica: La salud cardiovascular se puede ver afectada por distintos factores de riesgo, algunos que afectan más silenciosamente que otros. Estos factores influyen directamente en el tracto sanguíneo y el funcionamiento del corazón. Malos hábitos alimentarios pueden influir en padecimientos como la diabetes, la obesidad y altos niveles de colesterol. Estas consecuencias se ven reflejadas en la salud cardiovascular dado que pueden desencadenar numerosos problemas al corazón.

Ética: La información de como se puede ver afectada la salud cardiovascular debe ser precisa, basada en evidencia y respaldada por investigaciones científicas que sean congruente con los resultados.

Emocional: Las personas podrían verse bastante beneficiadas de conocer como se puede verse afectada su salud cardiovascular. Más allá de evitar enfermedades del corazón, un buen entendimiento de estas razones y sus posibles consecuencias pueden generar bienestar general de distintas maneras.

Contraargumentos

Lógica: La salud cardiovascular no se ve limitada sencillamente por las decisiones que se toman en la vida cotidiana, sino también hay factores que no se pueden modificar como, por ejemplo, la edad. A mayor edad, se es más propenso a este tipo de padecimientos y esto no tiene nada que ver con hábitos o genética.

Ética: Las enfermedades del corazón son realmente extensas y a pesar de tener fuentes confiables para realizar estudios, de igual manera se pueden presentar situaciones extraordinarias debido a que el cuerpo de cada persona es único y también reacciona distinto a diferentes factores.

Emocional: Sabiendo como puede verse afectada la salud cardiovascular, las personas podrían llegar a sentirse abrumadas o en sentimiento de negación. Más allá de brindar los resultados del análisis, también es importante proporcionar soluciones claras para la tranquilidad de la comunidad.

Conclusión

La combinación de un buen conocimiento de aquellos factores que pueden afectar la salud cardiovascular con información bien fundamentada, puede traer numerosos beneficios a todas las personas. Es racional determinar que hay hábitos menos saludables que otros y más allá de intentar generar preocupación, se pretende una motivación para hacer cambios para bien.

Argumentación a través de datos

Fuente de información: Los Centros para el Control y Prevención de Enfermedades (CDC, por sus siglas en inglés). Es la agencia de salud pública de los Estados Unidos. La información fue recolectada a través de Sistema de Vigilancia de Factores de Riesgo de Comportamiento (BRFSS, por sus siglas en inglés), que es el sistema de la CDC para la recolección de información a través de encuestas.

Contexto temporal y espacial de los datos: Los datos fueron recolectados entre el 2020 y el 15 de febrero del 2022 en los 50 estados de los Estados Unidos, el Distrito de Columbia y tres territorios del país.

Facilidad de obtener la información: La información se encuentra disponible en la página de los Centros para el Control y Prevención de Enfermedades (CDC, por sus siglas en inglés). Una versión con solo 18 variables se encuentra disponible en Kaggle.

Población de estudio: La población de los Estados Unidos.

Muestra observada: 400 000 adultos que formaron parte de la encuesta anual del CDC durante el año 2020.

Unidad estadística o individuos: Personas adultas que formaron parte de la encuesta anual del CDC durante el año 2020.

Descripción de las variables de la tabla:

- **HeartDisease:** Indica “Yes” o “No” para indicar si la persona padece o no una condición cardíaca o ha tenido un infarto de miocardio.
- **BMI:** Es el índice de masa corporal (IMC). Es un número positivo.
- **Smoking:** Indica “Yes” o “No” para indicar si la persona ha fumado por lo menos 100 cigarillos en su vida.
- **AlcoholDrinkning:** Indica “Yes” o “No” para indicar si la persona consume alcohol o no. Se define el consumo como tomar más de 14 veces a la semana para hombres y 7 para mujeres.
- **Stroke:** Indica “Yes” o “No” para indicar si la persona ha tenido un ataque cardíaco o no.
- **PhysicalHealth:** Indica durante cuántos días de los últimos 30 días la persona no ha tenido buena salud física.
- **MentalHealth:** Indica durante cuántos días de los últimos 30 días la persona no ha tenido buena salud mental.
- **DiffWalking:** Indica “Yes” o “No” para indicar si la persona tiene dificultades o no al caminar o subir escaleras.

