



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

FACULTAD DE CIENCIAS

Tarea 1

**Análisis descriptivo de pruebas para detectar  
Diabetes tipo 2**

Integrantes

**Aaron Mauricio Gómez Jiménez  
Bernardo Tobón Sánchez  
Israel Alejandro Altamirano García**

Profesora

**Ruth Fuentes García**

Asignatura

**Inferencia Estadística**

4 de septiembre de 2022

---

# 1. Introducción

La diabetes mellitus, o conocida simplemente como diabetes, es una enfermedad metabólica producida por deficiencias en la cantidad o en la utilización de la insulina (hormona que regula la concentración de azúcar en la sangre) y está asociada con factores genéticos o estilo de vida. Los efectos de la diabetes incluyen daños a largo plazo, disfunción e insuficiencia de varios órganos.

Existen dos principales tipos de diabetes: tipo 1 y tipo 2. La diabetes de tipo 1 que se presenta en pacientes jóvenes y es producida cuando el páncreas es incapaz de producir insulina. La diabetes de tipo 2 presente en la gran mayoría de los casos, que ocurre cuando no se produce suficiente insulina o cuando el organismo la utiliza eficazmente.

Por lo general, el diagnóstico de la diabetes en una persona es por medio de análisis de sangre que miden los valores de la glucosa plasmática o la hemoglobina glucosilada. Las tres pruebas son:

- **Glucosa en ayunas.** Esta prueba mide el nivel de glucosa en la sangre en un momento concreto, la cual se obtiene después de una noche de ayuno. El valor límite de diagnóstico es  $\geq 126 \text{ mg/dl}$ .
- **Glucosa pos carga (2h>75g).** El valor límite de diagnóstico es  $\geq 200 \text{ mg/dl}$ .
- **Hemoglobina glucosilada.** Esta prueba indica el nivel promedio de la glucosa sanguínea durante los últimos 2 a 3 meses. El valor para considerar a una persona con diabetes es si es mayor a 6.5 % en dos pruebas distintas.

En este trabajo se analizará los resultados de un grupo de personas de sus tres pruebas para diagnosticar diabetes de tipo 2.

## 2. Análisis

### 2.1. Limpieza de datos

Revisando la información, sabemos que contamos con 60 índices. De los cuales para la columna de Prueba de glucosa pos carga, 22 son no nulos, es decir el 63 % de las entradas de esa columna no se conocen. Eliminaremos esta columna pues no aporta suficiente información para rellenarla utilizando medidas de tendencia central.

Observando, notamos que los datos en el renglón 50 no tiene ningún dato en las primeras tres columnas, así que decidimos eliminar ese renglón por completo, ya que no proporciona ninguna información de las pruebas realizadas.

Decidimos completar la columna de Hemoglobina Glucosilada con la media de los datos de esa columna, ya que al ser solo 10 datos faltantes, tenemos el 91 % de datos con los cuales podemos hacer una buena aproximación para completar los datos faltantes. Completamos la columna con 2 datos diferentes: la media (8.177 %) de los datos de esa columna y una cifra que se considera normal (5.9 %) para una persona que no tiene diabetes de ningún tipo;

---

ya que al usar la media para rellenar toda la columna, no estaríamos tomando en cuenta si la persona es diabética, pues como vimos en la introducción la prueba de Hemoglobina Glucolisada con una prueba con resultado de 8.177 se considera una persona diabética, pero la columna de Tipo de diabetes nos dice que es una persona sana, es decir su prueba de Hemoglobina Glucolisada tendría que ser menor o igual a 5.9 %.

Conociendo ya la muestra, después de realizar la limpieza de datos, tenemos que está conformada por 60 personas, de las cuales 55 tienen diabetes tipo 2. Entonces la casi la mayoría de nuestra muestra es diabética.

