

UNIVERSIDAD DE COSTA RICA

ESCUELA DE MATEMÁTICA

DEPARTAMENTO DE MATEMÁTICA PURA Y CIENCIAS

ACTUARIALES

ESTADÍSTICA ACTUARIAL I - CA0303

**Relación entre diferentes factores y padecer una enfermedad
cardiovascular**

Anteproyecto

Jimena Chacón Chavarría - B81999

Maria Jimena Marchena Mendoza - B74425

Jorge Andreas Sanabria Arce - B97190



UNIVERSIDAD DE
COSTA RICA

EMat

Escuela de
Matemática

1. Anteproyecto	3
1.1. Pregunta de Investigación	3
1.2. Objetivos	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	3
1.3. Introducción al marco teórico	4
1.4. Descripción de los datos	5
1.5. Análisis descriptivo de los datos	6
1.6. UVE de Gowin	12

1.1. Pregunta de Investigación

¿Existe una relación entre el tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo, edad y las enfermedades cardiovasculares, en población adulta estadounidense en el período 2020-2022?

1.2. Objetivos

1.2.1. Objetivo General

Analizar la relación entre el tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo, edad y el surgimiento de las enfermedades cardiovasculares en población adulta estadounidense durante el período 2020-2022.

1.2.2. Objetivos Específicos

- Determinar la relación entre el tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo, edad y el surgimiento de las enfermedades cardiovasculares por medio del Test X^2 y el Test Wilcoxon-Mann-Whitney utilizando tablas de contingencia.
- Evaluar el impacto del tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo, edad y el surgimiento de las enfermedades cardiovasculares.
- Identificar el comportamiento de las variables de tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo y edad en relación con las variables de enfermedades cardiovasculares.

1.3. Introducción al marco teórico

Las enfermedades cardiovasculares son un grupo de desórdenes del corazón y los vasos sanguíneos, según WHO (2023). Los infartos y ataque isquémico son ambas enfermedades cardiovasculares, las cuales son variables a estudiar dentro de la base de datos. El riesgo de padecer una de estas se ve afectado por diferentes factores. La literatura dice que existe una relación entre factores como el peso, la presión arterial, la edad y la diabetes con las enfermedades cardiovasculares (Alam, J. et al., 2020). El tabaquismo, que es la adicción al consumo del tabaco (RAE, 2023), es otro factor que afecta ese riesgo (Peters, S., et al., 2013). Uno de los que llama más la atención es el índice de masa corporal (IMC), que es una manera de medir el peso de una persona y está directamente asociado al riesgo de accidente cerebrovascular isquémico en adultos jóvenes (Mitchell, A., et al., 2015). Estas relaciones ya se han comprobado para los Estados Unidos (Ramirez, L., et al., 2016) (Held, C., et al., 2022). Cabe mencionar que de vez en cuando se obtienen resultados poco intuitivos, como que hay un riesgo similar entre mujeres y hombres que fuman de un ataque isquémico (Peters, S., et al., 2013).

Considerando que las variables son categóricas, la manera más eficiente de acomodarlas para hacer un análisis estadístico es en tablas de contingencia (DeGroot, M. H. y Schervish, M. J., 2019). De esa manera, se crearán tablas para las variables Heart y Stroke y cada una de las variables categóricas restantes. De esa manera, se tendrían 5 tablas para cada una de las enfermedades cardiovasculares para un total de 10 tablas. Al tener listas las tablas de contingencia se puede realizar la prueba X^2 , la cual va a servir para establecer la independencia o dependencia de las variables categóricas dadas (DeGroot, M. H. y Schervish, M. J., 2019).

Para el caso del IMC, que es una variable continua, se creará una sucesión de IMC con Stroke positivo y una de IMC para Stroke negativo, de manera que se podrá establecer la relación entre el IMC y el padecimiento de un ataque isquémico a través de la prueba Wilcoxon-Mann-Whitney. El mismo proceso se va a realizar para el IMC y la variable HeartDisease, de manera que se tienen cuatro sucesiones de variables continuas en total (DeGroot, M. H. y Schervish, M. J., 2019).

