Universidad Veracruzana

Facultad de Estadística e Informática

Región Xalapa

Especialización en Métodos Estadísticos

Detección de puntos de cambio en series de tiempo con la prueba CUSUM utilizando R

Práctico educativo
para obtener el diploma de Especialista en
Métodos Estadísticos

Presenta:
Ana Lilia Aguilar Contreras

Director:
Dr. Sergio Francisco Juárez Cerrillo

Septiembre de 2024

"Lis de Veracruz: Arte, Ciencia, Luz"



Universidad Veracruzana

Facultad de Estadística e Informática
Región Xalapa

Especialización en Métodos Estadísticos

Detección de puntos de cambio en series de tiempo con la prueba CUSUM utilizando R

Práctico educativo para obtener el diploma de Especialista en Métodos Estadísticos

Presenta:
Ana Lilia Aguilar Contreras

Director:
Dr. Sergio Francisco Juárez Cerrillo

Codirector:
Dra. Cecilia Cruz López

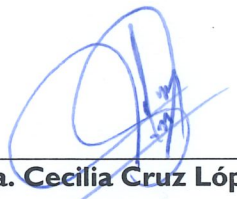
Sinodal:

UNIVERSIDAD VERACRUZANA

Facultad de Estadística e Informática

El Comité Académico de la Especialización en Métodos Estadísticos y el director de este trabajo recepcional intitulado: **Detección de puntos de cambio en series de tiempo con la prueba CUSUM utilizando R**, autorizan la impresión y la constitución del jurado para la defensa.

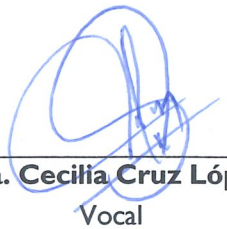
COMITÉ ACADÉMICO



Dra. Cecilia Cruz López
Coordinadora
Especialización en Métodos Estadísticos



Dr. Luis Gerardo Montané Jiménez
Director
Facultad de Estadística e Informática



Dra. Cecilia Cruz López
Vocal



Mtra. María de Lourdes Velasco Vázquez
Vocal



Dr. Sergio Francisco Juárez Cerrillo
Director del Trabajo Recepcional

GENERACIÓN: 2023-2024 **SEDE: Xalapa**

TÍTULO: Detección de puntos de cambio en series de tiempo con la prueba CUSUM utilizando R

AUTOR: Ana Lilia Aguilar Contreras

DIRECTOR Y CO-DIRECTOR: Dr. Sergio Francisco Juárez Cerrillo y Dra. Cecilia Cruz López

TIPO DE TRABAJO:

Reporte de aplicación		Trabajo Practico-Educativo		Desarrollo estadístico		Monografía	X	Artículo o Capítulo de libro	

TIPO DE ESTUDIO:

Exploratorio		Retrospectivo	
Descriptivo	X	Prospectivo	
Expostfaco o cuasiexperimental		Transversal	
Experimental		Longitudinal	

METODOLOGÍA ESTADÍSTICA:

A) Diseño:

Muestreo	
Experimento	
Estudio observacional	

B) Análisis

Exploratorio	
Descriptivo básico	
Inferencia básica	
Métodos multivariados	
Regresión	
ANOVA y ANCOVA	
Control de calidad	
Métodos no paramétricos	
Modelos especiales	
Técnicas avanzadas	

Series de tiempo	X

Dedicatoria

Quiero expresar mi agradecimiento al Conahcyt por la beca otorgada durante el periodo de marzo a agosto del 2024. Su apoyo ha sido fundamental para la culminación de mis estudios.

De la misma manera, agradezco a la Universidad Veracruzana y a la Especialización en Métodos Estadísticos por brindarme la oportunidad de continuar con mi desarrollo académico.

Agradezco profundamente a mis profesores de la Especialización, en especial al Dr. Sergio Francisco Juárez Cerrillo, cuya guía y conocimiento han sido invaluables para mi formación.

También quiero extender mi agradecimiento a todos mis compañeros de generación, quienes siempre estuvieron dispuestos a ayudarme. En especial, a mi compañero Ángel García Báez, quien generosamente compartió su tiempo y sus conocimientos conmigo.

Finalmente, dedico este trabajo a mi familia, cuyo apoyo constante ha sido esencial para alcanzar cada uno de mis logros. Este éxito también les pertenece a ellos.

Índice

INTRODUCCIÓN 4
CAPÍTULO I: Conceptos básicos sobre puntos de cambio 6
1.2 Definición de puntos de cambio 6
1.3 Tipos de punto de cambio 8
CAPÍTULO II: La prueba CUSUM para puntos de cambio 9
CAPÍTULO III: La prueba CUSUM para puntos de cambio en R 13
3.2 Ejemplo de detección de puntos de cambio en la tasa de desocupación en México. 19
3.3 Ejemplo de detección de puntos de cambio en el precio de la gasolina Magna en México. . 21
CAPÍTULO IV: Consideraciones finales 24
Referencias 25
Anexos 27

Resumen

El análisis de series de tiempo es un método cuantitativo que identifica patrones en datos a lo largo del tiempo, permitiendo detectar cambios o conservaciones en información estadística. Este análisis se basa en series de datos indexados en secuencias temporales, y se visualiza comúnmente mediante gráficos de línea. Es crucial detectar alteraciones en la media, varianza o coeficientes de regresión, ya que estas pueden provocar errores en estimaciones y predicciones. Los cambios abruptos en los datos, conocidos como Puntos de Cambio, son significativos para el modelado y pronóstico. La detección de estos puntos es aplicable en diversas áreas, incluyendo el monitoreo de condiciones médicas, cambios climáticos, análisis de voz e imagen, y la actividad humana. En este trabajo se aborda la prueba no paramétrica para detectar puntos de cambio en series temporales, conocida como la prueba CUSUM. Se describe de manera sencilla su aplicación, posterior a esto se muestra su aplicación en el lenguaje de programación R.

Palabras clave: Series de tiempo, detección de puntos de cambio, prueba CUSUM, métodos no paramétricos.

INTRODUCCIÓN

Los datos obtenidos de observaciones registradas a lo largo del tiempo de manera secuencial resultan ser muy comunes. Algunos ejemplos se observan en los negocios, donde se analizan las tasas de intereses semanales, los precios de cierre diarios de las acciones, los índices de precios mensuales. En meteorología se analizan las temperaturas máximas y mínimas diarias, las tasas anuales de lluvia y sequia las velocidades del viento por hora. En agricultura se registran cifras anuales de producción agrícola y ganadera, erosión del suelo y ventas de exportación en un determinado periodo de tiempo. En ciencias biológicas se observa la actividad eléctrica del corazón a intervalos de milisegundos. La lista de áreas en las que se estudian las series de tiempo es casi interminable (Cryer & Chan, 2008).

