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:

Juan Castillo phd.

Nombre:

Samir Fanilla 3-751-1012

Grupo:

1SF-131

Fecha de entrega:

26 de julio de 2023

# Exportaciones de Ron en zonas Francas

La elección de llevar a cabo el presente análisis surgió debido a mi inquietud por conocer los diversos comportamientos del arancel en las zonas francas de nuestro país. En esta sección, se expondrán los resultados del meticuloso estudio de los datos obtenidos

Esta base de datos posee información de las exportaciones de ron realizadas por diversas zonas francas de Panamá.

La tabla posee los siguientes atributos:

* Año: categórico
* Continente: dato Categórico
* Vía: dato Categórico
* Zonas francas: dato categórico
* Arancel: dato Categórico
* Valor FOB: dato numérico
* Peso Bruto: dato numérico

Análisis

1. ¿Qué zona franca exporta la mayor cantidad de peso bruto?

Para esto analizamos la cantidad total de peso bruto exportado por cada zona franca en el periodo de 2019 a 2020

En esta grafica podemos ver como la zona franca que más exporto peso bruto de ron en el periodo estudiado fue PanaPark Free Zone, seguido por Panexport.

Estas dos zonas francas, son las que más exportaron este producto.

Con las exportaciones de ambas zonas francas combinadas, tenemos el 98% de las exportaciones totales. Debido a esto, nos estaremos centrando de ahora en adelante, solo en las exportaciones realizadas por estas dos zonas.

Confirmamos nuestra selección de zonas francas para priorizar los datos de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cuál zona franca ha pagado el valor más alto de FOB en el periodo de tiempo estudiado?

Para estudiar esto, debemos ver los datos de valor FOB pagado por cada zona.

En esta grafica podemos ver como la zona franca que más FOB ha pagado ha sido Panexport. Esto es interesante, debido a que PanaPark Free Zone exporto más peso bruto que Panexport, como podremos ver en la siguiente gráfica:

1. ¿Cuál es el continente al que más se exporta Ron?

Esto lo podemos analizar con la siguiente grafica que nos muestra por continente cuanto peso bruto se exporto.

Los continentes a los que más ron se exportó por las zonas francas estudiadas en el periodo 2019 – 2020 fueron Europa y América

1. ¿Cuál es la vía preferida de transporte de este tipo de producto?

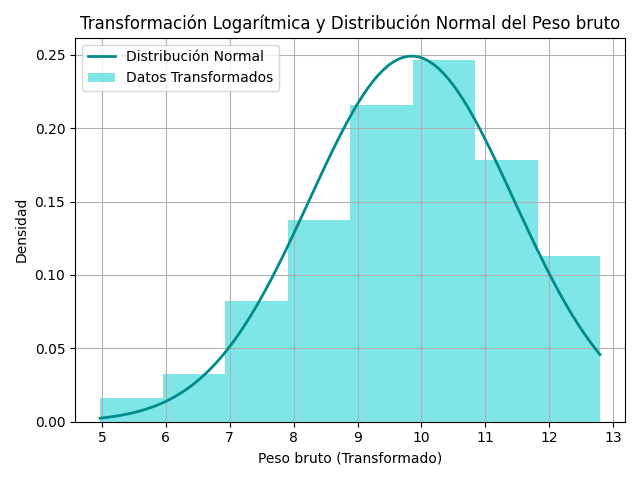
Esto se puede analizar con la siguiente grafica en donde se aprecia que la vía por la que más se exporta es la marítima seguida por la terrestre. Esto tiene mucho sentido debido a la cantidad de puertos que posee Panamá y que la mayoría de exportaciones se realiza a Europa

1. ¿Media de peso bruto exportado en el periodo estudiado?

Conseguimos los cálculos de la media, varianza y la desviación estándar del peso.



Con estos datos podemos realizar una gráfica de distribución normal de los datos de peso. Se transformó el peso a una escala logarítmica:



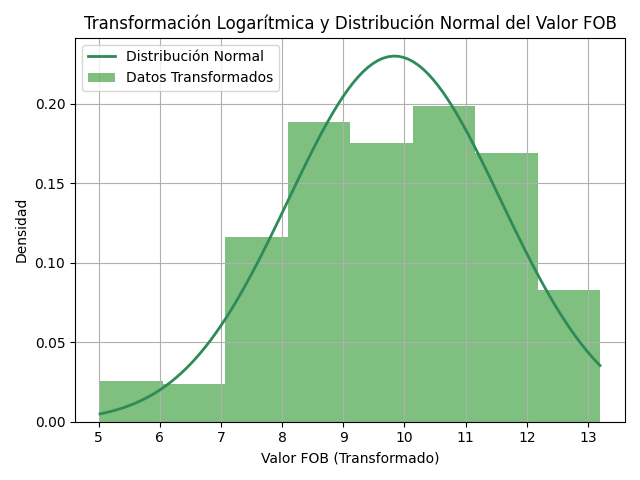
En este diagrama podemos observar como la mayoría de datos se encuentran hacia la derecha del pico de la distribución normal, lo que indica una asimetría hacia la izquierda (sesgo negativo)

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Conseguimos los cálculos de la media, varianza y la desviación estándar del FOB.



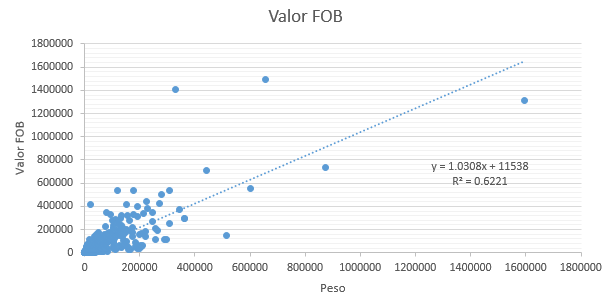
Con estos datos podemos realizar una gráfica de distribución normal de los datos del FOB, transformado a una escala logarítmica para poder apreciar el comportamiento mejor.



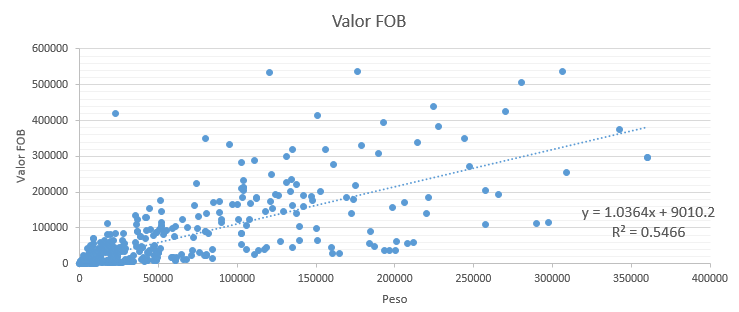
En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, por lo que podemos concluir que los datos no siguen esta distribución.

1. ¿Cuál es la correlación entre peso bruto y valor FOB

Para analizar esto saque una regresión lineal con las variables, buscando ver si había una relación entre el peso bruto y el valor FOB que se tuvo que pagar en la exportación. En la siguiente grafica podemos ver la regresión



En esta grafica podemos ver como existen unos puntos que se encuentran muy apartado de la mayoría de datos, por lo que realizamos una regresión lineal nuevamente, pero sin tener en cuenta estos puntos

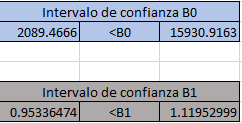


En esta grafica podemos observar de mejor manera el comportamiento de los datos, y vemos como la mayoría de datos se encuentran bastante dispersos de la línea de tendencia, lo que indica una baja relación entre las variables

La regresión lineal nos otorga la siguiente ecuación para y:

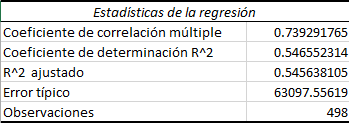
y = 1.0364x + 9010.2

A continuación conseguimos los intervalos de confianza al 95% para la intercepción y la variable x1



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente tabla, podemos observar ver como el R^2 se encuentra muy lejos de 1, lo que nos confirma el análisis de que no existe una relación muy fuerte entre las variables



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

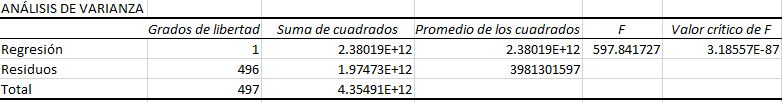
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos la prueba de F con los datos de la tabla ANOVA para comprobar la hipótesis nula



Podemos observar que el valor de F es muy grande mientras que el valor critico de F muy pequeño, por lo tanto podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.5465 también nos indica que alrededor del 54.65% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa y la regresión lineal consigue explicar gran parte de la variabilidad observada

# Exportaciones de Manufactura de piedra en zonas francas

La decisión de emprender el siguiente análisis se fundamentó en mi genuina curiosidad por comprender los distintos comportamientos del arancel en las zonas francas de nuestra nación. Con ello, en esta sección, se expondrán los resultados obtenidos tras minuciosa investigación.

Esta base de datos posee información de las exportaciones de manufactura de piedra realizadas por diversas zonas francas de Panamá.

Específicamente la manufactura de piedra que puede ser Abrasivos naturales o artificiales en polvo o gránulos con soporte de papel o cartón solamente, incluidos recortados, cosidos o unidos de otra forma

La tabla posee los siguientes atributos:

* Año: categórico
* Continente: dato categórico
* Mes: categórico
* Puerto de embarque: dato categórico
* Zonas francas: dato categórico
* Arancel: dato categórico
* Valor FOB: dato numérico
* Peso Neto: dato numérico

Análisis

1. ¿Qué zona franca exporta la mayor cantidad de peso neto?

Podemos analizar esto con la siguiente gráfica:

En esta grafica podemos apreciar como la mayoría de las exportaciones son realizadas por la zona franca de Panamá Pacifico.