1.4. Descripción de los 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 desde el 2020 al 2022.

Unidad estadística o individuos: Personas adultas que formaron parte de la encuesta anual del CDC desde el 2020 al 2022.

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.

1.5. ANÁLISIS DESCRIPTIVO DE LOS DATOS

- **AlcoholDrinking:** 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.
- **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).
- **PhysicalActivity:** Indica “Yes” o “No” para indicar si la persona ha realizado actividad física fuera de su trabajo durante los últimos 30 días.

Esta base de datos es útil para responder la pregunta de investigación porque incluye datos de enfermedades cardiovasculares (columnas HeartDisease y Stroke) y los potenciales factores de riesgo que se han elegido para analizar: alcohol (AlcoholDrinking), tabaquismo (Smoking), IMC (BMI) y edad (AgeCategory), sexo (Sex), actividad física (PhysicalActivity). La pregunta de investigación trata de ligar estas características de las personas con enfermedades cardiovasculares y la base de datos incluye esa información. Además, la cantidad de personas incluidas, que rondan las 400 000 personas, asegura que los resultados que se van a tener son confiables.

1.5. Análisis descriptivo de los datos

Para estudiar la relación entre el tabaquismo, alcoholismo, índice de masa corporal (IMC), actividad física, sexo, edad y las enfermedades cardiovasculares, inicialmente se pueden observar los siguientes gráficos y tablas de los datos a trabajar.

En el cuadro 2.1 se muestra que el número total de personas con enfermedad cardíaca es mucho mayor que el número de personas con historial de derrame cerebral. Sin embargo, el número de personas con derrame cerebral es más alto en el grupo de “Sin enfermedad cardíaca”. Esto podría indicar que, aunque la enfermedad cardíaca afecta a más personas, el riesgo de derrame cerebral puede ser relativamente mayor en personas sin enfermedad cardíaca.

Cuadro 2.1

Adultos estadounidenses por historial de derrame cerebral según historial de enfermedad cardíaca, 2020

Enfermedad Cardíaca	Sin Derrame Cerebral	Con Derrame Cerebral	Total
No	284 742	7 680	292 422
Si	22 984	4 389	27 373
Total	307 726	12 069	319 795

Fuente: Elaboración propia con datos de Centros para el Control y Prevención de Enfermedades (CDC, 2020). Personal Key Indicators of Heart Disease [Archivo de datos].

El cuadro 2.2 es de utilidad para estudiar la relación entre el tabaquismo, alcoholismo, actividad física y las enfermedades cardiovasculares. Se puede observar que la población con historial de enfermedades cardiovasculares, es mayor para aquellas consumidoras de tabaco, sin embargo, con el alcoholismo, se tiene el efecto contrario. Con respecto a la actividad física, la población estudiada que es físicamente activa, presenta una frecuencia mayor de enfermedades cardiovasculares, con respecto a las que no.

Cuadro 2.2

Adultos estadounidenses por historial de tabaquismo, alcoholismo y actividad física según historial de enfermedad cardíaca y derrame cerebral, 2020

Condición	Factores					
	Tabaquismo	No Tabaquismo	Alcoholismo	No Alcoholismo	Actividad Física	No Actividad Física
Enfermedad Cardíaca						
No	115 871	176 551	20 636	271 786	230 468	61 954
Si	16 037	11 336	1 141	26 232	17 489	9 884
Derrame Cerebral						
No	125 093	182 633	21 260	286 466	240 620	67 106
Si	6 815	5 254	517	11 552	7 337	4 732

Fuente: Elaboración propia con datos de Centros para el Control y Prevención de Enfermedades (CDC). (2020). Personal Key Indicators of Heart Disease [Archivo de datos].