El análisis de series de tiempo es un método cuantitativo que se utiliza para determinar patrones de comportamiento en los datos recopilados a lo largo del tiempo. El análisis de series de tiempo se utiliza para detectar patrones de cambio o conservación de información estadística en intervalos o periodos regulares. Estos patrones se proyectan para obtener estimaciones para el futuro (Levin y Rubin, 2010).

En matemáticas, una serie de tiempo es una serie de puntos de datos indexados (o calculados o graficados) en una secuencia de tiempo. Generalmente, una serie de tiempo es una serie tomada en puntos consecutivos que están espaciados uniformemente en el tiempo. Por consiguiente, es una secuencia de datos en tiempo discreto. Las series de tiempo se trazan con mayor frecuencia mediante un gráfico de ejecución (que es un gráfico de línea de tiempo) (*Serie de tiempo _ AcademiaLab*, s. f.).

Hay momentos en la serie de tiempo en los que se producen cambios en la media, en la varianza e incluso en los valores de los coeficientes de regresión, por ende, resulta de suma importancia detectar de manera oportuna esta alteraciones a tiempo, ya que son capaces de provocar errores en las estimaciones y en las predicciones de los modelos (Christen, 2022). Estas fluctuaciones abruptas en los datos de series temporales se les conocen como Puntos de Cambio. Estos cambios abruptos pueden representar transiciones entre estados. La detección de puntos de cambio (DPC) es útil para el modelado y pronóstico de series

temporales y tiene áreas de aplicación como monitoreo de condiciones médicas, detección de cambios climáticos, análisis de voz e imagen y análisis de la actividad humana (Aminikhanghahi & Cook, 2016).

La DPC es de gran relevancia y utilidad para estudiantes y profesionales que tengan la necesidad de analizar datos registrados a lo largo del tiempo. El problema presente para las personas interesadas en utilizar las técnicas en DCP es que existe una gran cantidad de información al respecto, lo que puede resultar abrumador para ellas. Principalmente, los estudiantes necesitan una guía que sea práctica y sencilla que les permita aplicar dichas técnicas, utilizando las herramientas que faciliten dicho análisis como el lenguaje de programación R.

Este trabajo tiene como objetivo generar una guía práctica y sencilla que aborda una de las técnicas en el análisis de DPC, conocida como la prueba del Control de Suma Acumulada (CUSUM). En los siguientes capítulos se abordan los conceptos básicos de puntos de cambio, así como la metodología y paquetería en el lenguaje de programación R de la prueba CUSUM , se incluyen también un par de ejemplos donde se utiliza dicha prueba haciendo una interpretación de los resultados y para finalizar se muestran las consideraciones finales del proyecto.

CAPÍTULO I: Conceptos básicos sobre puntos de cambio

En este apartado se abordan los principales conceptos sobre puntos de cambio cuyo conocimiento resulta indispensable para entender la metodología de la prueba CUSUM y su posterior aplicación en R.

1.2 Definición de puntos de cambio

Para Chen y Gupta (2012) Un punto de cambio en una serie de tiempo es el sitio o punto en el tiempo t, en una sucesión de datos $\{x_{ti}\}$ $i = 1,\ldots,n$ observados y ordenados respecto al tiempo tal que dichas observaciones siguen una distribución F_1, antes de un punto, y en otro posterior la distribución a seguir es F_2. En otras palabras, desde el punto de vista estadístico, la sucesión de observaciones muestra un comportamiento no homogéneo. El problema de punto de cambio es considerado como uno de los problemas centrales de inferencia estadística, pues relaciona a la teoría de control estadístico, a las pruebas de hipótesis y a la teoría de estimación. Todo lo anterior bajo el enfoque clásico y Bayesiano según Chen y Gupta (2012).

Chen y Gupta (2012) definen de manera general el problema de puntos de cambio de la siguiente manera:

Tenemos X_1, $X_2,\ldots,X_n$ que son una colección de vectores (variables) aleatorios independientes con funciones de distribución de probabilidad $F_1,F_2,\ldots,F_n$, respectivamente. En esta situación el problema de puntos de cambio consiste en probar la hipótesis nula H_0 de la no existencia de cambio contra la alternativa H_a de que existe al menos un punto de cambio; expresado lo anterior de la siguiente manera:

$$H_0: F_1 = F_2 = \ldots = F_n$$

$$vs$$

$$H_a: F_1 = \ldots = F_{(k1)} \neq F_{(k1+1)} = \ldots = F_{(k2)} \neq F_{(k2+1)} = \ldots = F_{(kq)} \neq F_{(kq+1)} = F_n,$$

Donde $1< k_1 < k_2 < \ldots < k_q < n$, q es el número desconocido de puntos de cambio y k_1, $k_2,\ldots,$ k_q son las posiciones desconocidas respectivas que deben ser estimadas. Si las distribuciones $F_1, F_2,\ldots,F_n$ pertenecen a una familia paramétrica común $F(\theta)$, donde $\theta \in \mathbb{R}^p$, entonces el problema de puntos de cambio reside en probar la hipótesis nula H_0 sobre la no existencia de cambio en los parámetros θ_i, $i = 1,\ldots,n$ de la población contra la alternativa H_a de que existe al menos un punto de cambio; por lo tanto se expresa de la siguiente forma:

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_n = \theta\,(desconocidos)$$

vs

$$H_a : \theta_1 = \ldots = \theta_{(k1)} \neq \theta_{(k1+1)} = \ldots = \theta_{(k2)} \neq \theta_{(k2+1)} = \ldots = \theta_{(kq)} \neq \theta_{(kq+1)} = \theta_n,$$

Donde q y $k_1, k_2, \ldots, k_q$ deben ser estimados, Estas hipótesis demuestran los aspectos de inferencia de puntos de cambio para concluir si existe algún punto de cambio en el proceso, estimar el número de ellos y sus respectivas posiciones. En diversos casos se acepta que las observaciones son independientes e idénticamente distribuidas (i.i.d.), el análisis resulta más complejo si se presenta dependencia entre las observaciones (Chen y Gupta, 2012).

De manera simple se define que un punto de cambio es el punto en el que cambian las propiedades estadísticas de una secuencia de observaciones. La detección de puntos de cambio es el nombre con el que se le conoce al problema de estimar los puntos donde ocurre dichos cambios (Killick & Eckley, 2014).

Para Schroth et al. (2024) los puntos de cambio normalmente se describen como cambios entre segmentos. Es decir, un punto de cambio divide una serie de tiempo en dos segmentos, cada uno con sus propias características estadísticas, ya sea la media o la varianza por dar un ejemplo. Por lo tanto, el punto de cambio es donde las características fundamentales cambian dramáticamente. A continuación se mencionan los tipos de punto de cambio más importantes.

1.3 Tipos de punto de cambio

Cambio en la media.