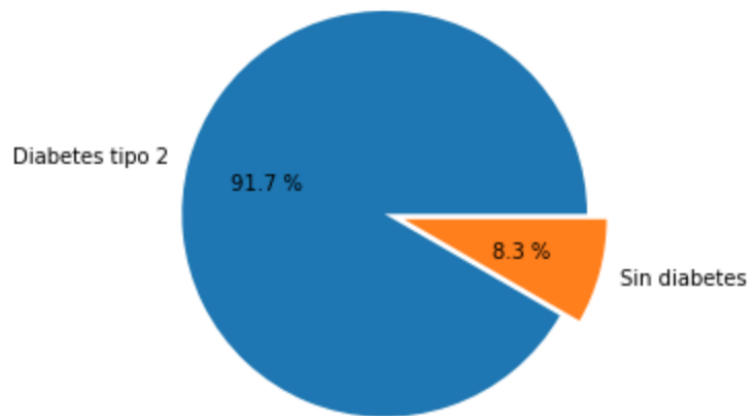


Figura 1: Gráfica de pastel Diabetes

## 2.2. Glucosa en ayunas

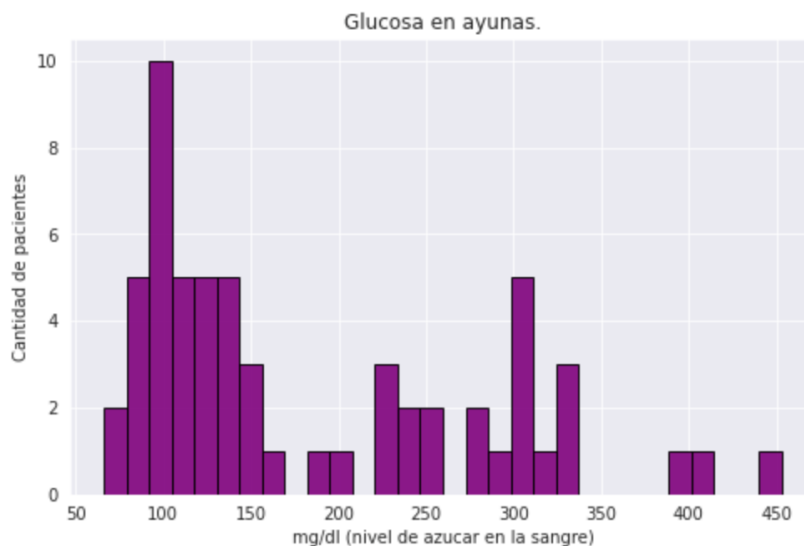


Figura 2: Histograma de Glucosa en ayunas

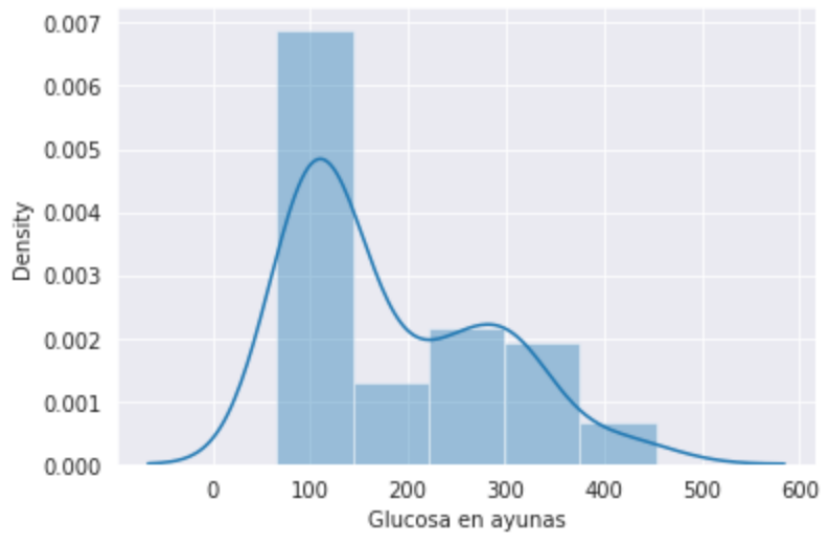


Figura 3: Histograma y densidad de Glucosa en ayunas

Al observar los datos en el histogramas (Figura 2 y 3) vemos que nuestra muestra no se asemeja a una variable aleatoria normal, sino a una variable aleatoria bimodal, puesto que podemos encontrar dos máximos dentro de la gráfica, aunque claramente la moda “principal” se aproxima a 100 podemos encontrar otro pico dentro de la distribución muy cerca del valor de 300.

