

Distribución de las personas con diagnóstico de diabetes

¿Qué es la diabetes —————→

La diabetes es una enfermedad crónica que ocurre cuando el páncreas no produce suficiente insulina o cuando el cuerpo no puede utilizar eficazmente la insulina que produce. Esta condición puede llevar a altos niveles de glucosa en sangre, lo que causa complicaciones graves a largo plazo en el cuerpo.

¿Ésta puede afectar a todas las personas, sin importar la edad, el sexo o su ubicación?

Contenidos —————→

Objetivos

Metodología

Analisis

Resultados

Objetivos



Identificar y analizar como es que los factores demográficos y geográficos están relacionados con el diagnóstico de diabetes.

Metodología (Código)

Pandas

- Importación de datos con `read_csv`.
- Transformación de columnas a numéricas con `pd.to_numeric`.
- Manejo de valores faltantes con `dropna` y `fillna`.
- Reemplazo de valores y renombrado de columnas.

Seaborn y Matplotlib

- Creación de gráficos estadísticos (dispersión, distribuciones y cajas).
- Personalización de visualizaciones con etiquetas y escalas.

Scipy.stats

- Operaciones numéricas (promedios, desviaciones estándar y normalización).

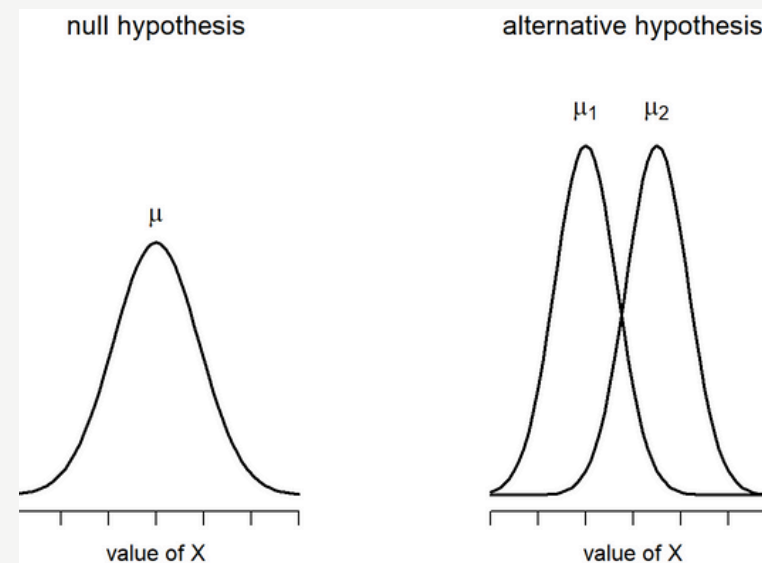
Metodología (Estadística)

Prueba Chi-cuadrada (χ^2)

- Evaluación de la independencia entre dos variables categóricas.
- Permite analizar si existe una relación significativa entre las categorías de las variables.

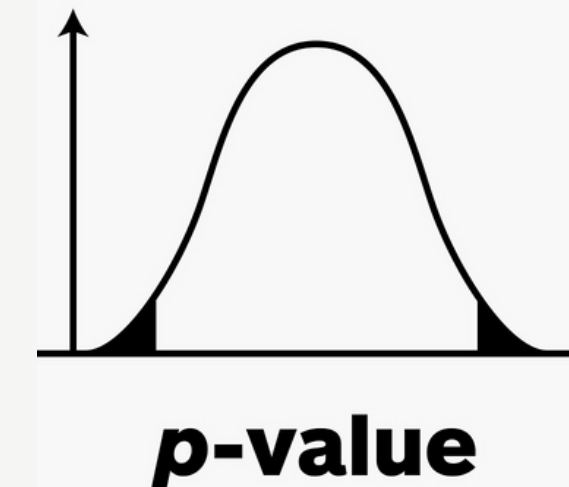
Prueba T ($t_{\text{test_ind}}$)

Comparación de medias entre dos grupos independientes.