Para Schroth et al. (2024) El cambio en la media es el más común, probablemente el más fácil de identificar de manera visual. El cambio en la media generalmente ocurre cuando una serie de tiempo se pude dividir en diferentes segmentos constantes que tienen diferentes valores medios. Uno de los primeros algoritmos para detectar tales cambios es el algoritmo de CUSUM, que fue desarrollado para detectar cambios en la media. Se aplicó para el control de calidad en la fabricación.

Cambio en la varianza.

El cambio en la varianza es otro tipo de cambio común. Aquí, el valor medio de la señal permanece constante, pero hay varios segmentos con diferentes valores de varianza. Esto puede interpretarse como un ruido repentino en la señal. Tanto el cambio en la media como el cambio en la varianza pueden ser detectados comparando las propiedades estadísticas a través de la señal (Schroth et al., 2024)

Cambio en la frecuencia.

El cambio en la frecuencia (También conocido como cambio en la ciudad periodística) se refiere a una serie de tiempo con propiedades cíclicas. En este caso, el cambio se produce cuando la frecuencia cambia repentinamente. La detección de este tipo de cambio se realiza generalmente en el dominio de la frecuencia, por ejemplo, mediante el uso de la transformada de Fourier o la transformada de ondículas (Schroth et al., 2024).

Cambio en el patrón.

El cambio de patrón es más difícil de abordar que los anteriores. Una forma de detectarlos es utilizar las distancias de Wasserstein entre distribuciones empíricas. En este punto, podemos ver que la detección de puntos de cambio está estrechamente relacionada con la detección de anomalías; La diferencia entre las dos tareas es a veces difusa (Schroth et al., 2024).

CAPÍTULO II: La prueba CUSUM para puntos de cambio

La prueba CUSUM para detectar puntos de cambio en series de tiempo se considera un método no paramétrico, así lo especifican Brosdsky y Darkhovsky (1993) y Carbonetti y Gende (2018). Dado que una prueba no paramétrica es una prueba de hipótesis que no requiere que la distribución de la población sea caracterizada por ciertos parámetros, por ejemplo, que partan del supuesto de que la población sigue una distribución normal con parámetros μ y σ, son útiles en los casos donde los datos son considerablemente no normales y resistentes a transformaciones (*Explicación de los Métodos No Paramétricos - Minitab*, s. f.).

La prueba CUSUM como su nombre lo indica, se basa en el concepto de sumas acumulativas, esta prueba tiene la capacidad de detectar pequeñas desviaciones en la media de la muestra. Muñiz (2021), define el siguiente procedimento para la elaboración del diagrama CUSUM. Es necesario calcular y graficar una suma acumulada de los datos. Sea $X_1, X_2, \ldots, X_N$ el conjunto de N datos. A partir de estos datos se calculan las sumas acumuladas $S_0, S_1, \ldots, S_i$ de la siguiente manera:

Paso 1. Calcular el valor medio de la muestra de observaciones fija.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N},$$

Donde N representa el número total de valores de la muestra.

Paso 2. Determinar la suma acumulada

Se comienza la suma acumulada estableciendo $S_0 = 0$, se calculan las otras sumas utilizando la siguiente fórmula:

$$S_i = S_{i-1} + (X_i - \bar{X}) \quad para \quad i = 1,2,\ldots N.,$$

Donde los X_i son los valores correspondientes de la muestra.Con la fórmula anterior podemos expresar la suma acumulativa de los datos de la siguiente manera

$$S_0 = 0\ S_1 = S_0 + (X_1 - \bar{X})\ S_2 = S_1 + (X_2 - \bar{X})$$

Continuando los cálculos de esta manera para todos los valores de la muestra.

Paso 3. Realizar el gráfico de sumas acumulativas de los S_i con respecto a los periodos de las observaciones obtenidas. Es importante tener claro que las sumas acumuladas representadas en el gráfico no son las sumas acumuladas de los valores, son las sumas acumuladas de las diferencias entre los valores y la media; por lo tanto estas diferencias siempre suman cero, por lo que se establece que $S_N = 0$.

Para la interpretación del gráfico CUSUM se toman las siguientes consideraciones: Un segmento con una pendiente ascendente indica un periodo en el cual los valores tienden a encontrarse por encima de la media general. Caso contrario, un segmento con una pendiente descendente indica un periodo de tiempo en el cual los valores están por debajo de la media general. Un cambio repentino en la dirección de las sumas acumulativas indica que existe un cambio en la media y si dicha dirección durante un periodo de tiempo se mantiene relativamente recta se define que no existe un cambio en la media.

Con las consideraciones anteriores se puede detectar de manera visual si existe un cambio en la media y en consecuencia si existe un punto de cambio en la serie de tiempo.

Paso 4. Estimación del punto de cambio. Posterior a detectar un cambio en el momento de realizar el análisis visual del CUSUM se determina el lugar donde ocurre dicho cambio utilizando el siguiente estimador. Sea m tal que,

$$|S_m| = \max_{i=0,1,\dots,N} |S_i|,$$

Donde S_m representa el punto más alejado del cero en el gráfico CUSUM y estima el último punto anterior al cambio, el punto posterior al cambio queda estimado por $m + 1$.

Paso 5. Estimar el nivel de confianza .Para determinar un nivel de confianza para el punto de cambio estimado se utiliza un análisis bootstrap. Este tipo de análisis es un método que estima la distribución de muestreo al seleccionar múltiples muestras con reemplazo de

una sola muestra aleatoria. Estas nuevas muestras se llaman muestras repetidas. Cada muestra es del mismo tamaño que la muestra original (¿Qué Es un Procedimiento Bootstrap? - Minitab, s. f.).

Para realizar el análisis bootstrap es necesario un estimador de la magnitud del cambio definido de la siguiente manera:

$$S_{diff} = S_{max} - S_{min}$$

Donde $S_{max} = \underset{i=0,1,\dots,N}{max} S_i$, $S_{min} = \underset{i=0,1,\dots,N}{min} S_i$

El estimador S_{diff} se pude utilizar independientemente del tipo de distribución que se tenga y una vez estimado se procede a realizar el análisis bootstrap, que reside en elaborar un gran número de bootstraps conocidos como individuales, un bootstrap individual se realiza de la siguiente manera:

1. Se elabora una muestra bootstrap de N unidades, dicha muestra queda denotada como $X_1^0, x_2^0, \dots, x_N^0$, reordenando de manera aleatoria los N valores originales.Esto se conoce como muestreo sin remplazo.

2. Tomando como base la muestra bootstrap, se calcula la suma acumulativa bootstrap, denotada como $S_0^0, S_1^0, \dots, S_N^0$.

3. Se calcula el valor máximo, mínimo y la diferencia de la suma acumulativa bootstrap, denotados como S_{max}^0, S_{min}^0, y S_{diff}^0.

