

**AJUSTE DE MODELOS AUTORREGRESIVOS QUE ASUMEN VALORES ENTEROS
A DATOS DE HURTO A BICICLETAS EN LA LOCALIDAD DE ENGATIVÁ DE LA
CIUDAD DE BOGOTÁ D.C**

MANUEL JAVIER TUNJANO

**Proyecto integral de grado para optar por el título de
PROFESIONAL UNIVERSITARIO EN ESTADÍSTICA Y CIENCIAS ACTUARIALES**

Director

DANIEL LEONARDO RAMÍREZ OROZCO

Mg.

**FUNDACIÓN UNIVERSIDAD DE AMÉRICA
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
PROGRAMA DE ESTADÍSTICA Y CIENCIAS ACTUARIALES
BOGOTÁ D.C**

2023

NOTA DE ACEPTACIÓN

Firma del presidente Jurado

Firma del Jurado

Firma del Jurado

Bogotá. D.C. agosto de 2023

DIRECTIVOS DE LA UNIVERSIDAD

Presidente de la Universidad y Rector del Claustro

Dr. Mario Posada García Peña

Consejero institucional

Dr. Luis Jaime Posada García Peña

Vicerrectoría Académica y de Investigaciones

Dra. Alexandra Mejía Guzmán

Vicerrector Administrativo y Financiero

Dr. Ricardo Alfonso Peñaranda Castro

Secretario General

Dr: José Luis Macías Rodríguez

Decano Facultad Ciencias Económicas y Administrativas

Dr. Esteban Durán Becerra

Las directivas de la Universidad de América, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN	8
OBJETIVOS	11
METODOLOGÍA	12
1. PRELIMINARES	14
1.1 Definiciones y Teoría Básica	14
1.1.1 Definición de Variable aleatoria	14
1.1.2 Definición Distribución Bernoulli	14
1.1.3 Definición de Distribución Binomial	15
1.1.4 Definición de Distribución Poisson	16
1.1.5 Definición de Proceso Estocástico	17
1.1.6 Definición de Series de Tiempo	18
1.1.7 Definición de Función de Autocorrelación (ACF)	21
1.1.8 Definición de Función de Autocorrelación Parcial (PACF)	22
1.1.9 Definición de Ruido Blanco	23
1.1.10 Definición de Proceso Autorregresivo de orden uno (AR(1))	25
1.1.11 Definición de Modelo de medias móviles (MA)	25
1.1.12 Definición de Modelo Autorregresivo de medias móviles AR, MA	25
1.1.13 Definición de Proceso Autorregresivo con Heterodeceasticidad Condicional	27
1.1.14 Definición de Criterio de Información de Akaike (AIC)	28
2. OPERADORES THINNING BINOMIAL Y THINNING POISSON	29
2.1 Definición de Thinning Binomial	29
2.2 Definición de Thinning Poisson	29
3. MODELO AUTORREGRESIVO PARA VALORES ENTEROS INAR(1)	31

3.1	Definición de Poisson INAR(1)	31
4.	PROCESO INARCH (1)	32
5.	PROCESOS AUTORREGRESIVOS MIXTOS	33
5.1	Pominar(1)	33
5.1.1	<i>Casos POMINAR(1)</i>	34
5.1.2	<i>Propiedades probabilísticas del proceso POMINAR(1)</i>	34
5.1.3	<i>Máxima Verosimilitud Condicional (MVC)</i>	36
6.	CONJUNTO DE DATOS SIMULADOS	38
6.1	Simulación	38
6.1.1	<i>Simulación 1</i>	39
6.1.2	<i>Simulación 2</i>	40
6.1.3	<i>Simulación 3</i>	41
7.	APLICACIÓN	42
7.1	Serie de datos	42
7.1.1	<i>Hurtos por año</i>	44
7.1.2	<i>Hurtos por mes (2010)</i>	45
7.1.3	<i>Hurtos por mes (2011)</i>	46
7.1.4	<i>Hurtos por mes (2012)</i>	47
7.1.5	<i>Hurtos por mes (2013)</i>	48
7.1.6	<i>Hurtos por mes (2014)</i>	49
7.1.7	<i>Hurtos por mes (2015)</i>	50
7.1.8	<i>Hurtos por mes (2016)</i>	51
7.1.9	<i>Hurtos por Genero</i>	52
7.1.10	<i>Hurtos por tipo de arma utilizada</i>	53
7.1.11	<i>Hurtos por rango de día</i>	54

7.2	Función de autocorrelación	55
7.3	Función de autocorrelación parcial	56
7.4	Estimación proceso POMINAR(1)	57
7.5	Estimación proceso INAR(1)	58
8.	CONCLUSIONES	59
	BIBLIOGRAFÍA	61

LISTA DE FIGURAS

	Pág.
Figura 1 Ejemplo gráfico Distribución Bernoulli	15
Figura 2 Ejemplo gráfico distribución Binomial	16
Figura 3 Ejemplo gráfico Distribución Poisson	17
Figura 4 Ejemplo de serie de tiempo 1	19
Figura 5 Ejemplo de serie de tiempo 2	20
Figura 6 Ejemplo de serie de tiempo 3	21
Figura 7 Ejemplo de Función de Autocorrelación	22
Figura 8 <i>Ejemplo de Función de Autocorrelación Parcial</i>	23
Figura 9 Ejemplo de ruido blanco	24
Figura 10 AR(1)	26
Figura 11 MA(1)	27
Figura 12 ARMA(1,1)	27
Figura 13 Casos POMINAR(1)	34
Figura 14 Propiedades	35
Figura 15 Esperanza y Varianza	35
Figura 16 Varianza	36
Figura 17 Probabilidades de transición	36
Figura 18 Función de verosimilitud	37
Figura 19 Simulación 1	39
Figura 20 Simulación 2	40
Figura 21 Simulación 3	41
Figura 22 Hurtos de bicicleta en la Localidad de Engativá	42
Figura 23 Estadísticas Descriptivas	43
Figura 24 Datos de Hurtos a bicicletas 2010-2016 localidad de Engativá	43
Figura 25 Hurtos por año	44
Figura 26 Estadísticas Descriptivas 2010-2016	44
Figura 27 Hurtos por mes (2010)	45
Figura 28 Estadísticas descriptivas	45
Figura 29 Hurtos por mes (2011)	46

Figura 30 Estadísticas Descriptivas	46
Figura 31 Hurtos por mes (2012)	47
Figura 32 Estadísticas Descriptivas	47
Figura 33 Hurtos por mes (2013)	48
Figura 34 Estadísticas Descriptivas	48
Figura 35 Hurtos por mes (2014)	49
Figura 36 Estadísticas Descriptivas	49
Figura 37 Hurtos por mes (2015)	50
Figura 38 Estadísticas Descriptivas	50
Figura 39 Hurtos por mes (2016)	51
Figura 40 Estadísticas Descriptivas	51
Figura 41 Hurtos por Sexo	52
Figura 42 Estadísticas Descriptivas	52
Figura 43 Hurtos por tipo de arma utilizada	53
Figura 44 Estadísticas Descriptivas	53
Figura 45 Rango del día	54
Figura 46 Estadísticas Descriptivas	54
Figura 47 Función de autocorrelación	55
Figura 48 Función de Autocorrelación Parcial	56
Figura 49 Estimación proceso POMINAR(1)	57
Figura 50 Criterio AIC POMINAR(1)	57
Figura 51 Estimación INAR(1)	58
Figura 52 Criterio AIC INAR(1)	58

INTRODUCCIÓN

El presente trabajo de grado se basa en el estudio de series de tiempo específicamente de procesos autorregresivos que asumen valores enteros. Se toma como referencia Orozco, (2017), que muestra de forma detallada la construcción de algunos de estos procesos.

Este trabajo inicia con la inquietud de ver teoría reciente en tópicos de series de tiempo, lo que conlleva a estudiar teoría avanzada y que sugiere preguntar:

¿Un modelo autorregresivo mixto que asume valores enteros con innovaciones Poisson de primer orden se ajusta mejor que un modelo autorregresivo para valores enteros de primer orden, a los datos de delitos de alto impacto, robo de bicicletas en la localidad de Engativá de la ciudad de Bogotá D. C entre los años 2010 a 2016?

OBJETIVOS

Objetivo General:

Ajustar dos modelos de series temporales para valores enteros de primer orden a los datos de los delitos de alto impacto en la localidad de Engativá de la ciudad de Bogotá D.C entre los años 2010 a 2016 para encontrar y comparar el mejor ajuste a los datos.

Objetivos específicos:

- Explicar las diferencias entre un modelo autorregresivo para valores enteros de primer orden y un modelo autorregresivo mixto para valores enteros con innovaciones Poisson de primer orden.
- Ajustar un modelo autorregresivo para valores enteros de primer orden a los datos del delito hurto de bicicletas en la localidad de Engativá de la ciudad de Bogotá para los años 2010-2016.
- Ajustar un modelo autorregresivo mixto para valores enteros con innovaciones Poisson de primer orden a los datos del delito hurto de bicicletas en la localidad de Engativá de la ciudad de Bogotá para los años 2010-2016.
- Comparar el ajuste de un modelo autorregresivo para valores enteros de primer orden y un modelo autorregresivo mixto para valores enteros con innovaciones Poisson de primer orden en la localidad de Engativá de la ciudad de Bogotá para los años 2010-2016.

METODOLOGÍA

En consecuencia, recopilar, almacenar y coleccionar datos cada vez es más común, en la actualidad se ha generado la necesidad de utilizar herramientas que ayuden con el procesamiento y el análisis de los datos para entender su comportamiento y poder predecir hechos o acontecimientos futuros. De la estructura de los datos depende la selección del instrumento con el que se puede realizar el análisis deseado. El registro de información en periodos o espacios que dependen del tiempo recibe el nombre de serie de tiempo Guerrero, 1990. Usualmente estos registros son observaciones puntuales en el tiempo. Algunas series temporales conocidas son por ejemplo, el número de días que llueve en un periodo de tiempo, el número de personas que entran a un banco en determinada hora, el número de pacientes del área de urgencias que recibe un hospital en un mes específico, entre otras. Para analizar estas series de tiempo usualmente se utilizan modelos para datos con valores enteros, pero se asume que la variable es de tipo continuo, un modelo autorregresivo para valores enteros de primer orden por sus siglas en inglés (INAR(1)) es una de las herramientas utilizadas. Los primeros en hablar sobre INAR(1) fueron McKenzie, (2003) y Al-Osh and Alzaid, (1987). Steutel and van Harn, (1979) introdujo el operador thinning Binomial que fue retomado por Ferland et al., (2006), básicamente tratan el producto entre una variable aleatoria X y un número real, dando como resultado un número entero. Después Ferland et al., (2006) propone un nuevo modelo basado en el operador thinning Poisson que muestra una alternativa para definir un modelo autorregresivo para valores enteros con heterocedasticidad condicional INARCH(1). Basado en los anteriores autores Orozco, (2017), construye un proceso mixto para valores enteros con innovaciones Poisson de orden uno denominado POMINAR(1) que significa PO=Poisson M=Mixed INAR=Interger Autorregresive, que demostró tener un mejor ajuste que el modelo INAR(1) y otros modelos para valores enteros que no serán tenidos en cuenta en este trabajo, allí se analizan 144 datos de robo reportados en la estación número 44 de la ciudad de Pittsburgh.

En este trabajo se utilizan los datos de hurto a bicicletas en la ciudad de Bogotá para el periodo comprendido desde el año 2010 hasta el año 2016 en la localidad de Engativá y que fueron extraídos de la página oficial de la Secretaría de Seguridad, Convivencia y Justicia de la Alcaldía Mayor de Bogotá D.C de seguridad convivencia y justicia, (2022), para estos datos se les ajusta un modelo autorregresivo para valores enteros INAR(1) y un modelo mixto autorregresivo POMINAR(1), con el objetivo de revisar cuál de los dos modelos arroja un mejor ajuste a los datos estudiados.

El trabajo se divide en capítulos de tal manera que al lector se le facilite la comprensión. En el capítulo 1 se definen las distribuciones Bernoulli, Binomial y Poisson, también se hace una introducción al campo de las series temporales como el proceso autorregresivo, la función de autocorrelación (FAC), la función de autocorrelación parcial (FACP), modelos autorregresivos $AR(p)$, de medias móviles $MA(q)$ y la combinación de estos $ARMA(p,q)$. En el capítulo 2 se define el proceso $INAR(1)$ con sus propiedades. Para el capítulo 3 se hace la explicación del operador thinning Poisson y thinning Binomial Ferland et al.,(2006). En el capítulo 4 se explica el proceso autorregresivo para valores enteros con heterocedasticidad condicional $INARCH(1)$. El proceso $POMINAR(1)$ creado por Orozco,(2017), incluye diferentes tipos $POMINAR(1)$ que se derivan del proceso creado. Las propiedades probabilísticas y la estimación por Máxima Verosimilitud también son presentadas en el capítulo 5. En el capítulo 6 se desarrollan 3 simulaciones de datos para observar el comportamiento del proceso $POMINAR(1)$, para finalmente en el capítulo 7 hacer la aplicación a datos reales y comparar el ajuste de los modelos $INAR(1)$ y $POMINAR(1)$ a los datos de hurtos de bicicleta con el criterio AIC Cavanaugh and Neath, (2019) que es utilizado para dar un valor de ajuste del modelo a los datos, a menor valor es mejor el ajuste del modelo a los datos. En el capítulo 8 se presentan las conclusiones y se deja una recomendación para los trabajos futuros.

1. PRELIMINARES

En este capítulo se definen las distribuciones Bernoulli, Binomial y Poisson Blanco Castañeda, (2013) y las definiciones básicas de series temporales Guerrero, (1990), Peña, (2005) y Hamilton, (2020) mencionadas en este trabajo.

1.1 Definiciones y Teoría Básica

1.1.1 Definición de Variable aleatoria

Una variable aleatoria X es cualquier función que asigna un número real a cada resultado dentro de un conjunto de posibles resultados Ω y definido por:

$$X : \Omega \rightarrow \mathbb{R}. \quad (1.1)$$

1.1.2 Definición Distribución Bernoulli

La distribución Bernoulli arroja una probabilidad de éxito θ o fracaso $1-\theta$.