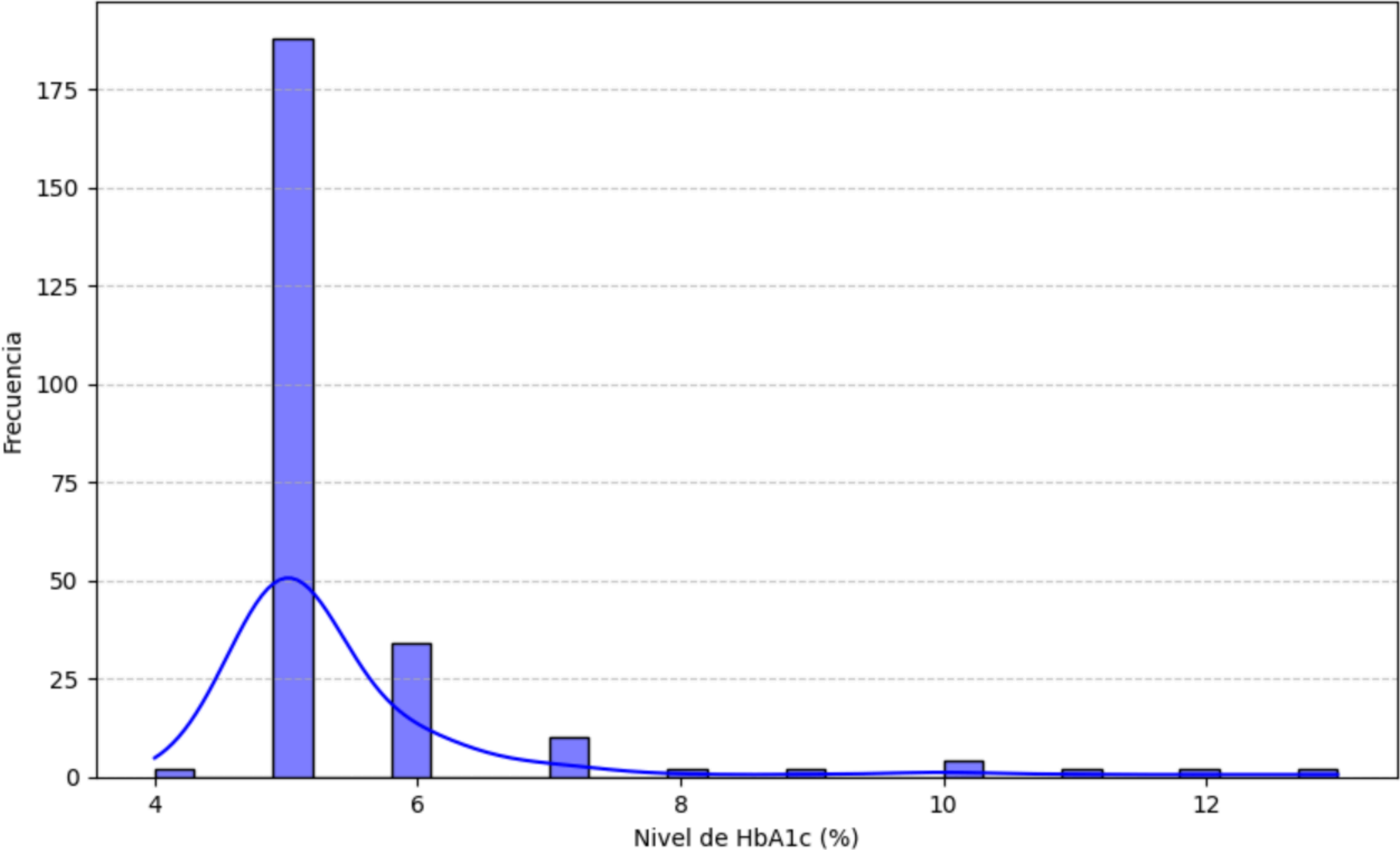
Valores P (Cálculo)

Evaluación de significancia estadística ($\alpha = 0.05$) para validar hipótesis.



(Devore, 2018)

Distribución de Glucosa en Sangre (HbA1c)



Interpretación

La gráfica nos da a entender tres regiones dentro de la población.

- Tendencia Central (5%):

Representa a la mayoría de la población con niveles normales de glucosa en sangre.

- Tendencia derecha (6%)

Incluye a personas con niveles de glucosa levemente elevados.

- Cola derecha (<6.5%)

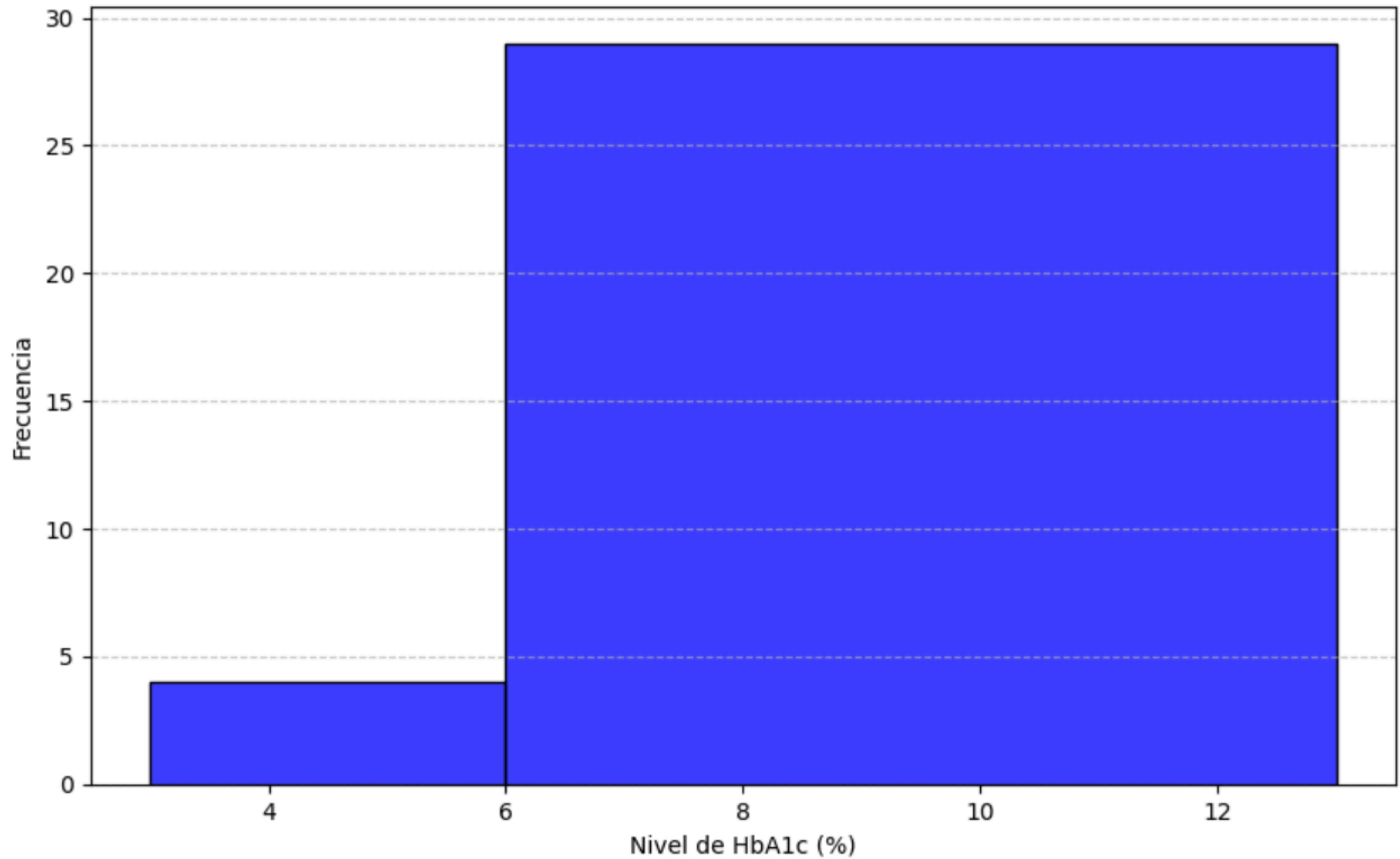
Muestra un grupo más pequeño, con niveles significativamente altos de glucosa.