4. Se determina si la diferencia S_{diff}^0 es menor que S_{diff}.

Una vez realizado el conteo se utiliza la siguiente fórmula para estimar el nivel de significancia del análisis de punto de cambio:

$$\text{Nivel de significancia} = \left(\frac{NX}{100}\right)\%,$$

Donde N es el número de muestras bootstrap individuales realizadas y X es el número de bootstrap individuales para los cuales $S_{diff}^{0} < S_{diff}$. Generalmente para afirmar que se tiene un punto de cambio se requiere un nivel de confianza de un 90% o un 95%.

Paso 6. Detección de múltiples puntos de cambio.

Para detectar si se han producido más puntos de cambio en la serie de tiempo se dividen los datos en dos segmentos, tomando como base de la segmentación el punto de cambio que ya se ha detectado, en un segmento se encontrarán los datos anteriores y en el otro segmento los datos posteriores a dicho punto de cambio. Se realiza de nuevo el CUSUM para cada segmento y de ser detectado otro punto de cambio significativo en algún segmento de datos se repite de nuevo la actividad de segmentación y prueba CUSUM.

CAPÍTULO III: La prueba CUSUM para puntos de cambio en R

La paquetería utilizada para la detección de puntos de cambio en este trabajo es "changepoint" (Killick & Eckley, 2014), ya que permite identificar cambios en la media y de manera adicional a otros paquetes existente en R tales como strucchange y cpm, los usuarios tienen la capacidad de seleccionar un método de búsqueda para múltiples puntos de cambio para su análisis. Este paquete contiene tres métodos: PELT, "BinSeg" y "SegNeigh" (Killick & Eckley, 2014).

En el paquete changepoint todos los métodos para detectar cambios en la media utilizan la función cpt.mean. Dicha función así como sus argumentos son descritos por Killick y Eckley (2014) de la siguiente manera:

Estructura de la función:

cpt.mean (data, penalty = "SIC", pen.value = 0, method = "AMOC", Q = 5, test.stat = "Normal", class = TRUE, param.estimates = TRUE)

Argumentos:

Data- Un vector que contiene los datos a los cuales se les busca un cambio en la media.

Penalty- Elección de penalizaciones "None", "SIC" (Criterio de información de Schwarz), "BIC" (Criterio de información bayesiano), "AIC" (Criterio de información akaike), "Asymptotic" (Penalización asintótica) y "Manual". La penalicación "Manual" se encuentra contenida en pen.value. Para la penaliación "Asymptotic", el error tipo 1 teórico, se encuentra contenido en pen.value.Las penalizaciones predefinidas enumeradas no cuentan el punto de cambio como un parámetro, para que el punto de cambio cuente como un parámetro se debe agregar un 1 como sufijo, ejemplo, "SIC1".

pen.value- Es el error tipo 1 teórico. Para la penalización "Asymptotic" es de 0.05. Para la penalización "Manual" es un valor númerico o texto que cuando se evalua da como resultado un valor de penalización.

Method- Método de cambio de punto único o múltiple. El valor predeterminado es

"AMOC (detecta un cambio como máximo), se puede seleccionar "PELT", "SegNeigh" o "BinSeg".

Q- Para el método "BinSeg" este argumento representa el número máximo de puntos de cambio para buscar. Para el método "SegNeigh" este argumento representa el número máximo de segmentos (número de puntos de cambio +1) a buscar. Para el método "PELT" no es necesario establecer ya que selecciona automáticamente el número de segmentos.

Test.stat- Es el estadístico de prueba, dicho de otra manera, es la distribución asumida o el método independiente de la distribución de los datos. Las opciones son "Normal" o "CUSUM".

Class- Lógico. Si es TRUE, se devuelve un objeto clase 'cpt' param.estimates- Lógico. Si es TRUE y class=TRUE, se devuelven las estimaciones de los parámetros. Si es FALSE y class=FALSE, no se devuelven estimaciones de los parámetros.

Para utilizar la función cpt.mean es necesario istalar en R la paqueteria "changepoint" y llamarla de la siguiente manera:

```
#install.packages("changepoint")
library(changepoint)
```

A continuación se genera un vector llamado x con una serie de datos que siguen una distribución normal para realizar un ejemplo de la función cpt.mean y detectar los puntos de cambio en los datos.

```
set.seed(1)
x=c(rnorm(100,0,1),rnorm(100,10,1), rnorm(100,20,1))
```

Posteriormente se utiliza la función cpt.mean estableciendo los argumentos de la siguiente manera:

```
ans=cpt.mean(x,penalty="Manual",pen.value=0.05,method="AMOC",test.stat="CUSUM")

print(ans)
```

Dicha función está contenida en el objeto ans. Como se puede observar en la función, se utilizó una penalización “Manual” asignando un valor de 0.05 en pen.value, el método seleccionado fue “AMOC”, y se asignó para test.stat la opción “CUSUM”. Al solicitarle a R mediante la función print el resultado de aplicar la función cpt.mean, contenido en el objeto ans se obtiene lo siguiente:

```
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## Changepoint type      : Change in mean
## Method of analysis    : AMOC
## Test Statistic  : CUSUM
## Type of penalty       : Manual with value, 0.05
## Minimum Segment Length :
## Maximum no. of cpts   : 1
## Changepoint Locations : 160
```

Al utilizar el método “AMOC” R localiza solo un punto de cambio localizado en la observación 160. Utilizando la fución plot se visualiza en la figura 1 el punto de cambio detectado de los datos analizados.

```
plot(ans)
```

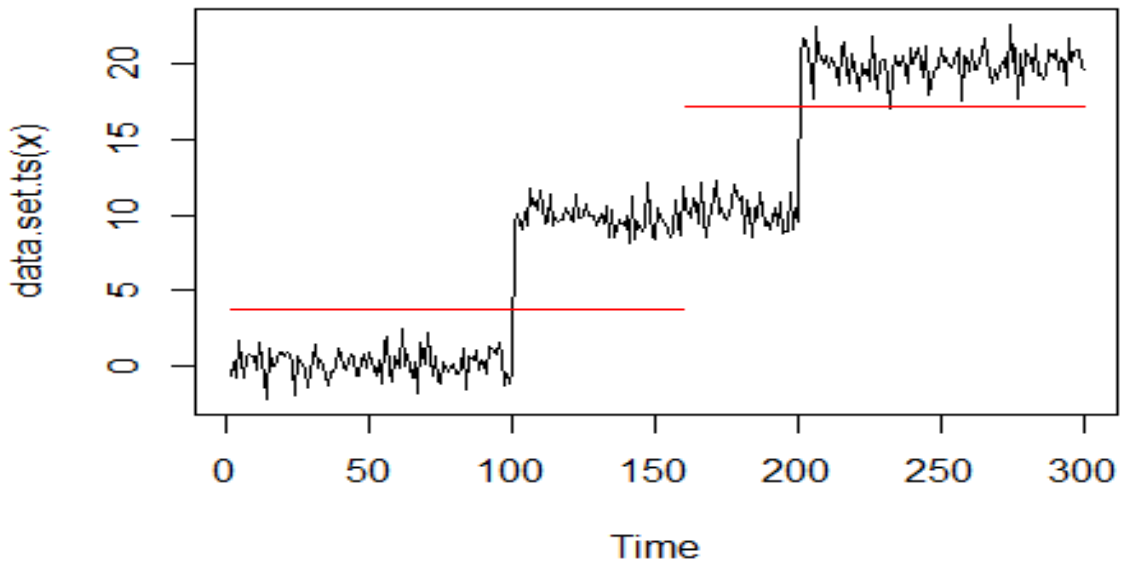


Figura 1. Gráfica donde se especifica el punto de cambio de los datos generados en R. Fuente: Generación propia.