Tabla 1

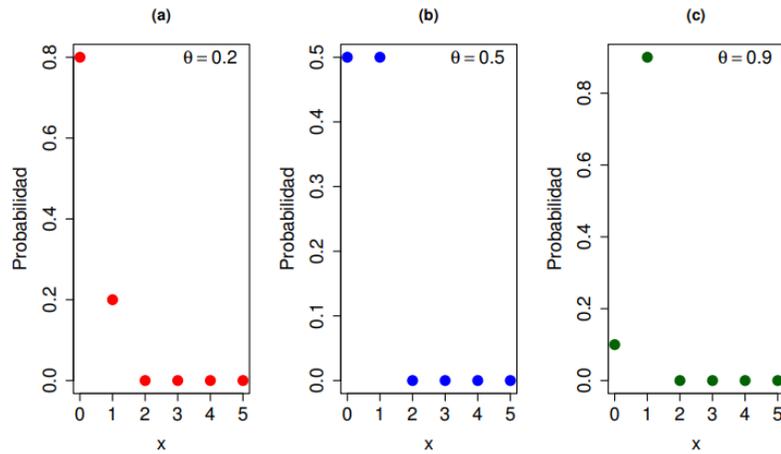
Distribución Bernoulli

Función de probabilidad	$f(x; \theta) = \theta^x(1 - \theta)^{1-x}$
Soporte	$x \in \{0, 1\}$
Parámetro	$0 \leq \theta \leq 1$
Valor Esperado	$\mathbb{E}(X) = \theta$
Varianza	$\mathbb{V}(X) = \theta(1 - \theta)$

Nota. se muestra la función de probabilidad, soporte, parámetro, valor esperado y la varianza de la distribución Bernoulli.

Figura 1

Ejemplo gráfico Distribución Bernoulli



Nota. Para ver el comportamiento de la distribución Bernoulli, se dan diferentes valores a los parámetros y se realizan los siguientes gráficos ilustrados en la figura 1.1. Para la parte (a) $\theta = 0,20$, (b) $\theta = 0,50$ y (c) $\theta = 0,90$.

1.1.3 Definición de Distribución Binomial

La distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos de n ensayos de Bernoulli independientes entre sí con una probabilidad θ de ocurrencia de éxito entre los ensayos.

Tabla 2

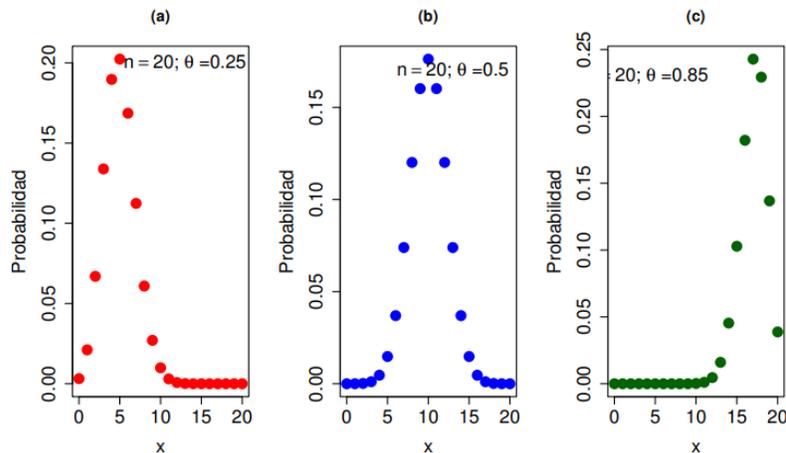
Distribución Binomial

Función de probabilidad	$f(x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$
Soporte	$x \in \{0, 1, 2, \dots, n\}$
Parámetros	$n \in \{0, 1, 2, \dots\}; 0 \leq \theta \leq 1$
Valor Esperado	$\mathbb{E}(X) = n\theta$
Varianza	$\mathbb{V}(X) = n\theta(1 - \theta)$

Nota. En la tabla se muestra la función de probabilidad, soporte, parámetro, valor esperado y la varianza de la distribución Binomial.

Figura 2

Ejemplo gráfico distribución Binomial I



Nota. Para ver el comportamiento de la distribución Binomial, se dan diferentes valores al parámetro θ y se realizan los siguientes gráficos ilustrados en la figura. Para la parte (a) $n = 20$ y $\theta = 0,25$, (b) $n = 20$ y $\theta = 0,50$ y (c) $n = 20$ y $\theta = 0,85$.

1.1.4 Definición de Distribución Poisson

La distribución Poisson es una distribución de probabilidad discreta que se basa en la frecuencia media en la que se presenta un evento durante cierto periodo de tiempo.

Tabla 3

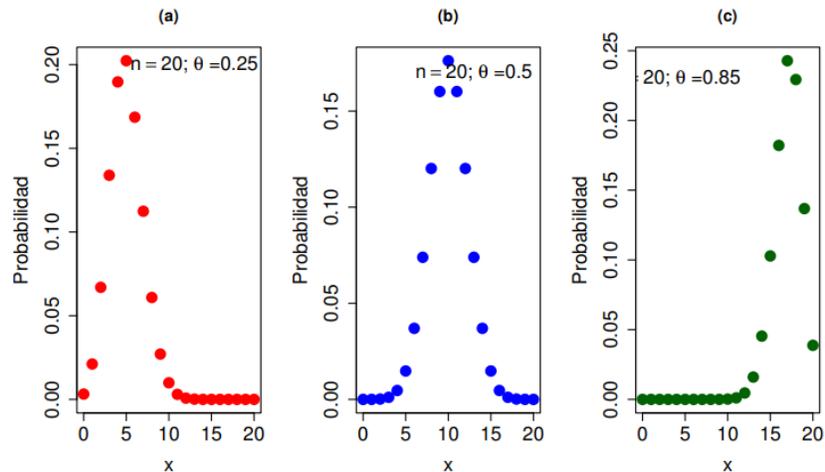
Distribución Poisson

Función de probabilidad	$f(x; \theta) = \frac{\theta^x e^{-\theta}}{x!}$
Soporte	$x \in \{0, 1, 2, \dots\}$
Parámetro	$\theta \in [0, \infty)$
Valor Esperado	$\mathbb{E}(X) = \theta$
Varianza	$\mathbb{V}(X) = \theta$

Nota. En la tabla se muestra la función de probabilidad, soporte, Parámetro, valor esperado y la varianza de la distribución Poisson.

Figura 3

Ejemplo gráfico Distribución Poisson



Nota. Para ver el comportamiento de la distribución Poisson, se dan diferentes valores al parámetro θ y se realizan los siguientes gráficos ilustrados en la figura. Para la parte (a) $\theta = 4$, (b) $\theta = 7$ y (c) $\theta = 10$.

1.1.5 Definición de Proceso Estocástico

Un proceso Estocástico es una colección de variables aleatorias $\{X_t\}_{t \in T}$ definidas sobre un espacio de probabilidad (Ω, \mathcal{F}, P) , en que Ω es el espacio muestral, \mathcal{F} es una sigma 'algebra y P es una función de probabilidad. T es un conjunto de 'índices del proceso o espacio paramétrico.

En nuestro caso Ω es un conjunto ordenado y corresponde a los instantes temporales t para cada valor t de Ω , está definida una variable aleatoria Y_t , en que los valores observados de las variables aleatorias en distintos instantes forman una serie temporal como se muestra en la siguiente definición.

1.1.6 Definición de. Series de Tiempo

Una serie de tiempo es una colección de observaciones medidas secuencialmente a lo largo del tiempo, Bisgaard and Kulahci, (2011). Otra definición para las series de tiempo es una observación numérica efectuada a intervalos de tiempo fijos, de tales características o variables económicas Guerrero, (1990). Una serie temporal también se puede definir como es el resultado de observar una variable al largo del tiempo en intervalos regulares Peña, (2005).

Estas colecciones u observaciones son denotadas por:

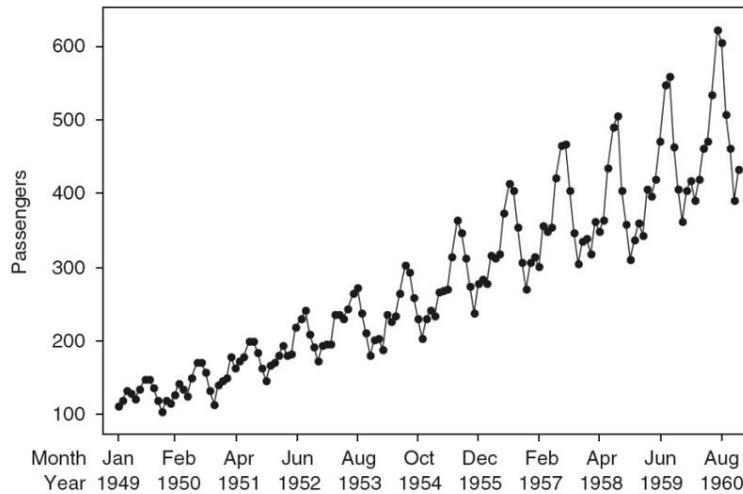
$$Y_t = Y_{t_1}, Y_{t_2}, \dots, Y_{t_n},$$

en que $t \in T \subseteq R$, Y_{t-i} es el valor de la variable en el instante $t - i$.

A continuación, se muestran algunos ejemplos para ilustrar lo que es una serie de tiempo.

Figura 4

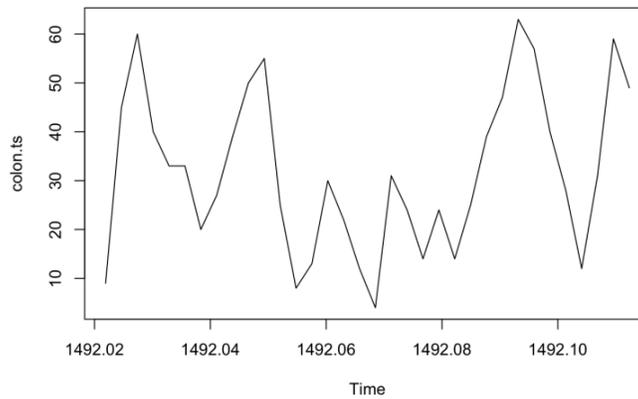
Ejemplo de serie de tiempo 1



Nota. En la figura extraída de Bisgaard and Kulahci, (2011) se muestra el número de pasajeros usuarios de las aerolíneas en Estados Unidos entre los años 1949 y 1960. Sobre la base de los patrones de viaje cíclicos, podemos ver que los datos exhiben un comportamiento estacional que es un comportamiento regular con subidas y bajadas periódicas en la serie temporal. Pero también podemos ver una tendencia al alza, lo que sugiere que los viajes aéreos son cada vez más y más populares a través del tiempo.

Figura 5

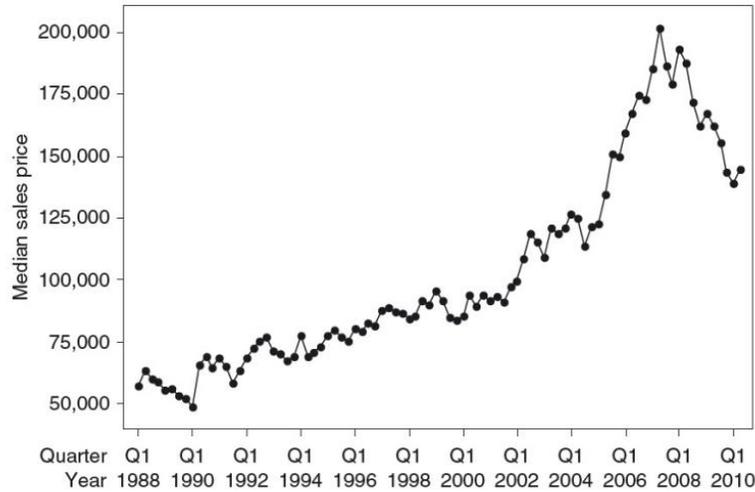
Ejemplo de serie de tiempo 2



Nota. En la figura extraída de Peña, (2005) se muestra el trayecto diario navegado, medido en leguas marinas, por la flota de Cristóbal Colón en su primer viaje al continente americano. Se observa que la serie es estable, con valores que oscilan alrededor de un recorrido de 30 leguas. Su gráfico no muestra tendencia creciente o decreciente clara.

Figura 6

Ejemplo de serie de tiempo 3



Nota. En la figura extraída de Bisgaard and Kulahci, (2011) se muestra el precio medio de venta de las casas en Estados Unidos desde 1988 hasta el segundo trimestre de 2010. Los precios venían en un ascenso hasta el año 2007 donde se produjo una noticia que advertía una crisis empezando el 2007.

1.1.7 Definición de Función de Autocorrelación (ACF)

La autocorrelación mide la correlación entre dos variables separadas por k periodos y se muestra a continuación:

Tabla 4

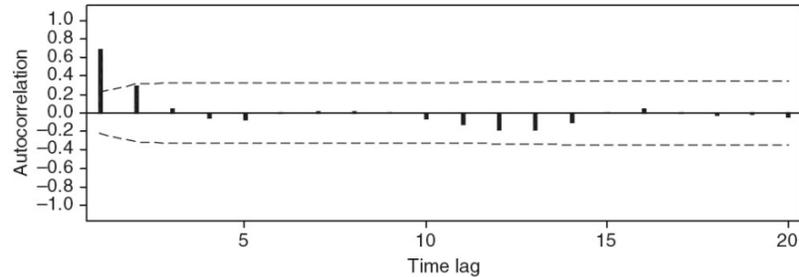
(AFC)

$$\rho(k) = \frac{\gamma(k)}{\sqrt{\gamma(0)} \sqrt{\gamma(0)}} = \frac{\gamma(k)}{\gamma(0)}$$

Nota. donde $\gamma(k)$ es la covarianza y $\gamma(0)$ la varianza.

Figura 7

Ejemplo de Función de Autocorrelación



Nota. En la figura extraída de Bisgaard and Kulahci, (2011) se muestra la Función de Auto- correlación de la temperatura de los hornos industriales. Las dos líneas horizontales punteadas indican los intervalos de confianza del 95 % alrededor de 0, la autocorrelación de retraso 1 es claramente significativamente diferente de 0, y el retraso 2 la autocorrelación es significativa en el límite.