Calculando la curtosis es  $-0.4363$ , que es negativa, y aunque es un valor pequeño esto nos dice que la distribución presenta un forma aplanada y con picos poco pronunciados y colas ligeras.

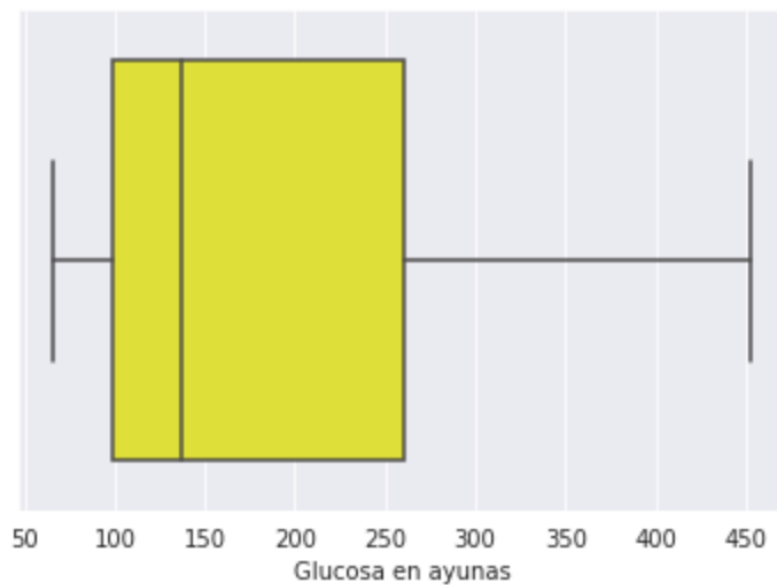


Figura 4: Gráfica boxplot de Glucosa en ayunas

---

Analizando el gráfico de boxplot (Figura 4) notamos que no existen puntos atípicos. Tenemos que la distribución es asimétrica y muy sesgada debido a que la parte más larga de la caja es la parte derecha, además el bigote derecho es mucho más largo, nos dice que la carga de la cola de la distribución se encuentra más larga del lado derecho.

Calculando los tres cuartiles tenemos que  $Q_1 = 99.5$ ,  $Q_2 = 137.0$ ,  $Q_3 = 260.5$ , entonces  $IQR = 161$  que es un número grande, que representa una mayor dispersión en los datos.