-
- Sex: Sexo de la persona. “Female” para femenino y “Male” para masculino.
 - AgeCategory: Rango etáreo en el que se encuentra la persona. Puede ser 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80 “or older” (o más viejo).
 - Race: Raza de la persona. Puede ser “Black” para negro, “White” para blanco, “Hispano” para hispano, “American Indian/Alaska Native” para indio americano o nativo de Alaska o “Asian” para asiático.
 - Diabetic: Indica “Yes” o “No” para indicar si la persona padece o no de diabetes.
 - PhysicalActivity: Indica “Yes” o “No” para indicar si la persona ha realizado actividad física fuera de su trabajo durante los últimos 30 días.
 - GenHealth: Indica la percepción que tiene la persona de su salud general. Puede ser “Very Good” (muy buena), “Good” (buena), “Excellent” (excelente), “Fair” (razonable) o “Poor” (pobre).
 - SleepTime: Indica en promedio cuántas horas de sueño tiene la persona en un periodo de 24 horas. Es un número entero.
 - Asthma: Indica “Yes” o “No” para indicar si la persona padece o no de asma.
 - KidneyDisease: Indica “Yes” o “No” para indicar si la persona padece o no de una enfermedad renal.
 - SkinCancer: Indica “Yes” o “No” para indicar si la persona padece o no de cáncer de piel.

Esta base de datos es útil para responder la pregunta de investigación porque incluye datos de enfermedades cardiovasculares (columnas HeartDiseas y Stroke) y potenciales factores de riesgo como consumo de alcohol (AlcoholDrinking), tabaquismo (Smoking), IMC (BMI) y edad (AgeCategory). La pregunta de investigación trata de ligar las características de una persona con enfermedades cardiovasculares y la base de datos incluye esa información.

Revisión Bibliográfica

Fichas

Ficha 1

Título: Trends in Acute Ischemic Stroke Hospitalizations in the United States

Autores: Lucas Ramirez, May A. Kim-Tenser, Nerses Sanossian, Steven Cen, Ge Wen, Shuhan He, William J. Mack y Amytis Towfighi.

Año: 2016

Nombre del tema: Tendencias en hospitalizaciones en los Estados Unidos.

Forma de organizarlo:

- Cronológico: 11 de mayo del 2016
- Metodológico: Número ponderado de hospitalizaciones utilizando datos del censo de los Estados Unidos.
- Temático: Estadísticas de salud.
- Teoría: Medicina y estadística.

Resumen en una oración: La raza, edad y sexo de una persona son características que se relacionan con el riesgo de tener un accidente cerebrovascular isquémico agudo.

Argumento central: La población negra tiene la mayor proporción de hospitalizaciones por accidentes cerebrovasculares isquémicos agudos, seguida de la población hispana y blanca.

Problemas con el argumento o el tema: La categoría de hispano no se refiere necesariamente a la raza de la persona. Pueden existir personas negras e hispanas. Las categorías de negro, hispano y blanco mezclan el concepto de raza y nacionalidad.

Resumen en un párrafo: El estudio recopila la tasa de hospitalizaciones por accidentes cerebrovasculares isquémicos. Un accidente cerebrovascular isquémico ocurre cuando una arteria que suministra sangre al cerebro se ve obstruida. Para esto se utilizaron datos del número de hospitalizaciones y el censo de Estados Unidos entre el 2000 al 2010. El estudio toma como variables de decisión la edad, la raza y el sexo para presentar sus resultados. El número de personas hospitalizadas por esta afección de edad entre 64 a 85 años disminuyó pero por otra parte, incrementó en personas de 25 a 64 años. Se concluyó además que las personas de raza negra tienen mayor incidencia de este padecimiento, seguidos por los hispanos y las personas de raza blanca.

