

# INFORME 1° PRACTICA

Paula Andrea Barrera Silva  
Universidad Autónoma de Bucaramanga  
Estadística inferencial y muestreo

Andrés Fabian Leal Archila

6 de agosto de 2025

**Base de datos:** Nacidos vivos en el municipio de Bucaramanga

**Variables analizadas:** Peso, Hora de nacimiento, Código de institución, Departamento, Área de nacimiento, Grupo Sanguíneo.

**1). Análisis de variables categóricas:** Se generaron diagramas de torta y barra para las siguientes variables

Departamento

-Todos los nacimientos registrados pertenecen al departamento de Santander, lo cual era esperado dado al enfoque municipal del conjunto de datos.

Área de nacimiento

-Se identificaron principalmente dos áreas: Cabecera municipal y centro poblado o rural.

-Predomina el área Cabecera municipal, lo cual es consistente con el mayor acceso a servicios de salud en zonas urbanas.

Grupo Sanguíneo

-Los grupos sanguíneos más frecuentes fueron O, A y B, con un claro predominio del grupo O.

-Se observaron algunas categorías con datos faltantes o poco comunes, que podrían estar relacionadas con errores de registro o casos raros.

**2). Análisis de Variables Numéricas:** Se exploraron dos variables continuas

Peso (gramos)

-Histograma inicial mostró una distribución con ligera asimetría a la izquierda.

- Boxplot evidenció la presencia de varios outliers, con valores menores a 2000g (posibles nacimientos prematuros o errores).

Hora de Nacimiento

-Aunque es una variable tipo tiempo, se transformó en horas numéricas para su análisis.

- El histograma mostró que los nacimientos ocurren con mayor frecuencia entre las 6:00 am y 11:59 am.
- También se observaron nacimientos durante la noche, aunque en menor cantidad.

### **3). Identificación y Eliminación de Outliers (IQR):**

- Aplicando el rango intercuartílico (IQR) al peso, se eliminaron los valores atípicos para mejorar la representatividad del análisis.
- El límite inferior fue de aproximadamente 2500g, por lo cual se eliminaron registros por debajo de este valor.
- Después de la limpieza, el nuevo histograma mostró una distribución más simétrica y ajustada.

### **4). Histogramas sin Outliers:**

Una vez depurados los datos:

- El nuevo histograma del peso mostró una \*\*distribución más concentrada alrededor de los 3000-3500g.
- Esto facilita la interpretación y evita sesgos debidos a nacimientos prematuros o errores de digitación.

### **5). Pruebas de Normalidad:**

Se aplicaron cuatro pruebas estadísticas de normalidad al peso (sin outliers):

1. Shapiro-Wilk
2. Kolmogórov-Smirnov
3. Anderson-Darling
4. Jarque-Bera

Además, se construyó el gráfico QQ.

**Resultados:**

Algunas pruebas (como Shapiro-Wilk) rechazaron la hipótesis de normalidad ( $p < 0.05$ ).

El gráfico QQ mostró ligeras desviaciones en las colas.

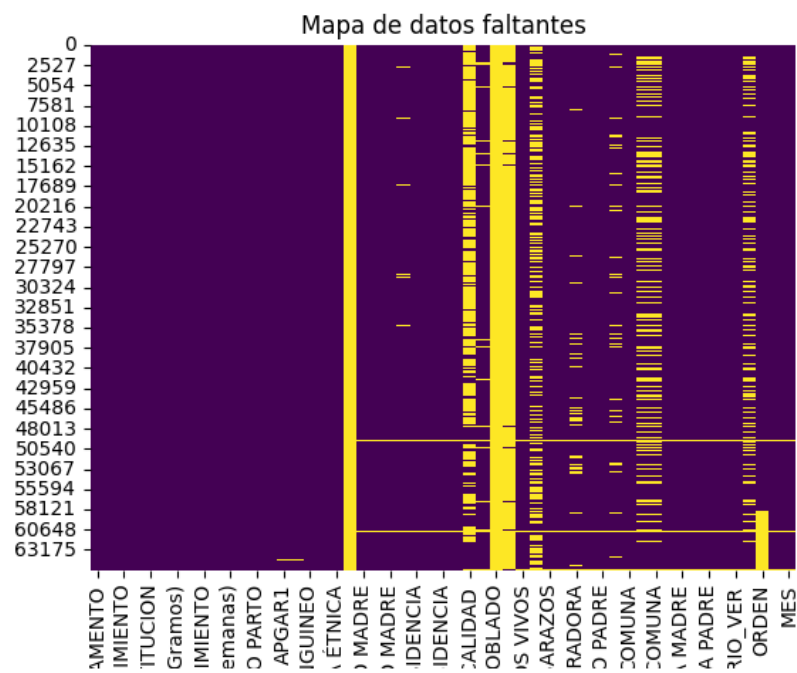
Por tanto, se concluye que la variable peso sin outliers no sigue una distribución completamente normal, aunque se aproxima visualmente.