Cuadro 2.3

Adultos estadounidenses por nivel de índice de masa corporal (IMC) según sexo, historial de enfermedad cardíaca y derrame cerebral, 2020

Condición	Niveles de IMC			
	< 18.5	18.5-24.9	25.0-29.9	30 ≥
Mujeres	6 994	114 528	104 138	109 950
Enfermedad Cardíaca				
No	3 228	54 273	48 627	50 443
Si	269	2 991	3 442	4 532
Derrame Cerebral				
No	3 307	55 526	50 041	52 504
Si	190	1 738	2 028	2 471
Hombres	3 226	75 558	125 556	99 640
Enfermedad Cardíaca				
No	1 481	34 632	56 235	43 503
Si	132	3 147	6 543	6 317
Derrame Cerebral				
No	1 548	36 514	60 521	47 765
Si	65	1 265	2 257	2 055

Fuente: Elaboración propia con datos de Centros para el Control y Prevención de Enfermedades (CDC). (2020). Personal Key Indicators of Heart Disease [Archivo de datos].

Según los datos, se puede observar en el cuadro 2.3, que a un mayor nivel de índice de masa corporal (IMC), se tiene una mayor incidencia en historial de enfermedad cardiovascular, esto tanto para hombres como mujeres. Además, como se observa en el gráfico 2.1, una mayor parte de la población se encuentra en los niveles altos de índice de masa corporal (IMC), de estos, como se puede notar en el gráfico 2.2, en el caso de un nivel de IMC de 25-29.9, existe una mayor cantidad de personas de sexo masculino, en caso contrario, para los niveles de 18.5-24.9, y más de 30, el sexo femenino presenta una mayor frecuencia.

Gráfico 2.1 Distribución de IMC en adultos estadounidenses por nivel, 2020

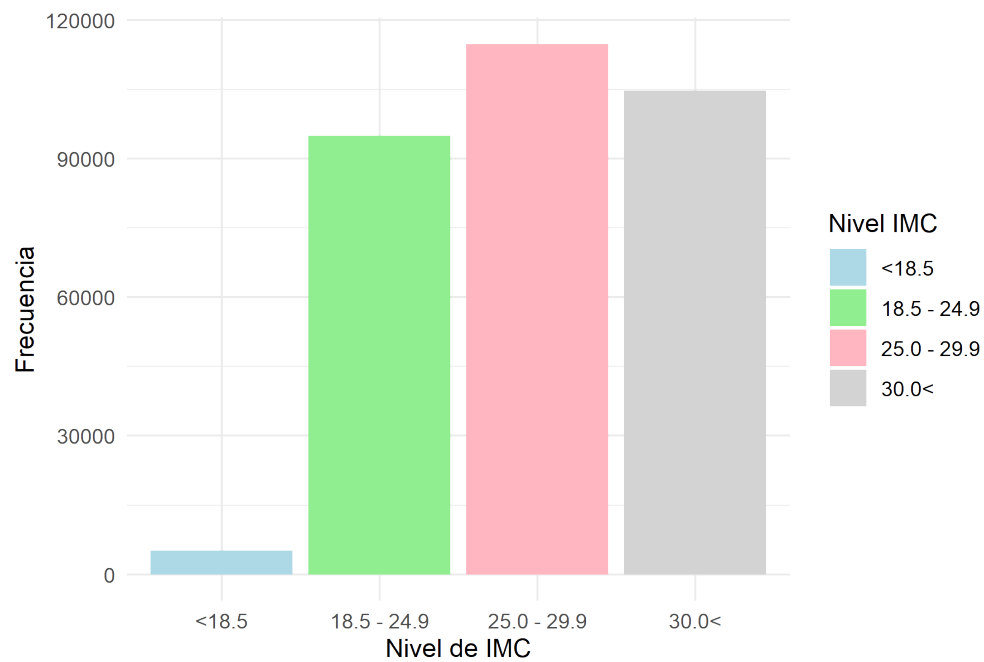
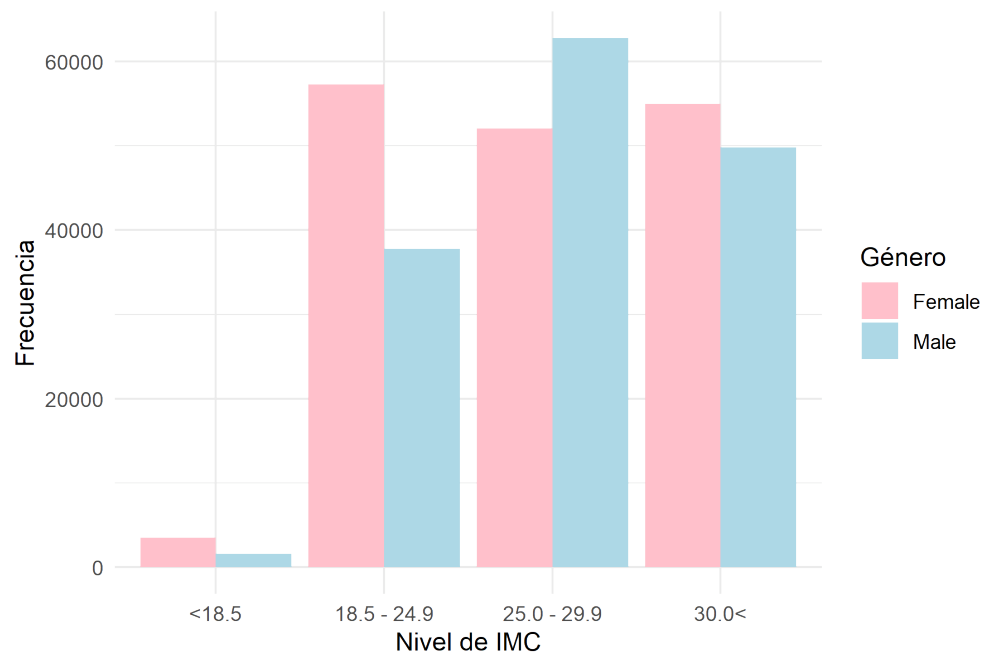


