

# Análisis estadístico basados en datos de la Federación Nacional de Cafeteros

Ballesteros Natalia, Castillos Maria Fernanda, Silva Santiago

Universidad Nacional de Colombia Sede la Paz

---

## Resumen

Tomando distintos datos brindados por la página de la federación Nacional del café colombiano se nos abre la oportunidad de analizar los de tal manera que se pueda percibir los distintos cambios que han sufrido los precios de dicho producto siendo este el producto de mayor exportación nacional teniendo en cuenta esta información previa procedimos a desglosar los valores contenidos en esta página para luego reacomodar y poder ejecutar y distintos tipos de gráficas que permitiera analizar y llegar a conclusiones de una manera objetiva

*Palabras clave:* Exportaciones, café colombiano, Analisis estadistico

---

## 1. Introducción

En el presente artículo se nos muestra cómo estudiantes de la universidad Nacional aplicando los conocimientos adquiridos desempeñan un estudio estadístico sobre algunos datos tomados de la página de la federación nacional del café colombiano de esta manera se nos narrará el recorrido que tomaron y los distintos procesos aplicados por ellos para poder culminar con conclusiones e influencias basadas en los datos y pruebas realizadas por ellos mismos

## 2. Análisis realizados

- 2.1. *Promedios y desviaciones estándar anuales o por décadas.*
- 2.2. *Gráficas de las variaciones de precio o producción a través del tiempo.*
- 2.3. *Proyecciones de precio o producción utilizando regresión lineal.*
- 2.4. *Gráficas de precios estandarizados a precios de hoy.*
- 2.5. *Pruebas estadísticas para comparar las medias de dos décadas.*

## 3. Materiales y Metodos

### 3.1. Técnicas estadísticas utilizadas

- **Promedio:** Esta es la media aritmética y se calcula agregando un grupo de números y dividiendo por el recuento de esos números.
- **Desviación estándar:** Es una medida que se utiliza para cuantificar la variación o la dispersión de un conjunto de datos numéricos.
- **Inflación:** Es el aumento generalizado y sostenido de los precios de los bienes y servicios existentes en el mercado durante un determinado período de tiempo. Cuando el nivel general de precios sube, con cada unidad de moneda se adquieren menos bienes y servicios.
- **Diagramas de dispersión:** Los gráficos de dispersión se usan para averiguar la intensidad de la relación entre dos variables numéricas. El eje X representa la variable independiente, mientras que el eje Y representa la variable dependiente.
- **Diagramas correlacionales :** Es una herramienta gráfica que permite demostrar la relación existente entre dos clases de datos y cuantificar la intensidad de dicha relación. Se utiliza para conocer si efectivamente existe una conexión entre dos magnitudes o parámetros de un problema y, en caso positivo, de qué tipo es la conexión.
- **Precios estandarizados:** Un producto estandarizado utiliza los mismos materiales independientemente del lugar del mundo en el que se venda, desde las materias primas utilizadas para crearlo hasta la marca, la denominación y el embalaje. Un servicio estandarizado puede prestarse a través del mismo medio o con características similares cada vez.
- **Regresión lineal:** Es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o analizar datos experimentales, financieros y biológicos, trazando un modelo lineal. Las técnicas de

regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente y (también conocida como la respuesta) como una función de una o varias variables independientes  $X_i$  (denominadas predictores).

### 3.2. Herramientas computacionales utilizadas

- **Rstudio:** Es un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráfica. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.
- **R:** Es un entorno de software libre (licencia GNU GLP) y lenguaje de programación interpretado, es decir, ejecuta las instrucciones directamente, sin una compilación previa del programa a instrucciones en lenguaje máquina. El término entorno, en R, se refiere a un sistema totalmente planificado y coherente, en lugar de una acumulación de herramientas específicas e inflexibles, como suele ser el caso en otros softwares de análisis de datos.
- **Excel:** Es una herramienta muy eficaz para obtener información con significado a partir de grandes cantidades de datos. También funciona muy bien con cálculos sencillos y para realizar el seguimiento de casi cualquier tipo de información. La clave para desbloquear todo este potencial es la cuadrícula de las celdas.