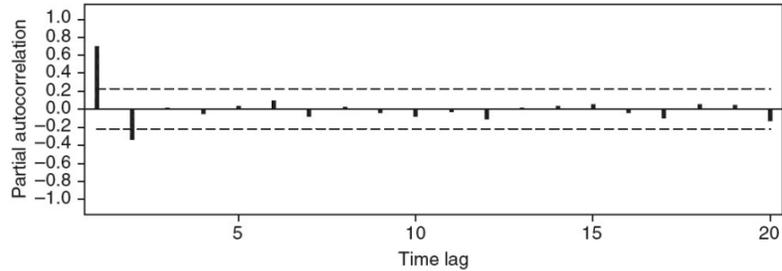
1.1.8 Definición de. Función de Autocorrelación Parcial (PACF)

Es la correlación que existe entre dos valores de la serie que se encuentran en un intervalo de distancia (k). La función de autocorrelación está definida por

$$\gamma(k) = E [(X_{t+k} - \mu)(X_t - \mu)](1.3)$$

Figura 8

Ejemplo de Función de Autocorrelación Parcial



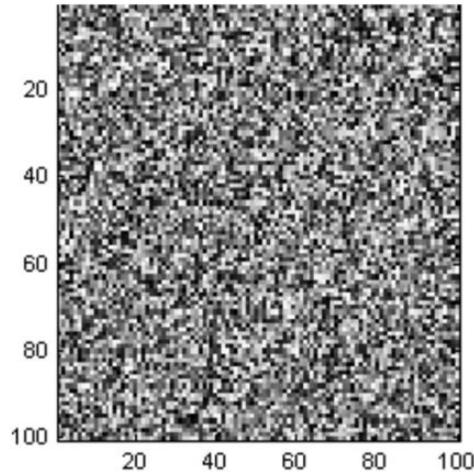
Nota. En la figura extraída de [Bisgaard and Kulahci, 2011] muestra una PACF de las diferencias (cambios) del número de usuarios del servidor de Internet durante un periodo de 100 minutos. Se observan ondas amortiguadas. Este patrón sugiere un modelo mixto ARMA que será definido más adelante.

1.1.9 Definición de Ruido Blanco

El proceso de ruido blanco es una señal aleatoria que se caracteriza por el hecho de que sus valores de señal en dos tiempos diferentes no guardan correlación estadística. Como consecuencia de ello, su densidad espectral de potencia (PSD, siglas en inglés de power spectral density) es una constante, es decir, su gráfica es plana. Esto significa que la señal contiene todas las frecuencias y todas ellas muestran la misma potencia, exactamente el mismo fenómeno que ocurre con la luz blanca, de allí la denominación. Además, se trata de un proceso en el que todas sus variables son independientes denotada por ε_t con media 0 y varianza constante.

Figura 9

Ejemplo de ruido blanco



Nota. La figura en blanco y negro extraída de Ruiz, (2022) es ruido blanco ya que sus píxeles no guardan correlación entre sí y por tanto su densidad espectral de potencia es constante. Si la imagen fuese en color, entonces la “nieve” sería de colores aleatorios.

Esta es la típica imagen que se ve en la pantalla de un televisor analógico cuando no está sintonizado en ningún canal. La señal que recibe en este momento el demodulador puede considerarse ruido blanco, ya que es el resultado de sumar el ruido electromagnético del canal de radio más el que generan los propios circuitos electrónicos del televisor, múltiples interferencias de baja intensidad todas ellas independientes entre sí. En este último caso, la “nieve” no permanecería estática, sino que cambiaría constantemente con el tiempo, porque la señal de televisión es una señal de video, por ejemplo, una sucesión de imágenes a más de 25 fotogramas por segundo.

1.1.10 Definición de. Proceso Autorregresivo de orden uno (AR(1))

El proceso autorregresivo depende del valor de la variable en un tiempo anterior, para describir cuál será el valor en el tiempo posterior.

$$X_t = \phi X_{t-1} + \varepsilon_t$$

en que X_{t-1} es el valor de la variable en el tiempo anterior, $\phi \in R$ y ε_t es el ruido blanco.

Algunas de las características de un AR(1) son:

La ACF será una función positiva y decreciente.

La ACF será una función alternada, y tendrá valores pares positivos, y valores. impares negativos.

La PACF tendrá un único valor, específicamente este valor será positivo.

1.1.11 Definición de Modelo de medias móviles (MA)

El modelo MA son promedios de valores pasados y presentes. Un modelo MA se puede escribir:

$$w_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

en dónde los coeficientes θ se determinan desde los datos.

1.1.12 Definición de Modelo Autorregresivo de medias móviles AR, MA

Los modelos de series de tiempo (AR), (MA), o una combinación de los dos, los modelos ARMA.

La serie de tiempo estacionaria está denotada por w_t y puede ser la primera diferencia de una serie temporal no estacionaria Z_t , *qué es*, $w_t = Z_t - Z_{t-1}$. . Por lo tanto, un modelo AR(p) se escribe como:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + a_t$$

en que a_t es un término de error que se supone que no está correlacionado, con media 0 y varianza constante y los coeficientes de regresión ϕ son los Parámetros que van a ser estimados.

Finalmente, el modelo ARMA se escribe:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} + a_t.$$

En la práctica nunca sabemos el tipo de modelo y en concreto los órdenes p y q necesarios para modelar adecuadamente un proceso dado. Necesitamos determinar el modelo. Con el siguiente cuadro 1.4 podemos ver las características más importantes para identificar el modelo:

Tabla 5

Modelos autorregresivos

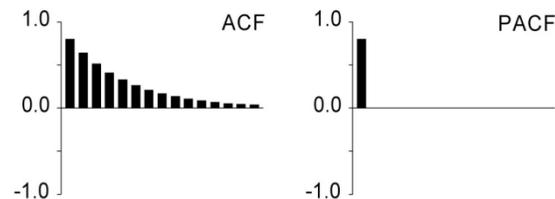
Modelos			
	AR(p)	MA(q)	ARMA(p,q)
Modelo	$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t$	$w_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$	$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} + a_t$
Función de autocorrelación (AFC)	Infinitas: Decaimiento exponencial con ondas senoidales.	Finitas: Cortes después de t retrasos	Infinitas: Decaimiento exponencial y/o ondas senoidales
Función de autocorrelación parcial (FACP)	Finitas: Cortes después de t retrasos	Infinitas: Decaimiento exponencial con ondas senoidales.	Infinitas: Decaimiento exponencial y/o ondas senoidales

Nota. El cuadro describe los modelos Autorregresivos AR(p), MA(q) y ARMA(p,q) formado por los dos anteriores donde p es la parte autorregresiva y q es el orden de la parte móvil.

Donde ϕ_1, \dots, ϕ_p son los parámetros del modelo y a_t son los errores con $t - n$ rezagos en el tiempo Bisgaard and Kulahci, (2011).

Figura 10

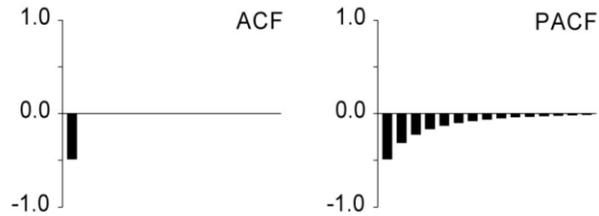
AR(1)



Nota. En la figura extraída de Mauricio, (2007) se puede observar un decaimiento exponencial en la ACF y el primer rezago significativo en la PACF .

Figura 11

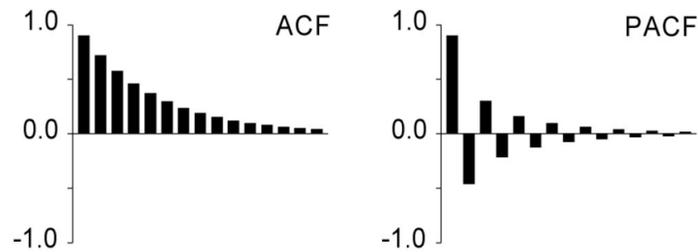
$MA(1)$



Nota. En la figura extraída de Mauricio, (2007) se puede observar un decaimiento en la PACF y el primer rezago significativo en la ACF.

Figura 12

$ARMA(1,1)$



Nota. En la figura 1.12 extraída de Mauricio, (2007) se puede observar un decaimiento exponencial en la PACF y en la ACF.

1.1.13 Definición de Proceso Autorregresivo con Heterodeceasticidad Condicional

El modelo de heterocedasticidad condicional autorregresiva (ARCH) Degiannakis and Xekalaki, (2010) es un modelo estadístico para datos de series de tiempo que modela la varianza del error actual como una función de los tamaños reales de los errores de periodos de tiempo anteriores. El modelo ARCH es apropiado cuando la varianza del error en una serie temporal sigue un modelo AR.

El proceso ARCH está definido por:

$$Y_t = \theta_0 + \sum_{i=1}^q Y_{t-i} + \varepsilon_t$$

en que ε_t denota el término de error, θ el parámetro a estimar y Y_t las observaciones.

1.1.14 Definición de Criterio de Información de Akaike (AIC)

El criterio de información para la identificación de modelos, conocido como criterio de información de Akaike (AIC) Akaike, 1987, es un estimador de la calidad relativa de los modelos estadísticos que representan un conjunto de datos. Dada la colección de modelos aplicables a esos datos, el AIC estima la calidad de cada modelo, en forma relativa con respecto a los otros modelos.

En general el AIC es determinado por:

$$AIC = 2k - 2\ln(L),$$

donde k es el número de Parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud que será explicada en la ecuación con la aplicación para series de tiempo.

2. OPERADORES THINNING BINOMIAL Y THINNING POISSON

En este capítulo se describen los operados thinning Binomial y thinning Poisson con sus respectivas propiedades. El operador thinning Binomial fue propuesto por Al-Osh and Alzaid, (1987) y en este trabajo especialmente se menciona el modelo poisson INAR(1). El operador thinning Poisson fue propuesto por [Ferland et al., 2006] con el objetivo de tener un modelo alternativo para los procesos INARCH(1).

2.1 Definición de Thinning Binomial

Sea X una variable aleatoria que asume valores enteros no negativos, α un número real tal que $\alpha \in [0, 1]$. El operador thinning binomial es definido como:

$$\theta \circ X = \sum_{i=1}^X Bi(\theta)$$

En donde $\{Bi(\theta)\}_{i=1}^X$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas con distribución Bernoulli de parámetro θ , independientes de X con función de probabilidad dada por $P(Bi(\theta) = 1) = \theta = 1 - P(Bi(\theta) = 0)$. Algunas propiedades del operador thinning Binomial son:

- 0 si $X = 0$
- 1 si $X = X$.
- $\theta \circ (\theta \circ X) \stackrel{d}{=} \theta$
- Si X y Y son independientes, $\theta \circ (X + Y) \stackrel{d}{=} \theta_1 \circ X + \theta_1 \circ Y$

El símbolo $\stackrel{d}{=}$ denota que las variables aleatorias siguen la misma distribución. Otras propiedades y demostraciones pueden ser consultadas en Orozco, (2017).

2.2 Definición de Thinning Poisson

Sea X una variable aleatoria que asume valores enteros no negativos, $\delta \geq 0$ un número real tal que $\delta \in [0, 1]$. El operador thinning Poisson “ $*$ ” es definido como:

$$\delta * X = \sum_{i=1}^X Pi(\delta)$$

en donde $\{Pi(\delta)\}_{i=1}^X$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas (*i.i.d.*) con distribución Poisson de parámetro δ e independientes de X . Algunas propiedades del operador thinning Poisson son:

- $0 * X = 0$.

- $\delta_1(X_1 + X_2) \stackrel{d}{=} \delta_1 * X_1 + \delta_1 * X_2$, si X_1, X_2 son independientes.
- $E[\delta_1 * X_1] = \delta_1 E[X_1]$.

El símbolo $\stackrel{d}{=}$ denota que las variables aleatorias siguen la misma distribución. Otras propiedades y demostraciones pueden ser consultadas en Orozco, (2017).

3. MODELO AUTORREGRESIVO PARA VALORES ENTEROS INAR(1)

El Modelo Autorregresivo para valores enteros INAR(1) es un proceso discreto para valores no negativos cuya ecuación es definida en este capítulo, complementando la definición con el proceso Poisson INAR(1) para valores enteros no negativos que satisface la siguiente ecuación:

$$X_t = \theta \circ X_{t-1} + \varepsilon_t \text{ para } t \geq 1$$

en donde $\theta \in [0, 1]$, $N_0 = \{0, 1, \dots\}$, " \circ " definido como el operador thinning binomial dado en la definición 2.1, $\{\varepsilon_t\}_{t \geq 1}$, $\varepsilon_t \in N_0$, una secuencia de variables aleatorias independientes e idénticamente distribuidas, para valores enteros no negativos con media μ_ε y varianza σ_ε^2 y la serie $\theta \circ X_{t-1}$ independientes de ε_t .

3.1 Definición de Poisson INAR(1)

Sea $\{X_t\}_{t \geq 1}$ un proceso INAR(1) y ε_t siguiendo una distribución Poisson con parámetro $\lambda \sim Po(\lambda)$, entonces $\{X_t\}_{t \geq 1}$ es un proceso Poisson INAR(1), en que $E[\varepsilon_t] = \mu_\varepsilon = \lambda$, $V(\varepsilon_t) = \sigma^2 = \lambda$.

La media y la varianza marginal del modelo Poisson INAR(1) están dadas por

$$E[X_t] = \frac{\lambda}{1-\theta} = Var(X_t)$$

y la media y la varianza condicional de X_t pueden ser escritas como

$$E[X_t | X_{t-1}] = \theta X_{t-1} + \lambda,$$

$$Var(X_t | X_{t-1}) = \theta(1 - \theta)X_{t-1} + \lambda,$$

Una interpretación del proceso Poisson INAR(1) según Ferland et al., (2006), puede ser interpretado como un proceso de nacimiento y muerte, en donde para cada individuo en el tiempo $t - 1$ tiene una probabilidad θ de continuar vivo en el tiempo t y en cada tiempo t ocurre un número de nacimientos que sigue una Distribución Poisson de parámetro λ .

4. PROCESO INARCH (1)

El proceso autorregresivo para valores enteros con heterodeceasticidad condicional, de- finido de forma alternativa por Fernández, (2013) basado en el operador thinning Poisson. Es definido como:

Un proceso para valores enteros no negativos $\{X_t\}t \geq 1, X_t \in N_0$, es un proceso INARCH(1) que satisface la siguiente ecuación:

$$X_t = \beta * X_{t-1} + \varepsilon_t, \text{ para } t \geq 1,$$

en donde $\beta \geq 0$, “ * ” es el operador thinning Poisson que asume una serie $\beta * X_{t-1}$ y es independiente de ε_t .