**Conclusiones**

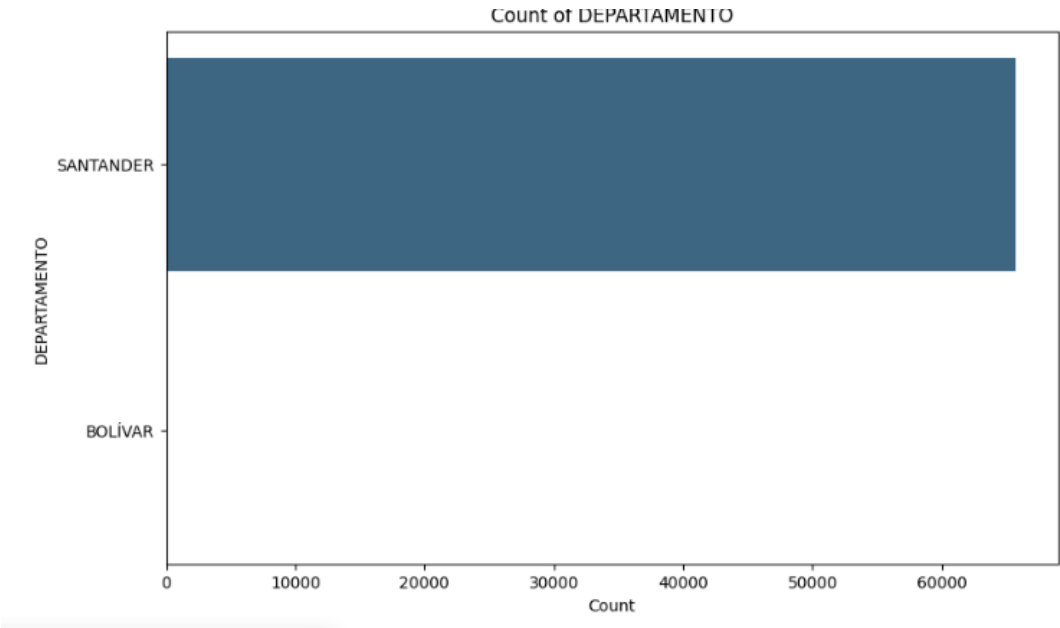
- El análisis permitió observar patrones significativos en variables como \*\*peso y hora de nacimiento.
- Se identificó que la mayoría de los nacimientos ocurre en la \*\*cabecera municipal\*\*, y en instituciones con un mismo código (lo cual indica concentración en pocos centros de salud).
- La eliminación de outliers fue clave para mejorar la calidad de los análisis.
- Se recomienda tener precaución al aplicar pruebas paramétricas, ya que no se cumple completamente la normalidad.

Anexos

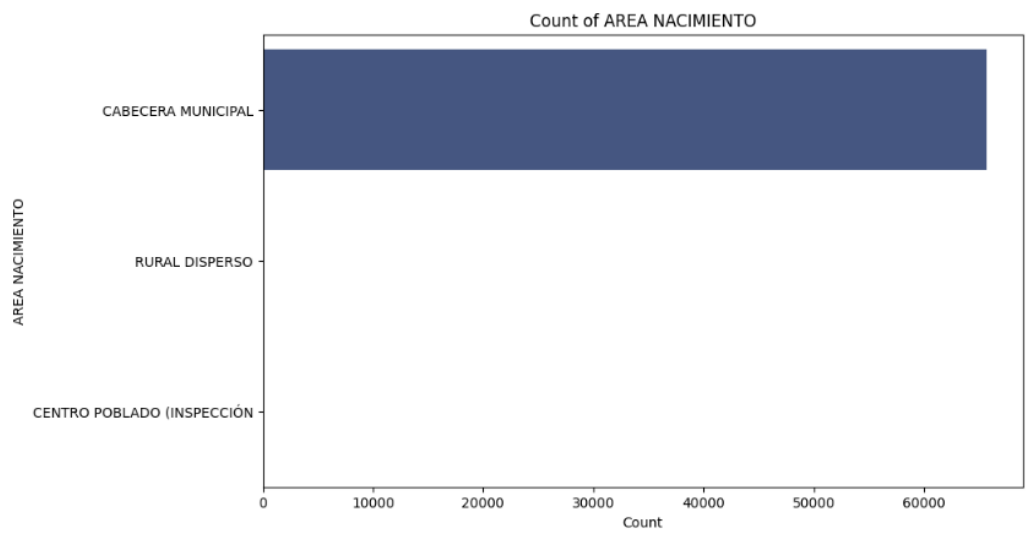
1.



