

# Factores demográficos en el síndrome metabólico

---

Mariano Puchades del Olmo

Bootcamp online Data Science

2024-2025



# INDICE

1. Introducción .....	3
2. Hipótesis.....	4
3. Exploración de los datos .....	5
4. Análisis de los datos .....	6
4.1. Análisis univariante.....	6
4.2. Análisis bivalente.....	6
4.3. Análisis en profundidad .....	9
5. Conclusiones .....	11

# 1. Introducción

El síndrome metabólico (SM) hace referencia al conjunto de alteraciones metabólicas formado por la obesidad de distribución central, disminución de la concentración de colesterol unido a lipoproteínas de alta densidad (HDL), aumento de la concentración de triglicéridos, hiperglucemia y aumento de la presión arterial (1,2).

Los sujetos diagnosticados de síndrome metabólico tienen un riesgo aumentado de desarrollar diabetes tipo 2 y enfermedad cardiovascular, convirtiendo este síndrome en un problema de salud pública a nivel mundial (1,2).

La diferenciación entre etnias es muy importante, ya que los criterios diagnósticos cambian de un grupo poblacional a otro. Por ejemplo, los valores máximos de circunferencia de la cintura son diferentes entre la población asiática y la europea (1).

El objetivo de este proyecto es indagar en la influencia de los factores demográficos en un grupo de sujetos en el que encontramos personas con síndrome metabólico diagnosticado y personas sin él.

Partimos de un *dataset* con 2401 entradas y 15 columnas, obtenido en Kaggle (3). Hay que tratarlo, puesto que contiene nulos en varias columnas. Es una muestra pequeña, pero podemos sacar una idea para profundizar más en un futuro.

## 2. Hipótesis

*El nivel de ingresos es un factor influyente en el desarrollo de síndrome metabólico.*

1. ¿Depende la existencia de síndrome metabólico del salario mensual del sujeto?
2. ¿Hay diferencias en esta relación entre hombres y mujeres?

*El estado civil del sujeto tiene relación con el desarrollo de síndrome metabólico.*

1. ¿Cómo influye el estado civil en el desarrollo de este cuadro clínico?
2. ¿Hay diferencias entre hombres y mujeres?
3. ¿Y entre etnias?

*La etnia del sujeto influye en la aparición de síndrome metabólico.*

1. ¿Hay más casos de síndrome metabólico en unas poblaciones que en otras?
2. ¿Está esto relacionado con los ingresos de los sujetos?

*La edad del sujeto influye en el diagnóstico de síndrome metabólico.*

1. ¿Hay relación entre edad y síndrome metabólico?

Con estas hipótesis, nuestro análisis se centrará en las variables demográficas, aunque también estudiaremos el resto de variables.

### 3. Exploración de los datos

Mostramos los datos de nuestro *dataset* utilizando Pandas. Nuestro grupo de datos está en inglés, por lo que primeramente traducimos las columnas y las variables categóricas de tipo *string*.

Comprobamos la presencia de nulos y encontramos 4 columnas afectadas: Estado civil, Ingresos, CircunferenciaCintura e IMC.

- EstadoCivil: tiene un 8.7% de nulos, pero para no eliminarlos vamos a crear una nueva categoría llamada “Desconocido”.
- Ingresos: 4.9% de nulos. Como es un porcentaje bajo, vamos a utilizar la mediana de Ingresos para completar esos nulos.
- CircunferenciaCintura: tiene un 3.5% de nulos. Usamos la mediana para completarlos.
- IMC: tiene un 1.1% de nulos, por lo que también usamos la mediana.

Al rellenar los nulos estamos introduciendo un sesgo, debemos tenerlo en cuenta.

A continuación, definimos y tipificamos las variables del *dataset*:

- Las clasificamos teniendo en cuenta su porcentaje de cardinalidad y el tipo de dato de la variable.
- Establecemos el orden de importancia de las variables, priorizando aquellas implicadas en la posible respuesta a nuestras preguntas.