Debido a esto, estaremos centrando el estudio solo en las exportaciones realizadas por esta zona franca.

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cuál ha sido el año en el que más se pagó FOB?

La siguiente grafica nos ayudara a entender en que año se pagó más FOB

En esta podemos ver como se ha mantenido estable la cantidad anual que se paga en FOB, con un aumento en el 2021. Podremos ver en la siguiente pregunta si esto se ha debido a mayor exportación en peso neto de este producto

1. ¿Ha crecido la demanda de este producto en el periodo de tiempo estudiado?

Para estudiar esto observaremos como se ha comportado la exportación realizada de este producto a través de los años.

En esta grafica podemos apreciar cómo se ha mantenido estable la exportación, bajando un poco en 2020 posiblemente debido a la pandemia, para luego dispararse nuevamente en 2021.

Debido a esto podemos asumir que el aumento de FOB pagado en el 2021, fue debido a un aumento en las exportaciones de ese año

1. ¿Cuál es la vía preferida de exportación de este tipo de producto?

Para esto analizaremos las distintas vías por las que ha sido exportado este producto:

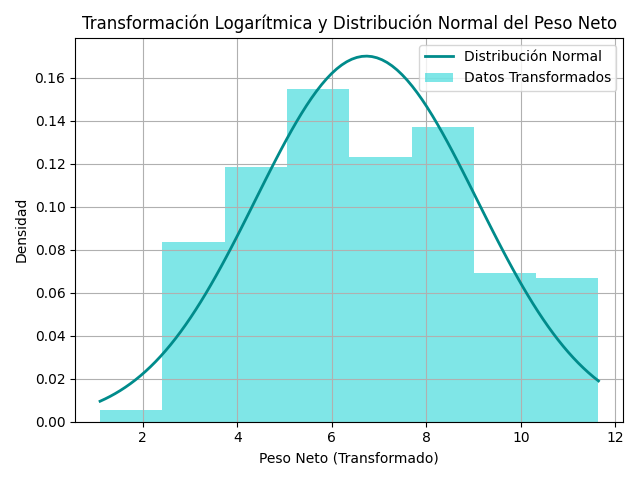
Podemos apreciar que la mayoría de exportaciones salen por vía Marítima a través del puerto de Manzanillo.

1. ¿Media de peso neto exportado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen

: 

Con estos datos podemos analizar la distribucion normal de el peso neto exportado en el perido estudiado



En esta grafica transformamos los valores del peso neto a una escala logarítmica para poder apreciar mejor el comportamiento de la distribución normal.

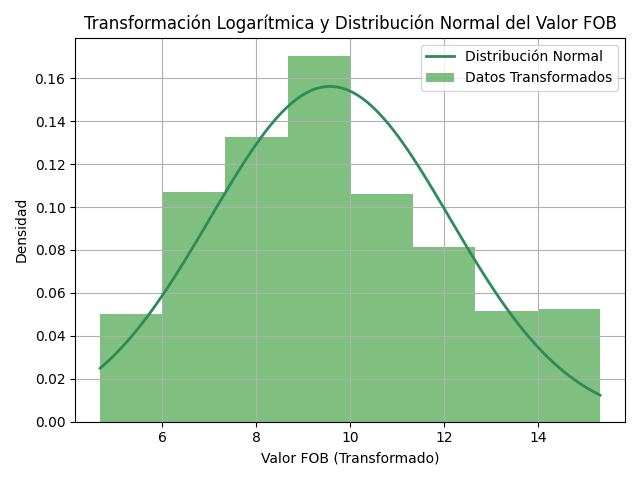
Podemos observar como los datos de peso que estamos analizando no parecen seguir una distribución normal debido a que el histograma no muestra la forma característica de campana que deben tener una distribución normal

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del FOB en la siguiente imagen



Con estos datos podemos analizar la distribución normal del valor FOB en el periodo estudiado



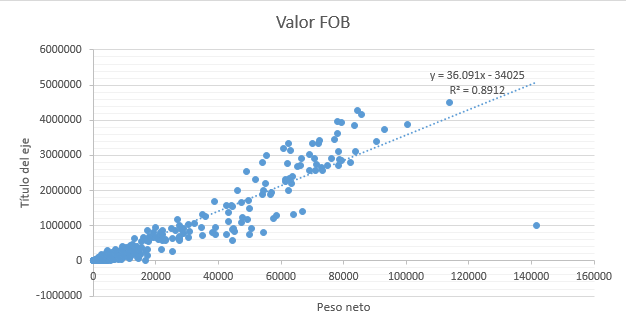
El valor FOB también fue transformado a una escala logarítmica para poder apreciar mejor el comportamiento

En este diagrama podemos observar como la mayoría de datos se encuentran hacia la izquierda del pico de la distribución normal, lo que indica que existe una Asimetría hacia la derecha (sesgo positivo)

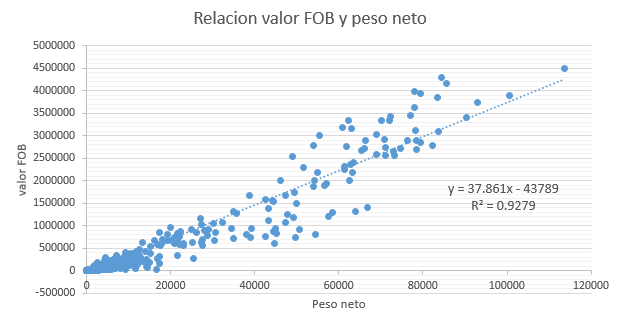
1. ¿Cuál es la correlación entre peso neto y valor FOB

Para poder observar la relación entre estas dos variables de peso neto y FOB, sacamos la regresión lineal.

En la siguiente grafica podemos observar el comportamiento de estas variables.



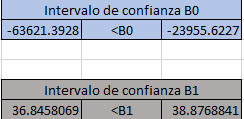
En esta grafica podemos observar como existe a primera vista una relación que indica que entre más peso neto exportado, más FOB se paga. Sin embargo también podemos visualizar un punto el cual se sale mucho del comportamiento de los otros puntos, por lo cual decidimos realizar nuevamente la regresión sin este punto.



En esta grafica se puede apreciar claramente, como a mayor cantidad de peso neto exportado, más se paga en valor FOB, ya que los puntos se acercan mucho a la línea de tendencia.

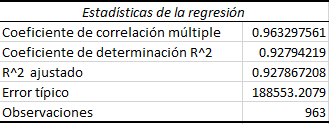
La regresión lineal nos da para la siguiente formula:

También podemos conseguir el intervalo de confianza al 95% para la intersección y la variable x1 de la fórmula de regresión



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente tabla podemos observar como el valor de R^2 está muy cerca de 1, lo que nos indica que existe una relación entre las variables analizadas



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

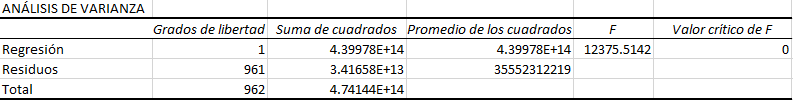
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos de igual manera la prueba de F para ver si podemos rechazar la hipótesis nula



Observamos que el valor de F es mucho más grande que el valor critico de F por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.9279 también nos indica que alrededor del 92.79% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa y la regresión lineal consigue explicar gran parte de la variabilidad observada

# Exportaciones de Placas metales en zonas francas

Se eligió realizar el siguiente análisis debido a que tuve una curiosidad en saber los diversos comportamientos de este arancel en las zonas francas de nuestro país, por lo cual estaremos analizando los datos de ello en esta sección.

Esta base de datos posee información de las exportaciones de manufactura diversas del arancel 8310.00.00.00, el cual categoriza el metal común.

Específicamente estudiaremos la subdivisión de “Placas indicadoras, placas rótulo, placas de direcciones y placas simil., cifras, letras y signos diversos, de metal común”.

La tabla posee los siguientes atributos:

* Año: dato categórico
* Continente: dato categórico
* Mes: dato categórico
* Puerto de embarque:
* Vía: dato categórico
* Zonas francas: dato categórico
* Arancel: dato categórico
* Valor FOB: dato numérico
* Peso Bruto: dato numérico

Análisis

1. ¿Qué zona franca exporta la mayor cantidad de peso bruto?

En el periodo estudiado podemos observar que las exportaciones se hacen por tres zonas francas, en la siguiente grafica podemos observar la cantidad de exportación total en el periodo de 2017 a 2021

Como se puede observar en la gráfica anterior, la gran mayoría de las exportaciones son realizadas por Panexport

Esta zona franca realizo exportaciones sumando 134,000 en peso net esto siendo un equivalente al 94% de las exportaciones totales en este periodo.

Las otras dos zonas francas, tienen un comportamiento más cercano, lo cual me parece más interesante para estudiar.

Debido a esto, los siguientes análisis se harán centrando en estas dos zonas francas

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cuál ha sido el año en el que más se pagó FOB?

Este análisis lo podemos realizar con la siguiente gráfica, en la cual ordenamos por año, la cantidad de FOB que fue pagado por las zonas francas que estudiamos

Podemos observar que el pago de FOB por año se mantiene bastante estable hasta el 2020, luego en el 2021 ocurre un alza bastante grande. Esto se debe a la cantidad de peso bruto de las exportaciones de estas dos francas.

1. ¿Cómo ha sido el comportamiento a través de los años de las exportaciones de este producto?

En la siguiente grafica podemos observar por año como se ha comportado la cantidad total de exportaciones de este producto.

Podemos observar que como el FOB, el peso neto exportado por año no cambiaba mucho hasta el 2020, luego en el 2021 se disparó.

Si observamos las exportaciones por año clasificado por zona franca podemos ver la razón de esto.

Aquí vemos como durante los años 2017 y 2020, solo exportaba Panamá pacífico y luego en 2021 Marpesca se unió a las exportaciones, debido a esto vemos el aumento grande en exportaciones este año

1. ¿Cuál es la vía preferida de transporte de este tipo de producto?