(HbA1c)

La HbA1C mide el promedio de la glucosa en sangre de los últimos 3 meses, los cuales tienen unos niveles establecidos.

El HbA1c, es uno de los factores que se usa en medicina para determinar el estado de salud de personas con diabetes, o para determinar el mal manejo de la diabetes en una persona.

Distribución de Glucosa en Sangre (HbA1c) para H0902A = 1



Visualización de Niveles de Glucosa (HbA1C)

Diagnóstico = 1 (diabetes)

Interpretación

En esta sección, se toma en cuenta solamente una muestra de personas con diabetes.

Muestra la distribución de los niveles de glucosa en personas que tienen un diagnóstico relacionado con diabetes.



Rangos

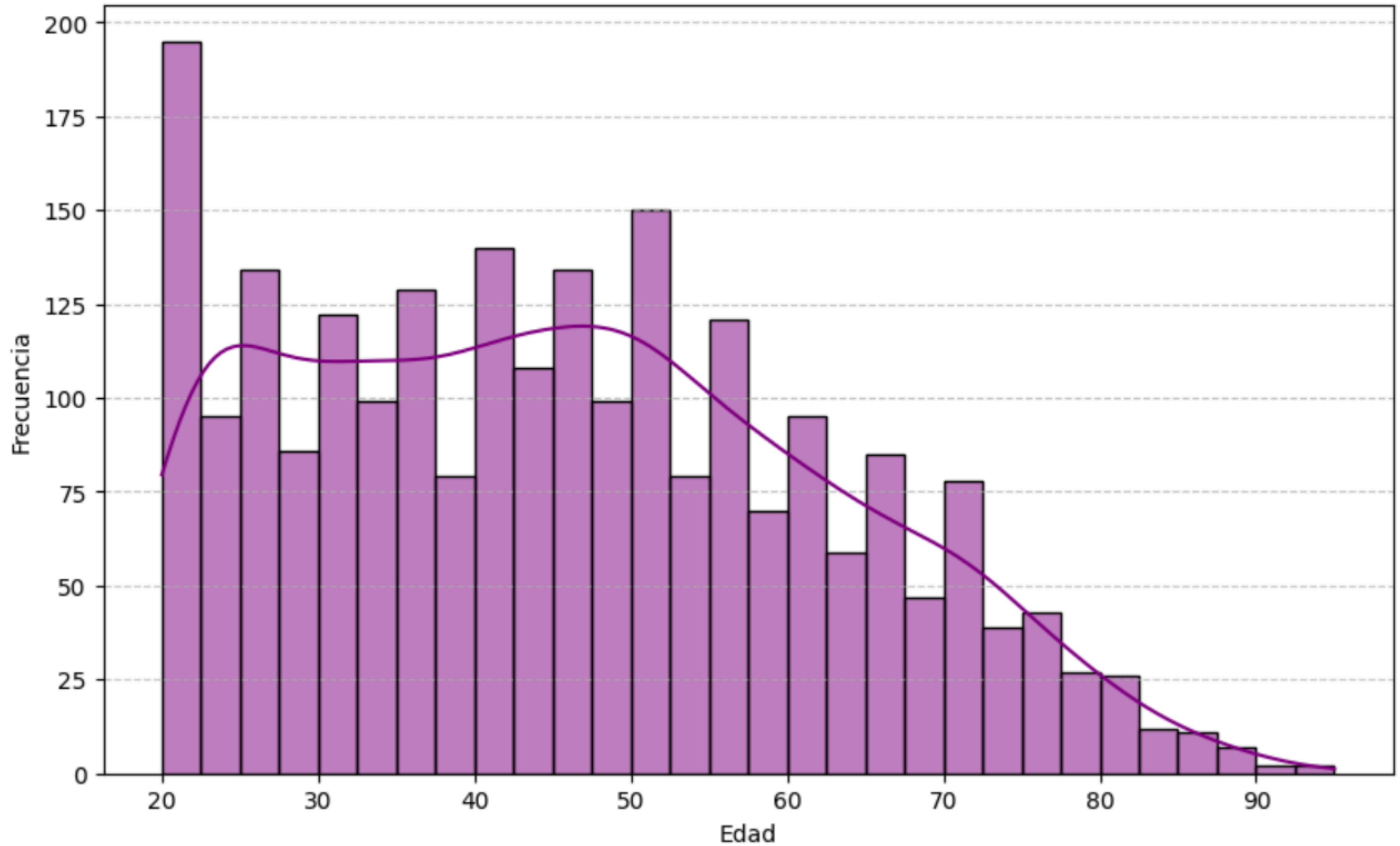
Rango 3-6%

Solo representa una pequeña proporción de la muestra, esto podría significar que son personas con diabetes bien controlada.

Rango >6%

Tienen niveles elevados de HbA1c, que podría indicar mal manejo de la diabetes.

Distribución de la Edad



Interpretación

El histograma muestra como están distribuidas las edades, dónde el pico más alto refleja el grupo de edad más común de la población.

Sesgo

Sin embargo la gráfica muestra un sesgo.

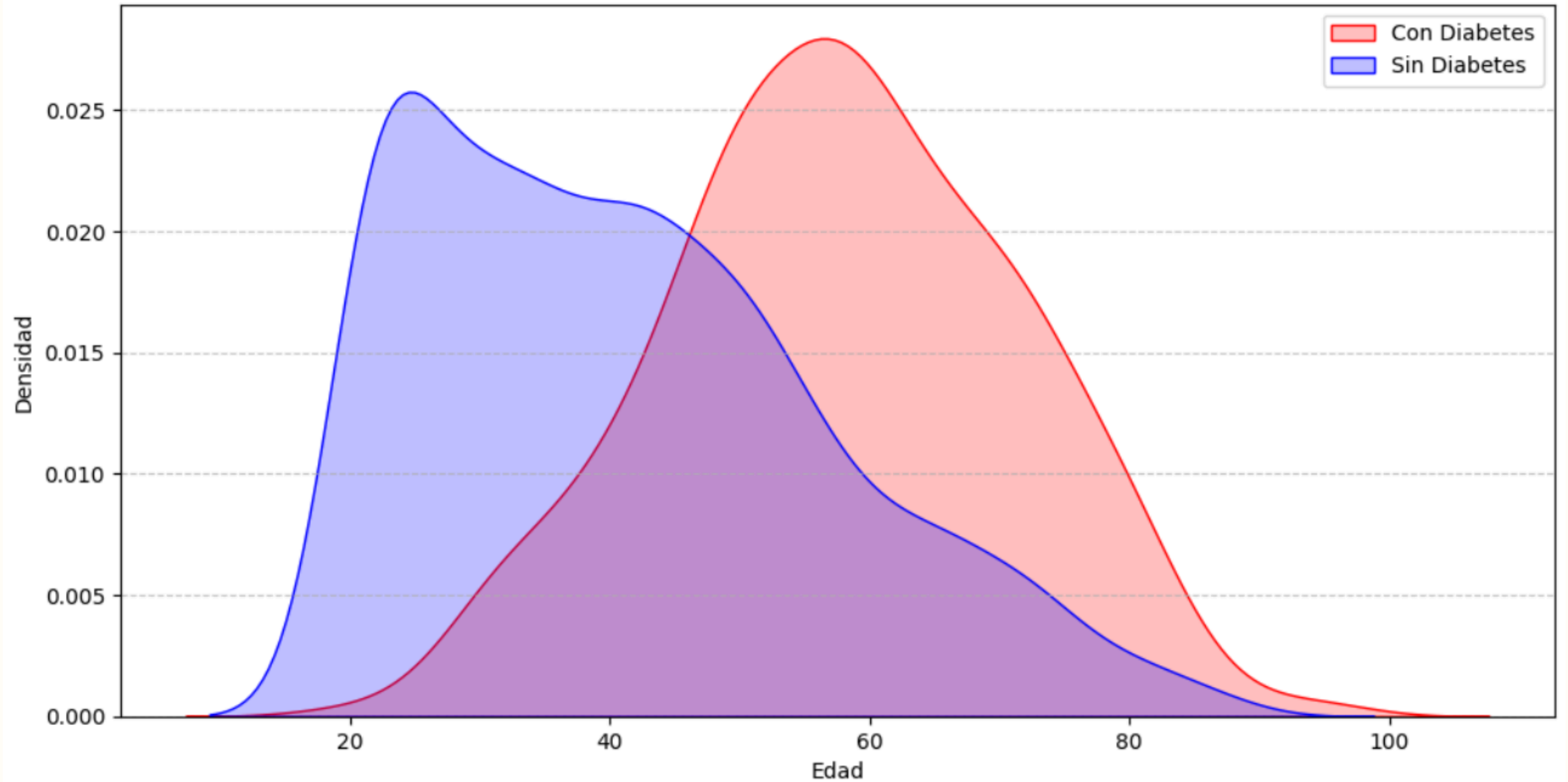
Este sesgo se denomina asimétrico positivo, lo que significa que hay más personas jóvenes y menos mayores , creando una cola hacia la derecha.