## 4. Resultados y discusión

Nuestros análisis parten de la información suministrada por la página “Federación Nacional de cafeteros de Colombia” de la cual se tomaron algunos datos en específico con el fin de poder enfocarnos en cómo el precio y volumen del producto de mayor exportación colombiano habría cambiado con el paso del tiempo, nuestra fuente ya nos entrega estos datos en forma de tabla por lo cual fue cuestión de escoger aquellas que nos permitieran llevar a cabo nuestra misión de la manera más óptima, estas fueron las escogidas para ello:

1. precio areas y producción de cafe
  - a) precio interno mensual
  - b) precio de exportación mensual
2. Exportaciones del café colombiano
  - a) volumen de exportaciones
  - b) valor de exportaciones anual

Empezando esta serie de estudios decidimos tener un muestreo por década que nos permitiera compararlas entre ellas pero las tabla escogidas para esto daban su

información de manera mensual por lo que se tomó la decisión de agrupar 120 meses para conseguir formar décadas dándonos un total de (8) décadas en la tabla de “Precio interno del café” y (7) décadas en la tabla de “Valor total de exportaciones”, luego de ya tener agrupado estos datos se le hallaron a cada década su respectiva media aritmética y desviación estándar dándonos los siguientes resultados :

	decadas	promedio	desviacion_Estandar
1	1	29.37899	8.349605
2	2	53.47479	36.500822
3	3	160.96891	54.732912
4	4	140.42857	45.798570
5	5	112.30587	39.331060
6	6	200.93801	46.445259
7	7	236.12328	46.423835

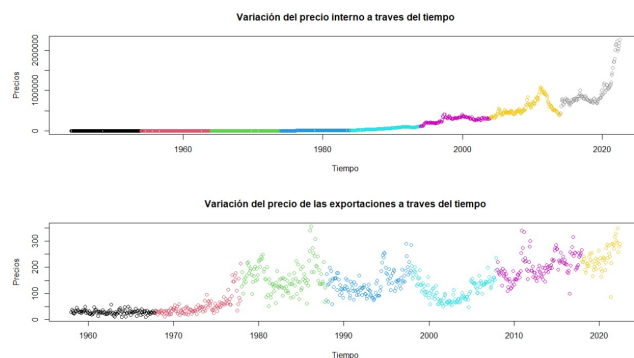
  

	decadas	promedio	desviacion_Estandar
1	1	144.0000	79.97023
2	2	425.4454	83.37710
3	3	1084.8151	394.84322
4	4	7497.4958	3260.91737
5	5	58587.5966	28826.80848
6	6	279784.4715	65276.39664
7	7	576137.7890	186939.36684
8	8	984614.7228	419661.88622

**Figura 1.** Precio interno y Precio de las exportaciones, ordenadas respectivamente .

Luego de analizar los datos arrojados podemos observar las variaciones de los precios internos y de exportaciones a través de las décadas.

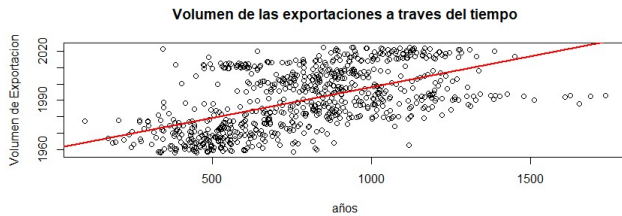
Ya tomando en cuenta los resultados anteriores se llevó a cabo la gráfica del trayecto de estos precios a través del tiempo abarcando los datos mes a mes, las gráficas resultantes se pueden observar a continuación:



**Figura 2.** Proyecciones de precio o producción utilizando regresión lineal..

Otro de los análisis realizados fue una regresión lineal que busca establecer un modelo que permita establecer una relación y predicción del comportamiento del volumen de las exportaciones a través del tiempo, para esta se

filtran 8 décadas partiendo desde 1958-01-01 hasta 2022-06-01 de esta manera se nos es posible contemplar las variaciones del volumen de este producto en momentos específicos y de manera más clara, el resultado de este proceso nos mostró la siguiente gráfica:



**Figura 3.** Proyecciones de precio o producción utilizando regresión lineal..

luego de lo antes planteado se realizó una estandarización de todos los precios contenidos en la tabla de precio de las exportaciones partiendo desde 1920 hasta 2021 usando con punto de guía para esta estandarización el año de 2021 tomando en cuenta la inflación con el pasar del tiempo, al graficar esto se nos permite esperar que precio tendría realmente dicha exportación actualmente tomando en cuenta que estos datos están en dólares americanos, esto se llevó a cabo estandarizando los precios de la lista creando una nueva lista de precios y con esa nueva lista creando la gráfica