Sea $\{X_t\}t \geq 1$ un proceso INARCH(1) y ε_t siguiendo una distribución Poisson con Parámetro λ $\sim Po(\lambda)$, entonces $\{X_t\}t \geq 1$ es un proceso Poisson INARCH(1), en que $E[\varepsilon_t] = \mu_\varepsilon = \lambda, V(\varepsilon_t) = \sigma^2 = \lambda$.

5. PROCESOS AUTORREGRESIVOS MIXTOS

Los procesos Autorregresivos Mixtos amplían el modelo general de manera que los datos puedan presentar variabilidad correlacionada y no constante. El modelo Autorregresivo Mixto proporciona, por tanto, la flexibilidad necesaria para modelar no sólo las medias sino también las varianzas y covarianzas de los datos. En el siguiente capítulo [Orozco, 2017] muestra la construcción de un proceso mixto para valores enteros con innovaciones Poisson de orden uno denominado POMINAR(1) PO=Poisson M=Mixed INAR=Interger Autorregresive con el objetivo de mejorar el ajuste de un modelo a datos enteros.

5.1 Pominar(1)

El proceso POMINAR(1), es un proceso autorregresivo mixto para valores enteros con innovaciones Poisson. El proceso considera dos operadores, el operador thinning Binomial “ \circ ” explicado en la definición 2.1 y el operador thinning Poisson “ $*$ ” mostrado en la definición 2.2. El objetivo es hacer un proceso mixto, uniendo dos procesos autorregresivos para valores enteros con innovaciones Poisson. Los procesos Poisson INAR(1) e INARCH(1) son casos particulares del modelo propuesto y descritos de forma general en la sección de definiciones.

Un proceso POMINAR(1) se define con la siguiente ecuación:

en donde $\theta, \delta \in [0,1), p \in [0,1), \{\varepsilon_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas que siguen una Distribución Poisson de Parámetro λ , ε_t es independiente de todo $\{Bi(\theta)\}$ y de todo $\{Pi(\delta)\}$ definidos en 1.3 y 1.4, respectivamente;

$\{Bi(\theta)\}$ y $\{Pi(\delta)\}$ son independientes entre sí; los operadores thinnings en el tiempo t son independientes uno de otro. Este tipo de procesos INAR(1) puede modelar varios tipos de casos

$$X_t = \begin{cases} \theta \circ X_{t-1} + \varepsilon_t, & \text{con probabilidad } p \\ \beta * X_{t-1} + \varepsilon_t, & \text{con probabilidad } 1 - p \end{cases}$$

5.1.1 Casos POMINAR(1)

Figura 13

Casos POMINAR(1)

- Si $p = 1$ entonces $X_t = \theta \circ X_{t-1} + \varepsilon_t$.
es un proceso POMINAR(1)-I.
- Si $p = 0$ entonces $X_t = \delta * X_{t-1} + \varepsilon_t$.
es un proceso POMINAR(1)-II.
- Si $\theta = 0$ entonces

$$X_t = \begin{cases} \varepsilon_t, \text{ con probabilidad } p, \\ \delta * X_{t-1} + \varepsilon_t, \text{ con probabilidad } 1 - p, \end{cases}$$

es un proceso POMINAR(1)-III.

- Si $\delta = 0$ entonces

$$X_t = \begin{cases} \theta \circ X_{t-1} + \varepsilon_t \text{ con probabilidad } p, \\ \varepsilon_t, \text{ con probabilidad } 1 - p, \end{cases}$$

es un proceso POMINAR(1)-IV.

- Si $\theta = 0$ y $\delta = 0$, $p = 1$ y $\theta = 0$, entonces $X_t = \varepsilon_t$, para todo t . Así X_t es una secuencia de variables aleatorias i.i.d. con una distribución Poisson de parámetro λ es un proceso POMINAR(1)-V.

Nota. Ilustración de los casos POMINAR

5.1.2 Propiedades probabilísticas del proceso POMINAR(1)

En esta sección se muestran algunas propiedades probabilísticas como la esperanza condicional k pasos adelante, una media y una varianza condicional k pasos adelante y probabilidades

Teorema suponga que X_t sigue un proceso POMINAR(1), entonces tiene una media condicional k pasos adelante y una varianza condicional k pasos al adelante.

Figura 14

Propiedades

$$\mathbb{E}[X_{t+k} | X_t] = C_1^k X_t + \frac{1 - C_1^k}{1 - C_1} \mathbb{E}[\varepsilon_t],$$

$$\begin{aligned} \text{Var}(X_{t+k} | X_t) &= (C_2^k - C_1^{2k})X_t^2 + (1 - C_2^k)E[X_t^2] + (C_1^{2k} - 1)(E[X_t^2])^2 \\ &+ \left(\frac{C_4(C_1^k - C_2^k)}{C_1 - C_2} - 2C_1^k(1 - C_1^k)\mathbb{E}[X_t] \right) (X_t - \mathbb{E}[X_t]), t \geq 1, \\ \text{en donde } C_1 &= p\theta + (1-p)\delta^2, C_2 = p\theta^2 + (1-p)\delta^2, C_3 = P\theta(1 - \theta) + 1(1-p)\delta, \\ C_4 &= C_3 + 2\mathbb{E}[\varepsilon_t]C_1, \mathbb{E}[X_t] = \lambda/(1 - C_1), \text{ y } \mathbb{E}[X_t^2] = (C_4\mathbb{E}[X_t] + \mathbb{E}[\varepsilon_t^2])/(1 - C_2). \end{aligned}$$

Nota. Propiedades probabilísticas POMINAR(1)

Para que un proceso POMINAR(1) sea estacionario de segundo orden debe cumplir que la media y la varianza estén en función del tiempo, tal que, estas expresiones además de existir, deben ser una función constante de t . Así como se observa en la siguiente ecuación:

Figura 15

Esperanza y Varianza

$$\mathbb{E}[X_t] = \frac{\lambda}{1 - C_1},$$

$$\mathbb{E}[X_t^2] = \frac{C_4\mathbb{E}[X_{t-1}] + \lambda^2 + \lambda}{1 - C_2},$$

$$\text{Var}(X_t) = \frac{\lambda^2[(1 - C_1)^2 - (1 - C_2)] + \lambda(1 - C_1)[C_4 + 1(1 - C_1)]}{(1 - C_2)(1 - C_1)^2}$$

Así para $C_1, C_2 \in [0,1)$, y como se define en 5.2.

$$\lim_{k \rightarrow \infty} \mathbb{E}[X_{t+k} | X_t] = \mathbb{E}[\varepsilon_t]/(1 - C_1) = \mathbb{E}[X_t],$$

Nota. Esperanza y Varianza POMINAR(1)

Figura 16

Varianza

$$\lim_{k \rightarrow \infty} \text{Var}(X_{t+k} | X_t) = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 = \text{Var}(X_t),$$

Nota. Varianza

es la varianza del proceso. Para más detalles de la construcción de este proceso, puede ser consultado Orozco, (2017).

5.1.3 Máxima Verosimilitud Condicional (MVC)

Un método de MVC en el condicionamiento de la primera observación de la Función de verosimilitud, se puede expresar con las probabilidades de transición, así:

$$L(\gamma, \delta, \lambda, p | X) = \prod_{t=2}^T P(X_t = j | X_{t-1} = i),$$

En que $\gamma, \delta, \lambda, p$ son los Parámetros a maximizar y X_t son los valores observados. Por lo tanto, con este método se busca estimar los parámetros de un modelo basado en una Distribución conjunta de un proceso $\{X_t\}$.

Figura 17

Probabilidades de transición

$$\mathbb{P}(X_t = j | X_{t-1} = i) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \theta^k (1-\theta)^{i-k} \left[\frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \right] + (1-p) \frac{e^{-(\delta+\lambda)} (\delta_i + \delta)^j}{j!}$$

en que las probabilidades de transición pueden ser explicadas así:

- X_t es la población en el tiempo t
- $\delta \circ X_{t-1}$ sobrevivientes en el tiempo t
- X_{t-1} es la población en el tiempo $t-1$
- ε_t son los nuevos miembros en el tiempo t

Nota. Probabilidades de transición

en que las probabilidades de transición pueden ser explicadas así:

X_t es la población en el tiempo

$\delta \circ X_{t-1}$ sobrevivientes en el tiempo t X_{t-1} es la población en el tiempo $t-1$

ε_t son los nuevos miembros en el tiempo t

Caso 1: $j \geq i$ en $(X_t = j | X_{t-1} = i)$. Para $X_{t-1} = i$, significa que el tamaño de la población en el tiempo $t-1$ es igual a $i \cdot \delta \circ X_{t-1}$, que son los sobrevivientes de X_{t-1} en el periodo $(t-1, t]$, una cantidad $k \leq i$. De tal forma que el valor observado para los nuevos miembros ε_t representa un valor mayor que cero y la suma de k con $j \geq i$.

Caso 2: $i < j$ en $(X_t = j | X_{t-1} = i)$. Aquí hay suficientes muertes para garantizar que no hay sobrevivientes ($\delta \circ X_{t-1} = 0$), o sobrevive una cantidad menor que i ($\delta \circ X_{t-1} < i$) entre los tiempos $(t-1, t]$, de tal manera que los valores observados de las innovaciones ε_t agregadas a los sobrevivientes en el tiempo $t-1$ con $j \geq i$ Orozco, (2017).

La Función de verosimilitud condicional $(\gamma, \delta, \lambda, p)$, del proceso POMINAR(1) está dada por:

Figura 18

Función de verosimilitud

$$\begin{aligned} l(\theta, \delta, \lambda, p) &= \log \left[\prod_{t=2}^T \mathbb{P}(X_t = j | X_{t-1} = i) \right] \\ &= \sum_{t=2}^T \log[\mathbb{P}(X_t = j | X_{t-1} = i)] \\ &= \sum_{t=2}^T \log \left[p \sum_{k=0}^{\min(i,j)} \binom{i}{k} \theta^k (1-\theta)^{i-k} \left[\frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \right] + (1-p) \frac{e^{-(\delta_i+\lambda)} (\delta_i+\lambda)^j}{j!} \right]. \end{aligned}$$

Nota. Función de verosimilitud

$\hat{\gamma}, \hat{\delta}, \hat{\lambda}, \hat{p}$ son los valores estimados que maximizan la función de verosimilitud condicional para los parámetros $\gamma, \delta, \lambda, p$. Para este trabajo las estimaciones serán realizadas a través del software Rstudio versión 4.2.1.

6. CONJUNTO DE DATOS SIMULADOS

6.1 Simulación

En esta sección se desarrolla una simulación para observar el comportamiento del modelo POMINAR(1). La dispersión está dada por σ^2 / μ_x , en donde σ^2 es la varianza del proceso y μ_x es la media del proceso, respectivamente. Con $Var(X_t)/E[X_t] > 1$, entonces el proceso POMINAR(1) es un modelo para datos superdispersos en donde la varianza es mayor que la media. En el siguiente cuadro se muestran diferentes valores generados aleatoriamente para θ, δ, λ y p , para cada caso se muestra el índice de dispersión:

Tabla 6

Valores simulados

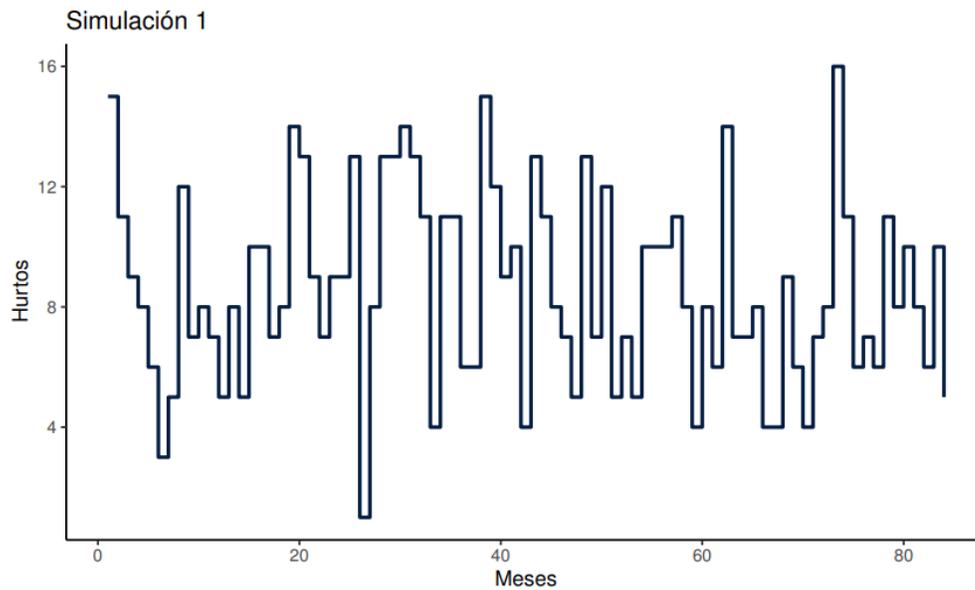
θ	δ	λ	p	σ_X^2 / μ_X
0.1	0.6	7	0.8	1.19
0.5	0.5	4	0.2	1.43
0.5	0.2	6	0.5	1.58

Nota. Simulación de valores

6.1.1 Simulación 1

Figura 19

Simulación 1

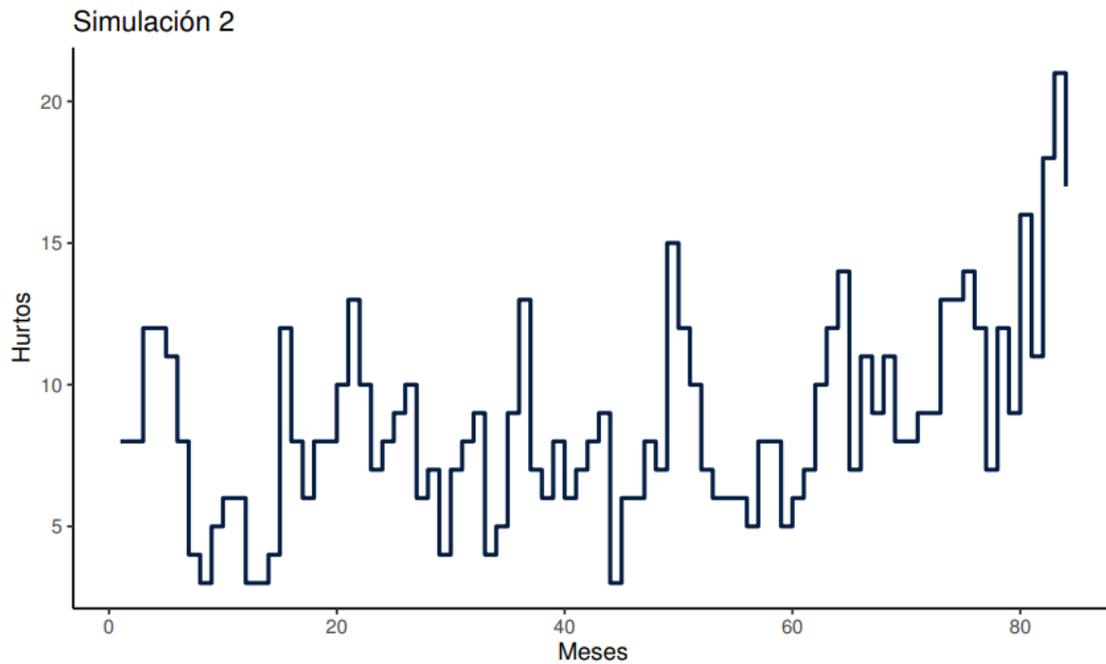


Nota. En la figura se simularon 84 datos para el POMINAR(1) con valores de los Parámetros $\theta = 0,1$, $\delta = 0,6$, $\lambda = 7$ y $p = 0,8$ como se puede observar en el cuadro 6.1, para estos datos se obtuvo una media de 8.58 y una varianza de 10.24. Los datos fueron simulados a través de un proceso POMINAR(1) en donde se generan datos aleatorios con ayuda del software R Studio Versión 4.2.1.