En la figura 1 se observa el punto de cambio localizado en la observación 160. El punto de cambio se indica mediante las líneas rojas horizontales que representan el valor medio en diferentes segmentos

```
ans=cpt.mean(x,penalty= "Manual",pen.value=0.05,method = "BinSeg", Q=5, t
est.stat = "CUSUM")
```

Utilizando la función print se visualiza el siguiente resultado:

```
print(ans)

## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 14 slots with names
##               cpts.full pen.value.full data.set cpttype method test.st
at pen.type pen.value minseglen cpts ncpts.max param.est date version
##
```

```
## Created on  : Thu Feb 29 03:43:10 2024
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## Changepoint type      : Change in mean
## Method of analysis    : BinSeg
## Test Statistic  : CUSUM
## Type of penalty       : Manual with value, 0.05
## Minimum Segment Length : 1
## Maximum no. of cpts   : 5
## Changepoint Locations : 100 133 159 180 200
## Range of segmentations:
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  159   NA   NA   NA   NA
## [2,]  159  100   NA   NA   NA
## [3,]  159  100  200   NA   NA
## [4,]  159  100  200  180   NA
## [5,]  159  100  200  180  133
##
##  For penalty values: 3.354865 2.260857 2.027006 0.177563 0.1514268
```

Como se ha mencionado en este ejemplo se asignó el valor de Q igual a 5, porque de esta manera R muestra en una matriz el orden en el que fue detectando los puntos de cambio, si el valor de Q es mayor que 5, R ya no muestra esta matriz, solo indica los puntos de cambio localizados.

Gracias a dicha matriz se observa que el primer punto de cambio que se detectó está localizado en la observación 159, el siguiente en la observación 100 y así, de manera consecutiva, la función localizó los puntos de cambio en las observaciones 200, 180 y 133. A continuación, utilizando la función plot, se puede apreciar la gráfica para este ejemplo en la figura 2.

```
plot(ans)
```

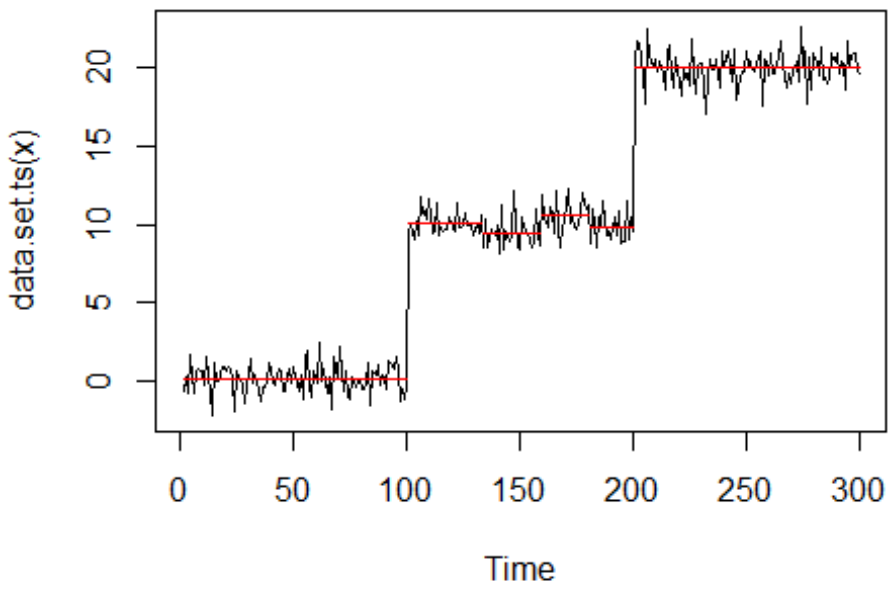


Figura 2. Gráfica donde se especifica los puntos de cambio detectados de los datos generados en R. Fuente: Generación propia.

En la figura 2, Se observan unas líneas horizontales rojas, éstas nos indican los periodos de tiempo en los cuales no se detectaron cambios en la media de las observaciones. El número de observación o el punto en el cual se suspende la continuidad de la línea roja es justo el punto de cambio detectado.

En el gráfico se puede apreciar que en los puntos de cambio detectados en las observaciones 200 y 100, los cambios en las medias son más pronunciados que en el resto de puntos de cambio detectados en las observaciones 133,159 y 180.

3.2 Ejemplo de detección de puntos de cambio en la tasa de desocupación en México.

Utilizando los datos obtenidos del Instituto Nacional de Estadística y Geografía sobre la tasa de desocupación (De Estadística y Geografía, s. f.), que van de enero del 2006 a enero del 2024, con observaciones mensuales, se realizó la detección de puntos de cambio utilizando la función cpt.mean de R. La variable continua tasa de desocupación está expresada en porcentaje. Los argumentos de la función cpt.mean para este ejemplo quedaron definidos de la siguiente manera:

```
diagrama=cpt.mean(tasa$Tasa_desocupa, penalty = "Manual", pen.value = 0.0
5, method ="BinSeg" , Q = 5,
test.stat = "CUSUM", class = TRUE, param.estimates = TRUE)
```

Utilizando la función print y plot respectivamente, se obtiene el resumen de la aplicación de la función cpt.mean y la gráfica respectiva, donde se visualizan los puntos de cambio detectados.

```
print(diagrama)
```

```
## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 14 slots with names
##               cpts.full pen.value.full data.set cpttype method test.st
at pen.type pen.value minseglen cpts ncpts.max param.est date version
##
## Created on  : Thu Feb 29 03:43:10 2024
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## Changepoint type      : Change in mean
## Method of analysis    : BinSeg
## Test Statistic  : CUSUM
## Type of penalty       : Manual with value, 0.05
## Minimum Segment Length : 1
## Maximum no. of cpts   : 5
## Changepoint Locations : 36 122 171 184 194
```

```
## Range of segmentations:
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  122   NA   NA   NA   NA
## [2,]  122   36   NA   NA   NA
## [3,]  122   36  194   NA   NA
## [4,]  122   36  194  171   NA
## [5,]  122   36  194  171  184
##
##  For penalty values: 0.2496501 0.2496501 0.1463878 0.1463878 0.1463878

plot(diagrama)
```

En la figura 3, se observa la gráfica que establece los puntos de cambio detectados en los datos de tasa de desocupación.