Ficha 2

Título: Obesity Increases Risk of Ischemic Stroke in Young Adults

Autores: Andrew B. Mitchell, John W. Cole, Patrick. F. McArdle, Yu-Ching Cheng, Kathleen A. Ryan, Mary J. Sparks, Braxton D. Mitchell y Steven J. Kittne

Año: 2015

Nombre del tema: Obesidad y ataques cardíacos.

Forma de organizarlo:

- Cronológico: 5 de mayo del 2015.
- Metodológico: Estudio de casos y controles basado en la población.
- Temático: Estadísticas de salud.
- Teoría: Medicina y estadística.

Resumen en una oración: La obesidad aumenta el riesgo de accidente cerebrovascular isquémico en adultos jóvenes

Argumento central: Hay una relación entre la obesidad y el riesgo de accidente cerebrovascular isquémico. Entre más IMC, más riesgo existe.

Problemas con el argumento o el tema: Es problemático definir la obesidad por medio del IMC. No hay una forma universalmente aceptada de hacerlo.

Resumen en un párrafo: Actualmente, una gran parte de la población estadounidense es obesa. Este artículo determina que las personas obesas tiene hasta un 70% de riesgo mayor de un accidente cerebrovascular que alguien con IMC más bajo. Por lo tanto, la relación casi lineal entre un IMC alto y el riesgo de accidente cerebrovascular isquémico es un problema a nivel nacional. En general, el término obesidad es un poco subjetivo pero definitivamente cada día se ven más y más niños y adultos jóvenes con obesidad. El tratamiento de la obesidad representaría una mejora en la calidad de vida de tales pacientes, además de una disminución en los accidentes cerebrovasculares isquémicos. Mucho más allá de tratar los accidentes cerebrovasculares además garantiza una mejor calidad de vida dado que previene otro tipo de afecciones como diabetes o hipertensión arterial.

Ficha 3

Título: Smoking as a Risk Factor for Stroke in Women Compared With Men

Autores: Sanne A.E. Peters, Rachel R. Huxley y Mark Woodward

Año: 2013

Nombre del tema: Fumar y accidentes cerebrovasculares.

Forma de organizarlo:

- Cronológico: 22 de agosto del 2013
- Metodológico: Agrupación mediante metanálisis de efectos aleatorios con ponderación de varianza inversa.
- Temático: Estadísticas de salud.
- Teoría: Medicina y estadística.

Resumen en una oración: Existe un riesgo similar entre mujeres y hombres que fuman de un ataque isquémico.

Argumento central: El riesgo de un ataque isquémico entre hombres y mujeres que fuman es muy similar. Existe evidencia de que es al menos tan grande entre las mujeres en comparación con los hombres. Los beneficios de dejar de fumar también son similares.

Problemas con el argumento o el tema: La definición de fumar podría generar resultados diferentes entre distintos estudios analizados.

Resumen en un párrafo: Esta investigación se basa en determinar si el tabaquismo afecta de igual manera a los hombres que a las mujeres como factor de riesgo a desarrollar un accidente cerebrovascular. Se concluye que fumar incrementa el riesgo de cualquier tipo de ataque isquémico independientemente del sexo. El riesgo y los beneficios de dejar de fumar son similares entre sexos. Estos resultados no coinciden con la expectativa de que hubiera más riesgo para los hombres. El tabaquismo siempre va a implicar un riesgo a la salud y bienestar de las personas.

Ficha 4

Título: Body Mass Index and Association With Cardiovascular Outcomes in Patients With Stable Coronary Heart Disease - A STABILITY Substudy

Autores: Claes Held, Nermin Hadziosmanovic, Philip E. Aylward, Emil Hagström, Judith S. Hochman, Ralph A. H. Stewart, Harvey D. White y Lars Wallentin

Año: 2022

Nombre del tema: IMC y enfermedades cardiovasculares

Forma de organizarlo:

- Cronológico: 21 de enero del 2022
- Metodológico: Análisis de regresión de Cox con ajustes multivariados.
- Temático: Estadísticas de salud.
- Teoría: Medicina y estadística.

Resumen en una oración: Las personas con obesidad tienen más riesgo de enfermedad cardiovascular.