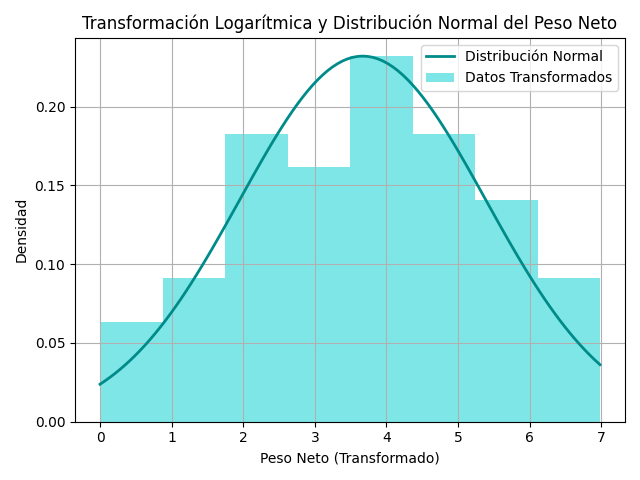
La vía preferida de exportación para este producto fue la marítima, como se puede observar en la siguiente gráfica:

1. ¿Promedio de peso bruto exportado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen



Con estos datos podemos sacar la siguiente distribución normal del peso



Para poder analizar mejor los datos de la distribución normal de los datos, decidí transformar los valores a una escala logarítmica para poder visualizar mejor el comportamiento de la distribución normal

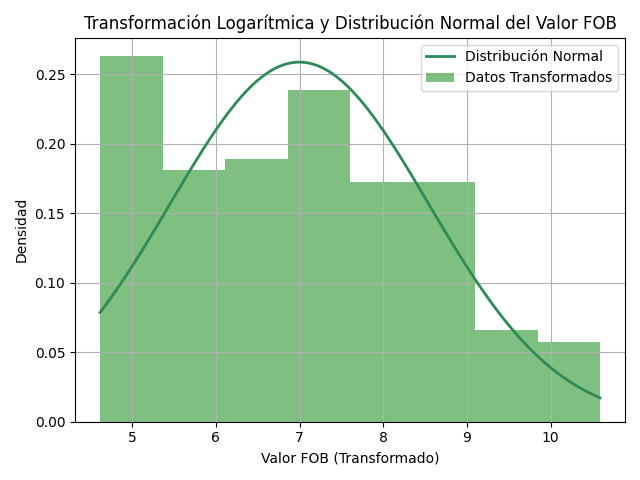
En este diagrama podemos observar como el histograma sigue de manera bastante parecida a la forma de campana que deben tener las distribuciones normales. Por lo que podemos concluir que estos datos si siguen una distribución normal

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del FOB en la siguiente imagen



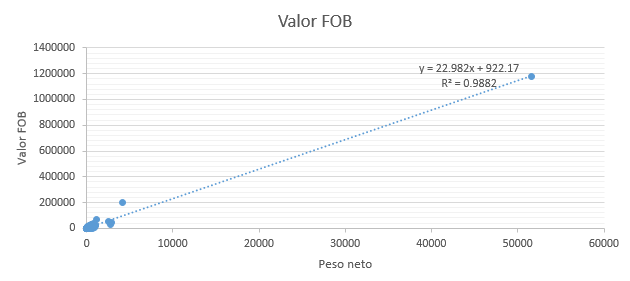
Con estos datos podemos analizar el comportamiento de los valores del FOB en una distribución normal



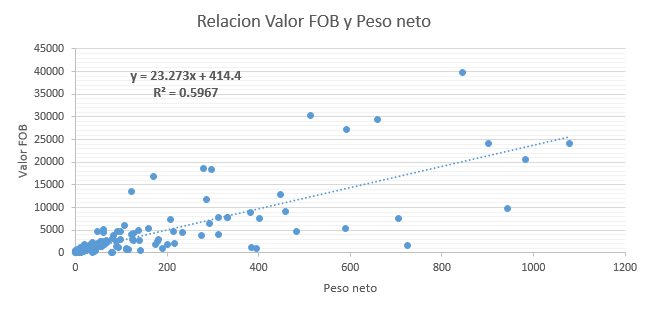
En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, por lo que podemos concluir que los datos no siguen esta distribución.

1. ¿Cuál es la correlación entre peso bruto y valor FOB

Sacamos una regresión lineal para poder observar la relación entre estas dos variables. En las siguientes graficas podemos observar el resultado



En esta grafica podemos apreciar como un solo punto se encuentra muy distante del resto, por lo cual decidimos realizar la regresión nuevamente eliminando este punto

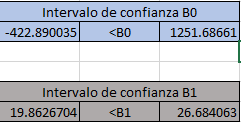


Con esta nueva regresión lineal podemos observar como la mayoría de puntos cuando se exporta más de 200 en peso neto, se encuentran muy dispersos de la línea de tendencia.

Para esta regresión lineal conseguimos la siguiente formula de y:

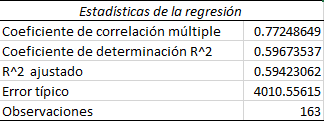
y = 23.273x + 414.4

Para esta fórmula podemos analizar los rangos de confianza al 95% para la intercepción y el la variable x1



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente tabla podemos ver los datos de esta regresión:



Podemos observar como el valor de R^2 está muy lejos de 1, lo que nos indica que no existe una relación fuerte entre las variables

La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

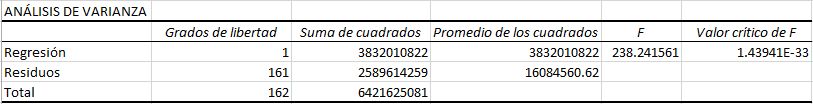
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

También podemos realizar la prueba de F para rechazar la hipótesis nula con los datos de ANOVA



En estos datos podemos observar como el valor de F es muy grande mientras que el valor critico de F es muy pequeño, muy por debajo de 0.05 por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.5967 también nos indica que alrededor del 59.67% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa y la regresión lineal consigue explicar gran parte de la variabilidad observada

# Exportaciones Medicinas como principales mercaderías

Motivado por el deseo de indagar en los variados comportamientos del arancel en las zonas francas del país, se procedió a realizar el presente análisis. A continuación, se presentarán los resultados derivados de un riguroso examen de los datos recopilados en esta materia.

Esta base de datos posee información acerca de las principales exportaciones realizadas en todo el País de Panamá. Estaremos estudiando de esta base de datos el arancel 3004.90.99.00.00 que corresponde a la clasificación de medicinas, específicamente:

“Medicamentos constituidos por productos mezclados o sin mezclar, preparados para usos terapéuticos o profilácticos, dosificados dosis.

La tabla posee los siguientes atributos:

* Año: Dato categórico
* Mes: Dato categórico
* Arancel: Dato categórico
* Valor FOB: Dato numérico
* Peso bruto: Dato numérico
* Vía: Dato categórico

Análisis

1. ¿Cuál es el país al que más se le exporta este arancel?

En la siguiente grafica podemos ver como el país al que más se le exporta este arancel es el país con código 236, el cual corresponde a Belice

De ahora en adelante el estudio se centrara en los datos de la exportación de este arancel a Belice

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cómo ha cambiado la cantidad de peso neto exportado en el tiempo estudiado?

Podemos observar, que la cantidad de peso neto exportado a Belice aumento mucho en el 2019, y se ha mantenido estable desde entonces

1. ¿En qué meses se exporta más de este arancel?

En la siguiente grafica podemos observar cual es el mes en el que más se exporta. La grafica esta ordenada en forma descendente y vemos que el mes en el que más se exporta es Agosto, mientras que marzo es el que menos exporta.

1. ¿Cuál es el mes en el que más FOB se paga?

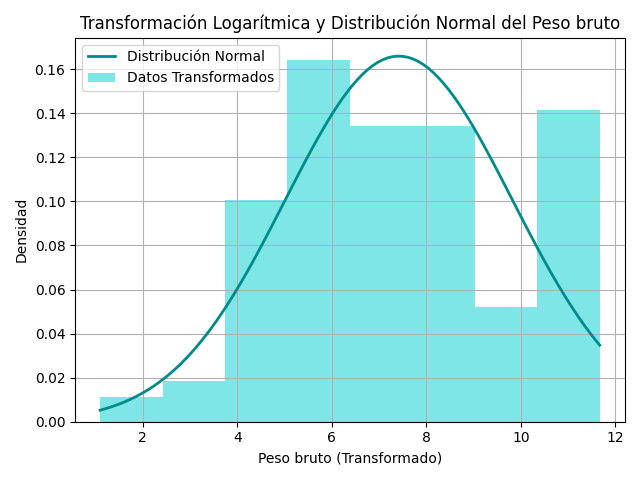
En la siguiente grafica de FOB pagado por mes, podemos observar como el comportamiento es bastante parecido a la gráfica de peso exportado por mes, esto debido a que entre más se exporta, más se está pagando por mes.