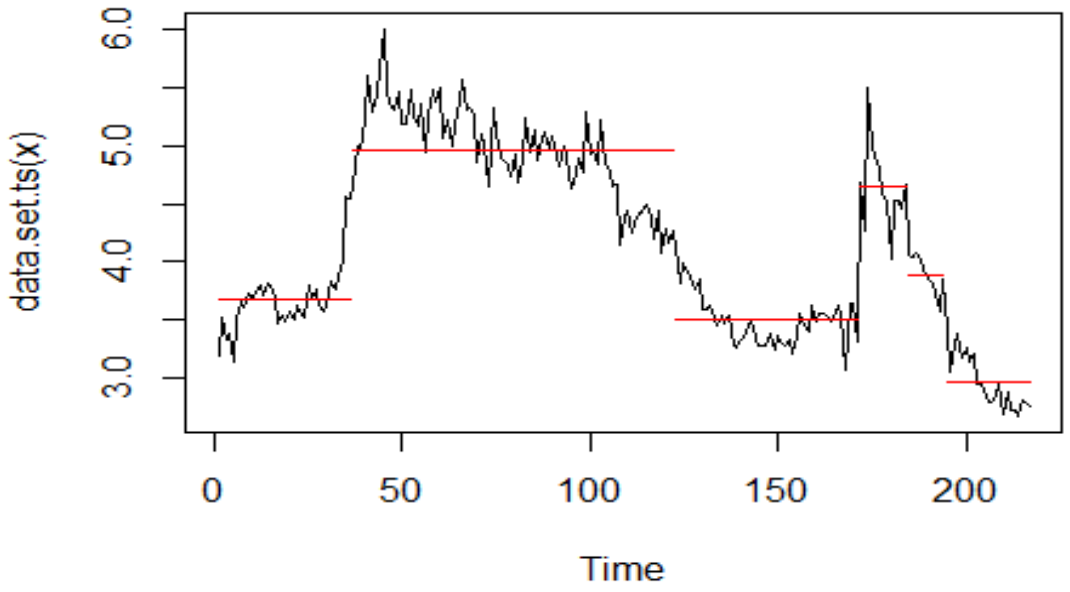


Figura 3. Gráfica de detección de puntos de cambio en la tasa de desocupación en México (Enero 2006-Enero 2024) Fuente: Elaboración propia con datos del INEGI (De Estadística y Geografía, s. f.)

En el eje horizontal se muestran las observaciones realizadas a través del tiempo y en el eje vertical los valores para la tasa de desempleo. Los puntos de cambio detectados en R

son los siguientes: 36, 122, 171, 184 y194, estos puntos indican las siguientes fechas abreviadas de manera correspondiente: nov/2011, ene/2019, mar/2020, abr/2021 y feb/2023.

A partir de noviembre del 2011 se observa un cambio brusco en la media de la tasa de desocupación, los valores de dicha variable sufren un incremento pronunciado a través del tiempo, llegando a un segmento donde no se detectan cambios bruscos en la media hasta enero del 2019, debido a una clara y pronunciada reducción de los valores registrados, se observa después del punto de cambio un segmento en el cual nuevamente la media no presenta cambios bruscos hasta marzo del 2020, fecha de inicio de la pandemia, lo cual provoca un aumento pronunciado en los valores de la tasa de desocupación, sin embargo, se vuelven a detectar otros dos puntos de cambio en abril del 2021 y febrero del 2024, donde se aprecia de manera clara una disminución en los valores de la tasa de desocupación.

Los puntos de cambio pueden indicar diferentes acciones o políticas públicas llevadas a cabo en diferentes sexenios, que generan cambios bruscos en la media a través del tiempo de la variable tasa de desocupación. A su vez, en este análisis se observa un punto de cambio específico generado por el inicio de la pandemia y las acciones que se llevaron a cabo en ese período de tiempo.

3.3 Ejemplo de detección de puntos de cambio en el precio de la gasolina Magna en México.

Utilizando los datos del precio de la gasolina Magna de la Procuraduría Federal del consumidor, mostrados en la Conferencia matutina del Presidente de la Republica del día viernes 12 de mayo del 2023 (Gobierno de México, 2023) se realizó la detección de puntos de cambio utilizando la función cpt.mean de R. La variable continua precio está especificada en pesos mexicanos. Las observaciones son anuales abarcando el periodo 2006-2023.

Los argumentos utilizados en la función cpt.mean quedaron determinados de la siguiente manera:

```
diagrama=cpt.mean(precio$precio, penalty = "Manual", pen.value = 0.05, me
thod ="BinSeg" , Q = 5,
test.stat = "CUSUM", class = TRUE, param.estimates = TRUE)
```

Utilizando la función print y plot respectivamente se obtiene el resumen de la aplicación de la función cpt.mean y la gráfica respectiva donde se visualizan los puntos de cambio detectados.

```
print(diagrama)
```

```
## Class 'cpt' : Changepoint Object
##        ~~   : S4 class containing 14 slots with names
##               cpts.full pen.value.full data.set cpttype method test.st
at pen.type pen.value minseglen cpts ncpts.max param.est date version
##
## Created on  : Thu Feb 29 03:43:10 2024
##
## summary(.)  :
## ----------
## Created Using changepoint version 2.2.4
## Changepoint type      : Change in mean
## Method of analysis    : BinSeg
## Test Statistic  : CUSUM
## Type of penalty       : Manual with value, 0.05
## Minimum Segment Length : 1
## Maximum no. of cpts   : 5
## Changepoint Locations : 5 7 8 11 12
## Range of segmentations:
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    8   NA   NA   NA   NA
## [2,]    8   12   NA   NA   NA
## [3,]    8   12    5   NA   NA
## [4,]    8   12    5    7   NA
## [5,]    8   12    5    7   11
##
##  For penalty values: 1.655802 0.7334 0.641875 0.4077778 0.393125
```

```
plot(diagrama)
```

En la figura 4 se observa la gráfica de los puntos de cambio detectados en los datos del precio de la gasolina Magna. En el eje horizontal se especifican las observaciones a través del tiempo, en este caso los años. En el eje vertical se especifica el precio de la gasolina.

