**Universidad Tecnológica de Panamá**

**Facultad de Ingeniería de Sistemas Computacionales**

**Licenciatura en Ingeniería de Software**

**Estadística con apoyo informático**

**Proyecto Final Tema:**

Análisis estadístico de bases de datos

**Facilitador:**

Ing. Juan Castillo

**Nombre:**

Diego Alvarado (8-997-1785)

**Grupo:**

1SF-131

**Fecha de entrega:**

26 de julio de 2023

Contenido

[Introducción 4](#_Toc141231382)

[Contenido 5](#_Toc141231383)

[CONSUMO DE COMBUSTIBLE EN NEW YORK 5](#_Toc141231384)

[Descripción de las variables: 5](#_Toc141231385)

[Tipo de análisis: 5](#_Toc141231386)

[Escogencia de la muestra: 5](#_Toc141231387)

[Histogramas 6](#_Toc141231388)

[Diagrama de Pareto 8](#_Toc141231389)

[Análisis de Regresión Lineal 9](#_Toc141231390)

[CANTIDAD DE CALORIAS EN COMIDAS RÁPIDAS 10](#_Toc141231391)

[Descripción de las variables: 10](#_Toc141231392)

[Tipo de análisis: 10](#_Toc141231393)

[Escogencia de la Muestra: 10](#_Toc141231394)

[Histogramas 10](#_Toc141231395)

[Diagrama de Pareto 12](#_Toc141231396)

[Análisis de Regresión Lineal 13](#_Toc141231397)

[PRECIOS DE LAPTOPS 15](#_Toc141231398)

[Descripción de variables 15](#_Toc141231399)

[Tipo de análisis: 15](#_Toc141231400)

[Escogencia de la muestra: 15](#_Toc141231401)

[Histogramas 15](#_Toc141231402)

[Diagrama de Pareto 17](#_Toc141231403)

[Análisis de Regresión Lineal 18](#_Toc141231404)

[IMPORTACIONES VALOR CIF Y PESO 20](#_Toc141231405)

[Descripción de las variables 20](#_Toc141231406)

[Tipo de análisis: 20](#_Toc141231407)

[Escogencia de la muestra: 20](#_Toc141231408)

[Histogramas 20](#_Toc141231409)

[Diagrama de Pareto 22](#_Toc141231410)

[Análisis de Regresión Lineal 23](#_Toc141231411)

[TAMAÑO Y PESO DE PECES 24](#_Toc141231412)

[Descripción de variables 24](#_Toc141231413)

[Tipo de análisis: 24](#_Toc141231414)

[Escogencia de la muestra: 24](#_Toc141231415)

[Histogramas 25](#_Toc141231416)

[Diagrama de Pareto 27](#_Toc141231417)

[Análisis de Regresión Lineal 27](#_Toc141231418)

[PIB\_PROVINCIA 29](#_Toc141231419)

[Descripción de las variables 29](#_Toc141231420)

[Tipo de análisis: 29](#_Toc141231421)

[Escogencia de Muestra: 29](#_Toc141231422)

[Histogramas 29](#_Toc141231423)

[STUDENT PERFORMANCE 33](#_Toc141231424)

[Descripción de las variables 33](#_Toc141231425)

[Tipo de análisis: 33](#_Toc141231426)

[Escogencia de la muestra: 33](#_Toc141231427)

[Histogramas 34](#_Toc141231428)

[Diagrama de Pareto 36](#_Toc141231429)

[MALLS 36](#_Toc141231430)

[Descripción de las variables 36](#_Toc141231431)

[Tipo de análisis: 36](#_Toc141231432)

[Escogencia de la muestra: 36](#_Toc141231433)

[Histogramas 37](#_Toc141231434)

[Análisis 39](#_Toc141231435)

[FECUNDIDAD 39](#_Toc141231436)

[Análisis de las variables 39](#_Toc141231437)

[Tipo de análisis 39](#_Toc141231438)

[Escogencia de la muestra 39](#_Toc141231439)

[Histogramas 40](#_Toc141231440)

[Conclusión 42](#_Toc141231441)

[Anexos 44](#_Toc141231442)

# Introducción

En este proyecto final de estadística, nos sumergiremos en el análisis detallado de 10 bases de datos utilizando los conocimientos adquiridos en nuestras clases. Aplicaremos herramientas estadísticas fundamentales como la media, mediana, varianza, histogramas y desviación estándar para obtener una visión completa de cada conjunto de datos. Aprovecharemos las ventajas de herramientas modernas como Python y Excel para facilitar y agilizar nuestro análisis, permitiéndonos procesar datos de manera eficiente y obtener resultados precisos.

Además de las técnicas tradicionales de estadística descriptiva, exploraremos la toma de muestras y el análisis de convergencia para garantizar que nuestras conclusiones sean representativas y confiables. Identificaremos el tipo de distribución presente en los datos, lo que nos brindará una comprensión más profunda de sus características. A través del análisis de correlación y el análisis lineal, estudiaremos las relaciones entre variables y podremos realizar predicciones valiosas basadas en datos históricos.

Al finalizar este proyecto, estaremos equipados con habilidades analíticas sólidas y la capacidad de tomar decisiones informadas y fundamentadas en datos. La combinación de técnicas estadísticas clásicas y herramientas modernas nos permitirá explorar y desbloquear información valiosa en estos 10 conjuntos de datos, proporcionándonos una ventaja significativa en cualquier campo donde la estadística sea un factor crítico para el éxito y la toma de decisiones acertadas.

# Contenido

## CONSUMO DE COMBUSTIBLE EN NEW YORK

### Descripción de las variables:

Name – Dato No Numérico (Tipo texto).

MPG – Dato Numérico (Tipo número).

Cylinders – Dato Numérico (Tipo número).

Displacement – Dato Numérico (Tipo número).

Horsepower: – Dato Numérico (Tipo número).

Weight – Dato Numérico (Tipo número).

Acceleration – Dato Numérico (Tipo número).

Model\_year – Dato Numérico (Tipo número).

Origin – Dato No Numérico (Tipo texto).

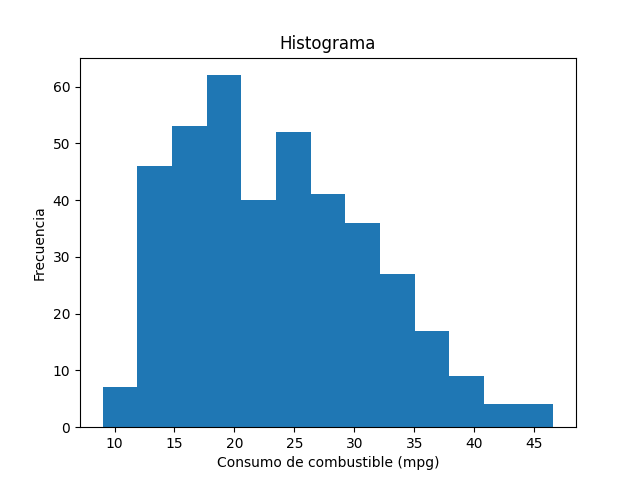
### Tipo de análisis:

El análisis realizado con esta base de datos es orientado a objetivos y está enfocado en comprender el consumo de combustible de los automóviles en la ciudad de New York. Se busca identificar patrones y relaciones entre las distintas variables para entender cómo influyen en el rendimiento del combustible de los vehículos. Además, se puede analizar la distribución de los valores de MPG para obtener información sobre la eficiencia general de los automóviles que circulan en la ciudad.

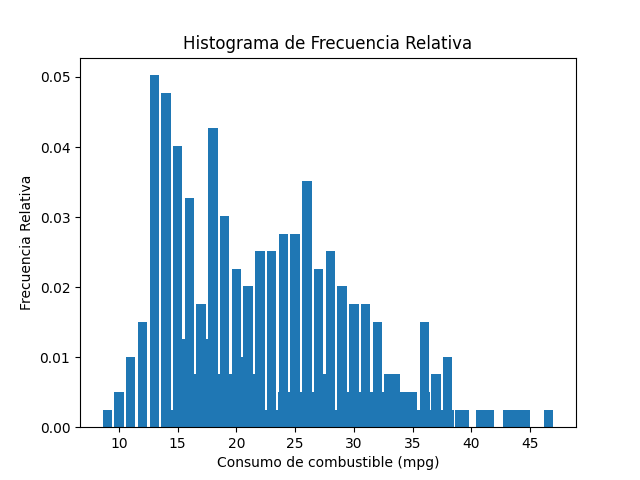
### Escogencia de la muestra:

La muestra fue escogida con la intención de obtener una representación significativa de los automóviles que se encuentran en New York y poder realizar un análisis adecuado sobre su consumo de combustible. Es importante contar con datos variados y representativos de diferentes marcas, modelos y características técnicas de los vehículos, para poder realizar conclusiones más sólidas sobre las tendencias y factores que afectan el consumo de combustible en la ciudad. Con esta información, se pueden tomar decisiones informadas para promover la adopción de automóviles más eficientes y reducir el impacto ambiental asociado con el consumo de combustible en New York.

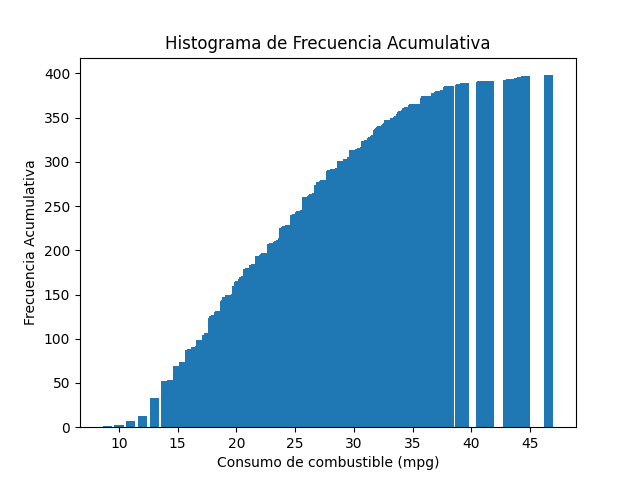
### Histogramas



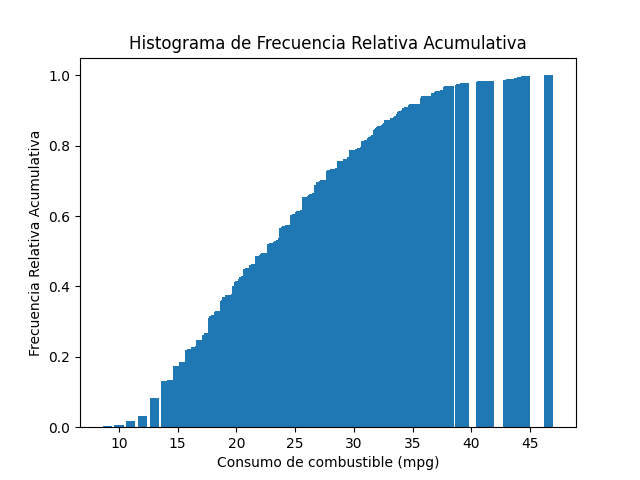
Este histograma muestra la distribución de las cantidades de millas recorridas por galón (MPG) en los automóviles de la muestra. En el eje horizontal se representan los distintos intervalos de MPG, mientras que en el eje vertical se muestra la cantidad de automóviles que caen dentro de cada intervalo. Este histograma permitirá visualizar la concentración de automóviles en diferentes rangos de eficiencia de combustible, lo que podría indicar si la muestra tiende a incluir más vehículos de alta o baja eficiencia en términos de consumo de combustible.



En este histograma, se muestra la distribución de la frecuencia relativa de las cantidades de millas recorridas por galón (MPG) en los automóviles de la muestra. La frecuencia relativa representa la proporción de automóviles que se encuentran en cada intervalo de MPG respecto al total de automóviles en la muestra. Este histograma permite visualizar las tendencias de eficiencia en el consumo de combustible de la muestra, mostrando qué intervalos de MPG son más frecuentes en comparación con otros.