## 4. Análisis de los datos

### 4.1. Análisis univariante

Realizamos un estudio de las medidas de tendencia central en todas las variables:

- Obtenemos la distribución de las variables categóricas como frecuencia relativa y su moda.
- Observamos la distribución de las variables numéricas mediante histogramas y diagramas de cajas y bigotes. Además, obtenemos media, mediana, percentiles mínimos y máximos.

Tras realizar ese análisis obtenemos bastantes *outliers* por lo que los estudiamos detenidamente y decidimos:

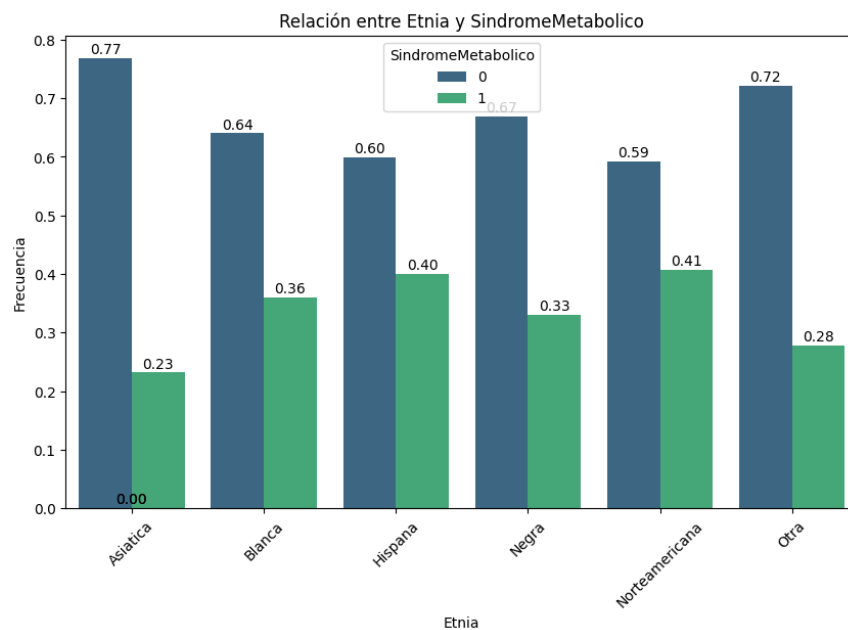
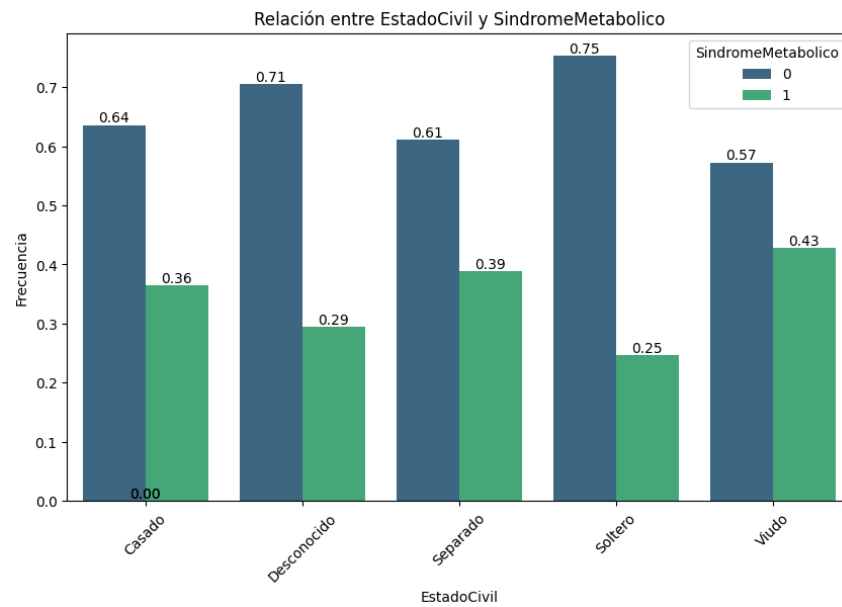
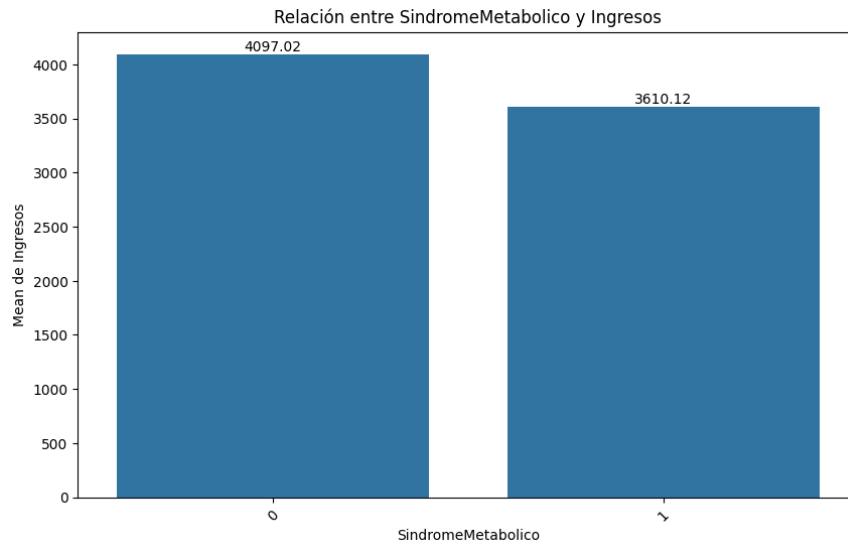
- Eliminar la columna CACu: tiene muchos *outliers* y sus valores están categorizados en la columna albuminuria.
- Los *outliers* de IMC y HDL cuadran con el resto de valores de esos sujetos, por lo que no los eliminamos, no parecen un error.
- En la columna de Triglicéridos encontramos varios *outliers* y eliminamos los dos más alejados, puesto que valores tan altos de triglicéridos indican patología que no recogemos en este *dataset*, no son de nuestro interés.