Argumento central: Hay una relación entre el IMC y el riesgo de enfermedad cardiovascular.

Problemas con el argumento o el tema: El IMC solo toma en cuenta el peso y la altura de una persona. Una persona podría pesar mucho por tener mucha masa muscular y no necesariamente tiene más riesgo de padecer una enfermedad cardiovascular.

Resumen en un párrafo: En este artículo se habla de un análisis comparativo entre un índice de masa corporal alto y enfermedades cardiovasculares. Se tomaron como muestra personas con enfermedades coronarias estables de alrededor de 3 a 5 años con el seguimiento y tratamiento debido. Se utilizó la regresión de Cox para hacer el análisis en los cuales se tenían en cuenta varias variables. Los resultados arrojaron que efectivamente el índice de masa corporal (IMC) puede incrementar el riesgo de complicaciones cardiovasculares como inflamatorias y cardiometabólicas para personas con enfermedad coronaria. Además distintos indicadores como la diabetes, hipertensión, triglicéridos, entre otros, se ven altamente asociados con un IMC alto. Sin embargo, también se concluyó que un IMC bajo, es decir, que una persona tenga bajo peso, también se ve asociado a insuficiencia cardíaca y a altos niveles de riesgo de muerte cardiovascular.

Ficha 5

Título: Inter-correlation of risk factors among heart patients

Autores: Jahoor Alam, Abdullah Ibrahim Alnafeesah y Mohd Saeed

Año: 2020

Nombre del tema: Análisis estadístico de factores de riesgo

Forma de organizarlo:

- Cronológico: 10 de junio del 2020
- Metodológico: Correlación de Pearson y medidas estadísticas de la prueba t.
- Temático: Estadísticas de salud.
- Teoría: Medicina y estadística.

Resumen en una oración: Hay una relación entre factores de riesgo como la diabetes y las enfermedades cardiovasculares.

Argumento central: Existe una relación entre factores como el peso, la presión arterial, la edad y la diabetes y las enfermedades cardiovasculares.

Problemas con el argumento o el tema: La diabetes en combinación con un estilo de vida saludable podría no estar tan relacionada con el padecimiento de enfermedades cardiovasculares.

Resumen en un párrafo: Las enfermedades cardiovasculares son la tercera más común razón de muerte en Arabia Saudita después de las muertes por vejez y accidente. Estas están asociadas a factores como el peso, la edad y el padecimiento de la diabetes. La diabetes es un factor que está particularmente relacionado con enfermedades cardiovasculares. Se recolectó información en la región central de Arabia Saudita y se separó la población. El peso y la presión están relacionados con enfermedades del corazón.