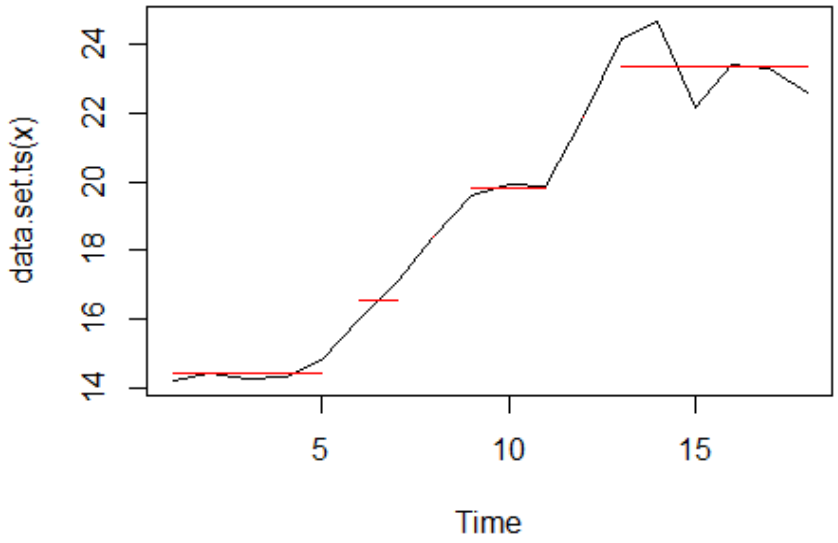


Figura 4. Gráfica de detección de puntos de cambio en el precio de la gasolina Magna en México del año 2006 al 2023 Fuente: Elaboración propia con datos de la PROFECO (Gobierno de México, 2023)

Los puntos de cambio detectados en R son los siguientes: 5, 7, 8, 11 y 12 y representan de manera correspondiente los años: 2010, 2012, 2013, 2016, 2017. Se puede observar en el gráfico como se detectaron cuatro cambios bruscos en la media producidos por un aumento en los valores del precio de la gasolina que van del año 2010 hasta el año 2017, a partir de esa fecha se puede notar un segmento en el cual no se presentan cambios bruscos en la media del precio de la gasolina a través del tiempo.

CAPÍTULO IV: Consideraciones finales

Como se muestra en el capítulo anterior, la prueba CUSUM para detectar puntos de cambio en series de tiempo con el lenguaje de programación R es relativamente sencilla, al ser una prueba no paramétrica no requiere del supuesto de que los datos provienen de una distribución normal. En el capítulo II, donde se hace la descripción de la prueba CUSUM, se explica cómo se puede obtener un nivel de significancia de la prueba, gracias a la aplicación de la técnica Bootstrap. En la función mpt.mean de R no se obtiene un p-valor que permita realizar inferencias sobre el resultado de la prueba, únicamente queda como un resultado descriptivo de la muestra de datos analizados.

La detección de puntos de cambio en series temporales es de gran utilidad en diferentes áreas del desarrollo humano. Se puede aplicar para observar si determinadas acciones destinadas a generar un cambio en una variable están o no teniendo resultado; por ejemplo, detectar si las acciones de Eficiencia Energética implementadas en determinados edificios están generando una disminución en el consumo de energía eléctrica. En este trabajo se desarrolló la explicación de la prueba y su aplicación en R con el objetivo de generar un documento de consulta para realizar proyectos que requieran detectar puntos de cambio en series temporales.

Para investigaciones posteriores se considera conveniente determinar diferentes funciones en el lenguaje de programación R, que permitan realizar estadística inferencial como complemento de la función cpt.mean.

Referencias

Aminikhanghahi, S., & Cook, D. J. (2016). A survey of methods for time series change point detection. *Knowledge And Information Systems*, *51*(2), 339-367. https://doi.org/10.1007/s10115-016-0987-z

Carbonetti, M., & Gende, M. A. (2018). Implentación y evaluación de métodos no paramétricos parra detectar variaciones bruscas en series de tiempo GNSS. Dialnet. https://dialnet.unirioja.es/servlet/articulo?codigo=8915955

Chen, J., & Gupta, A. K. (2012). Parametric Statistical Change point analysis. En *Birkhäuser Boston eBooks*. https://doi.org/10.1007/978-0-8176-4801-5

Christen, A. (2022, 11 enero). *Detección de puntos de cambio en series de tiempo utilizando análisis Wavelet*. https://repositoriobibliotecas.uv.cl/handle/uvscl/3952#:~:text=Es%20de%20mucha%20importancia%20detectar%20r%C3%A1pidamente%20estas%20alteraciones,y%2C%20posteriormente%2C%20en%20las%20predicciones%20de%20los%20modelos

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis whit Applications in R* (Second Edition). Springer. http://repository.cinec.edu/bitstream/cinec20/1228/1/2008_Book_TimeSeriesAnalysis.pdf

De Estadística y Geografía, I. N. (s. f.). *Empleo y ocupación*. https://www.inegi.org.mx/temas/empleo/

Explicación de los métodos no paramétricos - Minitab. (s. f.). (C) Minitab, LLC. All Rights Reserved. 2024. https://support.minitab.com/es-mx/minitab/help-and-how-to/statistics/nonparametrics/supporting-topics/understanding-nonparametric-methods/

Gobierno de México. (2023, 12 mayo). *#ConferenciaPresidente | Viernes 12 de mayo de 2023.* [Vídeo]. YouTube. https://www.youtube.com/watch?v=tsSz0lhWszw

Killick, R., & Eckley, I. A. (2014). changepoint: AnRPackage for Changepoint Analysis. *Journal Of Statistical Software*, *58*(3). https://doi.org/10.18637/jss.v058.i03

Levin, R. & Rubin, D. (2010). *Estadística para administración y economía* (Septima edición) PEARSON

Muñiz, L. (2021) Implementación de algunos procedimientos computacionales en aplicaciones para la detección de puntos de cambio. Retrieved from: https://hdl.handle.net/20.500.12371/11680.

Nonparametric Methods in Change Point Problems. (s. f.). Google Books. https://books.google.com.mx/books?hl=es&lr=&id=GLvwCAAAQBAJ&oi=fnd&pg=PR7&ots=IMqutP1KOn&sig=pVaH03JOw-Zp4yneXH0qZ6VnMjY&redir_esc=y#v=onepage&q&f=false

¿Qué es un procedimiento bootstrap? - Minitab. (s. f.). (C) Minitab, LLC. All Rights Reserved. 2024. https://support.minitab.com/es-mx/minitab/help-and-how-to/probability-distributions-random-data-and-resampling-analyses/supporting-topics/resampling-analyses/what-is-bootstrapping

Schroth, C., Siebert, J., & Groß, J. (2024, 12 enero). *Time Traveling with Data Science: Focusing on change point detection in time series Analysis (Part 2)*. Fraunhofer IESE. https://www.iese.fraunhofer.de/blog/change-point-detection/

Serie de tiempo _ AcademiaLab. (s. f.). https://academia-lab.com/enciclopedia/serie-de-tiempo/

Anexos

Instituto Nacional de Estadística y Geografía (s. f.) Tasa de desocupación en México con datos de enero del 2006 a enero del 2024, observaciones mensuales.