1. ¿Promedio de peso bruto exportado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen



Con estos datos podemos graficar la distribución normal del peso bruto en escala logarítmica



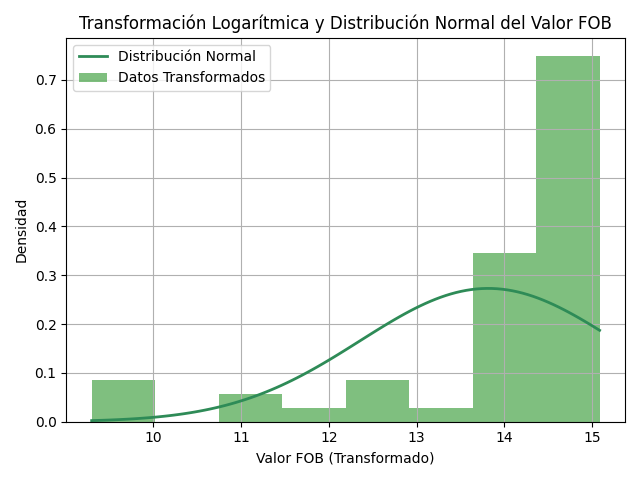
En esta grafica podemos observar como el histograma no sigue una distribución normal, debido a que no tiene la forma particular de campana que debería tener si su comportamiento fuera como una distribución normal

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del FOB en la siguiente imagen



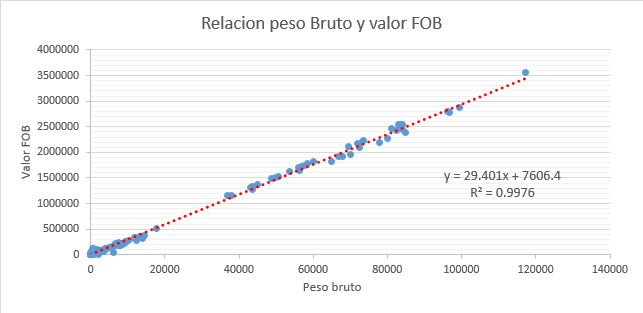
Con estos datos podemos graficar la distribución normal del peso bruto en escala logarítmica



En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, por lo que podemos concluir que los datos no siguen esta distribución.

1. ¿Cuál es la correlación entre peso bruto y valor FOB

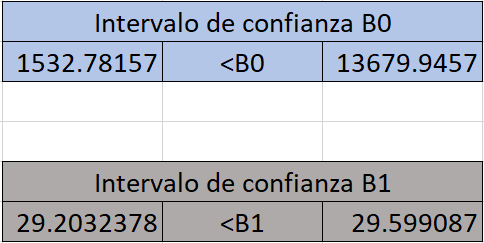
Esto lo podemos analizar con la siguiente gráfica. En ella podemos ver como se relaciona el peso bruto exportado, con el valor FOB que se tuvo que pagar



En esta grafica se puede apreciar, como la relación entre peso bruto y valor FOB se puede explicar muy bien con la regresión lineal, ya que los puntos están muy cerca de la línea de tendencia.

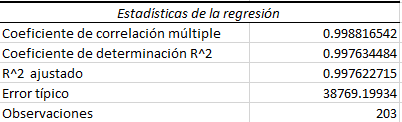
Esta regresión lineal nos devuelve la siguiente fórmula para y:

A estos valores de intercepción y variable x1 le sacamos los intervalos de confianza a 95%



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente tabla observamos como el valor de R^2 está bastante cerca de 1 validando nuestras observaciones



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis para comprobarlas con análisis a ANOVA y la prueba F

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

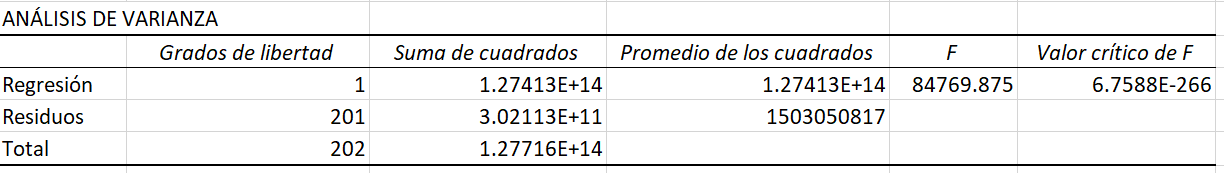
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

También realizamos la evaluación de F para comprobar la hipótesis nula con el cuadro de ANOVA



Observamos que el valor de F es mucho más grande que el valor critico de F por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.9976 también nos indica que alrededor del 99.76% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa y la regresión lineal consigue explicar gran parte de la variabilidad observada

# Importaciones de Cajas en zonas francas

Se eligió realizar el siguiente análisis debido a que tuve una curiosidad en saber los diversos comportamientos de este arancel en las zonas francas de nuestro país, por lo cual estaremos analizando los datos de ello en esta sección.

La base de datos de importaciones en zonas francas posee información de las importaciones que realizan estás en Panamá.

En el siguiente análisis nos estaremos centrando en las importaciones del arancel 4819.10.00.00.00, el cual corresponde a Cajas de papel o cartón corrugados

La tabla posee los siguientes atributos:

* Año: Dato categórico
* Continente: Dato categórico
* Vía: Dato categórico
* Zonas francas: Dato categórico
* Puerto de desembarque: Dato categórico
* Mes: Dato categórico
* Arancel: Dato categórico
* Valor FOB: Dato numérico
* Peso neto: Dato numérico

Análisis:

1. ¿Qué zona franca importa la mayor cantidad de peso neto?

En esta grafica podemos observar las zonas francas que más han importado de este arancel en el periodo estudiado de 2020 - 2021

Llama la atención como dos de estas zonas francas importan muchísimo más que las otras.

Panama Pacífico y PanaPark Free Zone, en conjunto tienen el 90% de las importaciones de este arancel, por lo cual no centraremos en ellas para las siguientes preguntas

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cómo se ha comportado la importación de este arancel?

El peso bruto que se ha ido importando de este arancel aumento bastante en ambas zonas francas en el 2021, de lo que se había importado en el 2020.

1. ¿A qué puerto de Panamá llega la mayoría de esta mercancía?

La mayoría de las importaciones de este producto llegan a Paso Canoas, por medio terrestre y al puerto de Manzanillo por medio marítimo

1. ¿Cómo se comporta el FOB de este arancel?

En la siguiente grafica podemos ver como en los meses de Marzo y Diciembre se ha pagado el mayor FOB de este arancel.

Se puede suponer que esto es debido a que se debe estar importando más en esos meses.

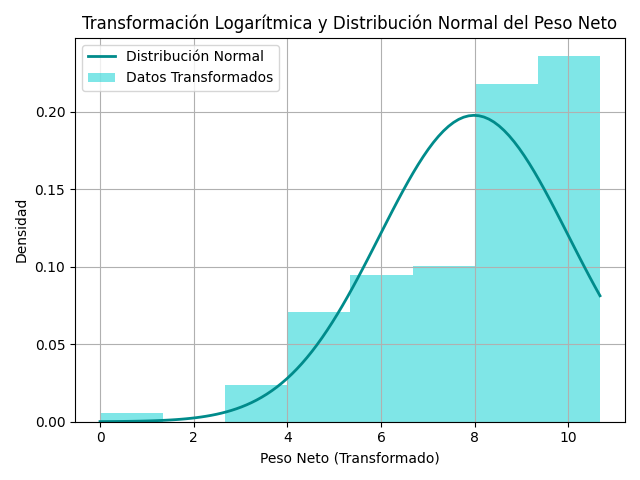
Si vemos la siguiente grafica que muestra la cantidad total de peso bruto importado por mes, observamos que realmente se importa más en octubre y junio que en marzo y diciembre.

1. ¿Promedio de peso neto importado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen



Analizando el peso como una distribución normal



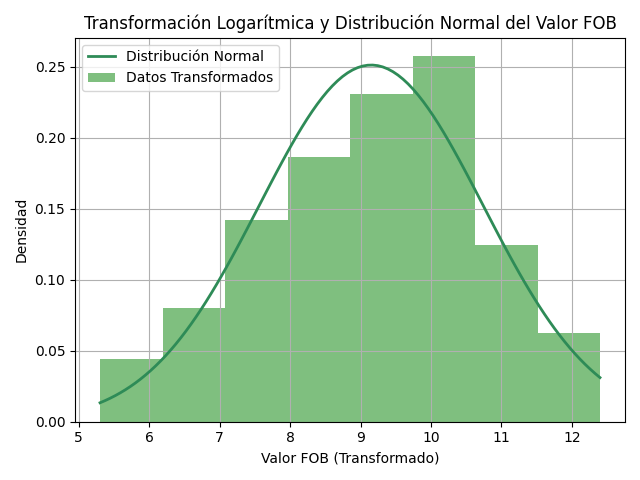
En este diagrama podemos observar como la mayoría de datos se encuentran hacia la derecha del pico de la distribución normal, lo que indica una asimetría hacia la izquierda (sesgo negativo). Podemos concluir que estos datos no siguen una distribución normal debido a como no forman la forma de campana de las distribuciones normales

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del FOB en la siguiente imagen



Con estos datos podemos analizar la distribución normal en escala logarítmica del peso

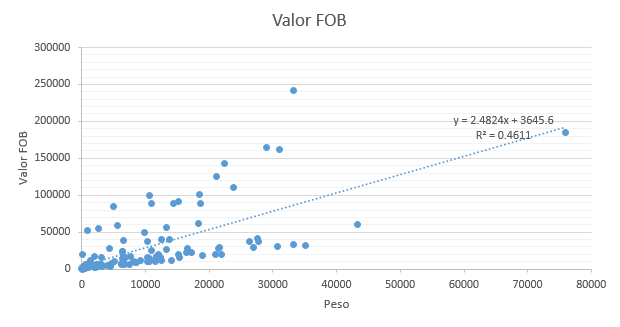


En este diagrama podemos observar como el histograma sigue de manera bastante parecida a la forma de campana que deben tener las distribuciones normales. Por lo que podemos concluir que estos datos si siguen una distribución normal

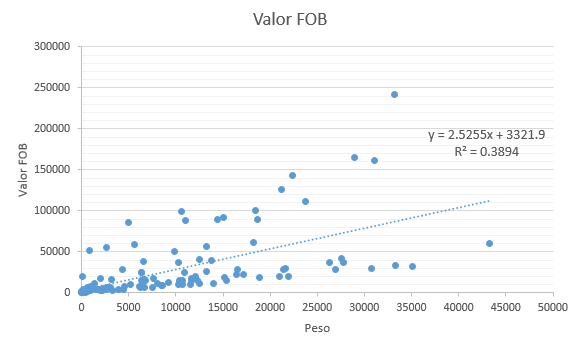
1. ¿Cuál es la correlación entre peso bruto y valor FOB?