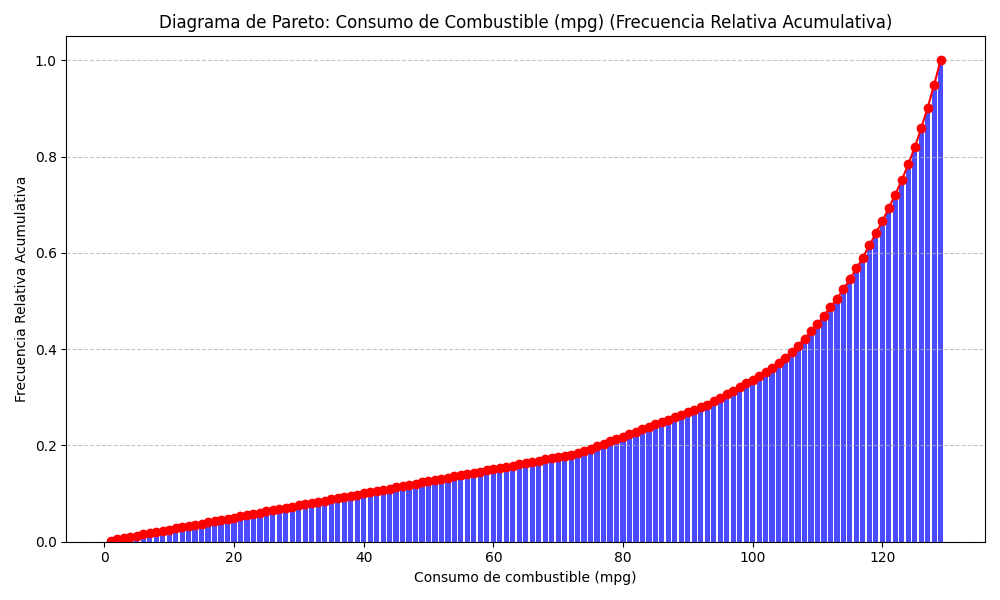


Este histograma muestra cómo se acumulan las cantidades de millas recorridas por galón (MPG) a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los distintos intervalos de MPG, mientras que en el eje vertical se muestra la cantidad acumulativa de automóviles hasta ese intervalo. Este histograma es útil para entender qué porcentaje del total de automóviles tienen un consumo de combustible igual o inferior a cierto valor de MPG, lo que ayuda a evaluar la eficiencia general de la muestra.



Este histograma muestra cómo se acumulan las frecuencias relativas de las cantidades de millas recorridas por galón (MPG) a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los distintos intervalos de MPG, mientras que en el eje vertical se muestra la frecuencia relativa acumulativa hasta ese intervalo. Este histograma es útil para entender qué porcentaje del total de automóviles tiene un consumo de combustible igual o inferior a cierto valor de MPG, brindando una visión general de la eficiencia promedio de la muestra y cómo se distribuyen los vehículos con respecto a sus consumos de combustible

### Diagrama de Pareto



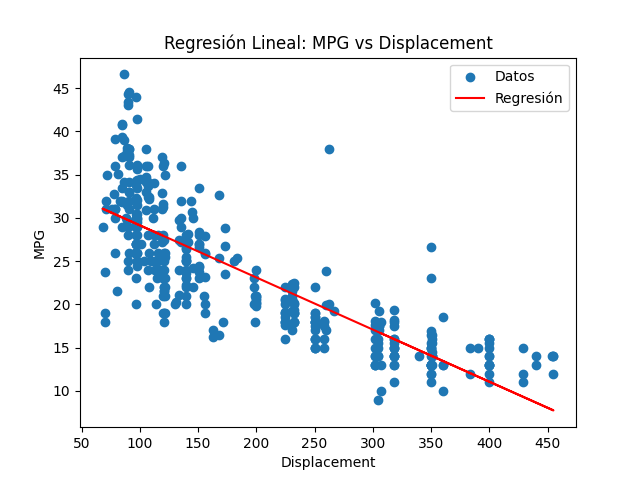
En este caso, se ha construido un diagrama de Pareto para el consumo de combustible de vehículos (medido en millas por galón, MPG) con una tendencia ascendente.

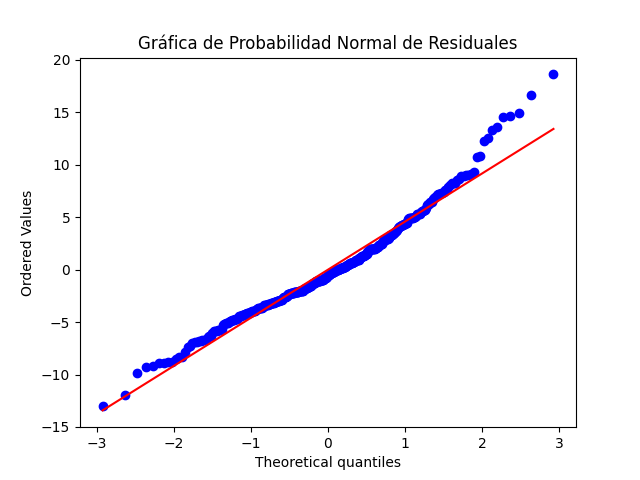
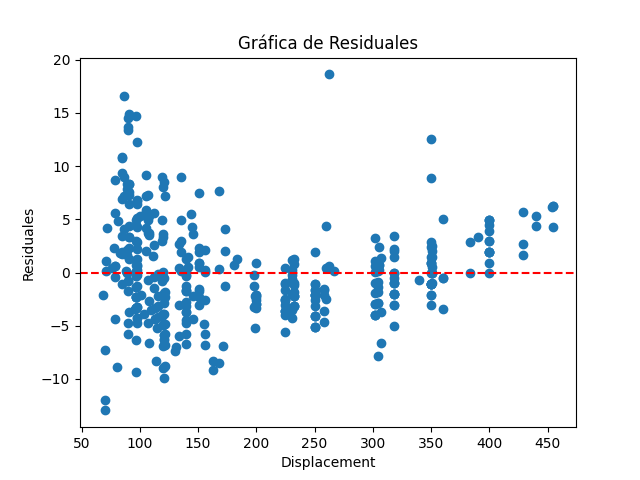
El eje vertical representa la frecuencia relativa acumulativa de los distintos valores de consumo de combustible (MPG) observados en una muestra o población de vehículos. La frecuencia relativa acumulativa indica la proporción de vehículos que tienen un consumo de combustible igual o menor que el valor en el eje horizontal. A medida que avanzamos hacia la derecha en el eje horizontal, los valores de MPG se ordenan de menor a mayor.

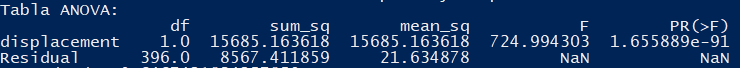
En este diagrama, se puede observar que la curva de frecuencia relativa acumulativa muestra una tendencia ascendente, lo que significa que hay una mayor concentración de vehículos con consumos de combustible más altos. Esto indica que la mayoría de los vehículos en la muestra están consumiendo más combustible por milla recorrida.

|  |  |
| --- | --- |
| Media: | 23.5145 |
| Varianza: | 61.08961 |
| Desviación Estándar: | 7.8156 |
| Coeficiente de correlación entre MPG y displacement: | -0.80420 |

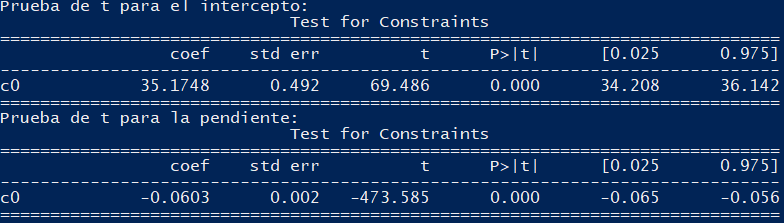
### Análisis de Regresión Lineal







R cuadrado: 0.6467421834257859





## CANTIDAD DE CALORIAS EN COMIDAS RÁPIDAS

### Descripción de las variables:

Restaurant – Dato No Numérico (Tipo texto).

Item – Dato No Numérico (Tipo texto).

Calories – Dato Numérico (Tipo número).

Cal\_fat – Dato Numérico (Tipo número).

Cholesterol – Dato Numérico (Tipo número).

Total\_carb – Dato Numérico (Tipo número).

Fiber – Dato Numérico (Tipo número).

Protein – Dato Numérico (Tipo número).

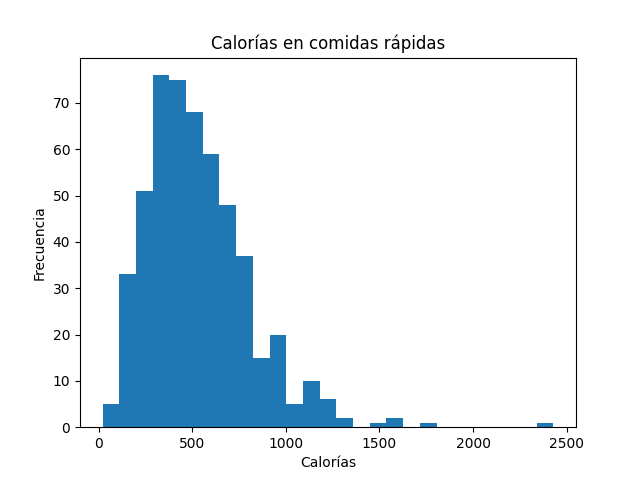
### Tipo de análisis:

El tipo de análisis realizado es orientado a objetivos, ya que el objetivo principal es determinar el contenido calórico y valores nutricionales de productos de comida rápida en diferentes restaurantes. Con este análisis, se busca identificar patrones y tendencias nutricionales en los alimentos ofrecidos por los distintos establecimientos, así como también evaluar la variabilidad entre los productos de diferentes restaurantes.

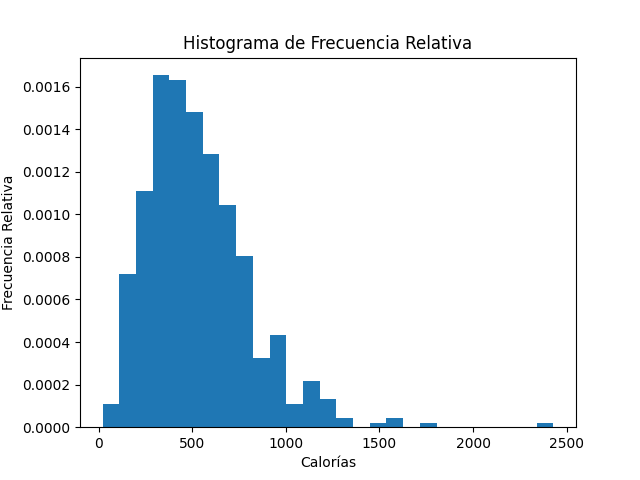
### Escogencia de la Muestra:

La muestra fue escogida para obtener una representación significativa de productos de comida rápida de distintos restaurantes, lo que permitirá obtener conclusiones más generales y relevantes sobre el contenido calórico y valores nutricionales en este tipo de comidas. Al abarcar diversos restaurantes y productos, el análisis puede proporcionar información valiosa para los consumidores, las autoridades sanitarias y los propios restaurantes, en términos de ofrecer opciones más saludables o comprender mejor el impacto nutricional de sus menús.