Utilidad

Entender cómo está dictada la frecuencia de edades, nos podrá ayudar a entender qué grupo de edad es más susceptible a la diabetes.

Y a su vez podremos saber si la diabetes tiene algún tipo de correlación con la edad.

Distribución de la Edad según Presencia de Diabetes



Distribucion de Edad segun presencialos de Diabetes

Interpretación



Analisis de Asimetría:

Skewness Con Diabetes: -0.05

Skewness Sin Diabetes: 0.64

Con esto vemos que las condicion de diabetes afecta de manera mas uniforme los rangos de edad. Mientras que las personas SIN diabetes se concentra mas en edades jovenes

Media CON diabetes: 57.68 años

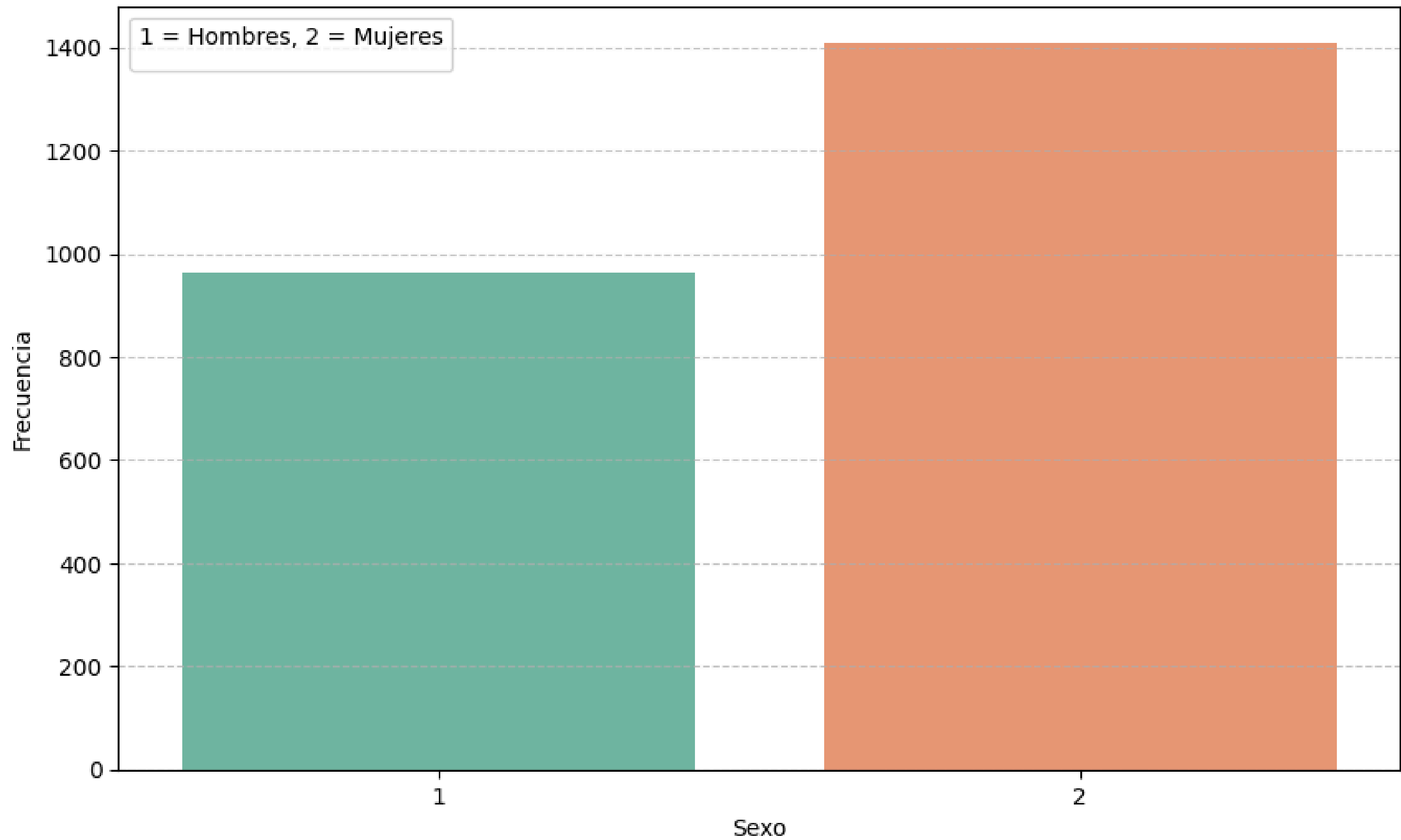
Media SIN diabetes: 41.09 años

T-statistic: 19.65

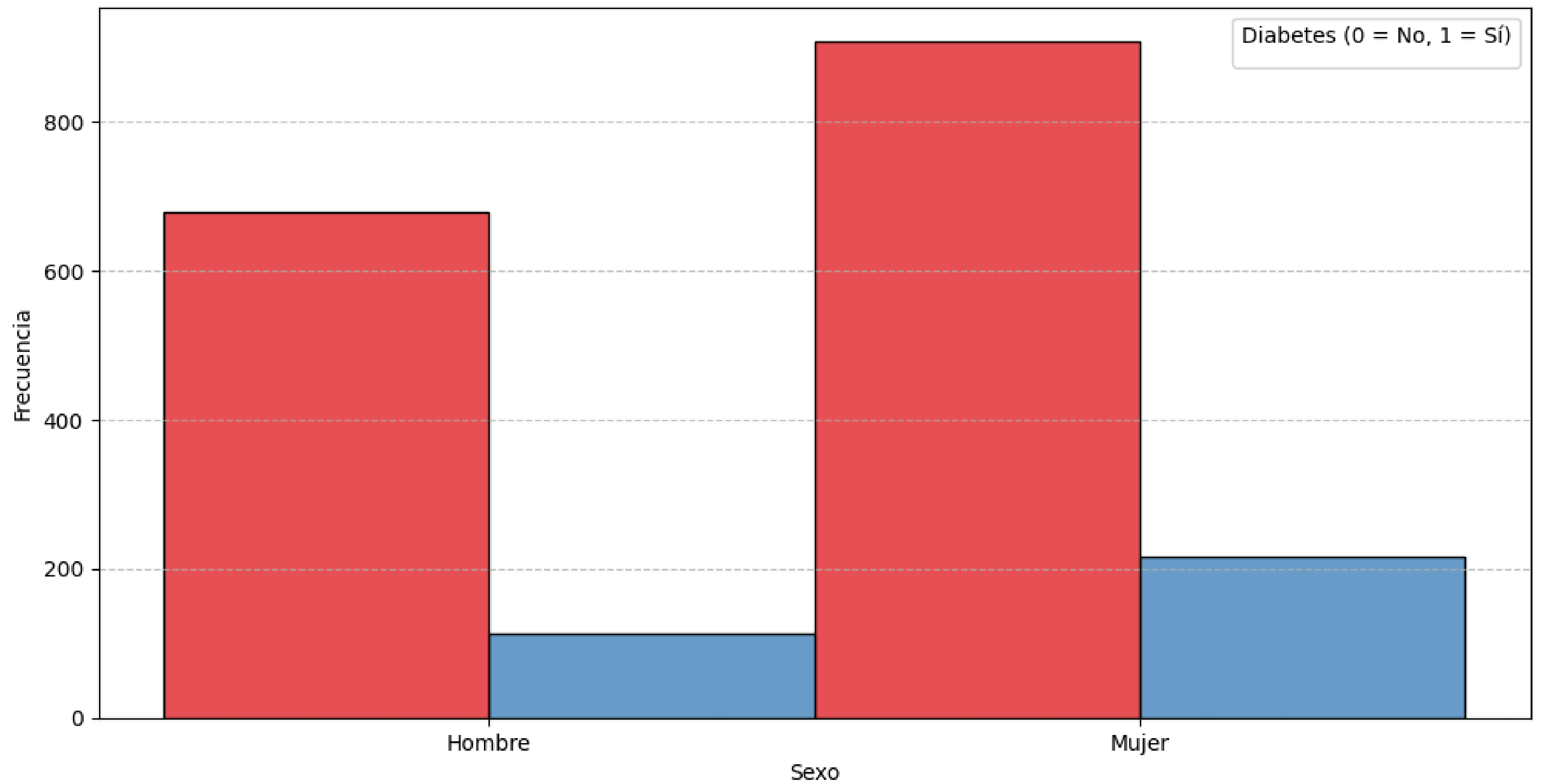
P-value: 5.24e-65 ($p < 0.05$)

La diferencia de medias entre los dos grupos es de 16.58 la cual es significativa. Esto se comprueba con el T-statistic de 19.65 y un p-value de 5.24e-65. Por lo que, la edad tiene una relación estadísticamente significativa en la presencia de diabetes.