Para poder analizar la relación entre las variables de peso neto y el valor FOB

Colocamos los diversos puntos que tenemos en una gráfica de dispersión, y sacamos la línea de tendencia.



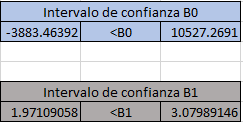
En esta grafica podemos apreciar como existe un punto que se sale por completo de la zona donde se encuentran la mayoría de datos, por lo que realizaremos nuevamente la regresión sin tomar en cuenta este punto



En esta nueva regresión podemos apreciar de mejor manera si existe una relación clara entre el peso y el valor FOB y según esta grafica no existe una relación fuerte entre estas variables, debido a como se encuentran muy dispersos los puntos de la línea de tendencia

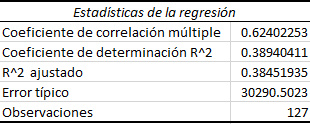
La regresión lineal realizada nos dio la siguiente fórmula para y:

Y para esta fórmula tenemos los siguientes intervalos de confianza al 95% para la intercepción y el valor de la variable x1



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente grafica podemos ver los datos de la regresión, incluyendo el R^2 el cual nos confirma la observación de que no existe una relación entre las variables estudiadas, debido a que el valor de R^2 se encuentra lejos de 1



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis para comprobarlas con análisis a ANOVA y la prueba F

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

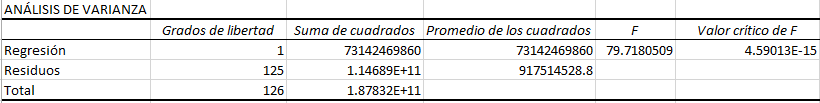
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos un análisis de F para comprobar la hipótesis nula, con la tabla de ANOVA



Podemos ver como el valor de F es muy grande y el del valor critico de F muy pequeño, mucho menor a 0.05, por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.3894 también nos indica que alrededor del 38.94% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa, sin embargo la regresión lineal no consigue explicar gran parte de la variabilidad observada

Importaciones zonas francas: Maquinas trituradoras

La elección de llevar a cabo el presente análisis surgió debido a mi inquietud por conocer los diversos comportamientos del arancel en las zonas francas de nuestro país. En esta sección, se expondrán los resultados del meticuloso estudio de los datos obtenidos.

En el siguiente análisis nos estaremos centrando en las importaciones del arancel 8474.90.00.00, el cual corresponde a Maquinas que se encargan de triturar, cribar, lavar y pulverizar tierra, piedra u otra materia

La tabla posee los siguientes atributos:

* Año: Dato categórico
* Continente: Dato categórico
* Vía: Dato categórico
* Zonas francas: Dato categórico
* Puerto de desembarque: Dato categórico
* Mes: Dato categórico
* Arancel: Dato categórico
* Valor FOB: Dato numérico
* Peso Neto: Dato numérico

Análisis:

1. ¿Qué zona franca importa la mayor cantidad de peso neto?

La zona franca que importa la mayor cantidad en peso neto de este arancel es Panamá pacífico.

Las importaciones de esta zona franca corresponden casi al 100% de las importaciones en el periodo estudiado de 2017 a 2021.

Debido a esto nos centraremos en el estudio de esta zona franca

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cómo se ha comportado la importación de este arancel?

Podemos observar el comportamiento de importación de este arancel por año en la siguiente gráfica.

En el vemos que la cantidad de piezas que se están importando se mantenía en niveles estables, hasta que disminuyo en gran parte a partir del año 2020, posiblemente debido a los efectos de la pandemia en las cadenas de suministros, pero pueden haber otras razones.

1. ¿Por cuál vía de transporte se suele importar este arancel?

Las importaciones de este arancel se suelen hacer en su mayoría por medios marítimos, seguido por la vía aérea

1. ¿Cómo se comporta el FOB de este arancel?

En la siguiente grafica podemos ver el FOB de las importaciones de este arancel por año. La grafica se comporta como se esperaría por la cantidad de peso importado cada año que vimos en la antes.

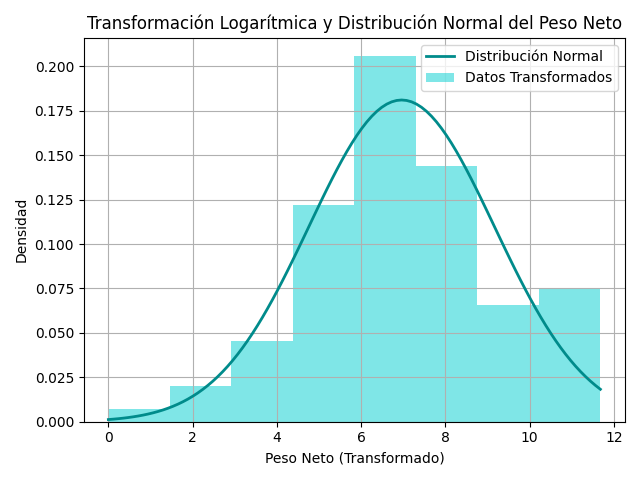
Sin embargo, un punto interesante es que se importó mucho más peso neto en 2017 y 2019, pero el FOB de 2018 es mayor al de estos dos años

1. ¿Promedio de peso neto importado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen



Con estos datos procedemos a analizar el comportamiento del peso como una distribución normal.



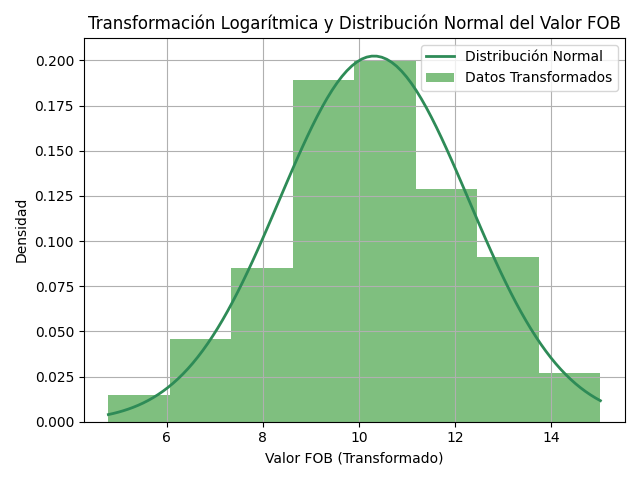
En este diagrama podemos observar como la mayoría de datos se encuentran hacia la derecha del pico de la distribución normal, lo que indica una asimetría hacia la izquierda (sesgo negativo)

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del FOB en la siguiente imagen



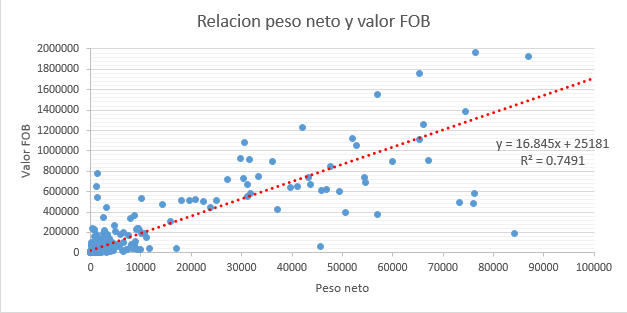
Con estos datos procedemos a analizar el comportamiento del FOB como una distribución normal.



En este diagrama podemos observar como el histograma sigue de manera bastante parecida a la forma de campana que deben tener las distribuciones normales. Por lo que podemos concluir que estos datos si siguen una distribución normal

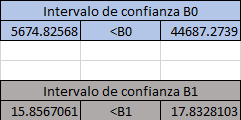
1. ¿Cuál es la correlación entre peso neto y valor FOB?

En la siguiente grafica podemos observar la regresión lineal entre el peso neto y el valor FOB



Esta regresión lineal nos da la siguiente ecuación para el valor de y:

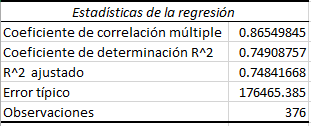
Y para la intercepción y la variable x1 tenemos los siguientes intervalos de confianza al 95%



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

De la gráfica podemos observar que los puntos están bastantes dispersos, pero se suelen mantener cerca de la línea de tendencia. Debido a esto podemos asumir que la regresión es aceptable.

El valor de R^2 no está muy lejos de 1, pero tampoco lo suficiente cerca como para poder decir que la regresión lineal para estas dos variables es buena



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis para comprobarlas con análisis a ANOVA y la prueba F

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

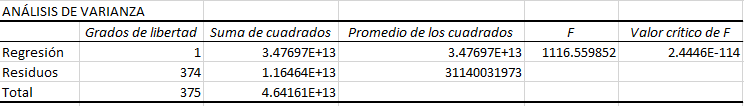
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos una prueba de F para comprobar la hipótesis nula. Para esto hacemos uso de una tabla de ANOVA



En esta tabla podemos observar que el valor de F es mucho más grande que el valor critico de F por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.7491también nos indica que alrededor del 74.91% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa y la regresión lineal consigue explicar gran parte de la variabilidad observada.

# Importaciones Centrifugadoras y aparatos para filtrar lubricantes o carburantes en motores en zonas francas

La decisión de emprender el siguiente análisis se fundamentó en mi genuina curiosidad por comprender los distintos comportamientos del arancel en las zonas francas de nuestra nación. Con ello, en esta sección, se expondrán los resultados obtenidos tras minuciosa investigación.

La base de datos de importaciones posee información de las importaciones que se realizan en Panamá por medio de zonas francas.