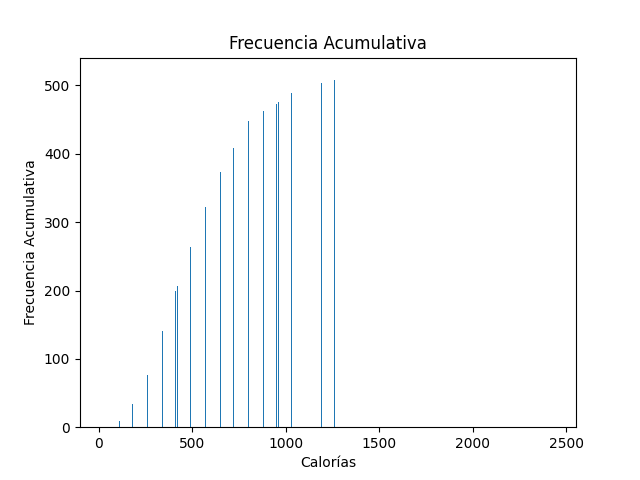
### Histogramas



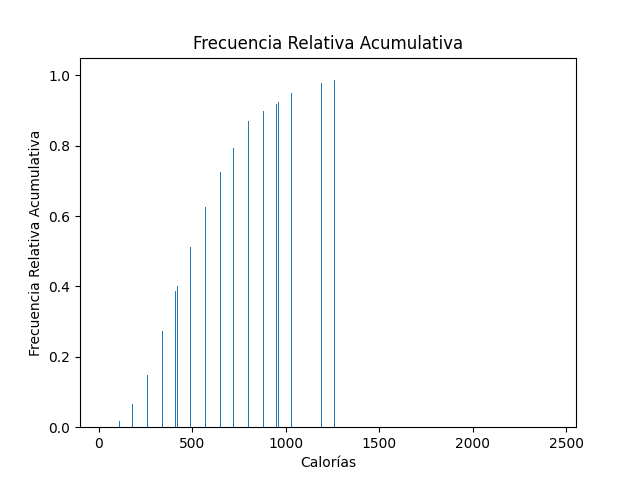
Este histograma muestra la distribución de la cantidad de calorías en los productos de comida rápida. En el eje horizontal se representan diferentes intervalos de calorías, mientras que en el eje vertical se muestra la cantidad de productos que caen dentro de cada intervalo. Este histograma permitirá visualizar la concentración de productos con cierta cantidad de calorías y ayudará a identificar patrones en la distribución calórica de los alimentos.



En este histograma se muestra cómo se acumulan las cantidades de calorías a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los distintos intervalos de calorías, mientras que en el eje vertical se muestra la cantidad acumulativa de productos hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de productos tiene una cantidad de calorías igual o inferior a cierto valor, lo que proporciona una visión general del contenido calórico promedio y la variabilidad en la muestra.

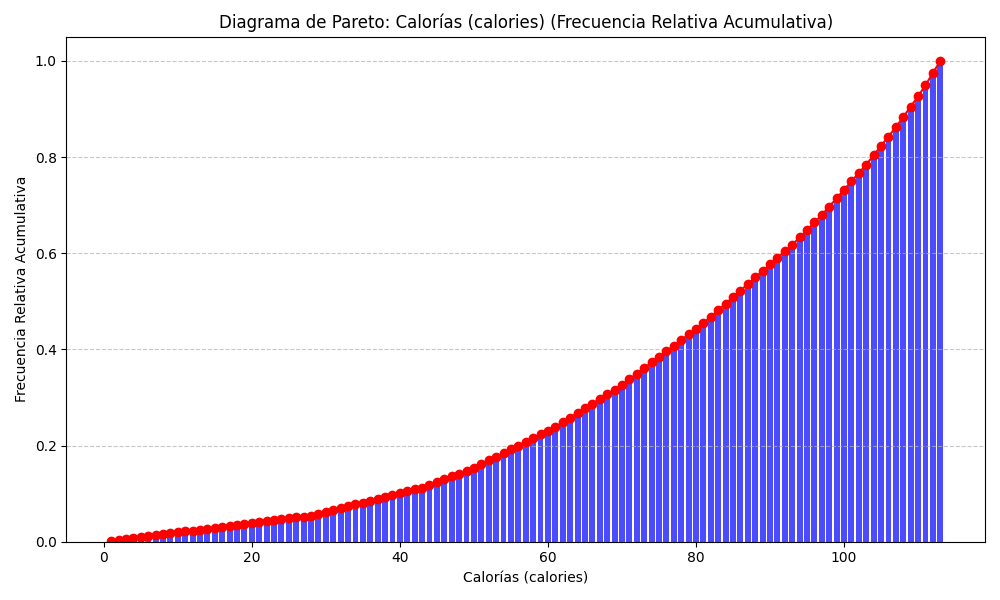


En este histograma se muestra la distribución de la frecuencia relativa de las cantidades de calorías en los productos de comida rápida. La frecuencia relativa representa la proporción de productos que se encuentran en cada intervalo de calorías respecto al total de productos en la muestra. Este histograma permitirá visualizar las tendencias en la cantidad de calorías que prevalecen en los productos y cuáles son más frecuentes en comparación con otros.



En este histograma se muestra cómo se acumulan las frecuencias relativas de las cantidades de calorías a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los distintos intervalos de calorías, mientras que en el eje vertical se muestra la frecuencia relativa acumulativa hasta ese intervalo.

### Diagrama de Pareto

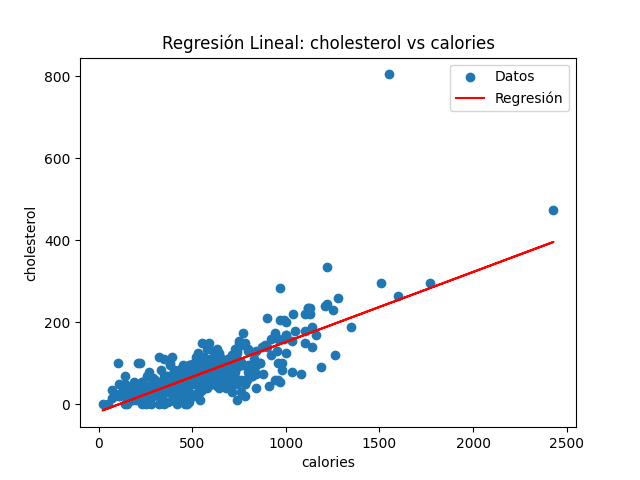


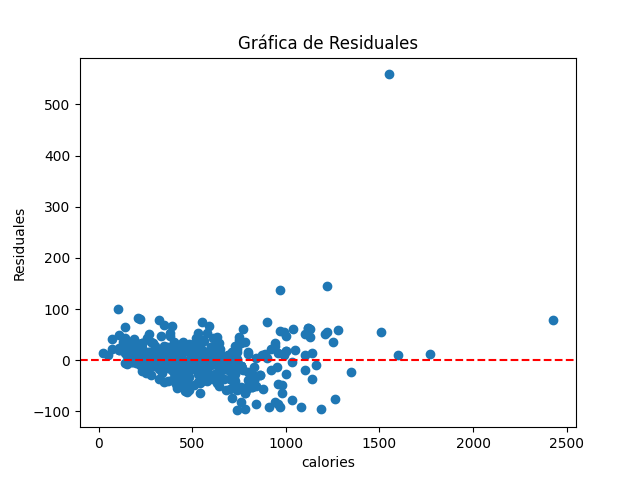
En este diagrama, la tendencia ascendente indica que un grupo reducido de productos de comida rápida tiene una cantidad significativamente alta de calorías en comparación con la mayoría de los productos. Los primeros elementos en el diagrama representan los productos con más calorías y, a medida que avanzamos hacia la derecha, las cantidades de calorías disminuyen gradualmente.

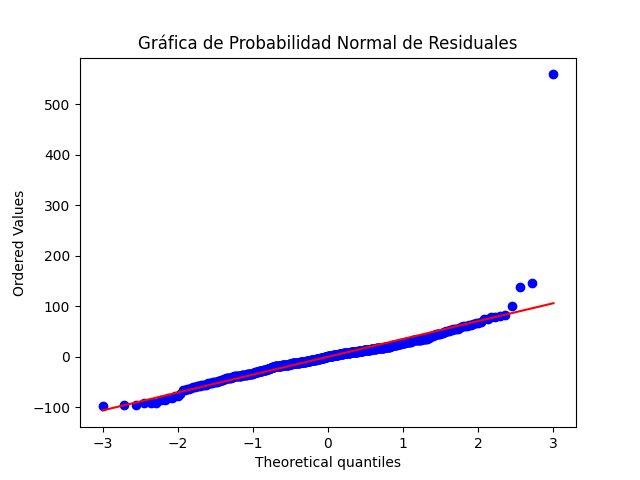
Este tipo de diagrama permite identificar rápidamente los productos que contribuyen más significativamente al contenido calórico total en la muestra, lo que podría ser útil para tomar decisiones informadas sobre opciones más saludables o para evaluar la oferta nutricional de los productos ofrecidos por los restaurantes.

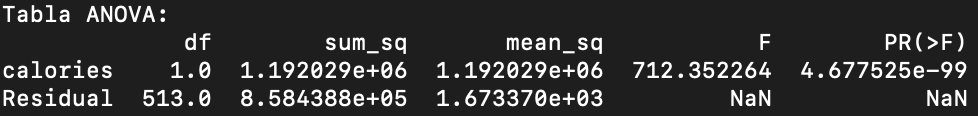
|  |  |
| --- | --- |
| Media: | 530.9126 |
| Varianza: | 79770.1771 |
| Desviación Estándar: | 282.4361 |
| Coeficiente de correlación entre calories y colesterol: | 0.76246 |

### Análisis de Regresión Lineal

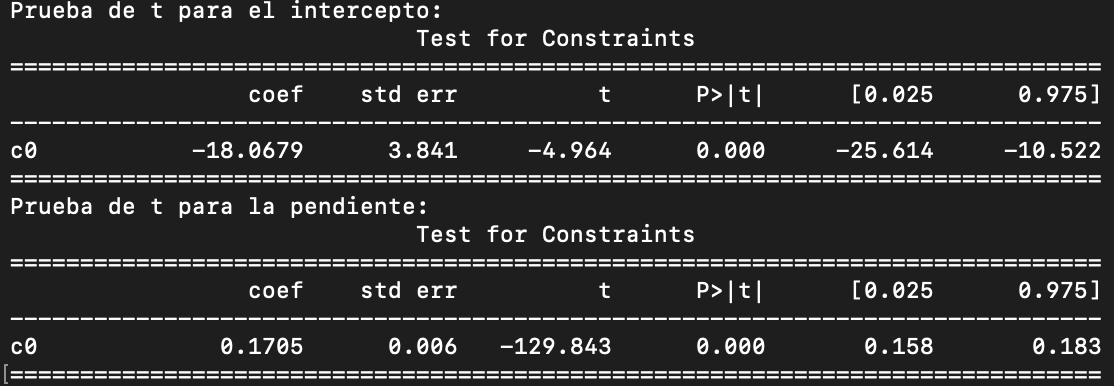


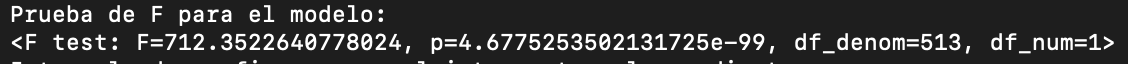






R cuadrado: 0.58134487931429





## PRECIOS DE LAPTOPS

### Descripción de variables

Laptop – Dato No Numérico (Tipo Texto).

Status – Dato No Numérico (Tipo Texto).

Brand – Dato No Numérico (Tipo Texto).

Model – Dato No Numérico (Tipo Texto).

CPU – Dato No Numérico (Tipo Texto).

RAM– Dato Numérico (Tipo número).

Storage – Dato Numérico (Tipo número).

GPU – Dato Numérico (Tipo número).

FinalPrice – Dato Numérico (Tipo número).

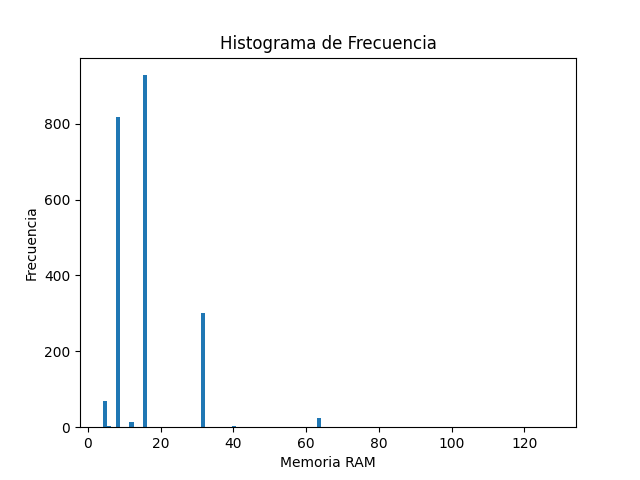
### Tipo de análisis:

El tipo de análisis realizado es orientado a objetivos, ya que el objetivo principal es evaluar y comparar los precios de las laptops en función de sus características técnicas. Se busca identificar patrones y tendencias en cómo las especificaciones de las laptops (RAM, almacenamiento, GPU, etc.) influyen en su precio final.