Distribución por Sexo

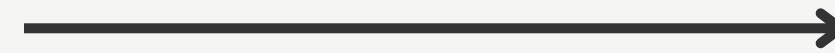


Distribución de la Diabetes según Sexo



Distribución de de Diabetes segun Sexo

Interpretación

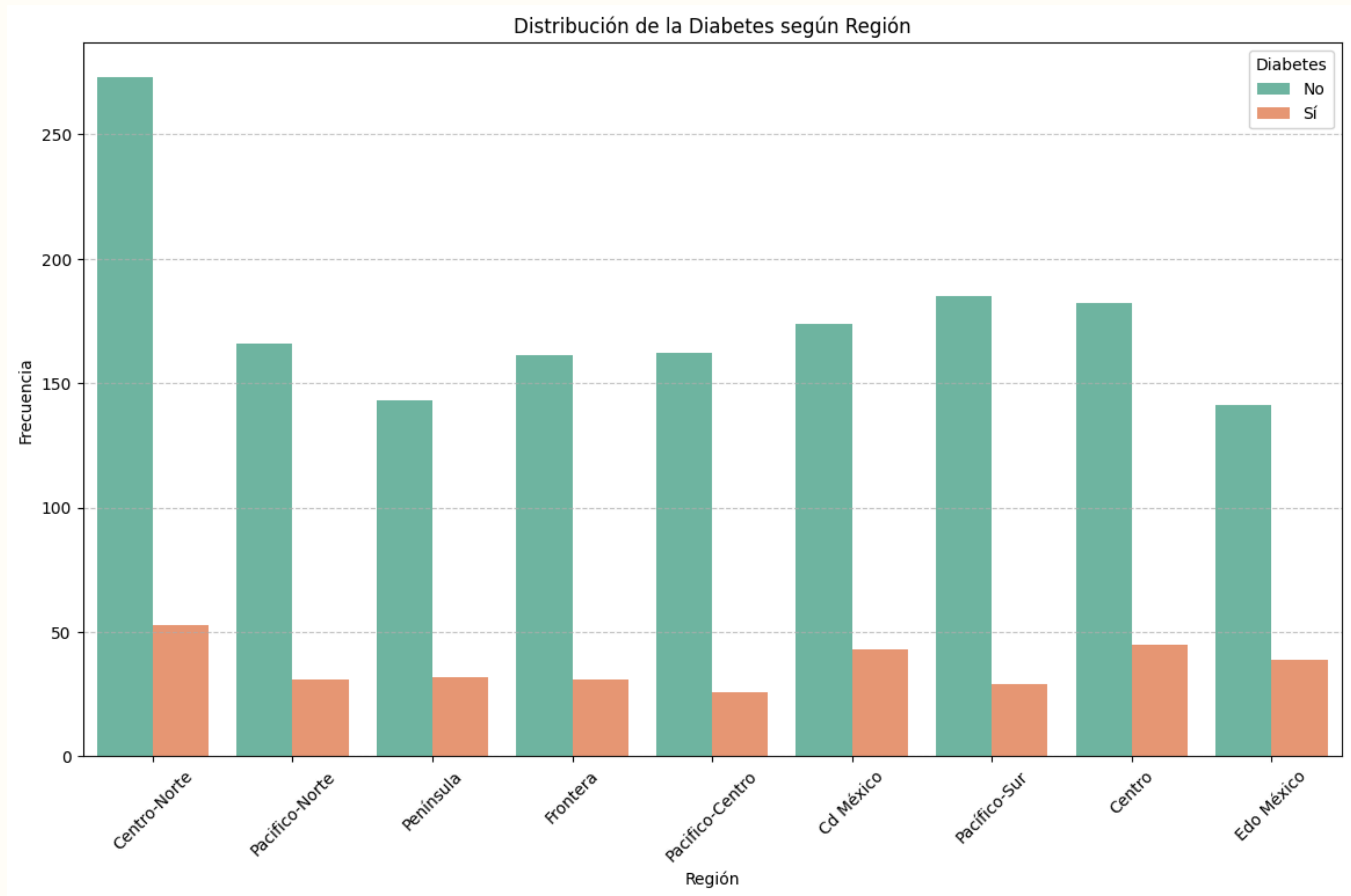


Chi² : 7.6581

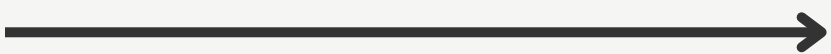
P-Value: 0.0056

Existe una asociación significativa entre el sexo y la presencia de diabetes ($p < 0.05$).

A partir de esto concluimos, con el p-value < 0.05 , la correlación es estadísticamente significativa. Lo que nos indica que hay una relación entre el sexo y la tenencia de diabetes.



Interpretación



Proporción de Diabetes por Region:

- 1. Edo México: 0.216667
- 2. Centro: 0.198238
- 3. Cd. México: 0.198157
- 4. Península: 0.182857
- 5. Centro-Norte: 0.162577
- 6. Frontera: 0.161458
- 7. Pacífico-Norte: 0.157360
- 8. Pacífico-Centro: 0.138298
- 9. Pacífico-Sur: 0.135514