En el siguiente análisis nos estaremos centrando en las importaciones del arancel 8421.23.00.00, el cual corresponde a Centrifugadoras y aparatos para filtrar lubricantes o carburantes en motores

La tabla posee los siguientes atributos:

* Año: Dato categórico
* Continente: Dato categórico
* Vía: Dato categórico
* Zonas francas: Dato categórico
* Puerto de desembarque: Dato categórico
* Mes: Dato categórico
* Arancel: Dato categórico
* Valor FOB: Dato numérico
* Peso Neto: Dato numérico

Análisis

1. ¿Qué zona franca importa la mayor cantidad de peso neto?

En la siguiente grafica podemos observar como las zonas de Panamá Pacifico y Chilibre son las que más peso han importado en el periodo estudiado

Estas dos zonas en conjunto suman en total un 96% del peso total importado, como se aprecia en la siguiente gráfica:

Me interesa estudiar el comportamiento del resto de las tres zonas francas que conforman el 3% restante

Por lo que las siguientes preguntas se responderán solo con los datos de estas tres zonas francas

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cómo se ha comportado la importación de este arancel a través de los años?

Este arancel ha sido importado generalmente en las mismas cantidades cada año, por diferentes zonas francas. En los años 2018 y 2019, solo fue importado por una zona franca

1. ¿Por cuál vía suele llegar la mayoría de esta mercancía?

La mayoría de esta mercancía suele llegar por vía marítima

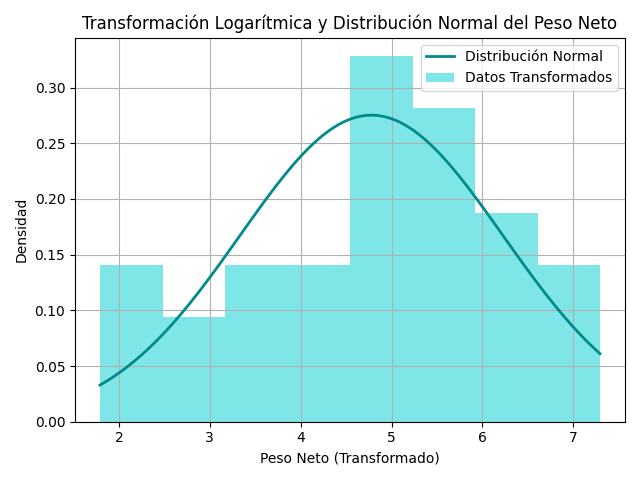
1. ¿Cómo se comporta el FOB de este arancel?

El FOB total pagado por año se comporta siguiendo el comportamiento de la cantidad de peso importado por año

1. ¿Promedio de peso neto importado en el periodo estudiado?



Con estos datos podemos observar el comportamiento del peso como una distribución normal

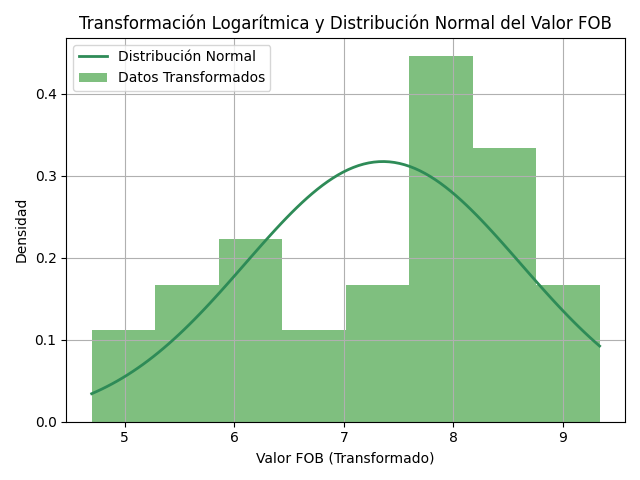


En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, esto lo podemos ver en los picos que hay en los extremos de la distribución de los datos. Debido a esto podemos concluir que los datos no siguen esta distribución.

1. ¿Promedio de valor FOB pagado en el periodo estudiado?



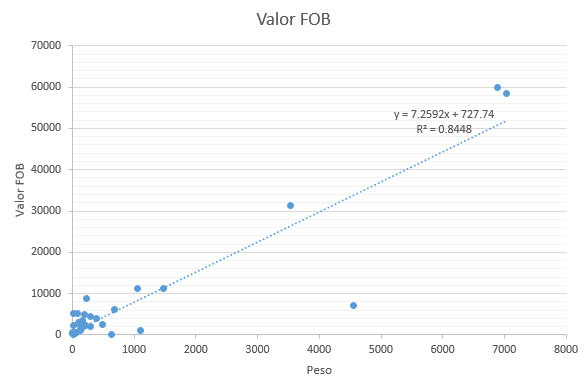
Con estos datos podemos analizar el comportamiento del FOB como una distribución normal



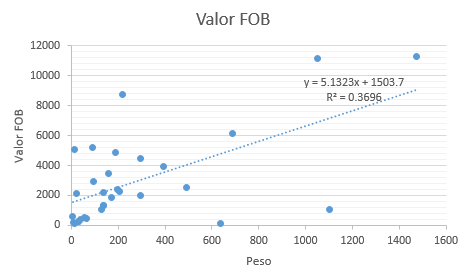
En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, por lo que podemos concluir que los datos no siguen esta distribución.

1. ¿Cuál es la correlación entre peso neto y valor FOB?

La relación de estas dos variables en esta base de datos la podemos observar en las siguientes gráficas.



Esta grafica muestra la regresión lineal de toda la muestra tomada, pero como se observa, existen unos puntos que se encuentran muy lejos de la mayoría de puntos, por lo cual realizamos la regresión nuevamente sin tomar en cuenta estos puntos.

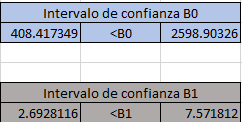


En esta nueva grafica podemos ver como realmente no existe una relación entre las dos variables estudiadas, debido a como los puntos no siguen realmente a la línea de tendencia

La regresión Lineal nos da la siguiente fórmula para el valor de y:

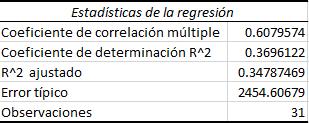
y = 5.1323x + 1503.7

Para esta fórmula sacamos el intervalo de confianza al 95% para la intercepción y la variable x1



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En la siguiente tabla vemos el valor de R^2, el cual está muy lejos de 1, lo cual indica que no existe una relación entre las variables estudiadas



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis para comprobarlas con análisis a ANOVA y la prueba F

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

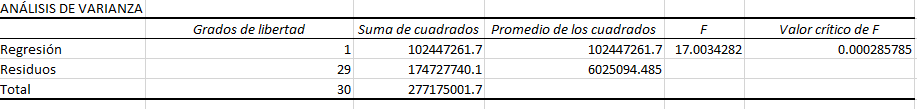
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos el análisis de la prueba F con el cuadro ANOVA



Con los valores de F y F critico Observamos que el valor de F es mucho más grande que el valor critico de F por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.3696 también nos indica que alrededor del 36.96% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

Por lo tanto podemos concluir que existe una relación significativa, sin embargo la regresión lineal no consigue explicar gran parte de la variabilidad observada

# Importaciones zonas francas: Proyectores

Motivado por el deseo de indagar en los variados comportamientos del arancel en las zonas francas del país, se procedió a realizar el presente análisis. A continuación, se presentarán los resultados derivados de un riguroso examen de los datos recopilados en esta materia.

La base de datos de importaciones posee información de las importaciones que se realizan en Panamá por medio de zonas francas.

En el siguiente análisis nos estaremos centrando en las importaciones del arancel 9007.20.00.00, el cual corresponde proyectores

La tabla posee los siguientes atributos:

* Año: Dato categórico
* Continente: Dato categórico
* Vía: Dato categórico
* Zonas francas: Dato categórico
* Puerto de desembarque: Dato categórico
* Mes: Dato categórico
* Arancel: Dato categórico
* Valor FOB: Dato numérico
* Peso Neto: Dato numérico

Preguntas

1. ¿Qué zona franca importa la mayor cantidad de peso neto?

La mayor parte de las importaciones de proyectores en zonas francas son realizadas por Albrook, como se observa en la siguiente grafica

La mayoría de las importaciones se dan sin Colon Maritme investor S.A,

Debido a esto no se tendrán en cuenta sus datos para las siguientes preguntas

Confirmamos nuestra selección de zonas francas para priorizar los datos que nos interesan de las más significativas con los siguientes histogramas de Frecuencia, Frecuencia relativa y frecuencias acumuladas.

Estos histogramas nos confirman la idea de que el mayor valor de datos que nos interesan se encuentra en las zonas francas que elegimos

1. ¿Cómo se ha comportado la importación de este arancel?

Las importaciones de este arancel han disminuido como se aprecia en la siguiente grafica

1. ¿Por cuál vía llega a Panamá la mayoría de esta mercancía?

La vía para las importaciones de esta dividida entre la terrestre utilizada por la zona franca de Panmá pacífico, mientras que la zona de Albrook utiliza la vía aérea para recibir las importaciones

1. ¿Cómo se comporta el FOB de este arancel?

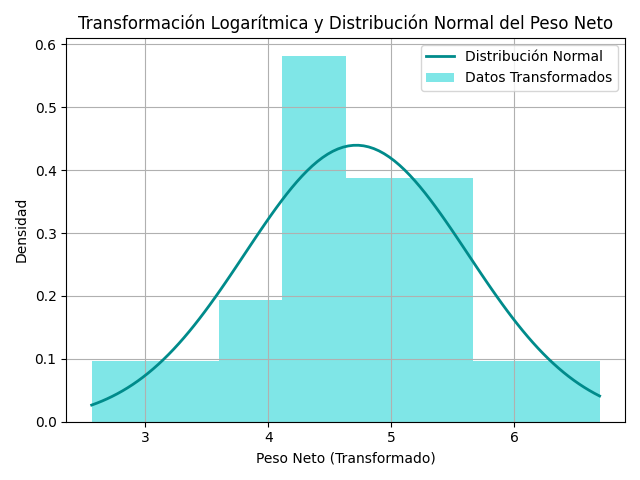
En la siguiente grafica podemos observar cómo se paga mucho más por las importaciones que ha recibido Panamá pacifico, la cuales han sido todas por via terrestre, mientras que Albrook que las recibe por vía aérea sus importaciones han pagado menos en FOB

1. ¿Promedio de peso neto importado en el periodo estudiado?