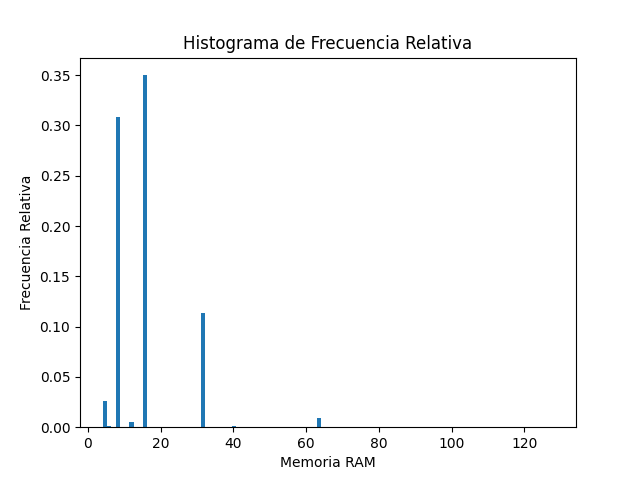
### Escogencia de la muestra:

La muestra fue escogida para representar una variedad de laptops de diferentes marcas y especificaciones técnicas. Al incluir diversos modelos y características, el análisis permitirá entender cómo factores como el rendimiento del procesador, la cantidad de RAM y el tipo de GPU afectan los precios. Esto podría ser útil tanto para los consumidores, al tomar decisiones informadas sobre qué laptop adquirir según sus necesidades y presupuesto, como para las empresas de tecnología, al comprender qué características son más valoradas por el mercado y cómo ajustar sus estrategias de precios.

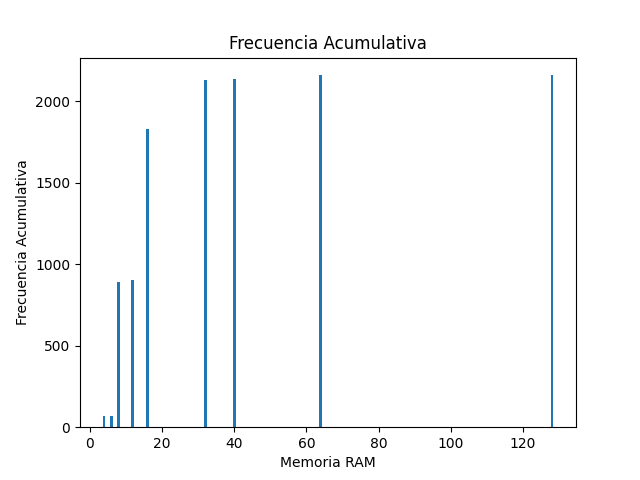
### Histogramas



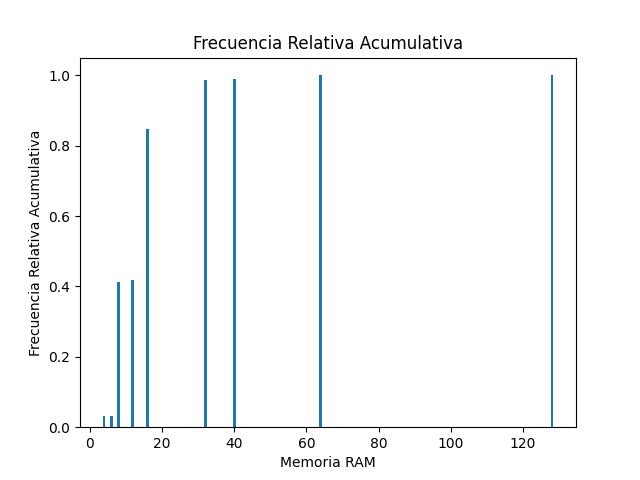
Este histograma muestra la distribución de los precios de las laptops en diferentes intervalos. En el eje horizontal se representan los intervalos de precios, mientras que en el eje vertical se muestra la cantidad de laptops que se encuentran en cada intervalo. Este histograma permitirá visualizar cómo se distribuyen los precios de las laptops en la muestra y ayudará a identificar los rangos de precios más comunes.



En este histograma se muestra la distribución de la frecuencia relativa de los precios de las laptops. La frecuencia relativa representa la proporción de laptops que se encuentran en cada intervalo de precios respecto al total de laptops en la muestra. Este histograma permitirá visualizar las tendencias en los precios de las laptops y cuáles son más frecuentes en comparación con otros rangos de precios.

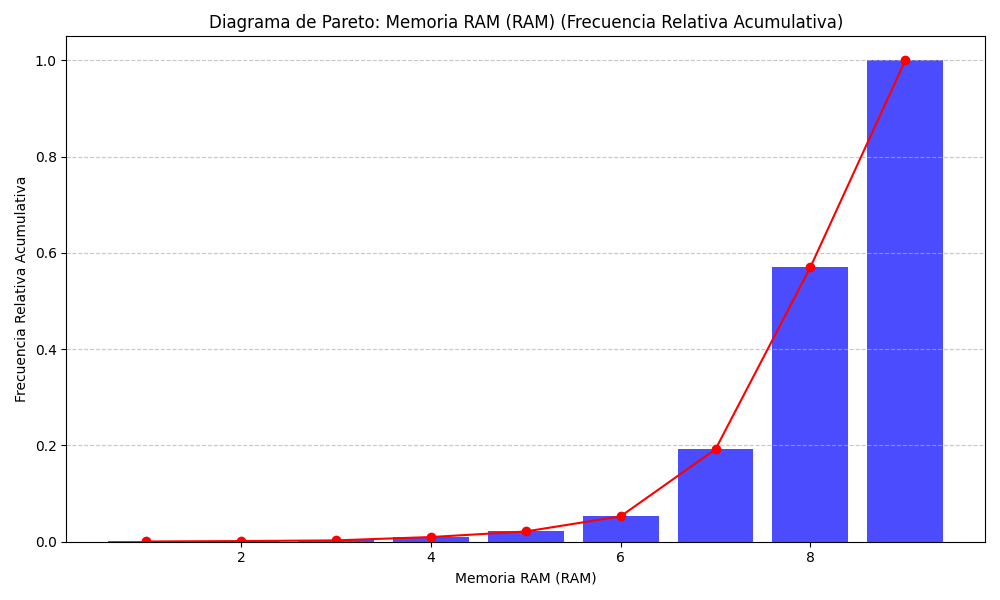


En este histograma se muestra cómo se acumulan los precios de las laptops a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de precios, mientras que en el eje vertical se muestra la cantidad acumulativa de laptops hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de laptops tiene un precio igual o inferior a cierto valor, brindando una visión general de la distribución de precios y su variabilidad.



En este histograma se muestra cómo se acumulan las frecuencias relativas de los precios de las laptops a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de precios, mientras que en el eje vertical se muestra la frecuencia relativa acumulativa hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de laptops tiene un precio igual o inferior a cierto valor, brindando una visión general de la distribución de precios y cómo se concentran los productos en diferentes rangos de precios.

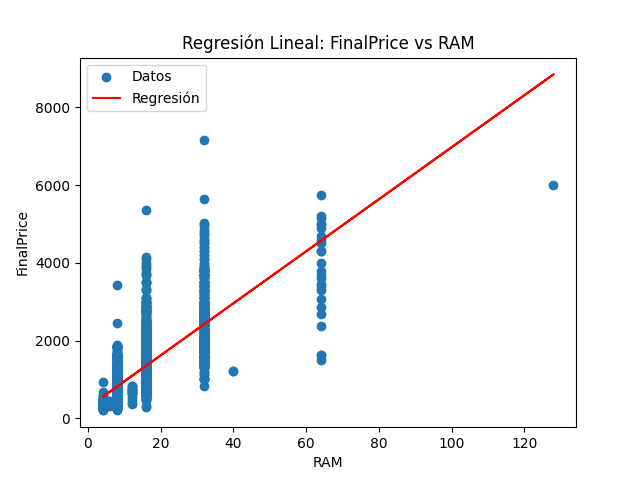
### Diagrama de Pareto

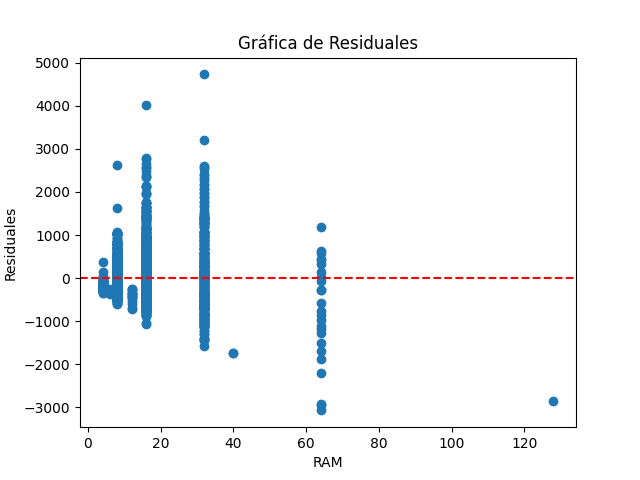


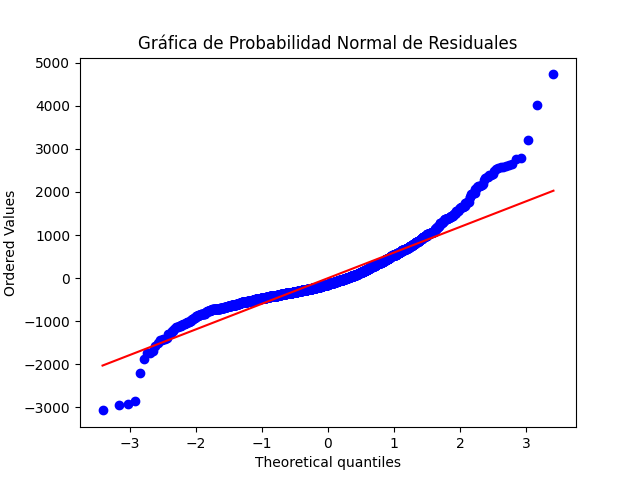
En este diagrama, la tendencia ascendente indica que un grupo reducido de laptops tiene una mayor capacidad de memoria RAM en comparación con la mayoría de las laptops. Los primeros elementos en el diagrama representan las laptops con más memoria RAM y, a medida que avanzamos hacia la derecha, la capacidad de memoria RAM disminuye gradualmente.

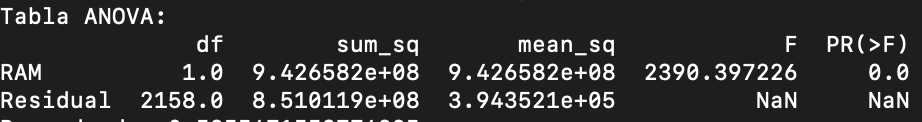
|  |  |
| --- | --- |
| Media: | 15.41388888888889 |
| Varianza: | 97.37377643970974 |
| Desviación Estándar: | 9.86781518066232 |
| Coeficiente de correlación entre RAM y FinalPrice: | 0.7249463124103899 |

### Análisis de Regresión Lineal

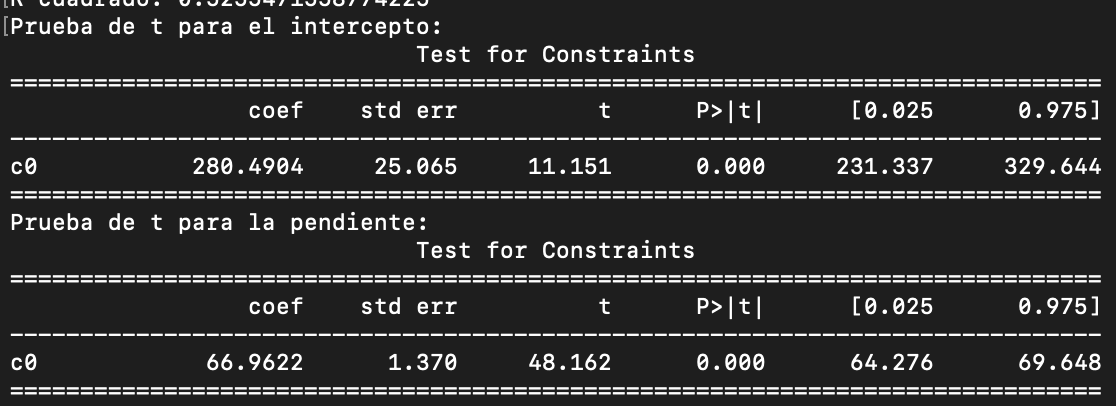


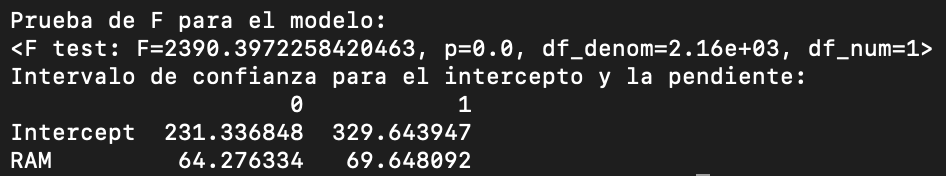






R cuadrado: 0.5255471558774225





## IMPORTACIONES VALOR CIF Y PESO

### Descripción de las variables

Arancel – Dato Numérico (Tipo número).