UVE de Gowin

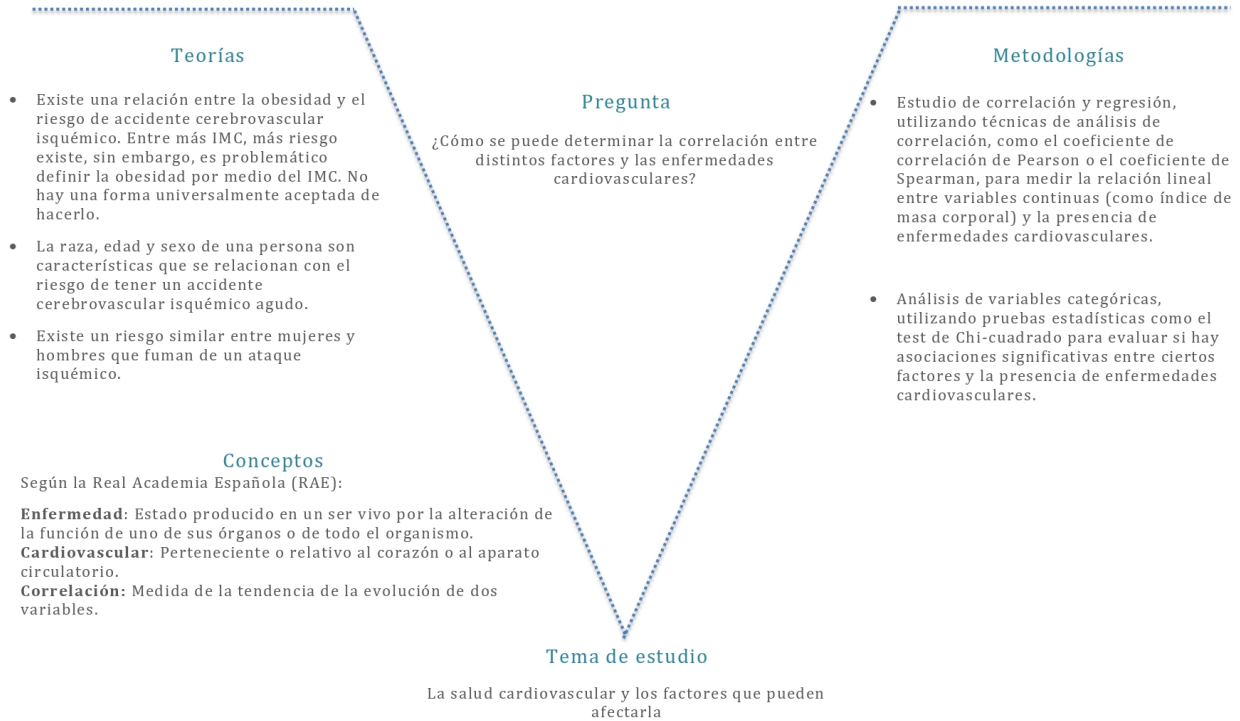


Figure 1: Elaboración propia

Parte de escritura

Pregunta

¿Como se puede determinar la correlación entre distintos factores y las enfermedades cardiovasculares?

Argumentación

La pregunta de investigación escogida puede verse contestada de distintas maneras. Para la finalidad de esta investigación, se comprende como la utilización de modelos estadísticos y distintos análisis para poder encontrar una respuesta más precisa y clara. Determinar la relación entre distintos factores y las enfermedades cardiovasculares es importante para identificar factores de riesgo. Después de haber identificados cuáles factores se correlacionan con enfermedades cardiovasculares, se pueden crear estrategias para mitigar el efecto que tienen sobre la población. Por ejemplo, si se descubriera que la obesidad (es decir, un IMC alto) es un factor que aumenta la probabilidad de padecer una enfermedad cardiovascular, se pueden hacer análisis posteriores para descubrir que está provocando obesidad en la población y qué se puede hacer para evitarlo.

Según Jahoor Alam et al., "la tasa de mortalidad por enfermedades cardiovasculares ha aumentado durante los últimos años". Por lo tanto, resulta relevante averiguar cuáles causas tienen. Ya se ha estudiado que variables incluidas en la base (raza, edad y sexo) se relacionan con el riesgo de tener un accidente cerebrovascular isquémico agudo (Lucas Ramirez et al., 2016). También, que las personas con obesidad tienen más riesgo de enfermedad cardiovascular (Glaes Held et al., 2022) y de accidente cerebrovascular isquémico (Andrew B. Mitchell et al., 2015). Por otro lado, se han encontrado resultados inesperados como que el tabaquismo tiene un efecto similar en hombres y mujeres de padecer cualquier tipo de ataque isquémico.

Métodos

Prueba X^2 : Según (DeGroot, M. y Schervish, M., 2012), si se tiene una población grande de elementos de k tipos diferentes, p_i denota la probabilidad de que un elemento escogido aleatoriamente va a ser de tipo $i = 1, \dots, k$ tal que $p_i \geq 0$ para cualquier i y $\sum_{i=1}^k p_i = 1$ y p_1^0, \dots, p_k^0 son números específicos tales que p_i^0 para cualquier i y $\sum_{i=1}^k p_i^0 = 1$, se tiene que

$$Q = \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{np_i^0}$$

tiene la propiedad de que si la hipótesis $p_i = p_i^0$ para $i = 1, \dots, k$ es cierta y el tamaño de la muestra $n \rightarrow \infty$, entonces Q converge en distribución a la distribución X^2 con $k - 1$ grados de libertad.