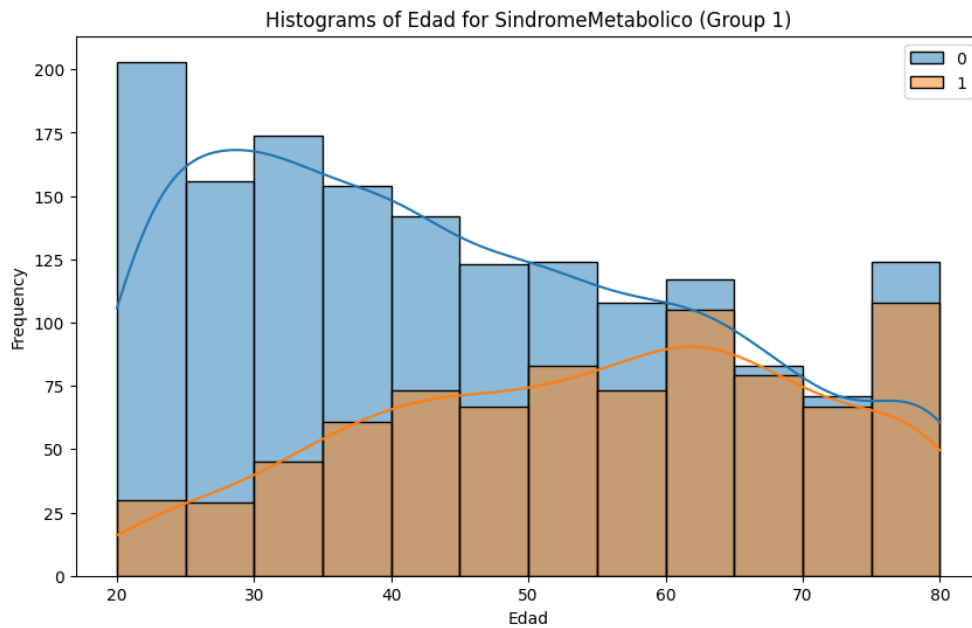
Tras realizar estos cambios, guardamos el *dataset* en un archivo csv final, con el que realizaremos el análisis bivariante.