Año – Dato Numérico (Tipo número).

MES – Dato Numérico (Tipo número).

ValorCIF – Dato Numérico (Tipo número).

PesoNeto – Dato Numérico (Tipo número).

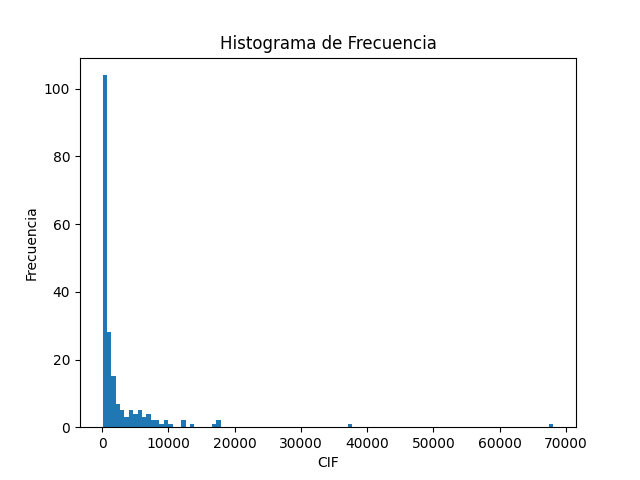
### Tipo de análisis:

orientado a objetivos y estar enfocado en evaluar el comportamiento y tendencias de las importaciones a lo largo de los años, identificar los productos con mayores aranceles, comparar la relación entre el valor CIF y el peso neto, y estudiar la evolución del comercio internacional en términos de cantidad y valor de los bienes importados.

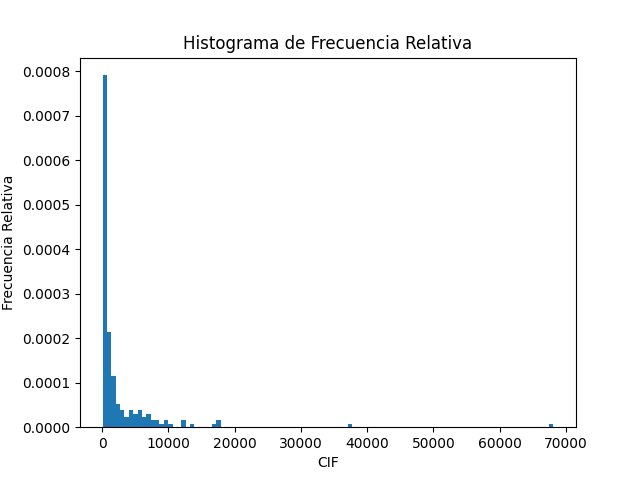
### Escogencia de la muestra:

Fue recopilada con el objetivo de analizar la situación económica y comercial de un país o región específica, para evaluar las políticas arancelarias y su impacto en el comercio internacional, y para identificar oportunidades de mejora en la eficiencia logística y en la optimización de costos en las operaciones de importación.

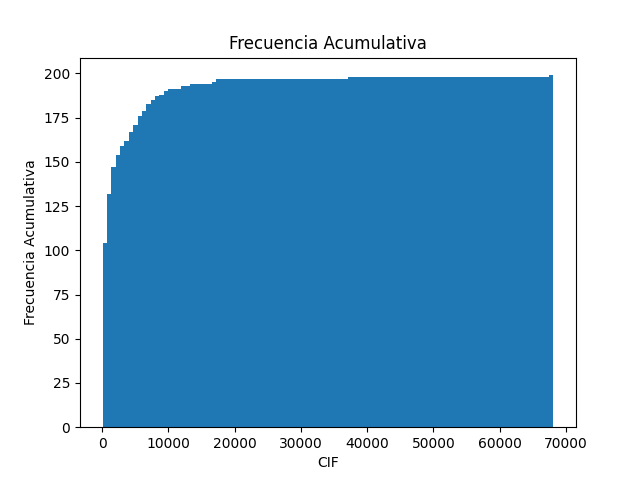
### Histogramas



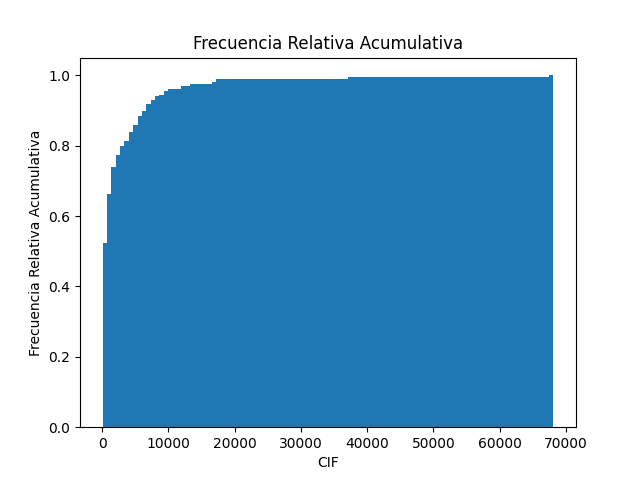
Este histograma muestra la distribución de la cantidad de importaciones en diferentes intervalos. En el eje horizontal se representan los intervalos de importaciones, mientras que en el eje vertical se muestra la cantidad de veces que ocurren importaciones en cada intervalo. Este histograma permitirá visualizar cómo se distribuyen las importaciones en la muestra y ayudará a identificar los rangos de cantidad de importaciones más comunes.



En este histograma se muestra la distribución de la frecuencia relativa de las importaciones. La frecuencia relativa representa la proporción de importaciones que se encuentran en cada intervalo respecto al total de importaciones en la muestra. Este histograma permitirá visualizar las tendencias en la cantidad de importaciones que prevalecen en comparación con otros rangos.

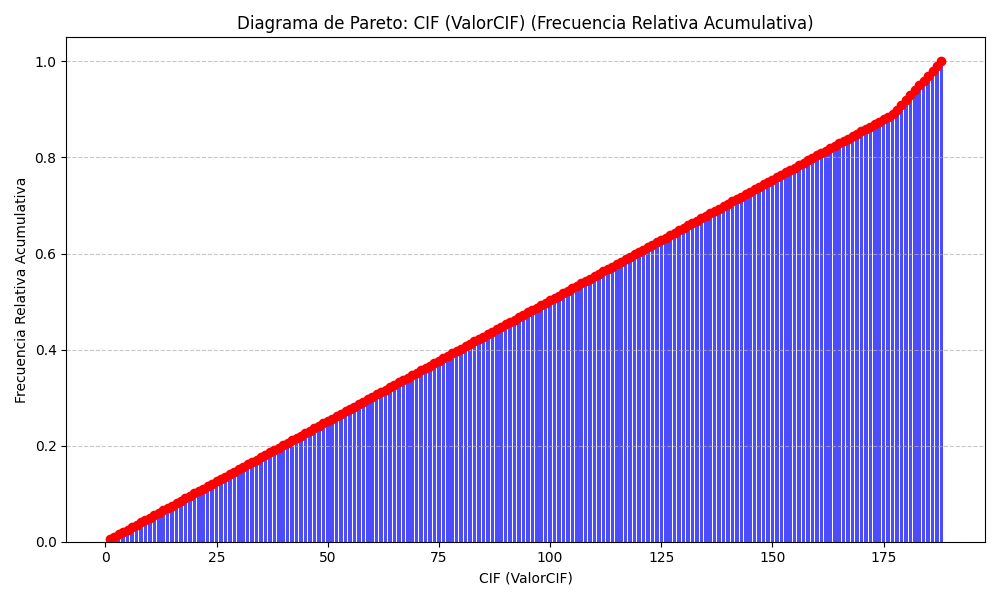


En este histograma se muestra cómo se acumulan las importaciones a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de importaciones, mientras que en el eje vertical se muestra la cantidad acumulativa de importaciones hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de importaciones corresponde a ciertas cantidades, brindando una visión general de la distribución de importaciones y su variabilidad.



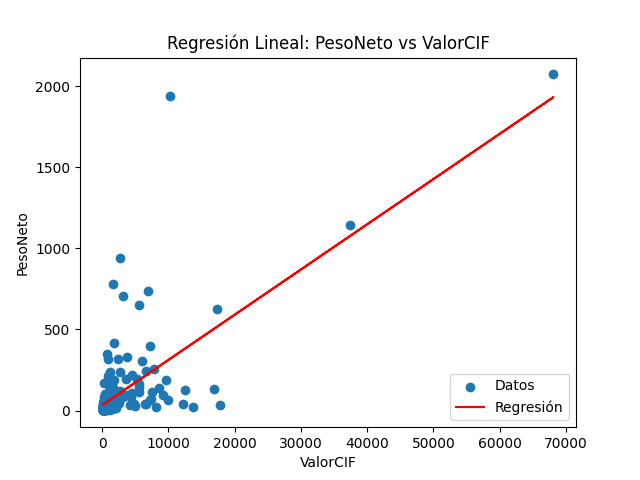
Este histograma permitirá entender qué porcentaje del total de importaciones corresponde a ciertas cantidades, brindando una visión general de la distribución de importaciones y cómo se concentran los valores de importaciones en diferentes rangos.

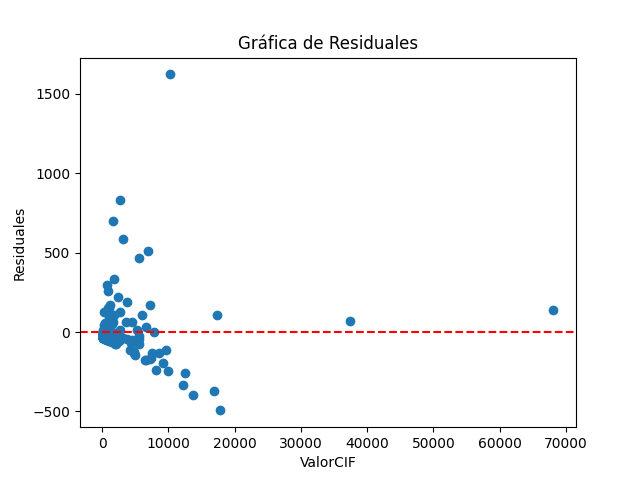
### Diagrama de Pareto

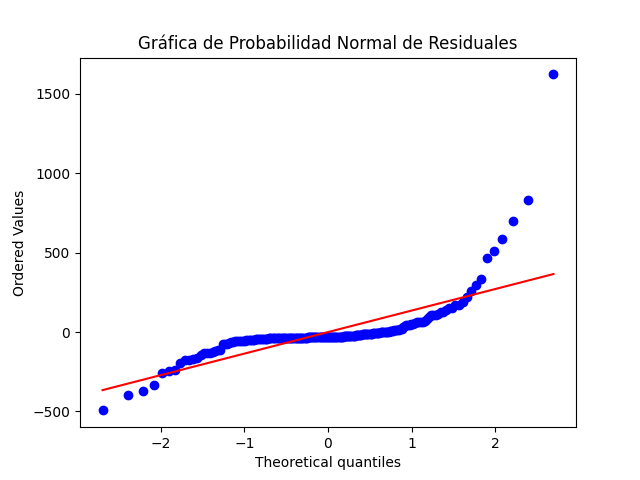


|  |  |
| --- | --- |
| Media: | 2539.758793969849 |
| Varianza: | 38736020.81021268 |
| Desviación Estándar: | 6223.826862165486 |
| Coeficiente de correlación entre ValorCIF y PesoNeto: | 0.6909290785344882 |

### Análisis de Regresión Lineal

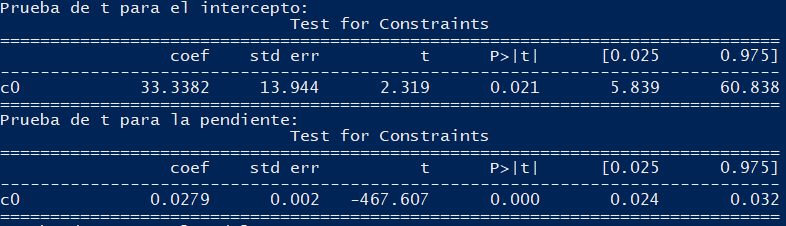


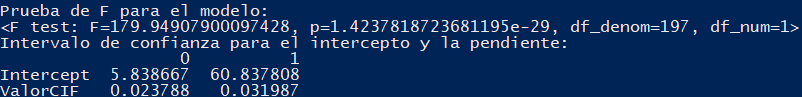






R cuadrado: 0.4773829915645167





## TAMAÑO Y PESO DE PECES

### Descripción de variables

Species – Dato No Numérico (Tipo Texto).