6.1.2 Simulación 2

Figura 20

Simulación 2

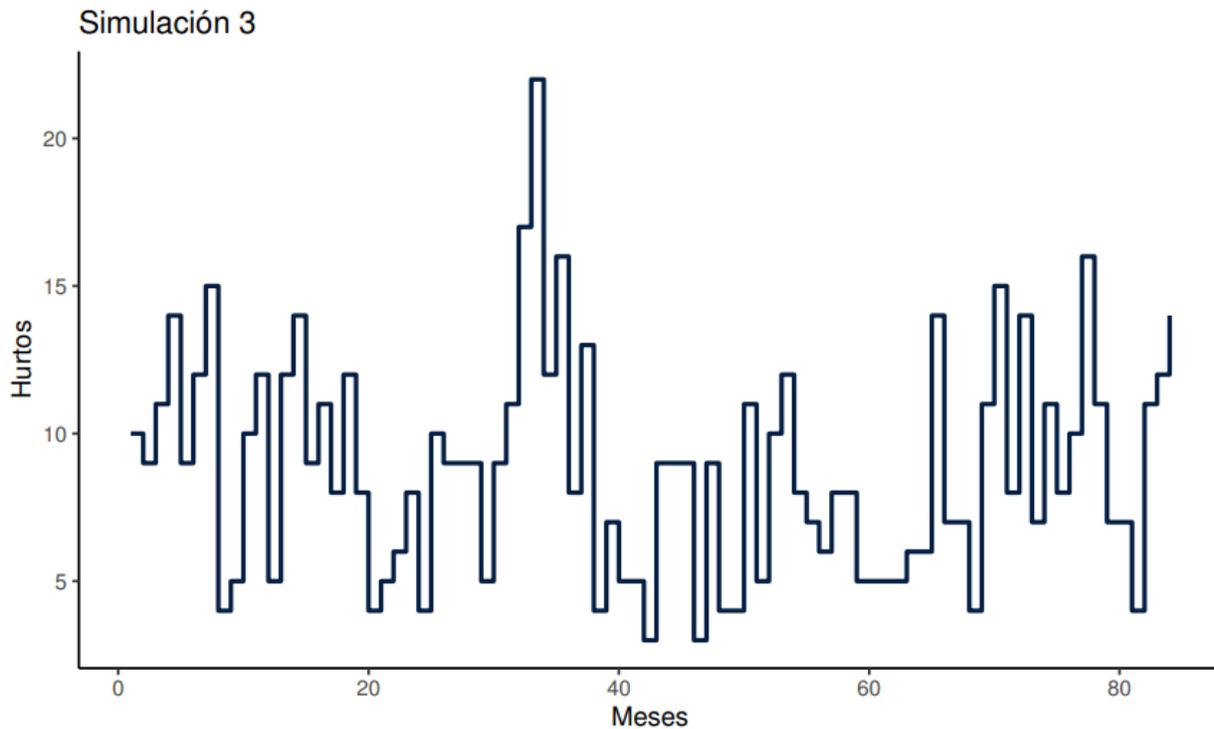


Nota. En la figura 6.2 se simularon 84 datos para el POMINAR(1) con valores de los Parámetros $\theta = 0,5$, $\delta = 0,5$, $\lambda = 4$ y $p = 0,2$ como se puede observar en el cuadro 6.1, para estos datos se obtuvo una media de 8.65 y una varianza de 12.44. Los datos fueron simulados a través de un proceso POMINAR(1) en donde se generan datos aleatorios con ayuda del software R Studio Version 4.2.1.

6.1.3 Simulación 3

Figura 21

Simulación 3



Nota. En la figura 6.3 se simularon 84 datos para el POMINAR(1) con valores de los Parámetros $\theta = 0,5$, $\delta = 0,2$, $\lambda = 6$ y $p = 0,5$ como se puede observar en el cuadro 6.1, para estos datos se obtuvo una media de 8.84 y una varianza de 14.01. Los datos fueron simulados a través de un proceso POMINAR(1) en donde se generan datos aleatorios con ayuda del software R Studio Version 4.2.1.

7. APLICACIÓN

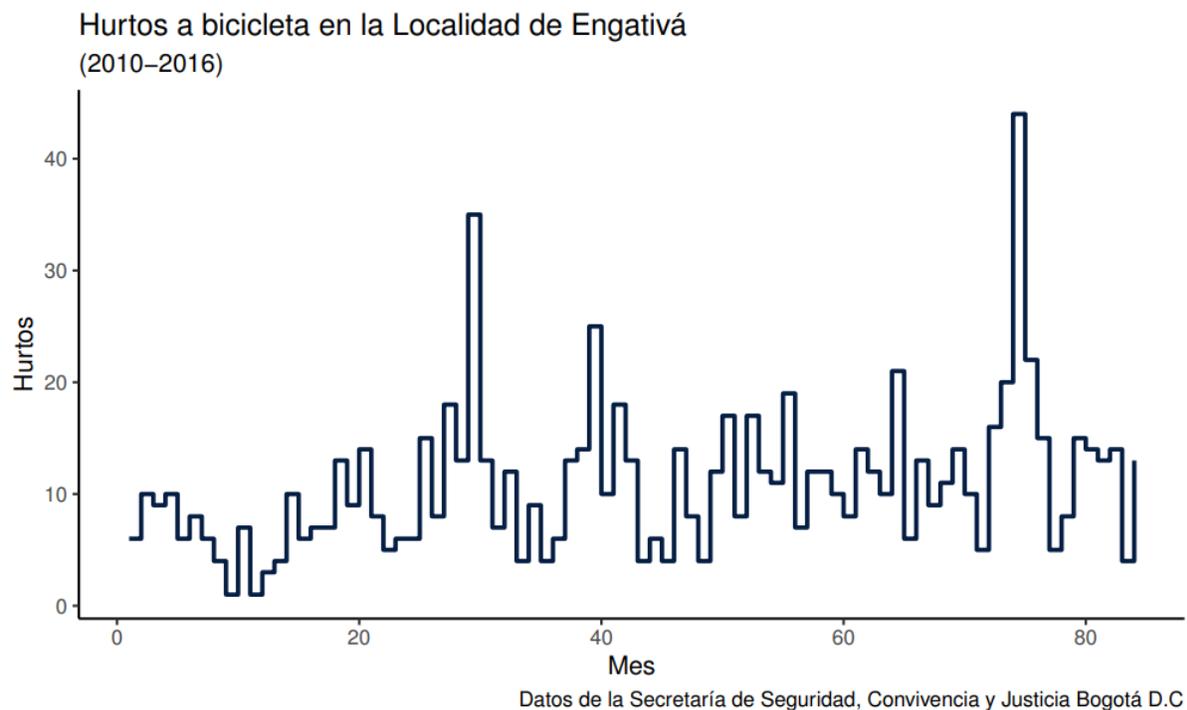
En este capítulo se presenta una aplicación de un proceso POMINAR(1) en un conjunto de datos reales para comparar con un proceso INAR(1). Los datos presentados contienen el historial de robos de bicicletas en Bogotá D.C desde el año 2010 hasta el año 2016 para la localidad de Engativá. En el siguiente cuadro se muestran las estadísticas de los datos analizados.

7.1 Serie de datos

A continuación se muestra el comportamiento de los hurtos de bicicleta en la localidad de Engativá para el periodo comprendido entre el año 2010 y 2016.

Figura 22

Hurtos de bicicleta en la Localidad de Engativá



Nota. Hurtos a bicicleta en la Localidad de Engativá (2010–2016). Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Para el periodo 2010-2016 se puede observar un comportamiento visualmente repetitivo cada 5 meses aproximadamente con algunos valores atípicos y superiores en los meses 29, 39 y 74 correspondientes a los meses de Mayo de 2012, Marzo de 2013 y Febrero de 2016 respectivamente.

Cuadro 7.1: Estadísticas Descriptivas

Figura 23

Estadísticas Descriptivas

T	Mínimo	Promedio	Máximo	Varianza
84	1	10.9	44	44.88

Nota. El cuadro muestra el número de datos T=84, el menor número de hurtos a bicicleta en un mes 1, el promedio de hurtos a bicicleta por mes 10.9 para el periodo 2010-2016, el máximo número de hurtos en un mes 44 y la varianza 44.8.

Figura 24

Datos de Hurtos a bicicletas 2010-2016 localidad de Engativá

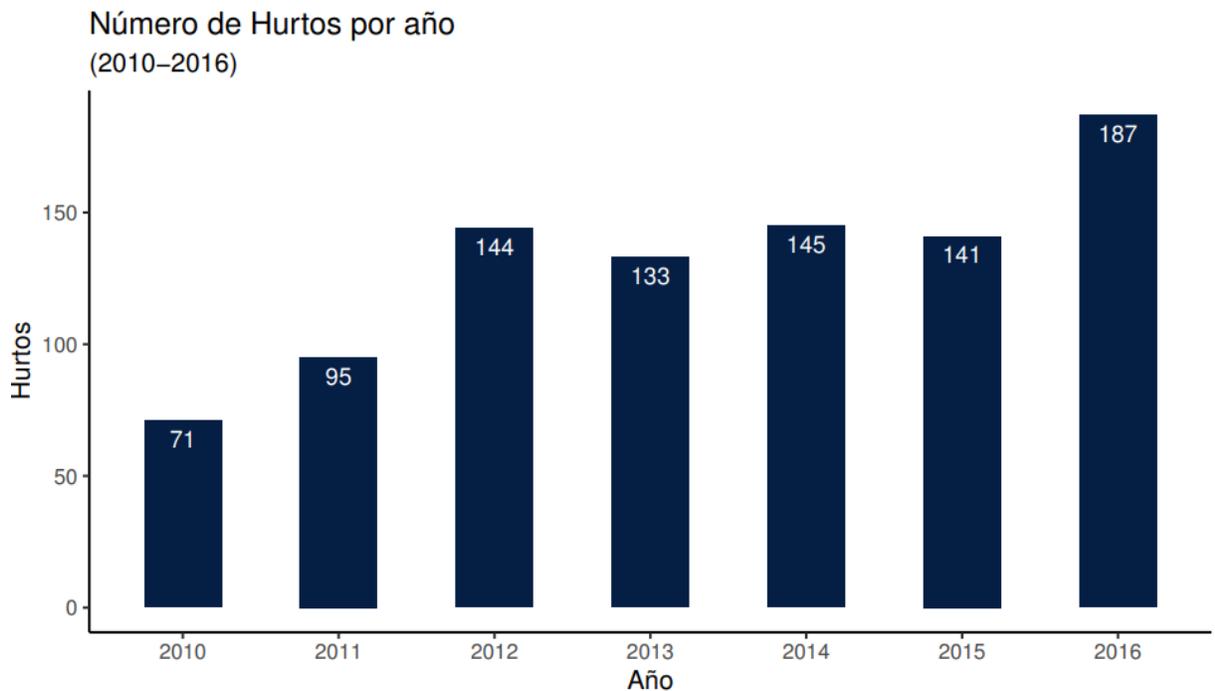
Hurtos			Hurtos			Hurtos			Hurtos		
N	Fecha	Hurtos									
1	2010M1	6	22	2011M10	5	43	2013M7	4	64	2015M4	21
2	2010M2	10	23	2011M11	6	44	2013M8	6	65	2015M5	6
3	2010M3	9	24	2011M12	6	45	2013M9	4	66	2015M6	13
4	2010M4	10	25	2012M1	15	46	2013M10	14	67	2015M7	9
5	2010M5	6	26	2012M2	8	47	2013M11	8	68	2015M8	11
6	2010M6	8	27	2012M3	18	48	2013M12	4	69	2015M9	14
7	2010M7	6	28	2012M4	13	49	2014M1	12	70	2015M10	10
8	2010M8	4	29	2012M5	35	50	2014M2	17	71	2015M11	5
9	2010M9	1	30	2012M6	13	51	2014M3	8	72	2015M12	16
10	2010M10	7	31	2012M7	7	52	2014M4	17	73	2016M1	20
11	2010M11	1	32	2012M8	12	53	2014M5	12	74	2016M2	44
12	2010M12	3	33	2012M9	4	54	2014M6	11	75	2016M3	22
13	2011M1	4	34	2012M10	9	55	2014M7	19	76	2016M4	15
14	2011M2	10	35	2012M11	4	56	2014M8	7	77	2016M5	5
15	2011M3	6	36	2012M12	6	57	2014M9	12	78	2016M6	8
16	2011M4	7	37	2013M1	13	58	2014M10	12	79	2016M7	15
17	2011M5	7	38	2013M2	14	59	2014M11	10	80	2016M8	14
18	2011M6	13	39	2013M3	25	60	2014M12	8	81	2016M9	13
19	2011M7	9	40	2013M4	10	61	2015M1	14	82	2016M10	14
20	2011M8	14	41	2013M5	18	62	2015M2	12	83	2016M11	4
21	2011M9	8	42	2013M6	13	63	2015M3	10	84	2016M12	13

Nota. El cuadro contiene los datos de hurtos a bicicleta mes a mes en la localidad de Engativá para el periodo comprendido entre el año 2010 y 2016.