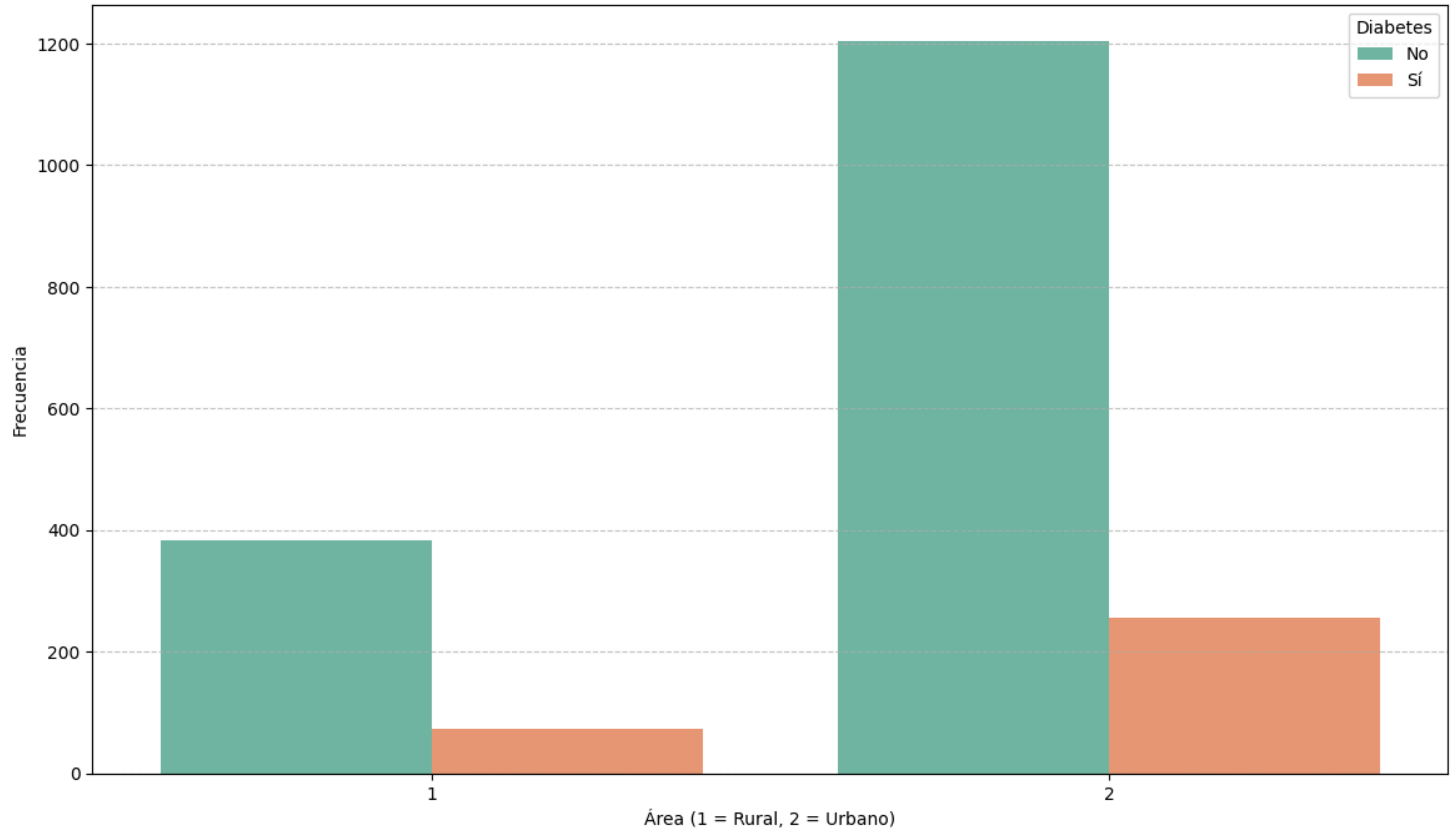
Chi2: 8.966321186777442

P-Value: 0.3451450335931555

No existe una asociación significativa entre la región y la presencia de diabetes ($p \geq 0.05$).

Dado que $p > 0.05$, no se rechaza la hipótesis anula. Lo que implica que no hay relación estadísticamente significativa

Distribución de la Diabetes según Área (Rural vs Urbano)



Distribución De La Diabetes según Área (Rural vs Urbano)

Interpretación

El análisis evalúa la proporción de personas con diabetes en áreas rurales y urbanas.

Valores Estadísticos

Chi2: 0.3188119795886933

P-Value: 0.5723224778733196

No existe una asociación significativa entre el área de residencia y la presencia de diabetes ($p \geq 0.05$).

Output

Muestra el promedio, de personas con diabetes en cada tipo de área, por proporciones:

- Rural

La proporción es: 0.1619 (16.2%)

- Urbano

La proporción es: 0.1747 (17.5%)

Cómo se puede apreciar las proporciones son bastante similares entre áreas rurales y urbanas.

Resultados

Infieren en la Diabetes

Con nuestro análisis la edad puede ser una factor en la tenencia o no de diabetes

De la misma manera el sexo.

NO infiere en la diabetes

De acuerdo a nuestro análisis, la región del país o la zona, ya sea Rural o Urbana. No son un factor determinante en la tenencia o no de diabetes.

Referencias

- User Guide — pandas 2.2.3 documentation. (n.d.-b).
https://pandas.pydata.org/docs/user_guide/index.html
- Devore, J. L. (2018). Probability and statistics for engineering and the sciences (9th ed.). Cengage Learning.
- National Library of Medicine. (s. f.). Prueba de hemoglobina glicosilada (HbA1c).
<https://medlineplus.gov/spanish/a1c.html>
- ¿Qué es la diabetes? - NIDDK. (s. f.). National Institute Of Diabetes And Digestive And Kidney Diseases. <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/que-es>
- Distribución de chi-cuadrada - Minitab. (s. f.). (C) Minitab, LLC. All Rights Reserved. 2024. <https://support.minitab.com/es-mx/minitab/help-and-how-to/probability-distributions-random-data-and-resampling-analyses/supporting-topics/distributions/chi-square-distribution/>