Weight: – Dato Numérico (Tipo número).

Height – Dato Numérico (Tipo número).

Width – Dato Numérico (Tipo número).

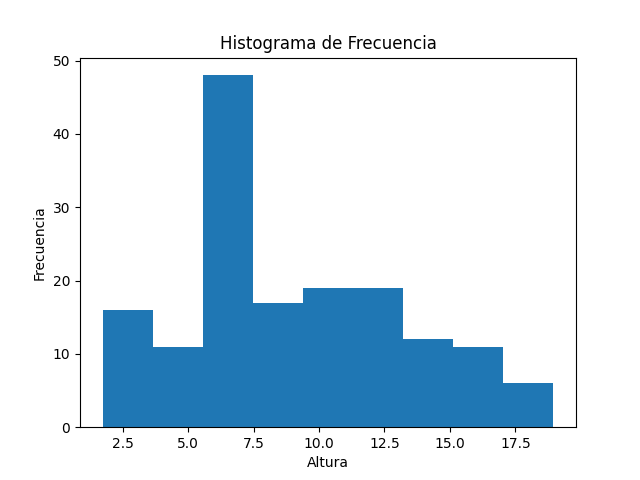
### Tipo de análisis:

El objetivo del análisis es el de explorar y entender cómo varía el peso, altura y anchura de diferentes especies de peces. Se busca identificar patrones y relaciones entre estas variables para obtener una mejor comprensión de las características físicas de cada especie.

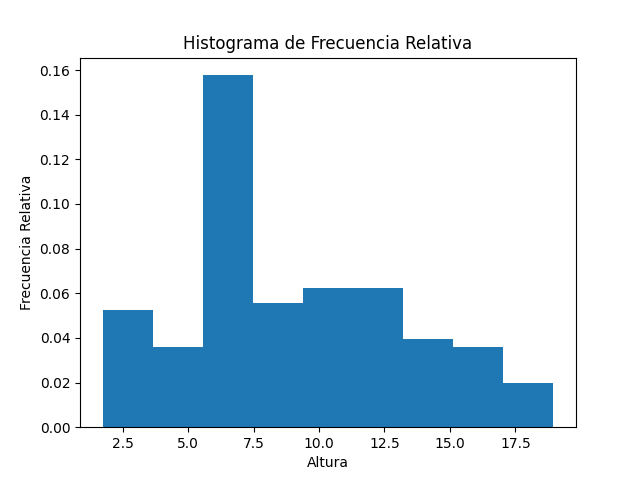
### Escogencia de la muestra:

Ha sido recopilada con el propósito de estudiar la biodiversidad de los peces y cómo diferentes factores, como la especie, pueden influir en su tamaño y peso. También podría ser útil para la investigación en acuicultura y conservación de especies, así como para ayudar a comprender las relaciones ecológicas entre diferentes especies de peces y su papel en los ecosistemas acuáticos.

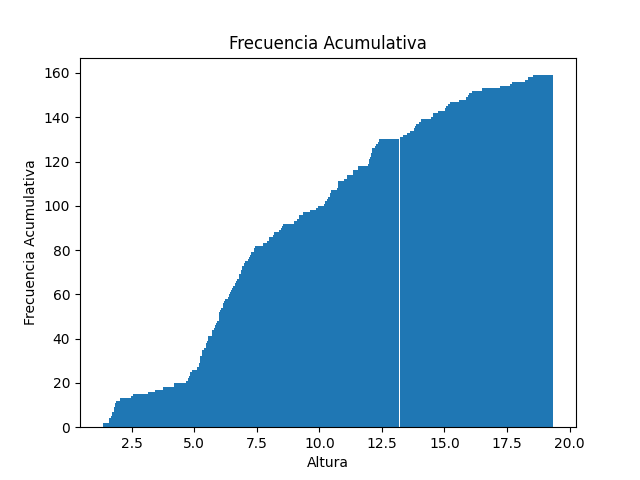
### Histogramas



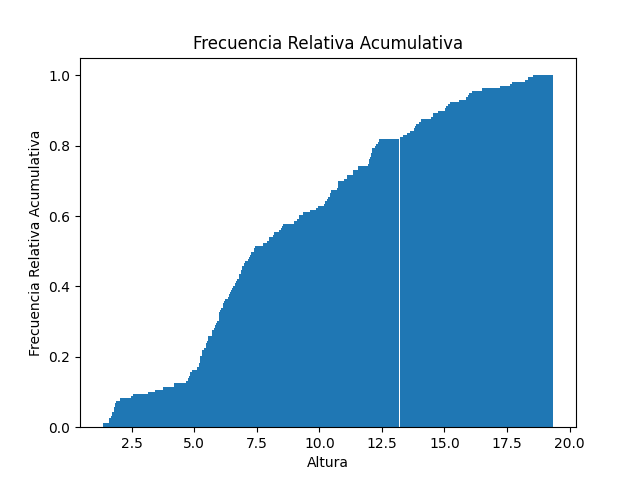
Este histograma muestra la distribución del peso de los peces en diferentes intervalos. En el eje horizontal se representan los intervalos de peso, mientras que en el eje vertical se muestra la cantidad de peces que caen dentro de cada intervalo. Este histograma permitirá visualizar cómo se distribuyen los pesos de los peces en la muestra y ayudará a identificar los rangos de peso más comunes.



En este histograma se muestra la distribución de la frecuencia relativa de los pesos de los peces. La frecuencia relativa representa la proporción de peces que se encuentran en cada intervalo de peso respecto al total de peces en la muestra. Este histograma permitirá visualizar las tendencias en los pesos de los peces que prevalecen en comparación con otros.

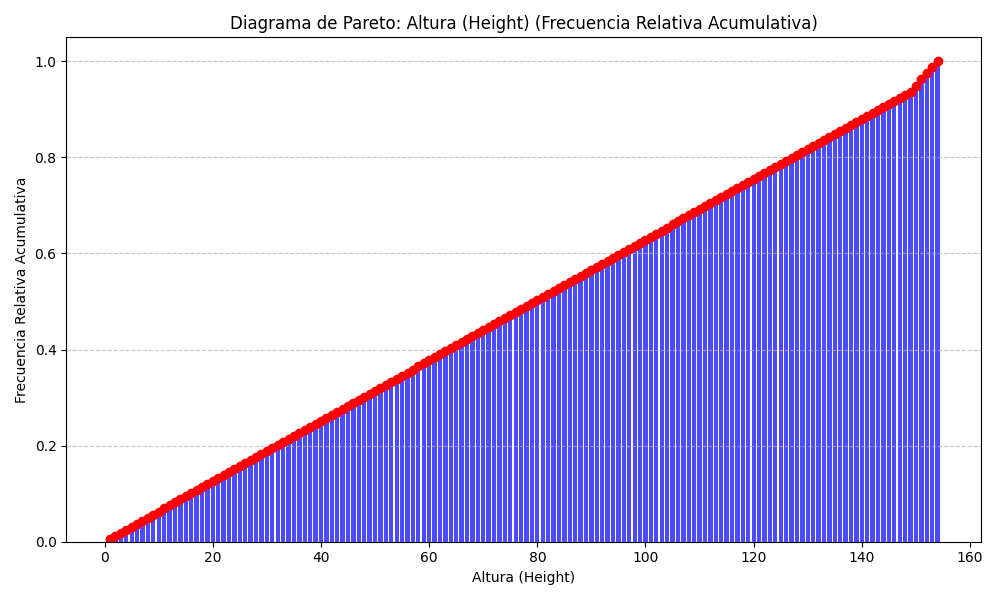


En este histograma se muestra cómo se acumulan los pesos de los peces a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de peso, mientras que en el eje vertical se muestra la cantidad acumulativa de peces hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de peces tiene un peso igual o inferior a cierto valor, brindando una visión general de la distribución de pesos y su variabilidad.



Este histograma permitirá entender qué porcentaje del total de peces tiene un peso igual o inferior a cierto valor, brindando una visión general de la distribución de pesos y cómo se concentran los valores de peso en diferentes rangos.

### Diagrama de Pareto

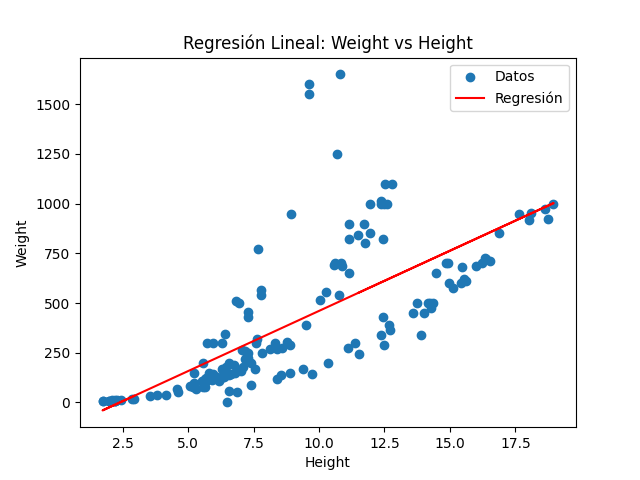


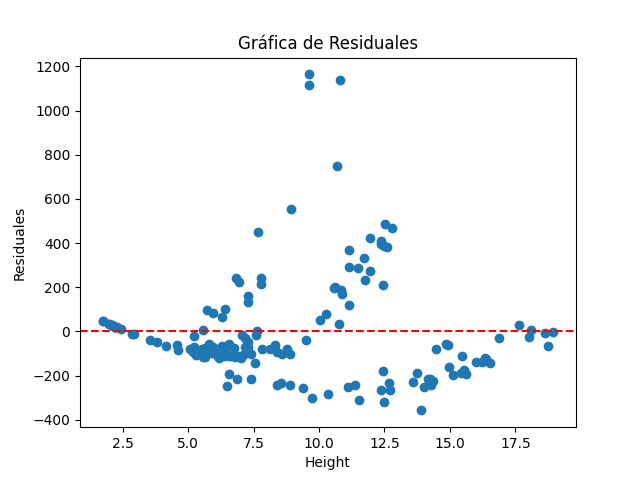
En este diagrama, la tendencia ascendente indica que un grupo reducido de peces tiene una altura significativamente mayor en comparación con la mayoría de los peces. Los primeros elementos en el diagrama representan los peces más altos y, a medida que avanzamos hacia la derecha, las alturas de los peces disminuyen gradualmente.

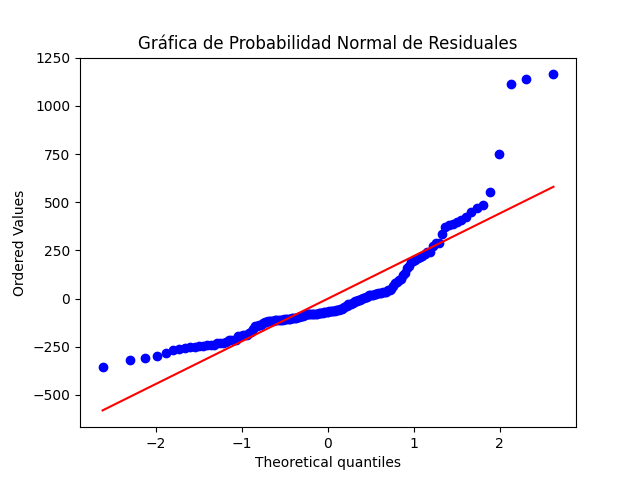
Este tipo de diagrama permite identificar rápidamente los peces que tienen una mayor altura, lo que puede ser relevante para aquellos que buscan especímenes más grandes o que estudian el crecimiento y desarrollo de ciertas especies.