7.1.1 Hurtos por año

Figura 25

Hurtos por año



Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Nota. Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Figura 26

Estadísticas Descriptivas 2010-2016

Total	Mínimo	Promedio	Máximo	Varianza
916	71	130.9	187	1416.81

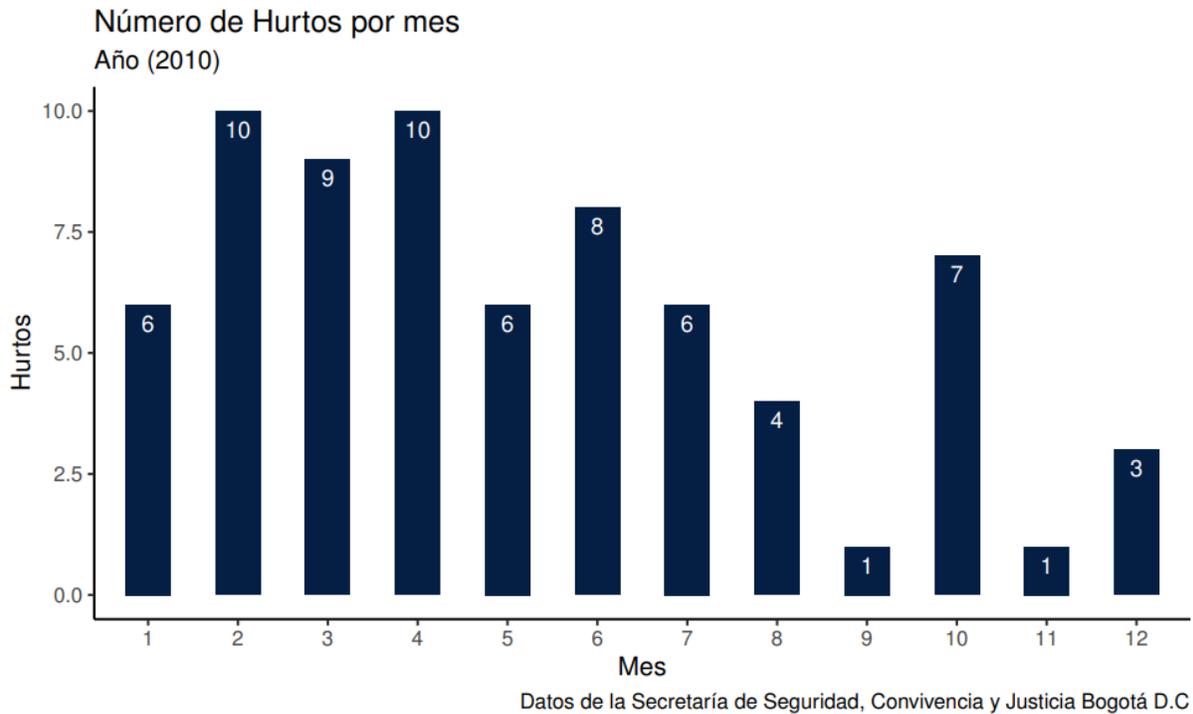
Nota. Estadísticas descriptivas

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá´ para el periodo comprendido entre el año 2010 y el año 2016, el año con mayor cantidad de hurtos fue el año 2016 con 187 casos y el menor 2010 con 71 hurtos, el promedio de hurtos por año fue de 130.9 como se muestra en el cuadro de estadísticas descriptivas

7.1.2 Hurtos por mes (2010)

Figura 27

Hurtos por mes (2010)



Nota. Número de Hurtos por mes Año (2010)

Figura 28

Estadísticas descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
71	1	5.917	10	9.90

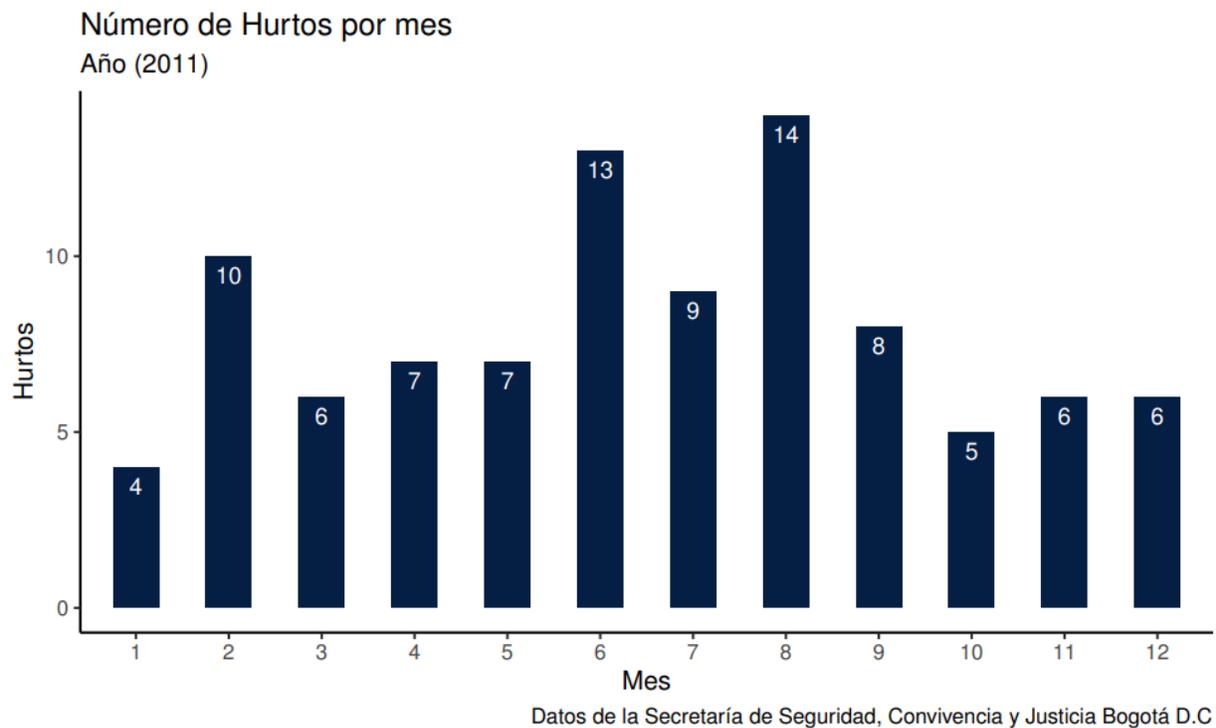
Nota. Estadísticas descriptivas

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2010 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 71 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 1 y la mayor 10 con un promedio de 5.917 hurtos como se muestra en el cuadro.

7.1.3 Hurtos por mes (2011)

Figura 29

Hurtos por mes (2011)



Nota. Número de Hurtos por mes Año (2011) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 30

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
95	4	7.917	14	9.53

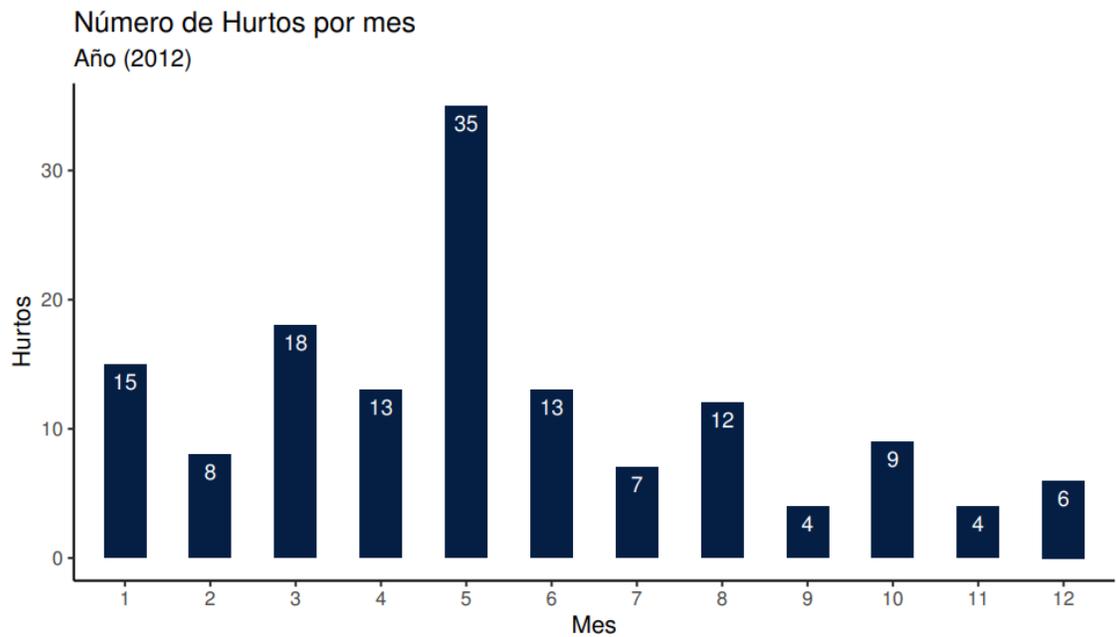
Nota. Estadísticas descriptivas

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2011 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 95 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 4 y la mayor 14 con un promedio de 7.917 hurtos como se muestra en el cuadro.

7.1.4 Hurtos por mes (2012)

Figura 31

Hurtos por mes (2012)



Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Nota. Hurtos por mes (2012) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 32

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
144	4	12.0	35	71.81

Nota. Estadísticas Descriptivas

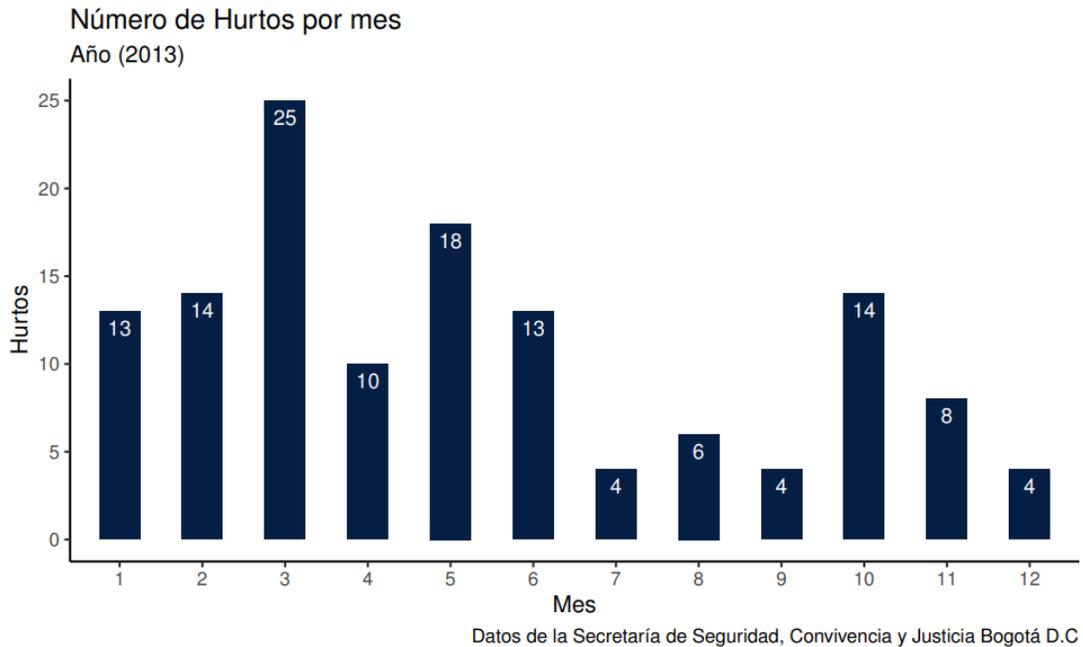
Se puede observar la cantidad de hurtos a bicicleta en la localidad de

Engativá' en el año 2012 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 144 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 4 y la mayor 35 con un promedio de 12 hurtos como se muestra en el cuadro.

7.1.5 Hurtos por mes (2013)

Figura 33

Hurtos por mes (2013)



Nota. Hurtos por mes (2013) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 34

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
133	4	11.08	25	41.17

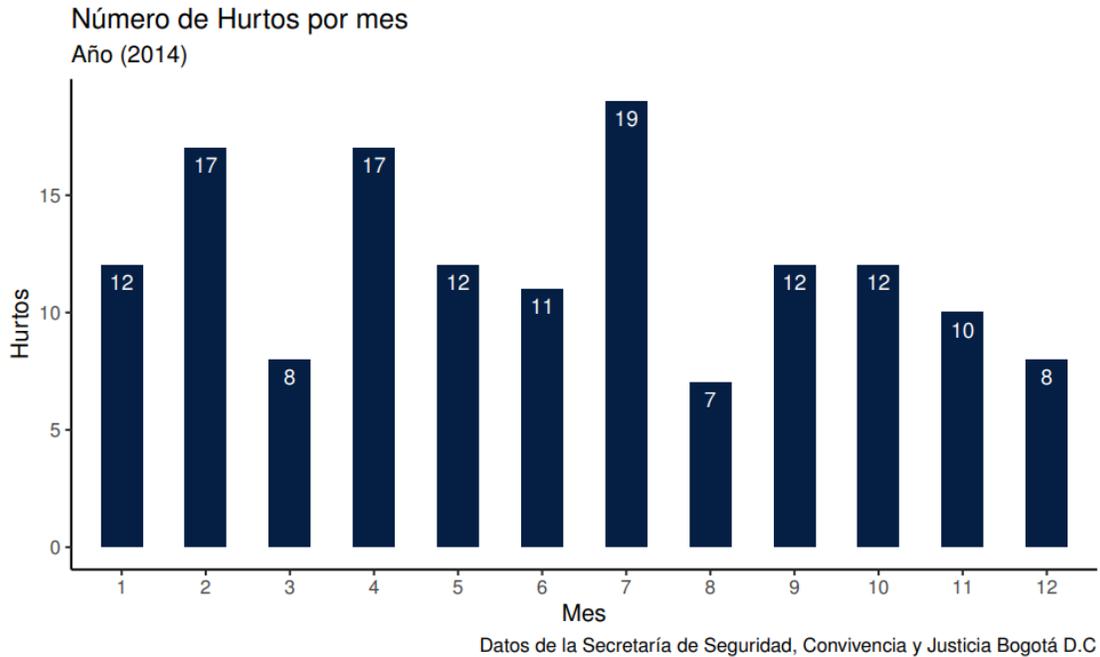
Nota. Estadísticas Descriptivas 2013

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2013 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 133 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 4 y la mayor 25 con un promedio de 11.08 hurtos como se muestra en el cuadro.