**Figura 4.** Gráficas de precios estandarizados a precios de hoy..

con el fin de comparar la información dada por la tabla generada con anterioridad de los precios de las exportaciones estandarizados se decide realizar una prueba estadística que nos permitiera evaluar dos décadas “1991-2000” y “2001-2010”, dicha prueba estadística es conocida

como t-test, esta nos mostró como resultado hipótesis alternativa ya que el p-valor es de 0,001 lo que nos indica que las medias de estas décadas son bastante distintas siendo estas respectivamente de 142.2215 y 116.6764, esto nos deja en evidencia que en la década transcurrida entre “1990-2000” los precios alcanzan mayores picos que los otros

## 5. Conclusiones

### 5.1. Promedios y desviaciones estándar anuales o por décadas.

Como se puede ver en los resultados de la primera tabla por década su promedio se tornó ascendente durante todo su trayecto y su desviación estándar o agrupación de sus datos también se aumentó de forma significativa desde los primeros datos para la segunda tabla los resultados se nos mostraron que en la primera, la segunda y la tercera década los promedios aumentaron mientras que en la cuarta y quinta descendieron para aumentar de manera rápida en la sexta y séptima, mientras que en la desviación estándar de esta tabla por décadas es similar se ve en aumento en el intervalo desde la primera a la tercera y luego en la cuarta y quinta disminuye para finalmente aumentar y estabilizarse entre las sextas y séptimas

### 5.2. Gráficas de las variaciones de precio o producción a través del tiempo.

En las gráficas ya podemos notar como hay un crecimiento poco significativo en las primeras cinco décadas y un crecimiento rápido con algunas variaciones hasta alcanzar los picos más altos al final esto con respecto a la primera tabla en cuanto a la segunda gráfica podemos ver con esta es más desordenada llegando a tener múltiples altibajos en periodos cortos de tiempo siendo sus picos más altos a finales de la tercera y entre la sexta y séptima

### 5.3. Proyecciones de precio o producción utilizando regresión lineal.

Al contemplar la gráfica expulsada en uno de los tests estadísticos realizados se puede notar como dicha gráfica nos revela la relación existente entre dos variables, analizando el cuadro completo se ve como los puntos marcan mayor tendencia entre los quinientos y los mil siendo así se deciden crear una regresión lineal para poder predecir el comportamiento de dichas variables y poder buscar otros datos más específicos de manera puntual ya teniendo esta línea podemos ver la tendencia de dichos valores dándonos como resultado que la sección de tiempo en la que los valores fueron más altos ha sido el área circundante a los mil

#### 5.4. *Gráficas de precios estandarizados a precios de hoy.*

Gráficas de precios estandarizados a precios de hoy, al partir de los años 1920 al 2021 los resultados que nos arrojó la estandarización de la inflación a pasar los años, recordemos que estos precios están valorados en dólares, lo cual nos da la siguiente información; en las primeras dos décadas el precio de las exportaciones con respecto a los años este no se ha visto afectado, con respecto a las siguientes dos décadas el valor de las exportaciones tuvo una devaluación aproximada de 600 en relación al dólar, en las décadas de 1960 a 1980 la inflación decayó pero al finalizar la segunda década el peso se vio devaluado hasta 1000 en relación al dólar, ya para finalizar en las últimas dos décadas de 1980 al 2000 la inflación disminuyó significativamente.

#### 5.5. *Pruebas estadísticas para comparar las medias de dos décadas.*

Con el fin de comparar la información dada por la tabla generada con anterioridad de los precios de las exportaciones estandarizados se decide realizar una prueba estadística que nos permitiera evaluar dos décadas “1991-2000” y “2001-2010”, dicha prueba estadística es conocida como t-test, esta nos mostró como resultado hipótesis alternativa ya que el p-valor es de 0,001 lo que nos indica que las medias de estas décadas son bastante distintas siendo estas respectivamente de 142.2215 y 116.6764, esto nos deja en evidencia que en la década transcurrida entre “1990-2000” los precios alcanzan mayores picos que los otros