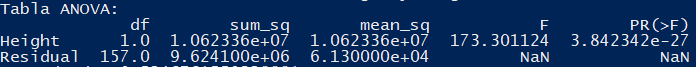
|  |  |
| --- | --- |
| Media: | 8.970993710691824 |
| Varianza: | 18.371575761479182 |
| Desviación Estándar: | 9.86781518066232 |
| Coeficiente de correlación entre Height y Weight: | 4.2862076199688675 |

### Análisis de Regresión Lineal

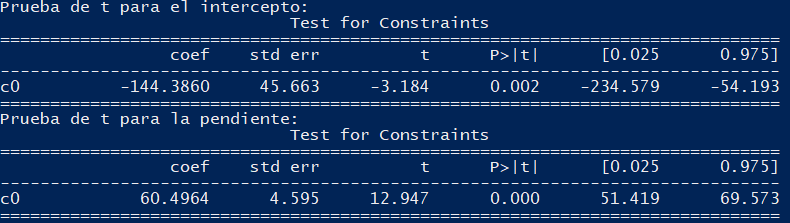


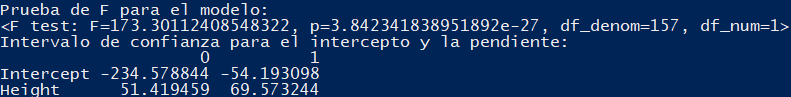






R cuadrado: 0.5246761559328881





## PIB\_PROVINCIA

### Descripción de las variables

Código Provincia – Dato Numérico (Tipo Número).

Provincia – Dato No Numérico (Tipo Texto).

Año – Dato No Numérico (Tipo Texto).

Composición Constante – Dato Numérico (Tipo Número).

Composición Corriente – Dato Numérico (Tipo Número).

Per Cápita Constante – Dato Numérico (Tipo Número).

Per Cápita Corriente – Dato Numérico (Tipo Número).

Valor Constante – Dato Numérico (Tipo Número).

Valor Corriente – Dato Numérico (Tipo Número).

Valores – Dato Numérico (Tipo Número).

Variación Porcentual Constante – Dato Numérico (Tipo Número).

Variación Porcentual Corriente – Dato Numérico (Tipo Número).

### Tipo de análisis:

El tipo de análisis realizado con esta base de datos fue orientado a objetivos. Se busca comprender y evaluar el comportamiento económico y el crecimiento del PIB en cada provincia a lo largo de los años.

### Escogencia de Muestra:

La muestra de datos se escogió con el propósito de estudiar y analizar el comportamiento económico de diferentes provincias en un país o región específica. Los datos del PIB son fundamentales para entender el desempeño económico, formular políticas y tomar decisiones estratégicas en áreas como desarrollo regional, inversión y bienestar económico de la población.

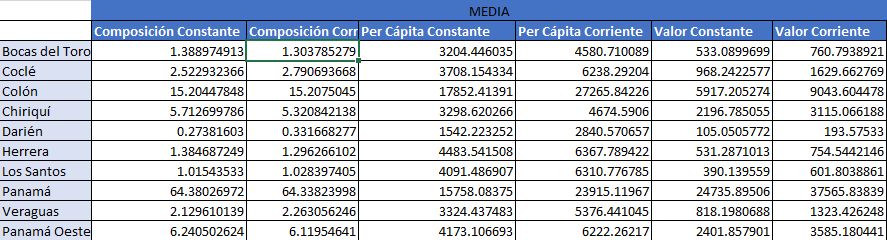
### Histogramas

Este histograma muestra la distribución del PIB per cápita constante en diferentes intervalos. En el eje horizontal se representan los intervalos de PIB per cápita constante, mientras que en el eje vertical se muestra la cantidad de provincias que tienen un PIB per cápita constante dentro de cada intervalo. Este histograma permitirá visualizar cómo se distribuyen los niveles de PIB per cápita constante en las provincias y ayudará a identificar los grupos de provincias con niveles similares de ingreso ajustado por inflación.

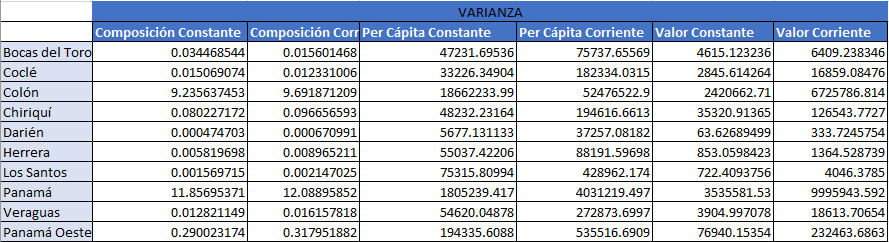
En este histograma se muestra cómo se acumulan los valores del PIB constante a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de PIB constante, mientras que en el eje vertical se muestra la cantidad acumulativa de provincias hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de provincias tiene un PIB constante igual o inferior a cierto valor, brindando una visión general de la distribución de ingresos ajustados por inflación y su variabilidad.

En este histograma se muestra la distribución de la frecuencia relativa del PIB corriente de las provincias. La frecuencia relativa representa la proporción de provincias que tienen un PIB corriente dentro de cada intervalo de valores respecto al total de provincias en la muestra. Este histograma permitirá visualizar las tendencias en los niveles de PIB corriente que prevalecen en comparación con otros.

Este histograma permitirá entender qué porcentaje del total de provincias tiene un PIB corriente igual o inferior a cierto valor, brindando una visión general de la distribución de ingresos sin ajustar por inflación y cómo se concentran los valores de PIB en diferentes rangos.



*ILUSTACION I. Media de las muestras*



*ILUSTACION II. Varianza de las muestras*



*ILUSTACION III. Desviación estándar de las muestras*

## STUDENT PERFORMANCE

### Descripción de las variables

Hours Studied – Dato Numérico (Tipo Número).

Previous Scores – Dato Numérico (Tipo Número).

Extracurricular Activities – Dato No Numérico (Tipo Texto).

Sleep Hours – Dato Numérico (Tipo Número).

Sample Question – Dato Numérico (Tipo Número).

Papers Practiced – Dato Numérico (Tipo Número).

Performance Index – Dato Numérico (Tipo Número).

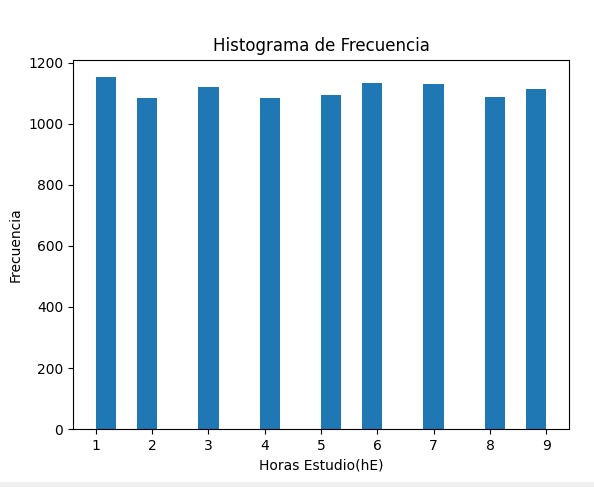
### Tipo de análisis:

El objetivo del análisis es el de identificar qué factores están asociados con un mejor rendimiento académico. Se busca entender cómo las horas de estudio, el rendimiento previo, la participación en actividades extracurriculares, el sueño y la práctica de ejercicios pueden influir en el rendimiento académico general.

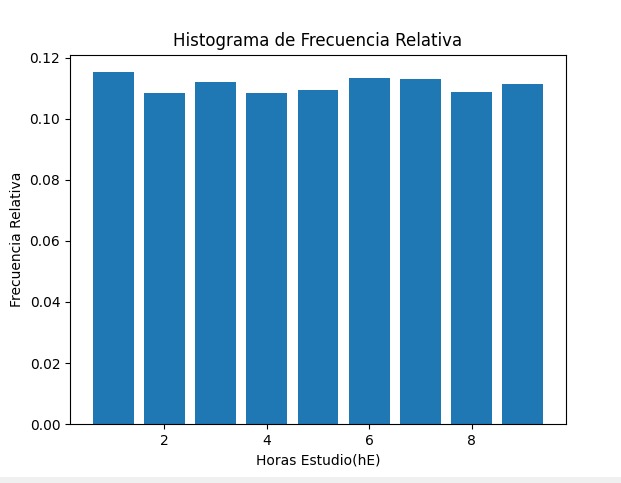
### Escogencia de la muestra:

Esta muestra de datos ha sido recopilada con el propósito de evaluar el desempeño académico de estudiantes, identificar posibles áreas de mejora y brindar información para la toma de decisiones educativas. Los datos sobre el tiempo de estudio, las horas de sueño y el rendimiento previo podrían ser utilizados por instituciones educativas para implementar estrategias que promuevan un mejor rendimiento estudiantil y para comprender cómo ciertas variables pueden afectar el éxito académico de los estudiantes.

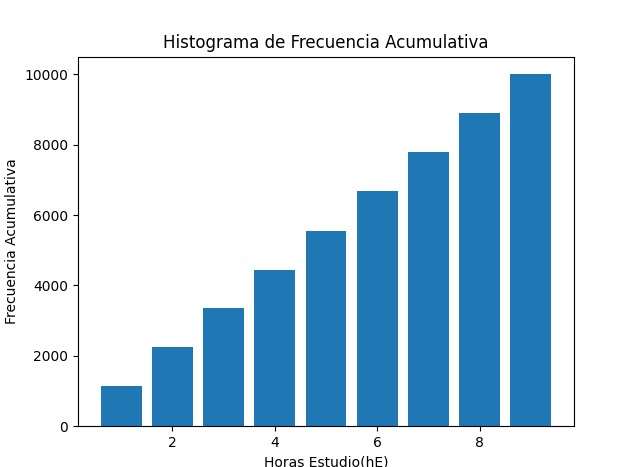
### Histogramas



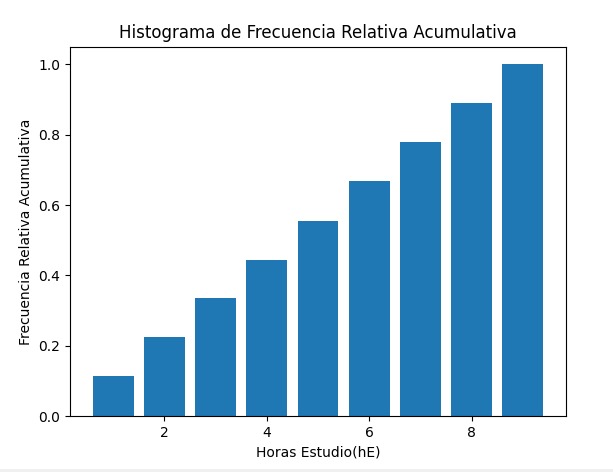
Este histograma muestra la distribución de la cantidad de estudiantes que han estudiado cierta cantidad de horas. En el eje horizontal se representan los intervalos de horas estudiadas, mientras que en el eje vertical se muestra la cantidad de estudiantes que han estudiado en cada intervalo. Este histograma permitirá visualizar cómo se distribuyen las horas de estudio entre los estudiantes y ayudará a identificar los intervalos de estudio más comunes.



En este histograma se muestra la distribución de la frecuencia relativa de las horas de estudio de los estudiantes. La frecuencia relativa representa la proporción de estudiantes que han estudiado cierta cantidad de horas respecto al total de estudiantes en la muestra. Este histograma permitirá visualizar las tendencias en la distribución de tiempo de estudio y cómo se concentran las horas de estudio en diferentes intervalos.