### 4.2. Análisis bivariante

Primero realizamos un análisis bivariante con la variable *SindromeMetabolico* como directora, cruzándola con *Ingresos*, *EstadoCivil*, *Etnia* y *Edad*. A través de los estadísticos Chi-cuadrado y U Mann-Whitney observamos que hay relación entre todas estas variables y la directora:

- Los sujetos con síndrome metabólico tienen una media menor de ingresos mensuales.
- Los grupos de sujetos separados y viudos tienen un mayor porcentaje de diagnóstico positivo, mientras que los solteros tienen el menor.
- Hay un mayor porcentaje de norteamericanos e hispanos con síndrome metabólico en comparación con el resto de etnias.
- Los sujetos con síndrome metabólico tienen una media mayor de edad.





También analizamos el resto de variables separando por poblaciones, pero no indagamos en exceso, ya que el resto de columnas están relacionadas con el diagnóstico: muchas forman parte de los criterios diagnósticos del síndrome.

De todas formas, observamos varias cosas:

- El IMC en la población negra es el más alto, sin embargo, no es el grupo que más diagnósticos tiene.
- La etnia norteamericana tiene una media de circunferencia de la cintura muy parecida a la que tienen otras etnias, pero es una de las que más sujetos con síndrome metabólico tiene.
- Con respecto a los indicadores sanguíneos y de orina:
  - Todos los valores sanguíneos cuadran con la cantidad de diagnósticos. Los hispanos y norteamericanos tienen peores valores analíticos. Destacamos que la etnia negra tiene valores de glucosa en sangre parecidos a los norteamericanos, pero su porcentaje de positivos es menor.
  - Los indicadores de orina son muy parecidos en todas las poblaciones, aunque sí hay relación entre etnia y valores en orina.

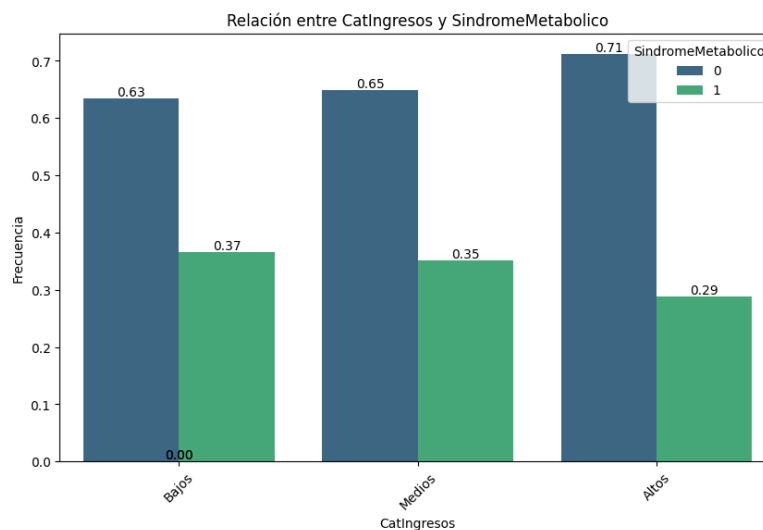
Estos datos destacados pueden deberse a diferencias genéticas entre etnias y a diferencias en la red de alimentación y hábitos de vida, pudiendo generar esos valores en los análisis de sangre.