Podemos ver el análisis de los datos numéricos del peso en la siguiente imagen



Con estos valores podemos analizar la distribución normal del peso

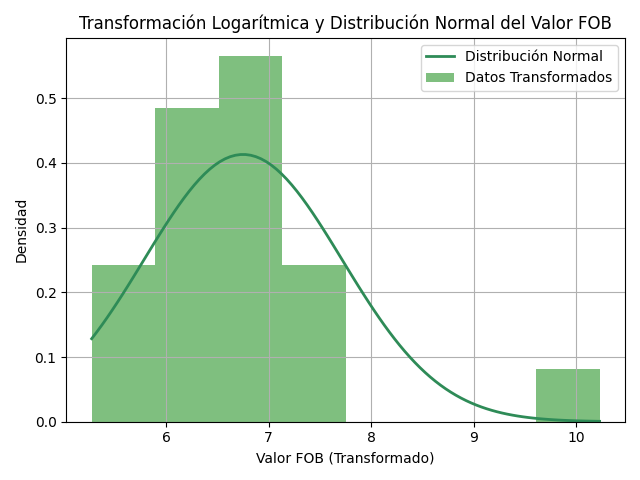


En este diagrama podemos observar como el histograma sigue de manera parecida a la forma de campana que deben tener las distribuciones normales. Por lo que podemos concluir que estos datos pueden seguir una distribución normal, pero sería bueno realizar otros análisis que pudieran confirmar esto.

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

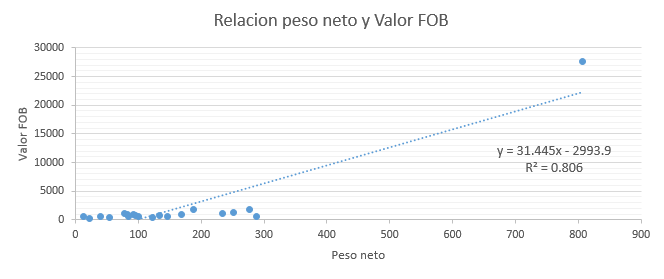


Con estos valores podemos analizar la distribución normal del Valor FOB

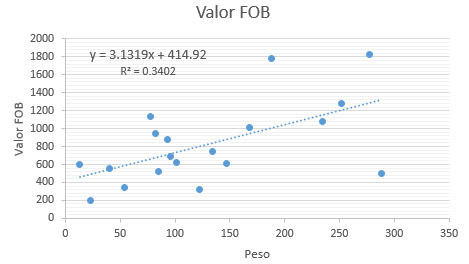


En este diagrama podemos observar como el histograma no sigue la forma característica de campana que poseen las distribuciones normales, por lo que podemos concluir que los datos no siguen esta distribución.

1. ¿Cuál es la correlación entre peso neto y valor FOB?

Para observar esto sacamos la gráfica con los diversos puntos de la muestra

Es complicado realizar el análisis de esta grafica como se nos da, debido a que hay un punto el cual mueve muchísimo la línea de tendencia. Si eliminamos este punto, podemos observar como la regresión lineal cambia por completo:

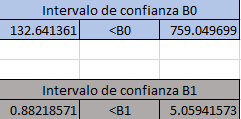


En esta nueva grafica se observa que no existe una relación fuerte entre las variables estudiadas

La regresión lineal nos da la siguiente ecuación para el valor de y:

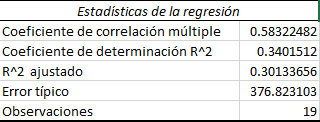
y = 3.1319x + 414.92

Podemos conseguir el intervalo de confianza al 95% para la intercepción y la variable x1



El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos indica que, si tomamos muchas muestras diferentes y calculamos el intervalo de confianza para cada muestra, aproximadamente el 95% de esos intervalos contendrían los verdaderos valores poblacionales de la intercepción y el coeficiente de x1.

En las siguiente imagen podemos observar los datos de la regresión lineal incluyendo el valor de R^2 el cual indica nos confirma que no existe una relación fuerte entre las variables estudiadas, debido a que el valor de R^2 se encuentra muy lejos de 1



La regresión también nos indica su grafico de residuales y el grafico de probabilidad normal, con los que podemos analizar mejor el comportamiento de la regresión.

A continuación formulamos nuestras hipótesis para comprobarlas con análisis a ANOVA y la prueba F

Hipótesis nula (H0):

No hay relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

H0: El coeficiente de la variable "peso" (β1) es igual a cero.

Esto significa que la variable "peso" no tiene un efecto significativo sobre el "valor FOB" y que cualquier variación en el valor FOB no se debe a cambios en el peso.

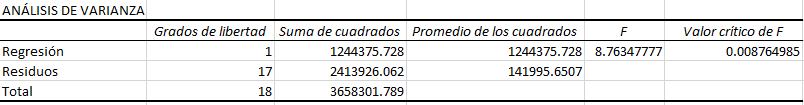
Hipótesis alternativa (H1):

Existe una relación significativa entre el peso y el valor FOB. En términos de coeficientes de regresión, esto se expresa como:

Ha: El coeficiente de la variable "peso" (β1) es diferente de cero.

Esto sugiere que el peso sí tiene un efecto significativo sobre el "valor FOB" y que cambios en el peso están relacionados con variaciones en el valor FOB.

Realizamos la prueba de F para comprobar la hipótesis nula



Con los valores de F y F critico Observamos que el valor de F es mucho más grande que el valor critico de F por lo que podemos rechazar la hipótesis nula de que todas las variables independientes sean irrelevantes y se acepta la hipótesis alternativa de que al menos una de las variables independientes es significativa en el modelo.

Por lo tanto, podemos concluir que hay una relación significativa entre las variables independientes y la variable dependiente en la regresión. En este contexto, el coeficiente de determinación (R²) de 0.3401 también nos indica que alrededor del 34.01% de la variabilidad en la variable dependiente puede ser explicada por las variables independientes en el modelo de regresión.

# Accidentes Conductores base de datos

El estudio que se realizara a los datos de esta base de datos se centrara en hacer una análisis descriptivo de los datos, ya que me pareció muy interesante el tema que manejaba esta base de datos, sin embargo no contenía suficientes columnas de tipo numérico para poder realizar un análisis de regresión y los análisis de los datos obtenidos de ella.

Para realizar este estudio decidí enfocarme en los datos de los el periodo de tiempo entre el 2016 y el 2021. De estos años me enfoque en los datos de las 10 provincias de Panamá.

Descripción de la base de datos

Provincia: Dato categórico

Año: Dato categórico

Alcohol: Dato categórico

Edad: Dato numérico

Al ver la tabla de información que tiene esta base de datos, me surgieron unas preguntas las cuales quería conseguir respuesta con un análisis a esta tabla.

Análisis a la Base de datos

1. ¿Qué provincia encabeza la mayor cantidad de accidentes?

Para dar respuesta a la pregunta realizamos el siguiente análisis:

Sacamos la frecuencia de todos los años desde 2016 hasta 2021, de cada accidente de tráfico en cada provincia y comarca, para saber cuál provincia tiene la mayor cantidad de accidentes.

En estas grafica observamos como la provincia de Panamá poesee el mayor impacto sobre las distintas frecuencias debido a que es la que posee la mayor cantidad de accidentes.

Si graficamos como es el comportamiento de la frecuencia de los accidentes en Panama contra el resto del pais obtenemos lo siguiente:

Debido esto, ara poder apreciar mejor el comportamiento de las demas provincias del pais y obtener una conclusiones más enriquecedoras eliminamos los datos de Panamá para la siguiente grafica

En esta grafica sin los datos de Panamá podemos apreciar de mejor manera como es el comportamiento de los accidentes en las diversas provincias

1. ¿La cantidad total de accidentes por año está aumentando o disminuyendo?

Para responder esta pregunta debemos analizar si la cantidad total de accidentes por año está aumentando o disminuyendo

Para ello graficamos la frecuencia de accidentes por año y obtenemos las siguientes graficas

Grafica de frecuencia de accidentes por año

En esta tabla podemos observar que durante los años 2016 a 2019, hubieron variaciones de ~2000 accidentes por año, por lo que no podemos sacar una conclusión confiable acerca de la tendencia a la subida o bajada. Cada año (excepción de 2020 y 2021) posee alrededor de 20% de la cantidad de accidentes totales en el periodo de 2016 a 2021.

1. ¿Se da la mayoría de accidentes por alcohol?

Para responder esta pregunta debemos analizar cuantos accidentes, de los registrados, el conductor estaba bajo los efectos del alcohol versus los accidentes en que no.



Tabla de frecuencia y frecuencia relativa de los accidentes y el nivel de alcohol en el conductor

En esta tabla podemos observar que la gran mayoría de accidentes no están relacionados con el conductor teniendo alcohol.

Podemos apreciar de mejor manera esto en las siguientes graficas

Donde podemos ver la gran cantidad de accidentes que se dan donde el alcohol no está involucrado.

Debido a esta información podemos concluir que la mayoría de accidentes no se dan por culpa del alcohol.

1. ¿De los accidentes provocados por alcohol, el nivel de alcohol es por lo general alto o bajo?

Para ver esto utilizamos los mismos datos que estudiamos en la pregunta anterior, pero esta vez centrándonos en los accidentes donde el alcohol jugo un papel.