Gráfico 2.2 - Distribución de IMC en adultos estadounidenses por nivel, y sexo 2020-2022



Fuente: Elaboración propia con datos de Centros para el Control y Prevención de Enfermedades (CDC). (2022). Personal Key Indicators of Heart Disease [Archivo de datos]

Por último, con los gráficos 2.3 y 2.4, se puede determinar que existe una diferencia entre la mediana de IMC de la población que, si presenta enfermedades cardiovasculares, con respecto a las que no, además esta diferencia es más significativa en enfermedad cardíaca.

Distribución de IMC en adultos estadounidenses por historial de enfermedades cardiovasculares, 2020-2022

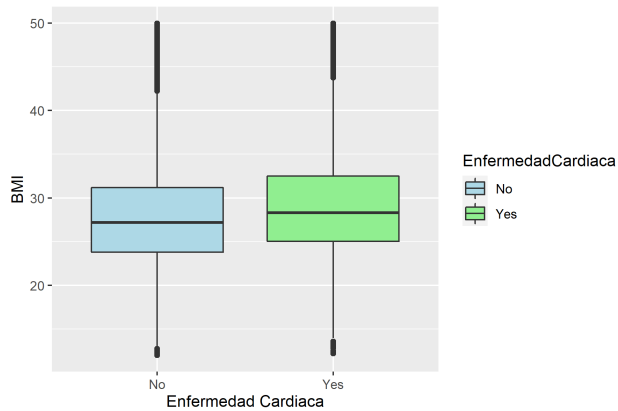


Gráfico 2.3 - Enfermedad cardíaca

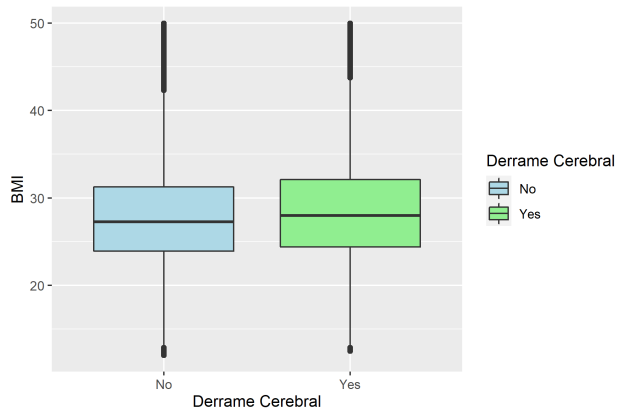


Gráfico 2.4 - Derrame cerebral

Fuente: Elaboración propia con datos de Centros para el Control y Prevención de Enfermedades (CDC). (2022). Personal Key Indicators of Heart Disease [Archivo de datos]