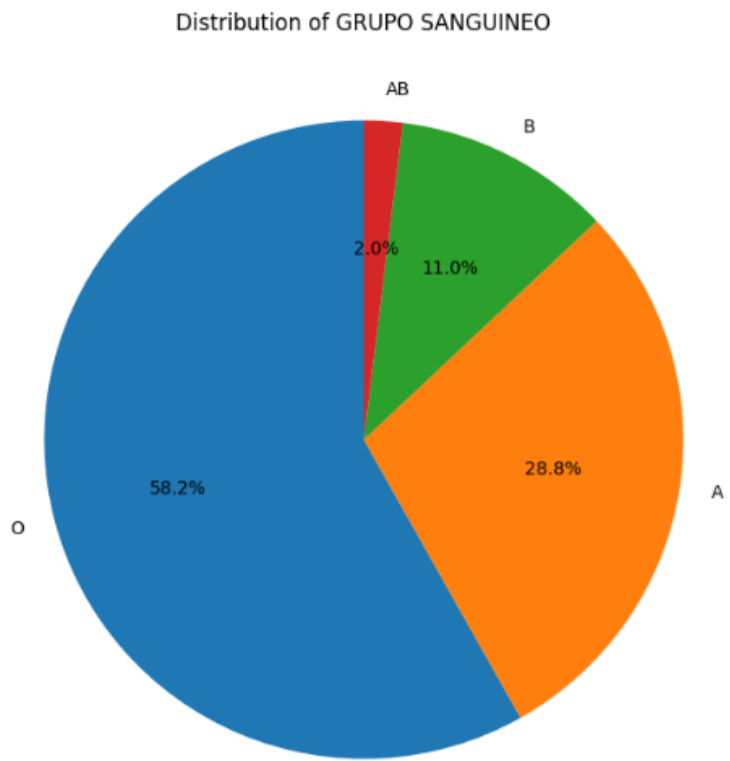
2.



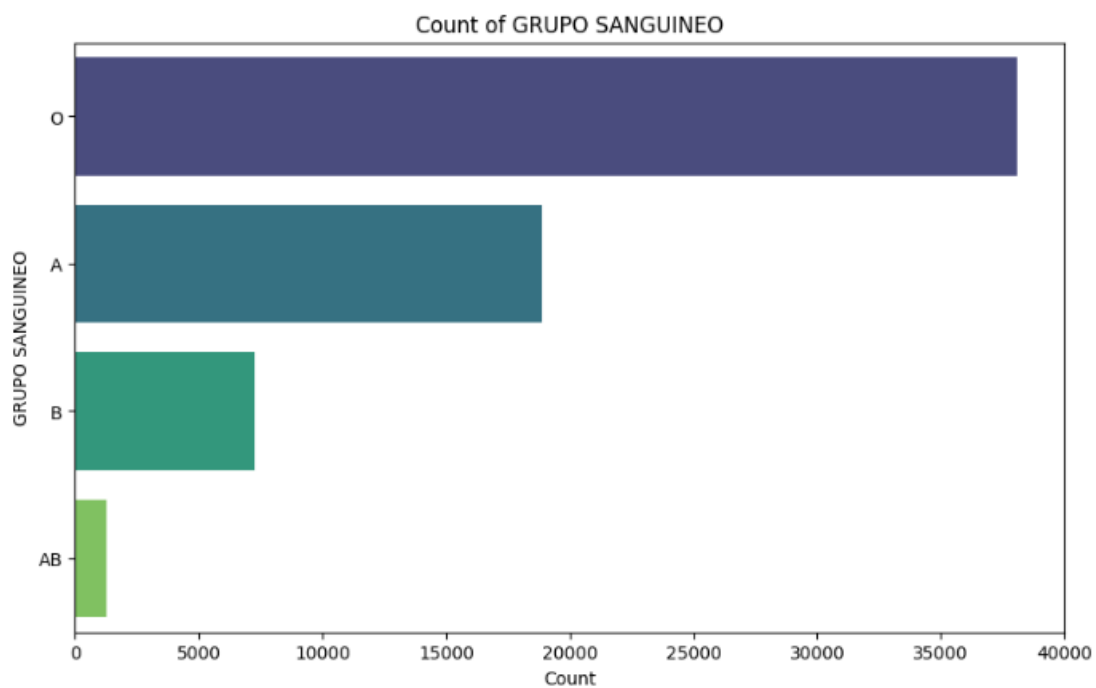
3.