Fecha	Tasa desocupación
2006/01	3.195447542
2006/02	3.521331407
2006/03	3.336206241
2006/04	3.383200428
2006/05	3.146090554
2006/06	3.509737887
2006/07	3.664939067
2006/08	3.615803716
2006/09	3.727749825
2006/10	3.676534413
2006/11	3.747474094
2006/12	3.790673635
2007/01	3.720183503
2007/02	3.814952359
2007/03	3.801973326
2007/04	3.691687979
2007/05	3.466774153
2007/06	3.544613972
2007/07	3.485133813
2007/08	3.568339669
2007/09	3.508952104
2007/10	3.625232573
2007/11	3.563920647
2007/12	3.526503114
2008/01	3.804524663
2008/02	3.677206984
2008/03	3.765026469
2008/04	3.624437713
2008/05	3.57823273
2008/06	3.620361743
2008/07	3.838202718
2008/08	3.761198037
2008/09	3.86334705
2008/10	4.012094801
2008/11	4.555848472

Fecha	Tasa desocupación
2008/12	4.54522591
2009/01	4.653561775
2009/02	4.99610555
2009/03	4.963169727
2009/04	5.218344653
2009/05	5.596860076
2009/06	5.298439156
2009/07	5.451635049
2009/08	5.674862739
2009/09	6.004866026
2009/10	5.471014061
2009/11	5.318638629
2009/12	5.30670713
2010/01	5.460958158
2010/02	5.194857318
2010/03	5.185982755
2010/04	5.481922233
2010/05	5.264060898
2010/06	5.168177377
2010/07	5.356485548
2010/08	4.950339905
2010/09	5.251995993
2010/10	5.486148465
2010/11	5.380162698
2010/12	5.496287914
2011/01	5.071062921
2011/02	5.226583229
2011/03	4.988253938
2011/04	5.179824018
2011/05	5.370505115
2011/06	5.563670325
2011/07	5.301772818
2011/08	5.325526802
2011/09	5.273501456
2011/10	4.857623367
2011/11	5.092234678
2011/12	5.012014626

Fecha	Tasa desocupación
2012/01	4.647803961
2012/02	5.328387962
2012/03	5.123786414
2012/04	4.896566813
2012/05	4.880609123
2012/06	4.848634775
2012/07	4.741731135
2012/08	4.933517374
2012/09	4.691206324
2012/10	4.877671576
2012/11	5.236143203
2012/12	4.944724829
2013/01	5.143483022
2013/02	4.870640413
2013/03	5.061267554
2013/04	5.123876305
2013/05	4.971808853
2013/06	5.077021683
2013/07	4.909425784
2013/08	4.825824642
2013/09	4.994036559
2013/10	4.906837126
2013/11	4.635030203
2013/12	4.746151318
2014/01	4.897098417
2014/02	4.768616104
2014/03	5.298668001
2014/04	4.918232946
2014/05	4.961952896
2014/06	4.831639301
2014/07	5.223157929
2014/08	4.84611523
2014/09	4.785490005
2014/10	4.654195111
2014/11	4.672639123
2014/12	4.146381928
2015/01	4.380624048

Fecha	Tasa desocupación
2015/02	4.440163647
2015/03	4.258277821
2015/04	4.340361078
2015/05	4.426309619
2015/06	4.415755377
2015/07	4.491450994
2015/08	4.386212342
2015/09	4.193098587
2015/10	4.437226303
2015/11	4.07027861
2015/12	4.292693769
2016/01	4.160963031
2016/02	4.269309292
2016/03	4.135578733
2016/04	3.81193747
2016/05	3.993148173
2016/06	3.920034114
2016/07	3.809245061
2016/08	3.76164882
2016/09	3.842520142
2016/10	3.584033485
2016/11	3.595843168
2016/12	3.619661645
2017/01	3.529198585
2017/02	3.453186977
2017/03	3.54074521
2017/04	3.468148272
2017/05	3.544123361
2017/06	3.281453549
2017/07	3.255787381
2017/08	3.332535423
2017/09	3.340491699
2017/10	3.448409158
2017/11	3.505003461
2017/12	3.343179312
2018/01	3.284055894
2018/02	3.285846725

Fecha	Tasa desocupación
2018/03	3.274044655
2018/04	3.381405208
2018/05	3.241811862
2018/06	3.368121068
2018/07	3.297310324
2018/08	3.272433309
2018/09	3.334487829
2018/10	3.20926928
2018/11	3.34878588
2018/12	3.556358408
2019/01	3.457636846
2019/02	3.407617404
2019/03	3.629859428
2019/04	3.487126747
2019/05	3.550697594
2019/06	3.55629693
2019/07	3.530675556
2019/08	3.487927715
2019/09	3.523432638
2019/10	3.624938733
2019/11	3.540379275
2019/12	3.074326515
2020/01	3.650482505
2020/02	3.649021342
2020/03	3.30841429
2020/04	4.675692899
2020/05	4.261749971
2020/06	5.499815129
2020/07	5.008916314
2020/08	4.914710463
2020/09	4.805443637
2020/10	4.602810203
2020/11	4.522686868
2020/12	4.017594091
2021/01	4.530859602
2021/02	4.528986114
2021/03	4.456485778

Fecha	Tasa desocupación
2021/04	4.662494696
2021/05	4.051079751
2021/06	4.04237318
2021/07	4.068280258
2021/08	4.017229224
2021/09	3.933573423
2021/10	3.853416529
2021/11	3.816666345
2021/12	3.734377012
2022/01	3.5692878
2022/02	3.847775446
2022/03	3.442350294
2022/04	3.050308727
2022/05	3.300683667
2022/06	3.386431752
2022/07	3.166125826
2022/08	3.267870047
2022/09	3.145947188
2022/10	3.213897921
2022/11	2.956029591
2022/12	2.942586344
2023/01	2.894693531
2023/02	2.79528823
2023/03	2.786578325
2023/04	2.845375638
2023/05	2.948757675
2023/06	2.685871455
2023/07	2.878658879
2023/08	2.726922739
2023/09	2.717228794
2023/10	2.671791337
2023/11	2.812234089
2023/12	2.786105714
2024/01	2.757186155

Gobierno de México (2023) Datos de la PROFECO del precio de la gasolina Magna en México del año 2002 al 2023

año	precio
2006	14.22
2007	14.43
2008	14.29
2009	14.32
2010	14.83
2011	16.00
2012	17.09
2013	18.38
2014	19.58
2015	19.94
2016	19.86
2017	21.89
2018	24.19
2019	24.66
2020	22.13
2021	23.39
2022	23.27
2023	22.60

“Lis de Veracruz: Arte, Ciencia, Luz”

www.uv.mx