1.6. UVE de Gowin

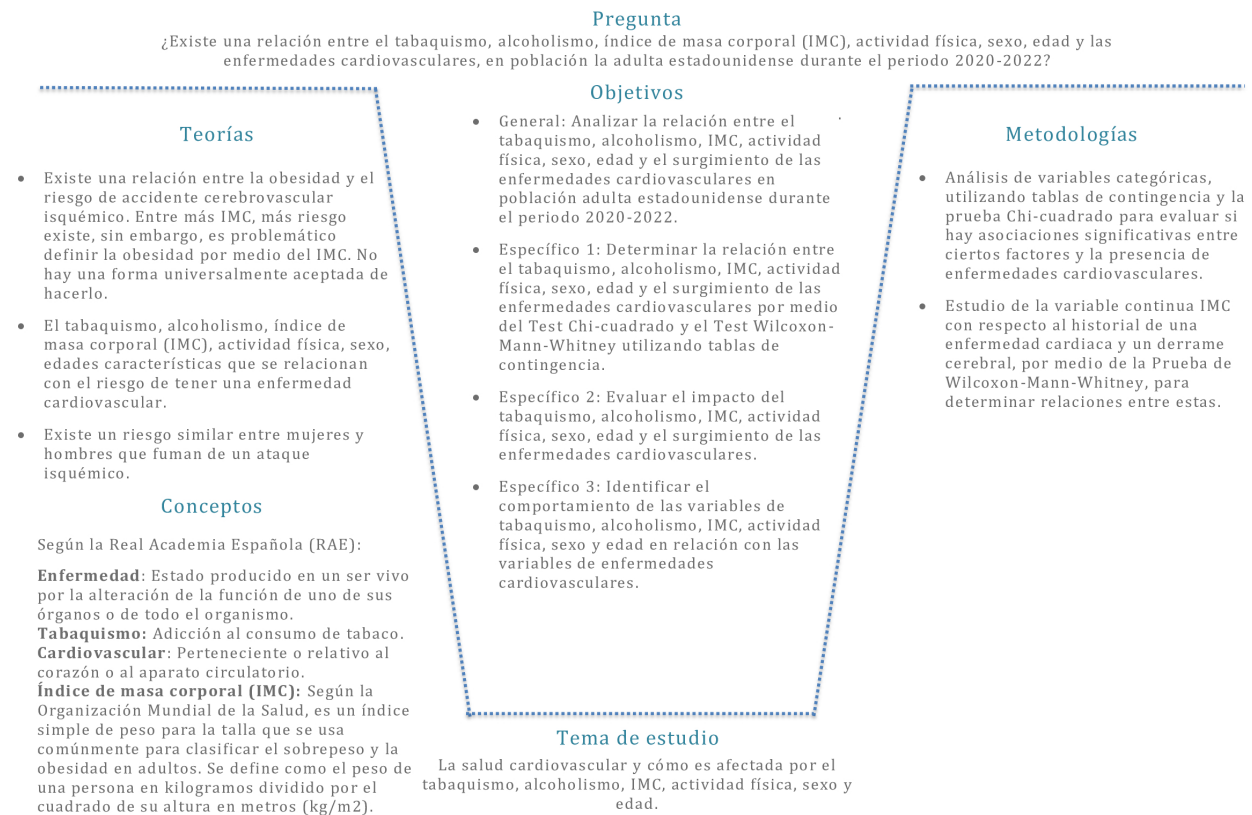


Figura 1.1: Elaboración propia

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

Centro Nacional para la Prevención de Enfermedades Crónicas y Promoción de la Salud, División de Nutrición, Actividad Física, y Obesidad. (2022) Acerca del índice de masa corporal para adultos. https://www.cdc.gov/healthyweight/spanish/assessing/bmi/adult_bmi/index.html#:~:text=El%20%C3%ADndice%20de%20masa%20corporal,saludable%2C%20sobrepeso%2C%20y%20obesidad.

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>

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). (2023). 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 (WHO). (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>

World Health Organization. (2023). Cardiovascular diseases. who.int/health-topics/cardiovascular-diseases#tab=tab1