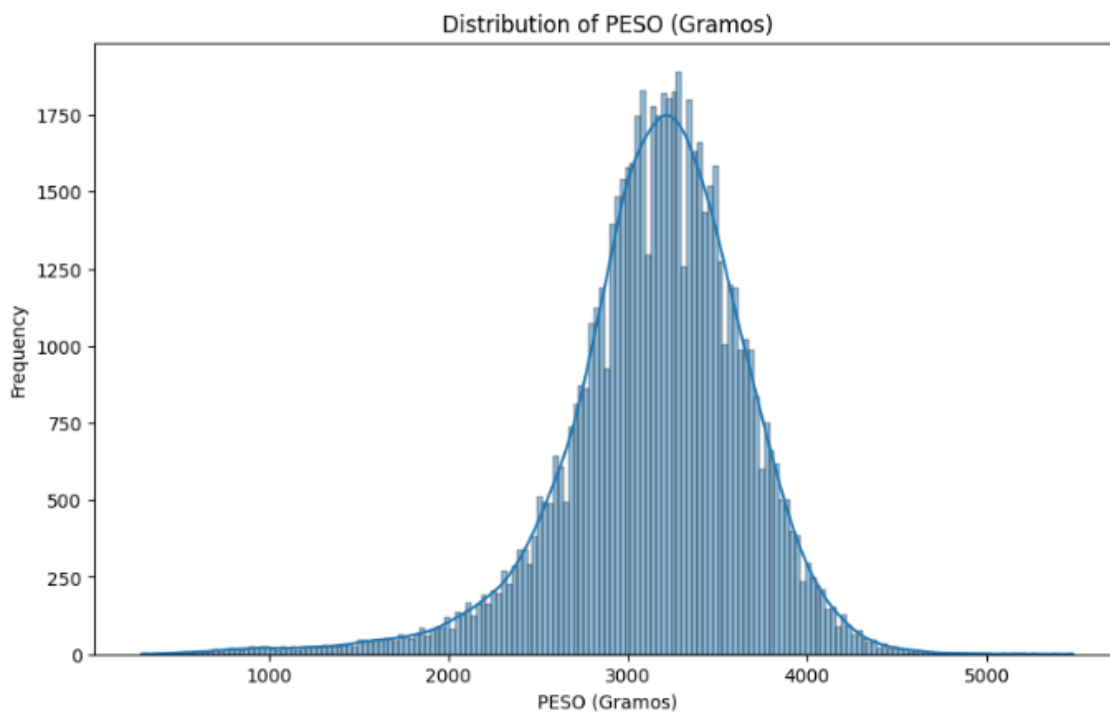
4.