### 2.3. Hemoglobina glucosilada

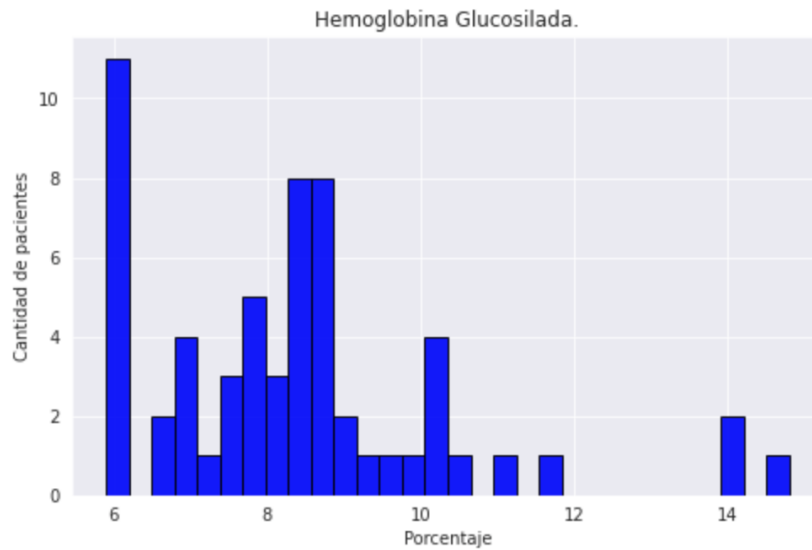


Figura 5: Histograma de Hemoglobina glucosilada

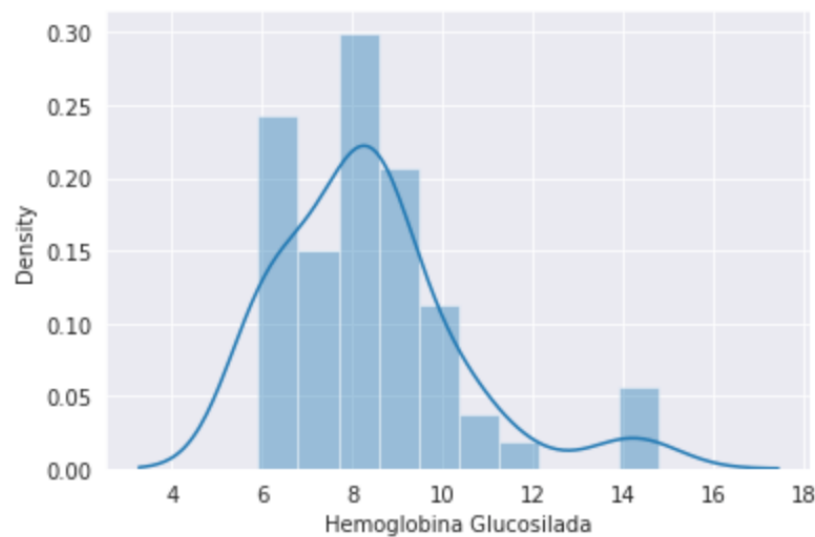


Figura 6: Histograma y densidad de Hemoglobina glucosilada

A diferencia de nuestro conjunto de datos anterior podemos observar que nuestra muestra sobre la Hemoglobina si presenta una similitud mayor a una variable aleatoria normal, podríamos decir que al mismo tiempo presenta una propiedad bimodal al ver un pico poco pronunciado cerca del valor de 14, a pesar de eso podemos distinguir que la moda "principal" se encuentra muy cerca del valor de 8.

Por otro lado la curtosis es 1.9035, presenta ser positiva (caso contrario al conjunto anterior de datos) esto nos dice que que la distribución cuenta con picos pronunciados y colas muy pesadas, que podemos observar claramente en la gráfica justo donde se encuentra la moda, el pico es pronunciado y que sus colas hacia ambos lados tienen una carga grande.

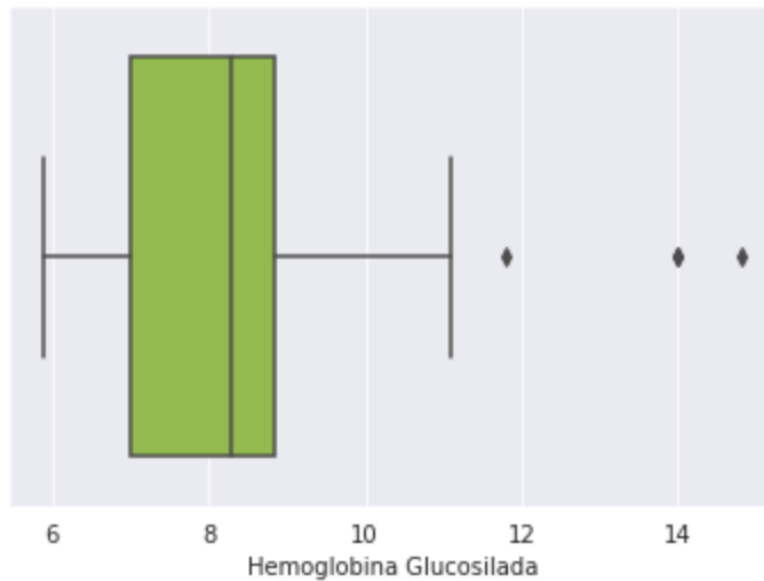


Figura 7: Gráfica boxplot de Hemoglobina glucosilada

Analizando el boxplot (Figura 7), que a diferencia del de la Glucosa, cuenta con tres puntos atípicos. También notamos que el sesgo es positivo, al igual que en el caso anterior podemos concluir que la carga de la cola de la distribución se encuentra hacia el lado derecho, lo cual es claro al ver un segundo pico poco pronunciado hacia el lado derecho de la moda. Calculando los tres cuartiles tenemos que  $Q_1 = 7.0$ ,  $Q_2 = 8.3$ ,  $Q_3 = 8.85$ , entonces  $IQR = 1.85$  que es un número pequeño, que significa que los datos no están dispersos.