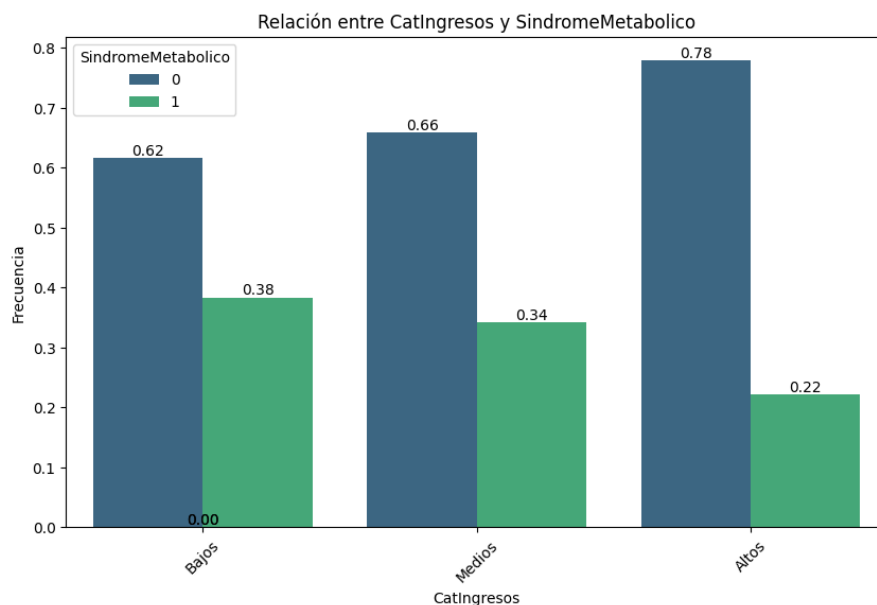
### 4.3. Análisis en profundidad

Tras obtener toda esa información, vamos a profundizar un poco más en nuestro análisis, filtrando nuestro *dataset* de diferentes maneras.

Vamos a crear una nueva columna llamada *CatIngresos*, clasificando a los sujetos en 3 categorías: ingresos bajos (menos de 3000), medios (entre 3000 y 6000) y altos (más de 6000). Cruzando esta columna con la de *SindromeMetabolico*, vemos que, en el grupo de ingresos altos, tan solo hay un 29% de sujetos con este síndrome, mientras que en el resto hay alrededor de un 36%.



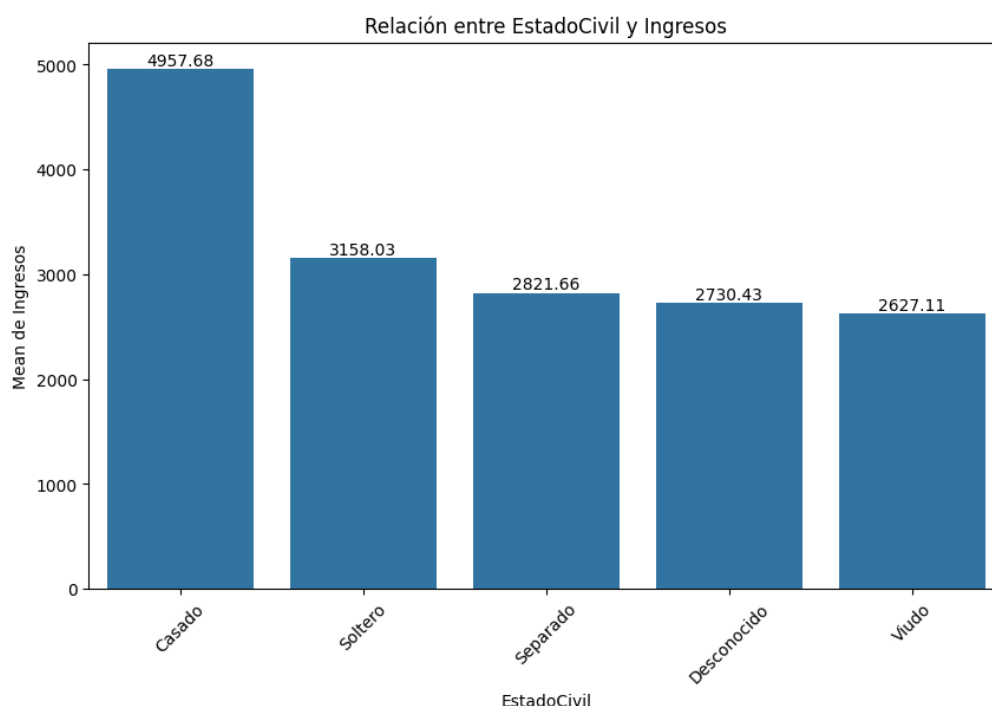
Para seguir profundizando, dividimos el *dataset* en dos: en hombres y mujeres. Volvemos a analizar las categorías de ingresos y obtenemos que en los hombres no hay diferencia entre grupos, mientras que, en las mujeres, sí hay una relación clara entre la categoría de ingresos: en el grupo de mujeres con ingresos altos tan solo un 22% tiene síndrome metabólico.