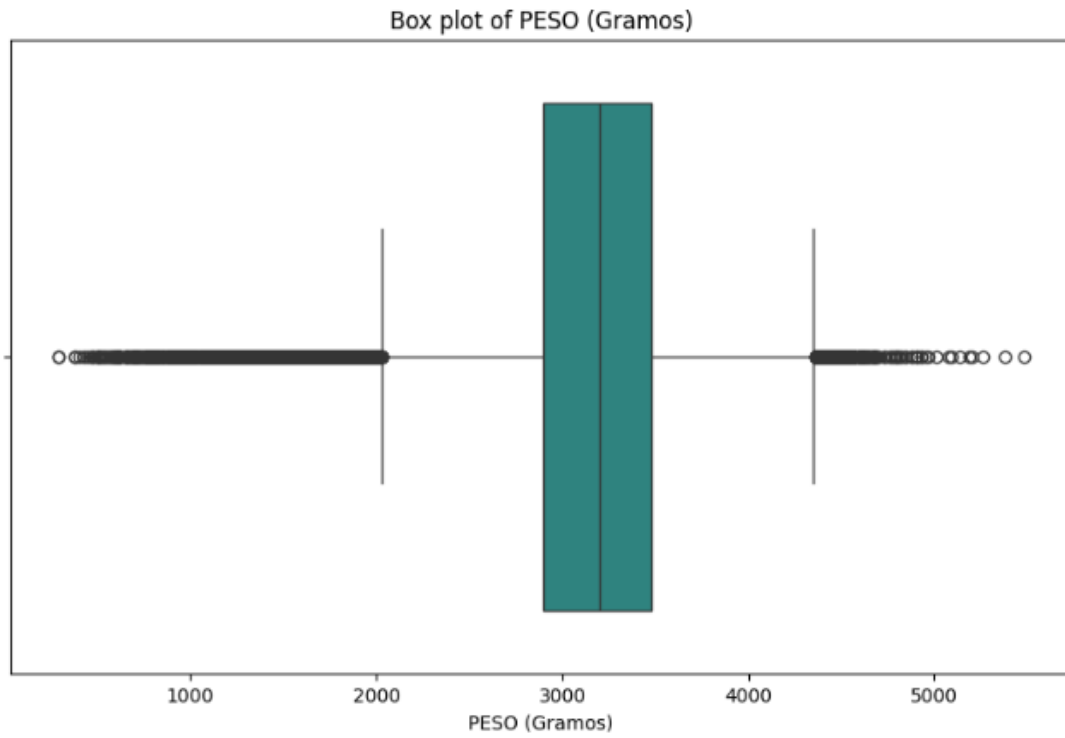
5.



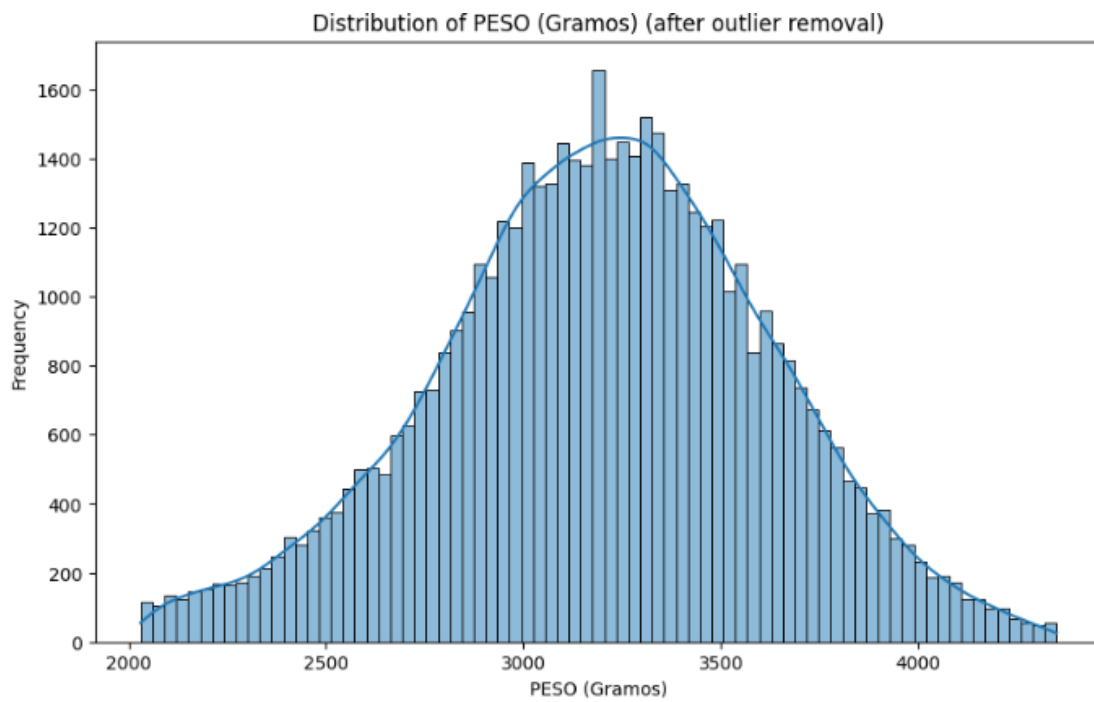
6.



7.