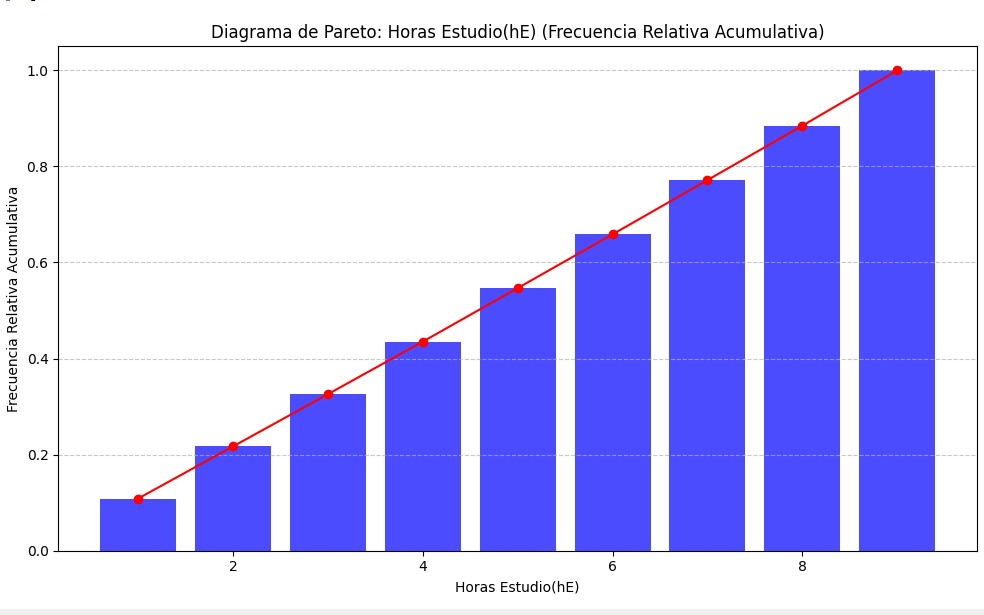


En este histograma se muestra cómo se acumulan las horas de estudio a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de horas estudiadas, mientras que en el eje vertical se muestra la cantidad acumulativa de estudiantes que han estudiado hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de estudiantes ha estudiado cierta cantidad de horas o menos, brindando una visión general de la distribución de tiempo de estudio y su variabilidad.



Este histograma permitirá entender qué porcentaje del total de estudiantes ha estudiado cierta cantidad de horas o menos, brindando una visión general de la distribución de tiempo de estudio y cómo se concentran los valores de horas de estudio en diferentes rangos.

### Diagrama de Pareto



En este diagrama, la tendencia ascendente indica que un grupo reducido de estudiantes ha dedicado más horas al estudio en comparación con la mayoría de los estudiantes. Los primeros elementos en el diagrama representan los estudiantes que han estudiado más horas y, a medida que avanzamos hacia la derecha, el número de estudiantes que han estudiado cierta cantidad de horas disminuye gradualmente.

Este tipo de diagrama permite identificar rápidamente los estudiantes que han dedicado más tiempo al estudio, lo que puede ser relevante para aquellos que buscan mejorar su rendimiento académico o para identificar ejemplos a seguir en términos de hábitos de estudio.

## MALLS

### Descripción de las variables

Country – Dato No Numérico (Tipo Texto).

City (metropolitan area) – Dato No Numérico (Tipo Texto).

Year opened – Dato No Numérico (Tipo Texto).

Gross leasable area (GLA) – Dato Numérico (Tipo Número).

Shops – Dato Numérico (Tipo Número).

### Tipo de análisis:

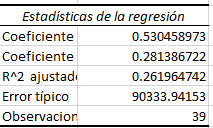
El objetivo del análisis es el de estudiar la evolución y características de los centros comerciales en diferentes países y ciudades. Se busca entender cómo el tamaño (GLA) y el número de tiendas (Shops) varían a lo largo del tiempo y en diferentes ubicaciones geográficas.

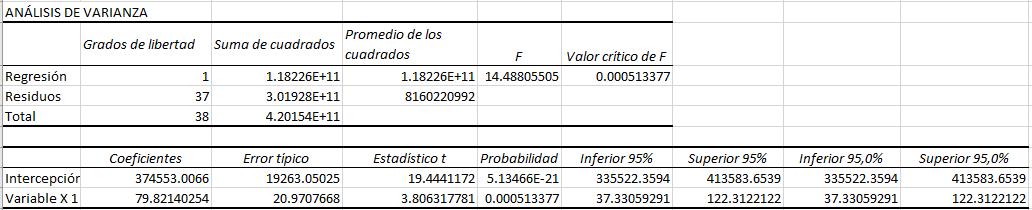
### Escogencia de la muestra:

Esta muestra de datos ha sido recopilada con el propósito de analizar el desarrollo y crecimiento de la industria de centros comerciales en diferentes regiones, así como para identificar las tendencias en términos de tamaño y oferta comercial de estos espacios. Los datos sobre el año de apertura permitirían comprender el desarrollo histórico de los centros comerciales, mientras que el tamaño (GLA) y el número de tiendas (Shops) podrían ser utilizados para comparar y clasificar diferentes centros comerciales según su tamaño y capacidad comercial. Esto podría ser útil para inversores y desarrolladores interesados en el mercado inmobiliario comercial, así como para estudios de mercado y análisis de tendencias en la industria minorista.

### Histogramas

### Análisis





## FECUNDIDAD

### Análisis de las variables

Ubicación – Dato No Numérico (Tipo Texto).

Año – Dato No Numérico (Tipo Texto).

Edad media de la fecundidad – Dato Numérico (Tipo Número).

Tasa bruta de natalidad – Dato Numérico (Tipo Número).

Tasa bruta de reproducción – Dato Numérico (Tipo Número).

Tasa de fecundidad general – Dato Numérico (Tipo Número).

Tasa global de fecundidad – Dato Numérico (Tipo Número).

### Tipo de análisis

El objetivo del análisis de la base de datos "FECUNDIDAD" es comprender y estudiar en profundidad los patrones de fecundidad y las tasas de natalidad en diferentes ubicaciones geográficas y a lo largo del tiempo. Al ampliar este objetivo, se busca obtener una visión completa de cómo la fertilidad y las tasas de natalidad impactan en la dinámica poblacional de distintas regiones, así como su relevancia en términos socioeconómicos y de políticas públicas.

### Escogencia de la muestra

Esta muestra de datos ha sido recopilada con el propósito de analizar la evolución de la fecundidad y la tasa de natalidad en diferentes países o regiones, así como para comprender el impacto de factores socioeconómicos y demográficos en estas tasas. Los datos sobre la edad media de la fecundidad y las tasas de fecundidad general y global podrían ser útiles para estudiar tendencias de población y planificar políticas de salud y bienestar. Además, las tasas brutas de natalidad y reproducción pueden ser relevantes para estudios de crecimiento poblacional y proyecciones demográficas.

### Histogramas

## CANCER – DEATH RATE

### Descripción de las variables

County – Dato No Numérico (Tipo Texto).

FIPS – Dato Numérico (Tipo Número).

Met Objective of 45.5? (1) – Dato No Numérico (Tipo Texto).

Age-Adjusted Death Rate – Dato Numérico (Tipo Número).

Lower 95% Confidence Interval for Death Rate

Upper 95% Confidence Interval for Death Rate

Average Deaths per Year – Dato Numérico (Tipo Número).

Recent Trend (2) – Dato No Numérico (Tipo Texto).

Recent 5-Year Trend (2) in Death Rates – Dato Numérico (Tipo Número).

Lower 95% Confidence Interval for Trend – Dato Numérico (Tipo Número).

Upper 95% Confidence Interval for Trend – Dato Numérico (Tipo Número).

### Tipo de análisis

El objetivo del análisis es el de estudiar y comprender la tasa de mortalidad por cáncer en diferentes condados y cómo ha variado a lo largo del tiempo. Se busca identificar tendencias en la mortalidad por cáncer, determinar si se han cumplido objetivos de reducción de la mortalidad y analizar la incertidumbre asociada a estas tasas a través de los intervalos de confianza.

### Escogencia de muestra

Esta muestra de datos ha sido recopilada con el propósito de evaluar la carga del cáncer en diferentes áreas geográficas, identificar áreas de mejora en términos de políticas de salud y bienestar, y comprender la eficacia de intervenciones o programas para reducir la mortalidad por cáncer. Los datos sobre las tendencias recientes y los intervalos de confianza ayudarán a evaluar la consistencia y la significancia de los cambios en la mortalidad por cáncer en cada condado. Esta información es crucial para informar estrategias de salud pública y para la toma de decisiones en la planificación de servicios médicos y programas de prevención y tratamiento del cáncer.

### Histogramas

Este histograma muestra la distribución de la tasa de mortalidad por cáncer ajustada por edad en diferentes condados. En el eje horizontal se representan los intervalos de tasa de mortalidad, mientras que en el eje vertical se muestra la cantidad de condados que tienen una tasa de mortalidad dentro de cada intervalo. Este histograma permitirá visualizar cómo se distribuyen las tasas de mortalidad por cáncer en los condados y ayudará a identificar los rangos más comunes de mortalidad.

En este histograma se muestra cómo se acumulan las tasas de mortalidad ajustadas por edad a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de tasa de mortalidad, mientras que en el eje vertical se muestra la cantidad acumulativa de condados que tienen una tasa de mortalidad igual o inferior a cada intervalo. Este histograma permitirá entender qué porcentaje del total de condados tiene una tasa de mortalidad por cáncer igual o inferior a cierto valor, brindando una visión general de la distribución de tasas de mortalidad y su variabilidad.

: En este histograma se muestra la distribución de la frecuencia relativa de las tasas de mortalidad ajustadas por edad en los condados. La frecuencia relativa representa la proporción de condados que tienen una tasa de mortalidad por cáncer dentro de cada intervalo respecto al total de condados en la muestra. Este histograma permitirá visualizar las tendencias en las tasas de mortalidad y cómo se concentran las tasas en diferentes intervalos.

En este histograma se muestra cómo se acumulan las frecuencias relativas de las tasas de mortalidad ajustadas por edad a medida que avanzamos hacia valores más altos. En el eje horizontal se representan los intervalos de tasa de mortalidad, mientras que en el eje vertical se muestra la frecuencia relativa acumulativa hasta ese intervalo. Este histograma permitirá entender qué porcentaje del total de condados tiene una tasa de mortalidad por cáncer igual o inferior a cierto valor, brindando una visión general de la distribución de tasas de mortalidad y cómo se concentran los valores en diferentes rangos.

# CONCLUSIÓN

En conclusión, este proyecto final de estadística nos ha brindado una valiosa oportunidad para aplicar y consolidar los conocimientos adquiridos en nuestras clases en un contexto práctico y real. A través del análisis detallado de 10 bases de datos, hemos explorado diversas técnicas estadísticas que nos han permitido comprender la naturaleza y comportamiento de los datos en profundidad. Desde calcular medidas de tendencia central y dispersión hasta visualizar distribuciones mediante histogramas, hemos obtenido una visión completa de los conjuntos de datos estudiados.

La utilización de herramientas como Python y Excel ha sido fundamental para facilitar y agilizar nuestro análisis, brindándonos la capacidad de manejar grandes volúmenes de datos con eficiencia y precisión. Estas herramientas nos han permitido realizar tareas complejas de manera más eficiente, lo que ha ampliado nuestras capacidades analíticas y nos ha proporcionado una ventaja en el manejo y comprensión de datos.

Además, el estudio de la toma de muestras, el análisis de convergencia y las técnicas de análisis de correlación y lineal han enriquecido nuestro enfoque en el análisis de datos. Estas técnicas nos han permitido obtener resultados más significativos y confiables, asegurando que nuestras conclusiones estén respaldadas por evidencia sólida y representativa.

En última instancia, este proyecto ha reforzado nuestra comprensión de la estadística como una herramienta esencial para la toma de decisiones informadas en cualquier campo. Hemos adquirido habilidades prácticas y transferibles que nos permitirán enfrentar desafíos futuros con una mentalidad analítica y crítica. Estamos preparados para aplicar nuestras capacidades en el análisis y visualización de datos, lo que nos convierte en profesionales más competentes y capacitados para abordar problemas complejos y tomar decisiones fundamentadas en datos en nuestra vida personal y profesional.

# ANEXOS

*GitHub* **:** <https://github.com/Iamdaac/ProyectoFinal_Estadistica>