Además, estudiamos la relación entre el estado civil y el síndrome separando entre hombres y mujeres, obteniendo los siguientes resultados:

- Los hombres casados, separados y viudos tienen un porcentaje de positivos muy parecido.
- Las mujeres viudas tienen un mayor porcentaje, seguido de las separadas y casadas.
- En ambos grupos, los solteros tienen un menor porcentaje.

Tras obtener esta información, cruzamos Ingresos con EstadoCivil, para ver si hay relación entre estas dos columnas, obteniendo que los separados y viudos tienen menor cantidad de ingresos de media. Además, hay relación estadística entre estas variables.



También encontramos relación entre etnia e ingresos y entre estado civil y etnias, destacando lo siguiente:

- Las poblaciones hispana y norteamericana tienen una menor media de ingresos, pudiendo relacionar este dato con su mayor porcentaje de positivos.
- La población asiática y negra tienen un porcentaje alto de solteros, siendo ambas poblaciones con un menor porcentaje de casos de síndrome metabólico (principalmente la etnia asiática).

## 5. Conclusiones

Respondiendo a nuestras hipótesis, podemos concluir:

*Existe relación entre los ingresos mensuales y el diagnóstico de síndrome metabólico.*

1. Los grupos con mayor porcentaje de síndrome metabólico son los que tienen una categoría de ingresos media y baja: < 6000\$ al mes.
2. El sexo también es decisivo: el grupo de mujeres con ingresos superiores a 6000\$ al mes tienen un menor porcentaje de casos de síndrome metabólico, mientras que en los hombres no parece ser significativo.

*El estado civil del sujeto tiene relación con el desarrollo de síndrome metabólico.*

1. Sí que hay relación entre el estado civil y el desarrollo del síndrome, destacando el caso del grupo de los viudos, que tienen un mayor porcentaje de positivos.
2. Los hombres casados, separados y viudos tienen un porcentaje muy parecido, mientras que las mujeres viudas tienen un mayor porcentaje. En ambos grupos, los solteros tienen un menor porcentaje. Es importante señalar que, tanto viudos como separados, tienen menos ingresos.
3. Con respecto a las etnias, destacamos que en la población asiática y negra hay un alto porcentaje de solteros, grupo que tiene un menor porcentaje de síndrome metabólico.

*La etnia del sujeto influye en la aparición de síndrome metabólico.*

1. Hay más casos de síndrome metabólico en las poblaciones hispana y norteamericana.
2. Podría estar relacionado con los ingresos de los sujetos, puesto que estas dos poblaciones tienen una media de ingresos más baja que el resto.

*La edad del sujeto influye en el diagnóstico de síndrome metabólico.*

1. La edad también es importante, a mayor edad, más casos de síndrome metabólico.

Esto abre puertas a futuras investigaciones, en las que habría que hacer hincapié en los hábitos de vida de cada grupo de sujetos:

- Indagar en la calidad y precios de los alimentos, intentando relacionarlo con los ingresos de cada grupo.
- Ver qué actividades físicas y deportivas hace cada grupo según su economía, estado civil y etnia.
- Estudiar los componentes genéticos de cada población para ver si las diferencias entre una etnia y otra se debe solo a hábitos de vida o a la fisiología de cada grupo.
- Relacionar alimentación, estilo de vida y actividad física o deportiva con la edad de los sujetos y con el lugar de residencia.

#### Fuentes:

1. Zimmet P, MM Alberti KG, Serrano Ríos M. Una nueva definición mundial del síndrome metabólico propuesta por la Federación Internacional de Diabetes: fundamento y resultados. Rev Española Cardiol [Internet]. 2005;58(12):1371–494. Available from: <https://www.revespcardiol.org/es-una-nueva-definicion-mundial-del-sindrom-articulo-13082533>
2. Rodolfo Lahsen M. Síndrome metabólico y diabetes. Rev Médica Clínica Las Condes [Internet]. 2014;25(1):47–52. Available from: <https://www.elsevier.es/es-revista-revista-medica-clinica-las-condes-202-articulo-sindrome-metabolico-diabetes-S0716864014700100>
3. Metabolic Syndrome: A Comprehensive Dataset on Risk Factors and Health Indicators [Internet]. 2023. Available from: <https://www.kaggle.com/datasets/antimoni/metabolic-syndrome>