N_i es el número de observaciones en una muestra aleatoria que son de tipo i , con $i = 1, \dots, k$, y tal que $\sum_{i=1}^k N_i = n$, donde n es el tamaño de la población aleatoria.

Coefficiente de Pearson: Según (Suárez, M., 2011) los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables, es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables. El coeficiente de correlación de Pearson está dado por

$$r_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}}$$

donde \bar{X} y \bar{Y} son las medias respectivas.

Coefficiente de Spearman: A diferencia del coeficiente de Pearson, esta medida, según DATAtab(2023), examina la relación de dos variables por medio de rangos, es decir, no es paramétrica y no se requiere una distribución normal de los datos. Según Martínez, R., Tuya, L., Martínez, M., Pérez, A., Cánovas, A. (2009), para calcular este coeficiente de correlación se utiliza la siguiente fórmula:

$$r_s = 1 - \frac{6 \times \sum d_i^2}{n(n^2 - 1)},$$

donde r_s corresponde al coeficiente de correlación que varía entre -1 y 1, d_i es la diferencia de rangos entre las dos variables, y n es el número de casos.

El coeficiente obtenido nos indica tanto la fuerza de la correlación así como la dirección en la que va la correlación. La fuerza de correlación se puede leer de la siguiente manera:

Valor r_s	Fuerza de Correlación
0 - 0.25	Nula o escasa
0.26 - 0.5	Débil
0.51 - 0.75	Moderada o fuerte
0.76 - 1	Alta o perfecta

Bibliografía

Centers for Disease Control and Prevention (CDC). (2020). 2020 BRFSS Survey Data and Documentation. https://www.cdc.gov/brfss/annual_data/annual_2020.html

DATAtab Team. (2023). Coeficiente de correlación de Spearman. <https://datatab.es/tutorial/spearman-correlation>

DeGroot, M. H., Schervish, M. J. (2019). Probability and Statistics. Pearson Education, Inc.

Held, C., Hadziosmanovic, N., Aylward, P. E., Hagström, E., Hochman, J. S., Stewart, R. A., White, H. D., Wallentin, L. (2022). Body mass index and association with cardiovascular outcomes in patients with stable coronary heart disease – a stability substudy. *Journal of the American Heart Association*, 11(3). <https://doi.org/10.1161/jaha.121.023667>

Jahoor Alam, M., Ibrahim Alnafeesah, A., Saeed, M. (2020). Inter-correlation of risk factors among heart patients. *AIMS Public Health*, 7(2), 354–362. <https://doi.org/10.3934/publichealth.2020030>

Kernan, W. N., Dearborn, J. L. (2015). Obesity increases stroke risk in young adults. *Stroke*, 46(6), 1435–1436. <https://doi.org/10.1161/strokeaha.115.009347>

Martínez, R., Tuya, L., Martínez, M., Pérez, A., Cánovas, A. (2009). El coeficiente de correlación de los rangos de Spearman. <http://scielo.sld.cu/scielo.php?script=sci.arttext&pid=S1729-519X2009000200017>

Peters, S. A. E., Huxley, R. R., Woodward, M. (2013). Smoking as a risk factor for stroke in women compared with men. *Stroke*, 44(10), 2821–2828. <https://doi.org/10.1161/strokeaha.113.002342>

Pytlak, K. (2021). Personal Key Indicators of Heart Disease. <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

Real Academia Española (RAE). (2018). Servicio de consulta electrónica al “Diccionario de la lengua española.”- Edición del Tricentenario. <https://dle.rae.es/>

Ramirez, L., Kim-Tenser, M. A., Sanossian, N., Cen, S., Wen, G., He, S., Mack, W. J., Towfighi, A. (2016). Trends in acute ischemic stroke hospitalizations in the United States. *Journal of the American Heart Association*, 5(5). <https://doi.org/10.1161/jaha.116.003233>

World Health Organization. (2021). Obesity and Overweight. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

Suárez, M. (2011). Coeficiente de correlación de Karl Pearson. <http://repositorio.utn.edu.ec/handle/123456789/766>