## 2.4. Correlación entre las variables

Una vez que tenemos el dataframe limpio podemos observar la correlación que existe entre los datos dispersos entre las dos pruebas, sacando la matriz de correlación notamos que su relación no es muy buena, puesto que los valores no son muy cercanos a 1.

|                         | Glucosa en ayunas | Hemoglobina Glucosilada |
|-------------------------|-------------------|-------------------------|
| Glucosa en ayunas       | 1                 | 0.645433                |
| Hemoglobina Glucosilada | 0.645433          | 1                       |

---

Para verlo de una manera más clara graficamos la matriz de correlación por medio de histogramas y una gráfica de dispersión.

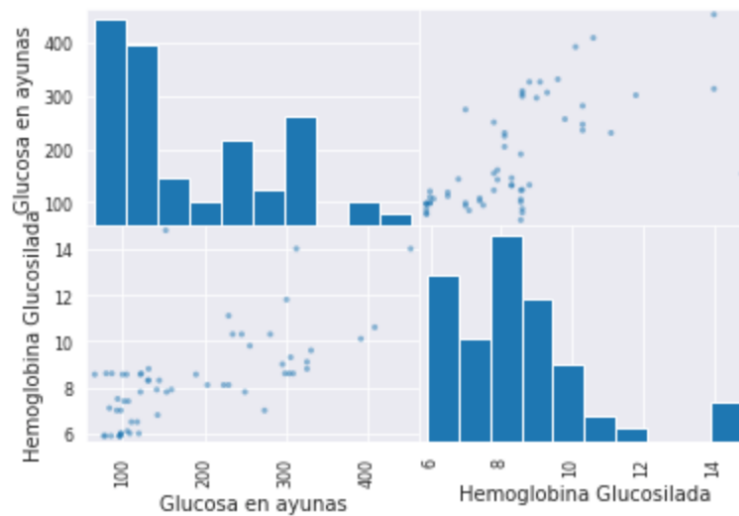


Figura 8: Correlación entre las variables

### 3. Conclusión

Revisando la información que se nos proporciono, nos percatamos que no conocíamos datos de los resultados, especialmente de la prueba de Glucosa pos carga, así que decidimos no tomarla en cuenta, ya que no aportaba gran información.

Ya limpia la muestra, tenemos que el 91.7% de la muestra tiene diabetes tipo 2, es decir del total de la muestra hay cinco personas sanas. Por lo que necesitaran seguir un tratamiento para evitar mayores complicaciones. Por consecuencia en las dos pruebas obtuvimos que sus medias se encontraban superiores a los valores limites de una persona no diabética; en el caso de la prueba de Glucosa en ayunas tenemos que media es de  $182.53 \text{ mg/dl}$  y en la de Hemoglobina glucosilada tenemos que la media es de 8.177%.

En el caso de la distribución de los datos a partir de las gráficas obtenidas tenemos que en la prueba de Glucosa en ayunas existe una mayor dispersión de los datos y su distribución es asimétrica. En la prueba de Hemoglobina presenta una similitud mayor a una variable aleatoria normal y a diferencia de la otra prueba los datos no están dispersos. Ambas pruebas tienen un sesgo a la derecha, que se justifica ya que los valores de las pruebas son altos en personas diabéticas. Tampoco podemos asegurar que exista una relación fuerte entre la Hemoglobina glucosilada y la Glucosa en ayunas.

#### Referencia:

★ Organización Panamericana de la Salud. (2020). *Diagnóstico y manejo de la diabetes de tipo 2 (HEARTS-D)*. Recuperado de <http://iris.paho.org>