Conseguimos con esto esta gráfica, donde podemos observar que de los accidentes en donde el conductor está bajo los efectos del alcohol, la gran mayoría de estos tienen un muy alto nivel de alcohol en su cuerpo.

1. ¿Cuál es el grupo de edad de conductor, que más accidentes de tráfico tiene?

Para analizar cuál es el grupo de edad más involucrado en accidentes de tráfico debemos dividir las edades en grupos de edad y observamos la frecuencia en la que se encuentran en accidentes.

En esta grafica podemos aprecia que la mayor cantidad de accidentes se da en el grupo de 30-34 años de edad, siendo seguido por el de 25-29 años.

La cantidad de accidentes tiene una tendencia a la baja cuando va aumentando la edad desde los 35 años, hasta que se llega al grupo de personas con 60 años donde nuevamente sube, igualando casi el grupo de edad de 45-49 años.

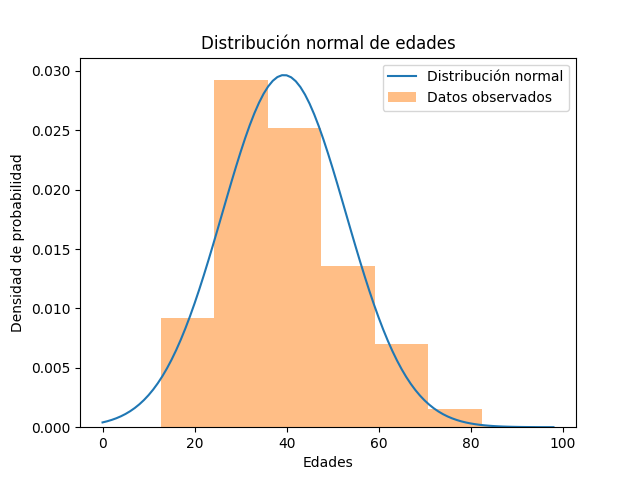
1. ¿Cuál es la media, varianza y desviación estándar de la edad de los conductores?

En el siguiente grafico podemos observar los valores



Podemos ver como la edad media de un conductor involucrado en un accidente son 39 años con una desviación de 13 años

Con estos datos podemos realizar una gráfica de la distribución normal de las edades



En este diagrama podemos observar como la mayoría de datos se encuentran hacia la izquierda del pico de la distribución normal, lo que indica que existe una Asimetría hacia la derecha (sesgo positivo)

# Embarazos base de datos

Se eligió realizar el siguiente análisis debido a que tuve una curiosidad en saber los diversos comportamientos de este arancel en las zonas francas de nuestro país, por lo cual estaremos analizando los datos de ello en esta sección.

Descripción de la tabla

1. Duración del embarazo: Dato numérico
2. Cantidad de nacidos: Dato numérico
3. Sexo: Dato categórico
4. Rango edad de la madre: Dato categórico
5. Edad de la madre: Dato numérico
6. Ocupación de la madre: Dato categórico
7. Escolaridad de la madre: Dato categórico
8. Estado conyugal: Dato categórico
9. Total hijos: Dato numérico

Análisis a la base de datos

1. ¿Qué provincia tiene la mayor cantidad de nacimientos?

Para poder hace este análisis conseguimos la frecuencia de nacimientos en cada provincia durante el periodo de 2016 a 2019

Con este análisis podemos sacar la siguiente gráfica, de frecuencia de nacimientos por provincias

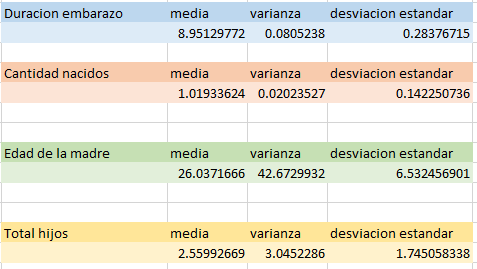
Observando esta grafica podemos ver que la provincia de Panamá es la que más nacimientos ha tenido en el periodo observado

Si decidimos sacar la provincia de Panamá para poder comprar las demás provincias observamos lo siguiente

En esta grafica de 2016 a 2019, podemos observar como el siguiente en la lista de provincias con mayor cantidad de nacimientos es Panamá Oeste, seguido de Chiriquí y después Colón.

1. ¿Cuál es la media, varianza y desviación estándar de los datos numéricos?

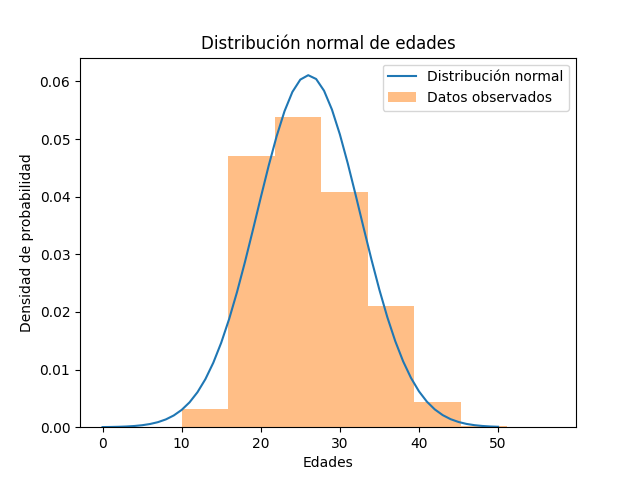
Para poder realizar los cálculos de más adelante estos valores son importantes por los que los sacamos de la muestra completa:



De estos datos podemos observar que la edad media de una madre es de 26 años, da a luz a un niño para tener en total 2 hijos.

1. ¿Cuál es la distribución normal de las edades de las madres?

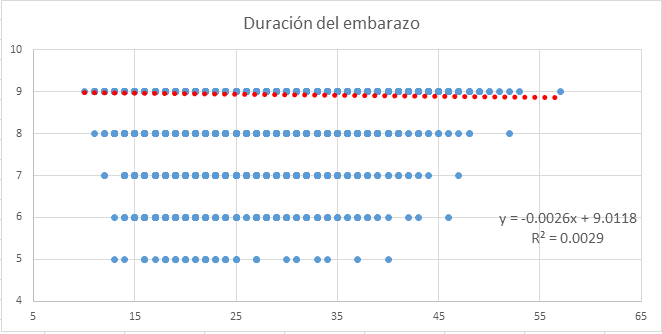
En la siguiente grafica podemos ver el comportamiento de distribución normal de las edades de las madres



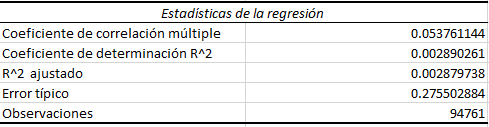
En este diagrama podemos observar como la mayoría de datos se encuentran hacia la izquierda del pico de la distribución normal, lo que indica que existe una Asimetría hacia la derecha (sesgo positivo). Esto se puede explicar debido a que a medida que aumenta la edad de una mujer después de cierto punto, es más complicado que lleguen a tener hijos

1. ¿Qué correlación hay entre la duración de los embarazos y la edad de la madre?

Para responder a esta pregunta realizamos una regresión línea con la edad de la madre en los valores de x y la duración de los embarazos como nuestra variable y.



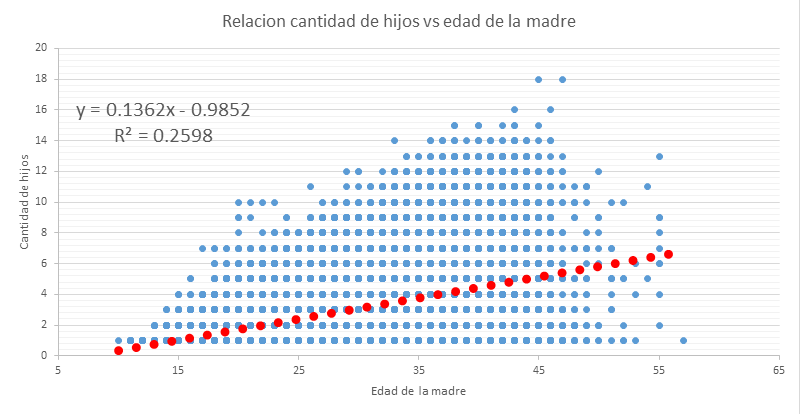
Como se observa en la gráfica, no parece existir a primera vista una relación entre ambas variables. Esto lo podemos confirmar con las siguientes estadísticas de la regresión:



En donde vemos que el valor de R^2 está muy cercano a 0 debido a esto concluimos que no hay una relación causa y efecto entre la edad de las madres y la cantidad de tiempo que dura el embarazo. Por esta falta de relación entre las variables, no existe valor en seguir estudiando los datos de la regresión

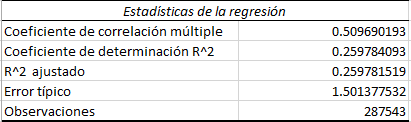
1. ¿Qué correlación hay entre la edad de la madre y la cantidad de hijos que tiene?

Para poder analizar esto conseguimos la regresión lineal de las variables de edad de la madre y la cantidad de hijos. En la siguiente grafica podemos ver el resultado obtenido.



Podemos observar como la línea de tendencia va subiendo, lo que nos indica que por lo general entre mayor la edad de la madre más hijos tiene, sin embargo la dispersión de los datos es mucha lo que nos da a entender que esto no es una relación causa efecto.

Podemos confirmar nuestro análisis de esta regresión con las siguientes estadísticas:



Podemos comprobar como el valor del coeficiente de determinación R^2 es muy bajo lo cual confirma nuestro análisis de que la edad no influye en la cantidad de hijos que tiene una madre. Por esta falta de relación entre las variables, no existe valor en seguir estudiando los datos de la regresión

**Repositorio Git hub**

https://github.com/kopolo435/proyectoEstadistica