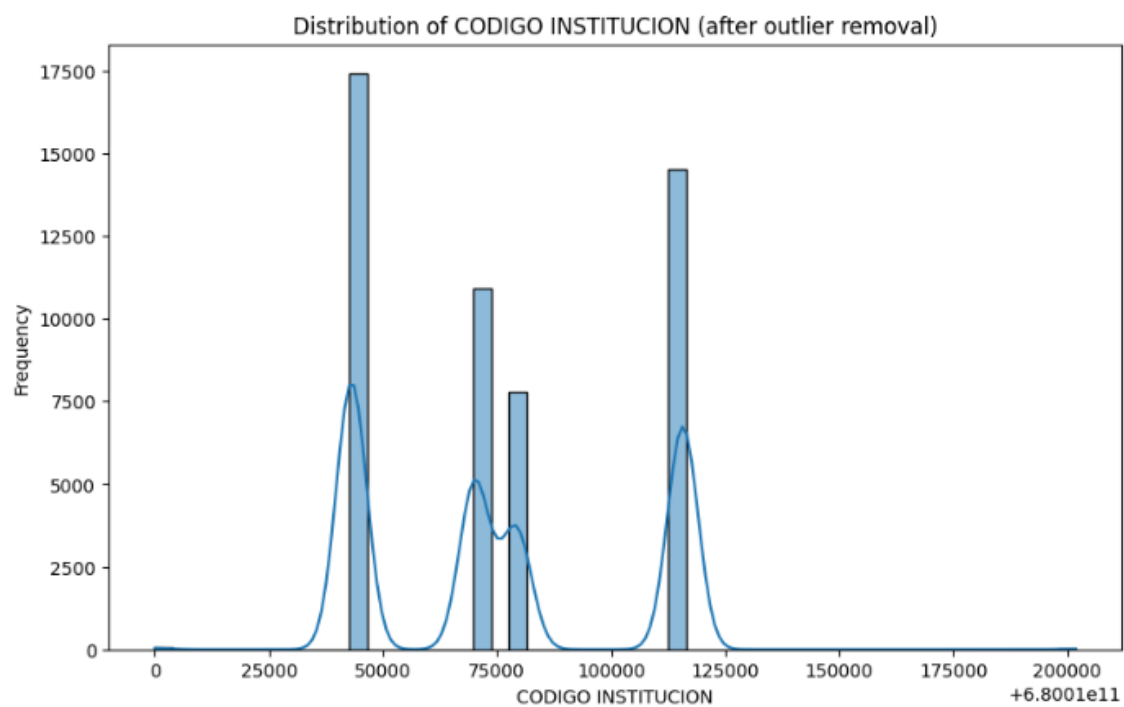


8.

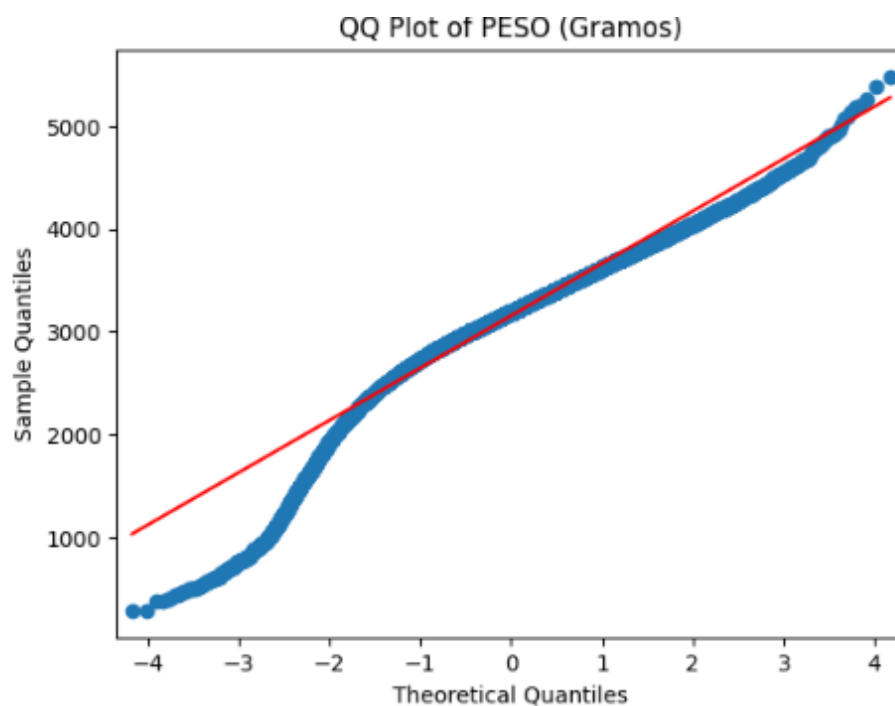




9.



10.



## **Bibliografías**

<https://datos.gov.co/browse?q=43.+Nacidos+Vivos+en+Municipio+de+Bucaramanga&sortBy=relevance&page=1&pageSize=20>