7.1.6 Hurtos por mes (2014)

Figura 35

Hurtos por mes (2014)



Nota. Hurtos por mes (2014) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 36

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
145	7	12.08	19	14.62

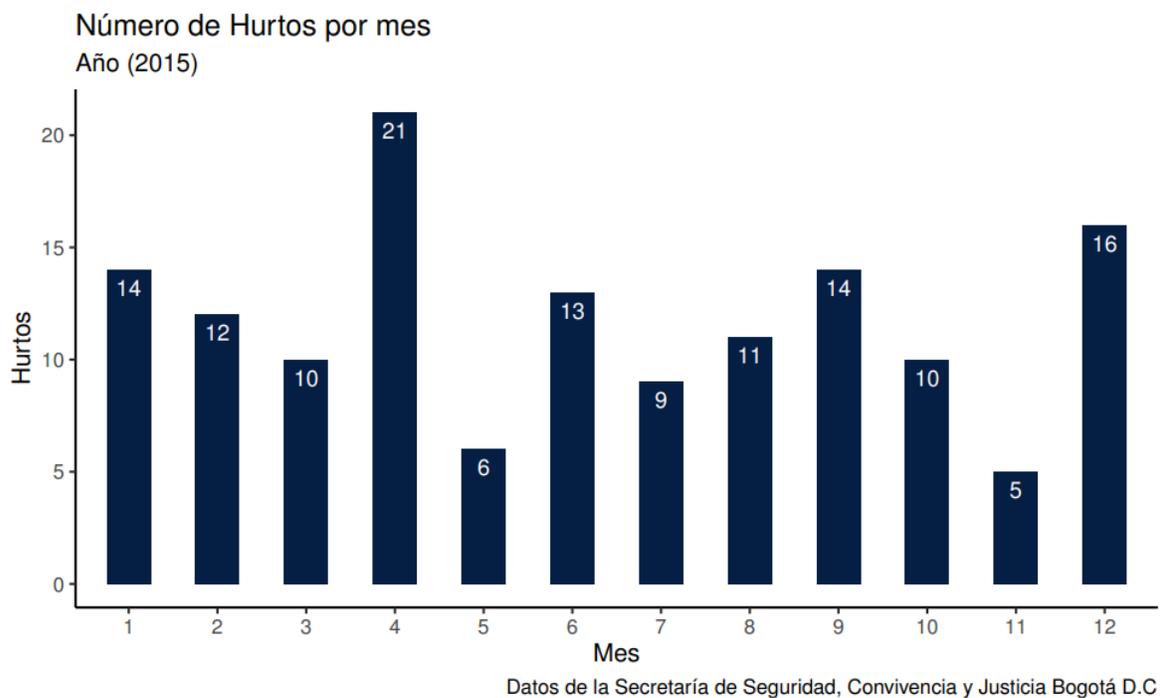
Nota. Estadísticas Descriptivas 2014

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2014 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 145 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 7 y la mayor 19 con un promedio de 12.08 hurtos como se muestra en el cuadro.

7.1.7 Hurtos por mes (2015)

Figura 37

Hurtos por mes (2015)



Nota. Hurtos por mes (2015) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 38

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
141	5	11.75	21	18.93

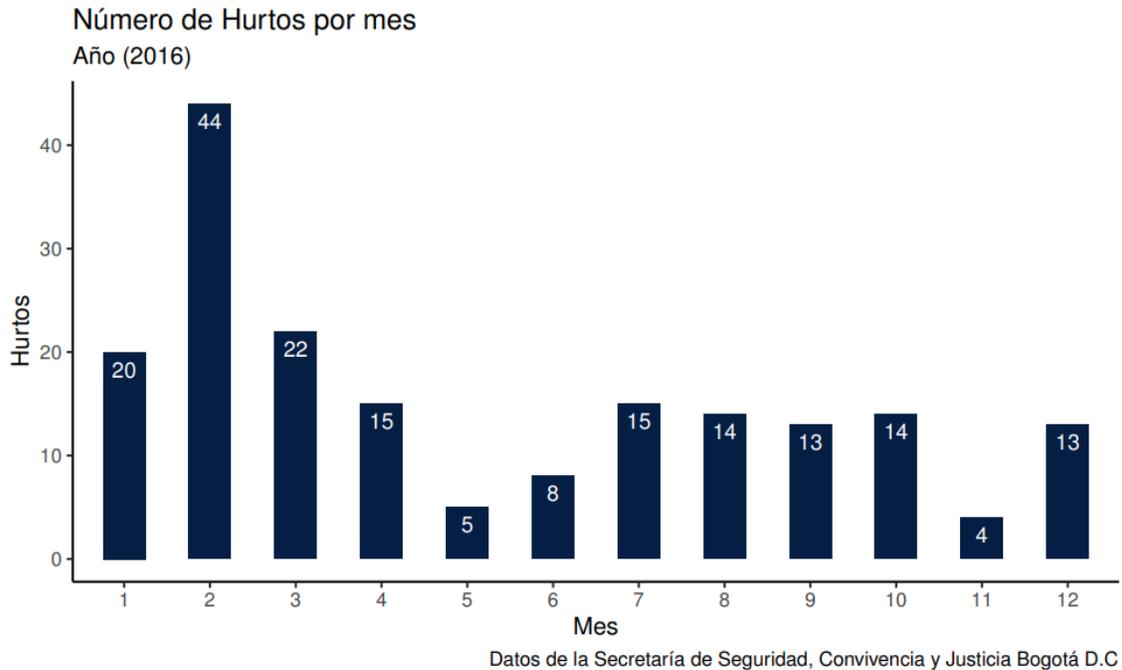
Nota. Estadísticas Descriptivas 2015

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2015 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 141 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 5 y la mayor 21 con un promedio de 11.75 hurtos como se muestra en el cuadro.

7.1.8 Hurtos por mes (2016)

Figura 39

Hurtos por mes (2016)



Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Nota. Hurtos por mes (2016) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 40

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
187	4	15.58	44	108.26

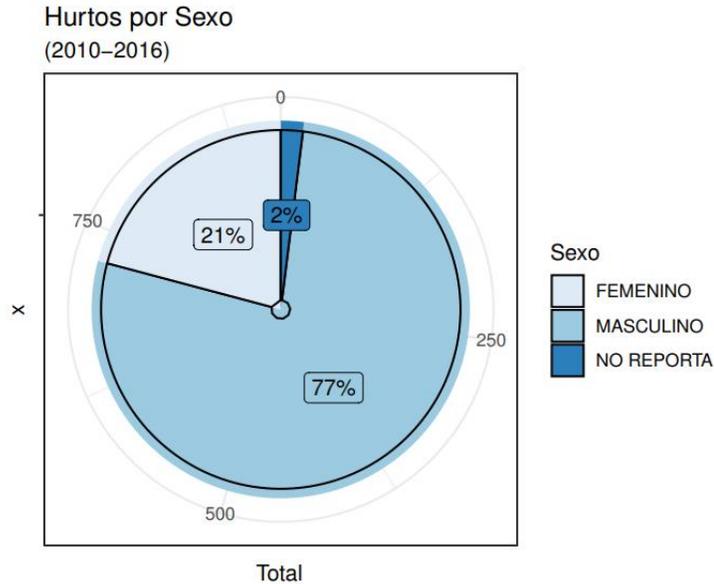
Nota. Estadísticas Descriptivas 2016

Se puede observar la cantidad de hurtos a bicicleta en la localidad de Engativá en el año 2016 para los meses de enero a diciembre, en donde Enero es el mes 1 y Diciembre es el mes 12, en total se presentaron 187 hurtos en este periodo de tiempo, la cantidad menor de hurtos en un mes fue de 4 y la mayor 44 con un promedio de 15.58 hurtos como se muestra en el cuadro.

7.1.9 Hurtos por Genero

Figura 41

Hurtos por Sexo



Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Nota. Hurtos por Sexo (2010–2016)

Figura 42

Estadísticas Descriptivas

Género	Total	Porcentaje
Masculino	707	77.18
Famenino	191	20.85
No reporta	18	0.19

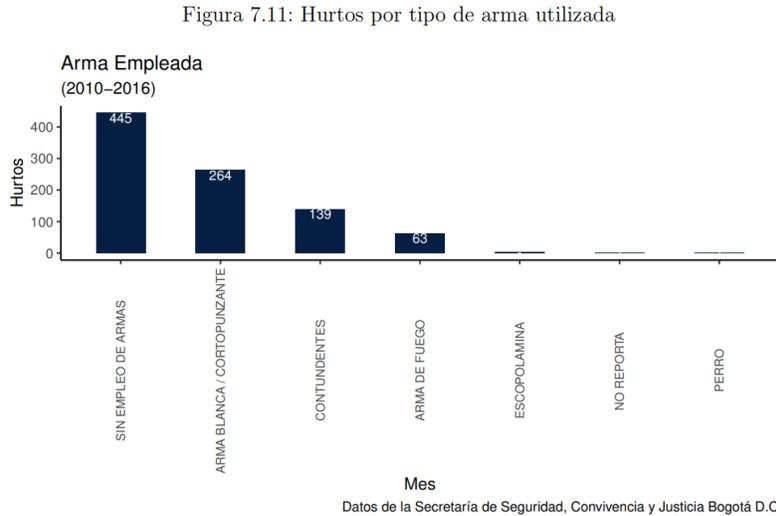
Nota. Estadísticas Descriptivas

Se puede observar el porcentaje de hurtos a bicicleta en la localidad de Engativá discriminados por género para el periodo comprendido entre el año 2010 al año 2016, en total se presentaron 916 hurtos en este periodo de tiempo, en mayor proporción las víctimas de hurto fueron hombres con un 77,18 % y 707 hurtos, las mujeres fueron víctimas en 191 ocasiones que representan el 20,85% y en 18 hurtos no se reportó el género de la víctima como se muestra en el cuadro.

7.1.10 Hurtos por tipo de arma utilizada

Figura 43

Hurtos por tipo de arma utilizada



Nota. Hurtos por tipo de arma utilizada (2010–2016) Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 44

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
916	1	130.9	445.0	28526.14

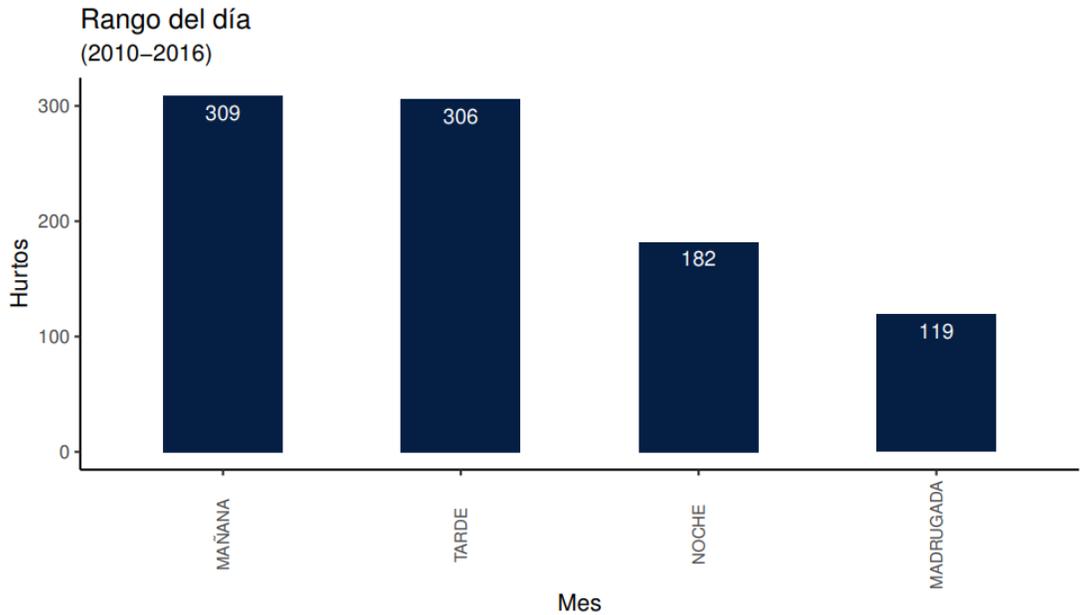
Nota. Hurtos por tipo de arma utilizada

En la figura y en el cuadro se puede observar que para los hurtos a bicicleta en la localidad de Engativá para el periodo 2010-2016 se emplearon las armas blanca o cortopunzante, contundente (cualquier tipo de instrumento usado para agredir), de fuego escopolaminas o perro y adicionalmente en algunos casos no usaron armas o no se reportaron. El arma que más se empleo fue la blanca o cortopunzante con 264 hurtos reportados con esta arma y solo 1 hurto reporta un perro utilizado como arma para cometer el hurto. De todos los hurtos, 445 fueron sin armas.

7.1.11 Hurtos por rango de día

Figura 45

Rango del día



Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C

Nota. Hurtos por rango de día Datos de la Secretaría de Seguridad, Convivencia y Justicia Bogotá D.C.

Figura 46

Estadísticas Descriptivas

Total	Mínimo	Promedio	Máximo	Varianza
916	119	229.0	309	8879.33

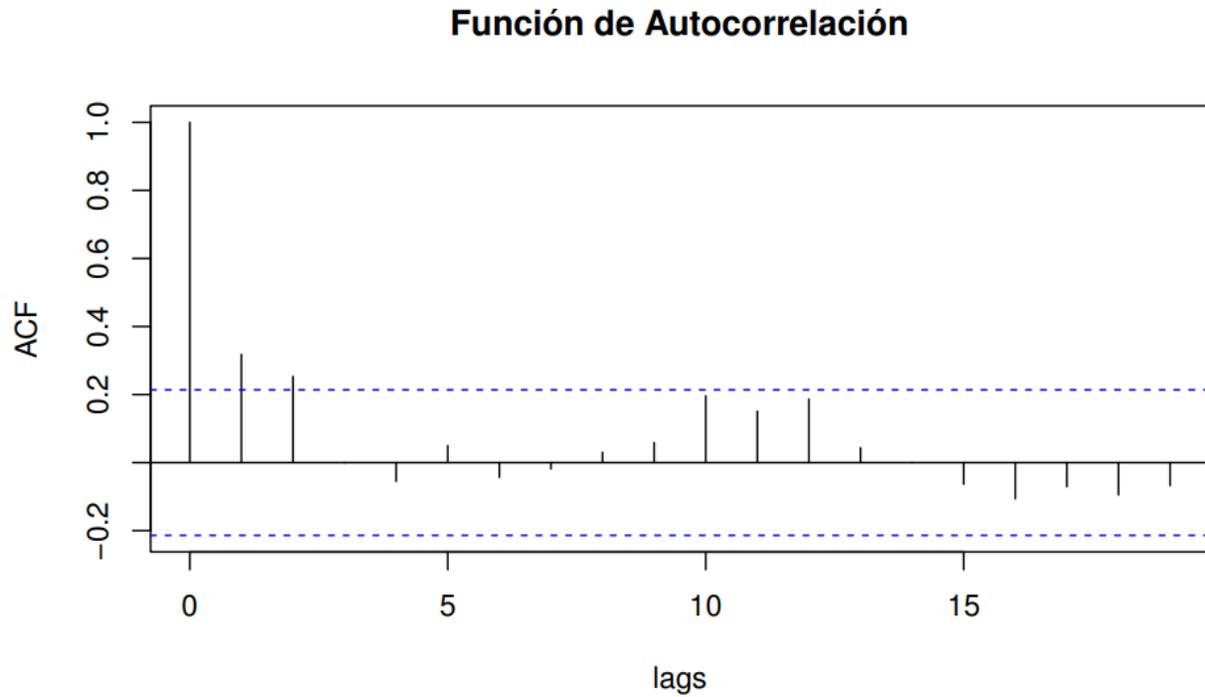
Nota Hurtos por rango del día

En la figura y en el cuadro se puede observar que los hurtos a bicicleta en la localidad de Engativá para el periodo 2010-2016 principalmente se presentaron en la mañana con 309 casos y en la tarde con 306 casos. En menor proporción los hurtos se presentaron en horas de la noche y la madrugada con 182 y 119 casos respectivamente.

7.2 Función de autocorrelación

Figura 47

Función de autocorrelación



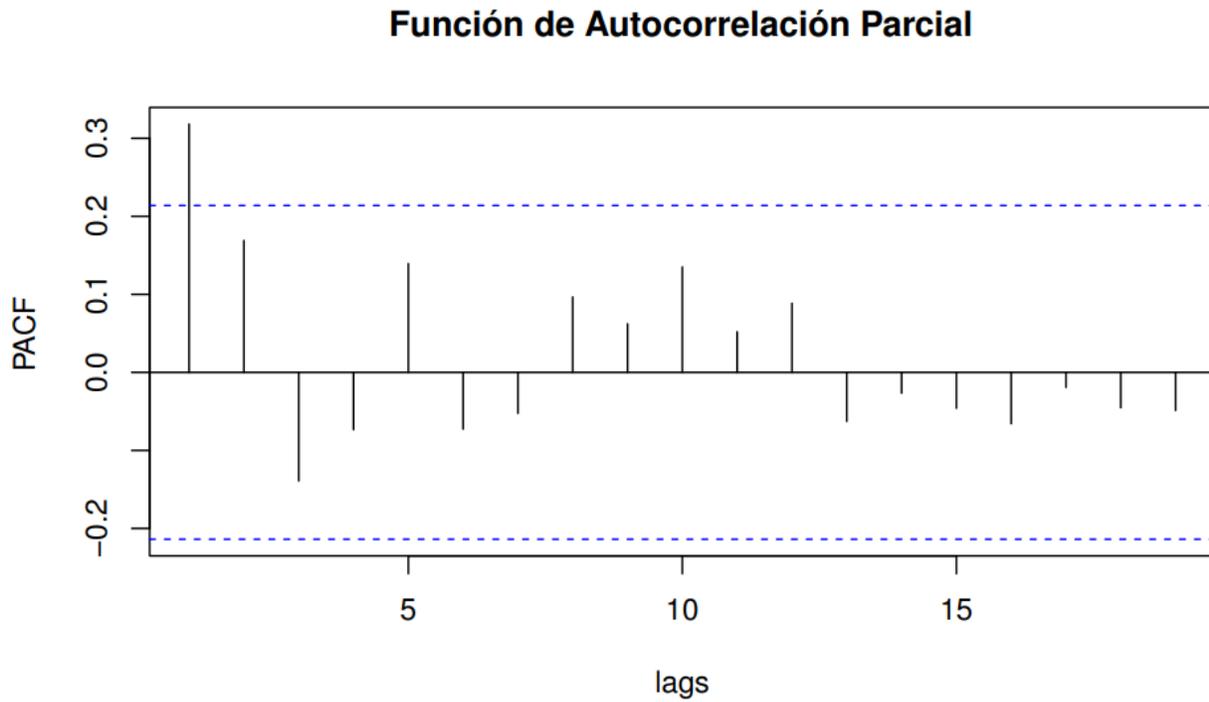
Nota. Función de autocorrelación

La figura muestra la función de autocorrelación que evidencia un corte significativo en el primer rezago que corresponde a la segunda línea, lo que indica que nuestra serie de tiempo es de orden 1 porque las observaciones dependen del valor anterior.

7.3 Función de autocorrelación parcial

Figura 48

Función de Autocorrelación Parcial



Nota. Función de autocorrelación parcial.

La figura muestra la Función de autocorrelación parcial que evidencia un corte significativo en el primer rezago que corresponde a la primera línea, lo que indica que nuestra serie de tiempo es de orden 1 porque las observaciones dependen del valor anterior.

7.4 Estimación proceso POMINAR(1)

Se procede a hacer una estimación de los parámetros con el proceso POMINAR(1) Cuadro 7.14:

Figura 49

Estimación proceso POMINAR(1)

θ	δ	λ	p
0.09	0.29	4.94	0.10

Nota. Estimación proceso POMINAR(1)

En el cuadro se muestran los valores estimados por el modelo POMINAR(1) para los parámetros $\theta = 0,09$, $\delta = 0,29$, $\lambda = 4,94$ y $p = 0,10$.

Después de estimar los parámetros del modelo, aplicamos el criterio AIC para determinar el ajuste del modelo POMINAR(1) a los datos de hurto en bicicleta para el periodo 2010-2016 en la localidad de Engativá´.

Figura 50

Criterio AIC POMINAR(1)

AIC
347.73

Nota. De acuerdo con lo definido en 5.6 el AIC para el modelo INAR es de 347.73

7.5 Estimación proceso INAR(1)

Se procede a hacer una estimación de los parámetros con el proceso INAR(1)

Figura 51

Estimación INAR(1)

θ	λ	p
0.18	8.9	1

Nota. Parámetros
estimados INAR(1)

Después de estimar los parámetros del modelo, aplicamos el criterio AIC para determinar el ajuste del modelo INAR(1) a los datos de hurto en bicicleta para el periodo 2010-2016 en la localidad de Engativá´.

Cuadro 7.17: Criterio AIC INAR(1)

Figura 52

Criterio AIC INAR(1)

AIC
620.56

Nota. De acuerdo con los
resultados el AIC para el
modelo INAR es de 620.56

Con los resultados obtenidos AIC INAR(1) de 620.56 y AIC POMINAR(1) 347.73, el modelo POMINAR(1) tiene un mejor ajuste a los datos de Hurto a bicicleta en la Localidad de Engativá´ para el periodo 2010-2016.

8. CONCLUSIONES

Después de haber hecho una revisión bibliográfica para comprender lo que es una serie de tiempo y en particular los procesos autorregresivos mixtos, se hace necesario verificar si en la aplicación a datos reales, el modelo tiene un mejor ajuste que otro modelo, en este caso se escogió el modelo INAR(1) debido a que en ausencia del operador thinning Poisson en el proceso POMINAR(1) esté quedaría reducido a un modelo INAR(1). En la selección de los datos se tuvo en cuenta las siguientes consideraciones:

Que gráficamente no se observara una tendencia clara.

Que su Función de autocorrelación fuera significativa en el rezago 1.

Que su Función de autocorrelación parcial fuera significativa en el rezago 1.

Con estas consideraciones se definió que la serie de datos seleccionada serían los Hurtos a Bicicleta en la Localidad de Engativá del Distrito Capital, una selección de 84 datos que corresponden a 7 años divididos en meses. Con las 84 observaciones seleccionadas y que cumplan con las consideraciones propuestas, se ajustó primero un modelo INAR(1) y un modelo POMINAR(1) a los datos seleccionados, encontrando que:

Gráficamente la serie no tiene una tendencia 7.2.

La función de autocorrelación es significativa al rezago 2, que corresponde a la observación inmediatamente anterior 7.13.47

La función de autocorrelación parcial es significativa al rezago 1, lo que indica una dependencia de la serie de la observación anterior 7.14.

El modelo INAR(1) obtuvo un valor en el criterio AIC de 620.56 7.16.

El modelo POMINAR(1) obtuvo un valor en el criterio AIC de 347.73 7.17.

El modelo POMINAR(1) se ajusta mejor a los datos de Hurto a bicicleta en la Localidad de Engativá para el periodo 2010-2016 que el modelo INAR(1), verificado a través del criterio AIC, el modelo INAR(1) arrojó un AIC de 620.56 y el modelo POMINAR(1) arrojó un AIC de 347.73.

Entonces con el modelo POMINAR(1) se pueden obtener unas predicciones más precisas a la hora de estimar los hurtos que van ocurrir en la localidad de Engativá en un futuro.

De acuerdo a los resultados obtenidos en este trabajo con los datos obtenidos de seguridad convivencia y justicia, (2022), también se observó que los hurtos a bicicleta se presentan principalmente en las horas de la mañana y la tarde, que en la mayoría de los hurtos no se emplea un arma o en su defecto se utiliza un arma contundente/cortopunzante y principalmente son las personas de género Masculino las víctimas del hurto, se necesita determinar en otro estudio que género hace mayor uso de la bicicleta para poder establecer si la diferencia en los hurtos, ocurre porque el género masculino es quien hace más uso de la bicicleta. Es posible que respecto a las horas en donde se presentan los hurtos, coincidan con el desplazamiento de las personas hacia sus lugares de trabajo o de regreso de sus jornadas laborales, una posible solución para mitigar los hurtos podría ser intensificar el acompañamiento de la policía en esos horarios. Para tener un análisis más completo, se podrían analizar detalladamente otras variables de interés. Por lo tanto, para fines de este trabajo de grado no se realizó dicho análisis, para un futuro trabajo de posgrado se recomienda hacer análisis a cualquier nivel de segregación como barrio, unidad de priorización zonal(UPZ), localidad o a nivel general en la ciudad de Bogotá. También, para que sean tenidas en cuenta todas las variables que acompañan a los reportes por cada hurto tales como, el rango del día en que sucedió el hurto, el arma utilizada en el hurto, el género de la persona que fue víctima de hurto y otros factores externos como por ejemplo el incremento del uso de la bicicleta para transporte después del 2019 por causa de la pandemia.

Como trabajos futuros se recomienda investigar si existen otros estudios en donde se pueda por ejemplo, verificar si el hurto al género Masculino es más alto porque los usuarios de bicicletas en su mayoría son de este género definitivamente existe una relación entre el género y el hurto. Otra sugerencia es utilizar otro tipo de modelos en series de tiempo, como también otro tipo de estimación como los estimadores de Yulie – Walker u otros.

BIBLIOGRAFÍA

- Akaike, H. (1987). Factor analysis and aic. In Selected papers of hirotugu akaike, pages 371–386. Springer. 15
- Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (inar (1)) process. *Journal of Time Series Analysis*, 8(3):261–275. 2, 16
- Bisgaard, S. and Kulahci, M. (2011). *Time series analysis and forecasting by example*. John Wiley & Sons. 7, 8, 9, 10, 11, 14
- Blanco Castañeda, L. (2013). *Probabilidad*. Editorial UN. 4
- Cavanaugh, J. E. and Neath, A. A. (2019). The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3):e1460. 3
- de seguridad convivencia y justicia, S. (2022). *Delitos de alto imp´acto*. PhD thesis. 3, 48
- Degiannakis, S. and Xekalaki, E. (2010). *ARCH models for financial applications*. John Wiley & Sons. 14
- Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued garch process. *Journal of time series analysis*, 27(6):923–942. 2, 3, 16, 19
- Fern´andez, L. M. Z. (2013). *Forecasting time series with integer values*. PhD thesis, Ph. D. thesis, Universidade Federal de Pernambuco. 20
- Guerrero, V. M. (1990). Desestacionalizaci´on de series de tiempo econ´omicas: introducci´on a la metodolog´ia. *Comercio Exterior*, 40(11):1035–1046. 2, 4, 7
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press. 4
- Mauricio, J. A. (2007). *Universidad complutense de madrid. Introducci´on al an´alisis de series temporales*. 14
- McKenzie, E. (2003). Ch. 16. discrete variate time series. *Handbook of statistics*, 21:573–606. 2

Orozco, D. L. R. (2017). Um novo processo autorregressivo misto para s eries temporais de valores inteiros de primeira ordem com inova oes poisson (pominar (1)). Master's thesis, Brasil. 1, 2, 3, 17, 21, 24, 25

Pen a, D. (2005). An alisis de series temporales. Alianza. 4, 7, 9

Ruiz, V. F. (2022). Ruido blanco. <<http://rstudio-pubs-static.s3.amazonaws.com/394452eaa513348793474b9d21505e2429bb05.html>>. 12

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. The Annals of Probability, pages 893–899. 2